

**UNA PROPUESTA PARA ELABORAR
PAQUETES DE LABORATORIO
ORIENTADOS AL PROCESO ACORDE
A LAS INSIGNIAS DE LA ACM EN
INGENIERÍA DEL SOFTWARE**

CATHY PAMELA GUEVARA VEGA

Tesis Doctoral

**Directores: Dra. Beatriz Bernárdez Jiménez y
Dr. Amador Durán Toro**



Universidad de Sevilla

septiembre 2022

First published in septiembre 2022 by

Cathy Pamela Guevara Vega

Copyright © MMXXII

<https://www.researchgate.net/profile/Cathy-Guevara-Vega>
cguevara@utn.edu.ec

This is a copyleft document but the content is copyrighted

Esta tesis doctoral ha sido parcialmente financiada por la Comisión Europea (FEDER) y por los proyectos de investigación EKIPMENT-PLUS (P18-FR-2895) del Plan Andaluz de I+D+i, MEMENTO (US-1381595) del Plan Andaluz de I+D+i, OPHELIA (RTI2018-101204-B-C22) del Plan Nacional, APOLO (US-1264651) del Plan Andaluz de I+D+i, HORATIO (RTI2018-101204-B-C21) y BELI (TIN2015-70560-R).

Dra. Beatriz Bernárdez Jiménez, Profesor Contratado Doctor del Área de Lenguajes y Sistemas Informáticos de la Universidad de Sevilla y Dr. Amador Durán Toro, Profesor Titular del Área de Lenguajes y Sistemas Informáticos de la Universidad de Sevilla,

HACEN CONSTAR

que D. Cathy Pamela Guevara Vega, Profesor Titular de la Facultad de Ingeniería en Ciencias Aplicadas de la Universidad Técnica del Norte, ha realizado bajo nuestra supervisión el trabajo titulado

Una propuesta para elaborar Paquetes de Laboratorio orientados al proceso acorde a las insignias de la ACM en Ingeniería del Software

Una vez revisado, autorizamos el comienzo de los trámites para su presentación como Tesis Doctoral al tribunal que ha de juzgarlo.

Fdo. Dra. Beatriz Bernárdez Jiménez y Dr. Amador Durán Toro
en la Universidad de Sevilla

09/09/2022



AGRADECIMIENTOS

Agradezco principalmente a Dios por bendecirme cada día y darme la sabiduría para finalizar este trabajo de tesis.

Durante el tiempo de dedicación a este trabajo, he recibido la guía, el conocimiento y colaboración de personas a quienes agradezco profundamente.

A mi tutor de tesis. Antonio Ruiz-Cortés, por darme la oportunidad de formar parte del grupo ISA y acogerme en su equipo de trabajo. Gracias Antonio por ser la primera persona que confió en mí para iniciar la presente investigación.

A mis directores de tesis. Bea, por guiarme, dedicar su tiempo durante todo el proceso de esta investigación. Ha sido muchos años de trabajo juntas, quien con su valioso conocimiento en la experimentación me ayudó a cumplir los objetivos planteados. Amador, por sus acertadas observaciones, consejos y soluciones oportunas, especialmente por su amplio conocimiento en la parte de modelado. Gracias Antonio, Bea, Amador por haber hecho posible la realización de esta tesis.

A los integrantes del grupo ISA, en especial a Margarita Cruz y José Antonio Parejo quienes han dedicado su tiempo para contestarme dudas acerca de las replicaciones y el proceso experimental.

Destacar mi agradecimiento a la Universidad Técnica del Norte por brindarme las facilidades económicas y logísticas para culminar la presente tesis doctoral, a través de sus autoridades en especial el apoyo del Dr. Miguel Naranjo Toro.

A mi familia, en especial a mi esposo Antonio por haber tomado el reto de seguir juntos este camino y mi hija Antonella por su infinito amor. A mi madre Cecilia y mis hermanos Alex y Jorge, por apoyarme en la dedicación a este trabajo. Gracias familia.

RESUMEN

En esta memoria, se presenta el trabajo de investigación llevado a cabo para fomentar las replicaciones de experimentos controlados con humanos mediante el desarrollo de una propuesta para la elaboración de un Paquete de Laboratorio orientado al proceso acorde a las insignias de la ACM en Ingeniería del Software.

La importancia de elaborar un Paquete de Laboratorio orientado al proceso ha sido motivado por tres aspectos. Primero la necesidad del propio investigador en tener una guía de cómo construir y estructurar el Paquete de Laboratorio de un experimento original. Segundo la importancia de integrar la elaboración del Paquete de Laboratorio con el proceso experimental. Tercero facilitar una guía paso a paso al experimentador para obtener una insignia de la ACM con respecto a la evaluación del Paquete de Laboratorio.

Como parte de la información identificada en el proceso de elaboración del paquete, se ha propuesto una especificación del proceso experimental de forma integrada con el Paquete de Laboratorio, la creación de un modelo de calidad para paquetes de laboratorio, un metamodelo que visualice de manera conceptual la estructura del contenido del Paquete de Laboratorio y la asignación de insignias de la ACM.

Basado en el metamodelo, se presenta la herramienta badgeGo que permite a los experimentadores automatizar el proceso de elaboración del Paquete de Laboratorio orientado al proceso e indique con alta probabilidad conseguir las insignias de la ACM.

Un estudio de caso enfocado en el área de Ingeniería del Software ha sido instanciado para validar la propuesta.

ABSTRACT

In this dissertation, we present the research work carried out to promote the replications of controlled experiments with humans by developing a proposal to elaborate a process-oriented Laboratory Package following the ACM insignia in Software Engineering.

The importance of developing a process-oriented Lab Package has been motivated by three aspects. First, the researcher needs to have a guide on how to build and structure the lab package of an original experiment. Second, the importance of integrating the elaboration of the Laboratory Package with the experimental process. Third to guide step-by-step the experimenter on achieving an ACM badge concerning the evaluation of the Lab Package.

As part of the information identified in the package development process, a specification of the experimental process in an integrated way with the lab package, the creation of a quality model for Lab Packages, a metamodel that conceptually visualizes the structure of the Lab Package content, and the assignment of ACM badges have been proposed.

Based on the metamodel, the badgeGo tool is presented that allows experimenters to automate the process-oriented Lab Package development process and achieve ACM badges with high probability.

A case study focused on the Software Engineering area has been instantiated to validate the proposal.

ÍNDICE GENERAL

I	Introducción	1
1.	Planteamiento del problema	3
1.1.	Motivación	4
1.2.	Relevancia del problema	7
1.3.	Objetivos de la investigación	13
1.4.	Metodología de investigación	15
1.5.	Resumen de la Contribución	16
1.6.	Contexto de la Tesis	20
1.7.	Estructura del documento	21
2.	Background	23
2.1.	Introducción	24
2.2.	Terminología	24
2.2.1.	Proceso experimental	24
2.2.2.	Tipos de replicación de experimentos	26
2.2.3.	Familia de experimentos	27
2.2.4.	Repetición, Reproducción y Replicación	28
2.2.5.	Sujetos y objetos experimentales	29
2.2.6.	Comunicación entre investigadores	29

2.2.7.	Paquete de Laboratorio	30
2.2.8.	Tipología de los Paquetes de Laboratorio	31
2.3.	Contenido de los Paquetes de Laboratorio	32
2.4.	Otras Propuestas de Paquetes de Laboratorio	34
2.5.	Open Science	37
2.5.1.	Transparencia en Open Science	38
2.6.	Amenazas a la validez en experimentación	39
2.6.1.	Relación entre validez interna y validez externa	40
2.7.	Resumen	40
II	Estado de la cuestión	43
3.	Estado actual de Paquetes de Laboratorio en IS	45
3.1.	Introducción	46
3.2.	¿Es una práctica habitual en la comunidad IS hacer públicos los PL? . . .	50
3.3.	¿Cuál es la tendencia en la publicación de PL tras el Open Science? . . .	53
3.4.	¿Cuál es el nivel de calidad de PL respecto a las insignias ACM?	55
3.5.	¿Cuál es el nivel de calidad de PL respecto a los módulos de Solari? . . .	57
3.6.	¿Qué guías utilizan los autores para reportar sus PL?	59
3.7.	¿Qué repositorios digitales se usan para que los PL estén disponibles? .	61
3.8.	Conclusiones	62
3.9.	Resumen	64
4.	Problemática en la elaboración de Paquetes de Laboratorio	67
4.1.	Introducción	68
4.2.	Poco uso de guías para reportar Paquetes de Laboratorio	69

- 4.3. Carencia de Paquetes de Laboratorio 71
- 4.4. El conocimiento tácito 72
- 4.5. Diferente terminología 73
- 4.6. Necesidad de clarificar el contexto de las replicaciones 74
- 4.7. El factor humano en las replicaciones 75
- 4.8. Resumen 76

- III Propuesta 79**

- 5. Paquete de Laboratorio orientado al proceso acorde a la ACM 81**

 - 5.1. Introducción 82
 - 5.2. Especificación del proceso experimental de forma integrada con PL 82
 - 5.3. Procedimiento para elaborar un PL acorde a la ACM 87
 - 5.3.1. Procedimiento para elaborar un PL durante la ejecución del experimento 88
 - 5.4. Modelo de calidad para Paquete de Laboratorio acorde a la ACM 91
 - 5.4.1. Introducción al modelo de calidad para Paquetes de Laboratorio acorde a la ACM 91
 - 5.4.2. Definición de los artefactos que forman parte del Paquete de Laboratorio acorde a la ACM 92
 - 5.4.3. Definición de las Características del Modelo de Calidad 94
 - 5.4.4. Características de calidad para la insignia Funcional 97
 - 5.4.5. Características de calidad para la insignia Reutilizable 109
 - 5.4.6. Características de calidad para la insignia Disponible 121
 - 5.4.7. Característica de calidad para la insignia Reproducido 124
 - 5.4.8. Característica de calidad para la insignia Replicado 132

5.5.	Estructura básica de un PL para tener un nivel de calidad.	140
5.6.	Resumen	143
6.	Automatización del proceso mediante badgeGo	145
6.1.	Introducción	146
6.2.	Requisitos	146
6.3.	Metamodelo de Paquete de Laboratorio acorde a la ACM	147
6.3.1.	Metamodelo de la especificación del proceso experimental de forma integrada con el Paquete de Laboratorio	147
6.3.2.	Metamodelo del modelo de calidad acorde a la ACM	150
6.4.	Aplicabilidad del Paquete de Laboratorio orientado al proceso	153
6.5.	Una Herramienta para evaluar Paquetes de Laboratorio	154
6.5.1.	Vista del registro de los experimentos	155
6.5.2.	Vista del registro de los experimentadores	155
6.5.3.	Vista del registro de los grupos y participantes	156
6.5.4.	Vista del registro de las tarea del experimento y sus artefactos . .	156
6.5.5.	Validación de las insignias de la ACM	156
6.5.6.	Vista del registro del Paquete de Laboratorio	159
6.6.	Resumen	159
IV	Validación	161
7.	Evaluación de la propuesta en un estudio de caso	163
7.1.	Introducción	165
7.2.	Estudio de caso	165
7.2.1.	Diseño del estudio de caso	166

- 7.2.2. Planificación, recopilación y análisis 168
- 7.2.3. Análisis e informe conjunto 170
- 7.3. Evaluación de la Viabilidad para la propuesta 171
 - 7.3.1. Definición de los atributos de Viabilidad 171
 - 7.3.2. Caso-APIs 171
 - 7.3.3. Evolución del metamodelo a partir de la instanciación del Paquete de Laboratorio 172
- 7.4. Evaluación de la Flexibilidad para la propuesta 176
 - 7.4.1. Definición de los atributos de Flexibilidad 177
 - 7.4.2. Caso-Mind 178
 - 7.4.3. Caso-Req 179
- 7.5. Evaluación de la Satisfacción para la propuesta 180
 - 7.5.1. Definición de los atributos de Satisfacción 180
- 7.6. Resultados del estudio de caso 182
 - 7.6.1. Resultados de la evaluación de Viabilidad 182
 - 7.6.2. Verificación de Completitud del Paquete de Laboratorio en el Caso-APIs 182
 - 7.6.3. Resultados de la evaluación de Flexibilidad 183
 - 7.6.4. Verificación de Completitud del Paquete de Laboratorio en el Caso-Mind y Caso-Req 183
 - 7.6.5. Resultados de la evaluación de Satisfacción 184
- 7.7. Conclusiones 185
- 7.8. Resumen 186

V Conclusiones	189
8. Conclusiones y trabajos futuros	191
8.1. Consecución de objetivos	192
8.2. Publicaciones	195
8.2.1. Revistas internacionales	195
8.2.2. Congresos internacionales	196
8.2.3. Participación en proyecto de investigación	196
8.3. Trabajo futuro	197
VI Anexos	199
A. Descripción del reporte NISO RP-31—2021	201
B. Instanciación la propuesta en el Caso-APIs	203
C. Instanciación de la propuesta en el Caso-Mind	221
D. Instanciación de la propuesta en el Caso-Req	235
E. Cuestionario de Satisfacción del experimentador	243
Bibliografía	245

ÍNDICE DE FIGURAS

1.1. Fases del proceso de DSR extraído de [400]	15
1.2. Resumen de contribuciones	18
3.1. Proceso de selección para los estudios primarios en revistas y paquetes de laboratorio en conferencias.	47
3.2. Artículos científicos que definen un Paquete de Laboratorio por revistas.	50
3.3. Paquete de Laboratorio por conferencias.	51
3.4. Artículos científicos que profundizan el contenido del Paquete de Laboratorio por revistas.	53
3.5. Evolución de la publicación de Paquetes de Laboratorio desde 2000 hasta 2021 por revista.	54
3.6. Insignias de la ACM para Paquetes de Laboratorio.	57
3.7. Insignias de la ACM para paquetes de laboratorio por conferencias.	57
3.8. Contenido del Paquete de Laboratorio clasificado en módulos propuestos por Solari <i>et al.</i> [394].	58
3.9. Repositorios digitales que contienen el Paquete de Laboratorio publicados en revistas y congresos	61
5.1. Fases y actividades del proceso experimental	83
5.2. Actividades de la fase de operación en el proceso experimental [435]	84
5.3. Integración de las tareas del experimento con el Paquete de Laboratorio	87
5.4. Procedimiento para elaborar un Paquete de Laboratorio acorde a la ACM	90
5.5. Ejemplo de la integración entre las tareas del experimento y los artefactos	93

6.1. Metamodelo de la especificación del proceso experimental de forma integrada con el Paquete de Laboratorio	148
6.2. Metamodelo del modelo de calidad acorde a la ACM	151
6.3. Pantalla de inicio	155
6.4. Vista de los experimentos registrados	156
6.5. Vista del registro de los experimentadores	156
6.6. Vista del registro de los grupos experimentales	157
6.7. Vista del registro de las tareas del experimento y sus artefactos	157
6.8. Vista de la validación de las insignias de la ACM para el Paquete de Laboratorio	158
6.9. Vista del registro del Paquete de Laboratorio	158
6.10. Vista del Paquete de Laboratorio en un archivo comprimido	159
7.1. Fases del estudio de caso basado en [355]	166
7.2. Paquete de Laboratorio de la familia de experimentos Mindfulness	167
7.3. Actualización de la clase "Persona" en el metamodelo	173
7.4. Actualización de la clase "Grupo" en el metamodelo	174
7.5. Actualización de la clase "Persona" en el metamodelo	175
7.6. Actualización en "Tipo_Artefacto" y "Tipo_Finalidad" en el metamodelo	177
7.7. Actualización en la clase "Artefacto" en el metamodelo	178

ÍNDICE DE TABLAS

1.1. Insignias definidas por la ACM para evaluar artefactos	10
1.2. Actividades de la metodología DSR y secciones correspondientes	17
2.1. Fases del proceso experimental en IS	25
2.2. Comparativa entre propuestas de paquetes de laboratorio para experi- mentos en Ingeniería del Software	36
3.1. Estudios primarios publicados en revistas y congresos en 2000–2021 or- denados por número de estudios.	52
3.2. ACM Badging para Paquetes de Laboratorio por revistas.	56
3.3. Guías para Paquetes de Laboratorio y Open Science en revistas y confe- rencias seleccionadas para este estudio.	60
5.1. Características de calidad para Paquetes de Laboratorio acorde a la ACM	96
5.2. Comparativa entre propuestas con respecto a los niveles de profundi- dad de Solari <i>et al.</i> [392] con respecto al contenido de los paquetes de laboratorio para experimentos en Ingeniería del Software	100
5.3. Modelo de calidad para Paquetes de Laboratorio (PL) acorde a la ACM - Artefactos Evaluados - Funcional	106
5.4. Artefactos opcionales del modelo de calidad para las insignias Funcional y Reutilizable	110
5.5. Modelo de calidad para Paquetes de Laboratorio (PL) acorde a la ACM - Artefactos Evaluados - Reutilizable	117
5.6. Modelo de calidad para Paquetes de Laboratorio (PL) acorde a la ACM - Artefactos Evaluados - Disponible	123

5.7. Características opcionales del modelo de calidad de Artefactos Disponibles	124
5.8. Modelo de calidad para Paquetes de Laboratorio (PL) acorde a la ACM - Resultados Validados - Reproducido	129
5.9. Características opcionales del modelo de calidad de los Resultados Validados - Reproducido	132
5.10. Modelo de calidad para Paquetes de Laboratorio (PL) acorde a la ACM - Resultados Validados - Replicado	137
5.11. Características opcionales del modelo de calidad de los Resultados Validados - Replicado	140
5.12. Estructura de PL para artefactos evaluados y disponibles relacionado con las insignias de la ACM.	141
5.13. Estructura de PL para resultados validados relacionado a las insignias de la ACM.	142
7.1. Actualización de la información del experimentador en la estructura de la primera instanciación del Caso-APIs	173
7.2. Actualización de la información de Grupos en la estructura de la primera instanciación del Caso-APIs	174
7.3. Actualización de la información de la tarea en la estructura de la primera instanciación del Caso-APIs	175
7.4. Actualización de la información de los artefactos en la estructura de la primera instanciación del Caso-APIs	176
7.5. Métricas de la calidad en uso para los atributos de Satisfacción	181
7.6. Resultado de los atributos de <i>Viabilidad</i> de la primera instanciación en el Caso-APIs	182
7.7. Resultado del cumplimiento del Paquete de Laboratorio para alcanzar las insignias de la ACM en el Caso-APIs	183
7.8. Resultado de los aspectos de flexibilidad de la instanciación en el Caso-Mind y Caso-Req	183

7.9. Resultado del cumplimiento del Paquete de Laboratorio para alcanzar las insignias de la ACM en el Caso-Mind y el Caso-Req 184

7.10. Resultados de los atributos de Satisfacción 185

A.1. Insignias definidas por la NISO para evaluar artefactos 202

A.2. Insignias definidas por la ACM para evaluar artefactos 202

B.1. Instanciación de la propuesta en el experimento Caso-APIs 204

C.1. Instanciación de la propuesta en el experimento Caso-Mind 221

D.1. Instanciación de la propuesta en el experimento Caso-Req 235

PARTE I

INTRODUCCIÓN

PLANTEAMIENTO DEL PROBLEMA

El comienzo es la parte más importante del recorrido

*Platón (427 a. C.– 347 a. C.),
Filósofo griego*

El primer capítulo, aborda el problema que se define en la presente memoria de tesis, justificando su relevancia, y las principales razones que han motivado su investigación. En primer lugar, en la sección §1.1 se presenta la motivación y planteamiento del problema. En la sección §1.2 se identifican algunas iniciativas, que destacan la importancia en la disponibilidad y evaluación de los Paquetes de Laboratorio experimentales, para facilitar las replicaciones, junto con autores que han resaltado la necesidad de documentar los artefactos experimentales. En la sección §1.3 se presenta el objetivo principal de este trabajo de investigación, y los objetivos parciales que permiten cumplir con dicho objetivo principal.

En la sección §1.4 se describe el método de investigación llevado a cabo. En la sección §1.5 se resumen las aportaciones como resultado de los objetivos planteados. En la sección §1.6 se enumeran los proyectos de investigación en los que se enmarca el presente trabajo. Finalmente, la sección §1.7 describe la estructura del resto de la memoria de tesis.

1.1 MOTIVACIÓN

La *Ingeniería del Software (IS)*, analiza las técnicas sistemáticas y económicas para el desarrollo del software. Estas técnicas ayudan a desarrollar software utilizando un enfoque de ingeniería. Al igual que cualquier otra disciplina de ingeniería, la IS hace un uso intensivo de los conocimientos acumulados a partir de las experiencias de profesionales. En casos cuando la teoría existente no es suficiente las experiencias pasadas se adoptan como regla general. Todas las soluciones científicas se construyen mediante la aplicación rigurosa de principios demostrables [259].

Basili *et al.* en 1991, consideraron que el proceso de IS debe desarrollarse de manera que los procesos, métodos y tecnologías sean delineados de forma clara para la construcción y evaluación del software, a este proceso lo definieron como la naturaleza experimental del desarrollo del software [44]. La experimentación contribuye a la maduración del cuerpo del conocimiento de la IS.

La *Ingeniería del Software Empírica (ISE)* permite la evaluación del producto del desarrollo del software [45]. Su evaluación comparativa a través de la experimentación en un entorno controlado permite manipular y observar la realidad en el comportamiento de un fenómeno. Durante el experimento se confirma y se refuta afirmaciones sobre técnicas y herramientas para la construcción del software [201]. Para generalizar los resultados obtenidos en un experimento es necesario replicarlo en diferentes contextos y condiciones [38]. La importancia de las repeticiones recae en la evolución del conocimiento científico experimental tanto en las ciencias exactas [94] como en las ciencias sociales [430].

En la actualidad, la comunidad de ISE ha considerado relevante la realización de familias de experimentos como mecanismo para evaluar métodos, técnicas y herramientas que permita conocer la conveniencia de utilizarlos en el proceso de desarrollo de software [389]. De inicio, Basili *et al.* en 1999 [46], manifiestan que una familia de experimentos consiste en un experimento original, seguido de un conjunto de repeticiones que responden a las mismas preguntas de investigación que el experimento original.

Posteriormente, Santos *et al.* [363] propusieron las siguientes premisas para considerar una serie de experimentos como una familia de experimentos: *i)* el acceso a los datos brutos está garantizado; *ii)* los investigadores conocen la configuración exacta de cada experimento y *iii)* al menos tres experimentos evalúan los efectos de al menos dos

tecnologías diferentes sobre la misma variable de respuesta. Por otro lado, los conocimientos obtenidos a partir de una familia de experimentos son más sólidos y fiables que los obtenidos a partir de un único experimento aislado, cuyos resultados pueden considerarse preliminares [381, 397].

Para Carver [86] un investigador novel que decide replicar un experimento original necesita revisar cuidadosamente toda la familia de estudios anteriores, a fin de adquirir los conocimientos para adaptar el experimento adecuadamente, mejorar su diseño o evitar cometer los mismos errores de experimentos anteriores.

En el mismo sentido, Carver *et al.* [87] señalan que la replicación de estudios empíricos en IS sigue sin atraer la suficiente atención de los investigadores y Kitchenham *et al.* [223] destacan la falta de incentivos y la percepción en los investigadores para llevar a cabo replications a pesar de su importancia como método científico.

Para solventar esta situación, han surgido varias iniciativas como la Ciencia Abierta traducida al Inglés como *Open Science* [268]. La Ciencia Abierta promulga que poner a disposición los conjuntos de datos, los análisis y una versión preimpresa de un experimento y su software relacionado, proporciona un valioso conocimiento para el experimentador, además cualquier parte interesada puede auditarlo. Esto da lugar a, que otros investigadores a través de la reutilización de los artefactos bien elaborados y documentados construyan sobre el trabajo anterior nuevo conocimiento que beneficie a la comunidad científica en IS [185].

Un enfoque complementario para aumentar la visibilidad y reproducibilidad de los estudios empíricos específicamente en el contexto de los experimentos controlados es promover la disponibilidad de los Paquetes de Laboratorio (PL) [46, 393]. Según Shull *et al.* [381] los artefactos que componen el PL deberían incluir no sólo el conjunto de datos, los análisis y el material experimental, sino también las guías para llevar a cabo una replicación y un resumen de la evolución del experimento en toda la familia [394]. La replicación es una tarea compleja por naturaleza y más aún cuando los paquetes de laboratorio no son lo suficientemente completos y fáciles de comprender [120]. A pesar de todos estos esfuerzos, Shepperd *et al.* [376] afirmaron que, es necesario reportar tanto los estudios originales como los replicados para consolidar los hallazgos de los estudios empíricos.

Para Timperley *et al.* [406] en la práctica, los artefactos sufren una variedad de problemas que impiden la realización de todo su potencial. En su estudio aplican una encuesta a 153 investigadores en IS durante el 2019 referente a la creación, publicación

y revisión de artefactos en estudios de IS. Concluyen que existe una serie de desafíos de alto nivel que afectan la creación, el intercambio, el uso y la revisión de artefactos. Estos desafíos incluyen, entre otros, la percepción de que el esfuerzo para crear y compartir artefactos no vale la pena, la falta de estándares y pautas comunitarias sobre los artefactos y cómo deben revisarse, y la necesidad de que los investigadores proporcionen un mantenimiento continuo en los PL.

Otros problemas asociados que destaca Timperley *et al.* [406] son una comunicación inadecuada, intereses de sistemas sociales como el empoderamiento, miedo a que los experimentos sean evaluados y los resultados se comprometan, poca recompensa por el esfuerzo que conlleva crear, publicar y evaluar los artefactos, efectos del tiempo, la facilidad de uso (usabilidad) entre otros.

En este contexto, la motivación para investigar los PL de experimentos controlados con humanos, específicamente, la manera como se reportan, su estructura, disponibilidad y calidad, surge como consecuencia de la práctica de construir PL de varios experimentos y sus posteriores replicaciones llevados a cabo por investigadores expertos en ISE que forman parte de nuestro grupo de investigación. Existen directrices para elaborar PL [21, 41, 194, 221, 381, 382, 393, 394, 416] sin embargo, no está claro la cantidad y calidad de los artefactos experimentales, tampoco cómo construirlos para alcanzar un paquete de calidad [392]. La información heterogénea y variada taxonomía para elaborar PL nos han generado un desgaste y demora en los tiempos de ejecución. En nuestra opinión, en parte, el problema se debe a que cuando se quiere construir el PL por lo general el experimento original ya se ha ejecutado, en consecuencia, el experimentador empieza la elaboración del PL, a posterior, incluso cuando ha finalizado de escribir el artículo y se siente agotado de tanto esfuerzo realizado.

Considerando los problemas expuestos anteriormente, surge tres preguntas que quizás engloban la dificultad de construir un PL de calidad. La primera pregunta se enfoca en si los artefactos que componen el material experimental serán suficientes para replicar el experimento. Una segunda pregunta que ha surgido es cómo definir la calidad de los mismos, de qué depende el proceso de su evaluación para que sean afines a ciertos estándares. La tercera pregunta que se plantea es, por tanto, cómo facilitar el proceso de elaboración del PL de un experimento original y posteriormente de una replicación.

Puede que sea suficiente poner a disposición de la comunidad científica ciertos artefactos que los experimentadores consideren suficientes para su replicación pero se ha pensado que llevar a cabo las actividades y tareas con tiempos de ejecución del

proceso experimental y que vaya de la mano al mismo tiempo con la integración de los artefactos permitirá solventar la elaboración *útil* y *ágil* del PL. En cualquier caso, la presente investigación quiere destacar la cantidad, el contenido y calidad de los artefactos que se publican en un PL. Sin una debida atención en la elaboración del paquete seguramente dificultará la ejecución de replicaciones posteriores.

En resumen, elaborar, documentar y publicar los artefactos de un experimento es vital, ya que permite transmitir la información del proceso experimental para fortalecer y apoyar a la replicación. Dada la complejidad del conocimiento experimental especialmente en el área de la IS que relativamente es joven con relación a otras ciencias maduras como la Medicina y la Psicología, si no existe una adecuada comunicación para que el replicador adquiera de alguna manera los elementos conceptuales y operacionales que fueron aplicados en el experimento original la tarea de replicar se convierte en un enorme reto [177].

Se ha pensado que la utilidad, la confianza y la comodidad en la elaboración de un PL experimental durante la ejecución del propio experimento aumenta: la comprensión de la replicación, la mejora de la calidad del PL y puede contribuir en el incremento del número de replicaciones.

1.2 RELEVANCIA DEL PROBLEMA

Para justificar la importancia de elaborar, documentar, evaluar y publicar los PL experimentales, a continuación, se presenta algunas organizaciones y específicamente investigadores que se han interesado por el tema en cuestión y sus aspectos relacionados.

Como se ha comentado en la sección §1.1, el movimiento *Open Science* [268] es una oportunidad para transparentar y evidenciar el trabajo científico. Permite que la comunidad conozca y haga uso del conocimiento científico de cualquier área sin restricciones. Eso da lugar a mejorar la información que se publica y en este contexto a perfeccionar los PL y facilitar la reproducibilidad y replicabilidad de los estudios empíricos [168].

Open Science está basada en tres aspectos: i) *Open Access*: acceso libre a los artículos, ii) *Open Data*: datos disponibles y iii) *Open Source Software*: software de código abierto. Los PL que contiene el material utilizado en el experimento, son un ejemplo de compartición de datos para facilitar las replicaciones y que están dentro del segundo

aspecto.

La publicación de los PL se realiza generalmente en repositorios en la nube, sitios web personales e institucionales que ayudan a la gestión de almacenar y compartir artefactos como GitHub y herramientas *Open Science* como Zenodo, OSF, Figshare. El uso de herramientas *Open Science* ayuda a solventar el acceso de los artefactos y la permanencia de alojamiento, un problema frecuente que presentan algunas investigaciones [269]. El Instituto de Ingenieros Eléctricos y Electrónicos (IEEE) incentiva el movimiento *Open Science* mediante algunas opciones como: *i*) Code Ocean [76] es una plataforma en la nube donde se comparte código, se ingresa datos para experimentar, reproducir y desarrollar las investigaciones con el fin de aumentar la visibilidad y el impacto de los estudios y publicaciones científicas. *ii*) IEEE Data-Port [112] es una plataforma de datos en la nube de fácil acceso que permite a los usuarios almacenar, buscar, acceder y administrar datos científicos, permite la colaboración entre investigadores en el análisis de datos y apoya la investigación reproducible. Y *iii*) TechRxiv [77] es un servidor de preimpresión abierto y moderado para investigaciones inéditas en las áreas de Ingeniería, Ciencias de la Computación y Tecnología relacionada. Los autores pueden difundir rápidamente su trabajo a una amplia audiencia y obtener comentarios de la comunidad sobre una versión preliminar de su investigación con marca de tiempo.

La *Sociedad Científica y Educativa en el Campo de la Computación ACM* (Association for Computing Machinery) [185] motiva a los investigadores a evaluar y validar los artefactos de investigación tanto para revistas como congresos a través de sus “Directrices y procedimientos de publicación” mediante un apartado dedicado a clasificar por insignias (Badges ACM) los estudios experimentales en *repetibles*, *reproducibles* y *replicables*; y a los artefactos en *artefactos evaluados*, *artefactos disponibles* y *resultados validados*.

Para la ACM un *artefacto* es un *objeto digital* creado por los autores de una investigación para ser utilizado como parte del estudio o generado por el experimento en sí, este puede ser un *script*, *dataset*, *formularios*, *vídeos*, *audio*, etc [185]. La relevancia de evaluar los artefactos de investigación es debido a un número importante de fallos como métodos experimentales descuidados, análisis estadísticos defectuosos o enlaces de internet no disponibles [185]. Es necesario realizar un proceso de arbitraje y auditar formalmente los artefactos para evitar errores en la reproducción y replicación de forma independiente.

Otros organismos como Sandia National Laboratories [178] propone un reporte acerca de una taxonomía de reproducibilidad compatible para las Ciencias de la Compu-

tación y la Informática, NISO RP-31-2021 Reproducibility Badging and Definitions [295] propone un reporte para Insignias y Definiciones de reproducibilidad. Para comprender de mejor manera el reporte de la NISO se hace una breve descripción en el Anexo §A.

Algunos congresos relevantes en el área de IS como ICSE'2021 (International Conference on Software Engineering), ESEM'2021 (International Symposium on Empirical Software Engineering and Measurement), RE'2021 (International Requirements Engineering Conference), ESEC/FSE'2021 (European Software Engineering Conference and Symposium on the Foundations of Software Engineering), ASE'2021 (IEEE/ACM International Conference on Automated Software Engineering) han generado un Track AE-Artifact Evaluation para evaluar los artefactos experimentales y asignar insignias de la ACM de los trabajos aceptados para incentivar a otros investigadores la probabilidad de reproducción y replicación de los resultados de las investigaciones.

La revista EMSE'2021 (Empirical Software Engineering) motiva y destaca el movimiento *Open Science* para hacer públicos los datos de investigaciones y aumentar la transparencia, la reproducibilidad y la replicabilidad de los estudios. Esto apoya la credibilidad inmediata del trabajo de los autores, y proporciona una base común para los esfuerzos conjuntos de la comunidad basados en datos compartidos. Además, construye un cuerpo de conocimiento general en la comunidad que conduzca a teorías de IS para su aplicabilidad [269].

A continuación, se muestra la clasificación de insignias de la ACM para la evaluación de los artefactos, a pesar que la terminología no es uniforme los términos están inspirados en el Vocabulario Internacional para Metrología (International Vocabulary for Metrology-VIM) [298]. En la tabla §A.2 aparecen las insignias de la ACM.

- *Artefactos evaluados*. Son los que han superado con éxito una auditoría independiente. Se distinguen dos niveles:
 - *Funcionales*. Los artefactos están documentados, son coherentes, completos, aplicables e incluyen pruebas apropiadas de verificación y validación.

Para comprender de mejor manera las definiciones de verificación y validación, se ha considerado explicar una definición estándar de la IEEE y la definición propuesta por el investigador Barry Boehm. Con respecto a la IEEE [187]: La *verificación*, es el acto de revisar, inspeccionar, probar, comprobar, auditar o establecer y documentar de otra manera si los elementos,

Tabla 1.1: Insignias definidas por la ACM para evaluar artefactos

Artefactos Evaluados		Artefactos Disponibles	Resultados Validados	
Funcional	Reutilizable	Disponible	Reproducido	Replicado
				
Artefactos documentados, consistentes, completos, ejercitables e incluyen evidencia apropiada de verificación y validación.	Funcional + Los artefactos asociados al documento son de una calidad que excede significativamente la funcionalidad mínima. Es decir, tienen todas las cualidades del nivel de Artefactos Evaluados-Funcionales, pero, además, están muy cuidadosamente documentados y bien estructurados hasta el punto de facilitar su reutilización y adaptación. En particular, se respetan estrictamente las normas y estándares de la comunidad investigadora para este tipo de artefactos.	Los artefactos creados por el autor que son relevantes para el artículo se han colocado en un repositorio de archivos de acceso público de forma permanente a través de un repositorio con una URL estable y un DOI (Digital Object Identifier).	Los principales resultados del artículo se han obtenido en un estudio posterior realizado por una persona o equipo distinto de los autores, utilizando, en parte, artefactos proporcionados por el autor.	Los principales resultados del artículo han sido obtenidos de forma independiente en un estudio posterior por una persona o equipo distinto de los autores, sin utilizar los artefactos suministrados por éstos.

procesos, servicios o documentos se ajustan a los requisitos especificados. La *validación*, es la evaluación del software al final del proceso del desarrollo del software para garantizar el cumplimiento de los requisitos del usuario. La validación es, por tanto, una verificación "de principio a fin".

Con respecto a Boehm [58]: La *verificación*, se relaciona con la pregunta: ¿Estamos construyendo el artefacto correctamente?. Mientras, que la validación se relaciona con la pregunta: ¿Estamos construyendo el artefacto correcto?. La diferencia entre ambas es que la verificación comprueba que las *especificaciones* de requisitos en la construcción del artefacto son implementadas correctamente, mientras que, la validación se refiere a las *necesidades de los usuarios*. La verificación comienza antes de la validación y luego se ejecuta en paralelo hasta que se pone en producción el artefacto [59]. Hay que considerar que la validación no sólo tiene que ver con los atributos de calidad,

como la funcionalidad y la fiabilidad, sino que también incluye los requisitos no funcionales.

- *Reutilizables*. Además de ser funcionales, están cuidadosamente documentados y bien estructurados facilitando la reutilización.
- *Artefactos disponibles*. Son los que además de ser funcionales son de acceso público y por tanto están *disponibles* de forma permanente a través de un repositorio con una URL estable y un DOI (Digital Object Identifier).
- *Resultados validados*. Se distinguen dos niveles:
 - *Reproducidos*. Experimentadores diferentes del equipo original han sido capaces de obtener los mismos resultados con la misma configuración experimental es decir, utilizando en parte, artefactos proporcionados por el autor.
 - *Replicados*. Experimentadores diferentes del equipo original han sido capaces de obtener los mismos resultados con diferente configuración experimental es decir, sin el uso de artefactos proporcionados por el autor.

La ACM considera que un resultado experimental no está plenamente establecido a menos que pueda ser *reproducido* de forma independiente. Por otra parte, los beneficios obtenidos al poner los *artefactos de la investigación* bien formados y documentados a disposición del público para que otros investigadores construyan directamente sobre el trabajo anterior mediante la reutilización dará lugar a generalizar los resultados de experimentos originales.

Además de las organizaciones presentadas anteriormente, diferentes investigadores destacan la importancia de crear, evaluar y publicar PL para experimentos en IS y disponer de directrices que mejoren la calidad y contenido del material experimental.

- Para Basili *et al.* [41] señalan la importancia de elaborar PL experimentales y el contenido que deben tener. Así como, la necesidad de directrices y metodologías para mejorar el proceso de elaboración del material experimental.
- En Shull *et al.* [382] discuten cómo las deficiencias de los documentos en los PL son uno de los puntos débiles dentro del área de IS que dificultan el uso de la replicación para avanzar en nuestro conocimiento. Su propuesta se enfoca en el uso de mecanismos de intercambio de conocimientos entre los investigadores.

- En [21] Almqvist considera que la cantidad de material disponible en los PL varía considerablemente. Un resultado de su estudio indica que varios investigadores que han realizado replications externas han señalado que los materiales proporcionados no eran suficientes para permitir una replicación cercana. Por lo que propone el contenido básico que debería tener un PL para su uso, que se detalla en la sección §2.3.
- Kitchenham *et al.* [221] proponen un checklist de contenido definido desde la perspectiva del replicador de experimentos para reportar estudios en IS empírica, destacando la necesidad de documentar claramente el material experimental.
- Para poder replicar un experimento, también hay que transmitir la información generada en el experimento original o su replicación así también lo manifiestan Vegas *et al.* [416] donde abordan la influencia del tipo de comunicación que existe entre los experimentadores en el éxito de una replicación.
- En Shull *et al.* [381] destacan la importancia de elaborar una documentación adecuada para un experimento (original o replicación) con el fin de realizar una nueva replicación. Concluyen que una replicación bien documentada proporciona los detalles necesarios para que el replicador siga los procedimientos establecidos.
- En [194] Jedlitschka *et al.* destacan que los materiales no deben presentarse textualmente, sino que se describan con todo el detalle necesario para que los replicadores comprendan su aplicación e interacción con los participantes durante el experimento.
- Con relación a la estructura y contenido de un PL, Solari *et al.* proponen una clasificación [393], una plantilla de módulos y un checklist [394] de instrucciones para el replicador. Además, ponen en manifiesto la necesidad de evaluar el contenido, su calidad y automatizar el proceso de desarrollo de un paquete experimental. Esta es una de las propuestas más inspiradora y completa que se ha encontrado a pesar de tener un número bajo de citas por parte de la comunidad científica, y que su propuesta no sea usada por los investigadores al momento de construir el PL.
- En [268] Mendez *et al.* reflexionan sobre los aspectos esenciales de *Open Science* para la IS, incluyendo conceptos, importancia e incentivos hacia la participación de la comunidad científica con el fin de establecer una base común y contribuir a que *Open Science* sea una norma también en la IS.

- Abrahão y Mendez [6] proponen una guía para envío y revisión de artefactos relacionados a investigaciones en IS utilizada específicamente en el Track AE-Artifact Evaluation del congreso ICSE'2021.
- Graziotin [168] propone políticas de *Open Science* e instrucciones para los revisores de congresos y revistas de investigación de la comunidad de IS y todas las disciplinas de la informática bajo la iniciativa de *Open Science ACM SIGSOFT*¹. Entre las políticas se describe el acceso abierto de los datos, artefactos y licencias para apoyar a la recuperabilidad, reproducibilidad y replicabilidad de los estudios.

A nuestro modo de ver, el hecho de que existan congresos, algunos de ellos específicos para evaluar los artefactos de estudios empíricos en IS, revela la importancia del problema para la comunidad investigadora. Dichos congresos y revistas, se identifican en la sección §3.2 donde se analiza el estado actual de los paquetes de laboratorio y la evaluación de sus artefactos.

1.3 OBJETIVOS DE LA INVESTIGACIÓN

En el presente trabajo se analiza el proceso de elaboración de un *Paquete de Laboratorio orientado al proceso acorde a la ACM* de experimentos controlados con humanos en el sentido de que, el PL es fruto de la definición de tareas que comprenden la ejecución del experimento. Es decir, el experimentador especifica qué tareas se realizan durante la ejecución y análisis del experimento y qué artefactos de entrada o salida participan en cada tarea, dando forma así al PL. La caracterización de cada tarea y cada artefacto debe permitirnos: *a)* por un lado, desarrollar un PL estructurado como un inventario de artefactos del experimento, *b)* generar una guía de experimentación fácil de usar para un investigador que desee realizar una replicación y *c)* predecir con una alta probabilidad que insignia de la ACM se otorgaría al PL fruto del proceso.

El objetivo principal de esta tesis es **fomentar las replications de experimentos controlados con humanos mediante la elaboración de un Paquete de Laboratorio orientado al proceso acorde a la ACM.**

Con este objetivo, se proponen las siguientes preguntas de investigación (PI):

¹<https://www.sigsoft.org>

- **PI₁**: ¿Cuál es el estado actual de las propuestas para la elaboración de los Paquetes de Laboratorio? (capítulo §3).
- **PI₂**: ¿Cuáles son los principales problemas identificados en la práctica de la elaboración de los Paquetes de Laboratorio? (capítulo §4).
- **PI₃**: ¿Cómo se están reportando actualmente los Paquetes de Laboratorio en la comunidad de IS? (capítulo §3).
- **PI₄**: ¿Se puede definir los artefactos suficientes para otorgar insignias de calidad al Paquete de Laboratorio? ¿Se puede definir un modelo para elaborar el Paquete de Laboratorio? (capítulo §5).

A su vez, la pregunta de investigación **PI₁** se puede concretar en las siguientes preguntas:

- **PI_{1.1}**: ¿Es una práctica habitual en la comunidad de IS hacer públicos los Paquetes de Laboratorio?
- **PI_{1.2}**: ¿Cuál es la tendencia en la publicación de los Paquetes de Laboratorio tras la iniciativa *Open Science*?
- **PI_{1.3}**: ¿Cuál es el nivel de calidad de los Paquetes de Laboratorio en relación con las insignias de la ACM?

Del mismo modo, la pregunta de investigación **PI₃** se puede concretar en las siguientes preguntas:

- **PI_{3.1}**: ¿Cuál es el nivel de calidad de los Paquetes de Laboratorio con respecto a la completitud de los módulos del estudio de Solari *et al.* [394]?
- **PI_{3.2}**: ¿Qué guías utilizan los autores para reportar sus Paquetes de Laboratorio?
- **PI_{3.3}**: ¿Qué repositorios digitales se suelen utilizar para que los Paquetes de Laboratorio estén disponibles en la comunidad de IS?

1.4 METODOLOGÍA DE INVESTIGACIÓN

La metodología aplicada en la presente investigación es *Ciencia del diseño* (DSR, por sus siglas en inglés de *Design Science Research*).

DSR crea y evalúa los criterios de valor y utilidad de los productos de software para resolver los problemas organizacionales identificados [423]. A continuación, se definen las siguientes fases de DSR propuestas en [316]. La figura 1.1 muestra las fases de DSR.

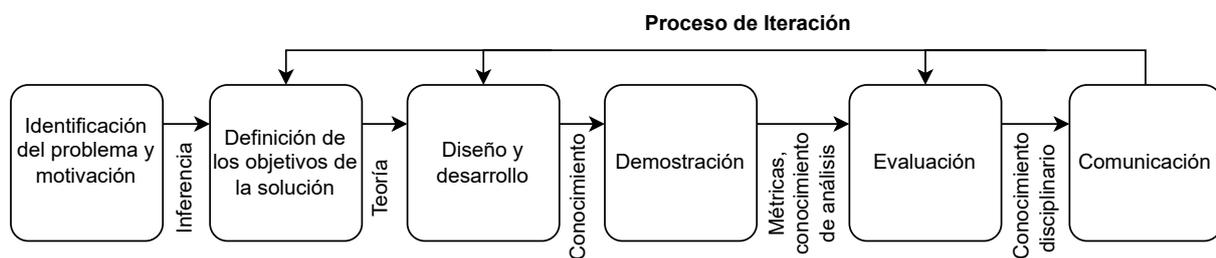


Figura 1.1: Fases del proceso de DSR extraído de [400]

1. *Identificación del problema y motivación.* Se define el problema que da lugar a una propuesta de investigación.

En nuestro caso, se analizan las dificultades encontradas en el contenido y estructura de los PL y en concreto su calidad, esto se debe a que se usan poco las directrices propuestas en la comunidad, y no se integran en la ejecución del experimento cuando el investigador está centrado en ello, es decir, en el momento adecuado. Esto provoca un desgaste para el investigador que desea elaborar el PL luego de finalizar el experimento.

2. *Definición de los objetivos de la solución.* Es necesario conocer otras propuestas y especificar qué se espera de la solución a elaborar.

Para ello, se revisa la literatura para: *i)* analizar cómo se están reportando los PL de experimentos controlados con humanos en el área de IS; *ii)* identificar la información involucrada; y, *iii)* conocer otras propuestas.

3. *Diseño y desarrollo.* Se da forma a la solución. Se concibe el PL orientado al proceso acorde a la ACM como una secuencia de tareas que el experimentador y los participantes llevan a cabo. Durante dichas tareas se utilizan artefactos de diferente tipo que el experimentador irá identificando y caracterizando. Se define la

especificación del proceso experimental de forma integrada con el PL. Se define el *procedimiento* para elaborar un PL acorde a la ACM. Se define un *modelo de calidad* para evaluar el grupo de artefactos que cumplen las políticas de evaluación para alcanzar las insignias de la ACM. Para expresar la información conceptualizada se define un *metamodelo* usando un diagrama de clases UML. Para automatizar el proceso de elaboración de un PL acorde a la ACM se presenta la *herramienta* badgeGo.

4. *Demostración*. Implica el uso de la solución desarrollada en una o más instancias del problema analizando el contexto de su aplicabilidad.
5. *Evaluación*. Mediante una prueba de concepto, se analiza si el uso de la solución cubre los objetivos esperados. Puede ser necesario realizar ajustes en la solución y volver, mediante un proceso iterativo, a la actividad de *diseño y desarrollo*.
6. *Comunicación*. Se publican los resultados de la investigación.

La tabla §1.2 muestra las actividades resumidas en las fases de la metodología DSR y su correspondencia con las secciones de esta memoria.

1.5 RESUMEN DE LA CONTRIBUCIÓN

La figura §1.2, explica de manera resumida las aportaciones de este trabajo.

A continuación, se describe las contribuciones relacionadas a las fases de DSR.

Para la fase de *Definición de los objetivos de la solución* se estable:

- *El análisis del estado de elaboración de los Paquetes de Laboratorio*. Por la afinidad en el contexto que existe con los experimentos controlados con humanos en IS se ha considerado estudiar como base la investigación realizada por Revoredo *et al.* [343] y posteriormente una actualización hasta el año 2021 llevada a cabo en esta memoria de tesis, se estudiaron 268 artículos que reportan experimentos controlados con humanos en el área de IS. Se analizaron aspectos relacionados con el PL como el uso de guías, disponibilidad y accesibilidad del material experimental, el contenido y su estructura, identificando características y elementos que forman parte del contenido del PL, además tendencias actuales como vacíos de investigación del tema en cuestión.

Tabla 1.2: Actividades de la metodología DSR y secciones correspondientes

Actividad	Descripción	Sección
<i>Identificación del problema y motivación</i>	Necesidad de abordar el proceso de elaboración de PL de experimentos controlados con humanos.	§1.1, §1.2
<i>Definición de los objetivos de la solución</i>	<i>i)</i> Integrar el proceso experimental con el desarrollo del PL. <i>ii)</i> Ayudar al experimentador a construir el PL y estructurarlo. <i>iii)</i> Guiar al experimentador para obtener una insignia de la ACM.	§3
<i>Diseño y desarrollo</i>	Se propone la <i>especificación</i> del proceso experimental de forma integrada con el PL, el <i>procedimiento</i> para elaborar un PL acorde a la ACM, el <i>modelo de calidad</i> , el <i>metamodelo</i> , y el desarrollo de la <i>herramienta</i> badgeGo.	§5.2 §5.3 §5.4 ?? ??
<i>Demostración</i>	Se instancia la primera versión del PL orientado al proceso acorde a la ACM que han dado origen a la propuesta.	§7.3
<i>Evaluación</i>	Se evalúa la solución, es decir el PL orientado al proceso, en un estudio de caso que abarca el área de conocimiento de IS. Se utiliza badgeGo como prueba de concepto.	§7.2, §7.3, §7.4, §7.5, §7.6
<i>Comunicación</i>	Se publican los resultados de la investigación.	

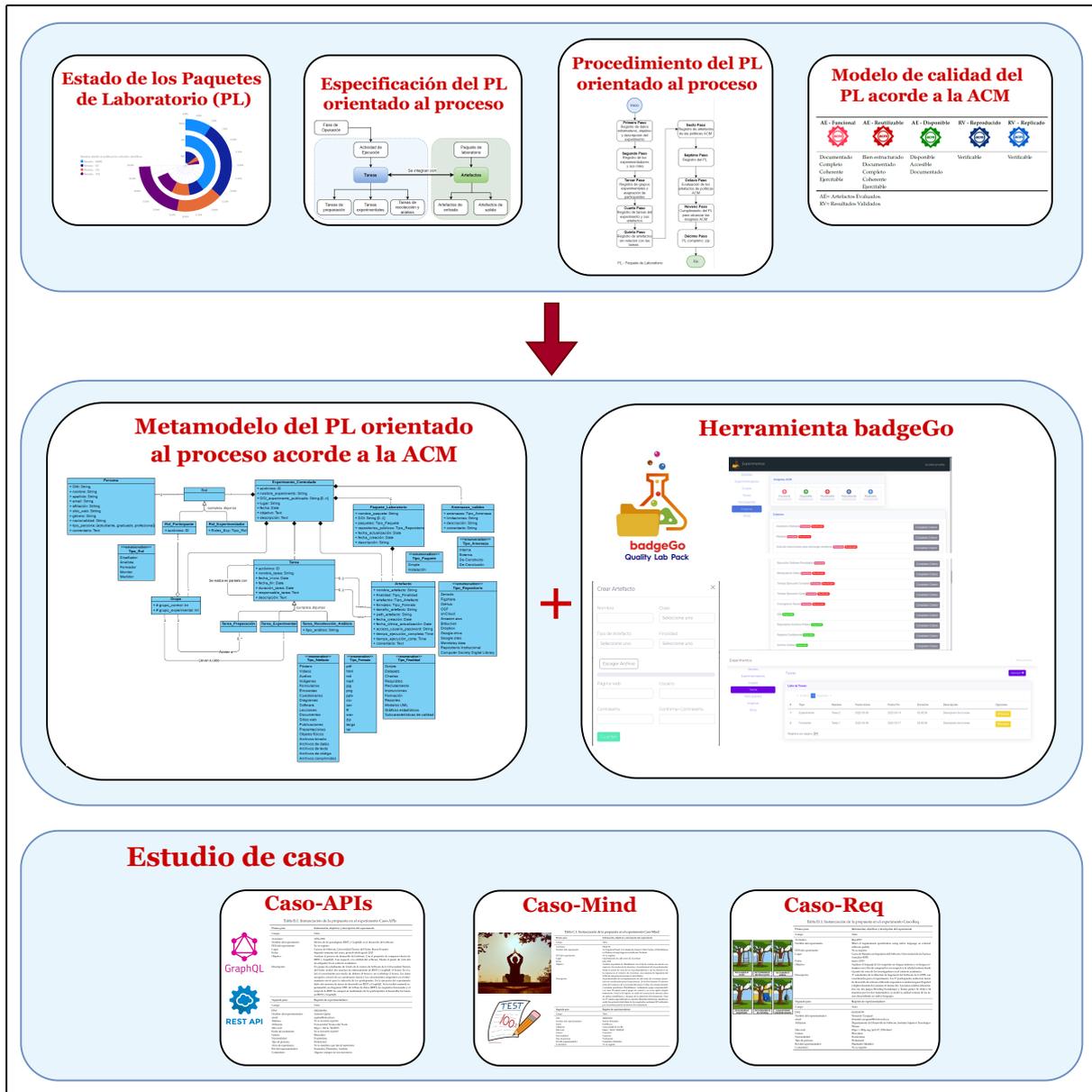


Figura 1.2: Resumen de contribuciones

- *La revisión de las insignias de la ACM para Paquetes de Laboratorio de los 268 artículos.* Se analizó el contenido y calidad de los artefactos del PL para asignar en caso que no la tengan y verificar si ya tienen una insignia basado en las directrices de la ACM [185] y el documento de revisión del Track de Artifact Evaluation del ICSE'2021 [6]. Esta clasificación permitió identificar cómo se han publicado los paquetes y la calidad de los mismos.

Posterior al análisis de los PL y la revisión de las insignias de la ACM de los 268 artículos se define los objetivos de la solución:

1. Integrar el proceso experimental con el desarrollo del PL.
2. Ayudar al experimentador a construir el PL y estructurarlo.
3. Guiar al experimentador para conseguir las insignias de la ACM.

Para la fase de *Diseño y desarrollo* se determina:

- *La Especificación* del proceso experimental de forma integrada con el PL. A partir de la necesidad de crear un PL de calidad surge la propuesta de integrar las tareas experimentales que se desarrollan en la ejecución del proceso experimental junto con los artefactos del PL.
- *Un Modelo de Calidad* para PL acorde a la ACM. A partir de la información sobre el contenido de los artefactos que conforman el paquete y la relación entre la importancia de evaluar la calidad del mismo, se ha propuesto un *modelo de calidad* que defina las *características de calidad* que debe tener un PL para ser evaluado.
- *Un Metamodelo de tareas experimentales.* A partir de la información sobre PL identificada en la fase de *Definición de los objetivos de la solución*, se ha formalizado el *metamodelo* que representa la información a gestionar y la relación entre los distintos conceptos relevantes. Sirve como base para la definición del PL orientado al proceso.
- *El Paquete de laboratorio orientado al proceso acorde a la ACM.* Es la contribución principal de la tesis. Ha pasado por varias versiones. Debido a las limitaciones encontradas en la versión inicial, se han ajustado el *modelo de calidad*, el *metamodelo*, y el *paquete* consecuentemente. Con la versión actual se han podido llevar a cabo las tareas experimentales, la integración de los artefactos y la asignación de las insignias de la ACM.

Para la fase de *Demostración* se utiliza:

La herramienta badgeGo que ha sido desarrollada para elaborar el PL orientado al proceso de manera automatizada.

Para la fase de *Evaluación* se realiza:

- *La evaluación del PL orientado al proceso acorde a la ACM.* Se ha evaluado mediante un *estudio de caso* que abarca la temática de conocimiento de IS: *i)* familias de experimentos de Mindfulness, *ii)* análisis multilingual de requisitos, *iii)* análisis de paradigmas de construcción de arquitecturas microservicios. Durante la validación con el *estudio de caso* se ha evaluado: *i)* el procedimiento para elaborar un PL acorde a la ACM, *ii)* el modelo de calidad para evaluar el grupo de artefactos que cumplen con las políticas de evaluación para alcanzar las insignias de la ACM y *iii)* el metamodelo. Debido a las limitaciones encontradas y la retroalimentación que ha generado la evaluación, han permitido modificar el metamodelo, dando lugar a establecer el PL orientado al proceso acorde a la ACM.

1.6 CONTEXTO DE LA TESIS

Esta tesis ha sido desarrollada en el contexto del grupo de investigación de Ingeniería de Software Aplicado (ISA) de la Universidad de Sevilla.

El trabajo relacionado con esta tesis se encuadra en el ámbito de los siguientes proyectos de investigación:

- **EKIPMENT-PLUS:** *Mejora del rendimiento de procesos basados en conocimiento: Un enfoque empírico multidisciplinar basado en personas, equipos, software y datos.* Proyecto del Plan Andaluz de I+D+i, financiado por la Junta de Andalucía (Consejería de Economía y Conocimiento) (P18-FR-2895, PAIDI 2020). Fecha de inicio: 01-01-2020. Fecha de finalización: 31-12-2022.
- **MEMENTO:** *Herramientas software para la colaboración y toma de decisiones en procesos basados en el conocimiento.* Proyecto del Plan Andaluz de I+D+i, financiado por la Junta de Andalucía (Programa operativo FEDER) (US-1381595). Fecha de inicio: 01-01-2021. Fecha de finalización: 31-12-2022.
- **OPHELIA:** *Optimización de servicios basados en conocimiento usando aplicaciones basadas en servicios.* Proyecto del Ministerio de Ciencia, Innovación y Universidades

(RTI2018-101204-B-C22). Fecha de Inicio: 01-01-2019. Fecha de Finalización: 21-12-2021.

- **APOLO:** *Tecnologías para sistemas orientados al servicio altamente fiables y regulados por contratos con los usuarios.* Proyecto del Plan Andaluz de I+D+i, financiado por la Junta de Andalucía (Programa operativo FEDER) (US-1264651). Fecha de inicio: 01-01-2019. Fecha de finalización: 21-12-2021.
- **HORATIO:** *Mejoras en la fiabilidad, la personalización y los costes de funcionamiento de los servicios de software regulados por los acuerdos de usuario.* Proyecto del Ministerio de Ciencia, Innovación y Universidades (RTI2018-101204-B-C21). Fecha de Inicio: 01-01-2019. Fecha de Finalización: 21-12-2021.
- **BELI:** *Tecnologías para Servicios Cloud Híbridos, Altamente Configurables y Regulados por acuerdo de nivel de servicio (SLA, por sus siglas en inglés de *Service Level Agreement*).* Proyecto del Ministerio Economía y Competitividad (TIN2015-70560-R). Fecha de Inicio: 01-01-2016. Fecha de Finalización: 31-12-2018.

1.7 ESTRUCTURA DEL DOCUMENTO

La organización de la memoria es la siguiente:

Parte I. Introducción. Consta de dos capítulos; en el capítulo §1, se plantea la motivación y relevancia del problema y se encuadra la investigación. En el capítulo §2, se presentan aspectos generales sobre los PL de experimentos y se explican los conceptos y terminología utilizados en el desarrollo de la propuesta.

Parte II. Estado de la cuestión. Se distinguen dos capítulos. En el capítulo §3 se muestra el estado actual de los PL identificando, entre otros, el contenido, su estructura, su publicación y su calidad. En el capítulo §4 se presentan los problemas detectados al abordar los PL, se analiza cómo se están especificando y reportando los artefactos de los PL y se identifican otras propuestas.

Parte III. Propuesta. Es el núcleo de la tesis. Incluye el capítulo §5 en el que se define: la *especificación* del proceso experimental de forma integrada con el PL, el *procedimiento* para elaborar un PL acorde a la ACM, el *modelo de calidad* para evaluar el grupo de artefactos que cumplen las políticas de evaluación para alcanzar las

insignias de la ACM, un *metamodelo* para expresar la información conceptualizada utilizando diagramas de clase UML . Finalmente, basado en el *metamodelo*, se presenta la herramienta badgeGo.

Parte IV. Validación. En el capítulo §7, mediante un estudio de caso se evalúa el PL orientado al proceso acorde a la ACM en el área de Ingeniería del Software.

Parte V. Conclusiones. En el capítulo §8 se analiza el grado de cumplimiento de los objetivos propuestos, se resumen las publicaciones derivadas de este trabajo, y se indican las principales líneas para el trabajo futuro.

Parte VI. Anexos. El anexo §A describe de forma resumida las insignias del reporte de la NISO y hace una comparación con las insignias de la ACM. El anexo §B contiene la Instanciación la propuesta en el Caso-APIs. El anexo §C contiene la Instanciación la propuesta en el Caso-Mind. El anexo §D contiene la Instanciación la propuesta en el Caso-Req. El anexo §E contiene el Cuestionario del aspecto de Satisfacción para la propuesta.

BACKGROUND

El método apropiado para obtener conocimiento de las cosas es deducirlas de experimentos

Isaac Newton (1643-1727),

Físico, teólogo, inventor, alquimista y matemático inglés

En este capítulo, se analizan aspectos generales de los PL experimentales en Ingeniería del Software, tales como su tipología, contenido y haciendo especial hincapié en los conceptos y terminología utilizados a lo largo de la memoria, y específicamente en la propuesta.

En la sección §2.1 se describe una breve introducción a este apartado acerca del PL. En la sección §2.2 se destaca la falta de acuerdo en la terminología a utilizar y calidad de los PL y se detalla los tipos de PL. En la sección §2.3 se presentan los distintos niveles de profundidad en el contenido de los PL. En la sección §2.4 se analiza la literatura y se resume otras propuestas que enfatizan la importancia de los PL para experimentos en IS. En la sección §2.5 se describe las principales características de los PL, la ética, la transparencia, su publicación y disponibilidad en Open Science. En la sección §2.6, se analizan las amenazas a la validez del experimento. Por último, la sección §2.7 resume el capítulo.

2.1 INTRODUCCIÓN

Dada la falta de acuerdo existente entre la terminología de experimentación en general y de PL en particular, en este capítulo se comenta la forma en qué se va a usar la terminología en el resto de la memoria. Para ello, se presenta una breve introducción sobre PL sin necesidad de profundizar a detalle, sino con el objeto de aclarar los términos que presentan más inconsistencias según la bibliografía consultada y abordando los principales conceptos utilizados en experimentación, que se interpretan de forma distinta según los autores.

A continuación, se presenta las definiciones de los términos y conceptos que los principales autores utilizan en el área de IS, considerando varios conceptos equivalentes que determine el término a utilizar en la presente memoria de tesis.

2.2 TERMINOLOGÍA

En esta sección, se abordan los principales conceptos y terminología sobre experimentos controlados y en particular para los PL.

Para ilustrar la explicación de algunos de los conceptos, se utiliza un escenario de experimentación simplificado (*Exp-aprendizaje*) conocido por haber instanciado el *Paquete de Laboratorio orientado al proceso acorde a la ACM* que se presenta en la sección §7.3. El objetivo de *Exp-aprendizaje* es la evaluación de la curva de aprendizaje de dos paradigmas de construcción de arquitecturas microservicios GraphQL y Rest por parte de alumnos de grado de la carrera de Software de la Universidad Técnica del Norte (UTN) en Ecuador. Para ello, cada alumno aplica uno de los paradigmas sobre tareas de desarrollo de software y se evalúa la completitud, el tiempo y el conocimiento adquirido en la etapa de entrenamiento.

2.2.1 Proceso experimental

El proceso experimental esta compuesto por fases y dentro de éstas se establece un conjunto de actividades, tareas, técnicas, herramientas que permiten alcanzar los objetivos de investigación.

Las fases del proceso experimental para algunos autores varia como se muestra en la tabla §2.1. Para este estudio de investigación se considera las fases del proceso ex-

Tabla 2.1: Fases del proceso experimental en IS

Autor	Año	Propuesta	No. Fases	Fases
Basili <i>et al.</i> [45]	1986	Marco de trabajo	4	Definición, planificación, operación, interpretación.
Pfleeger [323]	1995	Guía	6	Concepción, diseño, preparación, ejecución, análisis, difusión y toma de decisiones.
Juristo & Moreno [201]	2001	Libro	4	Definición del objetivo, diseño, ejecución, análisis.
Kitchenham <i>et al.</i> [224]	2002	Guía	6	Contexto experimental, diseño experimental, conducir el experimento y recogida de datos, análisis, presentación de resultados, interpretación de resultados.
Jedlitschka <i>et al.</i> [194]	2008	Guía	3	Planificación experimental, ejecución, análisis.
Wohlin <i>et al.</i> [435]	2012	Libro	5	Alcance, planificación, operación, análisis e interpretación, presentación y paquete.
Fonseca [152]	2014	Modelo	5	Planteamiento del experimento, diseño experimental, ejecución experimental, adquisición de los datos, análisis.
Genero <i>et al.</i> [158]	2015	Libro	5	Alcance, planificación, operación, análisis e interpretación, presentación y difusión.
Revoredo <i>et al.</i> [343]	2021	Modelo de procesos	3	Planificación experimental, ejecución y análisis.

perimental propuestas por Jedlitschka *et al.* [194] que son: la planificación, ejecución y análisis. Esta decisión porque el ámbito de estudio se centra en experimentos controlados con humanos según la clasificación de Zelkowitz *et al.* [439].

El proceso experimental en la IS se enfoca en un aprendizaje iterativo de sucesos, acontecimientos y comportamiento del fenómeno en estudio. La experimentación aporta en la evaluación y validación sobre el proceso y el producto de software [392].

2.2.2 Tipos de replicación de experimentos

La replicación experimental ha sido ampliamente reconocida en [220, 381]. La replicación involucra repetir la investigación en condiciones similares, con variaciones en ciertos aspectos o elementos que componen el proceso experimental como la población de sujetos con el fin de aumentar el grado de confianza y generalizar los resultados originales del experimento. La replicación es una tarea compleja por naturaleza, requiere de tiempo, esfuerzo y en ocasiones involucra altos costos, sin embargo, su ejecución es vital para contrastar los resultados experimentales [103].

Para Brooks *et al.* [68] definen las replications como *internas* cuando los mismos investigadores llevan a cabo la replicación y *externas* si lo realizan investigadores distintos.

Según el grado en que se siga el procedimiento en el experimento base, Lindsay *et al.* [244] clasifica a las replications en *cercanas* y *diferenciadas*.

Shull *et al.* [381] consideran dos tipos de replications *Exactas* y *Conceptuales* y esta última a su vez en *dependientes* e *independientes* según el grado de comunicación que exista entre los investigadores, su colaboración e interacción en las diferentes fases del proceso de replicación.

Basili *et al.* [46], propusieron una clasificación más profunda:

- Réplicas estrictas (sinónimo de cercanas y exactas).
- Réplicas que varían variables intrínsecas al estudio.
- Réplicas que varían variables intrínsecas al enfoque del estudio.
- Réplicas que varían las variables de contexto en el entorno en el que se evalúa la solución.

- Réplicas que varían la forma en que se realiza el experimento.
- Réplicas que amplían la teoría.

A partir de comparar diferentes tipos de replications en IS y en otras disciplinas Gómez *et al.* [164] identifican *dimensiones* en la configuración experimental: *operacionalización, población, protocolo y experimentadores* y establecen tres tipos de replications:

- *Literal*. Llevar a cabo una replicación lo más exacta posible al experimento base.
- *Operacional*. Variar algunas (o todas) las *dimensiones* de la configuración experimental.
- *Conceptual*. Se cambia el *protocolo y operacionalización* con diferentes *experimentadores* para definir la replicación. Solo se conoce un *hecho empírico* establecido por un experimentador con anterioridad.

2.2.3 Familia de experimentos

Cuando se realiza una serie de replications de un experimento base es necesario crear una familia de experimentos que a larga garantice la generalización de los resultados entre los estudios individuales.

Basili *et al.* [46] introducen el concepto de *familia de experimentos* mediante un marco para organizar el conjunto de estudios relacionados, facilitando la construcción de conocimiento de manera incremental y el trabajo entre experimentadores a través de la replicación de experimentos.

Ciolkowski *et al.* [90], determinan el término *familias de experimentos coordinada o concertada* que se caracteriza por un marco común para la recogida de datos de contexto de todos los estudios que la componen, considerando la medición del contexto, la preparación del estudio, el material y el análisis. Las *familias de experimentos concertada* son necesarias para poder extraer conclusiones con un nivel de detalle útil y analizar las variaciones entre los diferentes contextos experimentales.

Para Santos *et al.* [363, 365] Las *familias de experimentos* definidas como *grupos de replications* siguen el mismo objetivo de colaborar entre diferentes grupos de investigadores e instituciones que aporte al cuerpo del conocimiento en la replicación experimental.

La familia de experimentos resuelve problemas que se van encontrando en cada replicación de alguna manera atiende las amenazas a la validez que se reportan en el experimento. Cada experiencia permite mejorar en el proceso de replicación, alcanzar los objetivos de investigación y administrar mejor los recursos experimentales.

2.2.4 Repetición, Reproducción y Replicación

En la investigación experimental en el área de IS se utiliza los términos *reproducción* y *replicación* como sinónimos para referirse a la repetición de un experimento [200]. Para el sitio web de la ACM Artifact Evaluation versión 1.1 [185] los términos tienen diferencia. En la sección §1.2 de esta memoria de tesis ambos términos son descritos. Para la presente investigación se considera utilizar los significados de la terminología del sitio web de la ACM Artifact Evaluation en su versión 1.1 [185], de la siguiente manera:

- *Repetibilidad*. El mismo equipo de experimentadores y la misma configuración experimental. Es decir, lo que en la sección §2.2.2, se ha llamado *replicación interna*.
- *Reproducibilidad*. Diferente equipo de experimentadores, misma configuración experimental (utilizando en parte, artefactos proporcionados por el autor). Es decir, lo que en la sección §2.2.2, se ha llamado *replicación externa*.
- *Replicabilidad*. Diferente equipo de experimentadores, diferente configuración experimental (sin el uso de artefactos proporcionados por el autor). Es decir, lo que en la sección §2.2.2, se ha llamado *replicación externa*.

Para Gómez *et al.* [163] la reproducción, replicación y re-análisis son tres formas de verificar los hallazgos experimentales. Considerando la importancia de otros términos relacionados en la investigación experimental conviene tener claros algunos de ellos:

- *Re-Análisis*. Es utilizar los datos de un experimento ya ejecutado para verificar los resultados en lugar de volver a ejecutar el experimento. Se pueden utilizar nuevas técnicas de análisis o se repiten las técnicas ya utilizadas [163].
- *Meta-Análisis* o síntesis cuantitativa de experimentos. Consiste en combinar los resultados de varios experimentos para generar nuevo conocimiento [22].

- *Replicación Parcial*. Solo se replica una parte del experimento base [87]. Según Ramac *et al.* [337], las replications parciales tienen el mismo objetivo que el experimento base, pero se modifican el diseño y procedimiento experimental. Para Hendrick [176] en una replicación parcial se añaden o suprimen variables del experimento base en función del objetivo de investigación.

La verificación de los resultados experimentales desempeña un papel fundamental en el progreso científico, ya que consolida el conocimiento [163]. Para el filósofo Karl Popper [327] declara su valor y manifiesta: “No tomamos en serio ni siquiera nuestras propias observaciones, ni las aceptamos como observaciones científicas, hasta que las hemos repetido y comprobado. Sólo a través de tales repeticiones podemos convencernos de que no se trata de una mera coincidencia iso-lógica, sino de acontecimientos que, debido a su regularidad y reproducibilidad, son en principio intersubjetivamente comprobables”.

2.2.5 Sujetos y objetos experimentales

Los sujetos experimentales son las personas que participan en el experimento y las unidades experimentales son los objetos materiales, a quienes se aplica un tratamiento.

Según Wohlin *et al.* [435], el *sujeto experimental* es la persona que participa en un estudio empírico para evaluar un *objeto experimental*. Según Jedlitschka *et al.* [194], cuando los sujetos o en general participantes no son humanos, se denominan *unidades experimentales*.

Kitchenham [224], considera importante identificar la *población* de la que se seleccionan los *sujetos experimentales* y definir el proceso de selección para poder generalizar los resultados, la selección o *muestra* debe ser representativa de la *población* a estudiar.

- Para esta investigación, en *Exp-aprendizaje* los *sujetos experimentales*, *participantes* o *población* son los alumnos que evalúan los paradigmas de construcción de arquitecturas microservicios GraphQL y Rest.

2.2.6 Comunicación entre investigadores

Una replicación necesita de mecanismos e instrumentos que permitan a los investigadores ejecutar el proceso experimental con éxito. La comunicación entre investigadores es un mecanismo valioso para transmitir el conocimiento científico que por lo

general no se reporta por su dificultad de transcripción. La comunicación es una tarea compleja, requiere de una transferencia de conocimiento profunda y de recursos como tiempo y costo [416].

Existen varios mecanismos para realizar la comunicación entre investigadores como redes de investigación, congresos técnicos, reuniones menos formales como sesiones de grupos de investigación, consultas personales que consideren llegar a un mismo objetivo de investigación. Es recomendable mantener reuniones grupales físicas o al menos que la comunicación virtual utilice herramientas digitales que fluya sin inconvenientes la conversación entre los investigadores. En el área de ISE existe la Red Internacional de Investigación en IS, ISERN por sus siglas en Inglés (International Software Engineering Research Network [189], que es un medio para participar de las sesiones regulares de sus miembros.

Existe algunos factores importantes que benefician una replicación en especial cuando los investigadores son de distintos grupos. Un factor puede ser el *conocimiento tácito*, aquel conocimiento que surge de forma intuitiva e inconsciente, y que se adquiere mediante la propia experiencia, caracterizándose por ser personal y contextual [229, 380].

Por lo general el *conocimiento tácito* que el experimentador original adquiere cuando lleva a cabo el proceso experimental no es transmitido o descrito completamente en el reporte del artículo científico y menos se evidencia en el PL. Por lo que una comunicación directa entre experimentadores originales y nuevos, facilitará la recepción del *conocimiento tácito* [229] al nuevo investigador para clarificar el proceso de replicación. Durante el mecanismo de comunicación es importante destacar los artefactos que conforman el PL. Para llevar a cabo la replicación se necesita la disponibilidad del PL como un medio de soporte al proceso de investigación experimental.

2.2.7 Paquete de Laboratorio

En la presente investigación el Paquete de Laboratorio es el concepto fundamental por lo que es importante definir los términos y conceptos relacionados de dicho concepto.

Cómo se había mencionado en la sección §1.2, un artefacto es considerado por la ACM [185] como un *objeto digital* creado por los autores de una investigación para ser utilizado como parte del estudio o generado por el experimento en sí. A su vez, el conjunto de información y artefactos que se necesita para replicar el experimento es conocido por algunos investigadores [41, 46, 334, 394] en IS como PL.

Existe diferente terminología que los investigadores utilizan para nombrar el PL. En [97, 271] los autores definen como *Paquete experimental*, en [179, 349, 351, 393] como *Paquete de replicación*, en [21, 356] como *Material experimental*, en [365] como *Material suplementario*. En este trabajo de investigación usaremos el término "Paquete de Laboratorio". Para otras disciplinas maduras como Medicina, Biología y Psicología la replicación experimental se basa en un *Protocolo* el cual tiene un conjunto de instrumentos y guías basado estrictamente en procedimientos que se siguen para la replicación [394], y también de un Paquete de Laboratorio que contiene elementos (artefactos) para transmitir la información. El PL es clave para fortalecer y apoyar a la replicación de experimentos pero no garantiza el éxito de la replicación al cien por cien [381].

Un PL es un conjunto de artefactos que deben estar estructurados, validados y disponibles para la transferencia de conocimientos en relación a una familia de experimentos. El PL contiene las instrucciones y materiales necesarios para replicar un experimento y debe dar soporte a todas las actividades del proceso experimental que puede ser de un experimento original o de una replicación [392].

Entonces, al final el PL esta compuesto de: *i)* material experimental, *ii)* guías paso a paso, *iii)* otros artefactos que ayuden a comprender el trabajo, *iv)* el artículo científico (preprints) y *v)* las amenazas a la validez.

A continuación, se detallan los distintos tipos de PL identificados hasta el momento.

2.2.8 Tipología de los Paquetes de Laboratorio

Como ya se ha comentado, los PL son importantes y útiles para que otros investigadores repliquen el estudio y construyan la base del conocimiento para futuros trabajos. No obstante existe diversidad en el contenido, estructura y calidad de los PL para el área de ISE [382].

Para Solari *et al.* [394] los PL se clasifican en tres categorías:

Experimento individual: Se relaciona a los PL contruidos para un único experimento o una única familia de experimentos. Por lo general, suelen tener una estructura especializada, adecuada para describir las características particulares del experimento. Este tipo de paquete se muestra cuando por falta de espacio en los artículos científicos no se puede incluir todos los detalles del experimento. A manera de ejemplo, se puede observar este tipo de paquete en el estudio de Genero *et al.* [107] relacionado a un experimento de comprensión de UML.

Repositorios de material: Se focalizan en los objetos experimentales. Estos objetos pueden utilizarse en diferentes tipos de experimentos. Pueden contener diferentes versiones de programas y su respectiva información. Un ejemplo de este tipo de paquete es el repositorio de artefactos para pruebas de experimentación creado por Do *et al.* [125].

Infraestructuras de apoyo para propósitos generales: Apoyan las actividades del proceso experimental para cualquier familia de experimentos, es decir, no se limitan a una familia específica de experimentos. Las estructuras pueden ser conceptuales (por ejemplo, modelos de datos, ontologías) o estar implementadas en un sistema específico. Un ejemplo de este tipo de paquete es el marco computacional propuesto por Mian *et al.* [271]

Por otro lado, en relación a la naturaleza de los artefactos que componen el PL para el Track AE–Artifact Evaluation del ICSE’2021 [186] y la guía de envío y revisión de artefactos para el mismo congreso propuesta por Abrahão y Mendez [6] los paquetes son de dos tipos:

Paquete Simple: Contiene solo artefactos documentales que pueden estar en formato PDF, una hoja de Excel, o un archivo de texto, los autores pueden simplemente guardar todos los documentos en un solo archivo de paquete (zip o tar.gz).

Paquete de Instalación: Se elabora cuando el artefacto consiste en una herramienta o un sistema de software, contiene el material de instalación, configuración, una imagen de contenedor de Docker o de VirtualBox VM, es decir todas las herramientas necesarias para usar el artefacto.

Para la presente memoria de tesis el tipo de paquete que estudiamos es el *Experimento individual* esto se debe a la experiencia y aplicación de las familias de experimentos controlados con humanos que nuestro grupo de investigación ha estudiado a profundidad. Además, consideraremos los paquete *Simple* y de *Instalación* por la naturaleza de los artefactos.

2.3 CONTENIDO DE LOS PAQUETES DE LABORATORIO

En las replications de experimentos controlados con humanos en el área de IS la diversidad de investigaciones y la carencia de guías y directrices para llevar a cabo las replications hace que el contenido de los PL sea muy distinto.

El contenido de los paquetes puede presentarse de varias formas desde un contenido muy ligero como un breve resumen del experimento hasta presentar información e instrucciones detalladas del proceso experimental, con vídeos y tutoriales técnicos. Los PL no tienen una estructura normalizada y estandarizada, a pesar de que en los congresos y algunas revistas disponen de guías de presentación de los paquetes, los autores no siguen completamente las guías. Esto se ha evidenciado en algunos congresos como ICSE'2021. Existe un criterio de evaluación muy personal por parte de los investigadores que evalúan los PL, puede ser que para un evaluador sea suficiente la cantidad y contenido de los artefactos presentados y para otro evaluador sea todo lo contrario. A pesar de que existe criterios de evaluación definidos en los Track Artifact Evaluation de los congresos esto sigue siendo un reto y exige esfuerzo para llevar a cabo su revisión.

Solari [392] define distintos niveles de profundidad en el contenido de los PL para experimentos controlados en IS, categorizándoles de la siguiente manera:

- **Artículo.** Cuando el artículo científico no viene acompañado de un PL y este es el único que presenta la información para la replicación. Algunas revistas no limitan demasiado el espacio y formato para publicar el experimento esto ayuda a que se publique mayor información del estudio.
- **Operacional.** Cuando el PL incluye los artefactos que proporcionan los elementos de la producción del software, instrucciones para los sujetos, formularios, instrumentos de medición, modelos conceptuales o diagramas como UML que explican la teoría subyacente al experimento.
- **Descriptivo.** Cuando el PL incluye artefactos que describen los elementos que contiene el paquete para que el replicador pueda comprender y usarlos de manera adecuada. Se debe describir el diseño utilizado y explicar a detalle el contexto en el que fue realizado el experimento. Especialmente en la replicación externa es necesario contar con este tipo de información.
- **Procedimental.** Cuando el PL incluye guías para el proceso experimental. Estas guías pueden ser usadas paso a paso por los replicadores, de esa forma mejorar el nivel descriptivo definido anteriormente.
- **Fundamentado.** Cuando el PL contiene justificaciones con respecto a las decisiones tomadas por los experimentadores originales para realizar el experimento. Estas justificaciones permite direccionar a los replicadores el camino a seguir y

entender cuales fueron las razones que les llevaron a ejecutar el experimento y a la vez tomar sus propias decisiones.

El contenido de los PL acompañado de las características antes descritas probablemente cambien debido a los recursos disponibles y el contexto de la replicación que en ese momento se presente. Los paquetes son dinámicos por tanto la agregación y evolución del material experimental también se registra en cada replicación [392].

2.4 OTRAS PROPUESTAS DE PAQUETES DE LABORATORIO

Varios han sido los investigadores que han indagado el estudio e importancia acerca de los PL experimentales en el área de IS. Entre los cuáles se manifiestan los siguientes:

- Basili *et al.* [41], manifiestan que las replicaciones pueden ser promovidas por la disponibilidad de *kits experimentales* que recogen la experiencia de los investigadores originales. Su estudio propone un manual de laboratorio de un experimento controlado para evaluar la eficacia de una técnica de lectura de software conocida como lectura basada en la perspectiva (PBR). El manual de laboratorio incluye el diseño experimental, los instrumentos para su ejecución y los procedimientos para las pruebas estadísticas, con el fin de promover el empaquetado de experimentos y facilitar la replicación.
- Shull *et al.* [382], consideran que los resultados de cualquier estudio no pueden extrapolarse a todos los entornos porque hay muchas fuentes de variación incontrolables entre los distintos entornos. Los investigadores, concluyen que la disponibilidad de PL para los experimentos puede fomentar mejores replicaciones y estudios complementarios. Sin embargo, incluso con PL efectivamente especificados, la transferencia de conocimientos experimentales puede seguir siendo difícil. Además, discuten las estructuras de colaboración que existe en las replicaciones y transferencia de conocimientos experimentales para un proyecto bilateral entre Brasil y Estados Unidos.
- Almqvist [21], lleva a cabo una encuesta de literatura a 51 experimentos replicados en ISE. Revisa a detalle las directrices existentes para la experimentación y la

replicación, tanto dentro del área de IS como de otras disciplinas. Los experimentos son estudiados desde una perspectiva tanto cuantitativa como cualitativa. La encuesta también analiza cómo los investigadores describen su experiencia con la replicación, en lo que respecta al acceso del material original (Paquetes de Laboratorio) para motivar las replications.

- Kitchenham *et al.* [221], utilizan un método de lectura inspirado en revisiones basadas en perspectivas y listas de verificación para realizar una evaluación teórica de las directrices para experimentos controlados en el área de IS. Las perspectivas utilizadas fueron: Investigador, Practicante/Consultor, Meta-analista, Replificador, Revisor y Autor. Los investigadores detectaron 44 problemas en los que las directrices se beneficiarían de enmiendas o aclaraciones y 8 defectos. Finalmente, proponen una estructura de guía de evaluación para reportar estudios empíricos en IS.
- Vegas *et al.* [416], manifiestan que la replicación de experimentos es una tarea clave en la IS. Las replications exitosas permiten que el cuerpo de conocimientos de una disciplina crezca, ya que los resultados se suman a replications anteriores. Sin embargo, la replicación es extremadamente difícil en la IS, sobre todo porque es difícil conseguir un escenario exactamente igual al del experimento original. En consecuencia, hay que introducir cambios en el experimento para adaptarlo al nuevo entorno experimental. Para poder replicar un experimento, también hay que transmitir información (normalmente de forma oral y escrita) entre los investigadores que realizaron el experimento anterior y los que van a replicarlo. Los investigadores también analizaron la influencia del tipo de comunicación que existe entre los experimentadores en el éxito de una replicación.
- Solari *et al.* [393], consideran que los Paquetes de replicación son un medio para que los investigadores compartan artefactos que mejoren la replicación de los experimentos. Sin embargo, la replicación externa de experimentos de IS sigue siendo escasa. Además, manifiestan que a los investigadores que reciben Paquetes de replicación les resulta difícil evaluar su contenido tanto para adaptarlo al nuevo entorno de replicación como para estimar el esfuerzo que supondrá dicha adaptación. Los investigadores proponen un marco para mejorar la evaluación de los paquetes de replicación.
- Shull *et al.* [381], destacan la importancia de producir una documentación adecuada para un experimento (original o replicación) que permita una replicación.

Los investigadores manifiestan que, una replicación bien documentada proporciona los detalles necesarios para comprender suficientemente el estudio que se está reproduciendo sin exigir al replicador que siga servilmente los procedimientos establecidos.

- Jedlitschka *et al.* [194], definen que uno de los principales problemas para integrar los resultados de los estudios en un cuerpo común de conocimientos es la heterogeneidad de los estilos de presentación de informes para experimentos controlados en el área de IS. Los investigadores proponen una directriz para informar de los resultados de los experimentos controlados.

Tabla 2.2: Comparativa entre propuestas de paquetes de laboratorio para experimentos en Ingeniería del Software

Autor	Año	Estructura de contenido	Directriz	Guía paso a paso	Insignias ACM	Proceso automatizado
Basili <i>et al.</i> [41]	1996	✓	✓	x	x	x
Shull <i>et al.</i> [382]	2004	✓	✓	x	x	x
Almqvist [21]	2006	✓	✓	x	x	x
Kitchenham <i>et al.</i> [221]	2006	✓	x	x	x	x
Vegas <i>et al.</i> [416]	2006	✓	✓	x	x	x
Solari <i>et al.</i> [393]	2006	✓	✓	x	x	x
Shull <i>et al.</i> [381]	2008	x	✓	x	x	x
Jedlitschka <i>et al.</i> [194]	2008	✓	x	x	x	x
Solari <i>et al.</i> [394]	2018	✓	✓	x	x	x
Propuesta de esta tesis	2022	✓	✓	✓	✓	✓

- Solari *et al.* [394], manifiestan que los experimentos de IS se componen de conceptos, procedimientos y artefactos complejos. Los PL son un medio de transferencia de conocimientos entre investigadores para facilitar las réplicas de experimentos. Los investigadores proponen una estructura y una plantilla (checklist) del contenido de los PL para experimentos de IS.

La tabla §2.2 muestra una comparativa entre características definidas como: *i)* Directriz, *ii)* Guía paso a paso, *iii)* Estructura de contenido, *iv)* Insignias de la ACM y *v)* Proceso automatizado, de otras propuestas identificadas para PL con la propuesta que se presenta en esta tesis. Estas características permiten identificar de qué manera los

investigadores han logrado profundizar la importancia, el desarrollo y evaluación de los PL experimentales en área de IS.

2.5 OPEN SCIENCE

A inicios del siglo XVII los científicos tenían problemas con dar a conocer sus descubrimientos debido al sistema aristocrático que lidiaban en ese entonces. Científicos como Galileo, Kepler, Newton, Huygens y Hooke, reclamaban sus descubrimientos para probar su autoría ante un sistema que demoraba la publicación [114]. La tendencia de acceso a las revistas científicas crecía hasta que en 1660 Inglaterra estableció la Royal Society y en 1666 los franceses establecieron la Academia Francesa de las Ciencias. En 1665 Henry Oldenburg se convirtió en el editor de la primera revista científica de ciencia "Philosophical Transactions" de la Royal Society, que marcó el inicio de la publicación científica [352].

Open Science traducido al Español como *Ciencia Abierta* comienza a finales de la década del 2000, bajo la premisa de Ciencia 2.0 [379]. A partir del 2012 el New York Times acuña a esta premisa el término de *Open Science* [243]. Varios organismos mundiales han puesto interés en promulgar la *Ciencia Abierta* empezando desde la Royal Society británica en el 2012, mediante un documento titulado "Science as an Open Enterprise" [353].

Open Science fomenta el movimiento de poner a disposición del público cualquier artefacto de investigación. Integra cuatro aspectos importantes: *i) Open Access*: acceso libre a los artículos, *ii) Open Data*: datos disponibles, *iii) Open Material*: material usado para analizar los datos y *iii) Open-Source Software*: software de código abierto.

Posteriormente, esto fue seguido por la publicación de libros [272, 294, 431], una lista de libros blancos, documentos de gobierno como las política establecidas en la Comisión Europea [96] e institutos y grupos de expertos como el CERN (La Organización Europea para la Investigación Nuclear) [89], el Centro de Ciencia Abierta [102], entre otros.

Open Science esta tomando fuerza en la comunidad científica, aplicada de manera frecuente en otras ciencias maduras como la Medicina y la Psicología [111, 281, 434], donde es admitida como una norma. En la Ingeniería del Software aún esta iniciando y a sido aceptada con bastante interés por los investigadores, es un nicho de investigación que recién empieza.

El concepto de *Open Science* abarca desde la divulgación del código fuente del software (*código abierto*), pasando por los propios datos (*datos abiertos*) y el material utilizado para analizar los datos (*material abierto*), hasta los documentos que informan los resultados del estudio (*acceso abierto*) [268].

La divulgación pública de los artefactos de investigación aumenta la *transparencia* y, por tanto, la reproducibilidad y replicabilidad del proceso científico. *Open Science* es un medio para colaborar con la comunidad científica y fomentar las redes de investigación y co-crear conocimiento. Los datos abiertos y el código abierto fortalece la comprensión y el entendimiento para los investigadores que no participan en la realización de esos estudios. *Open Science* permite compartir preprints, detalles de infraestructura, modelos de licencia y derechos de autoría, modelos de revisión como la revisión a doble ciego [168]. A manera de ejemplo, cuando se comparte paquetes de laboratorio se aumenta la reproducibilidad del estudio, a su vez, refuerza la *credibilidad* de las conclusiones expuestas al final del estudio y permite que otros construyan su propio trabajo basándose en esas conclusiones.

2.5.1 Transparencia en Open Science

Para Grand *et al.* [167] *Open Science* abre la oportunidad de *transparentar* y revelar el proceso científico en tiempo real. La autenticidad y actualidad de las afirmaciones que publican los investigadores es un potencial para que la confianza de la comunidad científica en la ciencia se fortalezca y viceversa. *Open Science* está cambiando todos los aspectos del método científico para que sea más abierto, inclusivo e interdisciplinario [153].

La transparencia de los métodos y datos en la publicación científica da la posibilidad de evaluar no solo los resultados de los estudios sino también debatir y evidenciar las pruebas que subyacen las afirmaciones [353].

Es necesario que la comunidad científica ponga a disposición la mayor parte de evidencia para transparentar sus estudios, como una regla personal de los investigadores para elevar la calidad de los resultados. Algunas revistas de otras áreas del conocimiento como la *Journal of Personality de Psicología* [436], aplica normas con respecto a la transparencia analítica y de datos.

La recopilación y el análisis de datos son dos de los pasos más importantes del proceso científico y también los más vulnerables para cometer errores. Los informes totalmente transparentes de la investigación también incluirían materiales abiertos y

accesibles para todos los aspectos relevantes del estudio, con sus respectivas políticas de derechos de autor [436].

2.6 AMENAZAS A LA VALIDEZ EN EXPERIMENTACIÓN

Las *amenazas a la validez* son el conjunto de situaciones, factores, debilidades y limitaciones que interfieren en los resultados de un determinado estudio empírico y arriesgaría los hallazgos del mismo. La importancia de reportar la validez experimental es alta, así lo manifiestan algunos investigadores por ejemplo de las ciencias sociales específicamente de Psicología: Campbell [82], Reis y Judd [342], Cook *et al.* [98], Shadish *et al.* [372]. De igual forma sucede en el área de IS investigadores como Feldt *et al.* [146], Mustafa *et al.* [284], Teixeira *et al.* [401] manifiestan que se necesita enfatizar en la validez experimental para mejorar los experimentos y así fomentar las replicaciones.

Wohlin *et al.* [435] clasifican las *amenazas a la validez* de la siguiente manera:

Validez de la conclusión: Se relaciona con el grado de confianza de un nivel determinado de significación estadística con relación a las conclusiones del estudio. Puede verse afectada cuando el poder estadístico de una prueba es relativamente baja.

Validez interna: Es la confianza que existe en la relación de causa y efecto entre el tratamiento y el resultado. Es decir entre las variables independientes y dependientes. Se presenta amenazas a la validez interna cuando el resultado viene afectado por algún factor no deseado.

Validez del constructo: Está relacionado con la capacidad de las variables independientes y dependientes para reflejar o medir el constructo que se va a estudiar. Algunas amenazas están relacionadas con el diseño del experimento, otras con factores sociales.

Validez externa: Es el grado con que pueden generalizarse los resultados de un experimento. Algunos ejemplos de amenazas a la validez externa se evidencia en la selección no representativa de los sujetos experimentales o un entorno experimental que no es adecuado.

Para Shull *et al.* [381] la *validez externa* con respecto a las replicaciones determinan que los resultados experimentales no dependen de las condiciones específicas del ex-

perimento base. Mientras que, para la *validez interna*, permiten conocer el rango de condiciones bajo las cuales se mantienen los resultados experimentales.

En [291] se citan otros tipos de amenazas, aunque menos utilizadas en IS. La *validez ecológica* se relaciona con la similitud entre el entorno experimental y la realidad investigada. Otro tipo de amenaza es la *validez aparente* o *face validity* se relaciona con el grado en que una medida mide lo que se quiere medir según el observador. Es decir, importa más la apariencia de validez que presenta un estudio que su validez real.

2.6.1 Relación entre validez interna y validez externa

Según Arnau [33] una característica relevante del diseño experimental se manifiesta en la validez de inferir las hipótesis, es decir, comprobar la supuesta relación entre las variables dependientes e independientes. Todo diseño experimental debería presentar tanto *validez interna* como *validez externa*. Ambos tipos determinan si el experimento tiene validez. Por lo general las condiciones experimentales no son compatibles para alcanzar dichos tipos de validez. Los rígidos controles para incrementar la *validez interna* limitan la *validez externa* de los resultados. La *validez externa* o posibilidad de generalización de los resultados de un experimento presenta más retos en el contexto de las ciencias humanas que en otros sectores científicos, debido a la complejidad del ser humano, esta conjetura puede considerarse también para el área de IS.

Para Siegmund *et al.* [384] cuando los investigadores se centran en la *validez interna* intentan asegurar que el resultado sólo es causado por el tratamiento recibido. Si se centran en la *validez externa*, analizan el efecto en el mundo real, pero sin saber qué factores causan realmente el resultado observado.

En el estudio de Zhou *et al.* [444] analizan amenazas comunes que pueden afectar la validez del estudio. Concluyen que los investigadores prestan mucha atención a la validez interna y de conclusiones, mientras que la validez de constructo y externa reportan con menos frecuencia.

2.7 RESUMEN

En este capítulo, se han presentado los principales conceptos sobre experimentos controlados utilizados en esta memoria, destacando la terminología relacionada con los PL y las replicaciones. Seleccionando la terminología que se va a utilizar, justifican-

do los motivos en caso de que haya varias alternativas.

Se ha destacado la aplicación de una mínima comunicación entre investigadores como un mecanismo importante para apoyar la replicación de experimentos. El *conocimiento tácito* de los investigadores originales con frecuencia es el que menos se obtiene en los artículos científicos y en sí el que más se necesita para comprender a profundidad la motivación que les llevo a ejecutar dicho experimento.

Para completar la introducción a los PL, se ha revisado la *tipología* de los paquetes por su *contenido y estructura* siguiendo la propuesta de Solari *et al.* [394] y por la naturaleza de los artefactos propuesto por Abrahão y Mendez [6]. Además, se ha revisado la literatura acerca de otras propuestas que estudian la importancia de los PL.

Se ha enfatizado la importancia del movimiento *Open Science* para generar conocimiento a partir de la colaboración y publicación de contenido científico de acceso abierto.

Finalmente, se aborda de manera general las amenazas a la validez, una clasificación y la importancia de ser consideradas en el reporte del experimento.

PARTE II

ESTADO DE LA CUESTIÓN

ESTADO ACTUAL DE PAQUETES DE LABORATORIO EN IS

Para adquirir conocimiento uno debe estudiar, pero para adquirir sabiduría uno debe observar

Marilyn vos Savant (1946 –),

Escritora, columnista, matemática, literata y novelista estadounidense

En este capítulo, se muestra el estudio realizado sobre el estado actual de los PL de experimentos realizados en IS y se responde a cada una de las preguntas de investigación, que componen la pregunta PI_1 y la pregunta PI_3 .

En la sección §3.1, se presentan las preguntas de investigación. En la sección §3.2, se contesta a la pregunta sobre la práctica habitual en la comunidad de IS de hacer públicos los PL. En la sección §3.3, se presenta la tendencia en la publicación de PL tras la iniciativa de Open Science. En la sección §3.4, se presenta el nivel de calidad de los PL en relación con las insignias de la ACM. En la sección §3.5, se presenta el nivel de calidad de los PL con respecto a los módulos del estudio de Solari et.al [394]. En la sección §3.6, se contesta a la pregunta sobre guías y directrices que los autores utilizan para reportar los PL. En la sección §3.7, se contesta la pregunta sobre los repositorios digitales que se suelen utilizar para que los PL estén disponibles en la comunidad de IS. La sección §3.8 presenta las conclusiones. Por último, la sección §3.9 resume el capítulo.

3.1 INTRODUCCIÓN

Para comprender el estado actual de los PL de experimentos controlados con humanos publicados en el área de IS, se ha llevado a cabo una revisión bibliográfica sobre los PL en general.

Se ha considerado analizar 168 artículos científicos del estudio realizado por Revoredo *et al.* [343], debido a la relación que existe con el reporte de estudios de experimentos controlados con humanos en IS. Además, fue una sugerencia de los expertos del Track Simposio Doctoral del CIBSE 2021 (*Conferencia Iberoamericana de Ingeniería de Software*), donde el autor de la presente tesis expuso la idea de investigación.

Luego de revisar los 168 artículos científicos decidimos actualizar el método de revisión de mapeo sistemático propuesto por Revoredo *et al.* [343] realizado entre 2000 y 2020¹, motivados por la tendencia del movimiento *Open Science* y el llamado a evaluar los PL en las revistas y congresos de la comunidad científica en el área de IS.

El estudio de Revoredo *et al.* [343] está enfocado en un análisis mixto de investigación cualitativa y cuantitativa, en concreto un diseño de método-mixto secuencial propuesto por Venkatesh *et al.* [419]. Además, para la codificación cualitativa de los documentos de experimentación dicho estudio ha sido inspirado en los estudios de mapeo sistemático de Kitchenham y Charters [222].

Durante la actualización, se ha aplicado los mismos criterios de inclusión y exclusión, la misma cadena de búsqueda con el término ("*experiment*") y las mismas revistas científicas: Empirical Software Engineering (EMSE)², Information and Software Technology (IST)³, Journal of Systems and Software (JSS)⁴, IEEE Transactions on Software Engineering (TSE)⁵ entre el año 2020 y 2021⁶. El total de estudios de la búsqueda actualizada en las cuatro revistas es de 895 artículos y el resultado de la selección es de 87 artículos científicos.

Para completar la selección de los estudios, se ha considerado seguir la recomendación de Revoredo *et al.* [343] con respecto a seleccionar conferencias relevantes para aplicar el método de selección. Se ha escogido cuatro conferencias (ICSE'2021, ESEC/F-

¹Fecha final de la revisión bibliográfica del estudio de Revoredo *et al.* marzo 2020.

²<https://bit.ly/3wZY142>

³<https://bit.ly/30EtQUa>

⁴<https://bit.ly/3npZWLR>

⁵<https://bit.ly/2YZWH14>

⁶Fecha final de la revisión bibliográfica de la presente memoria de tesis diciembre 2021.

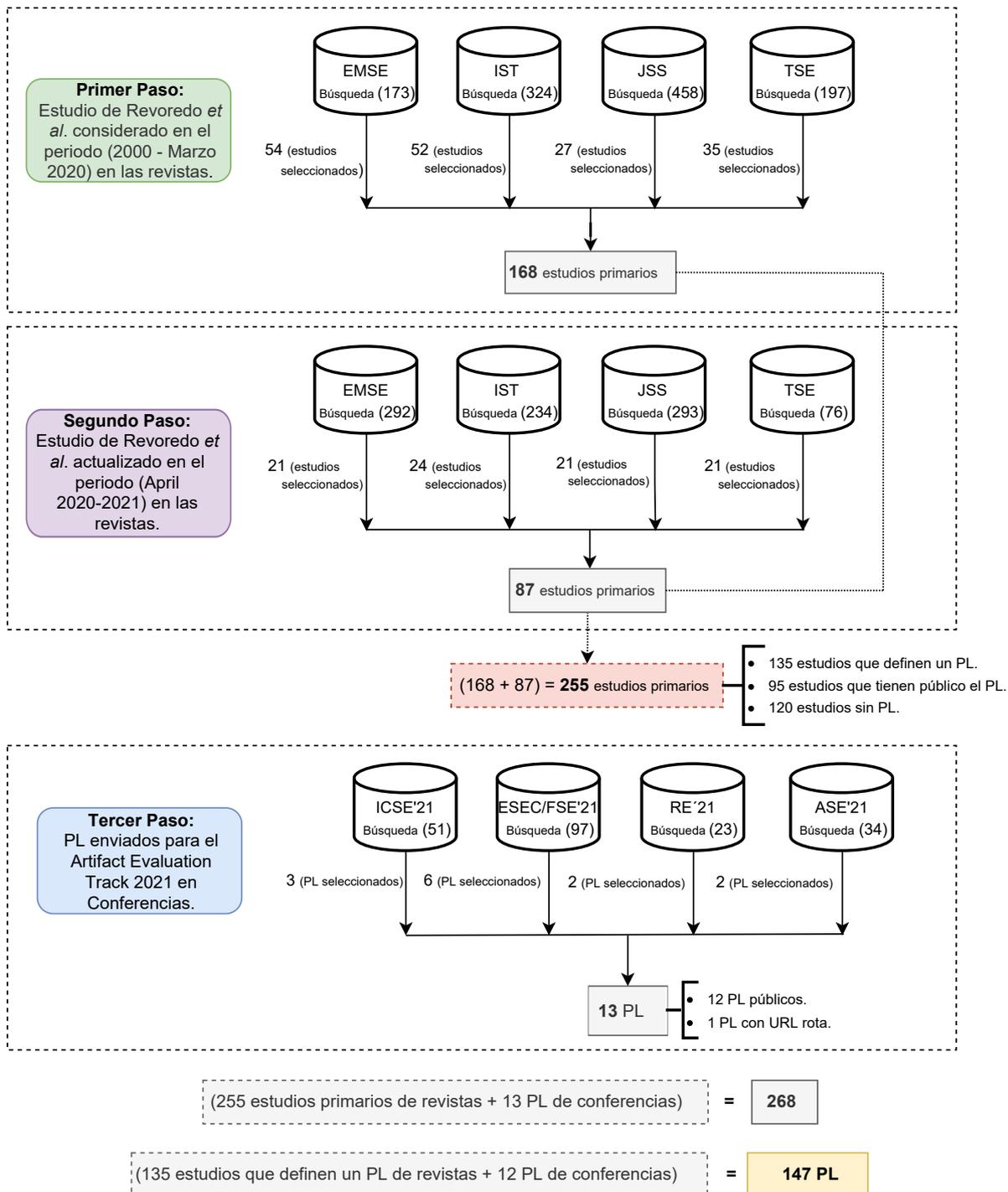


Figura 3.1: Proceso de selección para los estudios primarios en revistas y paquetes de laboratorio en conferencias.

SE'2021, RE'2021, ASE'2021), específicamente revisamos los PL aprobados en el Track de Artifact Evaluation. El total de PL de la búsqueda en las conferencias es de 205 PL y el resultado de selección es de 13 PL.

Finalmente, el total de estudios seleccionados para analizar los PL es de 268 (168 artículos de Revored, 87 artículos actualizados de las revistas y 13 PL de las conferencias). La figura §3.1 muestra el proceso de especificación y selección de los estudios primarios y los paquetes de laboratorio.

Para el análisis de los estudios seleccionados se han aplicado algunos criterios de categorización con valores y métricas cualitativas. Para definir dichos criterios se ha utilizado los documentos: *Artifact Review and Badging – Current* de la ACM [185] y el *Artifact Evaluation Track-Submission and Reviewing Guidelines* [6] de la conferencia ICSE'2021. Primero, se han considerado los siguientes criterios generales:

- El artículo científico publica el PL.
- El artículo científico sigue alguna guía para reportar el PL.
- El artículo científico profundiza el contenido y estructura del PL.

Segundo, los siguientes criterios se basan en los siete módulos del estudio realizado por Solari *et al.* [394] con respecto al contenido y estructura del PL:

- El Paquete de Laboratorio presenta el módulo de introducción.
- El Paquete de Laboratorio presenta el módulo de teoría.
- El Paquete de Laboratorio presenta el módulo de entrenamiento.
- El Paquete de Laboratorio presenta el módulo de experimento.
- El Paquete de Laboratorio presenta el módulo de evolución.
- El Paquete de Laboratorio presenta el módulo de replicación.
- El Paquete de Laboratorio presenta el módulo de agregación.

Tercero, los siguientes criterios se basan en el documento de *Artifact Review and Badging – Current* de la ACM [185] y la guía de envío de PL del ICSE'2021 [6] para la evaluación y asignación de insignias:

- El Paquete de Laboratorio presenta la insignia Funcional.
- El Paquete de Laboratorio presenta la insignia Reutilizable.

- El Paquete de Laboratorio presenta la insignia Disponible.
- El Paquete de Laboratorio presenta la insignia Reproducido.
- El Paquete de Laboratorio presenta la insignia Replicado.

La pregunta de investigación **PI₁**: ¿Cuál es el estado actual de las propuestas para la elaboración de los Paquete de Laboratorio? se ha concretado en las siguientes preguntas:

- **PI_{1.1}**: ¿Es una práctica habitual en la comunidad de IS hacer públicos los Paquetes de Laboratorio?
- **PI_{1.2}**: ¿Cuál es la tendencia en la publicación de los Paquetes de Laboratorio tras la iniciativa *Open Science*?
- **PI_{1.3}**: ¿Cuál es el nivel de calidad de los Paquetes de Laboratorio en relación con las insignias de la ACM?

La pregunta de investigación **PI₃**: ¿Cómo se están reportando actualmente los Paquete de Laboratorio en la comunidad de IS? se ha concretado en las siguientes preguntas:

- **PI_{3.1}**: ¿Cuál es el nivel de calidad de los Paquetes de Laboratorio con respecto a la completitud de los módulos del estudio de Solari *et al.* [394]
- **PI_{3.2}**: ¿Qué guías utilizan los autores para reportar sus Paquetes de Laboratorio?
- **PI_{3.3}**: ¿Qué repositorios digitales se suelen utilizar para que los Paquetes de Laboratorio estén disponibles en la comunidad de IS?

En las siguientes secciones, se abordan cada una de las preguntas de investigación utilizando gráficos para visualizar los datos recopilados y presentando las interpretaciones del análisis de resultados.

3.2 ¿ES UNA PRÁCTICA HABITUAL EN LA COMUNIDAD IS HACER PÚBLICOS LOS PL?

Para responder a esta pregunta, se han considerado los dos tipos de foros establecidos: revistas y conferencias. De los 255 estudios primarios que pertenecen a las revistas (EMSE, IST, JSS, TSE), 120 estudios (47,06 %) no tienen PL, 135 estudios definen un PL (los autores indican que presentan un PL en el artículo científico) y sólo 95 estudios de ellos (70,38 %) tienen PL públicos. Del resto, en 33 estudios (24,45 %) el enlace web a los PL está roto y en 7 estudios (5,18 %) se requiere autorización del autor para acceder al PL. La figura 3.2 muestra un gráfico de columnas apiladas sobre el número de artículos que definen los PL relacionados con las cuatro revistas (EMSE, IST, TSE, JSS) y se observa que la EMSE es la revista que contiene el mayor número de estudios cuyos PL son públicos (25,19 %).

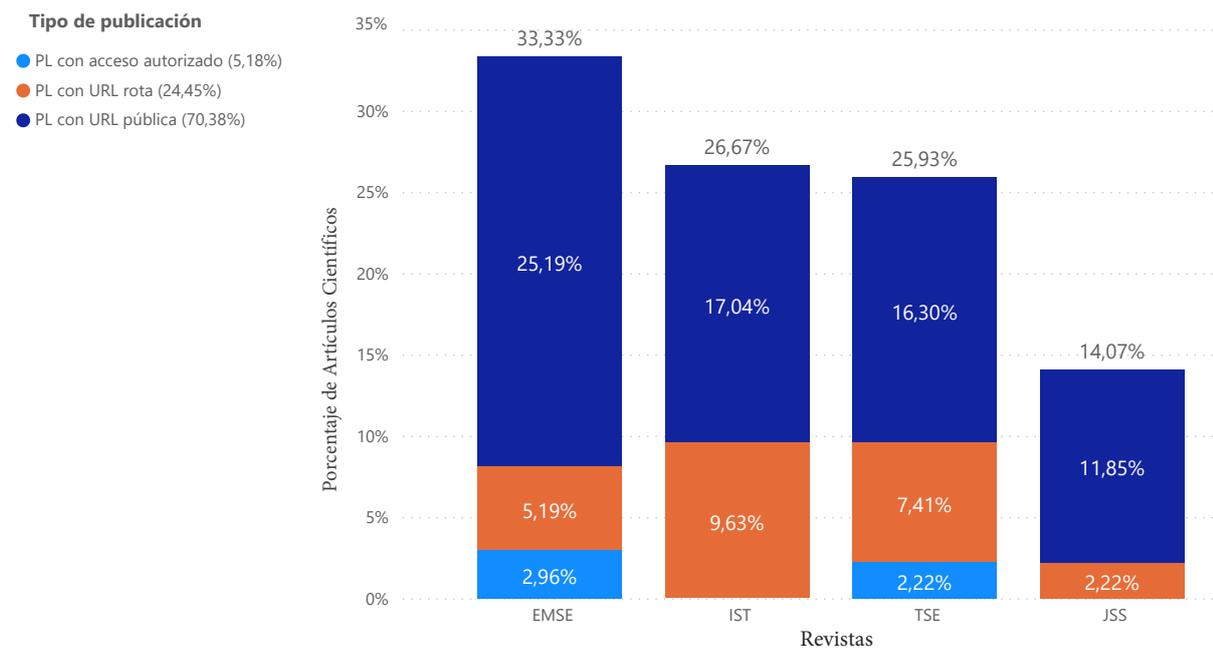


Figura 3.2: Artículos científicos que definen un Paquete de Laboratorio por revistas.

En lo que respecta a las conferencias con un Track AE en 2021 que reportan experimentos orientados a humanos, se informa de que 12 PL tienen un PL *público*. La figura 3.3 muestra la distribución del número de PL relacionados con las conferencias, por ejemplo, ICSE tiene 6 PL, ESEC/FSE tiene 3 PL, ASE tiene 2 PL y RE tiene 1 PL.

Además, se consideraron 255 estudios de las revistas seleccionadas para saber cómo

3.3. ¿CUÁL ES LA TENDENCIA EN LA PUBLICACIÓN DE PL TRAS EL OPEN SCIENCE?

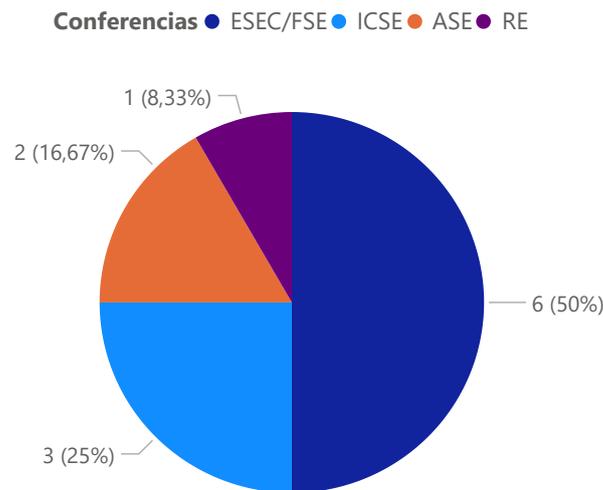


Figura 3.3: Paquete de Laboratorio por conferencias.

se describen los contenidos de los PL en el artículo científico. La figura §3.4 muestra que la revista EMSE es la que publica más artículos que profundizan en los contenidos de los PL en el mismo documento, es decir, que el artículo publicado explica con detalle los artefactos, materiales e información del proceso experimental. Por lo tanto, el número de páginas que contiene cada artículo se ha relacionado y el texto publicado se ha revisado en detalle para determinar el nivel de contenido de los paquetes *bajo* (1- 12 páginas), *medio* (13-25 páginas), *alto* (26 en adelante). Como ejemplo, en EMSE se observa que 70 (27,45%) de los artículos tienen entre 30 y 57 páginas; esto permite identificar que los autores explican con detalle el contenido de del paquete, principalmente en la sección de Anexos. Por otro lado, con un nivel *medio* de publicación del contenido del paquete en el propio documento del artículo, la revista IST tiene 57 (22,35%) artículos con entre 14 y 20 páginas publicadas. La figura §3.4 muestra la relación entre los artículos que profundizan en el contenido del paquete en el mismo documento que informa del estudio y las revistas científicas donde se publican el estudio.

A continuación, la tabla §3.1 presenta los 268 estudios primarios por revistas científicas y conferencias donde han sido publicados. El símbolo (*) representa los estudios actualizados en las cuatro revistas científicas.

Tabla 3.1: Estudios primarios publicados en revistas y congresos en 2000–2021 ordenados por número de estudios.

Revista/Congreso	Estudios primarios	# Estudios
Empirical Software Engineering (EMSE)	[14, 18, 24, 25, 26, 36, 70, 71, 72, 74, 75, 106, 118, 119, 137, 143, 147, 155, 169, 170, 174, 175, 180, 191, 192, 197, 203, 218, 225, 227, 228, 256, 258, 262, 266, 290, 312, 325, 333, 341, 344, 347, 361, 362, 367, 383, 407, 409, 420, 422, 426, 433, 438, 442]	54
Empirical Software Engineering (EMSE) *	[4, 16, 20, 109, 113, 142, 165, 184, 198, 238, 242, 277, 302, 360, 364, 366, 385, 414, 415, 417, 437]	21
Information and Software Technology (IST)	[7, 8, 28, 54, 73, 79, 81, 83, 95, 105, 133, 145, 149, 150, 159, 162, 173, 181, 183, 188, 204, 207, 209, 214, 254, 257, 261, 263, 286, 292, 293, 296, 297, 299, 300, 303, 305, 310, 313, 322, 340, 346, 357, 358, 371, 387, 390, 396, 403, 404, 429, 446]	52
Information and Software Technology (IST) *	[15, 17, 29, 110, 116, 117, 123, 126, 132, 151, 160, 182, 208, 217, 234, 239, 248, 250, 264, 301, 326, 335, 386, 412]	24
Journal of Systems and Software (JSS)	[47, 51, 53, 80, 84, 115, 122, 130, 135, 148, 205, 212, 213, 216, 233, 235, 240, 255, 270, 276, 280, 306, 307, 330, 359, 398, 440]	27
Journal of Systems and Software (JSS) *	[9, 37, 49, 56, 101, 129, 134, 206, 215, 247, 279, 288, 289, 309, 315, 319, 370, 374, 402, 410, 427]	21
IEEE Transactions on Software Engineering (TSE)	[5, 27, 30, 31, 32, 39, 61, 64, 65, 78, 88, 100, 108, 128, 140, 154, 211, 230, 237, 252, 265, 278, 311, 318, 320, 328, 329, 331, 332, 338, 345, 405, 408, 421, 432]	35
IEEE Transactions on Software Engineering (TSE) *	[19, 50, 57, 131, 136, 144, 161, 226, 245, 249, 273, 283, 285, 314, 348, 350, 368, 395, 413, 443, 445]	21
ICSE' 2021	[62, 99, 317]	3
ESEC/FSE' 2021	[139, 196, 246, 373, 377, 428]	6
RE' 2021	[166, 324]	2
ASE' 2021	[308, 378]	2
Total		268

3.3. ¿CUÁL ES LA TENDENCIA EN LA PUBLICACIÓN DE PL TRAS EL OPEN SCIENCE?

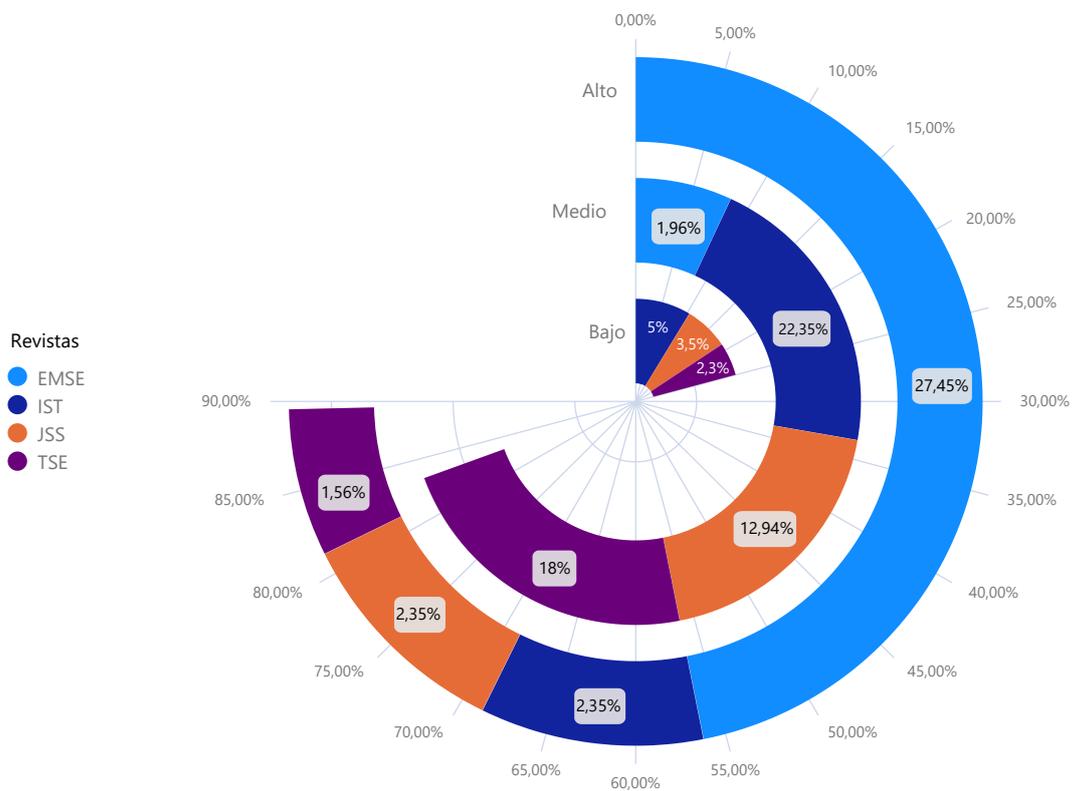


Figura 3.4: Artículos científicos que profundizan el contenido del Paquete de Laboratorio por revistas.

3.3 ¿CUÁL ES LA TENDENCIA EN LA PUBLICACIÓN DE PL TRAS EL OPEN SCIENCE?

En primer lugar, se destaca algunas de las iniciativas promovidas por las revistas que han propiciado el aumento del número de PL. EMSE [11] motiva y destaca la iniciativa del *Open Science* para hacer públicos los datos de investigación y aumentar la transparencia, reproducibilidad y replicabilidad de los estudios. Esto apoya la credibilidad inmediata del trabajo de los autores y proporciona una base cotidiana para los esfuerzos conjuntos de la comunidad basados en datos compartidos. Además, construye un cuerpo general de conocimientos en la comunidad que conduce a teorías de la IS para su aplicabilidad [26].

IST [12] y JSS [13] apoyan la iniciativa de *Open Science* ofreciendo a sus autores un servicio gratuito de publicación de preimpresos en el SSRN (Open Access Comunidad de Preprints Online) servidor que funciona como punto de partida para la publicación gratuita de las primeras investigaciones. Los preprints proporcionan un registro y una

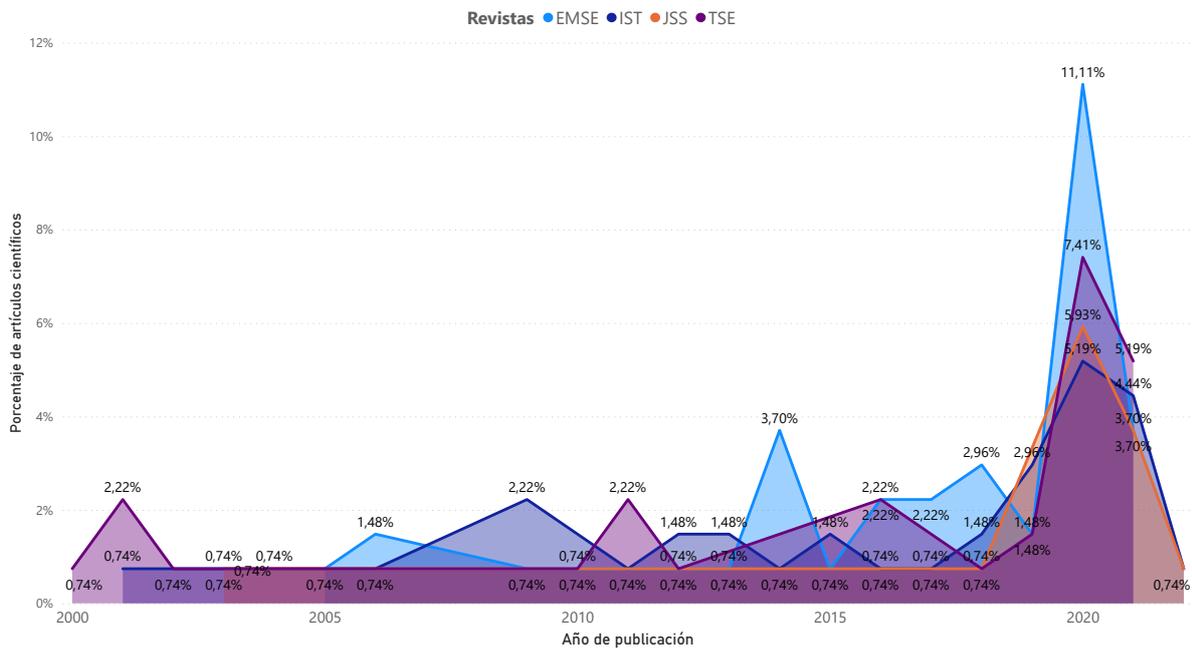


Figura 3.5: Evolución de la publicación de Paquetes de Laboratorio desde 2000 hasta 2021 por revista.

difusión temprana de la investigación, facilitando citación y colaboración tempranas [57]. Además, otras revistas como la TSE [14] apoyan el capítulo del “Comité ad hoc 2020 sobre ciencia abierta y Reproducibilidad” [58], para analizar modelos, prácticas y experiencias en el apoyo a *Open Science* y la Reproducibilidad dentro de la IEEE Computer Society (CS) y en las sociedades homólogas y editores en el contexto de la recomendación del informe de la NASEM titulado “Reproducibilidad y replicabilidad en Science” [59].

De los 135 estudios de las revistas seleccionadas que definen un PL, se observa que hay una media de seis estudios por año que publican un PL, por lo que no se observa una variación significativa debido al bajo número de estudios que publican su PL. Sin embargo, en 2020, hubo un aumento considerable de 40 estudios (29,64%) en la publicación de PL en las cuatro revistas (EMSE, IST, JSS, TSE). Un factor que puede haber influido en este elevado número es la motivación del movimiento *Open Science* desde los diferentes organismos y eventos científicos; por ejemplo, se han abierto conferencias con mayor énfasis en Track o sesiones exclusivas para evaluar PL e incentivar la calidad de los mismos mediante la concesión de insignias ACM [3]. Dentro del análisis de la tendencia de las publicaciones de PL, se ha identificado que era necesario el proceso de actualización del estudio de Revoredó *et al.* [55]. Como resultado, el número

de estudios publicados en 2020 y 2021 supera con creces los años anteriores. La figura §3.5, muestra la tendencia de las publicaciones de los estudios con un PL definido, entre el 2000 a 2021 en relación con las revistas seleccionadas. La línea de tiempo es cada cinco años, y los porcentajes se calculan en función del número de estudios publicados cada año.

3.4 ¿CUÁL ES EL NIVEL DE CALIDAD DE PL RESPECTO A LAS INSIGNIAS ACM?

En esta pregunta, ha sido necesario analizar por separado los PL de los estudios de las revistas seleccionadas y los PL de los AE Tracks de las conferencias. En lo que respecta a los estudios de las revistas, se ha utilizado el documento de las insignias de la ACM [185] que evalúa los artefactos, y la guías del ICSE'2021 [186] para la presentación y revisión del AE Track para identificar las insignias en cada PL. De los 147 PL analizados, sólo 95 estudios que pertenecen a las revistas seleccionadas tienen PL *público*. Al analizar los 95 PL de los estudios, fue necesario categorizar la insignia debido al variado contenido que los autores publican en el paquete para establecer qué insignia ACM alcanzaría cada PL. Para la ACM, las insignias son: *disponible*, *funcional*, *reutilizable*, *reproducido* y *replicado*. Cabe señalar que, según la ACM, en el caso de *funcional* y *reutilizable* sólo se puede aplicar una de las insignias, es decir, son mutuamente excluyentes.

Sin embargo, hay un número importante de PL que no completan las características de la insignia *funcional* y se ha considerado un nivel *parcial* que indica que sólo una parte del contenido se detalla en el PL y debe tenerse en cuenta para la evaluación y la validación del PL. Del mismo modo, se ha considerado un nivel *parcial* para la insignia *disponible*. En el caso de la insignia *reutilizable*, no es necesario considerar este nivel *parcial* ya que se puede otorgar la insignia *funcional*. Tampoco es necesario considerar el nivel *parcial* para las insignias *replicado* y *reproducido*.

Se ha identificado que a 84 PL (88,42%) se les asignaría la insignia *disponible*. Los 11 PL restantes (11,58%) sólo alcanzan el nivel *parcial* de la insignia *disponible*. En cuanto a las insignias *funcional* y *reutilizable*, 44 PL (46,31%) sólo obtendrían la insignia *funcional* a nivel *parcial*, 28 PL (29,47%) la insignia *funcional* y 16 PL (16,84%) la insignia *reutilizable* han sido asignados. El resto, 7 PL (7,36%) no recibirían la insignia *funcional* tampoco la insignia *reutilizable*. Aunque las insignias de *replicado* y *reproducido* son las

menos asignadas, el análisis muestra que 2 PL (2,10%) son reproducidos y 1 PL (1,05%) es replicado. Con estos resultados, se determina la necesidad de mejorar el proceso de elaboración y publicación de PL para apoyar el proceso de replicación en el área de IS. La Tabla §3.2, muestra las insignias ACM de los PL relacionados con los 95 estudios de las revistas, considerando el nivel *parcial* para *funcional* y la insignia *disponible*. El porcentaje presentado en la última columna de la Tabla §3.2, se muestra un resumen de forma gráfica en la Figura §3.6. De los 147 PL analizados, 12 PL pertenecen a las conferencias seleccionadas. Se reporta la insignia que los evaluadores han otorgado a los 12 PL en la presentación de los Track AE. La insignia *reutilizable* se concedió a 6 PL (50%). La insignia *disponible* se concedió a 7 PL (58,32%). Las insignias *funcional*, *reproducido* y *replicado* no se da a ningún PL. En el caso de los 3 PL que corresponden a ICSE'2021 no reciben la insignia *funcional* porque esta insignia se incluye cuando el PL es *reutilizable*. En cambio, el resto de los PL correspondientes a ESEC/FSE'2021 [141], RE'2021 [339] y ASE'2021 [35] la insignia *funcional* no se considera en la lista de categorías de insignias. La Figura §3.7 muestra las insignias ACM relacionadas con los 12 PL de las conferencias.

Tabla 3.2: ACM Badging para Paquetes de Laboratorio por revistas.

ACM Badge	Nivel	EMSE	IST	JSS	TSE	Total	Porcentaje
Disponible		34	23	16	22	95	100%
	Completo	27	19	16	22	84	88.42%
	Parcial	7	4	0	0	11	11.58%
Funcional		23	20	12	17	72	75.78%
	Completo	9	8	3	8	28	29.47%
	Parcial	14	12	9	9	44	46.31%
Reutilizable		11	1	0	4	16	16.84%
	Completo	11	1	0	4	16	16.84%
	Parcial	0	0	0	0	0	0%
Reproducido		1	1	0	0	2	2.10%
	Completo	1	1	0	0	2	2.10%
	Parcial	0	0	0	0	0	0%
Replicado		1		0	0	1	1.05%
	Completo	1	0	0	0	1	1.05%
	Parcial	0	0	0	0	0	0%

3.5. ¿CUÁL ES EL NIVEL DE CALIDAD DE PL RESPECTO A LOS MÓDULOS DE SOLARI?

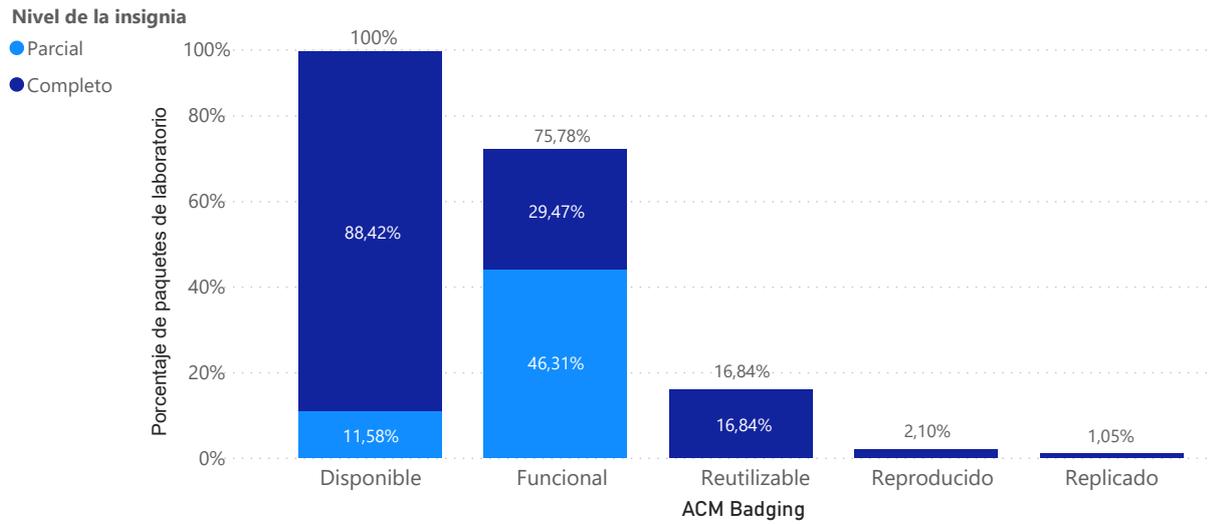


Figura 3.6: Insignias de la ACM para Paquetes de Laboratorio.

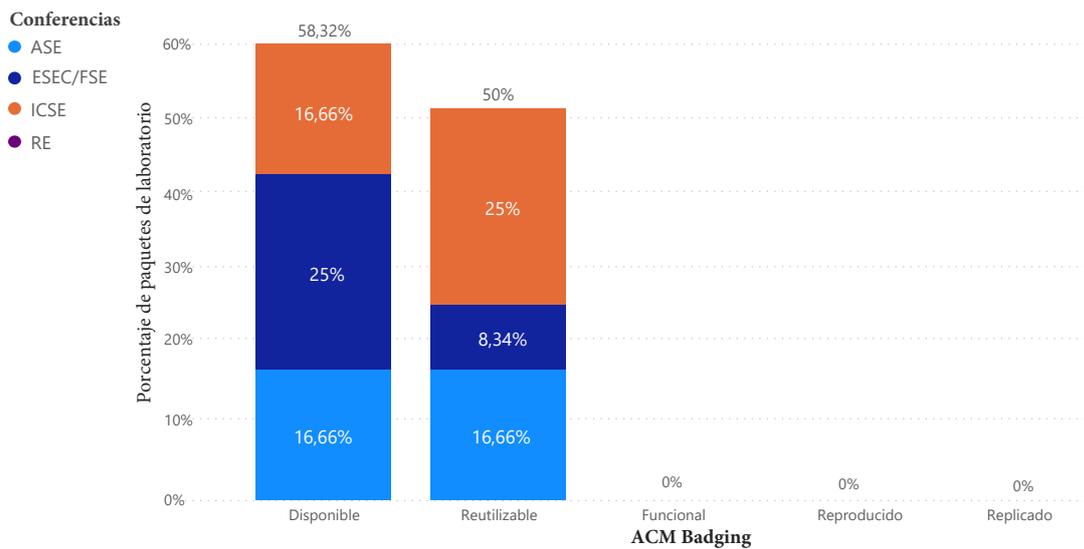


Figura 3.7: Insignias de la ACM para paquetes de laboratorio por conferencias.

3.5 ¿CUÁL ES EL NIVEL DE CALIDAD DE PL RESPECTO A LOS MÓDULOS DE SOLARI?

En esta pregunta, se ha estudiado la completitud de los PL respecto a la propuesta de Solari *et al.* [394] sobre los contenidos del PL para los experimentos orientados al ser humano en el área de IS. Siete módulos clasifican dicho contenido: *i)* experimento, *ii)* entrenamiento, *iii)* introducción, *iv)* teoría, *v)* agregación, *vi)* replicación y *vii)* evolución. Cada módulo contiene artefactos que han sido cuidadosamente desarrollados

para que los investigadores puedan utilizarlos y entender el experimento o replicarlo. Se ha considerado 95 PL públicos de los estudios de las revistas seleccionadas y 12 PL de las conferencias seleccionadas, en total 107 PL. Es decir, de los 147 PL analizados, sólo 107 son públicos. En el análisis de cada PL, el contenido del PL debía ser categorizado en dos niveles: *completo* cuando el PL incluye todo el material que Solari *et al.* [394] recomendó y *parcial* cuando los autores comparten una parte del contenido de los artefactos identificados por Solari *et al.* [394].

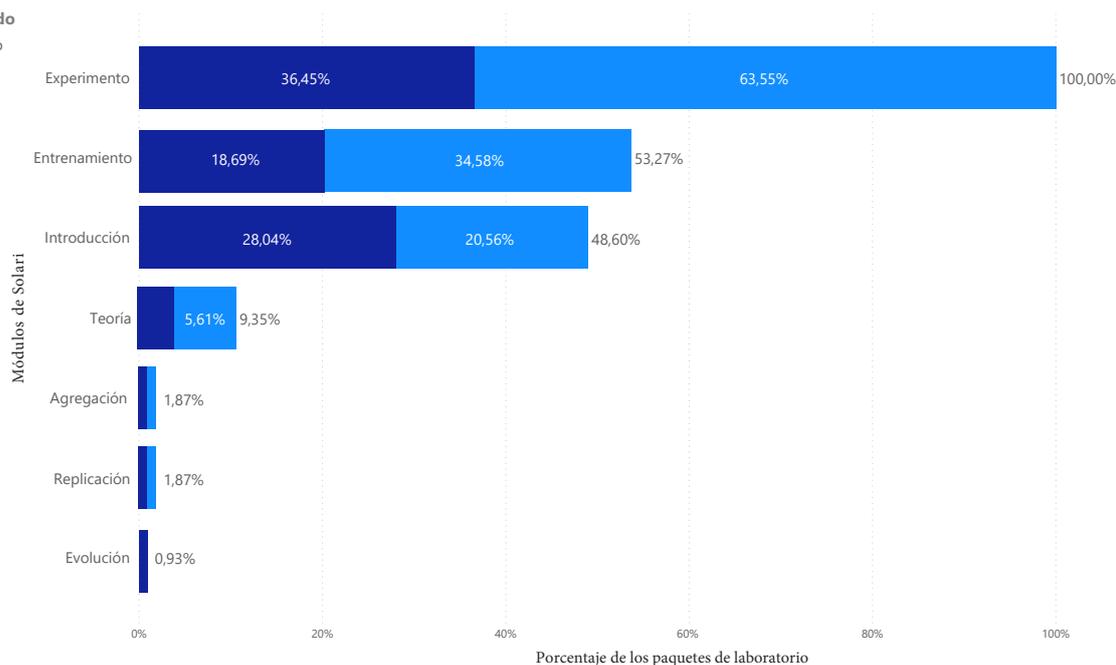


Figura 3.8: Contenido del Paquete de Laboratorio clasificado en módulos propuestos por Solari *et al.* [394].

Como se puede observar en la Figura 3.8, dentro del conjunto de módulos de Solari *et al.* [394], tres son los que más frecuentemente se ponen a disposición de la comunidad: experimento, entrenamiento e introducción. Sin embargo, los módulos de agregación, replicación y evolución sólo aparecen en los PL pertenecientes a la replicación o a una familia de experimentos, no a un experimento original. Dado que la mayoría (67,29%) de los PL analizados reportan material de experimentos originales, los tres módulos mencionados son menos comunes. Por otro lado, el módulo de teoría es raramente reportado porque los autores reportan muy pocos modelos conceptuales, diagramas, referencias bibliográficas, descripciones de las técnicas y fórmulas utilizadas en el experimento. En los PL analizados, todos los PL cubren el módulo de experimentación, ya sea *parcial* o *completo*. El contenido más común subido por los autores relacionado con el módulo de experimento es el material operativo utilizado en

el experimento, por ejemplo, formularios, recogida de datos, casos de prueba, cuestionarios y resultados de análisis estadísticos. Ejemplos de paquetes de módulos de experimentación son [383], [328]. Por otro lado, en el módulo de entrenamiento, hay un porcentaje de trabajos que no publican nada relacionado con el entrenamiento, debido a dos factores, el primero porque los experimentos no necesitan ejecutar tareas de entrenamiento para los sujetos y el segundo, porque los experimentadores no suben artefactos relacionados aunque realicen tareas de entrenamiento.

En el caso de publicar el módulo de formación, lo más habitual es compartir presentaciones teóricas y guías de formación, registro de ejercicios preliminares antes de ejecutar el experimento y registro de sesiones de feedback. Algunos ejemplos de paquetes de módulos de formación son [347], [142]. Del mismo modo y por las mismas razones que el módulo anterior, el módulo de introducción presenta un porcentaje de trabajos que no publican nada relacionado con la introducción.

En el caso de publicar el módulo de introducción, lo más habitual es que los experimentadores suban un registro de la estructura organizativa del PL, los instrumentos de comunicación entre los experimentadores y la información de contacto de los autores del experimento original y de la réplica. Algunos ejemplos de paquetes de módulos de introducción son [264], [147]. Se ha encontrado un trabajo de Basili *et al.* [42] cuyo PL completa los módulos y su contenido ha sido utilizado por otros investigadores.

3.6 ¿QUÉ GUÍAS UTILIZAN LOS AUTORES PARA REPORTAR SUS PL?

Se ha estudiado las guías proporcionadas por las revistas y los Track AE de las conferencias, para que los experimentadores puedan desarrollar sus PL. Tras analizar las guías de los foros estudiados en este trabajo, tanto en las revistas como en las conferencias, la Tabla §3.3 resume las principales conclusiones. En cuanto a las revistas, cabe destacar que: i) las guías son muy generales; ii) se centra en el material abierto; iii) los editores de las revistas tienen páginas web en las que hacen referencia a la iniciativa *Open Science* y a su relación con la reproducibilidad; iv) una de las revistas -EMSE- trabaja con un distintivo y otorga un premio llamado *Open Science badge*. Al estudiar los libros blancos publicados por EMSE relacionados con *Open Science*, se identificó la intención de esta revista en adoptar el sistema de insignias ACM en un futuro próximo.

Las conferencias, por su parte, siguen la iniciativa del ACM Badging con algunas

Tabla 3.3: Guías para Paquetes de Laboratorio y Open Science en revistas y conferencias seleccionadas para este estudio.

Foros	Año propuesta	Open Data	Open Material	Open-Source Software	Open Access	Badging	Prescripciones
TSE	2020	✓	✓	✓	✓	-	Web site [304]
EMSE	2019	✓	✓	✓	✓	Open Science badge	Web site [199]
IST	2021	-	✓	-	✓	-	Web site [138]
JSS	2021	-	✓	-	✓	-	Web site [138]
ICSE	2021	✓	✓	✓	-	ACM	AE Track [12]
ASE	2021	✓	✓	✓	-	ACM	AE Track [10]
ESEC/FSE	2021	✓	✓	✓	-	ACM	AE Track [11]
RE	2021	✓	✓	✓	-	ACM	AE Track [13]

variaciones. En el ICSE'2021 AE Track [186] se han seguido las especificaciones de la ACM, aunque se han definido criterios específicos para cada una de las insignias. En concreto: (i) no se considera la insignia *funcional*; y (ii) se requiere una insignia *reutilizable* para obtener la insignia *disponible*. En cuanto a ESEC/FSE'2021 [141], no se han encontrado diferencias en la aplicación de las insignias con respecto a las especificaciones de ACM. A diferencia del ICSE'2021 AE Track, se considera la insignia *funcional* y no es necesario que los artefactos hayan sido evaluados formalmente (insignias *funcional* y *reutilizable*) para recibir la insignia *disponible*.

Tras analizar detenidamente los 147 PL, se observa que no siguen totalmente los consejos propuestos en las prescripciones de *Open Science* de la editorial de la revista correspondiente. En el caso de los PL de las conferencias, los autores siguen las pautas ofrecidas en el AE Track pero los contenidos de los PL presentan una estructura bastante diferente y en algunos casos, parecen estar incompletos a pesar de la disponibilidad de las guías de presentación de artefactos para los autores. Por otra parte, el uso o la referencia a cualquiera de las guías conocidas para desarrollar PL, como la de Basili *et al.* [41] o la de Solari *et al.* [394], también se echan de menos.

En conclusión, a pesar de la existencia de propuestas completas y correctas para elaborar el PL, existe cierta resistencia a su uso por parte de los autores. Timperley *et al.* [406] destacan la percepción de que el esfuerzo de crear y compartir artefactos no merece la pena debido, entre otras cosas, a la falta de normas para la revisión de artefactos y a la necesidad de que los investigadores se encarguen del mantenimiento continuo del PL, junto con el temor a que los experimentos sean evaluados y los resultados puedan verse comprometidos. Una posible solución podría ser desarrollar

3.7. ¿QUÉ REPOSITORIOS DIGITALES SE USAN PARA QUE LOS PL ESTÉN DISPONIBLES?

el PL a medida que se realiza el estudio. En este sentido, detectamos como nicho de investigación la necesidad de desarrollar propuestas encaminadas a proporcionar procedimientos sistemáticos que complementen las directrices para la elaboración de PL y que se apoyen en sistemas de información, con el fin de evitar que los PL se desarrollen mediante procedimientos ad-hoc.

3.7 ¿QUÉ REPOSITORIOS DIGITALES SE USAN PARA QUE LOS PL ESTÉN DISPONIBLES?

De los 147 PL analizados en esta pregunta, 107 PL son públicos y pertenecen a las revistas y conferencias seleccionadas. Se ha encontrado 13 repositorios digitales donde los autores han publicado su PL. Se observa que la mayoría de los PL están publicados en el repositorio digital de *GitHub* 36 estudios (33,64%), seguido de 30 estudios (28,03%) en *Repositorios Institucionales* y 15 estudios (14,01%) en *Zenodo*. En el caso de *Zenodo*, se destaca que cuenta con más de 10.000 comunidades creadas por los usuarios, siendo una de ellas *Empirical Software Engineering* [441].

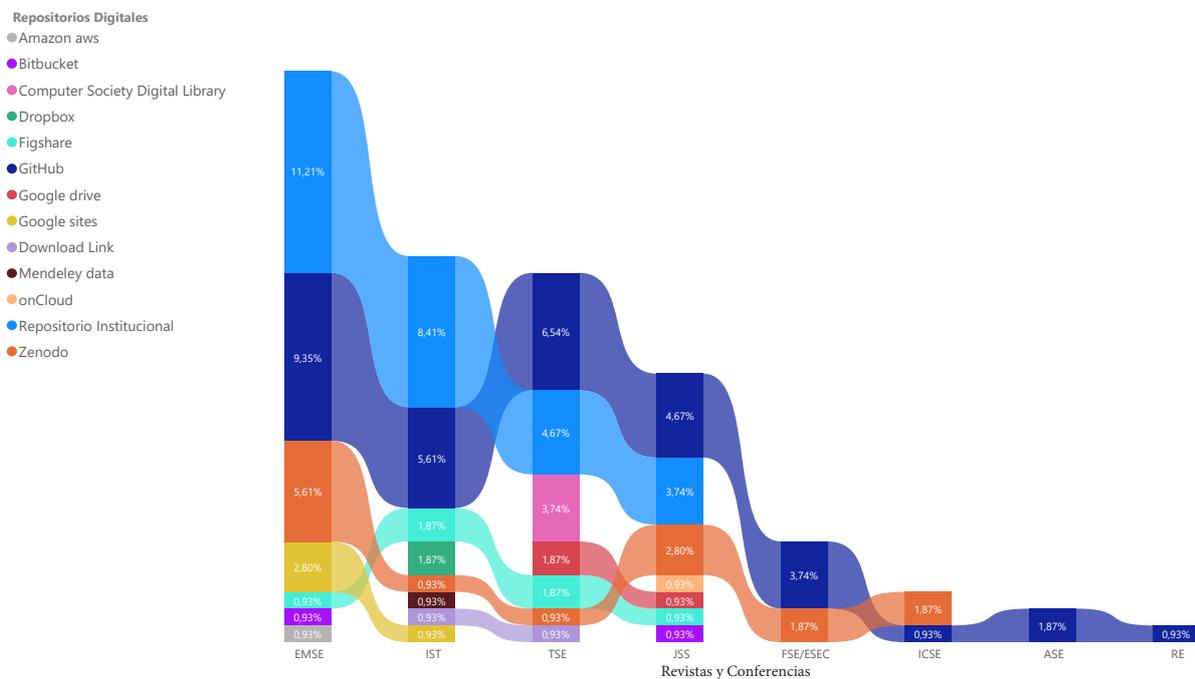


Figura 3.9: Repositorios digitales que contienen el Paquete de Laboratorio publicados en revistas y congresos

Esta comunidad es específica para compartir datos, código y artículos sobre el conocimiento empírico en IS. El aumento de los repositorios digitales disponibles actualmente en Internet facilita la subida y publicación de PL, concretamente la motivación de la comunidad científica para la adopción de la *Open Science*.

En dos revistas, EMSE e IST, predomina el *Repositorio Institucional* con un 19,62%, seguido de *GitHub* con un 14,96%. EMSE es la revista donde se suben los PL en mayor proporción con un 45,78%, seguida de IST con un 21,48%. En cuanto a las conferencias, la mayoría de los PL se suben a *GitHub* con un 7,47%. Entre los repositorios específicos que han surgido en los últimos años para *Open Science*, *Zenodo* es el más utilizado con un 14,01%, seguido de *Figshare* con un 5,6% según los datos reportados en nuestro estudio. La figura §3.9 muestra la lista de repositorios digitales que los autores utilizan para publicar PL con el respectivo análisis.

3.8 CONCLUSIONES

En esta sección se presentan algunas conclusiones obtenidas como respuesta a las preguntas de investigación:

- **PI_{1.1}**. Se ha encontrado que el número de estudios publicados en la revista EMSE 34 (25,19%) tienen el Paquete de Laboratorio *Disponible*, seguido de la revista IST 23 (17,04%) de los estudios publicados. La revista EMSE es donde más estudios con Paquete de Laboratorio *Disponible* se encuentra. Con respecto a las conferencias se identificó que el ICSE'2021 es el que reporta en su mayoría 6 PL. La revista EMSE es donde se publican más artículos científicos que profundizan el contenido de los PL en el mismo artículo con el 27,45% los mismos que tienen entre 30 y 57 páginas.
- **PI_{1.2}**. En el año 2020 es donde existe un significativo aumento 40 (29,64%) de experimentos controlados que publican su PL en el área de IS. Para el año 2021 la tendencia sigue en aumento según el análisis se observa que existe mayor importancia en publicar los PL posiblemente motivado por las revista y congresos que han afianzado el uso de *Open Science*. Se ha observado que en los últimos cinco años la evolución de publicar PL tiene mayor crecimiento 86 (63,70%) de paquetes publicados. Eso pone en manifiesto la importancia de replicar experimentos para validar los resultados del experimento original y aumentar el cuerpo de conocimiento.

- **PI_{1.3}**. Con respecto al nivel de calidad de los PL en relación con las insignias de la ACM para las revistas seleccionadas, el 88,42% alcanzarían la insignia *disponible*. Las insignias *funcional* y *reutilizable* sólo obtendrían a nivel *parcial* el 46,31%. Con respecto a los congresos, la insignia *funcional* no es asignada en el Track AE de las conferencias. La insignia *reutilizable* y la insignia *disponible* son las que en su mayoría se asignan.
- **PI_{3.1}**. Con respecto al nivel de completitud del contenido del PL relacionado a los módulos del estudio de Solari *et al.* [394], los módulos *experimento*, *entrenamiento* e *introducción* son los que mayor se reportan (67,29%). Se concluye que existe un porcentaje muy bajo de estudios que reportan los módulos de: *agregación*, *replicación* y *evolución*, por lo cual, es necesario que los autores reporten estos módulos cuando realizan replicaciones experimentales.
- **PI_{3.2}**. En cuanto a las guías y directrices que los autores usan para reportar PL de experimentos controlados en el área de IS, se concluye que los congresos en el Track AE-Artifact Evaluation han elaborado una guía de contenido y subida del PL, muy específica para cada congreso por ejemplo, en el ICSE'2021 el documento guía de Abrahão y Mendez [6]. Otros congresos como ESEC/FSE'2021 también ponen a disposición una extensión de la guía propuesta por Abrahão y Mendez, el congreso RE'2021 y ASE'2021 no presenta una guía pero si detalla en el sitio web las políticas y directrices para la evaluación de artefactos. Para las revistas, las guías son muy generales y se centran en el material abierto, además hacen referencia al apoyo de la iniciativa de Open Science.
- **PI_{3.3}**. Con la apertura y motivación del movimiento *Open Science* la comunidad científica ha presentado cierto interés en publicar contenido científico. En este caso se ha analizado la publicación de los PL para experimentos controlados en el área de IS. En el presente análisis se ha identificado 13 repositorios digitales, entre los cuáles, el repositorio de *GitHub* 36 estudios (33,64%) tiene la mayoría de publicaciones, seguido de los *Repositorios institucionales* 30 estudios (28,03%) y de *Zenodo* con 15 estudios (14,01%). Durante la crisis sanitaria de la Pandemia por el Covid-19 es posible que el acceso a la información y la facilidad de comunicación de los investigadores haya sido causa del uso frecuente de herramientas *Open Science*, además la motivación de revistas y congresos del área de IS en su uso y aplicación para compartir y evaluar contenido científico.

3.9 RESUMEN

Para conocer el estado actual de la disponibilidad de los PL en IS, se han analizado 268 artículos científicos, publicados entre 2000 y 2021 en las revistas (EMSE, IST, JSS y TSE) y congresos (ICSE '2021, ESEC/FSE '2021, RE '2021, ASE '2021) en el área de IS. Dichos artículos reportan experimentos controlados con humanos. Se han identificado: *i*) principales foros donde se han publicado experimentos controlados con sus respectivos PL o incluyen un Track AE-Artifact Evaluation; *ii*) estudios que han sido asignados con alguna insignia de la ACM referente al PL; *iii*) evolución en el tiempo de los estudios que publican PL; *iv*) guías y directrices utilizadas en el reporte de PL; *v*) módulos de contenido del PL que publican los autores con mayor frecuencia; *vi*) módulo de contenido del PL que los autores publican dentro del artículo científico; y *vii*) repositorios digitales *Open Science* que los autores usan con más frecuencia para reportar los PL.

Según los resultados, los foros más relevantes son las revistas *Empirical Software Engineering* (EMSE) y *Information and Software Technology* (IST).

Las insignias que la mayoría de los PL presentan son *Disponibles* y *Funcionales*. Las insignias *Reproducido* y *Replicado* son las que menos se evidencian en el análisis de los estudios.

Durante el periodo de los años 2000 al 2019 la evolución en el número de estudios que publican PL en experimentos controlados con humanos no presentan mayor variación, sin embargo en a partir del 2020 se ha observado una tendencia con mayor crecimiento, posiblemente por el impacto de la pandemia del Covid-19 debido al cambio de la comunicación virtual al igual que la motivación del movimiento *Open Science* en la comunidad científica.

No se evidencia que los estudios analizados sigan una guía especializada para elaborar PL experimentales. Destacando que para los paquetes que se presentan en los Track AE-Artifact Evaluation de los congresos y algunas revistas en el área de IS siguen las guías que cada foro dispone.

Los módulos del contenido del PL que mayor reportan los autores es el módulo de *experimento, entrenamiento e introducción*. Estos módulos están basados en el estudio de Solari *et al.* [394].

El contenido del PL se reporta en algunos casos dentro de la descripción del artículo científico como las tareas y material experimental. En el presente análisis se ha identificado que en la revista EMSE es donde más se publica artículos que profundizan

contenido del PL seguido de la revista IST con menos cantidad.

GitHub es el principal repositorio digital que los autores han publicado el PL en el presente estudio, la consecuencia posiblemente de que los autores escogen *GitHub* es para publicar y compartir código fuente y demás artefactos que pertenecen al paquete de instalación. Segundo, esta los *Repositorios institucionales* por ejemplo en las instituciones académicas como la Universidad de Maryland [42], Universidad de Castilla-La Mancha [107], Universidad Tecnológica de Chemnitz [143] entre otras. Y como tercer repositorio que utilizan se encuentra *Zenodo*.

PROBLEMÁTICA EN LA ELABORACIÓN DE PAQUETES DE LABORATORIO

No importa lo lento que vayas mientras no te detengas

*Confucio (551 a. C. – 479 a. C.),
Filósofo, escritor, pensador y maestro chino*

En este capítulo, respondiendo a la pregunta de investigación **PI₂**, se analizan algunos de los problemas detectados al abordar los PL de experimentos, y que contribuyen a que sean pocos los experimentos replicados.

En la sección §4.1, se destaca el bajo número de PL publicados. En la sección §4.2, se señala el poco uso de guías específicas para reportar PL. En la sección §4.3, se remarca la carencia de PL para facilitar las replications. En la sección §4.4, se señala el problema de la transmisión del conocimiento. En la sección §4.5, se presenta el problema de la falta de acuerdo en la terminología utilizada por los autores. En la sección §4.6, se describe la necesidad de clarificar el contexto de las replications. En la sección §4.7, se advierte sobre la dificultad añadida del factor humano en las replications. Por último, la sección §4.8 resume los problemas identificados.

4.1 INTRODUCCIÓN

Si bien realizar experimentos es una actividad que en el área de IS se realiza cada vez con mayor frecuencia, algunas tareas involucradas en el proceso experimental también conlleva tres efectos poco deseables: *i*) aumento considerable del tiempo invertido en tareas que aún siguen siendo manuales, aunque sean pequeñas tomas de decisión, *ii*) aumento considerable del riesgo de cometer errores en el proceso y *iii*) aumento del grado de conocimiento y experiencia necesarios por parte de los investigadores sobre experimentación en general y sobre estadística en particular [219].

Los primeros estudios que publican PL en IS son relativamente recientes. Entre los primeros artículos que reportan un PL de un estudio empírico fue publicado en 1995 por Kamsties *et al.* [210]. Más tarde, Basili *et al.* [41] en 1996 publica un manual de laboratorio que incluye el diseño experimental, los instrumentos para su ejecución y los procedimientos para las pruebas estadísticas de un experimento controlado para evaluar la eficacia de una técnica de lectura de software conocida como lectura de requisitos basada en la perspectiva (PBR).

A pesar de la importancia de publicar PL para posiblemente aumentar las repeticiones [41], el número de repeticiones y a su vez de publicaciones de PL en IS sigue siendo bajo (Solari *et al.* [394]). La falta de repeticiones *externas* que son las de mayor poder de confirmación es evidente. Sin repeticiones externas los resultados son sólo provisionales [67, 68]. A este problema se ajusta la necesidad de publicar PL válidos que tengan contenido disponible, completo y de fácil acceso para el uso y aplicación de los investigadores, como hojas de cálculo, datos acerca del análisis estadístico, dataset, guía del proceso de replicación, formularios, herramientas de software, herramientas estadísticas, inclusive vídeos, audio, fotos, y demás objetos digitales que ayude a bajar la incertidumbre del replicador. Sin embargo, existe limitaciones en la estandarización de la estructura y evaluación de un PL especialmente para el área de IS, esto hace aún más complejo el proceso de replicación [394].

En las políticas de evaluación de artefactos experimentales de ACM [185] se explica que en el campo biomédico existe estudios que publican métodos inadecuados, análisis estadísticos incorrectos e incluso se han detectado casos de fraude. En otras áreas de conocimiento distintas a la IS como Psicología pero que se tiene una estrecha relación por la experimentación con humanos, esta situación se conoce como *crisis de replicabilidad* [55].

Por otra parte, los problemas identificados que pueden contribuir al bajo número de publicación de los PL, se pueden resumir en:

- *Poco uso de guías para reportar los Paquetes de Laboratorio* [394]). Esto se justifica en la sección §3.6.
- *Carencia y debilidades de los Paquetes de Laboratorio* que facilitan las replicaciones [394].
- *Dificultad para transmitir el conocimiento tácito* [380].
- *Falta de acuerdo sobre la terminología*, los términos y conceptos relativos a los PL [394].
- *Dificultad para clarificar el contexto de los experimentos y las replicaciones.*
- *Complejidad de utilizar sujetos humanos en los experimentos originales y replicaciones.*
- *Demasiado esfuerzo, tiempo, coste y recursos* para elaborar, publicar y evaluar el PL. Además, este trabajo suele realizarse una vez que ya el investigador ha preparado el experimento, lo ha ejecutado, analizado e incluso ha escrito un artículo para darle difusión. En consecuencia, el investigador se encuentra con ganas de finalizar el trabajo en vez de comenzar a construir un buen PL.

En las siguientes secciones se analizan, en mayor detalle, cada uno de estos problemas.

4.2 POCO USO DE GUÍAS PARA REPORTAR PAQUETES DE LABORATORIO

El poco uso de *guías* para reportar PL dificulta la comprensión tanto del experimento original y de la replicación a reportar, como de las replicaciones previas. Solari *et al.* [394] señala la necesidad de *guías* y directrices para garantizar la información que se publica sobre el experimento original y cada replicación.

Para Almqvist [21] la necesidad de elaborar reportes de PL es una posibilidad de simplificar el proceso de replicación del experimento, ayuda a los investigadores a bajar los niveles de incertidumbre que siempre existen cuando se empieza al ejecución de un experimento. En su estudio realizado en el 2006 agrupa en 20 series de experimentos a los estudios encontrados luego de una búsqueda manual en revistas y actas de

congresos centrada en replications en IS, para clasificarlos en replications internas y externas. Posteriormente, analizar y confirmar los resultados del estudio original. Entre las principales recomendaciones está la de mejorar la presentación de informes experimentales, elaborar *guías* y directrices para llevar a cabo las replications y disponer de PL para promover la replicación. La problemática asociada a la carencia de PL se verán en el siguiente apartado §4.3.

En la comunidad científica se conoce de *guías* y directrices propuestas por algunos investigadores para reportar experimentos controlados en IS como: Jedlitschka *et al.* [195]), Juristo *et al.* [201]), Wohlin *et al.* [435]). De la misma manera se conoce de *guías* y directrices para reportar los PL de experimentos controlados en IS como: Basili *et al.* [41], Shull *et al.* [381, 382], Almqvist [21], Kitchenham *et al.* [221], Vegas *et al.* [416], Solari *et al.* [393, 394], Jedlitschka *et al.* [194]. Sin embargo, las *guías* para elaborar PL son poco usadas en las investigaciones. En la sección §2.4 de la presente memoria de tesis se resume y se compara cada una de dichas *guías* y directrices para PL. En los resultados de la revisión literaria del presente estudio, en la sección §3.6, se explica que ningún estudio hace uso de *guías* o directrices al respecto.

La propuesta de Solari *et al.* [394] presenta una plantilla que tiene la estructura y detalla el contenido del PL, esta compuesta por 7 módulos: (*Introducción, Teoría, Formación, Experimento, Evolución, Replicación y Agregación*) y de la cuál la presente investigación hace referencia. Cada uno de estos módulos contiene 5 secciones y en cada sección se explica el contenido que debería tener el paquete. Sin embargo, los investigadores indican que teniendo en cuenta la variedad de tipos de replicación, se necesitaría cierta flexibilidad en la plantilla. Si bien es cierto, dicha propuesta hace un estudio profundo del contenido del PL la comunidad científica no la usa para construir el PL eso se ha evidenciado en la sección §3.6 de la presente memoria de tesis.

4.3 CARENCIA DE PAQUETES DE LABORATORIO

Por el momento aún existe poco interés de los investigadores en realizar replications de estudios empíricos en IS, sea por la complejidad técnica, procedimental, el tiempo y alto coste que genera el proceso de replicación y posiblemente el poco interés de replicar los experimentos originales [220]. Esto da lugar a que los PL tampoco se elaboren y se publiquen como se debería.

En 1999 Basili *et al.* [46] realizaron el análisis de una serie de familias de experimentos para estudiar a detalle características que ayuden a generar más replications en el área de IS. Como una estrategia que se identificaron en dicho análisis para apoyar a las replications en la práctica fue la creación de PL, que recogen información sobre el diseño experimental, los artefactos y procesos utilizados en el experimento, los métodos de análisis experimental y la motivación de las principales decisiones de diseño. Concluyendo que la existencia de estos paquetes simplificará el proceso de replicación en el algún momento.

Para llevar a cabo una replicación de experimentos, es necesario la transferencia compleja de conocimientos entre los investigadores. Este conocimiento debe ser documentado y estructurado para su fácil uso y comprensión [394]. De manera similar, Vegas *et al.* [416] señalan que la comunicación con los experimentadores originales favorece que la replicación sea exitosa.

Por otra parte, Shull *et al.* [381], analizan el uso de PL como una manera de comunicación entre los investigadores del estudio original y los miembros del equipos que realizan la replicación. Basili *et al.* [46], recomiendan el uso de PL por parte de los experimentadores incluyendo el material experimental junto con detalles de las técnicas utilizadas y Brooks *et al.* [69] señalan la necesidad de facilitar la reutilización de los PL, consideran que sin esto es poco probable que un experimento sea verificado mediante su replicación.

Durante el análisis de los PL en el presente estudio se ha observado algunas debilidades como: *i)* algunos paquetes sólo contienen el material experimental; *ii)* no siguen guías y directrices para llevar a cabo la elaboración del paquete; *iii)* no se proporciona suficiente información sobre la agregación y la evolución del experimento; y *iv)* no presentan una estructura estándar; la mayoría se basan en una página web y enlaces a otros archivos con distintos formatos.

A lo anterior se une la carencia de PL y a su vez de datos disponibles en algunos

experimentos. En 2014, Larsson *et al.* [236], realizaron una revisión bibliográfica sobre la *accesibilidad a los datos* de experimentos para su replicación. Se encontró que sólo el 37% de los estudios revisados tenían datos disponibles.

En ese mismo sentido, Kitchenham en [220], advierte sobre el uso de los PL argumentando en contra de las replicaciones exactas. Indica que se pueden encontrar replicaciones que confirman los resultados debido a reutilización de materiales o procesos disponibles y de fácil acceso.

Otra carencia que se identifica es la falta de un procedimiento coherente para publicar un artefacto de software, lo que dificulta la capacidad de replicar o reutilizar ese artefacto. Este problema también se evidencia en el uso de un archivo no modificable (es decir, persistente) o la adhesión a “prácticas de Ingeniería del Software” [394].

Trabajos recientes como Collberg *et al.* [92, 93] sugieren que muchos artefactos sufren de una variedad de problemas por ejemplo, falta de documentación y dependencias no declaradas, que impiden que otros los reutilicen, amplíen y reproduzcan. Krishnamurthi [231] explica que no existen estándares claros dentro de la comunidad sobre cómo crear y evaluar artefactos.

Se puede pensar que el PL orientado al proceso acorde a la ACM propuesto en esta memoria puede ayudar a facilitar la elaboración del PL del experimento.

4.4 EL CONOCIMIENTO TÁCITO

Llevar a cabo una replicación no sólo requiere de la disponibilidad del PL sino de otros factores como el *conocimiento tácito* de los autores originales del experimento.

Para Shull *et al.* [380] sostienen que hay dos tipos de conocimiento: el *conocimiento tácito* y el conocimiento explícito que permiten la transferencia de conocimientos entre los experimentadores. El conocimiento explícito es parte del PL y es accesible a los replicadores que necesitan comprender las técnicas y métodos que se aplican junto con los conocimientos necesarios para llevar a cabo el experimento. El *conocimiento tácito* es conocimiento necesario para llevar a cabo la replicación que tiene el experimentador y que por lo general no se hace explícito al reportar la replicación. Este tipo de conocimiento, regularmente no está incluido en el PL posiblemente porque los autores no consideraron su relevancia o porque es difícil de transmitir. Existe dificultad de reconocer que conocimiento es *tácito* para los autores como para los replicadores, debido a

la variedad de comprensión que cada persona involucrada en el proceso de replicación necesite para continuar con dicho proceso.

Shull *et al.* [380] especifican que el factor humano influye considerablemente en generar *conocimiento tácito*. Además, establecen que la falta de *conocimiento tácito* puede producir ciertas diferencias en el proceso experimental, puede suceder que altere los resultados del experimento original y la replicación y que no se puedan comparar. En los experimentos controlados con sujetos humanos, estas pequeñas variaciones pueden tener grandes efectos en los resultados. La influencia del factor humano en las replications se analiza en la sección §4.7

En el 2018 von Nostitz-Wallwitz *et al.* [424, 425] realizan dos estudios empíricos utilizando un *Survey* para determinar algunas buenas prácticas e información que se debe tomar en cuenta para transferir conocimiento entre profesionales y artefactos de software. Los estudios concluyen la importancia de continuar con reuniones y establecer colaboración entre los involucrados en el proceso de investigación, las publicaciones tienen limitado *conocimiento tácito* y es importante mantener la relación entre investigadores con el objetivo de ayudar a la replicación.

La propuesta presentada en esta memoria de tesis contribuye a disminuir el *conocimiento tácito* ya que permite al mismo tiempo especificar las tareas y registrar el material experimental. Este accionar contribuye a la transferencia del conocimiento facilitando su transmisión y aprovechando el conocimiento que posee el experimentador original en el momento de ejecución de cada tareas durante la operación del experimento.

4.5 DIFERENTE TERMINOLOGÍA

Respecto de la terminología utilizada en los PL, los investigadores han utilizado diferentes términos relativos a las replications y específicamente para elaborar el PL. La problemática relativa a la clasificación de PL se ha analizado en la sección §2.2.8.

En la revisión de experimentos controlados llevada a cabo en el 2020 por Heumüller *et al.* [179], además de la necesidad de usar directrices para la realización y publicación de los artefactos del PL comentada anteriormente (ver sección §4.2), se detectó falta de consistencia en la terminológica utilizada que dificulta la comprensión de los experimentos y sus replications.

Al momento es difícil aplicar una estructura universal para la elaboración de un PL, existe dificultad en determinar un nivel de profundidad, es decir, hasta donde llega el contenido del paquete. Se ha observado de algunas propuestas pero aún sigue siendo un reto la elaboración del PL experimental para el área de IS.

En 2014–2016, los estudios realizados por Diebold *et al.* [124], Lo *et al.* [251] y Garousi *et al.* [156] han analizado los desafíos y buenas prácticas con respecto a la transferencia y comunicación en proyectos y estudios del área de IS. En sus conclusiones especifican la necesidad y dificultad de colaborar juntos, la academia y la industria por la diversa taxonomía, estructura y formalidad del uso de normas y estándares que contempla todo el proceso de investigación en el área de IS. Estos estudios coinciden en las necesidades que existe para llevar a cabo y publicar las replicaciones de estudios empíricos.

4.6 NECESIDAD DE CLARIFICAR EL CONTEXTO DE LAS REPLICACIONES

Briand *et al.* [63] consideran que se debe fomentar la investigación basada en el *contexto* para que la investigación en IS aumente su impacto y mejore la calidad de la misma. Dicha investigación se centra en problemas definidos en colaboración con la industria y está impulsada por necesidades concretas en dominios específicos y proyectos de desarrollo.

Clarificar el *contexto* de la investigación permite a los replicadores interpretar la realidad de conducir el experimento en las condiciones originales, considerando factores humanos, organizativos, el tiempo, el coste, el nivel de criticidad y cumplimiento de las normas entre otros contextos de dominio que sea relevante. Inclusive durante la replicación existe cambios que pretenda mejorar los resultados y llegar a conclusiones válidas para la comunidad científica.

Una ventaja de clarificar el contexto en las replicaciones es que ayuda a identificar problemas que se ha pasado por alto en el experimento original, pero que son importantes considerarlos.

Para Juristo *et al.* [202] controlar adecuadamente el contexto de un experimento facilita a los investigadores que trabajan en los mismos escenarios replicar los experimentos. Sin embargo, es muy difícil controlar adecuadamente el contexto de replica-

ción en campos experimentalmente inmaduros como el área de IS que tratan con seres humanos y que trabajan en entornos complejos. El contexto de un experimento de IS abarca cientos de variables, que requieren ser definidas para ayudar a la replicación.

Cartaxo *et al.* [85] identifican los mecanismos de apoyo a la caracterización del *contexto* mediante una revisión bibliográfica. Consideran que hay pocos estudios que describen el *contexto* y que la falta de información al respecto dificulta la replicación, la interpretación de resultados y la transferencia de conocimientos entre la academia y la industria.

Otros autores que también han señalado la importancia de describir el *contexto* son Murphy [282], Sjøberg *et al.* [388], Dyba *et al.* [127], Jedlitschka *et al.* [193], Petersen *et al.* [321] y Shepperd [375] señalan que los estudios empíricos especifican el *contexto* de forma genérica muchos de los experimentos enfatizan describir la *validez externa*. Además, coinciden en la necesidad de describir el *contexto* con la mayor claridad y especificación para obtener conclusiones válidas y aumentar las replicaciones.

4.7 EL FACTOR HUMANO EN LAS REPLICACIONES

Todo proceso realizado por humanos esta sujeto a error. Los experimentos con humanos en el área de IS y sus replicaciones cambian a los largo del tiempo debido a las características diferentes que cada persona tiene por naturaleza. Los sujetos experimentales adquieren experiencia o se agotan en las tareas asignadas, el hecho de estar sometidos a cambios en su entorno hace que no se controle del todo el *contexto* experimental.

En 2020 Storey *et al.* [399] analizan que la IS es un esfuerzo socio-técnico en donde generalmente la mayoría de los estudios esta orientado a la parte técnica. Además, revisan 151 artículos científicos y determinan cuánto se estudia los aspectos humanos y sociales en los diseños de investigación. En ese sentido, proponen un marco de investigación socio-técnica para capturar el beneficiario principal de un estudio de investigación. En sus conclusiones definen que existe la necesidad de estrategias que apunten a una comprensión más profunda de los aspectos humanos y sociales de la práctica del desarrollo de software para equilibrar el diseño y la evaluación de las innovaciones técnicas.

Para las Ciencias Humanas que son aquellas ciencias encargadas del estudio del ser Humano, la sociedad y su cultura, los objetos de estudio no parecen seguir reglas

predefinidas para ser descubiertas. Uno de los problemas en la investigación de las Ciencias Humanas y sociales es que las preguntas de investigación inician como una percepción de los investigadores [394]. Este problema se evidencia también en el área de IS debido a la estrecha relación que se tiene con dichas ciencias por el componente humano propiamente dicho.

Para Seaman [369], la influencia del *factor humano* en los experimentos en el área de IS necesita nuevos métodos de investigación como incorporar métodos cualitativos aplicados en otras disciplinas como las Ciencias Sociales para analizar la complejidad del *comportamiento Humano*.

Lung *et al.* [253], consideran que la variabilidad del comportamiento humano, las variables no controladas y el posible sesgo de los investigadores dificulta el proceso de replicar los experimentos controlados con humanos. Concluyen que para superar estas limitaciones, los estudios deberían ser replicados con un diseño mejorado o incluso con un método de investigación diferente.

4.8 RESUMEN

Los PL de experimentos controlados en el área de IS juegan un papel importante en el proceso de replicación, así lo menciona Timperley *et al.* [406]. El PL ayuda a comprender, entender y guiar al replicador como se ha conducido el experimento original, para luego establecer sus propias tomas de decisiones en la evolución del experimento.

En las anteriores secciones se han identificado algunos de los problemas que contribuyen de forma conjunta al bajo número de replications y por consiguiente pocos PL publicados. En resumen los problemas identificados son: *i)* el poco uso de guías y directrices para reportar Paquetes de Laboratorio; *ii)* carencia de los Paquetes de Laboratorio; *iii)* dificultad para transmitir el conocimiento tácito; *iv)* falta de acuerdo sobre la terminología; *v)* dificultad para clarificar el contexto de las replications; *vi)* la complejidad de utilizar sujetos humanos; y *vii)* el alto grado de esfuerzo y recursos necesarios para elaborar Paquetes de Laboratorio.

El PL orientado al proceso acorde a la ACM que se propone en la presente memoria de tesis, puede contribuir a paliar parte de los problemas identificados. En particular, a que el contenido y estructura de los artefactos que componen el PL se elaboren de mejor manera y posiblemente aumente la calidad de los mismos. Aspectos como la *estructura del paquete*, la *integración de las tareas junto con la ejecución del proceso experi-*

mental, la guía paso a paso para replicaciones externas y la calidad del contenido del paquete es considerado en la presente propuesta.

Los problemas identificados en este apartado, han servido como base para la propuesta de la presente tesis haciendo hincapié en apoyar a la elaboración de PL de calidad para experimentos controlados con humanos en el área de IS.

PARTE III

PROPUESTA

PAQUETE DE LABORATORIO ORIENTADO AL PROCESO ACORDE A LA ACM

El experimento científico es la más rica de todas las formas de experiencia humana: añade a la observación el control de ciertos factores en base a supuestos teóricos y, cuando es preciso, supone medición.

*Mario Augusto Bunge (1919 – 2020),
Físico, filósofo y epistemólogo argentino*

En este capítulo se presenta la contribución principal de este trabajo de tesis, cuyo objetivo principal es elaborar un PL orientado al proceso que posea un nivel de calidad adecuado según las recomendaciones existentes en la comunidad de la Ingeniería del Software [185, 295]. Como se ha argumentado en capítulos anteriores, el hecho de elaborar un buen PL es beneficioso ya que facilita la replicación del estudio en cuestión y contribuye a la transparencia y a la reproducibilidad de resultados. En la sección de introducción §5.1, se destacan las cualidades de la propuesta. En la sección §5.2, se describe la especificación del proceso experimental de forma integrada con la elaboración del PL. En la sección §5.3, se determina el procedimiento para elaborar un PL acorde a la ACM. En la sección §5.4, mediante un modelo de calidad, se determina características de calidad para evaluar los artefactos del PL y alcanzar las insignias de la ACM. En la sección §5.5 se propone una estructura básica con respecto al contenido que un PL debe presentar para tener un adecuado nivel de calidad. Finalmente, la sección §5.6 resume el capítulo.

5.1 INTRODUCCIÓN

La elaboración de un PL es un tarea que demanda esfuerzo y recursos más aún cuando su elaboración se realiza al finalizar el proceso experimental. Llevar a cabo un experimento controlado con humanos, por lo general trae consigo tareas, cualquier tarea, por pequeña que sea debe registrarse, al igual que los artefactos que se crean y que se utilizan en cada tarea.

El objetivo de este trabajo es fomentar las replicaciones de experimentos controlados con humanos mediante la elaboración de un PL orientado al proceso acorde a la ACM proponiendo: *i)* una especificación del proceso experimental de forma integrada con el PL, *ii)* un procedimiento para elaborar un PL acorde a la ACM, *iii)* un modelo de calidad para evaluar los artefactos del PL y alcanzar las insignias de la ACM y *iv)* una estructura básica que un PL debe presentar para tener un adecuado nivel de calidad acorde al estudio de Solari *et al.* [393] y las insignias de la ACM.

El uso de nuestra propuesta en el desarrollo del PL persigue una doble finalidad, por un lado, permite al investigador la elaboración del paquete mientras está ejecutando el experimento, y no a posteriori. Por otro lado, permite al replicador comprender mejor los artefactos que se usarán en la replicación mediante un paquete de artefactos bien estructurados y documentados.

5.2 ESPECIFICACIÓN DEL PROCESO EXPERIMENTAL DE FORMA INTEGRADA CON PL

Un proceso experimental proporciona una base para el necesario avance en el conocimiento y la comprensión. Como cualquier otro procedimiento experimental los experimentos controlados en IS, implica una iteración de un proceso de hipótesis y pruebas [45]. El proceso experimental en IS contempla un conjunto de actividades, tareas, técnicas, artefactos y herramientas que se utilizan para cumplir con los objetivos de la investigación. El marco conceptual más importante para este proceso lo establecen las actividades y en sí las tareas que lo componen [392].

Las tareas que se realizan con mayor intensidad en el proceso experimental corresponden a la planificación y realización del experimento. Para la presente investigación se ha considerado estudiar el proceso experimental propuesto por Basili *et al.* [45],

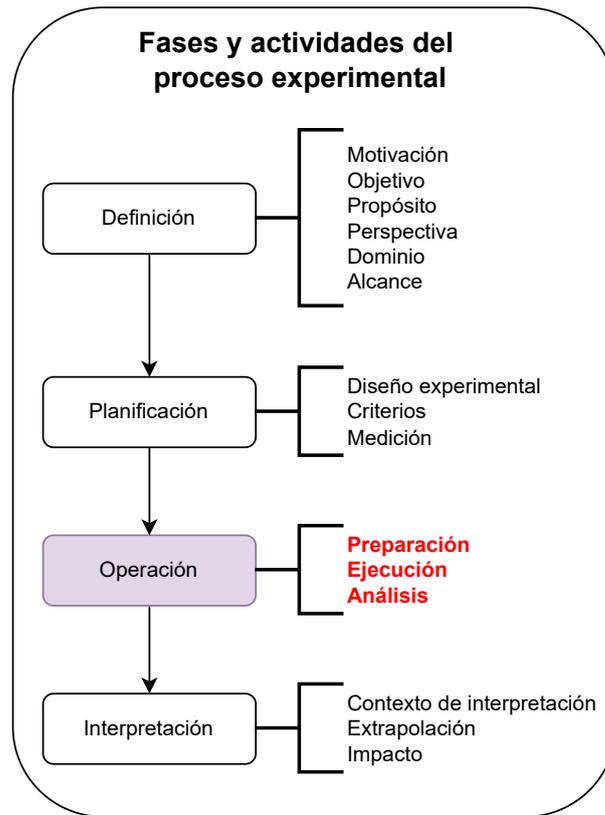


Figura 5.1: Fases y actividades del proceso experimental

Wohlin *et al.* [435] y Jedlitschka *et al.* [194], con el objetivo de identificar las fases del proceso experimental y las tareas de la ejecución del experimento.

Las fases del proceso experimental propuestas por investigadores en el área IS se describen en la tabla §2.1 (Pág.25). Se observa distintos nombres que las propuestas de la tabla §2.1 dan a la fase de "operación" en la cuál está investigación se enfoca, por ejemplo: "ejecución", "conducir el experimento" y el mismo nombre de "operación".

De manera general se explica cada unas de las fases del proceso experimental basándose en los estudios antes mencionados. La figura §5.1 resume las actividades de las fases experimentales:

- **Definición.**- Por lo general se contempla seis partes: *i)* la motivación, *ii)* el objetivo, *iii)* el propósito, *iv)* la perspectiva, *v)* el dominio y *vi)* el alcance. Estas partes especifican las siguientes cuestiones: el ¿Porqué?, ¿Cómo?, ¿Para qué? y ¿Dónde? se lleva a cabo el experimento, con la finalidad de definir su alcance. Esta fase es importante que sea estudiada y analizada en el equipo experimental porque de ésta definición depende que el experimento tenga un exitoso final.

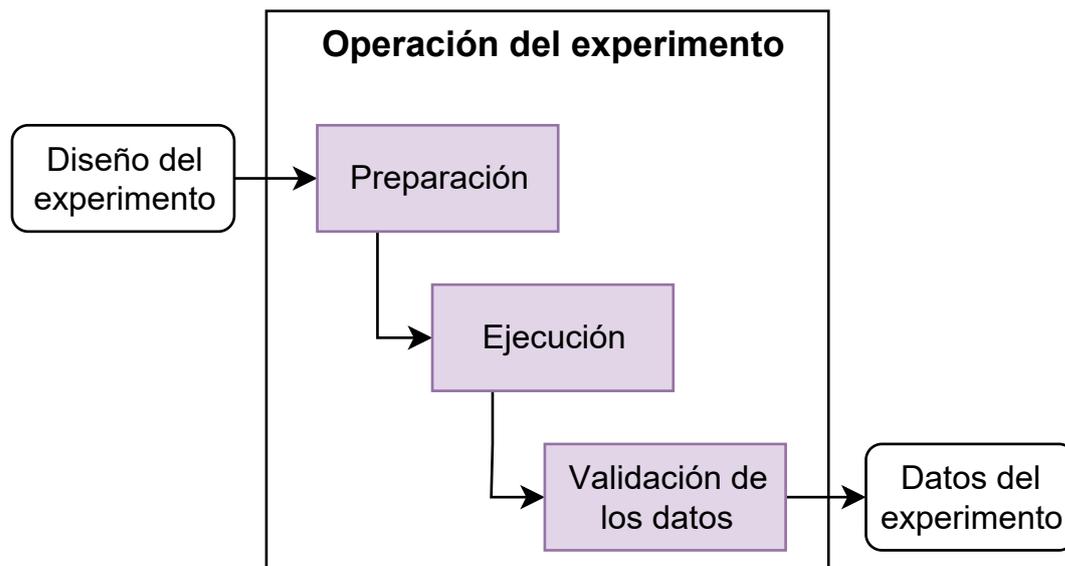


Figura 5.2: Actividades de la fase de operación en el proceso experimental [435]

- **Planificación.**- Esta fase discute tres aspectos: el *i)* diseño experimental, los *ii)* criterios y la *iii)* medición. Durante el diseño se asocia el ámbito del estudio con los métodos analíticos e indica las muestras de dominio que se van a examinar. Se analiza los criterios de calidad, coste, errores, cambios y complejidad del experimento. La medición capta las manifestaciones concretas de los aspectos de coste/calidad y lleva a especificar la manera como se evaluará y validará los resultados experimentales [91].
- **Operación.**- La fase de operación consiste de: la *i)* preparación, la *ii)* ejecución y del *iii)* análisis. Durante la preparación se confirma el escenario experimental mediante un estudio piloto, se organiza los factores experimentales por ejemplo la experiencia de los sujetos. En la ejecución se lleva a cabo el experimento ejecutando las tareas y aplicando los artefactos que forman parte del material experimental. En el análisis, los experimentadores recogen y validan los datos definidos durante la ejecución del estudio. El análisis de los datos puede incluir una combinación de métodos cuantitativos y cualitativos [60]. En la figura §5.2 se muestra de forma resumida las actividades de la fase de operación.
- **Interpretación.**- La interpretación del experimento consta de: el *i)* contexto de interpretación, la *ii)* extrapolación y el *iii)* impacto. Los resultados del análisis de los datos de un estudio se interpretan en una serie de contextos cada vez más amplios. Estos contextos de interpretación son el marco estadístico en el que se deriva el resultado, el propósito del estudio concreto, y los conocimientos en el

campo de la investigación. La extrapolación de los resultados a otros entornos representa el muestreo analizado. El impacto está dado por el seguimiento de varias actividades como el de presentar y publicar los resultados para su retroalimentación, la reproducción y replicación del experimento [43].

Otro estudio que favorece la comprensión en el reporte del proceso experimental de experimentos controlados en el área de IS es el estudio de Revoredo *et al.* [343], donde especifican el proceso y directrices para el reporte de experimentos controlados desde un modelo de procesos. En su estudio los autores concluyen que no se reporta de manera completa los experimentos por ejemplo las tareas en el experimento se explican de forma limitada, en la mayoría los estudios no siguen ninguna directriz para llevar a cabo el proceso experimental e inconsistencia en la presentación de informes.

Durante el reporte del proceso experimental, un factor a considerar es el desarrollo del PL, el mismo que está compuesto de artefactos, en los cuáles se registra y documenta las tareas, instrucciones y material necesario para replicar un experimento. El paquete debe dar soporte a todas las actividades del proceso experimental. La ACM ha definido el concepto de artefacto y ha establecido criterios de evaluación para asignar insignias de calidad al PL [185].

Inicialmente, para especificar el proceso experimental de forma integrada con el PL fue necesario comprender las fases del proceso experimental y analizar a profundidad la actividad de "ejecución" de la fase de operación, con la finalidad de integrar las tareas que contribuyen a la realización del experimento y el desarrollo del PL, siendo ésta finalidad la propuesta de la presente investigación. Considerando, que un experimento controlado es exclusivamente procedimental, la integración de las tareas con el PL van de la mano.

En la fase de operación se ha considerado analizar tres tipos de tareas: de *i)* preparación, *ii)* experimentales y de *iii)* recolección y análisis.

- Las tareas de preparación.- Involucran estrictamente tres aspectos: *i)* la captación de los sujetos experimentales, *ii)* la información que se comunica a los sujetos experimentales con respecto a la naturaleza y objetivos del experimento y *iii)* la enseñanza de los conceptos, teorías y demás conocimiento necesario dirigida a los sujetos para llevar cabo el experimento con el fin de tener un grupo de sujetos nivelado. Dicha enseñanza se realiza mediante material complementario que los investigadores crean y usan en las sesiones de formación como también la creación, uso y aplicación de formularios, por ejemplo: la configuración de una

herramienta, administración de inicios de sesión, permisos de acceso a una herramienta. Sin duda, a pesar de que los experimentos con humanos en IS requieren de un diseño perfectamente experimental y el uso adecuado de métodos de análisis para los datos recogidos, el resultado no será válido si los sujetos experimentales no han participado seriamente en el experimento. Es necesario motivar a los sujetos y nivelar sus conocimientos, estos aspectos, de manera similar se manifiesta en el campo de la Psicología experimental donde también se ocupa de los experimentos con seres humanos y se establece directrices para la realización de experimentos en ese campo [23, 391] que bien se pueden estudiar en la IS. Otra manera, que la comunidad del área de IS Empírica trata de motivar a los participantes es mediante los incentivos financieros, es decir, remunerar por el trabajo realizado durante la participación en el experimento. Aunque, esta manera de incentivo no es totalmente específica y no hay una guía ni lineamientos para su aplicación en el área de IS Empírica, puede ayudar a mejorar la validez imitando y siendo fiel al mundo real [232].

- Las tareas experimentales.- Son las tareas de ejecución propiamente dichas durante el experimento. Involucran dos aspectos: *i*) los sujetos realizan sus tareas según los distintos tratamientos y *ii*) la recogida de datos que es ejecutada por los investigadores y/o por los propios sujetos experimentales.
- Las tareas de recolección y análisis.- Consisten en la revisión y comprobación de cualquier tipo de procesamiento de datos necesario, además dichos datos sean razonables y que se han recogido correctamente, estas tareas son realizadas por el experimentador. Es importante revisar que el experimento se haya realizado realmente de la manera que se pretendía. Posteriormente, se realizará el análisis de los datos, para ver los hallazgos del estudio empírico.

La figura §5.3 muestra la especificación de la fase de "operación" del proceso experimental de forma integrada con el PL para estructurar el contenido del paquete y proporcionar un esquema de clasificación para entender y evaluar los artefactos.

Por consiguiente, la presente propuesta de tesis esta direccionada a que la operación experimental en la cual, se lleva a cabo una secuencia de tareas y que puede dar lugar a tareas simultáneas, es decir, similar a un gestor de tareas, se integrará con el material experimental que se va generando y usando durante la fase de operación del experimento. La finalidad de esta propuesta es doble: *i*) se genere una *guía paso a paso* para futuros experimentadores que quieran hacer reproducciones y/o replicaciones del estudio experimental y *ii*) permitir al experimentador la creación de un PL de calidad.

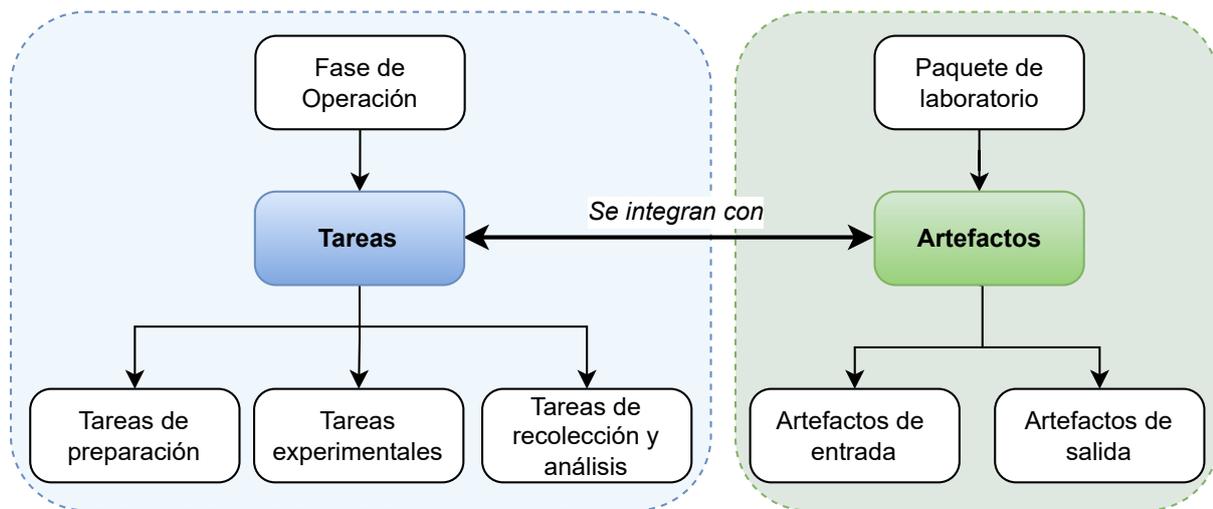


Figura 5.3: Integración de las tareas del experimento con el Paquete de Laboratorio

Finalmente, para conseguir que el PL generado tuviera un nivel de calidad aceptable y pudiera alcanzar alguna de las insignias de la ACM se desarrolla un modelo de calidad para PL.

5.3 PROCEDIMIENTO PARA ELABORAR UN PL ACORDE A LA ACM

Procedimiento es un término que hace referencia al accionar de *proceder*, que dicho de otra manera, es la forma determinada de actuar y esta relacionado con la acción de *ejecutar* algo en un tiempo determinado. En este sentido, un procedimiento consiste en seguir ciertos pasos predefinidos para ejecutar una acción de manera eficaz. Es posible, que para alcanzar un mismo objetivo existe uno o más procedimientos con estructuras, etapas, pasos diferentes y que resulten ser más o menos eficientes [34].

Un procedimiento contienen información cronológica y secuencial de las actividades que se realizan de manera lógicamente. Además, tiene la finalidad de establecer un orden y no alterar las definiciones ya establecidas [267].

En el presente estudio de investigación, luego de haber establecido la especificación del proceso experimental de forma integrada con el PL en la sección §5.2, se ha propuesto establecer un procedimiento que permita elaborar un PL acorde a la ACM.

5.3.1 Procedimiento para elaborar un PL durante la ejecución del experimento

Un procedimiento contiene la narrativa de cada una de los pasos a seguir, explicando detalles del ¿Qué?, ¿Quién?, ¿Cómo?, ¿Cuándo?, ¿Dónde?, ¿Con qué?, es necesario para que se lleven a cabo. Para la presente investigación se ha considerado establecer los siguientes pasos que debe contener el procedimiento para elaborar un PL acorde a la ACM:

- Primer paso.- Registrar datos informativos del experimento como: nombre, DOI, lugar, fecha, definir el objetivo del experimento siguiendo de preferencia la estructura propuesta por Basili *et al.* [40] Goal-Question-Metric (GQM) o, en su defecto, describir el objetivo y una breve descripción del mismo. Este primer paso permite conocer el alcance y la meta que se quiere cumplir una vez ejecutado el experimento. Esta actividad es registrada por el experimentador al igual que todas las actividades que se describen a continuación.
- Segundo paso.- Registrar a los experimentadores mediante el ingreso de datos personales y el rol que van a desempeñar en el experimento. En un experimento controlado puede haber uno o más investigadores que tienen roles asignados por ejemplo: de diseñador, analista, formador, monitor, medidor que hacen actividades específicas. Se registra datos personales como el DNI, los nombres completos, email, sitio web, afiliación, género, nacionalidad, el tipo de persona si es un estudiante, graduado o un profesional.
- Tercer paso.- Identificar los grupos es importante porque las tareas que realizan los diferentes grupos pueden ser distintas y, en consecuencia, el material experimental usado es distinto para cada uno de ellos. Los grupos pueden ser de control y/o experimental, el experimentador asigna a los participantes a cada grupo, luego de hacer una selección de los participantes. Los grupos se establecen en base al diseño experimental que se ha definido.
- Cuarto paso.- Se trata de un paso con varias iteraciones, ya que se va a identificar cronológicamente todas las tareas que se van desarrollando durante el experimento, se han contemplado tres tipos de tareas: de preparación, experimentales, de recolección y análisis. Para cada tarea el experimentador debe registrar los artefactos que se van generando sean estos de entrada y/o salida. Se debe registrar las premisas para una determinada tarea, por ejemplo la fecha de inicio, fecha de

fin, duración de la tarea, con el fin de identificar todos los detalles de la tarea. Las tareas pueden suceder secuencialmente, pero puede haber tareas que se realicen de manera simultánea. Por ejemplo, la tarea experimental de los grupos bajo tratamiento y el grupo de control, que previsiblemente usarán artefactos distintos durante la ejecución. Dicha situación se ilustra en la figura §5.5. Los artefactos que pertenecen a las tareas deben ser registrados con un nombre, la finalidad que persigue, un formato, el tamaño, fecha de creación y actualización así como también la URL o dirección web (PATH) donde se encuentra disponible.

- Quinto paso.- Registrar aquellos artefactos del PL que no están vinculados directamente a una tarea concreta del experimento, que puede ser una planificación de actividades de la fase de operación del experimento, el preprint del artículo donde se reporta el experimento, guías de actividades dentro de las fases del proceso experimental, amenazas a la validez del experimento que no necesariamente están ligadas a las tareas definidas en el cuarto paso.
- Sexto paso.- Registrar aquellos artefactos adicionales del PL que permiten cumplir con las políticas de evaluación para alcanzar las insignias de la ACM, por ejemplo: los archivos *readme*, *install*, *licence*, *status*, *guías de instalación*. Los grupos de artefactos que se mencionan en el cuarto, quinto y sexto paso están definidos y descritos en la sección §5.4.2 de la presente memoria de tesis.
- Séptimo paso.- Establecer la configuración del PL dónde el experimentador elige la estructura de carpetas, y cuál será el criterio para organizarlas (cronológico, por formato, por finalidad).
- Octavo paso.- Evaluar los artefactos del PL para alcanzar las insignias de la ACM.
- Noveno paso.- Generar una Guía paso a paso (con cronograma), para futuros replicadores del experimento. Como complemento a este trabajo de investigación se está desarrollando una herramienta de tal manera que esta guía se genere de forma automática y evidencie si se han registrado correctamente las tareas del proceso experimental.
- Décimo paso.- El porcentaje de cumplimiento de los artefactos que conforman el PL se establece en referencia al modelo de calidad y cumplimiento de las métricas establecidas en la sección §5.4 con la finalidad de conocer la insignia de la ACM que puede alcanzar el paquete.

- Undécimo paso.- Se registra información del PL orientado al proceso como el nombre, un DOI en el caso de tener disponible, de la misma manera el nombre del repositorio de archivos público donde se aloja el paquete, fechas de creación y actualización del paquete.

La figura 5.4 ilustra el procedimiento para elaborar un PL acorde a la ACM, mediante un diagrama de flujo que resume los pasos a seguir.

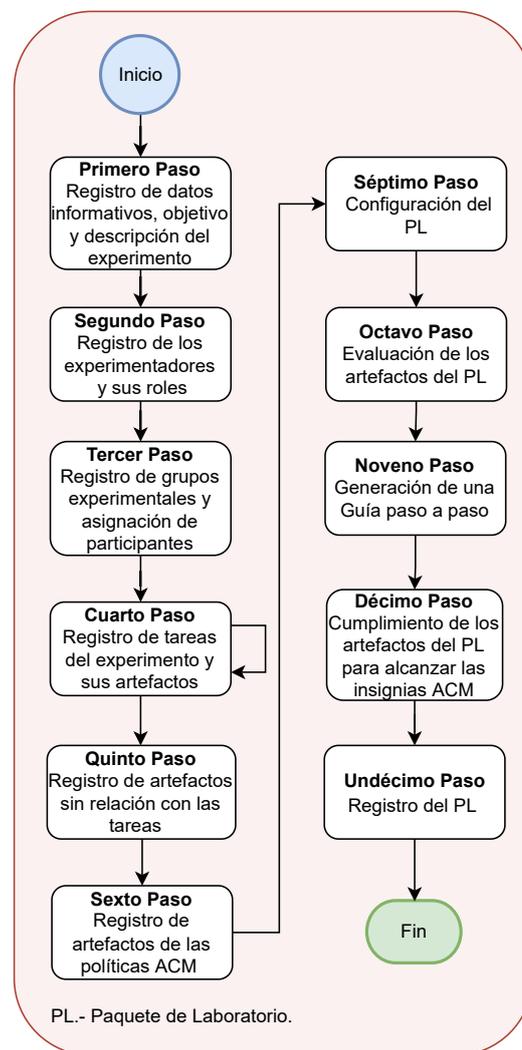


Figura 5.4: Procedimiento para elaborar un Paquete de Laboratorio acorde a la ACM

5.4 MODELO DE CALIDAD PARA PAQUETE DE LABORATORIO ACORDE A LA ACM

A continuación, se presenta el desarrollo del modelo de calidad para PL acorde a la ACM para experimentos controlados con humanos en el área de IS. La finalidad del modelo de calidad es establecer cuáles son las propiedades deseables de un PL para que este sea bueno dentro de los PL de su clase.

5.4.1 Introducción al modelo de calidad para Paquetes de Laboratorio acorde a la ACM

La calidad en la IS es un concepto dependiente de la perspectiva del usuario, el uso del producto, el contexto y entorno donde se ejecuta. La calidad depende en gran medida de las habilidades y destrezas del equipo humano que lo desarrolla. Puede definirse como un conjunto de características o cualidades, tales como: *eficiencia, fiabilidad, usabilidad, funcionalidad, mantenibilidad, portabilidad*, entre otras, como lo indica la familia de la norma ISO/IEC 25000, conocida como SQuaRE (System and Software Quality Requirements and Evaluation) [190]. Esta familia tiene por objetivo la creación de un marco de trabajo común para evaluar la calidad del producto software, es decir, es el cumplimiento de los requisitos contractuales por parte del producto de software desarrollado, así como durante el proceso de desarrollo [418].

En las empresas de IS, la calidad resulta de la mejora diaria en el proceso de producción, mantenimiento y gestión del software [157]. Para optimizar la calidad de los productos y/o servicios es preciso conocer al cliente y sus necesidades, conocer a la competencia y poseer un *modelo de calidad*. Un *modelo de calidad* permitirá incrementar la fiabilidad, reducir el mantenimiento, aumentar la satisfacción del cliente, mejorar la dirección del proyecto, detectar errores de manera temprana e incrementar el beneficio [260].

Considerar a la calidad en los PL en IS, es una tarea que esta tomando fuerza por parte de la comunidad científica. Inicialmente, la evaluación de artefactos se ha introducido en la IS como piloto en la conferencia ESEC/FSE 2011. En los últimos cinco años el Track AE-Artifact Evaluation dentro de los congresos y actualmente las revistas científicas del área de IS han impulsado evaluar los PL de los estudios, y más aún cuando el incentivo del movimiento *Open Science* está en crecimiento.

Si bien es cierto las directrices de revisión y evaluación de los artefactos en los congresos y revistas están disponibles, no existe un consenso sobre los umbrales de calidad de los artefactos en general. Durante el estudio de mapeo sistemático presentado en la sección §3, se ha observado que cada congreso de la comunidad en IS considera directrices de evaluación muy específicas para determinados tipos de artefactos, esto se ve reflejado en la diversa presentación de los PL que los autores reportan especialmente en su estructura y contenido. Otro fenómeno que se ha identificado es que a pesar de estar las directrices disponibles en los congresos los PL son distintos, posiblemente construir un paquete conlleva demasiado esfuerzo y más aún para cumplir con criterios de evaluación. La mayoría de las directrices están basadas en las políticas de evaluación e insignias de la ACM [185].

5.4.2 Definición de los artefactos que forman parte del Paquete de Laboratorio acorde a la ACM

Una vez que se ha especificado la integración del proceso experimental con el PL descrito en la sección §5.2 y el análisis realizado en el estudio de Solari *et al.* [392], se ha considerado que un "Buen" PL experimental esta compuesto de la estructura del contenido del PL propuesto por Solari *et al.* [394], y del registro de tres grupos de artefactos: *i)* artefactos relacionados con las tareas del experimento, *ii)* artefactos que no están relacionados con las tareas del experimento, es decir, cualquier artefacto que los experimentadores crean relevantes y necesarios para completar el PL. Y *iii)* artefactos que pertenecen a las políticas de evaluación para alcanzar las insignias de la ACM definidas en la sección §1.1.

- El *primer grupo*, hace referencia a los artefactos que se relacionan directamente con las tareas del experimento, estas tareas son de *preparación, experimentales* y de *recolección y análisis* definidas por Wohlin [435] y resumidas en la sección §5.2. Cada una de estas tareas puede registrar uno o más artefactos ya sean de entrada o de salida de acuerdo a la finalidad que tengan los artefactos. Sin embargo, el contenido de los artefactos depende de cómo el experimentador elabora cada uno de los artefactos.

A manera de ejemplo, para ilustrar este grupo de artefactos, la figura §5.5 muestra un modelo de procesos (BPMN - Business Process Model and Notation) definido para un experimento llevado a cabo en la Universidad Técnica del Norte.

Este experimento tiene el objetivo de evaluar dos paradigmas de construcción de

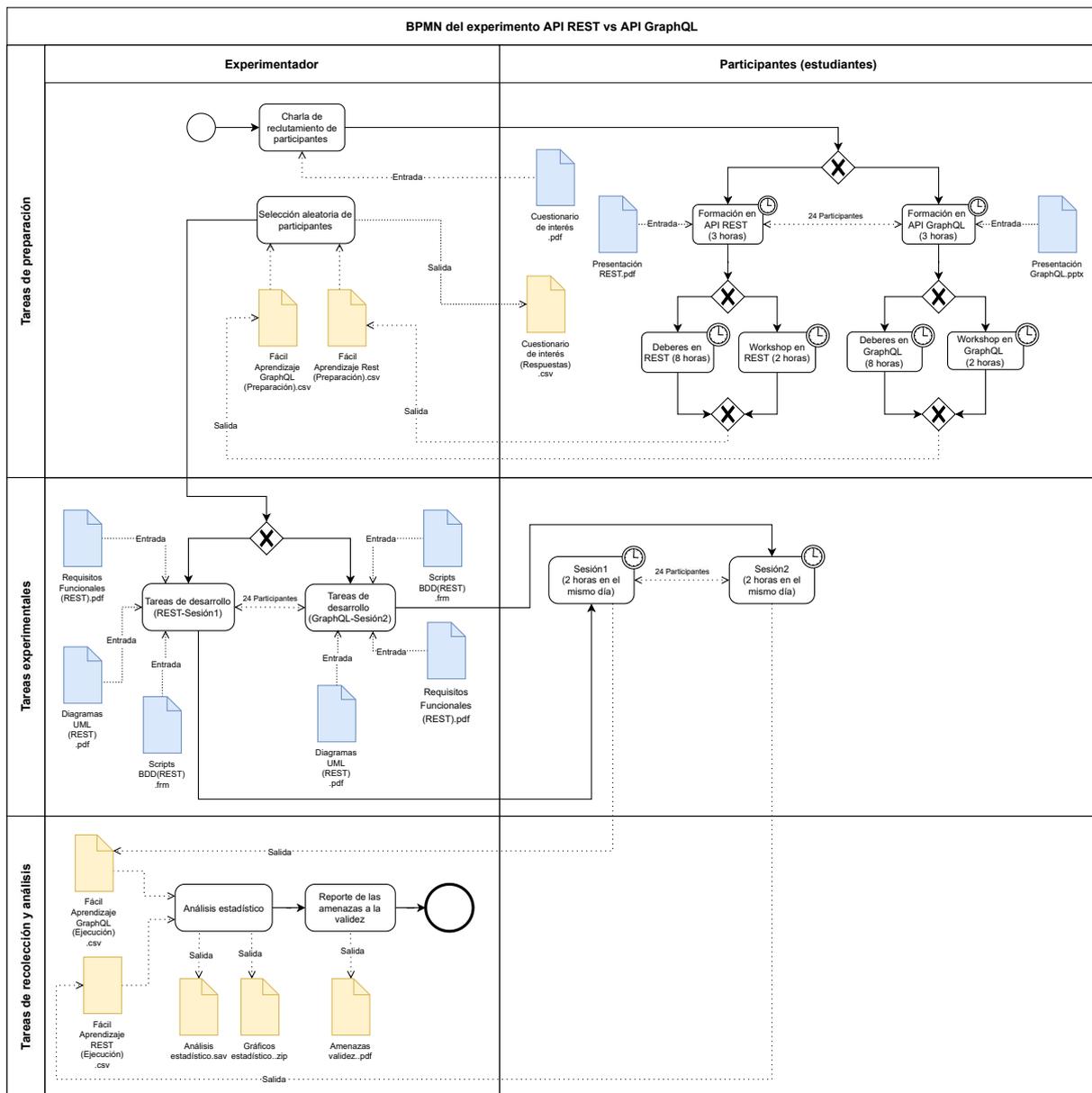


Figura 5.5: Ejemplo de la integración entre las tareas del experimento y los artefactos

arquitecturas microservicios API REST con API GraphQL. El diagrama BPMN indica las tareas y los artefactos de entrada y salida generados durante las tareas de preparación, experimentales, de recolección y análisis.

Para modelar la ejecución experimental se ha recurrido ha BPMN, de tal modo que mediante Actividades se muestran las tareas de preparación, experimentales, de recolección y análisis. Mediante *Data Objects* se representa los artefactos de cada tarea (que puede ser un artefacto de entrada o de salida) y a través de los *Swimlanes* se organizan las diferentes actividades junto con los responsables

(Experimentador y Participantes) que pertenecen a cada tarea.

El uso de la notación BPMN para representar la ejecución del experimento es adecuada en IS porque, en general la comunidad en el área de IS esta acostumbrada a ella, pero no es así en otras áreas. No obstante, en este trabajo no se aborda en profundidad la adecuación de BPMN para representar dicha información, solo se ha usado por familiaridad con ella y por claridad.

- El *segundo grupo*, corresponde a los artefactos que no necesariamente se relacionan con las tareas del experimento. Por ejemplo, la planificación de actividades para la ejecución del experimento, el reporte de las amenazas a la validez que han identificado los experimentadores.
- El *tercer grupo*, corresponde a los artefactos que se generan en el cumplimiento de las políticas de evaluación para alcanzar las insignias de la ACM. Por ejemplo, los archivos *readme*, *install*, *licence*, *status*, *contact*, *guías de instalación* (en el caso de que el experimentador implique la instalación de un software), que contienen información específica para alcanzar cada insignia.

Luego de analizar los artefactos que componen un "Buen" PL, se propone establecer un modelo de calidad, como un conjunto de propiedades deseables para PL acorde a la ACM. Se ha considerado que los artefactos definidos anteriormente en el primero, segundo y tercer grupo son comunes para la mayoría de los experimentos que desean alcanzar las insignias de la ACM y que de alguna manera se pueden categorizar. Para definir el modelo de calidad el cuál sirva como guía para alcanzar una insignia de la ACM, se establece un conjunto de características de calidad.

5.4.3 Definición de las Características del Modelo de Calidad

Las características del modelo de calidad para PL acorde a la ACM, han sido definidas luego de hacer un estudio profundo y haber analizado las políticas de envío y evaluación de artefactos que vienen acompañadas con el movimiento *Open Science* de varias organizaciones relacionadas al área de IS, entre las cuáles se ha estudiado: las políticas de revisión de artefactos e insignias de la ACM [185], las políticas para el Track AE-Artifact Evaluation de los congresos ICSE'2021 [186], ESEC/FSE'2021 [141], RE'2021 [339], ASE'2021 [35], ROSE'2018 [1], la guía de envío de los artefactos en el ICSE'2021 propuesta por Abrahão y Mendez [6], el reporte NISO RP-31-2021 [295], el estudio de Monperrus *et al.* sobre repositorios *Open Science* [274], lista de tipos de artefactos en IS propuesto por Monperrus & Menzies [275] la iniciativa de *Open Science*

de la revista EMSE [287] y los estudios de Solari *et al.* sobre PL [393, 394], en total trece estudios.

Para la presente investigación se ha considerado estudiar qué características de calidad debe cumplir un PL para alcanzar alguna de las cinco insignias de la ACM: *i) Funcional, ii) Reutilizable, iii) Disponible, iv) Reproducido y v) Replicado* las cuáles se explican a detalle en la sección §1.2.

La definición de las características de calidad se establecen para los artefactos descritos en la sección §5.4.2. Estos artefactos deben cumplir como mínimo las características de calidad establecidas en la tabla §5.1. Estas características están compuestas de artefactos los cuáles, son definidos y asignados una medida y peso para evaluar la calidad del PL.

Es importante indicar que, para las insignias de Resultados Validados *Reproducido y Replicado* se considera que los resultados hayan sido validados por experimentadores externos que no sean los autores originales y la diferencia entre ambas insignias es que para *Reproducido* los experimentadores utilicen parte del material del experimento original y para la insignia de *Replicado* los experimentadores no hagan uso de ningún material del experimento original. En este sentido, la característica de calidad para ambas insignias es *Verificable*. La tabla §5.1 muestra la lista de las características de calidad para las insignias de la ACM. En vista que la ACM no define valores cuantitativos para obtener las insignias de los artefactos, en la presente memoria de tesis se establece medidas para cada parámetro de calidad.

Los elementos que conforman el modelo de calidad propuesto en la presente investigación son:

Características de calidad. Se refiere a las propiedades deseables que debería tener el PL para obtener las insignias de la ACM.

Parámetros de calidad. Representa los criterios que se desea evaluar en relación a una medida definida para obtener las insignias de la ACM.

Evidencia de medida. Se relaciona con los artefactos que permitan demostrar el cumplimiento de cada parámetro de calidad.

Peso del Parámetro de calidad. Para cada parámetro de calidad se asigna un valor (peso) en porcentaje. El peso será asignado dependiendo del número de evidencias de medida. El sumatorio de los pesos de cada parámetro de calidad para alcanzar alguna insignia será del 100%.

Tabla 5.1: Características de calidad para Paquetes de Laboratorio acorde a la ACM

AE - Funcional	AE - Reutilizable	AE - Disponible	RV - Reproducido	RV - Replicado
				
Documentado Coherente Completo Ejercitable	Documentado* Coherente Completo Ejercitable* Bien estructurado	Accesible	Verificable (usando material original)	Verificable (sin usar material original)

AE= Artefactos Evaluados.

RV= Resultados Validados.

Documentado*= Contiene Parámetros de calidad adicionales a la característica Documentado de la insignia Funcional.

Ejercitable*= Contiene Parámetros de calidad adicionales a la característica Ejercitable de la insignia Funcional.

Métrica Constituye las medidas cuantitativas del comportamiento de los parámetros de calidad.

En la presente propuesta, es posible calcular la calidad del Paquete de Laboratorio ($Q(LP)$ en la ecuación (5.1) para cada una de las insignias que aparecen en la tabla §5.1. De forma que, la calidad de un Paquete de Laboratorio dado, de acuerdo con una insignia concreta de la ACM (ACMb en la ecuación (5.1)), viene dada por el sumatorio del *Peso de cada Parámetro de calidad* en porcentaje (%) multiplicado por el *Resultado de cada Métrica*. Siendo $i = (1...n)$ la característica de calidad i -ésima de la correspondiente insignia de la ACM. Siendo PP_i = el Peso del Parámetro de calidad i -ésimo. Siendo Rm_i) = el Resultado de cada métrica. La ecuación (5.1) es la siguiente,

$$Q(LP)_{ACMb} = \sum iPP_i * Rm_i \tag{5.1}$$

Para darle peso a los parámetros de calidad de las insignias de la ACM, se consideran dos casos: *i*) Experimentos que incluyen scripts, software y código fuente y *ii*) Experimentos que no incluyen scripts, software y código fuente. Al conjunto de parámetros requeridos en los dos casos se le denomina (PR), al conjunto total de parámetros de cada insignia se le denomina (TP) y al conjunto de parámetros condicionados, es decir, que sólo existe en el primer caso se denomina (PC) ya que su existencia es cuando el experimento incluye scripts, software y código fuente.

De forma que, el peso de cada parámetro (PP_i en la ecuación (5.2)), viene dada por la división entre el número de parámetros PR dividido para la diferencia entre TP menos el número de parámetros PC; todo esto dividido para el número de parámetros PR. El resultado se multiplica por 100 para representar en porcentaje. La ecuación (5.2) es la siguiente,

$$PP_i = (((PR)/(TP - PC))/PR) * 100(\%) \quad (5.2)$$

5.4.4 Características de calidad para la insignia Funcional

Para la insignia *Funcional* se determinan cuatro características de calidad: *Documentado*, *Coherente*, *Completo* y *Ejercitable*. Para estas características se identifican en total 16 parámetros de calidad a los cuales se asigna un peso (valor en porcentaje) para cumplir con el 100% y otorgarle al PL la insignia *Funcional*.

A continuación, se define las características mínimas del modelo de calidad para la insignia *Funcional*.

- Documentado.- Un PL se considera *Documentado* cuando el conjunto de los artefactos que lo componen cumplen los siguientes Parámetros de calidad: *i)* Los artefactos están registrados de manera detallada y su contenido es explicado mediante una lista de artefactos. Esta lista está representada por un inventario de artefactos estructurado y organizado que proporciona una descripción suficiente para poder ejercitarlos [185]; *ii)* La descripción de los artefactos se presenta en un archivo "readme", y que contiene los apartados siguientes: *Qué hace* y *Cómo se obtiene* los artefactos, el *Contexto*, la *Procedencia de datos*, *Claves* y *contraseñas* ocultas de ser necesario, *Declaraciones éticas y legales*, *Requisitos de almacenamiento*, *Aspectos de instalación* incluido un ejemplo mediante formularios registrados por el autor, toda ésta información se debe registrar y documentar [6].

Dado que los artefactos deben estar documentados, una forma de medir es registrar ambos artefactos en el PL. La *Métrica* para cada parámetro de calidad de la característica "Documentado" hace referencia al álgebra de Boole [121].

Archivo "inventario". La Métrica para el inventario de artefactos es:

$$Q_{1.1} = \{0,1\} \quad (5.3)$$

El valor es 0 cuando el (archivo "inventario" no se registra) y 1 cuando el (archivo "inventario" sí se registra).

Archivo "readme". La Métrica para la descripción de artefactos es:

$$Q_{1.2} = \{0,1\} \quad (5.4)$$

El valor es 0 cuando el (archivo "readme" no se registra) y 1 cuando el (archivo "readme" sí se registra).

- **Coherente.-** Según la ACM [185], un PL se considera *Coherente* cuando los artefactos son relevantes para el estudio empírico realizado y contribuyen de alguna manera inherente, es decir, esencial y permanente a la generación de sus resultados principales. Para definir la coherencia de los artefactos se considera que si los artefactos pertenecen a las tareas que se llevan a cabo en el experimento por naturaleza serán relevantes y en si necesarios para ejecutar el experimento. En este sentido, el hecho de que en la presente propuesta de tesis se vinculen los artefactos a las tareas del experimento trae como consecuencia que el PL sea coherente, ya que no debería aparecer ningún artefacto extraño es decir, artefactos que no contribuyan a la ejecución o análisis del experimento.

Tareas que necesitan artefactos. Dado que los artefactos deben ser coherentes, una forma de medir es evidenciar que debe existir un número considerable de artefactos generados y que pertenecen a las tareas. Dónde, n_{ta} representa el número de tareas con artefactos y n_{tt} representa el total de tareas del proceso experimental que necesitan artefactos. La métrica se define de la siguiente manera:

$$Q_2 = \frac{n_{ta}}{n_{tt}} \quad (5.5)$$

El rango de medición para Q_2 oscila entre 0 (ninguna de las tareas que necesitan artefactos se registra) y 1 (sí se registra tareas que necesitan artefactos), es decir, $0 \leq Q_2 \leq 1$. Se considera que los artefactos son coherentes para el experimento, cuando el valor de Q_2 es 1 o se acerca a este valor, esto significa que los artefactos son suficientes para completar el PL y por lo tanto la característica Coherente es aceptable.

- **Completo.-** Según la ACM, un PL se considera *Completo* cuando en la medida de lo posible, se incluyen todos los componentes relevantes para el artículo científico

en cuestión. Se considera componentes relevantes aquellos artefactos indispensables para llevar a cabo el experimento, es decir, todo lo que el experimentador debe hacer está incluido en el PL. Se debe conocer como obtener, descargar y ejecutar estos artefactos mediante directrices paso a paso fáciles de comprender para el experimentador. No es necesario incluir artefactos patentados. En el caso que si se requieren para ejercitar el paquete, esto debe documentarse, junto con instrucciones sobre cómo obtenerlos. Además, deben incluirse representantes de datos patentados para demostrar el análisis [185].

Obviamente, si los artefactos están incompletos, el experimentador no estará satisfecho cuando haga uso de los mismos, esto generará probablemente inconvenientes y suposiciones sobre el proceso experimental, y estas suposiciones pueden ser falsas, lo que lleva una vez más a experimentadores insatisfechos.

Medir la completitud de los artefactos es complejo, debido a que puede inferir el número de artefactos que necesariamente serán suficientes para llevar a cabo el experimento y aún más la relevancia de los mismos. Esta complejidad puede profundizarse cuando el experimentador decide crear el PL al finalizar las fases del proceso experimental, lo que probablemente puede generar que el paquete quede incompleto por artefactos que han sido olvidados o que posiblemente para ese momento ya no serán relevantes.

Para evitar de alguna manera esta complejidad la propuesta presentada en esta tesis esta enfocada en crear el PL orientado al proceso de manera que en la fase de ejecución del experimento vaya de la mano el desarrollo del PL. El número de artefactos en este caso serán los que el autor esta usando en el momento preciso de la ejecución y probablemente no deje de lado en documentarlos y completarlos. La relevancia estará menos discutida en ese instante porque el autor podrá determinar si tal o cuál artefacto vale la pena incluir en el PL.

Por lo tanto, se puede considerar que un PL es *Completo* si cumple con las características relacionadas a los niveles de profundidad del contenido del PL definidos en los estudios de Solari *et al.* [392, 393] y explicados en la sección §2.3. Además, el registro de los grupos de artefactos definidos en la sección §5.4.2. Para el presente estudio estas características serán consideradas como niveles de profundidad en el contenido del PL. Los niveles son: Artículo, Operacional, Descriptivo, Procedimental y Fundamentado. Teniendo en cuenta dichos niveles, la integración de la ejecución del experimento con la elaboración del PL permitirá que se cumpla la *Compleitud* de los niveles Operacional y Descriptivo. A nivel Procedimental, se propone generar una guía paso a paso que sea consecuencia

de haber especificado la secuencia de tareas realizadas durante la ejecución experimental. Para el nivel Fundamentado se propone que, la motivación que le llevó a los investigadores conducir el experimento deberá registrarse en el PL. La tabla §5.2 muestra una comparativa entre otras propuestas con respecto al contenido del PL definidas en la sección §2.4 y la que se presenta en esta tesis, para esclarecer la característica de "Compleitud" de un "Buen" Paquete de Laboratorio.

Además, para que un PL cumpla la característica de calidad "Completo", es necesario se registren tres artefactos considerados en la guía de evaluación de artefactos propuesta por Abrahão y Mendez [6], que son: *i)* una guía de instrucciones para obtener y descargar los artefactos, *ii)* una guía de instrucciones para ejecutar los artefactos, ambas guías estarán descritas paso a paso explicando cada instrucción y *iii)* el archivo del artículo científico donde se reporta el experimento que será proporcionado por el autor. En el caso de no existir todavía el artículo publicado puede ser un preprint del mismo.

Tabla 5.2: Comparativa entre propuestas con respecto a los niveles de profundidad de Solari *et al.* [392] con respecto al contenido de los paquetes de laboratorio para experimentos en Ingeniería del Software

Autor	Año	Artículo	Operacional	Descriptivo	Procedimental	Fundamentado
Basili <i>et al.</i> [41]	1996	✓	✓	✓	x	✓
Shull <i>et al.</i> [382]	2004	✓	✓	✓	x	✓
Almqvist [21]	2006	✓	✓	✓	x	✓
Kitchenham <i>et al.</i> [221]	2006	x	✓	✓	x	✓
Vegas <i>et al.</i> [416]	2006	x	✓	✓	x	x
Solari <i>et al.</i> [393]	2006	✓	✓	✓	✓	✓
Shull <i>et al.</i> [381]	2008	x	✓	✓	x	x
Jedlitschka <i>et al.</i> [194]	2008	x	✓	✓	x	x
Solari <i>et al.</i> [394]	2018	✓	✓	✓	✓	✓
Propuesta de esta tesis	2022	✓	✓	✓	✓	✓

Para calcular el porcentaje de cumplimiento de la característica de calidad *Completo* se definen en total siete *Parámetros de calidad*.

Artículo científico. Representa el primer nivel de completitud, se refiere a registrar el artículo científico o preprint del experimento original. La *Métrica* hace

referencia al álgebra de Boole [121] y se define de la siguiente manera:

$$Q_{3.1} = \{0,1\} \quad (5.6)$$

El valor es 0 cuando el (artículo científico o preprint no se registra) y 1 cuando el (artículo científico o preprint sí se registra).

Nivel Operacional. Representa el segundo nivel de completitud, se refiere a registrar los artefactos de las tareas que pertenecen al nivel operacional del PL. Los artefactos en este nivel pueden ser: artefactos para la producción del software, instrucciones para los sujetos, formularios, instrumentos de medición, modelos conceptuales, diagramas UML. Dónde, n_{no} representa el número de las tareas con artefactos del nivel operacional y n_{to} representa el total de tareas que necesitan artefactos del nivel operacional. La métrica se define de la siguiente manera:

$$Q_{3.2} = \frac{n_{no}}{n_{to}} \quad (5.7)$$

El rango de medición para $Q_{3.2}$ oscila entre 0 (ninguno de los artefactos de las tareas que pertenecen al nivel operacional se registra) y 1 (sí se registra los artefactos de las tareas que pertenecen al nivel operacional), es decir, $0 \leq Q_{3.2} \leq 1$. Se considera que los artefactos para el nivel operacional son completos para el experimento, cuando el valor de $Q_{3.2}$ es 1 o se acerca a este valor.

Nivel Descriptivo. Representa el tercer nivel de completitud, se refiere a los artefactos que describen el diseño utilizado, explicación del contexto. Dónde, n_{nd} representa el número de las tareas con artefactos del nivel descriptivo y n_{td} representa el total de tareas que necesitan artefactos del nivel operacional. La métrica se define de la siguiente manera:

$$Q_{3.3} = \frac{n_{nd}}{n_{td}} \quad (5.8)$$

El rango de medición para $Q_{3.3}$ oscila entre 0 (ninguno de los artefactos de las tareas que pertenecen al nivel descriptivo se registra) y 1 (sí se registra los artefactos de las tareas que pertenecen al nivel descriptivo), es decir, $0 \leq Q_{3.3} \leq 1$. Se considera que los artefactos para el nivel descriptivo son completos para el experimento, cuando el valor de $Q_{3.3}$ es 1 o se acerca a este

valor.

Nivel Procedimental. Representa el cuarto nivel de completitud, se refiere a registrar guías paso a paso para el proceso experimental para ser usadas por los replicadores. Dónde, n_{np} representa el número de las tareas con artefactos del nivel procedimental y n_{tp} representa el total de tareas que necesitan artefactos del nivel procedimental. La métrica se define de la siguiente manera:

$$Q_{3.4} = \frac{n_{np}}{n_{tp}} \quad (5.9)$$

El rango de medición para $Q_{3.4}$ oscila entre 0 (ninguno de los artefactos de las tareas que pertenecen al nivel procedimental se registra) y 1 (sí se registra los artefactos de las tareas que pertenecen al nivel procedimental), es decir, $0 \leq Q_{3.4} \leq 1$. Se considera que los artefactos para el nivel procedimental son completos para el experimento, cuando el valor de $Q_{3.4}$ es 1 o se acerca a este valor.

Nivel Fundamental. Representa el quinto nivel de completitud, se refiere a registrar el archivo "decision" en el cual se justifique las decisiones tomadas por los experimentadores originales para realizar el experimento con la finalidad de direccionar al replicador el camino a seguir y a la vez que tomen sus propias decisiones. La *Métrica* hace referencia al álgebra de Boole [121] y se define de la siguiente manera:

$$Q_{3.5} = \{0,1\} \quad (5.10)$$

El valor es 0 cuando el (archivo "decision" no se registra) y 1 cuando el (archivo "decision" sí se registra).

Guía de instrucciones para descargar los artefactos. El *sexto parámetro*, se refiere a registrar la guía de instrucciones para descargar los artefactos. La *Métrica* hace referencia al álgebra de Boole [121] y se define de la siguiente manera:

$$Q_{3.6} = \{0,1\} \quad (5.11)$$

El valor es 0 cuando la (guía de instrucciones para descargar los artefactos no se registra) y 1 cuando la (guía de instrucciones para descargar los artefactos sí se registra).

Guía de instrucciones para ejecutar los artefactos. El *séptimo parámetro*, se refiere a registrar la guía de instrucciones para ejecutar los artefactos. La *Métrica* hace referencia al álgebra de Boole [121] y se define de la siguiente manera:

$$Q_{3.7} = \{0,1\} \quad (5.12)$$

El valor es 0 cuando la (guía de instrucciones para ejecutar los artefactos no se registra) y 1 cuando la (guía de instrucciones para ejecutar los artefactos sí se registra).

- Ejercitable.- Según la ACM, se considera *Ejercitable* cuando los scripts y el software incluidos que se utilizan para generar los resultados en el documento asociado se pueden ejecutar con éxito, y se puede acceder a los datos incluidos y manipularlos adecuadamente [185].

Los scripts y software para generar los resultados se describirán sistemáticamente con el siguiente contenido: nombre script y software, enlace (link) de descarga o información de los script y software, requisitos y recomendaciones de los scripts y software. La descripción de los scripts y software se registrará en archivos, mientras que, el evaluador definirá si o no el acceso y manipulación adecuada de los datos incluidos es posible.

Cuando el PL no necesite describir los scripts al igual que el software porque seguramente en el experimento no se han utilizado, se debe omitir la descripción de los scripts y software al igual que, el registro de su ejecución con éxito.

Dado que los artefactos deben ser ejecutables, una forma de medir es que se incluya seis *Parámetros de calidad*: *i)* el archivo de descripción de los scripts, *ii)* el archivo de descripción del software, *iii)* el registro que se confirme que los scripts se ejecutan con éxito, *iv)* el registro que se confirme que el software se ejecutan con éxito, *v)* se confirme que los datos incluidos son accesibles y *vi)* se confirme que los datos incluidos se manipulan adecuadamente.

Descripción de los scripts. El *primer parámetro*, se refiere a describir los scripts que generan los resultados. Dónde, n_{sd} representa el número de scripts des-

critos sistemáticamente y n_{ts} representa el total de scripts registrados. La métrica se define de la siguiente manera:

$$Q_{4.1} = \frac{n_{sd}}{n_{ts}} \quad (5.13)$$

El rango de medición para $Q_{4.1}$ oscila entre 0 (ninguno de los scripts se describe sistemáticamente) y 1 (sí se describen sistemáticamente los scripts), es decir, $0 \leq Q_{4.1} \leq 1$. Cuando el valor de $Q_{4.1}$ es 1 o se acerca a este valor, significa que los scripts para generar los resultados se describen sistemáticamente en su totalidad o en su gran mayoría.

Descripción del software. El *segundo parámetro*, se refiere a describir el software que generan los resultados. Dónde, n_{sf} representa el número de software descritos sistemáticamente y n_{tf} representa el total de software registrados. La métrica se define de la siguiente manera:

$$Q_{4.2} = \frac{n_{sf}}{n_{tf}} \quad (5.14)$$

El rango de medición para $Q_{4.2}$ oscila entre 0 (ninguno de los software se describe sistemáticamente) y 1 (sí se describen sistemáticamente los software), es decir, $0 \leq Q_{4.2} \leq 1$. Cuando el valor de $Q_{4.2}$ es 1 o se acerca a este valor, significa que los software para generar los resultados se describen sistemáticamente en su totalidad o en su gran mayoría.

Scripts que se ejecutan con éxito. El *tercer parámetro*, se refiere a evidenciar que los scripts que generan los resultados se ejecutan con éxito. Dónde, n_{se} representa el número de scripts ejecutados con éxito y n_{ts} representa el total de scripts registrados. La métrica se define de la siguiente manera:

$$Q_{4.3} = \frac{n_{se}}{n_{ts}} \quad (5.15)$$

El rango de medición para $Q_{4.3}$ oscila entre 0 (ninguno de los scripts registrados se ejecutan con éxito) y 1 (sí se ejecutan con éxito los scripts registrados), es decir, $0 \leq Q_{4.3} \leq 1$. Cuando el valor de $Q_{4.3}$ es 1 o se acerca a este valor, significa que los scripts para generar los resultados se ejecutan con éxito en su totalidad o en su gran mayoría.

Software que se ejecuta con éxito. El *cuarto parámetro*, se refiere a evidenciar que el software que generan los resultados se ejecutan con éxito. Dónde, n_{fe} representa el número de software ejecutados con éxito y n_{tf} representa el total de software registrados. La métrica se define de la siguiente manera:

$$Q_{4.4} = \frac{n_{fe}}{n_{tf}} \quad (5.16)$$

El rango de medición para $Q_{4.4}$ oscila entre 0 (ninguno de los software registrados se ejecutan con éxito) y 1 (sí se ejecutan con éxito los software registrados), es decir, $0 \leq Q_{4.4} \leq 1$. Cuando el valor de $Q_{4.4}$ es 1 o se acerca a este valor, significa que los software para generar los resultados se ejecutan con éxito en su totalidad o en su gran mayoría.

Datos incluidos accesibles. El *quinto parámetro*, se refiere a evidenciar que los datos incluidos son accesibles. Dónde, n_{da} representa el número de datos incluidos accesibles y n_{td} representa el total de datos incluidos. La métrica se define de la siguiente manera:

$$Q_{4.5} = \frac{n_{da}}{n_{td}} \quad (5.17)$$

El rango de medición para $Q_{4.5}$ oscila entre 0 (ninguno de los datos incluidos es accesible) y 1 (sí son accesibles los datos incluidos), es decir, $0 \leq Q_{4.5} \leq 1$. Cuando el valor de $Q_{4.5}$ es 1 o se acerca a este valor, significa que los datos incluidos son accesibles en su totalidad o en su gran mayoría.

Datos incluidos se manipulan adecuadamente. El *sexto parámetro*, se refiere a evidenciar que los datos incluidos se manipulan adecuadamente, es decir, que el investigador pueda hacer cambios, editar los datos incluidos en el PL. Dónde, n_{dm} representa el número de datos incluidos que se manipulan adecuadamente y n_{td} representa el total de datos incluidos. La métrica se define de la siguiente manera:

$$Q_{4.6} = \frac{n_{dm}}{n_{td}} \quad (5.18)$$

El rango de medición para $Q_{4.6}$ oscila entre 0 (ninguno de los datos incluidos se manipula adecuadamente) y 1 (sí se manipulan adecuadamente los datos incluidos), es decir, $0 \leq Q_{4.6} \leq 1$. Cuando el valor de $Q_{4.6}$ es 1 o se acerca a

este valor, significa que los datos incluidos se manipulan adecuadamente en su totalidad o en su gran mayoría.

Para resumir el modelo de calidad para paquetes de laboratorio orientado al proceso acorde a la ACM con respecto a la insignia *Funcional*, se presenta la tabla §5.3 dónde se definen las características de calidad.

Tabla 5.3: Modelo de calidad para Paquetes de Laboratorio (PL) acorde a la ACM - Artefactos Evaluados - Funcional

Comienzo de tabla §5.3

Artefacto Evaluado - Funcional


Característica	Parámetro de calidad	Evidencia de medida	Peso parámetro	Métrica
Documentado	Inventario de artefactos.	Archivo "inventario".	{PP ₁ , PP ₂ , PP ₃ }	Q _{1.1} = {0,1}
	Descripción de artefactos.	Archivo "readme".	{PP ₁ , PP ₂ , PP ₃ }	Q _{1.2} = {0,1}
Coherente	Los artefactos son relevantes y contribuyen a los principales resultados.	Tareas que necesitan artefactos.	{PP ₁ , PP ₂ , PP ₃ }	Q ₂ = $\frac{n_{ta}}{n_{tt}}$ 0 ≤ Q ₂ ≤ 1
	Reporte del experimento original.	Artículo científico.	{PP ₁ , PP ₂ , PP ₃ }	Q _{3.1} = {0,1}
Completo	Artefactos del nivel operacional.	Artefactos de las tareas que pertenecen al nivel operacional.	{PP ₁ , PP ₂ , PP ₃ }	Q _{3.2} = $\frac{n_{no}}{n_{to}}$ 0 ≤ Q _{3.2} ≤ 1
	Artefactos del nivel descriptivo.	Artefactos de las tareas que pertenecen al nivel descriptivo.	{PP ₁ , PP ₂ , PP ₃ }	Q _{3.3} = $\frac{n_{nd}}{n_{td}}$ 0 ≤ Q _{3.3} ≤ 1
	Artefactos del nivel procedimental.	Artefactos de las tareas que pertenecen al nivel procedimental.	{PP ₁ , PP ₂ , PP ₃ }	Q _{3.4} = $\frac{n_{np}}{n_{tp}}$ 0 ≤ Q _{3.4} ≤ 1
	Artefactos del nivel fundamentado.	Archivo "decision".	{PP ₁ , PP ₂ , PP ₃ }	Q _{3.5} = {0,1}
	Instrucciones para descargar los artefactos.	Guía de instrucciones de descarga.	{PP ₁ , PP ₂ , PP ₃ }	Q _{3.6} = {0,1}
	Instrucciones para ejecutar los artefactos.	Guía de instrucciones de ejecución.	{PP ₁ , PP ₂ , PP ₃ }	Q _{3.7} = {0,1}

Continuación de la Tabla §5.3				
Característica	Parámetro de calidad	Evidencia de medida	Peso parámetro	Métrica
Ejercitable	Descripción sistemática de los scripts para generar los resultados (PC_{sc}).	Descripción de los scripts.	$\{0, PP_1, PP_3\}$	$Q_{4.1} = \frac{n_{sd}}{n_{ts}}$ $0 \leq Q_{4.1} \leq 1$
	Descripción sistemática del software para generar los resultados (PC_{sf}).	Descripción del software.	$\{0, PP_1, PP_3\}$	$Q_{4.2} = \frac{n_{sf}}{n_{tf}}$ $0 \leq Q_{4.2} \leq 1$
	Ejecución exitosa de los scripts para generar los resultados (PC_{sc}).	Scripts que se ejecutan con éxito.	$\{0, PP_1, PP_3\}$	$Q_{4.3} = \frac{n_{se}}{n_{ts}}$ $0 \leq Q_{4.3} \leq 1$
	Ejecución exitosa del software para generar los resultados (PC_{sf}).	Software que se ejecutan con éxito.	$\{0, PP_1, PP_3\}$	$Q_{4.4} = \frac{n_{fe}}{n_{tf}}$ $0 \leq Q_{4.4} \leq 1$
	Acceso a datos incluidos.	Datos accesibles.	$\{PP_1, PP_2, PP_3\}$	$Q_{4.5} = \frac{n_{da}}{n_{td}}$ $0 \leq Q_{4.5} \leq 1$
	Manipulación adecuada de los datos incluidos.	Datos manipulados adecuadamente.	$\{PP_1, PP_2, PP_3\}$	$Q_{4.6} = \frac{n_{dm}}{n_{td}}$ $0 \leq Q_{4.6} \leq 1$
Total			100%	

Continuación de la Tabla §5.3				
Característica	Parámetro de calidad	Evidencia de medida	Peso parámetro	Métrica
				{0,1} = Se verifica el registro del parámetro de calidad.
				Parámetro condicionado scripts= Se verifica si el experimento necesita o no registrar scripts= PC_{sc}
				Parámetro condicionado software= Se verifica si el experimento necesita o no registrar software = PC_{sf}
				Peso Parámetro = PP
				Número de Parámetros que no están condicionados, es decir, son requeridos = PR
				Número de Parámetros que están condicionados = PC
				Número total de Parámetros de cada insignia = TP
				Cálculo del Peso Parámetro con TP en {1} = $PP_1 = (((PR)/(TP - PC))/PR)*100(\%) = \{6.25\%$
				Cálculo del Peso Parámetro con PC en {0} = $PP_2 = (((PR)/(TP - PC))/PR)*100(\%) = \{8.34\%$
				Cálculo del Peso Parámetro con PC en {0,1}, es decir,
				el experimento registra scripst o software = $PP_3 = (((PR)/(TP - PC))/PR)*100(\%) = \{7.15\%$.
				{ PP_1, PP_2, PP_3 } = A cada parámetro de calidad se asigna uno de los tres pesos dependiendo si registra o no PC_{sc} y PC_{sf} .
				n_{no} = número de tareas con artefactos del nivel operacional.
				n_{to} = total de tareas que necesitan artefactos del nivel operacional.
				n_{nd} = número de tareas con artefactos del nivel descriptivo.
				n_{td} = total de tareas que necesitan artefactos del nivel descriptivo.
				n_{np} = número de tareas con artefactos del nivel procedimental.
				n_{tp} = total de tareas que necesitan artefactos del nivel procedimental.
				n_{ta} = número de tareas con artefactos.
				n_{tt} = total de tareas que necesitan artefactos.
				n_{sd} = número de scripts descritos sistemáticamente.
				n_{ts} = total de scripts registrados.
				n_{sf} = número de software descritos sistemáticamente.
				n_{tf} = total de software registrados.
				n_{se} = número de scripts ejecutados con éxito.
				n_{fe} = número de software ejecutados con éxito.
				n_{da} = número de datos accesibles.
				n_{td} = total de datos incluidos.
				n_{dm} = número de datos manipulados adecuadamente.

Fin de tabla §5.3

Según la guía de envío de los artefactos en el ICSE'2021 propuesta por Abrahão y Mendez [6], otros artefactos *opcionales* se hacen presentes en el PL. En la presente propuesta del modelo de calidad aunque los artefactos *opcionales* de la insignia *Funcional* no tienen un peso de medida definidos en este momento, son considerados importantes y necesarios que se registren en el PL, con el fin de proporcionar información adicional y generar más artefactos a los investigadores que deseen replicar el experimento. Sin embargo, la influencia de dichos artefactos en la calidad del PL será estudiado en

profundidad como trabajo futuro de la presente investigación.

Los artefactos *opcionales* para la insignia *Funcional* son:

- **Artefactos Opcionales.**- Se considera como artefactos *opcionales* de la insignia *Funcional*, al registro de indicaciones del tiempo necesario para ejecutar los artefactos de manera completa y también el registro de indicaciones del tiempo de ejecución de una versión más corta de los artefactos. Estos artefactos opcionales pueden presentarse cuando el PL registre Software. Estos artefactos *opcionales* apoyarán a los experimentadores en conocer la duración de ejecución de los artefactos.

Una forma de evidenciar dichos artefactos *opcionales* es incluir el registro de dos *Evidencias de medida*:

Indicadores de tiempo para ejecución completa. La *Métrica* para la primera evidencia de medida hace referencia al álgebra de Boole [121] y se define de la siguiente manera:

$$Q_{5.1} = \{0,1\} \quad (5.19)$$

El valor es 0 cuando los (indicadores de tiempo para ejecución completa de los artefactos no se registra) y 1 cuando los (indicadores de tiempo para ejecución completa de los artefactos sí se registra).

Indicadores de tiempo para ejecución corta. La *Métrica* para la segunda evidencia de medida hace referencia al álgebra de Boole [121] y se define de la siguiente manera:

$$Q_{5.2} = \{0,1\} \quad (5.20)$$

El valor es 0 cuando los (indicadores de tiempo para ejecución corta de los artefactos no se registra) y 1 cuando los (indicadores de tiempo para ejecución corta de los artefactos sí se registra).

Para resumir los artefactos *opcionales* con respecto a la insignia *Funcional*, se presenta la tabla §5.4.

5.4.5 Características de calidad para la insignia Reutilizable

Según la ACM, se considera que la insignia *Reutilizable* excede significativamente la funcionalidad mínima de los artefactos. Es decir, tiene y adopta todas las cualida-

Tabla 5.4: Artefactos opcionales del modelo de calidad para las insignias Funcional y Reutilizable

Artefacto opcional	Evidencia de medida	Métrica
Indicadores de tiempo para ejecución completa de los artefactos.	Registro de indicadores de tiempo para ejecución completa.	$Q_{5.1} = \{0,1\}$
Indicadores de tiempo para ejecución corta de los artefactos.	Registro de indicadores de tiempo para ejecución corta.	$Q_{5.2} = \{0,1\}$

Nota.- Estos artefactos opcionales son los mismos para ambas insignias "Funcional" y "Reutilizable".
 $\{0,1\}$ =Se verifica el registro del parámetro de calidad.

des de los Artefactos Evaluados *Funcional*, pero, además, están muy cuidadosamente documentados y bien estructurados en la medida en que se facilita la reutilización y readaptación [185]. Por consiguiente, todas las características de calidad de la insignia *Funcional* descritas anteriormente forman parte de la insignia *Reutilizable* y se definirá únicamente una característica de calidad adicional (*Bien estructurado*) y algunos parámetros adicionales de las características de calidad *Documentado* y *Ejercitable* de la insignia *Funcional*.

Para la insignia *Reutilizable* se determinan cinco características de calidad: *Bien estructurado*, *Documentado*, *Coherente*, *Completo* y *Ejercitable*. Para estas características se identifican en total 28 parámetros de calidad a los cuales se asigna un peso (valor en porcentaje) para cumplir con el 100% y otorgarle al PL la insignia *Reutilizable*.

A continuación, se definen las características mínimas del modelo de calidad para la insignia *Reutilizable*.

- **Bien estructurado.**- Los artefactos estarán estructurados en un conjunto de directorios para favorecer la comprensión por parte del nuevo investigador, la reutilización del mismo, y su adaptación al nuevo entorno del estudio en caso de ser necesario. Cada directorio contendrá los artefactos que poseen la misma finalidad, por ejemplo: todos los cuestionarios que cumplimentarán los participantes del estudio están juntos en un directorio. Otra posibilidad es organizar los artefactos por tarea experimental (se puede indicar el nombre de la tarea por ejemplo "Programación con GraphQL"). Otras formas de organización de los artefactos aparte de la finalidad o la tarea, son posibles como: orden cronológico, orden alfabético, formato de archivo o tamaño.

Artefactos bien estructurados. Dado que los artefactos deben estar bien estructurados, una forma de medir es que se registre el siguiente parámetro de

calidad: registro del directorio de los artefactos mediante un árbol bien estructurado. La estructura del directorio organizado con todos los artefactos estará disponibles en un PATH o una URL de acceso público.

La métrica para esta característica de calidad *Bien estructurado* hace referencia al álgebra de Boole [121]:

$$Q_6 = \{0,1\} \quad (5.21)$$

El valor es 0 cuando (no esta disponible mediante un PATH o URL de acceso público la estructura del directorio organizado con todos los artefactos) y 1 (sí esta disponible mediante un PATH o URL de acceso público la estructura del directorio organizado con todos los artefactos).

- Documentado*.- A más de los parámetros de calidad presentados en la característica *Documentado* de la insignia *Funcional* se suman las siguientes: *i)* Las normas y estándares de la comunidad investigadora para los artefactos de este tipo se respetarán estrictamente [6]. Esto conlleva a que en el área de IS, se analice y se ponga en práctica las recomendaciones y guía de la National Information Standards Organization (NISO RP-31-2021) para Insignias y Definiciones de Reproducibilidad [295]. Además, se hace referencia a la taxonomía de la reproducibilidad compatible para las Ciencias de la Computación y la Informática del reporte de Sandia National Laboratories [178]. De la misma manera, se hace uso de la guía propuesta por Abrahão y Mendez para el envío y revisión de los artefactos en el Track de AE-Artifact Evaluation del congreso ICSE 2021. *ii)* Los datos informativos como nombre, email, sitio web, afiliación de los autores del experimento serán registrados en un documento de acceso público para ser contactados por otros investigadores, además se debe marcar el autor correspondiente, esta información se registrará en un archivo denominado "*contact*". *iii)* Este parámetro se considerará o no cuando el PL registre software. es decir, si los artefactos se centran en software se describirá los requisitos para el entorno de hardware (rendimiento, almacenamiento, periféricos). Además, se describirá los requisitos para el entorno de software (Docker, VM, Sistema Operativo). Se describirá la información de control de versiones y se justificará cualquier desviación en dichos entornos, esta información se registrará en un archivo denominado "*requirements*". *iv)* Los autores describirán las razones por las que deberían otorgarles la insignia de *Reutilizable* a su PL, es decir, justificarán por qué su PL merece la insignia de *Reutilizable*. Esta información se registrará en un archivo denominado "*status*". *v)* Los derechos de distribución de los artefactos serán descritos junto

con la licencia de código abierto, esta información se registrará en un archivo denominado "licence" [6].

Los Parámetros de calidad de la característica *Documentado* definidos en la insignia *Funcional* (2 Parámetros) y los adicionales (5 Parámetros) que se describen anteriormente para la característica *Documentado** de la insignia *Reutilizable* deben registrarse en el PL. En total se registra siete Parámetros de calidad.

Archivos "inventario" y "readme". Estos archivos constituyen el primer y segundo parámetros y son idénticos a los descritos en la insignia *Funcional* (Sección §5.4.4).

Normas y estándares. Es el tercer parámetro y se refiere a respetar estrictamente las normas y estándares de la comunidad investigadora para los artefactos del PL. Dónde, n_{nr} representa el número de normas y estándares que se respetan estrictamente y n_{te} representa el total de normas y estándares registrados. La métrica se define de la siguiente manera:

$$Q_{1.3} = \frac{n_{nr}}{n_{te}} \quad (5.22)$$

El rango de medición para $Q_{1.3}$ oscila entre 0 (ninguna de las normas y estándares se respetan estrictamente) y 1 (sí se respetan estrictamente las normas y estándares de los artefactos), es decir, $0 \leq Q_{1.3} \leq 1$. Cuando el valor de $Q_{1.3}$ es 1 o se acerca a este valor, significa que las normas y estándares se respetan estrictamente en su totalidad o en su gran mayoría.

Archivo "contact". El *cuarto parámetro*, se refiere a registrar el archivo "contact", que contiene información como nombre, email, sitio web, afiliación de los autores externos del experimento y el registro del autor correspondiente. La *Métrica* hace referencia al álgebra de Boole [121] y se define de la siguiente manera:

$$Q_{1.4} = \{0,1\} \quad (5.23)$$

El valor es 0 cuando el (archivo "contact" no se registra) y 1 cuando el (archivo "contact" sí se registra).

Archivo "requirements". El *quinto parámetro*, se refiere a registrar el archivo "requirements", que describe los requisitos de los artefactos que se centran en el software y hardware, es decir, las restricciones técnicas de software y hardware como el sistema operativo o el entorno de desarrollo. Este parámetro tiene sentido si el experimento controlado incluye hardware o software. La *Métrica* hace referencia al álgebra de Boole [121] y se define de la siguiente manera:

$$Q_{1.5} = \{0,1\} \quad (5.24)$$

El valor es 0 cuando el (archivo "requirements" no se registra y el experimento si tiene hardware y software) y 1 cuando el (archivo "requirements" sí se registra).

Archivo "status". El *sexto parámetro*, se refiere a registrar el archivo "status", en el cual, los autores del PL indican qué insignia le otorgarían ellos mismos al PL explicando los motivos y las razones, es decir, describir una justificación al respecto. La *Métrica* hace referencia al álgebra de Boole [121] y se define de la siguiente manera:

$$Q_{1.6} = \{0,1\} \quad (5.25)$$

El valor es 0 cuando el (archivo "status" no se registra) y 1 cuando el (archivo "status" sí se registra).

Archivo "licence". El *séptimo parámetro*, se refiere a registrar el archivo "licence", que describe los derechos de distribución del PL junto con la licencia de código abierto. Este parámetro tiene sentido si el experimento controlado incluye código fuente. La *Métrica* hace referencia al álgebra de Boole [121] y se define de la siguiente manera:

$$Q_{1.7} = \{0,1\} \quad (5.26)$$

El valor es 0 cuando el (archivo "licence" no se registra y el experimento si tiene código fuente) y 1 cuando el (archivo "licence" sí se registra).

- **Coherente y Completo.**- Estas dos características de calidad coinciden en la definición con la insignia *Funcional* (Sección §5.4.4), en este punto sólo cambia el peso, como se muestra en la tabla §5.5.

- Ejercitable*.- A más de los parámetros de calidad presentados en la característica de calidad *Ejercitable* de la insignia *Funcional* se suman las siguientes: *i)* Se debe registrar el tipo de paquete, que será un paquete simple y/o un paquete de instalación. El paquete simple en donde están los artefactos documentales en un formato legible para documentos, además todos los artefactos están guardados en un sólo archivo de paquetes (zip o tar.gz). El paquete de instalación se elabora cuando el artefacto consiste en una herramienta o un sistema de software, contiene el material de instalación, configuración, una imagen de contenedor de Docker o de VirtualBox VM, es decir todas las herramientas necesarias para usar el artefacto. *ii)* Se debe indicar si el paquete de instalación incluye entornos virtuales. *iii)* Se debe indicar si el paquete de instalación tiene Scripts de instalación e instrucciones paso a paso para la instalación automática de herramientas y entornos necesarios, es decir, presentar una guía de instalación. *iv)* Los artefactos del paquete simple estarán comprimidos en un sólo archivo. *v)* Las instrucciones de instalación serán descritas mediante un ejemplo ilustrado para probar la instalación, ésta información se registrará en un archivo denominado "install". *vi)* El evaluador deberá revisar que la configuración e instalación de los artefactos dure menos de 30 minutos, y se registrará Sí o No el cumplimiento del mismo [6].

Cuando no se registra el PL de instalación en el experimento, es decir, no se registra software, no es necesario que los *Parámetros de calidad* asociados al PL de instalación sean considerados, por ejemplo: incluir los entornos virtuales, guías de instalación, el archivo "install", evidenciar que la instalación de los artefactos duren menos de 30 minutos.

Los Parámetro de calidad de la característica *Ejercitable* definidos en la insignia *Funcional* (6 Parámetros) y los adicionales (6 Parámetros) que se describen anteriormente para la característica *Ejercitable** de la insignia *Reutilizable* deben registrarse en el PL.

Descripción de los scripts y Scripts que se ejecutan con éxito. Estos parámetros son idénticos a los descritos en la insignia *Funcional* (Sección §5.4.4). El peso para cada parámetro esta sujeto a tres excepciones debido a que existe la posibilidad que el experimento no registre software y no registre scripts.

Descripción del software y Software que se ejecutan con éxito. Estos parámetros son idénticos a los descritos en la insignia *Funcional* (Sección §5.4.4). El peso para cada parámetro esta sujeto a tres excepciones debido a que existe la

posibilidad que el experimento no registre software y no registre scripts.

Datos incluidos accesibles. Este parámetro es idéntico al descrito en la insignia Funcional (Sección §5.4.4).

Datos incluidos se manipulan adecuadamente. Este parámetro es idéntico al descrito en la insignia Funcional (Sección §5.4.4).

Tipo Paquete. Este parámetro se refiere a registrar el tipo de paquete, si es un paquete simple y/o de instalación. La *Métrica* hace referencia al álgebra de Boole [121] y se define de la siguiente manera:

$$Q_{4.7} = \{0,1\} \quad (5.27)$$

El valor es 0 cuando el (tipo de paquete no se registra) y 1 cuando el (tipo de paquete sí se registra).

Registro de entornos virtuales. Este parámetro se refiere a indicar si el paquete de instalación incluye entornos virtuales. La *Métrica* hace referencia al álgebra de Boole [121] y se define de la siguiente manera:

$$Q_{4.8} = \{0,1\} \quad (5.28)$$

El valor es 0 cuando (no se indica los entornos virtuales del paquete de instalación) y 1 cuando (sí se indica los entornos virtuales del paquete de instalación).

Guía instalación para herramientas y entornos virtuales. Este parámetro se refiere a registrar una guía de instalación para las herramientas y entornos necesarios. La *Métrica* hace referencia al álgebra de Boole [121] y se define de la siguiente manera:

$$Q_{4.9} = \{0,1\} \quad (5.29)$$

El valor es 0 cuando (no se registra una guía de instalación) y 1 cuando (sí se registra la guía de instalación).

PL comprimido en un sólo archivo. Este parámetro se refiere a evidenciar que los artefactos del PL están comprimidos en un sólo archivo. La *Métrica* hace referencia al álgebra de Boole [121] y se define de la siguiente manera:

$$Q_{4.10} = \{0,1\} \quad (5.30)$$

El valor es 0 cuando (no se registra el archivo comprimido que contiene los artefactos del PL) y 1 cuando (sí se registra el archivo comprimido que contiene los artefactos del PL).

Archivo "install". Este parámetro se refiere a registrar el archivo "install", que describe las instrucciones de instalación mediante ejemplos de las herramientas y entornos necesarios. La *Métrica* hace referencia al álgebra de Boole [121] y se define de la siguiente manera:

$$Q_{4.11} = \{0,1\} \quad (5.31)$$

El valor es 0 cuando el (archivo "install" no se registra) y 1 cuando el (archivo "install" sí se registra).

Instalación de artefactos menos de 30 minutos. Este parámetro se refiere a evidenciar que la configuración e instalación de los artefactos duran menos de 30 minutos. La *Métrica* hace referencia al álgebra de Boole [121] y se define de la siguiente manera:

$$Q_{4.12} = \{0,1\} \quad (5.32)$$

El valor es 0 cuando (no se registra que la configuración e instalación de los artefactos duran menos de 30 minutos) y 1 cuando (sí se registra que la configuración e instalación de los artefactos duran menos de 30 minutos).

Para resumir el modelo de calidad para paquetes de laboratorio orientado al proceso acorde a la ACM con respecto a la insignia *Reutilizable*, se presenta la tabla §5.5 dónde se define las características de calidad.

Tabla 5.5: Modelo de calidad para Paquetes de Laboratorio (PL) acorde a la ACM - Artefactos Evaluados - Reutilizable

Comienzo de tabla §5.5

Artefacto Evaluado - Reutilizable


Característica	Parámetro de calidad	Evidencia de medida	Peso parámetro	Métrica	
Bien estructurado	Artefactos bien estructurados y organizados para reutilización y adaptación.	Árbol bien estructurado con todos los artefactos disponibles.	$\{PP_1, PP_2, PP_3, PP_4, PP_5\}$	$Q_6 = \{0,1\}$	
	Inventario de los artefactos.	Archivo "inventario".	$\{PP_1, PP_2, PP_3, PP_4, PP_5\}$	$Q_{1.1} = \{0,1\}$	
	Descripción de los artefactos.	Archivo "readme".	$\{PP_1, PP_2, PP_3, PP_4, PP_5\}$	$Q_{1.2} = \{0,1\}$	
	Normas y estándares para los artefactos de este tipo se respetarán estrictamente.	Normas y estándares que se respetan estrictamente.	$\{PP_1, PP_2, PP_3, PP_4, PP_5\}$	$Q_{1.3} = \frac{n_{nr}}{n_{te}}$ $0 \leq Q_{1.3} \leq 1$	
	Datos informativos de los autores del experimento y el autor correspondiente serán registrados en un documento de acceso público.	Archivo "contact".	$\{PP_1, PP_2, PP_3, PP_4, PP_5\}$	$Q_{1.4} = \{0,1\}$	
	Documentado*	Describir los requisitos de los artefactos centrados en software y hardware (PC_{sf}).	Archivo "requirements".	$\{0, PP_1, PP_3, PP_5\}$	$Q_{1.5} = \{0,1\}$
		Describir las razones por las que los autores creen se debe otorgarles la insignia que reclaman.	Archivo "status".	$\{PP_1, PP_2, PP_3, PP_4, PP_5\}$	$Q_{1.6} = \{0,1\}$
		Describir los derechos de distribución del PL junto con la licencia de código abierto (PC_{cf}).	Archivo "licence".	$\{0, PP_1, PP_3, PP_4\}$	$Q_{1.7} = \{0,1\}$

Continuación de la Tabla §5.5

Característica	Parámetro de calidad	Evidencia de medida	Peso parámetro	Métrica	
Coherente	Los artefactos son relevantes y contribuyen a los principales resultados.	Tareas que necesitan artefactos.	$\{PP_1, PP_2, PP_3, PP_4, PP_5\}$	$Q_2 = \frac{n_{ta}}{n_{tt}}$ $0 \leq Q_2 \leq 1$	
	Reporte del experimento original.	Artículo científico.	$\{PP_1, PP_2, PP_3, PP_4, PP_5\}$	$Q_{3.1} = \{0,1\}$	
Completo	Artefactos del nivel operacional.	Artefactos de las tareas que pertenecen al nivel operacional.	$\{PP_1, PP_2, PP_3, PP_4, PP_5\}$	$Q_{3.2} = \frac{n_{no}}{n_{to}}$ $0 \leq Q_{3.2} \leq 1$	
	Artefactos del nivel descriptivo.	Artefactos de las tareas que pertenecen al nivel descriptivo.	$\{PP_1, PP_2, PP_3, PP_4, PP_5\}$	$Q_{3.3} = \frac{n_{nd}}{n_{td}}$ $0 \leq Q_{3.3} \leq 1$	
	Artefactos del nivel procedimental.	Artefactos de las tareas que pertenecen al nivel procedimental.	$\{PP_1, PP_2, PP_3, PP_4, PP_5\}$	$Q_{3.4} = \frac{n_{np}}{n_{tp}}$ $0 \leq Q_{3.4} \leq 1$	
	Artefactos del nivel fundamentado.	Archivo "decision".	$\{PP_1, PP_2, PP_3, PP_4, PP_5\}$	$Q_{3.5} = \{0,1\}$	
	Instrucciones para obtener y descargar los artefactos.	Guía de instrucciones de descarga.	$\{PP_1, PP_2, PP_3, PP_4, PP_5\}$	$Q_{3.6} = \{0,1\}$	
	Instrucciones para ejecutar los artefactos.	Guía de instrucciones de ejecución.	$\{PP_1, PP_2, PP_3, PP_4, PP_5\}$	$Q_{3.6} = \{0,1\}$	
	Ejercitable*	Descripción sistemática de los scripts para generar los resultados.	Descripción de los scripts (PC_{sc}).	$\{0, PP_1, PP_4, PP_5\}$	$Q_{4.1} = \frac{n_{sd}}{n_{ts}}$ $0 \leq Q_{4.1} \leq 1$
		Descripción sistemática del software para generar los resultados (PC_{sf}).	Descripción del software.	$\{0, PP_1, PP_3, PP_5\}$	$Q_{4.2} = \frac{n_{sf}}{n_{tf}}$ $0 \leq Q_{4.2} \leq 1$
Ejecución exitosa de los scripts para generar los resultados (PC_{sc}).		Scripts que se ejecutan con éxito.	$\{0, PP_1, PP_4, PP_5\}$	$Q_{4.3} = \frac{n_{se}}{n_{ts}}$ $0 \leq Q_{4.3} \leq 1$	
Ejecución exitosa del software para generar los resultados (PC_{sf}).		Software que se ejecutan con éxito.	$\{0, PP_1, PP_3, PP_5\}$	$Q_{4.4} = \frac{n_{fe}}{n_{tf}}$ $0 \leq Q_{4.4} \leq 1$	
Acceso a datos incluidos.		Datos accesibles.	$\{PP_1, PP_2, PP_3, PP_4, PP_5\}$	$Q_{4.5} = \frac{n_{da}}{n_{td}}$ $0 \leq Q_{4.5} \leq 1$	
Manipulación adecuada de los datos incluidos.		Datos manipulados adecuadamente.	$\{PP_1, PP_2, PP_3, PP_4, PP_5\}$	$Q_{4.6} = \frac{n_{dm}}{n_{ts}}$ $0 \leq Q_{4.4} \leq 1$	

Continuación de la Tabla §5.5				
Característica	Parámetro de calidad	Evidencia de medida	Peso parámetro	Métrica
Ejercitable*	Registrar el tipo de paquete, si es un paquete simple y/o un paquete de instalación.	Registro del tipo de paquete.	{ $PP_1, PP_2, PP_3, PP_4, PP_5$ }	$Q_{4.7} = \{0,1\}$
	Indicar si el paquete de instalación incluye entornos virtuales (PC_{sf}).	Describir los entornos virtuales del paquete de instalación.	{ $0, PP_1, PP_3, PP_5$ }	$Q_{4.8} = \{0,1\}$
	Registrar sí o no el paquete de instalación tiene scripts e instrucciones paso a paso para la instalación automática de herramientas y entornos necesarios (PC_{sf}).	Guía de instalación.	{ $0, PP_1, PP_3, PP_5$ }	$Q_{4.9} = \{0,1\}$
	Los artefactos del paquete simple deben estar comprimidos en un sólo archivo.	Artefactos en un sólo archivo comprimido.	{ $PP_1, PP_2, PP_3, PP_4, PP_5$ }	$Q_{4.10} = \{0,1\}$
	Las instrucciones de instalación serán descritas mediante un ejemplo ilustrado para probar la instalación (PC_{sf}).	Archivo "install".	{ $0, PP_1, PP_3, PP_5$ }	$Q_{4.11} = \{0,1\}$
	La configuración e instalación de los artefactos debe durar menos de 30 minutos (PC_{sf}).	Confirmar la configuración e instalación de los artefactos duran menos de 30 minutos.	{ $0, PP_1, PP_3, PP_5$ }	$Q_{4.12} = \{0,1\}$
	Total			100%

Continuación de la Tabla §5.5				
Característica	Parámetro de calidad	Evidencia de medida	Peso parámetro	Métrica
				{0,1} =Se verifica el registro del parámetro de calidad.
				Parámetro condicionado scripts= Se verifica si el experimento necesita o no registrar scripts= PC_{sc}
				Parámetro condicionado software= Se verifica si el experimento necesita o no registrar software= PC_{sf}
				Parámetro condicionado código fuente= Se verifica si el experimento necesita o no registrar código fuente= PC_{cf}
				Peso Parámetro= PP
				Número de Parámetros que no están condicionados, es decir, son requeridos = PR
				Número de Parámetros que están condicionados = PC
				Número total de Parámetros de cada insignia = TP
				Cálculo del Peso Parámetro con TP en {1} = $PP_1 = (((PR)/(TP - PC))/PR)*100(\%) = \{3.6\%$
				Cálculo del Peso Parámetro con PC en {0} = $PP_2 = (((PR)/(TP - PC))/PR)*100(\%) = \{5.6\%$
				Cálculo del Peso Parámetro con $PC_{sc} = PP_3 = (((PR)/(TP - PC))/PR)*100(\%) = \{3.85\%$
				Cálculo del Peso Parámetro con $PC_{sf} = PP_4 = (((PR)/(TP - PC))/PR)*100(\%) = \{4.8\%$
				Cálculo del Peso Parámetro con $PC_{cf} = PP_5 = (((PR)/(TP - PC))/PR)*100(\%) = \{3.71\%$
				{ $PP_1, PP_2, PP_3, PP_4, PP_5$ } = A cada parámetro de calidad se asigna uno de los cinco pesos dependiendo si registra o no PC_{sc}, PC_{sf} y PC_{cf} .
				n_{nr} = normas y estándares que se respetan estrictamente.
				n_{te} = total de normas y estándares registrados en el PL.
				n_{ta} = número de tareas con artefactos.
				n_{tt} = total de tareas que necesitan artefactos.
				n_{no} = número de tareas con artefactos del nivel operacional.
				n_{to} = total de tareas que necesitan artefactos del nivel operacional.
				n_{nd} = número de tareas con artefactos del nivel descriptivo.
				n_{td} = total de tareas que necesitan artefactos del nivel descriptivo.
				n_{np} = número de tareas con artefactos del nivel procedimental.
				n_{tp} = total de tareas que necesitan artefactos del nivel procedimental.
				n_{sd} = número de scripts descritos sistemáticamente.
				n_{ts} = total de scripts registrados en el PL.
				n_{sf} = número de software descritos sistemáticamente.
				n_{tf} = total de software registrados.
				n_{se} = número de scripts ejecutados con éxito.
				n_{fe} = número de software ejecutados con éxito.
				n_{da} = número de datos accesibles.
				n_{td} = total de datos incluidos.
				n_{dm} = número de datos manipulados adecuadamente.
Fin de tabla §5.5				

Los artefactos *opcionales* con respecto a la insignia *Reutilizable* son los mismos que se presentan en la insignia *Funcional* y se presenta en la tabla §5.4.

5.4.6 Características de calidad para la insignia Disponible

Para la insignia *Disponible* se determina la característica de calidad: *Accesible*. Para esta característica se identifican tres parámetros de calidad a los cuales se asigna un peso (valor en porcentaje) para cumplir con el 100% y otorgarle al PL la insignia *Disponible*.

A continuación, se define la característica mínima del modelo de calidad para la insignia *Disponible*.

- **Accesible.**- Según la ACM [185], la accesibilidad se aplica a los documentos en los que los artefactos asociados se han puesto a disposición de forma permanente para su recuperación [185]. La accesibilidad de los artefactos se obtienen con el registro de tres *Parámetros de calidad*: *i*) Se proporciona un enlace al repositorio de archivos público junto con un identificador único y permanente para el objeto, que seguirá la nomenclatura de un código alfanumérico que sigue la norma NISO Z39.84 [411], es decir, el registro DOI del PL para su identificación y disponibilidad. *ii*) Los artefactos creados por el experimentador relevantes para el artículo científico donde reporta el experimento se han colocado en un repositorio de archivo de acceso público para su descarga y utilidad de forma permanente. No se obliga el uso de repositorios específicos. Se aceptan repositorios de editores (como la Biblioteca digital de ACM), repositorios institucionales o repositorios comerciales abiertos, por ejemplo: Figshare, Zenodo, Dryad, etc. En todos los casos, los repositorios utilizados para archivar datos deben tener un plan declarado para permitir la accesibilidad permanente. Las *páginas web personales* no son aceptables para este propósito. *iii*) El registro de acceso confidencial de los artefactos como claves y contraseñas son detalladas para ser utilizadas en caso de ser necesario. *Disponible*.

DOI del PL. El DOI del PL será proporcionado para acceso público de los artefactos. La métrica para esta característica de calidad *Disponible* hace referencia también al álgebra de Boole [121]:

$$Q_7 = \{0,1\} \quad (5.33)$$

El valor es 0 cuando (no se registra el DOI) y 1 (si registra el DOI).

Repositorio de archivos público. Este parámetro se refiere a registrar el repositorio de archivos público. La *Métrica* hace referencia al álgebra de Boole [121] y se define de la siguiente manera:

$$Q_8 = \{0,1\} \quad (5.34)$$

El valor es 0 cuando (no se registra el repositorio de archivos público) y 1 cuando (sí se registra el repositorio de archivos público).

Acceso confidencial. Este parámetro se refiere a registrar sí o no el acceso confidencial de los artefactos como claves y contraseñas, es decir, este parámetro no podría registrarse. La *Métrica* hace referencia al álgebra de Boole [121] y se define de la siguiente manera:

$$Q_9 = \{0,1\} \quad (5.35)$$

El valor es 0 cuando (no se registra el acceso confidencial) y 1 cuando (sí se registra el acceso confidencial).

Para resumir el modelo de calidad para paquetes de laboratorio orientado al proceso acorde a la ACM con respecto a la insignia *Disponible*, se presenta la tabla §5.6 dónde se define las características de calidad.

En la presente propuesta del modelo de calidad aunque los artefactos *opcionales* de la insignia *Disponible* no tienen un peso de medida definidos en este momento, son considerados importantes y necesarios que se registren en el PL, con el fin de proporcionar información adicional y generar más artefactos a los investigadores que deseen replicar el experimento. Sin embargo, la influencia de dichos artefactos en la calidad del PL será estudiado en profundidad como trabajo futuro de la presente investigación. Los artefactos *opcionales* para la insignia *Disponible* son:

- **Artefactos Opcionales.-** Se presenta dos artefactos opcionales para la característica de calidad *Disponible*: i) Las citas bibliográficas de los autores se presentará en formato bibtex y APA, esta información se registrará en un archivo denominado "*citation*". ii) Los datos informativos como nombre, email, sitio web, afiliación de los autores del experimento se registrarán en un archivo de acceso público para ser contactados por otros investigadores denominado "*authors*" [6]. Una forma de medir los dos artefactos *opcionales* es incluir el registro de los archivos "*citation*" y "*authors*" en el PL.

Tabla 5.6: Modelo de calidad para Paquetes de Laboratorio (PL) acorde a la ACM - Artefactos Evaluados - Disponible

Artefacto Evaluado - Disponible				
Característica	Parámetro de calidad	Evidencia de medida	Peso parámetro	Métrica
Accesible	Cada artefacto tendrá un identificador único y permanente que seguirá la nomenclatura de un código alfanumérico que sigue la norma NISO Z39.84.	Registro del DOI.	{PP ₁ , PP ₂ }	Q ₇ = {0,1}
	Los artefactos estarán en un repositorio de archivo de acceso público para su descarga y utilidad de forma permanente por lo que las páginas web no son aceptadas para este propósito.	Registro de repositorio.	{PP ₁ , PP ₂ }	Q ₈ = {0,1}
	Registrar sí o no el acceso confidencial de los artefactos como claves y contraseñas son detalladas para ser utilizadas en caso sea necesario (PC _{ac}).	Registro de acceso confidencial.	{0, PP ₁ }	Q ₉ = {0,1}
Total			100%	

{0,1}= Se verifica el registro del parámetro de calidad.

Peso Parámetro = PP

Número de Parámetros que no están condicionados, es decir, son requeridos = PR

Número de Parámetros que están condicionados = PC

Número total de Parámetros de cada insignia = TP

Cálculo del Peso Parámetro con TP en {1} = PP₁ = (((PR)/(TP - PC))/PR)*100(%) = {33.4%}

Cálculo del Peso Parámetro con PC en {0} = PP₂ = (((PR)/(TP - PC))/PR)*100(%) = {50%}

Parámetro condicionado acceso confidencial= Se verifica si el experimento necesita o no registrar el acceso confidencial de los artefactos como claves y contraseñas= PC_{ac}

{PP₁, PP₂} = A cada parámetro de calidad se asigna uno de los dos pesos dependiendo si registra o no PC_{ac}.

Tabla 5.7: Características opcionales del modelo de calidad de Artefactos Disponibles

Artefacto opcional	Evidencia de medida	Métrica
Las citas bibliográficas de los autores se registrará en formato bibtex y APA.	Archivo "citation"	$Q_{10} = \{0,1\}$
Los datos informativos como nombre, email, sitio web, afiliación de los autores del experimento serán registrados en un documento de acceso público.	Archivo "authors"	$Q_{11} = \{0,1\}$

$\{0,1\}$ =Se verifica el registro del parámetro de calidad .

Archivo "citation". La *Métrica* para el archivo "citation" hace referencia al álgebra de Boole [121] y se definen de la siguiente manera:

$$Q_{10} = \{0,1\} \tag{5.36}$$

El valor es 0 cuando el (archivo "citation" no se registra) y 1 cuando el (archivo "citation" sí se registra).

Archivo "authors". La *Métrica* para el archivo "authors" hace referencia al álgebra de Boole [121] y se definen de la siguiente manera:

$$Q_{11} = \{0,1\} \tag{5.37}$$

El valor es 0 cuando el (archivo "authors" no se registra) y 1 cuando el (archivo "authors" sí se registra).

Los artefactos *opcionales* con respecto a la insignia *Disponible*, se presenta en la tabla §5.7.

5.4.7 Característica de calidad para la insignia Reproducido

Para la insignia *Reproducido* se determina una característica de calidad: *Verificable (con uso de material)* y se han definido nueve parámetros de calidad a los cuales se asigna un peso (valor en porcentaje) para cumplir con el 100% y otorgarle al PL la insignia *Reproducido*.

Cabe destacar que esta insignia se otorga a un experimento original cuando hay evidencia de que ha sido reproducido. Es decir, existe al menos una replicación externa del mismo usando el material experimental proporcionado por los autores del experimento original.

A continuación, se define la característica mínima del modelo de calidad para la insignia *Reproducido*.

- **Verificable (con uso de material).**- Según la ACM, la insignia *Reproducido* se aplica a los trabajos en los que los principales resultados del experimento han sido obtenidos con éxito por una persona o equipo que no sea el autor del experimento original utilizando, en parte, artefactos proporcionados por dicho autor [185]. Para verificar que los resultados de la reproducción han sido validados, se debe registrar ocho *Parámetros de calidad* que ayuden a cumplir con la característica de calidad *Verificable (con uso de material)*.

Artículo científico original. El *primer parámetro*, se refiere a registrar el archivo del artículo científico original revisado por pares donde reporta el experimento. Se ingresará el DOI o enlace público del archivo. La *Métrica* hace referencia al álgebra de Boole [121] y se define de la siguiente manera:

$$Q_{12} = \{0,1\} \quad (5.38)$$

El valor es 0 cuando (no se registra el artículo científico original) y 1 cuando (sí se registra el artículo científico original).

Artículo científico de la reproducción. El *segundo parámetro*, se refiere a registrar el archivo del artículo científico revisado por pares donde se reporta la reproducción. Se ingresará el DOI o enlace público del archivo. La *Métrica* hace referencia al álgebra de Boole [121] y se define de la siguiente manera:

$$Q_{13} = \{0,1\} \quad (5.39)$$

El valor es 0 cuando (no se registra el artículo científico de la reproducción) y 1 cuando (sí se registra el artículo científico de la reproducción).

Archivo "original authors". El *tercer parámetro*, se refiere a registrar los nombres y emails de los autores del experimento original o de la reproducción actual. Para lo cuál, se debe registrar el archivo "original authors". La *Métrica* hace referencia al álgebra de Boole [121] y se define de la siguiente manera:

$$Q_{14} = \{0,1\} \quad (5.40)$$

El valor es 0 cuando (no se registra el archivo "original authors") y 1 cuando (sí se registra el archivo "original authors").

Archivo "reproduction authors". El *cuarto parámetro*, se refiere a registrar los autores del artículo reproducido con información relacionada a sus nombres y emails con la finalidad de distinguir la autoría del estudio reproducido. Para lo cuál, se debe registrar el archivo "reproduction authors". La *Métrica* hace referencia al álgebra de Boole [121] y se define de la siguiente manera:

$$Q_{15} = \{0,1\} \quad (5.41)$$

El valor es 0 cuando (no se registra el archivo "reproduction authors") y 1 cuando (sí se registra el archivo "reproduction authors").

Archivo "abstract" de la reproducción. El *quinto parámetro*, se refiere a describir un resumen mediante un archivo "abstract", con información detallada de cómo se llevó a cabo la reproducción, el porqué es importante la reproducción, cuál es el propósito de los artefactos usados en la reproducción, qué ambientes tecnológicos, software y hardware se utilizó en la reproducción. En este resumen es importante incluir una discusión que indique los aspectos que hicieron fácil o difícil reproducir, lecciones aprendidas, en el caso que la reproducción fue parcial se explicará las partes que se lograron y cuáles faltan. El contenido del resumen empezará por registrar el Título del artículo. En el título se indicará si la reproducción es parcial por ejemplo: Título: Un (Parcial)? (Reproducción) de XYZ. Se nombrará a los autores originales, autores de la reproducción y se distinguirá el autor correspondiente con emails, afiliaciones, sitios web. Se escribirá un enlace web disponible públicamente del artículo original y de los documentos e informes posteriores a la reproducción. Se describirá ¿Cuál? es la insignia que va a reclamar y finalmente cuáles son las habilidades tecnológicas que debe tener el revisor [6]. La *Métrica* hace referencia al álgebra de Boole [121] y se define de la siguiente manera:

$$Q_{16} = \{0,1\} \quad (5.42)$$

El valor es 0 cuando (no se registra el archivo "abstract") y 1 cuando (sí se registra el archivo "abstract").

Artefactos con pruebas sustanciales. El *sexto parámetro*, se refiere a facilitar otros documentos o evidencias que considere pruebas sustanciales sobre la reproducción. Dónde, n_{ps} representa el número de artefactos con pruebas sustanciales sobre la reproducción, por ejemplo: resultados del análisis de datos de la reproducción y n_{tp} representa el total de artefactos para evidenciar pruebas sustanciales sobre la reproducción. La métrica se define de la siguiente manera:

$$Q_{17} = \frac{n_{ps}}{n_{tp}} \quad (5.43)$$

El rango de medición para Q_{17} oscila entre 0 (no se registra artefactos con pruebas sustanciales) y 1 (sí se registra artefactos con pruebas sustanciales), es decir, $0 \leq Q_{17} \leq 1$. Cuando el valor de Q_{17} es 1 o se acerca a este valor, significa que los artefactos con pruebas sustanciales se registran en su totalidad o en su gran mayoría.

Artefactos que evidencian el respeto a la reproducción. El *séptimo parámetro*, se refiere a evidenciar que los autores prestan el debido respeto a los demás trabajos relacionados con la reproducción, por ejemplo: registrar las citas de los experimentos originales, las citas relacionadas con las reproducciones previas. Dónde, n_{ar} representa el número de artefactos que evidencian el respeto a trabajos relacionados con la reproducción y n_{tr} representa el total de artefactos para evidenciar el respeto a trabajos relacionados con la reproducción. La métrica se define de la siguiente manera:

$$Q_{18} = \frac{n_{ar}}{n_{tr}} \quad (5.44)$$

El rango de medición para Q_{18} oscila entre 0 (no se registra artefactos que evidencien el respeto a trabajos relacionados con la reproducción) y 1 (sí se registra artefactos que evidencien el respeto a trabajos relacionados con la reproducción), es decir, $0 \leq Q_{18} \leq 1$. Cuando el valor de Q_{18} es 1 o se acerca a este valor, significa que los artefactos que evidencian el respeto a trabajos relacionados con la reproducción se registran en su totalidad o en su gran mayoría.

Resumen no crítico hacia otros estudios. El *octavo parámetro*, se refiere a evidenciar que el resumen de la reproducción no es necesariamente crítico hacia

otros estudios relacionados en la comunidad de investigación. La *Métrica* hace referencia al álgebra de Boole [121] y se define de la siguiente manera:

$$Q_{19} = \{0,1\} \quad (5.45)$$

El valor es 0 cuando (no se evidencia que el resumen de la reproducción no es necesariamente crítico hacia otros estudios relacionados) y 1 cuando (sí se evidencia que el resumen de la reproducción no es necesariamente crítico hacia otros estudios relacionados).

Artefactos con un margen de tolerancia de los resultados de la reproducción. El *noveno parámetro*, se refiere a que los artefactos evidencien un margen y/o tolerancia de los resultados del estudio original para reclamar la insignia *Reproducido*, es decir, que coinciden en parte los resultados del experimento original con los resultados de la reproducción. Dónde, n_{ai} representa el número de artefactos que evidencian un margen y/o tolerancia de los resultados del estudio original para reclamar la insignia *Reproducido* y n_{ti} representa el total de artefactos para evidenciar un margen y/o tolerancia de los resultados del estudio original para reclamar la insignia *Reproducido*. La métrica se define de la siguiente manera:

$$Q_{20} = \frac{n_{ai}}{n_{ti}} \quad (5.46)$$

El rango de medición para Q_{20} oscila entre 0 (no se registra artefactos que evidencien un margen y/o tolerancia de los resultados del estudio original para reclamar la insignia *Reproducido*) y 1 (sí se registra artefactos que evidencien un margen y/o tolerancia de los resultados del estudio original para reclamar la insignia *Reproducido*), es decir, $0 \leq Q_{20} \leq 1$. Cuando el valor de Q_{20} es 1 o se acerca a este valor, significa que los artefactos que evidencian un margen y/o tolerancia de los resultados del estudio original para reclamar la insignia *Reproducido* se registran en su totalidad o en su gran mayoría.

Para resumir el modelo de calidad para paquetes de laboratorio orientado al proceso acorde a la ACM con respecto a la insignia *Reproducido*, se presenta la tabla §5.8 dónde se define la característica de calidad.

Tabla 5.8: Modelo de calidad para Paquetes de Laboratorio (PL) acorde a la ACM - Resultados Validados - Reproducido

Comienzo de tabla §5.8

Resultados Validados - Reproducido



Característica	Parámetro de calidad	Evidencia de medida	Peso parámetro	Métrica
Verificable (con uso de material)	Registrar el artículo científico original revisado por pares donde se reporta el experimento. Ingresar el DOI o enlace público del archivo.	Artículo científico original.	$\{PP_1\}$	$Q_{12} = \{0,1\}$
	Registrar el artículo científico revisado por pares donde se reporta la reproducción. Ingresar el DOI o enlace público del archivo.	Artículo científico de la reproducción.	$\{PP_1\}$	$Q_{13} = \{0,1\}$
	Registrar nombres y emails de los autores originales del estudio.	Archivo "original authors".	$\{PP_1\}$	$Q_{14} = \{0,1\}$
	Registrar nombres y emails de los autores de la reproducción actual con la finalidad de distinguir la autoría del estudio reproducido.	Archivo "reproduction authors".	$\{PP_1\}$	$Q_{15} = \{0,1\}$
	Describir un resumen con información detallada de cómo se llevó a cabo la reproducción, empezando por el título en el cuál se indique si fue parcial o completa la reproducción, la importancia y relevancia de la reproducción, describir los datos informativos de los autores, se facilita los enlaces web del artículo original y reproducido, se escribirá el propósito de los artefactos usados en la reproducción y las amenazas a la validez.	Archivo "abstract".	$\{PP_1\}$	$Q_{16} = \{0,1\}$
	Facilitar otros documentos o evidencias que considere pruebas sustanciales sobre la reproducción.	Artefactos con pruebas sustanciales.	$\{PP_1\}$	$Q_{17} = \frac{n_{ps}}{n_{tp}}$ $0 \leq Q_{17} \leq 1$

Continuación de la Tabla §5.8				
Característica	Parámetro de calidad	Evidencia de medida	Peso parámetro	Métrica
	Revisar el resumen y documentos necesarios para evidenciar que los autores prestan el debido respeto a los demás trabajos relacionados con la reproducción.	Artefactos que evidencien el respeto a trabajos relacionados con la reproducción.	{PP ₁ }	$Q_{18} = \frac{n_{ar}}{n_{tr}}$ $0 \leq Q_{18} \leq 1$
Verificable (con uso de material)	Revisar el resumen de la reproducción para evidenciar que no es necesariamente crítico hacia otros en la comunidad de investigación.	El resumen de la reproducción no es crítico hacia otros estudios relacionados.	{PP ₁ }	Q ₁₉ = {0,1}
	Revisar el resumen y documentos necesarios para evidenciar que basta con que los resultados están dentro de un margen/tolerancia y se desvían ligeramente de los resultados del estudio original para reclamar la insignia Reproducido.	Artefactos que evidencien un margen y/o tolerancia de los resultados del estudio original para reclamar la insignia Reproducido.	{PP ₁ }	$Q_{20} = \frac{n_{ai}}{n_{ti}}$ $0 \leq Q_{20} \leq 1$
Total			100%	

{0,1} =Se verifica el registro del parámetro de calidad .
 Peso Parámetro = PP
 Número de Parámetros que no están condicionados, es decir, son requeridos = PR
 Número de Parámetros que están condicionados = PC
 Número total de Parámetros de cada insignia = TP
 Cálculo del Peso Parámetro con TP en {1} = $PP_1 = (((PR)/(TP - PC))/PR)*100(\%) = \{11.2\%$
 n_{ps} = artefactos con pruebas sustanciales sobre la reproducción.
 n_{tp} = total de artefactos para evidenciar pruebas sustanciales sobre la reproducción.
 n_{ar} = artefactos que evidencien el respeto a trabajos relacionados con la reproducción.
 n_{tr} = total de artefactos para evidenciar el respeto a trabajos relacionados con la reproducción.
 n_{ai} = artefactos que evidencien los resultados para reclamar la insignia Reproducido.
 n_{ti} = total de artefactos para reclamar la insignia de Reproducido.

Fin de tabla §5.8

En la presente propuesta del modelo de calidad aunque los artefactos *opcionales* de la insignia *Reproducido* no tienen un peso de medida definidos en este momento, son considerados importantes y necesarios que se registren en el PL, con el fin de proporcionar información adicional y generar más artefactos a los investigadores que deseen reproducir el experimento. Sin embargo, la influencia de dichos artefactos en la calidad del PL será estudiado en profundidad como trabajo futuro de la presente investigación. Los artefactos *opcionales* para la insignia *Reproducido* son:

- **Artefactos Opcionales.**- Se presenta tres artefactos opcionales para la insignia *Reproducido*: *i)* Los autores de la reproducción presentarán una reflexión crítica de las dificultades, mejoras, limitaciones, lecciones aprendidas durante la reproducción. *ii)* El autor presentará una justificación del porqué la insignia *Reproducido* debe ser otorgada. *iii)* Se describirá de manera comprensible cómo se ha conseguido la Reproducción, escribiendo una narrativa de cómo fueron sucediendo los acontecimientos durante la Reproducción [6].

Una forma de medir los tres artefactos *opcionales* es incluir en el PL: el registro del archivo de reflexión crítica, el archivo de justificación para otorgarles la insignia *Reproducido* y el archivo de narrativa de sucesos de la reproducción, respectivamente.

Archivo de reflexión crítica. La *Métrica* para el archivo de reflexión crítica hace referencia al álgebra de Boole [121] y se define de la siguiente manera:

$$Q_{21} = \{0,1\} \quad (5.47)$$

El valor es 0 cuando el (archivo de reflexión crítica no se registra) y 1 cuando el (archivo de reflexión crítica sí se registra).

Archivo de justificación. La *Métrica* para el archivo de justificación para otorgarles la insignia de *Reproducido* hace referencia al álgebra de Boole [121] y se define de la siguiente manera:

$$Q_{22} = \{0,1\} \quad (5.48)$$

El valor es 0 cuando el (archivo de justificación para otorgarles la insignia de *Reproducido* no se registra) y 1 cuando el (archivo de justificación para otorgarles la insignia de *Reproducido* sí se registra).

Archivo de narrativa de sucesos de la reproducción. La *Métrica* para el archivo de narrativa de sucesos de la reproducción hace referencia al álgebra de Boole [121] y se define de la siguiente manera:

$$Q_{23} = \{0,1\} \quad (5.49)$$

El valor es 0 cuando el (archivo de narrativa de sucesos de la reproducción no se registra) y 1 cuando el (archivo de narrativa de sucesos de la reproducción sí se registra).

Los artefactos opcionales con respecto a la insignia *Reproducido*, se presentan en la tabla §5.9.

Tabla 5.9: Características opcionales del modelo de calidad de los Resultados Validados - Reproducido

Artefacto opcional	Evidencia de medida	Métrica
En el caso que los autores no son los del trabajo original, se presentará una reflexión crítica de las dificultades, mejoras, limitaciones, lecciones aprendidas durante la reproducción.	Archivo de reflexión crítica.	$Q_{21} = \{0,1\}$
El autor presentará una justificación y documentos adecuados que demuestre que merecen la insignia Reproducido.	Archivo de justificación que merece la insignia.	$Q_{22} = \{0,1\}$
Se describirá de manera comprensible cómo se ha conseguido la Reproducción, escribiendo una narrativa de cómo fueron sucediendo los acontecimientos durante la Reproducción.	Archivo de narrativa de sucesos de la reproducción.	$Q_{23} = \{0,1\}$

$\{0,1\}$ =Se verifica el registro del parámetro de calidad .

5.4.8 Característica de calidad para la insignia Replicado

Para la insignia *Replicado* se determina una característica de calidad: *Verificable (sin uso de material)* y se han definido nueve parámetros de calidad a los cuales se asigna un peso (valor en porcentaje) para cumplir con el 100% y otorgarle al PL la insignia *Replicado*.

Cabe destacar que esta insignia se otorga a un experimento original cuando hay evidencia de que ha sido replicado. Es decir, existe al menos una replicación externa del mismo usado material experimental independiente del usado en el experimento original.

A continuación, se define la característica mínima del modelo de calidad para la insignia *Replicado*.

- **Verificable (sin uso de material).**- Según la ACM, la insignia *Replicado* se aplica a los trabajos en los que los principales resultados del experimento han sido obtenidos con éxito por una persona o equipo que no sea el autor del experimento original sin hacer uso de los artefactos proporcionados por dicho autor [185]. Para verificar que los resultados de la replicación han sido validados, se debe registrar

ocho *Parámetros de calidad* que ayuden a cumplir con la característica de calidad *Verificable (sin uso de material)*.

Artículo científico original. El *primer parámetro*, se refiere a registrar el archivo del artículo científico original revisado por pares donde reporta el experimento. Se ingresará el DOI o enlace público del archivo. La *Métrica* hace referencia al álgebra de Boole [121] y se define de la siguiente manera:

$$Q_{24} = \{0,1\} \quad (5.50)$$

El valor es 0 cuando (no se registra el artículo científico original) y 1 cuando (sí se registra el artículo científico original).

Artículo científico de la replicación. El *segundo parámetro*, se refiere a registrar el archivo del artículo científico revisado por pares donde se reporta la replicación. Se ingresará el DOI o enlace público del archivo. La *Métrica* hace referencia al álgebra de Boole [121] y se define de la siguiente manera:

$$Q_{25} = \{0,1\} \quad (5.51)$$

El valor es 0 cuando (no se registra el artículo científico de la replicación) y 1 cuando (sí se registra el artículo científico de la replicación).

Archivo "original authors". El *tercer parámetro*, se refiere a registrar los autores del experimento original o de la replicación actual. Para lo cuál, se debe registrar el archivo "original authors". La *Métrica* hace referencia al álgebra de Boole [121] y se define de la siguiente manera:

$$Q_{26} = \{0,1\} \quad (5.52)$$

El valor es 0 cuando (no se registra el archivo "original authors") y 1 cuando (sí se registra el archivo "original authors").

Archivo "replication authors". El *cuarto parámetro*, se refiere a registrar los autores del artículo replicado con información relacionada a sus nombres y emails con la finalidad de distinguir la autoría del estudio replicado. Para

lo cuál, se debe registrar el archivo "replication authors". La *Métrica* hace referencia al álgebra de Boole [121] y se define de la siguiente manera:

$$Q_{27} = \{0,1\} \quad (5.53)$$

El valor es 0 cuando (no se registra el archivo "replication authors") y 1 cuando (sí se registra el archivo "replication authors").

Archivo "abstract" de la replicación. El *quinto parámetro*, se refiere a describir un resumen mediante un archivo "abstract", con información detallada de cómo se llevó a cabo la replicación, el porqué es importante la replicación, cuál es el propósito de los artefactos usados en la replicación, qué ambientes tecnológicos, software y hardware se utilizó en la replicación.

En este resumen es importante incluir una discusión que indique los aspectos que hicieron fácil o difícil replicar, lecciones aprendidas, en el caso que la replicación fue parcial se explicará las partes que se lograron y cuáles faltan. El contenido del resumen empezará por registrar el Título del artículo. En el título se indicará si la replicación es parcial por ejemplo: Título: Un (Parcial)? (Replicación) de XYZ. Se nombrará a los autores originales, autores de la replicación y se distinguirá el autor correspondiente con emails, afiliaciones, sitios web. Se escribirá un enlace web disponible públicamente del artículo original y de los documentos e informes posteriores a la replicación. Se describirá ¿Cuál? es la insignia que va a reclamar y finalmente cuáles son las habilidades tecnológicas que debe tener el revisor [6]. La *Métrica* hace referencia al álgebra de Boole [121] y se define de la siguiente manera:

$$Q_{28} = \{0,1\} \quad (5.54)$$

El valor es 0 cuando (no se registra el archivo "abstract" de la replicación) y 1 cuando (sí se registra el archivo "abstract" de la replicación).

Artefactos con pruebas sustanciales. El *sexto parámetro*, se refiere a facilitar otros documentos o evidencias que considere pruebas sustanciales sobre la replicación, por ejemplo: resultados del análisis de datos de la replicación. Dónde, n_{pr} representa el número de artefactos con pruebas sustanciales sobre la replicación y n_{tu} representa el total de artefactos para evidenciar prue-

bas sustanciales sobre la replicación. La métrica se define de la siguiente manera:

$$Q_{29} = \frac{n_{pr}}{n_{tu}} \quad (5.55)$$

El rango de medición para Q_{29} oscila entre 0 (no se registra artefactos con pruebas sustanciales) y 1 (sí se registra artefactos con pruebas sustanciales), es decir, $0 \leq Q_{29} \leq 1$. Cuando el valor de Q_{29} es 1 o se acerca a este valor, significa que los artefactos con pruebas sustanciales se registran en su totalidad o en su gran mayoría.

Artefactos que evidencian el respeto a la replicación. El *séptimo parámetro*, se refiere a evidenciar que los autores prestan el debido respeto a los demás trabajos relacionados con la replicación, por ejemplo: registrar las citas de los experimentos originales, las citas relacionadas con las replications previas. Dónde, n_{ra} representa el número de artefactos que evidencian el respeto a trabajos relacionados con la replicación y n_{te} representa el total de artefactos para evidenciar el respeto a trabajos relacionados con la replicación. La métrica se define de la siguiente manera:

$$Q_{30} = \frac{n_{ra}}{n_{to}} \quad (5.56)$$

El rango de medición para Q_{30} oscila entre 0 (no se registra artefactos que evidencien el respeto a trabajos relacionados con la replicación) y 1 (sí se registra artefactos que evidencien el respeto a trabajos relacionados con la replicación), es decir, $0 \leq Q_{30} \leq 1$. Cuando el valor de Q_{30} es 1 o se acerca a este valor, significa que los artefactos que evidencian el respeto a trabajos relacionados con la replicación se registran en su totalidad o en su gran mayoría.

Resumen no crítico hacia otros estudios. El *octavo parámetro*, se refiere a evidenciar que el resumen de la replicación no es necesariamente crítico hacia otros estudios relacionados en la comunidad de investigación. La *Métrica* hace referencia al álgebra de Boole [121] y se define de la siguiente manera:

$$Q_{31} = \{0,1\} \quad (5.57)$$

El valor es 0 cuando (no se evidencia que el resumen de la replicación no es necesariamente crítico hacia otros estudios relacionados) y 1 cuando (sí se evidencia que el resumen de la replicación no es necesariamente crítico hacia otros estudios relacionados).

Artefactos con un margen de tolerancia de los resultados de la replicación. El *décimo parámetro*, se refiere a que los artefactos evidencien un margen y/o tolerancia de los resultados del estudio original para reclamar la insignia *Replicado*, es decir, que coinciden en parte los resultados del experimento original con los resultados de la replicación. Dónde, n_{ia} representa el número de artefactos que evidencian un margen y/o tolerancia de los resultados del estudio original para reclamar la insignia *Replicado* y n_{tl} representa el total de artefactos para evidenciar un margen y/o tolerancia de los resultados del estudio original para reclamar la insignia *Replicado*. La métrica se define de la siguiente manera:

$$Q_{32} = \frac{n_{ia}}{n_{tl}} \quad (5.58)$$

El rango de medición para Q_{32} oscila entre 0 (no se registra artefactos que evidencien un margen y/o tolerancia de los resultados del estudio original para reclamar la insignia *Replicado*) y 1 (sí se registra artefactos que evidencien un margen y/o tolerancia de los resultados del estudio original para reclamar la insignia *Replicado*), es decir, $0 \leq Q_{32} \leq 1$. Cuando el valor de Q_{32} es 1 o se acerca a este valor, significa que los artefactos que evidencian un margen y/o tolerancia de los resultados del estudio original para reclamar la insignia *Replicado* se registran en su totalidad o en su gran mayoría.

Para resumir el modelo de calidad para paquetes de laboratorio orientado al proceso acorde a la ACM con respecto a la insignia *Replicado*, se presenta la tabla §5.10 dónde se define las características de calidad.

Tabla 5.10: Modelo de calidad para Paquetes de Laboratorio (PL) acorde a la ACM - Resultados Validados - Replicado

Comienzo de tabla §5.10					
Resultados Validados - Replicado					
					
Característica	Parámetro de calidad	Evidencia de medida		Peso parámetro	Métrica
Verificable (sin uso de material)	Registrar el artículo científico original revisado por pares donde se reporta el experimento. Ingresar el DOI o enlace público del archivo.	Artículo científico original.	científico	$\{PP_1\}$	$Q_{24} = \{0,1\}$
	Registrar el artículo científico revisado por pares donde se reporta la replicación. Ingresar el DOI o enlace público del archivo.	Artículo científico de la replicación.		$\{PP_1\}$	$Q_{25} = \{0,1\}$
	Registrar nombres y emails de los autores originales del estudio.	Archivo "original author".		$\{PP_1\}$	$Q_{26} = \{0,1\}$
	Registrar nombres y emails de los autores de la replicación actual con la finalidad de distinguir la autoría del estudio replicado.	Archivo "replication authors".		$\{PP_1\}$	$Q_{27} = \{0,1\}$
	Describir un resumen con información detallada de cómo se llevó a cabo la replicación, empezando por el título en el cuál se indique si fue parcial o completa la replicación, la importancia y relevancia de la replicación, describir los datos informativos de los autores, se facilita los enlaces web del artículo original y replicado, se escribirá el propósito de los artefactos usados en la replicación y las amenazas a la validez.	Archivo "abstract".		$\{PP_1\}$	$Q_{28} = \{0,1\}$
	Se facilitará otros documentos o evidencias que considere pruebas sustanciales sobre la replicación.	Artefactos con pruebas sustanciales.		$\{PP_1\}$	$Q_{29} = \frac{n_{pr}}{n_{tu}}$ $0 \leq Q_{29} \leq 1$

Continuación de la Tabla §5.10				
Característica	Parámetro de calidad	Evidencia de medida	Peso parámetro	Métrica
Verificable (sin uso de material)	Revisar el resumen y demás documentación necesaria de la replicación para evidenciar que los autores prestan el debido respeto a los demás trabajos relacionados con la replicación.	Artefactos que evidencien el respeto a trabajos relacionados con la replicación.	{PP ₁ }	$Q_{30} = \frac{n_{ra}}{n_{to}}$ $0 \leq Q_{30} \leq 1$
	Revisar el resumen de la replicación para evidenciar que no es necesariamente crítico hacia otros en la comunidad de investigación.	El resumen de la reproducción no es crítico hacia otros estudios relacionados.	{PP ₁ }	$Q_{31} = \{0,1\}$
	Revisar el resumen y documentos necesarios para evidenciar que basta con que los resultados están dentro de un margen/tolerancia y se desvían ligeramente de los resultados del estudio original para reclamar la insignia Replicado.	Artefactos que evidencien un margen y/o tolerancia de los resultados del estudio original para reclamar la insignia Replicado.	{PP ₁ }	$Q_{32} = \frac{n_{ia}}{n_{il}}$ $0 \leq Q_{32} \leq 1$
Total			100%	

{0,1} =Se verifica el registro del parámetro de calidad .

Peso Parámetro = PP

Número de Parámetros que no están condicionados, es decir, son requeridos = PR

Número de Parámetros que están condicionados = PC

Número total de Parámetros de cada insignia = TP

Cálculo del Peso Parámetro con TP en {1} = $PP_1 = (((PR)/(TP - PC))/PR)*100(\%) = \{11.2\%$

n_{pr} = artefactos con pruebas sustanciales sobre replicación.

n_{tu} = total de artefactos para evidenciar pruebas sustanciales sobre replicación.

n_{ra} = artefactos que evidencien el respeto a trabajos relacionados con la replicación.

n_{to} = total de artefactos para evidenciar el respeto a trabajos relacionados con la replicación.

n_{ia} = artefactos que evidencien los resultados para reclamar la insignia Replicado.

n_{il} = total de artefactos para reclamar la insignia de Replicado.

Fin de tabla §5.10

En la presente propuesta del modelo de calidad aunque los artefactos *opcionales* de la insignia *Replicado* no tienen un peso de medida definidos en este momento, son considerados importantes y necesarios que se registren en el PL, con el fin de proporcionar información adicional y generar más artefactos a los investigadores que deseen replicar el experimento. Sin embargo, la influencia de dichos artefactos en la calidad del PL será estudiado en profundidad como trabajo futuro de la presente investigación. Los artefactos *opcionales* para la insignia *Replicado* son:

- **Artefactos Opcionales.**- Se presenta tres artefactos *opcionales* para la característica de calidad *Verificable (sin uso de material)*: *i)* Los autores de la replicación presentarán una reflexión crítica de las dificultades, mejoras, limitaciones, lecciones aprendidas durante la replicación. *ii)* El autor presentará una justificación del porqué la insignia *Replicado* debe ser otorgada. *iii)* Se describirá de manera comprensible cómo se ha conseguido la Replicación, escribiendo una narrativa de cómo fueron sucediendo los acontecimientos durante la Replicación [6]. Una forma de medir los tres artefactos *opcionales* es incluir en el PL: el registro del archivo de reflexión crítica, el archivo de justificación para otorgarles la insignia de *Replicado* y el archivo de narrativa de sucesos de la replicación, respectivamente.

Archivo de reflexión crítica. La *Métrica* para el archivo de reflexión crítica hace referencia al álgebra de Boole [121] y se define de la siguiente manera:

$$Q_{33} = \{0,1\} \quad (5.59)$$

El valor es 0 cuando el (archivo de reflexión crítica no se registra) y 1 cuando el (archivo de reflexión crítica sí se registra).

Archivo de justificación. La *Métrica* para el archivo de justificación para otorgarles la insignia de *Replicado* hace referencia al álgebra de Boole [121] y se define de la siguiente manera:

$$Q_{34} = \{0,1\} \quad (5.60)$$

El valor es 0 cuando el (archivo de justificación para otorgarle la insignia de *Replicado* no se registra) y 1 cuando el (archivo de justificación para otorgarle la insignia de *Replicado* sí se registra).

Archivo de narrativa de sucesos de la replicación.

$$Q_{35} = \{0,1\} \quad (5.61)$$

El valor es 0 cuando el (archivo de narrativa de sucesos de la replicación no se registra) y 1 cuando el (archivo de narrativa de sucesos de la replicación sí se registra).

Los artefactos *opcionales* con respecto a la insignia *Replicado*, se presenta en la tabla §5.11.

Tabla 5.11: Características opcionales del modelo de calidad de los Resultados Validados - Replicado

Artefacto opcional	Evidencia de medida	Métrica
En el caso que los autores no son los del trabajo original, se presentará una reflexión crítica de las dificultades, mejoras, limitaciones, lecciones aprendidas durante la replicación.	Archivo de reflexión crítica.	$Q_{33} = \{0,1\}$
El autor presentará una justificación y documentos adecuados que demuestre que merecen la insignia Replicado.	Archivo de justificación que merece la insignia.	$Q_{34} = \{0,1\}$
Se describirá de manera comprensible cómo se ha conseguido la Replicación, escribiendo una narrativa de cómo fueron sucediendo los acontecimientos durante la Replicación.	Archivo de narrativa de sucesos de la replicación.	$Q_{35} = \{0,1\}$

$\{0,1\}$ =Se verifica el registro del parámetro de calidad .

Es importante, explicar que, para obtener un PL de calidad y alcanzar las insignias de la ACM se debe registrar los artefactos de las tareas experimentales, los artefactos que no necesariamente pertenecen a las tareas y cumplir con las características del modelo de calidad propuesto en el presente trabajo de tesis.

5.5 ESTRUCTURA BÁSICA DE UN PL PARA TENER UN NIVEL DE CALIDAD.

Tras el estudio de las guías de presentación y revisión de artefactos de las cuatro conferencias (ICSE [186], ESEC/FSE [141], RE [339], ASE [35]), las Políticas de Insignia de la ACM [185], y los niveles de madurez definidos por Solari *et al.* [393], proponemos una estructura del contenido de los PL para experimentos orientados al ser humano en IS que debe presentar un PL para tener un nivel de calidad adecuado (Tabla §5.12 y Tabla §5.13). La estructura ayudará al experimentador a desarrollar un PL organizado y estructurado con un nivel aceptable de integridad para otorgar las insignias de la ACM. Además, la Tabla §5.12 muestra la estructura del paquete de laboratorio para los artefactos evaluados y disponibles relacionados con las insignias de la ACM, como los *funcionales*, los *reutilizables* y los *disponibles*. La estructura comprende una lista de artefactos clasificados en seis secciones, es decir, Artículo, Material experimental, Datasets, Sistema, Scripts y Diccionario de PL. Además, la estructura presenta una lista

Tabla 5.12: Estructura de PL para artefactos evaluados y disponibles relacionado con las insignias de la ACM.

Sección	Contenido	Insignias ACM			Niveles de madurez por Solari.
		Funcional 	Reutilizable 	Disponible 	
Artículo	Artículo científico	✓	✓	-	Artículo
	Descripción del diseño	✓	✓	-	Descriptivo
	Descripción del contexto	✓	✓	-	Descriptivo
	Registro de autor	-	✓	-	Descriptivo
Material experimental	Instrucciones de participante	✓	✓	-	Operacional
	Encuesta demográfica de participantes	✓	✓	-	Operacional
	Artefactos de formación	✓	✓	-	Operacional
	Instrumentos de medida	✓	✓	-	Operacional
	Modelos conceptuales	✓	✓	-	Operacional
Datasets	Raw data	✓	✓	-	Operacional
	Curated data	✓	✓	-	Operacional
	Datos accesibles	✓	✓	-	Operacional
	Datos manipulables	✓	✓	-	Operacional
Sistema	Descripción del Software	✓	✓	-	Operacional
	Ejecución exitosa del Software	✓	✓	-	Operacional
	Requisitos técnicos del Software	-	✓	-	Operacional
	Requisitos técnicos del Hardware	-	✓	-	Operacional
Scripts	Descripción de Scripts	✓	✓	-	Descriptivo
	Análisis de scripts	✓	✓	-	Operacional
	Ejecución exitosa de scripts	✓	✓	-	Operacional
Diccionario de PL	Inventario de artefactos	✓	✓	-	Descriptivo
	Descripción de artefactos	✓	✓	-	Descriptivo
	Archivo comprimido de PL	-	✓	-	Descriptivo
	Entornos virtuales	-	✓	-	Descriptivo
	Derechos de distribución de PL	-	✓	-	Descriptivo
	Guías de instalación de artefactos	-	✓	-	Procedimental
	Licencia disponible para los artefactos	-	✓	-	Descriptivo
	Tipo de PL (Simple y/o instalación)	-	✓	-	Descriptivo
	Justificación para reclamar el distintivo	-	✓	-	Descriptivo
	Configuración e instalación de artefactos	-	✓	-	Descriptivo
	DOI	-	-	✓	Descriptivo
	Repositorio público del PL	-	-	✓	Descriptivo
	Acceso confidencial al PL	-	-	✓	Descriptivo
	Guía de instrucciones para la descarga de artefactos	✓	✓	-	Procedimental
	Guía de instrucciones para la ejecución de artefactos	✓	✓	-	Procedimental
	Toma de decisiones para realizar el experimento	✓	✓	-	Toma de decisiones
Total de artefactos		22	33	3	

de comprobación de los contenidos del PL cuyo cumplimiento contribuirá a la conce-

Tabla 5.13: Estructura de PL para resultados validados relacionado a las insignias de la ACM.

Sección	Contenido	Insignias ACM		Niveles de madurez por Solari.
		Reproducido 	Replicado 	
Artículo	Artículo científico de la reproducción	✓	-	Artículo
	Autores de la reproducción	✓	-	Descriptivo
	Autores originales del experimento	✓	✓	Descriptivo
	Artículo científico de la replicación	-	✓	Artículo
	Autores de la replicación	-	✓	Descriptivo
Diccionario de PL	Resumen de cómo se llevó a cabo la reproducción	✓	-	Descriptivo
	Evidencias que considere pruebas sustanciales sobre la reproducción	✓	-	Descriptivo
	Evidencias sobre el trabajo relacionado con la reproducción	✓	-	Descriptivo
	Evidencia de que el resumen no es crítico con otros estudios relacionados con la reproducción	✓	-	Descriptivo
	Evidencia del margen de tolerancia en los resultados de la reproducción	✓	-	Descriptivo
	Resumen de cómo se llevó a cabo la replicación	-	✓	Descriptivo
	Evidencias que considere pruebas sustanciales sobre la replicación	-	✓	Descriptivo
	Evidencias sobre el trabajo relacionado con la replicación	-	✓	Descriptivo
	Evidencia de que el resumen no es crítico con otros estudios relacionados con la replicación	-	✓	Descriptivo
	Evidencia del margen de tolerancia en los resultados de la replicación	-	✓	Descriptivo
Total de artefactos		8	8	

sión de la insignia correspondiente. El Tabla §5.13 muestra una estructura de PL para los resultados validados relacionados con las insignias ACM, es decir, *reproducidos* y *replicados*. En consecuencia, estos contenidos tienen sentido cuando cualquier replicación (de un experimento original) se ha llevado a cabo. La estructura comprende una lista de artefactos de replicación clasificados en dos secciones (Artículo y Diccionario de PL) que tienen el contenido de PL. Asimismo, la Tabla §5.12, pero relacionada con las insignias *reproducidos* y *replicados*, y la Tabla §5.13 muestran una lista de verificación de los contenidos del PL cuyo cumplimiento dará lugar a la concesión de estas insignias. Por último, además de la información sobre las insignias de la ACM, las Tablas §5.12 y §5.13 muestran, para cada elemento del PL estructurado, qué nivel de madurez permite alcanzar, según [394].

5.6 RESUMEN

En este capítulo, se ha presentado una de las principales aportaciones del trabajo recogido en esta memoria para fomentar las replicaciones y contribuir a la elaboración de PL de calidad en experimentos controlados con humanos en el área de IS. El enfoque propuesto consta de cuatro partes: *i)* una especificación del proceso experimental de forma integrada con el Paquete de Laboratorio, *ii)* un procedimiento que describe paso a paso las actividades del proceso de elaboración del Paquete de Laboratorio, *iii)* un modelo de calidad para Paquete de Laboratorio acorde a la ACM y *iv)* una estructura básica que un PL debe presentar para tener un adecuado nivel de calidad.

La *especificación* del proceso experimental de forma integrada con el PL ha permitido construir una estructura de información específica y detallada acerca del contenido del paquete orientado al proceso. Además, evita redundancias, ambigüedades y la falta de información relevante. Se ha logrado especificar tres tipos de tareas que se llevan a cabo en un experimento: *i)* de preparación, *ii)* experimentales y *iii)* de recolección y análisis. Se ha logrado especificar tres grupos de artefactos: *i)* aquellos que están relacionados con las tareas del experimento, *ii)* artefactos que no están relacionados con las tareas del experimento y *iii)* artefactos que cumplen las políticas de evaluación para alcanzar las insignias de la ACM.

Un *procedimiento* para elaborar un PL acorde a la ACM, que permita establecer paso a paso las actividades de manera secuencial que debe seguir el experimentador para construir un paquete de calidad y que alcance a obtener las insignias de la ACM.

El *modelo de calidad* propuesto para el grupo de artefactos que cumplen las políticas de evaluación para alcanzar las insignias de la ACM, ha permitido establecer características de calidad para el contenido del PL y establecer el camino para alcanzar las insignias de la ACM.

De manera subsecuente al modelo de calidad, se propone una *estructura*, que sirva básicamente al experimentador empezar a elaborar el contenido del PL con un nivel aceptable de completitud.

AUTOMATIZACIÓN DEL PROCESO MEDIANTE BADGE GO

La primera regla de cualquier tecnología que se utiliza en una empresa es que la automatización aplicada a una operación eficiente magnificará la eficiencia.

*Bill Gates (1955 –),
Empresario, informático y filántropo estadounidense*

En este capítulo se presenta un complemento a la contribución principal de este trabajo de tesis, cuyo objetivo principal es automatizar el procedimiento de elaborar un PL orientado al proceso que posea un nivel de calidad adecuado según las recomendaciones existentes en la comunidad de la IS [185, 295] mediante una herramienta (badgeGo).

En la sección de introducción §6.1, se destacan las cualidades de la automatización mediante la herramienta. En la sección §6.2, se describe los requisitos generales de la automatización del procedimiento de la elaboración del PL. En la sección §6.3, se conceptualiza la información del PL orientado al proceso acorde a la ACM mediante un metamodelo. En la sección §6.4, se describe el contexto de la aplicabilidad de la propuesta. En la sección §6.5 se presenta la herramienta badgeGo para asistir la elaboración del PL orientado al proceso y la asignación de insignias de la ACM sobre la calidad de su contenido. Finalmente, la sección §6.6 resume el capítulo.

6.1 INTRODUCCIÓN

Elaborar un PL que presente un nivel de completitud acorde a las insignias de la ACM requiere conocer los artefactos mínimos que se vuelven indispensables para apoyar a la replicación de experimentos. Luego de profundizar el procedimiento que se debe llevar a cabo para la elaboración de los PL y proponer un modelo de calidad para alcanzar las insignias de la ACM se ha propuesto automatizar dicho procedimiento mediante una herramienta informática (badgeGo) que ayude al experimentador a elaborar un PL acorde a la ACM.

Cómo complemento a la propuesta de la presente memoria de tesis se desarrolló una herramienta badgeGo que permita al experimentador la asistencia automatizada para elaborar el contenido de un PL acorde a la ACM. Esta sección tiene el siguiente contenido: *i*) una serie de requisitos generales para construir la herramienta badgeGo, *ii*) un metamodelo para formalizar y clasificar la información registrada en la estructura propuesta y *iii*) una herramienta que automatice la elaboración del PL y la asignación con una alta probabilidad de obtener una insignia concreta de la ACM.

6.2 REQUISITOS

Los requisitos funcionales que se han descrito de manera explícita para el comportamiento que debe tener la herramienta badgeGo y la información que debe gestionarse han sido considerados de manera general para la construcción de badgeGo. Los requisitos están sujetos a las siguientes funcionalidades. La herramienta badgeGo permite:

- Registrar información del experimento como el objetivo, el contexto, fecha de ejecución y descripción del experimento.
- Registrar información de los experimentadores con sus diferentes roles y datos necesarios para el PL.
- Registrar los grupos experimentales y la asignación de participantes a dichos grupos.
- Registrar cada una de las tareas del experimento junto con sus artefactos. En el caso de que las tareas no tengan artefactos disponibles la herramienta sugiera la subida de artefactos mediante alertas.

- Registrar aquellos artefactos que no tengan relación con alguna tarea específica durante el proceso experimental, pero que el experimentador cree necesario que se deben subir para facilitar la replicación del experimento. Algunos artefactos pueden ser relacionados con las políticas de la ACM.
- Registrar información del PL como nombre, DOI en el caso de estar disponible, nombre del repositorio digital donde se aloja el PL, fecha de creación y actualización del PL.
- badgeGo evalúa el contenido del PL registrado en la herramienta de acuerdo a las insignias de la ACM. Indicará un porcentaje de cumplimiento de los parámetros de evaluación que pertenecen a cada insignia de la ACM.
- El experimentador ejecuta el archivo comprimido del contenido del PL de forma automática.

La lista de requisitos definidos anteriormente han sido aplicadas para determinar las funcionalidades descritas en un conjunto de entradas, comportamiento y salidas que tiene que cumplir badgeGo.

6.3 METAMODELO DE PAQUETE DE LABORATORIO ACORDE A LA ACM

Para clarificar los conceptos, las relaciones y las reglas estructurales definidas en la propuesta de la presente memoria de tesis se desarrolla un metamodelo utilizando la notación de diagrama de clases UML. Para una mejor comprensión, el metamodelo se visualiza en dos partes: *i*) para la especificación del proceso experimental de forma integrada con el Paquete de Laboratorio, detallado en la sección §5.2 y *ii*) para el modelo de calidad de Paquete de Laboratorio acorde a la ACM, detallado en la sección §5.4.

6.3.1 Metamodelo de la especificación del proceso experimental de forma integrada con el Paquete de Laboratorio

A continuación se describe el contenido del metamodelo con respecto a la especificación del proceso experimental de forma integrada con el PL. La figura §6.1 muestra el metamodelo utilizando la notación de diagrama de clases UML con respecto a la especificación del proceso experimental de forma integrada con el PL.

objetivo del experimento y se registra una breve descripción.

- Se registra información de los experimentadores por ejemplo, los datos personales como: la *identificación, nombre, apellido, email, afiliación, sitio web, género, nacionalidad*, se identifica el tipo de persona para el experimentador que puede ser un: *estudiante, graduado o profesional* y un comentario de ser necesario. Se define un *Tipo_Rol* para los experimentadores, con la finalidad de determinar las tareas y funciones asignadas durante la ejecución del experimento, por ejemplo: *Diseñador, Analista, Formador, Monitor y Medidor* [164]. Los participantes del experimento serán registrados únicamente con un identificador, para conservar la privacidad y anonimización de los datos. El experimentador asignará a los participantes en los diferentes *grupos experimentales* para que cumplan con las tareas asignadas.
- Se registra información de las tareas: *nombre* de la tarea, tipo de tarea identificadas como: de *preparación, experimentales* y de *recolección y análisis*. Se registra la fecha de *inicio*, fecha de *fin*, la *duración* de la tarea, se registra el *responsable* de la tarea y una breve descripción. Puede existir más tareas algunas pueden realizarse en paralelo.
- Se registra información de los artefactos que forman parte del PL. Algunos artefactos están relacionados con las tareas y otros no necesariamente pertenecen a dichas tareas, además aquellos artefactos que se integran al PL para alcanzar las insignias de la ACM. Se registra el *nombre* del artefacto, la clase del artefacto si es de *entrada* o *salida*, el *tamaño*, la *fecha de creación* y *última actualización* del artefacto. Además, se registra la *finalidad* que tiene el artefacto por ejemplo, si forma parte de los Scripts, Datasets, Charlas, Requisitos, Reclutamiento o Captación de sujetos, Instrucciones, Formación, Reportes, Modelos UML, Gráficos estadísticos e Insignias de la ACM. Se registra el *tipo de artefacto* donde se lista los siguientes tipos: Póster, Vídeos, Audio, Imágenes, Formularios, Encuestas, Cuestionarios, Diagramas, Software, Lecciones, Documentos, Sitios web, Publicaciones, Presentaciones, Objetos físicos, Archivos binario, Archivos de datos, Archivos de texto, Archivos de código y Archivos comprimidos. Se registra el *tipo de formato* por ejemplo, si es un documento PDF, html, md, mp4, jpg, png, pptx, csv, sav, R, wav, zip, tar.gz, rar. Se ha listado los nombres con relación al tipo de finalidad, tipo de artefacto y tipo de formato, sin embargo, se debe indicar que a futuro podría extenderse dichas listas. Finalmente, se registra el *Path* que es la dirección (URL) donde está ubicado el artefacto, el acceso a los artefactos mediante el registro de

un *usuario y contraseña* en caso de ser necesario, el *tiempo de ejecución completo y corto* del artefacto y un breve comentario.

- Se registra la información de las amenazas a la validez como el tipo de amenaza si es *Interna, Externa, De Constructo, De Conclusión*, las *limitaciones* que se presentan en el experimento y un breve comentario.
- Se registra la información del PL como el *nombre* del paquete, en caso de existir se registra el *DOI* del paquete. Se registra el tipo de paquete si es *simple* o de *instalación*, el *repositorio* de archivos público donde se encuentra el PL por ejemplo: Zenodo, Figshare, GitHub, OSF, onCloud, Amazon aws, Bitbucket, Dropbox, Google drive, Google sites, Mendeley data, Repositorios institucionales, Computer Society Digital Library. Así, mismo se registra la fecha de *creación y actualización* del paquete y una breve descripción.

6.3.2 Metamodelo del modelo de calidad acorde a la ACM

A continuación, se describe el contenido del metamodelo utilizando la notación de diagrama de clases UML con respecto al modelo de calidad acorde a la ACM. La figura §6.2 muestra el metamodelo utilizando la notación de diagrama de clases UML con respecto al modelo de calidad.

- Se registra las características de calidad para obtener la insignia de Artefactos Evaluados - Funcional: *i) Documentado, ii) Completo, iii) Coherente y iv) Ejercitable*. Cada característica esta determinada por una lista de parámetros que deben cumplirse y que se encuentran registradas con un tipo de dato lógico: *Boolean*. Para *Documentado* los parámetros son: el *i) archivo readme* y el *ii) inventario de artefactos*. Para *Completo* los parámetros son: el *i) artículo científico*, la *ii) guía de instrucciones para obtención y descarga de los artefactos* y la *iii) guía de instrucciones para ejecución de los artefactos*. Para *Coherente* el parámetro es: la *i) relevancia de los artefactos*. Para *Ejercitable* los parámetros son: el *i) registro de software de los resultados*, la *ii) ejecución de software de los resultados*, la *iii) accesibilidad de los archivos de datos incluidos*, y la *iv) manipulación de los datos incluidos*. Algunos de éstos parámetros son derivados de otros componentes del metamodelo, por ejemplo para que se cumpla el inventario de artefactos es necesario el registro completo y documentado de los artefactos que forman parte de todas las tareas experimentales y de aquellos artefactos que no necesariamente

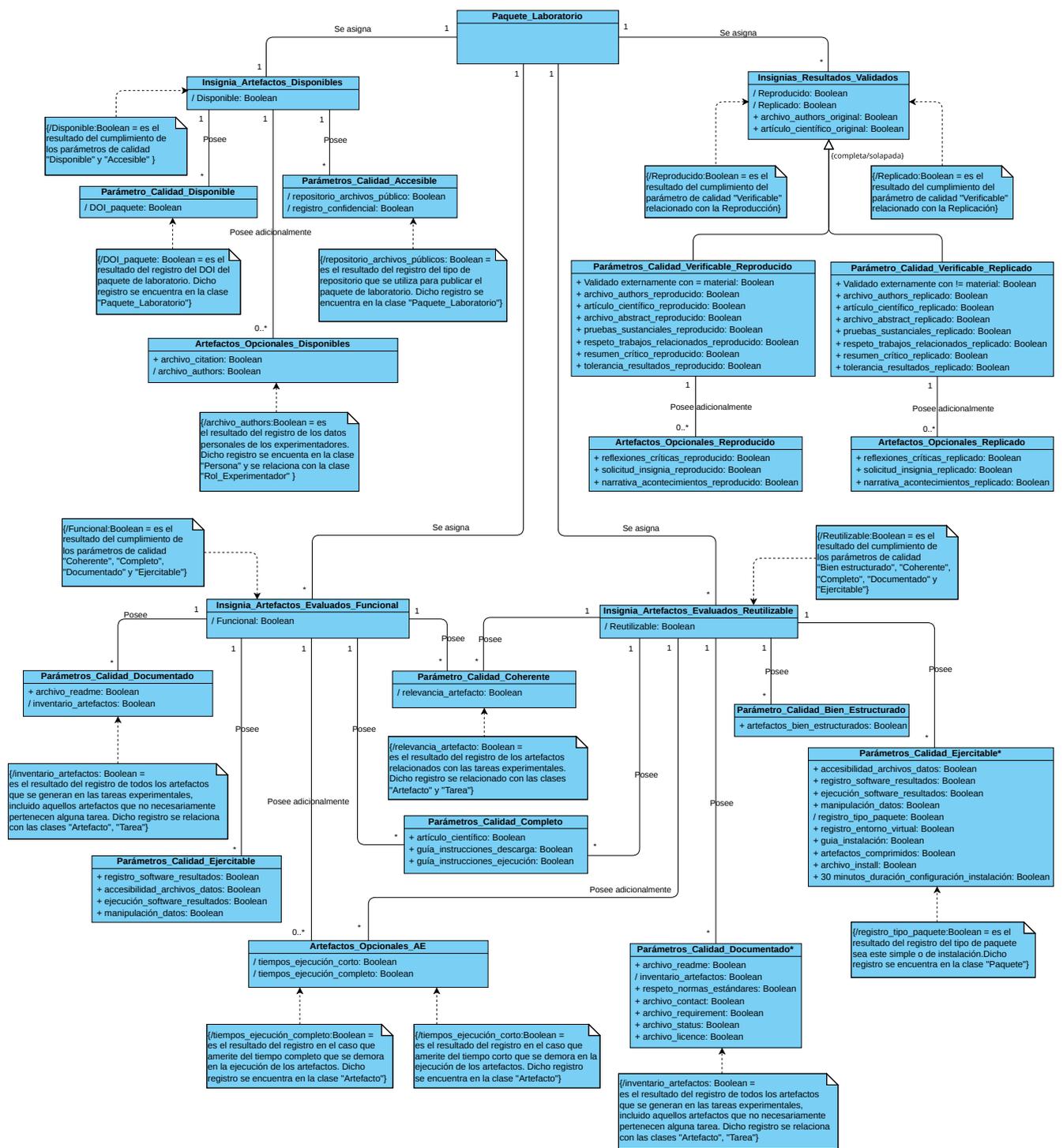


Figura 6.2: Metamodelo del modelo de calidad acorde a la ACM

forman parte de las tareas, los grupos de artefactos que forman el PI esta explicada en la sección §5.4.2. Se registran parámetros *opcionales* que a pesar de ser *opcionales* deben ser considerados que se incluyan en e PL para apoyar en el desarrollo

de un PL de calidad. Estos parámetros *opcionales* para la característica *Funcional* son: los *i*) tiempos de ejecución corto y los *ii*) tiempos de ejecución completo de los artefactos.

- Para obtener la insignia de Artefactos Evaluados - Reutilizable, se debe cumplir con todos los parámetros de la insignia Funcional al igual que los *opcionales* descritos anteriormente, y adicional los siguientes parámetros, que se encuentran registradas con un tipo de dato lógico *Boolean*. Para la característica *Bien estructurado* el parámetro es: *i*) artefactos bien estructurados. Para *Documentado* los parámetros son: el *i*) respeto de normas y estándares relacionado a la calidad de artefactos e insignias de la ACM, el *ii*) archivo contact, el *iii*) archivo requirement, el *iv*) archivo status y el *v*) archivo licence. Para *Ejercitable* los parámetros son: el *i*) registro del tipo de paquete, el *ii*) registro del entorno virtual, la *iii*) guía de instalación de los artefactos, los *iv*) artefactos estén comprimidos, el *v*) archivo install y el *vi*) registro de máximo 30 minutos de duración en la configuración e instalación de los artefactos del paquete de instalación.
- Se registra las características de calidad para obtener la insignia de Artefactos Disponibles: *i*) Disponible y *ii*) Accesible. Cada característica esta determinada por una lista de parámetros que deben cumplirse, que se encuentran registradas con un tipo de dato lógico *Boolean*. Para *Disponible* el parámetro es: el *i*) registro del DOI del PL. Para *Accesible* los parámetros son: el *i*) registro de un repositorio de archivos público y el *ii*) registro confidencial de credenciales y contraseñas de acceso a los artefactos en el caso de existir. Se registra dos parámetros *opcionales* de la característica *Disponible*: el *i*) archivo citation y el *ii*) archivo authors.
- Se registra la característica de calidad para obtener la insignia de Resultados Validados - Reproducido: *Verificable (con uso de material)*. La característica esta determinada por una lista de parámetros que deben cumplirse, que se encuentran registradas con un tipo de dato lógico *Boolean*. Se registra dos parámetros que son comunes para *Reproducido* y *Replicado*: el *i*) archivo authors y el *ii*) artículo científico, originales del experimento. Para la característica *Verificable (con uso de material)* de la insignia Reproducido, los parámetros son: *i*) Validado externamente usando parte del material experimental, el *ii*) archivo authors de la reproducción, el *iii*) artículo científico reproducido, el *iv*) archivo abstract de la reproducción, las *v*) pruebas sustanciales de la reproducción, el *vi*) respeto a los trabajos relacionados con la reproducción, el *vii*) resumen crítico de la reproducción y la *viii*) tolerancia de los resultados reproducidos. Se registran parámetros opcionales que a pesar

de ser opcionales deben ser consideradas para completar el PL de calidad. Algunos de estos parámetros son derivados de otros componentes del metamodelo, por ejemplo para que se cumpla el respeto de trabajos relacionados con la reproducción, es necesario registrar el archivo abstract en el cuál se explique los resultados de la reproducción. Estos parámetros opcionales están relacionadas con la característica de *Verificable (con uso de material)* y son: *i)* reflexiones críticas de la reproducción, la *ii)* solicitud de la insignia Reproducido y la *iii)* narración de acontecimientos de la reproducción.

- Se registra la característica de calidad para obtener la insignia de Resultados Validados - Replicado: *Verificable (sin uso de material)*. La característica esta determinada por una lista de parámetros que deben cumplirse, que se encuentran registradas con un tipo de dato lógico: *Boolean*. Para *Verificable (sin uso de material)* los parámetros son: *i)* Validado externamente sin usar el material experimental, el *ii)* archivo authors de la replicación, el *iii)* artículo científico replicación, el *iv)* archivo abstract de la replicación, las *v)* pruebas sustanciales de la replicación, el *vi)* respeto a los trabajos relacionados con la replicación, el *vii)* resumen crítico de la replicación y la *viii)* tolerancia de los resultados replicados. Algunos de éstos parámetros son derivados de otros componentes del metamodelo, por ejemplo para que se cumpla el respeto de trabajos relacionados con la replicación, es necesario registrar el archivo abstract en el cuál se explique los resultados de la replicación. Se registran parámetros opcionales que a pesar de ser opcionales deben ser consideradas para completar el PL de calidad. Estos parámetros opcionales están relacionadas con la característica de *Verificable (sin uso de material)* y son: *i)* reflexiones críticas de la replicación, la *ii)* solicitud de la insignia Replicado y la *iii)* narración de acontecimientos de la replicación.

6.4 APLICABILIDAD DEL PAQUETE DE LABORATORIO ORIENTADO AL PROCESO

En esta sección, se analiza el contexto de la aplicabilidad del PL orientado al proceso acorde a la ACM.

La propuesta del PL esta definida para experimentos controlados con humanos en el área de IS. Para ello, la *aplicabilidad* de la propuesta desde el punto de vista del experimentador se enfoca en dos contextos, por un lado cuando se ha realizado un

experimento controlado y el experimentador desea elaborar el PL, y por otro cuando un experimentador lleva a cabo la ejecución del experimento controlado y de manera paralela elabora el PL.

Además, la *aplicabilidad* de la propuesta también se enfoca cuando el experimentador desea evaluar el paquete laboratorio y conocer la posible insignia que alcanzaría dicho paquete. En base al cumplimiento de características de calidad de los artefactos que forman parte del contenido del PL establecido en el modelo de calidad, el experimentador puede conocer con alta probabilidad la asignación de la lista de insignias de la ACM y conocer el porcentaje de cumplimiento.

6.5 UNA HERRAMIENTA PARA EVALUAR PAQUETES DE LABORATORIO

En este apartado se describe la herramienta badgeGo, desarrollada en el grupo ISA de la Universidad de Sevilla y el grupo GISIA de la Universidad Técnica del Norte. badgeGo esta basada en el *metamodelo* propuesto en la presente memoria de tesis, para elaborar y evaluar el PL orientado al proceso acorde a la ACM de experimentos controlados con humanos en el área de IS. Ha sido desarrollada usando los frameworks para Backend Laravel 8.0 ¹ y para Frontend Angular 11.0 ²; el almacenamiento de archivos en la nube con Firebase ³, la base de datos con Postgresql ⁴ y desplegada en Heroku ⁵.

El actual prototipo consta de 10 vistas principales que permiten: *i)* describir el experimento original, *iii)* registrar a los experimentadores, *iii)* registrar los grupos experimentales junto con los participantes distribuidos en cada grupo, *iv)* registrar las tareas del experimento junto con los artefactos que se generan, *v)* registrar los artefactos que no pertenecen a las tareas, *vi)* registrar los artefactos de las insignias de la ACM, *vii)* establecer la configuración del PL, *viii)* evaluar los artefactos del PL para alcanzar las insignias de la ACM, *ix)* establecer una guía paso a paso para futuros replicadores, *x)* establecer el porcentaje de cumplimiento de los artefactos del PL y *xi)* registrar información del PL.

¹<https://laravel.com/>

²<https://angular.io/>

³<https://firebase.google.com/>

⁴<https://www.postgresql.org/>

⁵<https://www.heroku.com/>

La herramienta está disponible en <https://badgeGo.herokuapp.com/>. Cómo trabajo futuro se plantea seguir trabajando para que la herramienta evolucione como asistente y permita a los experimentadores generar de forma automática el PL en un archivo comprimido.

La figura §6.3 muestra la pantalla de inicio de badgeGo. A continuación se describen, de forma somera, las principales características de sus vistas.

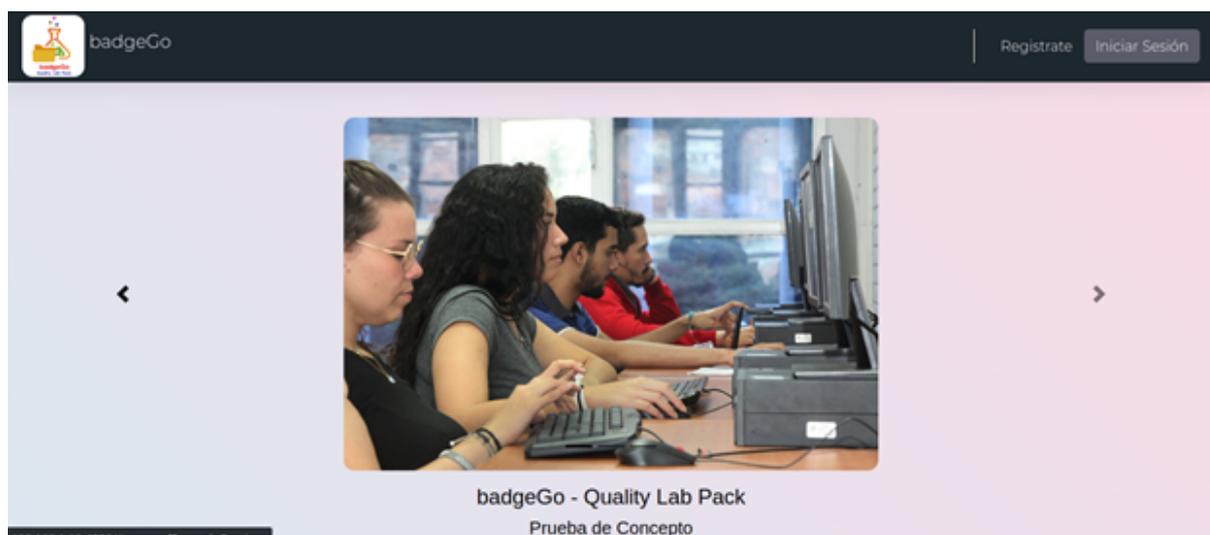


Figura 6.3: Pantalla de inicio

6.5.1 Vista del registro de los experimentos

Desde esta vista se visualizan los experimentos que se han registrado. La figura §6.4 muestra la lista de experimentos, permitiendo conocer campos como el contexto, el lugar, la fecha. Al seleccionar el botón de Opciones, se puede editar los campos.

6.5.2 Vista del registro de los experimentadores

En esta vista se visualiza el registro de datos informativos de los experimentadores como el nombre, género, email, afiliación, Rol que desempeña en la ejecución del experimento. Los campos del formulario son editables. La figura §6.5 muestra el registro de los experimentadores.

#	Código	Lugar	Descripción	Fecha	Opciones
1	EXP3	Cotacachi	descripcion del experimento	2022-02-02	[Opciones]
2	EXP002	Quito	Esta es una descripción	2021-11-22	[Opciones]
3	EXP001	Ibarra	Esta es una descripción	2021-11-20	[Opciones]

Figura 6.4: Vista de los experimentos registrados

#	Nombre	Apellido	Email	Pais	Organización	Experiencia	Opciones
1	Jhony	Martinez	mmartinez@gmail.com	Ecuador	Universidad Técnica del Norte	2 Años	[Detalles]

Figura 6.5: Vista del registro de los experimentadores

6.5.3 Vista del registro de los grupos y participantes

Desde esta vista se visualiza el registro de los grupos experimentales. Para cada grupo se asigna el número de participantes identificados por un acrónimo. La figura §6.6 muestra el registro de los grupos y los participantes.

6.5.4 Vista del registro de las tarea del experimento y sus artefactos

Desde esta vista se visualiza el registro de las tareas del experimento y los artefactos que se van generando en cada tarea. Para cada tarea y artefacto se registra datos informativos. La figura §6.7 muestra el registro de las tareas del experimento y sus artefactos.

6.5.5 Validación de las insignias de la ACM

Desde esta vista se visualiza la lista de artefactos que deben cumplirse para alcanzar las insignias de la ACM. Se observa el porcentaje de cumplimiento de las insignias y



Figura 6.6: Vista del registro de los grupos experimentales

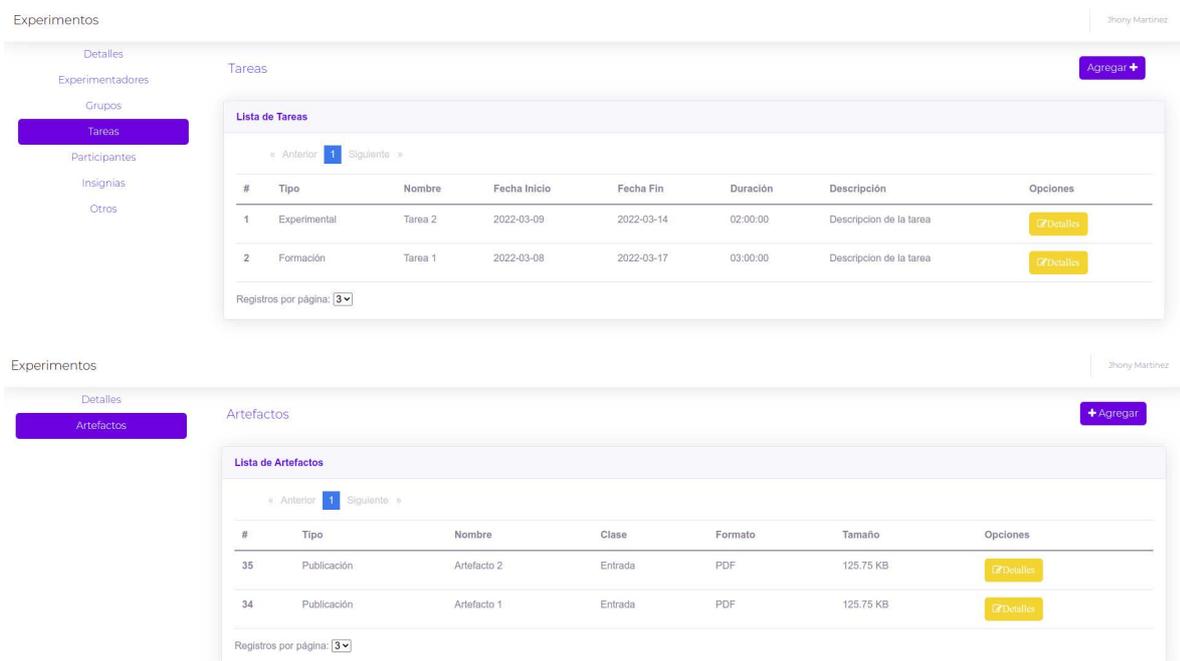


Figura 6.7: Vista del registro de las tareas del experimento y sus artefactos

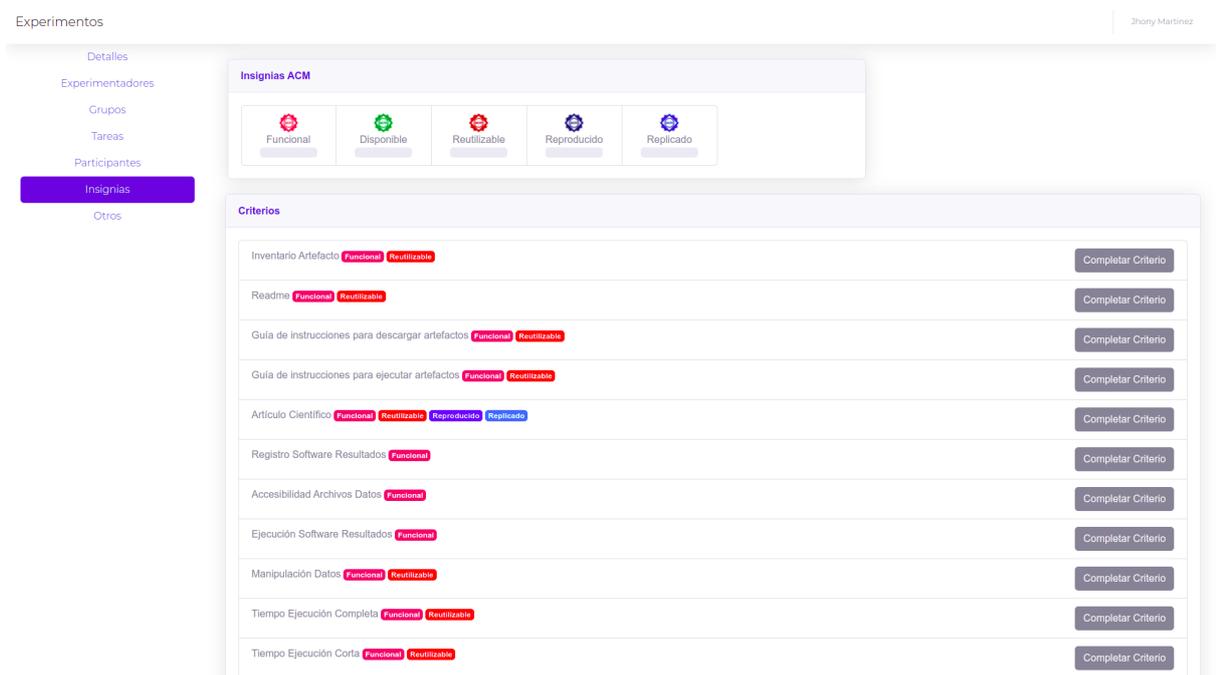


Figura 6.8: Vista de la validación de las insignias de la ACM para el Paquete de Laboratorio

se exporta el PL en un archivo comprimido. El paquete contiene los artefactos que se han registrado junto con las tareas del experimento, los artefactos que no pertenecen a las tareas y los artefactos que se relacionan con el cumplimiento de las insignias de la ACM. La figura §6.8 muestra la validación del las insignias de la ACM del PL.



Figura 6.9: Vista del registro del Paquete de Laboratorio

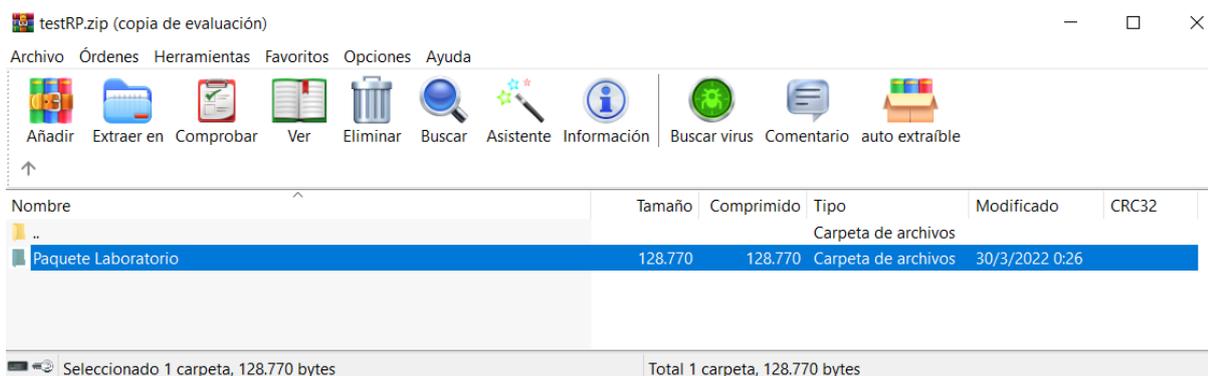


Figura 6.10: Vista del Paquete de Laboratorio en un archivo comprimido

6.5.6 Vista del registro del Paquete de Laboratorio

Desde esta vista se visualiza el registro del PL con datos informativos como el DOI, nombre, tipo de paquete, repositorio de archivos público donde se encuentra el paquete, fecha de creación y actualización del paquete. La figura §6.9 muestra el registro del PL. La figura §6.10 muestra el PL en un archivo comprimido.

6.6 RESUMEN

En este capítulo, se ha presentado la automatización del procedimiento para elaborar el PL acorde a la ACM mediante la herramienta badgeGo como un complemento a la principal aportación del trabajo recogido en esta memoria con la finalidad de fomentar las replicaciones y contribuir a la elaboración de PL de calidad en experimentos controlados con humanos en el área de IS. La automatización consta de tres partes: *i*) los requisitos generales que describen las características y las funcionalidades de la herramienta, *ii*) un metamodelo del Paquete de Laboratorio orientado al proceso acorde a la ACM y *iii*) una *herramienta* para automatizar el metamodelo llamado badgeGo.

Los *requisitos* generales que han permitido especificar y describir las funcionalidades principales de badgeGo.

El *metamodelo* que ha permitido identificar información correspondiente a la propuesta del PL orientado al proceso de manera conceptual. El metamodelo se conceptualiza utilizando la notación de diagrama de clases UML.

Basado en el *metamodelo* propuesto, como prueba de concepto más que como un prototipo, se presenta la *herramienta* badgeGo. Dicha herramienta, permite a los expe-

rimentadores la elaboración y evaluación del PL, que pasa a dar información sobre la posible insignia que el paquete cumple.

Se espera que la solución propuesta para la elaboración y evaluación del PL, redunde en beneficio de las investigaciones en IS.

PARTE IV

VALIDACIÓN

EVALUACIÓN DE LA PROPUESTA EN UN ESTUDIO DE CASO

Mide lo que es medible, y haz medible lo que no lo es

Galileo Galilei (1564 – 1642),

Astrónomo, ingeniero, filósofo, matemático y físico italiano

En este capítulo, se valida la propuesta aplicándola a experimentos ya realizados, lo que permite ver si la propuesta es usable y se obtiene un PL de calidad. Sin embargo, lo ideal sería realizar una validación cuyo proceso comenzara durante la realización del experimento. En este sentido, los experimentadores, ayudados por la herramienta badgeGo, irían registrando las tareas realizadas con sus artefactos. Además, badgeGo va guiando al experimentador para que el PL resultante sea de calidad y pueda obtener alguna insignia de la ACM. Queda como trabajo futuro esta validación. Se describe la evaluación de la propuesta mediante una prueba de concepto utilizando: i) el procedimiento para elaborar un PL acorde a la ACM definido en la sección §5.3, ii) el modelo de calidad para PL definido en la sección §5.4 y iii) el metamodelo del PL definido en la sección ???. La evaluación consiste en determinar los aspectos de Viabilidad, Flexibilidad y Satisfacción para la elaboración de un PL acorde a la ACM. La evaluación se realiza mediante un estudio de caso múltiple que incluye los PL de una familia de experimentos y dos experimentos originales que han sido ejecutados por los investigadores del grupo ISA, todos orientados a experimentos controlados con humanos en el área de la IS. En la sección §7.1, se presenta el estudio de caso. En la sección §7.2, se detallan las fases del estudio de caso. En la sección §7.3, se evalúa el aspecto de Viabilidad de la propuesta instanciando el PL en el Caso-APIs. En la sección §7.4, se evalúa el aspecto de Flexibilidad de la propuesta instanciando el PL en el Caso-Mind y el Caso-Req. En la sección §7.5, se evalúa el aspecto de Satisfacción de la propuesta aplicando un cuestionario a los experimentadores que instanciaron la propuesta. En la sección §7.6, se responde a las preguntas de investigación del estudio de

caso. En la sección §7.7, se muestran las conclusiones. Por último, la sección §7.8 resume el capítulo.

7.1 INTRODUCCIÓN

En la comunidad científica en el área de IS el *estudio de caso* se especifica como una estrategia empírica y que se usa con frecuencia [172]. El *estudio de caso* se lleva a cabo en su contexto natural [354]. Los casos pueden variar desde estudios bien organizados y robustos hasta pequeños ejemplos en un laboratorio académico como un ambiente universitario [355]. La estrategia empírica usada en este estudio se basa en el proceso de investigación de *estudios de caso* propuesto en Runeson y Höst [354] y Runeson *et al.* [355].

El presente *estudio de caso* abarca las temática de conocimiento de IS: *i)* familia de experimentos de Mindfulness, *ii)* análisis multilingual de requisitos y *iii)* análisis de paradigmas de construcción de arquitecturas microservicios; con el objetivo principal de determinar los aspectos de *Viabilidad, Flexibilidad y Satisfacción* para la elaboración del PL orientado al proceso acorde a la ACM dirigido a experimentos controlados con humanos en el área de IS.

Es importante indicar que la presente propuesta de tesis esta orientada a que, la elaboración del Paquete de Laboratorio orientado al proceso pueda ejecutarse para dos contextos: *i)* cuando el experimento, reproducción o replicación se ha finalizado, entonces se da inicio a la elaboración del PL y *ii)* cuando el experimento, reproducción o replicación se ha iniciado, entonces la elaboración del PL se hace al mismo tiempo que se va ejecutando el proceso experimental. En esta memoria de tesis la validación de la propuesta del PL orientado al proceso se aplica para experimentos que ya han sido finalizados. Considerando que, como trabajo futuro la validación del PL sea aplicado para experimentos que den inicio al proceso experimental.

7.2 ESTUDIO DE CASO

Para el presente *estudio de caso*, se identifican tres fases principales: *i)* Diseño del estudio de caso; *ii)* Planificación, recopilación y análisis; y *iii)* Análisis e informe conjunto, que se describen a continuación. La figura §7.1 muestra dichas fases.

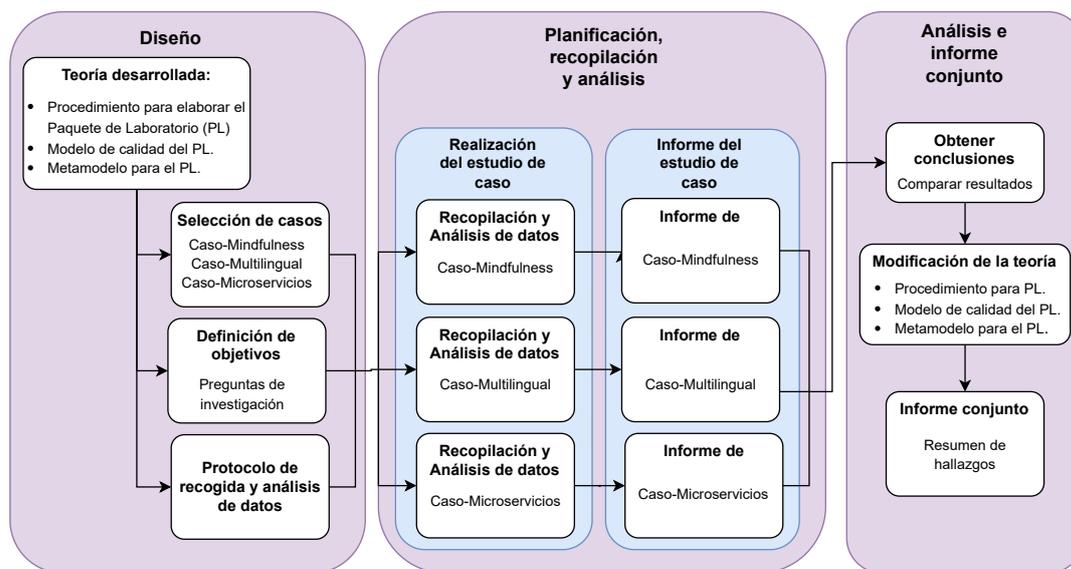


Figura 7.1: Fases del estudio de caso basado en [355]

7.2.1 Diseño del estudio de caso

Esta fase permite clarificar el contexto de investigación del estudio. En secciones anteriores, se ha desarrollado la teoría presentando: *i)* un procedimiento del PL acorde a al ACM, *ii)* un modelo de calidad para PL y *iii)* un metamodelo como propuesta para elaborar y evaluar PL orientados al proceso en IS. Para validar la propuesta mediante un *estudio de caso múltiple*, se definen los objetivos que permiten proponer las preguntas de investigación y establecer el protocolo para la recopilación y el análisis de datos.

Los pasos de esta primera fase se detallan a continuación:

1. *Teoría desarrollada.* Se ha presentado: *i)* un procedimiento para elaborar un PL acorde a la ACM (ver sección §5.3), *ii)* un modelo de calidad para PL (ver sección §5.4) y *iii)* un metamodelo que formaliza la información involucrada en el proceso experimental de forma integrada con el PL (ver sección ??).
2. *Selección de casos.* Para evaluar la teoría desarrollada, se seleccionan una familia de experimentos y dos experimentos originales que cubren varias temáticas de conocimiento en el área de la IS:
 - *Estudio de caso en la familia de experimentos de Mindfulness (Mind).* Consiste en un experimento original y dos replicas internas. El experimento evalúa la práctica de Mindfulness en el rendimiento de los participantes durante el

desarrollo de modelos conceptuales para IS. El experimento de Mindfulness también ha sido conducido en la industria para buscar evidencias empíricas de los beneficios del Mindfulness aplicado en un entorno de la industria del software, sobre el cual, ha sido construido el PL, que se aprecia en la figura §7.2 [52].

October 28, 2021

An Empirical Study to Evaluate the Impact of Mindfulness at a Software Company (Laboratory Package)

Beatriz Bernárdez; J. Ignacio Panach; Jose A. Parejo; Amador Durán; Natalia Juristo; Antonio Ruiz-Cortés

For the empirical study carried out in an Andalusian software company to evaluate the effects of the practise of mindfulness, this laboratory package contains the experimental material used during its execution, the data sets (both "csv" and "sav" files) derived from the experimental tasks, and the results of the statistical analysis (both scripts and plots).

Preview

Bernardez_et_al_2021_Mindfulness_EmpiricalStudy.zip

- Labpack
 - 1. Experimental Material
 - 1. Slides
 - 1. Recruitment [Talk].pdf 540.8 kB
 - 2. Mindfulness [Talk].pdf 2.2 MB
 - 3. To continue with Mindfulness [Talk].pdf 672.2 kB
 - 2. Questionnaires
 - Demographic and consent [Quest].pdf 184.5 kB
 - MAAS [Quest].pdf 647.0 kB
 - Satisfaction [Quest].pdf 301.3 kB
 - 2. Datasets
 - csv Files
 - KPIs-rawdata.csv 3.2 kB
 - MAAS-rawdata.csv 5.3 kB
 - sav Files
 - KPIs rawdata.sav 47.0 kB
 - MAAS rawdata.sav 1.6 kB
 - 3. Analysis

Figura 7.2: Paquete de Laboratorio de la familia de experimentos Mindfulness

- *Estudio de caso en el experimento Multilingual (Req)*. Consiste en un experimento original para comparar los requisitos escritos en lengua materna con los escritos en lengua extranjera y analizar el impacto en la calidad externa del desarrollo de software.
 - *Estudio de caso en el experimento Microservicios (APIs)*. Consiste en un experimento original para comparar dos paradigmas de construcción de arquitecturas microservicios API REST vs API GraphQL para comparar los efectos en la calidad del software.
3. *Definición de objetivos*. El objetivo general de la validación es evaluar empíricamente que la propuesta del PL orientado al proceso acorde a la ACM puede ser

instanciado y utilizado en experimentos controlados con humanos en el área de IS.

Previamente, se ha realizado una revisión de la literatura para analizar cómo han abordado otros investigadores el reporte de los PL, que guías usan, en qué repositorios de archivos públicos reportan el paquete y que insignias de la ACM han alcanzado (ver sección §3).

Los aspectos que se analiza en la presente evaluación son: *Viabilidad*, *Flexibilidad* y *Satisfacción* de la propuesta, lo que permite plantear el conjunto de preguntas de investigación (*PIs*) que se presentan a continuación:

- **PI_{3.1}:** *Viabilidad*. ¿Ha sido posible instanciar el Paquete de Laboratorio orientado al proceso acorde a la ACM en un experimento en concreto?
 - **PI_{3.2}:** *Flexibilidad*. ¿Ha sido posible instanciar el Paquete de Laboratorio orientado al proceso acorde a la ACM en distintos experimentos controlados con humanos?
 - **PI_{3.3}:** *Satisfacción*. ¿En qué medida la propuesta de desarrollo de un Paquete de Laboratorio orientado al proceso acorde a la ACM permite ser útil, confiable y cómodo para los experimentadores.
4. *Protocolo de recogida y análisis de datos*. Se trata de identificar las fuentes para recopilar los datos y definir los procedimientos para llevar a cabo el análisis de los resultados.

En el presente *estudio de caso*, el protocolo de recolección de datos consiste en: *i*) contactar con los experimentadores; *ii*) registrar el contenido de la estructura del paquete; *iii*) registrar los detalles de cómo, cuándo y quién llevó a cabo este proceso; y *iv*) pasar un cuestionario a los experimentadores después de cumplimentar el registro.

7.2.2 Planificación, recopilación y análisis

En esta segunda fase se aplica el protocolo anterior al estudio de caso y se escribe su informe.

Comprende, para cada caso, los procesos de: *i*) *recopilación de datos y análisis de los datos recogidos* y *ii*) *elaboración del informe del estudio de caso*. Los pasos se detallan a continuación:

1. *Proceso de recopilación de datos.* Lethbridge *et al.* [241] definen tres grados para las técnicas de recolección de datos: *i)* el primer grado requiere *acceso directo* a la población participante, *ii)* el segundo grado requiere *acceso al entorno de trabajo* de los participantes, y *iii)* el tercer grado requiere *acceso sólo a los artefactos de trabajo*. A continuación, para cada caso, se especifican las técnicas de recolección de datos utilizadas y las actividades involucradas.

- En el *Caso-Mind*, se mantiene varias reuniones con los experimentadores quienes pertenecen al área de conocimiento en IS para dar a conocer la propuesta y como van usar la estructura del PL orientado al proceso acorde a la ACM. Luego de cumplimentar la estructura se pasa un cuestionario para recoger la opinión de los experimentadores y medir el aspecto de *Satisfacción*.
- En el *Caso-Req* y el *Caso-APIs* el procedimiento es similar al caso anterior.

Se utilizan *técnicas de recogida de datos* de los tres grados: *i)* se recogen opiniones sobre la *Satisfacción* de la propuesta mediante un cuestionario. Para los experimentadores que están fuera del alcance geográfico se trabaja de manera virtual y para los experimentadores locales se les visita en los centros de trabajo (primer y segundo grado); y *ii)* para seleccionar los casos de los experimentos se utilizan datos de archivo (*archival data*, es decir, tipo de datos que se puede recopilar mediante técnicas de tercer grado en un estudio de caso), revisando las principales publicaciones de los experimentadores implicados.

2. *Análisis de los datos recogidos.* Una vez que se han recopilado los datos, es necesario su análisis tanto *cualitativo* como *cuantitativo*.

Se analizan los datos recopilados fundamentalmente al cumplimentar la evaluación de la propuesta y los cuestionarios que se han pasado a los experimentadores.

Así, tras recopilar la información de la evaluación cumplimentada, se analiza si los campos sin cumplimentar (en blanco) son debidos a la falta de información. Igualmente, se analiza si la propuesta recoge toda la información que el experimentador necesita reflejar.

Hay otros detalles de interés sobre la propuesta como: *quién la cumplimenta* (por ejemplo, los autores del experimento), *cómo* (por ejemplo, a partir del informe del experimento y/o el PL), *cuándo* (por ejemplo, después de que se lleve a cabo el experimento) en caso de no ser cumplimentada por los autores del experimento.

La instanciación completa por primera vez del PL orientado al proceso acorde a la ACM aplicada en el *Caso-APIs* para evaluar el aspecto de *Viabilidad*, se presentan en el Anexo §B.

La instanciación completa del PL orientado al proceso acorde a la ACM aplicada en el *Caso-Mind* para evaluar el aspecto de *Flexibilidad*, se presentan en el Anexo §C.

La instanciación completa del PL orientado al proceso acorde a la ACM aplicada en el *Caso-Req* para evaluar el aspecto de *Flexibilidad*, se presentan en el Anexo §D.

El cuestionario utilizado para el aspecto de *Satisfacción*, se presenta en el Anexo §E.

3. *Informe del Estudio de caso*. Es la redacción del informe de cada caso. Las secciones §7.3, §7.4 y §7.5 reportan en detalle cada uno de los tres casos.

7.2.3 Análisis e informe conjunto

Una vez analizado cada caso en su entorno, el siguiente paso es integrar los resultados obtenidos de manera que se puedan comparar y conseguir el objetivo del *Estudio de caso*, es decir, obtener conclusiones más sólidas. Los pasos se detallan a continuación:

1. *Obtener conclusiones*. Se extraen conclusiones cruzadas y se comparan los resultados de la aplicación de la propuesta en los tres casos. Las conclusiones se presentan en la sección §7.7.
2. *Modificación de la teoría*. Las limitaciones encontradas en la aplicación de la propuesta posibilitan que evolucione la teoría desde su versión inicial incorporando las mejoras pertinentes. Los ajustes realizados al aplicar la propuesta se detalla en cada caso.
3. *Informe conjunto*. Finalmente, se escribe el informe que se presenta en la sección §7.6.

7.3 EVALUACIÓN DE LA VIABILIDAD PARA LA PROPUESTA

En este apartado se describe la evaluación del aspecto de *Viabilidad* de la propuesta del Paquete de Laboratorio orientado al proceso acorde a la ACM. Se lleva a cabo la instanciación de un PL en concreto a partir de la propuesta de la presente tesis. La *Viabilidad* es la posibilidad técnica y administrativa de que la propuesta sea factible.

7.3.1 Definición de los atributos de Viabilidad

La instanciación se realiza para el *Caso-APIs*. Para obtener el PL orientado al proceso acorde a la ACM, se utiliza la estructura propuesta definida a partir de la teoría desarrollada. Los resultados se analizan para evaluar el aspecto de *Viabilidad* de la propuesta y contestar a la pregunta **PI_{3,1}** descrita en el caso de estudio. La instanciación es realizada por uno de los experimentadores, con el fin de encontrar sugerencias, observaciones y limitaciones de la propuesta durante la instanciación.

La evaluación de *Viabilidad* se realiza formalmente una solo vez, es importante como primer paso evaluar la *Viabilidad* porque demuestra que la propuesta de la investigación es posible. Los atributos de *Viabilidad* han sido adoptados del estudio de Solari [392] y estos son:

- Obtención de la instancia.- Indica si obtiene o no un Paquete de Laboratorio orientado al proceso acorde a la ACM a partir de la información recogida.
- Esfuerzo de instanciación.- Esfuerzo medido en horas-persona que dedica el investigador durante la instanciación.
- Aceptación.- Determina la aceptación de la instancia por parte de los autores originales del experimento.

7.3.2 Caso-APIs

En este caso, se evalúa la propuesta por primera vez en un experimento controlado del área de IS, analizando la *Viabilidad* que implica obtener una instancia en un tiempo razonable para el investigador y que el resultado sea aceptado por los autores del experimento. El experimento *Caso-APIs* ha sido seleccionado por ser conocido ya

que ha dado origen a la propuesta y por tener acceso al material experimental. Este experimento se trata de comparar dos paradigmas de construcción de arquitecturas microservicios API REST vs API GraphQL para medir los efectos en la calidad del software.

Para evaluar la *Viabilidad* de la propuesta, se especifica y se estructura la información de la teoría desarrollada. En el Anexo §B se presenta la tabla §B.1, que muestra, la instanciación de la propuesta del Caso-APIs mediante el desarrollo de los pasos del procedimiento para elaborar un PL acorde a la ACM definido en la sección §5.3 y los atributos del metamodelo propuesto en la sección ???. El *Campo* representa los atributos del metamodelo y el *Valor* llenado de la información del Paquete de Laboratorio del *Caso-APIs*.

En los siguientes apartados se detallan las limitaciones encontradas y los ajustes realizados específicamente en el *metamodelo*. La instanciación ha permitido, por un lado, la evolución del *metamodelo*, se han añadido nuevos atributos y se han modificado algunos de ellos; y por otro, se ha demostrado que existe información relevante que se debe considerar.

7.3.3 Evolución del metamodelo a partir de la instanciación del Paquete de Laboratorio

En la presente sección, se detallan los ajustes introducidos en el *metamodelo* como consecuencia de los problemas encontrados al instanciar por primera vez la propuesta en el *Caso-APIs*.

Necesidad de actualizar información de los experimentadores

- Años de experiencia de los experimentadores: Existe confusión en registrar los años de experiencia de los experimentadores, por lo cuál, este campo, se elimina del *metamodelo*.
- Teléfono del experimentador: No es necesario registrar el número de teléfono del experimentador, por lo cuál, este campo se elimina en el *metamodelo*.
- Fecha de nacimiento del experimentador: No es necesario registrar la fecha de nacimiento del experimentador, por lo cuál, este campo se elimina en el *metamodelo*.

La tabla §7.1 muestra la versión actualizada de la estructura presentada en el *Se-*

gundo paso del procedimiento de la propuesta referente a los *Experimentadores*, para la primera vez de la instanciación del *Caso-APIs*.

Tabla 7.1: Actualización de la información del experimentador en la estructura de la primera instanciación del *Caso-APIs*

Tercer paso	Registro de experimentadores
Campo	Valor
DNI	1002322384
Nombre del experimentador	Antonio Quiña
email	aquina@utn.edu.ec
Afiliación	Universidad Técnica del Norte
Sitio web	https://bit.ly/3Jx88T3
Género	Masculino
Nacionalidad	Ecuatoriana
Tipo de persona	Profesional
Rol del experimentador	Formador, Diseñador, Analista
Comentario	No se registra

La figura §7.3 muestra la actualización del antes y después de la clase "Persona" en el *metamodelo*.

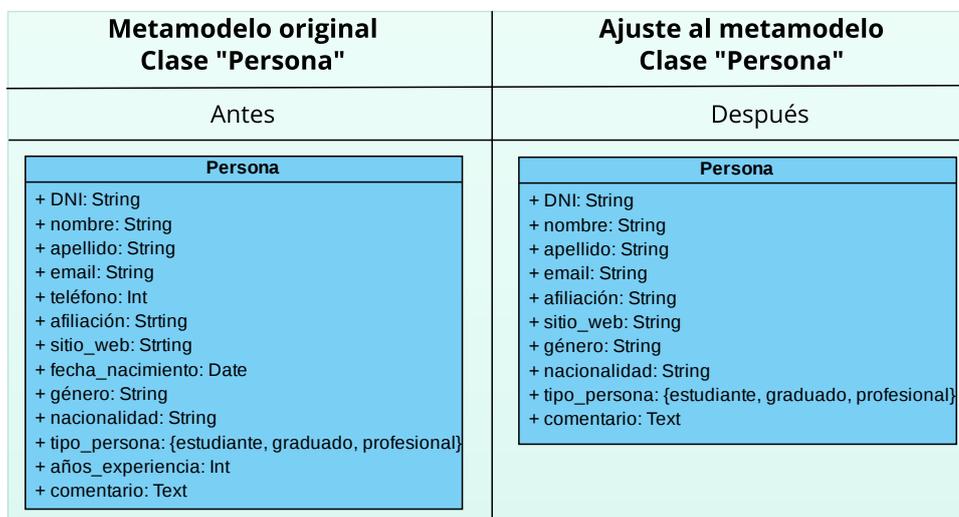


Figura 7.3: Actualización de la clase "Persona" en el metamodelo

Necesidad de añadir datos de los grupos

- # Grupos de control y experimental: El número de grupos de control y experimental no se han registrado en el *metamodelo*, al igual, que el número de participantes que conforman cada grupo.

- Ajuste en el metamodelo: Se agregaron los atributos *# grupo de control*, *# grupo experimental*, y se conserva la agregación con la clase "Rol_Participante" para asignar el *# participantes del grupo de control* y el *# participantes del grupo experimental* en el *metamodelo*.

La tabla §7.2 muestra la versión actualizada de la estructura presentada en el *Tercer paso* del procedimiento de la propuesta referente a los *Grupos*, para la primera vez de la instanciación del *Caso-APIs*.

Tabla 7.2: Actualización de la información de Grupos en la estructura de la primera instanciación del Caso-APIs

Segundo paso	Grupos y asignación de participantes
Campo	Valor
# grupos de Control	0
# participantes x grupo de Control	0
# grupos Experimentales	2
# participantes x grupo Experimental	24

La figura §7.4 muestra la actualización del antes y después de la clase "Grupo" en el *metamodelo*.

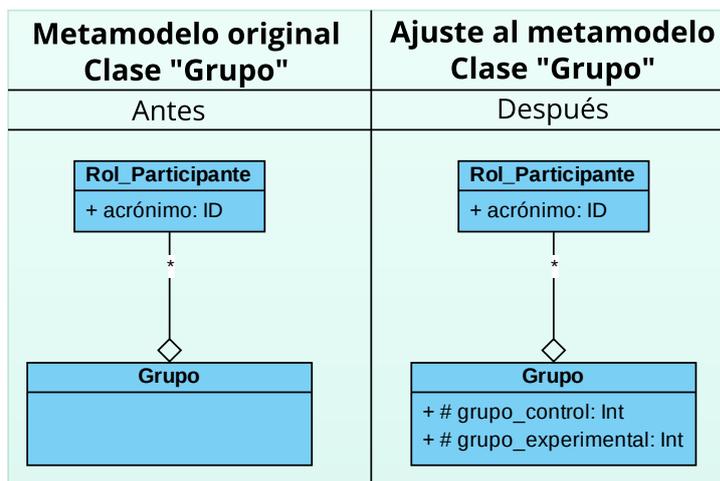


Figura 7.4: Actualización de la clase "Grupo" en el metamodelo

Necesidad de añadir datos a las tareas

- Responsable: El responsable no es registrado en la clase "Tarea", por cuál, es necesario agregar este campo en el *metamodelo*.

La tabla §7.3 muestra la versión actualizada de la estructura presentada en el *Cuarto paso* del procedimiento de la propuesta referente a la *Tarea*, para la primera vez de la instanciación del *Caso-APIs*.

Tabla 7.3: Actualización de la información de la tarea en la estructura de la primera instanciación del Caso-APIs

Cuarto paso	Registro de tareas del experimento y sus artefactos
Campo	Valor
Acrónimo	T01
Nombre de la tarea	Encuesta demográfica/diagnóstica
Tipo de tarea	Preparación
Fecha de inicio	19 de junio 2021
Fecha de fin	19 de junio 2021
Duración	4 minutos
Responsable	Formador
Descripción	Contiene información demográfica de los participantes para diagnosticar el conocimiento y percepción de la facilidad de uso, consumo y aprendizaje de los paradigmas REST y GraphQL

La figura §7.5 muestra la actualización del antes y después de la clase "Tarea".

Metamodelo original Clase "Tarea"	Ajuste al metamodelo Clase "Tarea"
Antes	Después
<div style="border: 1px solid black; padding: 5px; width: fit-content; margin: auto;"> <p style="text-align: center; margin: 0;">Tarea</p> <ul style="list-style-type: none"> + acrónimo: ID + nombre_tarea: String + fecha_inicio: Date + fecha_fin: Date + duración_tarea: Date + descripción: Text </div>	<div style="border: 1px solid black; padding: 5px; width: fit-content; margin: auto;"> <p style="text-align: center; margin: 0;">Tarea</p> <ul style="list-style-type: none"> + acrónimo: ID + nombre_tarea: String + fecha_inicio: Date + fecha_fin: Date + duración_tarea: Date + responsable_tarea: String + descripción: Text </div>

Figura 7.5: Actualización de la clase "Persona" en el metamodelo

Necesidad de actualizar datos a los artefactos

- Encuesta: El tipo de finalidad "Encuesta" debería formar parte del tipo de artefacto. Se realizar el cambio en el *metamodelo*.
- Cuestionario: El tipo de finalidad "Cuestionario" debería formar parte del tipo de artefacto. Se realizar el cambio en el *metamodelo*.

- Formación: El tipo de finalidad "Formación" no ha sido registrado, por lo cuál, se agrega este tipo de finalidad en el *metamodelo*.
- Reportes: El tipo de finalidad "Reportes" no ha sido registrado, por lo cuál, se agrega este tipo de finalidad en el *metamodelo*.

La tabla §7.4 muestra la versión actualizada de la estructura presentada en el *Cuarto paso* del procedimiento de la propuesta referente a los *Artefactos*, para la primera vez de la instanciación del *Caso-APIs*.

Tabla 7.4: Actualización de la información de los artefactos en la estructura de la primera instanciación del Caso-APIs

Artefacto de la T01	
Nombre del artefacto	encuesta-apis
Tipo de finalidad	Reclutamiento
Tipo de artefacto	Encuesta
Tipo de formato	PDF
Tamaño	90 KB
Clase de artefacto	Entrada
Path	https://bit.ly/3to1QiY
Fecha de creación	20 mayo 2021
Fecha de la última actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	4 minutos
Tiempo corto de ejecución	2 minutos
Comentario	No se registra

La figura §7.6 muestra la actualización del antes y después de "Tipo_Artefacto" y "Tipo_Finalidad" en el *metamodelo*.

La figura §7.7 muestra la actualización del antes y después de la clase "Artefacto" en el *metamodelo*.

7.4 EVALUACIÓN DE LA FLEXIBILIDAD PARA LA PROPUESTA

En este apartado se describe la evaluación del aspecto de *Flexibilidad* de la propuesta del PL orientado al proceso acorde a la ACM. La evaluación de la *Flexibilidad* se realiza

Metamodelo original "Tipo_Artefacto" y "Tipo_Finalidad"		Ajuste al metamodelo "Tipo_Artefacto" y "Tipo_Finalidad"	
Antes		Después	
<p><<enumeration>> Tipo_Artefacto</p> <ul style="list-style-type: none"> Pósters Videos Audios Imágenes Formularios Diagramas Software Lecciones Documentos Sitios web Publicaciones Presentaciones Objetos físicos Archivos binario Archivos de datos Archivos de texto Archivos de código Archivos comprimidos 	<p><<enumeration>> Tipo_Finalidad</p> <ul style="list-style-type: none"> Scripts Datasets Charlas Requisitos Reclutamiento Instrucciones Encuestas Cuestionarios Modelos UML Gráficos estadísticos 	<p><<enumeration>> Tipo_Artefacto</p> <ul style="list-style-type: none"> Pósters Videos Audios Imágenes Formularios Encuestas Cuestionarios Diagramas Software Lecciones Documentos Sitios web Publicaciones Presentaciones Objetos físicos Archivos binario Archivos de datos Archivos de texto Archivos de código Archivos comprimidos 	<p><<enumeration>> Tipo_Finalidad</p> <ul style="list-style-type: none"> Scripts Datasets Charlas Requisitos Reclutamiento Instrucciones Formación Reportes Modelos UML Gráficos estadísticos Insignias ACM

Figura 7.6: Actualización en "Tipo_Artefacto" y "Tipo_Finalidad" en el metamodelo

instanciando la propuesta para distintos experimentos. El procedimiento de instanciación es similar al seguido en la evaluación de viabilidad. La *Flexibilidad* es la capacidad que debe tener la propuesta para poder adaptarse a los continuos cambios que pueden afectar al contexto interno y externo del experimento. Sin embargo, la *Flexibilidad* debe seguir evaluándose con el mayor número de experimentos posibles para solventar los cambios.

7.4.1 Definición de los atributos de Flexibilidad

Para la evaluación de *Flexibilidad* se instancia la propuesta en dos casos adicionales al utilizado en el aspecto de *Viabilidad*. Los casos seleccionados son: *Caso-Mind* y *Caso-Req*. El contexto de la presente propuesta, es específicamente experimentos controlados con humanos para el área de IS, sin embargo, por más que los experimentos sean controlados con humanos cada uno conlleva particularidades propias de su contexto, de ahí la importancia de clarificar el contexto del experimento, lo cuál ha sido definido en la sección §4.6 de esta memoria de tesis.

Para obtener el PL orientado al proceso acorde a la ACM, se utiliza la estructura propuesta definida a partir de la teoría desarrollada y que ha sido actualizada luego

Metamodelo original Clase "Artefacto"	Ajuste al metamodelo Clase "Artefacto"																										
Antes	Después																										
<table border="1"> <thead> <tr> <th>Artefacto</th> </tr> </thead> <tbody> <tr><td>+ nombre_artefacto: String</td></tr> <tr><td>+ finalidad: Tipo_Finalidad</td></tr> <tr><td>+ artefactos: Tipo_Artefacto</td></tr> <tr><td>+ formatos: Tipo_Formato</td></tr> <tr><td>+ tamaño_artefacto: String</td></tr> <tr><td>+ path_artefacto: String</td></tr> <tr><td>+ fecha_creación: Date</td></tr> <tr><td>+ fecha_actualización: Date</td></tr> <tr><td>+ acceso_usuario_password: String</td></tr> <tr><td>+ tiempo_ejecución_completa: Time</td></tr> <tr><td>+ tiempo_ejecución_corta: Time</td></tr> <tr><td>+ comentario: Text</td></tr> </tbody> </table>	Artefacto	+ nombre_artefacto: String	+ finalidad: Tipo_Finalidad	+ artefactos: Tipo_Artefacto	+ formatos: Tipo_Formato	+ tamaño_artefacto: String	+ path_artefacto: String	+ fecha_creación: Date	+ fecha_actualización: Date	+ acceso_usuario_password: String	+ tiempo_ejecución_completa: Time	+ tiempo_ejecución_corta: Time	+ comentario: Text	<table border="1"> <thead> <tr> <th>Artefacto</th> </tr> </thead> <tbody> <tr><td>+ nombre_artefacto: String</td></tr> <tr><td>+ finalidad: Tipo_Finalidad</td></tr> <tr><td>+ artefactos: Tipo_Artefacto</td></tr> <tr><td>+ formatos: Tipo_Formato</td></tr> <tr><td>+ tamaño_artefacto: String</td></tr> <tr><td>+ path_artefacto: String</td></tr> <tr><td>+ fecha_creación: Date</td></tr> <tr><td>+ fecha_última_actualización: Date</td></tr> <tr><td>+ acceso_usuario_password: String</td></tr> <tr><td>+ tiempo_ejecución_completa: Time</td></tr> <tr><td>+ tiempo_ejecución_corta: Time</td></tr> <tr><td>+ comentario: Text</td></tr> </tbody> </table>	Artefacto	+ nombre_artefacto: String	+ finalidad: Tipo_Finalidad	+ artefactos: Tipo_Artefacto	+ formatos: Tipo_Formato	+ tamaño_artefacto: String	+ path_artefacto: String	+ fecha_creación: Date	+ fecha_última_actualización: Date	+ acceso_usuario_password: String	+ tiempo_ejecución_completa: Time	+ tiempo_ejecución_corta: Time	+ comentario: Text
Artefacto																											
+ nombre_artefacto: String																											
+ finalidad: Tipo_Finalidad																											
+ artefactos: Tipo_Artefacto																											
+ formatos: Tipo_Formato																											
+ tamaño_artefacto: String																											
+ path_artefacto: String																											
+ fecha_creación: Date																											
+ fecha_actualización: Date																											
+ acceso_usuario_password: String																											
+ tiempo_ejecución_completa: Time																											
+ tiempo_ejecución_corta: Time																											
+ comentario: Text																											
Artefacto																											
+ nombre_artefacto: String																											
+ finalidad: Tipo_Finalidad																											
+ artefactos: Tipo_Artefacto																											
+ formatos: Tipo_Formato																											
+ tamaño_artefacto: String																											
+ path_artefacto: String																											
+ fecha_creación: Date																											
+ fecha_última_actualización: Date																											
+ acceso_usuario_password: String																											
+ tiempo_ejecución_completa: Time																											
+ tiempo_ejecución_corta: Time																											
+ comentario: Text																											

Figura 7.7: Actualización en la clase "Artefacto" en el metamodelo

de haber evaluado el aspecto de *Viabilidad*. Los resultados se analizan para evaluar el aspecto de *Flexibilidad* de la propuesta y contestar a la pregunta **PI**_{3.2} descrita en el caso de estudio. La instanciación de cada caso es realizada por uno de los experimentadores, con el fin de encontrar sugerencias, observaciones y limitaciones de la propuesta durante la instanciación.

Los atributos de *Flexibilidad* considerados para esta evaluación han sido adoptados del estudio de Solari [392] y se definen a continuación:

- Diversidad de experimentos: Indica la *Flexibilidad* de la propuesta por el número de experimentos y/o familias de experimentos instanciados.
- Uso en replicaciones futuras: Considera si la instancia obtenida es satisfactoria para ser utilizada en reproducciones y/o replicaciones futuras. Este atributo es valorado por los mismos investigadores responsables del experimento.

7.4.2 Caso-Mind

En este caso, se evalúa la *Flexibilidad* mediante la instanciación del *Caso-Mind* utilizando la propuesta que fue actualizada en la evaluación del aspecto de *Viabilidad*. El *Caso-Mind* esta compuesto por la familia de experimentos de Mindfulness llevado a cabo en la academia y un experimento de Mindfulness conducido en la industria. El *Caso-Mind* ha sido seleccionado por ser conocido ya que ha dado origen a la propuesta y por tener acceso al material experimental.

La familia de experimentos de Mindfulness (*Mind*) [48, 51], consiste en un experimento y sus dos replicaciones internas realizadas en la Universidad de Sevilla, la segunda replicación se ha definido sobre la base de la primera replicación. La familia de experimentos trata sobre el efecto de la práctica de *Mindfulness* en el rendimiento de los estudiantes al momento de desarrollar modelos conceptuales. El experimento de *Mindfulness* también ha sido conducido en la industria para buscar evidencias empíricas de los beneficios del *Mindfulness* aplicado en un entorno de la industria del software, de lo cuál, se elabora un PL basado en la familia de *Mindfulness*.

Mindfulness, es un método para conseguir la atención plena, centrándose en lo que está sucediendo “aquí y ahora”, su significado es plena consciencia. El acrónimo usado para hacer referencia al PL del experimento en la industria es *Mind-PL*.

En el Anexo §C se presenta la tabla §C.1, que muestra, la instanciación de la propuesta del *Caso-Mind* aplicando la propuesta actualizada del PL orientado al proceso acorde a al ACM. La instanciación del *Caso-Mind* ha sido realizado por el experimentador, con la finalidad de encontrar limitaciones a la propuesta.

7.4.3 Caso-Req

En este caso, se evalúa la *Flexibilidad* mediante la instanciación del *Caso-Req* utilizando la propuesta que fue actualizada en la evaluación del aspecto de *Viabilidad*. El *Caso-Req* ha sido seleccionado por ser conocido ya que ha dado origen a la propuesta y por tener acceso al material experimental. Este experimento se trata de comparar los requisitos escritos en lengua materna con los escritos en lengua extranjera y analizar el impacto en la calidad externa del desarrollo de software. El experimento se realizó con 17 estudiantes que cursaban la Maestría de Ingeniería del Software en la Universidad de las Fuerzas Armadas - ESPE durante el primer periodo académico del 2019.

En el Anexo §D se presenta la tabla §D.1, que muestra, la instanciación de la propuesta del *Caso-Req* aplicando la propuesta actualizada del PL orientado al proceso acorde a al ACM. La instanciación del *Caso-Req* ha sido realizado por el experimentador, con la finalidad de encontrar limitaciones a la propuesta.

7.5 EVALUACIÓN DE LA SATISFACCIÓN PARA LA PROPUESTA

En este apartado se describe la evaluación del aspecto de satisfacción de la propuesta del PL orientado al proceso acorde a la ACM. La satisfacción es la capacidad que debe tener la propuesta para satisfacer las diferentes necesidades mínimas de los usuarios al momento de ser utilizada.

7.5.1 Definición de los atributos de Satisfacción

Para definir los atributos de *Satisfacción* se ha estudiado las métricas para la calidad en uso de la ISO/IEC 25022 [3], específicamente las métricas de las subcaracterísticas *Utilidad, Confianza y Comodidad* que son medidas de la característica de *Satisfacción* del usuario. La ISO/IEC 25022 es parte de la familia de la ISO/IEC 25000 definida por SQuaRE (System and Software Quality Requirements and Evaluation) se centra en la especificación, medición y evaluación de los requisitos de calidad de los productos de software. SQuaRE establece un modelo para medir la calidad interna, externa y de uso del software. Los requisitos de calidad de uso especifican el nivel de calidad requerido desde el punto de vista del usuario final; es una validación del producto de software desde las necesidades de un contexto de uso [336].

Para evaluar las métricas se aplicó una Encuesta de Satisfacción a los experimentadores que tienen conocimiento de la instanciación del PL orientado al proceso acorde a la ACM. La encuesta ha sido formulada bajo los parámetros de la herramienta "Escala de Usabilidad del Sistema - (SUS)". SUS consiste en un cuestionario de 10 ítems con cinco opciones de respuesta para los encuestados; de Totalmente de acuerdo a Totalmente en desacuerdo. Creado originalmente por John Brooke en 1986 [66].

Los resultados de la encuesta se analizan para evaluar los atributos de *Satisfacción* de la propuesta y contestar a la pregunta **PI**_{3.3} descrita en el caso de estudio. Los atributos de *Satisfacción* considerados para esta evaluación han sido adoptados del estudio de Quiña [336] y se definen a continuación:

- *Utilidad*: Determina la cantidad de usuarios satisfechos con el uso de la propuesta del PL orientado al proceso acorde a la ACM. Mediante una encuesta se determina el atributo de utilidad.
- *Confianza*: Determina la confianza de los usuarios, mediante la medición de las

quejas presentadas por alguna falla al momento de instanciar la propuesta del PL orientado al proceso acorde a la ACM.

- *Comodidad*: Determina la facilidad y poco esfuerzo en el uso de la propuesta el PL orientado al proceso acorde a la ACM. Mediante una encuesta se determina el atributo de comodidad.

La tabla §7.5 muestra las métricas de los atributos de Satisfacción que fueron evaluados en la encuesta.

Tabla 7.5: Métricas de la calidad en uso para los atributos de Satisfacción

Característica	Subcaracterística	Función de medida	Elementos de medida
Satisfacción	Utilidad	$X = N/T$	N=Número de usuarios satisfechos. T=Número de usuarios encuestados.
	Confianza	$C = A / T,$ $X=1-C$	X=% Quejas. C=% Confianza. A=Número de quejas presentadas. T=Total de número de usuarios encuestados.
	Comodidad	$X=A$	A = Valor de la escala psicométrica de un cuestionario (Escala de Likert).

7.6 RESULTADOS DEL ESTUDIO DE CASO

En esta sección se abordan cada una de las preguntas de investigación propuestas en la sección §7.2.1, resumiendo los resultados de los aspectos de la *Viabilidad*, *Flexibilidad* y *Satisfacción*, una vez instanciada la propuesta del PL orientado al proceso acorde a la ACM.

7.6.1 Resultados de la evaluación de Viabilidad

La instanciación del PL en el *Caso-APIs*, ha sido obtenido de manera favorable durante la aplicación de la propuesta del PL orientado al proceso acorde a la ACM. Se registró limitaciones por parte del experimentador, especificadas en la sección §7.3.3. Encontrar limitaciones en la propuesta permitió actualizar el *metamodelo* e identificar campos necesarios que no habían sido considerados por el autor de esta tesis. EL tiempo que se llevó a cabo para instanciar el PL aplicando la propuesta fue de 20 horas.

La tabla §7.6 muestra los resultados de los atributos de *Viabilidad* obtenidos luego de haber instanciado por primera vez la propuesta del PL:

Tabla 7.6: Resultado de los atributos de *Viabilidad* de la primera instanciación en el *Caso-APIs*

Atributo	Niveles alcanzados
Obtención de la instancia	Si se obtiene un Paquete de Laboratorio orientado a proceso acorde a la ACM.
Esfuerzo de instanciación	20 horas - 1 persona
Aceptación	El experimentador Acepta la instanciación del Paquete de Laboratorio.

7.6.2 Verificación de Completitud del Paquete de Laboratorio en el *Caso-APIs*

La *Completitud* del PL se encuentra especificado en la sección §5.4.3 de la presente memoria de tesis. Durante la instanciación de la propuesta para el *Caso-APIs*, se verificó limitaciones en la disponibilidad de los artefactos con relación a las políticas ACM. Por lo que el PL no cumplió en su totalidad la característica de *Completitud*. La tabla §7.7 muestra el resumen del cumplimiento del PL para alcanzar las insignias de la ACM..

Tabla 7.7: Resultado del cumplimiento del Paquete de Laboratorio para alcanzar las insignias de la ACM en el Caso-APIs

Cumplimiento del Paquete de Laboratorio para alcanzar las insignias de la ACM	
Funcional	38.55%
Disponible	0%
Reutilizable	57.30%
Reproducido	0%
Replicado	0%

7.6.3 Resultados de la evaluación de Flexibilidad

La instanciación del PL en el *Caso-Mind* y el *Caso-Req*, ha sido obtenido de manera favorable durante la aplicación de la propuesta del PL orientado al proceso acorde a la ACM. No se registró limitaciones por parte de los experimentadores, debido a que se utilizó la propuesta actualizada del PL corrigiendo las limitaciones de la primera instanciación en el *Caso-APIs*. EL tiempo que se llevó a cabo para instanciar el PL aplicando la propuesta fue de 16 horas en el *Caso-Mind* y 12 horas para el *Caso-Req*.

La tabla §7.8 muestra los resultados del aspecto de *Flexibilidad* obtenidos luego de haber instanciado en ambos casos la propuesta del PL:

Tabla 7.8: Resultado de los aspectos de flexibilidad de la instanciación en el Caso-Mind y Caso-Req

Atributo	Niveles alcanzados
Diversidad de experimentos	Se evidencia el atributo de diversidad al instanciar tres experimentos con la propuesta del Paquete de Laboratorio orientado al proceso acorde a la ACM.
Uso en replicaciones futuras	Los experimentadores si consideran que la propuesta del Paquete de Laboratorio orientado al proceso acorde a la ACM puede ser utilizada en futuras replicaciones.

7.6.4 Verificación de Completitud del Paquete de Laboratorio en el Caso-Mind y Caso-Req

La *Completitud* del PL se encuentra especificado en la sección §5.4.3 de la presente memoria de tesis. Durante la instanciación de la propuesta para el *Caso-Mind* y el *Caso-*

Req, se verificó pocas limitaciones en la disponibilidad de los artefactos con relación a las políticas ACM. Sin embargo, el PL no cumplió en su totalidad la característica de *Compleitud*. La tabla §7.9 muestra el resumen del cumplimiento del PL para alcanzar las insignias de la ACM.

Tabla 7.9: Resultado del cumplimiento del Paquete de Laboratorio para alcanzar las insignias de la ACM en el Caso-Mind y el Caso-Req

Cumplimiento del Paquete de Laboratorio para alcanzar las insignias de la ACM del Caso-Mind	
Funcional	52.61 %
Disponibile	100 %
Reutilizable	71.36 %
Reproducido	0 %
Replicado	0 %
Cumplimiento del Paquete de Laboratorio para alcanzar las insignias de la ACM del Caso-Req	
Funcional	52.61 %
Disponibile	100 %
Reutilizable	77.61 %
Reproducido	0 %
Replicado	0 %

7.6.5 Resultados de la evaluación de Satisfacción

Para obtener los resultados del aspecto de *Satisfacción*, se realizó un cuestionario que fue contestado por seis experimentadores que conocen la instanciación de la propuesta. El cuestionario esta disponible en el Anexo §E de la presente memoria de tesis. Una vez que se aplicó el cuestionario las respuestas fueron analizadas presentando los siguientes resultados:

- *Utilidad*: Para medir este atributo, se determinó las preguntas *P1*, *P3*, *P6* y *P8* del cuestionario. Se analizó los datos obtenidos en el cuestionario, y se utilizó los parámetros del modelo de calidad en uso de la ISO/IEC 25010 [2] para definir los pesos de medida de los atributos *Utilidad*, *Confianza* y *Comodidad*. Estos pesos están relacionados al contexto de uso y aplicación en este caso de la propuesta del PL orientado al proceso acorde a la ACM. Los cálculos obtenidos durante el análisis presenta como resultado el 88,5% de *Utilidad* para la propuesta.
- *Confianza*: Para medir este atributo, se determinó las preguntas *P5* y *P9* del cuestionario. Se analizó los datos obtenidos en el cuestionario y se preguntó a los

experimentadores si tienen quejas durante la instanciación de la propuesta. Los cálculos obtenidos durante el análisis presenta como resultado el 92% de *Confianza* para la propuesta y no se registra ninguna queja por parte de los experimentadores.

- *Comodidad*: Para medir este atributo, se determinó las preguntas *P2, P4, P7 y P10* del cuestionario. Los cálculos obtenidos durante el análisis presenta como resultado el 71,20% de *Comodidad* para la propuesta.

Luego de obtener los resultados de cada atributo *Utilidad, Confianza y Comodidad* se hizo el análisis de resultados en conjunto, obteniendo el 86,09% de *Satisfacción* para la propuesta. La tabla §7.10 muestra el resumen de la evaluación del aspecto de *Satisfacción*.

Tabla 7.10: Resultados de los atributos de Satisfacción

Característica	Subcaracterística	Peso de medida	Σ Resultado	% Resultado	Total*
Satisfacción*	Utilidad	50%	0,88	44,25%	86,09%
	Confianza	30%	0,92	27,6%	
	Comodidad	20%	0,71	14,24%	

7.7 CONCLUSIONES

El estudio de caso ha permitido la evaluación y evolución de la propuesta mediante su aplicación a casos reales de experimentos controlados con humanos en el área de IS. A continuación, se destaca la principal aportación de cada uno de los casos.

Caso-APIs Ha permitido detectar la falta de información relevante, así como la necesidad de mejorar la redacción y aumentar y/o eliminar campos que se han propuesto en el *metamodelo*. Una vez incorporadas las mejoras pertinentes, la propuesta ha evolucionado hasta su versión actual. Con la instanciación por primera vez de la propuesta se logró establecer el aspecto de *Viabilidad*.

Caso-Mind y Caso-Req Ha permitido instanciar de manera favorable la propuesta en diferentes experimentos, por lo que ha reforzado la idea de llevar a cabo un PL orientado al proceso acorde a la ACM. Con la instanciación de los dos casos se ha evaluado el aspecto de *Flexibilidad*.

Caso-APIs , **Caso-Mind** y **Caso-Req** Ha servido para confirmar el aspecto de *Satisfacción* de la propuesta por parte de los experimentadores mediante un cuestionario.

Cabe señalar, que la propuesta ha demostrado su utilidad al especificar la teoría desarrollada: el *procedimiento*, el *modelo de calidad* y el *metamodelo* del PL orientado al proceso acorde a la ACM aplicado al *estudio de caso*, sin embargo es importante extender la aplicación de la propuesta a más experimentos.

La instanciación de la propuesta en el estudio de caso ha permitido clarificar el proceso de automatización del PL orientado al proceso acorde a la ACM, mediante el desarrollo de una herramienta (Prototipo inicial).

Como conclusión, se recomienda el uso de la propuesta, por un lado: *i)* para elaborar un PL de calidad durante la ejecución del experimento y por otro lado: *ii)* alcanzar las insignias de la ACM en lo que se refiere a la evaluación de los artefactos.

7.8 RESUMEN

Para evaluar la propuesta, se ha instanciado mediante un *estudio de caso* los siguientes experimentos controlados con humanos en el área de IS: *i) Caso-APIs*; *ii) Caso-Mind*; y *iii) Caso-Req*.

Se han analizado los siguientes aspectos de la propuesta:

Viabilidad. Se han modificado algunos campos del *metamodelo* para expresar lo que el experimentador a sugerido incorporar en el PL propuesto. Una vez realizado la actualización de las observaciones y ajustes oportunos en el paquete, se ha podido establecer el aspecto de *Viabilidad* de la propuesta. Se ha logrado obtener una instancia favorable a partir de la información recogida, con un esfuerzo razonable medido en 20 horas- 1 persona y la aceptación del experimentador durante la aplicación de la instancia en el *Caso-APIs*.

Flexibilidad. La instanciación del *Caso-Mind* y el *Caso-Req* han permitido establecer de manera favorable la elaboración del PL orientado al proceso acorde a la ACM en experimentos con contextos diferentes, a pesar de estar relacionados con experimentos controlados con humanos en el área de IS cada experimento se presenta diferencias y aspectos con ciertas particularidades. La instanciación de ambos casos ha logrado considerar por parte de los experimentadores la aplicación de la propuesta en repeticiones futuras.

Satisfacción. Luego de instanciar los experimentos del estudio de caso, se ha logrado evaluar el aspecto de *Satisfacción* mediante un cuestionario aplicado a los experimentadores, específicamente se ha evaluado los atributos de *Utilidad, Confianza y Comodidad* de manera favorable.

Aparte de la evaluación mediante el estudio de caso, y como complemento, se ha implementado la herramienta badgeGo (ver sección ??), como prueba de concepto de la corrección del metamodelo y su implementación mediante la propuesta.

Más adelante, la intención es seguir profundizando para que la herramienta badgeGo sirva como asistente a los experimentadores noveles, sobre todo, para automatizar el procedimiento de la elaboración del PL de calidad y alcanzar las insignias de la ACM con respecto a la evaluación de los artefactos. De esta forma, contribuir al fomento de las replicaciones experimentales en el área de IS.

PARTE V

CONCLUSIONES

CONCLUSIONES Y TRABAJOS FUTUROS

Cualquier momento es perfecto para aprender algo nuevo

Albert Einstein (1879 – 1955),

Físico alemán

En este capítulo se concluye la presente memoria de tesis, dando a conocer una visión general del trabajo realizado mediante el análisis de los objetivos y las conclusiones.

En la sección §8.1, se analiza el cumplimiento de los objetivos propuestos. En la sección §8.2, se resumen las publicaciones asociadas a este trabajo. Finalmente, en la sección §8.3, se determinan las líneas de trabajo futuro.

8.1 CONSECUCIÓN DE OBJETIVOS

Al principio de la memoria (sección §1.3) se planteó el principal objetivo de este trabajo de **fomentar las replicaciones de experimentos controlados con humanos mediante la elaboración de un Paquete de Laboratorio orientado al proceso acorde a la ACM**. En esta sección se analiza el cumplimiento de dicho objetivo principal mediante la contestación a las preguntas de investigación propuestas en esta tesis.

1. *Estudiar el estado de los Paquete de Laboratorio de experimentos en el área de IS y evaluar la calidad de su contenido y como han evolucionando durante los últimos años.* En el capítulo §3 se han analizado estudios empíricos que reportan experimentos controlados con humanos en el área de IS publicados entre 2000 y 2021. El estudio ha permitido:
 - i) Identificar los principales foros para conocer las revistas y congresos donde publicar futuros estudios de PL o incluyen un Track AE-Artifact Evaluation.
 - ii) Conocer las insignias de la ACM que adquieren los estudios y PL en dichos foros.
 - iii) Detectar vacíos de investigación. Al analizar la calidad del contenido de los PL y la manera como son publicados, se ha identificado una diversidad en la forma de elaborar y reportar el contenido del PL. Considerando que aún existe un bajo interés en llevar a cabo replicaciones en el área de IS, es necesario fomentar dicho interés hacia la comunidad científica.

2. *Principales problemas identificados en la elaboración y publicación de los Paquete de Laboratorio.* Algunos de los problemas detectados en el capítulo §4, se pueden paliar en parte mediante la propuesta descrita en la presente memoria de tesis. De esta forma:
 - i) El *Poco uso de guías para reportar los Paquete de Laboratorio* puede mitigarse con el uso de la propuesta, como una alternativa para el experimentador en el proceso de elaboración del PL mediante una guía paso a paso.
 - ii) La *Carencia y debilidades de los Paquete de Laboratorio* puede mitigarse mediante la especificación del proceso experimental de forma integrada con el PL, es decir, durante la fase de ejecución del experimento se registre las tareas junto con los artefactos que se van utilizando, para generar de alguna manera un buen PL. La elaboración de un PL orientado al proceso como se propone en esta memoria de

tesis permite mitigar el esfuerzo, tiempo, coste y recursos que los experimentadores deben realizar al momento de elaborar el PL específicamente cuando luego de haber finalizado con la ejecución y reporte del experimento o la replicación.

iii) Ante la *Dificultad para clarificar el contexto de los experimentos y las replicaciones*, la propuesta motiva a que los experimentadores registren información sobre el contexto, es decir, clarificar las circunstancias del ¿Dónde? y ¿Cómo? se lleva a cabo el experimento, además, ¿Quién? o ¿Quiénes? son la comunidad interesada en el estudio. Esto permite, comprender de mejor manera el comportamiento del fenómeno en cuál ocurre el estudio.

3. *Identificar cómo se reportan actualmente los Paquete de Laboratorio*. En el capítulo §3 se han identificado aspectos del PL que han sido analizados como: la estructura y contenido, las guías y repositorios digitales que se usa para el reporte y la evolución en el tiempo sobre la publicación de los PL. El conjunto de Datasets es de 255 estudios, publicados en las revistas (EMSE, IST, JSS y TSE) y los congresos que tienen el Track AE-Artifact Evaluation (ICSE '2021, ESEC/FSE'2021, RE'2021, ASE'2021). Los principales hallazgos han sido:

i) La mayoría de los estudios analizados (36,45%) reportan el contenido relacionado al módulo *experimento*. Por el contrario, los módulos de *agregación, replicación y evolución* se reportan menos del (2%).

ii) Las insignias de la ACM para las revistas, en su mayoría alcanzan los PL son *disponible* con el (88,42%), seguido de la insignia *funcional* en (29,47%). Si embargo, se evidencia que el porcentaje de estudios con las insignias *reproducido y replicado* apenas es del (3,15%).

iii) El análisis de los estudios ha evidenciado que ninguno sigue una guía para reportar los PL y únicamente existe directrices generales que los congresos publican en el Track AE-Artifact Evaluation. Es evidente que la manera de reportar los paquetes no sigue una formalidad definida por tanto la heterogeneidad del contenido y estructura del paquete es variada.

iv) Los repositorios que se usan con mayor frecuencia para reportar los PL son *GitHub* (33%), *Repositorios institucionales* (28%) y *Zenodo* (14%). En el año 2020 se presenta un aumento significativo (29,62%) de experimentos controlados que publican un PL en el área de IS, la tendencia para el 2021 va en crecimiento. Con la motivación del movimiento *Open Science* en estos dos últimos años y acompañados de la emergencia sanitaria del Covid-19 las políticas de acceso abierto y

divulgación científica seguramente marcaran una subida en el uso de los repositorios digitales.

4. *Elaborar un Paquete de Laboratorio orientado al proceso acorde a la ACM.* En el capítulo §5 se establece el enfoque de la propuesta de la presente memoria de tesis y que consta de cuatro partes: *i)* una especificación del proceso experimental de forma integrada con el PL, *ii)* un modelo de calidad para Paquete de Laboratorio acorde a la ACM, *iii)* un metamodelo del Paquete de Laboratorio orientado al proceso acorde a la ACM y *iv)* una herramienta para automatizar el metamodelo.

El objetivo principal de la propuesta es fomentar las replicaciones de experimentos controlados con humanos mediante la elaboración de un PL orientado al proceso. La propuesta tiene doble finalidad, por un lado, permite al experimentador elaborar un PL bien estructurado y documentado, y por otro, presenta una alta probabilidad de adquirir una insignia de la ACM.

Basada en la propuesta y como prueba de concepto de la corrección del *metamodelo*, se ha desarrollado la herramienta badgeGo que permite a los experimentadores elaborar un PL orientado al proceso acorde a las insignias de la ACM.

5. *Instanciar la propuesta.* En el capítulo §7 se ha evaluado la propuesta mediante un *estudio de caso* que abarca tres experimentos controlados *Caso-APIs*, *Caso-Mind* y *Caso-Req* en el área de IS.

Se ha identificado limitaciones y observaciones con respecto algunos campos faltantes y terminología variada. La instanciación ha sido de ayuda para actualizar el *metamodelo* y por consiguiente la propuesta.

La propuesta se ha evaluado, mediante el análisis de los siguientes aspectos:

i) Viabilidad. Se han modificado algunos campos en el *metamodelo* para expresar lo que el experimentador necesita dejar reflejado. Una vez realizados los ajustes oportunos, se han podido instanciar la propuesta en el *estudio de caso*.

Algunos campos del *metamodelo* han sido más difíciles de cumplimentar por razones de terminología o posiblemente cuando la propuesta es cumplimentada por alguien que no es el autor del experimento.

ii) Flexibilidad. La instanciación de la propuesta en más contextos ha permitido observar que el procedimiento del PL se lleve a cabo sin dificultad. Es evidente que instanciar la propuesta en más experimentos siempre ayudará a confirmar la

Flexibilidad, significa que siempre se encontrará mejoras en la propuesta, esto es algo particular de este aspecto.

iii) *Satisfacción*. Al finalizar las instanciaciones del *estudio de caso*, se ha aplicado un cuestionario de satisfacción a los experimentadores que utilizaron la propuesta. Los atributos que se evaluaron fueron *Utilidad, Confianza y Comodidad*. Se establece de manera favorable la evaluación de este aspecto por parte de los experimentadores.

Aparte de la evaluación de la propuesta mediante el estudio de caso, el desarrollo de badgeGo, ha permitido identificar campos que eran de difícil cumplimentación en el proceso de elaboración del PL y que han sido automatizados para facilidad del experimentador.

8.2 PUBLICACIONES

Los resultados parciales obtenidos durante la realización del presente trabajo de investigación han dado lugar a las publicaciones que se enumeran a continuación.

8.2.1 Revistas internacionales

- C. Guevara-Vega, B. Bernárdez, A. Durán Toro, M. Cruz, A. Ruiz-Cortés. *Laboratory Packages and ACM Badging for Open Science. A Systematic Mapping Study*. IEEE Access. ISSN: 21693536, 2022. Índice de impacto: 0,93 (SJR: Q1). 3,476 (JCR: Q2). [En revisión].

En este trabajo se presenta la *revisión sistemática* llevada a cabo para conocer el estado actual de los PL de experimentos controlados con humanos en el área de IS e identificar los nichos de investigación para trabajos futuros.

- F. Uyaguari, C. Guevara-Vega, A. Quiña, C. Acosta y A. Uyaguari. *Effect of requirements specification using native language on external software quality*. IET Software. ISSN: 17518806, 2022. Índice de impacto: 0,31 (SJR: Q3). 1,40 (JCR: Q4). [En revisión].

En este trabajo se presenta una comparativa entre los requisitos escritos en lengua materna con los escritos en lengua extranjera y analiza el impacto en la calidad externa del desarrollo de software, mediante un experimento controlado con humanos en el área de IS.

- M. Cruz, B. Bernárdez, A. Durán, C. Guevara-Vega, y A. Ruiz-Cortés. *A Model-Based Approach for Specifying Changes in Replications of Empirical Studies in Computer Science*. El artículo científico está disponible en arxiv.org [104]. Enviado a Computing journal. ISSN: 0010485X, 14365057. 2022. Índice de impacto: 0,68 (SJR: Q2). 2,420 (JCR: Q2). [En revisión].

En este trabajo se presenta un enfoque sistemático y apoyado por herramientas para la especificación y el informe de los cambios en las replicaciones de estudios empíricos en Ciencias de la Computación.

- El *estudio de caso* aquí presentado que ha permitido la evaluación de la propuesta, será próximamente enviado a una revista científica.

8.2.2 Congresos internacionales

- [171] C. Guevara. *Publicación en Open Science de paquetes de laboratorio experimentales para Ingeniería del Software*. Simposio Doctoral del XXIV Conferencia Iberoamericana de Ingeniería de Software 2021 (CibSE). Costa Rica, Agosto-Septiembre 2021. Certificado de presentación.

- [172] C. Guevara, B. Bernárdez, A. Durán, A. Quiña, M. Cruz, y A. Ruiz-Cortés. *Empirical Strategies in Software Engineering Research: A Literature Survey*. Actas de: Second International Conference on Information Systems and Software Technologies 2021 (ICI2ST). Ecuador, Marzo 2021. IEEE Xplore. DOI:10.1109/ICI2ST51859.2021.00025

Se establece una visión general y se identifica el estado actual de las estrategias empíricas que se utilizan en el área de IS. Además, se propone una lista de características para fomentar la selección adecuada durante la concepción de una investigación empírica en el área de IS.

- Ponencia en el VI Congreso Internacional de Investigación Científica 2022, con el Tema: "*Estrategias Empíricas en la Ingeniería del Software*". La Maná, Ecuador. Certificado de ponencia.

8.2.3 Participación en proyecto de investigación

- Directora del proyecto de investigación interno de la Facultad de Ingeniería en Ciencias Aplicadas, Universidad Técnica del Norte. *Mejoras en la automatización*

del proceso experimental en Ingeniería Informática (Primera Fase). Duración 12 meses. Integrantes: C. Guevara, B. Bernárdez, A. Durán, A. y A. Ruiz-Cortés. Fecha inicio: 01 enero 2019, Fecha fin: 31-diciembre-2019. Aprobado con resolución HCD Nro. UTN-FICA-2020-023.

- Directora del proyecto de investigación interno de la Facultad de Ingeniería en Ciencias Aplicadas, Universidad Técnica del Norte. *Mejoras en la automatización del proceso experimental en Ingeniería Informática (Segunda Fase)*. Duración 12 meses. Integrantes: C. Guevara, B. Bernárdez, A. Durán, A. Quiña, M. Cruz, y A. Ruiz-Cortés. Fecha inicio: 17 enero 2020, Fecha fin: 17-enero-2021. Aprobado con resolución HCD Nro.UTN-FICA-2021-03-011.
- Directora del proyecto de investigación interno de la Facultad de Ingeniería en Ciencias Aplicadas, Universidad Técnica del Norte. *Mejoras en la automatización del proceso experimental en Ingeniería Informática (Tercera Fase)*. Duración 12 meses. Integrantes: C. Guevara, B. Bernárdez, A. Durán, A. Quiña, M. Cruz, y A. Ruiz-Cortés. Fecha inicio: 01 febrero 2021, Fecha fin: en proceso de cierre.

8.3 TRABAJO FUTURO

Una vez finalizado el presente estudio de investigación, ha surgido nuevas ideas y a su vez nuevas líneas de investigación que deberán incorporarse como un trabajo futuro a mediano y largo plazo. Entre las cuáles se establecen las siguientes:

- Ampliar la *revisión de la literatura* realizada en el presente estudio mediante una encuesta dirigida a los investigadores para conocer los aspectos que limita la realización de replicaciones y la publicación de PL.
- Profundizar en el análisis de definir un peso para los artefactos *opcionales* de las insignias de la ACM, y considerar el impacto e influencia que tendrán en la calidad del PL.
- Incrementar la *instanciación* de la propuesta en otros experimentos controlados con diferente contexto para ampliar los aspectos de *Viabilidad* y *Flexibilidad* de la propuesta.
- Incrementar la *instanciación* de la propuesta en experimentos controlados que estén iniciando con el proceso experimental, de manera que la elaboración del PL orientado a proceso se lleve a cabo durante la ejecución del experimento.

- Continuar con la *automatización* de la propuesta mediante el desarrollo de badgeGo, que permita al experimentador elaborar el PL orientado al proceso desde la fase de inicio del experimento, para disminuir el trabajo extra de elaborar el PL cuando ya está el artículo escrito. Similar a un cuaderno de bitácora. Y registrando cada paso del proceso. Además, tenga la posibilidad de conocer la insignia ACM que probablemente obtenga su PL.
- Implementar en badgeGo, la organización de artefactos y colaboración entre experimentadores especialmente cuando los datos necesitan ser anonimizados y re-identificados, si no se ha hecho antes de realizar la limpieza o análisis de los mismos. Por lo regular, los datos y análisis están segmentados/repartidos entre varios experimentadores y organizar los archivos no es simple.
- Evaluar la herramienta badgeGo para establecer su *aplicabilidad* y uso en la comunidad científica, específicamente para los investigadores que deseen enviar su PL a los congresos que presenten el Track AE-Artifact Evaluation.
- Pensar a largo plazo incrementar la propuesta hacia otro tipo de experimentos controlados como los *computacionales*, debido al aumento de la tendencia en evaluar los PL en los Track AE-Artifact Evaluation de los congresos.

Se considera que la presente lista de trabajos futuros permitirá establecer mejoras en el proceso de elaboración de los PL, a la vez, fomentar las replicaciones de experimentos en el área de IS. Además, establecen un horizonte de posibles temas de investigación que aportarán al cuerpo del conocimiento en el área de la IS.

PARTE VI

ANEXOS

DESCRIPCIÓN DEL REPORTE

NISO RP-31—2021

En este anexo se muestra la descripción de las insignias que se reportan en la NISO (National Information Standards Organization) y la ACM (Association for Computing Machinery), con el fin de diferenciar su aplicación en la evaluación de artefactos para la reproducibilidad.

El reporte de la NISO para Insignias y Definiciones de Reproducibilidad es un esfuerzo para desarrollar prácticas de reconocimiento, vocabulario e iconografía comunes que faciliten el intercambio de datos y métodos científicos. La reproducibilidad es tanto un proceso como un fin. Es fundamental definir los distintos niveles de reproducibilidad y acordar un sistema de insignias estandarizado que pueda aplicarse en el proceso de publicación, que al momento quizá pueda utilizarse como moneda de cambio en la cultura de las recompensas académicas.

Las revistas y congresos académicos utilizan varios sistemas de insignias para indicar la reproducibilidad, la apertura y la condición de artefacto, lo que indica un nivel general de alto interés en la actividad científica en las ciencias.

Las insignias que se identifican en la NISO son: i) ROR – Research Objects Reviewed, ii) ORO - Open Research Objects, iii) ROR-R - Results Reproduced; iv) RER - Results Replicated. Sus acrónimos están definidas en el idioma inglés.

En la tabla §A.1 se muestra las insignias definidas por la NISO, que son utilizadas para evaluar los artefactos y definir la reproducibilidad de estudios científicos.

En la tabla §A.2 se muestra las insignias definidas en la ACM que son utilizadas para el mismo fin que la NISO.

En ambas tablas §A.1 y §A.2 se puede observar las características que comparten cada una de las insignias. No se observa mayor diferencia en la descripción de cada una de las insignias NISO y ACM, sin embargo, quizá la diferencia está en el campo de acción donde se aplican. Los congresos y revistas del área de IS actualmente utilizan en mayor parte las insignias de la ACM, que trabaja directamente con investigadores y personas del área de Ciencias de la Computación para evaluar los artefactos.

Tabla A.1: Insignias definidas por la NISO para evaluar artefactos

Objeto de Investigación Revisado (ROR) - Reutilizable	Objeto de Investigación Abierto (ORO)	Resultados Reproducidos (ROR-R)	Resultados Replicados (RER)
Funcional + documentado con mucho cuidado y bien estructurado en la medida en que se facilita la reutilización y la adaptación. En particular, se cumplen estrictamente las normas y estándares de la comunidad de investigadores para artefactos de este tipo.	Funcional + colocado en un repositorio de archivo de acceso público. Se proporciona una URL estable al repositorio y un DOI (Digital Object Identifier). Esto coincide con el distintivo "Disponible" de la ACM.	Los principales resultados de ROR + ORO + del artículo se han obtenido en un estudio posterior por una persona o equipo distinto a los autores, utilizando, en parte, artefactos proporcionados por el autor.	ROR + ORO + los principales resultados del artículo han sido obtenidos de forma independiente en un estudio posterior por una persona o equipo distinto a los autores, sin el uso de artefactos proporcionados por el autor.

Tabla A.2: Insignias definidas por la ACM para evaluar artefactos

Artefactos Evaluados		Artefactos Disponibles	Resultados Validados	
Funcional	Reutilizable	Disponible	Reproducido	Replicado
				
Artefactos documentados, consistentes, completos, ejercitables e incluyen evidencia apropiada de verificación y validación.	Funcional + Los artefactos asociados al documento son de una calidad que excede significativamente la funcionalidad mínima. Es decir, tienen todas las cualidades del nivel de Artefactos Evaluados-Funcionales, pero, además, están muy cuidadosamente documentados y bien estructurados hasta el punto de facilitar su reutilización y adaptación. En particular, se respetan estrictamente las normas y estándares de la comunidad investigadora para este tipo de artefactos.	Los artefactos creados por el autor que son relevantes para el artículo se han colocado en un repositorio de archivos de acceso público de forma permanente a través de un repositorio con una URL estable y un DOI (Digital Object Identifier).	Los principales resultados del artículo se han obtenido en un estudio posterior realizado por una persona o equipo distinto de los autores, utilizando, en parte, artefactos proporcionados por el autor.	Los principales resultados del artículo han sido obtenidos de forma independiente en un estudio posterior por una persona o equipo distinto de los autores, sin utilizar los artefactos suministrados por éstos.

INSTANCIACIÓN LA PROPUESTA EN EL CASO-APIs

En este anexo se muestra la instanciación por primera vez de la propuesta Paquete de Laboratorio orientado al proceso acorde a la ACM sobre el *Caso-APIs*

Tabla B.1: Instanciación de la propuesta en el experimento Caso-APIs

Primer paso	Información, objetivos y descripción del experimento
Campo	Valor
Acrónimo	APIs-2021
Nombre del experimento	Efectos de los paradigmas REST y GraphQL en el desarrollo del software.
DOI del experimento	No se registra
Lugar	Carrera de Software, Universidad Técnica del Norte, Ibarra-Ecuador
Fecha	Segundo semestre del curso, periodo abril-agosto 2021
Objetivo	Analizar el proceso de desarrollo del software. Con el propósito de comparar efectos de REST y GraphQL. Con respecto a la calidad del software. Desde el punto de vista del investigador. En el contexto académico.
Descripción	Un grupo de estudiantes de Grado de la carrera de Software de la Universidad Técnica del Norte, realizó dos sesiones de entrenamiento de REST y GraphQL (3 horas). Se evaluó el conocimiento por medio de deberes (8 horas) y un workshop (2 horas). Los datos recogidos a través de un cuestionario inicial y los conocimientos adquiridos en el entrenamiento sirvió para la selección de los participantes. En la ejecución del experimento hubo dos sesiones de tareas de desarrollo en REST y GraphQL. Se les facilitó material experimental, un diagrama UML de la Base de datos (BDD), los requisitos funcionales y el script de la BDD. Se comparó el rendimiento de los participantes al desarrollar las tareas en REST y GraphQL.
Segundo paso	Registro de experimentadores
Campo	Valor
DNI	1002322384
Nombre del experimentador	Antonio Quiña
email	aquina@utn.edu.ec
Teléfono	<i>No es necesario registrar</i>
Afiliación	Universidad Técnica del Norte
Sitio web	https://bit.ly/3Jx88T3
Fecha de nacimiento	<i>No es necesario registrar</i>
Género	Masculino
Nacionalidad	Ecuatoriana
Tipo de persona	Profesional
Años de experiencia	<i>No se especifica a que tipo de experiencia</i>
Rol del experimentador	Formador, Diseñador, Analista
Comentario	Algunos campos no son necesarios
DNI	1004100903
Nombre del experimentador	Fernando Chulde
email	jfchuldeb@utn.edu.ec
Teléfono	<i>No es necesario registrar</i>
Afiliación	Universidad Técnica del Norte
Sitio web	No se registra
Fecha de nacimiento	<i>No es necesario registrar</i>
Género	Masculino
Nacionalidad	Ecuatoriana
Tipo de persona	Graduado
Años de experiencia	<i>No es necesario registrar</i>
Rol del experimentador	Medidor, Monitor
Comentario	Algunos campos no son necesarios

DNI	1002334835
Nombre del experimentador	Cathy Guevara
email	cguevara@utn.edu.ec
Teléfono	<i>No es necesario registrar</i>
Afiliación	Universidad Técnica del Norte
Sitio web	https://bit.ly/3twx87y
Fecha de nacimiento	<i>No es necesario registrar</i>
Género	Femenino
Nacionalidad	Ecuatoriana
Tipo de persona	Profesional
Años de experiencia	<i>No es necesario registrar</i>
Rol del experimentador	Diseñador, Monitor
Comentario	Algunos campos no son necesarios
Tercer paso	Grupos y asignación de participantes
Campo	Valor
<i>Se debería especificar el número de grupos y sus participantes</i>	
Cuarto paso	Registro de tareas del experimento y sus artefactos
Campo	Valor
Acrónimo	T01
Nombre de la tarea	Encuesta demográfica/diagnóstica
Tipo de tarea	Preparación
Fecha de inicio	19 de junio 2021
Fecha de fin	19 de junio 2021
Duración	4 minutos
Descripción	Se realiza una encuesta demográfica de los participantes antes de ejecutar el experimento, para diagnosticar el conocimiento y percepción de la facilidad de uso, consumo y aprendizaje de los paradigmas REST y GraphQL.
Artefacto de la T01	
Nombre del artefacto	encuesta-apis
Tipo de finalidad	Reclutamiento
Tipo de artefacto	<i>Debería ser una Encuesta</i>
Tipo de formato	PDF
Tamaño	90 KB
Clase de artefacto	Entrada
Path	https://bit.ly/3to1QiY
Fecha de creación	20 mayo 2021
Fecha de actualización	<i>No se especifica si es la última actualización</i>
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	4 minutos
Tiempo corto de ejecución	2 minutos
Comentario	No se registra
Acrónimo	T02
Nombre de la tarea	Introducción teórica del ambiente de desarrollo
Tipo de tarea	Preparación
Fecha de inicio	20 de junio 2021
Fecha de fin	20 de junio 2021
Duración	1 hora
Descripción	Presentación teórica de arquitecturas orientada a microservicios y tecnologías usadas en el ambiente de desarrollo.

Artefacto de la T02	
Nombre del artefacto	01 introducción-ambiente-desarrollo
Tipo de finalidad	<i>Debería ser de Formación</i>
Tipo de artefacto	Presentación
Tipo de formato	.pptx
Tamaño	300 KB
Clase de artefacto	Entrada
Path	Directorio interno del paquete
Fecha de creación	01 de abril 2021
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	No se registra
Tiempo corto de ejecución	No se registra
Comentario	No se registra
Acrónimo	T03
Nombre de la tarea	Introducción teórica al paradigma REST
Tipo de tarea	Preparación
Fecha de inicio	21 de junio 2021
Fecha de fin	22 de junio 2021
Duración	4 horas
Descripción	Presentación teórica del estilo arquitectónico REST que sirve para el desarrollo de APIs REST.
Artefacto de la T03	
Nombre del artefacto	02 introducción-rest
Tipo de finalidad	<i>Debería ser de Formación</i>
Tipo de artefacto	Presentación
Tipo de formato	.pptx
Tamaño	432 KB
Clase de artefacto	Entrada
Path	Directorio interno del paquete
Fecha de creación	03 abril 2021
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	No se registra
Tiempo corto de ejecución	No se registra
Comentario	No se registra
Acrónimo	T04
Nombre de la tarea	Introducción teórica al paradigma GraphQL
Tipo de tarea	Preparación
Fecha de inicio	23 de junio 2021
Fecha de fin	24 de junio 2021
Duración	4 horas
Descripción	Presentación teórica del estilo arquitectónico GraphQL que sirve para el desarrollo de APIs GraphQL.

Artefacto de la T04

Nombre del artefacto	03 introducción-graphql
Tipo de finalidad	<i>Debería ser de Formación</i>
Tipo de artefacto	Presentación
Tipo de formato	.pptx
Tamaño	535 KB
Clase de artefacto	Entrada
Path	Directorio interno del paquete
Fecha de creación	05 de abril 2021
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	No se registra
Tiempo corto de ejecución	No se registra
Comentario	No se registra

Acrónimo	T05
Nombre de la tarea	Requisitos: Gestión de Pizzas
Tipo de tarea	Preparación
Fecha de inicio	28 de junio 2021
Fecha de fin	28 de junio 2021
Duración	No se registra
Descripción	Especifican requisitos de proyectos de desarrollo de APIs

Artefacto de la T05

Nombre del artefacto	pizza
Tipo de finalidad	Requisitos
Tipo de artefacto	Archivo binario
Tipo de formato	.backup
Tamaño	8 KB
Clase de artefacto	Entrada
Path	Directorio interno del paquete
Fecha de creación	7 de abril 2021
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	15 minutos
Tiempo corto de ejecución	10 minutos
Comentario	Es una copia de seguridad de la base de datos Pizza

Nombre del artefacto	pizza
Tipo de finalidad	Requisitos
Tipo de artefacto	Archivo de código
Tipo de formato	.sql
Tamaño	2 KB
Clase de artefacto	Entrada
Path	Directorio interno del paquete
Fecha de creación	8 de abril 2021
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	15 minutos
Tiempo corto de ejecución	10 minutos
Comentario	Script de la estructura de la BDD y datos de prueba

Nombre del artefacto	pizza-diagrama
Tipo de finalidad	Requisitos
Tipo de artefacto	Diagrama
Tipo de formato	.png
Tamaño	8 KB
Clase de artefacto	Entrada
Path	Directorio interno del paquete
Fecha de creación	10 de abril 2021
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	No se registra
Tiempo corto de ejecución	No se registra
Comentario	Diagrama relacional de la estructura de la base de datos Pizza
Acrónimo	T06
Nombre de la tarea	Requisitos: Gestión de Libros
Tipo de tarea	Preparación
Fecha de inicio	28 de junio 2021
Fecha de fin	28 de junio 2021
Duración	No se registra
Descripción	Requisitos de operaciones básicas CRUD, y diseño de base de datos para administrar Libros.
Artefacto de la T06	
Nombre del artefacto	libros
Tipo de finalidad	Requisitos
Tipo de artefacto	Archivo binario
Tipo de formato	.backup
Tamaño	8 KB
Clase de artefacto	Entrada
Path	Directorio interno del paquete
Fecha de creación	11 de abril 2021
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	15 minutos
Tiempo corto de ejecución	10 minutos
Comentario	Es una copia de seguridad de la base de datos Libros
Nombre del artefacto	libros
Tipo de finalidad	Requisitos
Tipo de artefacto	Archivo de código
Tipo de formato	.sql
Tamaño	2 KB
Clase de artefacto	Entrada
Path	Directorio interno del paquete
Fecha de creación	11 de abril 2021
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	15 minutos
Tiempo corto de ejecución	10 minutos
Comentario	Script de la estructura de la BDD y datos de prueba

Nombre del artefacto	libros-diagrama
Tipo de finalidad	Requisitos
Tipo de artefacto	Diagrama
Tipo de formato	.png
Tamaño	10 KB
Clase de artefacto	Entrada
Path	Directorio interno del paquete
Fecha de creación	12 de abril 2021
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	No se registra
Tiempo corto de ejecución	No se registra
Comentario	Diagrama relacional de la estructura de la base de datos Libros
Nombre del artefacto	libros-requisitos-registro
Tipo de finalidad	Requisitos
Tipo de artefacto	Documento
Tipo de formato	.xlsm
Tamaño	41 KB
Clase de artefacto	Entrada
Path	Directorio interno del paquete
Fecha de creación	12 de abril 2021
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	1 hora
Tiempo corto de ejecución	0:45 minutos
Comentario	Hoja de cálculo que contiene los requisitos para el desarrollo de un API, consta de diez operaciones de datos CRUD. Además, contiene una sección para registrar la hora de inicio, fin, y el porcentaje de solución de cada requisito.
Acrónimo	T07
Nombre de la tarea	Requisitos: Gestión de Blogs
Tipo de tarea	Preparación
Fecha de inicio	28 de junio 2021
Fecha de fin	28 de junio 2021
Duración	No se registra
Descripción	Requisitos de operaciones básicas CRUD, y diseño de base de datos para administrar Blogs
Artefacto de la T07	
Nombre del artefacto	blogs
Tipo de finalidad	Requisitos
Tipo de artefacto	Archivo binario
Tipo de formato	.backup
Tamaño	14 KB
Clase de artefacto	Entrada
Path	Directorio interno del paquete
Fecha de creación	13 de abril 2021
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	15 minutos
Tiempo corto de ejecución	10 minutos
Comentario	Es una copia de seguridad de la base de datos Blogs

Nombre del artefacto	blogs
Tipo de finalidad	Requisitos
Tipo de artefacto	Archivo de código
Tipo de formato	.sql
Tamaño	5 KB
Clase de artefacto	Entrada
Path	Directorio interno del paquete
Fecha de creación	13 de abril 2021
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	15 minutos
Tiempo corto de ejecución	10 minutos
Comentario	Script de la estructura de la BDD y datos de prueba
Nombre del artefacto	blogs-diagrama
Tipo de finalidad	Requisitos
Tipo de artefacto	Diagrama
Tipo de formato	.png
Tamaño	21 KB
Clase de artefacto	Entrada
Path	Directorio interno del paquete
Fecha de creación	14 de abril 2021
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	No se registra
Tiempo corto de ejecución	No se registra
Comentario	Diagrama relacional de la estructura de la base de datos Blogs
Nombre del artefacto	blogs-requisitos-registro
Tipo de finalidad	Requisitos
Tipo de artefacto	Documento
Tipo de formato	.xlsm
Tamaño	41 KB
Clase de artefacto	Entrada
Path	Directorio interno del paquete
Fecha de creación	14 de abril 2021
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	1 hora
Tiempo corto de ejecución	0:45 minutos
Comentario	Contiene los requisitos para el desarrollo de un API, consta de 10 operaciones de datos CRUD. Contiene una sección para registrar la hora de inicio, fin, y el porcentaje de solución de cada requisito.
Acrónimo	T08
Nombre de la tarea	Requisitos: Gestión de Eventos
Tipo de tarea	Experimental
Fecha de inicio	28 de junio 2021
Fecha de fin	28 de junio 2021
Duración	No se registra
Descripción	Requisitos de operaciones básicas CRUD, y diseño de base de datos para administrar Eventos.

Artefacto de la T08

Nombre del artefacto	eventos
Tipo de finalidad	Requisitos
Tipo de artefacto	Archivo binario
Tipo de formato	.backup
Tamaño	10 KB
Clase de artefacto	Entrada
Path	Directorio interno del paquete
Fecha de creación	15 de abril 2021
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	15 minutos
Tiempo corto de ejecución	10 minutos
Comentario	Es una copia de seguridad de la base de datos Eventos

Nombre del artefacto	eventos
Tipo de finalidad	Requisitos
Tipo de artefacto	Archivo de código
Tipo de formato	.sql
Tamaño	6 KB
Clase de artefacto	Entrada
Path	Directorio interno del paquete
Fecha de creación	15 de abril 2021
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	15 minutos
Tiempo corto de ejecución	10 minutos
Comentario	Script de la estructura de la BDD y datos de prueba

Nombre del artefacto	eventos-diagrama
Tipo de finalidad	Requisitos
Tipo de artefacto	Diagrama
Tipo de formato	.png
Tamaño	17 KB
Clase de artefacto	Entrada
Path	Directorio interno del paquete
Fecha de creación	15 de abril 2021
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	No se registra
Tiempo corto de ejecución	No se registra
Comentario	Diagrama relacional de la estructura de la base de datos Eventos

Nombre del artefacto	eventos-requisitos-registro-graphql
Tipo de finalidad	Requisitos
Tipo de artefacto	Documento
Tipo de formato	.xlsm
Tamaño	42 KB
Clase de artefacto	Entrada
Path	Directorio interno del paquete
Fecha de creación	15 de abril 2021
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	1 hora
Tiempo corto de ejecución	0:45 minutos
Comentario	Contiene los requisitos para el desarrollo de un API GraphQL, consta de 10 operaciones de datos CRUD. Contiene una sección para registrar la hora de inicio, fin, y el porcentaje de solución de cada requisito.

Nombre del artefacto	eventos-requisitos-registro-rest
Tipo de finalidad	Requisitos
Tipo de artefacto	Documento
Tipo de formato	.xlsm
Tamaño	42 KB
Clase de artefacto	Entrada
Path	Directorio interno del paquete
Fecha de creación	15 de abril 2021
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	1 hora
Tiempo corto de ejecución	0:45 minutos
Comentario	Contiene los requisitos para el desarrollo de un API-REST, consta de 10 operaciones de datos CRUD. Contiene una sección para registrar la hora de inicio, fin, y el porcentaje de solución de cada requisito.
Acrónimo	T09
Nombre de la tarea	Requisitos: Gestión de Facturas
Tipo de tarea	Experimental
Fecha de inicio	28 de junio 2021
Fecha de fin	28 de junio 2021
Duración	No se registra
Descripción	Requisitos de operaciones básicas CRUD, y diseño de base de datos para administrar Facturas.
Artefacto de la T09	
Nombre del artefacto	facturas
Tipo de finalidad	Requisitos
Tipo de artefacto	Archivo binario
Tipo de formato	.backup
Tamaño	10 KB
Clase de artefacto	Entrada
Path	Directorio interno del paquete
Fecha de creación	16 de abril 2021
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	15 minutos
Tiempo corto de ejecución	10 minutos
Comentario	Es una copia de seguridad de la base de datos Facturas
Nombre del artefacto	facturas
Tipo de finalidad	Requisitos
Tipo de artefacto	Archivo de código
Tipo de formato	.sql
Tamaño	5 KB
Clase de artefacto	Entrada
Path	Directorio interno del paquete
Fecha de creación	16 de abril 2021
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	15 minutos
Tiempo corto de ejecución	10 minutos
Comentario	Script de la estructura de la BDD y datos de prueba

Nombre del artefacto	facturas-diagrama
Tipo de finalidad	Requisitos
Tipo de artefacto	Diagrama
Tipo de formato	.png
Tamaño	17 KB
Clase de artefacto	Entrada
Path	Directorio interno del paquete
Fecha de creación	16 de abril 2021
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	No se registra
Tiempo corto de ejecución	No se registra
Comentario	Diagrama relacional de la estructura de la base de datos Facturas
Nombre del artefacto	facturas-requisitos-registro-graphql
Tipo de finalidad	Requisitos
Tipo de artefacto	Documento
Tipo de formato	.xslm
Tamaño	42 KB
Clase de artefacto	Entrada
Path	Directorio interno del paquete
Fecha de creación	16 de abril 2021
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	1 hora
Tiempo corto de ejecución	0:45 minutos
Comentario	Contiene los requisitos para el desarrollo de un API GraphQL, consta de 10 operaciones de datos CRUD. Contiene una sección para registrar la hora de inicio, fin, y el porcentaje de solución de cada requisito.
Nombre del artefacto	facturas-requisitos-registro-rest
Tipo de finalidad	Requisitos
Tipo de artefacto	Documento
Tipo de formato	.xslm
Tamaño	42 KB
Clase de artefacto	Entrada
Path	Directorio interno del paquete
Fecha de creación	16 de abril 2021
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	1 hora
Tiempo corto de ejecución	0:45 minutos
Comentario	Contiene los requisitos para el desarrollo de un API-REST, consta de 10 operaciones de datos CRUD. Contiene una sección para registrar la hora de inicio, fin, y el porcentaje de solución de cada requisito.
Acrónimo	T10
Nombre de la tarea	Requisitos: Gestión de Notas
Tipo de tarea	Experimental
Fecha de inicio	28 de junio 2021
Fecha de fin	28 de junio 2021
Duración	No se registra
Descripción	Requisitos de operaciones básicas CRUD, y diseño de base de datos para administrar Notas.

Artefacto de la T10

Nombre del artefacto	notas
Tipo de finalidad	Requisitos
Tipo de artefacto	Archivo binario
Tipo de formato	.backup
Tamaño	10 KB
Clase de artefacto	Entrada
Path	Directorio interno del paquete
Fecha de creación	17 de abril 2021
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	15 minutos
Tiempo corto de ejecución	10 minutos
Comentario	Es una copia de seguridad de la base de datos Notas

Nombre del artefacto	notas
Tipo de finalidad	Requisitos
Tipo de artefacto	Archivo de código
Tipo de formato	.sql
Tamaño	5 KB
Clase de artefacto	Entrada
Path	Directorio interno del paquete
Fecha de creación	17 de abril 2021
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	15 minutos
Tiempo corto de ejecución	10 minutos
Comentario	Script de la estructura de la BDD y datos de pruebas

Nombre del artefacto	notas-diagrama
Tipo de finalidad	Requisitos
Tipo de artefacto	Diagrama
Tipo de formato	.png
Tamaño	19 KB
Clase de artefacto	Entrada
Path	Directorio interno del paquete
Fecha de creación	17 de abril 2021
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	No se registra
Tiempo corto de ejecución	No se registra
Comentario	Diagrama relacional de la estructura de la base de datos Notas

Nombre del artefacto	notas-requisitos-registro-graphql
Tipo de finalidad	Requisitos
Tipo de artefacto	Documento
Tipo de formato	.xlsm
Tamaño	42 KB
Clase de artefacto	Entrada
Path	Directorio interno del paquete
Fecha de creación	17 de abril 2021
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	1 hora
Tiempo corto de ejecución	0:45 minutos
Comentario	Contiene los requisitos para el desarrollo de un API GraphQL, consta de 10 operaciones de datos CRUD. Contiene una sección para registrar la hora de inicio, fin, y el porcentaje de solución de cada requisito.

Nombre del artefacto	notas-requisitos-registro-rest
Tipo de finalidad	Requisitos
Tipo de artefacto	Documento
Tipo de formato	.xlsm
Tamaño	42 KB
Clase de artefacto	Entrada
Path	Directorio interno del paquete
Fecha de creación	17 de abril 2021
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	1 hora
Tiempo corto de ejecución	0:45 minutos
Comentario	Contiene los requisitos para el desarrollo de un API-REST, consta de 10 operaciones de datos CRUD. Contiene una sección para registrar la hora de inicio, fin, y el porcentaje de solución de cada requisito.
Acrónimo	T11
Nombre de la tarea	Encuesta de autoevaluación
Tipo de tarea	Recolección y Análisis
Fecha de inicio	28 de junio 2021
Fecha de fin	28 de junio 2021
Duración	4 minutos
Descripción	Los participantes completaron la encuesta al finalizar el experimento, para evaluar su percepción de facilidad de uso, consumo y aprendizaje de los paradigmas estudiados.
Artefacto de la T11	
Nombre del artefacto	encuesta-auto-evaluación
Tipo de finalidad	Datasets
Tipo de artefacto	Encuesta
Tipo de formato	.PDF
Tamaño	78 KB
Clase de artefacto	Entrada
Path	Directorio interno del paquete
Fecha de creación	18 de abril 2021
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	4 minutos
Tiempo corto de ejecución	2 minutos
Comentario	No se registra
Acrónimo	T12
Nombre de la tarea	Análisis estadístico
Tipo de tarea	Recolección y Análisis
Fecha de inicio	12 de julio 2021
Fecha de fin	28 de septiembre 2021
Duración	No se registra
Descripción	No se registra

Artefacto de la T12	
Nombre del artefacto	análisis
Tipo de finalidad	Gráficos estadísticos
Tipo de artefacto	Documento
Tipo de formato	.R
Tamaño	300 KB
Clase de artefacto	Salida
Path	Directorio interno del paquete
Fecha de creación	20 de julio 2021
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	No se registra
Tiempo corto de ejecución	No se registra
Comentario	Gráficos del análisis estadístico en R studio
Quinto paso	Registro de artefactos sin relación con las tareas del experimento
Campo	Valor
Nombre del artefacto	planificación-entrenamiento
Tipo de finalidad	Instrucciones
Tipo de artefacto	Documento
Tipo de formato	.PDF
Tamaño	25 KB
Clase de artefacto	Entrada
Path	Directorio interno del paquete
Fecha de creación	18 de marzo 2021
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	No se registra
Tiempo corto de ejecución	No se registra
Comentario	No se registra
Nombre del artefacto	Facilidad_Aprendizaje
Tipo de finalidad	Datasets
Tipo de artefacto	Archivos de datos
Tipo de formato	.xlsx
Tamaño	35 KB
Clase de artefacto	Salida
Path	Directorio interno del paquete
Fecha de creación	04 de septiembre 2021
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	No se registra
Tiempo corto de ejecución	No se registra
Comentario	Resultados de los casos de prueba

Nombre del artefacto	Facilidad_Aprendizaje-00
Tipo de finalidad	Datasets
Tipo de artefacto	Archivos de datos
Tipo de formato	.xlsx
Tamaño	1.121 KB
Clase de artefacto	Salida
Path	Directorio interno del paquete
Fecha de creación	8 de septiembre 2021
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	No se registra
Tiempo corto de ejecución	No se registra
Comentario	Conjunto de datos de los resultados
Nombre del artefacto	Amenazas-validez
Tipo de finalidad	<i>Debería ser Reportes</i>
Tipo de artefacto	Documento
Tipo de formato	.PDF
Tamaño	80 KB
Clase de artefacto	Salida
Path	Directorio interno del paquete
Fecha de creación	10 de septiembre 2021
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	No se registra
Tiempo corto de ejecución	No se registra
Comentario	Reporte de amenazas a la validez del experimento
Sexto paso	Registro de artefactos de las políticas ACM
No se registra ningún artefacto relacionado	
Séptimo paso	Se establece la configuración del Paquete de Laboratorio
No se registra ninguna configuración del PL	
Octavo paso	Evaluación de los artefactos de las políticas ACM
Campo	Valor
Artefacto Evaluado - Funcional	
Campo	Valor
Archivo "Inventario_artefactos"	No
Archivo "Readme"	No
Guía de instrucciones para descarga de artefactos	No
Guía de instrucciones para ejecución de artefactos	No
Relevancia de artefactos	Si
Artículo científico	No
Artefactos de nivel operacional	Si
Artefactos de nivel descriptivo	Si
Artefactos de nivel procedimental	No
Archivo "decision"	No
Descripción de scripts para generar los resultados	No
Descripción de software para generar los resultados	No
Accesibilidad de los archivos de datos	Si
Verificar que los scripts se ejecutan con éxito	Si
Verificar que el software se ejecuta con éxito	Si
Manipulación adecuada de los datos	Si
Registro de tiempos de ejecución completa (Opcional)	No
Registro de tiempos de ejecución corta (Opcional)	No

Artefacto - Disponible	
Campo	Valor
DOI del Paquete de Laboratorio	No
Repositorio de archivos público	No
Registro confidencial claves y contraseñas (De ser necesario)	No
Archivo "Citation" (Opcional)	No
Archivo "Authors" (Opcional)	No
Artefacto Evaluado - Reutilizable	
Campo	Valor
Artefactos bien estructurados	Si
Respeto de normas y estándares	No
Archivo "Inventario_artefactos"	No
Archivo "Readme"	No
Archivo "Contact"	No
Archivo "Requirements"	No
Archivo "Status"	No
Archivo "Licence "	No
Guía de instrucciones para obtención y descarga de artefactos	No
Guía de instrucciones para ejecución de artefactos	No
Relevancia de artefactos	Si
Artículo científico	No
Artefactos de nivel operacional	Si
Artefactos de nivel descriptivo	Si
Artefactos de nivel procedimental	No
Archivo "decision"	No
Descripción de scripts para generar resultados	No
Descripción de software para generar resultados	No
Accesibilidad de los archivos de datos	Si
Verificar que los scripts se ejecutan con éxito	Si
Verificar que los scripts se ejecutan con éxito	Si
Manipulación adecuada de los datos	Si
Registro del tipo de paquete	Si
Registro del entorno virtual (De ser necesario)	No
Guía de instalación de los artefactos	No
Artefactos comprimidos en un solo archivo	Si
Archivo "Install"	No
Verificar que la instalación de los artefactos dure menos de 30 minutos	No
Registro de tiempos de ejecución completa (Opcional)	No
Registro de tiempos de ejecución corta (Opcional)	No
Artefacto Evaluado - Reproducido	
No se registra ningún artefacto relacionado	
Artefacto Evaluado - Replicado	
No se registra ningún artefacto relacionado	
Noveno paso	Generación de una guía paso a paso para futuros replicadores
No se registra una guía paso a paso	

Décimo paso	Cumplimiento del Paquete de Laboratorio para alcanzar las insignias de la ACM
Funcional	38.55%
Disponibile	0%
Reutilizable	57.30%
Reproducido	0%
Replicado	0%
Undécimo paso	Registro del Paquete de Laboratorio
Campo	Valor
Nombre del paquete	Facilidad de aprendizaje entre GraphQL y REST: Paquete de Laboratorio
DOI	No se registra
Tipo de paquete	Simple
Repositorio de archivos público	No se registra
Fecha de creación	10 febrero 2022
Fecha de actualización	15 marzo 2022
Descripción	No se registra

INSTANCIACIÓN DE LA PROPUESTA EN EL CASO-MIND

En este anexo se muestra la instanciación de la propuesta Paquete de Laboratorio orientado al proceso acorde a la ACM sobre el *Caso-Mind*

Tabla C.1: Instanciación de la propuesta en el experimento Caso-Mind

Primer paso	Información, objetivos y descripción del experimento
Campo	Valor
Acrónimo	Mind-PL
Nombre del experimento	An Empirical Study to Evaluate the Impact of the Practice of Mindfulness in Workers Giving Support on Software Products.
DOI del experimento	No se registra
Lugar	Departamento de call center de Accenture
Fecha	julio 2021
Objetivo	Analizar la práctica de Mindfulness con el fin de evaluar sus efectos con respecto a la conciencia de atención y el rendimiento de los profesionales desde el punto de vista de los experimentadores y de los directivos de la empresa en el contexto de Accenture, una empresa de Ingeniería del Software que proporciona apoyo tecnológico.
Descripción	56 profesionales de un departamento de call center de Accenture (muestra) son considerados para el experimento. 26 de ellos durante 15 minutos antes del comienzo de su jornada laboral por 4 días a la semana durante 6 semanas practicaron Mindfulness (tratamiento grupo experimental). Los otros 30 sujetos eran el grupo de control, y no se les aplicó ningún tratamiento. Para los 56 sujetos, se midió la conciencia de atención antes de aplicar mindfulness y después de la aplicación del tratamiento. Para los 27 sujetos especializados en atender llamadas telefónicas, también se midió la productividad diaria de los empleados mediante KPI utilizados por Accenture para las revisiones de rendimiento.
Segundo paso	Registro de experimentadores
Campo	Valor
DNI	48805299N
Nombre del experimentador	Beatriz Bernárdez
email	beat@us.es
Afiliación	Universidad de Sevilla
Sitio web	https://bit.ly/36dZjQ0
Género	Femenino
Nacionalidad	Española
Tipo de persona	Profesional
Rol del experimentador	Formador, Diseñador
Comentario	No se registra

DNI	33472034L
Nombre del experimentador	Ignacio Panach
email	j.ignacio.panach@uv.es
Afiliación	Universidad de Valencia
Sitio web	http://www.uv.es/joigpana
Género	Masculino
Nacionalidad	Española
Tipo de persona	Graduado
Años de experiencia	No se registra
Rol del experimentador	Medidor, Monitor
Comentario	No se registra
DNI	44602941F
Nombre del experimentador	José Parejo
email	japarejo@us.es
Afiliación	Universidad de Sevilla
Sitio web	https://bit.ly/3L7XKSs
Género	Masculino
Nacionalidad	Española
Tipo de persona	Profesional
Rol del experimentador	Analista
Comentario	No se registra
DNI	27308303S
Nombre del experimentador	Amador Durán
email	amador@us.es
Afiliación	Universidad de Sevilla
Sitio web	https://bit.ly/3uE0vEv
Género	Masculino
Nacionalidad	Española
Tipo de persona	Profesional
Rol del experimentador	Medidor
Comentario	No se registra
DNI	No se registra
Nombre del experimentador	Natalia Juristo
email	natalia@fi.upm.es
Afiliación	Universidad Politécnica de Madrid
Sitio web	https://bit.ly/3KXPPaa
Género	Femenino
Nacionalidad	Española
Tipo de persona	Profesional
Rol del experimentador	Diseñador
Comentario	No se registra
DNI	28478419A
Nombre del experimentador	Antonio Ruiz-Cortés
email	aruiz@us.es
Afiliación	Universidad Politécnica de Madrid
Sitio web	https://bit.ly/3KXPPaa
Género	Masculino
Nacionalidad	Española
Tipo de persona	Profesional
Rol del experimentador	Diseñador
Comentario	No se registra

Tercer paso	Grupos y asignación de participantes
Campo	Valor
# grupos de Control	1
# participantes x grupo de Control	26
# grupos Experimentales	1
# participantes x grupo Experimental	30
Cuarto paso	Registro de tareas del experimento y sus artefactos
Campo	Valor
Acrónimo	T01
Nombre de la tarea	Presentación del Taller de Mindfulness
Tipo de tarea	Preparación
Fecha de inicio	No se registra
Fecha de fin	No se registra
Duración	1 hora
Responsable	Formador
Descripción	Información sobre el taller de Mindfulness.
Artefacto de la T01	
Nombre del artefacto	1. Recruitment [Talk]
Tipo de finalidad	Reclutamiento
Tipo de artefacto	Presentación
Tipo de formato	PDF
Tamaño	540.801 KB
Clase de artefacto	Entrada
Path	https://bit.ly/3qyK4YD
Fecha de creación	No se registra
Fecha de la última actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	20 minutos
Tiempo corto de ejecución	15 minutos
Comentario	No se registra
Acrónimo	T02
Nombre de la tarea	Cuestionario de interés
Tipo de tarea	Preparación
Fecha de inicio	No se registra
Fecha de fin	No se registra
Duración	4 minutos
Responsable	Formador
Descripción	Información demográfica de los sujetos y el consentimiento informado.
Artefacto de la T02	
Nombre del artefacto	Demographic and consent [Quest]
Tipo de finalidad	Reclutamiento
Tipo de artefacto	Cuestionario
Tipo de formato	PDF
Tamaño	184.5 KB
Clase de artefacto	Entrada
Path	https://bit.ly/3qyK4YD
Fecha de creación	No se registra
Fecha de la última actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	4 minutos
Tiempo corto de ejecución	2 minutos
Comentario	No se registra

Acrónimo	T03
Nombre de la tarea	Charlas sobre Mindfulness
Tipo de tarea	Preparación
Fecha de inicio	No se registra
Fecha de fin	No se registra
Duración	15 minutos
Responsable	Formador
Descripción	No se registra

Artefacto de la T03

Nombre del artefacto	2. Mindfulness [Talk]
Tipo de finalidad	Charlas
Tipo de artefacto	Presentación
Tipo de formato	PDF
Tamaño	2.239.058 KB
Clase de artefacto	Entrada
Path	https://bit.ly/3qyK4YD
Fecha de creación	No se registra
Fecha de la última actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	20 minutos
Tiempo corto de ejecución	15 minutos
Comentario	No se registra

Nombre del artefacto	3. To continue with Mindfulness [Talk]
Tipo de finalidad	Charlas
Tipo de artefacto	Presentación
Tipo de formato	PDF
Tamaño	672.172 KB
Clase de artefacto	Entrada
Path	https://bit.ly/3qyK4YD
Fecha de creación	No se registra
Fecha de la última actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	20 minutos
Tiempo corto de ejecución	15 minutos
Comentario	No se registra

Acrónimo	T04
Nombre de la tarea	Cuestionario de estrés antes de las sesiones de Mindfulness
Tipo de tarea	Preparación
Fecha de inicio	No se registra
Fecha de fin	No se registra
Duración	8 minutos
Responsable	Formador
Descripción	No se registra

Nombre del artefacto	MAAS-rawdata
Tipo de finalidad	Datasets
Tipo de artefacto	Archivo de datos
Tipo de formato	.csv
Tamaño	5.334 KB
Clase de artefacto	Salida
Path	https://bit.ly/3qyK4YD
Fecha de creación	No se registra
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	No se registra
Tiempo corto de ejecución	No se registra
Comentario	Resultado del cuestionario de estrés

Artefacto de la T04

Nombre del artefacto	MAAS [Quest]
Tipo de finalidad	Reclutamiento
Tipo de artefacto	Cuestionario
Tipo de formato	PDF
Tamaño	647.049 KB
Clase de artefacto	Entrada
Path	https://bit.ly/3qyK4YD
Fecha de creación	No se registra
Fecha de la última actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	8 minutos
Tiempo corto de ejecución	4 minutos
Comentario	No se registra

Acrónimo T05

Acrónimo	T05
Nombre de la tarea	Sesiones de Mindfulness
Tipo de tarea	Experimental
Fecha de inicio	No se registra
Fecha de fin	No se registra
Duración	15 minutos
Responsable	Formador
Descripción	Sesiones de Mindfulness 4 días a la semana durante 6 semanas.

Acrónimo T06

Acrónimo	T06
Nombre de la tarea	Cuestionario de estrés después de las sesiones de Mindfulness
Tipo de tarea	Recolección y Análisis
Fecha de inicio	No se registra
Fecha de fin	No se registra
Duración	8 minutos
Responsable	Medidor
Descripción	No se registra.

Artefacto de la T06

Nombre del artefacto	MAAS [Quest]
Tipo de finalidad	Datasets
Tipo de artefacto	Cuestionario
Tipo de formato	PDF
Tamaño	647.049 KB
Clase de artefacto	Entrada
Path	https://bit.ly/3qyK4YD
Fecha de creación	No se registra
Fecha de la última actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	8 minutos
Tiempo corto de ejecución	4 minutos
Comentario	No se registra

Nombre del artefacto	MAAS-rawdata
Tipo de finalidad	Datasets
Tipo de artefacto	Archivo de datos
Tipo de formato	.sav
Tamaño	1.616 KB
Clase de artefacto	Salida
Path	https://bit.ly/3qyK4YD
Fecha de creación	No se registra
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	No se registra
Tiempo corto de ejecución	No se registra
Comentario	Resultado del cuestionario de estrés

Acrónimo	T07
Nombre de la tarea	Cuestionario de Satisfacción
Tipo de tarea	Recolección y Análisis
Fecha de inicio	No se registra
Fecha de fin	No se registra
Duración	10 minutos
Responsable	Medidor
Descripción	No se registra.

Artefacto de la T07

Nombre del artefacto	Satisfaction [Quest]
Tipo de finalidad	Reporte
Tipo de artefacto	Cuestionario
Tipo de formato	.PDF
Tamaño	301.258 KB
Clase de artefacto	Entrada
Path	https://bit.ly/3qyK4YD
Fecha de creación	No se registra
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	10 minutos
Tiempo corto de ejecución	8 minutos
Comentario	No se registra

Acrónimo	T08
Nombre de la tarea	Análisis estadístico
Tipo de tarea	Recolección y Análisis
Fecha de inicio	No se registra
Fecha de fin	No se registra
Duración	No se registra
Responsable	Medidor
Descripción	No se registra.
Artefacto de la T08	
Nombre del artefacto	BarPlotSatisfaction
Tipo de finalidad	Gráficos estadísticos
Tipo de artefacto	Archivo comprimido
Tipo de formato	.rar
Tamaño	91.254 KB
Clase de artefacto	Salida
Path	https://bit.ly/3qyK4YD
Fecha de creación	No se registra
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	No se registra
Tiempo corto de ejecución	No se registra
Comentario	Gráficos Plot del análisis estadístico en SPSS
Nombre del artefacto	Results GLM mixed ANOVA
Tipo de finalidad	Gráficos estadísticos
Tipo de artefacto	Documento
Tipo de formato	.spv
Tamaño	72.287 KB
Clase de artefacto	Salida
Path	https://bit.ly/3qyK4YD
Fecha de creación	No se registra
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	No se registra
Tiempo corto de ejecución	No se registra
Comentario	Gráficos del análisis estadístico en SPSS
Nombre del artefacto	.Rhistory
Tipo de finalidad	Script
Tipo de artefacto	Archivo de código
Tipo de formato	.R
Tamaño	22.802 KB
Clase de artefacto	Salida
Path	https://bit.ly/3qyK4YD
Fecha de creación	No se registra
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	No se registra
Tiempo corto de ejecución	No se registra
Comentario	Script del análisis estadístico en R Studio

Nombre del artefacto	figures-generation
Tipo de finalidad	Script
Tipo de artefacto	Archivo de código
Tipo de formato	.R
Tamaño	5.659 KB
Clase de artefacto	Salida
Path	https://bit.ly/3qyK4YD
Fecha de creación	No se registra
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	No se registra
Tiempo corto de ejecución	No se registra
Comentario	Script análisis estadístico en R Studio
Nombre del artefacto	Answered phone calls-profile-plot
Tipo de finalidad	Gráficos estadísticos
Tipo de artefacto	Imagen
Tipo de formato	.emf
Tamaño	6.164 KB
Clase de artefacto	Salida
Path	https://bit.ly/3qyK4YD
Fecha de creación	No se registra
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	No se registra
Tiempo corto de ejecución	No se registra
Comentario	Plot en R Studio
Nombre del artefacto	Answered phone calls-profile-plot
Tipo de finalidad	Gráficos estadísticos
Tipo de artefacto	Imagen
Tipo de formato	.PDF
Tamaño	5.258 KB
Clase de artefacto	Salida
Path	https://bit.ly/3qyK4YD
Fecha de creación	No se registra
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	No se registra
Tiempo corto de ejecución	No se registra
Comentario	Plot en R Studio
Nombre del artefacto	Errors-profile-plot
Tipo de finalidad	Gráficos estadísticos
Tipo de artefacto	Imagen
Tipo de formato	.emf
Tamaño	7.820 KB
Clase de artefacto	Salida
Path	https://bit.ly/3qyK4YD
Fecha de creación	No se registra
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	No se registra
Tiempo corto de ejecución	No se registra
Comentario	Plot en R Studio

Nombre del artefacto	Errors-profile-plot
Tipo de finalidad	Gráficos estadísticos
Tipo de artefacto	Imagen
Tipo de formato	.PDF
Tamaño	5.680 KB
Clase de artefacto	Salida
Path	https://bit.ly/3qyK4YD
Fecha de creación	No se registra
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	No se registra
Tiempo corto de ejecución	No se registra
Comentario	Plot en R Studio
Nombre del artefacto	maas-box-and-profile-plot
Tipo de finalidad	Gráficos estadísticos
Tipo de artefacto	Imagen
Tipo de formato	.emf
Tamaño	9.140 KB
Clase de artefacto	Salida
Path	https://bit.ly/3qyK4YD
Fecha de creación	No se registra
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	No se registra
Tiempo corto de ejecución	No se registra
Comentario	Plot en R Studio
Nombre del artefacto	maas-box-and-profile-plot
Tipo de finalidad	Gráficos estadísticos
Tipo de artefacto	Imagen
Tipo de formato	.PDF
Tamaño	6.298 KB
Clase de artefacto	Salida
Path	https://bit.ly/3qyK4YD
Fecha de creación	No se registra
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	No se registra
Tiempo corto de ejecución	No se registra
Comentario	Plot en R Studio
Nombre del artefacto	Percentage of available time
Tipo de finalidad	Gráficos estadísticos
Tipo de artefacto	Imagen
Tipo de formato	.PDF
Tamaño	4.997 KB
Clase de artefacto	Salida
Path	https://bit.ly/3qyK4YD
Fecha de creación	No se registra
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	No se registra
Tiempo corto de ejecución	No se registra
Comentario	Plot en R Studio

Nombre del artefacto	Percentage of resting-profile-plot
Tipo de finalidad	Gráficos estadísticos
Tipo de artefacto	Imagen
Tipo de formato	.PDF
Tamaño	5.390 KB
Clase de artefacto	Salida
Path	https://bit.ly/3qyK4YD
Fecha de creación	No se registra
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	No se registra
Tiempo corto de ejecución	No se registra
Comentario	Plot en R Studio
Nombre del artefacto	Percentage of time waiting for call-profile-plot
Tipo de finalidad	Gráficos estadísticos
Tipo de artefacto	Imagen
Tipo de formato	.emf
Tamaño	7.152 KB
Clase de artefacto	Salida
Path	https://bit.ly/3qyK4YD
Fecha de creación	No se registra
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	No se registra
Tiempo corto de ejecución	No se registra
Comentario	Plot en R Studio
Nombre del artefacto	Percentage of time waiting for call-profile-plot
Tipo de finalidad	Gráficos estadísticos
Tipo de artefacto	Imagen
Tipo de formato	.PDF
Tamaño	5.432 KB
Clase de artefacto	Salida
Path	https://bit.ly/3qyK4YD
Fecha de creación	No se registra
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	No se registra
Tiempo corto de ejecución	No se registra
Comentario	Plot en R Studio
Nombre del artefacto	Rate_of.Calls-profile-plot
Tipo de finalidad	Gráficos estadísticos
Tipo de artefacto	Imagen
Tipo de formato	.emf
Tamaño	7.496 KB
Clase de artefacto	Salida
Path	https://bit.ly/3qyK4YD
Fecha de creación	No se registra
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	No se registra
Tiempo corto de ejecución	No se registra
Comentario	Plot en R Studio

Nombre del artefacto	Rate_of.Calls-profile-plot
Tipo de finalidad	Gráficos estadísticos
Tipo de artefacto	Imagen
Tipo de formato	.PDF
Tamaño	5.600 KB
Clase de artefacto	Salida
Path	https://bit.ly/3qyK4YD
Fecha de creación	No se registra
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	No se registra
Tiempo corto de ejecución	No se registra
Comentario	Plot en R Studio
Nombre del artefacto	Resolution-profile-plot
Tipo de finalidad	Gráficos estadísticos
Tipo de artefacto	Imagen
Tipo de formato	.emf
Tamaño	7.760 KB
Clase de artefacto	Salida
Path	https://bit.ly/3qyK4YD
Fecha de creación	No se registra
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	No se registra
Tiempo corto de ejecución	No se registra
Comentario	Plot en R Studio
Nombre del artefacto	Resolution-profile-plot
Tipo de finalidad	Gráficos estadísticos
Tipo de artefacto	Imagen
Tipo de formato	.PDF
Tamaño	5.765 KB
Clase de artefacto	Salida
Path	https://bit.ly/3qyK4YD
Fecha de creación	No se registra
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	No se registra
Tiempo corto de ejecución	No se registra
Comentario	Plot en R Studio
Quinto paso	
Registro de artefactos sin relación con las tareas del experimento	
Campo	Valor
Nombre del artefacto	KPIs-rawdata
Tipo de finalidad	Datasets
Tipo de artefacto	Archivo de datos
Tipo de formato	.csv
Tamaño	3.223 KB
Clase de artefacto	Salida
Path	https://bit.ly/3qyK4YD
Fecha de creación	No se registra
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	No se registra
Tiempo corto de ejecución	No se registra
Comentario	Resultado de los KPIs

Nombre del artefacto	KPIs-rawdata
Tipo de finalidad	Datasets
Tipo de artefacto	Archivo de datos
Tipo de formato	.sav
Tamaño	46.995 KB
Clase de artefacto	Salida
Path	https://bit.ly/3qyK4YD
Fecha de creación	No se registra
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	No se registra
Tiempo corto de ejecución	No se registra
Comentario	Resultado de los KPIs
Sexto paso	Registro de artefactos de las políticas ACM
Campo	Valor
Nombre del artefacto	1.readme
Tipo de finalidad	Insignias ACM
Tipo de artefacto	Archivo de texto
Tipo de formato	.txt
Tamaño	201 KB
Clase de artefacto	Entrada
Path	https://bit.ly/3qyK4YD
Fecha de creación	No se registra
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	No se registra
Tiempo corto de ejecución	No se registra
Comentario	Instrucciones del artefacto BarPlotSatisfaction.rar
Nombre del artefacto	1.readme
Tipo de finalidad	Insignias ACM
Tipo de artefacto	Archivo de texto
Tipo de formato	.txt
Tamaño	160 KB
Clase de artefacto	Entrada
Path	https://bit.ly/3qyK4YD
Fecha de creación	No se registra
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	No se registra
Tiempo corto de ejecución	No se registra
Comentario	Instrucciones para los R Script
Séptimo paso	Se establece la configuración del Paquete de Laboratorio
No se registra ninguna configuración del PL	

Octavo paso	Evaluación de los artefactos de las políticas ACM
Artefacto Evaluado - Funcional	
Campo	Valor
Archivo "Inventario_artefactos"	No
Archivo "Readme"	Si
Guía de instrucciones para obtención y descarga de artefactos	No
Guía de instrucciones para ejecución de artefactos	No
Relevancia de artefactos	Si
Artículo científico	Si
Artefactos de nivel operacional	Si
Artefactos de nivel descriptivo	Si
Artefactos de nivel procedimental	No
Archivo "decision"	No
Descripción de scripts para generar resultados	No
Descripción de software para generar resultados	No
Accesibilidad de los archivos de datos	Si
Verificar que los scripts se ejecutan con éxito	Si
Verificar que los software se ejecutan con éxito	Si
Manipulación adecuada de los datos	Si
Registro de tiempos de ejecución completa (Opcional)	No
Registro de tiempos de ejecución corta (Opcional)	No
Artefacto - Disponible	
Campo	Valor
DOI del Paquete de Laboratorio	Si
Repositorio de archivos público	Si
Registro confidencial claves y contraseñas (De ser necesario)	No
Archivo "Citation" (Opcional)	No
Archivo "Authors" (Opcional)	No
Artefacto Evaluado - Reutilizable	
Campo	Valor
Artefactos bien estructurados	Si
Respeto de normas y estándares	No
Archivo "Inventario_artefactos"	No
Archivo "Readme"	Si
Archivo "Contact"	No
Archivo "Requirements"	No
Archivo "Status"	No
Archivo "Licence "	No
Guía de instrucciones para obtención y descarga de artefactos	No
Guía de instrucciones para ejecución de artefactos	No
Relevancia de artefactos	Si
Artículo científico	Si
Artefactos de nivel operacional	Si
Artefactos de nivel descriptivo	Si
Artefactos de nivel procedimental	No
Archivo "decision"	No
Descripción de scripts para generar resultados	No
Descripción de software para generar resultados	No
Accesibilidad de los archivos de datos	Si
Verificar que los scripts se ejecutan con éxito	Si
Verificar que los software se ejecutan con éxito	Si
Manipulación adecuada de los datos	Si

Registro del tipo de paquete	Si
Registro del entorno virtual (Si en caso amerita)	No
Guía de instalación de los artefactos	No
Artefactos comprimidos en un solo archivo	Si
Archivo "Install"	No
Verificar que la instalación de los artefactos dure menos de 30 minutos	No
Registro de tiempos de ejecución completa (Opcional)	No
Registro de tiempos de ejecución corta (Opcional)	No
Artefacto Evaluado - Reproducido	
No se registra ningún artefacto relacionado	
Artefacto Evaluado - Replicado	
No se registra ningún artefacto relacionado	
Noveno paso	Generación de una guía paso a paso para futuros replicadores
No se registra una guía paso a paso	
Décimo paso	Cumplimiento del Paquete de Laboratorio para alcanzar las insignias de la ACM
Funcional	52.61%
Disponibile	100%
Reutilizable	71.36%
Reproducido	0%
Replicado	0%
Undécimo paso	Registro del Paquete de Laboratorio
Campo	Valor
Nombre del paquete	An Empirical Study to Evaluate the Impact of Mindfulness at a Software Company (Laboratory Package)
DOI	https://doi.org/10.5281/zenodo.5723957
Tipo de paquete	Simple
Repositorio de archivos público	Zenodo
Fecha de creación	28 octubre 2021
Fecha de actualización	28 octubre 2021
Descripción	Si registra descripción en Zenodo

INSTANCIACIÓN DE LA PROPUESTA EN EL CASO-REQ

En este anexo se muestra la instanciación de la propuesta Paquete de Laboratorio orientado al proceso acorde a la ACM sobre el *Caso-Req*

Tabla D.1: Instanciación de la propuesta en el experimento Caso-Req

Primer paso	Información, objetivos y descripción del experimento
Campo	Valor
Acrónimo	Req-2019
Nombre del experimento	Effect of requirements specification using native language on external software quality.
DOI del experimento	No se registra
Lugar	Curso de Maestría en Ingeniería del Software, Universidad de las Fuerzas Armadas-ESPE.
Fecha	marzo 2019
Objetivo	Analizar el lenguaje de los requisitos en lengua materna y en lengua extranjera con el fin de compararlos con respecto a la calidad externa desde el punto de vista de los investigadores en el contexto académico.
Descripción	17 estudiantes de la Maestría de Ingeniería del Software de la ESPE son considerados para el experimento. Los 17 participantes realizaron tareas de desarrollo de software utilizando requisitos en ambas lenguas (Español e Inglés) durante dos sesiones el mismo día. Las tareas estaban relacionadas con dos juegos (Bowling Scorekeeper y Tennis game). Se obtuvo 34 muestras por los dos tratamientos, se midió la calidad externa de las tareas desarrolladas en ambos lenguajes.
Segundo paso	Registro de experimentadores
Campo	Valor
DNI	0103109799
Nombre del experimentador	Fernando Uyaguari
email	fernando.uyaguari@wissen.edu.ec
Afiliación	Departamento de Desarrollo de Software, Instituto Superior Tecnológico Wissen
Sitio web	https://dblp.org/pid/67/2436.html
Género	Masculino
Nacionalidad	Ecuatoriana
Tipo de persona	Profesional
Rol del experimentador	Diseñador, Medidor
Comentario	No se registra

DNI	1002334835
Nombre del experimentador	Cathy Guevara
email	cguevara@utn.edu.ec
Afiliación	Universidad Técnica del Norte
Sitio web	https://bit.ly/3twx87y
Género	Femenino
Nacionalidad	Ecuatoriana
Tipo de persona	Profesional
Rol del experimentador	Medidor
Comentario	No se registra
DNI	1002322384
Nombre del experimentador	Antonio Quiña
email	aquina@utn.edu.ec
Afiliación	Universidad Técnica del Norte
Sitio web	https://bit.ly/3Jx88T3
Género	Masculino
Nacionalidad	Ecuatoriana
Tipo de persona	Profesional
Rol del experimentador	Medidor
Comentario	No se registra
DNI	0104281944
Nombre del experimentador	Cristina Acosta
email	beat@us.es
Afiliación	Departamento de Desarrollo de Software, Instituto Superior Tecnológico Wissen
Sitio web	https://bit.ly/383XcyH
Género	Femenino
Nacionalidad	Ecuatoriana
Tipo de persona	Profesional
Rol del experimentador	Monitor
Comentario	No se registra
DNI	0103411112
Nombre del experimentador	Alvaro Uyaguari
email	beat@us.es
Afiliación	Universidad de las Fuerzas Armadas-ESPE
Sitio web	https://bit.ly/37RivDk
Género	Masculino
Nacionalidad	Ecuatoriana
Tipo de persona	Profesional
Rol del experimentador	Analista
Comentario	No se registra
Tercer paso	Grupos y asignación de participantes
Campo	Valor
# grupos de Control	0
# participantes x grupo de Control	0
# grupos Experimentales	1
# participantes x grupo Experimental	17

Cuarto paso	Registro de tareas del experimento y sus artefactos
Campo	Valor
Acrónimo	T01
Nombre de la tarea	Presentación del experimento
Tipo de tarea	Preparación
Fecha de inicio	9 marzo 2019
Fecha de fin	9 marzo 2019
Duración	15 minutos
Responsable	Diseñador
Descripción	Información sobre las tareas del experimento
Acrónimo	T02
Nombre de la tarea	Cuestionario de interés
Tipo de tarea	Preparación
Fecha de inicio	9 marzo 2019
Fecha de fin	9 marzo 2019
Duración	4 minutos
Responsable	Medidor
Descripción	Información demográfica de los sujetos y el consentimiento informado
Artefacto de la T02	
Nombre del artefacto	Cuestionario
Tipo de finalidad	Reclutamiento
Tipo de artefacto	Cuestionario
Tipo de formato	PDF
Tamaño	184.5 KB
Clase de artefacto	Entrada
Path	Directorio interno del paquete
Fecha de creación	5 febrero 2019
Fecha de la última actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	4 minutos
Tiempo corto de ejecución	2 minutos
Comentario	No se registra
Acrónimo	T03
Nombre de la tarea	Requisitos de las tareas
Tipo de tarea	Experimental
Fecha de inicio	9 marzo 2019
Fecha de fin	9 marzo 2019
Duración	15 minutos
Responsable	Medidor
Descripción	Requisitos de la tareas para los juegos Bowling Scorekeeper y Tennis game en Inglés y Español.

Artefacto de la T03

Nombre del artefacto	Template-Bowling
Tipo de finalidad	Requisitos
Tipo de artefacto	Archivos comprimidos
Tipo de formato	.zip
Tamaño	2.127.068 KB
Clase de artefacto	Entrada
Path	https://doi.org/10.5281/zenodo.4950367
Fecha de creación	8 febrero 2019
Fecha de la última actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	10 minutos
Tiempo corto de ejecución	8 minutos
Comentario	No se registra

Nombre del artefacto	Template-Tennis
Tipo de finalidad	Requisitos
Tipo de artefacto	Archivos comprimidos
Tipo de formato	.zip
Tamaño	6.464 KB
Clase de artefacto	Entrada
Path	https://doi.org/10.5281/zenodo.4950367
Fecha de creación	9 febrero 2019
Fecha de la última actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	10 minutos
Tiempo corto de ejecución	8 minutos
Comentario	No se registra

Quinto paso **Registro de artefactos sin relación con las tareas del experimento**

Campo	Valor
Nombre del artefacto	Software Testing-Bowling
Tipo de finalidad	Datasets
Tipo de artefacto	Archivos de datos
Tipo de formato	.xlsx
Tamaño	25.293 KB
Clase de artefacto	Salida
Path	https://doi.org/10.5281/zenodo.4950367
Fecha de creación	19 mayo 2019
Fecha de la última actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	No se registra
Tiempo corto de ejecución	No se registra
Comentario	Resultados de los casos de prueba

Nombre del artefacto	Software Testing-Tennis
Tipo de finalidad	Datasets
Tipo de artefacto	Archivos de datos
Tipo de formato	.xlsx
Tamaño	26.903 KB
Clase de artefacto	Salida
Path	https://doi.org/10.5281/zenodo.4950367
Fecha de creación	19 mayo 2019
Fecha de la última actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	No se registra
Tiempo corto de ejecución	No se registra
Comentario	Resultados de los casos de prueba
Sexto paso	Registro de artefactos de las políticas ACM
Campo	Valor
Nombre del artefacto	README
Tipo de finalidad	Insignias ACM
Tipo de artefacto	Archivo de texto
Tipo de formato	.md
Tamaño	1.196 KB
Clase de artefacto	Entrada
Path	https://doi.org/10.5281/zenodo.4950367
Fecha de creación	17 mayo 2019
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	No se registra
Tiempo corto de ejecución	No se registra
Comentario	Contenido de los artefactos del paquete
Nombre del artefacto	Requirements
Tipo de finalidad	Insignias ACM
Tipo de artefacto	Archivo de texto
Tipo de formato	.txt
Tamaño	364 KB
Clase de artefacto	Entrada
Path	https://doi.org/10.5281/zenodo.4950367
Fecha de creación	17 mayo 2019
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	No se registra
Tiempo corto de ejecución	No se registra
Comentario	Requisitos para el desarrollo de las tareas del experimento
Nombre del artefacto	Descriptive RP
Tipo de finalidad	Insignias ACM
Tipo de artefacto	Documento
Tipo de formato	.PDF
Tamaño	1.947.345 KB
Clase de artefacto	Entrada
Path	https://doi.org/10.5281/zenodo.4950367
Fecha de creación	14 junio 2019
Fecha de actualización	No se registra
Usuario y contraseña de acceso	No se registra
Tiempo completo de ejecución	No se registra
Tiempo corto de ejecución	No se registra
Comentario	Descripción del diseño y contexto del experimento

Séptimo paso	Se establece la configuración del Paquete de Laboratorio
No se registra ninguna configuración del PL	
Octavo paso	Evaluación de los artefactos de las políticas ACM
Artefacto Evaluado - Funcional	
Campo	Valor
Archivo "Inventario.artefactos"	No
Archivo "Readme"	Si
Guía de instrucciones para obtención y descarga de artefactos	No
Guía de instrucciones para ejecución de artefactos	No
Relevancia de artefactos	Si
Artículo científico	Si
Artefactos de nivel operacional	Si
Artefactos de nivel descriptivo	Si
Artefactos de nivel procedimental	No
Archivo "decision"	No
Descripción de scripts para generar resultados	No
Descripción de software para generar resultados	No
Accesibilidad de los archivos de datos	Si
Verificar que los scripts se ejecutan con éxito	Si
Verificar que los software se ejecutan con éxito	Si
Manipulación adecuada de los datos	Si
Registro de tiempos de ejecución completa (Opcional)	No
Registro de tiempos de ejecución corta (Opcional)	No
Artefacto - Disponible	
Campo	Valor
DOI del Paquete de Laboratorio	Si
Repositorio de archivos público	Si
Registro confidencial claves y contraseñas (De ser necesario)	No
Archivo "Citation" (Opcional)	No
Archivo "Authors" (Opcional)	No
Artefacto Evaluado - Reutilizable	
Campo	Valor
Artefactos bien estructurados	Si
Respeto de normas y estándares	No
Archivo "Inventario.artefactos"	No
Archivo "Readme"	Si
Archivo "Contact"	No
Archivo "Requirements"	Si
Archivo "Status"	No
Archivo "Licence "	No
Guía de instrucciones para obtención y descarga de artefactos	No
Guía de instrucciones para ejecución de artefactos	No
Relevancia de artefactos	Si
Artículo científico	Si
Artefactos de nivel operacional	Si
Artefactos de nivel descriptivo	Si
Artefactos de nivel procedimental	No
Archivo "decision"	No
Descripción de scripts para generar resultados	No
Descripción de software para generar resultados	No
Accesibilidad de los archivos de datos	Si

Verificar que los scripts se ejecutan con éxito	Si
Verificar que los software se ejecutan con éxito	Si
Manipulación adecuada de los datos	Si
Registro del tipo de paquete	Si
Registro del entorno virtual (De ser necesario)	No
Guía de instalación de los artefactos	No
Artefactos comprimidos en un solo archivo	Si
Archivo "Install"	No
Verificar que la instalación de los artefactos dure menos de 30 minutos	No
Registro de tiempos de ejecución completa (Opcional)	No
Registro de tiempos de ejecución corta (Opcional)	No
Artefacto Evaluado - Reproducido	
No se registra ningún artefacto relacionado	
Artefacto Evaluado - Replicado	
No se registra ningún artefacto relacionado	
Noveno paso	Generación de una guía paso a paso para futuros replicadores
No se registra una guía paso a paso	
Décimo paso	Cumplimiento del Paquete de Laboratorio para alcanzar las insignias de la ACM
Funcional	52.61%
Disponible	100%
Reutilizable	77.61%
Reproducido	0%
Replicado	0%
Undécimo paso	Registro del Paquete de Laboratorio
Campo	Valor
Nombre del paquete	Replication Package - Effect of requirements specification using native language on external software quality
DOI	https://doi.org/10.5281/zenodo.4950367
Tipo de paquete	Simple
Repositorio de archivos público	Zenodo
Fecha de creación	14 junio 2021
Fecha de actualización	14 junio 2021
Descripción	Si registra descripción en Zenodo

CUESTIONARIO DE SATISFACCIÓN DEL EXPERIMENTADOR

Este anexo presenta el cuestionario para medir el aspecto de *Satisfacción*, consta de 10 ítems con 5 opciones de respuesta considerando los parámetros de la escala SUS.

Cuestionario para medir la satisfacción del experimentador

Objetivo del cuestionario: Determinar el nivel de satisfacción del experimentador luego de haber instanciado la propuesta: "Paquete de laboratorio orientado al proceso acorde a la ACM", para experimentos controlados con humanos en la Ingeniería del Software.

Instrucciones: Responda las siguientes preguntas, seleccionando una opción de escala de 1- 5 donde: 5: Muy de acuerdo 4: Algo de acuerdo 3: Ni de acuerdo ni en desacuerdo 2: Algo en desacuerdo 1: Muy en desacuerdo Los resultados que se obtengan serán de ayuda a los investigadores para determinar el nivel de satisfacción del experimentador con respecto a la aplicación de la propuesta. Los datos proporcionados serán manejados de forma confidencial y segura para uso exclusivamente académico.

Observaciones: Los investigadores extendemos un agradecimiento por el tiempo dedicado en contestar el cuestionario.

1. ¿Considera usted qué la instanciación de la propuesta es compleja?

5	4	3	2	1
<input type="radio"/>				

2. ¿Considera usted qué la propuesta fue fácil de usar?

5	4	3	2	1
<input type="radio"/>				

3. ¿Considera usted qué necesitaría ayuda de una persona con conocimientos avanzados del área de estudio para instanciar la propuesta?

5	4	3	2	1
<input type="radio"/>				

4. ¿Considera usted qué los pasos del procedimiento de la propuesta están bien establecidos?

5	4	3	2	1
<input type="radio"/>				

5. ¿Considera usted qué la propuesta es inconsistente?

5	4	3	2	1
<input type="radio"/>				

6. ¿Considera usted qué la mayoría de la gente aprendería a usar la propuesta en forma rápida?

5	4	3	2	1
<input type="radio"/>				

7. ¿Considera usted qué la propuesta es difícil de usar?

5	4	3	2	1
<input type="radio"/>				

8. ¿Se siente confiado al usar la propuesta?

5	4	3	2	1
<input type="radio"/>				

9. ¿Considera usted que necesita aprender algún conocimiento extra antes de usar la propuesta?

5

4

3

2

1

10. ¿Considera usted que usaría la propuesta frecuentemente?

5

4

3

2

1

BIBLIOGRAFÍA

Al final de cada referencia bibliográfica aparece un enlace a cada página en que se cita dicha referencia.

- [1] ROSE Festival 2018. Recognizing and Rewarding Open Science in Software Engineering. <https://bit.ly/3QNpm2R>, November 2018. (page 94).
- [2] ISO/IEC 25010. System and software quality model - "quality in use measures" of systems and software quality requirements and evaluation (square). <https://bit.ly/3NtEQHB>, 2011. (page 184).
- [3] ISO/IEC 25022. "quality in use measures" of systems and software quality requirements and evaluation (square). <https://www.iso.org/standard/35746.html>, 2016. (page 180).
- [4] R. Abdalkareem, V. Oda, S. Mujahid, and E. Shihab. On the impact of using trivial packages: an empirical case study on npm and pypi. *Empirical Software Engineering*, 25, 03 2020. (page 52).
- [5] S. Abrahão, C. Gravino, E. Insfran, G. Scanniello, and G. Tortora. Assessing the effectiveness of sequence diagrams in the comprehension of functional requirements: Results from a family of five experiments. *IEEE Transactions on Software Engineering*, 39(3):327–342, 2013. (page 52).
- [6] S. Abrahão and D. Mendez. ICSE'2021 Artifact Evaluation Track-Submission and Reviewing Guidelines. <https://bit.ly/3yyfh32>, 2021. (pages 13, 19, 32, 41, 48, 63, 94, 97, 100, 108, 111, 112, 114, 122, 126, 131, 134, 139).
- [7] S. Abrahão, E. Insfran, L. Guevara, F. González, M. Fernández, C. Cano, and R. Oliveira. Assessing the effectiveness of goal-oriented modeling languages: A family of experiments. *Information and Software Technology*, 116, 07 2019. (page 52).

- [8] S. Abrahão and G. Poels. Experimental evaluation of an object-oriented function point measurement procedure. *Information and Software Technology*, 49:366–380, 04 2007. (page 52).
- [9] L. Addazi and F. Ciccozzi. Blended graphical and textual modelling for uml profiles: A proof-of-concept implementation and experiment. *Journal of Systems and Software*, 175:110912, 2021. (page 52).
- [10] AE-Track ASE’2021. Accepted research artifacts. <https://bit.ly/3bgG3Ud>, 2021. (page 60).
- [11] AE-Track ESEC/FSE’2021. Accepted research artifacts, 2021. (page 60).
- [12] AE-Track ICSE’2021. Accepted research artifacts, 2021. (page 60).
- [13] AE-Track RE’2021. Accepted research artifacts. <https://bit.ly/3bc5x1Q>, 2021. (page 60).
- [14] W. Afzal, A. Ghazi, J. Itkonen, R. Torkar, A. Andrews, and K. Bhatti. An experiment on the effectiveness and efficiency of exploratory testing. *Empirical Softw. Engg.*, 20(3):844–878, jun 2015. (page 52).
- [15] M. Ahrens and K. Schneider. Improving requirements specification use by transferring attention with eye tracking data. *Information and Software Technology*, 131:106483, 2021. (page 52).
- [16] E. Aktaş and C. Yilmaz. Automated issue assignment: results and insights from an industrial case. *Empirical Software Engineering*, 07 2020. (page 52).
- [17] A. Alazba and H. Aljamaan. Code smell detection using feature selection and stacking ensemble: An empirical investigation. *Information and Software Technology*, 138:106648, 2021. (page 52).
- [18] Ö Albayrak and J. Carver. Investigation of individual factors impacting the effectiveness of requirements inspections: A replicated experiment. *Empirical Softw. Engg.*, 19(1):241–266, feb 2014. (page 52).
- [19] V. Alizadeh, M. Kessentini, M. Mkaouer, M. Ocinneide, A. Ouni, and Y. Cai. An interactive and dynamic search-based approach to software refactoring recommendations. *IEEE Transactions on Software Engineering*, 46(9):932–961, 2020. (page 52).

- [20] L. Allodi, M. Cremonini, F. Massacci, and W. Shim. Measuring the accuracy of software vulnerability assessments: experiments with students and professionals. *Empirical Software Engineering*, 25, 03 2020. (page 52).
- [21] Jpf. Almqvist. *Replication of controlled experiments in empirical software engineering—a survey*. PhD thesis, Lund University, 2006. (pages 6, 12, 31, 34, 36, 69, 70, 100).
- [22] H. G. Amatriain, O. Dieste, E. Fernández, and R. García Martínez. Evaluación de la aplicabilidad del meta-análisis de efectos aleatorios en ingeniería del software. In *XVII Congreso Argentino de Ciencias de la Computación*, 2011. (page 28).
- [23] Ethics Committee American Psychological Association. Ethical principles of psychologists and code of conduct. *American Psychologist*, 2017. (page 86).
- [24] N. Amálio, L. Briand, and P. Kelsen. An experimental scrutiny of visual design modelling: Vcl up against uml+ocl. *Empirical Software Engineering*, 25, 03 2020. (page 52).
- [25] B. Anda and D. Sjøberg. Investigating the role of use cases in the construction of class diagrams. *Empirical Softw. Engg.*, 10(3):285–309, jul 2005. (page 52).
- [26] C. Apa, O. Dieste, E. Espinosa, and E. Fonseca. Effectiveness for detecting faults within and outside the scope of testing techniques: an independent replication. *Empirical Software Engineering*, 19:378–417, 2014. (page 52).
- [27] A. Aranda, O. Dieste, and N. Juristo. Effect of domain knowledge on elicitation effectiveness: An internally replicated controlled experiment. *IEEE Transactions on Software Engineering*, 42(5):427–451, 2016. (page 52).
- [28] P. Ardimento, M. Baldassarre, D. Caivano, and G. Visaggio. Assessing multiview framework (mf) comprehensibility and efficiency: A replicated experiment. *Information and Software Technology*, 48:313–322, 05 2006. (page 52).
- [29] L. Ardito, R. Coppola, G. Malnati, and M. Torchiano. Effectiveness of kotlin vs. java in android app development tasks. *Information and Software Technology*, 127:106374, 2020. (page 52).
- [30] E. Arisholm, L. Briand, S. Hove, and Y. Labiche. The impact of uml documentation on software maintenance: an experimental evaluation. *IEEE Transactions on Software Engineering*, 32(6):365–381, 2006. (page 52).

- [31] E. Arisholm, H. Gallis, T. Dyba, and D. Sjöberg. Evaluating pair programming with respect to system complexity and programmer expertise. *IEEE Transactions on Software Engineering*, 33(2):65–86, 2007. (page 52).
- [32] E. Arisholm and D. Sjöberg. Evaluating the effect of a delegated versus centralized control style on the maintainability of object-oriented software. *IEEE Transactions on Software Engineering*, 30(8):521–534, 2004. (page 52).
- [33] J. Arnau. *Psicología Experimental. Un enfoque metodológico*, volume 1. Trillas, 1980. (page 40).
- [34] M. Asanza, M. Miranda, R. Ortiz, and J. Espín. Manual de procedimiento en la empresa. *Revista Caribeña de Ciencias Sociales*, 15, 2016. (page 87).
- [35] ASE'2021. AE-Artifact Evaluation. <https://bit.ly/3uIfLBk>, 2021. (pages 56, 94, 140).
- [36] M. Babar, B. Kitchenham, and R. Jeffery. Comparing distributed and face-to-face meetings for software architecture evaluation: A controlled experiment. *Empirical Softw. Engg.*, 13(1):39–62, feb 2008. (page 52).
- [37] M. Baldassarre, D. Caivano, D. Fucci, N. Juristo, S. Romano, G. Scanniello, and B. Turhan. Studying test-driven development and its retainment over a six-month time span. *Journal of Systems and Software*, 176:110937, 2021. (page 52).
- [38] M. Baldassarre, J. Carver, O. Dieste, and N. Juristo. Replication types: Towards a shared taxonomy. In *Proceedings of EASE '14*, pages 18:1–18:4, 2014. (page 4).
- [39] R. Bandi, V. Vaishnavi, and D. Turk. Predicting maintenance performance using object-oriented design complexity metrics. *IEEE Transactions on Software Engineering*, 29(1):77–87, 2003. (page 52).
- [40] V. Basili, G. Caldiera, and H. D. Rombach. Goal question metrics paradigm. *Encyclopedia of Software Engineering*, pages 528–53, 1994. (page 88).
- [41] V. Basili, S. Green, O. Laitenberger, F. Lanubile, F. Shull, S. Sorumgard, and M. Zelkowitz. Packaging researcher experience to assist replication of experiments. In *Proc. of the ISERN meeting 1996*, 1996. (pages 6, 11, 30, 34, 36, 60, 68, 70, 100).
- [42] V. Basili, S. Green, O. Laitenberger, F. Lanubile, F. Shull, S. Sørungård, and M. Zelkowitz. Lab package for the empirical investigation of perspective-based reading, 2006. (pages 59, 65).

- [43] V. Basili and W. Reiter. A controlled experiment quantitatively comparing software development approaches. *IEEE Transaction Software Engineering*, SE-7, 1981. (page 85).
- [44] V. Basili and R. Selby. Paradigms for experimentation and empirical studies in software engineering. *Reliab. Eng. Syst. Saf.*, 1(32):171–191, 1991. (page 4).
- [45] V. Basili, R. Selby, and D. Hutchens. Experimentation in software engineering. *IEEE Transactions on software engineering*, 1(7):733–743, 1986. (pages 4, 25, 82).
- [46] V. Basili, F. Shull, and F. Lanubile. Building knowledge through families of experiments. *IEEE Transactions on Software Engineering*, 25(4):456–473, 1999. (pages 4, 5, 26, 27, 30, 71).
- [47] P. Berander and M. Svahnberg. Evaluating two ways of calculating priorities in requirements hierarchies – an experiment on hierarchical cumulative voting. *Journal of Systems and Software*, 82(5):836–850, 2009. (page 52).
- [48] B. Bernárdez, A. Durán, J. A Parejo, and A. Ruiz-Cortés. A controlled experiment to evaluate the effects of mindfulness in software engineering. In *Proceedings of ESEM*, pages 17–27. ACM, 2014. (page 179).
- [49] K. Bernsmed, D. Soares Cruzes, M. Gilje Jaatun, and M. Iovan. Adopting threat modelling in agile software development projects. *Journal of Systems and Software*, 183:111090, 2022. (page 52).
- [50] B. Bernárdez, A. Durán, J. Parejo, N. Juristo, and A. Ruiz-Cortés. Effects of mindfulness on conceptual modeling performance: a series of experiments. *IEEE Transactions on Software Engineering*, pages 1–1, 2020. (page 52).
- [51] B. Bernárdez, A. Durán, J. Parejo, and A. Ruiz-Cortés. An experimental replication on the effect of the practice of mindfulness in conceptual modeling performance. *Journal of Systems and Software*, 136:153–172, 2016. (pages 52, 179).
- [52] B. Bernárdez, I. Panach, J. Parejo, A. Durán, N. Juristo, and A. Ruiz-Cortés. An Empirical Study to Evaluate the Impact of Mindfulness at a Software Company (Laboratory Package). <https://doi.org/10.5281/zenodo.5723957>, October 2021. (page 167).
- [53] S. Biffel. Evaluating defect estimation models with major defects. *Journal of Systems and Software*, 65(1):13–29, 2003. (page 52).

- [54] T. Bipp, A. Lepper, and D. Schmedding. Pair programming in software development teams – an empirical study of its benefits. *Information and Software Technology*, 50:231–240, 02 2008. (page 52).
- [55] F. Blanco, J. Perales, and M. Vadillo. ¿puede la psicología rescatarse a sí misma? incentivos, sesgos y replicabilidad, 2017. (page 68).
- [56] J. Blanco and D. Lucrédio. A holistic approach for cross-platform software development. *Journal of Systems and Software*, 179:110985, 09 2021. (page 52).
- [57] A. Blouin and J. Jezequel. Interacto: A modern user interaction processing model. *IEEE Transactions on Software Engineering*, pages 1–1, 2021. (page 52).
- [58] B. Boehm. Verifying and validating software requirements and design specifications. *IEEE Software*, pages 75–88, 1984. (page 10).
- [59] B. Boehm. Value-based software engineering. *ACM Software Engineering Notes*, 2003. (page 10).
- [60] R. Bogdan and S. Biklen. *Qualitative Research for Education: An Introduction to Theory and Methods*. Boston, MA: Allyn and Bacon, 1982. (page 84).
- [61] L. Bratthall and C. Wohlin. Is it possible to decorate graphical software design and architecture models with qualitative information?-an experiment. *IEEE Transactions on Software Engineering*, 28(12):1181–1193, 2002. (page 52).
- [62] L. Braz, E. Fregnan, G. Çalikli, and A. Bacchelli. Data and materials for: Why don't developers detect improper input validation?; drop table papers; -. In *2021 IEEE/ACM 43rd International Conference on Software Engineering: Companion Proceedings (ICSE-Companion)*, pages 203–204, 2021. (page 52).
- [63] L. Briand, D. Bianculli, S. Nejati, F. Pastore, and M. Sabetzadeh. The case for context-driven software engineering research: Generalizability is overrated. *IEEE Software*, 34(5):72–75, 2017. cited By 50. (page 74).
- [64] L. Briand, C. Bunse, and J. Daly. A controlled experiment for evaluating quality guidelines on the maintainability of object-oriented designs. *Software Engineering, IEEE Transactions on*, 27(6):513 – 530, 07 2001. (page 52).
- [65] L. Briand, Y. Labiche, M. Di Penta, and H. Yan-Bondoc. An experimental investigation of formality in uml-based development. *IEEE Transactions on Software Engineering*, 31(10):833–849, 2005. (page 52).

- [66] J. Brooke. Sus: A retrospective. *Journal of Usability Studies-JUS*, 8(2):29–40, 2013. (page 180).
- [67] A. Brooks, J. Daly, J. Miller, M. Roper, and M. Wood. Replication’s role in experimental computer science. *EfoCS–5–94 (RR/94/171)*, 1994. (page 68).
- [68] A. Brooks, J. Daly, J. Miller, M. Roper, and M. Wood. Replication of experimental results in software engineering. *Technical Report ISERN-96-10, University of Strathclyde*, 1996. (pages 26, 68).
- [69] A. Brooks, M. Roper, M. Wood, J. Daly, and J. Miller. *Replication’s Role in Software Engineering*. Springer London, London, 2008. (page 71).
- [70] A. Brooks, F. Utbult, C. Mulligan, and R. Jeffery. Early lifecycle work: Influence of individual characteristics, methodological constraints, and interface constraints. *Empirical Softw. Engg.*, 5(3):269–285, nov 2000. (page 52).
- [71] D. Budgen, A. Burn, and B. Kitchenham. Reporting computing projects through structured abstracts: A quasi-experiment. *Empirical Software Engineering*, 16:244–277, 04 2011. (page 52).
- [72] D. Budgen, B. Kitchenham, S. Charters, M. Turner, P. Brereton, and S. Linkman. Presenting software engineering results using structured abstracts: A randomised experiment. *Empirical Softw. Engg.*, 13(4):435–468, aug 2008. (page 52).
- [73] J. Bulegon Gassen, J. Mendling, A. Bouzeghoub, L. Thom, and J. Palazzo M. de Oliveira. An experiment on an ontology-based support approach for process modeling. *Information and Software Technology*, 83:94–115, 2017. (page 52).
- [74] C. Bunse. Using patterns for the refinement and translation of uml models: A controlled experiment. *Empirical Software Engineering*, 11:227–267, 2006. (page 52).
- [75] J. Burkhardt, F. Détienne, and S. Wiedenbeck. Object-oriented program comprehension: Effect of expertise, task and phase. *Empirical Softw. Engg.*, 7(2):115–156, jun 2002. (page 52).
- [76] Code Ocean by IEEE. IEEE Xplore/Code Ocean FAQs. <https://codeocean.com/signup/ieee>, 2021. (page 8).
- [77] TechRxiv by IEEE. TechRxiv preprint server-Browse. <https://bit.ly/3EOCSMD>, 2021. (page 8).

- [78] J. Börstler and B. Paech. The role of method chains and comments in software readability and comprehension—an experiment. *IEEE Transactions on Software Engineering*, 42(9):886–898, 2016. (page 52).
- [79] C. Cachero, S. Meliá, and J. Hermida. Impact of model notations on the productivity of domain modelling: An empirical study. *Information and Software Technology*, 108, 12 2019. (page 52).
- [80] D. Caivano, M. Fernández-Roper, R. Pérez-Castillo, M. Piattini, and M. Scalera. Artifact-based vs. human-perceived understandability and modifiability of refactored business processes: An experiment. *Journal of Systems and Software*, 144:143–164, 2018. (page 52).
- [81] C. Calero, M. Piattini, and M. Genero. Empirical validation of referential integrity metrics. *Information and Software Technology*, 43:949–957, 12 2001. (page 52).
- [82] D. Campbell. Factors relevant to the validity of experiments in social settings. *Psychological bulletin*, 54(4):297, 1957. (page 39).
- [83] M. Campusano, J. Fabry, and A. Bergel. Live programming in practice: A controlled experiment on state machines for robotic behaviors. *Information and Software Technology*, 108:99–114, 2019. (page 52).
- [84] G. Canfora, A. Cimitile, F. Garcia, M. Piattini, and C. Visaggio. Evaluating performances of pair designing in industry. *Journal of Systems and Software*, 80(8):1317–1327, 2007. The Impact of Barry Boehm’s Work on Software Engineering Education and Training. (page 52).
- [85] B. Cartaxo, A. Almeida, E. Barreiros, J. Saraiva, W. Ferreira, and S. Soares. Mechanisms to characterize context of empirical studies in software engineering. In *Experimental Software Engineering Latin American Workshop (ESELAW 2015)*, pages 1–14, 2015. (page 75).
- [86] J. Carver. Towards reporting guidelines for experimental replications: a proposal. In *1st International Workshop on Replication in Empirical Software Engineering*, 2010. (page 5).
- [87] J. Carver, N. Juristo, M. Baldassarre, and S. Vegas. Replications of software engineering experiments, 2014. (pages 5, 29).

- [88] J. Carver, N. Nagappan, and A. Page. The impact of educational background on the effectiveness of requirements inspections: An empirical study. *IEEE Transactions on Software Engineering*, 34(6):800–812, 2008. (page 52).
- [89] CERN. European Organization for Nuclear Research. <https://home.cern/>, 1954. (page 37).
- [90] M. Ciolkowski, F. Shull, and S. Biffi. *A family of experiments to investigate the influence of context on the effect of inspection techniques*. na, 2002. (page 27).
- [91] W. Cochran and G. Cox. *Experimental Designs*. New York: Wiley, 1950. (page 84).
- [92] C. Collberg and T. Proebsting. Repeatability in computer systems research. *Communications of the ACM*, 59(3):62–69, 2016. (page 72).
- [93] C. Collberg and A. Warren. Repeatability and Benefaction in Computer Systems Research: A Study and Modest Proposal, TR 14-04. Technical report, University of Arizona, Arizona, feb 2015. (page 72).
- [94] H. Collins. The seven sexes: A study in the sociology of a phenomenon, or the replication of experiments in physics. *Sociology*, 9(2):205–224, 1975. (page 4).
- [95] M. Colosimo, A. Lucia, G. Scanniello, and G. Tortora. Evaluating legacy system migration technology through empirical studies. *Information and Software Technology*, 51:433–447, 02 2009. (page 52).
- [96] European Commission. Policy for the European Commission. <https://bit.ly/3TJ14IF>, 2020. (page 37).
- [97] R. Conradi, V. Basili, J. Carver, F. Shull, and G. Travassos. A Pragmatic Documents Standard for an Experience Library: Roles, Documen, Contents and Structure. Technical report, University of Maryland, 2001. (page 31).
- [98] T. Cook, D. Campbell, and A. Day. *Quasi-experimentation: Design & analysis issues for field settings*, volume 351. Houghton Mifflin Boston, 1979. (page 39).
- [99] N. Cooper, C. Bernal-Cárdenas, O. Chaparro, K. Moran, and D. Poshyvanyk. A replication package for it takes two to tango: Combining visual and textual information for detecting duplicate video-based bug reports. In *2021 IEEE/ACM 43rd International Conference on Software Engineering: Companion Proceedings (ICSE-Companion)*, pages 160–161, 2021. (page 52).

- [100] B. Cornelissen, A. Zaidman, and A. van Deursen. A controlled experiment for program comprehension through trace visualization. *IEEE Transactions on Software Engineering*, 37(3):341–355, 2011. (page 52).
- [101] F. Corradini, A. Morichetta, A. Polini, B. Re, L. Rossi, and F. Tiezzi. Correctness checking for bpmn collaborations with sub-processes. *Journal of Systems and Software*, 166:110594, 2020. (page 52).
- [102] COS. Center for Open Science. <https://www.cos.io/>, 2017. (page 37).
- [103] M. Cruz, B. Bernárdez, A. Durán, J. Galindo, and A. Ruiz-Cortés. Replication of studies in empirical software engineering: A systematic mapping study, from 2013 to 2018. *IEEE Access*, 8:26773–26791, 2020. (page 26).
- [104] M. Cruz, B. Bernárdez, A. Durán, C. Guevara-Vega, and A. Ruiz-Cortés. A model-based approach for specifying changes in replications of empirical studies in computer science, 2022. Submitted to *Computing*. (page 196).
- [105] J. Cruz-Lemus, M. Genero, D. Caivano, S. Abrahão, E. Insfrán, and J. Carsí. Assessing the influence of stereotypes on the comprehension of uml sequence diagrams: A family of experiments. *Information and Software Technology*, 53(12):1391–1403, 2011. (page 52).
- [106] J. Cruz-Lemus, M. Genero, M. Manso, S. Morasca, and M. Piattini. Assessing the understandability of uml statechart diagrams with composite states—a family of empirical studies. *Empirical Software Engineering*, 14(6):685–719, 2009. (page 52).
- [107] J. A. Cruz-Lemus, M. Genero, D. Caivano, S. Abrahão, E. Insfrán, and J. Carsí. Assessing the influence of stereotypes on the comprehension of UML sequence diagrams: A family of experiments. *Information and Software Technology*, 53(12):1391–1403, 2011. (pages 31, 65).
- [108] C. Czepa and U. Zdun. On the understandability of temporal properties formalized in linear temporal logic, property specification patterns and event processing language. *IEEE Transactions on Software Engineering*, 46(1):100–112, 2020. (page 52).
- [109] J. da Costa, R. Gheyi, M. Ribeiro, S. Apel, V. Alves, B. Fonseca, F. Medeiros, and A. Garcia. Evaluating refactorings for disciplining #ifdef annotations: An eye tracking study with novices. *Empirical Software Engineering*, 26, 09 2021. (page 52).

- [110] F. Dalpiaz, P. Gieske, and A. Sturm. On deriving conceptual models from user requirements: An empirical study. *Information and Software Technology*, 131:106484, 2021. (page 52).
- [111] P. Dasgupta and Paul D. Toward a new economics of science. *Research Policy*, 23(5):487–521, 1994. (page 37).
- [112] IEEE DataPort. Search standard and open access datasets-ieee dataport. <https://ieee-dataport.org/datasets>, 2021. (page 8).
- [113] M. Daun, J. Brings, P. Aluko, and V. Stenkova. Reliability of self-rated experience and confidence as predictors for students' performance in software engineering: Results from multiple controlled experiments on model comprehension with graduate and undergraduate students. *Empirical Software Engineering*, 26, 07 2021. (page 52).
- [114] P. David. The Historical Origins of 'Open Science': An Essay on Patronage, Reputation and Common Agency Contracting in the Scientific Revolution. *Capitalism and Society*, 2(2), 2008. (page 37).
- [115] M. de Almeida Maia and R. Raquel Fialho Lafetá. On the impact of trace-based feature location in the performance of software maintainers. *Journal of Systems and Software*, 86(4):1023–1037, 2013. SI : Software Engineering in Brazil: Retrospective and Prospective Views. (page 52).
- [116] M. de Freitas Farias, M. de Mendonça Neto, M. Kalinowski, and R. Oliveira. Identifying self-admitted technical debt through code comment analysis with a contextualized vocabulary. *Information and Software Technology*, 121:106270, 2020. (page 52).
- [117] J. de la Vara, B. Marín, C. Ayora, and G. Giachetti. An empirical evaluation of the use of models to improve the understanding of safety compliance needs. *Information and Software Technology*, 126:106351, 2020. (page 52).
- [118] A. De Lucia, C. Gravino, R. Oliveto, and G. Tortora. An experimental comparison of er and uml class diagrams for data modelling. *Empirical Software Engineering*, 15(5):455–492, oct 2010. (page 52).
- [119] A. De Lucia, R. Oliveto, and G. Tortora. Assessing ir-based traceability recovery tools through controlled experiments. *Empirical Software Engineering*, 14:57–92, 2009. (page 52).

- [120] V. De Magalhães and F. Da Silva. Towards a taxonomy of replications in empirical software engineering research: A research proposal. In *Proceedings - 2013 3rd International Workshop on Replication in Empirical Software Engineering Research, RESER 2013*, pages 50–55, 2013. (page 5).
- [121] Notas de Teórico Versión 1.0 del Dpto. de Arquitectura-InCo-FIng. Aljebra de boole, 2000. (pages 97, 101, 102, 103, 109, 111, 112, 113, 115, 116, 121, 122, 124, 125, 126, 128, 131, 133, 134, 135, 139).
- [122] I. Deligiannis, I. Stamelos, L. Angelis, M. Roumeliotis, and M. Shepperd. A controlled experiment investigation of an object-oriented design heuristic for maintainability. *Journal of Systems and Software*, 72(2):129–143, 2004. (page 52).
- [123] T. Dewi Saputri and S. Lee. Integrated framework for incorporating sustainability design in software engineering life-cycle: An empirical study. *Information and Software Technology*, 129:106407, 2021. (page 52).
- [124] P. Diebold and A. Vetrò. Bridging the gap: Se technology transfer into practice: Study design and preliminary results. In *Proceedings of the 8th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement, ESEM '14*, New York, NY, USA, 2014. Association for Computing Machinery. (page 74).
- [125] H. Do, S. Elbaum, and G. Rothermel. Supporting Controlled Experimentation with Testing Techniques: An Infrastructure and its Potential Impact. *Empirical Software Engineering* 2005 10:4, 10(4):405–435, oct 2005. (page 32).
- [126] Á. Domingo, J. Echeverría, Ó. Pastor, and C. Cetina. Evaluating the influence of scope on feature location. *Information and Software Technology*, 140:106674, 2021. (page 52).
- [127] T. Dybå, D. I. K. Sjøberg, and D. S. Cruzes. What works for whom, where, when, and why?: on the role of context in empirical software engineering. In *Proceedings of the ACM-IEEE international symposium on Empirical software engineering and measurement*, pages 19–28. ACM, 2012. (page 75).
- [128] W. Dzidek, E. Arisholm, and L. Briand. A realistic empirical evaluation of the costs and benefits of uml in software maintenance. *IEEE Transactions on Software Engineering*, 34(3):407–432, 2008. (page 52).
- [129] E. Díaz, J. Panach, S. Rueda, and D. Distanto. A family of experiments to generate graphical user interfaces from bpmn models with stereotypes. *Journal of Systems and Software*, 173:110883, 2021. (page 52).

- [130] L. D'Avila, K. Farias, and J. Victória Barbosa. Effects of contextual information on maintenance effort: A controlled experiment. *Journal of Systems and Software*, 159:110443, 2020. (page 52).
- [131] Z. Eberhart, A. Bansal, and C. Mcmillan. A wizard of oz study simulating api usage dialogues with a virtual assistant. *IEEE Transactions on Software Engineering*, pages 1–1, 2020. (page 52).
- [132] J. Echeverría, F. Pérez, J. Panach, and C. Cetina. An empirical study of performance using clone & own and software product lines in an industrial context. *Information and Software Technology*, 130:106444, 2021. (page 52).
- [133] H. Eichelberger and K. Schmid. Guidelines on the aesthetic quality of uml class diagrams. *Information and Software Technology*, 51:1686–1698, 12 2009. (page 52).
- [134] B. Eken and A. Tosun. Investigating the performance of personalized models for software defect prediction. *Journal of Systems and Software*, 181:111038, 2021. (page 52).
- [135] M. El-Attar. Using smcd to reduce inconsistencies in misuse case models: A subject-based empirical evaluation. *Journal of Systems and Software*, 87:104–118, 2014. (page 52).
- [136] M. El-Attar. Empirically evaluating the effect of the physics of notations on model construction. *IEEE Transactions on Software Engineering*, pages 1–1, 2021. (page 52).
- [137] M. El-Attar and J. Miller. A subject-based empirical evaluation of ssucd's performance in reducing inconsistencies in use case models. *Empirical Softw. Engg.*, 14(5):477–512, oct 2009. (page 52).
- [138] Elsevier. Free preprint posting service in ssrn (open-access online pre-print server). <https://bit.ly/3RYMgom>, 2021. (page 60).
- [139] M. Endres, M. Fansher, P. Shah, and W. Weimer. To read or to rotate? comparing the effects of technical reading training and spatial skills training on novice programming ability. In *To Read or to Rotate? Comparing the Effects of Technical Reading Training and Spatial Skills Training on Novice Programming Ability*, ESEC/FSE 2021, page 754–766, New York, NY, USA, 2021. Association for Computing Machinery. (page 52).

- [140] H. Erdogmus, M. Morisio, and M. Torchiano. On the effectiveness of the test-first approach to programming. *IEEE Transactions on Software Engineering*, 31(3):226–237, 2005. (page 52).
- [141] ESEC/FSE'2021. AE-Artifact Evaluation. <https://bit.ly/3jJGJCo>, 2021. (pages 56, 60, 94, 140).
- [142] S. Fakhoury, D. Roy, Y. Ma, V. Arnaoudova, and O. Adesope. Measuring the impact of lexical and structural inconsistencies on developers' cognitive load during bug localization. *Empirical Software Engineering*, 25, 05 2020. (pages 52, 59).
- [143] J. Feigenspan, C. Kästner, S. Apel, J. Liebig, M. Schulze, R. Dachsel, M. Papendieck, T. Leich, and G. Saake. Do background colors improve program comprehension in the #ifdef hell? *Empirical Softw. Engg.*, 18(4):699–745, aug 2013. (pages 52, 65).
- [144] D. Feitelson, A. Mizrahi, N. Noy, A. Ben, O. Eliyahu, and R. Sheffer. How developers choose names. *IEEE Transactions on Software Engineering*, pages 1–1, 2020. (page 52).
- [145] M. Felderer and A. Herrmann. Manual test case derivation from uml activity diagrams and state machines: a controlled experiment. *Information and Software Technology*, 61, 05 2015. (page 52).
- [146] R. Feldt and A. Magazinius. Validity threats in empirical software engineering research-an initial survey. In *Seke*, pages 374–379, 2010. (page 39).
- [147] A. Fernández-Sáez, M. Genero, D. Caivano, and M. Chaudron. Does the level of detail of uml diagrams affect the maintainability of source code? a family of experiments. *Empirical Softw. Engg.*, 21(1):212–259, feb 2016. (pages 52, 59).
- [148] A. Fernández, S. Abrahão, and E. Insfran. Empirical validation of a usability inspection method for model-driven web development. *Journal of Systems and Software*, 86(1):161–186, 2013. (page 52).
- [149] A. Fernández-Sáez, M. Genero, M. Chaudron, D. Caivano, and I. Ramos. Are forward designed or reverse-engineered uml diagrams more helpful for code maintenance?: A family of experiments. *Information and Software Technology*, 57:644–663, 2014. (page 52).

- [150] J. Ferreira, S. Acuña, O. Dieste, S. Vegas, A. Santos, F. Rodríguez, and N. Juristo. Impact of usability mechanisms: An experiment on efficiency, effectiveness and user satisfaction. *Information and Software Technology*, 117:106195, 2020. (page 52).
- [151] D. Firmenich, S. Firmenich, G. Rossi, M. Wimmer, I. Garrigós, and C. González-Mora. Engineering web augmentation software: A development method for enabling end-user maintenance. *Information and Software Technology*, 141:106735, 2022. (page 52).
- [152] E. Fonseca. *Conceptualización e Infraestructura para la Investigación Experimental en Ingeniería del Software Parte 2*. Tesis de doctorado, Universidad Politécnica del Madrid, 2014. (page 25).
- [153] European Commission-Directorate-General for Research and Innovation. Next-generation metrics: Responsible metrics and evaluation for open science. <https://bit.ly/3TKgdJK>, 2017. (page 38).
- [154] D. Fucci, G. Scanniello, S. Romano, and N. Juristo. Need for sleep: The impact of a night of sleep deprivation on novice developers' performance. *IEEE Transactions on Software Engineering*, 46(1):1–19, 2020. (page 52).
- [155] D. Fucci and B. Turhan. On the role of tests in test-driven development: A differentiated and partial replication. *Empirical Software Engineering*, 04 2014. (page 52).
- [156] V. Garousi, K. Petersen, and B. Ozkan. Challenges and best practices in industry-academia collaborations in software engineering: A systematic literature review. *Information and Software Technology*, 79:106–127, 2016. (page 74).
- [157] D. Garvin. What does “product quality” really mean?, 1984. (page 91).
- [158] M. Genero, J. Cruz-Lemus, and M. Piattini. *Métodos de investigación en ingeniería de software*. Ra-Ma, Bogotá, ra-ma edition, 2015. (page 25).
- [159] B. George and L. Williams. A structured experiment of test-driven development. *Information and Software Technology*, 46:337–342, 04 2004. (page 52).
- [160] M. Gil, M. Albert, J. Fons, and V. Pelechano. Engineering human-in-the-loop interactions in cyber-physical systems. *Information and Software Technology*, 126:106349, 2020. (page 52).

- [161] D. Girardi, F. Lanubile, N. Novielli, and A. Serebrenik. Emotions and perceived productivity of software developers at the workplace. *IEEE Transactions on Software Engineering*, pages 1–1, 2021. (page 52).
- [162] C. Glezer, M. Last, E. Nachmany, and P. Shoval. Quality comprehension of uml interaction diagrams—an experimental comparison. *Information and Software Technology*, 47:675–692, 07 2005. (page 52).
- [163] O. Gómez, N. Juristo, and S. Vegas. Replication, reproduction and re-analysis: Three ways for verifying experimental findings. In *Proceedings of the 1st international workshop on replication in empirical software engineering research (RESER 2010), Cape Town, South Africa, 2010*. (pages 28, 29).
- [164] O. Gómez, N. Juristo, and S. Vegas. Understanding replication of experiments in software engineering: A classification. *Information and Software Technology*, 56(8):1033–1048, 2014. (pages 27, 149).
- [165] C. Gralha, M. Goulão, and J. Araújo. Are there gender differences when interacting with social goal models? *Empirical Software Engineering*, 25:1–38, 11 2020. (page 52).
- [166] C. Gralha, R. Pereira, M. Goulão, and J. Araujo. On the impact of using different templates on creating and understanding user stories. In *2021 IEEE 29th International Requirements Engineering Conference (RE)*, pages 209–220, 2021. (page 52).
- [167] A. Grand, C. Wilkinson, K. Bultitude, and A. Winfield. Open Science: A New “Trust Technology”? *Science Communication*, 34(5):679–689, sep 2012. (page 38).
- [168] D. Graziotin. open science policies. <https://bit.ly/31UNn39>, 2021. (pages 7, 13, 38).
- [169] J. Grigera, A. Garrido, J. Panach, D. Distanto, and G. Rossi. Assessing refactorings for usability in e-commerce applications. *Empirical Softw. Engg.*, 21(3):1224–1271, jun 2016. (page 52).
- [170] L. Guerrouj, M. Penta, Y. Guéhéneuc, and G. Antoniol. An experimental investigation on the effects of context on source code identifiers splitting and expansion. *Empirical Softw. Engg.*, 19(6):1706–1753, dec 2014. (page 52).

- [171] C. Guevara-Vega. Publicación en open science de paquetes de laboratorio experimentales para ingeniería del software. In *2021 XXIV Ibero-American Conference on Software Engineering (CIBSE) - Doctoral Symposium*, 2021. (page 196).
- [172] C. Guevara-Vega, B. Bernárdez, A. Durán, A. Quiña-Mera, M. Cruz, and A. Ruiz-Cortés. Empirical strategies in software engineering research: A literature survey. In *2021 Second International Conference on Information Systems and Software Technologies (ICI2ST)*, pages 120–127, 2021. (pages 165, 196).
- [173] I. Hadar, I. Reinhartz-berger, T. Kuflik, A. Perini, F. Ricca, and A. Susi. Comparing the comprehensibility of requirement models expressed in use case and tropes: Results from a family of experiments. *Information and Software Technology*, 55, 10 2013. (page 52).
- [174] S. Hanenberg, S. Kleinschmager, R. Robbes, É. Tanter, and A. Stefik. An empirical study on the impact of static typing on software maintainability. *Empirical Software Engineering*, 19(5):1335–1382, oct 2014. (page 52).
- [175] C. Hannebauer, M. Hesenius, and V. Gruhn. Does syntax highlighting help programming novices? In *Proceedings of the 40th International Conference on Software Engineering, ICSE '18*, page 704, New York, NY, USA, 2018. Association for Computing Machinery. (page 52).
- [176] C. Hendrick. Replications, strict replications, and conceptual replications: are they important? *Journal of Social Behavior and Personality*, 5(4):41, 1990. (page 29).
- [177] B. Hermann, S. Winter, and J. Siegmund. Community expectations for research artifacts and evaluation processes. In *Proceedings of the 28th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering, ESEC/FSE 2020*, page 469–480, New York, NY, USA, 2020. Association for Computing Machinery. (page 7).
- [178] M. A Heroux, L. A Barba, M. Parashar, and V. Stodden. Toward a compatible reproducibility taxonomy for computational and computing sciences. Technical report, Sandia National Laboratories, 2018. (pages 8, 111).
- [179] R. Heumüller, S. Nielebock, J. Krüger, and F. Ortmeier. Publish or perish, but do not forget your software artifacts. *Empirical Software Engineering*, 25(6):4585–4616, nov 2020. (pages 31, 73).

- [180] R. Hofman. Behavioral economics in software quality engineering. *Empirical Softw. Engg.*, 16(2):278–293, apr 2011. (page 52).
- [181] L. Huang and M. Holcombe. Empirical investigation towards the effectiveness of test first programming. *Information and Software Technology*, 51:182–194, 01 2009. (page 52).
- [182] F. Hujainah, R. Binti Abu Bakar, A. Nasser, B. Al-haimi, and K. Zamli. Sr-ptackle: A semi-automated requirements prioritisation technique for scalable requirements of software system projects. *Information and Software Technology*, 131:106501, 2021. (page 52).
- [183] F. Häser, M. Felderer, and R. Breu. Is business domain language support beneficial for creating test case specifications: A controlled experiment. *Information and Software Technology*, 79, 07 2016. (page 52).
- [184] P. Hübner and B. Paech. Interaction-based creation and maintenance of continuously usable trace links between requirements and source code. *Empirical Software Engineering*, 25, 09 2020. (page 52).
- [185] ICSE'2021. Artifact Review and Badging-Current. <https://bit.ly/3oQqJB0>, 2020. (pages 5, 8, 19, 28, 30, 48, 55, 68, 81, 85, 92, 94, 97, 98, 99, 103, 110, 121, 125, 132, 140, 145).
- [186] ICSE'2021. AE-Artifact Evaluation. <https://bit.ly/3JxuQeu>, 2021. (pages 32, 55, 60, 94, 140).
- [187] IEEE. Standard Glossary of Software Engineering Terminology. *IEEE Std 610.12-1990*, pages 1–84, 1990. (page 9).
- [188] A. Inge Wang and E. Arisholm. The effect of task order on the maintainability of object-oriented software. *Information and Software Technology*, 51:293–305, 2009. (page 52).
- [189] ISERN. International Software Engineering Research Network. <https://isern.iiese.de/>, 2021. (page 30).
- [190] ISO/IEC. 25000 SQuaRE System and Software Quality Requirements and Evaluation. <https://iso25000.com/index.php/normas-iso-25000>, 2021. (page 91).

- [191] J. Itkonen and M. Mäntylä. Are test cases needed? replicated comparison between exploratory and test-case-based software testing. *Empirical Softw. Engg.*, 19(2):303–342, apr 2014. (page 52).
- [192] A. Jbara and D. Feitelson. How programmers read regular code: A controlled experiment using eye tracking. *Empirical Softw. Engg.*, 22(3):1440–1477, jun 2017. (page 52).
- [193] A. Jedlitschka and M. Ciolkowski. Towards evidence in software engineering. In *Proceedings. 2004 International Symposium on Empirical Software Engineering, 2004. ISESE'04.*, pages 261–270. IEEE, 2004. (page 75).
- [194] A. Jedlitschka, M. Ciolkowski, and D. Pfahl. Reporting experiments in software engineering. In *Guide to advanced empirical software engineering*, pages 201–228. Springer, 2008. (pages 6, 12, 25, 26, 29, 36, 70, 83, 100).
- [195] A. Jedlitschka and D. Pfahl. Reporting guidelines for controlled experiments in software engineering. In *Empirical Software Engineering, 2005. 2005 International Symposium on*, pages 10–pp. IEEE, 2005. (page 70).
- [196] Y. Jiang, H. Liu, Y. Zhang, N. Niu, Y. Zhao, and L. Zhang. Which abbreviations should be expanded? In *Which Abbreviations Should Be Expanded?*, ESEC/FSE 2021, page 578–589, New York, NY, USA, 2021. Association for Computing Machinery. (page 52).
- [197] A. Johanson and W. Hasselbring. Effectiveness and efficiency of a domain-specific language for high-performance marine ecosystem simulation: A controlled experiment. *Empirical Softw. Engg.*, 22(4):2206–2236, aug 2017. (page 52).
- [198] R. Jolak, M. Savary-Leblanc, M. Dalibor, A. Wortmann, R. Hebig, J. Vincur, I. Polášek, X. Pallec, S. Gérard, and M. Chaudron. Software engineering whispers: The effect of textual vs. graphical software design descriptions on software design communication. *Empirical Software Engineering*, 25, 11 2020. (page 52).
- [199] EMSE journal. The open science initiative of empirical software engineering. <https://bit.ly/3zgdeiz>, 2019. (page 60).
- [200] N. Juristo and O. Gómez. Replication of software engineering experiments. In *Empirical software engineering and verification*, pages 60–88. Springer, 2010. (page 28).

- [201] N. Juristo and A. Moreno. *Basics of software engineering experimentation*. Kluwer Academic Publishers, 2001. (pages 4, 25, 70).
- [202] N. Juristo and S. Vegas. Using differences among replications of software engineering experiments to gain knowledge. In *Using differences among replications of software engineering experiments to gain knowledge*, pages 356–366, 2009. cited By 55. (page 74).
- [203] J. Jurkiewicz, J. Nawrocki, M. Ochodek, and T. Głowacki. Hazop-based identification of events in use cases. *Empirical Softw. Engg.*, 20(1):82–109, feb 2013. (page 52).
- [204] M. Jørgensen. Identification of more risks can lead to increased over-optimism of and over-confidence in software development effort estimates. *Information and Software Technology*, 52:506–516, 05 2010. (page 52).
- [205] M. Jørgensen. Unit effects in software project effort estimation: Work-hours gives lower effort estimates than workdays. *Journal of Systems and Software*, 117:274–281, 2016. (page 52).
- [206] M. Jørgensen and T. Halkjelsvik. Sequence effects in the estimation of software development effort. *Journal of Systems and Software*, 159:110448, 2020. (page 52).
- [207] J. Kabeli and P. Shoval. Comprehension and quality of analysis specifications - a comparison of foam and opm methodologies. *Information and Software Technology*, 47:271–290, 03 2005. (page 52).
- [208] S. Kadri, S. Aouag, and D. Hedjazi. Ms-quaaf: A generic evaluation framework for monitoring software architecture quality. *Information and Software Technology*, 140:106713, 2021. (page 52).
- [209] E. Kamsties, A. Antje von Knethen, and R. Reussner. A controlled experiment to evaluate how styles affect the understandability of requirements specifications. *Information and Software Technology*, 45:955–965, 2003. (page 52).
- [210] E. Kamsties and C. Lott. An empirical evaluation of three defect-detection techniques. In *Proceedings of the Fifth European Software Engineering Conference, ESEC*, 1995. (page 68).
- [211] E. Karac, B. Turhan, and N. Juristo. A controlled experiment with novice developers on the impact of task description granularity on software quality in test-

- driven development. *IEEE Transactions on Software Engineering*, 47:1315–1330, 2019. (page 52).
- [212] A. Karahasanović, A. Levine, and R. Thomas. Comprehension strategies and difficulties in maintaining object-oriented systems: An explorative study. *Journal of Systems and Software*, 80(9):1541–1559, 2007. Evaluation and Assessment in Software Engineering. (page 52).
- [213] P. Karpati, A. Opdahl, and G. Sindre. Investigating security threats in architectural context: Experimental evaluations of misuse case maps. *Journal of Systems and Software*, 104:90–111, 2015. (page 52).
- [214] P. Karpati, Y. Redda, A. Opdahl, and G. Sindre. Comparing attack trees and misuse cases in an industrial setting. *Information and Software Technology*, 56, 01 2014. (page 52).
- [215] O. Karras, K. Schneider, and S. Fricker. Representing software project vision by means of video: A quality model for vision videos. *Journal of Systems and Software*, 162:110479, 2020. (page 52).
- [216] M. Keil, L. Li, L. Mathiassen, and G. Zheng. The influence of checklists and roles on software practitioner risk perception and decision-making. *Journal of Systems and Software*, 81(6):908–919, 2008. Agile Product Line Engineering. (page 52).
- [217] F. Kifetew, A. Perini, A. Susi, A. Siena, D. Muñante, and I. Morales-Ramirez. Automating user-feedback driven requirements prioritization. *Information and Software Technology*, 138:106635, 2021. (page 52).
- [218] J. King, J. Stallings, M. Riaz, and L. Williams. To log, or not to log: Using heuristics to identify mandatory log events — a controlled experiment. *Empirical Softw. Engg.*, 22(5):2684–2717, oct 2017. (page 52).
- [219] B. Kitchenham. Procedures for Performing Systematic Reviews. Technical report, Keele University, Australia, 2004. (page 68).
- [220] B. Kitchenham. The role of replications in empirical software engineering—a word of warning. *Empirical Software Engineering*, 13(2):219–221, 2008. (pages 26, 71, 72).
- [221] B. Kitchenham, H. Al-Khilidar, M. Babar, M. Berry, K. Cox, J. Keung, F. Kurniawati, M. Staples, H. Zhang, and L. Zhu. Evaluating guidelines for empirical

- software engineering studies. In *Proceedings of the 2006 ACM/IEEE international symposium on Empirical software engineering*, pages 38–47, 2006. (pages 6, 12, 35, 36, 70, 100).
- [222] B. Kitchenham and S. Charters. Guidelines for performing systematic literature reviews in software engineering, 2007. (page 46).
- [223] B. Kitchenham, T. Dyba, and M. Jorgensen. Evidence-based software engineering. In *Proceedings of the 26th international conference on software engineering*, pages 273–281. IEEE Computer Society, 2004. (page 5).
- [224] B. Kitchenham, S. Pfleeger, L. Pickard, P. Jones, D. Hoaglin, and K. El Emam. Preliminary Guidelines for Empirical Research in Software Engineering. *IEEE Transactions on Software Engineering*, 28(8):721 – 734, 2002. (pages 25, 29).
- [225] J. Knodel, D. Muthig, and M. Naab. An experiment on the role of graphical elements in architecture visualization. *Empirical Softw. Engg.*, 13(6):693–726, dec 2008. (page 52).
- [226] P. Koch, K. Schekotihin, D. Jannach, B. Hofer, and F. Wotawa. Metric-based fault prediction for spreadsheets. *IEEE Transactions on Software Engineering*, 47(10):2195–2207, 2021. (page 52).
- [227] T. Kosar, S. Gaberc, J. Carver, and M. Mernik. Program comprehension of domain-specific and general-purpose languages: Replication of a family of experiments using integrated development environments. *Empirical Softw. Engg.*, 23(5):2734–2763, oct 2018. (page 52).
- [228] T. Kosar, M. Mernik, and J. Carver. Program comprehension of domain-specific and general-purpose languages: Comparison using a family of experiments. *Empirical Softw. Engg.*, 17(3):276–304, jun 2012. (page 52).
- [229] J. Krein. *Replication and Knowledge Production in Empirical Software Engineering Research*. PhD thesis, Brigham Young University, 2014. (page 30).
- [230] J. Krein, L. Prechelt, N. Juristo, A. Nanthamornphong, J. Carver, S. Vegas, C. Knutson, K. Seppi, and D. Eggett. A multi-site joint replication of a design patterns experiment using moderator variables to generalize across contexts. *IEEE Transactions on Software Engineering*, 42(4):302–321, 2016. (page 52).
- [231] S Krishnamurthi. Examining “Reproducibility in Computer Science”, 2013. (page 72).

- [232] J. Krüger, G. Çalıkılı, D. Bershadskyy, R. Heyer, S. Zabel, and S. Otto. Registered report: A laboratory experiment on using different financial-incentivization schemes in software-engineering experimentation, 2022. (page 86).
- [233] A. Kumar Thurimella and B. Brügge. A mixed-method approach for the empirical evaluation of the issue-based variability modeling. *Journal of Systems and Software*, 86(7):1831–1849, 2013. (page 52).
- [234] S. Kuttal, X. Chen, Z. Wang, S. Balali, and A. Sarma. Visual resume: Exploring developers' online contributions for hiring. *Information and Software Technology*, 138:106633, 2021. (page 52).
- [235] O. Laitenberger, C. Atkinson, M. Schlich, and K. El Emam. An experimental comparison of reading techniques for defect detection in uml design documents. *Journal of Systems and Software*, 53(2):183–204, 2000. (page 52).
- [236] H. Larsson, E. Lindqvist, and R. Torkar. Outliers and replication in software engineering. In *Software Engineering Conference (APSEC), 2014 21st Asia-Pacific*, volume 1, pages 207–214. IEEE, 2014. (page 72).
- [237] R. Latorre. Effects of developer experience on learning and applying unit test-driven development. *IEEE Transactions on Software Engineering*, 40(4):381–395, 2014. (page 52).
- [238] T. LaToza, M. Arab, D. Loksa, and A. Ko. Explicit programming strategies. *Empirical Software Engineering*, 25, 07 2020. (page 52).
- [239] A. Lavallo, A. Maté, J. Trujillo, M. Teruel, and S. Rizzi. A methodology to automatically translate user requirements into visualizations: Experimental validation. *Information and Software Technology*, 136:106592, 2021. (page 52).
- [240] T. Lehtinen, M. Mäntylä, J. Itkonen, and J. Vanhanen. Diagrams or structural lists in software project retrospectives – an experimental comparison. *Journal of Systems and Software*, 103:17–35, 2015. (page 52).
- [241] T. Lethbridge, S. Sim, and J. Singer. Studying software engineers: Data collection techniques for software field studies. *Empirical software engineering*, 10(3):311–341, 2005. (page 169).
- [242] B. Li, X. Diao, W. Gao, and C. Smidts. A requirements inspection method based on scenarios generated by model mutation and the experimental validation. *Empirical Software Engineering*, 26, 09 2021. (page 52).

- [243] T. Lin. ‘Open Science’ Challenges Journal Tradition With Web Collaboration - The New York Times, 2012. (page 37).
- [244] R. Lindsay and A. Ehrenberg. The design of replicated studies. *The American Statistician*, 47(3):217–228, 1993. (page 26).
- [245] D. Liu, Y. Feng, X. Zhang, J. Jones, and Z. Chen. Clustering crowdsourced test reports of mobile applications using image understanding. *IEEE Transactions on Software Engineering*, pages 1–1, 2020. (page 52).
- [246] M. Liu, X. Peng, A. Marcus, C. Treude, X. Bai, G. Lyu, J. Xie, and X. Zhang. Learning-based extraction of first-order logic representations of api directives. In *Learning-Based Extraction of First-Order Logic Representations of API Directives*, ESEC/FSE 2021, page 491–502, New York, NY, USA, 2021. Association for Computing Machinery. (page 52).
- [247] S. Liu. A three-step hybrid specification approach to error prevention. *Journal of Systems and Software*, 178:110975, 2021. (page 52).
- [248] S. Liu, H. Li, Z. Jiang, X. Li, F. Liu, and Y. Zhong. Rigorous code review by reverse engineering. *Information and Software Technology*, 133:106503, 2021. (page 52).
- [249] S. Liu and S. Nakajima. Automatic test case and test oracle generation based on functional scenarios in formal specifications for conformance testing. *IEEE Transactions on Software Engineering*, pages 1–1, 2020. (page 52).
- [250] S. Liu, H. Wang, and C. Xu. Tidy: A pbe-based framework supporting smart transformations for entity consistency in powerpoint. *Information and Software Technology*, 138:106611, 2021. (page 52).
- [251] D. Lo, N. Nagappan, and T. Zimmermann. How practitioners perceive the relevance of software engineering research. In *Proceedings of the 2015 10th Joint Meeting on Foundations of Software Engineering*, ESEC/FSE 2015, page 415–425, New York, NY, USA, 2015. Association for Computing Machinery. (page 74).
- [252] K. Lui, K. Chan, and J. Nosek. The effect of pairs in program design tasks. *IEEE Transactions on Software Engineering*, 34(2):197–211, 2008. (page 52).
- [253] J. Lung, J. Aranda, S. Easterbrook, and G. Wilson. On the difficulty of replicating human subjects studies in software engineering. In *Proceedings of ICSE’08*, pages 191–200. IEEE, 2008. (page 76).

- [254] I. Lytra, P. Gaubatz, and U. Zdun. Two controlled experiments on model-based architectural decision making. *Information and Software Technology*, 63:58–75, 2015. (page 52).
- [255] E. Løhre and M. Jørgensen. Numerical anchors and their strong effects on software development effort estimates. *Journal of Systems and Software*, 116:49–56, 2016. (page 52).
- [256] P. Mäder and A. Egyed. Do developers benefit from requirements traceability when evolving and maintaining a software system? *Empirical Softw. Engg.*, 20(2):413–441, apr 2015. (page 52).
- [257] L. Madeyski. The impact of test-first programming on branch coverage and mutation score indicator of unit tests: An experiment. *Information and Software Technology*, 52:169–184, 02 2010. (page 52).
- [258] J. Maldonado, J. Carver, F. Shull, S. Fabbri, Dória; E., L. Martimiano, M. Mendonça, and V. Basili. Perspective-based reading: A replicated experiment focused on individual reviewer effectiveness. *Empirical Software Engineering*, 11:119–142, 2006. (page 52).
- [259] R. Mall. *Fundamentals of Software Engineering, Fifth Edition*. Eastern Economy Edition. PHI Learning Private Limited, 2018. (page 4).
- [260] J. Marcos, A. Arroyo, J. Garzás, and M. Piattini. La norma ISO/IEC 25000 y el proyecto KEMIS para su automatización con software libre. In *La norma ISO/IEC 25000 y el proyecto KEMIS para su automatización con software libre*, 2008. (page 91).
- [261] B. Marculescu, S. Poulding, R. Feldt, K. Petersen, and R. Torkar. Tester interactivity makes a difference in search-based software testing: A controlled experiment. *Information and Software Technology*, 78:66–82, 2016. (page 52).
- [262] A. Martens, H. Koziolk, L. Prechelt, and R. Reussner. From monolithic to component-based performance evaluation of software architectures. *Empirical Softw. Engg.*, 16(5):587–622, oct 2011. (page 52).
- [263] Y. Martínez, C. Cachero, and S. Meliá. Mdd vs. traditional software development: A practitioner’s subjective perspective. *Information and Software Technology*, 55:189 – 200, 02 2013. (page 52).

- [264] M. Matias Urbieto, L. Leandro Antonelli, G. Gustavo Rossi, and J. Sampaio do Prado Leite. The impact of using a domain language for an agile requirements management. *Information and Software Technology*, 127:106375, 2020. (pages 52, 59).
- [265] R. Maxion and R. Olszewski. Eliminating exception handling errors with dependability cases: A comparative, empirical study. *Software Engineering, IEEE Transactions on*, 26:888 – 906, 10 2000. (page 52).
- [266] I. Mcchesney and R. Bond. Eye tracking analysis of computer program comprehension in programmers with dyslexia. *Empirical Softw. Engg.*, 24(3):1109–1154, jun 2019. (page 52).
- [267] A. Medina, D. Nogueira, A. Hernández-Nariño, and R. Comas. Procedimiento para la gestión por procesos: métodos y herramientas de apoyo. *Ingeniare. Revista chilena de ingeniería*, 27:328 – 342, 04 2019. (page 87).
- [268] D. Méndez, D. Graziotin, S. Wagner, and H. Seibold. Open science in software engineering. *Contemporary Empirical Methods in Software Engineering*, pages 477–501, 2020. (pages 5, 7, 12, 38).
- [269] D. Méndez, M. Monperrus, R. Feldt, and T. Zimmermann. The open science initiative of the Empirical Software. *Empirical Software Engineering*, 1(24):1057–1060, 2019. (pages 8, 9).
- [270] P. Meso, G. Madey, M. Troutt, and J. Liegle. The knowledge management efficacy of matching information systems development methodologies with application characteristics—an experimental study. *Journal of Systems and Software*, 79(1):15–28, 2006. (page 52).
- [271] P. Mian, G. Travassos, A. Regina, and C. Da Rocha. A Computerized Infrastructure for Supporting Experimentation in Software Engineering. In *ESELAW 2005*, 2005. (pages 31, 32).
- [272] F. Miedema. *Open Science: the Very Idea*. Springer Dordrecht, 2022. (page 37).
- [273] R. Mohanani, B. Turhan, and P. Ralph. Requirements framing affects design creativity. *IEEE Transactions on Software Engineering*, 47(5):936–947, 2021. (page 52).
- [274] M. Monperrus. How to make a good open-science repository? <https://bit.ly/3HWafjy>, December 2019. (page 94).

- [275] M. Monperrus and T. Menzies. List of known research artifact types. <https://bit.ly/3AartaR>, January 2018. (page 94).
- [276] M. Morales, E. Navarro, P. Sánchez, and D. Alonso. A controlled experiment to evaluate the understandability of kaos and i* for modeling teleo-reactive systems. *Journal of Systems and Software*, 100:1–14, 2015. (page 52).
- [277] R. Morales, F. Khomh, and G. Antoniol. Repor: Mimicking humans on refactoring tasks. are we there yet? *Empirical Software Engineering*, 25, 07 2020. (page 52).
- [278] S. Mouchawrab, L. Briand, Y. Labiche, and M. Di Penta. Assessing, comparing, and combining state machine-based testing and structural testing: A series of experiments. *IEEE Transactions on Software Engineering*, 37(2):161–187, 2011. (page 52).
- [279] A. Mukhtar, B. Hofer, D. Jannach, and F. Wotawa. Spreadsheet debugging: The perils of tool over-reliance. *Journal of Systems and Software*, 184:111119, 2022. (page 52).
- [280] M. Müller. Two controlled experiments concerning the comparison of pair programming to peer review. *Journal of Systems and Software*, 78(2):166–179, 2005. (page 52).
- [281] M. Munafò, B. Nosek, D. Bishop, K. Button, C. Chambers, N. Percie Du Sert, U. Simonsohn, E. Wagenmakers, J. Ware, and J. Ioannidis. A manifesto for reproducible science. *Nature Human Behaviour* 2017 1:1, 1(1):1–9, jan 2017. (page 37).
- [282] G.C. Murphy. Beyond integrated development environments: Adding context to software development. In *Beyond integrated development environments: Adding context to software development*, pages 73–76. Institute of Electrical and Electronics Engineers Inc., 2019. cited By 12. (page 75).
- [283] E. Murphy-Hill, J. Dicker, M. Hodges, C. Egelman, C. Jaspan, L. Cheng, E. Kammer, B. Holtz, M. Jorde, A. Knight, and C. Green. Engineering impacts of anonymous author code review: A field experiment. *IEEE Transactions on Software Engineering*, pages 1–1, 2021. (page 52).
- [284] N. Mustafa, Y. Labiche, and D. Towey. Mitigating threats to validity in empirical software engineering: A traceability case study. In *2019 IEEE 43rd Annual Com-*

- puter Software and Applications Conference (COMPSAC)*, volume 2, pages 324–329, 2019. (page 39).
- [285] P. Mäder, T. Kuschke, and M. Janke. Reactive auto-completion of modeling activities. *IEEE Transactions on Software Engineering*, 47(7):1431–1451, 2021. (page 52).
- [286] M. Mäntylä and J. Itkonen. More testers – the effect of crowd size and time restriction in software testing. *Information and Software Technology*, 55:986–1003, 06 2013. (page 52).
- [287] D. Méndez, M. Monperrus, R. Feldt, and T. Zimmermann. The open science initiative of the empirical software engineering journal. *Empirical Software Engineering*, 24:1057–1060, 2019. (page 95).
- [288] K. Nafi, B. Roy, C. Roy, and K. Schneider. A universal cross language software similarity detector for open source software categorization. *Journal of Systems and Software*, 162, 2020. (page 52).
- [289] M. Nasir, N. Ikram, and Z. Jalil. Usability inspection: Novice crowd inspectors versus expert. *Journal of Systems and Software*, 183:111122, 2022. (page 52).
- [290] J. Natt Och Dag, T. Thelin, and B. Regnell. An experiment on linguistic tool support for consolidation of requirements from multiple sources in market-driven product development. *Empirical Software Engineering*, 11(2):303–329, 2006. (page 52).
- [291] D. Navarro and D. Foxcroft. Learning statistics with jamovi: A tutorial for psychology students and other beginners (version 0.70), 2018. (page 40).
- [292] T. Ng, Y. Yu, S. Cheung, and W. Chan. Human and program factors affecting the maintenance of programs with deployed design patterns. *Information and Software Technology*, 54:99–118, 01 2012. (page 52).
- [293] V. Nguyen, B. Boehm, and P. Danphitsanuphan. A controlled experiment in assessing and estimating software maintenance tasks. *Information and Software Technology*, 53:682–691, 06 2011. (page 52).
- [294] M. Nielsen. *Reinventing discovery: The new era of networked science*. Princeton University Press, 2012. (page 37).

- [295] NISO. Reproducibility badging and definitions. Technical report, National Information Standards Organization (NISO), 2021. (pages 9, 81, 94, 111, 145).
- [296] A. Nugroho. Level of detail in uml models and its impact on model comprehension: A controlled experiment. *Information and Software Technology*, 51:1670–1685, 12 2009. (page 52).
- [297] M. Ochodek, B. Alchimowicz, J. Jurkiewicz, and J. Nawrocki. Improving the reliability of transaction identification in use cases. *Information and Software Technology*, 53:885–897, 08 2011. (page 52).
- [298] Organization International of Legal Metrology. International vocabulary of metrology – basic and general concepts and associated terms (vim). <https://bit.ly/31YLSRa>, 2007. (page 9).
- [299] F. Olav Bjørnson, A. Inge Wang, and E. Arisholm. Improving the effectiveness of root cause analysis in post mortem analysis: A controlled experiment. *Information and Software Technology*, 51(1):150–161, 2009. Special Section - Most Cited Articles in 2002 and Regular Research Papers. (page 52).
- [300] R. Oliveira, R. de Mello, E. Fernandes, A. Garcia, and C. Lucena. Collaborative or individual identification of code smells? on the effectiveness of novice and professional developers. *Information and Software Technology*, 120:106242, 12 2019. (page 52).
- [301] R. Oliveira, R. de Mello, E. Fernandes, A. Garcia, and C. Lucena. Collaborative or individual identification of code smells? on the effectiveness of novice and professional developers. *Information and Software Technology*, 120:106242, 2020. (page 52).
- [302] J. Olsson, E. Risfelt, T. Besker, A. Martini, and R. Torkar. Measuring affective states from technical debt: A psychoempirical software engineering experiment. *Empirical Software Engineering*, 26, 09 2021. (page 52).
- [303] A. Opdahl and G. Sindre. Experimental comparison of attack trees and misuse cases for security threat identification. *Information and Software Technology*, 51:916–932, 05 2009. (page 52).
- [304] Open Science IEEE Computer Society. The ad hoc committee on open science and reproducibility chapter. <https://bit.ly/3P7272e>, 2020. (page 60).

- [305] M. Otero and J. Dolado. Evaluation of the comprehension of the dynamic modeling in uml. *Information and Software Technology*, 46(1):35–53, 2004. (page 52).
- [306] M. Otero and J. Dolado. An empirical comparison of the dynamic modeling in oml and uml. *Journal of Systems and Software*, 77(2):91–102, 2005. (page 52).
- [307] A. Ottensooser, A. Fekete, H. Reijers, J. Mendling, and C. Menictas. Making sense of business process descriptions: An experimental comparison of graphical and textual notations. *Journal of Systems and Software*, 85(3):596–606, 2012. Novel approaches in the design and implementation of systems/software architecture. (page 52).
- [308] M. Paltenghi and M. Pradel. Thinking like a developer? comparing the attention of humans with neural models of code. In *2021 36th IEEE/ACM International Conference on Automated Software Engineering. (ASE)*, 2021. (page 52).
- [309] M. Pan, Y. Lu, Y. Pei, T. ZHANG, J. Zhai, and X. Li. Effective testing of android apps using extended ifml models. *Journal of Systems and Software*, 159:110433, 10 2020. (page 52).
- [310] I. Panach, S. España, O. Dieste, O. Pastor, and N. Juristo. In search of evidence for model-driven development claims: An experiment on quality, effort, productivity and satisfaction. *Information and Software Technology*, 62, 02 2015. (page 52).
- [311] J. Panach, Ó. Dieste, B. Marín, S. España, S. Vegas, Ó. Pastor, and N. Juristo. Evaluating model-driven development claims with respect to quality: A family of experiments. *IEEE Transactions on Software Engineering*, 47(1):130–145, 2018. (page 52).
- [312] S Panagiotis, S. Ioannis, A. Lefteris, and D. Ignatios. An experimental investigation of personality types impact on pair effectiveness in pair programming. *Empirical Softw. Engg.*, 14(2):187–226, apr 2009. (page 52).
- [313] M. Pancur and M. Ciglarič. Impact of test-driven development on productivity, code and tests: A controlled experiment. *Information and Software Technology*, 53:557–573, 06 2011. (page 52).
- [314] B. Papis, K. Grochowski, K. Subzda, and K. Sijko. Experimental evaluation of test-driven development with interns working on a real industrial project. *IEEE Transactions on Software Engineering*, pages 1–1, 2020. (page 52).

- [315] P. Paulweber, G. Simhandl, and U. Zdun. On the understandability of language constructs to structure the state and behavior in abstract state machine specifications: A controlled experiment. *Journal of Systems and Software*, 178:110987, 2021. (page 52).
- [316] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee. A design science research methodology for information systems research. *Journal of management information systems*, 24(3):45–77, 2007. (page 15).
- [317] N. Peitek, S. Apel, C. Parnin, A. Brechmann, and J. Siegmund. Program comprehension and code complexity metrics: An fmri study. In *2021 IEEE/ACM 43rd International Conference on Software Engineering (ICSE)*, pages 524–536, 2021. (page 52).
- [318] M. Peleg and D. Dori. The model multiplicity problem: Experimenting with real-time specification methods. *Software Engineering, IEEE Transactions on*, 26:742 – 759, 09 2000. (page 52).
- [319] R. Pereira, T. Carção, M. Couto, C. Jácome, J. Fernandes, and J. Saraiva. Spelling out energy leaks: Aiding developers locate energy inefficient code. *Journal of Systems and Software*, 161:110463, 2020. (page 52).
- [320] M. Perepletchikov and C. Ryan. A controlled experiment for evaluating the impact of coupling on the maintainability of service-oriented software. *IEEE Transactions on Software Engineering*, 37(4):449–465, 2011. (page 52).
- [321] K. Petersen and C. Wohlin. Context in industrial software engineering research. In *2009 3rd International Symposium on Empirical Software Engineering and Measurement*, pages 401–404. IEEE, 2009. (page 75).
- [322] Dietmar Pfahl, Oliver Laitenberger, Guenther Ruhe, Jörg Dorsch, and Tatyana Krivobokova. Evaluating the learning effectiveness of using simulations in software project management education: Results from a twice replicated experiment. *Information and Software Technology*, 46:127–147, 02 2004. (page 52).
- [323] S. Pfleeger. Experimental design and analysis in software engineering. *Annals of Software Engineering*, 1(1):219–253, 1995. (page 25).
- [324] Y. Pham, A. Bouraffa, M. Hillen, and W. Maalej. The role of linguistic relativity on the identification of sustainability requirements: An empirical study. In *2021*

- IEEE 29th International Requirements Engineering Conference (RE)*, pages 117–127, 2021. (page 52).
- [325] G. Polančič, G. Jošt, and M. Heričko. An experimental investigation comparing individual and collaborative work productivity when using desktop and cloud modeling tools. *Empirical Softw. Engg.*, 20(1):142–175, feb 2013. (page 52).
- [326] C. Politowski, F. Khomh, S. Romano, G. Scanniello, F. Petrillo, Y. Guéhéneuc, and A. Maiga. A large scale empirical study of the impact of spaghetti code and blob anti-patterns on program comprehension. *Information and Software Technology*, 122:106278, 2020. (page 52).
- [327] K. Popper. *The Logic of Scientific Discovery*. Hutchinson & Co, 1959. (page 29).
- [328] L. Prechelt. Plat forms: A web development platform comparison by an exploratory experiment searching for emergent platform properties. *IEEE Transactions on Software Engineering*, 37(1):95–108, 2011. (pages 52, 59).
- [329] L. Prechelt and B. Unger. An experiment measuring the effects of personal software process (psp) training. *IEEE Transactions on Software Engineering*, 27(5):465–472, 2001. (page 52).
- [330] L. Prechelt, B. Unger, M. Michael, and W. Tichy. A controlled experiment on inheritance depth as a cost factor for code maintenance. *Journal of Systems and Software*, 65(2):115–126, 2003. (page 52).
- [331] L. Prechelt, B. Unger, W. Tichy, P. Brossler, and L. Votta. A controlled experiment in maintenance: comparing design patterns to simpler solutions. *Software Engineering, IEEE Transactions on*, 27:1134–1144, 01 2001. (page 52).
- [332] L. Prechelt, B. Unger-Lamprecht, M. Philippsen, and W. Tichy. Two controlled experiments assessing the usefulness of design pattern documentation in program maintenance. *IEEE Transactions on Software Engineering*, 28(6):595–606, 2002. (page 52).
- [333] A. Przybyłek. An empirical study on the impact of aspectj on software evolvability. *Empirical Softw. Engg.*, 23(4):2018–2050, aug 2018. (page 52).
- [334] J. Pucci, L. Passos, R. García, R. Messias, and C. Olivete. Exptool : a tool to conduct , package and replicate controlled experiments in software engineering. In *ExpTool : a Tool to Conduct , Package and Replicate Controlled Experiments in Software Engineering*, 2014. (page 30).

- [335] F. Pérez, J. Echeverría, R. Lapeña, and C. Cetina. Comparing manual and automated feature location in conceptual models: A controlled experiment. *Information and Software Technology*, 125:106337, 2020. (page 52).
- [336] A. Quiña-Mera, P. Fernández-Montes, J. García, E. Bastidas, and A. Ruiz-Cortés. Quality in use evaluation of a graphql implementation. In M. Botto-Tobar, H. Cruz, A. Díaz, and B. Durakovic, editors, *Emerging Research in Intelligent Systems*, pages 15–27, Cham, 2022. Springer International Publishing. (page 180).
- [337] R. Ramac, B. Turhan, I. Karac, N. Juristo, and V. Mandic. Lessons learned from a partial replication of an experiment in the context of a software engineering course. In *International Scientific Conference on Industrial Systems (IS)*, pages 198–203, 2017. (page 29).
- [338] N. Ramasubbu, C. Kemerer, and J. Hong. Structural complexity and programmer team strategy: An experimental test. *IEEE Transactions on Software Engineering*, 38(5):1054–1068, 2012. (page 52).
- [339] RE'2021. AE-Artifact Evaluation. <https://bit.ly/3rxWdgX>, 2021. (pages 56, 94, 140).
- [340] I. Reinhartz-berger, K. Figl, and Ø. Haugen. Investigating styles in variability modeling: Hierarchical vs. constrained styles. *Information and Software Technology*, 87, 02 2017. (page 52).
- [341] I. Reinhartz-Berger and A. Sturm. Comprehensibility of uml-based software product line specifications. *Empirical Softw. Engg.*, 19(3):678–713, jun 2014. (page 52).
- [342] H. Reis and C. Judd. *Handbook of research methods in social and personality psychology*. Cambridge: Cambridge University Press, 2014. (page 39).
- [343] K. Revoredo, D. Djurica, and J. Mendling. A study into the practice of reporting software engineering experiments. *Empirical Software Engineering*, 26(113), 2021. (pages 16, 25, 46, 85).
- [344] M. Riaz, J. King, J. Slankas, L. Williams, F. Massacci, C. Quesada-López, and M. Jenkins. Identifying the implied: Findings from three differentiated replications on the use of security requirements templates. *Empirical Softw. Engg.*, 22(4):2127–2178, aug 2017. (page 52).

- [345] F. Ricca, M. Di Penta, M. Torchiano, P. Tonella, and M. Ceccato. How developers' experience and ability influence web application comprehension tasks supported by uml stereotypes: A series of four experiments. *IEEE Transactions on Software Engineering*, 36(1):96–118, 2010. (page 52).
- [346] F. Ricca, M. Torchiano, M. Di Penta, M. Ceccato, and P. Tonella. Using acceptance tests as a support for clarifying requirements: A series of experiments. *Information and Software Technology*, 51:270–283, 02 2009. (page 52).
- [347] F. Ricca, M. Torchiano, M. Leotta, A. Tiso, G. Guerrini, and G. Reggio. On the impact of state-based model-driven development on maintainability: A family of experiments using unimod. *Empirical Softw. Engg.*, 23(3):1743–1790, jun 2018. (pages 52, 59).
- [348] N. Ritschel, V. Kovalenko, R. Holmes, R. Garcia, and D. Shepherd. Comparing block-based programming models for two-armed robots. *IEEE Transactions on Software Engineering*, pages 1–1, 2020. (page 52).
- [349] M. Robillard, M. Nassif, and S. McIntosh. Threats of aggregating software repository data. In *Proceedings - 2018 IEEE International Conference on Software Maintenance and Evolution, ICSME 2018*, pages 508–518. IEEE, 2018. (page 31).
- [350] S. Romano, C. Vendome, G. Scanniello, and D. Poshyvanyk. A multi-study investigation into dead code. *IEEE Transactions on Software Engineering*, 46(1):71–99, 2020. (page 52).
- [351] M. Roper, M. Wood, and J. Miller. An empirical evaluation of defect detection techniques. *Information and Software Technology*, 39(11):763–775, 1997. Evaluation and Assessment in Software Engineering. (page 31).
- [352] Royal Society. *Philosophical Transactions of the Royal Society of London*, 1665. (page 37).
- [353] Royal Society. *Science as an open enterprise*. Royal Society, 2012. (pages 37, 38).
- [354] P. Runeson and M. Höst. Guidelines for conducting and reporting case study research in software engineering. *Empirical software engineering*, 14(2):131, 2009. (page 165).
- [355] P. Runeson, M. Höst, A. Rainer, and B. Regnell. Case study research in software engineering. In *Guidelines and examples*. Wiley Online Library, 2012. (pages XVI, 165, 166).

- [356] P. Runeson, A. Stefik, and A. Andrews. Variation factors in the design and analysis of replicated controlled experiments: Three (dis)similar studies on inspections versus unit testing. *Empirical Software Engineering*, 19(6):1781–1808, oct 2014. (page 31).
- [357] G. Sabaliauskaite, S. Kusumoto, and K. Inoue. Assessing defect detection performance of interacting teams in object-oriented design inspection. *Information and Software Technology*, 46:875–886, 10 2004. (page 52).
- [358] G. Sabaliauskaite, F. Matsukawa, S. Kusumoto, and K. Inoue. Further investigations of reading techniques for object-oriented design inspection. *Information and Software Technology*, 45:571–585, 06 2003. (page 52).
- [359] L. Sabatucci, M. Ceccato, A. Marchetto, and A. Susi. Ahab’s legs in scenario-based requirements validation: An experiment to study communication mistakes. *Journal of Systems and Software*, 109:124–136, 2015. (page 52).
- [360] W. Said, J. Quante, and R. Koschke. Mining understandable state machine models from embedded code. *Empirical Software Engineering*, 25:1–46, 11 2020. (page 52).
- [361] V. Sakhnini, L. Mich, and D. Berry. Group versus individual use of power-only epmcreate as a creativity enhancement technique for requirements elicitation. *Empirical Softw. Engg.*, 22(4):2001–2049, aug 2017. (page 52).
- [362] I. Salman, B. Turhan, and S. Vegas. A controlled experiment on time pressure and confirmation bias in functional software testing. *Empirical Softw. Engg.*, 24(4):1727–1761, aug 2019. (page 52).
- [363] A. Santos, O. S. Gómez, and N. Juristo. Analyzing families of experiments in se: a systematic mapping study. *IEEE Transactions on Software Engineering*, 2018. (pages 4, 27).
- [364] A. Santos, S. Vegas, O. Dieste, F. Uyaguari, A. Tosun, D. Fucci, B. Turhan, G. Scanniello, S. Romano, I. Karac, M. Kuhrmann, V. MandiÄ†, R. RamaÄ†, D. Pfahl, C. Engblom, J. Kyykka, K. Rungi, C. Palomeque, J. Spisak, M. Oivo, and N. Juristo. A family of experiments on test-driven development. *Empirical Software Engineering*, pages 1573–7616, 2021. (page 52).
- [365] A. Santos, S. Vegas, M. Oivo, and N. Juristo. A Procedure and Guidelines for Analyzing Groups of Software Engineering Replications. *IEEE Transactions on Software Engineering*, 5589(FEBRUARY):1–1, 2019. (pages 27, 31).

- [366] M. Sayagh, N. Kerzazi, F. Petrillo, K. Bennani, and B. Adams. What should your run-time configuration framework do to help developers? *Empirical Software Engineering*, 25:1259–1293, 2020. (page 52).
- [367] G. Scanniello, C. Gravino, M. Genero, J. Cruz-Lemus, G. Tortora, M. Risi, and G. Doderio. Do software models based on the uml aid in source-code comprehensibility? aggregating evidence from 12 controlled experiments. *Empirical Softw. Engg.*, 23(5):2695–2733, oct 2018. (page 52).
- [368] G. Scoccia, I. Malavolta, M. Autili, A. Di Salle, and P. Inverardi. Enhancing trustability of android applications via user-centric flexible permissions. *IEEE Transactions on Software Engineering*, 47(10):2032–2051, 2021. (page 52).
- [369] C. Seaman. Qualitative methods in empirical studies of software engineering. *IEEE Transactions on software engineering*, 25(4):557–572, 1999. (page 76).
- [370] K. Sedighiani, S. Shokrollahi, and F. Shams. Basba: A framework for building adaptable service-based applications. *Journal of Systems and Software*, 179:110989, 09 2021. (page 52).
- [371] G. Sedrakyan, S. Poelmans, and M. Snoeck. Assessing the influence of feedback-inclusive rapid prototyping on understanding the semantics of parallel uml statecharts by novice modellers. *Information and Software Technology*, 82, 04 2017. (page 52).
- [372] W. Shadish, T. Cook, and D. Campbell. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Unknown Publisher, 2002. (page 39).
- [373] B. Shen, W. Zhang, C. Kästner, H. Zhao, Z. Wei, G. Liang, and Z. Jin. Smart-commit: A graph-based interactive assistant for activity-oriented commits. In *SmartCommit: A Graph-Based Interactive Assistant for Activity-Oriented Commits*, ESEC/FSE 2021, page 379–390, New York, NY, USA, 2021. Association for Computing Machinery. (page 52).
- [374] Q. Shen, S. Wu, Y. Zou, Z. Zhu, and B. Xie. From api to nli: A new interface for library reuse. *Journal of Systems and Software*, 169:110728, 2020. (page 52).
- [375] M. Shepperd. Replication studies considered harmful. In *2018 IEEE/ACM 40th International Conference on Software Engineering: New Ideas and Emerging Technologies Results (ICSE-NIER)*, pages 73–76. IEEE, 2018. (page 75).

- [376] M. Shepperd, N. Aijenka, and S. Counsell. The role and value of replication in empirical software engineering results. *Information and Software Technology*, 99:120–132, 2018. (page 5).
- [377] L. Shi, X. Chen, Y. Yang, H. Jiang, Z. Jiang, N. Niu, and Q. Wang. A first look at developers' live chat on gitter. In *A First Look at Developers' Live Chat on Gitter*, ESEC/FSE 2021, page 391–403, New York, NY, USA, 2021. Association for Computing Machinery. (page 52).
- [378] L. Shi, Z. Jiang, Y. Yang, X. Chen, Y. Zhang, F. Mu, H. Jiang, and Q. Wang. Ispy: Automatic issue-solution pair extraction from community live chats. In *2021 36th IEEE/ACM International Conference on Automated Software Engineering (ASE)*, 09 2021. (page 52).
- [379] B. Shneiderman. Computer science: Science 2.0. *Science*, 319(5868):1349–1350, mar 2008. (page 37).
- [380] F. Shull, V. Basili, J. Carver, J. Maldonado, G. Travassos, M. Mendonça, and Fabbri S. Replicating software engineering experiments: Addressing the tacit knowledge problem. In *Proc. In International Symposium on Empirical Software Engineering*, pages 7–16, 2002. (pages 30, 69, 72, 73).
- [381] F. Shull, J. Carver, S. Vegas, and N. Juristo. The role of replications in empirical software engineering. *Empirical Software Engineering*, 13(2):211–218, 2008. (pages 5, 6, 12, 26, 31, 35, 36, 39, 70, 71, 100).
- [382] F. Shull, M. Mendonça, V. Basili, J. Carver, J. C. Maldonado, S. Fabbri, G. Travassos, and M. Ferreira. Knowledge-sharing issues in experimental software engineering. *Empirical Software Engineering*, 9(1-2):111–137, 2004. (pages 6, 11, 31, 34, 36, 70, 100).
- [383] J. Siegmund, C. Kästner, J. Liebig, S. Apel, and S. Hanenberg. Measuring and modeling programming experience. *Empirical Softw. Engg.*, 19(5):1299–1334, oct 2014. (pages 52, 59).
- [384] J. Siegmund, N. Siegmund, and S. Apel. Views on internal and external validity in empirical software engineering. In *Software Engineering (ICSE), 2015 IEEE/ACM 37th IEEE International Conference on*, volume 1, pages 9–19. IEEE, 2015. (page 40).

- [385] R. Silva, C. Roy, M. Rahman, K. Schneider, K. Paixão, C. Dantas, and M. Maia. Crokage: effective solution recommendation for programming tasks by leveraging crowd knowledge. *Empirical Software Engineering*, 25:1–52, 11 2020. (page 52).
- [386] S. Silvia Rueda, J. Panach, and D. Distanto. Requirements elicitation methods based on interviews in comparison: A family of experiments. *Information and Software Technology*, 126:106361, 2020. (page 52).
- [387] F. Siqueira. Comparing the comprehensibility of requirements models: an experiment replication. *Information and Software Technology*, 96, 11 2018. (page 52).
- [388] D. I. K. Sjøberg, T. Dyba, and M. Jorgensen. The future of empirical methods in software engineering research. In *2007 Future of Software Engineering*, pages 358–378. IEEE Computer Society, 2007. (page 75).
- [389] D. I. K. Sjøberg, J. E. Hannay, O. Hansen, V. B. Kampenes, A. Karahasanovic, N. K. Liborg, and A. C. Rekdal. A survey of controlled experiments in software engineering. *Software Engineering, IEEE Transactions on*, 31(9):733–753, 2005. (page 4).
- [390] C. Snook and R. Harrison. Experimental comparison of the comprehensibility of a z specification and its implementation in java. *Information and Software Technology*, 46, 11 2004. (page 52).
- [391] British Psychological Society. Guidance: Ethical principles for conducting research with human participants. *American Psychologist*, 2021. (page 86).
- [392] M. Solari. *Propuesta de paquete de laboratorio para experimentos de Ingeniería del Software*. PhD thesis, Universidad Politécnica de Madrid, Madrid, 2011. (pages xvii, 6, 26, 31, 33, 34, 82, 92, 99, 100, 171, 178).
- [393] M. Solari and S. Vegas. Classifying and analysing replication packages for software engineering experimentation. In *7th International Conference on Product Focused Software Process Improvement (PROFES 2006)-Workshop Series in Empirical Software Engineering (WSESE)*, 2006. (pages 5, 6, 12, 31, 35, 36, 70, 82, 95, 99, 100, 140).
- [394] M. Solari, S. Vegas, and N. Juristo. Content and structure of laboratory packages for software engineering experiments. *Information and Software Technology*,

- 97(December 2016):64–79, 2018. (pages xv, 5, 6, 12, 14, 30, 31, 36, 41, 45, 48, 49, 57, 58, 60, 63, 64, 68, 69, 70, 71, 72, 76, 92, 95, 100, 142).
- [395] M. Soltani, A. Panichella, and A. van Deursen. Search-based crash reproduction and its impact on debugging. *IEEE Transactions on Software Engineering*, 46(12):1294–1317, 2020. (page 52).
- [396] E. Souza, A. Moreira, J. Araújo, S. Abrahão, E. Insfran, and D. Silveira. Comparing business value modeling methods: A family of experiments. *Information and Software Technology*, 104, 08 2018. (page 52).
- [397] J. Spence and D. Stanley. Prediction interval: What to expect when you’re expecting... a replication. *PloS one*, 11(9):e0162874, 2016. (page 5).
- [398] M. Staron, L. Kuzniarz, and C. Wohlin. Empirical assessment of using stereotypes to improve comprehension of uml models: A set of experiments. *Journal of Systems and Software*, 79(5):727–742, 2006. Quality Software. (page 52).
- [399] M Storey, N. Ernst, C. Williams, and E. Kalliamvakou. The who, what, how of software engineering research: a socio-technical framework. *Empirical Software Engineering*, 25(5):4097–4129, 2020. (page 75).
- [400] G. Tebes, D. Peppino, M. Rivera, P. Becker, F. Papa, and L. Olsina. Especificación del proceso de design science research: Caso aplicado a una ontología de testing de software. In *7mo Congreso Nacional de Ingeniería Informática-Sistemas de Información (CoNaIISI)At: La Matanza, BsAs, Argentina*, 11 2019. (pages xv, 15).
- [401] E. Teixeira, L. Fonseca, and S. Soares. Threats to validity in controlled experiments in software engineering: What the experts say and why this is relevant. In *Proceedings of the XXXII Brazilian Symposium on Software Engineering, SBES '18*, page 52–61, New York, NY, USA, 2018. Association for Computing Machinery. (page 39).
- [402] S. Teixeira, B. Agrizzi, J. Pereira, S. Rossetto, I. Pereira, P. Costa, A. Branco, and R. Martinelli. Laura architecture: Towards a simpler way of building situation-aware and business-aware iot applications. *Journal of Systems and Software*, 161:110494, 12 2020. (page 52).
- [403] M. Teruel, E. Navarro, V. López-Jaquero, F. Montero, J. Jaen, and P. González. Analyzing the understandability of requirements engineering languages for

- csw systems: A family of experiments. *Information and Software Technology*, 54:1215–1228, 11 2012. (page 52).
- [404] T. Thelin, P. Runeson, and B. Regnell. Usage-based reading—an experiment to guide reviewers with use cases. *Information and Software Technology*, 43:925–938, 12 2001. (page 52).
- [405] T. Thelin, P. Runeson, and C. Wohlin. An experimental comparison of usage-based and checklist-based reading. *IEEE Transactions on Software Engineering*, 29(8):687–704, 2003. (page 52).
- [406] C. Timperley, L. Herckis, C. Le Goues, and M. Hilton. Understanding and improving artifact sharing in software engineering research. *Empirical Software Engineering*, 26(4):1573–7616, 2021. (pages 5, 6, 60, 76).
- [407] A. Tosun, O. Dieste, D. Fucci, S. Vegas, B. Turhan, H. Erdogmus, A. Santos, M. Oivo, K. Toro, J. Jarvinen, and N. Juristo. An industry experiment on the effects of test-driven development on external quality and productivity. *Empirical Softw. Engg.*, 22(6):2763–2805, dec 2017. (page 52).
- [408] A. Tosun, O. Dieste, S. Vegas, D. Pfahl, K. Rungi, and N. Juristo. Investigating the impact of development task on external quality in test-driven development: An industry experiment. *IEEE Transactions on Software Engineering*, 47(11):2438–2456, 2019. (page 52).
- [409] Y. Tu, E. Tempero, and C. Thomborson. An experiment on the impact of transparency on the effectiveness of requirements documents. *Empirical Software Engineering*, 21, 02 2016. (page 52).
- [410] K. Tuma, C. Sandberg, U. Thorsson, M. Widman, T. Herpel, and R. Scandariato. Finding security threats that matter: Two industrial case studies. *Journal of Systems and Software*, 2021. (page 52).
- [411] National Information Standards Organization (Estados Unidos). ANSI/NISO z39.84 : Syntax for the Digital Object Identifier. <http://hdl.handle.net/10421/8954>, 2005. (page 121).
- [412] P. Valderas, V. Torres, and V. Pelechano. A microservice composition approach based on the choreography of bpmn fragments. *Information and Software Technology*, 127:106370, 2020. (page 52).

- [413] D. van der Linden, E. Williams, J. Hallett, and A. Rashid. The impact of surface features on choice of (in)secure answers by stackoverflow readers. *IEEE Transactions on Software Engineering*, pages 1–1, 2020. (page 52).
- [414] C. Vassallo, S. Panichella, F. Palomba, S. Proksch, H. Gall, and A. Zaidman. How developers engage with static analysis tools in different contexts. *Empirical Software Engineering*, 25, 03 2020. (page 52).
- [415] C. Vassallo, S. Proksch, T. Zemp, and H. Gall. Every build you break: Developer-oriented assistance for build failure resolution. *Empirical Software Engineering*, 25, 05 2020. (page 52).
- [416] S. Vegas, N. Juristo, A. Moreno, M. Solari, and P. Letelier. Analysis of the influence of communication between researchers on experiment replication. In *ISESE'06 - Proceedings of the 5th ACM-IEEE International Symposium on Empirical Software Engineering*, volume 2006, pages 28–37, 2006. (pages 6, 12, 30, 35, 36, 70, 71, 100).
- [417] S. Vegas, P. Riofrío, E. Marcos, and N. Juristo. On (mis)perceptions of testing effectiveness: an empirical study. *Empirical Software Engineering*, 25, 07 2020. (page 52).
- [418] M. Velando. *La calidad y la ingeniería de software*, 1997. (page 91).
- [419] V. Venkatesh, S. Brown, and H. Bala. Bridging the qualitative-quantitative divide: Guidelines for conducting mixed methods research in information systems. *MIS Quarterly: Management Information Systems*, 37(1):21–54, 2013. (page 46).
- [420] J. Verelst. The influence of the level of abstraction on the evolvability of conceptual models of information systems. *Empirical Software Engineering*, 10:467–494, 2005. (page 52).
- [421] P. Vitharana and K. Ramamurthy. Computer-mediated group support, anonymity, and the software inspection process: an empirical investigation. *IEEE Transactions on Software Engineering*, 29(2):167–180, 2003. (page 52).
- [422] A. Viticchié, L. Regano, C. Basile, M. Torchiano, M. Ceccato, and P. Tonella. Empirical assessment of the effort needed to attack programs protected with client/server code splitting. *Empirical Software Engineering*, 25, 01 2020. (page 52).
- [423] R. Von Alan, S. March, J. Park, and S. Ram. Design science in information systems research. *MIS quarterly*, 28(1):75–105, 2004. (page 15).

- [424] I. von Nostitz-Wallwitz, J. Krüger, and T. Leich. Towards improving industrial adoption: The choice of programming languages and development environments. In *Proceedings of the 5th International Workshop on Software Engineering Research and Industrial Practice, SER&IP '18*, page 10–17, New York, NY, USA, 2018. Association for Computing Machinery. (page 73).
- [425] I. von Nostitz-Wallwitz, J. Krüger, J. Siegmund, and T. Leich. Knowledge transfer from research to industry: A survey on program comprehension. In *Knowledge Transfer from Research to Industry: A Survey on Program Comprehension, ICSE '18*, page 300–301, New York, NY, USA, 2018. Association for Computing Machinery. (page 73).
- [426] C. von Wangenheim, M. Thiry, and D. Kochanski. Empirical evaluation of an educational game on software measurement. *Empirical Software Engineering*, 14:418–452, 2009. (page 52).
- [427] H. Vrzakova, A. Begel, L. Mehtätalo, and R. Bednarik. Affect recognition in code review: An in-situ biometric study of reviewer's affect. *Journal of Systems and Software*, 159:110434, 10 2020. (page 52).
- [428] Z. Wang, H. Zhang, T. Chen, and S. Wang. Would you like a quick peek? providing logging support to monitor data processing in big data applications. In *Would You like a Quick Peek? Providing Logging Support to Monitor Data Processing in Big Data Applications, ESEC/FSE 2021*, page 516–526, New York, NY, USA, 2021. Association for Computing Machinery. (page 52).
- [429] R. Webb Collins, A. Hevner, G. Walton, and R. Linger. The impacts of function extraction technology on program comprehension: A controlled experiment. *Information and Software Technology*, 50:1165–1179, 10 2008. (page 52).
- [430] M. Webster and J. Sell. *Laboratory Experiments in the Social Sciences, Second Edition*. Elsevier, 2014. (page 4).
- [431] D. Weinberger. *Too Big to Know: Rethinking Knowledge Now That the Facts Aren't the Facts, Experts Are Everywhere, and the Smartest Person in the Room Is the Room*. Basic Books, Inc., USA, 2014. (page 37).
- [432] J. Wilkerson, J. Nunamaker, and R. Mercer. Comparing the defect reduction benefits of code inspection and test-driven development. *IEEE Transactions on Software Engineering*, 38(3):547–560, 2012. (page 52).

- [433] K. Wnuk, M. Höst, and B. Regnell. Replication of an experiment on linguistic tool support for consolidation of requirements from multiple sources. *Empirical Software Engineering*, 17(3):305–344, 2012. (page 52).
- [434] M. Woelfle, P. Olliaro, and M. Todd. Open science is a research accelerator. *Nature Chemistry* 2011 3:10, 3(10):745–748, sep 2011. (page 37).
- [435] C. Wohlin, P. Runeson, H. Martin, M. Ohlsson, B. Regnell, and A. Wesslén. *Experimentation in Software Engineering*. Springer Heidelberg New York Dordrecht London, London, 2012. (pages xv, 25, 29, 39, 70, 83, 84, 92).
- [436] A. Wright, J. Adler, C. DeYoung, C. Emily Durbin, R. Edelstein, C. Jordan, K. Locke, S. Luo, D. Lynam, S. von Stumm, V. Zeigler-Hill, and H. Tennen. Transparency and Open Science at the Journal of Personality. *Journal of Personality*, 89(2):171–174, apr 2021. (pages 38, 39).
- [437] D. Wu, X. Jing, H. Zhang, B. Li, Y. Xie, and B. Xu. Generating api tags for tutorial fragments from stack overflow. *Empirical Software Engineering*, 26, 07 2021. (page 52).
- [438] W. Wu, A. Serveaux, Y. Guéhéneuc, and G. Antoniol. The impact of imperfect change rules on framework api evolution identification: An empirical study. *Empirical Softw. Engg.*, 20(4):1126–1158, aug 2015. (page 52).
- [439] M. Zelkowitz and D. Wallace. Experimental Models for Validating Technology. *Computer*, 31(5):23–31, 1998. (page 26).
- [440] A. Zender, T. Pfeiffer, M. Eicks, and F. Lehner. Experimental comparison of coarse-grained concepts in uml, oml, and tos. *Journal of Systems and Software*, 57(1):21–30, 2001. (page 52).
- [441] Zenodo. Official website zenodo. <https://zenodo.org/>, 2021. (page 61).
- [442] J. Zettel. Methodology support in case tools and its impact on individual acceptance and use: A controlled experiment. *Empirical Software Engineering*, 10:367–394, 2005. (page 52).
- [443] J. Zhang, H. Jiang, Z. Ren, T. Zhang, and Z. Huang. Enriching api documentation with code samples and usage scenarios from crowd knowledge. *IEEE Transactions on Software Engineering*, 47(6):1299–1314, 2021. (page 52).

- [444] X. Zhou, Y. Jin, H. Zhang, S. Li, and X. Huang. A map of threats to validity of systematic literature reviews in software engineering. In *Proceedings - Asia-Pacific Software Engineering Conference, APSEC*, volume 0, pages 153–160. IEEE Computer Society, jul 2016. (page 40).
- [445] Y. Zhou, C. Wang, X. Yan, T. Chen, S. Panichella, and H. Gall. Automatic detection and repair recommendation of directive defects in java api documentation. *IEEE Transactions on Software Engineering*, 46(9):1004–1023, 2020. (page 52).
- [446] J. Zubcoff, I. Garrigós, S. Casteleyn, J. Mazón, J. Aguilar, and F. Castillo. Evaluating different i*-based approaches for selecting functional requirements while balancing and optimizing non-functional requirements: A controlled experiment. *Information and Software Technology*, 106, 09 2019. (page 52).