

# HEMEROGRAFÍA DIGITAL EN LA INVESTIGACIÓN DE LOS CIBERMEDIOS

**Javier Díaz Noci y Koldobika Meso Ayerdi**

**Facultad de Ciencias Sociales y de la Comunicación.**

**Dpto. Periodismo II. Universidad del País Vasco, España.**

## **1. Introducción**

En diez años de presencia de los medios de comunicación en Internet, los títulos se han multiplicado. Si al principio sólo eran unas decenas, pronto no hubo medio convencional que no pusiese también su versión en la Red. A ellos se sumaron otros medios completamente nuevos. Las propias características de la comunicación digital, el efecto palimpsesto de las informaciones que se dan a conocer en Internet, el cambio de título, razón social, URL, la desaparición y aparición constante de medios son algunas de las razones que hacen difícil un catálogo exacto y una conservación completa de todos los cibermedios.

Sin embargo, para investigar esos medios, se hace imprescindible disponer tanto de un catastro lo más completo y científico posible de los mismos, como de colecciones, a ser posibles accesibles a través de Internet (aunque también en soporte físico), que permitan conocer lo que se ha producido. En este último punto, es ya mucho lo que se ha perdido, y no siempre la política de las instituciones públicas o de las propias empresas que producen esos medios ha sido la más acertada para el investigador.

Expondremos aquí los principales problemas de la hemerografía de los cibermedios, así como las características de algunos de los trabajos que sobre esto se han hecho. El objetivo es que quien quiera investigar sobre ciberperiodismo sepa confeccionar su propia base de datos, donde recoja el universo de estudio, y enfrentarse a las carencias que ya existen en este campo.

## **2. Información en formato digital: Catalogación**

### *2.1. Problemas generales*

La primera característica de la información digital es que carece, en principio, de soporte físico, aunque pueda ser almacenada por medios materiales. Sin embargo, la

información que contienen esos soportes materiales debe ser descodificada mediante medios informáticos. Por otra parte, la mayoría de esa información digital fluye libremente por el ciberespacio; no sólo los medios de comunicación, también, por ejemplo, la comunicación epistolar en forma de correos electrónicos, o la comunicación interpersonal en forma de foros de debate, *chats* o respuestas en las bitácoras o *blogs*.

Planteamos así un primer problema de investigación: ¿Se están archivando de alguna forma, pública o privadamente, los epistolarios que con este nuevo medio están surgiendo, las “cartas” que muchísima gente se está cruzando? ¿Algo que nos permita hacer, como se ha hecho con las cartas escritas por gente importante, una edición posterior de los epistolarios?

Hace tiempo que la historiografía de la comunicación pretende superar una fase meramente positivista y heurística, sin que, por otra parte, dicho enfoque tenga nada de malo, ni de reduccionista, si es que se emplea especialmente en la primera fase de los estudios de que se trate. Los estudios hemerográficos corresponderían a esta primera fase. Una catalogación de fondos, que nos permita estudiar un universo determinado de estudio, es fundamental, como lo es que esté actualizado. Sobre todo, en un mundo como el digital, donde existe el efecto palimpsesto (unas informaciones sustituyen a otras, superponiéndose a ellas), la acumulación, o donde se abandonan muchas páginas Web sin renovar. De todos esos fenómenos, así como de las características morfológicas de esos medios, deben dar cuenta – como lo hacen para el medio impreso– los registros hemerográficos, en forma de bases de datos que luego nos permitan obtener conclusiones cuantitativas o cualitativas.

Disponemos, sin embargo, de algunos materiales que ya ahora nos permiten hacer algunos estudios comparativos. A pesar de la muy diversa metodología empleada por esos estudios, algo se ha hecho. Por ser realistas, no pretendemos que nuestro modesto librito de 1997, *Medios de comunicación en Internet*, sea un catálogo con características científicas, aunque, a pesar de todo, intentaba ser un repertorio lo más completo posible de los cibermedios españoles de la primera hornada. Le han sucedido, para el País Vasco, otros trabajos del profesor Koldo Meso, publicados en diferentes números de la revista *Mediatika*, de la Sección de Medios de Comunicación de la Sociedad de Estudios Vascos.

## 2.2. Definición de cibermedio

Para catalogar los medios de comunicación en Internet, o cibermedios, primero tenemos que definir qué entendemos por cibermedio. No hay una respuesta definitiva, aunque

sí algunas características en las que basarnos. Establecer un catálogo con unos criterios establecidos y marcar una estrategia que permita ir alimentando esa base de datos sería una buena idea. Naturalmente, eso exige un trabajo continuo en el tiempo y que involucre a varios equipos de investigadores, de manera que el surgimiento, desarrollo y extinción de cada cibermedio pueda ser constatado en una base de datos con un cierto detalle. Las tecnologías informáticas, que permiten volcar los datos desde diversos puntos geográficos, hacen desde luego más realizable un empeño tan titánico. Es la metodología que seguimos en el proyecto que, en estos momentos, se halla en su última fase y que, a partir de la coordinación de cuatro equipos de otras tantas universidades españolas (Málaga, Navarra, País Vasco y Santiago de Compostela) ha pretendido hacer un catálogo lo más exhaustivo posible de los cibermedios españoles entre 2002 y 2005. Dicha base de datos se ha ido completando con aportaciones de investigadores de todos los grupos, y ha sido diseñada por investigadores de la Universidad de Santiago (Xosé López, Xosé Pereira y Manuel Gago).

El empleo de esa base de datos ha dado ya algunos resultados conjuntos. Por ejemplo, el censo de cibermedios gallegos, el denominado Arquivo da Comunicación del Consello da Cultura Galega, dirigido por Xosé López, que ya ha dado como resultado un completo catálogo de medios impresos y digitales, clasificados por comarcas y puestos a disposición del público en el sitio Web del Consello, de toda Galicia. En el apartado metodológico se da la siguiente definición de cibermedio: “Un cibermedio é un emisor de contidos que ten vontade de mediación entre feitos e público, utiliza fundamentalmente criterios e técnicas xornalísticas, usa a linguaxe multimedia, actualízase e publícase na rede Internet”.

Se dejan fuera, a la vez, “todos aqueles sitios que, malia reuniren as condicións para seren considerados cibermedios, non están a renovar os seus contidos de maneira periódica, non foron incluídos nesta base de datos. Asemade, tampouco foron considerados cibermedios aqueloutros servicios de noticias elaborados por gabinetes de comunicación ou, doutro xeito, aqueles que unicamente dan conta da información relacionada coa súa propia entidade”.

Con los datos contenidos en esa base de datos, los diferentes investigadores confeccionamos una monografía que, coordinada por el profesor Ramón Salaverría, de la Universidad de Navarra, recoge las aportaciones de los diversos grupos y subgrupos encargados de analizar temas tan diversos como los géneros periodísticos, los derechos de autor, el asociacionismo y los rasgos de la profesión, y donde se explica con detalle el criterio seguido para confeccionar dicha base de datos en incluir en ella determinadas Web que se han considerado periodísticas, mientras que se han dejado de lado otras. Al final de este capítulo

se reproduce una breve comunicación en forma de póster presentada por Xosé Pereira y Manuel Gago en el 2º Congreso *New research for new media*, celebrado en la Universitat Rovira I Virgili de Tarragona en 2004, en el que se muestran los pasos seguidos en la definición de la metodología de la base de datos, así como en la propia arquitectura de la base de datos.

En definitiva, los campos que se han tenido en cuenta para confeccionar esa base de datos morfológica de cibermedios españoles –la clásica ficha de Jacques Kayser, ideada en un momento histórico y un lugar geográfico concretos, y para cumplir un determinado objetivo, censar los diarios franceses de la década de los años 60, no sirve ya del todo satisfactoriamente para los nuevos medios– son los siguientes:

Nombre.  
Razón social.  
Comunidad autónoma.  
Ciudad.  
Provincia.  
URL.  
Correo(s) electrónico(s).  
Contacto.  
Tipo de medio (generalista/especializado).  
Actualización.  
Otros soportes (impreso, audiovisual).  
Contenido.  
Idioma(s)  
Descripción.

A partir de esa misma base de datos, y con el respaldo de un nuevo proyecto de investigación, en este caso ceñido al impacto de Internet en los medios de comunicación vascos<sup>1</sup>, bajo la supervisión del profesor Koldo Meso se esta llevando a cabo un censo de los cibermedios vascos de acuerdo a los siguientes campos:

---

<sup>1</sup> 1/UPV 00017.323-H-16005/2004.

Nombre del medio  
URL  
Razón social.  
Dirección postal.  
Municipio.  
Territorio histórico.  
Correo(s) electrónico(s).  
Fecha de inicio (y de finalización, si es el caso).  
Director del medio.  
Webmaster.  
Persona de contacto.  
Tipo de medio (generalista/especializado).  
Correspondencia con otros soportes.  
Medio matriz (si lo hay)  
Régimen de propiedad del medio.  
Actualización.  
Contenidos.  
Elementos multimedia.  
Interactividad.  
Publicidad.  
Idioma(s)  
Observaciones.

Pretendemos con esta ficha constatar los elementos mensurables que nos permitan luego investigar diversos aspectos, desde los más puramente morfológicos (nombre, URL...) a los cronológicos (fecha de inicio y finalización, como en la ficha de Jacques Kayser), pero también sobre la titularidad de los medios y su estructura empresarial, utilización de variedades lingüísticas, y también que nos permita interpretar hasta qué punto se utilizan las características intrínsecas del nuevo medio (interactividad, multimedialidad, etc.).

### **3. Clasificación de cibermedios**

En Cataluña también se han hecho algunos intentos de catalogar los cibermedios de aquella comunidad autónoma española. Hay un catálogo breve, referido sólo a los medios en

catalán, coordinado por Jordi Fortuny, de 2003, aunque el intento más sistemático y completo de catalogación y análisis es el que David Domingo, profesor de la Universitat Rovira i Virgili de Tarragona y secretario del Grup de Periodistes Digitals de Catalunya llevó a cabo por encargo, precisamente, de esta asociación. En su libro, David Domingo recoge 445 medios de comunicación catalanes.

El libro de David Domingo comienza por hacer un breve recorrido teórico por las características del ciberperiodismo y una pequeña historia de la presencia de los medios catalanes en Internet. Sin embargo, lo más interesante para las cuestiones hemerográficas que nos interesan aquí se recogen en el capítulo 3 de su libro, referido a los criterios y variables empleadas para confeccionar el censo. Éste es amplio, pues recoge toda Web que contenga información de actualidad general. David Domingo busca el mínimo común denominador. A los cibermedios los denomina “publicaciones digitales”. Para localizar esos medios o publicaciones, David Domingo siguió una estrategia triple:

1. Recurrir a los buscadores más habituales.
2. Directorios de medios de las asociaciones profesionales.
3. Seguimiento de enlaces en cada Web hallada, que a su vez podía remitir a otras similares o relacionadas.

La hemerografía, por lo tanto, debe partir de los repertorios ya establecidos, pero también buscar sus propios recursos con el objetivo de hacer catálogos lo más completos, ordenados y útiles posibles. También, y dada la volatilidad de los datos y de los propios contenidos en Internet, debe fijarse unos límites temporales, a ser posible acordes con la periodización escogida en función de variables históricas explícitas.

En cuanto a las variables analizadas por Domingo, son las siguientes, divididas en variables independientes, variables dependientes, contenidos informativos y otros contenidos:

1) Variables independientes:

\* Promotor: Persona o entidad que tiene la iniciativa de crear y mantener la página Web informativa en Internet. A su vez, puede ser:

Una empresa privada.

Una empresa pública.

Un proveedor de servicios.

Una entidad cívica.

Un colectivo virtual.

Una persona individual.

\* Tipo de Web. David Domingo distingue categorías según los medios (radio, televisión, prensa) de que son versiones dichas Web, o, si son iniciativas que no parten de otros medios, entonces las divide en:

Portales, que deben incluir al menos tres de los siguientes conjuntos de elementos: información de actualidad, información de servicios, buscador o directorio de páginas Web, foros y otros servicios en línea (por ejemplo, correo electrónico).

Publicación digital autónoma, no vinculada a una empresa previa dedicada a la información periodística.

\* Ámbito geográfico. Naturalmente, esta clasificación debe atender a la división administrativa de cada lugar, David Domingo sigue la estructura catalana:

Barrio.

Municipal.

Supramunicipal.

Comarcal.

Supracomarcal.

Provincial.

Autonómico.

Estatal.

Global.

\* Localización: Lugar concreto donde se publica la Web catalogada.

## 2) Variables dependientes:

Idiomas utilizados (idioma principal / otros idiomas).

### 3) Contenidos informativos:

\* Grado de desarrollo de las noticias. Domingo lo mide en función de que se cumpla o no:

- Texto completo (sobre todo si reproduce el de un medio impreso).

- Piezas seleccionadas (también en el caso de medios derivados).

- Titulares.

- Texto y audiovisual, sobre todo en el caso de páginas Web derivadas de radios y televisiones.

- Audiovisual, sin acompañamiento de texto.

- Sin noticias de actualidad.

\* Producción de las noticias: Determina su origen:

- Propia *online*, redactada por una redacción específica para Internet.

- Propia *offline*, producida por el medio matriz en el caso de las páginas Web derivadas.

- Colaboración con un medio local, del que se reproducen las noticias.

- Informaciones suministradas por los usuarios; el modelo de *OhMyNews*.

- Servicio de noticias en línea: las noticias las suministra una agencia externa al medio.

- Selección de noticias tomadas de otros medios de comunicación.

\* Última hora.

\* Proveedores de noticias: determina quiénes son los que suministran la información que publica el medio.

\* Actualización: ritmo con que se renuevan las informaciones.

\* Enlaces a noticias relacionadas (del propio medio).

\* Enlaces a otras páginas Web externas.

\* Comentarios de los lectores a las noticias.

\* Personalización.

#### 4) Otros contenidos

- \* Directorio de páginas Web, si el medio ofrece una lista de enlaces relacionados.
- \* Tipología de las páginas Web, si las que figuran en ese directorio son:
  - \* Generales
  - \* Especializadas, categoría a su vez dividida en:
    - \* Comerciales
    - \* Medios
    - \* Personales
    - \* Entidades
  - \* Información de servicio: agenda, cartelera de cines, transportes, etc.
  - \* Foros, *chats*.
- \* Activos del último mes: indicación acerca de si ha habido intervención de los usuarios en esos foros o *chats*.

La propuesta de David Domingo, aunque muy completa, deja de lado otras cuestiones que sí se reflejan en otras bases de datos, como las que hemos visto. Eso obedece al interés diferente de cada estudio y a las propias características de la comunicación en cada lugar. Sin embargo, sería deseable un cierto acuerdo en la comunidad académica para lograr una ficha más o menos estándar que, como la de Jacques Kayser para la prensa –francesa y de los años 60, todo hay que decirlo– sea aceptada por la mayoría de la comunidad académica. En ese sentido, la adopción de la norma ISO 690-2 <<http://www.collectionscanada.ca/iso/tc46sc9/standard/690-2e.htm>> puede ayudar bastante, así como la estandarización de los metadatos, tanto si se siguen los estándares del *Dublin Core* <<http://dublincore.org>>, que permiten describir la estructura lógica de los documentos digitales, especialmente si se emplea el lenguaje XML.

#### **4. Hemerotecas digitales**

La Historia –ese breve periodo de la Humanidad que conocemos y estudiamos precisamente porque disponemos de documentos para estudiarlo– se ha caracterizado precisamente por una lucha contra el olvido. La invención de la escritura y su plasmación en

diversos soportes más o menos duraderos –más o menos perecederos– es el resultado de ese empeño que, a la manera de Sísifo, busca una y otra vez subir esa masa de recuerdos que queremos que se conserven y se transmitan más allá de nuestro efímero paso físico y temporal por este mundo, hasta un lugar seguro, protegidos del olvido. Y aún buscamos algo más: relacionar todos esos conocimientos, formar todo un mundo de relaciones complejas en que cada cosa tenga que ver con otras muchas.

Internet ha supuesto la materialización, o el comienzo de la misma al menos, de ese viejo sueño de la Humanidad de ampliar el conocimiento humano a través de herramientas que permiten la recuperación del mismo a través de una inmensa red distribuida. El viejo sueño de la biblioteca global se antoja ahora posible, y de hecho en la Red pueden encontrarse catálogos enteros de bibliotecas, colecciones de periódicos, páginas personales, catálogos de empresas y otras muchas *sedes* que han hecho de la edición digital un sustituto de la analógica, y que además permite trascender las barreras del espacio y del tiempo, al hacer inmediatamente accesible a, potencialmente, todo el mundo (todo el mundo que esté conectado a Internet, al menos) la información que en un momento y lugar determinados colocamos en ese inmenso zoco.

Eso ha hecho que sea posible plantearse la formación de enormes bibliotecas digitalizadas, y también de hemerotecas digitales. En España, ha tratado sobre esta cuestión el profesor Lluís Codina (Codina 2000, 2001), y, a la hora de enfrentarse a la ingente tarea de hacer digitales los viejos papeles, son muy recomendables los libros de Herrera Morillas, así como el de Townsend *et al.*

A la hora de acometer la digitalización de hemerotecas, hay que tomar algunas decisiones previas:

- \* El formato a emplear para reproducir los documentos originales (¿PDF? ¿Formato de imágenes?), tamaño y resolución de esos facsímiles digitales.

- \* Si se han de reproducir los objetos como tales.

- \* Si han de reproducirse en el estado en que se encuentran o bien merece la pena intentar restituirlos a su aspecto original;

- \* Si es preferible reproducir su contenido textual (de tal manera que tengamos un archivo susceptible de buscarse), o ambos, el facsímil y el texto.

Entre estas hemerotecas digitales, merece la pena mencionar el pionero proyecto Internet Library of Early Journals (ILEJ), llevado a cabo por las universidades británicas de Oxford, Manchester, Leeds y Birmingham (<http://www.bodley.ox.ac.uk/ilej>) bajo la dirección de Thaddeus Lippinski. Le han seguido otros, hasta el punto de que se ha generado toda una metodología para la digitalización de prensa histórica (Mühlberger, 1999; Deegan *et al.*, 2002). La necesidad de digitalizar y poner a disposición de los estudiosos los productos informativos audiovisuales ha dado lugar a proyectos como los descritos en Lange, 2001.

En España, *El País* y *La Vanguardia* incluyen entre sus servicios toda su hemeroteca, digitalizada en formato PDF; esta última, probablemente la más extensa hemeroteca digital de Europa, ya que el diario data de 1883, más de 120 años de periódicos. Ya en 1995, *Abc* se había planteado un proyecto similar (Armañanzas *et al.*, 1996: 147), que diez años después continuaba sin realizarse. La envergadura de estos proyectos –sobre todo en el caso de diarios con muchos años de existencia cuya colección más antigua, en algunos casos pliegos de gran tamaño y deficiente estado de conservación, ha de fotografiarse o escanearse digitalmente– hace que muchas empresas se lo piensen dos veces antes de realizar la inversión. El formato elegido, el Portable Document Format de Adobe, está resultando ser prácticamente un estándar, pero en el mundo digital todos los formatos y tipos de archivos acaban por volverse obsoletos y obligan a una nueva inversión si se quiere convertirlos en nuevos archivos legibles por las nuevas máquinas y programas.

En todo caso, no son pocos los diarios digitales que ofrecen, la mayoría de las veces mediante pago, una versión en PDF del diario impreso de esa jornada. En su día, fue pionero *El Periódico de Catalunya* (de hecho, parece ser que fue el primer diario en utilizar este formato en Europa), le siguió *El País*, y sobre todo a partir de 2004, se les van uniendo otros diarios, como *El Correo* de Bilbao, que de momento sólo cobra por ese servicio, y que va confeccionando así un archivo digital de sus ediciones impresas.

Las hemerotecas digitales de medios impresos o audiovisuales son una cuestión que aquí sólo dejamos planteada y que requiere de inversiones cuantiosas y de estrategias claras, a medio y largo plazo, por parte de grupos de investigación, universidades, instituciones públicas, bibliotecas, medios de comunicación e incluso particulares <<http://www.ehu.es/diaz-noci/hemeroteca>>. Lo que nos interesa más ahora (y sobre eso hay, sorprendente y lamentablemente, poco camino hecho) es el archivo de los propios cibermedios, que plantean problemas propios. El primero de ellos, y quizá el más importante: a falta de una política por parte de las empresas, instituciones, bibliotecas, ya no son posibles, salvo que se haga una

labor de arqueología digital (y no sería nada descabellado), las ediciones digitales de buena parte de la primera década de vida de los medios de comunicación en Internet. Se argumentará que los propios medios ofrecen esta posibilidad; no siempre. Hay medios que sí disponen de un servicio de hemeroteca digital, pero no hay más que solicitar, a través del motor de búsqueda, cualquiera de los primeros ejemplares de la edición digital para comprobar que el sistema no nos va a devolver aquellas ediciones con aquel aspecto primitivo, sino una versión digital, en PDF o en HTML, del periódico impreso, siempre con el aspecto que tiene la edición actual. Otros medios, ni siquiera eso.

Otro de los problemas fundamentales es: ¿Qué versión, de las que a lo largo del día ofrece un medio que se actualiza con más o menos frecuencia, ha de tomarse como objeto de investigación? ¿Hay alguna versión que un ciberdiario considere como definitiva? ¿La última del día? ¿Guardan los propios medios una versión de los diferentes archivos? Si no es así, el historiador que fije sus ojos en las rutinas profesionales y el cambio que en esta década ha comenzado a producirse al respecto –del cual la actualización más o menos constante es un ejemplo– no tendrá materiales sobre los que comprobar sus hipótesis.