

## **Análisis del Consumo de Información de una Revista Electrónica:**

### **Análisis de Ficheros Log de *Cybermetrics*.**

José Luis Ortega

Laboratorio de Internet

Centro de Información y Documentación Científica-CSIC

Joaquín Costa, 22 28002 Madrid

[jortega@cindoc.csic.es](mailto:jortega@cindoc.csic.es)

#### **Resumen**

El propósito de este trabajo es estudiar el consumo de información que durante el periodo 2001-2003 se realiza de la revista electrónica *Cybermetrics*. Como metodología se ha analizado las transacciones realizadas a dicha revista recogidas en los ficheros web log. Los resultados nos proporcionan información sobre el origen de los accesos, secciones más visitadas y otros aspectos de su usabilidad. Seguidamente, se discute la problemática existente en la utilidad y uso de esta fuente y particularmente los problemas derivados de esta metodología en este trabajo. Por último, se concluye con la internacionalización de la revista y su posición en el mundo de la Bibliometría y Cienciometría.

**Palabras Clave:** Revistas electrónicas, Consumo de información, Análisis de ficheros log, *Cybermetrics*.

### **Abstract**

The aim of this paper is to study the information consumption made in the electronic journal *Cybermetrics* for 2001-2003. As methodology, the web log files transactions made to this journal were analysed. The results provide information about access source, sections more visited and other usability aspects. Next, the paper discuss the utility and use of this information source and concretely the methodological problems raised in this study. Finally, it conclude with the international position of the journal and its visibility in the Bibliometrics and Scientometrics field.

**Keywords:** Electronic journals, Information consumption, Web log analysis, Cybermetrics.

### **Introducción**

Björneborn e Ingwersen (1), en un trabajo sobre el estado de la cuestión de la Cibermetría, apuntaron que uno de los hitos fundamentales en el nacimiento de la disciplina, fue la aparición de la revista *Cybermetrics*. La revista nació en 1997 de la mano de Isidro Aguillo en el Laboratorio de Internet del Centro de Información y Documentación Científica del CSIC, convirtiéndose en la primera revista científica versada exclusivamente en estudios cuantitativos de la producción científica en Internet. La revista adoptó el formato electrónico con la pretensión de valerse de Internet como el mejor medio de difusión de la producción científica, además de ser la propia Internet el objeto de estudio de la revista. Por otro lado, al ser pionera en su campo adoptó el inglés como única lengua de comunicación, lo que le reportaría una mayor difusión. La revista

no sólo pretendía ser una revista científica sino que pretendía también ser un portal de acceso al mundo de la Cienciometría, Bibliometría y Cibermetría en Internet. En ella encontramos una sección dedicada a la bibliografía existente en Internet sobre estos campos de investigación, accesos a sedes web de distintas entidades significativas de Internet o en el ámbito de la investigación cuantitativa (instituciones, grupos de investigación, buscadores, estadísticas, etc.). Se encuentran agrupadas en distintas secciones como *Scientometrics* (literatura sobre la Cienciometría), *Tools* (distintas herramientas como software, buscadores, etc., para la actividad cibermetra) o *R&D Policy and Resources* (enlaces sobre I+D). Aún más, junto a la revista, se desarrollan seminarios específicos sobre Cibermetría paralelos a las Conferencias Internacionales en Cienciometría e Informetría (ISSI) donde se discute el desarrollo y evolución de esta reciente disciplina. Estas reuniones también están reflejadas en la revista en la sección *The Seminars*. Distintas investigaciones sobre el fenómeno de las revistas electrónicas (e-journals) han tomado a esta publicación en cuenta. Koehler et al. (2) concluyen que *Cybermetrics* está en fase de maduración. Hawkins (3), sin embargo, apunta a esta misma como la revista electrónica en Documentación con más citas del SCI por artículos publicados.

Desde la revisión de Peters (4) sobre la literatura en análisis de ficheros log o archivos de acceso (5), estos se han convertido en una fuente de información muy valiosa para conocer distintos aspectos del comportamiento de los usuarios en Internet. El tipo de consultas, el uso de distintos operadores o la navegación entre los resultados de un motor de búsqueda son aspectos cuantificables a través de estos ficheros y claves para el diseño y evaluación de estos buscadores (6,7,8) y en concreto, de servicios bibliotecarios (9,10). Sin embargo estos ficheros también proporcionan información sobre conductas y motivaciones a la hora de usar un sitio web lo que ayuda a su diseño

y usabilidad (11,12). Este hecho también está fomentando los estudios de mercado (E-metrics) investigando el comportamiento de los clientes a la hora de realizar compras on-line (13).

Por otro lado, el fenómeno de las revistas electrónicas han generado diversos estudios que se han centrado en la problemática del formato electrónico frente al formato impreso (14,15,16). La normalización y el uso y consumo de revistas electrónicas también han centrado diferentes trabajos (17).

## **Objetivos**

En este trabajo se pretende estudiar el consumo de información que durante el periodo 2001-2003 se realiza de una revista electrónica. Como medio representativo de las investigaciones cibernéticas actuales, la revista *Cybermetrics* (<http://www.cindoc.csic.es/cybermetrics>) ha sido considerada para este estudio. A través de la información que los ficheros log de dicha revista nos proporciona se estudiará los accesos a qué elementos de la revista, perfil de las visitas y origen de éstas. Por último, se discute la problemática existente en la utilidad y uso de esta fuente y particularmente los problemas derivados de esta metodología.

## **Metodología**

Los ficheros log son archivos generados por un sistema operativo o por un determinado programa donde se reflejan sucesos ocurridos dentro de una aplicación. Existen diferentes tipos de ficheros log dependiendo de quién los genera y qué tipo de acciones

recoge. En este estudio nos centraremos en los web log, o ficheros log generados por un servidor web con el formato *Common Log Format* el cual genera los siguientes datos:

- IP: es el número que identifica el ordenador, o servidor desde donde se accede a un recurso web.
- RFC nombre: permite identificar al ordenador remoto que está haciendo la petición, pero en este caso el servidor no ha recogido este tipo de información
- Autenticación: recoge los datos de autenticación en servidores con permisos de acceso. Al ser un servidor de libre acceso no ha recogido esta información.
- Fecha: momento en que se realiza el acceso, recoge tanto la fecha como la hora
- Acción: es el tipo de acción que se realiza en un servidor. La más usual es GET, la acción de obtener un tipo de fichero. Existen otros tipos de acciones como PUT, POST, HEAD, etc. dependiendo de la acción que se realice.
- Ruta: recoge el archivo que ha sido solicitado y su ruta dentro del servidor.
- Protocolo: es el protocolo de acceso al servidor, en este caso HTTP
- Código: se trata de una convención de W3.org donde se codifica el estado de la acción, así un 200 es un acceso satisfactorio, un 404 es todo lo contrario.
- Bytes: cantidad de bytes que solicita el ordenador cliente.
- Referencia: recoge la url de la página web que ha enlazado con el recurso solicitado.
- Navegador: tipo de navegador usado para hacer el acceso.

Para este trabajo se ha tenido solo en cuenta los datos: IP, Fecha, Ruta, Código y Referencia.

### ***Obtención de datos***

En un primer momento se realizó un proceso de depuración de datos, ya que al estar la revista incluida en un directorio de [www.cindoc.csic.es](http://www.cindoc.csic.es) los ficheros log recogían a la url en su totalidad. Una vez filtrados los accesos a [www.cindoc.csic.es/cybermetrics/](http://www.cindoc.csic.es/cybermetrics/) se exportaron a una base de datos Access 2000, donde se contabilizaron un total de 615.194 accesos durante el periodo mayo de 2001 hasta octubre de 2003. A continuación, una serie de consultas se fueron realizando y posteriormente volcando en Excel 2000 para su posterior tratamiento estadístico.

## **Resultados**

### ***Accesos en el tiempo***

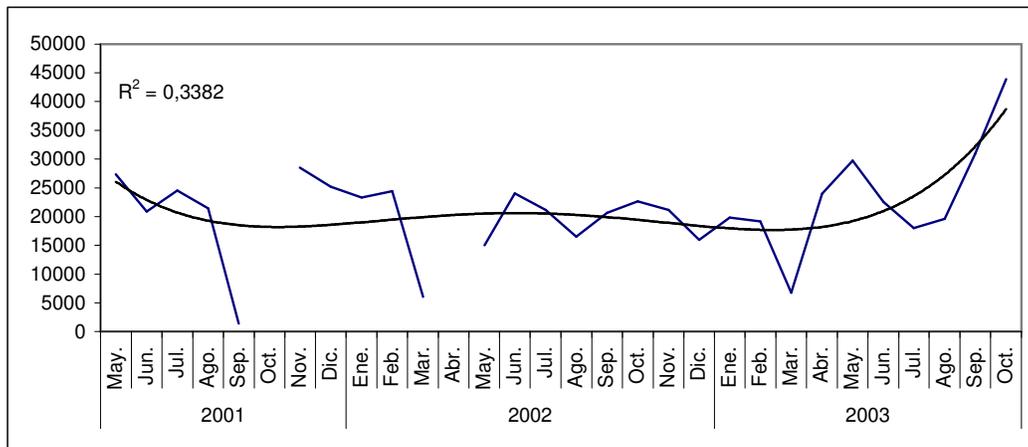


Figura 1. Accesos mensuales en el tiempo

<i>Año</i>	<i>Meses</i>	<i>Accesos</i>	<i>Periodo</i>
2001	Abr.	1957	27-30
	May.	27320	01-31
	Jun.	20909	01-30
	Jul.	24548	01-31
	Ago.	21418	01-31
	Sep.	1385	01-03
	Oct.	0	
	Nov.	28534	01-30
	Dic.	25199	01-31

2002	Ene.	23330	01-31
	Feb.	24408	01-29
	Mar.	6038	01-07
	Abr.		0
	May.	15040	01-31
	Jun.	24037	01-30
	Jul.	21134	01-31
	Ago.	16539	01-31
	Sep.	20686	01-30
	Oct.	22680	01-31
	Nov.	21178	01-30
	Dic.	15942	01-31
2003	Ene.	19873	01-31
	Feb.	19176	01-28
	Mar.	6753	01-10
	Abr.	23938	01-30
	May.	29735	01-31
	Jun.	22574	01-30
	Jul.	17991	01-31
	Ago.	19575	01-31
	Sep.	30942	01-30
	Oct.	43882	01-31

Tabla I. Número de registros por meses

Antes de todo, se ha detectado que los datos de los ficheros log poseen ciertas lagunas temporales. Por ejemplo, en la figura 1, las líneas discontinuas en los meses de octubre y abril expresan ausencia de datos y los bruscos descensos en los meses de septiembre y marzo representan meses incompletos. Los datos por cada mes pueden verse mejor representados en la tabla I. Aún así, los resultados ofrecen una cierta tendencia a la baja en los años 2001 y 2002, pero a lo largo del 2003 encontramos un repunte en los accesos, llegando a duplicarse con respecto a los años anteriores. Esto es debido a la inactividad de la revista durante estos dos primeros años donde no se publicaron números nuevos. Sin embargo, el repunte de 2003 es debido a la aparición del último número y el cambio de diseño de la revista. El estancamiento de los dos primeros años no se debe entender como un signo de estabilidad, todo lo contrario, si tenemos en cuenta el crecimiento exponencial de Internet se debe interpretar como un aspecto negativo, ya que la línea de tendencia tendría que adoptar un valor más elevado

( $R^2=0,338$ ). Por otro lado, el hecho de no contar con datos de algunos meses y algunos de ellos no completos, provoca estos descensos bruscos del gráfico.

### ***Accesos por Países***

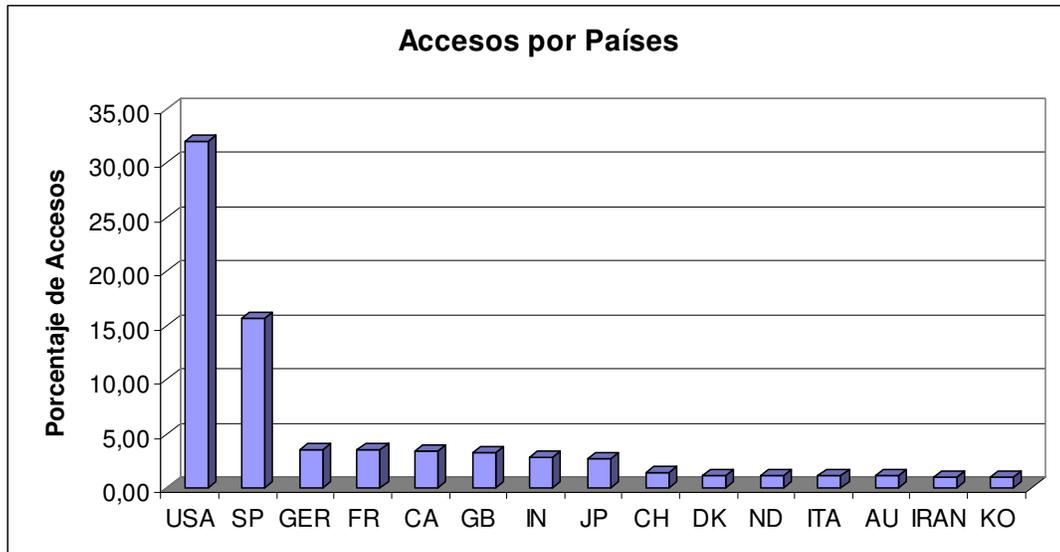


Figura 2. Porcentaje de accesos por países

En esta fase se ha utilizado el software IPLocator 1.21 para conocer el origen geográfico y titularidad de las IPs registradas. Este programa realiza de forma automática consultas a distintas bases de datos WHOIS, generando como resultado una hoja de Excel con el origen de las IPs.

Los accesos realizados a la revista ascienden a un total de 151 países. Debido a la gran dispersión, sólo hemos considerados los 15 primeros (ver Figura 2) con un acceso mayor del 0,9%, siendo estos 15 el 74,85% del total de los accesos. De los 15 cabe destacar en un principio el elevado peso de Estados Unidos (32%) y España (15,65%) muy alejados de los siguientes países como Alemania (3,58%), Francia (3,49%) y Canadá (3,37%). La desproporcionada presencia de Estados Unidos sólo es explicable por el desarrollo e implantación que existe de la red Internet en este país, estimándose que el 29% de los internautas son de este país (18). Por otro lado, el porcentaje de

España (excluyéndose los accesos internos desde el Cindoc y el CSIC) demuestra el grado de implantación local de la revista.

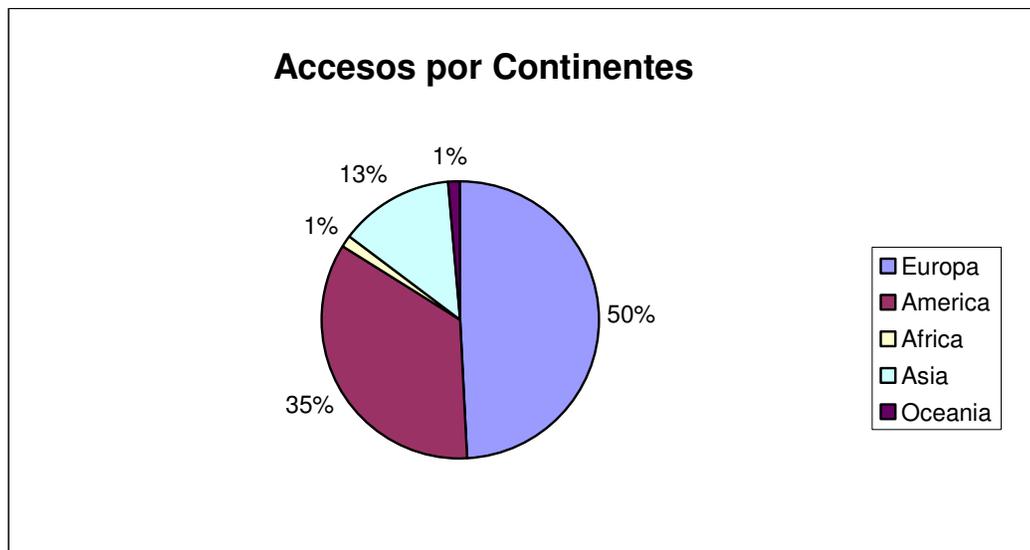


Figura 3. Accesos realizados por Continentes

Como muestra la figura 3, en el acceso por continentes destaca una mayoritaria presencia europea (50%), zona donde se concentran los principales grupos de investigación y donde la Cibermetría están tomando más arraigo. Le sigue América con un 35%, justificado escasamente por el alto porcentaje de acceso de Estados Unidos y Canadá. En tercer lugar destaca la presencia de Asia, y más concretamente el Sudeste Asiático con un 13%, lo que demuestra la fuerte incorporación que estos países están realizando al mundo de Internet y concretamente al mundo de la Cibermetría, como ejemplo, la celebración en Pekín en 2003 de la Conferencia Internacional de Cienciometría e Informetría.

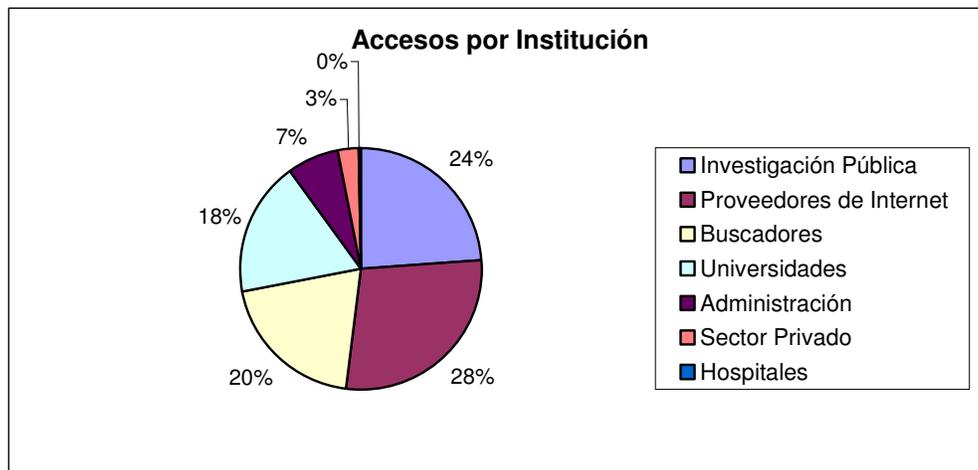


Figura 4. Accesos realizados por Institución

La distribución de las visitas es tan dispersa que el 90% de los accesos tiene una frecuencia inferior a 25. Estas escasas frecuencias, a nuestro modo de ver, son poco significativas ya que muchas de ellas pueden deberse a un acceso casual dirigido por buscadores o por la simple navegación. Aún así, sólo han sido considerados las entidades con más de 50 accesos (46,5%).

Como se puede apreciar en la figura 4, el 23,94% de los accesos ha sido realizado por organismos de investigación pública (descontando a los accesos realizados desde el CINDOC que por sí solos constituyen el 35% de la investigación pública). El porcentaje en Proveedores de Servicios Web como Telefónica o America Online es bastante elevado (28,12%), estos proveedores suelen recoger a la mayoría de los usuarios particulares pero el uso de IPs dinámicas y los proxis dificultan conocer el usuario final. También es notable la presencia de *robots* o *spiders* de buscadores con un 19,77%; su abundancia y la creciente frecuencia de actualización de estos servicios está incrementando estos accesos. A continuación siguen las Universidades (17,83%), la Administración Pública (7,08%) y el sector privado (2,59%).

De las instituciones más significativas, dejando a un margen proveedores y *robots*, sobresalen entre las universidades la Universidad Carlos III de Madrid y entre la Administración pública la Comunidad de Madrid. Ello destaca la cercanía física de ambas instituciones y las relaciones de colaboración que existe entre estos centros y el CINDOC.

### **Secciones**

Como habíamos comentado antes, *Cybermetrics* no sólo es una revista sino que se presenta como un portal de acceso al mundo de la Bibliometría, Cienciometría y Cibermetría. Por ello, además de la propia revista, posee distintas secciones sobre bibliografías, herramientas, seminarios, etc.

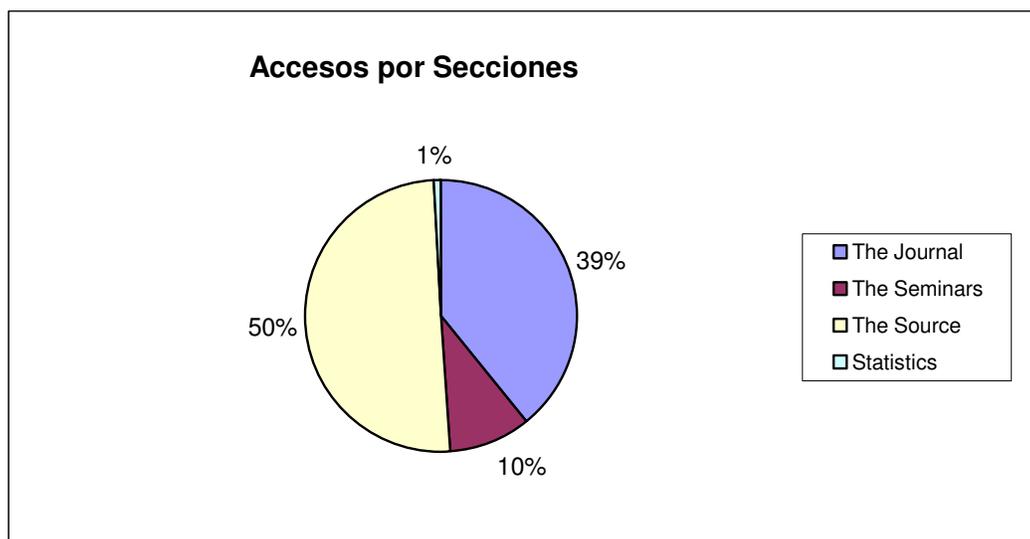


Figura 5. Accesos realizados las distintas secciones de *Cybermetrics*

Como podemos ver en la figura 5. La parte *The Source*, donde se recoge distintos enlaces en el ámbito de la Cibermetría, I+D, Políticas científicas, herramientas, etc., posee la mitad de los accesos (50%). A continuación, *The Journal*, la revista en sí, con un 39%, le sigue *The Seminars* con un 10% y finalmente *Statistics* con apenas un 1%.

Pero estos datos pueden resultar engañosos si tenemos en cuenta el total de páginas que posee cada sección y la estructura web de cada sección, con páginas-pasarela que dirigen de un contenido a otro. Por ello, hemos estudiado los promedios de acceso a distintas secciones. Los resultados difieren considerablemente.

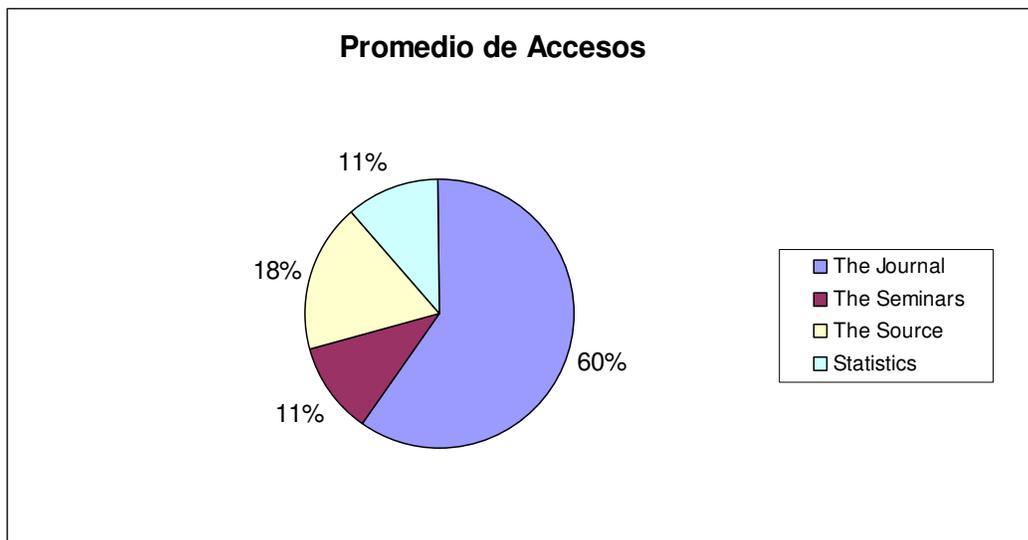


Figura 6. Promedio de accesos a las distintas secciones

Como se puede apreciar ahora en la figura 6 en *The Journal* recae el gran peso de los accesos (60%), siendo esta la principal sección del sitio web. La significativa presencia de *The Source* (18%) nos informa de esa doble función de *Cybermetrics*, revista y fuente, convirtiéndose también en medio de acceso a otras páginas y contenidos. El menor peso de *The Seminars* (11%) muestra poco interés en los accesos ya que no contiene resúmenes ni documentos sobre estos seminarios. Por último, y con un mismo porcentaje, *Statistics* (11%) aporta estadísticas del sitio.

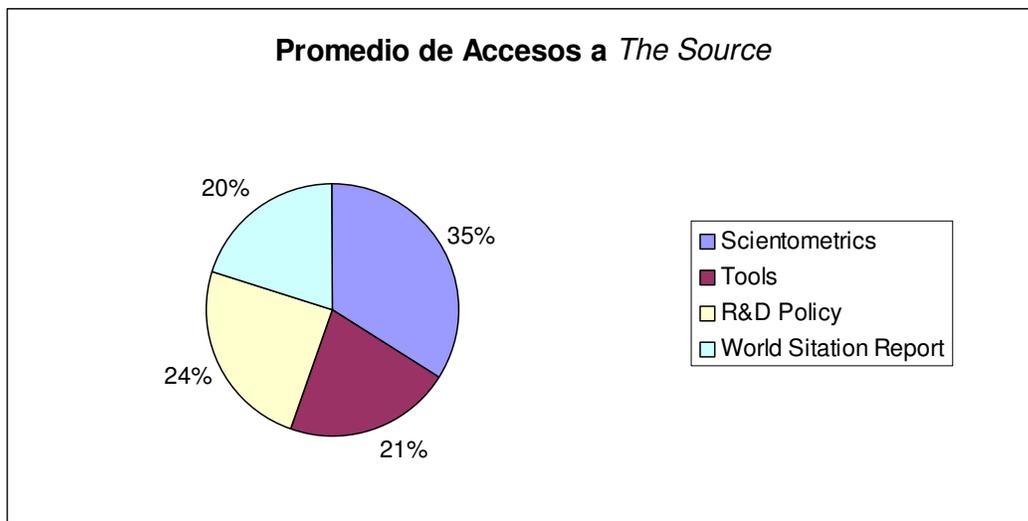


Figura 7. Promedio de accesos a la sección *The Source*

Dentro de la sección *The Source* (figura 7), la cual posee mayor número de páginas, podemos ver que las distintas sub-secciones no difieren mucho entre sí. La más visitada es Scientometrics con un 35% frente a la menos *World Sitation Reports* con un 20%. Esto nos lleva a pensar que todas las sub-secciones poseen el mismo interés para las visitas y se accede a todas por igual.

### ***The Journal***

Debido a su importancia merece una mención aparte los accesos realizados a los artículos de la revista ya que estos son el centro del sitio web y además el centro de nuestro estudio. Como habíamos dicho anteriormente, la revista nació en 1997. Desde ese momento ha sacado de forma intermitente 6 números hasta el momento del estudio (recientemente han aparecido los números 6-7 de 2003 y 8 de 2004). El número de artículos por cada número también ha variado, encontrándonos números con sólo un trabajo y otros con dos o cuatro.

Artículos	Accesos html	Accesos pdf	TOTAL	%	Citas
ROUSSEAU, R. <i>Sitation, an exploratory study</i> (1997)	1270	5416	6686	13,85	45
LEYDESDORFF, L. <i>Indicators of Innovation in a Knowledge-based Economy</i> (2001)	3194	2818	6012	12,46	1
THELWALL, M. <i>The Responsiveness of Search Engine Indexes</i> (2001)	2855	2226	5081	10,53	16
BAR-ILAN, J. <i>Search Engine Results over Time -A Case Study on Search Engine Stability</i> (1999)	2135	1374	3509	7,27	19
LEYDESDORFF, L. And CURRAN, M. <i>Mapping University-Industry-Government Relations on the Internet: The Construction of Indicators for a Knowledge-Based Economy</i> (2000)	1984	1442	3426	7,10	13
KOEHLER, W. <i>A Profile in Statistics of Journal Articles: Fifty Years of American Documentation and the Journal of the American Society for Information Science</i> (2000)	1619	1777	3396	7,04	1
ROUSSEAU, B. and ROUSSEAU, R. <i>LOTKA: A program to fit a power law distribution to observed frequency data</i> (2000)	2028	1232	3260	6,75	4
MANNINA, B. And QUONIAM, L. <i>How to hold a Virtual Library Active?</i> (2000)	1576	1067	2643	5,48	1
ROUSSEAU, R. <i>Daily time series of common single word searches in AltaVista and Northern Light</i> (1999)	1584	1029	2613	5,41	35
ARCHAMBAULT, E. <i>Comments to the article by Rousseau &amp; Rousseau</i>	1091	1053	2144	4,44	0
ROUSSEAU, B. and ROUSSEAU, R. <i>Rejoinder</i>	1030	994	2024	4,19	0
KATZ, S. L. <i>Software and Peer-Review: The Rousseau Case</i>	1113	781	1894	3,92	0
NEWMAN, M. <i>Comments to the article by Rousseau &amp; Rousseau</i>	1016	756	1772	3,67	0
<b>TOTAL</b>	<b>25578</b>	<b>21965</b>	<b>48263</b>	<b>100</b>	<b>135</b>

Tabla II. Acceso a los artículos publicados en *Cybermetrics*

Como se puede apreciar en la tabla II, el artículo más significativo de esta publicación es el trabajo de Rousseau *Sitation: an exploratory study*. Es el artículo más visitado en este periodo con un 13,85% a la vez que el más citado por el SSCI, con un total de 45 citas (19). Su importancia, además de ser pionero en la Cibermetría, estriba en la aportación del concepto *sitation* como término paralelo al *citation* bibliométrico, mostrando la aplicabilidad de las técnicas bibliométricas a la Cibermetría.

Aparte a esta excepción, los artículos más visitados son los de más reciente publicación, como son los artículos del número 5, el trabajo de Leydesdorff (12,46%) y Thelwall (10,53%). Se percibe también que los trabajos más antiguos tienden con el tiempo a incrementar el número de accesos.

El número de accesos en formato pdf puede interpretarse como un grado de consumo mayor que el html ya que usualmente se utiliza este formato para descargas e impresiones. Teniendo en cuenta este hecho podemos destacar la desproporción de descargas en el artículo más visitado de Rousseau, con una relación de 1 a 4,

demostrando el gran interés que este trabajo muestra. Con respecto a los demás la relación de accesos html es superior a pdf aunque poseen un alto grado de explicación ( $R^2=0,82$ ).

Con respecto a la relación citas y accesos, se detecta que, a excepción del trabajo de Leydesdorff del 2001 y el de Rousseau de 1999 existe una alta correlación ( $R^2=0,81$ ). Aún así, el pequeño número de trabajos con el que contamos nos hace ser cautos con este resultado.

### ***Estado del acceso***

Siempre que se produce una conexión el protocolo HTTP nos informa del estado de la acción, a través unos códigos convención de w3.org, con los que podemos obtener información relevante sobre navegabilidad y usabilidad del sitio.

<i>Cod</i>	<i>Significado</i>	<i>Frecuencia</i>	<i>Porcentaje</i>
200	OK	467436	79,93
304	No modificado	98287	16,81
206	Contenido parcial	10638	1,82
301	Eliminado	4683	0,80
404	No se encuentra	3737	0,64
400	Respuesta errónea	21	0,00
403	Prohibido	4	0,00
405	Acción no permitida	3	0,00

Tabla III. Estado de los accesos a la revista

Como podemos ver en la tabla III, cerca del 80% de los accesos han sido satisfactorios (código 200), lo cual nos aporta un porcentaje muy bueno sobre la conexión. Esto se confirma al contar tan sólo un 0,6% de códigos 404 (No se encuentra). Pero quizás es más interesante el porcentaje significativo de código 304 (No modificado) con un 16,8% y el 206 (Contenido Parcial) con un 1,81%. El código 304 nos advierte de un redireccionamiento, el cual se produce cuando entramos en el directorio

[www.cindoc.csic.es/cybemetrics](http://www.cindoc.csic.es/cybemetrics) que automáticamente carga la página principal [www.cindoc.csic.es/cybermetrics/cybermetrics.html](http://www.cindoc.csic.es/cybermetrics/cybermetrics.html). El código 206 nos informa que el documento no ha sido cargado en su totalidad. Este hecho se ha comprobado que usualmente sucede con la transacción de ficheros .pdf y .xls ya que poseen un mayor tamaño y su carga es más dificultosa.

### ***Páginas de referencia***

En este apartado veremos cuáles han sido las páginas de referencia u origen desde donde se ha accedido a la revista *Cybermetrics*.

La mayoría de los accesos se ha producido como efecto de la navegación, por lo que es lógico pensar que el 52,6% se ha realizado desde el Cindoc, de los cuales el 80,4% son de la propia revista *Cybermetrics*. También localizamos páginas de origen significativas con más de 1000 accesos origen como la International Atomic Energy Agency o la Electronic Journal of Sociology. La otra gran parte lo representa accesos de *robots* y accesos directos sin paso previo por enlaces o buscadores.

Desde el punto de vista de los buscadores, sólo representan el 3,8% de los accesos. Los más importantes son Google con el 66,9% de las consultas, Yahoo con el 16% y AltaVista con el 9%. Con respecto a los términos de búsqueda hemos recuperado los términos de búsqueda lanzados a Google que se han utilizado para acceder a la revista. Se han recuperado 4819 consultas y los términos más frecuentes de consulta son mostrados en la tabla IV.

<i>Término</i>	<i>Frecuencia</i>	<i>Posicionamiento</i>
Cybermetrics	651	1
Scientometrics	616	3
scientific definitions	373	0
siting websites	258	0
Bibliometrics	224	1
Sitations	221	1
Webometrics	96	25
Informetrics	88	1
invisible internet	52	20
Lotka	48	14
Scientometrics journal	45	1
Colima México	37	297
Bibexcel	34	13
bibliometric software	34	5
Isidro Aguillo	30	4

Tabla IV. Frecuencia de términos utilizados en Google para acceder a la revista

Todo los términos se inscriben perfectamente en el ámbito léxico de la revista. Cabe mencionar que “Colima México” corresponde a la sede de uno de los seminarios de *Cybermetrics* y “bibexcel” a un software para cálculos bibliométricos. Se destaca también que 7 de los términos más frecuentes posicionan a la revista entre los 5 primeros puestos, y que estos términos suelen ser los conceptos más apropiados para describir a la revista, con lo que podemos considerar a la revista con un alto grado de visibilidad. También es reseñable que los términos “scientific definitions” y “siting websites” generan una posición muy alejada (fuera de los 1000 resultados) debido a lo genérico de palabras como “scientific” o “websites”. Por otro lado, también se puede apreciar que no existe relación alguna entre el posicionamiento y la frecuencia de los términos de consulta. Este hecho puede verse afectado por el número de documentos recuperados, filtros, país de consulta, etc.

## Discusión

El uso de ficheros log como herramienta de *web data mining* genera una amplia discusión en distintos aspectos (5,20,21,22).

Por un lado, el alto número de IPs dinámicas y el creciente uso de proxies dificulta la identificación del usuario real que está realizando el acceso. En nuestro caso, el alto grado de proveedores de Internet y servicios intermedios no nos permite precisar a la población que consulta o accede a la revista, generándonos un margen considerable de error (24, 9%). Se puede convenir que la mayoría de estas conexiones sean particulares, pero también es posible que instituciones o entidades con menos recursos utilicen a estos intermediarios para su acceso a Internet, por lo que estos datos sólo pueden tener un valor orientativo.

Por otro lado, es difícil cuantificar el consumo de información en Internet. La visibilidad en los buscadores y el número de enlaces que un web recibe son factores que hacen aumentar las visitas. Pero estos accesos o visitas pueden ser motivados por un interés específico o simplemente por una casualidad de la navegación, por lo que la relación acceso/consumo debe contemplarse con cautela. Sin embargo, ciertas publicaciones electrónicas están contabilizando como consumo los accesos no a documentos web, como el html, sino a otros formatos como pdf (Portable Document Format) y ps (PostScript). La necesidad de un software añadido para leer estos documentos y su idoneidad para su impresión hacen pensar que este tipo de accesos sí está más correlacionado con el consumo de información. En el Institute for Future (17) estiman que el acceso a pdf es una señal de valoración, ya que se realiza después de su exploración en html. En nuestro caso, no se ha apreciado una disparidad excesiva entre accesos en pdf y en html, salvo en el mencionado primer artículo de Rousseau.

También se ha podido apreciar que el número de accesos a distintas secciones depende del número de páginas que albergue esa sección o directorio. Como se ha podido ver en el apartado Secciones, hemos optado más por el análisis de los promedio que por los valores totales, lo cual nos ha valido para encontrar una mayor claridad en los resultados.

El hecho de contar con una publicación irregular e intermitente, con escaso número de artículos, y con los ficheros log de sólo los tres últimos años, no nos ha permitido detectar conductas o patrones en este trabajo que nos permitiera extrapolarlos al resto de la realidad web y en concreto al de las publicaciones electrónicas. Por ejemplo, no se ha podido detectar relación consistente entre el consumo y el número de citas que reciben todos los trabajos publicados (19), lo que nos hubiera servido como indicador en la valoración de los trabajos.

Sin embargo, las características de difusión y acceso de estas publicaciones pueden modificar las formas clásicas de comunicación científica (aumento de la visibilidad y con ello de las citas, bajo o nulo coste de distribución, rapidez en el proceso de evaluación y publicación, etc.) y con ello la evolución e interrelación de los mismos agentes científicos. Por ejemplo, Institute for Future (17) ha detectado que las revistas electrónicas que aportan información añadida a modo de portal, incrementa considerablemente sus visitas y que los usuarios valoran sus enlaces en la misma medida que valoran sus artículos.

## Conclusiones

Como primera conclusión podemos destacar que el uso de ficheros log ha permitido de forma satisfactoria conocer el consumo que se ha realizado de la revista a lo largo de tres años. Concretamente, ha permitido conocer la evolución temporal de los accesos y confirmar un relanzamiento de la publicación en los últimos meses de 2003, pese al leve descenso que apuntaba en los dos años anteriores.

Por otro lado, la gran dispersión de los accesos permite concluir que la revista *Cybermetrics* posee una gran presencia internacional, confirmándolo el hecho de que sólo el 15,65% de los accesos son españoles.

También podemos constatar que la sección *Journal* es la más visitada, confirmándose como la más relevante dentro de la publicación. Aún así, merece la pena destacar el hecho de que el 40% de los accesos se dirigen a otras secciones. Esto nos lleva a la conclusión de que no sólo estamos hablando de una revista científica sino también del principal portal en Internet de Bibliometría y Cienciometría. Este hecho queda reforzado cuando examinamos el posicionamiento de la revista en Google y apreciamos que los términos más significativos y que mejor describen a la revista la posicionan en los primeros lugares de este buscador.

Por último, concluimos que los artículos más visitados son los de reciente publicación, por lo que el factor novedad es un elemento clave en los accesos a una revista electrónica. Pese a esto, también se aprecia que determinados trabajos con más antigüedad, los cuales podemos considerar “claves” en una disciplina como el trabajo de Rousseau de 1998, acumulan una gran cantidad de accesos convirtiéndose así en un factor que incrementa considerablemente las visitas a una revista.

## Referencias

1. Björneborn, L. & Ingwersen, P. Perspectives of webometrics. *Scientometrics*, 2001, 50(1): 65-82
2. Koehler, W., Aguilar, P., Finarelli, S., Gaunce, C., Hatchette, S., Heydon, R., McEwen, E., Mahsetky-Poolaw, W., Melson, C. T., Patterson, R., Stahl, M., Walker, M. A., Wall, J. & Wingfield, G. A bibliometric analysis of select information science print and electronic journals in the 1990s. *Information Research*, 2000, 6(1). Available at: <http://informationr.net/ir/6-1/paper88.html>
3. Hawkins, D. T. Bibliometrics of electronic journal in information science. *Information Research*, 2001, 7(1). Available at: <http://informationr.net/ir/7-1/paper120.html>
4. Peters, T. The history and development of transaction log analysis. *Library High Tech*, 1993, 11(2), 41-58
5. Baeza-Yates, R. Excavando en la web. *El Profesional de la Información*, 2004, 13(1): 4-10
6. Silverstein, C., Henzinger, M., Marais, H. & Moricz, M. Analysis of a very large Web search engine query Log, *SIGIR Forum*, 1999, 33(1), 6-12.
7. Spink, A., Jansen, B. J. & Ozmultu, H. C. Use of query reformulation and relevance feedback by Excite users, *Internet Research: Electronic Networking Applications and Policy*, 2000, 10(4), 317-328.
8. Spink, A., Wolfram, D., Jansen, B. J. & Saracevic, T. Searching the Web: the public and their queries, *Journal of the American Society for Information Science and Technology*, 2001, 52(3), 226-234.

9. Blečić, D.D., Dorsch, J., Koenig, M. & Bangalore, N. A longitudinal study of the effects of OPAC screen changes on searching behavior and searcher success. *College & Research Libraries*, 1999, 60: 515-530
10. Rozic-Hristovski, A., Hristovski, D. & Todorovski, L. Users' information-seeking behavior on a medical library Web site. *Journal of the Medical Library Association*, 2002, 90: 210-217
11. Hochheiser, H. & Schneiderman, B. Using Interactive Visualizations of WWW Log Data to Characterize Access Patterns and Inform Site Design, *Journal Of American Society for Information Science and Technology*, 2001, 52(4), 331-343
12. Cooley, R., Mobasher, B. & Srivastava, J. Data Preparation for Mining World Wide Web Browsing, *Patterns Knowledge and Information Systems*, 1999, 1(1)
13. Gomory, S., Hoch, R., Lee, J., Podlaseck, M., Schonberg, E. Analysis and Visualization of Metrics for Online Merchandizing. In: WebKDD, Springer, San Diego, CA, 1999
14. Hitchcock, S., Carr, L. & Hall, W. A survey of STM Online Journals 1990-95: the Calm Before the Storm. In: Directory of Electronic Journals, Newsletters and Academic Discussion Lists, edited by D. Mogge, sixth edition (Washington, D.C.: Association of Research Libraries), 1996, pp. 7-32.  
Available: <http://journals.ecs.soton.ac.uk/survey/survey.html>
15. Okerson, A. Recent Trends in Scholarly Electronic Publishing. Seminar on Multimedia Scholarly Publishing, Helsinki, May 29, 1997 Available: <http://www.library.yale.edu/~okerson/recent-trends.html>

16. Huber, C. F. Electronic Journal Publishers: A Reference Librarian's Guide. *Issues in Science and Technology Librarianship*, Summer 2000 Available: <http://www.library.ucsb.edu/istl/00-summer/article2.html>
17. Institute for the Future. E-journal user study. Report of Web log data mining (SR-786). Menlo Park, CA: Stanford University, 2002 Available: <http://ejust.stanford.edu/logdata.html>
18. Nielsen//NetRanking. The global standard for digital media measurement and analysis. [Milpitas, CA: Nielsen//NetRanking], 2002 Available at: <http://www.nielsen-netratings.com/>
19. Thomson ISI. Web of Knowledge. [Philadelphia, PA: ISI], 2004. <http://isiwebofknowledge.com/>
20. Thelwall, M., Vaughan, L. & Björneborn, L. Webometrics. *ARIST 39*, 2003 (in press)
21. Pitkow, J. In Search of Reliable Usage Data on the WWW, *Proc. of 6th Int. World Wide Web Conference, 1997*
22. Thelwall, M. Web log file analysis: Backlinks and queries. *ASLIB Proceedings*, 2001, 53, 217-223