Artículo

Archivos de datos en línea para ciencias sociales

Por **Óscar Roba Stuart**

Óscar Roba Stuart, licenciado en sociología por la Univ. Católica de Lovaina, Bélgica, en 1984, es actualmente prof. adj. responsable de la Unidad de Computación del Dpto. de Sociología y Coordinador del Área Socio-demográfica del Banco de Datos, ambos de la Facultad de Ciencias Sociales, Univ. de la República, Uruguay. Doctorando en la Universidad de Zaragoza. Ha sido ponente en congresos y ha dictado numerosos cursos sobre métodos estadísticos e informática aplicada a las ciencias sociales.

Resumen: *Actualmente existen herramientas para localizar fuentes de datos múltiples en archivos de ciencias sociales, visualizar metadatos, consultar y analizar datos en línea y descargar subconjuntos de los mismos en diversos formatos. En este artículo se hace su estudio descriptivo, sistematizando la información referida a ellos según lo presentado en sus páginas web, analizando la información que ofrecen, y considerando las perspectivas de evolución de los servicios. Los archivos se presentan agrupados en 3 grupos: 20 europeos, 16 norteamericanos y 10 del resto del mundo. Los norteamericanos son los mayores, y los que tienen mejores sistemas y facilidades de búsqueda. También están a la vanguardia en el tratamiento de información estadística en línea. Los europeos han desarrollado más el acceso en línea a facetas técnicas básicas como son los catálogos, las variables utilizadas en los estudios y las preguntas formuladas en las encuestas.*

Palabras clave: *Archivos de datos en ciencias sociales, Servicios de datos y estadísticos, Servicios de datos electrónicos, Centro de datos, Microdatos, Sistemas de información, Datos estadísticos, Encuestas, Censos.*

Title: Online data archives for the social sciences

Abstract: *There currently exist tools for locating multiple data sources in social sciences archives, visualising metadata, searching and visualising online data, and downloading subsets in different formats. This study describes these archives by systematising the information found on their web pages, analysing the information offered, and considering the potential for further developing these services. The archives are divided into 3 groups: European (20), North American (16), and other countries from around the world (10). The North American archives are the largest and have the best systems and the most user-friendly search tools. They are also leaders in the use of statistical information online. The Europeans have developed greater online access to basic technical facets such as catalogues, variables used in the studies, and questions asked in the surveys.*

Keywords: *Social sciences data archives, Data and statistical services, Electronic data services, Data centres, Microdata, Information systems, Statistical data, Surveys, Censuses.*

Roba Stuart, Óscar. *"Archivos de datos en línea para ciencias sociales". En: El profesional de la información, 2003. septiembre-octubre, v. 12, n. 5, pp. 400-410.*

I. Introducción

Este trabajo tiene su origen en la inquietud por conocer más profundamente lo ofrecido por los bancos de datos para ciencias sociales en sus páginas web.

Estos recursos informativos se han constituido con la finalidad de conservar y difundir los resultados de encuestas, censos e investigaciones en general, ya sean cuantitativas o cualitativas, a efectos de hacerlas disponibles para la comunidad científica y especialmente para los investigadores y docentes en ciencias sociales. Pueden ser utilizados para los siguientes fines: análisis secundario en investigación básica, práctica o de políticas; obtención de información precisa sobre características, actitudes y comportamientos de actores sociales; comparaciones en el tiempo y en el espacio entre diferentes estudios; testar nuevas teorías e hipótesis; realizar estudios piloto de preguntas y técnicas aplicables en una futura investigación; estudios metodológicos sobre datos perfectamente estructurados; formación de estudiantes en los métodos de análisis en ciencias sociales; corregir y evaluar muestras.

Los archivos de datos acumulan información generalmente en soporte magnético. De esta manera se

Artículo recibido el 05-06-2003
Aceptación definitiva: 06-08-2003

garantiza mejor que el resultado de costosas investigaciones pueda ser utilizado nuevamente por otros usuarios. Para cumplir sus funciones deben realizar tareas de adquisición, almacenamiento, validación, catalogación y difusión. Pueden ser utilizados por departamentos académicos, institutos de investigación y oficinas gubernamentales que desarrollen programas y políticas de investigación y enseñanza.

«Los archivos analizados se encuentran en diferentes fases de evolución, desde los que ofrecen una simple presentación estática y sin posibilidad de búsqueda hasta los que permiten procesar datos en línea»

Aunque comenzaron sus actividades hace varias décadas, no ha sido hasta hace unos pocos años que tienen presencia en internet. Aún hoy es considerable la diferencia que existe entre ellos en la oferta de servicios: los hay desde la simple presentación estática y sin posibilidad de búsqueda hasta los que han desarrollado herramientas para procesar datos en línea. Desde el punto de vista institucional dependen de institutos privados autofinanciados, universidades, o de consejos de investigación en ciencias sociales. Algunos están muy vinculados a las cátedras de sociología y ciencia política pero otros forman parte de bibliotecas. Generalmente tienen en común alguna relación con la enseñanza de las ciencias sociales y su ubicación en el hemisferio norte. Si bien son instituciones autónomas, a veces únicas por país (especialmente en Europa), intercambian información entre sí. Por otro lado ya se han desarrollado también bancos transnacionales.

El mayor archivo de datos en castellano es el *Archivo de estudios sociales* (*Arces*) del *Centro de Investigaciones Sociológicas* (*CIS*) de España[1]. Éste permite acceder vía web a metadatos de: su base (más de 1.600 estudios con alrededor de 75.000 preguntas), investigaciones del *Cires* (*Centro de Investigación de la Realidad Social*), *Issp* (*International Social Survey Program*), *Icpsr* (*Inter-university Consortium for Political and Social Research*), *Roper Center*, *Cessda* (*Council of European Social Science Data Archives*), encuesta nacional de valores, y los latinobarómetros, entre otros. Se pueden solicitar distribuciones de frecuencia marginales unidimensionales, tabulaciones pluridimensionales, archivos de microdatos, series temporales y estudios cualitativos, por teléfono, fax, correo, correo-e o personalmente.

El catálogo del banco de datos del *CIS* permite consultar las encuestas realizadas en el *Centro* mediante distintos tipos de búsquedas: por índice temático (estado, constitución e instituciones políticas, justicia, comunidades autónomas, elecciones, temas internacionales, economía, trabajo, medio ambiente y urbanismo, educación, ciencia y tecnología, sanidad, población, problemas sociales, medios de comunicación, ocio, distribución del tiempo, cultura, fiestas de Navidad, religión, valores, cuestiones de actualidad, prospectiva, barómetros de opinión), por índice anual, número de estudio y búsqueda avanzada. En ésta encontramos los siguientes campos: referencias con fecha (intervalo de realización del trabajo de campo o relevamiento de información), tamaño de la muestra, existencia de publicación, tipo de investigación (cualitativa o no), ámbito (comunidad, etc.), buscar palabras en el título, en el contenido, universo (amas de casa, jóvenes, etc.).

Las fichas técnicas de los estudios del *CIS* presentadas en su página contienen: título del trabajo, periodicidad, donante, lengua, número de submuestras, fecha, universo, ámbito, número de observaciones, método, muestreo, tamaño del archivo, número y lista de variables.

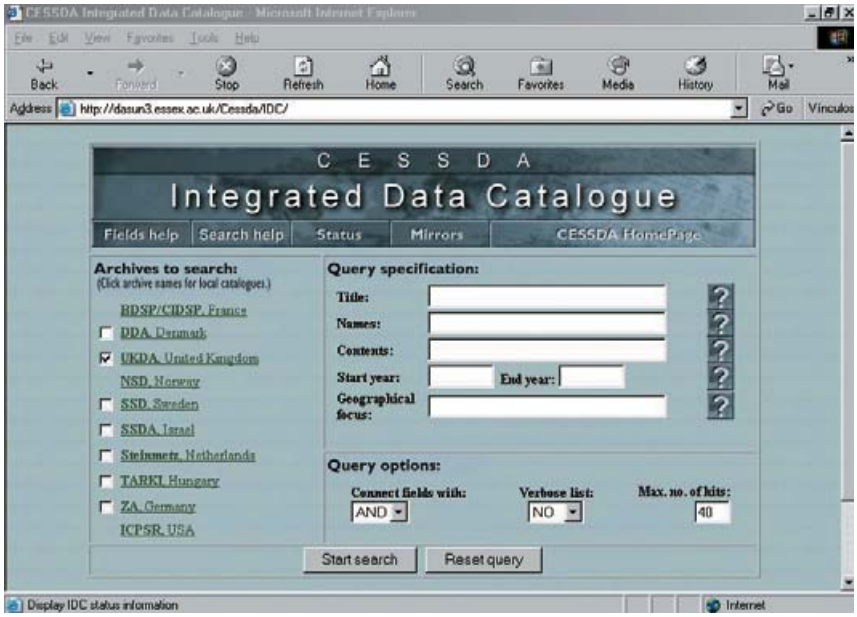
El web del CIS no ha avanzado en la posibilidad de realizar operaciones con los datos, elegir entre diversas opciones de presentación o bajar información directamente, pero en cambio ha realizado un gran esfuerzo en la actualización de sus barómetros. Presenta un avance provisional de resultados —disponible luego de 15 días de finalización del trabajo de campo, codificación y digitación— para los barómetros mensuales y para las encuestas de carácter electoral y político. El avance incluye la distribución de frecuencias de las preguntas.

II. Antecedentes

Según **Cordero** (1998), los archivos se iniciaron en 1945, cuando **Elmo Roper** en Estados Unidos dona datos estadísticos que había acumulado durante 10 años a una biblioteca universitaria. Pero las cajas de tarjetas perforadas se trataron como si fueran libros: se acumuló información pero no se avanzó en su recuperación para otros usuarios. Esto recién ocurrirá en 1957 cuando se constituye el *Roper Public Opinion Research Center*. En los años posteriores se fundaron otros archivos: en 1962 en la *Universidad de Michigan*, Ann Arbor, se reúnen 24 universidades para constituir el *Icpsr* (*Inter-university Consortium for Political and Social Research*) —hoy son más de 500—, y en la *Universidad de California Berkeley* se crea en 1964 un banco de datos censales. Con el correr de los años se irán agregando otros en diversas universidades norteamericanas.

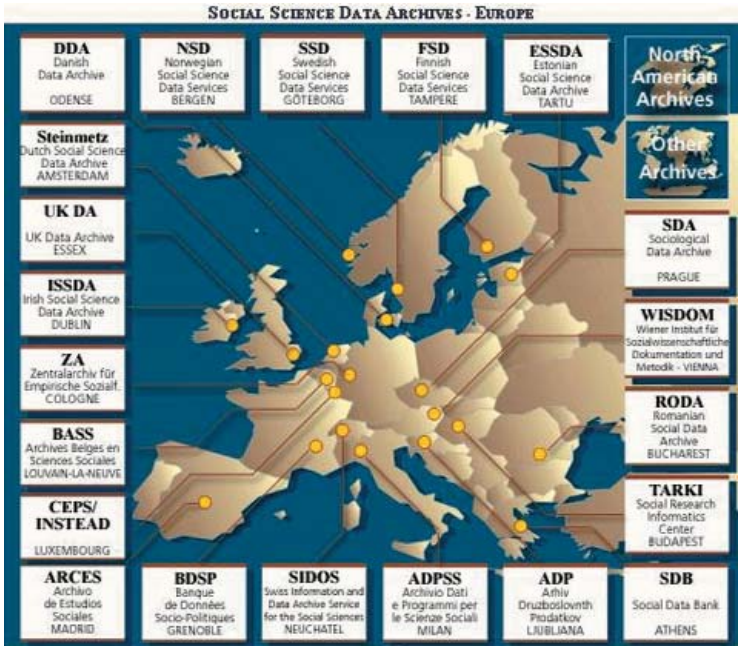
En Europa el primero fue el *Archivo Central* (*ZA*, *Zentralarchiv für Empirische Sozialforschung*), Colo-

nia, Alemania, creado en 1960. En los años posteriores se irán creando otros en diversos países. En 1968 la *Unesco* autoriza el estudio de un centro internacional de archivos de encuestas (**Nasatir**, 1973). En 1976 se funda en Amsterdam el *Council of European Social Science Data Archives* (*Cessda*)[2] y en 1977 se crea en Lovaina la Nueva, Bélgica, la *Ifdo* (*International Federation of Data Organizations*)[3], con la ayuda del *Comité Permanente de Archivos de Datos de Ciencias Sociales* del *Consejo Internacional en Ciencias Sociales* de la *Unesco*.



Desde mediados de la década de los 90 se comienza a pensar en Estados Unidos la necesidad de integrar las ciencias sociales en la revolución que se estaba viviendo en la informática y las comunicaciones. **Rockwell** (1994) plantea la necesidad de documentación electrónica en lugar de impresa y de integrar en red el acceso a los datos y su documentación. Su ejemplo de búsqueda usando texto libre y expresiones booleanas constituye un modelo que tendrán como referencia los proyectos posteriores. Más adelante **Rockwell**, **Hardin** y **Loots** (1995) analizan la muerte de los estándares aceptados hasta ese momento, la necesidad de otros y enumeran los servicios que mejorarán la relación entre los usuarios de datos y una combinación de centro de computación, biblioteca de datos y biblioteca regular, que usa las nuevas herramientas: www, sgml, html y el protocolo de búsqueda Z39.50.

Los investigadores y profesionales de la información pueden ahora buscar en internet las variables e investigaciones de su interés. Para ello se han elaborado catálogos conjuntos entre varios archivos —en su mayoría europeos— donde se realizan búsquedas de la misma forma que en las bases de datos bibliográficas. Un ejemplo es el catálogo del citado *Cessda*[4] y el proyecto *Nesstar* (*Networked social science tools and resources*). Pero mientras el primero es un catálogo integrado de datos de los archivos de Francia, Dinamarca, Reino Unido, Noruega, Suecia, Holanda, Hungría, Alemania, Estados Unidos e Israel, donde es posible buscar investigaciones por título, nombres de los investigadores, resumen, año y región geográfica, el segundo es un sistema integrado de búsqueda, análisis y difusión.



El proyecto *Nesstar* ha sido realizado dentro del *4º Programa Marco* de la *Comisión Europea* y tiene como finalidad que sus usuarios puedan localizar fuentes de datos múltiples, independientemente de su ubicación nacional, visualizar metadatos, consultar y analizar datos en línea —mediante tabulados y gráficas— y descargar subconjuntos de los mismos en diversos formatos. De una manera relativamente fácil, los archivos ofrecen la posibilidad de trabajar conjuntos de datos en diversos formatos y en el PC de los usuarios. El acceso puede ser libre o restringido a determinado tipo de usuarios registrados.

En su versión inicial de 1999, la herramienta de usuario final del *Nesstar*, el Explorer, podía ser instalado en cualquier plataforma Java, incluyendo Windows y Mac, con un buscador web y una conexión a internet.

El programa, que instalaba una conexión directa desde el escritorio, permitía realizar búsquedas de información en los cuatro archivos que inicialmente trabajaron en este proyecto: *Danish data archive* (*DDA*), *Finnish social science data services* (*FSD*), *Norwegian social science data services* (*NSD*) y *UK data archives* (*Ukda*). Una vez realizada la búsqueda se seleccionaban las variables para un análisis. El programa permitía seleccionar casos, ponderar variables, distribuir frecuencias, tabular y hacer gráficos. También era posible bajar el archivo de microdatos correspondiente, si se encontraba disponible. El entorno era muy amigable y estaba previsto que se incorporaran otros archivos, así como versiones del programa en varias lenguas europeas y la posibilidad de realizar búsquedas en ellas, con traducción automática.

El proyecto *Nesstar* ha experimentado varios cambios y recién en este año se espera una nueva versión que cumpla en su totalidad con lo estipulado anteriormente. Mientras tanto los usuarios pueden seguir usando el catálogo integrado de datos del *Cessda*. Una completa descripción del proyecto *Nesstar* se encuentra en **Ryssevik** y **Musgrave** (2001) y en su sitio web[5].

En EUA también se han llevado a cabo proyectos para administrar y difundir datos cuantitativos así como procesar datos en línea. En particular merecen destacarse *VDC* (*Virtual Data Center*) y *SDA* (*Survey documentation and analysis*). El primero es un software de código abierto desarrollado en la *Universidad de Harvard* con el apoyo de la *National Science Fundation* que facilita la preparación de catálogos de estudios en diferentes formatos mediante la creación de colecciones virtuales de datos, administración independiente de archivos de datos y documentación, soporte de diferentes protocolos de búsqueda, extracción de datos y análisis exploratorio de los mismos, conversión entre formatos, distribución flexible de colecciones y mejora de las citas sobre datos en periódicos académicos. Más información puede encontrarse en **Altman** y otros (2001) y en su web[6].

«Los bancos de datos se han creado para asegurar que datos numéricos y textuales de interés para las ciencias sociales se conserven frente a cambios tecnológicos y daños físicos»

SDA ha sido elaborado por el *Computer-assisted survey methods program* de la *Universidad de California*, en Berkeley[7]. Es un conjunto de programas para la documentación y análisis de datos de encuestas a tra-

Lista de archivos

EUROPEOS

Archives belges en sciences sociales (Bass)
Université Catholique de Louvain, Belgique.
http://logi.rspo.ucl.ac.be/

Archivio dati e programmi per le scienze sociali (Adpss), Sociodata
Dipartimento di Sociologia e Ricerca Sociale dell'Università degli Studi di Milano-Bicocca, Italia.
http://www.sociologia.unimib.it/sociodata/

Banco de datos del CIS
Centro de Investigaciones Sociológicas (CIS), Mº de la Presidencia, España.
http://www.cis.es/

Banque de données socio-politiques (Bdps)
Unidad de Servicio e Investigación del Cnrs conjuntamente con la Université Pierre Mendès France y el IEP de Grenoble, France.
http://www-bdsp.upmf-grenoble.fr/contenfr.htm

Central archive for empirical social research (ZA)
Instituto de The Cologne Association for Social Research e.V. (KGS) en estrecha cooperación con la Universidad de Colonia, Alemania.
http://www.gesis.org/en/za/index.htm

Danish data archive (DDA)
Unidad independiente en el grupo Danish State Archives, Dinamarca.
http://www.dda.dk/

Estonian social science data archive (Essda)
Faculty of Social Sciences, Tartu University, Estonia.
http://psych.ut.ee/esta/

Finnish social science data archive (FSD)
Universidad de Tampere, Mº de Educación, Finlandia.
http://www.fsd.uta.fi/

Irish social science data archive (Issda)
Auspiciado por el Institute for Study of Social Change y administrado conjuntamente por University College Dublin y el Economic and Managed Social Research Institute, Irlanda.
http://www.ucd.ie/~issda/

Latvian social science data archive (Lszda)
Fundado por el Latvian Council of Science (Social development and social security program), Lituania.
http://www.lszda.lv/

Norwegian social science data services (NSD)
Research Council of Norway, universidades noruegas, Mº de Educación, Investigación y Asuntos Eclesiásticos y el Mº de Sanidad y Asuntos Sociales, Noruega
http://www.nsd.uib.no/

Service suisse d'information et d'archivage de données pour les sciences sociales (Sidos)
Fundación de la Académie Suisse des Sciences Humaines et Sociales (Assh), Suiza.
http://www.sidos.ch/

Slovenian social science data archive (ADP)
Fundado por la Facultad de Ciencias Sociales, Universidad de Ljubljana y subsidiado por el Mº de Ciencia y Tecnología, Slovenia.
http://www.adp.fdv.uni-lj.si/

Social data bank (SDB)
National Center for Social Research (Ekke), Grecia.
http://www.ekke.gr/databank/

Sociological data archive (SDA)
Institute of Sociology, Academy of Sciences of the Czech Republic, República Checa.
http://archiv.soc.cas.cz/enindex.htm

Steinmetz archive (Star)
Netherlands Institute of Scientific Information Services (Niwi), Royal Netherlands Academy of Arts (Knaw), Holanda.
http://www.niwi.knaw.nl/us/dd_star/dd_star.htm

Swedish social science data service (SSD)
Universidad de Göteborg, Suecia.
http://www.ssd.gu.se/enghome.html

Tárki data archive
Tárki Social Research Centre Inc., Hungría.
http://www.tarki.hu/index-e.html

UK data archive (Ukda)
University of Essex, Economic and Social Research Council (Esrc) y Joint Information Systems Committee (Jisc) de la Higher Education Funding Councils, Reino Unido.
http://www.data-archive.ac.uk

Wiener Institut für Sozialwissenschaftliche DOkumentation und Methodik (Wisdom)
Parcialmente financiado por el Mº de Ciencia e Investigación, Austria.
http://www.wisdom.at

NORTEAMERICANOS

Data and program library service (Dpls)
Data and Computation Center (Dacc), College of Letters and Science, University of Wisconsin-Madison, EUA.
http://dpls.dacc.wisc.edu/

Data and statistical services (DSS)
Firestone Library's Social Science Reference Center, Princeton University, EUA.
http://www.princeton.edu:80/~data/index.html

Data Center, Carleton University Library
Madgic (Maps, Data and Government Information Centre), Canadá.
http://www.carleton.ca/~ssdata/

Data resources library (DRL)
Social Science Computing Laboratory (Sscl), Faculty of Social Science, University of Western Ontario, Canadá.
http://www.sscl.uwo.ca/sscl/drl/

Electronic data service (EDS)
Operado conjuntamente por Columbia University Libraries and Academic Information Systems (AcIS), EUA.
http://www.cc.columbia.edu:80/acis/eds/

Geospatial & Statistical Data Center (Geostat)
University of Virginia, Library, EUA.
http://fisher.lib.virginia.edu/

Harvard-MIT Data Center (Hmdc)
Harvard University y Massachusetts Institute of Technology, EUA.

http://data.fas.harvard.edu/hdc/

Inter-University Consortium for Political and Social Research (Icpsr)
Institute for Social Research de la University of Michigan, EUA.
http://www.icpsr.umich.edu/

Issr data archives
Institute for Social Science Research, University of California, Los Angeles, EUA.
http://www.sscnet.ucla.edu/issr/da/

Louis Harris Data Archive
Howard W. Odum, Institute for Research in Social Science (Irss), University of North Carolina at Chapel Hill, EUA.
http://www.irss.unc.edu/data_archive/home.asp

Social science data archives (Ssda)
Main Library, University of California, Irvine, EUA.
http://data.lib.uci.edu/

Social sciences data collection (Ssdc)
Data Services Unit, Biblioteca, University of California, San Diego, EUA.
http://ssdc.ucsd.edu/

The Roper Center for Public Opinion Research
University of Connecticut, EUA.
http://www.ropercenter.uconn.edu/

TriUniversity Data Resources
University of Guelph, University of Waterloo y Wilfrid Laurier University, Canadá.
http://drc.uoguelph.ca/

UBC Numeric Data Services
Humanities and Social Sciences, Government Publications Division, Library, University of British Columbia, Canadá.
http://data.library.ubc.ca/

University of Alberta Data Library
Humanities & Social Sciences Library, University of Alberta, Canadá.
http://www.library.ualberta.ca/datalibrary/index.cfm

RESTO DEL MUNDO

Área socio-demográfica del banco de datos
Servicio Técnico de la Facultad de Ciencias Sociales, Universidad de la República, Uruguay.
http://www.rau.edu.uy/fcs/banco/banco.htm

Icssr data archives
Indian Council of Social Science Research, India.
http://www.icssr.org/data_archive.htm

Israel Social Sciences Data Center (Isdc)
Faculty of Social Sciences, Hebrew University of Jerusalem, Israel.
http://isdc.huji.ac.il/

Korean Social Science Data Center (Ksdc)
http://www.ksdc.re.kr/

National survey data bank on public opinion
Dependencia del Centro de Estudos de Opinião Pública (Cesop) de la Universidade Estadual de Campinas, Brasil.
http://www.unicamp.br/cesop/Bancodados.htm

New Zealand social research data archives (Nzsrda)
Massey University, Nueva Zelanda.
http://www.massey.ac.nz/~nzsrda/

Social science data archives (Ssda)
Research School of Social Sciences, Australian National University, Australia.
http://ssda.anu.edu.au/

Social science Japan data archive (Ssjda)
Integrante del Information Center for Social Research on Japan, Institute of Social Science, University of Tokyo, Japón.
http://ssjda.iss.u-tokyo.ac.jp/pages/ssjda-e/

South African data archive (Sada)
Integrante del National Research Foundation (NRF), South Africa.
http://www.nrf.ac.za/sada/

Survey research data archive (Srda)
Integrante de la Office of Survey Research (OSR), Academia Sinica, Taiwan, R.O.C.
http://www.sinica.edu.tw/as/survey/srda/english/tot_list.htm

vés del web. Los programas permiten realizar búsquedas de datos, localizarlos, crear subconjuntos, obtener sus códigos así como calcular frecuencias de variables, tablas, comparación de medias, correlaciones, regresiones y listas de valores de casos individuales. Otras técnicas están previstas en el futuro.

Los programas del *SDA* están diseñados para correr desde un navegador web y pueden obtener resultados de procesamiento de datos en pocos segundos. También pueden crear subconjuntos de variables y/o casos y descargar los datos en formato ascii, *Spss*, *Sas* o *Stata*, conjuntamente con toda la documentación.

Los archivos europeos se han diferenciado de los norteamericanos en su fuente de financiamiento. Los primeros la obtienen del presupuesto universitario y/o gubernamental, aunque algunas actividades pueden ser financiadas con fondos externos. Los segundos en cambio deben obtener recursos para financiarse (se espera que se autofinancien aunque en muchos casos obtengan fondos públicos para hacerlo). Esta diferencia tiene sus consecuencias al emprender tareas costosas de recuperación de datos históricos.

III. Objetivos de la investigación

Aunque *Cessda* y la mayoría de archivos mantienen páginas con enlaces hacia otros archivos —en particular se destaca la del *Social sciences data collection* (*Ssdc*) de la *Universidad de California*[8] y la de los miembros de *Ifdo*—, los listados no contienen mucho más que el nombre, ubicación y alguna característica del archivo. No se encuentran sistematizados de forma tal que se facilite su lectura y carecen de información fundamental.

Un buen intento de confeccionar un cuadro comparativo entre archivos lo encontramos en el trabajo de **Guy** y **Haley** (1995), pero con la expansión de los mismos y la evolución de internet este trabajo se encuentra hoy totalmente superado. Además, la cantidad

de variables tenidas en cuenta era bastante pequeña, por lo que consideramos que había llegado el momento de actualizar tal trabajo, integrando nuevas variables.

Este artículo presenta una versión resumida de los resultados del análisis de archivos de ciencias sociales en línea realizado en el marco del *Doctorado en Sistemas de Información y Documentación* de la *Universidad de Zaragoza*.

Se trata de un estudio descriptivo que posee como primer objetivo sistematizar la información referida a los archivos, según lo presentado en sus páginas web, a efectos de obtener una rápida visualización de sus actividades y posibilidades. El segundo objetivo es analizar la información ofrecida con la finalidad de mejorar la difusión de sus actividades de almacenamiento y asistencia a la investigación y enseñanza.

En este artículo presentamos solamente los datos de identificación —nombre, afiliación institucional y URL— de los archivos y el análisis cuantitativo de las variables estudiadas. Se incluye igualmente un análisis de las perspectivas de desarrollo de los servicios.

IV. Marco conceptual y funcional de los archivos de datos para ciencias sociales

Actualmente existe una importante cantidad de oferta de datos en el mercado, accesible gracias a internet. Oficinas de estadística, organizaciones gubernamentales, empresas de opinión pública e instituciones académicas producen grandes cantidades de datos. Pero muchas veces los posibles usuarios no logran conocer los mismos, sus posibilidades de búsqueda son reducidas y los datos terminan usándose únicamente como fuente primaria.

«El mayor de ellos en castellano es el *Archivo de estudios sociales* (*Arces*) del *Centro de Investigaciones Sociológicas* (*CIS*), España»

A pesar de cierta confusión en la terminología utilizada, la definición más ajustada a nuestro trabajo es que un archivo de datos es un sistema de información que integra datos y metadatos. Entre los últimos encontramos en las ciencias sociales:

—significado de las variables: definiciones, etiquetas, textos de preguntas;

—características técnicas relacionadas con la producción del dato: ámbito, universo, tipo de muestreo, tamaño muestral;

—cambios metodológicos relevantes a lo largo del tiempo: diferentes clasificaciones, cambios de definición;

—información de interés para el investigador: fuentes, referencias, valoraciones realizadas por otros investigadores, estimaciones.

La información se puede clasificar según su forma, exactitud, origen, temporalidad, tipo de acceso, etc. Por su forma se puede considerar según su nivel de medición, presentación, soporte y su nivel de agregación. Por su nivel de medición la clasificamos en cuantitativa o cualitativa. De la primera los archivos de datos almacenan números y de la segunda, textos e imágenes. La información numérica y textual codificada puede presentarse bajo la forma de microdatos —elementos mínimos de información provenientes de registros, encuestas o censos— o datos agregados —aquellos que provienen de los microdatos mediante alguna transformación matemática—. El resto de la información textual se presenta como documentos completos, informes, leyes, noticias de prensa, etc. Si consideramos a los datos según su soporte encontramos que se pueden hallar impresos o en un medio magnético-óptico. También se pueden clasificar según se consideren tratables o no por máquinas.

El usuario de los archivos de datos puede acceder a los siguientes productos:

1. Sobre la información depositada: microdatos, datos agregados; frecuencias de variables, cuadros; selección de subconjuntos; variables derivadas y estimadores; gráficos.

2. Sobre la metainformación e información referencial: base de datos de preguntas y variables; cuestionarios, libros de códigos; ficha técnica de los estudios, catálogo indexado; tesauro desarrollado; especificaciones técnicas sobre el formato informático.

De acuerdo con **Scheuch** (1990), los archivos de datos se diferencian de las bibliotecas, además del soporte generalmente utilizado, en que para los primeros la unidad de almacenamiento y recuperación es la variable mientras que para las segundas es el estudio.

V. Material y método

Para confeccionar la lista de variables se consideró el trabajo de **Guy** y **Haley** (1995) y el de **Anagnostelis** y **Cooke** (1997).

Para cumplir el primer objetivo decidimos realizar la búsqueda en internet de los archivos de datos existentes en el mapa de la página del *Cessda*, complementando con la de *Ifdo* y las de otros archivos. Siguiendo la división de *Cessda*, los reunimos en 3 grupos cuyo resultado fue: 20 europeos, 16 norteamerica-

nos y 10 del resto del mundo. De esta forma se cubre la mayor cantidad y los más importantes archivos de datos para ciencias sociales.

La información se obtuvo durante el mes de mayo y primera quincena de junio del año 2002.

Para ingresar la información relevada se creó un archivo en *Excel* con las variables en las columnas y los archivos en las filas. De esta forma se introdujo en línea la información de base directamente desde la página web del archivo y la requerida por las restantes variables.

Para alcanzar el segundo objetivo transformamos ordenadamente los datos de cada archivo en un listado tipo catálogo, en *Word*. Las variables se procesaron en *Spss*, considerando además el comportamiento cualitativo de algunos archivos y de las áreas geográficas.

Dificultades halladas

Cuando alguna de las variables lo requería, se realizaron búsquedas de datos en la página del archivo a efectos de responder los interrogantes planteados. En general se ha elegido hacer búsquedas de archivos censales, dada su universalidad.

Como los archivos de lengua no inglesa presentan una versión en esa lengua, se ha elegido trabajar con ella. En algunos casos se ha debido recurrir a la versión en idioma original ya que la versión en inglés presentaba notorias insuficiencias y atrasos en su actualización.

La variable "número de archivos depositados" ha sido muy difícil de discernir. A pesar de su importancia en el momento de establecer un ranking de archivos generalmente no se encuentra disponible en las páginas, debido quizá a su constante variación. De toda maneras, la mayoría de los casos en que pudimos construir el dato se debió a la posibilidad de enumerar la totalidad de archivos almacenados o a la consulta en las páginas de *Ifdo*.

En relación con una de las variables más importantes del trabajo, posibilidad de descarga de archivos de datos, se adoptó el criterio afirmativo cuando así se indicaba en el web del archivo, aunque se necesitara autorización especial. Esto podría subvalorar aquellos casos en los cuales se entró desde la primera página como usuario común y no registrado, pero este sesgo es inevitable.

El hallazgo de las áreas y temas trabajados por los archivos no fue totalmente satisfactorio, pues aunque en general se encuentran dentro de las ciencias sociales, a lo largo de sus años de formación se han ido es-

pecializando en algunos temas, sin que sean mencionados de forma específica.

Debido a la variación de precios en los archivos de datos y a la necesidad de presentar un valor indicativo, se adoptó el criterio especificado en la variable. En los casos en que existían diferentes categorías según el número de bases o su tamaño, se adoptó la media. Los valores que se encontraban en la moneda local fueron convertidos a dólares americanos.

VI. Análisis de datos

Presentamos a continuación una síntesis del análisis realizado para las principales variables.

Sede. Indica la ubicación geográfica por país. La mayor parte de los archivos reside en el hemisferio norte, fundamentalmente EUA y Europa. En el resto del mundo sólo hay archivos en China, Sudáfrica, Israel, Australia, Nueva Zelanda, India, Brasil, Japón, Corea y Uruguay.

Fecha de fundación. Año de creación institucional del archivo y/o de comienzo de la acumulación de ficheros. Existen notorias diferencias en las fechas de fundación de los archivos. Aunque el dato es limitado —28% no presentaba información—, se puede establecer que los norteamericanos comenzaron a fundarse lentamente entre las décadas de los 40, 50 y 60, los europeos del oeste en los 60, 70 y 80; y los del este y resto del mundo en las últimas dos décadas del siglo. Estas fechas no son homogéneas para ningún grupo.

Afiliación institucional. Nombre de la institución a la cual el archivo pertenece. En su gran mayoría los archivos están vinculados a universidades y en particular a las facultades de ciencias sociales. El resto se encuentra en consejos nacionales de investigación y en ministerios (especialmente en Europa), así como en academias privadas. Los archivos estadounidenses forman parte de bibliotecas, lo que se refleja en su denominación: *data library*. Los europeos se han constituido en unidades independientes, generalmente únicas por país, lo que los diferencia notablemente de los norteamericanos e implica políticas de relaciones entre ellos muy diferentes.

Fecha de actualización del sitio web. Las fechas de las páginas web no pudieron ser encontradas en un 40% de los casos, lo que no implica necesariamente que el web no estuviera actualizado, finalidad de este indicador. De todos modos los archivos que tenían fechas demostraron estar bastante al día, especialmente los de América del Norte.

Número de archivos disponibles. Cantidad total de ficheros de datos depositados. Esta variable también presentó problemas, puesto que a pesar de que se consultaron varias fuentes, sólo se obtuvieron resultados en la mitad de los casos. Como era de esperar, los archivos mayores se encuentran en Estados Unidos: el *DSS* de Princeton y el *Roper Center*. Los mayores europeos son el *ZA* de Alemania y el *Ukda* del Reino Unido. En el otro extremo se encuentran los archivos de Nueva Zelanda y Uruguay. Pero no se pudo obtener esta información sobre el que se considera el mayor archivo del mundo: el *Icpsr* de Michigan.

Descarga de archivos. Posibilidad de enviar ficheros de datos desde una máquina del archivo a la máquina del usuario. Un 46% de los archivos investigados permite que los usuarios puedan bajar ficheros de datos desde sus páginas. A través de la propia página o por ftp, los archivos habilitan que algunas de sus bases de datos se encuentren disponibles ya sea para el público en general o para usuarios que reúnen determinados requisitos.

«Un 65% de los archivos de datos ha desarrollado sistemas de búsqueda de información en línea; de ellos, un 48% empleando lenguaje natural»

Documentación aportada. Se refiere a los metadatos (catálogo, cuestionario y libro de códigos) que acompañan a los ficheros de datos. La cantidad de archivos que permiten la descarga de documentación es mayor que los que permiten descarga de datos. Un 63% autoriza la descarga de catálogos, cuestionarios o libro de códigos de las investigaciones almacenadas.

Procesamiento de datos en línea. Posibilidad de generar en línea datos agregados, seleccionar subconjuntos y calcular variables derivadas, a partir de ficheros de microdatos. Son muy pocos los archivos que permiten algún tipo de procesamiento estadístico de sus datos. Sólo el 15% tiene habilitada esta opción, en su mayoría de América del Norte.

Existencia de limitaciones de acceso a datos. Restricciones a la consulta en línea de los ficheros de microdatos. Casi 3/4 partes de los archivos establecen alguna limitación de acceso a sus datos. Así no podrá obtenerlos quien no pertenezca a su comunidad académica o no se encuentre afiliado o no haya pagado por ellos, etc.

Existencia de publicaciones del archivo. Relativo a la disponibilidad regular y actual de publicaciones del archivo (newsletter, etc.). La mayoría de los archivos (54%) posee algún tipo de publicación propia.

Asistencia a usuarios en línea. Ayuda a las búsquedas de información. La ofrece un 60% de los archivos.

Búsqueda en línea. Posibilidad de realizar consultas en línea a catálogos, índices, variables y ficheros. Un 65% de los archivos tiene sistemas de búsqueda de información en línea, siendo éstos muy variados, desde un simple listado hasta el desarrollo de tesauros y búsqueda en archivos múltiples.

Acceso a catálogos. Consulta en línea de catálogos que contengan las fichas técnicas de las investigaciones.

Acceso a variables. Consultar en línea de la lista de variables de las investigaciones.

Acceso a preguntas. Consulta en línea de la lista de preguntas de los cuestionarios de investigaciones.

Acceso a índices. Consultar en línea de los listados de nombres de las investigaciones.

De las cuatro posibilidades de acceso consideradas, la de catálogos es la más frecuente (63%), seguida de la de índices con un 59%. El acceso a variables y preguntas es considerablemente menor (22 y 9% respectivamente). Es interesante notar que los archivos europeos tienen la proporción más alta en acceso a catálogos, variables y preguntas.

Método de búsqueda. Se refiere a la forma utilizada para hacer ordenadamente la recuperación de información deseada, bien mediante lenguaje natural o lenguaje documental. Casi la mitad de los archivos (48%) ofrece el primero, un 13% ha realizado una combinación entre lenguaje natural y documental; y un 35% no ha diseñado ningún sistema. Los archivos norteamericanos son los que más han implementado sistemas de búsqueda.

Facilidades de búsqueda. Utilización de operadores booleanos, archivos múltiples, detección de duplicados, operadores de proximidad, etc. La mayoría de los archivos (52%) tiene alguna facilidad de búsqueda, siendo la más frecuente los operadores booleanos [solos (17%) o combinados con los de proximidad (13%)]. Las otras combinaciones son de menor uso. La mejor elaboración corresponde a archivos norteamericanos.

Utilización de lenguajes documentales. Posibilidad de utilizar términos y operadores propios que permiten representar el contenido de los documentos. El 78% de los archivos no utiliza lenguajes documentales. Sólo un 13% presenta una lista de palabras claves.

Posibilidad de guardar la sesión de búsqueda. Opción de salvar una búsqueda y poderla ejecutar posteriormente. Solamente la han establecido 3 archivos norteamericanos.

Costos por base. Costo de los ficheros de datos para el usuario no comercial o perteneciente a una institución educativa. La mitad de los archivos no brinda información acerca de los costos que representa para los usuarios el acceder a sus bases de datos y entre los que lo hacen existe una gran dispersión de situaciones. Sólo un 17% menciona sus tarifas. Así observamos que los costos más altos corresponden a los archivos de Holanda y Suiza mientras que el menor a Brasil. Un 15% establece que es libre para el personal universitario (principalmente entre los archivos norteamericanos). Un 11% cobra solamente gastos de soporte y envío.

VIII. Conclusiones

Los archivos de datos para ciencias sociales se ha incrementado en los últimos años, seguramente gracias a la contribución de la informática e internet, pero esto no asegura un aumento formidable en su utilización. Hay trabajos recientes que tratan este tema en las universidades del Reino Unido: **Burnhill** y otros (2001). El trabajo ha permitido estudiar con mayor profundidad algunas de las características de los archivos de datos, tal cual se presentan en las páginas web, y sacar algunas conclusiones.

La diferenciación por áreas que hemos realizado no es meramente geográfica, pues a través de ella se pueden estudiar la evolución y el estado actual de los archivos. Los norteamericanos son los que poseen mayor almacenamiento de información, no sólo sobre su país sino sobre el resto del mundo. Su desarrollo ha sido tal que uno de ellos —el *Icpsr*— vende sus servicios a otros. Entre las características que los diferencian se encuentra el pertenecer casi exclusivamente a universidades, cosa que se nota en sus webs, muy orientados a los principales usuarios, los estudiantes. Quizá sea por ello que han implementado más los sistemas y facilidades de búsqueda de información y también estén a la vanguardia en el tratamiento de información estadística en línea, innovación tecnológica que rápidamente se irá extendiendo por el resto del mundo y que es fundamental para los estudiantes, ya que a veces no poseen equipo ni software para tratar sus datos. Un antecedente de las posibilidades de las tecnologías en línea puede encontrarse en **Clark** y **Maynard** (1998).

«Un 46% de los archivos permite que los usuarios puedan bajar conjuntos de datos a su PC»

Los archivos europeos se encuentran en universidades, consejos de investigación en ciencias sociales y ministerios. Su naturaleza parecería ser más planificada y de alcance nacional. Han desarrollado un acceso mayor en línea a dominios técnicos esenciales como

son los catálogos, las variables usadas y las preguntas formuladas en las encuestas.

El resto del mundo no posee una tendencia definida y sus archivos se encuentran vinculados a universidades y a consejos de investigación en ciencias sociales. Parecería que su evolución depende de las circunstancias locales.

En los próximos años los archivos deberán extender sus actividades. Hacerlo entre los estudiantes de todos los niveles parece un buen camino. Estos tendrán la posibilidad de usar múltiples conjuntos de datos, recogidos en diferentes puntos del tiempo, de manera que puedan emplear varios estudios para tratar un tema y puedan ver el fenómeno desde varios ángulos. Los archivos del norte de Europa tienen gran experiencia en la preparación de datos para estudiantes de primaria y secundaria. Un buen análisis sobre estas perspectivas ha sido realizado por **Ryssevik** (1994). También **Frota** (2000) ha estudiado las ventajas de la utilización de archivos de datos en la enseñanza de estadística.

«En el futuro los estudiantes tendrán la posibilidad de usar múltiples conjuntos de datos, recogidos en diferentes momentos, de manera que podrán ver un fenómeno desde varios ángulos»

El avance tecnológico está facilitando que los usuarios puedan procesar información según sus necesidades. El servicio mínimo ofrecido es la consulta de catálogos, ya sea en el ámbito nacional (por ejemplo el *CIS*) o internacional (por ejemplo el catálogo integrado del *Cessda*). De esta forma no sólo pueden elegir investigaciones sino países, períodos, variables, casos. Con la creación de nuevas herramientas se pueden localizar fuentes de datos de diferentes países, visualizar metadatos detallados, procesar y visualizar datos en línea y si se desea, descargar conjuntos de datos en diversos formatos.

Los archivos norteamericanos se han llevado a cabo para sus estudiantes universitarios. Resta ahora continuar con las innovaciones tecnológicas y mejorar la utilización de lenguajes documentales para favorecer el acceso y penetración de los archivos en el resto del mundo.

Notas

1. *http://www.cis.es*

2. *http://www.nsd.uib.no/Cessda/index.html*

3. *http://www.ifdo.org*

4. *http://dasun3.essex.ac.uk/Cessda/IDC/*

5. *http://www.nesstar.com*

6. *http://TheData.org*

7. En el año 2000 el *SDA* obtuvo premios de la *American Association for Public Opinion Research* y de la *American Political Science Association*. *http://sda.berkeley.edu*

8. *http://odwin.ucsd.edu/idata/*

Referencias

Altman, Micah; **Andreev, Leonid**; **Diggory, Mark**; **King, Gary**; **Sone, Akio**; **Verba, Sidney**; **Kiskis, Daniel**; **Krot, Michael**. "A digital library for the dissemination and replication of quantitative social science research. The virtual data center". En: *Social science computer review,* 2001, Winter, v. 19, n. 4, pp. 458-470.

Anagnostelis, Betsy; **Cooke, Alison**. "Evaluation criteria for different versions of the same database. A comparison of Medline services available via the World Wide Web". *21st. Intl. online information meeting*, London, 9-11 December 1997. Consultado en: 21-03-02. *http://biome.ac.uk/sage/iolim97.html*

Burnhill, Peter; **Rice, Robin**; **Wright, Melanie**; **Townsend, Sean**; **Cole, Keith**. "An enquiry into the use of numeric data in learning & teaching". September, 2001. Consultado en: 6-02-02. *http://datalib.ed.ac.uk/projects/datateach.html*

Clark, Rich; **Maynard, Marc**. "Research methodology. Using online technology for secondary analysis of survey research data-'act globally, think locally'". En: *Social science computer review,* 1998, Spring, v. 16, n. 1, pp. 58-71.

Cordero, Magdalena. "Bancos de datos". En: *Cuadernos metodológicos*, 1998, mayo, n. 24, Centro de Investigaciones Sociológicas, Isbn 84-7476-256-1.

Frota, Álvaro. "Apresentação: aprenda estatística a través da pesquisa". Consultado en: 8-04-2000. *http://members.es.tripod.de/AlvaroFrota/MaterialDidatico.htm*

Guy, Laura; **Haley, Greg**. "Archive comparison chart. international social science and economic data archives". En: Strategies for navigating the internet to locate and access data from foreign sources. *Conference workshop on international data resources. Part II, Section III: International data from foreign sources*. 29 October, 1995. Washington, D. C. Consultado en: 22-02-99. *http://dpls.dacc.wisc.edu/apdu/*

Nasatir, David. "Archives de données de sciences sociales: objectifs, fonctionnement et problèmes". En: *Rapports et documents de sciences sociales,* 1973, n. 26, Unesco, Ibsn 92-3-201052-6.

Rockwell, Richard. "An integrated network interface between the research and social science data resources: in search of a practical vision". En: *Social science computer review,* 1994, Summer, v. 12, n. 2, pp. 202-214.

Rockwell, Richard; **Hardin, Joseph**; **Loots, Melanie**. "Surviving the three revolutions in social science computing". En: *Social science computer review,* 1995, Summer, v. 13, n. 2, pp. 149-162.

Ryssevik, Jostein. "Laboratorios de datos para la ciencia social comparada". En: *Revista internacional de ciencias sociales,* 1994, diciembre, n. 142, pp. 665-675.

Ryssevik, Jostein; **Musgrave, Simon**. "The social science dream machine. Resource discovery, analysis, and delivery on the Web". En: *Social science computer review,* 2001, Summer, v. 19, n. 19, pp. 163-174.

Scheuch, Erwin. "De los archivos de datos a una infraestructura para las ciencias sociales". En: *Revista internacional de ciencias sociales,* 1990, marzo, pp. 101-120. Citado por **Cordero** (1998).

Óscar Roba Stuart, *Facultad de Ciencias Sociales, Universidad de la República, Uruguay.*
ors@fcsbd.edu.uy