The Spanish digitization project DML-E and related topics

Enrique Macías Virgós
Institute of Mathematics,
Universidade de Santiago de Compostela, Spain
xtquique@usc.es

This chapter is an extended version of the talk I gave at the workshop *Communicating Mathematics in the Digital Era* (Aveiro, August 2006), as a part of the section on Electronic Publishing. I would like to thank the organizers, and especially Eugénio Rocha, José Francisco Rodrigues and Bernd Wegner.

The aim of my talk was to present the current situation and circumstances of mathematical research in Spain, some bibliometric indices, the (short) history of the Spanish digitization project DML-E, and a personal reflection about digitization matters. I hope that this report will help teams facing analogous problems in other countries. At the end of the report I provide a rather complete bibliography (mostly in Spanish) which I hope will be of interest to our colleagues and librarians in Latin America.

1 Mathematical research in Spain

1.1 Some figures

The situation of mathematical research in Spain can be summarized as follows. There are 350 research groups, mainly working at universities, which receive an average of 50,000 euros every three years. The mean size of groups is $5-6$ researchers. This investment is made available under the so called National R+D+i Program [20]; in many cases, there are supplementary funds from the governments of local autonomous regions or from universities.

Several recent reports all underline the remarkable evolution of mathematics in Spain [1, 4]. Mathematical research has rapidly increased in quantity as well as in quality and impact. As a matter of fact, the Spanish contribution to the worldwide published papers in mathematics (measured by papers in ISI journals having a Spanish author) rose from 0.3% in 1980 to the current 5.0%. As a reference, the overall Spanish percentage share for all scientific fields is 3.18%. Spain now occupies the 9th world position in terms of contribution to mathematics, which more or less corresponds to our economical status.

93–97	94–98	95–99	96–00	97–01	98–02	99–03	00–04
3.46	3.66	3.88	4.18	4.42	4.53	4.65	4.82
-17	-14	-15	-16	-13	-13	-6	-3

Table 1: Spanish % production in mathematics and relative impact factor in comparison with the world mean. Years 1993–2004. Source: Web of Knowledge Thomson-ISI. Taken from [4].

1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
1.7	2.0	2.2	2.2	2.1	2.3	2.4	2.7	2.8	3.2
1.7	1.7	2.1	2.6	2.5	2.9	3.0	3.3	3.6	3.9

Table 2: Percentage comparison Spain/World production in mathematics. Years 90–99. Source MathSciNet (first row) and ISI (second row). Taken from [1].

Accordingly to [1], in the period 1996–2001 there were 5,600 different Spanish authors of mathematical papers (60% of them published just one paper). The median number of authors was two; and 40% of papers were written in international collaboration (mainly with European countries and North America, and increasingly with Latin America).

On the other hand [4], for the period 1994–2004 the total number of papers (in ISI journals) was 6,220. Approximately 20% of them were in the first quartile of journals ranked by impact factor (IF, number of cites per document) in JCR Web of Science. The mean IF was slightly below the world mean (2.28 versus 2.51).

This rather impressive development has received international recognition, for instance with the organization of the 3rd. European Congress of Mathematics (Barcelona, 2000) and the choice of Spain by IMU as the venue of its 15th. General Assembly (Santiago de Compostela) and the International Congress of Mathematicians (ICM Madrid, 2006).

1.2 Consolider Mathematica [9]

However, several weaknesses have been detected in the Spanish mathematical research, including the lack of an adequate contact with industry; the small presence of computational infrastructures to support research (with some brilliant exceptions [7]); the insufficient long term planning; and a low leading role in many fields.

Hence, a new program from the Ministry of Education and Science has been approved [9] under the name of CONSOLIDER - INGENIO 2010. The mathematics branch of this program, called MATHEMATICA (not a very original name!), will offer the mathematical community a supplement of 7.5 million euros for the next 5 years (prorogueable for 5 more years).

The project aims to develop strategic measures in order to provide cohe-

sion to the activity of the Spanish mathematicians, to increase the weight of mathematics in the Spanish system of science, technology and industry, and to improve our situation in the international panorama.

MATHEMATICA is structured in five nodes [1] with a Board of Directors formed by fourteen leading Spanish mathematicians.

The main proposed measures include semesters devoted to specific subjects; internationally recognized doctorate courses; fellowships for post-docs and researchers (from Spain and abroad). The tools at disposal will be of three types:

- *Platforms*, named accordingly to their activities: FUTURE (for identifying emergent research areas of strategic interest); CONSULTING (for applicability of results in a non-academic context); COMPUTING (a network infrastructure of computational tools and software repositories); EDU (for Mathematical Education); and WEB (see below);

- *Thematic Actions*, such as an International Graduate School or several Programs of Intensive Research;

- *Cross-sectional Services*, e.g. to support meetings and encounters, with a recently inaugurated facility located in Cantabria.

1.3   The platform MATHEMATICA WEB

In order to increase the accessibility of the results of mathematical research, a specific platform will be created. It will present the activities and services of the project MATHEMATICA, as well as the corresponding links to services, research programs, thematic networks and platforms. This *virtual house* of MATHEMATICA will also give access to such tools of general interest as databases, specific software libraries, digitized publications, preprint servers, MathNet webpages or research journals. Also, divulgation, difussion and prospective will be taken into account for public awareness of the mathematical activity.

2   Access to scholarly literature

2.1   Open Access

Dissemination of knowledge is not complete if the information is not readily available. In March 2006, the European Commission issued an important study on the economic and technical evolution of the scholarly publication system and on the scientific publication markets in Europe. The EC recommendations include *to guarantee public access to publicly-funded research, shortly after publication* [27].

---

[1]Centre de Recerca Matemática (CRM Barcelona), Institute of Mathematical Sciences (ICMM Madrid), Institute of Mathematics (IMUB Barcelona), Centre of Supercomputation of Galicia (CESGA Santiago de Compostela) and Centre for International Mathematics Meetings (CIEM Cantabria).

Accordingly to [21] (see also [3]), Open Access in Spain is still an emerging movement. Currently, the 180 registered signatories of the Berlin Declaration include 18 Spanish universities or research institutions. ROAR (Registry of Open Access Repositories [25]) and OpenDOAR (Directory of Open Access Repositories [23]) have in their records 17 and 11 Spanish open access repositories, respectively, which represent less than 2% of the whole institutional repositories registered in those directories.

However, in recent years more and more initiatives related either to institutional open access repositories or open/free journals have been reported. There is an intense debate among librarians about the impact and visibility of scientific publications and the restrictions imposed by monopolistic publishers. Last summer, a course [29] about Open Access was organized by FECYT (*Fundación Española para la Ciencia y la Tecnología* [15]). Very recently, the symposium *Libraries and Digital Objects* [22] organized by CSIC joined many international experts.

Open archives serve to store papers, preprints, conference proceedings, reports and other research documents, and include metadata which are accessible through an interface like OAI-PMH (Protocol for Metadata Harvesting). Most current repositories started as repositories for doctorate theses, but some of them have evolved to institutional repositories (IR), which serve to collect the scientific production of an institution in such a way that it remains distributed in multiple archives but with a common interconnecting tool, ensuring long term preservation of academic work.

Open/free access journals were formerly printed journals that are currently free online [13]. In most cases they are published by public institutions or learned societies. Examples of open access journals directories or digital libraries are Scielo Spain (Scientific Electronic Library on-line[26]); Dialnet [12]; or Tecnociencia *e-revistas* (see below). There are other projects in several universities.

The Ministry of Culture has a directory of digitization projects and OAI harvesters [19] with basic information about the projects and initiatives of digitization existing in Spain.

2.2 Tecnociencia [28]

Tecnociencia is a website about Science and Technology in Spain, which depends on CINDOC-CSIC (see Section **??**). It was created after a long discussion about the need of good scientific journals in the Latin-American world, and it aims to provide unified access to electronic journals, under the auspices of FECYT. *Tecnociencia* includes news, research projects, and an OAI server called *e-revistas* offering e-journals.

2.3 ReviCien [24]

Currently, almost 30 Spanish scientific journals are included in the Thomson-ISI Science Journal Citation Index. Four of them are from mathematics.

Most of these journals share a common on-line platform for promotion and difussion, which is called ReviCien (*Revistas Científicas*, Scientific Journals). It offers links to 35 Spanish scientific journals from agricultural sciences, Earth sciences, environmental sciences, life sciences, chemistry, engineering and technology, materials science, computer science, mathematics (12 journals), medicine, neuroscience, pharmacology and toxicology, physics and astronomy.

While full texts of papers remain in the website of each journal, users of ReviCien have access to the abstracts and basic information like publisher, editor, ISSN, scope of the journal, cover, and table of contents of current or back issues, as well as quick/advanced search by author/title/keywords.

Subscribers receive e-alerts with tables of contents and announcements, and journalists have a dedicated Press Room. Finally, journal editors can access a restricted area inside the server.

3 Organizations involved in DML-E

3.1 The Spanish IMU Committee CEMAT

In order to vertebrate the Spanish mathematical community, a very important role is reserved to the Spanish Mathematics Committee (CEMAT, *Comité Español de Matemáticas* [5]), which coordinates the activities of international scope in Spain related to the International Mathematical Union (IMU). It also advises the Ministry of Education and Science of Spain about IMU recommendations on education and research in mathematics, and ensures coordination with ICSU (International Council for Science). CEMAT publishes a web page [5] and a regular newsletter

Spain entered IMU in 1952 (group II), passed to group III in 1986 and entered group IV in 2004. The current CEMAT was founded in 2004, as a continuation and extension of the former Spanish IMU Committee. Today, the seven main Spanish mathematical societies[2] are represented in CEMAT. There are also representatives of the Ministry of Education and Science.

CEMAT has an Executive Committee, a General Council and four Commissions, each being a counterpart of the analogous IMU Committee, that is:

- Commission on Development and Cooperation;

- Commission on Education;

- Commission on History;

- Commission on Electronic Information and Communication.

---

[2]RSME Royal Spanish Mathematical Society; SCM Catalan Society for Mathematics; SEMA Spanish Society for Applied Mathematics; SEIO Society for Statistics and Operations Research; FESPM Spanish Federation of Associations of Mathematics Teachers; SEIEM Spanish Society for Research in Mathematics Education; SEHCYT Spanish Society for History of Science and Technology.

3.2 The Spanish CEIC

The purpose of the Commission on Electronic Information and Communication [3] is to advise the Spanish Committee of Mathematics on all subjects related to electronic communication, and to represent Spain in the activities of the Committee on Electronic Information and Communication (CEIC) of the IMU.

The Commission maintains a web page with links to interesting issues such as digitized journals, literature related to electronic scientific edition, or the Spanish TeX users group (CervanTeX [6]).

3.3 CINDOC

The Spanish Center for Scientific Information and Documentation (*Centro de Información y Documentación Científica* [8]) aims to provide adequate documental support and scientific information to the Spanish users in all areas of knowledge, to collect the Spanish scientific output and to make it widely available.

CINDOC also undertakes research projects in the field of scientific documentation; carries out bibliometric studies; develops systems, methods, tools and techniques for information processing, storage, retrieval and dissemination of information; and organizes training courses in order to encourage the use of the information technologies.

CINDOC depends on CSIC (*Consejo Superior de Investigaciones Científicas*), the Spanish National Research Council [10]. Its personnel is formed by about 100 librarians and technicians.

3.4 A brief history [8]

The *Information and Documentation Center (CID)* was created in 1953 as a bibliographic enquiry service which began the publication of an Index to Scientific and Technical Journals, later replaced by the Abstracts of Scientific and Technical Papers.

In the 70s a new body was created, with the name of National Centre for Scientific Information and Documentation (CENIDOC). CENIDOC was conceived as a coordinating body, operating through three sectorial institutes: ICYT (Institute for Information and Documentation in Science and Technology), ISOC (Institute for Information and Documentation in Social Sciences and Humanities); and IBIM (Institute for Information and Documentation in Biomedicine).

In 1976, ISOC began to publish both the Spanish Index on Humanities and the Spanish Index on Social Sciences, while ICYT began the publication of the Spanish Index on Science and Technology in 1979. All three indexes covered the bibliographic references for all papers published in Spanish scientific journals.

---

[3]Its present composition is as follows: Jaume Amorós (Secretary, Universitat Politécnica de Barcelona), Manuel González Villa (Universidad Complutense de Madrid), Rafael de la Llave (University of Texas at Austin), Juan Luis Varona (Universidad de La Rioja) and Enrique Macías (Chairman, Universidade de Santiago de Compostela).

Since 1989, these databases are available on-line. In 1990, the CSIC databases were the first bibliographic information product ever released on CD-ROM in Spain.

Back in 1992, ICYT and ISOC merged to form the new Scientific Information and Documentation Centre (CINDOC).

4 The digitization project

DML-E is the Spanish counterpart of the WDML project of digitization of scholarly literature in mathematics supported by IMU. Since 2000 we have participated in several meetings sponsored by the EMS and were actively involved in the preparation of DML-EU proposals. Finally it was clear that the European Commission did not intend to cover digitization fees, but only coordination tasks, definiton of standards, and long-term archiving. Hence national funds were necessary to support the physical process of digitization, metadata capture, server access and linking to databases.

In 2005, the Ministry of Education and Science agreed to fund a first digitization project in mathematics by means of a so-called MEC Special Action 2005-2007. The objective is to retro-digitize all mathematical research journals published in Spain, starting from 1980 (sometimes 1940, depending on the scientific interest), which means about 100,000 pages. The scientific responsibility relies on CEMAT while the technical part is being developed by CINDOC (see section 3). There is a representative of the Editors of the involved journals.

4.1 Standards

Thirteen research journals with a recognized scientific quality in mathematics participate in DML-E. The project is digitizing about 50,000 pages/year and by mid-2007 we expect to have a dedicated web portal, with Open Access metadata and summaries, stable url addresses, and pointers to MathSciNet, Zentralblatt and catalogs from universities.

The standards follow WDML and CEIC recommendations. Resolution is 600 ppi, with software correction of text orientation and margins. Each page is archived in a tiff bitonal lossless compression file of the type journal/ year/ volume / number/ paper/ page.tif. A pdf file for each paper is composed from the TIFF files of its pages.

4.2 Metadata

CINDOC has developed a tool for capturing data and a web interface with a search engine. The structure of the metadata is as follows: title (original language, English, Spanish); author; journal; organization(s); keywords (English, Spanish); abstract (English, French, Spanish); UNESCO classification; publication year; collation; type of document (book, article, report); archive.

Most papers are written in English (see Table 4.2).

English	Spanish	French	Italian	German	Catalan
79.2	12.9	5.9	1.4	0.4	0.1

Table 3: Percentage of original languages in digitized papers in DML-E.

4.3 Journals

The following is a comprehensive list of mathematics research journals published in Spain. More information about these journals appears in my Stockholm report [18] and their web pages.

- *Applied General Topology*, Universidad Politécnica de Valencia;

- *Collectanea Mathematica*, Universitat de Barcelona UB;

- *Extracta Mathematicae*, Universidad de Extremadura;

- *Matematiques*, Universitat de València;

- *Mathware and soft computing*, Universidad de Granada and Universitat Politécnica de Catalunya UPC;

- *Publicacións Matemàtiques*, Universitat Autónoma de Barcelona UAB;

- *Qualitative theory of dynamical systems*, Universitat de Lleida;

- *RACSAM Revista de la Real Academia de Ciencias Exactas, Físicas y Naturales*, Serie A: Matemáticas;

- *Revista Estadística Española*, Instituto Nacional de Estadística INE;

- *Revista Matemática Complutense*, Universidad Complutense de Madrid. Old series: *Revista Matemática de la UCM*;

- *Revista Matemática Iberoamericana*, CSIC-RSME. Old series: *Revista Matemática Hispano-americana*;

- *SORT Statistics and Operations Research Transactions*, Institut d'Estadística de Catalunya. Old series: *Questiió: Quaderns d'Estadistica i Investigació Operativa*;

- *TEST*, Sociedad de Estadística e Investigación de Operaciones SEIO. Old series: *Trabajos de estadística*;

- *TOP*, Sociedad de Estadística e Investigación de Operaciones SEIO. Old series: *Trabajos de investigación operativa*;

- *Disertaciones Matemáticas*, Departamento de Matemáticas Fundamentales, Universidad Nacional de Educación a Distancia UNED.

I have not included journals from mathematical societies, as for instance *La Gaceta de la RSME* or *Butlletí de la SCM*, among others.

5   Previous reports

The following references are pertinent for understanding the evolution of the digitization iniciatives for mathematics in Spain. First, my report to the Royal Spanish Mathematical Society RSME [17] was largely inspired on John Ewing's report [14], but including my own experience when preparing the EMS applications for EU funds.

In the same issue of *La Gaceta* there is a very good paper from Rafael de la Llave [11] explaining several key ideas about scientific publication, universal access, the role of librarians, and copyright issues. Rafael de la Llave makes very clever comments on distribution and archiving of scientific information, editorial policies and evaluation of journals' quality.

Another interesting report is the CINDOC publication [2] about the state of the art of electronic publishing.

Finally, there is an excellent book by José Luis González Quirós and Karim Gherab [16] about the so-called Universal Digital Library. The authors (a philosopher and a theoretical physicist) give a very detailed account on the current status of electronic publishing and related topics, but mainly they explore with great insight the profound changes that digitized repositories will introduce in our experience as scientists and readers.

References

[1] Carlos Andradas and Enrique Zuazua, *Informe sobre la investigación matemática en España en el período 1990-1999*, Tech. report, CEAMM, 2002, `www.rsme.es/inicio/informem.pdf`.

[2] *Revistas científicas electrónicas: estado del arte*, Tech. report, CINDOC, 2004, `www.tecnociencia.es/e-revistas/especiales/revistas/pdf/e-revistas_informe.pdf`.

[3] J. M. Barrueco Cruz and I. Subirats Coll, *Open archives en las universidades españolas*, III REBIUN Workshop, Barcelona, 2003, `http://eprints.rclis.org/archive/00000384/`.

[4] María Bordons et al., *La investigación matemática española de difusión internacional: estudio bibliométrico (1996-2001)*, Tech. report, MEC-RSME, 2005, `www.ce-mat.org/documentos/informe_csic.pdf`.

[5] CEMAT: `www.ce-mat.org`.

[6] CervanTeX: `http://filemon.mecanica.upm.es/CervanTeX/`.

[7] CESGA: `www.cesga.es/content/view/740/1/lang,en/`.

[8] CINDOC: `www.cindoc.csic.es/eng/`.

[9] Consolider Mathematica: `www.mathematica.unican.es/`.

[10] CSIC: `www.csic.es/index.do?lengua=en`.

[11] Rafael de la Llave, *La publicación electrónica de la investigación matemática: Apuntes para un debate sobre su economía y sociología*, La Gaceta de la RSME **6** (2003), no. 2, 307–350.

[12] DIALNET: `http://dialnet.unirioja.es/`.

[13] DOAJ: `www.doaj.org/doaj?func=subject&cpid=57`.

[14] John Ewing, *Twenty centuries of mathematics : digitizing and disseminating the past mathematical literature*, Tech. report, 2002, `http://eprints.rclis.org/archive/00001381/`.

[15] FECYT: `www.fecyt.es/`.

[16] K. Gherab and J. L. González Quirós, *El templo del saber: hacia la biblioteca digital universal*, Ediciones Deusto/Planeta DeAgostini, 2006.

[17] E. Macias Virgós, *Un gran proyecto de cooperación internacional: la biblioteca digital de matemáticas*, La Gaceta de la RSME **6** (2003), no. 2, 351–366.

[18] E. Macias Virgós, *Some digitization initiatives in Spain*, New Developments in Electronic Publishing of Mathematics, Stockholm, June 25–27, 2004, FIZ Karlsruhe. Eggenstein-Leopoldshafen (Germany), 2005.

[19] MCU: `www.mcu.es/roai/en/inicio/inicio.cmd`.

[20] MEC. Plan Nacional de I+D+i: `www.mec.es/ciencia/jsp/plantilla.jsp?area=plan_idi&id=2`.

[21] R. Melero, *Open access environment in Spain: How the "movement" has evolved and current emerging initiatives*, Workshop Open Access and Information Management, Oslo, 2006, `http://eprints.rclis.org/archive/00006668/`.

[22] *Bibliotecas y Objetos Digitales*, `www.bibliotecasdigitales.es/`.

[23] OpenDOAR: `www.opendoar.org/`.

[24] REVICIEN: `www.revicien.net/`.

[25] ROAR: `http://archives.eprints.org/`.

[26] SCIELO: `www.scielo.org/index.php?lang=en`.

[27] *Study on the economic and technical evolution of the scientific publication markets in Europe*, Tech. report, European Commission. Directorate-General for Research, 2004, `http://ec.europa.eu/research/science-society/pdf/scientific-publication-study_en.pdf`.

[28] Tecnociencia: `www.tecnociencia.es`.

[29] Los archivos abiertos y los repositorios institucionales, Curso de verano. El Escorial, 3–7 julio, 2006, `www.ucm.es/BUCM/biblioteca/11807.php`.