

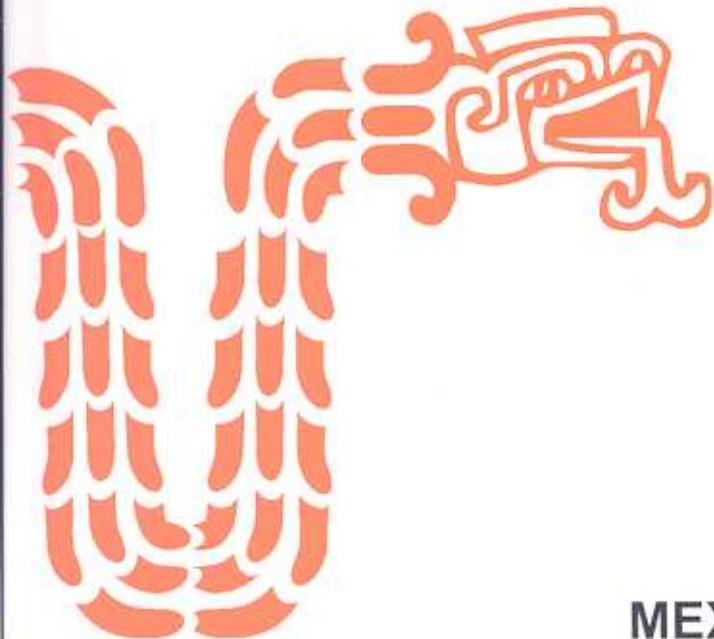


DC-2006

Proceedings of the International Conference on Dublin Core and Metadata Applications

Metadata for Knowledge and Learning

3-6 October, 2006 • Manzanillo, Colima, Mexico



MEXICO



UNIVERSIDAD
DE COLIMA

Cátedra UNESCO en Tecnologías de Información



Dublin Core Metadata Initiative

Universidad de Colima
Publicaciones



All rights reserved. No part of this book may be reproduced in any form without the written permission of the publisher.

The texts of the papers published in this volume are as submitted by the authors. No editorial intervention was undertaken by the Editor unless absolutely necessary and only as the production schedule allowed. The Editor task was to solicit contributions and to assist in the editorial matters.

Edition ended on October 2006 at University of Colima, México. 300 books were printed plus extra copies for replacement.

Printed and made in Mexico

© Copyright 2006
Metadata for Knowledge and Learning

Universidad de Colima
Avenida Universidad 333
Colima, México, 28040

Derechos reservados. No se permite la reproducción total o parcial de este material sin la autorización previa y por escrito del editor.

Los textos publicados en este volumen son responsabilidad de los autores.

Se terminó de imprimir en octubre de 2006 en la Dirección General de Publicaciones de la Universidad de Colima, Colima, México. Se imprimieron 300 ejemplares más sobrantes para reposición.

Impreso y hecho en México
ISBN 970-692-268-7

Editora responsable / *Editor:*
Myriam Cruz Calvario, Dirección General de Publicaciones de la Universidad de Colima.

Diseño / *Design:*
Carmen Millán, Dirección General de Publicaciones / Miguel Ávila, CENEDIC

Contents

Session 1. Metadata models	11
Towards an interoperability framework for metadata standards	13
<i>Mikael Nilsson, Pete Johnston, Ambjörn Naeve and Andy Powell</i>	
DC-Text: a simple text-based format for DC metadata	24
<i>Pete Johnston and Andy Powell</i>	
Toward a collection-based metadata maintenance model	31
<i>Martin Kurbis and Jim LeBlanc</i>	
Using the DC abstract model to support application profile developers	42
<i>Sarah Purvis and Liddy Nevile</i>	
Session 2. Ontologies and controlled vocabularies	53
SKOS: requirements for standardization	55
<i>Alistair Miles</i>	
A metadata registry from vocabularies up: the NSDL registry project	65
<i>Diane L. Hillmann, Stuart A. Sutton, Jon Phipps and Ryan Laundry</i>	
From AGROVOC to the Agricultural Ontology Service/concept server.	
An OWL model for creating ontologies in the agricultural domain	76
<i>Boris Lauer and Maryberita Sutti</i>	
Encoding Library of Congress subject headings in SKOS: authority control for the semantic web	89
<i>Cory All Harper</i>	
Session 3. Implementations	95
Implementing an institutional repository for digital archive communities: experiences from National Taiwan University	97
<i>Chiung-min Tsai, Jieh Hsiang and I-Hsueh-kua Chen</i>	
Dublin Core as a tool for interoperability common presentation of data from archives, libraries and museums	109
<i>Leif Andreassen</i>	
Eureka! Dublin Core based metadata supports the archimedes palimpsest manuscript imaging program	120
<i>Michael B. Toth, William A. Christens-Barry and Roger L. Easton, Jr.</i>	
Session 4. Metadata for education	131
Metadata towards an e-research cyberinfrastructure: The case of French PhD theses	133
<i>Jacques Ducloy, Jean-Paul Ducaze, Muriel Foulonneau, Luc Grivell, Diane Le Henaff, Yann Nicolas</i>	

Vocabularies for describing pedagogical approach in e-learning: a scoping study <i>Sarah Currier, Sheila MacNeill, Lisa Corley, Lorna Campbell, Helen Beetham</i>	149
Virtual reality learning objects of molecular structures <i>Miguel A. García-Ruiz, Arthur Edwards, Jorge R. Gutiérrez-Pulido, Ricardo Acosta-Díaz</i>	162
Memex metadata (M2) for reflective learning <i>Jane Greenberg, Abe Crystal, Anuj Sharma, Eva Méndez, John Oberlin and Michael Shoffner</i>	169
Session 5. Adapting and sharing	181
Using FRBR for the selection and adaptation of accessible resources <i>Ayako Morozumi, Mitsuaki Nagamori, Liddy Nevile and Shigeo Sugimoto</i>	183
Semantic mediacasting and collaborative feed sharing <i>Narendra Kumar Shukla and Anand Agrawal</i>	195
Development of a wiki-based, expert community-driven nanosystem vocabulary <i>Laura M. Bartolo, Cathy S. Lowe, Sharon C. Glatzer and Christopher Iacovella</i>	211
Identity matters: constructing social identities through ontology-based metadata <i>Maria Cristina Pattielli and Lisa R. Norberg</i>	215
Session 6: Linking and citation	221
DCMI abstract model analysis: resource model <i>Vicente Palacios, Jorge Morato, Juan B. Llorens and Jose A. Moreira</i>	223
Developing cultural heritage preservation databases based on Dublin Core data elements <i>Fenella G. France and Michael B. Coth</i>	233
Redalyc OAI-PMH the open archives initiative protocol for metadata harvesting (protocol version 2.0) <i>Eduardo Aguado López, Rosario Rogel Salazar, Arianna Becerril García and Honorio García Flores</i>	244
Development of a controlled vocabulary for learning objects' functional description in an educational repository <i>Miguel Ángel Marzá, Javier Calzada-Prado, M. Jesús Colmenero Ruiz and Aurora Cuevas Cerveri</i>	253
Session 7: Resources in context	259
Customising location of knowledge <i>Ann Apps and Ross MacIntyre</i>	261
Linking legal sources in a shared web environment <i>E. Francesconi and G. Peruginelli</i>	273
The Semantic Network Service-supporting heterogeneous environmental information Systems <i>Matthias Menger, and Maria Rüther</i>	286
Signed metadata: method and application <i>Emma Tonkin</i>	292
Authors' index	299

Redalyc OAI - PMH

The open archives initiative protocol for metadata harvesting (protocol version 2.0)

Eduardo Aguado López,¹ Rosario Rogel Salazar,²
Arianna Becerril García³ and Honorio García Flores⁴

Abstract

With the arrival of the new communication systems and the *disappearance* of the borderlines, new challenges emerged to scientific production and spreading. We recognized that great part of that production was unknown and its area of influence is limited. We stated what we already knew, that social sciences were under-represented in databases that determined *the great current of science* and that Latin American scientific journals would hardly enter the legitimized databases. The adoption of the Dublin Core format in Redalyc - that administers a wide database of arbitrated scientific articles of and on Ibero-America it has required to analyze the structure of metadata, in order to implement the protocol OAI-PMH, to fortify the search, recovery and indexing the database through harvesters and search engines over the Internet by making all the articles of its collection available, facing the challenge to give visibility to the Ibero-American scientific production.

Keywords

Dublin Core, OAI-PMH protocol, Redalyc, Ibero-American journals, full text articles, metadata format, scientific journals, arbitrated articles, OAI data provider, OAI software implementation.

1. Antecedents

Since its creation in October 2002, the *Red de Revistas Científicas de América Latina y El Caribe, España y Portugal (Redalyc)*, set a goal: to give visibility to the scientific production generated in Ibero-America, that is underestimated in the world-wide scene due to different factors like low investment in science and technology, the low participation of Latin American

¹ Universidad Autónoma del Estado de México, Director General de Redalyc.

Tel: +52 (722) 2 15 04 94. eal@uaemex.mx

² Universidad Autónoma del Estado de México, Directora Editorial de Redalyc.

Tel: +52 (722) 2 15 04 94. rrs@uaemex.mx

³ Universidad Autónoma del Estado de México, Directora de Sistemas de Redalyc.

Tel: +52 (722) 2 15 04 94. aribg@uaemex.mx

⁴ Universidad Autónoma del Estado de México, Coordinador de la Hemeroteca Redalyc.

Tel: +52 (722) 2 15 04 94. hgf@políticas.uaemex.mx

scientists in the *main current of science*, measured by the percentage of articles signed by Latin American authors in main databases and the low impact of that production.

The main communication vehicle of academic and scientific means is the publication in scientific journals. In this sense, the Latin American participation of authors or institutions in principal journals or the impact of journals produced in the region allows to know the effect of Latin American science in the world. According to RICYT (2002), the participation of the Latin American scientists in the *main current of science*, measured by the percentage of articles signed by authors of Latin America in the main databases that register scientific publications was practically null, less than 3% in the important repositories, although there was a growing, for example, 2.7% in the Science Citation Index (SCI).

The under-representation of the Latin American production is shown in the composition of the Journal Citation Reports (JCR) of ISI - Thomson Scientific, about 9.000 journals and more than 200 disciplines based on four journal parameters (higher impact, greater frequency of use, greater demand and greater) where social sciences only participate with 1712 journals of that universe, the under-representation is evident. On the other hand, the low weight of the Latin American publications is exhibited when we identified that only 12 Latin American journals –of the 1712– have been able to be indexed in the JCR. If we consider that it is the main base to measure the impact, the conclusion is clear: the Latin American production is not present.

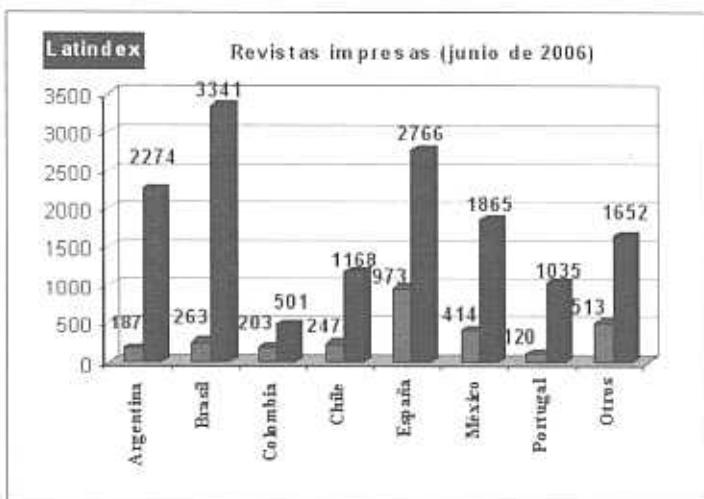
Recent researches show that the cost efficiency in investment and development depends, in a big way, on the possibility that the results are transmitted and consumed by the academic actors in general. The breach between *releases* (published and cited articles) is greater than *entries* (investment and development costs). This polarization demands to modify the levels of *releases* (Dickson, 2004) and sets the need to participate actively in the construction-validation of the science of those countries that are not included in the *great current of science*. The enlargement of the breach shows that increasing the relative weight of the investment in science and development is not enough, but it is equal or more important, the reached effectiveness of the communication of the scientific production.

The low efficiency in the communication of the scientific production has taken some specialists to affirm that the major part of the scientific information of the third world countries stays in the penumbra (Garfield, 1999)¹ situation that is turning worst, when it is identified that a researcher coming from the countries in development, is publishing in high international prestige journals, his impact factor is smaller than one of a researcher coming from the countries that control the scientific production (Gibbs, 2001). Therefore, the weakness, the lack of presence and international visibility, the ignorance and the inadequate valuation that journals are experiencing, produced in the Latin American region a central worry of the scientific production's actors (Cetto, 1998a, b).

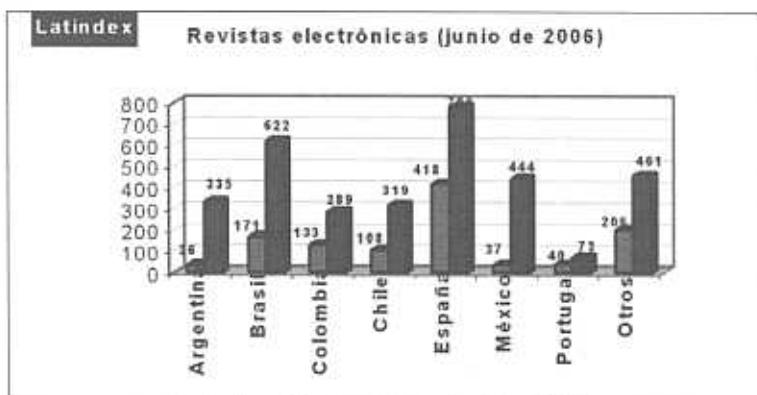
¹ Available at: www.isinet.com/isi/

An example that evidences the former premise, can be observed in one of the most important aggregators of content: Oaister, in which, is shown clearly a minimal presence of data providers with Ibero-American collection. There are within the portal, providers like Scielo enriching the collection with more than 50,000 records to the date on its public health database, besides that Dialnet contributes with about 60,000 records, mostly in a referential level. This shows that it doesn't exist a database that contributes with a significantly number of OAI records with link to full text, and focus on the Ibero-American scientific production.

United to this, the lack of technological tools has originated that few journals can offer their contents in an electronic format, even when the efforts have been made to revert this situation; there are still few periodic publications that achieve the access to the electronic format. From the 21,994 registered journals in Latindex just 4,472 belong to electronic journals.

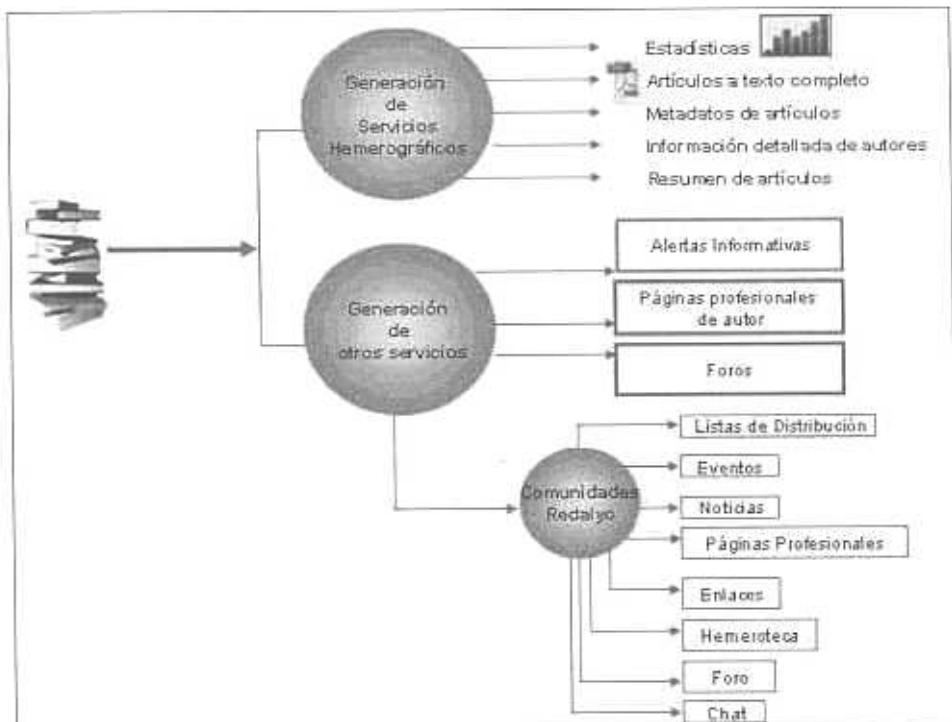


Latindex. Printed journals (June 2006).



Latindex. Electronic journals (June 2006).

Taking into consideration that situation Redalyc tries to integrate a greater number of journals than the big databases, keeping a standardized metadata format. This issue proposes a great challenge, to give visibility to the scientific production of the region; as a result Redalyc contributes designing a web site that includes library services of interaction for the publishers, editors, authors and users in general.



Redalyc's architecture.

The different search engines and information interchange protocols search and harvest over metadata to find the documents of their interest, for the metadata description Redalyc uses a code of its own, which includes all the required fields in the description of journals and articles.

Even when the necessary elements are contemplated, at the level of search engines, automation systems, harvesters and information interchange protocols like the Open Archive Initiative Protocol for Metadata Harvesting (OAI-PMH) and the Z39.50, it's necessary the use of standard formats, such is the case of the OAI, that requires necessarily the metadata format Dublin Core, to be able to use them.

Dublin Core is the most divulged and cited meta-information format, at least in the strictly librarian scope. It was designed to promote a general purpose standard, simple and descriptive of the electronic resources of any subject (Manso-Rodríguez, 2002).

Nowadays, the adequacy to interoperability standards is a fundamental requirement for the scientific production spreading that made necessary the adoption of the Dublin Core format, with the intention of being harvested in the OAI official website, besides the distinct search engines like Google Scholar and OA-HERMES and databases, among these we can find mainly the OCLC's WorldCat.

To carry on with this process, it was necessary to analyze the Redalyc's database structure, which in essence works under a relational model. Once identified these elements the next step for the metadata migration process, consisted in the determination of the format to use, choosing to implement the Simple Dublin Core format.

Establishing the requirements and the coincidences by part of Redalyc, to get incorporated to the open archive initiative, the next step was the transformation of the metadata to the Dublin Core, this process implied the programming of a tool that allows to send SQL requests to a relational database, homologate the retrieved records to the Dublin Core format and generate the XML output.

2. Incorporation of the OAI protocol

The essence of the open access initiative consists in making available to any user, the existing materials on the Web through interoperable repositories to share metadata. This allows that such material can be consulted widely, supporting the basic Redalyc objectives referent to increase the visibility of the scientific production in the Ibero-America countries and collaborating to decrease the notorious disadvantage that exists world-wide in the Internet presence of the scientific collection generated in the region.

The immersion in the open access world, give us the opportunity to participate in the data provider community, contributing to the repositories conformation that will serve as a basis to operate different services to satisfy in a better way, the growing user information demands.

Therefore, the Redalyc OAI - PMH protocol implementation will enable the international access to the library newspaper showing the institutional research outputs and will improve the citation impact in what we call lost science.

Keeping the Redalyc presence in the OAI ambit, important aspect, like it's mentioned in different studies (Brody, 2004; Harnad and Brody, 2004) the number of oa-articles citations increases between a 300% and 500% compare to non oa-articles, we have the expectation to have a definitive raise to the consulted content of at least the triple (read articles-citation correlation). In this way Redalyc not only will become an OAI official data provider (setting the UAEM as one of the three institutions having official participation in the OAI-PMH join to UDLA and the ITESO) but also could compete in the content aggregators positioning like Laoap, OA-Hermes or Oaister, contributing Ibero-American journals to be more known and consulted through the Web and proportionally more cited.

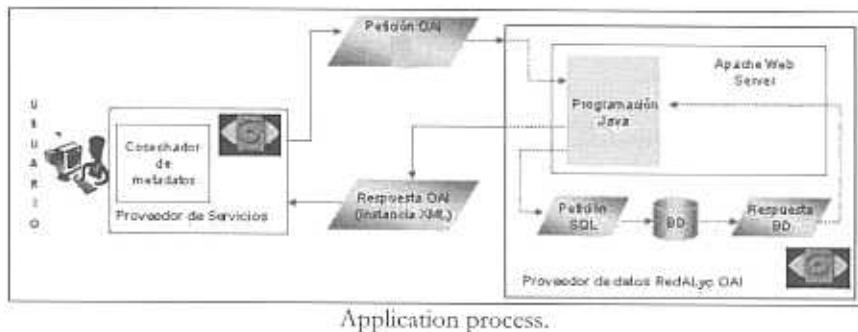
About the protocol implementation it was accomplish by means of the Web development supported by Voai software elaborated by the Universidad de Las Américas, Puebla, that helps to generate an OAI metadata server using the Java programming language and Oracle database to produce XML instances. These servers are made to be harvested by other application and the own performance depends on the Web Server, the application code and the operation of the database.

Nevertheless a harvest process is heavy-weighted and server resources demanding. Such performance oscillates between 20 and 200 records by minute of download (Operation of official OAI servers) depending in the amount of metadata that is extracted among other factors.

3. Redalyc OAI-PMH implementation and software architecture

This software makes a data processing that begins with an OAI-PMH request sent to the server through a metadata harvester and finishes with an XML output instance.

The OAI-PMH gives a simple technical option for Redalyc to make its metadata available to services, based on the open standards Hypertext Transport Protocol (HTTP) and Extensible Markup Language (XML). The metadata, that is harvested, might be in any format that is agreed by a community, although unqualified Dublin Core is specified to provide a basic level of interoperability.



4. Protocol implementation details

- *Implemented Tool.* VOAI, software developed by Udlap, Universidad de las Américas, Puebla.
- *Programming Interface.* Java-Servlet 2.3
- *Record.* Redalyc has nowadays more than 27,000 records ready to be available through OAI-PMH. A record has three parts, a header and metadata, both of which are mandatory, and an optional about statement. In Redalyc we cover the following:
 - *Header.* Identifier: composed of a prefix followed by a Redalyc's id
 - timestamp: the day since it was available on the web
 - setSpec: the Redalyc's id of the logical partition of the repository.
 - *Metadata.* Redalyc covers the following DC elements.

- <dc:title>
- <dc:creator>
- <dc:subject>
- <dc:contributor>
- <dc:publisher>
- <dc:date>
- <dc:type>
- <dc:format>
- <dc:identifier>
- <dc:relations>
- <dc:rights>
- *Sets.* Sets enable a logical partitioning of repositories; the case of Redalyc includes 44 knowledge areas.

```

<?xml version="1.0" encoding="UTF-8"?>
<OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd"
  responseDate="2006-07-25T05:41:45Z"><responseInfo>
<request verb="ListSets">http://redalyc.uam.mx/me/redalyc/servlet/Oai_handler</request>
</responseInfo>
<listSets>
  <setSpec>1</setSpec>
  <setName>Comunicación</setName>
</set>
<set>
  <setSpec>2</setSpec>
  <setName>Cultura</setName>
</set>
<set>
  <setSpec>3</setSpec>
  <setName>Demografía</setName>
</set>
<set>
  <setSpec>4</setSpec>
  <setName>Derecho</setName>
</set>
<set>
  <setSpec>5</setSpec>
  <setName>Administración Pública</setName>
</set>
<set>
  <setSpec>6</setSpec>
  <setName>Antropología</setName>
</set>
<set>
  <setSpec>7</setSpec>
  <setName>Divulgación Científica</setName>
</set>
<set>
  <setSpec>8</setSpec>
  <setName>Economía</setName>
</set>
<set>
  <setSpec>9</setSpec>
  <setName>Educación</setName>
</set>
<set>
  <setSpec>10</setSpec>
  <setName>Estudios Territoriales</setName>
</set>

```

Example: Redaly's ListSets.

- *Request format.* Requests are submitted using GET/POST methods of HTTP.
- *Responses.* Responses are formatted as HTTP responses. The response format is well-formed XML with markup as follows:
 - XML declaration
(<?xml version="1.0" encoding="UTF-8"?>)
 - root element named OAI-PMH with three attributes
(xmlns, xmlns:xsi, xsi:schemaLocation)

- three child elements
 - o responseDate
 - o request
 - o error (in case of error exception) or element with the name of the OAI-PMH request.

```
<?xml version="1.0" encoding="UTF-8"?>
<oai:PMH xmlns="http://www.openarchives.org/OAI/2.0/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/ http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2006-04-19T10:19:53Z</responseDate>
  <request verb="GetRecord" identifier="oai:redalyc.uaemex.mx:10201906"
    metadataPrefix="oai_dc">http://redalyc.uaemex.mx/redalyc/servlet/Oai_handler</request>
  <GetRecords>
    <record>
      <headers>
        <identifier>oai:redalyc.uaemex.mx:10201906</identifier>
        <datestamp>2004-10-15</datestamp>
        <setSpec>10</setSpec>
      </headers>
      <metadata>
        <oai_dc:dc xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
          xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
          xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/
            http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
          <dc:title>Reseña de "Breve Historia de Sinaloa" de Sergio Ortega Noriega</dc:title>
          <dc:creator>Hiram Felix Rosas</dc:creator>
          <dc:subject>Estudios Territoriales</dc:subject>
          <dc:contributor>Hiram Felix Rosas</dc:contributor>
          <dc:publisher>Colegio de Sonora</dc:publisher>
          <dc:date>2004-10-15</dc:date>
          <dc:type>articulo_cientifico</dc:type>
          <dc:format>application/pdf</dc:format>
          <dc:identifier>http://redalyc.uaemex.mx/redalyc/src/inicio/ArtPdfRed.jsp?icve=10201906</dc:identifier>
          <dc:relation>Región y Sociedad</dc:relation>
          <dc:rights>Región y Sociedad</dc:rights>
        </oai_dc:dc>
      </metadata>
    </record>
  </GetRecords>
</oai:PMH>
```

Example: Redalyc's response.

5. Conclusions

With the migration of our metadata to the Dublin Core format, we had the possibility to incorporate Redalyc as a data provider using protocol OAI-PMH, and to increase the visibility, access and diffusion of the scientific production of Ibero-America, besides we are available of being located and indexed of automatic way by different search engines and data bases, contributing to the Redalyc's objective to make scientific resources accessible through the Web.

The internationalization of knowledge is an urgent task of the scientific communities: Redalyc is making its part to reach this goal.

Belong to different communities whose objectives are to offer free access to the electronic resources, will allow to set Redalyc as a safe and reliable option and with effective technological developments to spread and to make accessible the information contained to a greater quantity of users contributing, with it, to give greater presence to the scientific production generated in our region.

6. References

- Brody, T. (2004). The Effect of Open Access on Citation Impact. In: National Policies on Open Access (OA) Provision University Research Output: an International Meeting. Southampton.
- Cetto, A.M. (1998a). Ciencia y producción científica en América Latina. El proyecto Latin-dex. *Internat. Microbiol.*, Springer-Verlag, Ibérica, 1:181-182.
- Cetto, A.M. (1998b). Las revistas científicas como fuentes de bases de datos. Experiencias del Taller de Guadalajara. In: *Taller de obtención de indicadores bibliométricos*, RICYT-Cindoc, Madrid, 23-25 de febrero.
- Dickson, D. (2004). Scientific output: the real *knowledge divide*. Ed. SciDerNet, 19 de julio de 2004.
- Gibbs, W.W. (2001). Ciencia del tercer mundo. In: Eduardo Loria Díaz (ed.). *Viejos y nuevos dilemas de las revistas académicas*, UAEM, México, 101-115.
- Harnad, S. and Brody, T. (2004). Comparing the Impact of Open Access (OA) vs. Non-OA articles in the Same Journals. *DLib Magazine* 10 (6).
- Manso-Rodríguez, R.A. (2002). Aplicación del Formato Dublin Core para la descripción de los recursos en la Biblioteca Virtual del CDICT-UCLV. In: *Revista Latina de Comunicación Social*, 51, junio-septiembre, 2002. La Laguna (Tenerife), Available at: (URL):<http://www.ull.es/publicaciones/latina/2002mansojunio5104.htm>
- RICYT (2002). El estado de la ciencia. Principales indicadores de ciencia y tecnología iberoamericanos/interamericanos 2001. Red Iberoamericana de Indicadores de Ciencia y Tecnología (RICYT), Buenos Aires.

7. Other useful references

- Budapest Open Access Initiative (BOAI). www.soros.org/openaccess/
- Cetto, A.M. (2001). El impacto de las revistas y cómo incrementarlo. In: *Seminario Conacyt-UNAM para editores de revistas académicas*, 3-4 de octubre, México.
- Dublin Core Metadata Initiative. Retrieved April 19, 2006, available at:
<<http://es.dublincore.org/index.shtml>>
<http://ict.udlap.mx/software/voai_dist/index.htm>
<<http://www.oaforum.org/tutorial/>> <<http://www.openarchives.org/>>
- II Taller Sobre Publicaciones Científicas en América Latina, conclusiones y recomendaciones (1997). Documento de trabajo. Universidad de Guadalajara, México, 27-29 de noviembre.
- ISI (2004). www.isinet.com/isi/. Retrieved July, 2004.
- King, D. (2004). The scientific impact of nations. *Nature*, 430. Retrieved July 15, 2004, available at: www.nature.com/nature.
- Metainformación - Dublin Core (2006). Elementos del conjunto de metadatos de Dublin Core: Descripción de Referencia. Retrieved April 19, 2006, available at:
http://www.rediris.es/metadata/dublin_core_elements.es.html