



# **SciELO artículos disponibles para la comunidad agraria utilizando OAI-PMH y formato AGRIS AP XML**

**Presentadora: Stefka Kaloyanova,([stefka.kaloyanova@fao.org](mailto:stefka.kaloyanova@fao.org)),  
División de Intercambio de conocimiento y creación de capacidades,  
KCEW, FAO de la ONU**

# La cooperación entre las iniciativas de las redes SciELO y AGRIS

Dado que las áreas de agricultura y salud están conectadas muy directamente, la cooperación entre BIREME y FAO incluye:

- intercambio de metadatos, estándares, herramientas y metodologías;
- identificación de los participantes comunes;
- creación de infraestructuras y capacidades;
- acceso abierto a publicaciones, etc.

# Acceso Abierto

Acceso Abierto (OA) se define como: permanente, libre disponibilidad y acceso en línea (Internet) a la literatura científica de investigación en texto completo.

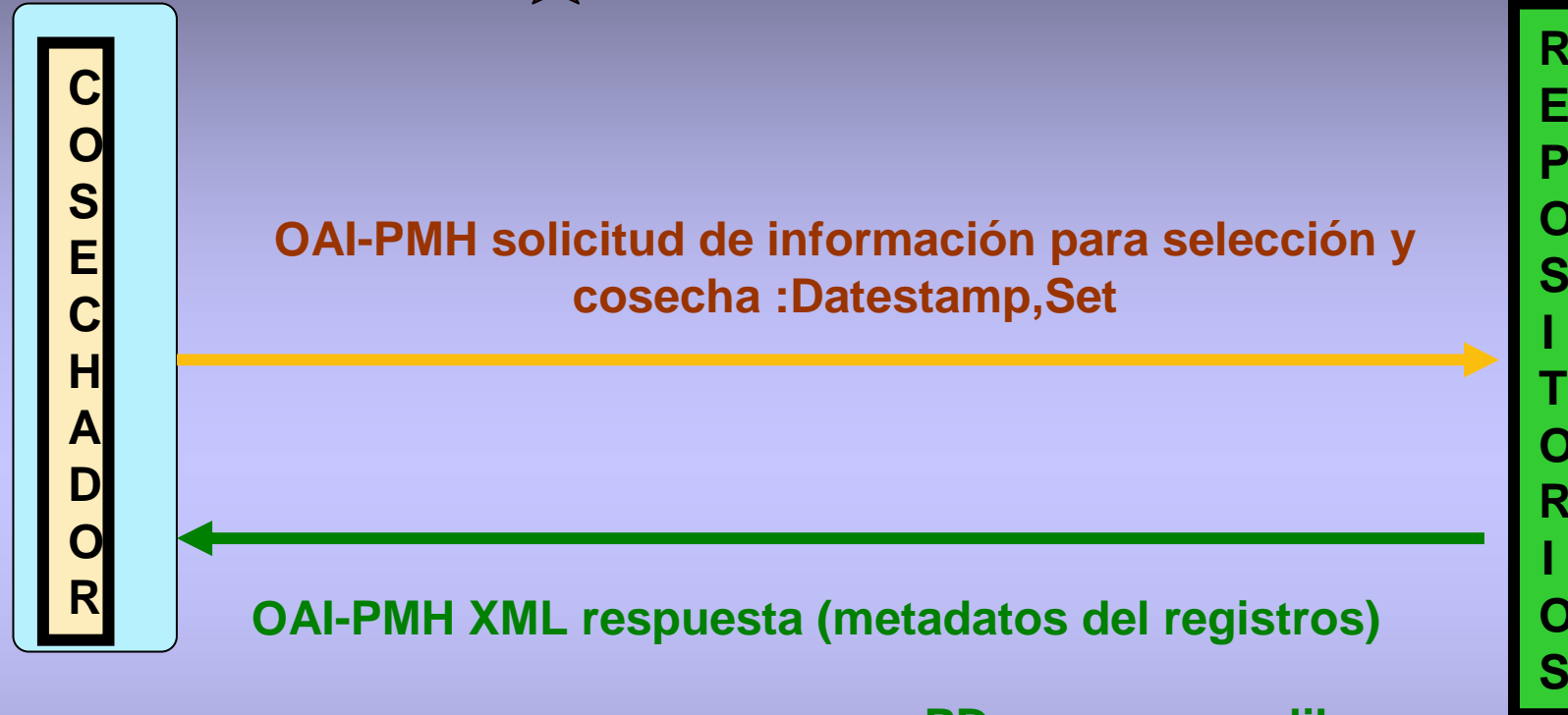
# Marco OAI (Iniciativa de Acceso Abierto): participantes

- Proveedor de datos (repositorios)
  - Provee libre acceso a metadatos, y puede, pero no es necesario, ofrecer libre acceso al texto completo
- Proveedor de servicios
  - Usa la interfaz OAI de los proveedores de datos para recolectar y almacenar metadatos;
  - Puede seleccionar determinadas colecciones;
  - Usa los metadatos de los proveedores de datos como base para construir servicios de valor agregado, y puede enriquecer los datos.

# Participantes en el Marco OAI

Proveedor de servicios ★

Proveedor de datos ★



**PS** – opera el cosechador para recolectar y almacenar metadatos, pueden seleccionar determinadas colecciones. Construyen servicios de valor añadido sobre metadatos recogidos

**PD** – proveen libre acceso en línea desde el cosechador al repositorio y cosecha de metadatos y texto completo del recurso, siguiendo las normas de protocolo OAI-PMH

La calidad del servicio es proporcional a la calidad de los datos recolectados.

# Prerequisitos para OAI-PMH

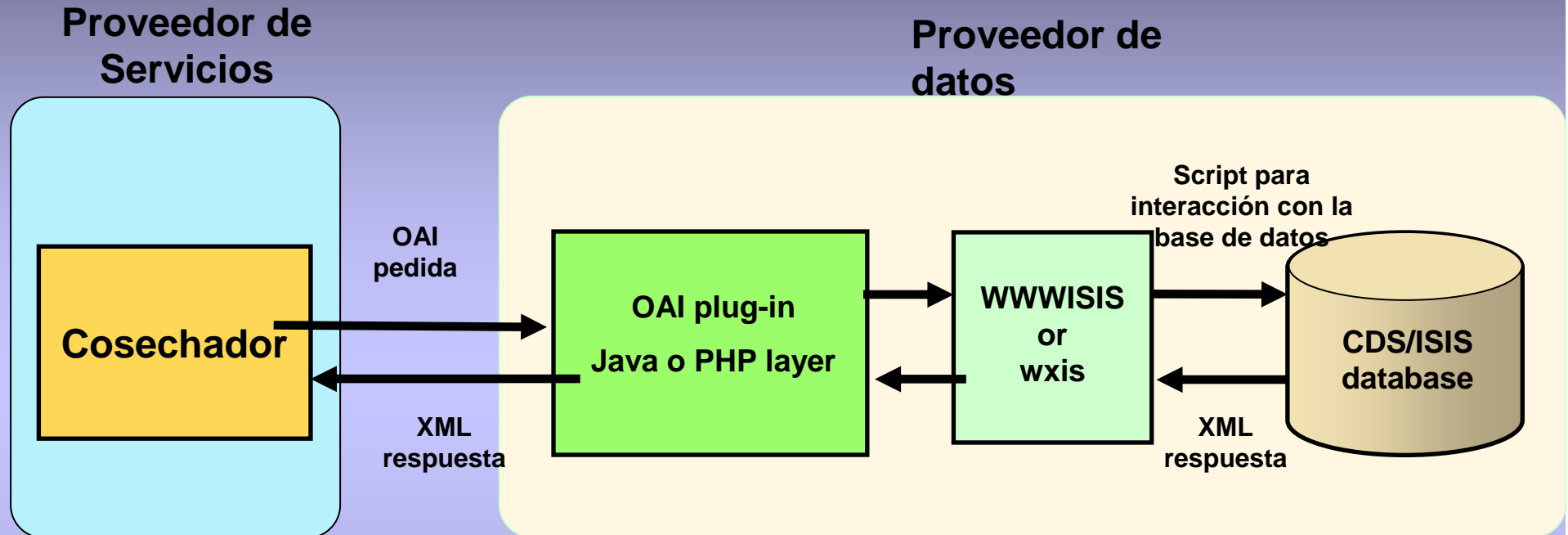
- Metadatos sobre documentos almacenados en una base de datos.
- Servidor Web, accesible vía Internet,
- API, interfaz (ej. Perl, PHP, Java-Servlet que reciba los pedidos OAI-PMH, consulte la base de datos y devuelve la respuesta).
- Base URL
- Un único identificador para cada documento
- Formato de metadatos (uno o más, como mínimo: DC no calificado)
- *Datestamps* para los metadatos (identifica creación/última modificación)
- *Sets* (aunque no es obligatorio)

## CDS/ISIS base de datos - Proveedor de datos

Combinando la experiencia de CDS/ISIS con el nuevo paradigma de Archivos Abiertos y la Iniciativa de Archivos Abiertos tenemos dos opciones:

- Incorporar una interfaz (capa) OAI para poner a disposición de la comunidad Internet los datos que hasta ahora estaban escondidos en bases de datos/ catálogos o
- cambiar la herramienta

# CDS/ISIS e OAI-PMH



**Solicitud:**

[http://www4.fao.org:8080/oaiagris/OAIHandler?verb=GetRecord&metadataPrefix=oai\\_dc&identifier=oai%3Aagris.uruguay%3AUY2006005761](http://www4.fao.org:8080/oaiagris/OAIHandler?verb=GetRecord&metadataPrefix=oai_dc&identifier=oai%3Aagris.uruguay%3AUY2006005761)

**Script:** [http://www4.fao.org/cgi-](http://www4.fao.org/cgi-bin/oaiagris.exe?database=agris&search_type=query&query=ID=UY2006005761&table=mont&lang=oai&format_name=oaidc)

[bin/oaiagris.exe?database=agris&search\\_type=query&query=ID=UY2006005761&table=mont&lang=oai&format\\_name=oaidc](http://www4.fao.org/cgi-bin/oaiagris.exe?database=agris&search_type=query&query=ID=UY2006005761&table=mont&lang=oai&format_name=oaidc)



# Logros de trabajo realizado:

- Recuperación de metadatos semánticamente ricos para las ciencias agrarias basada en “Perfil de Aplicación AGRIS” (AGRIS-AP) desde el repositorio de SciELO, usando el protocolo OAI-PMH.
- Selección de metadatos, creados por la aplicación ISIS para la cosecha y búsqueda con los datos de AGRIS.
- Utilizando las dos metodologías SciELO y AGRIS juntas para mayor visibilidad de las publicaciones científicas agrícolas y de la salud.

# El trabajo comprende:

1. Selección de 43 títulos de revistas con temas sobre agricultura desde el sitio Web de SciELO.
2. Definición de la metodología de la cosecha
3. Desarrollo de OAI-PMH plug-in para la cosecha: adaptación del programa OAI-PMH plug-in de BIREME para exponer los metadatos en formato AGRIS AP en adición de DC
4. Cosecha de 15000 artículos de las revistas seleccionadas
5. Transformación XSLT
6. Validación y verificación de datos
7. Registro e inclusión en la base de datos AGRIS
8. Implementación del cosechador para cosecha automática de los datos nuevos

# 1. Selección de las revistas

Selección de las revistas pertinentes a agricultura y biología que se encuentran publicadas desde una lista alfabética en el sitio Web de SciELO

[http://www.scielo.br/scielo.php?script=sci\\_alphabetic&lng=en&nrm=iso](http://www.scielo.br/scielo.php?script=sci_alphabetic&lng=en&nrm=iso)

De 211 fueron seleccionados 43 títulos de revistas.

# Selección

	Exists in FAO catalogue	ISSN of the journal
<b><u>Agricultural Science</u></b>		
1. Agricultura Técnica	FAO	0365-2807
2. Anais da Academia Brasileira de Ciências		0001-3765
3. Anais da Sociedade Entomológica do Brasil		0301-8059
4. Arquivo Brasileiro de Medicina Veterinária e Zootecnia		0102-0935
5. Bioagro		1316-3361
6. Bosque (Valdivia)		0717-9200
7. Bragantia	FAO	0006-8705
8. Brazilian Archives of Biology and Technology		1516-8913
9. Brazilian Journal of Microbiology		1517-8382
10. Brazilian Journal of Veterinary Research and Animal Science		1413-9596
11. Ciência e Tecnologia de Alimentos		0101-2061
12. Ciência Rural		0103-8478
13. Engenharia Agrícola		0100-6916
14. Fitopatologia Brasileira		0100-4158
15. Horticultura Brasileira		0102-0536
16. Interciencia		0378-1844
17. Journal of Venomous Animals and Toxins		0104-7930
18. Journal of Venomous Animals and Toxins including Tropical Diseases		1678-9199
19. MADERAS: Ciencia y Tecnología		0718-221X

## 2. Selección de la metodología de la cosecha:

1. **Utilizando XML formatos existentes de SciELO , que representan artículos de SciELO en:**

- **Dublin core DC XML**

[http://www.scielo.br/oai/scielo-oai.php?verb=ListRecords&metadataPrefix=oai\\_dc&set=0001-3765](http://www.scielo.br/oai/scielo-oai.php?verb=ListRecords&metadataPrefix=oai_dc&set=0001-3765)

- **Formato XML con más detalles en datos para PubMed Central**

<http://artigos.scielo.br/S0001-371419980003.xml>

2. **Creación directa de AGRIS AP XML desde aplicación OAI-PMH de SciELO**

<http://www.scielo.br/oai>

# Resultados de la cosecha en DC

```
http://www.scielo.br/oai/scielo-oai.php?verb=ListRecords&metadataPrefix=oai_dc&set=0001-3765 - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Forward Stop Home Search Favorites
Address http://www.scielo.br/oai/scielo-oai.php?verb=ListRecords&metadataPrefix=oai_dc&set=0001-3765 Go Links
Google G Go Bookmarks 56 blocked Check AutoLink AutoFill Send to Settings
<?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>2008-08-26T06:06:56Z</responseDate>
  <request verb="ListRecords" metadataPrefix="oai_dc" set="0001-3765">http://www.scielo.br/oai/scielo-oai.php</request>
- <ListRecords>
  - <record>
    - <header>
      <identifier>oai:scielo:S0001-37652000000100001</identifier>
      <datestamp>2000-03-01</datestamp>
      <setSpec>0001-3765</setSpec>
    </header>
    - <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>
          <![CDATA[ Some reflections on science, industry and education ]]>
        </dc:title>
        - <dc:creator>
          <![CDATA[ GENNES,PIERRE G. DE ]]>
        </dc:creator>
        - <dc:publisher>
          <![CDATA[ Anais da Academia Brasileira de Ciências ]]>
        </dc:publisher>
        <dc:date>2000-03-01</dc:date>
        <dc:type>journal article</dc:type>
        <dc:format>text/html</dc:format>
        <dc:identifier>http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0001-
        37652000000100001</dc:identifier>
        <dc:language>en</dc:language>
      </oai_dc:dc>
    </metadata>
  </record>
</ListRecords>
Done Internet
```

# Testeo utilizando el cosechador (PKP Harvester 2) de artículos SciELO en DC

Los resultados en DC contienen menos elementos (solamente 15 DC) y no pueden ser incluidos en el repositorio de AGRIS (formato AGRIS AP)



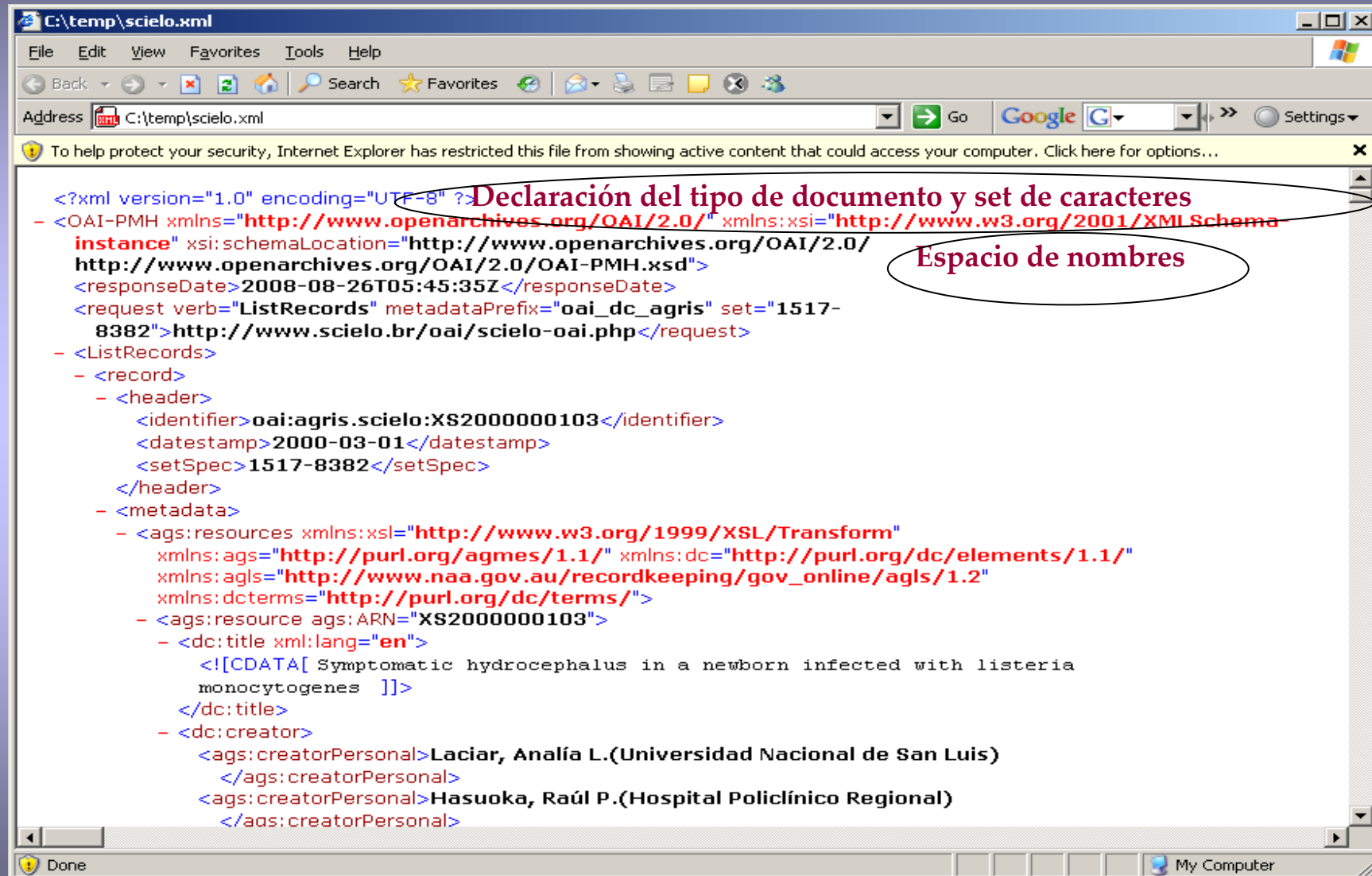
The screenshot displays the Open Archives Harvester 2 web interface. At the top, it reads "PUBLIC KNOWLEDGE PROJECT Open Archives Harvester<sub>2</sub>". Below this is a navigation menu with links for HOME, ABOUT, ADMINISTRATION, SEARCH, BROWSE, and HELP. The breadcrumb trail shows "Home > Browse > SciELO - Scientific Electronic Library Online > Harvester2". The main heading is "Harvester2" followed by "SciELO - Scientific Electronic Library Online". A link for "VIEW ARCHIVE INFO" is present. The "Sort By:" options are "TITLE", "DATE", and "None". Two search results are listed:

- » **On focal stability in dimension two**  
Peixoto, Mauricio M.  
Pugh, Charles C.  
2007-03-01  
[VIEW RECORD](#) | [VIEW ORIGINAL](#)
- » **Finitely curved orbits of complex polynomial vector fields**  
Mafra, Albetã C.  
2007-03-01  
[VIEW RECORD](#) | [VIEW ORIGINAL](#)

On the right side, there is an "ADMINISTRATION" section with the text "You are logged in as... harvester2" and links for "Administration" and "Log Out". Below that is a "CONTENT" section with a search input field, a checkbox for "in all archives", and a "Search" button. At the bottom of the right sidebar is a "FONT SIZE" section with three icons for adjusting text size.



# Resultados de la cosecha en AGRIS AP



```
<?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>2008-08-26T05:45:35Z</responseDate>
  <request verb="ListRecords" metadataPrefix="oai_dc_agris" set="1517-8382">http://www.scielo.br/oai/scielo-oai.php</request>
  <ListRecords>
    <record>
      <header>
        <identifier>oai:agris.scielo:XS2000000103</identifier>
        <timestamp>2000-03-01</timestamp>
        <setSpec>1517-8382</setSpec>
      </header>
      <metadata>
        <ags:resources xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
          xmlns:ags="http://purl.org/agmes/1.1/" xmlns:dc="http://purl.org/dc/elements/1.1/"
          xmlns:agls="http://www.naa.gov.au/recordkeeping/gov_online/agls/1.2"
          xmlns:dcterms="http://purl.org/dc/terms/">
          <ags:resource ags:ARN="XS2000000103">
            <dc:title xml:lang="en">
              <![CDATA[ Symptomatic hydrocephalus in a newborn infected with listeria
                monocytogenes ]]>
            </dc:title>
            <dc:creator>
              <ags:creatorPersonal>Laciar, Analía L.(Universidad Nacional de San Luis)
              </ags:creatorPersonal>
              <ags:creatorPersonal>Hasuoka, Raúl P.(Hospital Policlínico Regional)
              </ags:creatorPersonal>
            </dc:creator>
          </ags:resource>
        </ags:resources>
      </metadata>
    </record>
  </ListRecords>
</OAI-PMH>
```

**Declaración del tipo de documento y set de caracteres**

**Espacio de nombres**



# Más detalles sobre AGRIS AP están disponibles en documentos de AGRIS

- Guía del usuario de AGRIS application profile (<http://www.fao.org/docrep/008/ae909e/ae909e00.htm>)
- Generando AGRIS AP XML desde bases de datos locales (<http://www.fao.org/docrep/008/ae908e/ae908e00.htm>)
- DTD para validar el XML (<http://www.purl.org/agmes/agrisap/dtd/>)

# Especificaciones y requerimientos de AGRIS AP (elementos obligatorios y esquemas)

- Campos obligatorios:
  - Título
  - Fecha
  - Idioma
  - Tema
  - Disponibilidad
- Esquemas
  - Códigos de Idioma (Códigos de tres letras)
  - Vocabularios controlados por tema (AGROVOC, CABI)
  - Formato de fecha (YYYY-MM-DD)

### 3. Adaptación del programa OAI-PMH plug-in en SciELO :

- Actualización de la interfaz OAI-PMH plug-in existente para DC, para permitir creación directa de AGRIS AP XML desde aplicación SciELO
- AGRIS AP formato fue incluido como un parametro adicional en los verbos de OAI-PMH  
<http://www.scielo.br/oai>
- ISSN fue usado para el parametro set
- Necesidad de un cosechador para proseguir con la cosecha automatica en la FAO utilizando el verbo ListRecords para cada revista seleccionada y por fecha.

# OAI funcionalidades: seis verbos de OAI

Verbo	Función
Identify	Describe el repositorio
ListMetadataFormats	Da todos los formatos de metadatos soportados por este repositorio
ListSets	Describe los posibles subgrupos definidos por repositorio (semántica o tipo de doc.)
ListIdentifiers	Lista registro de identificadores que dan set/rango de fechas/formatos de metadatos de este repositorio
ListRecords	Da todos los registros para set/rango de fechas/formato de metadatos de este repositorio
GetRecord	Obtener un solo registro de identificación

# URL de OAI-PMH

## OAI URL

URL:  http://www.scielo.br/oai/scielo-oai.php

## verbs

Identify

ListMetadataFormats

ListIdentifiers

ListSets

ListRecords

GetRecord

## setup parameters

verb

MetadataPrefix

oai\_dc | oai\_dc\_agris

set

ISSN

identifier

from

YYYY-MM-DD

until

YYYY-MM-DD

resumptionToken

do harvesting

## verb description

It retrieves the header of articles in SciELO.

## Compulsory parameters

- verb

## output

Harvested URL:

[http://www.scielo.br/oai/scielo-oai.php?verb=ListRecords&metadataPrefix=oai\\_dc\\_agris&set=0001-3714](http://www.scielo.br/oai/scielo-oai.php?verb=ListRecords&metadataPrefix=oai_dc_agris&set=0001-3714)

Result:

## 4.Cosecha de los datos existentes

Uso del verbo ListRecords con parámetros:

- **set** (setSpec) que identifica el ISSN de la revista;
- **metadataPrefix**: formato AGRIS AP XML (oai\_dc\_agris)
- **from/until**

[http://www.scielo.br/oai/scielo-oai.php?verb=ListRecords&metadataPrefix=oai\\_dc\\_agris&from=1998-01-01&until=1999-01-01&set=0001-3714](http://www.scielo.br/oai/scielo-oai.php?verb=ListRecords&metadataPrefix=oai_dc_agris&from=1998-01-01&until=1999-01-01&set=0001-3714)

La cosecha se hace en grupos de 100 registros cada una y después se usa “Resumption Token” para el próximo grupo de registros

[http://www.scielo.br/oai/scielo-oai.php?verb=ListRecords&metadataPrefix=oai\\_dc\\_agris&from=1999-01-01&until=1999-12-31&set=0001-3714&resumptionToken=HR\\_S0001-37141999000400013:0001-3714:1999-01-01:1999-12-01](http://www.scielo.br/oai/scielo-oai.php?verb=ListRecords&metadataPrefix=oai_dc_agris&from=1999-01-01&until=1999-12-31&set=0001-3714&resumptionToken=HR_S0001-37141999000400013:0001-3714:1999-01-01:1999-12-01)

## 5. Validación de AGRIS AP XML

- Fue utilizado el editor Altova XMLSpy
- Cuando XML es validado y reúne los requerimientos de AGRIS puede ser agregado al repositorio AGRIS

# Problemas encontrados

- Metadatos: faltan elementos obligatorios para AGRIS AP (tema, título)
- AGROVOC no fue usado por SciELO y por eso con los datos de SciELO no se puede utilizar herramientas de extensión de la semántica basada en AGROVOC.
- A menudo falta identificación del idioma (e.j. en título, tema);
- Los estándares utilizados por AGRIS y SciELO para la descripción bibliográfica son diferentes
- El formato de fechas incluye datos que no se puede utilizar con OAI-PMH (e.j.2008-00-00)



## 6. Inclusión en la base de datos AGRIS

1. Cosecha de más de 15 000 artículos de SciELO
2. Inclusión en el portal de AGRIS para búsqueda común.

Se puede buscar en el sitio AGRIS :

<http://www.fao.org/agris/search/search.do?query=%2Bcenter%3A%28XS%29>



[Search tips](#)

+center:(XS)

Search Assistant

Search Reset  
[Set Preferences](#)  
[Clear preferences](#)

Your Search For +center:(XS) found 1758 results . Showing Results 1-10

Save selected Save all Save results as: XML

Go to Page 1 2 3 4 5 6 7 8 9 10 Next

MICROBIAL COUNTS OF DARK RED LATOSOL SAMPLES STORED AT DIFFERENT TEMPERATURES

**Vieira, Francisco Cleber Sousa ; Nahas, Ely(Universidade Estadual Paulista) ; ;**

The number of colony forming units (CFU) of different groups of bacteria and fungi in samples stored at temperatures of 5 to -12oC for 0-32 weeks was evaluated. The number of CFU obtained after the different periods of storage of red latossol soil was compared with the number of colonies obtained immediately after removal of soil samples (time zero). The number of total bacteria and actinomycetes in the samples remained practically unchanged throughout the storage period.



MICROBIAL ACTIVITY OF DARK RED LATOSOL SAMPLES STORED AT DIFFERENT TEMPERATURES

**Vieira, Francisco Cleber Sousa ; Nahas, Ely(Universidade Estadual Paulista) ; ;**

The enzymatic activity of soil samples stored at temperatures of 5 to -12oC and at room

# Conclusiones y perspectivas

- El nuevo formato AGRIS AP para la representación de metadatos de los artículos de SciELO enriquece la colección AGRIS dando más visibilidad y acceso a los artículos con textos completos de SciELO para la comunidad agraria.
- Los centros participantes de SciELO no tienen la necesidad de enviar datos a dos sistemas paralelos y también están integrados utilizando reglas y estándares comunes (AGRIS AP and OAI-PMH)
- La integración en el portal AGRIS, hace posible la búsqueda única por todos los usuarios.
- El usuario puede efectuar búsquedas más precisas: por autor, título, etc.

## Conclusiones (cont.)

- El programa OAI-PMH plug-in de BIREME está disponible (en línea) por los centros AGRIS para no duplicar la catalogación de registros de SciELO en las bases de datos locales.
- Esta experiencia es un ejemplo y un paso adelante hacia el Acceso Abierto y el intercambio de metodologías y recursos comunes.

# Agradecimientos:

- Mi agradecimiento a todos los que me acompañaron en el equipo de BIREME y FAO:
- por SciELO: Gustavo Fonseca, Fabio Batalha, Solange Santos, Renato Murasaki (para OAI-PMH plug-in y posibilidad de cosecha desde SciELO),
- por FAO: Irene Onyanha y Imma Subirats (selección de las revistas), Maria Folch (cosecha de artículos, validación y edición), Stefano Anibaldi (validación y publicación en AGRIS para búsqueda),
- Abel Packer, Steve Katz y Johannes Keizer por apoyar la iniciativa.

Gracias !