

# **Approaches to make CDS/ISIS databases interoperable with OAI-compliant digital libraries**

F. Jayakanth, K. Maly, M. Zubair, and L Aswath

## **Authors:**

F. Jayakanth is a visiting Fulbright fellow at the Computer Science Department, Old Dominion University, Norfolk, Virginia, US. E-mail: [franc@cs.odu.edu](mailto:franc@cs.odu.edu);

K. Maly is the Kaufman Professor and Chair at the Computer Science Department, Old Dominion University, Norfolk, Virginia, US. E-mail: [maly@cs.odu.edu](mailto:maly@cs.odu.edu)

M. Zubair, is a professor at the Computer Science Department, Old Dominion University, Norfolk, Virginia, US. E-mail: [zubair@cs.odu.edu](mailto:zubair@cs.odu.edu)

Dr. L. Aswath is a professor at the Department of Library and Information Science Bangalore University, India. E-mail: [lalitha\\_aswath@hotmail.com](mailto:lalitha_aswath@hotmail.com)

## **Abstract**

Research Paper

**Purpose:** To make CDS/ISIS databases OAI-Compliant

**Design/Methodology/Approach:** One of the biggest obstacles for information dissemination to a user community is that many digital libraries or bibliographic databases use heterogeneous technologies that inhibit interoperability. The Open Archives Initiative (OAI) addresses interoperability by using a framework to facilitate the discovery of content stored in distributed archives or bibliographic databases through the use of the OAI Protocol for Metadata Harvesting (OAI-PMH). Though the OAI-PMH is becoming the de facto standard, many of the legacy databases or bibliographic systems, for various reasons, are not OAI-compliant. CDS/ISIS is one such database system. In

this paper, we discuss the static repository approach to make CDS/ISIS databases OAI-compliant. The records from a CDS/ISIS database are exported onto a file and the exported records are then converted to the static repository (SR) XML file format. The SR file is made OAI-compliant either by ingesting it into the Kepler Archivelet or through the intermediation of a static repository gateway.

**Findings:** The SR is a simple approach for making bibliographic records contained in non- OAI-compliant systems, OAI-compliant.

**Research Limitations/Implications :** The SR approach is meant for records which are relatively static. If the OAI-compliance is achieved through the SR gateway, then the number of records contained in a SR file should not exceed 5000. However, if the SR file is ingested into the Kepler Archivelet, then there is no restriction on the number of records contained in a SR file.

**Keywords:** CDS/ISIS; Digital library; OAI-PMH; Interoperability; Kepler Archivelet; Static repository; Static repository gateway

**Word length:** 2,370

## **1. Introduction**

Digital libraries (DLs) comprise of material stored in computer systems in a form that allows them to be manipulated (Noerr, 2003). One early example, from 1991, is the preprint server, arXiv.org (<http://www.arxiv.org>) at Los Alamos National Laboratory (LANL), which archives articles on high energy physics. ArXiv is now based at Cornell University and encompasses other fields such as mathematics, computer science and quantitative biology. Other examples of similar preprint collections include:

- Cogprints for Cognitive Sciences (<http://cogprints.ecs.soton.ac.uk/>)

- Research Papers for Economics (RePEC) (<http://repec.org>)
- Networked Computer Science Technical Reference Library (NCSTRL) (<http://www.ncstrl.org>)

As far as interoperability is concerned, each collection is mainly restricted to a specific subject domain. For example, RePEC has its own framework for interoperability amongst different RePEC nodes. There was, initially, no interoperability across the various preprint servers. As research activity is becoming more and more interdisciplinary, it was soon realised that the impact of preprint servers could be significantly enhanced, if a framework could be developed to support interoperability across the preprint servers (Van de Sompel and Lagoze, 2000). This led to the establishment of the Open Archives Initiative (OAI) and the OAI Protocol for Metadata Harvesting (OAI-PMH). OAI-PMH provides a mechanism for an application-independent interoperability framework based on metadata harvesting from OAI-compliant systems. OAI-PMH uses the concept of ‘data providers’ and ‘service providers’. Service providers harvest metadata using the OAI-PMH protocol requests from OAI-compliant data providers.

However, for some data providers, implementing the OAI-PMH protocol has either been not possible or has been difficult to achieve. The developers of the OAI-PMH framework were aware of this fact and researched further to find an OAI-PMH based solution for such data providers. This led to the development of the OAI-PMH static repository (SR) and static repository gateway (SRG) model. The SR is an XML file, available at a persistent URL, that conforms to the rules specified in the OAI specification document (Lagoze et al., 2004). SR provides a simple approach for exposing relatively static and small collections of records through the OAI-PMH. An SR XML file can also be

imported into the Kepler Archivelet, a light-weight, easy-to-install, and OAI-compliant archive developed by the DL research group at Old Dominion University (<http://dlib.cs.odu.edu/>).

The SR model helps in making data from legacy database systems, OAI-compliant. In this article we explain the approaches we have adopted to expose the metadata records from CDS/ISIS databases for harvesting through the intermediation of the Kepler Archivelet and the SR gateway.

## **2. The need for interoperability of DLs**

Interoperability is the ability of multiple systems with different hardware and software platforms, data structures, and interfaces to exchange data with minimal loss of content and functionality (NISO, 2003). Interoperability is all the more important currently where the nature of research work being carried out across the world is becoming more and more interdisciplinary. The impact of DLs will be much greater when individual DLs are interoperable and are able to provide a seamless interface for the end-users. Many interoperability frameworks have been developed to date including:

- Z39.50 (<http://www.loc.gov/z3950/agency/resources>)
- Simple Digital Library Interoperability Protocol (SDLIP) (<http://www-diglib.stanford.edu/~testbed/doc2/SDLIP> )
- Guildford protocol (<http://openlib.org/acmes/root/docu/guilp.html>)
- Dienst (<http://www.cs.cornell.edu/cdlrg/dienst/protocols/DienstProtocol.htm>).

These frameworks provide a variety of end-user services, such as searching and browsing across the repositories implementing the same protocol. However, little or no autonomous metadata sharing is supported across diverse environments. The end result is

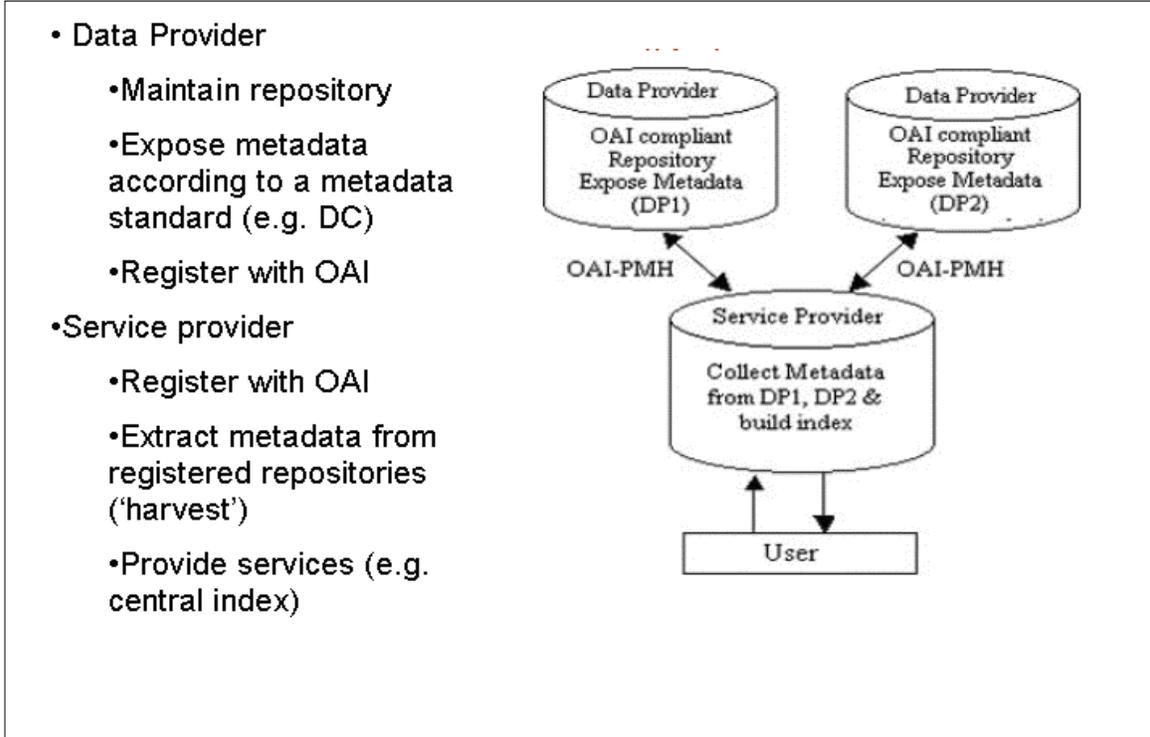
that the end-users are faced with multiple search interfaces making resource discovery harder and there is no machine-based way of sharing the metadata. These limitations in the existing protocols led to the development of the OAI-PMH.

### **3. The OAI-PMH protocol and the SR and SRG model**

The data providers within OAI-PMH are the creators and maintainers of OAI-compliant repositories, and expose their metadata using the Dublin Core (DC) metadata set or any other metadata format supported by the OAI-PMH protocol. The service providers harvest the exposed metadata from various registered data providers and provide federated DL services based on the harvested data as shown in Figure 1 ( Liu, Maly and Zubair, 2001).

Take in Figure 1

Figure 1. Interoperability through the OAI-PMH protocol



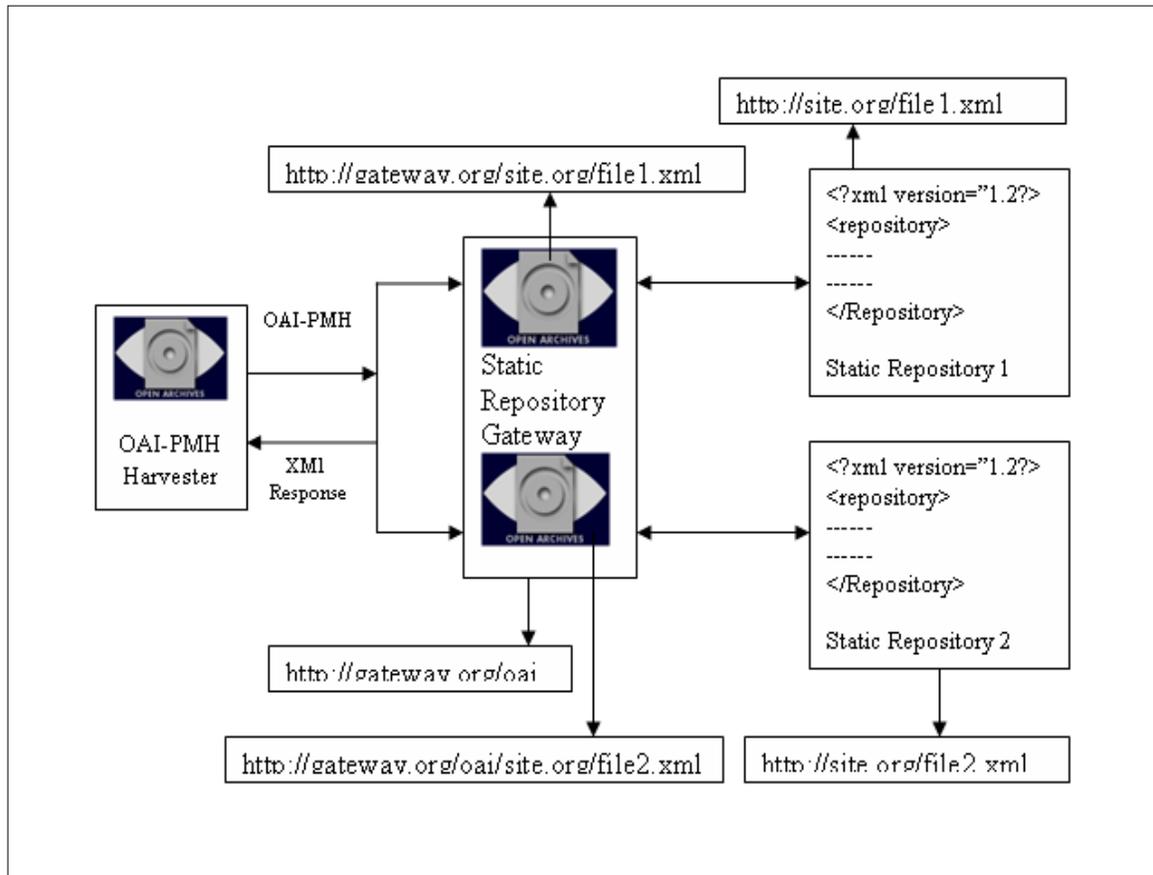
This mechanism greatly aids end-users as they are relieved of visiting individual repositories, as by using the federated search service, they can find publications available from various data providers. The SR and SRG model provides a solution to data providers who are unable to install and maintain a full fledged OAI-compliant systems.

A SR contains metadata records and supporting information required for the purpose of harvesting via the OAI-PMH through the intermediation of SRG (Hochstenbach et al.,2003). A SR is not an OAI-PMH repository by itself as it is just an XML file, not a server that can respond to OAI-PMH requests.

A SRG provides intermediation for one or more SRs. It assigns each repository a unique SR base URL, all with a common SRG URL prefix, and thereby expose each

individual SR as an individual OAI-PMH repository. The SRG is the software component, which needs to be downloaded and installed. Figure 2 illustrates the interaction of SRs, SRGs, and the OAI-PMH harvester.

Figure 2– Interaction of Static Repositories, Gateway, and OAI-PMH Harvester



A SR XML file can also be made OAI-compliant through the intermediation of Kepler archivelet, a light-weight, easy-to-install, and OAI-compliant archive.

#### 4. The Kepler Archivelet.

Self-archiving is a process whereby researchers deposit a copy of their final version of a paper, accepted for publication, on a computer server accessible on the Internet and

preferably conforming to the standards like the OAI-PMH framework. It is the easiest and fastest route to providing free access to the peer reviewed scholarly literature (Harnard, 1995). Studies have shown that free online availability of scholarly publications substantially increases their impact (Lawrence, 2001). Self-archiving approaches are mainly based on either discipline or institutional based repositories. Today, there are quite a few open source OAI-compliant DL software systems, which help in the creation and maintenance of OAI-compliant repositories. EPrints.org (<http://www.eprints.org>) and DSpace (<http://www.dspace.org>) are the two most popular ones. These are designed mainly to be used at the institutional/organisational level rather than by individuals on their desktop systems. The general experience of institutions which have set up self-archiving facilities for their researchers is that, many researchers fail to complete the last phase of their publication cycle, i.e., self-archiving research publications on institutional repositories. This led to the development of the Kepler framework, an OAI-compliant ‘personal data provider’ or ‘archivelet’ (Maly et al, 2001; Maly et al, 2004, Liu et al., 2002). Kepler is a light-weight utility that facilitates the deployment of an OAI-compliant DL easily and quickly on a desktop PC.

### **5. CDS/ISIS databases and interoperability**

CDS/ISIS is an information storage and retrieval software that has been developed by Unesco since 1985 and is available for free (<http://www.unesco.org/isis>). Many organisations and individuals across the world are using this package, either standalone or in a network, for developing databases. The software by itself does not facilitate publishing the database on the Web. However, a number of free gateway tools, such as WWWIsis (<http://productos.bvsalud.org/product.php?id=wwwisis&lang=en>) are

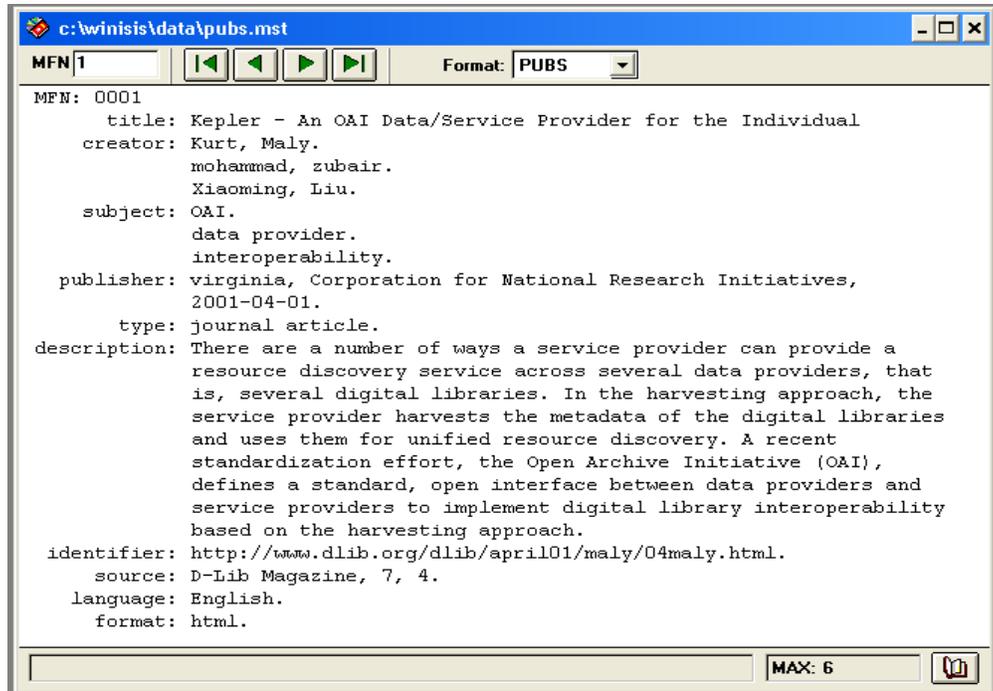
available for this (Jayakanth and Jayashree, 1999). At present, CDS/ISIS databases cannot participate in the interoperability with other OAI-compliant DLs. It is in this context that we are proposing two approaches for making CDS/ISIS databases interoperable with other OAI-compliant DLs. In both these approaches the SR XML file plays a key role. As the name suggests, the use of SR is ideal for repositories that are more or less static, i.e. the records in the repository do not change often. There are other approaches as well to make non-OAI resources available for federated searching (Coleman et al., 2004).

### *5.1 Approach 1 – SR XML File and the Kepler Archivelet*

The CDS/ISIS print formatting language can be used to a great advantage to print the records onto a file in any required format. Figure 3 shows the display, in browse view, of a sample record from a CDS/ISIS database.

Take in Figure 3

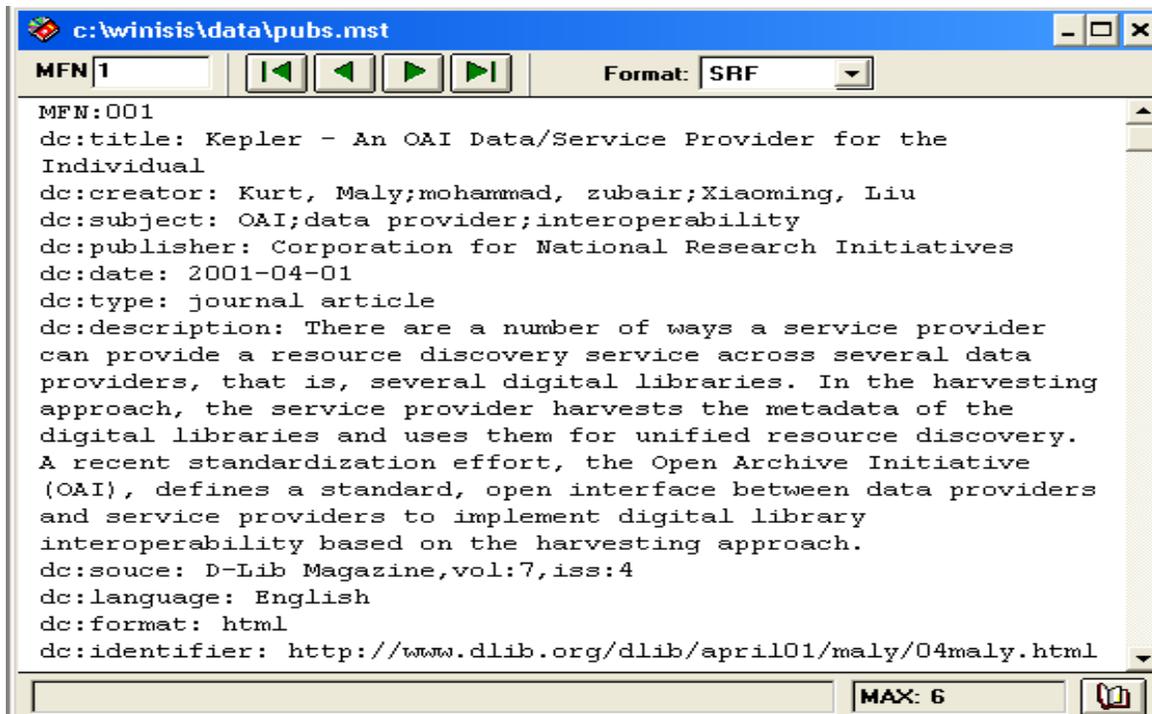
Figure 3. A sample CDS/ISIS record



While printing, it is necessary to ensure that the field names of CDS/ISIS records are translated into the equivalent DC metadata set (<http://dublincore.org>). Figure 4 shows a screenshot of the same record printed into a file in DC format.

Take in Figure 4

Figure 4. A sample CDS/ISIS record in Dublin Core format



The next step is to convert the records into the SR XML file as shown in Figure 5. This is achieved by writing a simple conversion program.

Take in Figure 5.

Figure 5. A sample record from an SR XML file

```

<?xml version="1.0" encoding="UTF-8"?>
<Repository xmlns="http://www.openarchives.org/OAI/2.0/static-repository"
xmlns:oai="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/static-repository
http://www.openarchives.org/OAI/2.0/static-repository.xsd">
<Identify>
<oai:repositoryName>Demo Static Repository</oai:repositoryName>
<oai:baseURL>http://localhost/cds.html</oai:baseURL>
<oai:protocolVersion>2.0</oai:protocolVersion>
<oai:adminEmail>franc@cs.odu.edu</oai:adminEmail>
<oai:earliestDatestamp>2004-09-15</oai:earliestDatestamp>
<oai:deletedRecord>no</oai:deletedRecord>
<oai:granularity>YYYY-MM-DD</oai:granularity>
</Identify>
<ListMetadataFormats>
<oai:metadataFormat>
<oai:metadataPrefix>oai_dc</oai:metadataPrefix>
<oai:schema>http://www.openarchives.org/OAI/2.0/oai_dc.xsd</oai:schema>
<oai:metadataNamespace>http://www.openarchives.org/OAI/2.0/oai_dc/oai_dc.xsd</oai:
i:metadataNamespace>
</oai:metadataFormat>
</ListMetadataFormats>
<ListRecords metadataPrefix="oai_dc">
<oai:record>
<oai:header>
<oai:identifier>oai:cds:mfn-1</oai:identifier>
<oai:datestamp>2004-09-25</oai:datestamp>
</oai:header>
<oai: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-in
stance" xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/ http://w
ww.openarchives.org/OAI/2.0/oai_dc.xsd">
<dc:title>Kepler - An OAI Data/Service Provider for the Individual </dc:title>
<dc:creator>Kurt, Maly</dc:creator>
<dc:creator>mohammad, zubair</dc:creator>
<dc:creator>Xiaoming, Liu</dc:creator>
<dc:subject>OAI</dc:subject>
<dc:subject>data provider</dc:subject>
<dc:subject>interoperability</dc:subject>
<dc:publisher>Corporation for National Research Initiatives</dc:publisher>
<dc:date>2001-04-01</dc:date>
<dc:type>journal article</dc:type>
<dc:description>There are a number of ways a service provider can provide a resource discovery service
across several data providers, that is, several digital libraries. In the harvesting approach, the service
provider harvests the metadata of the digital libraries and uses them for unified resource discovery. A
recent standardization effort, the Open Archive Initiative (OAI), defines a standard, open interface
between data providers and service providers to implement digital library interoperability based on the
harvesting approach.</dc:description>
<dc:language>English</dc:language>
<dc:identifier>http://www.dlib.org/dlib/april01/maly/04maly.html</dc:identifier>
<dc:format>html</dc:format>
</oai_dc:dc>
</oai:metadata>
</oai:record>
</ListRecords>
</Repository>

```

The SR XML file can now be imported into the Kepler Archivelet and thus the records from a CDS/ISIS database becomes OAI-compliant. Figure 6 is a screenshot of a Kepler Archivelet with some sample records from a CDS/ISIS database, imported through the SR XML file.

Take in Figure 6

Figure 6. Sample records from a CDS/ISIS database imported into a Kepler Archivelet



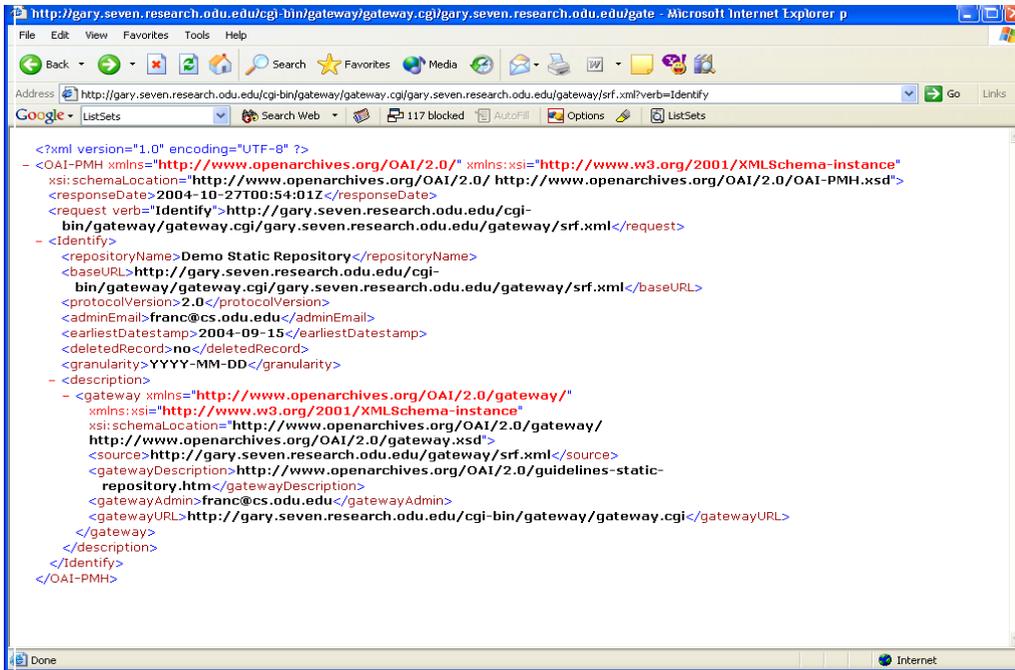
Identifier	Title	Creator	Publ. Date
oai.cds:mfn-2	Arc - An OAI Service Provide...	Xiaoming, Liu	2003-04-01
oai.cds:mfn-1	Kepler - An OAI Data/Service...	Kurt, Maly	2001-04-01
oai.cds:mfn-4	Light-Weight Communal Di...	kurt, maly	2004
oai.cds:mfn-6	Specification for an OAI Stat...	herbert, van, de, sompel	2004
oai.cds:mfn-3	The Dienst-OAI Gateway	Terry, L, Harrison	2003-01-01
oai.cds:mfn-5	The OAI-PMH Static Reposit...	Patrick, Hochstenbach	2003

### 5.2 Approach 2 - SR XML file and SRG

In this approach, a SR XML file becomes OAI-compliant through the intermediation of the SRG. The SRG is an easy to install open source software. After the successful installation and configuration of the SRG, it can respond to the OAI-PMH queries. Figures 7 and 8 are the screenshots of the SRG responses to two of the OAI-PMH verbs, namely, 'Identity' and 'GetRecord'.

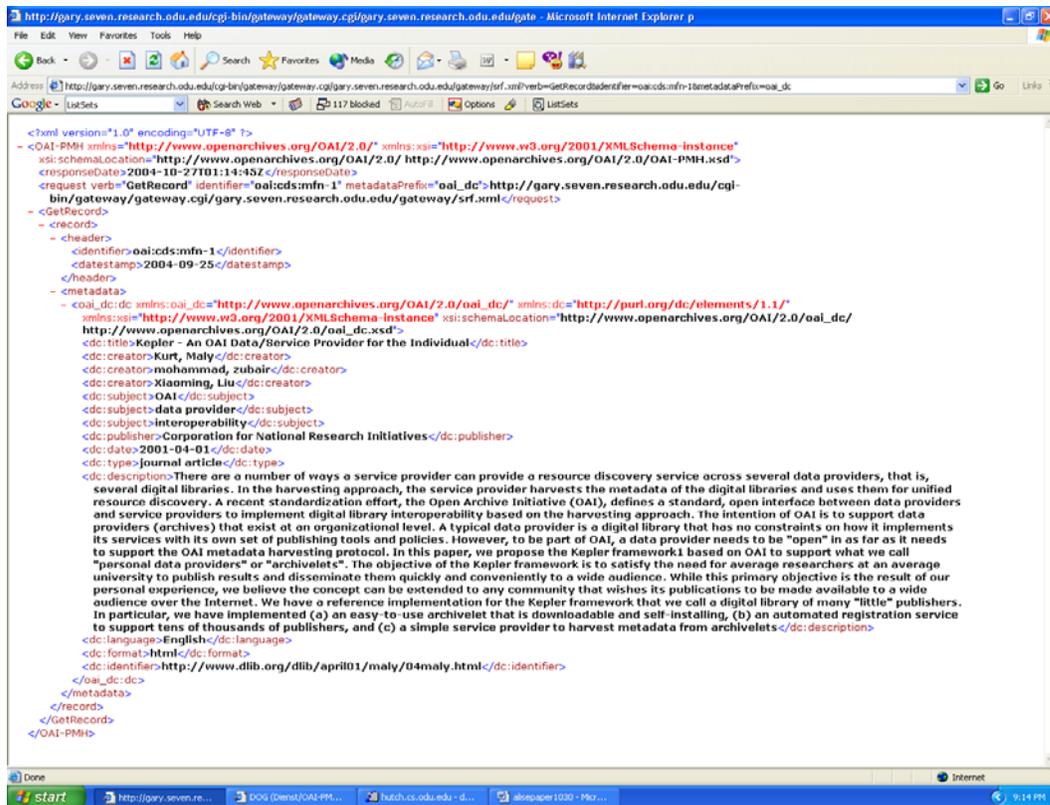
Take in Figure 7

Figure 7. Response from the SRG for the Identify verb



Take in Figure 8

Figure 8 Response from the SRG for the GetRecord verb



Federated search service providers can harvest metadata from the SRGs and build federated search and browse services. This approach, however, has certain limitations as compared to the standard OAI-PMH repositories:

- the SR file can have only up to 5000 records;
- the SRs do not support set, deleted records, response compression, harvesting date granularity other than YYYY-MM-DD, and resumptionTokens.

## 6. Conclusion

The static repository approach discussed in this paper enable bibliographic records, which are relatively static to become OAI-compliant. A better approach for making CDS/ISIS databases interoperable with other OAI-compliant DLs is to develop a general purpose OAI interface for the CDS/ISIS databases. This general purpose interface should respond directly to all the OAI-PMH verbs and in the process overcome the shortcomings of the

SR model. Work on the development of such an interface is being carried out by the authors.

## References

(All URLs were checked 22.4.05)

Coleman A., Bracke P. and Karthik S. (2004), “ Integration of non-OAI resources for federated searching in DLIST, an Eprints repository”, *D-Lib Magazine*, Vol 10, No. 7/8. Available at: <http://www.dlib.org/dlib/july04/coleman/07coleman.html>

Harnard S. (1995), *The Post Gutenberg Galaxy: How to Get There from Here*. Available at: <http://cogprints.ecs.soton.ac.uk/archive/00001689/00/thes.html>

Hochstenbach P., Jerez H. and Van De Sompel H. (2003), “The OAI-PMH Static Repository and Static Repository Gateway”, *Joint Conference on Digital Libraries 2003, Houston, Texas*. Available at: <http://public.lanl.gov/herbertv//papers/jcdl2003-submitted-draft.pdf>

Jayakanth F. and Jayashree S. (1999), “A tutorial on integrating CDS/ISIS databases on the WWW”, *Information Today and Tomorrow*. Vol 18, Issue 4,

Lagoze C., Van de Sompel H., Nelson M., and Warner S. (2004), *The Open Archives Initiative Protocol for Metadata Harvesting* . Available at:

<http://www.openarchives.org/OAI/openarchivesprotocol.html>

Lawrence S. (2001) , “Free online availability substantially increases a paper’s impact”, *Nature Webdebates*. Available at:

<http://www.nature.com/nature/debates/e-access/Articles/lawrence.html>

Liu X., Maly K. and Zubair M. (2001), “Arc – An OAI service provider for Digital

Library Federation”, *D-Lib Magazine*, Vol 7 No. 4, Available at :

<http://www.dlib.org/dlib/april01/liu/04liu.html>

Liu, X, Maly, K, Zubair, M. (2002), “Enhanced Kepler framework for self-archiving”,

*Proceedings of the 2002 International Conference on Parallel Processing*

*Workshops, Aug 18-21, 2002, Vancouver, Canada.* Available at:

<http://kepler.cs.odu.edu/publications/kepler.pdf>

\*\*\*\* please check this ref. is available – I failed to access it on 22.4.05\*\*\*\*

Maly K., Zubair M., Liu X. (2001), “Kepler – An OAI Data/Service Provider

for the Individuals”, *D-Lib Magazine*, Vol 7 No. 4. Available at:

<http://www.dlib.org/dlib/april01/liu/04liu.html>

Maly K and others (2004). Light-Weight Communal Digital Libraries. Paper presented at

Joint Conference on Digital Libraries, Tucson, Arizona, USA. Available at:

<http://csdl.computer.org/comp/proceedings/jcdl/2004/2493/00/24930237.pdf>

NISO (2003), *Understanding Metadata*, National Information Standards Organization,

Available at: <http://www.niso.org/standards/resources/UnderstandingMetadata.pdf>

Noerr, P. (2003), *The Digital Library Tool Kit*, 3<sup>rd</sup> edition. Available at:

[http://www.sun.com/products-n-solutions/edu/whitepapers/pdf/digital\\_library\\_toolkit.pdf](http://www.sun.com/products-n-solutions/edu/whitepapers/pdf/digital_library_toolkit.pdf)

Van De Sompel H. and Lagoze, C. (2000), “The Santa Fe Convention of the Open

Archives Initiative”, *D-Lib Magazine*, Vol 16, No 2. Available at:

<http://www.dlib.org/dlib/february00/vandesompel-oai/02vandesompel-oai.html>