Building a Digital Repository on a Shoestring Budget

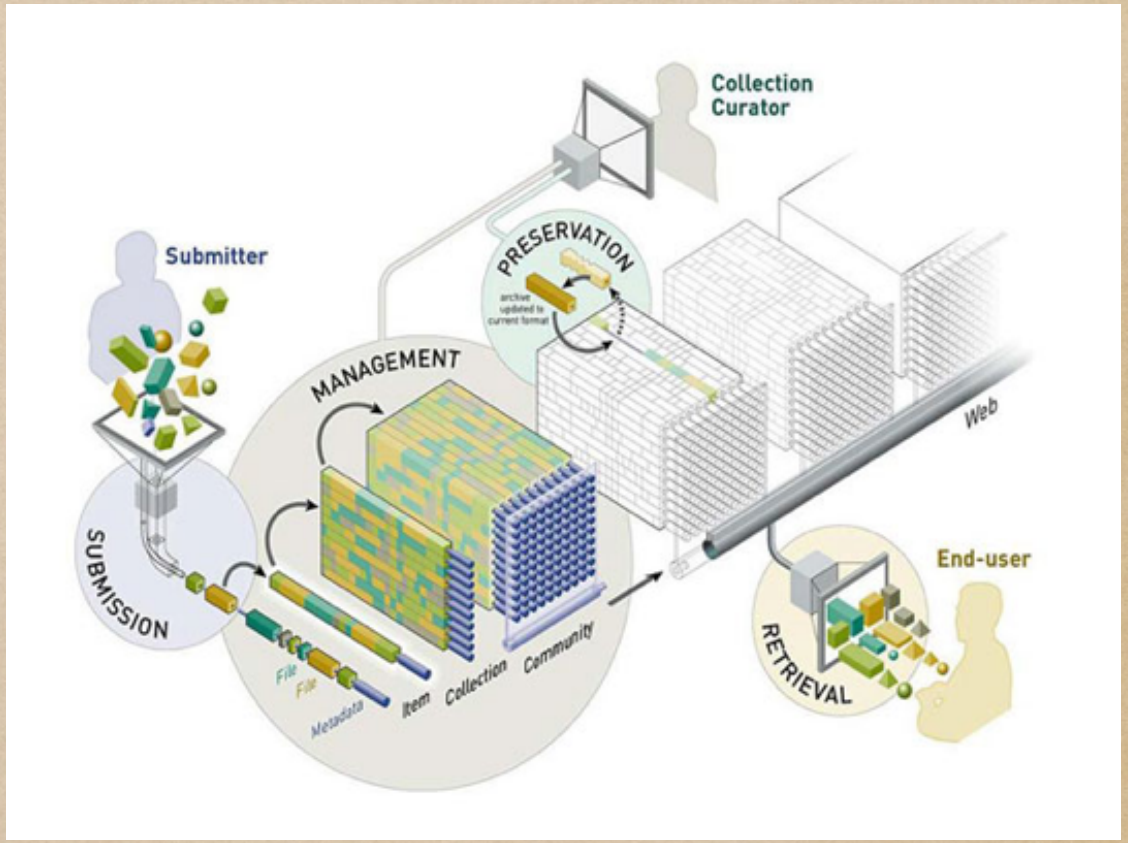
Christinger Tomer
University of Pittsburgh

PALA
September 30, 2014

A version this presentation is available at http://www.pitt.edu/~ctomer/shoestring/

Reasons for Building a Digital Repository

Repositories serve various purposes, but in the main they are built to guarantee the availability of and access to select materials. And, like libraries and archives, a more specific purpose is derived from the community that is served.



Types of Digital Repositories

- **Digital Archives** — complaint with archival standards and practices
- **Digital Repositories** — compliant with library standards for metadata
- **Digital Asset Management Systems** — generally outside the realm of standards for library and archival practice; may comply with standards for records management

Digital Preservation

What is the goal of digital preservation? The ultimate outcome of the preservation process should be authentic preserved objects; that is, the outputs of a preservation process ought to be identical, in all essential respects, to what went into that process.

The Trustworthy Repository

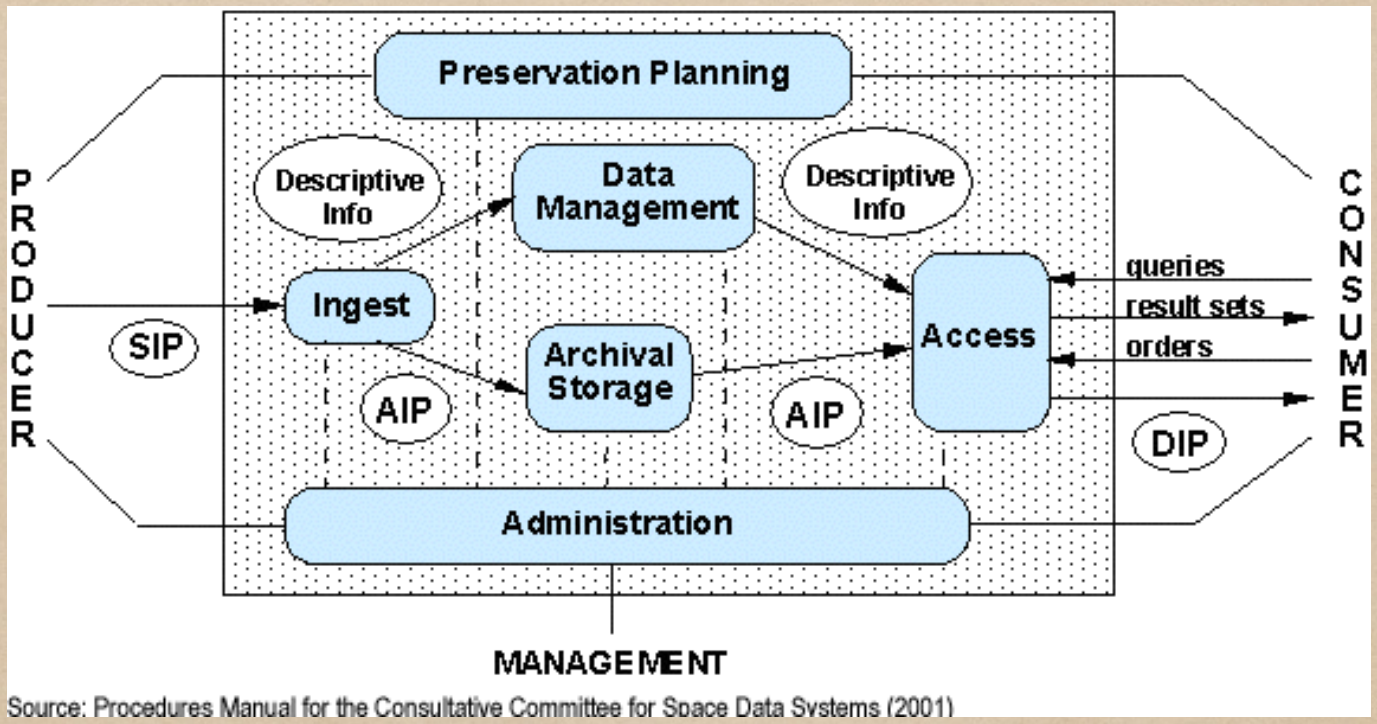
A trusted digital repository is one whose mission is to provide reliable, long-term access to managed digital resources to its designated community, now and in the future. Trusted digital repositories may take different forms: some institutions may choose to build local repositories while others may choose to manage the logical and intellectual aspects of a repository while contracting with a third-party provider for its storage and maintenance. Whatever the overall infrastructure, however, to meet expectations all trusted digital repositories must:

- accept responsibility for the long-term maintenance of digital resources on behalf of its depositors and for the benefit of current and future users;
- have an organizational system that supports not only long-term viability of the repository, but also the digital information for which it has responsibility;
- demonstrate fiscal responsibility and sustainability;
- design its system(s) in accordance with commonly accepted conventions and standards to ensure the ongoing management, access, and security of materials deposited within it;
- establish methods for system evaluation that meet community expectations of trustworthiness;
- be depended upon to carry out its long-term responsibilities to depositors and users openly and explicitly; and
- have policies, practices, and performance that can be audited and measured;

Attributes of a Trusted Repository

- Compliance with the Reference Model for an Open Archival Information System (OAIS)
- Administrative responsibility
- Organizational viability
- Financial sustainability
- Technological and procedural suitability
- System security
- Procedural accountability

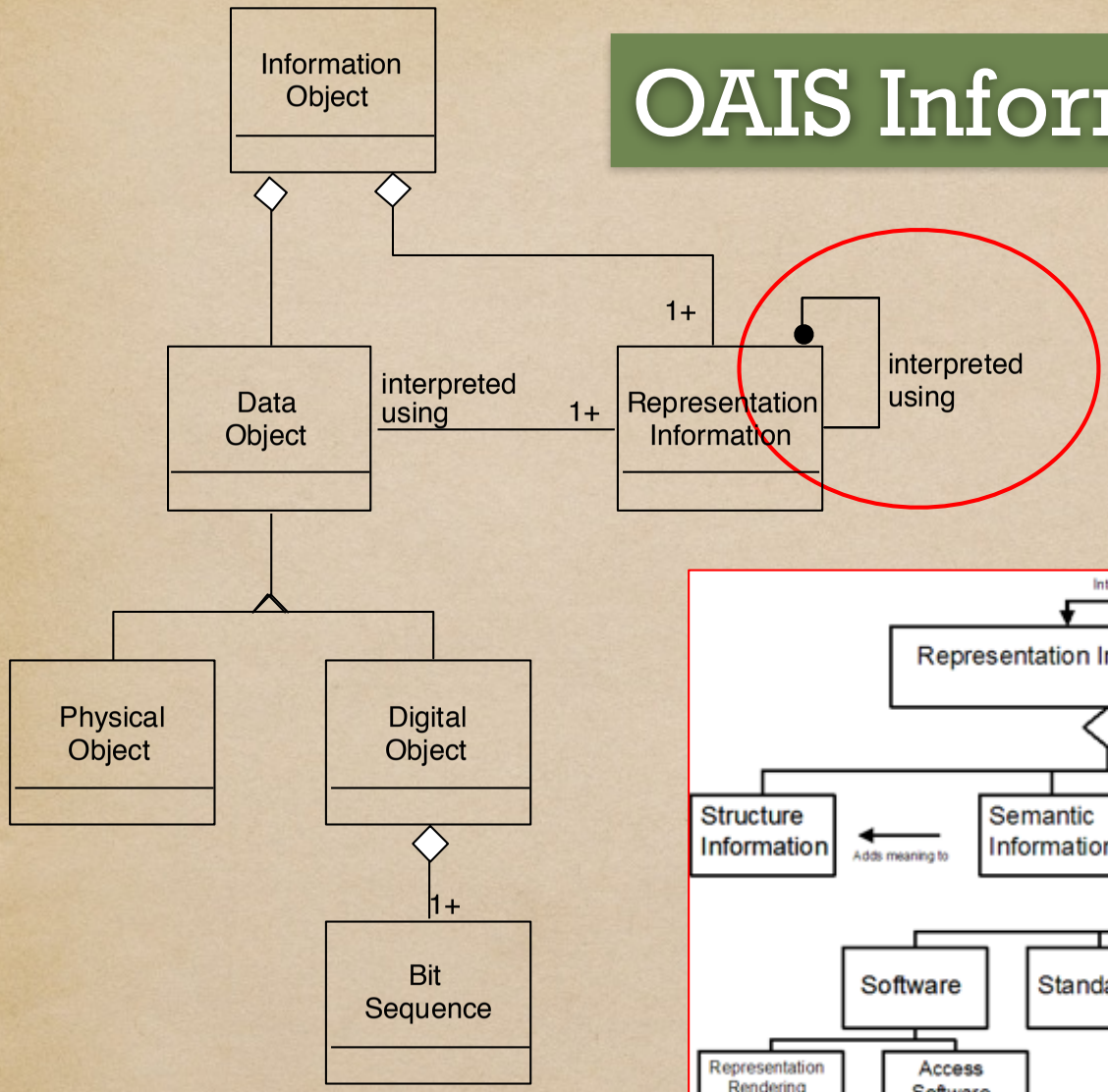
Compliance with the Reference Model for an Open Archival Information System (OAIS)

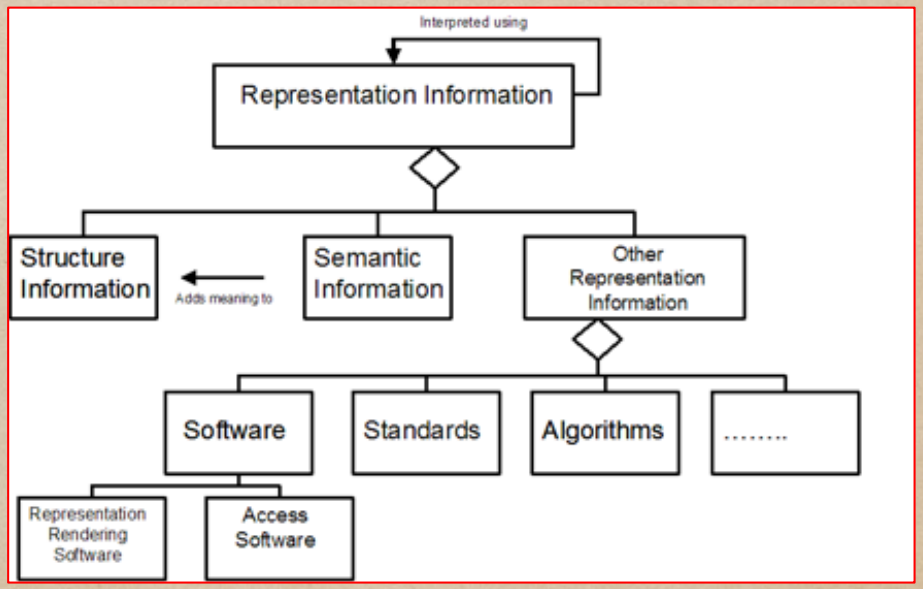


Source: Procedures Manual for the Consultative Committee for Space Data Systems (2001)

The OAIS Reference Model supplies a common framework, including terminology and concepts, for describing and comparing architectures and operations of digital archives. It also provides both a functional model —the specific tasks performed by the repository such as storage or access —and a corresponding information model that includes a model for the creation of metadata to support long-term maintenance and access.

OAIS Information Model





Responsibilities of a Trusted Repository

Research repositories need to understand fully what responsibilities they should assume for the preservation of digital materials. Organizational responsibility must be understood at three basic levels: understanding local requirements and how to meet them; understanding how other organizations might share responsibilities through geography or arrangements such as consortial agreements or shared user communities, disciplines, or format of materials; and understanding which responsibilities can be shared and how. In summary, the major factors in defining the responsibilities of a trusted repository are:

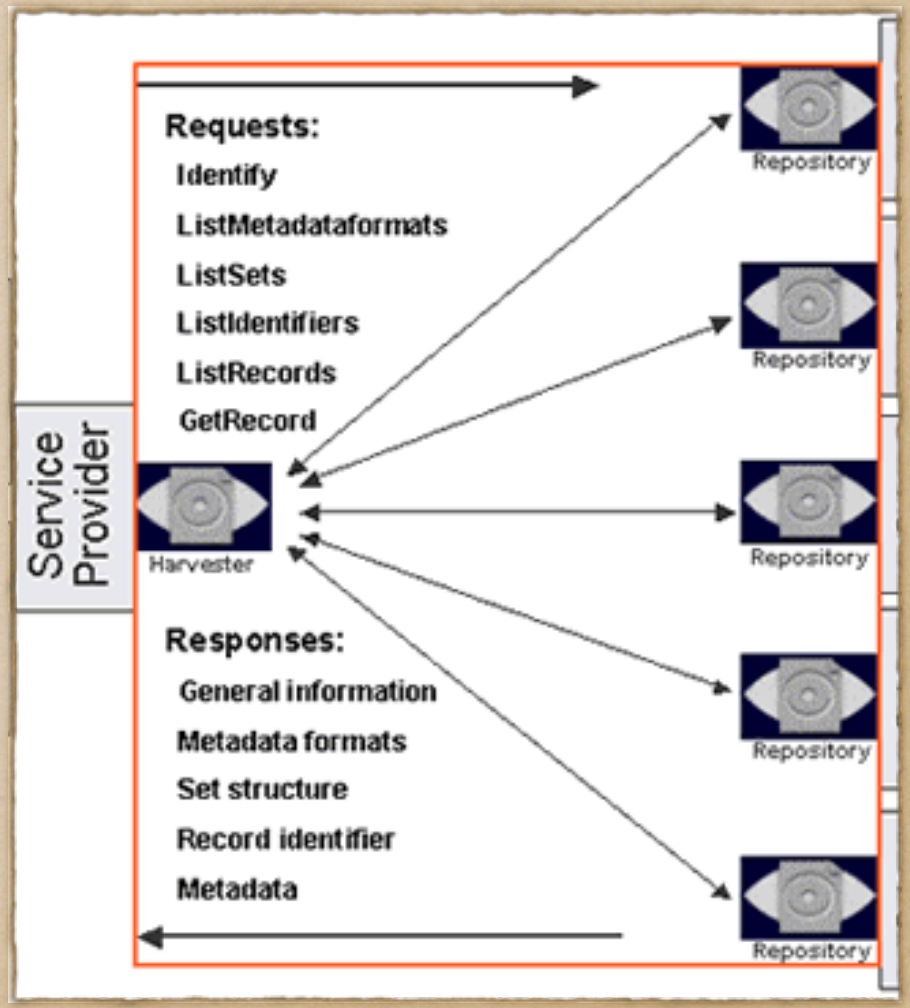
- the scope of collections;
- preservation and lifecycle management;
- the wide range of stakeholders;
- ownership of material and other legal issues; and
- cost implications

A Reliable Digital Repository

- negotiates for and accepts appropriate information from information producers and rights holders;
- obtains sufficient control of the information provided to support long-term preservation;
- determines, either by itself or with others, the users that make up its designated community, which should be able to understand the information provided;
- ensures that the information to be preserved is "independently understandable" to the designated community;
- follows documented policies and procedures that ensure the information is preserved against all reasonable contingencies and enables the information to be disseminated as authenticated copies of the original or as traceable to the original;
- ·makes the preserved information available to the designated community; and works closely with the repository's designated community to advocate the use of standard practice in the creation of digital resources, including outreach programs for potential depositors.

OAI Protocol for Metadata Harvesting

The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) is a protocol developed by the Open Archives Initiative. It is used to collect the metadata descriptions of the records in an archive so that services can be built using metadata from many archives. An implementation of OAI-PMH must support representing metadata in Dublin Core, but may also support additional representations. OAI-PMH uses XML over HTTP. Version is 2.0 was updated in 2008.



Open Source

Open source software is made available under a license — there are many variations — that allows the software to be used free of charge and makes its source code available for modification and improvement, provided that the improvements revert to the developer community. Open source systems predominate in this area of computing, largely because commercial developers have not viewed it as sufficiently lucrative

Key Factors in Adoption and Use

- Compliance with Technical Standards, including OAIS and OAI-PHM
- Workflows
- Import/Export Capabilities
- Extensibility
- Presentation Modes

Cost Factors

- **Labor** — setup/startup time is relatively short, but collection building is a labor-intensive process
- **Computer** — a small-to-medium sized Intel-based server is most appropriate
- **Disk Storage & Backup** — storage costs are low and continue to drop, but the storage system should be wholly redundant
- **Other software** — For collections that entail audio, video, and/or image files, utilities supporting the processing, conversion, etc., of such files will be needed; open source options are generally available, but not always sufficient

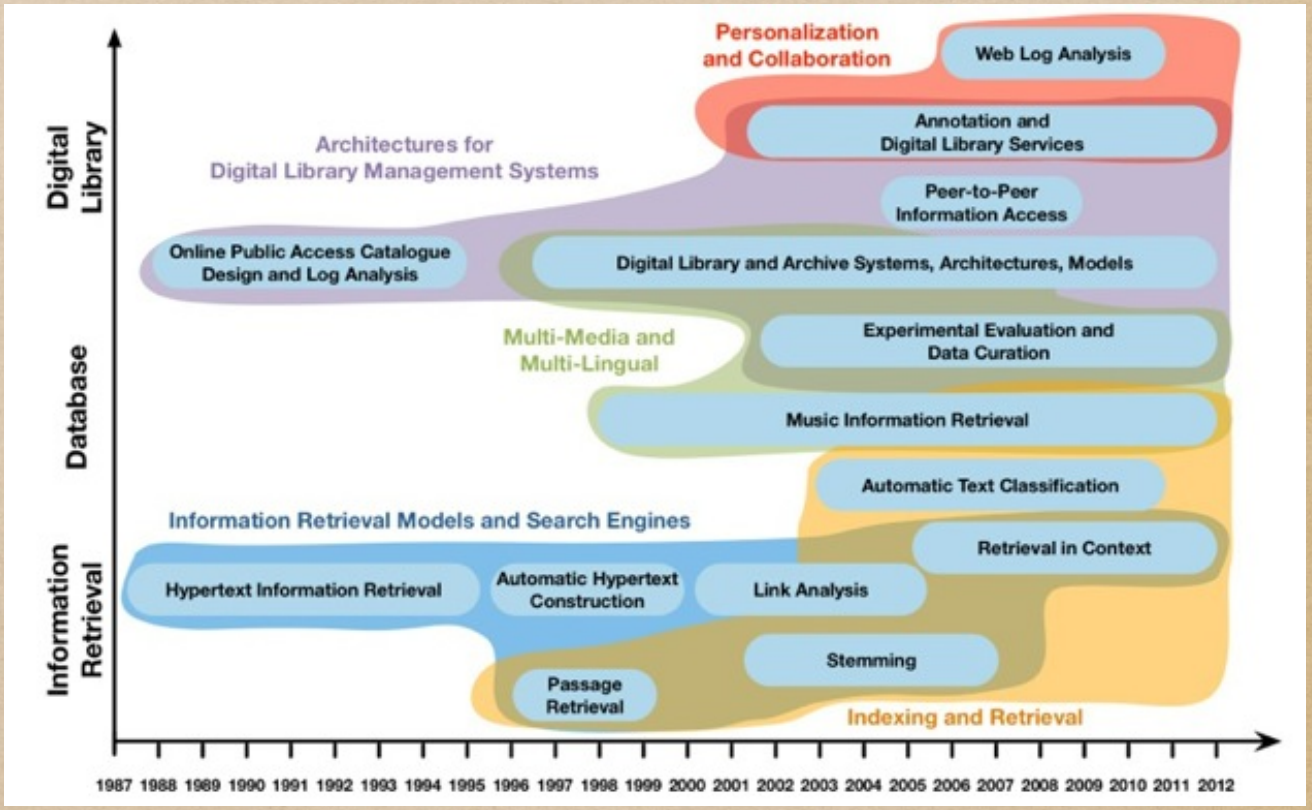
Costs of External Support

- **DuraCloud** — basic plan is $1800 a year, with 1TB storage
- **omeka.net** — plans range from $50-999 a year, with 1-25GB storage and support for unlimited sites, plugins, and themes on the high end
- **ePrints Services** — prices on request

Key Open Source Systems

Digital Archive

The digital archive is distinguished by the use of metadata schemes that have been developed within the archival community and by an emphasis on the collection, as opposed to discrete items, as the primary structure for organization and presentation.



Digital Asset Management

Digital asset management systems tend to place less emphasis on the support for formally defined metadata schemes, often relying instead on more ad hoc approaches to description, and more emphasis on the presentation of content.

Digital Repositories

Digital repositories are systems designed to store and provide access to digital objects of various types, and they are distinguished by the use of standardized metadata schemes, usually based on the Dublin Core elements, as a basis for description and information retrieval, and comparatively more sophisticated workflows.

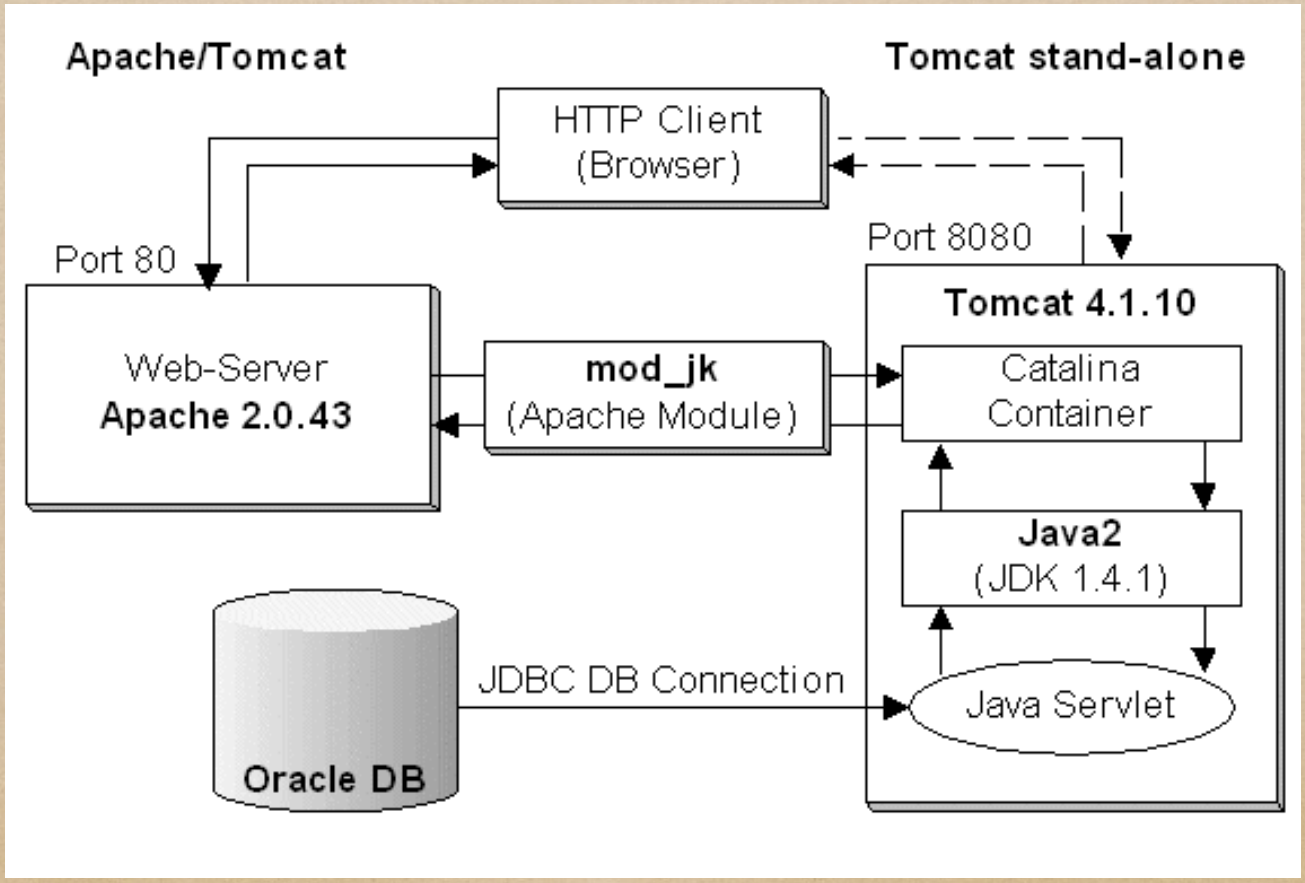
Of the current systems, ePrints is the most popular, because it supports self-archiving. DSpace is also popular. Its customizable workflow, increasing extensibility, and support for standardized formatting and packaging are main reasons.

L-A-M-P Model

The LAMP model is the most commonly used configuration, owing to the ease with which the underlying database and the Web server may integrated.

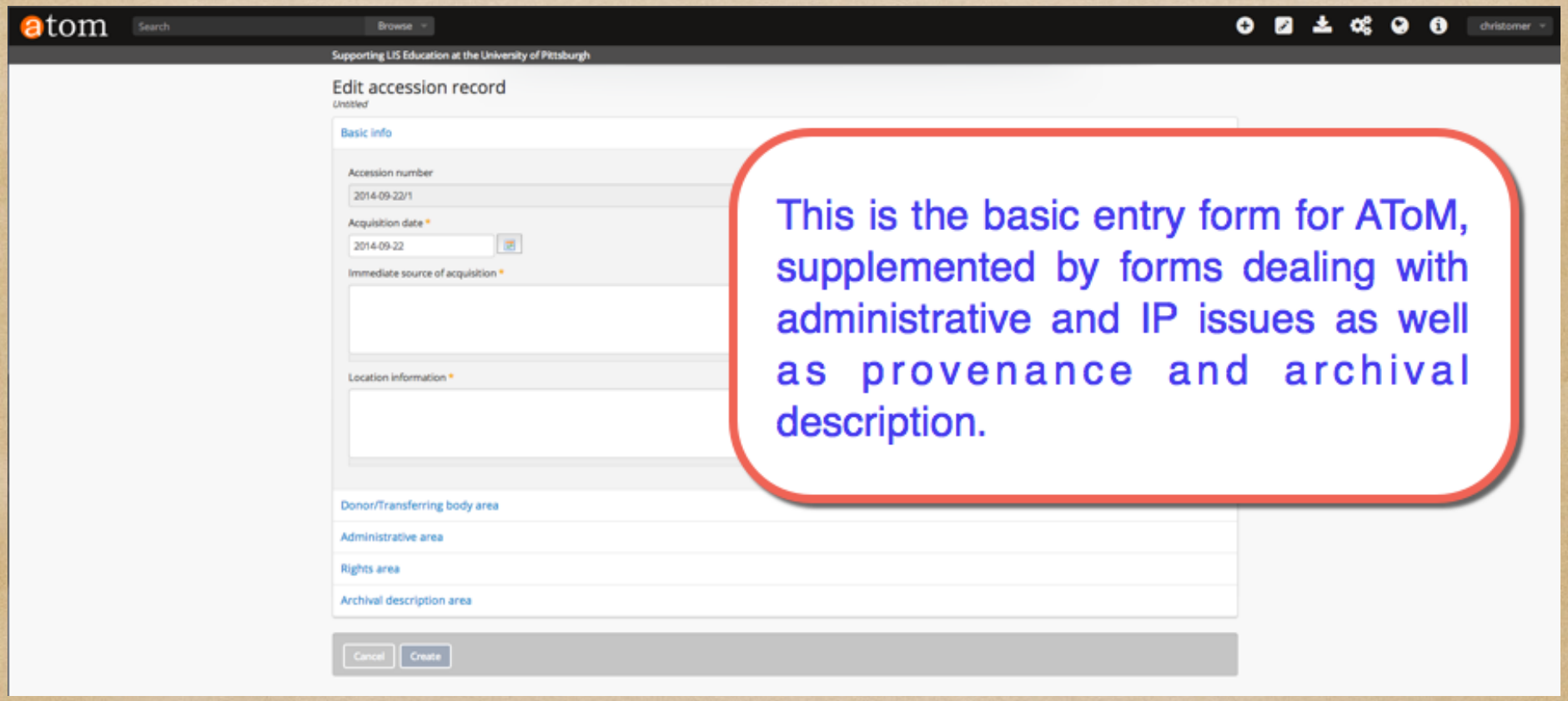
Apache Tomcat

Like LAMP configurations, Apache Tomcat may be configured to work with open source RDMS such as MySQL and PostgreSQL. Under heavier loads, Tomcat configurations are thought to be more stable

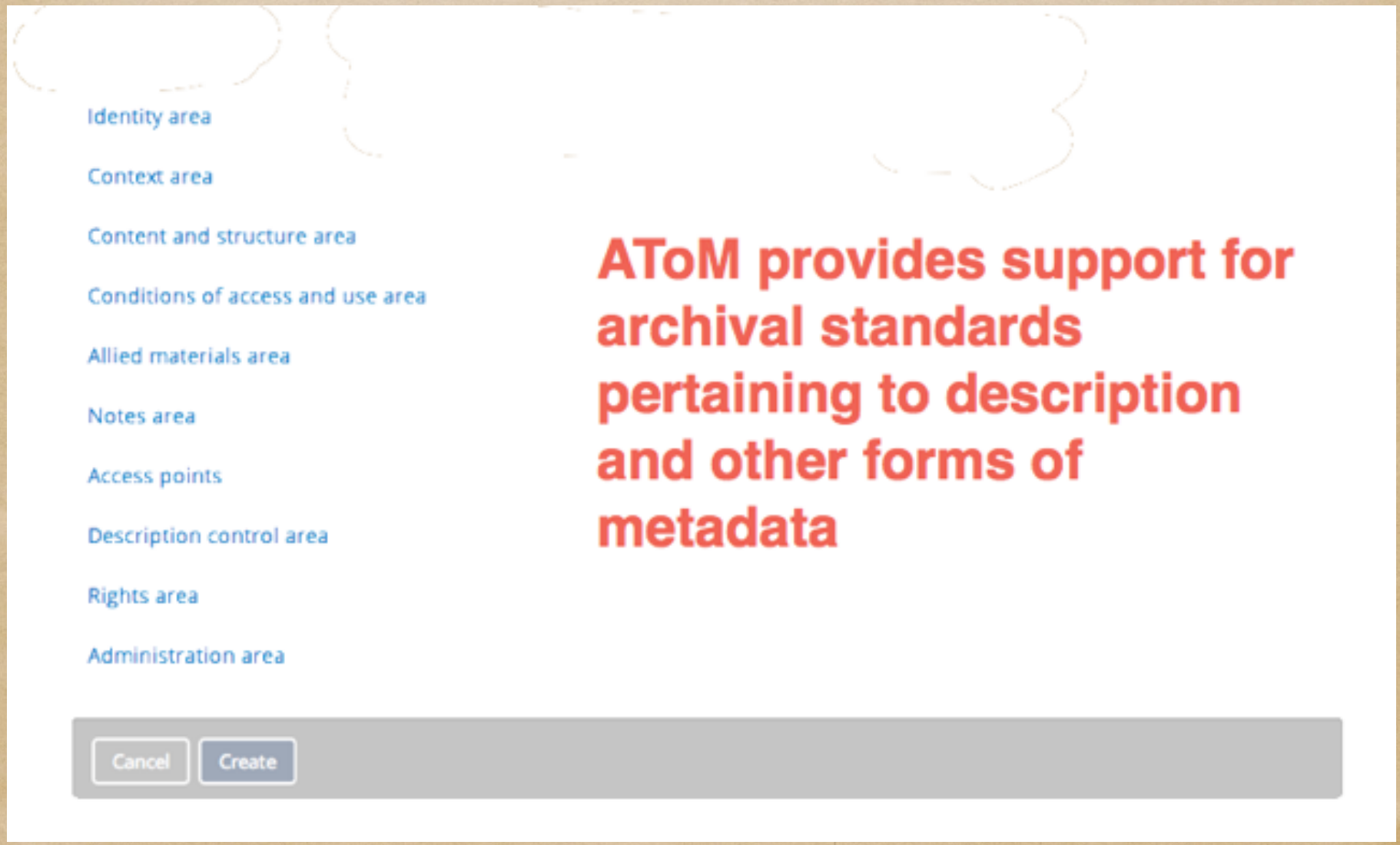


AToM — Access to Memory

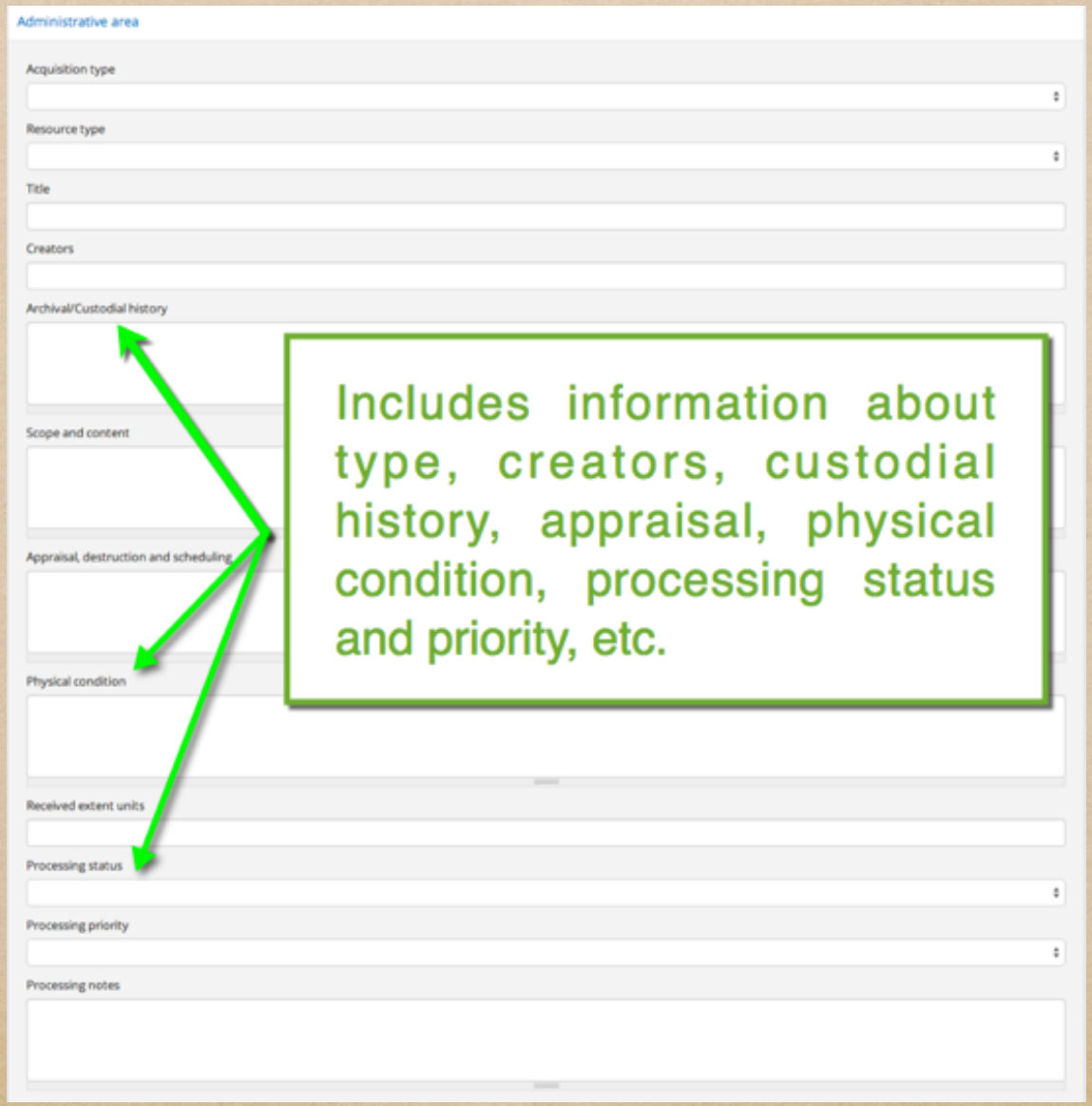
Basic Entry under AToM



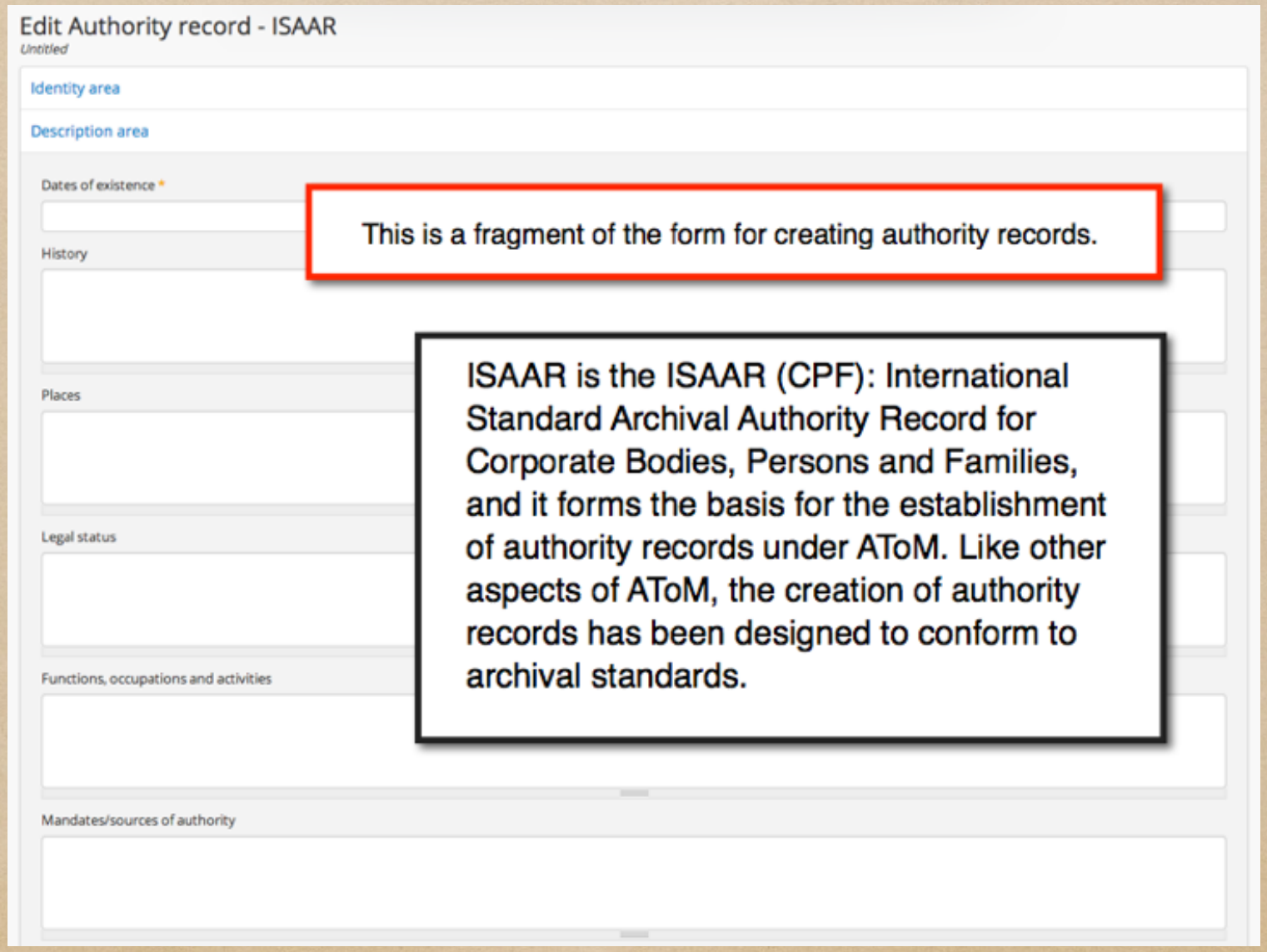
Support for Archival Standards



Template for AtoM's Administrative Area

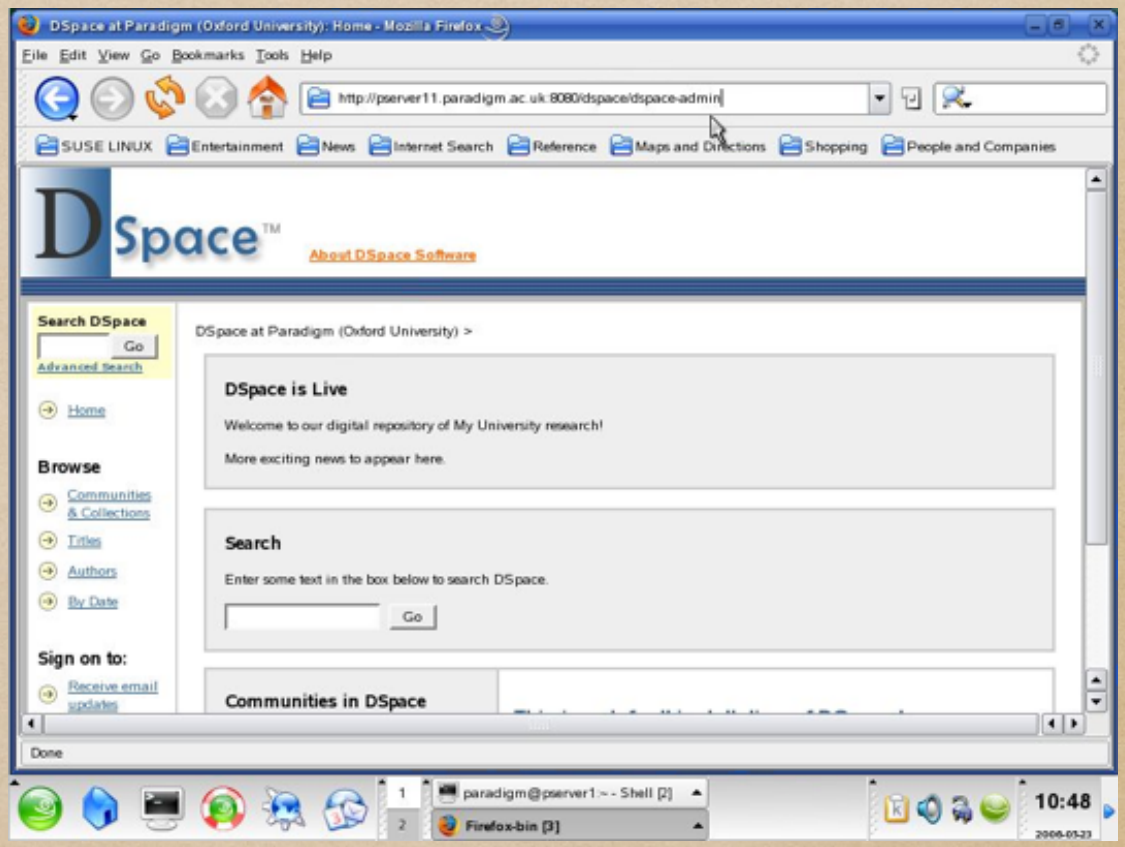


Authority Records under AToM



DSpace

DSpace, which began as a collaboration between MIT and Hewlett-Packard in 2002 and is now maintained by DuraSpace, supports key standards for metadata, packaging, and import/export.



DSpace Technology

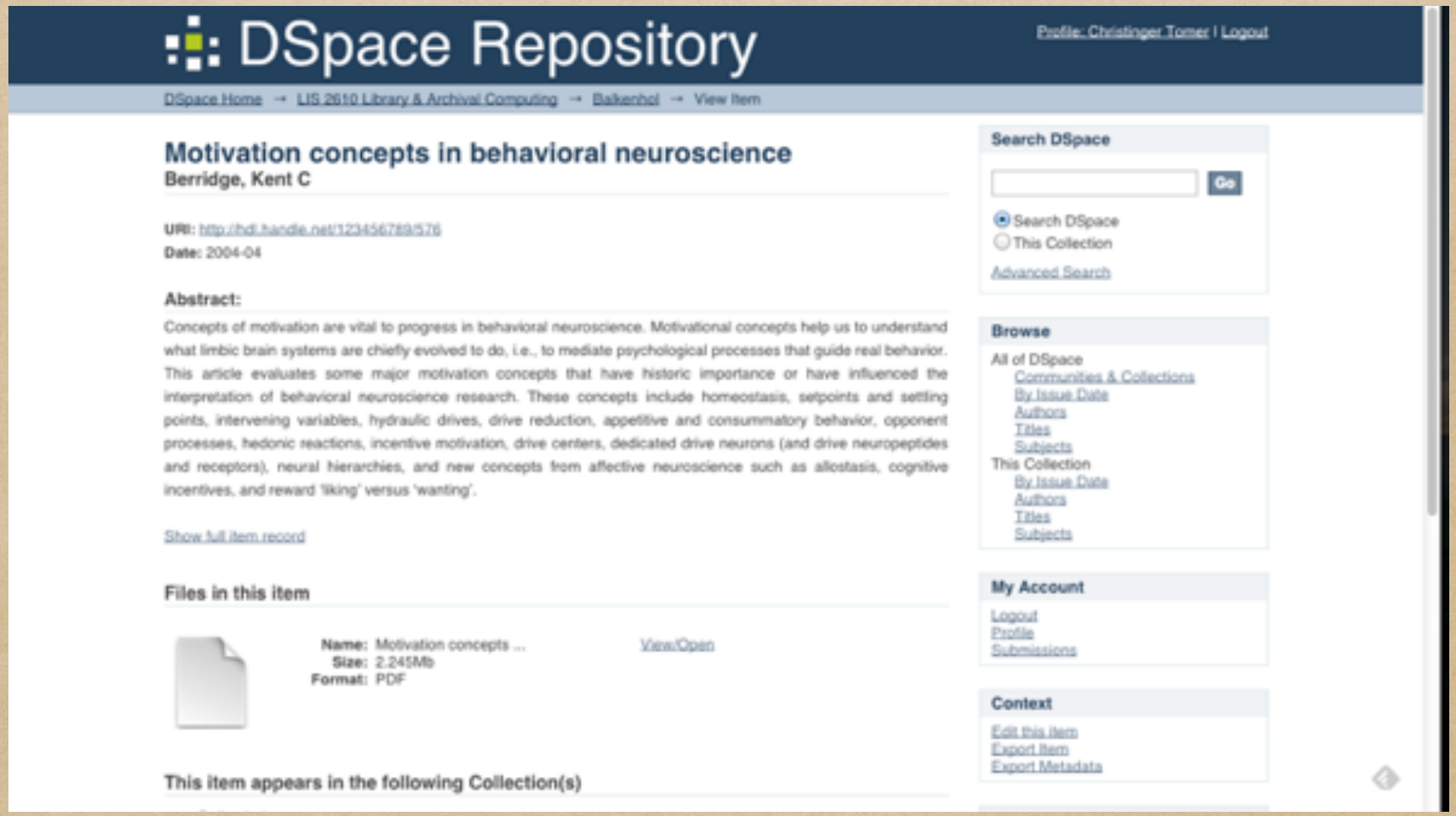
DSpace is a set of cooperating Java web applications and utility programs that maintain an asset store and an associated metadata store. The Web applications provide interfaces for administration, deposit, ingest, search and access. The asset store is maintained on a file system or similar storage system. The metadata, including access and configuration information is stored in a relational database and supports the use of PostgreSQL and Oracle database. DSpace currently support two primary web interfaces: JSPUI which uses JSP and the Java Servlet API and XMLUI (aka Manakin), which is based on Apache Cocoon, using XML and XSLT.

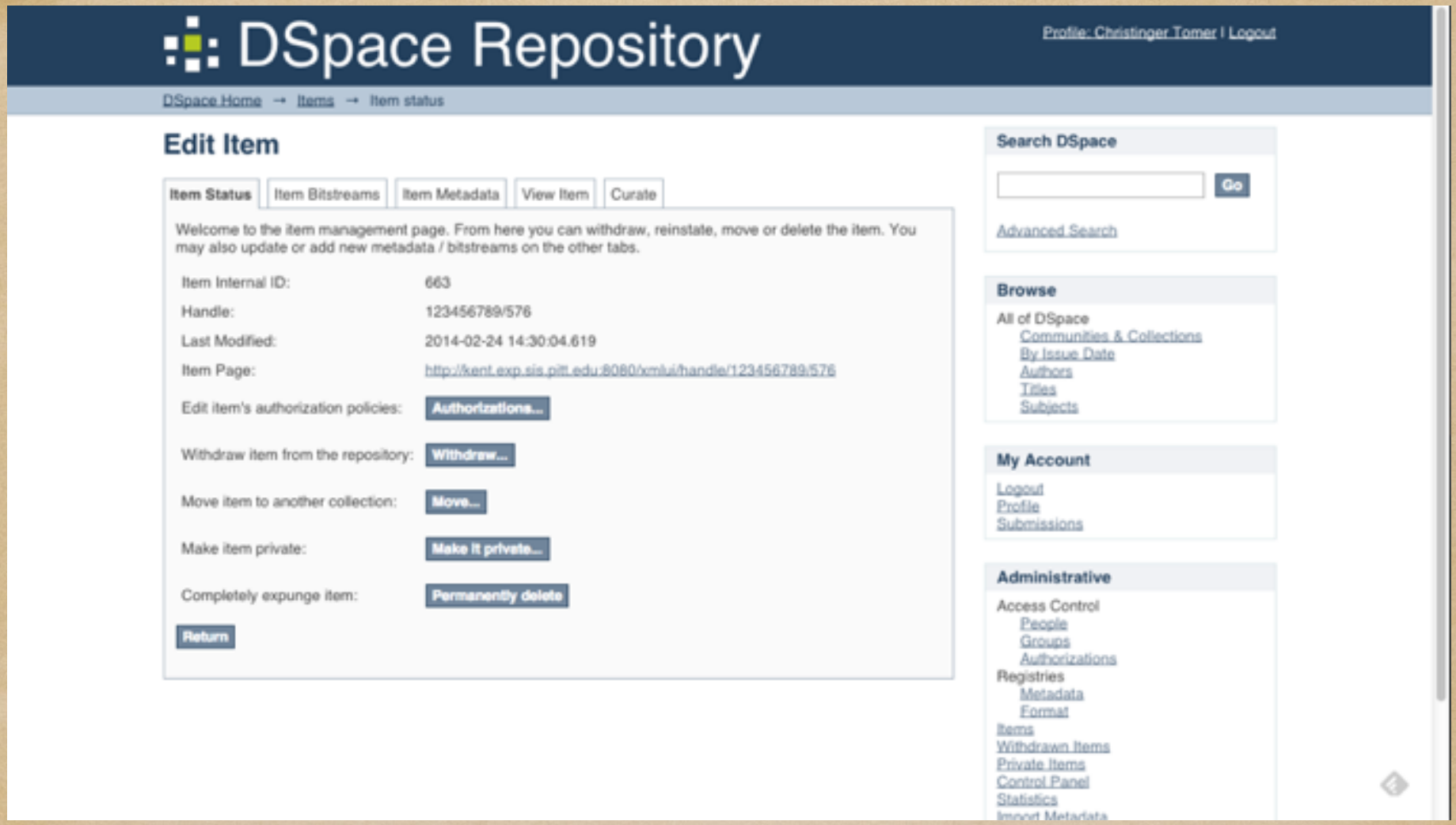
DSpace holdings are made available primarily via a web interface, but it also supports the OAI-PMH v2.0, and is capable of exporting METS (Metadata Encoding and Transmission Standard) packages. DSpace supports the common interoperability standards used in the Institutional repository domain, such as Open Archives Initiative Protocol for Metadata Harvesting, SWORD, OpenSearch, and RSS. More recent versions of DSpace also support faceted search and browse functionality using Apache Solr.

DSpace, with XMLUI

The DSpace interface can be enhanced; the most popular enhancement, also known as "Manakin," was developed at Texas A&M.

Standard Entry under DSpace





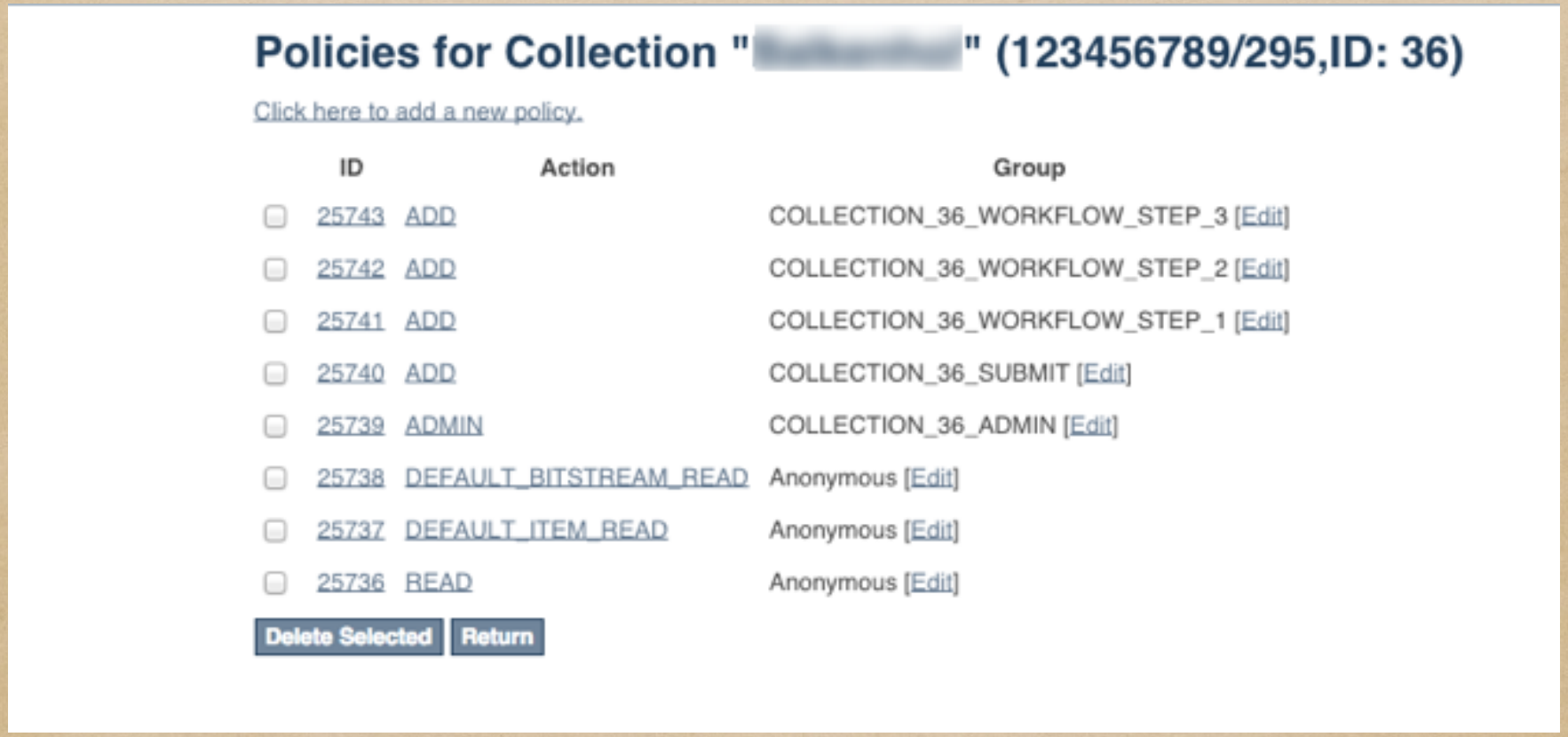
Editing a DSpace Item

Metadata under DSpace

Metadata

Remove	Name	Value	Language
☐	dc.contributor.author	Berridge, Kent C	
☐	dc.date.accessioned	2014-02-24T19:30:04Z	
☐	dc.date.available	2014-02-24T19:30:04Z	
☐	dc.date.issued	2004-04	
☐	dc.description.abstract	Concepts of motivation are vital to progress in behavioral neuroscience. Motivational concepts help us to understand what limbic brain systems are chiefly evolved to do, i.e., to mediate psychological processes that guide real behavior. This article evaluates some major motivation	en_US
☐	dc.description.provenance	Submitted by [redacted] (mab384@pitt.edu) on 2014-02-24T06:12:02Z No. of bitstreams: 1 Motivation concepts in behavioral neuroscience.pdf: 2354677 bytes,	en
☐	dc.description.provenance	Approved for entry into archive by Michael Balkenhol (mab384@pitt.edu) on 2014-02-24T19:01:06Z (GMT) No. of bitstreams: 1 Motivation concepts in behavioral neuroscience.pdf: 2354677 bytes, checksum: e3fad1d2aa9a9967b9ae9d26b6201fdd (MD5)	en

Default Policies Set



Policies for Collection "[redacted]" (123456789/295,ID: 36)

Click here to add a new policy.

	ID	Action	Group
☐	25743	ADD	COLLECTION_36_WORKFLOW_STEP_3 [Edit]
☐	25742	ADD	COLLECTION_36_WORKFLOW_STEP_2 [Edit]
☐	25741	ADD	COLLECTION_36_WORKFLOW_STEP_1 [Edit]
☐	25740	ADD	COLLECTION_36_SUBMIT [Edit]
☐	25739	ADMIN	COLLECTION_36_ADMIN [Edit]
☐	25738	DEFAULT_BITSTREAM_READ	Anonymous [Edit]
☐	25737	DEFAULT_ITEM_READ	Anonymous [Edit]
☐	25736	READ	Anonymous [Edit]

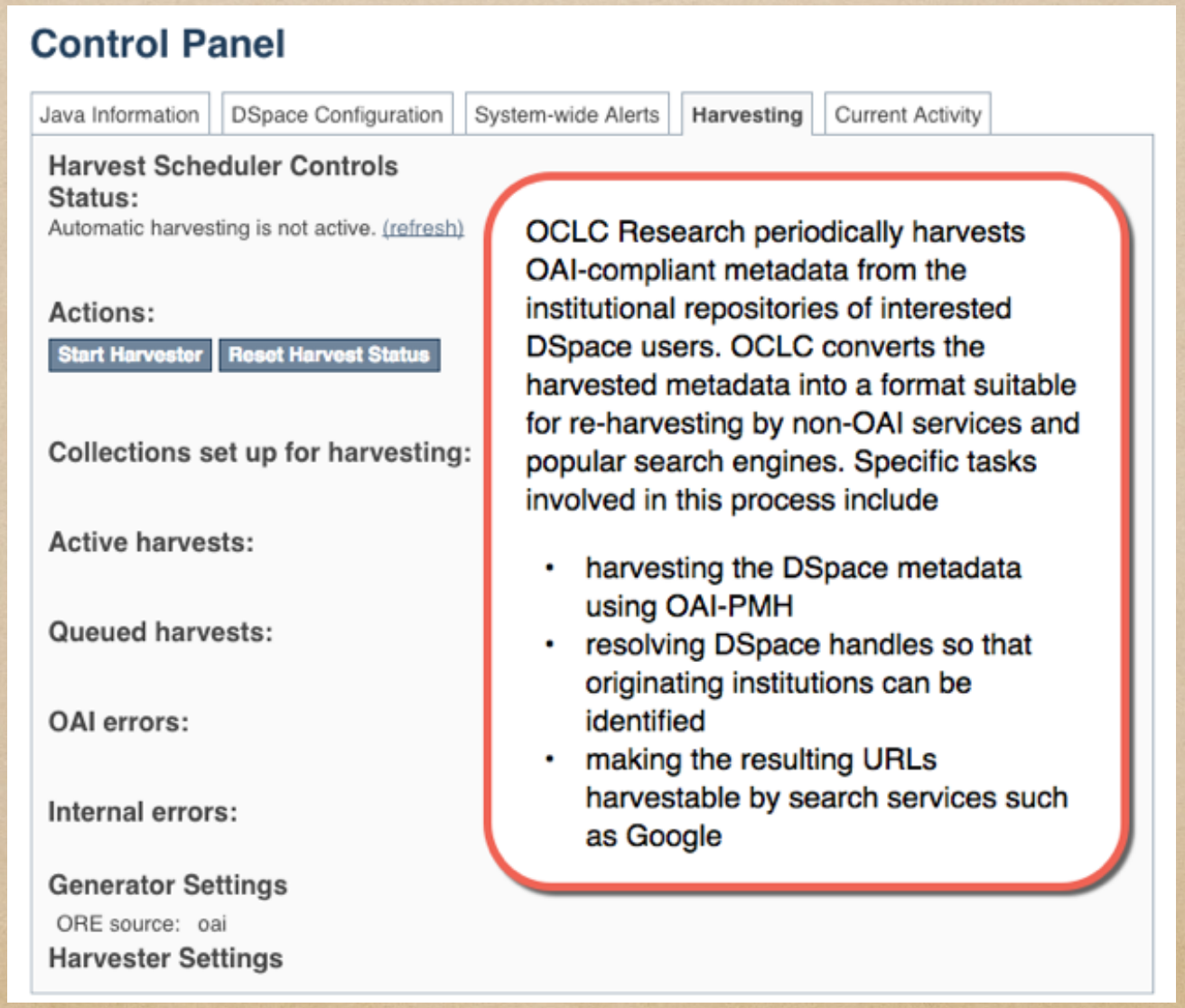
Delete Selected | Return

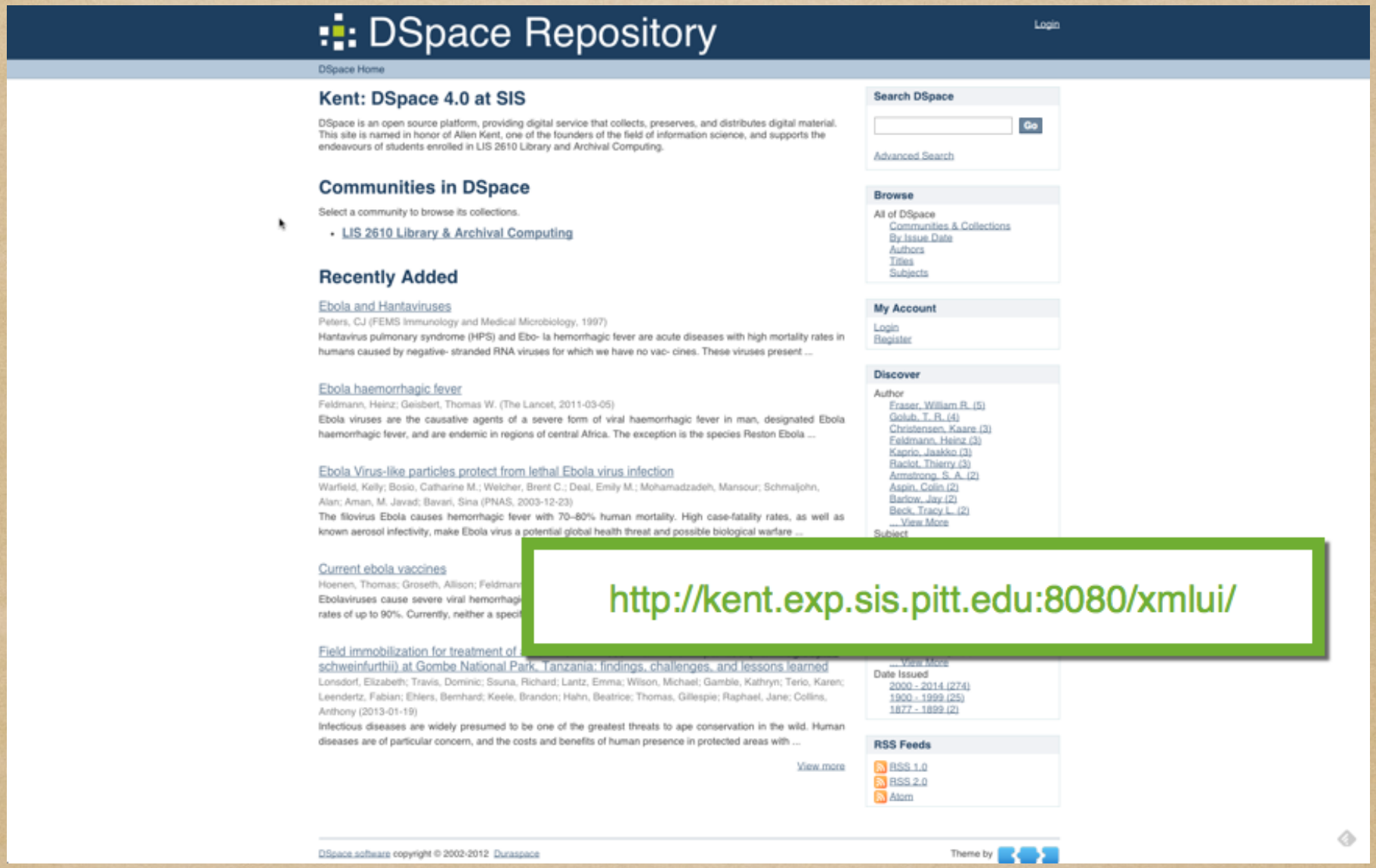
Metadata Fields

Schema metadata fields

	ID	Field	Scope Note
☐	2	dc.contributor.advisor	Use primarily for thesis advisor.
☐	3	dc.contributor.author	
☐	4	dc.contributor.editor	
☐	5	dc.contributor.illustrator	
☐	6	dc.contributor.other	
☐	1	dc.contributor	A person, organization, or service responsible for the content of the resource. Catch-all for unspecified contributors.
☐	7	dc.coverage.spatial	Spatial characteristics of content.
☐	8	dc.coverage.temporal	Temporal characteristics of content.
☐	9	dc.creator	Do not use; only for harvested metadata.
☐	11	dc.date.accessioned	Date DSpace takes possession of item.
☐	12	dc.date.available	Date or date range item became available to the public.
☐	13	dc.date.copyright	Date of copyright.
☐	14	dc.date.created	Date of creation or manufacture of intellectual content if different from date.issued.
☐	15	dc.date.issued	Date of publication or distribution.
☐	16	dc.date.submitted	Recommend for theses/dissertations.
☐	67	dc.date.updated	The last time the item was updated via the SWORD

OAI Harvesting




DSpace Repository
Login
DSpace Home
Kent: DSpace 4.0 at SIS
DSpace is an open source platform, providing digital service that collects, preserves, and distributes digital material. This site is named in honor of Allen Kent, one of the founders of the field of information science, and supports the endeavours of students enrolled in LIS 2610 Library and Archival Computing.
Communities in DSpace
Select a community to browse its collections.
LIS 2610 Library & Archival Computing
Recently Added
Ebola and Hantaviruses
Peters, CJ (FEMS Immunology and Medical Microbiology, 1997)
Hantavirus pulmonary syndrome (HPS) and Ebo- la hemorrhagic fever are acute diseases with high mortality rates in humans caused by negative- stranded RNA viruses for which we have no vac- cines. These viruses present ...
Ebola haemorrhagic fever
Feldmann, Heinz; Geisbert, Thomas W. (The Lancet, 2011-03-05)
Ebola viruses are the causative agents of a severe form of viral haemorrhagic fever in man, designated Ebola haemorrhagic fever, and are endemic in regions of central Africa. The exception is the species Reston Ebola ...
Ebola Virus-like particles protect from lethal Ebola virus infection
Warfield, Kelly; Bosio, Catharine M.; Welcher, Brent C.; Deal, Emily M.; Mohamadzadeh, Mansour; Schmaljohn, Alan; Aman, M. Javad; Bavari, Sina (PNAS, 2003-12-23)
The filovirus Ebola causes hemorrhagic fever with 70–80% human mortality. High case-fatality rates, as well as known aerosol infectivity, make Ebola virus a potential global health threat and possible biological warfare ...
Current ebola vaccines
Hoenen, Thomas; Groseth, Allison; Feldmann
Ebolaviruses cause severe viral hemorrhagi
rates of up to 90%. Currently, neither a speci
Field immobilization for treatment of
schweinfurthii) at Gombe National Park, Tanzania: findings, challenges, and lessons learned
Lonsdorf, Elizabeth; Travis, Dominic; Ssuna, Richard; Lantz, Emma; Wilson, Michael; Gamble, Kathryn; Terio, Karen; Leendertz, Fabian; Ehlers, Bernhard; Keele, Brandon; Hahn, Beatrice; Thomas, Gillespie; Raphael, Jane; Collins, Anthony (2013-01-19)
Infectious diseases are widely presumed to be one of the greatest threats to ape conservation in the wild. Human diseases are of particular concern, and the costs and benefits of human presence in protected areas with ...
View more
Search DSpace
Go
Advanced Search
Browse
All of DSpace
Communities & Collections
By Issue Date
Authors
Titles
Subjects
My Account
Login
Register
Discover
Author
Fraser, William R. (5)
Golub, T. R. (4)
Christensen, Kaare (3)
Feldmann, Heinz (3)
Kaprio, Jaakko (3)
Raclot, Thierry (3)
Armstrong, S. A. (2)
Aspin, Colin (2)
Barlow, Jay (2)
Beck, Tracy L. (2)
... View More
Subject
... View More
Date Issued
2000 - 2014 (274)
1900 - 1999 (25)
1877 - 1899 (2)
RSS Feeds
RSS 1.0
RSS 2.0
Atom
DSpace software copyright © 2002-2012 Duraspace
Theme by
http://kent.exp.sis.pitt.edu:8080/xmlui/

Omeka

Omeka mimics a development model that was devised by a number of open source content management systems, most notably Drupal, embracing a more modular approach to design and development.

Omeka surrounds a core set of functions and features with extensions that may be installed and used at the discretion of the system's administrators, and some of which may be deployed on a collection-by-collection basis.

Omeka Entry for an Image

Omeka at SIS

SEARCH

Browse Items | Browse Collections | Browse Exhibits | About | Collection Tree | Map

IMAGE FOR TIBBO'S "PLACING THE HORSE BEFORE THE CART: CONCEPTUAL AND TECHNICAL DIMENSIONS OF DIGITAL CURATION"

Collection

Embree, Digital Curation Images

Citation

The Phrase Finder, "Image for Tibbo's "Placing the Horse before the Cart: Conceptual and Technical Dimensions of Digital Curation"," *Omeka at SIS*, accessed September 21, 2014, http://50.17.193.184/omeka/items/show/1093.

Dublin Core

Title

Image for Tibbo's "Placing the Horse before the Cart: Conceptual and Technical Dimensions of Digital Curation"

Creator

The Phrase Finder

Source

http://www.phrases.org.uk/meanings/put-the-cart-before-the-horse.html

Format

JPEG Image

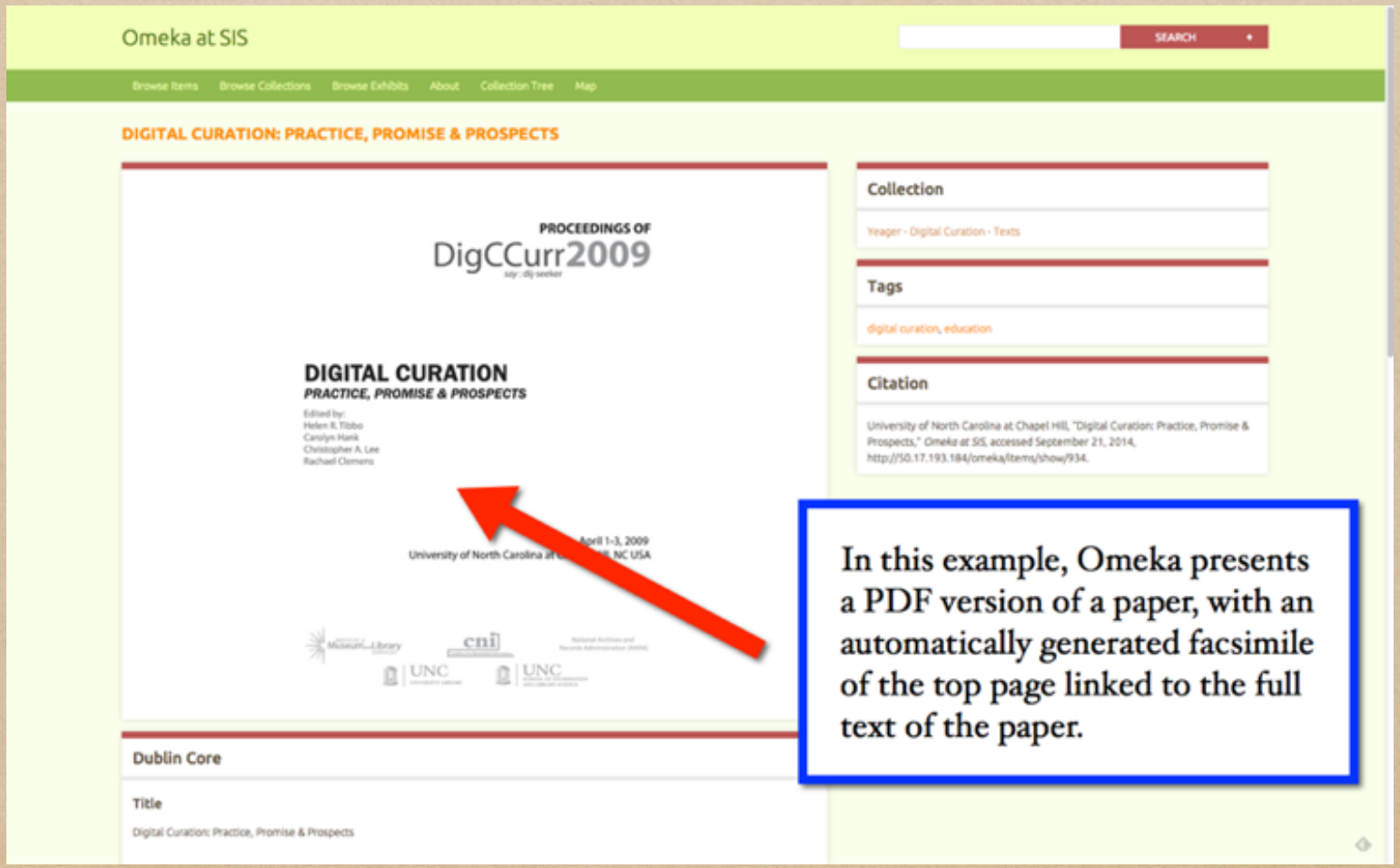
Embed

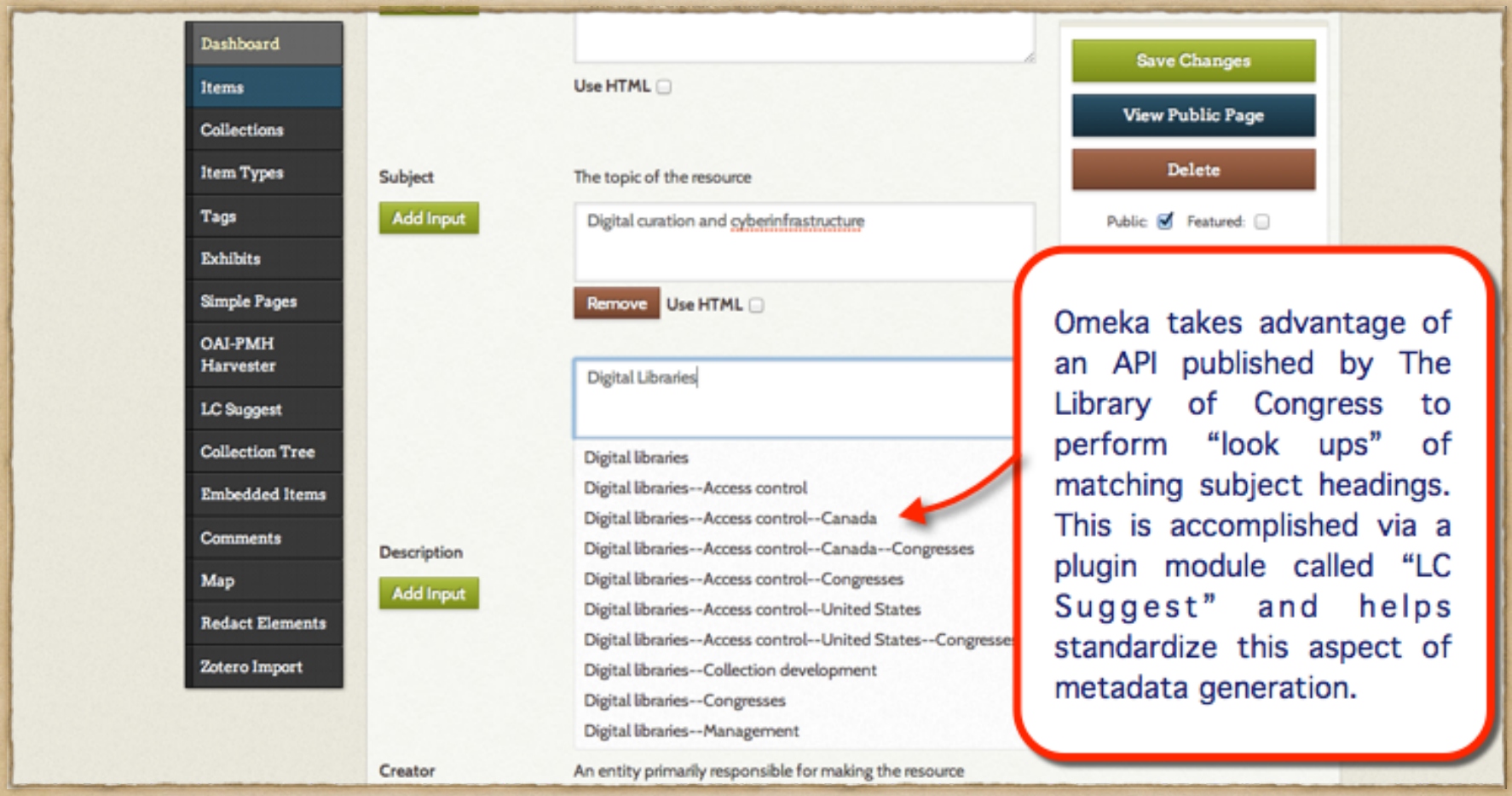
Copy the code below into your web page

```
<iframe class='omeka-embed' src='http://50.17.193.184/omeka/items/embed/1093' width='560px'
height='315px' frameborder='0' allowfullscreen></iframe>
```

This is a standard output for an entry under Omeka. Please note that it includes a description based on the Dublin Core, a citation in a standard format, and embedding code for transcluded use.

Omeka Entry for a PDF





Omeka's LC Suggest Module

http://50.17.193.184/omeka/

Omeka at SIS

SEARCH

Browse Items | Browse Collections | Browse Exhibits | About | Collection Tree | Map

Featured Item

EDUCATORS NEED TO BECOME DIGITAL CURATORS

The image displays different lenses indicating that educators are now required to become digital curators of a wide scope of disciplines throughout…

Featured Collection

RHOCK-DIGITAL IMAGES

Featured Exhibit

TESTING OMEKA

Recently Added Items

IMAGE FOR TIBBO'S "PLACING THE HORSE BEFORE THE CART: CONCEPTUAL AND TECHNICAL DIMENSIONS OF DIGITAL CURATION"

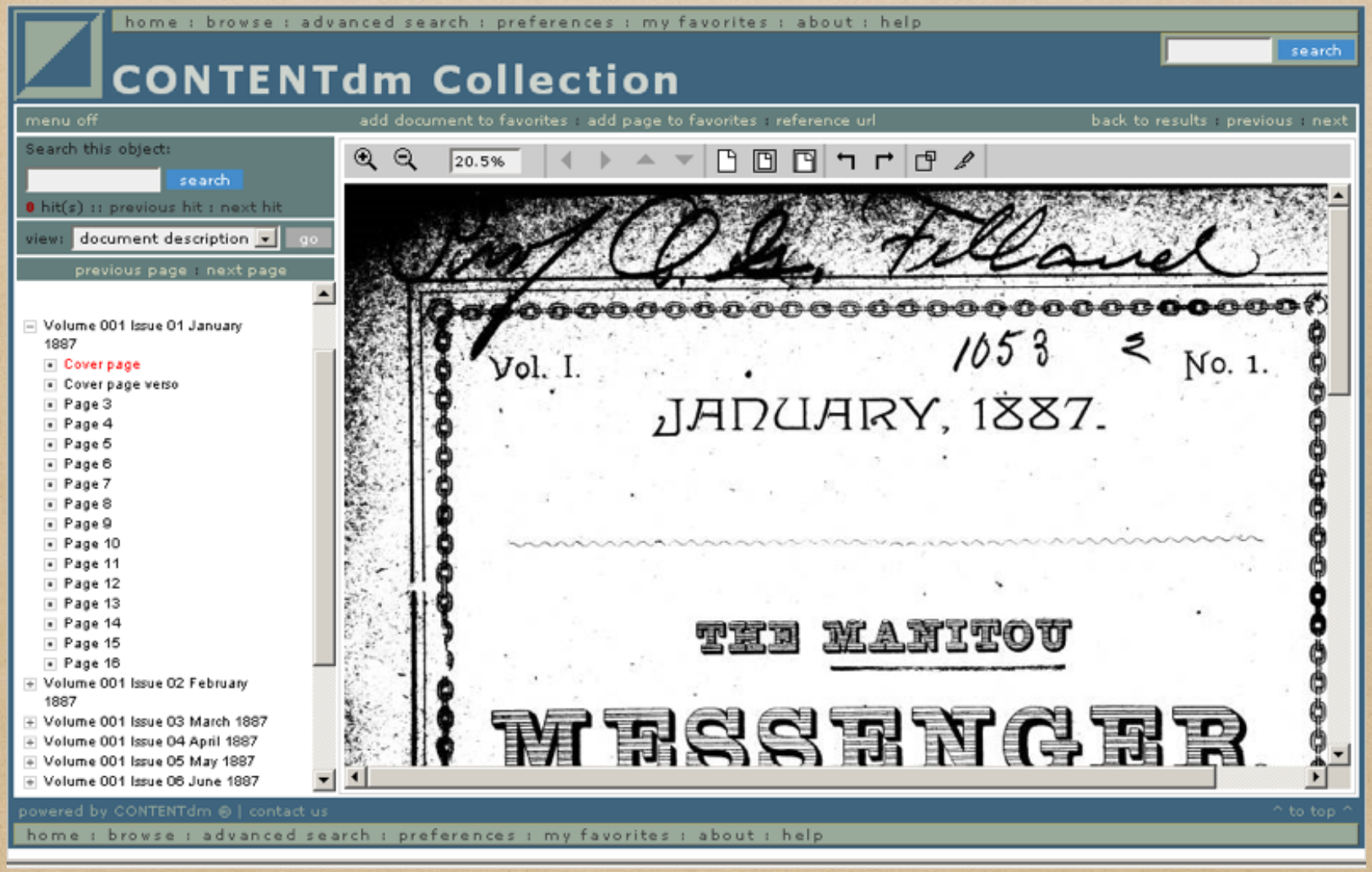
IMAGE FOR DIGITAL CURATION FOR DIGITAL NATIVES

ePrints

- EPrints is a Web and command-line application based on the LAMP architecture (but is written in Perl rather than PHP). It has been successfully run under Linux, Solaris and Mac OS X. A version for Microsoft Windows was released 17 May 2010.
- Version 3 of the software introduced a (Perl-based) plugin architecture for importing and exporting data, converting objects (for search engine indexing) and user interface widgets. Configuring an EPrints repository involves modifying configuration files written in Perl or XML. The appearance of a repository is controlled by HTML templates, CSS stylesheets and inline images. While EPrints is shipped with an English translation it has been translated to other languages through (redistributable) language-specific XML phrase files. Translations include Bulgarian, French, German, Hungarian, Italian, Japanese, Russian, Spanish and Ukrainian.

Key Proprietary Systems

CONTENTdm


home : browse : advanced search : preferences : my favorites : about : help
search
CONTENTdm Collection
menu off
add document to favorites : add page to favorites : reference url
back to results : previous : next
Search this object:
search
0 hit(s) :: previous hit : next hit
view: document description
go
previous page : next page
Volume 001 Issue 01 January 1887
Cover page
Cover page verso
Page 3
Page 4
Page 5
Page 6
Page 7
Page 8
Page 9
Page 10
Page 11
Page 12
Page 13
Page 14
Page 15
Page 16
Volume 001 Issue 02 February 1887
Volume 001 Issue 03 March 1887
Volume 001 Issue 04 April 1887
Volume 001 Issue 05 May 1887
Volume 001 Issue 06 June 1887
20.5%
Vol. I.
No. 1.
JANUARY, 1887.
THE MANITOU
MESSENGER
powered by CONTENTdm | contact us
^ to top ^
home : browse : advanced search : preferences : my favorites : about : help

Key Resources & Sites

- AtoM — https://www.accesstomemory.org/en/
- CONTENTdm
- DSpace — http://www.dspace.org/
- ePrints — http://www.eprints.org/software/
- Omeka — http://omeka.org/


ANY
QUESTIONS
?