



Low Resource and Loosely Coupled Archives

Hussein Suleman <hussein@cs.uct.ac.za> Digital Libraries Lab, Department of Computer Science, School of IT University of Cape Town 10 July 2019

Low Resource Archives



Digital Libraries

- A Digital Repository/Archive stores and provides access to digital objects/metadata.
 - Institutional repositories, heritage collections, etc.
- Digital Libraries is the meta-discipline that studies the development of digital analogues to non-digital libraries.
 - Usually through the creation and management of digital repositories/archives.



Major Goals

Preservation

- Maintaining digital collections in perpetuity.
- Active curation of software/hardware/data.
- Hard? Is digital really durable?

Engagement

- Connecting people with resources.
- Use and contribution/enrichment.
- Hard? People cannot find stuff.



Typical Services of Digital Repositories

- Store documents and metadata
- Search and Browse
- Submission of documents and metadata
- Moderation of content
- Access through Web interface
- Compliance with standards
- Enrichment: comments, reviews, etc.
- Linking with other systems



African Problems 1/3 Skills and Education



- Typical archivists are not as highly skilled as counterparts elsewhere.
- Digital media is still not the norm.
- Education levels of general population hinders preservation – end-user data curation is very difficult.



African Problems 2/3 Funding



- Typically, there is little.
- Many preservation projects are funded by external agencies, but with restrictions on data accessibility.
- There is a desperate need to do more with less.



African Problems 3/3

Internet Bandwidth (Digital Divide)



- Non-existent in some places and poor everywhere else.
- Preservation projects designed for high bandwidth are not suitable.
- All online solutions must be bandwidthfriendly.



Principles of DL for African Heritage

- Efficient bandwidth use
- Advanced technology
- Appropriate technology
- Local relevance
- Modernization instead of Africanization
- Global applicability of solutions
- Minimalism of staff/money
- Multicultural/multilingual inclusivity



Low Resource Archives

Design Goals:

- Simplicity
- Low cost
- Portability
- Easier maintenance
- Preservation: Backup/Migration/Rescue

Compromise:

Scalability



Technology Elements

XML/XSLT

- All data stored in simple files in hierarchical directories – no database.
- Static generation
 - Site is pre-created so software is not a long-term risk.
- Offline search/browse
 - Provides services through browser-based Javascript.



Approach 1: XML + XSLT \rightarrow XHTML

- Encode all descriptive information using XML files for objects, users, etc.
- Use XSL transformations to convert the XML metadata into corresponding HTML page views.
- Also, pre-process XML to create search/browse indices.



Approach 2: in-Browser Services





department of computer science

Approach 3: Selected Online Services

- Use online Web applications to add items, comments, annotation, etc.
- Web applications must modify XML and regenerate portions of static site.
 - No dynamic page generation.



FHYA Prototype 1/3



Logged in as: Hussein Suleman (edit profile)



About FHYA | Using the FHYA site | Contributors | Search | Contact Us

Search FHYA Collections

Enter your search terms:

Browse FHYA Collections

• JAG

About FHYA

In its current form FHYA is an archival exemplar that includes a sample selection of materials pertinent to a small geographic area (KwaZulu-Natal-Swaziland region) in a limited time frame (from about 1770 to various points in the nineteenth century).

The exemplar includes a highly diverse selection of materials in different media - documents, photographs, recordings - from a few different institutions - museums, libraries, archives - as well as personal collections.

This selection demonstrates that is possible to convene online materials that have been historically separated through institutional practice.

Materials have been processed to make them readily searchable. This creates opportunities for researchers to step beyond institutional categories and make new connections between diverse materials.



department of computer science

GO

FHYA Prototype 2/3



Metadata

Title	Beadwork
	[Source of title : Nessa Leibhammer using JAG materials]
Material Designation	Object
	Textual record
Repository	Johannesburg Art Gallery (JAG)
Identifier	JL-U-8
Arrangement	[Source - Nessa Leibhammer for FHYA, 2015: Accession numbers had already been allocated to the objects in the collection before it was sold to Harry Oppenheimer and the numbering system was retained by JAG: the initials JL stand



department of computer science

FHYA Prototype 3/3



About FHYA | Using the FHYA site | Contributors | Search | Contact Us

Search Results

Query	
zulu beadwork	1
repository	
AI	D
subcollection	
JAG/BRENTHURST	0
curationactor	
A	Ð
custodyactor	
Al	D

Results

- 1. Beadwork JAG/BRENTHURST/JL-U-8.xml
- 2. Beadwork
- JAG/BRENTHURST/JL-U-166.xml 3. Beadwork
 - JAG/BRENTHURST/JL-U-42.xml
- 4. Beadwork JAG/BRENTHURST/JL-U-88.xml
- Beadwork JAG/BRENTHURST/JL-U-14.xml
- Beadwork JAG/BRENTHURST/JL-U-17.xml
- 7. Beadwork
- JAG/BRENTHURST/JL-U-19.xml 8. Staff
- JAG/BRENTHURST/JL-C-23.xml
- 9. Snuff-spoon JAG/BRENTHURST/JL-H-1.xml



Loosely-Coupled Archives



Metadata

- Metadata refers to standardised descriptions of objects, digital or physical.
- Most digital repositories manipulate metadata records, which contain pointers to the actual data.



Types of Metadata

Descriptive

- title, author, type, format, ...
- Structural
 - part, subpart, relation, child, ...
- Administrative
 - Iocation, identifier, submitter, ...
- Preservation
 - resolution, capture device, watermark, ...
- Provenance
 - source archive, previous version, source format, ...





DC Metadata in Valid Qualified XML

<oaidc:dc xmlns="http://purl.org/dc/elements/1.1/"
xmlns:oaidc="http://www.openarchives.org/OAI/2.0/oai_dc/"
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">

<title>02uct1</title>

<creator>Hussein Suleman</creator>

<subject>Visit to UCT </subject>

<description>the view that greets you as you emerge from the tunnel under the
freeway - WOW - and, no, the mountain isnt that close - it just looks that way in
2-D</description>

<publisher>Hussein Suleman</publisher>

<date>2002-11-27</date>

<type>image</type>

<format>image/jpeg</format>

<identifier>http://www.husseinsspace.com/pictures/200230uct/02uct1.jpg
</identifier>

<language>en-us</language>

<relation>http://www.husseinsspace.com</relation>

<rights>unrestricted</rights>

</oaidc:dc>



Global Identifiers

- Identifier is a unique reference to a digital object.
 - Can be used to access object online.
 - Can be used as a reference.
 - Can be used for disambiguation.
 - Usually a URI (and sometimes a URL).

Examples: DOIs, HTTP URL, etc.



Interoperability

- Interoperability is the ability of systems to work together.
- It allows connecting of repositories so researchers can find data/documents across repositories.
- 3 major aspects: standard metadata, standard network protocols and global identifiers.
- Google: no metadata, HTTP, URLs
 - Can we do better?





High-Level History of Interoperability

- □ 1993/4: Web was invented
- mid-1990s: Early digital archives emerged
- Iate-1990s: First experiments to connect archives together together into central catalogues
 - NCSTRL, NDLTD, RePEc, arXiv
- 1998: XML invented
- 1999: Santa Fe Convention
 - Representatives of academic digital libraries agreed to set up a *low-barrier* interoperability solution
- 2000: Open Archives Initiative (OAI) formed out of Santa Fe Convention
- **2**2000-2002: Large-scale interoperability experiments
- 2002: v2.0 of OAI Protocol released for public use



OAI-PMH

Open Archives Initiative (OAI)

- Organisation created to solve problems of digital library interoperability by defining simple protocols, with most success in exchange of metadata.
- Protocol for Metadata Harvesting (PMH)
 - Network protocol to transfer metadata from a source archive to a destination archive.



OAI in Practice

Multiple independent university-based and university-controlled collections of electronic documents





Harvesting vs. Federation

Competing approaches to interoperability.

- Federation is when services are run remotely on remote data (e.g., Federated searching).
- Harvesting is when data/metadata is transferred from the remote source to the destination where the services are located (e.g., Union catalogues).
- Federation requires more effort at each remote source but is easier for the local system and vice versa for harvesting.



Metadata vs. Data

- Data refers to digital objects or digital representations of objects (e.g., a PDF version of an scanned book).
- Metadata is information about the objects (e.g., title, author).

OAI focuses on metadata, with the implicit understanding that metadata usually contains useful links to the source digital objects.



Data and Service Providers

- Data Providers refer to entities who possess data/metadata and are willing to share this with others (internally or externally) via welldefined OAI protocols (e.g., repositories).
- Service Providers are entities who harvest data from Data Providers in order to provide higherlevel services to users (e.g., search engines).



HTTP and XML

- OAI-PMH is an almost stateless request/response protocol.
- Requests and responses are sent through the WWW in standard URL-encoded formats.
- Responses are well-formed XML documents.



XML Namespaces and Schema

- Consistency and data quality is ensured by using XML Schema descriptions for each possible response.
- XML Namespaces are used where necessary to clearly define which parts of the responses are actual metadata and which support the Metadata Harvesting Protocol.



OAI Record

- A record refers to an independent XML structure that may be associated with digital or physical objects.
- Records are usually associated with metadata, not data.
- OAI advocates harvesting of records, which contain metadata and additional fields to support the harvesting operation.



Sample OAI Record

(note: schema and namespaces have been left out for clarity)



Multiplicity of Metadata

- Multiple formats of metadata allowed.
- Dublin Core is mandatory.
- Any other format allowed as long as it has an XML encoding.
 - MARC (Libraries), ETDMS (Theses/Dissertations), VRA-Core (for works of art), etc.



Sets

- Protocol mechanism to allow for harvesting of sub-collections.
- Useful to separate repositories based on type of data, subject area, etc.
- May be defined by arrangement between data providers and service providers.
- Allows a repository to supply one subset of data to one set of collaborators, and another subset to another set of collaborators.



OAI Protocol for Metadata Harvesting

Service Requests

- Identify
- ListMetadataFormats
- ListSets
- GetRecord
- ListIdentifiers
- ListRecords



UNIVERSITY OF CAPE TOWN

Sample Identify - Response

Address 🙆 http://scholar.lib.vt.edu/theses/OAI2/?verb=Identify

<?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>2003-10-14T20:01:48Z</responseDate> <request verb="Identify">http://scholar.lib.vt.edu/theses/OAI2/</request> - <Identify> <repositoryName>VT Electronic Thesis and Dissertation Archive</repositoryName> <baseURL>http://scholar.lib.vt.edu/theses/OAI2/</baseURL> <protocolVersion>2.0</protocolVersion> <adminEmail>mailto:webmaster@scholar.lib.vt.edu</adminEmail> <earliestDatestamp>1970-01-01T00:00:00Z</earliestDatestamp> <deletedRecord>no</deletedRecord> <granularity>YYYY-MM-DD</granularity> + <description> - <description> - <eprints xmlns="http://www.openarchives.org/OAI/1.1/eprints"</p> xsi:schemaLocation="http://www.openarchives.org/OAI/1.1/eprints http://www.openarchives.org/OAI/1.1/eprints.xsd*> - <content> <text>Theses and Dissertations produced by students</text> </content> - <metadataPolicy> <text>Metadata may be used by commercial and non-commercial users</text>

</metadataDolicy>



department of computer science

Advantages of OAI-PMH

- Lightweight protocol not network-intensive.
- Relatively easy to implement (8 hours?).
 - Lots of software.
- Very large community all modern archive software supports it.
- Allows any metadata.
- Allows archives to participate in multiple groups.



Global ETD Search

re search.ndltd.org

\$ 1



Global ETD Search

Search the 5,767,606 electronic theses and dissertations contained in the NDLTD archive:

Type something to start searching...

Q

advanced search tips 🗙 how to contribute records 🗲



department of computer science

Global ETD Search

Search New Search New Search

NDLTD

Refine Query	
hussein suleman	
	Apply
Source	
Filter by source	\$
Publication year	
1995 to	2002
Language	
🗆 English	58
🗆 French	5
Portuguese	5

Search results

Showing 1 to 10 of 139 (0.096 seconds) Spelling suggestions: "husserl suleiman" "husserl sulene"

Open Digital Libraries

Suleman, Hussein 26 November 2002 🔗 (has links)

Digital Libraries (DLs) are software systems specifically designed to assist users in information seeking activities. Stemming from intersection of library sciences and computer networking, traditional DL systems impose library philosophies of structure and ma on the sprawling collections of data that are made possible through the Internet. DLs evolve to keep pace with innovation on the there is little standardization in the architecture of such systems. However, in attempting to provide users with the highest possils service with the minimum possible effort, many systems work collaboratively with others, e.g., meta-search engines. This type of interoperability is encouraged by the emergence of simple data transfer protocols such as the Open Archives Initiative?s Protoco Metadata Harvesting (OAI-PMH). Open Digital Libraries are an extension of the work of the OAI. It is proposed in this dissertation philosophy and approach adopted by the OAI can easily be extended to support inter-component interaction within a component

✓ Read more

🗞 open archive 🕞 component 🕞 system architecture 🕞 digital library 🕞 interoperability 🕞 protocol

Mohammed Taha Hussein : ein künstlerischer Dialog zwischen Orient und Okzident / Thesing, Dagmar. January 2003 & (has links)

Thesis Ph. D--Universität, Köln, 2003. / Bibliogr. p. 233-242.

🗣 Ḥusayn, Muḥammad Ṭahá,

Missiles of Terror: Hitler's and Hussein's Use of Ballistic Missiles



SPES DOWN

Global ETD Search: How It Works

- Standard metadata
 - DC, ETDMS
- Efficient way to share metadata
 - OAI-PMH
- Collect updates from around the world every 12 hours
- Discovery through search engine

NDLTD Union Ar × n	The FAIR Guidin 🗙 🛛 🔓 citing datasets 🖙 🔺 🔄 DataCite Search 🗙 🚱 UCT Str	udent Ac 🗙 📔 🄇	
union.ndltd.org/portal/		\$	
Home			
nome	About		
FORMATION	This system collects metadata records for ETDs from institutions around th	ne world	
	and aggregates them into a single collection that can then be used by serv	rice	
Submit your site	providers.		
A In 2014	If you wish to search for ETDs, you can use the NDLTD Global ETD Search.		
ADOUT			
Admin	Recent Submissions		
	 The Effect of Time Heuristic Message <i>Tue</i>, <i>18 Jun</i> 2019 03:38:41 UTC To Explore the Feasibility of the Position Deviation Conversion between Ro Translational Axis in Cone-Beam Computerized Tomography for Breast Ca <i>Tue</i>, <i>18 Jun</i> 2019 03:38:41 UTC The relationship between accountant characteristics and key audit matters company at emerging stock market board <i>Tue</i>, <i>18 Jun</i> 2019 03:38:41 UTC Research on the Prevention of High-speed Rail Accidents <i>Tue</i>, <i>18 Jun</i> 2019 03:38:41 UTC Discussion on fire evacuation measures in operating room of medical institi example of a regional hospital in southern Taiwan examination databank <i>Tue</i>, <i>18 Jun</i> 2019 03:38:41 UTC 	ntation Axis and Incer Patients - listed tutions -	
	Collection Statistics		
	Collection	Total	
	UCF	6755	
	UBOLOGNA	0	
	CRANFIELD	22	
	Arizona State University	8244	
	Atlanta University Center	4332	
	Australiasian Digital Theses Program	56698	
	Poll State University	7507	



2430 2884

Bibliothèque interuniversitaire de la Communauté française de Belgique

that's all folks!

