Data standards: what we can learn from the information and library sciences

Lessons Learned and Future Directions

iDigBio - Leveraging Digitization Practices across Multiple Domains
6-9 October 2014 - Santa Barbara, CA

Larry Schmidt– University of Wyoming Libraries

Before Computers

- What is this?
- What did it Contain?
- Where are they now?

- If you remember them how were they organized?

Back in the day

http://loc.gov/pictures/resource/cph.3a09325/

Standards Made It Easier to Migrate from a text based system to a computer based system.

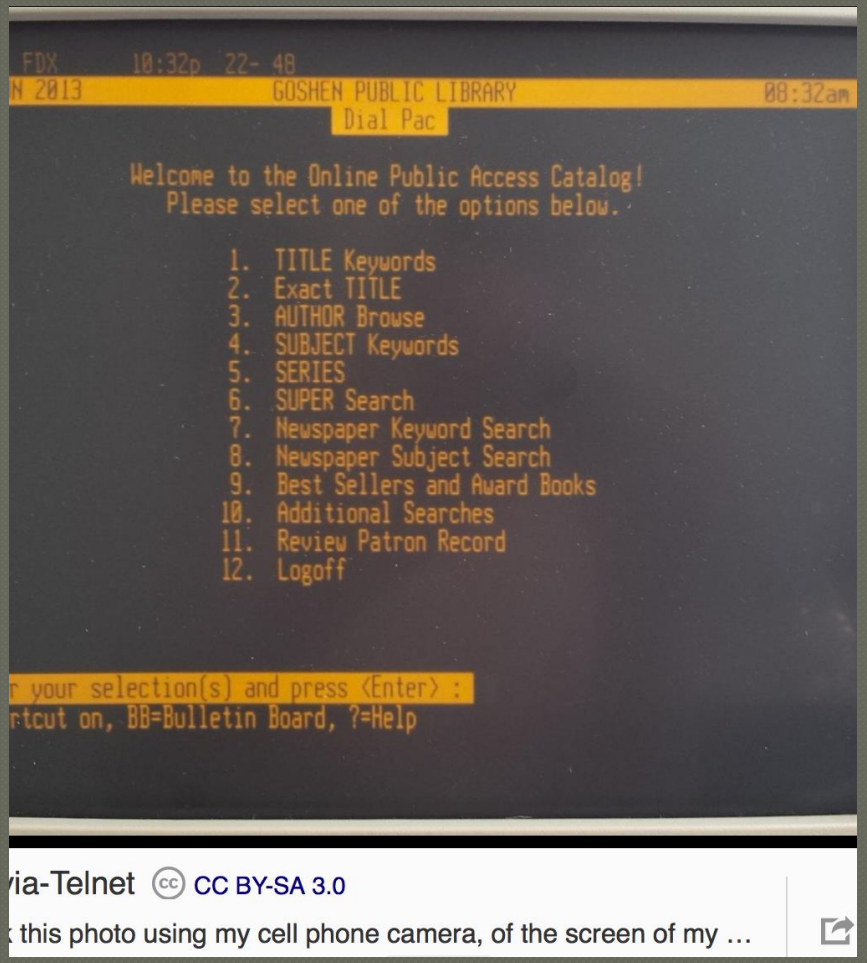
It is all about helping users find the object/information that they are looking for.

OPAC – first generation

- Online catalog
- MARC is a standard
  - Machine-readable Cataloging Record

- Side note:
  - Creative Commons
  - http://creativecommons.org/

http://en.wikipedia.org/wiki/Online_public_access_catalog#mediaviewer/File:Dynix-Main-Menu-via-Telnet.jpg



Dynix-Main-Menu-via-Telnet

```
LEADER 00000cam a2200313Ia 4500
001    40965961
005    20040720081032.0
008    990316s1999    enkaf    b    001 0 eng d
020    0750609613
035    (OCoLC)ocm40965961
040    EUN|cEUN|dYAM|dUKM|dMUQ|dWYU
049    WYUI
082 04 508.075|221
090    QH61 .C378 1999
090    QH61|b.C378 1999
100 1  Carter, David J.|q(David James),|d1943-
245 10 Care and conservation of natural history collections /
       |cDavid Carter, Annette K. Walker.
260    Oxford [England] ;|aBoston :|bButterworth-Heinemann,
       |c1999.
300    xxii, 226 p., [16] p. of plates :|bill. (some col.) ;|c26
       cm.
440  0 Butterworth-Heinemann series in conservation and museology
500    The Natural History Museum's logo on t.p.
504    Includes bibliographical references and index.
590    Roxanne
590    WYUI
650  0 Natural history|vCatalogs and collections.
650  0 Natural history museums.
650  0 Museum conservation methods.
700 1  Walker, Annette K.,|d1945-
710 2  Natural History Museum (London, England)
935    1306531
994    02|bWYU
```

Machine Readable

Who creates MARC records? Why?

AACR2 *Anglo-American Cataloguing Rules*

Designed for consistency between library catalogs

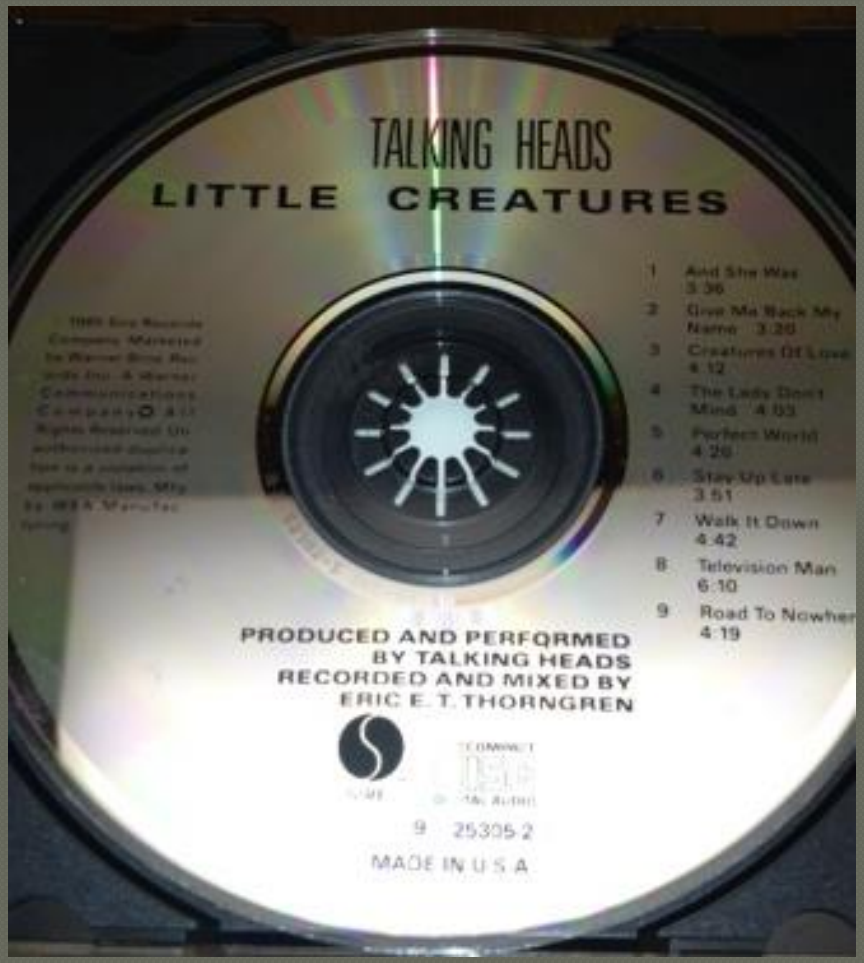
Through Library Standards

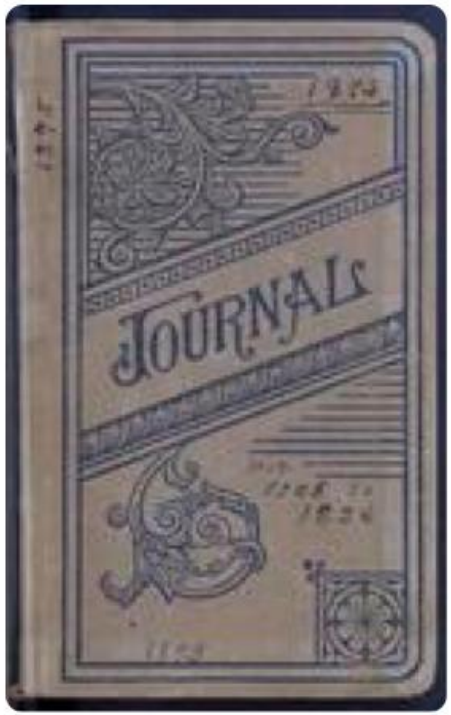
ONLINE COMPUTER LIBRARY CENTER, INC. (OCLC)

WORLD WIDE CATALOG OF LIBRARY HOLDINGS

Z39.50 is an international standard for communication between computer systems primarily, library and information related systems

Media types : File types

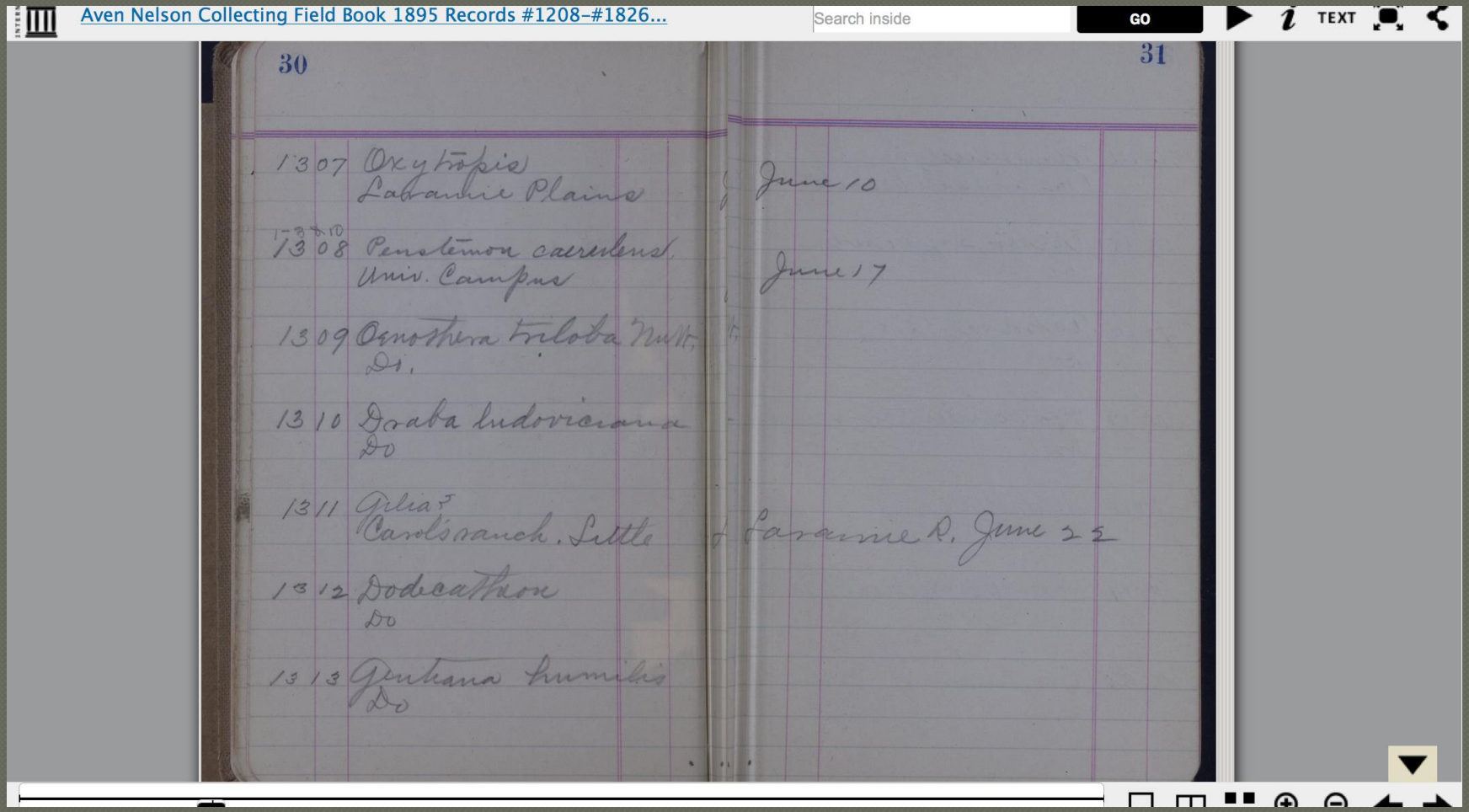




Aven Nelson Collecting Field Book

1895 Records #1208-#1826

Now we have digital formats of traditionally print materials



Traditional library standards are changing to address this

Abstract & revision notes

Abstract

Bibliographic Formats and Standards is a guide to the bibliographic information about records in WorldCat. It provides tagging conventions, input standards, and guidelines for entering information in WorldCat records.

Revision in progress

The OCLC WorldCat Quality Management Division is revising Bibliographic Formats and Standards (BFAS). We have finished incorporating the changes to BFAS that were part of the OCLC-MARC Updates for 2012 and 2013, which were documented in Technical Bulletins 261 and 262.

The larger job of revising BFAS in its entirety is ongoing and includes such changes as:

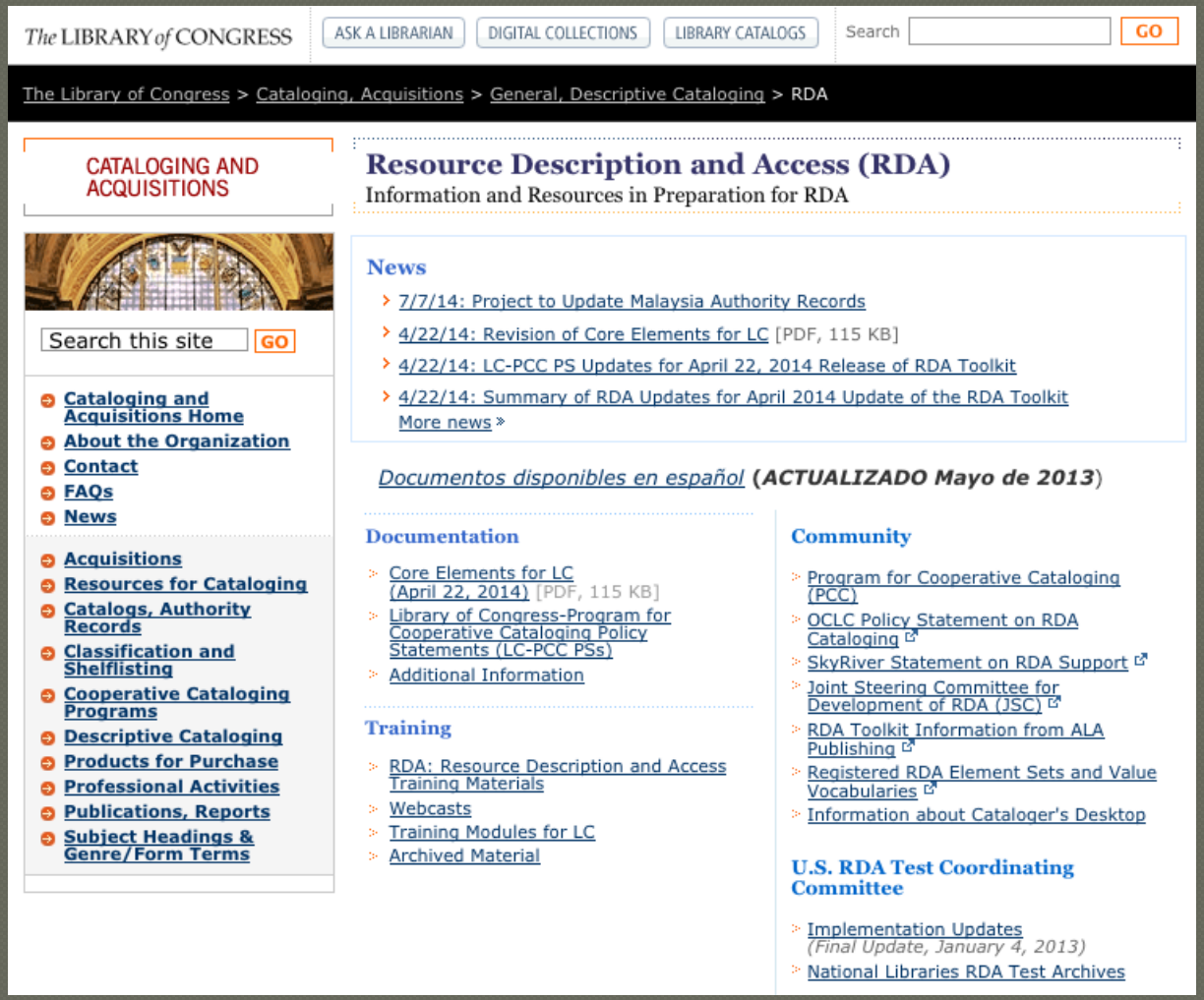
- Updating policies
- Incorporating references to RDA and the Library of Congress-Program for Cooperative Cataloging Policy Statements (LC-PCC PSs)
- Adding and revising examples (RDA and nonRDA)
- Adding links to Searching WorldCat Indexes

This process of review and revision is a huge undertaking that will happen gradually over time. This means you will see changes occurring throughout BFAS at frequent and irregular intervals (the date of the most recent update appears at the bottom of each page). As a result, there may be periods during which the book is inconsistent. Any such periods will be temporary.

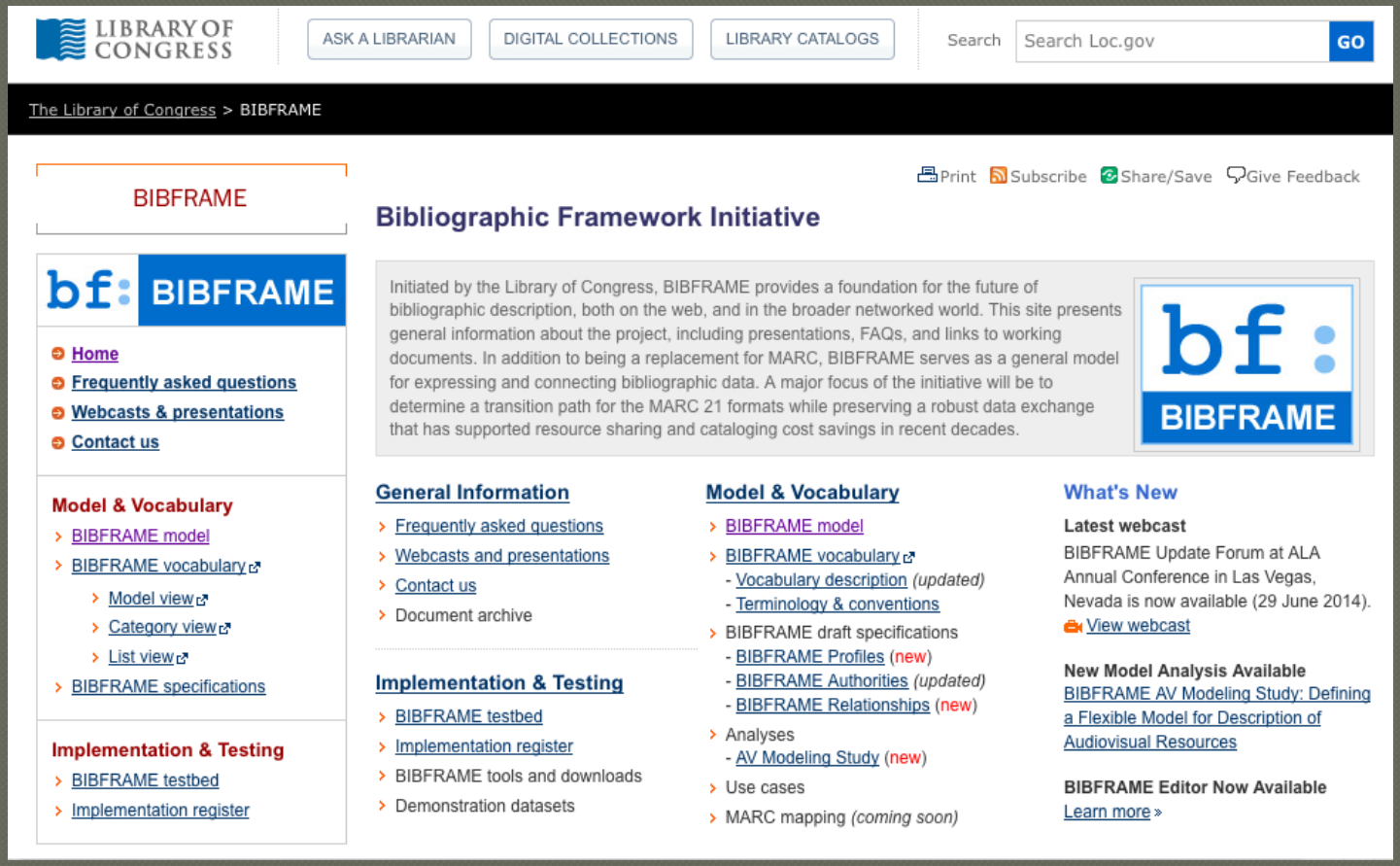
We have brought together on the OCLC About RDA page links to LC, OCLC, and other documentation about RDA. These include the current OCLC RDA Policy Statement and a recorded Webinar that takes an in-depth look at that policy statement. Many cataloging communities (including AV, music, maps, etc.) are working on their own RDA best practices documents and we will be providing links to those as they become available.

Please send any questions or concerns to AskQC@oclc.org.

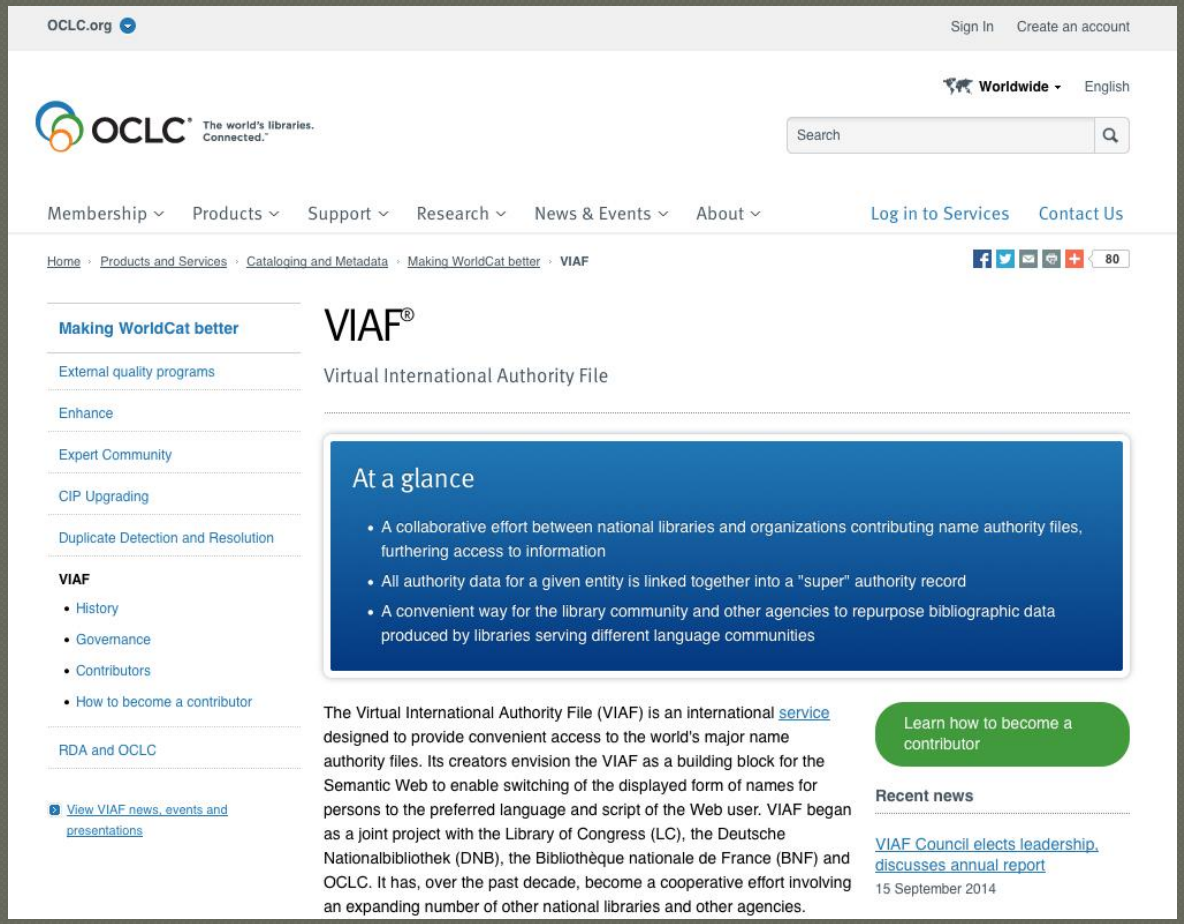
RDA - the new guidelines for bibliographic data - Libraries, Archives and Museums



Along with this the MARC record has to change - BIBFRAME



Authority Files



Links national and regional level authority files
*linking different names for the same person (languages)

Which brings us to this question. So How can Librarians Help you Today?

- You have your systems for physical objects as do libraries
- Libraries today are dealing with multiple materials and file types - you are too
- Linking digital "objects" creates unique problems and libraries have had a head start.

Let's start with

Dublin Core

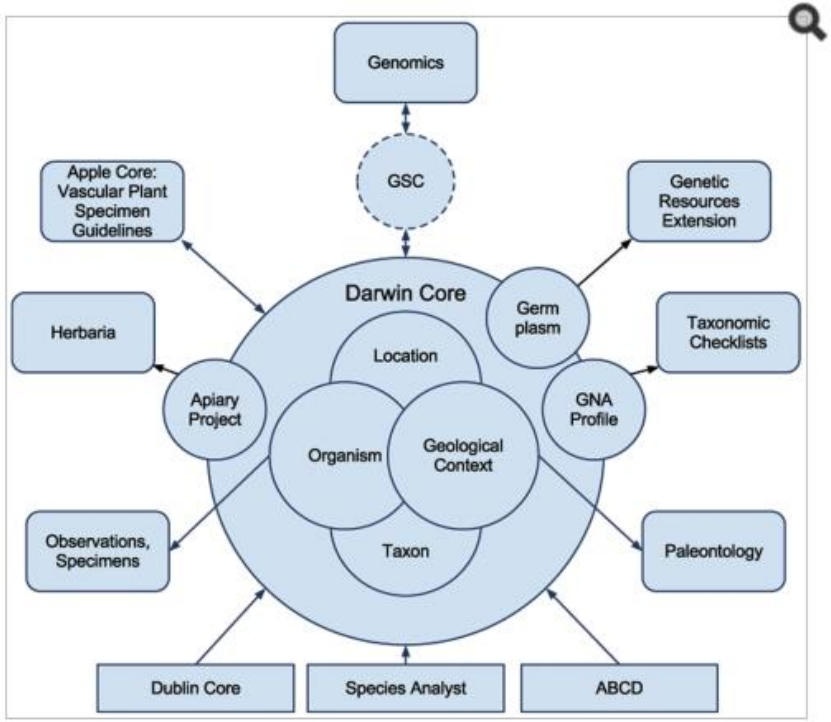
Defines fifteen metadata elements for resource description in a cross-disciplinary information environment.

The Dublin Core Metadata Elements
15 Simple Elements

- Title (Title):
- Author or Creator (Creator)
- Subject and Keywords (Subject)
- Description (Description):
- Publisher (Publisher)
- Other Contributors (Contributors)
- Date (Date)
- Resource Type (Type
- Format (Format)
- Resource Identifier (Identifier)
- Source (Source)
- Language (Language)
- Relation (Relation)
- Coverage (Coverage)
- Rights Management (Rights)

ANSI/NISO Z39.85 - The Dublin Core Metadata Element Set

Figure 1



Scope of Darwin Core: The Standard, deriving from previous standards work (e.g., Dublin Core), describes core sets (e.g., organismal, taxonomic) of characteristics of biodiversity, which are applicable in many biological domains (e.g., Paleontology, Botany).

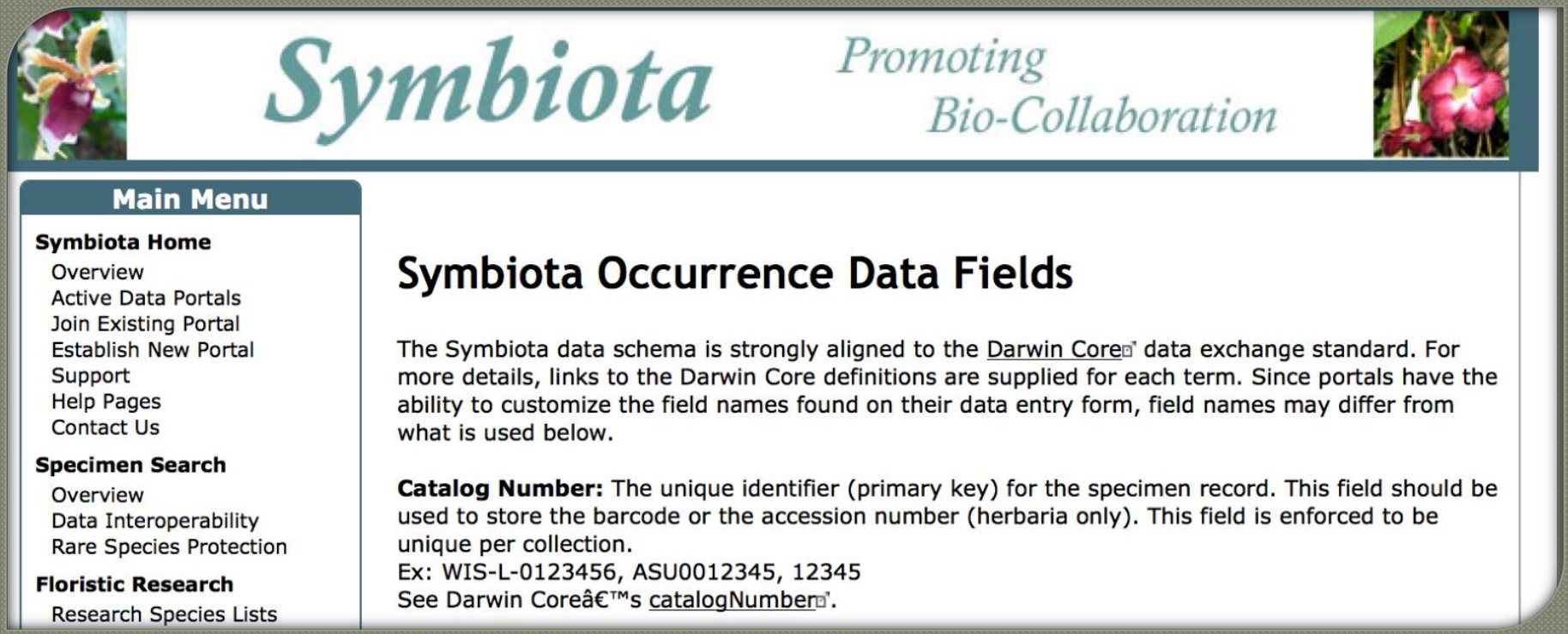
PLoS One

PLoS One

Modified Dublin Core = Darwin Core

http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3253084/figure/pone-0029715-g001/

So you have already used library methods for your online collections if you use



Barcodes, ISBN, ISSN, DOI's, catalog numbers

http://symbiota.org/tiki/tiki-index.php?page=Occurrence+Fields

More Standards

- Standard identifier registries
  - ISBN
  - ISSN
  - ISRC
  - PMID
- Uniform resource locator (URL) & more
  - DOI - ISO 26324:2012
  - GUID
  - ARK
  - PURL

ISSN & ISBN separate identifier for type of media

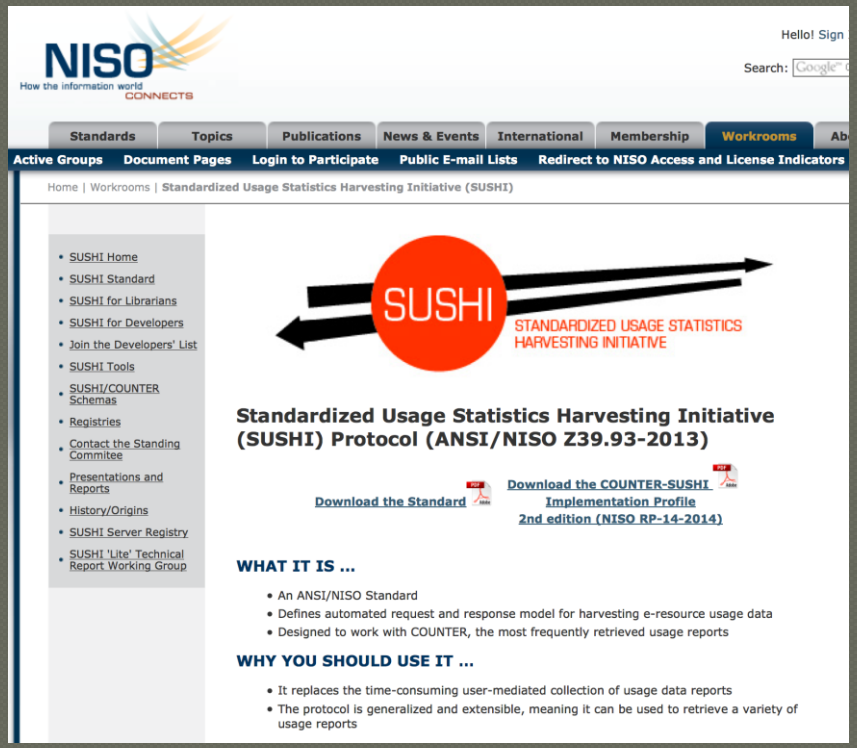
ORCID (**Open Researcher and Contributor ID**) the aim is to help with the transition from science to e-science – may be able to include data linked to the researcher as well

What is a DOI and why is it important?

- It is a character string that is used to identify and object on the web
- Metadata about the object is stored in association to the DOI name
- Includes location where it can be found
- DOI remains fixed but other metadata can change

Standards help us get to where we want and need to be!

- Usage statistics for library e-resources
  - SUSHI



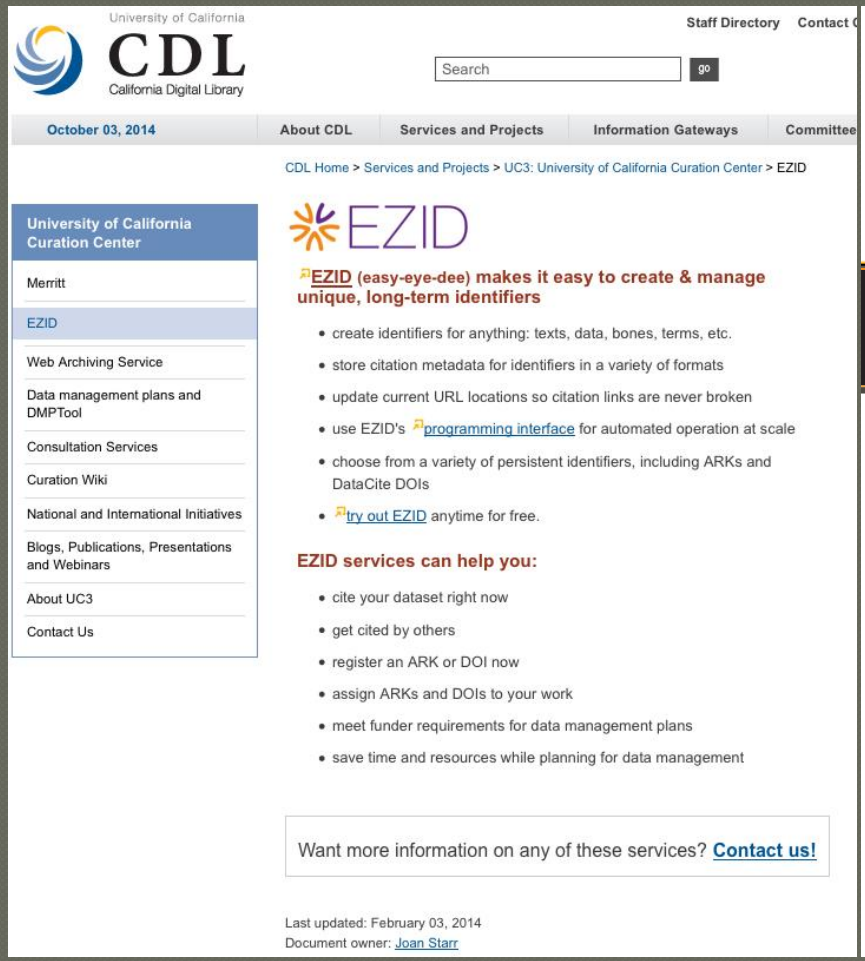
http://www.niso.org/workrooms/sushi

Standards help us get to where we want and need to be!

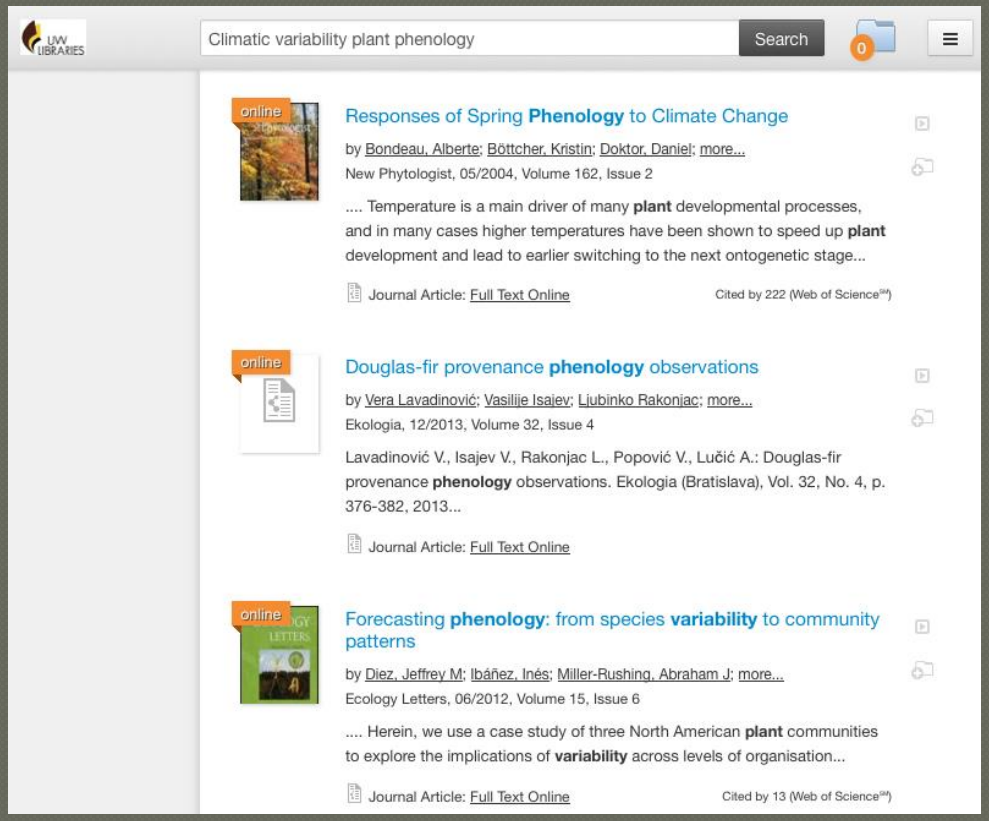
- Libraries and link resolvers
  - Library access to purchased content through open url linking
  - Knowledge base suppliers
  - Content providers

  - Links often did not work and took patrons to content they were asked to pay for

Standards help us get to where we want and need to be!



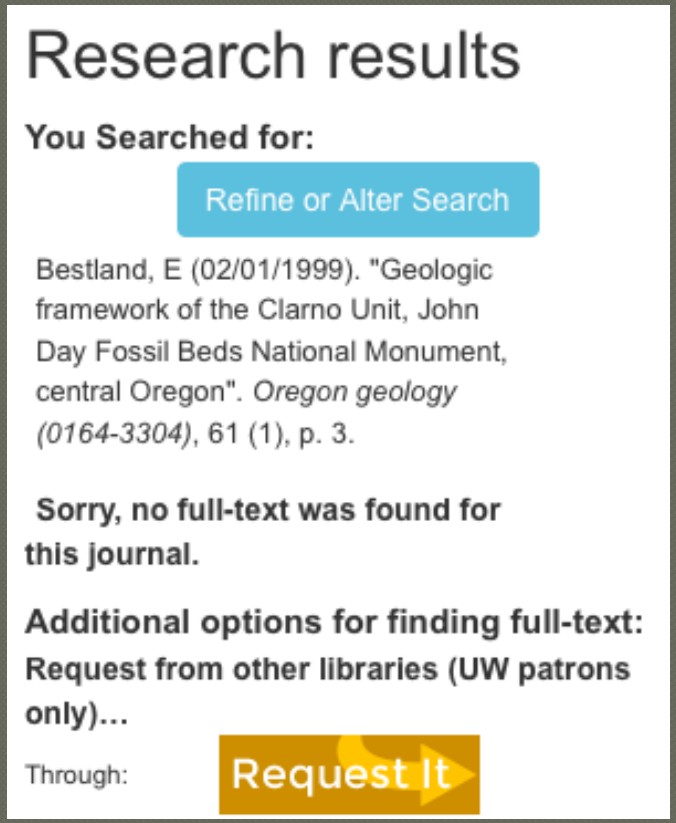
But even with URL's, DOI's, GUID's there are still problems



Article level access through vendor provided databases

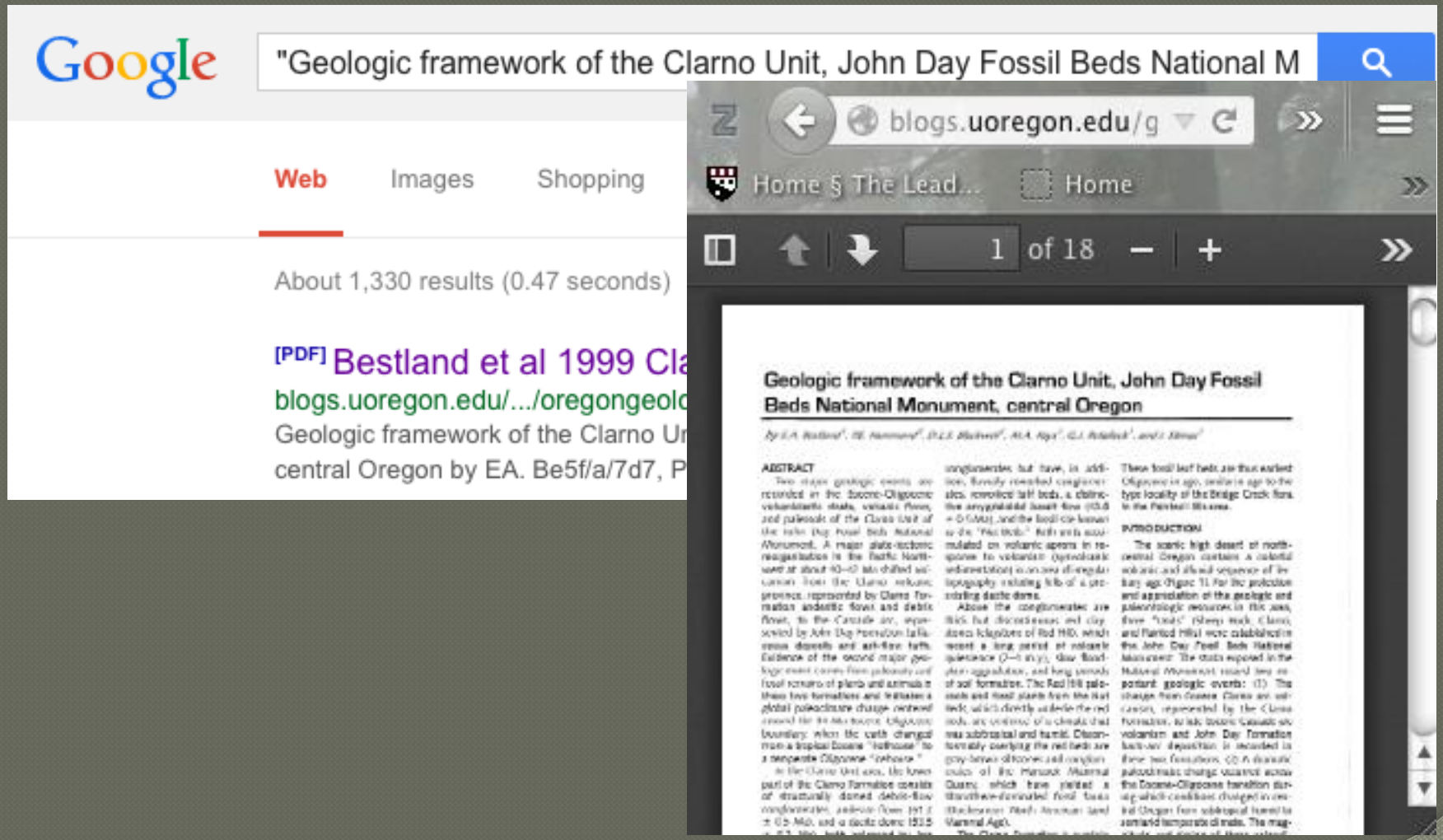
- CrossRef
- Digital Object Identifiers DOI's

We would rather have this

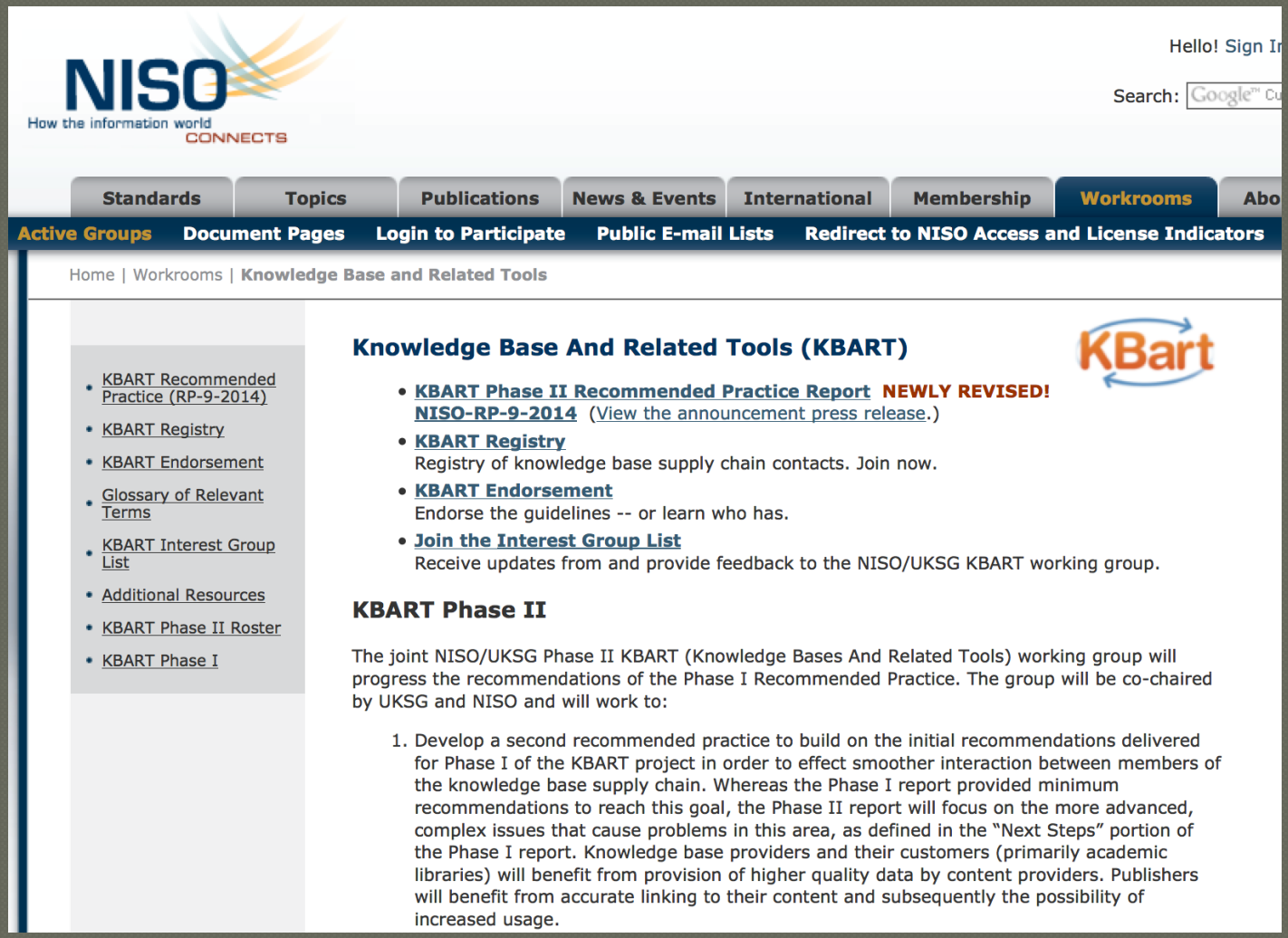


Not this

There is full text for this article



KBART – new recommendations



NISO
How the information world CONNECTS

Hello! Sign In

Search: Google™ Custom Search

Standards | Topics | Publications | News & Events | International | Membership | Workrooms | About

Active Groups | Document Pages | Login to Participate | Public E-mail Lists | Redirect to NISO Access and License Indicators

Home | Workrooms | **Knowledge Base and Related Tools**

- KBART Recommended Practice (RP-9-2014)
- KBART Registry
- KBART Endorsement
- Glossary of Relevant Terms
- KBART Interest Group List
- Additional Resources
- KBART Phase II Roster
- KBART Phase I

Knowledge Base And Related Tools (KBART)

KBart

- **KBART Phase II Recommended Practice Report** **NEWLY REVISED!** **NISO-RP-9-2014** (View the announcement press release.)
- **KBART Registry**
  Registry of knowledge base supply chain contacts. Join now.
- **KBART Endorsement**
  Endorse the guidelines -- or learn who has.
- **Join the Interest Group List**
  Receive updates from and provide feedback to the NISO/UKSG KBART working group.

KBART Phase II

The joint NISO/UKSG Phase II KBART (Knowledge Bases And Related Tools) working group will progress the recommendations of the Phase I Recommended Practice. The group will be co-chaired by UKSG and NISO and will work to:

1. Develop a second recommended practice to build on the initial recommendations delivered for Phase I of the KBART project in order to effect smoother interaction between members of the knowledge base supply chain. Whereas the Phase I report provided minimum recommendations to reach this goal, the Phase II report will focus on the more advanced, complex issues that cause problems in this area, as defined in the "Next Steps" portion of the Phase I report. Knowledge base providers and their customers (primarily academic libraries) will benefit from provision of higher quality data by content providers. Publishers will benefit from accurate linking to their content and subsequently the possibility of increased usage.

KBART – Knowledge Bases and Related Tools

Libraries and linking to purchased content (articles)

- Links often did not work and took patrons to content they were asked to pay for
  - Library access to purchased content through open url linking
  - Knowledge base suppliers
  - Content providers
- Lack of clarity on data standards
- Updates to knowledge base are not consistent

KBART - Group Members

- Librarians
- Publishers
- Aggregators
- NISO
- UKSG
- OCLC

Today it takes a team to:

- Build Collections
- Build Connections
- Organize
- Archive
- Create Access

More examples

- BHL Biodiversity Heritage Library
  - http://www.biodiversitylibrary.org/
- GBIF – Global Biodiversity Information Facility
  - http://www.gbif.org/
- Linked Data - Sharing data across multiple domains
  - http://lod-cloud.net/
- Schema.org

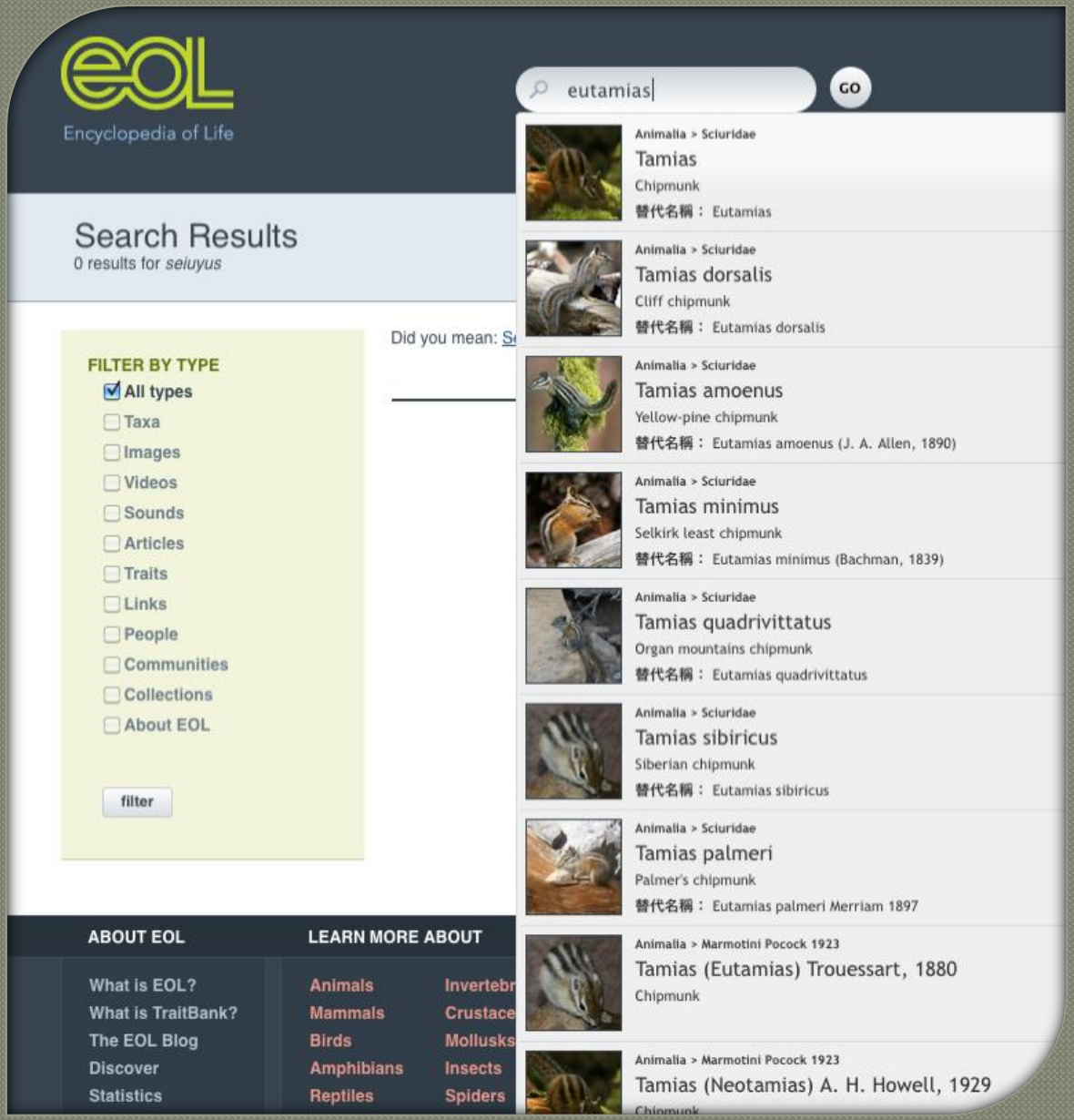
  Effort among search engine providers to promote better linking of WEB content through the use of metadata attributes and HTML markup

Libraries are all about storing, organizing, describing and providing access to information in whatever format it arrives in.

- Metadata
- Linking
- Access
- Archive

The work we do will get more complex as we move to new delivery methods and create new collections!

It is only going to get to be more fun!



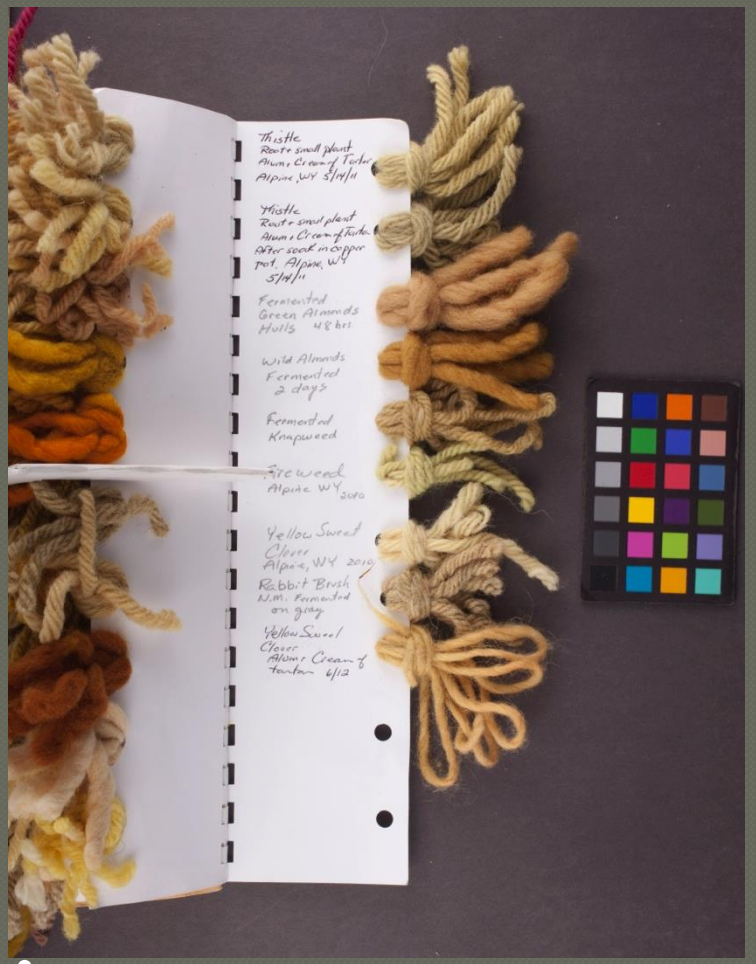
Encyclopedia of Life

Linking digital objects from different collections together

http://eol.org/search?q=seiuyus&search=Go&search_log_id=10445648

Think about it –

On your own



Or consult with a **Librarian**

Actually it will take a team