Digital Libraries: a quick overview by Jens Vigen - Jens.vigen@cern.ch

(11:20am)

- Web was born at CERN by Tim Berners-Lee.
- First American web site was a library catalog.
- This was demonstrated at a CERN conference making the web a reality.
- First smart screen developed by CERN in 1973 – paper on development published in Open Access (OA).

REPOSITORIES

Institutional Repositories VS Subject repositories

IR – features the academic production of an institution, makes research discoverable over the internet.

SR – features artifacts of a given academic discipline e.g. arXiv and inspire (inspirehep.net).

Spires – A database of grey literature

Has 1.3 million records and 50K users

Has digitized collections that are out of copyright – historical collections

Web of science – Quite expensive even for large research firms.

Open Science: Every new paper published in Europe will be OA by 2020

PLAN S: - EU president names OA specialist. A strategy of achieving full OA.

Author have to retain the copyright,

Author to publish in compliant OA journals and platforms especially when funded by certain donor agencies,

Fee for OA should be included in grant and capped.

INVENIO

- Open source digital-library infrastructure,
- Its available under GPL,
- Invenio-software.org,
- Serves as IR at CERN,
- Can handle physical (library and bookshop) collection and digital collection,
- It's a one stop shop for print, videos, animations, data etc.

WHERE TO SEARCH FOR INFORMATION

- 60% of CERN library users less than 35 years with 20% being women,
- 79% of researcher go straight to arXiv

OA MONOGRAPHYS

- A lot in the beginning and less with time

NOTE

- Multimedia: - photo, video and audio recordings for upload to IR,
- Digitization for preservation and improved access,
- CERN has digitalized most of their photo – both black and white and color.
- **CERN Courier** – fully digitized
- The Pauli archive – notes taken by Josef Jauch of Wolfgang Paulis (renown physicist)

SHERPA ROMEO

What can be put in repositories,

Maintained by University of Nottingham.

WORD PROCESSING

LaTex vs Ms Word

Ms word good for small small documents

LaTex better for large document and documents with formula,

OVERLEAF- Used by 2.9 students 3600 academics,

Free for individual users,

BIBTEX – Machine readable reference list

NOTE

Beware of publishers who are out to exploit new researcher,

Thesis should be freely available,

Predatory journals – Beall's list of predatory OA publishers,

Most owned by individuals not research institutions or education institution,

Not published on arXiv,

Very extra compliments,

Usually cheap,

Editorial board not authorities in the field,

Frequency of publication high.

- Librarians better placed to advice academics

INSIGHT INTO DIGITAL LIBRARY SYSTEMS BY JEAN-YVES LE MEUR

- Project leader of CERN digital library
- Set up first web server in 1993 – eprint,
- Currently working on preserving of digital content,
- Digital library – all the objects are digital, no print, microfilms or tapes.
    - Information retrieval systems
    - Virtual organization

Born digital vs converted to digital

- Digital library vs digital archive
- Green open access vs Gold open access
- Gold OA – published article in open access journal
- Green OA – published article in any journal then self-archives a copy in a freely accessible IR
- ILS vs Document management systems vs multimedia systems, research and data systems vs eprint repositories vs hybrid systems

STANDARDS

- Content representation: Marc, Dublin core, JSON,
- Exchange protocol: Z39.50, OAI-PMH between data and service provider
- Interoperability – SWORD- simple web-service offering repository deposit
- Identifier – ISBN, DOI, ORCHID
- Preservation of metadata: METS with descriptive, structural and administration of content in the OAIS

DIGITAL LIBRARIES

- usually have targeted communities
- Internet livestats.com – internet statistics
- Software:
    - Eprint – in pearl since 1990
    - Dspace, Fedora, Greenstone, koha, invenio
- Most participants using Dspace and Koha
- CERN lab started in 1954
- Spire was first database on the web
- Invenio – scalable, flexible, collaborative, books, open source and hybrid – repository, library system, multimedia and document management,
    - Can also be used for administrative documents,
    - Can be used in many different langaugaes,

WORKFLOWS

<u>Main component</u>

Ingestion – Different inputs, batch import and web submission

Processing – Classification and extraction, indexing

Page rank – clustering big data, like and hot trends (What documents to display first)

Curation –

Dissemination – Push and pull, alerts, tags, search ui, basket, discussion, newsletter, circulation

Access control: authentication and authorization management

NOTE

Technology evolution – Invenio ver1 2000's to ver3 2010's

PMB – repository software used by Cameroon university

A successful lib need good content and relevant services

Social media: - how best to engage internal users (Staff)?

Digital library with quality and relevant services capture the users' needs

Free eBooks: e.g. google books, amazon – the look inside function

PERSISTENT IDENTIFIERS: ORCID BY ANNETTE HOLTKAMP

The challenge – Unambiguously link researchers with their professional activities,

How to tell one researcher from another,

Name not enough to reliably identify an author,

Solution

Orcid – unique identify to identify researchers

Open, nonprofit and community driven

900+ organizational members

5.2million live iDs associated with researchers

Participants were able to create orcid IDs

DOI – Document Object Identify

- NO disambiguous
- Associated metadata

- Persistent identify of an object

Other identifies include:

- Uniform resource identifier/Uniform resource locator
- Uniform resource name

ZENODO

- A cloud infrastructure to archive research
- CERN developed together with EU for institutions without IR,
- Can also archive Github projects.
- It hosts researcher of any subject area to upload files.
- Can upload up to 100 files and 2GB at once,
- User have up to 50GB free space.
- Users have to describe their research data to enhance discoverability.

Felix Muriuki (Kenya) – bmarcfelix@gmail.com

Ateben Uta Rein (Cameroon) – rein.bcyui@gmail.com