



CERN Library Requirements

**T. Baron
CERN ETT-DH-CDS**



Outlines

- ◆ **CERN Library Wishes**
- ◆ **JACoW Records in CERN Library**
- ◆ **A Possible Solution: OAI**



CERN Library Wishes (1)

- ◆ **In CERN Library Mission Statement:**
 - ◆ **Keep track of all CERN production**
 - ◆ **Maintain it for long term**
 - ◆ **Produce an annual report each year**
- ◆ **Obligation of getting information from JACoW**
- ◆ **Up to 2002: “semi-manual” import**
 - ◆ **Lot of human efforts**
 - ◆ **Delay in publishing**
 - ◆ **Difficulty to achieve completeness**



CERN Library Wishes (2)

◆ Automatic Procedure

◆ To get metadata

- ◆ Language
- ◆ Author(s)
 - ◆ Name
 - ◆ Affiliation
- ◆ Title
- ◆ Conference:
 - ◆ Name
 - ◆ Place
 - ◆ Date
- ◆ Abstract
- ◆ [Number of pages]
- ◆ Link to fulltext document



CERN Library Wishes (3)

◆ Automatic Procedure

◆ Possibly to get Fulltexts – 2 solutions:

- ◆ Keep only the links to the JACoW fulltext archive
 - ◆ URL stability?
- ◆ Fetch the fulltexts and store them in the CERN Document Server
 - ◆ Long time storage place
 - ◆ Added Services



JACoW Records in CERN Library (1)

- ◆ **Metadata Enhancement:**
 - ◆ **Normalisation (authors, experiments...)**
 - ◆ **Completeness (reference numbers...)**
- ◆ **Added Services**
 - ◆ **Integration in a library catalogue**
 - ◆ Multiple catalogue search (eg.: search for “Evans” inside “Conference papers”, “pictures” and “journal articles”)
 - ◆ **CDS Personalization tools**
 - ◆ Alerts, baskets...
 - ◆ **On fulltexts:**
 - ◆ Keyword extraction
 - ◆ Fulltext indexing
 - ◆ Automatic Conversion
 - ◆ Citation Extraction



JACoW Records in CERN Library (2)

- ◆ **Increased Visibility:**
 - ◆ **156000 different users/year**
- ◆ **Interlinking CERN-> SLAC**
- ◆ **Launch search on other systems (Google, KEK, SLAC...)**
- ◆ **HEPDOC: parallel search on CDS, KEK and SLAC**



A Possible Solution: OAI (1)

- ◆ **Open Archives Initiative: Protocol for exposing/harvesting metadata initiated by LANL people in 1999.**
- ◆ **Aims at simplicity and easiness of implementation**
- ◆ **Widely supported and used**
- ◆ **Last workshop @ CERN in October with 140 participants from 20 countries**



A Possible Solution: OAI (2)

- ◆ **Two actors:**
 - ◆ **Repositories (also called data providers) (JACoW)**
 - ◆ **Harvesters (also called service providers) (SLAC, CERN...)**
- ◆ **Metadata representation:**
 - ◆ **Unqualified Dublin Core metadata format mandatory**
 - ◆ **Other metadata formats can be added**
 - ◆ Qualified DC
 - ◆ Marc 21
 - ◆ ...
- ◆ **Relies on the http protocol**



A Possible Solution: OAI (3)

◆ Repositories:

- ◆ Serves the metadata on request from the harvester
- ◆ Each document has a unique ID
- ◆ Some “verbs” are implemented (http cgi scripts)
- ◆ The verbs serve the requested metadata
- ◆ Can group records into data sets (conferences)

◆ Harvesters:

- ◆ Invoke repositories OAI verbs
- ◆ Time-driven requests (“I want all documents published/modified from this date”)
- ◆ Category-driven requests (“the documents should belong to the EPAC’2002 conf”)
- ◆ Can chose between the proposed metadata formats.
- ◆ Can directly access one record knowing its OAI id



A Possible Solution: OAI (4)

◆ Technical Implementation:

◆ Repository:

- ◆ Create unique references for each record
- ◆ Date stamp all records for creation and modification
- ◆ Implement the 6 OAI requests (verbs)
- ◆ Reduced work (CDS people can help)

◆ Harvester:

- ◆ Harvester programs are already available.



CONCLUSION

- ◆ **CERN Library Wishes and Services**
- ◆ **OAI Interface**



The End

- ◆ <http://www.openarchives.org>
- ◆ <http://www.dublincore.org>
- ◆ <http://cds.cern.ch>