Since the creation of CERN 50 years ago, library mission is the same: the dissemination and long term keeping of High Energy Physics results.

Only the means have changed.

1965: 1st computer ever used by CERN library

1990: 1st computer ever used by WWW

1993: Preprint
1996: Library
1999: Document Servers

2002: 1st release of CDSware to promote OAI movement
General description

- CERN Document Server Software (CDSware) is the software developed by, maintained by, and used at, the CERN Document Server (5 to 10 people)
- It allows you to run your own electronic preprint server, your own online library catalogue or a document system on the web.
- It has been deployed through an incremental organic-growth SW development model.
- It uses freeware technology: MySQL RDBMS; CDSware Indexes; Apache/Python; XML MARC
- The CDSware is free software, licensed under GNU General Public Licence (GPL).
- It complies with the Open Archives Initiative metadata harvesting protocol (OAI-PMH) and uses MARC 21 as its underlying bibliographic standard.
To or not to?

Cataloguers are becoming “batchers”!

- **CDSware** can harvest metadata from
  - OAI sources with BibHarvest
  - Non OAI sources can be “transformed” with BibConvert, allowing for example loading backlog of records

- **CDSware** provides its metadata via
  - OAI
    - Records can be private, public and “OAI-public”
    - OAI Sets can be defined using any search criteria
    - XML MARC; XML Dublin Core and more…
    - any query that is “OAI-ready”
      - Eg: OAI harvester could harvest only papers written by Ellis, J.
      - Eg: OAI harvester could harvest only title fields
  - Command Line Interface (APIs)
    - Eg: weekly bulletin, e-journal site and many applications

At CERN, 92% of the acquisition is done through automatic import from more than 80 different sources - not using OAI (yet) in most cases
Each collection can have its own submission policy
- Direct submission
- Submission with monitoring
- Submission with simple approval
- Submission with peer reviewing and editorial board

Each collection can have its own record definition
- Metadata fields (mandatory, optional, controlled at input time, ...)
- Full text formats
- Revised versions

Each submission has its own process management
- With an HTML administration interface
- To define submission screens
- To define actions to be applied

Eg:
- When finishing the submission of a videotape by video service, the label is created (PDF) to be stick on the tape
- When a note is submitted by ATLAS collaboration, the option of sending comments to its author is provided to the collaboration members
From the Archive to the end-user

- Google-like speed for up to 1,000,000 records
  - Web Application server ↔ DB server
  - DB insufficient: in-house performance-driven index design
  - Fast marshalling & fast set intersections:
    
    | query       | no. hits | search time |
    |-------------|----------|-------------|
    | cern        | 223,843  | 0.07 sec    |
    | of          | 439,793  | 0.07 sec    |
    | of cern     | 109,635  | 0.10 sec    |
    | of cern the | 11,940   | 0.17 sec    |

- Combined metadata/fulltext/reference search
  - Eg: title:higgs or reference:higgs or fulltext:higgs

- Multi-stage search guidance system

- Personalization: baskets, email alerts

- Navigable collection trees
  - Primary and Virtual orthogonal views

- Internationalization: multi-language interface

“When it was proclaimed that the Library contained all books, the first impression was one of extravagant happiness.”
Borges
Long Term Archive?

- CDSware at CERN
  - “Certified Information System” (CIS)
  - Considered as a long term electronic archive
  - Hosts the official CERN Archives
- MARC21 based: LOC standard
  - XML MARC is the internal representation of CDSware records
- Records deletion policy
  - Record IDs never change
- Full text automatically converted to PDF
  - CERN Conversion server can be plugged in (GNU GPL)
- Digital content disseminated… via OAI!
CERN Document Server

Over 630,000 bibliographic records, including 250,000 fulltext documents, of interest to people working in particle physics and related areas. Covers preprints, articles, books, journals, photographs, and much more.

Search 650,262 records for:

Narrow search:
- Articles & Preprints (521,214)
  - Published Articles (168,571) Preprints (289,833) Theses (29,534) Reports (25,652) CERN Internal Notes (8,145) CERN Committee Documents (24,105)
- Books & Proceedings (60,310)
  - Books (33,631) Proceedings (12,679) Standards (7,511)
- Presentations & Talks (13,263)
  - Conference Announcements (12,337) Academic Training Lectures (457) Summer School Lectures (286) General Talks (17) Videotapes (405)
- Periodicals & Progress Reports (2,934)
  - Periodicals (2,347) Progress Reports (587)
- Multimedia & Outreach (10,473)
- Archives (49,823)
  - CERN Archives (44,420) Pauli Archives (3,472) DSU Archives (701) SL Archives (1,026) AB Archives (203)

Focus on:
- CERN Yellow Reports (391)
- CERN Divisions (46,193)
  - Accelerator Sector (9,251) Administration Sector (22,222) Research Sector (12,348) Technology Sector (2,462)
- CERN Experiments (8,720)
  - LEP Experiments (2,687) LHC Experiments (6,055)
- CERN Projects (1,043)
  - LHC Project (1,043)

- 125,000 distinct hosts/clients in 2003
- 12,000 distinct hosts/clients per month
- 120,000 searches per month
- 5,000 OAI harvesting requests per month
Future: a Full Open Digital Library system

- Extending traditional library systems
- Designed to evolve
- Suitable for mid to large size repositories (1M recs)
- Dedicated support from CERN CDS team

- Used in more and more places, (or considered for use) by:
  - University of Missouri-Columbia, USA
  - Fundao Osqaldo Cruz (Ministry of Health) Rio de Janeiro, Brasilia
  - ISDN-ENSSIB, France - Montreal International - Bologna University, Italy
  - UN Population Fund, New York, USA
  - Instituto de investigacions Electrica, Mexico - Casalini Libri, Italy
  - HBZ-NRW, Germany - SDSC, USA
  - Aristotie University of Thessaloniki, Greece
  - RERO: Consortium des bibliothèques publiques de Suisse Romande, Switzerland
  - and: EPF Lausanne, Switzerland

- Package downloaded 510 times in 2004 (on mid-may)
Towards the Paperless Office for researchers?

We still have some work left…
Questions?

- Distribution site: http://cdsware.cern.ch
- At CERN: http://cdsweb.cern.ch
- Support email: cds.support@cern.ch

- Send me an e-mail: Jean-Yves.Le.Meur@cern.ch
- Questions?