Digital content sewed together within a library catalogue

WebLib - The CERN Document Server

- Jens Vigen
- Scientific Information Officer, CERN
- Workshop on Personalization Techniques in Electronic Publishing on the Web: Trends and Perspectives
- Malaga, Spain, 28 May, 2002
Integrating digital content

- Everybody can do it, or ...
  - Content and knowledge are though required
- Transfer of librarians knowledge
- Explore what you already are paying for
- Costs "nothing"
- Go home and do it!
Setting the scene

- CERN / particle physics community
  - Fast developing field
  - Expensive equipment
  - Well developed preprint culture
- Pre 1990
  - Traditional communication channels
  - Traditional libraries
December 1990 - WWW

"... the earth was formless and empty, darkness was over the surface of the deep ..."

... and all the world's web servers were at CERN - maybe even in the same building.
1990 - "stone age"
What happened since?

- 1990 - the Web was born
- 1991 - xxx.lanl.gov
- 1996 - first version of WebLib; IOP launches their e-journals
- 1997 - reference linking is getting on fashion
- 2000 - journal collections made available retrospectively
- 2001 - the start of the semantic web?
- 2002 - CDSware
Go direct - SFX@CERN

- Linking ISBNs to Amazon
  - ISBN 0-8493-0596-9
- Standards linked to original records
  - The new ISO 9000 series
  - Same principle as for ISBNs
- Linking with system numbers
  - i.e. SLAC and KEK or to abstract of UMI dissertations
Go direct - SFX@CERN

- Linking journal references
  - The triplet - has served us for 400 years
  - Phys. Rev. : 48 (1935) , pp.73-77
  - http://prola.aps.org/abstract/PR/v48/p73_1

- A too simple solution?
  - Massive lobby was required
  - Good collaboration with DataStarWeb (INSPEC)
  - Get in early - not wait for the market
References

- Correction

- Linking
  - Simply using Go Direct

- Example:
  - References in astro-ph-0203041
Matching

- Uploading publication references
  - Little or no help from the publisher side

- Results
  - **2000**: collected 38160 preprints; 40% detected as published
  - **1999**: collected 35115 preprints; 52% detected as published
  - **1998**: collected 28 922 preprints; 53% detected as published
Key wording

- Automatic indexing
  - Learning set based on DESY/SLAC SPIRES
  - Keywords created on the fly

- Opens to enhanced reader services
  - Linking extracted terms to dictionaries, encyclopaedias, tables etc.
    - Oxford English Dictionary
    - Encyclopaedia Britannica
    - Review of Particle Properties
    - Landolt Börnstein
Services

- Tailor making services at a personal level
  - Information provision
    - Alerts/searching
    - Creates close relations to the various users
    - "The Google-approach" is not always the best ....
  - Records management
    - Electronic shelves, list of loans
Conclusions

- **Trends**
  - Dynamic linking, authoring tools
  - First signs of leaving the PDF era?

- **Integrating digital content**
  - Partly a technical challenge
  - Strong strategy
  - Involved parties must get together
  - You have to believe in it!
Publicity:
Automation Tools soon available as OpenSource

- **CDSware**
  - Upload
  - Extraction/correction of references
  - Matching
  - Linking
  - Key wording

- Download from
  - [http://cdsware.cern.ch/](http://cdsware.cern.ch/)