document data & personal data

Thomas Krichel
Long Island University & Novosibirsk State University
2007-04-18
about me

• I am working for the College of Information and Computer Science of the CW Post campus of Long Island University and for the Faculty of Information Technology at Novosibirsk State University.

• Next year on sabbatical.

• I am a trained economist.

• My main claim to fame is the creation and coordination of the RePEc digital library for economics at http://repec.org.
about this talk

• My attendance today is sponsored by the conference. I am very grateful.
• I am mainly discussing work that has been funded by OSI.
• Since we are in the technical section I will limit myself to technical details.
• Ask me later: why?
origin of this work

• Most of the ideas are the work of my consultant Ivan V. Kurmanov.

• I showed him the draft slides this morning. He wrote

  “I looked at it. It is soooooo complicated, it is soooo much text. Many unimportant details.”

• I am sooo sorry!
general picture

• I discuss the interoperability between two types of services.
• One is a document-submission service henceforth a “dosus”.
• The other is an author-registration service henceforth a “aures”.

dosus

- A dosus may be an IR, or a subject-specific repository, or something else.
- Its main function is to collect document data.
- Author metadata comes as part of the metadata information with the document.
- The author may be identified within the scope of the collection.
- The user of a dosus is called a submitter.
An aures is a service that collects personal data and connects them with metadata about documents. The key data is author name data and document identifiers.

Authors can contact the aures to identify themselves.

Once they are registered they can say what documents they have authored.

The user an aures is called a registrant.
ACIS

• This is the academic contribution information system.
• It is a generic software to enable aures services that are somewhat more general.
• Work on ACIS is sponsored by the Open Society Institute.
dosus ↔ aures interoperability

• Interoperability comes in different levels.
• With each level up, we have more (better) interoperability.
• We have levels 0 to 4.
• At level zero, an aures and an dosus simply live side by side, and no interaction is happening.
level 1

• In level 1, a dosus provides metadata about its documents to aures.
  – The data is stored in files
  – In a compatible format. For ACIS this would be AMF or ReDIF.

• The aures processes the data periodically.
  – add new records to the document stock
  – perform probationary associations between documents and authors
A dosus delivers to the aures data for some of its authorships that point to data in the aures. ACIS will accept any of the following 3 identification avenues:

- an identifier known to the aures
- a shortid, previously generated by aures
- an email address, known to the aures as the login on a registrant.

This data will have to be entered by submitter.
level 3

- The dosus helps submitters to find the data required for level 2 interoperability.
- While submitters enter authorship data, the dosus performs searches in the aures data. If matching records are found, the submitter is invited to select them.
- The document data is the exported to the aures in the usual way.
implementing level 3

• The aures needs to expose registrants data to the dosus. The data can not be made available publicly if we want the email to be an avenue of identification.

• The dosus must search the aures data display optional matches in an unobtrusive way and give submitters an easy way to choose an option.
level 4

- The dosus immediately notifies the aures about a document submission.  
- The aures processes the notification, the document is added to the research profiles of its identified authors.

(you may argue: it's on a different level)
level dependency

• There is level dependency
  – Level 1 is really required for other levels.
  – Level 2 is a basis for level 3.
  – Level 4 can be done without either level 2 or level 3.

• It does not really matter current ACIS code can implement all four levels.

• There is ACIS code for Eprints 2.0 that implements the dosus side of the interoperability.
what other things does ACIS do

- Well ACIS does not do anything, it is a software.
- But let us think about an ACIS implementation of aures, an aiares.
- The RePEc author service at http://authors.repec.org is a aiares.
aiares name details

- It contains the name details as they may be found in the bibliographic data
  - Krichel, Thomas
  - T. Krichel
  - Томас Крихель
- “Bruno van Pottelsberghe de la Potterie”
- Sometimes a name of an author may not appear in the bibliographic data at all
  et. al.
aiareds contact details

• This is a set of trivial fields
  – email. This detail is required but not exported by default.
  – homepage. This detail is optional.
  – phone number. This detail is optional
  – postal address. This detail is optional.
affiliations profile

• This is more complicated.
• Institutional data is kept as separate records, not as string data.
• Registrants can search for existing institutional records to create an affiliation with.
• Or they can propose a new record to be added by filling out a form.
research profile

• This is a collection of metadata about research documents the registrant has written.

• Available functions include
  – display a list of works in the profile
  – search for new suggested works
  – manual search for works by title
  – display refused research documents
  – change preferences for automatic updates
  (next)
automatic updates

- By default, when a document record quotes an aiares short id, the document is added to the profile.
- By default, a regular search using the name variations profile identifies a set of potential new documents and reports them to the user.
- The registrant may choose to have exact matches of these searches being added to the research profile.
new research profile features

• Document to document links can be created for authors to say that two documents in the profile are related.

• Document full-text links can be confirmed or rejected.
  – Typically such full-text files would found by an automated search external to the aiares.
  – See the talk that I gave here in 2004.
citations profile

- Within this profile, author can partially manage citation information for items is the research profile.
- Like a dosus may submit data to a aures, a citation discovery service may take give citations data to a aiares.
- Such data can be maintained in the citations profile.
References processing

- References are processed to see if they may correspond to a document in the research profile.
- If a document in the profile has a potential citation it is called an “interesting” document.
- Once reference processing is done, registrants can navigate by decreasing level of interest.
Suggestions processing

- Registrants navigate the set of suggested citations to see if the reference string really matches the research profile item.
- If the registrant refuses a citations, there is a screen where she can later overturn such a decision.
automatic updates

• If the reference is very close to citation data, the registrant can have it added automatically.

• When a co-author has identified a citation to an item in her profile, the registrant can allow it to be added automatically.
Thank you for your attention!

http://openlib.org/home/krichel