

February 2007

Institutional Repositories Workshop Strand Report

Strand title: Author Identification

1. Executive Summary

2. Summary of recommendations

The group formulated two main recommendations:

Capitalise on knowledge from ongoing initiatives; not only from the library community but also from the Internet community and from standardisation activities within e.g. the universities.

Address the need for a sustainable model for author identification. The level of ambition should match the expected funding for the activity.

To support these recommendations, the group proposes four initiatives:

1. initiate two studies, one on the need for author Identification and one addressing the potential business models for a author ID based services,
2. establish a prototype for cross institution use of author ID, which can serve as a blueprint,
3. arrange workshops for experts from the library field and from universities and from the Internet community identifying relevant initiatives and potential architectures for a service and finally to
4. establish a working group looking into the flow of relevant metadata.

3. Discussion (including recommendations and items of interest)

The discussion was guided by the position paper written by Traugott Koch on author ID and some of the issues in it was modified – Traugott, did you rescue what you wrote????

Identified issues:

In the cause of discussion a number of relevant issues were identified:

1. Various ID systems
 - a. openID
 - b. ISPI
 - c. ID system developed by SURF
2. Business model (payment, DRM, homepage resolvers,.....) short term/long term
3. Metadata
 - a. actual data

- b. link/maintenance infrastructure
 - c. quality of data
 - d. who do you trust to produce and maintain data
4. Legal Issues (also on European level – EC activities?)
5. Should we develop something ourselves or rely on existing system (such as DAI)?
6. ID resolution (national, international resolver)
7. What is the need for author identification
8. Identify important stakeholders (universities, authentication (shibboleth), authors, ...SURF, DEFF,.. ISO,.. publishers,..libraries,...)
9. Technical issues – how can we facilitate
10. How to populate the ID database
11. identify critical success factors /killer application
 - a. short term critical success factors
 - i. used to improve retrieval
 - ii. used in citation /research statistics and evaluation
 - b. long term critical success factors
 - i. used to improve retrieval
 - ii. used in citation /research statistics and evaluation
 - iii. could be used as a basis for new services
12. influence/participate in international initiatives (such as ISPI)
13. Divide between open and commercial systems and communities
14. how to deal with heterogeneous systems (there will be more with different purposes and supporters)
15. global <-> local part of an identifier (e.g. country or organisation could have their own part – and it could be prefixed to give a global identifier)
16. Relation to the content infrastructure (e.g. Driver)
17. How can different metadata systems be based on a core author identification system -> kind of architecture question
18. Different layers

Five issues

For each of the issues relevant activities were discussed and for a number questions like local, national and international were touched.

Based on this list, the issues were grouped where relevant and be the group turned into six issues as described below.

Issue 1: Articulate the need for author ID in connection with IR

As a foundation of all the other steps it is suggested to produce a document investigating:

- What are the special requirements for IR
- Identify common ground with other activities
- Investigate failures and success of present and past activities
- Describe use cases for author IDs
 - Ensure input from use community (e.g. workshop)

The objective of this document is to get acceptance of the identified needs and anticipated usage of Author Identification in the KE community

Issue 2: Ensure that IR related ID developments are compatible with developments in other sectors

The developments of author identification is addressed by many parties and it is important to capitalise on the development, which happens in the Internet community. It is expected, that libraries should adopt standards and solutions developed elsewhere. We therefore propose to

- Identify stakeholders and other initiatives – identify key stakeholders for IR activities and to hold a
- Workshop with other ID initiatives such as ISPI, OpenID, authentication systems, etc.

The objective is to establish communication lines and active participation with relevant stakeholders/parties

Issue 3: Define metadata for author identification relevant for IR

The area of metadata is both simple and complicated. In principle author ID is in its pure form very simple, however use cases add to extended functionality. Additional metadata required to verify the identity is one example, its use in services across different repositories is another. Therefore there is a need for identifying requirements for different situations – and for identifying where required metadata best is maintained – e.g. by governmental parties, by universities or others.

It is therefore suggested to establish a working group addressing:

- Specify functional requirements (Verification, match between repositories,...)
- Identify and prioritise metadata elements for the different tasks (required metadata will depend on context) including legal and privacy aspects
- Identify roots for the metadata. Trusted infrastructure?

The objective is to establish an accepted (legal issues, privacy aspects) metadata specification and recommendation and identify sources for different parts.

Issue 4: Invite external experts to advise on resolving services

It is envisaged, that the underlying technology and architecture best is adapted from the general activities which is being developed and standardised as part of the Internet activities. As an example W3C supports many relevant activities, It is suggested to explore this road by inviting experts from outside the library sphere to discuss the technical aspects and to come with advice on a proper architecture.

It is suggested that KE heads a workshop, which should:

- identify the relevant Internet technologies,
- Identify the appropriate architecture and ideally
- Make decision about resolution system

To support the decision it might be relevant to perform experiments to prove the concepts. Finally, a author resolution service could be implemented.

The objective of this issue is to gain knowledge and experience operating a service based on standard Internet technology and standards.

Issue 5: Create Blueprint on author ID services

Identify author identifying services and specify requirements (person look-up, change management, techniques to use like SOA, SOAP ..) based on use cases

For each decide if it should result in a KE recommendation

Objective: To get a general understanding of ID services and what functions are required at what level: locally within universities, national, international)

Outcome: Blueprint which could be the foundation of a prototype

Identify (propose?) business model for sustainable services

Issue 6: Make report on models for ensuring a sustainable model

A document should be prepared to outline the various aspects and possibilities to ensure a sustainable model. This document can be the basis for further discussion in the KE community to get agreement on the most suitable model to implement together.

The document should contain aspects like:

- All activities cost money – who will pay?
- What is the estimated cost – for the different architectural models
- Based on national activities and funding
- Any services which has commercial value to other parties?
- Commitment to maintain IDs

4. Outcomes

The outcomes of the above if successful are

- Acceptance of the identified needs for an author ID in the KE community
- Establishment of communication lines with, and involvement of, the relevant stakeholders around technical issues
- A report outlining the potential for ensuring a sustainable model for author IDs
- A blueprint for an author ID service
- .

5. Annexes (including supporting papers, referenced materials, list of participants)

