

Optimising Interoperability in Multi-KOS Subject Searching: Framework for a Collaborative Approach?

Dennis Nicholson, Centre for Digital Library Research

Second International Seminar on Subject Access to Information, Helsinki, 29-30th November 2007

What I mean by this...

- ∅ I want to suggest that it is both desirable and possible to agree a model set of requirements for subject interoperability service design – and to use it as the basis of an associated common research and development agenda that will facilitate gradual cumulative and collaborative progress towards optimising subject interoperability across the networked world.

What is it?

- ∅ In rough outline, a subject interoperability service is a service designed to facilitate user working across information services using different KOS when applying subject descriptions to their resources
- ∅ Not simply a KOS/KOS crosswalk service – a service that uses data from such services (together with other relevant data) to transparently facilitate user working across multiple information services using multiple KOS

Clarification

- ∅ Not suggesting that every future interoperability service must adopt any model requirements set the community might agree
 - § **That the requirements set must be applied in every situation, or,**
 - § **That there aren't other ways of designing such services**
- ∅ Only that there are likely to be beneficial effects both locally and globally if significant numbers of services are designed in line with agreed model requirements

Why it is desirable

- ∅ Because optimising interoperability in multi-KOS subject searching is
 - § **A complex and difficult to tackle problem**
 - § **Of increasing concern to a wide and growing range of organisations**
 - § **Entails many common elements - so that a collaborative approach would optimise efforts by reducing duplication**
 - § **Entails many community-specific elements best tackled by and within the communities concerned**
 - § **Best tackled collaboratively**

Why a Model Requirements Set?

- ∅ Agreeing a model requirements set for subject interoperability service design would bring focus and structure to collaborative R&D efforts, by providing :
 - § **A generally accepted perspective on the nature and scope of the problem to be tackled**
 - § **Agreement on the problems and issues that need to be addressed if it is to be resolved**
 - § **An understanding of why, where, and how these relate to both the problem as a whole and each other**
 - § **A basis for reaching agreement on how success and failure should be measured in respect of both the problem itself and its individual elements**

Desirable and Possible?

Ø Difficult because no one community faces quite the same problem:

§ **The services, domains, KOS, languages, users, tasks, and crosswalk approaches mix different in every case**

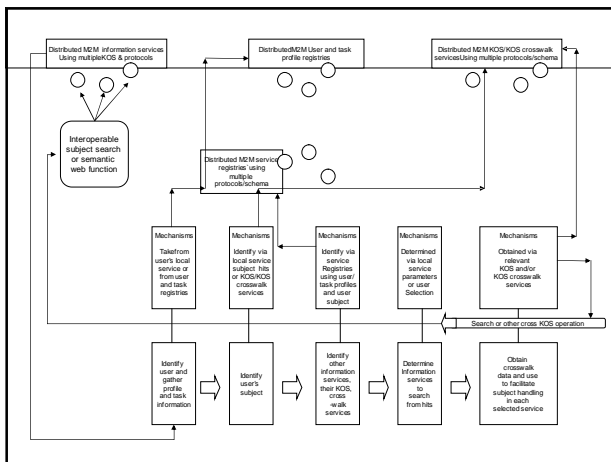
Ø However, it is possible if we aim to be inclusive and design a requirements set that assumes the use of:

§ **Work from different domains, with different, often unknown, sets of KOS and mark-ups, different approaches to 'crosswalking', different user types and tasks, different ways of finding services etc**

Ø Taking this inclusive approach leads to...

Tentative Model Requirements Set ; Ability to:

Accurately express a user's subject search in the KOS used by a given information service	Essence of any subject Interoperability service
Identify other information services in a subject area; their KOS; relevant KOS crosswalk services	Known? Need to discover unknown services growing
Identify a user's subject in relation to some standard scheme (KOS)	Needed to identify relevant I services and KOS terms
Offer a user and task adaptive approach	Possibly needed to identify Info/crosswalk services
Operate effectively whilst encompassing a range of distributed elements	Needed for wide registry/ service/crosswalk coverage
Offer multiple protocol and schema support	Needed to ensure inclusivity of domains & communities
Offer Machine-to-Machine (M2M) functionality	Needed for locally adaptive, user transparent services



R&D Agenda

- Ø Not rocket science – most requirements not especially new
- Ø Important because they offer possibility of allowing us to co-ordinate and direct future work in the area in a largely 'hands-off' fashion
- Ø The requirements set just described suggests a need for R&D in a wide variety of areas...
- Ø ...with the following sets being illustrative rather than necessarily complete and well-structured

Developments suggested includes:

- Ø Alternative mechanisms for or approaches to
 - § **Identifying, storing, and processing user profiling data** (e.g. profiles held in local information service, central service profiles, a mix of the two (local refines central))
 - § **Identifying, storing, and processing task profiling data** (e.g. profiles held in local information service, central service profiles, a mix of the two (local refines central))
 - § **Identifying the subject of a user's search** (e.g. user choice from a hierarchy, classification information in good local hits, or finding user input term in preferred or non-preferred terms in locally or remotely held KOS)

Developments suggested includes:

- Ø Alternative mechanisms for or approaches to
 - § **Identifying information services with subject coverage relevant to a given user's subject search** (e.g. Local list used for all; local list used for given user and task types; use of user and task profile and user subject to identify services in remote registry of information services classified according to subject coverage; user subject browse in a registry; use of a subject strength service)
 - § **Identifying which KOS and KOS version is in use in a given information service** (e.g. Locally held information; registry held information; information held at service site)

Developments suggested includes:

- Ø Alternative mechanisms for or approaches to
 - § **Identifying which KOS or KOS crosswalk service to use to identify the terms to use for the user's subject search in a given information service using a given unknown KOS** (e.g. Local data; service registry using KOS and user and task profile; using a terminology registry instead of a service registry)
 - § **Storing, serving up, or processing different types of KOS crosswalking data** (automated v intellectual; different spines or none, different levels of mapping granularity, and so on)

Developments suggested includes:

- Ø Alternative mechanisms for or approaches to
 - § **Describing, categorising, storing, serving up, and processing data about KOS and KOS crosswalking services** (automated v intellectual; different spines or none, different levels of mapping granularity, etc)
 - § **Handling interoperability data obtained from these KOS and KOS crosswalking services and presenting it helpfully to users**
 - § **User interaction (or lack of it) in all of these processes**
 - § **Screen design in all of these areas**
 - § **Providing all of the above in different programming environments**

Research suggested includes:

- Ø Investigations into
 - § **The user characteristics that are relevant to profiling user types from a subject retrieval perspective** (and how best to categorise and express them for operational purposes)
 - § **The various retrieval tasks facing users** (and how best to categorise and express them for operational purposes)
 - § **How user and task profiles impact on retrieval requirements** (in a variety of contexts)

Research suggested includes:

- Ø Investigations into
 - § **Mapping user input to subject in one or more standard schemes: determining what works best**
 - § **Using user's subject** (and user and task profiles) **to identify information services with subject coverage relevant to a user's subject search** (and their KOS)
 - § **Operational program-level requirements associated with KOS and KOS version identification and categorisation**

Research suggested includes:

- Ø Investigations into:
 - § **Using KOS** (and user and task profiles) **to identify appropriate KOS crosswalk services** (and their characteristics)
 - § **The relative effectiveness of a range of approaches to facilitating inter-KOS 'crosswalking'** (automated v intellectual; different spines or none, different levels of mapping granularity, and so on)

Research suggested includes:

- Ø Investigations into:
 - § **How distributed KOS/KOS crosswalk services** (utilising a range of different approaches and providing a range of different effectiveness levels) **can be described in a registry** (in such a way as to permit local user interfaces utilising M2M interaction to select, connect, and interact with each so as to best meet their users' needs)
 - § **The mechanics of requesting and obtaining interoperability data and using it** (with user and task profiles and interoperability service characteristics) **to facilitate subject interoperability**

Research suggested includes:

- ∅ Investigations into:
 - § **Inter-KOS mapping issues, such as the range of mapping types needed for an effective service**
 - § **The usefulness or otherwise of providing user feedback on mapping types**
 - § **The special challenges of M2M-based interaction, (including technical, operational, and user issues)**
 - § **The minimum protocol/schema requirements of a fully functional KOS/KOS crosswalk service**

Research suggested includes:

- ∅ Investigations into:
 - § **Cost-effectiveness.** How should we measure it? Does it vary according to user and task type?
 - § **Functional effectiveness.** Are there different levels? How do these compare for cost-effectiveness? How can they be usefully categorised to allow useful M2M interaction?
 - § **Manageability issues in a globally distributed as opposed to a centralised service.** How do they compare for cost and functional effectiveness?

Research suggested includes:

- ∅ Investigations into:
 - § **The different needs and perspectives in different communities** (Archives v Libraries v Museums and others; information retrieval v semantic web)
 - § **User interface design issues in relation to all or most of the above**
 - § **User behaviour issues for all or most of the above**

A Basis for Agreement?

- ∅ Is the model requirements set I described a basis for agreement?
- ∅ Unknown, but useful either way
- ∅ If it can be a basis for agreement, it can provide a basis for a common R&D agenda as we've just seen
- ∅ If not, it can, at least be an initial focus for discussion that might lead to a set we can all agree on, or, failing that, a set of options that we need further research to choose between

What are the benefits?

- ∅ The community can work together in a 'hands-off' and devolved and co-ordinated fashion
- ∅ Division of labour will facilitate and expedite R&D
- ∅ Reduced duplication of effort in respect of common mechanisms and research problems
- ∅ An inclusive approach ensuring involvement across all interested domains
- ∅ Each new KOS/KOS crosswalk service will enrich and extend the whole

A Practical Proposition?

- ∅ Are services based on the model requirements set known to be a feasible proposition?
 - § **At a basic level, yes. Phase III of HILT built a pilot subject interoperability service that included working mechanisms that:**
 - ü Identified a user's subject,
 - ü Used it to identify information services and their KOS in a registry,
 - ü Drew crosswalk data from a central terminology server,
 - ü Used it to facilitate user searches in a service using terms from the KOS appropriate to a given service

A Practical Proposition?

- § **It didn't implement user / task profiling** (but there is no reason to suppose that there are insurmountable technical barriers to that).
- § **It didn't implement distributed KOS crosswalk services, but the use of a service registry to identify information services by subject proved the mechanism** (and we intend to work with OCLC in HILT phase IV to implement a pilot version of distributed crosswalks in action).
- § **As just shown, a good deal of additional work is required on detailed workability** (but there are no grounds for believing that a service based on the requirements set isn't workable at a basic level)

Thanks for listening!

ØEmail: d.m.nicholson@strath.ac.uk

ØHILT: <http://hilt.cdlr.strath.ac.uk/>

ØHILT IV: <http://hilt.cdlr.strath.ac.uk/hilt4/index.html>

ØCDLR: <http://cdlr.strath.ac.uk/projects.html>