

Thomas Krichel (Long Island University*, NY, and RePEc)
<http://openlib.org/home/krichel>

Simeon S. Warner (Cornell University and arXiv)

* we are recruiting four faculty!

structure of talk

- general introduction
- history
- current development

background

Simeon Warner: second in command to arXiv.
<http://www.arXiv.org>

Thomas Krichel: founder of RePEc, the second-largest author-self archiving initiative.
<http://www.repec.org>

but in fact I don't like the term

a good word

Free online scholarship (Peter Suber)
Academic self-documentation (Thomas Krichel)
ASD is not per-se a metadata problem...

a good quote

One thing that I know about metadata, it is very expensive to produce...

William Y. Arms

so we need a strategy to make free metadata available.

We have a dream

In an ideal steady state, there a big data set of all that is relevant to academia.

Work that is not in the dataset is considered not exist.

All academics contribute to the collection.

Database is almost self-sustaining i.e. there is no need to fund the dataset.

a strategy for free metadata

- have contributors give the data for free
- have a large dataset through multiple contributors
- distribute data widely, as to maximize benefits from contribution

business model first, metadata design after.

business model

To understand the motivation of the contributor is crucial.

Metadata must be intuitive such that elementary data can be collected.

Metadata must be sufficiently rich to encode the whole of the academic reality, or at least the most important aspects.

This debate between simplicity and comprehensiveness is of course old.

DC / AMF strategy

DC: Simple, but extensible resource description.

AMF: Simple but relational description of resources and non-resource

resource: anything digital or digitisable.

In academia, the description of non-resources is crucial.

simple descriptive model of academia

resource collection

person institution

relations between elements in the table are crucial.

Each item in the metadata table a "describable entity".

AMF History

Comes from the Open Archives Initiative.

Santa Fe convention of 2000 had on "Open Archives Metadata Set".

OAI usage out of the scope of scholarly communication, OAI DC (simple DC with XML encoding) replaces OAMS.

At OAI technical group meeting Working Group is appointed, consisting of us.

Thomas becomes visiting professor at Hitotsubashi University, starts writing paper in THIS very building.

History obliges

to be used in OAI-PMH, we need to have

- XML encoding
- XML schema constraintability
- DC and VCard as re-usable vocabularies

and it must be simple, remember...

simple analogy with natural language

noun "describable entity"

adjective characteristic of a noun

verb links one noun to another

at least that is the basic theory. In theory, theory and practice are the same

but in practice they are not

resources are represented by different nouns by resource type, using the DC type vocabulary

currently only a "text" noun is defined

person and institution are different nouns but have same attributes

some attributes have to be grouped, e.g. citation data

AMF not final

official specification lives at

<http://amf.openlib.org/doc/ebisu.html>

suggestions and collaboration welcome, all implementors considered as creator of AMF.

testing

- RePEc OAI gateway at <http://oai.repec.openlib.org>
- OpCit export of citation data to arXiv
 - OpCit produce metadata for arXiv AMF
 - export AMF back to arXiv
 - all that through OAI

Thanks for your attention

Thomas Krichei
Palmer School of Library and Information Science
720, Northern Boulevard
Brookville, NY 11548-1300
USA
<http://openlib.org/home/krichei>