# Repositories thru the looking glass

Andy Powell
Eduserv Foundation
andy.powell@eduserv.org.uk
www.eduserv.org.uk

eduserv

eduserv

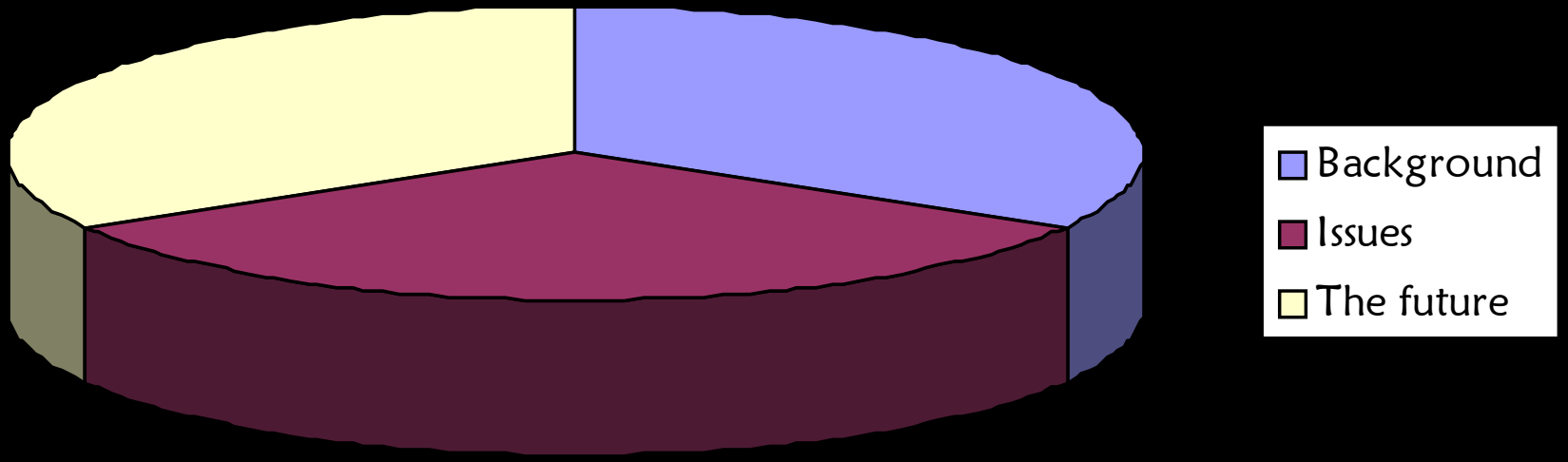
There are many methods for predicting the future. For example, you can read horoscopes, tea leaves, tarot cards, or crystal balls. Collectively, these methods are known as “nutty methods.” Or you can put well-researched facts into sophisticated computer models, more commonly referred to as “a complete waste of time”.

Dilbert

eduserv

Either that wallpaper goes or I do.

Oscar Wilde's last words


Background
Issues
The future


some background…

eduserv

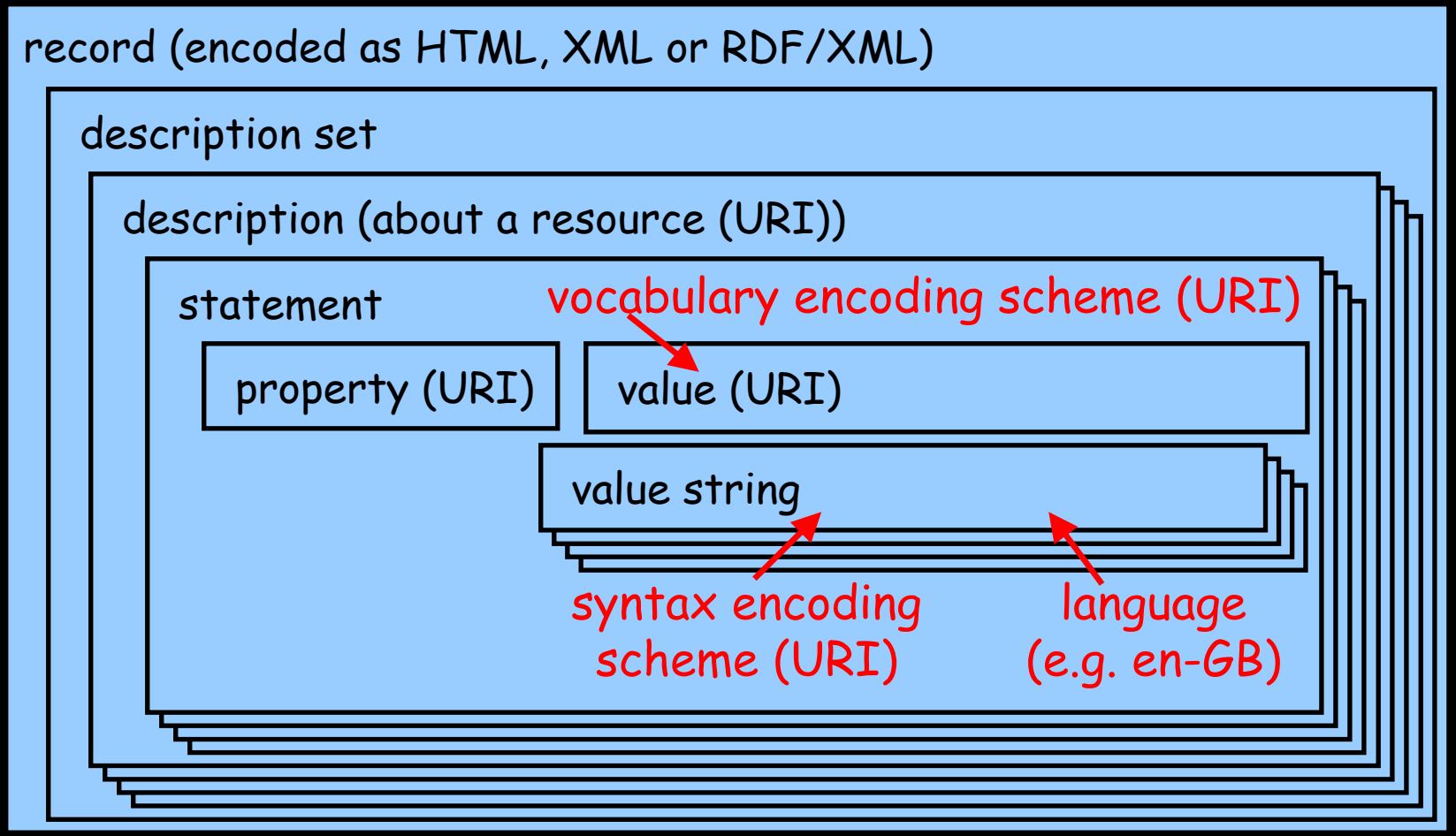
# The DCMI Abstract Model


- a set of rules defining how DC metadata descriptions are constructed
  - A description is made up of one or more statements …
  - Each statement instantiates a property/value pair and is made up of …
  - …
  - Each value string is a simple, human-readable string …
  - …
- a set of human-readable statements (as per above)
- also formalised using UML

# The DCMI Abstract Model

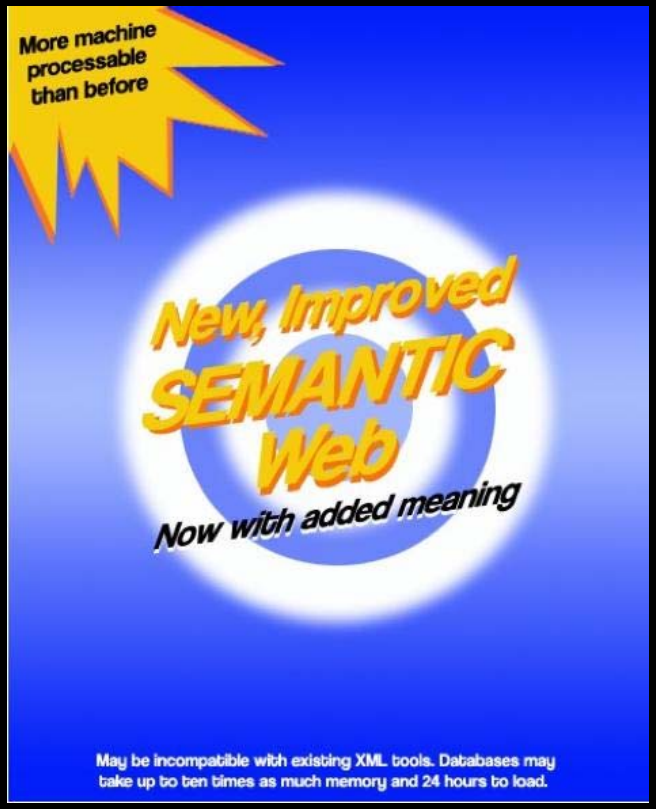

- independent of particular syntaxes
- but descriptions that comply with the model can be encoded using any of the recognised DCMI encodings
  - i.e. XHTML, XML and RDF
- simple
  - largely based on resource, property, value triple
  - formally mapped to the RDF model
- highly extensible

# The DCMI Abstract Model



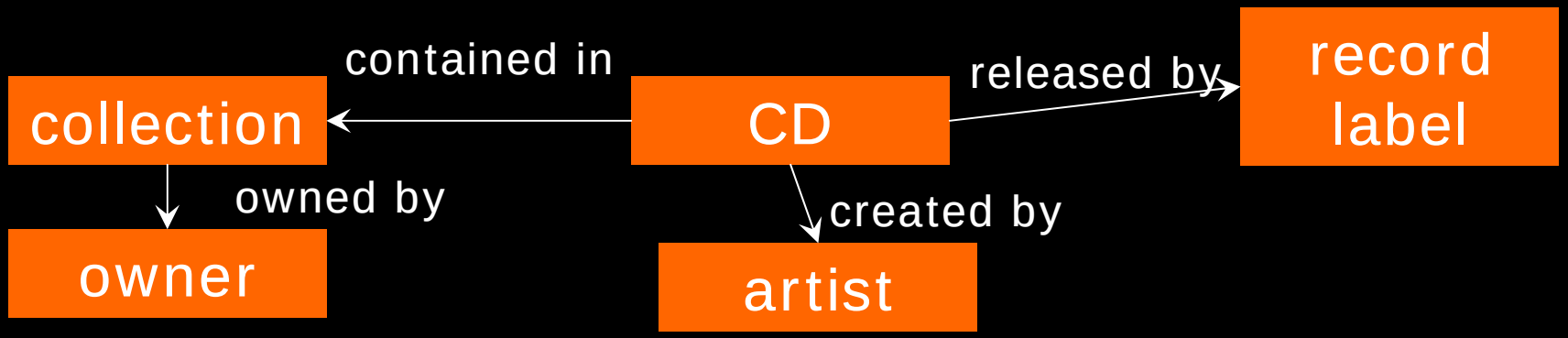
# The DCMI Abstract Model

- relationships between the **descriptions** in a **description set** and the **resources** being described made explicit
- oddly, most metadata standards do **not** do this
- DC application profiles now start by defining which set of resources are being described…
- …then assigning the set of properties and so on that will be used to describe them



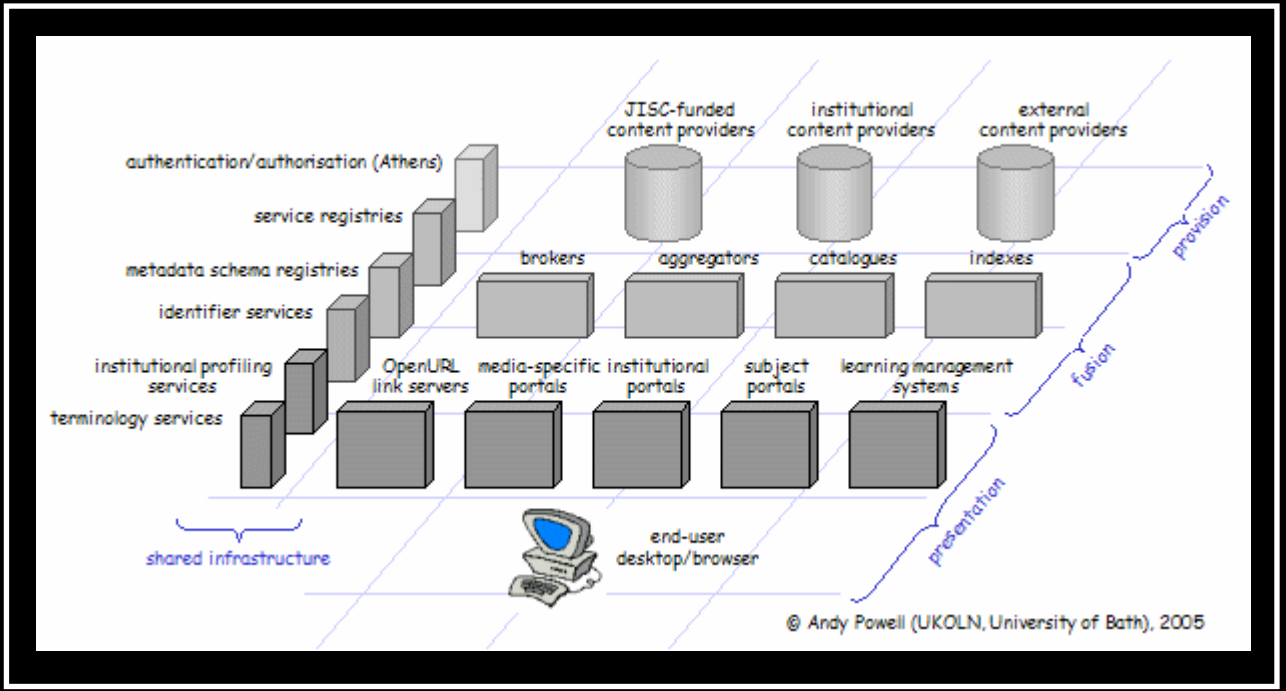
# E.g. an application profile for CDs

- start with the set of **entities** that we want to describe
- and the key **relationships** between those entities
- e.g. a CD collection entity/relationship model…




- then define a set of properties for each

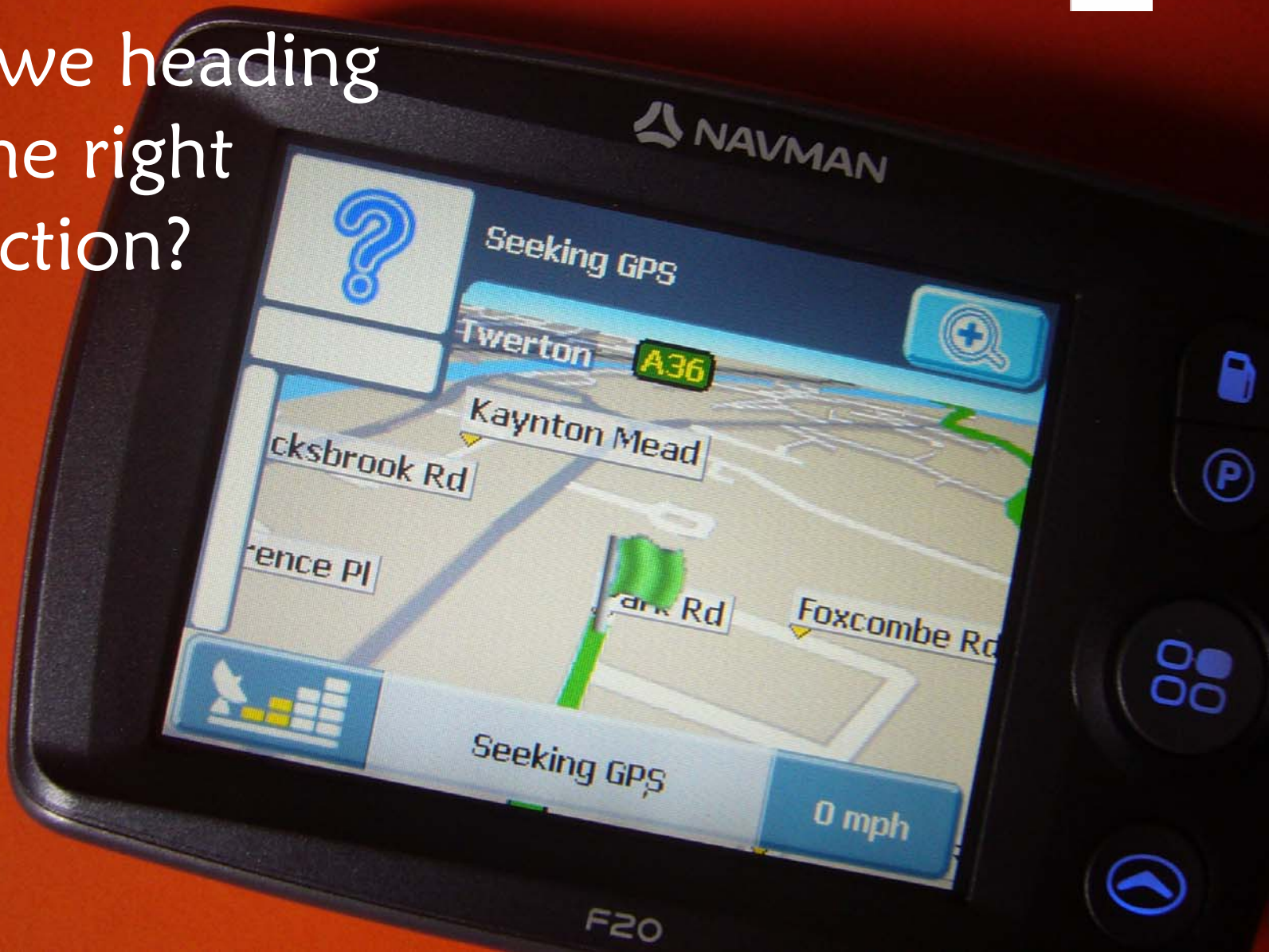
# JISC Information Environment




eduserv

eduserv

# are we heading in the right direction?




open access

not ‘if’ but
‘when’

eduserv

# 3 issues...

eduserv


issue #1
have we got our
terminology right?

eduserv

a university-based **institutional repository** is a set of services that a university offers to the members of its community for the **management and dissemination of digital materials** created by the institution and its community members. It is most essentially an organizational commitment to the **stewardship of these digital materials, including long-term preservation where appropriate, as well as organization and access or distribution.** … An institutional repository is **not simply a fixed set of software and hardware**

(Cliff Lynch, 2003)

eduserv

a focus on ‘making content available on the Web’ would be more intuitive to researchers

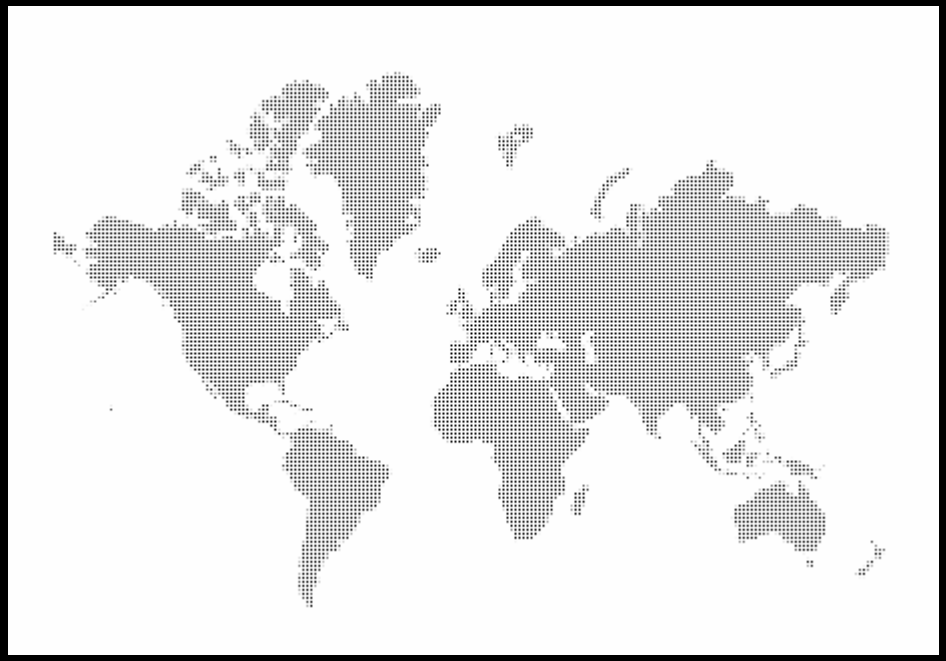
- a focus on '**content management**' would change our emphasis
- OAI-PMH out…
- search engine optimisation, usability, accessibility, tagging, information architecture, cool URIs in…

eduserv

# issue #2
# service oriented
# vs.
# resource oriented


REST = Representational State Transfer

an architectural style with a focus on resources, their identifiers (e.g. URIs), and a simple uniform set of operations that each resource supports (e.g. GET, PUT, POST, DELETE)



# issue #3
# national vs. global

# The impact of Web 2.0

prosumer

remote apps

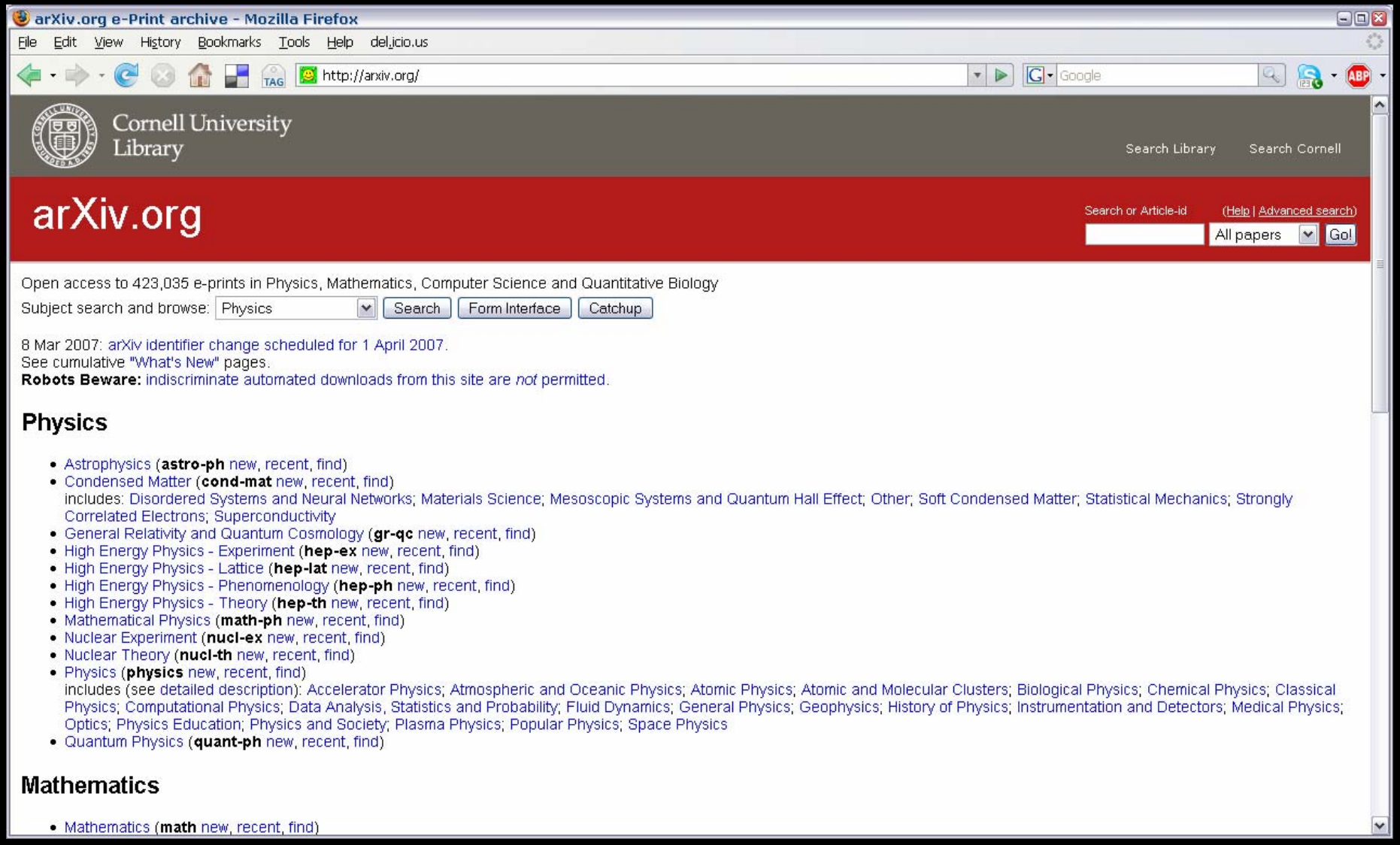
social

API

diffusion

concentration

eduserv


arXiv.org e-Print archive - Mozilla Firefox
File Edit View History Bookmarks Tools Help del.icio.us
http://arxiv.org/
Google
Cornell University Library
Search Library
Search Cornell
arXiv.org
Search or Article-id
(Help | Advanced search)
All papers
Go!
Open access to 423,035 e-prints in Physics, Mathematics, Computer Science and Quantitative Biology
Subject search and browse: Physics
Search
Form Interface
Catchup
8 Mar 2007: arXiv identifier change scheduled for 1 April 2007.
See cumulative "What's New" pages.
Robots Beware: indiscriminate automated downloads from this site are not permitted.
Physics
Astrophysics (astro-ph new, recent, find)
Condensed Matter (cond-mat new, recent, find)
includes: Disordered Systems and Neural Networks; Materials Science; Mesoscopic Systems and Quantum Hall Effect; Other; Soft Condensed Matter; Statistical Mechanics; Strongly Correlated Electrons; Superconductivity
General Relativity and Quantum Cosmology (gr-qc new, recent, find)
High Energy Physics - Experiment (hep-ex new, recent, find)
High Energy Physics - Lattice (hep-lat new, recent, find)
High Energy Physics - Phenomenology (hep-ph new, recent, find)
High Energy Physics - Theory (hep-th new, recent, find)
Mathematical Physics (math-ph new, recent, find)
Nuclear Experiment (nucl-ex new, recent, find)
Nuclear Theory (nucl-th new, recent, find)
Physics (physics new, recent, find)
includes (see detailed description): Accelerator Physics; Atmospheric and Oceanic Physics; Atomic Physics; Atomic and Molecular Clusters; Biological Physics; Chemical Physics; Classical Physics; Computational Physics; Data Analysis, Statistics and Probability; Fluid Dynamics; General Physics; Geophysics; History of Physics; Instrumentation and Detectors; Medical Physics; Optics; Physics Education; Physics and Society; Plasma Physics; Popular Physics; Space Physics
Quantum Physics (quant-ph new, recent, find)
Mathematics
Mathematics (math new, recent, find)

eduserv

# thinking about the future...

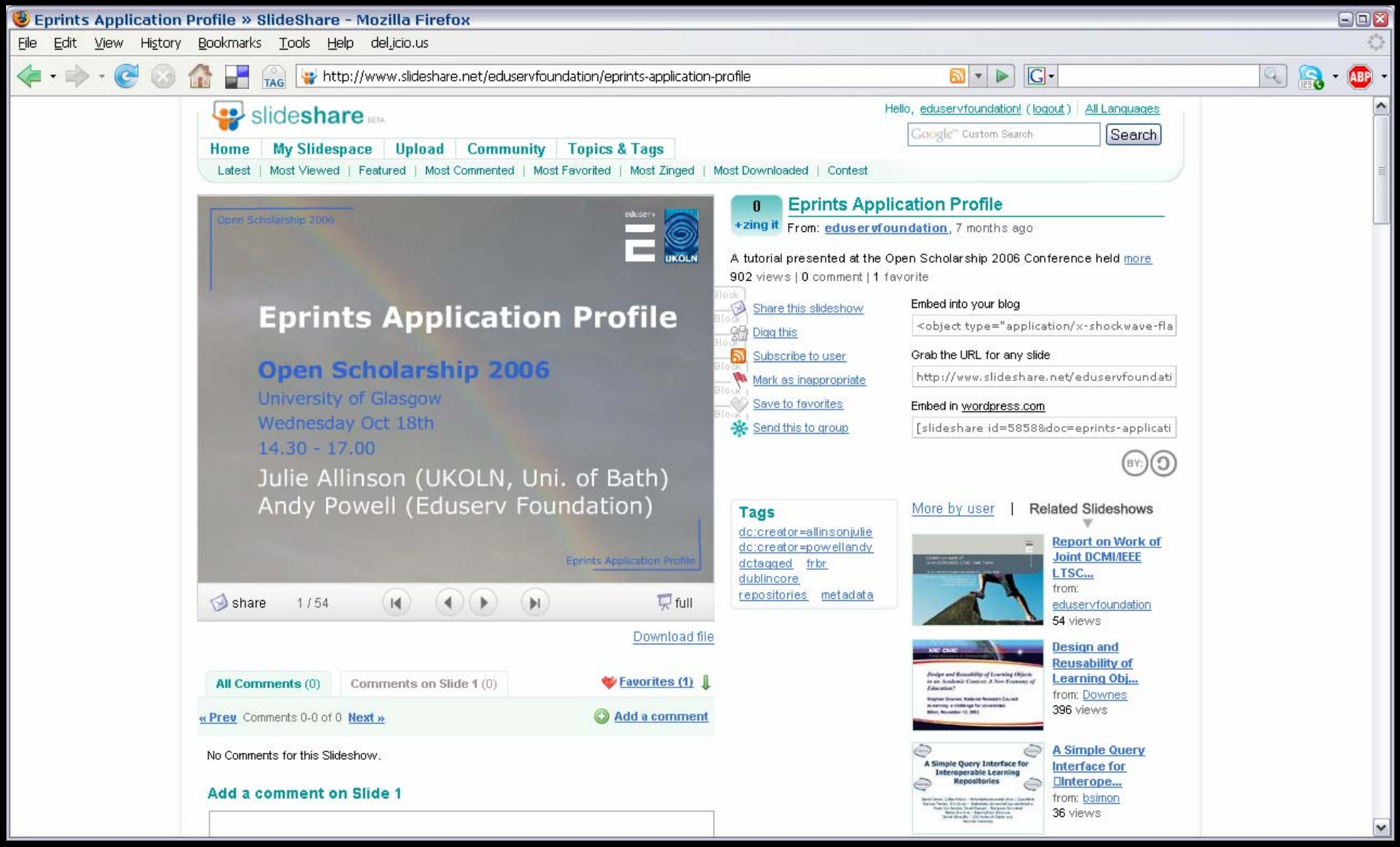
eduserv

1. what would a Web 2.0 repository look like?
2. potential impact of the Semantic Web on repositories

eduserv

1. what would a Web 2.0 repository look like?
2. potential impact of the Semantic Web on repositories

eduserv


Eprints Application Profile » SlideShare - Mozilla Firefox
File Edit View History Bookmarks Tools Help del.icio.us
http://www.slideshare.net/eduservfoundation/eprints-application-profile
slideshare BETA
Hello, eduservfoundation! (logout) | All Languages
Search
Home | My Slidespace | Upload | Community | Topics & Tags
Latest | Most Viewed | Featured | Most Commented | Most Favorited | Most Zinged | Most Downloaded | Contest
Open Scholarship 2006
Eprints Application Profile
Open Scholarship 2006
University of Glasgow
Wednesday Oct 18th
14.30 - 17.00
Julie Allinson (UKOLN, Uni. of Bath)
Andy Powell (Eduserv Foundation)
Eprints Application Profile
share 1 / 54 full
Download file
All Comments (0) Comments on Slide 1 (0) Favorites (1)
« Prev Comments 0-0 of 0 Next » Add a comment
No Comments for this Slideshow.
Add a comment on Slide 1
0 +zing it
Eprints Application Profile
From: eduservfoundation, 7 months ago
A tutorial presented at the Open Scholarship 2006 Conference held more
902 views | 0 comment | 1 favorite
Share this slideshow
Digg this
Subscribe to user
Mark as inappropriate
Save to favorites
Send this to group
Embed into your blog
<object type="application/x-shockwave-fla
Grab the URL for any slide
http://www.slideshare.net/eduservfoundati
Embed in wordpress.com
[slideshare id=5858&doc=eprints-applicati
Tags
dc:creator=allinsonjulie
dc:creator=powellandy
dctagged frbr
dublincore
repositories metadata
More by user | Related Slideshows
Report on Work of Joint DCMI/IEEE LTSC...
from: eduservfoundation
54 views
Design and Reusability of Learning Obj...
from: Downes
396 views
A Simple Query Interface for Interoperable Learning Repositories
A Simple Query Interface for Interope...
from: bsimon
36 views


- high-quality browser-based document viewer (not Acrobat!)
- tagging, commentary, more-like-this, favorites, …
- persistent (cool) URIs to content
- ability to form simple social groups
- ability to embed documents in other Web sites
- high visibility to Google
- offer RSS as primary API
- use of Amazon S3 to cope with scalability

eduserv

# a Web 2.0 repository would be a global service

global concentration is an
enabler
of social interaction

# But…

- they don't do preservation
- they don't handle complex workflows
- they don't expose rich metadata
  - yes, scholarly communication has some particular functional requirements which are not met by Google…
  - author searching, citation counting, object complexity
  - not handled well by the current Web
  - how are these requirements best met? thru richer metadata?

eduserv

1. what would a Web 2.0 repository look like?
2. potential impact of the Semantic Web on repositories

eduserv

# SWAP

## The Scholarly Works Application Profile

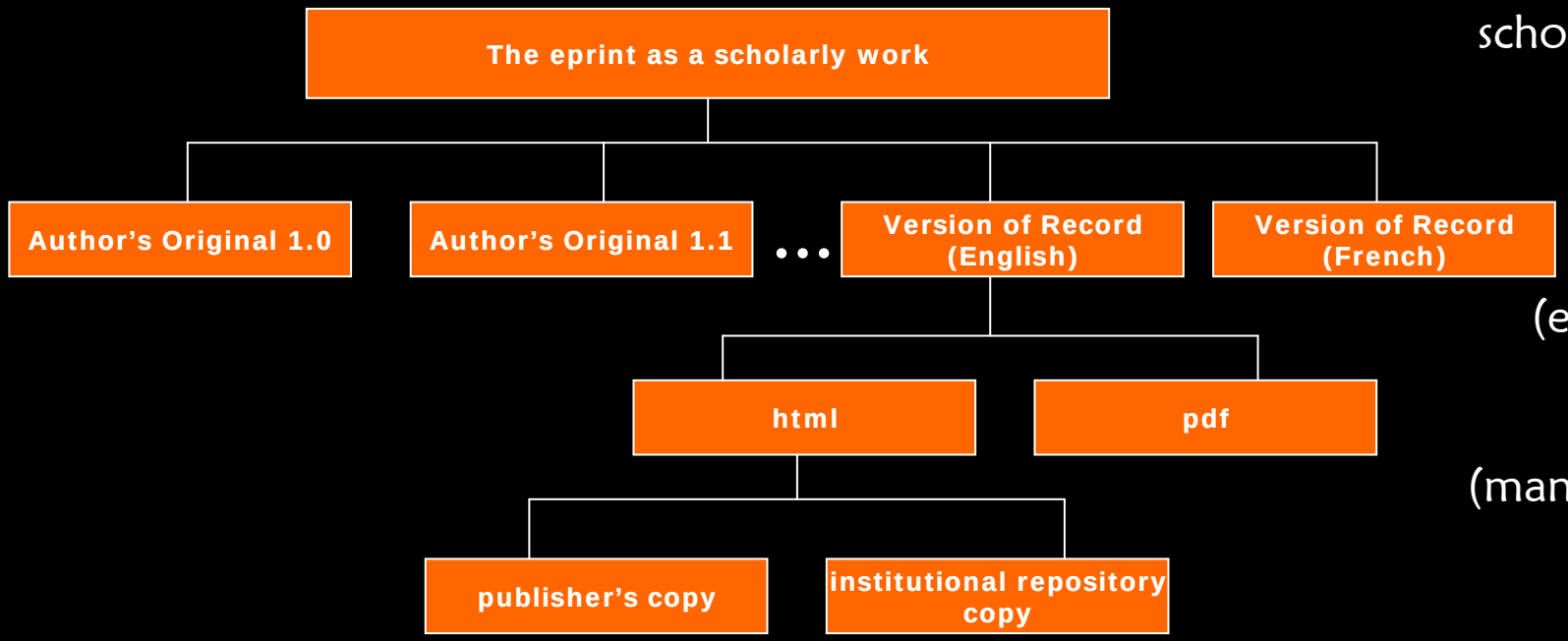
# A model based on FRBR

- Functional Requirements for Bibliographic Records
- an application model for the entities that bibliographic records are intended to describe
- FRBR models the world using 4 key entities
  - Work, Expression, Manifestation and Item

# FRBR and scholarly works

- FRBR is a useful model in the context of scholarly works (eprints) because it allows us to answer questions like
  - what is the URL of the most appropriate copy (an item) of the PDF format (a manifestation) of the pre-print version (an expression) for this eprint (the work)?
  - are these two copies related? if so, how?

# FRBR for scholarly works



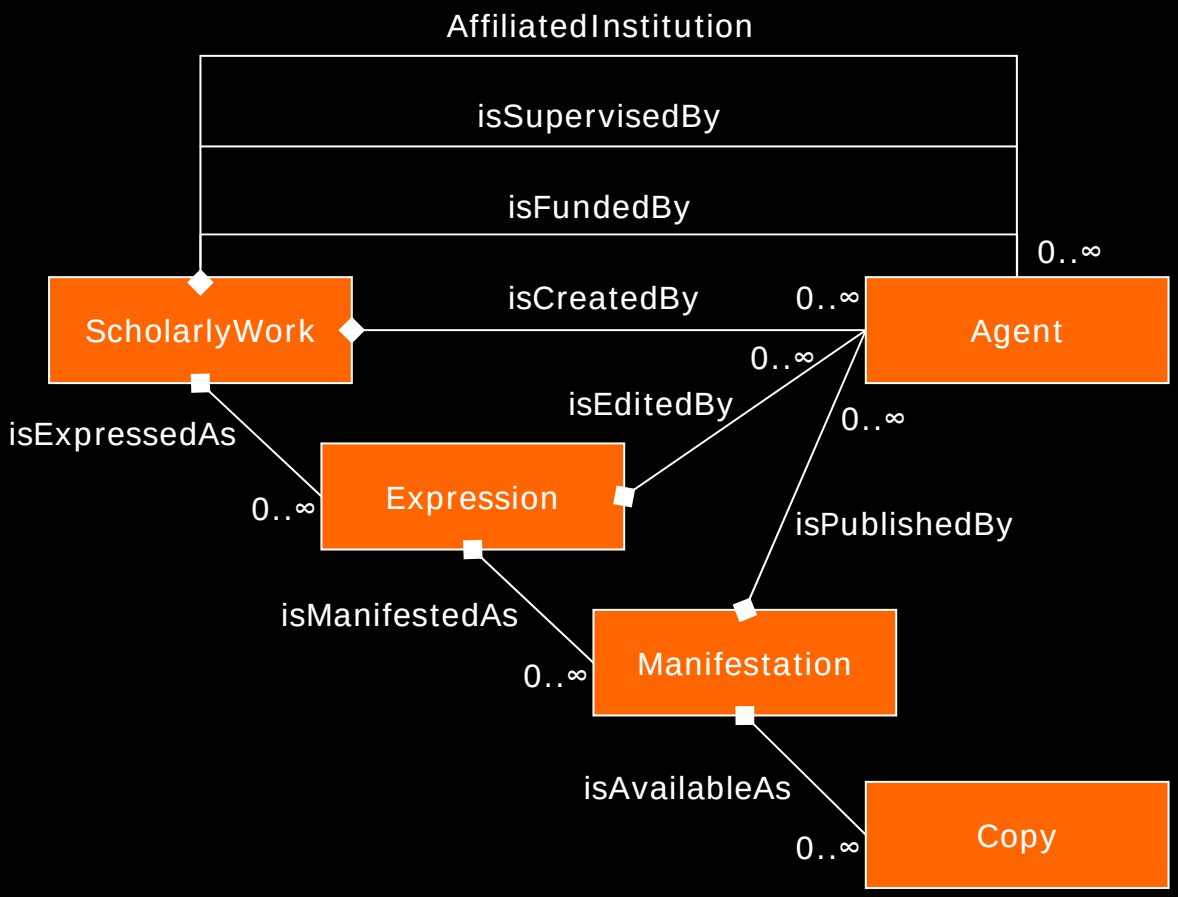

scholarly work (work)

version (expression)

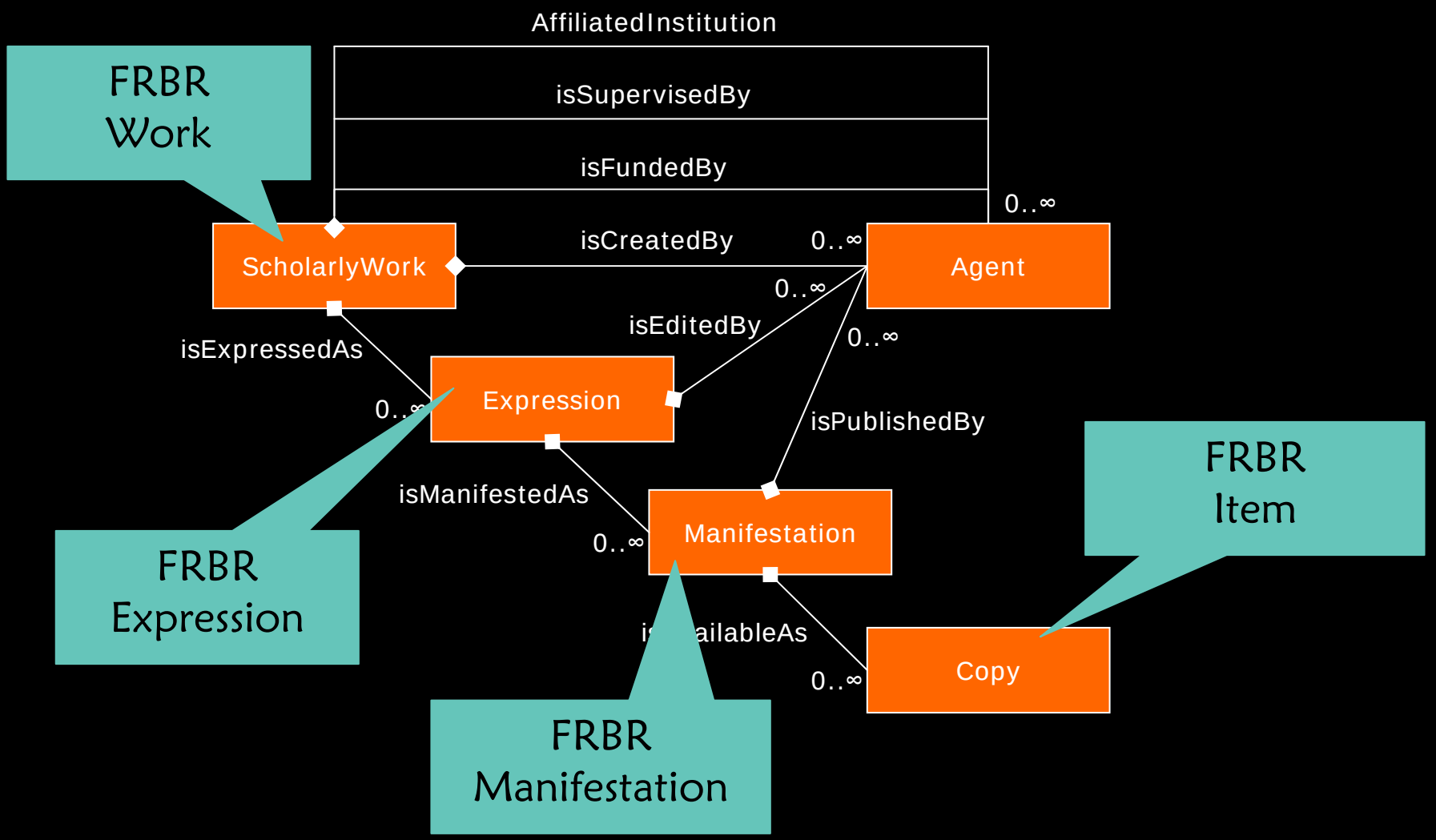
format (manifestation)

copy (item)

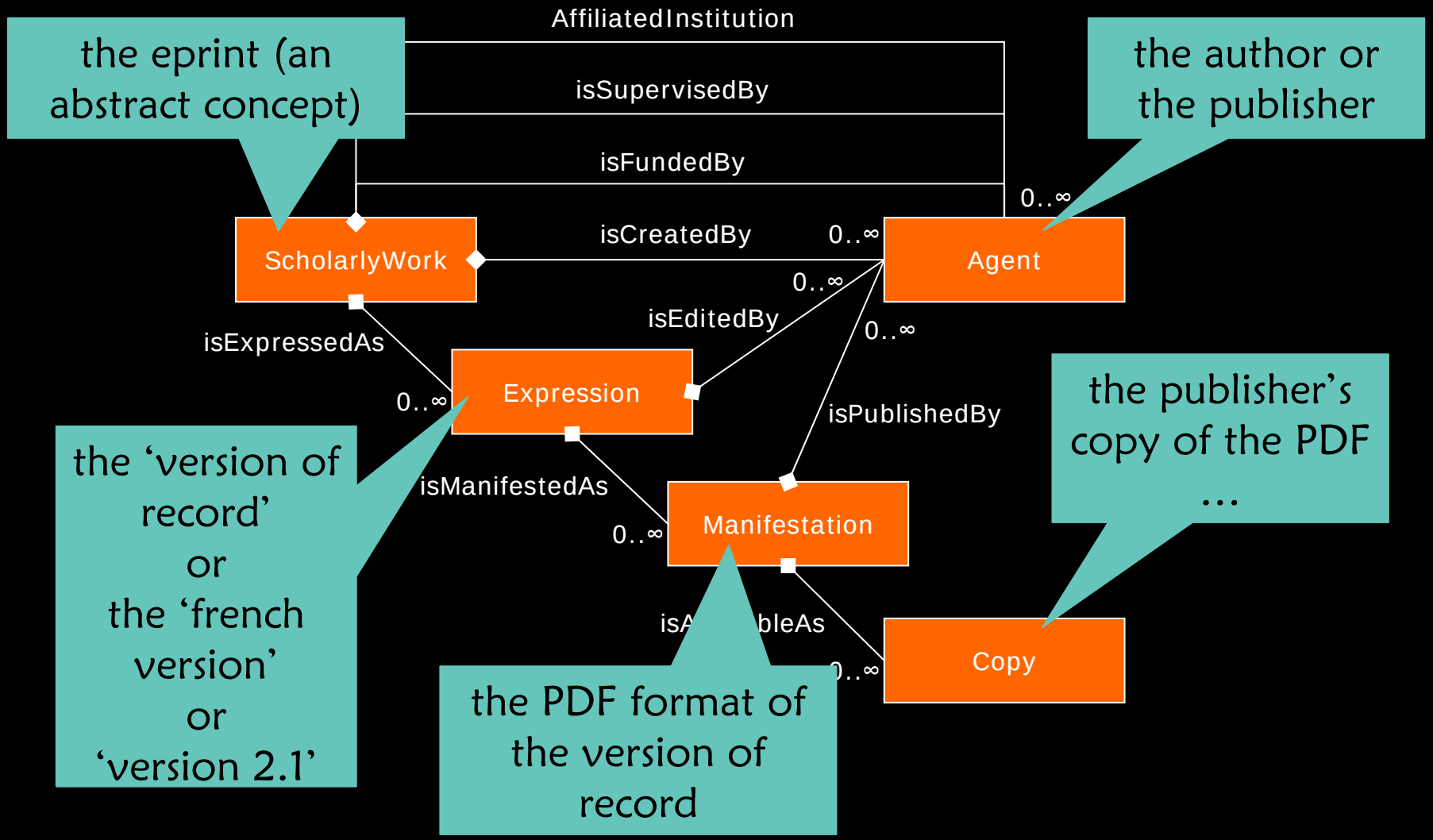
# SWAP application profile model



# SWAP and FRBR



# SWAP and FRBR



# Attributes

- the application model defines the entities and relationships
- each entity needs to be described using an agreed set of attributes

# Example attributes

**ScholarlyWork:**
title
subject
abstract
affiliated institu
identifier

**Expression:**
title
date available
status
version number
language
genre / type
copyright holder
bibliographic citation
identifier

**Manifestation:**
format
date modified

**Agent:**
name
type of agent
date of birth
mailbox
homepage
identifier

**Copy:**
date available
access rights
licence
identifier

# Final thoughts on the model

- this model makes it easier to rationalise 'traditional' and 'modern' citations
  - traditional citations tend to be made between eprint 'expressions'
  - hypertext links tend to be made between eprint 'copies' (or 'items' in FRBR terms)
- adopting a simple underlying model now may be expedient in the short term but costly to interoperability in the long term
  - the underlying model need to be as complex as it needs to be, but not more so!
- a complex underlying model may be manifest in relatively simple metadata and/or end-user interfaces
- existing eprint systems may well capture this level of detail currently – but use of simple DC stops them exposing it to others!


eduserv
time to reflect?

# Repositories

- what can we learn learn from Web 2.0?
  - user interface design matters
  - global ‘concentration’ is an enabler of social interaction
- simple DC is both too simple and too complex
- richer DC application profiles such as SWAP may be a way forward
- but need to ensure that their use does not over-complicate user interfaces and workflows

# Open Access

- in policy terms - talking about the aim, "**making content available on the Web**" would be much better than the objective, "**putting content in repositories**"


more generally…
resource orientation
REST
Semantic Web
Web architecture
…are important
digital libraries ignore
them at their peril

eduserv

# Thank you

images by eNil, Poppyseed Bandits, m o d e, striatic, estherase, Gen Kanai, //bwr - Hieronymus Karl Frederick, dullhunk, Today is a good day (all @Flickr), and yours truly