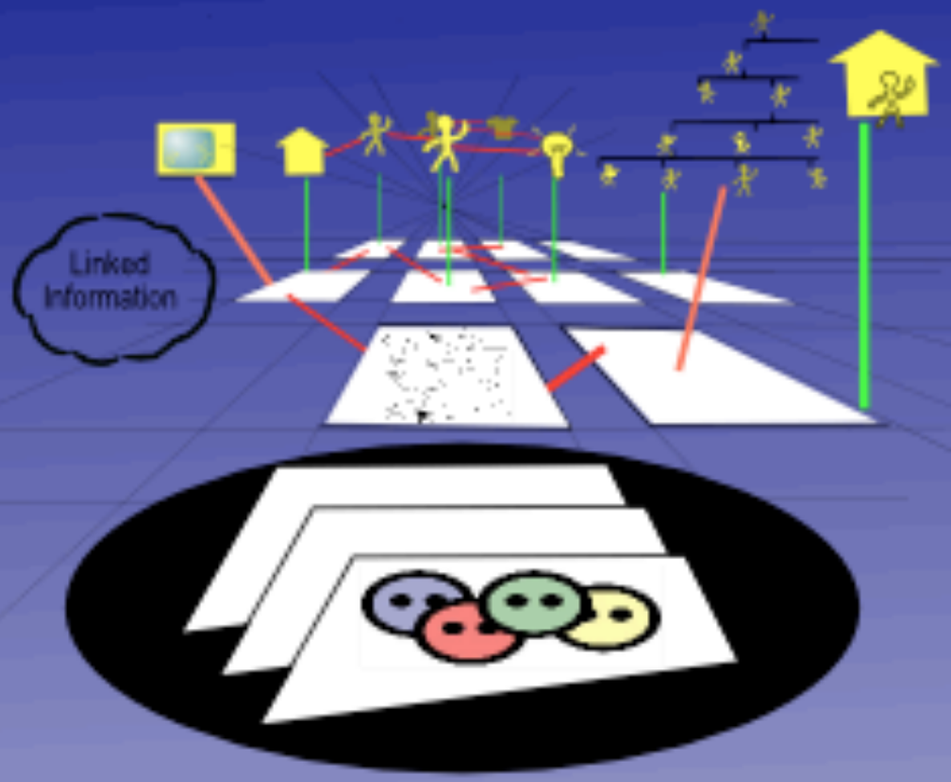


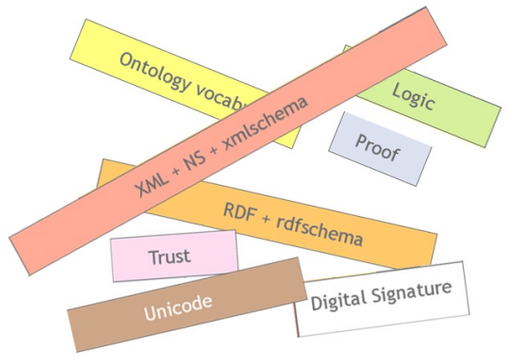
Understanding the standards gap

LIDA 2009 - Libraries in the Digital Age
Dublin Core Tools Community workshop
30 May, Zadar, Croatia

Dan Brickley

Dublin Core RDF tools in theory & practice....

Standards Gap =
what you want
minus
what you get



What RDF's standards do "out of the box"
...and what we have to build ourselves.

A brief introduction to Semantic Web thinking,
and a proposal for role of the DC Tools group

If you buy, install or build RDF/DC "standards-compliant" metadata tools, what's still missing? how to fill that gap?

Two conceptual tools:

1. Understand the RDF & Semantic Web style -

   Information-linking using Web technology

2. Understand Dublin Core history: DC as element set vs DC as a community meeting place -

   from 15 DC metadata elements as 'the' solution,
   to DC community as place to share & find solutions

The Semantic Web project

A true story about linked information systems:

From many small pieces of information ("*claims*")...

...to the total universe of information ("*The Web*").

Claims? as properties, relationships and attributes...

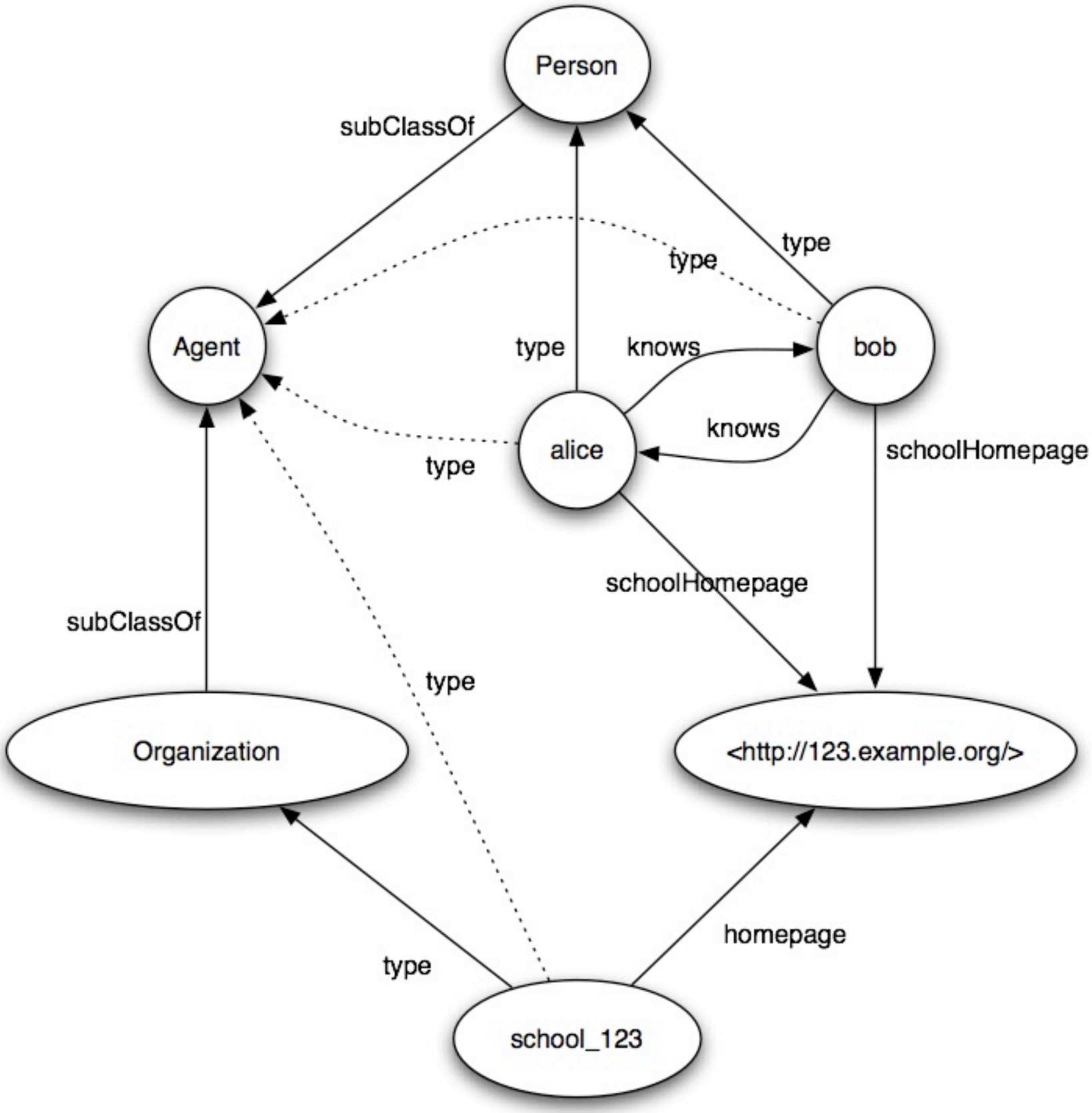
#me	**name**	‘Dan Brickley’
#me	**mbox**	danbri@danbri.org
#me	**workplaceHomepage**	http://www.vu.nl/
#me	**currentProject**	http://notube.tv/
#me	**workplaceHomepage**	http://www.fao.org/
#me	**currentProject**	http://www.fao.org/agris/
#me	**currentProject**	http://www.foaf-project.org/
#me	**homepage**	http://danbri.org/
#me	**interest**	http://www.dublincore.org/
#me	**interest**	http://www.w3.org/
#vu	**homepage**	http://www.vu.nl/
#vu	**type**	Organization
#vu	**name**	‘Vrije Universiteit Amsterdam’
#me	**type**	Person
http://www.dublincore.org/	**type**	Document
#me	**schoolHomepage**	http://hginfant.schoolwebbuilder.co.uk/
#me	**schoolHomepage**	http://www.westergate.w-sussex.sch.uk/
#me	**schoolHomepage**	http://www.bristol.ac.uk/

W3C Resource Description Framework (RDF): *simple factual claims in the Web.*

(Q: are these claims useful for anything unless you know who made them?)

(Q: how do these claims relate to documents? to their creators? to hyperlinks? provenance?)

The world according to a document: claims as triples



http://hginfant.schoolwebbuilder.co.uk/

http://www.westergate.w-sussex.sch.uk/

Westergate Community School

Specialists in Visual and Performing Arts

and Six Villages Community Campus

arts colleges

Westergate Community School

Six Villages Sports Centre

Westergate Adult Education Centre

Westergate Day Nursery

Westergate Youth Wing

Contact Us

LONDON GRID FOR LEARNING

Westergate Community School: The school is open as usual

ATOMWIDE network solutions

OpenCheck is a London Grid for Learning service supported by Atomwide Ltd

http://www.bristol.ac.uk/

University of BRISTOL

Swine flu advice

Top news stories

Lib Dem leader launches green jobs scheme in Bristol
26 May 09

Churches to chime for Charter Day
22 May 09

Green Impact Awards — winners announced
22 May 09

Latest news »

Research news »

Writing constitutions into British history

18 June

2009 CENTENARY LECTURE

Study at Bristol >>

- Undergraduate
  > including advice about money
- Postgraduate
- Lifelong learning

Information for:	Information about:
Prospective students	Faculties and departments
Current students	Research

http://www.notube.tv/

Acronym Guide part 1:

RDF/XML, RDFa, GRDDL, N3/Turtle standards

These are all ways of encoding simple 3-part claims in various Web documents.

That is all that they do.

So different documents can make different claims?

Yes.

knows

schoolHomepage

homepage

knows

schoolHomepage

homepage

schoolHomepage

€

homepage

What the computers see...

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
      xmlns="http://xmlns.com/foaf/0.1/">
  <Person rdf:about="#alice">
    <name>Alice</name>
    <knows>
      <Person rdf:about="http://localhost/notube/layer1.rdf#bob">
        <name>Bob</name>
      </Person>
    </knows>
    <schoolHomepage rdf:resource="http://lookingglass.example.org/" />
  </Person>
  <Organization>
    <homepage rdf:resource="http://lookingglass.example.org/" />
    <name>LookingGlass School</name>
  </Organization>
</rdf:RDF>
```

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
         xmlns="http://xmlns.com/foaf/0.1/">
  <Person rdf:about="#bob">
    <name>Bob</name>
    <knows>
      <Person rdf:about="http://localhost/notube/layer1.rdf#alice">
        <name>Alice</name>
      </Person>
    </knows>
    <schoolHomepage rdf:resource="http://lookingglass.example.org/" />
  </Person>
  <Organization>
    <homepage rdf:resource="http://lookingglass.example.org/" />
    <name>The LookingGlass School</n
  </Organization>
</rdf:RDF>
```

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
         xmlns="http://xmlns.com/foaf/0.1/"
     xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" >
  <Person rdf:about="http://localhost/notube/layer1.rdf#alice">
    <name>Alice</name>
    <schoolHomepage rdf:resource="http://lookingglass.example.org/" />
  </Person>
  <Organization rdf:about="#lcc">
    <homepage rdf:resource="http://lookingglass.example.org/" />
    <name>LookingGlass Community College (formerly LookingGlass School)</name>
    <rdfs:seeAlso rdf:resource="#online_store" />
    <rdfs:seeAlso rdf:resource="#alumni_directory" />
  </Organization>
</rdf:RDF>
```

3 sets of claims

Acronym Guide part 2:

SPARQL - an RDF query language

SPARQL lets you ask factual questions against a database of these claims. That is all that it does.

And the queries can talk about who made each claim?

Yes.

```
PREFIX : <http://xmlns.com/foaf/0.1/>
SELECT ?x  WHERE {
  GRAPH  <http://localhost/notube/layerlist.rdf> {
       <http://localhost/notube/layer1.rdf#alice> :made ?byalice .
    }
 GRAPH ?byalice {
      <http://localhost/notube/layer1.rdf#alice> :schoolHomepage ?x .
  }
}
```

What school does Alice say she went to?

```
PREFIX : <http://xmlns.com/foaf/0.1/>
SELECT ?g ?who WHERE {
  GRAPH  <http://localhost/notube/layerlist.rdf> {  ?who :made ?g . }
  GRAPH  ?g {    <http://localhost/notube/layer1.rdf#alice>
                 :schoolHomepage  <http://lookingglass.example.org/> .  }
}
```

Who has something to say about Alice's schooling?

- does Alice's school agree that she attended it? Does Bob's?

- when sources disagree, how to prioritise? who to believe?

This is RDF's approach - understand it and you'll understand what you get from RDF tools, RDF data, RDF people...

Linking claims with topics, authentication & authenticity....

prospects.html	dc:title	"Oil investigation - early findings"
prospects.html	dc:creator	person1
prospects.html	rdf:type	Document
person1	foaf:name	"John Smith"
prospects.html	dc:subject	sub1
person1	foaf:workplaceHomepage	<http://mining_company.example.com/>
person1	foaf:schoolHomepage	<http://miningschool.example.org/>
person1	foaf:openid	<http://miningschool.example.org/jsmith>
person1	rdf:type	Person
sub1	rdf:type	Concept
sub1	skos:prefLabel	"OFFSHORE DRILLING (MINING EXPLORATION)"
sub1	skos:notation	“62.001.4(26)” (UDC/lonclass)
sub2	skos:preLabel	"INDUSTRIAL TESTING"
sub2	skos:notation	"62.001.4" (UDC/lonclass)
sub1	skos:broader	sub2

SKOS in the Web: each concept gets a page...
SKOS is a set of terms for making claims about subjects/topics, their properties and relationships.

(Q: how might employers, academia or individuals use it to make Web-based claims about expertise?)

Beyond Toy Examples

Each Library of Congress Subject Heading has an RDF/SKOS page

... 6.5M bib records, ~200k auth records Swedish Union Catalog

Many thesauri. All of Wikipedia (dbpedia.org). Yahoo & Google.

From http://www.gutenberg.org/wiki/Gutenberg:Feeds

```
<dc:subject>
 <rdf:Bag>
  <rdf:li><dcterms:LCSH><rdf:value>Epic poetry, Greek -- Translations into English</rdf:value></dcterms:LCSH></rdf:li>
  <rdf:li><dcterms:LCSH><rdf:value>Achilles (Greek mythology) -- Poetry</rdf:value></dcterms:LCSH></rdf:li>
  <rdf:li><dcterms:LCSH><rdf:value>Trojan War -- Poetry</rdf:value></dcterms:LCSH></rdf:li>
 </rdf:Bag>
</dc:subject>
```

To using “http://id.loc.gov/authorities/**sh2007100520**”

DC Tools Community

People care about metadata standards for many reasons:

- long term archival
- integration with other collections
- lower cost of systems development
- reduce lock-in to commercial or opensource systems
- potential for alternative UIs
- consistency of interface for users

None of us get all we need from off-the-shelf tools...

So ... DC tools & the standards gap...

(the difference between what you want and what you get...)

Typical RDF off-the-shelf toolkit offers:

Parsers, XML & SQL mappings, query and rule systems, databases.

That's about it. But solid, standardised, well understood and with many independent implementations...

What more could you want?

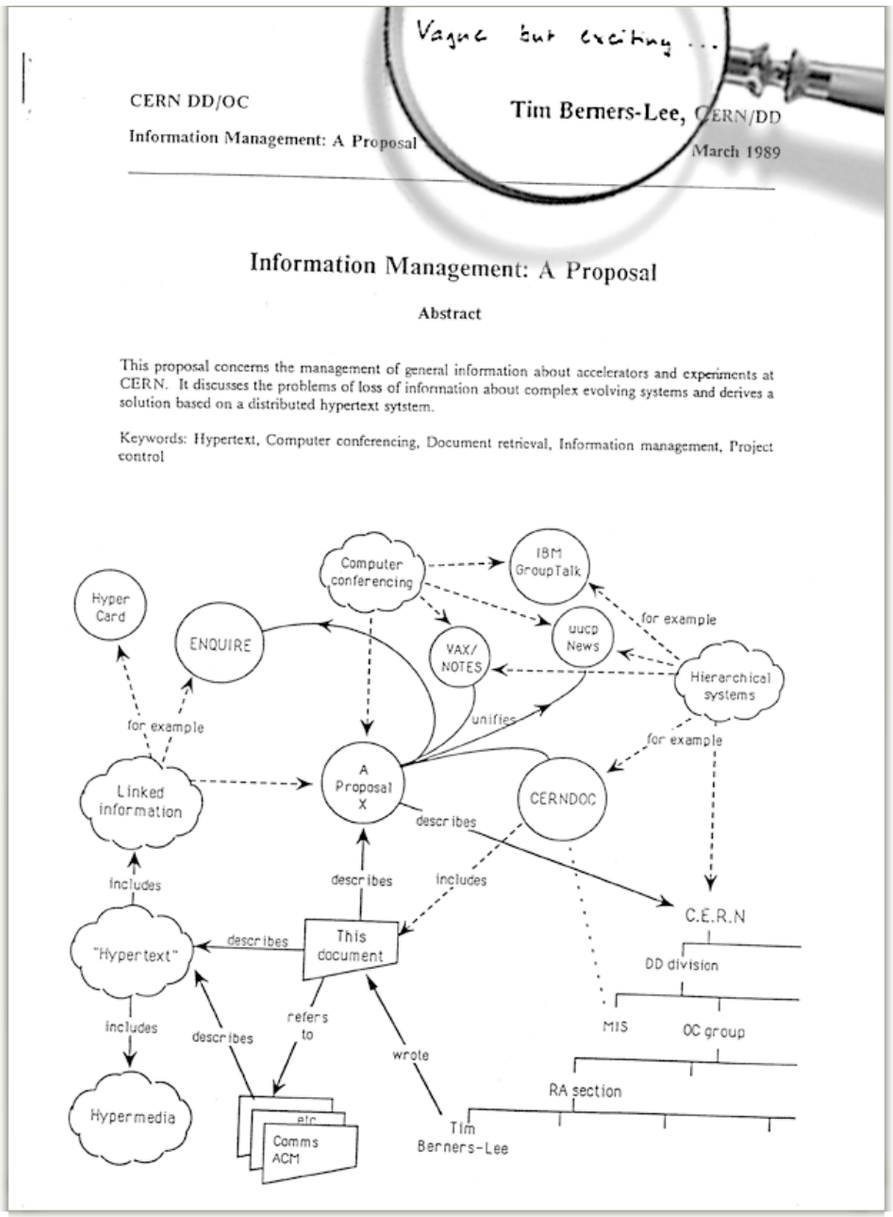
Under development... (wishlists!)

Full text search, autocompletion systems, sophisticated ranking algorithms, tagging, search and browse interfaces. Tools to check and validate conformance to 'application profiles' (ie. particular patterns of descriptions & claims). Conversion and mapping tools. Analytics. Stats. Tools that compensate for inconsistent, semi-chaotic data. Tools that help authors and publishers.[...]

... & .... documents that describe the available tools and their use cases.

Dublin Core is not a technical answer to all questions, but a place where practitioners can share answers, experiments, ideas...

Back to the future?



Vague but exciting ...

CERN DD/OC

Information Management: A Proposal

Tim Berners-Lee, CERN/DD

March 1989

Information Management: A Proposal

Abstract

This proposal concerns the management of general information about accelerators and experiments at CERN. It discusses the problems of loss of information about complex evolving systems and derives a solution based on a distributed hypertext sytstem.

Keywords: Hypertext, Computer conferencing, Document retrieval, Information management, Project control

TODO.txt

1. (re-)read "Information Management: A Proposal" (TimBL, 1989)

2. Document your own standards gap in the DC Tools community: what do you need that isn't ready yet? What have you tried? What works, what doesn't?

Questions?

mailto:danbri@danbri.org