

Using Standards-based Metadata to Support
the Software Systems Lifecycle

Anthony Finkelstein

Professor of Software Systems Engineering

University College London

A.Finkelstein@cs.ucl.ac.uk

<http://www.cs.ucl.ac.uk/staff/A.Finkelstein>



Centre for Systems Engineering

Do as I say

- **Not as I do ...**

This talk ...

- outlines the application of metadata standards to information resources in a large enterprise and presents an augmented metadata set
- shows how, with relatively little effort, the use of metadata of this form can generate significant benefits
- analyses experience to date and future direction

An industrial organisation ...

- **Challenge:** managing the information produced by the many hundreds of complex, interlocking, technical and business processes which large-scale systems engineering entails
- **Observation:** Lost documents, inconsistent information and version tangles

An industrial organisation ...

- Issue: loss of management oversight and the inability to diagnose and treat strategic issues which are raised by projects.
- Cause: the inability to be able to identify and retrieve the appropriate information resources quickly and accurately.

"Solutions"

- The "database" approach
 - All the documents are placed in a single database
 - Standard techniques from database management are used to provide views and reports on the database
 - All technical and business processes make reference to the database.

"Problems"

- vast amount of work to put in place
- assumes that it is straightforward to integrate existing processes with the database
- requires mature "local" information management practices be disrupted
- very sensitive to changes in the underlying data storage technology

"Problems"

- requires very strong centralised management control on a continuing basis
- takes no account of the scale of both the information to be stored and the diversity and complexity of the underlying processes
- presents particular difficulties where there are large volumes of "live" legacy information

"Solutions"

- The "repository" approach
 - A specialised repository is used to store fine-grain schemas for all the significant types of information
 - Can be used as a resource to perform schema integration.

"Solutions"

- Forms a valuable common reference point
- Can be used to construct an "enterprise data model" as a basis for the database approach.

"Problems"

- requires very high skills on the part of the analyst responsible for placing the schemas in the repository
- many of the information resource types are loosely structured free text
- templates and pro-formas are honoured more in the breach than in the observance

"Problems"

- there may be very many such templates for any activity, existing concurrently in many different versions, none of which has a privileged status
- a great number of document instances exist for which the original document templates or type descriptions have long disappeared

"Problems"

- there are powerful constituencies associated with specific information types and interpretations
- with large volumes of live legacy information it is not possible to start from scratch
- end result is that the repository runs the risk of becoming an idealisation bearing only tenuous relationship to the underlying information.

Metadata

- "Metadata is highly structured data that describes an information resource"
- Generally used more precisely to refer to either schema-level type descriptions of information resources or to the structured data tagged to documents and similar for the purposes of information retrieval and interpretation

Information Resource

- In the context of systems development this is a document (such as a business case, requirements specification, or process description) or composite data-object (such as UML model in a ROSE Petal file). It might even be a small database (such as an Access database belonging to an individual engineer).

Dublin Core

- Elements (label and value):
 - Title
 - Author or Creator
 - Subject and Keywords
 - Description
 - Publisher

Dublin Core

- Other Contributor
- Date
- Resource Type
- Format
- Resource Identifier
- Source

Dublin Core

- Language
- Relation
- Coverage
- Rights Management

Qualified Dublin Core

- Adds to the Dublin Core data model two concepts: the element qualifier that refines the meaning of the element (a type); and, the value qualifier that refines or enhances the interpretation of the value (a scheme)

Challenge

- Instantiate Qualified Dublin Core for an intranet (systems development) setting

Defining a Metadata Set

- Warwick Framework
- ISO11179 (ISO, 1997)
- Resource Description Format (RDF)
- eXtensible Markup Language (XML)

How I did it ...

- Title
 - As vanilla DC
- Author or Creator
 - Using vCards

How I did it ...

- **Subject and Keywords**

- **Reference to controlled ontologies**

- telecommunications products & systems (generic) - Termite/TMN
 - business processes (generic) - SAP Solution Maps
 - telecommunications products & systems (organisation-specific) - Atlas
 - business processes (organisation-specific) - Process Library

How I did it ...

- **Description**

- As vanilla DC

- **Publisher**

- URI to an organisational unit in an XML organisational chart (using Xpath)

How I did it ...

- Other Contributor
 - Used Gotel & Finkelstein "Contribution Structures"
- Date
 - ISO 8601 plus ODMA (Open Document Management API) attributes

How I did it ...

- Resource Type
 - DC working proposal plus URI to an XML internal coarse grain model of document types (using Xpath)
- Format
 - Internet Media Types (MIME values)

How I did it ...

- Resource Identifier
 - URNs plus Xpath (where appropriate)
- Source
 - As vanilla DC

How I did it ...

- Language
 - RFC 1766 plus abuse of sub-tag system
- Relation
 - Scrap DC proposal replace by XMI or simplified XMI

How I did it ...

- Coverage
 - Location from ERP service, expiry, size
- Rights Management
 - Use of Sloman domain expressions, use of organisational chart as domain service

Obtaining Metadata

- Extraction
 - Resources
 - Tools
- Potential of XSL

Storage Strategy

- Resource-side
- Database
- Mobile agent (?)
- Use of Rochade

Resource Discovery

- effectiveness of the metadata is dependent not just on its quality but on the granularity of the information resources that are being described
- 80% of the information resource discovery concerns only 20% of the resources
- 80% of the information resource discovery concerns a relatively small set of attributes of the resource

Questions

- RT — *retrieval* is about finding relevant resources
- RS — *responsibility* is about who is responsible for their content
- EX — *expertise* is about finding expertise on a particular topic or technique
- AC — *access* is about the means to access the content of a resource
- TE — *team* is about the structure and nature of the networks of agents associated with a resource
- TR — *traceability* is about the relationship between resources across the system lifecycle.

	Title	Creator	Subject	Description	Publisher	Contributor	Date	Type	Format	Identifier	Source	Language	Relation	Coverage	Rights
Title															
Creator	RT														
Subject	RT	EX													
Description	RT	EX	RT												
Publisher	RS	RS	RS	RS											
Contributor	EX	TE	EX	EX	TE										
Date	RT	RT	RT	RT	RS	RS									
Type	RT	EX	RT	RT	RS	EX	RT								
Format	AC	EX	AC	AC	EX	EX	-	AC							
Identifier	RT	RT	RT	RT	RT	RT	RT	RT	RT						
Source	TR	TR	TR	TR	TR	TR	TR	TR	AC	TR					
Language	AC	EX	AC	AC	EX	EX	-	AC	AC	AC	AC				
Relation	TR	TR	TR	TR	TR	TR	TR	TR	TR	TR	TR	TR			
Coverage	AC	-	AC	AC	-	AC	AC	AC	AC	AC	AC	AC	AC		
Rights	AC	AC	AC	AC	AC	AC	AC	AC	AC	AC	AC	AC	AC	AC	AC

Conclusion

- Cheap
- Stable (very important)
- Simple
- Flexible
- Accommodates lots of other information management and knowledge management techniques
- Leverages Web-tools

Problem

- Not really tested!
- Strong feedback to DC community ...