

Designing a shared representation

Em Tonkin

e.tonkin@ukoln.ac.uk



Who are UKOLN?

- Centre of expertise in digital information management
- Funded by JISC and MLA
- Supporting the Museums, Libraries and Archives contexts

(but also have a hand in e-Science)

- *In this context, metadata often means:*
 - *Dublin Core*
 - *IEEE Learning Object Metadata*
 - *etc...*



Just to be difficult (preface)

I'm not going to talk about ontologies.

Well, maybe sort of...

The DL domain has its own specialised language for this sort of thing...!

Based around a cataloguer/library approach



IEMSR

A machine-readable registry of schemas, profiles and elements

“Metadata schema registries enable the publication, navigation and sharing of information about metadata. The Information Environment Metadata Schema registry (IEMSR) will act as the primary source for authoritative information about metadata schemas recommended by the JISC IE Standards framework.”

IEMSR is developed by UKOLN, with the partnership of the ILRT.

Contributors to the project include CETIS, Becta and the British Library



<http://www.ukoln.ac.uk/projects/iemsr/>

Metadata schema registries

What?

- Centralised collection of:
 - Metadata schemas/application profiles
 - Vocabularies
 - Elements
- Via:
 - Creation,
 - Publication,
 - Aggregation
- For: DC/LOM



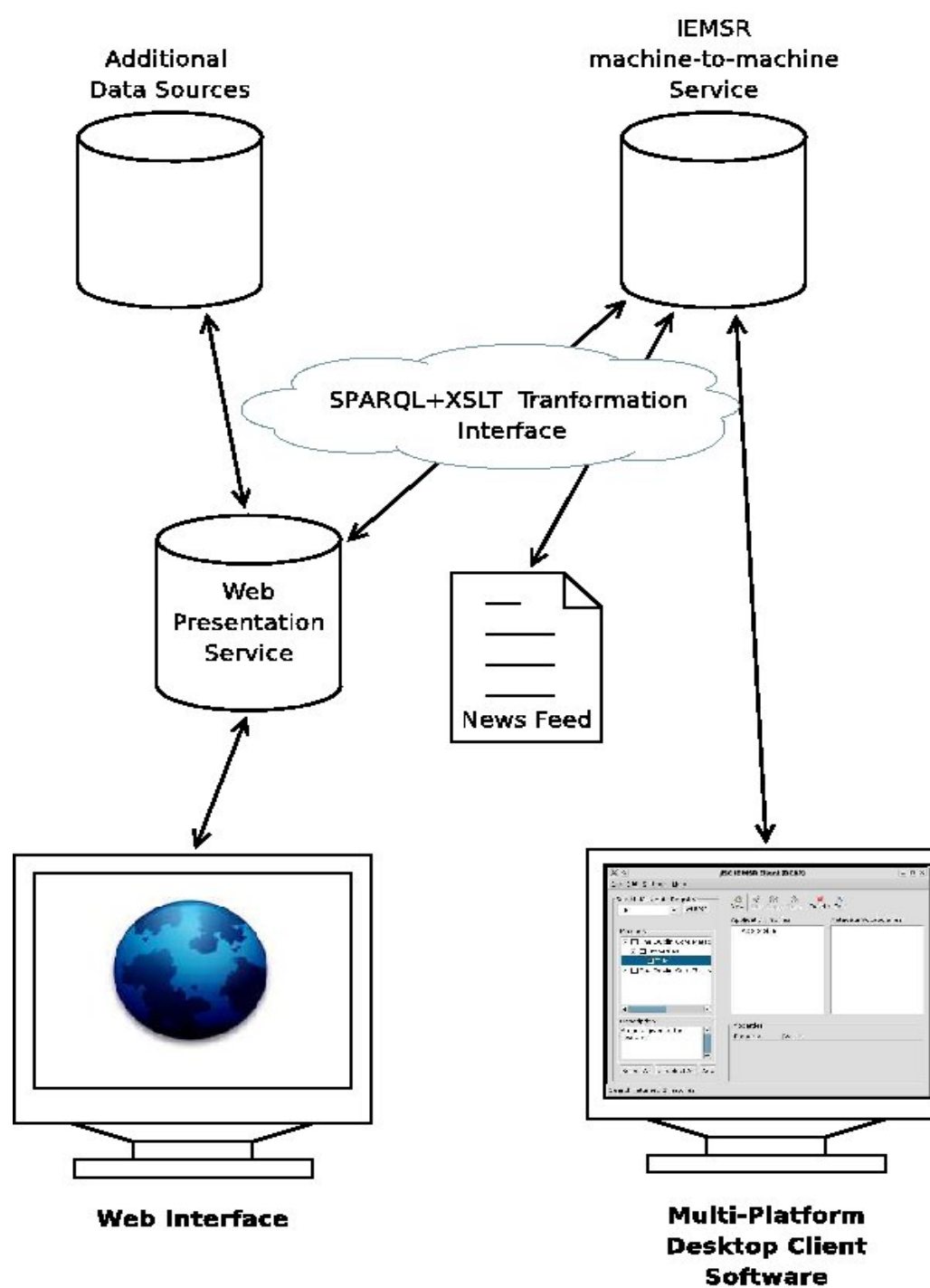
Aside: definitions

- Metadata: generic – data about data
- A metadata vocabulary: a collection of metadata elements
- A schema: (roughly) a set of elements that form a complete metadata record
- An application profile: a set of pragmatic instructions covering metadata application/use and interpretation in a given context
 - An 'accent' or 'dialect'



Comparing DC and RDF

- DC can be encoded in RDF
- Designed for resource description – and for
 - Simplicity
 - Widely-applicable, context-independent semantics
 - Corresponds to DC Abstract Model
- A small heresy: APs are
 - Contextualised
 - Often complex – in appearance at least
 - More implicit than explicit



Metadata schema registries

Why?

- Reduces duplication of effort
- Promotes reuse
- Demonstrates existing practice
- Machine-readable representation

- Collection is prerequisite for:
 - Mapping functionality
 - Shared 'concept spine'
 - Reuse



How?

- Schema descriptions in RDF
- Created via Java client or independently
- Stored in Redland-DB
- Queried via SPARQL
- XML-encoded response
 - transformed via XSL(T)
 - Easy to add output formats

DC as 'concept spine'

A vague 'feel' for the content of a term

Element: someSubHeading

Is an instance of: dc:title

Good enough?

Depends!



Schema development as CSCW problem

- Schema/AP development characteristics:
 - Partial encoding of domain-knowledge
 - Often distributed
 - Often have heterogeneous user population
 - Schemas generally evolve with time -
 - *Ongoing process...*

The importance of user and domain: why some people have it easy(?)



Open Questions

- How much difference does the choice of encoding make to schema development?
- Should schemas be designed with the end-user in mind?

If yes, how can this be done?

- Process
- Evaluation
- Ongoing maintenance

Usability & the schema

- The pathetic fallacy:
 - Karger and schraefel: `pathetic fallacy':
“the notion that data should be presented to the user as it is represented in the computer”
- APs are commonly used in this manner
- Assertion: *“As an integral part of the system, the data/metadata model is a core element. Hence, a usability evaluation of the system includes these elements.”*

Knowledge representation

Metadata model as shared conceptualisation

Stakeholders:

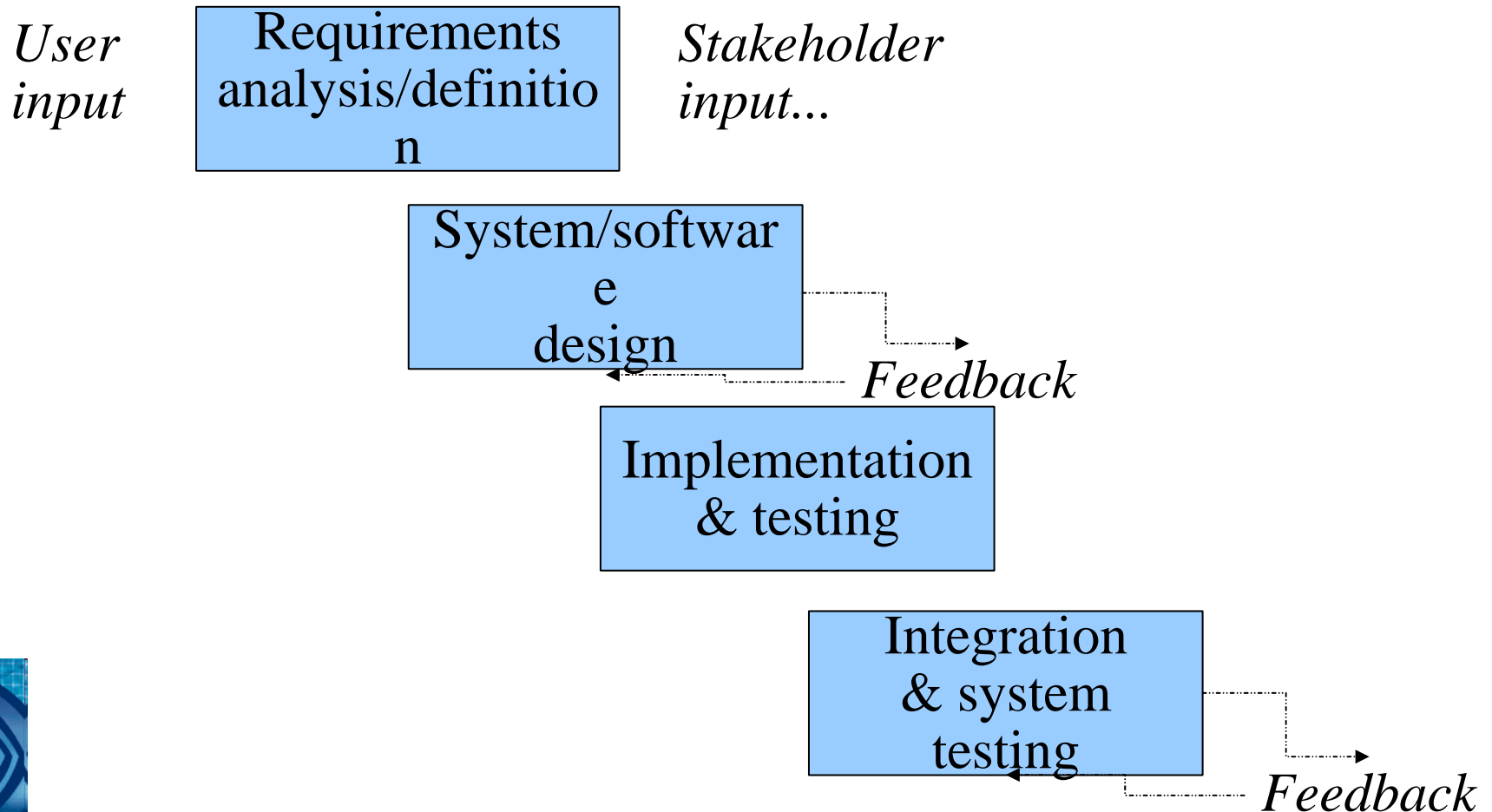
- Record creators
- System end-users
- App developers
- 'Intermediate developers/consumers'

Principle: Minimalism (or: *Don't fall in love with your model...*)



Observed design process

Observed design process typically follows a variation on the waterfall model:



Iterative design process

Iterative process, ie Hix & Hartson (1993)

Input from stakeholders & existing work

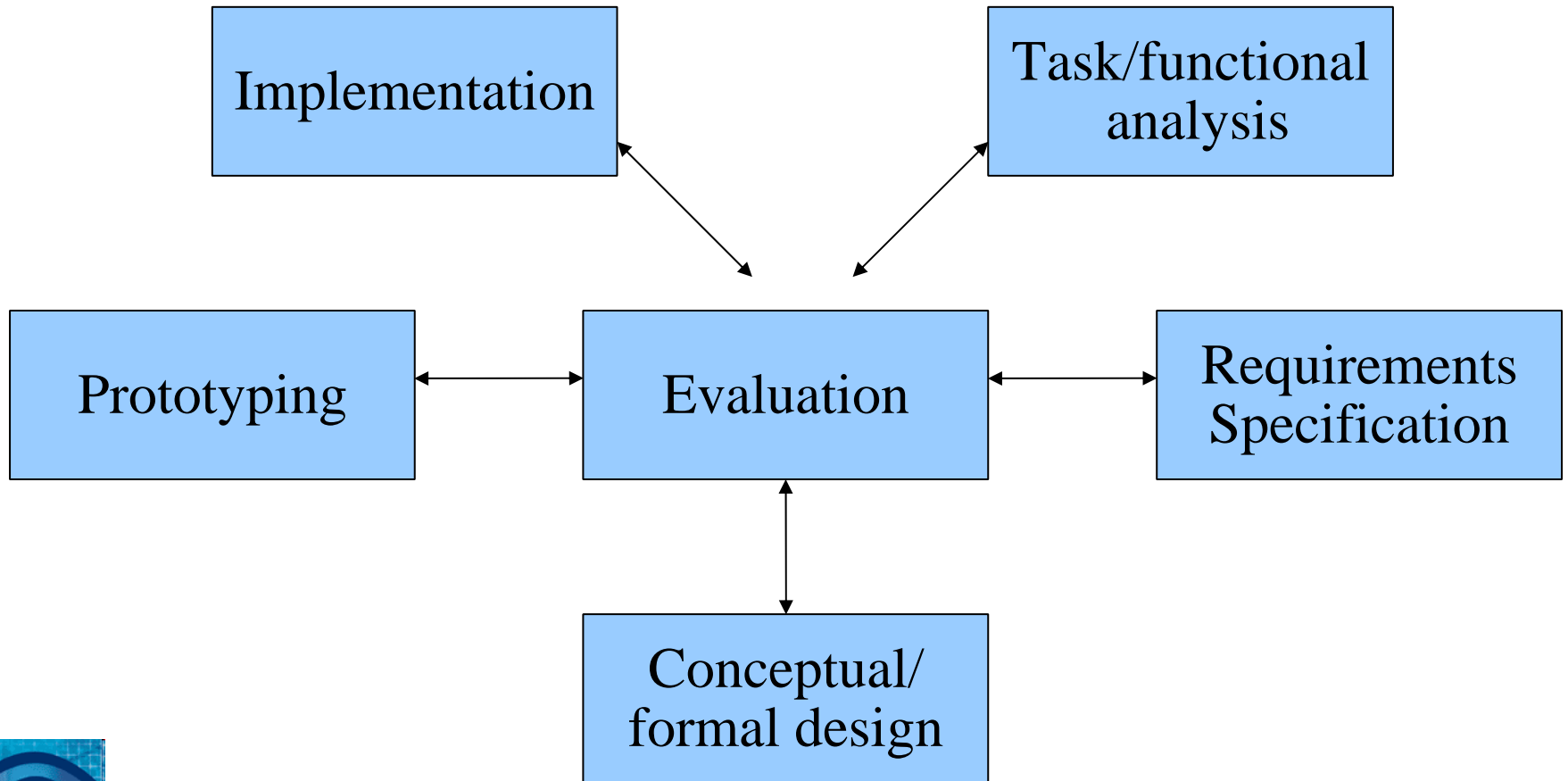
User-centred – *not user-controlled!*

Concentration on

- *solid user/task analysis*
- *Technical analysis*

Building clear requirements spec

Star life cycle



Evaluation

Develop realistic user scenarios

Employ 'paper prototyping'

Key issues:

- Is this conceptualisation expected to be shared, with whom and to what extent?
- In context, how much do the details matter?

Expectation management

Expectation management: Schemas can't brew coffee

...expectation management is vital.

- Is demand real or perceived?
- How complete/accurate is the result?
 - Will it still be accurate next month?
- How do schemas evolve? Can we support the process?
- Should it be allowed?
 - Prescriptivist approaches to metadata!
 - “We'll teach the users to see it our way”



Evaluation

- Sufficient unto the day use case
 - Can achieve all identified aims
- Alternative metrics:
 - Provides reproducible results
 - Is applied in a consistent manner
 - How can we tell? Through analysis of examples.

Is evaluating once sufficient?





*Information Environment
Service Registry*

A machine-readable registry of electronic resources

“The aim of the Information Environment Service Registry (IESR) is to make it easier for other applications to discover and use materials which will help their users' learning, teaching and research.”

*IESR is developed by
MIMAS, in partnership
with UKOLN*

*Read more at the Service
Registries Blog:*

*[http://iesr.ac.uk/service-
registries-blog/](http://iesr.ac.uk/service-registries-blog/)*



<http://iesr.ac.uk/>

MIMAS

JISC

Authority

- APs often contain references to:
 - Other metadata vocabularies (same standard) – RSLP, eGMS, etc
 - 'Informally imported' elements from different but familiar standards, such as MODS (an XML standard)
- A chicken-and-egg problem: to import this structure, every authority must add their terms before system becomes useful!
- But authorities may not be aware of 'informal imports'!



'Assertion-Without-Proof'

- An exercise in semiotic dynamics
- 'We use a term that means something like this – and I think Bob invented it'
- 'I disagree – it was Mary's'
- 'I think it means more like ...'
- '...and Laura agrees with me!'
- The wiki approach to achieving consensus?



The semantic mismatch

I want to say...

but I only have the option of saying ...

- Symptoms
- Effects
- Solutions



User annotation and tagging

Shortcut to 'speaking the user's language'?

- Various competing terminologies
- 'Lead into the system'
- Lightweight means to increase accessibility?

The domain expert vs the domain user



Dealing with evolution

- Language use evolves quickly
 - especially in specialised discourse!
- Language can act as 'branding'
- The Kuhnian paradigm
- To paraphrase UoC's John Kunze:
Opaque identifiers avoid arguments

Lessons learnt

- Language use very domain-specific
- More 'pragmatic' than 'semantic' web
- Sometimes it's good to be informal:
 - *You know you'll need to interpret the result*
- Works well in specialised domains
- Less appropriate in general use
- Realistic aims and expectations recommended
- Manage your committees with care



Conclusion

K.I.S.S. - the DC-Kernel project

Community consensus (1-2 people in every ten give an opinion) vs. community accessibility (hopefully, most people use the system).

The cappuccino heuristic: if you need to meet over coffee to explain it, you may need to do the same for each of your end-users...

Vision: a system with good metadata and relatively low cognitive load...

