



# **Terminology and Vocabulary Management**

## **Vocman's Background and Approach**

Mike Collett

Vocabulary Management Group





# Aims

- Background
- An overview of requirements
- Standards and modelling
- Summary





## Our Background

- Started with metadata
  - Curriculum Online then Adult Community Learning
- Developed Becta's UK vocabulary management services, now with DCSF
- European projects and standards
- Lexaurus Editor and Bank evolved

LEXAURUS  
EDITOR ■■■■

LEXAURUS  
BANK ■■■■





# Vocabularies

- A vocabulary is a set of terms
  - Each term is associated with information (elements in a term record)
  - A term can be related to other terms
- A vocabulary can have a structure
  - List
  - Thesaurus
  - Topic Map
  - Taxonomy
  - ...





# Purposes

- Vocabularies can be used to:
  - provide controlled lists for use in metadata
  - share structures for navigation - in tools or portals
  - carry additional information
  - provide mappings between terms





## Places

- Leeds
- Richmond
- ...

- Try them in [streetmap.co.uk](http://streetmap.co.uk)





You are trying to find "Leeds"

There are several possible locations, please select one...

- Leeds, Kent
- Leeds, West Yorkshire
- Leedstown, Cornwall
- etc





You are trying to find "Richmond"  
There are several possible locations, please select  
one...

- Richmond, Sheffield
- Richmond, Richmond upon Thames
- Richmond, North Yorkshire
- Richmond, Aberdeenshire
- etc





## More Precise Search

- Search IPSV for 'business' and there are 31 preferred terms
  - 96 with a wildcard and non-preferred terms

- The user can choose the best fit

E.g. *Businesses (companies)* has the preferred term of *Companies* that appears under both *Business and Industry* and *Organisations*

**BT** Business and industry

**BT** Organisations

**NT** Cooperatives

**NT** Small businesses

**RT** Employers

**UF** Businesses (companies)

**UF** Corporations

**UF** Enterprises (companies)

**UF** Firms (companies)





# Drivers

- Vocabularies change
  - Version control and history are important
  - Publishing is becoming a web service
- It is useful to publish vocabularies
  - Separate editing and publishing functions
- The fundamental model has to meet several requirements
  - Now and predicted





## Some requirements

- Polyhierarchy
  - terms can appear in more than one place
- Term reusability
  - the source vocabulary of a term is needed
- Multilingualism
  - for terms, descriptions or *any* text
- Data transfer
  - optimisation and partial transfer possible
- Controlled references
  - term facets
- Extensibility





# Standards

- Open formats preferred
- There are lots of different kinds of standards
  - Exchange - import or export
  - Search, query and retrieve
  - Presentation and style
  - Others?
    - Upload or ingest
    - Alerting





# Exchange

- Import formats should carry all the information the publisher intends
  - Can be very simple or more complex
- Export formats should be capable of carrying the information a consumer needs
  - Different purposes
- Internal formats should have everything!
  - Not so easy to get one size fits all
- Cannot assume that the information in one format is the same as in another format





## Exchange

- Started with Zthes 1.0 and extended and profiled the model
- Currently supported
  - Zthes, Extended Zthes, SKOS, ESD (a proprietary UK format), VDEX, JSON
- Others driven by demand
- And less specific formats such as spreadsheets (csv), documents, diagrams (mindmaps)





## Search and query

- A vocabulary service needs to be searched and browsed by machines as well as people
- Use an open standard with widespread support and rich search semantics (SRU does the job)
- Customisable searching e.g.
  - specific vocabulary
  - just preferred terms
  - description as well as term name





## Style

- Several standards provide guidance on how vocabularies:
  - can be displayed
  - are constructed e.g. plural-singular, caps or not
- Mostly guidelines not conformance
- Use stylesheets to give localised format





# Model

- We use a flexible model that will allow for anything to be represented
- There are only 3 basic things
  - Stuff
  - Information about stuff
  - Connections between stuffs
- Allow multiple schemas to be used
  - to define labels as well as elements (or properties)





## Summary

- Vocabulary Management Group have experience of vocabulary management
- We have identified a set of requirements - driven by customers
- Use open standards, or profiles of them, where appropriate
- Lexaurus uses a simple, extensible underlying model that is not format specific

