



Challenges and opportunities for KOS standards

Stella G Dextre Clarke

Convenor, IDT/2/2 Working Group of BSI
and Project Leader for ISO NP 25964



Summary

- BS8723 progress report
 - Background and overview of BS 8723
 - Highlight what's new in the five parts
 - Outline of ISO NP 25964 (adoption of BS 8723 as an ISO standard)
- Challenges and opportunities



Project background: Thesaurus standards back in 2001

- ISO 2788-1986 Guidelines for the establishment and development of monolingual thesauri
= BS 5723:1987
- ISO 5964-1985 Guidelines for the establishment and development of multilingual thesauri
= BS 6723:1985
- Biggest motivation for updating them is “Interoperability”



Procedure

- Small UK Working Group of unpaid volunteers forms the BSI drafting committee IDT/2/2/1 (Leonard Will, Alan Gilchrist, Ron Davies + convenor Stella Dextre Clarke)
- Comments have been received internationally wherever possible
- Once complete, BS 8723 gets offered for adoption as an international standard



Overview of BS 8723

BS 8723: Structured vocabularies for information retrieval – Guide

- Part 1: Definitions, symbols and abbreviations
- Part 2: Thesauri
- Part 3: Vocabularies other than thesauri
- Part 4: Interoperability between vocabularies
- Part 5: Exchange formats and protocols for interoperability



Progress to date

- Parts 1 and 2 published November 2005
- Drafts of Parts 3 and 4 have gone out for public comment; comments now received and incorporated; publication expected Oct-Nov 2007.
- Draft of Part 5 currently in development



New features in BS 8723-2: [monolingual] Thesauri

- Clearer guidance on applying facet analysis to thesauri
- Some changes to the 'rules' for compound terms
- More guidance on managing thesaurus development and maintenance
- Functional specification for software to manage thesauri
- General overhaul in all areas



Chapters of BS 8723-3:

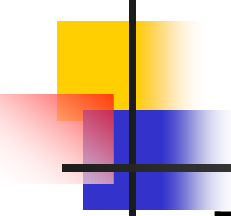
Vocabularies other than thesauri

- Classification schemes
- Business classification schemes for records management
- Taxonomies
- Subject heading schemes
- Ontologies
- Authority lists



BS 8723-4: Interoperability between vocabularies

- Huge demand for accessing information that has been indexed with another language and/or vocabulary. The buzzword is 'Mapping'. The Semantic Web is just one application.
- Part 4 includes multilingual thesauri as a special case of mapping between vocabularies



Simple stuff, e.g. What to map in different vocabulary types

Thesaurus	Preferred terms
Classification scheme	Notations
Taxonomy	Category labels, notations or identifiers
Subject heading scheme	Terms or pre-coordinated strings
Authority list	Terms or identifiers
Ontology	Terms or identifiers



Symbols and abbreviations recommended for mappings

EQ Indicates equivalence

Example:

lifts (UK) EQ elevators (USA)

+ When combining *aspects* of the concept

Example:

women executives EQ women + executives

| When combining *parts* of the concept

Example:

livestock EQ sheep | cattle | pigs | poultry

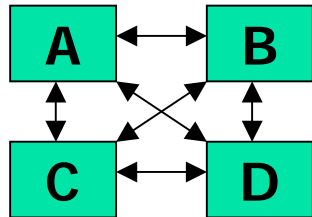


Why the fancy symbols?

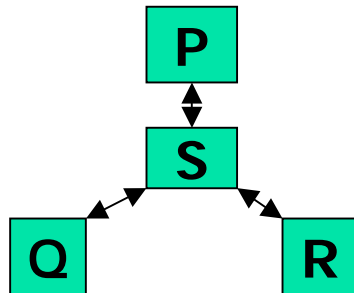
- It is not often appreciated that mappings can be used either for converting search statements or for converting index terms, and the way you implement combinations in these different situations depends on whether the elements represent *parts* or *aspects* of the original concept

Structural models

- Model 1: Structural unity
- Model 2: Non-equivalent pairs



- Model 3: Backbone structure





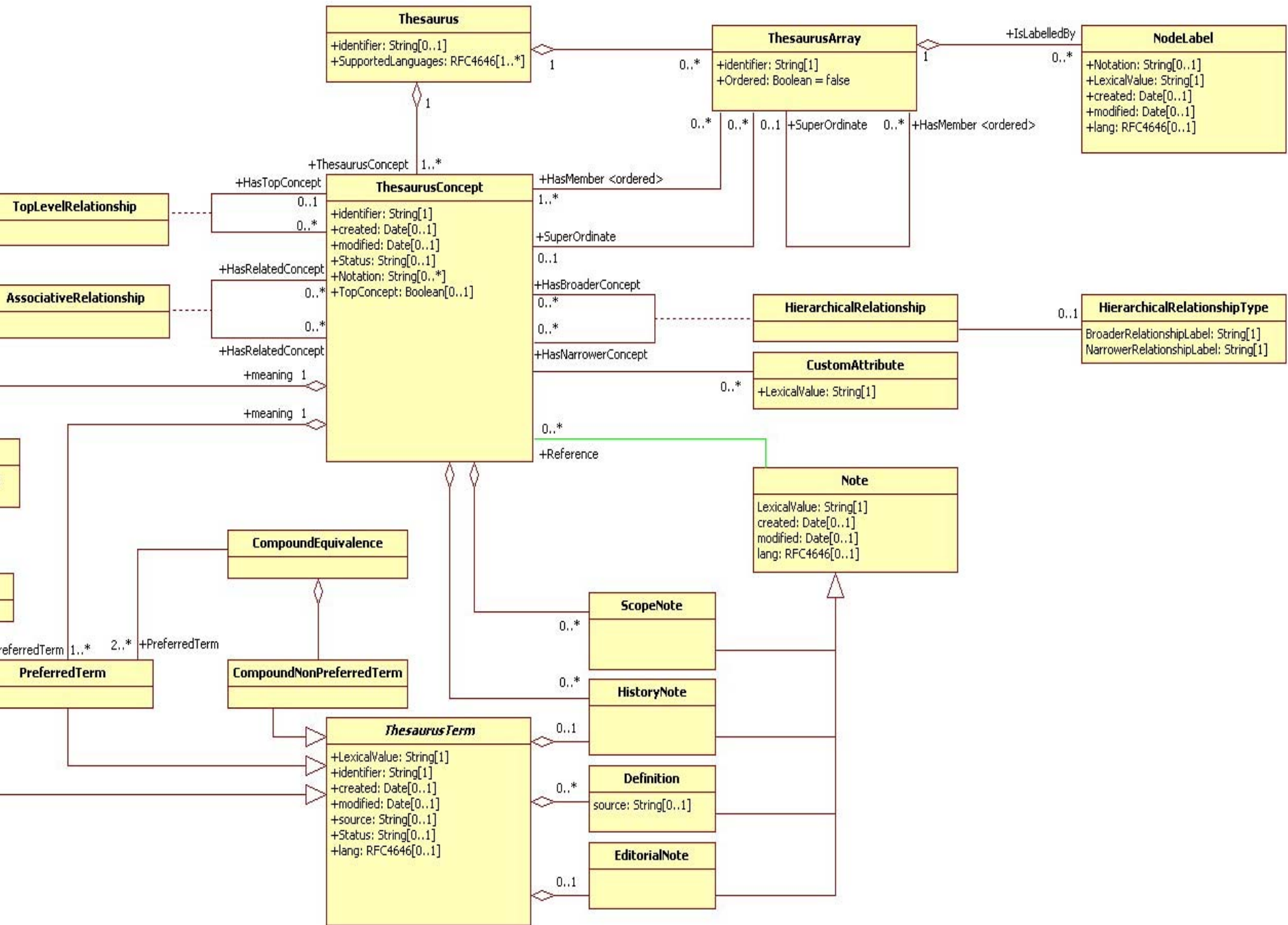
BS 8723-5: Exchange formats and protocols for interoperability

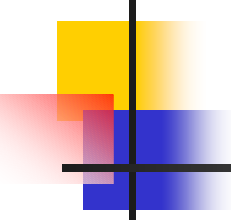
- In Oct 2006 a workshop was held, to seek advice from experts
- Key conclusions:
 - We need a model
 - We should recommend one standard format, not several, and it should be based on XML
 - The format should provide only for thesauri
 - We should not develop a new protocol



BS 8723-5 progress to date

- After much heated debate... and a **huge** amount of work...
- Model and schema available on development website
- But not everyone is confident this is the best, or the only, or the most practical solution
- It is hoped to issue a DD (Draft for Development) rather than a full BS standard, this November.





ISO NP 25964 (adoption of BS 8723 as an ISO standard)

- The proposal to revise ISO 2788 and ISO 5964, basing the work on BS 8723, was submitted to ISO TC 46/SC 9 members in April 2007
- Project now approved
- At least 9 countries participating:
France, Germany, Canada, Finland, New Zealand, Sweden, UK, Ukraine, USA



Next steps

- SC9 Secretariat will convert Parts 1-4 to the ISO template, and these will constitute first working drafts. Hopefully with a draft of Part 5 too.
- Funding for meetings is very limited.
- Most communication will be by email. Any meetings are likely to piggy-back other conferences.
- We have still to assess how much variation on BS 8723 will be requested.
- One day we shall have a new ISO standard covering the whole field!



Challenges and opportunities in a nutshell

- The theory: if only everyone would do it the same way, there would be huge gains for clarity, efficiency, economies of scale, interoperability, etc.
- In practice: differentiation is the lifeblood of innovation and competitive advantage
- How to find the balance?



A few specific challenges, or are they opportunities?

- Should we think of BS 8723-2 as needing rigorous implementation, or is it more of a springboard from which to launch new ideas?
- Facet analysis practice as developed by the CRG – will it come of age, or has the genie already left the bottle?
- Is it useful/practical to apply a tight definition to “taxonomy”?
- Which offers greater benefit: one exchange format for all applications, or a choice of formats, supported by transforms between them?
- Format extension: should it reach towards ontologies, or towards classification schemes, or what?
- And what about folksonomies – is there any benefit from considering them at all in BS 8723 or ISO 25964?