



Providing for interoperability between thesauri and classification schemes in ISO 25964

Stella G Dextre Clarke

Project Leader, ISO NP 25964



What is ISO 25964?

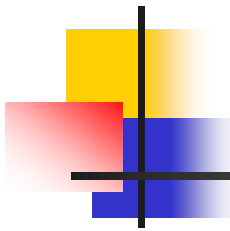
ISO 25964: Thesauri and interoperability with other vocabularies

- Part 1: Thesauri for information retrieval
- Part 2: Interoperability with other vocabularies



Where did it come from?

- ISO 2788-1986 Guidelines for the establishment and development of monolingual thesauri
- ISO 5964-1985 Guidelines for the establishment and development of multilingual thesauri



The relationship with the past is not quite what you'd expect...

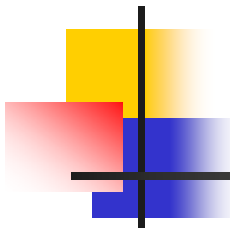
ISO 25964: Thesauri and interoperability with other vocabularies

- Part 1: Thesauri for information retrieval
- Part 2: Interoperability with other vocabularies

Part 1 covers monolingual **and** multilingual thesauri

Part 2 covers mapping between thesauri and other types of vocabulary

- *Both parts are still in draft. DIS/ISO 25964-1 will be issued in 2009; DIS/ISO 25964-2 is still in the early stages*



Where did it come from? - Alternative answer

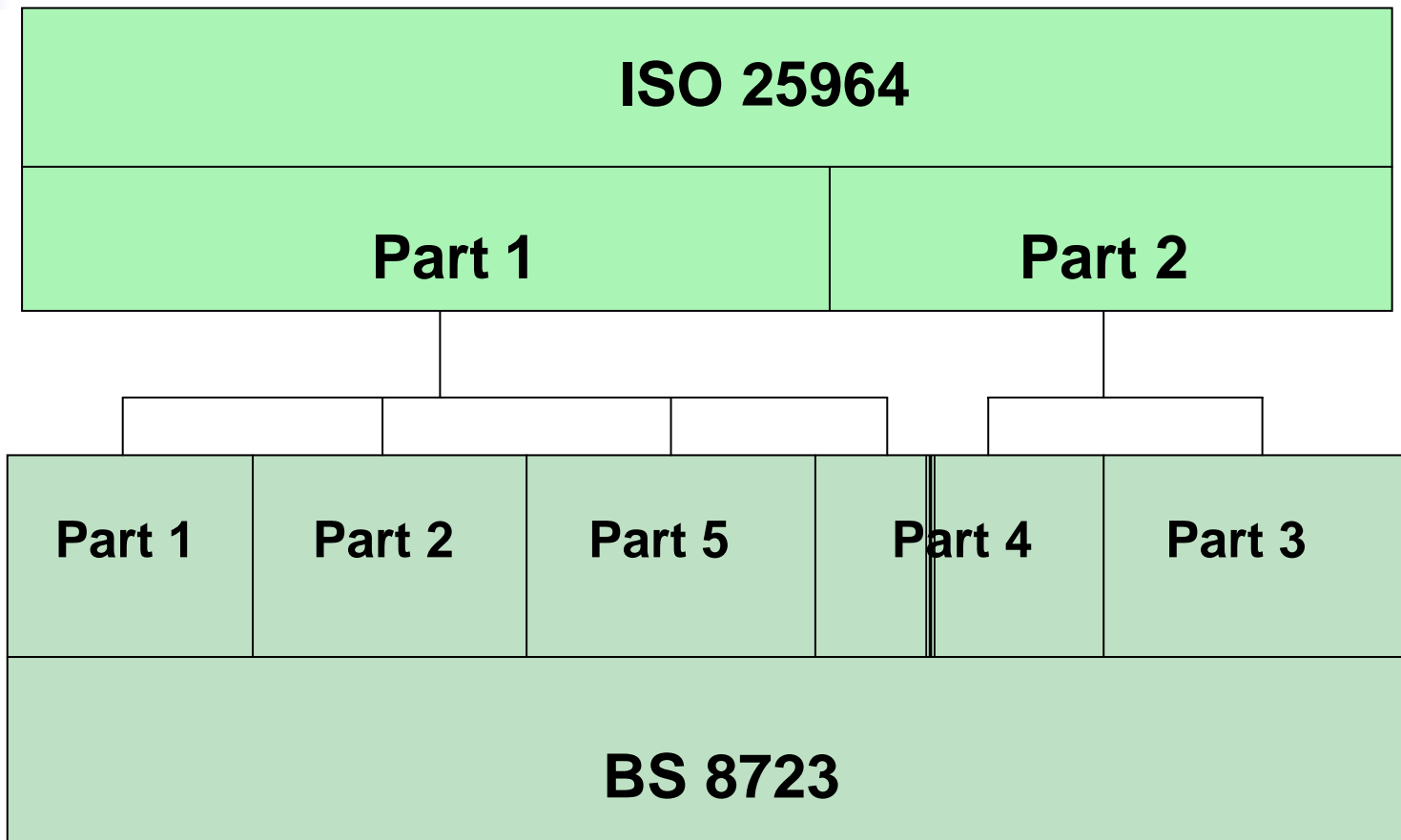
BS 8723: Structured vocabularies for
information retrieval – Guide

- 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

Published 2005-2008



How ISO 25964 rests on BS 8723

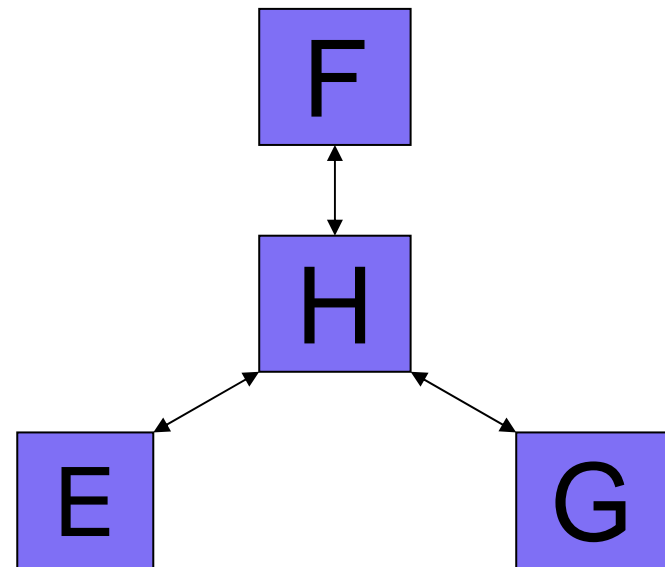
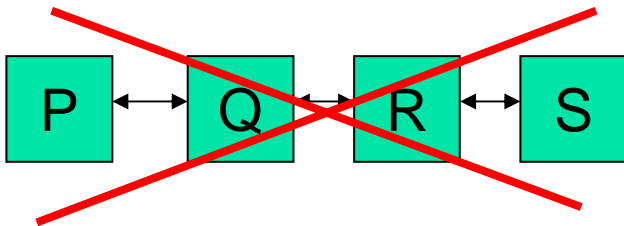
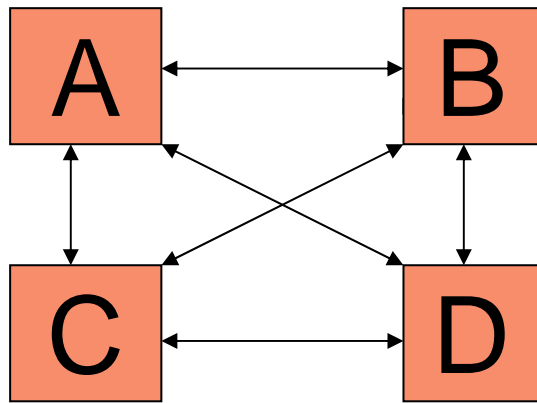




Planned content of Part 2

- Structural models for interoperability

Structural models for interoperability





Planned content of Part 2

- Structural models for interoperability
- Discussion of equivalence
- Principles of mapping
- One-to-many and many-to-one mappings
- Managing mappings data
- Display of mapped vocabularies
- Mapping system functionality
- Exchange formats for mappings
- Description of other vocabulary types



Issues for Part 2 (just some of them!)

- How much description of classification schemes (and other vocabularies)?
- Whether (and if so how) to include a data model for each type of KOS
- How to handle pre-coordinated classes
- How to provide for classes that are not enumerated in the scheme but synthesised on demand

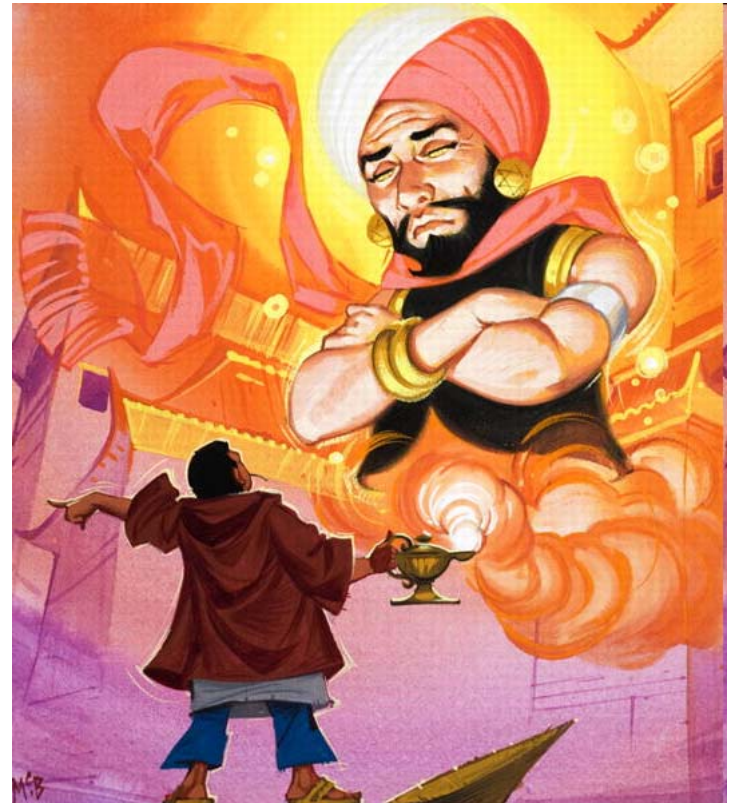
Vocabularies other than thesauri



- Classification schemes
- Business classification schemes for records management (aka file plans)
- Taxonomies
- Subject heading schemes
- Ontologies/Topic maps/Semantic networks
- Terminologies/Term banks
- Name authority lists
- Synonym rings

How much description of classification schemes?

- It will not be our aim to make a new standard for classification schemes
- (because the genie is already out of the bottle)





Some mapping/modelling challenges for classification schemes

- Notations:
 - Some are expressive; some not
 - Rules for sequence vary, especially with a mix of alpha, numeric and punctuation characters
 - Some (e.g. in BC2) are almost impossible to parse automatically, because of special rules to keep them short and memorable
- Synthesised classes pose extra problems
- Captions should not be equated with preferred terms in a thesaurus
- Notes beside the captions have different functions in different schemes
- How to encode systematic sequence of sibling classes
- How to provide for terms in the index, **if necessary**

Example: "academic library labor unions in Germany"

(- from Marcia Lei Zeng/FRSAD report)

DDC: "331.881102770943"

331.8811 – labor unions in industries and occupations other than extractive, manufacturing, construction

-027.7 – academic libraries

-0943 – Germany

LCSH:

"Library employees--Labor unions--Germany"

"Universities and colleges--Employees--Labor unions--Germany"

"Collective bargaining--Academic librarians--Germany"

"Libraries and labor unions--Germany"

UNESCO Thesaurus:

"Trade unions" "Academic libraries" "Germany"

ILO Thesaurus:

"Trade union" "library" "educational institution" "Germany"



How to map to and from pre-coordinated classes and synthesised notations?

- For vocabularies using postcoordination (esp thesauri) mappings between them look feasible
- Mapping *from* a precoordinated or synthesised class to a thesaurus looks feasible.
- Mapping *to* a precoordinated class looks more problematic!
- The same applies to mapping from a synthesised class in one scheme to a differently synthesised class in another scheme
- Comparing Subject headings with classification schemes, precoordination works in slightly different ways. Can we find common solutions?
- In any case, should the aim to be to map between schemes, or between the indexes of collections indexed/catalogued with the schemes?

Whether to include a data model?



- A unified model or metamodel to cover all vocabulary types?
- A model for each type of vocabulary?
- Do all classification schemes conform to the same model?
- To support mappings, maybe a model is unnecessary and a syntax would be enough? You may need a full model to support live retrieval, but not if you are just establishing and storing mappings?
- What are the use cases?



Want to get involved?

- Contact your national standards body, specifically the committee corresponding to ISO TC 46/SC 9/WG8
- 15 countries already participate:
Bulgaria, Canada, China, Denmark, France, Germany, Finland, Korea, New Zealand, South Africa, Spain, Sweden, UK, Ukraine, USA