



ISO 25964 – plans and progress towards the revised international standard for thesauri

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

NB Both parts are still in draft



Where did it come from?

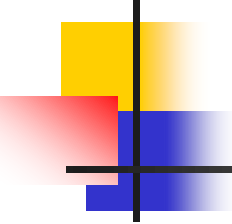
- ISO 2788-1986 Guidelines for the establishment and development of monolingual thesauri
= BS 5723:1987
- ISO 5964-1985 Guidelines for the establishment and development of multilingual thesauri
= BS 6723:1985



Moving on from ISO 2788 and ISO 5964...

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



Where did it come from? - Alternative answer

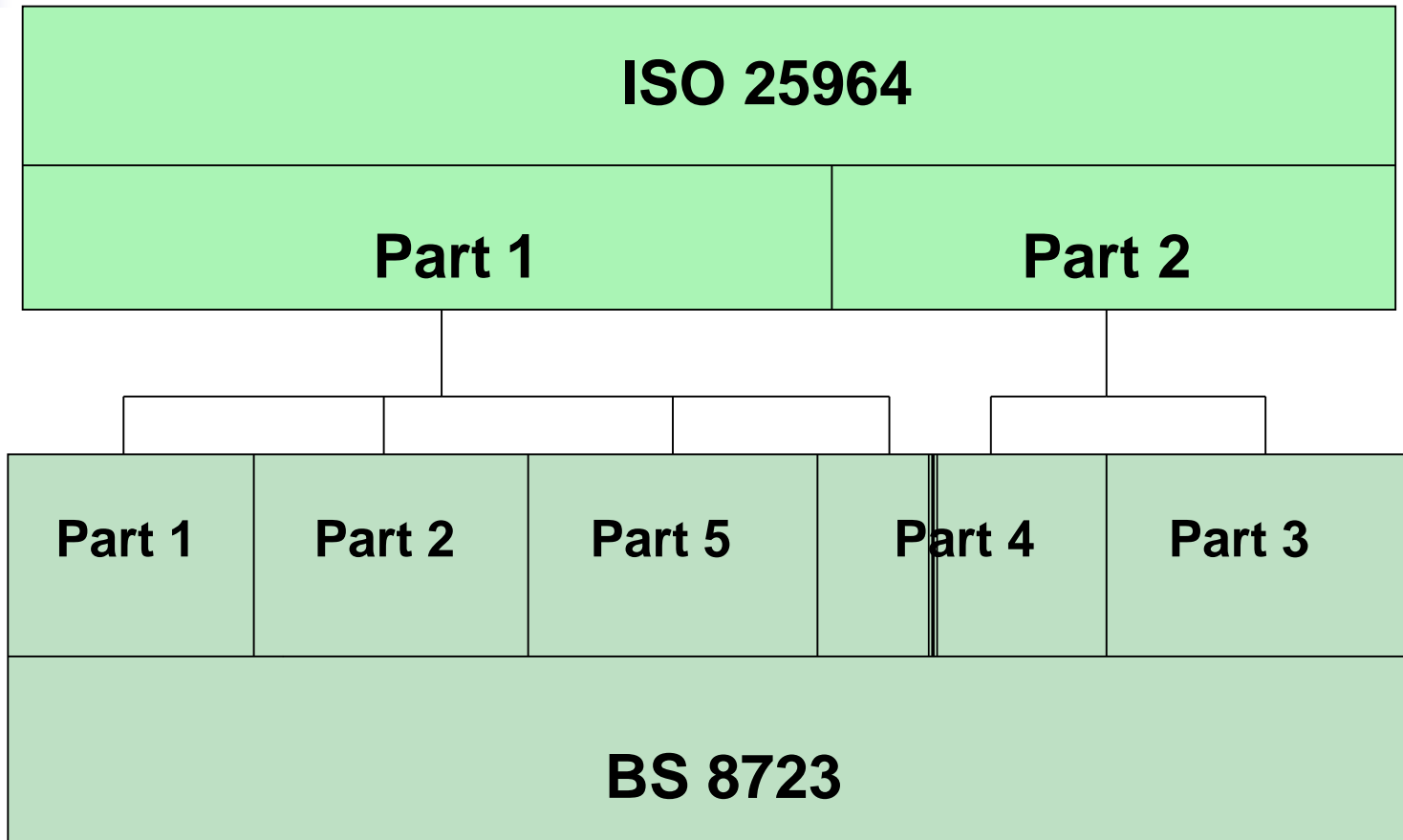
BS 8723: Structured vocabularies for
information retrieval – Guide

- 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



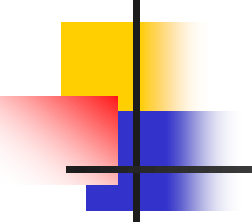
How ISO 25964 rests on BS 8723





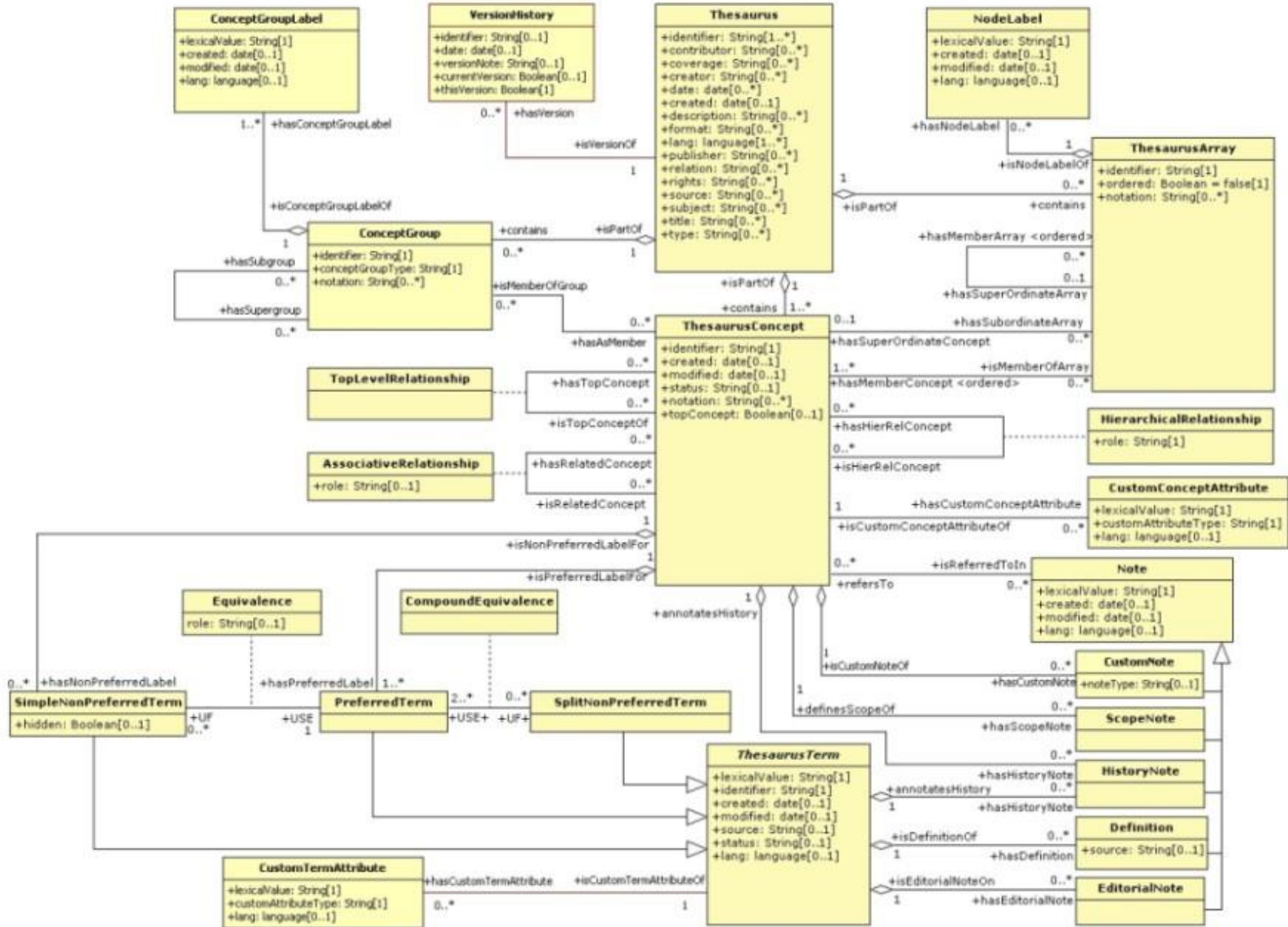
Progress with Part 1

- Some loose ends in BS8723 have been tidied up
- Data model has been extended
- Exchange format XML schema will be designated “informative” so that it can be accessed 24/7 without password control or copyright restrictions
- DIS 25964-1 to be issued in 2009



What distinguishes ISO 25964-1 from ISO 2788/5964?

- Clearer differentiation between terms and concepts
- 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
- Data model and XML schema for data exchange
- General overhaul in all areas, e.g. sweeping update of multilingual examples





What distinguishes ISO 25964-1 from SKOS?

- SKOS does not tell you how to build a vocabulary; only how to publish it.
- The data models for both standards are concept-based. At a simple level it is easy to convert between them.
- Whereas the ISO 25964 model provides for all features of a thesaurus, SKOS aims to serve several different vocabulary types and so does not provide for some thesaurus features.



Issues for Part 2 (ISO 25964-2)

- How much description of vocabularies other than thesauri?
- Whether and how to include “non-symmetrical multilingual thesauri”
- To provide for interoperability, what do we need in the way of data modelling, exchange formats and protocols?

Vocabularies other than thesauri



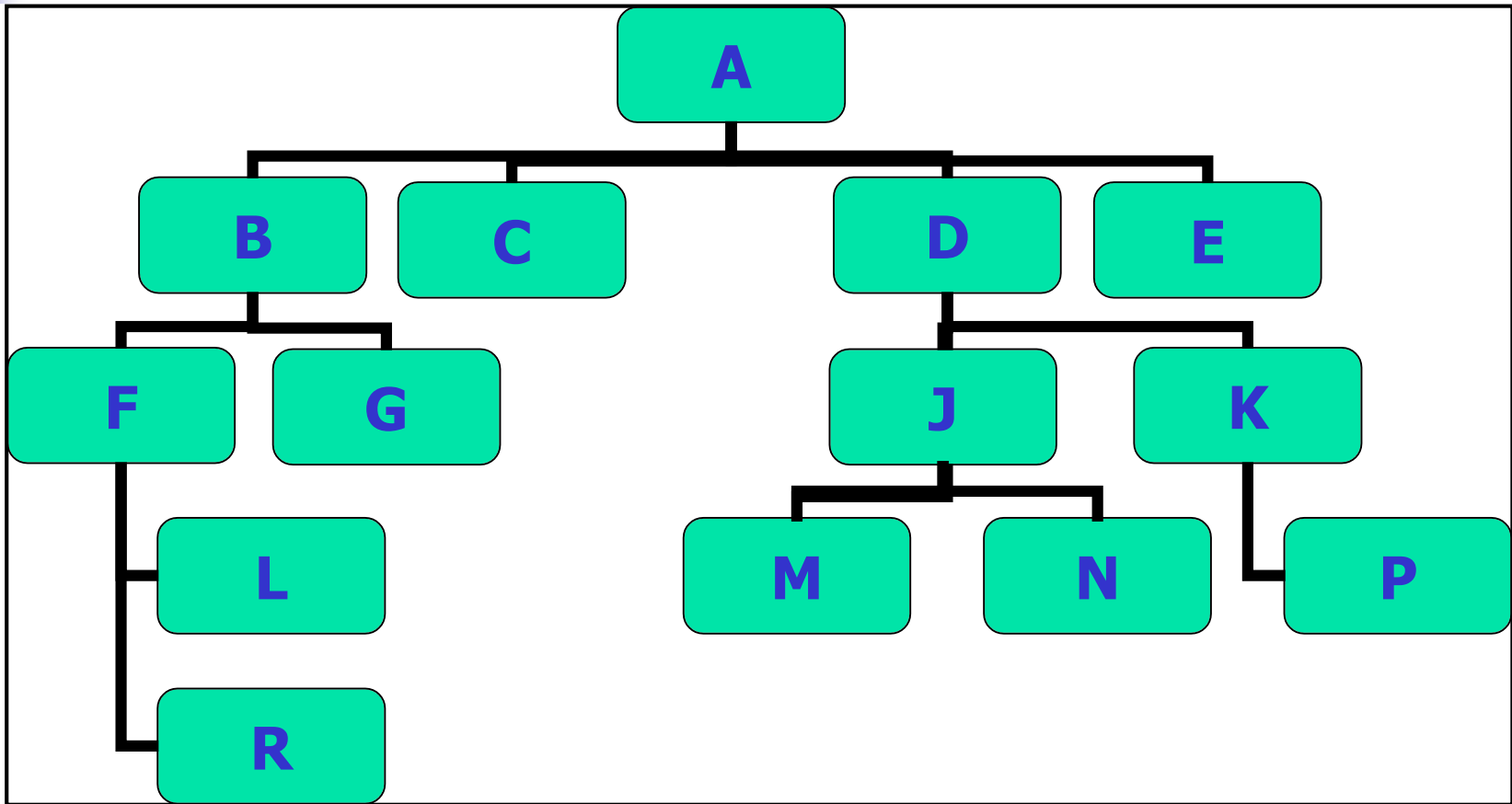
- Classification schemes
- Business classification schemes for records management (aka file plans)
- Taxonomies
- Subject heading schemes
- Ontologies/Topic maps/Semantic networks
- Terminologies/Term banks
- Name authority lists
- Synonym rings



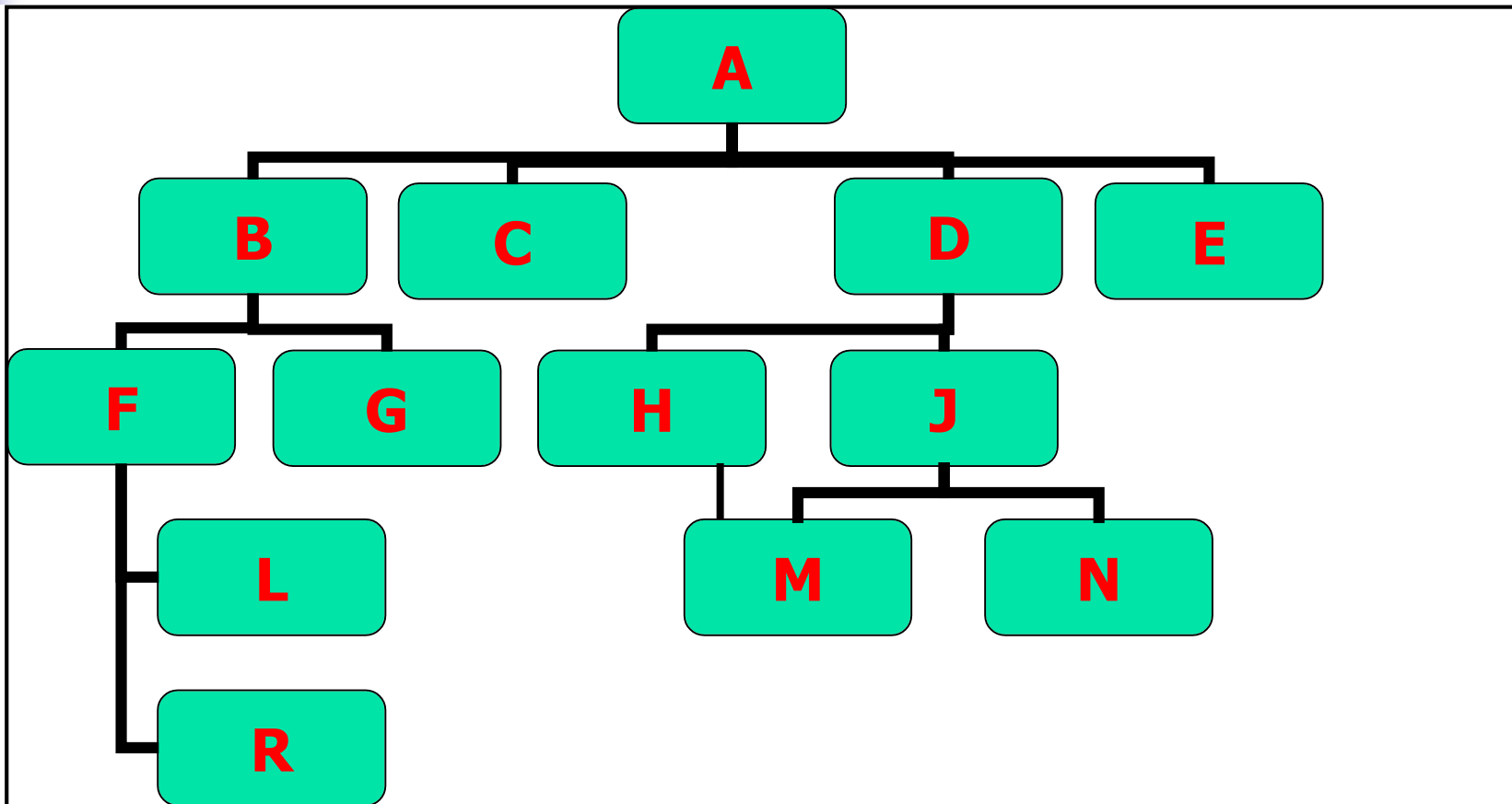
non-symmetrical multilingual thesaurus – what is it?

- A multilingual thesaurus in which the languages do not always share the same structure of hierarchical and associative relationships
- And what is the difference between a non-symmetrical thesaurus and two or more monolingual thesauri with mappings between them?

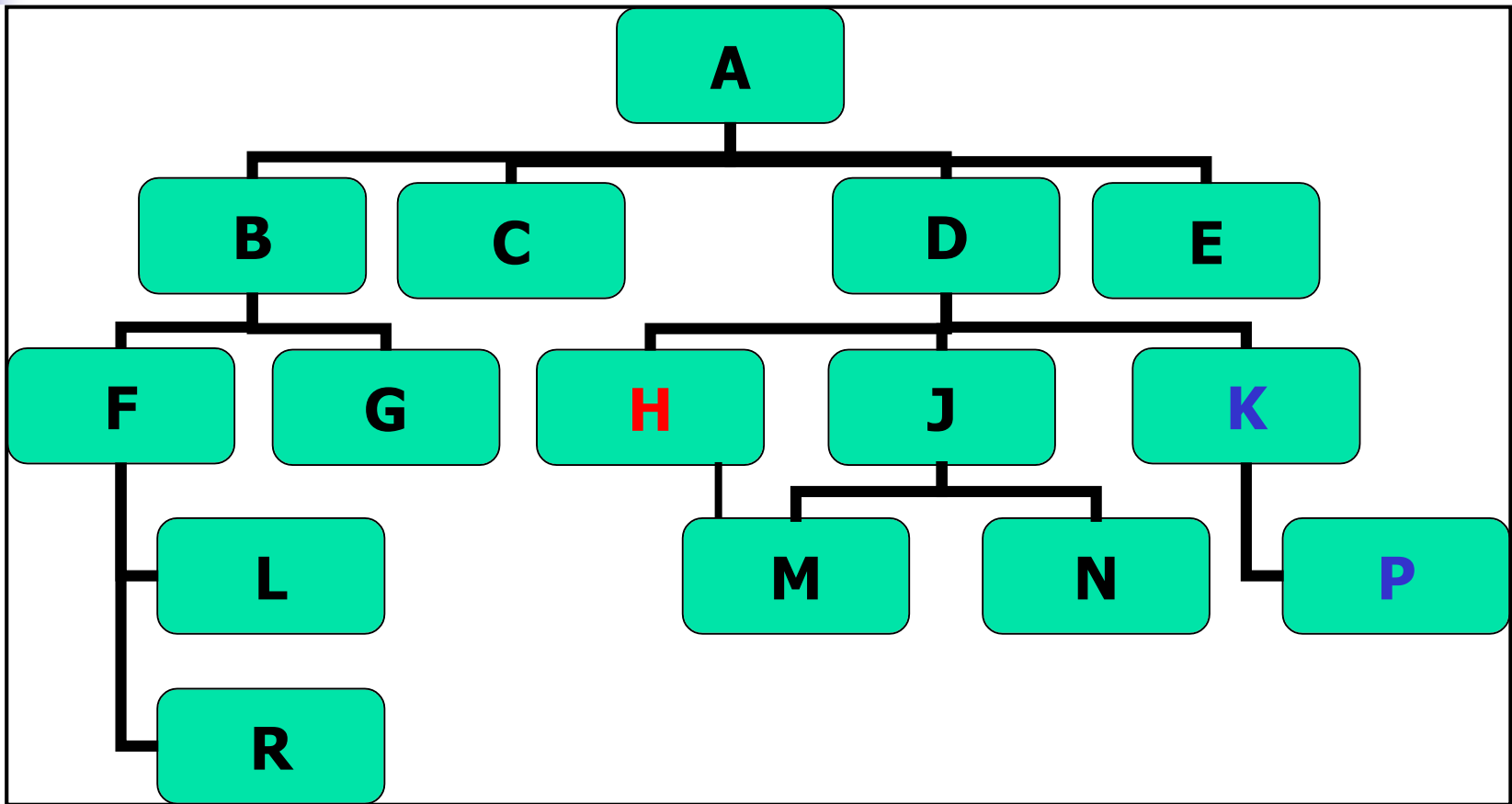
non-symmetrical thesaurus example – Greek in blue



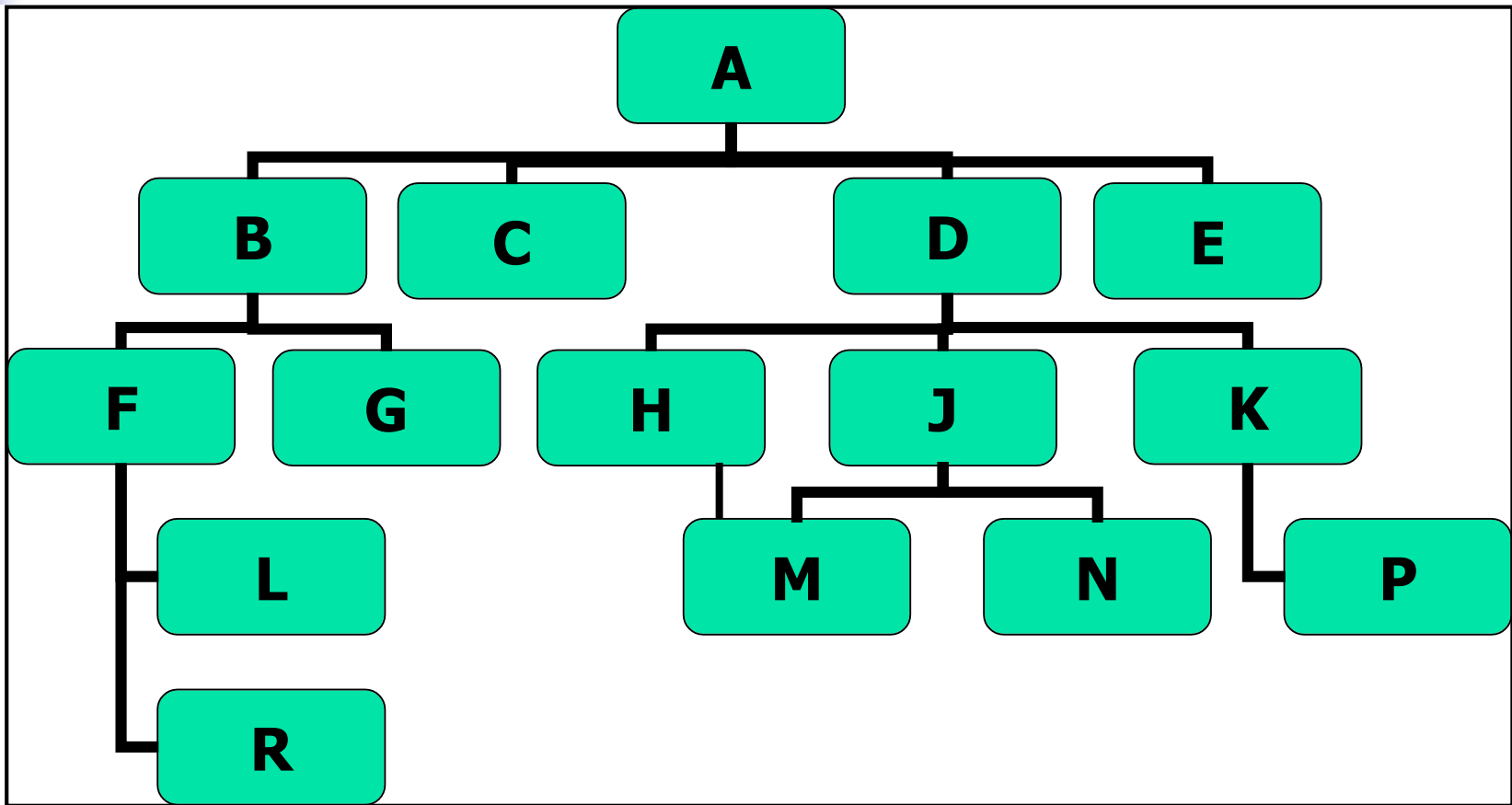
non-symmetrical thesaurus example – Spanish in red



non-symmetrical thesaurus example –
black where structure coincides



Compel both languages to accept loan terms, and you make it symmetrical!





Pros and cons of a non-symmetrical thesaurus

- Each language version retains its own individuality
- Less need to accept artificial concepts and terms
- Limitations on interoperability
- Does not conform to the same data model
- Hard to find software to manage all the language versions in concert



What needs for data models, formats and protocols?

- A unified model or metamodel to cover all vocabulary types?
- A model for each type of vocabulary?
- A syntax for specifying mappings between pairs of vocabularies? Or between tuples of vocabularies?
- 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