

A NEW STANDARD FOR THESAURUS CONSTRUCTION AND MANAGEMENT

Michèle Hudon, Ph.D.

Associate Professor

École de bibliothéconomie

et des sciences de l'information

EXISTING STANDARDS

- ISO2788-1986 *Guidelines for the establishment and development of monolingual thesauri*
- ISO5964-1985 *Guidelines for the establishment and development of multilingual thesauri*
- **BSI8723-2005 Structured vocabulary for information retrieval – Guide**
- ANSI/NISO Z39.19-2005 *Guidelines for the construction, format and management of monolingual controlled vocabularies*
- AFNOR Z47-100 1981 *Règles d'établissement des thésaurus monolingues* (+AFNOR Z47-101 1990)
- DIN 1463-1 and DIN 1463-2 *Erstellung und Weiterentwicklung von Thesauri*



OUTLINE

- ISO procedures
- TC46 SC9 WG8
- BS8723 as source
- ISO25964 Title and scope
- ISO25964-1
- ISO25964-2
- Difficulties and challenges
- Where we are at this point?



ISO PROCEDURES

- New work item proposal (Vote)
- Creation of a working group within a sub-committee
- Preparation of a working draft
- 1st Committee draft (International review – 3 months – Comments)
- 2nd Committee draft (International review, etc.)
- Draft international standard (Vote and comments – 5 months)
- Translation
- Final draft international standard (International review – 2 months)
- Publication



THE MAY 2007 PROPOSAL

Update and extend the scope of ISO2788 and ISO5964 to reflect the information environment of the 21st century. All of their existing scope would be retained and revised, with the following additional subjects to be added:

- guidance on electronic functions and displays
- functional specification for software to manage thesauri
- guidelines for certain additional types of vocabulary (e.g. classification schemes, taxonomies)
- interoperability (mapping) between vocabularies
- formats for exchange of thesaurus data

Proposal accepted in **July 2007**



TC46/SC9 WG8 1

- WG8 created officially in **October 2007**
 - Convenor : Stella Dextre Clarke (UK P)
 - Group members:
 - Leonard Will (UK P)
 - Sylvie Dalbin (France P)
 - Marianne Nielsen (Denmark A)
 - Esther Scheven (Germany L)
 - Marcia Lei Zeng (USA A)
 - Sophie Lessard (Canada P)
 - Michèle Hudon (Canada A)
 - External experts



TC46/SC9 WG8 2

○ Mandate

- develop a new ISO standard on structured vocabularies to replace the existing ISO standards on thesauri, i.e. ISO5964 (multilingual thesauri) published in 1985, and ISO2788 (monolingual thesauri) published in 1986

○ Timelines

- Original: October 2007 – October 2010
- Revised: January 2008 – December 2010 (Part 1 only)

○ Process

- Collaborative workspace provided by NISO
- Monthly teleconference (from July 2008-),
- Three face-to-face meetings to date for the whole group
- Two sub-groups to cover specific sections



USING BSI8723 AS WORKING DRAFT

British Standard 8723 used as the starting point for ISO Project 25964. BS 8723 consists of five parts, under the general title *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: Interoperability with applications



ISO25964 TITLE AND SCOPE 1

Information and Documentation – Thesauri and interoperability with other vocabularies

ISO25964 gives recommendations for the development and maintenance of structured vocabularies of subject terms and concepts for information retrieval. It also gives guidance on the exchange of data between multiple vocabularies and between a vocabulary and the software with which it interacts.



ISO25964 SCOPE 2

ISO25964 applies to vocabularies used for retrieving information from text resources, such as knowledge bases, bibliographic databases or collections of full text documents, and also catalogues of images or artefacts, such as those in a museum collection. Structured vocabularies are used in post-coordinate retrieval systems, hierarchical directories, pre-coordinate indexes and classification systems.



ISO25964 SCOPE 3

ISO25964 does not apply to the preparation of back-of-the-book indexes, *although many of its recommendations may be useful for that purpose.*

It does not cover the structure of authority files of proper names of people, organizations and places, while recognizing that these may sometimes interact with vocabularies of subject terms.



ISO25964 PART 1 1

- *Thesauri for information retrieval*
- General principles and rules for developing and managing monolingual and multilingual thesauri

- Contents
 - 2. Definitions
 - 5. Types of concepts in the thesaurus
 - 6. Thesaurus terms
 - Disambiguation
 - Grammatical form
 - Singular or plural form
 - Names
 - 7. Complex concepts (and factoring rules)



ISO25964 PART1 2

○ Contents

- 8. The equivalence relationship (monolingual context)
- 9. **Equivalence across languages**
 - Degrees of equivalence
 - Problems and solutions
 - Numerous links to Part 2 on mapping
- 10. Relationships between concepts
 - 10.2 Hierarchical relationship (incl. Multilingual thesauri)
 - 10.3 Associative relationship (incl. Multilingual thesauri)
- 12. Presentation and layout
- 13. Management issues (creation to updating)
- 14. Thesaurus management software
- 15. **Data model**

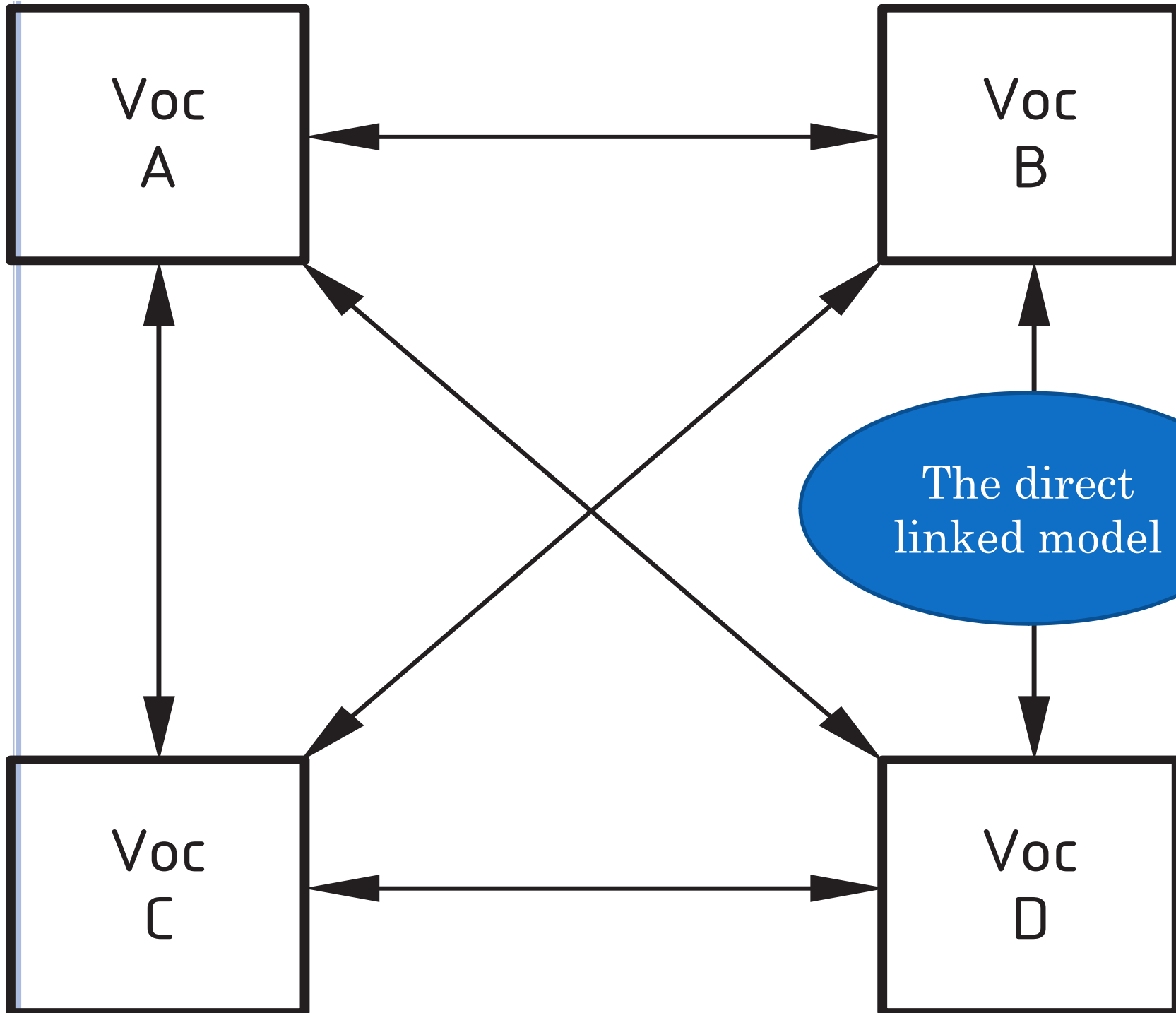


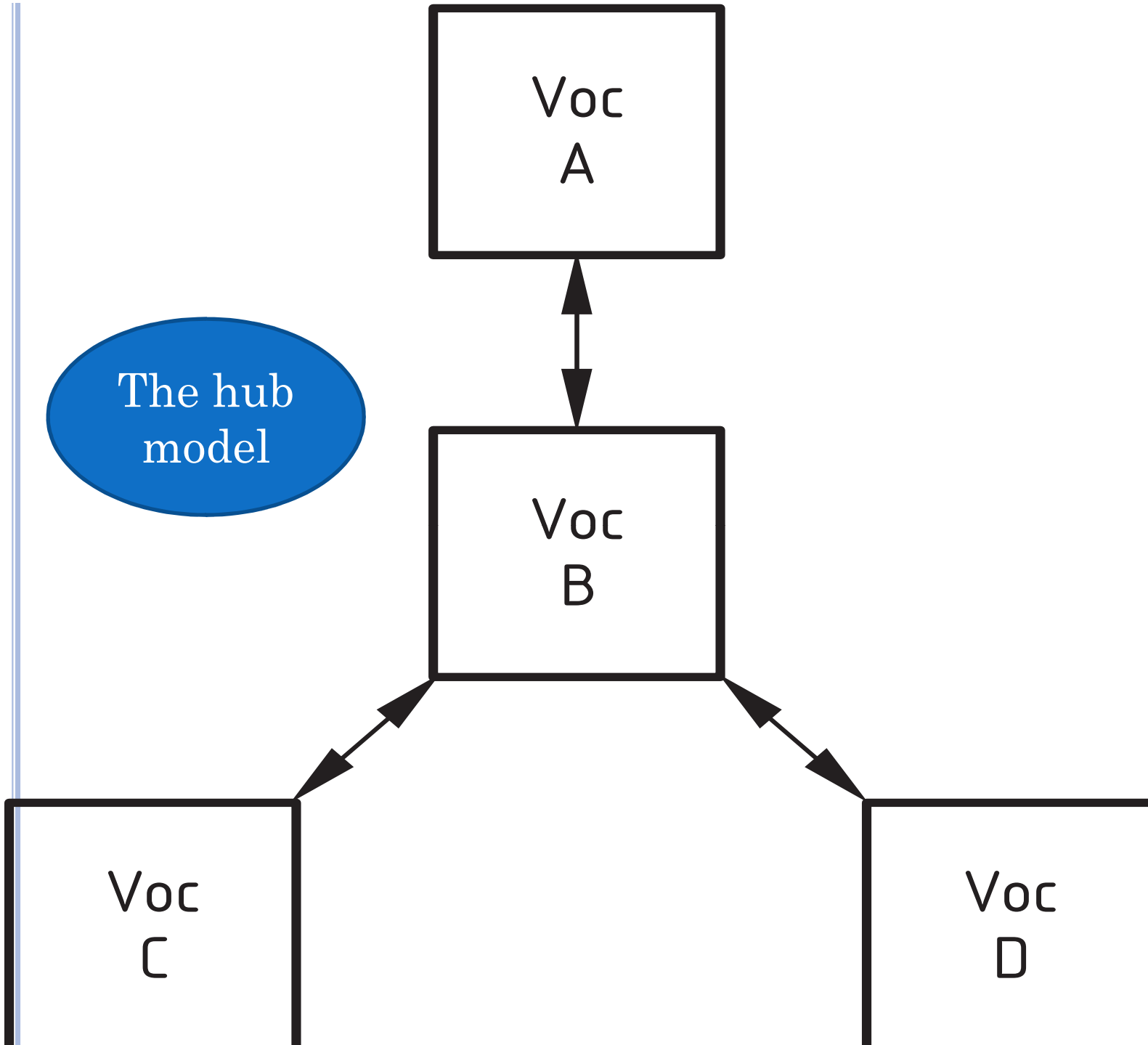
ISO25964 PART 2 1

- *Interoperability with other vocabularies*
- Addresses crosswalks and mapping

- Contents
 - 3. Definitions
 - *Interoperability* : ability of two or more systems to exchange information and to use the information that has been exchanged
 - *Mapping* : process of establishing relationships between the terms, notations or concepts of one vocabulary and those of another
 - 5. Objectives
 - 6. Structural models for mapping
 - Structural unity
 - Direct-linked
 - The hub







The hub model



ISO25964 PART 2 2

○ Contents

- 8. Equivalence mapping
 - Exact equivalence
 - Compound equivalence
 - Intersecting compound equivalence
 - Cumulative compound equivalence
 - Inexact equivalence
- 11. Implementation of mappings
- 12. Handling pre-coordination
- 15. Display of mapped vocabularies
- 16-23 Other vocabularies
 - Classification schemes
 - List of subject headings
 - Taxonomies
 - Terminologies
 - Authority files



DIFFICULTIES AND CHALLENGES 1

- 1. Team and project management
 - Number of people involved
 - Language
 - Pursuing a « conversation » through e-mail and a collaborative web site
 - Keeping track of documents and versions of documents
 - Dealing with suggestions from group members (all experts in their area, with extended experience)
 - Dealing with suggestions from ISO voting members (e.g. demand for more definitions)
 - Meeting deadlines



DIFFICULTIES AND CHALLENGES 2

ISC 27 May 2010

○ 2. Contents

- Standardizing terminology
- Defining terms
- What should be kept from previous standards?
- What should go in, what shouldn't? [e.g. Non symmetric structures]
- Multilingual and multicultural issues
- Data models
- Generally, how much detail to provide (is it necessary to anticipate all problems and issues...)



WHERE WE ARE AT THIS POINT?

- Committe draft (CD) circulated November 2008 for Part 1
- Draft international standard (DIS) issued August 2009 for Part 1
 - in English only; after consulting with AFNOR, it was decided that a French version would be produced at a later date
- Part 2 under development (June 2009)

