Typology for KRRs

Knowledge Representation Resources

Sue Ellen Wright
Kent State University, Kent, Ohio
NKOS-CENDI September 2008

Prescribed Parameters

- The range of “KOS” types
- Current standards and what is going on in some of these areas
- Efforts at harmonization & coordination
  - Issues in leveraging across system boundaries and between communities of practice
  - Examples: ISO TC 37 DCR, Terminology, SKOS, OWL-DL

©Sue Ellen Wright

Steps towards a Typology

- **Hodge, Gayle. 2000; 2004.**
  - Taxonomy of Knowledge Organization systems.
  - Types and Varieties of Terminology Resources: A Taxonomy of Knowledge Organization Systems.
- **Soergel, Dagobert. 2001**
  - Evaluation of Knowledge Organization Systems (KOS): Characteristics for describing and evaluating KOS.
- **Wright, Sue Ellen. 2006.**
  - A Typology for Knowledge Representation Resources, ISKO, Vienna
- **Tudhope, Doug. 2006.**
  - A Tentative Typology of KOS: Toward a KOS of KOS?
- **Hlava, Marjorie M.K. 2007.**
  - Insuring Compatibility and Crosswalks.

©Sue Ellen Wright

Steps towards a Typology

- **Hodge, Gayle. 2000; 2004.**
  - Taxonomy of Knowledge Organization systems. http://nkos.slis.kent.edu/KOS_taxonomy.htm
  - Types and Varieties of Terminology Resources: A Taxonomy of Knowledge Organization Systems. http://eea.eionet.europa.eu/Public/irc/envirowindows/jad/library?l=/ecoinformatics_indicator/thesaurusterminology/term-mtg-agenda-rev5doc/_EN_1.0_&a=d.
- **Soergel, Dagobert. 2001**
  - Evaluation of Knowledge Organization Systems (KOS): Characteristics for describing and evaluating KOS. http://nkos.slis.kent.edu/2001/SoergelCharacteristicsOfKOS.pdf

©Sue Ellen Wright

Steps towards a Typology

- **Wright, Sue Ellen. 2006.**
  - A Typology for Knowledge Representation Resources, ISKO, Vienna
- **Tudhope, Doug. 2006.**
  - A Tentative Typology of KOS: Toward a KOS of KOS? www.ukoln.ac.uk/nkos/nkos2006/presentations/tudhope.ppt
- **Hlava, Marjorie M.K. 2007.**
  - Insuring Compatibility and Crosswalks. www.comp.glam.ac.uk/pages/research/hypermedia/n**kos**/n**kos**2007/presentations/N**KOS**%202007-HLava.ppt

©Sue Ellen Wright

KOS

- **TC 37 concept system:**
  - set of **concepts (3.2.1) structured according to the** relations among them
- **Knowledge Organization Schemes**
  - **"a set of concepts, *optionally* including statements about semantic relationships between those concepts"**
  - **⁂ KOS under this definition potentially includes both systematic and non-systematic resources.**
- **Resources with latent concept systems**

©Sue Ellen Wright

Knowledge Representation Resources (KRRs)

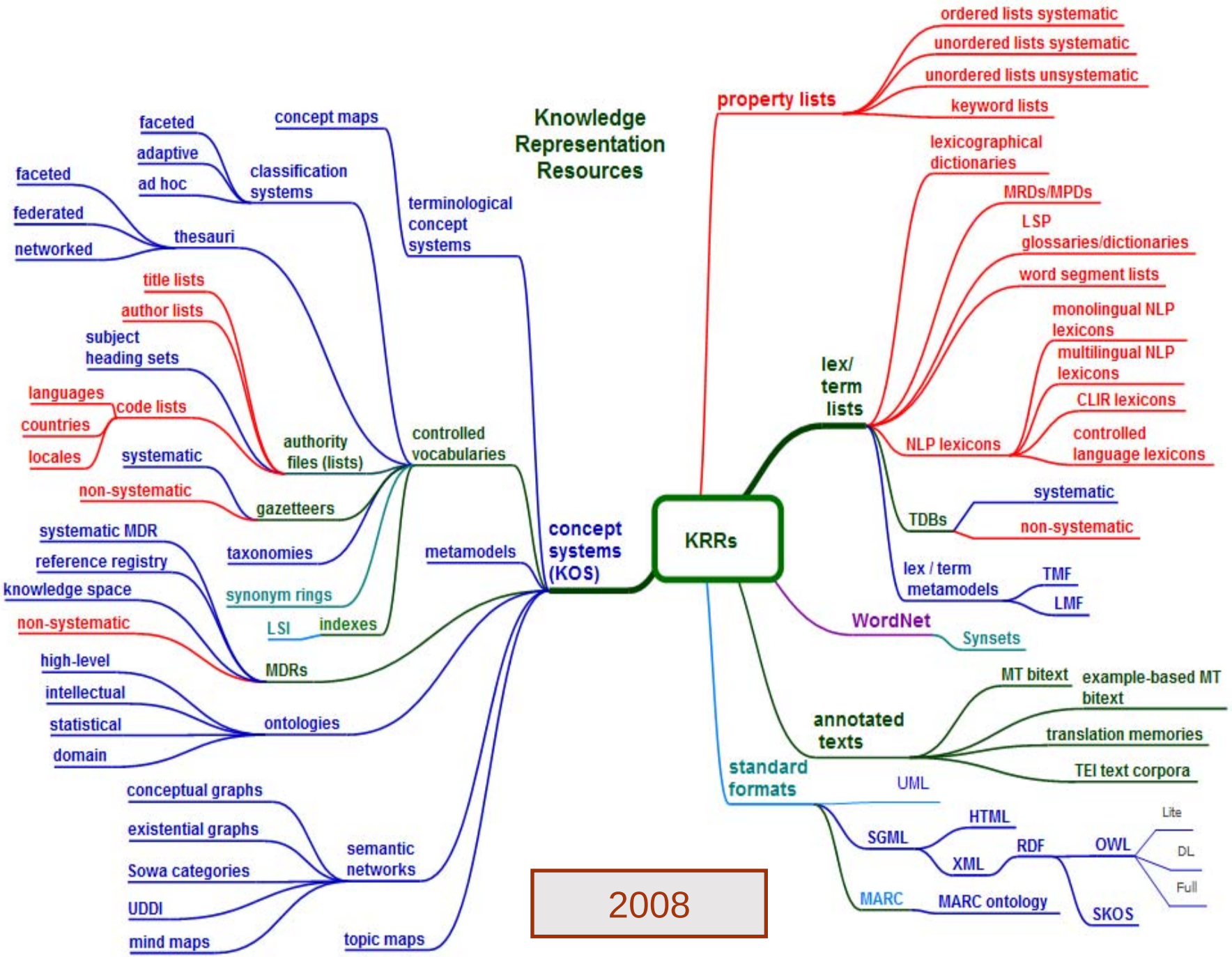
- Resources that are in themselves well "organized" (systematic)
- Resources that contain latent organizational elements (non-systematic)
- Resource that contains knowledge we can :
  - Manipulate
  - Mine or use to enrich other resources
  - Analyze and reuse (leverage)
  - Use to interact with various tools, either based on common environment planning (or not)

©Sue Ellen Wright

Principle of Systematicity

- Systematic resources include explicitation of relationships (parent-child, meronymy & metonymy, sequentiality, defined edges in logical triads (RDF), etc.)
- Non-systematic resources are not ordered or are conventionally ordered (alphabetical dictionaries, non-mnemonic numerical sequences, etc.)
- Some resources (terminologies, metadata registries [MDRs], etc.) are manifested in both variations
- Degrees of systematicity

©Sue Ellen Wright

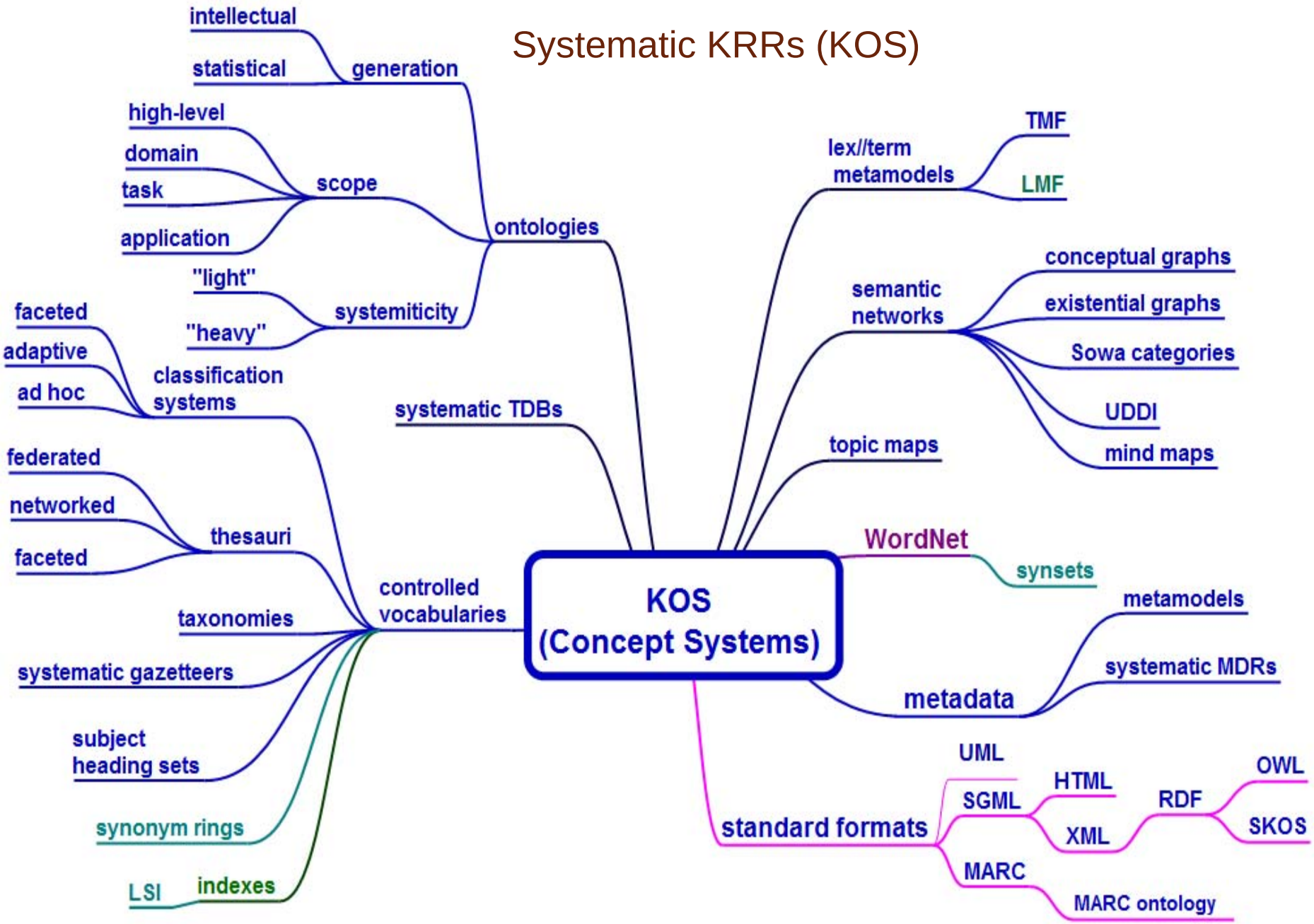

Knowledge Representation Resources
KRRs
concept systems (KOS)
terminological concept systems
concept maps
classification systems
faceted
adaptive
ad hoc
thesauri
faceted
federated
networked
controlled vocabularies
authority files (lists)
title lists
author lists
subject heading sets
code lists
languages
countries
locales
gazetteers
systematic
non-systematic
taxonomies
synonym rings
indexes
LSI
metamodels
MDRs
systematic MDR
reference registry
knowledge space
non-systematic
ontologies
high-level
intellectual
statistical
domain
semantic networks
conceptual graphs
existential graphs
Sowa categories
UDDI
mind maps
topic maps
property lists
ordered lists systematic
unordered lists systematic
unordered lists unsystematic
keyword lists
lex/ term lists
lexicographical dictionaries
MRDs/MPDs
LSP glossaries/dictionaries
word segment lists
NLP lexicons
monolingual NLP lexicons
multilingual NLP lexicons
CLIR lexicons
controlled language lexicons
TDBs
systematic
non-systematic
lex / term metamodels
TMF
LMF
WordNet
Synsets
annotated texts
MT bitext
example-based MT bitext
translation memories
TEI text corpora
standard formats
UML
SGML
HTML
XML
RDF
OWL
Lite
DL
Full
MARC
MARC ontology
SKOS
2008

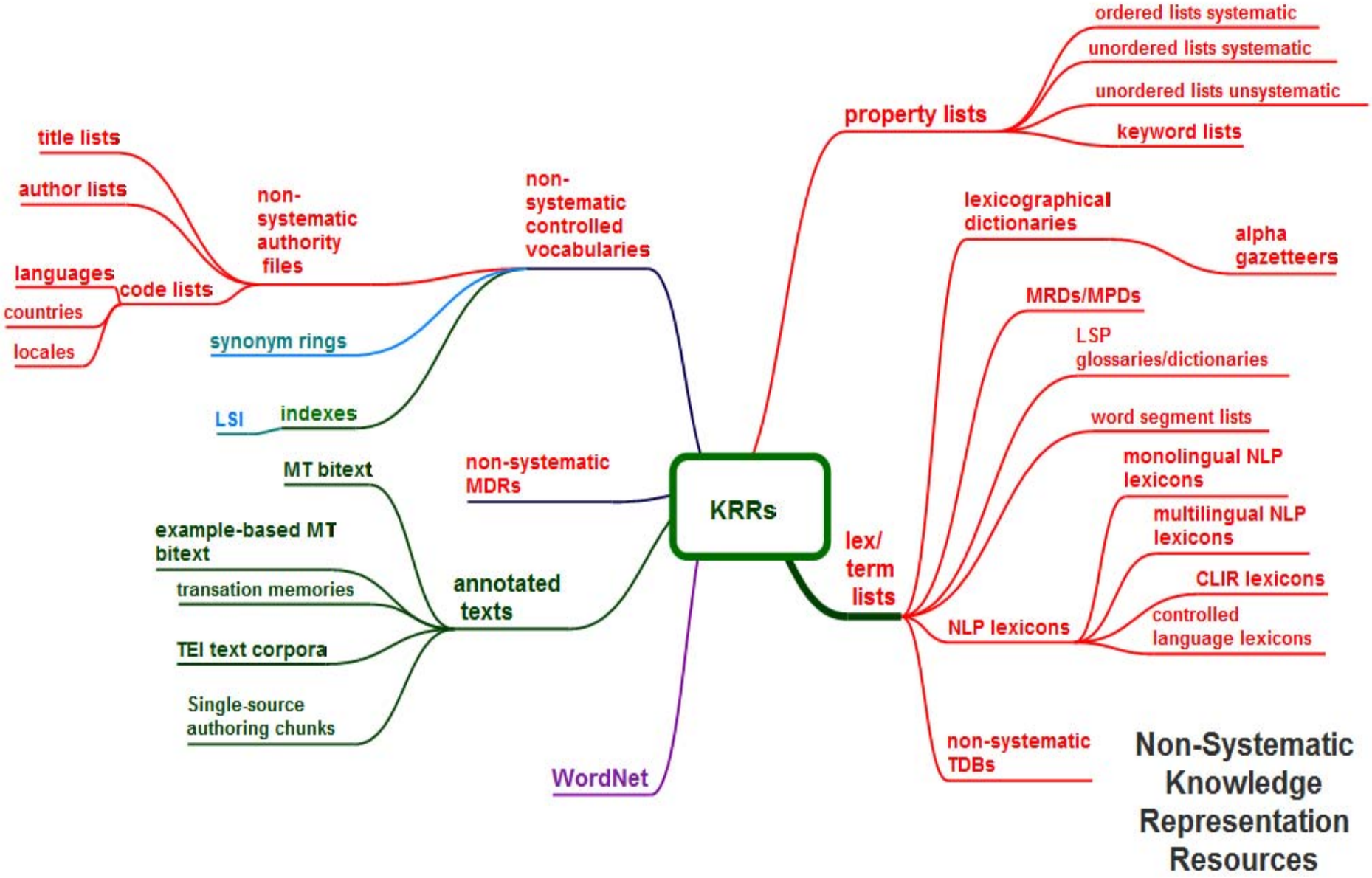
Colors

- **Blue: systematic, represents shallow to deep semantic structures**
- **Red: non-systematic, primarily lists with random or conventional (e.g., alphabetical) ordering principles**
- **Green: hybrid superordinate nodes with both systematic and non-systematic children; texts of various kinds**
- **Purple: WordNet: internally hybrid system; shallow systematics, lexicographical approach**

©Sue Ellen Wright

Systematic KRRs (KOS)

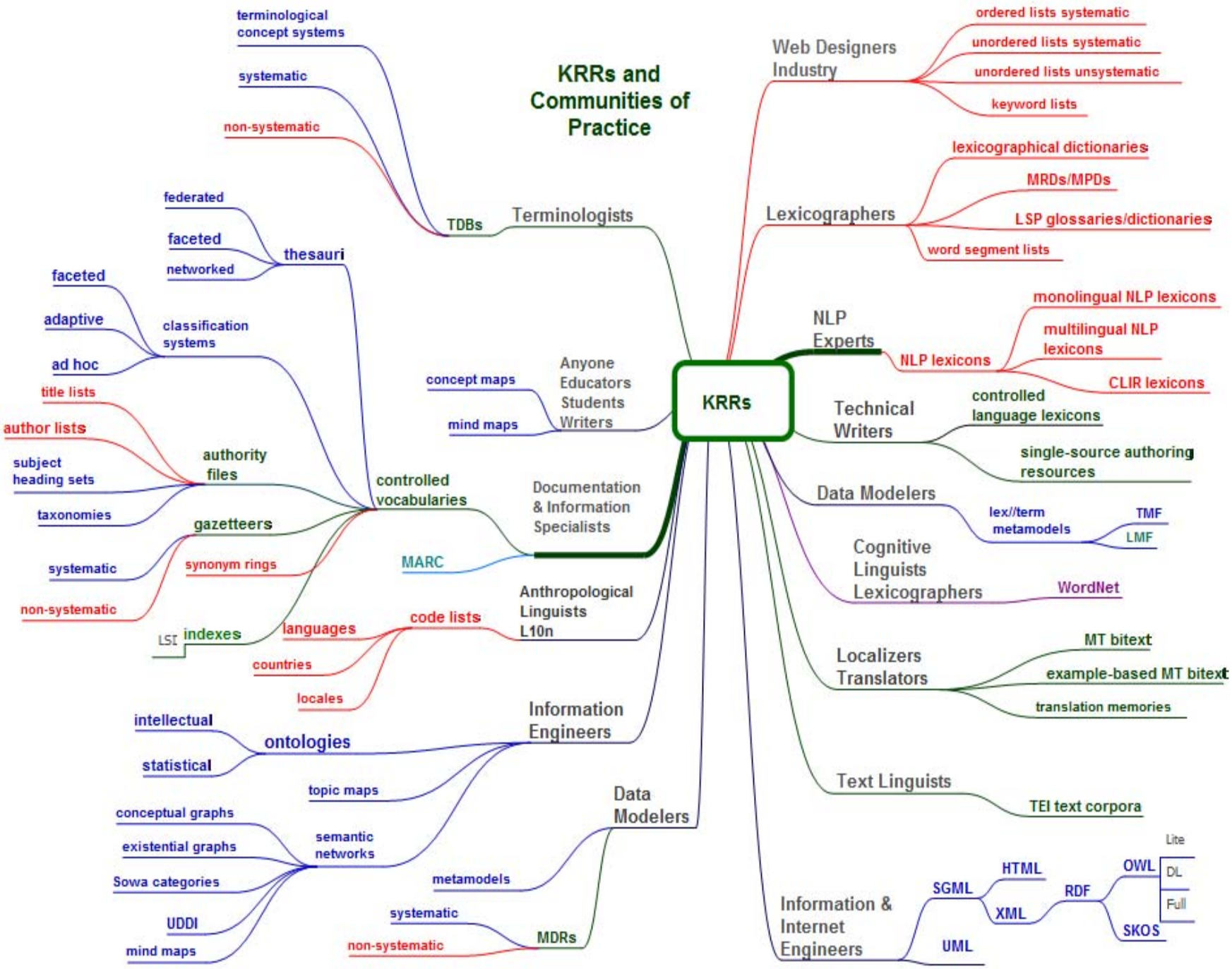



KRRs
non-systematic MDRs
WordNet
non-systematic controlled vocabularies
non-systematic authority files
title lists
author lists
code lists
languages
countries
locales
synonym rings
indexes
LSI
annotated texts
MT bitext
example-based MT bitext
transation memories
TEI text corpora
Single-source authoring chunks
lex/ term lists
property lists
ordered lists systematic
unordered lists systematic
unordered lists unsystematic
keyword lists
lexicographical dictionaries
alpha gazetteers
MRDs/MPDs
LSP glossaries/dictionaries
word segment lists
monolingual NLP lexicons
multilingual NLP lexicons
CLIR lexicons
controlled language lexicons
NLP lexicons
non-systematic TDBs
Non-Systematic Knowledge Representation Resources

Communities of Practice

- Different communities of practice (CoPs) using different terms for the same concepts and the same terms for different concepts
- KOS defined differently by different CoPs
- Critical concepts subject to indeterminacy
  - *e.g., term, terminology, metadata, ontology*
- Result: indeterminacy in the form of hidden polysemy and synonymy

©Sue Ellen Wright


KRRs and Communities of Practice
KRRs
Terminologists
TDBs
terminological concept systems
systematic
non-systematic
thesauri
federated
faceted
networked
classification systems
faceted
adaptive
ad hoc
controlled vocabularies
authority files
title lists
author lists
subject heading sets
taxonomies
gazetteers
systematic
non-systematic
synonym rings
indexes
LSI
Anyone
Educators
Students
Writers
concept maps
mind maps
Documentation & Information Specialists
MARC
Anthropological Linguists L10n
languages
code lists
countries
locales
Information Engineers
ontologies
intellectual
statistical
topic maps
semantic networks
conceptual graphs
existential graphs
Sowa categories
UDDI
mind maps
Data Modelers
metamodels
MDRs
systematic
non-systematic
Web Designers Industry
ordered lists systematic
unordered lists systematic
unordered lists unsystematic
keyword lists
Lexicographers
lexicographical dictionaries
MRDs/MPDs
LSP glossaries/dictionaries
word segment lists
NLP Experts
NLP lexicons
monolingual NLP lexicons
multilingual NLP lexicons
CLIR lexicons
Technical Writers
controlled language lexicons
single-source authoring resources
Data Modelers
lex//term metamodels
TMF
LMF
Cognitive Linguists Lexicographers
WordNet
Localizers Translators
MT bitext
example-based MT bitext
translation memories
Text Linguists
TEI text corpora
Information & Internet Engineers
SGML
HTML
XML
RDF
OWL
SKOS
Lite
DL
Full
UML

Language-Purposed Vocabulary: terminology

- ***Terminology (TC 37 for Terminology)***
  - **Definition: Set of designations belonging to a special language (reflecting the concepts used in that special language)**
  - **Terminologies (and glossaries) document the use of these designations in discourse, i.e., *uncontrolled vocabulary* used for communication rather than their use for documentation and retrieval as part of a *controlled vocabulary*.**
  - **They can contain concept systems.**
  - **They can also be used to enrich other resources.**

©Sue Ellen Wright

Subject-Purposed Vocabulary: terminology

- **thesaurus-related *term*:**
  - **One or more words designating a concept**
  - **A descriptor in a controlled vocabulary used for information management & retrieval**
  - **Thesaurus developers tend to use *terminology* as a synonym for *thesaurus*, whereas terminologists consider terminologies and even terminological concept systems to be different from thesauri.**
  - **This distinction led to a hypothesis: need to identify crosswalk nodes between the systems in order to achieve interoperability.**

©Sue Ellen Wright

Terminologies

- *SKOS: Subject-purposed vocabularies = Subject Language Terminologies*
  - *Svenonius: SLTs*
  - *E.g., thesauri: controlled vocabularies*
  - *Information storage & retrieval*
- *TBX: Language-purposed vocabularies*
  - *Tudhope*
  - *Terminological Databases (TDBs)*
  - *Uncontrolled vocabularies; discourse oriented*
- *Issues involving mapping data elements between the two CoPs*

Hypothesis: TDB Term = SKOS Term (Label)?

SKOS classes and properties	12620 data category → SKOS	12620 data category ≠ Not in SKOS
Labels and Terms		
label	N/A	term (A.1)

- Svenonius:
  - SLTs designed for the special purpose of retrieving information; extension = all *documents about a subject* (e.g., all documents about butterflies)
  - TDBs designed to document terms used in discourse; extension = the class consisting of *all objects covered by the concept* (e.g., all butterflies)

Hypothesis: Definition = Definition?

Definitions		
definition	definition (A.5.1)	(See discussion)
scopeNote	explanation (A.5.2)	(See discussion)

- SKOS definition: *A statement or formal explanation of the meaning of a concept*.
- SKOS scopeNote: *A note that helps to clarify the meaning of a concept*.
- These declarations coincide ostensibly with TBX */definition/* and */explanation/*, respectively.

Problems with Proposed Solution

- Skos examples:

```
<skos:Concept rdf:about="http://my.example.org/GCL/702#scopeNote">
 <skos:prefLabel xml:lang="en">    Competitiveness</skos:prefLabel>
 <skos:scopeNote xml:lang="en">The ability of businesses to compete in
local, national or international markets.</skos:scopeNote>
</skos:Concept>

<skos:Concept rdf:about="http://www.example.com/concepts#banana">
 <skos:prefLabel xml:lang="en">  banana</skos:prefLabel>
 <skos:definition xml:lang="en">A long curved fruit with a yellow skin
and soft, sweet white flesh inside.</skos:definition>
</skos:Concept>
```

- Problem: Both of these items are formal definitions in the sense of TBX (ISO 704)

SKOS Example

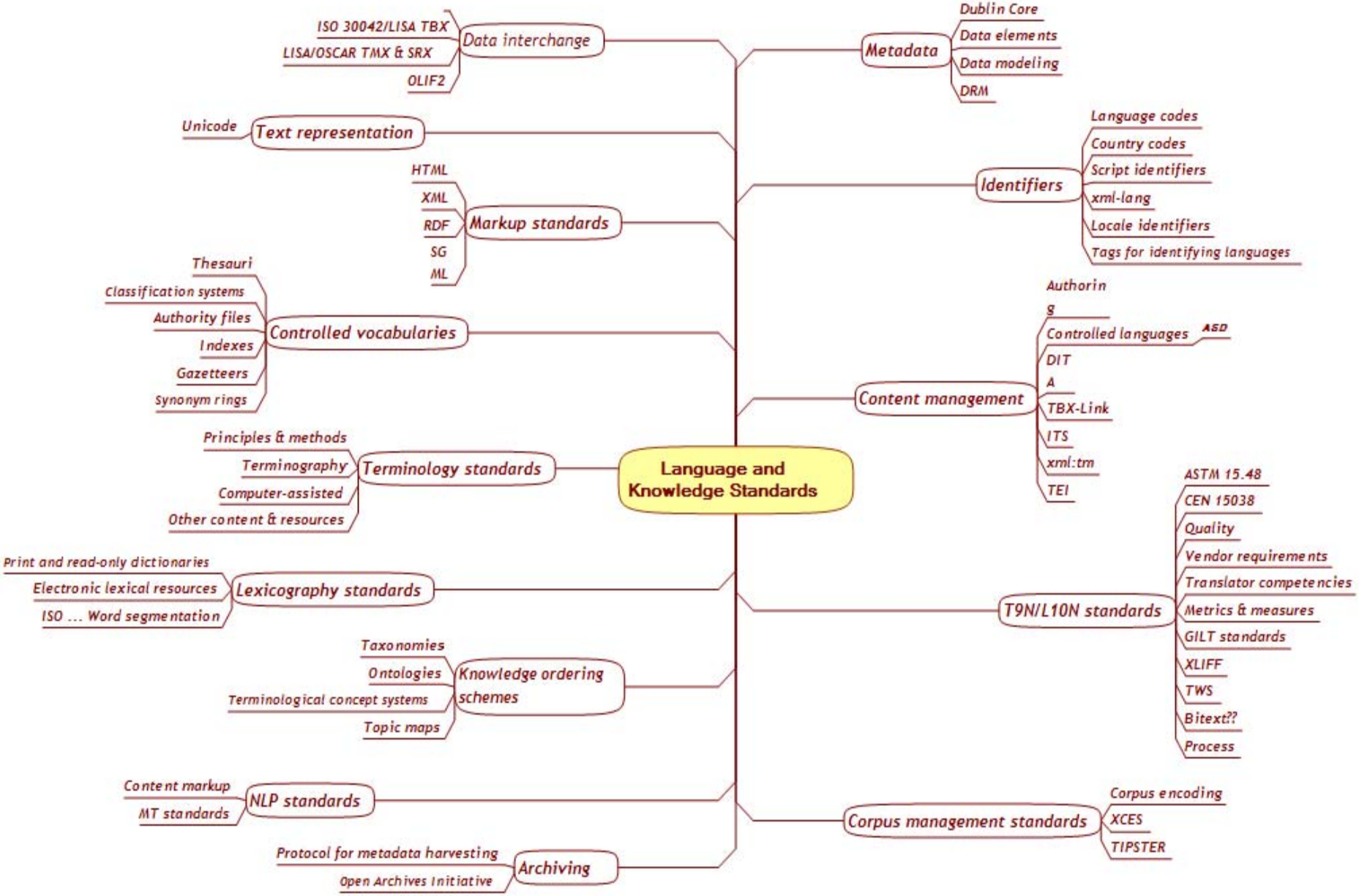
<skos:definition>A feature type category for places such as the Erie Canal </skos:definition>
<skos:scopeNote>Manmade waterway used by watercraft or for drainage, irrigation, mining, or water power</skos:scopeNote> (Mikhalenko 2005)

- This approach, if followed, would suggest flipping the crosswalk assignment, with TBX */definition/* mapping to SKOS *scopeNote*, and TBX */explanation/* mapping to SKOS *definition*!

Standards for Language & Knowledge Organization

- Wide range of subject areas
- Focus on the so-called “language industry,” much of which is potentially interesting for information science perspectives
- Wide range of Communities of Practice
- Wide range of standardizing & authoritative bodies
- Overlapping interests & projects

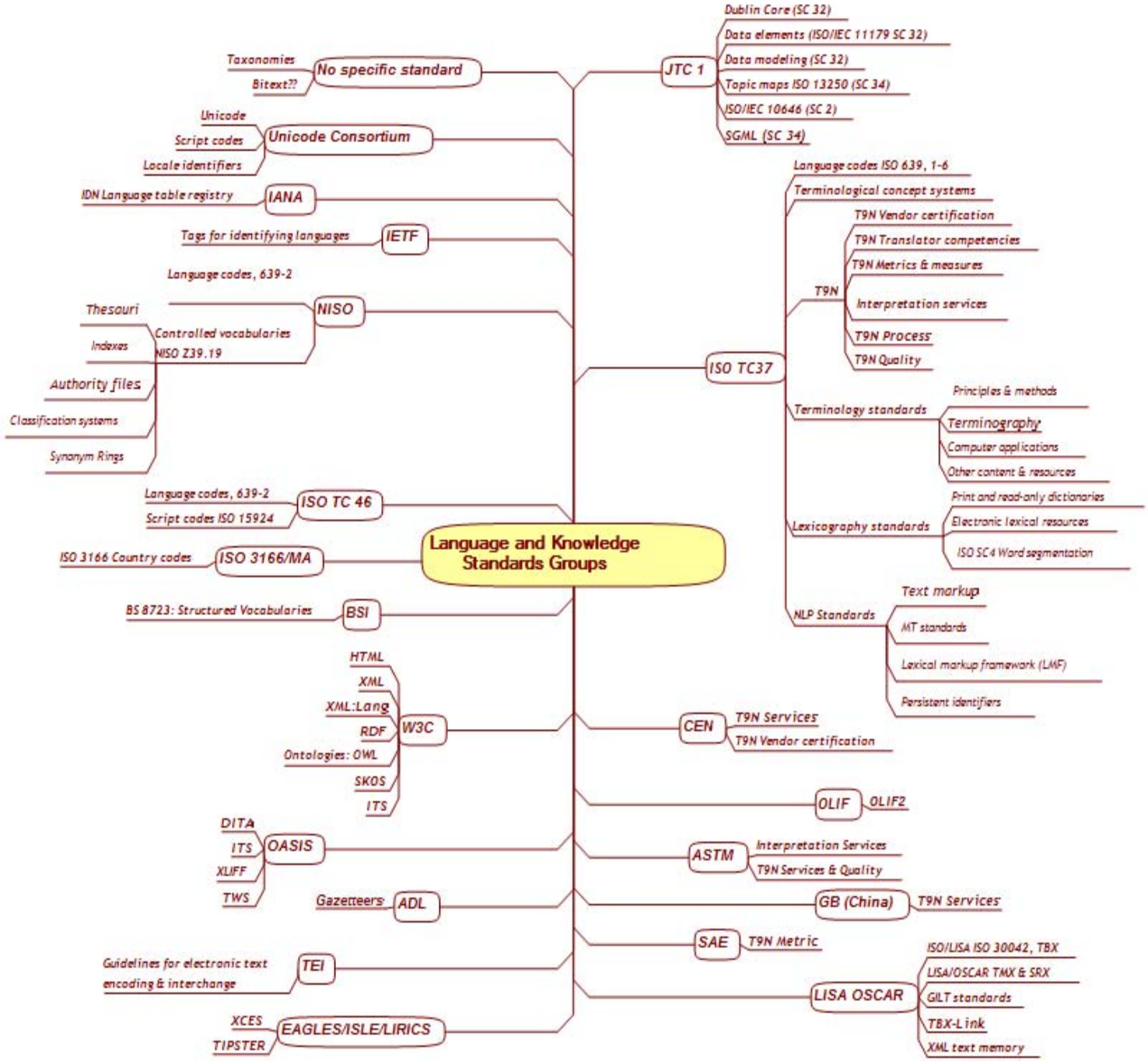
©Sue Ellen Wright


Language and Knowledge Standards
Data interchange
ISO 30042/LISA TBX
LISA/OSCAR TMX & SRX
OLIF2
Text representation
Unicode
Markup standards
HTML
XML
RDF
SG
ML
Controlled vocabularies
Thesauri
Classification systems
Authority files
Indexes
Gazetteers
Synonym rings
Terminology standards
Principles & methods
Terminography
Computer-assisted
Other content & resources
Lexicography standards
Print and read-only dictionaries
Electronic lexical resources
ISO ... Word segmentation
Knowledge ordering schemes
Taxonomies
Ontologies
Terminological concept systems
Topic maps
NLP standards
Content markup
MT standards
Archiving
Protocol for metadata harvesting
Open Archives Initiative
Metadata
Dublin Core
Data elements
Data modeling
DRM
Identifiers
Language codes
Country codes
Script identifiers
xml-lang
Locale identifiers
Tags for identifying languages
Content management
Authoring
Controlled languages
ASD
DITA
TBX-Link
ITS
xml:tm
TEI
T9N/L10N standards
ASTM 15.48
CEN 15038
Quality
Vendor requirements
Translator competencies
Metrics & measures
GILT standards
XLIFF
TWS
Bitext??
Process
Corpus management standards
Corpus encoding
XCES
TIPSTER

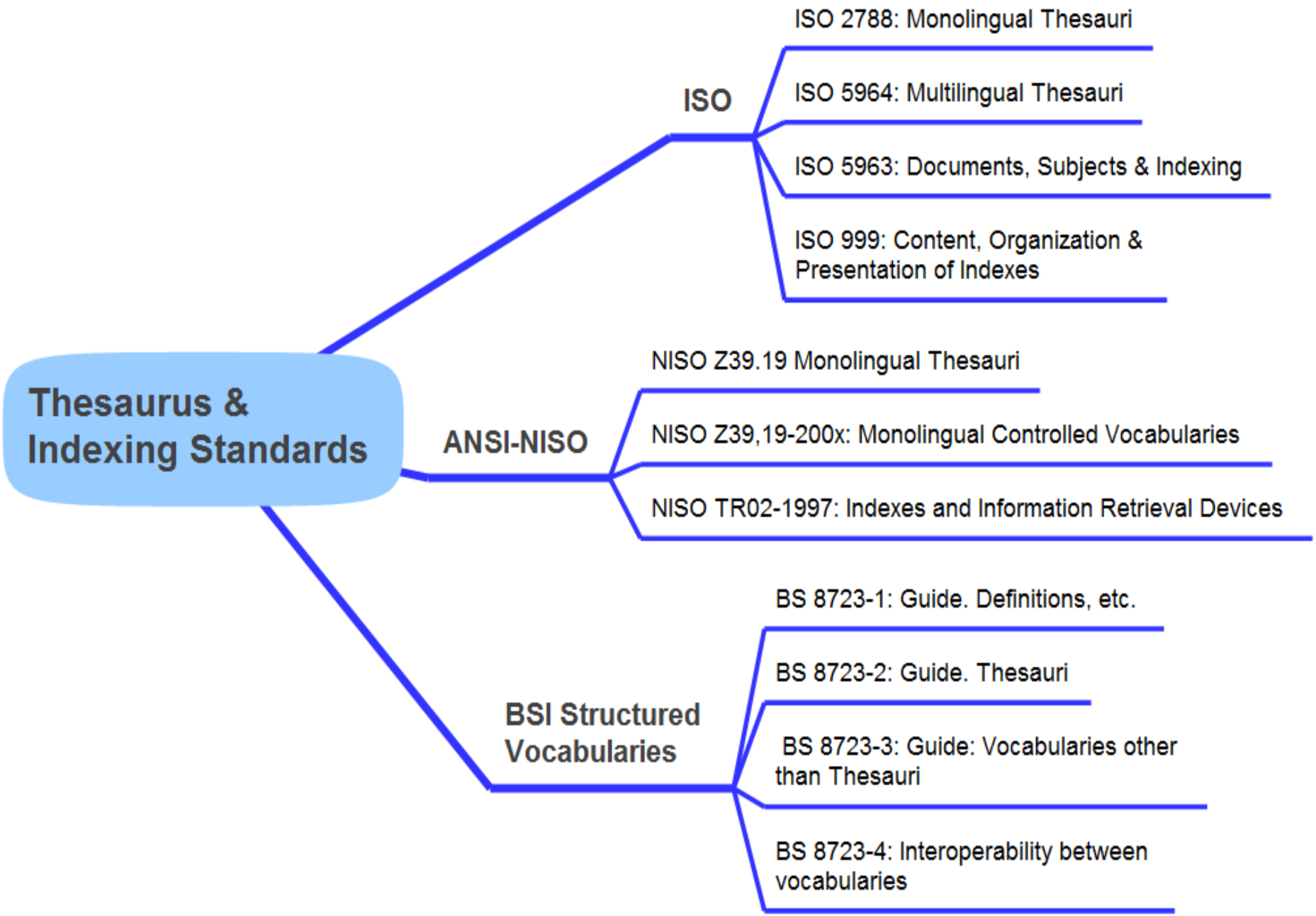
Standards Bodies

- ISO family of standards bodies
- National bodies (NISO, DIN, BSI, etc.)
- Web-oriented standards bodies (W3C, IETF, Unicode, etc.)
- Industry standards (OMG, OASIS, LISA, etc.)
- Professional organizations (ATA, FIT, etc.)
- Research groups and grant teams
- Others?

©Sue Ellen Wright


Language and Knowledge Standards Groups
No specific standard
Taxonomies
Bitext??
Unicode Consortium
Unicode
Script codes
Locale identifiers
IANA
IDN Language table registry
IETF
Tags for identifying languages
NISO
Language codes, 639-2
Thesauri
Controlled vocabularies
Indexes
NISO Z39.19
Authority files
Classification systems
Synonym Rings
ISO TC 46
Language codes, 639-2
Script codes ISO 15924
ISO 3166/MA
ISO 3166 Country codes
BSI
BS 8723: Structured Vocabularies
W3C
HTML
XML
XML:Lang
RDF
Ontologies: OWL
SKOS
ITS
OASIS
DITA
ITS
XLIFF
TWS
ADL
Gazetteers
TEI
Guidelines for electronic text encoding & interchange
EAGLES/ISLE/LIRICS
XCES
TIPSTER
JTC 1
Dublin Core (SC 32)
Data elements (ISO/IEC 11179 SC 32)
Data modeling (SC 32)
Topic maps ISO 13250 (SC 34)
ISO/IEC 10646 (SC 2)
SGML (SC 34)
ISO TC37
Language codes ISO 639, 1-6
Terminological concept systems
T9N
T9N Vendor certification
T9N Translator competencies
T9N Metrics & measures
Interpretation services
T9N Process
T9N Quality
Terminology standards
Principles & methods
Terminography
Computer applications
Other content & resources
Lexicography standards
Print and read-only dictionaries
Electronic lexical resources
ISO SC4 Word segmentation
NLP Standards
Text markup
MT standards
Lexical markup framework (LMF)
Persistent identifiers
CEN
T9N Services
T9N Vendor certification
OLIF
OLIF2
ASTM
Interpretation Services
T9N Services & Quality
GB (China)
T9N Services
SAE
T9N Metric
LISA OSCAR
ISO/LISA ISO 30042, TBX
LISA/OSCAR TMX & SRX
GILT standards
TBX-Link
XML text memory

Terminology Standards
Principles of terminology
ISO1087: Terminology work -- Vocabulary -- Part 1: Theory and application, Part 2: Computer applications
ISO 704: Terminology work -- Principles and Methods
ISO TR 24156: Guidelines for using UML notation in terminology work
Terminography
ISO 10241: Terminology entries in standards
ISO 1951: Presentation/representation of entries in dictionaries
Computer applications in terminology
ISO DIS 26162: Design, implementation and maintenance of Terminology Management Systems
ISO 12620:2009 Data Category Registry (DCR)
ISO 16642: Terminology Markup Framework (TMF)
ISO 30042: TBX TermBase eXchange
Other content & resources
ISO 24618: Language resource management -- Citation of Electronic Resources
ISO 24613: Lexical Markup Framework (LMF)
ISO 24615:. Word Segmentation of written texts for monolingual and multilingual information processing

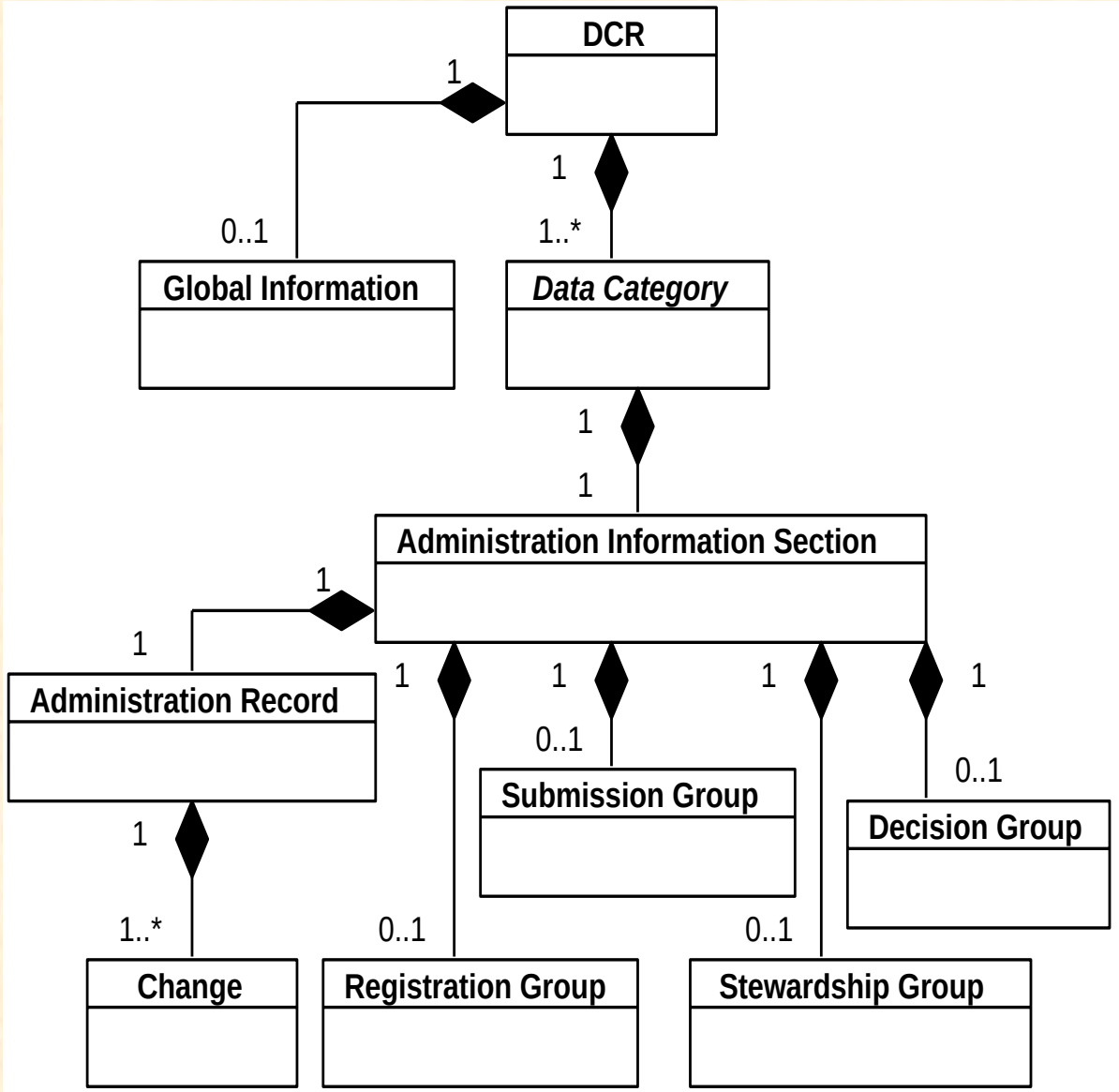

Thesaurus & Indexing Standards
ISO
ISO 2788: Monolingual Thesauri
ISO 5964: Multilingual Thesauri
ISO 5963: Documents, Subjects & Indexing
ISO 999: Content, Organization & Presentation of Indexes
ANSI-NISO
NISO Z39.19 Monolingual Thesauri
NISO Z39,19-200x: Monolingual Controlled Vocabularies
NISO TR02-1997: Indexes and Information Retrieval Devices
BSI Structured Vocabularies
BS 8723-1: Guide. Definitions, etc.
BS 8723-2: Guide. Thesauri
BS 8723-3: Guide: Vocabularies other than Thesauri
BS 8723-4: Interoperability between vocabularies

Current Projects

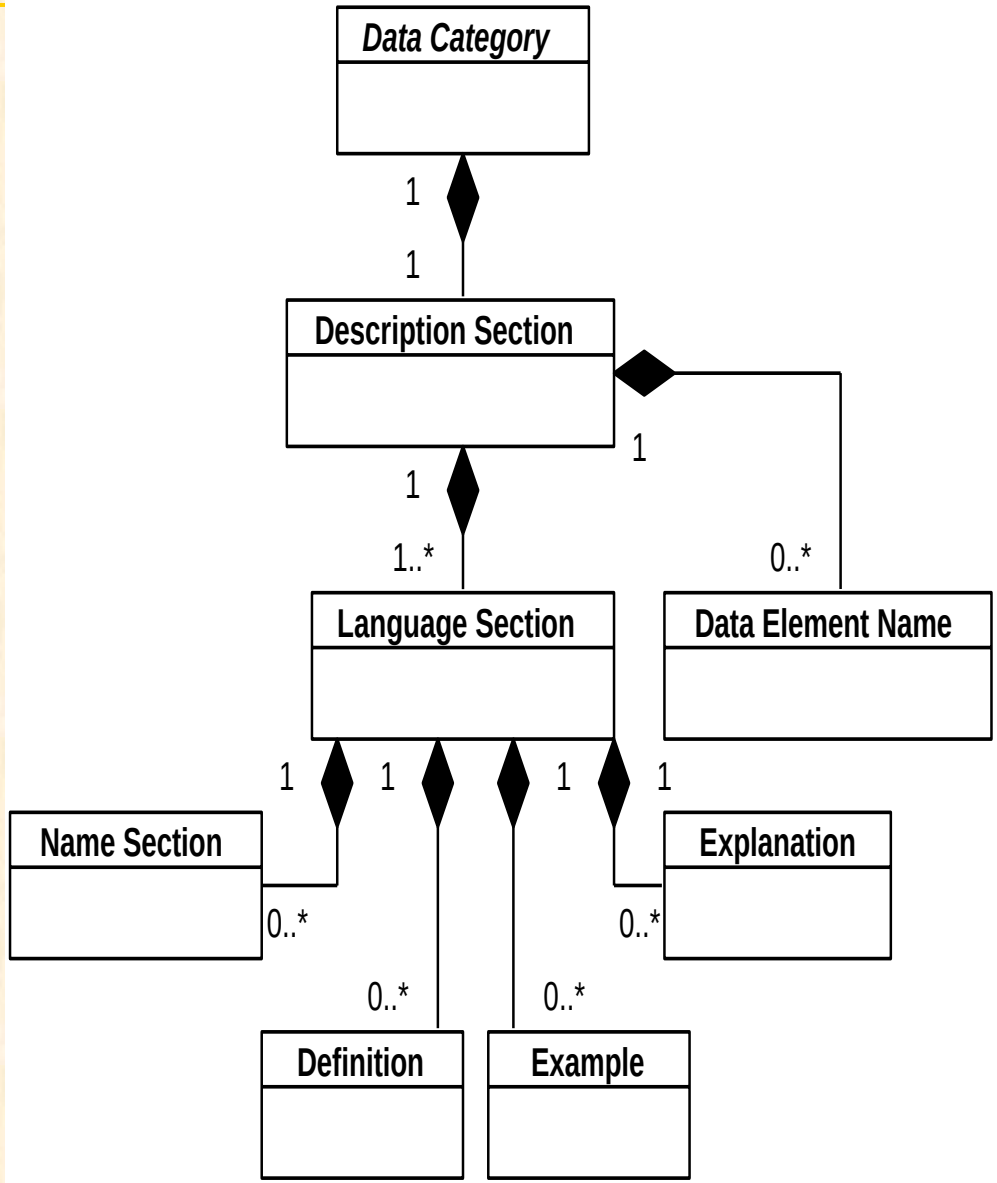
- Completion of the TC 37 Data Category Registry
- Provision for flexible RDF output from the DCR to support interchange and interoperability
- Experimentation with RDF-based views of critical standards, such as TBX

©Sue Ellen Wright

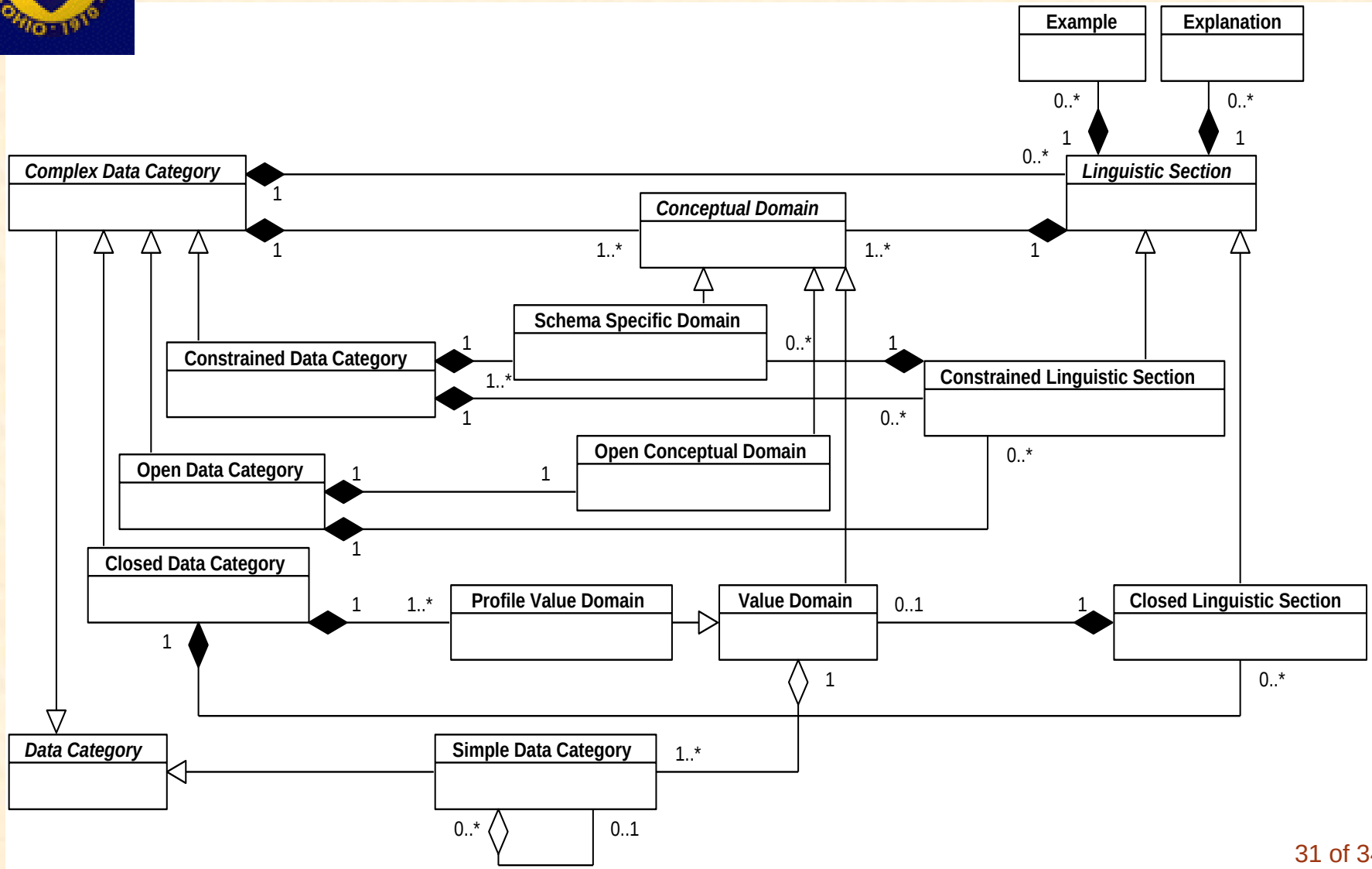
Data Category Specification Administration Section



Data Category Specification Description Section



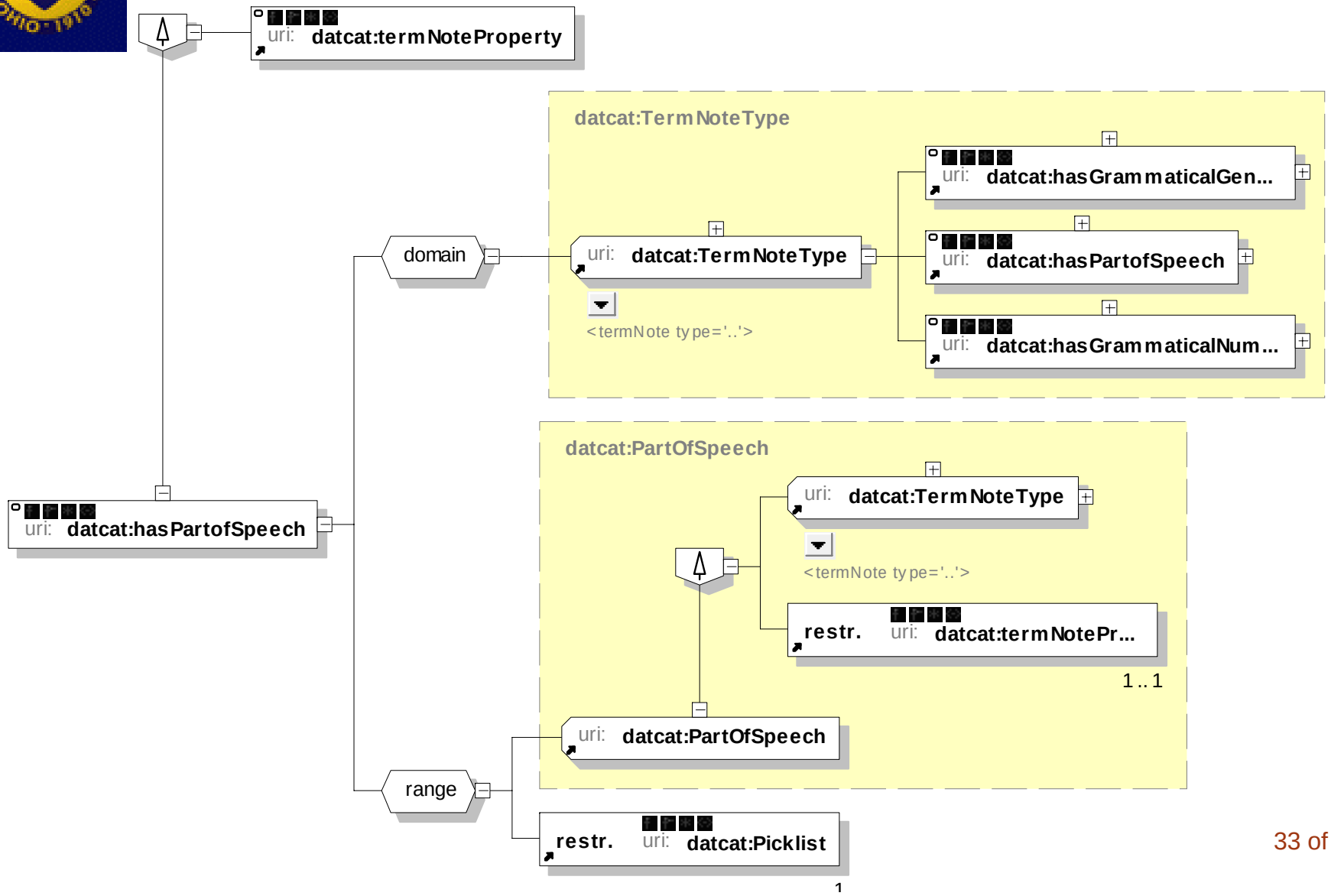
Data Category Specification Linguistic Section



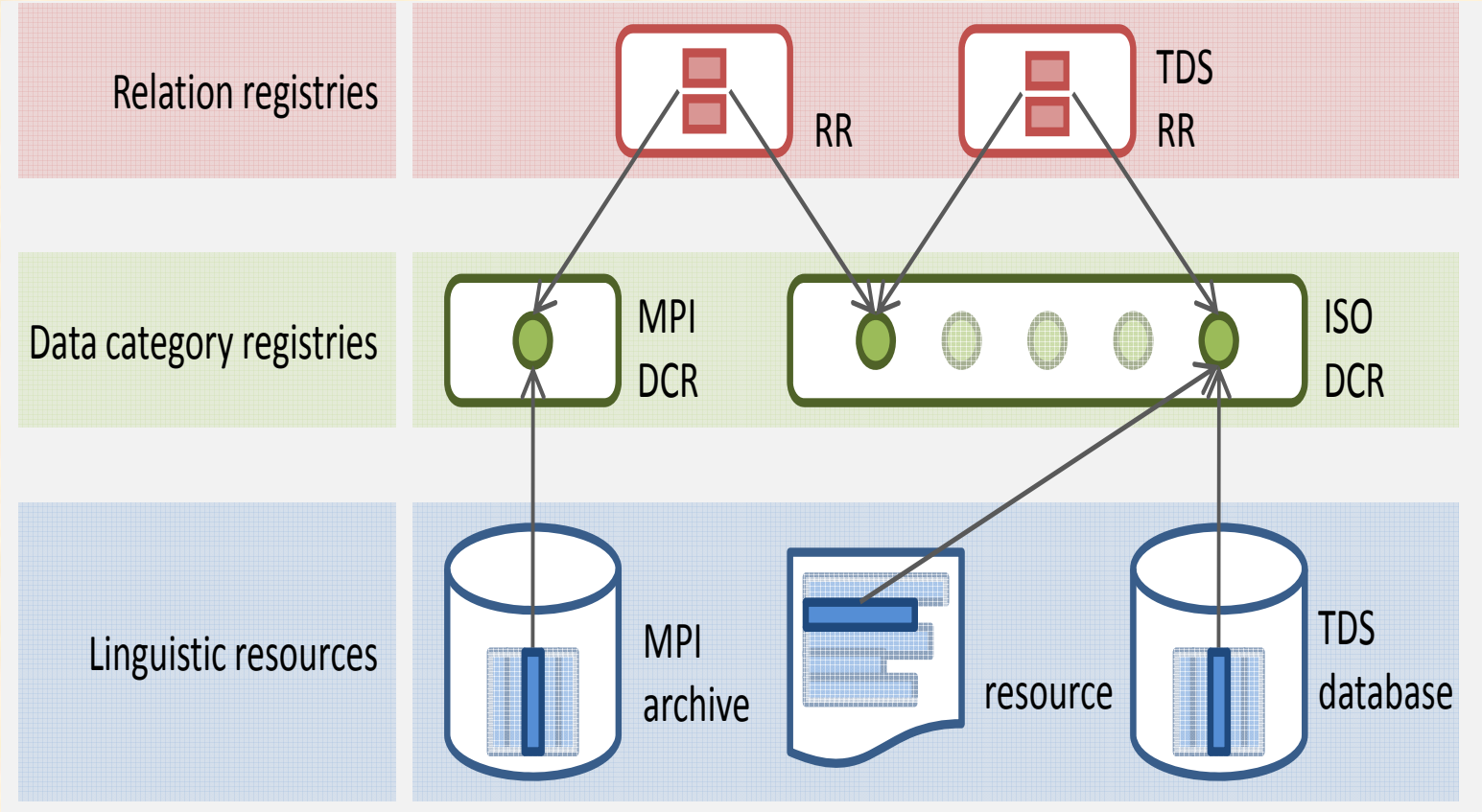
http://www.isocat.org

- Alpha/beta test due Oct. 2008
- Planned as open source software
- Special features:
  - ISO balloting procedures
  - Multilingual solutions
  - Metadata Registry In the spirit of ISO 11179

OWL-DL Representation of TBX



Navigating Relation Registries



©Sue Ellen Wright

For More Information

- Sue Ellen Wright
  Institute for Applied Linguistics
  Kent State University
  109 Satterfield Hall
  Kent, Ohio 44242, USA

  sellenwright@gmail.com

©Sue Ellen Wright