

# Metadata and Thesauri

What they are and how they can be used

Angela Vorndran  
vorndran@dipf.de

Semantic Analysis Meeting  
22 January, 2010, Grenoble

Metadata and Thesauri – Angela Vorndran – Semantic Analysis Meeting, 22 January, 2010, Grenoble



[www.eerqi.eu](http://www.eerqi.eu)



**A Project funded under the  
Socio-Economic Sciences  
and Humanities Theme.**

# Metadata

Metadata and Thesauri – Angela Vorndran – Semantic Analysis Meeting, 22 January, 2010, Grenoble



[www.eerqi.eu](http://www.eerqi.eu)



**A Project funded under the  
Socio-Economic Sciences  
and Humanities Theme.**



# Metadata

- Definition: data about data
- In our case: all information describing a document in a full text database
  - bibliographic metadata
  - Subject indexing
- Structured format to facilitate identification, retrieval and organization of documents and collections

Metadata and Thesauri – Angela Vorndran – Semantic Analysis Meeting, 22 January, 2010, Grenoble



EUROPEAN EDUCATIONAL RESEARCH  
**Quality Indicators**

[www.eerqi.eu](http://www.eerqi.eu)



SEVENTH FRAMEWORK  
PROGRAMME



**A Project funded under the  
Socio-Economic Sciences  
and Humanities Theme.**



# Role of metadata in EERQI

- Project proposal names metadata as one component of a multivariate quality indicator
- Correlations between metadata are supposed to give information on document characteristics
- Project workshops have yielded critical opinions on actual usability of metadata for quality assessment

Metadata and Thesauri – Angela Vorndran – Semantic Analysis Meeting, 22 January, 2010, Grenoble



[www.eerqi.eu](http://www.eerqi.eu)



A Project funded under the  
Socio-Economic Sciences  
and Humanities Theme.



# Options for the use of metadata

- General conclusion: metadata are not made for quality detection
- Possible hints at quality: peer review status, persistent identifiers
- Metadata categories to be explored further:
  - title
  - descriptors/keywords
  - peer review
  - publisher/journal
  - web address
  - number of pages

Metadata and Thesauri – Angela Vorndran – Semantic Analysis Meeting, 22 January, 2010, Grenoble



[www.eerqi.eu](http://www.eerqi.eu)



**A Project funded under the  
Socio-Economic Sciences  
and Humanities Theme.**

# Applications of metadata information

- References
  - Citation analysis
  - Co-citation analysis
- Affiliations
  - Cooperation analysis
  - Network analysis
  - Research Centers
- Index terms, title words, keywords
  - Co-word analysis
  - Cluster analysis
- Science-specific text structures
  - Existence of
    - Abstract
    - Bibliography
    - Footnotes

Metadata and Thesauri – Angela Vorndran – Semantic Analysis Meeting, 22 January, 2010, Grenoble



[www.eerqi.eu](http://www.eerqi.eu)



**A Project funded under the  
Socio-Economic Sciences  
and Humanities Theme.**

# Thesauri

Metadata and Thesauri – Angela Vorndran – Semantic Analysis Meeting, 22 January, 2010, Grenoble



[www.eerqi.eu](http://www.eerqi.eu)



**A Project funded under the  
Socio-Economic Sciences  
and Humanities Theme.**



# Thesauri

- **Controlled vocabulary**
  - Terms are selected as descriptors
- **Displays relationships between descriptors**
  - Hierarchical
  - Equivalency
  - Associative
- **Mostly serve for documentation purposes**
- **Describing a predefined topic area as comprehensively as possible**

Metadata and Thesauri – Angela Vorndran – Semantic Analysis Meeting, 22 January, 2010, Grenoble



[www.eerqi.eu](http://www.eerqi.eu)



**A Project funded under the  
Socio-Economic Sciences  
and Humanities Theme.**

# Thesaurus relations

- Hierarchical relations
  - -NT-narrower term
  - -BT-broader term
- Equivalency relations
  - -SY-synonym
- Associative relations
  - -RT-related terms

## special needs education

- .. NT1 curriculum support
- .... NT2 language support
- ..... RT foreign mother tongue student (11)
- .... RT curriculum support class
- .. NT1 inclusive education
- .... RT student integration (01)
- .. NT1 separate special education
- .... RT special school (06)
- .. RT correctional education (02)
- .. RT special educational needs
- .. RT special school (06)

Thesaurus for Education Systems in Europe (TESE)

Metadata and Thesauri – Angela Vorndran – Semantic Analysis Meeting, 22 January, 2010, Grenoble



[www.eerqi.eu](http://www.eerqi.eu)



A Project funded under the  
Socio-Economic Sciences  
and Humanities Theme.

# Thesaurus

## ➤ Descriptors

- Synonyms and quasi-synonyms are subordinated to one descriptor term
- Descriptor can be used as index term for described entities (USE or Use for (UF))

Scope notes (SN)  
explain the meaning  
and use of a descriptor

*adviser*

**USE** guidance officer

*ball game*

MT (09) curriculum

da: boldspil

de: Ballspiel

el: παιχνίδι με μπάλλα

es: juego de pelota

fi: palloilu

fr: jeu de balle

it: gioco della palla

nl: balspelen

pt: Jogo de bola

sv: bollspel

**SN** Includes baseball, golf, tennis, etc..

**UF** baseball

**UF** golf

**UF** tennis

**BT1** sport

European Education Thesaurus (EET)

Metadata and Thesauri – Angela Vorndran – Semantic Analysis Meeting, 22 January, 2010, Grenoble

# Use of Thesauri

- Use controlled terms as indexing vocabulary
  - To avoid the use of different synonyms
    - e.g. andragogy -> adult education
  - To establish links between broader and narrower terms
    - e.g. to differentiate homonyms bank <river>, bank <monetary institution>
  - To establish a structured overview over a topic area

Metadata and Thesauri – Angela Vorndran – Semantic Analysis Meeting, 22 January, 2010, Grenoble



[www.eerqi.eu](http://www.eerqi.eu)



A Project funded under the  
Socio-Economic Sciences  
and Humanities Theme.

# Thesauri in EERQI

- European Education Thesaurus (EET) is used for multilingual search
- Thesaurus terms as query terms can be expanded by their translations and related terms
- Example:  
query „adult education“

- ✓ Descriptor (en): adult education
- ✓ Descriptor (de): Erwachsenenbildung
- ✓ Descriptor (fr): éducation des adultes
- ✓ Descriptor (sv): vuxenutbildning
- ✓ Related term (en): study circle
- ✓ Related term (de): Studienkreis
- ✓ Related term (fr): cercle d'étude
- ✓ Related term (sv): studiecirkel
- ✓ Related term (en): adult education institute
- ✓ Related term (de): Volkshochschule
- ✓ Related term (fr): université populaire
- ✓ Related term (sv): vuxenutbildningsinstitut

Metadata and Thesauri – Angela Vorndran – Semantic Analysis Meeting, 22 January, 2010, Grenoble



EUROPEAN EDUCATIONAL RESEARCH  
**Quality Indicators**

[www.eerqi.eu](http://www.eerqi.eu)



**A Project funded under the  
Socio-Economic Sciences  
and Humanities Theme.**

# Extension to multilingual search

- In context of WP9: „Accommodating Europe´s multilingual environment“
- Thesaurus terms do not cover many possible query terms
- ->Construction of multilingual term networks
- Comprise as many terms as possible to describe a topic area

Metadata and Thesauri – Angela Vorndran – Semantic Analysis Meeting, 22 January, 2010, Grenoble



[www.eerqi.eu](http://www.eerqi.eu)



**A Project funded under the  
Socio-Economic Sciences  
and Humanities Theme.**

# Term networks

## ➤ Sources

- European Education Thesaurus (EET)
  - English, French, German, Swedish
- Thesaurus for Education Systems in Europe (TESE)
  - English, French, German
- IRDP term list
  - French
- DIPF list of controlled vocabulary
  - English, German
- Other Sources
  - Dictionaries, research publications, websites, own ideas

Metadata and Thesauri – Angela Vorndran – Semantic Analysis Meeting, 22 January, 2010, Grenoble



[www.eerqi.eu](http://www.eerqi.eu)



**A Project funded under the  
Socio-Economic Sciences  
and Humanities Theme.**

# Term networks

- Five exemplary topic areas
  - Professional development of teachers
  - Curriculum development
  - Educational effectiveness
  - Intercultural education
  - Higher education

Metadata and Thesauri – Angela Vorndran – Semantic Analysis Meeting, 22 January, 2010, Grenoble



[www.eerqi.eu](http://www.eerqi.eu)



**A Project funded under the  
Socio-Economic Sciences  
and Humanities Theme.**

# Evaluation of term networks

- Usefulness will be tested in comparison to thesaurus expansions after the implementation
  - Are more relevant documents retrieved?
  - Are more documents in other languages retrieved?
  - Which expanded terms are responsible for more relevant hits?

Metadata and Thesauri – Angela Vorndran – Semantic Analysis Meeting, 22 January, 2010, Grenoble



[www.eerqi.eu](http://www.eerqi.eu)



A Project funded under the  
Socio-Economic Sciences  
and Humanities Theme.



# Objective

- In contrast to the aim of a thesaurus to limit a variety of terms to one descriptor, here as many terms as possible of an area shall be collected
- Term networks stand between dictionaries and a thesaurus

Metadata and Thesauri – Angela Vorndran – Semantic Analysis Meeting, 22 January, 2010, Grenoble



EUROPEAN EDUCATIONAL RESEARCH  
**Quality Indicators**

[www.eerqi.eu](http://www.eerqi.eu)



SEVENTH FRAMEWORK  
PROGRAMME



**A Project funded under the  
Socio-Economic Sciences  
and Humanities Theme.**

# Use for semantic analysis

- Metadata carrying semantic qualities can be used for semantic text analysis
  - Title
  - Index terms
  - Abstract
- They contain information about the content of a document

Metadata and Thesauri – Angela Vorndran – Semantic Analysis Meeting, 22 January, 2010, Grenoble



[www.eerqi.eu](http://www.eerqi.eu)



A Project funded under the  
Socio-Economic Sciences  
and Humanities Theme.

# Use for semantic analysis

- Thesauri can provide vocabulary to define the frame for educational sciences concerning the content of publications
- Keywords (e.g. those taken from thesauri) can be indicators of the document's content -> important topics of the text are identified

Metadata and Thesauri – Angela Vorndran – Semantic Analysis Meeting, 22 January, 2010, Grenoble



[www.eerqi.eu](http://www.eerqi.eu)



A Project funded under the  
Socio-Economic Sciences  
and Humanities Theme.

**Thank you very much for  
your attention.**

Angela Vorndran  
vorndran@dipf.de



Metadata and Thesauri – Angela Vorndran – Semantic Analysis Meeting, 22 January, 2010, Grenoble



[www.eerqi.eu](http://www.eerqi.eu)



**A Project funded under the  
Socio-Economic Sciences  
and Humanities Theme.**