
Design Aspects of a Thesaurus for Thematic Search in Geospatial Metadata Services

Katja Korhonen
co-authored with Paula Ahonen-Rainio
Helsinki University of Technology,
Finland

ICC 2007, Moscow, 7.8.2007
SKS:n syysseminaari 11.10.2007

HELSINKI UNIVERSITY OF TECHNOLOGY
DEPARTMENT OF SURVEYING
LABORATORY OF GEOINFORMATION AND POSITIONING TECHNOLOGY



Paikkatietojen asiansanoituksessa käytettävän tesauuksen suunnittelu metatietopalvelun käyttöön

Katja Korhonen
SKS:n syysseminaari 11.10.2007

HELSINKI UNIVERSITY OF TECHNOLOGY
DEPARTMENT OF SURVEYING
LABORATORY OF GEOINFORMATION AND POSITIONING TECHNOLOGY



Outline

- Introduction:
 - requirements of thematic search for the keywords
 - the design project of a thesaurus for the Finnish National Metadata Service
- Aims of the study
- Lessons learned:
 - structuring of the thesaurus
 - use of source materials
 - role and cooperation of experts
- Design process
- Conclusions



Introduction: requirements for the keywords

Search

- to discover potentially suitable datasets
- by keywords (thematic, spatial, temporal)

Keywords have to be commonly understood by data suppliers (as authors of metadata) and by users who search datasets for various purposes

Challenge of rich diversity of keywords in the multi disciplinary context of geographic datasets ⇒ conflicting keywords (synonyms, homonyms, inconsistencies at different levels)



Introduction: the design project

Finnish National Metadata Service

- in Internet since 1994
- now re-engineering to conform to the ISO 19115

Thesaurus for the Metadata Service

- pragmatic approach: **discount engineering**
- ontology based solution would be ideal, but:
in a multi disciplinary context very complex, and more
so when individual disciplines have not their formal
ontologies specified yet



Aims of the study

Design an operational thesaurus

- to be used in an operational Metadata Service
- Hierarchical structure
 - instead of a list to support usability
 - instead of a network because of the simplistic solution of the software
 - relationships added as internal references
- Observations of the design project
 - critical stages of the design
 - role of experts



Lessons learned: structuring of the thesaurus

- You need a structuring of the field as a starting point
 - e.g. thematic fields of INSPIRE directive (2007/2/EC) of the European Commission
 - e.g. topic categories of ISO 19115:2003
 - modifications necessary for national purposes, e.g. in Finland forestry is combined with agriculture instead of economy & our seas are not oceans nor inland waters
- The level of detail has to be appropriate for discovery
 - not too general but not too detailed either!



Lessons learned: use of source materials

- Thesauri (or rather lists of keywords) available for many disciplines
 - the more “scientific” or legally controlled the field (e.g. geology, environment) the more consistent thesaurus or glossaries – but they may be too specific for users without expertise on this field
 - the diversity of the field (e.g. socio-economic activities) complicates the viewpoint
- Classifications and enumerations for the fields for which no thesauri were available
 - use of classifications should not confuse the thematic keyword with the actual data contents

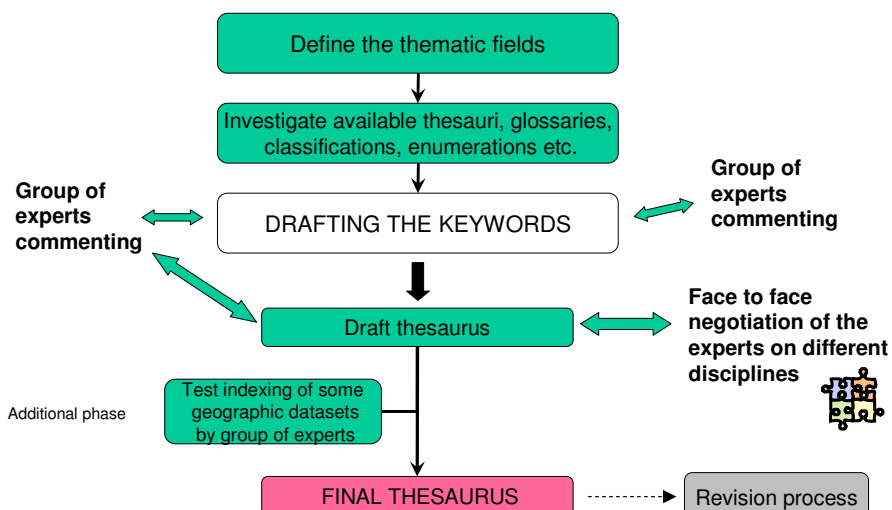


Lessons learned: role of experts

- The draft of a thesaurus has to be commented by experts in several stages of the design process
- The experts have to be chosen so that they cover all the disciplines of the thesaurus
- Face to face commenting is superior to virtual commenting – use of wiki environment failed in our project
- Important additional phase: Test indexing of (some) geographic datasets performed by experts who have good knowledge of these datasets



The design process



Conclusions

- A thesaurus for a national metadata service was designed successfully
- Multi disciplinary nature of GI is the main concern in thesaurus design for extensive metadata services
- Existing thesauri help as a starting point, but not much more
- Experts from various disciplines need to meet to negotiate
- **Revision of the thesaurus needs to be organised – perfect solution cannot be done without further processing**

