

Linked Open Ontology Cloud KOKO – a Decade's worth of Lessons Learned

Matias Frosterus

The National Library of Finland

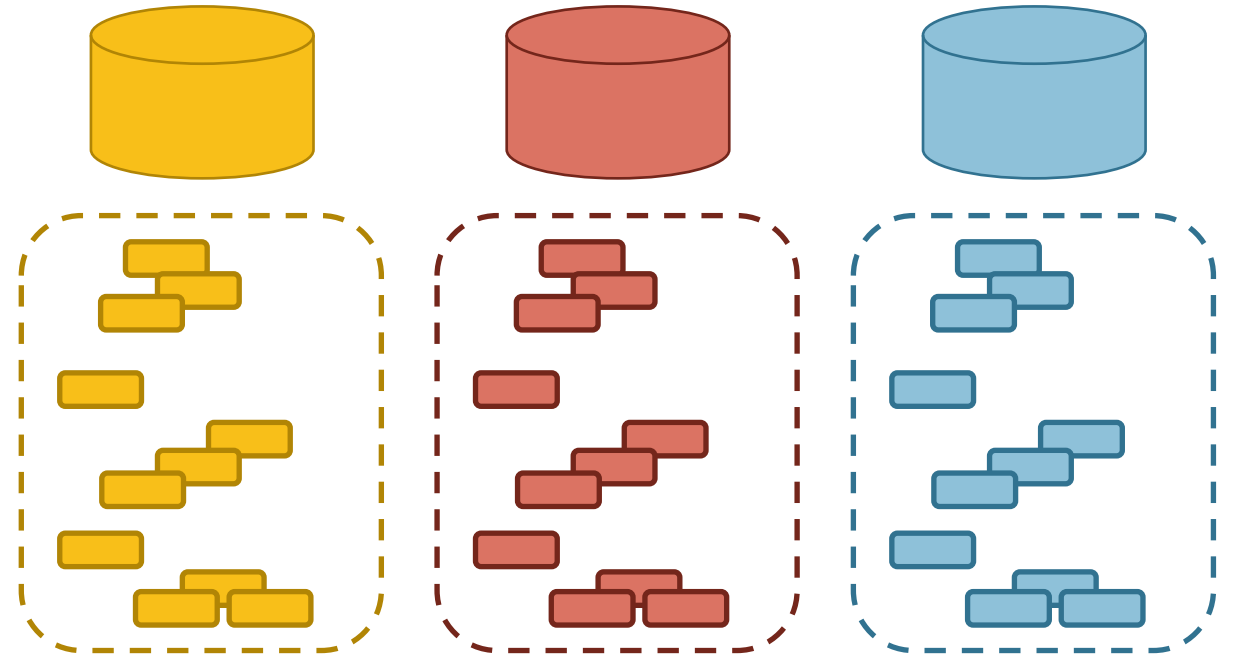
Special thanks to professor Eero
Hyvönen and information systems
specialist Joeli Takala

Outline of the presentation

- The concept of the linked open ontology cloud
- The application to practice
- Lessons learned

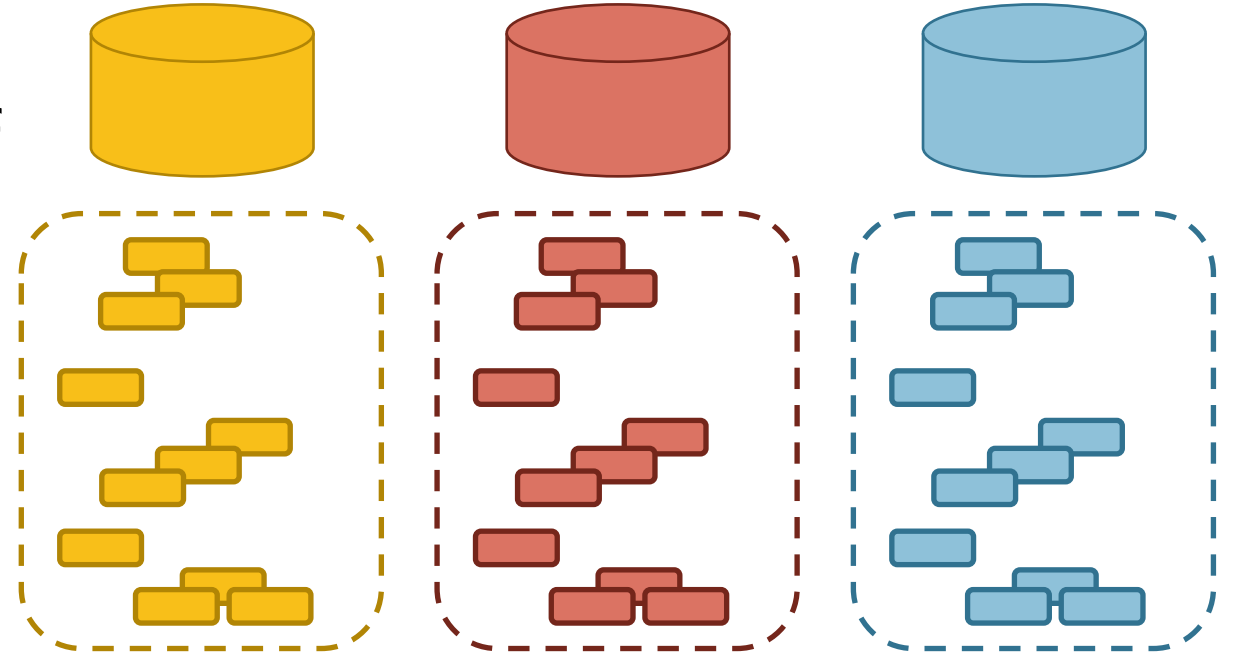
Data annotated using various thesauri

- Controlled vocabularies a mainstay of descriptive metadata
- Various expert organizations maintain their own thesauri
- A lot of high-quality annotations done using these thesauri



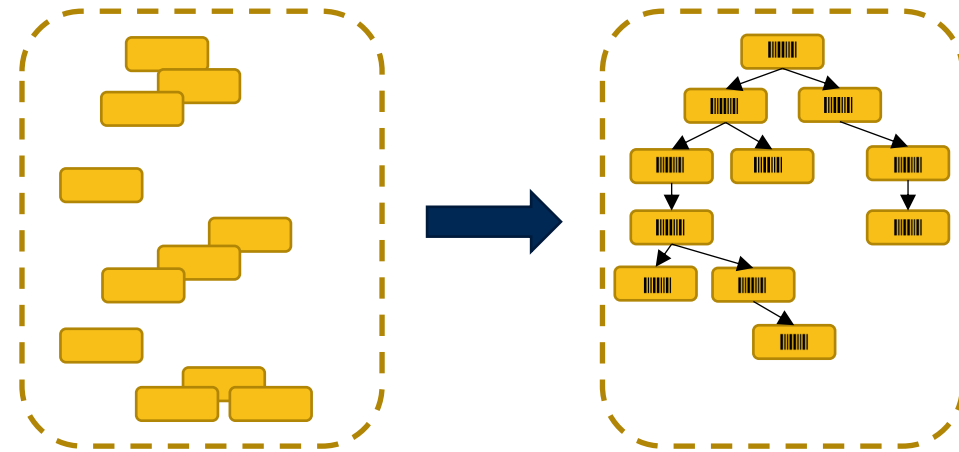
The core question

- How to search all this data simultaneously?
- How to retain the expertise of each organization?
- How to make all of it maintainable?



From thesauri to light-weight ontologies

- SKOS vocabularies
- From terms to concepts
- A complete is-a hierarchy
 - skos:broader but applying it more strictly



From terms to concepts – a note on language

- Interoperability difficult to achieve on the term level
 - A lot easier on the concept level
- Finland has two official languages
 - Finnish and Swedish used for conceptualization
 - Translated to English (and, as of late, in part to Northern Sami language)
- For example
 - linnat@fi, borgar och slott@sv, castles@en, šloahtat"@se

General Upper Ontology (GUO)

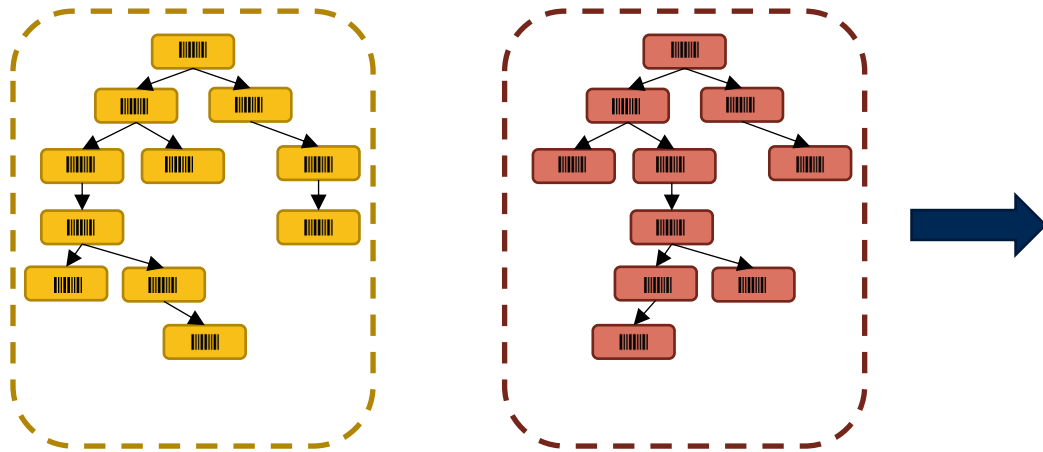
- There is a lot of overlap between the vocabularies
 - Especially if you want complete is-a hierarchies!

General Upper Ontology (GUO)

- There is a lot of overlap between the vocabularies
 - Especially if you want complete is-a hierarchies!
- We can build a General Upper (light-weight) Ontology (or GUO)
- The other ontologies expand on the GUO in their own domains
 - Thus we call them domain ontologies

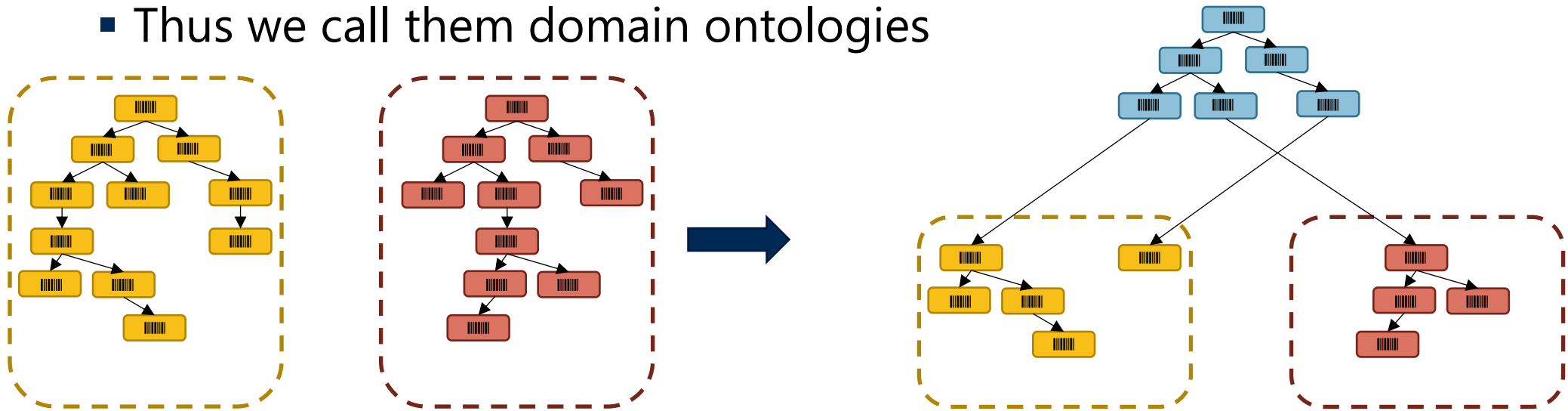
General Upper Ontology (GUO)

- There is a lot of overlap between the vocabularies
 - Especially if you want complete is-a hierarchies!
- We can build a General Upper (light-weight) Ontology (or GUO)
- The other ontologies expand on the GUO in their own domains
 - Thus we call them domain ontologies



General Upper Ontology (GUO)

- There is a lot of overlap between the vocabularies
 - Especially if you want complete is-a hierarchies!
- We can build a General Upper (light-weight) Ontology (or GUO)
- The other ontologies expand on the GUO in their own domains
 - Thus we call them domain ontologies



Application to practice

Finto thesaurus and ontology service

■ Finto

- <https://finto.fi/en/>
- A publication platform for SKOS vocabularies
- An ecosystem surrounding said vocabularies
- For the whole public sector in Finland

■ Skosmos

- <https://skosmos.org/>
- Open source software for setting up a SKOS vocabulary service
- Browsing interface as well as REST APIs
- A thriving community of users around the world



YSO - General Finnish ontology

Content language English Search

- A-Z
- Hierarchy
- Groups
- New and Deprecated
- amusement parks
- animal parks
- aquaculture ponds
- artists' homes
- battle sites
- bazaars (markets)
- boat ramps
- breakwaters
- bridges
- brothels
- buildings
 - abandoned buildings
 - abandoned houses
 - Academy House (Turku)
 - Aleksis Kivi Memorial Cottage
 - Allinna
 - Andersudde
 - animal housing
 - apartment blocks
 - Arktikum
 - Arppeanum
 - artists' houses
 - Aschan House (Helsinki)

objects > place > places defined by humans > buildings > abandoned buildings

PREFERRED TERM **abandoned buildings**

TYPE General concept

BROADER CONCEPT buildings

NARROWER CONCEPTS abandoned houses

RELATED CONCEPTS abandonment (depopulation)
urban exploration

BELONGS TO GROUP 37 Construction. Building Industry. Housing Construction. Earth Construction. Hydraulic Engineering. Road Construction

IN OTHER LANGUAGES hylätyt rakennukset Finnish
övergivna byggnader Swedish
ödebyggnader

URI <http://www.yso.fi/onto/yso/p28916>

DOWNLOAD THIS CONCEPT: RDF/XML TURTLE JSON-LD Created 5/19/16, last modified 5/10/17

GUO and domain ontologies

- General Finnish Thesaurus YSA and its Swedish-language counterpart Allärs were combined to the General Finnish Ontology YSO

Linked open ontology cloud KOKO

- Approximately 60,000 concepts
- GUO + 13 Domain ontologies

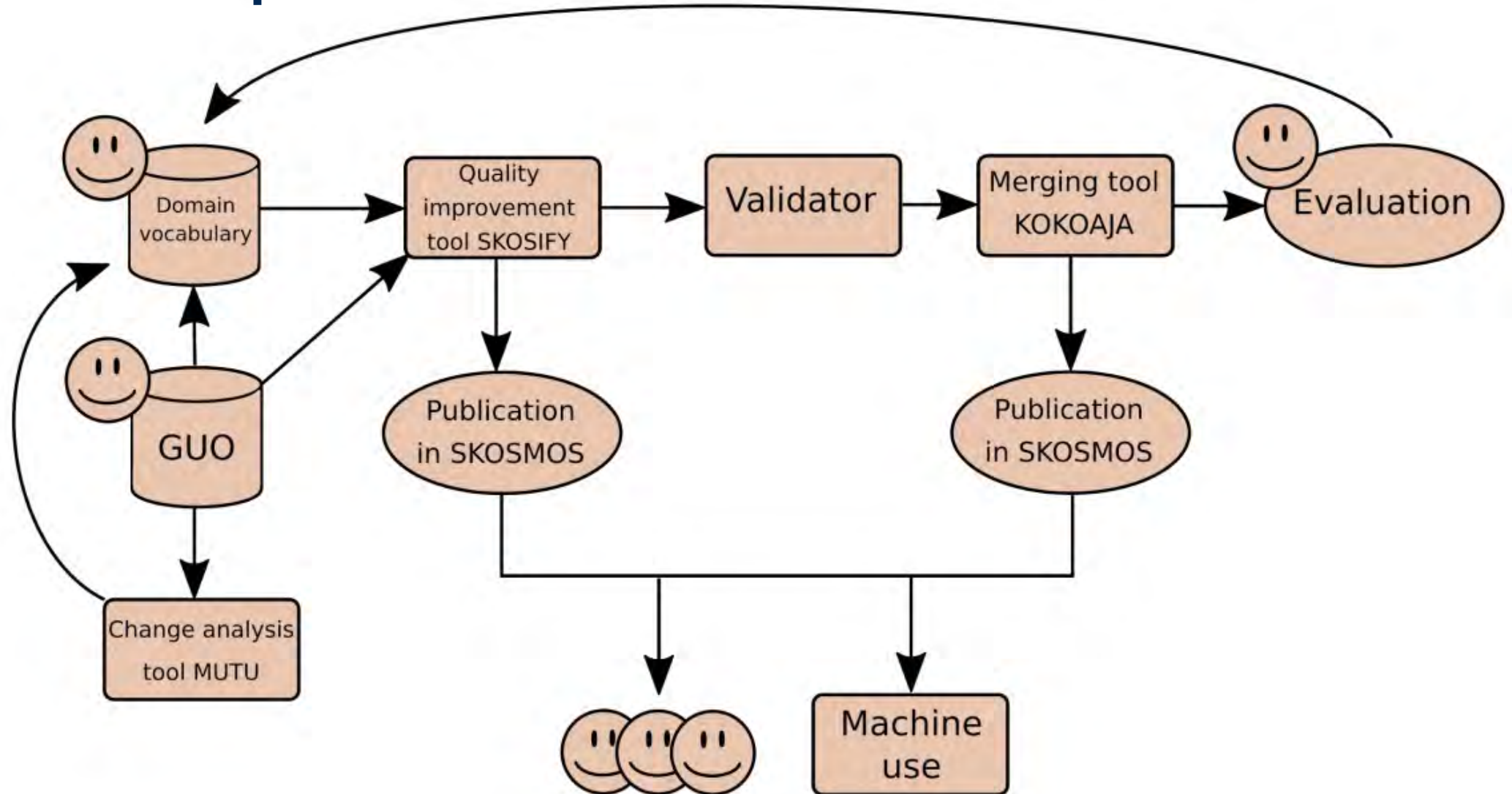
KOKO

- AFO - Natural resource and environment ontology
- JUHO - Finnish Ontology for Public Administration
- JUPO - Finnish Ontology for Public Administration Services
- KAUNO - Ontology for fiction
- KITO - Ontology of Literary Research
- KTO - Ontology for Linguistics
- LIIKO - Ontology of Transportation
- MAO/TAO - Ontology for Museum Domain and Applied Arts
- OMA - Ontology for Media Art
- SOTO - Ontology for Military Science
- TERO - Finnish Ontology of Health and Welfare
- TSR ontology
- VALO - The Finnish Ontology of Photography
- YSO - General Finnish ontology

Expert group of ontology developers

- The original expert organizations maintain the new domain ontologies
- A group that decides the principles behind the development of ontologies
- Ranging from technical details of deprecation to general guidelines on what to do with archaic offensive altLabels.

Publication process



Lessons learned

Over ten years' worth of experience

Finding balance

- Based on a modular set of tools to solve each subproblem
- Ongoing iterative development process
- Balance between manual labour, automated processes and proactive guidelines

Synchronizing changes

- Propagating changes is complex
- Changes in GUO
 - A new “frozen” version of GUO published twice a year and a new version of KOKO published quarterly
 - We developed a tool to help domain ontology developers
 - Differing resources and development cycles in different organizations
- Changes in domain ontologies that might affect other domain ontologies
 - More difficult to catch

Concepts and PIDs are forever, organizations less so

- Domain ontologies come and go
- New ones getting added is relatively straightforward
- Old ones leaving is challenging
 - Two choices
 - Kept in stasis inside KOKO – slowly deteriorates due to changes in GUO
 - Merging into the GUO or another domain ontology
 - Deprecation for singular concepts can be generalized for an entire vocabulary but poses challenges

Semantics are hard

- Semantic problems are difficult to find (let alone fix!) using automation
- In theory, there should be no overlap between the concepts in different domain ontologies that are not present GUO
 - In practice this is very difficult to ascertain
 - Domain ontologies each have distinct domains
 - Made to answer a specific annotation need as opposed to try to model reality as such
 - If there is overlap, ideally it should enter GUO
- Same prefLabels are automatically differentiated using suitable signifiers in parentheses

A success?

- KOKO has been in active use for over ten years
 - Especially useful for organizations that deal with "everything" such as museums and media companies
- Far from perfect
 - But maybe good enough?
 - Still improving!





www.kansalliskirjasto.fi

Matias Frosterus
matias.frosterus@helsinki.fi