GACS
Quality Improvement Project

Lori Finch, Sujata Suri, Tom Baker, Osma Suominen

GACS teleconference

5 December 2017

http://bit.ly/gacs_qip_report

GACS Quality Improvement Project (QIP)

- **May 2016** GACS Beta 3.1 launched at Open Harvest, Chania

- **June 2016** "GACS Phase 4" workplan lists improvements needed and proposes a stakeholder survey about hierarchy

- **November 2016** GACS meeting in Goettingen discusses survey results and GACS structure

- **April 2017** GACS QIP starts, Lori and Sujata tasked with improving GACS structure with support from Osma and Tom

Discussion of hierarchy in Goettingen

- Stakeholders surveyed about three scenarios
  - A. Focused top level with few concepts (like Finnish YSO)
  - B. Type-based top level, more extensive (like AGROVOC)
  - C. Thematic-based top level (like CAB Classified Thesaurus)
- Survey results ambiguous, with some conclusions
  - In principle, scenarios "all valid" (but maintainable?)
  - Hierarchy is needed, but should be cleaned up
  - "Enough hierarchy to help disambiguate or navigate" (but no more?)
  - Thematic groups should be kept as an additional view

GACS QIP process

- GACS Beta 3.1 in Vocbench frozen and exported for structural editing

- Changes made:
  - Polyhierarchy reduced
  - BT/NT checked for typological consistency
  - Top concepts pruned and relocated
  - Still-unassigned concepts assigned to thematic groups

Top concepts moved to the middle

- GACS Beta 3.1 had 578 top concepts
- QIP moved concepts from top to middle, such as:
  - plutonium BT transition elements
  - adenosine BT nucleosides
  - phenylketanuria BT metabolic disorders
  - life expectancy BT biological properties

Problems in hierarchy fixed

Before: people => groups => zoos, museums…

After

- people
  - athletes
  - food service workers
- groups
  - collections
    - animal collections
      - insect collection
      - **zoos**
    - archives
    - germplasm collections
    - museums
    - plant collections
      - botanical gardens
      - herbaria
      - type collections

- people
  - agricultural economists
  - athletes
  - audiences
  - consumers
  - farmers
    - growers
  - food service workers
  - immigrants
  - injecting drug users
  - landowners
  - migrants
  - patients
  - sexual partners
  - students

Problems in hierarchy fixed

Before, "structures" is ambiguous:
- **Socioeconomic structures**
- **Architectural structures**

After:
- **Inconsistent meanings are separated**
- **Broader hierarchy reduced from six to one**

occupations > engineering > agricultural engineering > agricultural structure
sciences > engineering > agricultural engineering > agricultural structure
occupations > engineering > manmade structures > agricultural structure
sciences > engineering > manmade structures > agricultural structure
structures > manmade structures > agricultural structure
analysis > economic analysis > economic structure > agricultural structure

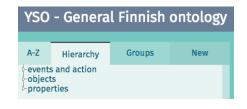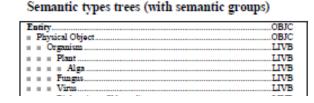| PREFERRED TERM	agricultural structure
TYPE	Topic
BROADER CONCEPT	agricultural engineering economic structure manmade structures
NARROWER CONCEPTS	abattoirs agrarian structure animal housing dairies farm buildings farm structure grain storage facilities

OBJECTS > ABSTRACT OBJECTS > SOCIETAL ABSTRACT OBJECTS > economic structure > agricultural structure

PREFERRED TERM	agricultural structure
TYPE	http://id.agrisemantics.org/vocab#Topic
DEFINITION	ⓘ A concept that can be freely defined as characteristics and patterns that describe the participants in agricultural production, including farm businesses, farm operators, and farm households. Structural indicators that are used in analyses such as tracking changes in farms and farming over time, and assessing the impact of changes in farm programs and policies. Source: www.ers.usda.gov
BROADER CONCEPT	economic structure
NARROWER CONCEPTS	agrarian structure farm structure production structure

... > PHYSICAL OBJECTS > MANMADE OBJECTS > FACILITIES AND STRUCTURES > factories > abattoirs

PREFERRED TERM	abattoirs
TYPE	http://id.agrisemantics.org/vocab#Topic
BROADER CONCEPT	factories
RELATED CONCEPTS	meat byproducts meat hygiene meat inspection meat production

Status of upper hierarchy

- Post-QIP, concepts clustered nicely under 3 top concepts
  - **Objects**: material things that can be seen or touched or visited (locations), including "abstract objects" such as ideas and models.
  - **Events and actions**: things that happen, such as processes or phenomena.
  - **Properties**: attributes, characteristics, or qualities of things.
- Lower levels adapt YSO and Unified Medical Language System (UMLS)

  - 100% of concepts aligned
  - Towards cleaner hierarchies that are BT/NT transitive ("work in progress")
  - First-draft scope notes for upper structure

Overly specific concepts weeded

- Moved to separate concept scheme
  - Concepts "not of global interest" in GACS due to mapping process
    - 4-H Youth Development Program
    - Food and Drug Administration
  - Concepts that "need more work" (or ambiguous)
    - Practice
  - For now just 10 (proof of concept), but more could be "weeded"
- ***Weeds cannot be eradicated, only transplanted!***
  - ***The Iron Principle of URI Persistence!***

GACS Thematic Classification

GACS Thematic Classification

- Before QIP
  - Based on CAB Classification and UAT
  - Circa 2,800 concepts unclassified
  - Frequency distribution was uneven
- After QIP
  - 100% of concepts under at least one category
  - Principle: Concept moved out of "common terms" if better found.
  - Principle: Concepts can be under more than one category.
  - First-draft Scope Notes
    - Categories list examples of concepts assigned
    - Underused concepts flagged for possible deletion

Classification categories adjusted

Before: **NY** category had just two concepts

After: Concepts moved to **NQ fungi**

NE organisms, taxonomically arranged
- NH prions
- NK viruses
- NN prokaryotes
- NQ fungi
- NT plants
- NW animals
- NY organisms of uncertain taxonomy
  - Pneumocystis
  - Pneumocystis carinii
- NZ organisms by alternative taxonomic

Pneumocystis carinii

PREFERRED TERM	Pneumocystis carinii
TYPE	http://id.agrisemantics.org/vocab#Organism
BROADER CONCEPT	Pneumocystis
ENTRY TERMS	*Pneumocystis carini*
SCOPE NOTE	Previously indexed as a protozoan but it is now classi…
BELONGS TO GROUP	fungi

Pneumocystis

PREFERRED TERM	
TYPE	http://id.agrisemantics.org/vocab#Organism
BROADER CONCEPT	coccidia
	Pneumocystidaceae
NARROWER CONCEPTS	Pneumocystis carinii
BELONGS TO GROUP	fungi

Classification categories underused

- Flagged for consolidation or deletion
  - NY organisms of uncertain taxonomy
  - SW vermiculture
  - WQ arts
  - WS history
  - WY philosophy and ethics
  - LU evolution
  - LX natural history

Concepts removed from "common terms"

PREFERRED TERM	**insulating materials**
TYPE	Topic
BROADER CONCEPT	materials
NARROWER CONCEPTS	asbestos polystyrenes vermiculite
RELATED CONCEPTS	construction materials energy conservation heat transfer rockwool
ALTERNATIVE LABEL	*insulation*
BELONGS TO GROUP	common terms materials

PREFERRED TERM	**insulating materials**
TYPE	http://id.agrisemantics.org/vocab#Topic
BROADER CONCEPT	materials
NARROWER CONCEPTS	asbestos polystyrenes vermiculite
RELATED CONCEPTS	construction materials energy conservation heat transfer rockwool
ENTRY TERMS	*insulation*
BELONGS TO GROUP	materials

Concepts assignable to multiple categories

PREFERRED TERM	**milking machines**
TYPE	http://id.agrisemantics.org/vocab#Topic
BROADER CONCEPT	milking equipment
	milking
RELATED CONCEPTS	milking parlours
ENTRY TERMS	*machine milking*
BELONGS TO GROUP	animal production
	equipment

PREFERRED TERM	**plant protection equipment**
TYPE	http://id.agrisemantics.org/vocab#Topic
BROADER CONCEPT	equipment
NARROWER CONCEPTS	weed control equipment
BELONGS TO GROUP	equipment
	health protection

Mappings to external vocabularies

540 concepts of type Geographical were mapped to Wikidata

CLOSELY MATCHING CONCEPTS	Alberta	www.wikidata.org
EXACTLY MATCHING CONCEPTS	Alberta	AGROVOC
	Alberta	CAB Thesaurus
	Alberta	NAL Thesaurus

Guidelines for GACS – a Github website

https://agrisemantics.github.io/gacs-qip/

Guidelines for GACS – a Github website

⌂ Guidelines for GACS

Search docs

About these guidelines

About GACS

Concepts

Concept relations

Defining concepts with semantic relations

Broader, narrower, and related concepts

Top-level concepts

Concept types

Custom (non-SKOS) relations between concepts

Mapping to concepts in external vocabularies

Concept definitions

Concept labels

Scientific and common names

Docs » Concept relations

⌗ Edit on Github

Defining concepts with semantic relations

A GACS concept is formally defined by its "semantic neighborhood", a web of associative and hierarchical relations with other concepts in GACS. The concept *alfalfa*, for example, is defined by relations such as:

- *alfalfa* `hasType` *Product*
- *alfalfa* `BT` *legumes*
- *alfalfa* `RT` *fodder plants*
- *fodder plants* `RT` *livestock feeding*
- *fodder plants* `RT` *forage*
- *fodder plants* `RT` *hay*

These relations are reflected in the GACS definition of *alfalfa* as: "A valuable *leguminous crop* for *forage* or *hay* used in *livestock feeding*". However, less than 20% of the concepts in GACS have definitions, and crafting definitions for the other 80% would be expensive, especially across translations in multiple languages. Moreover, narrative definitions are less useful than semantic relations for automated processing because relations between concepts can only roughly be computed from label strings, sentence parsing, and the co-occurrence of words.

Next steps

- Hierarchy
  - Further adjustments, especially in crucial middle levels
  - A solution for "part of" relations?
- Thematic classification
  - Consolidate or drop underused categories
  - Assign more concepts as needed (frequencies are still uneven)
  - Fill out scope notes
- Concept definitions (less than 20% of GACS)
  - Some added (43), possibly more needed
- Scope notes
  - Some added (8)
  - AGROVOC-, CABT-, and NALT-specific notes, out of context, could be consolidated or dropped
- More concepts to be weeded from GACS and "transplanted"

Links

- GACS QIP development – backend Github repo
  - https://github.com/agrisemantics/gacs-qip
  - https://github.com/agrisemantics/gacs-qip/tree/master/gacs-data - data
  - https://github.com/agrisemantics/gacs-qip/blob/master/gacs-data/webstudio-exports/NOTES.md - things changed
- GACS QIP development – Skosmos website
  - http://tester-os-kktest.lib.helsinki.fi/gacsdemo/gacs/en/
- Guidelines for GACS – Github website
  - https://agrisemantics.github.io/gacs-qip/
- GACS 4.0 – production website at Tor Vergata
  - http://browser.agrisemantics.org/gacs/en/ - to replace GACS Beta 3.1 (soon?)