

GACS Quality Improvement Project

Lori Finch, Sujata Suri, Tom Baker, Osmo 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 Osmo 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

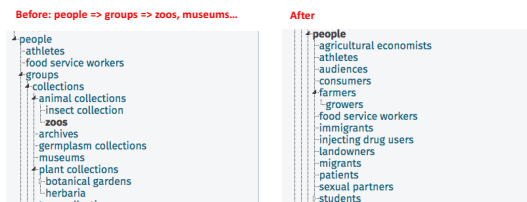
GACS QIP process

- 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



Problems in hierarchy fixed

Before, "structures" is ambiguous:

- Socioeconomic structures
- Architectural structures

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

BROADER CONCEPT

NARROWER CONCEPTS

Topic

agricultural engineering

economic structure

manmade structures

abattoirs

agrarian structure

animal housing

farm buildings

farm structure

grain storage facilities

PREFERRED TERM **agricultural structure**

TYPE

DEFINITION

(2) A concept that can be freely defined as characteristics and patterns that describe the participants in agriculture production, including farm buildings, farm equipment, 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

NARROWER CONCEPTS

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

PREFERRED TERM **abattoirs**

TYPE

BROADER CONCEPT

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)



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!*

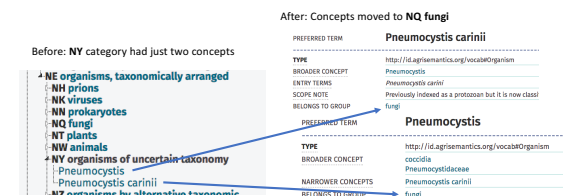
GACS Thematic Classification

Alphabetical	Hierarchy	Groups
CA GENERAL FA PHYSICAL SCIENCES JA EARTH SCIENCES LA LIFE SCIENCES PA APPLIED SCIENCE AND TECHNOLOGY WA SOCIAL SCIENCES AND HUMANITIES		

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



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

Concepts assignable to multiple categories

PREFERRED TERM	milking machines	PREFERRED TERM	plant protection equipment
TYPE	http://id.agrisemantics.org/vocab#topic	TYPE	http://id.agrisemantics.org/vocab#topic
BROADER CONCEPT	milking equipment	BROADER CONCEPT	equipment
NARROWER CONCEPTS	milking milking parlours	NARROWER CONCEPTS	weed control equipment equipment health protection
RELATED CONCEPTS	machine milking		
ENTRY TERMS	animal production equipment		
BELONGS TO GROUP			

Mappings to external vocabularies

540 concepts of type Geographical were mapped to Wikidata

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

Guidelines for GACS – a Github website

Guidelines for GACS

[Docs](#) • [About these guidelines](#) [Edit on Github](#)

About these guidelines

- About GACS
- Concepts
- Concept relations
- Concept definitions
- Concept labels
- Scientific and common names
- Thematic classification
- Scope notes

This website documents the principles and guidelines used to maintain the Global Agricultural Concept Scheme (GACS). It aims at providing enough explanation and historical context to convey the rationale for the design principles underlying GACS. It is intended to serve as a living document that will be amended as GACS evolves.

GACS is expressed as a SKOS concept scheme, so readers will benefit from a basic understanding of [Simple Knowledge Organization Scheme \(SKOS\)](#).

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

Guidelines for GACS – a Github website

Guidelines for GACS

[Docs](#) • [Concept relations](#) [Edit on Github](#)

About these guidelines

- About GACS
- Concepts
- Concept relations
- Concept definitions
- Concept labels
- Scientific and common names

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* **is a** *Product*
- *alfalfa* **is** *legumes*
- *alfalfa* **is** *fodder plants*
- *fodder plants* **is** *livestock feeding*
- *fodder plants* **is** *forage*
- *fodder plants* **is** *hay*

These relations are reflected in the GACS definition of *alfalfa* as: “A valuable leguminous crop for foraging 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 QJP development – backend Github repo
 - <https://github.com/agrisemantics/gacs-qjp>
 - <https://github.com/agrisemantics/gacs-qjp/tree/master/gacs-data> - data
 - <https://github.com/agrisemantics/gacs-qjp/blob/master/gacs-data/webstudio-exports/NOTES.md> - things changed
- GACS QJP development – Skosmos website
 - <http://tester-os-kktest.lib.helsinki.fi/gacsdemo/gacs/en/>
- Guidelines for GACS – Github website
 - <https://agrisemantics.github.io/gacs-qjp/>
- GACS 4.0 – production website at Tor Vergata
 - <http://browser.agrisemantics.org/gacs/en/> - to replace GACS Beta 3.1 (soon?)