

GACS Quality Improvement Project

Summer 2017

Lori Finch Lori.Finch@ars.usda.gov

Sujata Suri Sujata.Suri@ars.usda.gov

Tom Baker Tom.Baker@tombaker.org

Osma Suominen osma.suominen@helsinki.fi

GACS Beta 3.1 – motivation for QIP?

- GACS Beta 3.1 has not appreciably changed since May 2016.
- List of Quality Improvements
 - “GACS Phase 4: from soft launch to Version 1.0” document dated June 20, 2016.
 - Tasks not done
- NAL resources available
- GACS survey, autumn 2016

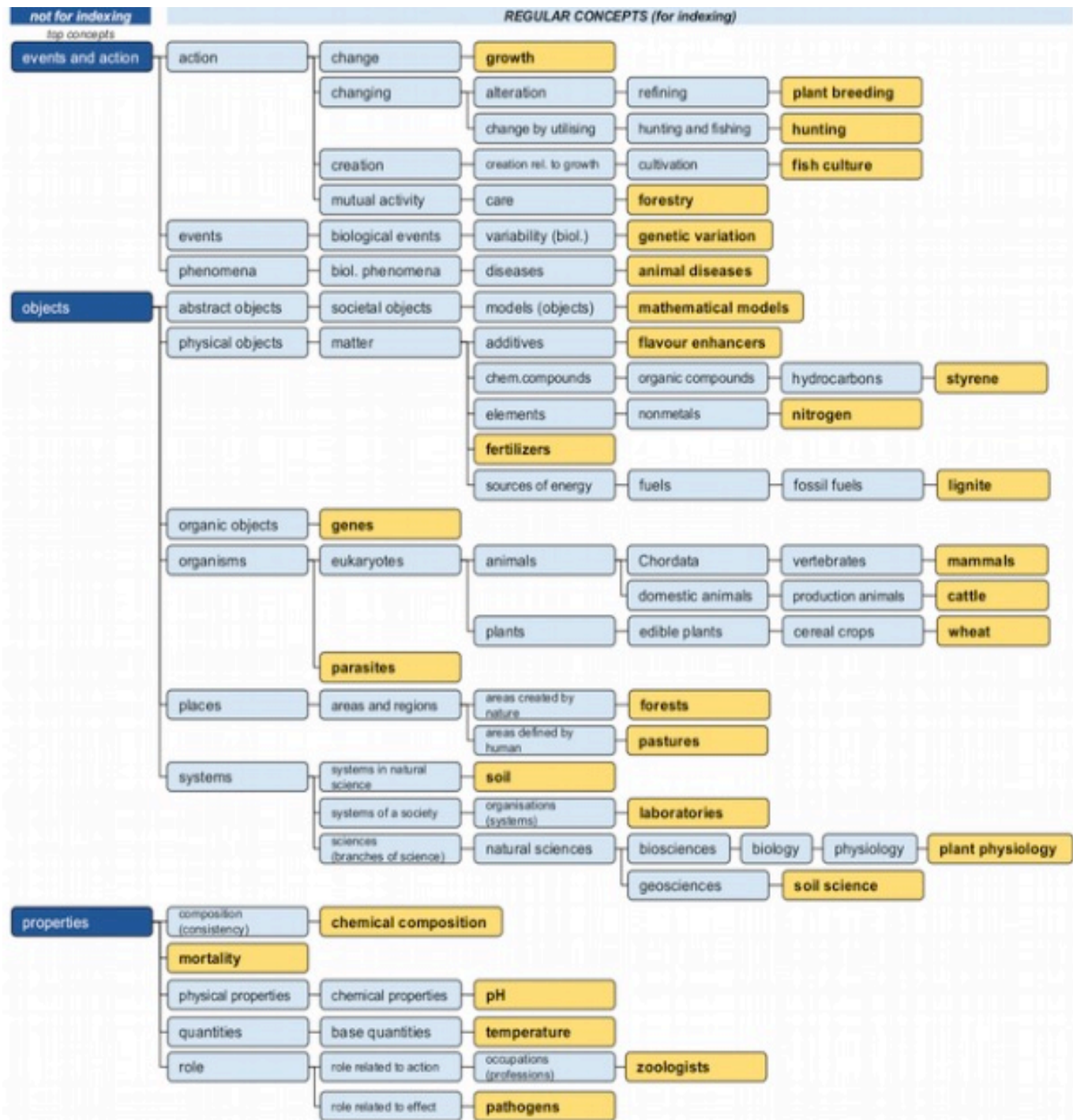
GACS survey – Autumn 2016

“We wanted to hear thoughts about how GACS should be **structured** so that it best serves the **needs of users** as well as **applications** that make use of GACS.”

https://github.com/agrisemantics/2016_11_goettingen/blob/master/presentations/2016-11-23.gacs_survey_structure.pdf

- Three hierarchical structures proposed

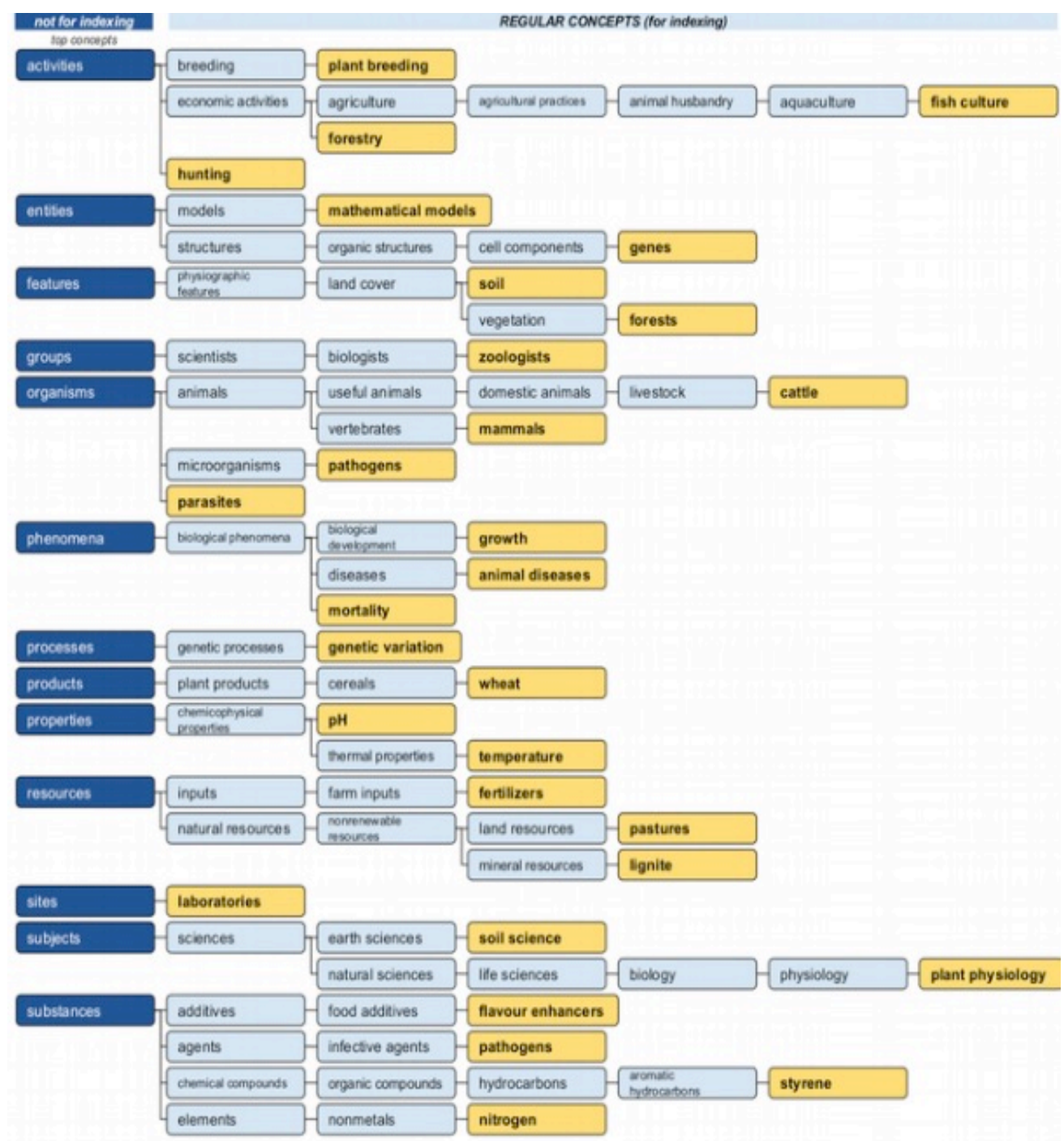
Scenario A – based on YSO and DOLCE



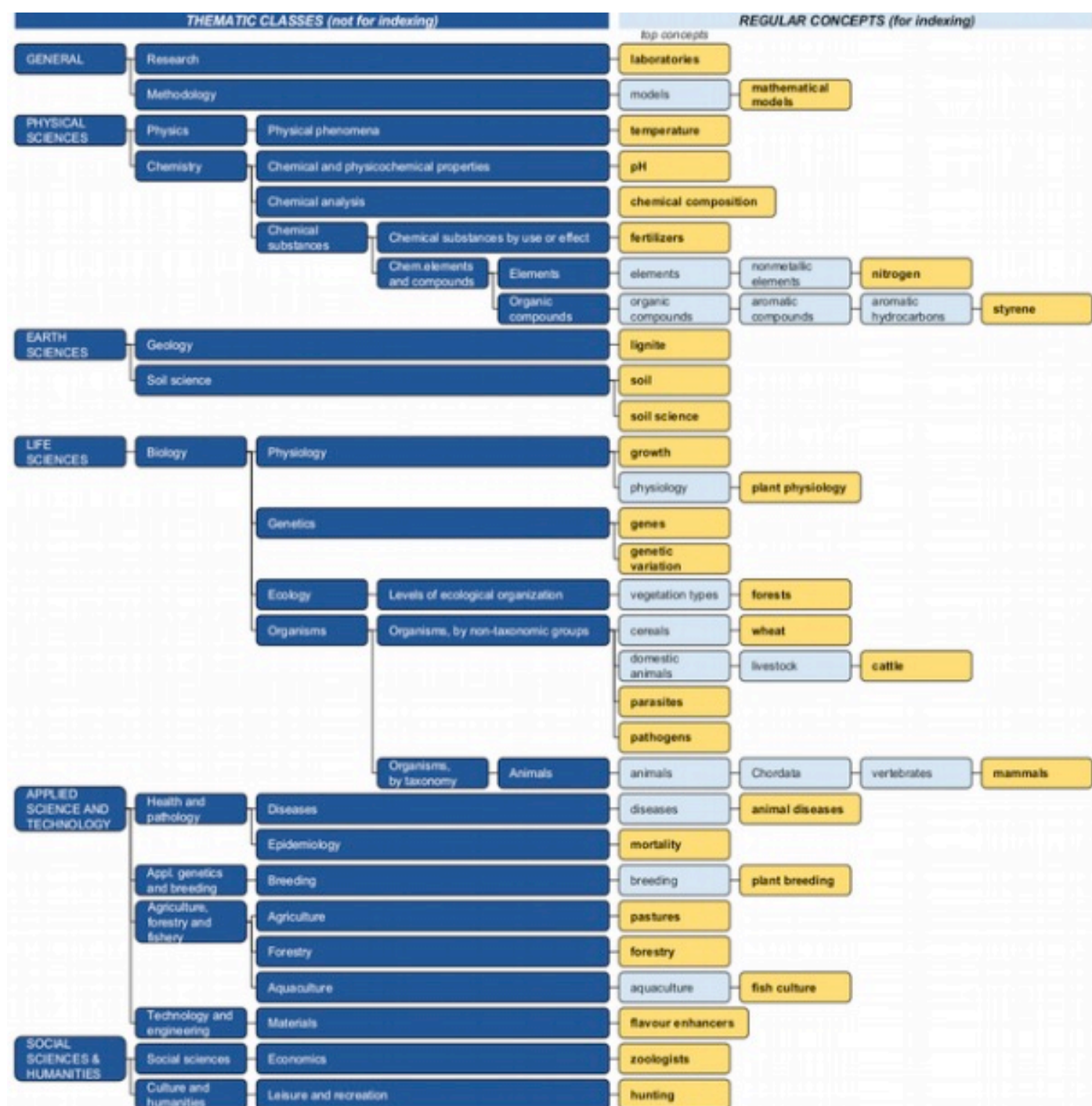
Scenario B

– based on

AGROVOC



Scenario C – thematic based on CABT classified



GACS survey: Conclusions and recommendations (Osma)

- 1) **We should keep the thematic groups as an additional view**, with possibly some tweaks
 - allow multiple groups for a concept, with guidance
 - **classify the remaining 20% of concepts** which currently are not
- 2) We should consider adopting the current concept types (Organism, Chemical, Geographical, Product, Topic) as top concepts
 - **maybe split off Property** (and others?) from Topic
 - **maybe create an additional layer of organization below these top concepts, especially for Topics which are quite many**
 - could mean dropping the notion of concept types, to avoid encoding the same information in two different ways
- 3) **In any case we need to continue cleaning up the hierarchy**
 - **reduce unwarranted polyhierarchy**, for example by calculating scores for each BT/NT relationships based on metrics (e.g. shared among source thesauri?) and removing the worst ones
 - **flag and correct situations where BT and NT have different concept types** (~1300)
 - **It would help to move some leaf concepts out of GACS Core**

https://github.com/agrisemantics/2016_11_goettingen/blob/master/presentations/2016-11-23.gacs_survey_structure.pdf

GACS QIP

- Resources
 - Lori Finch and Sujata Suri
 - Osma Suominen, GODAN funded
 - Tom Baker, GODAN funded
- Main goal:
 - Assure thematic assigned for 100% of terms
 - Cleanup hierarchy
 - Remove polyhierarchy
 - BT/NT of same type
 - Split off property, etc.

GACS Core Beta 3.1 hierarchy stats

- 15,428 concepts
- 20,072 BT/NT relationships (1.3 per concept)
- 4,226 concepts (27%) have more than one BT
- 578 top concepts

Beginning principles

- Starting with 15,428 concepts
- No deletion of concepts
- Working only in English
- Changes to SKOS file for import into Luxid Web Studio <https://github.com/agrisemantics/gacs-qip>
- Time limits: Working June, July, Aug (not Osma and Tom) and Sept

Hierarchy: “Finding homes for top terms” - moving to lower parts of hierarchy

Showing how hierarchy is
“before” – 578 top terms
in Web Studio,

terms with “>” have
narrower terms

4-H Youth Development Program
access
acclimatization
> accounting
action plans
> activity
> administration
> Africa
age differences
agricultural economists
agricultural production
> agriculture
> alkalis
allergic reactions
Alps
> Americas
aminolaevulinic acid
analogues
> analysis
animal growth forms
animal needs
> animal welfare
Antarctica
anthelmintic properties
> antituberculous agents
> application
> application rates
> applications
arctic regions
> Asia
aspiration
assortments
> atmospheric deposition
> atmospheric sciences
attenuation
audiences
> Balkans
bankruptcy
barriers
> basic needs

> people
> performance traits
> personality
phenylketonuria
photographs
> physical education
> physical states
> physiographic features
> physiological functions
plant animal relations
> plant condition
plant growth forms
plant-insect relations
> plant parts
plutonium
political power
population pressure
> postmortem changes
practice
> precursors
predisposition
pre-eclampsia
prescriptions
> prevention
primary production

Examples of terms removed as “top term” and given BT

- plutonium BT transition elements
- adenosine BT nucleosides
- phenylketonuria BT metabolic disorders
- life expectancy BT biological properties
- endocrine-disrupting chemicals BT chemical substances

Hierarchy problems

- Example: “reproduction control” with BT “Brachyspiraceae”
- Example: “people” has narrower terms including-
 - Zoos
 - Botanical gardens
 - Seedless cultivars

Alphabetical

Hierarchy

Groups

- aerobes
- anaerobes
- cellulolytic microorganisms
- denitrifying microorganisms
- genetically engineered microorganisms
- heterotrophic microorganisms
- inoculum
- keratinophilic microorganisms
- ligninolytic microorganisms
- Mesophile microorganisms
- methanotrophs
- microbial flora
- prokaryotes
 - Archaea
 - bacteria**
 - acetic acid bacteria
 - Acetobacter
 - animal pathogenic bacteria
 - coliform bacteria
 - cyanobacteria
 - Cytophaga-Flexibacter-Bacteroides group
 - denitrifying bacteria
 - entomopathogenic bacteria
 - Erysipelothrix
 - Fusobacterium
 - Gracilicutes
 - Bacteroidetes (phylum)
 - Chlamydiae
 - Proteobacteria
 - Spirochaetes
 - Spirochaetales
 - Brachyspiraceae
 - Brachyspira (Bacteria)
 - reproduction control**
 - abortion
 - artificial insemination
 - contraceptives
 - embryo transfer
 - family planning
 - induced ovulation
 - synchronization
 - Leptospiraceae
 - Spirochaetaceae
 - gram negative bacteria
 - gram positive bacteria
 - lactic acid bacteria

... > Gracilicutes > Spirochaetes > Spirochaetales > Brachyspiraceae > reproduction control

PREFERRED TERM	• reproduction control	
<hr/>		
TYPE	Topic	
BROADER CONCEPT	• Brachyspiraceae	
NARROWER CONCEPTS	abortion artificial insemination contraceptives embryo transfer family planning induced ovulation synchronization	
ALTERNATIVE LABEL	contraception in animals reproductive control	
IN OTHER LANGUAGES	ضبط التكاثر	Arabic
	繁殖控制	Chinese
	家畜繁育控制	
	kontrola reprodukce	Czech
	kontrola rozmnožování	
	zabránění březosti	
	Maîtrise de la reproduction	French
	Contraception (animaux)	
	Contrôle reproduction (animaux)	
	Fortpflanzungskontrolle	German
	Kontrazeption (tier)	
	प्रजनन नियंत्रण	Hindi
	गर्भनिरोधण (पशुओं में)	
	reprodukció szabályozása	Hungarian
	fogamzásgátlás (állat)	
	Controllo della riproduzione	Italian
	Contraccezione (animali)	
	繁殖調節	Japanese
	動物の繁殖抑制	
	ການຄວບຄຸມການແພ່ພັນ	Lao
	ການຄຸມກຳເນີດໃນສັດ	

“People” hierarchy at GACS Beta

narrower terms are nonsensical:

Zoos

Museums

Botanical gardens

Herbaria

Breeds

Colored varieties

Seedless varieties

GACS Beta

Alphabetical

Hierarchy

Groups

- people •
 - athletes
 - food service workers
 - groups
 - collections
 - animal collections
 - insect collection
 - zoos •
 - archives
 - germplasm collections
 - museums •
 - plant collections
 - botanical gardens •
 - herbaria •
 - type collections
 - lectotypes
 - committees
 - consumers
 - human communities
 - interest groups
 - landowners
 - population
 - risk groups
 - social groups
 - stakeholders
 - students
 - taxa
 - biotypes
 - breeds •
 - genetic lines
 - races
 - species
 - strains
 - subspecies
 - varieties
 - color varieties (en-us)
 - colored varieties •
 - cultivars
 - introduced varieties
 - landraces
 - seedless varieties •
 - vegetarians
 - tourists
 - volunteers
 - women
 - youth

Mixed hierarchy

example:

agricultural structure



occupations > engineering > agricultural engineering > agricultural structure
sciences > engineering > agricultural engineering > agricultural structure
occupations > engineering > manmade structures > agricultural structure
sciences > engineering > manmade structures > 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 greenhouses production structure silos
RELATED CONCEPTS	adjustment of production agricultural economics agricultural situation agriculture land reform structural change structural policies structure
ALTERNATIVE LABEL	agricultural structures and facilities facilities, agricultural land fragmentation structures, agricultural
BELONGS TO GROUP	economics

Added Upper Structure

- Enables “ease of first sorting”
- Like “Scenario A”
- Reuse of terms useful for intellectual organization e.g., “activity”, “characteristics” now have placement
- Additions, kept “separate” from GACS URIs
- Reuse of YSO and UMLS structures

Unified Medical Language System – Semantic types

Semantic types trees (with semantic groups)

Entity		[Entity] (continued)	
Physical Object	OBJC	Conceptual Entity	CONC
Organism	LIVB	Idea or Concept	CONC
Plant	LIVB	Temporal Concept	CONC
Alga	LIVB	Qualitative Concept	CONC
Fungus	LIVB	Quantitative Concept	CONC
Virus	LIVB	Functional Concept	CONC
Rickettsia or Chlamydia	LIVB	Body System	ANAT
Bacterium	LIVB	Spatial Concept	CONC
Archaeon	LIVB	Body Space or Junction	ANAT
Animal	LIVB	Body Location or Region	ANAT
Invertebrate	LIVB	Molecular Sequence	GENE
Vertebrate	LIVB	Nucleotide Sequence	GENE
Amphibian	LIVB	Amino Acid Sequence	GENE
Bird	LIVB	Carbohydrate Sequence	GENE
Fish	LIVB	Geographic Area	GEOG
Reptile	LIVB	Finding	DISO
Mammal	LIVB	Laboratory or Test Result	PHEN
Human	LIVB	Sign or Symptom	DISO
Anatomical Structure	ANAT	Organism Attribute	PHYS
Embryonic Structure	ANAT	Clinical Attribute	PHYS
Anatomical Abnormality	DISO	Intellectual Product	CONC
Congenital Abnormality	DISO	Classification	CONC
Acquired Abnormality	DISO	Regulation or Law	CONC
Fully Formed Anatomical Structure	ANAT	Language	CONC
Body Part, Organ, or Organ Component	ANAT	Occupation or Discipline	OCCU
Tissue	ANAT	Biomedical Occupation or Discipline	OCCU
Cell	ANAT	Organization	ORGA
Cell Component	ANAT	Health Care Related Organization	ORGA
Gene or Genome	GENE	Professional Society	ORGA
Manufactured Object	OBJC	Self-help or Relief Organization	ORGA
Medical Device	DEVI	Group Attribute	CONC
Drug Delivery Device	DEVI	Group	LIVB
Research Device	DEVI	Professional or Occupational Group	LIVB
Clinical Drug	CHEM	Population Group	LIVB
Substance	OBJC	Family Group	LIVB
Chemical	CHEM	Age Group	LIVB
Chemical Viewed Functionally	CHEM	Patient or Disabled Group	LIVB
Pharmacologic Substance	CHEM	Event	ACTI
Antibiotic	CHEM	Activity	ACTI
Biomedical or Dental Material	CHEM	Behavior	ACTI
Biologically Active Substance	CHEM	Social Behavior	ACTI
Neuroreactive Substance or Biogenic Amine	CHEM	Individual Behavior	ACTI
Hormone	CHEM	Daily or Recreational Activity	ACTI
Enzyme	CHEM	Occupational Activity	ACTI
Vitamin	CHEM	Health Care Activity	PROC
Immunologic Factor	CHEM	Laboratory Procedure	PROC
Receptor	CHEM	Diagnostic Procedure	PROC
Indicator, Reagent, or Diagnostic Aid	CHEM	Therapeutic or Preventive Procedure	PROC
Hazardous or Poisonous Substance	CHEM	Research Activity	PROC
Chemical Viewed Structurally	CHEM	Molecular Biology Research Technique	PROC
Organic Chemical	CHEM	Governmental or Regulatory Activity	ACTI
Nucleic Acid, Nucleoside, or Nucleotide	CHEM	Educational Activity	PROC
Organophosphorus Compound	CHEM	Machine Activity	ACTI
Amino Acid, Peptide, or Protein	CHEM	Phenomenon or Process	PHEN
Carbohydrate	CHEM	Human-caused Phenomenon or Process	PHEN
Lipid	CHEM	Environmental Effect of Humans	PHEN
Steroid	CHEM	Natural Phenomenon or Process	PHEN
Eicosanoid	CHEM	Biologic Function	PHEN
Inorganic Chemical	CHEM	Physiologic Function	PHYS
Element, Ion, or Isotope	CHEM	Organism Function	PHYS
Body Substance	ANAT	Mental Process	PHYS
Food	OBJC	Organ or Tissue Function	PHYS
		Cell Function	PHYS
		Molecular Function	PHYS
		Genetic Function	PHYS

YSO - General Finnish ontology

Content language

English

x

Search

- A-Z
- Hierarchy
- Groups
- New

- events and action
- action
- events
- periods of time
- phenomena
- objects
- abstract objects
- physical objects
- physical whole
- place
- systems
- properties
- abstract character
- accessibility
- activation level
- activity
- addiction
- advantages
- age
- agentivity
- ambiguity
- ambivalence
- appearance (looks)
- asymmetry

PREFERRED TERM

events and action

TYPE

Hierarchical concept

NARROWER CONCEPTS

- action
- events
- periods of time
- phenomena

NOTE

Events and action can happen or occur. They often have fairly definite temporal but indefinite spatial limits. A given event or action is only partially present at a given time, for example the French Revolution was not wholly present at the moment of the Tennis Court Oath. A part of an event or action can constitute an independent event or action at a given time. Unlike objects, entities in this class often have causal relationships. Several events or actions can occur simultaneously at the same spatial point. Entities that participate in events and actions are called objects.

IN OTHER LANGUAGES

tapahtumat ja toiminta	Finnish
händelser och handling	Swedish

Upper Structure

- **OBJECTS**

- material things that can be seen or touched or visited (locations), including “abstract objects” such as ideas, models

- **EVENTS AND ACTIONS**

- something that happens, such as processes or phenomenon

- **PROPERTIES**

- attributes / characteristics / qualities of something/objects

Hierarchy Status

- 100% of terms aligned to upper structure
- First draft of Scope Notes for upper structure (many taken from YSO or UMLS)
- Satisfactory progress to have cleaner hierarchies that are BT/NT transitive
- “Work in progress”/ not perfect, takes time

View the work done

- All the work on the data side is in GitHub
<https://github.com/agrisemantics/gacs-qip/tree/master/gacs-data>
- SKOSMOS installation (Osma)
<http://tester-os-kktest.lib.helsinki.fi/gacsdemo/gacs/en/>

Non-Core GACS

- Using top 10,000 terms from each thesaurus includes many terms very specific to one region, “not of global interest”
 - 4-H Youth Development Program
 - Food and Drug Administration
- Parking place for some terms ambiguous and need more work e.g., “practice”
- If desired, can move more terms to this location or eliminate (only 10 terms here)

GACS Core QIP-development

Content language

English



Search

Alphabetical

Hierarchy

Groups

- EVENTS AND ACTIONS
- NON-CORE GACS
- OBJECTS
- PROPERTIES

Vocabulary information

TITLE

GACS Core QIP-development

Global Agricultural Concept Scheme Core

DESCRIPTION

The Global Agricultural Concept Scheme (GACS) is a hub for concepts related to agriculture, in multiple languages, for use in Linked Data. The idea for GACS emerged out of discussions at the World Congress of IAALD, the International Association of Agricultural Information Specialists, in July 2013. The Food and Agricultural Organization of the United Nations (FAO), CAB International (CABI), and the National Agricultural Library of the USA (NAL) agreed in October 2013 to explore the feasibility of developing a shared concept scheme by integrating their three thesauri: the AGROVOC Concept Scheme, the CAB Thesaurus (CABT), and NAL Thesaurus (NALT). In the GACS vision, the integration of these three thesauri is but the first step towards the realization of a hub that links to and from the concept schemes beyond the initial three, and in multiple language areas.

Thematic Classification

- Based on CABT Classification / UAT
- Thematic classification not assigned for ~2800 concepts. Goal: complete for 100% of concepts.
- Frequency distribution of terms to categories very uneven, many categories with one or without concepts.
- There are no scope notes for the classification.

Classification Scheme applied to GACS Concepts

Two-Letter

<i>Code</i>	<i>Classification Heading</i>
CA	GENERAL
FA	PHYSICAL SCIENCES
JA	EARTH SCIENCES
LA	LIFE SCIENCES
PA	APPLIED SCIENCE AND TECHNOLOGY
WA	SOCIAL SCIENCES AND HUMANITIES

Thematic Classification progress

- 100% of concepts are sorted into classification
 - Principle: remove terms from “common terms” category if a better category found
 - Principle: can allow concepts to be in more than one category
- First draft of Scope Notes for Classification
 - “thin” but with examples of Concepts assigned
 - Some categories with no / few GACS Concepts, these are noted

NY Category – “organisms of uncertain taxonomy”

- Contained only 2 entries,
 - Pneumocystis
 - Pneumocystis carinii (aka P. jiroveci)

Moved to NQ category, “fungi”

Categories not used or with few terms

- **SW vermiculture**
- **WQ arts**
- **WS history**
- **WY philosophy and ethics**
- **LU evolution**
- **LX natural history**
- **NY organisms of uncertain taxonomy**

QC of Category work

- insulating materials

http://id.agrisemantics.org/gacs/G_TL ✓

http://id.agrisemantics.org/gacs/G_CC ✗

- Removal of second category “CC=common terms”, retention of “TL=materials”

QC of Category work

Double category assignment –

- plant protection equipment
 - “TN=equipment”
 - “PR=health protection”
- milking machines
 - “TN=equipment”
 - “SS=animal production”

plant protection equipment

http://id.agrisemantics.org/gacs/G_TN http://id.agrisemantics.org/gacs/G_PR

milking machines

http://id.agrisemantics.org/gacs/G_TN http://id.agrisemantics.org/gacs/G_SS

GACS Core QIP-development

Content language

English



Search

Alphabetical

Hierarchy

Groups

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

Vocabulary information

TITLE

GACS Core QIP-development
Global Agricultural Concept Scheme Core

DESCRIPTION

The Global Agricultural Concept Scheme (GACS) is a hub for concepts related to agriculture, in multiple languages, for use in Linked Data. The idea for GACS emerged out of discussions at the World Congress of IAALD, the International Association of Agricultural Information Specialists, in July 2013. The Food and Agricultural Organization of the United Nations (FAO), CAB International (CABI), and the National Agricultural Library of the USA (NAL) agreed in October 2013 to explore the feasibility of developing a shared concept scheme by integrating their three thesauri: the AGROVOC Concept Scheme, the CAB Thesaurus (CABT), and NAL Thesaurus (NALT). In the GACS vision, the integration of these three thesauri is but the first step towards the realization of a hub that links to and from the concept schemes beyond the initial three, and in multiple language areas.

Some analysis of changes done to-date

28-Aug-17 version

- <https://github.com/agrisemantics/gacs-qip/blob/master/gacs-data/webstudio-exports/NOTES.md>

Removed triples, by RDF property

1416 <<http://www.w3.org/2004/02/skos/core#broader>>
507 <<http://www.w3.org/2004/02/skos/core#topConceptOf>>
81 <<http://www.temis.com/luxid-schema#memberOf>>
18 <<http://www.w3.org/2004/02/skos/core#related>>
8 <<http://www.w3.org/2004/02/skos/core#scopeNote>>
8 <<http://www.w3.org/2004/02/skos/core#prefLabel>>
4 <<http://www.w3.org/2004/02/skos/core#definition>>

Added triples, by RDF property

2839 <<http://www.temis.com/luxid-schema#memberOf>>
1128 <<http://www.w3.org/2004/02/skos/core#broader>>
419 <<http://www.w3.org/2004/02/skos/core#related>>
43 <<http://www.w3.org/2004/02/skos/core#definition>>
31 <<http://www.w3.org/2004/02/skos/core#prefLabel>>
23 <<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>>
2 <<http://www.w3.org/2004/02/skos/core#scopeNote>>

Next Steps

- Agreement to continue work until November
- Refinement of scope notes, adherence to scope notes for both hierarchical and thematic classification
- Review and feedback by partners, stakeholders
- Beyond November, there is still quality work to be done on definitions, scope notes as well as refinement of hierarchy / classification

Discussion

Lori Finch

Lori.Finch@ars.usda.gov

Sujata Suri

Sujata.Suri@ars.usda.gov

Tom Baker

Tom.Baker@tombaker.org

Osma Suominen

Osma.Suominen@helsinki.fi

Discussion points

- WG needs time to review.
- Long term: Generally, where we do think GACS is going?
- Long term: maintenance? Sustainability?
- Short term: Do we replace GACS Beta with this new version? When is decision point? Who?
- When is next meeting GACS WG?