DCPAPERS
Proc. Int'l Conf. on Dublin Core and Metadata Applications 2016

The Global Agricultural Concept Scheme and Agrisemantics

Thomas Baker
Independent FAO consultant, Bonn, Germany
tom@tombaker.org

Caterina Caracciolo,
Food and Agriculture Organization of the UN,
Rome, Italy
caterina.caracciolo@fao.org

Anton Doroszenko,
CAB International,
Wallingford, UK
a.doroszenko@cabi.org

Lori Finch,
National Agricultural Library, USDA, USA
lori.finch@ars.usda.gov

Osma Suominen,
National Library of Finland,
Helsinki, Finland
osma.suominen@helsinki.fi

Sujata Suri
National Agricultural Library, USDA, USA
sujata.suri@ars.usda.gov

Abstract

Key concepts from three thesauri about agriculture and nutrition—AGROVOC, CAB Thesaurus, and NAL Thesaurus—have been merged into a Global Agricultural Concept Scheme (GACS). The respective partner organizations—Food and Agriculture Organization of the UN (FAO), CAB International (CABI), and the USDA National Agricultural Library (NAL)—undertook this initiative in 2013 with the goal of facilitating search across databases, improving the semantic reach of their databases by supporting queries that freely draw on terms from any mapped thesaurus, and achieving economies of scale from joint maintenance. The GACS beta release of May 2016 has 15,000 concepts and over 350,000 terms in 28 languages.

The creation of GACS began by mapping three sets of 10,000 frequently used concepts from the three thesauri to each other, pairwise. Mappings were vetted by experts; vetted mappings were algorithmically checked for awkward clusters, or "lumps"; and lumps were resolved through discussion on teleconferences and in meetings—for example, by drawing a line between "energy intake" (related to organisms) and "energy consumption" with the narrower "fuel consumption" (related to natural resources). Mappings were manually corrected, and GACS was iteratively re-generated, until the set of concepts was considered stable enough for publication as GACS Beta.

Some inevitable results of this process of aggregation, such as overlapping labels, have already been fixed. Other issues, such as concepts with multiple hierarchical relations ("polyhierarchy"), have yet to be tackled. The working group has revived a classification scheme, developed jointly in the 1990s, to tag concepts by thematic group. Concepts are being typed as chemical, geographical, organisms, products, or topics. Alongside generic thesaurus relations to broader, narrower, and related concepts, organisms will be related to relevant products.

GACS is seen as a first step for Agrisemantics, an emerging community network of semantic assets relevant to agriculture and food security. Within Agrisemantics, the general-purpose, search-oriented concepts of GACS are intended to serve as a hub for concepts defined, with more precision, in a multitude of ontologies modeled for specific domains. Ontologies, in turn, are intended to provide global identity to concepts used in a vast diversity of quantitative datasets, such as sensor readings and crop yields, defined for a multitude of software applications and serialization formats.

Such semantic authority control of data elements could support, for example, an analysis of the yield gap in sub-Saharan Africa. A wheat data element, labeled 'GW' in a phenotype dataset, could be mapped to the concept 'grain weight' as defined and globally identified in the CGIAR Crop Ontology. In turn, the Crop Ontology concept could be mapped to the broader concept 'Grain' in GACS. Searches could return not only datasets about grain weight, but references to published papers where the weight of the grain was studied.

OPEN ACCESS
This article is licensed under a **Creative Commons Attribution 4.0 International License**, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and cite the source.

https://doi.org/10.23106/dcmi.952137812

Agrisemantics aims at improving the discoverability and semantic interoperability of agricultural information and data for the benefit of researchers, policy-makers, and farmers with the ultimate goal of enabling innovative responses to the challenges of food security under conditions of climate change. Achieving these goals will require innovation in processes for the cooperative maintenance of linked semantic assets in the modern Web environment.

OPEN ACCESS
This article is licensed under a **Creative Commons Attribution 4.0 International License**, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and cite the source.

https://doi.org/10.23106/dcmi.952137812