Position paper and commentary on “GACS: A proposal for future development. V4”  
by Caterina Caracciolo, FAO, 18th October 2016

Anton Doroszenko, Phil Roberts, Brandon Whitehead, CABI

4th November 2016

Executive Summary

CABI strongly recommends that GACS remain a core concept scheme, with no subdivisions, which links out to other more detailed, more comprehensive or more focussed concept schemes or knowledge organization systems. The rationale is built on simplicity of maintenance, relative stability, and low resource demand.

Context

CABI has been participating in GACS on the understanding that it would be a tightly focussed, utilitarian, generally upper-level vocabulary, which would act as a hub connected to external, more specialised vocabularies, ontologies and other knowledge organization systems (KOS). The strategy was to select the most used terminology in the participants’ bibliographic databases and merge and harmonize them as best we could with the resources available. This was largely achieved, with some lesser issues still be resolved. FAO, NAL and CABI agreed to the idea of ‘GACS Core’ which would be a component within the much broader umbrella of Agrisemantics (<http://www.agrisemantics.org/>). Whether it is called ‘GACS Core’ or simply ‘GACS’ is immaterial to this concept of a connecting hub.

The advantages of GACS Core was that it was relatively small and that any evolution would also be relatively small and could be managed within the resources available. CABI currently has no funding devoted to GACS in the remainder of 2016 or in 2017.

It was recognised that there was no need to significantly expand GACS in order to maintain its utility and value. Specialist KOS could link through GACS and be maintained by the external participating organizations best placed to maintain them without burdening the GACS Editorial WG.

Commentary on FAO paper on GACS future development

This commentary goes through the paper “GACS: A proposal for future development. V4”. We will provide a quote from the paper under its appropriate section heading and follow it with our comments.

*Goals of the Global Agricultural Concept Scheme (page 2)*

*Quote: “Achieve efficiency in maintenance by removing duplication of efforts.”*

Comment: This anticipated removal of duplication of efforts is much overstated. If GACS Core remains small there will be little gain in efficiency since it will evolve relatively slowly. In any case we already ‘outsource’ several areas of general agreement, such as geopolitical concepts to ISO lists, and these don’t change much either. The work load would not change much for CABI. We will not merge CABT with GACS for the many reasons we have discussed previously.

*Quote: “Capitalize on existing resources”*

Comment: CABI would have to consider this very closely, particularly where there is a relinquishing of control of IP.

*What is GACS? (page 2)*

*Quote: “Content in GACS will be organized into different concept schemes”*

Comment: How is this compatible with the statement in the Executive Summary which says “abandon the notions of GACS extension, module and partitions”? This seems to indicate that some wish GACS to be more than a core vocabulary.

*Quote: “The loose semantics of thesauri provides no formal and general way to enforce ‘restrictions’ on topics, such as the definition of the admitted values for attributes and relationships between entities.”*

Comment: Generally, this is not how a thesaurus works; this is talking about ontologies, which have a different structure and rationale. This comment also betrays a lack of understanding of what GACS is for and how it will be used. Nevertheless, thesauri can be ‘ontologized’ to some extent, as has CABT by the use of custom relationships and category codes (=attributes). The hierarchy also imposes a context for concepts.

*Quote: “The creation of new concepts to meet any specific need can quickly make a thesaurus grow to sizes that hamper usability.”*

Comment: On the point of usability this is not necessarily true. But it is agreed that we need to keep GACS Core small and focussed. To illustrate, CABT has a ‘core’ of general concepts with a broad overlap with GACS, which evolves relatively slowly, in addition to a substantial taxonomic part. The taxonomic part is a challenge, but is rewarded by more accurate and up-to-date indexing/tagging of data to satisfy our business requirements. But that does not need to be in GACS Core. Where these more comprehensive source terminologies reside is less of an issue in an interconnected electronic landscape.

*Quote: “part of a community sharing the same view of semantic interoperability of agricultural data”.*

Comment: GACS is a controlled vocabulary and not about semantics *per se.*

*Quote: [GACS participants] share the same policies on vocabulary reuse.*

Comment: That cannot be assumed. For example, can one be sure that ‘wheat’ means the crop plant or the commodity? [See also Osma’s comment CC1] Different user communities might use it differently. In addition, policy is not fixed. CABI has more than 100 years of activity and use of specific terminology can and has changed over that time.

*Vocabularies as authoritative lists of entities (page 4)*

*Quote: “often a considerable amount of effort put into maintaining local vocabularies when reusing existing resources would be more efficient.”*

Comment: The three examples given immediately above this quote are exactly what GACS Core is about. These lists of crops, livestock and countries are relatively small and GACS provides a standardized naming resource. But it is unrealistic to expect farmers not to use local names for crops, pests, etc. and the need to maintain local vocabularies will not go away. GACS cannot impose the use of its terminologies, even if it is more efficient for data interchange. But at least GACS Core provides the link to these local resources.

*Quote: “Controlled vocabularies may be already available…maintained by authoritative institutions…but the format…may be sub-optimal for re-use.”*

Comment: This is a particular challenge in taxonomic fields, such as organisms and soils. It is good in principle to outsource these areas to specialist databases that link to GACS, but the reality is that these taxonomic resources rarely provide their names in a format we can link out to. The task to maintain just the taxonomic names is so huge that everything else gets relegated to the margins. In the meantime, thesauri like CABT can act as proxies to these specialist databases. CABI is willing to do that because we have a business need to do it. Therefore, again, the model of GACS Core linking to the major partners is the only viable option in the short term.

*Quote: “The intended goal of these vocabularies is to provide lists of fundamental concepts in various domains.”*

Comment: Yes, indeed. But the examples that follow demonstrate a narrow FAO world view, presumably related to major FAO programmes. The expectation that these lists are *“very focussed, possibly small in size”* is not the reality when it comes to things like ecology or weeds, pests, diseases and their control. In other words, trying to mould GACS to provide *“lists of fundamental concepts in various domains”* is problematic. Tom Baker has commented on the technical challenges of maintaining URIs in doing this and recommends we not pursue this idea. By all means keep the fundamental concepts within GACS, but create the various domains elsewhere.

*Quote: “It is not to be expected that GACS cover ‘all’ topics relevant to agriculture”.*

Comment: Absolutely! That is what GACS Core is about. Please don’t try to change GACS into something we did not agree to when FAO, CABI and NAL began our collaboration.

*Thesauri (page 5)*

*Quote: “Thesauri…are best used in applications where strict control on the allowed concepts and their structure and composition are not crucial.”*

Comment: We may not understand what is meant here, but a thesaurus is a controlled vocabulary and consistency is crucial for information retrieval. The hierarchical structure of thesauri also provides valuable context for a concept. It depends on the application.

*Ontologies (page 5)*

*Quote: “So no ‘upper ontology’ at all? No backbone?*

Comment: GACS is not an ontology. Upper ontologies are conceptually satisfying but of little practical use in the context of GACS. GACS Core *is* the backbone.

*Quote: “The point is not to destroy what we have, nor to decompose the whole GACS Core in mini vocabularies, but to define a strategy that allows GACS to be a reference outside the world of thesauri.”*

Comment: Yes, agreed. What is being described here is GACS Core.

*Envisaged GACS users and application (page 6)*

*Quote: “We have identified a number of use cases.”*

Comment: Yes, this is GACS Core.

*Content Maintenance – Editorial Board (page 6)*

Comment: Because of the current lack of funding CABI cannot commit effort to maintain anything other than GACS Core and even then at a low level of input.

*Infrastructure (page 7)*

*Quote: “Proposal – VocBench for content editing.”*

Comment: VocBench may be suitable if GACS Core does not expand much beyond its current size. There are performance issues with the online nature of VocBench and we find the editorial tools in VocBench to be limiting. We are happy with the concession that *“partners are not forced to switch to it”.*

*Some considerations on GACS Core – May 2016 release (page 8)*

*Quote: “The creation of a common pool of those concepts would improve data sharing and reuse.”*

Comment: This still ignores history. CABI has a database which documents resources going back more than 100 years. Imposing commonality on that history is not straightforward.

*Quote: “The underlying idea was that [GACS was] such a common pool.”*

Comment: Yes, indeed. Keep GACS Core *small*; 10K-15K concepts are manageable.

*Quote: “The resulting GACS Core consisted in a large pool of concepts with shallow organization.”*

Comment: It could be argued that the shallow organization was not by design. But is it a major issue that needs significant resources? The GACS WG did not have the resources or time to consider the hierarchy in detail in GACS Beta, but some tidying up may help improve its consistency.

*Known open issues in GACS Core (page 9)*

*Quote: “Holes…One example is the concept Soil Horizon A.”*

Comment: A solution might be to remove the narrow, specific concepts and substitute a broader one. Then GACS Core becomes a higher level concept scheme linking to other more specific schemes, such as a soils scheme in this example.

*Quote: “Inconsistent hierarchies.”*

Comment: This was inevitable given there were three source thesauri. We suggest a hierarchy is needed – it provides structure and context in the concept scheme. Effort here will pay off.

*Quote: “Top concepts.”*

Comment: A classification scheme similar to CABT Classified or NALT may be the answer to the proliferation of top concepts. We are unsure how much users browse concept schemes or thesauri, so it may not matter that there are a lot of top concepts. CABT uses category codes at the moment as a structural element, but intends to reinstate the full classified thesaurus in due course.

*Further considerations on current GACS (page 10)*

*Quote: “Their content evolves over time, making a GACS Core with those foundations soon outdated.”*

Comment: That is exactly why GACS Core should remain *small*. It evolves more slowly and we can manage that.

*Quote: “The approach proposed … is to keep vocabularies and classifications independent.”*

Comment: CABI has been working in this manner for a long time and it has served us well.

*Quote: “Delegating parts of GACS to external vocabularies.”*

Comment: One needs to realise that some of these resources are just lists of names. There is no indication of ‘preferred’ or ‘non-preferred’ neither do they provide any context to the names, e.g. is it a pest, weed, disease, etc.?

*Quote: “GACS partners are unlikely to have the resources or willingness to maintain all of the vocabularies by themselves.”*

Comment: In fact, CABI fully funds CABT because we have a *business need.* CABT supports our core publishing and research business. This is the main problem with GACS – what is the business case? Unless we can demonstrate what the business need is funding will dry up in the long term, in our opinion. GACS cannot be supported merely on admirable ideals.