

A Roadmap for the Future of Shared Vocabularies in Australia: Harmonising Across Sectors, Disciplines and Internationally

Steve McEachern (ADA) and Lesley Wyborn (ANU)
for the Vocabs Roadmap group

Today's agenda

1. Welcome
2. An overview of the Vocabularies Roadmap - our work so far
3. Open discussion
4. Next steps

This presentation is available from:

<http://bit.ly/vocabs2023>

What are vocabularies?

- An “organized arrangement of words and phrases used to index content and/or to retrieve content through browsing or searching” (Harpring, P., 2010)
- These vocabularies and support services form an ecosystem.
- But this ecosystem is diverse and largely uncoordinated.
- Vocabularies that are available are not necessarily findable, let alone It is this diversity and coordination that this Roadmap seeks to address.

Why a roadmap?

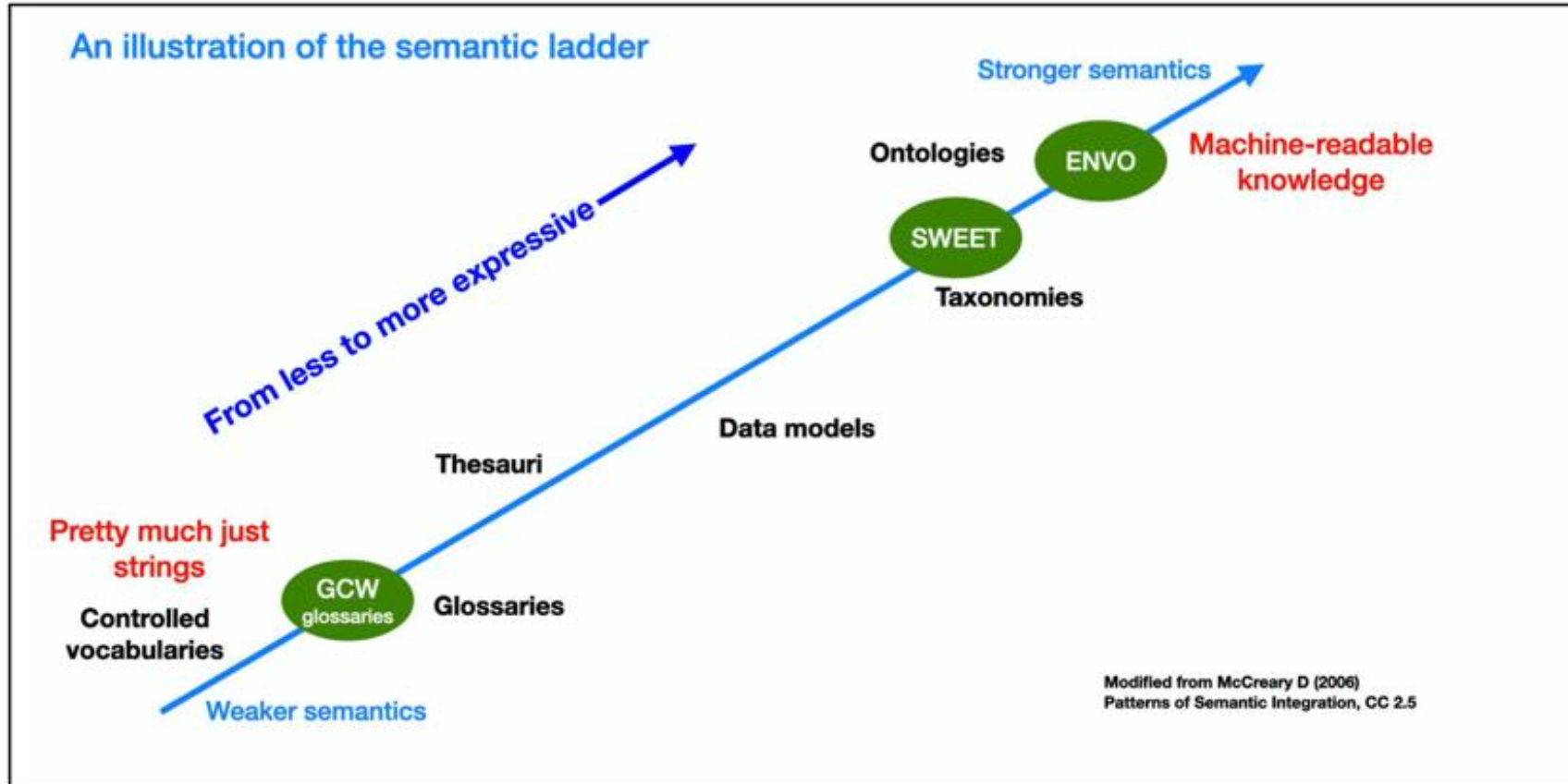
- In November 2022, a small group of colleagues with expertise and interest in vocabularies - their development, management, support and use - came together in Canberra, Australia to explore how the ecosystem of vocabularies in Australia could be made more sustainable and broadly used.
- This working group has met on a regular basis since that foundation workshop, with the aim of establishing a roadmap for the future of vocabularies in Australia and more broadly.
- This roadmap therefore lays out our shared view of what a future vocabulary ecosystem might look like, key elements of that ecosystem, and possible means for achieving that future state.

The vocabularies ecosystem

- A vocabulary in and of itself is an inherently useful thing.
- It helps to organise knowledge and terminology, and to enable access to and use of that knowledge in a consistent way between users, organisations and systems.
- The vocabulary however exists within a broader ecosystem of usage - from a simple controlled list within a single organisation or project, to a commonly held and used terminology used across increasingly broad stakeholder groups.

The Semantic Ladder

Source: Ruth Duerr, Pier Luigi Buttigieg, Gary Berg-Cross, Kai Lewis Blumberg, Brandon Whitehead, Nancy Wiegand, Kate Rose preprint: [Harmonizing GCW Cryosphere Vocabularies with ENVO and SWEET: Towards a General Model for Semantic Harmonization](#)



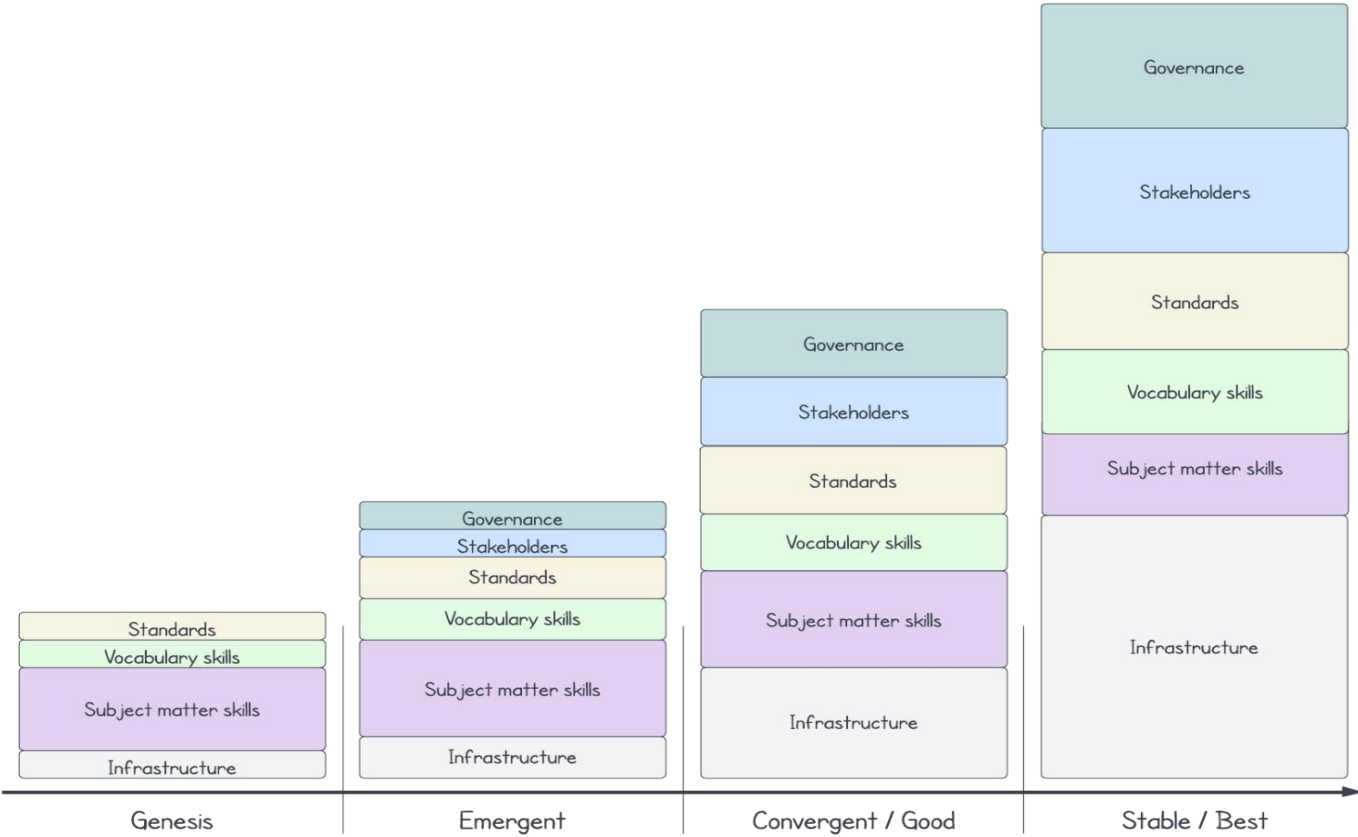
What's in a vocabulary ecosystem?

A “broadly distributed ecosystem for vocabulary creation, maintenance, and use based on a commonly agreed URI infrastructure, built to support distribution of terms to consumers based on their explicit preferences” ([NISO, 2017](#))

In this roadmap therefore we outline what **we** envisage as the core elements of such an ecosystem surrounding the vocabularies themselves:

- The stakeholders in the system and their relationships
- Governance arrangements
- Subject matter and technical skills
- Standards
- Policy
- Infrastructure, tools and technologies

Maturity and the vocabulary ecosystem



Vision and Mission

- “Our **vision** is that within the next 5-10 years the utility of vocabularies is well understood within Australia and there to be a burgeoning ecosystem of vocabularies and associated services that lower the barriers of using vocabularies as a key foundation of data assets, services and infrastructure across the research, government and broader 'data' sector.”
- “The group’s **mission** is to further the consistency and sustainability in the development, implementation, (management?) and use of vocabularies across domains to solve real world problems both nationally and internationally, through establishment of an effectively coordinated community and services

Principles (1)

1. Approach all **vocabularies as a representation of knowledge** in a particular domain that could potentially have utility across a broader community.
2. The **means through which a vocabulary is transmitted**, while important for their management and use, **should not be critical** to the end user.
3. **Data is more reusable** if it is well described and connected to **vocabularies that are FAIR, persistent and sustainable**.
4. We should be looking to **move from human to machine-interoperable services**. Our customers (e.g. data collectors) should become our customers' customers (e.g. users of data collection software that consumes a vocabulary). Direct consumers will increasingly become machines rather than humans

Principles (2)

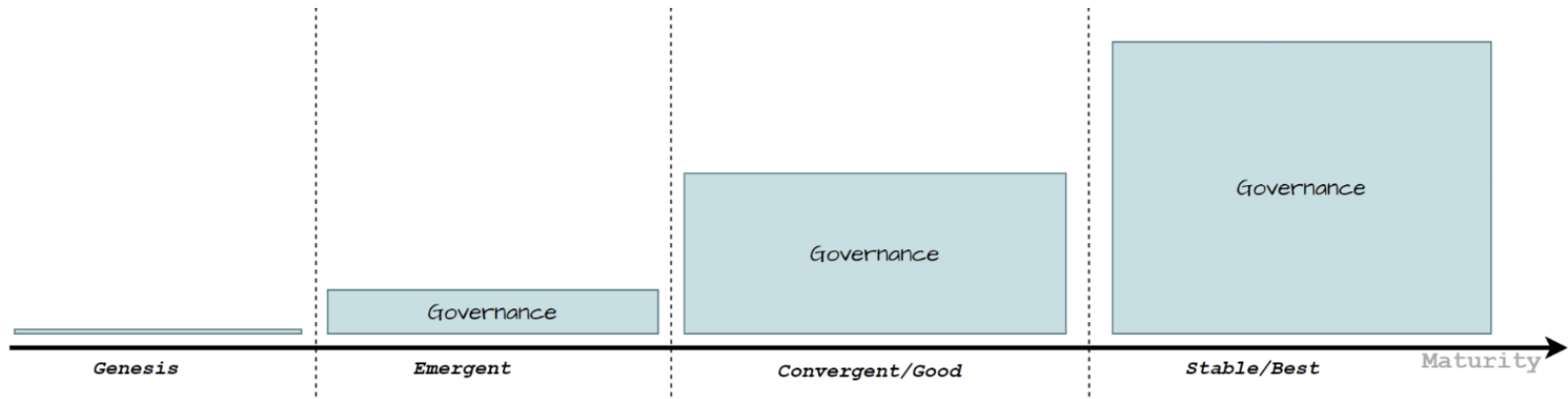
5. The creation, management and use of vocabularies requires **governance, infrastructure and funding** to ensure both sustainability and regular updating and maintenance.
6. We need **fewer but better vocabularies** - we should be supporting convergence and reuse of vocabularies where possible (including finding, mapping and harmonisation)
7. We need **means for bringing indigenous and other forms of knowledge** into vocabularies
8. Vocabularies should be **multilingual**, with translation based around equivalent concepts.
9. Something is needed about **building trust in the vocabs**. Social trust on the expertness of the creator of the vocab and system trust in that the vocab services are well managed and governed>

1. Stakeholders

1. Vocabulary **Content Creators**: The subject matter experts that are involved in distilling the knowledge in crafting the authoritative content of vocabularies, shaping the semantic foundation.
2. Vocabulary **Content Publishers**: This community takes on the pivotal task of making vocabularies available to a broader audience.
3. Vocabulary **Infrastructure Providers**: Offering the essential backbone, these stakeholders provide the necessary technical infrastructure and platforms that underpin the creation, storage, and retrieval of vocabularies.
4. Vocabulary **Tool Builders**: This community develops tools that facilitate the creation, enrichment, and management of vocabularies, enhancing the overall ecosystem.
5. Vocabulary **Consumers**: At the heart of this ecosystem, vocabulary consumers encompass a diverse range of users, from software applications to researchers, who rely on these semantic structures to derive meaningful insights from data.

2. Governance

- As a vocabulary evolves from its genesis to wider adoption, addressing the governance concerns becomes increasingly important for creators, publishers and users of vocabularies. At each stage of the maturity continuum, there's a need for a significant shift in governance requirements to effectively sustain and expand adoption.
- To foster the growth and effective use of vocabularies, it is crucial to address and adapt to key governance concerns as they evolve over time.
- To create a thriving ecosystem for vocabularies in Australia, we need to implement appropriate governance practices. This means establishing mechanisms that incentivize vocabulary professionals to consider and implement appropriate governance concerns, and to provide them with the support they need to do so.



3. Skills

- As a community advances up the semantic ladder or as a vocabulary gains wider adoption and use, the need for specialised skills in vocabulary development and management becomes more pronounced. These skills are essential to ensure the continued FAIRness and usability of the vocabulary.
- These skills span from technical proficiencies to abilities in bridging connections, delivering solutions, problem-solving, and managing change. In most cases, vocabulary workers draw from a combination of these skills when developing vocabularies to enrich, transfer, and harmonise data.

4. Standards

- Vocabularies themselves aspire to be standards.
- Emphasis here is on how technical and other standards deliver the ‘standard’
- Standards can include:
 - 1) Content standards**
 - a) Universal: e.g., Units of Measure, List of Countries, Place names
 - b) Domain Specific: e.g., GeoSciML, GeodesyML, DDI
 - 2) Technical standards**
 - a) e.g. OWL, SKOS, etc.
 - 3) Process standards**
 - a) e.g., PROV, BPML,
- Governance, process, curation and management, etc.
- (These are discussed here, but linked to related sections - e.g. Relationships section talks about roles)

5. Policy

Policy agendas set the framework within which vocabularies are designed and used. It is possible to articulate three broad forms of policy which impact upon the use of vocabularies and services

1. High level policies: set the direction for government and organisational activities, establishing core outcomes to be achieved, and frameworks for their implementation and evaluation. Australian examples include (e.g. alignment of activities with SDGs, biodiversity policy or Closing the Gap).
2. Information policies: establish the type of information that needs to be managed by organisations, and the means through which that information should be managed and used. For example, the National Archives of Australia information management policy
3. Internal policies: internal-facing policies that then influence internal direction, e.g. NHMRC policy, UNSW data governance policy

6. Infrastructure and technology (platforms and tools)

- Supporting the up-skilling of vocabulary builders and users suggests some convergence on technologies and platforms is required.
- Standards also support the concept of tooling convergence - so long as these support vocabulary and standardisation convergence.
- How can we advance across such a diverse range of areas?
- The challenge is not to have a single tool that is able to host vocabularies - our toolsets need to work across this whole problem space for all vocabularies.

Roadmap - setting directions

	Short term (1-2 year)	Medium term (2-3 year)	Long term (3-5 year)
Stakeholders and Relationships	<ul style="list-style-type: none"> ● Establish a working group and communication channels for development of the roadmap. ● Expand the focus of AVSIC from an operational knowledge sharing forum for vocabularies, to also include strategic considerations of the emerging needs of vocabularies content publishers and infrastructure providers. ● Identify effective modes of communication across the five different communities with a focus on the needs for FAIR vocabularies. <ul style="list-style-type: none"> ○ Communicate value through publishing a collection of high impact success (and failure) stories consumable in plain language across the sectors (balance across 	<ul style="list-style-type: none"> ● Establish long term leadership and governance arrangements that support a collective of individuals and organisations working toward a common vision for vocabularies 	<ul style="list-style-type: none"> ● Establish organisational and institutional structures and activities that act as scaffolding to enable the formation and sustainment of the relationship between the five different communities involved in the vocabulary ecosystem. ● Establish mechanisms for translation of community learnings to inform evolution of best practices at different scales (local, national, international)

Next steps

Publication of first draft of Roadmap

Presentation at Vocabularies workshop:

2023 Vocabulary Symposium: [Please register](#)



Newsletter Subscribe

Contact us



About us

Programs and Projects

Services

Resource Hub

Get Involved

News and Events

For Researchers

ARDC > News and Events > Events > 2023 Vocabulary Symposium: FAIR Vocabularies for All

14-15
NOV

2023 Vocabulary Symposium: FAIR Vocabularies for All

Spanning research, government and industry, this 2-day hybrid event will offer opportunities to discuss approaches to using, developing, publishing and maintaining vocabularies.



- 14-15 November 2023, in Canberra and online
- So far we have have 172 registrations (41 in person and 131 online)

Discussion

A: What are your initial reactions?

1. Have we left anything off?
2. Do you need clarity on anything?
3. Do you have any specific priorities?
4. Are we targeting the correct audience?