Australian National University

A STRATEGIC COMMUNITY ROADMAP FOR AN AUSTRALIAN FAIR VOCABULARY ECOSYSTEM

Wong, M., Atkins, N., Boyle, D., Dharmawardena, K., Hobern, D., McEachern, S., and Wyborn, L.

The Australian Data Archive

POLIS@ANU - The Centre for Social Policy Research

Research School of Social Sciences

College of Arts and Social Sciences

https://ada.edu.au/

RSSS Building | 146 Ellery Crescent

The Australian National University

Canberra ACT 2600 Australia

www.anu.edu.au

TEQSA Provider ID: PRV12002 (Australian University)

CRICOS Provider Code: 00120C

About POLIS

The Centre for Social Research and Methods has been rebranded as *POLIS: The Centre for Social Policy Research*. As part of this change, the Centre for Aboriginal Economic Policy Research (CAEPR) has joined POLIS and is being renamed the Centre for Indigenous Policy Research.

POLIS – which draws from the Ancient Greek for the administrative centre of the City-State – is designed to provide a designated space at the ANU for discussion, debate and research on the formulation of social policy. The rebrand will allow POLIS to better capture and market the key work of the centre in providing research and expertise on social policy in response to community and federal and state/territory government needs and requirements.

POLIS delivers exceptionally robust data and evidence-driven insights into the key challenges facing contemporary Australia. This provides the foundational cornerstones of informed social policy development amongst leading stakeholders within our modern policy: government, community groups, business representatives, and educators.

POLIS is home to seven research centres:

- Centre for Indigenous Policy Research
- Centre for Social Research
- Centre for Educational Equity
- Centre for Crime and Social Justice
- Centre for Gambling Research
- Centre for Data, Analytics, and Evaluation
- Social Impact Hub

Author affiliations:

Steve McEachern Associate Professor with POLIS: The Centre for Social Policy Research School of Social Sciences and Director of the Australian Data Archive.

Megan Wong Federation University, Centre for eResearch and Digital Innovation and Australian Research Data Commons ARDC, **Natalia Atkins** Integrated Marine Observing System (IMOS), Hobart, **Douglas Boyle** Melbourne Medical School, Department of General Practice and Primary Care, The University Melbourne, **Kheeran Dharmawardena** Cytrax, Melbourne, Australia, **Donald Hobern** University of Adelaide, Adelaide, Australia, and **Lesley Wyborn** National Computational Infrastructure, Australian

A Strategic Community Roadmap for an Australian FAIR Vocabulary Ecosystem

Wong, M., Atkins, N., Boyle, D., Dharmawardena, K., Hobern, D., McEachern, S. and Wyborn, L.

Abstract

Vocabularies serve as essential anchor points for both humans and machines in effective and efficient data processing. Vocabularies include controlled vocabularies, taxonomies, thesauri, ontologies, and metadata schemas, each of which contributes to an ecosystem that encompasses the people, resources, standards, tools, platforms, policies, and practices that make them accessible and useful for researchers.

Currently, Australia's vocabulary ecosystem is fragmented and lacks effective coordination. To address these challenges, a Vocabulary Workshop was held in 2022, sponsored by the Australian Data Archive, the Australian Research Data Commons, and CODATA. From this workshop, a proposal for a strategic roadmap emerged, followed by numerous community consultations conducted between 2022 and 2024.

The resulting Strategic Community Roadmap outlines a pathway for Australia to transition from its fragmented landscape to a cohesive and dynamic FAIR Vocabulary Ecosystem. It presents a Vision, Mission, and 57 recommendations categorised into seven key topics, organised around four Strategic Themes. Each recommendation is prioritised by its importance and urgency for implementation.

The goal is to promote wider adoption and greater community engagement with machine-actionable vocabularies, emphasising the social and technical support needed to address current data interoperability challenges. This serves as a call to action to maximise the societal, economic, and environmental benefits that can be derived from our national research and data initiatives.

National University, Auscope Limited,
Melbourne, Australia, Australian Research
Data Commons ARDC

Acknowledgments

Contents

Versioning History

Version	Date	Authors	Notes
Draft prepared at Workshop. Version 0.1	November 18 2022	Workshop attendees: Atkins, N., Birtles, S., Boyle, D., Brownlee, R., Burton, A., Car, N., Chuc, E., Cox, S., Dharmawardena, K., Druken, K., Hancock, A., Hoy, D., Kneebone, L., Lawley, M., McEachern, S., McLaughlin, A., Newman, P., Rigby, M., Smith, L., von Gavel, S., Williams, M., Wong, M., and Wyborn, L.	Initial draft prepared at the workshop
Version 0.2	October 29 2024	Wong, M., Hobern, D., McEachern, S., Atkins, N., Boyle, D., Dharmawardena, K., and Wyborn, L.	Released at eResearch Australasia 2024 and the December AeRO eResearch newsletter 2024
Version 0.3	May 2025	Wong, M., Hobern, D., McEachern, S., Atkins, N., Boyle, D., Dharmawardena, K., and Wyborn, L.	Incorporated feedback from eResearch Australasia 2024 and Australian Institute of Aboriginal and Torres Strait Islander Studies (AIATSIS) review.
Version 1.0	August 2025	Wong, M., Hobern, D., McEachern, S., Atkins, N., Boyle, D., Dharmawardena, K., and Wyborn, L.	Edited for clarity and updated images. Document deposited for DOI.

Acronyms

ADA	Australian Data Archive (ADA, n.d.).
AeRO	Australasian eResearch Organisation (AeRO, n.d.).
ARDC	Australian Research Data Commons (ARDC, n.d.-a).
AGLDWG	Australian Government Linked Data Working Group (AGLDWG, n.d.).
AVSIG	Australian Vocabulary Special Interest Group (AVSIG, ARDC, n.d.-e).
CARE	Collective Benefit, Authority to control, Responsibility, Ethics (Carroll et al., 2020).
CF Conventions	Climate and Forecasting Conventions (CF Metadata Conventions, n.d.).
CODATA	Committee on Data and the International Science Council (CODATA, n.d.).
DCAT	Data Catalog Vocabulary for the Web (Albertoni et al., 2024).
DOI	Digital Object Identifier (DOI Foundation, n.d.-a)
EOSC	European Open Science Cloud (European Commission, n.d.).
ESIP	Earth Science Information Partners (ESIPFed, n.d.-a).
FAIR	Findable, Accessible, Interoperable and Reusable (Wilkinson et al., 2016).
FHIR	Fast Healthcare Interoperability Resources (HL7, n.d.).
GUPRI	Globally unique persistent resolvable identifier (Schultes et al., 2022).
I-ADOPT	InteroperAble Descriptions of Observable Property Terminologies (Magagna et al., 2021, 2022)
I-ADOPT WG	InteroperAble Descriptions of Observable Property Terminology Working Group (RD-A I-ADOPT WG, n.d.).
IEEE	Institute of Electrical and Electronic Engineers (IEEE, n.d.).
IRI	Internationalised Resource Identifiers (Dürst and Suignard, 2005).
ISO	International Organisation for Standardisation (ISO n.d.-b).
JSON	JavaScript Object Notation (JSON, n.d.).
JSON.LD	JSON for Linking Data (JSON-LD, n.d.; Sporney et al., 2020).
KOS	Knowledge Organization Systems (Zeng, 2008).
LLMs	Large Language Models (Naveed et al., 2025)

LOV	Linked Open Vocabularies (LOV, n.d.).
NCRIS	National Collaborative Research Infrastructure Strategy (NCRIS). (Australian Government Department of Education, n.d.-b).
NDRI	National Digital Research Infrastructure (Australian Government Department of Education, n.d.-d).
NRI	National Research Infrastructure (Australian Government Department of Education, n.d.-a)
NRIAG	National Research Infrastructure Advisory Group (Australian Government Department of Education (n.d.-e).
OGC	Open Geospatial Consortium (OGC, n.d.).
OWL	Web Ontology Language (W3C OWL Working Group, 2012).
PID	Persistent Identifiers (CODATA RDM Terminology Working Group, 2024).
RD-A	Research Data Alliance (RD-A, n.d.).
RDA	Research Data Australia (ARDC, n.d.-f).
RDF	Resource Description Framework (Miller, 2001; RDF Working Group, 2014).
RDFa	RDF in Attributes (Herman et al., 2015).
RDFS	Resource Description Framework-Schema (Brickley and Guha, 2014).
RVA	Research Vocabularies Australia (ARDC, n.d.,-c).
SKOS	Simple Knowledge Organisation System (Isaac and Summers, 2009).
SNOWMED-CT	Systematised Nomenclature of Medicine Clinical Terms (Snowmed, n.d.).
SSSOM	Simple Standard for Sharing Ontological Mappings (Matentzoglu et al., 2022).
TERN	Terrestrial Ecosystem Research Network (TERN, n.d.).
URI	Uniform Resource Identifier (URI), formerly Universal Resource Identifier (Berners-Lee et al., 2005).
W3C	World Wide Web Consortium (W3C, n.d.-b).
WorldFAIR	WordFAIR - Global cooperation on FAIR data policy and practice (WorldFAIR, n.d.-a).
WorldFAIR+	WorldFAIR Plus (WorldFAIR, n.d.-b).
XML	eXtensible Markup Language (Bray et al., 2008).

Tables

Figures

1. Executive Summary

Knowledge representation artifacts, like controlled vocabularies, taxonomies, thesauri, ontologies, schemas, and standards, are crucial for organising, annotating, and sharing data in a way that makes it easy for systems to interoperate. For simplicity, this paper uses the term 'vocabulary' to refer to all these artifacts. Vocabularies are key to following the FAIR Guiding Principles for Scientific Data Management and Stewardship (FAIR Principle I2). They help describe data in a way that machines can read and act on, making this information accessible to users now and in the future. Vocabularies serve as essential reference points for both humans and machines, enabling effective data processing, integration, and modelling. In the age of AI, vocabularies support machine learning and artificial intelligence, helping us gain new insights and model complex systems. To make the most of our data, we need strong machine-readable metadata. Vocabularies are vital in this process as they help clarify the meaning of data, its connections, and its context.

Internationally, there is consensus across the research data landscape that to support the FAIR Principles, vocabularies must themselves be FAIR. They should be published with unique persistent identifiers, accompanied by rich metadata, and provided under standard Open licences where possible. Additionally, they should be available in community-agreed standard formats to maximise reuse. However, many vocabularies of interest to Australian researchers are not easily discoverable, reusable, or fit-for-purpose, and they do not fully align with FAIR Principles. In the Australian context, despite the importance of FAIR vocabularies in implementing the FAIR Data Principles and ensuring that data is sustainable and fit for purpose, their essential role in the data ecosystem is often overlooked. Enhancing the value of investments in research data requires improved tooling, support and governance for vocabularies.

> *This Strategic Community Roadmap outlines 57 recommendations and a strategy to enhance the Australian Vocabulary Ecosystem, which encompasses the people, resources, tools, platforms, policies, and practices that make vocabularies accessible and valuable for researchers and users of digital knowledge. The recommendations are organised under four strategic themes, emphasising their significance and the urgency for implementation.*

Whilst these recommendations may benefit parties in other countries, our primary focus is on the social and technical support required within the Australian research context to address data interoperability challenges effectively. The primary audience for this Community Roadmap includes individuals and organisations involved in the creation, management, and use of vocabularies and related services and tools within Australia's research and government sectors. This comprises those from the Australian Government's National Digital Research Infrastructure (NDRI), the National Collaborative Research Infrastructure Strategy (NCRIS) Facilities, National Research Infrastructure (NRI) Platforms, as well as Australian universities, research organisations, government agencies, researchers, and data scientists.

The recommendations and Community Roadmap outlined here were developed from a Vocabulary Workshop held at the Australian National University (ANU) from November 16 to 18, 2022, along with subsequent community consultation between 2022 to 2024. The objective of the 2022 workshop was to create a Roadmap that facilitates the transition of Australian vocabularies from their current state to a future state that fulfils next-generation requirements. The workshop brought together a diverse group of participants, including vocabulary creators, managers, service providers, tooling experts, and advocates from various fields such as health and medicine, Earth and environmental science, as

well as the humanities, arts, and social sciences. The recommendations presented reflect the broad range of experiences and perspectives of these participants, who have significant involvement in international data partnerships and a comprehensive understanding of the Australian research infrastructure landscape. Feedback from subsequent community consultations, particularly from the eResearch Australasia Conferences (2022-2024) and the Australasian eResearch Organisations (AeRO) Newsletter (AeRO, 2024), has been incorporated into these recommendations.

Many researchers and data managers create and utilise vocabularies and related resources in their work; however, these efforts often lack visibility. There are numerous challenges for those trying to discover, utilise, or repurpose the domain expertise documented in these vocabularies. Our goal is to enable any researcher to find the vocabularies that best represent the concepts they want to document, use them to plan data collection, and structure their data products and analyses — all without needing to be an expert in semantic technologies. Achieving this goal will require substantial investment in vocabulary tools and platforms, such as Research Vocabularies Australia (RVA), as well as the development of modules and plug-ins to simplify vocabulary use. Additionally, high-quality training materials and supportive social and governance frameworks will be necessary to manage the vocabulary lifecycle effectively.

Through the vocabulary workshop and subsequent community consultations, fifty-seven actionable recommendations were identified, focusing on seven key topics:
1. Technical standards adoption;
2. Vocabulary content creation;
3. Vocabulary publishing;
4. Vocabulary governance;
5. Mapping and annotation of vocabularies;
6. Finding and using FAIR vocabularies; and
7. Skills, training and outreach.

For implementation purposes, the recommendations were organised under four strategic themes and ranked by importance and urgency for implementation.
1. Foster a thriving vocabulary ecosystem;
2. Improve the consistency, content and context of vocabularies across Australia;
3. Improve the consistency of vocabulary tools and services; and
4. Lower barriers for using and maintaining vocabularies

Over the past 20 years, the Australian Government has made significant investments in the National Collaborative Research Infrastructure Strategy (NCRIS), leading to substantial advancements in research data management. These strategic roadmap recommendations for an Australian FAIR Vocabulary Ecosystem aim to build upon these advancements and promote greater reuse of research data, both within individual domains and across multiple fields. Additionally, they help address transdisciplinary challenges that cannot be effectively solved without the cross-domain interoperability of data.

2. Background

2.1. The Value of Vocabularies: Overview

Vocabularies are "organised arrangement[s] of words and phrases used to index content and/or to retrieve content through browsing or searching" (Harpring, 2013). Vocabularies provide structures for organising the knowledge of a scientific domain (e.g., the periodic table, social systems, the United Nations standard classifications, etc). Consequently, every well-managed vocabulary is a valuable component of our shared knowledge infrastructure. Vocabularies play a foundational role throughout the data-sharing landscape, both in Australia and internationally, and are used, often unnoticed, in most data assets, services, and infrastructure maintained by the research, government, and industry sectors. In this document, we use the word 'vocabularies' to broadly describe all 'semantic artifacts' that are machine-actionable (Wilkinson et al., 2016), machine interpretable (Elixir-Europe, n.d.) and machine-readable (CODATA RDM Terminology Working Group, 2024) formalisations of conceptual frameworks, including ontologies, taxonomies, glossaries, thesauri, metadata schemas, and standards (Hugo et al., 2020). Their use enables standard data representation, annotation, and knowledge organisation for robust data sharing and reuse by both humans and machines (Hugo et al., 2020; Corcho et al., 2024).

Vocabularies that are expressed using machine-actionable standards are fundamental to the 'FAIR Guiding Principles for Scientific Data Management and Stewardship' (Wilkinson et al., 2016). Vocabularies enhance data discovery and integration, facilitating cross-domain interoperability (WorldFAIR, n.d.-a). The FAIR Principles—Findable, Accessible, Interoperable, and Reusable—place particular emphasis on improving "the ability of machines to find and use the data automatically." They assert that all research objects should be FAIR for both machines and people.

FAIR Principle I2 states, "(Meta)data use vocabularies that follow FAIR principles" (Wilkinson et al., 2016). Therefore, for a dataset to be entirely FAIR compliant, any vocabularies used to describe the metadata and data must also be findable, accessible, interoperable, and reusable by both humans and machines (Wilkinson et al., 2016). Such vocabularies are a core part of cross-domain FAIR implementation, facilitating the potential for automated data integration in cross-domain research and enabling automated access control (Hodson and Gregory, 2024), ultimately allowing for better use of AI (Hodson and Gregory, 2024; ESIP Data Readiness Cluster, 2022). For example, vocabularies support knowledge graphs (Fensel et al., 2020) and applications that utilise graph technology, and in conjunction with Large Language Models (LLMs; Naveed et al., 2025), they enhance semantic understanding and knowledge processing.

There are numerous available best practice documents relating to creating and using vocabularies on the web (e.g., Hyland et al., 2014, Hugo et al. 2020, Garijo and Proveda-Villalón 2020; Cox et al. 2021, Lóscio et al., 2017). FAIR-compliant 'machine-actionable' vocabularies are created from information resources in hardcopy books, PDFs, and data sheets (.xls, .xlsx formats) using appropriate standard serialisations and formats (See Appendix 2 Recommendation 2.a.1). For uniqueness, persistence and re-use on the web the use of Globally Unique, Persistent and Resolvable Identifiers (GUPRIs: Wittenburg and Strawn, 2018; Schultes et al., 2022) for vocabularies and each term in the vocabulary is essential. As stated by Hugo et al. (2020): "Globally Unique, Persistent and Resolvable Identifiers must be used for Semantic Artefacts, their content and their versions" (P-Rec. 1, p21). Ideally, raw vocabulary data is readable by both humans (e.g., HTML documentation,

specification pages) and machines (RDF/XML, Turtle, JSON-LD), both resolvable from a single GUPRI such as an Internationalised Resource Identifier (IRI; Dürst and Suignard, 2005) (Rule 5 of Cox et al., 2021). For documentation and citation of the vocabulary as a whole (Rule 8 of Cox et al., 2021), assigning a separate GUPRI, such as a Digital Object Identifier (DOI), is useful (Rule 8 of Cox et al., 2021).

The complexity of the knowledge that can be encoded in a vocabulary varies along a 'semantic ladder' (e.g., McGuinness, 2023; Duerr et al., 2024; Figure 1 below), from controlled vocabularies, which tend not to include a structure of relations between their terms, through to ontologies which are shared and standardised list of words, terms and phrases to describe components of a particular discipline or domain, along with a taxonomy of their relations (Zeng, 2008; CODATA RDM Terminology Working Group, 2024).
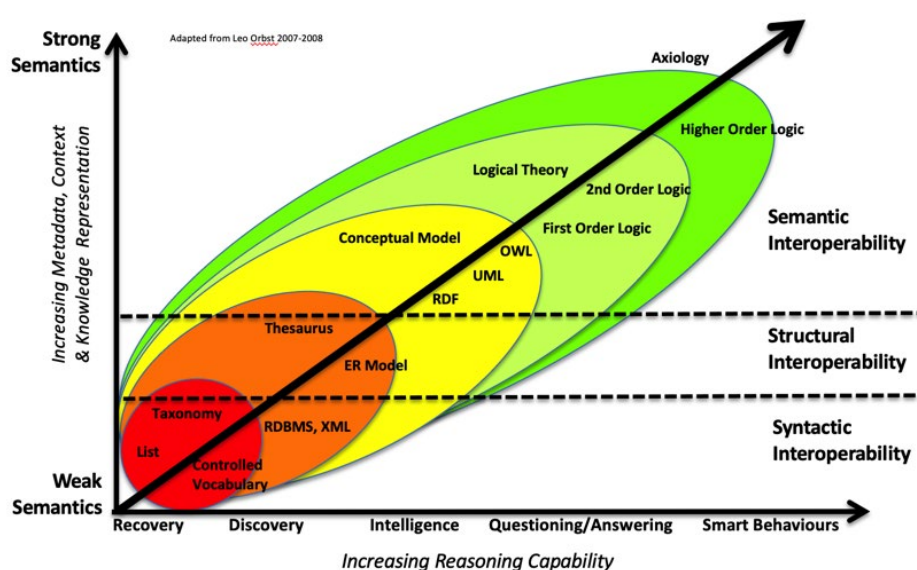
Figure 1. The Semantic Ladder
based on the Spectrum of Knowledge Representation and Reasoning Capabilities. Orbst, 2008. The arrow indicates increasing metadata, context and knowledge representation. See Glossary for definition of terms used in this figure (see also McCreary (2006) and Duerr et al. (2024)).
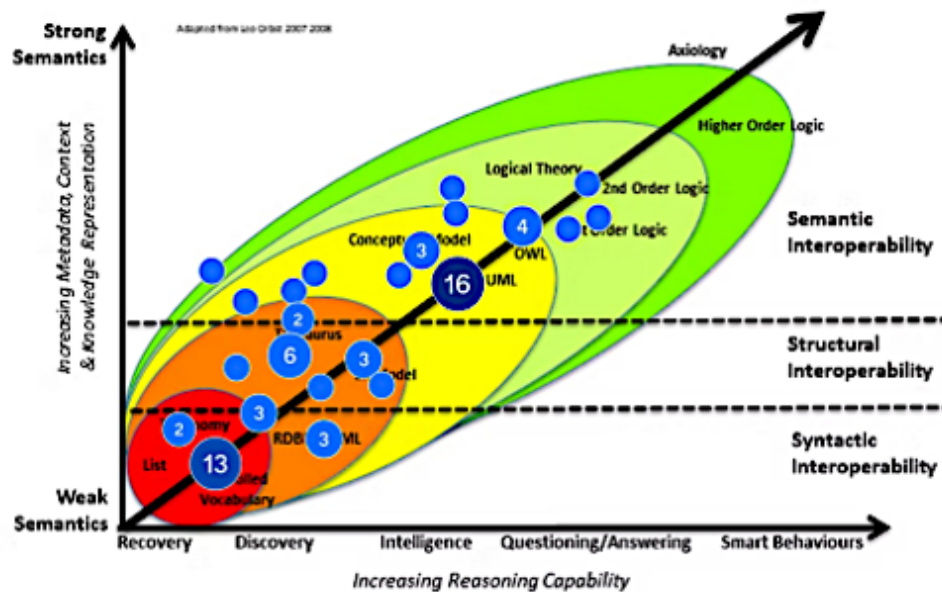
Figure 2. Semantic resources that are worked with, Australian Vocabulary Symposium 2022 responses as to the types of semantic resources participants worked with (Source: Figure 7 in Wong et al., 2023).

The 2022 Australian Vocabulary Symposium (Wong et al., 2023) featured a poll that asked participants about the vocabulary resources they were currently using (see Figure 2). The participants included vocabulary creators, managers, service providers, and advocates from various sectors, such as technology, research (including Federal, State, and Territory governments, organisations, and universities), national data infrastructures, consultancies, and service providers. Responses came from individuals working with a range of vocabulary types, from simple lists, taxonomies, and thesauri to more complex logics, such as ontologies. Figure 2 indicates there are diverse needs of parties in creating, managing, and reusing vocabularies that are purpose-driven and designed to be FAIR.

2.2. Vocabularies as enablers of interoperability

Data interoperability is a central focus in the development of research infrastructure. Interoperability cannot be significantly achieved without vocabularies, particularly in contexts where machines are involved in data processing. Data interoperability can be classified using different interoperability levels (Veltman, 2001; Goldbeck and Simperler, 2018):

1. **Structural and Syntactic Interoperability**: The first levels of interoperability level is achieved by adopting structural and syntactic standards that enable others to import and access the data. This involves, for example, using widely used encodings and well-known data formats, such as NetCDF or JSON (JSON; n.d.). Although both machines and human users can load and read this data, the meaning of the elements may be undefined or unclear, which can lead to potential misinterpretation.

2. **Semantic Interoperability**: The next level of interoperability involves the addition of semantic information to make the meaning of data understandable by both humans and machines. Semantic

Interoperability is the 'Ability of computer systems to transmit data with unambiguous, shared meaning. Semantic interoperability is a requirement to enable machine computable logic, inferencing, knowledge discovery, and data federation between information systems...in other words, what is sent is what is understood' (CODATA RDM Terminology Working Group, 2024). Vocabularies, along with supporting tools and services, are the means of achieving semantic interoperability.

Simple controlled vocabularies and lists (represented by the red and orange ellipses in Figure 1) allow users to recognise that multiple datasets reference the same concept, facilitating data integration. When hierarchical vocabularies and taxonomies are employed, simple reasoning based on set relationships becomes possible. Semantically annotated vocabularies and ontologies (depicted by the yellow and green ellipses in Figure 1) enable deeper integration and the ability to infer assertions that are not explicitly stated within a dataset. All these types of vocabulary serve as 'semantic glue', enhancing the broader application and reusability of datasets.

3. **Cross-domain interoperability**: Cross-domain interoperability leverages semantic interoperability, utilising vocabularies and semantic technologies to connect different terminologies across various domains. Cross-domain interoperability is enriched when mappings between disparate terms are documented and published in machine-readable formats (Goldbeck and Simperler, 2018; Figure 3). New projects are emerging that are focused explicitly on cross-domain interoperability, for example, the CODATA (Committee on Data and the International Science Council) and RD-A-led WorldFAIR: Global cooperation on FAIR data policy and practice project, which worked with a set of 11 disciplinary and cross-disciplinary case studies to improve interoperability and reusability of digital research objects (WorldFAIR, n.d.-a). The European Open Science Cloud (EOSC) is another example (European Commission, n.d.). Fundamental to cross-domain interoperability is the leverage of universal vocabularies that span across multiple domains, such as geography, units of measure and the periodic table.
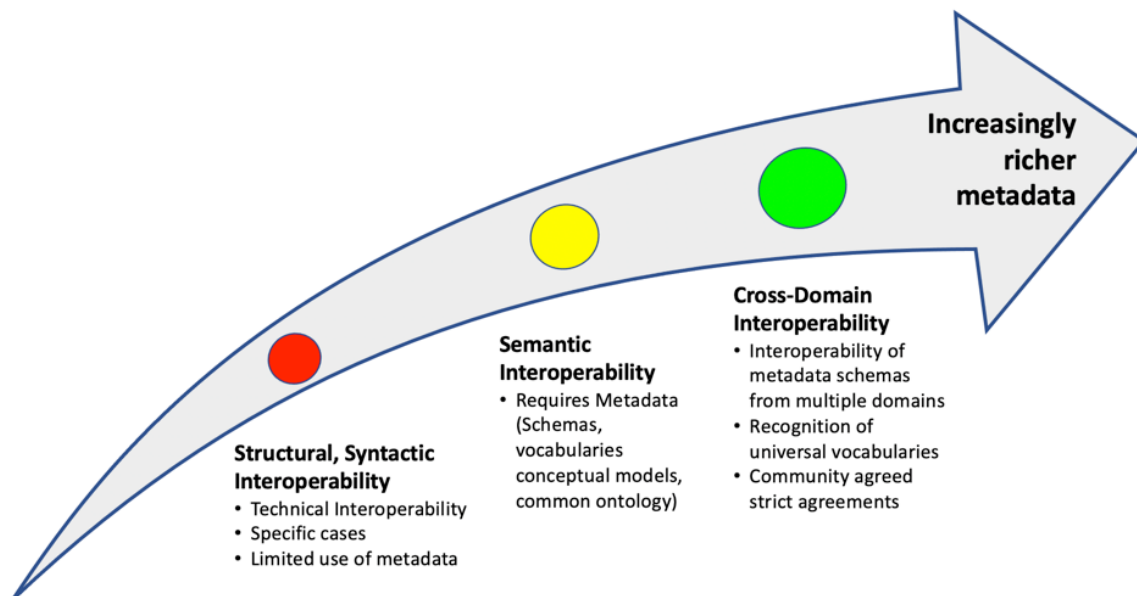
Figure 3. Types of interoperability
with increasing levels of interoperability achieved within and across domains. Richer metadata allows terms within and across datasets, as well as their relationships, to be understood by both humans and machines (based on Goldbeck and Simperler, 2018).

2.3. The Vocabulary Ecosystem

With the increasing number and complexity of vocabularies and the growing requirements for human and machine interoperability, vocabularies need to be understood as part of a broad ecosystem that includes governance models, parties, and technical and social systems.

Baker et al. (2013a) used the term 'Vocabulary Ecosystem' to encompass the links and dependencies between linked open vocabularies, as well as the global governance of the ecosystem in which vocabulary managers, standard bodies, and memory institutions should engage. Tamma and Poveda-Villalón (2021), in Boissier et al. (2021), more precisely defined a Vocabulary Ecosystem as "1) the minimum vocabulary or vocabularies needed, and 2) the possible relations amongst them to support discovery, access and interoperability amongst autonomous agents operating in the Web of Things (WoT)". Robust vocabulary governance is an essential component of "a more interconnected and sustainable data ecosystem" (Ramezani et al., 2024 p.8, Baker et al., 2013a). This includes the need for good governance and processes in the catalogues that preserve and maintain vocabularies (Corcho et al., 2024). The CARE Principles for Indigenous Data Governance also raise important governance considerations for innovation and Indigenous self-determination through the modern data ecosystem (Carroll et al., 2020).

For this Community Roadmap, we define the Vocabulary Ecosystem as the parties, resources, tools, platforms, policies and practices that together make vocabularies accessible and useful for all researchers and users of digitally encoded knowledge. No matter how valuable vocabularies may be as formalisations of expert knowledge, their value will not be realised unless all parties can discover and apply this encoded knowledge to their data. If best practice is not well understood, or if adoption

of vocabularies requires significant technical knowledge, most data will continue to be generated without the interoperability benefits provided by vocabulary use, and large-scale data integration, for example, for meta-analysis and time-series studies, will be costly activities.

A well-framed and properly resourced Vocabulary Ecosystem will facilitate the mainstreaming of encoding expert knowledge as vocabularies, maximise vocabulary reuse and application, and ultimately improve data quality and lower the transaction cost of data reuse for end users.

2.4. Vocabulary Ecosystem challenges: the Australian context

The National Digital Research Infrastructure Strategy (NDRI), released in July 2024, provides a user-centric vision for Australia's digital research infrastructure (Australian Government Department of Education, n.d.-d.) and lists six Outcomes. Outcome 3 states that by 2030, Australia's NDRI system should be "Consistent in its standards for data collection, curation, and access". The identified challenge for Outcome 3 is that "Increasing volumes of research data are being generated that are not FAIR/CARE compliant". Specific to vocabularies, FAIR Principle I2 states that for data to be interoperable, "(meta)data use vocabularies that follow FAIR Principles" (Wilkinson et al., 2016).

However, the reality in Australia is that shared vocabularies and services supporting them form a diverse ecosystem that is not effectively coordinated, meaning that many potentially valuable vocabularies are not discoverable or reusable (McEachern et al., 2023). In addition, existing vocabularies are not necessarily findable, let alone accessible, interoperable and reusable; that is, very few vocabularies comply with FAIR Principle I2. Tools and processes for vocabulary creation and maintenance can be esoteric and off-putting to data stewards and generators (such as researchers).

Recognising these challenges, a cross-domain workshop was held in person at the Australian National University (ANU) from 16-18 November 2022 (Wong et al., 2023). The workshop focused on identifying and addressing key issues and barriers to promote movement from the current state of Australian vocabularies to a future state that meets next-generation vocabulary requirements. From this workshop, the necessity for a Strategic Community Roadmap for an Australian FAIR Vocabulary Ecosystem was acknowledged.

2.5. Purpose of the Strategic Community Roadmap: Call to Action

This Community Roadmap outlines a strategic path forward to create an integrated FAIR Vocabulary Ecosystem in Australia. It was developed through a series of community engagements primarily within the Australian research sector and government from 2021 to 2024. It combines vocabulary-related infrastructure, services, tools, and practices within the Australian National Research Infrastructure (NRI) to enhance the development, maintenance, and use of vocabularies.

It is a **Call to Action** for NRIs and the Australian research sector, including both academic institutions and government bodies, to work towards a shared Vision and Mission (Section 4.2) for a FAIR Vocabulary Ecosystem. This requires coordination, investment, and significant efforts to reinforce and expand initial endeavours. The roadmap highlights specific, actionable recommendations that

parties can adopt, both individually and collectively, to overcome the critical technical and social barriers to help realise this Vision.

The Purpose of this Community Roadmap is to promote a more coordinated approach to overcoming key technical and social barriers towards establishing a FAIR Vocabulary Ecosystem in Australia.

The Aim is to facilitate wider adoption and broader community support for machine-readable and machine-actionable vocabularies, as a fundamental aspect of meeting the FAIR Data Principles.

The Goal is to make this process straightforward, clearly beneficial, and mainstream. Additionally, it seeks to raise awareness of the importance of vocabularies in realising the advantages of FAIR data. It presents improvements in the National Research Infrastructure (NRI; Australian Government Department of Education, n.d.-a) landscape that would support the creation and use of such data.

2.6. Audience

The primary audience for this Community Roadmap includes those involved in creating, managing, and using vocabularies within Australia's research and government sectors. Specifically, the Community Roadmap is aimed at:

- **NDRI Strategy Coordinators and Implementers**: The NDRI is coordinated through the National Research Infrastructure Advisory Group (NRAIG, Australian Government Department of Education, n.d.-e).

- **NCRIS and NRI Platforms**: Decision-makers and technical teams responsible for developing and maintaining Australia's NCRIS (National Collaborative Research Infrastructure Strategy: Australian Government Department of Education, n.d.-b) and related NRI platforms (current funded projects are listed in the Australian Department of Education n.d.-c);

- **Universities**: Academic institutions that can benefit from and contribute to the uptake of improved vocabulary practices and infrastructure.

- **Researchers and Data Scientists**: Individuals in academic, government, and non-governmental organisations (NGOs) who are directly involved in managing, curating, and using high-quality, reusable data assets. This includes those who are "hands-on" with data and metadata and who would benefit from enhanced vocabulary services and tools. Ready access to high-quality and widely adopted vocabularies will help researchers understand the structure of each research domain and facilitate communication using a shared conceptual framework.

- **Government Agencies**: Various government departments, such as the Department of Agriculture, Fisheries and Forestry (DAFF), the Department of Climate Change, Energy, the Environment and Water (DCCEEW), Geoscience Australia (GA), as well as State and Territory Geological Surveys and Geospatial agencies, have a vested interest in managing data and vocabularies for policymaking, reporting, and research purposes. Additionally, we include publicly funded research agencies, including the CSIRO.

Although industry and the public will benefit from the services and infrastructure improvements mentioned in this Roadmap, they are not the primary audience for this document.

3. Methodology of Roadmap development

This Community Roadmap emerged from the November 2022 Vocabulary Workshop (Wong et al., 2023), which focused on vocabularies that were machine-actionable (FAIR) and well-governed and built on two preceding cross-domain Vocabulary Symposia held in 2021 (Brownlee et al., 2021) and 2022 (Wong et al., 2023). The organisational leads for the 2022 Workshop were the Australian Data Archive (ADA; ADA, n.d.), the Australian Research Data Commons (ARDC; ARDC, n.d.-a) and the Committee on Data and the International Science Council (CODATA; CODATA n.d.).

Invited participants at the 2022 Workshop included vocabulary creators, managers, service/tooling providers or advocates who worked either within or across the domains in technology and service provision (including geospatial). They were from agencies and organisations across Research Agencies (Including Universities, Federal, State, and Territory governments), National Data Infrastructures, consultancies, and service providers. Focus was placed upon domains covered by the eleven case studies of the CODATA WorldFAIR Project (WorldFAIR, n.d.-a) and the three ARDC Thematic Research Data Commons Initiatives (ARDC, n.d.-b): People (health and medicine); Planet (Earth and environmental science); and HASS (humanities, arts, social science) and Indigenous.

On Day 1, participants at the 2022 Vocabulary Workshop were given a cross-domain problem and discussed components of the Vocabulary Ecosystem required to solve this problem. By Day 2 of the Workshop, broad issues had been identified, and separate cross-domain breakout groups were formed to brainstorm core challenges that could be addressed through a Community Roadmap (p15, Wong et al., 2023). From the discussions and activities, participants expressed interest in developing a Community Roadmap for the future state of the Australian Vocabulary Ecosystem. A Roadmap outline was drafted. A cross-domain working group was nominated to develop the roadmap further.

During the writing process after the workshop, community consultations were regularly held through three Birds of a Feather sessions (BoFs) at the annual eResearch Australasia Conference (McEachern et al., 2022; 2023; Boyle et al., 2024), at the 2023 Vocabulary Symposium (Australian Research Data Commons, 2023) and through the November 2024 eResearch Newsletter (AeRO, 2024).

Here, we present the output of these community efforts. Firstly, a Vision and a Mission statement are proposed to guide the actionable recommendations. Recommendations were then gathered and synthesised from across the consultative activities of 2022 to 2024. The recommendations were built into an Implementation Roadmap. To guide strategic investment and decision-making, as well as practical implementation actions, recommendations were grouped under seven key topics. For implementation purposes, the recommendations were organised into four Strategic Themes and ranked by suggested importance and urgency for implementation. Each recommendation has been assigned a prioritisation for implementation (high, medium, or low) and the timeframe within which it should be implemented (short-term, 1-2 years; medium-term, 2-3 years; or long-term, 3-5 years).

A note on policy, guidelines and related documents
In the 'Mapping the Landscape' session of the 2022 Vocabulary Workshop, participants developed a comprehensive list of sources of information that could influence the priorities for implementation, including documents, legislation, guidelines, groups, reviews, etc. (Brownlee and Wyborn, 2023; Appendix 2 in Wong et al., 2023). Participants felt that all the documents, initiatives, and/or directives identified generally supported, rather than constrained, the creation and use of vocabularies to enable machine-actionable FAIR data. Some may also inform how vocabularies and

related services are governed and managed to allow for best practice of equity and transparency, such as Codes of Ethics and Government Legislation.

Hence, specific policy and guidelines are not referenced in the roadmap. Although they are recognised as essential parts of the Vocabulary Ecosystem, their relevance will vary across parties and the vocabulary lifecycle and maturity. For example, the Australian Government Data Availability and Transparency Act (Australian Government, n.d.) and the Data and Digital Government Strategy: The Data and Digital Vision for a world-class APS to 2030 (Australian Government, 2023) are of relevance to Australian Government agencies, whilst the Australian Code for the Responsible Conduct of Research, in particular the Guide on Management and Data Research (NHMRC, 2018; 2019) and related institutional or organisational Data Management Plans are mainly relevant to individual researchers and Institutions conducting research in Australia or under the auspices of Australian Institutions.

4. Recommendations for a FAIR Vocabulary Ecosystem

4.1. The Vocabulary Ecosystem: Identified Components

Based on the provided definition (Section 2.3) and recurring themes identified during the consultations for developing this Roadmap, we define the Vocabulary Ecosystem as consisting of several elements and their interdependencies (Figure 4):

• The **Content** of the vocabularies. Representation of subject matter knowledge by domain experts, authorities, communities and knowledge-brokers;

• The **Community** that uses, delivers or influences vocabularies. These are the participants and influencers in the vocabulary ecosystem, and the relationships between them, including funders, domain experts, informaticians and data engineers, and users, including researchers and data stewards.

• **Standards** for publishing FAIR vocabularies, e.g., RDF, OWL, using XML, Turtle, JSON-LD (for the representation and encoding of the vocabularies);

• The **Governance** supporting the persistence, trust, currency and authority of vocabularies. Agreed community frameworks, best practices, and rules for developing, managing, using and citing vocabularies;

• The **Policy and Investment Context** driving vocabulary use. Investments and policy drivers such as the roadmap priorities of the National Research Infrastructure (Australian Government Department of Education, n.d.-d.) and the rules that are managed via institutional guidelines and policies through data management plans;

• **Infrastructure, tools and technologies** for the vocabulary life cycle, including creation, publishing, preservation and maintenance, and facilitating discovery, access, use and re-use.

These tools encompass the services for storing and/or sharing vocabularies such as repositories, catalogues, registries, libraries, and terminology/vocabulary servers, and associated tools and technologies, including editors, client applications, plug-ins, and APIs. Herein, we refer to

vocabulary services as a service that enables users to discover, browse, and download controlled vocabularies in a variety of languages (e.g., SKOS, RDF/XML, Turtle, RDF/JSON, Notation3). The service may also contain an editor, where authorised users create, manage, and translate the vocabularies (CESSDA, n.d.). These play a critical role in preserving and maintaining vocabularies, as well as addressing interoperability challenges (Corcho et al., 2024).

Other **Socio-technical drivers** of vocabularies include, for example, intrinsic motivations for knowledge sharing and the rise of AI.

Skills uplift and training materials for vocabularies in research and data management, and in support of the FAIR data principles, including outreach activities to support vocabulary creation, management and use.

Community participants and influencers within the ecosystem, along with their priorities and motivations, play crucial roles in the creation, management, dissemination, and reuse of vocabularies. This concept was highlighted during the cross-domain symposia on vocabulary governance held in 2021 (Brownlee et al., 2021) and in the "Mapping the Landscape" activity, which articulates all stakeholders involved in the Australian Vocabulary Ecosystem (p. 12, Wong et al., 2023). The interdependencies of stakeholder relationships throughout the vocabulary supply chain were emphasised.

In the context of this Community Roadmap, it is essential to recognise that any implementation activities must be tailored to the subcommunities within the Australian research community, each possessing different perspectives, priorities, needs, and skill sets. For instance:

- **Domain experts** and **knowledge holders (subject matter experts)** need collaborative tools for developing accessible vocabularies, ensuring their expertise is appropriately accessible and valuable to others.

- **Vocabulary service providers** and **informaticians** require clear guidance on standards, best practices, and futureproofing to deliver vocabularies, vocabulary services and tooling that are fit-for-purpose for end users; and

- **Users** (who may also be **knowledge holders**) require intuitive and straightforward ways to access and utilise the most appropriate and well-supported vocabularies for their data management and analyses. This includes researchers, who should not have to master semantic technologies to apply vocabularies to their research data effectively.
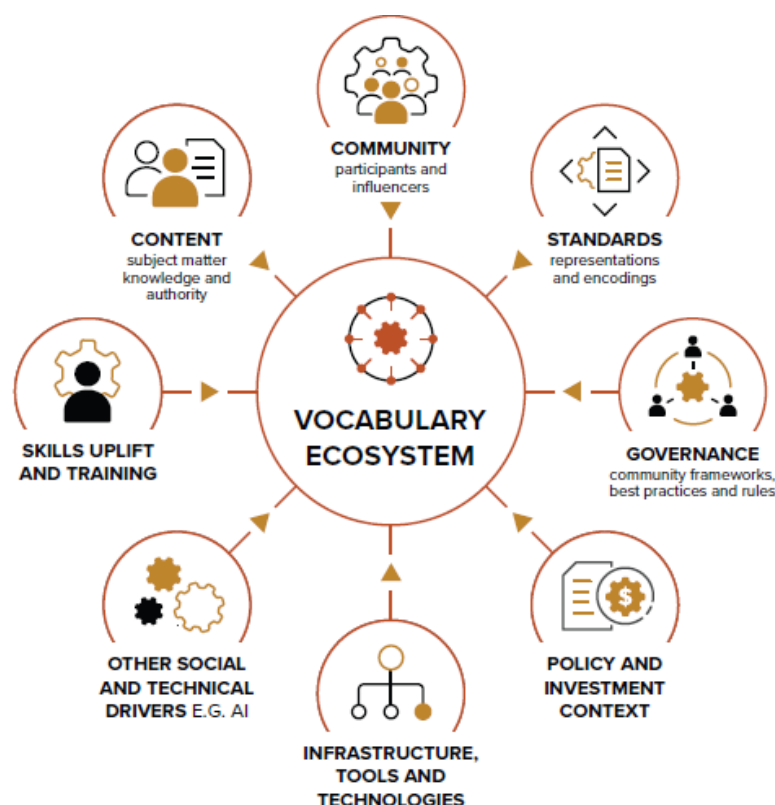
Figure 4. The elements of the Vocabulary Ecosystem

4.2. Vision and Mission

Shaped through the activities (a workshop, symposia, community consultations) across 2022 - 2024, the following Vision and Mission statements are proposed. The purpose of articulating these is to help foster a shared understanding of the stakeholder community's goals within the Australian Vocabulary Ecosystem. The Vision and Mission can be adopted and adapted at the level of individual organisations or collectively across several organisations or research groups.

The following **Vision** is proposed:

That shared vocabularies and their ecosystem form a key part of the Australian and international data landscape, enabling the impact of data for beneficial cultural, environmental, societal, economic and health outcomes.

The Outcome of this Vision would be:

In the next 5 to 10 years, the importance of vocabularies will be widely recognised in Australia. This will lead to a growing ecosystem of vocabularies, along with associated services and tools, that will make it easier to publish and use vocabularies. These vocabularies will serve as a fundamental component of data assets, services, and infrastructure across the research, government, and broader data sectors.

To achieve this Vision, the following **Mission** statement is proposed:

To further the consistency and sustainability in the development, implementation, governance and use of vocabularies across domains to solve real-world problems, both in Australia and internationally, through the establishment of an effectively coordinated community and services.

A group coordinating communications across the Vocabulary Ecosystem may facilitate working toward a shared Vision and Mission across organisations, institutes, communities, facilities, and service providers.

4.3. Recommendations and Strategic Priorities

The 2022 Vocabulary Workshop (Wong et al., 2023) and subsequent community consultations between 2021 - 2024 identified 57 Recommendations.

The 57 recommendations are categorised into seven Key Topics:

a)   Technical standards adoption;
b)   Vocabulary content creation;
c)   Vocabulary publishing;
d)   Vocabulary governance;
e)   Mapping and annotation of vocabularies;
f)   Finding and using FAIR vocabularies; and
g)   Skills, training and outreach.

To support strategic implementation, we have further organised the 57 recommendations around four Strategic Themes (Table 1).

1.   Foster a thriving vocabulary ecosystem;
2.   Improve the consistency, content and context of vocabularies across Australia;
3.   Improve the consistency of vocabulary tools and services; and
4.   Lower barriers for using and maintaining vocabularies.

Table 1. Description of the Four Strategic Themes

Strategic theme	Description
1. Foster a thriving Vocabulary Ecosystem	The development of a robust Vocabulary Ecosystem depends on creating a thriving network of interconnected parties and vocabulary tools and services. This strategic theme prioritises collaboration and community building, sustainability and longevity, demonstrating value, well-defined roles and responsibilities, and effective communication and knowledge sharing, supporting a thriving ecosystem. This theme aims to encourage the widespread adoption of FAIR vocabularies for data sharing and integration.
2. Improve the consistency, content and context of vocabularies across Australia	To ensure the effective use of vocabularies in data management and integration, it is essential to enhance their consistency, content, and context. This strategic theme focuses on promoting the development of high-quality vocabularies that are aligned with the needs of diverse

	stakeholders, emphasising governance, authority, inclusivity and representation. By improving vocabulary consistency, content, and context, this theme aims to reduce duplication, promote interoperability, and facilitate appropriate data integration and reuse, in line with the FAIR and CARE principles.
3. Improve the consistency of vocabulary tools and services	The consistency of vocabulary tools and services is essential for encouraging the widespread adoption of standardised vocabularies. This strategic theme focuses on developing interoperable and standardised vocabulary tools and services to enable seamless integration, facilitate data sharing, and promote discovery, accessibility and reuse. By fostering consistency in these vocabulary tools and services, this theme aims to enhance data quality, support the FAIR data principles, and establish a cohesive vocabulary ecosystem.
4. Lower barriers for using and maintaining vocabularies	To encourage the widespread use of vocabularies, it is crucial to simplify their usage and maintenance. This strategic theme aims to reduce the complexity involved in managing vocabularies, making them more accessible and user-friendly to a broader audience. By lowering barriers, this theme aims to promote the reuse and adoption of vocabularies, thereby supporting the FAIR principles and enhancing data interoperability.

The 57 recommendations are presented in the tables below, organised by Strategic Theme. Further details on each recommendation can be found in Appendix 1.

Each recommendation is coded, as an example:
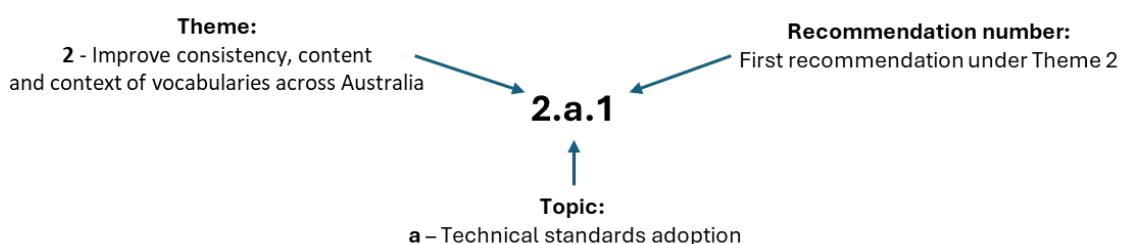
a) Technical standards adoption

i.   Recommendations

Number	Recommendations
2.a.1	Promote the use of well-governed and established, internationally endorsed technical standards

ii.   Rationale

To tackle the challenges of data interoperability, vocabularies need to be both machine-actionable and machine-readable. This allows for the sharing and reuse of concepts and knowledge by both

humans and machines. To achieve this, it is essential to implement the appropriate technical standards across the Australian Vocabulary Ecosystem. This includes encoding vocabularies in standard representation languages and formats, as well as supporting technical standards through vocabulary services and tools. Ensuring the technical accessibility of vocabularies is vital for promoting widespread adoption and integration into various data applications, products, and platforms. Whenever possible, it is important to utilise well-governed and internationally endorsed technical standards throughout the Australian Vocabulary Ecosystem.

b) Vocabulary content creation

i.   Recommendations

Number	Recommendations
2.b.1	Reduce vocabulary duplication by facilitating the search for existing resources, including across vocabulary services.
2.b.2	Allow vocabularies to be created at the local, regional, national and international levels (including at the project level if required), recognising that those at the global level can have a broad impact and utility but that these can take considerable coordination effort.
2.b.3	Support collaborations to extend or revise existing vocabularies, including work at national and international levels to merge semantically equivalent vocabularies.
2.b.4	Support creation of multilingual vocabulary content, including Indigenous languages and dialects.
2.b.5	Support vocabulary creation and management workflows that respect the role of knowledge holders and domain experts (e.g., communities, committees, bodies or individuals) in maintaining the content of vocabularies as authoritative.
2.b.6	Encourage the use of best practice standards in the 'profiling' of existing vocabularies.
3.b.1	Provide access to registers and services that handle the full spectrum of knowledge encoding, from simple controlled vocabularies to more expressive ontologies, encoded using the appropriate technical standards and standard schemas as applicable (e.g., Data Catalogue Vocabulary, DCAT).
3.b.2	Provide options for utilising open-source collaborative platforms such as GitHub for vocabulary creation and management.

ii.   Rationale

Recommendations for vocabulary content creation focus on the key topics of creating a more **interoperable, collaborative, and inclusive system for managing vocabularies**. The rationale of these recommendations is to make it easier for people to find, use, and contribute to vocabularies in a way that respects expertise and promotes wider use. The main themes are *Minimizing Duplication and Encouraging Reuse* making it easier to search for existing vocabularies and to collaborate on extending or merging them; *Promoting Collaboration and Inclusivity* emphasising supporting collaboration at all levels—local, national, and international; *Respecting Expertise and Authority* ensuring individuals, communities, or committees that hold expertise and knowledge have a central role in creating and maintaining the content of the vocabularies; and *Adopting Modern and Flexible Tools* using standards to encode different types of knowledge (from simple lists to complex ontologies).

Vocabularies are critical to ensuring shared understanding of any term used in data. The absence of controlled vocabularies in the terms used in metadata elements and/or data descriptions is often the main reason for the lack of machine-to-machine interoperability between datasets, as machines cannot easily understand text-based content in the same way that the human brain can decode, interpret and make sense of the free text (Wang et al., 2017). However, there is an unnecessary proliferation of vocabularies, many of which are developed in isolation. Facilitating the search for existing vocabularies across vocabulary services will help maximise re-use of existing vocabularies, reducing duplicity. There also needs to be a more precise articulation of roles and responsibilities within vocabulary metadata, as well as through creation and management workflows. This includes the identification of the governing authority of the vocabulary's content. Such information will enable users to evaluate which vocabulary is fit for their purposes, including considering which is more authoritative and likely to persist over time.

c) Vocabulary publishing

i.   Recommendations

Number	Recommendations
2.c.1	Offer tools that enable easy publication of a metadata record for each vocabulary, ideally with a DOI, including metadata on content and scope, governance and maintenance.
2.c.2	Enable vocabulary owners to express the licence and copyright for each vocabulary.
2.c.3	Vocabulary services should support the currency of vocabularies and mapping by enabling the easy management and curation by authorised parties, through community-driven and automated approaches.
3.c.1	Provide pathways to easily, efficiently and persistently publish a vocabulary.
3.c.2	Promote good practice in GUPRI allocation by using dedicated, persistent identifier services. Consider integrating the harvesting of GUPRIs through vocabulary tools and services.
3.c.3	Vocabulary services and registries must support version control and version management, including metadata and tools to assist mapping between versions.
3.c.4	Vocabulary services must offer the ability to notify users of changes, such as deprecating and replacing a term.
4.c.1	The ability to represent concepts in a form that accommodates the semantic complexity of the knowledge, and delivers it in a manner that is no more complicated than necessary for the use.

ii.   Rationale

A thriving vocabulary ecosystem requires that vocabularies are published in a manner that is both accessible and sustainable. Without streamlined processes for easy, efficient, and persistent publication, vocabularies risk remaining fragmented, hindering their utility and adoption.

The rationale behind the recommendations for the vocabulary publishing key topic is to ensure that vocabularies are **discoverable, authoritative, and sustainable** over time. They focus on making vocabularies easy to find and use while ensuring their integrity and long-term viability. Raised were recommendations for *Trust and Reliability* achieved through transparent governance, version control and the use of persistent identifiers; *Accessibility and Management* making vocabularies easy to publish, manage and re-use; *Flexibility and Adaptability* allowing vocabularies to evolve, such as

through version management; and *User Notification and Currency* highlighting the importance of keeping users informed.

d) Vocabulary governance

i.   Recommendations

Number	Recommendations
1.d.1	Support and advocate for persistent and sustainable vocabulary registers and services, including resourcing and contingency planning.
1.d.2	Reduce the inertia experienced in vocabulary creation and maintenance (e.g., no one wanting to 'own' the vocabulary) by providing clear examples of roles, responsibilities and resources required.
2.d.1	Engage with and respect the critical role of subject domain experts, committees or bodies in endorsing and governing authoritative vocabularies.
2.d.2	Support Indigenous-led scoping and activity for vocabularies of, by and for Australian Aboriginal and Torres Strait Islanders, including for use in decolonising data, promoting the expression of CARE through the vocabulary lifecycle (including in vocabulary metadata).
2.d.3	Promote authoritatively governed vocabularies within vocabulary services and registers.
2.d.4	Acknowledge that vocabularies are not neutral and support the right of reply, including in vocabulary notes, metadata and mappings.
2.d.5	The location of a vocabulary can change over time. Consider the responsible parties and resources required for ensuring that vocabulary GUPRIs consistently resolve to webpages containing the vocabulary information and metadata.
3.d.1	Understand and promote the expression of vocabulary governance in vocabulary services, tooling and workflows.
4.d.1	Document and share scenarios of good practice and protocols for GUPRI construction, allocation and endorsement.
4.d.2	Provide guidance and reference examples on fit-for-purpose vocabulary governance - including for creation, maintenance, versioning, authorisation, sustainability and persistence.

ii.   Rationale

The recommendations under the key topic of vocabulary governance are about building a **sustainable and respectful** Vocabulary Ecosystem. The rationale is to ensure that vocabularies and related tools, infrastructure and services are not only technically sound, but deliver authoritative, up-to-date content that is ethically governed and culturally aware. Raised were recommendations around *Sustainability and Ownership*; *Respectful and Authoritative Governance, Transparency and Trust*, and *Documenting and Sharing Best Practices*.

The FAIR Principles include the notion of "long-term care" of data, including vocabularies that are used to represent and annotate data in a standard way (Ramezani et al., 2024). Vocabularies that support meaningful knowledge representation through the FAIR data ecosystem are "particularly sensitive to changes, versioning, sharing, management, all of which are governance aspects" (Ramezani et al., 2024, p9). Therefore, robust vocabulary governance is an essential component of an interconnected, sustainable data ecosystem (Ramezani et al., 2024; Baker et al., 2013a). Semantic artifact governance (herein 'vocabulary governance') can be described as:

"A principled approach for standardising different aspects of semantic artifacts development through the semantic artifact lifecycle, from acquisition to use to disposal. It specifies decision rights and accountabilities for an organisation's decision making about semantic artefact" (Ramezani et al., 2024 p.12).

Governance requirements for vocabularies vary with maturity and user base, as well as between specific use cases, research communities, and projects. In addition to the recommendations identified here, the 2022 Workshop identified key concerns relating to the governance of FAIR vocabularies (Wong et al., 2023). This included the requirements to understand and communicate scope, context, roles, and responsibilities, manage change requests and communicate revisions, and consider matters of persistence and sustainability, including vocabulary services. Addressing the gap in the maturity and the skills of vocabulary governance across the Australian Vocabulary Ecosystem is necessary to empower parties to establish robust governance structures, ensuring that vocabularies are a trusted and persistent resource for diverse parties. There is a need for practical guidance and examples on governance patterns that can be adopted and adapted for various scenarios. One such resource in the context of the European Open Science Cloud (EOSC) is the models of vocabulary governance based on the diverse needs of different scientific communities and projects documented by Ramezani et al. (2023, 2024).

e) Mapping and annotation of vocabularies

i.   Recommendations

Number	Recommendations
2.e.1	Vocabulary services should include tools to provide additional text representations or labels for existing vocabulary terms.
3.e.1	Vocabulary services should include tools for providers and the community to define and publish mappings that establish the relationship between pairs of vocabulary terms using SSSOM, SKOS, OWL, FHIR or other well-established technical standards.
3.e.2	Vocabulary services, or specialist mapping registries, should provide stable, persistent and FAIR hosting for these mappings and their metadata (e.g., provenance and authority).
3.e.3	Vocabulary services and registries should provide access to mappings using a standardised set of formats and representations, as well as via APIs and software packages that facilitate their use.
3.e.4	Vocabulary services and registries should enable the discovery of known mappings that reference each vocabulary.
4.e.1	Vocabulary services or other associated software should facilitate semi-automated creation of (candidate) mappings based on term similarity, other mappings, etc., and practitioners should have the ability to edit and curate these candidate mappings.
4.e.2	Consider the feasibility of AI for creating, mapping and annotating vocabularies.

ii.   Rationale

Vocabularies are key building blocks for knowledge management, but additional effort is necessary to build mappings between vocabularies so applications can use them to aggregate, integrate and reason over data.

The rationale behind recommendations for the key topic on mapping and annotation of vocabulary is to support **dynamic, interconnected, and intelligent** vocabularies. They focus on making vocabularies more useful by linking them together and by leveraging automation to do so more efficiently. Raised were recommendations for *Interoperability Through Mappings*; the *Enrichment of Vocabulary Content*; *Discovery and Accessibility of mappings*; and *Making the Creation and Maintenance of Vocabularies and their Mappings Easier* through Automated and Semi-Automated Tooling.

The ability to contextualise vocabularies into a broader ecosystem supports several important use cases, including:

- A network of partners all collect data using related but different vocabularies and want to improve accessibility for their data by identifying how terms in different vocabularies relate to each other.

- Researchers wish to map vocabulary terms to domain ontologies so they can leverage the semantic power of these ontologies;

- A local network has developed a functional vocabulary to meet its needs, but a subsequent international standard vocabulary now exists, and the network wishes to document how the two vocabularies are related; and

- Communities wish to annotate vocabulary terms with known variant text representations (abbreviations, additional languages, misspellings, etc.).

All these use cases depend on the ability of researchers or networks to create sets of annotations for existing vocabulary terms, even in cases in which they have no rights to modify the original vocabulary, and to publish these in a form that enhances the usefulness and interoperability of the vocabulary and of data using the vocabulary.

f) Finding and using FAIR vocabularies

i.   Recommendations

Number	Recommendations
1.f.1	Understand the emergent role of vocabularies in AI optimisation and ethical data re-use, including in automated data tagging.
3.f.1	Support standard API usage, including the development of vocabulary APIs with standard and extensible search mechanisms.
3.f.2	Provide human and machine-readable views of vocabularies, of varying complexities, to standard formats (content negotiation by profile).
3.f.3	Vocabulary services should facilitate re-use (forking) of vocabularies where this is necessary for technical or governance reasons, including establishing links to the terms in the source vocabulary.
4.f.1	Enable users (both humans and machines) to easily find, access and reuse vocabularies (ie FAIR vocabularies).
4.f.2	Ensure vocabularies are appropriately licensed to enable and allow reuse.
4.f.3	Allow users to reference and cite vocabularies appropriately attributing curators and contributors.
4.f.4	Encourage interoperability between vocabulary services and registries where it will support vocabulary discovery and re-use within or across domains.

4.f.5	Vocabulary services and registries should deliver information in a way that is user-friendly to search, discover and access the granularity of information they require.
4.f.6	Vocabulary services and registries should include tools and APIs that allow vocabularies to be accessed and used directly within popular programming languages and software (e.g., spreadsheet and database tools).
4.f.7	Support the seamless integration of vocabularies into data management workflows and user interfaces to increase vocabulary reuse.

ii.   Rationale

The main recommendations in the key topic of finding and using FAIR vocabularies are about making vocabularies **discoverable, accessible, and useful** for both humans and machines in a modern, data-driven world. The rationale is to embed vocabularies into digital workflows, so they become an everyday, effortless part of how we manage data. Raised were recommendations focused on *Enabling Machine-Readable and AI-Friendly Vocabularies,* including in emergent roles in AI optimisation and ethical data re-use; *Promoting Interoperability and Reusability*, making it easier for people and systems to find and use vocabularies; *User-Centric Design and Integration* for a user-friendly experience; and *Attribution and Licensing,* ensuring vocabularies are appropriately licensed for reuse and cited.

There is a proliferation of vocabularies available online, many of which are local solutions for local vocabularies: researchers are confused as to which ones are reliable, endorsed, licensed for re-use, and have sustainable funding. Whilst FAIR-compliant vocabularies are essential for making any dataset FAIR-compliant for both human and machine readability and actionability (Interoperability principle I2: *"meta(data) use vocabularies that follow FAIR principles"*), many vocabularies are not entirely FAIR. With vocabularies developed by small groups at the local or regional scale, and/or dataset-specific profiles of existing vocabularies, there are often not the resources required to make these vocabularies FAIR compliant. Furthermore, many vocabularies are not described using standard metadata schema elements, and metadata for citation, such as DOIs and standard licensing. The machine-actionability of vocabularies needs to be addressed as a priority, as well as working towards the convergence and reuse of existing vocabularies where possible (including finding, mapping and harmonisation) to provide fewer but better vocabularies. Vocabulary services and registries play a crucial role in the Vocabulary Ecosystem, enabling the creation, discovery, and reuse of FAIR vocabulary. A dedicated program of work is required to address the challenges of vocabulary services federation, including articulating a shared vision.

g) Skills, training and outreach

i.   Recommendations

Number	Recommendations
1.g.1	Recognise the value of, and invest in, vocabulary professionals critical to the FAIR Vocabulary Ecosystem.
1.g.2	Encourage and support communities of practices for knowledge brokering, sharing, mentorship etc.
1.g.3	Use tangible case studies and vignettes that demonstrate the value of FAIR vocabularies to stakeholder communities, including the cost-benefit compared with using local lists of data systems.
1.g.4	Collaborate for co-investment across parties involved in the vocabulary lifecycle.

1.g.5	Leverage working groups, including international efforts that cut across/address pressing issues in the vocabulary data supply chain.
1.g.6	Establish communication and information dissemination channels across the Vocabulary Ecosystem in Australia, such as through the Australian Vocabulary Special Interest Group (AVSIG) and the Australian Government Linked Data Working Group (AGLDWG).
2.g.1	Foster linkage and collaboration with relevant international initiatives, for example, WorldFAIR+, Interoperable Descriptions of Observable Property Terminology Working Group (I-ADOPT WG), RD-A Vocabulary Services Interest Group, RD-A FAIR Mappings Working Group.
2.g.2	Provide the necessary support for knowledge holders and domain experts (e.g., communities, committees, bodies or individuals) critical in the governance of vocabularies as authoritative resources.
4.g.1	Provide a central touchpoint for those seeking guidance on the various tools and services for finding, creating, publishing, and reusing vocabularies.
4.g.2	Support the democratisation of vocabulary reuse through providing training and workshops on vocabularies tailored to communities' and domains' needs. Provide guidance on finding vocabularies across services and evaluating them as fit-for-purpose and likely persistent for reuse.
4.g.3	Provide access to a body of knowledge for education on vocabulary types, standards, best practices, guidelines, and examples critical to the FAIR vocabulary lifecycle.
4.g.4	Advocate for vocabularies as first-class 'digital citizens' in their own right, including support for citation and indexing.

ii.   Rationale

The core rationale of the skills, training and outreach recommendations is to **build a robust, sustainable, and collaborative ecosystem for creating, managing and using vocabularies** in Australia. The focus is on valuing the people and partnerships that make the Vocabulary Ecosystem work, as well as educating users about the benefits of using FAIR vocabularies. Recommendations focused on the areas of *Investing in People and Expertise,* emphasising that the success of a Vocabulary Ecosystem depends on the people who support it. This focus is on human infrastructure as much as on digital infrastructure; *Demonstrating Value and Promoting Collaboration* to grow buy-in through collaboration and co-investment; *Educating and Empowering Users* to democratise vocabulary reuse by empowering users to find, evaluate, and use vocabularies effectively; and *Elevating the Status of Vocabularies* as 'first-class digital citizens', including through ensuring they can be cited and indexed.

FAIR vocabularies and a vibrant Vocabulary Ecosystem hinge on a diverse pool of skills and expertise. There is a recognised shortfall in Australia that must be addressed in any National Digital Research Infrastructure workforce planning. Skills and education needs are evolving at a rapid pace, driven by the current upward trend in research practices such as the adoption of machine-actionable FAIR principles; vocabulary-backed technology solutions such as graph databases; making data assets AI-ready; etc.

There are few university courses or other training courses available that cover or specialise in FAIR vocabularies, and access to authoritative online training courses is desirable. Developing formalised

information models, ontologies and linked data implementations is usually beyond the skill set of the majority of research scientists, and they do not necessarily need to be able to develop these skills. Hence, training courses need to be targeted at the three communities: subject matter experts, users, and informaticians (Section 4).

5. Conclusions

The development of this Strategic Community Roadmap has been a collaborative effort, involving numerous parties from the Australian research and government sectors. Through a series of symposia, a workshop and community consultations held between 2021 and 2024, a comprehensive understanding of the challenges and opportunities within the Australian Vocabulary Ecosystem has been achieved. It is intended to complement, rather than compete with, vocabulary activities, outputs and recommendations that come out of international efforts initiatives such as the WorldFAIR Project (WorldFAIR, n.d.-a), working groups/interest groups and plenaries of the Research Data Alliance (RD-A; RD-A n.d.) and clusters within the US-based Earth Science Information Partners (ESIP; ESIPFed n.d.-a), such as the Semantics Cluster and the Data Readiness Cluster (ESIPFed, n.d.-b). It is anticipated that implementations of this Community Roadmap would continue to be informed by best practices emerging from such initiatives.

This Community Roadmap presents a critical step towards maturing the FAIR Vocabulary Ecosystem within Australia. By adopting and adapting the Vision and Mission and actioning recommendations outlined in this document strategically, parties can work collectively towards creating a thriving Vocabulary Ecosystem. This ecosystem will not only enhance the utility and impact of research data within Australia but also contribute to the global data landscape, supporting beneficial cultural, environmental, societal, economic, and health outcomes. The successful realisation of this Community Roadmap will cement Australia's advances in research data management, facilitate greater reuse of research data, and help address significant transdisciplinary challenges. By working together, parties can ensure that the Vocabulary Ecosystem in Australia is robust, sustainable, and capable of supporting the needs of a rapidly evolving data-driven landscape.

The roadmap has identified a clear Vision and Mission to guide the actions and investments necessary to strengthen the FAIR Vocabulary Ecosystem in Australia. The Vision emphasises that shared vocabularies and their related ecosystem should form a key part of the Australian and international data landscape, enabling the impact of data for beneficial outcomes across various sectors. The Mission focuses on furthering the consistency and sustainability in the development, implementation, governance, and use of vocabularies across domains, both nationally and internationally.

A significant outcome of this effort has been the identification of 57 recommendations, categorised into seven key topics and organised around four Strategic Themes ranked by suggested importance and urgency for implementation:

1. Foster a thriving vocabulary ecosystem;
2. Improve the consistency, content, and context of vocabularies across Australia;
3. Improve the consistency of vocabulary tools and services; and

4. Lower barriers for using and maintaining vocabularies.

The 57 recommendations provide a comprehensive framework for parties to address the critical technical and social barriers hindering the creation, adoption and effective use of vocabularies in Australia. They cover a wide range of aspects, from the adoption of technical standards and the creation of vocabulary content to publishing, governance, mapping, finding, and utilising FAIR vocabularies, as well as skills, training, and outreach.

The path forward involves implementing the 57 recommendations across the four Strategic Themes, with a focus on prioritisation, resource allocation, and collaborative action. As the FAIR Vocabulary Ecosystem evolves, it is anticipated that the utility of vocabularies will become well understood, and a burgeoning ecosystem of vocabularies and associated services will emerge, lowering barriers to publishing and using vocabularies as a foundational aspect of data assets, services, and infrastructure.

This Community Roadmap outlines a pathway for Australia to transition from its current fragmented landscape to a cohesive and dynamic FAIR Vocabulary Ecosystem. The transition will unlock the hidden value within our national data assets, empower researchers and data scientists with interoperable resources, and ensure that data is suitable for an AI-enabled future. Implementing these recommendations will require coordination, investment, and effort from various parties, including NDRI Strategy Coordinators and Implementers, NCRIS Facilities and NRI Platforms, universities, researchers, data scientists, and government agencies. The successful execution of this Roadmap will facilitate a more coordinated approach to vocabulary management, enhance data interoperability, and support the FAIR Data Principles.

Ultimately, this Strategic Community Roadmap aims to maximise the societal, economic, and environmental benefits derived from our national research and data initiatives. It is a **call to action** to collectively build the vocabulary ecosystem essential for Australia's data-driven future.

6. References

ADA (n.d.). Home – Australian Data Archive. Accessed on 18 January 2025. Available at
https://ada.edu.au/

AeRO (2024). Australasian eResearch Organisation, Australasian eResearch Newsletter, November
2024. Accessed on 12 March 2025. Available at https://us9.campaign-
archive.com/?u=2d3ef974f3a64989335c0e079&id=68217f08ec

AGLDWG (n.d.). Home - Australian Government Linked Data Working Group (AGLDWG). Accessed on 23
March 2025. Available at https://www.linked.data.gov.au/

Albertoni, R., Browning, D., Cox, S.J.D., Gonzalez-Beltran, A., Perego, A., and Winstanley P. (2024). Data
Catalog Vocabulary (DCAT) — Version 3 W3C Recommendation 22 August 2024. Accessed on 25
March 2025. Available at https://www.w3.org/TR/vocab-dcat/

Amano, T., Rios Rojas, C., Boum II, Y., Calvo, M., and Misra, B.B. (2021). Ten tips for overcoming language
barriers in science. *Nature Human Behaviour*, *5*(9), pp.1119-1122. https://doi.org/10.1038/s41562-
021-01137-1

Australian Research Data Commons (2019). ARDC Research Data Rights Management Guide. Zenodo.
https://zenodo.org/records/5091580

Australian Research Data Commons (2023). Event report: 2023 Vocabulary Symposium. 2023
Vocabulary Symposium, Onsite (Canberra, ANU) and online. ARDC.
https://doi.org/10.5281/zenodo.10145667

ARDC (n.d.-a). Home - Australian Research Data Commons (ARDC). Accessed on 18 January 2025.
Available at https://ardc.edu.au/

ARDC (n.d.-b). Australian Research Data Commons (ARDC) Thematic Research Data Commons.
Accessed on 18 January 2025. Available at https://ardc.edu.au/about-us/thematic-research-
data-commons/

ARDC (n.d.-c). Home – Australian Research Data Commons (ARDC) Research Vocabularies Australia.
Accessed on 29 March 2025. Available at https://vocabs.ardc.edu.au/

ARDC (n.d.-d). Australian Research Data Commons (ARDC) Vocabulary Service Extension (Health and
Medical). Accessed on 15 March 2025. Available at https://ardc.edu.au/project/ardc-vocabulary-
service-extension-health-and-medical/

ARDC (n.d.-e). Home - Australian Vocabulary Special Interest Group (AVSIG). Accessed on 30 March
2025. Available at https://sites.google.com/ardc.edu.au/avsig

Atkinson, R., Coyle, K., Isaac, A., and Car, N.J. (2024). Profile Guidance. W3C Editor's Draft 07 October
2024. Accessed on 28 March 2025. Available at https://w3c.github.io/dxwg/profiles/

Australian Government (n.d.). Data Availability and Transparency Act 2022. Accessed on 26 January
2025. Available at https://www.legislation.gov.au/C2022A00011/latest/text

Australian Government (2023). Data and Digital Government Strategy: The Data and Digital Vision for a world-class APS to 2030. Accessed on 27 January 2025. Available at https://www.dataanddigital.gov.au/strategy

Australian Government Department of Education (n.d.-a). Home - National Research Infrastructure (NRI). Accessed on 18 January 2025. Available at https://www.education.gov.au/national-research-infrastructure .

Australian Government Department of Education (n.d.-b). Home - National Collaborative Research Infrastructure Strategy (NCRIS). Accessed on 18 January 2025. Available at https://www.education.gov.au/ncris

Australian Government Department of Education (n.d.-c). Funded research infrastructure projects. Accessed on 18 January 2025. Available at https://www.education.gov.au/national-research-infrastructure/funded-research-infrastructure-project

Australian Government Department of Education (n.d.-d). National Digital Research Infrastructure Strategy (NDRI). Accessed on 18 January 2025. Available at https://www.education.gov.au/national-research-infrastructure/resources/national-digital-research-infrastructure-strategy

Australian Government Department of Education (n.d.-e). National Research Infrastructure Advisory Group (NRIAG). Accessed on 18 January 2025. Available at https://www.education.gov.au/national-research-infrastructure/advisory-group

Baker, T., Vandenbussche, P.Y., and Vatant, B. (2013a). Requirements for vocabulary preservation and governance. *Library Hi Tech, 31*(4), pp.657-668. https://doi.org/10.1108/LHT-03-2013-0027

Baker, T., Bechhofer, S., Isaac, A., Miles, A., Schreiber, G., and Summers, E. (2013b). Key choices in the design of Simple Knowledge Organization System (SKOS). *Journal of Web Semantics,* 20, 35-49. https://doi.org/10.1016/j.websem.2013.05.001

BARTOC.org (n.d.). Terminology Registries. Accessed on 29 August 2025. Available at https://bartoc.org/registries

Bastrakova, I.V. (2018). Geoscience Australia Community Metadata Profile of ISO 19115-1:2014. *Record 2018/26, Geoscience Australia, Canberra.* http://dx.doi.org/10.11636/

Beckett, D., Berners-lee, T., Prud'hommeaux, E., and Carothers, G. (2014). RDF 1.1 Turtle Terse RDF Triple Language. W3C recommendation 25 February 2014. Accessed on 23 March 2025. Available at https://www.w3.org/TR/turtle/

Belton, A., Boyle, D., Faux, N., Brownlee, R., Lovelace-Tozer, M., and Ward, R. (2024). Advancing Australian Vocabularies and Mappings. Accessed on 22 March 2025. Available at https://doi.org/10.26188/27913398.v2

Bioportal (n.d.). Home - Bioportal (biomedical ontologies). Accessed on 29 March 2025. Available at https://bioportal.bioontology.org/

Boissier, O., Ciortea, A., Harth, A., and Ricci, A. (2021). Autonomous agents on the web. In *Dagstuhl-Seminar 21072: Autonomous Agents on the Web* (p. 100p). https://doi.org/10.4230/DagRep.11.1.24

Boyle, D. R., Lawley, M., and Brownlee, R. (2023). Research Terminologies Australia (Rosetta). Vocabulary Symposium 2022, Onsite (Canberra, ANU) and online. Zenodo. https://doi.org/10.5281/zenodo.7834769

Boyle, D., McEachern, D., Wyborn, L., Hobern, D., Dharmawardena, K., Wong, M., and Atkins, N. (2024). The Research Vocabularies Ecosystem in Australia: Continuing the Conversation. Abstract, eResearch Australasia, 2024 Conference, Melbourne, Australia, October-November 2024. Accessed on 18 January 2025. Available at https://conference.eresearch.edu.au/the-research-vocabularies-ecosystem-in-australia-continuing-the-conversation/ .

Bray, T., Paoli,T., Sperberg-McQueen, C.M., Maler, E., and Yergeau. F. (2008). Extensible Markup Language (XML) 1.0 (Fifth Edition). Accessed on 18 January 2025. Available at https://www.w3.org/TR/xml/

Brickley, D., and Guha, R.F. (2014). RDF Schema 1.1. Accessed on 18 January 2025. Available at https://www.w3.org/TR/rdf-schema/

Brownlee, R., and Wyborn, L. (2023, April 17). Starting the roadmap - Documents and communities. Vocabulary Workshop 2022, Canberra, ANU. Zenodo. https://doi.org/10.5281/zenodo.7834372

Brownlee, R., Cox, S., Dharmawardena, K., Gregory. A., Hodson, S., McEachern, S., Orten, H., Wackero, J., and Wyborn, L. (2021). Interoperability for Cross-Domain Research: FAIR Vocabularies. Workshop Report. Data Documentation Initiative (DDI) Alliance and CODATA. Accessed on 28 October 2024. Available at https://ddi-alliance.atlassian.net/wiki/spaces/DDI4/pages/2681667690/2021+Interoperability+for+Cross-Domain+Research+FAIR+Vocabularies

Carroll, S.R., Garba, I., Figueroa-Rodríguez, O.L., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., Rowe, R., and Sara, R. (2020). The CARE Principles for Indigenous Data Governance. *Data Science Journal*, 19(1), p.43. https://doi.org/10.5334/dsj-2020-043

Chuc, E., and Car, N. (2024). VocPub Editor: A Free and Intuitive Browser Vocabulary Management Tool. Abstract, eResearch Australasia 2024, Brisbane, Australia, Accessed on 22 March, 2025. Available at https://conference.eresearch.edu.au/vocpub-editor-a-free-and-intuitive-browser-vocabulary-management-tool/.

CESSDA (n.d.). Home - CESSDA (Consortium of European Social Science Data Archives) Vocabulary Services. Accessed on 11 April 2025. Available at https://vocabularies.cessda.eu/about

CF Metadata Conventions (n.d.). Home - Climate and Forecasting Metadata Conventions. Accessed on 23 March 2025. Available at https://cfconventions.org/

Chuc, E., and Car, N. (2024). VocPub Editor: A Free and Intuitive Browser Vocabulary Management Tool. Abstract, eResearch Australasia 2024, Brisbane, Australia, Accessed on 22 March, 2025.

Available at https://conference.eresearch.edu.au/vocpub-editor-a-free-and-intuitive-browser-vocabulary-management-tool/.

CODATA (n.d.). Home - Committee on Data and the International Science Council (CODATA). Accessed on 18 January 2025. Available at https://codata.org

CODATA RDM Terminology Working Group (2024). CODATA Research Data Management (RDM) Terminology (2023 version): overview (Version 2023). Zenodo. https://doi.org/10.5281/zenodo.10626170

CoreTrustSeal (n.d.). Home - CoreTrustSeal. Accessed on 29 March 2025. Available at https://www.coretrustseal.org/

Corcho, O., Ekaputra, F.J., Heibi, I., Jonquet, C., Micsik, A., Peroni, S., and Storti, E. (2024). A maturity model for catalogues of semantic artefacts. *Scientific Data*, *11*(1), p.479. https://doi.org/10.1038/s41597-024-03185-4

Cox, S.J.D., Gonzalez-Beltran, A.N., Magagna, B., and Marinescu, M.C. (2021). Ten simple rules for making a vocabulary FAIR. *PLoS Computational Biology*, 17(6): e1009041. https://doi.org/10.1371/journal.pcbi.1009041

Crystal-Ornelas, R., Varadharajan, C., Bond-Lamberty, B., Boye, K., Burrus, M., Cholia, S., Crow, M., Damerow, J., Devarakonda, R., Ely, K.S., and Goldman, A. (2021). A guide to using GitHub for developing and versioning data standards and reporting formats. *Earth and Space Science*, 8(8), https://doi.org/10.1029/2021EA001797

DOI Foundation. (2025). DOI Foundation. Accessed on 10 September 2025. Available at https://www.doi.org/

David, R., Specht, A., O'Brien, M., Wyborn, L., Drummond, C., Edmunds, R., Filippone, C., Machicao, J., Miyairi, N., Parton, G., Pignatari Drucker, D., Stall, S., and Zimmer, N. (2022). Multilingual Data Challenges in Professionalizing Data Stewardship worldwide (V1.1). RD-A 19th Plenary Meeting, Part of International Data Week, 20–23 June 2022 (RD-A Plenary 19th), Seoul, South Korea. Zenodo. https://doi.org/10.5281/zenodo.6588167

Dillo, I., Treloar, A., Lusoli, W., Kupiainen, I., Asch, M., Diepenbroek, M., Hayashi, K., Noh, S.-Y., Oh, S. K., Skålin, R., Sithole, H., van Deventer, M., Golliez, A., Grolimund, P., Cudre- Mauroux, P., Dearry, A., Hodson, S., Neylon, C., Madalli, D. P., Bermann, F., and Harrower, N. (2017). Business models for sustainable research data repositories. (OECD Science, Technology and Innovation Policy Papers). OECD Publishing. Accessed on 29 March 2025. Available at https://doi.org/10.1787/302b12bb-en

Duerr, R., Buttigieg, P.L., Cross, G.B., Blumberg, K.L., Whitehead, B., Wiegand, N., and Rose, K. (2024). Harmonizing GCW Cryosphere Vocabularies with ENVO and SWEET. Towards a General Model for Semantic Harmonization. *Data Science Journal*, 23(1), p. 26. https://doi.org/10.5334/dsj-2024-026

Dürst, M.J., and Suignard, M. (2005). Internationalized Resource Identifiers (IRIs), Request for comments 3987. Accessed on 31 March 2025. Available at https://datatracker.ietf.org/doc/html/rfc3987

Dutta, B., Nandini, D., & Shahi, G. (2015). MOD: Metadata for Ontology Description and Publication. *International Conference on Dublin Core and Metadata Applications, 2015.* https://doi.org/10.23106/dcmi.952136974

Elixir-Europe (n.d.). Elixir Research Data Management Kit (RDM Kit). Accessed on 20 April 2025. Available at https://rdmkit.elixir-europe.org/machine_actionability

ESIP Data Readiness Cluster (2022). Checklist to Examine AI-readiness for Open Environmental Datasets. ESIP. Online resource. https://doi.org/10.6084/m9.figshare.19983722.v1

ESIPFed (n.d.-a). Home – Earth Science Information Partners (ESIP). ESIPFed. Accessed on 12 November 2023. Available at https://www.esipfed.org/

ESIPFed (n.d.-b). Collaboration is what we do. Accessed on 12 November 2024. Available at https://www.esipfed.org/collaboration-areas/

ESIPFed (n.d.-c). Home - Earth Science Information Partners (ESIP) Community Ontology Repository (COR). Accessed on 29 March 2025. Available at https://cor.esipfed.org/

European Commission (n.d.). Home - European Open Science Cloud (EOSC). Accessed on 18 January 2025. Available at https://research-and-innovation.ec.europa.eu/strategy/strategy-research-and-innovation/our-digital-future/open-science/european-open-science-cloud-eosc_en

Faircore4eosc (n.d.). Metadata Schema and Crosswalk Registry (MSCR). Accessed on 24 August 2025. Available at https://faircore4eosc.eu/eosc-core-components/metadata-schema-and-crosswalk-registry-mscr

FAO (n.d.). Home - Food and Agricultural Organisation of the United Nations (FAO) AgroPortal. Accessed on 29 March 2025. Available at https://aims.fao.org/agroportal

Fensel, D., Şimşek, U., Angele, K., Huaman, E., Kärle, E., Panasiuk, O., Toma, I., Umbrich, J., Wahler, A., Fensel, D., and Şimşek, U. (2020). Introduction: what is a knowledge graph?. *Knowledge graphs: Methodology, tools and selected use cases*, pp.1-10. Springer, Cham, Switzerland. https://doi.org/10.1007/978-3-030-37439-6_1

Gandon, F., and Schreiber, G. (2014). RDF 1.1 XML Syntax. W3C Recommendation 25 February 2014. Accessed on 25 March 2025. Available at https://www.w3.org/TR/rdf-syntax-grammar/

Garijo D, Poveda-Villalón M. (2020) Best Practices for Implementing FAIR Vocabularies and Ontologies on the Web. In: Cota G, Daquino M, Pozzato GL, editors. Studies on the Semantic Web [Internet]. IOS Press. Accessed on 31 August 2025. Available at http://ebooks.iospress.nl/doi/10.3233/SSW200034

Goldbeck, G., and Simperler, A. (2018). Report on Workshop on Interoperability in Materials Modelling. Zenodo. https://doi.org/10.5281/zenodo.1240229

Goleby, B., Rees, N., and Wyborn, L. (2023). Metadata Standards for Magnetotelluric Time Series Data: AuScope AusLAMP Profile. Accessed on 22 March 2025. Available at https://demo.vocabs.ardc.edu.au/viewById/1019

Gregory, A., Bell, D., Brickley, D., Buttigieg, P. L., Cox, S., Edwards, M., Doug, F., Gonzalez Morales, L. G., Heus, P., Hodson, S., Kanjala, C., Le Franc, Y., Maxwell, L., Molloy, L., Richard, S., Rizzolo, F., Winstanley, P., Wyborn, L., and Burton, A. (2024). WorldFAIR (D2.3) Cross-Domain Interoperability Framework (CDIF) (Report Synthesising Recommendations for Disciplines and Cross-Disciplinary Research Areas) (Version 1). Zenodo. https://doi.org/10.5281/zenodo.11236871

Guha, V., Brickley, D., and Macbeth, S. (2015). Schema.org: Evolution of structured data on the Web: Big data makes common schemas even more necessary. *Queue, 13*(9), pp.10-37. https://doi.org/10.1145/2857274.2857276

Harpring, P. (2013). Introduction to controlled vocabularies: terminology for art, architecture, and other cultural works. Getty Publications. Accessed on 18 January 2025. Available at https://www.getty.edu/publications/virtuallibrary/160606018X.html.

Heidorn, P. B. (2008). Shedding light on the dark data in the long tail of science. *Library Trends, 57*(2), 280-299. https://doi.org/10.1353/lib.0.0036

Herman, I., Adida, B., Sporny, M., and Birbeck, M. (2015). RDFa 1.1 Primer - Third Edition Rich Structured Data Markup for Web Documents. W3C Working Group Note 17 March 2015. Accessed on 23 March 2025. Available at https://www.w3.org/TR/rdfa-primer/

Hitzler, P., Krötzsch, M., Parsia, B., Patel-Schneider, P.F., and Rudolph S. (2012). OWL 2 Web Ontology Language Primer (Second Edition), World Wide Web Consortium; Accessed on 19 January, 2025. Available at https://www.w3.org/TR/owl-primer/

HL7 (n.d.). Health Level 7 International (HL7) Fast Healthcare Interoperability Resources (FHIR) overview. Accessed on 25 March 2025. Available at https://www.hl7.org/fhir/overview.html .

Hodson, S., and Gregory, A. (2024). WorldFAIR Final Policy Brief: Enabling Global FAIR Data - Recommendations for Research Infrastructures (Version 2). Zenodo. https://zenodo.org/records/14236140

Hugo, W., Le Franc, Y., Coen, G., Parland-von Essen, J., and Bonino, L. (2020). D2.5 FAIR Semantics Recommendations Second Iteration (1.0 DRAFT). Zenodo. https://doi.org/10.5281/zenodo.4314321

Hyland, B., Atemezing, G., and Villazón-Terrazas, B. (2014). Best Practices for Publishing Linked Data. W3C Working Group Note 09 January 2014. Accessed on 24 August 2024. Available at https://www.w3.org/TR/ld-bp/#how-to-find-existing-vocabularies

Isaac, A., and Summers, E. (2009). SKOS Simple Knowledge Organization System Primer. World Wide Web Consortium. Accessed on 19 January 2025. Available at https://www.w3.org/TR/skos-primer/

IEEE (n.d.). Home - IEEE (Institute of Electrical and Electronic Engineers). Accessed on 2 February 2025. Available at https://www.ieee.org/

ISO (n.d.). Home - ISO Global Standards for Trusted Goods and Services. Accessed on 2 February 2025. Available at https://www.iso.org/home.html

ISO (2004). ISO 19106:2004, Geographic Information – Profiles (reconfirmed 2023). Accessed on 22 March 2025. Available at https://www.iso.org/standard/26011.html#lifecycle

ISO (2014). ISO 19115-1:2014. Geographic Information - Metadata. Part 1: Fundamentals. Accessed on 22 March 2025. Available at https://www.iso.org/standard/53798.html#amendment

ISO (2022). ISO 704:2022 - Terminology work - Principles and Methods. Accessed on 2 February 2025. Available at https://www.iso.org/standard/79077.html

Jonquet, C., Toulet, A., Arnaud, E., Aubin, S., Yeumo, E.D., Emonet, V., Graybeal, J., Laporte, M.A., Musen, M.A., Pesce, V., and Larmande, P. (2018). AgroPortal: A vocabulary and ontology repository for agronomy. *Computers and Electronics in Agriculture*, 144, pp.126-143. https://doi.org/10.1016/j.compag.2017.10.012

Jonquet, C., and Poveda-Villalón, M. (2023). About versioning ontologies or any digital objects with clear semantics. HAL Open Science. Version 2. HAL Id: hal-04094847. Accessed on 5 September 2025. Available at https:hal.science/hal-04094847v2

Jonquet, C., Whitehead, B., Car, N.; Kokkinaki, A., Atkinson, R., Pierkot, C., Karam, N., Graybeal, J., and Wong, M. (2025). Interoperable and Federated Vocabulary Services. Abstract, SciDataCon 2025, Brisbane, Australia. Accessed on 1 September 2025. Available at https://scidatacon.org/event/9/contributions/54/

JSON (n.d.). Home – Java Script Object Notation. Accessed on 2 February 2025. Available at https://www.json.org/json-en.html

JSON-LD (n.d.). Home - JSON for Linking Data. Accessed on 19 January 2025. Available at https://json-ld.org/

Lifewatch (n.d.). Home - EcoPortal. Accessed on 29 March 2025. Available at https://ecoportal.lifewatch.eu/

Lin, D., Crabtree, J., Dillo, I., Downs, R.R., Edmunds, R., Giaretta, D., De Giusti, M., L'Hours, H., Hugo, W., Jenkyns, R., and Khodiyar, V. (2020). The TRUST Principles for digital repositories. *Scientific data*, 7(1), p.144. https://doi.org/10.1038/s41597-020-0486-7

Lóscio, B.F., Burle, C, and Newton, C. (2017). Data on the Web, Best Practices. W3C Recommendation 31 January 2017. Accessed on 24 August. Available at https://www.w3.org/TR/dwbp/

LOV (n.d.). Home - Linked Open Vocabularies (LOV). Accessed on 29 March 2025. Available at https://lov.linkeddata.es/dataset/lov/

Ma, X., Wyborn, L., and Hodson, S. (2022). Data sharing: more science unions must act. Nature, 610(7931), https://doi.org/10.1038/d41586-022-03233-2

Magagna, B., Rosati, I., Stoica, M., Schindler, S., Moncoiffe, G., Devaraju, A., Peterseil, J., and Huber, R. (2021). The I-ADOPT Interoperability Framework for FAIRer data descriptions of biodiversity. arXiv preprint arXiv:2107.06547. https://doi.org/10.48550/arXiv.2107.06547

Magagna, B., Moncoiffé, G., Devaraju, A., Stoica, M., Schindler, S., and Pamment, A. (2022). InteroperAble Descriptions of Observable Property Terminologies (I- ADOPT) WG Outputs and Recommendations. Research Data Alliance. Accessed on 30 March 2025. Available at https://doi.org/10.15497/RDA00071

Mahuika, J., Siddeswara, G., and Devaraju, A. (2021). Terrestrial Ecosystem Research Network (TERN) Metadata Profile of ISO 19115-3:2016 and ISO 19157-2:2016 (v1.0) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.5652221

Matentzoglu, N., Balhoff, J.P., Bello, S.M., Bizon, C., Brush, M., Callahan, T.J., Chute, C.G., Duncan, W.D., Evelo, C.T., Gabriel, D., and Graybeal, J. (2022). A Simple Standard for Sharing Ontological Mappings (SSSOM), *Database*, Volume 2022, 2022, baac035, https://doi.org/10.1093/database/baac035

McCreary, D. (2006). Patterns of semantic integration. April. Accessed on 18 January 2025. Available at https://www.slideshare.net/dmccreary/semint.

McEachern, S., Cox, S., Brownlee, R., Wyborn, L., Burton, A., and Wong, M. (2022). Enabling FAIR Vocabularies for Interoperability in Cross Domain Research in Australia. Abstract, eResearch Australasia, 2022 Conference, Brisbane, Australia, October 2022. Accessed on18 January 2025. Available at https://conference.eresearch.edu.au/events/enabling-fair-vocabularies-for-interoperability-in-cross-domain-research-in-australia/

McEachern, S., Wong, M., Cox, S., Hobern, D., Atkins, N., Dharmawardena, K., and Wyborn, L. (2023). A Roadmap for the Future of Shared Vocabularies in Australia: Harmonising Across Sectors, Disciplines and Internationally. Abstract, eResearch Australasia, 2023 Conference, Melbourne, Brisbane, Australia, October 2023. Accessed on 18 January 2025. Available at https://conference.eresearch.edu.au/a-roadmap-for-the-future-of-shared-vocabularies-in-australia-harmonising-across-sectors-disciplines-and-internationally/

McGuinness, D.L. (2003). Ontologies come of age. In: Fensel D., Hendler, J., Lieberman, H., and Wahlster, W., editors. *Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential.* MIT Press, 2003. https://doi.org/10.7551/mitpress/6412.001.0001

Miles, A., and Bechhofer, S. (2009). SKOS Simple Knowledge Organization System Reference [Internet]. Cambridge, Mass. USA: World Wide Web Consortium. Accessed on 19 January 2025. Available at http://www.w3.org/TR/skos-reference/

Miller, E.J. (2001). An Introduction to the Resource Description Framework. *Journal of Library Administration*, 34(3–4), 245–255. https://doi.org/10.1300/J111v34n03_04

Naveed, H., Khan, A.U., Qiu, S., Saqib, M., Anwar, S., Usman, M., Akhtar, N., Barnes, N. and Mian, A. (2025). A comprehensive overview of large language models. *ACM Transactions on Intelligent Systems and Technology*, *16*(5), pp.1-72. https://doi.org/10.1145/3744746

NERC (n.d.). Home - National Environmental Research Council (NERC) Vocabulary Server. Accessed on 29 March 2025. Available at https://www.bodc.ac.uk/resources/products/web_services/vocab/

NHMRC (National Health and Medical Research) Council (2018). Australian Code for the Responsible Conduct of Research, 2018. National Health and Medical Research Council, Australian Research Council and Universities Australia. Commonwealth of Australia, Canberra. Accessed on 26 January 2025. Available at https://www.nhmrc.gov.au/file/14384/download?token=gje4DNtT

NHMRC (National Health and Medical Research Council) (2019). Management of Data and Information in Research: A guide supporting the Australian Code for the Responsible Conduct of Research. National Health and Medical Research Council, Australian Research Council and Universities Australia. Commonwealth of Australia, Canberra. Accessed on 26 January 2025. Available at https://www.nhmrc.gov.au/file/14359/download?token=L5GTBw96

NLM (n.d.). Home - National Library of Medicine. RxNORM Overview. Accessed on 23 March 2025. Available at https://www.nlm.nih.gov/research/umls/rxnorm/overview.html

Obrst, L. (2008). From Searching to Knowing — Spectrum of Knowledge Representation and Reasoning Capabilities. In: Davis, M, Semantic Wave 2008 Report: Industry Roadmap to Web 3.0 & Multibillion Dollar Market Opportunities (2008), http://www.project10x.com Accessed on 3 April 2025. Available at https://www.scribd.com/document/2599547/Project10X-s-Semantic-Wave-2008-Report-Industry-Roadmap-to-Web-3-0/

OGC (n.d.). Home – OGC (Open Geospatial Organisation). Accessed on 2 February 2025. Available at https://www.ogc.org/

Peacock, J.R., Frassetto, A., Kelbert, A., Egbert, G., Smirnov, M., Schultz, A.C., Kappler, K.N., Ronan, T., and Trabant, C. (2021) Metadata Standards for Magnetotelluric Time Series Data: U.S. Geological Survey data release, https://doi.org/10.5066/P9AXGKEV

Peacock, J. R., Kappler, K., Ronan, T., Heagy, L., Kelbert, A., and Frassetto, A. (2022). MTH5: An archive and exchangeable data format for magnetotelluric time series data. *Computers and Geoscience*, 162, https://doi.org/10.1016/j.cageo.2022.105102

Prent, A., Wyborn, L., and Farrington, R. (2024). WorldFAIR (D5.2) Geochemistry Methodology and Outreach (Version 1). Zenodo. https://doi.org/10.5281/zenodo.10406332

Ramezani, P., Grau, N., Jonquet, C., and Fiore, N. (2023). M4.1 Semantic artefact governance models: example of community practices. Zenodo. https://doi.org/10.5281/zenodo.10287011.

Ramezani, P., Grau, N., Jonquet, C., and Fiore, N. (2024). D4.1 - Semantic artefact governance models and disciplinary approaches for inclusion within EOSC. Zenodo. https://doi.org/10.5281/zenodo.15113031.

RD-A (n.d.). Home - International Research Data Alliance (RD-A). Accessed on 12 November 2024. Available at https://www.rd-alliance.org/

RD-A FAIR Mappings WG (n.d.). Home - RD-A FAIR Mappings Group. Accessed on 18 January 2025. Available at https://www.rd-alliance.org/groups/fair-mappings-wg/activity/

RD-A I-ADOPT WG (n.d.). Home - InteroperAble Descriptions of Observable Property Terminology Working Group (I-ADOPT WG). Accessed on 30 on March 2025. Available at https://www.rd-

alliance.org/groups/interoperable-descriptions-observable-property-terminology-wg-i-adopt-wg/repository/

RD-A Metadata Standards Catalog (n.d.). Home - RD-A Metadata Standards Catalog. Accessed on 1 April 2025. Available at https://rdamsc.bath.ac.uk/

RD-A Vocabulary Services IG (n.d.). Home - RD-A Vocabulary Services Interest Group. Accessed on 30 March 2025. Available at https://www.rd-alliance.org/groups/vocabulary-services-interest-group/activity/

RDF Working Group (2014). Home - Resource Description Framework (RDF). Accessed on 18 January 2025. Available at https://www.w3.org/RDF/

Schultes, E., Roos, M., Bonino da Silva Santos, L.O., Guizzardi, G., Bouwman, J., Hankemeier, T., Baak, A., and Mons, B. (2022). Fair digital twins for data-intensive research. *Frontiers in Big Data*, 5, p.883341. https://doi.org/10.3389/fdata.2022.883341

Shepherd, A., Jones, M., Richard, S., Jarboe, N., Vieglais, D., Fils, D., Duerr, R., Verhey, C., Minch, M., Mecum, B., and Bentley, N. (2022). Science-on-Schema.org v1.3.1 (1.3.1). Zenodo. https://doi.org/10.5281/zenodo.7872383

SNOWMED (nd). What is SNOWMED CT (Systematized Nomenclature of Medicine Clinical Terms). Accessed on 25 March 2025. Available at https://www.snomed.org/what-is-snomed-ct

Specht, A., Stall, S., Machicao, J., Catry, T., Chaumont, M., David, R., Devillers, R., Edmunds, R., Jarry, R., Mabile, L., Miyairi, N., O'Brien, M., Margaret O'Brien, Pizzigatti Corrêa, P.L., Santos, S., Subsol, G., and Wyborn, L. (2024). Managing linguistic obstacles in multidisciplinary, multinational, and multilingual research projects. *PloS one*, 19(12). https://doi.org/10.1371/journal.pone.0311967

Sporny, M., Longley, D., Kellogg, G., Lanthaler, M., Champion, P.-A., and Lindström, N. (2020). JSON-LD 1.1. A JSON-based Serialization for Linked Data. W3C Recommendation 16 July 2020. Accessed on 25 March 2025. Available at https://www.w3.org/TR/json-ld/

Tamma, V. and Poveda-Villalón, M. (2021). 4.11 The Value of Knowledge: Opportunistic Knowledge Sharing. *Autonomous Agents on the Web*, p.52. Accessed on 7 April 2025. Available at https://doi.org/10.4230/DagRep.13.2.71

Tarallo, A., Pulieri, M., Ramezani, P., and Rosati, I. (2024). Advancements in EcoPortal: Enhancing functionalities for the ecological domain semantic artefacts repository. *FAIR Connect*, *2*(1), 1-7. https://doi.org/10.3233/FC-240002

TERN (n.d.). Home - Terrestrial Ecosystem Research Network (TERN). Accessed on 29 March, 2025. Available at https://www.tern.org.au/

Tomaszuk, D., and Haudebourg, T. (2025). RDF 1.2 Schema W3C Working Draft 24 February 2025. Accessed on 25 March 2025. Available at https://www.w3.org/TR/rdf12-schema/

UNESCO, Canadian Commission for UNESCO (2022). An introduction to the UNESCO Recommendation on Open Science. https://doi.org/10.54677/XOIR1696

Veltman, K.H. (2001). Syntactic and semantic interoperability: New approaches to knowledge and the semantic web. *New Review of Information Networking*, 7:1, 159-183. https://doi.org/10.1080/13614570109516975

Wackerow, J., Brownlee, R., Cox, s., Dharmawardena, K., Arofan, G., McEachern, Steve, Orten, H., Wyborn, L. 2021 Interoperability for Cross-Domain Research: FAIR Vocabularies. Accessed on 5 September 2025. Available at https://ddi-alliance.atlassian.net/wiki/spaces/DDI4/pages/2681667690/2021+Interoperability+for+Cross-Domain+Research+FAIR+Vocabularies

Wang, J., Aryani, A., Wyborn, L., and Evans, B. (2017). Providing research graph data in JSON-LD using Schema.org. In *Proceedings of the 26th International Conference on World Wide Web Companion* (pp. 1213-1218). https://doi.org/10.1145/3041021.3053052

W3C OWL Working Group (2012). OWL 2 Web Ontology Language Document Overview (Second Edition) [Internet]. W3C Recommendation. Cambridge, Mass. USA: World Wide Web Consortium. Accessed on 19 January 2025. Available at http://www.w3.org/TR/owl2-overview/.

W3C Semantic Web Standards (n.d.-a). Main Page - W3C Semantic Web Standards. Accessed on 26 January 2025. Available at https://www.w3.org/2001/sw/wiki/Main_Page

W3C (n.d.-b). Home - W3C (World Wide Web Consortium. Accessed on 2 February 2025. Available at https://www.w3.org/

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M.A., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., 't Hoen, P.A.C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., and Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data,* 3, 160018. https://doi.org/10.1038/sdata.2016.18.

Wittenburg, P., and Strawn, G. (2018). Common Patterns in Revolutionary Infrastructures and Data. *B2SHARE*. Accessed on 18 December 2024. Available at https://doi.org/10.23728/B2SHARE.4E8AC36C0DD343DA81FD9E83E72805A0.

Wong, M., Brownlee, R., Wyborn, L., Barlow, M., Dharmawardena, K., Atkins, N., and Cox, S.J.D. (2023). 2022 Vocabulary Symposium and Workshop Report. Zenodo. https://doi.org/10.5281/zenodo.8057904

WorldFAIR (n.d.-a). Home - The WorldFAIR project. Accessed on 18 December 2024. Available at https://worldfair-project.eu

WorldFAIR (n.d.-b). WorldFAIR Plus (WorldFAIR+). Accessed on 30 March 2025. Available at https://worldfair-project.eu/worldfair-plus/

Wu, M.F., Brandhorst, H., Marinescu, M.C., Lopez, J.M., Hlava, M., and Busch, J. (2023). Automated metadata annotation: What is and is not possible with machine learning. *Data Intelligence,* 5(1), 122-138. https://doi.org/10.1162/dint_a_00162

Zeng, M.L. (2008). Knowledge Organization Systems (KOS). *Knowledge Organization*, 35(3/2): 160–82. https://doi.org/10.5771/0943-7444-2008-2-3-160

Appendix 1. Implementation Roadmap for the Recommendations

The Implementation Roadmap outlines the actionable steps required to build and sustain a robust vocabulary ecosystem in Australia, organised around and colour-coded by the four Strategic themes:

1. Foster a thriving vocabulary ecosystem;

2. Improve the consistency, content and context of vocabularies across Australia;

3. Improve the consistency of vocabulary tools and services; and

4. Lower barriers for using and maintaining vocabularies.

The roadmap below presents priorities for action (High, medium and low). The timeframe that each recommendation should be scheduled for action to begin is suggested, with recommendations to schedule commencement immediately (1 - 2 years), through to medium-term (2-3 years) and long-term (3-5 years).

1. Foster a thriving vocabulary ecosystem

Recommendation	Topic	Recommendation Description	Timeframe	Priority
1.d.1	d) Vocabulary governance	Support and advocate for persistent and sustainable vocabulary registers and services, including resourcing and contingency planning.	1. Short	1. High
1.d.2		Reduce the inertia experienced in vocabulary creation and maintenance (e.g., no one wanting to 'own' the vocabulary) by providing clear examples of roles, responsibilities and resources required.	2. Medium	2. Medium
1.f.1	f) Finding and using FAIR vocabularies	Understand the emergent role of vocabularies in AI optimisation and ethical data re-use, including in automated data tagging.	1. Short	1. High
1.g.1	g) Skills, training and outreach	Recognise the value of, and invest in, vocabulary professionals critical to the FAIR data ecosystem.	2. Medium	1. High

A Strategic Community Roadmap for an Australian FAIR Vocabulary Ecosystem. Wong, M., Atkins, N., Boyle, D., Dharmawardena, K., Hobern, D., McEachern, S., and Wyborn, L.

38

1.g.2		Encourage and support communities of practices for knowledge brokering, sharing, mentorship etc.	3. Long	2. Medium
1.g.3		Use tangible case studies and vignettes that demonstrate the value of FAIR vocabulary to stakeholder communities, including the cost-benefit compared with using local lists of data systems.	1. Short	1. High
1.g.4		Collaborate for co-investment across parties involved in the vocabulary lifecycle	2. Medium	2. Medium
1.g.5		Leverage working groups, including international efforts that cut across/address pressing issues in the vocabulary data supply chain.	1. Short	1. High
1.g.6		Establish communication and information dissemination channels across the ecosystem, such as through the Australian Vocabulary Special Interest Group (AVSIG) and the Australian Government Linked Data Working Group (AGLDWG)	1. Short	1. High

2. Improve the consistency, content and context of vocabularies across Australia

Strategic theme	Topic	Recommendation Description	Timeframe	Priority
2.a.1	a) Technical standards adoption	Promote the use of well-governed and established, internationally endorsed technical standards.	1. Short	1. High
2.b.1	b) Vocabulary content creation	Reduce vocabulary duplication by facilitating the search for existing resources, including across vocabulary services.	2. Medium	1. High
2.b.2		Allow FAIR vocabularies to be created at the local, regional, national and international levels (including at the project level if required), recognising that those at the international level can have broad impact and utility but that these can take considerable coordination effort.	2. Medium	2. Medium
2.b.3		Support collaborations to extend or revise existing vocabularies, including work at national and international levels to merge semantically equivalent vocabularies.	2. Medium	3. Low
2.b.4		Support the creation of multilingual vocabulary content, including languages and dialects.	3. Long	2. Medium
2.b.5		Support vocabulary creation and management workflows that respect the role of knowledge holders and domain experts (e.g., communities, committees,	2. Medium	2. Medium

A Strategic Community Roadmap for an Australian FAIR Vocabulary Ecosystem. Wong, M., Atkins, N., Boyle, D., Dharmawardena, K., Hobern, D., McEachern, S., and Wyborn, L.

39

		bodies or individuals) in maintaining the content of vocabularies as authoritative.		
2.b.6		Encourage the use of best practice standards in the 'profiling' of existing vocabularies	1. Short	2. Medium
2.c.1	c) Vocabulary publishing	Offer tools that enable easy publication of a metadata record for each vocabulary, ideally with a DOI, including metadata on content and scope, governance and maintenance.	1. Short	1. High
2.c.2		Enable vocabulary owners to express the licence and copyright for each vocabulary.	2. Medium	1. High
2.c.3		Vocabulary services should support the currency of vocabularies and mapping by enabling the easy management and curation by authorised parties, through community-driven and automated approaches.	2. Medium	2. Medium
2.d.1	d) Vocabulary governance	Engage with and respect the critical role of subject domain experts, committees or bodies in governing authoritative vocabularies.	2. Medium	2. Medium
2.d.2		Support Indigenous-led scoping and activity for vocabularies of, by and for Australian Aboriginal and Torres Strait Islanders, including for use in decolonising data, promoting the expression of CARE through the vocabulary lifecycle (including in vocabulary metadata).	1. Short	1. High
2.d.3		Promote authoritatively governed vocabularies within vocabulary services.	1. Short	1. High
2.d.4		Acknowledge that vocabularies are not neutral and support the right of reply, including in vocabulary notes, metadata and mappings.	2. Medium	1. High
2.d.5		The location of a vocabulary can change over time. Consider the responsible parties and resources required for ensuring that vocabulary GUPRIs consistently resolve to webpages containing the vocabulary information and metadata.	2. Medium	2. Medium
2.e.1	e) Mapping and annotation of vocabularies	Vocabulary services should include tools to provide additional text representations or labels for existing vocabulary terms.	1. Short	1. High
2.g.1	g) Skills, training and outreach	Foster linkage and collaboration with relevant international initiatives, for example, WorldFAIR+, RD-A Interoperable Descriptions of Observable Property Terminology Working Group (I-ADOPT), RD-A Vocabulary Services Interest Group, RD-A FAIR Mappings Working Group.	1. Short	1. High

A Strategic Community Roadmap for an Australian FAIR Vocabulary Ecosystem. Wong, M., Atkins, N., Boyle, D., Dharmawardena, K., Hobern, D., McEachern, S., and Wyborn, L.

40

2.g.2		Provide the necessary support for knowledge holders and domain experts (e.g., communities, committees, bodies or individuals) critical in the governance of vocabularies as authoritative resources	2. Medium	2. Medium

3. Improve the consistency of vocabulary tools and services

Strategic theme	Topic	Recommendation Description	Timeframe	Priority
3.b.1	b) Vocabulary content creation	Provide access to registers and services to handle the full spectrum of knowledge encoding, from simple controlled vocabularies through to more expressive ontologies, encoded using the appropriate technical standards and standard schema as appropriate (e.g., Data Catalogue Vocabulary (DCAT)).	2. Medium	1. High
3.b.2		Provide options for utilising open-source collaborative platforms such as GitHub for vocabulary creation and management.	2. Medium	2. Medium
3.c.1	c) Vocabulary publishing	Provide pathways to easily, efficiently and persistently publish vocabulary.	2. Medium	1. High
3.c.2		Promote good practice in GUPRI allocation by using dedicated, persistent identifier services. Consider integrating the harvesting of GUPRIs through vocabulary tools and services.	1. Short	1. High
3.c.3		Vocabulary services and registries must support version control and version management, including metadata and tools to assist mapping between versions.	1. Short	1. High
3.c.4		Vocabulary services must offer the ability to notify users of changes, such as deprecating and replacing a term.	3. Long	1. High
3.d.1	d) Vocabulary governance	Understand and promote the expression of vocabulary governance in vocabulary systems, tooling and workflows	3. Long	2. Medium
3.e.1	e) Mapping and annotation of vocabularies	Vocabulary services should include tools for providers and the community to define and publish mappings that establish the relationship between pairs of vocabulary terms using SSSOM, SKOS, OWL, FHIR or equivalent standards.	1. Short	1. High
3.e.2		Vocabulary services, or specialist mapping registries, should provide stable, persistent and FAIR hosting for these mappings and their metadata (e.g., provenance and authority).	2. Medium	2. Medium

A Strategic Community Roadmap for an Australian FAIR Vocabulary Ecosystem. Wong, M., Atkins, N., Boyle, D., Dharmawardena, K., Hobern, D., McEachern, S., and Wyborn, L.

41

3.e.3		Vocabulary services and registries should offer access to the mappings using an agreed set of standard formats and representations and via APIs and software packages that facilitate their use.	2. Medium	2. Medium
3.e.4		Vocabulary services and registries should enable the discovery of known mappings that reference each vocabulary.	3. Long	2. Medium
3.f.1	f) Finding and using FAIR vocabularies	Support standard API usage, including the development of vocabulary API with standard and extensible search mechanisms	1. Short	1. High
3.f.2		Provide human and machine-readable views of vocabularies, of varying complexities, to standard formats (content negotiation by profile)	2. Medium	1. High
3.f.3		Vocabulary services should facilitate re-use (forking) of vocabularies where this is necessary for technical or governance reasons, including establishing links to the terms in the source vocabulary.	3. Long	3. Low

4. Lower barriers for using and maintaining vocabularies.

Strategic theme	Topic	Recommendation Description	Timeframe	Priority
4.c.1	c) Vocabulary publishing	The ability to represent concepts in a form that accommodates the semantic complexity of the knowledge, and delivers it in a manner that is no more complicated than necessary for the use	1. Short	1. High
4.d.1	d) Vocabulary governance	Document and share scenarios of good practice and protocols for GUPRI construction, allocation and endorsement.	2. Medium	2. Medium
4.d.2		Provide guidance and reference examples on fit-for-purpose vocabulary governance - including for creation, maintenance, versioning, authorisation, sustainability and persistence.	2. Medium	2. Medium
4.e.1	e) Mapping and annotation of vocabularies	Vocabulary services or other associated software should facilitate semi-automated creation of (candidate) mappings based on term similarity, other mappings, etc., and practitioners should have the ability to edit and curate these candidate mappings.	2. Medium	2. Medium
4.e.2		Consider the role of AI in creating, mapping and annotating vocabularies.	1. Short	2. Medium
4.f.5	f) Finding and using FAIR vocabularies	Vocabulary services and registries should deliver information in a way that is user-friendly to search, discover and access the granularity of information they require.	1. Short	1. High

A Strategic Community Roadmap for an Australian FAIR Vocabulary Ecosystem. Wong, M., Atkins, N., Boyle, D., Dharmawardena, K., Hobern, D., McEachern, S., and Wyborn, L.

42

4.f.1		Enable users (both humans and machines) to easily find, access and reuse vocabularies (ie FAIR vocabularies)	2. Medium	1. High
4.f.2		Ensure vocabularies are appropriately licensed to enable and allow reuse	1. Short	1. High
4.f.3		Allow users to reference and cite vocabularies appropriately attributing curators and contributors.	2. Medium	2. Medium
4.f.4		Encourage interoperability between vocabulary services and registers where it will support vocabulary discovery and re-use within or across domains.	1. Short	2. Medium
4.f.6		Vocabulary services and registries should include tools and APIs that allow vocabularies to be accessed and used directly within popular programming languages and software (e.g., spreadsheet and database tools).	3. Long	2. Medium
4.f.7		Support the seamless integration of vocabularies into data management workflows and user interfaces to increase vocabulary reuse.	2. Medium	2. Medium
4.g.1	g) Skills, training and outreach	Provide a central touchpoint for those seeking guidance on the various tools and services for finding, creating, publishing and re-using vocabularies.	2. Medium	2. Medium
4.g.2		Support the democratisation of vocabulary reuse through providing training and workshops on vocabularies tailored to communities' and domains' needs. Provide guidance on finding vocabularies across services and evaluating them as fit-for-purpose and likely persistent for reuse.	2. Medium	2. Medium
4.g.3		Access to a body of knowledge for education on vocabulary types, standards, best practices, guidelines, and examples critical to the FAIR vocabulary lifecycle.	2. Medium	2. Medium
4.g.4		Advocate for vocabularies as first-class 'digital citizens' in their own right, including support for citation and indexing.	1. Short	2. Medium

A Strategic Community Roadmap for an Australian FAIR Vocabulary Ecosystem. Wong, M., Atkins, N., Boyle, D., Dharmawardena, K., Hobern, D., McEachern, S., and Wyborn, L.

43

Appendix 2. More Detailed Explanations of Individual Recommendations

In this appendix, we include various supporting ideas presented at different forums for individual recommendations; some are more comprehensive than others. These ideas can be used in scoping activities for implementing the recommendations.

As a reminder, each recommendation is coded as follows: the first number is the theme, the second is the topic, and the third number is the number of the recommendation within that topic.

a) Technical standards adoption

Recommendation 2.a.1: Promote the use of well-governed and established, internationally endorsed technical standards

Technical standards are endorsed by major International standards bodies such as the International Organisation for Standardisation (ISO; ISO n.d), Open Geospatial Consortium (OGC, OGC n.d.), World Wide Web Consortium (W3C, W3C n.d.-b), or Institute of Electrical and Electronics Engineers (IEEE; IEEE, n.d.). They are largely domain agnostic and are foundational to being able to integrate data across domains. The creation and governance of technical standards are managed mainly at the international level by standards authorities and are beyond the scope of this paper.

It is critical to the operation of the Australian Vocabulary Ecosystem that vocabulary and service creators, users, managers and implementers are supported in practices that use well-governed and established, internationally endorsed technical standards. Technical standards in the context of vocabularies are often referred to as 'Semantic Standards' (W3C, n.d.-a).

Some commonly adopted technical standards for representing and delivering vocabularies include:

- RDF: Resource Description Framework (RDF Working Group, 2014) and extensions to RDF Schema (RDFS) (Brickley and Guha, 2014; Tomaszuk and Haudebourg 2025);
- Description of RDF data using semantic standards such as SKOS (Simple Knowledge Organisation System, Miles and Bechhofer, 2009; Baker et al., 2013b) and OWL (Web Ontology Language: Hitzler et al., 2012; W3C OWL Working Group, 2012; Cox et al., 2021), and widely adopted schemas such as schema.org (Guha et al., 2018), and domain standards e.g. science-on-schema.org for geoscience/geospatial data (Shepherd et al., 2022) and FHIR for health care data exchange (HL7 - n.d.);
- Standard serialisations for linked data such as Turtle (Beckett et al., 2014); RDFa (Herman et al., 2015); RDF/XML (Gandon and Schreiber, 2014); and JavaScript Object Notation for Linked Data (JSON-LD, n.d.; Sporny et al., 2020);
- DCAT (Data Catalog Vocabulary for the Web), an RDF vocabulary designed to facilitate interoperability between data catalogues published on the Web (Albertoni et al., 2024);
- XML (eXtensible Markup Language) (Bray et al., 2009);

A Strategic Community Roadmap for an Australian FAIR Vocabulary Ecosystem. Wong, M., Atkins, N., Boyle, D., Dharmawardena, K., Hobern, D., McEachern, S., and Wyborn, L.

44

- ISO 704: Terminology work — Principles and methods (ISO, 2022); and
- I-ADOPT (InteroperAble Descriptions of Observable Property Terminologies (Magagna et al., 2021, 2022).

For persistent identification on the web, Hugo et al. (2020) state that "Globally Unique, Persistent and Resolvable Identifiers must be used for Semantic Artefacts, their content and their versions" (P-Rec. 1). This includes reference to several relevant recommendations, for example Data on the Web Best Practices (Lóscio et al., 2017) that state:

- Reuse vocabularies, preferably standardized ones (Best Practice 15);
- Use persistent URIs as identifiers of datasets (Best Practice 9);
- Use persistent URIs as identifiers within datasets (Best Practice 10);
- Assign URIs to dataset versions and series (Best Practice 11) ; and
- Use content negotiation for serving data available in multiple formats (Best Practice 19).

For the raw vocabulary data, ideally, both Human-Readable (e.g., HTML documentation, specification pages) and Machine-Readable representations (RDF/XML, Turtle, JSON-LD) can be served from the same IRI using HTTP Content Negotiation, with separate stable URLs (e.g. DOIs) for documentation, indexing and citation.

b) Vocabulary content creation

Recommendation 2.b.1: Reduce vocabulary duplication by facilitating the search for existing resources, including across vocabulary services.

Quite often, duplicates are created because individual vocabularies themselves are not FAIR and cannot be discovered or reused. As well as providing the support and tools to make vocabularies FAIR, registers and services are critical to enabling discovery, sharing and reuse of any form of vocabulary. Dedicated work is required to overcome the challenges of vocabulary service federation. For example, many vocabulary services lack standards-compliant metadata records that enable searching both within and across vocabulary services. One such example is that EcoPortal has adopted MOD - the Metadata for Ontology Description and Publication (Dutta et al., 2017).

Recommendation 2.b.2: Allow vocabularies to be created at the local, regional, national and international levels (including at the project level if required), recognising that those at the international level can have a broad impact and utility but that these can take considerable coordination effort.

It is recognised that any stakeholder community involved in creating and using vocabulary resources can be highly variable in skills and can be geographically distributed, ranging from International, National, Regional and Local Communities (Veltmann, 2001; UNESCO, 2022). It can also take years to develop agreed content standards that are widely accepted by the international community.

Prent et al. (2024), who reviewed the types and sizes of communities that were developing vocabularies to describe geochemical data, defined four levels of community engagement (Figure 5):

- The Local community, which can consist of a single institution or group;
- The Regional community would be a combination of groups from universities, government agencies, a national standards group, or a group of discipline experts;
- An international community would involve collaboration between two or more countries; and
- A global community acts on behalf of all communities. It is endorsed by an International Authoritative body such as an International Science Union, Association or Society, or an International Standards Body.
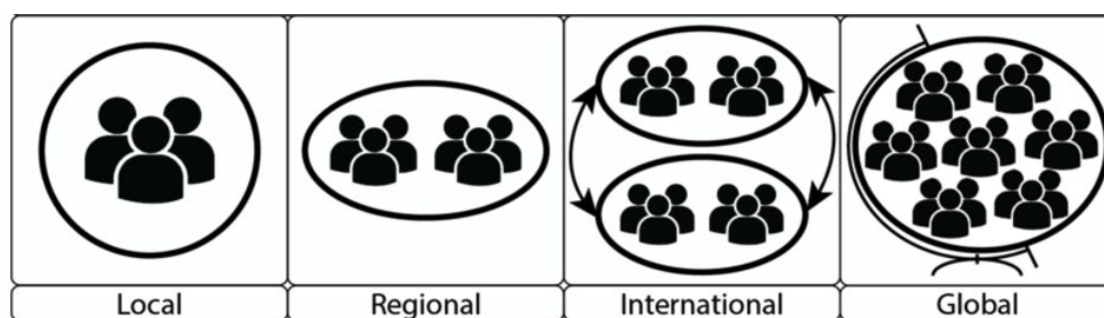
A Strategic Community Roadmap for an Australian FAIR Vocabulary Ecosystem. Wong, M., Atkins, N., Boyle, D., Dharmawardena, K., Hobern, D., McEachern, S., and Wyborn, L.

45

Figure 5. The communities involved in developing vocabularies in geochemistry
(Source: Prent et al., 2024).

Geochemical data is typical of long-tail data communities (Heidorn, 2008) and is generated by many small, disconnected groups around the globe and across many sectors (research, government and industry). As it will be years before globally agreed, endorsed vocabularies can be agreed to, Prent et al. (2024) recommended publication of any semantic resource that provides a guide to the definition, meaning and/or relationships of any term used with a published dataset no matter what the size of the community was that created it and will enable the dataset itself to become compliant with Principle I2 of Wilkinson et al. (2016). As semantically equivalent, internationally endorsed standards become available, the URIs used in the local vocabularies can be redirected.

Recommendation 2.b.3: Support collaborations to extend or revise existing vocabularies, including work at national and international levels to merge semantically equivalent vocabularies.
Once developed, publishing and then maintaining internationally agreed vocabularies requires dedicated funding and time resources to be effective. Currently, few resources are available for this work; it is mainly done voluntarily. In Australia, prohibitive travel costs and inconvenient time zones make participation in most international meetings, either in person/or virtual, very difficult.

Recommendation 2.b.4: Support creation of multilingual vocabulary content, including Indigenous languages and dialects.
Multilingual translations of content, including those in Indigenous languages, are crucial for enabling access to and sharing of data and information across geographic and cultural boundaries. This was advocated in the UNESCO Recommendation on Open Science (UNESCO, 2022), which also emphasises the need to ensure that multilingual scientific knowledge is openly available, accessible, and reusable to ensure that members of a wider community are involved including citizen scientists and Indigenous knowledge-holders (Amano et al., 2021; Specht et al., 2024). Good translations of scientific content into multiple languages should occur at the concept level and be done by the domain experts in each language, rather than as a simple word-for-word translation (David et al., 2022). Most online generic automatic translation tools simply translate each word, and this could also be a limitation of many online domain-specific translation tools. Each language version should be published and linked to the source version, allowing any updates to be propagated back into each language version.

Recommendation 2.b.5: Support vocabulary creation and management workflows that respect the role of knowledge holders and domain experts (e.g., communities, committees, bodies or individuals) in maintaining the content of vocabularies as authoritative.

A Strategic Community Roadmap for an Australian FAIR Vocabulary Ecosystem. Wong, M., Atkins, N., Boyle, D., Dharmawardena, K., Hobern, D., McEachern, S., and Wyborn, L.

46

Developing internationally endorsed standards can be a lengthy process, and for content standards, it is best done under the support and endorsement of International Unions, Professional Societies and equivalent bodies (Ma et al., 2022). The metadata of any vocabulary should indicate who supports and endorses a particular vocabulary and recommends its use.

Recommendation 2.b.6: Encourage the use of best practice standards in the 'profiling' of existing vocabularies.

Many vocabulary content standards are very comprehensive, listing definitions, conditionality for each element (mandatory/optional or true/false), and other variables. A content standard may contain many more elements than are necessary to define the metadata/data in a particular dataset/data collection, whilst the required elements may reside within more than one published content standard. In this case, a specific 'profile' can be created and then published and made FAIR.

A profile is defined as "*a set of one or more base standards or subsets of base standards, and, where applicable, the identification of chosen clauses, classes, options and parameters of those base standards, that are necessary for accomplishing a particular function*" (ISO, 2004). "*A profile is derived from base standards so that by definition, conformance to a profile is conformance to the base standards from which it is derived*" (ISO, 2004, Section 4.5). Best practices in creating, describing, and publishing profiles of existing vocabularies are defined in ISO 19106:2004, Geographic Information – Profiles (ISO, 2004), and in Atkinson et al. (2024). Essentially, a profile can use terms, definitions and IRIs from one or multiple source vocabularies. Conditionality of a select set of metadata elements may be raised but never reduced.

Examples of published vocabulary profiles include:
- The ISO 19115-1 metadata standard for fundamentals of Geographic Metadata has over 300 elements, most of which are optional (ISO, 2014). The Geoscience Australia profile of this standard (Bastrakova, 2018) contains 26 elements of ISO 19115-1, and the conditionality of some elements has been raised;
- The Terrestrial Ecosystems Research Network (TERN; TERN, n.d.) have published profiles of ISO 19115-3:2016 and of ISO 19157-2:2016 (Mahuika et al., 2021); and
- The mt_metadata standard, a metadata schema for magnetotelluric geophysical data acquisition and processing, currently has 370 defined elements and conditionalities (Peacock, 2021, 2022). Many of these elements are not relevant to legacy data collections. Goleby et al. (2023) have published a profile of the mt_metadata standard that contains 75 elements relevant to Australian legacy datasets funded by AuScope.

Recommendation 3.b.1: Provide access to registers and services to handle the full spectrum of knowledge encoding, from simple controlled vocabularies through to more expressive ontologies, encoded using the appropriate technical standards and standard schema as applicable (e.g., Data Catalogue Vocabulary (DCAT)).

There are a few service providers of vocabulary infrastructures that can enable the range of semantic resources (Figure 1) to be entirely FAIR compliant, particularly in Australia. For example, in the Australian Research Community, the ARDC Research Vocabularies Australia (RVA: ARDC, n.d.-c) currently (2025) only hosts simple SKOS vocabularies that contain values for each term, where each value is a label for a SKOS concept, and each concept has a definition and an IRI. RVA presently does not take thesauri, taxonomies, terminology mappings, ontologies, metadata schemes, subject heading systems, Climate and Forecasting (CF) Conventions (CF Metadata Conventions, n.d.), identifier schemas and other metadata schemas that are required for describing Research input

A Strategic Community Roadmap for an Australian FAIR Vocabulary Ecosystem. Wong, M., Atkins, N., Boyle, D., Dharmawardena, K., Hobern, D., McEachern, S., and Wyborn, L.

47

artefacts and ensuring that they are fully compliant with the FAIR principles of Wilkinson et al. (2016), nor does it take mappings and crosswalks.

To address these gaps for the Australian medical and clinical research domain, in 2024, ARDC trialled extending the Research Vocabulary Australia (RVA) service to boost support for health and medical research (ARDC, n.d.-d). This extension, termed Research Terminologies Australia (Boyle et al., 2023; Belton et al., 2024), focused on the gap in the data market by enabling the curation, publication and reuse of critical information resources for medical and clinical data interoperability, including:

- Curation of a look-up list: for example, a general practitioner database list of drugs and internal ID numbers associated with them;
- Creating custom mappings: e.g., free text strings to standard terminology such as RxNORM, a medication standard (NLM, n.d.) or custom terminology mapping to a standard terminology such as SNOMED CT (Systematised Nomenclature of Medicine Clinical Terms, SNOWMED, n.d.); and
- Text mapping to a Single concept: e.g., free text diagnosis of 'Type 2 Diabetes' to a single code/definition such as SNOMED-CT term codes, for example 'T2DM', 'Type two diabetes', 'NIDDM'.

There are increasingly more examples of vocabulary registers and services globally, such as listed in BARTOC (BARTOC.org, n.d.). These include generic vocabulary services such as Linked Open Vocabularies (LOV; LOV, n.d.) , and domain-specific repositories and services including the ESIP Community Ontology Repository (ESIPFed, n.d.-c), the UK National Environmental Research Council (NERC) Vocabulary Server (NERC, n.d.), OntoPortal and its various instances, including BioPortal for biomedical ontologies (Bioportal, n.d.), Agroportal (FAO, n.d.; Jonquet et al., 2018) and Ecoportal (Lifewatch, n.d.; Turallo et al., 2024). For metadata schemas there is the RD-A Metadata Schema Catalog (RD-A Metadata Standards Catalog, n.d.) and the EOSC Metadata Schema and Cross Walk Registry (Faircore4eosc, n.d.).

Recommendation 3.b.2: Provide options for utilising open-source collaborative platforms such as GitHub for vocabulary creation and management.
Many tools exist to support components of creating and managing vocabularies. Choosing the tool that is fit for purpose and suitable for the available level of expertise is challenging. Open-Source tools include methods to use GitHub for the creation and management of vocabularies (e.g., Crystal-Ornelas et al., 2021; Chuc and Car, 2024).

c) Vocabulary publishing

Recommendation 2.c.1: Offer tools that enable easy publication of a metadata record for each vocabulary, ideally with a DOI, including metadata on content and scope, governance and maintenance.
As far as possible, tools for this purpose should not require special understanding of knowledge engineering concepts and technologies. Spreadsheets and simple form-based interfaces with clear guidance are likely to be most appropriate. Resulting issues will need to be detected and diagnosed on entry/upload, and clear assistance must be given on how to fix them. Documentation and clear reference examples should target the needs of all users.

A Strategic Community Roadmap for an Australian FAIR Vocabulary Ecosystem. Wong, M., Atkins, N., Boyle, D., Dharmawardena, K., Hobern, D., McEachern, S., and Wyborn, L.

48

Recommendation 2.c.2: Enable vocabulary owners to express the licence and copyright for each vocabulary.
Each published vocabulary must have a human and a machine-readable licence and copyright statement (Recommendation P-Rec. 16, Hugo et al., 2020) and preferably an open license for users (e.g., CC0 or CC-BY).

If a legacy vocabulary is to be republished, it is essential that it is possible to produce a derivative product and that the conditions for derivation are known. It may be necessary to contact the copyright holder directly for a list of terms to be republished (Hugo et al., 2020; Cox et al., 2021).

Recommendation 2.c.3: Vocabulary services should support the currency of vocabularies and mapping by enabling the easy management and curation by authorised parties, through community-driven and automated approaches.

Once a vocabulary is published through a service, it is essential that only authorised parties can update that vocabulary.

Recommendation 3.c.1: Provide pathways to easily, efficiently and persistently publish a vocabulary.
Many researchers find both publishing and updating their vocabularies an onerous task. As they may not do this regularly, it must be easy to publish and modify a vocabulary. The alternative is to have dedicated staff whose role is to either assist staff in publishing vocabularies or offer a service to publish vocabularies.

Recommendation 3.c.2: Promote good practice in GUPRI allocation by using dedicated, persistent identifier services. Consider integrating the harvesting of GUPRIs through vocabulary tools and services.
Globally unique, persistent and resolvable identifiers should be assigned to (a) the vocabulary and (b) each term in the vocabulary (Cox et al., 2021) and to their versions (P-Rec.1, Hugo et al., 2020), to support the retrieval of both the metadata of the vocabulary and of each term. Assigning globally unique, persistent and resolvable identifiers (GUPRIs) will facilitate vocabulary harvesting via standardised tools and services into vocabulary registries, which in turn will increase their findability.

Recommendation 3.c.3: Vocabulary services and registries must support version control and version management, including metadata and tools to assist mapping between versions.
and
Recommendation 3.c.4: Vocabulary services must offer the ability to notify users of changes, such as deprecating and replacing a term.
This relates to the maturity dimension of Transparency (Corcho et al. 2024) for resource versioning within a workflow for data curation; records on previous versions of vocabularies must be made. See Jonquet and Poveda-Villalón (2023) for recommendations regarding versioning principles. It also relates to recommendation P-Rec.1 (Hugo et al. 2020) that all versions of vocabularies and their concepts should have GUPRIs. This enables versioning systems for ensuring backward compatibility and documentation when a change is made, including deprecations and replacements.

Recommendation 4.c.1: The ability to represent concepts in a form that accommodates the semantic complexity of the knowledge, and delivers it in a manner that is no more complicated than necessary for the user.
For some use cases, particularly where the community of data producers is well-defined and connected, it may be possible to integrate and/or interpret a dataset using simple code lists without

A Strategic Community Roadmap for an Australian FAIR Vocabulary Ecosystem. Wong, M., Atkins, N., Boyle, D., Dharmawardena, K., Hobern, D., McEachern, S., and Wyborn, L.

49

requiring a formal ontology. It is essential to understand what can be done with a vocabulary as a simple list, and what use cases require greater knowledge representation and reasoning capabilities (Figure 1).

d) Vocabulary governance

Recommendation 1.d.1: Support and advocate for persistent and sustainable vocabulary registers and services, including resourcing and contingency planning.
Vocabularies must persist throughout the life of the datasets, databases, services, and related implementations that use and rely on them. Therefore, establishing, continually updating, improving and persistently sustaining a vocabulary register and service over time requires long-term funding and support.

There are software stacks that enable the creation of local- and/or project-based registries and services, but there have been cases of repositories that have ceased operations, without organising any redirects, leaving users with loss of access to content. Hence, when a repository or vocabulary service is established, they should have business models for sustainability (Dillo et al., 2017) and planning for contingencies and emergencies.

Corcho et al. 2024 state that, as part of a maturity model for a catalogue of semantic artifacts, recommendations for governance and processes for preserving and maintaining the artifacts are required. They include an assessment of the sustainability of the models in place to sustain services financially and preserve the catalogue over the longer term.

Criteria for trustworthy data repositories (e.g., CoreTrustSeal (CoreTrustSeal, n.d.); the TRUST principles (Lin et al., 2020)), could be adapted and applied to semantic repositories, so that the both the developer and the user of the semantic resource can have trust that any semantic resource that they have published or used will persist and improve over the lifetime of their data resources, which in longitudinal research can be over decades.

Trustworthy data repositories, focused on specific disciplines, are available for data curation and preservation, and these repositories may be amenable to preservation of copies of vocabulary data files in the short term. In the medium term, development of criteria for Trustworthy Semantic Repositories is a community responsibility, as is the establishment of mechanisms for certification of compliance. This recommendation does not endorse any specific technology. Still, it emphasises the necessity to share, publish, and preserve vocabularies in such repositories to improve both findability and reuse over time.

Recommendation 1.d.2: Reduce the inertia experienced in vocabulary creation and maintenance (e.g., no one wanting to 'own' the vocabulary) by providing clear examples of roles, responsibilities and resources required.
Vocabulary creation and maintenance are essential for a robust Vocabulary Ecosystem, but often stall due to unclear roles, responsibilities, and resource allocation. To overcome this, clear examples and guidelines are needed to empower parties and promote collaboration. By defining key parties, their roles, and required resources, we can establish a framework for vocabulary governance that supports the development and maintenance of high-quality vocabularies. An Australian workshop supported through CODATA, ARDC and the Data Documentation Initiative (DDI) produced a snapshot of key concerns for the governance of FAIR vocabularies, including key stakeholders and their roles

A Strategic Community Roadmap for an Australian FAIR Vocabulary Ecosystem. Wong, M., Atkins, N., Boyle, D., Dharmawardena, K., Hobern, D., McEachern, S., and Wyborn, L.

50

(Wackerow et al., 2021). Ramezani et al. (2024) identify the types of actors involved in semantic artifact governance, synthesise overarching critical components of governance frameworks (principles, standards, and quality), and propose governance models for pre-defined target groups.

Recommendation 2.d.1: Engage with and respect the critical role of subject domain experts, committees or bodies in endorsing and governing authoritative vocabularies.

For a vocabulary to be suitable for the widest adoption, it needs to be legitimised through ownership by a trusted community that adequately and transparently represents the needs of the stakeholder community — broader representation, and where relevant international contributions, will increase both trust and sustainability. Ma et al. (2022) advocate for Science Unions to be more involved in the development and endorsement of domain-specific vocabularies.

Recommendation 2.d.2: Support Indigenous-led scoping and activity for vocabularies of, by and for Australian Aboriginal and Torres Strait Islanders, including for use in decolonising data, promoting the expression of CARE through the vocabulary lifecycle (including in vocabulary metadata).

The CARE Principles for Indigenous Data Governance address "important considerations in modern data ecosystems and across data lifecycles that support both innovation and Indigenous self-determination" (Carroll et al., 2020); thus, the rights and interests of Indigenous Peoples require consideration through the life cycle of vocabularies.

Particularly relevant to vocabularies is CARE Principle E3: Ethics - For future use. This states "Data governance should take into account the potential future use and future harm based on ethical frameworks grounded in the values and principles of the relevant Indigenous community" and that "Metadata should acknowledge the provenance and purpose and any limitations or obligations in secondary use, inclusive of issues of consent." Vocabularies are used to describe and contextualise data. They are also a type of dataset. Thus, CARE should be taken in the governance and use of vocabularies about, for and by Australian Aboriginal and Torres Strait Islanders.

Recommendation 2.d.3: Promote authoritatively governed vocabularies within vocabulary services.

Automated scoring of vocabularies based on FAIR metrics, documented governance processes, continuity of maintenance communities, past use in multiple datasets and applications, and peer-review mechanisms can all contribute to these evaluations. Recommenders also have value.

Recommendation 2.d.4: Acknowledge that vocabularies are not neutral and support the right of reply, including in vocabulary notes, metadata and mappings.

Vocabularies that are intended for adoption and reuse by broader audiences should have established and well-documented mechanisms for community members to contribute to the future of the vocabulary. Software communities such as the Apache Software Foundation and GitHub offer examples of how tools can support collaborative development.

Recommendation 2.d.5: The location of a vocabulary can change over time. Consider the responsible parties and resources required for ensuring that vocabulary GUPRIs consistently resolve to webpages containing the vocabulary information and metadata.

and

Recommendation 3.d.1: Understand and promote the expression of vocabulary governance in vocabulary systems, tooling and workflows

Many vocabulary infrastructures and services cannot document governance alongside the vocabularies, which hinders trust in the use of the vocabularies and discourages vocabulary

A Strategic Community Roadmap for an Australian FAIR Vocabulary Ecosystem. Wong, M., Atkins, N., Boyle, D., Dharmawardena, K., Hobern, D., McEachern, S., and Wyborn, L.

51

professionals from establishing effective governance and reporting it. Vocabularies require sustainable governance that has dedicated personnel and resources to continually ensure the vocabularies are up-to-date and can accommodate location changes, as well as updates to the underlying software. Additionally, tooling and workflows must be updated as required.

Recommendation 4.d.1: Document and share scenarios of good practice and protocols for GUPRI construction, allocation and endorsement.
and
Recommendation 4.d.2: Provide guidance and reference examples on fit-for-purpose vocabulary governance - including for creation, maintenance, versioning, authorisation, sustainability and persistence.
One of the strengths of an interconnected Vocabulary Ecosystem should be the ability to facilitate structural changes that make it easier for users to find and select the best existing vocabularies for their purposes, to contribute improvements where necessary, and to develop new vocabularies only when no suitable equivalents exist. Just as for publishing new vocabularies, the barriers should also be low for domain knowledge-holders to contribute their expertise to improve an existing vocabulary.

e) Mapping and annotation of vocabularies

Recommendation 2.e.1: Vocabulary services should include tools to provide additional text representations or labels for existing vocabulary terms.
and
Recommendation 3.e.1: Vocabulary services should include tools for providers and the community to define and publish mappings that establish the relationship between pairs of vocabulary terms using SSSOM, SKOS, OWL, FHIR or other well-established technical standards.
Standards and tools are already emerging to facilitate or enable annotation and mappings. For example, relevant work carried out under SSSOM (Simple Standard for Sharing Ontological Mappings; Matentzoglu et al., 2022) and the proposed RD-A FAIR Mappings Working Group (RD-A FAIR Mappings WG, n.d.). SSSOM can be adopted as a standard model for representing all categories of mapping discussed here, along with the metadata on sets of mappings (MappingSets) and on registries of mapping sets (MappingRegistries).

Recommendation 3.e.2: Vocabulary services, or specialist mapping registries, should provide stable, persistent and FAIR hosting for these mappings and their metadata (e.g., provenance and authority).
and
Recommendation 3.e.3: Vocabulary services and registries should offer access to the mappings using an agreed set of standard formats and representations and via APIs and software packages that facilitate their use.
and
Recommendation 3.e.4: Vocabulary services and registries should enable discovery of known mappings that reference each vocabulary.
Most vocabulary services are developed in isolation, and there are no agreed interconnections between them. Often, when an organisation chooses to adopt a subset of terms from multiple vocabulary services, it is usually done as an ad hoc copy-paste, with poor documentation of provenance and version tracking; in some cases, new PIDs are minted for the same term.

A Strategic Community Roadmap for an Australian FAIR Vocabulary Ecosystem. Wong, M., Atkins, N., Boyle, D., Dharmawardena, K., Hobern, D., McEachern, S., and Wyborn, L.

52

There are many challenges in federating vocabulary services and the semantic resources they deliver. Dedicated work towards solving the challenges of vocabulary federation is required, in a way that focuses on articulating a shared vision and enabling discovery, access and reuse of vocabularies and concepts within individual vocabularies.

Recommendation 4.e.1: Vocabulary services or other associated software should facilitate semi-automated creation of (candidate) mappings based on term similarity, other mappings, etc., and practitioners should have the ability to edit and curate these candidate mappings.
and
Recommendation 4.e.2: Consider the feasibility of AI for creating mappings and annotating vocabularies.
Current methods of mapping and annotation are mainly manual, but automated methods are emerging (e.g., Wu et al., 2025). Investment in projects that propose automated mapping and annotation should be encouraged.

f) Finding and using FAIR vocabularies

Recommendation 1.f.11: Understand the emergent role of vocabularies in AI optimisation and ethical data re-use, including in automated data tagging.
Including how AI can and could be integrated with vocabularies to improve data quality, integration, and reasoning, thereby enhancing the utility of research data and services. Additionally, vocabularies can support ethical data reuse, for example, through the standardised machine-readable expression of permissions, restrictions, and provenance.

Recommendation 3.f.2: Provide human and machine-readable views of vocabularies, of varying complexities, to standard formats (content negotiation by profile).
This requirement is essential to making vocabularies FAIR.

Recommendation 3.f.3: Vocabulary services should facilitate re-use (forking) of vocabularies where this is necessary for technical or governance reasons, including establishing links to the terms in the source vocabulary.
This is a requirement for building vocabulary services

Recommendation 4.f.1: Enable users (both humans and machines) to easily find, access and reuse vocabularies (ie FAIR vocabularies).
This broad recommendation aligns with many of the recommendations outlined in this roadmap. Vocabularies themselves need to be Findable by both humans and machines and need to be assigned a globally unique and persistent identifier (F1), described with rich metadata (F2), that includes the identifier of the data it describes (F3). To enhance their discoverability, they must be registered or indexed in a searchable resource (F4).

Recommendation 4.f.2: Ensure vocabularies are appropriately licensed to enable and allow reuse.
With the content owner/copyright holder, agree on the license for the FAIR vocabulary, preferably an open license for users (e.g., CC0 or CC-BY). If you are reusing a vocabulary, it may not have a licence. If it does have a licence and you are making an existing vocabulary FAIR, or making a profile of an existing vocabulary, you must take into account the licence of the source vocabulary (Hugo et al., 2020; Cox et al., 2021). If you are making a profile from multiple vocabularies that have different

A Strategic Community Roadmap for an Australian FAIR Vocabulary Ecosystem. Wong, M., Atkins, N., Boyle, D., Dharmawardena, K., Hobern, D., McEachern, S., and Wyborn, L.

53

licences, the final profile must take on the most restrictive licence of all the source vocabularies (ARDC, 2019).

Recommendation 4.f.3: Allow users to reference and cite vocabularies appropriately, attributing curators and contributors.
Users need to be able to reference and cite vocabularies, which will enable the tracking of adoption and subsequent usage, including:

- The ability to include links (relatedIdentifiers or similar) to (external, public) vocabularies from the metadata for datasets that use the vocabulary; and
- GUPRIs for each vocabulary itself (such as DOIs) to facilitate discoverability and enable indexation and citation.

Recommendation 4.f.4: Encourage interoperability between vocabulary services and registers to support vocabulary discovery and reuse within or across domains.
and
Recommendation 3.f.1: Support standard API usage, including the development of vocabulary APIs with standard and extensible search mechanisms.
These recommendations point to the need to continue to overcome vocabulary service federation challenges (Jonquet et al., 2025), including potentially defining and communicating a standard architecture ('best practice') for implementation of infrastructures (Best Practice 13, Hugo et al. 2020).

Recommendation 4.f.5: Vocabulary services and registries should deliver information in a way that is user-friendly to search, discover and access the granularity of information they require.
and
Recommendation 4.f.6: Vocabulary services and registries should include tools and APIs that allow vocabularies to be accessed and used directly within popular programming languages and software (e.g., spreadsheet and database tools).
These recommendations relate to requirements for building a user-friendly vocabulary service and registry. They relate to the Technology requirements of mature semantic artifact catalogues to "enable users to have a better experience in exploring the data, such as REST APIs, Web search interfaces and SPARQL endpoints, etc" (Corcho et al., 2024).

Recommendation 4.f.7: Support the seamless integration of vocabularies into data management workflows and user interfaces to increase vocabulary reuse.
Integrating vocabularies into workflows would increase their reuse. Raised was the requirement to point to and access authoritative vocabularies through machine-to-machine interoperability - rather than downloading, adapting, and using them locally. This included the need to share tools and libraries for interacting with vocabularies (e.g., Python libraries, R packages and Excel plug-ins).

g) Skills, training and outreach

Recommendation 1.g.1: Recognise the value of, and invest in, vocabulary professionals critical to the FAIR Vocabulary Ecosystem.
Organisations should recognise the importance of vocabulary professionals in their organisations. Position descriptions that describe the required skills, roles and functions could be shared.

Recommendation 1.g.2: Encourage and support communities of practices for knowledge brokering, sharing, mentorship, etc.

A Strategic Community Roadmap for an Australian FAIR Vocabulary Ecosystem. Wong, M., Atkins, N., Boyle, D., Dharmawardena, K., Hobern, D., McEachern, S., and Wyborn, L.

54

Those involved in the vocabulary lifecycle work across various sectors, including academia and government, and need to rely on external networks for assistance. Those working with vocabularies have responsibilities encompassing tasks such as creating, updating, discovering, disseminating, and governing vocabularies. This can require conceptual, informatics, technical and sometimes content skills and training, including a grasp of technical tools, workflows and infrastructures. Creating communities of practice amongst vocabulary professionals to share experiences and offer mentoring is desirable.

Recommendation 1.g.3: Use tangible case studies and vignettes that demonstrate the value of FAIR vocabularies to stakeholder communities, including the cost-benefit compared with using local lists of data systems.
To support existing professionals, documented case studies/research papers of the return on investment in making vocabularies FAIR are required.

Recommendation 1.g.4: Collaborate for co-investment across parties involved in the vocabulary lifecycle.
As the investment in developing new vocabulary technologies and skills is high, to accelerate change, parties could consider co-investment across multiple organisations. This could include participation in collaborative external projects (e.g., NCRIS National Digital Research Infrastructure projects).

Recommendation 1.g.5: Leverage working groups, including international efforts that cut across/address pressing issues in the vocabulary data supply chain.
and
Recommendation 1.g.6: Establish communication and information dissemination channels in Australia across the ecosystem, such as through the Australian Vocabulary Special Interest Group (AVSIG) and the Australian Government Linked Data Working Group (AGLDWG).
and
Recommendation 2.g.1: Foster linkage and collaboration with relevant international initiatives, for example, WorldFAIR+, RD-A Interoperable Descriptions of Observable Property Terminology Working Group (I-ADOPT WG), RD-A Vocabulary Services Interest Group, RD-A FAIR Mappings Working Group.
Community Working Groups, Interest Groups and other communities of practice exist both nationally and internationally to develop new technologies and address pressing issues in the operationalisation of various components of the Vocabulary Ecosystem. They also enhance and accelerate communication and information dissemination channels across the ecosystem. Leading Australian communities of practice include the Australian Vocabulary Special Interest Group (AVSIG, ARDC, n.d.-e) and the Australian Government Linked Data Working Group (AGLDWG, AGLDWG, n.d.), whilst international groups include WorldFAIR Plus (WorldFAIR, n.d.-b), RD-A Interoperable Descriptions of Observable Property Terminology Working Group (I-ADOPT: RD-A I-ADOPT WG, n.d), RD-A Vocabulary Services Interest Group (RD-A Vocabulary Services IG, n.d.) and RD-A FAIR Mappings WG (RD-A FAIR Mappings WG, n.d.). There is a need to encourage greater participation of Australians in these, particularly the international ones, although time zone constraints and the cost of travel can make participation difficult.

Recommendation 2.g.2: Provide the necessary support for knowledge holders and domain experts (e.g., communities, committees, bodies or individuals) critical in the governance of vocabularies as authoritative resources.

A Strategic Community Roadmap for an Australian FAIR Vocabulary Ecosystem. Wong, M., Atkins, N., Boyle, D., Dharmawardena, K., Hobern, D., McEachern, S., and Wyborn, L.

55

Knowledge domain experts and field participants may need to be engaged to create and manage vocabularies for specific purposes. In this kind of engagement, those working to develop, manage and deliver FAIR vocabularies can play a pseudo 'business analyst' role, where the key frameworks, activities and concepts of a domain are elicited, extracted, and documented. The end users and potential re-users of vocabularies require a basic understanding of the value in applying vocabularies in their data assets and systems. They need skills in finding, selecting, and applying vocabularies.

Recommendation 4.g.1: Provide a central touchpoint for those seeking guidance on the various tools and services for finding, creating, publishing and re-using vocabularies.
Develop community websites dedicated to this or enhance existing sites such as the AVSIG community site (ARDC, n.d.-e).

Recommendation 4.g.2: Support the democratisation of vocabulary reuse through providing training and workshops on vocabularies tailored to communities' and domains' needs. Provide guidance on finding vocabularies across services and evaluating them as fit-for-purpose and likely persistent for reuse.
and
Recommendation 4.g.3: Provide access to a body of knowledge for education on vocabulary types, standards, best practices, guidelines, and examples critical to the FAIR vocabulary lifecycle.
Both recommendations require the development of dedicated training programs and indexed online resources that can be easily discovered and accessed.

Recommendation 4.g.4: Advocate for vocabularies as first-class 'digital citizens' in their own right, including support for citation and indexing.
This should be supported by institutions and organisations recognising vocabularies as 'first-class' research objects. As such, the professional recognition and development of professionals who work to create, manage, and deliver vocabularies should help support their sustainability.

A Strategic Community Roadmap for an Australian FAIR Vocabulary Ecosystem. Wong, M., Atkins, N., Boyle, D., Dharmawardena, K., Hobern, D., McEachern, S., and Wyborn, L.

56

Appendix 4. Glossary

Controlled vocabulary: List of standardised terminology, words, or phrases, used for indexing or content analysis and information retrieval, usually in a defined information domain (CODATA RDM Terminology Working Group, 2024). The term 'controlled' indicates that there has been some governance of its creation, but it does not necessarily mean that the governance is being sustained.

Data Model: A model that specifies the structure or schema of a dataset. The model provides a documented description of the data and thus is an instance of metadata. It is a logical, relational data model showing an organised dataset as a collection of tables with entity, attributes and relations (CODATA RDM Terminology Working Group, 2024).

DCAT: Data Catalog Vocabulary for the Web. DCAT is an RDF vocabulary designed to facilitate interoperability between data catalogs published on the Web (Albertoni et al., 2024).

FAIR Vocabulary: A vocabulary that in itself complies with the FAIR principles (Wilkinson et al., 2016) and allows references to the terms to be resolved to learn what they mean, using standard web technology through an individual persistent resolvable unique web identifier (web link) per term (i.e. a HTTP (Hypertext Transfer Protocol) IRI (Internationalised Resource Identifier) (Cox et al., 2021).

Glossary: Alphabetical list of terms with definitions (Zeng, 2008).

International standard: Standard that is used in multiple nations and whose development process is open to representatives from all countries (CODATA RDM Terminology Working Group, 2024).

IRI: Internationalised Resource Identifiers, an internet protocol standard which builds on the Uniform Resource Identifier (URI) protocol by greatly expanding the set of permitted characters (Dürst and Suignard, 2005).

Knowledge Graphs: Very large semantic nets that integrate various and heterogeneous information sources to represent knowledge about certain domains of discourse (Fensel et al., 2020).

Legacy Vocabulary: A list of terms available in a print-based glossary or table not accessible using web standards (Cox et al., 2021).

Machine Actionable: Indicate a continuum of possible states wherein a digital object provides increasingly more detailed information to an autonomously-acting, computational data explorer. This information enables the agent — to a degree dependent on the amount of detail provided — to have the capacity, when faced with a digital object never encountered before, to: a) identify the type of object (with respect to both structure and intent), b) determine if it is useful within the context of the agent's current task by interrogating metadata and/or data elements, c) determine if it is usable, with respect to license, consent, or other accessibility or use constraints, and d) take appropriate action, in much the same manner that a human would (Wilkinson et al., 2016).

Machine interpretable: Machines can put the provided information into context and "understand" the meaning (semantics) and relations contained in the digital object. This concept is related to the Semantic Web vision and the Linked Data concept (Elixir-Europe, n.d.).

A Strategic Community Roadmap for an Australian FAIR Vocabulary Ecosystem. Wong, M., Atkins, N., Boyle, D., Dharmawardena, K., Hobern, D., McEachern, S., and Wyborn, L.

57

Machine Readable: Data in a form that can be used and understood by a computer (CODATA RDM Terminology Working Group, 2024).

Ontology: Shared and standardised list of words, terms and phrases to describe components of a particular discipline or domain, along with a taxonomy of their relations. Compare this to a controlled vocabularies, which tend not to include a structure of relations between their terms. Ontologies are typically developed by domain-specific institutions or communities to aid in the precise referencing of elements (CODATA RDM Terminology Working Group, 2024).

OWL: Web Ontology Language, a computational logic-based language for ontologies (Hitzler et al., 2012; W3C OWL Working Group, 2012; Cox et al., 2021).

PID: Persistent Identifier - long lasting digital reference to an object that gives information about that object regardless of what happens to that object. Developed to address link rot, a persistent identifier can be resolved to provide an appropriate representation of an object whether that object changes its online location or goes offline (CODATA RDM Terminology Working Group, 2024).

RDF: Resource Description Framework is the core data model of the Semantic Web (Miller, 2001; RDF Working Group, 2014; Cox et al., 2021).

RDF/XML: is a syntax, defined by the W3C, to express (serialise) an RDF graph as an XML document (Gandon and Schreiber, 2014).

RDFS: Resource Description Framework-Schema, is an extension of RDF and is used for representing simple RDF vocabularies on the Web (Brickley and Guha, 2014; Cox et al., 2021; Tomaszuk and Haudebourg, 2025).

Semantic Artefacts: A semantic artefact is a machine-actionable and -readable formalisation of a conceptualisation, enabling sharing and reuse by humans and machines. Semantic artefacts may have a broad range of formalisation, which include ontologies, terminologies, taxonomies, thesauri, vocabularies, metadata schemas, and standards Semantic artefacts are serialised using a variety of digital representation formats, e.g., RDF Turtle, and OWL, using XML (RDF) and JSON-LD (Hugo et al., 2020; Corcho et al., 2024).

Semantic Data: Data that are tagged with particular metadata that can be used to derive relationships between data (CODATA RDM Terminology Working Group, 2024).

Semantic interoperability: Ensures that the precise format and meaning of exchanged data and information are preserved and understood throughout exchanges between parties; in other words, what is sent is what is understood (CODATA RDM Terminology Working Group (2024).

Semantic Ladder: Synonym of Semantic Spectrum.

Semantic Spectrum: Sometimes referred to as the ontology spectrum, the smart data continuum, or the semantic ladder, is a series of increasingly precise or rather semantically expressive definitions for data elements in knowledge representations ranging from: 1) simple controlled vocabularies that list terms for use as standard labels to support indexing, content analysis and information retrieval, through to 2) ontologies that encode complex domain knowledge and concept relationships as part

A Strategic Community Roadmap for an Australian FAIR Vocabulary Ecosystem. Wong, M., Atkins, N., Boyle, D., Dharmawardena, K., Hobern, D., McEachern, S., and Wyborn, L.

58

of the definition of each term, especially for machine readability (McGuiness, 2003; McCready 2006; Duerr et al., 2024).

Semantic Standards: See Semantic Web Technologies.

Semantic Web Technologies: Semantic Web Technologies enable people to create data stores on the Web, build vocabularies, and write rules for handling data (W3C Semantic Web Standards, n.d.-a).

SKOS: Simple Knowledge Organisation System, is a simple OWL ontology to represent Knowledge Organization Systems (KOS) such as thesauri, term lists and controlled vocabularies (Isaac and Summers, 2009).

Standard: Set of agreed-upon and documented guidelines, specifications, accepted practices, technical requirements, or terminologies that have been prepared by a standards developing organisation or group, and published in accordance with established procedures. These can be mandatory or voluntary and are distinct from Acts, regulations, and codes, although standards can be referenced in those legal instruments (CODATA RDM Terminology Working Group, 2024).

Taxonomy: Divisions of terms into ordered, hierarchical groups, or categories based on particular characteristics (Zeng, 2008).

Thesaurus: Sets of terms representing concepts and the relationships connecting them (Zeng, 2008).

Turtle: A textual syntax for RDF that allows an RDF graph to be completely written in a compact and natural text form, with abbreviations for common usage patterns and datatypes (Beckett et al., 2014).

URI: Uniform Resource Identifier (URI), formerly Universal Resource Identifier, is a unique sequence of characters that identifies an abstract or physical resource, such as resources on a webpage, mail address, phone number, books, real-world objects such as people and places, concepts (Berners-Lee et al., 2005).

Vocabulary: Organised arrangement[s] of words and phrases used to index content and/or to retrieve content through browsing or searching (Harpring, 2013).

Vocabulary Ecosystem: The set of resources, tools, platforms, policies and practices that together make vocabularies accessible and useful for all researchers and users of digitally encoded knowledge (this paper).

Vocabulary Services: A vocabulary service enables users to discover, browse, and download controlled vocabularies in a variety of languages (e.g., SKOS, RDF/XM, Turtle, RDF/JSON, Notation3). The service may also contain an editor, where authorised users create, manage, and translate the vocabularies (CESSDA, n.d.). Corcho et al., 2024 p.3 define a vocabulary service as "a dedicated web-based system that fosters the availability, discoverability and long-term preservation and maintenance of semantic artifacts".

Vocabulary Registry: The terms "collection", "listing", or "registry" can be used to describe ontology libraries. These systems help reuse or find ontologies by simply listing them or offering structured metadata to describe them. Yet, those systems do not support additional services beyond the description of the items (Corcho et al., 2024).

A Strategic Community Roadmap for an Australian FAIR Vocabulary Ecosystem. Wong, M., Atkins, N., Boyle, D., Dharmawardena, K., Hobern, D., McEachern, S., and Wyborn, L.

59