



Australia's National
Science Agency

A Data Assessment Framework for Data Users

David Lemon, Peter Fitch, and Andrew Schwenke

August 2024

Version 1.1

Citation:

Lemon D., Fitch P., and Schwenke A. (2024). A Data Assessment Framework for Data Users. CSIRO, Australia.

Copyright

With the exception of the Commonwealth Coat of Arms and where otherwise noted, all material in this publication is provided under a Creative Commons Attribution 4.0 International (CC BY 4.0) license, available at <https://creativecommons.org/licenses/by/4.0/>. If material is copied or redistributed, the authors request attribution as follows: include the citation as shown above, plus the following copyright notice: '© Commonwealth Scientific and Industrial Research Organisation'

Important disclaimer

CSIRO advises that the information contained in this publication comprises general statements based on scientific research. The reader is advised and needs to be aware that such information may be incomplete or unable to be used in any specific situation. No reliance or actions must therefore be made on that information without seeking prior expert professional, scientific and technical advice. To the extent permitted by law, CSIRO (including its employees and consultants) excludes all liability to any person for any consequences, including but not limited to all losses, damages, costs, expenses and any other compensation, arising directly or indirectly from using this publication (in part or in whole) and any information or material contained in it.

CSIRO is committed to providing web accessible content wherever possible. If you are having difficulties with accessing this document please contact David.Lemon@csiro.au.

Version control tracking table

Title	A Data Assessment Framework for Data users			
Description	A framework for assessing datasets for 'fitness for purpose' in a range of activities. Includes instructions for use.			
Created By	David Lemon, Peter Fitch and Andrew Schwenke			
Date Created	May 2024			
Version Number	Released By	Description	Date Released	Status
1.0	David Lemon	Initial draft for comment	17/5/24	Draft
1.0.1	David Lemon	Released as part of May deliverable for AASF Data Ecosystem Project	31/5/25	Final Draft
1.0.2	David Lemon	Minor updates after internal CSIRO reviews		Final
1.1	David Lemon	Reorder assessment dimensions and renumber questions for ease of use Add of Obligations of Use question in Accessibility Dimension Addition of advice around assessment along data supply chains	30/8/24	Final
1.2	David Lemon	Update Acknowledgements	8/11/24	Final

Acknowledgements

This work has been developed with support from two projects:

- The *Australian Agricultural Sustainability Framework (AASF)* is a joint initiative led by the National Farmer's Federation (NFF) and supported by the Australian Government Department of Agriculture, Fisheries and Forestry through the Agriculture Traceability Grants Program
- The Ecological Knowledge System for the Nature Repair Market is funded by the Department of Climate Change, Energy, the Environment and Water (DCCEE). The work is being undertaken through a partnership between the Australian Government Department of Climate Change, Energy, the Environment and Water, and CSIRO.

CSIRO acknowledges the Traditional Owners of the land, sea, and waters, of the area that we live and work on across Australia. We acknowledge their continuing connection to their culture and pay our respects to their Elders past and present.

Contents

1	Introduction	1
1.1	Background	1
1.2	Purpose	2
1.3	Principles	2
2	Existing Frameworks, Principles and Standards	3
2.1	UN Stats DQAF	3
2.2	ABS Data Quality Framework	4
2.3	FAIR, CARE and SAFE	5
2.4	Metadata Standards	8
3	The Data Assessment Framework	10
3.1	Framework Structure	10
3.2	Using the Framework	11
4	Assessment Dimensions and Questions	15
4.1	Accessibility	15
4.2	Institutional Environment	16
4.3	Relevance	16
4.4	Timeliness	17
4.5	Accuracy	17
4.6	Coherence	18
4.7	Interpretability	18
	Appendix 1 - Example Framework Profiles	20

1 Introduction

1.1 Background

1.1.1 Drivers for the Need

As programs and projects increasingly aim to reuse or repurpose data collected by others, the need to create efficiencies, realise greater value and manage risks from expensive data collection exercises becomes paramount. This growing trend necessitates that data users—those seeking to reuse or repurpose data—can effectively determine whether a dataset meets their needs.

1.1.2 Challenges with Existing Metadata Statements

Datasets are typically accompanied by metadata statements that describe the content of the dataset and the context in which it was created. These metadata statements aim to assist potential data users in deciding whether the dataset is fit for their purposes. Understandably, they are written from the perspective of the data creator who is unlikely to understand any one potential user's specific requirements. Thus, for data users to make informed decisions, they must have a clear understanding of their specific needs. Therefore, a complementary mechanism is required for data users to articulate their needs, enabling a thorough assessment of candidate datasets.

1.1.3 The CSIRO Initiative

The Commonwealth Scientific and Industrial Research Organisation (CSIRO) has been approached by two concurrent projects — the Ecological Knowledge System (EKS) and the Australian Agricultural Sustainability Framework (AASF) Data Ecosystem — to develop such a mechanism. After conducting research, it became apparent that no existing framework fully met the requirements as current frameworks were found to be either: developed by and for data creators or publishers (e.g., FAIR principles); tailored towards specific types of data (e.g., statistics from surveys); or overly theoretical and difficult to use in practice.

1.1.4 Developing a Generic Framework

Given these findings, CSIRO recognised the need to develop a generic framework that could be tailored to individual circumstances. This framework aims to encompass all potential considerations an end-user might have when assessing a dataset for use, while being adaptable to various contexts and specific needs. The goal is to create a comprehensive, user-friendly tool that facilitates the assessment of dataset suitability, thereby supporting the efficient and effective reuse of data across diverse projects and initiatives.

1.2 Purpose

A general framework which individual communities/initiatives/projects can use to enable them to assess the fitness for purpose of candidate datasets for use within their context.

1.3 Principles

1. Aimed at end users of data (NOT data publishers)

Specifically in the agriculture/environment sectors where:

- There is a heavy reliance on modelling,
 - A corresponding heavy reliance on geospatial (often Earth Observation) data,
 - And a heavy reliance on repurposing/reusing data collected for by others for different purposes.
 - There can also be a reliance on surveys of stakeholders, to gather enterprise or individual's data.
2. Simple to use – ideally a set of question to which the answer is “Yes” or “No”.
 3. Comprehensive – covers a range of considerations and not just data quality.
 4. Can be used in a range of circumstances (research projects, ongoing programmes of work, reporting)

2 Existing Frameworks, Principles and Standards

The need to be able assess datasets for suitability, or assess a datasets ‘fitness-for-purpose’, has been previously recognised in several domains. The health sector, for example, has recognised that if health data is to be reused/repurposed, there needs to exist a structured method for undertaking this assessment in a consistent way¹.

More broadly, there exists several data assessment frameworks (sometimes referred to as data quality assessment frameworks or data quality statements) which have potential relevance to this work.

The following describes some relevant initiatives including data quality frameworks, data use principles and frameworks and metadata standards. Note that the following is not a detailed academic literature review and is not intended as such. Rather, it is a review of existing work that has been used to guide the development of the framework described in later sections.

2.1 UN Stats DQAF²

The United Nations Statistics (UN Stats) Data Quality Assessment Framework (DQAF) was developed with a focus on statistical data. The framework enables users to specify requirements for and assess against 5 dimensions of data quality. The dimensions are:

1. assurances of integrity – covering the professionalism, transparency, and ethical standards of the data collection, and its production and publishing processes
2. methodological soundness – ensuring the methodological basis for the dataset is sound
3. accuracy and reliability – ensuring that the source data statistical techniques are sound, and outputs portray reality
4. serviceability – ensuring the data has adequate periodicity (collection return rate) and timeliness (time from collection to publishing), is consistent within the dataset and with other datasets and has a predictable revision policy
5. accessibility – ensuring the data and metadata are easily accessed and assistance to users is adequate.

For each dimension, the framework identifies elements of good practices and within each of these elements, relevant indicators. For individual datasets, the framework adds focal issues and key points as shown in Figure 1.

¹ <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9006677/>

² https://unstats.un.org/unsd/dnss/docs-nqaf/IMF-dqrs_factsheet.pdf

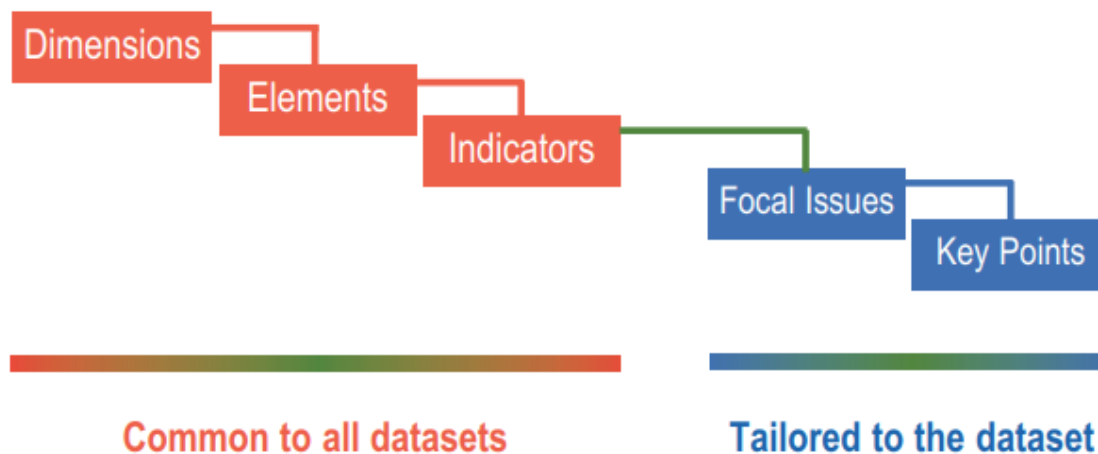


Figure 1 DQAF Hierarchy³

The DQAF has been developed for statistical data products and so has some limitations when trying to apply it to other initiatives. For instance, the framework does not directly address issues around the useability (interpretability) of data.

However, the dimensions provide excellent guidance on the topics that need to be addressed when assessing a dataset. Furthermore, the hierarchical structure is useful for framing an assessment and allows for flexibility with implementation. Similarly, the acknowledgement that an initiative will have common requirements as well as dataset specific requirements is useful.

2.2 ABS Data Quality Framework⁴

The Australian Bureau of Statistics (ABS) Data Quality Framework (DQF) was developed to provide a standard for assessing and reporting on the quality of statistical information. Like the DQAF, the ABS DQF is comprised of dimensions (in this case 7):

1. Institutional Environment – focussing the impartiality, objectivity, professionalism, mandate, capability, and commitment of the organisation(s) that collect, produce and publish the dataset.
2. Relevance – focussing on how well the dataset meets the needs of users in terms of concepts measured and populations represented.
3. Timeliness – focussing on the time delays between data collection, the advertised data of publication and the actual date of publication
4. Accuracy – focussing the degree to which the data adequately describes the phenomenon it is designed to measure.

³ From https://unstats.un.org/unsd/dnss/docs-nqaf/IMF-dqrs_factsheet.pdf

⁴ <https://www.abs.gov.au/ausstats/abs@.nsf/Latestproducts/1520.0Main%20Features3May%202009?opendocument&tabname=Summary&prodno=1520.0&issue=May%202009&num=&view=>

5. Coherence – refers to both the internal consistency of the dataset as well as the ability to compare the data with other sources.
6. Interpretability – refers to the availability of information to help provide insight into the data.
7. Accessibility – refers to the ease of access to the data by users.

The ABS DQF is a general framework and so, to be applied, users need to consider what aspects of the framework are important to their circumstance. The recommendation is that all dimensions should be considered but the importance placed upon each, when deciding, may vary.

The ABS DQF has been adopted by several organisations including the Australian Institute for Health and Welfare (AIHW)⁵. Here, a modified version of the ABS DQF is used to assess data quality of a dataset to determine ‘fitness-for-purpose’. Of interest here is the data quality assessment template which, for each dimension, poses a set of questions to which the answer can only be: “Yes”, “Partially”, or “No”.

The ABS DQF dimensions provide a greater breadth of considerations than the DQAF and have been applied in Australia. The AIHW approach of posing questions with only ‘Yes’, ‘No’ or ‘Maybe’ answers meet the desire for the framework to be easy to use.

2.3 FAIR, CARE and SAFE

While not data assessment frameworks *per se*, the FAIR⁶ and CARE⁷ principles, as well as the Shared Analytics Framework for the Environment⁸ (SAFE) can be used to provide guidance on the issues and concerns that should be considered when developing such an assessment framework.

2.3.1 FAIR

The FAIR (Findable, Accessible, Interoperable, and Reusable) principles are primarily for data owners and have been adopted widely throughout the research sector to support the sharing of research data.

From a data users’ perspective, the FAIR Principles provide guidance around subjects such as data accessibility and interpretability. They require that:

- Data be retrievable by humans and machines through standardised communication protocols with appropriate authentication and authorisation where necessary;
- Data have clear licence and provenance information;
- There be clarity and transparency around the conditions governing access to and use of the data;

5 <https://nla.gov.au/nla.obj-788584958/view>

6 <https://force11.org/info/guiding-principles-for-findable-accessible-interoperable-and-re-usable-data-publishing-version-b1-0/>

7 <https://www.gida-global.org/care>

8 <https://ardc.edu.au/resource/shared-analytic-framework-for-the-environment-safe-2-0/>

- Data uses community accepted languages, formats, and vocabularies; and
- Metadata provides rich and accurate information.

2.3.2 CARE

The CARE (Collective benefit, Authority to control, Responsibility, Ethics) Principles for Indigenous Data Governance were developed in recognition that the open data movement and adoption of principles such as FAIR (above) were focussed primarily on facilitating greater sharing, and hence, exploitation of data without considering power differentials between stakeholders. This impacts indigenous peoples in particular who are seeking to assert greater control over the application and use of their data and knowledge.

Thus, the CARE principles complement the FAIR principles by seeking consideration of people and purpose in decision making around the use of datasets.

For the purposes of a data assessment framework, they ask users to consider:

- Where indigenous data is to be used, any value generated should benefit indigenous communities in an equitable manner;
- Indigenous people's rights and interests in their data must be recognised and their authority to control it, empowered;
- Information on how indigenous data is used within the project need to be shared with the data owners; and
- Indigenous rights and wellbeing should be the primary concern at all stages of the data lifecycle.

2.3.3 SAFE⁹

The Shared Analytic Framework for the Environment (SAFE) has been developed jointly by the Western Australian Biodiversity Science Institute (WABSI) and the Western Australian Marine Science Institute (WAMSI). It was created to describe the capabilities needed in an information supply chain to support environmental assessments.

As with the FAIR and CARE Principles, SAFE is not a data assessment framework. However, it can provide guidance on the subjects that should be considered when assessing data and data supply chains.

The framework consists of tiers and capabilities. These are reproduced in Figure 2

⁹ <https://wabsi.org.au/wp-content/uploads/2021/07/SAFE-Guide-V1.1P.pdf>

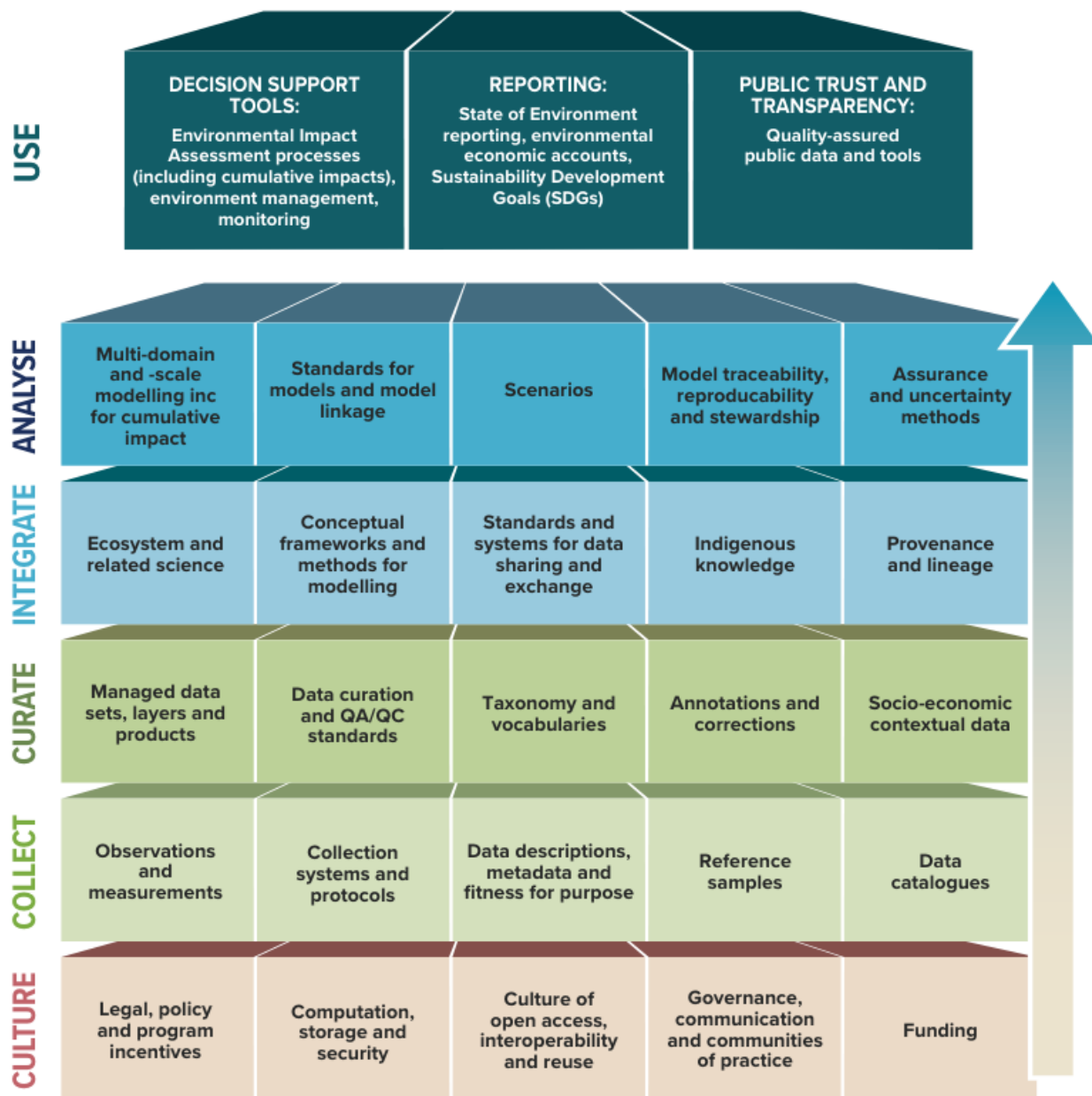


Figure 2 SAFE Tiers and Capabilities¹⁰

From a data assessment framework perspective, not all capabilities described in SAFE are relevant. However, many are and so, when developing the framework, it is useful to consider each to understand what aspects of that capability might be important to a potential data user.

¹⁰ From <https://wabsi.org.au/wp-content/uploads/2021/07/SAFE-Guide-V1.1P.pdf>

2.4 Metadata Standards

There are many metadata standards in active use and a review of each is not relevant here. Nor are metadata standards a substitute for a data assessment framework. The statements created using metadata standards are generally created by the data owner/producer and reflect the dataset with which it is associated rather than a potential data user's needs.

However, as with the FAIR and CARE principles and SAFE, the type of information captured within a metadata statement can guide the creation of a data assessment framework as it provides guidance as to what the developers of the standard determined to be of important to potential users.

2.4.1 Dublin Core Metadata Initiative

Dublin Core is a basic, domain-agnostic metadata standard developed and maintained by the Dublin Core Metadata Initiative (DCMI). It is one of the best known and most widely used metadata standards. The specification consists of 15 metadata elements¹¹:

- Contributor – An entity responsible for making contributions to the resource.
- Coverage – The spatial or temporal topic of the resource, the spatial applicability of the resource, or the jurisdiction under which the resource is relevant.
- Creator – An entity primarily responsible for making the resource.
- Date – A point or period of time associated with an event in the lifecycle of the resource.
- Description – An account of the resource.
- Format – The file format, physical medium, or dimensions of the resource.
- Identifier – An unambiguous reference to the resource within a given context.
- Language – A language of the resource.
- Publisher – An entity responsible for making the resource available.
- Relation – A related resource.
- Rights – Information about rights held in and over the resource.
- Source – A related resource from which the described resource is derived.
- Subject – The topic of the resource.
- Title – A name given to the resource.
- Type – The nature or genre of the resource.

¹¹ <https://www.dublincore.org/specifications/dublin-core/dces/>

Since publication, the standard has been refined with by adding qualifiers for some of the elements. For a data assessment framework, knowledge of the datasets coverage, key dates, format, language, subject, rights for use, and more will likely be of interest.

2.4.2 ISO TC 211

The International Standards Organisation's TC 211 is responsible for standards within the domain of digital geographic information. Within this group's suite of 97 standards are a number relating to data quality and data description. These are:

- ISO 19113 – Quality Principles
- ISO 19114 – Quality Evaluation Procedures
- ISO 19115 – Metadata
- ISO 19131 – Data Product Specification
- ISO 19138 – Data Quality Measures
- ISO 19157 – Data Quality
- ISO 19158 – Quality Assurance of Data Supply

Once again, none of these is a data assessment framework *per se*. However, they do provide guidance on the information that might be needed to undertake a data assessment.

3 The Data Assessment Framework

3.1 Framework Structure

The proposed data assessment framework consists of two high level concepts: a general framework; and activity profiles.

The **general framework** adopts the ABS DQF dimension structure as well as the AIHW approach of posing a set of questions within each dimension to which the answer can only be “Yes” or “No”.

The development of the questions has been guided by analysis of existing data quality frameworks, the FAIR and CARE principles, the Shared Analytic Framework for the Environment (SAFE) and a number of metadata and data quality standards (DCMI, ISO TC211)

Activity profiles recognise that no two activities (projects, initiatives, programmes of work) will have the same set of requirements when it comes to data. Activity profiles are used by individual activities to: identify what, from the general framework, is important to their activity; capture key requirements of data for the activity; and provide guidance to data users as to how to proceed with a dataset.

Key elements of the framework are described in Figure 3 and Table 1.

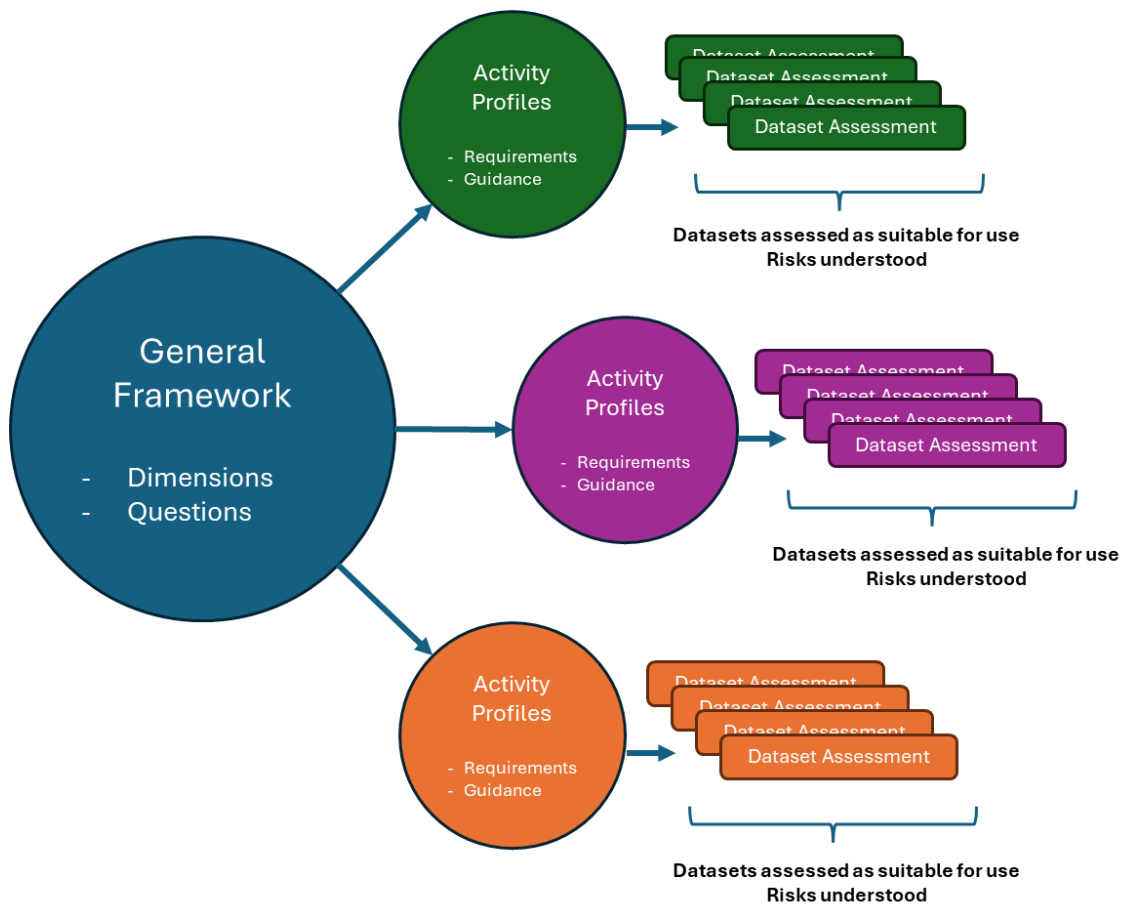


Figure 3 Data Assessment Framework

Table 1: Data Assessment Framework key concepts

COMPONENT	DESCRIPTION
General Framework	
<i>Dimensions</i>	Dimensions are areas of concern related to the assessment of the dataset. For the framework, the dimensions are: <ul style="list-style-type: none"> • Accessibility - Ease of access to data by users. • Institutional Environment - Institutional and organisational factors that may have influence on the credibility of the data. • Relevance - How well the dataset meets user needs. • Timeliness - The latency between dataset data collection, collation, supply and use as well as update frequency. • Accuracy - The degree of correctness of the data for the estimate provided. • Coherence - The comparability of the dataset to other data of similar type, as well as to prior version of the same dataset. • Interpretability - The availability of information to aid interpretation of that data to generate insights.
<i>Questions</i>	A set of questions, relevant to the dimension which are designed to evaluate the suitability for use of a particular dataset. The answer to a question can only be “Yes”, “No” or “Not applicable”. See Section 4
Activity Profiles	
<i>Activity</i>	An activity, project, initiative, programme of work which will seek to use one or more datasets as inputs and will likely have similar requirements of these datasets with respect to the general assessment framework
<i>Relevance Assessment</i>	An initial assessment of the general framework questions to determine which are relevant to the activity profile.
<i>Activity Requirements</i>	The set of requirements that define ‘fit for use’ for each question within the activity profile. In other words, the “bar” that needs to be met to achieve ‘Yes’ for the question.
<i>Activity Guidance</i>	For each question in an activity profile, advice needs to be provided on how to proceed if the answer to a question is ‘No’. Specific guidance is needed for each question.

3.2 Using the Framework

Use of the framework occurs in two phases:

1. Activity Profile development
2. Data Assessment

3.2.1 Activity Profile Development

The requirements that any activity will have of data will vary. For example, different activities may require different data formats or use different definitions of key concepts or have different expectations in regarding data quality. Thus, prior to using the data assessment framework, an activity will need to create its own profile of it. This activity profile will determine:

1. Which of the framework questions are relevant (important) to the activity;
2. For the relevant questions, the requirement to be met for each question to pass (achieve ‘Yes’); and

3. What guidance is to be provided to a user of the profile when a dataset fails ('No') a question during an assessment.

Ideally, this task will be undertaken by those involved in the use of data within the project.

Determining Relevance

Not all questions in the framework will be relevant to all activities. For example, for an activity that uses geospatial data, Question 2.4: *Appropriate geospatial scale* is important. However, an activity that does not use any geospatial data will not regard Question 2.4 as not relevant.

The initial step for an activity in developing its profile is to determine which of the framework questions are relevant to the activity.

Identifying Expectations/Requirements

While the answer to every framework question is either "Yes", "No" or, in a small number of cases, "Not Applicable", the expectation or requirement to be tested for a question may not always be the same. For example, Question 7.7 (*Are the data available in suitable and widely used formats?*) requires prior identification of what formats are considered suitable within the context of the activity.

Thus, a key step in developing an activity profile, is to articulate the expectation or requirement to be met for each of the relevant framework questions.

When describing individual expectations, it is useful to indicate the obligation to meet the expectation using the words "Must" and "Should". That is, if a requirement is compulsory, it "Must" be met. Whereas, if an expectation is recommended but not compulsory, it "Should" be met.

Providing Guidance

It is likely that, when undertaking an assessment, a dataset will not meet the requirements for one or more questions. The framework allows for this outcome by including, in the activity profile, the opportunity to provide guidance on how to proceed in these situations. In some cases, this guidance might direct that the dataset should or must not be used. In other cases, the guidance might recommend remedial actions to be taken prior to using it or it define circumstances in which the data can and should not be used.

Similar to the description of expectations/requirements, when providing guidance, it is useful to indicate the obligation to meet the expectation using the words "Must" and "Should". That is, if the guidance is compulsory, it "Must" be followed. Whereas, if the guidance is a recommendation but not compulsory, it "Should" be followed.

An example of part of a profile is shown in Table 2. Further examples are provided in Appendix A

Table 2 Snippet of a completed profile. In this example, Q 1.7 and Q 1.8 have been deemed to be relevant to the activity. Q 6.6a, however, has been determined not relevant as no reference to individual features will be used.

ID	QUESTION	MATERIAL	REQUIREMENT	GUIDANCE
1.7	Access methods: Does the organisation that collects, produces and disseminates the dataset have acceptable protocols/systems in place to facilitate physical data access?	Yes	Data should be available via a standard API	If not available via API, others forms of digital data can be used but should be sourced prior to use. The original data needs to be archived and the date and form of access documented
1.8	Formats: Are the data available in suitable and widely used formats?	Yes	Data must be in one of the following formats: CSV, JSON, XML	If possible, data should be converted to one of these formats prior to use. If not possible, another source for the data should be found.
6.6a	Identification of Features: Does the data source use published/well known identifiers for features within the data set?	No	N/A	N/A

3.2.2 Assessing a dataset

To use the framework, the assessor (the individual or individuals performing the assessment), needs the completed activity profile along with any available documentation associated the dataset to be assessed. They may wish to undertake the assessment with an owner of the dataset or someone with great knowledge of the dataset. However, this may not always be possible.

For the dataset being assessed, the assessor determines whether, or not, the dataset meets the activity’s requirements against each of the relevant questions within the activity profile. The result for each question can be either:

- “Yes” – the data set meets the requirement;
- “No” – the dataset does not meet the requirement; or
- “Not Applicable” – in a small number of situations the question may not be applicable. These situations will be clear.

If a dataset meets all requirements, it is deemed to be ‘fit for purpose’ to be used within the activity.

If a dataset does not meet one or more requirements, the activity profile will provide guidance on what actions to take for each unmet requirement. Note that it is likely that, in some circumstances, a data set may still be used within an activity, despite not meeting the requirements for a question, so long as the guidance for that question is followed. There will however, be circumstances where failure to meet the requirements for a question will render the dataset unusable.

Just as no two activities are alike, the purposes to which a dataset may be put within a project may also differ. Thus, when undertaking an assessment, it is important to consider the purpose for which the dataset is to be used. For example, an assessment might set a higher expectation of

datasets that are to be used within some important analysis, than it might for datasets that are used to provide context in a map for a report.

The data assessment framework is a tool for those proposing to use a dataset within their activity asking them to consider a range of issue about the data set prior to use. The framework guides the user through exploration of a set of dimensions that should be considered before committing to use a dataset. Ultimately, the decision to use a dataset within an activity is the responsibility of the end user and it may be that other factors, beyond those within the framework, need to be considered.

Assessing Input Supply Chains

Its important to note that many datasets are the product of analysis of other datasets potential produced by third parties. This analysis may be as simple as the combination of datasets (eg. the National Vegetation Information System (NVIS), produced y the Dept of Climate Change, Energy, Environment and Water, is a compilation of data sets, each of which is produced by a state or territory agency) or more sophisticated (eg. The Habitat Condition Assessment System (HCAS), produced by CSIRO, involves sophisticated analysis of remotely sensed imagery, sourced from third party organisations).

Any assessment should be aware of these input supply chains and, where appropriate, consider any risks they may introduce for the data set being assessed as well as the activity in which the dataset will be used. For example, an activity that requires ongoing updates to a dataset should consider risks to the supply of inputs and ensure that the data producer has a plan in case there is a disruption along their input supply chain. Similarly, where a dataset has been created through compilation of other data sets created at varying times or with varying levels of uncertainty, the assessor should consider the impact of this on their needs.

Ideally, the assessment process should not undertake a full assessment of input data sets along a supply chain, and this is not recommended. However, for assessment questions where input datasets will have a material impact on the answer to the question, the nature of these datasets should be considered.

4 Assessment Dimensions and Questions

4.1 Accessibility

QUESTION	COMMENTS
1.1 Licence: Does the dataset have a licence agreement?	If the dataset does have a licence, it is important to check if the licence is acceptable (see following questions).
1.2 Licence acceptability: If YES to 7.1, are the conditions of the licence acceptable?	Things to consider with this licence are the nature of the licence (Does it require the same licence be applied to derivative products (known as a copyleft or viral licence)?, Does it transfer ownership of IP?) This question also asks the activity to consider what expectations the licence places on the activity and whether the activity is willing and able to meet these obligations. For example, data from first nations communities can come with a number of obligations on use.
1.3 License restriction: If YES to 7.1, do the licence conditions allow for the intended use of the dataset?	This question asks the activity to consider its intended use for the data set within the activity, an whether the licence allows for this? For example, does the licence allow for publication of outputs if this is required? Does it allow for commercialisation of outputs if this is planned? (For example, some licences allow for data to be used for research but not other purposes)
1.4 Permission to use data: Has permission to use the dataset been given?	This is important in all circumstances but especially where the data comes from less powerful providers (eg, small business, individuals, first nations communities). Some licences will grant permission automatically (eg. "open data") while others will not. Consideration should also be given to formalising and recording permission for use in a suitable forum.
1.5 Obligations of use: Are obligations associated with use of the dataset understood and can they be met?	This question is particularly important with respect to the use of data from first nations peoples. These datasets are often made available with specific obligations to be met by the user. Here, the question is asking if these obligations are known and understood as well as if the user has systems in place to ensure they are fulfilled.
1.6 Cost: Is the dataset available at an acceptable cost?	The activity will need to determine what 'acceptable' means in their context as this will vary.
1.7 Access methods: Does the organisation that collects, produces and disseminates the dataset have acceptable protocols/systems in place to facilitate physical data access?	This question covers how the data is actually accessed. What is, and isn't, acceptable will vary by project and be a function of what is intended for the data set, how and how often access is needed, etc.
1.8 Formats: Are the data available in suitable and widely used formats?	This will vary by activity. Some may require specific formats, others may be less concerned. Consider here the work necessary to convert data from non-digital (paper) or non-analysable forms.

4.2 Institutional Environment

QUESTION	COMMENT
<p>2.1 Objectivity and transparency: Does the organisation or entity that collects, produces and publishes the dataset, do so in an objective, professional, transparent and legislatively responsible manner?</p>	<p>Can the organisations(s)/entity(s) that collect, produce and/or publish the data be trusted to be objective? Why do they collect, produce and/or publish the data? Do they have permission/authority to collect, produce and/or publish the data? Are they free from potential conflicts of interest? Do they comply with privacy and legislative requirements for managing data?</p>
<p>2.2 Capability to produce: Does the organisation that collects, produces and publishes the dataset have the processes, staff and facilities in place to ensure required data quality?</p>	<p>This question assesses the quality of the subject dataset with respect to the capabilities of the organisation producing it. Is the organisation able to produce the dataset at the required data quality level? This ability includes having appropriate systems in place, appropriate skills and appropriate resources.</p> <p>Also consider here any data supply chains that the data producers rely on to produce the dataset.</p>
<p>2.3 Resources to maintain: Does the organisation that collects, produces, and disseminates the dataset have sufficient resources for ongoing maintenance if applicable?</p>	<p>While an organisation may have the resources to produce a dataset, they may not have resources to maintain that dataset. This might include the need to update the dataset or to continue to make the dataset accessible.</p>
<p>2.4 Resources to sustain collection: Does the organisation that collects, produces and disseminates the dataset have sufficient resources for the ongoing collection, production and dissemination if applicable?</p>	<p>If the dataset is a series and/or the activity requires ongoing data supply, is the organisation that collects, produces and publishes the dataset, resourced to continue to do this. This includes financial resources, staffing, mandate, etc.</p> <p>Of interest here are datasets created through one-off, limited term (research) projects which, if operationalised, might be collected on an ongoing basis but resources to do this are yet to be secured.</p>

4.3 Relevance

QUESTION	COMMENT
<p>3.1 Applicability: Does the dataset contain measures/estimates of the phenomena needed for the activity?</p>	<p>This question is essentially asking ‘does the dataset measure what you need it to measure or what you think it measures?’ There are a number of elements to assessing this:</p> <ol style="list-style-type: none"> 1. Does the data within the dataset relate to features of interest to the activity? 2. Does the definition of the concepts represented within the dataset match the activity’s definitions of that concept? 3. Is the data representative of the target population identified by the end user of the data? <p>For example: Given a dataset of species presence/absence, surveyed in location X, is this data representative of the target population? In some cases this will be difficult to answer - it could depend on the type of data, the collection methods used, sample strategy, time period, repetitions, weather conditions, etc.</p>
<p>3.2 Methods Do the data reflect the condition or situation it was designed to measure?</p>	<p>This question focuses on the validity of the methods used to create the dataset to measure the phenomenon being measured.</p>
<p>3.3 Collection methods: Were the data collection methods applicable to the data being collected?</p>	<p>Related to 2.2, this question focusses on the data collection methods. In particular, was the method used valid in the context it was used. For example, while a particular method may be valid under laboratory conditions, it may not be valid for use in the field.</p>

QUESTION	COMMENT
<p>3.4 Appropriate geospatial scale: If a geospatial dataset, is the dataset applicable at the scale at which it is intended for use?</p>	<p>Can the dataset be scaled up or down to the geographic scale needed for the activity?</p> <p>If Raster: Is the spatial resolution of the dataset suitable for use. Note that this resolution can be too high as well as too low.</p> <p>If Vector: Are the precision and scale of vector features suitable for use.</p>
<p>3.5 Appropriate temporal coverage: Is the dataset relevant within the time-period of use?</p>	<p>Is there a consequential time difference between the needs of the project/activity and the time the dataset was collected? Another way to ask this is: Is the dataset representative of the measures within it, at the time in question for the project/activity?</p> <p>This will vary depending upon the needs of the project and the purpose the dataset will be used for. For example, if the data measures some parameter that varies over time and the dataset is to be used for analysis that is time critical, this question will be important.</p> <p>Also consider here the date of collection of any input datasets if appropriate.</p>

4.4 Timeliness

QUESTION	COMMENT
<p>4.1 Minor update: Are there likely to be subsequent additions or changes to this dataset?</p>	<p>For some datasets, producers may provide minor changes or additions. Users will need to consider the implications of such changes on their activity. What are likely timeframes for minor updates?</p>
<p>4.2 Major update frequency: Are there likely to be updates or revisions to the data?</p>	<p>More generally relevant, the activity needs to factor that there might be updates to the dataset and have a plan to accommodate these. If there are to be updates, what is the frequency of updates?</p>
<p>4.3 Latency: For time sensitive activities, is the dataset published in an appropriate time for use?</p>	<p>Similar to 2.5, however here this question speaks to the timeliness of data supply. Will the dataset be available when needed?</p>

4.5 Accuracy

QUESTION	COMMENT
<p>5.1 Uncertainty sources: Are potential or acknowledged sources of error described sufficiently?</p>	<p>Here “sufficiently” means there is enough information about uncertainty such that fitness for purpose of the dataset can be assessed. This will depend on the needs of the activity and the purpose for which the dataset will be used. For example, a dataset which is intended for use in analysis may require more detailed information about uncertainty than a dataset to be used for creating context maps.</p> <p>Note also that uncertainty can take different forms. Geospatial data sets, for example, should describe positional accuracy as well as accuracy of other parameters.</p> <p>Also consider here the description of uncertainty for any input datasets if appropriate.</p>
<p>5.2 Bias and error: Are biases and errors within the dataset acceptable for the purpose of the dataset?</p>	<p>This question is related to 4.1. However, here the question is testing that reported uncertainty within the dataset meets the needs/expectations of the activity. That is, from an uncertainty perspective, the dataset is fit for purpose.</p>

4.6 Coherence

QUESTION	COMMENT
6.1 Inter-timestep coherence: If a time series, is the dataset coherent between time steps	Does the dataset measure the same concepts between time steps? Does the dataset use the same features of interest between time steps?
6.2 Conceptually consistent through time: Is the dataset internally coherent?	Are concepts used consistently throughout? If the dataset has been compiled from multiple sources, were these sources consistent?
6.3 Completeness: Is the dataset complete?	Does the dataset contain a full coverage of the phenomena being measured or are there gaps. These gaps might be spatial or temporal or the data may not cover individual populations. It may also be that gaps in the data have been 'filled' in some way by the data producer. Have these processes been adequately described?

4.7 Interpretability

QUESTION	COMMENT
7.1 Metadata for interpretability: Is metadata available to support correct interpretation of the data?	Is there enough information to be able to confidently use the dataset without having to infer (or guess) particular aspects of the data? The minimum metadata needed to be able to interpret a dataset will vary and may include information such as: units of measure; spatial and/or temporal extent, spatial reference system, definition of concepts.
7.2 Documented methods: Are the dataset methods documented in sufficient detail?	Related to Q 2.2, here the test is whether there is sufficient information to be able to understand and have confidence in the methods to the level needed by the activity. Sufficient detail will be dependent on the use of the dataset within the activity. Also consider here documentation of any input datasets used in the creation of the dataset.
7.3 Standardised concepts: Does the data source use standard concepts, classifications, and vocabularies?	To use a dataset requires understanding the language of the dataset. That is potential users need to understand the producer's definition of the various concepts and other terms used in the dataset. Datasets that use standardised concepts, classifications, and vocabularies enable the 'understand' process to be more efficient.
7.4 Acronym explanations: Are all acronyms and abbreviations used in the dataset clearly defined?	While many acronyms may be commonly understood within individual communities, the same acronym within other disciplines may refer to something completely different. For example, the acronym MVP often means 'Minimum Viable Product', particularly in the IT sector. However, in sport, it often refers to 'Most Valuable Player'. It is not essential that acronyms in a dataset be defined. However, if they are not, it is incumbent on the user to be confident in their meaning.
7.5 Method comparison: Does the data source use methods comparable with other data collections?	This question refers to situations where, between data collection activities, the methods used are changed. It is important to consider if these methods are comparable.

QUESTION	COMMENT
<p>7.6a Identification of features: Does the dataset use published/well known identifiers for features within the dataset?</p>	<p>The Observations and Measurements Standard (O&M) recognises that a key aspect of any measurement is the feature (thing) to which the parameter being measured, applies. For example, if flow in a stream is being measured, it is necessary to know in which stream the flow is being measured.</p> <p>As with standardised concepts and vocabularies, using published/well-known identifiers for features within a dataset enables users to use the dataset with confidence. They also enable integrating the data set with other datasets that have the same features.</p>
<p>7.6b Identifier consistency: Are the identifiers used within the dataset consistent with those used by related datasets which describe the same features?</p>	<p>Following Q6.6a, here the question is testing whether the identifiers used for features within the dataset can be related to or align with the identifiers used for the same features in other datasets. (This is regardless of whether the identifiers are well known or not). An important risk here is that the same feature is identified in different ways between datasets.</p>
<p>7.7 Definition accessibility: If concepts and identifiers are defined outside of the dataset, are the definitions accessible / retrievable / maintained / versioned?</p>	<p>Related to Q6.3 and 6.6a, this question is ensuring that the definition of concepts and identifiers are well managed.</p>

Appendix 1 - Example Framework Profiles

A.1.1 AASF Model Report

The following is an example Activity Profile for AASF Whole of Australian Agriculture Report use case. Note that this is an example of what an Activity Profile might look like for two dimensions “Institutional Environment” and “Relevance” and the information here has been fabricated. Ideally, this Activity Profile would be developed in collaboration with those developing the report.

Within this use case (activity), datasets are being collated to produce a report. The report is to be produced on an annual basis at the national scale for use within trade negotiations or similar. The information contained within the report will undergo a high level of scrutiny.

DIMENSION	QUESTIONS		ASSESSMENT		GUIDANCE
	Question		Relevant	Activity Requirements	
Institutional environment	2.1: Objectivity and transparency	Does the organisation or entity that collects, produces, and publishes the dataset, do so in an objective, professional, transparent and legislatively responsible manner?	Yes	For all data uses: Data collection, production and dissemination agency MUST meet legislative requirements on data privacy.	Datasets from agencies that do not meet legislative requirements MUST not be used.
				For data to be used in production of indicators: Data SHOULD be from agencies with no vested interest in the outcome of the activity.	For data to be used in production of indicators: Datasets from agencies with a vested interested can be used. However, attention SHOULD be paid to the methods used within the dataset as well as the transparency of the organisations processes.
	2.2: Capability to produce	Does the organisation that collects, produces and publishes the dataset have the processes, staff and facilities in place to ensure required data quality?	Yes	For data to be used in production of indicators: production agencies SHOULD have the skills and resources to produce data to quality required to produce the report	For data to be used in production of indicators: it is unlikely that datasets from agencies without appropriate skills or resources are likely to meet requirements and SHOULD not be used.
	2.3: Resources to maintain	Does the organisation that collects, produces, and disseminates the dataset have sufficient resources for ongoing maintenance if applicable?	Yes	For data to be used in production of indicators: production agencies SHOULD have the skills and resources to maintain the data	For data to be used in production of indicators: where the production agency does not have the resources to maintain a dataset, consideration SHOULD be given to ensuring the data is maintained in some other way to ensure it remains accessible.

	2.4: Resources to sustain collection	Does the organisation that collects, produces and disseminates the dataset have sufficient resources for the ongoing collection, production and dissemination if applicable.	Yes	For data to be used in production of indicators: production agencies SHOULD have the skills and resources to continue to produce the data	For data to be used in production of indicators: this represents a significant risk to the activity as future supply of the data is not assured. The activity MUST consider where future supplies of the data will be sourced.
Relevance	3.1: Applicability	Does the dataset contain measures/estimates of the phenomena needed by the end user of the data?	Yes	For data to be used in production of indicators: the dataset MUST contain measures of the indicator in question	For data to be used in production of indicators: the dataset SHOULD not be used
	3.2: Suitability of methods	Do the data reflect the condition or situation it was designed to measure?	Yes	For data to be used in production of indicators: the dataset method MUST be valid for the indicator in question	For data to be used in production of indicators: the dataset SHOULD not be used
	3.3: Collection methods	Were the data collection methods applicable to the data being collected?	Yes	For data to be used in production of indicators: the dataset collection method MUST be valid for the context in which it was used	For data to be used in production of indicators: the dataset SHOULD not be used
	3.4: Appropriate spatial scale	If a geospatial dataset, is the dataset applicable at the scale at which it is intended for use?	Yes	For data to be used in production of indicators: the dataset MUST be applicable at a national scale	For data to be used in production of indicators: if the data can be scaled up to the national scale, this should be done prior to use. Otherwise, the dataset SHOULD not be used
	3.5: Appropriate temporal coverage	Is the dataset relevant within the time period of use?	Yes	For data to be used in production of indicators: the dataset collection method MUST be relevant to the time period of the report.	For data to be used in production of indicators: the dataset SHOULD not be used

A.1.2 Ecological Knowledge System

The following is an example Activity Profile from the Ecological Knowledge System project. Note that this is an example of an Activity Profile, presented here for all dimensions of the framework.

Within this use case (activity), datasets are being collated support various functions of a market. Ongoing access to scientifically rigorous data as well as data from first nations people are important in this activity. The information contained within the report will undergo a high level of scrutiny.

DIMENSION	QUESTIONS		ASSESSMENT		GUIDANCE
	Question		Relevant	Activity Requirements	
Accessibility	1.1: Licence	Does the data set have a licence agreement?	Yes	The dataset must have an acceptable licence agreement.	Datasets that do not meet these requirements must not be used.
	1.2: Licence acceptability	If YES to 7.1, are the conditions of the licence acceptable?	Yes	The dataset must have an acceptable licence conditions.	Datasets that do not meet these requirements must not be used.
	1.3: License restriction	If YES to 7.1, do the licence condition allow for the data sets intended use?	Yes	The licence must allow for the data sets intended use.	Datasets that do not meet these requirements must not be used.
	1.4: Permission to use data	Has permission to use the data set been given?	Yes	We must have permission to use the data sets.	Datasets that do not meet these requirements must not be used.
	1.5: Obligations of use	Are obligations associated with use of the dataset understood and can they be met?	Yes	Obligations for use must be clearly understood and, if they can be met, systems must be in place to ensure they are met.	Datasets that have obligations for use that cannot be met, MUST not be used.
	1.6: Cost	Is the data set available at an acceptable cost?	Yes	Cost to use the data must be acceptable	Datasets that do not meet these requirements must not be used.
	1.7: Access methods	Does the organisation that collects, produces and disseminates the data set have acceptable protocols/systems in place to facilitate physical data access?	Yes	Data should be in digital form and publicly and or directly accessible over the web via systems such as APIs or direct downloads.	If the data is private access only assess the system in place for access, and whether it is appropriate/operational. If not, the dataset should not be used.

Institutional environment	1.8: Formats	Are the data available in suitable and widely used formats?	Yes	<p>Examples of suitable and widely used formats:</p> <p>General: CSV, JSON, XML</p> <p>Raster data: GeoTIFF, Cloud optimised GeoTIFF, GRID (being one of the two types: ASCII Grid (.ASC), or Esri Grid (proprietary); within ESRI grid: Arc/Info Binary Grid Format and Arc/Info ASCII Grid Format</p> <p>Vector data: Esri Shapefile, JSON, GeoJSON, Google Keyhole Markup Language (.KML/.KMZ)</p> <p>Others - Esri File Geodatabase, NetCDF, HDF, GRIB</p>	If data is not available in an acceptable format for use by the EKS, assess whether it can be converted appropriately for use. If not, it should not be used.
	2.1: Objectivity and transparency	Does the organisation or entity that collects, produces and publishes the data set, do so in an objective, professional, transparent and legislatively responsible manner?	Yes	<p>Purpose of the dataset must be transparent.</p> <p>The organisation needs to have acknowledged any conflicts of interest.</p> <p>Organisations must comply with privacy and legislative requirements.</p>	Data that does not meet these requirements must not be used.
	2.2: Capability to produce	Does the organisation that collects, produces and publishes the data set have the processes, staff and facilities in place to ensure required data quality?	Yes	Organisation is known and reputable and has demonstrated their ability to deliver appropriate quality datasets.	Dataset must have independent assessment of data quality including exploratory data analysis and validation against known benchmarks or alternative sources.
	2.3: Resources to maintain	Does the organisation that collects, produces, and disseminates the data set have sufficient resources for ongoing maintenance if applicable?	Yes	The organisation should be resourced to maintain access to the data set if remote access is necessary.	If the organisation doesn't have the resources to maintain access, then the EKS must consider maintaining its own copy.
	2.4: Resources to sustain collection	Does the organisation that collects, produces and disseminates the data set have sufficient resources for the ongoing collection, production and dissemination if applicable?	Yes	Where a timeseries of the dataset is required, the organisation must be resourced to produce the next / subsequent versions (depending on the frequency required for each dataset).	Undertake assessment of sensitivity of Activity needs to interruption in data supply. In high sensitivity situations, an alternative must be identified, or the data set must not be used.

Relevance	3.1: Applicability	Does the data set contain measures/estimates of the phenomena needed by the end user of the data?	Yes	All necessary variables, measures and attributes are present. Our definition of the phenomena required matches the dataset definition. Metadata must provide clear documentation on the intended use and limitations of the dataset.	Datasets that do not meet these requirements must not be used.
	3.2: Methods	Do the data reflect the condition or situation it was designed to measure?	Yes	The variables and metrics captured in the dataset must be directly relevant to the intended phenomena. Appropriate and reliable methods were used to measure the intended phenomena.	Datasets that do not meet these requirements must not be used.
	3.3: Collection methods	Were the data collection methods applicable to the data being collected?	Yes	The data collection methods align with the specific condition or situation the dataset is meant to measure. Methods must have been used in the correct context.	Datasets that do not meet these requirements must not be used.
	3.4: Appropriate spatial scale	If a geospatial data set, is the data set applicable at the scale at which it is intended for use?	Yes	Is the spatial scale appropriate for the intended use? Or else, can it be appropriately scaled to meet the needs of the Activity? If the data has been aggregated or generalized this must be appropriate for the intended scale of use.	Datasets that do not meet these requirements must not be used.
	3.5: Appropriate temporal coverage	Is the data set relevant within the time period of use?	Yes	Data needs to be representative of the time relevant to the purpose (e.g. if we need to infer a present state from this data, the data should be relevant to inferring the present state).	If data is not representative of the time period (e.g. a consequential time difference given the Activity needs) it should not be used. Also see Institutional environment - resources to sustain (1.4). If applicable, consider where future updates are coming from.
Timeliness	4.1: Minor update	Are there likely to be subsequent data additions or to changes this data set?	Yes	Metadata must describe update frequency, last update date, next scheduled update, and latency of updates.	Develop a plan that includes governance based on different scenarios including: minor updates, major updates, updates to EKS, and potential interactions between data updates and EKS.

Accuracy	4.2: Major update frequency	Are there likely to be updates or revisions to the data?	Yes	Metadata must categorize and document the types of updates applied to the data including whether the update is to spatial features, attributes, structural updates, or metadata updates.	Develop a plan that includes governance based on different scenarios including: minor updates, major updates, updates to EKS, and potential interactions between data updates and EKS.
	4.3: Latency	For time sensitive activities, is the data set published in an appropriate time for use?	Yes	The latency between data collection and availability must be known and considered as per 2.5.	If latency is unacceptable, the dataset should not be used.
	5.1: Uncertainty sources	Are potential or acknowledged sources of error described sufficiently?	Yes	Metadata should describe errors comprehensively including positional accuracy, attribute accuracy, temporal accuracy (related to the timing of data collection and any discrepancies), data collection method errors and processing and transformation sources of error.	If potential sources of error are not adequately described, EKS must source additional information on errors/uncertainty and/or do independent analysis. If still not transparent or unclear the dataset must not be used.
	5.2: Bias and error	Are biases and errors within the data set acceptable for the purpose of the data set?	Yes	The metadata must identify sources of error (see 4.1) and bias such as sampling bias and measurement bias and these must be within acceptable tolerance levels.	Datasets that do not meet these requirements must not be used.
Coherence	6.1: Inter-timestep coherence	If a time series, is the data set coherent between time steps	Yes	The data set must use the same/comparable concepts and features between timesteps. The data must be coherent over the specified time period. Changes in data collection methods, definitions, or classifications must be noted in the metadata.	If the dataset is not coherent between time steps it requires reassessment as a new dataset under the data assessment framework.
	6.2: Conceptually consistent through time	Is the data set internally coherent?	Yes	The dataset must maintain consistent data formats, structures, and conventions across all variables and records.	Inconsistencies must be documented including the source of the inconsistency e.g. in compilation, classification, method etc. EKS must assess the impact of the inconsistencies before making the decision to use.

	6.3: Completeness	Is the data set complete?	Yes	Metadata must describe where data has been filled and by what technique (interpolation, extrapolation, estimation).	If gap filling techniques are not adequately described EKS must source additional information on errors/uncertainty and/or do independent analysis. If still not clear, the dataset must not be used.
Interpretability	7.1: Metadata for interpretability	Is metadata available to support correct interpretation of the data?	Yes	Metadata should provide clear documentation on the intended use and limitations of the dataset. There must be enough information to be able to confidently use the data set without having to infer (or guess) particular aspects of the data. Units of measure, CRS, key concepts must be defined, compliant with ISO19115, ANZLIC	Where metadata doesn't reference a standard framework the EKS must seek input from domain experts or stakeholders familiar with the data source to validate any unclear aspects. If still not transparent or clear the dataset must not be used.
	7.2: Documented methods	Are the dataset methods documented in sufficient detail?	Yes	The metadata must document the methods including specifying the sources from which data was collected, including original datasets, surveys, sensors, or other instruments; data acquisition method; data preparation, transformation, aggregation and processing.	Where metadata doesn't reference a standard framework the EKS must seek input from domain experts or stakeholders familiar with the data source to validate any unclear aspects. If still not transparent or clear the dataset must not be used.
	7.3: Standardised concepts	Does the data source use standard concepts, classifications and vocabularies?	Yes	The metadata should reference standard frameworks, ontologies, or controlled vocabularies (e.g., ISO standards, domain-specific terminologies).	Where metadata doesn't reference a standard framework the EKS must seek input from domain experts or stakeholders familiar with the data source to validate the use of standard concepts and classifications and their adequacy and appropriateness for the intended use.
	7.4: Acronym explanations	Are all acronyms and abbreviations used in the dataset clearly defined?	Yes	The metadata should define acronyms etc	Where metadata doesn't describe acronyms, the data producer must be contacted for the information, or the acronym must be verified by other means.
	7.5: Method comparison	Does the data source use methods comparable with other data collections?	Yes	(see also 5.1) The metadata should describe potential variability in methods due to factors such as geographic location, technological advancements, or changes in sampling procedures.	Evaluate whether the methods used align with recognised standards and guidelines relevant to the domain (and that they align with Section 2). If the new methods can't be validated and compared, the dataset should not be used and/or should be treated as a new dataset in the data assessment framework.
	7.6a: Identification of features	Does the data set use published/well known identifiers for features within the data set?	Yes	The data set should use the same identifier for features and be aligned with standard identifies schemes or systems or adhere to standard naming conventions, formats and protocols.	If data does not use a published identifier standard, there will be a requirement to standardise before use. The data must be attached to the same feature. If the dataset cannot be rectified, it must not be used.

7.6b: Identification of features	Are the identifiers used within the data set consistent with those used by related data sets which describe the same features?	Yes	If 6.6a is no, then 6.6b is very important.	The dataset must be rectified to be consistent/standardised. If the dataset cannot be rectified, it must not be used.
7.7: Availability of definitions and identifiers	If concepts and identifiers are defined outside of the dataset, are the definitions accessible / retrievable / maintained / versioned?	Yes	Must be able to access definitions.	Datasets that do not meet these requirements must not be used.

As Australia's national science agency, CSIRO is solving the greatest challenges through innovative science and technology.

CSIRO. Creating a better future for everyone.

Contact us

1300 363 400
+61 3 9545 2176
csiro.au/contact
csiro.au

For further information

Environment
David Lemon
+61 417 880 758
David.Lemon@csiro.au
csiro.au/environment