



# Digital Antarctica

## A1. High-level Scope

April 2021

## Contents

Executive Summary .....	3
Digital Antarctica Initiative overview .....	3
Background .....	3
Vision.....	3
Goals .....	3
Deliverables.....	4
Digital Data Model.....	6
Data map documentation .....	7
Architectural standard.....	7
A hosted Digital Antarctica .....	7
Hosting a centralised library of metadata .....	7
Hosting services.....	8
Hosting a portal.....	8
Hosting data.....	8
Prototype .....	8
Scope.....	8
Data Scope .....	9
Research Scope .....	10
Geographical Scope .....	10
Geopolitical Scope .....	11
Glossary .....	11

## Version information

Version	Description	Author	Date
1.0	Release version	Rob Jennings	23/2/2021
2.0	Release version after revision	Rob Jennings	13/04/2021

© Copyright Australian Antarctic Program Partnership 2021



This work by the Australian Antarctic Program Partnership is licensed under a Creative Commons Attribution (CC BY) 4.0 International License.

Details of the licence are available at: <https://creativecommons.org/licenses/by/4.0/>

## Executive Summary

*Digital Antarctica* has the goal of improving the lives of anyone who benefits from Australian Antarctic research by ensuring that the data used to create that research – and the data products created from that research – are easier to find, collate and integrate, which leads to richer science and stronger decision making. The vision of the initiative is to provide a definitive answer to the question “Where do I go to find Australian Antarctic research data?”

The scope of *Digital Antarctica* is to deliver access to all Australian Antarctic and Southern Ocean research data and data products stored by the AAPP partner organisations.

*Digital Antarctica* will be a digital data model, acting as a schema of the data across the disparate organisations within the AAPP. While the exact form of the model is not yet decided, it will be used to help consumers of data find the data they are looking for and to improve the quality of the work they are doing. The initiative will also provide standards in defining, storing and sharing data, to ensure that Australian research organisations are generating data that meets the FAIR data principles where data is Findable, Accessible, Interoperable and Reusable.

The first phase of *Digital Antarctica* is an analysis phase which, over the next 18 months, will deliver a series of documents detailing the scope of the whole initiative, the current state of Antarctic research data practices, the requirements and desired state, and recommended pathways to achieve that state. As a whole, the documents will be presented as a recommended strategy to building a fully realised digital data model of Australian Antarctic data. As part of the analysis phase, a prototype will also be scoped, which will test core outputs of *Digital Antarctica* using real-world data.

This first document outlines a high-level scope, which will be used to guide the overall initiative.

## Digital Antarctica Initiative overview

### Background

In 2017, Drew Clarke released the Australian Antarctic Science Program Governance Review (the Clarke review), which called for a number of measures to be taken in regards to Antarctic research, including “a comprehensive data model of the Australian Antarctic Territory”.

The Australian Antarctic Program Partnership (AAPP) was created from the recommendations in the Clarke review, and in 2020 work began on the data model. During the foundation work, the model took the name *Digital Antarctica*.

### Vision

---

*To provide a definitive answer to the question:  
“Where do I go to find Australian Antarctic research data?”*

---

Currently, the question “Where do I go to find Australian Antarctic research data?” generally has the answer “it depends on what you’re looking for”. The vision of *Digital Antarctica* is to be the definitive answer to that question.

### Goals

The primary goal of *Digital Antarctica*, as taken from the Clarke review, is to enable “a world-class centre for Antarctic data analytics” which demonstrates that Australia is a leader in Antarctic data, bolstering Australia’s position in Antarctic negotiations. A core concept within the review is that data

should be seen as core science infrastructure with the same importance as other infrastructure (e.g. a ship or a station). This strong focus on data will help us not only meet our Antarctic Treaty obligations, but exceed them by showing respect and understanding of the region and by displaying a pro-active use of data to meet policy and research needs.

Another primary goal of *Digital Antarctica* is to assist data users in all forms. This includes data creators, data managers and data consumers – such as researchers, data officers, educators, industry and government departments – including policy makers and their advisors. A comprehensive model of Antarctic data will assist data creators and data consumers in accessing the most relevant data for their needs, in a way that gives them the most benefit. The *Digital Antarctica* initiative will assist data managers by providing ongoing guidance on best practice in capturing, storing and sharing data. This will ensure that the data within *Digital Antarctica* is Findable, Accessible, Interoperable and Reusable (see more on FAIR in the Glossary below). This, in turn will:

- facilitate data analytics by integrating with wider datasets, which will improve the breadth and effectiveness of the data products for end users.
- increase the value of data. Because of its remoteness, Antarctic data is expensive to collect. A comprehensive data model which enables data consumers to bring together science from decades of research across a number of organisations will distribute the value of each of those collections across multiple use cases.
- ensure confidence in the data. The strong standards employed to produce data through *Digital Antarctica* will ensure that a data consumer will have confidence in the completeness, and the provenance, of that data.

Another goal of *Digital Antarctica* is to foster the partnership between AAPP organisations. The AAPP, through the *Digital Antarctica* initiative, will work together to create a common and agreed set of standards for Antarctic data and metadata, and combine resources to assist in the creation and ongoing activity that supports the future of this digital model.

## Deliverables

The first phase of the *Digital Antarctica* initiative will be an analysis and design phase, resulting in a series of documents defining various aspects of the initiative (see below). Over the course of this analysis phase, a number of meetings and workshops will be held with relevant stakeholders, the outputs of which will be reflected in the documentation. As part of this analysis a prototype is planned, which will also require a scope and definition document.

The documents are:

- A. *Scope* – This document will be split into two, a High-level and Refined, and will describe the projects goal and vision, as well as the definitions and scope.
  - A1. *High-level Scope (this document)* – As well as defining the initiatives goal and vision, this document will broadly describe the scope, and will explore possible attributes of a digital data model. This document will guide the creation of the refined scope and current state documents.
  - A2. *Refined Scope* – This document will take the broad concepts discussed in the high-level scope and drill down into specifics regarding data, research, geographical and geopolitical scope, and will further specify data model concepts to be explored during requirements and future state definition.
- B. *Current State* – This document will also be split in two, but will, as a whole, describe the current state of data capture and sharing within the AAPP

*B1. High-level current state* – This document will give a general overview, not bounded by the scope document. This will include describing in general terms the remit of each data centre, their services and the standards they use. This document will guide the creation of the refined current state.

*B2. Refined current state* – This document will more specifically define the current state within the parameters of the scope defined by document A. It will be used to help identify the areas of concern to be addressed in the Requirements & Future State document, as well as inform the Transition document. It will include

- detailed discussion on the remit, services and standards of the data centres
- Descriptions of data interactions, at the individual data centre level, the AAPP as a whole, and from a data user’s perspective (including use cases)
- Descriptions and applicability of other services that provide a similar function to *Digital Antarctica*

*C. Prototype Scope Definition* – This document will describe the function and scope of a proposed prototype (see Prototype below)

*D. Requirements & Future State* – This document will be the core description of *Digital Antarctica* and will include the requirements, results and findings from the prototype, high-level design concepts and future use cases and scenarios

*E. Transition* – This document will detail the effort required to transition from current state to future state, including the contributions that AAPP partners can make in the transition. It will also explore standards to be applied and discuss how new data can be made ready for *Digital Antarctica*.

*F. Summary/Recommendations* – This document will be a summary of all of the above, presented as a recommendation. This may be used to present the *Digital Antarctica* concept to a wider audience for project continuation.

The timeline for the delivery of these documents is below:

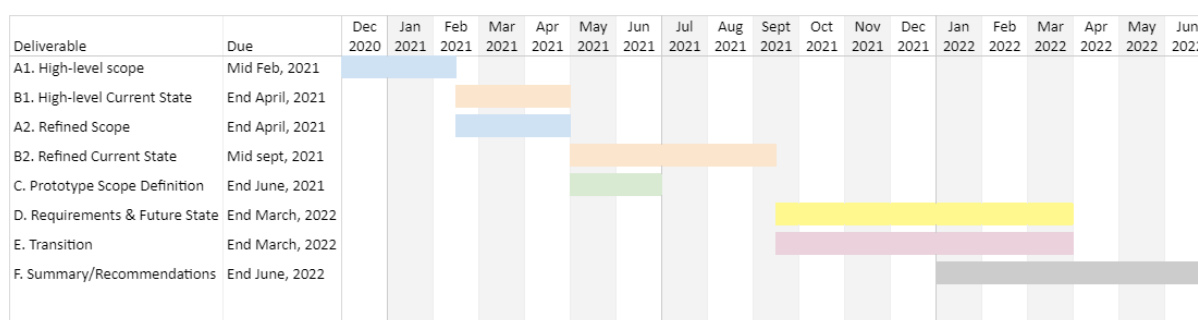


Figure 1 - Deliverables and timeline

Once each document is completed and reviewed internally, it will be forwarded to relevant stakeholders for review before being signed off by the AAPP management committee.

The timeline for the delivery of the actual *Digital Antarctica* product, and for the prototype, will be determined during the course of the project when more information is available regarding:

- The full scope of *Digital Antarctica*
- The scope of the prototype
- The budgets involved
- The designs of both *Digital Antarctica* and the prototype

- The outcomes from the prototype exercise (see Prototype below)

## Digital Data Model

*Digital Antarctica* is a digital data model of the Australian Antarctic Territory and Southern Ocean (see Geographical Scope below). The model's primary purpose will be to enable the interoperability of Antarctic data provided by AAPP organisations (see Glossary below).

“Data model” as a term has a number of different definitions, especially within the research data community. For the purposes of *Digital Antarctica*, the kind of data model being discussed is, at its most basic level, a tool which identifies the location and organisation of data within one or more data sources to create a cohesive view of the data as a whole. This can range from a simple map or schema diagram identifying the data and data sources, through to a fully interconnected infrastructure of data access services and portals.

The data within the data model is explored further in the Data Scope section below, but refers to any item available for retrieval within a data centre. This could be a dataset of raw observations, a document or paper, or a full simulation.

Exactly what form *Digital Antarctica* will take as a data model will be determined by the requirements and design to be explored as part of this initiative and will be documented in the *Refined Current State* and *Requirements & Future State* documents. In determining the full scope of delivery, considerations will need to include (but not be limited to) such elements as utility, performance, interoperability and cost. Figure 2 below shows a full potential implementation of a delivery model with the data model at the centre, enabling a mapping of data. While the full scope of *Digital Antarctica* is not yet known, a core outcome is to synthesise disparate data and provide it to a consumer of that data.

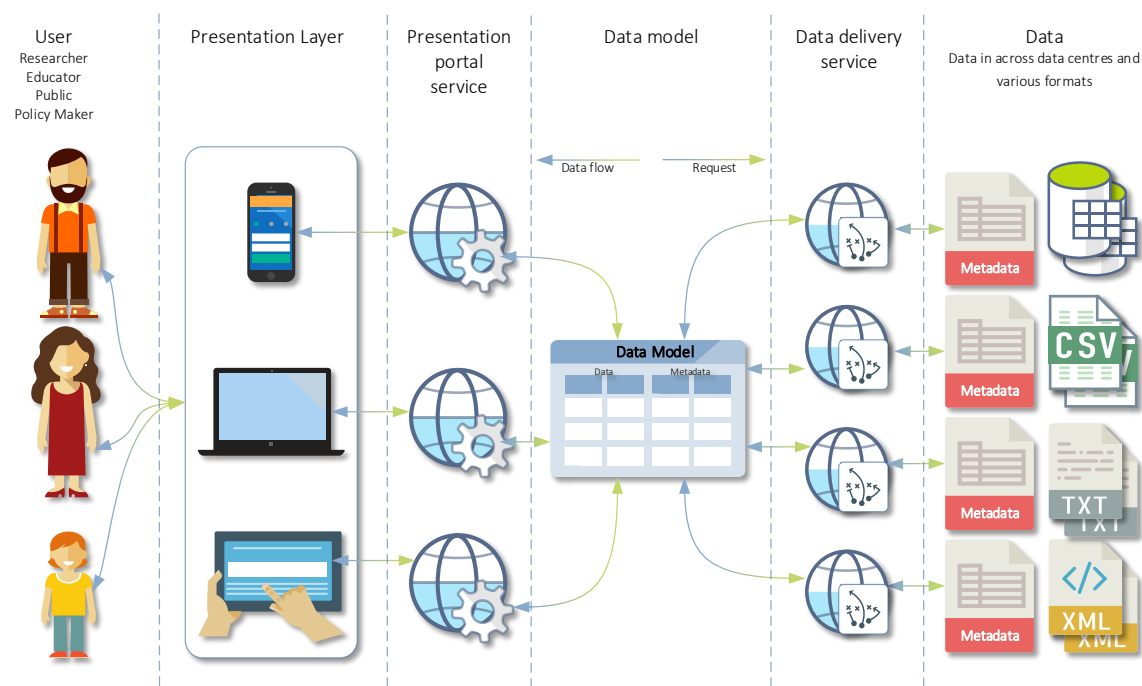


Figure 2 – A full potential implementation of a delivery model. A user accesses a front-end interface (via an interactive display, mobile device or app, web page or some other means), providing some form of input. That interface then calls one or more APIs which uses the data model to locate data relevant to the input. The API then pulls the data from those data sources, to return to the interface in the required format.

Below are a number of data model concepts or approaches that may be undertaken in creating a completed *Digital Antarctica*. Note that these approaches represent the possible directions that *Digital Antarctica* may take. They are not mutually exclusive and, in some cases are built upon each other. A fully realised *Digital Antarctica* is likely to draw on a number of these concepts.

## Data map documentation

For a data model to function, the location and accessibility of the data must be known and documented. A documented map of Antarctic data across the AAPP's data centres (detailing the location and type of data and metadata, as well as the existing services to serve the data) will be an essential step, regardless of further development into a more comprehensive model.

## Architectural standard

*Digital Antarctica* may provide a collection of documented architectural standards for web services and web service designs which will enable anyone to create an index used for finding and accessing Antarctic data. Various organisations' data and metadata is served through their own services, meeting a centralised standard guided by the *Digital Antarctica* architecture, ensuring that their data is available. Likewise, anyone wishing to build a portal (or some other form of access) can use the *Digital Antarctica* architectural standards to design and build their own web services for that portal to draw Australian Antarctic data from the data source web services.

As an architectural standard, *Digital Antarctica* would, in essence, be a methodology of collating the data (metadata and/or actual data) served by those services to an interface, but will not be a hosted service nor provide an interface itself. Its primary role is to provide guidance on how data within the data model can interoperate. It will also provide guidance on data and metadata standards to ensure that data and metadata are ready for incorporation into a federated model.

## A hosted Digital Antarctica

There are a number of ways *Digital Antarctica* could be provided as a hosted solution. These will be explored as potential options, considering the following:

- Data considerations
  - Performance
  - Data Integrity
  - Security
- Technical considerations
  - Feasibility
  - Resources
  - Supportability
  - Technical expertise available
  - On premises, cloud hosting or a combination
- Practical considerations
  - Cost to build
  - Ongoing maintenance
  - Ongoing funding

The following is a selection of hosting options that may be looked at for *Digital Antarctica*.

### ***Hosting a centralised library of metadata***

All data available within *Digital Antarctica* will have fully realised metadata, which will be used to search for data. While the metadata will, usually, be sourced from the same data centre as the data (or, at least, within the same data centre ecosystem as the data) it may also require replication at a centralised host, for the purposes of facilitating searches within *Digital Antarctica*. These metadata



may be the metadata records drawn directly from the source, or they may be manipulated to fit the *Digital Antarctica* and may even be enriched or truncated for the purpose of enhancing the search facilities (e.g. enriched by also including column names of gridded data, or truncated by only including certain elements within the metadata record). In this model the metadata and the data themselves would still be hosted at their source, however the metadata (or information contained within the metadata) will be replicated at the central host, ready to be accessed by the search and retrieve web services.

### **Hosting services**

A model of hosting may include hosting of any required services used to search for and serve data to any external portal. These services would be built to the *Digital Antarctica* architectural standard, and could:

- Take data and/or data search requests from a user source (e.g. a data portal, or directly) and search for data across the data model using the parameters in the request; and/or
- Send requests to, and retrieve data from, the data centres hosting the required data.

### **Hosting a portal**

There are many existing research data portals that may incorporate data from *Digital Antarctica* however it may be necessary to create a data portal specifically for users to search Australian Antarctic data. This data portal could be hosted as part of the *Digital Antarctica* solution; however, many users may not require a data portal but will instead interface directly with the services. *Digital Antarctica* should enable this functionality.

### **Hosting data**

Based on a number of factors to be discovered and investigated during analysis, it may be necessary to mirror metadata AND data from primary data sources to enable data to be served efficiently to users of *Digital Antarctica*. While the master data would remain with the original source, this may involve caching data periodically to improve performance or full data replication.

## **Prototype**

As part of the initiative (as laid out in the timeline above), a prototype will be designed and built to test the core principles behind *Digital Antarctica*. This prototype will endeavour to take disparate, but related, data from multiple sources (preferably from multiple organisations) and bring them together in a way that proves one or more design concepts being considered for *Digital Antarctica*. The scope of this prototype will be explored further in future documents.

## **Scope**

In a very broad sense:

---

*Digital Antarctica will deliver access to all Australian Antarctic and Southern Ocean research data and data products stored by the AAPP partner organisations.*

---

There are a number of dimensions of scope that impact the overall breadth of the initiative. These will be explored in further depth below, and include:

- Data Scope
- Research Scope
- Geographical Scope
- Geopolitical Scope



It is important to note that this is a high-level overview of scope. As such the dimensions discussed below will remain at the high level for the purposes of this document. These will be further developed during the analysis process and published in a later scope document. However, the dimensions discussed below are considered a minimum to meet the overall goals of the initiative. These may be refined and expanded over time, either during development or as part of *Digital Antarctica* operations.

## **Data Scope**

---

*Digital Antarctica will be primarily concerned with **research data and data products stored by the AAPP partner organisations.***

---

Data in this context is any item available from a data centre, and may include any forms of data along a spectrum between raw data and fully realised data models and simulations (see <https://earthdata.nasa.gov/collaborate/open-data-services-and-software/data-information-policy/data-levels> for examples of different data levels). Data in this context also includes analytical data products created for policy and decision making.

Pragmatically, the following factors will also be considered:

- The readiness of the data to be shared. This includes considerations such as
  - whether the data is digitised
  - compliance with agreed standards, including structure and quality
  - the data's FAIRness (see FAIR in the Glossary below)
  - the existence, and state, of the data's metadata
  - whether there are services available to share the data
- The data currently accessible via the data centres of the AAPP partner organisations. As a first step, the data available through *Digital Antarctica* will be sourced from the AAPP partner organisations, but (due to the services and filtering available at the time) may also include:
  - Data collected by non-AAPP programs within those organisations.
  - Data that is otherwise out of the scope areas outlined in this document (e.g. non-Antarctic Data).

### **Metadata**

In context of research data, a metadata record details information about an associated dataset such as a description of the data and/or research, the time and location of the data collection, keywords etc. While metadata can be recorded in any format (and any data about data is considered to be a form of metadata) there are a number of metadata standards specifically used for recording research data. There are two standards that are in primary use across the AAPP organisations:

- Directory Interchange Format (DIF) – this is a metadata standard introduced by NASA for describing earth science data. It is the standard used by the Australian Antarctic Data Centre<sup>1</sup>
- ISO 19115 – This is the metadata standard introduced by ISO, the International Organisation for Standardisation for describing geographical information and services. It is the standard used by many organisations within the AAPP.<sup>2</sup>

Both of these standards have attributes which help describe data, however the way those standards are implemented can vary from organisation to organisation. Keywords, for example, might be drawn from a known vocabulary, or might be unique to an individual dataset. Likewise, the depth of detail within the metadata records varies across datasets. Also, while DIF and ISO 19115 share many attributes, they do not have a 1:1 compatibility. As such, strategies will need be developed for *Digital*

---

<sup>1</sup> For more information on DIF, see <https://earthdata.nasa.gov/esdis/eso/standards-and-references/directory-interchange-format-dif-standard>

<sup>2</sup> For more information on ISO 19115, see <https://www.iso.org/standard/26020.html>

*Antarctica* in regards to handling multiple metadata formats and levels of richness. This may involve a mapping between metadata standards, as well as developing strategies and best-practice guidance on metadata usage.

## Research Scope

*Digital Antarctica will be primarily concerned with Antarctic and Southern Ocean data.*

Specifically, this means any research of any type (e.g. earth, physical, life, environmental and social science etc as referenced in the Clarke review) either:

- focussed on the areas defined within the geographical scope (below). This includes research conducted outside that geographical scope (e.g. research conducted in mainland Australia, about the Australian Antarctic Territory); or
- conducted within the bounds of the geographical scope (below). This includes research that is conducted within, but not specifically about, the Antarctic and Southern Ocean region (e.g. astronomical observations recorded from instruments located in Antarctica).

## Geographical Scope

*Digital Antarctica will be primarily concerned with data from the Australian Antarctic Territory, the Southern Ocean and Australian subantarctic islands.*

While the Australian Antarctic Territory is clearly defined, the exact geographical boundaries of the Southern Ocean that *Digital Antarctica* will use (including any necessary exceptions), will be defined at a deeper level in the future, and will consider such sources as:

- Oceanographic definitions of “Antarctic and Southern Ocean”
- Data outside of the Antarctic and Southern Ocean that “overlaps” (i.e. impacts or is impacted by) the defined geographical scope.
- Official Australian government sources, such as the Australian Hydrographic office:

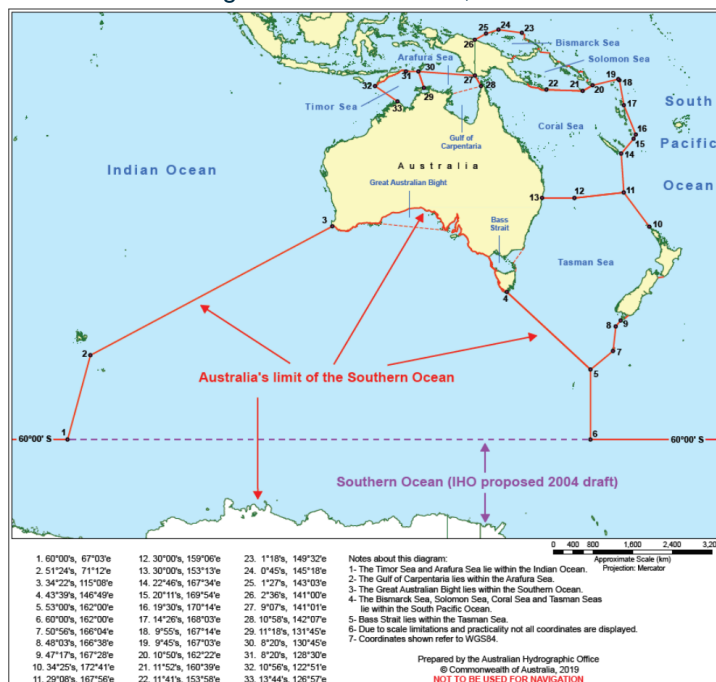


Figure 3 - Australian Hydrographic Office, Names and limits of oceans and seas around Australia

- Non-scientific uses of Southern Ocean data (e.g. Australia’s search and rescue region as defined here: <https://www.amsa.gov.au/safety-navigation/search-and-rescue/australias-search-and-rescue-region>)
- Any non-spatial data that may be otherwise relevant.

## Geopolitical Scope

*Digital Antarctica will be primarily concerned with **Australian** research data.*

The source of data for *Digital Antarctica*, at least initially, will be the data centres of the AAPP partner organisations, all of whom are Australian organisations.

Due to the nature of the organisations and research, data stored within those data centres may not be wholly Australian and in some cases may not be Australian at all. While this data may still be available to *Digital Antarctica*, the primary concern of the initiative will be Australian data.

The architecture of *Digital Antarctica*, however, will be designed for scalability to ensure that, in the future, additional datasets from wider sources (either other Australian Antarctic data repositories, or non-Australian Antarctic data repositories), can be integrated.

All data available through *Digital Antarctica* will be legally allowed to be shared under licensing laws and will, where applicable, be subject to the data policy of Australian Antarctic Data Centre ([https://data.aad.gov.au/aadc/about/data\\_policy.cfm](https://data.aad.gov.au/aadc/about/data_policy.cfm))

## Glossary

Term	Description
AAPP	<p>The Australian Antarctic Program Partnership. A partnership of Australian Antarctic research organisations with the goal of better understanding the role of the Antarctic Region. The partnership includes the following partner agencies</p> <ul style="list-style-type: none"> <li>• University of Tasmania</li> <li>• Institute for Marine &amp; Antarctic Studies</li> <li>• The Australian Antarctic Division</li> <li>• CSIRO</li> <li>• Bureau of Meteorology</li> <li>• Geoscience Australia</li> <li>• Tasmanian Government</li> </ul>
Data User	<p>A data user is anyone that interacts with data. Specifically, there are 3 broad categories of data user within the <i>Digital Antarctica</i> ecosystem.</p> <ul style="list-style-type: none"> <li>• <i>Data creator</i> – this is anyone who generates research data for the purpose of sharing within <i>Digital Antarctica</i></li> <li>• <i>Data manager</i> – this is anyone involved in the curation and maintenance of data. E.g. Anyone who facilitates the upload of data, ensures the quality of data and metadata, and who maintains hardware and services used in capturing serving and otherwise sharing data</li> <li>• <i>Data consumer</i> – this is anyone who uses data that has been shared. E.g. researchers,</li> </ul>

	government departments, policy makers and advisors, educators.
FAIR	<p>FAIR is an acronym that describes attributes of data in terms of its shareability. The acronym stands for:</p> <ul style="list-style-type: none"> <li>• <i>Findable</i> – This attribute defines how easily the data can be found based on its metadata. Data that is richly described and tagged, and that has unique identifiers (such as DOIs) is considered findable.</li> <li>• <i>Accessible</i> – This attribute defines how easily the data can be accessed, based on where and how it is shared. For data to be accessible it must be able to be retrieved by both humans and machines</li> <li>• <i>Interoperable</i> – This attribute describes how well the data can be integrated with other data and data centres.</li> <li>• <i>Reusable</i> – This attribute describes how ready a dataset is to be re-used or repurposed. This includes determining how applicable the data is outside its own initial purpose, as well as its provenance and how attributable it is</li> </ul>
Service	<p>The word service has a number of real-world definitions, usually regarding an amenity or facility that is performed for someone (e.g. a cleaning service or a ride-sharing service).</p> <p>However, in terms of systems and data delivery (and in terms of <i>Digital Antarctica</i>), the term “service” refers to a piece of software that exposes and delivers data or functions from a system to an external source. A web service hosted by a data centre, for example, allows a person or system to access some of that data centre’s data via the web without being granted access to the data centre’s whole systems.</p>
The Clarke Review	<p>The Australian Antarctic Science Program Governance Review, published in 2017, by Drew Clarke. Available at: <a href="https://www.environment.gov.au/antarctic-review">https://www.environment.gov.au/antarctic-review</a></p>