Digital Antarctica

Refined Scope

August 2021

Contents

Version information

Version	Description	Author	Date
1.0	Initial draft	Rob Jennings	17/8/2021
2.0	Release Version	Rob Jennings	18/10/2021

Executive Summary

Digital Antarctica is an Australian Antarctic Program Partnership (AAPP) initiative defining standards which will bolster Australian Antarctic research by ensuring 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.

This document defines what will be delivered as part of *Digital Antarctica*, the scope of the content within *Digital Antarctica*, and the responsibilities of the organisations within the AAPP towards creating a fully realised and interoperable data landscape. It also touches on the pathway to get to that point, and the users it will benefit.

The vision of the *Digital Antarctica* initiative is to support Australia's Antarctic treaty interests and obligations by defining an interoperable, comprehensive and sustainable approach to data sharing. This vision will keep the initiative working towards its goals of creating a comprehensive model of Australian Antarctic and Southern Ocean data; providing guidance to data creators and data curators; encouraging the use of best-practice principles; and fostering collaboration between the AAPP partners.

Digital Antarctica is a series of standards which define how data providers can share their data in a "Findable, Accessible, Interoperable and Reusable" (FAIR) way. Using the *Digital Antarctica* standards, data centres will be able to build services to serve their data, and anyone wishing to access those data will be able to build services into their tools to retrieve the data. The *Digital Antarctica* standards will also be an ongoing guide to the best-practices around serving Antarctic data.

Initially, *Digital Antarctica* will facilitate access to all publishable Australian Antarctic and Southern Ocean research data and relevant ancillary data stored by the AAPP partner organisations. This includes all observations and data products created by Australian research organisations which focus on, or are created in, the Australian Antarctic Territory; the Southern Ocean; and the regions that impact, or are impacted by these regions. As *Digital Antarctica* is an AAPP initiative it will initially be limited to the data served by the AAPP organisations, however as it matures, the scope may move beyond these organisations, and beyond any of the scope elements. Additionally, each data centre will have full control over what data are available via *Digital Antarctica*. The *Digital Antarctica* standards will detail the technical parameters the services will use for automatically including or excluding data, and will also provide guidelines and describe functionality for data managers to manually include or exclude data as required.

Digital Antarctica is a collaborative effort involving all of the AAPP partners. The creation of the *Digital Antarctica* standards is the first step towards helping the partners create a fully realised and interoperable ecosystem of services that will enable a deep and comprehensive sharing of data among all user types across the whole Australian Antarctic data landscape.

Introduction

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"[1].

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

In 2020 the Digital Antarctica Reference Group was created, primarily consisting of representatives from the data centres of each of the AAPP organisations as well as additional interested parties. The project has so far produced high-level scope and current state documents to help set the foundation for the work to come, which will include documents to describe

- a more in-depth state of the data capabilities relevant to the scope of *Digital Antarctica*;
- prototype definition;
- requirements;
- high-level design concepts;
- transition plans;
- recommendations.

Purpose of this document

This document defines the scope of *Digital Antarctica*. It is the culmination of a number of workshops held with the reference group and is based off the earlier high-level scope document, which was used to generate in-principle agreement to the concepts explored in this document. The prior high-level document is not a pre-requisite to this current document: all content and concepts from the high-level document which are relevant to the scope of *Digital Antarctica* are represented in this refined document.

The main purpose of this document is to define the boundaries within which the *Digital Antarctica* project will function. This document is not a requirements or design document. While the document does cover some elements of the requirements and functionality of *Digital Antarctica*, it does this with the purpose of illustrating its scope, rather than as proposed design.

This document is also intended to communicate the general *Digital Antarctica* concepts and may be distributed to anyone with an interest in *Digital Antarctica*.

---

[1] *Australian Antarctic Science Program Governance Review (2017) by Drew Clarke AO PSM FTSE. Available at: https://www.environment.gov.au/antarctic-review*

Digital Antarctica overview

Digital Antarctica is proposed as a series of standards and guidelines that define how services within a data centre are used to present and share Antarctic data. Once services are built to the *Digital Antarctica* standards they will, as a whole, create a federated collection of independent, but interoperable, services across the data centres within the AAPP. This collection of services will be used by any tool requiring access to Antarctic research data. This could include a data portal or an analytics tool. Each tool will use services, also built to the *Digital Antarctica* standards, to interrogate the *Digital Antarctica* data delivery services to find and access Antarctic data from those data centres.
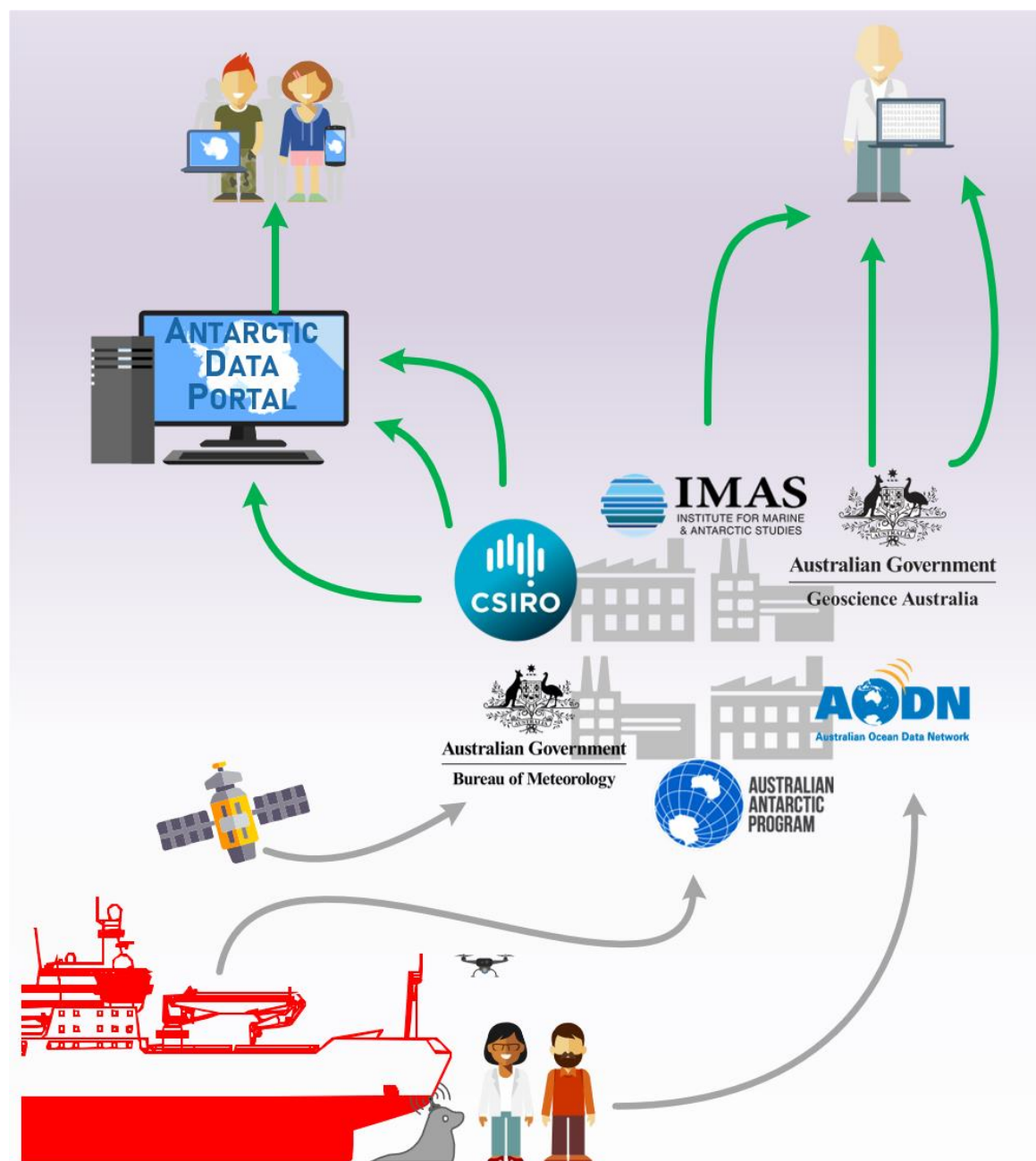
Figure 1 - Overall flow of data from collection to consumption

Figure 1 shows the overall flow of research data that *Digital Antarctica* will facilitate. Data are collected by researchers and devices, and archived at one or more of the AAPP Data centres. Those data centres make their data available via services. A tool, such as a portal, uses its own services to retrieve the data from the data centre services before displaying them to its users. Researchers also use services to connect their analytical tools directly to the data centres.

Figure 2 - A basic view of the Digital Antarctica model

Figure 2 (above) shows a basic view of the *Digital Antarctica* model. Each data centre builds the web services it requires (shown as a green box with arrows) for the data that it houses, built to the *Digital Antarctica* standards. A data portal wishing to retr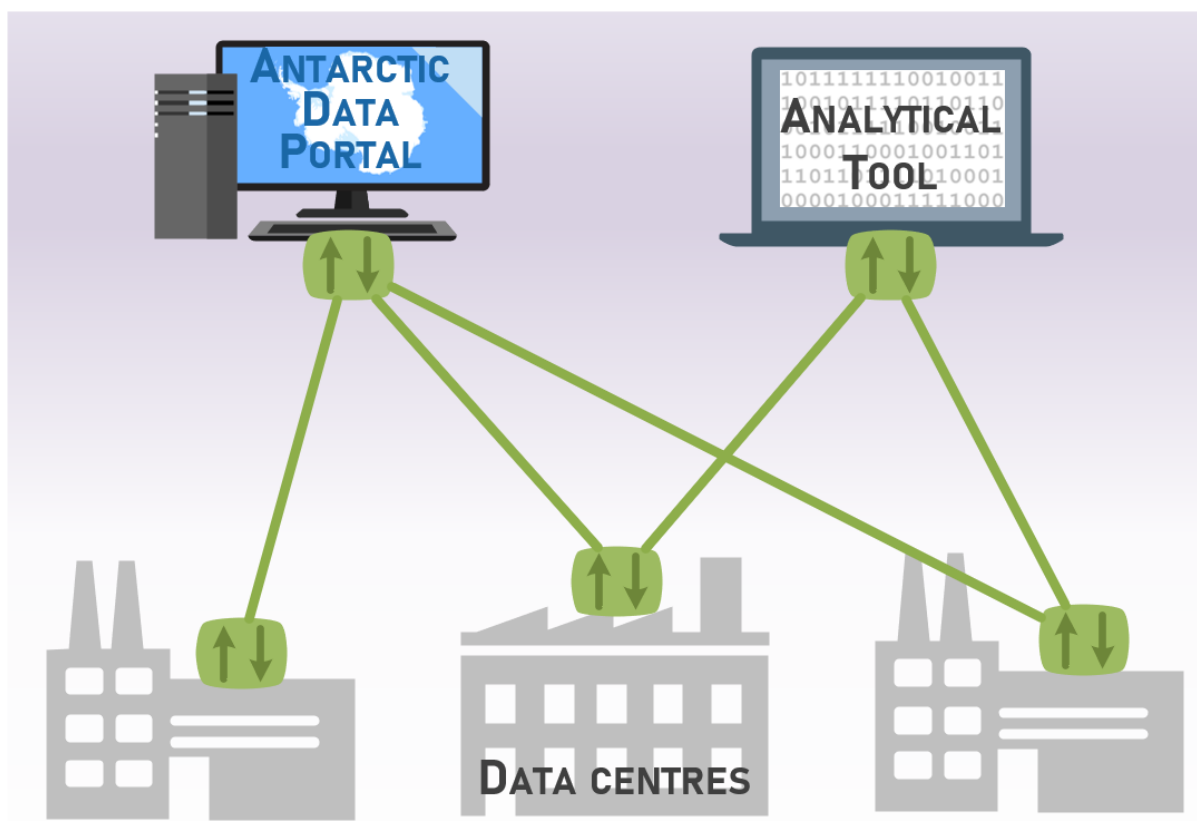ieve Antarctic data also uses web services built to the *Digital Antarctica* standards to access those services that match the features of the portal. A researcher using an analytical tool may also access a subset of the data (relevant to the requirements of the analysis) using services built to the *Digital Antarctica* standards.

See The Digital Antarctica services and considerations section below for more details on the service model.

Vision

The primary vision of the *Digital Antarctica* initiative is:

> *To support Australia's Antarctic treaty interests and obligations by defining an interoperable, comprehensive and sustainable approach to data sharing.*

In itself, the vision is not intended to be a definitive statement of scope, but instead should be used to give direction and guidance to the work being performed through *Digital Antarctica*. Every component of *Digital Antarctica* should contribute in some way to the vision, and should be undertaken with the vision in mind.

Goal

The primary goal of *Digital Antarctica*, as taken from the Clarke review, is to create "a data model of the Australian Antarctic Territory"[2] which demonstrates that Australia is a leader in Antarctic data,

---

[2] *Australian Antarctic Science Program Governance Review (2017) by Drew Clarke AO PSM FTSE*

supporting Australia's Antarctic Treaty obligations and bolstering its position in Antarctic negotiations. A major concept within the review is that data should be seen as core science infrastructure with the same regard as other infrastructure (e.g. a ship or a station). This strong focus on data will help Australia not only meet its 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. *Digital Antarctica* aims to contribute to this by providing facility for data centres to standardise their data delivery and create services that can interoperate to provide a cohesive view of the Australian Antarctic data landscape.

Another primary goal of *Digital Antarctica* is to assist data users in all forms. "Data users" in this sense includes data creators - such as researchers; data managers - including anyone involved in the curation and maintenance of data; and data consumers - such as researchers, data officers, policy makers and their advisors, educators, and government departments. A comprehensive model of Antarctic data will assist data users in publishing and accessing the most relevant data for their needs, in a way that gives them the most benefit.

The outputs of the *Digital Antarctica* initiative will assist data managers by providing ongoing guidance on best practice in capturing, storing and sharing data so that they are accessible via *Digital Antarctica*. This will help and encourage data providers to ensure that their data are Findable, Accessible, Interoperable and Reusable (see more on these concepts, and how they apply to *Digital Antarctica*, in the FAIR section below) in order to be made available via *Digital Antarctica*. This, in turn will

- facilitate data analytics. This will improve effectiveness of the data for end users;
- increase the value of data. Because of its remoteness, Antarctic data are 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.

Digital Antarctica Scope

Digital Antarctica will facilitate access to all publishable Australian Antarctic and Southern Ocean research data and relevant ancillary data stored by the AAPP partner organisations.

Any set of data that is currently shared by the AAPP organisations will be available to the *Digital Antarctica* services as long as at least some of the data within the dataset fall completely within the boundaries of the statement of scope. This scope statement can be broken down into a number of terms that can be further defined:

- Publishable
- Australian (from a geopolitical point of view)
- Australian Antarctic and Southern Ocean research
- Research data and relevant ancillary data
- Stored by AAPP partner organisations

For data to be made available to a *Digital Antarctica* service, they must meet the criteria for each of these terms, as defined below:

Publishable data

Publishable data in this context are data that meet the criteria for public access. Each data centre has its own data policy that covers how data are licensed and the criteria for publication. Each centre also provides different levels of access to data depending on the data content and the intended audience. For example, some data may be commercially sensitive and only available to members of a particular organisation.

Digital Antarctica will only provide access to data that are licensed to be freely available to the public.

Australian data

Australian data, in a geopolitical context, are data that are owned, or have been generated, by an Australian research program or a team that includes Australian researchers.

Most data in the AAPP data centres will meet this criterion, however if a data centre currently serves data that has no Australian contributors or components, the data should not be available through *Digital Antarctica*.

Australian data products available to *Digital Antarctica* may be based on non-Australian source data. There will be a requirement for all data, including these, to be shared with full provenance, with links or guidance leading to source data, where available.

Australian Antarctic and Southern Ocean data

Australian Antarctic and Southern Ocean data are data produced by research performed in, or research focussed on, the Australian Antarctic Territory; Australian sub-Antarctic islands; the Southern Ocean; and atmosphere and ocean waters which impact on, are impacted by, or geographically overlap with, the Southern Ocean.

The Antarctic Treaty is limited to any area south of 60°S. Likewise, the Southern Ocean as defined by the Australian Hydrographic office (Figure 3)[3], is the body of water south of 60°S but which extends north of that between the Australian Antarctic Territory and Australia.

---

[3] *https://www.hydro.gov.au/factsheets/FS_Hydrographic_Surveying-Limits_of_oceans_and_seas_around_Aus.pdf*
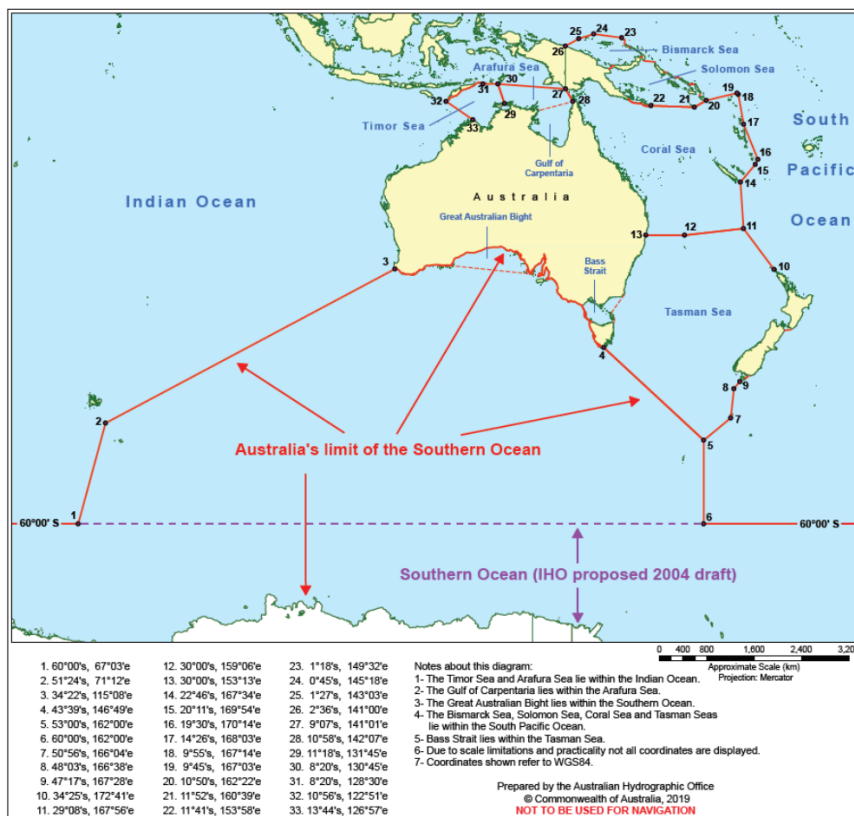
Figure 3 - The Southern Ocean limits as defined by the Australian Hydrographic Office

National geographic has recently recognised the Southern Ocean as "Most of the waters that surround Antarctica out to 60 degrees south latitude, excluding the Drake Passage and Scotia Sea"[4].

Figure 4 - The Southern Ocean limits as defined by National Geographic. Matthew W. Chwastyk and Greg Ugiansky, National Geographic Staff. Sources: NASA/JPL; International Hydrographic Organization (IHO)

---

[4] *https://www.nationalgeographic.com/environment/article/theres-a-new-ocean-now-can-you-name-all-five-southern-ocean*

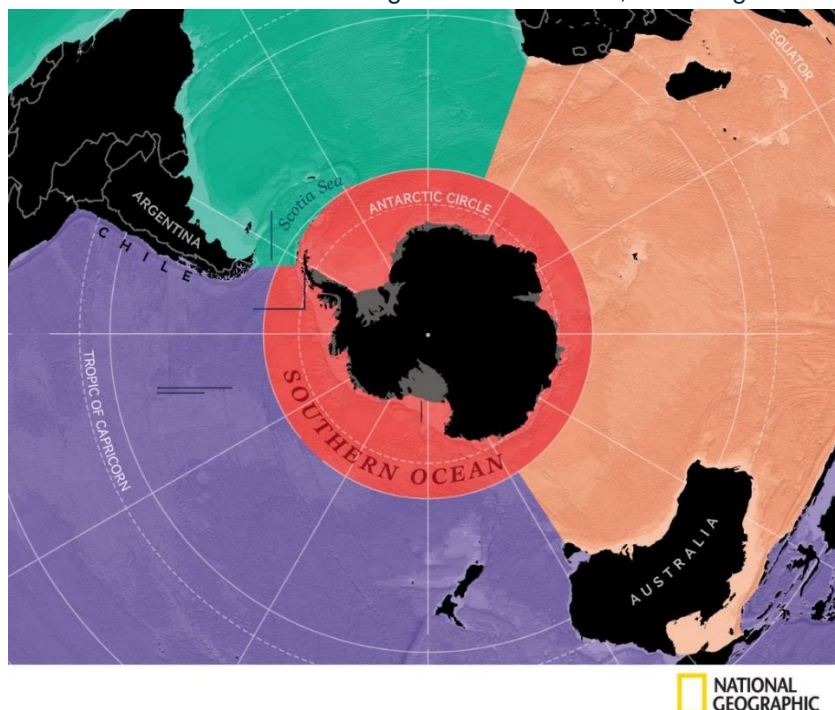However, while these definitions consider geopolitical and hydrographic properties of the ocean they do not reflect the bounds of research which are conducted in studying the Southern Ocean. As such, the *Digital Antarctica* scope will include not only the official definition of the Southern Ocean, but also the waters that geographically overlap with the Southern Ocean and which have an impact on, or are impacted by, the Southern Ocean.

Most Antarctic and Southern Ocean research data are spatial; however *Digital Antarctica* should also include and cater for any non-spatial data that are captured and relevant to Antarctica, the sub Antarctic islands, and the Southern Ocean. These data may not have metadata that make them immediately findable using co-ordinate-based parameters.

Research data and relevant ancillary data

Research data are any data that could be used for research purposes, and may include raw data; calibrated data; analysed data (at varying levels); data products; models; and model output data.

Ancillary data may include code; configuration files and specifications; and supporting documentation. Some data centres do not currently store or share all relevant ancillary data. Data will only be made available to *Digital Antarctica* if they meet the data policy of its data centre. In most cases, when ancillary data are stored, they would normally accompany the research data that they support.

Only ancillary data that are relevant to research should be shared. Operational data (such as ship logs) and business data (such as HR and finance data) would not normally be archived or shared, however depending on the research performed or the data in question, these data may be included, but only if they could conceivably be of scientific interest (either now or in the future). This determination would be made by the data centre in question, in consultation with the research team.

While data from the *Digital Antarctica* services will be used for decision making, the focus of the data centres serving the data, and therefore of *Digital Antarctica* itself is research data. Other decision support data (e.g. economic data) would not be available via *Digital Antarctica* unless they would also be classified as research data.

Data stored by AAPP partner organisations

In its initial stages, *Digital Antarctica* will only include data stored and served by AAPP Partner organisations. These organisations will be primarily responsible for building the services required to contribute to *Digital Antarctica*, and are major stakeholders in the development of the *Digital Antarctica* standards.

Data managers at AAPP organisations will curate the data that contribute to *Digital Antarctica*, ensuring that they meet the standards at the point of service and that the data that are served fall within the scope elements defined above.

By default, data that are available through an AAPP data centre but are not hosted or managed by that data centre should not be shared via *Digital Antarctica*.

Additional notes on scope
- *Digital Antarctica* will be designed in such a way that it will be scalable beyond the bounds of this scope; however, these definitions of scope apply to its initial design and release.
- In cases where only some data within a complete dataset falls within scope (for example, where data are collected from a network of devices, some of which are within the geographic scope and some of which are not) all data from the dataset will be included. This is to ensure that complete datasets are not divided based on the requirements of *Digital Antarctica*.
- See Data Inclusion below for more details on inclusion and exclusion.

The Digital Antarctica services and considerations

To deliver data to *Digital Antarctica*, a data centre will require one or more data delivery services. The services required will be defined in the Future State document later in the project, and the number of services and their features may depend on the requirements of the data centre and the current state of its data. However, each service's primary function will be to enable external interaction with the centre's data at various levels. A service may provide multiple functions and examples include:

- metadata search
- standardised metadata delivery
- standardised data delivery (this may require a number of services, depending on the types and formats of the data)
- data interoperability

Services may also be defined that enable analytics. These could come in the form of services that are designed for analytics tools to connect to, or services that themselves provide a very high level of analytics (e.g. summaries of data based on pre-defined parameters). See Practical analytics below for more on analytics.

Any product needing to access these functions may also require services to perform each of the above functions.

The main output of the *Digital Antarctica* initiative will be the definition of all of the services required and the standards those services should meet. These definitions should be comprehensive enough to enable Australian Antarctic data centres, and other organisations, to build data delivery and retrieval services that can interact with all other standardised *Digital Antarctica* services.

Together, these services built to the *Digital Antarctica* standards will create a federation of Antarctic data, accessible by any tool that is configured to interact with the services.

Most data centres have existing services that serve data to the public. These services may already meet, or may be updated to meet, the *Digital Antarctica* standards. In these cases, those existing services may be used. Likewise, the standards agreed upon by the *Digital Antarctica* initiative may be built upon existing standards. There is no default requirement to build new services or standards specifically for use in *Digital Antarctica*.

To be delivered via the data centre's *Digital Antarctica* service, all data will require metadata and must be available digitally. There may be datasets that meet these criteria, but do not meet the specifications of a specific *Digital Antarctica* service due to the state or format of the data or metadata. In these cases, the data centre may need to uplift the data or metadata to meet the standards. Alternatively, intermediate services may be created between the data and the *Digital Antarctica* services to help "translate" data to meet the required specification.
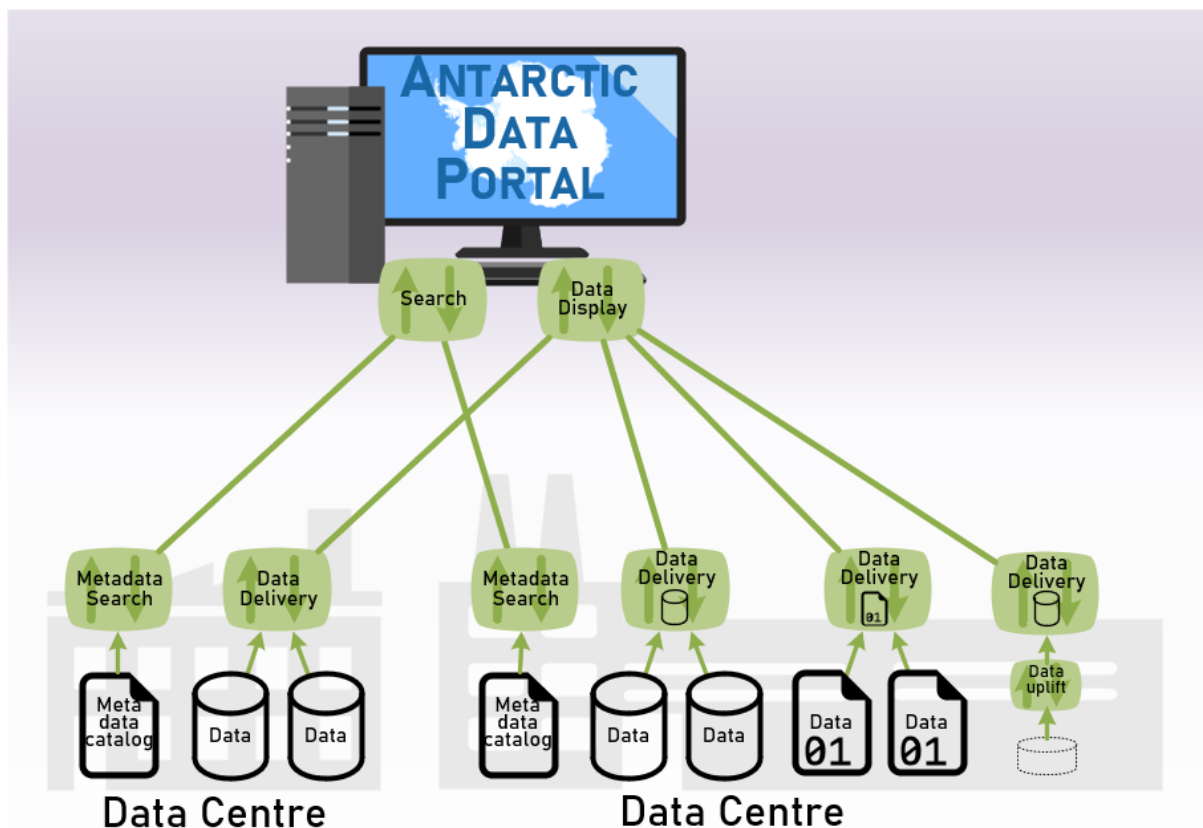
Figure 5 Service interaction

Each data access tool (e.g. a portal) will capture a user's requests and convert those requests via its own service/s into a format that the data delivery services will understand. The data centre services will take those requests and perform the actions across their data and send the results back to the access tool's services.

For example, if someone is using a portal to search for data, the portal will convert their search terms into parameters and pass them to the data centre metadata services, which will search the metadata catalogue and return the appropriate metadata records.

If someone is using a tool to download a specific dataset, the tool will convert that request into input parameters and pass those parameters to the data centre services, which will access the dataset and return the data to the portal for viewing or download.

Data Inclusion

Where practical, data within scope should be included in *Digital Antarctica* automatically through elements already defined within standard metadata that reflect the scope criteria. These criteria elements will include geographical boundaries, contribution partners, licensing et cetera. This will ensure a minimum amount of manual intervention is required to include data in the *Digital Antarctica* network of services.

The standards will also include the facility to manually override a dataset's inclusion or exclusion regardless of whether it meets the criteria for inclusion. This may be achieved through the use of a new or existing metadata tag. For example, a dataset that might meet the standard inclusion criteria may need to be excluded, and may have a specific metadata tag applied to it to mark it for exclusion. Services within the *Digital Antarctica* ecosystem will recognise this tag, and exclude the data from being served.

While the scope of *Digital Antarctica* defines the data to be included, the decision to include an individual dataset may ultimately fall to the data centre itself. To help with these decisions, *Digital Antarctica* will define a set of inclusion guidelines to act as a heuristic in helping with any decisions around the inclusion of data. This will be based around a "traffic light" system of red, indicating out-of-scope data; green, indicating in-scope data; and amber, indicating data that may be in or out of scope based on the data manager's discretion.

Out of scope	Data in this category would not, under normal circumstances, be visible via *Digital Antarctica* services. Examples include: • Data that fall entirely outside the geographic scope of Digital Antarctica and has no impact on Antarctica or the Southern Ocean • Data that are deemed commercial in confidence • Data that should not be published for ethical reasons • Ship log data that have no bearing on research data
May be in scope	Data in this category may need to be treated on a case-by-case basis to determine its availability via *Digital Antarctica* services. This determination may happen in consultation with subject matter experts, including the project and researchers involved with the data. Examples may include: • Including data that are collected outside of the geographic scope of Digital Antarctica but have significant (but not immediately obvious) impact • Excluding data that are collected within the geographic boundaries of *Digital Antarctica* but should be excluded from publication. • Excluding data collected by Australian researchers within Antarctica but are not, for other reasons, considered Australian data
In scope	Data in this category should, under normal circumstances, be visible via *Digital Antarctica* services. Examples include: • Data fully within the scope of Digital Antarctica • Data collected outside the Southern Ocean but with a direct impact on the Southern Ocean

FAIR

Digital Antarctica will assist data centres in making their data more Findable, Accessible, Interoperable, and Reusable (FAIR). The *Digital Antarctica* standards will be designed so that data served using those standards align, where practicable, with the FAIR data principles.

Findable

Digital Antarctica will enhance findability by defining data and metadata search service standards that ensure those services can be connected to from a single access point. A search portal using the *Digital Antarctica* standard services will be able to search across multiple data sources for Antarctic data.

The standards will ensure that the metadata available via the services are machine readable, so that they may be found by systems as well as people.

Accessible

The *Digital Antarctica* standards will ensure that data delivery meets the requirements of the data centres' data policies as well as the Antarctic Treaty, ensuring that data can be accessed directly and are available to the users that need them and have permission to view them. As all Antarctic data available via the *Digital Antarctica* services should be publicly available there should be no requirement to enable authentication within the services themselves.

Interoperable

The standardised inputs and outputs of the *Digital Antarctica* data delivery services will ensure that data obtained from different sources via *Digital Antarctica* can be brought together with a minimum of effort. Additionally, *Digital Antarctica* aims to enable searches for data at the individual parameter level, so that disparate datasets that contain similar data can be found and used together.

Re-usable

The standardised inputs and outputs of the *Digital Antarctica* data delivery services data will also improve data re-usability by providing data in a standardised format that can be incorporated into future work that also relies on those standards.

Additionally, services that serve data and metadata via the *Digital Antarctica* services will include the original licence and full provenance, ensuring that the data retrieved via *Digital Antarctica* services can be traced to their sources.

Appendix 1 – Pathway and Future

Digital Antarctica is a set of agreed standards to which data centres and other organisations can build services to create an interoperable model of Australian Antarctic data. The *Digital Antarctica* project itself includes a number of activities up to the inclusion of the standards and services, and is the responsibility of the AAPP, with the participation of all AAPP partners. Once that project is completed, however, the responsibility for the implementation of the *Digital Antarctica* standards and the creation and ongoing maintenance of the services will shift to the AAPP partners (with assistance where required from the AAPP).
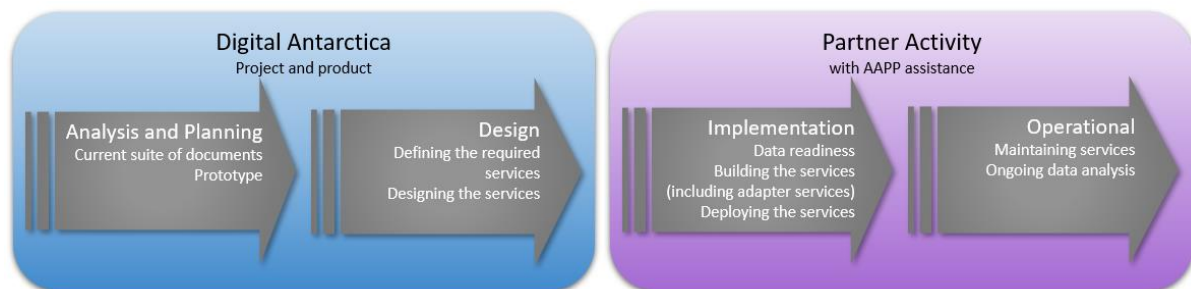

Figure 6 - Digital Antarctica project responsibilities and continuing partner activity

Many of these tasks will be outlined in later documents in the Analysis and Planning phase, but are presented here to help define the scope of the *Digital Antarctica* project, and the responsibilities that lay within and without the project. These activities may happen concurrently, depending on the approaches taken during design and implementation.

Prototype

Part of the analysis and planning phase of the *Digital Antarctica* project is a prototype to prove the concepts raised and the proposed service standards that underpin *Digital Antarctica*. The prototype will be more comprehensively defined in a separate document, and will be built within the *Digital Antarctica* project, but will not be a fully functioning releasable product itself. Rather, it will be a very narrowly focused product that will demonstrate interoperable data being served through standardised services, into an interface also using standardised services.

This prototype will endeavour to take disparate, but related, data from multiple sources across multiple organisations and bring them together in a way that highlights the interoperability features of *Digital Antarctica*. It will endeavour to cover a thin sliver of most of the functionality that a fully realised *Digital Antarctica* can provide. This should include:

- data search across multiple data sources
- data retrieval (from a user interface as well as a machine interface
- data interoperability
- analytics capability (see Practical analytics).

The results of this prototype will inform later design decisions.

Design

Once analysis and planning of *Digital Antarctica* is complete and agreed upon, work will begin on defining and designing the required services. This will use the current state documentation to create a definitive list of the service and service types required by each organisation, based on the data they plan to serve. Design can then begin on those services.

Knowledge gained from the prototyping activity will inform the design of the services.

Implementation

Implementation of the *Digital Antarctica* standards will involve not only building the services, but also analysing the data holdings of a data centre to determine

- what data should be made available via *Digital Antarctica* (see Data Inclusion above);
- the readiness of those data to be served via *Digital Antarctica;*
- the best strategy for serving data that doesn't currently meet the *Digital Antarctica* standards.

This analysis work will require a strong understanding of the data as they are stored, and will therefore be best performed by the data custodians (for the most part, the data centres), in consultation with the data experts and the *Digital Antarctica* project.

Data readiness and strategies

As the *Digital Antarctica* services are standard across organisations, they will have standardised inputs and outputs. Data at each organisation, and datasets within organisations, will be in varying states of readiness in relation to the input requirements of the standardised services.

Depending on the type and the state of the data, it made be relatively simple to alter the data to bring them into a standardised state ready for *Digital Antarctica*. However, it will not be practical to update all data to meet the standard of the service. In many cases it may be more practical to create and employ adapter services between the data and the *Digital Antarctica* services. These adapter services will be custom built to the data they adapt and will present the data to the standardised service in the format required. This will allow data to remain untouched in the archive, while ensuring they are brought to *Digital Antarctica* in a standardised from. Adapter services may be employed across a number of datasets that share attributes. This will be also determined during the analysis of the data.

Building services

Services for *Digital Antarctica,* including primary *Digital Antarctica* services as well as any required adapter services, will primarily be built by the data centre hosting the service. During implementation, the *Digital Antarctica* project will assist in the creation of the services, however responsibility for the creation and maintenance of any *Digital Antarctica* services will sit with the data centre. This ensures that a data centre has full control of the data they are serving; can make the decision to use or adapt existing services for the purpose of *Digital Antarctica* rather than create new services; and can build them in a manner that complements the data centres ongoing development strategy.

Collaborative network of experts

As mentioned above, an important part of implementation is data analysis. The scale of *Digital Antarctica* is broad, covering a wide range of data types. There will be opportunities throughout the analysis and development of *Digital Antarctica* to create a network of experts in various data fields. This could be in technical data types (e.g. file types or databases) or types of scientific data (e.g. genomics or aerosol data). This network can be co-ordinated by the AAPP, but made up of staff from the partner organisations, and used by the partners when creating, implementing, maintaining, and refining the *Digital Antarctica* standards.

Operational

Once services are implemented they will need to be maintained. Ongoing maintenance will be the responsibility of the data centre hosting the services contributing to *Digital Antarctica*.

The *Digital Antarctica* standards will also require ongoing refinement to ensure their currency. The ongoing maintenance of the standards will initially be the responsibility of the AAPP, with contributions from the AAPP partner organisations to ensure that any required changes are agreed, communicated and implemented.

Additionally, the standards implemented for *Digital Antarctica* may have an impact on ongoing data collection and storage standards. Once the standards are established and shown to have benefit, standards on how data are created, collected, and stored can be established using the *Digital Antarctica* standards as a baseline. This would not replace existing international standards, such as ISO19115, but could help guide data users in their processes.

Practical analytics

One driver of *Digital Antarctica* is to enable analytics of Antarctic and Southern Ocean data. Analysis and analytical reports use data from varying levels of detail, depending on the purpose of the analysis.

By covering all research in the Antarctic and Southern Ocean regions, *Digital Antarctica's* data offering is very broad, and encompasses a wide variety of disciplines and user types. As such it is not practical, at the first instance, to provide generic analytical tools to serve its users' needs.

In its early stages, *Digital Antarctica's* services will provide end points for analytical reporting tools to gather data for analytics. As *Digital Antarctica* matures and a deeper understanding of the required analytics is developed, data centres will be able to create services to provide data specifically for analytics (e.g. summaries of themed datasets). This deeper understanding will arise from user feedback as well as analysis of usage.

Appendix 2 – User Stories

The following scenarios, or user stories[5], show how data served via *Digital Antarctica* will provide benefits across the Antarctic research, policy, and data community.

As a researcher, I want to supplement my own data with other data that has already been collected so that my research can be well supported.

A researcher looking to find data related to their current research will be able to use a fully implemented *Digital Antarctica* by

- looking for all data of a particular type collected at a particular region and time;
- looking for all data containing a particular parameter or set of parameters from a particular time and/or place;
- looking for intersections of data types, such as bathymetry and AUV imagery; and
- looking for all data collected by a particular voyage or project.

As a research planner, I want to see what data have already been collected so that I can plan my observations around the gaps in existing understanding.

A fully implemented *Digital Antarctica* will provide functionality for anyone planning research, projects or expeditions to

- look for existing data in the discipline of the research or project;
- find areas of data within those fields that are under-represented and could be bolstered by further research.

As a researcher, I want to ensure that my data are findable, accessible, and reusable by as many people as possible so that they can help others and so that interest in this field is maintained.

Digital Antarctica will consider the FAIR principles when developing its standards to ensure that the data served via its services will

- be findable using various search parameters, including contributor or project;
- show full provenance;
- be available to other data providers by standardised service end points.

As a policy maker, I want to find data that supports the decisions I make, so that policy can be backed by empirical evidence.

A fully implemented *Digital Antarctica* will provide functionality for policy makers or their staff to

- find data of a particular type that shows change happening over time;
- find research relating to policy based on categories or themes;
- find researchers associated with particular types of data.

As a member of the general public, I want to be able to find Antarctic and Southern Ocean data and information so that I can use them in my school, work or area of general interest.

The data served via *Digital Antarctica* services will be available to all members of the public. And while the general public use of data is difficult to anticipate, the *Digital Antarctica* services will be able to be used to provide data for a number of personal uses. Examples include:

- searching for and accessing imagery for use in art projects;
- finding information about special areas of interest for personal education or school assignments.

---

[5] *See the glossary for more details on user stories*

Appendix 3 – Glossary

Term	Description
AAPP	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: • University of Tasmania • Institute for Marine & Antarctic Studies • The Australian Antarctic Division • CSIRO • Bureau of Meteorology • Geoscience Australia • Tasmanian Government
Data User	A data user is anyone that interacts with data. Specifically, there are 3 broad categories of data user within the *Digital Antarctica* ecosystem: • *Data creator* – this is anyone who generates research data for the purpose of sharing within *Digital Antarctica* • *Data manager* – 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 • *Data consumer* – this is anyone who uses data that has been shared e.g. researchers, government departments, policy makers and advisors, educators.
FAIR	FAIR is an acronym that describes attributes of data in terms of shareability. The acronym stands for: • *Findable* – This attribute defines how easily the data can be found based on their metadata. Data that are richly described and tagged, and that have unique identifiers (such as DOIs) are considered findable. • *Accessible* – This attribute defines how easily the data can be accessed, based on where and how they are shared. For data to be accessible they must be able to be retrieved by both humans and machines • *Interoperable* – This attribute describes how well the data can be integrated with other data and data centres. • *Reusable* – This attribute describes how ready a dataset is to be re-used or repurposed. This includes determining how applicable the data are outside their own initial purpose, as well as their provenance and how attributable they are.
Service	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). However, in terms of systems and data delivery (and in terms of *Digital Antarctica*), 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.
The Clarke Review	The Australian Antarctic Science Program Governance Review, published in 2017, by Drew Clarke. Available at: https://www.environment.gov.au/antarctic-review
User story	A user story is a software development tool that describes a user, a task that user wishes to perform, and a reason that the user wants to perform the task. They are usually written in the format of "As a [user or type of user], I want to [an action that the user would like to perform] so that [a goal that performing that action will achieve]." User stories highlight the various users of a tool, and the benefits that the tool provides them. They can be used during development as a measure of progress or success, and also help to give a personal perspective to user requirements.