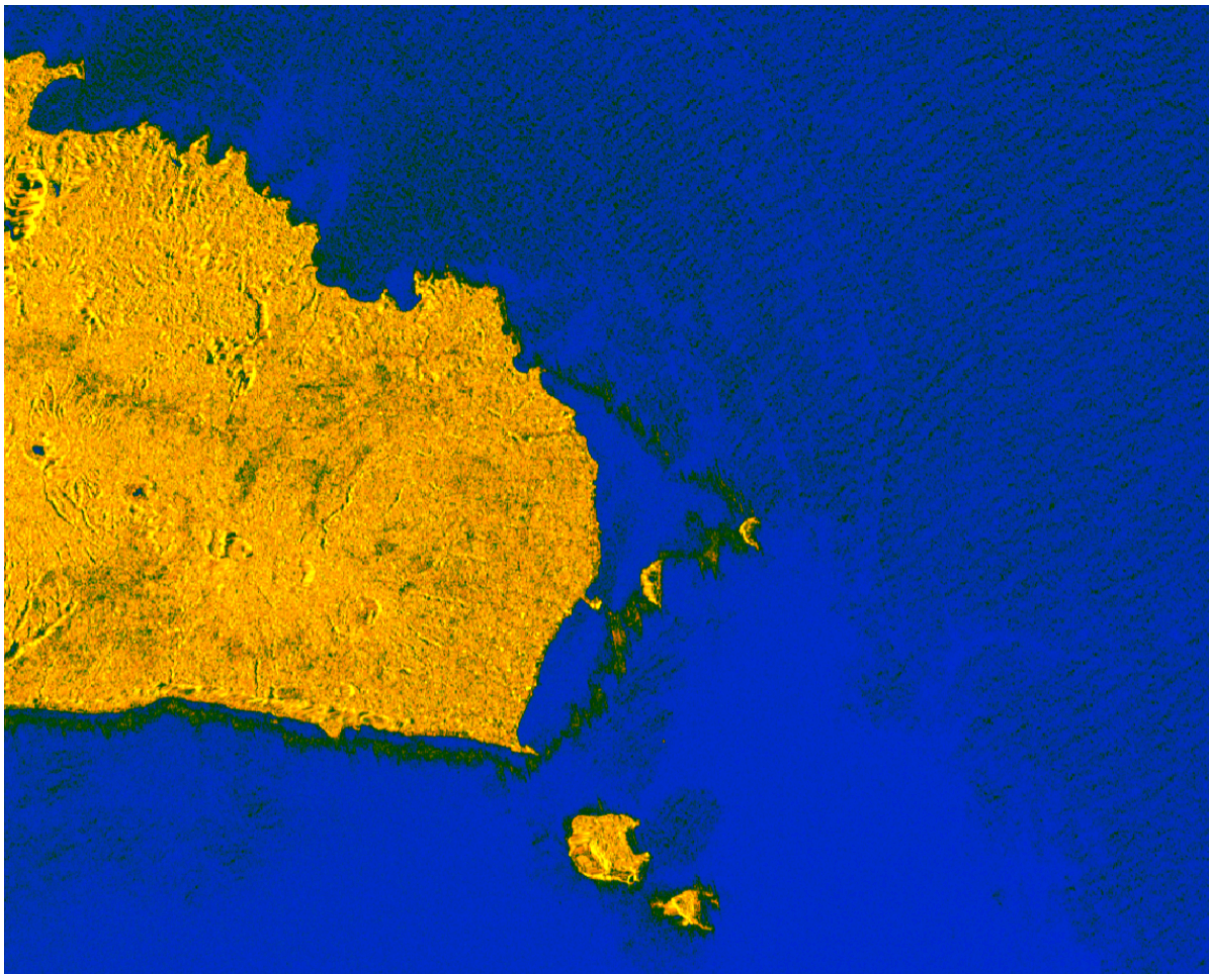
Digital Earth PACIFIC



Digital Earth Pacific Technical Roadmap

Version 1.0, September 2023

Cover image from Sentinel-1 over Samoa.

Document control

Version	Date	Change	Authors
0.1	7 July 2023	Initial draft	Alex Leith
0.2	7 August 2023	Second draft incorporating feedback from across the team	DEP team
1.0	29 September 2023	Version 1.0 including feedback from Andiswa	Alex Leith and Andiswa Mlisa

Table of Contents

Executive Summary

The Digital Earth Pacific Technical Roadmap outlines a strategic plan to empower the Pacific region through enhanced Earth and Ocean Observation data access.

The Pacific region's diverse geography and needs necessitate a flexible and scalable platform for data. Existing reports highlight opportunities in expanding regional capability with respect to the use of Earth observation (EO) tools, alleviating data accessibility challenges, and supporting improved regional collaboration. Key thematic needs encompass land use, marine environments, water security, food security, disaster management, climate adaptation and mitigation, and sovereignty.

The DEP program envisions a collaborative ecosystem based on open data principles and analysis-ready cloud optimised (ARCO) data formats. The program seeks to empower diverse users through the provision of documentation, training and support for accessing, visualising, and analysing large-scale data. The platform will integrate with existing technology and data infrastructure, follow cybersecurity best practices, and foster a culture of knowledge sharing and community building.

The roadmap is divided into three phases as follows:

1. Empower more Pacific people to access data:
    - Phase 1 (2023-2024): Develop capacity through delivery of interactive workshops and provision of scientific computing resources.
    - Phase 2 (2025-2027): Enhance engagement, documentation, and training and support the establishment of a community of practice.
    - Phase 3 (2028-2030): Expand capacity and training, catering to evolving needs.
2. Simplify access to Earth and Ocean Observation data:
    - Phase 1 (2023-2024): Create a secure and maintainable public digital data infrastructure platform with core functionality.
    - Phase 2 (2025-2027): Improve technology, automate data pipelines, and ensure cybersecurity best practices are being met.
    - Phase 3 (2028-2030): Maintain and expand infrastructure.
3. Deliver new products specific to the needs of the Pacific:
    - Phase 1 (2023-2024): Define a product strategy and develop initial offerings (e.g., coastlines, land cover, mangroves and seagrass).
    - Phase 2 (2025-2027): Implement strategy, review and update strategy, introduce more products.
    - Phase 3 (2028-2030): Continuously review the strategy and deliver new data products based on user needs.

The DEP Technical Roadmap is a comprehensive guide that envisions a collaborative and accessible ecosystem for EO data in the Pacific. By embracing open data and cloud native paradigms, and with a commitment to capacity development, DEP aims to empower people from across the Pacific to more easily access and analyse Earth and Ocean Observation data.

Overview

The Digital Earth Pacific (DEP) Technical Roadmap is intended to be a high level plan that guides the DEP team in terms of technical direction. The document is broken up into three sections: this overview, covering the background and justification for the program, the technology vision, which talks about our goals and finally, the roadmap, which talks about how we're going to achieve the goals.

Background

There have been a range of reports prepared over recent years that help set the background for the DEP program and a selection of these are summarised below.

According to the CSIRO EO Platform Report from 2019, the Pacific region exhibits significant diversity in geography, governance, needs, capacity, and infrastructure. The report recommends that a flexible and scalable platform with associated Earth Observation (EO) data archives is necessary to support this diversity. It is noted that decision-makers in the region often lack knowledge about the capabilities and limitations of EO tools and data for achieving desired outcomes. Additionally, the report highlights the absence of regional communities of practice to share expertise, tools, and data, along with challenges in accessing available data. Supporting the further development of a community of practice and a geospatial market and industry is a significant opportunity for DEP.

Thematic needs identified in the platform report include the following:

- Make land use cover and potential information available for Pacific users
- Empower Pacific Islands people and develop more resilience for coastal and marine environments
- Leverage EO to improve water security, quality and quantity for Pacific countries and territories
- To use and leverage available data for food security
- To better integrate EO science in the marine domain to improve sovereignty and security
- [To investigate] how to best manage disasters in a coordinated way.

In the Pacific Earth Observation Coordination Meeting Report from 2019, it was emphasised that training programs should be tailored to meet specific needs and capacities of local communities. It suggests that the focus should shift from process-based training to core principles and theoretical knowledge that can be applied in new areas. Furthermore, the inclusion of Unmanned Aerial Vehicle (UAV) data, was recommended to enhance the range of available data sources.

The more recent Digital Earth Pacific Needs Assessment from 2021 identified primary use cases in agriculture, climate change, conservation, disaster management, and urban development. Tier I and II needs for data products were captured based on a national workshop conducted across four countries: Republic of the Marshall Islands, Vanuatu, Tonga and Fiji. Each multi-stakeholder workshop

included representatives across government where policy, current initiatives, needs and challenges, and opportunities were discussed.

Why Digital Earth Pacific

Increasing access to data and supporting increased collaboration on the use of data will lead to better decisions being made and will enable businesses to innovate. DEP is needed because, while people are using EO data now, there is a clear need for more collaborative and scalable solutions. The emergence of cloud-native infrastructure and global analysis-ready data catalogues means that it has never been easier to access vast stores of Earth and Ocean Observation data, but this infrastructure and the training on how to use it needs to be more easily accessible to the people of the Pacific.

The Digital Earth Australia and Digital Earth Africa programs are both excellent examples that support user communities in the use of EO data through the provision of a multi-level platform. These platforms provide direct access to data as well as to web services and a visualisation platform, enabling people with different technical capabilities to exploit the data and information in a way that works for them.

Through the provision of infrastructure and data, but more importantly, by delivering a program of capacity development including documentation, self-paced training material and workshops, DEP will make it easier to use Earth and Ocean Observation data across the Pacific. DEP will also develop new decision-ready products, prioritising on what is needed by people in the Pacific.

Technology Vision

Core to the success of the Digital Earth programs is the culture of collaboration and support, including sharing of methodologies and the ability to "stand on the shoulders" of others in the community. Where possible, DEP should aim to share software under an open licence, such as Apache 2.0 and data under a permissive licence, such as the Creative Commons by attribution 4.0. The program should aim, where possible, to leverage existing work from other EO programs, and feed the work DEP does back into the broader community.

There is an existing global community who collaborate on an ecosystem of tools focussed on the new data delivery paradigm that has been made available through the use of the public cloud. Microsoft's Planetary Computer[1], for example, is made up of an extensive open data repository as well as a compute environment and simple visualisation tools. Pangeo[2], which provides the foundation for Planetary Computer, is another community platform that supports data science, although it is more of a community of practice than a specific technical solution.

[1] https://planetarycomputer.microsoft.com/
[2] https://pangeo.io/

And the Open Data Cube is a third example of a global community who collaborate on the application of software to Earth observation product development.

What these three examples have in common is the underlying open source Python libraries as well as infrastructure-as-code templates and examples that empower individuals to do work that used to be the domain of large organisations. This contrasts with commercial platforms, where usage examples may be open, but key capability is impossible to reproduce. While it can be expedient to use an existing platform, maintaining a sovereign capability or at least some level of ownership will empower people of the Pacific to reproduce the platform at a national level, or to deliver their own commercial applications using the same patterns.

The other key underlying concept is called Analysis-Ready Cloud Optimised (ARCO) data, which means storing data in formats that make it possible to access large volumes of data by streaming it into memory without needing to download it onto a local disk. There are now a number of ARCO data repositories that contain reference data products, such as the Landsat and Sentinel-2 archives as analysis-ready data stored in a cloud optimised format.

By using simple, established open technologies, and leveraging the global ARCO data archives, DEP can deliver a platform that can be used by many types of users with different levels of capability, from those consuming visualisations in a web application, through to those consuming data in integrated commercial applications.

As the DEP platform will be implemented and hosted by Pacific Community (SPC), it should be compatible with their existing infrastructure and applications and will aim to be used internally throughout the organisation. The SPC is the principal scientific and technical organisation supporting development in the Pacific region. DEP will strengthen the way SPC applies its collective capabilities in science, knowledge and innovation to serve the people of the Pacific in reaching their sustainable development goals and aspirations. In addition to serving SPC, DEP will be accessible to people from government, industry and academia from across the Pacific. The DEP platform will use the cloud native tools and processes for infrastructure and will follow cyber security best practices.

The biggest measure of the success of DEP will be how many people from the Pacific use the platform, both in a technical sense, but also as a tool to aid in decision-making. So capacity development is core to the delivery of the technology of DEP and the program must work closely with people who want to use the tools and data across SPC, government ministries, academia, technical organisations and private entities, to ensure that it can meet their needs. Ideally, new products will be developed in the Pacific, for the Pacific by Pacific people.

Context and Values

Key values that are important to the DEP program can be summarised as follows:

- User-driven and community led. Supporting people to do the work they already need to do using common tools and methods and with expanded access to data.
- A unified data lake. Prioritising interoperability with existing data catalogues and repositories rather than replicating or reproducing data.
- Incorporation of the GEO Data Sharing Principles, GEO Management principles and the FAIR principles (findable, accessible, interoperable and reproducible) supporting open data
- Appreciation and consideration of the CARE principles (collective benefits, authority to control, responsibility and ethics) supporting indigenous values around data
- Delivery through the use of open source software and open standards, including meaningful engagement with the open source geospatial community
- While working using agile methodologies, remain focused on delivering a secure, reliable and maintainable technology platform that is useful.

Goals

1. **Empower** more Pacific people to access and analyse data. Chart a course towards a culture of scientific collaboration that is self-sustaining and meaningful.
2. **Simplify** access to Earth and Ocean Observation data. Deliver a technology platform that enables people to access, visualise and analyse data.
3. **Deliver** new products specific to the needs of the Pacific. Research and develop new and adapt existing derived data products and deliver them routinely and reliably.

Roadmap

The roadmap section below is a high-level overview of the work required as part of the DEP program. These are functional requirements, rather than specific technologies, and although data products have been listed, they should be considered more of a guide and will adapt based on what is achievable and what is in demand. The roadmap is structured in three sections aligning with the goals outlined above.

The timing of the work outlined below has been revised from that in the Digital Earth Pacific Business Case Version 3 from 2022, which outlines a roadmap for the DEP project, and includes three phases, as follows:

- **Phase 1 (2023-2024)** focuses on establishing the operational, institutional, governance, and technical capabilities
- **Phase 2 (2025-2027)** aims to increase capacity and usage by engaging stakeholders, developing the platform and its use cases, and facilitating uptake
- **Phase 3 (2028-2030)** envisions a data ecosystem with networked data, users, and applications, enabling users to create their own innovations within the platform.

Empower more people from the Pacific to access data

Phase 1

- Draft a capacity development strategy
- Deliver notebooks for the scientific data computing platform
- Undertake workshops with the user community to build awareness, engagement and to identify specific use cases and needs.

Phase 2

- Continue to implement the recommendations of the capacity development strategy
- Deliver DEP documentation that supports discovery of DEP capabilities and helps different types of users to use the platform
- Create and support and develop a community of practice around the DEP Analytics Hub both within SPC and more broadly
- Design a process for power users to request expanded access to resources on the DEP Analytics Hub
- Deliver an online open courseware application, similar to DE Africa's Learning Platform or the Ocean Teacher Global Academy, empowering people to learn in their own time or to run a course in their community
- Integrate DEP visualisations in the PacificMap, TerriaMap or an equivalent easy-to-use web mapping application, to enable non-technical users to interact with the DEP information products.

Phase 3

- Continue to deliver capacity development program
- Broaden use cases and continue to improve training material.

Simplify access to Earth and Ocean Observation data

Phase 1

- Establish a secure, maintainable technology platform including the following core functions:
    - A Data Lake, or well organised store of data including appropriate metadata
    - DEP Analytics Hub, a data science environment with compute available and learning material included
    - Open Geospatial Consortium (OGC) Web Services, enabling visualisation and data access as an API
    - A Spatio-Temporal Asset Catalog (STAC) API for data discovery
- Set up a cloud-native data pipeline or workflow tool and establish patterns to follow for operationalisation of data pipelines
- Start writing architectural decisions systematically and agree on a common way of product development to use in notebooks and data pipelines.

Phase 2

- Continue to improve the technology platform including clear separation between staging and production environments
- Establish and exploit new baseline products, examples could include:
    - Water reflectance radiance
    - Sentinel-3 ARD
- Automate data pipelines and support the operational delivery of data products
- Deliver a cyber security strategy and operationalise the delivery of its recommendations.

Phase 3

- Maintain and enhance the DEP infrastructure platform based on further technical needs identified.
- Support existing data pipelines and processing, ensuring reliable and timely delivery of data
- Continue to add new baseline data products as appropriate.

Deliver new products specific to the needs of the Pacific

Phase 1

- Draft a product development roadmap
- Deliver products, including:
    - Coastlines
    - Mangroves and seagrass

    - Land use/land cover
    - Digital elevation and bathymetric models.

Phase 2

- Implement the recommendations of the product development strategy, including products such as:
    - Monitoring of extractive industries (mining and logging)
    - Forest cover
    - Crop mask, or a map of crop/non-crop land and probabilities
    - Shallow water bathymetry
    - Water observations from Space daily and annual summaries
    - Fractional cover daily and annual summaries
    - Tasseled cap annual summaries
    - Geometric median and absolute deviations on Landsat and Sentinel-2
    - Coastal water quality
    - Continuous change detection and classification
    - Flood Detection and Monitoring.

Phase 3

- Deliver further products as determined by the needs of the community

References

- The Pacific Community's GEM Division Business Plan 2023-2027
- Digital Earth Pacific Business Case v3_2022
- Digital Earth Pacific: Needs Assessment 2021
- Digital Earth Africa Technical Roadmap 2021
- CSIRO EO Platform Report 2019
- Pacific Earth Observation Coordination Meeting Report 2019
- The CARE Principles