Australian Government

Australian Centre for International Agricultural Research

Final report

Project full title

An assessment of data management and FAIR data principles across the ACIAR research portfolio

project ID	SLAM/2021/156
date published	3/05/2023
prepared by	Ben Searle
co-authors/ contributors/ collaborators	Arun Jadhav, Ashitha Ganapathy, Akanksha Nagpal Sally Stone, Martin Parr, Ruthie Musker
approved by	Dr James Quilty, Belinda Nielsen
final report number	FR2023-012
ISBN	978-1-922983-04-6
published by	ACIAR GPO Box 1571 Canberra ACT 2601 Australia

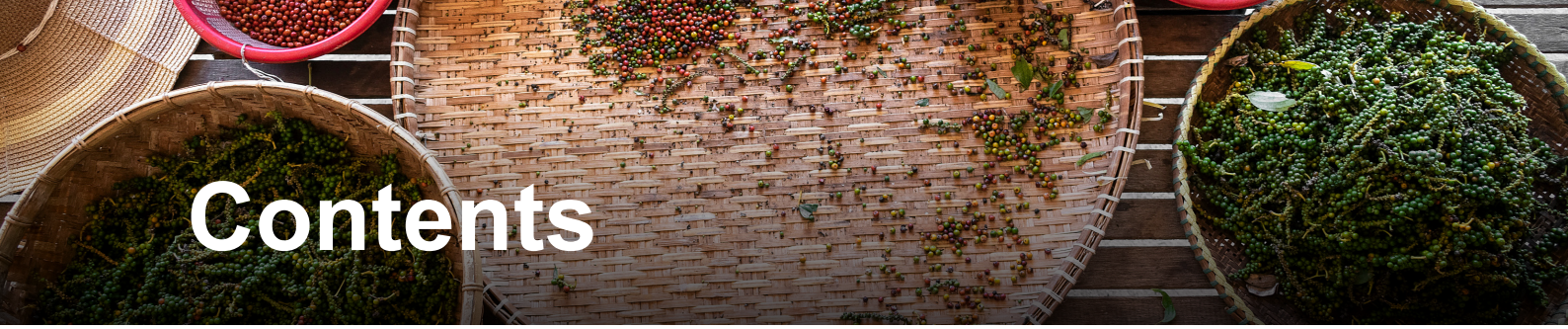
Australian Centre
for International
Agricultural Research

An assessment of data management and FAIR data principles across the ACIAR research portfolio

SLAM/2021/156

Final Report

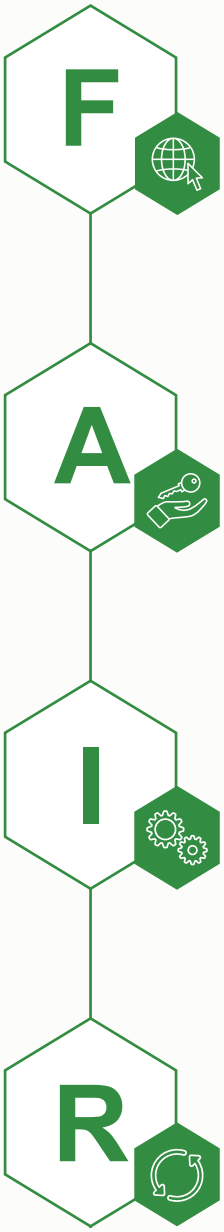
Project Name	An assessment of data management and FAIR data principles across the ACIAR research portfolio
Project Number	SLAM/2021/156
Date Published	October 2022
Prepared by	Ben Searle
Co-authors/ Contributors/ Collaborators	Arun Jadhav Ashitha Ganapathy Akanksha Nagpal Sally Stone Martin Parr Ruthie Musker
Approved By	James Quilty Belinda Nielsen
Final Report Number	
DOI	
Published by	ACIAR and CABI



Contents

Acronyms

ACIAR	Australian Centre for International Agriculture Research
APAARI	Asia-Pacific Association of Agricultural Research Institutions
ARC	Australian Research Council
ARDC	Australian Research Data Commons
ASTI	Agricultural Science and Technology Indicators
BIM	Building Information Management
CABI	Centre for Agriculture and Bioscience International
CARDI	The Cambodian Agricultural Research and Development Institute
CGIAR	Consortium of International Agricultural Research Centres
CRC	Cooperative Research Centre
CSIRO	Commonwealth Scientific and Industrial Research Organisation
DAFWA	Department of Agriculture and Food, Western Australia
DFAT	Department of Foreign Affairs and Trade
DTL	Dutch Techcentre for the Life Sciences
EAC	Eastern Australian Current
EU	European Union
FAIR	Findable, Accessible, Interoperable and Reusable
FAO	United Nations Food and Agriculture Organisation
GCOS	Global Climate Observing System
GDA	General Directorate of Agriculture (Cambodia)
GOOS	Global Ocean Observing System
IDRC	Canadian International Development Research Centre
IEAG	Independent Expert Advisory Group on a Data Revolution for Sustainable Development
IOC	Intergovernmental Oceanographic Commission
ISRIC	International Soil Reference and Information Centre
KFA	Key Focus Areas
MAFF	Ministry of Agriculture and Food, Forests and Fisheries (Tonga)
MoA	Ministry of Agriculture (Fiji)
MRC	Mekong River Commission
MWLR	Manaaki Whenua – Landcare Research (New Zealand)
NMC	National Mekong Committees
NOAA	USA National Oceanic and Atmospheric Administration
NSF	US National Science Foundation
ODA	Official Development Assistance
OSTP	The White House Office of Science and Technology Policy
PGPA Act	Public Governance, Performance and Accountability Act
PPP	Preliminary Project Proposal
R&I	Research and Innovation
ROI	Return on Investment
RPM	Research Program Manager
SDGs	Sustainable Development Goals
SPC	The Pacific Community
SRA	Small Research and Development Activity
UNESCO	United Nations Educational, Scientific and Cultural Organisation
WA	Western Australian
WMO	World Meteorological Organisation



FAIR data is Findable, Accessible, Interoprerable, and Reusable

Executive summary

CABI has conducted a Small Research Activity (SRA) comprising **"An assessment of data management and FAIR data principles across the ACIAR research portfolio"**. The SRA has consisted of an extensive study of the data management aspects within the Australian Centre for International Agricultural Research (ACIAR) and its investment projects to gather evidence of the perceptions, awareness and implementation of the FAIR data principles. Based on the evidence gained, the study developed a number of recommendations. These recommendations describe how the investment outcomes could benefit from improved data management and how the FAIR data principles could be implemented in ACIAR and across its investments to:

- improve ACIAR's return on investment
- increase investment efficiencies and in-country benefits
- provide cross-cutting approaches to break down 'silo-like' approaches to investment research
- increase the digital capabilities and capacity in recipient organisations and countries
- increase access to investment data
- enable in-country partners to lead in the publication of research activities, supporting the process of decolonisation[1] that can only occur if in-country researchers have easy access to project data
- provide a foundation data platform for supporting multidisciplinary and larger collaborative research investments to address complex issues
- meet the Australian Government's increasing policy and legislative focus on ensuring publicly funded data are open

The review was extensive and involved using a range of techniques to identify the application and awareness of FAIR data principles within ACIAR.

This report provides a broad description of what FAIR means, and CABI's approach to FAIR, including a summary of the differences between FAIR and Open Data. It also describes the methodology used by CABI to gain the evidence necessary to understand the attitudes towards FAIR within ACIAR, and the potential benefits of applying FAIR concepts within ACIAR and its investments.

[1] https://frompoverty.oxfam.org.uk/the-local-researcher-merely-a-data-collector/

The report identifies the wide-ranging, multisectoral and rapid movement towards FAIR data principles at a global level within the research community. It shows how this movement links to key aspects of ACIAR's 10-Year Vision. We discuss the increasing reliance on FAIR data principles within ACIAR's current and potential research-funding partners, and how significant globally-focused science funding organisations are adopting and, in some cases, mandating FAIR data principles in their research contracts.

Details are provided on the FAIR data principles in the context of organisational and researcher relationships, and research agreements, together with evidence of the benefits of applying FAIR in agricultural research investments. Evidence on the movement towards FAIR, and its benefits, are also presented, together with the Australian Government's view on data and how data collected with public funds is seen as a public asset. Additionally, evidence is provided on how the Australian and global research communities value FAIR, and also demonstrates the substantial scientific, economic and social benefits arising from sharing science data.

The most significant evidence and insights obtained in the study were gained through a series of interviews and discussions held with a range of ACIAR personnel and with investment project stakeholders, including several in-country project participants. As a result of these discussions, clear evidence was obtained about the level of awareness and understanding of FAIR, together with the actual application of FAIR data principles within ACIAR and its investments. Additionally, the views expressed by the investment project researchers show the high level of desire to improve current data management practices and data sharing activities within and between investments. Underpinning this perspective is a strong conviction regarding the need for guidance and suitable frameworks, together with training, and other capacity building approaches, to improve investment data management and facilitate data sharing.

“
The most significant evidence and insights obtained in the study were gained through a series of interviews and discussions”

Strong views were expressed on the potential benefits of ACIAR having a data strategy to underpin and better exploit cross-cutting capabilities within its investment portfolio. Some of these benefits that were highlighted include leveraging and building on existing agreements, and reusing data and relevant tools created in previous investments to reduce costs and improve research outcomes. It was suggested that an effective data management (FAIR-based) platform would provide a powerful cross-cutting platform between investments and would facilitate cross-cutting activities between the 10 ACIAR program areas. Within ACIAR there is already a recognition of the benefits of implementing cross-cutting initiatives, such as those achieved through gender and monitoring and evaluation activities.

In the interviews and discussions, participants also highlighted the considerable diversity in regard to data literacy and data management capability, and the capacity gaps. The information provided suggests that strengthening data management capabilities would enable de-colonisation in recipient countries, by improving the ability of in-country researchers to access well-managed project data more easily, thereby enabling them to lead on the publication of research outcomes. Applying FAIR principles would also help in providing equal access to the research data to enable analysis and interpretation of the research outputs, which would further strengthen in-country capabilities.

Implementation by ACIAR of a FAIR-based data management approach that facilitates data sharing is seen as critical to reducing the current inward focus of investments. Adopting a FAIR perspective in ACIAR's investments, by demonstrating the application of FAIR principles to enhance multidisciplinary and multi-agency projects aimed at addressing complex issues, would be viewed favourably by other potential investment partners. FAIR has the potential to provide a framework that underpins and supports integration across the 10 ACIAR program areas, which would enhance the support for multidisciplinary investments.

Leveraging FAIR to deliver the 10 Year Strategy

The CABI study was undertaken in the context of ACIAR's 10-Year Strategy (2018-2027), which aims to develop more collaborative and holistic projects and platforms to solve complex problems. Other reasons for moving towards a FAIR data principles-informed approach to investments articulated in the 10-Year Strategy (2018–2027) include the following:

- It would support ACIAR's aim to increase its collaboration and joint funding with other global and regional research donors, in line with the ambition voiced by Hon. Julie Bishop MP, the Minister of Foreign Affairs, regarding "Building a scientific platform for tackling some of the biggest issues facing our region".
- It would support building a knowledge base for tackling the biggest and most complex challenge of our time.
- It would be in line with the recognition that any ACIAR investment in development needs to be highly strategic and catalytic.
- It would support ACIAR to invest jointly with larger development donors in co-designed and co-managed initiatives, in a single country, across several countries or across a region.
- It would support co-investment partnerships. Such partnerships signify strong trust between institutions and can enable ACIAR to leverage its resources, access complementary expertise, as well as engage in larger and more ambitious research programs than ACIAR could fund alone.

The observations made, evidence gathered and multiple drivers identified within ACIAR and at national and international levels led to the development of a comprehensive set of recommendations. These recommendations aim to provide a pathway forward for ACIAR to commence a journey towards applying a FAIR data-compliant approach to its research investments while acknowledging the broad diversity of data literacy, capability and capacity that exists within ACIAR investment participants. The recommendations provide a stepped approach that is achievable for each investment, but one that provides a data management stretch target commensurate with the investment participants' level of data management maturity.

CABI's recommendations propose a comprehensive ACIAR data framework that will provide a portfolio cross-cutting capability that will benefit all ACIAR investments and that will improve the effectiveness and outputs of its investment projects. To achieve this, CABI suggests a multifaceted approach to implementing an ACIAR interpretation of FAIR data principles and the creation of an ACIAR data strategy. The recommendations cover the following:

- a data strategy
- data management policies
- project data governance and sharing agreements
- business process changes
- increasing project setup effort
- ACIAR data personnel
- data management tools
- data management training and education
- a FAIR implementation road map

A transition to an effective FAIR-compliant method of operation for ACIAR will require activities across several key areas, including:

- data management processes
- knowledge and skills
- strategic oversight
- investment
- monitoring, evaluation and compliance

With an ACIAR data strategy providing a vision for the future and providing the appropriate tools, guidance and support, ACIAR will have the capacity to further strengthen its leadership role in the agricultural research investment space.

It is anticipated that this report will provide a stimulus and a roadmap on how ACIAR can usefully establish and implement its own interpretation of the FAIR data principles to help it meet its corporate strategies and goals, while supporting the Australian Government's data sharing objectives. This report can provide a starting point for the journey to FAIR compliance through providing guidance on applying FAIR as a flexible strategy to improve the impact of research and to address many of the issues relating to data sharing that concern project participants.

> "More usable data from projects for policy makers is required"

The journey towards FAIR data compliance is considered critical for ACIAR to be able to meet broader science and governmental objectives and expectations, and to support the increasing focus on FAIR data that is developing rapidly across international funding entities and within the Australian Government. A move to FAIR is also considered to be essential for improving ACIAR's return on investment, and for helping achieve ACIAR's strategic objectives by helping to support larger, multi-organisation and trans-disciplinary projects to address complex problems, and to enhance its support of the Government's foreign policy agenda.

Introduction

This report sets out the findings, analysis and recommendations from the SRA in relation to the perceptions, awareness and implementation of the FAIR data principles within ACIAR and its investments.

ACIAR is an entity of the Australian Government's Department of Foreign Affairs and plays a major role in supporting Australian foreign policy. Over a span of 40 years, ACIAR has commissioned and managed more than 1,500 research projects in 36 countries and has partnered with 150 institutions, along with more than 50 Australian research organisations. By creating opportunities for international agricultural research, ACIAR provides research organisations and Australian scientists with the chance to use their skills to benefit partner countries and in turn to contribute to Australia's foreign aid program and the achievement of its goals.[2]

> *"ACIAR's core business is to identify research priorities collaboratively with partner countries, commission research and broker research partnerships to tackle those priorities, and then manage and monitor these investments throughout the research process to maximise impact and return on investment."*[3]

To maximize ACIAR's return on investment, researchers from the various commissioned organisations must be able to collect high-quality data, feel comfortable sharing that data, and reuse others' data in order to make more informed decisions. They need to be able to cultivate the fertile medium that is data.

[2] https://www.aciar.gov.au/our-work

[3] https://www.aciar.gov.au/publication/corporate-publications/aciar-10-year-strategy-2018-2027

Data is the ‘New Soil’

“Farmers don’t grow crops simply because they have fertile soil. They have a clear and compelling goal: to feed people. Data and analytics initiatives are no different. They should be driven by a clear and compelling business goal.”[4]

[4] https://endjin.com/blog/2021/05/data-is-the-new-soil

Report elements

This introduction is followed by sections that cover:

- **Background** – an introduction to FAIR principles, CABI's approach to FAIR, and the difference between FAIR and open data.
- **Methodology** – how the SRA was undertaken through desktop studies, a review of ACIAR project documentation and interviews with key personnel.
- **Why FAIR** – describes the science communities' movement towards FAIR, the linkage of FAIR to ACIAR's 10-Year Strategy, ACIAR partners' views of data sharing, other funding organisations' data policies, the Australian Governments view on data sharing, and how these external views link to ACIAR's 10-Year Strategy.
- **Findings from interviews** – a high-level overview of the outcomes of the interviews and discussions with ACIAR and project personnel, together with an assessment of how FAIR is perceived across ACIAR projects and a 'deep dive' into three data-focused investment projects.
- **How ACIAR might proceed with FAIR implementation** – the drivers and potential incentives that could be used to implement FAIR concepts in ACIAR, and a discussion on a broad approach for ACIAR's implementation of FAIR.
- **Context** – the development of the report's recommendations based on the review of data management activities within ACIAR projects and discussions with ACIAR and the investment project personnel, together with national and international drivers for moving towards FAIR.
- **Recommendations** – the steps ACIAR could follow to implement FAIR if it sees value in working towards a FAIR data principle-based research environment.

Background

What is FAIR?

FAIR represents data that is **Findable**, **Accessible**, **Interoperable** and **Reusable**.

The FAIR data principles were developed by the Dutch Techcentre for the Life Sciences and the FORCE11 community and were published by Springer Nature in 2016.[5] The principles provide a framework to guide data producers towards effective data management and high-quality publishing, use, and reuse of research data. Since the inception of FAIR, the UK Data Service[6], Australian Research Data Commons[7], European Union (EU)[8] and the Wellcome Trust[9], among many others, have adopted the FAIR principles to encourage transparency and reproducibility. FAIR has also been a 'guiding star' for many large organisations, such as the Bill & Melinda Gates Foundation, in regard to how they conceptualise best data practices in their funded projects.

CABI's approach to FAIR

CABI's experience of working to embed FAIR with the Gates Foundation[10] and building FAIR policy for the CGIAR[11] (Section 3.4.1) has shown how FAIR frameworks are more widely accepted by stakeholders than a simple exhortation to make data 'open', which might cause concerns regarding the risk of protocol breaches and intellectual property challenges. In CABI's experience, very detailed technical definitions of FAIR can also discourage research teams, who may believe that FAIR cannot or does not apply to them if they do not already have access to highly technical data management systems.

[5] *Wilkinson, M., Dumontier, M., Aalbersberg, I.* et al. (2016) *The FAIR Guiding Principles for scientific data management and stewardship*, Sci Data ***3****, 160018 (2016)*. https://doi.org/10.1038/sdata.2016.1

[6] https://ukdataservice.ac.uk/about/

[7] https://ardc.edu.au/about_us/policies-and-guidelines/fair-policy-ardc-or-ardc-co-invested-materials/

[8] European Commission, Directorate-General for Research and Innovation, Horizon Europe, open science: early knowledge and data sharing, and open collaboration, 2021, https://data.europa.eu/doi/10.2777/18252

[9] https://wellcomeopenresearch.org/for-authors/data-guidelines

[10] https://gatesopenresearch.org/for-authors/data-guidelines

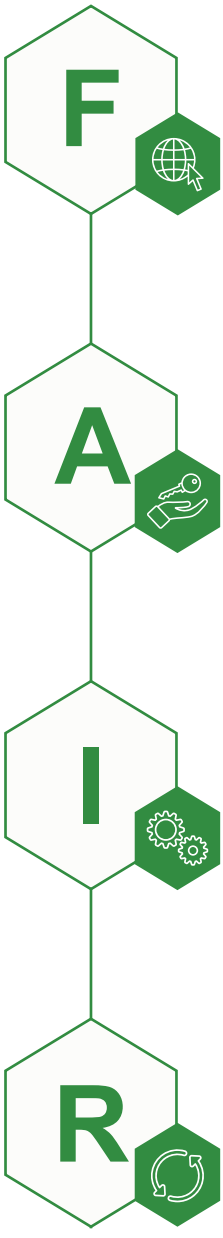
[11] https://cgspace.cgiar.org/bitstream/handle/10947/4488/Open%20Access%20Data%20Management%20Policy.pdf?sequence=1

For example, the Force 11 definition of the Findable element of FAIR is:

- F1. (meta)data are assigned a globally unique and eternally persistent identifier.
- F2. data are described with rich metadata.
- F3. (meta)data are registered or indexed in a searchable resource.
- F4. metadata specify the data identifier.

This level of detail is complex and can be overwhelming for some audiences. As different organisations have different levels of data management capability, having rigid principles can often deter some organisations from starting on their FAIR data journey. To overcome this perceived complexity some organisations are establishing their own interpretations of FAIR.

In response to this understanding, CABI more usefully defines the FAIR concept of Findable as “Can the data be found by the stakeholder who needs it?”. By following this approach, the definition of what it means to be FAIR can be tailored to different needs. CABI also routinely uses the following broader FAIR definition to guide its work with partners:



Data should be
Findable

If the implementing partner has a data requirement and the same or similar data has already been created or collected, then they should be able to find out that it exists.

Data should be
Accessible

Once the implementing partner has found out that the data they need exists, they should be able to access it so that they do not have to re-create or re-collect it.

Data should be
Interoperable

When the implementing partner accesses the data, they should be able to easily use it, so it needs to be in a format that can be readily used and, if necessary, align with other data.

Data should be
Reusable

Once the implementing partner has the data they need, it should be clear to them what they can and cannot do with it.

FAIR versus Open

While the terms Open Data and FAIR appear to be very similar in meaning, there is a subtle difference to the approach each 'classification' takes in sharing data. The GO-FAIR[12] organisation provides the following definitions and gives some clarity around the differences between Open Data and FAIR data principles.

> *"FAIR is not equal to Open: The 'A' in FAIR stands for 'Accessible under well-defined conditions'. There may be legitimate reasons to shield data and services generated with public funding from public access. These include personal privacy, national security, and competitiveness. In the envisioned Internet of FAIR Data and Services, the degree to which any piece of data is available, or even advertised as being available (via its metadata) is entirely at the discretion of the data owner."*

These differences are particularly relevant to ACIAR investment projects, since several investment participants have suggested that 'open' access to data may be problematic. Thus, the focus for ACIAR should be on the FAIR principles, since these can facilitate data sharing without making sharing mandatory or open.

The Open Data Institute views data sharing as spanning a spectrum that ranges from closed to shared to open, as shown in Figure 1 below.

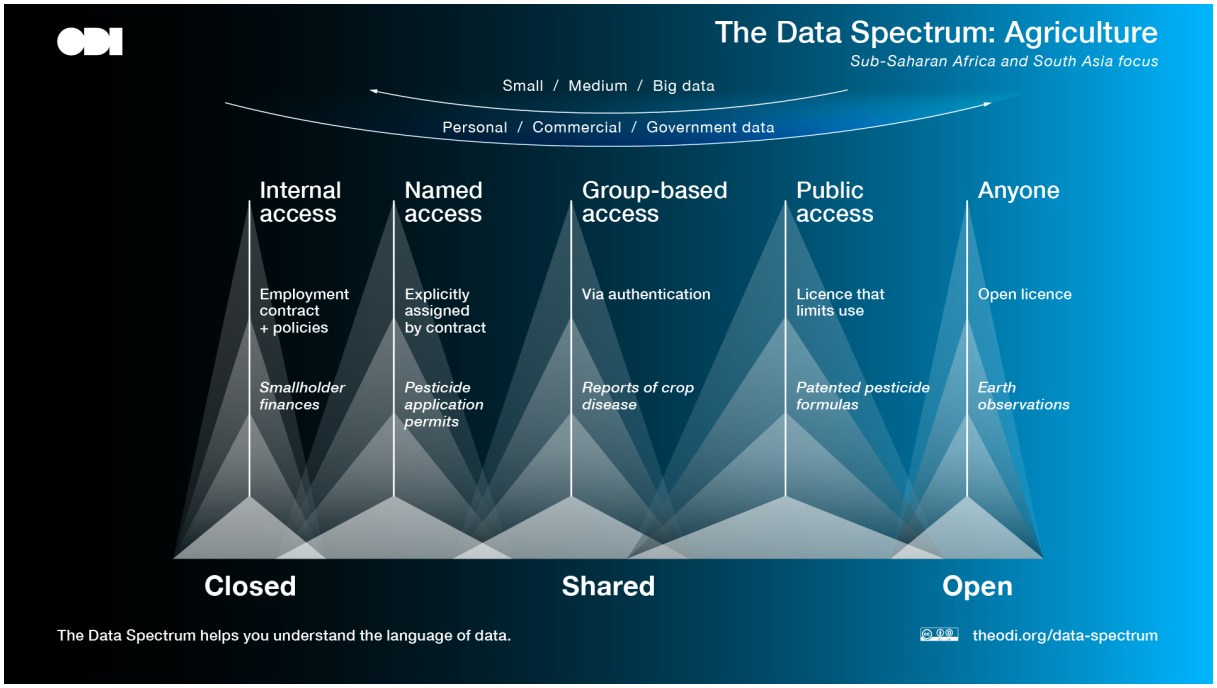


Figure 1: The data spectrum: agriculture (Open Data Institute [CC-BY])

> *"Open data is available for anyone to access, use and share for any purpose. Shared data is only available to certain people, e.g. researchers. It might contain sensitive information about individuals or groups."*[13]

The data spectrum view guides thinking about what level of access to data is appropriate in a given context, and the FAIR Data Principles guides thinking about how to maximise the value of data without prescribing that the data must be 'open'. The FAIR data approach acknowledges the 'real-world' privacy challenges that exist. Thus, restricted access data, shared data or open data can still comply with the FAIR data principles.

[12] https://www.go-fair.org/fair-principles/

[13] https://theodi.org/about-the-odi/the-data-spectrum/

Methodology

During 2022 CABI has engaged across a range of ACIAR organisational levels and with its partners to understand, analyse and document data management and data sharing activities and challenges across ACIAR's geographic regions of interest and the different scientific domains ACIAR covers.

The research methods applied included the following:

- **Desk research:** This included reviewing project proposals and related documentation, mid-term and annual reports, documents mentioned/shared during our interactions with the project teams, and existing literature, to better understand the regional context.
- **Investment-specific questionnaires:** Based on our pre-analysis of the given investment and an early assessment of the stakeholder and the organisation they represented; we created tailored questionnaires that were used to conduct our interviews with a variety of project stakeholders.
- **Semi-structured interviews:** We conducted semi-structured interviews with a variety of stakeholders within the projects, to gather diverse perspectives and to elicit relevant responses.
- **Analysing the content of reports and other project documentation using a data lens:** The focus here was on extracting information on the data-related activities of ACIAR projects.
- **Conducting a Deep Dive into three specific data-rich investments:** This involved conducting a comprehensive series of interviews and discussions with stakeholders, focusing on data and data management within the following three ACIAR investments:
  - Climate-smart landscapes for promoting sustainability of Pacific Island agricultural systems – ASEM/2016/101
  - Soil management in the Pacific Islands: investigating nutrient cycling and development of the soils portal – SMCN/2016/111
  - Land suitability assessment and site-specific soil management for Cambodian uplands – SMCN/2016/237
- **Ecosystem mapping to visualize the current state:** Ecosystem maps were used to better understand the existing relationships among stakeholders and the flow of assets among them.
- **Ecosystem mapping to visualize the envisioned future state:** Ecosystem maps were used to better understand the potential relationships, infrastructure, processes, systems, cultural changes that will be potentially required for the Deep Dive reviewed projects to achieve their vision and for the stakeholders' needs to be met.

CABI's activities were divided into the following six workstreams:

- Discovery phase: ACIAR situational analysis/alignment.
- An assessment of the state of FAIR implementation across ACIAR investments.
- A Deep Dive into regional investments with systemic-level FAIR interventions.
- An assessment of the need/demand for FAIR data implementation guidelines and resources among Research Program Managers (RPMs) and grantees
- Vision and recommendations for the next steps on how best to overcome barriers to data sharing within the scope of ACIAR grant-making practices.
- Project Management.

The details of these workstreams can be found in Annex 1.

The Movement Towards FAIR

Why FAIR?

Recent studies have estimated the annual opportunity cost of not applying FAIR data principles to be at least €10.2 billion for the European economy, while the positive impact of FAIR on potential economic annual growth is estimated to be €16 billion annually.[14]

Today, there is a considerable movement towards the application of FAIR data principles to science data. This movement is being driven by many funding entities, including national governments, and is a result of the realisation of the value of research data and the significant potential arising from its reuse. For example, Australia, through the Australian Research Data Commons, is a signatory to the Organisation for Economic Co-Operation and Development's revised version of Principles and Guidelines for Access to Research Data from Public Funding, which states:

> *"Both research data and associated digital objects must be FAIR"*[15]

Additionally, the Universities Australia Deputy Vice Chancellors (Research) Committee has developed a FAIR Access Policy.[16]

These perspectives regarding research data's scientific, social and economic value, especially for sharing and reuse, can be considered as significant drivers for beginning a journey towards FAIR data compliance within ACIAR and its investments.

A further incentive for ACIAR to move towards adopting the FAIR data principles is provided by the strong desire of the Australian Government and ACIAR to build and improve regional partnerships and improve regional capability. The establishment of deep partnerships with national organisations and scientists in other countries can create the trust required to establish the sharing of data, and to generate the many benefits that arise from this sharing. Building this 'trust' also supports Australia's foreign policy objectives.[17]

[14] European Commission 2019 Commissioned Report - http://publications.europa.eu/resource/cellar/d375368c-1a0a-11e9-8d04-01aa75ed71a1.0001.01/DOC_1

[15] https://ardc.edu.au/news/australia-signs-up-to-expanded-oecd-recommendation/

[16] https://www.fair-access.net.au/fair-statement

[17] https://www.dfat.gov.au/sites/default/files/corporate-plan-2020-21.pdf

The introduction of FAIR data principles into ACIAR projects will facilitate access to project data for in-country researchers, resulting in data being shared equitably within project teams. Through this, there is also an opportunity for capacity building and career development for partner country scientists, as it will enable them to truly engage in the analysis of data and the publication of research findings. Ideally, this will result in in-country researchers being able to more readily lead the publication of research outcomes.

> “ACIAR has a lot of data collected by projects, [but] no systems or rules for how to make data accessible.”

The benefits of embedding FAIR are not solely economic (i.e., relating to the return on investment): there is also a direct linkage to capacity building. Improving the ability of in-country researchers to publish will support the process of ‘decolonisation’ and build capacity, which is a key ACIAR objective.

> *“Research is an activity that involves several actors at different levels – both international and local. Although as part of their jobs they are all called upon to collaborate, the actors in the latter group at times find themselves locked into an imbalanced power dynamic. This can have an impact on the quality of their work.”*[18]

Additionally, in-country project participants who collect data will also be able to see the results of their data collection efforts and appreciate the value of bringing together data from multiple sources to create larger and more significant datasets more easily. Being able to integrate and merge data from different sources will increase the value and validity of the research outcome.

Global context – data sharing is evolving

With more and more scientific data being produced, perspectives on data sharing have evolved and are continuously evolving. Scientific data is beginning to be recognised as a scholarly object in its own right and academic journals are slowly making changes to adjust to this view. Nature – Scientific Data is a journal dedicated to data.[19]

> *“Scientific Data is a peer-reviewed, open-access journal for descriptions of scientifically valuable datasets, and research that advances the sharing and reuse of scientific data.”*[20]

Other journals are now strongly encouraging that a copy of research datasets be submitted along with journal articles, or that researchers submit their data to the appropriate public repository.

Additionally, many funding organisations support data sharing. For example, the European Commission’s research-funding principles provide a range of rules regarding open access to scientific publications and open access to research data (with opt-outs being possible).

Considerable effort in many sectors and within many countries is now being applied to re-aligning the focus of research from being totally output-oriented (i.e. producing scientific papers) to considering the data as an equal part of the research process. Many governments, including the Australian Government, are seeking a greater utility and return on investment from their research-funding, and projects supporting the FAIR data principles are seen as one mechanism for achieving this.

[18] https://frompoverty.oxfam.org.uk/the-local-researcher-merely-a-data-collector/

[19] (Atkins et al. 2003; Hey 2009)

[20] https://researchdata.springernature.com/users/69239-scientific-data

Supporting ACIAR's 10-Year Strategic Plan

> *ACIAR's enduring research collaborations within the region and globally are among the most effective, innovative and promising science partnerships, underpinning far-sighted policy, community and industry responses to complex challenges.*

The ACIAR Strategic Plan aims to apply three distinct partnership models:

- bilateral country partnerships
- multilateral research collaborations
- co-investment with research and development partners

Applying these models requires building strong collaborative relationships with other government agencies, agricultural research and funding organisations and universities. An ACIAR priority is to apply a more programmatic approach to its investment projects, with a focus on bringing individual projects together in a more holistic way to respond to complex problems. When such an approach is applied to multi/trans-disciplinary projects the need for data integration becomes critical.

"Across the board, trans-disciplinary approaches are needed to reduce the vulnerability of the natural resource base, and to create climate-smart agricultural landscapes. Using national policy, land-use planning and community engagement to manage water, soils, livestock, crops, forests, natural vegetation and coastal marine resources, from 'ridge to reef', in an integrated manner can increase resilience and sustainably improve livelihoods. But achieving this will require numerous and well-coordinated innovations in technology and ways of working."

One interpretation of the ACIAR Strategy is that it aims to strengthen collaborative relations with other funding and research organisations to create more significant and broader projects, leveraging ACIAR's and other partners' funding, and leveraging the achievements and outcomes from other projects to achieve more significant outcomes. As part of this process ACIAR intends to be "effective and innovative" in its approach. In this regard, ACIAR can demonstrate its leadership role through ensuring science data is sharable. This, in turn, will improve its investment projects to better support *"far-sighted policy, community and industry responses to complex challenges"*.

As ACIAR expands this collaborative approach, it is critical that its investments have an effective level of data management that can support data sharing as a key outcome focus. Strong data management will be necessary to facilitate the exchange of data and information between the organisations and researchers involved in the larger projects that are envisaged. Without effective data management and FAIR compliance, it will be difficult for other funding organisations and their research teams to use and leverage ACIAR-funded datasets, thus reducing their ability to increase knowledge and achieve greater research outcomes. Additionally, it is probable that similar types of data will be collected by multiple researchers on these larger projects. Being able to access and integrate data easily will lead to larger datasets providing more valid analysis and outputs that bring greater benefit. As projects get larger and run for longer periods, it becomes even more critical for data to be able to be shared easily throughout each project, and even more effective data management is required to ensure the data can be easily shared.

The larger, more collaboration-focused projects envisaged by the ACIAR Strategy, which are expected to address complex issues, will involve multiple researchers – often from different science fields. In order to integrate data from a range of different sources easily, the data needs to be managed and documented effectively in a form that is suitable for analysis. If the data are not in a suitable (FAIR-compliant) form, this will greatly increase the effort and cost required to integrate that data and to undertake this analysis.

Existing research partners' data sharing agendas

Some existing ACIAR research partners have their own data and data management policies aimed at supporting data sharing. Additionally, there are various mechanisms within these organisations to assist with compliance with these policies. One example is the concept developed by Monash University of the "Data Concierge" (see Annex 10), to support early career researchers with their data management activities.

Existing ACIAR partners with relevant data policies include:

- CGIAR
- the Asia-Pacific Association of Agricultural Research Institutions (APAARI)
- the Pacific Community (SPC)
- Australian universities

Their data policies and strategies are described briefly in the subsections below.

CGIAR's Open and FAIR Data Assets Policy

CGIAR has an Open and FAIR Data Assets Policy[21] (developed with assistance from CABI) that came into effect in April 2021. The background and justification for this policy is considered highly relevant to ACIAR. Extracts of the CGIAR policy are provided below.

"CGIAR's Commitment

1. As set out in the CGIAR Principles on the Management of Intellectual Assets, CGIAR adheres to the principle that the results of its research and development activities are international public goods and is committed to their widespread dissemination and use to achieve the maximum impact to advantage the poor, especially smallholder farmers in developing countries. CGIAR considers Open Access (as defined in Annex 1) to be an important practical application of this commitment as it enhances the visibility, accessibility and impact of its research and development activities. CGIAR is therefore committed to sharing outputs of its research that are as open as possible and always Findable, Accessible, Interoperable and Reusable (FAIR)[1], advancing CGIAR's aspirations to digital transformation and data-driven innovation.
2. Open and FAIR data assets improve the speed, efficiency, and efficacy of research; they facilitate interdisciplinary research; assist data aggregation, computation, and the derivation of new insights; and allow the global public to benefit from CGIAR Research. They enable CGIAR to collectively leverage the infrastructure, data pools, and new data science capacities necessary for innovation and for effective and agile responses to global challenges. They facilitate text and data mining and analysis to derive insights, recognizing that these are dependent not only on access to high quality data, but also on that data being well-contextualized (through rich metadata and relevant open materials), interoperable, and reusable.
3. CGIAR is also committed to managing data assets responsibly, with regard to privacy, ethical, and any other nationally or internationally determined considerations, in accordance with the CGIAR Research Ethics Code.

[21] https://cgspace.cgiar.org/bitstream/handle/10568/113623/CGIAR_OFDA_Policy_Approved_16April2021.pdf?sequence=1&isAllowed=y

Purpose

The purpose of the CGIAR Open and Fair Data Assets Policy ("this Policy") is to clarify expectations regarding the management and dissemination of data assets (as defined in Annex 1) to ensure that they are as open as possible, always FAIR, and managed responsibly. This Policy should be read in conjunction with the CGIAR Open and FAIR Data Assets Implementation Guidelines which may be updated from time to time to reflect current recommended practices[2].

Scope

1. This Policy applies to the management and dissemination of data assets (as defined in Annex 1) produced by CGIAR entities3 and CGIAR implementing partners in the context of CGIAR Research4, and by CGIAR's advisory bodies.
2. Indicative types of data assets include peer-reviewed journal articles; non-reviewed articles, reports, briefs, extension and training and other materials; books and book chapters; data and databases; data collection and analysis tools (e.g. models and survey tools); video, audio and images; computer software (e.g. models, APIs); web tools (e.g. data portals, dashboards). This Policy also applies to the metadata associated with data assets.

Policy

Management of data assets according to FAIR Data Principles

Data assets shall be managed in compliance with the FAIR Data Principles, as appropriate for the particular data asset type and as elaborated upon below, per agreed targets.

a. **Findability:** data assets shall be made findable within and beyond CGIAR.
b. **Accessibility:** access levels to data assets shall be consciously considered according to content and made as open as possible.
c. **Interoperability:** data assets shall be made easy to interpret by humans and machines.
d. **Reusability:** the conditions for reuse of data assets shall be clearly established upon publication."

CGIAR's Commitment statement above could be considered equally applicable to ACIAR. CGIAR has a clear understanding of the benefits that data management and open access brings to research activities and their outcomes. There may also be an expectation from CGIAR for its partners to at least be aware of this policy, and to be attempting to implement similar policies within their own or co-funded projects.

APAARI

APAARI's Strategic Plan 2017–2022[22] includes a range of activities, one of which is specifically data management-focused. Within the 'Knowledge Management' theme, Key Strategy 1.2.3 states the following:

> *"1.2 Agri-food Research and Innovation Systems (FRIS) strengthened through more effective knowledge management.*
>
> *Strategy 1.2.3 is: Improve scientific data management to make it available for analysis and knowledge creation. This objective is to be achieved by:*
>
> - *Support projects that improve data collection for research, compilation, management, analysis, evaluation and application e.g., Agricultural Science and Technology Indicators (ASTI) Project*
> - *Improve APAARI publication quality, relevance, packaging and outreach."*

APAARI acknowledges that effective data management is necessary to create knowledge and to derive the full value from data.

Pacific Community (SPC)

SPC has recently developed its Strategic Plan 2022–2031:[23] - Sustainable Pacific development through science, knowledge and innovation. This outlines seven Key Focus Areas (KFAs). Within each of these KFAs there is reference to data, statistics and knowledge. For example, the KFA regarding food systems states:

> *"Data collection and analysis support evidence-based policy making across the food–water–health–trade nexus to create healthier, more equitable food systems."*

Each of the seven KFAs indicates the importance and significance of data and how it benefits each KFA.

Australian universities

Australian universities view the application of the FAIR data principles as a key objective for their research data. This direction has been driven through the Universities Australia Deputy Vice Chancellors (Research) Committee's FAIR Access Policy, which states:

> *"All Australian publicly funded research outputs will be Findable, Accessible, Interoperable and Reusable. Access to research outputs will accord with international practices that are well defined, secure and trusted, and delivered through sustainable, fair, and efficient dissemination models. Publicly funded researchers will be expected, supported and rewarded to disseminate their work in such a way that anyone can find and re-use research publications and research data for further research, policy development, innovation, education and public benefit."*[24]

As an example, two of ACIAR's major Australian university partners, the University of the Sunshine Coast[25] and the University of Queensland,[26] have specific research data access policies.

[22] http://www.apaari.org/web/wp-content/uploads/2017/APAARI%20Strategic%20Plan%202017-2022.pdf

[23] https://www.spc.int/strategic-plan

[24] https://www.fair-access.net.au/fair-statement

[25] https://www.usc.edu.au/about/policies-and-procedures/research-data-management-procedures

[26] https://ppl.app.uq.edu.au/content/4.20.06-research-data-management

Compliance with Australian university data policies

Australian universities have research data policies in place, although the current level of compliance with these policies is not clear. Some recent studies have looked at university research data management and have concluded that research data management is now an institutional priority within Australian universities. A survey conducted by the University of Newcastle Library also shows there is a growing demand from researchers for research data management support.[27]

Over the last decade Australian universities have developed policies and procedures to address research data management. A few universities, like the Australian National University, the University of Queensland, and Queensland University of Technology, have had data management policies in place since at least 2008. However, only a few universities have developed additional policies that consider data sharing and access-related implications.

Marita Shelly and Margaret Jackson, in their article "Research data management compliance: is there a bigger role for university libraries?"[28] explore the support structures that exist to assist researchers in managing research data better and making it FAIR. Their study covers 13 Australian universities and finds that a variety of approaches are followed by universities in supporting researchers. While data storage options often exist, there appears to be a lack of support for researchers in regard to how to use such options. Furthermore, while there is infrastructure that can be used to save the data beyond the life of projects, efforts to make that data accessible for further research is limited.

These conclusions align with the CABI findings in relation to ACIAR's investments in general, and especially through the evidence gained from the 'deep dive' into three data-rich investments. In these investments the level of support for research data management and data sharing provided by the commissioned organisations to the researchers was found to vary significantly. In organisations where data was viewed as essential to the research outcome (i.e., an organisational priority), the researchers demonstrated a high level of data maturity and received organisational support. With investments that were unable to get supporting guidance from their organisation on how best to implement data management the researchers and their projects exhibited a lower level of data maturity, and this reduced the level of potential benefits (and return on investment) that was possible through the investment.

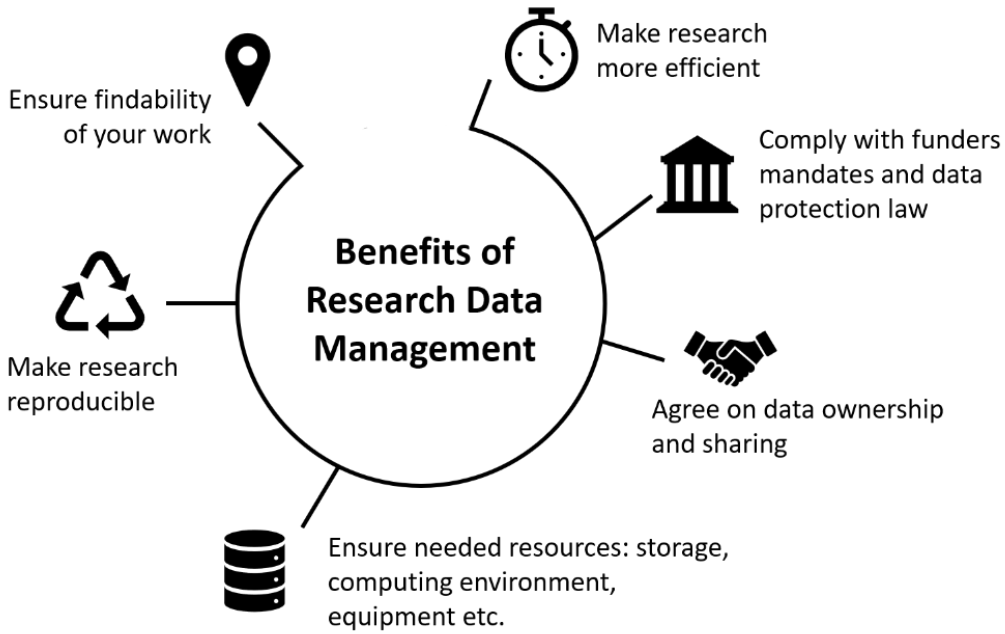


Figure 2: The Benefits of Research Data Management

https://research.csc.fi/documents/48467/0/Benefits_of_RDM_v4.png/a6898987-c297-e391-ebb9-894bd1df5f4b?t=1594195892267&imagePreview=1

[27] https://www.infodocket.com/2018/01/10/council-of-australian-university-librarians-publishes-caul-research-publications-repository-survey-2017/

[28] https://www.tandfonline.com/doi/full/10.1080/24750158.2018.1536690

Other Research Funding Organisations Data Policies

Many science and research-funding organisations now have strong data management and data sharing policies. Details of some of these are provided below, indicating the significant movement towards data sharing that is taking place. The organisations discussed are:

- the Bill and Melinda Gates Foundation
- Horizon Europe
- the US National Science Foundation (NSF)
- the World Bank
- the Canadian International Development Research Centre (IDRC)
- the United Nations Food and Agriculture Organisation (FAO)

Bill and Melinda Gates Foundation[29]

The Gates Foundation have a strong policy on Open Access to data that is created from the science projects they fund. Their Foundations policy states:

> *"We have adopted an Open Access policy that enables the unrestricted access and reuse of all peer-reviewed published research funded, in whole or in part, by the foundation, including any underlying data sets."*

Item 3 in the Foundation's policy states:

> *"3. Publications and Underlying Data Will Be Accessible and Open Immediately. All Funded Research including articles accepted for publication shall be available immediately at publication, without any embargo period. Each accepted article must be accompanied by a Data Availability Statement that describes where any primary data, associated metadata, original software, and any additional relevant materials necessary to understand, assess, and replicate the reported study findings in totality can be found."*

The Bill and Melinda Gates Foundation policy provides clear guidelines on what is expected in terms of data management activities and open data requirements for projects receiving its funding. Gates Foundation funding is dependent on meeting certain data management obligations, such as those outlined in the policy quoted above. While this may appear to be a 'big stick' approach to ensuring effective data management, the Open Access policy has been included in the project documentation because of the recognised benefits arising from effective data management, and in particular data sharing. There is a strong view within the Bill and Melinda Gates Foundation of the benefits of Open Data compliance within its projects.

Horizon Europe

Horizon Europe was established by the European Commission to provide research and innovation funding until 2027, with a budget of some €95.5 billion. Horizon Europe has published a document titled "Early knowledge and data sharing, and open collaboration".[30] This document states:

> *"The challenge is for Europe to embrace open science as the modus operandi for all researchers. Open science consists in the sharing of knowledge, data and tools as early as possible in the Research and Innovation (R&I) process, in open collaboration with all relevant knowledge actors, including*

[29] https://www.gatesfoundation.org/about/policies-and-resources/open-access-policy

[30] European Commission, Directorate-General for Research and Innovation, Horizon Europe (2021) "Open science: early knowledge and data sharing, and open collaboration". https://data.europa.eu/doi/10.2777/18252

academia, industry, public authorities, end users, citizens and society at large. Open science has the potential to increase the quality, efficiency and impact of R&I, lead to greater responsiveness to societal challenges, and increase trust of society in the science system.

***Legal provisions in the grant agreements will strengthen open access rights and obligations** for beneficiaries. Horizon Europe will require immediate open access to all scientific publications and responsible research data management so that data are **Findable, Accessible, Interoperable and Re-usable (FAIR)**. Data will be made 'as open as possible but will be allowed to stay as closed as necessary', safeguarding legitimate interests or constraints. However, access to research outputs shall be provided for third parties to be able to verify or validate publications.*

*The **criteria for evaluating research proposals** under Horizon Europe will take into account the quality and appropriateness of the open science practices in the submitted proposals. This will be assessed as part of the project's methodology, under the excellence award criterion. This will provide a strong incentive for applicants and beneficiaries to practice open science."*

These are strong statements about mandating open access and FAIR data principles within the European Commission's science funding processes. The Horizon Europe document acknowledges certain situations where there may legitimately be restricted access to data, which is why the Horizon Europe applies the FAIR principles, rather than simply mandating that data is Open. They expect FAIR compliance in all research that the European Commission funds.

Open science is embedded throughout Horizon Europe: in its work programs, in proposal evaluation, in its grant agreements, in project execution and follow-up, and in program evaluation.

Horizon Europe expects, among other things, the following results, and impacts from its open science policy.

- improved overall capacities within the European Research Area to conduct open science and implement it as a modus operandi of modern science
- better and more societal engagement in research and innovation
- increased trust in science by society
- improved reproducibility of research results and minimised duplication of effort

***Comment:** The benefits and outcomes arising from the implementation of FAIR data principles under Horizon Europe can be reasonably be expected from implementing those principles within ACIAR investments.*

USA NSF[31]

Since 2011, the USA's NSF has provided guidelines and mandates concerning the sharing of data created in the research activities it funds. NSF documentation regarding proposals for funding states that:

"All proposals must describe plans for data management and sharing of the products of research or assert the absence of the need for such plans. NSF's electronic systems will not permit submission of a proposal that is missing a Data Management Plan. The Data Management Plan will be reviewed as part of the intellectual merit or broader impacts of the proposal, or both, as appropriate."

The NSF Data Sharing Policy[32] states:

"Investigators are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants. Grantees are expected to encourage and facilitate such sharing".

[31] https://www.nsf.gov/bfa/dias/policy/papp/pappg22_1/faqs22_1.pdf

[32] https://www.nsf.gov/bfa/dias/policy/dmp.jsp

The World Bank

The World Bank is aware of, and acknowledges, the benefits of making data accessible and sharable. To this end it has created a Data Catalogue[33] that aims to: "make World Bank's development data easy to find, download, use, and share."

A goal of the Data Catalogue is:

> *"to maximize the value and investment in data by increasing the potential for the data to be shared and reused, to minimize transaction costs in finding relevant data and data methodologies, and to prevent duplication."*

Through the Data Catalogue, the World Bank is implementing FAIR principles as an aid to supporting research and knowledge creation. The World Bank clearly sees the benefits of data sharing and is investing in resources to facilitate this.

IDRC

The IDRC published its Open Data Statement of Principles[34] in November 2018. IDRC says that it embraces the principle of sharing research data and supports and proactively works with researchers to make their data open and accessible. IDRC's Open Data Principles are as follows:

- "Open access to research data is of fundamental importance to IDRC's mandate and to developing-country researchers;
- Data gathering and data management are to be conducted in a manner that maximizes the potential for data sharing;
- Promoting open data should not place an undue burden on IDRC's grantees or IDRC itself;
- Not all research data should be shared. IDRC acknowledges that the nature of the research data itself due to ethical, cultural, legal or intellectual property considerations may create legitimate grounds for data protection or limited access"

The IRDC provides guidance and assistance to researchers to develop data management plans within their research proposals.

FAO

The FAO operates CountryStat, a web-based source of food and agricultural statistical data. FAO views data as being highly critical in the development of policies and in supporting the achievement of Sustainable Development Goals (SDGs). In relation to "Data for Policy",[35] the FAO states:

> *"In today's data-driven world, policy-makers rely more than ever on accurate information and analysis to make informed decisions. Yet, missing and poor-quality data can hamper the design of successful policies – thus, jeopardizing the achievement of the SDGs".*

The FAO further adds that:

> *"By providing free access to reliable data, the Organisation helps to make informed policy decisions and supports civil society, the private sector and policy practitioners to undertake evidence-based research".*

[33] https://datacatalog.worldbank.org/getting-started

[34] https://idrc.ca/en/opendata-statement-principles

[35] https://www.fao.org/policy-support/policy-themes/data-for-policy/en/#c869797

Improving research outcomes through applying FAIR data principles

The organisations discussed in the preceding subsections recognise the significant benefits that come from the ability to find, access and reuse data collected through their funded science programs. While many of these benefits may be difficult to quantify, all of these organisations acknowledge the scientific, social and economic value that through the application of FAIR data principles can provide in the science domain and also the general community. They also recognise the benefits that can arise within and between their research programs.

Having data that can be easily shared results in greater quantities of data being available for research and analysis, which improves research outcomes. Additionally, effectively undertaking research into complex real-world problems requires the ability to bring together data from multiple sources and multiple disciplines. For example, when looking at field trials on crop productivity, if researchers only consider details regarding the genetic makeup of a particular crop species, results will likely be of marginal impact. If instead, data on soils, weather, cultivation methods and fertilizer usage are included, the research outputs can be significantly more reliable and accurate, leading to improved practices and policy recommendations. Being able to undertake research using such a range of variables requires data to be well-managed; and the ease of integrating the data is enhanced when it is findable, accessible, interoperable and reusable.

A Broad Perspective on Implementing FAIR

Relationships and agreements leading to data sharing

At a conceptual level, the management of data is a "relatively" simple task: all that is required is to determine the research project's specifications for covering several data variables, including:

- data requirements such as:
  - type of data
  - data formats
  - metadata structures
- accuracy and quality requirements (ensuring data is "fit for purpose")
- a glossary of terms

Once these have been determined and agreed by the relevant stakeholders and potential participants, there are a wide range of data management tools and technological capabilities that will support the collection and management of the required data, ranging from the very simple (Excel spreadsheets) through to highly sophisticated databases. By reaching an agreement on these data elements, data sharing can be relatively easily achieved, both within the project and also for external users of the data. An agreement on the conditions for data sharing is also required.

The key principle here is **"agreement"**. Agreement is a relationship concept. Without a strong and trusting relationship between those collecting and analysing the data, it can be difficult to secure agreement so as to create data that is easily sharable, and for all parties in the project to be willing to share their data.

Once a strong relationship has been built, and once agreement has been reached on the core principles, it is possible to apply an appropriate data management regime to the project – ideally one that is FAIR-compliant. Regardless of the "maturity level" of FAIR compliance within a project, any improvement in the project's data management will increase the ability of data to be more easily shared. Obviously, effective data management and full FAIR compliance is the desired outcome, but any movement towards increased "FAIRness", when starting from a low base, is desirable. ACIAR's focus on partnerships, and collaborative proposal development, puts the organisation in a good position to develop these types of agreements, which will provide the ability for ACIAR to be aware of what data is collected, and to support commissioned organisations and in-country stakeholders to share this data more easily.

While ACIAR itself will not reuse the data collected by an investment project, having an awareness of what has been collected can help inform other, similar types of research projects that may benefit from access to the data collected through previous investments.

Agreements supporting a Project

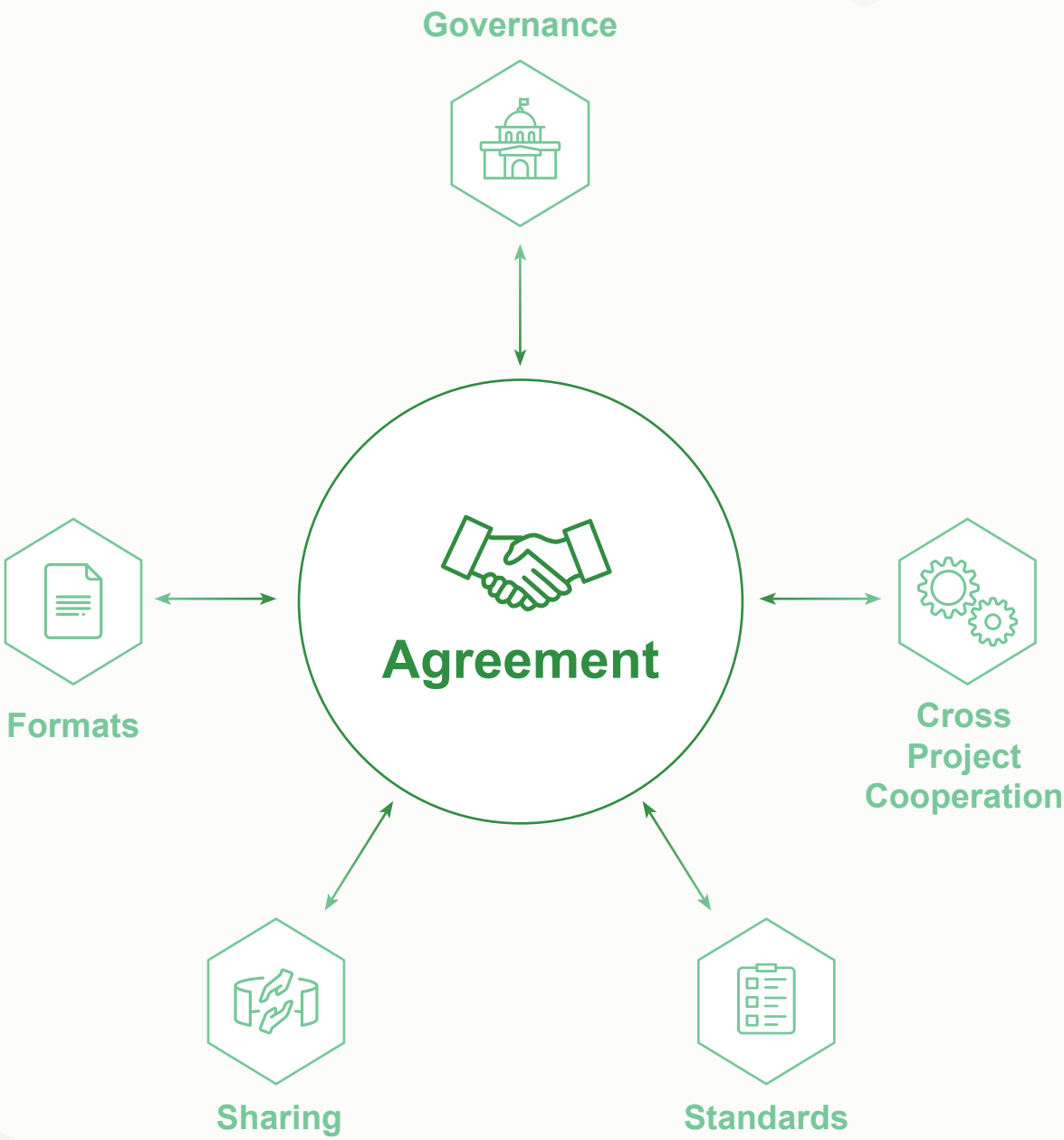


Figure 3: Visualization of agreements supporting a project

Data sharing benefits within the science community

The value of data can be more clearly recognised when significant benefits can be demonstrated to have resulted from data sharing. When such recognition occurs between research partners, this supports the building of relationships and enables data sharing agreements to be reached. Making data sharable, and then allowing for easy access to this data, can provide numerous benefits. For example, through data sharing scientists will have a larger “pool” of data to analyse than they could create on their own, improving their science. Additionally, in some of the countries that receive ACIAR research funding, it is often difficult for local researchers to publish their findings since it may not be easy or even possible for them to access project data. To enhance the process of decolonisation in ACIAR investment-recipient countries, the provision of easy access to project data provides the opportunity for in-country researchers to lead the publication process more readily. Equal access to data allows in-country researchers to undertake analysis and interpretation of research results, further strengthening in-country capabilities.

Examples of global data sharing programs

A brief description of a significant global program that has established effective data sharing arrangements (and operationalised this data sharing), resulting in massive social, economic and scientific benefits on a global scale, is provided below.

Prior to the formation of the World Meteorological Organisation (WMO) over 70 years ago, minimal amounts of meteorological data were shared between countries. Data was used for research purposes and national weather forecasting. However, national meteorological services began to recognise the benefits of sharing their data and accessing data from other countries and the formation of the WMO lead to the establishment of data sharing agreements and mechanisms between its member countries.

Today, the collection and exchange of a wide range of meteorological data has been operationalised and is shared in real-time globally. This data sharing has resulted in significant national benefits for WMO member countries. An acknowledged outcome is the huge financial benefits the USA government and other countries have created in their economies by making meteorological data open to the commercial sector and the public. The economic benefits from this in the USA alone run into billions of dollars annually and have resulted in the creation of whole new industry sectors, such as the multi-billion-dollar weather derivatives financial industry.

Annex 3 provides some examples of how some science programs have evolved as a result of the recognition of the value of data sharing. These programs have moved from conducting individual research activities to creating data sharing arrangements, which has resulted in significant local, national and global benefits. Annex 3 describes how individual Australian scientists working on their own to study the East Australian Current began sharing their data to provide a more comprehensive picture of the current. This sharing developed as their activities became integrated with many other national and regional oceanographic research programs to create the Global Ocean Observing System (GOOS). Annex 3 also describes the WMO example (below) in more detail, and sets out the significant social, economic and scientific benefits created in the USA through opening the data collected by the National Oceanic and Atmospheric Administration (NOAA).

Potential benefits from sharing agricultural research data

ACIAR's **mission** is to achieve more productive and sustainable agricultural systems, for the benefit of developing countries and of Australia, through entering into international agricultural research partnerships.

ACIAR's **vision or purpose** is to contribute to reducing poverty and improving the livelihoods of people in the Indo-Pacific region through more productive and sustainable agriculture emerging from collaborative international research. The potential arising from being able to easily find, access and reuse data collected by ACIAR investment projects is significant. Data sharing can assist in providing a mechanism for linking investment activities to establish more diverse, larger-scale and multidisciplinary research programs focusing on complex issues. Data sharing will be essential in these larger-scale and more multidisciplinary investments and will be a fundamental requirement when working with other research-funding organisations on large projects. The implementation of effective data management practices that lead to a data sharing culture based on FAIR data principles can thus help assist ACIAR in achieving its mission and vision.

The implementation of FAIR principles is potentially just a beginning on the digital journey for ACIAR and the many countries it works with. Increasing its emphasis on data management will support ACIAR in meeting its objectives of reducing poverty and improving livelihoods within the Indo-Pacific region. This can be realised through assisting recipient countries to move gradually into more digitally-rich methods of agricultural operation. Such an approach will significantly improve their agricultural capabilities, but will also ensure the gains made are maximised through increasing the efficiency of the agriculture and food supply chain process.

The potential of digital agriculture and research

Australian examples

Cooperative Research Centre (CRC) for Food Agility

The potential of digital (data) integration within the agriculture sector is well recognised in Australia, as highlighted by the establishment of the CRC on Food Agility in 2017. This research organisation has in place a very strong data management focus, including FAIR-based approaches, to underpin its research and to increase the benefits arising from that research. The Food Agility CRC has a program called Better Data for Better Decisions,[36] and has on staff a Digital Policy Manager and a Digital Content Manager to support this data focus. The Food Agility CRC operates a FAIR-based approach to data as an effective mechanism that aims to make data as shareable as possible, but also to provide protection to farmers and other stakeholders by considering who can access their data.

The CRC operates several highly focused research activities on the use of data to drive productivity. For example, one recently completed research program is a program called *Data-driven grain profitability through 'on-farm experimentation'*.[37]

[36] https://www.foodagility.com/research/better-data-for-better-decisions-constellation

[37] https://www.foodagility.com/posts/final-report-data-driven-grain-profitability-through-on-farm-experimentation

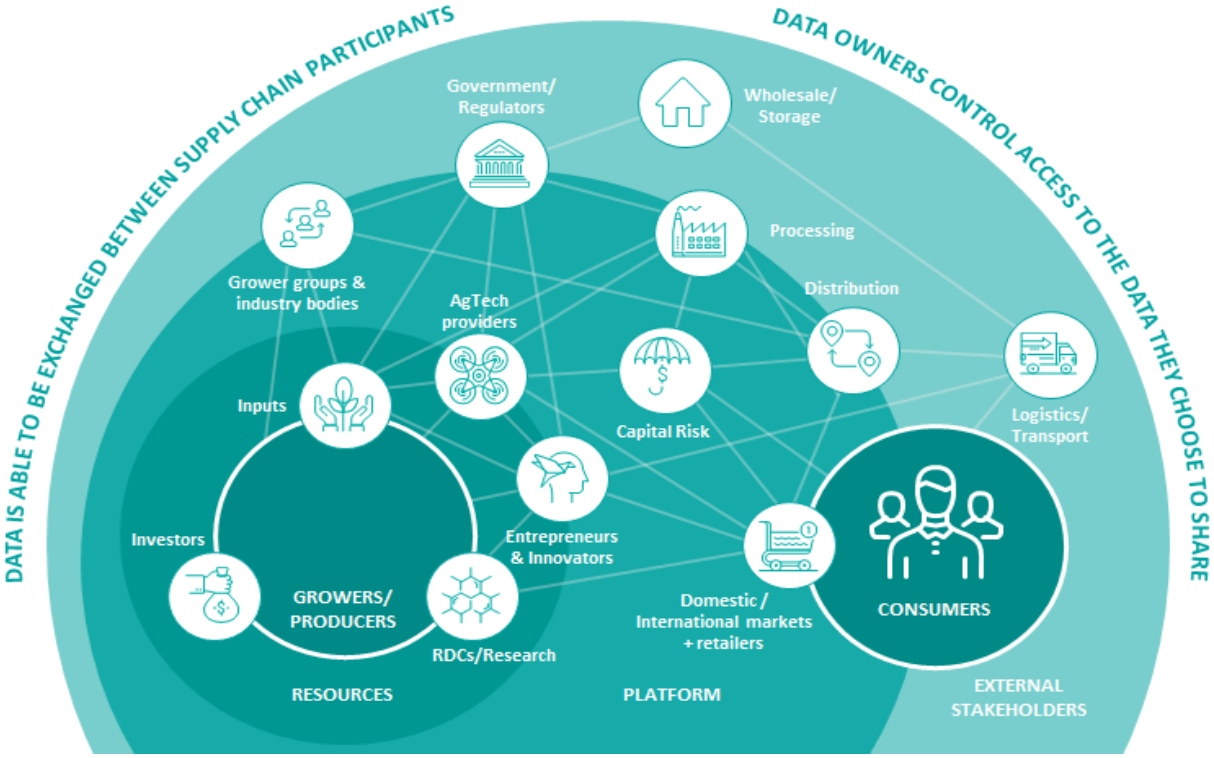


Figure 4: The flow of data between supply chain participants, data owners, and data consumers.

Australian AgriFood Data Exchange

A further Australian example of making data FAIR-compliant in order to support multisectoral needs is the newly developed concept of the Australian AgriFood Data Exchange.[38] This exchange aims to hold data coming from agricultural researchers, growers and producers, data about the supply chain, and data relating to consumers (refer to the diagram above). This offers enormous social, economic and commercial potential.

The vision for the Australian AgriFood Data Exchange[39] is to create:

> *"An interconnected data highway for Australia's AgriFood value chain.*
>
> *An absence of well-managed national agricultural and food data exchange for use by the various stakeholders within agriculture industry is limiting Australia's production potential, causing significant supply side inefficiencies and leaving Australia trailing behind competing nations. This problem is likely to get worse as the world becomes increasingly reliant on technology and access to information.*
>
> *Finding and preparing data for decision-making purposes consumes a significant amount of management capacity. Producers and researchers spend time manually reconciling data and hand coding integrations and transformations. Knowledge of existing data sets and tools is limited, making it challenging to generate meaningful insights."*

A component of the Australian AgriFood Data Exchange's Vision, specifically: "Knowledge of existing data sets and tools is limited, making it challenging to generate meaningful insights."

This statement is considered highly relevant to ACIAR investments. Improving data management practices and moving towards a FAIR data principles-compliant approach to ACIAR's investments will help generate meaningful insights for recipient countries.

[38] https://www.integritysystems.com.au/ozdata

[39] https://www.frdc.com.au/sites/default/files/inline-files/ozag-data-exchange-prospectus.pdf

Australian Farm Data Code

Edition 1 of the Australian Farm Data Code was adopted by the National Farmers' Federation, in consultation with industry, in February 2020.[40] The code is intended to inform the policies of service providers who manage data on behalf of farmers. It is also a yardstick by which farmers can evaluate the policies of those providers.

The National Farmers' Federation considers the benefits to farmers of the code to include:

- increased awareness and understanding of the ways in which providers are collecting, using and sharing their farm data
- providing a framework by which to compare providers and inform negotiations about data policies
- encouraging improvements to industry-wide data practices over time
- The benefits for service providers are considered to include:
- providing clear and agreed guidance on data policies
- providing a helpful framework to inform discussions with farmers about data

A key focus of the Farm Data Code is building trust with farmers in regard to ensuring that their data is only used for purposes that they agree with and approve. The code provides mechanisms to help build this trust with the farming community, as the "owners" of the data, and it is farmers themselves who determine who can access their data.

Extrata

The Western Australian (WA) Government has recently launched a similar initiative to JoinData: a data transfer pilot for farm data, which was announced in May 2022. This initiative is called Extrata and is considered by the Western Australian Government to be:

> *"A new 'gold standard' data sharing platform is about to be piloted in Western Australia, enabling producers to transfer information through highly secure and trusted channels – helping to take agribusinesses to the next level."*

A key element of Extrata is the ability it provides for farmers to take control of who accesses and uses data they generate. This aims to increase trust and encourage increased data sharing as a result.

International examples

Mekong River Commission

The application of FAIR is essential when operating data-related programs across national borders. In particular, data sharing is essential when supporting multi-jurisdictional management of critical resources. A regional example of the effective operation of a multi-country program is the Mekong River Commission. One of its four Key Result Areas is:

- "Better monitoring and communication of the basin conditions
  - Member Countries strengthen basin-wide monitoring, forecasting, impact assessment and dissemination of results for better decision-making."

To underpin this Key Result Area, an Agreement and Procedures Document (1995) provides considerable detail on data sharing, under the section on 'Procedures for Data and Information Exchange and Sharing'. The objectives for this data sharing are:[41]

[40] https://nff.org.au/wp-content/uploads/2020/02/Farm_Data_Code_Edition_1_WEB_FINAL.pdf

[41] https://www.mrcmekong.org/assets/Publications/MRC-1995-Agreement-n-procedures.pdf

- "Operationalise the data and information exchange among the four MRC Member Countries.
- Make available, upon request, basic data and information for public access as determined by the NMCs concerned.
- Promote understanding and cooperation among the MRC Member Countries in a constructive and mutually beneficial manner to ensure the sustainable development of the Mekong River Basin."

The Mekong River Commission could not exist or operate without these data sharing agreements in place, and the benefits resulting from these agreements were recognised early on in the establishment of the Commission. The benefits of data sharing are also acknowledged at the highest levels of the five member countries, with the signatories being either the member countries' minister of foreign affairs or their deputy prime minister.

JoinData

Another example of how making data easily shareable leads to a range of benefits is provided by JoinData,[42] which is a non-profit organisation that has been developed in the Netherlands. JoinData is an independent data platform focused on farmers. A significant investment focus into the security, privacy and farm-friendliness of the platform. It facilitates secured data sharing along the value chain and connects the data to allow better insights and smarter decision making. This is similar in concept to the **Australian AgriFood Data Exchange** described above.

Data sharing supporting ACIAR investment outcomes

The data generated through ACIAR's agricultural research investments forms an important starting point for the vision of creating an agricultural data highway, such as those highlighted above. The realisation of such a vison is potentially a long way in the future for ACIAR, given the current data literacy and data management capability of many of the partner scientists, organisations and countries. However, it does provide a potential view of the future for digital agriculture and could possibly be considered a future vision for ACIAR. Like many other data-driven initiatives across the commercial and government communities, this vision is based on the huge benefits that can be generated through improved data sharing, and FAIR compliance is seen as a critical element in supporting this data sharing.

Data sharing benefits beyond agriculture

Data sharing can provide a wide range of benefits well beyond the original research focus. For example, the measurement of a country's progress towards meeting the SDGs relies on accessible and reusable data from a wide range of sources, including agricultural research. Inputs into greenhouse gas emissions, and targeting resource allocations in efforts to mitigate climate change or address soil, land and water degradation, as well as tracking progress on social and environmentally focused SDGs, all require access to data from many different sources. If this data is not FAIR-compliant there is the possibility that it will not be used in supporting these national needs since it will be too difficult to locate and analyse in a cost-effective manner.

By adopting a FAIR-compliant approach to research, ACIAR investments can provide many benefits beyond simply improving research objectives. These include supporting higher-level needs, such as enhancing access to data that can be used in measuring national SDG-relevant activities. In addition, when properly considered, FAIR also enables the security of, and the reuse of, data in a controlled manner, ensuring that potential ethical, proprietary and commercial issues are effectively managed.

[42] https://join-data.nl/en/about-joindata/

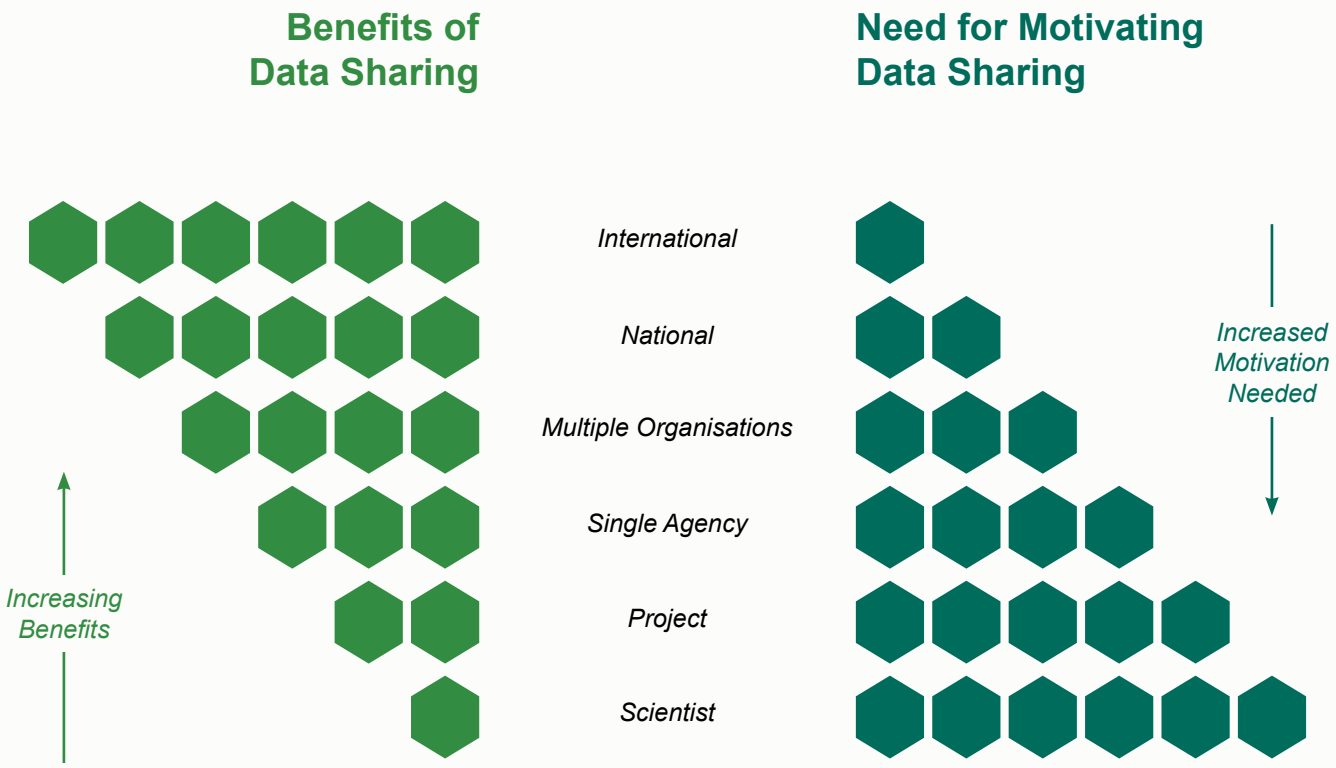


Figure 5: Data sharing – benefits versus motivation

The need for incentives to support data sharing

The benefits from data sharing increase as the scale of a project increases and/or as projects become more complex, with a larger number of parameters – such as under the global activities of the WMO as described above. At the other end of the spectrum is a scientist working on their own and on a small scale. At this individual level the benefits of data sharing are perceived to be, and often are, rather limited. The scientist is focused on their own research project and may see no immediate benefit in supporting national or global needs for data. Therefore, applying effective data management principles to this level of research can be counterproductive and may be seen as a burden for the individual scientist. However, one motivation for the individual scientist to share their data is the potential to benefit from a much greater "pool" of data they could potentially use to improve their own research outcomes.

"...the smaller the project, the greater the need for motivations"

While ACIAR investments are not at the single researcher level, it is nevertheless important to recognise that, in general terms, the smaller the project, the greater the need for motivations to encourage and support the scientist(s) to share their data. Motivations can include financial inducement, acknowledging the use of scientists' data, increased resource support, and training assistance. These approaches can be used to improve data management and facilitate data sharing at the small-scale end of the research project spectrum.

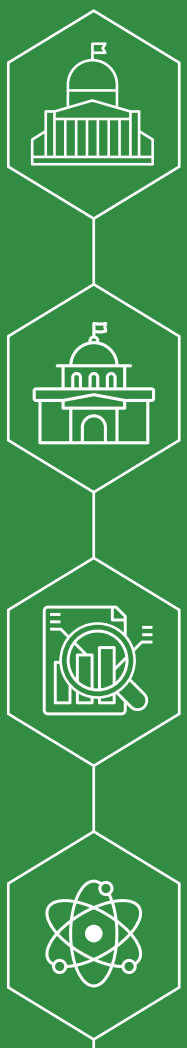
The diagram on the previous page shows the inverse relationship between the benefits of data sharing and the need for motivations to share data. This relationship is dependent on the scale and complexity of the research activity. When common goals exist at the international level (such as in the WMO example) we see very high levels of easily recognised benefits coming from data sharing. At this level, motivations are not generally required since the benefits are obvious. As the scale moves down to smaller projects the benefits are less obvious to the project participants. At this level there is a need to provide sufficient motivations to encourage and support data management – and therefore data sharing.

A move by ACIAR towards applying the FAIR data principles is seen as being beneficial to future ACIAR investments, especially in regard to jointly funded, multi-donor and multidisciplinary collaborative projects. When multiple variables, such as data about soil, water, crop types, topography, farming methods, etc., are collected, these will need to be easily accessed and structured in a way that enables their use within multivariant analysis tools. Investments in larger projects to address more complex issues will only be successful if the data collected by different researchers and across different disciplines can be effectively bought together to enable analysis. This will only occur if the data meets predefined formats and quality, which can be achieved by having effective data management (FAIR-oriented) practices in place across the project. If the researchers are required to spend time seeking out the data and reformatting it to get it into common structures before any analysis can be undertaken, then this will use up valuable time and will add additional costs, reducing the potential impact of the research outcomes.

ACIAR should therefore consider the benefits of implementing FAIR data principles in its investments and of leveraging this approach to help engage in larger, multi-donor collaborative projects.

A key driver for improving data sharing is understanding the benefits that will arise from each project. These benefits may differ for each potential participant, but they need to be recognised in order to establish agreement on a project's data management and sharing activities. Once this agreement has been reached, it then becomes possible to move towards a FAIR-based data management regime, or at least to introduce some of the FAIR principles if the level of participant data literacy and capability is low. It is considered that all project participants will receive some benefits if they are encouraged to move a little further down the path of data management towards FAIR compliance. Providing a small stretch target (and support for reaching the target) for increasing an investment's data management abilities will provide numerous benefits and will increase host agency, and potentially in-country, capacity.

Ensuring that individual scientists are sufficiently incentivised and supported to undertake effective data management will benefit everyone involved in a research project through enabling more effective data sharing, as well as benefiting others in follow-up projects and other external groups.



ACIAR's 10 Year Strategy

Data as a public asset

In recent years, research funders, in conjunction with academic institutions and policymakers, have been working towards increasing awareness around differentiating between a researcher's publication and the related data. The publication and knowledge derived from research data pertain to the authors, but the research data itself is often considered as a public asset, with the aim of maximising its value.[43] The case for data being considered a public asset is even stronger for ACIAR, as it is funded by public money, and because the Australian Government is strengthening its commitment to the sharing of publicly funded data from all sources. Details of these views are provided below.

[43] European Commission 2014; OECD 2015; Järvenpää and Markus 2018; Vassilakopoulou, Espen, and Aanestad 2016

Australian Government's view on data

The Australian Government's Data Strategy,[44] developed by the Department of the Prime Minister and Cabinet, was released in December 2021. The Minister for Employment, Workforce, Skills, Small and Family Business, the Hon Stuart Robert MP, has been quoted as saying:

> *"The Data Strategy is part of our commitment to deliver better services to all Australians, and it will power our national ambition to become a modern, data-driven society by 2030"*

The Australian Data Strategy states that:

> *"Data is a valuable national asset that, when leveraged effectively, can bring transformative benefits to its users and to individuals and the economy more broadly.*

The Australian Government proved this in its response to the initial COVID-19 outbreak, when it leveraged private and public data to respond to the health and economic effects of the virus. The private sector also has a long history of using data to benefit its clients through better and more tailored services and offerings.

Enshrining the effective, safe, ethical and secure use of data as an important foundational tool for businesses, individuals, the non-government and government sectors in an Australian Data Strategy will help to support the Government's vision to become a modern data-driven society by 2030.

The Australian Data Strategy signposts the Australian Government's data intent and efforts over the period to 2025. It focuses on three key themes:

1. ***maximising the value of data*** *– describes why data is important, its economic and social value, its use in responding to priority issues, and the benefit that can be gained through using and safely sharing data. Data can create new value when shared between different levels of government, and the private and non-government sectors.*

[44] https://ausdatastrategy.pmc.gov.au

2. ***trust and protection*** *– describes the settings that can be adopted in the private and public sectors to keep data safe and secure, and the frameworks available to protect Australians' data and ensure its ethical use through the entire data lifecycle.*

3. ***enabling data use*** *– sets out approaches and requirements to leverage the value of data, such as capabilities, legislation, management and integration of data, and engaging internationally.*

The document considers both public sector data, which is managed by the government, and data in the broader economy, where the Australian Government both enables data users and regulates its use and sharing to provide greater certainty in how people deal with their data."

In 2015 the Australian Government released its Public Data Policy Statement,[45] which forms part of the National Innovation and Science Agenda. The Policy Statement commits Commonwealth Government entities to:

- undertake specific actions designed to optimise the use and reuse of public data
- release non-sensitive data as open by default
- collaborate with the private and research sectors to extend the value of public data for the benefit of the Australian public

Several documents have been produced and released to support this Public Data Policy, including:

- the Australian Government Public Data Policy Statement (2015)
- Guidance on Data Sharing for Australian Government Entities (2016)
- a Review of Australian Government Data Activities (2018)

The Australian Government has continued to push for a greater level of transparency and accountability in how data is shared and released. Most recently, in April 2022, the Government passed the Data Availability and Transparency Act,[46] which provides a legislative framework for sharing government data. The Act sees three purposes for sharing public sector data:

- the delivery of government services
- to inform government policies and programs
- for research and development.

The requirement for good research data management within government-funded entities is increasing domestically in Australia. A similar focus is occurring in relation to Australia's participation in international activities, with the Department of Foreign Affairs and Trade partnering with several other large international funders of agricultural research for development in the "50 X 2030 Initiative".[47] This investment, and numerous others, is representative of the growing recognition that data generation is less of a concern (quantity) than the ability to categorise and refine what is useful (quality) out of large-scale data generation.

The Australian Research Council (ARC) advises the Australian Government on research matters, and administers the National Competitive Grants Program, which is a significant component of Australia's investment in research and development. The ARC recently (in 2020) made the requirement that researchers outline how they plan to manage research data arising from ARC-funded research a part of the agreement for funding under the National Competitive Grants Program.[48] Since 2002, the ARC has provided more than $13.6 billion to over 29,000 projects.

[45] https://www.pmc.gov.au/public-data/public-data-policy

[46] https://www.legislation.gov.au/Details/C2022A00011

[47] https://www.50x2030.org/

[48] https://www.arc.gov.au/policies-strategies/strategy/research-data-management

The Australian Research Data Commons view of data sharing

In June 2022, the Australian Research Data Commons (ARDC) released a document titled "Accelerating Research and Innovation Through Data".[49] This paper is a response to the Australian Government's National Research Infrastructure Roadmap (2023–2028). The ARDC document states:

> *"The roadmap reinforces the critical nature of data – for decision making, for commercialisation and for fundamental research. It supports our (ARDC) focus on ensuring Australian researchers are internationally competitive through having access to high quality data assets, platforms, infrastructure, policies, people and training. …*
>
> *A focus of the ARDC is the maximising the value of Australia's data assets – providing competitive advantage to Australian researchers by improving the discoverability, accessibility and usability of Australia's research data assets."*

Linking the Australian Government's views on Data to ACIAR's 10-Year Strategy

> *"ACIAR is a high performing research investor within the Australian Government, a skilled broker of collaborations across governments and between government and industry, and a crucial asset in Australia's global science diplomacy effort."*

As an Australian Government entity, ACIAR has an obligation to support the Government's philosophy, strategies and legislation regarding data sharing. A step forward in regard to supporting the government's data agenda can be made through the establishment of an ACIAR data strategy. Over time, such a strategy can build capability to ensure that, as much as possible, ACIAR, through its investment projects, can apply the FAIR data principles and achieve the Government's objectives in regard to making data sharable. Details of how this can be achieved are included under the recommendations in Section 8 of this report.

USA Government perspective on access to publicly funded research

As recently as August 2022, the US Government, through the White House Office of Science and Technology, released updated policy guidance regarding taxpayer-supported research. This update requires the results of such research to be made immediately available to the public at no cost.

> *"When research is widely available to other researchers and the public, it can save lives, provide policymakers with the tools to make critical decisions, and drive more equitable outcomes across every sector of society," said Dr. Alondra Nelson, head of OSTP. "The American people fund tens of billions of dollars of cutting-edge research annually. There should be no delay or barrier between the American public and the returns on their investments in research."*[50]

[51] https://ardc.edu.au/about_us/case-studies/accelerating-research-and-innovation-through-data/

[50] https://www.whitehouse.gov/ostp/news-updates/2022/08/25/ostp-issues-guidance-to-make-federally-funded-research-freely-available-without-delay/

The science community’s views on sharing data

Over the last several decades the science community has developed numerous data sharing arrangements, across many different disciplines. Some examples of cases where scientists have transitioned from engaging in individual research activities to leveraging the benefits of sharing their data are provided in Annex 3. An example showing the very significant economic and social benefits that result from this sharing is also provided in Annex 3.

Levels of data sharing vary considerably across each scientific discipline, due to numerous factors. Some of these variabilities are well described in the paper “Data sharing by scientists: Practices and perceptions” (2011).[51] This work, while now over 10 years old, still provides valuable insights into the issues around, and barriers to, sharing data, as perceived by a range of scientists. In CABI’s view, many of the concepts described in the paper are still current today and are relevant to ACIAR investments. The paper states:

> *“Scientific research in the 21st century is more data-intensive and collaborative than in the past.”*

It goes on to argue that:

> *“Barriers to effective data sharing and preservation are deeply rooted in the practices and culture of the research process as well as the researchers themselves.”*

The paper continues by suggesting that:

> *“Data are the infrastructure of science. Sound data are critical as they form the basis for good scientific decisions, wise management and use of resources, and informed decision-making. Moreover, science is becoming data-intensive and collaborative.”*

The paper points to various benefits that arise from sharing scientific data, including the following:

- “re-analysis of data helps verify results data, which is a key part of the scientific process;
- different interpretations or approaches to existing data contribute to scientific progress – especially in an interdisciplinary setting;
- well-managed, long-term preservation helps retain data integrity;
- when data is available, (re-)collection of data is minimized; thus, use of resources is optimized;
- data availability provides safeguards against misconduct related to data fabrication and falsification;
- replication studies serve as training tools for new generations of researchers.”

[51] **Data Sharing by Scientists: Practices and Perceptions** Carol Tenopir, Suzie Allard, Kimberly Douglass, Arsev Umur Aydinoglu, Lei Wu, Eleanor Read, Maribeth Manoff, Mike Frame. Published: June 29, 2011 https://doi.org/10.1371/journal.pone.0021101

Supporting ACIAR's 10-Year Vision through sharing data

"ACIAR is a trusted science partner in the Indo-Pacific region"

"ACIAR is building a knowledge base for tackling the biggest and most complex challenge of our time: how to achieve food and nutrition security, poverty reduction and better human health in ways that also improve biosecurity, water and energy security, while reducing net greenhouse emissions and adapting to more difficult climates."

ACIAR seeks to build a knowledge base that can underpin the achievement of six high-level objectives:

- ensuring food security and reducing poverty
- ensuring better management of natural resources and a more effective responses to climate change
- improving human nutrition and health
- empowering women and girls
- ensuring inclusive agrifood and forestry market chains
- building science capacity in our region

To achieve this outcome, ACIAR will need to explore what constitutes knowledge and what is required to build a knowledge base. If the final outputs and publications of ACIAR investments do not also include the underlying data, this may not be viewed as creating sufficient knowledge, with due rigour and replicability. The ambition of establishing an effective ACIAR knowledge environment is totally reliant on having access to quality information, which is in turn generated by quality data. The creation of high-quality data can only be achieved through effective data management and by supporting data sharing capabilities.

ACIAR should embrace the move being made by both the science community and the Australian Government towards open or FAIR data. It can contribute effectively to developing knowledge in partnership with the numerous science and science funding organisations that have developed data management policies and that, in some cases, mandate the management and sharing of research data. The objective of moving in this direction would be to realise greater benefits to science, industry and the community.

The global application and benefits of introducing these data management policies are described by the Open Data Policy Lab, in its online repository, which reviews almost 50 examples of legislation, directives and other policy documents from around the world concerning Open Data.[52]

[52] https://opendatapolicylab.org

Interview findings

As part of the study, CABI undertook many interviews and discussions across ACIAR and with a number of investment project personnel, including in-country project stakeholders and participants. The identities of the stakeholders who were interviewed are indicated in the detailed summaries of the interviews provided in Annexes 5, 6, 7 and 8. These interviews provided insights into the awareness and knowledge of, and issues around, the application of the FAIR data principles within ACIAR and its investments. The discussions also provided an understanding of where underlying data management activities fit into research priorities of, and the data management activities undertaken by, these projects and within ACIAR.

The interviews captured the understanding that exists of the FAIR concept, of data management in general, and of the potential benefits arising from making improvements to data management within ACIAR investment projects. Additionally, the interviewees identified barriers to implementing FAIR concepts and also provided views on what they see as being required for projects to improve their data management practices.

What this investigation made very clear is the broad spectrum of views that exist on data management, with a diverse range of understandings and thoughts expressed by stakeholders, from ACIAR executives through to project stakeholders and in-country researchers. While at this stage there is no specific ACIAR data policy, there is ample evidence from projects and participants that effective data management processes are well understood and are being implemented. For example, the ACIAR investment "Climate-smart landscapes for promoting sustainability of Pacific Island agricultural systems" clearly recognises the benefits of data sharing, and in fact data sharing is a key premise of this project. The project's aim of developing a collaborative mapping approach, leading to the creation of the Geospatial Platform, was highly dependent on establishing data sharing and on effective data management. The success and value achieved by this project to date underlines the benefits that can come from sharing data.

The views captured by CABI during the interviews suggest that each investment project has its own thoughts and priorities regarding data management, and the strength or otherwise of any data management activity undertaken within that project is usually the result of the individuals involved, rather than a top-down ACIAR "corporate" data philosophy or policy directive. This situation is common in research-driven projects across many disciplines where data management is frequently seen as not being of importance and/or can be viewed by some as an impediment to achieving the research outcome. However, researchers working on larger, multi-agency projects generally do appreciate the value of sharing data and therefore implement approaches to manage data more effectively to meet this data sharing need.

“Each project follows their own data management process. Difficult to access data. Lot of inconsistency.”

Key interview insights

- While data is recognised as a key and generally critical component within ACIAR investment projects, there is ambiguity around defining what data is, how it is created, what is required to effectively manage data, and who owns the data and data-related intellectual property.
  - This impacts the desire to share data since there is confusion over ownership, confidentiality and other related issues.
- The respondents were aware of the value of data management and saw it as being especially important in large or complex projects with multiple participants, across disciplines and/or regions, where data sharing is required for the success of the project.
- There was limited awareness of the FAIR data principles. However, respondents did often refer to a range of data management concepts relating to the terms findable, accessible, interoperable and reusable. While not necessarily being aware of the FAIR principles per se, the interviewees certainly understood the need to be able to support the concepts that constitute FAIR.
- Respondents were generally not familiar with existing national and international standards for data formats. However, the lack of application of standards to data was seen by many as a barrier to exchanging and reusing data collected within and across various investment projects.
- Privacy, confidentiality and security issues around data were highlighted by in-country participants and those deeply engaged in projects.
  - A view expressed by some was that “data is power” and so it should not be shared.
- There was a broad consensus by regional and country-level managers regarding the need to have high-quality data in each project. How this can be achieved was not clearly articulated and while mention was made of data standards, awareness of what these are and how they might be applied was also limited.

Summary of findings, with commentary

A high-level overview of the analysis of the interview responses is provided below. This has been broken down into the following data-related categories (a more detailed summary of the interviews according to the different roles of participants in ACIAR and ACIAR projects is provided in Annexes 5, 6, 7 and 8).

The interview categories consisted of:

Awareness

- the value of data and its management
- awareness of, and the application of, FAIR

Data challenges

- data Literacy
- data Formats
- privacy, confidentiality and security
- data quality
- standards, specifications and protocols

Within these categories, the views of those interviewed have been captured, and these are summarized below. We also provide comments from the CABI interview teams on these summaries, to provide additional context and interpretation.

Awareness	
The value of data and its management	The interview respondents all appreciated that data is a key – and generally critical – component of ACIAR investment projects. Data is seen as the primary driver for most project outputs and the analysis of this data is used to generate an understanding of the issues and ways forward to drive policy decisions and improve agricultural capability. There seems to be a general and broad understanding and awareness of how data is essential in all ACIAR investment projects, and that data management is required to ensure the effectiveness of these projects.
Comment: *There was a level of confusion about what constitutes data and what constitutes information that is created through analysing data. This confusion existed within ACIAR above the project level, although there was still some ambiguity at the project level. While there may be some grey areas concerning what is data and what is information, many of those who were interviewed frequently conflated the two. This is an important distinction since the management of data and the management of information is generally a different activity, requiring different tools and skills. Refer to the section on "Data literacy" below for more detail on "what is data".*	
Awareness of, and the application of, FAIR	A high percentage of those interviewed were not aware of FAIR concepts. A small number did have some understanding but in general the concept of FAIR was not well known. However, the general principles of good data management were referenced by various people, without them being aware that these concepts are included in the FAIR principles. Without specifically being aware of FAIR and the FAIR terms, some comments were aligned to the importance of project data being findable, accessible, interoperable and reusable.

Data challenges	
Data literacy	While the significance and value of data is generally well understood across the wide range of ACIAR and the project personnel, the general level of data literacy was found to be low. An exception to this was often among project leaders and Australia or New Zealand-based researchers, due to their awareness of their own organisations data policies. In a number of cases, it was found that the following questions could not be easily answered • What actually is data? • How is it created? • What is effective data management? • What is data governance? • Who owns the data and its intellectual property? These concepts were not always clearly understood by ACIAR management and by Regional and Country Managers, although there was generally a greater awareness of them at the project level.
Comment: *The issues raised around what is data, and around data ownership and data intellectual property, were generally caused by a confusion or a lack of clarity around what constitutes data. Data is considered to be a fact, as in the following definition from the Oxford English Dictionary:* *"Data: A thing given or granted; something known or assumed as fact, and made the basis of reasoning or calculation; an assumption or premise from which inferences are drawn."*[53] *It would be useful for ACIAR to provide a definition of what constitutes data within its research projects, to create greater clarity in this area.*	
Data formats	A number of those who were interviewed referred to the need to have effective – and possibly standardised – data formats. The lack of such standards was seen by many as a barrier to exchanging and reusing data collected within and across various investment projects. There was limited awareness of national or international standards for particular data types. If such standards were applied, at least in some projects, this would assist in data exchange (interoperability) and would facilitate reuse.
Comment: *The subject of data formats can be a very complex topic and a desire to achieve something like a "gold standard" can be off-putting to those seeking to achieve meaningful progress – there are many approaches to supporting the creation of datasets that are more easily exchanged, short of moving to a point of applying a fully agreed international standard. Achieving some form of agreement at the commencement of complex projects regarding basic data management principles, including data formats and data sharing, would be beneficial in improving project outcomes. Additionally, the creation of effective metadata, where data formats are well described, can alleviate the need to strictly follow international data format standards.*	
Privacy, confidentiality and security	Privacy, confidentiality and security were issues that were raised by a number of those who were interviewed, and particularly by people who were more closely involved in projects such as the RPMs and in-country networks that are more concerned with these issues. It is likely that the countries in which the ACIAR investment projects are carried out demonstrate a range of legislative environments as relates to privacy and (probably) security. Thus, the legislative environment is likely to be different for each country. Local project participants and stakeholders are closer to these "constraints" and are more aware of any privacy, security or confidentiality requirements. Additionally, some stakeholders hold the view that "**data is power**", and believe that sharing the data reduces that power.

Continued

Privacy, confidentiality and security

Comment: *Confidentiality and security are legitimate issues, although unfortunately they are often used as a justification for NOT sharing data when in reality there are no real issues. The legitimate restriction of access to data for confidentiality and security reasons obviously need to be adhered to; however, it is important that, unless datasets are highly confidential, at the very least the existence of these datasets is known, by listing them in metadata inventories. There are unlikely to be many ACIAR investment projects with datasets whose actual existence needs to be kept secret. Documenting the existence of datasets (even if the dataset cannot be shared) should and could be achieved through descriptions of these datasets being available in appropriate metadata or inventory systems. Access to such datasets might then be gated by strong password control, with access being sought via named individuals, etc. It is important to recognise that the FAIR principles provide a framework for controlling access to, and safeguarding, data if the originator is unwilling to share it openly.*

Data quality	There was a level of both interest in, and awareness of, data quality among many of those who were interviewed. There was a broad consensus among regional and country-level managers of the need to have high-quality data in each project. Some mention was made of data standards but a detailed awareness of what these are, and how they might be applied, was mostly absent. The responses around data quality are not considered unusual within an investment and research project environment. It was obvious to all those who were interviewed that projects should collect and use the highest quality data possible to ensure that research outcomes are robust and reflect the situation being examined.

Comment: *Data quality is a difficult and complex topic and is not easy to articulate and define. The reason for this relates to the fact that "quality" is defined in part by the purpose for which the data is being applied. Within a research project, data is normally collected for a very specific research purpose, and the proposed purpose and intended analysis activities assist in determining the "quality" of the data required. The level of data "quality" required is therefore usually determined in the initial development of the research program. However, what is actually happening in this case is not defining the quality of data, but rather defining what data is "fit for purpose". What this means is that concepts such as the resolution of the data, the level of detail, and a range of other data variables, must be considered in relation to achieving the research objectives (for example: is 1.0-degree temperature resolution required or 0.1 degree?) The capability of the instruments and the techniques used in collecting the data influence the level of detail of the data, at the same time as the project's requirements also impact the data "quality". Once the specifications of the data are determined (fit for purpose, i.e., "quality") the project can proceed. The project data "specification" provides data for the project that is fit for purpose.*

Once the data has been collected it may have value to other projects. This value will be dependent on its fitness for purpose for other projects. Sticking with the temperature example above, if the initial project collects temperature data to an accuracy of 1.0 degree, and another affiliated or subsequent project requires data accurate to 0.1 degree, the 1.0-degree dataset may be of limited use – it is not totally "fit for purpose" for that second project.

What is "quality" data? – This can be defined as data that is as accurate and complete as possible as regards achieving the project objectives (fit for purpose). This is likely to mean that each data record is complete, with no, or minimal, gaps in any of the data fields. For example, a "quality" dataset is one where a spreadsheet of the data has all fields (in both columns and rows) filled in, and where these values have been validated in some manner; and where obvious errors (instrument or human-created) are corrected, edited or deleted. Data meeting these criteria could be considered "quality" data for a specific project. As discussed above, though, this data may or may not be "fit for purpose" or considered a quality dataset for another project.

Like the numerical data in the example described above, social data has similar issues, and determining "data quality" in regard to fitness for purpose should be done in the design phase of all projects. The approach to determining if data is fit for purpose for other projects (reuse) can only be determined if the data is appropriately documented, with adequate information about the context provided. For example, a given survey response may simply be recorded as "YES" in a dataset. Without the context and the ability to cross-reference the actual question, who was interviewed, where they were interviewed, and perhaps the purpose of the interview, "YES" will be meaningless for other potential users.

Continued

Standards, specifications and protocols	The respondents spoke in general terms about standards, specifications and data protocols. These views on standards were generally very broad. The desire to implement data standards within projects was not strong. However, it was recognised that relevant standards could improve a project's ability to share data – and therefore improve the research outcome. FAIR compliance means ensuring data are findable, accessible, interoperable and reusable. All respondents, while not necessarily being aware of FAIR, certainly understood the need to be able to support the concepts that underpin FAIR, and that standards, etc., provide a mechanism to do so.

Comment: *Standards can apply to a range of elements in relation to data. The most common use relates to data formats: there are many internationally agreed data format standards for a very wide range of data types. The application of these standards when creating datasets facilitates the exchange, interoperability and reuse of this data. Knowing that a dataset is standards-compliant means a researcher can – relatively – easily get the data into their own systems. Other standards and protocols relate to metadata, and also to data management governance processes. The greater the application of relevant standards the easier it is to share the data, and the closer to meeting FAIR compliance a dataset will be.*

> "Better data infrastructre and data management processes are needed to access project data"

[53] https://www.oed.com/oed2/00057804;jsessionid=F1B19E9B19C4EA7FB6B0CE2A68B04E4B

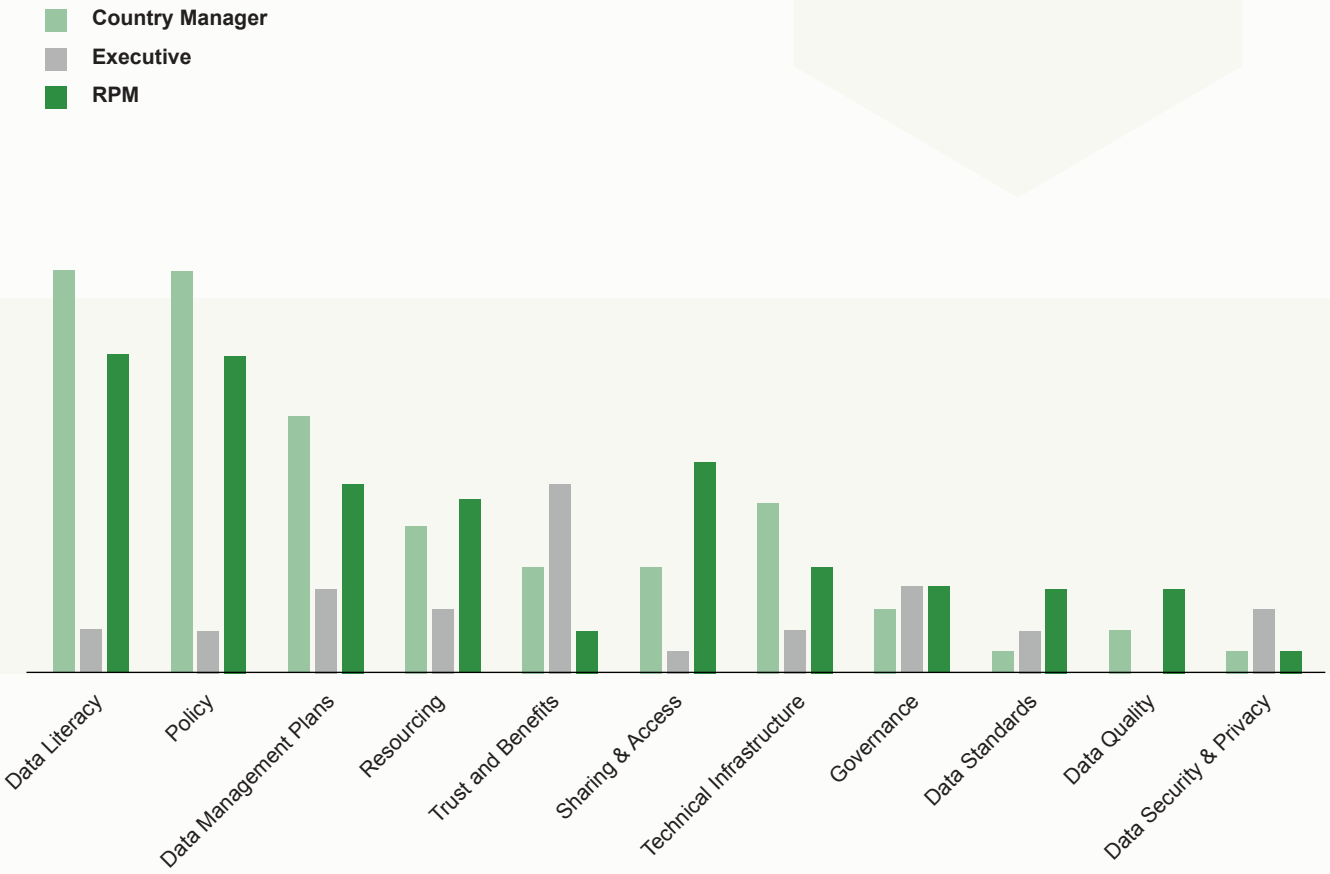


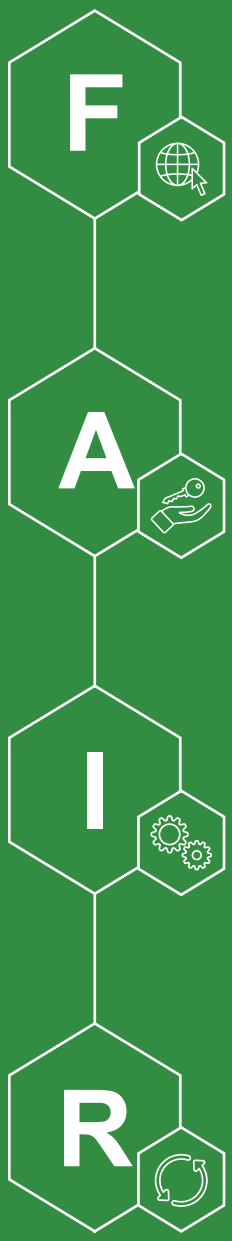
Figure 3: Interview responses

The figure above summarises the number of responses to the questions across the three groups interviewed:

- Executives
- RPMs
- Country Managers

The height of the bars shows the numbers within each of the three groups that responded in relation to the specific question. For example, “Data literacy” received a high number of responses from country managers and RPMs but few responses from ACIAR executives.

All groups of interviewees frequently referred to policy, data management plans and resourcing as constraints. With respect to other constraints, the responses from RPMs and Country Managers were very similar. For example, data literacy and sharing and access were frequently brought up as issues by RPMs and Country Managers, but less so by executives. On the other hand, executives viewed trust and benefits as a significant constraint, whereas RPMs and Country Managers did not highlight this to the same extent.



Perspectives on the status of the FAIR principles in ACIAR

The interviews and discussions with a wide range of stakeholders across ACIAR staff, together with researchers and other investment-related stakeholders, have resulted in a clearer picture of the existing capabilities, understanding, views, perceptions, barriers and needs as regards data management and data sharing within ACIAR and its projects.

Some of the responses provided during the interviews are shown below in the figures described as covering

- FAIR as seen through the lens of ACIAR staff
- observations of data collection and usage within investments in ACIAR
- FAIR as seen through the lens of observed current practices

> “Not sure where to look for data within ACIAR?”

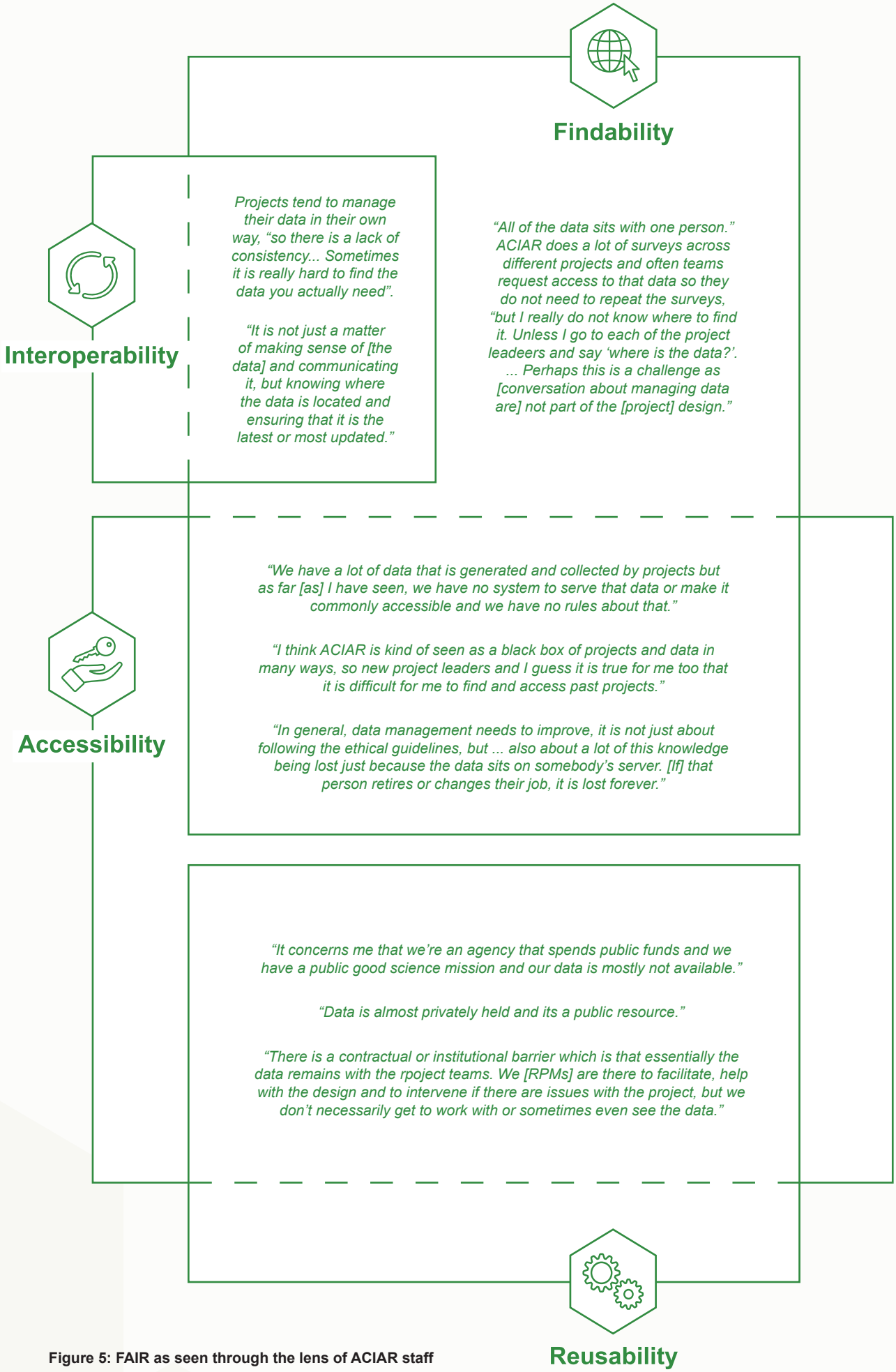


Figure 5: FAIR as seen through the lens of ACIAR staff

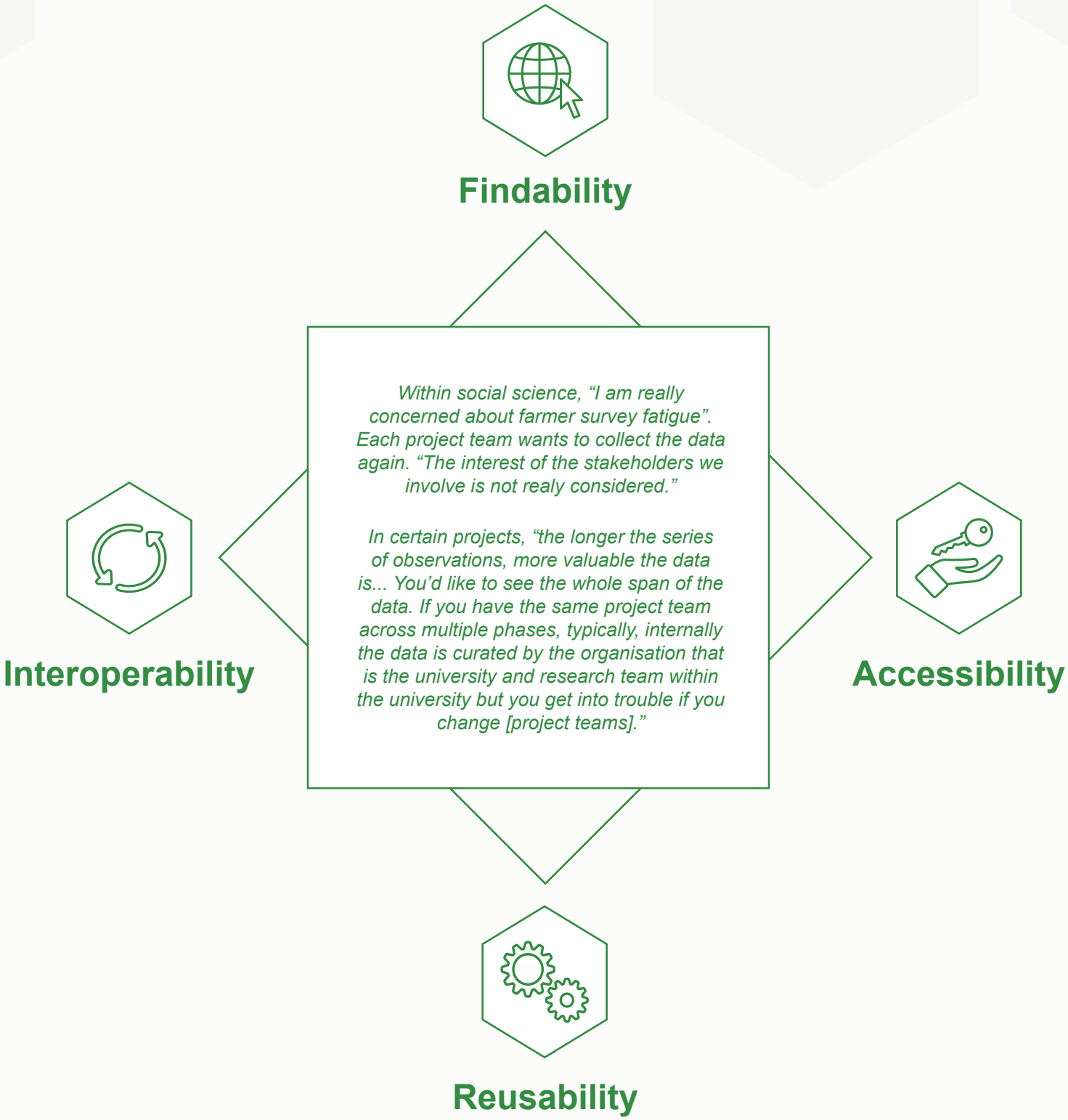


Figure 6: FAIR as seen through the lens of ACIAR staff

While all interviewees may not have been aware of, or had limited knowledge of, the concept of the FAIR data principles, concerns around the findability, accessibility, interoperability and reusability of data were expressed, and these are captured in the images given above.

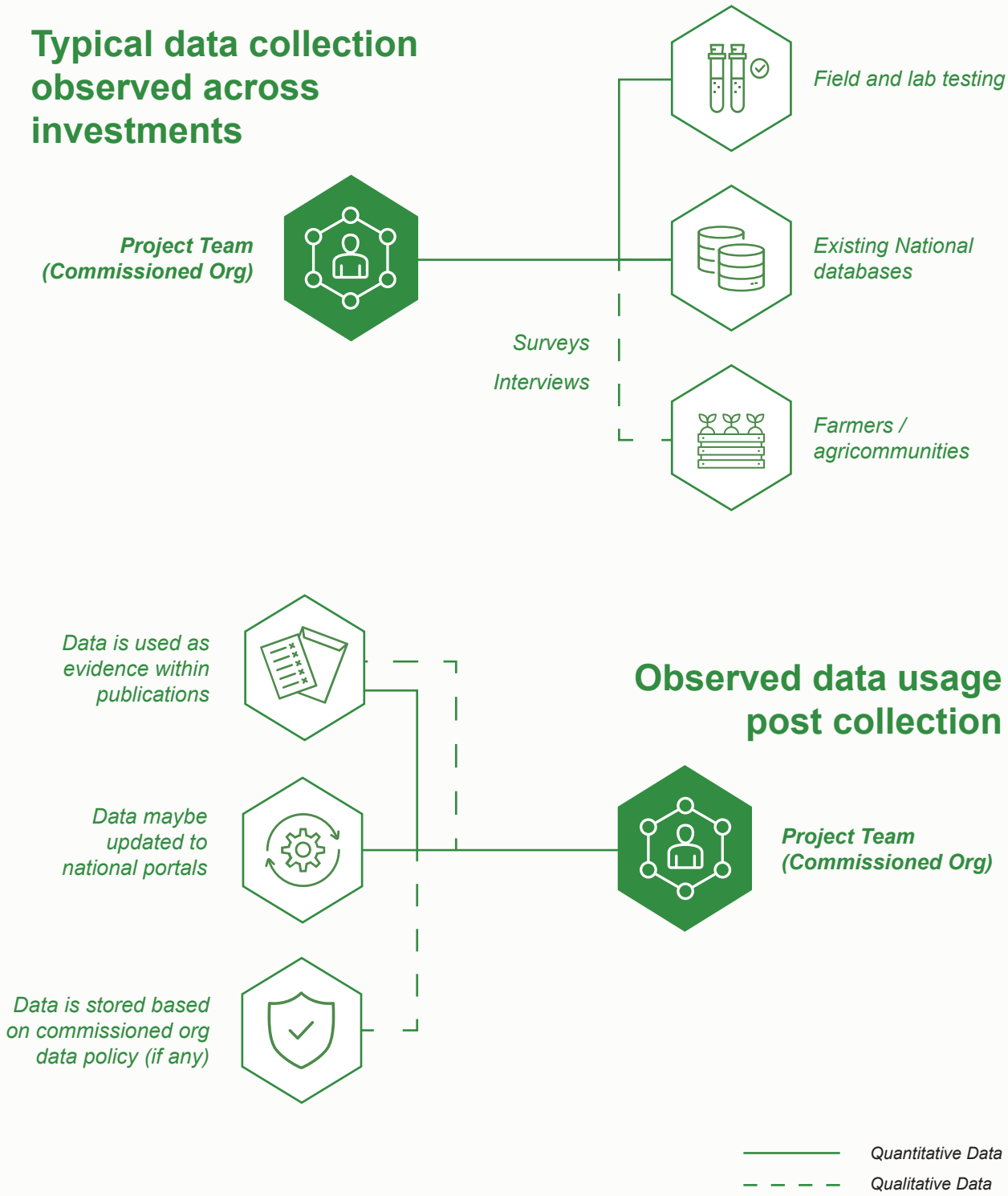
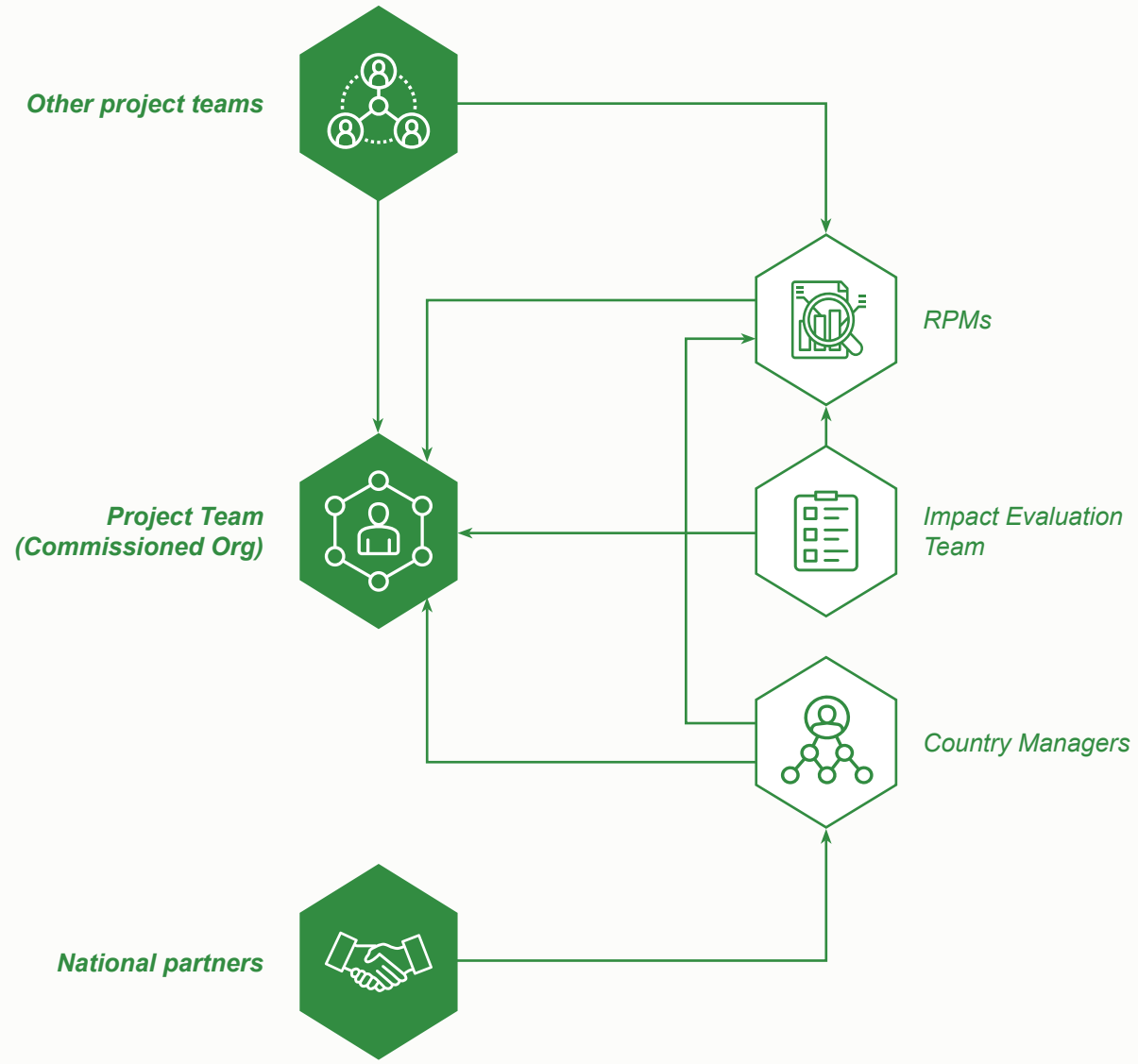


Figure 7: Observations of data collection and usage within investments in ACIAR

The results reported in Figure 5 show the broad observations made by interviewees regarding the way data is typically collected and used across ACIAR investments.

Observed status of Findability & Accessibility



External Partners

ACIAR staff

Request for data

The level of findability is highly dependant on the project team. Other stakeholders and internal teams of ACIAR are completely reliant on the project leader's response to understand whether the data exists.

The level of accessibility is highly dependant on relationships and the level of responsiveness of the project team. Other stakeholders and internal teams of ACIAR are completely reliant on the projcet leader/ team's response to get access to the data.

Figure 8: Findability and Accessibility through the lens of observed current practices

Observed status of Interoperability and Reusability

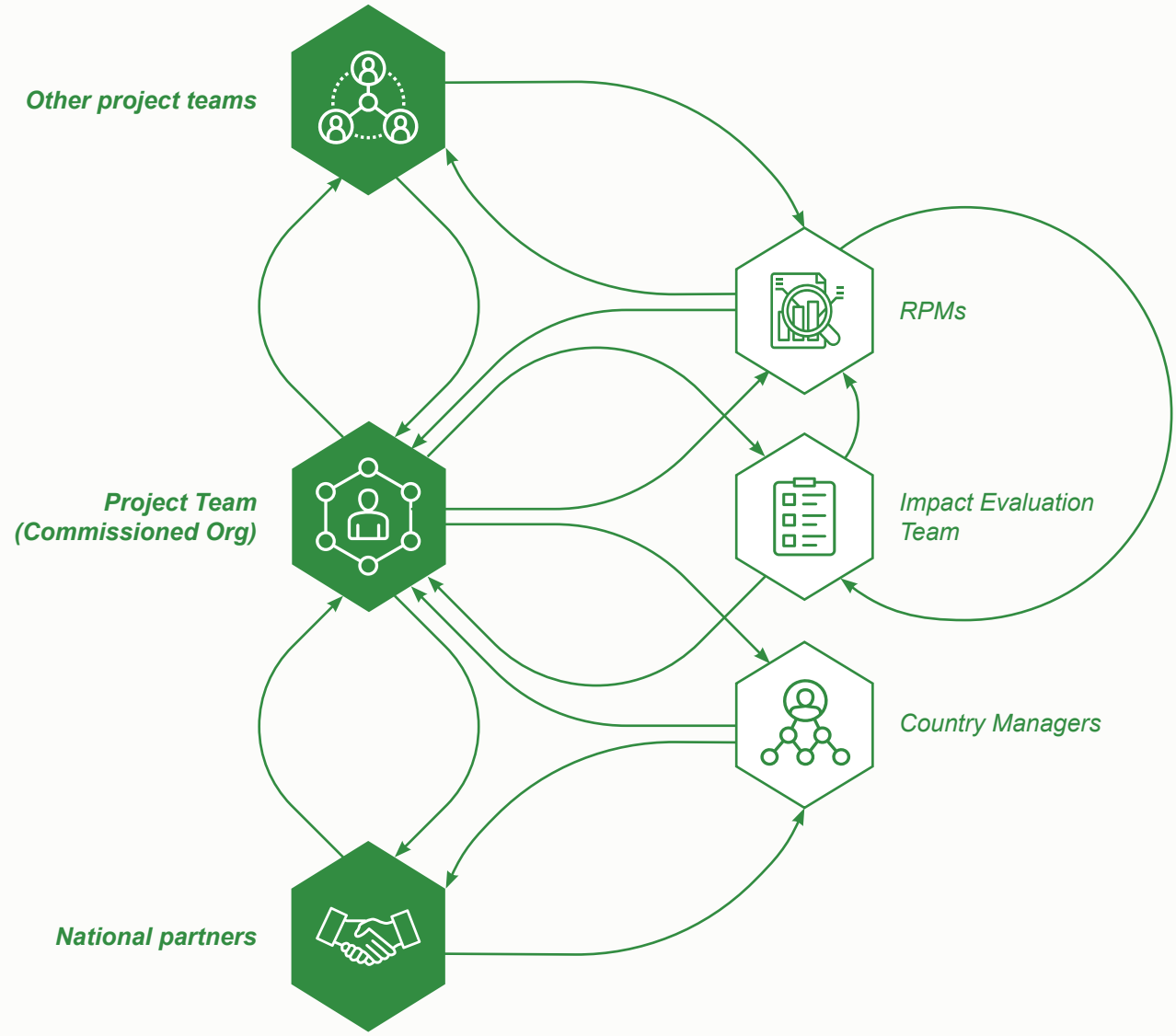


Figure 9: Interoperability and Resuability through the lens of observed current practices

The process flow diagrams on the proceeding two pages show how finding, accessing, interoperating and reusing data are people and relationship-dependent. Currently significant back and forth communication is required before data can be found, accessed, interoperated and reused.

Deep-dive interviews relating to three data-rich investments

Introduction

To gain more in-depth and focused evidence on the application of data management and the FAIR data principles across the ACIAR research portfolio, CABI undertook a deep dive into three data-intensive investments. This deep dive focused specifically on use cases where data and data sharing are an integral component of the investment. The aim of this "deep dive" was to better understand how these investments consciously or unconsciously follow good data management and sharing practices, and what barriers to data utility exist, and to assess the culture of data sharing within the investments.

ACIAR suggested eight possible deep-dive investments and CABI undertook an evaluation of these using specifically developed selection criteria; this resulted in the selection of three investments for further examination. The intention of the deep dives was not to review the entire investment in each case, but to focus solely on the specific data-intensive use case aspects.

The three investments reviewed in the deep-dive exercise were the following:

- Climate-smart landscapes for promoting the sustainability of Pacific Island agricultural systems – focusing on the ***Geospatial Platform*** *(ASEM/2016/101)*.
- Soil management in Pacific Islands: investigating nutrient cycling and the development of a soils portal – focusing on the ***Pacific Soils Portal*** *(SMCN/2016/111).*
- Land suitability assessment and site-specific soil management for Cambodian uplands – focusing on the ***Cambodian Soils Database*** *(SMCN/2016/237).*

Figure 10: Key recommendation drivers from deep-dive interview findings

Key deep-dive interview findings

For each investment, the data management, data sharing and data utility activities were discussed with a number of investment participants. These discussions provided considerable information, which provides the evidence on which many of this report's recommendations have been developed. The main concepts and understandings gained, and which underpin some of the recommendations, include the following:

- An identification of the value of **upfront investment** in key aspects of the project prior to project commencement. Important activities identified include the development of data governance and data sharing arrangements before commencement – this is found to significantly improve the project outcome.
- Working closely with key in-country stakeholders to **co-develop and test any technology** used in the investment. This was especially evident for data collection tools and ensures that the data collected fully meets local user needs.
- Identifying a **local "champion"** who can assist in supporting the in-country research activities, which leads to improved data usability and increases the effectiveness of the project outcomes.
- Embedding relevant elements of the research activities as **"business as usual" processes** within the in-country host organisations, which can ensure the longevity of the project's activities. The embedding of aspects of the project in the organisation's workflow will occur if the function is seen to add value – in other words, if it is successful. This process will increase the likelihood of the capability being sustained.
- **Reusing relevant data, tools and agreements** from previous investments to improve project outcomes. This reduces the time required to develop new data governance and data sharing agreements, and reusing relevant tools and other capabilities also reduces time and costs for new projects.
  - In the context of data governance and data sharing agreements, if an agreement already exists (from another investment) and is between the same organisations then adding some new data types to the existing agreement should be much simpler than trying to implement a separate new agreement.
  - Leverage capabilities, such as data repositories, databases or portals, that have been developed to manage similar types of data.
  - Leverage similar data collected by previous projects to increase the size of the dataset and to improve the analysis outcome.
- Assist investment participants who wish to improve their data management and data sharing activities but are unsure of what tools to use, and how to use them, through the provision of **technical guidance, training and resource support**.
- Ensure a level of **flexibility in the investment** to allow small changes to occur as relevant information is gained through the project – for example, the target users of the Geospatial Platform moved from initially being farmers to being Ministry of Agriculture and Food, Forests and Fisheries (MAFF) staff in Tonga, which resulted in the platform being highly successful. The original target (farmers) were supported through the connections of the extension officers.

How is FAIR perceived by ACIAR now, and how will it be perceived moving forward?

The benefits arising from data management, data sharing and the application of the FAIR data principles within ACIAR and its investment projects were acknowledged by the stakeholders who were interviewed. While the value arising from being able to interrogate data from previous projects seems to be commonly understood, there appears to be a disconnect between the motivation for a given project to share their data and the benefits that might accrue for other projects having access to that data. It was observed that projects often operate in silos, with little or no linking to other related projects, to leverage data or research outcomes from similar types of projects. Data sharing and data management activities are often not seen as adding value to an investment project that is already underway, and implementing data management protocols is seen by some as potentially being an inhibitor to the effective execution of investment projects. For example, where intellectual property issues are flagged, or when data specialists are not directly engaged and supporting a project, a suggestion to improve data management is seen as a cost and administrative burden to the researchers.

While the interviews found that there was a broad awareness that improved data management could be important, there was a limited demonstration of actual implementation of data management processes and data governance structures within some of the projects examined. In general, the Australian and New Zealand research leaders were significantly more aware of the value of data management, and generally had more data management capability in place, than the in-country participants.

To some degree, the level of perceptions and understanding of FAIR concepts can be linked to the levels of capability within the organisations working on projects, especially the in-country components of projects. It is important to note that interpretation of the degree of understanding of the value of data management took place within the context of the investment partners' and participants' technical abilities, and organisational facilities and tools (their data maturity level). Many of these in-country recipient organisations have very limited data management capability and are operating from a relatively low base of awareness. Additionally, while some partner organisations do have some data management capability, their budgets and priorities often force them to focus on direct (immediate) outcomes from their research. They currently see improved data management as only providing secondary benefit. The ability to reuse research data was not commonly seen as being beneficial in the context of the specific and urgent issues that any given research investment is aiming to address.

These views – which can become embedded in organisational culture – might best be described as symptomatic of many researchers' views that data management is a burden and not a benefit. These attitudes, as well as organisational data management capacity issues, will need to be addressed (through the provision of suitable training and resources) before effective FAIR-based capabilities can be implemented in many ACIAR investments. Consequently, moving forward to improve data management and support the FAIR principles will require careful management, education and resourcing support.

For this to occur it will be necessary for ACIAR to articulate the benefits of supporting FAIR, and to demonstrate where some projects have achieved success as a result of sharing data. An awareness of these benefits is required across all levels of ACIAR, and needs to be incorporated into formal and informal communications and relevant documentation coming from ACIAR in relation to its investment projects.

Influencing factors for recommendations

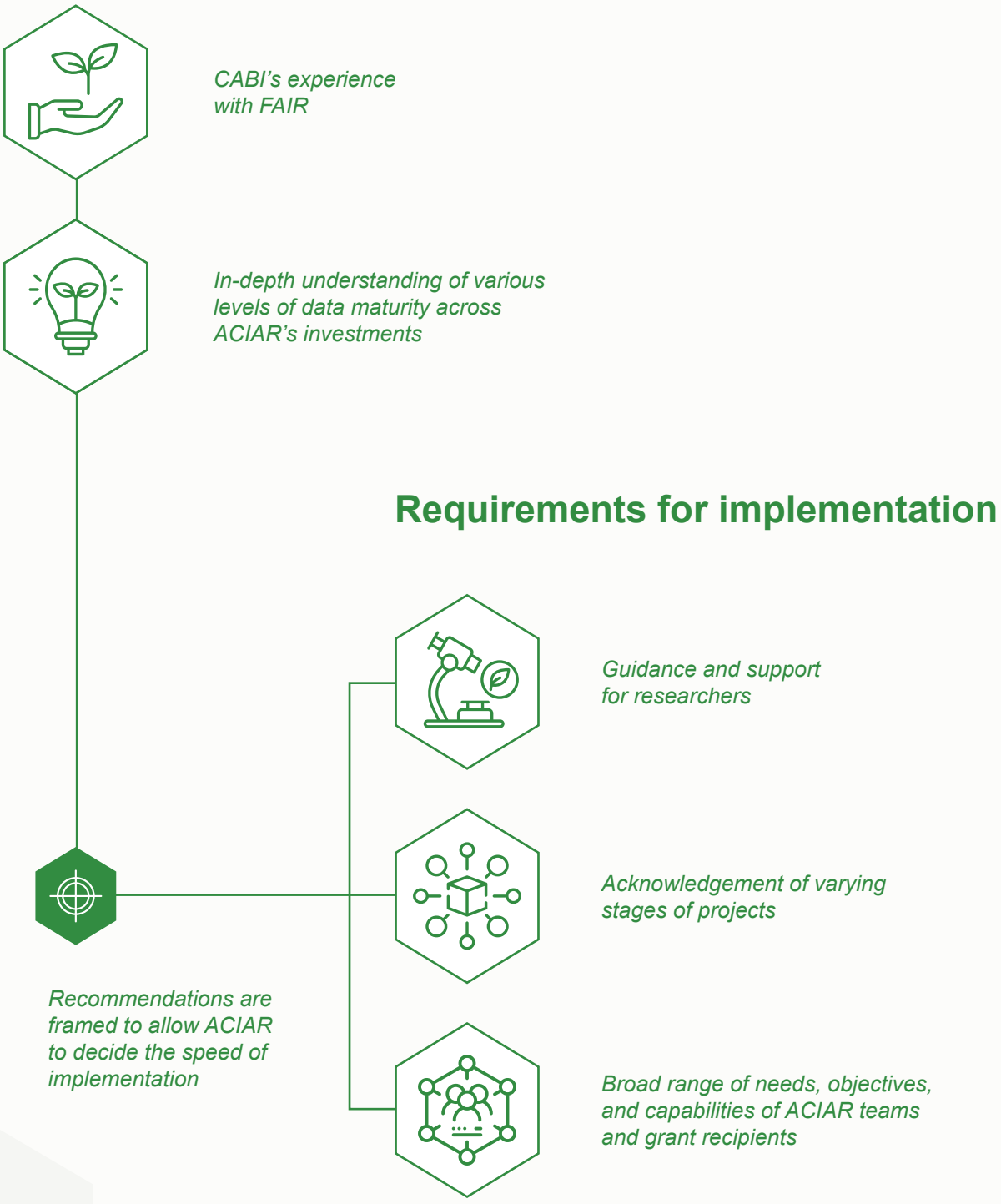
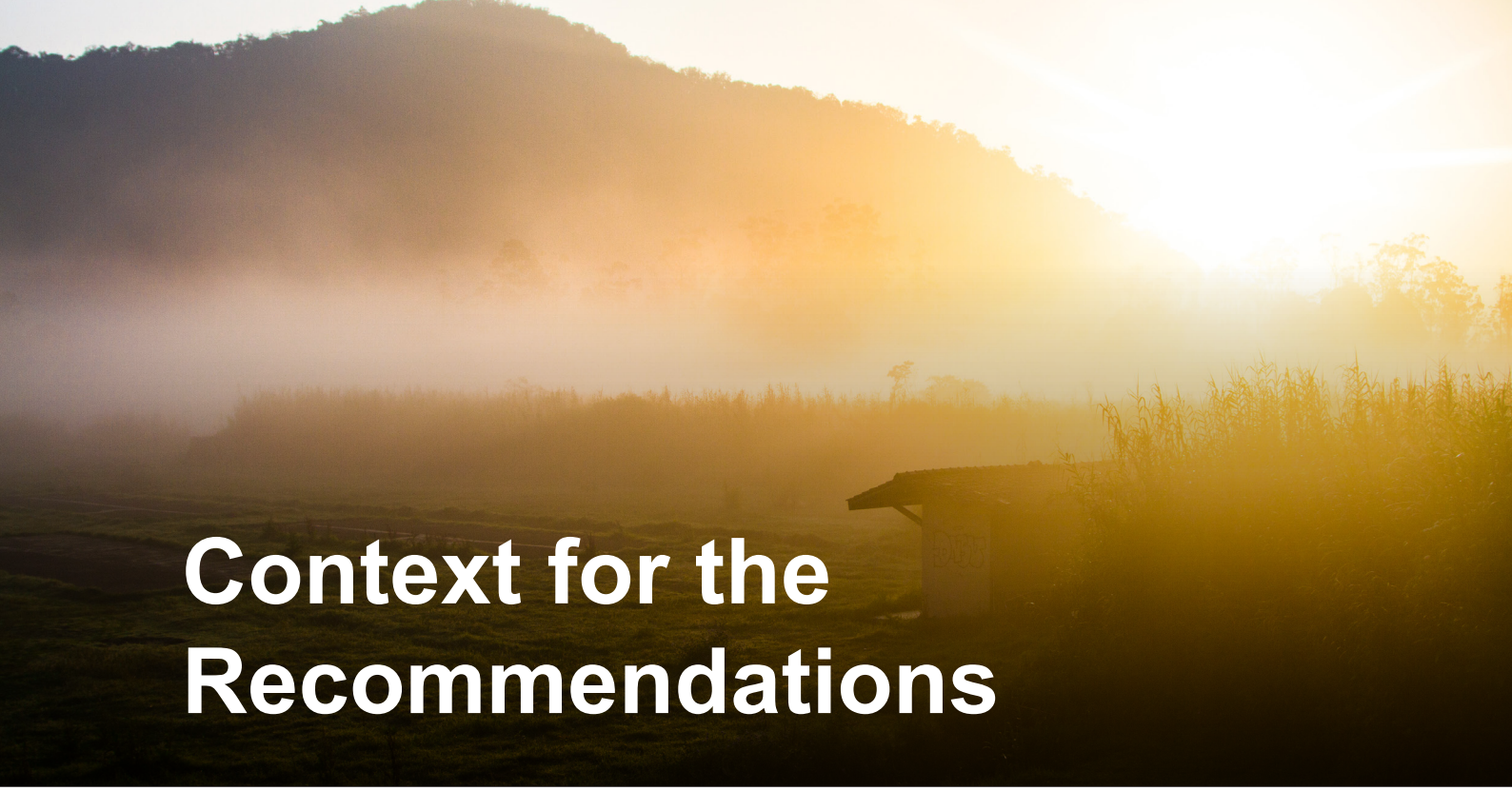


Figure 11: Recommendations and Solution Pathways



Context for the Recommendations

Development of the recommendations

CABI undertook a comprehensive assessment of the application of data management and FAIR principles within ACIAR and its investments. This assessment applied the following methodology:

- desk research
- investment-specific questionnaires
- semi-structured interviews
- analysing the content of reports and other project documentation using a data lens
- a deep dive into three specific data-rich investments
- ecosystem mapping to visualize the current state
- ecosystem mapping to visualize the envisioned future state

The evidence obtained from these different research methods was used to develop recommendations for ACIAR. In particular, significant input was gained from the comprehensive interviews and discussions held with individuals at different levels within ACIAR and through interviews with investment stakeholders, as part of the deep dive into the three data-rich projects. Summaries of the results of these interviews can be found in Section 6 and in Annexes 5, 6, 7 and 8.

The recommendations also draw on CABI's experience with FAIR data principles, together with data sharing developments at national and international levels.

The aim of the recommendations is to assist ACIAR and its partner organisations support its grantees through providing them with a FAIR-based data framework they can follow during their data journey, as appropriate for their level of data maturity and capability. The end goal would be a data strategy based on an ACIAR-specific interpretation of the FAIR data principles (refer to CABI's interpretation of the FAIR principles in Section 2.2).

CABI recognizes that if ACIAR wishes to move towards FAIR data compliance there must be a clear and, most importantly, achievable pathway forward. This pathway must recognise the widely diverse levels of data literacy and data management capabilities, the range of cultural and organisational issues, and the significantly different levels of data management experience across the ACIAR grantees and their host organisations. Because of this diversity, the recommendations aim to provide an appropriate

pathway that meets the broad range of needs, objectives and capabilities of the ACIAR teams and grant recipients. Such a stepped and "tailored" approach will enable ACIAR to move progressively towards the vision of full implementation of FAIR data compliance over time.

The recommendations are also built on the premise that any additional burden on the researcher through implementing improved data management capabilities to facilitate data sharing must be accompanied by appropriate guidance, education and resource support.

"National partners are increasingly asking for support for capacity building on data – it used to be just for stats, now more broadly regarding data management."

It is anticipated that the recommendations will enable ACIAR to work with each new project at a level suitable to that project's capabilities. The intention of the recommendations is to be able to put in place data management "stretch targets" for each project to lift the project participants and host organisations to a higher level of data management capability and data maturity. This "case by case" approach to supporting investment data management improvements will provide an achievable path forward, gradually building capability suited to each project and their partner organisations and participating scientists, but with the common objective of increasing the ability to share project data.

The recommendations are structured in such a way that while proposing a clear pathway forward they also enable ACIAR to control the speed of implementation in a way that ensures progress but without putting significant or additional burdens on the grantees and on ACIAR and its staff. The pace with which each ACIAR investment project moves towards FAIR compliance can be managed by ACIAR through the implementation of these recommendations.

Annex 9 provides some additional context for these recommendations, in relation to the assessment findings, and also discusses a number of external drivers for the recommendations that are considered influential in terms of ACIAR moving towards a FAIR-compliant approach to its investment projects.

It is important to understand that ACIAR is not on its own in pursuing a path towards a FAIR principled data regime for its investments. Many research and donor organisations pursuing this path. Consequently, there are many examples of capabilities available that can be used or modified to support the development of ACIAR-specific FAIR data principles, data management frameworks, data governance and data maturity models. These existing tools and support can be used to benefit ACIAR investments. For example, the Data Sharing Tool Kit[54] provides a powerful example of resources researchers and project managers can use to improve many aspects of their data management activities, benefiting the investments and outcomes. The Tool Kit and other existing tools can be modified to meet specific ACIAR investment data management and FAIR-compliant needs.

A FAIR-based data strategy for ACIAR would support a significantly improved interaction between similar investments, leveraging various elements of previous projects to increase the value and success of subsequent projects. Some projects are conducted with similar stakeholders and collecting similar data, and while the specific outcomes of these projects may be different this study has found there is a level of commonality that could result in increased value of investment outcomes if effectively leveraged. A data strategy and associated data management and data sharing improvements would drive improved outcomes, providing greater benefits to the recipient countries, their host organisations, and the in-country researchers.

[54] https://www.datasharingtoolkit.org/

Setup for success

During the deep dive into data-rich use cases of three identified projects a clear picture emerged, highlighting the significant benefits that can be achieved by undertaking several key and focused steps at the commencement of a project. The outcome of this “setup for success” activity was clearly demonstrated in the Pacific Soils Portal component of the “Soil management in Pacific Islands: investigating nutrient cycling and development of the soils portal” project. The initial focus during the commencement of this project was on establishing a governance framework to resolve issues around data and data exchange activities.

> “National partners are increasingly asking for support for capacity building on data – it use to be just for stats, now more broadly regarding data management.”

While it took some time to get the soils portal project governance arrangements in place, once this was achieved the project had a solid governance structure that allowed it to operate effectively. Consequently, the project participants had clear guidelines regarding data exchange, which enabled them to operate comfortably. With the data exchange agreements in place, the personnel responsible for data collection did not need to worry about whether they could or could not share their data. What they were “allowed” to do with the data was clear. Without this clarity, people generally default to not sharing data, as they are afraid of doing something that is not approved. Developing the governance agreements at the setup stage removed this potential barrier across the Pacific Soils Portal activities.

“Setting up for success” within a project brings considerable downstream benefits. Many industries are now applying this philosophy of increasing their investment efforts upfront to gain significant downstream benefits, including: the car manufacturing, building and construction, and aeronautical and space manufacturing sectors. While these industries are obviously very different from the agricultural research sector, the philosophy of improving their projects and activities through an increased investment at project commencement can bring similar beneficial results. Each of these industries are briefly discussed below.

Vehicle manufacture

Today's car manufacturing industry undertakes significant digital development of a new car and invests heavily in this process before any prototypes are built. This initial investment reduces the need for multiple and expensive prototypes to be constructed since many of the building and design issues that can arise when building prototypes have already been resolved through the digital process.

Building and construction sector

The building and construction sector has introduced a concept called Building Information Management (BIM):

> *"BIM is a broad term that describes the process of creating and managing the digital building model of any construction site or building object at all its stages. The use of a digital model greatly facilitates the design, construction and operation processes in addition to providing an accurate and reliable multi-component information base for decision-making."*[55]

> *"According to the Royal Institution of Chartered Surveyors, the use of BIM on UK Government-funded projects has been reported to deliver savings in the region of £3 (€3.4) billion."*[56]

While there is an acknowledged additional cost at project commencement with BIM, the downstream benefits are significant and exceed the initial expenses. The realisation of these benefits has resulted in the UK government making BIM mandatory for government funded construction projects since 2016. The key barrier to this increased investment upfront is purely 'cultural', with the concept of 'we have always done it this way' being the main obstacle.

ACIAR has the potential to gain benefits and improve investment outcomes through adopting a similar philosophy. Focusing on project setup can help realise downstream benefits through resolving issues early and providing clarity for the project participants. Additionally, focusing effort at the beginning of a new project can help identify activities from previous projects that can be leveraged to reduce time and costs. These might include such things as building on existing data sharing and data governance agreements, as well as re-purposing any relevant tools and IT capabilities. "Setting up for success" has the potential to bring a range of benefits to ACIAR investments.

[55] https://www.planradar.com/gb/bim-adoption-in-europe/

[56] https://irishadvantage.co.uk/the-transformative-benefits-of-bim/

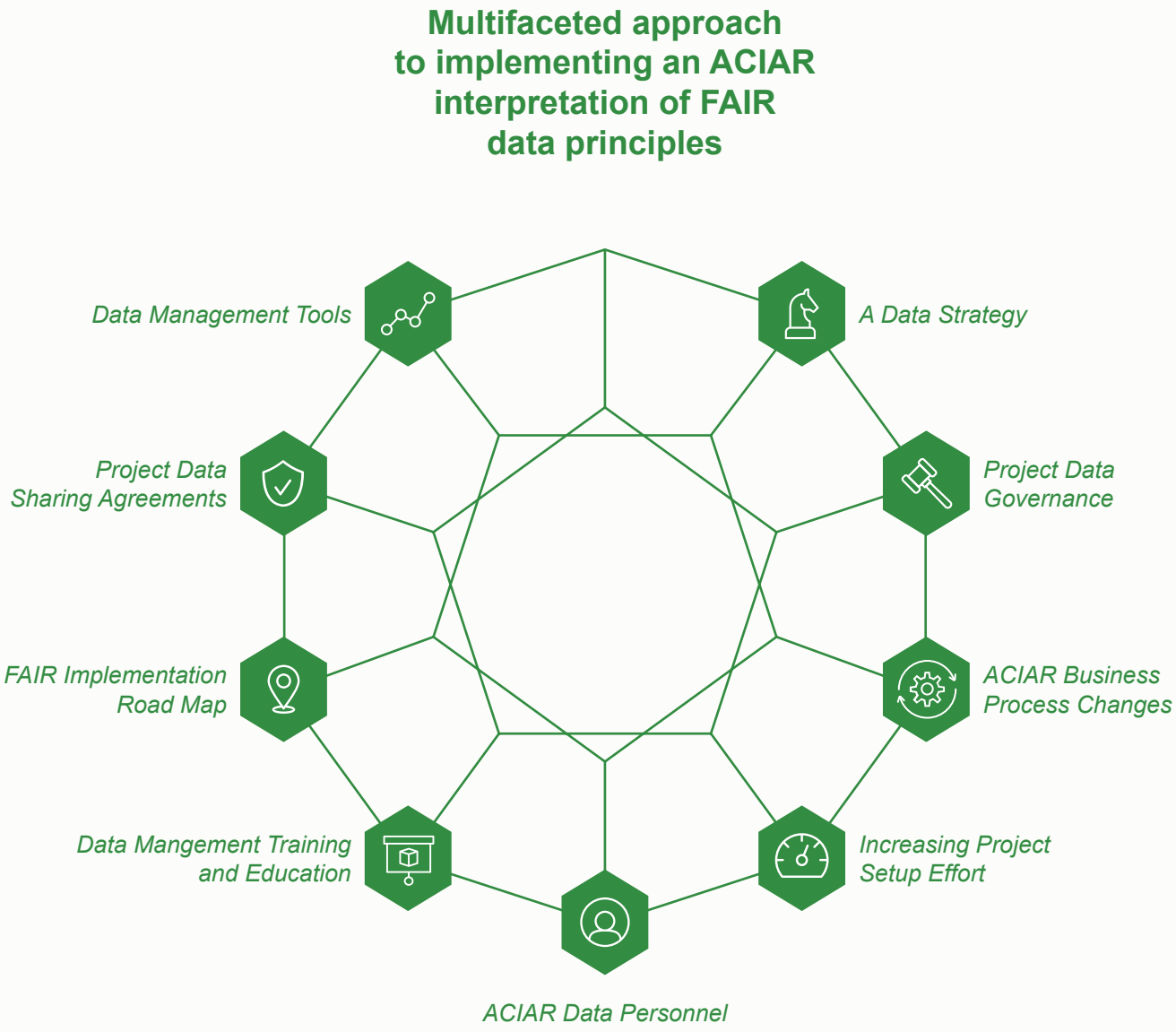


Figure 12: Identified Elements leading to development of Enabling Environment

Recommendations

FAIR implementation within ACIAR

CABI believes that the development of an ACIAR data framework will provide a cross-cutting capability across ACIAR's portfolio that will benefit all ACIAR investments and that will improve the effectiveness and outputs of its investment projects. To achieve this, CABI recommends applying a multifaceted approach to implementing an ACIAR interpretation of the FAIR data principles. The components of this approach could include:

- a data strategy
- data management policies
- project data governance and sharing agreements
- ACIAR business process changes
- increasing project setup effort
- ACIAR data personnel
- data management tools
- data management training and education
- a FAIR implementation road map

A transition to an effective FAIR-compliant method of operation will require activities across several key areas, including:

- data management processes
- knowledge and skills
- strategic oversight
- investment
- monitoring, evaluation and compliance

The recommendations below provide an ACIAR-focused interpretation of these key areas.

Within the recommendations are several "bubble" inserts. These contain statements from individuals obtained during the interview phase of this analysis activity. These statements provide specific ACIAR and research partner perspectives on various data management and data sharing issues, needs and concerns. They are included to demonstrate that the recommendations are closely aligned to the views, thoughts and concerns of ACIAR, and its investment stakeholders.

Data strategy

CABI considers that the highest priority is to develop an ACIAR data strategy, building on an ACIAR-specific interpretation of the FAIR data principles. The interpretation of FAIR that would best suit ACIAR's needs will likely be similar to that used by CABI when supporting FAIR implementation within the Gates Foundation.

> "I think it [the current environment] is crying out for some good policy and structure."

A data strategy would provide an overarching and strategic perspective and vision. It would help drive ACIAR towards a greater digital and data focus in its investments. A data strategy would highlight ACIAR's vision for data management and data sharing objectives, which would help to clearly demonstrate its intent to meet the obligations of the Australian Government's Data Strategy and the Public Data Policy. An ACIAR data strategy would also send a strong message to potential funding partners by describing ACIAR's recognition of the value and benefits of data sharing through a FAIR-based approach within its research investment program.

An ACIAR data strategy should be aspirational and should provide a clear vision. It may take some time to fully develop this strategy, given the diversity of data management capabilities that currently exist in ACIAR's partner organisations and grantees. An extensive consultation will also be required across ACIAR, its co-funding partners and grantees.

Once developed, an ACIAR data strategy would provide the framework for, and would include clear objectives supporting, the long-term implementation of a FAIR data approach, supported by ACIAR.

An ACIAR data strategy would describe the "end game" and overall objectives, identify the various data management drivers, and identify benefits to the researchers, research funding agencies, the research communities and the public in both Australia and the countries supported by ACIAR.

The data strategy could be developed by building on and leveraging other relevant polices, such as those of the Australian Government and relevant international models, such as the EU Horizon's FAIR Data Maturity Model.[57]

> "Stakeholders wish for clarity – need clear rules to follow, with appropriate systems and adequate technical support."

[57] https://ec.europa.eu/isa2/actions/measuring-maturity-data_en/

Data strategy components

The Australian Data Commissioner[58] suggests that a government entity should develop a data strategy that contain the following elements:

- vision statement
- alignment with government data policy
- alignment with internal policies and strategies
- strategic objectives
- initiatives and areas of focus
- measurable outcomes and indicators of success
- a road map for implementation

ACIAR may wish to follow these recommendations, although obviously other structures would also be appropriate. It is considered that a data strategy should also provide a range of relevant information that provides the reader with details regarding the how, when and why. Some of these types of information can be summarised as follows:

- Objectives (supporting three key themes seen as significant by the Australian Government):
  - maximising the value of data
  - ensuring trust and protection
  - enabling data reuse
- Drivers
  - Why have a Strategy?
  - What is "driving" ACIAR to create a data strategy?
- Benefits
  - Who benefits from the data strategy?
  - What are these benefits?
- Approach (FAIR compliance)
  - How will this data strategy be developed?
  - What are the timelines?
  - How will the data strategy be implemented?
- Outcomes
  - What capabilities, tools and related support will underpin this data strategy?
  - How will ACIAR support its research partners?

The development of an ACIAR data strategy is best seen as an iterative process. While the outcome and objectives of the strategy should be clearly articulated, it may be necessary to develop the data strategy in a series of phases. A vision statement should be developed first. The initial data strategy phase should support the first year or two of any proposed implementation activities, with a second and possibly third phase of the strategy leading to meeting the initial vision statement.

[58] https://www.datacommissioner.gov.au/data-management/foundational-four/strategy

Data management policies

It is considered that an ACIAR data strategy would be usefully underpinned by several data-related policies. These policies would provide the mechanisms, tools and specific requirements for implementing the strategy using an approach that is in line with the wide-ranging capability levels, skills and issues faced by the researchers and their host organisations. The data policies would establish suitable mechanisms that are achievable for each individual research project, regardless of the data management maturity and capability of the project participants.

> “Such digital platforms [the Soil Portal] are seen as important to ensure data is available and safeguarded for scientists and policymakers in the future.”

CABI suggests that several data-related policies should be developed, supporting specific mechanisms for building ACIAR’s and the investment projects’ data management ability to achieve the data strategy vision. There are several areas and activities that could be supported through the development of one or more ACIAR data policies.

> “It is for ACIAR to look into a plan that tackles the storage of data, access to the data, the use of data, and then transforming that data to information and analysis that will be useful to end users.”

Those activities and areas considered to require a policy or similar mechanisms include the following:

- A definition of data and the outcomes of the analysis of any data in terms of intellectual property and ownership issues.
- The requirement to include a data management section in all grant applications, covering project data management and data sharing practices, together with the ability of those submitting the grant proposals to include relevant data management resource costings.
- Inclusion of a section in the grant contract covering the projects’ responsibilities for data management and data sharing at a level that is suitable to the project partners and participants’ capabilities.
- Project-related metadata requirements to ensure the recording of basic, but important, details covering the overall project activities and describing the datasets collected.
- The establishment of a process for developing a project-specific glossary of definitions for each project, especially relating to those project data items that are required when a project includes a number of researchers and organisations collecting and potentially sharing their data.
- A mechanism for determining the level of FAIR data compliance in each project, via a suitable FAIR data maturity model.
- Establishment of a project data decision tree to provide a mechanism for assisting with determining the level of individual project proposals’ data literacy and capability (data maturity) in order to include a small “stretch target” for the projects’ data management activities to improve grantee and host organisations data capability.
- Establishment of a data governance policy for both ACIAR and, over time, to cover its investment projects.
- Development and provision of suitable data management and analysis tools, and training and education, commensurate with the capability levels of each project’s researchers.
- Establishment and definition of the concept of data custodianship.

> “There have been no discussions till now on the protocols and standards of data collection.”

Some of the activities and mechanisms described above could be included in several policies to provide a more agile approach when change to the policies are required, or they could be developed as a single umbrella ACIAR investment project data policy. Considerable consultation and discussions will be required to determine the most suitable approach for ACIAR and its partner organisations and to determine what these policies should contain.

As part of developing data-related policies, ACIAR could examine what relevant policies and standards already exist in this area, within the agricultural science community. While it would not be considered appropriate to immediately implement data format standards, being aware of existing international standards and establishing procedures to gradually move in the direction of these standards would be beneficial in the longer term: for example, looking at the work of the ISRIC[59] and its soil standards to see if adoption of their approach would be of benefit to ACIAR and grantees in the long term.

> “If we had a unified data storage system for all the projects, it would be amazing because you could do meta-analysis and analyse across projects by gathering different data sources.”

Cross-project collaboration

The extensive review of ACIAR investments conducted as part of this study identified that there are several potential overlaps between projects that could be exploited through improved collaboration. This would enhance the outcomes of each project and would greatly benefit project stakeholders. Where commonality exists in the projects reviewed by the deep-dive exercise, each project demonstrated strengths and weaknesses in some areas of common interest. For example, in the data-rich projects some emphasised the user needs, with limited focus on how the data would be managed, while a similar project focused on establishing an effective data management platform, with limited effort to understand the user needs.

The areas where closer collaboration between projects could bring significant benefits include:

- data sharing
- governance and sharing agreements
- data collection tools
- data management tools
- standards
- repositories or portals

Cross-project collaboration, building on and exploiting the strengths of similar projects, would increase the value and outcomes of each project. This collaboration could leverage the activities of other existing or previous (historical) projects. For example, if a data governance or data sharing agreement has been reached with stakeholders of a previous investment, and if the new investment includes some or all of these stakeholders, then it would be more efficient to simply add new types of data to the existing agreement, rather than negotiate a separate agreement.

The deep-dive investigation into data-focused use cases within three data-rich investments highlighted the significant potential that could be achieved through closer collaboration. For example, within the

[59] https://www.isric.org

Pacific Soils Portal and the Geospatial Platform examples of two Pacific Island investments there are activities and capabilities in each project that could benefit the other investment. The Geospatial Platform showed the benefit of significant user consultation and co-development to create a highly effective and well-used data collection application based on an open-source platform. This could be relatively easily modified to include new data types, such as soils data. The Pacific Soils Portal is also based on an open-source platform and perhaps could be modified to manage data collected using the Geospatial Platform. Integrating the data from each of these projects would provide significantly richer information to the in-country users.

> “The establishment of the project Governance Group, with representatives from all participating countries, provides a forum to discuss a range of issues, such as data access and data sharing arrangements.”
> **(Pacific Soils Portal)**

ACIAR can identify ways to improve the outcomes of future projects by examining the work of other investments and determining areas for potential collaboration and the sharing of both data and other capabilities (such as data governance and data sharing agreements). While effectively implementing this type of collaboration would require resources, many savings can be made through leveraging existing capabilities and agreements. Additionally, this would provide a mechanism for assisting ACIAR to move towards more complex trans-disciplinary projects. Perhaps a “small” integration project could be funded to underpin this increased collaboration, with appropriate personnel supported to create the linkages and to modify existing capabilities to provide a level of integration.

Business process changes

Some changes are suggested to allow ACIAR’s investment-related business processes to begin the implementation of a FAIR-based approach to research projects. Through some simple changes ACIAR can begin to create an increased emphasis on improving project data management, leading to an increase in data sharing, through several simple steps. These steps could be supported by a more formal policy, as described in the section above.

Preliminary Project Proposal and contract changes

The Preliminary Project Proposal (PPP) should have an additional sub-section within *Section 3 – Implementation*, covering some of the following details:

- What types of data is it anticipated will be collected?
- Are there any specific data standards that will be followed?
- Will a data governance and/or a data sharing agreement be developed?
- Will this data be shared (with other project partners or externally)?
- How will this data be stored and/or managed and who will do this?
- Skills development requirements.

Additions are suggested for inclusion within *Section 4 – Resourcing – What inputs are needed?* These additions should include a corresponding section to that described above, covering the resourcing and costs necessary to achieve an appropriate level of data management capability within the project. The inclusion of a section on data management costs, etc., in the Resourcing section of the PPP would clearly indicate to the researcher that ACIAR recognises that data management is an additional cost but one that can be reduced through the provision of additional resources and training. The Resourcing section of the PPP may also be an appropriate section in which to include any data management education, training or other capability needs of the researchers and their host organisations.

In the evaluation of the PPPs, ACIAR can make determinations as to the capabilities of the potential researchers and their organisations using a data management maturity model or decision tree approach, which would help identify the projects' existing and also potential data capability and data maturity levels. This would assist in determining what level of resourcing and training may be required to provide a stretch target to improve the recipient's data management capabilities.

> "Some direction could/should be provided in project specifications. Basic data sharing agreements in principle would be beneficial and would give guidance to where the data should be delivered to."

When a project has been approved for funding a project contract is developed. In a similar manner to the PPP, it is recommended that this contract includes a section on the data management obligations and data-related deliverables for the project, including timeframes for these deliverables. As has been mentioned previously, the expectations and contracted deliverables must align with the capabilities of the recipients and their host organisations, although there will ideally be a level of "stretch" for these data management activities to ensure there is some improvement in capability and skills at the end of the project.

How these data-related sections of the proposals and contracts will be specified without placing an increased burden on both the potential grantee and ACIAR evaluation processes will need to be carefully determined through consultation.

> "We have nothing in contracts to say how researchers should collect or store data."

The PPPs and subsequent project contracts could include sections on data that describe aspects such as:

- data quality
- data standards
- data management plans
- sharing and access arrangements
- data governance

Comment: *It was evident from the deep-dive interviews that when any agreement, especially data sharing agreements, are made before actual project commencement the projects benefit considerably. Additionally, it was clear that some projects are currently operating in a siloed environment and would benefit considerably from leveraging previous and similar projects, especially when they have common stakeholders. For example, if a data sharing agreement exists between stakeholders, and these are the same stakeholders in a new project, then it may be easier to simply add new data types to the earlier negotiated agreement, rather than creating a separate data sharing agreement. Likewise, there may be data from the previous project that could be leveraged for the new project. Also, any tools or other capabilities developed by other similar projects could possibly be used or modified to meet the new projects' needs at less cost and time than creating new tools. A useful practice would be to review previous projects where similar data is collected, and ideally use the same data formats in the new project, enabling data from both projects to be easily integrated, providing greater value to the researchers.*

“No matter how much you make it clear that it is a contractual obligation, unless people have time and the systems setup to manage data and clean it and publish it, it doesn’t happen. … as a funder we need to be aware of managing and resourcing so that something closer to an ideal state can exist.”

“If ACIAR mandated the project team to align budget and resources to build the capability and technical infrastructure with proper support and direction this would definitely help them to plan better.”

Increasing project setup effort

In a similar manner in which BIM has been implemented to significantly benefit the construction industry, as described above, an increased focus on "investing more effort" during the setup phase of a project would help advance data management capability within ACIAR investments. Conducting several activities during the setup and early phases of a project will provide significant improvements downstream as the project matures. By undertaking some additional work during project commencement, a number of potential barriers and inefficiencies in the project can be eliminated. Areas that would benefit from increased focus include:

- project data governance
- data sharing agreements
- working with the key in-country stakeholders to pinpoint the user needs
- investigate the opportunities for co-development of any relevant capability – such as data collection, data management and data access tools
- identifying and working closely with an in-country "champion"
- investigating how a project-developed procedure or technical capability could be embedded in the daily work activity of the host organisation.

"This project focused on understanding the need of the region before designing an intervention, an approach that has been integral to the project's success."
(Geospatial Platform)

"Co-development of the geospatial platform was essential to get better engagement and buy-in."

Personnel

It is recommended that ACIAR look at recruiting personnel who are capable of supporting data management. Adoption of the newly developed concept of a data concierge[60] – a person who provides advice and support to researchers on data management activities – could be critical in improving project data management and data sharing. A number of those who were interviewed expressed a desire to improve their data management practice but did not know how to go about this. They require guidance in the form of strategies and policies, as well as assistance – which could be provided by a data concierge-type role.

> “If we were going to add a requirement around FAIR data, we would need somewhere a capacity to keep an eye on it.”

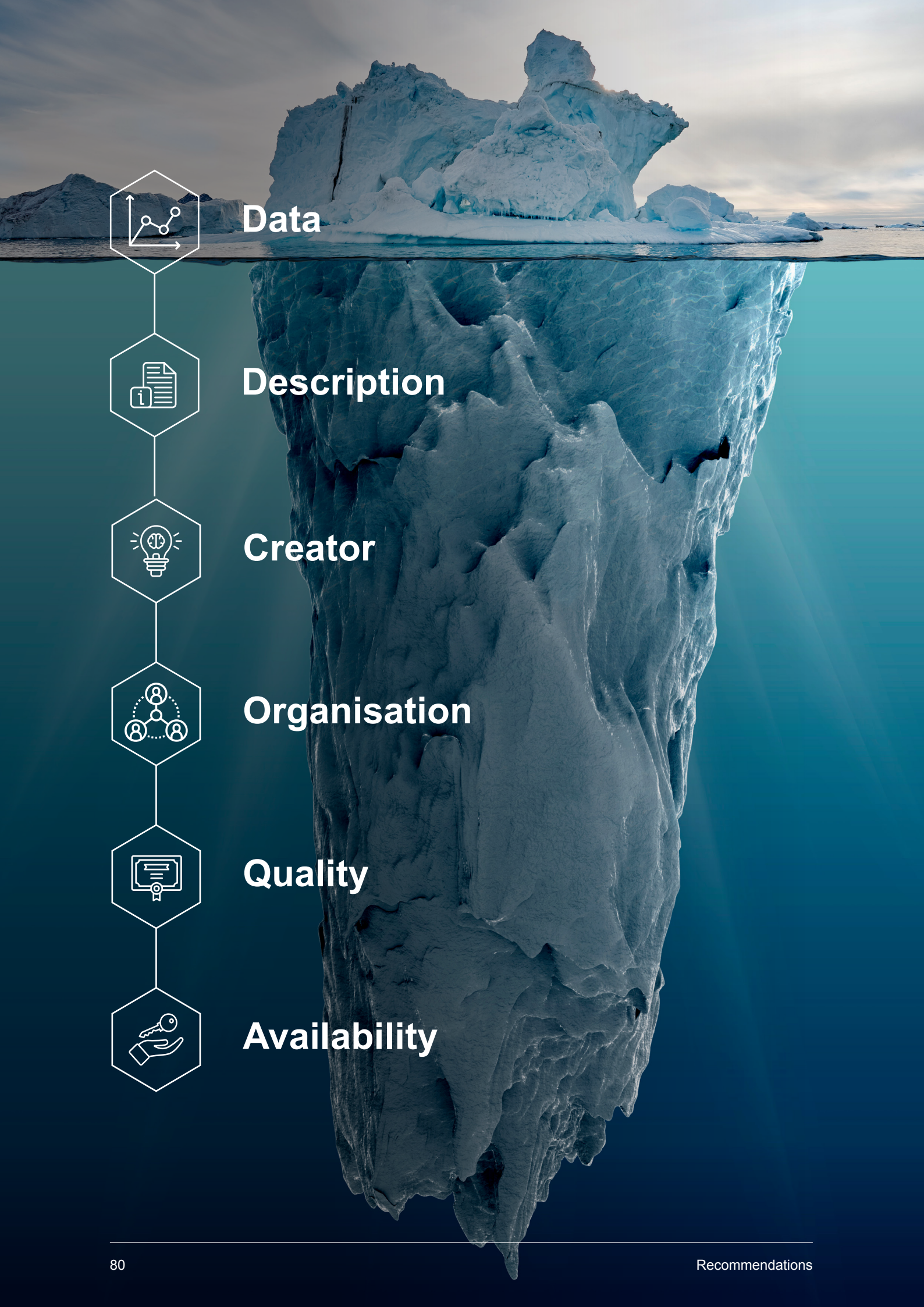
An ACIAR data concierge would be a “hands on” practitioner who understands data management issues in the agricultural space, how different levels of technologies can be applied, what types of data management approaches would be suitable for specific projects, and who would be able to provide practical help and advice during the project proposal phase and during the research activity phase. The data concierge paper referenced above is the result of a study at Monash University, one aim of which was:

> *“to determine what targeted research data management support researchers require to enable them to spend more time on research, or to undertake their research more efficiently.”*

More details on this paper and the role of a data concierge are provided in Annex 11.

In the longer term, ACIAR may look to recruit a data scientist who can provide higher-level strategic data management support, including introducing data analytics concepts that would support larger, multidisciplinary and multi-research teams and regional projects. A data scientist would be able to build on an improved data sharing capability to assist and advise on longer term and higher-level data management strategies, data integration and analysis, using advanced data analytic tools. As data management and data sharing improves, the value of a data scientist would be demonstrated through improving the return on investment of each project’s data, supporting possible integration of data from other relevant projects, leading to new areas of research and innovation. Such an approach is considered a longer-term activity, but this direction would significantly boost ACIAR investment projects’ contributions to the achievement of national and global SDGs.

[60] Splawa-Neyman, P. (2022) ‘What the dickens is a data concierge? Researcher interviews and data management reviews: misinformation, appreciation and remediation’, figshare. Conference contribution. https://doi.org/10.6084/m9.figshare.20104631.v1


Data
Description
Creator
Organisation
Quality
Availability

Data management tools

CABI recommends developing some basic and simple tools that could be made available at no cost to the researchers. An initial suggestion would be the creation of a very simple metadata tool tailored to meet agricultural research needs, to make data **findable**. Such a capability could perhaps be based on the Dublin Core metadata model, although some investigation may be able to identify an agricultural data-oriented metadata approach – possibly something similar to the Pacific Data Hub[61] or the Pacific Soils Portal – to simplify the findability of data.

To be of value, such a tool would need to capture three aspects of each investment project:

- A simple description of the project (outcomes, etc.), personnel and organisations involved.
- A basic metadata record covering the different types of data collected in the project – this could include details such as whether the data were sharable, or other relevant information for potential re-users of the data to inform them of the "suitability" (fitness for purpose) of the data for their project.
- An ACIAR staff-only section providing a range of "project-related intelligence information".

> "ACIAR does not have a system in place. A unified data storage system is not available which can allow metadata analysis."

An ACIAR-only section could contain additional information to support the development of pipeline projects, improve the focus, and reduce setup effort, if there was a greater awareness of previous activities. This additional material could contain a number of details about the project, including the following:

- Descriptions of data governance and data exchange agreements between stakeholders that could be leveraged to create agreements for new projects with similar stakeholders.
- Key individuals and points of contact, such as known "in-country champions" and perhaps other significant "experts" who could assist in supporting or providing in-country influence in getting traction on new projects.
- Details describing data collection and data management tools that could be used by other similar projects.
- Known data repositories to support similar types of data from new projects.
- Project ecosystem diagrams showing the relationships between stakeholders and potential key users, to assist in identifying possible overlaps and leverage points to improve new projects.
- Evidence of impact, to support the breaking down of project silos.
- Documentation indicating the details of data and any relevant tools and capabilities from other investments that have been used in the new project. Such examples could be subsequently used to communicate the benefits of interactions with other investments.

Most of the information required for the project-level layer could be obtained from the PPP and/or the project contract. Some details of the data component could also be populated from the PPP if this type of information was required in the PPP.

[61] https://pacificdata.org

> “ACIAR has a lot of data collected by projects, no systems or rules for how to make data accessible.”

To be of value, the “public” metadata information would need to be made easily accessible to other ACIAR researchers and the broader community. A simple online capability on the ACIAR website could hold these metadata records and provide an easily usable tool for seeing what similar projects have already been undertaken, and what data was collected, together with sufficient details to provide a researcher or organisational point of contact for those seeking access to the data. It is important to remember that providing details of the existence of data does not automatically mean that this data is available, but it does mean that if the data is of relevance, a discussion can be entered into to determine its availability. If there is no knowledge of this data, then no discussions can take place and data may be re-collected, at extra cost to the investment.

Additionally, such a metadata capability would provide a range of valuable information that could be leveraged in a new project, including the existence of relevant data repositories or portals, data collection tools and data management systems. In some cases, knowing about the existence of these capabilities could reduce the effort of a new project, enabling them to focus more on core research activities.

> “Maybe we should manage the data ourselves or stipulate where it should be put so we know where we can find it in [five, ten, or fifteen] years.”

An additional tool that could be used to assist projects, or review PPPs to determine their data management capabilities, would be a data maturity model, such as the EU Horizon FAIR Data Maturity Model.[62]

A matrix providing a number of questions on stakeholders’ data management capabilities, knowledge and skills could be created to help quickly evaluate the level of data maturity that exists. Once this level is determined a data management stretch target can be developed and resourced to help build in-country capacity by improving the current level of data maturity.

Data management skills and knowledge

Education, training and tools will be required to increase grantee researchers’ skills in data management, data sharing and FAIR if ACIAR is to achieve its FAIR objectives. This skill development activity could potentially be a role for the suggested data concierge position mentioned above.

The development of such training and education tools will take some time and consultation, especially given the diversity of data literacy and data management capabilities among grantees and host organisations. However, such tools need not be developed from scratch. Many sectors have gone through similar processes and will have suitable types of tools already available that, with minimal modification, could meet ACIAR grantee needs. Refer to the Data Sharing Tool Kit discussed in Section 7.1.[63] An initial role of the data concierge would be to develop and run basic data management training for new grantees, tailored to their level of knowledge and capability.

A further activity to support project data management would include holding country or regional workshops where data management training and support could be provided to local grantees. Again, these would be tailored to suit the level of capability of these researchers – and would perhaps be delivered by an ACIAR data concierge.

[62] https://datascience.codata.org/articles/10.5334/dsj-2020-041/

[63] https://www.datasharingtoolkit.org/

FAIR implementation road map

The concepts described above cover a broad range of suggested activities and deliverables. Some of these, such as the data strategy, should be developed with a high level of priority. The strategy is seen as the high-level framework that will inform the development of the other recommended concepts outlined in the sections above.

Some of the recommendations could be implemented concurrently with or shortly after the data strategy is developed – such as the inclusion of data management sections in the PPPs and research grant contract. The timely recruitment of a data concierge would provide a suitable level of expertise to work on several of the recommendations, and would reduce possible workload increases for other ACIAR personnel.

Communicate the benefits of sharing data

Based on observations made during this study, data sharing does occur within and among current programs, but when sharing occurs it is generally ad-hoc and there is no standard approach to it. For ACIAR to drive organisational-level change, it will be important to identify and communicate the benefits of data sharing. Benefits can accrue to ACIAR and partners when the following takes place:

- communicating an investment's data sharing successes
- establishing a method of acknowledging data sources to increase recognition of the contribution made by sharing datasets
- publishing papers and communicating in other ways that give recognition within science communities
- highlighting benefits to host organisations/countries – promoting data for public good
- leveraging data sharing to enable access to larger datasets than would otherwise have been possible within a single project

Articulating and communicating these benefits will provide clear acknowledgement that ACIAR values data sharing and will set the expectation that project teams are willing to reflect and rethink their current data sharing practices.

ACIAR could investigate different approaches and channels for communicating stories where shared data was beneficial and could openly acknowledge and recognise teams that have practised good data sharing practices. Currently, there is limited visibility of data sharing activities among teams and across projects. Effectively communicating the practices and learnings of project teams that have achieved good data management will encourage a more transparent and open culture. Furthermore, celebrating a team for applying good practices may motivate them to continue this practice within current and new projects, and may inspire other project leaders to follow suit.

> "An institutional change is required. We have a focus on the quality of science at the expense of adherence to some of the management details in our contracts. The projects need to be managed in alignment with larger organisational needs."

Annex 1 CABI's Project Methodology

CABI – ACIAR Project Workstream Details

Workstream 1: Discovery Phase - ACIAR Situational Analysis / Alignment

Summary of activity

Workstream 1, the discovery phase, focussed on preliminary alignment between the CABI team and ACIAR team to gain a clear understanding of the ACIAR team's expectations, as well as become more familiar with ACIAR's investment data-related strategies, grants and personnel.

Objective

To ensure that the project team has a clear understanding of ACIAR programmes, strategies and desired outcomes of the project, through desk-based research, remote consultations, examining pre-existing documentation, and documenting overall ACIAR context.

Activities

- Compiled and reviewed programme, strategy, policy and investment documents from ACIAR investments regarding FAIR data, open data, data access, or other data-related grants and concepts (CABI/ACIAR).
- Reviewed expectations of project with key ACIAR personnel (CABI/ACIAR).
- Finalised investment selection criteria for Workstreams 2 and 3 (CABI/ACIAR).
- Engaged with additional ACIAR teams, as needed, such as legal, policy development, and program coordination (CABI/ACIAR).

Deliverables

- Documented clear expectations to ensure understanding between the project team and the ACIAR team.
- Finalised investment selection criteria.

Workstream 2: Assessment of the state of FAIR implementation across ACIAR investments

Summary of activity

Workstream 2 focussed on gathering evidence of the current state of FAIR understanding and implementation across an agreed sample of ACIAR investments. The CABI team worked with ACIAR RPMs and grantees within this workstream.

Objective

- To determine evidence on the current state of FAIR implementation across ACIAR investments. This includes, from a high-level perspective, how successful FAIR implementation is when the investment is systemic, or not-systemic.
- To demonstrate the demand for FAIR implementation as well as the demand for support to overcome challenges in FAIR implementation.

Activities

- Selected 12 investments that are representative of ACIAR research (CABI/ACIAR).
- Selection criteria were created in Workstream 1 and covered geographies, domains, investment stage, type of data collection, sharing, and management, presence of sensitive data, systemic approaches and influences and other criteria that ACIAR prioritised in Workstream 1.

- Examination of the 12 investments through desk research, including reading documents and understanding investment context (CABI).
- Research involved developing a series of questions to ensure the CABI team understood the problem to be solved, risks, assumptions, and buy-in from the intended beneficiaries.
- Interviewed the RPMs and grantees to discuss challenges and successes in FAIR data implementation.
- Each (remote) interview was scheduled to take about 1-2 hours.
- Analysed each investment from an individual perspective.
- Analysed results across investments to present a holistic picture of the current state of FAIR implementation across ACIAR investments.

Deliverables

- Evidence-based assessment of the state of FAIR implementation across ACIAR investments.

Assumptions

- Investments covered a wide range of criteria, however, the most important was if FAIR has been implemented from a systemic perspective.
- RPMs and grantees engage with the CABI process.

Workstream 3: Deep-dive into regional investments with systemic-level FAIR interventions

Summary of activity

When implementing FAIR, often the success of implementation depends on situations out of the specific investment's control. Within CABI's work, we call this the "enabling environment", or how strong the data sharing culture exists within the context of the investment. This workstream delved down deeply into several regional investments with systemic level FAIR interventions to determine challenges, opportunities, needs, and situational context. The deep-dive was in three investments in ACIAR priority regions, and was a sub-set of the number in Workstream 2.

Objective

To understand specific regional nuance and context to FAIR implementation in select systemic-level ACIAR investments.

Activities

- Relationship building with regional partners, grantees, and ACIAR RPMs (CABI).
- Undertook interviews via Zoom with key stakeholders in order to understand common barriers to FAIR data implementation within the investment, and within the system (CABI).
- Conducted data ecosystem mapping and use system thinking informed approaches with partners (CABI).
- Identified potential capacity development opportunities for FAIR implementation.
- Developed an investment-specific roadmap which will highlight our findings, and clarified demand for potential future engagement and barrier remediation.

Deliverables

- Identified common, and nuanced regional technical, cultural, and institutional barriers, and the necessary infrastructure, training, and capacity development needs documented.
- Developed a clear understanding of the purpose, actionability and sustainability of an investment or project specific roadmap for successful FAIR implementation.
- Synthesized findings of regional assessment and feedback to ACIAR team.

Workstream 4: Needs-based assessment (demand) of RPMs and grantees for FAIR data implementation guidelines and resources

Summary of activity

The workstream occurred concurrently throughout Workstreams 2 and 3, and consolidates information from RPMs and grantees on common constraints to FAIR implementation that are faced, and what type of support they need to implement good FAIR data practice in their current and future investments. This work included the building of understanding of incentives for following good practice.

Objective

To determine common constraints that are faced, and what type of support RPMs and grantees need in order to implement FAIR data in active and future investments, including what incentive systems may need to be in place.

Activities

- Throughout Workstreams 1 and 2 activities, questions were asked of RPMs and grantees concerning what constraints they have faced, and what their needs are for support. They were also be asked about incentives to manage and share data (CABI).
- Draft needs-based assessment (demand) component in final Report to show common constraints faced and potential support options for RPMs and grantees (CABI).

Deliverables

- Needs-based assessment component in final Report.

Assumptions

None noted.

Workstream 5: Vision and Recommendations for next steps on how best to overcome barriers within the scope of ACIAR grant making practices

Summary of activity

This workstream consolidates the evidence gathered from Workstreams 2, 3, and 4 into a cohesive package of conclusions and insights for ACIAR. CABI and ACIAR reviewed the package together to co-create a vision for FAIR implementation in ACIAR investments. Based on the vision, CABI provides recommendations for next steps to achieve this vision.

Objective

To clearly provide ACIAR with evidence-based recommendations to achieve their future vision for FAIR implementation across investments.

Activities

- Consolidated evidence from Workstreams 2, 3, and 4 into a clear report of conclusions and insights for ACIAR (CABI).
- Reviewed work with ACIAR and co-create a vision for FAIR implementation.
- Based on the vision, CABI provides recommendations for next steps to achieve this vision, which will include changing current business processes, leveraging existing tools, guidelines, capacity building and resources, and potential incentives.

Deliverables

- Articulated vision and recommendations for next steps on how best to overcome barriers within the scope of ACIAR grant making practices.

Assumptions

- Evidence-based report of project conclusion and insights.
- Future vision for FAIR implementation.
- Recommendations for next steps including use of tools, guidelines, capacity building, and resources, and potential incentives.

Workstream 6: Project Management

Summary of activity

This workstream ensured clear communication, delivery and reporting between project partners, ACIAR stakeholders and other stakeholders as required. Project finances and timely completion of deliverables and milestones were managed as per project plans. This workstream was completed by CABI.

Objective

To successfully manage the overall project and to maintain a strong working relationship with ACIAR.

Activities

- Create and manage a project plan and risk log.
- Fortnightly status reviews between project team and key ACIAR stakeholders.
- Creation of an end project report as per pre-agreed narrative and financial reporting.
- Determination of appropriate asynchronous meeting and communication structures.
- Attendance at key meetings regarding data initiatives as requested.

Deliverables

- Agreed project plan and amendments as applicable.
- Risk register which will identify in each case key risk name, area to which the risk applies in project plan, detail of specific risk and implications, risk score based on impact and likelihood, management strategy for dealing with risk, risk indicators, and responsible individual.
- End project report including narrative and financial reporting, and lessons learned.

Assumptions

None noted.

Annex 2 ACIAR Desk Research

ACIAR Operating Structure

The **Australian Centre for International Agricultural Research** (ACIAR) is the Australian Government's specialist **agricultural research-for-development** agency. ACIAR operates solely on budget appropriation from Australia's **Official Development Assistance** (ODA).

ACIAR is established by the Australian Centre for International Agricultural Research Act 1982 (ACIAR Act) and is an agency of the **Department of Foreign Affairs and Trade** (DFAT) portfolio. ACIAR has an executive management structure, headed by the Chief Executive Officer, who is accountable directly to the Minister for Foreign Affairs.

ACIAR is a **non-corporate Commonwealth entity** under the Public Governance, Performance and Accountability Act 2013 (PGPA Act). Its staff in Australia and overseas are employees of the Australian Public Service under the Public Service Act 1999.

ACIAR **mission** is to achieve more productive and sustainable agricultural systems, for the benefit of developing countries and Australia, through international agricultural research partnerships.

The **vision or purpose** of ACIAR is to contribute to reducing poverty and improving the livelihoods of many in the Indo-Pacific region through more productive and sustainable agriculture emerging from collaborative international research.

ACIAR 10-year strategy 2018-2027 is to:

- Broker, facilitate, invest in and manage strategic partnerships in agricultural research-for-development.
- Ensure that its research for development programs are equitable, inclusive and empowering

Corporate plan outlines their work in 3 key areas[64]

- **Global research collaborations** – We develop and foster partnerships and relationships with other international research and development agencies, the most significant being CGIAR. We also develop and foster partnerships with development donors and the private sector to pursue shared goals and ensure that ACIAR-funded research results are implemented at scale.
- **Bilateral and regional research projects** – We generate knowledge from ACIAR projects and programs to empower smallholder farmers, extension agents, scientists and policymakers to take on the intersecting challenges of growing more and healthier food and reducing poverty while using less land, water and energy.
- **Scientific and policy capacity-building activities** – We identify and establish opportunities for individuals and institutions in partner countries to boost technical, policy and management skills in agriculture, fisheries, forestry and management of land and water resources.

Strategic Objectives:

ACIAR brokers and invests in **research partnerships** with developing countries in the Indo-Pacific region to build knowledge to support crucial development objectives:

- **Food Security and Policy Reduction** – Improving food security and reducing poverty among small holder farmers and rural communities
- **Natural Resources and Climate Change** – Managing natural resources and producing food more sustainably, adapting to climate variability and mitigating climate change
- **Human Health and Nutrition** – Enhancing human nutrition and reducing risks to human health

[64] https://www.aciar.gov.au/publication/corporate-publications

ACIAR works to ensure that its **research-for-development programs** are equitable, inclusive and empowering:

- **Gender Equity and Women's Empowerment** – Improving gender equity and empowerment of women and girls
- **Inclusive Value Chains** – Fostering more inclusive agri-food and forestry values chains, engaging the private sector wherever possible
- **Capacity Building** – Building scientific and policy capability within our partner countries

Research Programs (Investment FY 2021-22):

- **Agribusiness (AUD 5.3 mil)** – Unlocking economic opportunities for farmers
- **Climate Change (AUD 1 mil)** – Adapting to and mitigating climate change effects
- **Crops (AUD 3.6 mil)** – Improving food security and farmer's livelihood through more productive and sustainable crops
- **Fisheries (AUD 10.1 mil)** – Improving livelihoods from productive aquatic farming systems and sustainable fisheries
- **Forestry (AUD 5.4 mil)** – Scientific support to establish, manage and sustainably use forests
- Horticulture (AUD 7.8 mil) - Improving fruit, vegetable and ornamental crop production
- **Livestock Systems (AUD 8.5 mil)** – Developing more productive, profitable and sustainable livestock systems
- **Social Systems (AUD 4.6 mil)** – Putting people at the centre of agricultural research-for-development
- **Soil & Land Management (AUD 6.6 mil)** – Introducing conservation agriculture
- **Water (AUD 4.2 mil)** – Improving agricultural water management through innovative technical and policy approaches.

Relationship between ACIAR Portfolio Budget Statements, Corporate Plan and Annual Performance Statement

Portfolio Budget Statements

Outcome 1 - Mission

To achieve more productive and sustainable agricultural systems, for the benefit of developing countries and Australia, through international agricultural research and training partnerships

Program 1

International agricultural research-for-development for more productive and sustainable agriculture

Corporate Plan

Purpose - Vision

To contribute to reducing poverty and improving the livelihoods of many in the Indo-Pacific region through more productive and sustainable agriculture emerging from collaborative international research

Target

Through annual project progress reports, mid- and end-project reviews, long-term adoption studies and impact pathway analyses, case studies (quantitative and qualitative), and financial activity indicators, as appropriate to each performance criteria, ACIAR will measure the extent to which these criteria are:

- Exceeding Expectations
- Meeting Expectations
- Not Meeting expectations

Performance criteria

Delivery of programs in line with the ACIAR 10-Year Strategy 2018–2027:

- Global research collaborations
- Bilateral and regional research projects through 10 research programs
- Scientific and policy capacity building activities

Annual Performance Statement

Report achievement of targets set for performance criteria

Geographies Covered

In 2021–22, ACIAR will work in 31 countries in the Indo-Pacific region, with 375 research organisations, universities, government agencies and private sector organisations from Australia and partner countries.

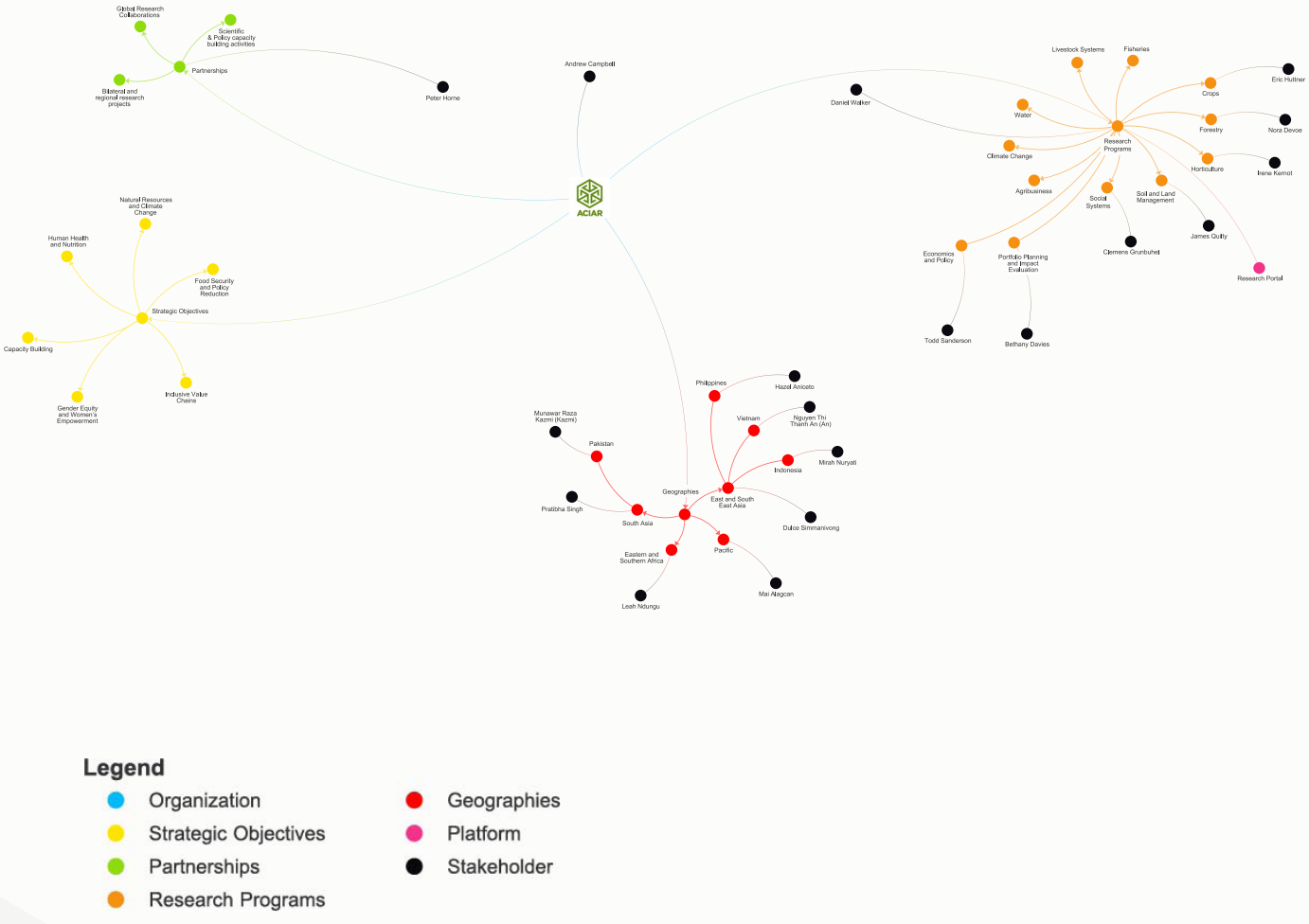
Pacific	East and South-East Asia	South Asia	Eastern and Southern Africa
• **Fiji** • Kiribati • **Papua New Guinea** • Samoa • Solomon Islands • Timor-Leste • Tonga • Vanuatu	• Cambodia • **China** • **Indonesia** • **Laos** • Myanmar • **Philippines** • **Vietnam**	• Bangladesh • **India** • Nepal • **Pakistan** • Sri Lanka	• Burundi • Ethiopia • **Kenya** • Malawi • Mozambique • Rwanda • South Africa • Tanzania • Uganda • Zambia • Zimbabwe

*Countries highlighted in bold are ACIAR Country Offices

Partnerships and Engagement

Area of partnerships and engagement	Partners and stakeholders
ACIAR **engage** with **experts and governments** to ensure our work is attuned to the needs of our partner countries, embracing new methods and in step with Australian Government policy	1. Australian Government – Department of Foreign Affairs and Trade 2. Commission for International Agricultural Research 3. Policy Advisory Council
ACIAR **establish partnerships** with **Australian and international organisations** to jointly address shared priorities in agricultural research-for-development	1. Australian Government departments, universities 2. International research-for-development agencies 3. International agricultural research centres 4. Private-sector and not-for-profit foundations
ACIAR **commission technical experts** to implement our research portfolio and capacity building programs, according to well-defined priorities and well-developed project scopes	1. Universities 2. CSIRO 3. State and federal government departments 4. Private-sector specialists and consultants
ACIAR **facilitate** the delivery of new knowledge and technologies via our **commissioned research partners**	1. Smallholder farmers in partner countries 2. Small and medium community-based enterprises in partner countries 3. Australian agricultural industries and farmers
ACIAR c**ommunicate** project news and results to extend the reach of our work, and increase our understanding of the impact of Australia's investment in agricultural research-for-development	1. Project and program partners 2. Industry peak bodies 3. Australian public 4. International public

ACIAR overview



Legend

- Organization
- Strategic Objectives
- Partnerships
- Research Programs
- Geographies
- Platform
- Stakeholder

Figure 13: Ecosystem map (link: https://tinyurl.com/SLAM-2021-156-annex-fig-13)

Annex 3 Examples of the Evolution of Scientific Data Sharing

How Has Data Sharing Evolved in Global Science Programs

The sharing of data occurs when the outcomes of this sharing produce recognised benefits and achieve specific goals. Within science most research follows a common set of processes. A theory is contemplated, data collected and analysed to prove (or disprove) that theory. A scientific paper is often the main outcome of the research activity and while the data is a critical element of this process it is not usually the main objective. When scientists work alone, the issue of data sharing is not considered a major factor in the research process. As research projects look at larger, more complex issues, and research becomes a team effort, the need for data sharing begins to become an increasing factor in generating a successful research outcome. To ensure effective data sharing, agreement must be reached on common data management protocols for the research project.

An example of how sharing of data becomes important and provides substantial benefits to the contributors is demonstrated in the field of oceanography. Prior to the mid 1980's oceanographers around the world were undertaking studies into various aspects of the ocean's physical structures, including ocean dynamics. In Australia, studies occurred with scientists looking at major ocean currents, such as the Eastern Australian Current (EAC). Initially, research was undertaken by individual scientists with minimal data sharing. As knowledge of the EAC developed, scientists recognised that larger, more complex studies were required involving multiple scientists. The scientific benefits of sharing individual data was then seen as significantly outweighing any negative impacts arising from data sharing.

The international oceanographic community understood the interconnectedness of the oceans, and that one small research program could not answer local, regional, or global scale questions. It became obvious that the sharing of data on a global scale would bring massive scientific benefit. In March 1991 the Intergovernmental Oceanographic Commission (IOC) of UNESCO established the Global Ocean Observing System (GOOS)[65].

GOOS was built on the original efforts of individual researchers, who then collaborated at the national level and finally collaborated internationally, to collect and share data resulting in access to extensive global datasets that could be used by hundreds of ocean and climate scientists. The significant benefits arising from data sharing were recognised by this sector of the science community. Those participating may have had different goals at local, regional, or global levels but they all gained through access to larger datasets than they could collect individually. The benefits of sharing were obvious for marine scientists and also benefited governments and the general community.

In a similar series of steps, climate scientists also worked initially at the individual level and then moved to collaboration at national, regional, and then global levels. This led to the creation of the Global Climate Observing System (GCOS) by the World Meteorological Organisation (WMO) in 1992[66]. Because of this sharing GCOS and GOOS scientists had access to consistent, quality data in a usable form and could undertake small scale, local, regional, or global research activities leveraging the data available through both international programs. Scientists were able to move from the analysis of hundreds of observations they could collect themselves to hundreds of thousands of observations available through these data sharing programs, massively improving their research outcomes.

These global scientific programs provided huge gains for the scientific community who greatly benefited from this data sharing. The programs leveraged the relationships developed between scientists, governments, and the international bodies. To indicate the power and significance of these relationships, many Navies around the world (including the Navy's of Australia, Russia, the UK and the US) contributed oceanographic data to both GOOS and GCOS. The potential consequences of national security were seen to be easily outweighed by the benefits coming from data sharing.

[65] https://goosocean.org/index.php?option=com_content&view=article&id=296&Itemid=427

[66] https://gcos.wmo.int/en/about/gcos-story

The sharing of data at this scale required significant work, building on the relationships between the contributing scientists, their research organisations, and their countries. It required trust and agreement to create the data sharing mechanisms, and this resulted in the development of a range of data standards and sharing protocols. Significant resources and effort went into developing these protocols and while the individual goals of the participants may have been different around the world everyone recognised the substantial benefits from creating these agreements.

The scientific community across a wide range of disciplines has and continues to play a leading role in the development of data sharing.

Data Sharing Exemplars - World Meteorological Organisation (WMO)

For over 70 years the WMO has been making weather observations collected by member countries available to all other member countries. At the beginning of this data sharing journey much of the data was collected for scientific purposes. A major reason for beginning to share these weather observations is because no one country was able to collect sufficient data to make accurate weather forecasts or undertake meteorological research. The sharing of data was seen 70 years ago as a significant mechanism to improve both national weather services and weather-related research. Recently the WMO has created the Unified Data Policy[67], which is an update of the original data sharing policies. The Unified Data Policy says that:

> *"The last decades have seen explosive growth in the demand for weather, climate and water monitoring and prediction data to support essential services needed by all sectors of society, as they face issues such as climate change, increasing frequency and impact of extreme weather, and implications for food security.*
>
> *The free and unrestricted exchange of observational data from all parts of the world and of other data products among all WMO Members must be updated and strengthened to accommodate this growing demand."*

Data Sharing Exemplars - USA National Oceanic and Atmospheric Administration (NOAA)

Building on the WMO concept of sharing data NOAA has recognised the considerable national benefits arising from the concept of 'open data'. In regard to NOAA's data sharing activities the Open Data Institute has described the benefits in the following way:

> *Since its inception, NOAA has boasted a strong open data culture and is considered a leader in open data, if not the leading open data example among government agencies. When the Obama administration launched data.gov as part of its flagship Open Government Initiative in January 2009, NOAA was cited as the paradigmatic example as to how government agencies can both publish data and make that data accessible for the private sector to use and build a multi-billion-dollar industry.*[68]
>
> *Opening up weather data through the United States National Oceanic and Atmospheric Administration (NOAA) has significantly lowered the economic and human costs of weather-related damage through more accurate forecasts; the development of a multi-billion-dollar weather derivatives financial industry; and the growth of a million-dollar industry of tools and applications derived from NOAA's real- time data. In many ways, the industry built around NOAA's weather data is seen as the paradigmatic example of how the release of open data can yield major economic impacts.*[69]

Some specific economic benefits resulting from NOAA's open data policy[70] are given below:

- NOAA real-time data supplies a burgeoning private weather service industry with well over $700 million in value added annually
- US's $8-10 billion and growing annual Weather Derivatives financial industry relies on NOAA's seasonal weather data and records

[67] https://public.wmo.int/en/our-mandate/what-we-do/observations/Unified-WMO-Data-Policy-Resolution

[68] https://odimpact.org/case-united-states-noaa-opening-up-global-weather-data-in-collaboration-with-businesses.html

[69] https://odimpact.org/files/case-studies-noaa.pdf

[70] https://www.oecd.org/sti/ieconomy/40066192.pdf

Annex 4 FAIR Principles

Background to FAIR

Below are some details regarding the formation of the FAIR principles, its application within the community and also CABI's views and experience with the application of FAIR. The concept of Findable, Accessible, Interoperable, Reusable (FAIR) has been developed and endorsed through the collaboration of a diverse set of stakeholders—representing academia, industry, funding agencies, and scholarly publishers. FAIR consists of a concise and measurable set of principles that are referred to as the FAIR Data Principles. Data in this context refers not only to numerical datasets, but also to social science and qualitative data, where it becomes even more essential for contextual information and narrative perspectives need to be captured.

FAIR Data Principles

The FAIR Data Principles were developed by the Dutch Techcentre for the Life Sciences (DTL) and the FORCE11 community. In 2016, the 'FAIR Guiding Principles for scientific data management and stewardship' were published in *Scientific Data*[71]. The authors intended to provide guidelines to improve the **Findability, Accessibility, Interoperability,** and **Reuse** of digital assets. The principles emphasise machine-actionability (i.e., the capacity of computational systems to find, access, interoperate, and reuse data with none or minimal human intervention) because humans increasingly rely on computational support to deal with data because of the increase in volume, complexity, and creation speed of data.

The principles provide a framework to guide data producers towards effective data management leading to high-quality publishing, use, and reuse of research data. Since the inception of FAIR, the UK Data Service[72], Australian Research Data Commons[73], and EU Horizon 2020[74] among others have adopted the principles to encourage transparency and reproducibility. FAIR has also been a "guiding star" for many large organisations, such as the Bill & Melinda Gates Foundation, for how they conceptualize best data practices.

Detailed technical definitions of FAIR can often discourage initiatives to implement FAIR, or project participants can believe that FAIR cannot or does not apply to them if they already have highly technical systems. Therefore, CABI routinely uses a broader FAIR definition to guide its work:

Data should be **Findable**	If the implementing partner has a data requirement and the same or similar data has already been created or collected, then they should be able to find out that it exists.
Data should be **Accessible**	Once the implementing partner has found out that the data they need exists, they should be able to access it so that they do not have to re-create or re-collect it.
Data should be **Interoperable**	When the implementing partner accesses the data, they should be able to easily use it, so it needs to be in a format that can be readily used and, if necessary, align with other data.
Data should be **Reusable**	Once the implementing partner has the data they need, it should be clear to them what they can and cannot do with it.

[71] Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

[72] https://ukdataservice.ac.uk/about/

[73] https://ardc.edu.au/about_us/policies-and-guidelines/fair-policy-ardc-or-ardc-co-invested-materials/

[74] European Commission, Directorate-General for Research and Innovation, Horizon Europe, open science : early knowledge and data sharing, and open collaboration, 2021, https://data.europa.eu/doi/10.2777/18252

CABI's experience, from working to embed FAIR with the Gates Foundation and building FAIR policy for the CGIAR, has shown how FAIR frameworks are more widely accepted by stakeholders than a simple exhortation to make data 'open', which can seem daunting.

The FAIR Principles:[75]

The Go-FAIR organisation considers that FAIR:

> *"emphasise machine-actionability (i.e., the capacity of computational systems to find, access, interoperate, and reuse data with none or minimal human intervention) because humans increasingly rely on computational support to deal with data as a result of the increase in volume, complexity, and creation speed of data."*

The FAIR Guiding Principles

To be Findable:

- F1. (meta)data are assigned a globally unique and persistent identifier
- F2. data are described with rich metadata (defined by R1 below)
- F3. metadata clearly and explicitly include the identifier of the data it describes
- F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:

- A1. (meta)data are retrievable by their identifier using a standardized communications protocol
- A1.1 the protocol is open, free, and universally implementable
- A1.2 the protocol allows for an authentication and authorization procedure, where necessary
- A2. metadata are accessible, even when the data are no longer available

To be Interoperable:

- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles
- I3. (meta)data include qualified references to other (meta)data

To be Reusable:

- R1. meta(data) are richly described with a plurality of accurate and relevant attributes
- R1.1. (meta)data are released with a clear and accessible data usage license
- R1.2. (meta)data are associated with detailed provenance
- R1.3. (meta)data meet domain-relevant community standards

[75] https://www.go-fair.org/fair-principles/

Annex 5 ACIAR Interviews - Detailed Analysis

Introduction

Interviews and discussions have been held with a wide range of 'groups' within ACIAR and with representatives from some of its investment projects. These interviews have been aimed at gaining evidence of the understanding of the perspectives of different levels of management, and what data management and FAIR means to them. The interviews have also helped in determining the present status of data management and FAIR implementation within ACIAR projects.

The interviews had similar themes across the different cohorts and covered the respondents':

- perception of the current data management situation;
- concerns regarding data management (in their programs) and barriers impacting efforts to improve it;
- views on the need to, and interest in, improving data management activities; and
- what incentives could be used to increase interest and application of data management and FAIR principles.

A summary has also been extracted from the interviews demonstrating the priorities in the interview responses to the issue of data utility. These are provided for each organisational group interviewed and with an overall set of priorities that cover all those interviewed.

The interview summaries have been divided into the following organisational groups:

- ACIAR Executives
- RPMs
- Regional and Country Project Managers
- Focussed Project (Deep Dive) Respondents

As considered appropriate, some comments have been made throughout these interview summaries to provide both context and a perspective of good data management and data governance principles.

Additionally, where constraints to implementing FAIR principles and constraints around good data management practises have been identified, some high-level solution pathways have also been included.

Executive Interviews

The following Executives were interviewed:

- Andrew Campbell — CEO
- Daniel Walker — Chief Scientist
- Peter Horne — General Manager Country Partnerships

Perspective

Internally, ACIAR is seen as a research funding organisation and not as an agency that undertakes development. ACIAR executives are of the view that while research is funded through their programs it is not the principal role of their investment projects to correct any data management issues that may exist in recipient organisations and countries.

The Executive consider that it is the national partners role to provide effective data management capabilities for the research investments. In addition to this view of data management responsibilities it was thought that researchers are generally on strict timeframes and could be diverted from delivering the project reports if they were required to support an increase in data management responsibilities. A view was also expressed that acknowledges the perceived lack of direct value from implementing improved

data management within the individual projects. It was felt by some that improved data management is unlikely to benefit the actual research project. There was acknowledgment that the greatest benefit from effective data management was delivered to external users and other related projects, not the project itself.

A counter view regarding data management was also expressed suggesting that improving data management is not difficult and could be achieved with better planning. Also, there were views that the concept of extra time being required to undertake effective data management should not be considered a valid argument for not improving the management of project data to improve data utility.

The Executive did not raise or explicitly acknowledge the principles of FAIR or the need to have FAIR principles adopted within ACIAR projects.

> ***Comment:*** *The view that further investment in data management does not benefit some projects has some partial validity. When research activities are small with only a limited number of participants and limited data is collected then the researchers on such a project are likely to know where the data is, how it was collected and what it means. The data can be accessed 'easily', if not formally managed. However, with larger projects where data is collected by numerous researchers and perhaps across different regions and/or by different organisations, effective data management does provide direct project benefit since data will need to be shared and exchanged within the project. If that data is well managed it saves considerable time and effort when processing or analysing the data to meet the projects research objectives.*
>
> *As a broad generalisation, more benefit from effective data management is likely to realised outside the projects themselves. The benefits are achieved through the data's future reuse and may add value to a future project investigating similar things.*

Executive Concerns

The ACIAR Executive expressed a range of concerns around data management. These included concerns relating to increasing the burden on projects if additional data management activities were introduced as this was felt, if data management was not adequately resourced, to have potential to slow research progress. Additionally, there was concern that were ACIAR to impose tight data management requirements within projects with multiple partners this would pass an additional burden to project managers to ensure (or enforce) data management protocols on those collecting the data.

A common concern across the executive related to issues for the researchers around privacy and security especially if the data is 'released' prior to publication. It was considered that such a release of data may (if not done with due care) result in the data being exploited for commercial or other scientific purposes. The Executive indicated that these types of concerns have also been raised by country partners.

One specific and very valid concern was raised about what happens when researchers leave a project. The potential for data to be lost if it has not been effectively managed and easily accessible to the project could result in significant issues for the project and the effectiveness of the research outcome.

There was a broad view of the lack of direct benefit to many research projects arising from effective data management during the lifetime of the project. It was understood that there was longer term value coming from effective data management but that this would be realised outside the specific project and often only realised in the future (often when funding had been pulled and when impacts were no longer being tracked). In addition, it was felt that due to the 'long term nature' of the benefits from effective data management other stakeholders may not be willing to participate in improved data management aspects of the project, because the immediate benefits may not be apparent or be hard to articulate.

The Executive also expressed concerned over a lack of clarity about the Australian Governments views and expectations around data management needs. *(Refer to Section 7 - Context for Recommendations that provides a brief view of the Australian Government's Data Policy's and legislation).*

Additionally, it was acknowledged that there is nothing in the present project Contracts that inform, obligate or guide the researchers in respect to managing or sharing their data.

Resolving Concerns

Some of the concerns expressed by the Executive can be relatively easily resolved. The specific concerns and some basic approaches for how these concerns can be addressed are provided below.

Issue: **Data management increases the burden on researchers.**

> ***Response:*** Ensure that data management is included as part of the research project proposal and the investment contract so it is costed, time is allocated to undertake data management and it is resourced effectively.

Issue: **It is difficult for CRMs and others to enforce strict data management protocols.**

> ***Response:*** As above, ensuring resources are available and time allocated in the project plan for data management together with education/training, tools and other resources to assist the researcher will make the 'enforcement' task less difficult.

Issue: **Concern about the privacy and security of data if released.**

> ***Response:*** Some data should not be released for genuine security and privacy reasons and data that is 'releasable' does not need to be released until the project has been completed. However, all datasets should have basic descriptions (metadata) available so that this can be found through appropriate search engines. Also, adopting FAIR principles, rather than Open Data concepts, means data access can be controlled.

Issue: **When researchers leave a project, they may take the data with them, or leave it at a location such that it is difficult to find.**

> ***Response:*** Effective data management with a focus on ensuring that project data is locatable (and not just held on a specific individual's computer) will reduce this impact.

Issue: **Limited direct benefit to the individual researcher from improving data management.**

> ***Response:*** In some cases, this concern is valid. However, in projects involving multiple researchers and organisations, data sharing is often essential to ensure the best outcome from the project. The ability to effectively share the data without adding to other researchers' burden, can only be achieved when data is well managed. Additionally, fully FAIR compliant data can be easily used by others in the future for follow on or similar studies, reducing their work burden. Releasing well-managed data (FAIR compliant) has potential social and economic benefits to the local and (possibly) global community.

Another important approach to resolving these various issues is through effecting a cultural change and attitude towards the benefits of data management. This is something that could be established by providing basic training in data management. The concept of data governance should be something understood at all levels of ACIAR, and its significance reflected in ongoing conversations with all major investment participants. An attitude that data management adds value coming from 'on top' will help change the culture and perspectives of those below, and of those participating in the investment projects.

Part of creating this cultural change can be through the regular provision of and access to up-to-date data management training material and related resources. Additionally communicating the benefits arising from data sharing by ACIAR projects and identifying data sharing success stories will also assist in gradually changing organisational culture.

Perceived needs for data management in the Executive

There was some general agreement across the Executive on the need for data management and how data management activities within the projects could be improved. Suggestions included:

- Building trust in country partnerships to facilitate improved data management and data sharing.
- Developing a cost/benefit understanding of effective (FAIR) data management based on evidence and show relevant executives the value arising from effective data management.
- Working with Australian stakeholders and partners to increase interest in ACIAR projects data management activities.
- Better communication of the benefits and value of data management with national partners.
- Including new approaches to developing data management strategies with country partners.
- Leveraging existing data management structures (e.g., Australian National Soils Strategy).
- Improve ACIAR's understanding of the pushbacks and the 'mood' of stakeholders in relation to investing in improved data management.
- Include data management as a section within ACIAR PPP's and Project Contracts.
- Provide data management support (training, resources and a Data Concierge) in the project development phase.

This list of perceived needs would suggest that the ACIAR Executive does see value in improving data management within their projects but what appears to be lacking is agreement on a way forward, and specifically a plan for how to achieve this given the numerous barriers that exist.

Views do differ somewhat amongst the executive and responses depended on exactly how questions were framed in the interviews. Some went as far to suggest that because ACIAR principally focusses on funding research that data management improvements are considered to be the concerns only for those undertaking the research. Data management capacity building and 'development' more generally and are not considered by some to be a primary responsibility of ACIAR.

> ***Comment:*** *There appears to be differing views across the Executive on what constitutes data management. It would be important for ACIAR moving forward to have a clear, commonly agreed stance on what data management means both within the organisation, and outside to those it defines as key stakeholders.*

Incentives

There were several views and suggestions expressed regarding likely incentives to improve data management within ACIAR and its funded projects. Consensus was greatest around the approach to incentives and an understanding that the incentives needed to focus on predominantly social or cultural levers.

A 'big stick' approach that might involve explicitly mandating data management procedures, was not seen as useful. Several suggestions regarding potentially effective incentives were made; some of these suggestions could be implemented through a data management education process. Such a process would highlight the benefits of effective project data management and provide some guidance on how improvements could be made.

Incentive suggestions included:

- Informing the Australian partners that the research is funded through public money and therefore there was a certain obligation regarding the data collected and its availability to others (refer to the Australian Government's Data Policy's View on Data – Section 5.1)
- Show how teams that exchange data with, and received data from, other projects have improved their research outcomes.
- Engage with early career scientists who may be more open to data science, sharing and co-creative approaches with a view to building out champions for effective data management.
- Encourage the use of data standards rather than mandating data management activities.

Data Utility Issue Responses

A qualitative analysis of the responses from the Executives to the interviews regarding their concerns about constraints to data utility are given in the graph below.

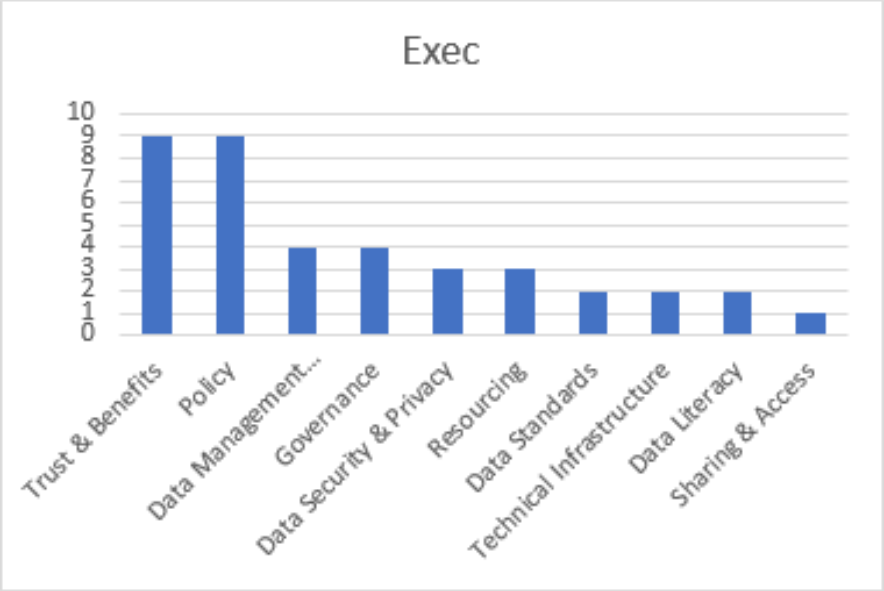


Figure 14: Count of Responses by Executives to Data Utility Issues

Comment: *The order of priority of these areas of concern reflects the strategic and 'big picture' perspective of the Executives' views of data management within the context of ACIAR and its research investments. Issues such as trust and policy are of concern at the executive level, while an understanding of things like data standards, technical infrastructures and data literacy would not be considered as high from their strategic perspective.*

Annex 6 Interviews with Regional Project Managers

The following Managers were interviewed

• James Quilty	Research Programme Manager	Soil & Land Management
• Eric Huttner	Research Programme Manager	Crops
• Irene Kernot	Research Programme Manager	Horticulture
• Nora Devoe	Research Programme Manager	Forestry
• Clemens Grunbuhel	Research Programme Manager	Social Sciences
• Bethany Davies	Programme Manager	Portfolio Planning & Impact Evaluation
• Todd Sanderson	Programme Manager	Economics & Policy

Background

The same topics of discussions and questions that were used in the interviews of ACIAR executives formed the basis of the discussions held with the ACIAR Regional Project Managers RPM). An analysis of these discussions is provided below. As would be expected, the RPM's, while still having a strategic understanding, are coming at the issues of data management and FAIR with a more focussed and project-oriented perspective than the strategic level views of ACIAR Executives.

Perspective

The RPMs indicated that many of the investment projects included social science objectives, resulting in more qualitative (rather than quantitative) data being collected. Consequently, the sharing of such data was seen as being more difficult because the social data needed a significant amount of contextualisation to be of value to other projects. Nonetheless, it was felt that the social science related data could provide valuable insights into a range of human dimensions and therefore provide value to other relevant projects. In other words, the 'social science' data would provide benefits through adopting FAIR principles where it could be effectively leveraged in other projects.

The RPMs feel that they were dealing with large portfolios of projects and so for them to add more focus on project data management may be problematic from a resource and workload perspective. In conjunction with this, it was felt that some of the ACIAR investment projects were not data-rich and therefore the benefits gained from improved data management may not provide any significant benefit or improve the projects' ROI. RPMs suggested that while they have a reasonable level of influence over the quality of science within projects, they did not have any direct control on how project data is collected or managed.

One view provided was that it is probably better for the outcome of a project to work with smaller amounts of quality data than large quantities of poor-quality data. It was also pointed out that data is generally not provided as a research outcome other than perhaps in an appendix of a project report, although this is not common. Based on the RPMs' experience they suggested that it was the final project reports that were asked to be shared but not any raw project data. RPMs indicated that they could influence projects but had no direct decision vis a vis inputs to the allocation of funds and so suggesting resources be expended on improving data management would be difficult. There was also a view that conversations around each project were more about project design than on data and its management. There appears to be a broad view across the RPMs that rigid implementation of FAIR and FAIR principles would be hard and that attempting to do so would be 'painful' to the projects as well as frustrating for the researchers.

Overall, the concept of FAIR was not generally acknowledged in the context of ACIAR investment projects, and the general views were that while data management could provide some benefits, having a greater focus on improving data management may well hinder the outcomes of projects through increased workload if project budgets were not increased.

RPM Concerns

The RPMs raised a considerable number of concerns around the present state of data management within ACIAR projects. These included:

- Knowledge and data management is recognised as a large challenge within ACIAR and its projects given the lack of available resources to help understand and clean project data.
- There is a lack of a clear data management policy or even guidelines for projects to follow to improve project data management.
- RPMs suggested that there are considerable concerns within the projects regarding data privacy and data security which results in project personnel showing a lack of interest in or ability to share project data.
- A lack of an information or data infrastructure to even share project reports easily let alone project data is seen as constituting a significant barrier.
- ACIAR does not have access to project-level raw data and there are institutional barriers preventing data access due to the high level of dependence on individual institutions hosting project data.
- There is a great variability of data management capability and interest in data management within project host institutions. Those with limited or no data management capabilities results in project data being managed by individual researchers (and therefore is prone to getting lost over time).
- The RPMs felt that once a project had been completed, access to data becomes dependent on good will and personal contacts.
- Data can easily be lost when key project people leave the project.
- The lack of a data management policy and the breadth of research project topics results in multiple data formats being used within each project and this significantly complicates the ability of a project to effectively share data with reuse in mind.

Comment: *The lack of data management policy, structure, agreed processes and guidelines means that the researchers and the host institutions go about data management in a manner that best suits them (and not necessarily donors or other partners). There are no formal or informal rules governing the projects to make data accessible and reusable. Additionally, there are no mechanisms for training in data management or guidance in relation to privacy and confidentiality of data. This lack of clarity for data management is a major barrier to FAIR principles within ACIAR investment projects. However, this can only be addressed through changing the research culture, education and providing appropriate tools together with the creation of suitable data management policies and guidelines.*

Resolving Concerns

Concern: **The challenge of data and knowledge management given the lack of resources.**

Resolution: Ensure that data management is included as part of the research project proposal and the investment contract, so it is costed, and time is allocated to undertake data management and it is also resourced effectively.

Concern: **A lack of data management policy and guidelines.**

Resolution: The creation of an ACIAR Data Strategy.

Concern: **Data privacy and security.**

Resolution: Some data should not be released for genuine security and privacy reasons and data that is 'releasable' does not need to be released until the project has been completed. However, all datasets should have basic descriptions (metadata) available so they can be found through appropriate search engines and potential users can be aware of its existence.

Concern: **A lack of capability to share project reports and project data.**

Resolution: The establishment of a metadata capability that can provide links to project reports and also, where available the actual project datasets.

Concern: **ACIAR not having access to project level data due to host institutions data access barriers.**

Resolution: Ensuring data management and FAIR concepts are included in the investment contract. Additionally, work with the host institutes to determine the reasons for withholding data access. It may be as simple as their lack of capability or understanding of the benefits and value of sharing data.

Concern: **The variability in data management capabilities and interest in data sharing across project host institutions.**

Resolution: Similar responses to the previous concern and one addressed through communication and also education about the value of data management and data sharing.

Concern: **Once a project has been completed, access to data becomes dependant on good will.**

Resolution: Including data access as part of the investment contract and providing education to demonstrate the value proposition to other researchers and the community of releasing data.

Concern: **Project data can be lost when a researcher leaves.**

Resolution: The addition of project management policies and procedures within ACIAR in general and in investment project contracts in particular to ensure data is effectively managed.

Concern: **A lack of a data policy and the complexity arising from the breadth of research topics resulting in multiple formats being used.**

Resolution: Developing an ACIAR Data Strategy and ensuring that data management is included in the project proposal and investment contract and ensure the issues of data sharing have been resolved to an adequate degree prior to project commencement.

Comment: *Another important approach to resolving many of these various issues is through ACIAR having a Data Strategy, that sets out a Vision for how project data can be leveraged to benefit other projects and a wide range of external users. Supporting national SDG measurement for example. A Data Strategy will also help with the required 'cultural' change that can be further supported through the provision of training in data management. The concept of data governance should be something understood at all levels of ACIAR and its significance reflected in all relevant on-going conversations with all investment participants. An attitude that data management adds value should come from 'on top'. This will help gradually change the culture and perspectives of those below, and of those participating in the investment projects.*

Needs

The concepts of policy, infrastructure, guidelines and rules as well as incentives to underpin and inform data management activities in projects are all seen as important needs. Likewise, there is a perceived need for suitable tools for the researchers and perhaps a centralised infrastructure for at least the capture and sharing of project metadata. In this way at a minimum the presence of a description of project data and details of how it was collected, descriptions of the data itself and where it is located would provide a significantly greater level of knowledge about ACIAR project data and could go some way to facilitating reuse of this data.

The RPMs felt that a much higher level of awareness of both the principles of data management and the benefits of effective data management by all project stakeholders was needed. The building of this awareness through education and other support mechanisms should be commencing at project design stages with oversight, support and input from ACIAR and RPMs during project design and while the project is operating. Adding a data management element and data sharing agreement into the initial project proposal would also increase the awareness of ACIAR's view of the importance of data management.

Incentives

The RPMs provided some valuable thoughts in relation to incentives for implementing FAIR principals. These incentives revolved around two main themes.

The first covered the concept of rewarding, acknowledging and supporting those data collectors that create good quality data. It was felt that recognition of the efforts of those researchers that applied effective data management in their projects should be acknowledged in some manner. There was also a view that joint workshops could be held between past and present teams to ensure knowledge transfer and ideally this would include a component regarding the benefits of data sharing.

The second area of providing incentives revolved around processes and procedures, with concepts such as mandating data sharing, using project milestones and resourcing as mechanisms for improving data management compliance. It was pointed out that the project data is collected using public funds and therefore there was a responsibility that the data should be made available (refer section on Australian Government's View on Data– Section 5.1). There was also a view that well-designed project milestones covering data management and together with the provision of relevant data management tools and adequate resourcing would provide benefits and help move ACIAR investments towards FAIR compliance.

The RPMs indicated that they are very busy and would require some form of increased capacity to monitor any increased emphasis on data management within projects.

Data Utility Issue Responses

A qualitative analysis of the RPMs responses to the interviews regarding their concerns about constraints to data utility are given in the graph below.

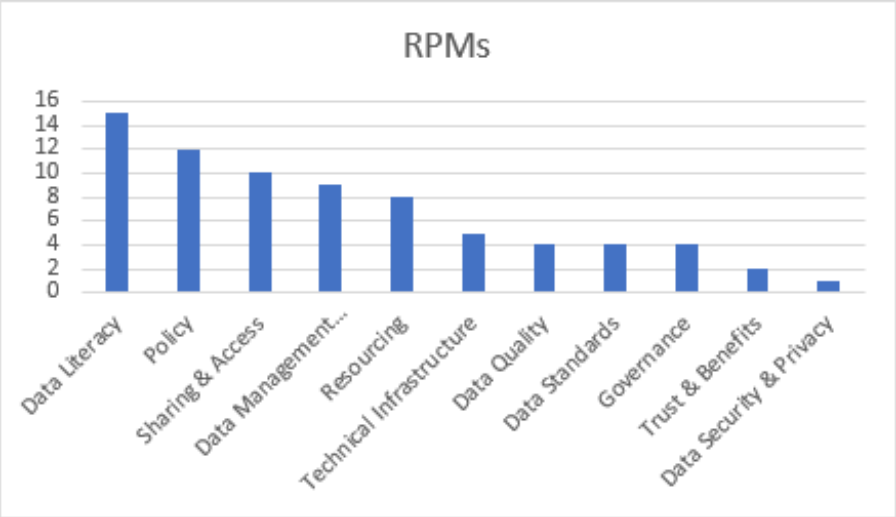


Figure 15: Count of Responses by RPMs to Data Utility Issues

Comment: *The RPMs had slightly different views from the Executives on the priority issues around data utility than the Executives. The RPMs are closer to the research activities and have a more project focussed view that enables them to be able to identify the lack of data management literacy within projects. Not having an ACIAR Data Strategy to 'govern' data management activities was commonly seen as a barrier and as a consequence the sharing and access to data was seen as limiting a projects potential. Interestingly, the concept of trust and benefits together with data security measured lower from the RPM's perspective. This is likely due to their perspective which is closer to the projects and they can see examples of the level of trust within projects.*

Annex 7 Discussions with In-country Networks (Country and Regional Managers)

The following Country and Regional Mangers were interviewed:

• Nguyen Thi Thanh An	Country Network	Vietnam (An)
• Dulce Simmanivong	Regional Manager	Cambodia, Lao PDR, Myanmar, Thailand
• Mirah Nuryati	Country Network	Indonesia
• Pratibha Singh	Regional Manager	South Asia
• Munawar Raza Kazmi	Country Network	Pakistan (Kazmi
• Mai Alagcan	Regional Manager	Pacific & PNG
• Hazel Aniceto	Country Network	Philippines

Perspective

As would possibly be expected the responses from those closest to the projects and the in-county networks had a more focussed perspectives on data management activities.

Several respondents recognised the value of data and the benefits of its effective management. They understood how access to good data could and should promote its use to inform country stakeholders and for policy formation in participating countries. Data was seen as being useful as potential input to future projects.

An interesting comment concerned the need for having appropriate contextual information to accompany the data, although there was the view that perhaps project reports may normally provide some of this context. The need for contextual support was considered more important for projects collecting social (qualitative) data.

A need to have projects and the resulting data aligned with country needs was identified as being significant. There was also a recognition that generally project data was not easily accessible since there were often no specific project portals where data was held and could be accessed. The data skillsets of participants in the project team was also seen as a key element that determined the likelihood of the development of good data management practices. It was suggested that there needed to be a balance between project data requirements and its subsequent management and accessibility of project data and the countries' policy around data accessibility.

Country and Regional Managers Concerns

The respondents raised a number of concerns regarding project data and its management. These included a level of confusion about how to find existing data and a general view that most projects do not have good data management practices. Also, there appears to be little consistency in project data management, with each project going their own way. It was felt that data/information is sometimes 'confined' to one person and therefore access to data becomes a personal interaction. A key issue raised was that of different languages used to describe data which can create difficulties when preparing country strategies and project reports. Language can often provide a barrier for the exchange and reuse of data.

While it is viewed that data is often shared between project team members, it was not more widely shared. This is due to data requiring a certain level of processing to make it suitable for external use. Other concerns were raised that related to the difficulty of getting projects to effectively manage their data and also the effort required to do this to ensure the data was suitable for reuse. This 'additional effort' resulted in considerable negotiations to establish agreements to ensure data could be shared. Another area of concern coming from potential project partners was about the misuse of data if it was shared, even within the project team.

The concerns raised indicate several significant barriers to implementing more effective data management and the FAIR principles and also indicates some smaller but possibly equally important barriers.

Resolving Concerns

Concern: **There is difficulty in knowing where to look for data.**

> ***Resolution:*** An effective data management environment that includes a metadata repository will provide a record and appropriate details of data collected during ACIAR investment projects, where this data is located and sufficient information to determine if it is fit for purpose for another project.

Concern: **There is little consistency in how projects manage data, with a number of projects have poor data management practices.**

> ***Resolution:*** Provide an ACIAR Data Strategy and guidelines to build greater consistency between and within projects. Provide data management training, tools and suitable resources to undertake more effective data management activities in each project.

Concern: **Data management can be confined to a single person and so access to data becomes a personal interaction.**

> ***Resolution:*** Strengthening data management procedures, implementing metadata capabilities and ensuring a single or central project data repository for each project. Identify where this will be held as part of the initial project negotiations.

Concern: **The impact of projects that cover more than one country can result in issues of language, especially around describing data.**

> ***Resolution:*** Developing data management plans for each project and ensuring that an agreed glossary of terms, especially those describing data be established in the project planning stage. Such a glossary could be translated to the languages within participating countries

Concern: **Getting data ready to share requires processing and therefore resources to ensure effective interoperability and reuse**.

> ***Resolution:*** An effective project data management plan that includes the necessary level of 'processing' data (which generally is provided by having effective metadata) to make it suitable for sharing and including the resources to get the data to a sharable state.

Concern: **A general lack of data management activities within the projects is a consequence of a lack of resources.**

> ***Resolution:*** A data management plan supporting data sharing and adequate resources identified in the project contract to ensure the data can be easily shared.

Concern: **Potential misuse of data if shared.**

> ***Resolution:*** Determining what data will be collected during the project, identifying any data that may legitimately have restrictions placed on its release (privacy issues etc) and documenting this in the project plan so any legitimate restrictions for exchange are agreed to upfront.

Needs

Interviews with in-country Networks (Country and Regional Managers) raised a large number of perceived needs in relation to FAIR data principles. Most of those interviewed agreed that the general concept of FAIR was understood but they provided a large number of reasons why it was not implemented. One comment did suggest that FAIR was a familiar concept but had never been used.

Many of the 'needs' expressed covered two major and related themes. The first relates to having a suitable data strategy and appropriate data management infrastructure in place and the second need related to having adequate resources (and time) to support effective project data management activities. Several respondents suggested the need for suitable data management infrastructures to be available so authenticated data can be easily accessed. Additionally, there was also a need for metadata systems to provide an accessible way to determine what data existed and where to find this data.

The concept of having a 'data champion' within each institution was expressed as was the need to build trust between researchers and institutions to help facilitate data sharing. Project partners should have an awareness of any in country data-related legislation or policies to ensure they comply appropriately with any requirements.

It was felt that data needed to be processed in different ways to be suitable for different project stakeholders. For example, policy users would need a different perspective of the data than researchers. While this may relate more to the final research report, it does indicate that any project data needs to be well documented to ensure it is fit for purpose. This will help potential users of this data (or information) to determine if the material will meet their needs. These comments also indicate a misunderstanding of data vs information. It is not considered appropriate to process data to meet the needs of a range potential re-users. Instead, the data should be adequately described to enable different stakeholders to evaluate if the data is 'fit for purpose' to meet their needs and determine if further processing or integration of the data is required.

Aligned to this was a view that there should be a 'dedicated' section for communicating and reporting to meet specific country network needs. Again, this view is probably more an information (research report) outcome rather than concerning project data, but it does reflect the general idea of supporting the concept of 'fit for purpose'.

While the issue of 'language' itself wasn't raised as a need, the concept of having an agreed naming convention for files was. This is not dissimilar to the concept raised above in the section on Concerns where a standardised glossary was discussed. A common naming convention for files and also for data elements is considered as a component of an effective data management framework. The sharing of data is greatly facilitated by having common naming conventions which also provides information on what the data is actually describing.

Incentives

The interviews with the in-country Networks - Country and Regional Managers resulted in only a few thoughts on incentives to improve data management. The views that were expressed covered the concept of 'cross-learning' between different projects through workshops to spread and develop ideas and data management improvements. It was also seen that an incentive for improved data management would be to clearly articulate the benefits that would flow to the development of in-country policies.

Data Utility Issue Responses

A qualitative analysis of the Regional and Country Managers responses to the interviews regarding their concerns about constraints to data utility are given in the graph below.

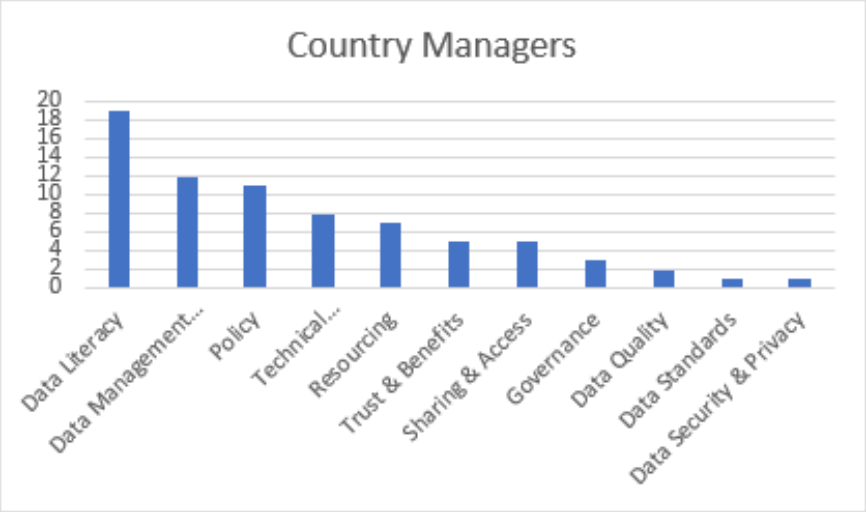


Figure 16: Count of Responses by Country Managers to Data Utility Issues

Annex 8 Deep Dive Review of Three Data Rich Investments

Summary of the Deep Dives

The outcomes from the analysis of the discussions and semi-structured interviews with three specific investments are documented below. The approach undertaken and level of detail investigated within each deep dive reflects the general levels of data maturity that existed within each of the three projects. As with the previous interview analysis CABI has included comments to provide additional context and interpretation where appropriate.

Deep Dive Investment 1 - Land suitability assessment and site-specific soil management for Cambodian uplands – SMCN/2016/237

The focus of this investments deep dive was on the Cambodian Soils Database, a key outcome of the project and the following stakeholders were interviewed.

- Wendy Vance – Project Manager, Murdoch University
- Noel Schoknecht - Soil Scientist, Murdoch University
- Paul Galloway - Soil Surveyor, DAFWA
- Sarith Hin – Director, CARDI

Data Governance

The Cambodian Soils Database is built on a MS ACCESS platform and was initially focussed on simply being a repository for soils data held and collected by CARDI. Because of this initial focus there is a lack of data governance or data sharing agreements in place. As the project progressed and the concept of the database forming the national Cambodian Soils Database within GDA evolved, and also as other organisations began contributing soils data a need for data governance has become clear.

In-Country Expertise

The in-country participants, especially those involved in data entry and data management have limited access to the data management knowledge required to enhance the database to better meet user needs. Guidance and support are required to assist in-country teams to resolving matters concerning data sharing agreements, data ownership, database development and management and data licencing for re-use. The stakeholders indicated that they did want to improve their data management activities, but they were unsure of how to go about doing this. They suggested that advice, guidelines and training would assist them significantly.

The CARDI database was originally developed as a simple repository, but as the project developed its potential to support a wider range of users and also to transition into the national Soils Database the current limitations have become more obvious. Additionally, the CARDI database holds complex soils data but to provide effective outputs to meet users' needs necessitates considerable interpretation by soil experts. The database was not designed to provide direct access to meet user needs.

> ***Comment:*** *By thinking a little beyond the original investment scope and visualising the potential longer-term use of the soils data in Cambodia more focus on data governance and sharing could have made a difference at the commencement of the project. Also, with more support and training during the database development phase could have assisted in supporting the longer-term needs and use, improving the investments impact.*

Lack of Alignment

Understanding of the objective of the data management aspect of the project varied among stakeholders. A lack of clarity of the investments objectives has resulted in some fragmentation of the project's activities causing some duplication of effort. Additionally, if there had been an understanding of the potential of the

database to become the Cambodian Soil Database early in the project then more effort may have been provided for broader data management considerations including establishing data sharing agreements. Clarity around the most suitable soils data standards from the commencement of the project would also have assisted with improving data utility.

Key Learnings

A shared understanding of objective of the project together with a shared perspective of the potential for the database by all stakeholders could have improved the data management aspects of the investment's outcome.

> ***Comment:*** *Developing a greater level of clarity and consistency of the projects objectives across all stakeholders and developing effective data governance and sharing arrangements early in the project would help achieve an improved investment outcome. Additionally, capacity building through training and providing data management guidelines suitable for the stakeholder's data maturity level would also benefit the investment's outcomes.*

Deep Dive Investment 2 – Soil management in Pacific Islands: investigating nutrient cycling and development of the soils portal - SMCN/2016/111

The focus of deep dive for this investment was on the Pacific Soils Portal aspect of the investment and the following stakeholders were interviewed:

- James Barringer – GIS Scientist, Manaaki Whenua – Landcare Research (MWLR), New Zealand
- Vuvinesi Minoneti – Economist, Ministry of Agriculture and Food, Forests and Fisheries (MAFF), Tonga
- Ami Sharma – Principal Scientist, Ministry of Agriculture (MoA), Fiji
- Ben MacDonald – Group Leader, Process and Function, Soils and Landscape Group - CSIRO

Data Governance

A data governance mechanism was established early in the project with sharing arrangements put in place through agreements reached by high-level country representatives. This effort took considerable time but supported the data sharing processes from the commencement of the investment which has contributed to the successful establishment of the portal.

While this data governance mechanism was established the roles and responsibilities for the operation of the portal by in-country departments is still unclear. There is a limited level of understanding around who should support different aspects of the project such as data collection standards and other protocols. A relatively low level of use of the portal has resulted from the limited communication between in-country stakeholders concerning the portal and its potential benefits to a broader user community. This low level of usage also reflects the lack for consultation with the stakeholders to ensure their data needs were supported.

Longevity of the portal

The portal is currently run and managed by MWLR with an objective of moving its management to an in-country organisation, possibly the Pacific Community (SPC). This move requires political and institutional challenges to be overcome to ensure the portals sustainability. Additionally, key in-country contacts often move to other roles and would possibly take any specialised knowledge of the portal with them. The portal's longevity is seen to be dependent on having it become part of the day-to-day operations of the key stakeholders' organisations which would then reduce the dependence on individuals.

Harmonization of datasets

Existing data within the Portal was predominantly collected by MWLR and other New Zealand organisations and was easily integrated as they followed appropriate standards. Data collected from previous ACIAR investments in the region and new data collected will need to be harmonized before it can be added to the portal. This will be challenging without establishing agreed data formats and standards.

Increasing usage of the portal

There is a relatively low level of awareness of the portal and on how users could gain benefit from the data in the portal. A lack of focus on user needs when developing the portal has resulted in a lack of clarity of its purpose so it is difficult to embed the portal into the regular day-to-day operations of the stakeholder organisations.

Key Learnings

A perception of the negative impacts of having data 'open' has resulted in a phased approach for sharing of data within the portal with the initial access being a user simply viewing rather than being able to download data. This starting point is aimed at changing stakeholders' attitudes over time and encouraging them to begin sharing data with external users. Seeing the potential of data sharing through the portal and understanding how the portal works is beginning to change attitudes that may support the introduction of FAIR data principles. This would enable data contributors to control access to their data based on their individual views of releasing data.

The development of the initial data governance mechanisms at the commencement of the project has shown many benefits for the project.

The sustainability of the portal is dependent on overcoming political and institutional challenges.

> ***Comment:*** *Developing a data governance agreement at the commencement of this project has provided a range of benefits. Improvements in this could be made through the introduction of FAIR data principles where individual users with concerns of having their data released could still contribute to the portal. Those stakeholders that were happy to have their data released could also operate through the portal.*
>
> *Increasing the awareness of the portal's capabilities has the potential to move the portal into the operational, (day-to-day) activities of some stakeholder organisations and this will increase the sustainability of the portal.*

Deep Dive Investment 3 – Climate-smart landscapes for promoting sustainability of Pacific Island agricultural systems – ASEM/2016/101

The focus of the deep dive for this investment was on the Geospatial Platform component and the following stakeholders were interviewed.

- Elanor Bruce – Investment Project Co-Lead - University of Sydney
- Bryan Boruff – Senior Lecturer - University of Western Australia
- Ahi Saipaia – Research Assistant - MAFF (Tonga)
- John Duncan – Research Fellow - University of Western Australia
- Kevin Davies – Lecturer and Research Associate - University of Sydney

Background

The Geospatial Platform is an open-source mobile platform developed for collecting data for agricultural landscape mapping and monitoring. A key driver for developing this platform was to move from the previous paper-based surveys to digital surveys.

Outcome Focus

A detailed review of the user community resulted in the target group being narrowed to focus initially on Tonga where the in-country stakeholders had expressed a very clear view of their needs. Also, during the development phase it was recognised that having farmers as the target user group would not be effective and so the target group was changed to the MAFF and its extension officers.

An important element in the success of the platform was that it was co-developed by in-country (Tonga) and external developers and was also extensively tested in-country and modified to ensure it fully met the local needs.

The success of the platform in MAFF has resulted in other Ministries now wishing to use it. Additionally, the project is also being scaled to other Pacific countries driven by its success in Tonga.

Platform Longevity

The likelihood of the platform being used over a long period of time has been enhanced through MAFF embedding the capability into their day-to-day work stream. This is also a clear reflection of how useful the platform is because of both the co-development and the extensive testing to ensure it met MAFF and their extension officer's needs.

> ***Comment:*** *Getting the CEO of MAFF on board provided this project with a high-level champion in Tonga. This resulted in a strong MAFF buy-in and assisted in having the platform embedded within their work practices. Having good in-country technical capacity is also a key to the platforms success and should ensure the longevity of the platform.*

Several activities have given the Geospatial Platform aspect of this investment a high level of success. These include:

- Clear expression and support of user needs.
- High level in-country champion.
- Co-development – enhancing in-country capability.
- Flexibility during development to ensure user needs were fully met.
- Embedding the platform in day-to-day MAFF operations
- Demonstrating the platforms success giving confidence to other Ministries and other countries to seek to use the platform.

Annex 9 Additional Context for Recommendations

Recommendations and Assessment Findings

The extensive interviews and discussions with a wide range of ACIAR staff and project participants indicate that there is a diverse range of understanding, awareness, and the application of data management processes within the ACIAR investment projects. The reasons for this are discussed in Section 6 – Interview Findings. However, one important theme extracted from these discussions was the need for a strategic level of guidance and direction for data management activities within ACIAR investments.

It is acknowledged that given the existing diversity of capability and understanding, it is not possible to provide a 'one size fits all' approach for these Recommendations. Each project will have different data management requirements and different data management capabilities. Additionally, taking each project 'one step' higher up the ladder of the FAIR maturity model will require different levels and types of support.

This variability necessitates the development of an approach that supports all projects and ACIAR's data management implementation at a speed and scale appropriate to each project. The Recommendations aim to reflect this diversity and provide a consistent platform for all ACIAR projects to increase their data management capabilities to facilitate data sharing regardless of what level they may be at within a FAIR Data Maturity model.

Recommendations and External Drivers

There are a range of external activities that have the potential to significantly influence ACIAR and its approach to data management and data sharing. Some of these factors cover an intention to align and work with other organisations to expand ACIAR's collaborative and multi-organisation research approach to meet its strategic goals. The second and perhaps equally as significant is the rapidly increasing focus and specific directives coming from the Australian Government in relation to publicly funded data.

The Recommendations are also driven by these external drivers. These external influences come from several organisational and government initiatives and include:

- UNESCO Sustainable Development Goals
- Global Partnership for Sustainable Development Data
- Australian Government Data Strategies and Policies

Details of how these have influenced the Recommendations are provided below.

UNESCO Sustainable Development Goals

In 2015, UNESCO commenced creating a new development agenda that is underpinned by Sustainable Development Goals. UNESCO states that:

> *"Achieving these goals requires integrated action on social, environmental and economic challenges, with a focus on inclusive, participatory development that leaves no one behind.*
>
> *Critical data for global, regional and national development policymaking is still lacking. Many governments still do not have access to adequate data on their entire populations. This is particularly true for the poorest and most marginalized, the very people that leaders will need to focus on if they are to achieve zero extreme poverty and zero emissions by 2030, and to 'leave no one behind' in the process.*
>
> *The Secretary-General's Independent Expert Advisory Group on a Data Revolution for Sustainable Development (IEAG) has made specific recommendations on how to address these challenges,*

calling for a UN-led effort to mobilise the data revolution for sustainable development, by:

- *Fostering and promoting innovation to fill data gaps.*
- *Mobilising resources to overcome inequalities between developed and developing countries and between data-poor and data-rich people.*
- *Leadership and coordination to enable the data revolution to play its full role in the realisation of sustainable development."*[76]

The UNESCO SDG initiatives together with their views on the value and use of data can be seen as important drivers for an ACIAR to move towards FAIR principle compliance within its investment projects to better support the recipient countries.

Global Partnership for Sustainable Development Data

As a consequence of UNESCO's Global Sustainability Goals, the Global Partnership for Sustainable Development Data has been established to provide a specific focus on data for sustainable development. They state that:

Data is power, but our data systems are failing people across the globe.

Governments want to make better use of data to tackle the world's toughest challenges, transform economies and protect people and the planet. Companies are constantly innovating and want to contribute to public good. Community groups are using data to serve their people and hold their leaders accountable.

Yet action is fragmented. Changemakers don't know how to find each other. When they do, they don't speak the same language.

The Global Partnership for Sustainable Development Data bridges these divides - as a connector, a facilitator, a broker, and an advocate. Our vibrant global network of nearly 300 partners is collaborating to put the power of data behind delivering the Sustainable Development Goals by:

- *Building a global movement and voice advocating on data for development*
- *Influencing and shaping global policies and frameworks on data*
- *Getting data to where it is needed, in the hands of decision makers and governments.*[77]

CABI appreciates that ACIAR projects generally are not considered as 'Big Data' activities. However, the benefit of any data sharing is that data from a range of disciplines can be used to support research and other activities on a much broader scale. Bringing data together from multiple sources (including ACIAR projects) provides an opportunity for countries to better support their sustainable development goals.

CABI suggests that by moving towards a FAIR compliant data management regime ACIAR can become a significant catalyst for recipient nations through helping them to uplift their digital capabilities. For example, helping recipient countries move towards having such capability as the Australian AgriFood Data Exchange described above. This will support these countries contributing more effectively to both national and global UNESCO Sustainable Development Goals. ACIAR's current role in capacity building can be elevated by creating improved investment outcomes through supporting the sharing of valuable agricultural science data.

[76] https://www.un.org/en/global-issues/big-data-for-sustainable-development

[77] https://www.data4sdgs.org/about-gpsdd

Recommendations and Government Data Strategies and Policies

CABI believes that the current direction of the Australian Government, regarding publicly funded data related activities, are increasingly moving towards a FAIR or Open Data stance. The Government's movement in this direction is seen as continually being strengthened with an emphasis on how its publicly funded entities support data management with a focus on data reuse. Consequently, the Recommendations aim to reflect these Australian Government's directions and where applicable address the concepts and requirements expressed in the relevant Government data policies[78].

- Public Data Policy Statement
- Government response to the Productivity Commission's Data Availability and Use report
- Digital Continuity 2020 Policy
- Protective Security Policy Framework
- Australian Government Information Security Manual
- Privacy Act 1988 and Australian Government Agencies Privacy Code
- Data Sharing Principles
- De-identification Decision Making Framework
- Delivering for Australians - A world-class Australian Public Service: The Government's APS public reform agenda

The current relevant Australian Government Policies include:

A key element of the Public Data Policy Statement is that it…

> *"commits Commonwealth Government entities to:*
>
> - *specific actions designed to optimise the use and reuse of public data*
> - *release non-sensitive data as open by default*
> - *collaborate with the private and research sectors to extend the value of public data for the benefit of the Australian public.*
>
> *The data security and privacy of all Australians is of the highest importance. The government will always adhere to privacy laws and the highest possible security standards. Non-sensitive public data can, however, be of enormous benefit to the Australian economy."*[79]

The Australian Government is committed to making publicly funded data created by Government entities available as much as possible to stimulate the digital economy and realise a range of social, economic, and scientific benefits.

ACIAR has a role to play in this space and will be able to assist its partner organisations and grantees across its regions of interest in developing and improving their data management capabilities, bringing similar national benefits to those recognised by the Australian Government.

[78] https://www.datacommissioner.gov.au/data-management/foundational-four/strategy

[79] https://www.pmc.gov.au/public-data/public-data-policy

The Requirements for Effective Data Utility

There are many different factors and activities that are needed to establish effective data management. The table below provides a brief overview of these various elements that provide a framework for establishing a FAIR data compliant capability. Reference to these elements will be provided as appropriate in the Section 8 - Recommendations.

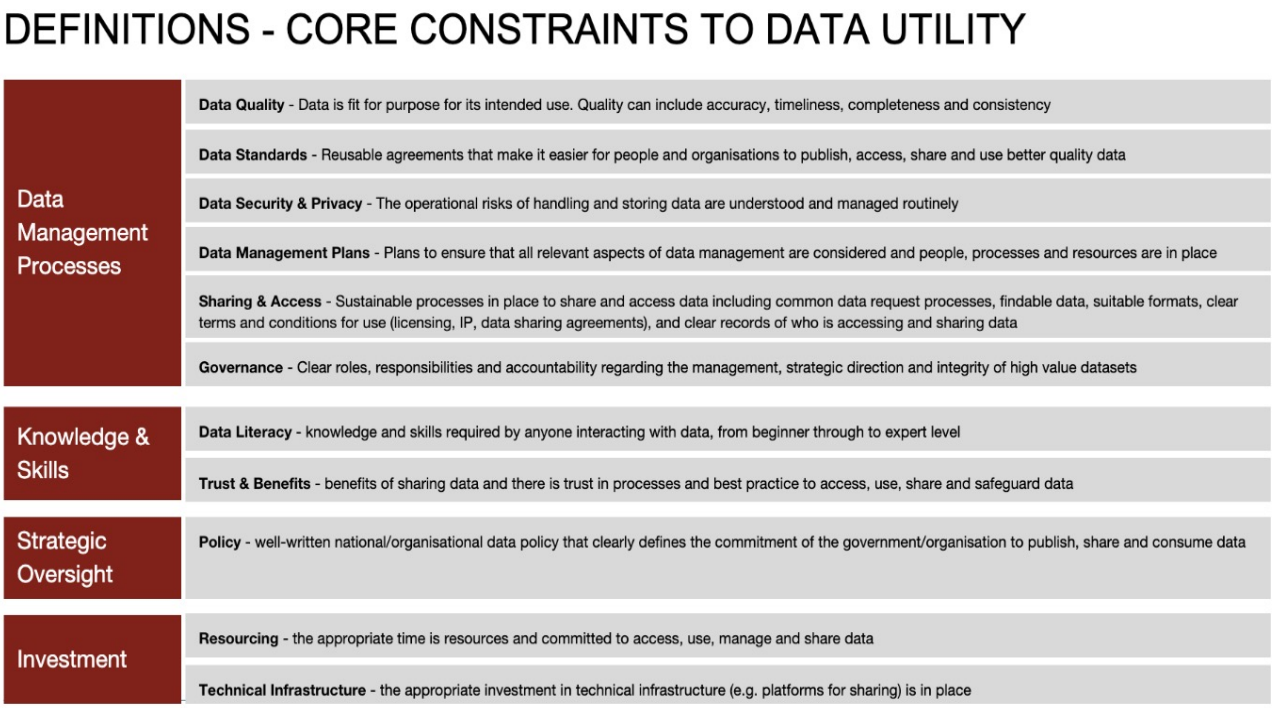


Figure 17: Definitions – Core Constraints to Data Utility - Day C, Musker R, Parr M, Smith F, Dodds L. Final Report: Enabling Data Access for Decision Agriculture for Gates Foundation Soil and Agronomy Programmes. September 2018.

Annex 10 What Is a Data Concierge?

A paper published in June 2022 by Patrick Splawa-Neyman from Monash University titled "What the dickens is a data concierge? Researcher interviews and data management reviews: misinformation, appreciation and remediation"[80] provides a range of interesting perspectives for ACIAR in relation to addressing the potential challenges of adopting a FAIR data principle framework. This paper was developed through discussing data management practices and needs with 232 researchers and graduate research students at Monash University. One of the aims of the project was:

> *"to determine what targeted research data management support researchers require to enable them to spend more time on research, or to undertake their research more efficiently."*

This aim aligns with the concerns identified in the interviews of ACIAR staff and researchers regarding increasing the data management burden for researchers undertaking ACIAR projects. Increasing the data management activity is seen as a 'zero sum' gain for individual researchers if they devote more time to data management, giving them less time to spend on their actual research. The Data Concierge paper indicates that about 25% of those interviewed were appreciative that the University was interested in their research data management experience. What was identified in the paper was a lack of clarity on what information should be included in a data management plan and the broad diversity of tools and capabilities available to support data management.

Other areas of concern or confusion included:

- the Intellectual Property ownership of the data,
- a lack of file naming conventions,
- a lack of a data management plan and framework for projects, and
- issues of data storage, data sharing and data backup

Many of the issues and concerns faced by the Monash researchers align with those of the ACIAR researchers. A way forward suggested in the Monash paper is the creation of an effective research data management support mechanism, labelled in the paper as a Data Concierge. A person (or persons) who provides advice and direction to researchers to help resolve their data management issues. The paper identifies 10 core principles for those aiming to assist researchers with data management activities. These are:

- Time for researchers is a zero-sum game
  - In the context of research, a zero-sum game is one where a researcher sees his or her time as a finite resource, and spending time on one activity takes away from another activity. Researchers are focussed on completing research and gaining credit for their research outputs, and activities such as research data management that indirectly contribute to research are not given prominence.
- Research data management is low priority
  - Related to core principle one, a researcher's goal is to create research outputs such as journal articles and the focus is on activities that directly contribute to that output at the expense of other activities.
- Researchers use their own language
  - Every industry has its own jargon and research is no different. For example, the FAIR data principles encourage access to data, but FAIR is not a term that researchers use. Researcher engagement needs to be worded in researchers' terms, so when promoting FAIR talking about increasing the impact of digital data rather than interoperability is more likely to be well received by researchers.

[80] Splawa-Neyman, Patrick (2022): What the dickens is a data concierge? Researcher interviews and data management reviews: misinformation, appreciation and remediation. figshare. Conference contribution. https://doi.org/10.6084/m9.figshare.20104631.v1

- If advice is not based on a single, written source of truth it has limited value
  - At Monash University, too often researchers are provided information that exists only in the minds of a select few individuals. This diminishes the soundness of the message and leaves research support staff in a difficult position. A single, written source of truth is required to give researchers confidence in the advice they receive.
- Researchers need to associate data management with protection and control
  - A strategy that focuses on compliance is likely to fail for two reasons: policing of compliance is largely non-existent, and compliance is associated with administrivia, not value. Librarians need to work with researchers to understand their research as this will help them to protect and control their data.
- Librarians help researchers protect and control their data
  - The role of librarians in research support is to work with researchers to protect their data. It is not enough that researchers associate data management with data protection, as they often do not have the time to learn how best to manage their data, and they will not risk changing their data management practice unless they are confident a better option can be implemented and maintained.
- Make it very easy for researchers
  - Research data management resources at large research institutions can often be spread across various seemingly unrelated work groups. Researchers will not spend the time to learn best practices for research data management if it will take them away from their research. The information provided must be simple, complete, accurate, relevant and easily accessible from a known location.
- Researchers need one-on-one assistance
- Providing in-depth and useful research data management support takes time and cannot adequately be done on a group basis unless providing a very general overview. Researchers may find generic advice useful; however, there will always be a large group that needs more personalised service. A personalised service will also improve the expertise of librarians.
- Researchers need information that is simple
  - Research data management can be very complex and researchers need assistance to understand the complexity. Matrices, flowcharts, sequencing, infographics and related ways to present complex information often makes complex information simplistic rather than simple, further confusing researchers
- Ultimately, librarians must create value for researchers
  - Irrespective of library strategy, team direction or individual roles, librarians need to ensure value is being created or co-created for researchers. Expertise is needed by researchers and rather than being seen as an add-on, our expertise needs to be embedded in research.

While this paper is focussed on library support for researchers' data management activities, the concepts and findings of this paper are considered applicable to the findings of this ACIAR FAIR review and should be viewed in this broader context. The paper clearly identifies issues similar to those identified by the ACIAR and the investment researchers and proposes some very useful solutions. These solutions would need to be tailored to meet ACIAR's needs but are considered generic enough to provide a useful starting point. This view is reflected in the Recommendations of this Report where the concept of a Data Concierge is proposed for consideration.