Environmental Data Management Policy Statement

mrst

Contents

Purpose

The government is working to make the New Zealand science system deliver greater benefits for New Zealand by becoming more efficient and effective. More and better domestic and international collaboration, greater use of e-research services and applications, better use of large-scale research infrastructure, and greater connectedness to end-users are strong themes. e-Research – real-time data sharing using web-based tools – is a powerful vehicle for these science system improvements. Publicly-funded science data, highly mobile and heavily reused, will fuel more efficient and powerful science that maximises value for New Zealand.

A comprehensive national policy on management and reuse of publicly-funded science data will open the door for New Zealand's science sector to move into this future. The Environmental Data Management Policy Statement (the Policy Statement) is the first stage in this important development.

Context

The data deluge – worldwide and in science

Since the Digital Revolution, a 'data deluge' has begun sweeping the globe (see infographic below). In the scientific domain, it is opening the door to a new world: a 'Fourth Paradigm'[1] of research dominated by data-driven methods, collaboration and connection.   The advent of data-driven science has been likened to the invention of the printing press – it promises to lift the world into an entirely new era of efficiency, connectedness, knowledge and insight.
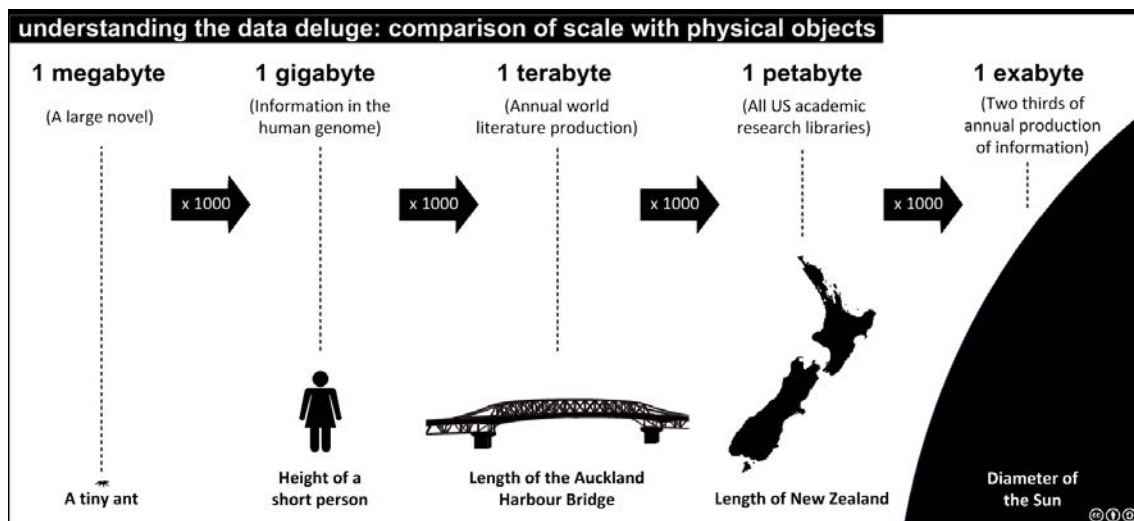
This infographic is licensed by Julian Carver under the Creative Commons Attribution-ShareAlike 3.0 New Zealand License. The original figures are from *The Data Deluge*, JISC 2004

[1] *The Fourth Paradigm: Data-Intensive Scientific Discovery* Tony Hey, Stewart Tansley, Kristin Tolle Microsoft Research 2009

Worldwide, the scientific community is striving to develop the technological tools and ways of working that will enable us to harness the power of the data deluge.[2] Governments have been backing their efforts by investing in data infrastructures and e-research architecture, and setting policy directions so that the scientific community can move forward into the Fourth Paradigm.

The data deluge is sweeping with similar force through the non-scientific community, and worldwide there is a steadily rising wave of citizen demand for greater openness, sharing and reuse of government-funded data. The public, aware that technology is well ahead of policy, are pressuring governments to open access to and mobilise publicly-funded data.[3]

The recent global economic crisis saw many nations focus attention on economic growth and value for money from public spending. Governments started looking to data reuse as a powerful way to maximise return on public investment, as well as make governments' activity more transparent. An indicator of the significance of this trend is that the Obama administration's first initiative on taking office was to sign the high-profile Memorandum on Transparency and Open Government.[4]

New Zealand context – efficiency and democracy

Recognising these trends, the Government's attention to data management and reuse has steadily intensified. There has been a significant increase in departmental activity under the Government's Open Government Information and Data Reuse work-programme, driven by the Ministerial Committee on Government ICT.[5]

In the science arena, New Zealand has invested heavily in the technological infrastructure for mobilising and reusing science data – such as the KAREN, grid computing, repositories and supercomputers.

Watching out for best practice

New Zealand is looking at international counterpart agencies, both within and beyond the scientific community (such as the OECD), for examples of data management policies and systems to learn from. Recently, policy signals have been converging, and prioritising mobilisation and reuse of publicly-funded data. New Zealand is now beginning to develop the frameworks of data management practices, policies and infrastructures that harness the data deluge and tap into its benefits.

---

[2] See the Definition of Concepts and Glossary sections for terms used in this Policy Statement.
[3] An example is the open.org movement: h*ttp://en.wikipedia.org/wiki/Open_science_data*
[4] http://www.whitehouse.gov/the_press_office/TransparencyandOpenGovernment/
[5] The Hon. Maurice Williamson (Lands) at the ESRI User Conference 3.11.2009
http://www.beehive.govt.nz/speech/nz+esri+user+conference ; The Hon. Steven Joyce at the TUANZ Telecommunications Day, 20.4.2010
http://www.beehive.govt.nz/speech/speech+tuanz+telecommunications+day+1

The New Zealand science community has nodes of significant progress, principally in environmental research, science and technology (RS&T) and also in the biosciences. Nevertheless the Ministry of Research, Science and Technology (MoRST) considers the lack of comprehensive national policy on science data management to be a critical gap. It is impeding our science sector's progress towards the new paradigm of greater efficiencies, better connectedness and more collaboration in RS&T.

Data management policy

A variety of national and international developments are coinciding to provide a significant opportunity for New Zealand to start making tangible progress on science data management. The government wishes to see New Zealand develop both the sociological and technological frameworks that will improve how publicly-funded science data are managed and used. This means setting a clear policy direction for improving data management, and collaboratively developing pragmatic implementation frameworks to underpin it.

MoRST's approach: prototyping and learning

With its data management policy work, MoRST[*] is setting a clear direction of travel for science data management. We are taking a long-term, holistic approach to this journey. This approach recognises the importance of working with the whole 'ecosystem' of players involved with science data.

In partnership with practitioners from the environmental RS&T sector, we are prototyping the sensible implementation frameworks to underpin high-level data management policy.

A vision

The first stage was identifying a generic vision and high-level principles, tested with expert practitioners. The 2008 *Environment Data 2.0 r*eport[6] set out a vision and a set of underpinning principles for managing science data. Based on the OECD's generic *Principles and Guidelines for Access to Research Data from Public Funding* these formed the foundation for subsequent development of this Environmental Data Policy Statement (the Policy Statement).

The vision is:

> *In 2015, open access to environmental research data from public funding is easy, timely, user-friendly and preferably web-based.*

---

[*] MoRST and FRST – At the time of writing, it has been announced that the Ministry for Research, Science and Technology (MoRST) and the Foundation for Research, Science and Technology (FRST) will merge, but the new agency is not yet in existence. For the purposes of the Policy Statement, 'MoRST' will be used to denote the new, merged organisation.

[6] 'Environment Data 2.0: A Draft Environment Sector Data Management Vision and Action Plan' discussion document prepared for MoRST by Julian Carver April 2008

The underpinning set of principles are ways to manifest the vision. In summary, these are:

- *data are well stored*
- *advanced collection, aggregation and federation methods are commonplace*
- *data are reused*
- *New Zealand science is a strong international collaborator*
- *IP, privacy and safety are well-managed*
- *data can be found and interpreted.*

The Policy Statement: prototyping the vehicles

The next stage is embarking New Zealand's science system on the collaborative journey towards the vision. This Policy Statement signals how MoRST intends to contribute to practically improving New Zealand's environmental data management, providing the stepping-stones for progress in the wider science data 'ecosystem'.

This Policy Statement outlines MoRST's intention to work simultaneously on three fronts: Vote RS&T funding and system settings, Advice and assistance, and Culture change and awareness-raising.   We will be partnering with leading practitioners – primarily from the environmental RS&T community – to build and test prototypes of the implementation approaches that will encourage real progress in science data management.   We will use the experiences and lessons from the environmental RS&T sector in future work on data management with other sectors of the science community.

Our operating principles for this work are:

- clear, well-aligned policy settings at the national level
- an holistic 'system approach' to achieving progress
- a commitment to collaborating with the ecosystem's other players.

These principles will be guiding MoRST's work across a variety of current and ongoing initiatives. We look forward to a collaborative journey working with players from throughout the science data 'ecosystem'.

Managing Publicly-Funded Environmental Data

1.   Vote: RS&T funding and system settings

MoRST* will work with the RS&T sector and the Foundation to integrate clearly the government's expectation for better data management into Vote RS&T funding and system settings.

Empowering CRIs to drive better data management

MoRST will work with CRIs to include reference to data management in CRIs' Statements of Core Purpose (SCPs) and Statements of Corporate Intent (SCIs). This will be at levels appropriate to the 'empowering' approach underpinning the implementation of the CRI Taskforce's recommendations. Performance would be assessed through the overall performance framework and assessment process that is being developed as part of implementing the CRI Taskforce recommendations.

This approach will enable CRIs to develop responses to government data management policy that are appropriate to their institution and are strategic across the RS&T system.   SCPs' articulations of data management will be as outcome-level statements, generic to management of all publicly-funded science data.

Contestable funding

All new environmental science data generated with contestable Vote RS&T funds should meet minimum expectations of good data management. We will work with the Foundation and the RS&T sector to identify these expectations and incorporate them appropriately into processes and mechanisms for the contestable part of the Vote.

This work will start with the Environment research priority area but is designed to eventually be applied to other areas of science. In the medium term it will identify government's minimum expectation for management of publicly-funded data, which will provide a point of reference for CRIs to work with in their institutional data management policy.

Evaluation

MoRST will be turning to the RS&T sector to help ensure there is sufficient expertise in data management built into the planned assessment and evaluation structures to ensure good performance is recognised. Potential mechanisms might include the use of expert panels for the five-year reviews of CRIs, and expertise in panels assessing proposals for contestable funding.

Infrastructure

MoRST will also work to ensure national large-scale research and e-science infrastructure investment strategies encourage and support effective data management.

2.   Advice and assistance – The Best Steps Guide web resource

MoRST plans to construct a service-oriented 'portal' page on MoRST's website dedicated to the basics of science data management. This will be regularly updated and will provide generic information and assistance applicable to all domains of science.

It will prepare the ground for players in the broader New Zealand science data 'ecosystem' to connect, build partnerships and lift their own data management performance over time.

It will do this by:

i.   Pointing players to expert assistance

In one place, the web-page will enable practitioners to find links to New Zealand information sector agencies and to useful resources[7] that can guide them with practical steps to improve their data management.

Agencies with data management expertise will provide links to their key assistance contacts and will provide their and others' most useful resources.

ii.   Outlining roles, rights and responsibilities

MoRST will adapt existing resources to create a New Zealand-specific resource that indicates the basic roles of the different players in a well-functioning science data 'ecosystem'. This will enable practitioners to begin conversations about data management on the same page, with a shared understanding of the different roles that need to be fulfilled.

iii.   Providing tangible illustrations of good practice

MoRST is already gathering anecdotes and exemplars of good data management practice from across the New Zealand science data ecosystem. Partner agencies will help grow this stock of exemplars, so eventually we can illustrate each data ecosystem role with a range of real New Zealand case-studies and personal testimonies.

---

[7] Such as those provided by StatsNZ, Archives NZ, the SSC's NZGOAL team, and the Australian National Data Service

3.   Culture change and awareness-raising

MoRST will be working across the whole science data ecosystem to build capacity and help shift players' attitudes into the 'open data' paradigm. We will work to raise awareness of the government's focus on data management and elevate data management's status as a valued activity – for scientists and all other players in the science data ecosystem.

Website banner advertisement

MoRST will develop a small web advertisement that will ideally be added to websites that regularly experience heavy traffic of players in the ecosystem. It will click viewers directly through to the Policy Statement's web-page.

Awareness-raising and promotion

MoRST will use its networks across government and the science system to draw attention to and promote data management. This will include communicating key messages at key events and conferences, and showcasing New Zealand exemplars of good practice in these and other high-exposure messaging (such as science news bulletins).

Accounting for data

MoRST and partners will embark on a collaborative investigation and discussion process (likely to involve a facilitated workshop) investigating and discussing with CFOs and CIOs[*] how science data assets and data management activity are and should be accounted for.

Community-building and capability development

Greater community cohesion and capability is essential for building New Zealand's scientific information services sector from its current infancy relative to other nations'. MoRST will actively encourage collaboration both across the RS&T sector and between it and the information sector, and facilitate the building of communities of expertise.

This is likely to involve hosting an event such as a practitioner workshop, then encouraging practitioners to take the lead. Where appropriate we will start with the environmental RS&T sector, but where development can be genuinely generic we will take an RS&T-sector-wide approach.

---

[*] Chief Financial Officers and Chief Information Officers

One example of a sector-wide initiative is current work by MoRST, ARMS and CONZUL[*] investigating how New Zealand can develop an equivalent of the 'Intersect' model of shared end-to-end data management and e-research services.

Innovative and long-term work

Investigating innovative ways to encourage better data management, embed it into science and professional culture, and align key political, financial and institutional settings across the science ecosystem.

Examples of potential work include:

- working with the tertiary, RS&T and information sectors on how to embed data management education into scientists' training

- investigating innovative ways to encourage progress, such as exploring the concept of a data management 'capability fund' or rotating Data Chair position (held in turn by leaders from the RS&T system)

- working to develop or adopt reward / attribution systems for data reuse for scientists (such as dataset citation indices)

- working to measure and quantify the benefits of different kinds of data management effort to guide future work.

---

[*] Australasian Research Management Society and the Council of New Zealand University Librarians

Definition of concepts

See the Glossary (follows) for specific definitions of terms. Important concepts for this Policy Statement are defined in this section.

Environmental data

In the Policy Statement we use this as a broad term for publicly-funded data about natural phenomena, collected by regularised and repeatable methods. This is a deliberately broad definition, covering a vast range of subject areas and types of data. The data management policy statement is designed on the premise that there are certain basic, generic characteristics of well-managed, highly reusable data bodies that are common across disciplines.

Data management

This document uses 'data management' as an umbrella term for the technological, institutional, financial, policy and cultural decisions that determine whether and to what degree science data produced are sharable and reusable.

These decisions happen at all levels, throughout the ecosystem of agencies involved with environmental data. They are made by a wide range of bodies and people, including (but by no means limited to) scientists and researchers, librarians and IT managers, contract managers, CIOs and CFOs, research and policy strategists, investors, and end-users.

New data or archives?

The Policy Statement focuses on building better data management practices, philosophies and policies into the ecosystem of science data. It is likely to be easier for people to make shifts in current practice, so we focus on new data being produced with public funding. However the characteristics of well-managed data apply equally to 'back catalogues'.

MoRST and FRST

At the time of writing, it has been announced that the Ministry for Research, Science and Technology (MoRST) and the Foundation for Research, Science and Technology (FRST) will merge, but the new agency does not yet exist .

MoRST will be used in this document to denote the new, merged organisation of MoRST and FRST.

Glossary of terms

Data ecosystem

The organisations that produce, fund, store, curate, handle, transmit, analyse, present and consume data or the products thereof, interacting through infrastructure, interpersonal and inter-institutional relationships and connections.

For the purposes of this statement, data means publicly-funded data (q.v.)

Environmental science data

Data about natural phenomena, collected by regularised and repeatable methods. This includes everything from physical specimens through to machine-readable instrumental data, and includes living and non-living (such as climate and geological) phenomena. It includes both raw and QA/QC'd data, collected manually or automatically, by people or instruments, for research or for monitoring.

This document uses this definition for 'science data' but without the qualification of 'natural phenomena'.

e-science

Also called e-research, this is real-time data sharing in research, science and technology using web-based tools. e-science means scientists and researchers can:

- share and analyse data and information in real-time

- create new research collaborations

- use re-use data and information in novel ways

- keep accurate archives of research information to share later

- share equipment and infrastructure

- participate in solving complex global problems through international collaborations.

Informatics

A field of study concerned with the gathering, manipulating, storing, retrieving, sharing and mobilisation of recorded information, with particular focus on the use of technology for improving access to and utilisation of information.

Information (services) sector

The agencies and organisations expert in the management and movement of information.

Specifically they include organisations with technical expertise – such as libraries and librarians (including the National Library, CONZUL and CRI Librarians), IT, communication technology and computing services organisations, and agencies with political mandates for data and information management, such as Archives New Zealand, the State Services Commission and the Department of Internal Affairs.

For the purposes of this document, 'information sector' means organisations with significant expertise in data management and informatics (q.v.), and willingness to provide assistance outside their organisations (although their specific mandate may not include this external 'services' component). Statistics New Zealand, Archives New Zealand, the National Library and Landcare Research's informatics expertise are key examples.

Publicly-funded

Paid for wholly with the public purse – i.e. with funding from taxes or from rates.

Science data ecosystem

As for 'data ecosystem', but the data flowing through are science data.