

Research Data and the *Code*: The Institutional Perspective

Outline

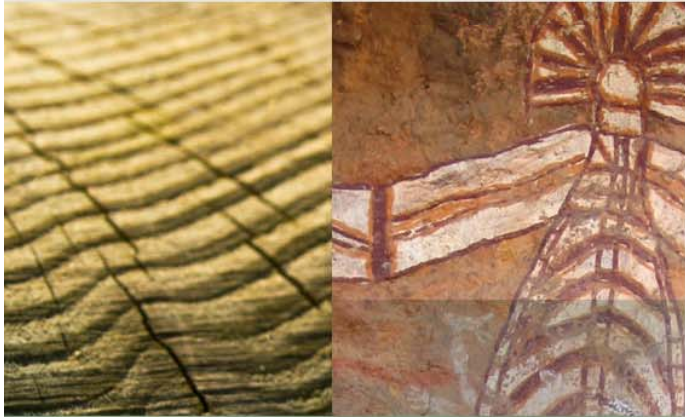
- About ANDS
- The *Code*
- The institutional response to the *Code*
- Strategies for institutions to address the *Code*
- Morning tea
- Martin Borchert: The QUT response
- Discussion
- Report back
- Summing up

What this workshop is **not** about

- Provisions in the Code that don't relate to data
- Data that isn't digital
- The researcher perspective (come back tomorrow for that one)

About us

- Ian Barnes
 - Research Analyst
- Margaret Henty
 - Leader, Capabilities Program



About ANDS



Australian Government

Department of Innovation, Industry, Science and Research

venturousaustralia

building strength in innovation August 2008 (1)

- *“We should aim to create an open scientific culture where as much information as possible is moved out of people’s heads and labs, onto the network, and into tools which can help us structure and filter the information.”* Michael Nielsen





Australian Government

Department of Innovation, Industry, Science and Research

Powering Ideas

An Innovation Agenda for the 21st Century

May 2009

- As the problems studied become more complex, so does the infrastructure needed to develop collaborative and multidisciplinary solutions...
- For example, climate change, astronomy, genomics and other data-intensive fields of inquiry all depend on large-scale ICT capabilities in data computation, integration, storage, and distribution.



The Rationale for ANDS

- Research data intensity is increasing
- Problems are at a larger scale
- A research group approach is inefficient and variable
- An institutional approach to data leads to greater research efficiency
- A national approach to data is an enabler for tackling larger problems, more efficiency and greater synergy
- A combined approach provides Australia with a means of engaging with the world, but also provides Australia a research collaboration advantage

About ANDS

- An initiative of the Australian Government being conducted as part of the National Collaborative Research Infrastructure Strategy (NCRIS) and the Education Investment Fund (EIF)
- A collaboration between Monash University, the Australian National University and CSIRO
- Locations at Monash & ANU

What is ANDS trying to do?

ANDS aims to:

- transform the disparate collections of research data around Australia into a cohesive collection of research resources
- influence national policy in the area of data management in the Australian research community
- inform best practice for the long-term sustainability of data

ANDS funding

- ANDS was established in 2008 under the NCRIS Platforms for Collaboration with initial funding of \$24m to June 30, 2011.
- EIF funding is another \$48 Million.

ANDS Response

- The combined funding means that this is the largest project devoted to research data in the world
- The funding is for 2 years
- ANDS will work with research institutions rather than directly with researchers
- Data access will be a priority
- ANDS will seek to create the Australian Research Data Commons (ARDC)

Government expectations

- The establishment of the Australian Research Data Commons will support the discovery of, and access to, research data held in Australian universities, publicly funded research agencies and government organisations for the benefit of research.
- This investment will enable the construction of a range of ICT utilities to capitalise on and ensure greater use and re-use of data resources.

The Australian Research Data Commons

- The set of shareable Australian research collections
- Descriptions of those collections
- The infrastructure needed to enable, populate and support the Commons
- ANDS does not actually hold the data, but points to the location of the data

Path: Home > IVEC MEST - Western Australian Marine Data and Projects > Collections >

Collection

[View the complete record in the ANDS Collections Registry](#)

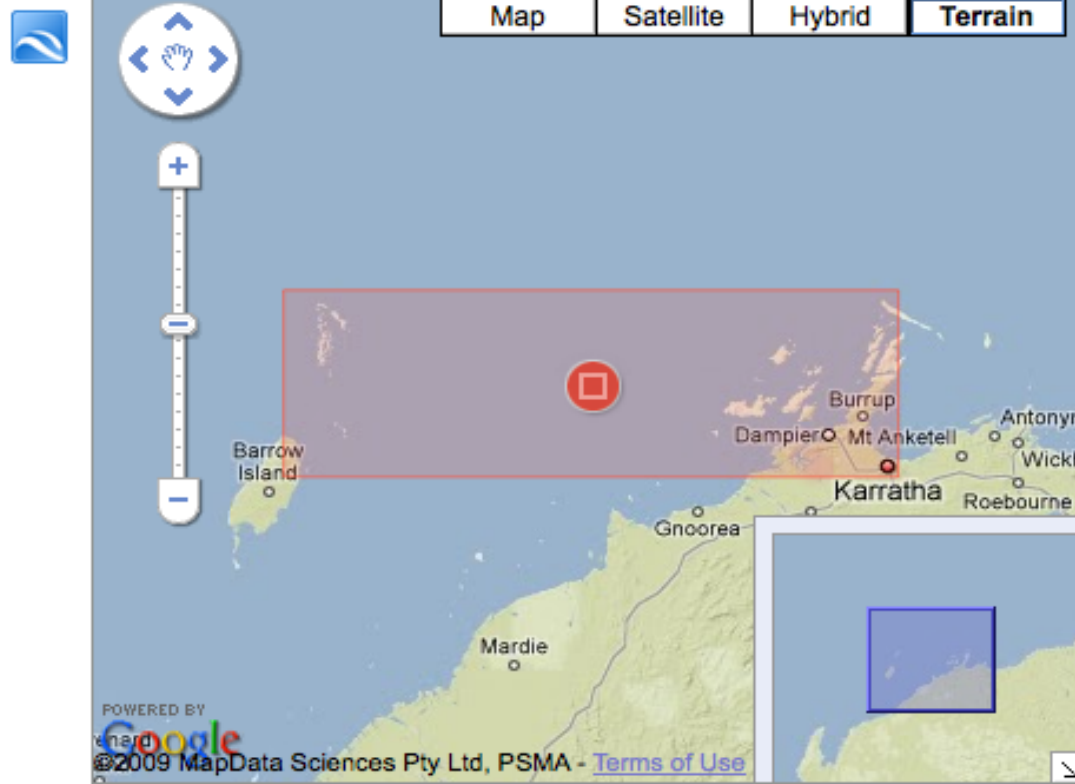
Type: dataset

Title: A preliminary population estimate of the Group IV breeding stock of humpback whales off Western Australia

Descriptions: *brief:* This record is derived from DEC Marine Policy Branch Endnote library and spatially referenced SIER Database. **KEYWORDS:** Western Australia/Marine Fauna/Whale/Humpback/Biology /Movement/Breeding/Population

- Locations:**
- url: <http://mest.ivec.org:80/...>
 - coverage; from 1994-01-01T00:00:00Z; to 1994-12-31T23:59:59Z; iso19139dcmiBox: northlimit=-20.33; southlimit=-20.76; westlimit=115.44; eastLimit=116.87; projection=WGS84

Coverage:



Search Research Data Australia

whales

Search X

Results 1 to 6 of 40.

[Whale sharks, migration and ecology movements and behaviour of Whale Sharks with use of acoustic, satellite, PAT, crittercam genetic tagging.](#)

Whale sharks migration, ecology movements and behaviour of whale sharks were measured with the use of acoustic, satellite, PAT, crittercam genetic tagging technology. Keywords: Ningaloo...

[Whale Shark Monitoring](#)

The Western Australian Department of Environment and Conservation together with the Ningaloo whale shark tourism industry have been closely involved with monitoring whale sharks at...

[Whale Shark photo ID collection program](#)

To determine Whale Shark numbers at Ningaloo Marine Park by working with whale shark industry videographers to collect footage - part of the Whale Shark Management Program. KEYWORDS:...

[Whale shark population genetics](#)

In collaboration with the Shark Research Institute, Roland Mau and the rest of the Exmouth whale shark research group have been collecting tissue biopsy samples from whale sharks at...

[Population Structure and Management of the Humpback Whale \(Megaptera novaeangliae\)](#)

This project consists of a large-scale genetic study of the Western Australian Humpback (Antarctic Area Group IV). The primary aim of this study is to determine whether humpback whales...

[Whale shark ecology](#)

Baseline data collection on Whale Sharks by photo-identification, Whale Shark sex & size and collection of

Path: Home > IVEC MEST - Western Australian Marine Data and Projects > Collections >

Collection

[View the complete record in the ANDS Collections Registry](#)

Type: dataset

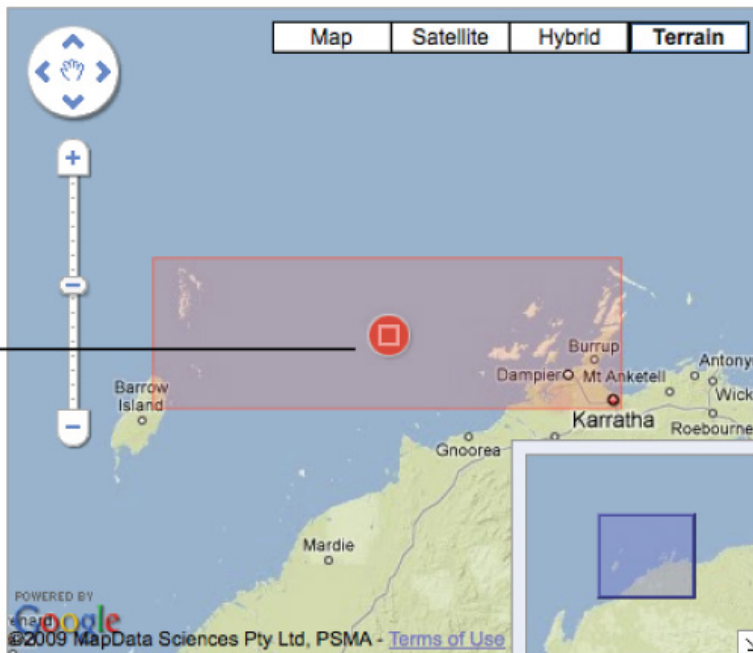
Title: A preliminary population estimate of the Group IV breeding stock of humpback whales off Western Australia

Descriptions: **brief:** This record is derived from DEC Marine Policy Branch Endnote library and spatially referenced SIER Database. **KEYWORDS:** Western Australia/Marine Fauna/Whale/Humpback/Biology /Movement/Breeding/Population

Locations:

- url: <http://mest.ivec.org:80/...>
- coverage; from 1994-01-01T00:00:00Z; to 1994-12-31T23:59:59Z; iso19139dcmiBox: northlimit=-20.33; southlimit=-20.76; westlimit=115.44; eastLimit=116.87; projection=WGS84

Coverage:



Subjects: local: biota

Relationships: distributor:

pointOfContact:

Search Research Data Australia

whales

Search x

Results 1 to 6 of 40.

[Whale sharks, migration and ecology movements and behaviour of Whale Sharks with use of acoustic, satellite, PAT, crittercam genetic tagging.](#)

Whale sharks migration, ecology movements and behaviour of whale sharks were measured with the use of acoustic, satellite, PAT, crittercam genetic tagging technology. Keywords: Ningaloo...

Whale Shark Monitoring

The Western Australian Department of Environment and Conservation together with the Ningaloo whale shark tourism industry have been closely involved with monitoring whale sharks at...

Whale Shark photo ID collection program

To determine Whale Shark numbers at Ningaloo Marine Park by working with whale shark industry videographers to collect footage - part of the Whale Shark Management Program. KEYWORDS:...

Whale shark population genetics

In collaboration with the Shark Research Institute, Roland Mau and the rest of the Exmouth whale shark research group have been collecting tissue biopsy samples from whale sharks at...

Population Structure and Management of the Humpback Whale (*Megaptera novaeangliae*)

This project consists of a large-scale genetic study of the Western Australian Humpback (Antarctic Area Group IV). The primary aim of this study is to determine whether humpback whales...

Whale shark ecology

Baseline data collection on Whale Sharks by photo-identification, Whale Shark sex & size and collection of plankton samples to determine Whale Shark prey items.

Title of the data

Description of the data

Link to original metadata record

Google Earth (KML) Link

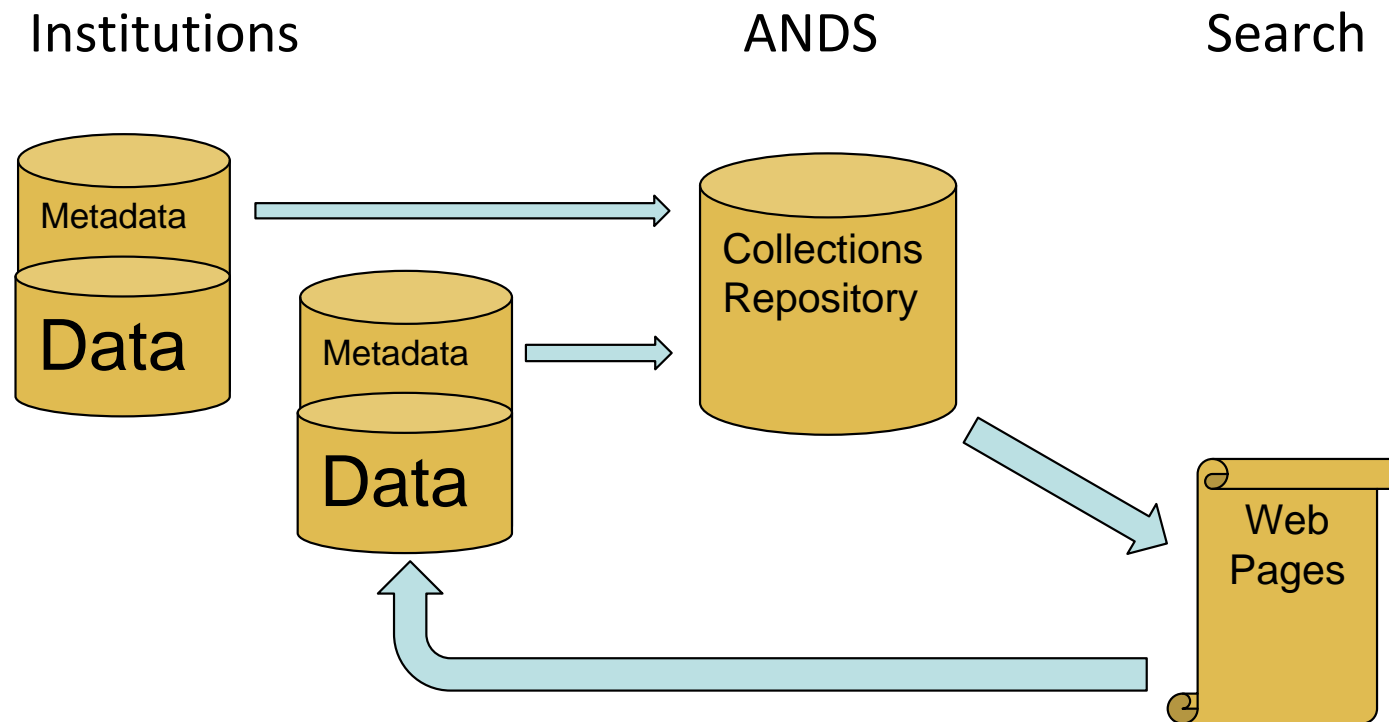
Google Map, data coverage

Search

Search Results

SAMPLE RECORD

Data Flows – Data stays with Institutions

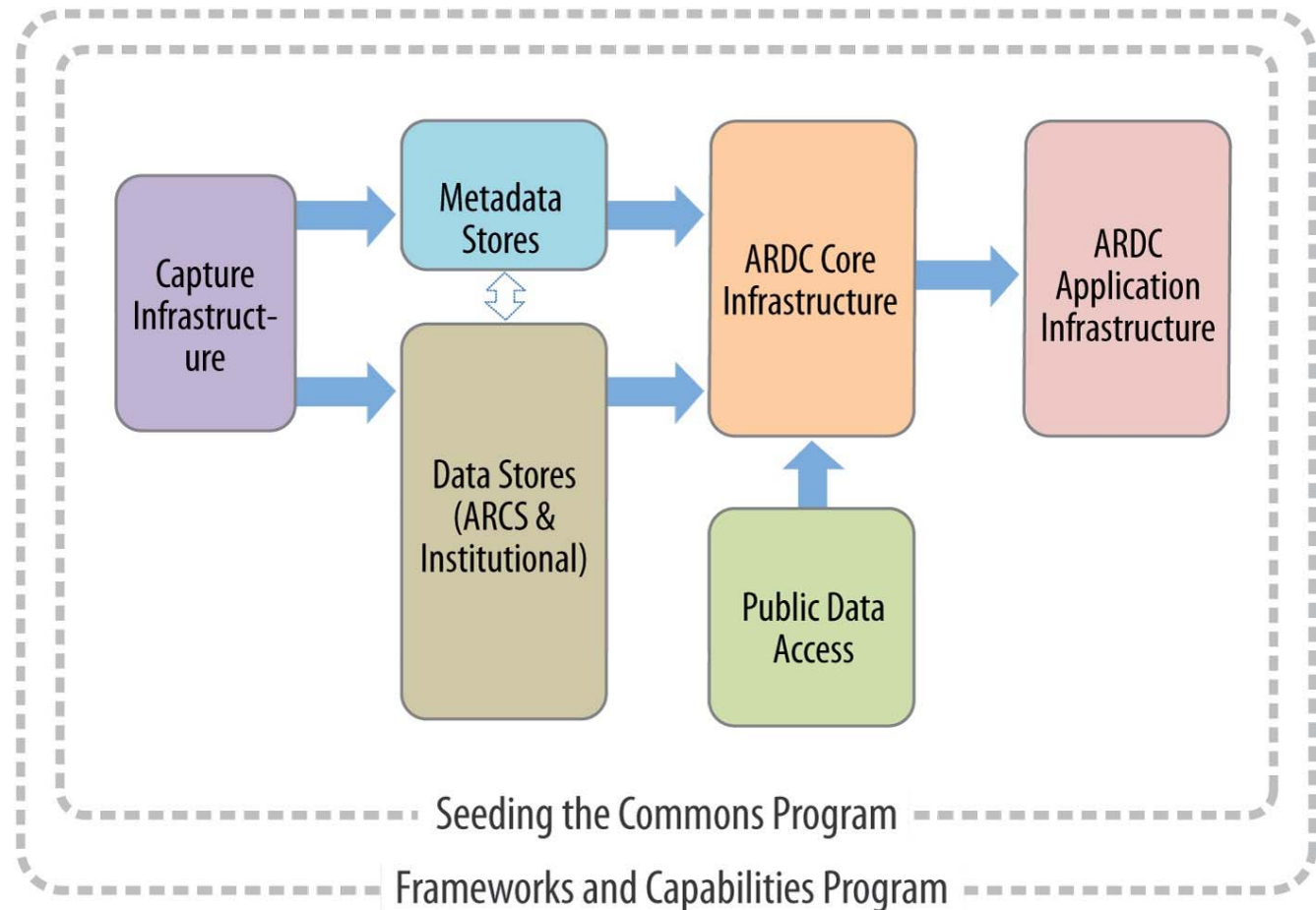


ANDS in context

- ARCS
 - State based members – iVec (WA), eRSA (SA), Intersect (NSW), VPAC (Vic), QCIF (Qld), TPAC (Tas)
- Australian Access Federation (AAF)
 - Currently in trial
- ANDS
 - Funding program to introduce data management practices
 - Provision of advice
 - Services and product provision

ANDS is being structured as seven co-ordinated inter-related service delivery programs:

- *Frameworks and Capabilities*
- **Data Capture**
- *Seeding the Commons*
- **Public Data Access**
- **Metadata storage**
- **ARDC Core**
- **ARDC Applications**



ANDS engagements

- Our role is to help researchers and institutions achieve their research data ambitions whilst creating the ARDC
- Our model is to partner at the institutional level and through local partners
- We don't have 'best practice' to distribute, but are sharing good practice
- It works best when administration, records, the research office, the library and ITS work together to meet mutually desired outcomes

Early Deliverables

- A national collections registry service – *Register My Data*
<http://www.ands.org.au/services/register-my-data.html>
- A national persistent identifier service – *Identify My Data*
<http://www.ands.org.au/services/identify-my-data.html>
- A national discovery service – *Research Data Australia*

ANDS is not a national storage and management facility, and does not provide storage hardware

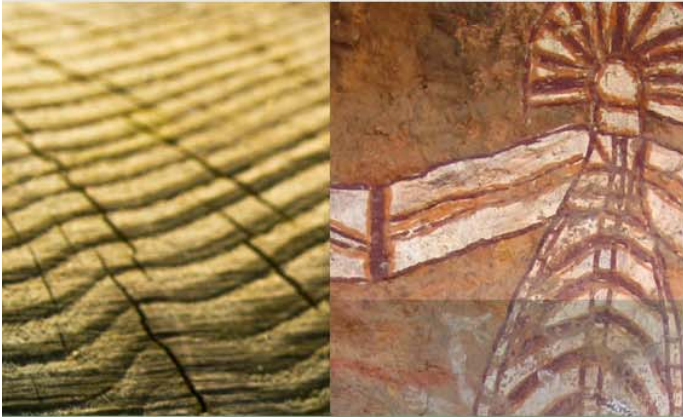
Other focus areas

- Exemplar discoverable data spaces
 - Tropics, Crystallography, Law, Bio-security, Coast
- Early advice on data capture and management practice
- Partnering with selected institutions
- Developing and contracting software solutions
- Building shared practice in response to the *Code*
- Partner with institutional ITS and ARCS for local repositories and storage

Success:

More researchers re-using data more often...

1. More data sharable
2. More data discoverable
3. Easier to share
4. Easier to find
5. More researchers engaged
6. More institutions engaged



The *Code*

What is the *Code*?

- ‘The purpose of the *Australian Code for the Responsible Conduct of Research* (the Code) is to guide institutions and researchers in responsible research practices.’
- ‘It is written specifically for universities and other public sector research institutions.’
- Covers all aspects of research practice including honesty and integrity, good stewardship of resources, responsible communications, etc.

What does the *Code* say about data?

- ‘The responsible conduct of research includes the **proper management and retention** of the research data.’

Responsibilities of institutions

- Retain research data and primary materials
- Provide secure research data storage and record-keeping facilities
- Identify ownership of research data and primary materials
- Ensure security and confidentiality of research data and primary materials

ANDS and the *Code*

- ANDS has no formal responsibility for the *Code*
- However many of the aims of the *Code* coincide with the aims of ANDS
- The *Code* makes significant references to research data and responsibilities related to it
- ANDS is therefore interested in promoting awareness and supporting the uptake of the *Code*
 - ANDS has conducted fora on institutional responses to the data aspects of the *Code*

Implications for institutions

- Data must be:
 - ‘Owned’
 - Retained
 - Stored
 - Retrievable
 - Accessible
- Institutions need to provide the policies, infrastructure, support services and facilities to do this
- Researchers cannot meet their obligations without institutional support

Key considerations

- Recognising that the *Code* is about cultural change as well as research practice
- Ensuring an appropriate governance framework with representation from all interested parts of the institution — research office, library, IT, archives & researchers
- Ensuring buy-in from senior administrators
- Providing both data management infrastructure and services to support compliance

The Challenge

- Effective collaboration between these services and areas
- Getting the institution to accept the need for it
- Using or adapting existing policy and services to enable compliance

Getting started

- High level support for the work – who is the ‘Champion’?
- Steering process to coordinate work
- Identification and review of what policy and infrastructure is already in place
- Identification of resources required to provide support – this will vary between institutions

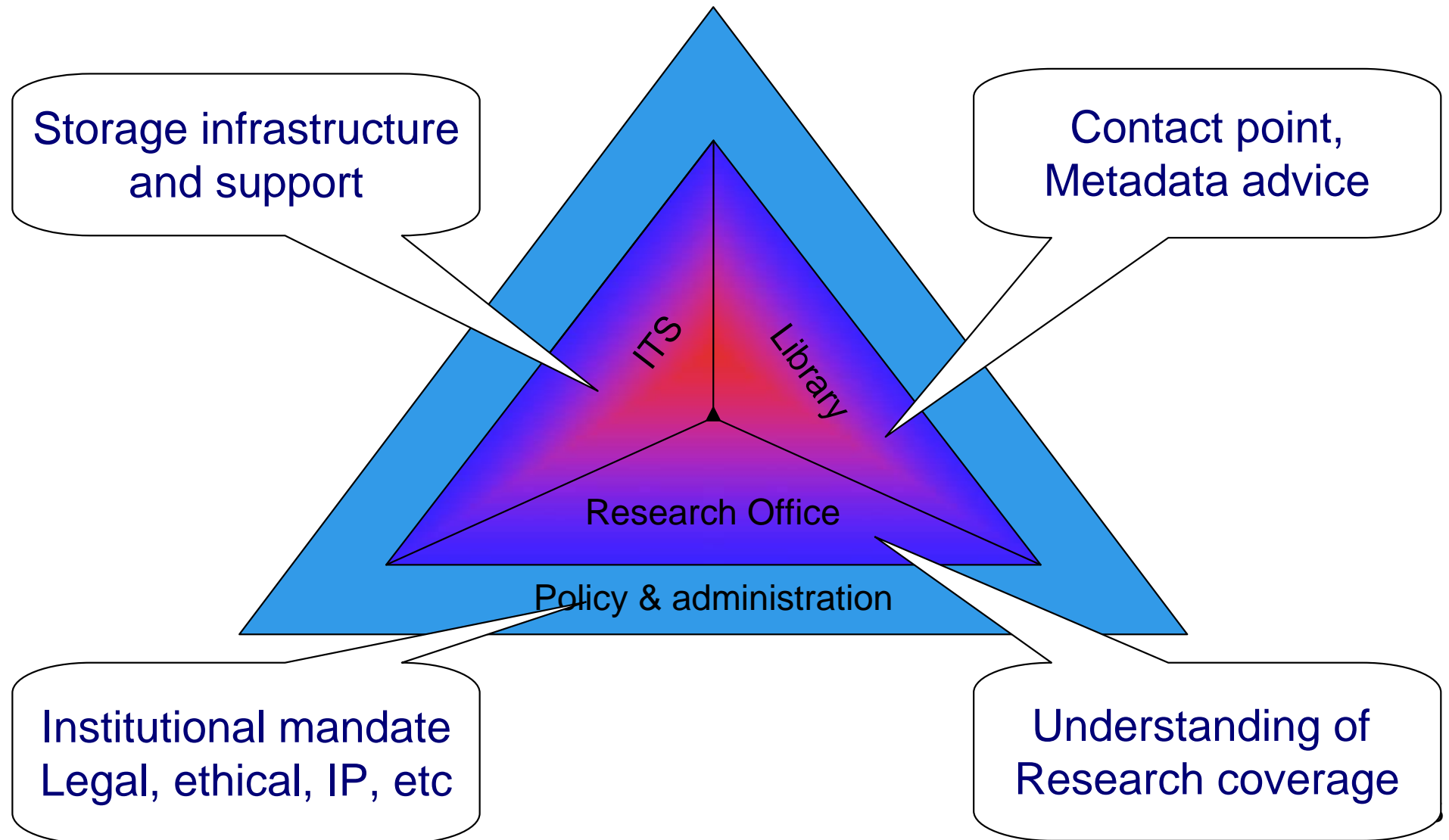
First steps in responding to the *Code* (1)

- Activities for assessing the readiness and capability of the institution (as opposed to the individual researcher):
 - an assessment of institutional commitment and priorities
 - a review of all existing relevant policies and responsibilities, especially information management policies, practices, planning and responsibilities

First steps in responding to the *Code* (2)

- a review of data storage capacity, including possible use of cloud services
- an assessment of capability of individual researchers to respond to the *Code*, through, for example, a survey of data management practices
- an audit of data currently held within the institution
- consideration of the financial implications

Institutional support areas



Potential role of Research Office

- Policy creation & oversight
- Identifying significant researchers and data producers
- Connecting data with publications
- Leadership and advocacy role
- Research practice training and oversight

Potential roles of library/archives/records

- Contact and outreach to researchers
- Information gathering about available data
- Advice on metadata, descriptions, disposal policy and sustainability
- Possible use of institutional repository for holding descriptions or connecting publications to data
- Training and support

Potential role of ITS/ICT

- Provider and supporter of storage hardware, back-ups
- Provider and supporter of tools for data capture, metadata capture and discovery
- Access control
- Connections to the ARCS Data Fabric

Policy and administration

- Institutions already have policies and procedures covering research, records management, intellectual property and other research-related activities.
- These can be adapted for *Code* compliance purposes.

Related policies (1)

1. Intellectual property — covering copyright, moral rights, patents
2. Collaboration and contractual agreements — do these include provision for data sharing after the research is complete and agreement on who will host and store the data
3. Conflict of interest — does this impact on ability to make data available

Related policies (2)

4. Data management, including:

- Storage — how to provide appropriate storage and under what conditions
- Retention — how long data is to be stored
- Disposal — how this is to be recorded and managed
- Access — how to make data available to those engaged in the research project; how data is to be “published” and made available more widely; how the institution is to keep a record of its data assets; the institutional position on open access of both data and publications

Related policies (3)

5. Ethics and privacy — how these affect of the length of data storage and the ability to share
6. Compliance — what measures are in place to track this.

Each of these needs to be considered in terms of both institutional and researcher responsibilities.

Not all are specific to data management, but each should have a data management component and requires a procedure translating policy into practice.

The Strategic Response to the *Code*

- Once an appropriate policy framework is in place, there will be a need to develop strategies to cover:
 - publicising the policies within the community
 - providing appropriate infrastructure to support researcher requirements and training in data management and use of infrastructure
 - establishing support services related to the infrastructure
 - establishing record-keeping procedures
 - demonstrating compliance through review or audit frameworks.

Costs of *Code* compliance

- Policy review and amendment
- Legal review and changes
- Potential view as a burden
- More work for existing services
- Creation of compliance monitoring
- IT infrastructure

Benefits of *Code* compliance

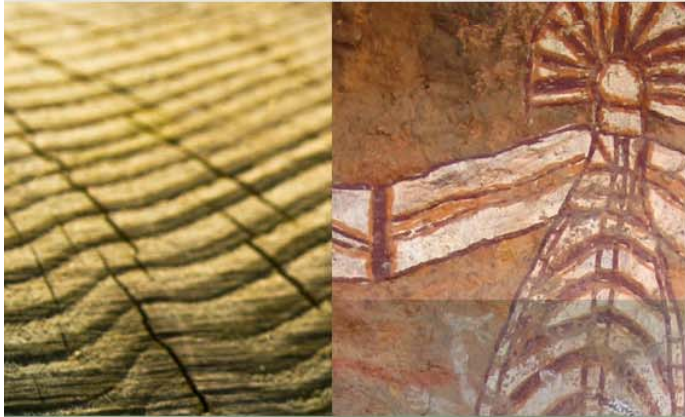
- Security of storage
- Enhanced ability to share (including “sharing with yourself”)
- Better understanding of what your researchers are doing
- Establishing a corpus of past research to build upon
- Readiness if the *Code* does become mandatory
- Readiness as funders start to require it

What do researchers get?

- Greater recognition of their work through increased citations and follow-on research
- May lead to new collaborations with data users
- Demonstrates the value of their work by allowing continued re-use of the data, which in turn may lead to further funding
- Leads to satisfaction as a result of benefiting the broader community
- Meets the contractual obligations of research funders who have data sharing policies
- Multiple copies protect the data from accidental loss

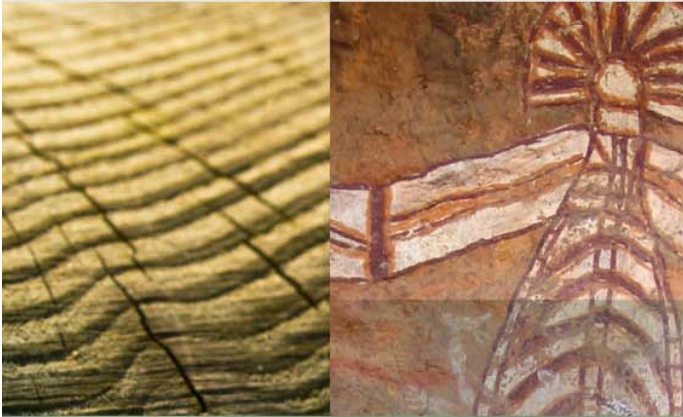
Key messages

- If it is not easy and clear, researchers won't/can't do it
- There are benefits for the organisation in doing this
- This process is about change and it will be gradual
- Effective data management comes from collaboration across the organisation
- ANDS can help



QUT's Responses to the *Code* [Data Management] – Martin Borchert

Slides available separately



Small group discussion

Discussion topics

- Where is your organisation up to?
- What connections do you have with other relevant areas in your organisation?
- Is the *Code* being discussed?
- Is Data Management being discussed?
- Are they being discussed by the same people?
- What do you see as the major impediments to *Code* compliance in your organisation?
- How can you engage with your researchers?

Thank you for your attention

Questions?

Contact ANDS via:

<http://www.ands.org.au/contact.html>

ANDS Code Resources:

<http://www.ands.org.au/resource/code.html>