Carnegie Mellon University

Curating the Scholarly Record: Data Management and Research Libraries

20 October 2016

Keith Webster
Dean of University Libraries
Director of Emerging and Integrative Media Initiatives

@cmkeithw

- The rapid development in computing technology and the Internet have opened up new applications for the basic sources of research — the base material of research data — which has given a major impetus to scientific work in recent years.

- Access to research data increases the returns from public investment in this area; reinforces open scientific inquiry; encourages diversity of studies and opinion; promotes new areas of work and enables the exploration of topics not envisioned by the initial investigators.

- The value of data lies in their use. Full and open access to scientific data should be adopted as the international norm for the exchange of scientific data derived from publicly funded research.

More data will be created in the next five years than has been collected in the whole of human history. Properly managed, this data will form a major resource for Australian researchers.

Research Data Australia

Explore ▾ About MyRDA Login

Find data for research

Find, access, and re-use data for research - from over one hundred Australian research organisations, government agencies, and cultural institutions

All Fields ▾ *Search for Data* Search

Publicly accessible online

Advanced Search Map Search

Browse By Subjects

Humanities and Social Sciences

Business, Economics and Law

Medical and Health Sciences

Engineering, Computing and Technology

Built Environment and Design

Biological Sciences

Agricultural and Veterinary Sciences

Environmental Sciences

Earth Sciences

Physical, Chemical and Mathematical Sciences




Data for researchers everywhere

Research Data Australia caters specifically for researchers but also has broader relevance to others including policy makers, educators and business people.

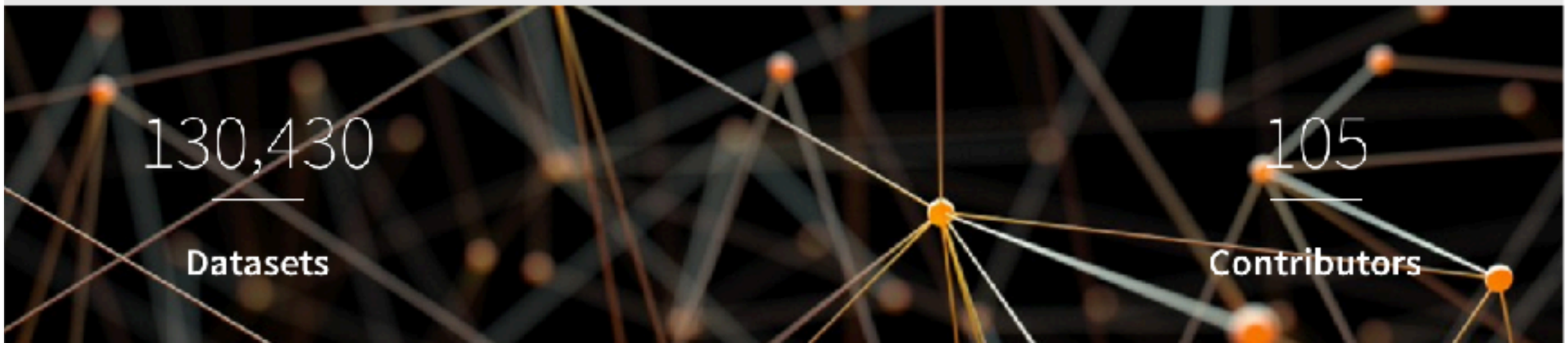
Research Data Australia covers a broad spectrum of research fields - across sciences, social sciences, arts and humanities. Much of the data you can discover here is immediately accessible online via our partners and free to use (subject to any licence conditions).

More than just a search engine

A single search in Research Data Australia retrieves data resources across a wide range of subjects and providers, so that you can:

- **reuse existing data**
- **explore beyond your discipline**
- **assemble data resources to solve big problems**

For example a search for the town "Wagga Wagga" will return results for data from a number of fields of research including: Earth Sciences, Agriculture and Veterinary sciences, Environmental Sciences, Built Environment and Design, Biological Sciences, Studies in Human Society, Studies in Creative Arts, History and Archaeology, and Philosophy and Religious Studies.

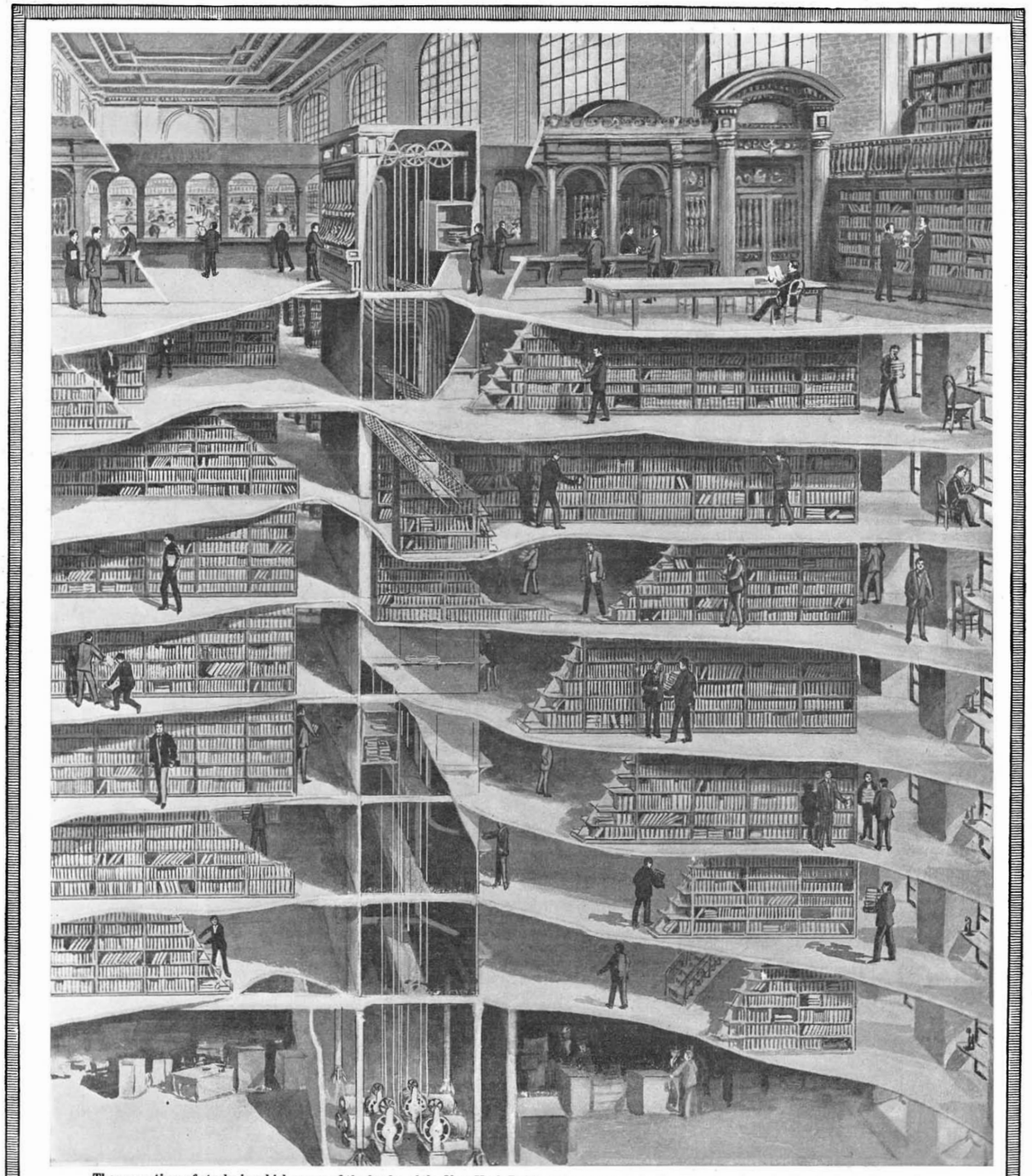


SIXTY-SEVENTH YEAR

SCIENTIFIC AMERICAN

THE WEEKLY JOURNAL OF PRACTICAL INFORMATION

VOLUME CIV.] NUMBER 21 NEW YORK, MAY 27, 1911 [10 CENTS A COPY $3.00 A YEAR

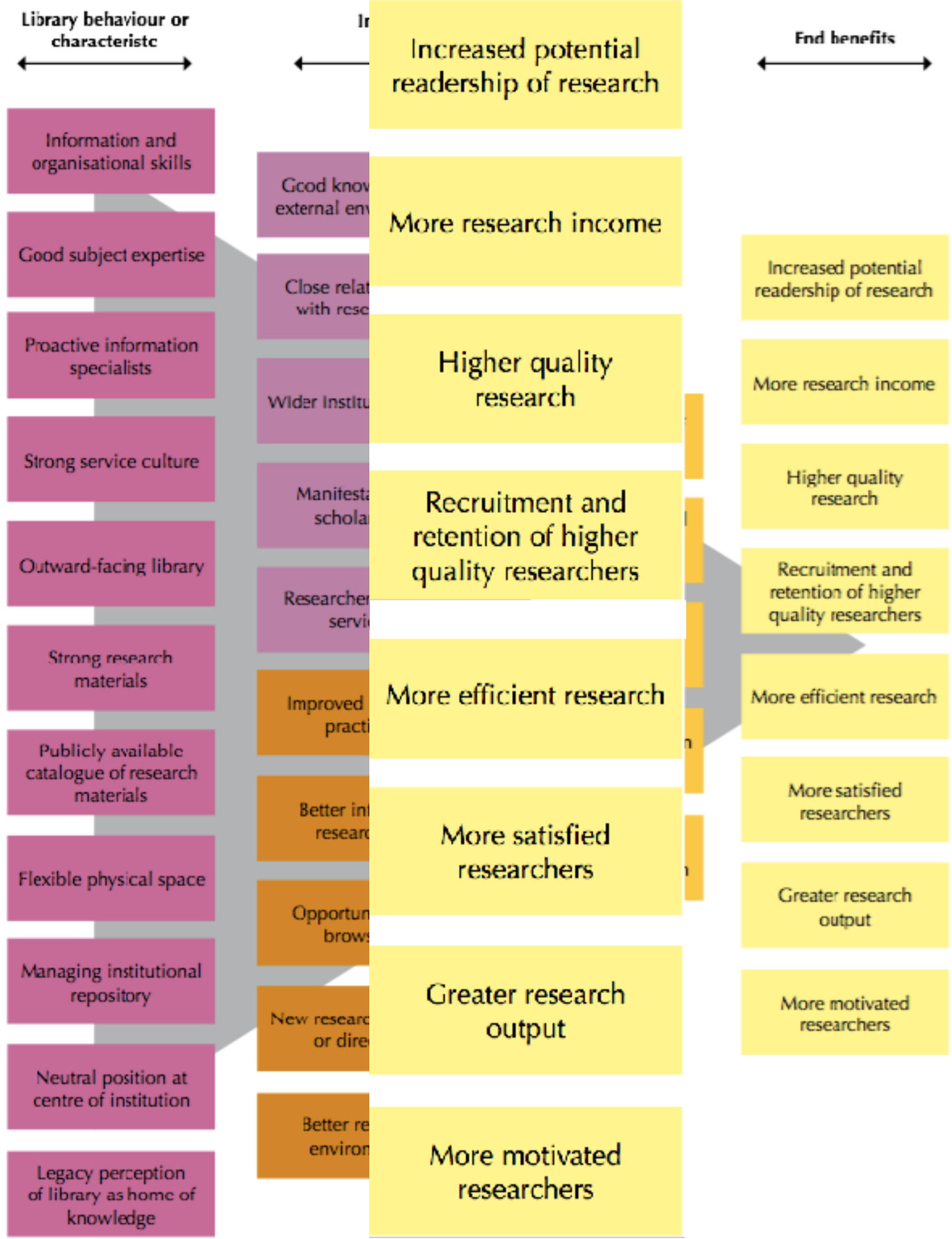


The seven tiers of stacks in which many of the books of the New York Public Library are shelved. Elevators transfer the books to the reading room above as they are called for.

A SECTIONAL VIEW OF THE NEW YORK PUBLIC LIBRARY.—[See page 527.]

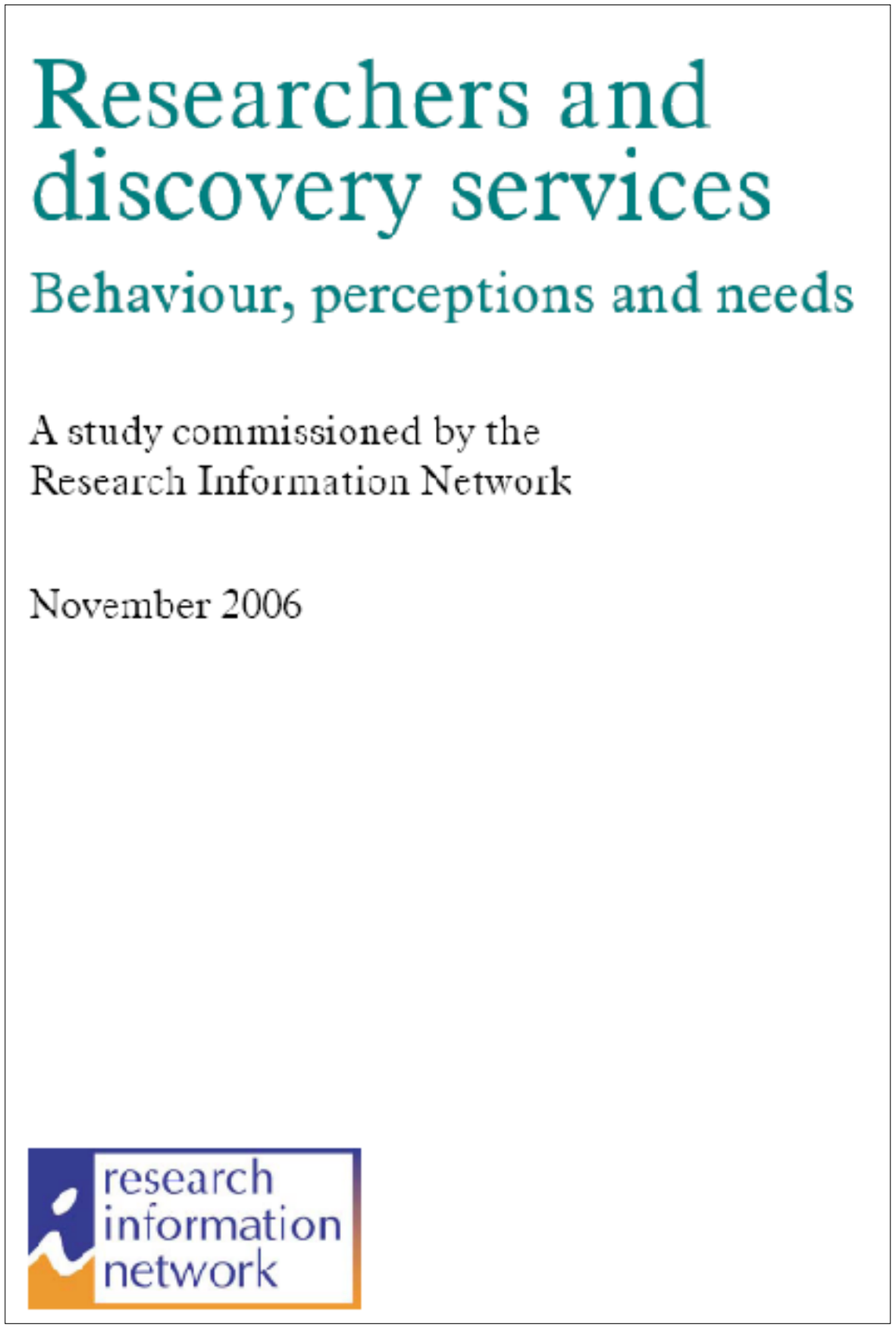

Display a menu

http://d7.library.gatech.edu/research-data/home


Library behaviour or characteristc
End benefits
Information and organisational skills
Good subject expertise
Proactive information specialists
Strong service culture
Outward-facing library
Strong research materials
Publicly available catalogue of research materials
Flexible physical space
Managing institutional repository
Neutral position at centre of institution
Legacy perception of library as home of knowledge
Increased potential readership of research
More research income
Higher quality research
Recruitment and retention of higher quality researchers
More efficient research
More satisfied researchers
Greater research output
More motivated researchers
Increased potential readership of research
More research income
Higher quality research
Recruitment and retention of higher quality researchers
More efficient research
More satisfied researchers
Greater research output
More motivated researchers

What is happening in the world is bypassing university libraries

Peter Murray-Rust
The scientist's view
JISC Libraries of the future debate, April 2009

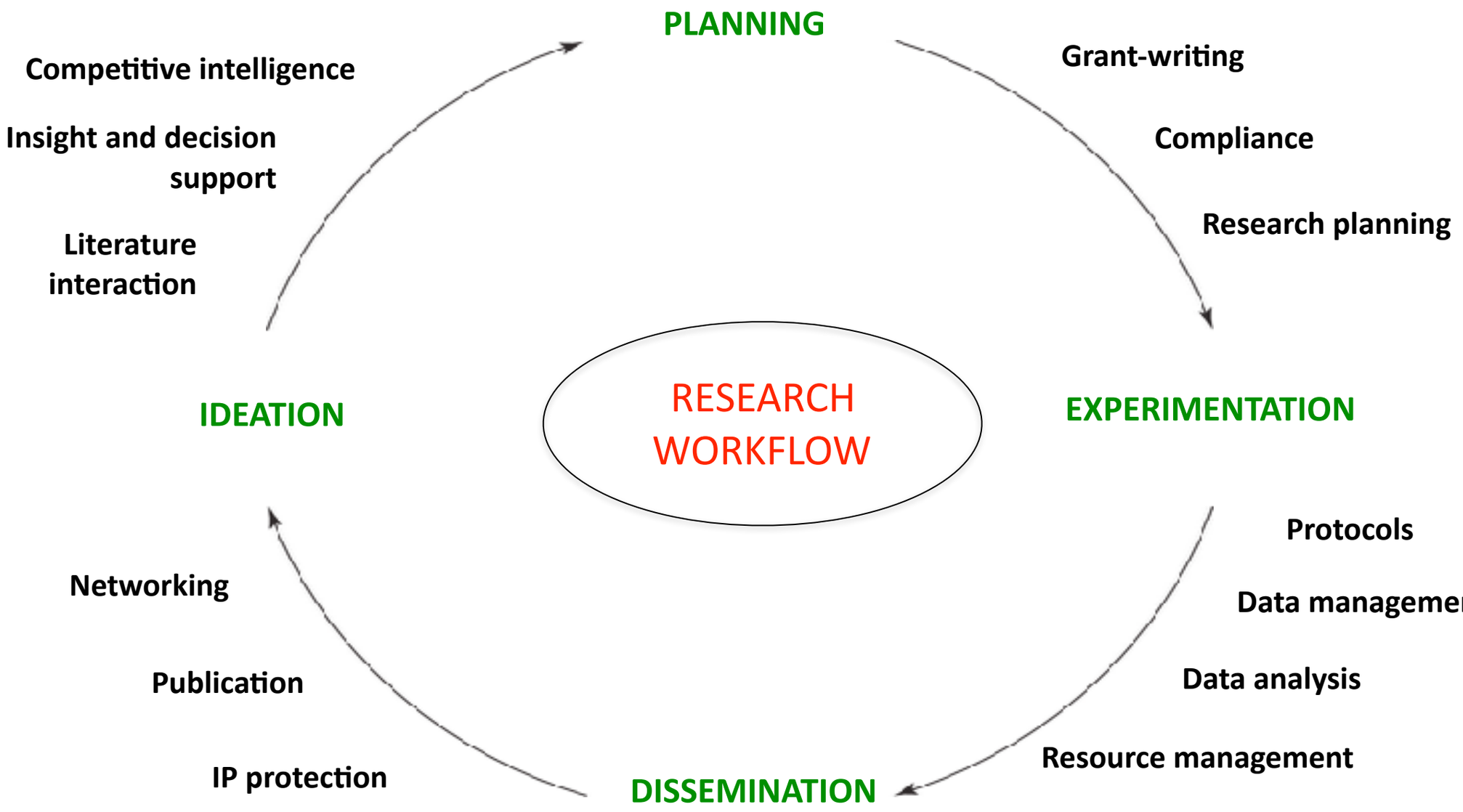


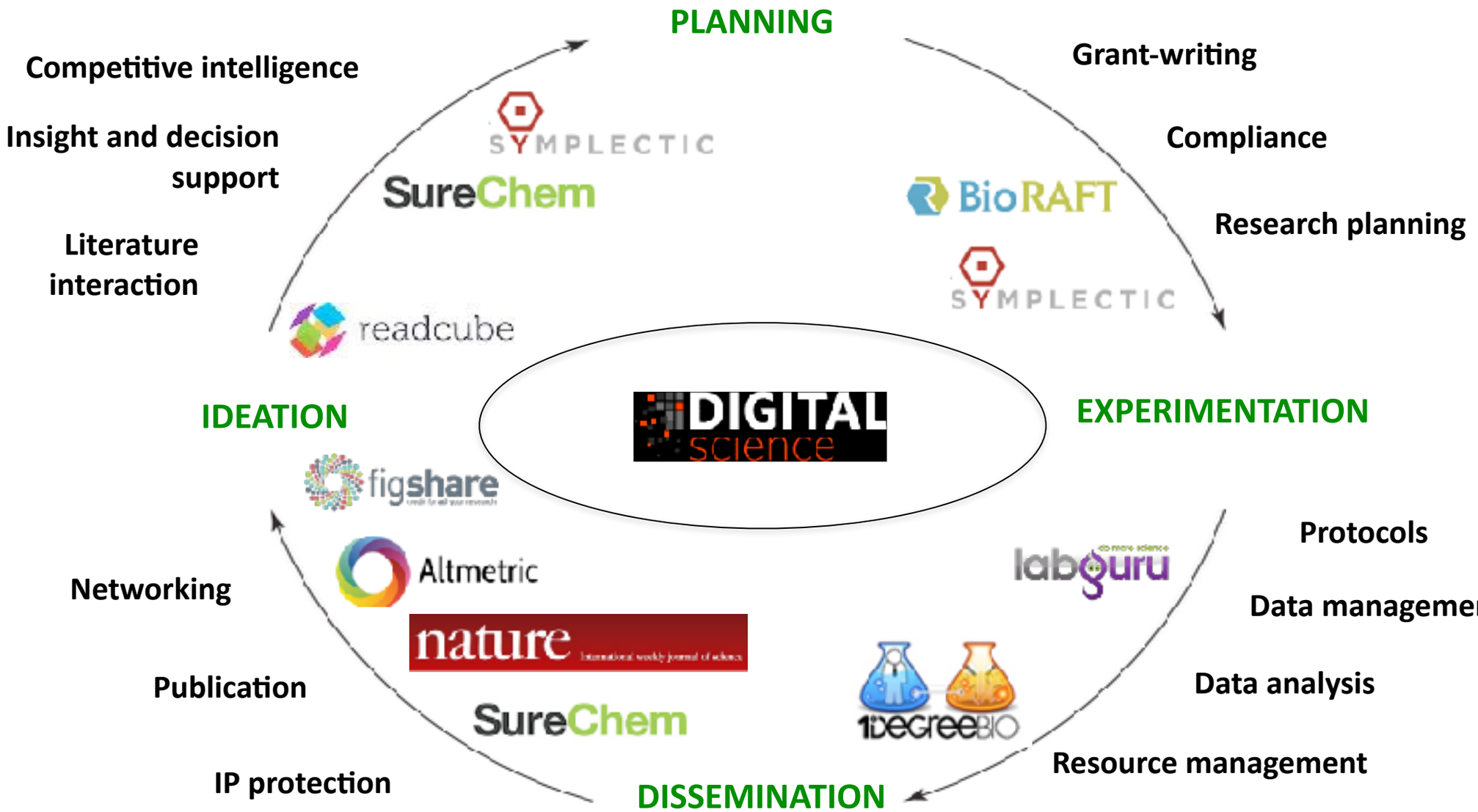
"…contact with librarians and information professionals is rare"

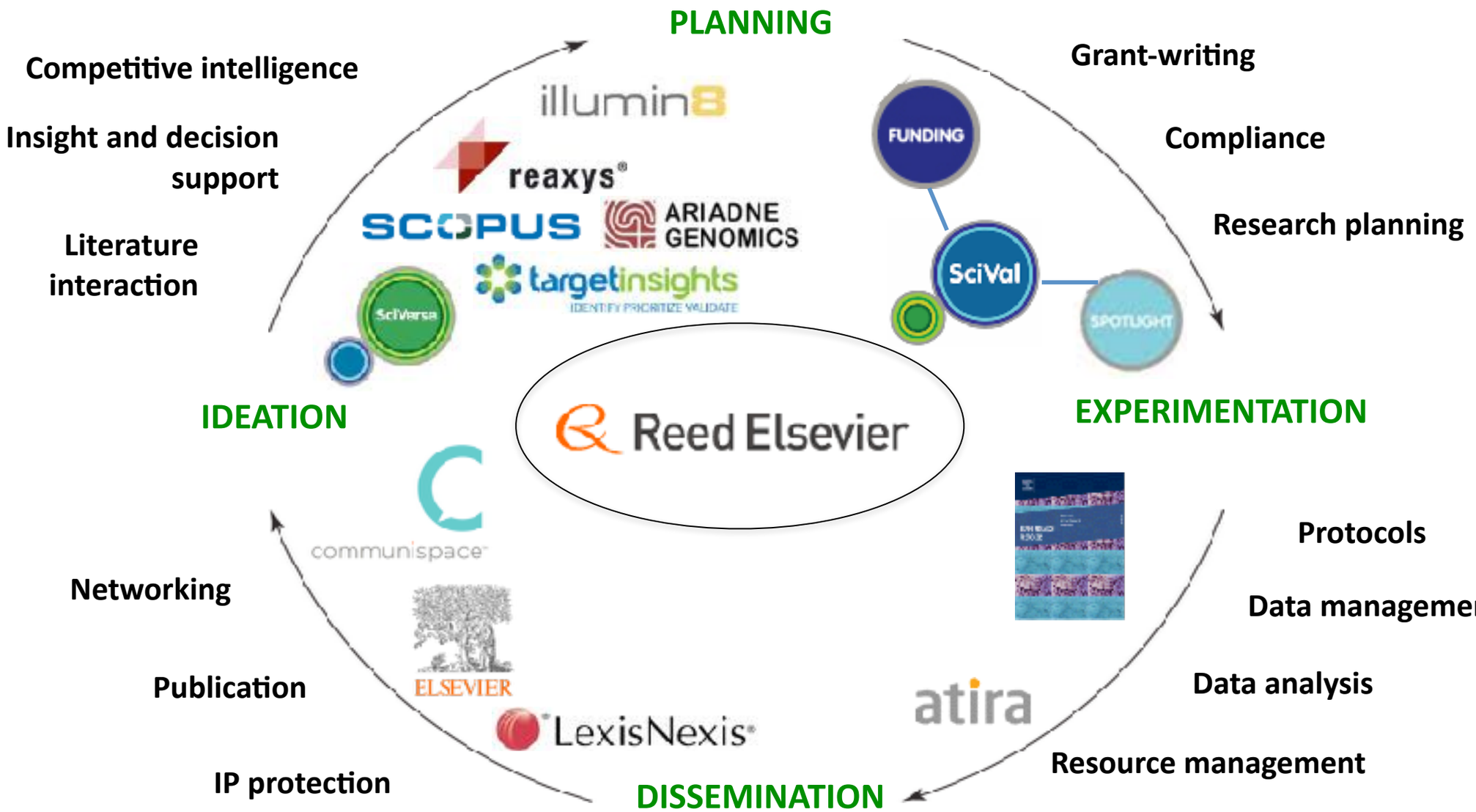
"…researchers are generally confident in their [self-taught] abilities.., librarians see them as..relatively unsophisticated"

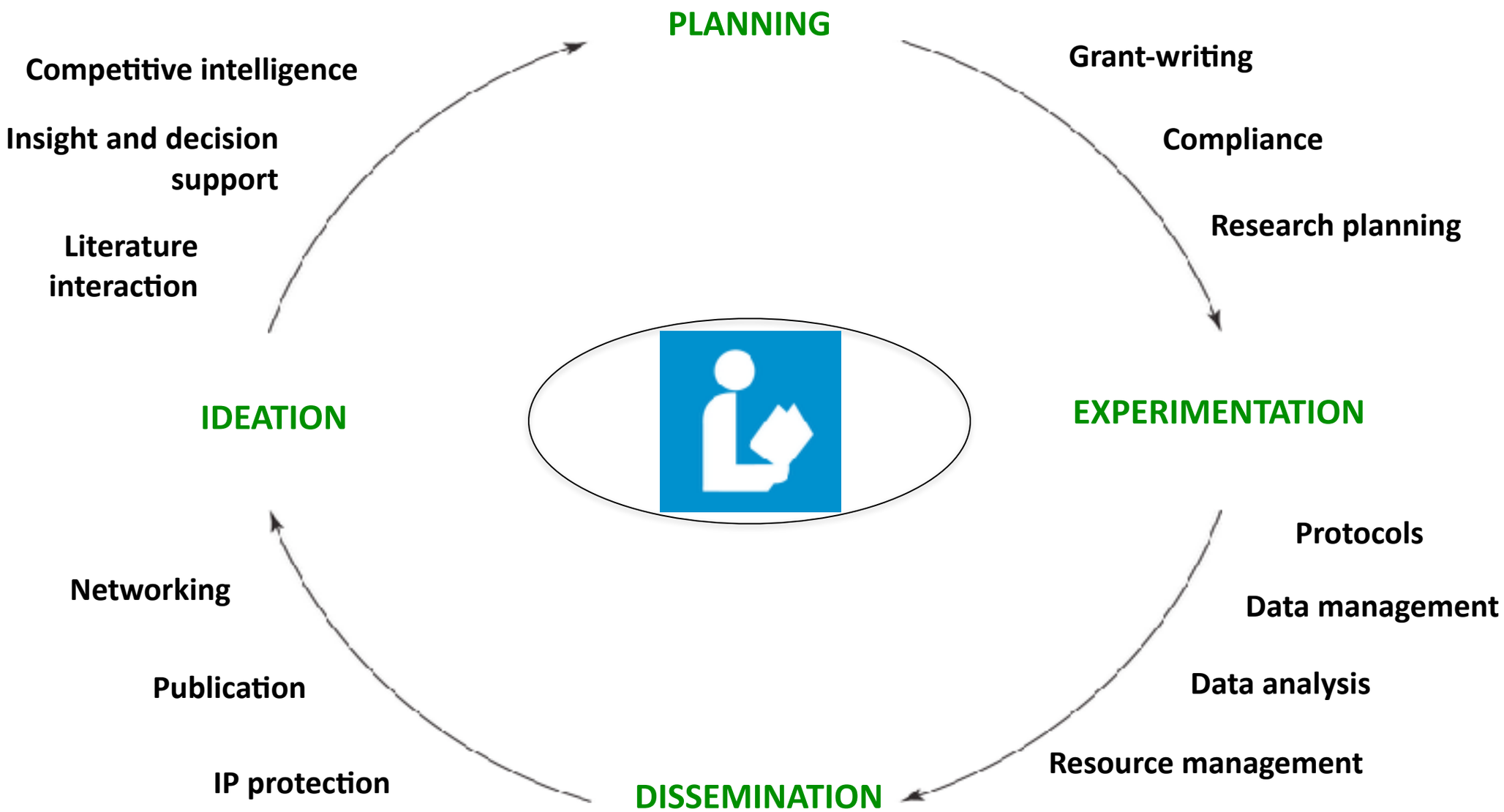
"…librarians see it as a problem that they are not reaching all researchers with formal training, whereas most researchers don't think they need it"

- The part that academic librarians should play remains unclear
- Raise awareness of eResearch amongst library staff
- Provide advice on data management to eResearchers
- Data curation is vast, complex and requires subject input


PLANNING
Competitive intelligence
Insight and decision support
Literature interaction
Grant-writing
Compliance
Research planning
IDEATION
RESEARCH WORKFLOW
EXPERIMENTATION
Networking
Publication
IP protection
DISSEMINATION
Protocols
Data management
Data analysis
Resource management


PLANNING
Competitive intelligence
Insight and decision support
Literature interaction
SYMPLECTIC
SureChem
readcube
Grant-writing
Compliance
Research planning
BioRAFT
SYMPLECTIC
IDEATION
EXPERIMENTATION
DIGITAL science
figshare
Altmetric
nature
SureChem
labguru
1DegreeBio
Networking
Publication
IP protection
Protocols
Data management
Data analysis
Resource management
DISSEMINATION


PLANNING
Competitive intelligence
Insight and decision support
Literature interaction
illumin8
reaxys®
SCOPUS
ARIADNE GENOMICS
targetinsights
IDENTIFY PRIORITIZE VALIDATE
SciVerse
Grant-writing
Compliance
Research planning
FUNDING
SciVal
SPOTLIGHT
IDEATION
Reed Elsevier
EXPERIMENTATION
communispace
Networking
Publication
IP protection
ELSEVIER
LexisNexis
DISSEMINATION
atira
Protocols
Data management
Data analysis
Resource management


PLANNING
Competitive intelligence
Insight and decision support
Literature interaction
Grant-writing
Compliance
Research planning
IDEATION
EXPERIMENTATION
Networking
Publication
IP protection
Protocols
Data management
Data analysis
Resource management
DISSEMINATION


Method
Evidence
Discussion
PROCESS
Outcomes
AFTERMATH
Discussion
Revision
Re-use

METHODSX
You've done the work. Now get the credit.

DRYAD

arXiv.org

PubPeer

slideshare
Present Yourself

figshare
credit for all your research

OXFORD UNIVERSITY PRESS online resource centres

F1000 Posters

101 Innovative tools and sites in 6 research workflow phases (< 2000 - 2015)

January 2015

http://figshare.com/articles/101_Innovations_in_Scholarly_Communication_the_Changing_Research_Workflow/1286826

Fix

Create

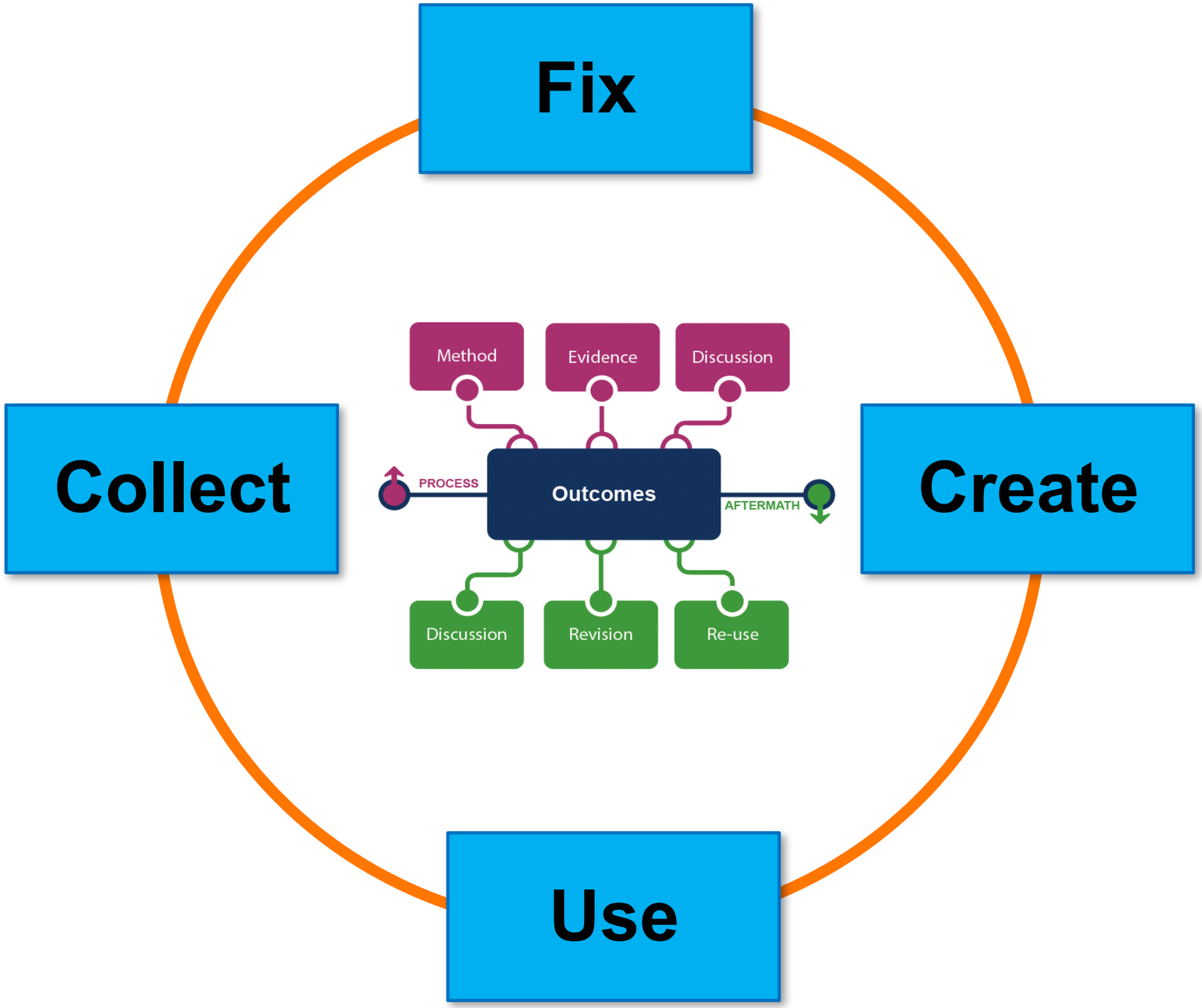
e-literature

Social media (blogs, Twitter)

Social storage (SlideShare, YouTube, Flickr)

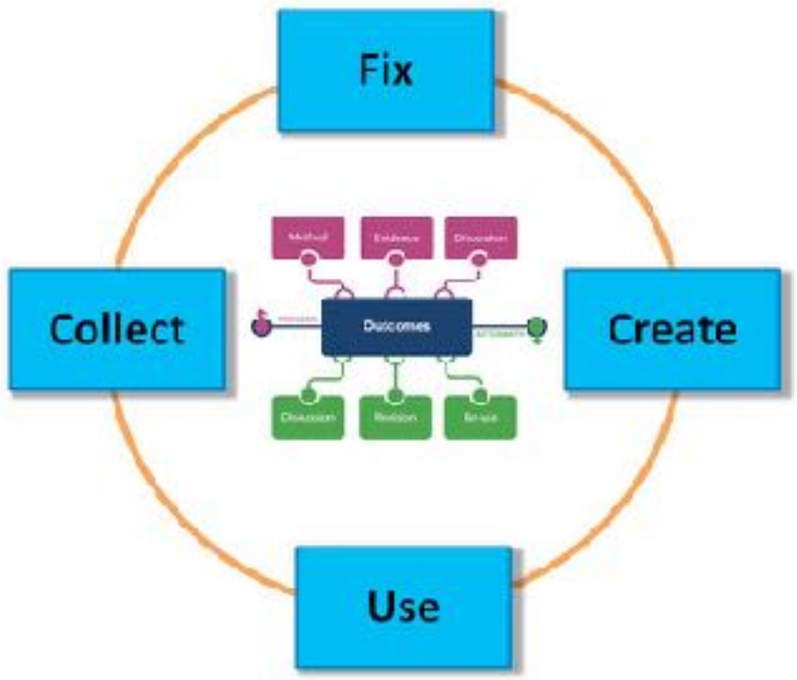
Collect

Use


Fix
Collect
Create
Use
Method
Evidence
Discussion
PROCESS
Outcomes
AFTERMATH
Discussion
Revision
Re-use

A selection of scholarly content …

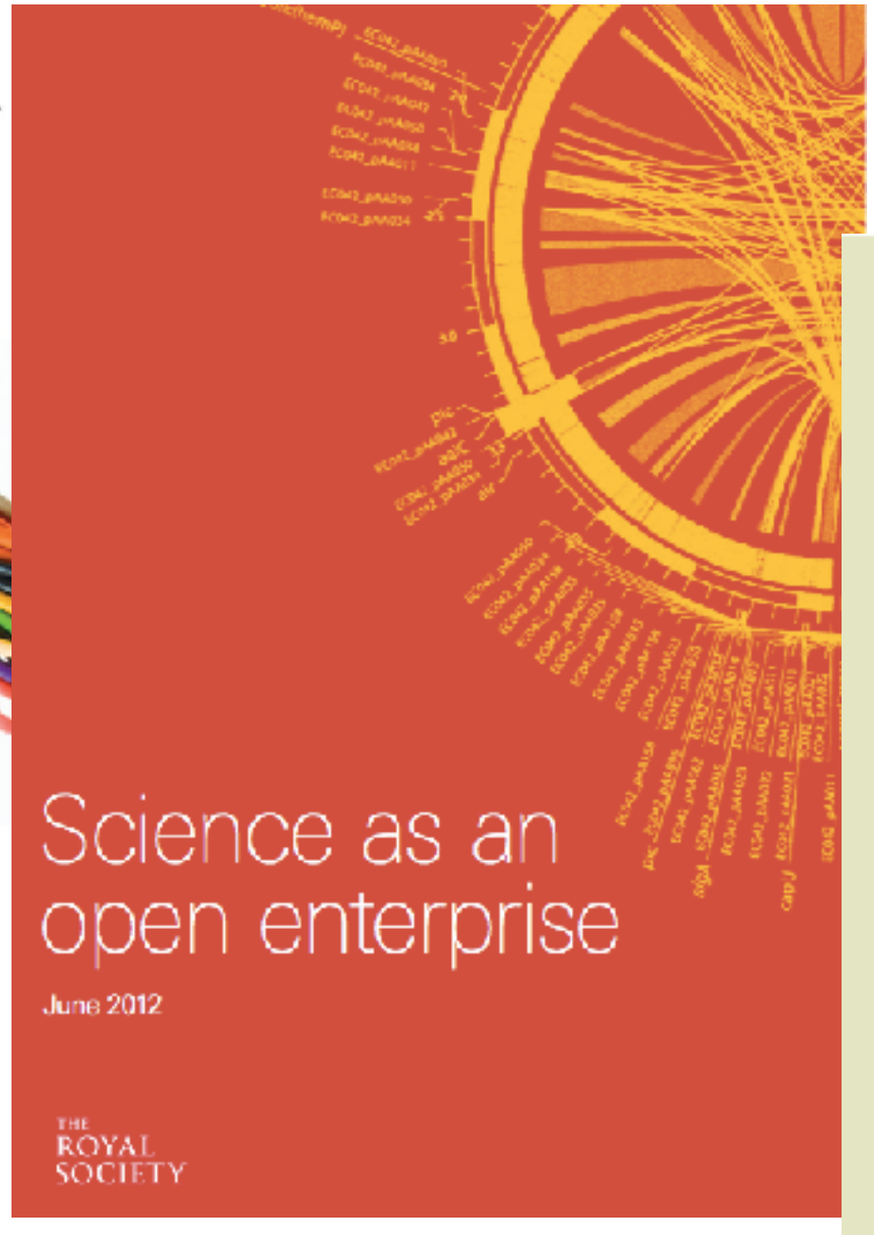
… supported by stable configurations of stakeholder roles



REINVENTING
DISCOVERY

The New Era of Networked Science

MICHAEL NIELSEN



Science as an
open enterprise

June 2012

THE
ROYAL
SOCIETY

Sönke Bartling & Sascha Friesike

opening science

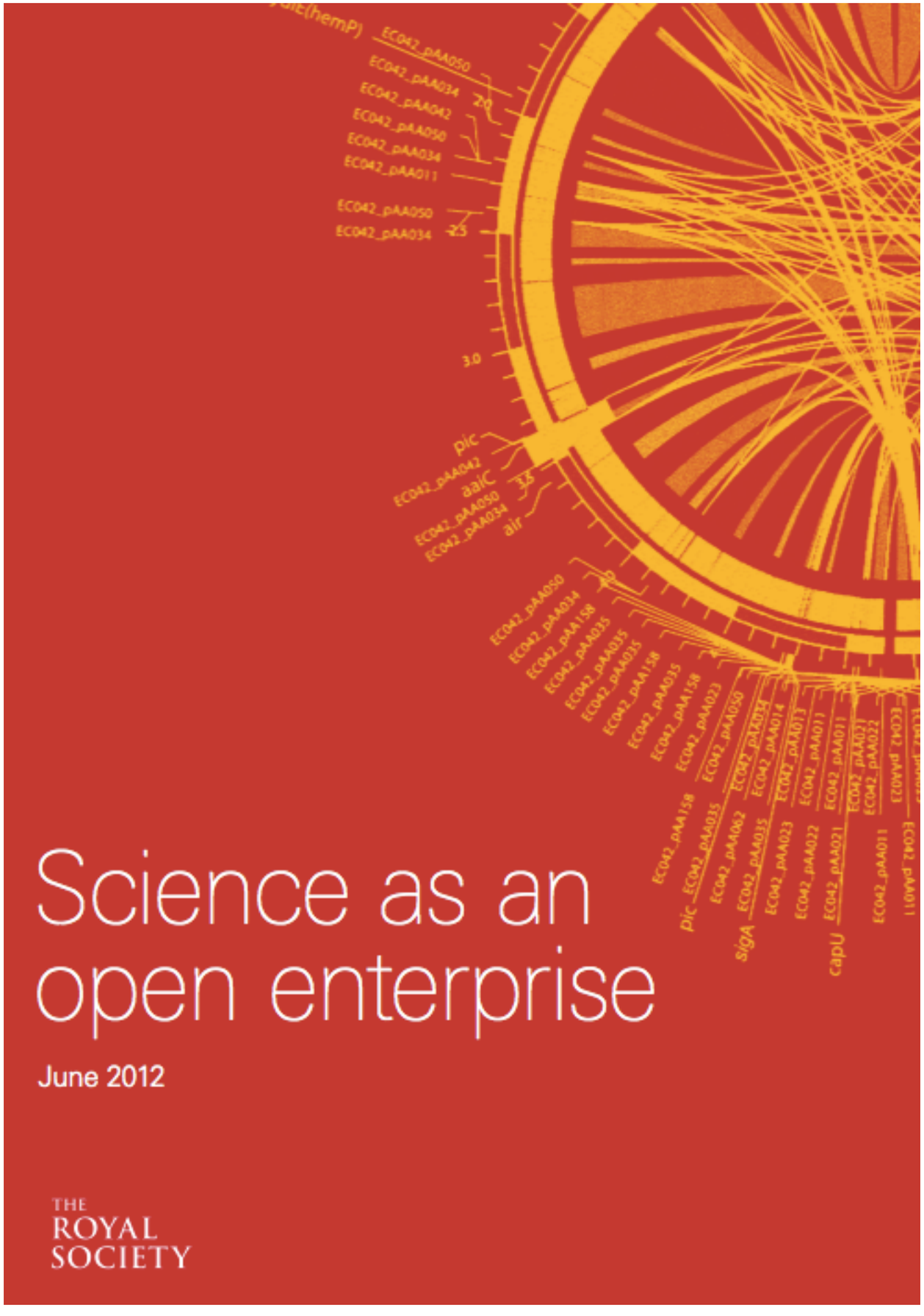
The Evolving Guide on How the Internet is Changing Research,
Collaboration and Scholarly Publishing

Springer Open

Open access, open data, open science

Increasingly, the “private” nature of academic science is being displaced by a culture of openness - ideas, approaches and observations are shared at the earliest opportunity with colleagues - and sometimes the world at large.

Whilst the ‘version of record’ approach to journal article creation retains validity, this is increasingly seen as a compliance matter - required to meet career objectives and funder/government requirements



The New York Times

January 16, 2012

Cracking Open the Scientific Process

By THOMAS LIN

The New England Journal of Medicine marks its 200th anniversary this year with a timeline celebrating the scientific advances first described in its pages: the stethoscope (1816), the use of ether for anesthesia (1846), and disinfecting hands and instruments before surgery (1867), among others.

For centuries, this is how science has operated — through research done in private, then submitted to science and medical journals to be reviewed by peers and published for the benefit of other researchers and the public at large. But to many scientists, the longevity of that process is nothing to celebrate.

The system is hidebound, expensive and elitist, they say. Peer review can take months, journal subscriptions can be prohibitively costly, and a handful of gatekeepers limit the flow of information. It is an ideal system for sharing knowledge, said the quantum physicist Michael Nielsen, only "if you're stuck with 17th-century technology."

Dr. Nielsen and other advocates for "open science" say science can accomplish much more, much faster, in an environment of friction-free collaboration over the Internet. And despite a host of obstacles, including the skepticism of many established scientists, their ideas are gaining traction.

Open-access archives and journals like arXiv and the Public Library of Science (PLoS) have sprung up in recent years. GalaxyZoo, a citizen-science site, has classified millions of objects in space, discovering characteristics that have led to a raft of scientific papers.

On the collaborative blog MathOverflow, mathematicians earn reputation points for contributing to solutions; in another math experiment dubbed the Polymath Project, mathematicians commenting on the Fields medalist Timothy Gower's blog in 2009 found a new proof for a particularly complicated theorem in just six weeks.

And a social networking site called ResearchGate — where scientists can answer one another's questions, share papers and find collaborators — is rapidly gaining popularity.

theguardian

News | US | World | Sports | Comment | Culture | Business | Environn

Comment is free

Open, free access to academic research? This will be a seismic shift

Opening up access to academic research will put more data and power in the hands of the people who pay for it

David Willetts

The Guardian, Tuesday 1 May 2012 16.00 EDT

Jump to comments (141)

Wikipedia's Jimmy Wales will be helping ensure that the publicly funded portal promotes collaboration and engagement. Photograph: Peter Macdiarmid/Getty

My department spends about £5bn each year funding academic research – and it is because we believe in the fundamental importance of this research that we have protected the science budget for the whole of this parliament.

Adam Bly

About 35 percent of scientists are using things like blogs to consume and produce content. There is an explosion of online tools and platforms available to scientists, ranging from Web 2.0 tools modified or created for the scientific world to Web sites that are doing amazing things with video, lab notebooks, and social networking.

The next generation of PIs is already establishing new behaviors. They feel comfortable blogging, using social media tools, and using wikis to advance their research. It will take the big institutions to support open-access journals, for example. And it will take technological innovation in the form of software that is purpose-built for this unique community and its set of challenges.

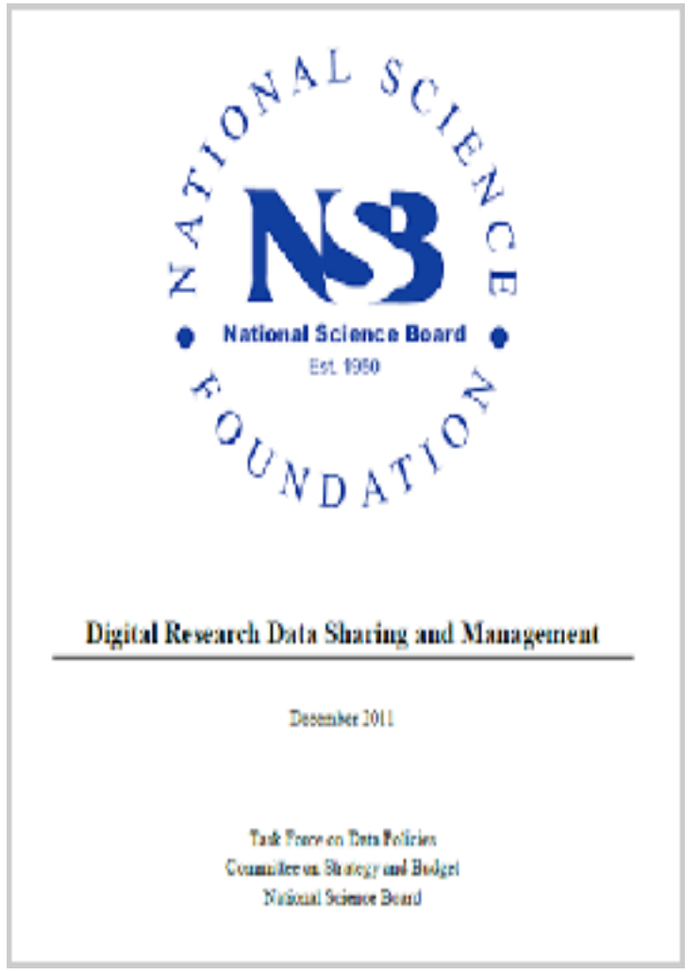
We're talking about something as fundamental and important as modernizing the architecture of science.

http://seedmagazine.com/content/article/science_2.0_pioneers/

Why Data Management Services?

"The Board believes that **timely attention to digital research data sharing and management** is fundamental to supporting U.S. science and engineering in the twenty-first century.

...strong and **sustainable data sharing and management policies [are] a critical national need**."



Digital Research Data Sharing and Management
December 2011
Task Force on Data Policies
Committee on Strategy and Budget
National Science Board

Sharing data?

- Create opportunities
  - For re-analysis and re-use
  - To facilitate collaboration
- Solve problems
  - Waste of money, people and effort
  - Loss of irretrievable data
  - Inability to verify research
- Issues and challenges
  - Patient confidentiality
  - IP and discovery protection
- Promote curation rather than sharing?

nature Vol 461|10 September 2009

OPINION

Prepublication data sharing

Rapid release of prepublication data has served the field of genomics well. Attendees at a workshop in Toronto recommend extending the practice to other biological data sets.

Open discussion of ideas and full disclosure of supporting facts are the bedrock for scientific discourse and new developments. Traditionally, published papers combine the salient ideas and the supporting facts in a single discrete 'package'. With the advent of methods for large-scale and high-throughput data analyses, the generation and transmission of the underlying facts are often replaced by an electronic process that involves sending information to and from scientific databases. For such data-intensive projects, the standard requirement is that all relevant data must be made available on a publicly accessible website at the time of a paper's publication[1].

One of the lessons from the Human Genome Project (HGP) was the recognition that making data broadly available before publication can be profoundly valuable to the scientific enterprise and lead to public benefits. This is particularly the case when there is a community of scientists that can productively use the data quickly — beyond what the data producers could do themselves in a similar time period, and sometimes for scientific purposes outside the original goals of the project.

The principles for rapid release of genome-sequence data from the HGP were formulated at a meeting held in Bermuda in 1996; these were then implemented by several funding agencies. In exchange for 'early release' of their data, the international sequencing centres retained the right to be the first to describe and analyse their complete data sets in peer-reviewed publications. The draft human genome sequence[2] was the highest profile data set rapidly released before publication, with sequence assemblies greater than 1,000 base pairs usually released within 24 hours of generation. This experience demonstrated that the broad and early availability of sequence data greatly benefited life sciences research by leading to many new insights and discoveries[2], including new information on 30 disease genes published prior to the draft sequence.

At a time when advances in DNA sequencing technologies mean that many more laboratories can produce massive data sets, and when an ever-growing number of fields (beyond genome sequencing) are grappling with their own data-sharing policies, a Data Release Workshop was convened in Toronto in May 2009 by Genome Canada and other funding agencies. The meeting brought together a diverse and multinational group of scientists, ethicists, lawyers, journal editors and funding representatives. The goal was to reaffirm and refine, where needed, the policies related to the early release of genomic data, and to extend, if possible, similar data-release policies to other types of large biological data sets — whether from proteomics, biobanking or metabolite research.

Building on the past

By design, the Toronto meeting continued policy discussions from previous meetings, in particular the Bermuda meetings (1996, 1997 and 1998)[3–5] and the 2003 Fort Lauderdale meeting, which recommended that rapid prepublication release be applied to other data sets whose primary utility was a resource for the scientific community, and also established the responsibilities of the resource producers, resource users, and the funding agencies[6]. A similar 2008 Amsterdam meeting extended the principle of rapid data release to proteomics data[7]. Although the recommendations of these earlier meetings can apply to many genomics and proteomics projects, many outside the major sequencing centres and funding agencies remain unaware of the details of these policies, and so one goal of the Toronto meeting was to reaffirm the existing principles for early data release with a wider group of stakeholders.

In Toronto, attendees endorsed the value of rapid prepublication data release for large reference data sets in biology and medicine that have broad utility and agreed that prepublication data release should go beyond genomics and proteomics studies to other data sets — including chemical structure, metabolomic and RNA interference data sets, and to annotated clinical resources (cohorts, tissue banks and case-control studies). In each of these domains, there are diverse data types and study designs, ranging from the large-scale 'community resource projects' first identified at Fort Lauderdale (for which meeting participants endorsed prepublication data release) to investigator-led hypothesis-testing projects (for which the minimum standard should be the release of generated data at the time of publication).

Several issues discussed at previous data-

EXAMPLES OF PREPUBLICATION DATA-RELEASE GUIDELINES

Project type	Prepublication data release recommended	Prepublication data release optional
Genome sequencing	Whole-genome or mRNA sequence(s) of a reference organism or tissue	Sequences from a few loci for cross-species comparisons in a limited number of samples
Polymorphism discovery	Catalogue of variants from genomic and/or transcriptomic samples in one or more populations	Variants in a gene, a gene family or a genomic region in selected pedigrees or populations
Genetic association studies	Genomewide association analysis of thousands of samples	Genotyping of selected gene candidates
Somatic mutation discovery	Catalogue of somatic mutations in exomes or genomes of tumour and non-tumour samples	Somatic mutations of a specific locus or limited set of genomic regions
Microbiome studies	Whole-genome sequence of microbial communities in different environments	Sequencing of target locus in a limited number of microbiome samples
RNA profiling	Whole-genome expression profiles from a large panel of reference samples	Whole-genome expression profiles of a perturbed biological system(s)
Proteomic studies	Mass spectrometry data sets from large panels of normal and disease tissues	Mass spectrometry data sets from a well-defined and limited set of tissues
Metabolomic studies	Catalogue of metabolites in one or more tissues of an organism	Analyses of metabolites induced in a perturbed biological system(s)
RNAi or chemical library screen	Large-scale screen of a cell line or organism analysed for standard phenotypes	Focused screens used to validate a hypothetical gene network
3D-structure elucidation	Large-scale cataloguing of 3D structures of proteins or compounds	3D structure of a synthetic protein or compound elucidated in the context of a focused project

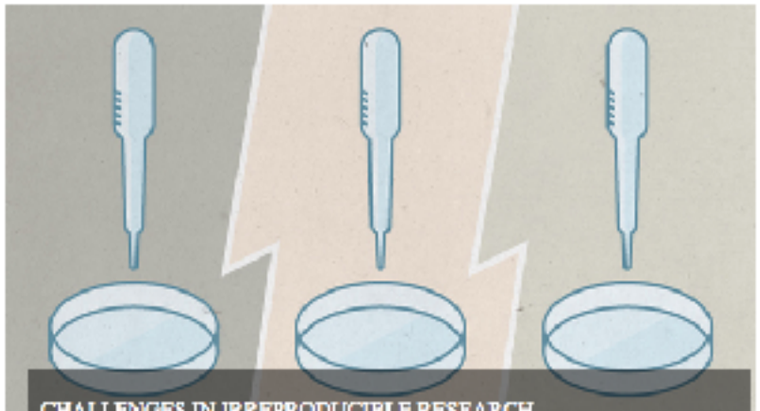
- Builds upon work in Fort Lauderdale biological data sharing principles

http://www.nature.com/nature/journal/v461/n7261/pdf/461168a.pdf

nature
International weekly journal of science

SPECIAL

See all specials



CHALLENGES IN IRREPRODUCIBLE RESEARCH

No research paper can ever be considered to be the final word, and the replication and corroboration of research results is key to the scientific process. In studying complex entities, especially animals and human beings, the complexity of the system and of the techniques can all too easily lead to results that seem robust in the lab, and valid to editors and referees of journals, but which do not stand the test of further studies. Nature has published a series of articles about the worrying extent to which research results have been found wanting in this respect. The editors of Nature and the Nature life sciences research journals have also taken substantive steps to put our own houses in order, in improving the transparency and robustness of what we publish. Journals, research laboratories and institutions and funders all have an interest in tackling issues of irreproducibility. We hope that the articles contained in this collection will help.

Free full access

Editorial | Features | News and analysis | Comment

The Washington Post

Health & Science

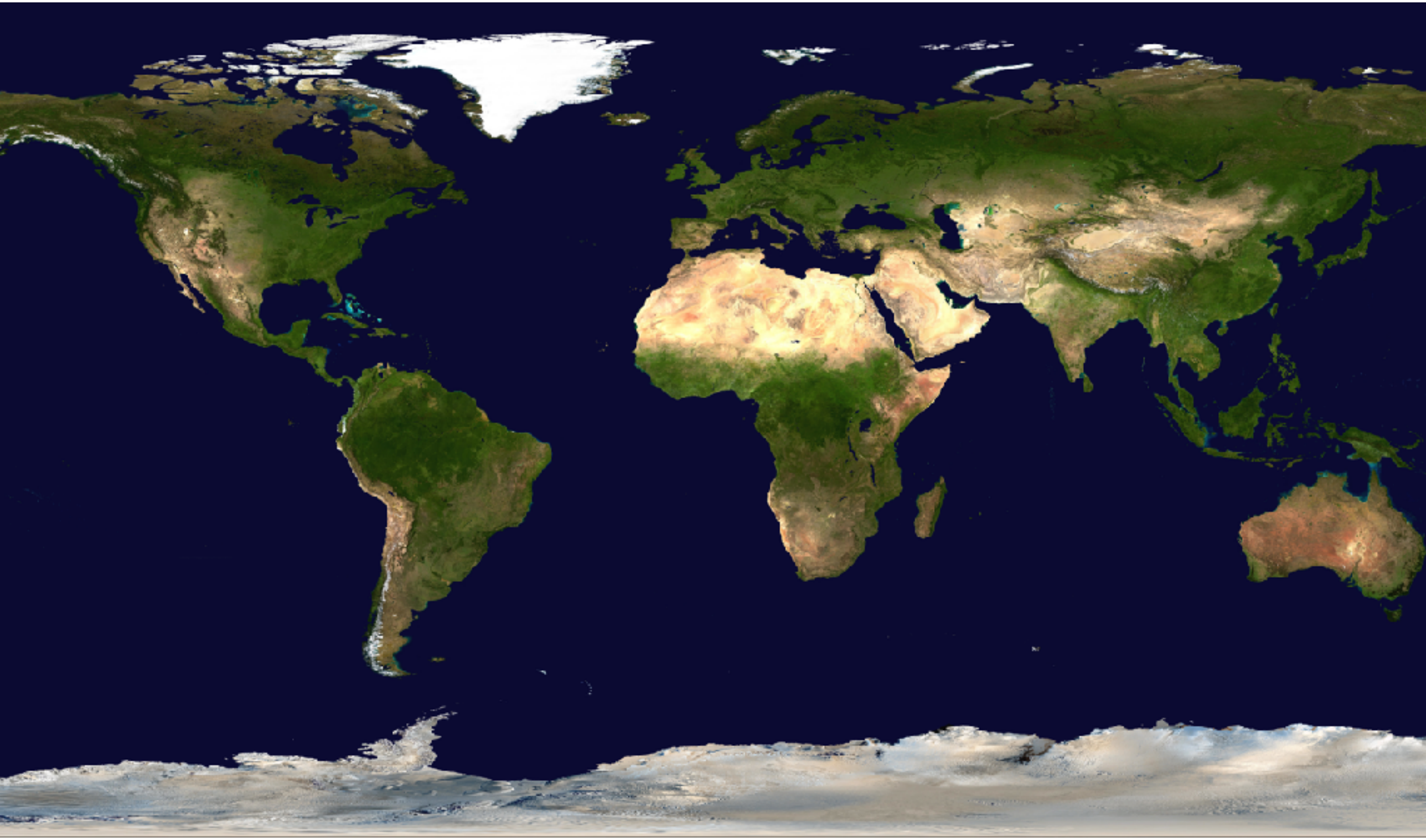
The new scientific revolution: Reproducibility at last

By Joel Achenbach January 27 Follow @joelachenbach

Diederik Stapel, a professor of social psychology in the Netherlands, had been a rock-star scientist — regularly appearing on television and publishing in top journals. Among his striking discoveries was that people exposed to litter and abandoned objects are more likely to be bigoted.

And yet there was often something odd about Stapel's research. When students asked to see the data behind his work, he couldn't produce it readily. And colleagues would sometimes look at his data and think: It's beautiful. Too beautiful. Most scientists have messy data, contradictory data, incomplete data, ambiguous data. This data was too good to be true.

REPRODUCIBILITY INITIATIVE



List of all research funders

Click on the relevant column heading to re-sort by Country

Funding Organisation (*linked to their JULIET summary*)	Publications Policy: Open Access Publishing	Publications Policy: Open Access Archiving	Data Archiving Policy	Country
中国科学院 **see**: Chinese Academy of Sciences (CAS)				China
国家自然科学基金委员会 **see**: National Natural Science Foundation of China (NSFC)				China
Académie suisse des sciences humaines et sociales (ASSH) **see**: Schweizerische Akademie der Geistes- und Sozialwissenschaften (SAGW)				Switzerland
Academy of Finland	OA	OA — —	OA —	Finland
Action Medical Research				United Kingdom
Action on Hearing Loss		OA OA —		United Kingdom
Age UK				United Kingdom
Agence Nationale de la Recherche (ANR)		OA — —		France
Alliance canadienne pour la recherche sur le cancer du sein (ACRCS) **see**: Canadian Breast Cancer Research Alliance (CBCRA)				Canada
Alzheimer's Research UK				United Kingdom
Alzheimer's Society	—	OA OA —		United Kingdom
Arthritis Research UK		OA OA		United Kingdom
Arts and Humanities Research Council (AHRC)	OA	OA OA —	OA —	United Kingdom
Arts and Humanities Research Council [2006-2012] (AHRC [2006-2012])		OA — —	OA —	United Kingdom
Asthma UK				United Kingdom
Ataxia UK				United Kingdom
Australian Research Council (ARC)		— — —	— —	Australia
Australian Research Council [2013] (ARC)	OA	OA OA		Australia
Austrian Science Fund (FWF)	OA	OA OA —	OA —	Austria
Autism Speaks		OA — —		United States
Avon Foundation for Women		OA OA		Canada
Bill and Melinda Gates Foundation	OA	OA OA —	OA —	United States
Biotechnology and Biological Sciences Research Council (BBSRC)	OA	OA OA —	OA OA	United Kingdom
Biotechnology and Biological Sciences Research Council [2006-2012] (BBSRC [2006-2012])		OA	OA OA	United Kingdom

JULIET assigns up to two Open Access ticks for data archiving policies:

OA Data archiving is required.

OA Data must be deposited within five years.

RCUK Common Principles on Data Policy

Making research data available to users is a core part of the Research Councils' remit and is undertaken in a variety of ways. We are committed to transparency and to a coherent approach across the research base. These RCUK common principles on data policy provide an overarching framework for individual Research Council policies on data policy.

Principles

- Publicly funded research data are a public good, produced in the public interest, which should be made openly available with as few restrictions as possible in a timely and responsible manner.
- Institutional and project specific data management policies and plans should be in accordance with relevant standards and community best practice. Data with acknowledged long-term value should be preserved and remain accessible and usable for future research.
- To enable research data to be discoverable and effectively re-used by others, sufficient metadata should be recorded and made openly available to enable other researchers to understand the research and re-use potential of the data. Published results should always include information on how to access the supporting data.
- RCUK recognises that there are legal, ethical and commercial constraints on release of research data. To ensure that the research process is not damaged by inappropriate release of data, research organisation policies and practices should ensure that these are considered at all stages in the research process.
- To ensure that research teams get appropriate recognition for the effort involved in collecting and analysing data, those who undertake Research Council funded work may be entitled to a limited period of privileged use of the data they have collected to enable them to publish the results of their research. The length of this period varies by research discipline and, where appropriate, is discussed further in the published policies of individual Research Councils.
- In order to recognise the intellectual contributions of researchers who generate, preserve and share key research datasets, all users of research data should acknowledge the sources of their data and abide by the terms and conditions under which they are accessed.
- It is appropriate to use public funds to support the management and sharing of publicly-funded research data. To maximise the research benefit which can be gained from limited budgets, the mechanisms for these activities should be both efficient and cost-effective in the use of public funds.

Further guidance and information on the individual principles can be found within the guidance documentation.

A blog piece on supporting research data management costs through grant funding, including responses to questions raised at the DCC/RDMF Special Event on funding for Research Data Management, is available here.

This policy was published in April 2011 and revised July 2015.

Key points

- Publicly funded research data are a public good, produced in the public interest, which should be made openly available with as few restrictions as possible in a timely and responsible manner that does not harm intellectual property.
- To ensure that the research process is not damaged by inappropriate release of data, research organisation policies and practices should ensure that these are considered at all stages in the research process.

Policy on data management and sharing

The Wellcome Trust is committed to ensuring that the outputs of the research it funds, including research data, are managed and used in ways that maximise public benefit. Making research data widely available to the research community in a timely and responsible manner ensures that these data can be verified, built upon and used to advance knowledge and its application to generate improvements in health.

We believe that success in maximising the value of research data depends crucially on fostering a culture in which both data generators and data users adopt good research practice, and act with integrity and transparency in managing, using and sharing research data. Researchers, research institutions, resource providers, funders and publishers all have important roles to play in helping to develop this enabling environment and in developing the resources and systems required.

There is a growing international consensus on the need to preserve and share research datasets in a manner that maximises their long-term value. This has been articulated in key documents such as the UK Concordat on Open Research Data (2016), the OECD Principles and Guidelines for Access to Research Data from Public Funding (2007); the Fort Lauderdale Principles (2003) and the Toronto Statement (2009).

Our policy on data management and sharing reflects the principles set out in these statements. It is consistent with the Trust's position on open and unrestricted access to published research, our guidelines on good research practice and our policy on intellectual property and patenting. The policy should be read in conjunction with the associated guidance. We will keep both our policy and guidance under review.

Policy statement

1. The Wellcome Trust expects all of its funded researchers to maximise the availability of research data with as few restrictions as possible.

CONTACT US

If you have questions about any of our policies, contact the policy team.

● Full Coverage ◐ Partial Coverage ○ No Coverage

	Policy Coverage		Policy Stipulations					Support Provided			
Research Funders	Published outputs	Data	Time limits	Data plan	Access/ sharing	Long-term curation	Monitoring	Guidance	Repository	Data centre	Costs
AHRC	●	●	●	●	●	◐	○	●	○	◐	◐
BBSRC	●	●	●	●	●	●	●	●	●	◐	●
CRUK	●	●	●	●	●	●	●	◐	●	○	○
EPSRC	●	●	●	◐	●	●	●	◐	○	○	●
ESRC	●	●	●	●	●	●	●	●	●	●	◐
MRC	●	●	●	●	●	●	○	◐	●	○	◐
NERC	●	●	●	●	●	●	●	●	●	●	◐
STFC	●	●	●	●	●	●	●	◐	●	◐	◐
Wellcome Trust	●	●	●	●	●	●	●	●	●	◐	●

UK • DATA ARCHIVE

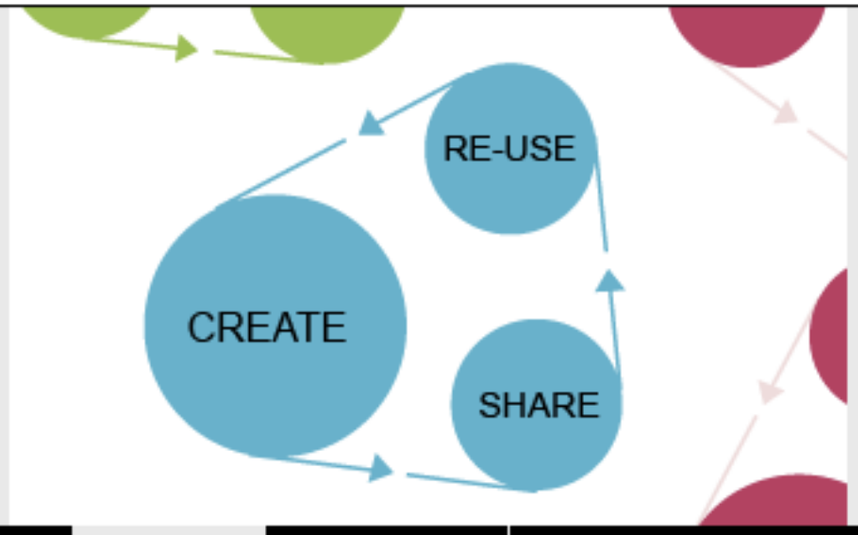
THE UK'S LARGEST COLLECTION OF DIGITAL RESEARCH DATA IN THE SOCIAL SCIENCES AND HUMANITIES

HOME | ABOUT US | CREATE & MANAGE DATA | DEPOSIT DATA | HOW WE CURATE DATA | FIND DATA | NEWS & EVENTS

SEARCH OUR SITE GO

FECYCLE

search data lifecycle
nd technology to
process of data
and use



FIND DATA

We provide acces
data collections fo
teaching covering
key economic and

READ ON

What's new | Data lifecycle | Find data | Who are we?

FIRST TIME HERE?

HELPFUL INFORMATION

A QUICK GUIDE TO THE ARCHIVE

9 of 10: You won't be able to find your family records on our site. You should visit a genealogical site instead

WHO GIVES US DATA?

Find out what kind of data are available to you

DEPOSITING YOUR DATA

Depositing your data with the Archive ensures that they will be professionally curated and accessible

DEPOSIT DATA

FINDING DATA TO USE

We can help you find data for research and teaching with our catalogue of over 5,000 data collections

OUR CATALOGUE

OUR DATA IN USE

Labour Force Survey: How does the health service distribute resources equitably...

OUR DATA IN USE

OUR SERVICES

SQB
providing access to key survey questionnaires

VIEW SITE

CITING DATA

Proper citation should be a significant feature of research publications.

PREPARING YOUR DATA

Prepare your data from the start of research so that they can be shared and re-used in the future

LATEST NEWS & EVENTS

A new way to search for data
Visitors to the ESDS website will, from today, ...

Staff vacancies Census Microdata
The Enhancing and Enriching Historic Census ...

Social science challenge
Comedienne Hazel Humphreys will be testing the ...

User Interface Developer

EUROPEAN COMMISSION
Directorate-General for Research & Innovation

H2020 Programme

Guidelines on

FAIR Data Management in Horizon 2020

Version 3.0
26 July 2016

Open Research Data Pilot in Horizon 2020
How can OpenAIRE help?

Briefing paper for Researchers, Project coordinators, and Research Managers participating in the EC Open Research Data Pilot in Horizon 2020.

What is the Open Research Data Pilot?

Open data is data that is free to use, reuse, and redistribute. The Open Research Data Pilot aims to make the research data generated by selected Horizon 2020 projects open. It will be carefully monitored and used to inform future EC policy.

If your Horizon 2020 project is part of the pilot, you must:

- Develop (and keep up-to-date) a Data Management Plan (DMP).
- Deposit your data in a research data repository.
- Make sure third parties can freely access, mine, exploit, reproduce and disseminate it.
- Make clear what tools will be needed to use the raw data to validate research results (or provide the tools themselves).

The pilot applies to (1) the data (and metadata) needed to validate results in scientific publications, and (2) other curated and/or raw data (and metadata) that you specify in the DMP.

What's in it for you?

- Be part of the new era of Open Science, integrating transparency, effectiveness and timeliness into all areas of scientific methods and processes.
- Reach more people, have greater impact.
- Avoid duplication of effort and help preserve data for future researchers.
- Simplify final reporting thanks to an up-to-date DMP.

OpenAIRE

Do you have a Horizon 2020 project grant?
Check your grant agreement to see if you are part of the Pilot.

Are you part of the Pilot?

If your project stems from one of these Horizon 2020 areas, you are automatically part of the pilot:

- Future and Emerging Technologies
- Research infrastructures – part e-Infrastructures
- Leadership in enabling and industrial technologies – Information and Communication Technologies
- Societal Challenge: Secure, Clean and Efficient Energy – part Smart cities and communities
- Societal Challenge: Climate Action, Environment, Resource Efficiency and Raw materials – with the exception of raw materials topics
- Societal Challenge: Europe in a changing world – inclusive, innovative and reflective Societies
- Science with and for Society

Can you opt in? Yes, projects in other areas can participate on a voluntary basis.

Can you opt out? We hope you don't, but projects may at any stage opt out of the Pilot. See the EC Guide on OA (link overleaf) for eligible reasons.

How can OpenAIRE help?

OpenAIRE provides a range of resources, FAQs, webinars and support pages. OpenAIRE has local representatives in all EU countries: the National Open Access Desks or NOADs. Contact them via our helpdesk system at

www.openaire.eu

If you have no other suitable place to deposit your data, you can use OpenAIRE's catch-all repository, Zenodo

zenodo

Institutions are to retain research data, provide secure data storage, identify ownership, and ensure security and confidentiality of research data

Researchers are to retain research data and primary materials, manage storage of research data and primary materials, maintain confidentiality of research data and primary materials.

Australian requirements

1. Intellectual property
2. Data management, including:
    - Storage
    - RetentionDisposal
    - Access, publication, description

3. Conflict of interest — do all parties have the same understanding about the use of the data?

3. Collaboration and contractual agreements
4. Ethics and privacy Compliance

About ANDS Events | Contact Us

About us | News & Events | Partners & Communities | Working with data | Online Services | Guides

ANDS is enabling the transformation of:

Data that are:	to	Structured Collections that are:
Unmanaged	→	Managed
Disconnected	→	Connected
Invisible	→	Findable
Single Use	→	Reusable

What we do

Find out how ANDS is making Australia's research data assets more valuable and why it matters

Search for Research Data

Enter Keywords

Search the ANDS Site

Enter Keywords

23 Things

Get involved, learn new skills

23 (research data) Things is a self-guided training concept for anybody interested in data, with activities throughout 2016.

More

Data impact campaign

Tell us your stories

Has your research data led to a real-life benefit for Australia? Tell us how to help us show the value of research data.

More

Open data stories

Australia institutions are supporting innovation and new discovery by sharing their data with other researchers. See what ANDS-supported projects have achieved.

23 [research data] things Australia

Do as many or as much as you want; do them in any order; do them by yourself, or form a group to learn together.

There are many activities to choose from — whether you are new to data; need to extend your knowledge, or want a challenge!

Extend your knowledge and skills about research data by exploring activities, links and much more at ands.org.au/23-things

Ready, set, data!

Build your knowledge of key concepts and issues in research data management

1 Getting started with research data

2 Issues in research data management

3 Data in the research lifecycle

Repositories for data

Learn about repositories for depositing, managing and discovering research data.

4 Data discovery

5 Data sharing

6 Long-lived data: curation & preservation

Data citation & impact

Extract value from research data: data citation, impact and metrics.

7 Data citation for access & attribution

8 Citation metrics for data

Rights, ethics & sensitive data

Learn about responsible sharing and reuse, and the importance of licensing research data.

9 Licensing data for reuse

10 Sharing sensitive data

Metadata & more

Resource description, controlled vocabularies, linked data and crosswalks.

11 What's my metadata schema?

12 Vocabularies for data description

13 Walk the crosswalk

14 Identifiers and linked data

Let's talk data!

Start a data conversation: data interviews and data management plans. Data stories from publishers and research funders.

15 Data management plans

16 What are publishers & funders saying about data?

17 Data literacy & outreach

18 Data interviews: talk the talk

Hands on with data & tools

Delve into domain specific data and use software tools for research data. Challenge yourself to try something new!

19 Exploring APIs & apps

20 Find it with dkrnet

21 Tools of the trade

Data communities

Who's who in the research data management zoo? Connecting with colleagues and continuing to learn.

22 What's in a number?

23 Making connections

ands.org.au/23-things

NCRIS National Research Infrastructure for Australia
An Australian Government Initiative

ands

What's in Research Data Australia

Collections (94835)
Research datasets or collections of research materials.

Parties (25487)
Researchers or research organisations that create or maintain research datasets or collections.

Activities (40726)
Projects or programs that create research datasets or collections.

Services (188)
Services that support the creation or use of research datasets or collections.

Spotlight on research data

Professor Jamie Rossjohn

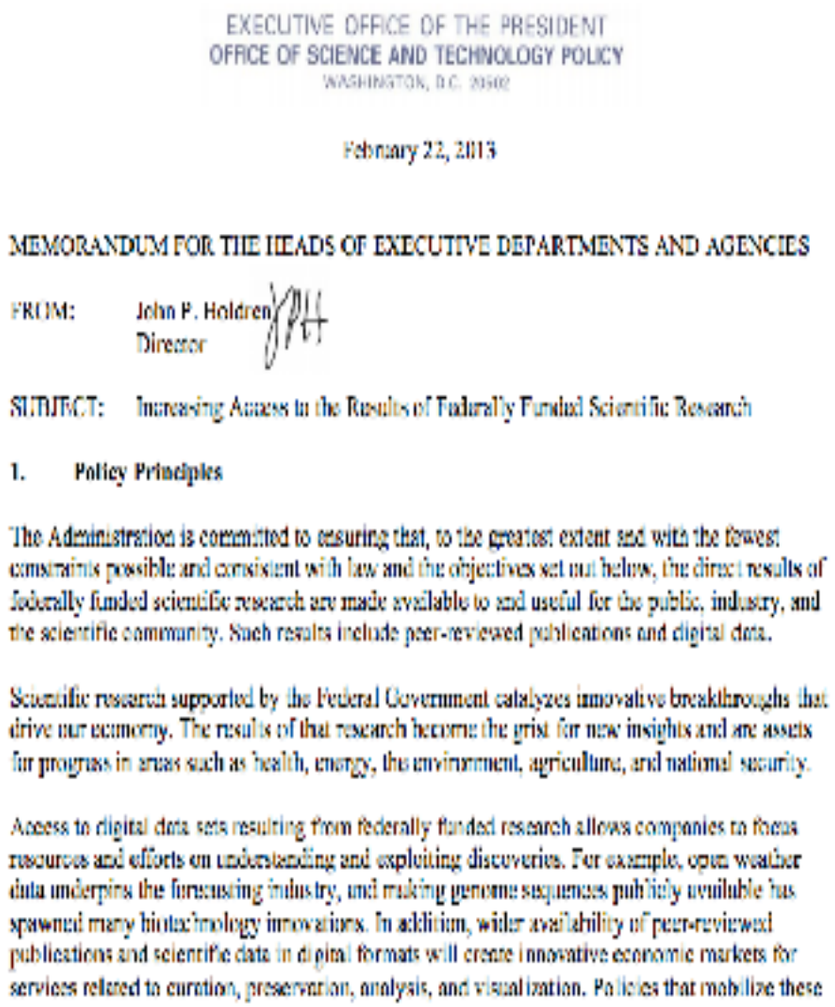
The Australian Academy of Science has recently elected Professor Rossjohn as a Fellow in recognition of his research into the structural basis for T-cell recognition of foreign antigens. His research has had a profound impact on the understanding of immune recognition, particularly in autoimmunity and drug and food hypersensitivities. Professor Rossjohn is based at Monash University's School of Biomedical Sciences researching the basis of infection and immunity, specifically host recognition, responses developed by the pathogen and drug design to modulate and/or counteract these events.

Discover collections contributed by Professor Rossjohn in Research Data Australia >>>

“The Holdren Memo”

To achieve the Administration’s commitment to increase access to federally funded published research and digital scientific data, Federal agencies investing in research and development must have **clear and coordinated policies for increasing such access.**

Memo on Increasing Access to the Results of Federally Funded Scientific Research
White House Office of Science and Technology Policy
February 22, 2012



EXECUTIVE OFFICE OF THE PRESIDENT
OFFICE OF SCIENCE AND TECHNOLOGY POLICY
WASHINGTON, D.C. 20502

February 22, 2013

MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

FROM: John P. Holdren
Director

SUBJECT: Increasing Access to the Results of Federally Funded Scientific Research

1. Policy Principles

The Administration is committed to ensuring that, to the greatest extent and with the fewest constraints possible and consistent with law and the objectives set out below, the direct results of federally funded scientific research are made available to and useful for the public, industry, and the scientific community. Such results include peer-reviewed publications and digital data.

Scientific research supported by the Federal Government catalyzes innovative breakthroughs that drive our economy. The results of that research become the grist for new insights and are assets for progress in areas such as health, energy, the environment, agriculture, and national security.

Access to digital data sets resulting from federally funded research allows companies to focus resources and efforts on understanding and exploiting discoveries. For example, open weather data underpins the forecasting industry, and making genome sequences publicly available has spawned many biotechnology innovations. In addition, wider availability of peer-reviewed publications and scientific data in digital formats will create innovative economic markets for services related to curation, preservation, analysis, and visualization. Policies that mobilize these

AND THEN THEY SAID

“BUT YOU CAN’T GIVE US ANOTHER UNFUNDED MANDATE”

Office of Budget Finance & Award Management (BFA)

- Office of Budget, Finance, & Award Management
- Budget Division
- Division of Acquisition and Cooperative Support
- Division of Financial Management
- Division of Grants & Agreements
- Division of Institution & Award Support
- Large Facilities Office

Advisory Committees

- Business and Operations Advisory Committee

External Links

- Chief Financial Officer Council

Contact BFA

View BFA Staff

Search BFA Staff

GO

Follow us BFA

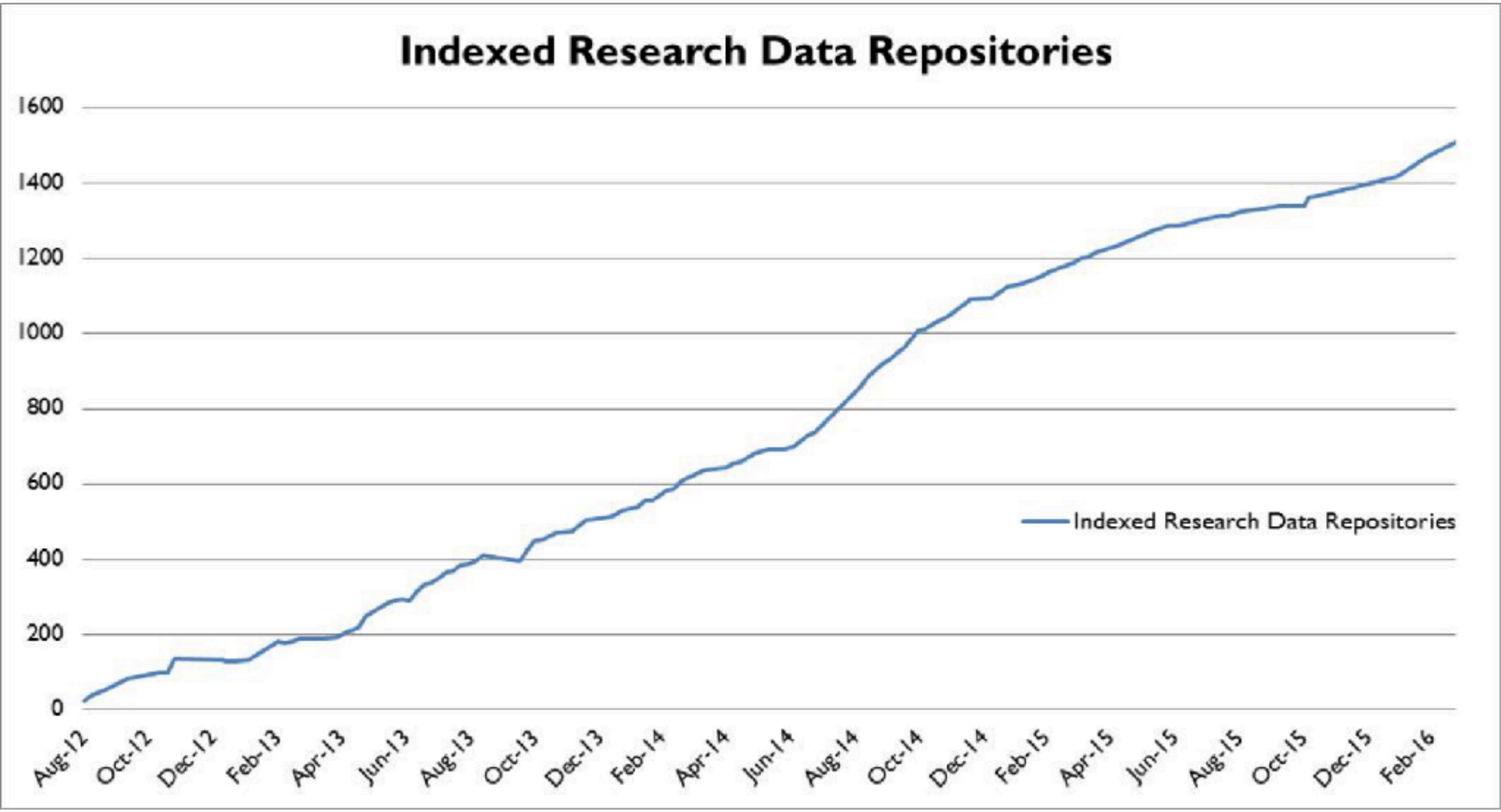
Home

Email Print Share

Data Management & Sharing Frequently Asked Questions (FAQs) -

UPDATED NOVEMBER 30, 2010

1. What constitutes "data" covered by a Data Management Plan?
2. Is a plan for Data Management required if my project is not expected to generate data or samples?
3. Am I required to deposit my data in a public database?
4. There is no public database for my type of data. What can I do to provide data access?
5. Should the budget and its justification specifically address the costs of implementing the Data Management Plan?
6. My institution's policy is that the data and all supporting materials from all research are owned and must remain with the institution if I leave. How does this policy affect what I can say about data management and access?
7. Does data management and access include supporting documentation and metadata, such as validation protocols, field notebooks, etc.?
8. How long should data be archived and made accessible?
9. Does this policy mean that I must make my data available immediately, even before publication?
10. What are NSF's expectations regarding the release of data that include sensitive information (e.g., information about individuals or locations of endangered species)?
11. My data include information of potential commercial value. Am I required to make that information available?
12. Does NSF have particular requirements for archiving and accessibility of samples, physical collections and so forth?
13. Does NSF have particular requirements for what types of samples, physical collections, and so forth should be saved?
14. If data or samples are requested before I have completed all analyses on them, must I share them?
15. How does this policy relate to the issue of open access publishing?
16. If I participate in a collaborative international research project, do I need to be concerned with data management policies established by institutions outside the United States?



re3data.org

REGISTRY OF RESEARCH DATA REPOSITORIES

Research Data Management: Home

Data Planning Checklist

Data Planning Questions:

Data description

1. What type of data will be produced? Will it be reproducible? What would happen if it got lost or became unusable later?
2. How much data will there be? How quickly will it grow? How often will it change? Once archives/stored, what kind of access will be needed to use it?
3. Who will use the data now, and in the future?
4. Who controls the data (PI, student, lab, Caltech, funding agency)? What intellectual property considerations might apply?
5. How long should the data be retained? How long would you expect it to be useful, e.g. through the end of grant/experiment, 3-5 years, 10-20 years, permanently?

Standards

1. Is there good project and data documentation?
2. What directory and file naming conventions will be used?
3. What project and data identifiers will be assigned?
4. What file formats are used? Are they standards-based or proprietary?
5. Are there tools or software needed to create/process/visualize the data? Are the tools or software proprietary?
6. Is there an ontology or other community standard for data sharing/integration?

Access, Sharing, and Re-use

1. Any special privacy or security requirements? e.g., personal data, high-security data
2. Any sharing requirements? e.g., funder data sharing policy
3. Any other funder requirements? e.g., data management plan in grant proposals
4. What is your storage and backup strategy?
5. When will it be shared and where? How broadly will it be shared? Are there I/O throughput issues with respect to the size of the datasets?

Head of Research Services

Gail Clement

Email Me

Offline

Contact:

Sherman Fairchild Library, 3rd floor

626-395-1203

Social:

Data Management Planning: Home

Information on best practices and standards for data management planning.

General Information

NYU DataServices

NYU LIBRARIES & IT

Location

Staffed Hours: Fall 2018

Contact Us

@NYUDataServices

Tweets by @nyudataservices

Data Services
@nyudataservices
.@Qualtrics is back up! Thank you everyone for your patience!
1h

Data Services Retweeted

Sarah DeMott
@311Sjd
@nyudataservices @Qualtrics UPDATE: NYU Qualtrics logins have been restored, although users may still experience issues. twitter.com/nyudataservice...
1h

Embed View on Twitter

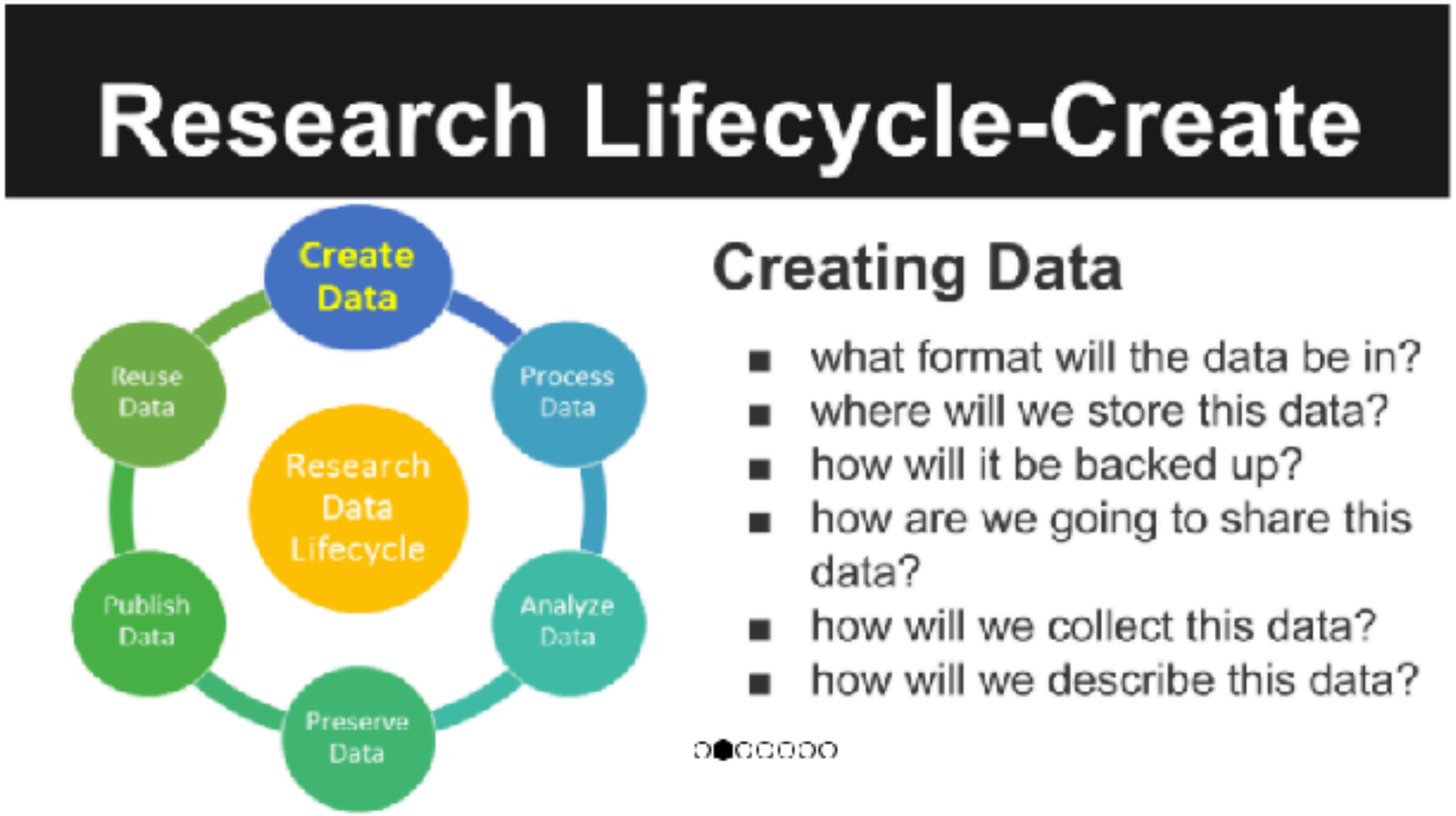
Welcome

Many key granting organizations, like NSF, NIH, NEH and more, now require submitters to include a Data Management Plan as part of their application. In short, these plans outline what steps the applicant will take to collect, safeguard, archive, and make available the data used for the research in question.

This guide lays out practical considerations and information to aid you in preparing your Data Management Plan, and leads you to a Data Management Plan Tool that you can use to create your plan. You can see some more background on this issue, or get started by selecting a tab at the top of the page.

Good luck in your application!

Data Lifecycle



Yale University Library

Research Data Management: Overview

Search this Guide | Search

Resources for learning about best practices in research data management across a variety of disciplines.

Research Data Consultation Group

For help finding, using, managing, or archiving your research data, contact the Research Data Consultation Group.

Quick data management checklist

1. always keep original data
2. back up regularly (automate this if at all possible)
3. document your data thoroughly (metadata, data dictionary)
4. name and organize files according to a schema
5. use version control
6. secure the data appropriately
7. cite any secondary data you use
8. consider your long-term plan
    1. What will you keep, for how long, where, and who will pay for it?
    2. What kinds of reuse or sharing will be allowed? In what timeframe?

What is research data?

Research data is loosely defined as **information collected, observed, or created for purposes of analysis to produce original research.**

This includes observational variables like rainfall, wind speed, water quality, or survey data; simulated data from earthquake models; experimental data from lab instruments; and derived or compiled data for text mining or testing algorithms. Research data can take almost any digital file format (video, text, photographs, numbers), so managing it effectively can be a challenge.

If you've found this page, it's likely that you manage some form of research data, either your own, your lab's, or your advisor's. Many researchers are not taught data management skills in their graduate courses. This guide and accompanying workshops aim to help fill this gap with the expertise that librarians and data specialists can offer.

Why is managing research data important?

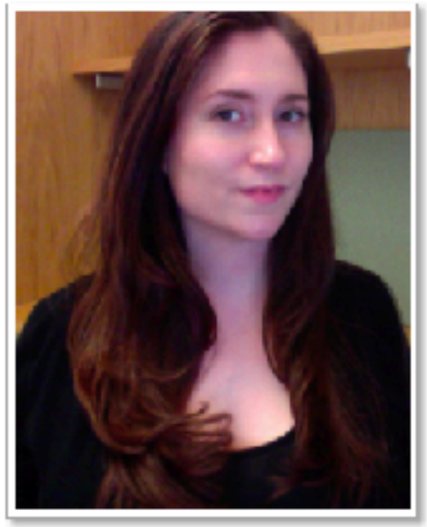
Good data management:

- ensures integrity of data
- ensures that data is findable and usable when grad students leave projects over the years
- makes the data of a project readily understandable to people outside the project
- enables the sharing of data within and across disciplines
- makes it easier to archive and preserve data in the long term
- encourages data citation to increase the impact of the research

Public Access to Federally Funded Research

- Expanding Public Access to the Results of Federally Funded Research
  White House Office of Science & Technology Policy memo requiring all Federal agencies with more than $100M in R&D expenditures to develop plans to make the published results of federally funded research freely available to the public within one year of publication and requiring researchers to better account for and manage the digital data resulting from federally funded scientific research.

Science & Social Science Data Librarian



Michelle Hudson

Email Me

Schedule Appointment

Contact:
Center for Science & Social Science Information
203.432.4587

Subjects:
Data And Statistics

Data management

Helping MIT faculty and researchers manage, store, and share data they produce

We help MIT faculty and researchers manage, store, and share data they produce.

Make a plan

- Why manage your data?
- Write a data management plan
- Other guides to data management
- Meet funder requirements

Store your data

- Organize your files
- Back up & secure your data
- Use documentation & metadata
- Use file formats for long-term access

Share your data

- Find a data repository
- Meet journal requirements
- Learn about confidentiality & intellectual property

Learn more about our services

Questions?

MIT users

Send email to: data-management@mit.edu

Learn how we can help: Services

Visitors

Information for visitors

News

- Data Management: File Organization
- Patent Searching Fundamentals
- Data Management Planning & the DMPTool

STRATEGIC PLAN 2025

"The result is a plan that you would expect of Carnegie Mellon; it does not look like any other university's plan, but it does look like CMU—creative, thoughtful, pragmatic, and ambitious. **A plan, of course, is only a beginning.**"

READ MORE

CREATING A 21ST CENTURY LIBRARY

INDIVIDUAL EXPERIENCE

UNIVERSITY COMMUNITY

SOCIETAL IMPACT

STRATEGIC RECOMMENDATION

Create a 21st century library that serves as a cornerstone of world-class research and scholarship.

In a world where digital is becoming the default format for information, the library will remain a vital presence on campus, sustaining serious scholarship and providing opportunities for interactive research and study environments. To support this important work for students, faculty, and staff, and to create 21st century library spaces for 21st century learners, the library will:

- Develop information specialists as partners in research, teaching, and learning.
- Collaborate with peer institutions to provide coordinated access to a global collection of information resources.
- Steward the evolving scholarly record, and champion new forms of scholarly communication.
- Be recognized globally as a leader in the development of the scholarly information ecosystem.

Learning Spaces

Create 21st-century library spaces for 21st-century learners

Global Collections

Collaborate with peer institutions to provide coordinated access to a global collection of information resources.

Scholarly Record Curation

Steward the evolving scholarly record, and champion new forms of scholarly communication.

Information Specialists

Develop information specialists as partners in research, teaching, and learning

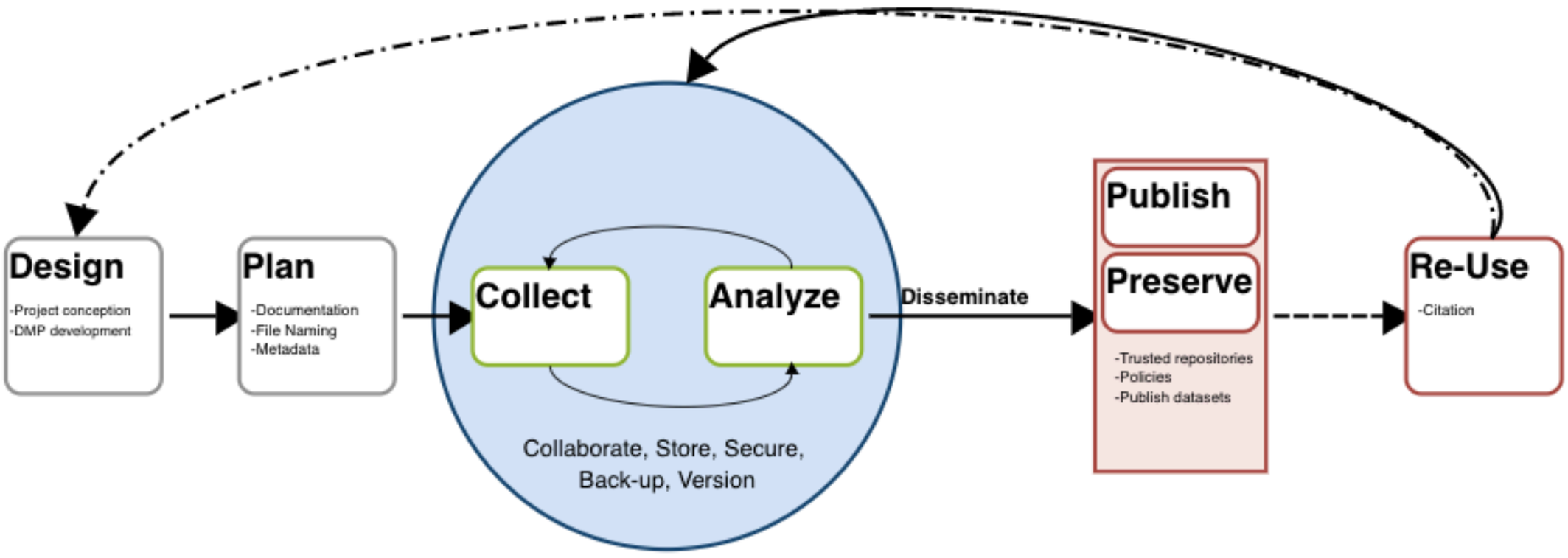
Intro to data management

Data 101 | Document | Metadata | Secure | Publish | DMP basics | CMU resources

Data management 101

This guide will provide general information about data management, including an overview of Data Management Plans (DMPs), file naming conventions, documentation, security, backup, publication, and preservation. We have included the CMU data life cycle to put the pieces in context.

The CMU Libraries provides research data management resources for guidance on data management, planning, and sharing for researchers, faculty, and students.



Elements

figshare

SciVal Experts

Identify expertise and enable collaboration

Dimensions

KUDOS

Altmetric

Carnegie Mellon University

kgw@cmu.edu

cmkeithw

Keith Webster

cmkeithw