Strategies for Open Science and Research Data

Dr Simon Hodson
Executive Director, CODATA
www.codata.org
Research data and Open Science: Towards a national strategy

- Identify...
  - the main strategic areas to build national and institutional roadmaps for RDM services
  - the main / priority requirements to be addressed

- A great honour to be invited to share my experience.
- UK national programme to develop RDM capacity in institutions;
- CODATA works with national members of open science strategies (process of co-design)
CODATA: Committee on Data of the International Council for Science

- Established by the International Council of Science to address issues of data availability and quality.
- Remit has broadened over the years.
- New Executive Committee: includes members from Kenya and South Africa, will co-opt a member from Latin America.
- **Increased orientation towards playing a coordinating role on national and regional Open Science strategies.**
- CODATA President, Geoffrey Boulton, was lead author and chair of Royal Society Report: Science as an Open Enterprise.
- Identifies challenges and opportunities for science systems, technical and human.
- Fundamental methodological issues for reproducibility and transparency.
- **Publications and data should be Intelligently Open and available concurrently.**
- **Report with very significant impact: G8, H2020**
CODATA

Principles, Policies and Practice

Frontiers of Data Science

Capacity Building

Data Science Journal

IDW 2016, 11-17 Sept, Denver, CO.
Research data and Open Science: Towards a national strategy

- Essential to be aware of international, national and institutional dimensions
- We must address the human dimensions (and we neglect them at our peril)

- Proposed CODATA collaboration on Open Science strategy in **Poland** addresses stakeholder responsibilities and enabling practices.

- Current ICSU – CODATA Open Data Platform initiative in **Africa** addresses:
  - Co-development of data policies
  - Incentives and culture
  - Training and skills
  - Roadmap for research data infrastructure
The Open Data Iceberg

Technology

The Technical Challenge

Processes & Organisation

The Ecosystem Challenge
The Funding Challenge
The Support Challenge
The Skills Challenge
The Incentives Challenge
The Mindset Challenge

People

motivation and ethos.

Where should research data go?

- **Homogenous data collections essential for research**
  - Earth observation data;
  - Genetic data;
  - Social science survey data…

- **Significant data outputs of publicly funded research**
  - Significant data outputs from funded projects;
  - Raw and analysed experimental data…

- **Data underpinning research publications**
  - Raw and analysed data for reproducibility (evidence);
  - Data behind the graph…

- **National and international data archives**
- **National or institutional data archives; data papers**
- **Dedicated data archives (e.g. Dryad)**
The Case for Open Data in a Big Data World

- Presents a powerful case that the profound transformations mean that data should be:
  - Open by default
  - Intelligently open
- Supported by four major international science organisations.
- Lays out a framework of principles, responsibilities and enabling practices for how the vision of Open Data in a Big Data World can be achieved.
- Campaign for endorsements: over 100 organisations so far. Please consider endorsing the Accord.
- Translations: Chinese, Russian, Polish, Spanish, French.
An Open Research Data Strategy for Poland

- Collaboration on a national workshop to develop a national open research data/open science strategy for Poland.
- CODATA leads met earlier this year with representatives from Ministry of Science and Education and with Open Science Centre to plan a workshop for Feb/March 2017.
- Draws strongly on the approach of the accord.

**Stakeholders and Responsibilities:** governments/funders, universities and research institutions, institutional libraries, national academies and learned societies, national and international research and data infrastructures, publishers and journal editorial boards.

**Working Groups on Enabling Practices:** boundaries of open, normative values (sharing, timeliness), non-restrictive reuse and TDM, incentives, interoperability, sustainability of data infrastructure, data literacy.
African Open Data Platform Initiative
ICSU-CODATA

- Proposals for Open Data Platform initiatives, Africa and Latin America and Caribbean.
- Holistic ‘science systems’ approach: policies, procedures, incentives, data infrastructure, scholarly communications, skills and training.
- Keystone is to establish an Open Data Platform with a coordinating role.
- Pilot initiative funded by Department of Science and Technology in South Africa: nearly 500K euros over three years.
- Implemented by staff from South African Academy of Sciences, under direction from ICSU-CODATA.
- Currently undertaking preparatory study to identify partners.
Building the Initiative

Establish African Open Data Forum / Platform

- Co-design African Open Data Policies
- Develop Incentives Frameworks
- Develop Research Data Science Training
- African Research Data Infrastructure Roadmap

Activities require low funding for coordination, secondment, contributions in kind and evaluation.

Activities require higher investment for coordination, co-design implementation and evaluation.

Funded Research Data Infrastructure Initiatives

Funded, co-designed transdisciplinary research projects
CODATA in Kenya

- International workshop on open data for science in developing countries, UNESCO, Nairobi, August 2014.
- Strong endorsement for the workshop from Kenyan Cabinet Secretary and from local universities and research institutes.
- Cabinet Secretary Dr. Fred Matiang’i: called on CODATA and other international organisations to 'become more visible in education and capacity-building, by developing science and educational programs and activities that focus on data and information’ in developing countries.
- Announced data centre to be established at Jomo Kenyatta University of Agriculture and Technology.
- ‘JKUAT has now established an ICT Centre of Excellence and Open Data (iCEOD) that was part of the Nairobi-CODATA conference recommendation’
- Working with CODATA on data management policies and development of iCEOD: http://www.codata.org/membership/national-members/kenya
Challenges and developments JUAT / Kenya

1. Lack of national legal/policy framework for open data: e.g. FOI Act ... still Bill
   - JLUAT enacted JORD Policy...as part of implementation of CODATA strategy
   - Undertaking research in various domains
   - Utilizing Collaborations eg CODATA, CAS etc

2. Big Data / Open Data still a new concept ... data reuse and sharing minimal
   - Developed courses for PhD IT- Business Analytics and reviewed undergraduate courses
   - Supply (motivation) and demand balancing

3. Still building and Integrating infrastructure
   - Building iCEOD Open Data Platform to support data reuse, preservation, innovation

4. IP issues ... not enough legal and institutional instruments to encourage more open approaches

5. Cultural practices ... data private by default

6. Capacity building ... organizing short courses & new curriculum for Data Science.
Resources: Current Best Practice for Research Data Management Policies

- Expert report commissioned by CODATA member.
- Provides comprehensive summary of best practice in funder data policies.
- Identifies key elements to be addressed:
  1. Summary of policy drivers
  2. Intelligent openness
  3. Limits of openness
  4. Definition of research data
  5. Define data in scope
  6. Criteria for selection
  7. Summary of responsibilities
  8. Infrastructure and costs
  9. DMP requirements
  10. Enabling discovery and reuse
  11. Recognition and reward
  12. Reporting requirements, compliance monitoring

- Zenodo: http://dx.doi.org/10.5281/zenodo.27872
- See also RECODE Report, Annex on Policy Development: http://recodeproject.eu/
Developments: Journal Data Policies

- **Dryad Joint Data Archiving Policy, Feb 2010:** [http://datadryad.org/jdap](http://datadryad.org/jdap)
  - This journal **requires, as a condition for publication**, that data supporting the results in the paper should be archived in an appropriate public archive, such as GenBank, TreeBASE, Dryad, or the Knowledge Network for Biocomplexity.

- **PLOS Data Availability Policy, revised Feb 2014:** [http://www.plosone.org/static/policies.action#sharing](http://www.plosone.org/static/policies.action#sharing)
  - PLOS journals require authors to make all data underlying the findings described in their manuscript fully available without restriction, with rare exceptions.

- **Jisc work to develop registry of journal data policies; BioSharing [https://biosharing.org/](https://biosharing.org/)**
  - Likely new initiative through RDA to encourage development and adoption of journal data policies.
  - CODATA working with ICSU to encourage ISUs to address data policy from disciplinary perspective.
Barriers to Data Availability / Publication

Researchers concerns:

- Concern that data may be misused or misunderstood.
- Concern that will lose scientific edge if sharing before fully exploited.
- Desire to retain control of a professional asset.
- **Concern that will not be credited.**
- **Lack of career rewards for data publication.**


- Culture in particular research disciplines; availability of infrastructure.
- **Fundamentally, researchers are reluctant to expend effort sharing data because they do not feel that data is adequately exposed or credited.**

Citation advantage of having archived Gene Expression Omnibus data

Examined 10,555 studies that created gene expression microarray data, comparing those that made data available and those that didn’t.

Studies that made data available in a public repository received 9% more citations than similar studies for which the data was not made available.

Increased citation of 30% for those published 2004-5.

Piwowar and Vision (2013), PeerJ DOI:10.7717/peerj.175
Data Policies: Data Citation

If publications are the stars and planets of the scientific universe, data are the ‘dark matter’ – influential but largely unobserved in our mapping process.

Task Group on Data Citation Principles and Practices

Out of Cite, Out of Mind


Joint Declaration of Data Citation Principles:
https://www.force11.org/datalcitation

Background and Developments:
Data Citation as a Recognised Scientific Responsibility


- Endorses the OECD Principles and Guidelines on Access to Data from Publicly Funded Research (2007)

- Recommendation 4: ‘Science publishers and chief editors of scientific publications should require authors to provide explicit references to the datasets underlying published papers, using unique persistent identifiers. They also should require clear assurances that these datasets are deposited and available in trusted and sustainable digital repositories. Citing datasets in reference lists using an accepted standard format should be considered the norm.’

- **Accord on Open Data in a Big Data World**, Principle ix; Paras 54-57 on Citation and Provenance: ‘When, in scholarly publications, researchers use data created by others, those data should be cited with reference to their originator, to their provenance and to a permanent digital identifier.’
CODATA Task Group on Data Citation ‘Data Citation: From Principles to Practice, A Focus on the Research Policy and Funding Community’: [http://www.codata.org/task-groups/data-citation-standards-and-practices](http://www.codata.org/task-groups/data-citation-standards-and-practices)

- Organising an international series of implementation and adoption workshops.
- Promote the implementation of data citation principles in the research policy and funding communities throughout the world.
- **Stakeholders** include: government, funders, research performing institutions, research administrators, research librarians, researchers, learned societies, publishers, data archives, journal editors ...
  - What is the policy environment for data citation?
  - What are current attitudes to data citation?
  - What infrastructure currently exists to support data citation?
  - What specific plans for implementation were identified?
We are taking Data Citation workshops on a world tour!

2015: China, Australia, Japan, India and South Africa.
2016: USA, Israel, Russia + Finland (Nov) and Taiwan (Dec).
2017: France, Korea, Indonesia, Brazil...

Synthesis Report of first 8 workshops to be published in soon!
Contemporary research – particularly when addressing the most significant, transdisciplinary research challenges – increasingly depends on a range of skills relating to data. These skills include the principles and practice of Open Science and research data management and curation, the development of a range of data platforms and infrastructures, the techniques of large scale analysis, statistics, visualisation and modelling techniques, software development and data annotation. The ensemble of these skills, relating to data in research, can usefully be called ‘Research Data Science’.
Foundational Research
Data Science Curriculum

Seven components: open science, data management and curation; software carpentry; data carpentry; data infrastructures; statistics and machine learning; visualisation.

Builds on much existing courses to create something more than the sum of its parts:

- **Open Science** – reflection on ethos and requirements of sharing/openness
- **Open Research Data** – Basics of data management, DMPs, RDM life-cycle, data publishing, metadata and annotation
- **Software Carpentry** – Introduction to programming in R, the Unix shell and Git (sharing software and data)
- **Data Carpentry** – Introduction to SQL databases
- **Visualisation** – Tools, Critical Analysis of Visualisation
- **Analysis** – Statistics and Machine Learning (Clustering, supervised and unsupervised learning)
- **Computational Infrastructures** – Introduction to cloud computing, launching a Virtual Machine on an IaaS cloud
CODATA-RDA School of Research Data Science

- First School of Research Data Science, 1-12 August 2016, ICTP, Trieste
- Funding for students and tutors provided by ICTP, TWAS, CODATA, ACU, RDA Europe, GEO and GODAN.
- Attended by 70 students from all around the world.
Programme for #datatrieste

- School will repeat at Trieste in 2017 and 2018, at least...
- Possibly with addition of one week more advanced on Big Data.
- Will run foundational two week course at ICTP INESP in Sao Paolo, Brazil, December 2017.
- Schools can be run with a greater or lesser degree of support and coordination from the international convenors.
- Keen to encourage a network of schools, but also local schools with lower central input.
- Discussions with possible partners in South Africa and India.
- Keen to explore opportunities with CODATA National and Union Members.
Research Data Infrastructure Roadmaps

- Data Intensive Research Infrastructure SA (DIRISA) and SA Research Infrastructure Roadmap (SARIR) identify key issues in development of research infrastructure.
  - SARIR identifies 17 key research infrastructures for SA, based on an ESFRI-type methodology.
  - ESFRI approach and the development of the ERICs/research infrastructures important.
  - What is the data landscape and ecosystem? What is provided by national and international infrastructures and what by research institutions?
  - Comparable exercise will be performed with partners of Data Science Capacity Building Initiative.
- Important to ensure that research infrastructure adequately addresses Open Science requirements.
  - Develop roadmap Research Data Infrastructures
- What are the RDI requirements at a regional level and for African nations?
- What is the role of disciplinary infrastructures and of research institutions?
- Importance of a full-lifecycle approach.
Supporting the Research Data Lifecycle

- Plan
  - Create
    - Store
    - Annotate
  - Use
    - Discard
  - Appraise
    - Select
    - Hand Over?
  - Publish
    - Describe
    - Identify
  - Discover
    - Access
Where should research data go?

**Homogenous data collections essential for research**
- Earth observation data;
- Genetic data;
- Social science survey data…

**Significant data outputs of publicly funded research**
- Significant data outputs from funded projects;
- Raw and analysed experimental data…

**Data underpinning research publications**
- Raw and analysed data for reproducibility (evidence);
- Data behind the graph…

**National and international data archives**

**National or institutional data archives**
- Data papers

**Dedicated data archives (e.g. Dryad)**
Research Data Infrastructure Roadmaps

- **Research priorities and gap analysis.**
- **Ecosystem:** what is the provision of RDIs for particular disciplines through national and international initiatives?
- **Role of Research Institutions:** Is lifecycle support and long tail being supported in institutions.
- RDIs are not just hardware, but ‘part of a research ecosystem’, so must address: governance; training, personnel and career structures’ sustainable funding; access and outreach to national, public and commercial partners.
- **Roadmap for Data Infrastructure**
  - Co-design to meet national needs and priorities.
  - Research priorities, opportunities for shared infrastructures, examples of good governance and sustainable funding models.
  - **Sustainable Business Models for RDIs**
The Challenge: Sustainable Business Models for Data Repositories

- Research funder policies – quite rightly – mandate data stewardship.
  - OECD Principles and Guidelines, 2007
  - G8 Science Ministers Statement, 2013
  - Major funders in US, UK, EC Horizon 2020 data policy etc.
- Increasing need for data repositories and data stewardship.
  - Increasing volume presents a challenge.
  - Requirements for stewardship present a greater challenge.
- Sustaining digital data infrastructure is a major issue for science policy!
- Genuine concern that current funding models will prove inelastic and not meet the growing requirements – concern on the part of repositories and funders.
- Witnessing Innovation
  - Changes in funding / business models (ADS, TAIR; DANS, ICPSR)
  - Innovative business models (Dryad, FigShare)
OECD Global Science Forum Project:
Sustainable Business Models for Data Repositories

- Questions to address:
  1. How are data repositories currently funded?
  2. What innovative income streams are available?
  3. What means of restraining costs are available?
  4. How do income streams match willingness/ability to pay of various stakeholders?
  5. How do income streams/willingness to pay fit together into a sustainable business model?

- Builds on previous work of RDA-WDS Interest Group:
  http://dx.doi.org/10.5281/zenodo.46693

- Broader landscape survey of current funding models, May-Sept 2016.
- Focus group on innovative income streams and on cost restraint, workshop Nov 2016.
- Test business models with stakeholder groups, workshop April 2017.
- Policy recommendations based on concrete business model options, April-June 2017.
iCEOD Value Chain: Data and Society Agriculture and Nutrition

Value Chain – iCEOD Open Data Cloud

Ecosystem of Open Data Partners

JKUAT iCEOD Open Data Cloud

Data Scientists & Analysts

Developers, Innovators & Entrepreneurs

End Users Consumers

Open Data from iCEOD and their partner ecosystem including:
- Hospitals
- Research Partners
- Agriculture Partners
- Governments

JKUAT Staff and PhD. students expose and monetize APIs across IBM enabled platforms

iCEOD Staff, JKUAT faculty, PhD students and other Research Partners enrich and extend the Open Data by applying analytics to unlock valuable insights

Developers, ISVs, Entrepreneurs access Open Data through a catalogue of free and monetized APIs using these to create innovative apps and solutions

End Users take advantage of these innovative Apps to address their needs and problems, commercialised paid and free Apps with in-app purchases
Data Revolution: how can we improve ... with open data?

- GODAN-ODI Report: improving agriculture, food and nutrition with open data.
- ‘Although the amount of data openly available is constantly increasing, there are still challenges related to data management, licensing, interoperability and exploitation. There is a need to evolve policies, practices and ethics around closed, shared, and open data.’
- Enabling more efficient and effective decision making > lowers cost of accessing information and underpins tools that farmers themselves can use.
- Fostering innovation to benefit everyone > an opportunity that must not be missed for creating new businesses and jobs in ‘new data-powered innovation ecosystems’.
- Driving organisational and sector change through transparency > open data is essential to understanding complex systems, interventions, targets, change.
- Availability is not enough > essential that the data be interoperable and machine-readable.
- Problem oriented and solution-based data strategies.
- Develop infrastructure and human capacity.
The Value of Open Data Sharing

- Report by CODATA for GEO, the Group on Earth Observation.
- Provides a concise, accessible, high level synthesis of key arguments and evidence of the benefits and value of open data sharing.
- Particular, but not exclusive, reference to Earth Observation data.
- Benefits in the areas of:
  - Economic Benefits
  - Social Welfare Benefits
  - Research and Innovation Opportunities
  - Education
  - Governance
- Available at http://dx.doi.org/10.5281/zenodo.33830
- GEO DSWG is building on this work with further examples: would be valuable to work with this community.
Thank you for your attention!

Credits for slides: inc. Geoffrey Boulton, Joseph Muliaro Wafuila
Credit for photos: Andjani Gatzweiler

Simon Hodson
Executive Director CODATA
www.codata.org
http://lists.codata.org/mailman/listinfo/codata-international_lists.codata.org
Email: simon@codata.org
Twitter: @simonhodson99
Tel (Office): +33 1 45 25 04 96 | Tel (Cell): +33 6 86 30 42 59

CODATA (ICSU Committee on Data for Science and Technology), 5 rue Auguste Vacquerie, 75016 Paris, FRANCE
Extra Slide

- XXXX
- XXXX
Motivations and Drivers

- Rigour and reproducibility
- Research benefits of data reuse
80% of ecology data irretrievable after 20 years (516 studies)

Vines TH et al. (2013) Current Biology DOI: 10.1016/j.cub.2013.11.014
Data Revolution: A World that Counts!

- **Creating a world that counts**: Mobilising the Data Revolution for Sustainable Development.
- To meet the new sustainability goals ‘there is an urgent need to mobilise the data revolution for all people and the whole planet in order to monitor progress, hold governments accountable and foster sustainable development.’
- Without immediate action, gaps between developed and developing countries, between information-rich and information-poor people, and between the private and public sectors will widen, and risks of harm and abuses of human rights will grow.
  - Data quality and integrity
  - Data disaggregation (no-one should be invisible)
  - Data timeliness
  - Data transparency and openness
  - Data usability and curation
  - Data protection and privacy
  - Data governance and independence
  - Data resources and capacity
  - Data rights

*Source: UN Global Pulse (http://www.unglobalpulse.org/nowcasting-food-prices)*
Improving crop varieties with open data on breeding trials: AgTrials

Cultivar testing is an important means of improving crop varieties. A wide range of trials are taking place on sites all over the world, addressing issues such as drought tolerance, heat stress, and soil management. However, almost all of the data generated has been inaccessible to other researchers – filed away on laboratory hard drives, or sometimes lost completely due to bad data management.

By compiling data from agronomic and plant breeding trials and making it open, the Global Agricultural Trial Repository (AgTrials) hosted by a CGIAR Research Programme on Climate Change, Agriculture and Food Security (CCAFS) offers a rich knowledge base to inform ongoing, collaborative research, while eliminating unnecessary and costly duplication of efforts.

Scientists used 250 open AgTrials datasets to build crop models specific to the West Africa region. The models are used to project the local impacts of climate change, and define breeding programmes for adaptation.
Barbara Ryan, Director of Secretariat GEO, TED-X Talk Barcelona

In 2008 US Government was convinced to make Landsat Data openly available, for free.

Under charging, the highest number of downloads was 53 scenes per day.

Now over 5700 scenes per day are downloaded.

Spanish deforestation research: under the charging regime data access alone would have cost €260M

CODATA produced a White Paper on the Value of Data Sharing for the GEO-XII Plenary:
http://dx.doi.org/10.5281/zenodo.33830

https://www.youtube.com/watch?v=9umWTFgFIVs
Economic Benefits of Data Sharing: LandSat

- **2006 Study** estimated the loss in case of a data gap as equivalent to US$935 M.
- **2011 Study** estimated benefits of landsat-sourced information for agriculture as US$858 M just for the state of Iowa.
- **2015 Study** estimated worldwide economic benefit of US$2.19 BN.
- Estimated benefit in US of US$1.8 BN.
- Valuing Geospatial Information: Using the Contingent Valuation Method to Estimate the Economic Benefits of Landsat Satellite Imagery: [http://dx.doi.org/10.14358/PERS.81.8.647](http://dx.doi.org/10.14358/PERS.81.8.647) (Paywall... Irony...)
- Open data and open data infrastructure has a significant economic benefit.
### Figure 3: The value and impacts of the three UK data centres

<table>
<thead>
<tr>
<th>Data Centre</th>
<th>Investment &amp; Use Value (Direct)</th>
<th>Contingent Value (Stated)</th>
<th>Efficiency Impact (Estimated)</th>
<th>Return on Investment (Estimated)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ESDS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Based on Depositor &amp; User Counts</td>
<td></td>
<td></td>
<td>Based on Data Spend</td>
</tr>
<tr>
<td></td>
<td><strong>Investment Value</strong> £23m per annum</td>
<td>Willingness to Pay £25m per annum</td>
<td>Survey Community £68m - £112m per annum</td>
<td>Increased RoI [From Additional Use] £58m - £233m 2.5 to 10-fold RoI</td>
</tr>
<tr>
<td></td>
<td><strong>Use Value</strong> £24m per annum</td>
<td>Willingness to Accept £83m per annum (£111m per annum)</td>
<td>Wider Community £100m plus per annum</td>
<td></td>
</tr>
<tr>
<td><strong>ADS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Based on Deposit &amp; Download Counts</td>
<td></td>
<td></td>
<td>Based on Data Spend</td>
</tr>
<tr>
<td></td>
<td><strong>Investment Value</strong> £1.1m per annum</td>
<td>Willingness to Pay £1.1m per annum (Constrained by capacity to pay)</td>
<td>User Community [ADS data use] £13m per annum</td>
<td>Increased RoI [From Additional Use] £2.4m - £9.7m 2.1 to 8.3-fold RoI</td>
</tr>
<tr>
<td></td>
<td><strong>Use Value</strong> £1.4m per annum</td>
<td>Willingness to Accept £7.4m per annum</td>
<td>User Community [All activity time] £58m per annum</td>
<td></td>
</tr>
<tr>
<td><strong>BADC</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Investment Value</strong> £2.8m per annum</td>
<td>Willingness to Pay £5.2m per annum</td>
<td>User Community [BADC data use] £10m per annum</td>
<td>Increased RoI [From Additional Use] (Non Recreate) £11m - £34m 4 to 12-fold RoI</td>
</tr>
<tr>
<td></td>
<td><strong>Use Value</strong> £2.3m per annum</td>
<td>Willingness to Accept £16m per annum</td>
<td>User Community [All activity time] £58m per annum</td>
<td></td>
</tr>
</tbody>
</table>
Economic Benefits of Data Sharing

- ‘Many studies and reports have documented the positive value of openness for EO data, specifically, and for various other types of data and information, more generally.’

- Weiss 2002: quantified considerable economic benefits of making meteorological data open ($400-700M in gross receipts; businesses and employment).


- Houghton 2014: Estimate unrealised benefits of research data of AU$1.4-4.9BN set against estimated AU$130-200M cost of data infrastructure.

- Interested to know what studies of the benefits of data availability have been conducted in this area of research?