

# Publishing data and Trustworthy Data Repositories

Mustapha Mokrane  
WDS Executive Director

Belmont Forum e-Infra & DM workshop:  
Open Data Principles in Publishing

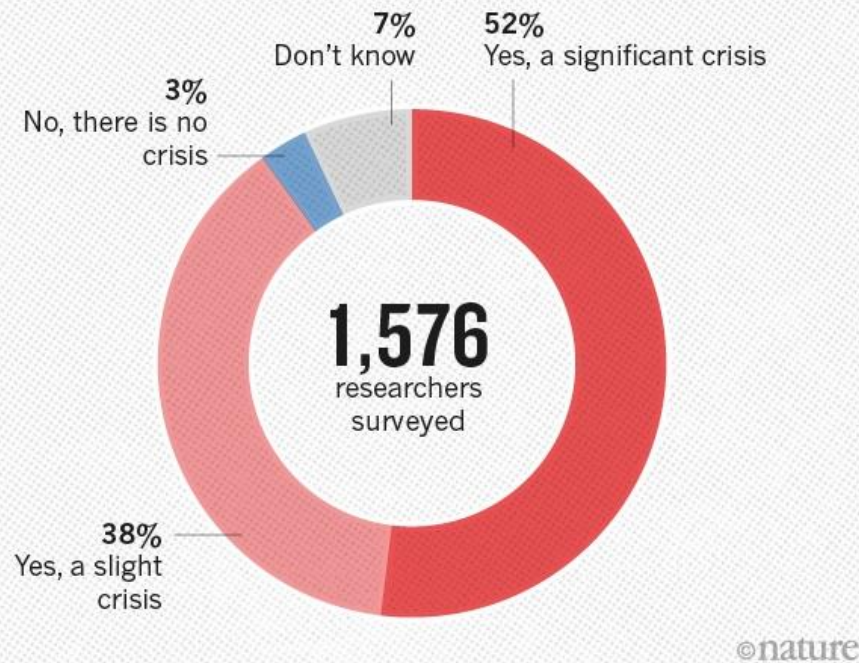
London, 23 June 2017



**ICSU**  
WORLD DATA SYSTEM

# Open Science?

## IS THERE A REPRODUCIBILITY CRISIS?

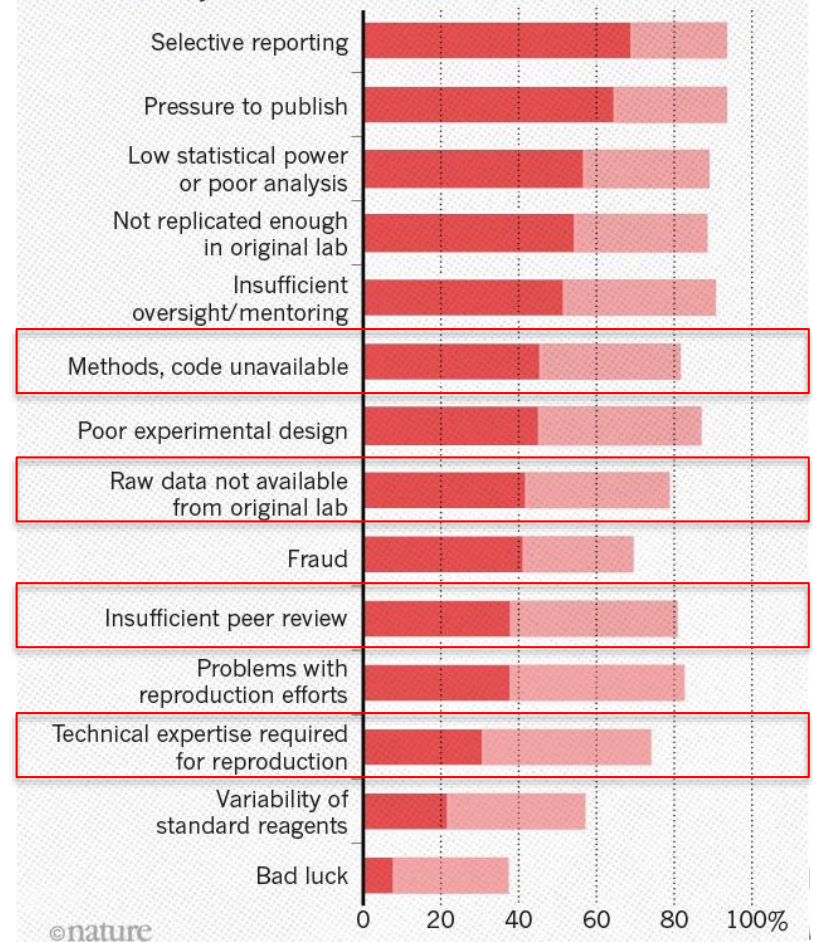


Nature 533, 452–454 (26 May 2016)  
doi:10.1038/533452a

## WHAT FACTORS CONTRIBUTE TO IRREPRODUCIBLE RESEARCH?

Many top-rated factors relate to intense competition and time pressure.

● Always/often contribute ● Sometimes contribute



# Role of (Domain) Data Repositories

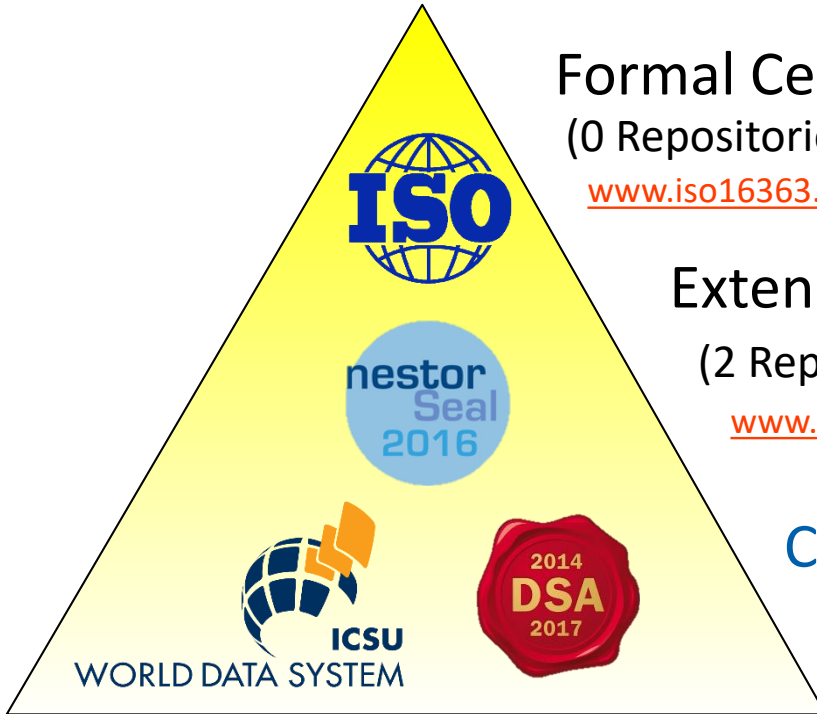
- Data publishers: data underlying experiments, scientific papers, projects, etc.
- Data practitioners: involved in early stages of research planning (DMP)
- Data managers: workflows, metrics, linked objects (PIDs), etc.
- Long-term preservation: reproducibility!



# Data Repositories Stakeholders

- Researchers/projects/private sector: users of repositories
- Scientific journal publishers: recommend/use repositories (e.g. ESSD, Nature Scientific Data)
- Funders: support repositories
- Policymakers: provide a mandate

# Trustworthy Data Repositories



**Formal Certification ISO 16363:2012**  
(0 Repositories, 100+ requirements)

[www.iso16363.org/](http://www.iso16363.org/)

**Extended Certification DIN 31644**  
(2 Repositories, 34 requirements)

[www.langzeitarchivierung.de](http://www.langzeitarchivierung.de)

**Core Certification WDS / DSA**  
(150+ Repositories, 16 requirements)

[www.datasealofapproval.org](http://www.datasealofapproval.org)

[www.icsu-wds.org](http://www.icsu-wds.org)

# Core certification requirements

## 16 requirements

- Context
- Organizational infrastructure (6)
- Digital object management (8)
- Technology (2)
- Applicant feedback

## F.A.I.R and beyond

25/08/2015

Common Requirements/V2.1



### DSA–WDS Partnership Working Group Catalogue of Common Requirements

#### Introduction

#### Importance of Certification

National and international funders are increasingly likely to mandate open data and data management policies that call for the long-term storage and accessibility of data.

If we want to be able to share data, we need to store them in a trustworthy digital repository. Data created and used by scientists should be managed, curated, and archived in such a way to preserve the initial investment in collecting them. Researchers must be certain that data held in archives remain useful and meaningful into the future. Funding authorities increasingly require continued access to data produced by the projects they fund, and have made this an important element in Data Management Plans. Indeed, some funders now stipulate that the data they fund must be deposited in a trustworthy repository.

Sustainability of repositories raises a number of challenging issues in different areas: organizational, technical, financial, legal, etc. Certification can be an important contribution to ensuring the reliability and durability of digital repositories and hence the potential for sharing data over a long period of time. By becoming certified, repositories can demonstrate to both their users and their funders that an independent authority has evaluated them and endorsed their trustworthiness.

#### Basic Certification and its Benefits

Nowadays certification standards are available at different levels, from a basic level to extended and formal levels. Even at the basic level, certification offers many benefits to a repository and its stakeholders.

[DOI.org/10.17026/dans-22n-gk35](https://doi.org/10.17026/dans-22n-gk35)





re3data.org Search Browse Suggest Resources Contact DataCite

Repository details

### NASA Socioeconomic Data and Applications Center

General Institutions Terms Standards

Name of repository: NASA Socioeconomic Data and Applications Center

Additional name(s): SEDAC

Repository URL: <http://sedac.ciesin.columbia.edu/>

Subject(s): Social Sciences, Human Geography, Geophysics and Geodesy, Geosciences (including Geography), Social and Behavioural Sciences, Humanities and Social Sciences, Geography, Natural Sciences, Political Science, Soil Sciences, Economics, Urbanism, Spatial Planning, Transportation and Infrastructure Planning, Landscape Planning, Agriculture, Forestry, Horticulture and Veterinary Medicine, Life Sciences, Construction Engineering and Architecture, Engineering Sciences

Description: SEDAC, the Socioeconomic Data and Applications Center, is one of the Distributed Active Archive Centers (DAACs) in the Earth Observing System Data and Information System (EOSDIS) of the U.S. National Aeronautics and Space Administration. SEDAC is a regular member of the World Data System and focuses on human interactions in the environment. Its mission is to develop and operate applications that support the integration of socioeconomic and Earth science data and to serve as an "Information Gateway" between the Earth and social sciences.

Contact: <http://sedac.uservice.com/>

Content type(s): Standard office documents, Scientific and statistical data formats, Structured graphics

Certificates and Standards: World Data System

Keyword(s): environmental assessment, environmental health, environmental treaties, environmental indicators, energy infrastructure, poverty, environmental sciences, water, remote sensing, population, marine and coastal, land use, infrastructure, health, hazards, conservation, climate, agriculture, socioeconomics, global environment, sustainability, governance

Repository size: 37 data collections; 196 datasets

Repository type(s): disciplinary

Mission statement for designated community: <http://sedac.ciesin.columbia.edu/about>

Research data repository: eng

The Data Repository is certified

Certificates and Standards World Data System

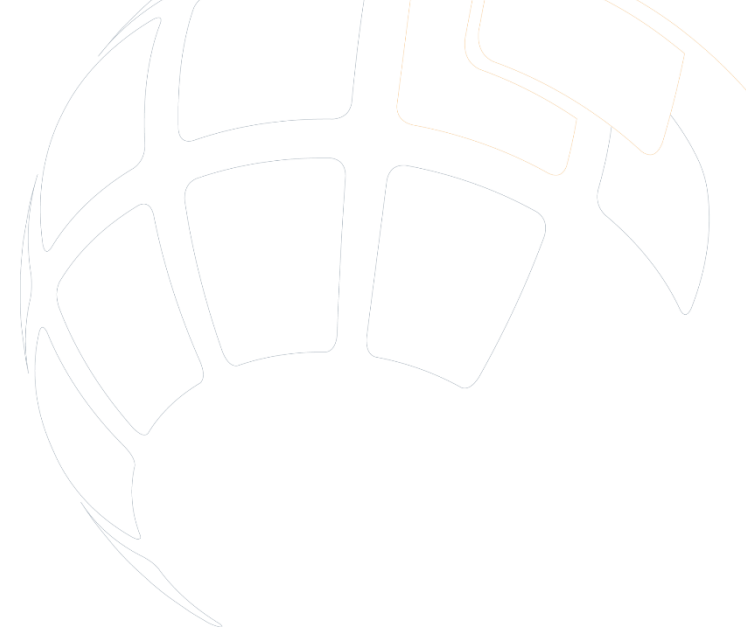


# What's next ?

- Belmont Forum recognizes the important role of Trustworthy Data Repositories in its *Data Policy*
- Other funders (NIH, NSF, EC) are also showing interest in data repositories certification
- Several scientific journal publishers already recommend trustworthy/reliable data repositories
- Data policies of scientific journal publishers should make clear statements about the use of **certified** trustworthy data repositories







# Thank you for listening!

Email: [ipo@icsu-wds.org](mailto:ipo@icsu-wds.org)

Twitter: [@ICSU\\_WDS](https://twitter.com/ICSU_WDS)