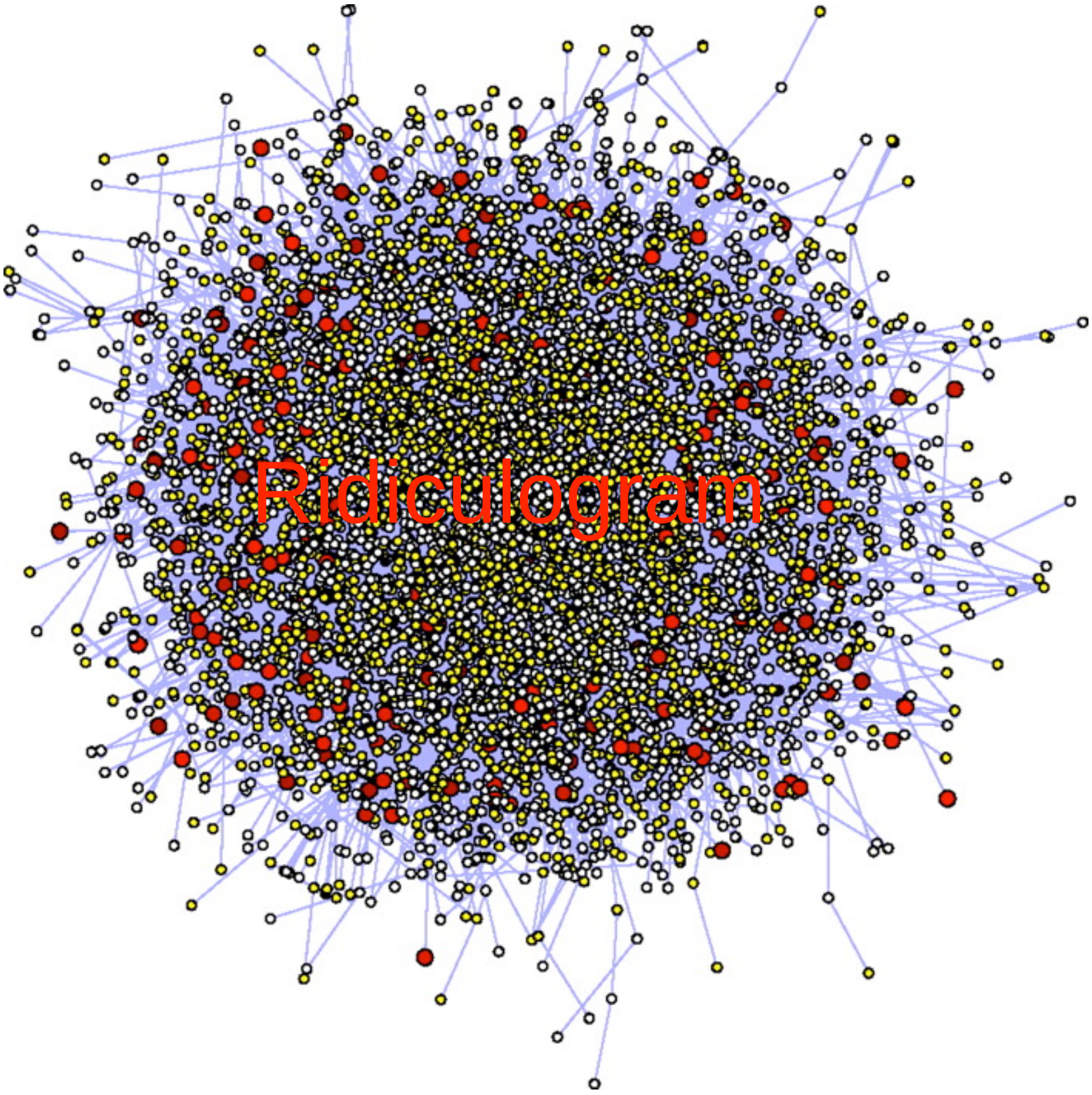
The European Open Science Cloud

e-IRG
Amsterdam

JC Burgelman, W. Lusoli, B. Mons

Key challenges

- Still a lack of widespread **awareness** of the value of data and of **incentives** for data sharing.
- Lack of common standards to ensure **inter-operability** of data.
- **Not enough hardware capacity** for scientific computing, storage, connectivity.
- **Fragmentation and lack of coordination** over different scientific communities and countries.
- Need to translate recent **changes in privacy, data protection and copyright rules** to the research data domain.

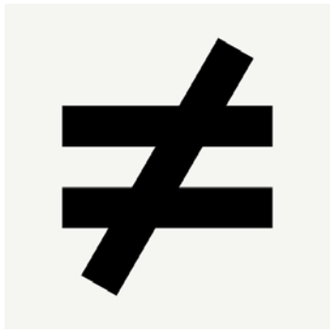

Ridiculogram

- European
- Open
- Science
- Cloud

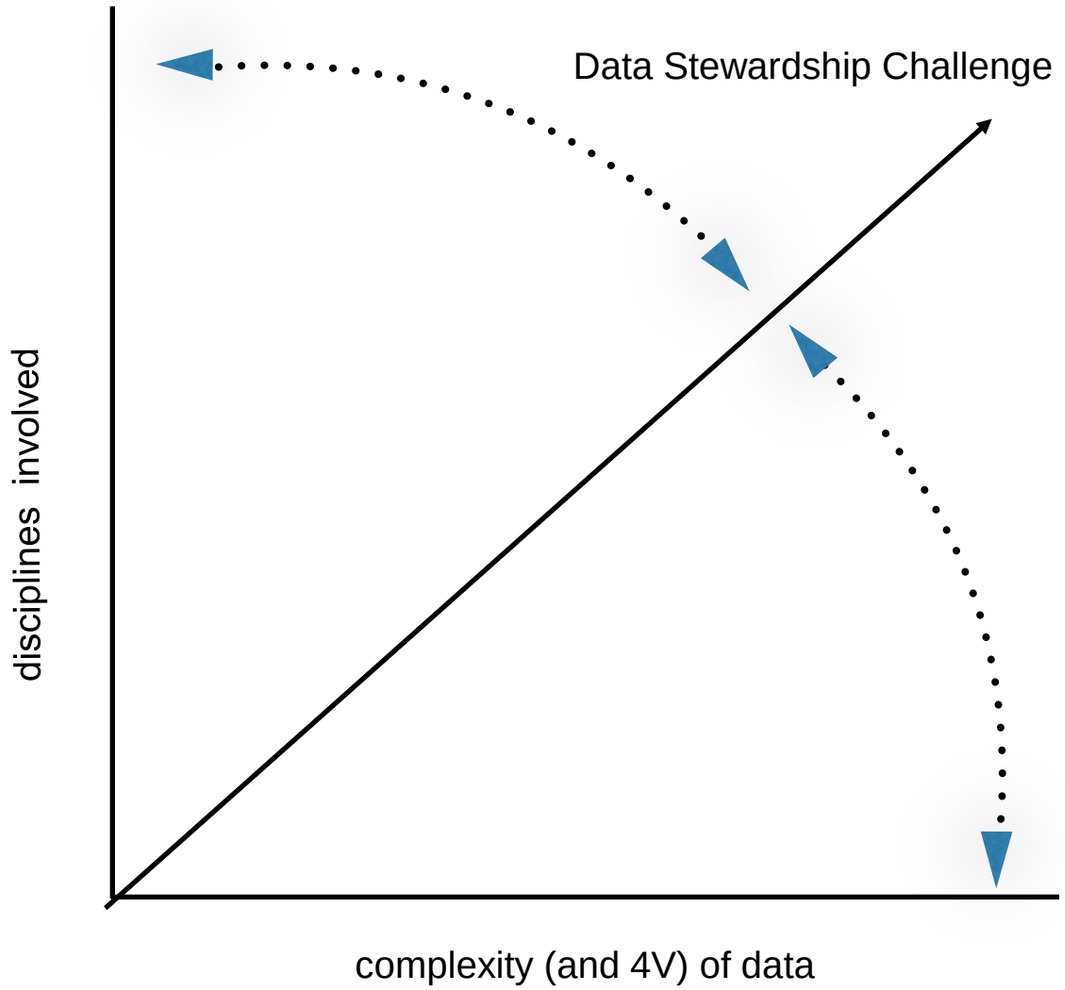
- European
- Open Science
- Cloud

The EOSC

Open Science



OA (articles)


Data Stewardship Challenge
disciplines involved
complexity (and 4V) of data

something to refer to

SCIENTIFIC DATA

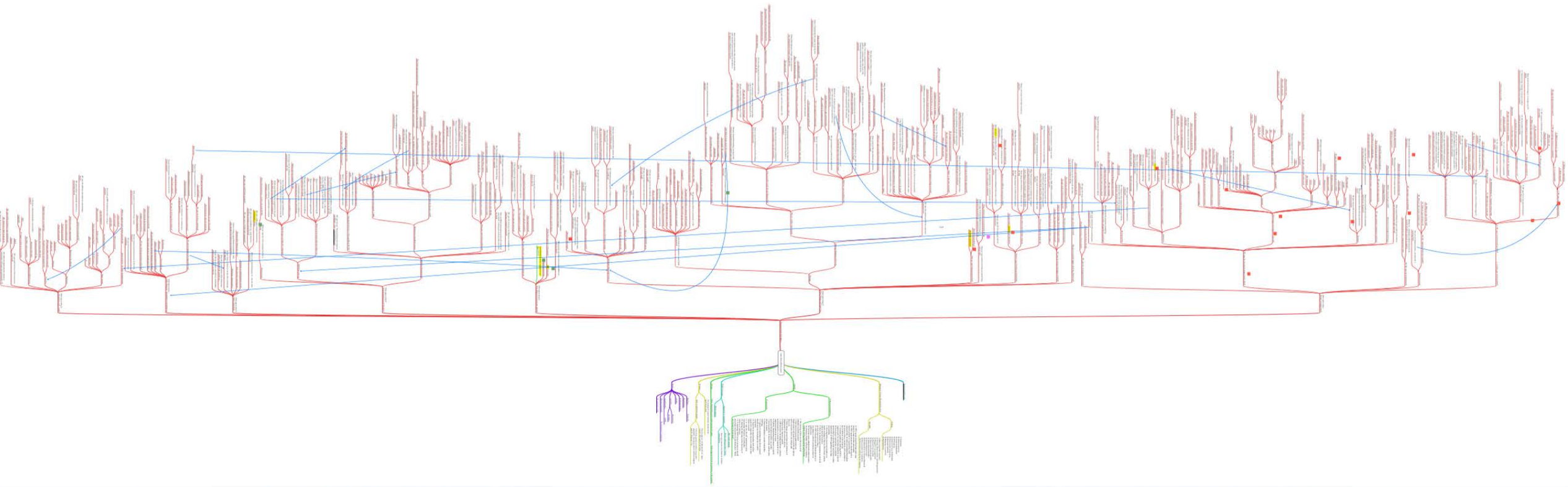
IN PRESS

The FAIR Guiding Principles for scientific data management and stewardship

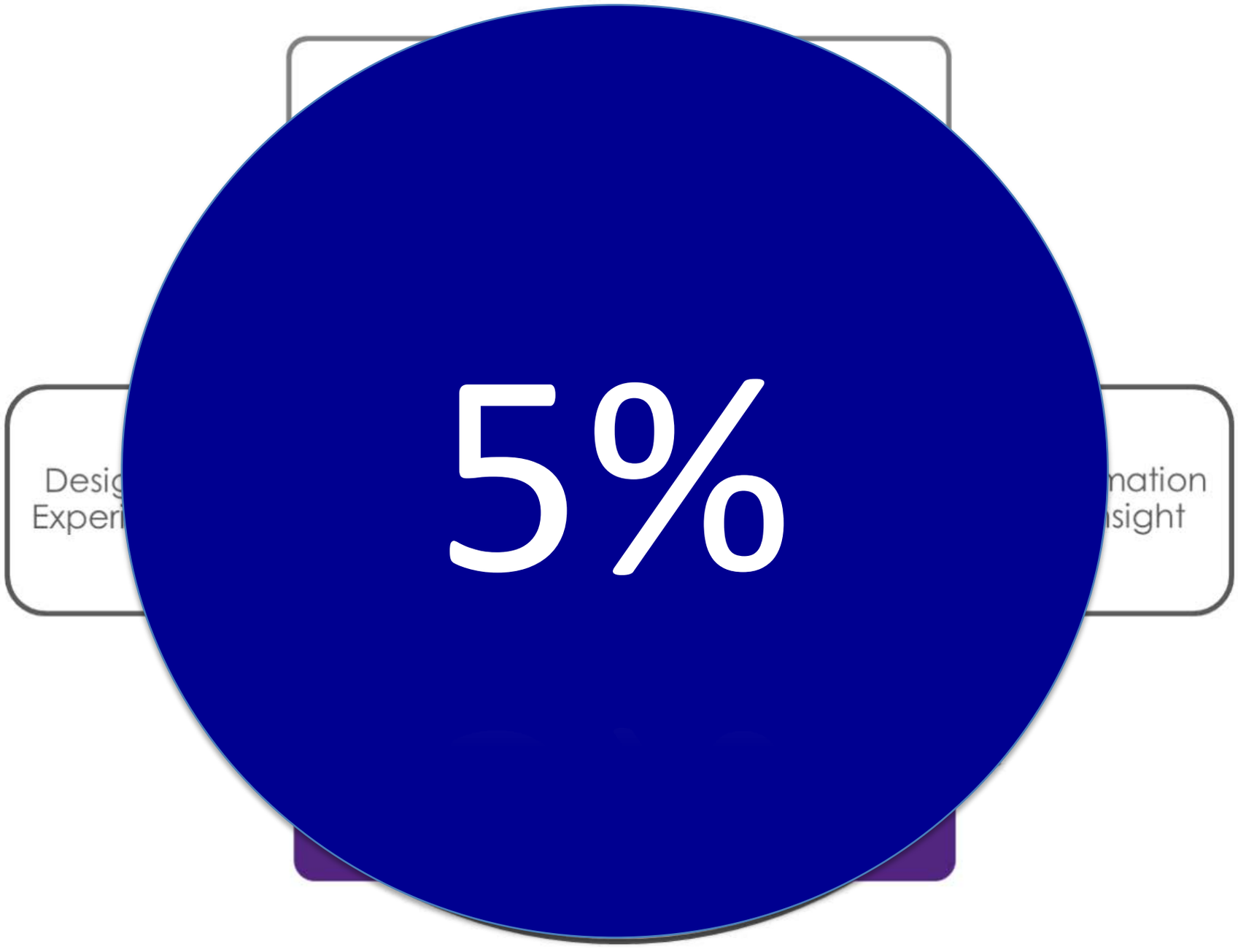
Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E Bourne, Jildau Bouwman, Anthony J Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J G Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A.C. 't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao, and Barend Mons

http://www.nature.com/sdata/ nature publishing group npg

Data Stewardship is NOT simple !!!!



The Data Stewardship Cycle



EOSC: **Framing**

- **Trusted access** to **services** & **systems**
- **Re-use** of shared **data**
- **Across** disciplinary, social and geographical **borders**
- **Federated** environment, across Member States

EOSC: **‘Internet approach’**

- **Minimal** international guidance and governance
- **Maximum** freedom to implement.
- **Globally** interoperable and accessible
- **Globally** embedded in a **‘Commons’**

EOSC: **Scope**

- **Human expertise**
- **Core resources**
- **Standards, best practices**
- **underpinning technical infrastructures**
- **A web of Data and Services**

EOSC: **Supports**

- **Open Science**
- **Open Innovation**
- **Systematic and professional data management**
- **Long term data stewardship**

EOSC: **Challenges and Observations**

- The majority of the challenges are **social** rather than **technical**
- Not just t**he size of data,** but in particular **complex data** and **analytics across domains**.
- Shortage of **data experts** globally and in the European Union
- **Archaic system of rewards** and **funding** of science and innovation
- '**Valley of death**' between **(e-)infrastructure providers** and **domain specialists.**
- **Short funding cycles** of **core research infrastructures** are **not fit for purpose**
- **Fragmentation** between domains causes **repetitive** and **isolated** solutions
- Distributed data sets increasingly **do not move** (**size & privacy** reasons)
- Centralised HPC is **insufficient** to support **distributed meta-analysis and learning**.
- However, the **major components** for a **first generation EOSC** are largely 'there'
- But '**lost in fragmentation**' and spread over 28 Member States.

EOSC: **Key requirements**

- **New modes** of scholarly communication
- **Modern reward** and recognition practices need to support data sharing and re-use
- **Innovative**, fit for purpose **funding schemes** for sustainable underpinning infrastructures
- Core **data experts** need to be trained and their career perspective significantly improved
- Cross-disciplinary **collaboration-specific measures** for review, funding and infrastructure
- Support for the transition from **scientific insights** towards **societal innovation**
- The EOSC needs to be developed as an **eco-system of infrastructures**
- Key Performance Indicators should be developed for the EOSC
- The EOSC should **enable automation of data processing** and thus **machine actionability** is key.
- FAIR principles

EOSC: **Policy Recommendations**

- P1: Take immediate, affirmative action in close concert with Member States
- P2: Close discussions about the ‘perceived need’
- P3: Build on existing capacity and expertise where possible
- P4: Frame the EOSC as supporting Internet based protocols & applications

EOSC: **Governance Recommendations**

- G1: Aim at the lightest possible, internationally effective governance
- G2: Guidance only where guidance is due
- G3: Define Rules of Engagement for formal participation in the EOSC
- G4: Federate the Gems across Member States

EOSC: **Implementation Recommendations**

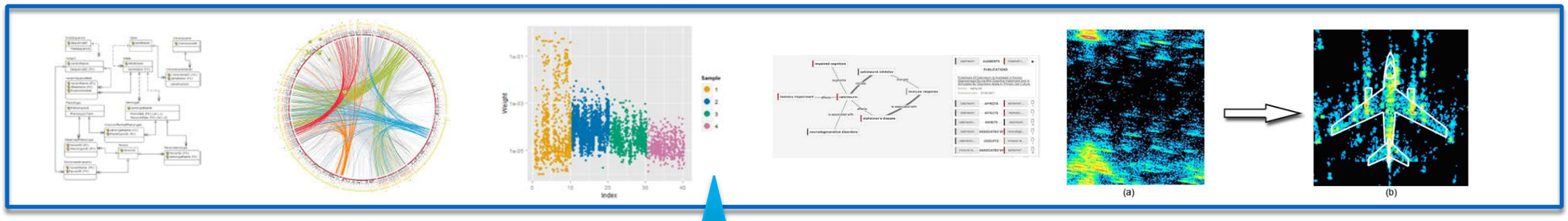
- I1: Turn this report into an EC approved White Paper to guide EOSC initiative
- I2: Develop, Endorse and implement a Rules of Engagement scheme
- I3: Fund a concentrated effort to locate and develop Data Expertise in Europe
- I4: Install a highly innovative guided funding scheme for the preparatory phas
- I5: Make adequate data stewardship mandatory for all research proposals
- I6: Install an executive team to deal with international coherence of the EOSC
- I7: Install an executive team to deal with the preparatory phase of the EOSC

PDF

confirmational reading

Actionable Knowledge

Pattern recognition



Analysis transformation (mostly interlinking and optimisation)

EOSC API's

FAIR (meta)data
(RDF, XML etc.)

FAIR transformation

FAIR download
(in local format)

processed data
(any storage format)

provenance

provenance

Raw data
(many formats)

Hard infrastructure and repositories

?