European Commission

RESEARCH & INNOVATION

Open Science Cloud

Open Science as a Social Machine

&

the support through

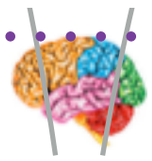
The European Open Science Cloud

2 min.

of

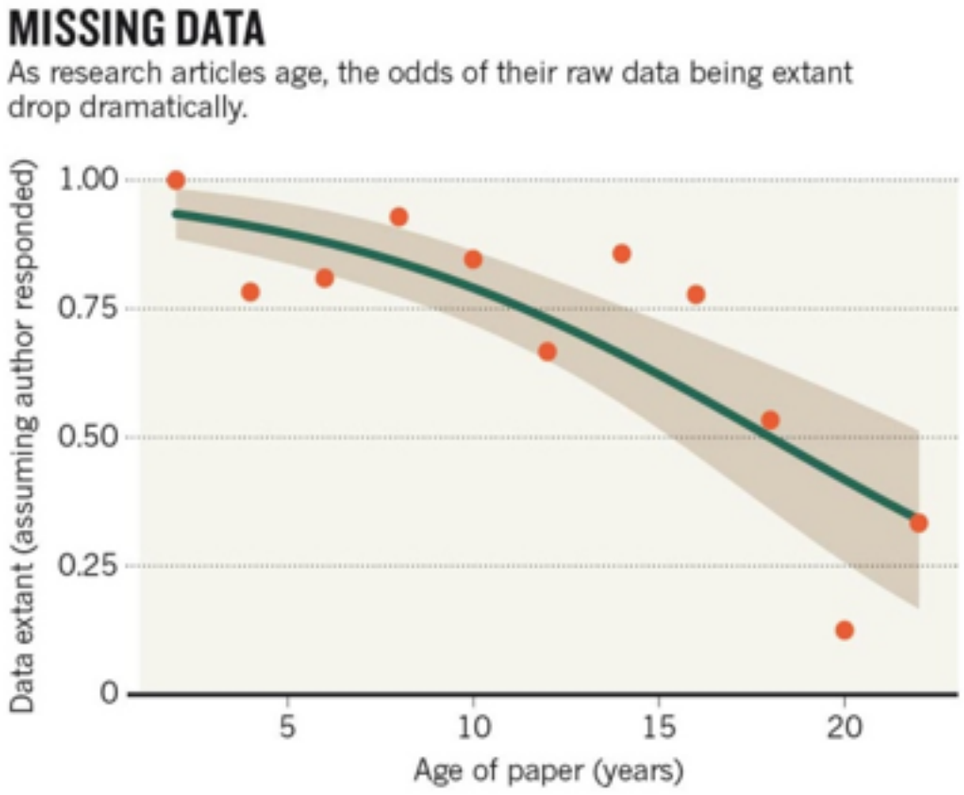
Lamenting

From Thomas Young to Social Machines

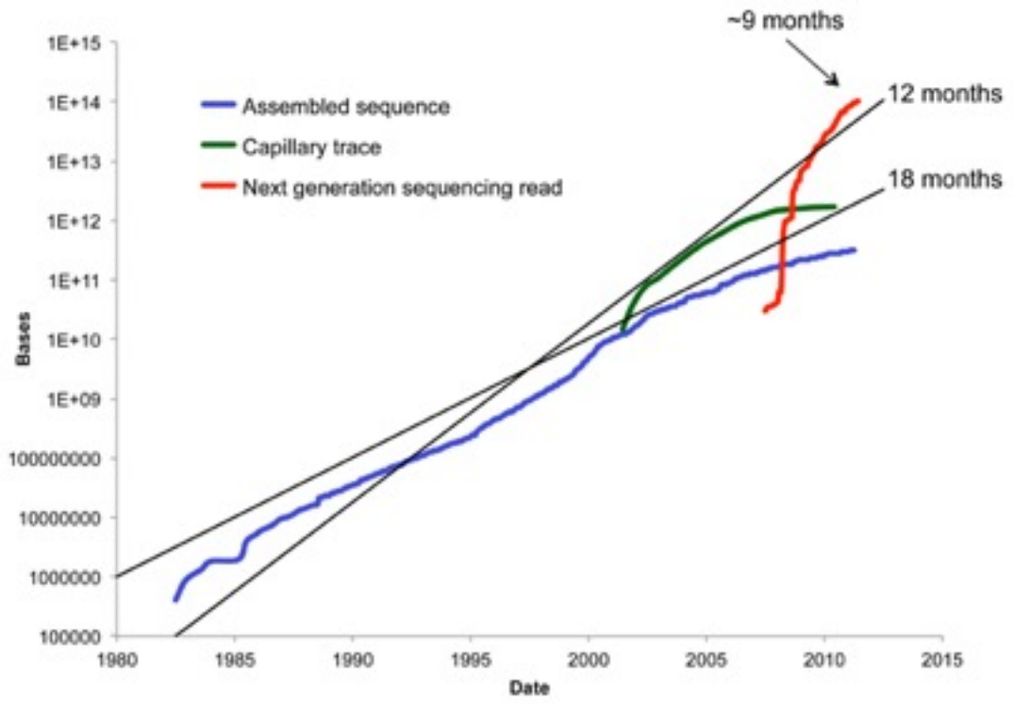


Science1.0

Data loss is real and significant, while data growth is staggering



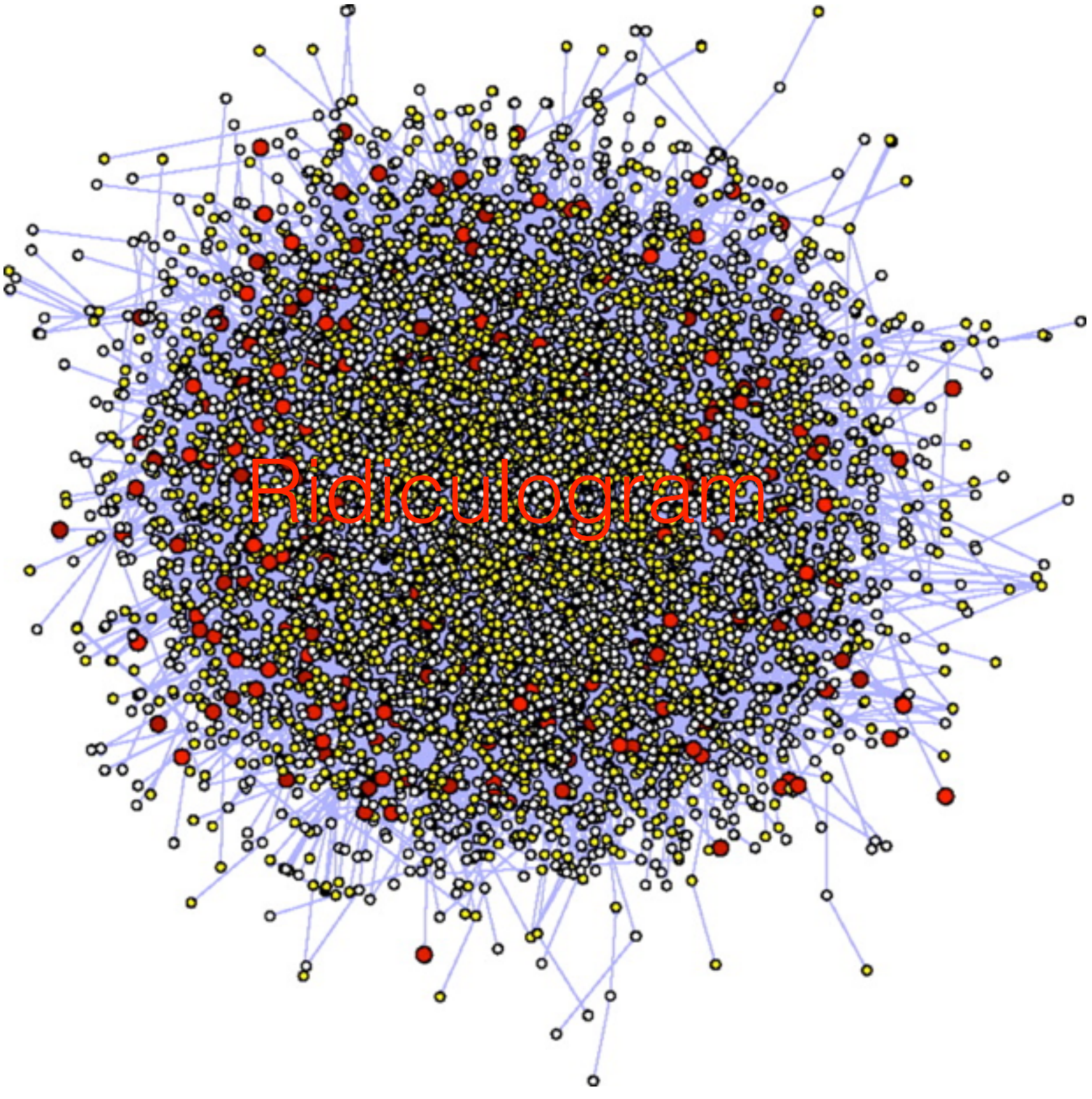
Nature news, 19 December 2013



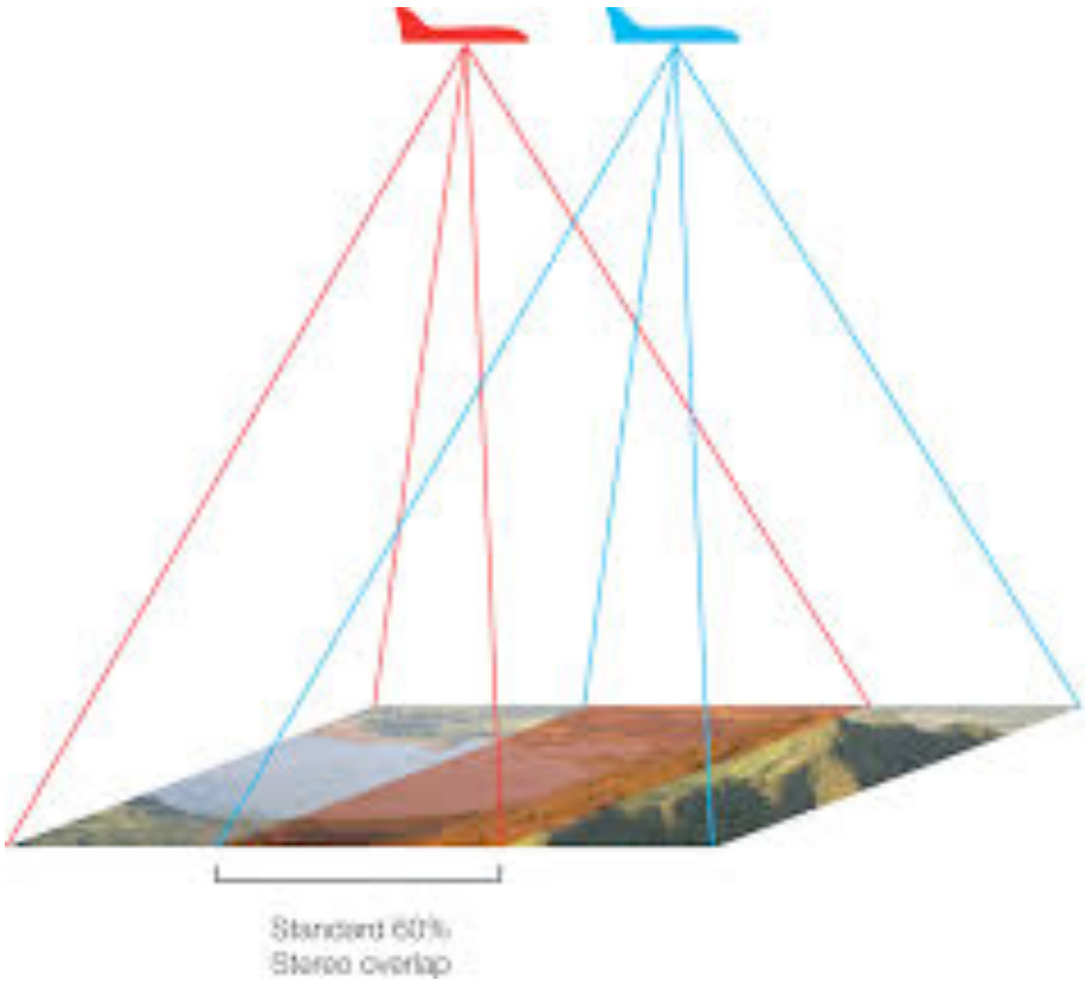
'Oops, that link was the laptop of my PhD student'

- Computer speed and storage capacity is **doubling every 18 months** and this rate is steady
- DNA sequence data is **doubling every 6-8 months** over the last 3 years and looks to continue for this decade

Simplified e-Science



The Explicitome is estimated at 10^{14} assertions



The Explicitome is spread over > 1600 databases

The ‘Cardinal Explicitome’ is estimated to be ‘only’ 10^{11} assertions

- European
- Open
- Science
- Cloud

- European
- Open Science
- Cloud

The **EOSC**

Open Science

OA (articles)

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 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

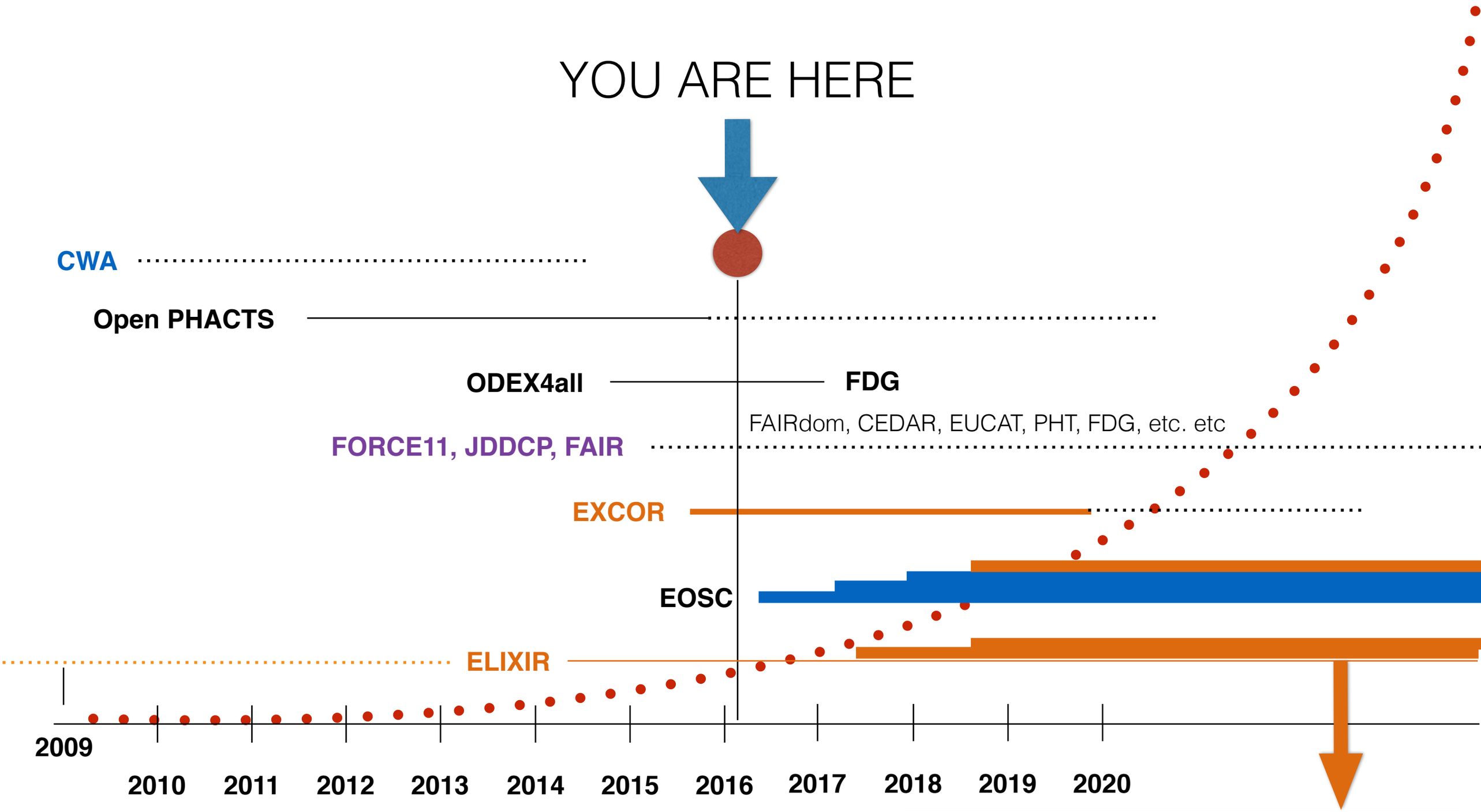
EOSC: **Governance Recommendations**

- 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

YOU ARE HERE



95% MS

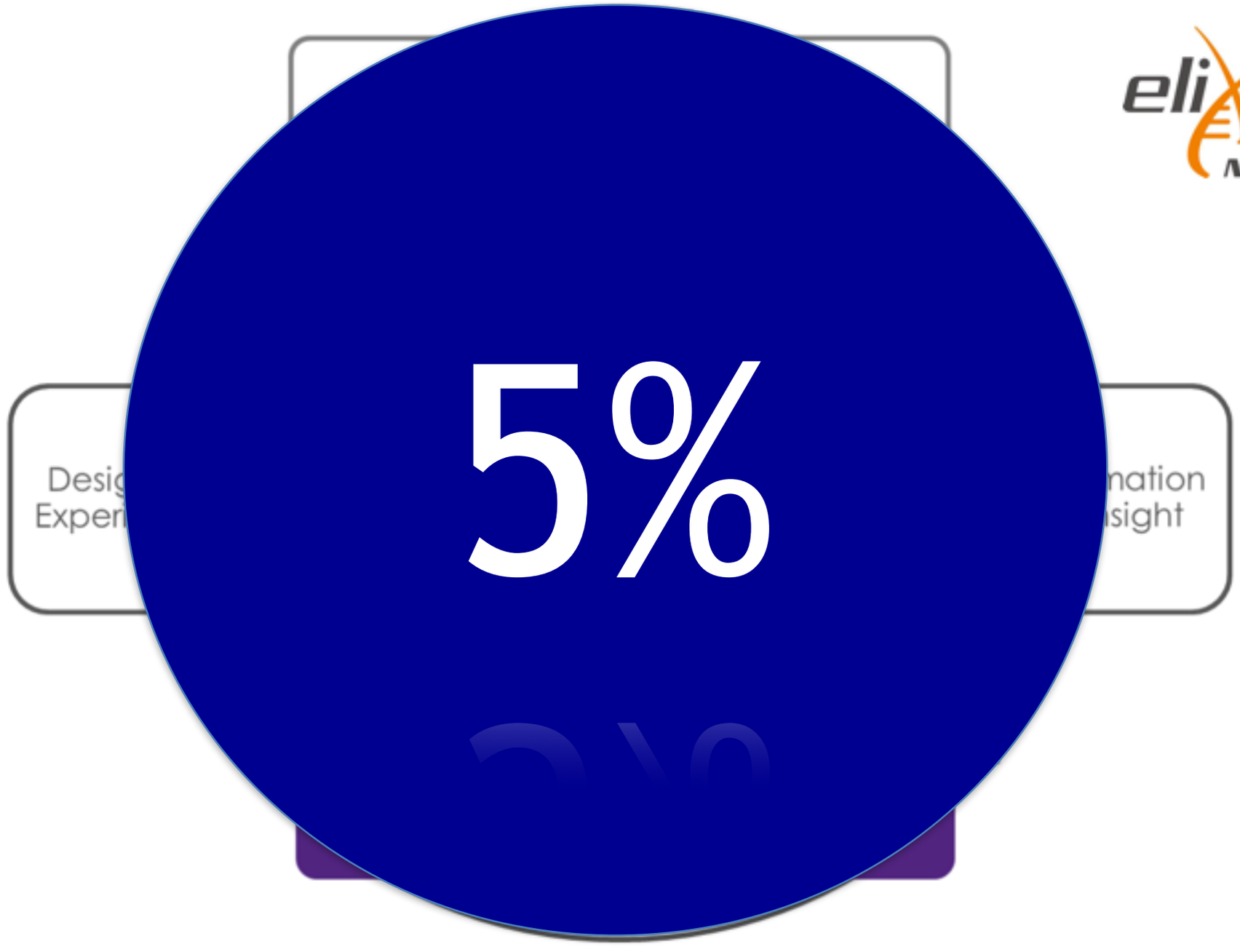
5% EC

Data Management Plans
Mandatory for Research Projects H2020 & Member States

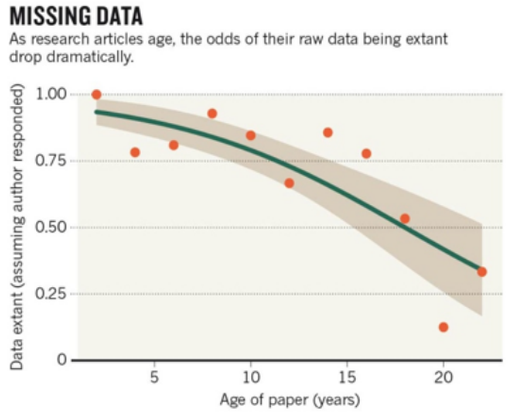
Long-Term Data Stewardship
How to finance ESFRI's and EBI SIB type + infra
Mainly private for reliability

Interoperability Backbones, Standards, Procedures
Mainly H2020 + ESFRI-type domain expertise

The Data Stewardship Cycle



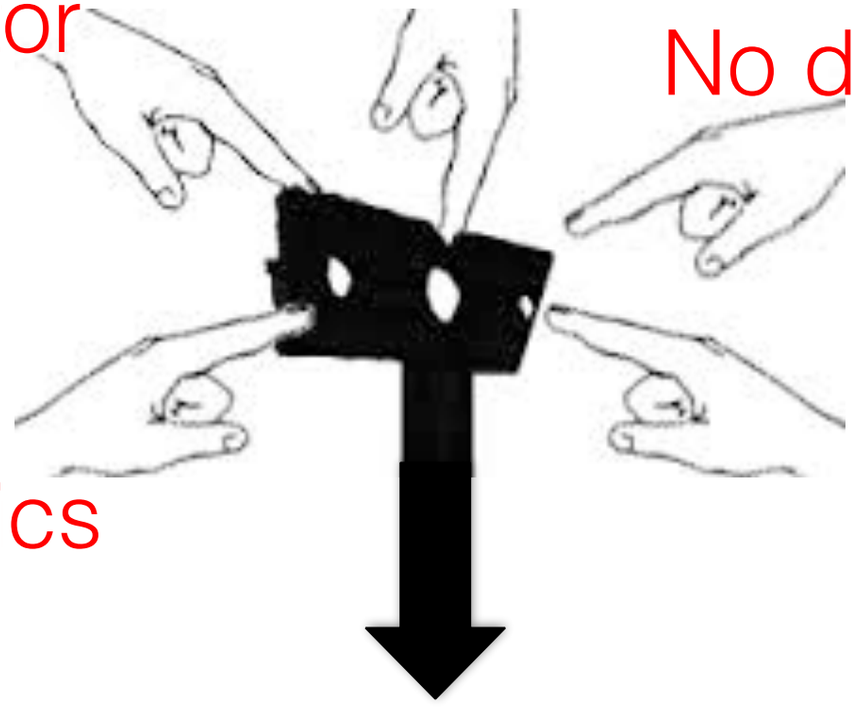
Malpractices…….



‘supplementary data’

Journal Impact Factor

No data stewardship plan



Ignore relevant metrics

Obstruct Tenure
Data Experts

Knowledge Sharing Impaired

?