Open Data: a scientist's perspective

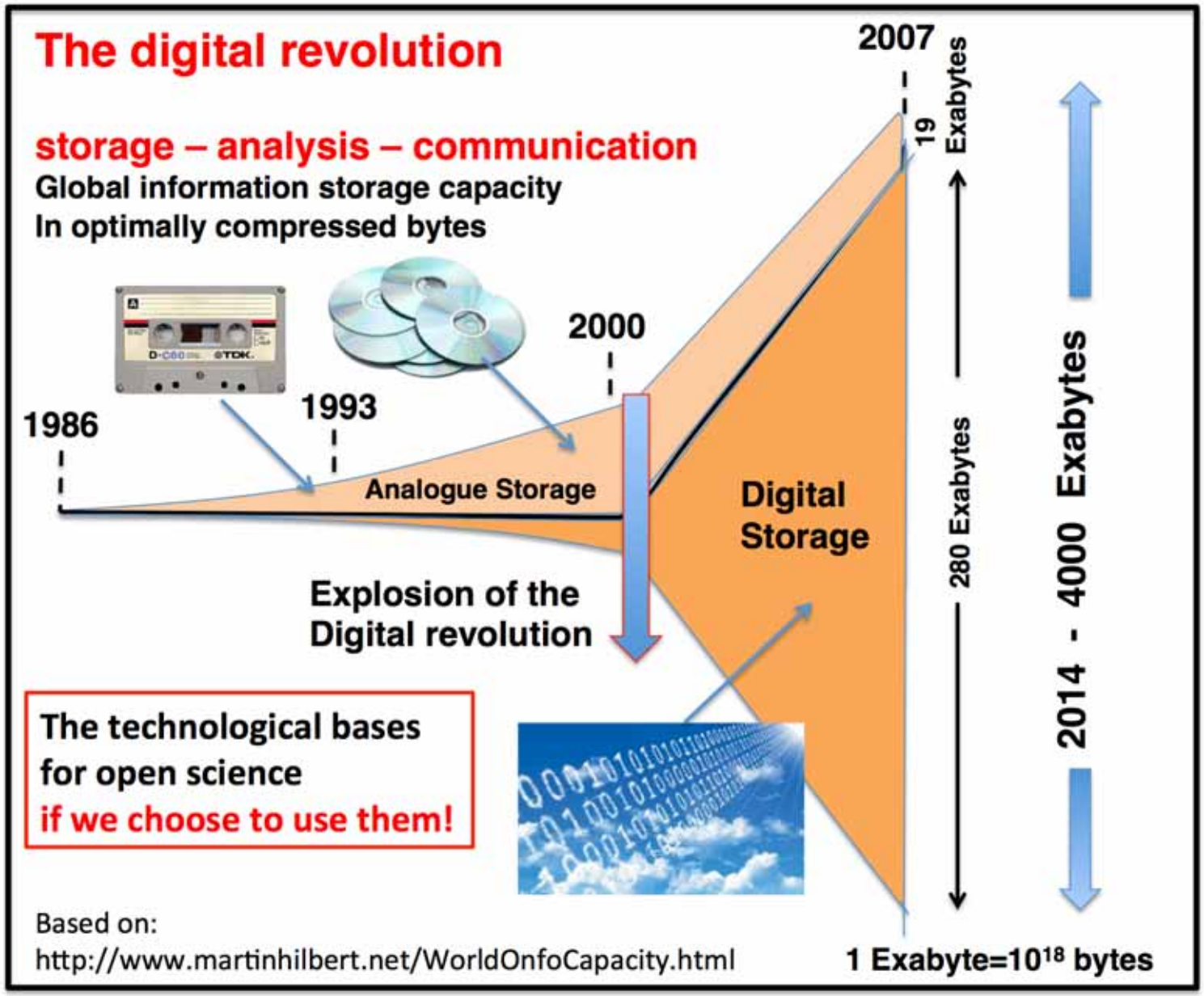
Geoffrey Boulton

CODATA & the University of Edinburgh

Argument:

- A world historical event: the "digital revolution"
- Response of scientific community thro' "Open Data"
- Fundamental opportunities/challenges for science
- Roles for European ICSU members?

The Digital Explosion



Another Gutenberg Revolution

The technologies by which knowledge is acquired, stored and communicated always been essential drivers of human material and social progress

Johannes Gutenberg
1400-1468

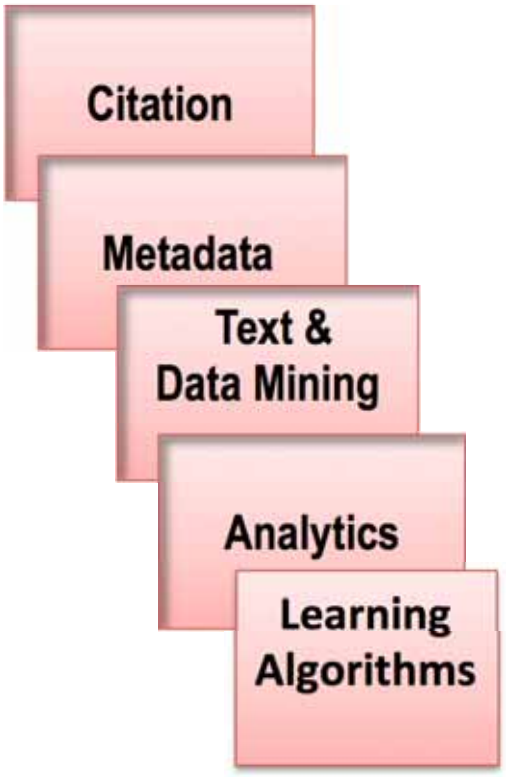
Response from the Science Community: Open Science

Conservative definition: new efficiencies and opportunities for discovery based on open data and open access – but that is still science talking to itself.

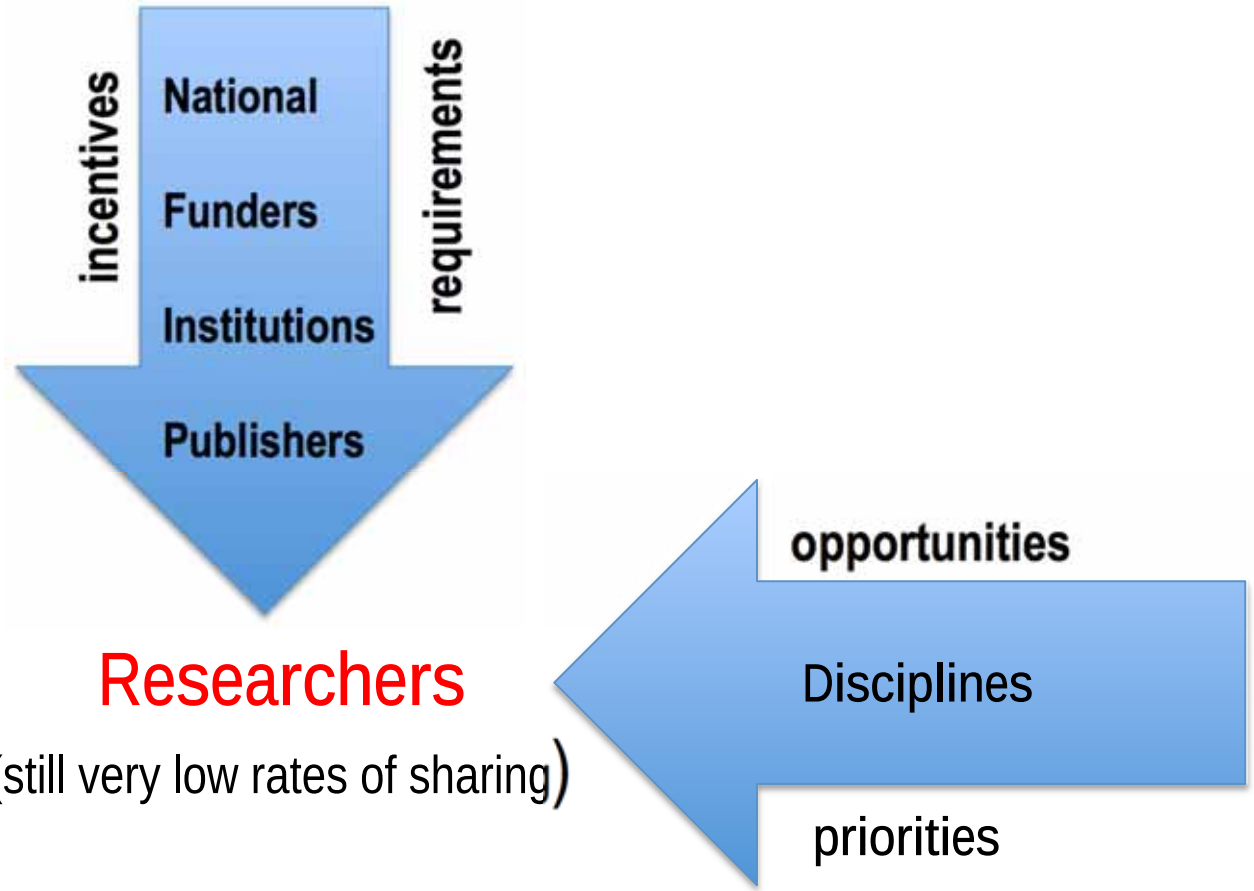
Radical definition: they are foundations for an "Open Science", that is integrated across the disciplines, societally engaged, open to all.

Enabling requirements

Principles-Policies-Management Services **Tools**



How are researchers influenced?



Researchers

(still very low rates of sharing)

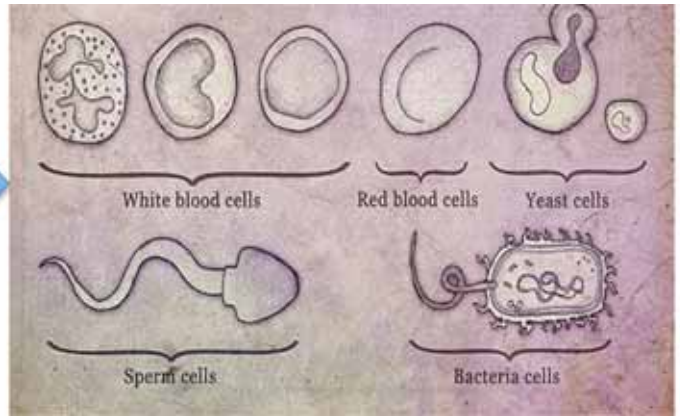
Q: So what does all this data offer to researchers?

A: Not causality, but the capacity to identify patterns in multi/hyper-dimensional data

Q: Is that all?

A: Recognition of new patterns leads to new questions and new answers – its what Darwin, Newton and Marx did

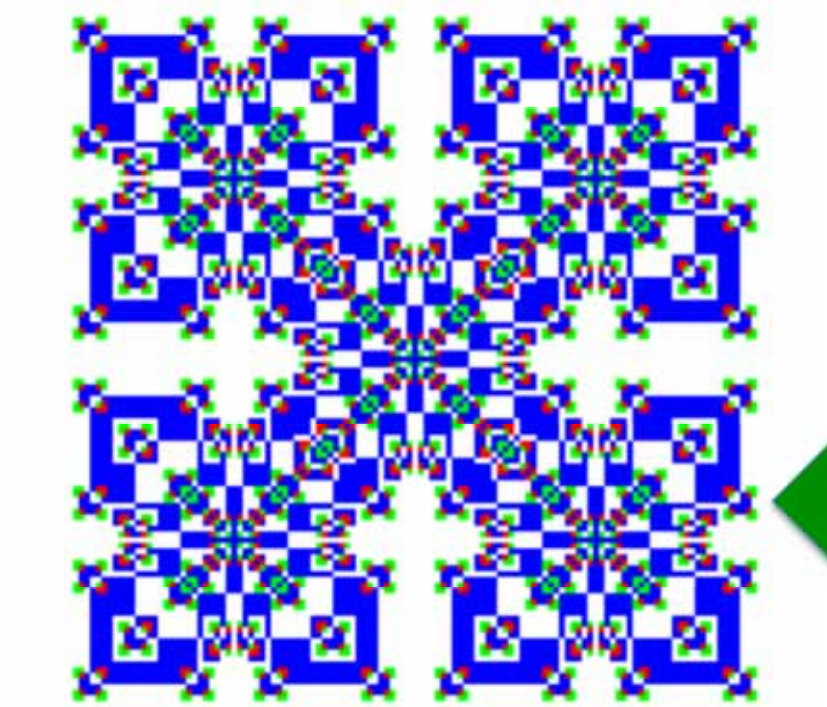
Remember Anthonie van Leeuwenhoek (1632-1723)



Leeuwenhoek saw a microscopic world that had not been seen before. We now have a "Digital Microscope".

Complexity: system state & dynamic evolution

Simulating system dynamics



Emergent behaviour of a specific 6-component coupled system

Mapping a complex state

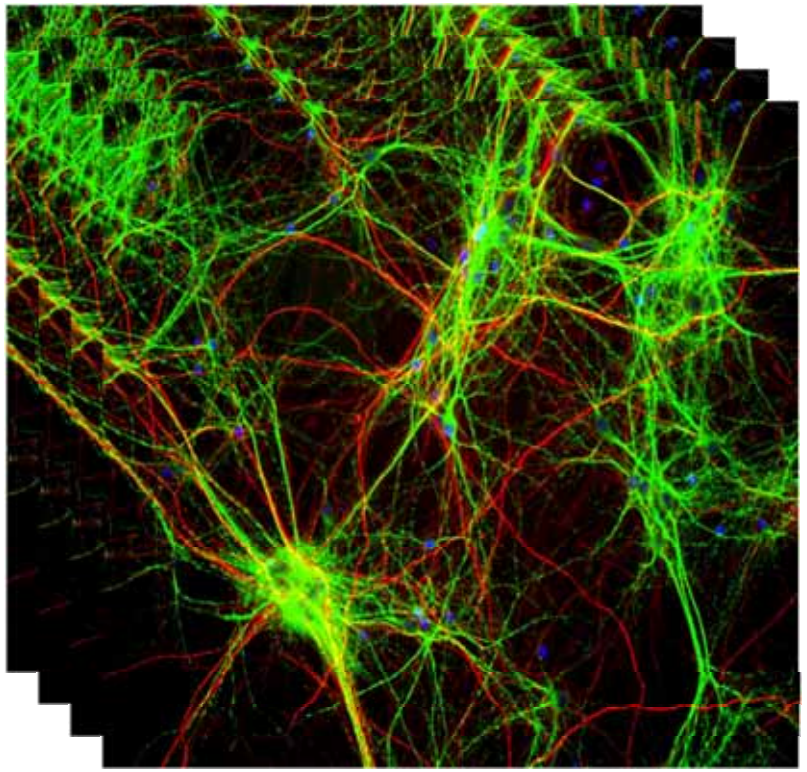
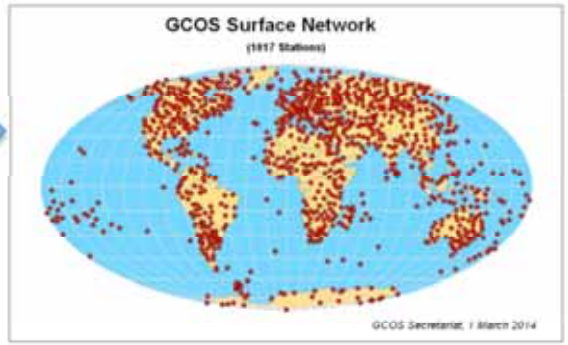


Image of brain cells in a rat

An example: data-modelling: iterative integration

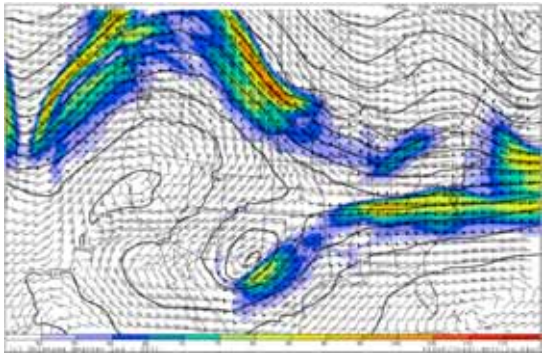
Satellite observation

Surface monitoring

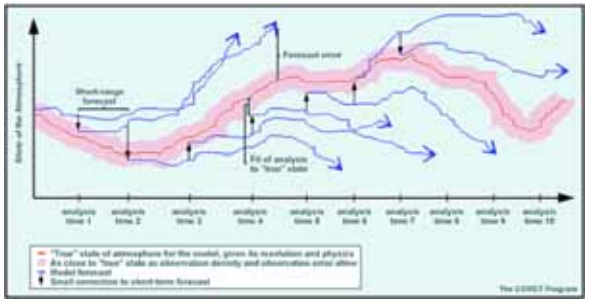


Initial conditions

Model forecast

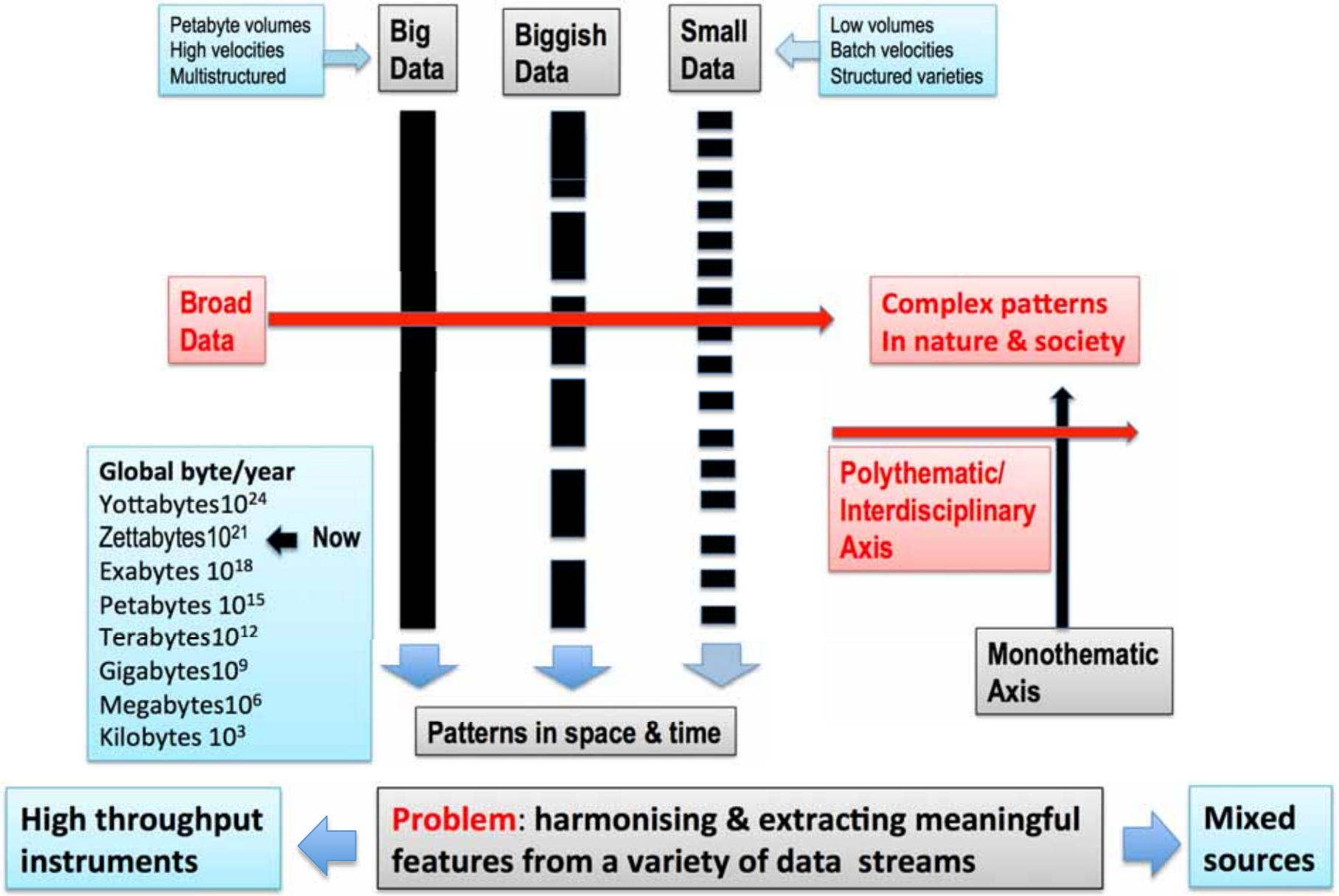


Model-data iteration - forecast correction using learning algorithms



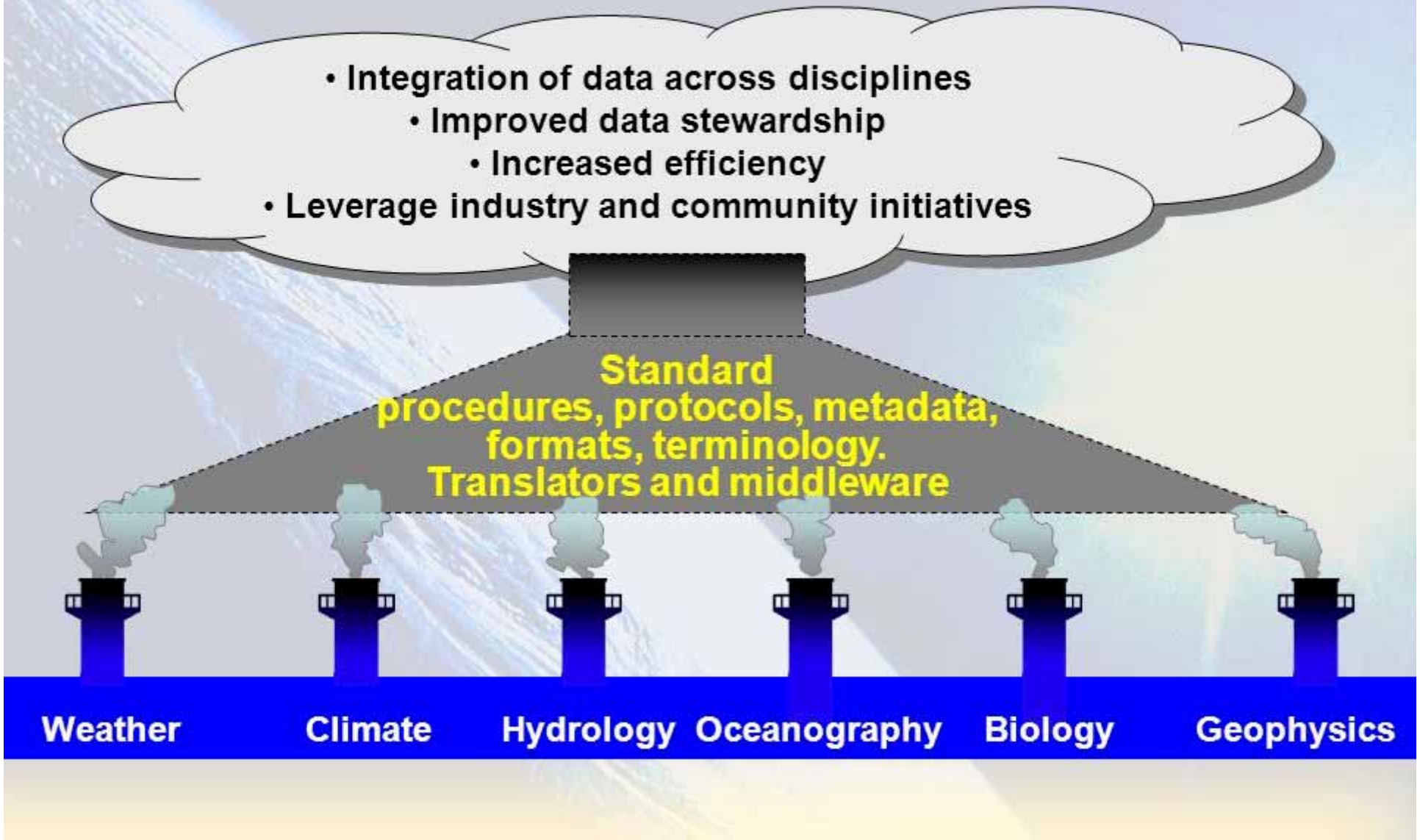
The Data Universe

Implications for interdisciplinary science



Integrated Data Environment

Bridging the gaps between stove-pipe systems

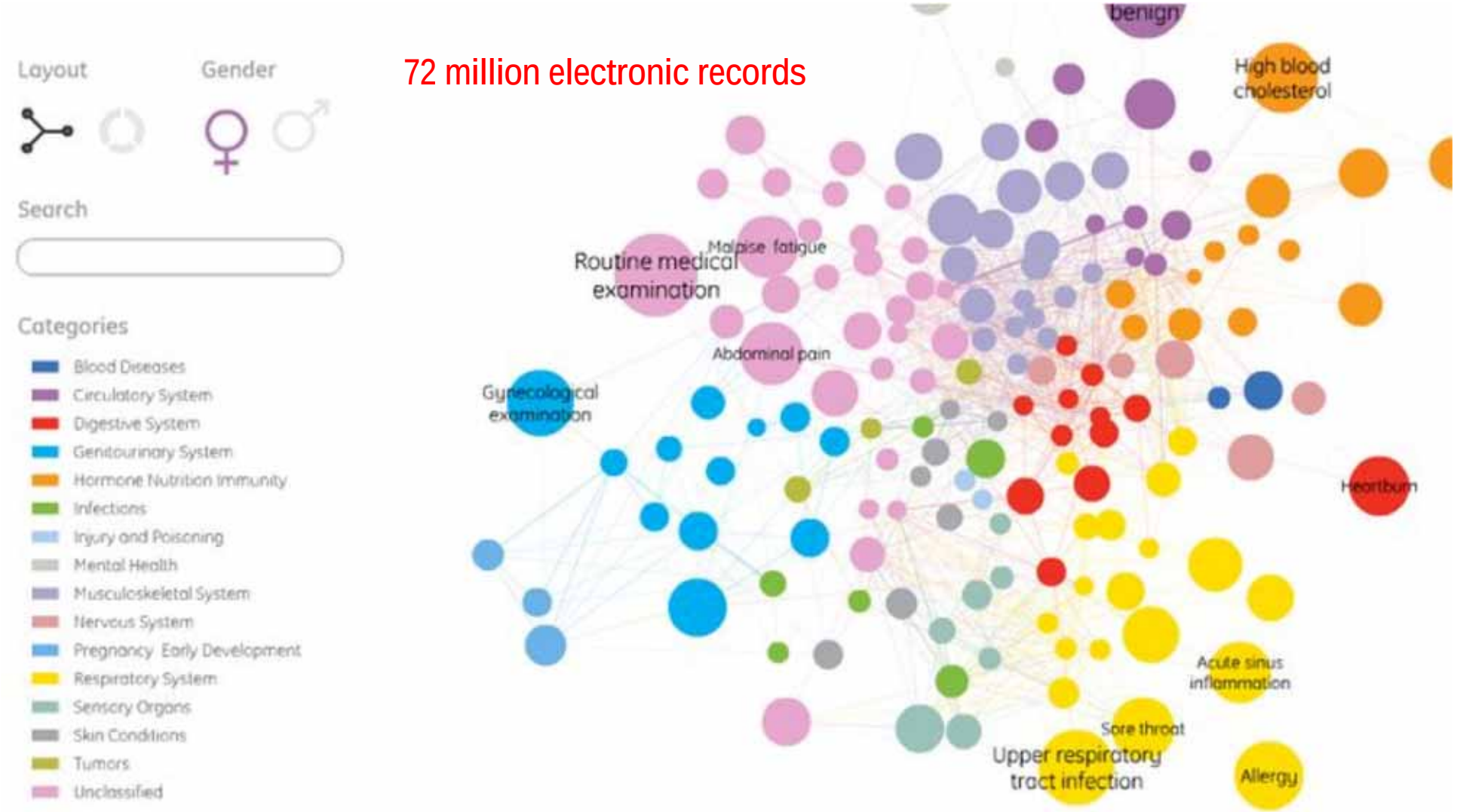


Big/Broad Data reveals patterns in nature and society have been beyond resolution

Example: North Atlantic Ocean Circulation

Example 2: You have a disease – what is the probability that you will suffer from another?

72 million electronic records



Lesson: The data integration needed to produce such results is achieved by information communities that know what they want.

The big problem – the last examples cover narrow disciplinary ranges broadband interdisciplinary complexity is much more demanding

In modelling

Integrated, coupled simulation is becoming the norm

In data

Integration tends to occur in relatively narrow siloes of limited disciplinary range that share common standards

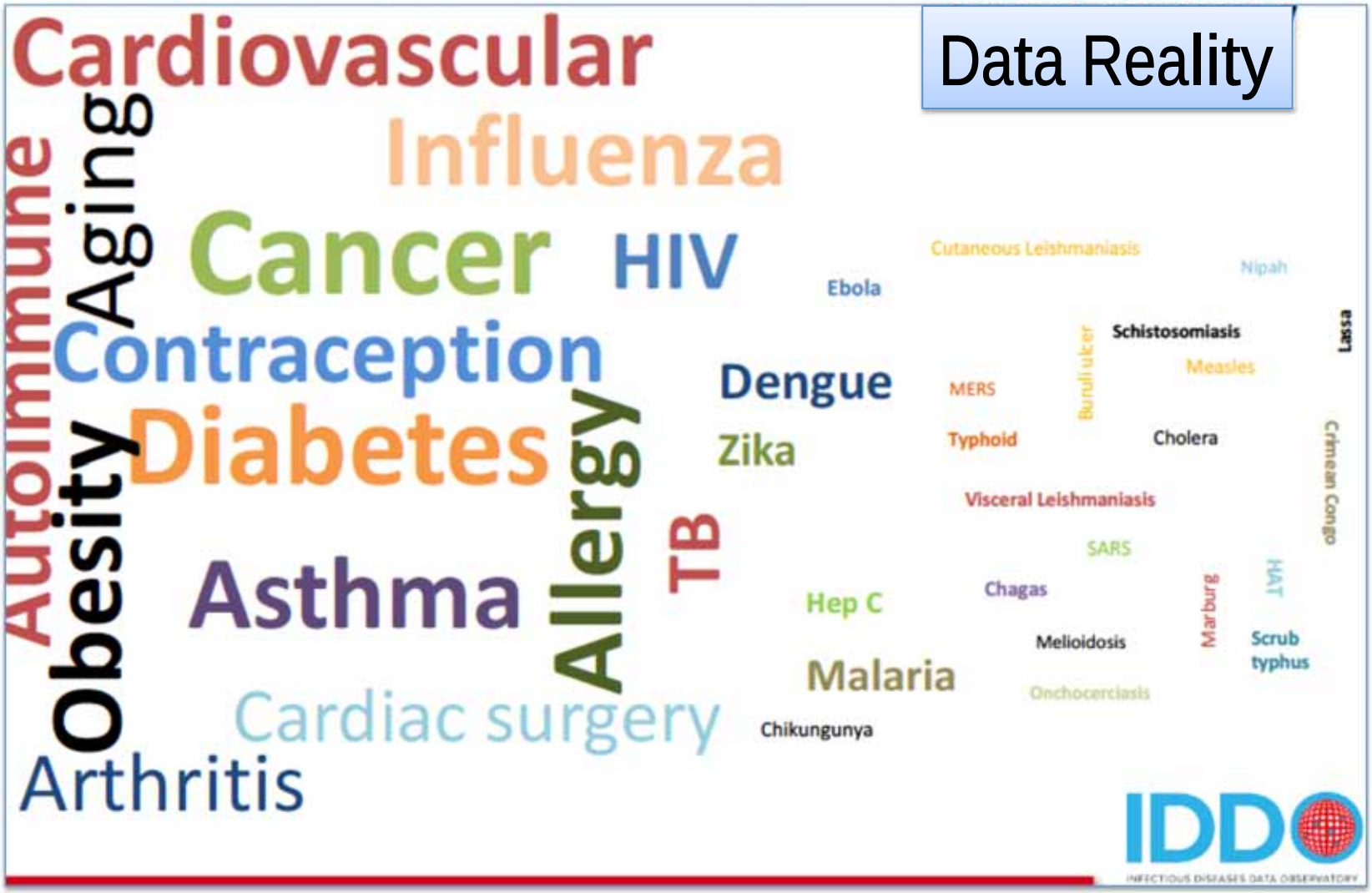
When using data from multiple sources and multiple disciplines, discovery & integration is a major effort because of the diversity of standards & ill-defined vocabularies

Shared, standardised terminologies would enable the broadband integration that is vital in deducing patterns in the data relating to major interdisciplinary challenges

Implementing standards can be costly …

… but not implementing standards is even more costly – the potential of our investments and of the EOSC will be reduced unless we do so.

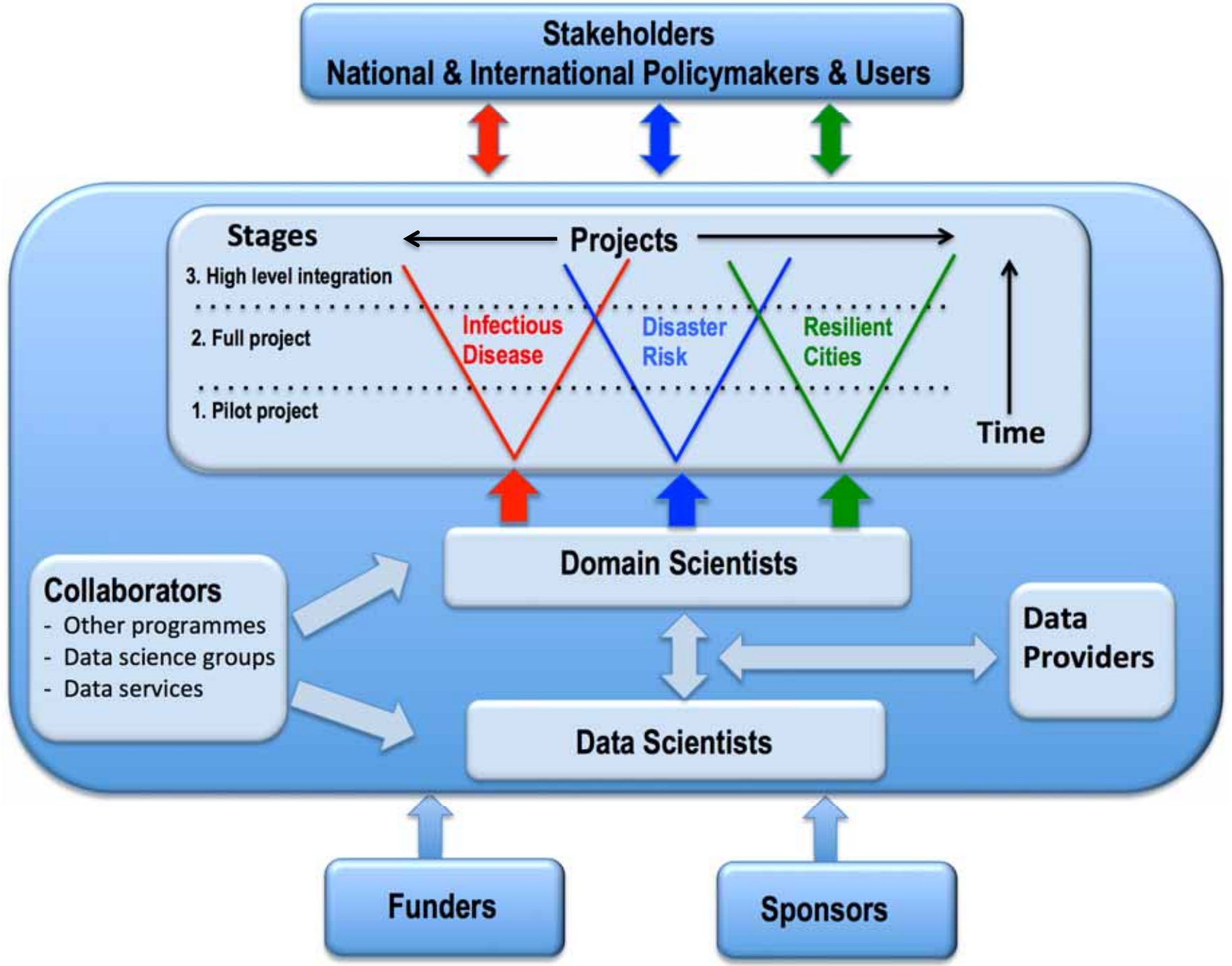
Example: infectious disease is a major global issue that requires broadband data integration



Vector-borne Infectious Diseases

Data relevant to vector-borne disease

- Agent biology
- Population genetics
- Water supply
- Sanitary engineering
- Travel & transport
- Social Habits
- Climate
- Population dynamics
- Faunal & floral dynamics
- Atmospheric dynamics
- Microbiotic dynamics


Stakeholders
National & International Policymakers & Users
Stages
Projects
3. High level integration
2. Full project
1. Pilot project
Infectious Disease
Disaster Risk
Resilient Cities
Time
Domain Scientists
Collaborators
- Other programmes
- Data science groups
- Data services
Data Providers
Data Scientists
Funders
Sponsors

African Open Science Platform

A managed virtual digital ecosystem coordinating resources to deliver to African researchers, institutions, communities, citizens

- Capacity building
- Principles, policies, practices, services software tools, analytic approaches
- Computational access
- Database access
- Networking with the international community
- Interaction with societal stakeholders

Recommendations to ICSU National Members

1. They should be advocates of the well-known Merton "norms of modern science" to this new world of EOSC:
   - Common ownership of research outputs.
   - Communalism: act for the community rather than personal gain.
   - Transparent.
   - Collective.
   - Open by default (therefore "Open Science" is a tautology).
2. Advocacy for "Open Science" in their national systems.
3. Some key concerns:
   - Data governance.
   - Maintaining data resources.
   - Privatisation of knowledge.
   - Enabling text and data mining.

Recommendations for EOSC

1. Funding to developing functionalities for broadband interdisciplinarity
2. Ensure that an Open Science strategy is not a means of commercial publishers accessing public funds