



Towards collaborative and FAIR science for a FACT



Jaap Heringa, ELIXIR-NL HoN

j.heringa@vu.nl

December 12th 2018

Modern life sciences are information sciences..

..and are becoming ever more inter-disciplinary

WHAT IS DRIVING LIFE SCIENCES

Data acquisition by high-throughput measurements

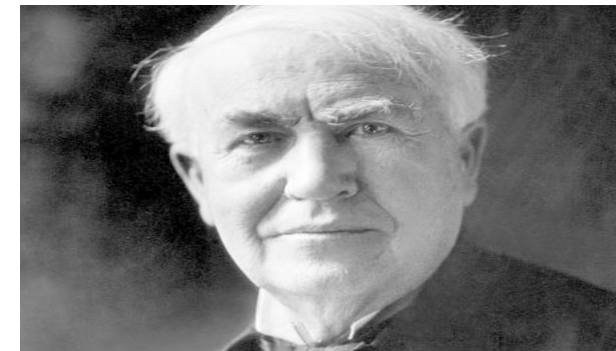
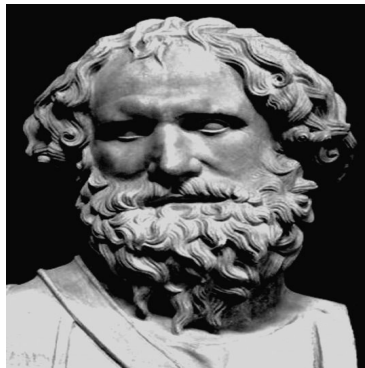
- Bio-sciences
 - Genomics: HTP measurements, e.g. Sequencing (NGS), Chip-seq
 - X-ray, NMR, Mass Spec
 - Imaging, optical measurement techniques, single cell measurements, single molecule tracking
 - Lots of new stuff coming up...

Enabled by IT in general

DATA

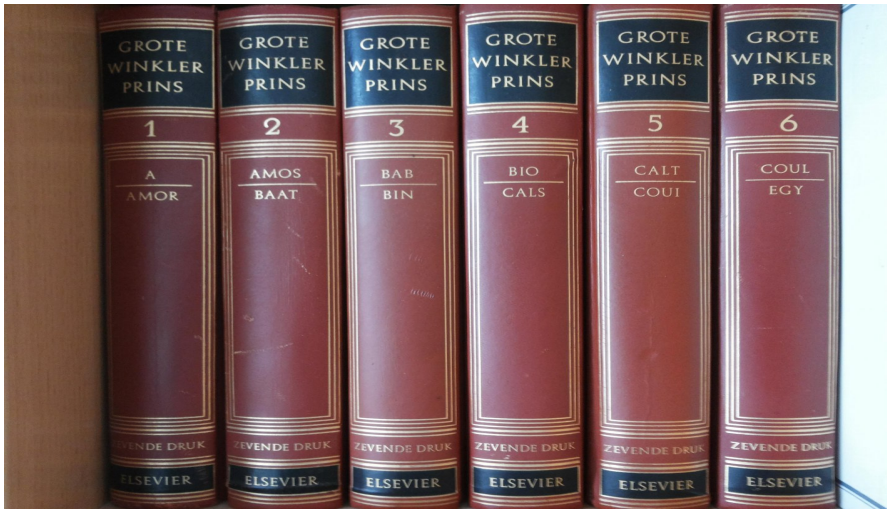


“Too often we forget that genius, too, depends upon the data within its reach, that even Archimedes could not have devised Edison’s inventions.” Ernest Dimnet.



DATA ACCESS AND FAKE NEWS

- Over the centuries, **access** to information has been a privilege of the elite.
- Now virtually everyone has access to information



- But how do we deal with the tsunami of information?
- And how do we treat our data?
- How accessible is scientific data?
- How do we know what is true and what is fake?

STEVEN KEATING – PHD STUDENT @MIT

GRADUATED 2016



STEVEN KEATING – BRAIN TUMOR

- Brilliant MIT student
- Participated in a brain study
 - MRI scan revealed a small abnormality.
- Brain rescanned in 2010, which confirmed that nothing had changed.
- In 2014, he started smelling a faint vinegar-like scent for several seconds each day.
- Keating remembered that the abnormality on his scan was near the smell centre of his brain.
- Doctors were not very concerned, but another scan was done, showing an extremely large tumor that covered around 10 percent of his brain.
- Three weeks later, Keating had a 10-hour “awake surgery”

STEVEN KEATING ON ACCESSING HIS OWN MEDICAL DATA

- Decided to research his own medical data
- It turned out to be **shockingly difficult** to obtain his own medical data.
 - Received dozens of CDs in the mail, but **no tools to use** the highly complex information they contained.
- **Legal gray zones** presented an even bigger challenge.
 - For instance, federal policies on clinical lab tests performed on humans prevented Keating from **accessing his own tumor's genome sequence** from a research study.

STEVEN KEATING ON ACCESSING HIS OWN MEDICAL DATA

“Why couldn’t there be a hospital “Share” button, under the control of the patient? A Google Maps for health? A hospital App Store—with third-party tools enabled by APIs—that would let patients easily access, use, and even share their data?”



DIRK-JAN VAN DER POL

Health-RI Conference,
December 2017



Dirk-Jan van der Pol is an ICT consultant with over 18 years of experience in industry and healthcare. In 2003 he was diagnosed with cancer, and a few years later with COPD. At the Health-RI conference on 8 December 2017, Van der Pol will take the audience on his journey from ambitious consultant to insecure patient, to director of his own health(care), and finally to researcher on his own data.

DIRK-JAN VAN DER POL

After three-four years of hard work he has established various links and is now effectively a world expert on the link Thymus gland-COPD.



→ with a FAIR data infrastructure in place a scientist (or citizen!) could do this research and reproduce the results in 15 minutes....

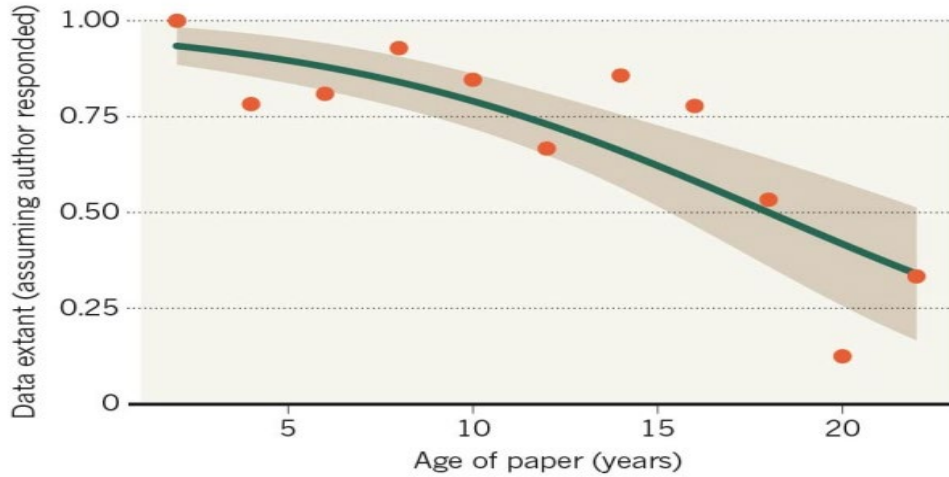
→ Also: Scientific databases contain a lot more information than scientific literature repositories - A FAIR infrastructure will open this up.



BIG DATA: Two problems - data loss & growth

MISSING DATA

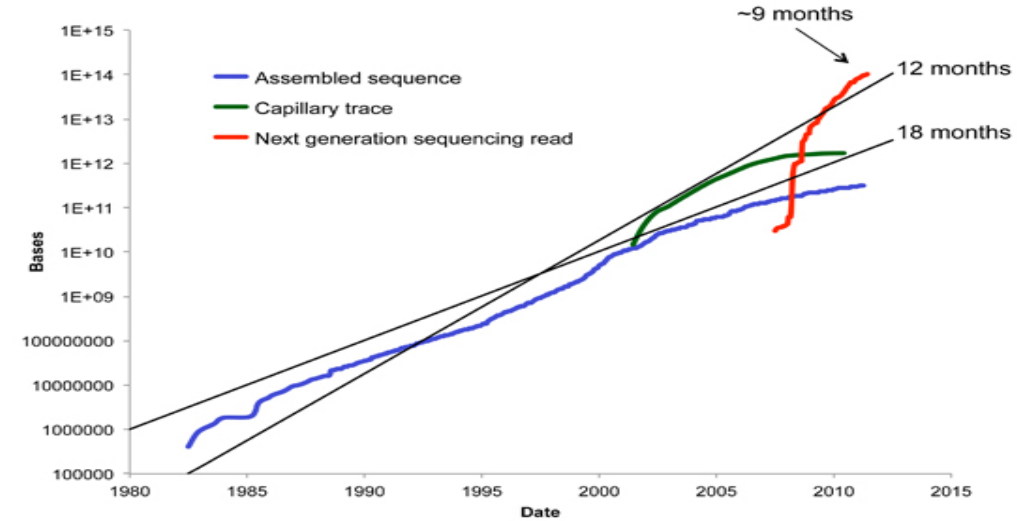
As research articles age, the odds of their raw data being extant drop dramatically.



Nature news, 19 December 2013



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



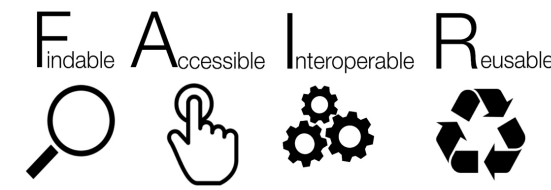
- Computer speed and storage capacity is **doubling every 18 months** and this rate is steady (Moore's law)
- DNA sequence data is **doubling every 6-8 months** over the last 3 years and looks to continue for this decade

WHAT DO WE NEED GIVEN THESE STORIES

- We need a data services infrastructure able to facilitate and use FAIR (*Findable, Accessible, Interoperable, Reusable*) data.
- We need an analytics ecosystem with FAIR-compliant tools to unleash the potential of collective data repositories.
- We need secure, scalable, flexible and hybrid cloud solutions to be able to serve the research community and operate at scale
- The infrastructure should be ELSI (*Ethical, Legal & Social Implications*) aware
- We need experts (data stewards, translation experts) to prepare and assist the (research) community

- 2014: FAIR (Findable, Accessible, Interoperable, Reusable) data principles launched at Leiden Lorentz meeting.
- 2016: G20 adopt FAIR Principles
- 2017: Open European Science Cloud stipulates FAIR principles
- 2017: G7 adopt FAIR principles
- 2017: ELIXIR ESFRI bases its platforms on FAIR principles
- 2017: Science funders (NWO in The Netherlands) stipulate adherence
- 2017: GO-FAIR initiative endorsed by Dutch, German and French Governments

WHAT IS FAIR DATA?



FAIR Data aims to support existing communities in their attempts to enable valuable scientific data and knowledge to be published and utilised in a 'FAIR' manner.

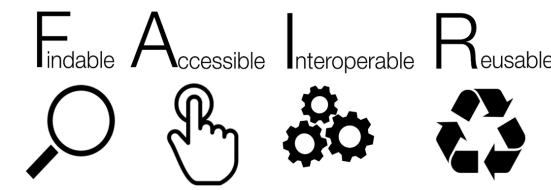
Findable - (meta)data is uniquely and persistently identifiable. Should have basic machine readable descriptive metadata.

Accessible - data is reachable and accessible by humans and machines using standard formats and protocols.

Interoperable - (meta)data is machine readable and annotated with resolvable vocabularies/ontologies.

Reusable - (meta)data is sufficiently well-described to allow (semi)automated integration with other compatible data sources.

FAIR DATA PRINCIPLES



Findable

- F1.** (meta)data are assigned a globally unique and persistent identifier;
- F2.** data are described with rich metadata;
- F3.** metadata clearly and explicitly include the identifier of the data it describes;
- F4.** (meta)data are registered or indexed in a searchable resource;

Interoperable

- I1.** (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2.** (meta)data use vocabularies that follow FAIR principles;
- I3.** (meta)data include qualified references to other (meta)data;

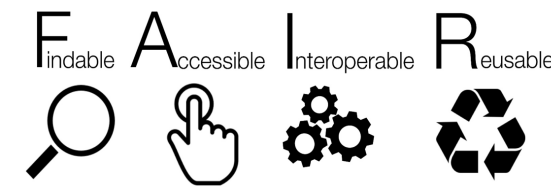
Accessible

- A1.** (meta)data are retrievable by their identifier using a standardized communications protocol;
 - A1.1** the protocol is open, free, and universally implementable;
 - A1.2.** the protocol allows for an authentication and authorization procedure, where necessary;
- A2.** metadata are accessible, even when the data

Reusable

- R1.** meta(data) are richly described with a plurality of accurate and relevant attributes;
 - R1.1.** (meta)data are released with a clear and accessible data usage license;
 - R1.2.** (meta)data are associated with detailed provenance;
 - R1.3.** (meta)data meet domain-relevant community standards;

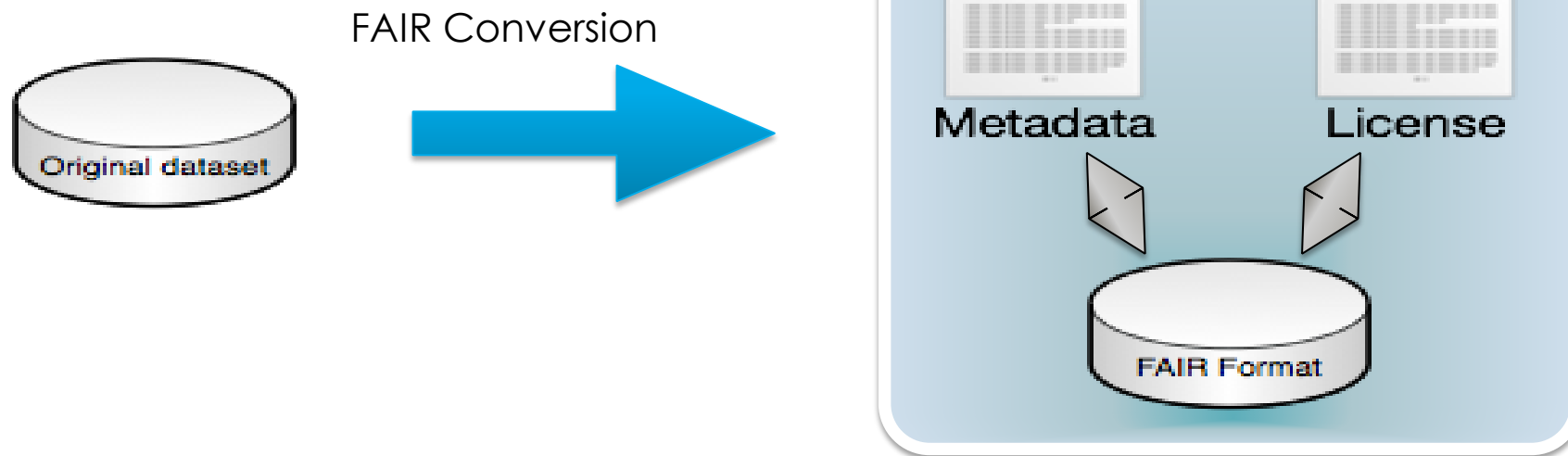
FAIR DATA RESOURCE



Datasets expressed using one of the prescribed standards of the FAIR Data Protocol.

The original dataset is transformed into a FAIR format and proper metadata and license are added to produce a FAIR Data Resource.

Original and the FAIR version can co-exist, each one fulfilling its own purpose.



FAIR DATA

Turning FAIR into reality

Final Report and Action Plan
from the European Commission Expert Group
on FAIR Data

European Commission Expert Group on FAIR Data

Sandra Collins, National Library of Ireland, Ireland: <https://orcid.org/0000-0003-2286-8540>

Françoise Genova, Observatoire Astronomique de Strasbourg, France: <https://orcid.org/0000-0002-6318-5028>

Natalie Harrower, Digital Repository of Ireland, Ireland: <https://orcid.org/0000-0002-7487-4881>

*Simon Hodson, CODATA, France, **Chair of the Group**: <https://orcid.org/0000-0003-3179-7270>*

*Sarah Jones, Digital Curation Centre, UK, **Rapporteur**: <https://orcid.org/0000-0002-5094-7126>*

Leif Laaksonen, CSC-IT Center for Science, Finland: <https://orcid.org/0000-0002-2161-4461>

Daniel Mietchen, Data Science Institute, University of Virginia, USA: <https://orcid.org/0000-0001-9488-1870>

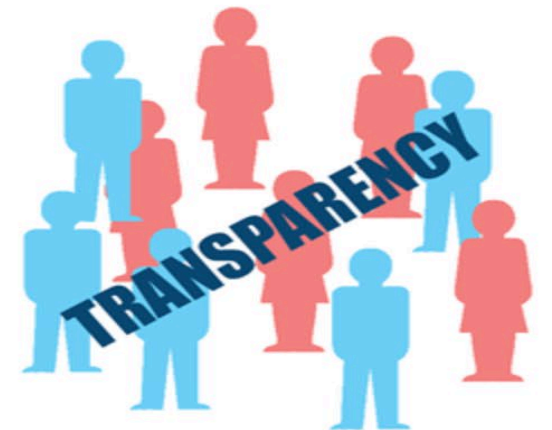
Rūta Petrauskaitė, Vytautas Magnus University, Lithuania: <http://orcid.org/0000-0002-6948-3202>

Peter Wittenburg, Max Planck Computing and Data Facility, Germany: <https://orcid.org/0000-0003-3538-0106>

RESPONSIBLE DATA SCIENCE - FACT

WE NEED A GENERAL DATA INFRASTRUCTURE
TO ENABLE COMPLEX COLLABORATIVE
SCIENCE PROJECTS

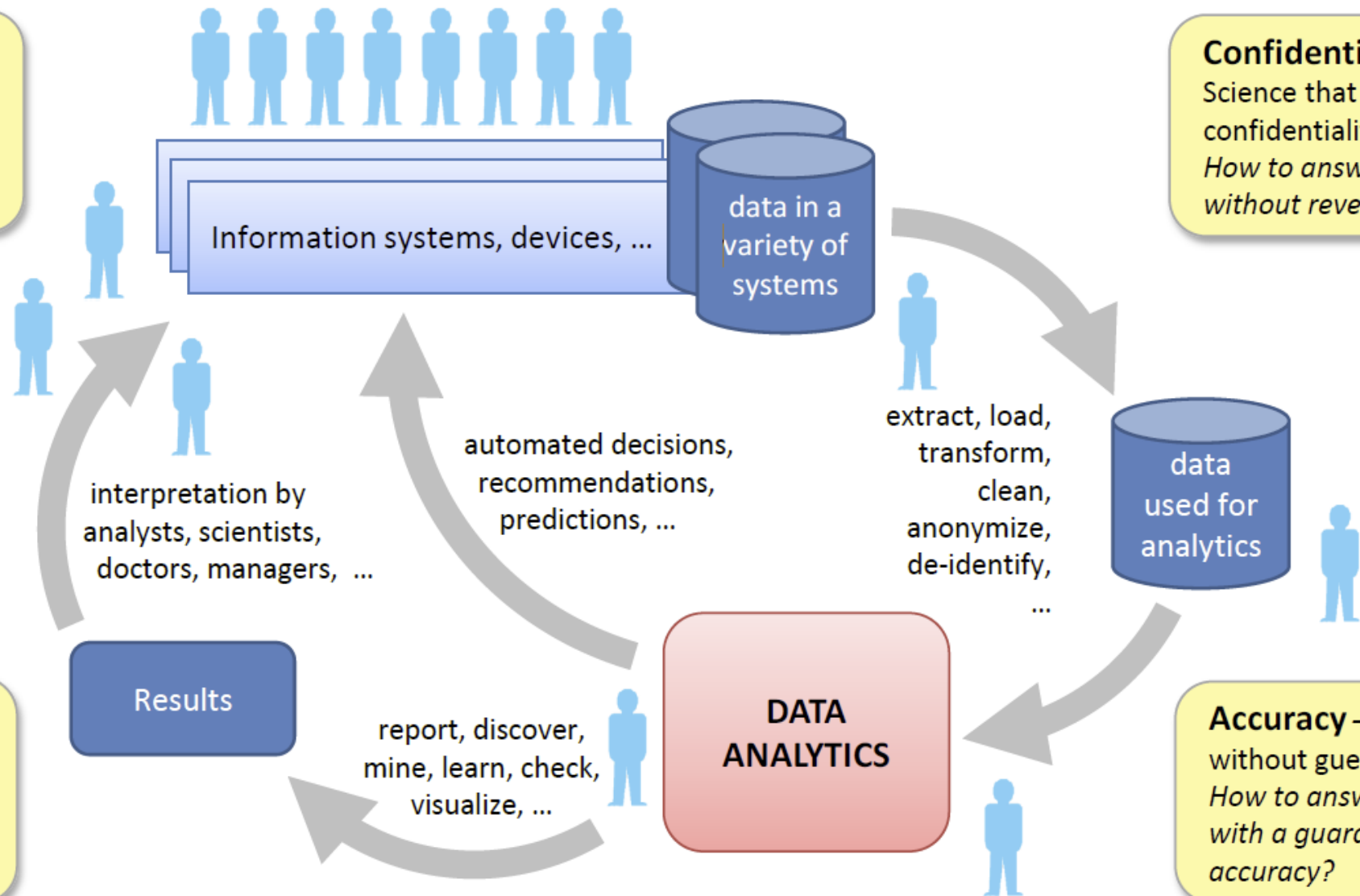
RESPONSIBLE DATA SCIENCE



RESPONSIBLE DATA SCIENCE - FACT

Transparency – Data Science that provides transparency: *How to clarify answers such that they become indisputable?*

Confidentiality – Data Science that ensures confidentiality: *How to answer questions without revealing secrets?*



Fairness – Data Science without prejudice: *How to avoid unfair conclusions even if they are true?*

Accuracy – Data Science without guesswork: *How to answer questions with a guaranteed level of accuracy?*

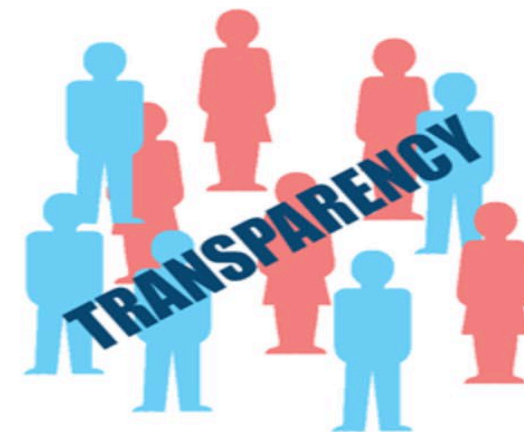
RESPONSIBLE DATA SCIENCE

RDS community working on four main remits:

- Responsible Science,
- Responsible Health,
- Responsible Business,
- Responsible Government

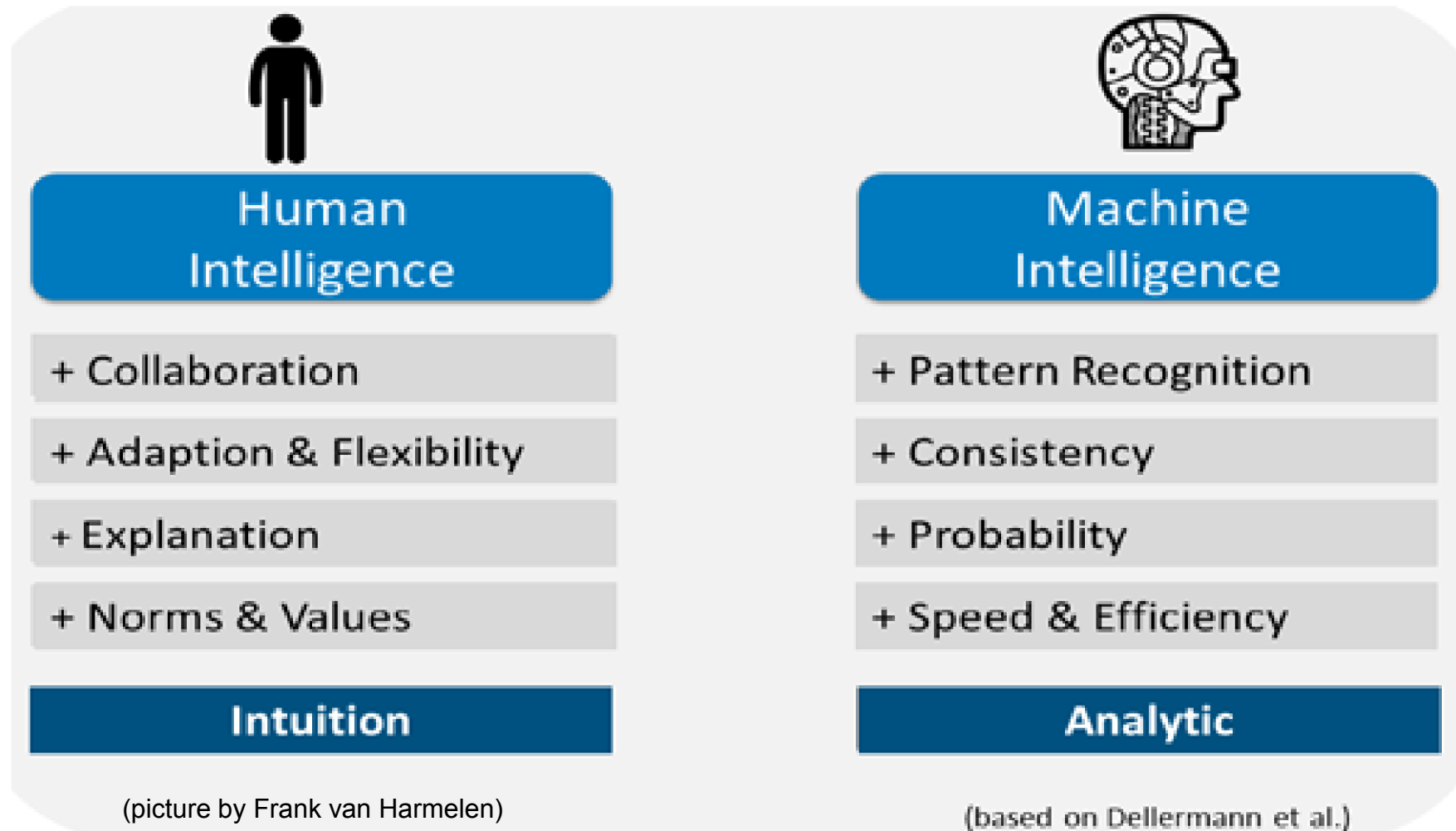


RESPONSIBLE DATA SCIENCE



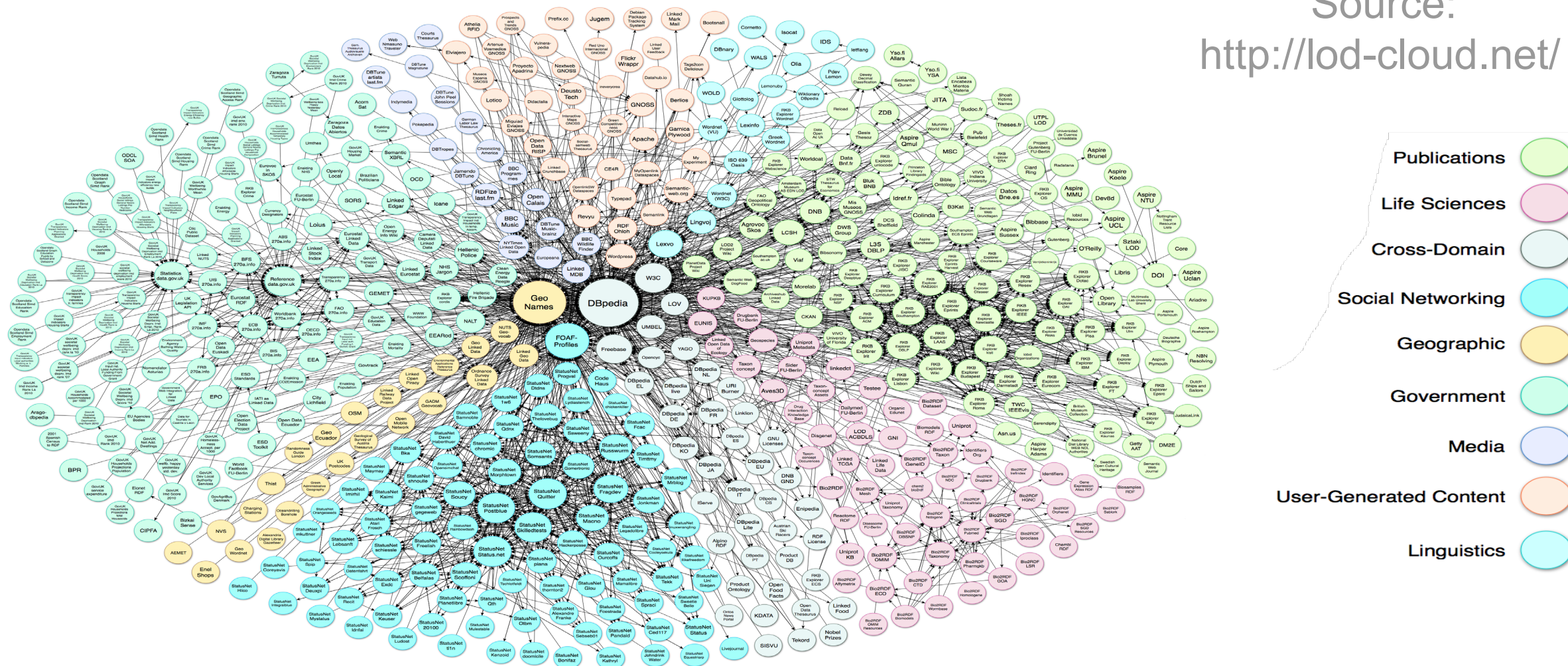
HYBRID INTELLIGENCE – HUMANS IN THE LOOP

Viewing AI systems not as independent “thinking machines,” but as cognitive prostheses that can help humans think better (Guszcza, 2018).



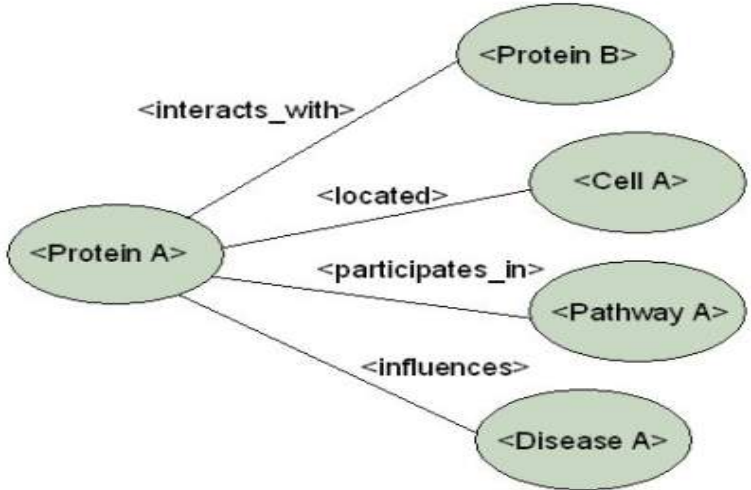
LINKED OPEN DATA

Source:
<http://lod-cloud.net/>

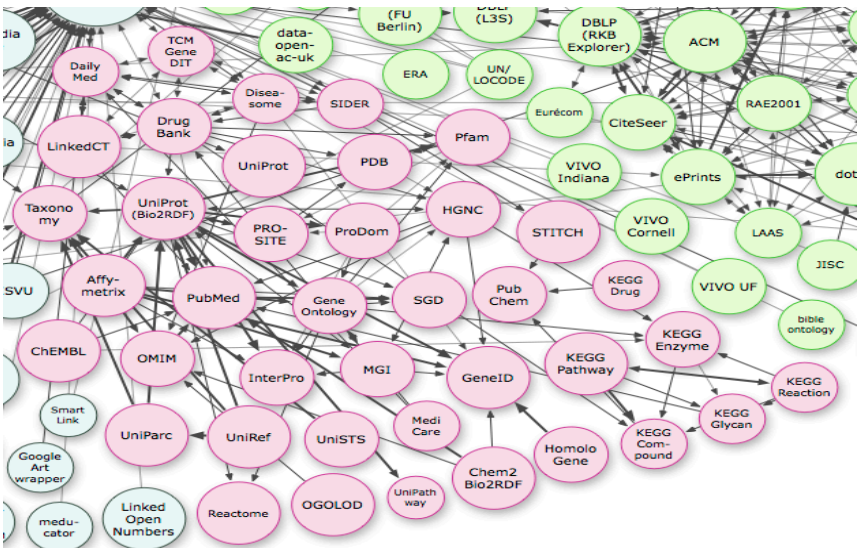
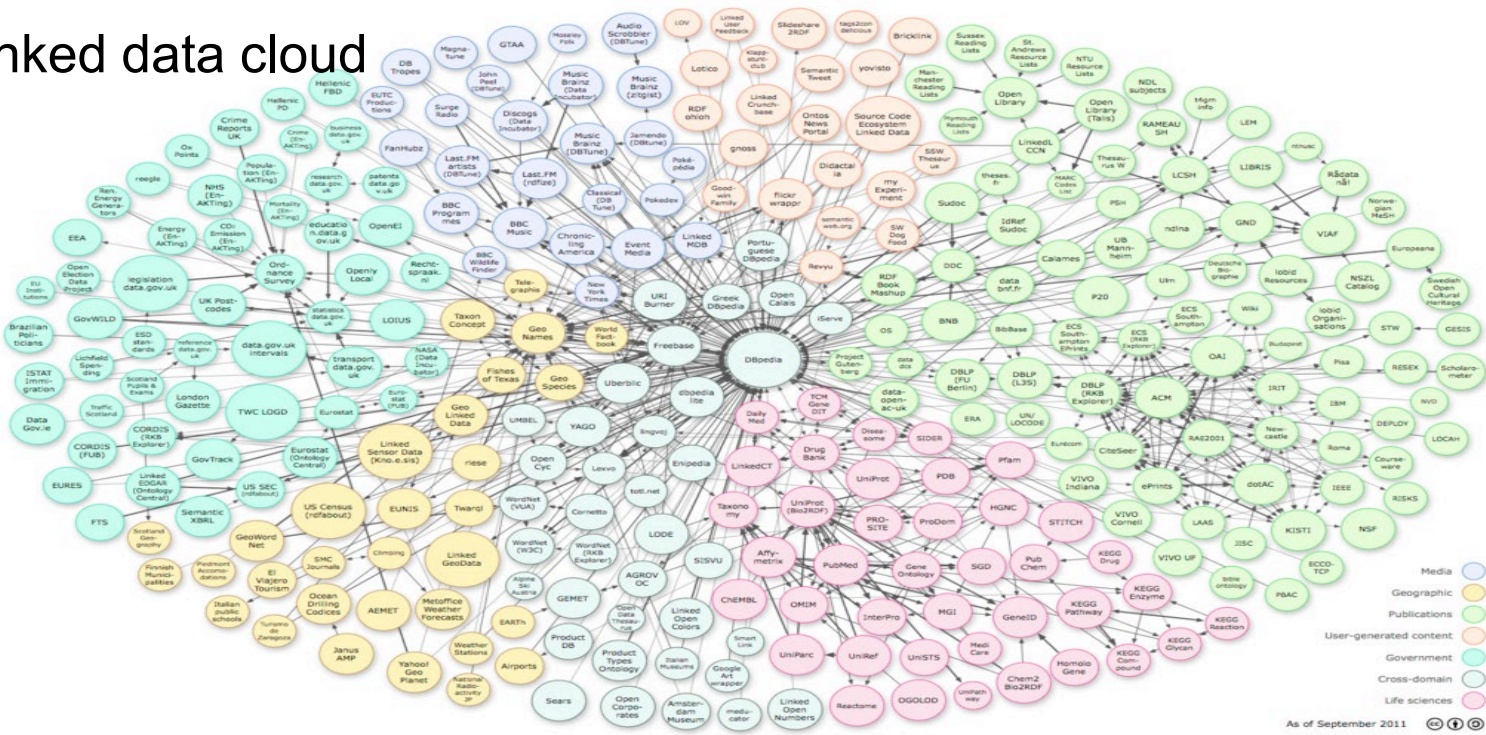


SEMANTIC INTEROPERABILITY THROUGH RDF (RESOURCE DESCRIPTION FRAMEWORK)

Unit of information: the data triple

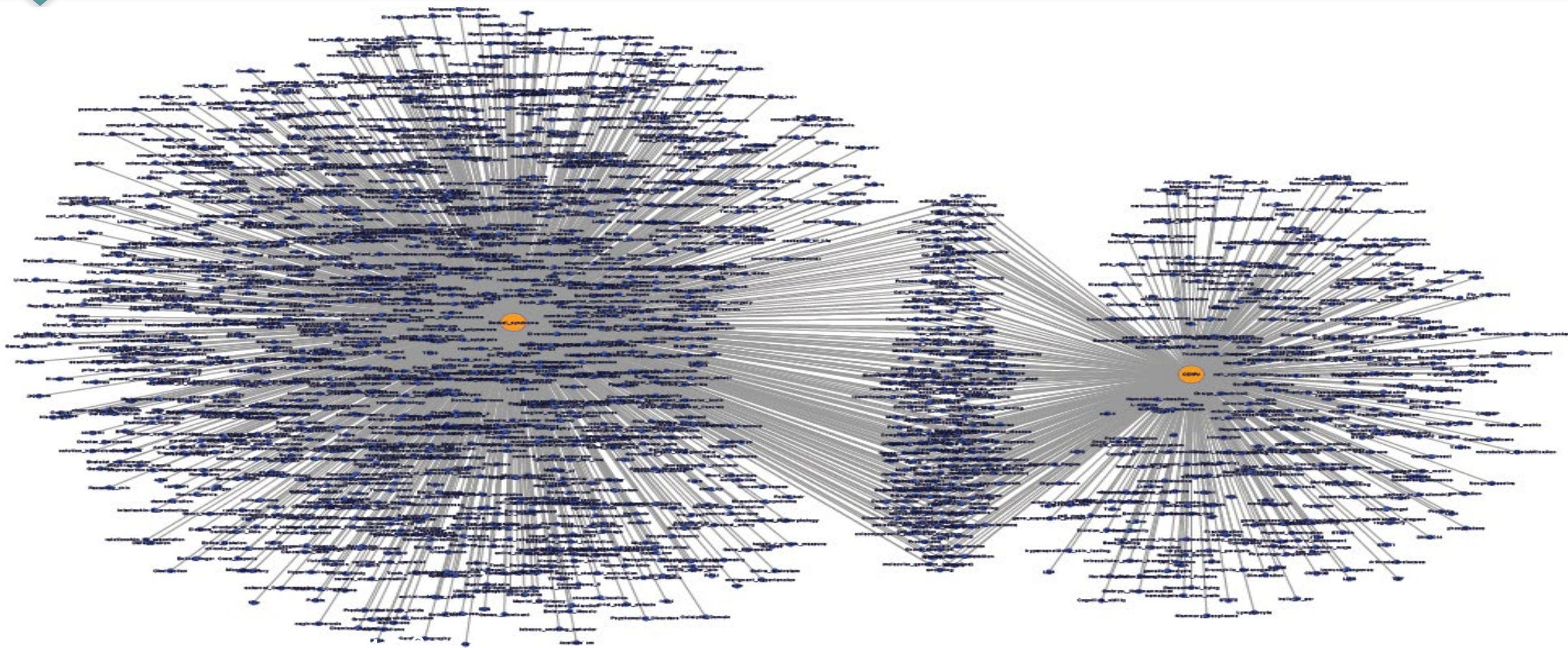


Linked data cloud



Data interoperability

TRAFFICKING THE DATA HIGHWAY: THE POWER OF INTEROPERABILITY



EURETOS **BRAIN**

<http://www.euretos.com/>

DTL | 
DUTCH TECHCENTRE FOR LIFE SCIENCES

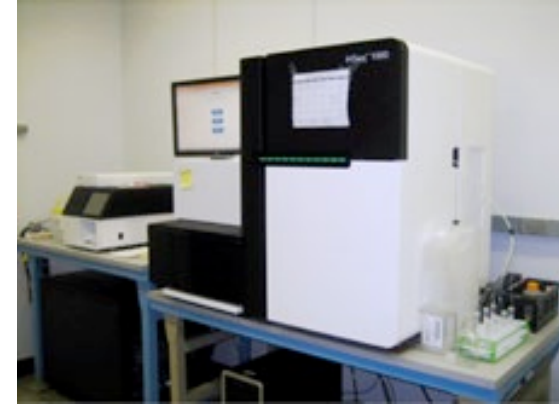
WHERE ARE WE HEADING WITH BIG DATA SCIENCE

- Translational medicine
 - “From bench to bedside”
 - Personalised/precision medicine
 - Mental problems and weather data
- Metagenomics
 - gut/oral bacteria
 - soil/water ecogenomics
- Bio-based economy

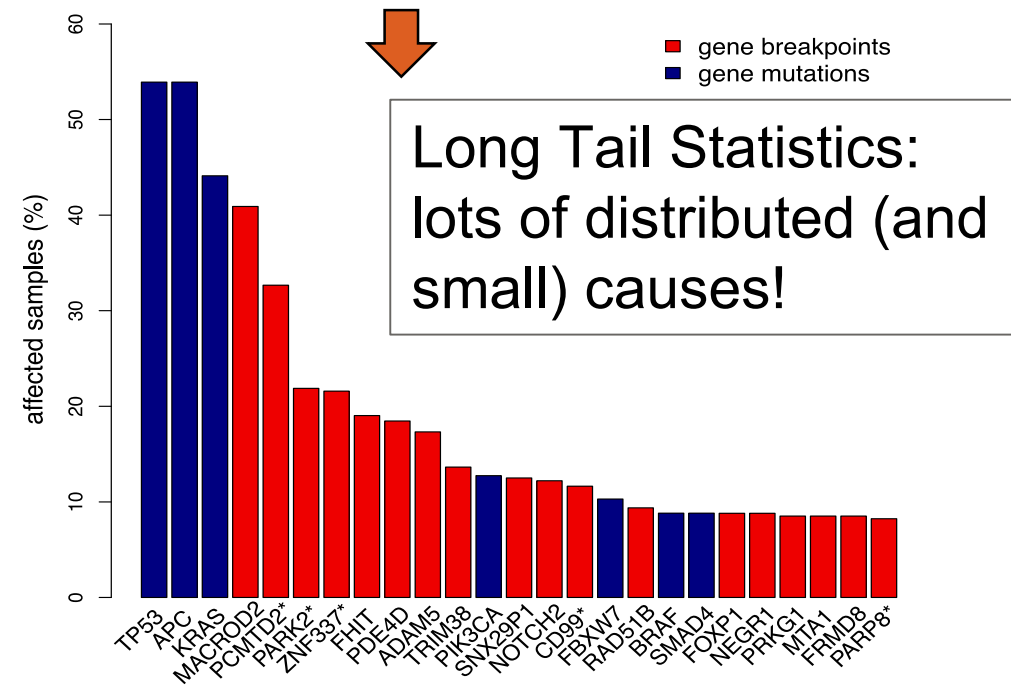
Ivacaftor (trade name **Kalydeco**) by Vertex Pharmaceuticals (for CF)

Trastuzumab (trade name **Herceptin**) by Genentech/UCLA (for breast cancer)

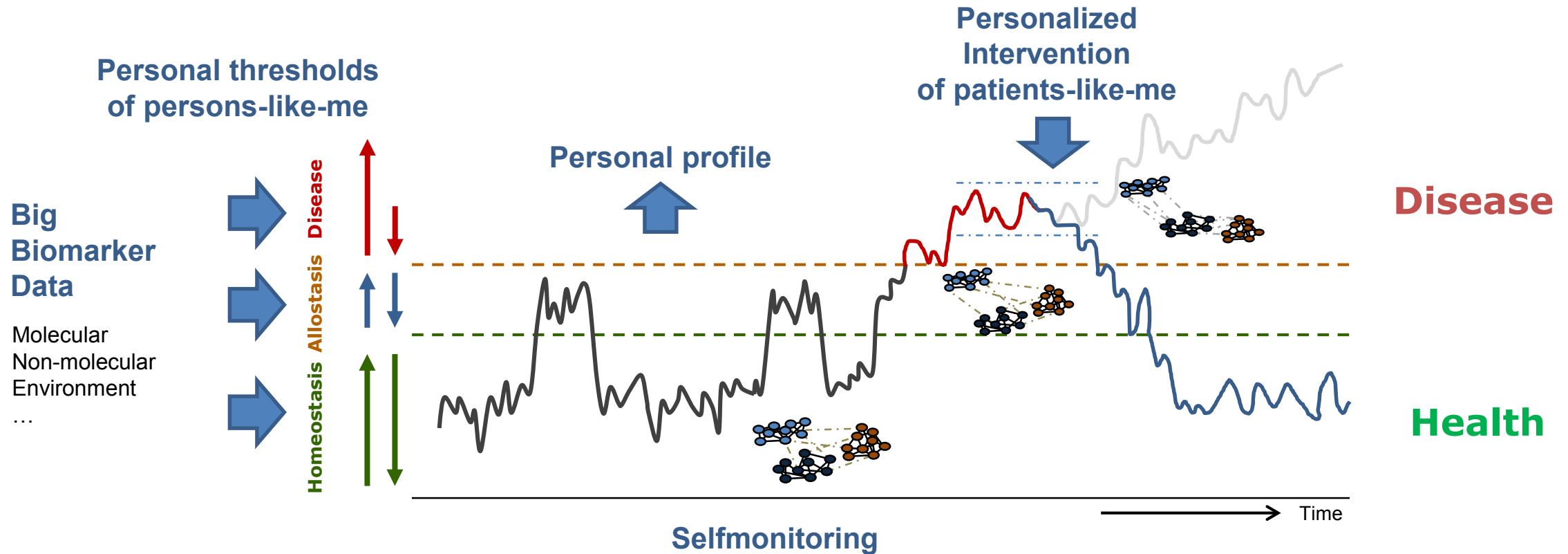
So far: Targeted drug development based upon intimate knowledge of mechanisms (e.g. protein structure) rather than unbiased patient sequencing and GWAS



experimental data
(1G -1TB per sample)



PERSONALIZED HEALTH(CARE) MODEL



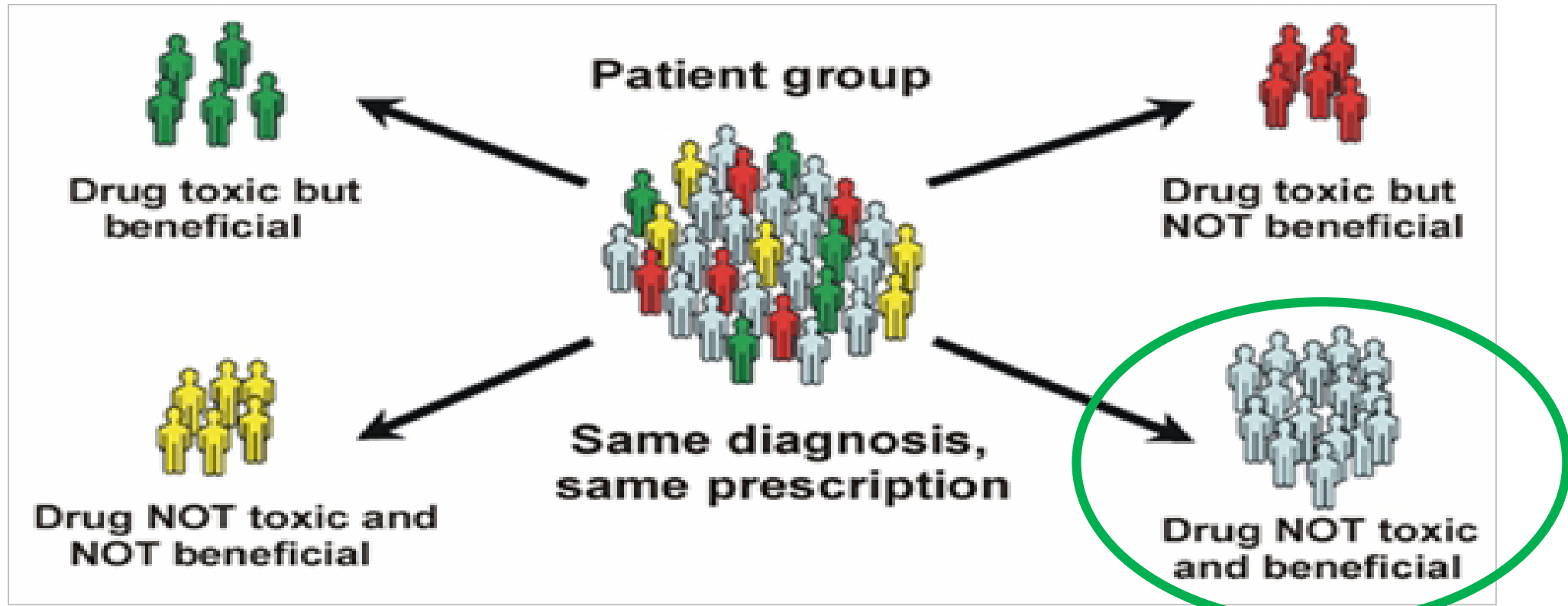
Personalized health

Personalized medicine

Adapted from Jan van der Greef, TNO

{See eg Chen ... Snyder, Cell 2012, 148: 1293}

PRINCIPLE OF PERSONALIZED MEDICINE

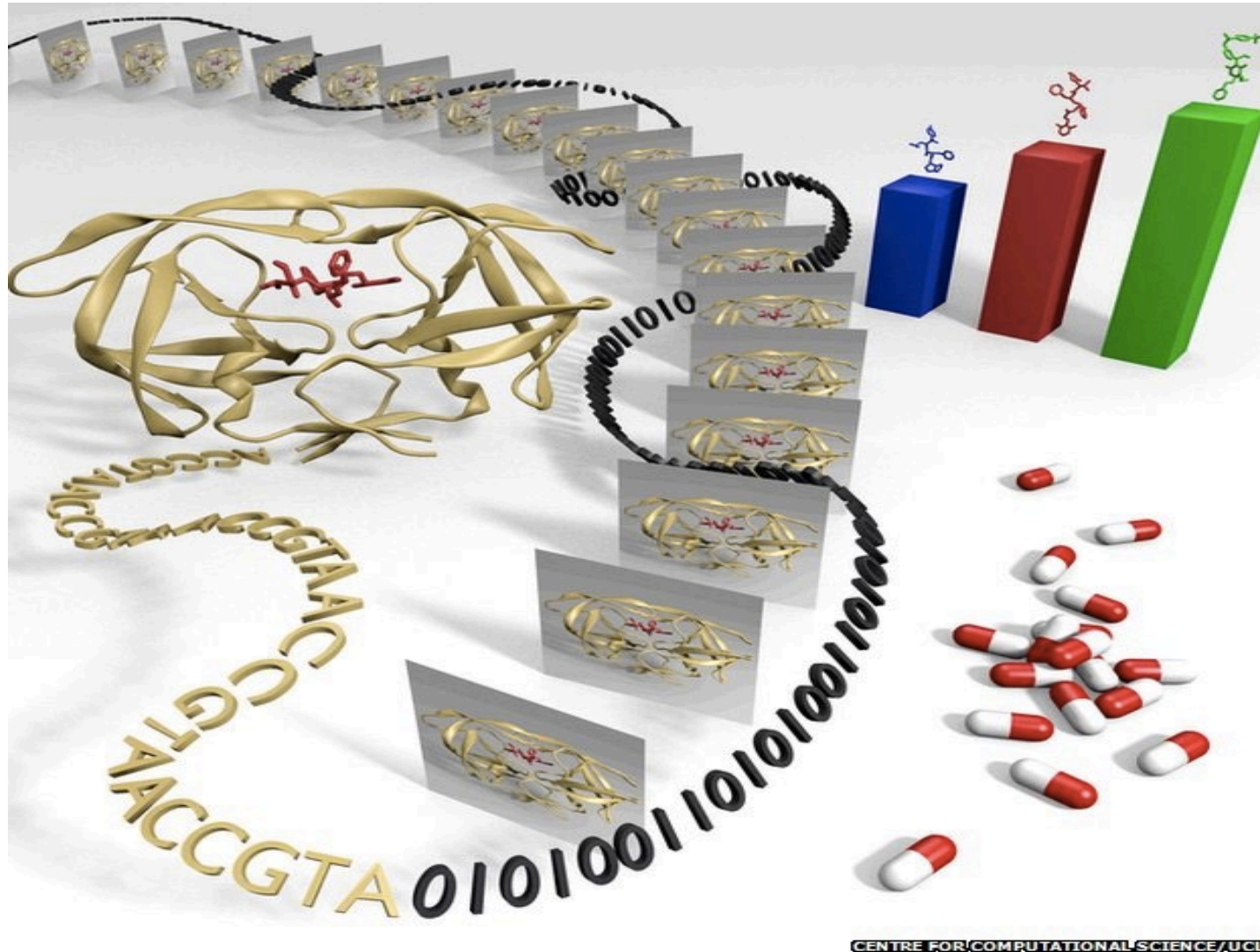


Source: Chakma, *Journal of Young Investigators*, 16, 2009

- The right drug for right patient at right dose at right time
- Molecular biomarkers as key drivers of patient selection
- = Precision medicine or Targeted medicine



HIV PERSONALISED/PRECISION MEDICINE

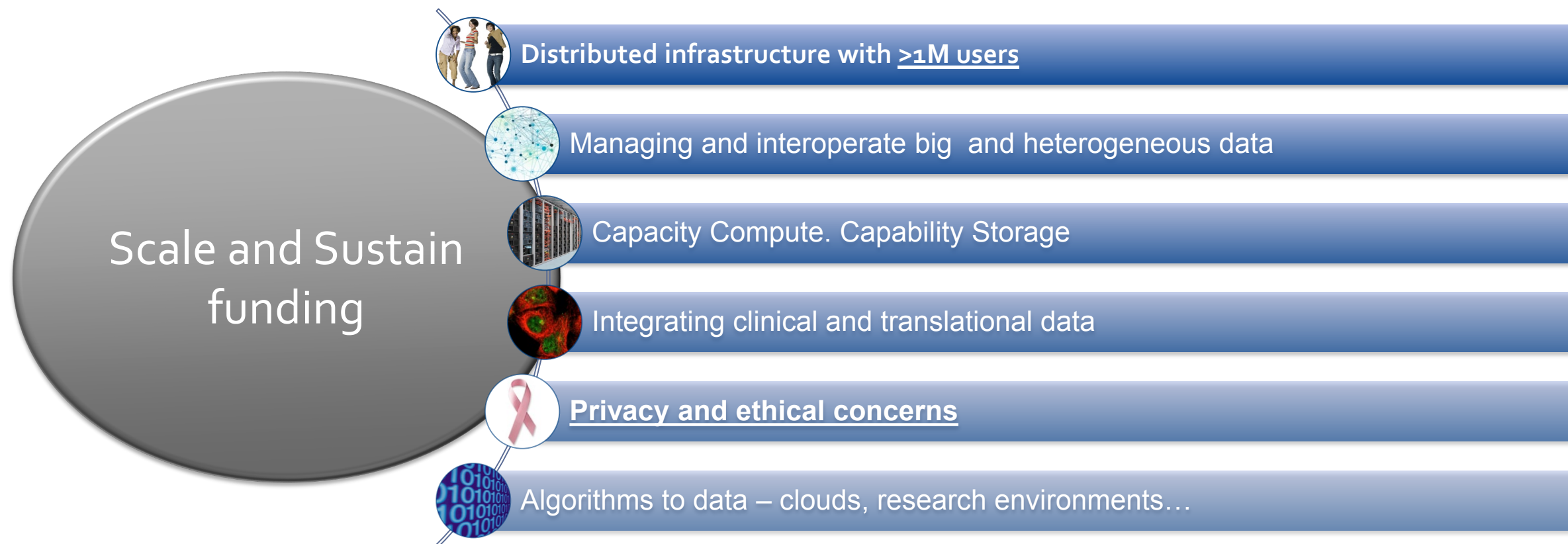


- Sequencing
 - Get HIV protease sequence
- Personal protease model building
- Docking across drugs library
- Prioritise

(Coveney and co-workers, UCL)



Challenges for life-science data services



ELIXIR Members



Belgium



Czech Republic



Denmark



EMBL



Estonia



Finland



France



Germany



Greece



Hungary



Ireland



Italy



Luxembourg



Israel



Netherlands



Norway



Slovenia



Portugal



Spain



Sweden



Switzerland



United Kingdom

ELIXIR Observers



Cyprus



ELIXIR Members



Belgium



Czech Republic



Denmark



EMBL



Estonia



Greece



Luxembourg



Slovenia



Switzerland



United Kingdom

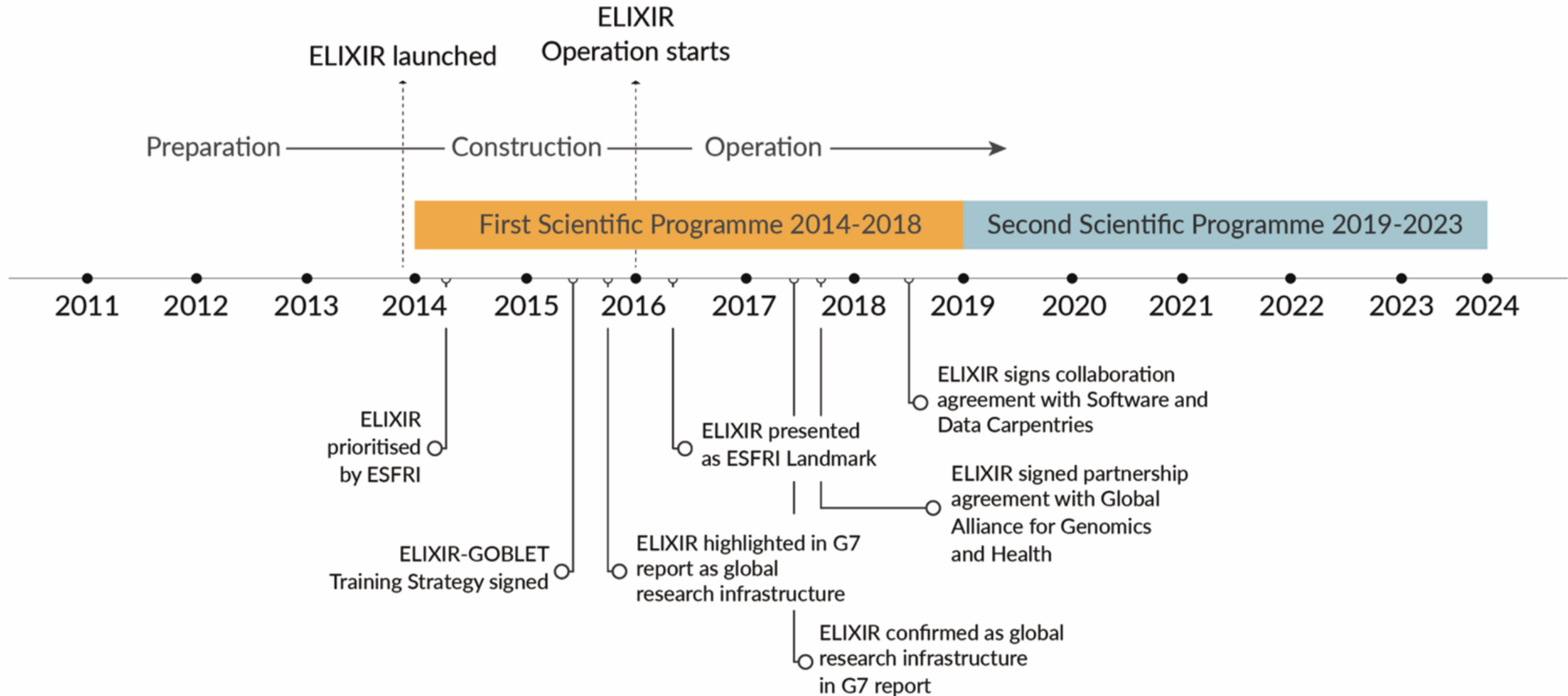
ELIXIR Observers



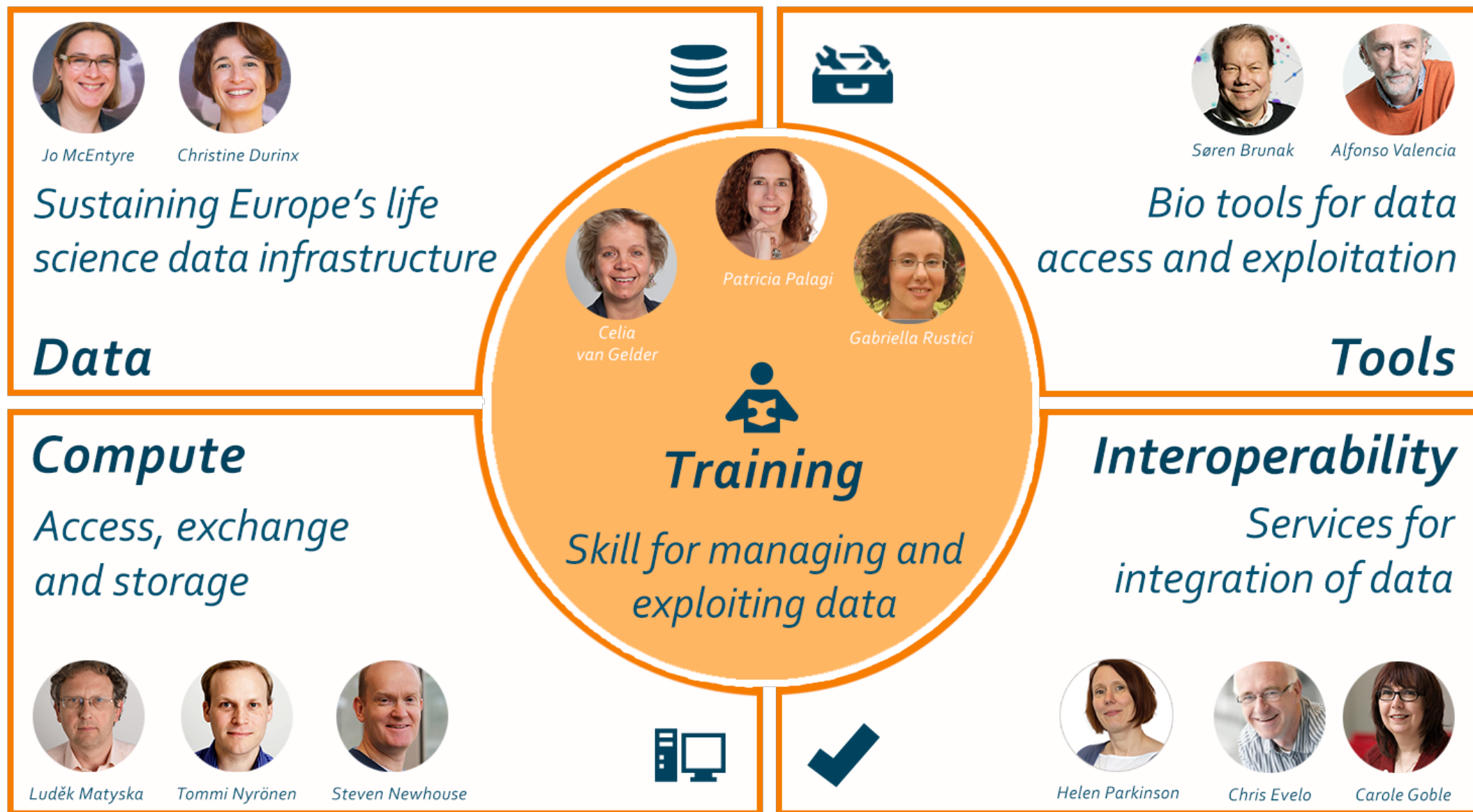
Keyword: co-creation
Started with 5 nodes, now 23 nodes
(22 nations + EBI)
Hub (located at Wellcome Genome
Campus) and spokes model



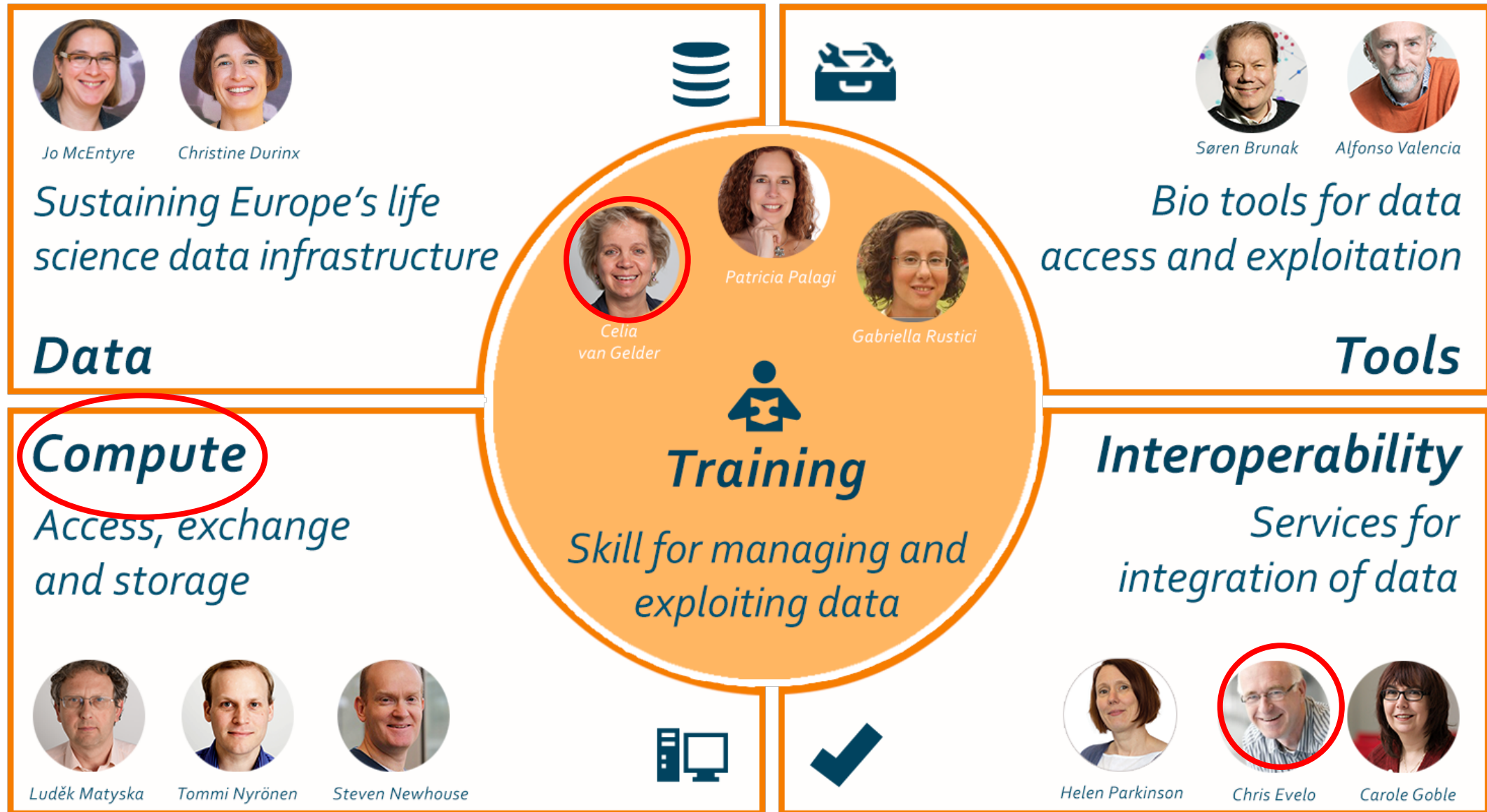
ELIXIR's Timeline



ELIXIR : 5 platforms of shared services led by leading European scientists



ELIXIR : 5 platforms of shared services led by leading European scientists



ELIXIR Communities – Current Status - Dutch Node

Community	Status	Community	Status
Human Data ✓	Approved	Food & Nutrition	Awaiting Comm. Mtg.
Rare Diseases ✓	Approved	Human Cell Atlas	Status under discussion
Plant Science ✓	Approved	Human CNV ✓	Meeting 28 Sep 18
Marine Metagenomics ✓	Approved	Intrinsically Disordered Proteins	Meeting 31 Oct 18
Galaxy ✓	Approved	Microbial Biotechnology ✓	Meeting 8 Sep 18 Expecting white paper by end October
Metabolomics ✓	Approved	Structural Bioinformatics	Awaiting Full Application Pre-meeting 19 Oct 18
Proteomics ✓	Approved	Toxicology* ✓	Awaiting Comm. Mtg.

* Asked to confirm Node involvement of membership by HoNs

✓ - This community has engagement from the Dutch Node

ELIXIR Core Data Resources



Rachel Drysdale



Jo McEntyre



Christine Durinx

ELIXIR Core Data Resource list

Core Data Resource	Data type
ArrayExpress	Functional Genomics Data from high-throughput functional genomics experiments.
CATH	A hierarchical domain classification of protein structures in the Protein Data Bank.
ChEBI	Dictionary of molecular entities focused on 'small' chemical compounds.
ChEMBL	Database of bioactive drug-like small molecules, it contains 2-D structures, calculated properties and abstracted bioactivities.
EGA	Personally identifiable genetic and phenotypic data resulting from biomedical research projects.
ENA	Nucleotide sequencing information, covering raw sequencing data, sequence assembly information and functional annotation.
Ensembl	Genome browser for vertebrate genomes that supports research in comparative genomics, evolution, sequence variation and transcriptional regulation.
Ensembl Genomes	Comparative analysis, data mining and visualisation for the genomes of non-vertebrate species.
Europe PMC	Europe PMC is a repository, providing access to worldwide life sciences articles, books, patents and clinical guidelines.
Human Protein Atlas	The Human Protein Atlas contains information for a large majority of all human protein-coding genes regarding the expression and localization of the corresponding proteins based on both RNA and protein data.
The IMEx Consortium: represented by IntAct and MINT	IntAct provides a freely available, open source database system and analysis of molecular interaction data. MINT focuses on experimental interactions mined from the scientific literature.
InterPro	Functional analysis of protein sequences, predicting the presence of

Note: This

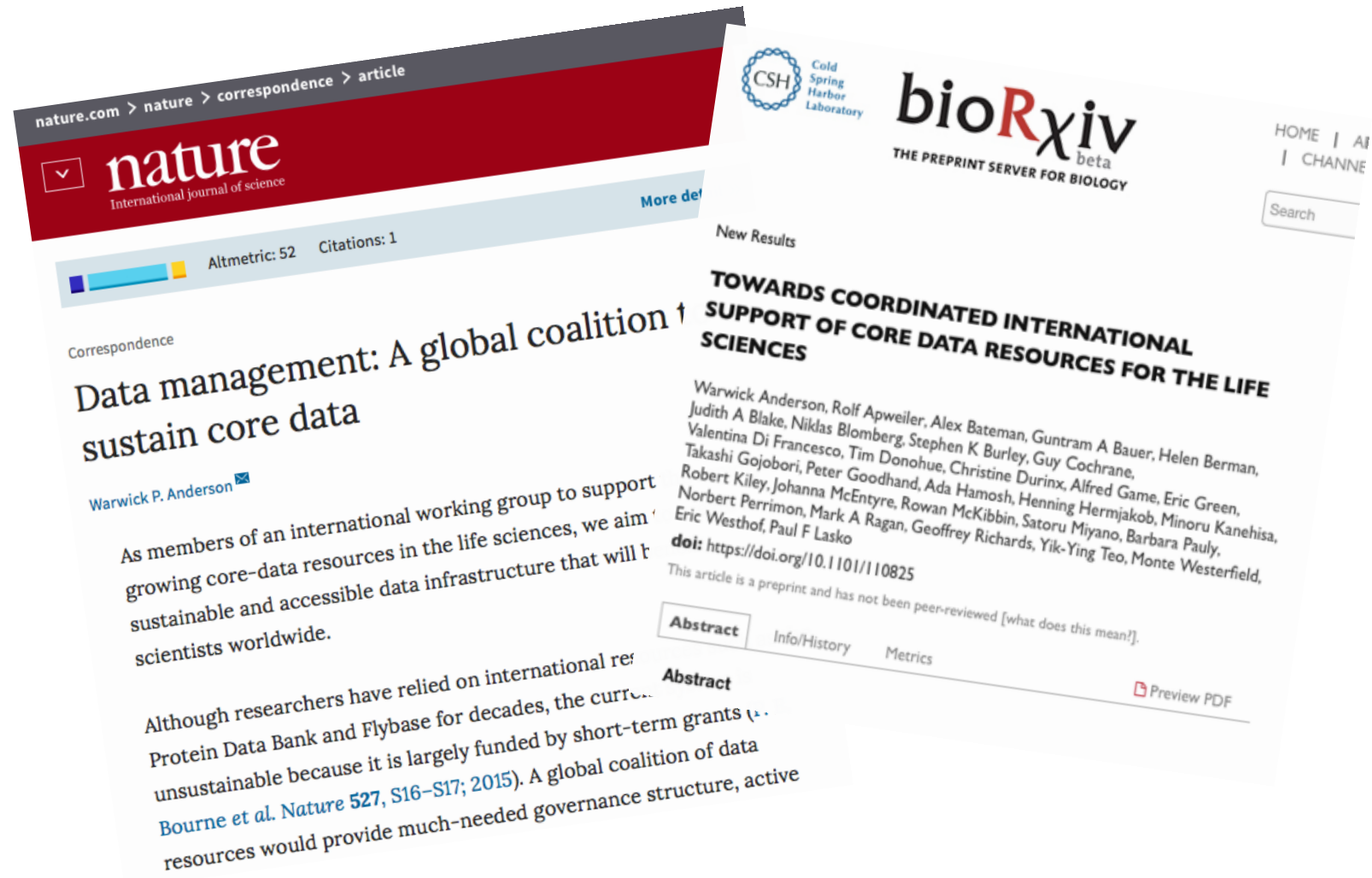
- Set of European data resources of fundamental importance to the wider life-science community and the long-term preservation of biological data.
- Agreed collectively by 21 Heads of Nodes
- International collaborative effort

<https://elixir-europe.org/platforms/data/core-data-resources>

Durinx C, McEntyre J, Appel R *et al.* Identifying ELIXIR Core Data Resources *F1000Research* 2016, **5**(ELIXIR):2422

Towards a Global coalition to sustain Core Data Resources

- Call for Action published in Nature in March 2017
- Full text of article available as pre-print in bioRxiv
- June workshop in London with international funders
- Great interest in Core Data Resources (outcome and method)
- Outcomes taken into HIRO meeting following day
- Working Group established to take forward next steps



Core Data Resources

Document with five main criteria compiled by Christine Durinx (SIB) and Jo McIntyre (EBI)

- F1000 ELIXIR track article 'Identifying ELIXIR Core Data Resources'.
1. Scientific focus and quality of science
 2. Community served by the resource
 3. Quality of service
 4. Legal and funding infrastructure, and governance
 5. Impact and translational stories

Commissioned Service

ELIXIR Authorisation and Authentication Infrastructure

Enables life science researchers to use their institutional IDs to

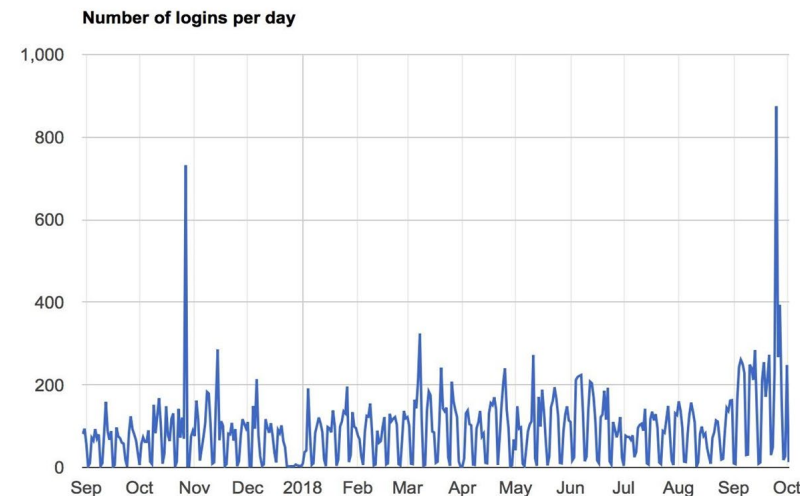
STATS

Production services: 67

Services in test environment: 65

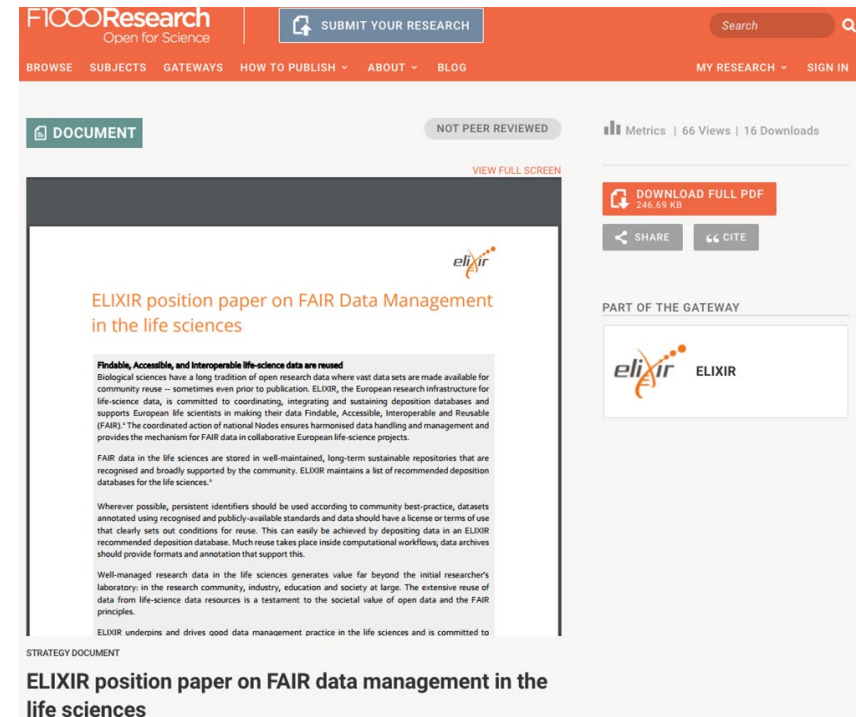
- Reduced bureaucracy
- Improved vetting: confidence to the
- Regular updates: their affiliation inf
- Improved access to accounts allows se use of their service
- Applicable to oth

Number of logins



ELIXIR Position Paper on FAIR data management in the life sciences

1. Open sharing of research data is a core principle
2. Data Management is crucial to science
3. Data should be submitted to deposition databases
4. All data submitted to Open Data archives should align with community-defined standards
5. ELIXIR Nodes implement FAIR for their respective nations
6. Professional skills, adequate resources and appropriate funding are needed for Data Management and infrastructure



STRATEGY DOCUMENT

ELIXIR position paper on FAIR data management in the life sciences

Blomberg N and ELIXIR Consortium. ELIXIR position paper on FAIR data management in the life sciences. **F1000Research 2017**, 6(ELIXIR):1857 (document) (doi: [10.7490/f1000research.1114985.1](https://doi.org/10.7490/f1000research.1114985.1))



-> **Note:** ELIXIR NL have played a key role in developing FAIR through DTL (Barend Mons)



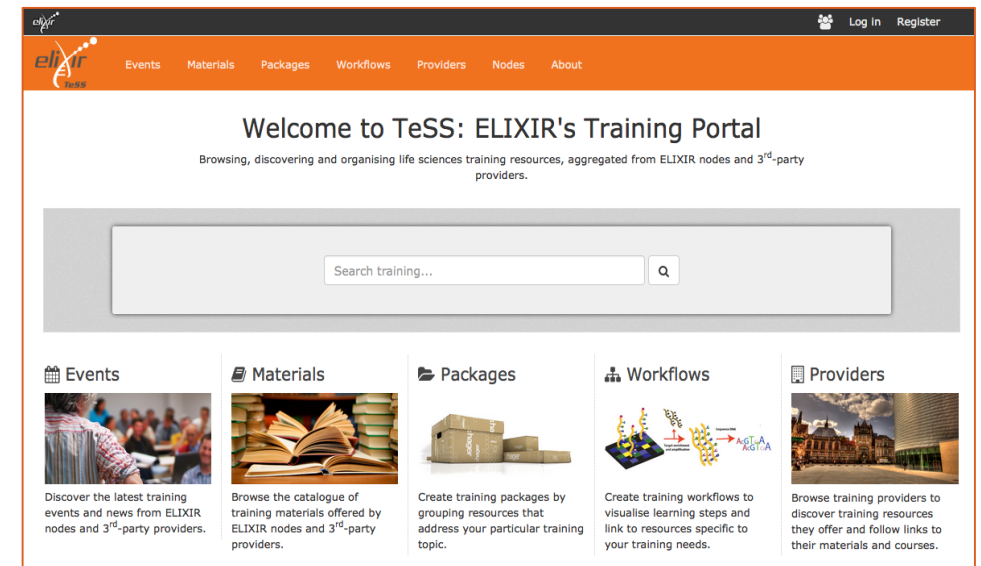
ELIXIR Training Portal TeSS

Aggregate, disseminate and coordinate information on ELIXIR's training activities/material and beyond



- 1023 training materials
- 48 content providers: 21 scraped automatically, 13 ELIXIR Nodes
- 4,934 unique users (2017-18)
- 308 upcoming events listed
- AAI and  Bioschemas integration
- Linked to other ELIXIR registries: bio.tools and  FAIRsharing
- Events annotated with EDAM topics terms

• <https://tess.elixir-europe.org>



ELIXIR Implementation Studies with a NL involvement

Title	Platform/Use Case
Using clouds and VMs for bioinformatics training (Workshop as a Service)	Training, Compute
Developing learning paths for users of ELIXIR services	Training
Towards Data Stewardship in ELIXIR: Training & Portal	Training
FAIRness of the current ELIXIR Core resources: Application (and test) of newly available FAIR metrics, and identification of steps to increase interoperability	Data
Bioschemas: Community Adoption and Training	Interoperability
Metabolite Identification	Metabolomics
Crowd-sourcing the annotation of public proteomics datasets to improve data reusability	Proteomics
Beacon & Beacon Network As A Service	Human Data
A Scalable approach to Personal FAIR Data Management and Analysis	Human Data
Towards Data Stewardship in ELIXIR: Training and Portal	Training

ELIXIR strategic objectives 2019-23



ELIXIR operates a portfolio of integrated services that meets the data needs of life scientists at a European scale



ELIXIR Communities drive service uptake, support standards development and connect ELIXIR's experts in life-science disciplines



ELIXIR Core Data Resources are the global standard for bioinformatics resource management and the foundation for an international funding and life cycle management strategy that secures the long-term sustainability of those resources.



ELIXIR is the recognised and trusted life science foundation of the European Open Science Cloud



All ELIXIR Nodes function as effective national centres of excellence for life-science data management and provide valued support to local users and industry.

Major Instruments in 2019-23 Programme to deliver objectives

ELIXIR Platforms	ELIXIR Infrastructure Services	Strategic Implementation studies	Community-led implementation studies	Mature Communities Support
<ul style="list-style-type: none">• Align Services across Nodes• Connect expertise in Nodes to Communities• Establish Service Frameworks	<ul style="list-style-type: none">• Commitment from ELIXIR to the operation of services• Foundation for our architecture	<ul style="list-style-type: none">• Targeted ELIXIR developments in key areas• Expect that they will lead to Infrastructure Services	<ul style="list-style-type: none">• Science led challenges with objective to create “service bundles” for community• Competitive RFPs; Expect at least 1 Community / 2 Platforms in the proposal• Fund the use and service adaptation to community need	<ul style="list-style-type: none">• Initial roadmap implementation study• Annual Workshop• Build towards grant submissions – ELIXIR is infrastructure provider



EXTENDING ELIXIR-NL REMIT

ELIXIR's activities are divided into five areas called 'Platforms'. These are Data, Tools, Interoperability, Compute and Training. The Platforms are managed by Platform leaders and the work is carried out by groups within the Platforms.



Tools Platform

Helps researchers find the best software tools to analyse their data.



Data Platform

Aims to identify key data resources across Europe and support the linkages between data and literature



Compute Platform

Develops services to make it easier to store, share and analyse large datasets.



Interoperability Platform

Develops and encourages the adoption of standards to describe life science data.



Training Platform

Helps scientists and developers find the training they need, and also provides that training.

EXTENDING ELIXIR-NL REMIT

ELIXIR's activities are divided into five areas called 'Platforms'. These are Data, Tools, Interoperability, Compute and Training. The Platforms are managed by Platform leaders and the work is carried out by groups within the Platforms.



Data Platform

Aims to identify key data resources across Europe and support the linkages between data and literature



Tools Platform

Helps researchers find the best software tools to analyse their data.



Compute Platform

Develops services to make it easier to store, share and analyse large datasets.



Interoperability Platform

Develops and encourages the adoption of standards to describe life science data.



Training Platform

Helps scientists and developers find the training they need and also provides that training.

ELIXIR Human Data

Resources and projects in human genomics and translational data

EXTENDING ELIXIR-NL REMIT

Potential NL core data
resources?

Potential NL core tools /
pipelines?



Int

Dev
of sta
data.

resources and projects in human
genomics and translational data
training Platform
With scientists and developers find
the training they need and also
provides that training.

ELIXIR-NL - DATA STEWARDSHIP WIZZARD



https://www.dtls.nl/2017/11/11/data-stewardship-wizard-enhance-quality-data-management-plan/



[Home](#) » [News](#) » A Data Stewardship Wizard to enhance the quality of your Data Management Plan

A Data Stewardship Wizard to enhance the quality of your Data Management Plan

Posted on 11 November 2017

DTL's Rob Hooft has developed a Data Stewardship Wizard in collaboration with colleagues from the Czech ELIXIR node. The Data Stewardship Wizard converts a lengthy data management questionnaire into an effective flowchart, saving you research time and money, and enhancing the quality of your Data Management Plan.

[Use the Data Stewardship Wizard.](#)

A modern complete research proposal includes a Data Management Plan (DMP). However, regular DMP questionnaires consist of a series of hundreds of open-ended questions. Answering these requires a combination of expertise from different people. As a result, many researchers perceive writing a DMP as an unpleasant obligation and a time claimer at the cost of research time. In addition, regular DMP questionnaires necessarily have a restricted scope because including all potential topics would create even longer lists of questions that may largely be irrelevant to many projects.

Data Stewardship Wizard

DTL's Rob Hooft has developed a Data Stewardship Wizard in collaboration with colleagues from the Czech ELIXIR node. He explains: "The Wizard uses a logical flowchart to guide you through a questionnaire that incorporates the full breath of relevant topics, producing answers that are relevant to your DMP as well as realistic research budgets. The Wizard offers an efficient way of working, with lower risk of forgetting essential components of data management, thereby reducing risks for the project and leaving more effective time for the research itself."

Why make a Data Management Plan

Save time & money

Hooft: "Nowadays, handling digital data consumes around 25% of the total budget of an average life science research project, so proper data management planning from the start is worthwhile."



Dr Rob Hooft

FAIR Data approach



A key enabler to achieve international-grade data stewardship is for research data and information to be published in a 'FAIR' manner. Data should be: Findable, Accessible, Interoperable and Reusable. DTL actively promotes FAIR Data Stewardship of life science information.

 [Read more](#)

ELIXIR-NL - DATA STEWARDSHIP WIZZARD

https://www.dtls.nl/2017/11/11/data-stewardship-wizard-enhance-quality-data-management-plan/



Home » News » A Data Stewardship Wizard to enhance the quality of your Data Management Plan

A Data Stewardship Wizard to enhance the quality of your Data Management Plan

FAIR Data approach

- Stepwise protocol guiding users to accomplish Data Management Planning and Data Stewardship
- Developed by ELIXIR-NL and ELIXIR-CZ



re international-grade data
arch data and information to be
anner. Data should be: Findable,
ole and Reusable. DTL actively

promotes FAIR Data Stewardship of life science
information.

[Read more](#)



Dr Rob Hooft

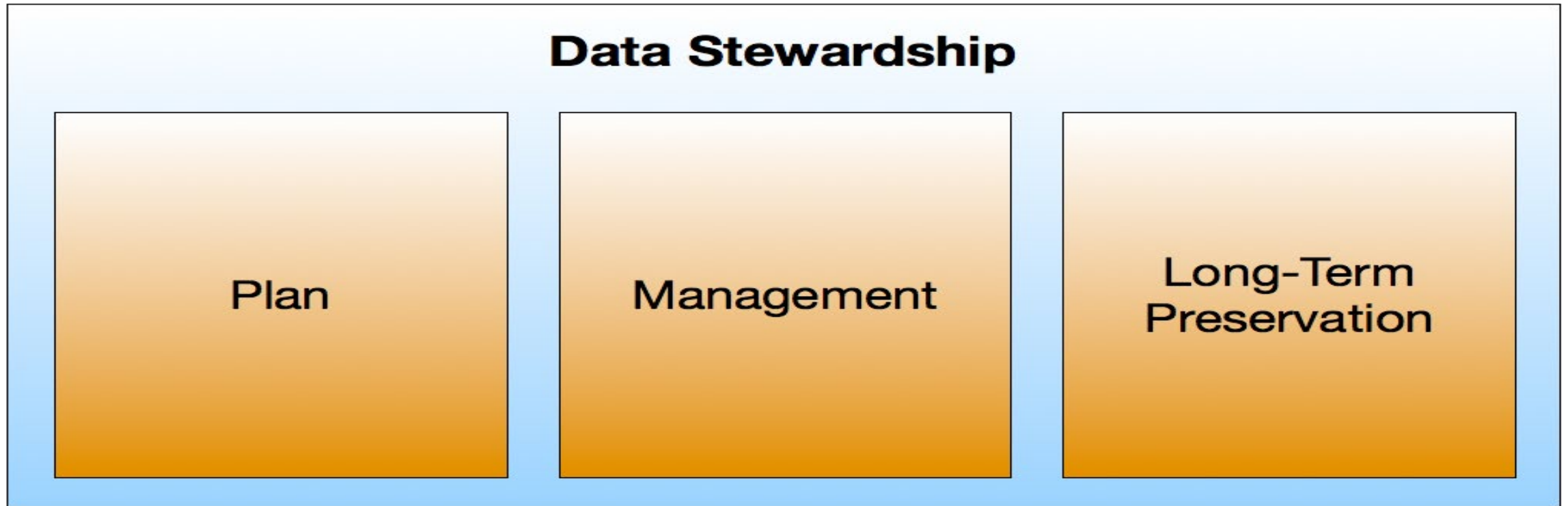
Data St
DTL's R
from the Czech ELIXIR node. He explains: "The wizard uses a logical flowchart to guide you through a questionnaire that incorporates the full breath of relevant topics, producing answers that are relevant to your DMP as well as realistic research budgets. The Wizard offers an efficient way of working, with lower risk of forgetting essential components of data management, thereby reducing risks for the project and leaving more effective time for the research itself."

Why make a Data Management Plan

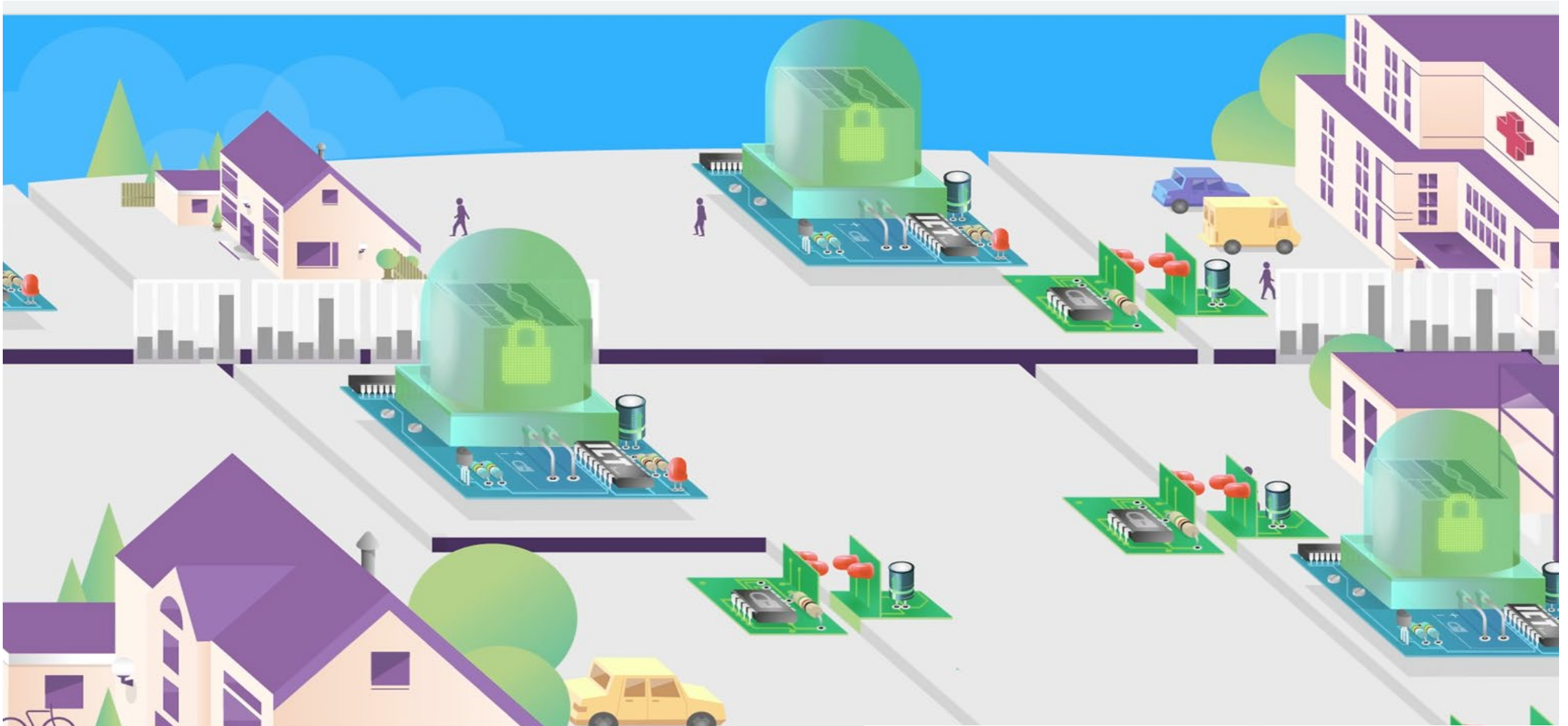
Save time & money

Hooft: "Nowadays, handling digital data consumes around 25% of the total budget of an average life science research project, so proper data management planning from the start is worthwhile."

Data Stewardship – Process View



DISTRIBUTED LEARNING - PERSONAL HEALTH TRAIN



DISTRIBUTED LEARNING - PERSONAL HEALTH TRAIN

Distributed data (scalability issues)
Privacy-restricted data
- Learning algorithms visit data sources,
learning iteratively
Health / Life Sciences (Farm Data Train)

FAIR DATA ECOSYSTEM – FAIRification software



Create

FAIRIFIER



Publish

FAIR
DATA POINT



Find

FAIR
DATA 



Annotate

ORKA Open RDF
Knowledge Annotator



DTL

Data FAIRport

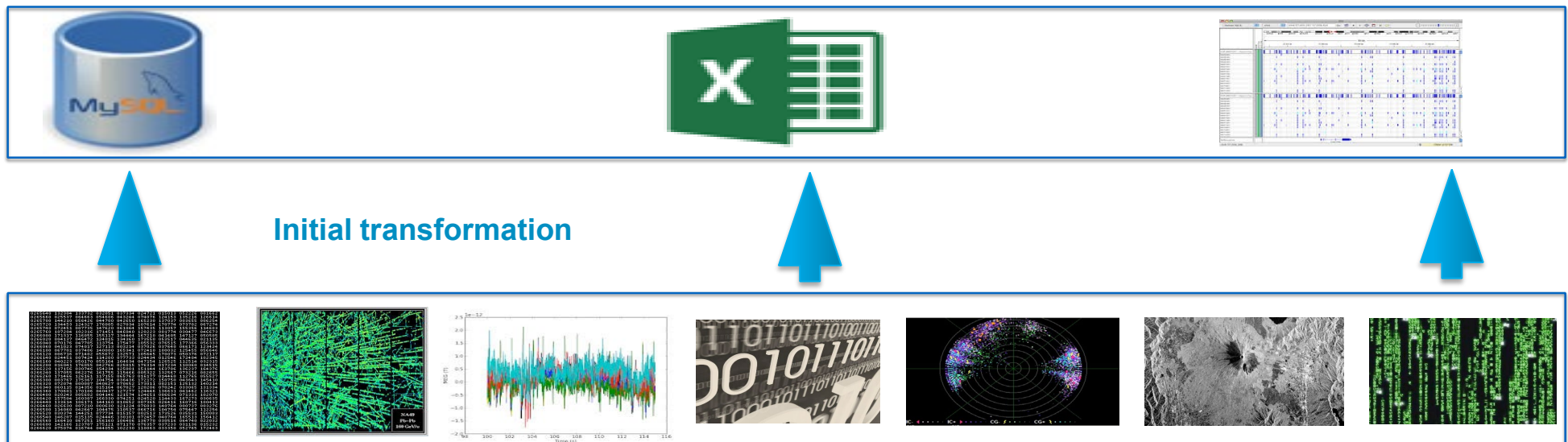
Raw data
(many formats)



Processed data
(primary storage format)



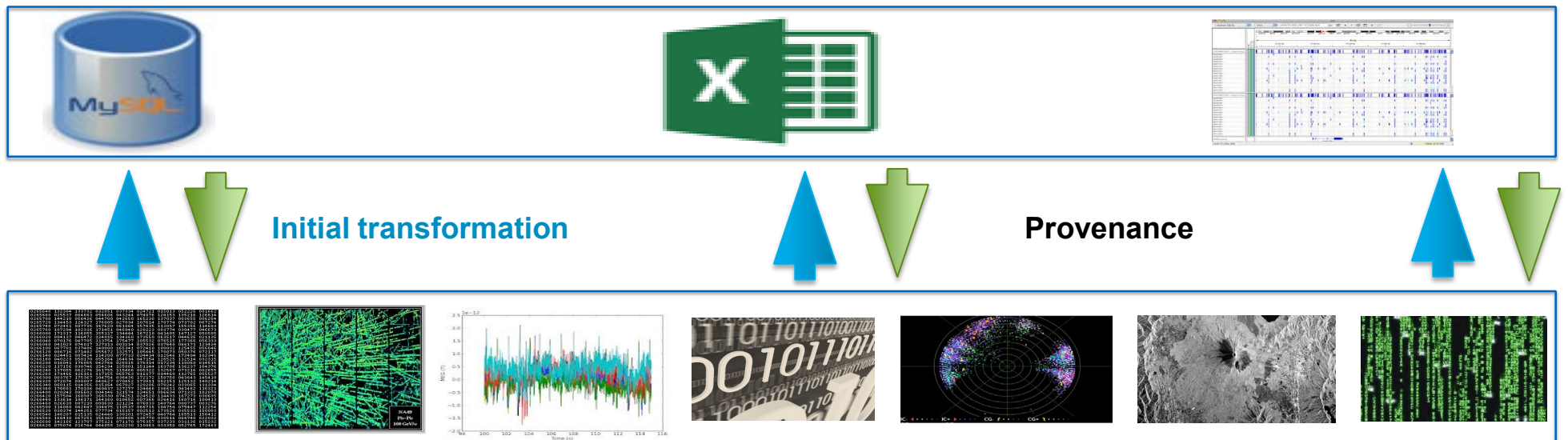
Raw data
(many formats)



Processed data
(primary storage format)



Raw data
(many formats)





Three interlocking gears, symbolizing the interconnectedness of the three pillars of the business model.



Three interlocking gears, symbolizing the interconnectedness of the three pillars of the business model.

The diagram illustrates the FAIR transformation process. It shows a flow from raw data (represented by a MySQL database icon) through an 'Initial transformation' (represented by a blue arrow) to a structured data format (represented by an Excel spreadsheet icon). This structured data then undergoes a 'Provenance' step (represented by a green arrow) to become a network graph (represented by a complex graph icon). The final output is a 'FAIR transformation' (represented by a blue arrow) leading to a highly structured, interconnected network graph (represented by a complex graph icon).

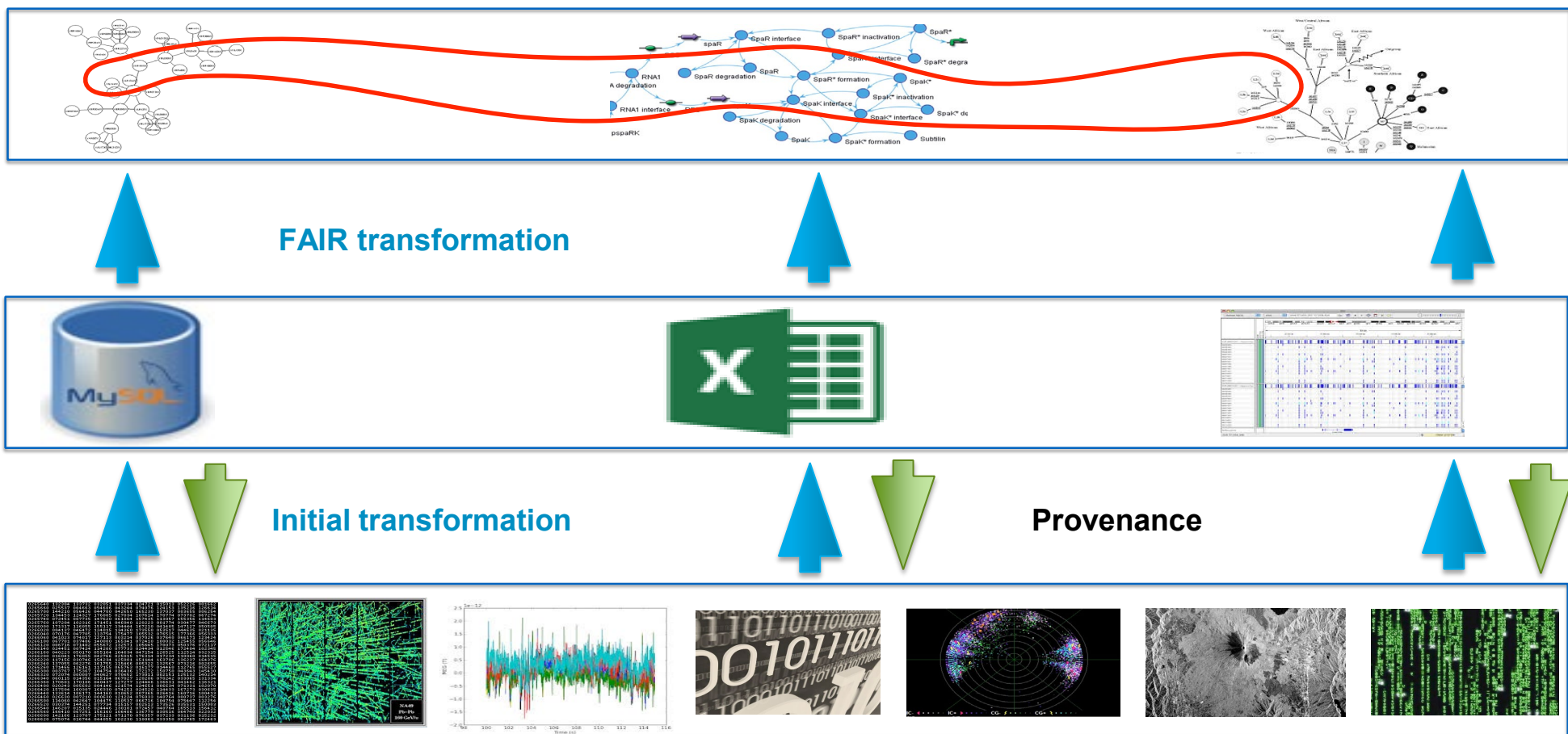
FAIR (meta)data
(RDF,XML etc.)



Processed data
(primary storage format)



Raw data
(many formats)



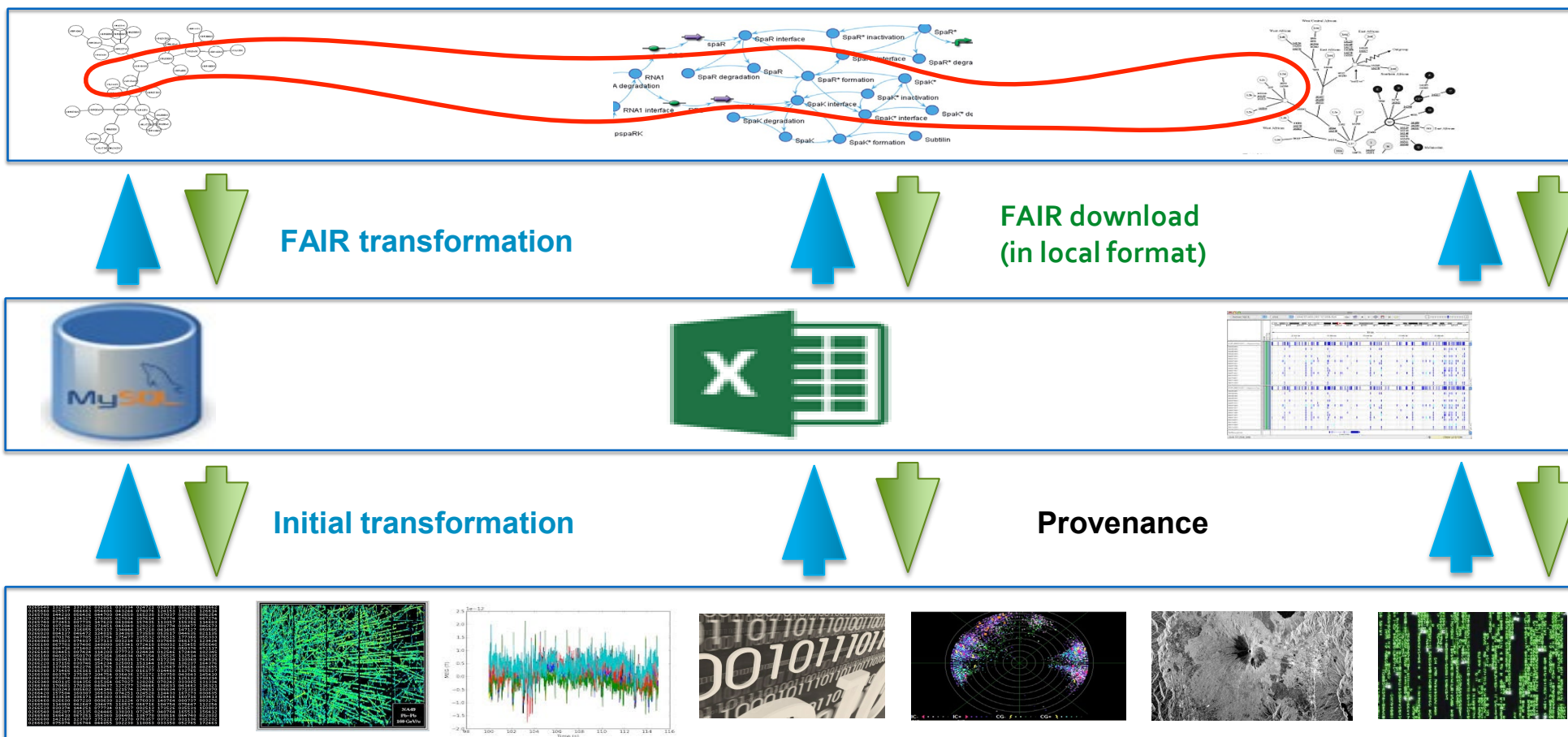
FAIR (meta)data
(RDF,XML etc.)



Processed data
(primary storage format)



Raw data
(many formats)



High-Performance
Analysis



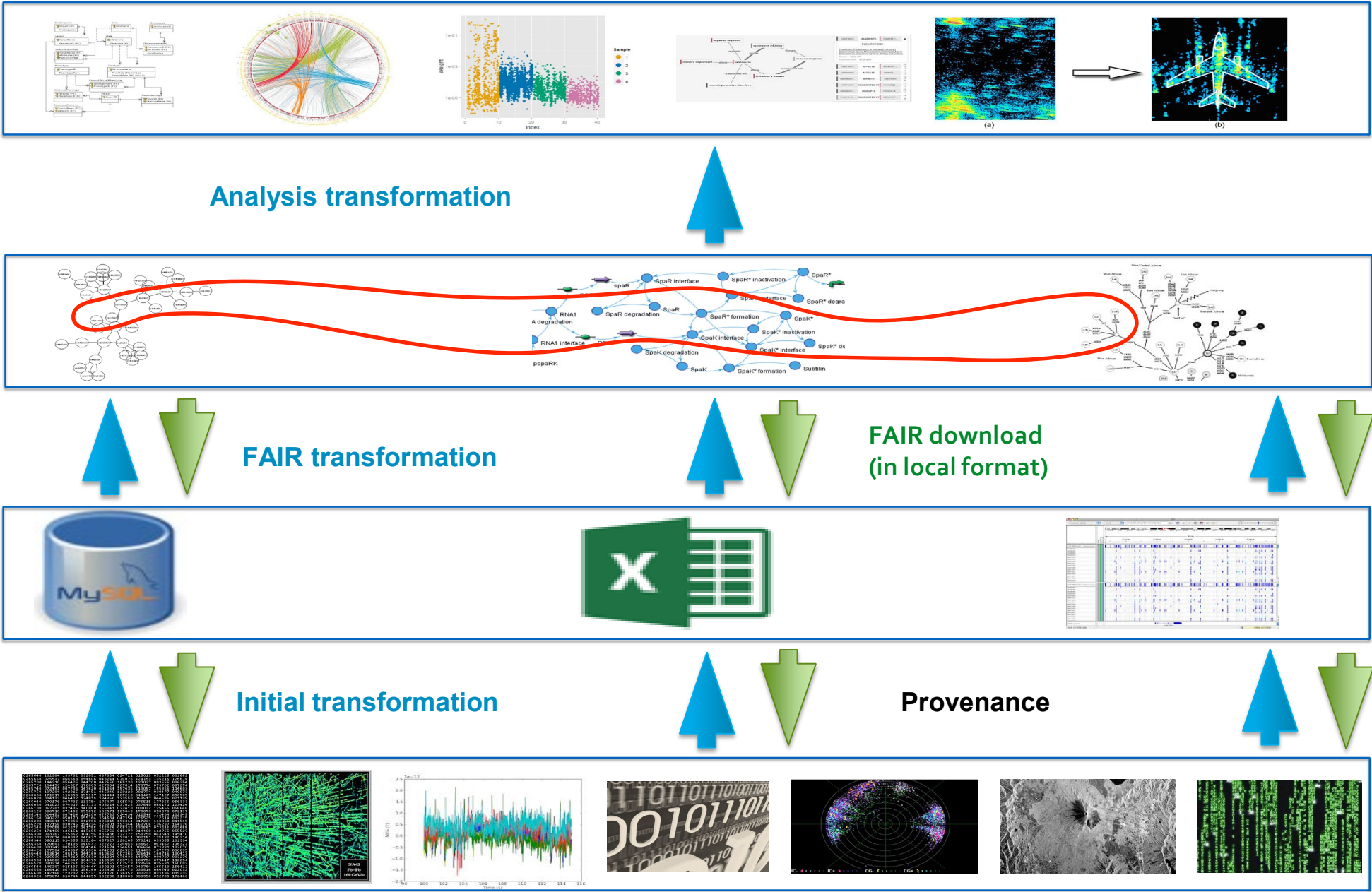
FAIR (meta)data
(RDF,XML etc.)



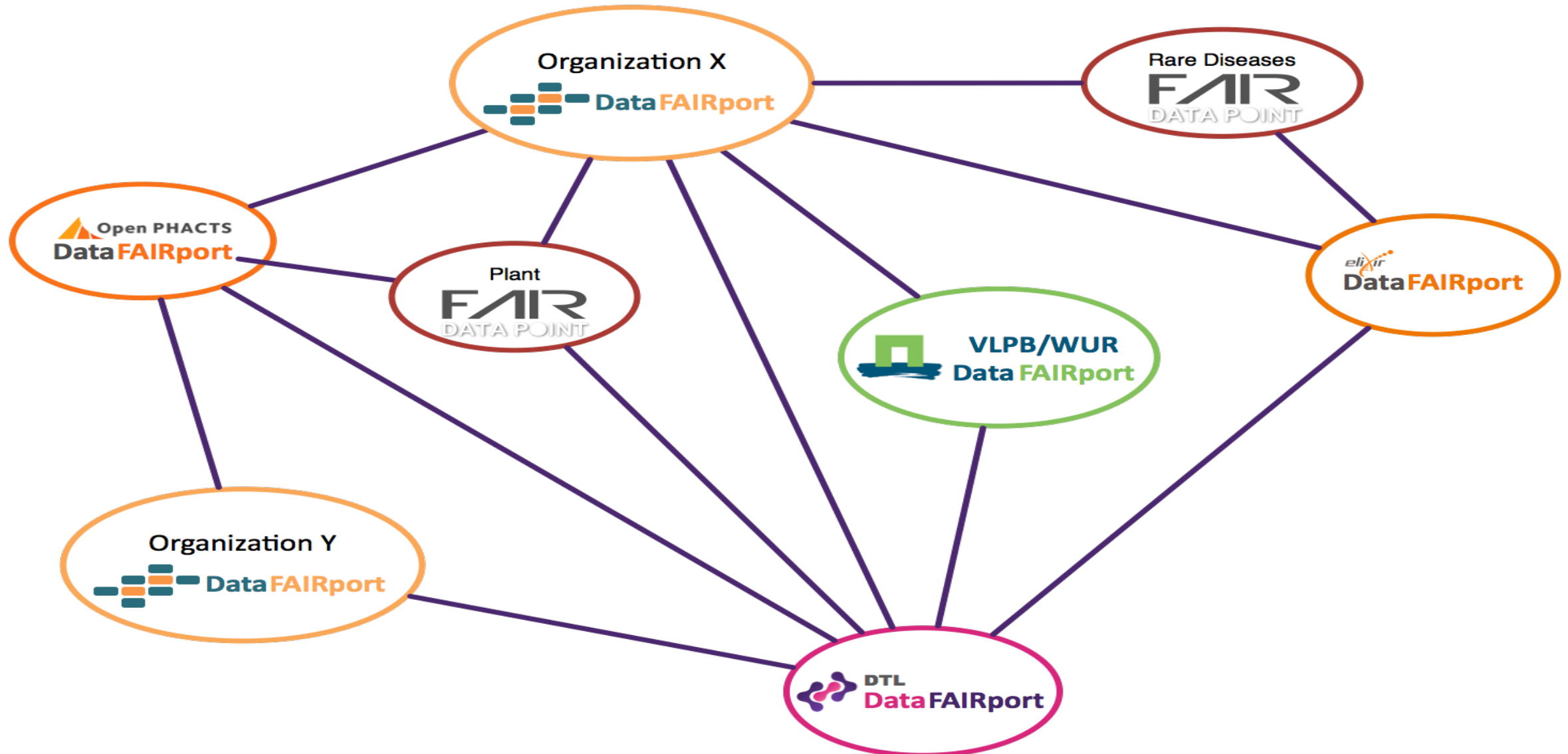
Processed data
(primary storage format)



Raw data
(many formats)



Distributed Architecture

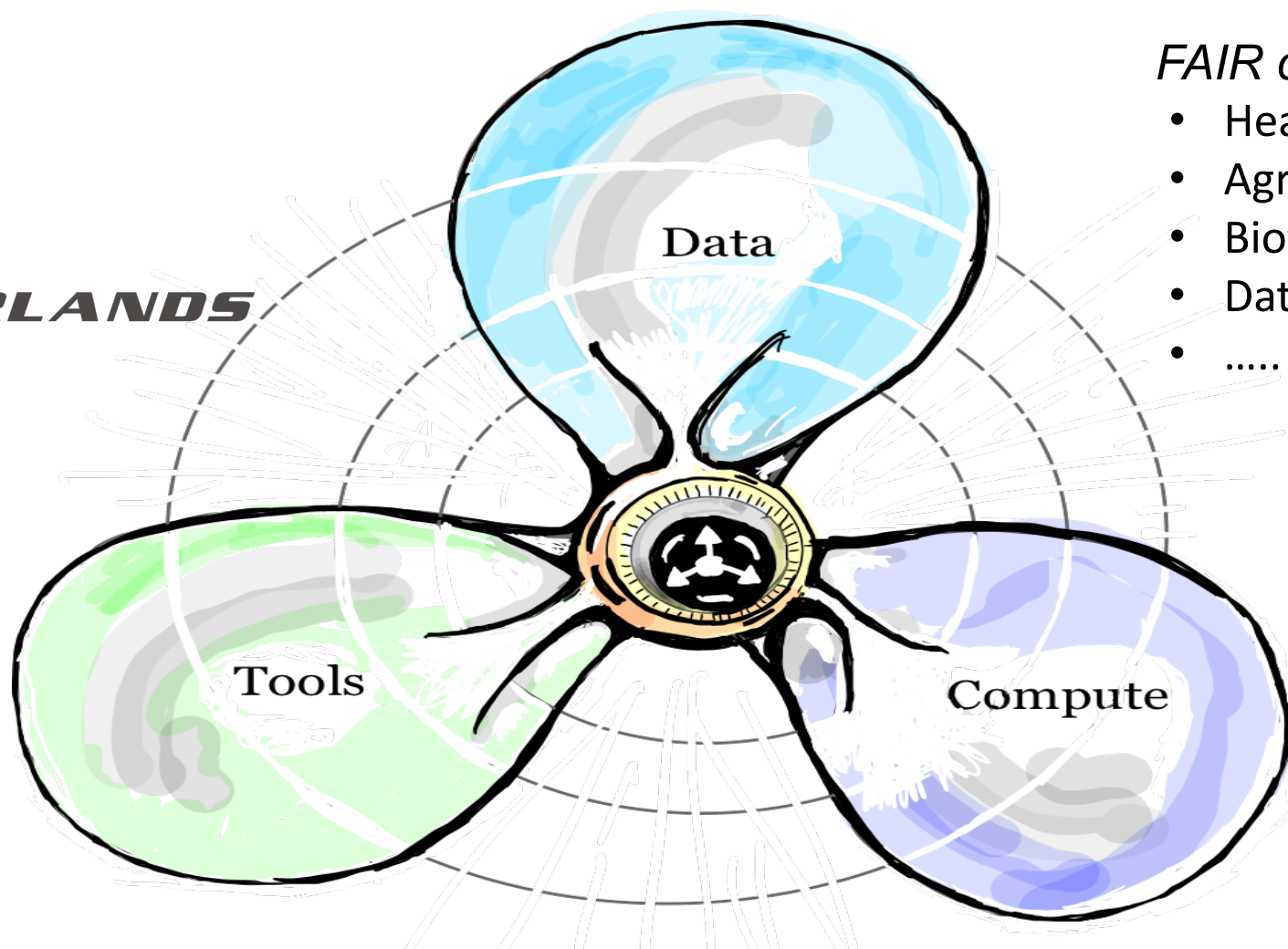


FAIR INFRASTRUCTURE

DATA, TOOLS, WORKFLOWS, MODELS AND COMPUTE



*FAIR-compliant
tooling
(workflows,
models)*



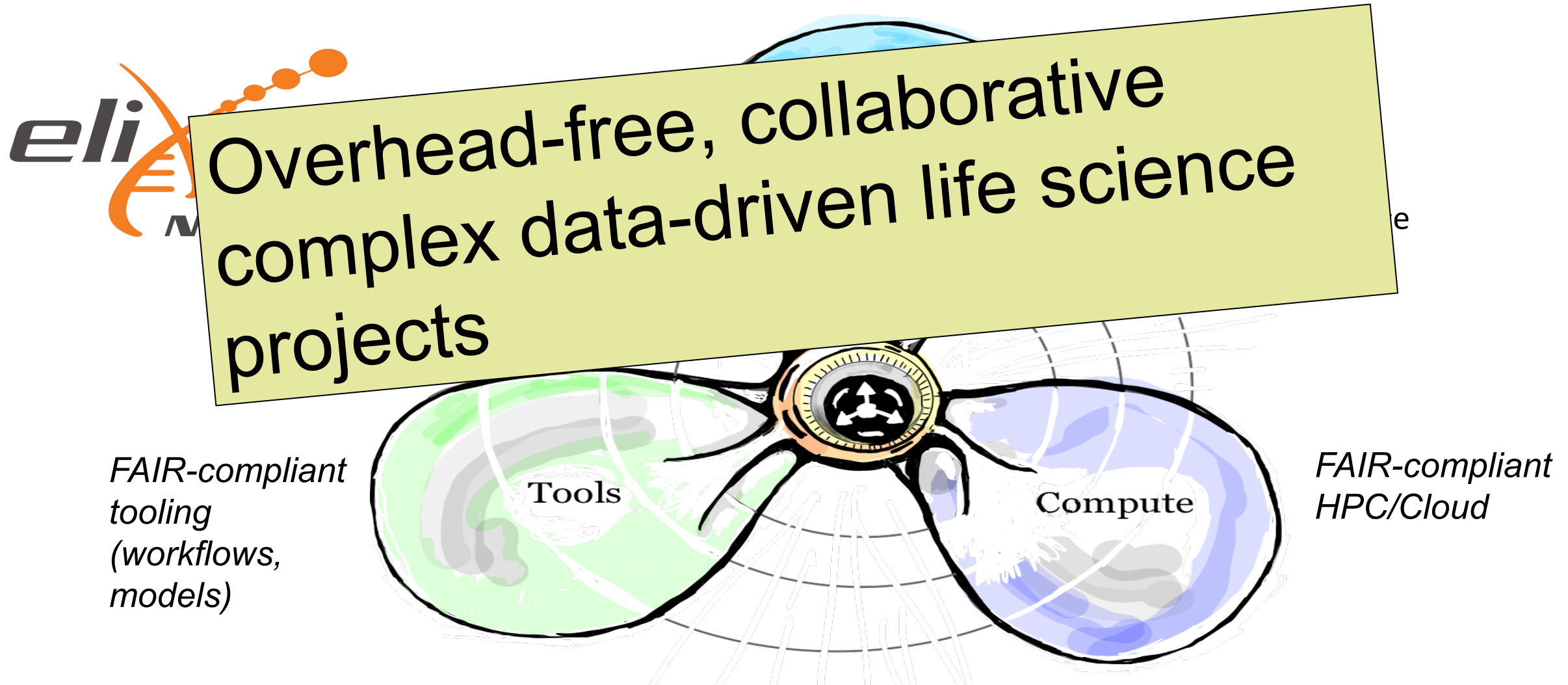
FAIR data

- Health
- Agrofood
- Biotech
- Data science
-

*FAIR-compliant
HPC/Cloud*

FAIR INFRASTRUCTURE

DATA, TOOLS, WORKFLOWS, MODELS AND COMPUTE



DTL/ELIXIR-NL/BioSB: data-related training for the life science research community

Training & Community Building for researchers, developers and trainers

- FAIR Data
- Data Stewardship



- Software and Data Carpentry
- Galaxy



- Next Generation Sequencing

- Infrastructure



- Bioinformatics, Systems Biology, Metabolomics



- FAIR Data Interest Group
- Data Stewards Interest Group

- Carpentries-nl Interest Group
- Galaxy Interest Group

- NGS Interest Group

- Programmers Meeting
- Compute Resources for LS Research Interest Group

Galaxy



DTL Interest Group Galaxy: Towards national strategy

- F2F Meetings in 2016 and 2017. 2018 being planned
- Chair: *Saskia Hiltemann*
- www.choosegalaxy.nl
- ELIXIR-NL Galaxy course (Oct 2018).
Led by *Cico Zhang* and *Saskia Hiltemann*
- Liaising with ELIXIR Europe Galaxy activities; actively involved in co-organizing workshops and teaching in workshops
- Working to realise national Galaxy server





Netherlands Bioinformatics and Systems Biology Research School

- Representing Dutch Bioinformatics & Systems Biology **research community**
- Bioinformatics and Systems Biology **education programme**
 - Research committee
 - Education committee
 - Student committee (RSG/ YCB)
 - Industrial platform
- Introductory, core and specific courses
- Hot topics meetings
- Well-attended annual BioSB Conference

www.biosb.nl



ISSN: 2590-0633

- Open Access
- Guide for Authors
- Track Your Paper

- Your Research Data
- Share your research data
 - Visualize your data

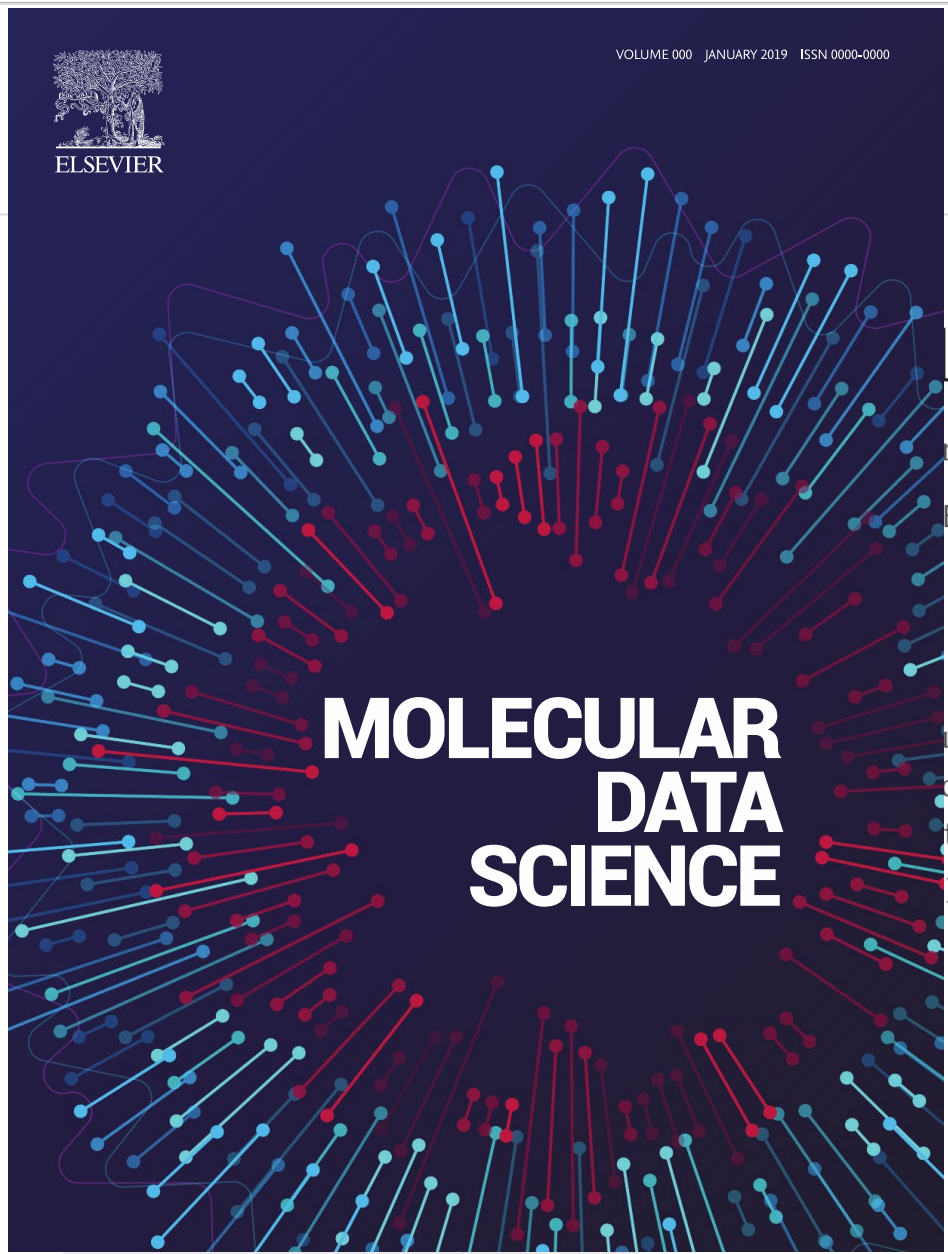
Molecular Data Science

Editor: Jaap Heringa

> View Editorial Board

The vast accumulation of health-related and biomedical data resources and the rapid proliferation of technological developments in data analytics are opening up new avenues to gaining insight in complex biological processes. High-throughput and precision measurement technologies are generating at a...

Read more



> Visualize your data



Molecular Data Science

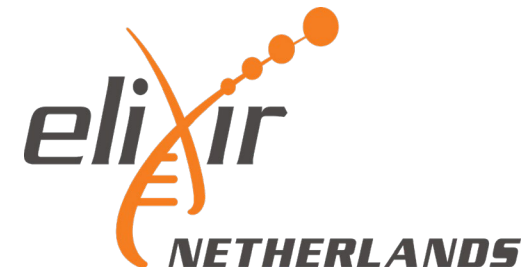
Editorial Board

Open Access

Integration of health-related and biomedical data resources and the application of technological developments in data analytics are opening up new opportunities to gaining insight in complex biological processes. High-precision measurement technologies are generating at a...

Feedback

Thank you



THANK YOU

MedUni Vienna researchers have developed a way to use retinal imaging to diagnose diabetes, explaining how the technique could be used for other conditions...

MedUni Vienna's Department of Ophthalmology and Optometrics has revealed its progress in using digital retinal screening to diagnose patients with diabetes.

"Exactly one year ago, we spoke about the fact that, in future, it will be possible to diagnose diabetes from the eye using automatic digital retinal screening, without the assistance of an ophthalmologist – 12 months on, MedUni Vienna is right in the middle of this digital revolution," said Dr Ursula Schmidt-Erfurth, Head of MedUni Vienna's Department of Ophthalmology and Optometrics.

One of the latest developments is automatic diabetes screening, which was recently implemented at MedUni Vienna.



.One of the latest developments is automatic diabetes screening, which was recently implemented at MedUni Vienna.

The Division of Cardiology led by Dr Christian Hengstenberg within the Department of Medicine II is working on how digital retinal analysis can also be used in future for the early diagnosis of cardiovascular diseases.



ELIXIR International Goals

- By the end of 2019, ELIXIR AAI and Registries have a defined role in the EOSC architecture.
- By the end of 2021, ELIXIR can routinely deploy containerised workflows for federated data in harmonised clouds that support transnational user access.
- By end of 2023, all ELIXIR Nodes can routinely support their users with access to EOSC via a set of established ELIXIR Infrastructure Services that provide access to clouds, workflows and ELIXIR data resources.
- By the end of 2021, at least 8 ELIXIR Nodes will have established national initiatives for data management and hosted events to support industry and SME usage of open, public data resources.
- By the end of 2023, ELIXIR Nodes provide FAIR data management support for LS including actively supporting submissions to the ELIXIR Deposition Databases.
- By end of 2023, ELIXIR Nodes understand the national landscape of open data use in SMEs and actively support reuse of data by these in partnership with national innovation programmes.

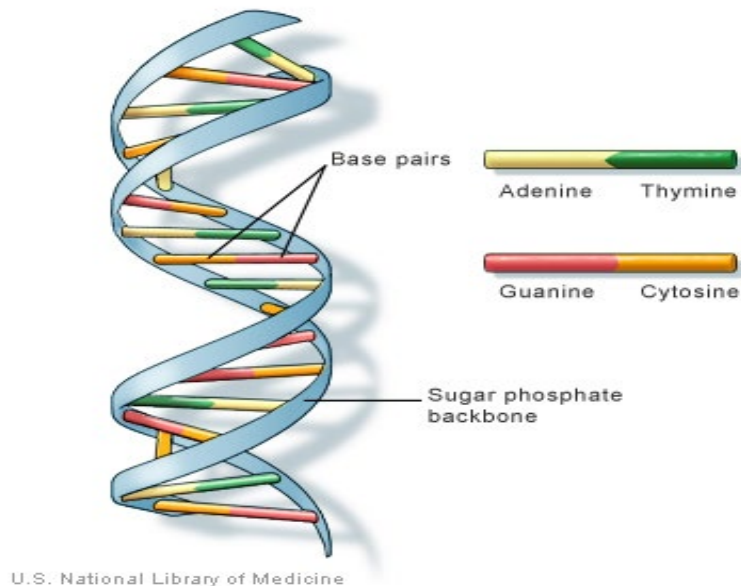
The champion of data storage?

- Storing all data of 2017 (9 yottabyte)?
- Ultra modern disk technology?
- Or a molecule that evolved over about 4.2 billion years...



The champion of data storage?

- Storing all data of 2017 (9 yottabyte)?
- Ultra modern disk technology?
- Or a molecule that evolved over about 4.2 billion years...



**DNA can store 1
yottabyte of data on
roughly 1 gram!**

George Church, Harvard Univ.

