DTL

DUTCH TECHCENTRE FOR LIFE SCIENCES

LUMC

DATASTEWARDSHIP FOR DISCOVERY
NEW ROLES FOR OLD VACCINES?
CORRELATION > MECHANISMS

Pensieres June 10, 2015

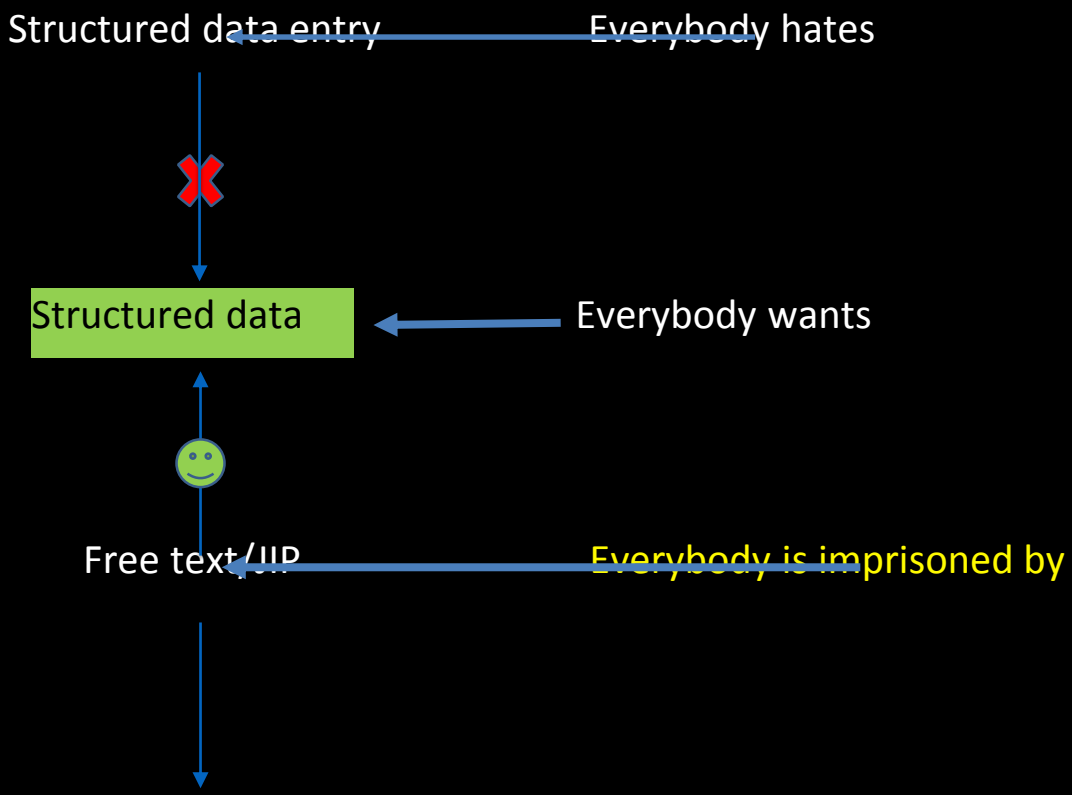
elixir NETHERLANDS

5 min.

of

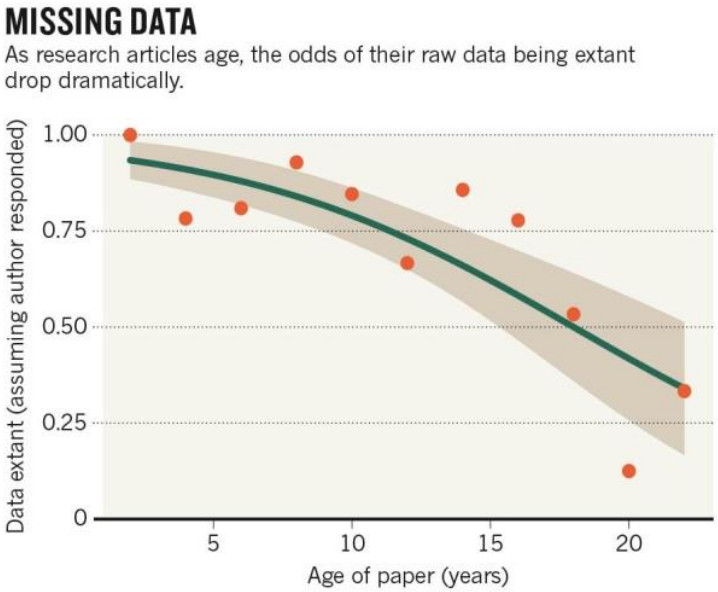
Lamenting

The problem (also with community annotation) in a nutshell

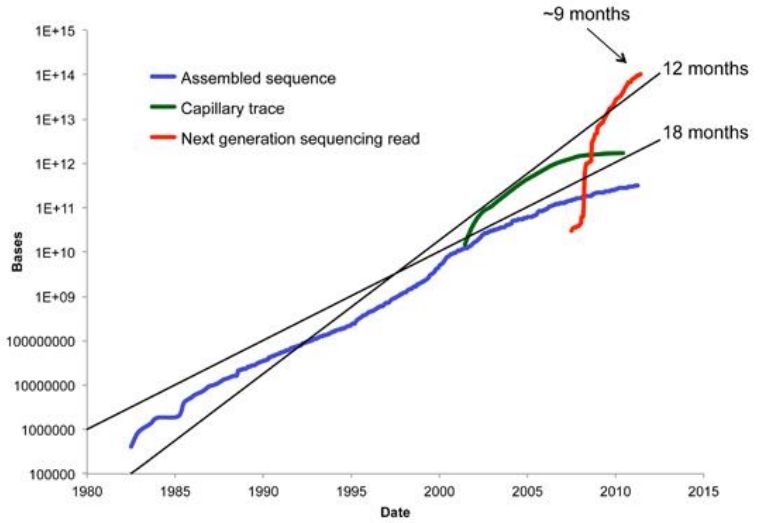


We collectively impair machine-assisted knowledge discovery !

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

AND

then came

ELIXIR

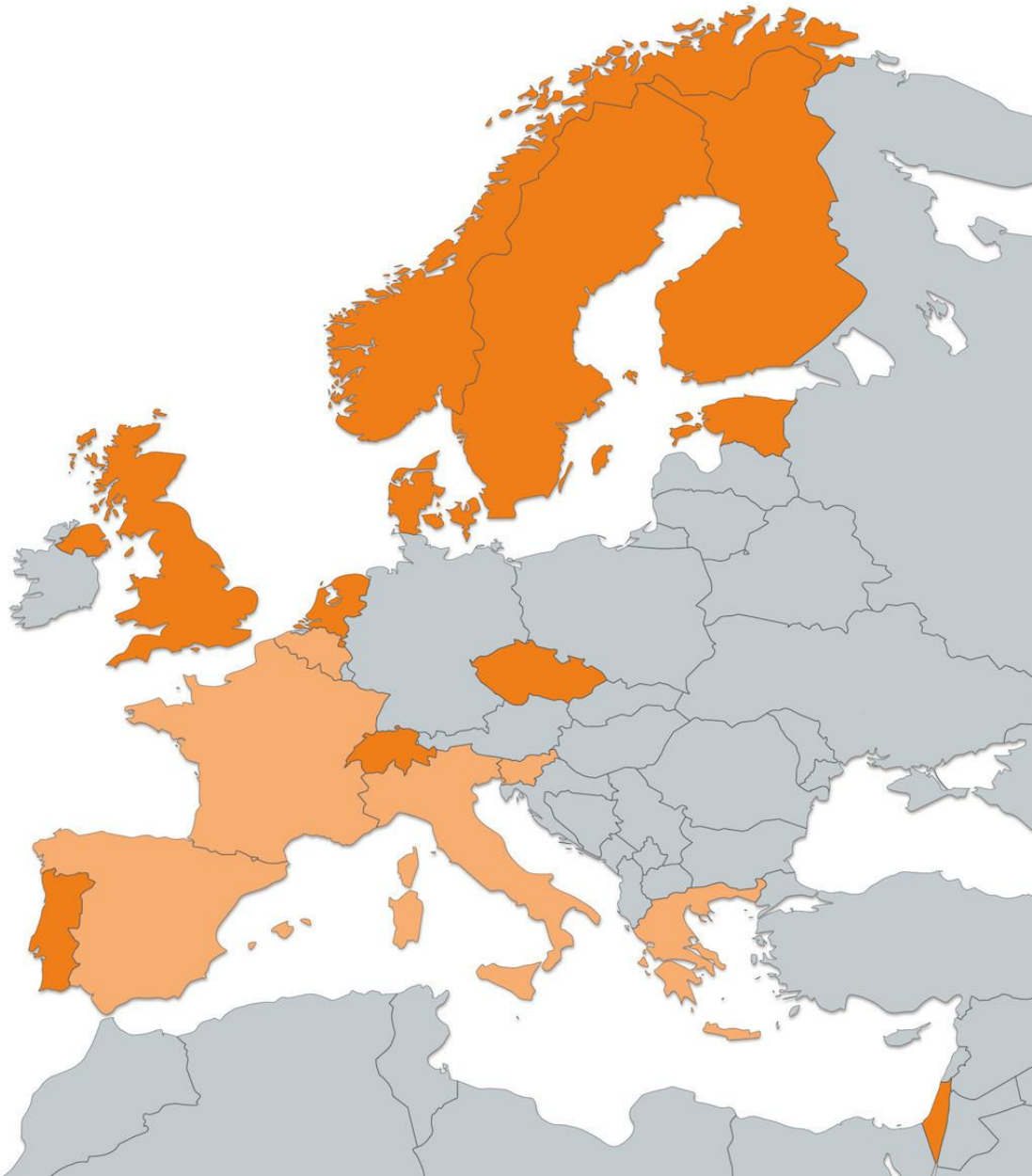
ELIXIR: the European Research Infrastructure for biological data

ECA signed:

Czech Republic
Denmark
EMBL
Estonia
Finland
Israel
Netherlands
Norway
Portugal
Sweden
Switzerland
United Kingdom

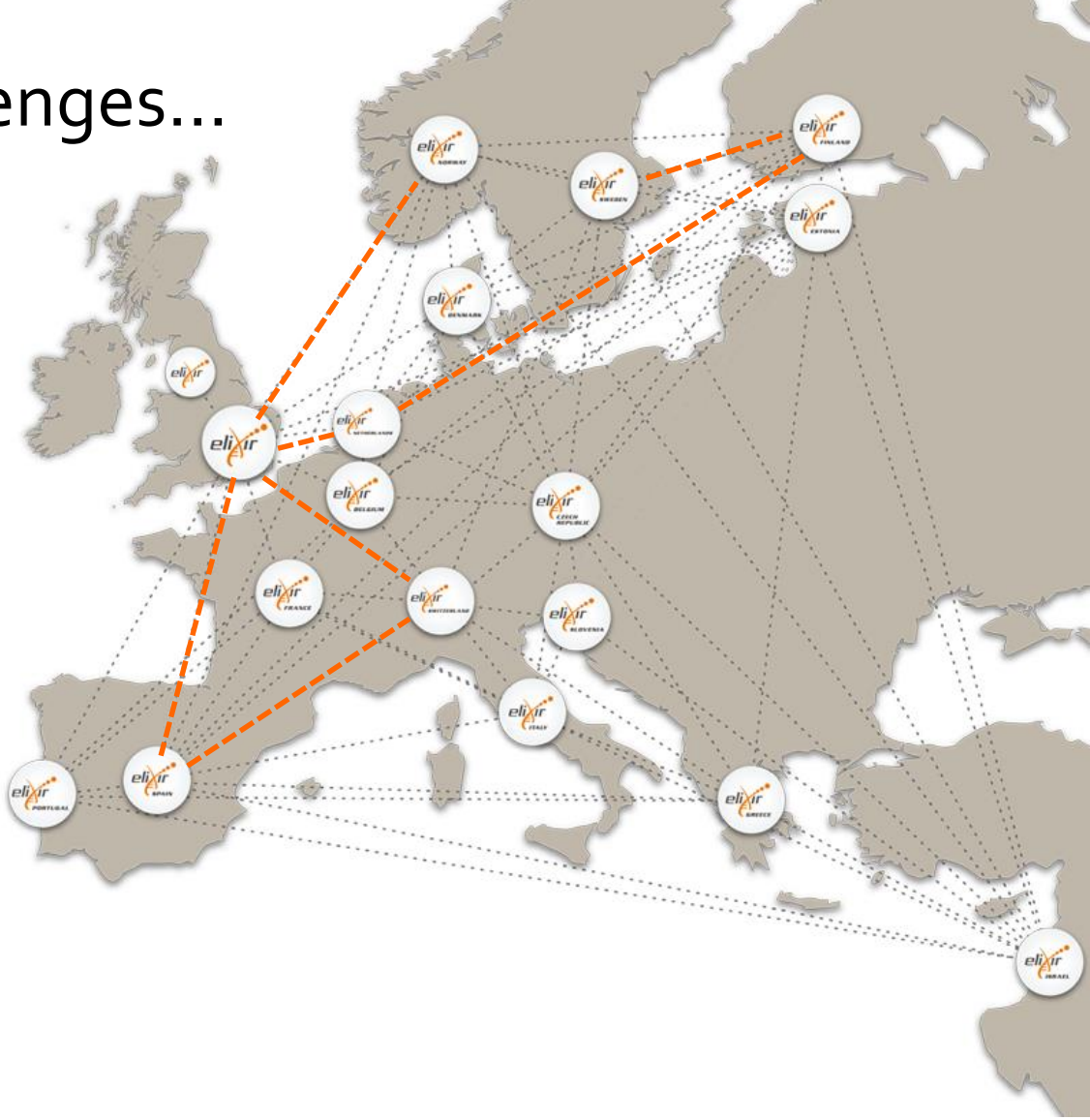
MoU signed:

Belgium
France
Greece
Italy
Slovenia
Spain



Data Roaming Challenges…

- Submission of large data-sets
  - Standards, Curation, Deposition
- Access of large and growing reference data sets
- Replication of large and growing reference data sets
- Secure access to human reference and study data



The Dutch Node

DTL
DUTCH TECHCENTRE FOR LIFE SCIENCES

ELIXIR's NL node is hosted by the Dutch Techcenter for Life sciences (DTL), a public private partnership that aims to jointly establish a world-class Next Generation Life Sciences cross technology & cross sector capability including a federated data infrastructure.

The ELIXIR NL node acts as the gateway of ELIXIR capabilities and expertise to all the associated partners in DTL. The NL node focuses its contribution to ELIXIR in three core areas: data interoperability, compute & storage infrastructure services and training.

Collaborating organisations

University Medical Centers
- Academic Medical Centre (AMC)
- Erasmus Medical Centre Rotterdam (EMC)
- Leiden University Medical Centre (LUMC)
- Radboud University Nijmegen Medical Centre (UMCN)
- University of Groningen Medical Centre (UMCG)
- Utrecht University Medical Centre (UMCU)
- VU University Medical Centre (VUMC)
- Maastricht UMC+

Institutes
- Centrum voor Wiskunde en Informatica (CWI)
- CBS-KNAW
- Hubrecht Institute
- Netherlands Cancer Institute (NKI)
- Netherlands eScience Centre
- Plant Research International (PRI)
- RIKILT – Institute of Food Safety
- Royal Tropical Institute (KIT)
- SURFnet & SURFsara

Universities
- Delft University of Technology (TU-Delft)
- Eindhoven University of Technology (TUe)
- Leiden University (UL)
- Maastricht University (UM)
- Radboud University Nijmegen (RU)
- University of Amsterdam (UvA)
- University of Groningen (RUG)
- Utrecht University (UU)
- VU University of Amsterdam (VU)
- Wageningen University (WU)

Private sector partners
- DSM
- Philips
- TNO
- Unilever
- SME's

Data interoperability and exchange

Compute and storage infrastructure services

Training & Education

Data interoperability and exchange

Several Dutch groups have specialized in data capture standards, software, semantic web standards and formats to enable meaningful exchange and integration of biological information. ELIXIR NL will focus on implementing and developing professional capturing, publishing and hosting of data in standard (semantically interoperable) format that will be offered in a public-private partnership in close collaboration with other ELIXIR nodes and the Hub

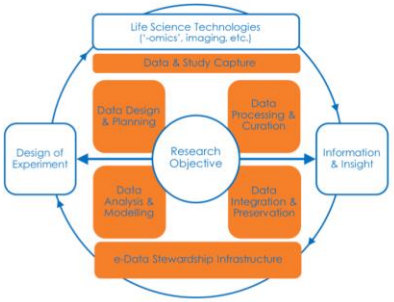
Compute and storage infrastructure services

The e-infrastructure capabilities of the Dutch national compute, data and ultra high speed network infrastructure are a clear strength of the ELIXIR NL Node, with extensive experience in running a shared compute and storage environment for collaborative life science projects. The ELIXIR NL node will focus on supporting complex data/compute-intensive life science projects, in collaboration with, and complementary to the offerings of other ELIXIR nodes.

Training

ELIXIR-NL will contribute extensive experience and capacity in bioinformatics training built up within NBIC, and will leverage broad education & training capabilities of the broader DTL partnership in a comprehensive portfolio in the broader scope of the ELIXIR train programme.

ELIXIR NL: focus within the Data Cycle



NWO Netherlands Organisation for Scientific Research · ZonMw · DTL DUTCH TECHCENTRE FOR LIFE SCIENCES · netherlands eScience center · SURF

ELIXIR Infrastructure = Nodes + ELIXIR Platforms

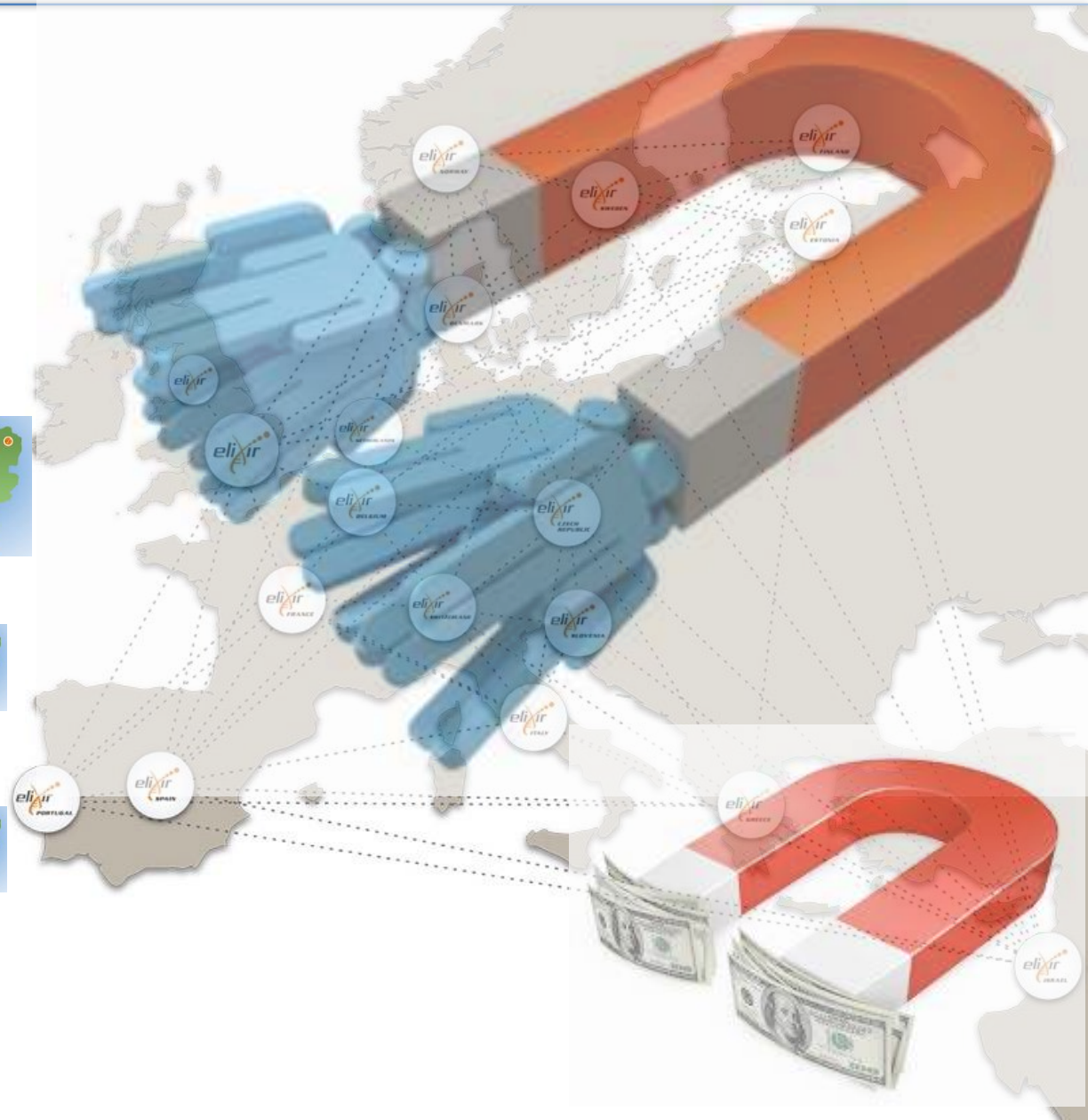
Data
Sustain core data resources

Tools
Services & connectors to drive access and exploitation

Interoperability
Integration and interoperability of data and services.

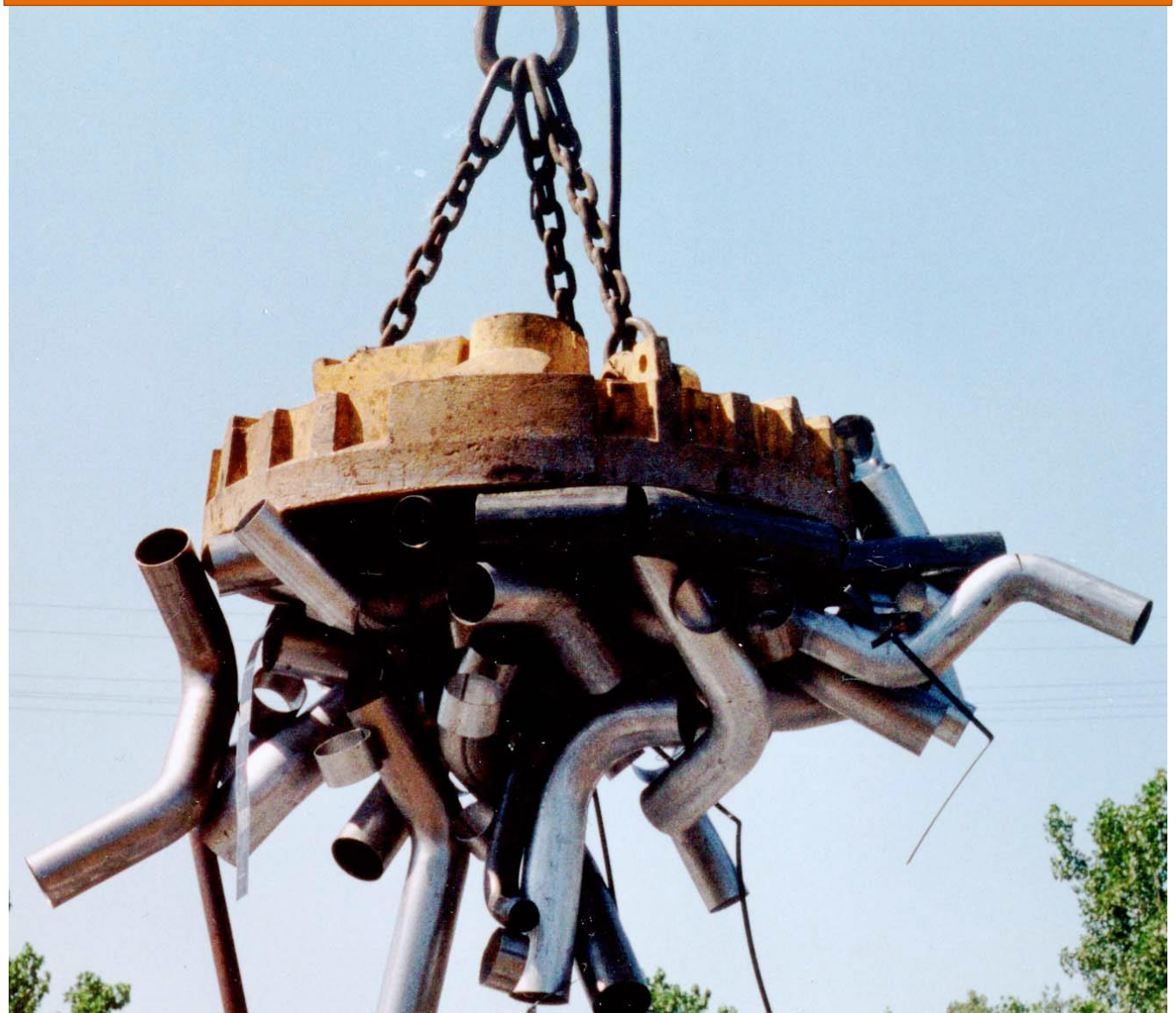
Compute
Access, Exchange & Compute on sensitive data

Training
Professional skills for managing and exploiting data

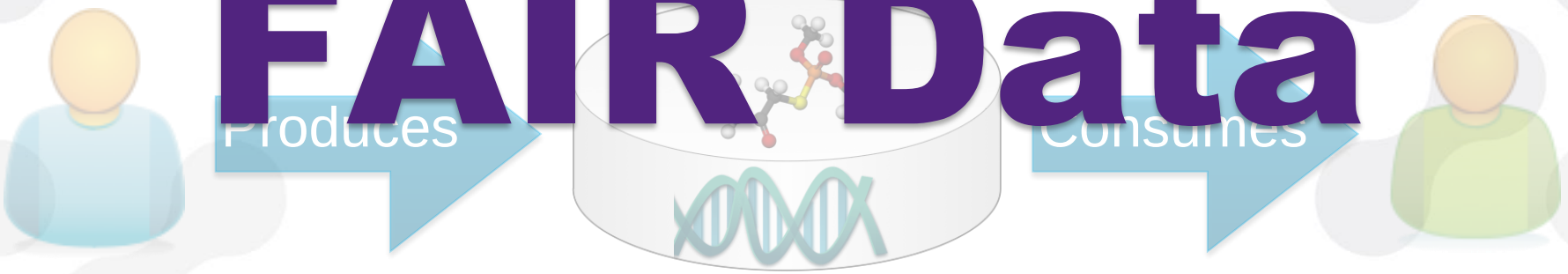


Tools

Services & connectors to drive access and exploitation



FAIR Data



DTL

WHAT IS FAIR DATA?

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

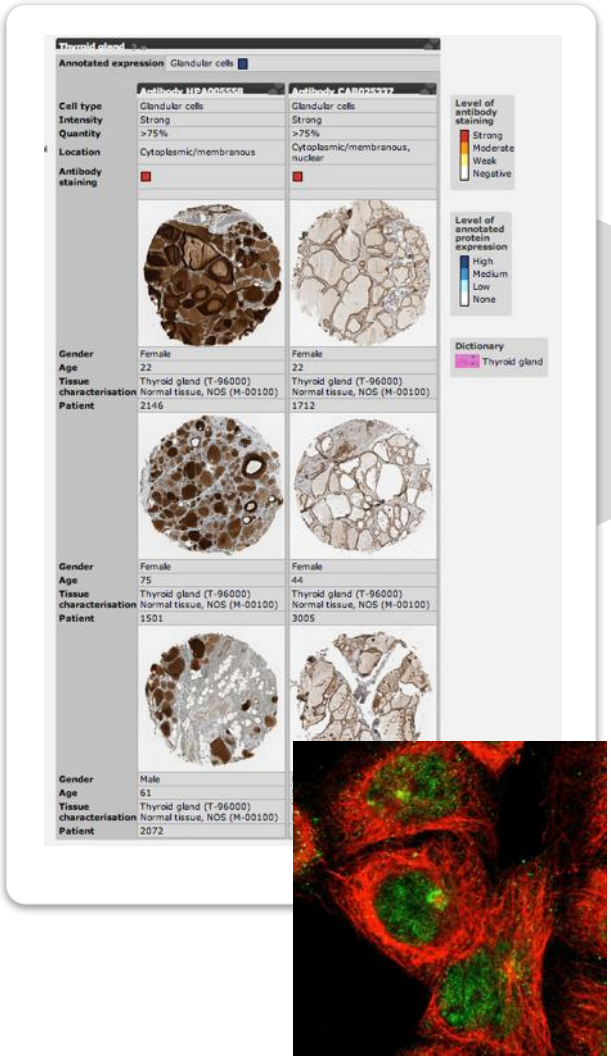
Findable - (meta)data is uniquely and persistently identifiable. Should have basic machine readable descriptive metadata.

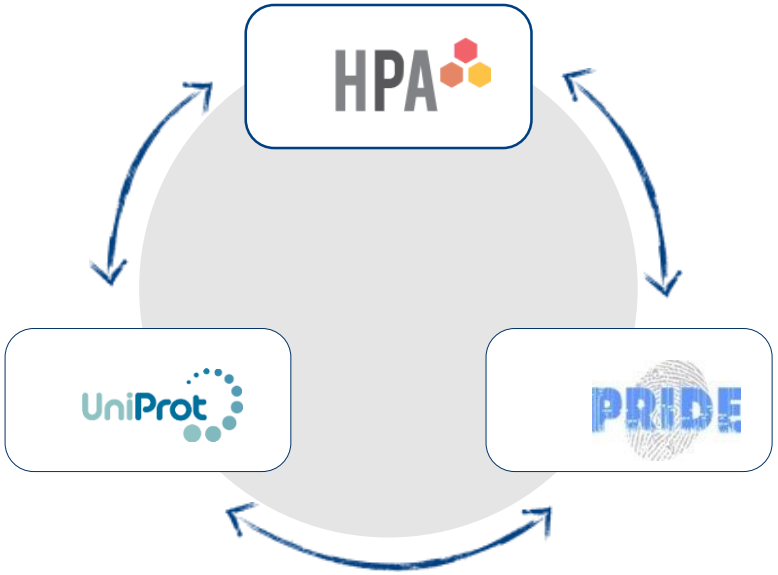
Accessible - identifiers should provide a mechanism for (meta) data access, including authentication, access protocol, license, etc.

Interoperable - (meta)data should be 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.

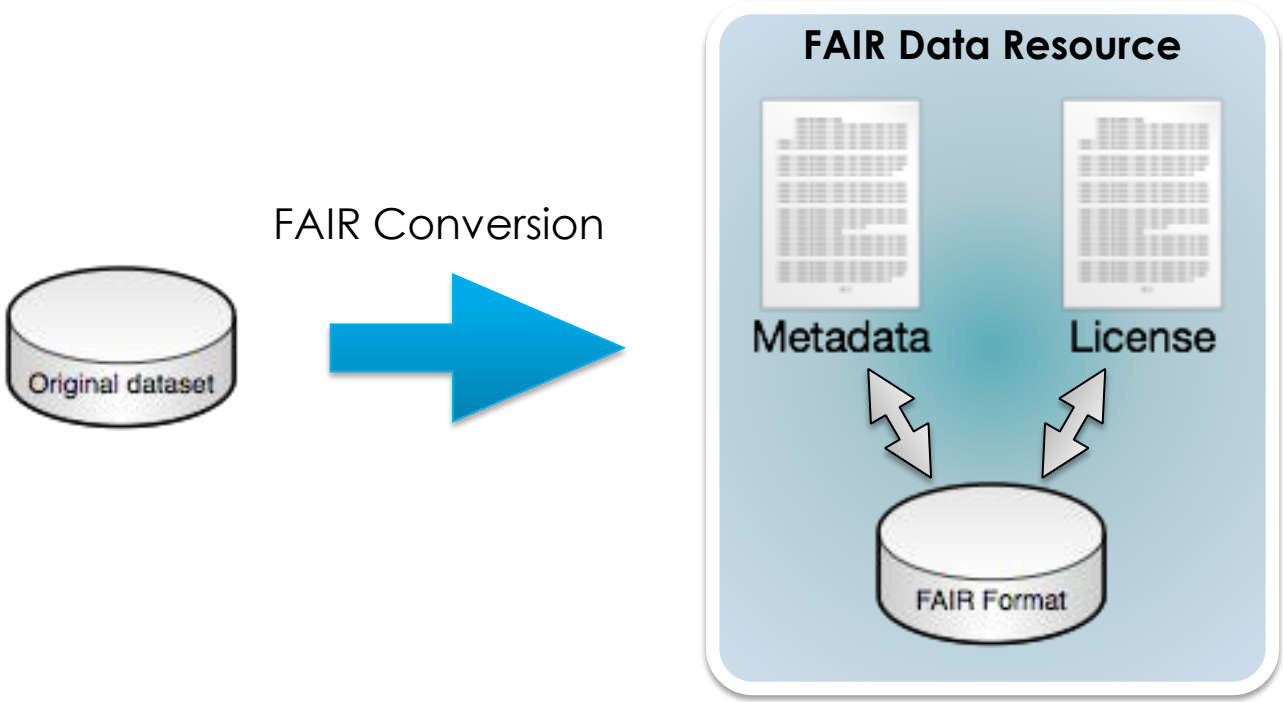
Data interoperability – Human Protein Atlas



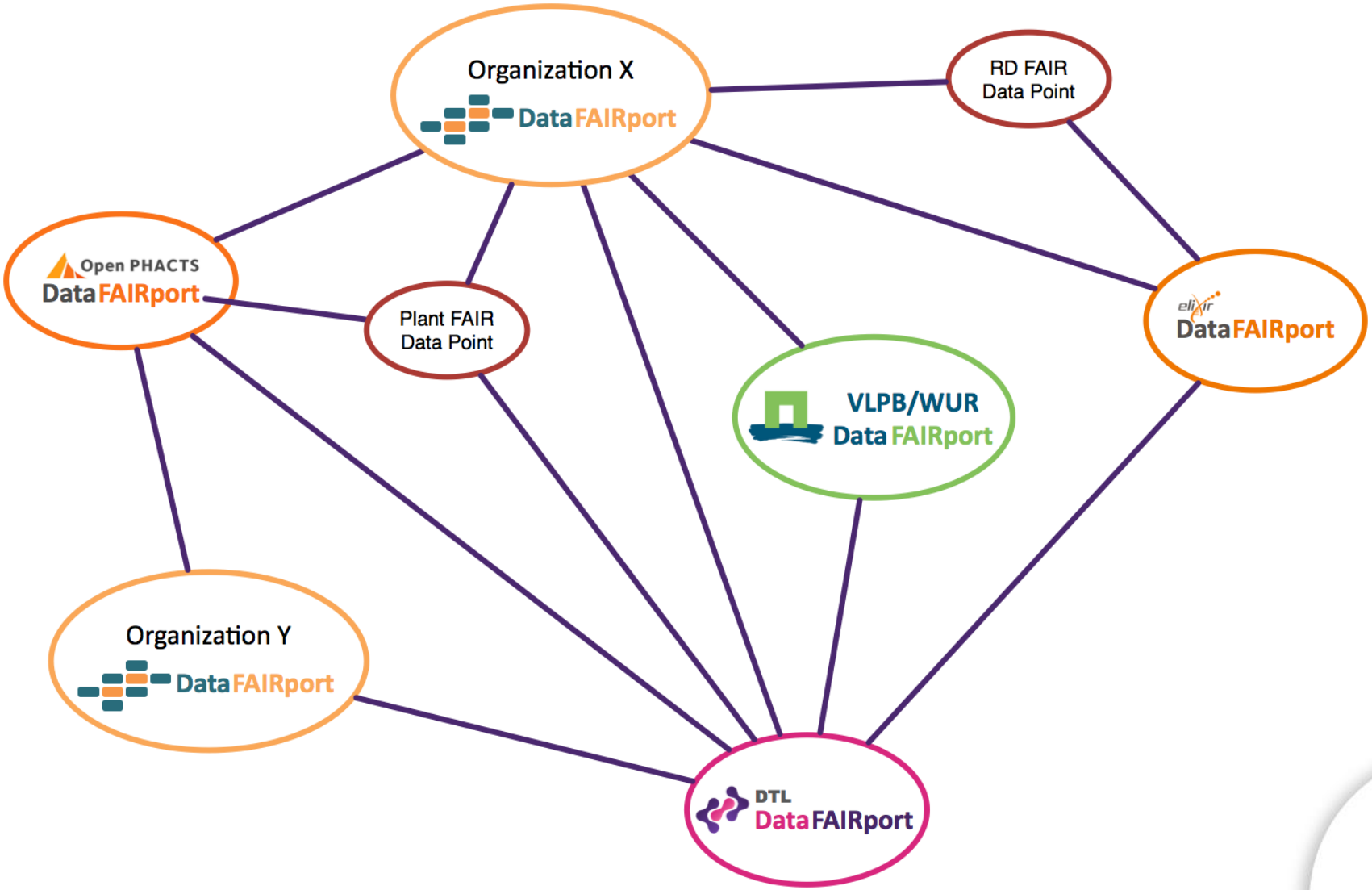


FAIR DATA RESOURCE

Datasets expressed using one of the prescribed standards of the FAIR Data Protocol, with metadata complying with the protocol and license. The original dataset is transformed into a FAIR format and proper metadata and license are added to produce a FAIR Data Resource. The original and the FAIR version can co-exist, each one fulfilling its own purpose.



DISTRIBUTED ARCHITECTURE: EUROPEAN OPEN SCIENCE CLOUD

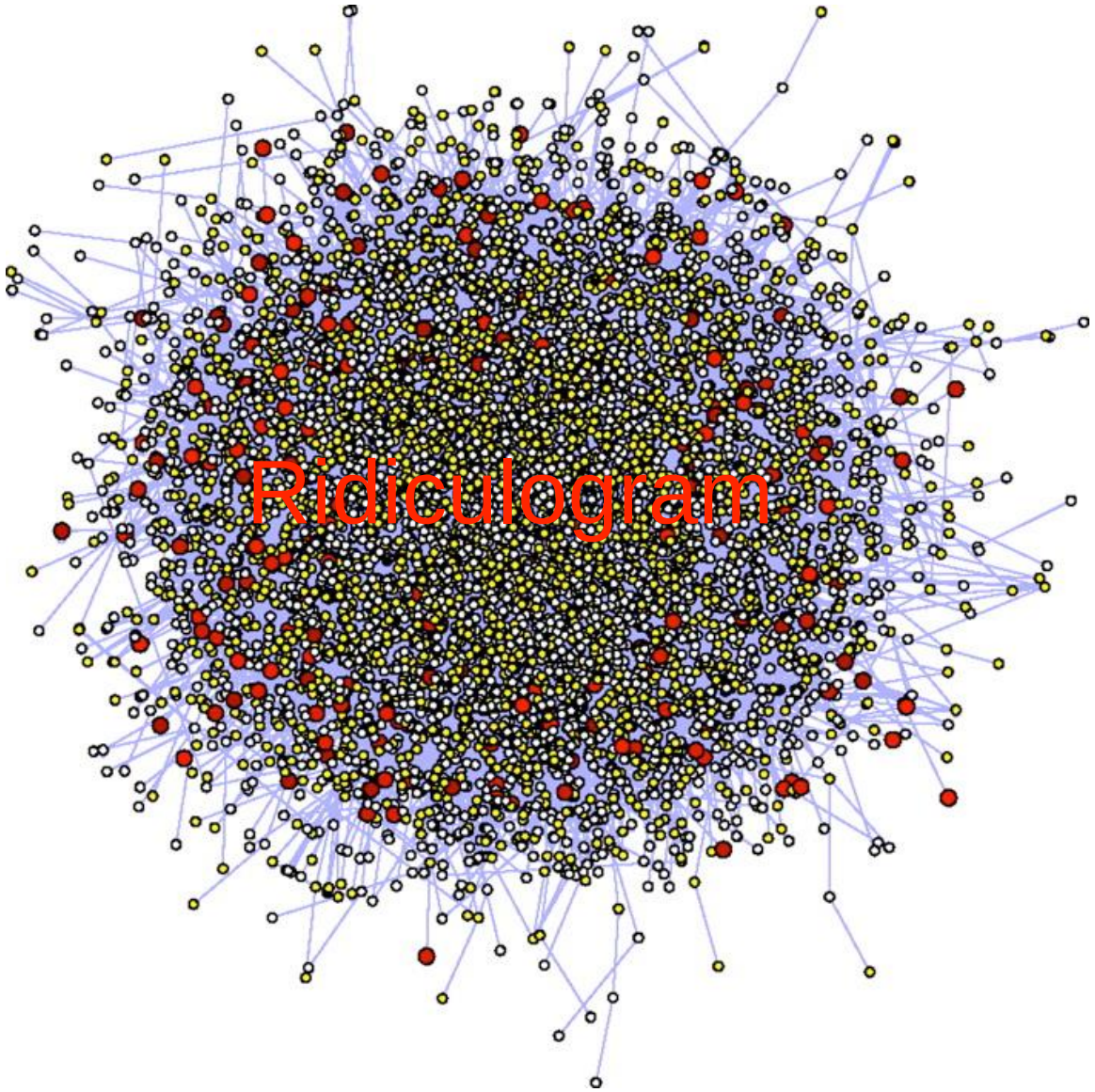


So

what about

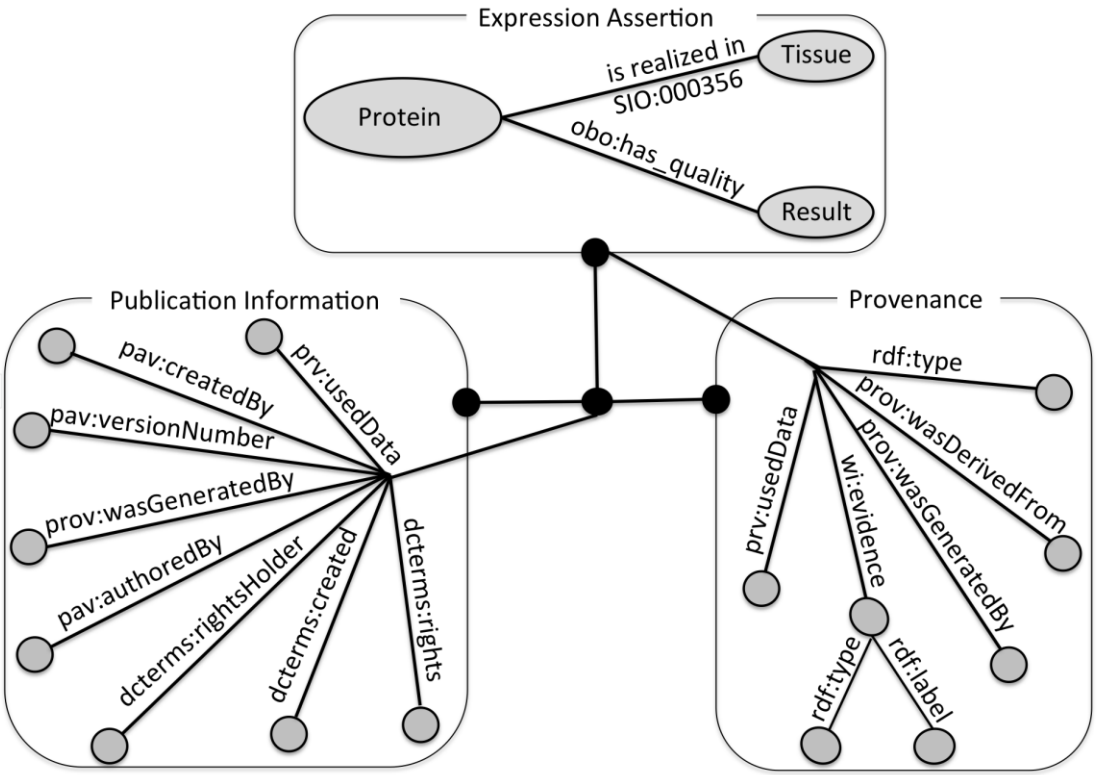
Vaccines?

Simplified eScience



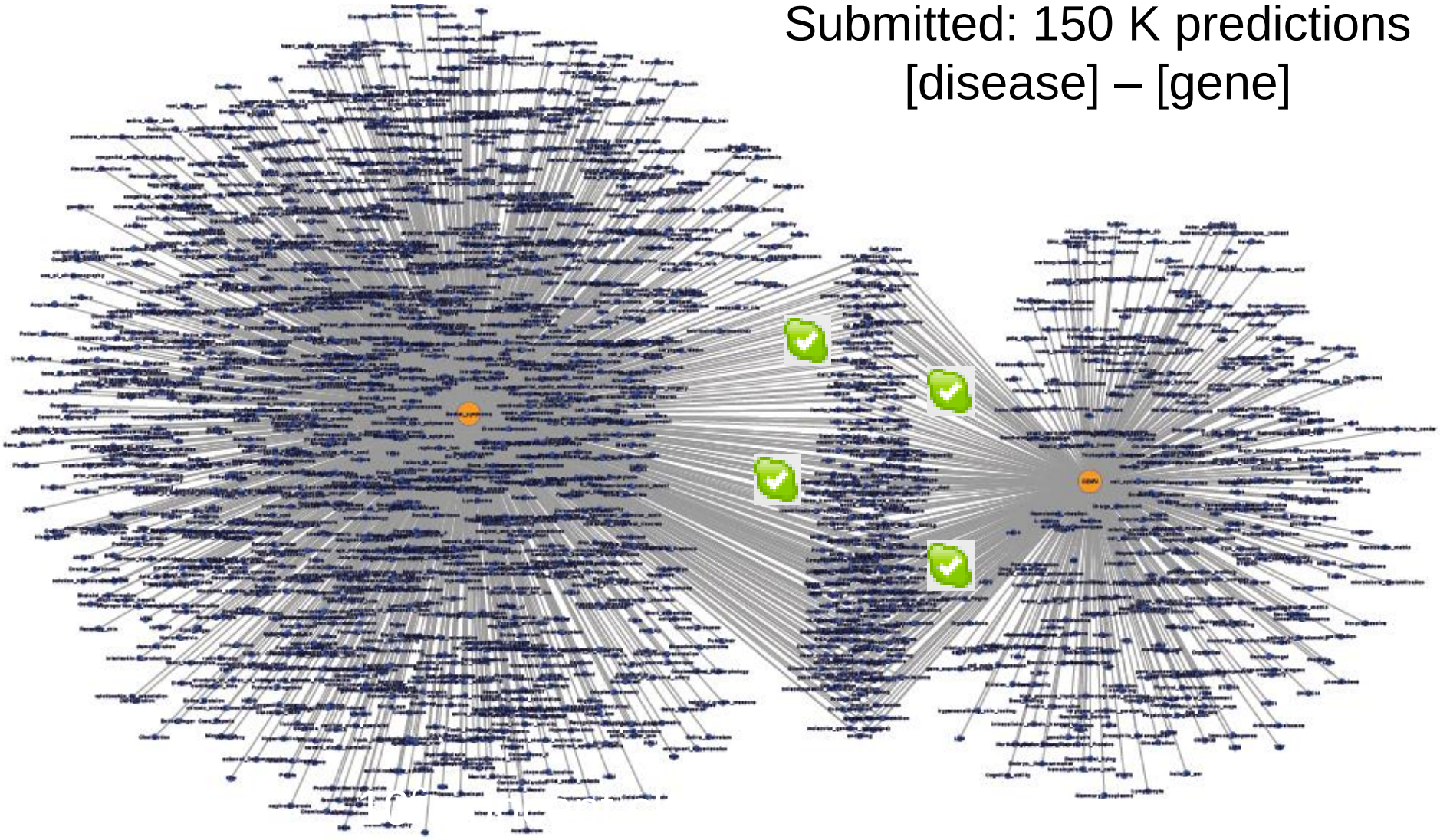
Nanopub.org

Learn how to create, find, use and cite nanopublications



We publish about less than a million LSConcepts !

Submitted: 150 K predictions
[disease] – [gene]

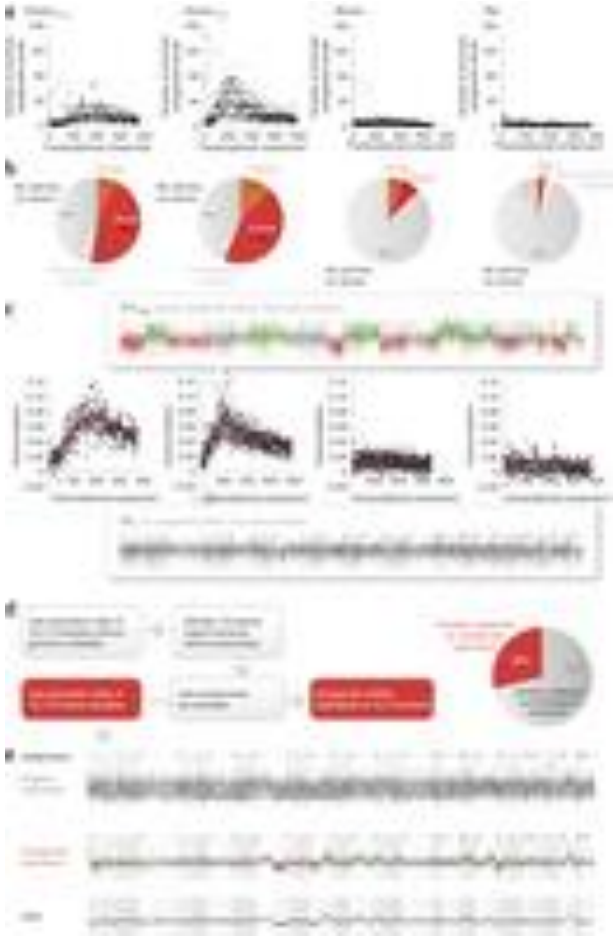


Gene expression analysis identifies global gene dosage sensitivity in cancer

Rudolf S N Fehrmann, Juha M Karjalainen, Małgorzata Krajewska, Harm-Jan Westra, David Maloney, Anton Simeonov, Tune H Pers, Joel N Hirschhorn, Ritsert C Jansen, Erik A Schultes, Herman H H B M van Haagen, Elisabeth G E de Vries, Gerard J te Meerman, Cisca Wijmenga, Marcel A T M van Vugt & Lude Franke

Affiliations | Contributions | Corresponding authors

Nature Genetics **47**, 115–125 (2015) | doi:10.1038/ng.3173
Received 29 July 2014 | Accepted 02 December 2014 | Published online 12 January 2015



We reanalyzed **77,840 expression profiles** and observed a limi

1: concept profiles were successfully
Used to annotate protein-function
(2008, 2009)

Betaine-homocysteine methyltransferase-2: cDNA cloning, gene sequence, physical mapping, and expression of the human and mouse genes.

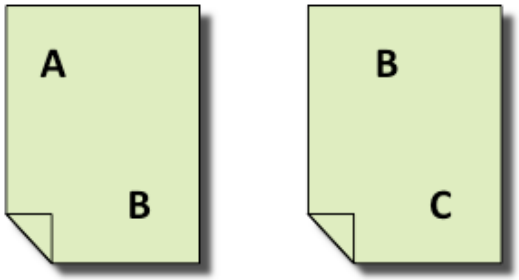
Chadwick LH, McCandless SE, Silverman GL, Schwartz S, Westaway D, Nadeau JH.

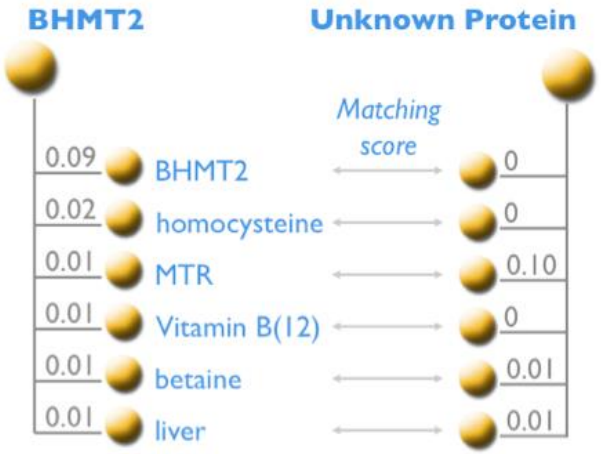
Department of Genetics, Case Western Reserve University School of Medicine, Cleveland, Ohio 44106, USA.

Anomalies in folate and homocysteine metabolism can result in homocysteinemia and are implicated in disorders ranging from vascular disease to neural tube defects. Two enzymes are known to methylate homocysteine, vitamin B(12)-dependent methionine synthase (MTR) and betaine-homocysteine methyltransferase (BHMT). BHMT uses betaine, an intermediate of choline oxidation, as a methyl donor and is expressed primarily in the liver and kidney. We report the discovery of a novel betaine-homocysteine methyltransferase gene in humans and mice. The human BHMT2 gene is predicted to encode a 363-amino-acid protein (40.3 kDa) that shows 73% amino acid identity to BHMT. The BHMT2 transcript in humans is most abundant in adult liver and kidney and is found at reduced levels in the brain, heart, and skeletal muscle. The mouse Bhmt2 gene shows 69% amino acid identity and 79% similarity to the mouse Bhmt gene and 82% amino acid identity and 87% similarity to the human BHMT2 gene. Bhmt2 is expressed in fetal heart, lung, liver, kidney and eye. The discovery of a third gene with putative homocysteine methyltransferase activity is important for understanding the biochemical balance in using methyltetrahydrofolate and betaine as methyl donors as well as the metabolic flux between folate and choline metabolism in health and disease. Copyright 2000 Academic Press.

PMID: 11087663 [PubMed - indexed for MEDLINE]

Methyltransferase A is shown to have B property
Protein C is shown to have B property.
Therefore, Protein C may be a methyltransferase.





Concept profiles

Direct co-occurrence

Peptide A is shown to bind receptor B
Binding of receptor B is shown to affect disease C
Therefore, Peptide A may affect disease C.

PLoS ONE: Novel Protein-Protein Interactions Inferred from Literature Context

http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0007894

Twitter / Home Google Agenda Apple Yahoo! Google Maps YouTube Wikipedia News (90) Popular

Home Browse Articles About For Readers For Authors and Reviewers Journals Hubs PLoS.org

RESEARCH ARTICLE OPEN ACCESS

Novel Protein-Protein Interactions Inferred from Literature Context

Article Metrics Related Content Comments: 0

Herman H. H. B. M. van Haagen[1*], Peter A. C. 't Hoen[1], Alessandro Botelho Bovo[2], Antoine de Morrée[1], Erik M. van Mulligen[1], Christine Chichester[1], Jan A. Kors[1], Johan T. den Dunnen[1], Gert-Jan B. van Ommen[1], Silvère M. van der Maarel[1], Vinícius Medina Kern[2], Barend Mons[1], Martijn J. Schuemie[1]

1 Biosemantics Association, Department of Human Genetics, Leiden University Medical Center, Leiden, and Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, The Netherlands, **2** Post-Graduate Program in Knowledge Engineering and Management (EGC), Federal University of Santa Catarina (UFSC), Florianópolis, Brazil

To **add a note**, highlight some text. Hide notes

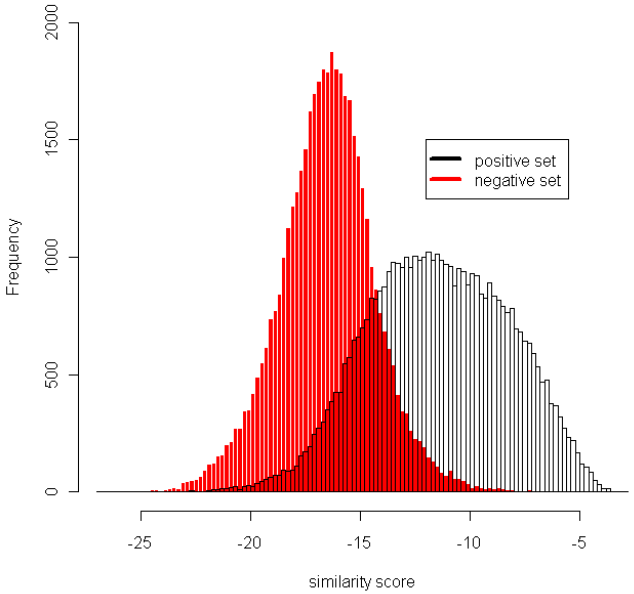
Make a general comment

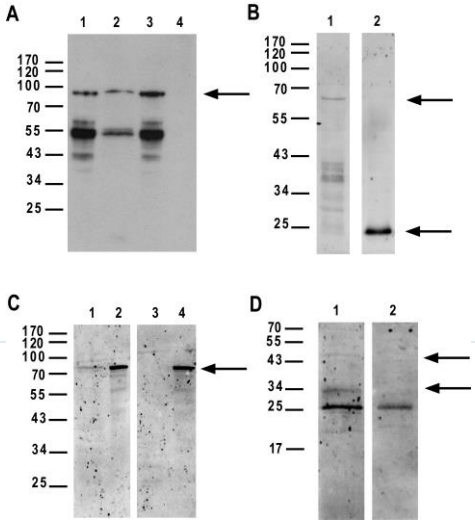
Jump to

Abstract
Introduction
Results
Discussion
Methods
Supporting Information
Author Contributions
References

Abstract

We have developed a method that predicts Protein-Protein Interactions (PPIs) based on the similarity of the context in which proteins appear in literature. This method outperforms previously developed PPI prediction algorithms that rely on the conjunction of two protein names in MEDLINE abstracts. We show significant increases in coverage (76% versus 32%) and sensitivity (66% versus 41% at a specificity of 95%) for the prediction of PPIs currently archived in 6 PPI databases. A retrospective analysis shows that PPIs can efficiently be predicted before they enter PPI databases and before their interaction is explicitly described in the literature. The practical value of the method for discovery of novel PPIs is illustrated by the experimental confirmation of the inferred physical interaction between CAPN3 and PARVB, which was based on frequent co-occurrence of both proteins with concepts like Z-disc, dysferlin, and alpha-actinin. The relationships between proteins predicted by our method are broader than PPIs, and include proteins in the same complex or pathway. Dependent on the type of relationships deemed



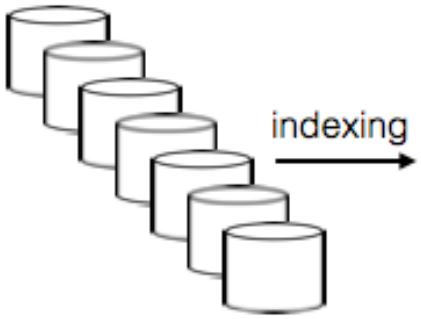


www.biosemantics.org

LUMC - LIACS EURETOS

BioSemantics Knowledge Discovery Pipeline

data sources



The END of Professorware

Euretos Product Solution: BRAIN

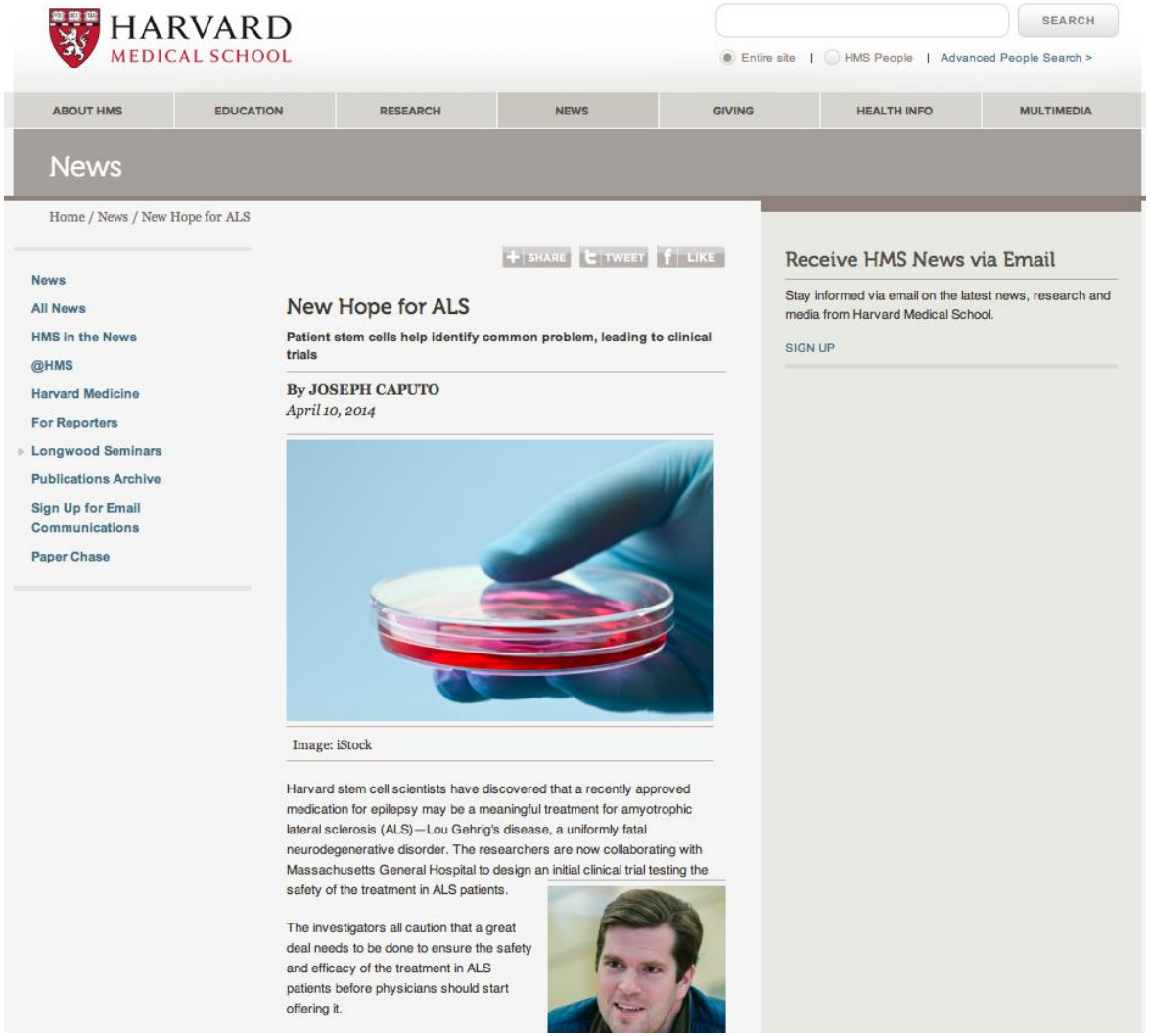
- **Search** 50+ databases at once
- **Discover** indirect relationships
- **Predict** their likeliness
- **Evaluate** evidence instantly
- **Minimize Time to knowledge**

PubMed DrugBank ChEMBL ChemSpider OMIM PubChem uspto.gov ELSEVIER UniProt WikiPathways GWAS CENTRAL

BRAIN

Copyright Euretos b.v. 2014

Knowledge prediction



HARVARD MEDICAL SCHOOL

SEARCH

Entire site | HMS People | Advanced People Search >

ABOUT HMS | EDUCATION | RESEARCH | NEWS | GIVING | HEALTH INFO | MULTIMEDIA

News

Home / News / New Hope for ALS

News
All News
HMS in the News
@HMS
Harvard Medicine
For Reporters
Longwood Seminars
Publications Archive
Sign Up for Email Communications
Paper Chase

SHARE TWEET LIKE

New Hope for ALS

Patient stem cells help identify common problem, leading to clinical trials

By JOSEPH CAPUTO
April 10, 2014

Image: iStock

Harvard stem cell scientists have discovered that a recently approved medication for epilepsy may be a meaningful treatment for amyotrophic lateral sclerosis (ALS)—Lou Gehrig's disease, a uniformly fatal neurodegenerative disorder. The researchers are now collaborating with Massachusetts General Hospital to design an initial clinical trial testing the safety of the treatment in ALS patients.

The investigators all caution that a great deal needs to be done to ensure the safety and efficacy of the treatment in ALS patients before physicians should start offering it.

Receive HMS News via Email

Stay informed via email on the latest news, research and media from Harvard Medical School.

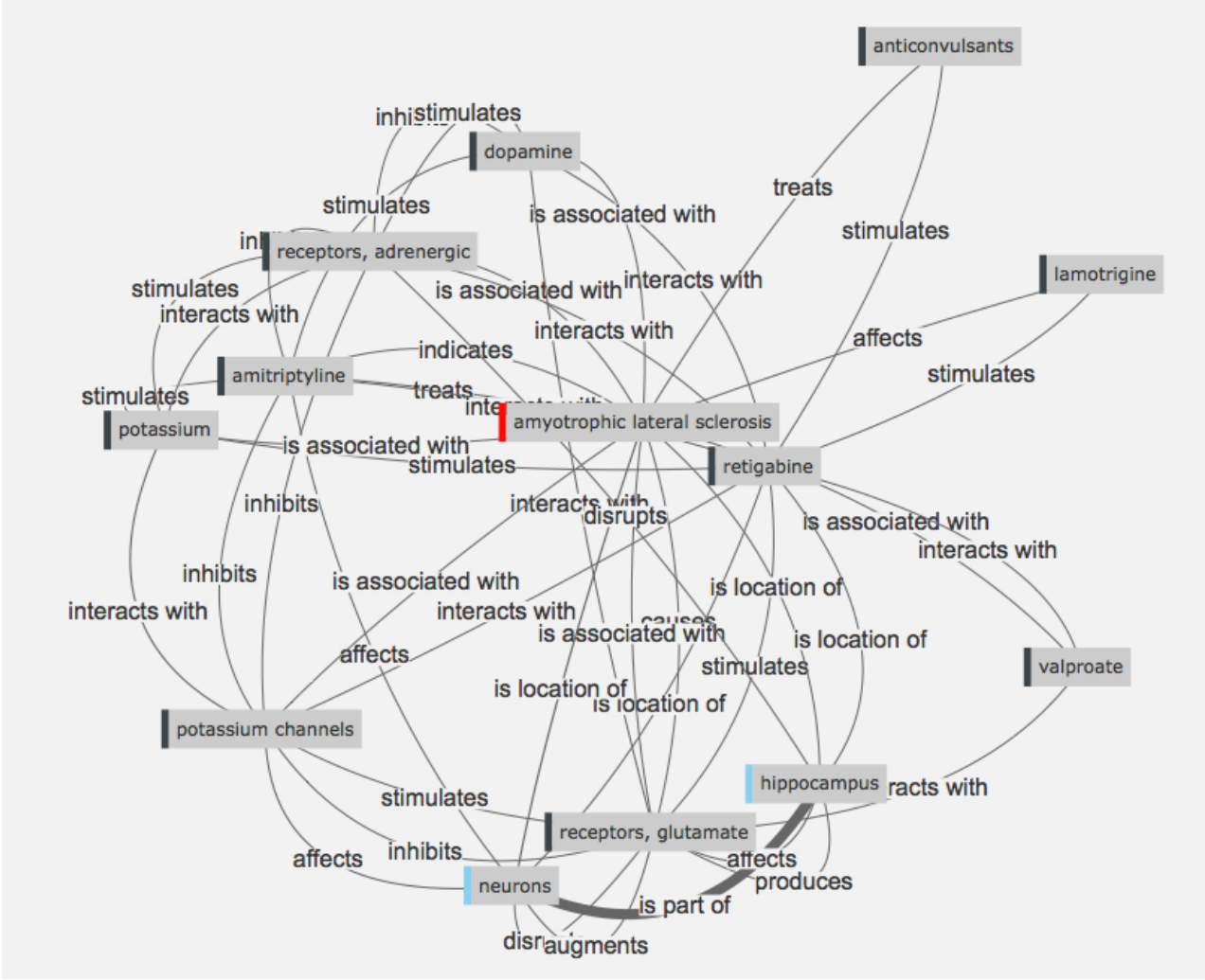
SIGN UP

"Also just saw this article on ALS and Retigabine, a drug that affects Potassium channels and might work for ALS. Would BRAIN have predicted this relationship?

EURETOS

Copyright Euretos b.v. 2014

The power of Self-organising Graph of Cardinal Assertions: BRAIN

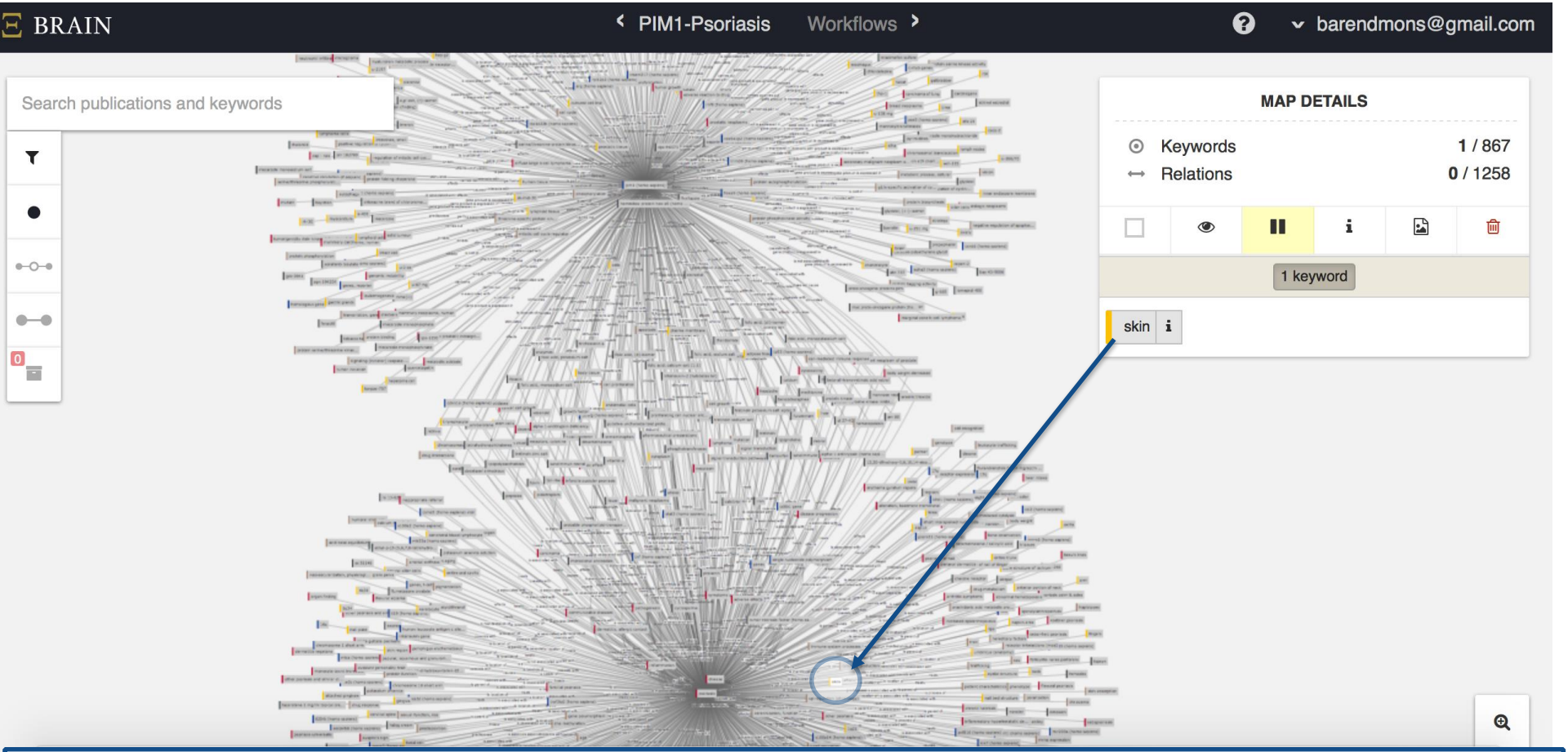


Before this paper was published, BRAIN1.0 contained sufficient indirect relationships between **Retigabine** and **ALS** to predict it as a potential drug.

Discovery that several mutations associated with ALS cause abnormally high activity in **motor neurons** > early degradation. There is a deficit in open **potassium channels** in ALS motor neurons. **Retigabine**, a drug that **opens potassium channels** and approved for human use, seems to **normalize ALS cells *in vitro***, reducing their hyperexcitability > clinical trials in preparation.

Copyright Euretos b.v. 2014

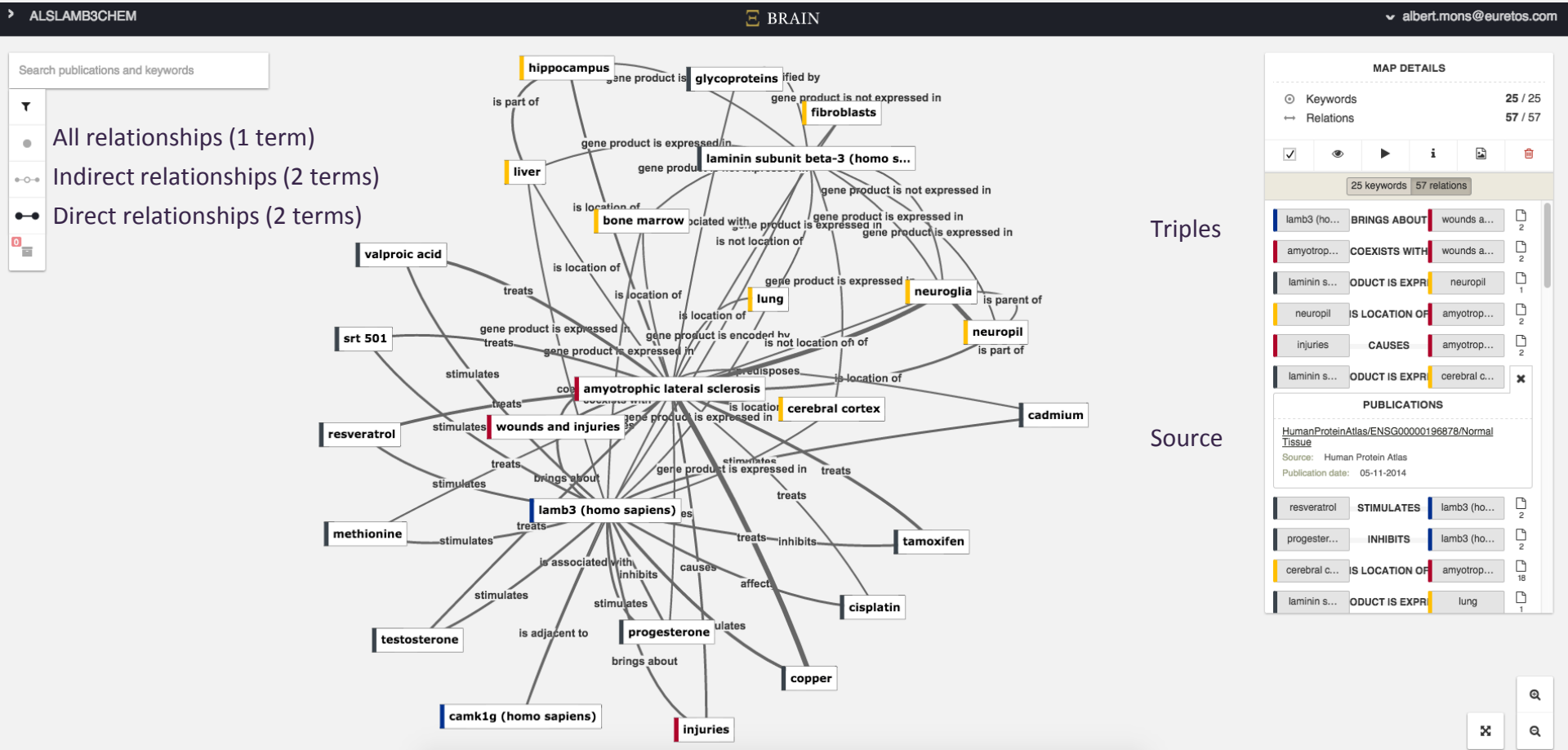
Collaboration DTL/LUMC/EURETOS/EMC



PIM1 <> Psorias pre-publication Frank Nestle expression in Skin (HPA)

The BRAIN Mindmap

Example: validation of relationship ALS and LAMB3



Copyright Euretos b.v. 2014

EURETOS

Collaboration DTL/LUMC/EURETOS/EMC

The Gene Disease Association Study (GDAS) Initiative

One of our key ambitions is to make industrial grade knowledge discovery available to all. We believe knowledge should be shared as early on as and widely as possible. To enable this type of knowledge sharing on genetic diseases we have started a public private initiative, the 'Gene Disease Association Study (GDAS)initiative with the Leiden University Medical Center (LUMC), the Erasmus Medical Centre Rotterdam (EMC) and the Dutch Techcentre for Life Sciences (DTL)

The **Gene Disease Association Study (GDAS)** Initiative

- Top 1% of most likely potential **gene-disease** associations
- **'Crowd' validation** by experts
- **Hundreds of diseases** planned for 2015

GDAS

Reference: GDAS-2015.LAMB3.000002.01

SEED ARTICLE: SUGGESTION FOR FURTHER RESEARCH

Is LAMB3 (homo sapiens) biologically associated with Amyotrophic Lateral Sclerosis (ALS)?

GDAS Initiative[1,2,3,4]

1 Leiden University Medical Centre (LUMC), **2** Erasmus University Medical Centre (EMC), **3** Dutch Techcentre for Life Sciences (DTL), **4** Euretos.

The Gene-Disease Association Studies (GDAS) initiative publishes potential biological mechanisms for gene-disease associations resulting from (1) clinical sequencing, (2) transcriptomics, (3) Genome Wide Association Studies or (4) literature analysis where no explanatory mechanism has been published. The associations derived from literature are based on the Indirect Conceptual Association score (ICA score), a measure developed by the LUMC and the EMC[1]. The potential disease mechanisms are provided by Euretos.

The most promising gene-disease mechanisms, preferably having multiple types of associations, are published for review by experts to assess their biological meaning. By focusing on mechanisms supported by multiple types of associations, the likelihood that an association may be biologically relevant is significantly increased.

ABSTRACT

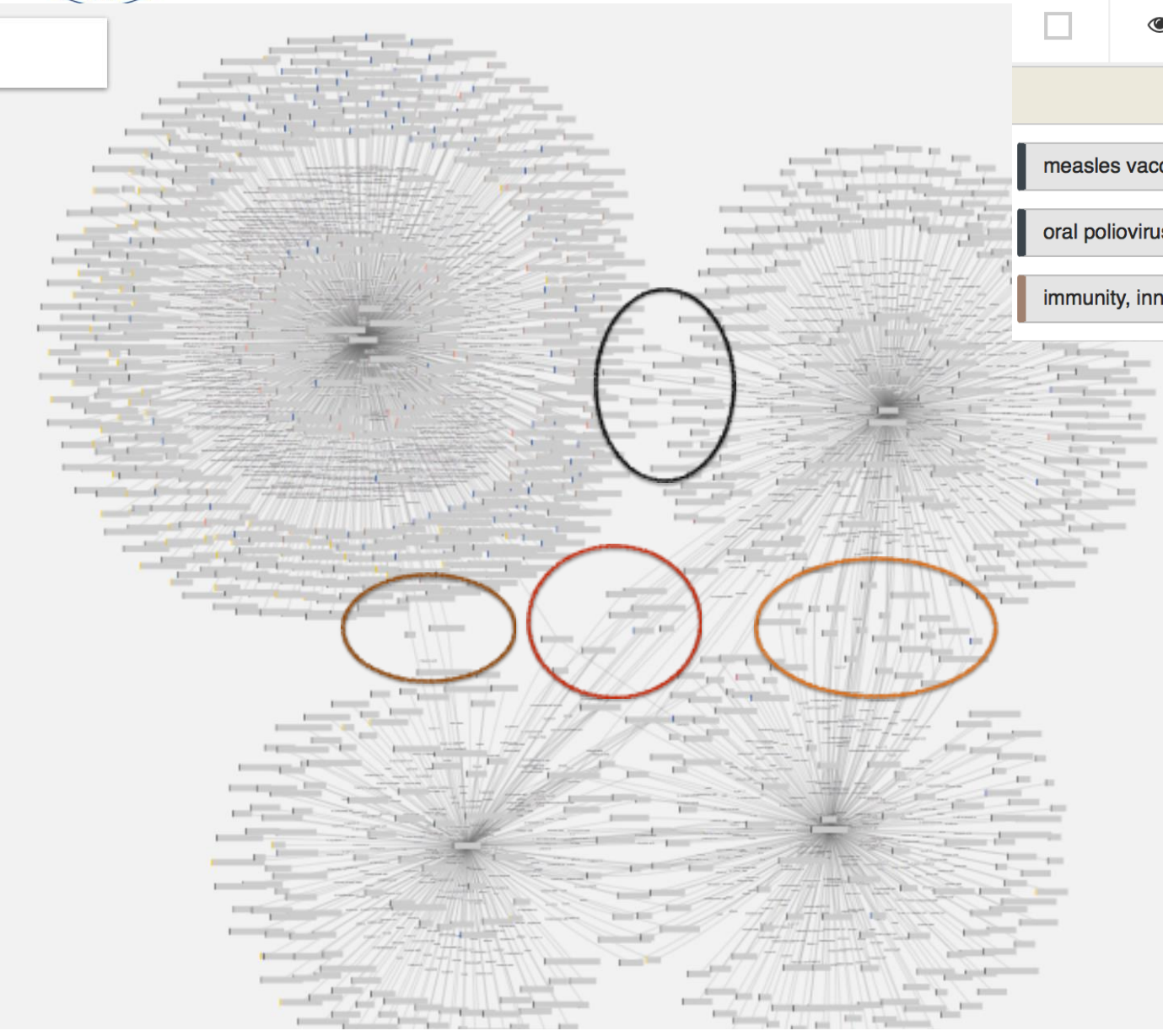
A combination of ICA score[1] and GWAS[2] associations suggests that an interaction between LAMB3 (homo sapiens) may exist. Further analysis using the BRAIN knowledge discovery platform[3] shows that LAMB3 (homo sapiens) has 22 concepts indirectly associated with Amyotrophic Lateral Sclerosis (ALS) via 8 anatomical locations, 2 disorders, 11 biologically active molecules and 1 gene. These relations are supported by 101 references to publications or databases. Based on this we believe assessment by experts for biological meaning is justified.

Type of observation	Value
ICA score	99% percentile
GWAS	8 studies[3]
Clinical Sequencing	0
Transcriptomics	0

The 21 indirect relations that have contributed to the high ICA score

Key-note address:
Bacillus Calmette-Guérin immunotherapy for bladder cancer: overview of an "off-target" effect of BCG immunotherapy

Gary D. Steinberg

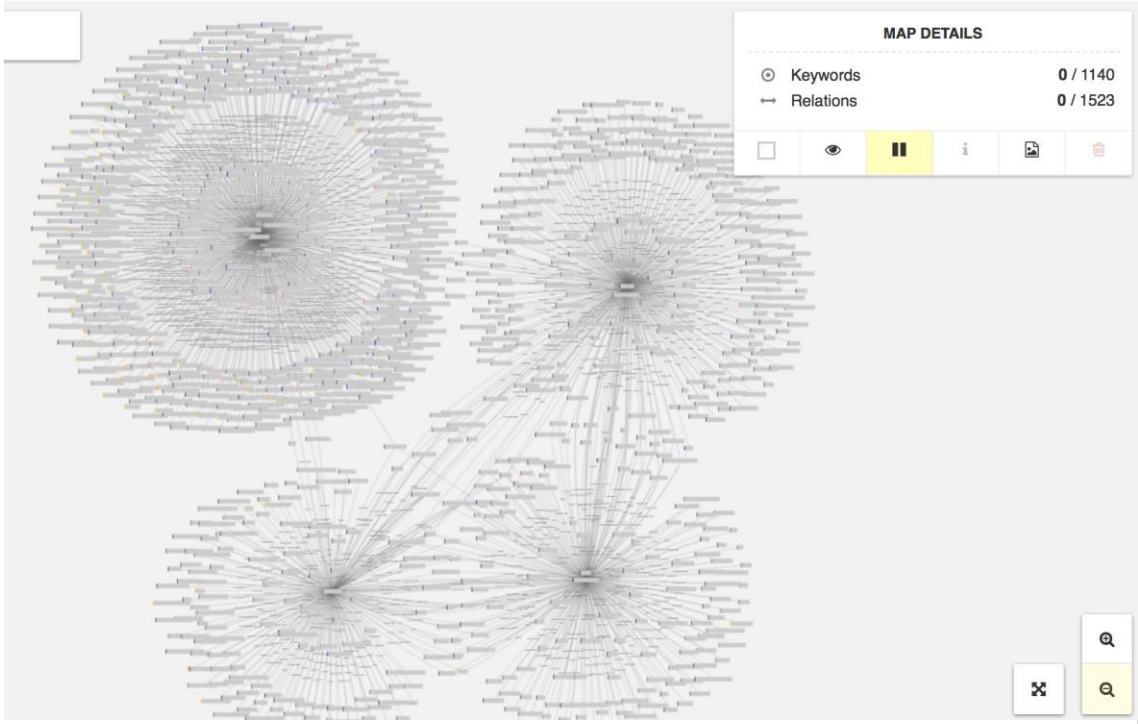


MAP DETAILS

Keywords 4 / 1140
Relations 1 / 1523

4 keywords | 1 relation

- measles vaccine
- bcg vaccine
- oral poliovirus vaccine
- immunity, innate



MAP DETAILS

Keywords **18** / 1140

Relations **1** / 1523

18 keywords | 1 relation

- respiratory tract infections
- vaccines, attenuated
- tumor necrosis factor (homo sapiens)
- butocin
- cytotoxic t-lymphocytes
- cd8b1 gene
- cd4 (homo sapiens)
- immunity, mucosal
- cd46 (homo sapiens)
- liver
- verruca vulgaris
- hodgkin's disease
- internal ribosome entry site
- putative uncharacterized protein encode
- latent infection
- lung
- intestines
- immunity, innate

BRAIN

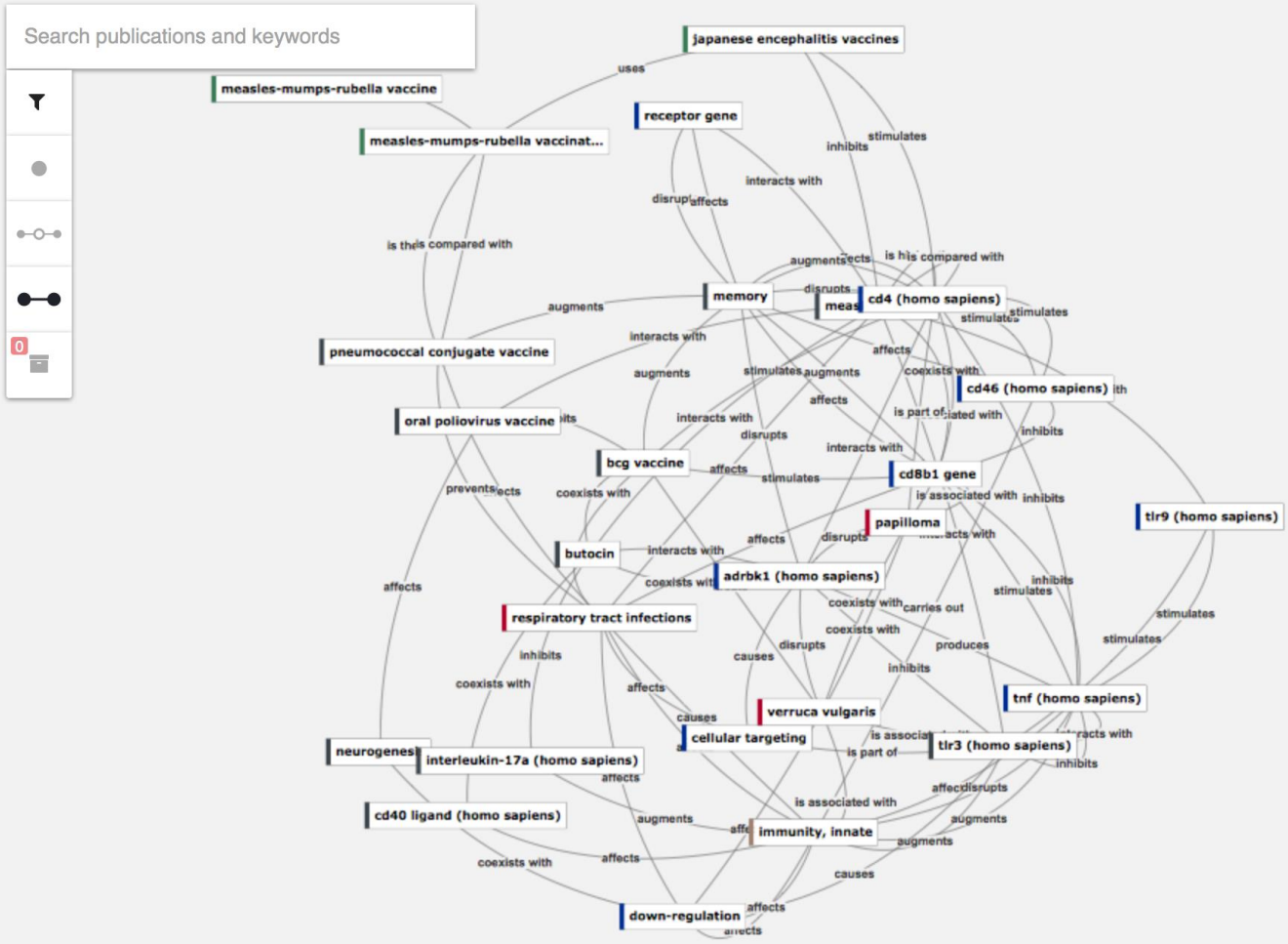
Live vaccines and off-target effects: BCG and measles vaccines

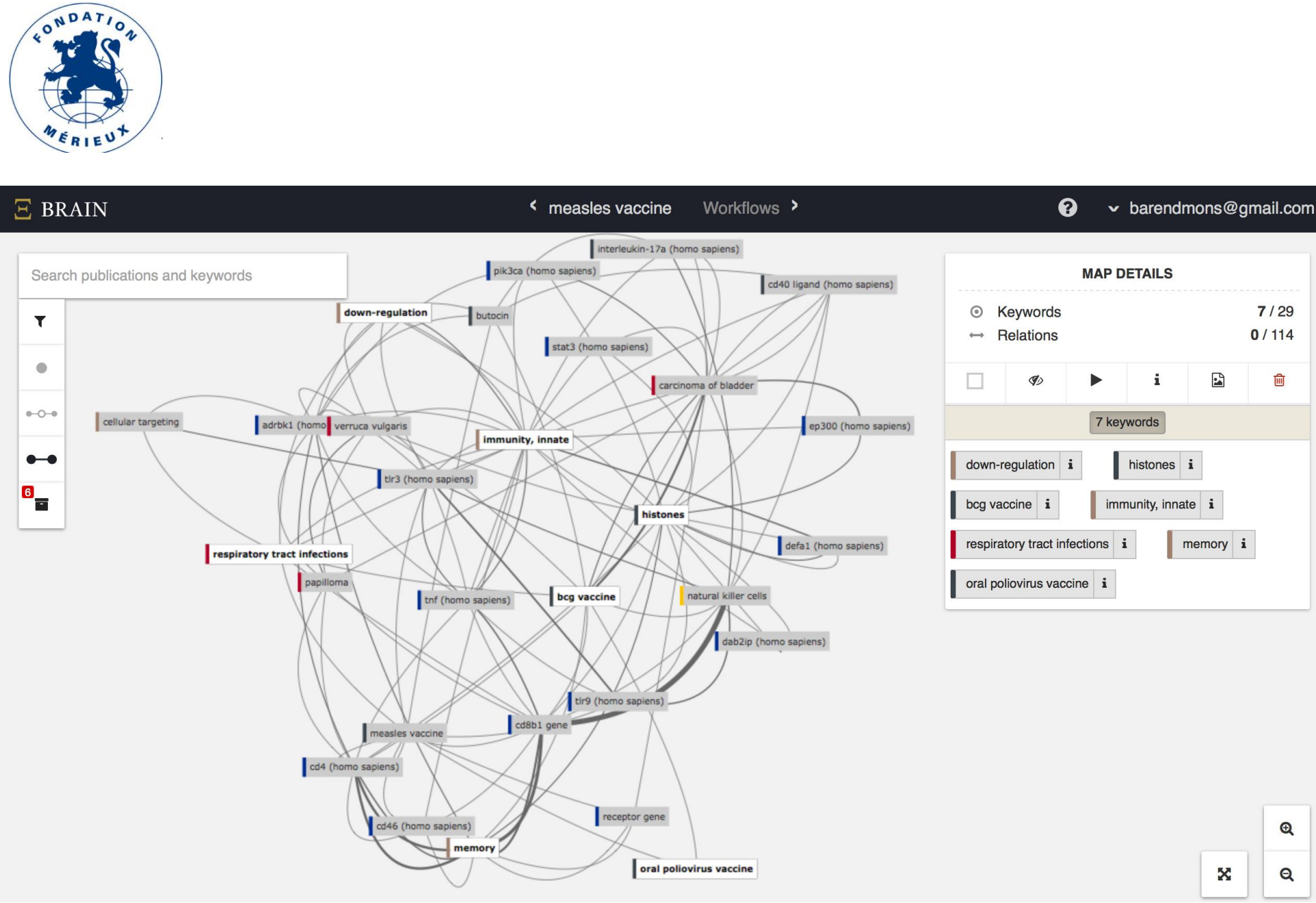
Peter Aaby

Discussion

Inactivated vaccines and off-target effects: DTP and interaction between live and inactivated vaccines

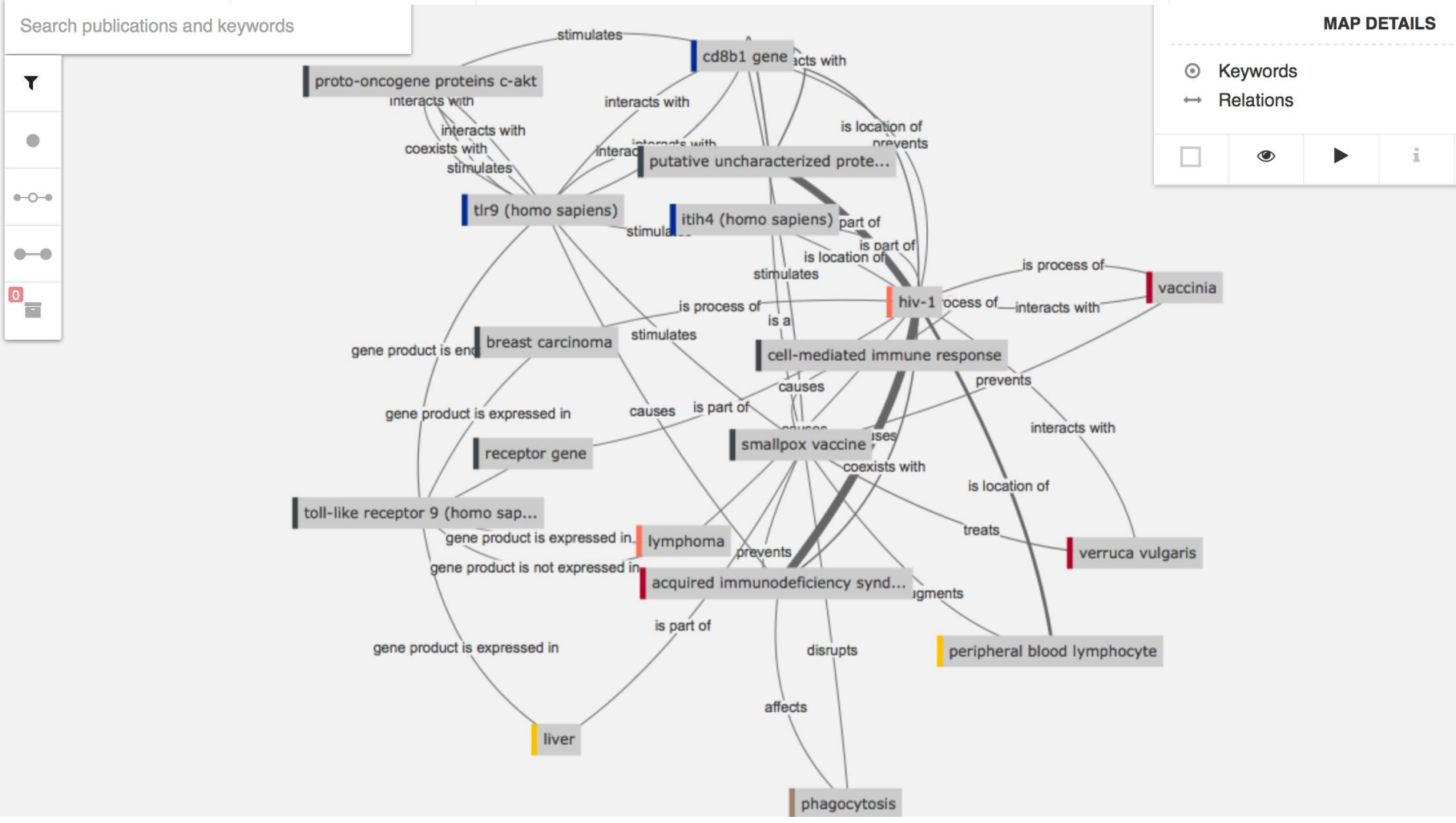
Henrik Ravn




BRAIN
measles vaccine
Workflows
barendmons@gmail.com
Search publications and keywords
interleukin-17a (homo sapiens)
pik3ca (homo sapiens)
cd40 ligand (homo sapiens)
down-regulation
butocin
stat3 (homo sapiens)
carcinoma of bladder
cellular targeting
adrbk1 (homo
verruca vulgaris
immunity, innate
ep300 (homo sapiens)
tlr3 (homo sapiens)
histones
defa1 (homo sapiens)
respiratory tract infections
papilloma
tnf (homo sapiens)
bcg vaccine
natural killer cells
dab2ip (homo sapiens)
tlr9 (homo sapiens)
measles vaccine
cd8b1 gene
cd4 (homo sapiens)
receptor gene
cd46 (homo sapiens)
memory
oral poliovirus vaccine
MAP DETAILS
Keywords 7 / 29
Relations 0 / 114
7 keywords
down-regulation
histones
bcg vaccine
immunity, innate
respiratory tract infections
memory
oral poliovirus vaccine

16:45 - 17:05	Implications of beneficial off-target effects for upcoming eradication campains (measles and polio)	Ane Fisker
17:05 - 17:20	Discussion	
17:20 - 17:40	Vaccinations against smallpox and tuberculosis are associated with better long-term survival	Christine Stabell-Benn

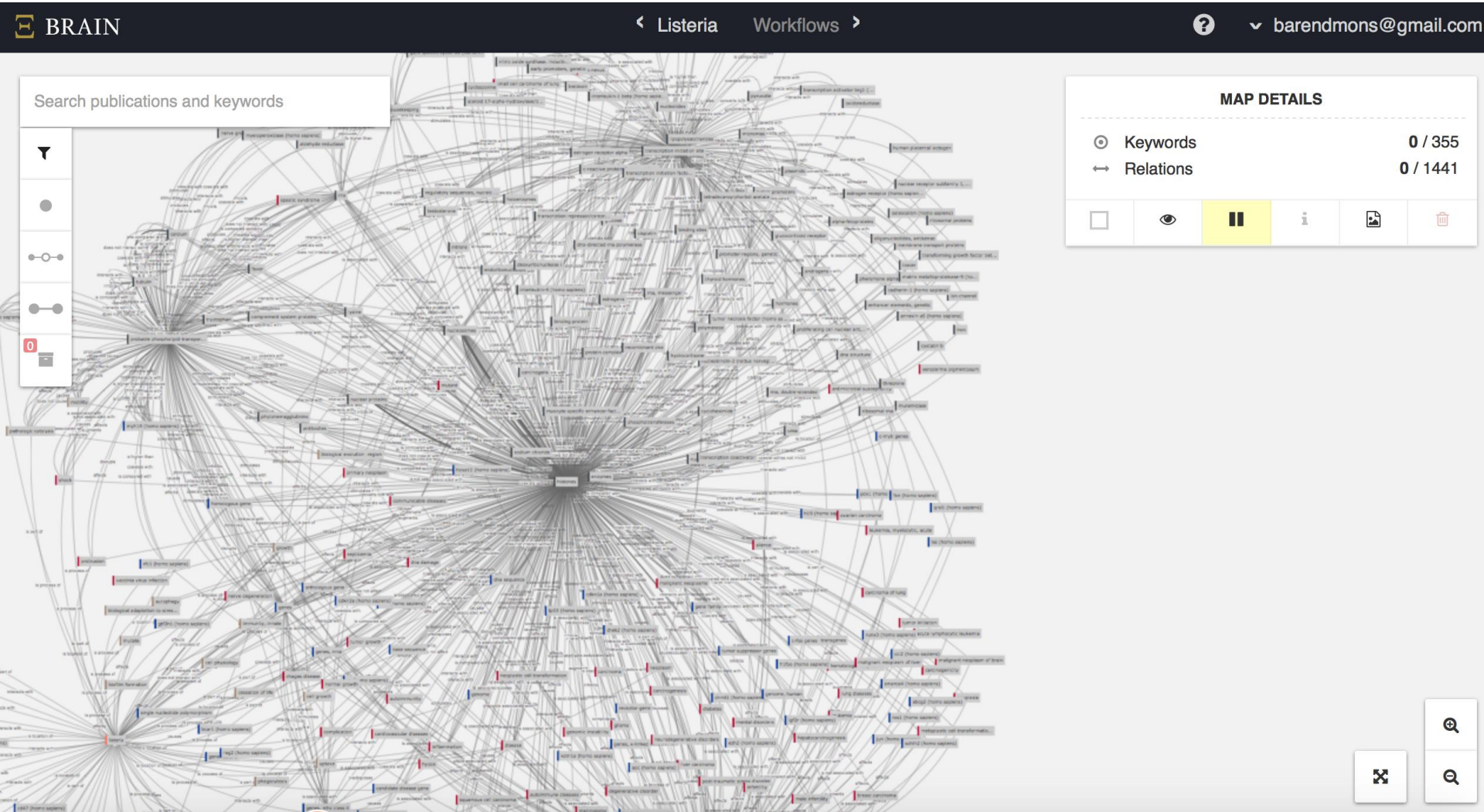
BRAIN



Basic biological mechanisms accounting for the off-target effects of vaccination

Chair: William Warren

11:20 - 11:40	Epigenetic reprogramming by listeria monocytogenes	Melanie Hamon



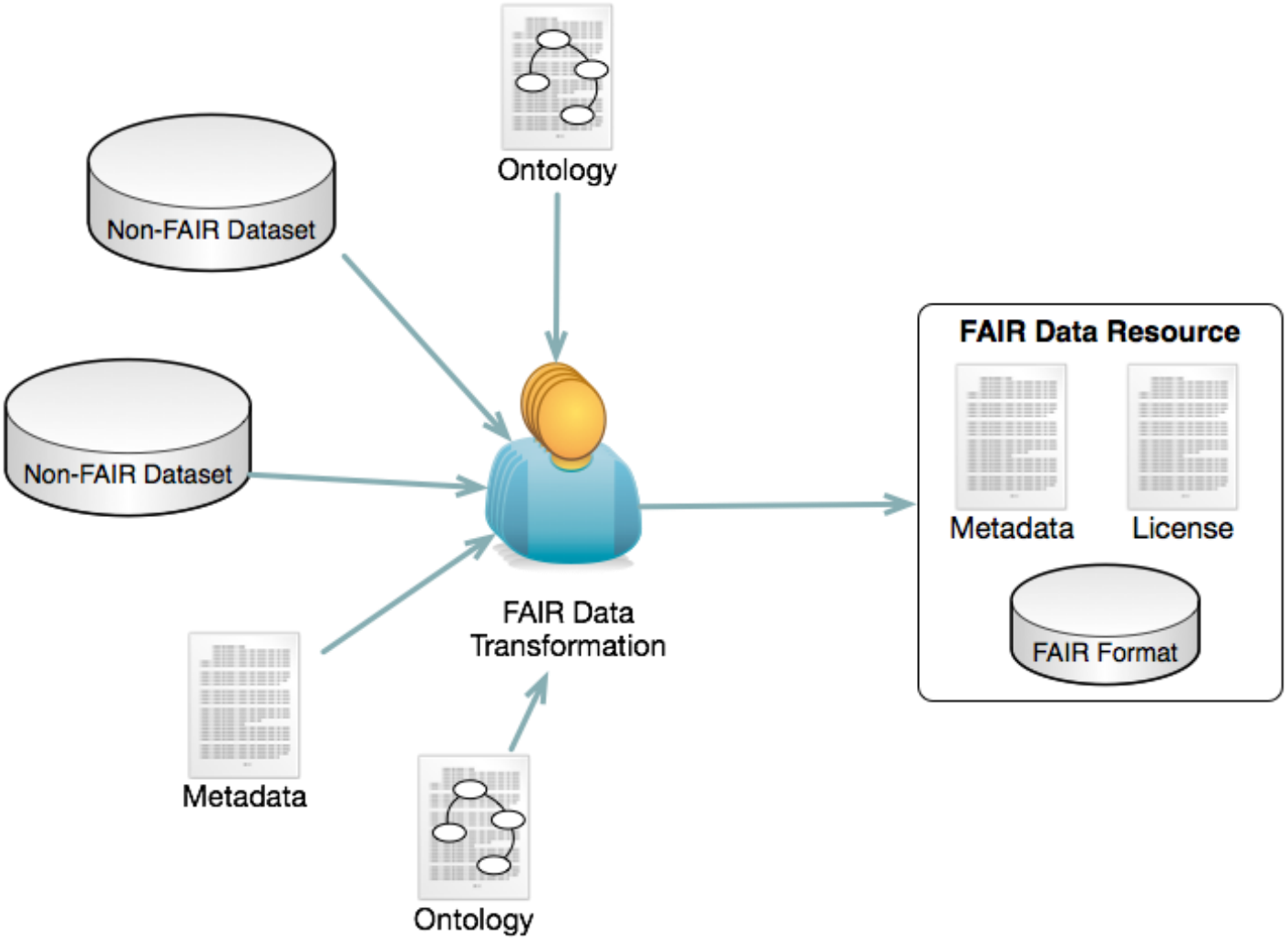
Of

Humans

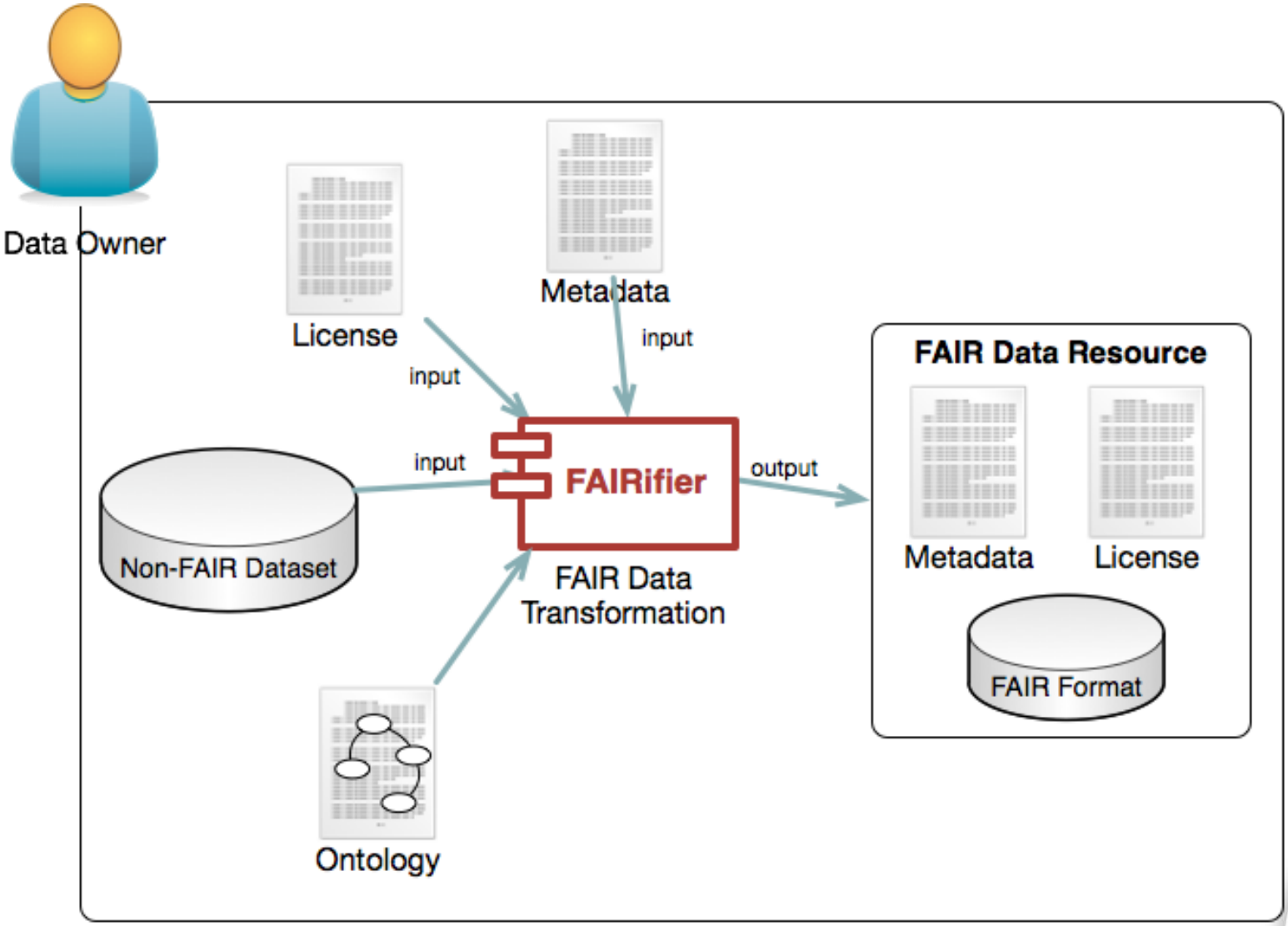
and

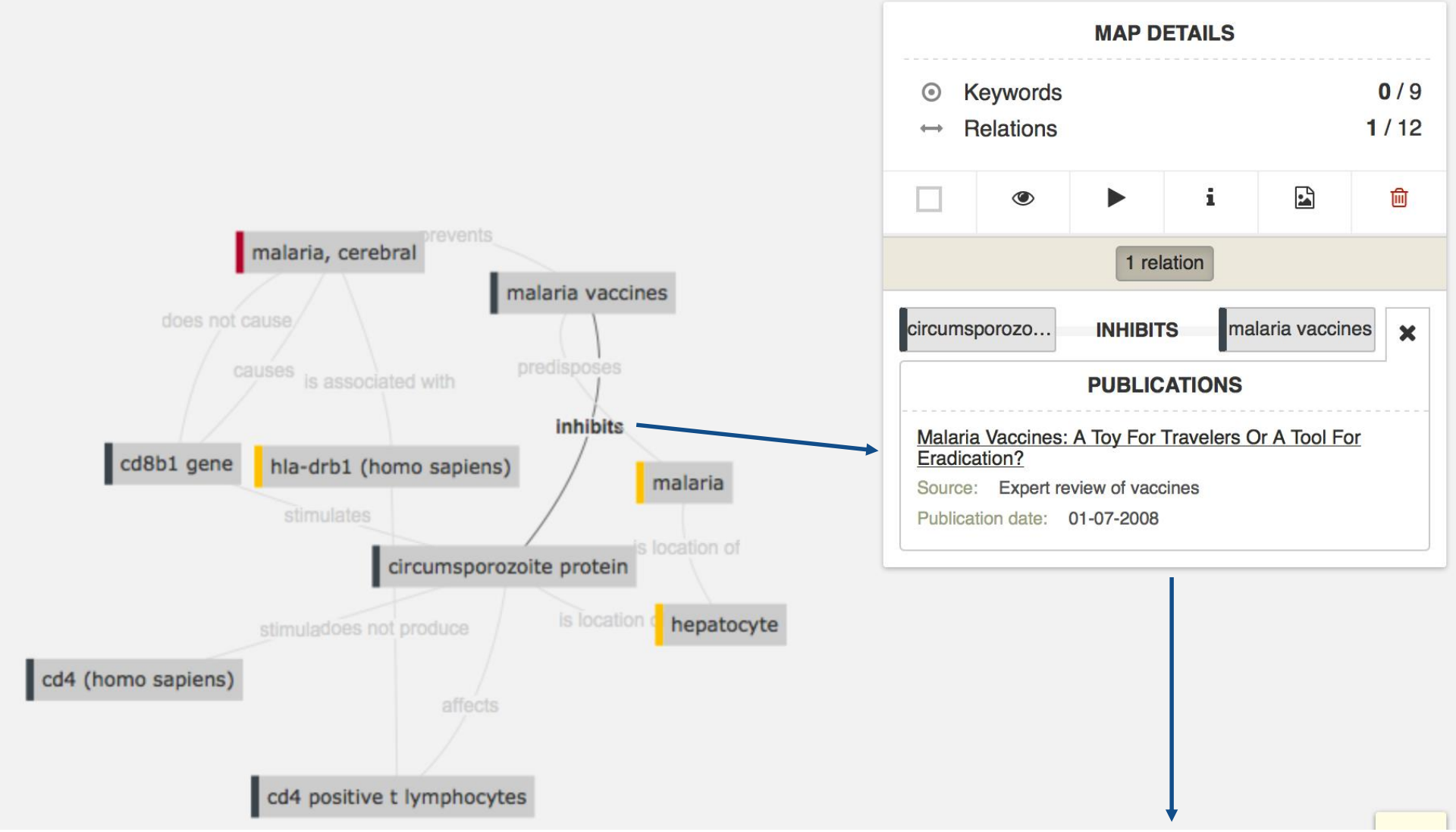
Machines

BRING YOUR OWN DATA (BYOD)



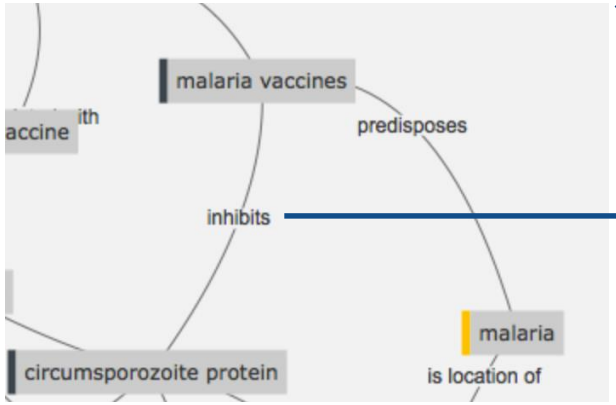
FAIRIFIER

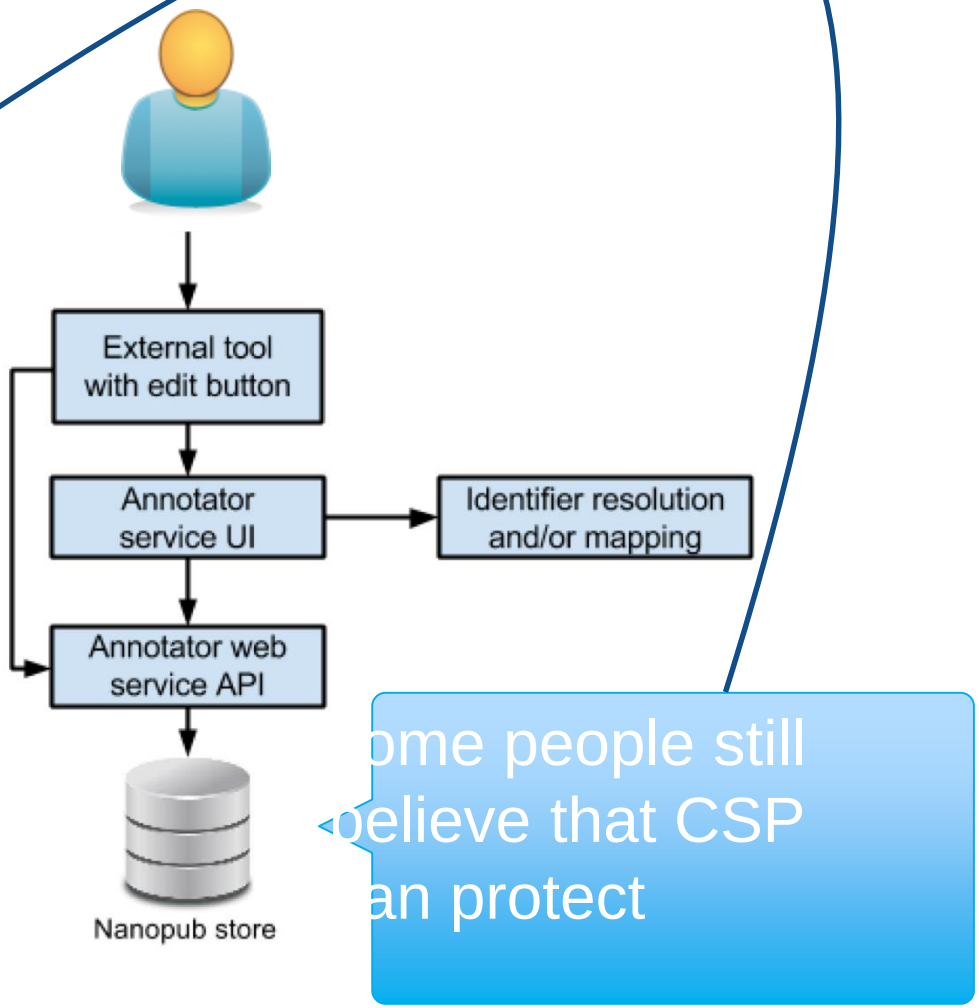

Data Owner
License
Metadata
input
input
input
FAIRifier
output
Non-FAIR Dataset
FAIR Data
Transformation
Ontology
FAIR Data Resource
Metadata
License
FAIR Format



Only long-term travellers, expatriates and soldiers might realistica

Don’t panic….annotation is coming





“Bring your own data” - Hands on interoperability

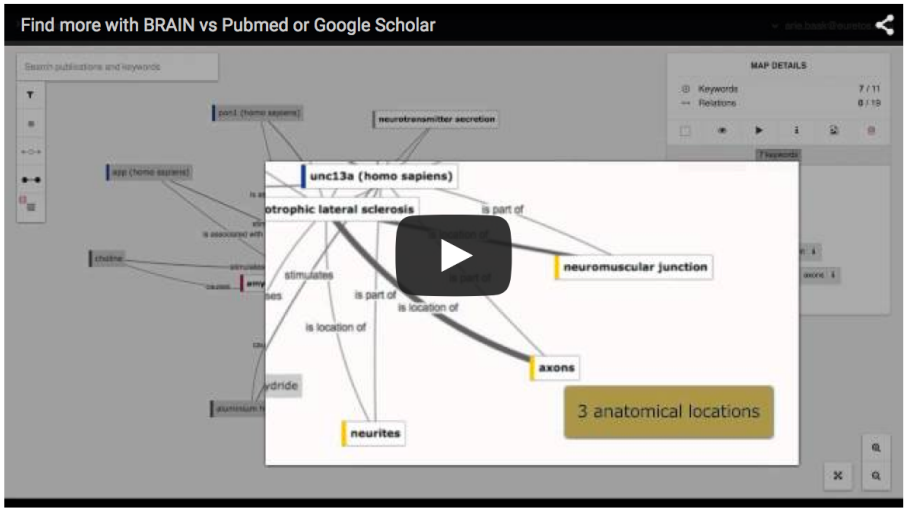
- Problem-centered workshops
- Integration experts - Data resources –Users
- ELIXIR funds external trainers
- Great feedback –turning into programme

http://www.euretos.com/brain

Home BRAIN News GDAS Euretos Contact

account.

Get Started
Click here to use BRAIN



Prof. David Webb (Scripps Research Institute)

BRAIN is nothing less than an in silico way to save all drug and biotech researchers an infinite amount of blood, sweat and tears by providing them with a very powerful knowledge discovery platform.

BRAIN allows the researcher to rapidly assess whether a given target, research area, etc., is related, even distantly, to another such entity thus allowing one to uncover hidden connections, leading to new knowledge. In the drug discovery arena, this thing is a **gold mine**