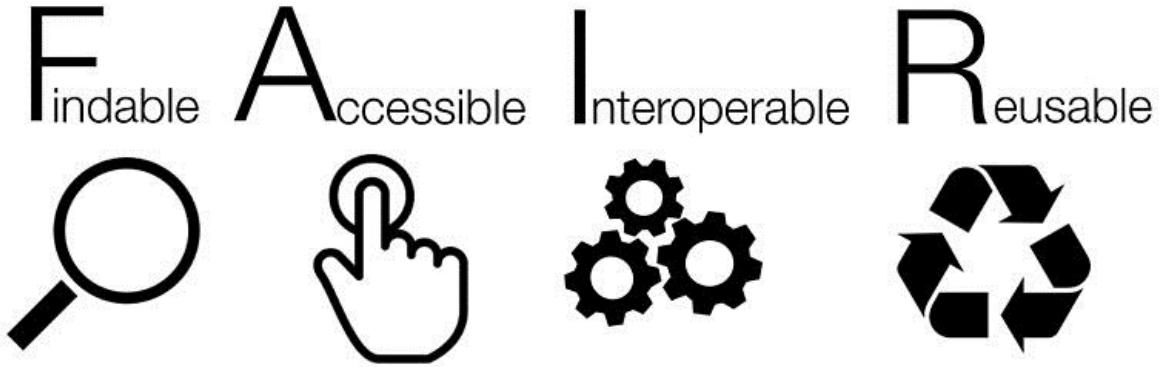
Disclosure belangen spreker

(potentiële) belangenverstrengeling	Zie hieronder
Voor bijeenkomst mogelijk relevante relaties met bedrijven	Ontoforcen NV Data2Discovery Inc
• Sponsoring of onderzoeksgeld • Honorarium of andere (financiële) vergoeding • Aandeelhouder • Andere relatie, namelijk …	Shareholder

Accelerating Drug Discovery with an Internet of FAIR Data and Services



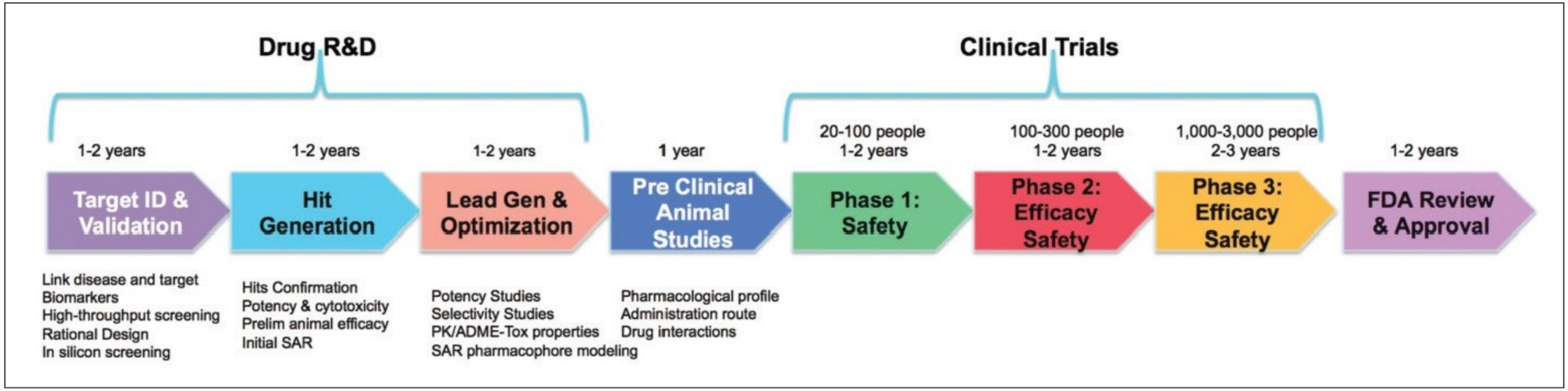
Michel Dumontier, Ph.D.
Distinguished Professor of Data Science
Director, Institute of Data Science

Maastricht University

Drug development time consuming

- ~10-20 years to develop a new drug

Drug discovery pipeline



The drug discovery value chain. Estimates of timings and resources based on classical small molecule drug discovery.

Source: https://steveblank.files.wordpress.com/2013/08/drug-discovery-pipeline.jpg

Drug development
time consuming and expensive

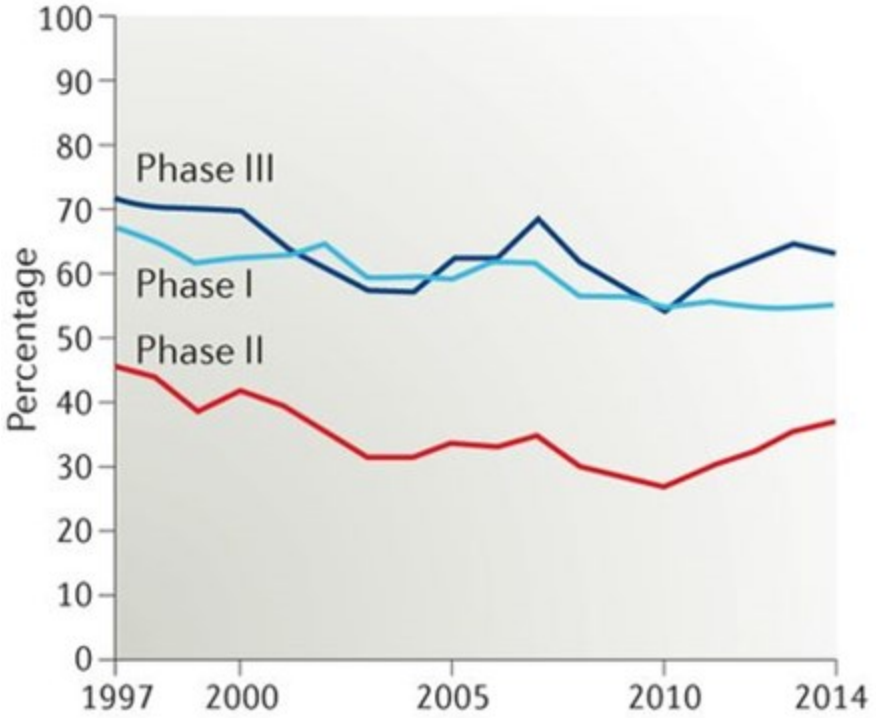
- ~15-20 years to develop a new drug
- Reporting of cost between $1.2B-$2.4B
  - 70% costs in clinical trials.
  - 75% of this price tag is attributable to failure

THE CLINICAL-TRIAL CLIFF

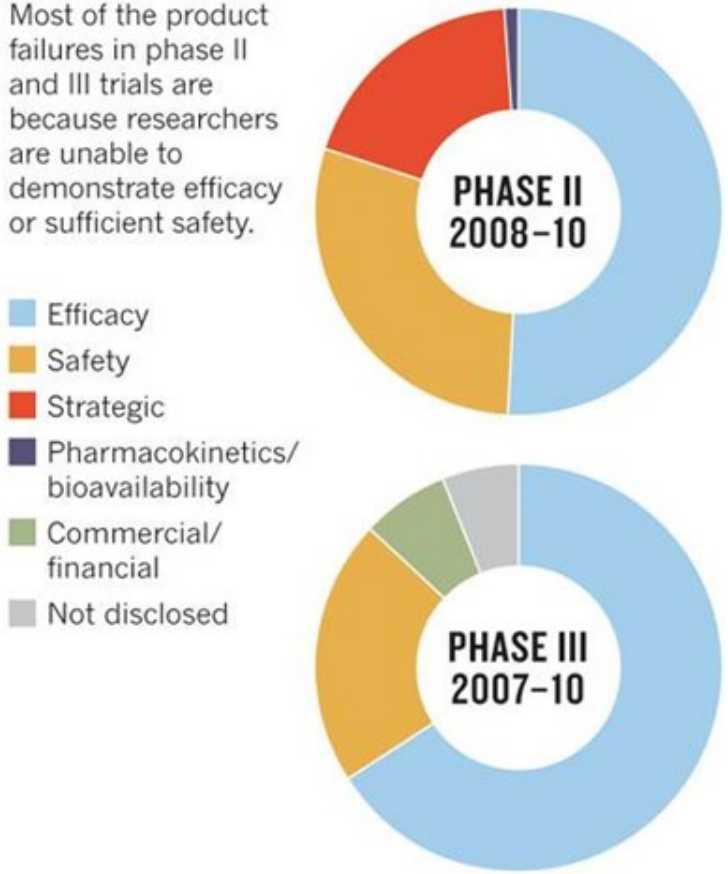
Drug companies are removing more compounds from the pipeline at all levels of testing than ever before.

Success rates by phase

Percentage likelihood of moving to next phase, 3-year rolling average*

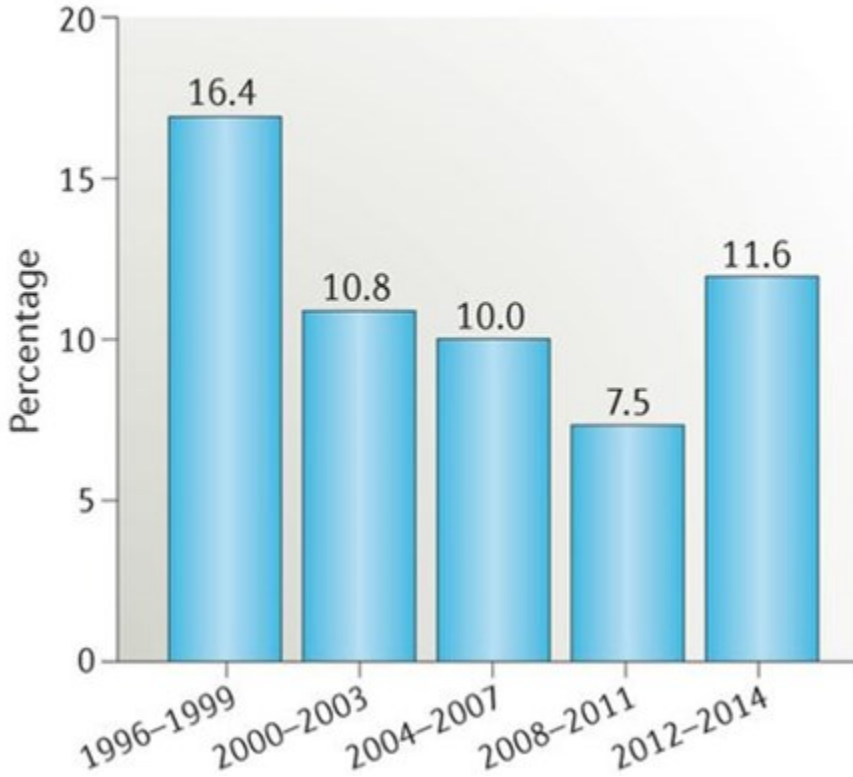


Most of the product failures in phase II and III trials are because researchers are unable to demonstrate efficacy or sufficient safety.



Cumulative success rate Phase I to launch

Percentage likelihood of moving from Phase I to launch



Nature Reviews | Drug Discovery

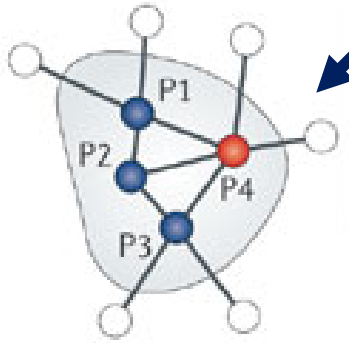
Drug development
time consuming and expensive and prone to failure

- ~15-20 years to develop a new drug
- Reporting of cost between $1.2B-$2.4B
  - 70% costs in clinical trials.
  - 75% of cost attributable to failure
    - 90% of all drugs developed don't make it to market
    - 60% transition from phase i, 35% make phase ii, 62% make phase III
  - Successful drugs - $500M-$2B

Massive data generating methods now operating across the discovery pipeline

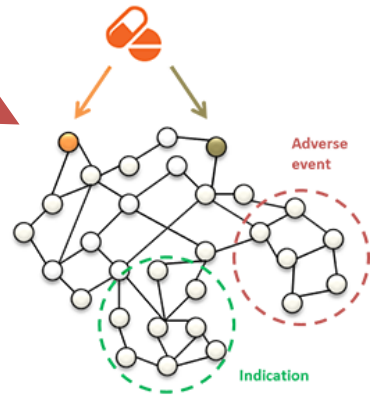
OMICs - Genomics, Transcriptomics, Proteomics, Metabolomics

Identify a wide range of relevant and even personalized targets



High Throughput Screening Screen 100k+ compounds/day

In silico profiling Use computers to predict activity

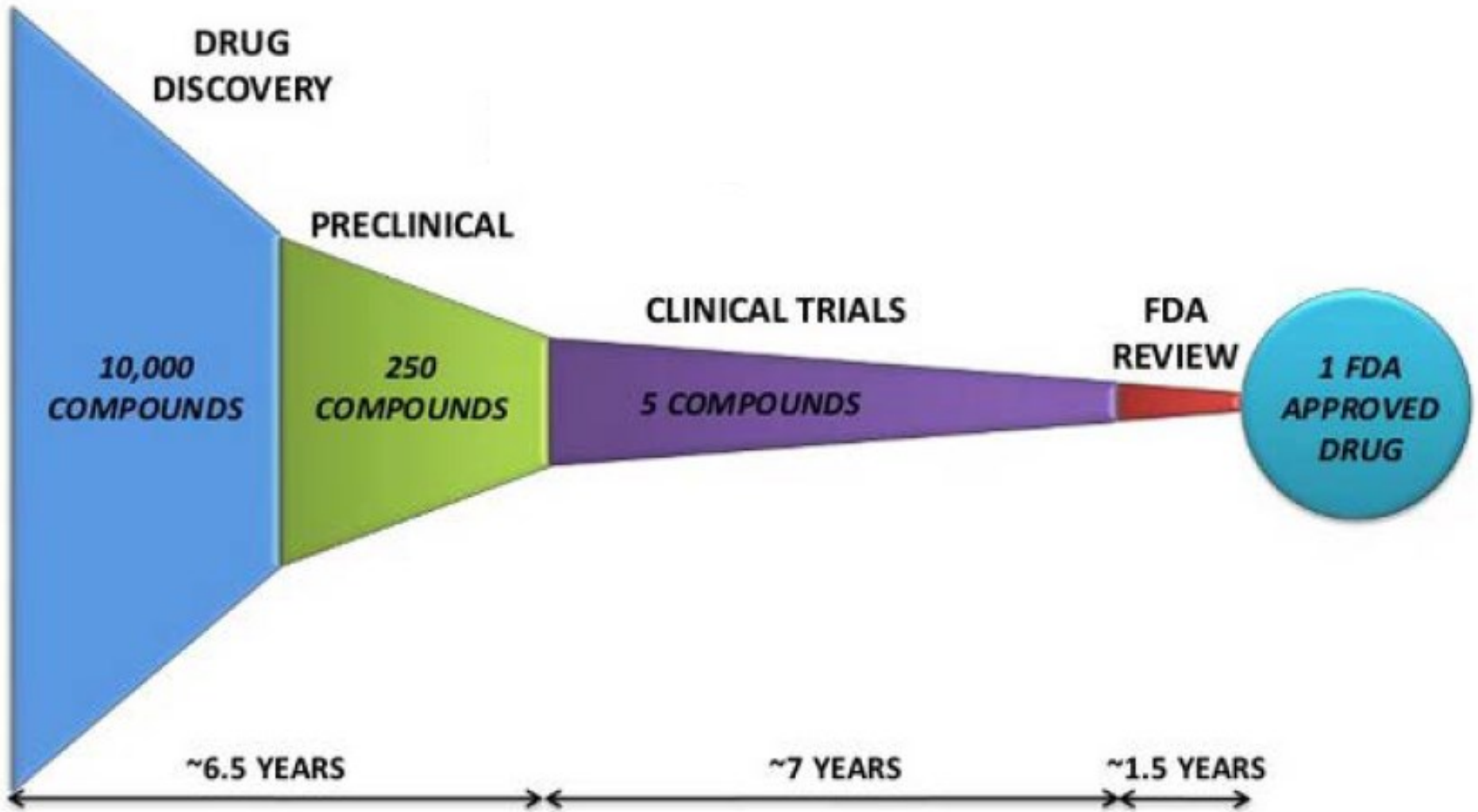


drug candidates

Molecular modeling
Optimize structure and activity

In vitro and in silico ADME models
Assess biological feasibility

preclinical testing


DRUG DISCOVERY
PRECLINICAL
CLINICAL TRIALS
FDA REVIEW
10,000 COMPOUNDS
250 COMPOUNDS
5 COMPOUNDS
1 FDA APPROVED DRUG
~6.5 YEARS
~7 YEARS
~1.5 YEARS

Bottom Line

Despite our collective knowledge, experience, improved technology and vast amounts of data…

We still don’t know what we are doing.

Moreover, it’s unethical do subject people to chronic failure.

Most published research findings are false.

- John Ioannidis, Stanford University

PLoS Med 2005;2(8): e124.

Reproducibility of landmark studies is shockingly low:

39% (39/100) in psychology[1]
21% (14/67) in pharmacology[2]
11% (6/53) in cancer[3]

[1]doi:10.1038/nature.2015.17433 [2]doi:10.1038/nrd3439-c1 [3]doi:10.1038/483531a

Maybe it’s time to completely rethink how we perform drug discovery

Maybe it's time to rethink how we perform drug discovery

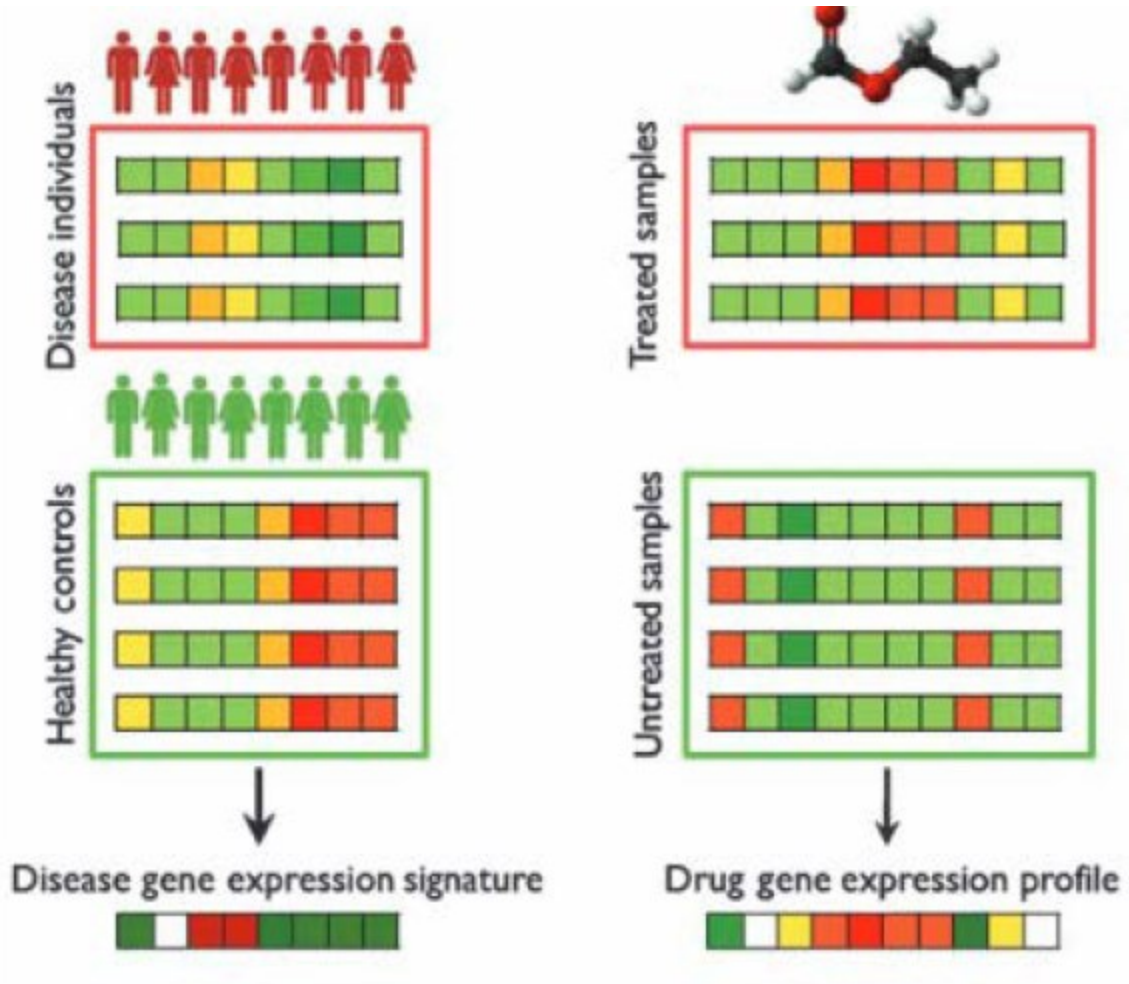
We need to use **machines** to improve our ability to identify a successful drug candidate **as early as possible** by aggregating and analyzing **data across the drug discovery pipeline**

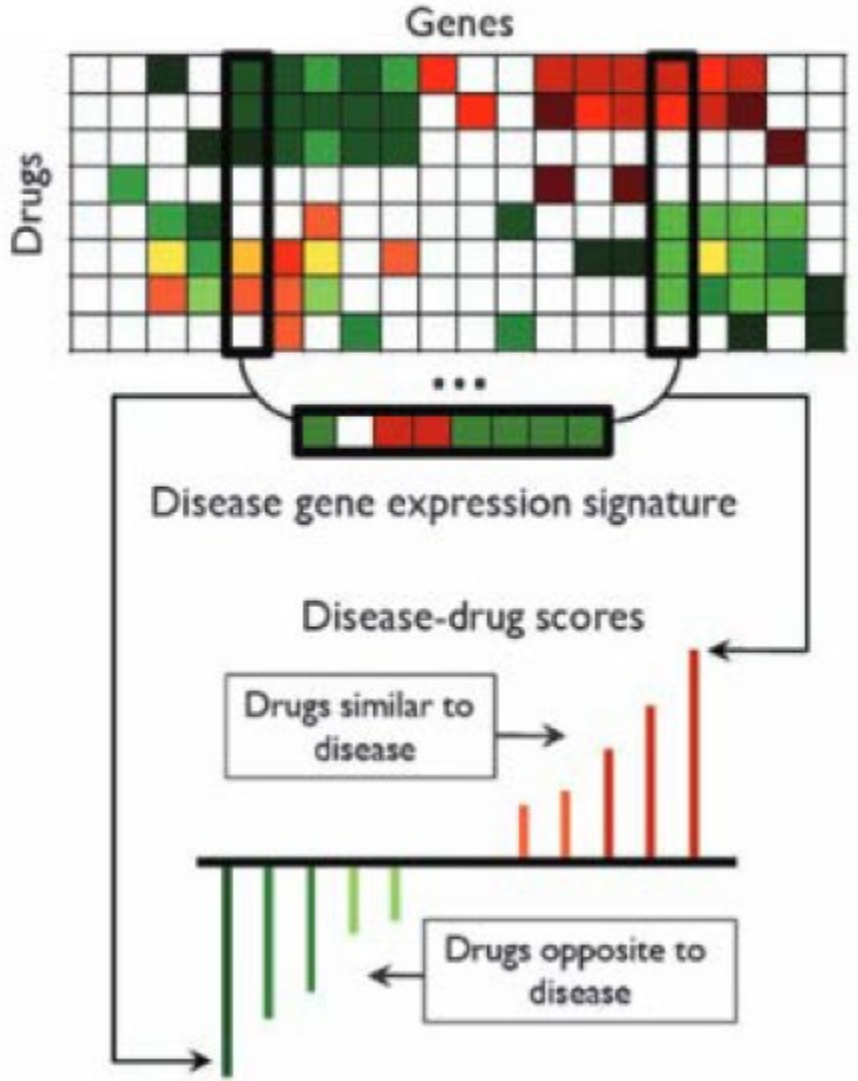
Discovery and Preclinical Validation of Drug Indications Using Compendia of Public Gene Expression Data

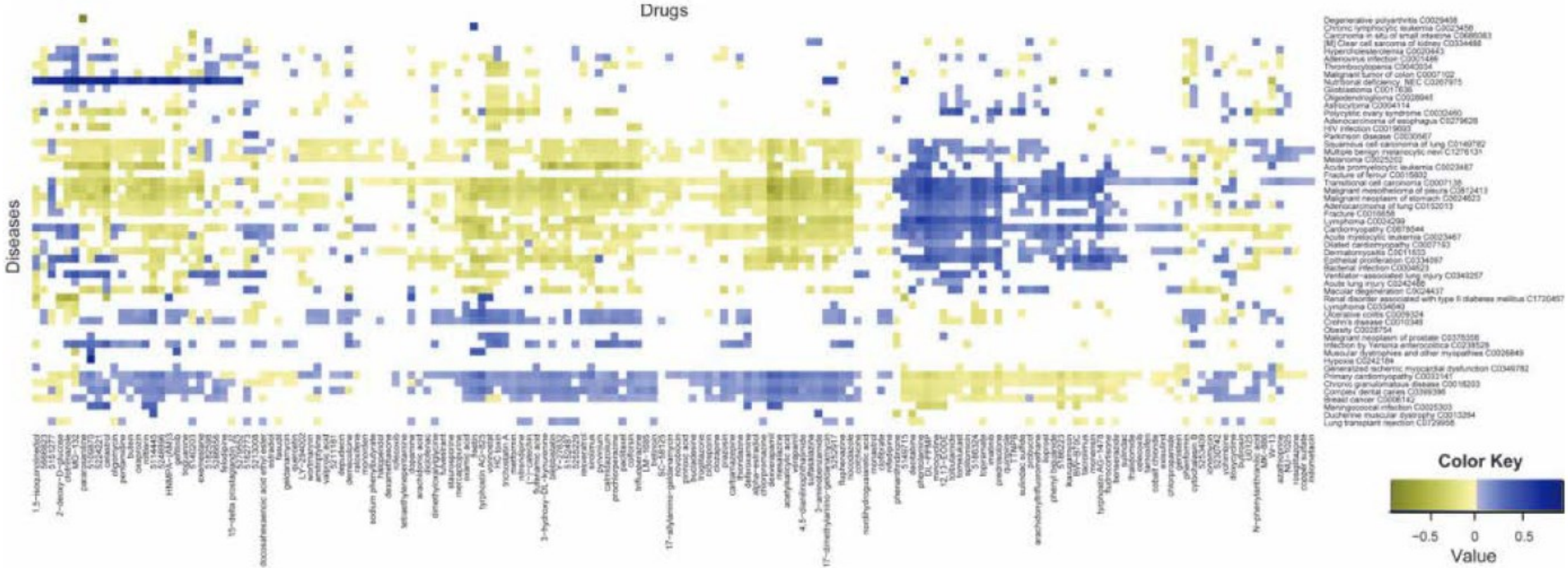
Marina Sirota,[1,2,3]* Joel T. Dudley,[1,2,3]* Jeewon Kim,[4] Annie P. Chiang,[1,2,3] Alex A. Morgan,[1,2,3] Alejandro Sweet-Cordero,[1,5] Julien Sage,[1,5,6] Atul J. Butte[1,3,5]†

Published 17 August 2011; revised 28 September 2011

ScienceTranslationalMedicine. Vol 3 Issue 96

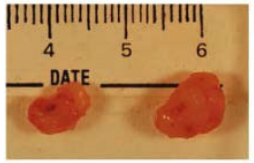





Drugs
Diseases
Color Key
−0.5
0
0.5
Value

Evaluation of cimetidine for lung adenocarcinoma

- Lung cancer contributes the greatest burden of cancer mortality and incidence in Europe and the United States
- Evaluated therapeutic predictions for lung adenocarcinoma
  - cimetidine - off-patent and inexpensive OTC to treat heartburn and ulcers with favorable side effect profile
  - tested three doses of cimetidine in mice implanted with a human A549 LA cell line and followed the growth of these tumors for 12 days
  - 2.3x growth of original compared to 2.8x for control



2012 paper suggests utility as adjuvant in colorectal cancer

Inflammatory bowel diseases

- Inflammatory bowel diseases (IBDs) ulcerative colitis (UC) and Crohn's disease (CD) form cluster for which corticosteroids and other immunosuppressive drugs are broadly indicated.
- **Topiramate**, an anticonvulsant drug currently used to treat epilepsy, had a strong therapeutic score for CD and UC
- topiramate performed favorably in a preclinical rodent colitis model

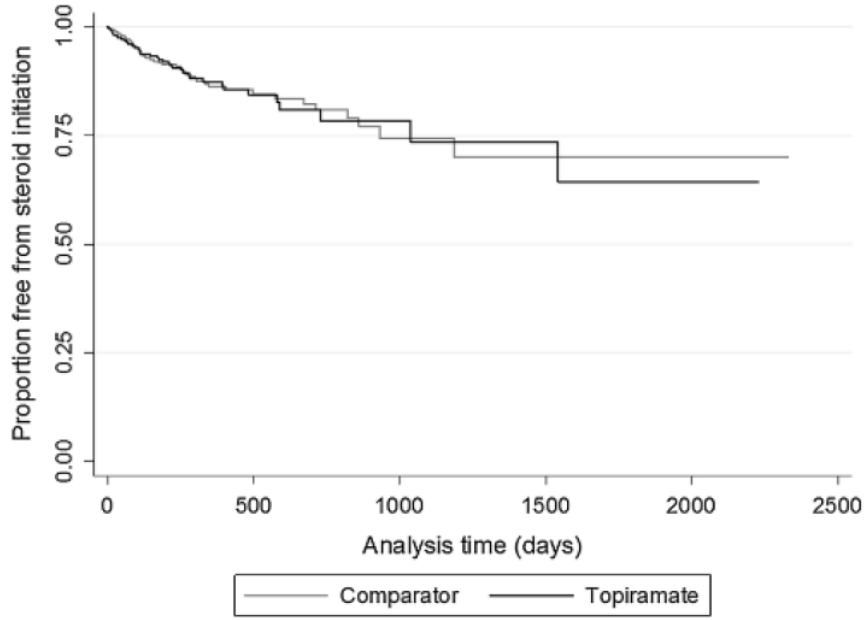
Sci Transl Med. 2011 Aug 17;3(96):96ra76.
doi: 10.1126/scitranslmed.3002648

Digestive Diseases and Sciences
July 2014, Volume 59, Issue 7, pp 1535-1543

First online: 07 February 2014

Topiramate Use Does Not Reduce Flares of Inflammatory Bowel Disease

Seth D. Crockett ✉ , Robin Schectman, Til Stürmer, Michael D. Kappelman



pharmacoepidemiology study using **<u>administrative claims data</u>** to determine whether topiramate exposure is associated with a reduced rate of disease flares in subjects with IBD.

Maybe it's time to rethink how we perform drug discovery

We need to use machines to improve our ability to identify a successful drug candidate as early as possible by aggregating and analyzing **a complementary set of positive and negative findings** across the drug discovery pipeline

MENU

nature
International journal of science

Subscr

SPOTLIGHT · 30 MAY 2018

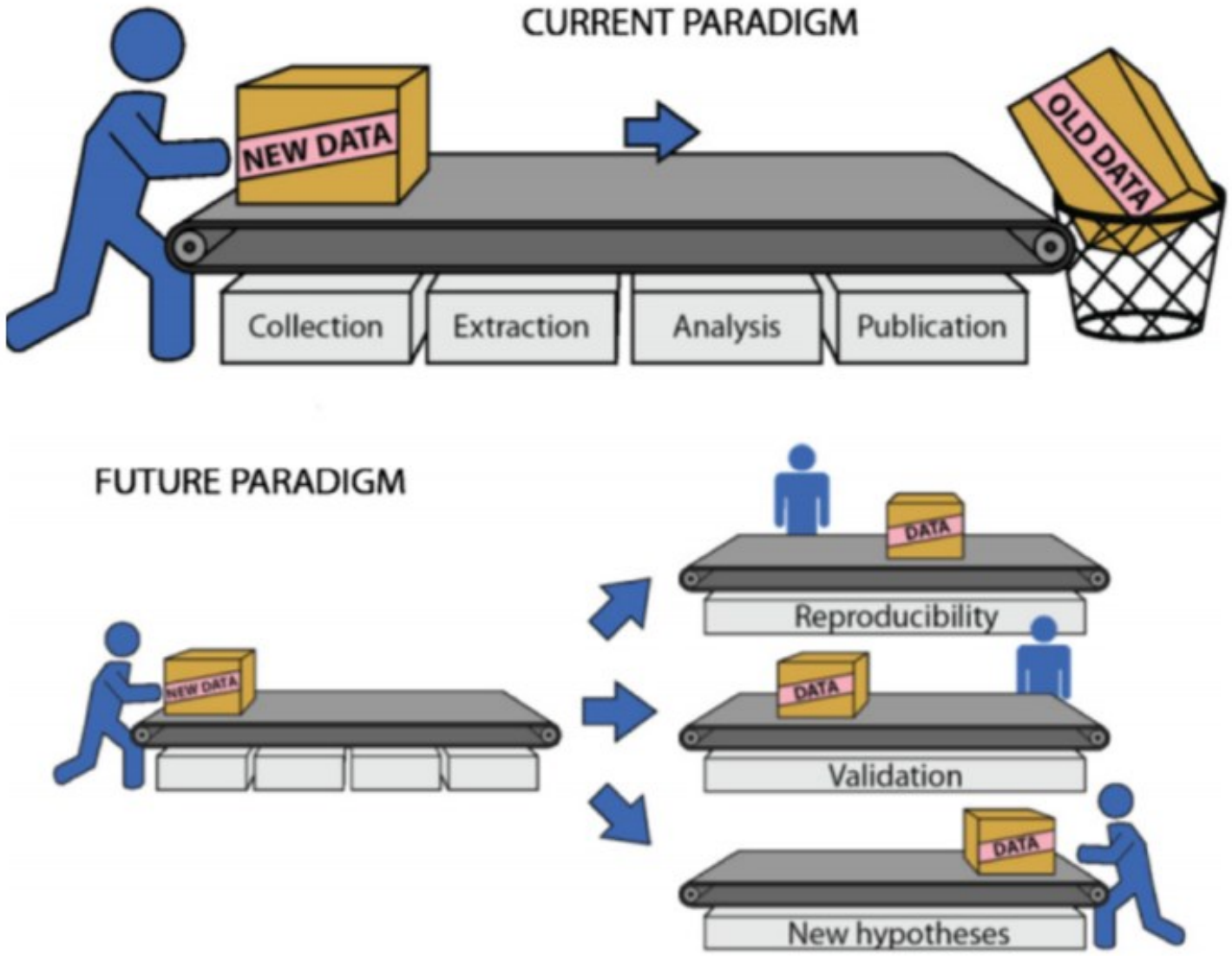
How artificial intelligence is changing drug discovery

Machine learning and other technologies are expected to make the hunt for new pharmaceuticals quicker, cheaper and more effective.

However, *significant effort* is needed to find these datasets, make sense of them, and ultimately use them for a new purpose

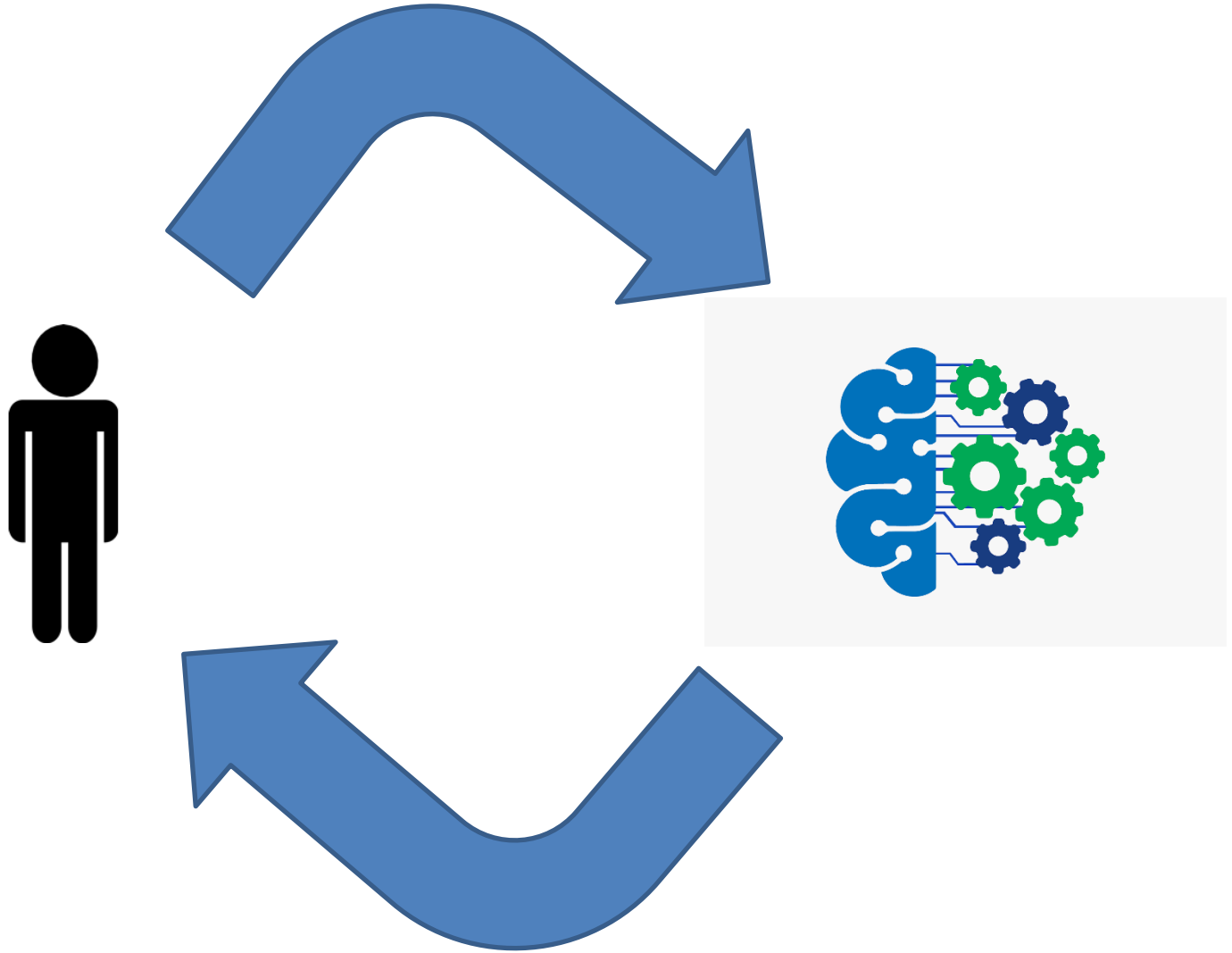
We need a new *social contract*, supported by *legal* and *technological* infrastructure to make digital resources available and accessible to

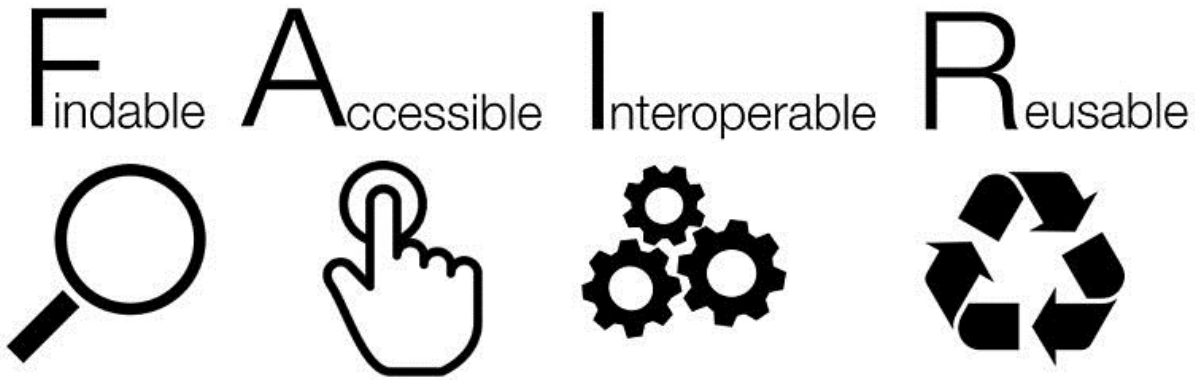
<u>people *and* machines</u>

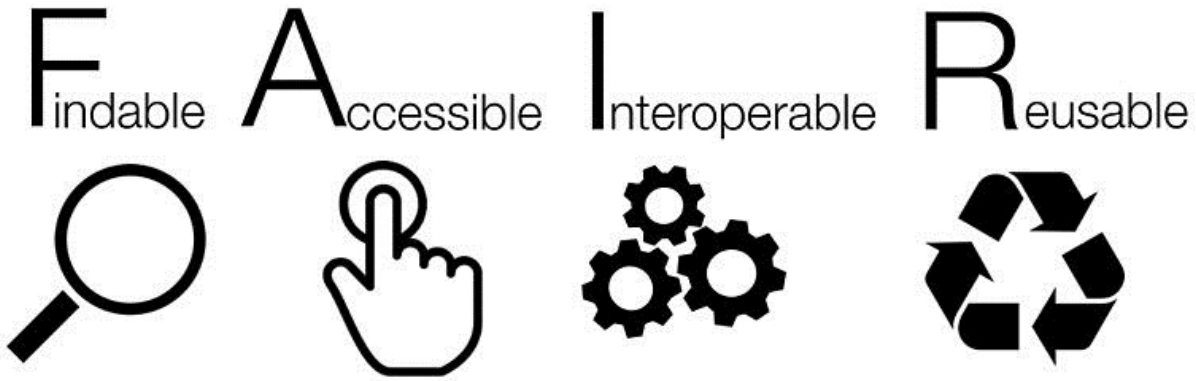


Lambin et al. Radiother Oncol. 2013. 109(1):159-64. doi: 10.1016/j.radonc.2013.07.007

Human machine cooperation




Findable
Accessible
Interoperable
Reusable



An international, bottom-up paradigm for the discovery and reuse of digital content

by people who want to use machines

SCIENTIFIC DATA

The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson, Michel Dumontier [...] Barend Mons

Affiliations | Contributions | Corresponding author

Scientific Data **3**, Article number: 160018 (2016) | doi:10.1038/sdata.2016.18

Received 10 December 2015 | Accepted 12 February 2016 | Published online 15 March 2016

http://www.nature.com/articles/sdata201618

Box 2 | The FAIR Guiding Principles

To be Findable:
F1. (meta)data are assigned a globally unique and persistent identifier
F2. data are described with rich metadata (defined by R1 below)
F3. metadata clearly and explicitly include the identifier of the data it describes
F4. (meta)data are registered or indexed in a searchable resource

To be 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 are no longer available

To be 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

To be 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 in a nutshell

FAIR aims to create **social** and **economic impact** by facilitating the discovery and reuse of **digital resources** through a set of basic requirements:

- **unique identifiers** to retrieve all forms of digital content and knowledge
- **high quality meta(data)** to enhance discovery of digital resources
- **registered in appropriate repositories** to make sure they can be found
- **use of common vocabularies** to create shared meaning and facilitate search
- **adherence to community standards** for common representations
- **detailed provenance** to provide context and facilitate reproducibility
- **social and technological commitments** to realize reliable access
- **simpler terms of use** to clarify expectations and intensify innovation

FAIR: Impact

EUROPEAN COMMISSION

Press Release Database

European Commission > Press releases database > Press Release details

European Commission - Statement

G20 Leaders' Communique Hangzhou Summit

Hangzhou, 5 September 2016

1. We, the Leaders of the G20, met in Hangzhou, China on 4-5 September 2016.

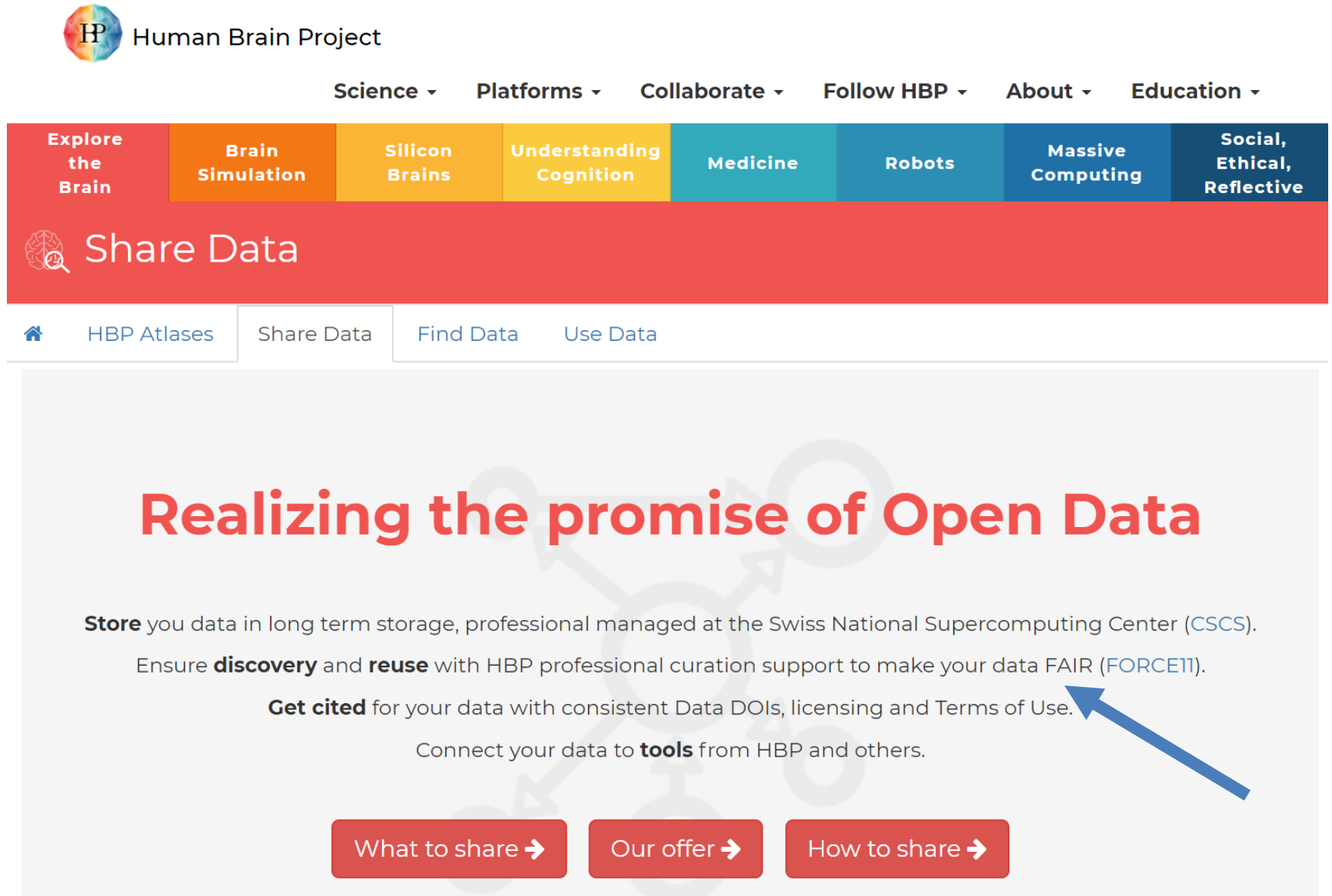
12. To achieve innovation-driven growth and the creation of innovative ecosystems, we support dialogue and cooperation on innovation, which covers a wide range of domains with science and technology innovation at its core. We deliver the G20 2016 Innovation Action Plan. We commit to pursue pro-innovation strategies and policies, support investment in science, technology and innovation (STI), and support skills training for STI - including support for the entry of more women into these fields - and mobility of STI human resources. We support effort to promote voluntary knowledge diffusion and technology transfer on mutually agreed terms and conditions. Consistent with this approach, we support appropriate efforts to promote open science and facilitate appropriate access to publicly funded research results on findable, accessible, interoperable and reusable (FAIR) principles. In furtherance of the above, we emphasize the importance of open trade and investment regimes to facilitate innovation through intellectual property rights (IPR) protection, and improving public communication in science and technology. We are committed to foster exchange of knowledge and experience by supporting an online G20 Community of Practice within the existing Innovation Policy Platform and the release of the 2016 G20 Innovation Report.

NIH

Big Data to Knowledge(BD2K)

DCPPC

DATA COMMONS PILOT PHASE CONSORTIUM

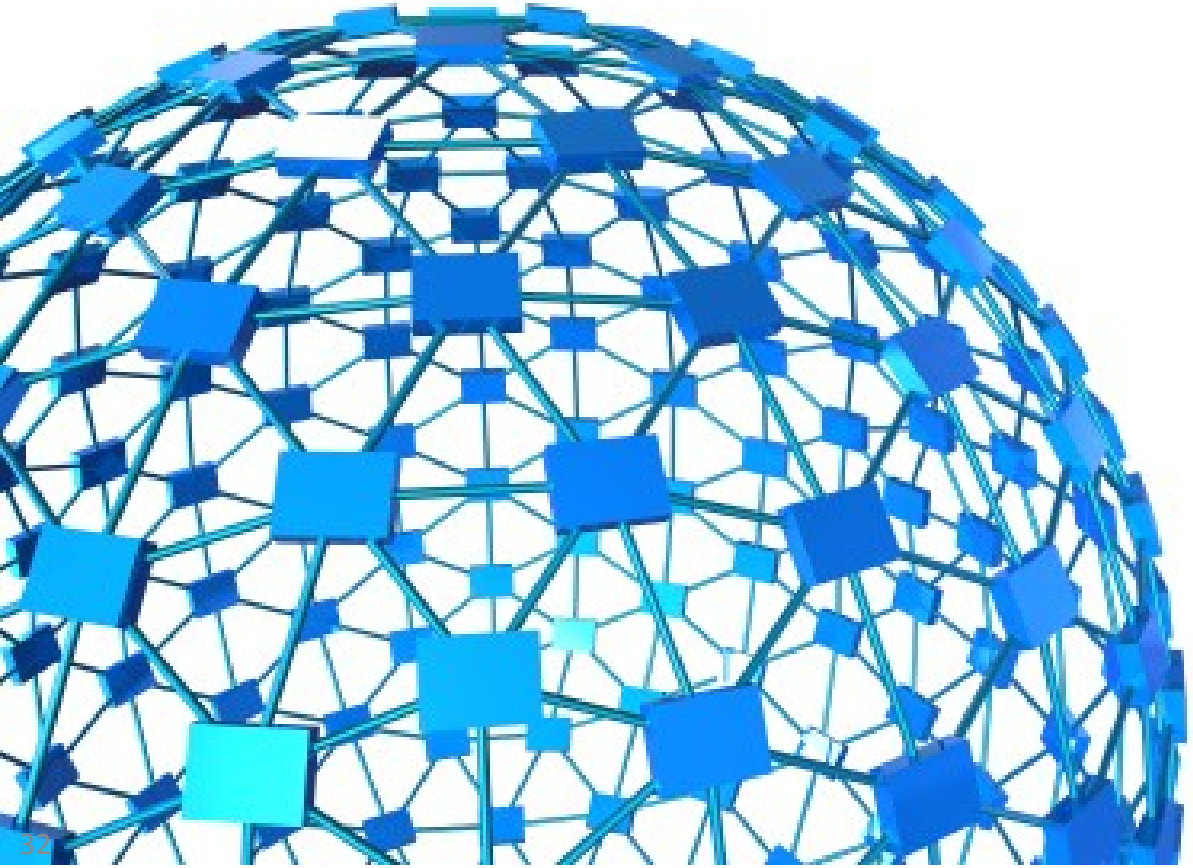

Human Brain Project
Science
Platforms
Collaborate
Follow HBP
About
Education
Explore the Brain
Brain Simulation
Silicon Brains
Understanding Cognition
Medicine
Robots
Massive Computing
Social, Ethical, Reflective
Share Data
HBP Atlases
Share Data
Find Data
Use Data
Realizing the promise of Open Data
Store you data in long term storage, professional managed at the Swiss National Supercomputing Center (CSCS).
Ensure discovery and reuse with HBP professional curation support to make your data FAIR (FORCE11).
Get cited for your data with consistent Data DOIs, licensing and Terms of Use.
Connect your data to tools from HBP and others.
What to share
Our offer
How to share

Improving the FAIRness of digital resources will *increase* their potential for and ease in reuse

The Semantic Web
is a portal to the **web of knowledge**

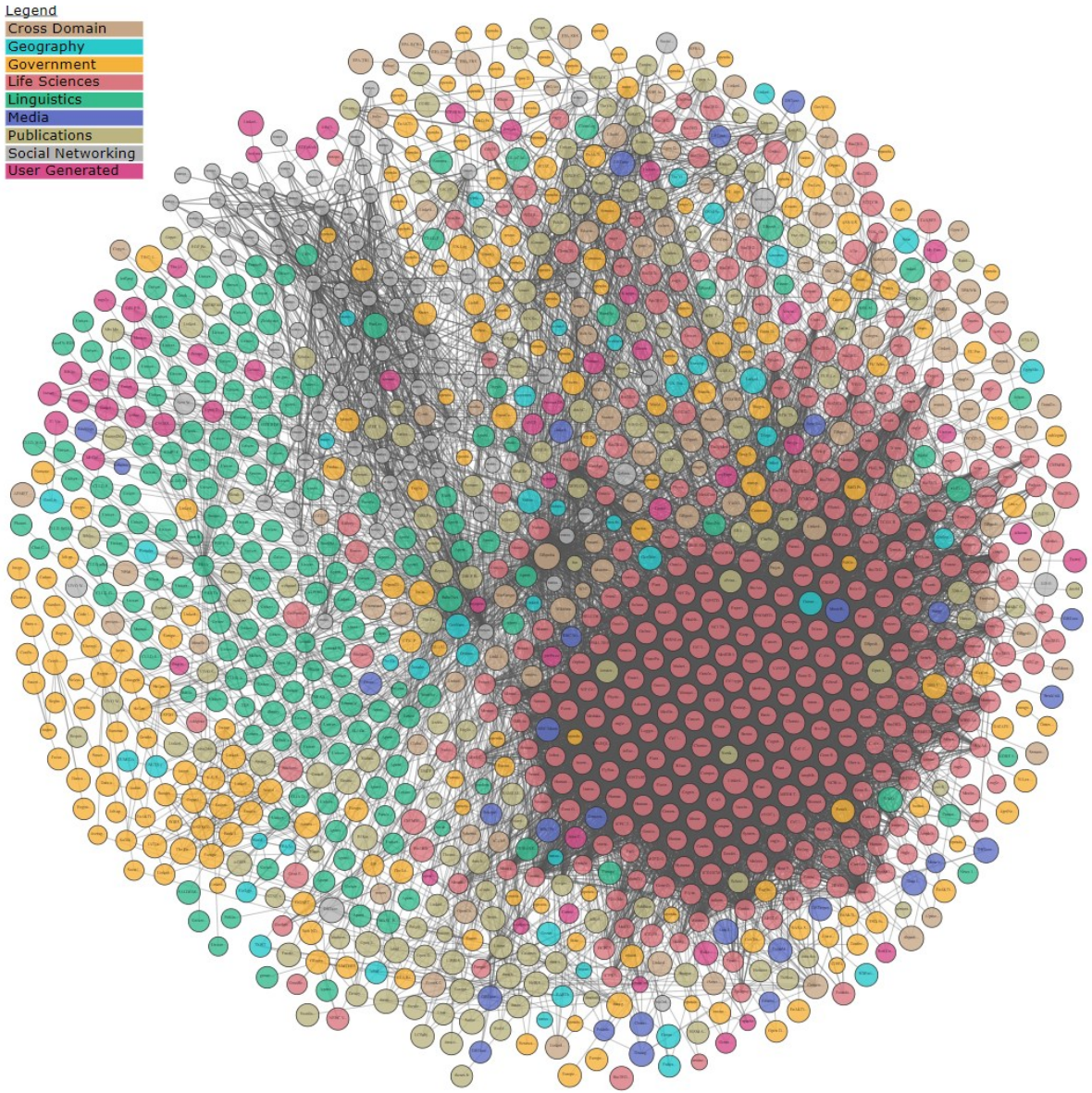
standards for publishing, sharing and querying
facts, expert knowledge and services

scalable approach for the discovery
of *independently constructed,*
collaboratively described,
distributed knowledge



W3C Semantic Web

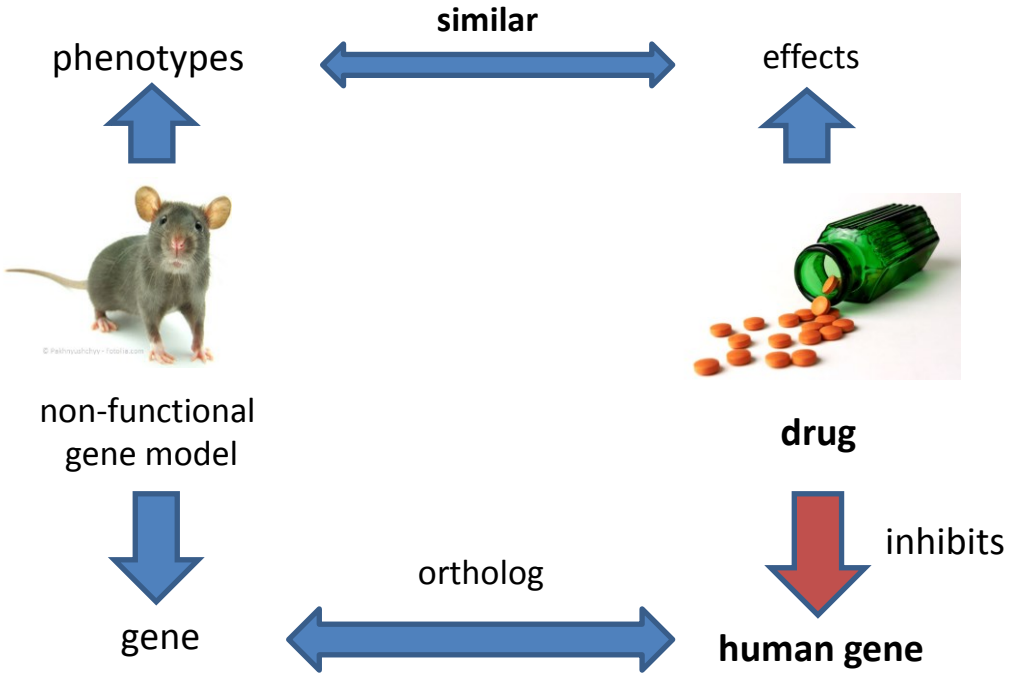
The semantic web community has built a massive open and decentralized knowledge graph



The Linked Open Data Cloud from lod-cloud.net

Identifying drug targets from mouse knock-out phenotypes

Main idea: we compare the phenotypes of knockout mouse models with the effects of drugs. When similar, we hypothesize that the drug acts as an inhibitor of the gene, thereby mimicking its phenotypic effect.



14,682 drugs; 7,255 mouse genotypes
Validation against known and predicted inhibitor-target pairs
- 0.82 ROC AUC for mouse targets
- 0.76 ROC AUC for human targets

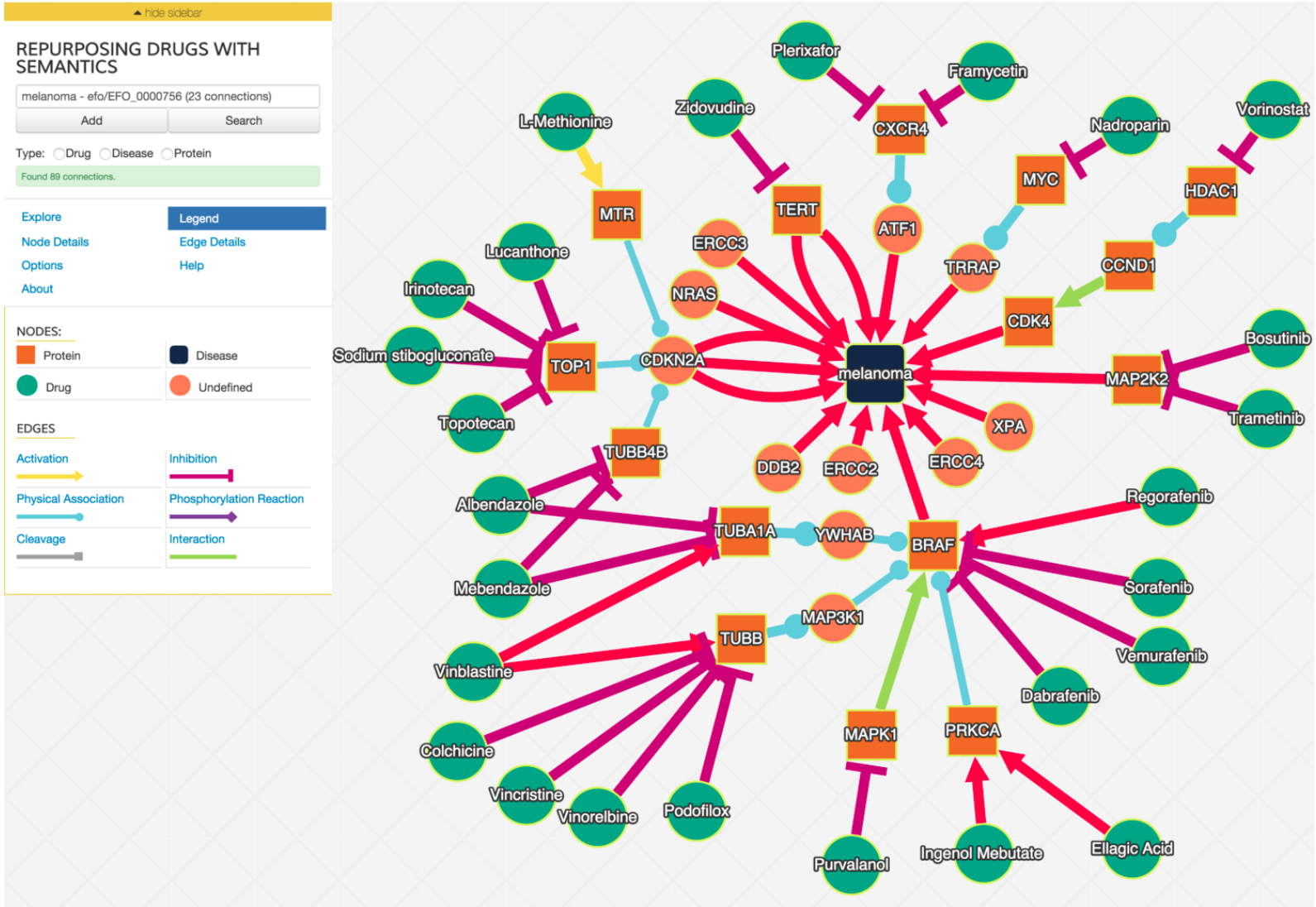
Diclofenac: NSAID used to treat pain, osteoarthritis and rheumatoid arthritis

- 46% drug effects explained by **COX-2** knockout: inflammation, gastritis, constipation, upper GI tract pain
- 49% drug effects explained by **PPARg** knockout: peroxisome proliferator activated receptor gamma (PPARg) regulates metabolism, proliferation, inflammation and differentiation

Mouse model phenotypes provide information about human drug targets. Bioinformatics, 2013. doi:10.1093/bioinformatics/btt613

Find new uses for existing drugs

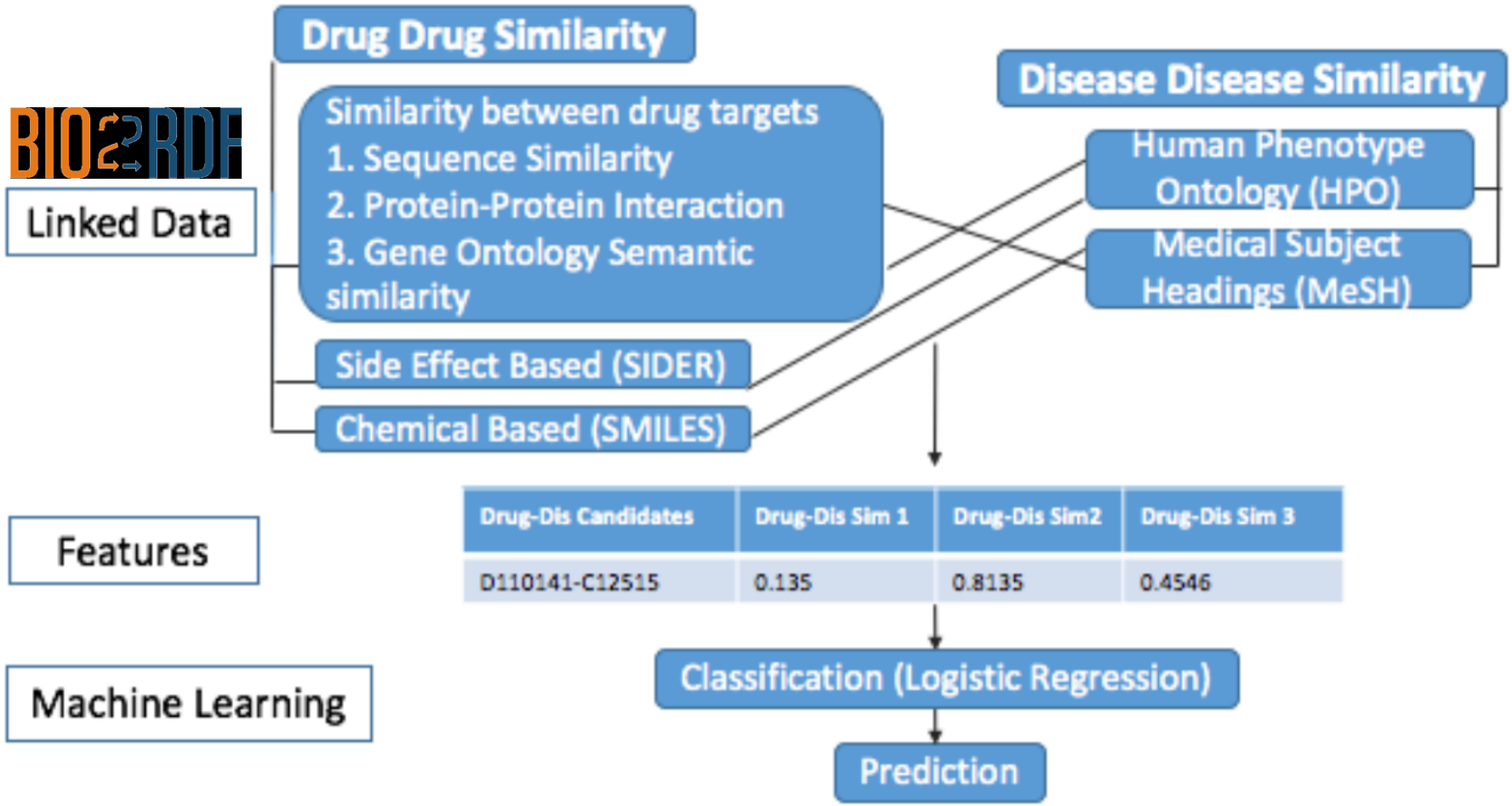
by exploring a probabilistic semantic knowledge graph



And validate them against pipelines for drug discovery

Status	Drug	Pathway	Steps	Joint p
Approved	Vemurafenib[2]	BRAF	2	0.98
Phase III	Dabrafenib[13]	BRAF	2	0.98
	Sorafenib[14]	BRAF	2	0.98
	Vinblastine[18]	MAP kinase	3	0.93
Phase II	Zidovudine[29]	TERT	2	0.98
	Trametinib[19]	MAP kinase	2	0.98
	Regorafenib[15]	BRAF	2	0.98
	Nadroparin[30]	MYC	3	0.97
	Vinorelbine[20]	MAP kinase	3	0.93
	Irinotecan[43]	CDKN2A	3	0.93
	Topotecan[44]	CDKN2A	3	0.93
Phase I	Sodium stibogluconate[45]	CDKN2A	3	0.93
Case Study	Ingenol Mebutate[46]	PRKCA/BRAF	3	0.95
In Vitro	Bosutinib[17]	MAP kinase	2	0.98
	Purvalanol[21]	MAP kinase/TP53	3	0.97
	Ellagic Acid[47]	PRKCA/BRAF	3	0.95
	Albendazole[48]	CDKN2A	3	0.93
	Colchicine[22]	MAP kinase	3	0.93
In Vivo	Plerixafor[27]	CXCR4	3	0.97
	Vincristine[23]	MAP kinase	3	0.93
	L-Methionine[49]	CDKN2A	3	0.93
	Mebendazole[50]	CDKN2A	3	0.93

Finding melanoma drugs through a probabilistic knowledge graph.
PeerJ Computer Science. 2017. 3:e106 https://doi.org/10.7717/peerj-cs.106

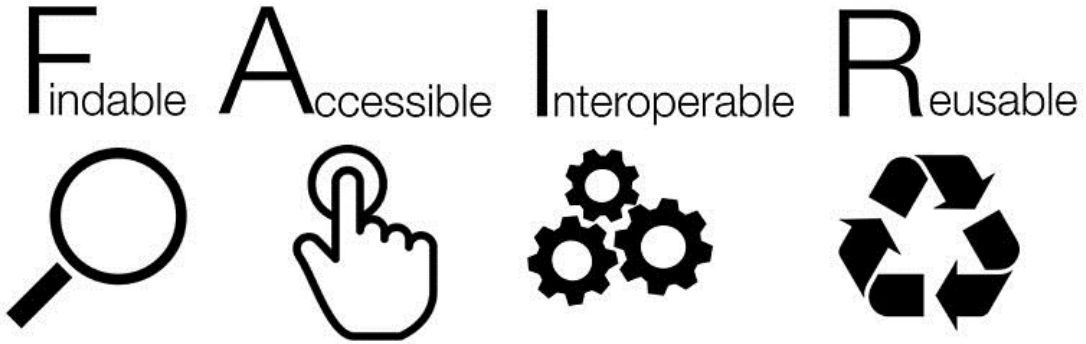


accuracy 0.80-0.89
(depends on gold standard)

Summary

- **Drug discovery remains an error failure prone endeavor that will benefit from more effective selection of early candidate molecules based on smart algorithms operating over vast amounts of data across the discovery pipeline**

- **The FAIR principles** offers a way forward: a global initiative to enhance the discovery and reuse of all kinds of digital resources of benefit to drug discovery, and biomedical research in general



The mission of the **Institute of Data Science at Maastricht University** is to foster a collaborative environment for multi-disciplinary data science research, interdisciplinary training, and data-driven innovation .

We tackle key **scientific, technical, social, legal, ethical issues** that advance our understanding across a variety of disciplines and strengthen our communities in the face of these developments.

michel.dumontier@maastrichtuniversity.nl

Website: http://maastrichtuniversity.nl/ids

Maastricht University