



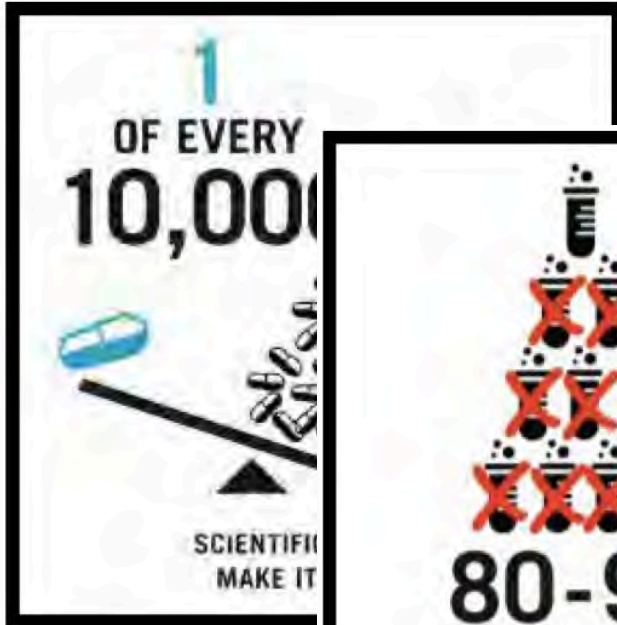
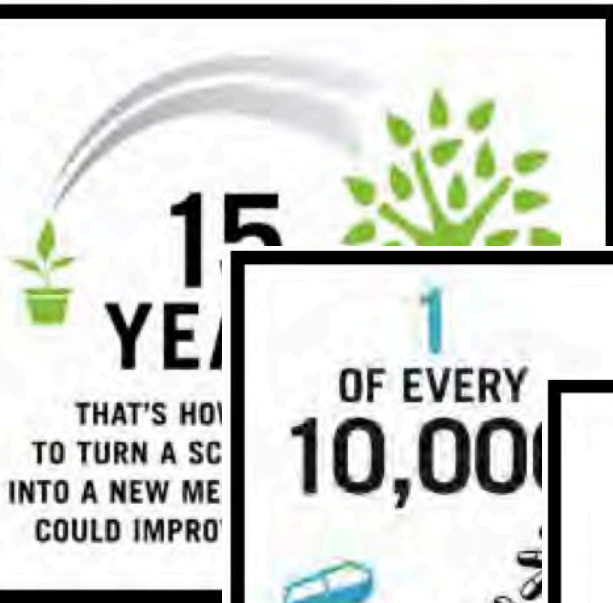
tranSMART
FOUNDATION

The tranSMART Foundation, tranSMART Platform, and Open Data Collaborations

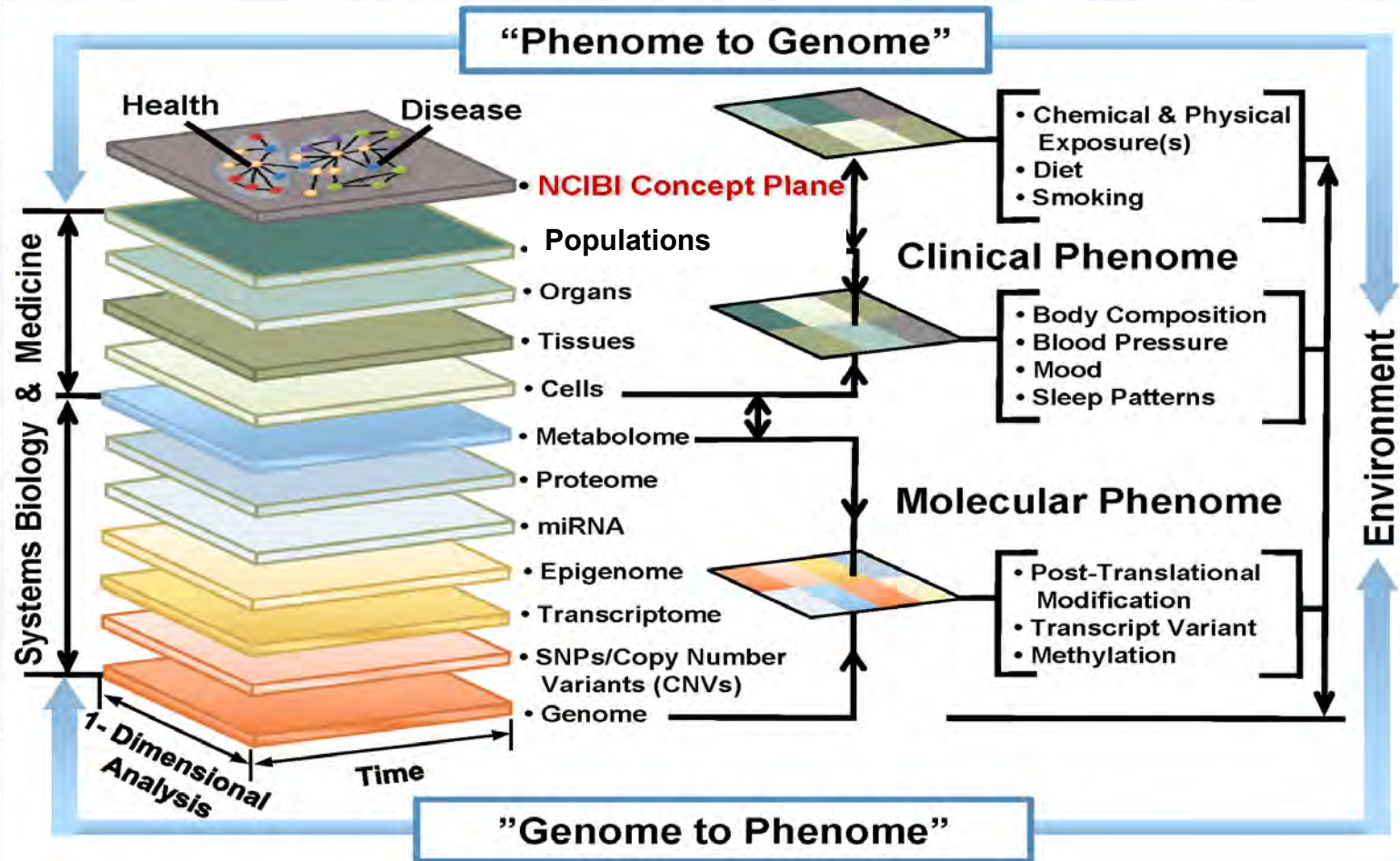
Translational Research

Translational Research Is the process of applying ideas, insights, and discoveries generated through basic scientific inquiry to the treatment and prevention of human disease – the critical bridge between basic and clinical research.

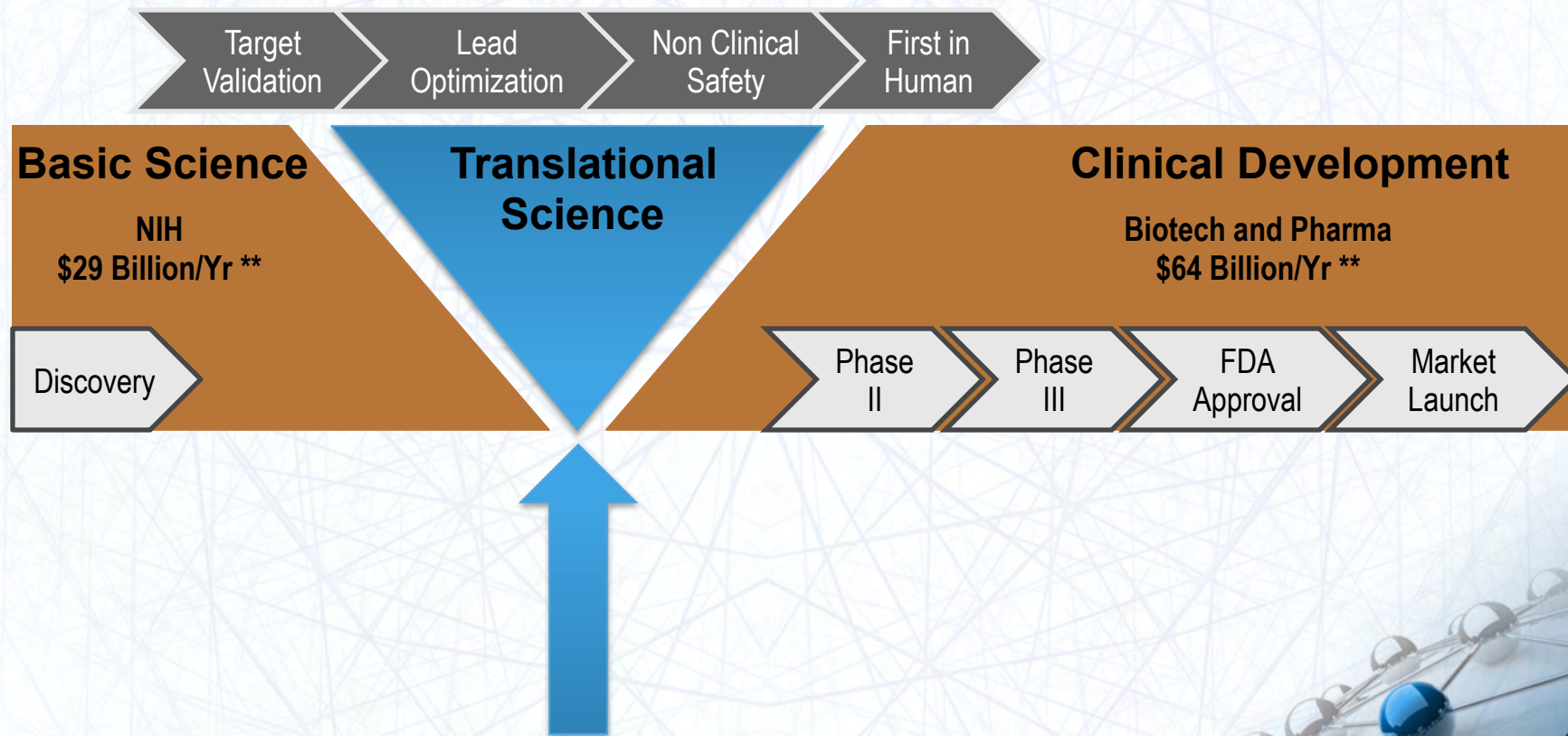
The steps in translational research are designed to ensure that the discoveries that advance into human trials have the highest possible chance of success in terms of both safety and efficacy.



Translational Research Data is Big Data

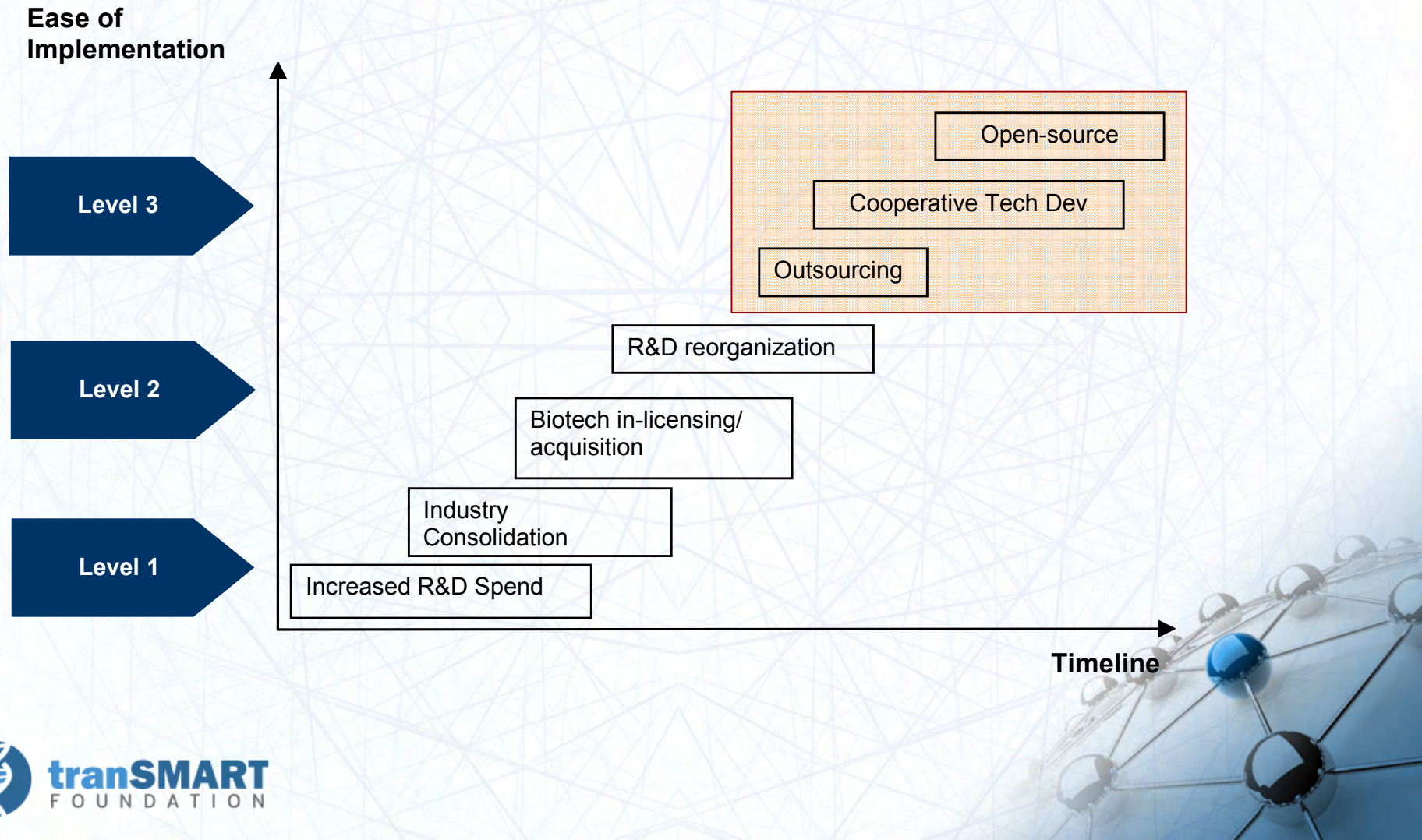


The Translational Challenge



The Translational Gap

The Innovation Challenge



The tranSMART Foundation

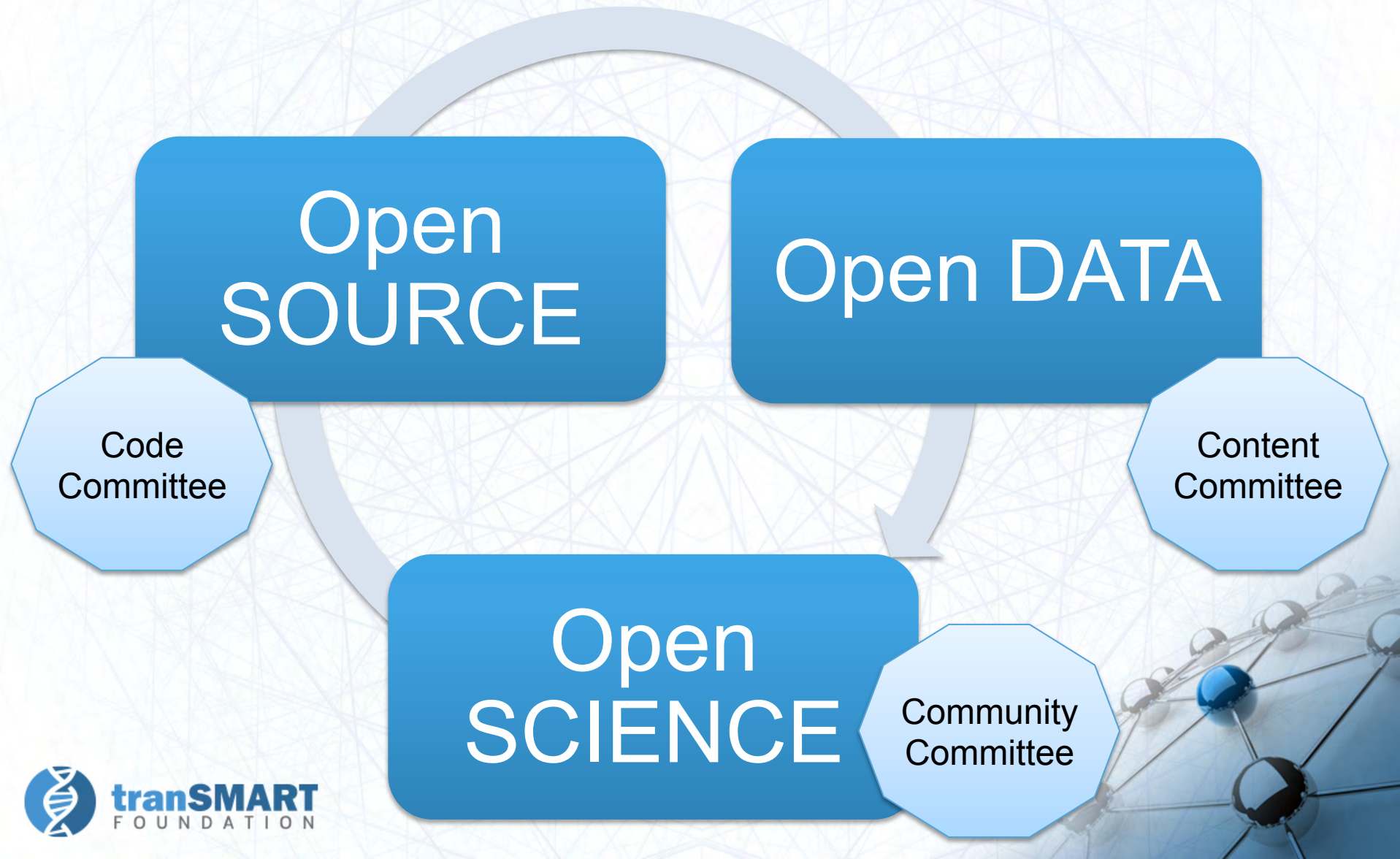


tranSMART
FOUNDATION

- The tranSMART Foundation is a member-driven non-profit foundation developing an ***open-source / open-data*** community around the tranSMART translational research platform.
- The mission of the Foundation is to stimulate the growth and development of the translational research community and the tranSMART platform, and to enable the development of ***precision medicine***



tranSMART Foundation Missions



tranSMART Foundation Ethos:

From the Linux Foundation Playbook

Open communication

- With an open community and publicly visible and accessible communication channels, anyone can join the community and meet hundreds of other community members just like them.

Licensing of work

- Every contribution to the community is licensed in such a way that it benefits the entire community. The fair licensing of all contributions adds a strong sense of confidence to the security of the community.

Open tools

- Anyone with an Internet connection and a computer can contribute. All of the development tools and documentation are entirely free and open to access. This provides a low barrier to entry, and lets new users play with the technology.

The tranSMART Foundation History

- 2012
 - The tranSMART Platform v1.0 released under GPL license by J&J in February 2012
 - An open Community Meeting was held on tranSMART at the BioIT World in April 2012
- 2013
 - An organizing meeting was held at the **University of Michigan** in Ann Arbor in February, 2013
 - The Foundation was incorporated on April 15, 2013
 - Founders include **University of Michigan**, Imperial College and the Pistoia Alliance
 - Foundation initiates membership program – October 2013
 - First fully open-stack release of tranSMART – v1.1 – November 2013
 - The first Annual Meeting of the Foundation was held in Paris at Sanofi in November 2013
 - The plans for v1.2 were laid out
- 2014
 - Foundation receives 501c3 tax-exemption from the IRS
 - Version 1.2, a broad community effort, released in Aug 2014
 - The Second Annual Meeting at **University of Michigan**, Oct 2014
 - Foundation surpasses 20 members – December 2014
- 2015
 - Requirements gathering for v1.3 – April 2015
 - Foundation initiates v1.3 development program
 - Third Annual Meeting to be held in Amsterdam at the Netherlands Cancer Institute – October 2015

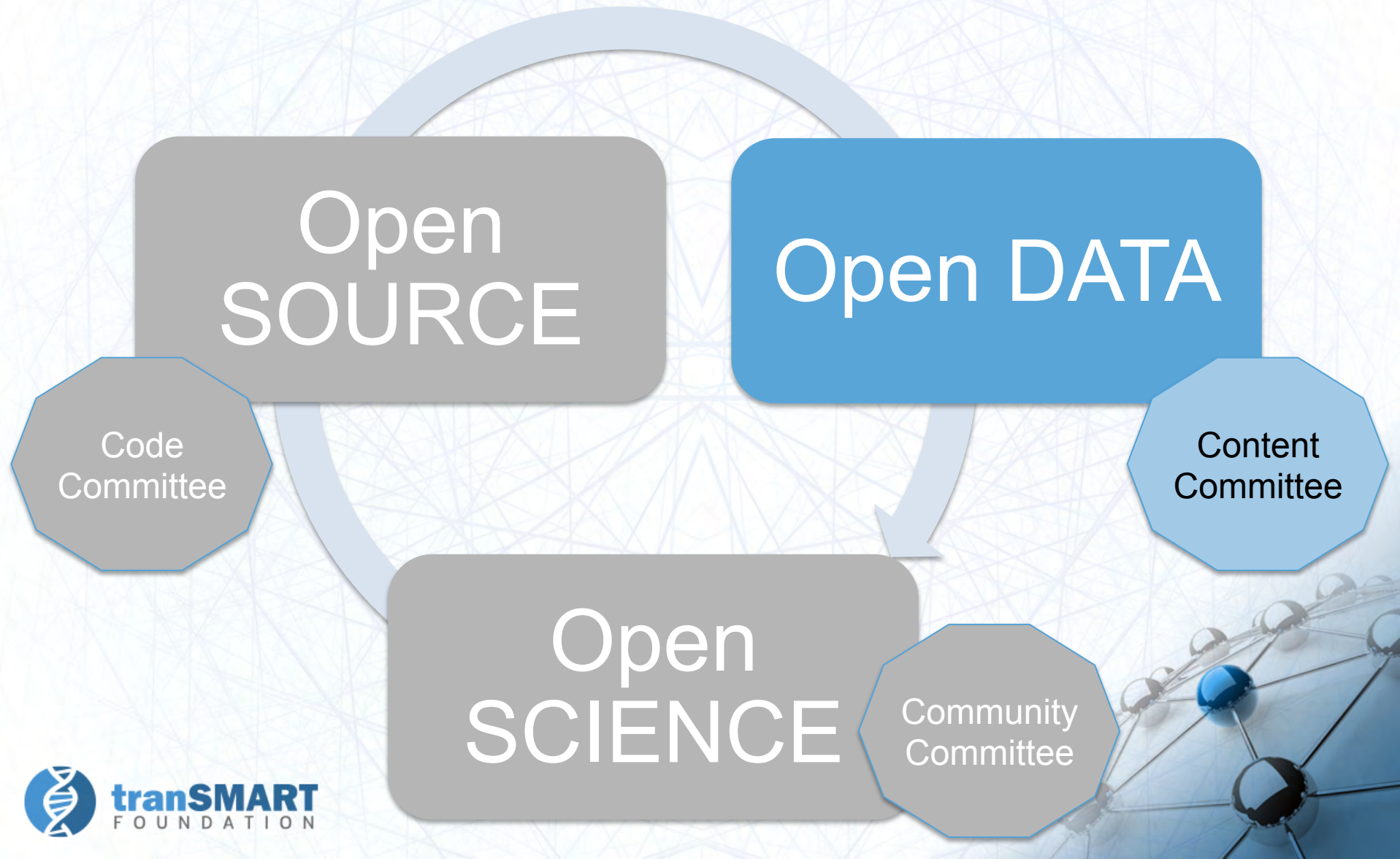


The 'Mortar between the Bricks'

- The Foundation provides the structure that links the community members together, to build a strong community and platform



tranSMART Foundation Missions





tranSMART
FOUNDATION

Open-data

Data Availability and Open-Data

Open data is the idea that certain [data](#) should be ***freely available to everyone to use and republish as they wish, without restrictions*** from [copyright](#), [patents](#) or other mechanisms of control.^[1]

The goals of the open data movement are similar to those of other "Open" movements such as [open source](#), [open hardware](#), [open content](#), and [open access](#).

The philosophy behind open data has been long established (for example in the [Mertonian tradition of science](#)), but the term "open data" itself is recent, gaining popularity with the rise of the [Internet](#) and [World Wide Web](#) and, especially, with the launch of open-data government initiatives such as [Data.gov](#) and [Data.gov.uk](#).



Open Big Data: The Reality

Alzheimer's Disease Neuroimaging Initiative (ADNI) DUA

I **will not further disclose these data** beyond the uses outlined in this agreement and my data use application and understand that redistribution of data in any manner is Prohibited.

Parkinson's Progression Markers Initiative (PPMI) DUA

I **will not further disclose these data** beyond the uses outlined in this agreement and my data use application.

I will do my best to ensure that Investigators who utilize PPMI data use appropriate administrative, physical and technical safeguards **to prevent use or disclosure of the data** other than as provided for by this Agreement and will promptly report any use or disclosure of data that does not comply with the guidelines established by this Agreement

The Cancer Genome Atlas (TCGA) DUA

The Requester and Approved Users agree to retain control over the data and further **agree not to distribute data** obtained through this Data Access Request to any entity or individual not covered in the submitted Data Access Request.

Key Barriers to Open Data

- **Technological**

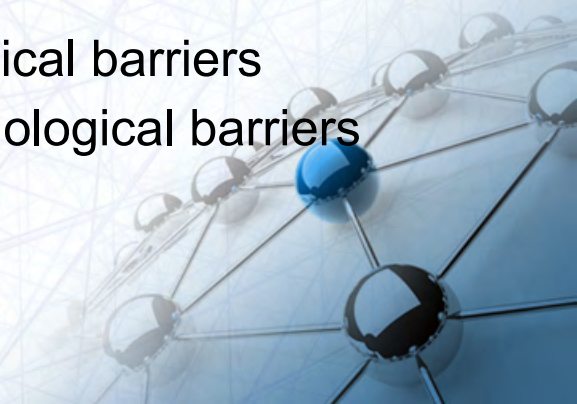
- Platforms to load data into
- Methods and technology for data integration
- Analytics for integrated analysis

- **Sociological**

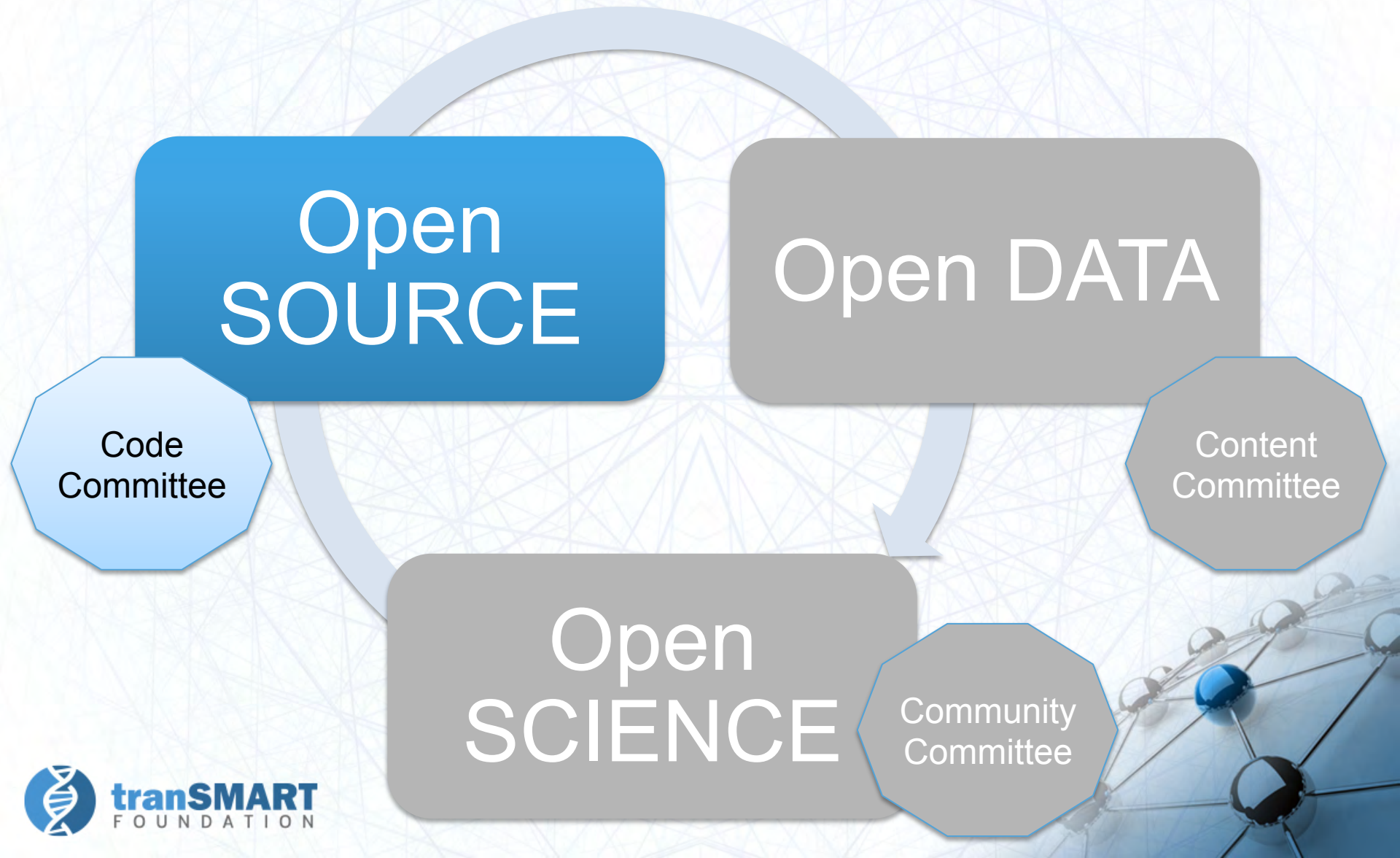
- Incentives for sharing data are misaligned
- Peer review does not require full sharing
- Individuals are rewarded for not sharing data

- **Approaching Solutions**

- The tranSMART Platform addresses the technological barriers
- The tranSMART Community is addressing the sociological barriers



tranSMART Foundation 3C Committees

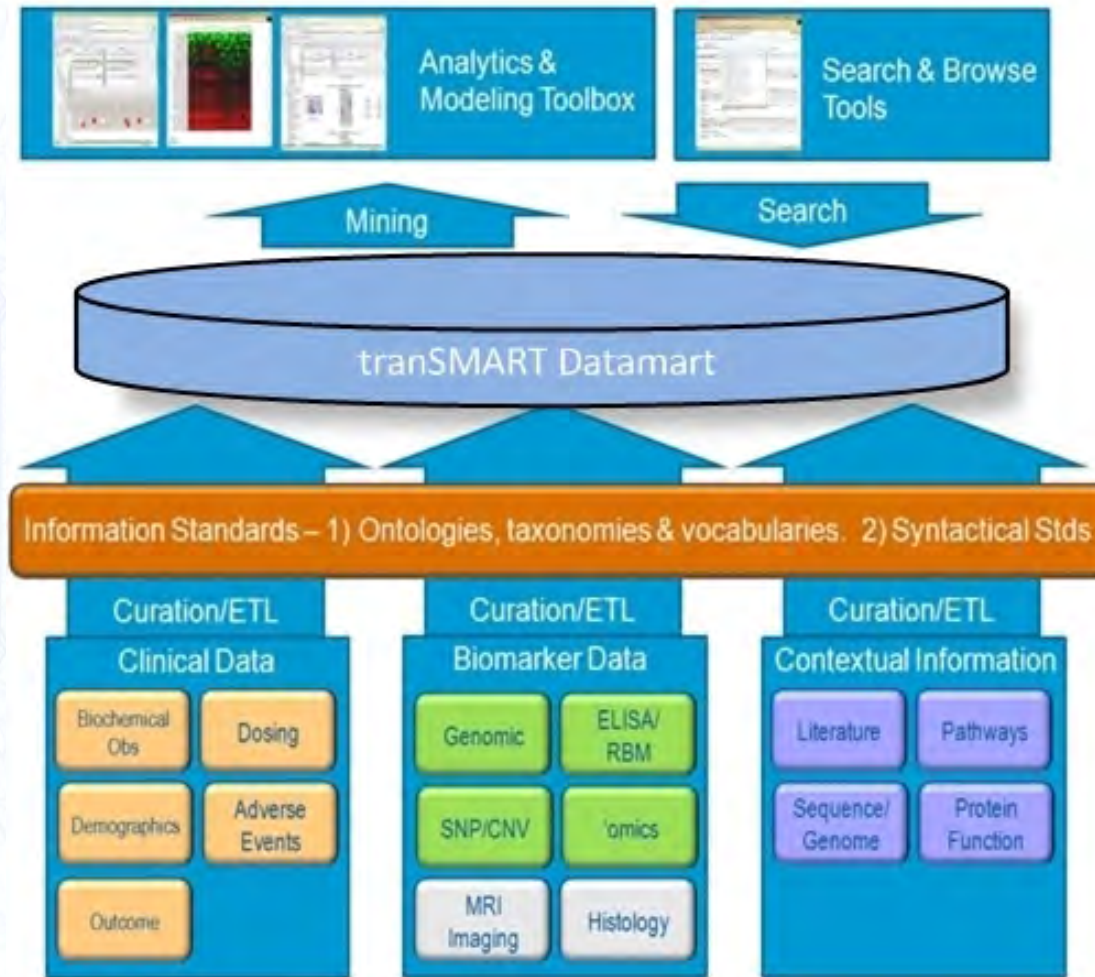




tranSMART
FOUNDATION

tranSMART Roadmap

The tranSMART Platform: an open architecture to enable translational research



tranSMART – A platform and Community

- Open-source and open-data translational biomedical research community
- Scientists, Developers, Service Providers, Clinicians

The tranSMART Platform Roadmap

- The Mission:
 - Get *Big Data* in; organize and harmonize it; browse, analyze and select it; export it for further analysis
- v1.x Roadmap – tranSMART Workgroup Edition
 - Fully open stack
 - Integration of features across disparate branches
 - Increasing genome variant capabilities
 - Integration of primary imaging data
- v2.x Roadmap – tranSMART Enterprise Edition
 - Commercial Grade, Production Quality core
 - Plug-in architecture, excellent export functions
 - Big data capabilities
 - Full Genome Sequences
 - Wearable Sensor data
 - Full time-course support
 - Technologies to support stability and scalability



V1.2 Development Summary



**Prioritized
features in
Paris, Nov
2013**

**Summer
2014 -
Organized
Hackathons,
Testathons**

**Production
v1.2
released
(Aug'14)**

**Implemented Code
Governance (Sep
2014)**

- Regular bug fixes and patches
- Documentation
- Tutorials
- Training

**Regular
Scheduled
Updates to
v1.2 (2015)**

tranSMART v1.2 Contributors



More than 100 individual developers

tranSMART Foundation Awards - 2015

Bio-IT World
CONFERENCE & EXPO '15



The tranSMART Roadmap

tranSMART Workgroup

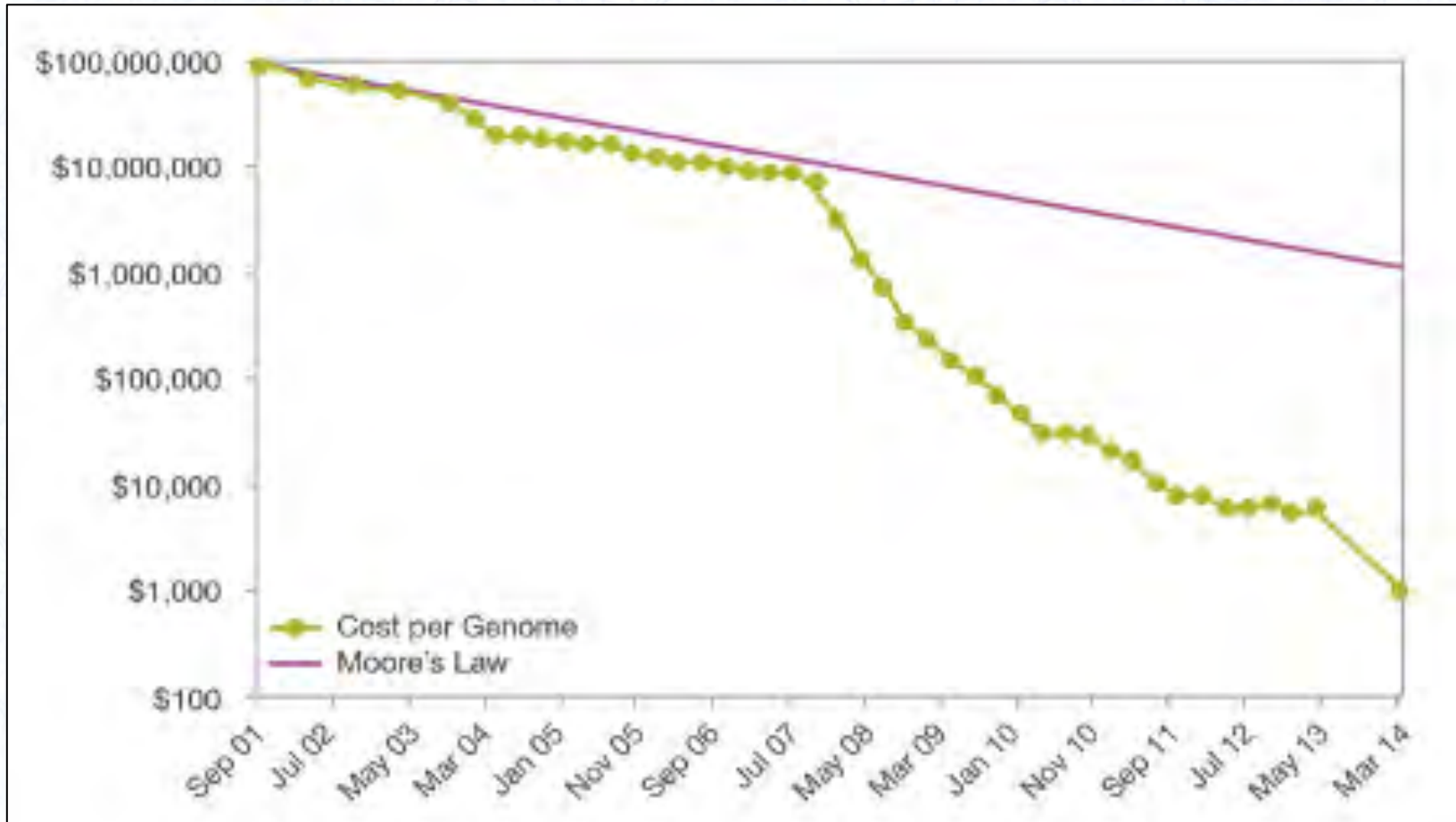
- ✓ V1.1 – an open stack
 - Support for a fully open-source stack
- ✓ V1.2 – multiomics capabilities
 - Support for new data types and analyses
- V1.3 – scaling the platform
 - Large scale variants and more

tranSMART Enterprise

- Commercial quality core
- Production-ready Enterprise platform
- Professional testing and release program
- Plug-in architecture
- Support for large scale full genomes
- Support for large-scale sensor data



Moore's law and Next Gen Sequencing



Applications of Next Generation Sequencing

- Profiling of Variation
 - Genetic variation
 - Transcript variation
 - Epigenetic variation
 - Metagenomic variation
- Discovery
 - Novel genomes
 - Novel genes
 - Novel transcripts
 - Small / long non-coding RNA

RNA Sequencing (RNASeq)

- Coding and non-coding transcript profiling
- Dynamic and Context dependent

Epigenomics

- Genome-wide protein-DNA interactions, DNA modifications
- Heritable and reversible regulation of gene expression

NGS= Not just Genome Sequence

Sensor Big Data

Activity Details

Sunday 1 Dec

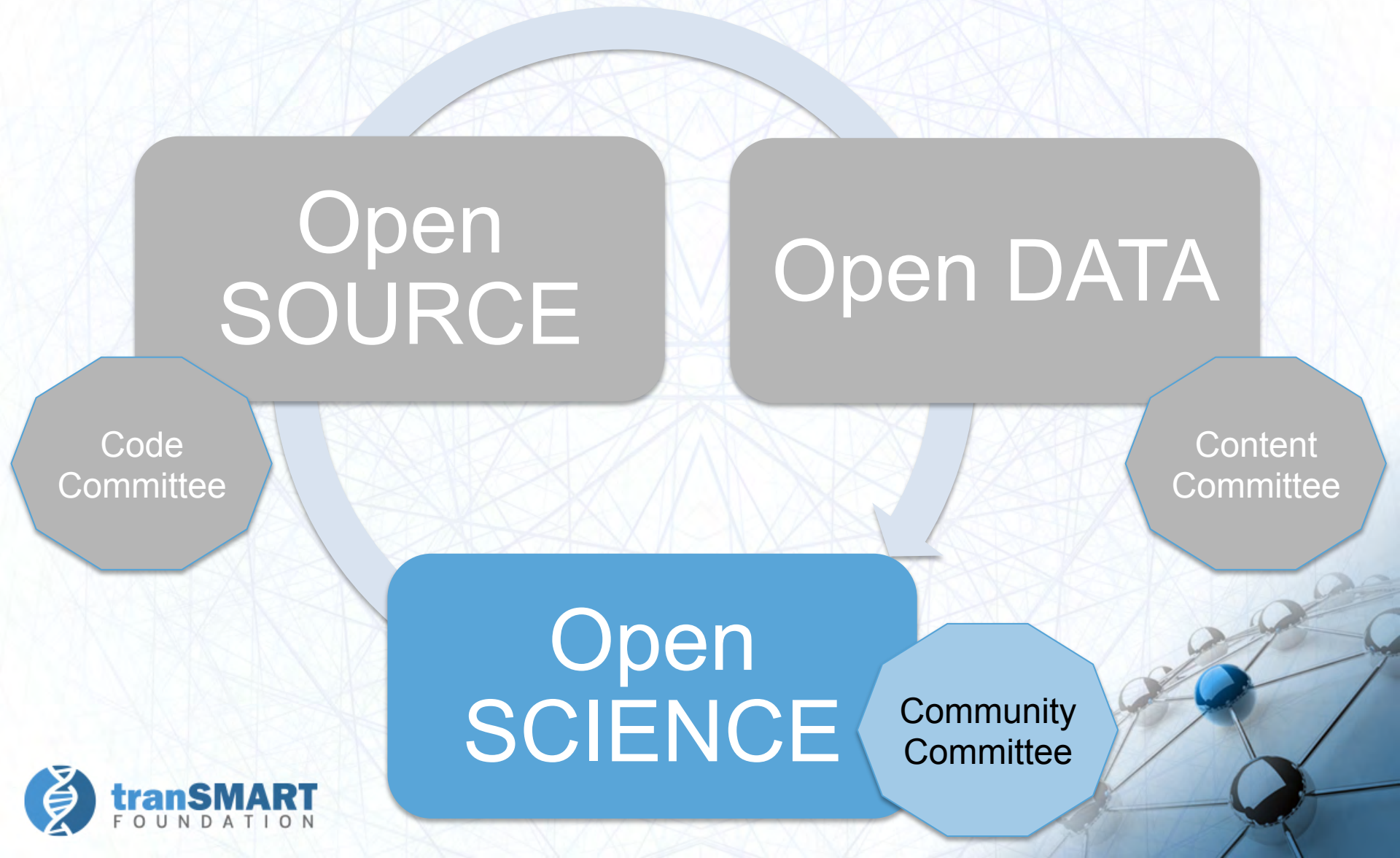


Research Activities

- tranSMART development requires proactive research on new technology, algorithms and coupling systems
- Research in world class universities (Imperial, UMICH, ULux) is one of the unique advantage of the tranSMART community
- Achievement made :
 - **BigTranSMART** (Imperial College) : tranSMART on Hbase and fully compatible with 1.2
 - **Cognitive tranSMART** (Imperial College with IBM) : BigtranSMART integration with Watson
 - **NeuroSMART** (Imperial College with Biogen): tranSMART with XNET and neuroimaging processing
 - **tranSMART-R** (University of Luxembourg): tranSMART with SMARTR integration for interactive analysis



tranSMART Foundation 3C Committees



Cross Neurodegenerative Disease Datathon

- Co-developed with the Michael J. Fox Foundation
- 3 Day data science hackathon on neuroscience data
- Alzheimers
 - ADNI
- Parkinsons
 - PPMI, LRRK2, BIOFIND
 - Plus 10 open GEO datasets
- Objectives
 - Identify cross disease biomarkers
 - Define similarities and differences between diseases
 - Enable new scientific discoveries

A 5-star Open Data Project



What is a “Datathon”?

- A datathon is an intense 3-day workshop that challenges researchers to turn data and information into knowledge.
- The datathon format is modeled after hackathons, which are focused on software development.
 - The difference is that datathons use research questions and datasets to advance knowledge, not to develop applications.
- At a datathon, participants work in teams to frame research questions, create and implement a research design, mobilize data resources and present their findings in front of a panel of judges.

Datathons are much harder to run than Hackathons!

Why is a Datathon a Good Idea?

- Datathons enable scientists to test new research ideas and meet potential collaborators in a working environment without requiring a great deal of commitment.



Potential Outcomes...

- Demonstrate the value of integrated data
 - Show the value of bringing these disparate datasets together in a single analytical platform
- Generate interest in these datasets
 - Demand to bring these data 'back home' to integrate with more data
- Bring Awareness to the need for 'Open Data'
 - Integration projects like this should be much easier to accomplish and much more commonly executed
- Develop new collaborations
 - Develop new relationships that extend beyond the 3-day event
- Define new data, new experiments and new methods
 - Contribute to the data and code roadmaps for the Foundation



Dataset Explorer View

The screenshot displays the Dataset Explorer web application. The browser address bar shows the URL <https://ts1.loni.usc.edu/transmart/datasetExplorer/index>. The user is logged in as Terry (Gmail). The interface includes a top navigation bar with tabs: Browse, Analyze, Gene Signature/Lists, GWAS, Upload Data, Admin, and Utilities. Below this is a secondary navigation bar with links: Comparison, Summary Statistics, Grid View, Advanced Workflow, Data Export, Export Jobs, Analysis Jobs, Workspace, and Sam. On the left, there is a sidebar with 'Active Filters' (showing 'and') and 'Navigate Terms'. The 'Navigate Terms' section lists various study categories and specific datasets: 'Across Trials', 'PD Related GEO Studies' (including GSE20146 (20), GSE20168 (29), GSE20291 (35), GSE20292 (29), GSE20295 (93), GSE20333 (12), GSE35642 (18), GSE54282 (33), GSE6613 (105), and GSE7621 (25)), 'Private Studies' (including ALZHEIMERS_ADNI (3150), PARKINSONS_BIOFIND (232), PARKINSONS_LRRK2C (2841), and PARKINSONS_PPMI (1307)), and 'Public Studies'. The main content area is divided into two columns for 'Subset 1' and 'Subset 2'. Each column has a 'Save Subset' button and a 'Clear' button. Below these are three rows of input fields for each subset, with 'AND' operators and 'Exclude' buttons between them. The status bar at the bottom right indicates 'Ready'.

**Clinical and High Dimensional Data
from more than 8,000 patients**

Key Outcomes:

- Very interesting preliminary findings
 - 4 SNP's that predict Parkinson's progression BETTER than Age
 - 3 Biomarkers that cross Alzheimers and Parkinsons
 - Sets of biomarkers that predict disease stage
- More “Open” Neurodegenerative Disease Data
 - tranSMART Foundation, MJFF and LONI are working together to enable the distribution and access of ADNI, PPMI, LRRK2 and BioFIND datasets in tranSMART
 - tranSMART Ready™ datasets
 - Preloaded Virtual Machine Images
 - Online platform hosted at LONI



Summary

- Translational research IS data science
- tranSMART is a key part of a translational research infrastructure
- The tranSMART platform is an open-collaborative platform for translational research
- Open-data is critical to realizing the promise of translational research
- The tranSMART Foundation is a condensation nucleus for the translational research community
- Joining the tranSMART Foundation and community will accelerate the implementation of translational research



What Can you do: CALL TO ACTION!

*How You
Can
HELP
Participate!*

BECOME a FOUNDATION MEMBER!!!

Provide funding for specific deliverables!!!
tranSMART releases, development of services
Provide in-kind contributions to workgroups and general services!!!

ATTEND a FOUNDATION Webinar!!!

**The third Tuesday of every month, we hold a
Community Webinar, open to all!**

Third Tuesday of every month at 11AM Eastern
Register through our website

Use the Platform!!!

**Free to use; download it today
Watch the Tutorials
Take the free Training**

Check out our website

Visit our Website

www.transmartfoundation.org

Attend the Annual Meeting

Transmart.eventbrite.com