

The Yale Open Data Access (YODA) Project: 10 Years of Clinical Trial Data Sharing

PHUSE US Connect 2026
Data Transparency Pavillion
March 23, 2026



Joseph S. Ross, MD, MHS

Section of General Internal Medicine, School of Medicine
Center for Outcomes Research and Evaluation, Yale-New Haven Hospital



Strengthening Science through Data Sharing

- Ensures all data can be used for secondary research to advance science and improve public health and healthcare
- Positions research as a public good
- Respects contributions of participants:
 - maximizing value of collected data, while
 - minimizing duplicative data collection
- Promotes responsible conduct of research through transparency and reproducibility:
 - sample, design, and analysis



“

The scientific community [is] arguably the most powerful collective enterprise in human history.

Atul Gawande



Our Mission

The Yale University Open Data Access (YODA) Project's mission is to advocate for the responsible sharing of clinical research data, open science, and research transparency. The Project is committed to supporting research focused on improving the health of patients and informing science and public health. The YODA Project can only improve with your feedback. Please share your comments and ideas.

CONTACT US

Our Model

The YODA Project seeks mutually beneficial partnerships with Data Partners, promoting independence, responsible conduct of research, good stewardship of data, and the generation of knowledge in the best interest of society. To participate, each Data Partner must transfer full jurisdiction over data access to the YODA Project.

HOW IT WORKS

Request Data

Are you ready to request data? To date, 499 trials have been identified as available. The YODA Project and Data Partners continue to identify and add more.

GET STARTED

Johnson & Johnson Partnership

- Initiated in 2014 after proof-of-concept effort with Medtronic, Inc.
- Focused on promoting and facilitating access to clinical trial data:
 - All pharmaceutical products (including legacy trials)
 - Device and diagnostic products as of 2015
 - Consumer products as of 2017
- Established data access policy and procedures, with input from Steering Committee, experts, stakeholders, and public comment

Trials By Generic Name

Below is a list of trials that have been identified as available. This is not a complete list of the trials that are available for sharing. Before a trial can be shared, Data Partners must confirm data location and availability in an electronic format, and confirm that data availability conforms to any applicable partner agreements. All trials listed below have gone through this process. We continue to add trials to this list on a regular basis.

Don't see the trial(s) you are looking for?

[Submit an Inquiry](#)
[GENERIC NAME](#)
[PRODUCT CLASS](#)
[THERAPEUTIC AREA](#)
[CONDITION STUDIED](#)
[ADVANCED SEARCH](#)

Abiraterone acetate

[VIEW TRIALS](#)

Ablation Catheter

[VIEW TRIALS](#)

Acetaminophen

[VIEW TRIALS](#)

Bedaquiline/TMC207

[VIEW TRIALS](#)

Bosentan

[VIEW TRIALS](#)

Canagliflozin

[VIEW TRIALS](#)

Daratumumab

[VIEW TRIALS](#)

Darunavir

[VIEW TRIALS](#)

Doxorubicin hydrochloride

[VIEW TRIALS](#)

Epoetin alfa

[VIEW TRIALS](#)

Ethinyl estradiol

[VIEW TRIALS](#)

Etravirine

[VIEW TRIALS](#)

STUDY
PHASE
3

A Randomized, Double-Blind Trial of Anti-TNF Chimeric Monoclonal Antibody (Infliximab) in Combination With Methotrexate for the Treatment of Patients With Polyarticular Juvenile Rheumatoid Arthritis



CSR Summary



NCT00036374



Primary Citation



Data Specification



Annotated CRF
Available upon data
request approval

[Add Trial to Data Request](#)

PRODUCT INFO

Generic Name
Infliximab

Product Name
REMICADE®

Therapeutic Area
Muscle, Bone, and Cartilage
Diseases

Enrollment
123

% Female
85.9%

% White
92.1%

Product Class

Antirheumatic Agents - Biologic
Response Modifiers

Sponsor Protocol Number
C0168T32

Data Partner
Johnson & Johnson

Condition Studied
Arthritis, Juvenile

Mean/Median Age (Years)
11.2

SUPPORTING DOCUMENTATION

- Clinical Study Report
- Collected Datasets
- Data Definition Specification
- Annotated Case Report Form
- Protocol with Amendments
- Statistical Analysis Plan

APPROVED DATA REQUESTS ASSOCIATED WITH THIS TRIAL

Impact of the dose of Immunomodulators on pharmacokinetics of biologics: Patient level meta-analysis of randomized controlled trials
Impact of Biologic Therapy on the Risk of Arterial and Venous Thromboembolic Events In Chronic Autoimmune Diseases: A Post-Hoc Analysis of RCTs

Requests Submitted Online

- Investigator names, affiliations, funding
- Narrative summary / public abstract
- Detailed research proposal, including:
 - Project background, clear objectives
 - Trials, eligibility criteria, variables
 - Primary and secondary endpoints
 - Statistical analysis plan
- Timeline and dissemination plan
- Data use agreement training

The image shows a white application form titled "APPLICATION" in large, bold, black letters. The form is laid out on a wooden surface. A black pen with a gold-colored tip is resting on the form. The form has several sections with labels and input fields:

- Personal Information:** Includes fields for Name (Last), First, Middle Initial, and Date of Birth.
- Address (Mailing Address):** Includes fields for Address, City, State, and Zip.
- E-Mail Address:** A single line for email.
- Services needed:** A large open area for text.
- Current Income:** A section with a table for income details.
- High School Graduate Or General Education (GED) Test Passed?** A section with "Yes" and "No" options.
- Business School, Military (Most recent first):** A section with a table for business school or military education.
- Credits Earned:** A section with a table for credits earned.
- Graduate:** A section with "Yes" and "No" options.
- Degree & Year:** A section for degree and year.

YODA Project Review

The YODA Project reviews proposals to ensure that each proposal has scientific merit, specifically verifying:

- **Scientific purpose is clearly described**
- **Data requested will be used to create or materially enhance generalizable scientific and/or medical knowledge to inform science and public health**
- **Proposed research can be pursued using the requested data**
- **Appropriateness of requested data (e.g. CSR vs IPD)**

Ensuring Data Stewardship



- **Once approved, require signed DUA**
- **Investigators gain access to data maintained on secure platform via VPN**
- **Prevents re-distribution, protects patient privacy**

YODA Protocol Number

2022-5104

Product(s) of Interest

REMICADE® (Infliximab)
SIMPONI® (Golimumab)
STELARA® (Ustekinumab)

No. Trials Provided

6

Published

 Publications Available! See more below

**Biomarkers for Predicting Long-term Outcomes in
Ulcerative Colitis**

Principal Investigator

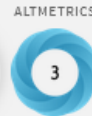
Rirong Chen

Reports & Publications

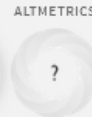
 UEG Journal 2023



 Precision Clinical Medicine 2023



 Int. Immunophar 2024



Protocol & Associated Materials

 Protocol #2022-5104 Proposal (PDF)

 Protocol #2022-5104 Review (PDF)

 Protocol #2022-5104 Due Diligence Assessment (PDF)

Fostering Collaboration and Responsible Research

Experience so far ...

- Of 499 trials currently available, 90.9% have thus far been requested
- Of 578 requests submitted, 557 (96.4%) approved, 6 (1.0%) remain under review; 13 (2.2%) withdrawn/closed, 2 (0.3%) rejected
 - Usually because data not available/cannot be adequately de-identified
- Nearly all require some administrative revision, but 108 (19.0%) required scientific revision after review for clarity
- Median number of trials per request: 3 (IQR, 1.25-8); 96.7% for IPD
- 257 manuscripts and 133 abstracts have been submitted, 220 and 92 of which have been published or presented, respectively

Strengthening Science through Data Sharing

- Numerous studies that might not otherwise have been feasible to pursue, some of which have impacted health policies and guidelines
- Facilitated direct collaborations with original investigators
- Developed efficiencies (J&J now conducts all trials intending to share)
- Replication studies have supported – not undermined – original study
- No instances of patient privacy breaches
- No publications of spurious safety findings that received unwarranted attention or disrupted patient care
- No data have been used for commercial or litigious purposes

What is the Scientific Impact of Data Sharing?

- Effort underway to characterize the scientific impact of primary and secondary publications resulting from 336 clinical trials for which data were shared on the YODA Project platform as of December 31, 2021

	Primary pubs (n=336)	Secondary pubs (n=1167)
Median journal impact factor (IQR)	7.5 (3.8-21.0)	4.7 (3.0-10.6)
Median citations per year (IQR)	9.8 (4.1-28.5)	3.3 (1.5-7.0)
Cited by clinical practice guidelines, No. (%)	233 (72.4)	256 (25.9)
Cited by policy documents, No. (%)	200 (62.1)	220 (22.2)
Median Altmetric score per year (IQR)	0.9 (0.4-2.9)	0.8 (0.3-2.9)
Median Mendeley reader count per year (IQR)	9.4 (5.3-16.9)	7.9 (4.9-13.5)

What is the Scientific Impact of Data Sharing?

- Effort underway to characterize the scientific impact of primary and secondary publications resulting from 336 clinical trials for which data were shared on the YODA Project platform as of December 31, 2021

	Primary pubs (n=336)	Secondary pubs (n=1167)	Internal 2° pubs (n=958)	External 2° pubs (n=209)
Median journal impact factor (IQR)	7.5 (3.8-21.0)	4.7 (3.0-10.6)	4.6 (2.9-10.2)	6.7 (3.4-16.6)
Median citations per year (IQR)	9.8 (4.1-28.5)	3.3 (1.5-7.0)	3.4 (1.6-7.5)	2.7 (1.1-5.6)
Cited by clinical practice guidelines, No. (%)	233 (72.4)	256 (25.9)	235 (29.2)	21 (11.4)
Cited by policy documents, No. (%)	200 (62.1)	220 (22.2)	206 (25.6)	14 (7.6)
Median Altmetric score per year (IQR)	0.9 (0.4-2.9)	0.8 (0.3-2.9)	0.6 (0.3-2.3)	2.1 (0.7-7.1)
Median Mendeley reader count per year (IQR)	9.4 (5.3-16.9)	7.9 (4.9-13.5)	8.0 (5.1-13.6)	7.4 (3.9-13.0)

Challenges Remain

- Broadening awareness of data availability
- Fostering expertise in using data from clinical trials (*it's complicated*)
- Making older trial data available in contemporary formats
- Adopt data standards, across sponsors, to enable meta-analyses
- Sustainable model that covers the cost of data sharing, including centralized platform (especially for NIH)
- Data Use Agreements ...
- Establish standards: when should data be available, for how long, how to reward those who share data?
- Many large pharma sharing, perhaps NIH, what about other sponsors?

SCIENTIFIC DATA

OPEN

Overview and experience of the YODA Project with clinical trial data sharing after 5 years

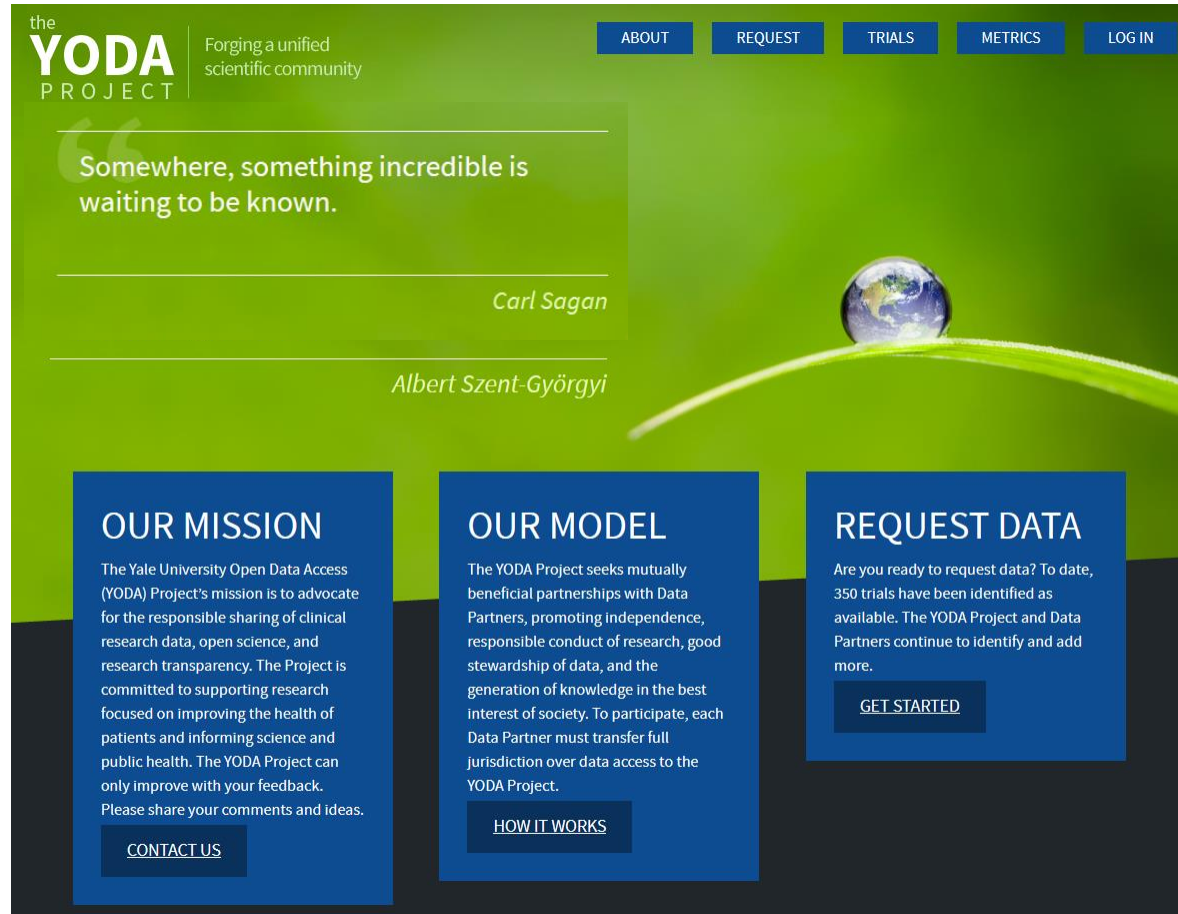
Joseph S. Ross^{1,2,3,4}, Joanne Waldstreicher⁵, Stephen Bamford⁶, Jesse A. Berlin⁵, Karla Childers⁵, Nihar R. Desai^{4,7}, Ginger Gamble⁴, Cary P. Gross^{1,2,4,8}, Richard Kuntz⁹, Richard Lehman¹⁰, Peter Lins⁵, Sandra A. Morris⁵, Jessica D. Ritchie⁴ & Harlan M. Krumholz^{2,3,4,7}

Received: 28 March 2018

Accepted: 24 October 2018

Published: 27 November 2018

The Yale University Open Data Access (YODA) Project has facilitated access to clinical trial data since 2013. The purpose of this article is to provide an overview of the Project, describe key decisions that were made when establishing data sharing policies, and suggest how our experience and the experiences of our first two data generator partners, Medtronic, Inc. and Johnson & Johnson, can be used to enhance other ongoing or future initiatives.



<http://yoda.yale.edu>

@YODAProject



@jsross119

joseph.ross@yale.edu



CENTER FOR GLOBAL CLINICAL RESEARCH DATA

Enabling Data Sharing in a Complex Global Environment

Rebecca Li, CEO

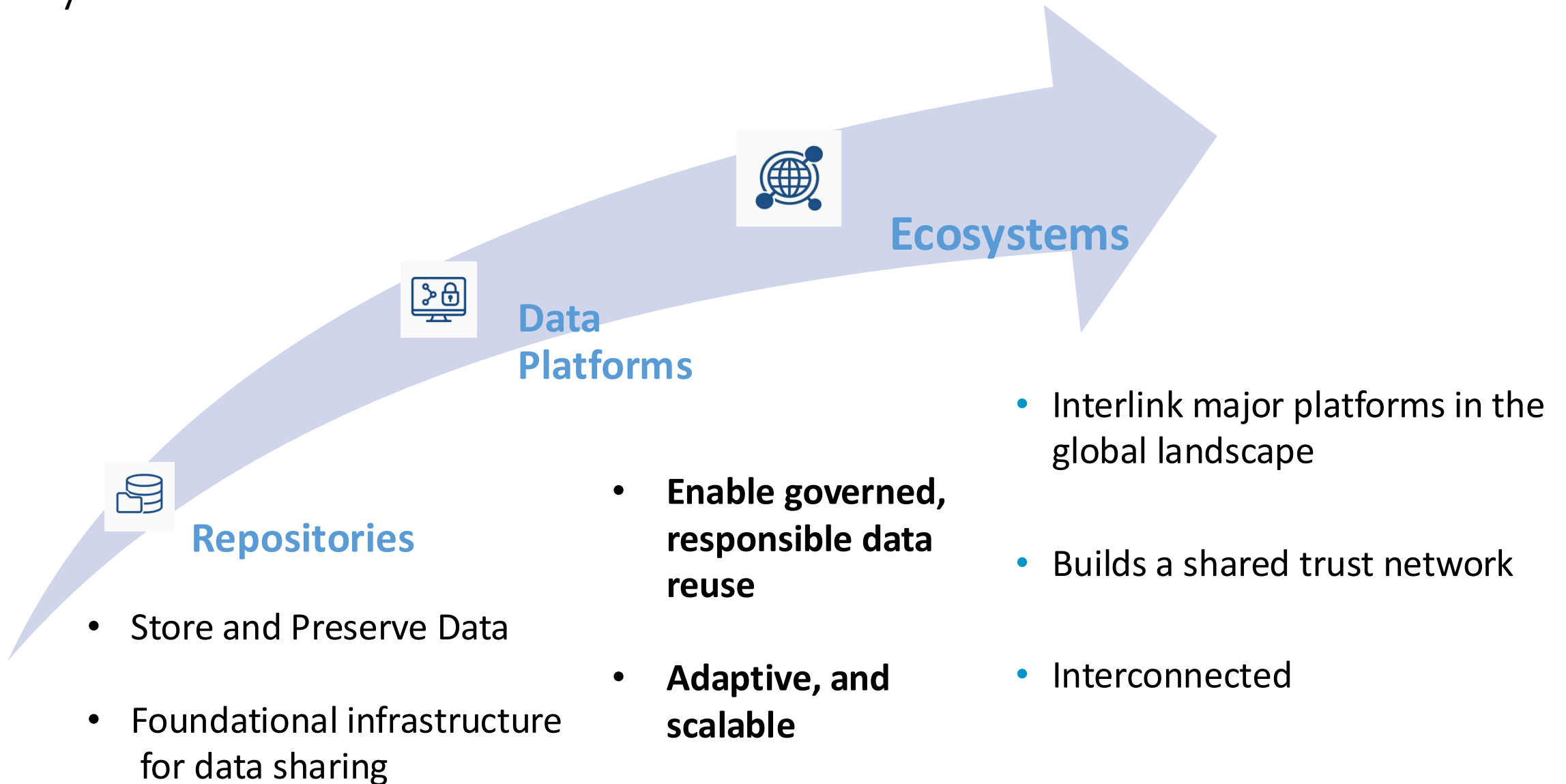
Julie Wood, COO

PHUSE US Connect 2026

Data Transparency Pavillion

March 23, 2026

Evolution of Data Sharing: Building Toward Connected, Adaptive Ecosystems



The Changing Research Landscape



Regulatory Shift

EHDS, GDPR, and new interoperability rules are redefining access and secondary use



Rise of AI/ML

Data are becoming more complex, and the governance of models is now as important as data itself



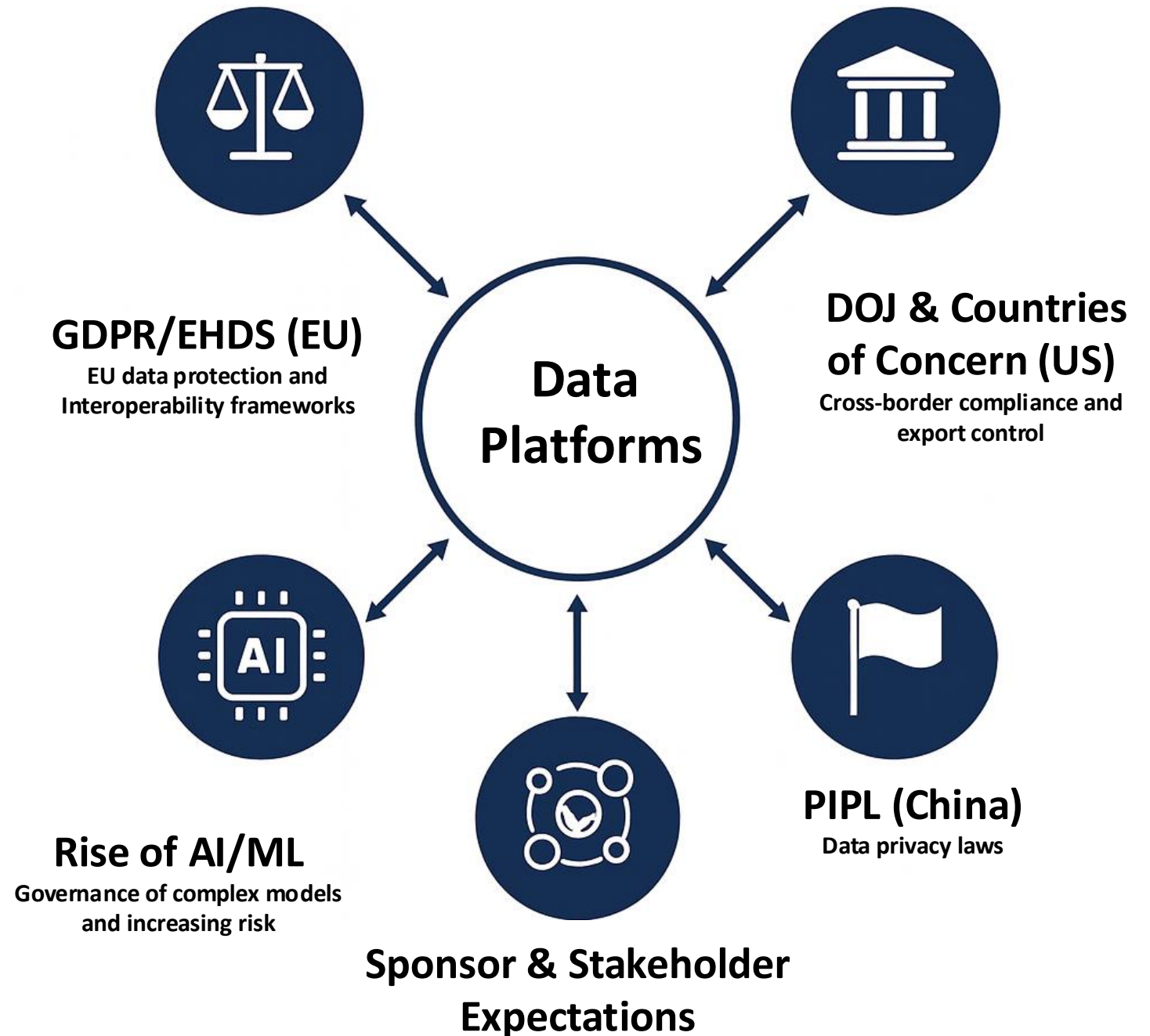
Trust & Collaboration

Research requires trust across borders—between sponsors, regulators, and academia

Adapting to Regional Data Sharing Requirements

- **Establishing regional infrastructure** – Vivli Amsterdam-based non-profit launched to support European collaborations and data sharing
- **Aligning with evolving policy frameworks** – Active engagement with EHDS and TEHDAS 2 shaping secondary use of health data
- **Building local Leadership and presence** – Dedicated leadership to navigate regional regulatory and operational requirements

Examples of Forces Driving Our Vivli Platform Adaptation



Introducing Vivli

THE ENTITY

- Trusted global non-profit
- 8 years' experience facilitating secure, sensitive data sharing
- Support Various Stakeholders
 - Academia
 - Non-profit foundations
 - Government
 - Biopharma
- Governance and Policy
 - Move culture of data sharing
 - Protecting Patient data
 - Promoting incentives for sharing and data re-use

THE PLATFORM

- An established, secure, and flexible data sharing and computing platform
- Serving the international community, including trials from any disease, country, sponsor, funder, or investigator
 - Open search
 - 400 active SPEs at any given time
 - Robust security
 - Modern tools and technologies

Vivli Industry Members

abbvie



KYOWA KIRIN



SHIONOGI

ALFASIGMA



Lilly

SERVIER
moved by you

Alnylam
PHARMACEUTICALS



Lundbeck



astellas



AstraZeneca

GSK



REGENERON

TEMPUS



IONIS



ucb
Inspired by patients.
Driven by science.

Biogen

sanofi



BIONTECH

Johnson & Johnson

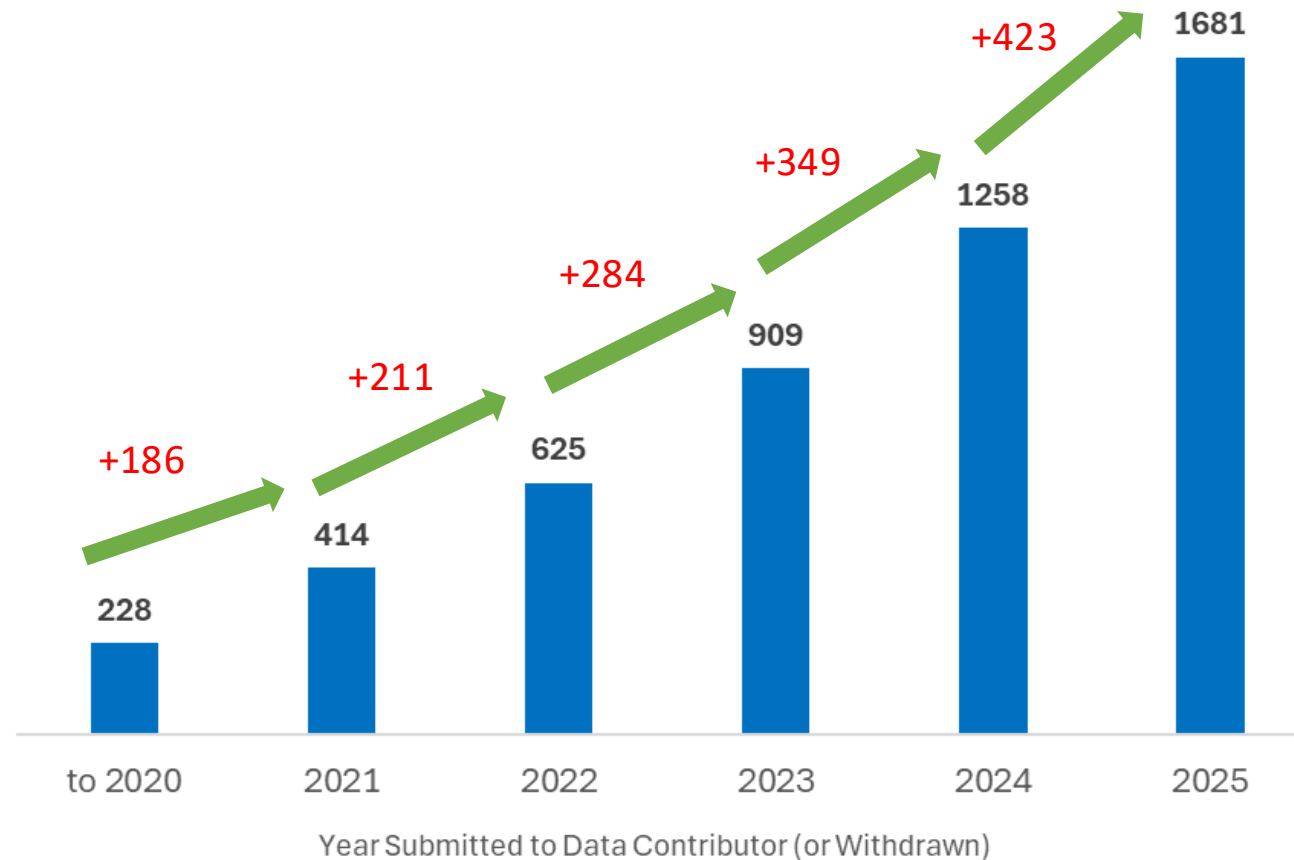
NOVARTIS

Vivli by the numbers ...TODAY



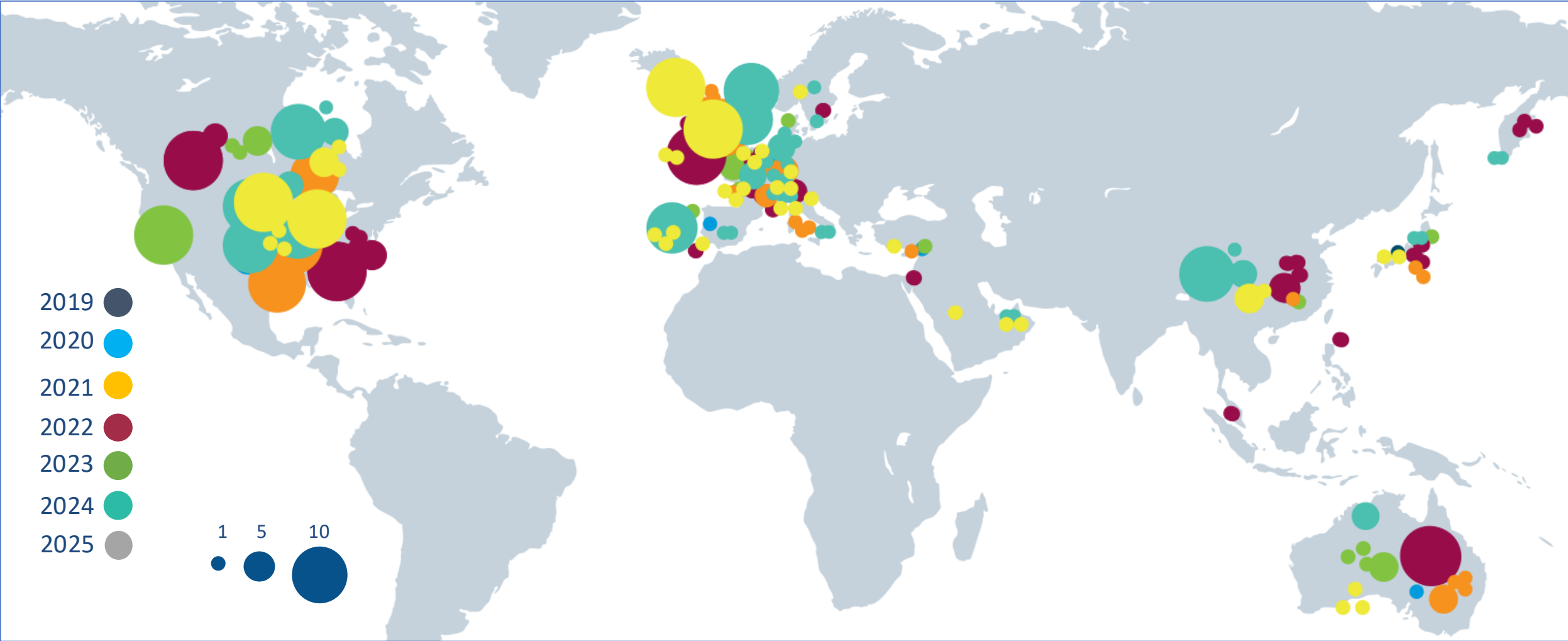
~5 datasets
requested
per project

Data Request Cumulative Totals— Submitted Requests are Increasing Annually



The number of submitted data requests increased by 33%

Data Re-use – Publications from Vivli – Global Reach



Enterprise Impact: Enabling Data Sharing in a Complex Regulatory Environment

- Reduce governance risk through a trusted multi-stakeholder approach
- Manage complexity and operational burden in lean operating environments
- Navigate risk through supporting evolving regulatory policies and responsible use of AI
- Drive forward new science and support multi-sponsor research

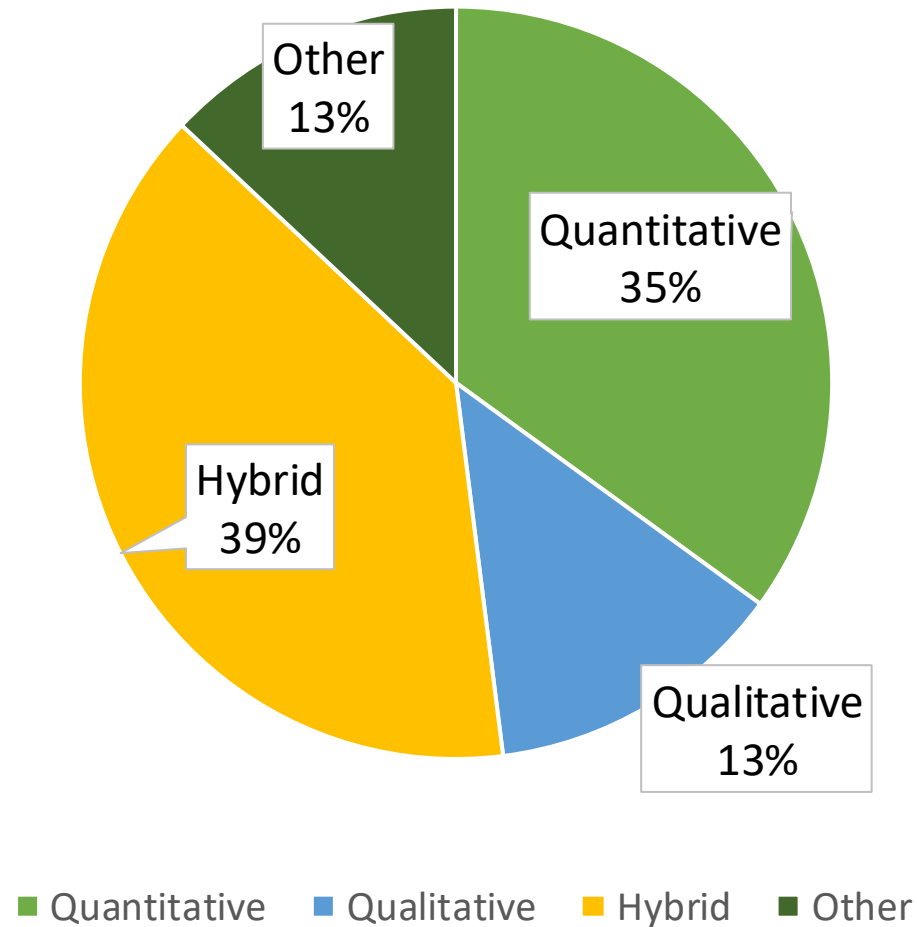
2025 Key Accomplishments – Navigating Data Sharing Compliance Under the DOJ Final Rule on Transfers of Sensitive Personal Data to Countries of Concern

- **Context** - In April 2025, new U.S. DOJ guidance on countries of concern created immediate uncertainty for many members around how to handle ongoing and future data sharing without increasing legal or reputational risk
- **Key accomplishment**
Vivli worked with our legal firm Ropes to interpret the guidance and operationalize a **consistent, compliant approach** to managing secondary use requests impacted by the new requirements

What This Enables

- **Consistent governance across organizations**
A shared approach avoided fragmented or ad hoc responses across contributors.
- **Rapid adaptation without disruption**
Members were able to respond to new guidance without pausing data sharing activities.
- **Clear risk identification and management**
Enabled proportionate controls based on jurisdiction, data type, and access conditions.
- **Enterprise Outcome**
Vivli reduced legal and reputational risk for members across both platforms while increasing regulatory readiness allowing organizations to respond confidently to new requirements without rebuilding internal governance processes.

Do you use a
quantitative,
qualitative, or
hybrid approach
for clinical trial
data
anonymization?



Member Anonymization Survey Results: Takeaways

- Overall, variability in the answers to each of the 16 questions
- Majority of members (65%) are anonymizing datasets one time, not customizing the data per request
- Overall Anonymization is highly variable, resource-intensive, and not standardized across organizations
- Our role is to ensure that researchers understand the anonymization process especially for multi-sponsor requests

Key Takeaways

- Data sharing has expanded beyond access now requiring responsible use in a complex, global environment
- Managing risk is now central to data sharing organizations who must navigate regulatory, operational, and AI-related risks in increasingly lean environments
- Platforms and the organizations behind them play a critical role enabling transparency, governance, and collaboration at scale