Open Science, Data Access, and Reproducibility: History, Current State and Ways Forward for Clinical Trials

Frank W. Rockhold, PhD
Professor of Biostatistics and Bioinformatics

Biostatistics and Bioinformatics
June 4, 2021

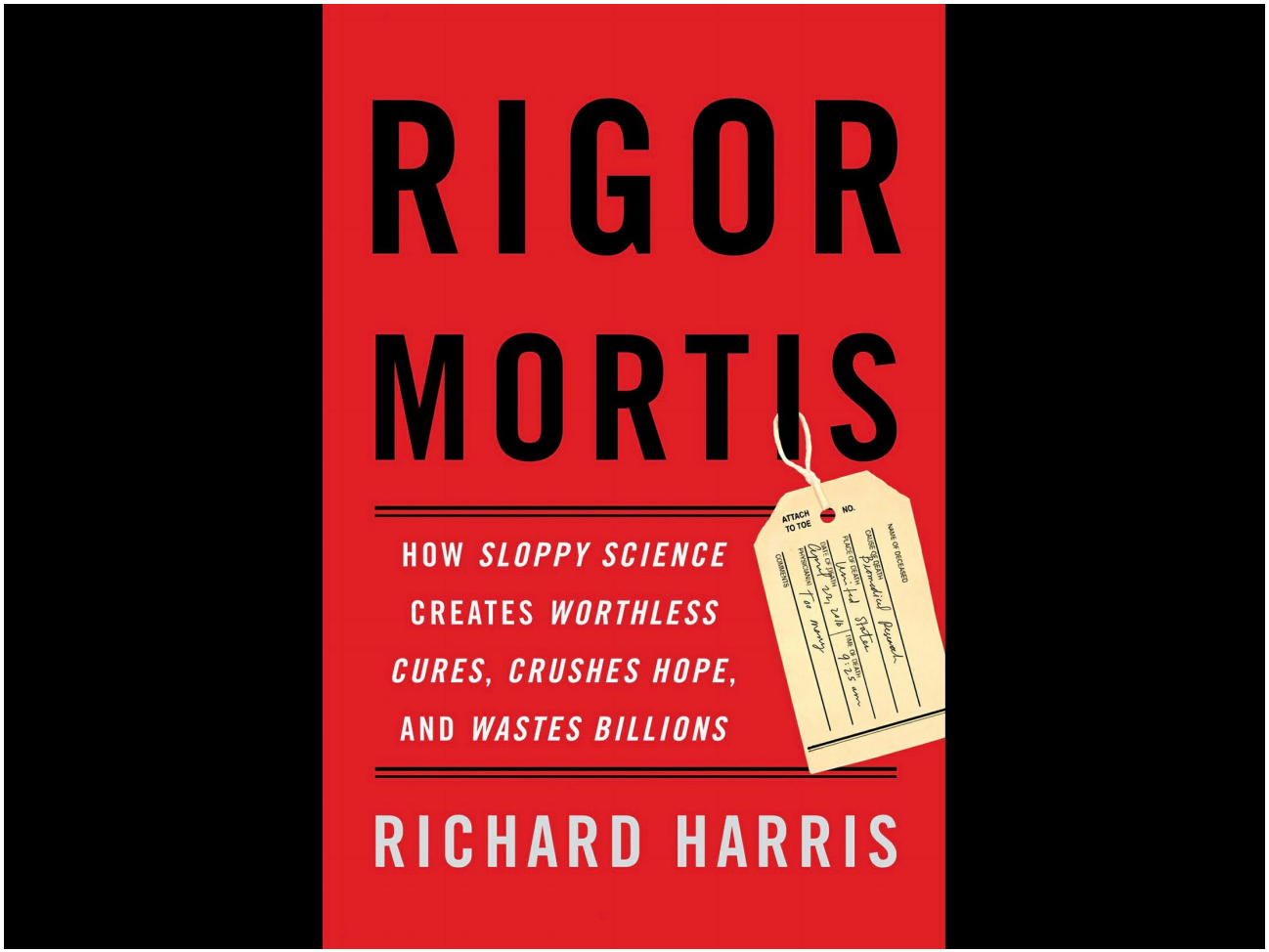
Duke Clinical Research Institute | FROM THOUGHT LEADERSHIP TO CLINICAL PRACTICE

Calls for Open Science

- Calls for greater transparency and 'open data access' in clinical research continue actively.
  - "Open science is the movement to make scientific research, data and dissemination accessible to all levels of an inquiring society"*
  - Open Science Project**: "If we want open science to flourish, we should raise our expectations to: Work. Finish. Publish. *Release.*"
  - FAIR Principles: Findability, Accessibility, Interoperability, and Reusability***

*https://www.fosteropenscience.eu/resources
**(http://openscience.org/)
*** https://www.nature.com/articles/sdata201618

Rigor Mortis



Open vs Transparent vs Access vs Sharing

- What does it mean to be “open” or “transparent” and why is it important?
- Transparency and openness are strategy or belief systems.
- Disclosure and access are actions which are necessary steps on that journey.
- What is the difference between “access” and “sharing”?
- Disclosure or access without transparency, might check a requirements box, but not help patients, healthcare practitioners or other researchers.
- Transparency can only be achieved if people disclose in a manner digestible by the recipient (data utility).

Important Issues Not Covered Today

- Synthetic data
- RWD data sharing
- Machine learning and AI
- Anonymization vs deidentification
- Nonclinical data
- Analysis code access

Open Science and Clinical Trial Data Access

- Some of the challenges are:
  - Patient privacy
  - Academic credit and commercial sensitivity
  - Data utility and standards
  - Resources (money and people)
  - Fear of the “Rogue” analysis
- There should be room for researchers and patients alike to gain from this effort.
- Informatics experts and data scientists are essential elements of this discussion.
- This talk will cover the challenges and imperatives that are at the core of this critical effort.

Quote from Dr. Robert Califf

"When human research participants give consent to participate in research, there is a commitment to create generalizable knowledge. Given the purpose and risk of clinical trials, it is sensible that the data generated from the research should be available for learning.

With openness and transparency of the primary researchers, reciprocal obligations are being delineated for secondary users of the data. Future patients will benefit from optimizing the use of information through appropriate sharing"

Robert Califf, MD, Vice-Chancellor, Director of the Duke Health Science Institute and Professor of Cardiology, Duke University Medical School, and Former Commissioner, US Food and Drug Administration.

The Challenge for Clinical Trials

- **"The tendency for researchers to "sit" on their data for an unduly long period of time is neither desirable from a scientific point of view nor acceptable from an ethical perspective."**
- **"After all, the data belong to the patients who agreed to participate in the research, not just to the investigators who coordinated it, as the new European General Data Protection Regulation emphasizes." ***

*Rockhold, F, et al. Open science: The open clinical trials data journey, *Clinical Trials*, Vol 16 (5) 1-8, 2019

Access to individual patient data (IPD) from clinical trials is important for future research

- There are certainly challenges, the question is not *whether* data should be shared, but rather *how and when.*
- *Responsible* open access enables secondary analyses which:
  - Enhance reproducibility of clinical research
  - Honor the contributions of trial participants
  - Improve the design of future trials
  - Generate new research findings
- This journey of making individual patient data (IPD) available is an evolution and not a sudden awakening.

Clinical Trials: The Journey to “Transparency”

- ICMJE 2005, CT.Gov and WHO ICRTP
- FDAAA 2007
- IOM 2015 Report
- EMA Policy 70
- ICMJE Proposal 2016
- FDA and NIH Final Rules 2016
- ICMJE 2017 Requirements
- US Office of Human Research Protection 2017 Revised Informed Consent Rule
- NLM/NIH Meeting 2017 on Open Science
- AAMC Meeting 2018 on Academic Incentives
- National Academy of Medicine Meetings (2) 2019
- COVID Data Alliance (Gates Foundation) 2021
- ……..

EMA Policy

"As of October 2016, the European Medicines Agency (EMA) publishes clinical data submitted by pharmaceutical companies to support their regulatory applications for human medicines under the centralised procedure. This is based on EMA's flagship policy on the publication of clinical data."

European Medicines Agency Policy 0070

International Committee of Medical Journal Editors (ICMJE) Requirements*

- The ICMJE expects that the Data Sharing Statement and the Data Sharing Plan will include the items listed below.
  - Whether individual de-identified IPD (including data dictionaries) will be shared
  - What data will be shared
  - Whether additional, related documents will be available
  - When the data will become available and for how long
  - What access criteria will be used to decide if data will be shared (e.g., with whom, for what types of analyses, and by what mechanism).

*Taichman DB, et al. Data sharing statements for clinical trials: a requirement of the International Committee of Medical Journal Editors. Ann Intern Med. 2017;167:63–5.

ICMJE Continued…

- **Thus, if the authors of a manuscript are not prepared to share their data, a short statement, such as, "Data will not be shared", should satisfy the new requirements.**
- **Nevertheless, …. the authors' response to Data Sharing Statement may affect the editorial decision.**
- **"The technical and statistical challenges of accessing research data for re-analyses and other secondary uses are not trivial." ***

*Thomas and Paarlberg, ICMJE requirements for data sharing Individual participant data from interventional clinical trials. June 2019 *Medical Writing* | Vol 28 Number 2

The Patient's View- NEJM

The NEW ENGLAND JOURNAL of MEDICINE

SPECIAL ARTICLE

Clinical Trial Participants' Views of the Risks and Benefits of Data Sharing

Michelle M. Mello, J.D., Ph.D., Van Lieou, B.S., and Steven N. Goodman, M.D., Ph.D.

ABSTRACT

BACKGROUND

Sharing of participant-level clinical trial data has potential benefits, but concerns about potential harms to research participants have led some pharmaceutical sponsors and investigators to urge caution. Little is known about clinical trial participants' perceptions of the risks of data sharing.

METHODS

From the Department of Health Research and Policy, Stanford University School of Medicine (M.M.M., V.L., S.N.G.) and Stanford Law School (M.M.M.) — both in Stanford, CA. Address reprint requests to Dr. Mello at Stanford Law School, 559 Nathan Abbott Way, Stanford, CA 94305, or at mmello@law.stanford.edu.

Patient Perspective

- Imperative to put patient data in the public domain: "Fail fast and fail once. This is particularly important in early development."

  Donna Cryer, Global Liver Institute

- Subjects aren't in favor of keeping their data locked up in the files of the doctor doing the study. "If you ask the patient is it OK to share your data with every scientist who's working on your type of cancer, of course they'll say yes. That's why they're doing it. But they [researchers] don't ask that question! I'd like to see that change."

  Steven Salzburg Johns Hopkins University. *(excerpted from Rigor Mortis)*

- Patient panels expressed consensus that their data should be shared (with their privacy protected) and made available and patients were *shocked* that there was any debate over this issue.

  NEJM Meeting on Data Sharing 2017

Creating Access to IPD Generates Value

- **Data access** is the practice of making scientific, clinical or other data available to other investigators, researchers or the public and:

  - ***Enables new discovery*** and new research questions through using and combining existing data with increased statistical power.
  - ***Validates*** existing research results by peer review and reanalysis.
  - ***Broadens*** research by enabling appropriate aggregation of data derived from disparate data generators.
  - ***Prevents*** repetitive trials and patients at risk when data already exist.
  - ***Targets*** analysis of benefit to risk of therapies in subgroups ("precision medicine").
  - ***Strengthens*** trust in clinical research through enhanced transparency.
  - ***Honors*** the commitment to patients who volunteer for trials.

Main Considerations Around IPD Access

- Safeguarding the privacy of research participants
  - Continues to evolve. Impact of GDPR, etc.
- Providing data in ways that enable external researchers to understand and navigate the information- data utility
- Creating incentives for academic researchers to share
- Balancing the needs of the data generators and secondary researchers
- Ensuring the data are used for scientific purposes?
- Protecting commercial interests?

Data Utility EMA

- **EMA POLICY 0070**
  - **4.3. Data utility: Different anonymisation techniques will lead to different levels of data utility in the anonymized reports. Applicants/MAHs should take into consideration the impact of the data transformations/redactions on the scientific usefulness of the data.**

Data Utility Definition

"A summary term describing the value of a given data release as an analytical resource. This comprises the data's analytical completeness and its analytical validity. Disclosure control methods usually have an adverse effect on data utility. Ideally, the goal of any disclosure control regime should be to maximize data utility whilst minimizing disclosure risk. ***In practice disclosure control decisions are a trade-off between utility and disclosure risk."****

*OECD: Organization for Economic Co-operation and Development

A First Step in 2013

- In May 2013, GSK launched a system to provide greater access to anonymized patient level data from our clinical trials.

2021: Numerous Other Platforms in Place!

- Clinical Study Data Request: multi-sponsor request site (13 companies), managed by the Wellcome Trust
- YODA: Yale Open data Access for two sponsors (Janssen/Medtronic)
- Project Data Sphere (CEO roundtable on cancer)
- INSPIIRE : Integrated System for Pfizer Investigator Initiated Research
- SOAR: Bristol Myers Squibb and Duke Data Strategic Initiative (DCRI)
- Celgene's Clinical Trial Data Sharing (Now part of BMS system)
- NIH BioLiNCC
- Vivli.org
- And many, many, others either in place or in development
- So good news and in some ways but a fractured, disconnected approach

Spectrum of Data Sharing Models

Immune Tolerance Network- Trial share

- Open access to ITN data after registration and agreement to terms of use
- No further approval process
- Downloadable data

ClinicalStudyDataRequest.com

- Multiple industry sponsors; governance ranges by sponsor
- Secure interface, DUA, IRP
- IRP considers scientific relevance, COIs, and investigator expertise
- Some sponsors may review requests, and veto based on data specific considerations, competitive risk etc.

SOAR

- DUA, IRP
- Evaluates for COIs and research quality
- Requirement for detailed statistical analysis plan, evaluated for major design flaws
- Final analyses are reviewed by the IRC prior to publication

AHA Precision Medicine Initiative

- CV and stroke data
- Cloud-based, secure sharing environment
- Forum for collaboration
- Data access is granted in private workspaces by data contributor

Project data sphere (PDS)

- Oncology research
- Downloadable data
- DUAs
- Open to all
- **Control Group Only**

Vivli

- Attempting to harmonizing data sharing governance
- Secure interface, DUAs, IRP
- Review process considers the research plan, team, statistician, and COIs
- Contributors can veto requests, but number and reasons for rejections will be made public

YODA

- Generally, data is not downloadable
- DUA, IRP
- Data requests evaluated for scientific merit and COIs
- Restrictions to data access for legal or commercial purposes

Open Access

Restricted Access

Data Sharing at a Crossroads

The NEW ENGLAND JOURNAL of MEDICINE

HOME | ARTICLES & MULTIMEDIA | ISSUES | SPECIALTIES & TOPICS | FOR AUTHORS | CME

Perspective

Data Sharing at a Crossroads

Frank Rockhold, Ph.D., Perry Nisen, M.D., Ph.D., and Andrew Freeman, B.Sc.

N Engl J Med 2016; 375:1115-1117 | September 22, 2016 | DOI: 10.1056/NEJMp1608086

Comments open through September 28, 2016

Share:

Article | References | Citing Articles (4) | Comments (2) | Metrics

Sharing patient-level data from clinical trials can improve the quality of research and our understanding of disease and medical treatments. Various concerns have been voiced about data sharing; they involve privacy, consent, intellectual property, costs, infrastructure, data standards, free-riding researchers, and potentially erroneous conclusions. Many of these concerns cannot be totally eliminated, but they can be mitigated and managed.

The clinical research community is at an important crossroads. We believe that sharing data is the right thing to do and that we need to find the best ways to realize the benefits while minimizing the risks. Multiple different approaches and systems may be creating a fragmented, complex, and confusing landscape in which data sharing's full benefits will not be realized.

AUDIO INTERVIEW

Interview with Dr. Jeffrey Drazen on the future of sharing clinical trial data. (9:31)

Listen

Download

Incentives for Data Generators*

- Clinical trialists could be given incentives to share data. Trialists could receive appropriate acknowledgment when other researchers use “their” shared data to
- Secondary investigators analyzing shared clinical trial data should provide a research question and data-analysis plan when requesting data access, publish their findings
- Data sharing does not have to be a zero-sum game where trialists lose if others perform secondary analyses. Trialists and secondary investigators could be collaborators rather than antagonists

*Lo, B. and DeMets, D.L., *Incentives for Clinical Trialists to Share Data.* NEJM, 2016. **375**(12): p. 1112-1115
*Bierer, B.E., *Academic Incentives for data sharing.* New England Journal of Medicine, 2017.

Goals and Principles of Open Science at Duke

- Advance Science
  - Answer new questions
  - Combine data to increase it's power
  - Avoid duplication of effort
  - Foster new connections and collaborations
- Improve Research Integrity
  - Validate original analyses
  - Promote transparency and trustworthiness

- Principles
  - Facilitate appropriate access
  - Ensure appropriate governance
  - Maintain data utility
  - Share results (and data) from secondary analyses
  - ***Acknowledge and provide academic credit for those who share original data***

The open science roadmap was introduced by Adrian Hernandez, MD

https://medschool.duke.edu/about-us/news-and-communications/med-school-blog/introducing-duke-school-medicine-roadmap-open-science

SOAR

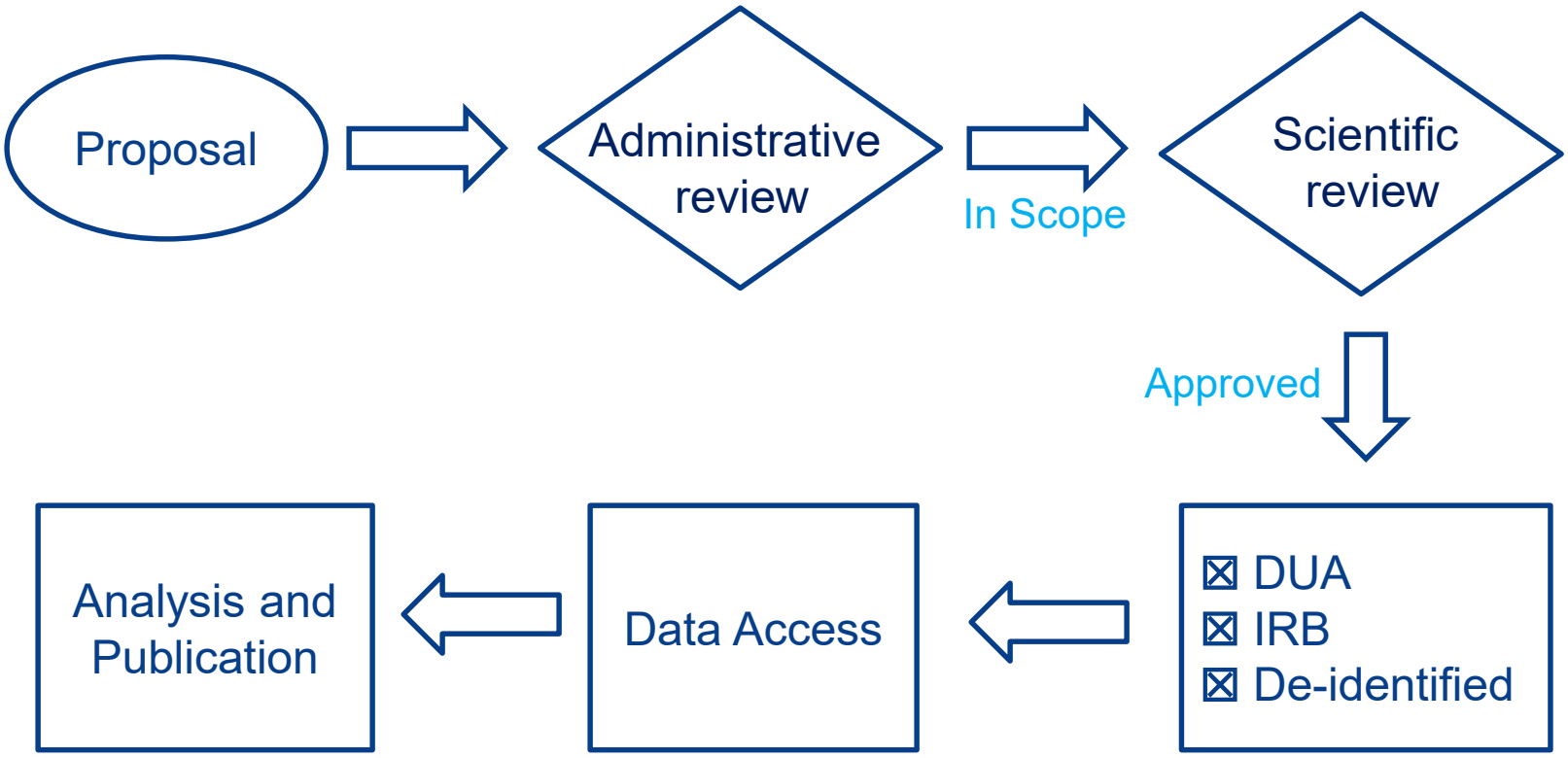
……Through SOAR, DCRI offers access to:

1. **An Independent Review Committee (IRC) with subject matter experts**
   a) Duke Clinical Research Institute (DCRI) & Bristol-Myers Squibb (BMS)
      a) BMS shares its Clinical Trial datasets
      b) DCRI's academic experts review proposals

2. **DCRI Datasets**
   a) **Duke Cardiac Catheterization Research Dataset** (DukeCath) (de-identified)
   a) **Duke Cardiac Catheterization Educational Dataset** (DukeCathR) (anonymized and modified)

3. **Other Data Sharing Resources**
   a) Links to the Aggregate Analysis of ClinicalTrials.gov (AACT) database & dataset
   b) Links to request datasets shared through other platforms (i.e., NIH/DASH)
   c) DCRI Dataset Catalog

SOAR Research Dataset Request Process



* Feedback cycle is activated when IRC requests clarification or additional information from researcher

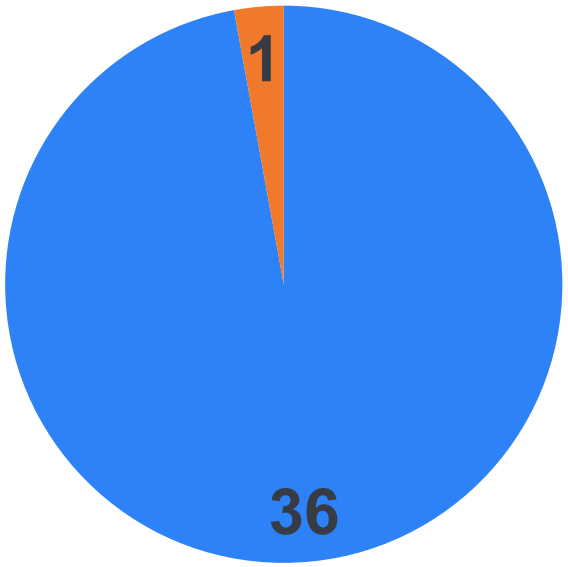
What Information Should Be Made Available?

- The following deidentified or anonymized data and documents (with personally identifiable information redacted) can be made accessible in a secure system:
  - Raw dataset
  - Analysis-ready dataset
  - Protocols with any amendments
  - Annotated case report form
  - Reporting and analysis plan
  - Dataset specification
  - Redacted (for privacy) Clinical Study Report including modular appendices (potentially identifiable information, including patient level data and patient narratives are removed)

Who's Asking for Data?

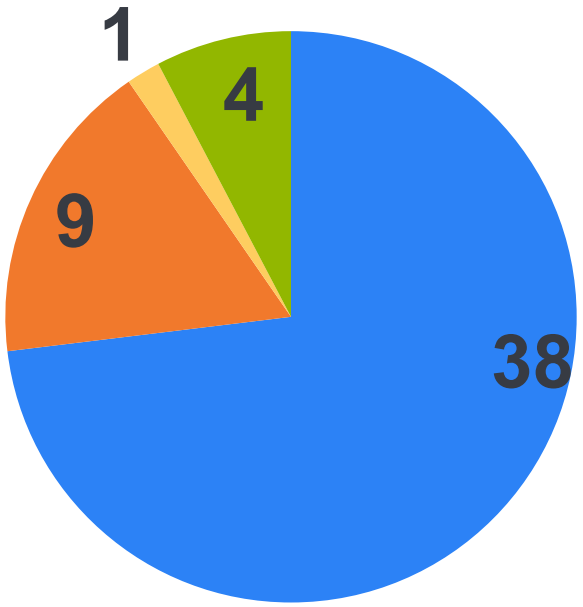
Reviewed BMS-DCRI Proposals

n = 37



DukeCath Datasets

n = 50



Academia Industry Non-profit Other

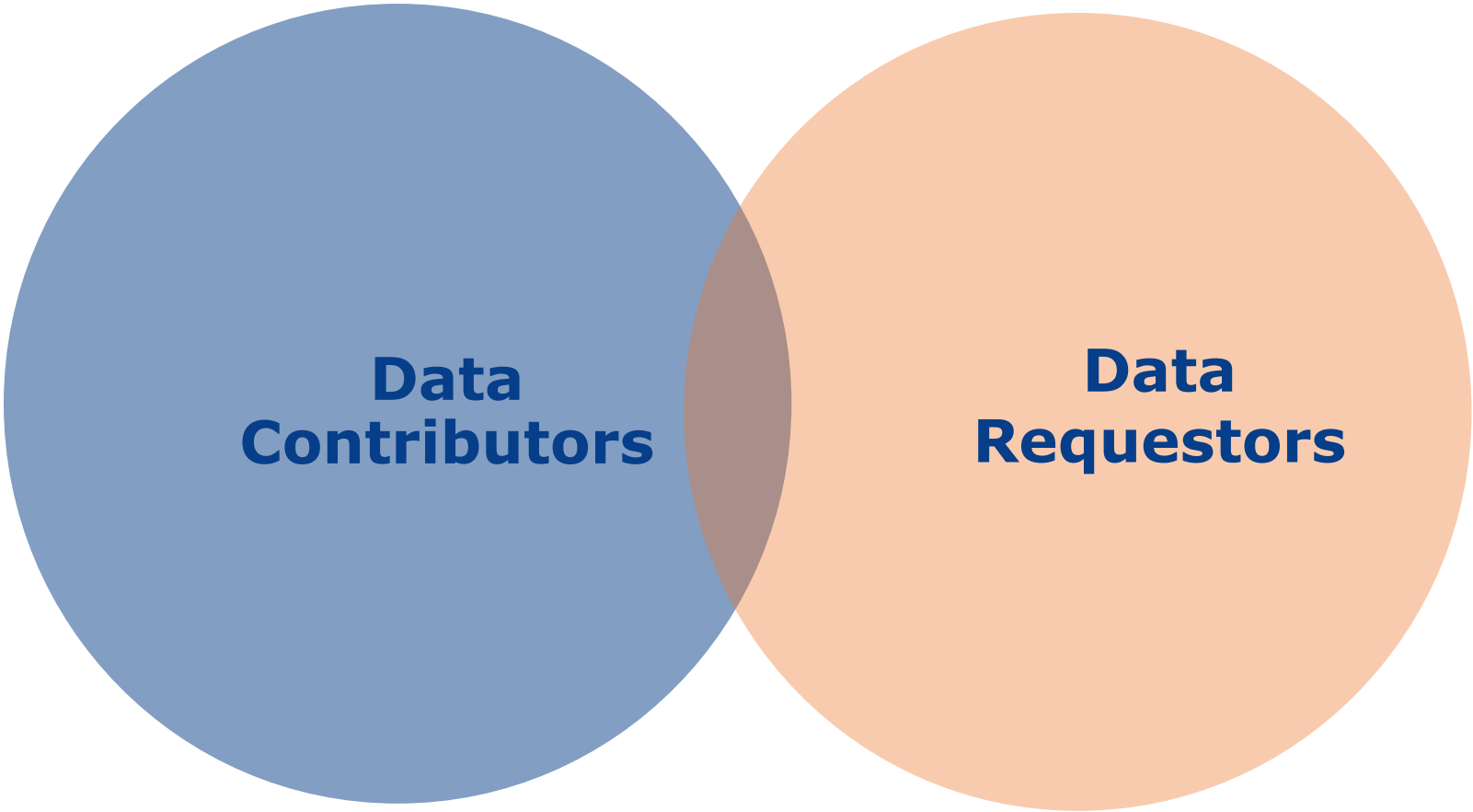
One Potential Solution: A trusted global neutral data sharing platform or portal

- Minimize risks to all stakeholders
- Ensure maximum participant safety and privacy
- Allow participation of academia, industry, government, non-profits, and others
- Inventory of trials can be searched in one place and Integrated search functionality
- Enable transparent and inclusive governance
- Provide interoperability of data
- Ensure Data security

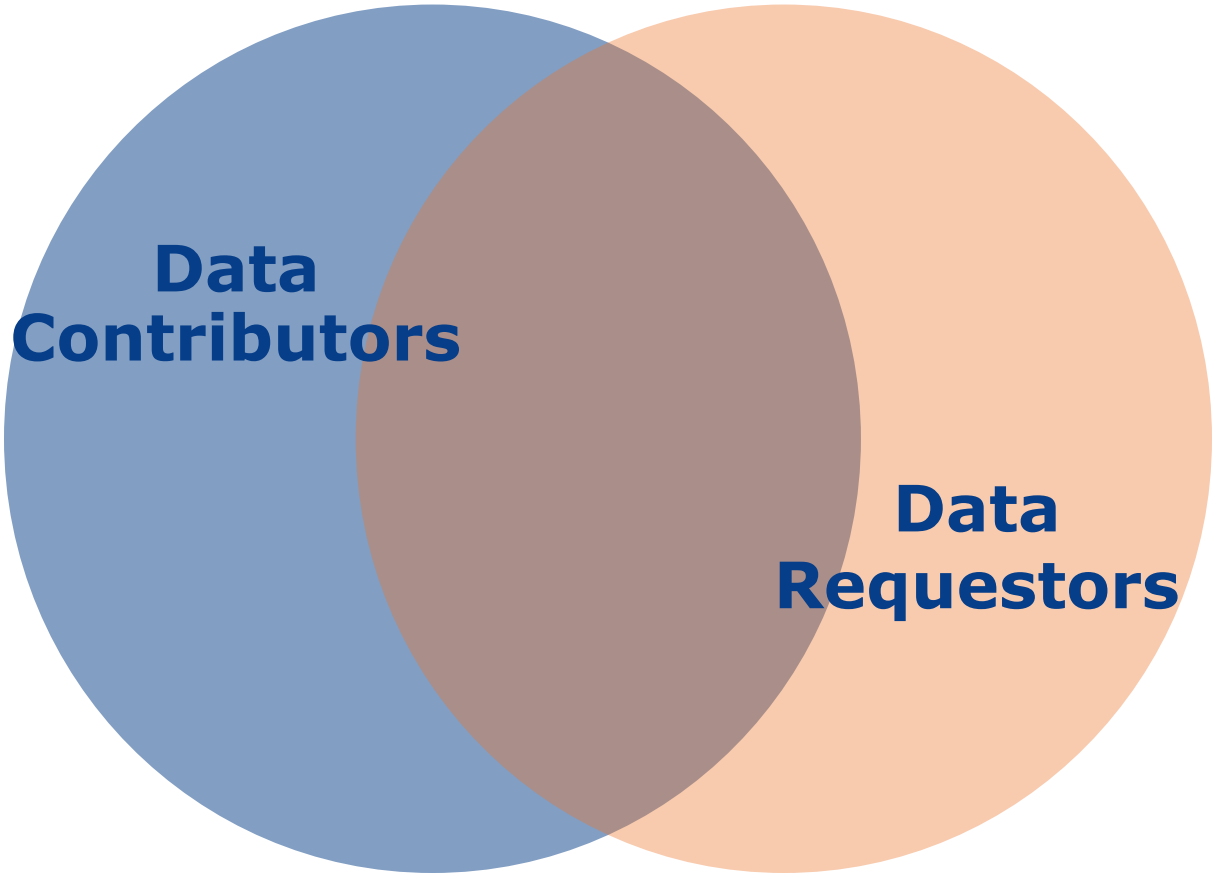
The Vivli Solution (in partnership with Microsoft and The Cochrane Collaboration)

- Independent non-profit organization bringing together multiple stakeholders in a neutral entity.
- Harmonized policies and processes for data submission, requests, access, and academic credit
- More user-friendly data sharing platform providing:
  - easy process for cataloging which studies are available for sharing
  - powerful precise searching of study features to identify studies of interest
  - Automated data set and citation tracking

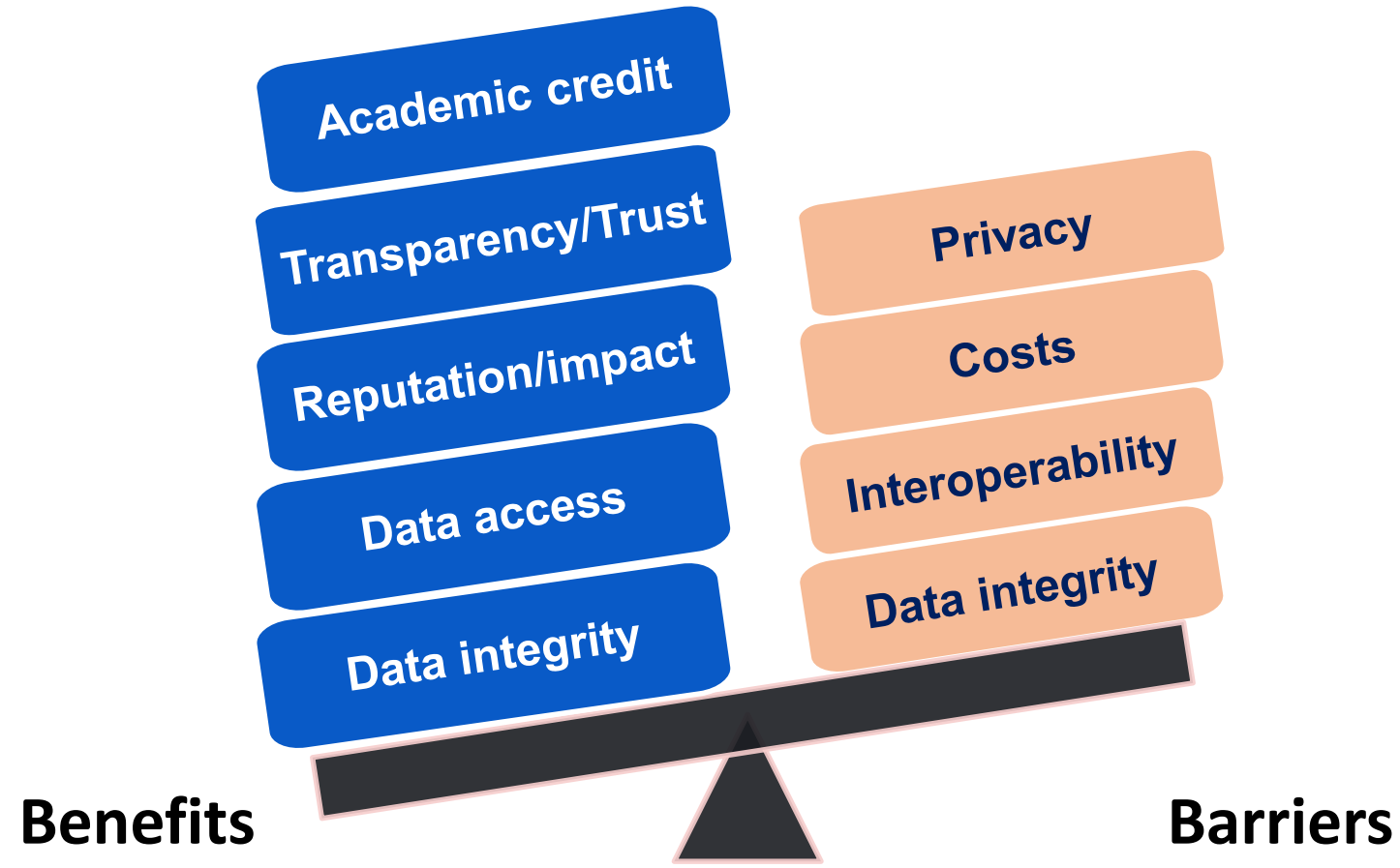
Data Sharing: Current State



Data Sharing: Desired Future State



Benefits & Barriers: Desired Future State



Suggested Guidelines for Data Sharing

1) Plan for data sharing and involve patients--
   a. Generate a plan as early as possible to make the data accessible
   b. Ensure that consent documents do not prohibit data sharing
2) Protect patient rights and make end users responsible-- Protect the rights of patients during every aspect of data sharing by minimizing their risk
3) Credit data source-- Give credit to the data source in any presentation or publication of analyses using shared data.
4) Acknowledge funding-- The source of funding should be acknowledged in any presentation or publication of analyses using shared data.
5) Rare diseases and unique subgroups-- Ensure patient privacy and data utility are kept in balance

Key Summary Points

- We owe it to patients who enroll in trials to make their data available.
- We need to view ”Data Access” effort as a journey that benefits patients and meets other requirements from journals and legislators.
- There are manageable challenges and risks:
  - Understanding of benefits “Openness” and “Transparency”
  - Patient Privacy and Informed Consent
  - Data Utility
  - Data Standards and interoperability
  - Incentives for researchers
  - Proof of value
- Engagement of statisticians, informatics experts, and data scientists is critical for this to yield value to society and patients.
- Promoting data access promotes better data stewardship up front.

Questions and Discussion