Responsible data-sharing to improve research integrity

November 2020
Julie Wood, Vivli
Vivli: The Difference Data Re-Use Can Make
Balancing Risks and Benefits

Openness
Maximizing the value of the data collected respects participants' contributions

Privacy
Protecting participant privacy, particularly when sharing individual patient-level data (IPD)
## Approaches to Sharing Human Data

<table>
<thead>
<tr>
<th>Type</th>
<th>Key Features</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open access</td>
<td>Anyone can access, simple account creation, simple on-line Data Use Agreement (DUA)</td>
<td>Health CAN, EMA, Project Datasphere</td>
</tr>
<tr>
<td>Managed access</td>
<td>• for scientific purposes only (standard request form)</td>
<td>Vivli, CSDR, SOAR, VISTA</td>
</tr>
<tr>
<td></td>
<td>• (independent) review process</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• secure environment for data access</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• clear legal framework</td>
<td></td>
</tr>
<tr>
<td>Restricted access</td>
<td>Invitation only, access only to those who provide data</td>
<td>DataCelerate, IBD Plexus</td>
</tr>
</tbody>
</table>
Challenges with Data Sharing

HEALTH

Scientists Have a Sharing Problem

Competition and disorganization within their disciplines prevent many researchers from making their data publicly available, which is stunting scientific progress.

MAGGIE PUNIEWSKA  DEC 15, 2014

When it comes to sharing information, there seems to be quite a difference of opinion—across areas both trivial and serious—as to how much is enough. Some people broadcast their lives on Facebook; others poke fun at the passive, private data collector.
Setting your data free

As science becomes more open, researchers who share data are reaping the benefits.

BY GABRIEL POPKIN

Ecologist Thomas Crowther knew that scientists had already collected a vast amount of field data on forests worldwide. But almost all of those data were sequestered in researchers' notebooks or personal storage drives. Now, in CSV files (plain-text files that contain a list of data) on servers at Crowther's present laboratory at the Swiss Federal Institute of Technology in Zurich and on those of a collaborator at Purdue University in West Lafayette, Indiana, he hopes to outsource database storage to a third-party organization with expertise in managing large-scale data. He calls his project Open Earth, and he has tried to make it answer the following questions:

- How much has the Earth's atmosphere changed over the past century?
- How much of that change is attributed to human activity?
- How much of the world's forests have been cleared? What is the current state of science on climate change?

Crowther's project is partly open, partly closed, and with unclear and inconsistent policies and expectations on data sharing that are still in flux. High-level bodies such as the US National Academies of Sciences, Engineering, and Medicine and the European Commission have called for science data to be made FAIR (findable, accessible, interoperable, and reusable). His project is an attempt to reach that goal.

"A love letter to your future self": What scientists need to know about FAIR data
Challenges with Data Sharing

• “The secondary requester will be unfamiliar with the data set structure or analysis unduly alarm the public or hinder science rather than advance science”
• “What if they find something a mistake in my trial dataset?”
• Providing academic credit
• Clinical trials take years to design, conduct and analyze is this fair?
• “Won’t this stifle innovation and new science?”
Developing a data sharing program

1. Why should we share?
2. What are the key components of a data sharing program?
3. When should we begin a program?
4. How can we manage a data sharing program?
5. What is the role of a partner such as Vivli?
1. Why Should My Organization Share Its Data?

- Ethical obligations to trial participants
- Journal requirements
- BIO, EFPIA, PhRMA, IFPMA publicly stated commitments for members
Patients Expect Data Sharing and Reuse

Roxana Mehran @Drroxmehran · 4 Apr 2017
trial participants-"share the data as widely as possible and as soon as possible to advance human health" #NEJMDataSummit #cancer

Vinay Prasad MD MPH @VPplenaryshesh · 4 Apr 2017
Patients listened to trialists fears for one day and then it's supposed to go #nejmdatasummit

Anna McCollister @annamcslipp · 4 Apr 2017
@JeffDrazen living in time where "trust me I'm a doctor I have data out there & let people see themselves #I

Sharon F. Terry @sharonfterry · 4 Apr 2017
Love idea that next generation is open to openness-will we watch people die meanwhile? Do we have appetite for such waiting? #NEJMDataSummit

P. F. Anderson @pfanderson · 4 Apr 2017
OUTCOMES of patient panel > Share early, often, with me, responsibly, understandably #NEJMDataSummit

Aaron Eisman @aaroneisman · 4 Apr 2017
Patients incredulous that sharing data isn't the norm, speaking loud and clear: "share my data!" #NEJMDataSummit
What Do Surveys Show Regarding Patient and Participant Preference Regarding Data Sharing?

• High levels of support for data sharing; however, patients are reluctant to have their data “commodified” purely for commercial gain ¹

• If adequate safeguards were in place, trial participants are willing to share their data²


Not only is it the right thing to do, it also helps increase citation rate

“The 48% of trials with publicly available microarray data received 85% of the aggregate citations. Publicly available data was significantly (p = 0.006) associated with a 69% increase in citations”

https://doi.org/10.1371/journal.pne.0000308
Trial Registration

- Data sharing plan is part of the ClinicalTrials.gov registration record

12. IPD Sharing Statement

Plan to Share IPD
Definition: Indicate whether there is a plan to make individual participant data (IPD) collected in this study, including data dictionaries, available to other researchers (typically after the end of the study). Select one.
- Yes: There is a plan to make IPD and related data dictionaries available.
- No: There is not a plan to make IPD available.
- Undecided: It is not yet known if there will be a plan to make IPD available.

IPD Sharing Plan Description
Definition: If Plan to Share IPD is "Yes," briefly describe what specific individual participant data sets are to be shared (for example, all

- As of 1 January 2019, ICMJE requires registration of your data sharing plan at time of trial registration
What are Journals Requiring as of July 1, 2018?

- Trial manuscripts must be submitted with a data sharing statement
  - Must describe how you will share your Individual participant-level data (IPD), including who, what, when, where, and why
- IPD sharing is not (yet) required but “editors may take into consideration data sharing statements when making editorial decisions”

Providing Access to IPD Generates Value

- **Honors** the commitments of participants
- **Strengthens trust** in the clinical research enterprise
- **Prevents repetitive trials** and putting additional patients at risk
- **Enables new discovery** and scientific insights through by combining data from disparate sources
Perhaps most importantly for participants if the data is not shared...

It is used only one time to answer one question (the primary endpoint) rather than leveraging participants’ contributions to answer multiple scientific lines of inquiry thereby advancing science.
Developing a data sharing program

1. Why should we share?
2. **What are the key components of a data sharing program?**
3. When should we begin a program?
4. How can we manage a data sharing program?
5. What is the role of a partner?
3. How to Share: 3 key elements to consider

- POLICY
- MECHANISM
- RESOURCES
Data sharing governance

Key considerations when formulating your policy

- In data sharing, transparent decision-making equals good public policy
- Data Sharing Policies vary based on a sponsor or funders current portfolio, experience with data sharing and risk tolerance
1. Which studies will you share?
2. Are there exceptions to sharing?
3. On a specific request, who makes the final decision on whether to share?
1. What studies will you share?
   - From when? (Date)
   - Which Phases?
   - Both Submitted and Approved products?
1. What studies will you share?
   - From when?
   - Which Phases?
   - Both Submitted and Approved products?

Example 1: we will share interventional clinical trials conducted in patients (Phase I-IV) for products and indications submitted and approved, since 2002 will be shared.

Example 2: Company-sponsored studies supporting indications approved in both the United States and European Union after January 1, 2013 will be considered for sharing.
2. Exceptions to sharing

- Practical constraints
- Legal or contractual constraints
- Language
- Anonymization
2. Exceptions to sharing
- Practical constraints
- Legal or contractual constraints
- Language
- Anonymization

Examples:
- Studies where there is reasonable likelihood that patients’ anonymity cannot be maintained. For example, in very rare diseases, studies with very low patient numbers or studies performed at a single center.
- There are practical constraints to providing the data (for example, issues related to the format of the databases, and/or resources (costs) are considerable to retrieve data)
- Study documentation is not in English.
1. What studies will you share?
2. Are there exceptions to sharing?
3. On a specific request, who makes the final decision on whether to share?
3. On a specific request, who makes the final decision on whether to share?

- Internal Approving Entity
- External Review Panel (ERP)
- Independent Review Panel (IRP)
Agenda

1. Why should we share?
2. What are the key components of a data sharing program?
3. **When should we begin a program?**
4. How can we manage a data sharing program?
5. What is the role of a partner?
3. Recommendation - When to share what

**Trial Registered**
- Data sharing plan at registration

**Study Completed/Terminated**

**Publication**
- 6 months after publication, share post-publication data package – this includes the data that underlies the tables, figures, graphs in the paper (typically a subset of the entire dataset)

**Regulatory Application**
- 18 months after product abandonment
- OR 30 days after regulatory approval share post-regulatory data package

Recommendations based upon IOM report *Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk*, Jan. 2015
3. Recommendation - When to share what

At least 18 months before regulatory submission or a major publication is when institutions should begin their data sharing program planning.

Recommendations based upon IOM report *Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk*, Jan. 2015.
What data will be shared?

<table>
<thead>
<tr>
<th>Item</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Recommended Set</strong></td>
<td></td>
</tr>
<tr>
<td>Study protocol</td>
<td>Final protocol with all amendments</td>
</tr>
<tr>
<td>Data dictionary</td>
<td>Detailed descriptions of each variable in the dataset, including the definition, source, coding, etc. of the variable</td>
</tr>
<tr>
<td>Statistical Analysis Plan</td>
<td>Description of the principal features of the analyses described in the protocol</td>
</tr>
<tr>
<td>Clinical Study Report (CSR)</td>
<td>Report that summarizes the efficacy and safety data from the study (after regulatory decision)</td>
</tr>
<tr>
<td>IPD dataset</td>
<td>Final cleaned individual participant-level data, de-identified/anonymized</td>
</tr>
<tr>
<td><strong>Optional</strong></td>
<td></td>
</tr>
<tr>
<td>Analytic code</td>
<td>Software code used to carry out prespecified and additional analyses</td>
</tr>
<tr>
<td>Analysis ready IPD dataset</td>
<td>Dataset in a format used to carry out a sponsor’s analyses</td>
</tr>
<tr>
<td>Case report forms</td>
<td>Forms used to collect the data that is described in the protocol for each trial participant</td>
</tr>
</tbody>
</table>
4. How can we manage a data sharing program?

- **Manage in-house:**
  - *Mechanism for sharing* – build, management and updating of a platform
  - *Team* – internal resources to maintain the platform; negotiate legal agreements; user queries, generate metrics, data anonymization and data preparation
  - *Policy* – draft and manage data sharing policies

- **Trusted partner to manage and assist with:**
  - Mechanism
  - Team
  - Policy
5. What can partners like Vivli do for us?
Introducing Vivli

THE ENTITY

• Non-profit organization
• Convening function
  - Biomedical industry (pharma, bio, device)
  - Academia
  - Non-profit funders and foundations
  - Government (funders and regulators)
  - Patient/patient advocates
• Governance and policy
  - Harmonizing language & agreements
  - Move culture of data sharing
• Advocacy
  - Lowering barriers
  - Promoting incentives
• Oversight of Implementation

THE PLATFORM

• A user-friendly, secure, state-of-the art data sharing and computing platform
• Serving the international community, including trials from any disease, country, sponsor, funder, or investigator
  - Open search
  - Robust security
  - Modern tools and technologies
Vivli by the numbers ...TODAY

5,700+ Trials

3.0M Participants from 119 countries

29 Members
Vivli Members

AstraZeneca
Boehringer Ingelheim
Biogen
BioLINCC
Celgene
CRITICAL PATH INSTITUTE
Cure Duchenne

Daiichi-Sankyo
Duke University
Harvard University
Helmsley Charitable Trust
IMMPORT

Johns Hopkins University
Johnson & Johnson FAMILY OF COMPANIES
Lilly
Lundbeck
Mitsubishi Tanabe Pharma
Pfizer
Roche
Taiho
Tempus
UCB
UCSF
Vivli

Project Data Sphere
Regeneron
Takeda

Inspired by patients. Driven by science.
Vivli is a Global Data Platform – Agnostic to Disease, Funder or Data Contributor
How Vivli works

**SEARCH**

Search Vivli platform for information about available studies.

**REQUEST**

Request IPD Data sets. Each Data Request will be **reviewed** according to contributors' publicly stated requirements.

**ACCESS**

Data from approved requests can be **accessed** in Vivli's secure research environment or **downloaded** with permission.

**ANALYZE**

Use robust **analytical tools** to combine and analyze multiple data sets.

**DISSEMINATE**

Completed **research results** will be assigned a DOI. Researchers may use the Vivli platform to meet their **publication** requirements.
Vivli’s Secure Environment Bridges Multiple Platforms

- STATA
- MS Office
- R
- Jupyter Notebook
- Python
- SAS
Explore Vivli

Log on to Vivli.org
To explore the ~5,800 trials available
Thank you!