Clinical trial data sharing
Clinical trial data sharing
Aims of the presentation

• Introduction

• Present and discuss the data sharing life cycle

• Summary/conclusions
Clinical trial data sharing
Introduction: Cultural change

- Strong promotion of a culture of openness and sharing of research data
- Commitment to open science research data and definition of principles and proposals for implementation
- Discussion of full transparency from different perspectives, including public health, human rights and economic perspectives
- Clinical trials more and more considered as a public good and assess to IPD seen as part of a fundamental right to health
Clinical trial data sharing
Introduction: Cultural change

Transparency/open science/FAIR

Moving Towards Transparency of Clinical Trials
Deborah A. Zarin1 and Tony Tse
National Library of Medicine, National Institutes of Health, US Department of Health and Human Services, Bethesda, MD 20894 USA

RESEARCH & INNOVATION
Open Science
THE FAIR DATA PRINCIPLES

Waste in research

REWARD
REduce research WAste and Reward Diligence
http://researchwaste.net/

Efficiency of knowledge generation

Cochrane
Individual Participant Data Meta-analysis

Reproducability crisis

1,500 scientists lift the lid on reproducibility
Survey sheds light on the ‘crisis’ rocking research.
Monya Baker
25 May 2016 | Corrected: 28 July 2016
Clinical trial data sharing
Introduction: Stakeholder engagement
Clinical trial data sharing

Introduction: Potential benefits, risks and issues

• Potential benefits
  (reduce unnecessary duplication and exposure of participants, create new hypotheses, better evidence)

• Potential risks
  (participant privacy, unfair commercial use, invalid secondary analysis, no credit for trialists)

• Major issues to be tackled
  (e.g. fairness, trust & security)
Clinical trial data sharing
Use of shared data*

Background for research, usually not leading to research outputs

• Education
• Researcher training
• Data understanding

Studies that should lead to new research outputs

• Validation/reproducability of results
• IPD meta-analyses
• Further additional analyses (prognostic models, decision-support, subgroup analyses, etc.)

*modified according to Thelwall & Kousha, AJIM, 2017; 69:36
Clinical trial data sharing
Main processes in sharing IPD of clinical trials*

1. Preparation for data sharing, in general
2. Plan for data sharing, in the context of a specific trial
3. Preparation of data for sharing, after data collected
4. Transferring data objects to an external repository
5. Repository data and access management
6. Access to individual participant data and associated data objects
7. Discovering the data objects available
8. Publishing results of re-use
9. Monitoring data sharing

*Ohmann et al., F1000Research 2018, 7:138
Clinical trial data sharing
Main processes in sharing IPD from clinical trials*

*Ohmann et al., F1000Research 2018, 7:138
Clinical trial data sharing

1. Preparation for data sharing

1.1 Learn about data sharing

1.2 Clarify own’s institution’s requirements for data sharing

1.3 Develop SOPs/related quality documents to support data sharing
Clinical trial data sharing

1.1 Learn about DS: recommendations

- **MRC, UKCRC, CRU, Wellcome**: Good practice principles for sharing IPD from publicly funded clinical trials (*Tudor Smith et al.*, April 2015)
- **MRCT, Wellcome, Arnold**: Launch of data sharing working groups 2015 *(based on workshop, Harvard, 30-31, March 2015)*
- **Wellcome**: Assessing the research potential of access to clinical trial data. A report to the Wellcome Trust (*Varnai et al.*, *Technopolis Group*, 2014)
- **Institute of Medicine**: Sharing clinical trial data: maximizing benefits, minimizing risk (*IOM*, 2015)
- **Nordic Trial Alliance Group**: Transparency and registration in Clinical Research in the Nordic Countries (2015)
Clinical trial data sharing

1.1 Learn about DS: CORBEL/ECRIN approach

BMJ Open

Sharing and reuse of individual participant data from clinical trials: principles and recommendations

Christian Ohmann,1 Rita Banzi,2 Steve Canham,3 Serena Battaglia,4 Mihaela Matei,4 Christopher Ariyo,5 Lauren Becnel,9 Barbara Bierer,7 Sarion Bowers,5 Luca Clivo,7 Monica Dias,8 Christiane Druml,13 Hélène Faure,11 Martin Fenner,12 Jose Galvez,13 Davina Ghersi,10 Christian Gluud,16 Trish Groves,16 Paul Houston,6 Ghassan Karam,17 Dipak Kaira,18 Rachel L Knowles,19 Karmela Krehja-Jeric,20 Christine Kubiak,4 Wolfgang Kuchinke,21 Rebecca Kush,22,23 Ari Lukkarinen,5 Pedro Silverio Marques,24 Andrew Newbigging,25,26 Jennifer O’Callaghan,27 Philippe Ravaud,28 Irene Schlünder,29 Daniel Shanahan,11,30 Helmut Sitter,31 Dylan Spalding,32 Catrin Tudur-Smith,33 Peter van Reusel,6 Evert-Ben van Veen,34,35 Gerben Rienk Visser,36 Julia Wilson,9 Jacques Demotes-Mainard4


Pre-publication history and additional material for this paper are available online. To view these files, please visit

ABSTRACT

Objectives. We examined major issues associated with sharing of individual clinical trial data and developed a consensus document on providing access to individual participant data from clinical trials, using a broad interdisciplinary approach.

Design and methods. This was a consensus-building process among the members of a multistakeholder task force, involving a wide range of experts (researchers, patient representatives, methodologists, information technology experts, and representatives from funders, sponsors, and regulators).

Strengths and limitations of this study

- An effective and formal consensus-building process among a large group of very experienced researchers and others involved in clinical trials.
- A unique perspective: Europe-wide, non-commercial, with a focus on the particular needs of researchers.
- A large number of practical recommendations set against an overarching framework of principles.
- The recommendations now need to be implemented and tested in practice, and feasibility and usability

*doi:10.1136/bmjopen-2017-018647
Clinical trial data sharing
Consensus process

Ohmann C, et al., BMJ Open 2017;7:e018647
Clinical trial data sharing
1.3 SOP for DS within a non-commercial CTU*

A) During study planning and setup

- Investigating existing data controller policies
- Establishing the data sharing decision making group
- Identifying the data to be made available for sharing, and when it will be available
- Identifying the nature of the data sharing possible, given the available consent
- Identifying the study documents to be made available for sharing, and when they will be available
- Identifying the likely long term storage location.
- Identifying the likely work / resources required to make datasets sharable
- Identifying the likely access arrangements
- The use of data standards
- Study documents should include appropriate reference to the data sharing plans
- The data sharing plan should be approved

*Canham, proposal for ECRIN
B) At study end and beyond

- Review of planned decisions
- Preparation for data sharing
- Generation of discovery metadata
- Data transfer to a data repository
Clinical trial data sharing

1.3 SOPs/related quality documents (example)

University of Warwick Sponsored Studies
Standard Operating Procedure 15
Information Handling, part 3: Data Transfer

STANDARD OPERATING PROCEDURE 15 part 3

Information Handling: Data Transfer

Version: 1.1

Issue Date: 19 February 2018

Effective Date: 5 March 2018

https://warwick.ac.uk/fac/sci/med/research/ctu/conducting/during/damnagement/sop_15_part_3__v1.1.pdf
Clinical trial data sharing

2. Plan for data sharing

2.1 Decide strategy for data sharing

2.2 Document strategy for this trial in trial documents

2.3 Incorporate information on data sharing plan into participant documents of clinical trials

2.4 Check and align data sharing plans of collaborators who are also generating data

2.5 Ensure that data and metadata standards have been used as far as possible in the database design
Clinical trial data sharing
2.1 Decide strategy for DS: funder requirements

<table>
<thead>
<tr>
<th>Noncommercial Funder</th>
<th>Trial Transparency Policies</th>
<th>Summary Results Sharing</th>
<th>IPD Sharing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Trial Registration</td>
<td></td>
<td>Has Policy</td>
</tr>
<tr>
<td>NIH United States</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>EC Europe</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>MRC United Kingdom</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Inserm France</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>US DoD</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Wellcome Trust</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>CIHR Canada</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>NHMRC Australia</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>DFG Germany</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>NSFC China</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>CNRS France</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>NIHR United Kingdom</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>JSPS Japan</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>BMBF Germany</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Gaisers Foundation</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>United States</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>MoH Italy</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>ISCII Spain</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>MoH China</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Total, No. (%)</td>
<td>14 (78)</td>
<td>10 (56)</td>
<td>9 (50)</td>
</tr>
</tbody>
</table>

*de Vito et al., JAMA, 2018; 319: 1723
2.1 Decide strategy for DS: publisher requirements

**International Committee of Medical Journal Editors (ICMJE), January, 2016:**

As a condition of consideration for publication of a clinical trial report in our member journals, the ICMJE proposes to require authors to share with others the deidentified individual-patient data (IPD) underlying the results presented in the article no later than 6 months after publication.

The ICMJE also proposes to require that authors include a plan for data sharing as a component of clinical trial registration.

**ICMJE, June 2017:**

Therefore, ICMJE will require the following as conditions of consideration for publication of a clinical trial report in our member journals:

1. As of 1 July 2018 manuscripts submitted to ICMJE journals that report the results of clinical trials must contain a data sharing statement as described below.

2. Clinical trials that begin enrolling participants on or after 1 January 2019 must include a data sharing plan in the trial's registration.

---

**Offline: Data sharing—why editors may have got it wrong**

*Richard Horton*

Published: 17 September 2016
Clinical trial data sharing

2.2 Document strategy for DS

2.2.1 Incorporate data sharing details within the data management plan

2.2.2 Incorporate data sharing summary in section of the study protocol

2.2.3 Incorporate data sharing summary within trial registration data
Clinical trial data sharing
2.2.1 Data management plan (example)

A Model Data Management Plan Standard Operating Procedure: Results From the DIA Clinical Data Management Community, Committee on Clinical Data Management Plan

Article in Therapeutic Innovation and Regulatory Science · April 2015
DOI: 10.1177/2168479015579529
5. Data sharing and access
Identify any data repository (-ies) that are, or will be, entrusted with storing, curating and/or sharing data from your study, where they exist for particular disciplinary domains or data types. Information on repositories is available here.

5.1 Suitability for sharing
Is the data you propose to collect (or existing data you propose to use) in the study suitable for sharing? If yes, briefly state why it is suitable. If No, indicate why the data will not be suitable for sharing and then go to Section 6.

5.2 Discovery by potential users of the research data
Indicate how potential new users (outside of your organisation) can find out about your data and identify whether it could be suitable for their research purposes, e.g. through summary information (metadata) being readily available on the study website, in the MRC gateway for population and patient research data, or in other databases or catalogues. How widely accessible is this depository?

Indicate whether your policy or approach to data sharing is (or will be) published on your study website (or by other means).

5.3 Governance of access
Identify who makes or will make the decision on whether to supply research data to a potential new user.

For population health and patient-based research, indicate how independent oversight of data access and sharing (please see page 10 of PDF file generated by selecting the above or adjacent link) works (or will work) in compliance with MRC policy.

Indicate whether the research data will be deposited in and available from an identified community database, repository, archive or other infrastructure established to curate and share data.

*MRC: Template for a Data Management Plan, v01-1, 10 March 2017
Clinical trial data sharing

2.2.1 DMP, section data sharing (template*)

5.4 **The study team’s exclusive use of the data**

MRC’s requirement is for timely data sharing, with the understanding that a limited, defined period of exclusive use of data for primary research is reasonable according to the nature and value of the data, and that this restriction on sharing should be based on simple, clear principles. What are the timescale/dependencies for when data will be accessible to others outside of your team? Summarize the principles of your current/intended policy.

5.5 **Restrictions or delays to sharing, with planned actions to limit such restrictions**

Restriction to data sharing may be due to participant confidentiality, consent agreements or IPR. Strategies to limit restrictions may include data being anonymised or aggregated; gaining participant consent for data sharing; gaining copyright permissions. For prospective studies, consent procedures should include provision for data sharing to maximise the value of the data for wider research use, while providing adequate safeguards for participants. As part of the consent process, proposed procedures for data sharing should be set out clearly and current and potential future risks associated with this explained to research participants.

5.6 **Regulation of responsibilities of users**

Indicate whether external users are (will be) bound by data sharing agreements, setting out their main responsibilities (please see page 13 section 7, titled Data-sharing agreements of the PDF file generated by selecting either of two links above).

*MRC: Template for a Data Management Plan, v01-1, 10 March 2017*
IPD Sharing Statement: Plan to Share IPD (ClinicalTrials.gov*)

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.

If yes,

- IPD Sharing Plan Description
- IPD Sharing Supporting Information Type
- IPD Sharing Time Frame
- IPD Sharing Access Criteria
- IPD Sharing URL

https://prsinfo.clinicaltrials.gov/definitions.html
Clinical trial data sharing
2.3 Information on DS in participant documents*

An appropriate consent process for secondary use of data should ensure the following:

The reasons for asking about data sharing, and the general benefits of data sharing in clinical research, are made clear to the trial participant. This information is likely to be part of the patient information sheets.

The nature of data preparation, storage and access is explained to the trial participant, so far as they are known at the time the patient documents are produced. It will also be important to describe, in broad terms, how and where the data will be stored, and how confidentiality will be maintained (e.g., by de-identification measures).

The information provided should be clear and concise, and couched in vocabulary understood by the trial participants (or if applicable their legal representatives).

The explicit consent for data sharing should be reflected in the layout of the consent forms. A request for consent to secondary use of data must be clearly distinguishable from any other matters in the informed consent document.

Although data participants should have the right to withdraw their consent for data sharing, the practical difficulties in implementing this should be made clear.

*Ohmann C, et al., BMJ Open 2017;7:e018647
## Clinical trial data sharing

### 2 Checklist for DS – Study planning stage*

<table>
<thead>
<tr>
<th>#</th>
<th>Decision</th>
<th>Done</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Investigation of any existing data sharing policies / procedures from the Data Controller. Please summarise details of any policies / procedures found</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Data Sharing decision group established. Please list members of the group (name and role)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Decision about datasets to be shared, and estimated dates to be made available. Please provide details below</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Decision taken on the need for anonymisation or additional pseudonymisation. Nature of decision</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Steps required for any anonymisation or additional pseudonymisation</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Identification of study documents to be made available for sharing. Details of the documents and estimated dates to be made available</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Decision about long term storage location (e.g. a data repository). Details of the selection</td>
<td></td>
</tr>
</tbody>
</table>

*Canham, proposal for ECRIN*
Clinical trial data sharing
2 Checklist for DS – Study planning stage*

<table>
<thead>
<tr>
<th>#</th>
<th>Decision</th>
<th>Done</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>Decision about long term storage location (e.g. a data repository)</td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Details of the selection</em></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Estimate of costs of data preparation and storage carried out</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Costs included in bids for funding</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Access arrangements (e.g. managed access versus public) decided</td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Details of access arrangements selected (for datasets and documents)</em></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Decision taken on use of data standards in study design</td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Details of standards to be used</em></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Summary description of data sharing in study protocol</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Description of data sharing strategy in DMP</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Text prepared for Trial Registry entries on data sharing</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Text on data sharing included in Patient Information Sheets and Informed Consent Form.</td>
<td></td>
</tr>
</tbody>
</table>

*Canham, proposal for ECRIN*
Clinical trial data sharing
3. Preparation of data sharing after data collected

3.1 Decide upon strategy for data sharing
3.2 Carry out strategy for data preparation
3.3 Document data preparation process
3.1 Decide strategy for data preparation for DS

3.1.1 Decide if pseudonymisation or anonymisation required (take care of legal requirements)

3.1.2 Assess the risk of re-identification with existing datasets
Decide on de-identification required
Clinical trial data sharing

3.2.1 Practical guidance on anonymizing

<table>
<thead>
<tr>
<th>Direct identifiers</th>
<th>Indirect identifiers</th>
</tr>
</thead>
<tbody>
<tr>
<td>01. Name</td>
<td>A. Place of treatment or health professional responsible for care</td>
</tr>
<tr>
<td>02. Initials</td>
<td>B. Sex</td>
</tr>
<tr>
<td>03. Address, including full or partial postal code</td>
<td>C. Rare disease or treatment</td>
</tr>
<tr>
<td>04. Telephone or fax numbers or contact information</td>
<td>D. Sensitive data, such as illicit drug use or ‘risky behaviour’</td>
</tr>
<tr>
<td>05. Electronic mail addresses</td>
<td>E. Place of birth</td>
</tr>
<tr>
<td>06. Unique identifying numbers</td>
<td>F. Socioeconomic data, such as occupation or place of work, income, or education</td>
</tr>
<tr>
<td>07. Vehicle identifiers</td>
<td>G. Household and family composition</td>
</tr>
<tr>
<td>08. Medical device identifiers</td>
<td>H. Anthropometry measures</td>
</tr>
<tr>
<td>09. Web or internet protocol addresses</td>
<td>I. Multiple pregnancies</td>
</tr>
<tr>
<td>10. Biometric data</td>
<td>J. Ethnicity</td>
</tr>
<tr>
<td>11. Facial photograph or comparable image</td>
<td>K. Small denominators – population size of &lt; 100</td>
</tr>
<tr>
<td>12. Audiotapes</td>
<td>L. Very small numerators – event counts of &lt; 3</td>
</tr>
<tr>
<td>13. Names of relatives</td>
<td>M. Year of birth or age</td>
</tr>
<tr>
<td>14. Dates related to an individual (including date of birth)</td>
<td>N. Verbatim responses or transcripts</td>
</tr>
<tr>
<td>Superfluous</td>
<td></td>
</tr>
<tr>
<td>02. Superfluous information (audit trail data, administration data)</td>
<td></td>
</tr>
</tbody>
</table>

*Keerie et al., Trials. 2018; 19: 25*
Clinical trial data sharing

3.2.1 De-identification tools (examples*)

- The PARAT tool from Privacy Analytics Inc. implements comprehensive risk management for three types of identity disclosure risk.

- mu-Argus, developed by the Netherlands national statistical agency.

- The Cornell Anonymization Toolkit (CAT) implements a k-anonymity algorithm. It is an open source tool.

- The University of Texas at Dallas Anonymization Toolbox, which contains open source Java implementations of some k-anonymity and attribute disclosure control algorithms, with documentation.

- The sdMicro package in R provides some basic de-identification functions.

*Electronic Health Laboratory
http://www.ehealthinformation.ca/faq/de-identification-software-tools/
<table>
<thead>
<tr>
<th>#</th>
<th>Decision</th>
<th>Done</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Review of Data sharing strategy carried out</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Results of the review, with details of any changes</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>De-identification (and if necessary anonymisation) applied</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Impact of de-identification on the analyses carried out.</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Data Privacy Impact Assessment (DPIA) carried out on the prepared dataset.</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Preparation of descriptive metadata for the de-identified dataset.</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Documentation of steps 3, 4 and 5 (as part of the metadata for the dataset)</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Discovery metadata generated for the dataset(s)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Please provide further details if multiple timepoints involved (which datasets / documents when)</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>(if applicable) Data transferred to a data repository</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Please provide further details if multiple timepoints involved (which datasets / documents when)</td>
<td></td>
</tr>
</tbody>
</table>

*Canham, proposal for ECRIN*
Clinical trial data sharing

4. Transferring data objects to external repository

4.1 Select repository (within institutional constraints)

4.2 Transfer the data sets under a formal data transfer agreement

4.3 Monitor repository and status of datasets transferred to the repository
## Clinical trial data sharing

### 4.1 Select repository

<table>
<thead>
<tr>
<th>types of repositories</th>
<th>examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>general scientific repositories</td>
<td>Zenodo, Figshare, Dryad, B2Share, Drum, Easy, OSF, etc.</td>
</tr>
<tr>
<td>general clinical trial repositories</td>
<td>Vivli, UMIN</td>
</tr>
<tr>
<td>disease-, funder-, stakeholder-, institutional specific repositories</td>
<td>CancerData.Org, Project Datasphere, EBCTCG, Freebird, ITN Trialshare, Melanoma MMP, NDACAN, NDCT NIMH, NIDDK, BioLINCC, ProAct, TBI-Impact, WWARN, etc.</td>
</tr>
</tbody>
</table>
Clinical trial data sharing

4.1 Select repository: Quality indicators*

<table>
<thead>
<tr>
<th>Guidelines for upload and storage</th>
<th>De-identification</th>
<th>Data quality control</th>
<th>Contract for upload and storage</th>
<th>Application of metadata</th>
<th>Application of identifiers</th>
<th>Flexibility of Access</th>
<th>Long term preservation</th>
</tr>
</thead>
</table>

*Banzi et al., Evaluation of repositories for sharing individual participant data from clinical studies (to be submitted to Trials, 2018)*
Clinical trial data sharing

4.2 Data transfer agreement (example*)

Template:
Data Transfer Agreement (DTA) for Personal Data

Clinical trial data sharing
5. Repository data and access management

Tasks of repository managers

5.1 Maintain highly granular access control to IPD
5.2 Maintain mechanisms to set up and apply authentication and authorisation
5.3 Provide a protected temporary analysis environment
5.4 Supply discovery data for IPD and data objects on a regular basis to metadata repositories
5.5 Provide an expert advisory panel
5.6 Provide data request forms
5.7 Provide data use agreement templates
5.8 Provide usage reports for data repositories
Clinical trial data sharing

6. Access to IPD and associated data objects

6.1 Manage direct responses to sponsors/PIs

6.2 Manage access to data in a repository
Clinical trial data sharing
6.1.5 Data use agreement

• Partners and bodies involved
• Definitions
• The purpose of the request and possible restrictions
• Agreement to acknowledge and give credit to the original data generators
• Public dissemination of the results of the re-analyses
• Consent issues
• Terms and conditions of control over the data within the requesting organisation
• Terms and termination of the agreement
Clinical trial data sharing

6.2 Data request form (example*)

DATA RELEASE REQUEST FORM FOR MRC CTU AT UCL STUDIES

*http://www.ctu.mrc.ac.uk/research/documents/data_sharing_application_form
Clinical trial data sharing

7. Discovering the data

7.1 Agree on a common metadata standard
7.2 Agree on ID generation scheme for data objects
7.3 Agree on ID generation scheme for clinical studies
7.4 Collect metadata repository under a single portal
7.5 Search for the data objects concerned with a trial or clinical study
## Clinical trial data sharing

### 7.1 Metadata standard*

<table>
<thead>
<tr>
<th>Mandatory</th>
<th>Recommended</th>
<th>Optional</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>The Source Study</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A.1 Source Study Title</td>
<td>A.2 Study Identifiers*</td>
<td>A4. Other Study Titles*</td>
</tr>
<tr>
<td></td>
<td>A.3 Study Topics*</td>
<td></td>
</tr>
<tr>
<td><strong>Data Object Identifiers</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B.1 DOI (1)</td>
<td>B.3 Version</td>
<td>B.4 Object Other Identifiers*</td>
</tr>
<tr>
<td>B.2 Object Title</td>
<td></td>
<td>B.5 Object Additional Titles*</td>
</tr>
<tr>
<td><strong>Data Object Provenance</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C.1 Creators*</td>
<td></td>
<td>C.2 Contributors*</td>
</tr>
<tr>
<td>D.1 Creation Year</td>
<td></td>
<td>D.2 Dates*</td>
</tr>
<tr>
<td><strong>Data Object Attributes and Descriptors</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E.1 Resource Class</td>
<td>E.6 Description*</td>
<td>E.9 Topics (of data object) e*</td>
</tr>
<tr>
<td>E.2 Resource Type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E.3 Record key type (3)</td>
<td>E.7 Language</td>
<td></td>
</tr>
<tr>
<td>E.4 Identifier type (3)</td>
<td>E.8 Related Resource Identifiers*</td>
<td></td>
</tr>
<tr>
<td>E.5 Associated consent (3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Location and Access Details</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F.1 Publisher / Provider</td>
<td></td>
<td>F.6 Rights*</td>
</tr>
<tr>
<td>F.2 Access Type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F.3 Access Details (2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F.4 Access Contact (2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F.5 Resources*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


Canham & Ohmann, Trials, 2016; 17:557
Clinical trial data sharing

7.4 Metadata repository*

A **metadata repository**, implemented as single portal system, should allow

- Identification of clinical research studies *(e.g. via clinical trial registries)*

- Display and selection of data objects related to a specific clinical study

- Display of metadata of a data object, including nature of the object *(e.g. name, type, provenance)* and the mode of access *(public, restricted, private)*.

- Link to the data object *(if possible)*

---

*developed by ECRIN/OneData in the H2020-funded project eXtreme DataCloud (XDC) (11/17-1/20) [http://www.extreme-datacloud.eu/](http://www.extreme-datacloud.eu/)
Clinical trial data sharing
7.4 Metadata repository

Unique identifier

Clinical trial

Patient-level data in a repository

Data in regulatory databases (e.g. EMA)

Summary results in clinical trial registry

Statistical analysis plan on trial webpage

Trial protocol published in journal

Data objects

PID

Metadata (intrinsic)

'provenance' (user defined)

Data (elements)

Fair principles
Clinical trial data sharing
8. Publishing results of re-use

8.1 Carry out secondary use and publish results

8.1.1 Publish re-analysis preferably open

8.1.2 Ensure proper citation of data and credit to data generators

8.1.3 Publish summary results and relevant datasets – usually in source repository

8.1.4 Apply metadata to new data objects, ensure harvesting into metadata system
Clinical trial data sharing

9. Monitor data sharing

9.1 Gather and disseminate data on data requests

9.2 Gather and disseminate data on reasons for request refusal

9.3 Gather and disseminate data on data accesses, downloads, etc.

9.4 Attempt to monitor products of secondary use (papers, datasets)
Clinical trial data sharing

9. Monitor data sharing: CSDR metrics*

<table>
<thead>
<tr>
<th>Requirements check</th>
<th>Number of Research Proposals submitted up to 31 August 2018</th>
</tr>
</thead>
<tbody>
<tr>
<td>In process</td>
<td>30</td>
</tr>
<tr>
<td>Withdrawn by the requestor</td>
<td>42</td>
</tr>
<tr>
<td>No response received</td>
<td>3</td>
</tr>
<tr>
<td>Did not meet require</td>
<td>2</td>
</tr>
<tr>
<td>Potentially agreed (signed) View details of these Research Proposals</td>
<td>217</td>
</tr>
<tr>
<td>Met requirement</td>
<td>12</td>
</tr>
<tr>
<td>Data preparation</td>
<td></td>
</tr>
<tr>
<td>In process</td>
<td>12</td>
</tr>
<tr>
<td>Withdrawn by the requestor</td>
<td>1</td>
</tr>
<tr>
<td>No response received</td>
<td>2</td>
</tr>
<tr>
<td>Complete (data available)</td>
<td>202</td>
</tr>
<tr>
<td>IRP review</td>
<td></td>
</tr>
<tr>
<td>No response received</td>
<td>113</td>
</tr>
<tr>
<td>Rejected</td>
<td></td>
</tr>
<tr>
<td>Data Access</td>
<td></td>
</tr>
<tr>
<td>Access provided</td>
<td>113</td>
</tr>
<tr>
<td>Withdrawn by the requestor</td>
<td>5</td>
</tr>
<tr>
<td>Access closed</td>
<td>84</td>
</tr>
<tr>
<td>Publication</td>
<td></td>
</tr>
<tr>
<td>Publication anticiped</td>
<td>31</td>
</tr>
<tr>
<td>Published View details of these Research Proposals</td>
<td>29</td>
</tr>
<tr>
<td>Publication not received after 18 months of Access Closed</td>
<td>26</td>
</tr>
</tbody>
</table>

*https://www.clinicalstudydatarequest.com/Metrics.aspx
Clinical trial data sharing

9.4 Monitor products of secondary use

Figure 1: The data sharing pipeline

Status of sharing of Individual Participant Data from clinical trials, use of shared data and impact of research outputs of sharing data: a scoping review protocol

Contributors: NAUDET Florian, David Moher, Ohmann Christian, Motschall Edith
Date created: 2018-09-12 04:54 PM | Last Updated: 2018-09-13 03:58 PM
Clinical trial data sharing
Summary/conclusions

• Data sharing should become the norm in clinical research

• Plans for data sharing should be described prospectively, and be part of study development from the earliest stages

• Data sharing should be formalised through SOPs, data sharing plans and agreements (data transfer, data use)

• Data made available for sharing should be prepared for that purpose, with de-identification of data sets and adequate metadata

• Shared data should, as far as possible, be structured, described and formatted using widely recognised data and metadata standards

• Data and trial documents made available for sharing should be transferred to a suitable data repository
Clinical trial data sharing
Sharing within the Trial Reporting System*

*Zarin & Tee, Plos Med, 2016; 13; e1001946
P1: The provision of individual participant data should be promoted, incentivised and resourced so that it becomes the norm in clinical research. Plans for data sharing should be described prospectively, and be part of study development from the earliest stages.

P2: Individual participant data sharing should be based on explicit broad consent by trial participants (or if applicable by their legal representatives) to the sharing and reuse of their data for scientific purposes.

P3: Individual participant data made available for sharing should be prepared for that purpose, with de-identification of data sets to minimise the risk of re-identification. The de-identification steps that are applied should be recorded.

P4: To promote interoperability and retain meaning within interpretation and analysis, shared data should, as far as possible, be structured, described and formatted using widely recognised data and metadata standards.

P5: Access to individual participant data and trial documents should be as open as possible and as closed as necessary, to protect participant privacy and reduce the risk of data misuse.

Ohmann C, et al., BMJ Open 2017;7:e018647
P6: In the context of managed access, any citizen or group that has both a reasonable scientific question and the expertise to answer that question should be able to request access to individual participant data and trial documents.

P7: The processing of data access requests should be explicit, reproducible and transparent, but, so far as possible, should minimise the additional bureaucratic burden on all concerned.

P8: Besides the individual participant data sets, other clinical trial data objects should be made available for sharing (eg, protocols, clinical study reports, statistical analysis plans, blank consent forms) to allow a full understanding of any data set.

P9: Data and trial documents made available for sharing should be transferred to a suitable data repository to help ensure that the data objects are properly prepared, are available in the longer term, are stored securely and are subject to rigorous governance.

P10: Any data set or document made available for sharing should be associated with concise, publicly available and consistently structured discovery metadata, describing not just the data object itself but also how it can be accessed. This is to maximise its discoverability by both humans and machines.