Proceedings
Promoting Clinical Trial Data Transparency Conference

30-31 March, 2015

Harvard Faculty Club • Cambridge, MA

Table of Contents

Executive Summary – Committed Leaders Gather to Drive Clinical Trial Data Transparency Solutions

Introduction

On March 30-31st, 2015, over seventy committed international stakeholders joined together at the Harvard Faculty Club in Cambridge Massachusetts to deliberate how to promote and coordinate clinical trials data transparency. Over the two day conference, renowned experts from academia, industry and others shared their organizational experiences in data sharing, distilled best practices, described lessons learned from case studies, reviewed recommendations from the Institute of Medicine report: *Strategies for Responsible Sharing of Clinical Trial Data*, and considered how best to put those recommendations into practice.

The explicit goal of the conference was to develop an approach whereby:

- Expectations and practices of registration and results reporting of all clinical trials would be regularized among industry and academia;
- Greater access to participants-level clinical trial data could be facilitated using a common portal;
- Researchers would be able to access and combine data across various platforms and sponsors, to multiply opportunities for data analysis; and
- Research participant privacy can be safeguarded

The assembled participants separated into four parallel break-out sessions to discuss lessons learned in their ongoing data sharing initiatives and to consider options for promoting clinical trials data sharing and bringing consistency to data sharing efforts. The data sharing models that emerged from these discussions shared similar characteristics including:

- Organizational structure - A centralized, international, not-for-profit organization responsible for a coordinated data sharing initiative;
- A centralized and single portal - A central user interface with a robust search engine functionality, including information on trials around the world, with the capability to grow and add data from new sponsors;
- Governance – Creation of an empowered central multi-stakeholder body with authority and accountability to enable the long term vision wherein a not-for-profit entity may promote and oversee the data sharing enterprise end-to-end;
- Data requirements – Sufficient data pedigree including, data definition and metadata to enable the appropriate integration of datasets across studies and sponsors for analysis;
- Shared or common services – Efficient shared or common services across data generators / sponsors (policy setting, data de-identification, and when appropriate, criteria for independent review panel decisions); and

- Flexibility- A data platform that accommodates differing expectations and research needs, including the ability to download data if freely available and the ability to host data for those data generators that do not wish to do so themselves.

Consensus in four primary action areas was identified, with a goal of working toward implementation over the next 18 months. These areas encompass:

- The creation of a working group tasked with development of specific principles, operating guidelines, and characteristics of the suggested not-for-profit organization and its governance. The goal of the not-for-profit organization would be to oversee, create, implement and direct a sustainable data-sharing platform.
- Initially this working group will be led by MRCT Center at Harvard and the Wellcome Trust, with active participation from the John and Laura Arnold Foundation.
- The results and recommendations of the working group would be reviewed and commented upon by a variety of public and private stakeholders, not limited to the participants in the March 2015 Harvard meeting. Discussions and follow-up will be expanded to include the perspectives of trial registries, start-up companies, university leaders, public sponsors, disease foundations and medical journals.
- Policy initiatives and a communication plan will be a created to promote of, and incentives to promote clinical trials data sharing. This will potentially include journal publishing requirements, academic and faculty acknowledgements of contributions relevant for promotions, and others, to promote practices of open data in the academic setting.

The participants of the data-sharing workshop demonstrated their personal commitments to data transparency and agreed that progress toward a sustainable solution may be challenging but is an important transformative goal.

This conference, which followed from an earlier multi-stakeholder conference held at Harvard in May 2013, was supported and hosted by The MRCT Center at Harvard, the Laura and John Arnold Foundation, and the Wellcome Trust.

Pre-Conference Survey

A pre-conference survey of 21 conference participants demonstrated that 100% of respondents supported the value of data sharing and the aggregation of data, but specifics of how data should be shared varied. On the one extreme, the most open vantage sought to allow all data to be available while a more conservative position supported a 'learned intermediary model' of making data available upon request after review. All respondents agreed that a centralized system that allowed interoperability and data integration for analysis was the preferred option.

Session I: Lessons Learned and Current Needs – Session #1A: Experiences and best practices from current initiatives

Five current initiatives were presented (for speaker biographies, see Appendix 3):

- The Clinical Study Data Request (CSDR) platform, a multi-sponsor request site in which data from 11 companies are anonymized and used in a secure environment
- The Duke Clinical Research Institute – Bristol Myers Squibb Strategic Initiative (SOAR), which supports open access to clinical trials data for researchers
- The Merck data sharing process, which prepares data on demand after internal review of data requests, or, if access declined, external adjudication
- Pfizer's Integrated System for Pfizer Investigator Initiated Research (INSPIIRE [IIR]) Portal, which subjects data requests to an internal review, followed by an external review if the decision is declined
- The YODA project, in which Yale, an academic institution, is partnering with Johnson & Johnson and Medtronic.

The panel discussion addressed the challenges of: the complexity of device trials in data sharing, the evaluation of validity of a proposed data analysis, assessing legal compliance, the review process of data requests, the ability to merge data across different platforms, the challenges of releasing data to a student class for reanalysis, and the types of requests to date for data sharing. These issues need to be considered in establishing a data sharing platform.

Session #1B: Forward looking strategies and initiatives

Five initiatives were shared:

- Clinical Data Interchange Standards Consortium or CDISC, that has developed foundational standards for clinical and non-clinical research from protocol, data collection, tabulation & analyses to submission/publication & reporting
- The National Institutes of Health (NIH) perspective that reviewed the poor results reporting rates of clinical trials and the gap in time between end point of study and publication, if published at all
- Patient-Centered Outcomes Research Institute (PCORI) that aims to make study results and data more widely available and still has to address some open questions
- A Clinical Trial Data Exchange idea that envisions an exchange in which data sharer and data user come together for the purpose of making clinical trial data available for secondary analysis
- The Wellcome Trust that recently commissioned a research report that outlined several barriers to data access as well as made recommendations

The panel discussion addressed the challenges of: differences and considerations in making all data available vs only requested data, proposing a portal that is different from the place where data would be stored, providing access to international trials through WHO platform, requiring data sharing plan as part of grant application, preventing unnecessary future trials, different cost models for data sharing, and reporting requirements. Panel members suggested creating a culture of data sharing.

Case Studies: history, lessons learned, best practices

Four case studies were presented:

- Global Alliance for Genomics & Health, which has as its mission to accelerate progress in human health by helping to establish a common framework of harmonized approaches to enable effective and responsible sharing of genomic and clinical data, and by catalyzing data sharing projects that drive and demonstrate the value of data sharing.
- European Federation of Statisticians in the Pharmaceutical Industry (EFSPI), which has a data transparency working group with five work streams to provide input into EMA/EFPIA-related data transparency activities, develop recommendations for re-analysis practices, assess the future impact to biostatistics with increased data transparency, establish the minimum requirements for sharing data, and ensure patient confidentiality.
- Laboratory of Neuro Imaging (LONI) at the University of Southern California, which provides open access to neuroimaging scans. LONI has learned that design must be prospective and clear, technology and tools must keep pace with development, the size of data needs to be considered, duration and sustainability must match, the mission needs to be clear, and individuals must have the "right personality" and commitment.
- The National Heart, Lung and Blood Institutes's (NHLBI's) Biologic Specimen and Data Repository Information Coordinating Center (BioLINCC) which provides biologic specimen and data repository. BioLINCC has developed a relatively simple application process, the ability to combine data from multiple studies and access to data from closed studies, but noted limited support and limited access to ancillary data, issues with de-identification that may limit the range of applicable research questions, and newly collected data that may still be in 'protected time' during which they are not yet allowed to be released.

The panel discussion addressed concerns about re-identification, informed consent, availability of the statistical code, ability to understand the wealth of data available, challenges of what data to share, complexity of interoperability, and cost model for making data available. These issues need to be considered in establishing a data sharing platform.

Session II: The Way Forward

Dr. Bernard Lo summarized relevant recommendations from the recent IOM Report entitled Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk. Stakeholders in clinical trials should foster a culture in which data sharing is the expected norm. Sponsors and investigators should share the various types of clinical trial data no later than the times specified in the report. Data generators should employ appropriate privacy protections, learn from experience by collecting outcomes of data sharing, and have an independent review panel that includes members of the public that review data requests. It is also important to include other stakeholders.

Dr. Vasee Moorthy informed the conference participants that the World Health Organization (WHO) will convene a meeting on data sharing in the context of public health on September 1-2, 2015. WHO is

working with many different stakeholders to advocate that all clinical trials are registered and results publicly disclosed.

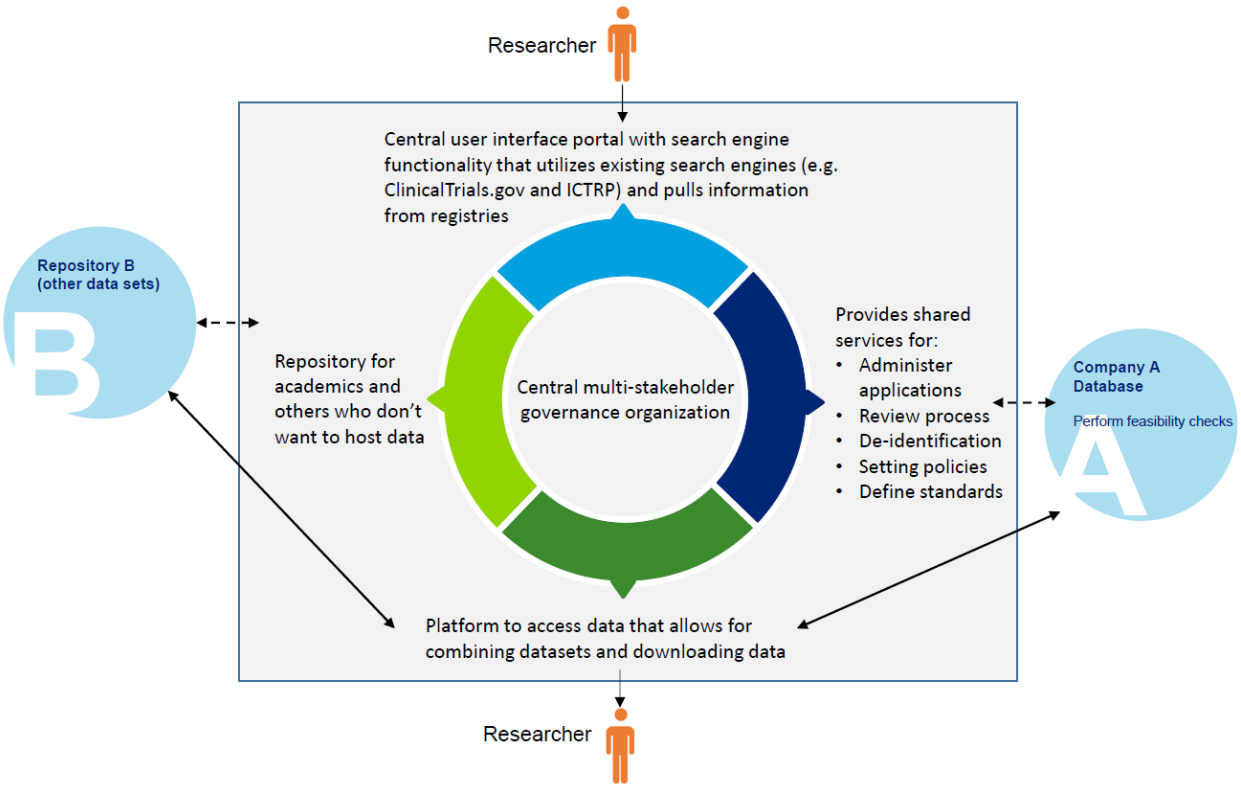
Breakout Sessions

Participants were then divided into four multi-stakeholder groups to discuss the following topics: (1) Vision of the ideal platform, (2) Barriers and incentives, (3) Characteristics of the platform, and (4) Getting data into common platform and accessing data from common platform.

Session III: Commitments and Next Steps

Reporting back and input from multi-stakeholder panels culminated in the vision of the ideal platform.

Topic 1: Vision of ideal platform



The multi-stakeholder panel emphasized the importance of an independent review panel, building on existing infrastructure (registries), having a federated model, considering open access, realistically calculating the costs, and clarifying the principles.

Topic 2: Barriers and Incentives

Breakout groups and multi-stakeholder panel emphasized:

- the importance of professional recognition, including the concerted action of academics and professional journals to give credit and visibility to data generators and data sharers

- creating a culture for data sharing partly driven by journal publishing requirements, academic institution and faculty expectations, and learned society standards
- developing case studies where data sharing has led to good science

Topic 3: Characteristics of the platform

Breakout groups and multi-stakeholder panel recommended:

- the criticality of a centralized, single portal of entry
- the importance of having a global platform and a search engine
- the importance of a self-regulating system
- the importance of data standards and data sharing protocols
- the need for investment in infrastructure, governance, and sustainability including multi-stakeholder commitment to building the IT infrastructure
- offsetting entry barriers for small biotech/device companies and academics
- multiple funding channels including fee for service, subscription-based service, mini grants, and the option for data generators to donate costs for specific projects
- a suggestion of a publically-available, data-sharing transparency ranking index

Topic 4: Getting data into and accessing data from a common platform

Breakout groups and multi-stakeholder panel suggested:

- methods of data standardization
- the default of non-downloadable data, with case-by-case exceptions
- the importance of agreeing on Data Use Agreements (DUAs) and Informed Consent Forms (ICF) for a streamlined process
- the necessity of protecting research participants and build-in safeguards to reduce the re-identification risk
- the importance of having a dedicated staff/head to lead this
- the importance of having a flexible platform that allows for different mechanisms of data access, data analysis, and of the ability to host data
- the consideration of a public-private partnerships to move this forward

Summary

After engaged discussions, Mr. Mark Barnes, co-chair of the Multi-Regional Clinical Trial Center, summarized:

We need a federated model “with many flowers blooming” to improve opportunities for research.

We were charged to continue the conversation in a structured way, involving all interested stakeholders.

The discussed governance model has the following requirements:

- Determine if a new or existing entity shall govern and take responsibility to convene
- Primary purpose has to be to improve collaborative science in the interest of human health:
    - Efficient – collaboration within initiative must be more efficient than functioning outside the initiative
    - Sustainable – might be started with any sources of funds, has to be sustainable
    - Voluntary – needs to foster a self-culture of improving the conditions under which science is done, bolstered under FDAAA, EMA, NIH, ensuring participation is done within regulations
- Needs to be broad enough in design to accommodate those in the forefront and those who lag behind
    - Must incorporate and respect existing systems, e.g., Framingham Study
    - Broad enough to accommodate national requirements, e.g. anonymization of data
    - Must allow companies to comply with their own particular regulatory regimes
- Must interact and enhance regulatory initiatives (e.g., CT.gov, WHO platform)
    - Expand the knowledge base of what exists
- Protect study subjects (participants) and re-identification
    - Improve Informed Consent process and Data Use Agreements
    - Articulate best practices

Welcome and Introduction

Introduction
- Barbara Bierer (MRCT Center, Brigham & Women's Hospital)

Dr. Barbara Bierer introduced the purpose of the Multi-Regional Clinical Trials (MRCT) Center at Harvard which is to improve the design, conduct, and oversight of multi-regional clinical trials, especially trials sited in or involving the developing world; to simplify research through the use of best practices; and to foster respect for research participants, efficacy, safety and fairness in transnational, trans-cultural human subjects research. Clinical Trial Data Sharing and Transparency is one of a number of current MRCT Center initiatives. The MRCT Center's focus on data sharing included a conference on "Issues and Case Studies in Clinical Trial Data Sharing: Lessons and Solutions" in May 2013, and developing templates for Data Use Agreements (DUA) and Informed Consent Forms (ICF).

Dr. Bierer also presented the conference objectives:

- To discuss high-level principles with the explicit goal of developing an approach whereby:
    - Expectations of registration and results reporting of all clinical trials would be intensified
    - Greater access to participant-level clinical trial data could be facilitated
    - Researchers would be able to access and combine data across various platforms and sponsors
    - Patient privacy and confidentiality would be respected
- To present and deliberate models for data access to promote interoperability including a discussion of the merits of centrally-managed vs federated systems for hosting data
- To discuss requirements and incentives to enable harmonization and broad data sharing across sponsors (academic, industry, biotech, not-for-profit sponsors)

Dr. Bierer explained that all the assembled conference participants are conversant on current efforts in data sharing and the goal of each short talk is to review from the experiences to date what worked, what did not work, and what is still needed. The conference is a cooperative effort to synthesize elements of commonality and important differences; discussion is essential and differing views are anticipated in order to develop an agenda for further action and our commitment to the future.

Issues for harmonization of data sharing and transparency efforts and Analysis of Data Use Agreements
- Mark Barnes (MRCT Center, Ropes & Gray)

Mr. Mark Barnes presented issues that would be addressed during the conference in an effort to advancing and optimizing data sharing efforts. Issues included:

- Characteristics of an ideal platform
- Review of data request
- Access to data
- Use of data
- Anonymization Standards
- Governance
- Incentives

Mr. Barnes further presented the current versions of Informed Consent Forms (ICF) and Data Use Agreements (DUA) that have been developed by MRCT Center working groups. He also introduced an analysis of DUAs that were collected in preparation for the conference and looked at the main elements of:

- Permitted uses of data
- Downloading data
- Safety concern notification
- Use of confidential information
- Re-identification
- New intellectual property
- Publication
- Public register
- Acknowledgements

The DUA analysis showed that all of them included rigorous requirements for permitted use of data (i.e., sole purpose for data use is analysis according to approved research plan) and for use of confidential information (i.e., may use confidential information only for approved analysis). DUAs from industry were most rigorous, while DUAs from foundations and nonprofit health care providers were less rigorous and were silent on more of the issues.

Pre-Conference Survey Findings

- Pete Lyons and Nick Lingler (Deloitte)

Mr. Lyons and Mr. Lingler presented a total of 21 interviews conducted over two months to identify commonalities, limitations and gaps of current data sharing approaches.

Common principles were identified for sharing of participant level clinical trial data. Interviewees promoted the sharing of data to advance science and improve public health while

- Protecting patient rights and privacy
- Ensuring responsible conduct of research and good stewardship of data
- Maintaining incentives for those who generate data to conduct new research

The survey also showed that 100% of respondents supported the value of data sharing and the aggregation of data, but specifics of what data to share varied from making all data available to making data available on request after internal review for legal and competitive risk. A central system that allows interoperability was widely considered to be the preferred option, however, many challenges need to be addressed, including:

- Use of common data standard
- Resources to anonymize data
- Data repository for academics
- User friendly system
- Ability to combine data easily
- Common criteria or common review board

Moreover, to improve transparency, sponsors plan to or are currently publishing metrics of requests and track reasons for denial. Data from respondents show that 93% of data requests have been approved. Reasons for denying a request for data include: proposal lacks clear scientific merit, data requested is not appropriate for the study proposal, DUA was not signed, out of scope for informed consent, etc. To achieve a shared vision for an interoperable system, key process points should be harmonized such as proposal review, uploading data, system framework, and end of study requirements.

Questions to the presenters clarified that the survey sample included individuals who are participating in the current conference, who are on the forefront of data sharing. The respondents did not specify who would run a centralized system, however, it would not necessarily be run by a data generator. The time for the approval of data requests varied widely in the sample.

Session I: Lessons Learned and Current Needs

Experiences and best practices from current initiatives

Clinical Study Data Request (CSDR) platform
- Frank W. Rockhold (GSK)

Dr. Rockhold presented that in May 2013, GSK launched a 'controlled access' system for researchers to request data to conduct research that can help advance medical science. In January 2014, this has expanded to a new multi-sponsor request site, in which data from 11 companies are anonymized and used in a secure environment, and the scientific rational for data requests is reviewed by an Independent Review Panel. Currently, eleven sponsors are using the request site.

GSK is committed to listing all interventional trials that were ongoing or started after the formation of GSK in 2000. Trials are listed after the primary manuscript has been accepted for publication AND the medicines has been approved / terminated.

Researchers are able to request data from multiple sponsors by using one proposal, one review, and one common access system. The Wellcome Trust is managing the review of proposals and the operation of the review panel. There is a need for alignment with different systems and with different approaches to data access and a need for common data standards.

Taking this first step of creating a platform showed that "it can be done" and encouraged access to data. It also provided practical experience in managing requests, preparing data and providing request and access systems. Future priorities include improving usability and meeting needs of non-industry sponsors. GSK's long term perspective is to help realize a broad, independent solution to allow access to data from clinical trials conducted by multiple companies and organizations.

Supporting Open Access to Clinical Trials Data for Researchers: The Duke Clinical Research Institute – Bristol Myers Squibb SOAR Initiative
- Eric D. Peterson (Duke University)

Dr. Peterson referred to calls for increased transparency and open access to clinical trials data as well as key recommendations from the Institute of Medicine (IOM) report. Currently, most pharmaceutical companies are implementing data sharing strategies, but they vary greatly in terms of data request review, requirements for data access, and who provides the data.

The Supporting Open Access for Researchers (SOAR) is focused on clinical trial transparency and provides independent scientific review of Bristol-Myers Squibb data sharing request through the Duke Clinical Research Institute (DCRI). The Independent Review Committee (IRC) includes specialists for statistics/data, clinical aspects, and bioethical/protection of human subjects. The process includes administrative assessment and scientific assessment.

Review criteria for data requests include: clearly defined and scientifically valid research question, rigorous statistical analysis plan, clear rationale for combining data sets, adequate publication plan for disseminating findings, sufficient experience of the research team, and adequate plan for protection of human subjects and data privacy.

To date, DCRI has received several proposals for IRC review of which most were approved for expedited review. The average time between receipt of the proposal and submission of formal review summary was 11 business days.

Key advantages of SOAR include that is offers an efficient model for independent review and easy access to diverse faculty. Each review committee can be customized which allows for more detailed evaluation based on subject matter expertise, and it has a built in final review to ensure the published analysis is consistent with the proposed analysis plan.

Merck & Co. Data Sharing Process

- Barbara Kress (Merck)

Ms. Kress presented that Merck has been committed to data sharing since 2008 and has received 59 data sharing requests since 2012/2013 of which three were declined because of informed consent and resource issues. Data for which they received requests date back to the late 1990s. Last year, 18 requests were received of which two were declined. One was declined because the hypothesis did not match the data requested, the other was declined because of a substantial resource issue. Merck has its own data request portal, which started in 2012 for investigator interaction. They loaded data to an SAS platform two weeks ago.

Data requests are reviewed internally at Merck. If requests are declined, they are passed to an external review board. The review committee makes a recommendation, and an internal Merck executive steering committee makes a final decision. The external board also reviews manuscripts.

Merck prepares data on demand, not proactively. They are still waiting for standards to be established and hope to learn more at this conference.

Merck has only one staff member dedicated to this, in addition to clinical teams in the biostatistics department. They find the expense of SAS a barrier and need a more cost-effective platform. Joint

ventures pose interesting challenges to data sharing. Merck staff believes that contracts should define the data sharing parameters up front. They also perceive a need for keeping metrics and continuously improving the data sharing process and experience.

Pfizer Experience and Practices in Clinical Trial Data Sharing
- Justin McCarthy (Pfizer)

Mr. McCarthy explained Pfizer's approach to sharing patient level data. Available data include data from authorized or terminated medicines two years after trial completion and from completed trials since September 2007, except Phase I studies in healthy volunteers. All data request research proposals will be assessed to ensure there is a sound scientific rationale for the research, a well-documented statistical analysis plan, and a commitment to publish resulting findings. Researchers will be able to request access to data to answer scientific questions.

Pfizer uses a SAS platform for data sharing. Researchers submit their request through the INSPIIRE Portal, the Pfizer Review Committee conducts an internal review and makes a decision. If the decision is approved, the researcher will obtain access to the data. If the decision is declined, the proposal will be submitted for adjudication by an Independent Review Panel. If the request is subsequently approved, the researcher then obtains access to the data. If the request is declined, Pfizer notifies researcher and posts the outcome.

Mr. McCarty presented several lessons learned:

- Data sharing is here to stay
- The risks are real but can be managed
- Purpose driven data sharing is effective and balances risk
- Policies need to address different audiences
- Legislation is not an appropriate vehicle for governance

In the future, we need to look beyond clinical trials and also consider genomics/omics data and big data from EMR and health care systems. The future state will likely be: collaborative, end-to-end, purpose-driven, integrates health care, networked, structured and standardized.

The YODA Project
- Joseph S. Ross (Yale University)

Dr. Ross explained the principles of the YODA Project, which include promoting the sharing of clinical research data to advance science and improve public health and health care and to promote responsible conduct of research. Unique features include that an independent, academic organization is partnering with data holders, making all decisions transparent, while data holders have given full jurisdiction to make decisions regarding data access, removing the perception of influence.

Currently, the YODA Project is partnering with Johnson & Johnson and Medtronic, Inc. It has established policies and procedures after soliciting feedback from stakeholders, experts, and the public. More than 110 trials are available. The YODA project has received less than 20 submitted requests, all of which have been approved.

Required standards for data requests include: a clear scientific purpose, requested data will be used to create or materially enhance generalizable scientific and/or medical knowledge to inform science and public health, and proposed research can be reasonably addressed using the requested data.

The YODA Project functions as an independent third party reviewing requests for data access. It is not responsible for preparing data, housing data, or maintaining secure data sharing server—these tasks fall on partnering data holders. The YODA Project strives to be an innovative leader to set standards in the field. It has capacity to do more, but its purpose is to demonstrate the potential and value of their approach. Data are currently being shared through a secure data sharing platform maintained by SAS.

Lessons learned include:

- Creating a web platform that facilitates research: what trials are or can be made available?
- Resources are not unlimited: should there be a fee?
- Patient privacy & secure data analytic platform: how easy can it be?
- Maintaining public input, transparency
- Scope and intensity of YODA Project review and due diligence assessment by data holder
- Understanding value of one data sharing approach versus another, demonstrating value to data holders

Ross shared the perspective that the goal for data sharing is to be as least burdensome as possible, for patients, investigators (data collectors) and research funders (data holders), which requires:

- Universal informed consent language
- Data standards
- Clear data definitions
- Platforms that secure data, protect patients, but can accommodate added data and analytics

Moderated Panel #1A: Experiences and best practices from current initiatives

- Frank Rockhold, Eric Peterson, Barbara Kress, Justin McCarthy, Joseph Ross; Moderators: Mark Barnes & Barbara Bierer

The panel discussion included the following issues:

- Device trials complexity in data sharing:
  - Device trials are more complicated than biomedical drug trials
  - Standardization in device trials is different and complex
  - Organizational structures can create complexity, as well as multiple regulatory pathways and regulatory complexity
- Evaluating the validity of proposed data analysis ("analysis police") in data requests:
  - Some organizations have persons appointed to determine adherence to standards
  - Others have peer review process for review of applications and requests
  - Others appoint reviewers that are independent from the company, which reduces bias
  - Some have internal review process and sample request form posted on company website
- Legal compliance
  - Assess if proposal includes a valid study design but not to influence the design of the question or trial
  - Issues in interacting with data requester if they ask for a study not listed as available
  - Review by independent review panel
- Review process
  - Proposals for data use are submitted via website
  - Journals have not yet asked for statistical plan to be submitted with article manuscript
- Merging data across different platforms:
  - Need to work closely with different funders in order to facilitate release of data for meta-analysis
  - Need to meet data privacy requirements
  - Data needs to be uploaded or accessed for the researcher to use
  - Providing summary analysis may be more expedient but serves a different purpose
- Types of requests for data sharing
  - Broad range of requests, many analyzing a subset of the data
  - Rarely receive requests for re-analysis of original data report
  - Most requests are for original, academic (non-commercial) research
- Hypothetical: would releasing data to an entire class of students for reanalysis by every member of the class be a proposal that would be approved by an independent review panel?
  - This issue has not been sufficiently considered, but might say no if request is to "play with the data"

- Individual students could submit scientific proposals
- Alternatively, an overarching proposal could come from professor/teacher who is then the locus of authority and accountable to the sponsor
- Must review how original Informed Consent Form (ICF) is written as many did not anticipate 'student research' or 'education' as a potential future use

Forward looking strategies and initiatives

CDISC

- Rebecca D. Kush (CDISC)

Dr. Kush reminded us that patients participating in clinical research expect their data to be used wisely, efficiently, and accurately. In the words of Drs. Kathy Hudson and Francis Collins, NIH, "to honor these participants' commitment to advancing biomedical science, researchers have an ethical obligation to sharing the results of clinical trials in a swift and transparent manner." In order to do so, the implementation of global consensus-based standards is critical. A learning health system should not only support big data for data mining and analytics, but also the collection of high quality research data for data science.

Over the last few years, CDISC has developed global clinical research standards that are harmonized through the BRIDG model. Foundational standards for clinical and non-clinical research from protocol and data collection through tabulation/analyses and submission/publication have been developed. Currently, CDISC is working on therapeutic area standards to augment the foundational standards.

When standards are not used, even simple questions often cannot be answered readily because data are collected and reported in different formats. Without standards, different studies might use different coding and terminology for gender or sex, for instance. Another example was the use of the Alzheimer's cognition scoring questionnaire, which was implemented differently across organizations. There is now a CDISC Standard for Alzheimer's disease. If this is used for data collection and reporting, the results can now be compared directly to the Critical Path Institute's research database with data in the CDISC format from more than 6,000 patients. According to Enrique Aviles of C-Path, the CDISC Alzheimer's Disease standard provides a faster path to gaining insights from the data. CDISC Standards also enable research study start-up in 70-90% less time than without standards.

Common data elements (CDEs) are not necessarily standards. They must be developed through an open, consensus-based standards development process and should not be redundant with existing global standards. The CDISC standards will be required by U.S. FDA and PMDA in Japan. Another opportunity

is to use EHRs for clinical research to streamline the process. Synergistic standards (from HL7, IHE and CDISC) are available to link healthcare and research.

Going forward, sponsors need to use the standards and eSource Guidance to streamline research and more therapeutic area standards need to be developed. CDISC is also creating a Shared Health and Research Electronic Library, which will make the CDISC standards more readily accessible (electronically) and accelerate the development of therapeutic area standards enabling re-use of common components across the therapeutic areas.

The NIH Perspective
- Kathy Hudson (NIH)

Dr. Hudson presented that 12% of the NIH FY 2014 budget went to clinical trials. Sharing summary and participant level data from clinical trials builds public trust in clinical research by:

- Informing future research and research funding decisions
- Mitigating bias (e.g., non-publication of results, especially negative results)
- Preventing the duplication of unsafe trials
- Meeting ethical obligation to human subjects
- Increasing access to data about marketed products

However, the rate of reporting in ClinicalTrials.gov is low: while 17.0% of industry trials reported results within 12 months, only 8.1% of NIH trials did so. The gap in time to publication increases with time after trial completion. Clinical trials with clinical endpoints were more likely to be published than trials with surrogate end points.

While NIH supports only a small fraction of clinical trials initiated annually, the sharing of data for trials of approved products as well as for unapproved products and non-intervention clinical trials is an issue of public trust. NIH received close to 300 comments on its proposed policy on dissemination of data from NIH-funded clinical trials.

Sharing and reporting the results of clinical trials has been an increasing focus in the United States across public and private sectors, even though it adds an additional burden to investigators who already spend half of their time doing non-clinical work. In regard to clinical trials data sharing of participant level data:

- NIH is focused on 100% compliance with results reporting
- NIH is beginning to think about policies for participant-level data

- IOM Report includes strategies for responsible sharing of clinical trial data
- Evaluation of the utility of individual level data is needed
- Several programs at NIH collect participant level data

PCORI Proposed Policies for Open Science and Data Sharing

- Steven Goodman (Stanford University)

Dr. Goodman presented PCORI's efforts to make study results and data more widely available. PCORI (Patient-Centered Outcomes Research Institute) is authorized under the Affordable Care Act (ACA) through 2019. The hope is that PCORI's models will persist beyond the ACA purview. The scope of PCORI's studies goes beyond the portfolio represented in this conference and includes, for instance, clinician reminder systems.

While PCORI's open science policy has not been fully instituted yet, it aims to be harmonized with the recent IOM report in terms of fostering a culture in where data sharing is the expected norm (Recommendation 1), sharing data according to particular timelines (Recommendation 2), and access to data and governance (Recommendation 3).

As proposed elements of its policy, PCORI requires:

- All studies to prepare for possible future requests for data sharing
- All applicants to describe the methods and plans for sharing data in their application for funding
- All investigators to share the initial and final study protocols as study deliverables

The proposed procedure includes:

- Application and Study Conduct: requirement for PIs to lay out plans for sharing requests
- Study Completion: indication of requirements for applicants at the end of the study to facilitate sharing
- Data Storage: data, meta-data, analytic code and study protocols will be deposited in a stable repository
- Requests for Sharing: process by which requests for sharing will be vetted and approved
- Study Reproduction: process by which PCORI may choose to reproduce select studies

There are still Open Questions that need to be addressed:

- What % of funds be allocated to data sharing versus new research? Should they be provided on a "just in time" basis or for all studies, for all studies regardless of requests, or just for high priority studies?
- How can we distinguish activities related to sound data management and curation from those uniquely required for data sharing?
- Should PCORI commit resources to study reproduction? If so, how much, and how are studies selected?
- What platforms / models can be used for: Data and meta-data storage? Adjudicating data sharing requests?
- What criteria should be used for data sharing requests?
- Are requestors bound by PCORI publication policies?
- Who "reproduces" analyses, and what are the consequences/actions if meaningful differences are found?

A Clinical Trial Data Exchange

- Ronald L. Krall (University of Pittsburgh)

Dr. Krall presented a concept idea of how data could be shared which he developed, based on conversations with multiple stakeholders. It is envisioned as an exchange where data sharer and data user come together for the purpose of making clinical trial data available to enable secondary research.

The Exchange would offer a set of services for the data sharer that allows him to formally record data available, determine conditions for access including user and research qualifications and access methods, and deposit the data. User and Research Qualifications would be addressed through Data Use Agreements and Research Review Committees. The Exchange would operate a data repository that would offer data access in closed analytical environments and direct download. Through the Exchange, the data user would be able to register, determine data availability, satisfy conditions of access, and achieve access to the data. This would be governed by a Board of Directors, with established procedures for management and operations. This would establish an infrastructure for data sharing and secondary research with transparency, compliance, and investment so that secondary research becomes part of the research process.

In conversations with potential data sharing entities, Dr. Krall and his colleague found a small and growing number of companies having high commitment to data sharing. The remainder is unsure about the need to share or unaware. There is an interest in and support for concepts embodied in the described Exchange, and concern that the existing platform is unworkable, inappropriate and too expensive for most clinical data sharers. Even for pharma companies, the current platform is expensive and alternative options are desired.

The envisioned long-term goal is to create a vehicle that enables data sharing so that sharing of data becomes possible. In the shorter term, this would involve creating a mandate that data be shared and enhancing the methodology of the conduct of clinical trials so that data sharing is a minimal additional burden and part and parcel of conducting an investigation in humans and secondary research is more efficient and informative.

Furthermore, Dr. Krall found that adopting standards is hard and expensive and the case for investment is beyond most who conduct clinical trials. The proposed business model should be global, sustainable, scalable, with minimal additional cost to the clinical trial enterprise. If data sharing is mandated and the data from the 18,000 trials completed each year are shared, a fee-based model is viable.

Wellcome Trust Activities

- Nicola Perrin (Wellcome Trust)

Ms. Perrin gave a short introduction to The Wellcome Trust, which is a global charitable foundation dedicated to improving health.

Recently commissioned research by Technopolis showed several barriers to data access: current access to relevant datasets blocks research (66%), incomplete knowledge of what data currently exists (53%), data are not mapped to a common standard (42%), concerns with participants' consent (41%), and restrictions if data are not downloadable (40%). Key findings of the commissioned research include:

- Ability to access IPD through a central access model would enhance quality of research, and influence the direction
- Future model should include both commercial and academic trial data (71%)
- Reviewed access is most suitable mechanism (61%)
- Datasets should be downloadable for analysis (68%)
- Should include data from all geographic regions (60%)

The commissioned research report made the following recommendations:

- Link current data sharing initiatives and prevent further fragmentation of data landscape
- Establish a global discussion forum of potential funders of IPD sharing initiatives to ensure a joined-up approach
- Establish a central repository or data portal, and scale over time

This is in line with the recent Institute of Medicine recommendations for "trusted impartial organization(s), to convene a multi-stakeholder body with global reach and broad representation to address, in an ongoing process, the key infrastructure, technological, sustainability, and workforce

challenges associated with the sharing of clinical trial data." The discussion at this meeting should help to move in this direction.

ClinicalStudyDataRequest.com currently already provides access to data from eleven pharmaceutical companies. The Wellcome Trust has just taken over the secretariat for the Independent Review Panel to CSDR, which is providing some very useful lessons about how one model of data access works. Utilizing an Independent Review Panel can protect patient confidentiality, ensure appropriate secondary uses of data, review scientific and analytical robustness of research proposals, assess the competence of the data requestor, and maintain public confidence.

But CSDR at the moment only includes data from commercial trials. It will be important to avoid a siloed approach, and to facilitate access to data from both academic and commercial trials. To consider what a future model might look like brings up the following questions:

- Data In: What data, from where? Are data deposited, or kept by the data generator until requested?
- Repository or federated portal: What is the access mode (e.g., open access / reviewed access / stricter controls)? Who operates this? What is the governance model?
- Data Out: How does an IRP operate? Is there some resource to help with curation and de-identification? Can data be downloaded? Who pays?

Moderated Panel #1B: Forward looking strategies and initiatives

- Rebecca Kush, Kathy Hudson, Steven Goodman, Ronald Krall, Nicola Perrin; Moderators: Mark Barnes & Barbara Bierer

The panel discussed the following issues:

- Whether to make all data available or only requested data
    - Not all trials have data that are usable for secondary purposes
    - Of those trials from which data are requested, only a fraction of data are requested; therefore, make data available "just in time" to save resources
    - Need good data standards and curation at data generation stage, then use just in time model
    - Answer might look different for pharma vs academic
- Proposing a portal that is separate from the database
    - Plan where to share data
    - Use registries to find data elements of trials and summary results, and to request subsets with participant level data
    - Note that there are studies for which no data are available

- International trials
  - WHO sets standards for international trials
  - WHO International Clinical Trials Registry Platform (ICTRP) registry to agree on 21st data element [there are currently 20 items that must appear in a register in order for a trial to be fully registered in the WHO Trial Registration Data Set, a proposed 21st item could cover data sharing]
  - Portal should have robust searchable registry with a link to where to get the data
- Data sharing plan
  - NIH will soon require data sharing plan as part of grant application
- Creating a culture of data sharing
  - Involves changing all the different elements of the ecosystem
  - Is an area of learning
  - Requires looking through the lens of communal benefit of sharing
  - Involves looking at the issue of selling data, where that is part of the business model
- Preventing unnecessary future trials
  - Need more research when individual data vs summary data are needed
  - "Just in time" may not work for academic investigators
  - Evaluate utility and value of data being used
  - Determine if there are trials of high value vs lower priority
  - Consider interventional vs observational studies (e.g., Framingham study)
- Cost model
  - Add some cost to trial budgets (e.g. 10% of costs) for data sharing
  - Redcap (Research Electronic Data Capture) for academic trials does not include funding to put data into standard data or record forms
- Identifiable data
  - With more rigorous review and informed consent, consider moving to identifiable data
- Reporting requirements
  - Requirement for open clinical research could lead to individuals not enrolling
  - Many trials are never analyzed nor results reported: Small trials, grant applications without protocol, internally funded studies
  - Shift in behavior and culture should put effort to report every trial

Case studies: history, lessons learned, best practices

Global Alliance for Genomics & Health
- Peter Goodhand (Global Alliance)

Mr. Goodhand outlined the challenge of unparalleled generation of human genetic data and how to unlock its potential in a way that allows data to be shared on a global level and thus empowering new knowledge, new diagnostics and new therapeutics for patients and populations. Data from millions of samples may be needed to achieve results and progress in showing patterns that would otherwise be obscure. Right now, data are typically in silos, analysis methods are non-standardized, and approaches to regulation, consent and data sharing limit interoperability. If we do not act, we risk an overwhelming mass of fragmented data such as electronic medical records in many countries.

What we can do to address these challenges includes:

- working together internationally to ensure interoperability of data and of methods, to harmonize approaches to ethics and regulation, and to promote participant autonomy
- support pilot projects that responsibly and effectively harmonize, analyze and share genomic and clinical data
- engage in professional communities and the public, build trust and encourage appropriate sharing and learning

The mission of the Global Alliance is to accelerate progress in human health by helping to establish a common framework of harmonized approaches to enable effective and responsible sharing of genomic and clinical data, and by catalyzing data sharing projects that drive and demonstrate the value of data sharing. The role of the Global Alliance is to convene stakeholders, catalyze sharing of data, create harmonized approaches, act as clearinghouse, foster innovation, and commit to responsible data sharing. The Global Alliance will NOT directly generate or store data, perform research or care for patients, interpret genomes, and be exclusive to entities that have and share data.

The Global Alliance has 270+ organizational members in 29 countries. Members include universities and research institutes, academic medical centers and health systems, disease advocacy organizations and patient groups, consortia and professional societies, funders and agencies, life science and information technology companies. The organizational structure of the Global Alliance includes a Steering Committee, a Secretariat and various working groups and demonstration projects that involve more than 2,000 individuals. Current working groups include: the Clinical Working Group, the Data Working Group, the Regulatory and Ethics Working Group, and the Security Working Group. Working groups work in task teams on flagship demonstration projects and work products (e.g. Framework for Responsible Sharing of Genomic and Health-Related Data, Catalogue of Activities – eHealth, Consent Tools) which then become publicly available.

Current data sharing projects are undertaken by the members, not the Global Alliance as an organization, though they are catalyzed and supported by Global Alliance coordinators and Working Groups. Their purpose is to drive learning, to identify requirements, to evaluate value and to coordinate activity. Current data sharing projects include the Beacon Project, Matchmaker Exchange, and BRCA Challenge.

International data sharing can be achieved by federation and use of metadata while respecting national and regional restrictions. To realize the benefits, a new and widespread willingness to share data for the greater good and to learn from data is required. Public attitudes towards personal data differ between countries and are changing: there is an increasing awareness of the benefit of sharing and increasing attention to privacy. Collaboration on interoperability is important while there might be competition on implementation.

We need to fully engage with individuals and organizations in **all** continents to be truly global. Individuals are key to creating new tools, frameworks, enablers, projects and opportunities. Organizations are key to ensuring the adoption of best practices and support/reward or responsible data sharing.

Working Group on Data Sharing of the European Federation of Statisticians in the Pharmaceutical Industry (EFSPI)

- Christoph Gerlinger (Bayer)

Dr. Gerlinger presented the European Federation of Statisticians in the Pharmaceutical Industry (EFSPI) that has ten country member groups with about 2,800 pharmaceutical statisticians. The objectives of the EFSPI are: to promote professional standards of statistics and the standing of the statistical profession, to offer a collective expert input on statistical matters to national and international authorities and organizations, and to exchange information on and harmonize attitudes to the practice of statistics in the European pharmaceutical industry and within member groups.

The EFSPI working group was founded after an EMA workshop on data transparency in November 2012. EFSPI council members were represented in various EMA advisory groups on data transparency. EFSPI authored a position paper in 2013, “European Federation of Statisticians in the Pharmaceutical Industry’s position on access to clinical trial data,” and EFSPI commented on EMA’s draft policy 70 on data transparency.

The objectives of the EFSPI data transparency working group are:

- To identify and prospectively prioritize statistical issues in data transparency
- To co-ordinate statistical contributions across Europe to the data transparency debate

- To disseminate relevant information on the topic across the statistical community
- To develop and share a vision of the potential longer term impact of data transparency

Five work streams work on:

- providing input into EMA/EFPIA related data transparency activities
- developing recommendations for re-analysis practices
- assessing the future impact to Biostatistics with increased data transparency
- establishing what are the minimum requirements for sharing data
- ensuring patient confidentiality

Dr. Gerlinger presented three scenarios of best practice for analyses of shared data:

1. To replicate and verify the results in the study report
2. To investigate the original research questions differently or more thoroughly
3. To use the data for a research question that is different from the original objective of the trial(s)

Minimal requirements for data sharing include: Considerations for independent group to evaluate research proposals, considerations for type of access, inter-company collaboration versus separate solutions, minimal details to be included in a research proposal, minimal information data owners share with researchers when data access is approved, recommendations on collaborating between data owners and researchers to improve proposed research request details. On the other hand, patient confidentiality has to be guaranteed and the limits of the patients' informed consent have to be respected.

EFSPI Working groups currently work on best practice documents and aim for presentations at scientific meetings and publication(s) or position paper.

Principles for Integrating and Sharing Clinical Trial Data: Laboratory of Neuro Imaging

- Arthur W. Toga (University of Southern California)

Dr. Toga addressed issues related to scale as data sharing models have to accommodate different types of data that measure different things.

New imaging technology is emerging that characterizes patients and combines 4-dimensional data. The latest imaging technology gives extraordinary resolution to look at brain structure. Neuroimaging study size has increased manifold in the last 25 years. Each neuroimaging scan can spawn many derived images leading to exponential growth. This amplification of data size has to be accommodated. It is also important to link data to tools so they can be searched in a database.

With big data comes big responsibility: artificial correlations such as the per capita consumption of margarine's relation to the divorce rate in Maine have to be avoided as well as incompatible nomenclature, hasty users, and reluctant collaborators.

To share genome sequencing, it does not scale to have centralized systems as for clinical trials and academic studies. Rather, these genetic data are available to the public, and everybody can get an account to access the data within 10 minutes. There is a low bar to the requests which just filters out spam, and there have been no negative consequences.

LONI's infrastructure can support large-scale data efforts. It takes a lot of technology to support this, and sustainability plans. The current lab has 26 Terabytes of memory and 2.4 Petabytes of storage. The new lab that will open this summer will double this capacity.

Data has been uploaded from 17,200 subjects, including 200,000 scans, 121,000 processed scans, and 64 million image files. Over 200,000 neuroimages were archives from 2005 through 2014. From 2006 through 2014, 6,175 data use applications were received which resulted in more than 6 million downloads, with an upward trend. Thus, if data are adequately described, people will come and download it. From 2009 through 2014, 789 manuscripts have been submitted online on one project, mostly by people who were not involved in gathering the data. This shows that getting the data to the scientific community is advantageous.

Models such as GAAIN.org (The Future of Alzheimer's Disease Research) connect with archives around the world. They respect the rules of engagement and do not own the data. Since data are heterogeneous; a mapping strategy had to be developed to integrate the different data standards (e.g., fields labeled Sex, Gender, PTGender, MF). The number of subjects searchable in GAAIN in the first year reached almost 120,000. It includes patients from around the world; adhering to data use requirements of the respective archives. This interface allows doing discovery to see how many subjects match. One cannot download the data, but determine if the data are available which is an important first step to data sharing. The working interface couples simple-to-use tools that talk directly to the database and facilitate utilization with tools that do preliminary analysis such as correlations.

Lessons learned of what works and doesn't work include:

- Data are critical, bad data make bad results
- Design must be clear up front
- Technology and tools must keep pace with development
- Scientific focus should be clear
- Size of data need to be considered
- Duration and sustainability must match with how long data will be accessible
- Mission needs to be clear and must have buy-in
- People must have right personality
- Sociology will not be appropriate if not the right people are involved

- Sharing has to ensure that everybody gets something out
- Expertise is needed at technical, legal, ethical levels
- Leadership must ensure somebody to drive it who is goal-directed
- Funding has to be ensured and might be expensive

NHLBI Data Repository (BioLINCC)

- Michael S. Lauer (NIH/NHLBI)

Dr. Lauer explained that in 1999 he published a paper on heartrate dynamics related to exercise and, through a colleague, found out where other data were available that could be used to replicate findings, which resulted in subsequent publications. This eventually led to BioLINCC.

BioLINCC is a biologic specimen and data repository information coordinating center which was established in 2008 to facilitate access to, and promote use of, the NHLBI repositories. It links phenotypic data to the study vials; provides online access to biospecimens and data; assists investigators with sample selection; and assists NHLBI with the review of requests, the assessment of collection utilization, and the promotion of resource use. It combines the NHLBI data repository which stores and distributes data sets from NHLBI clinical studies with the NHLBI BioRepository which stores and distributes biospecimens from NHLBI clinical studies and provides online access for researchers.

Data from clinical trials have been released 3 years after the final visit or 2 years after publication of primary outcome paper and data from observational epidemiology studies have been released 3 years after the completion of each examination or 2 years after data sets were finalized for analysis. The intent is to have a 2 year protected period of time.

Data de-identification is consistent with informed consent forms, removes obvious identifiers and geography, removes sensitive data when not integral, and other data may be recoded.

To request data from the repository, researchers have to register on the website, fill in a data request form and get IRB approval. After NHLBI administrative review under NIH's IRB is completed favorably, a Research Materials Distribution Agreement is signed that prohibits data transfer and limits use to 3 years, requires acknowledgement in manuscripts/abstracts, and requires notification of any new projects.

Advantages include the relatively simple application process, the ability to combine data from multiple studies, and access to data from closed studies. Disadvantages include limited support and limited access to ancillary data, de-identification may limit the range of applicable research questions, and newly collected data may still be in 'protected time.' Major burdens of the data repository include the

review for completeness and de-identification, investigator support post approval, tracking of three year use limitation, manual surveillance for peer reviewed publications, website development and maintenance, and the annual cost of $1 million.

There has been an increasing trend of new investigators using BioLINCC and international interest, as well as more and more requests. In 2014, roughly half of the requests were for clinical trials, and half for observational studies. The trial with the most requests is ACCORD, a diabetes trial. This has yielded an increasing number of publications, with about 80 publications in 2014.

A study by Gina Wei and Sean Coady has looked at papers published by investigators of original studies vs. investigators who downloaded the study from BioLINCC who did comparable studies and papers. This study found that papers from original investigators were four times more common (~800) than papers with data from the BioLINCC repository (~200). While this quantity of papers was not as high as anticipated, the impact appears to be comparable. The distribution of citation impact from BioLINCC papers and original study papers was about the same.

Panel with Case Study Presenters

- Peter Goodhand, Christoph Gerlinger, Arthur Toga, Michael Lauer; Moderator: Rebecca Li

The panel discussion included the following issues:

- Concern about re-identification (of images)
    - Informed Consent Form (ICF) should make clear that data will be shared
    - Data Use Agreements prohibit users from re-identification, however there is no enforcement process in place
- Informed Consent and BioLINCC
    - The ICF for each study, some of which are decades old, are reviewed to ensure concordance of the ICF with the data to be shared
    - Having a more unified ICF could make it less complex
    - To date, it does not seem secondary analysis has been averted because of informed consent
- Availability of statistical code
    - Suggestion to post the statistical analysis code, so that secondary analysis does not have to start anew
    - The National Health and Nutrition Examination Survey (NHANES), an series of studies designed to assess the health and nutritional status of adults and children in the United States combining interviews and physical examinations, is a model to emulate in that it provides a code library for own data sets

- Ability to understand the wealth of data available
    - Young, aspiring researchers take the time to understand the available data
    - The presence of well-annotated coding libraries helps
    - Dialogue between data generators and those who intend to perform the secondary analysis is helpful and recommended
    - Workshops with young investigators and original investigators (data generators) to understand the data have been conducted and appear to enhance data requests
    - More tools, including start-up tools, and user-friendly interface ideally should be provided
    - Descriptions of the data, such as data dictionaries and FAQs, can help
- Challenge of which data to share
    - If the code is too sophisticated, statisticians have to work to make it easier
- Framework for data sharing
    - There is a need for best practices, especially if a trial is reanalyzed with different findings
- Biostatistics platform
    - Argued for closed system or data use agreement since anonymization is challenging
    - Need efforts to combine data from two trials
- Complexity of interoperability
    - Interoperability is an ongoing process that requires persistent effort
    - Some data types e.g. genomics and images, can be integrated through networks of machines
    - Technological solutions may help to make the data look unified and allow integrated analyses (pulls data from different sources, user does not know from which machine it comes from)
    - Interoperability of clinical data is more complicated
- Cost model for making data available
    - The costs of BioLINCC have been calculated: the budget for 110 studies is $110 million/year, more precise data available (biospecimen vs data): includes cost of working with investigator to prepare data and putting data into platform, does not include time of investigator
    - There is a process by which people have to apply to deposit data in BioLINCC: an external review group has to approve it and the approval takes approximately 3-6 months. The external review group evaluates the significance and likely utility of the proposed dataset

Session II: The Way Forward

Implications of the IOM Report *Sharing Clinical Trial Data*

- Bernard Lo (Greenwall Foundation)

Dr. Lo shared that responsible clinical trial data sharing is in the public interest and there is a momentum for data sharing. The question is not whether to share, but ***what*** types of clinical trial data to share, ***when*** to share, and ***how*** to share. Dr. Lo particularly encouraged coordination across additional stakeholders and providing incentives that can help draw people in.

The IOM Report had 23 public and private sponsors, a committee with a diverse, balanced background, and included IOM peer review. The report included four recommendations with the following relevant points:

- Recommendation 1: Stakeholder Responsibilities
    - Stakeholders in clinical trials should foster a **culture** in which data sharing is the **expected norm**.
    - **Specific incentives** rather than regulations should be provided.
    - **Funders and Sponsors** should require data sharing and provide appropriate support
    - **Journals** should require sharing of analytic data set supporting the published results of a trial
    - **Universities** should require data sharing and consider in promotion
    - **Institutional Review Boards (IRBs)** should consider data sharing when reviewing clinical trials, provide guidance and templates for informed consent, recognize the importance of education and discussion as well as consent forms

- Recommendation 2: What data should be shared and When?
    - Sponsors and investigators should share the various types of clinical trial data no later than the times specified:
        - No later than 6 months after publication: subset of the analyzable data set supporting the findings, tables, and figures in the publication and full protocol, full statistical analysis plan and analytic code (post-publication data package)
        - 18 months after study completion: full analyzable data set, full protocol, full statistical analysis plan, and analytic code (full data package)
        - 30 days after regulatory approval or 18 months after abandonment: full analyzable data set, redacted CSR, full protocol, full statistical analysis plan, and analytic code (post-regulatory data package)
    - Distinguish data sharing from sharing summary results

- Recommendation 3: With whom should data be shared and under what conditions?
  - Current data use agreements
    - Reduce risk to stakeholders
      - Prohibit re-identification
      - Prohibit support competitor's application or unfair commercial use
      - Acknowledge original clinical trial investigators
    - Enhance scientific value of secondary analysis
      - Publish in peer-reviewed journals, share statistical analysis plan
    - Protect public health by notifying sponsors and regulatory agencies of significant safety concerns
    - Role for MRCT to develop and **evaluate**
  - **Employ appropriate privacy protections**, in addition to de-identification and data security
    - Risk in "big data" era is additional information combined with study data
    - MRCT could describe current practice, identify challenges and responses, develop best practices, discuss how to **evaluate and update**
  - **Learn from experience** by collecting data on outcomes of data sharing and disseminating information / lessons learned
    - GSK: N Engl J Med 2014;371:2052-4.
    - Develop case studies, FAQs, common misunderstandings and problems
    - Address common challenges: How to combine data across different sponsors; What is impact of having data sets that cannot be downloaded?
  - Holders of clinical trial data should
    - Have **independent review panel** that includes **members of the public** review data requests
      - Overly restrictive controls inhibit secondary analyses and innovative proposals
  - Key Challenges
    - Technical
      - Current platforms not discoverable, searchable, and interoperable
      - Common data model
      - Common data elements
    - Sustainability
      - Small subset of sponsors, funders and trialists cannot continue to bear costs.
      - Reach out to **nonprofits, device, biologicals**
      - For data sharing to be sustainable, those who benefit from data sharing will need to bear fair share of costs: First need **accurate data on costs;** Limitations of 10% estimate
      - Costs can be reduced by **innovations in data collection and management:** Pre- or non-competitive collaboration?

- Recommendation 4: Stakeholders should work together on key challenges toward a vision for data sharing
    - The take home message was to carry on and **bring in other stakeholders:**
        - Devices, biologicals, start ups, nonprofit and public sponsors, medical journals, universities, disease groups, public

Update from the World Health Organization
- Vasee Moorthy (WHO)

Dr. Moorthy explained his two roles at the World Health Organization: (1) in vaccines development, which includes many clinical trials, and (2) on the Ebola Research & Development (R&D) team, for which no clinical trials were available, but for which clinical trials were commissioned with a very tight timeframe, with the highest standards in place to safeguard quality and safety. Some issues have arisen in regard to information sharing in the context of Ebola, with examples of good and unhelpful practices coming to light.

WHO will be convening a meeting on September 1-2, 2015 on data sharing in the context of public health emergencies, focusing on a global public health perspective and on low- and middle-income countries.

WHO has six core roles:

- providing leadership on matters critical to public health and engaging in partnerships
- shaping the research agenda and stimulating the generation, translation and dissemination of knowledge
- setting norms and standards and promoting and monitoring their implementation
- articulating ethical and evidence-based policy options
- providing technical support, and building sustainable institutional capacity
- monitoring the health situation and assessing health trends

This current data-sharing conference relates to many of these core functions and is thus important to WHO.

In the context of Ebola:

- Research & Development in the context of the Ebola emergency was akin to “building a bridge while walking across it,” trying to get answers as soon as possible. The timelines for generating new information with clinical trials for vaccines broke records: from no clinical trial data with

Ebola vaccine candidates on 1 September 2014 to having many clinical trials fully enrolled in North America, Europe and Africa by the end of December 2014

- Many lessons were learned on how product development could be accelerated, which can possibly be related to other experimental fields
- New pieces of information were available on a daily and weekly basis, but it was variable how results were publicly disclosed and shared
- There a perception by some scientists of an obstacle for sharing information before it has been published in a scientific journal; however, nothing in journals policies precludes sharing of aggregate data on meeting websites, for example, to Dr Moorthy's knowledge

Dr. Moorthy acknowledged an unfinished agenda in regard to results reporting of clinical trials. WHO is working with many different stakeholders to advocate that all clinical trials are registered and results publicly disclosed. This is a very important area for evidence-based medicine, with ethical, scientific and decision-making justifications for requiring that all interventional clinical trials are publicly disclosed. Since much information confirms that many trials remain unreported, WHO is calling for incentives and legislation to ensure that all interventional clinical trials are reported in reasonable timeframes.

Dr. Moorthy stated that the participant level data sharing momentum is important. It is very important that public disclosure of summary results is considered an essential first step for any interventional clinical trial, prior to considerations of participant level data sharing.

WHO has been engaging with data sharing initiatives on several fronts, and will continue to do so. Dr Moorthy congratulated those at the meeting for being in the vanguard of the participant level data sharing movement, which has the potential for many benefits.

Breakout Groups

Participants were divided into four multi-stakeholder groups for two hours to discuss the following topics:

1. Vision of the ideal platform
2. Barriers and incentives
3. Characteristics of the platform
4. Getting data into common platform and accessing data from common platform

Facilitators from the MRCT Center and Deloitte were provided with Facilitator Guides, and all participants received Worksheets (Appendix 4). Groups were asked to report back on the most important items per topic.

Reporting Back: Orange Group

Topic 1: Ideal Platform

- **Characteristics of Common Search Engine Platform** (e.g. the search engines Kayak or Expedia)
- Ability to search for trials residing in databases from all sponsors (academia or industry) and in all locations
    - Advanced Search (beyond simple key words) that meet criteria
    - IRP – tiered approach – when a sponsor places a trial in they denote, tier 1, 2, level of review
    - Limitations for use of data will be specified using search filters
- Use common language and common standards for data
- Non-downloadable – all analysis and data lives in the cloud
- Ability to use the data and easy access to legacy data
- A platform in which all Data is deposited and held seems impractical as there is too much data for a single repository and there are concerns over safeguards and security
- A learning system – constant analysis / safeguards
- Potential for certain researchers to be fast-tracked, pre-qualified "rewarded" and incentivized for sharing data
- Potential for fee based on user load

Topic 2: Barriers and Incentives

Barriers to data sharing	**Incentive**
ACADEMICS - Not obtaining adequate credit (for academics)	• List datasets on NIH biosketches – track citations for utilization of shared datasets • If a click on every figure or data point in a journal publication could open the dataset, then the journal could help track the number of datasets downloaded • Credit and citations for use of datasets to data generator for promotions and tenure
SMALL COMPANIES / BIOTECH - Cost of time, upfront resources (especially for smaller companies, biotech)	Large pharma is proposed to fund the platform disproportionately to biotech and small companies. However, cost of time and resources would need to be factored into costs of clinical trials

Topic 3: Characteristics of the platform

1. Single search engine and single portal

2. Independent, federated platform with a central repository component (for those who prefer not to house their own data i.e. academics, small biotechs).
3. From requester or user perspective the portal should appear seamless - one request could emanate to multiple sponsors
4. Proposed business model – multiple sources or a subscription model funds the non-profit which runs the portal, IRP, data management from generators and repository for those who request it, and implementation and operations of data sharing
5. Platform should be global
6. Stakeholder compliance / adherence - metrics and public reporting and visibility of the lifecycle of the proposal. Transparency of the process. Journals to require citations for journals and enable them to access protocols.

Examples of other Business models

A. NIH –mini-grant 10K fast applications, to pay for doing the data analysis projects. Fund many secondary analyses. Enabler of research progress.

Reporting Back: Blue Group

Topic 1: Elements of Ideal Platform: all are essential

- Global model; need national or regional reporting
- Academic & commercial
- Not Disease specific: but to be able to pull data for a specific disease
- Federated model, not centralized data warehouse
- Independent review panel has final say
- Non-downloadable, but easy to work, with potential exceptions: ad-hoc request; controlled release
- Sharing with approved researchers only
- Registry of approved requests
- Search function: Description of data & availability

Topic 2: Barriers and Incentives

- Cost is primary issue
- Provide recognition of data provider / Credit (negative & positive)
    - Requires culture change, including culture change of data use
- Need regulation / requirements for universal data sharing (a major funder, trade organization can drive regulation)
- Patient privacy is a “solvable” problem, has a cost, and is essential to address
- Consent is essential going forward
- Harmonization of standards is important but will take time

Topic 3: Characteristics of the Platform

- Global trusted entity must drive this; need associated regional authorities
- Business model: need vision & communication strategy
- Need common process, multiple players can drive it
- Feasibility assessment needs to be made, may involve multiple companies; then independent review
- Business model considerations: (1) fee for request, (2) public-private partnership models (all stakeholders pay a percentage of the cost), (3) data sharing required as part of grant proposal, (4) initial set-up to leverage best practices of IT, (5) subscription model

Topic 4: Data Principles and Access

- Availability: Going forward: all data available, ad-hoc for older data
- Curation (going forward): anonymizing data, adhering to standards at collection, provide associated meta data
- Allow for harmonization & merging of data sets
- Data from all regions to be included
- Use of data: include hypothesis generation, reproducing results, hypothesis testing, meta-analyses, secondary analysis, NOT for educational use, NOT for competitive reasons & criminal investigation
- Independent review board's roles: 1) Yes/No decision, 2) facilitate between researcher and data provider

Reporting Back: Yellow Group

- Governance model:
    - **Organized multi-stakeholder group is needed,** associated with existing or new organization
    - Have an organized way to make decisions going forward
    - Possible to make SAS compete with other platforms

- Preference for **open access** model and downloadable data,
    - Need for DUA
    - Safeguard privacy
    - Could bring civil action against data user who re-identified data

- Incentivizing:
    - Develop a **culture of data sharing**:
    - Journals: as a condition of results of research to be published, commit to data sharing plans

  - Give people credit for re-use of data: the better the data set, the more use of the data set
  - Reward people for ability to access data, not reward "free rider"

- **Downloading**
  - As long as integrity of individual was protected, downloading data was not alarming

- Preference for common data platform and common elements
- Cost:
  - For industry: cost of doing business
  - For academia: fee-for-service

Reporting Back: Green Group

- Cost sharing
  - Who pays for time & resources & opportunity cost
  - How to lower cost going forward: commitment to data standards & curation of data would bend cost curve
- Data standards:
  - What is in CDISC and what is not
  - Do we want to promote trials that can't align with CDISC Standards?
- Different kinds of data:
  - Safety data
  - Endpoint & efficacy data
  - Understand what is in data standards
- Data quality depends on what it is used for
- Data standards & meta data standards:
  - Common terminology base for search engine
  - Underlying data set: what data sets would be available to the public and would be in the platform?
    - Data by which an adjudication panel renders its decisions, for instance, is not generally available in the final dataset
  - Sponsor should provide data for reasonable request
- Platform:
  - Data generator should be allowed to hold their own data, some academics & biotechs may not want to
  - Should have ability to be a **mixed model: search data generators and be able to host data**
  - IT platform: search engine capability, e.g., clinicaltrials.gov has the search ability, WHO platform has capacity for 20 countries/registries

- **Ability to search** needs to be a piece of the platform
- System includes communication and tracking & security elements, enforcement potential is important
- Portal held by independent not-for-profit needs to be built
- **Cost model:**
  - initial capital shared amongst industry, government, foundations, not borne by academics, afterwards there would be a user fee for all users;
  - if additional data sets requested & data needs to be prepared: data requester should pay, unless data generator wants to make a gift of the data; default should be that data user pays for development of data;
  - If same dataset is used subsequently, it would be possible for a fixed proportion of the initial cost is levied to offset database/platform costs or for providing funds for education or development of future issues
- Journals:
  - not all journals require the same standardization, each journal may have a different expectation & standard;
  - request for one standard for data sharing, limited to the data set that informs the primary endpoints that are reported
- Data sharing plan:
  - Concern about having to fill out data sharing plans for multiple trials = companies could generate commitments to always share a certain kind of data in a certain way
  - Create an **expectation** where transparency is normative, if data are not shared, provide a reason why not = negative visibility to inform culture

- Barriers: all barriers listed on the worksheet (see Appendix 4) are essential to address

- IRP: independent body to manage:
  - Principles and guidelines, framework, data generators could sign on to be a member
  - Commitment to principles and guidelines, not necessarily to data platform
  - IRP to include scientific procedure, meta data, analysis plan, etc.
  - Capability of team to execute the proposal
  - Not duplicative review
  - Ability of data generator to review manuscript before publication, if so desired

- Platform elements
  - Agreed with everything said earlier
  - Exception: platform should be more flexible:
    - Review board
    - Non-downloadable

        - Sometimes comfortable with data downloadable, one system should incorporate those abilities
        - Should not be shared with general public, for approved proposal only
        - Principles of platform and of data sharing cannot be changed retrospectively: when data are posted as non-downloadable, governance cannot change this retrospectively

- Topic 4: questions were general
    - Anonymization: de-identification: defer to our statistical colleagues
    - Make provision in informed consent that people understand the small risk

Session III: Commitments and Next Steps

Facilitators outlined themes from breakout sessions. A number of multi-stakeholder panels were asked to respond to these themes and to make any appropriate commitments for the future.

Topic 1: Vision of ideal platform

Summary of input from breakout groups

Platform

- An integrated solution is necessary, and the platform should be sufficiently flexible to accommodate and link different kinds of data (clinical structured data, imaging, 'omics, etc.)
- Platform should provide a single portal for entry
- Platform should be transparent, discoverable
- Platform to provide search engine, utilizing CT.gov or ICTRP in background if sufficient (after testing) or creating search engine.
- Data generators should have the option of hosting their own data (accessible through portal) or delivering data to the platform to host and maintain
- Data should not be downloadable (default position), but option to download should be provided
- Robust metadata should be accessible through portal
- System should provide:
    - a single portal for entry with search engine as described above
    - tools for viewing, analyzing, manipulating, integrating data as a potential side-car to portal
    - visibility to policies, forms, data-use agreements, tools, other common or accessible attributes
    - ability to submit proposals, request review, etc.
    - tracking and accountability
    - communications (including listserv)
    - security
- Global federated model facilitated by single, trusted, not-for-profit entity
- Governance is critical:
    - Representative academic, industry, patient/patient advocates, journal, not-for-profit organizations participation, reflecting consensus model
    - Regulatory agencies as advisors/observers (to be discussed)
    - Responsible for policies, and for providing oversight of management, operations, finances, and enforcement

    - Policy changes, if any, are not retroactive.
- Principles and guidelines of the system and process to be defined, and membership of the "federation" offered
- Incentives are needed to drive involvement and commitment to single model

Data Access

- Data are available ad-hoc for completed studies, and for all new studies going forward after date certain
- Data generator may make some or all data open access, but default is review required
- Data generator may make data downloadable, but default is data not to be downloadable
- Data generator responsible for ensuring data use is consistent with informed consent provisions
- Datasets of interest go beyond that which informs publication tables.
- Individual participant-level data required
- In principle other data may be requested; reasonable request will be honored.

Data Request Review

- Data generator may review and approve request, or decline opportunity to review
- If data generator does review or does not provide access, request reviewed by Independent Review Panel (IRP)
- IRP
    - Has final authority
    - Will review:
        - Scientific proposal
        - Statistical analysis plan
        - Credentials and capability of research team to execute proposal
        - Data availability to execute scientific project as outlined
        - Participant privacy protections and ICF concordance

- Criteria for review and approval to be defined and uniform
- Avoidance of duplicative review, collaborative approach

Data Use Agreements

- Review prototype provided, and please provide comments
- To include:
    - Requirement to make secondary data analysis available
    - Publication in peer-reviewed journal as goal
    - Accountability

- Enforcement important, and to be discussed

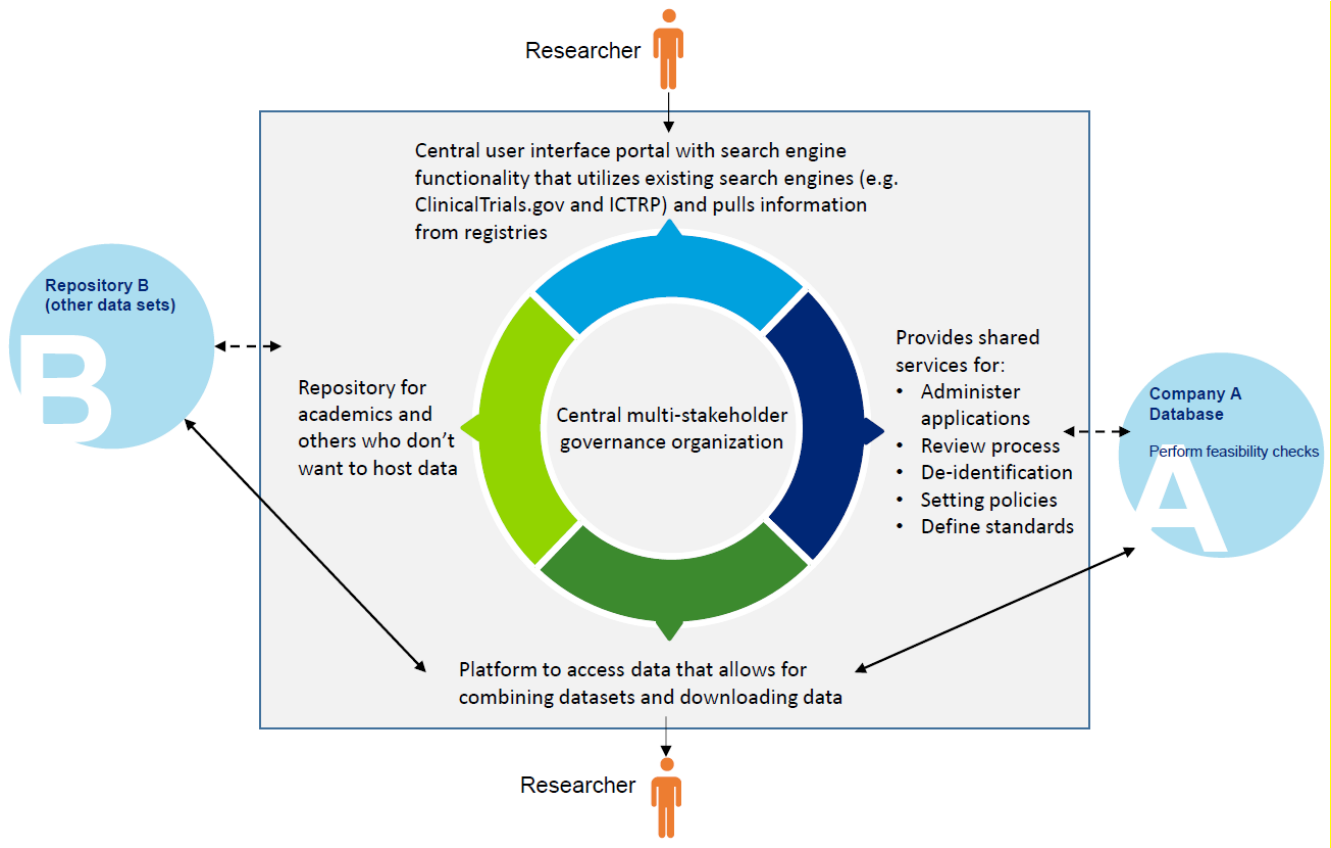
Data Standards

- Go-forward commitment to CDISC
- Some data elements conform more readily to CDISC than others, but commitment to enhance CDISC
- Iterative vision, not to exclude current datasets that do not conform to CDISC
- Global regulatory agencies to require CDISC
- Metadata standards must be developed

Data Sharing Plan

- All grants, clinical trial contracts, and proposals will include data sharing plan
- 21st data element in WHO registry an expectation: to define what data to be shared, when and how
- Organization (company, consortium) can define and commit to terms of data sharing, and reference such commitment

Proposed Model



Response from key stakeholders and discussion

-Joseph Ross (Yale), Jessica Scott (GSK), Deborah Zarin (NIH), Thomas Peppard (Gates Foundation), Julie Ingelfinger (NEJM), Kay Dickersin (JHU); Moderator: Mark Barnes

Responses from stakeholders included the following issues:

- No trials should be left out, including Phase I trials and legacy trials
- Move forward with a review process by an **independent panel** rather than data generator review
- Disease specificity seems to be a complication
- Consider **existing infrastructure (registries)** that can be utilized
    - Fix the current systems first (i.e. trials not registered, too much diffuse information)
    - Aggregate what is already out there
    - Consider lessons learned from ClinicalTrials.gov
    - Make existing global registries interoperable
- Don't combine features of platform with features of **policy**
- Journals don't want to be holders of data, but require protocols, statistical analysis plan, sometimes consent form, and raw data that goes into figures and will likely ask for 21st element
- **Federated model** would be received enthusiastically
- Consider **open access**
- Don't reject data that don't conform to standards
- Consider how to provide access to legacy trials for meta-analyses
- Do more research on what it realistically **costs**
- Clarify the **principles** that we are aspiring to. "What are we doing this for?"

Discussion points from the audience:

- Ensure the new model encompasses the **full scope of clinical trials,** even those not adhering to standards and legacy trials
- Important to have a **collective movement of all journals** requiring protocols
- **Draw on existing forms** such as data request forms and distribution agreements from 67 years of Framingham study
- Definition of which "interventional studies" to include has to be clarified
- **Take advantage of existing resources**: WHO/Australia(?) has a platform with a search function
- Realize that most countries do not have a trial register

Topic 2: Incentives to address barriers and gaps

Summary of input from breakout groups

Incentives to Drive Non-Adopters

- **ICMJE and domain-specific professional societies** can be used to leverage researchers and research institutions into data sharing commitments
- **Professional recognition** (e.g. NIH, academic institution) by number of citations for dataset being used
- Those who share data: **priority access** to other data, and/or have access fees discounted or waived?
- **Preference** to researchers and entities that buy into the system
- Provide **credit on grant applications** based on history of data sharing.
- Create culture driven partly by **journal publishing requirements, academic institution and faculty expectations, learned society standards**
- Transparency is normative
- Organizations, institutions--and individuals--will commit to transparency
- Organizations and institutions will publish their policies and procedures
- All have a right to say "no" but reason for refusal or declination should be made public.

ICMJE

- Request standard policy and expectations across journals
- Minimal data required to be shared includes datasets that inform data tables and figures publication
- Preferable to require data and individual patient data concordant with primary outcome of registered trial, with secondary endpoint data if reported.
- Timing of release in relation to publication to be discussed
- Process for consideration of rebuttable presumption to be discussed

Response from key stakeholders and discussion

-Pamela Gavin (NORD), Benjamin Rotz (Lilly), Frances Rawle (Med Res Council, UK), Heather Pierce (AAMC), Elizabeth Loder (BMJ), Stuart Buck (Arnold), Andrew Emmett (BIO), Moderator: Barbara Bierer

Responses from stakeholders included the following issues:

- **Concerted action of journals** is important; small journals are particularly vulnerable
- Journals could **give credit to data generators and data sharers** that translates into academic credit
- Would be helpful for Independent Review Panel to consider what to do if people do not abide by Data Use Agreement
- An incentive would be to see the **benefits of data sharing** ("what's in it for me?" for spending millions of dollars for data sharing)
- Industry could **partner with academic institutions** for joint publications
- This will take a huge **culture change,** with 4 elements:
    - Costs: resources need to be allocated for preparing data
    - Culture: data sharing has to be valued
    - Credit: data sharing has to be credited
    - Community: without a community movement, it will be difficult to move one institution at a time
- Everyone has to feel they have "skin in the game"
- Shifting culture takes time
- Barriers include **opportunity cost**, coupled with **unclear perception of the benefits**
- Recognition is a key issue for academics
- Some companies have published **principles on data sharing**
- **Bring in all stakeholders**, including smaller journals and biomedical companies
- Realize that this is a very **competitive** field
- **Develop case studies** where data sharing has led to good science
- **Quick wins**: develop a meta data model in 12 months to capture attention
- **Harmonization** of data standards is important
- Patients are interested in sharing information, especially if there is an unmet medical need
- Perhaps have a task force that looks at **de-identification** of data
- Governance model: **Multi-stakeholder model to learn from** : Medical Device Innovating Consortium (mdic.org)
- "Don't let the perfect be the enemy of the good"
- Need buy-in from other players and financial sustainability
- Incentives can be provided via:
    - federal law
    - medical journal
    - automated production of data dictionary
    - university tenure and promotion committees

Discussion points from the audience:

- Promote data sharing as an **ethical and scientific requirement**, not a legal and bureaucratic requirement

- University promotion committees include anything that is **easily measurable**; thus, if "data sharing" can be measured, it will be incorporated
- There is some flexibility with the timelines proposed in the IOM report (e.g., 18 months)
- **Patients are important stakeholders**
- Consider whether the problem is **creating a platform or incentivizing its use**: some have invested millions of dollars and get limited use
- If investment is too high, people will not utilize the system: use an **open model** that does not require people to fill out forms

Topic 3: Characteristics of platform

Summary of input from breakout groups

Business Proposition

Points of Agreement

- Investment in infrastructure, governance, and sustainability is necessary
    - Need for value based incentives and a fair/equitable funding model
    - Cost efficiency appear over time, e.g. evolving data standards will improve efficiency
- Entry barriers for small biotech/device and academics need to be offset
    - Requires a cultural shift in how data usage is viewed, e.g. Investigator recognition based on the citation frequency of data set
    - Journals should require data sharing as condition for consideration for publication

Areas for Discussion

- What is the sustainable value proposition?
- Multiple funding channels available
    - Industry pay-to-play, subscription-based service, fees for data preparation, federal funding, foundations, NIH $10K fast track mini-grant, tiered model based on data your organization makes available

Financial Considerations

- Build (capital costs) of system and IT infrastructure to be multi-stakeholder commitment including government(s), foundations, and industry
- Running costs will assume decreasing dependence on philanthropy as use increases
- Long term: costs borne by user fees

- Additional costs of preparing unique datasets to be assessed and paid by secondary data user
- If same dataset used subsequently, then fractional costs of development will return to offset costs of system or as discretionary funds
- Data generators have ability to donate costs for specific projects
- Assumption that in silico research will be appropriate for grant funding support as research projects

Response from key stakeholders and discussion

- Daniel Moreira (Mayo), Jennifer Miller (Harvard), Hanns-Georg Leimer (Boehringer), Kathy Hudson (NIH), Davina Ghersi (Natl Health & Med Res Council, Australia), Moderator: Mark Barnes

Responses from stakeholders included the following issues:

- Distributing the cost: fee for subscription needs more discussion
- Create **data standards** with the use in mind
- Allow room for sharing raw data as well as processed data
- **Data Sharing Protocol:** evaluate the risk as not all data have the same risk associated
- **Search engine** is key to show what data sets are available and not
- Platform should be customizable to the needs of the user and have flexibility
- **Costs** include: cost to prepare data (which is cost of business for industry) and cost of establishing and maintaining the platform (users should contribute)
- Does not need to be a SAS platform
- Consider **self-regulating** system
- Platform needs to be **global**
- For governance of platform, consider Wellcome Trust, MRCT, Gates Foundation, or a combination
- Need incentives for data sharing and recognizing best practices
- Suggested **transparency ranking** to be included in index: add column on data sharing
- Increase trust/trustworthiness and awareness
- Users must have “skin in the game”
- **Data standards** should not be too prescriptive that it prevents people from using it
- Encourage collaboration with existing **global agencies** (e.g., WHO , Cochrane, OECD)
- Look for global standards and value added to existing initiatives
- Australia has a Code of Conduct for Research which could be edited to include requirements of making results publicly available

Discussion points from the audience:

- Valuable studies that do not conform to standards should be listed to inform the user
- A federated model would be broad enough to **incorporate different approaches** and adjust to the needs of the user
- Move to a system of **metrics**: e.g., University of Oxford conducts audit at university level re which clinical trials are registered, reported, audited

Topic 4: Getting data into platform and accessing data from platform

Summary of input from breakout groups

Characteristics of the Platform / Access Mechanisms

- **Level of Data detail** made available
- Method of **data standardization**
- Data **non-downloadable**, with case-by-case exceptions
- **Data are stored** by governing body or by the data-generator
- Who handles **data curation responsibility**?
- Sponsors agree on **DUA and ICF template for streamlined process**
- **Transparent review system**
    - **IRP review** (but what are criteria for review?)
    - Registry of denied requests with explanations
    - IRP – tiered approach – when a sponsor places a trial in they denote, tier 1, 2, level of review

Protecting Research Participants

- ICF should disclose the breadth of data sharing requirements and other data sharing obligations
- DUA with pledge not to re-identify and not to re-use or re-distribute data
- DUA pledge so clear that aggrieved participants could have basis for lawsuit against offending data user
- Consider keeping data sets that could readily identify participants (e.g., peds studies, rare disease studies) behind firewall and do not allow downloading
- Statute with civil/criminal penalties for re-identifying and causing harm to research participants

Anonymization and Biostatistics

- Common anonymization standards
- Biostatistical tools developed
- Cooperative statistical methodologies

Response from key stakeholders and discussion

- Catrin Tudur Smith (U. Liverpool), Sandra Morris (J&J) Caroline Stockwell (Pfizer), Martha Brumfield (Critical Path), Ronald Krall (U. Pittsburgh), Moderator: Barbara Bierer

Responses from stakeholders included the following issues:

- Make sure that indexing **catalogue** is complete
- Build on the good efforts already out there
- Governance is critical: need a **dedicated staff/head** to lead this
- Provide flexibility for data generator and don't insist on data standards
- Trials must be easy to identify and link within the registry
- The key is to have a **flexible platform** that allows for different mechanisms, not one-size-fits-all
- Consider structures currently in place to share data
- Some existing systems are disease-specific, but open to adding flexibility
- Data Use Agreement is very important
- Some analysis tools are cost prohibitive, built them into the system
- Build in additional safeguards to reduce **re-identification risk**, especially if data from multiple data sets are combined
- Consider **public-private partnership** to move this forward
- Define specific objectives early on
- **Costs** have to be shared, not rely on industry, everyone needs "skin in the game"
- **Need for flexibility** (for parameters for data input) is paramount
- Need common principles for data input and use of data
- Consider a **self-regulatory** mechanism
- Constraint around anonymization of legacy trial data
- Europe has different privacy laws than HIPAA
- **Costs** include 4 categories: preparing data, running this entity, administrative costs, analysis
- **Mini grants** could stimulate secondary research
- Consider the culture of responsible human subject data sharing: responsible conduct includes investing In preparing data properly
- This entity has a **power of community,** which is a major benefit of creating the entity

Discussion points from the audience:

- Use mini grants to get academic ideas on existing platforms
- Granting mechanism: attract people early for good quality of research, don't make grants too big or too small

Summary

Mr. Mark Barnes thanked funders of this conference and summarized:

We need a federated model "with many flowers blooming" to improve opportunities for research.

We were charged to continue the conversation in a structured way, involving all interested stakeholders.

The discussed governance model has the following requirements:

- Determine if a new or existing entity shall govern and take responsibility to convene
- Primary purpose has to be to improve collaborative science in the interest of human health:
    - Efficient – collaboration within initiative must be more efficient than functioning outside the institution
    - Sustainable – might be started with any sources of funds, has to be sustainable
    - Voluntary – needs to foster a self-culture of improving the conditions under which science is done, bolstered under FDAAA, EMA, NIH, ensuring participation is done within regulations
- Needs to be broad enough in design to accommodate those in the forefront and those who lag behind
    - Must incorporate and respect existing systems, e.g., Framingham Study
    - Broad enough to accommodate national requirements, e.g. anonymization of data
    - Must allow companies to comply with their own particular regulatory regimes
- Must interact and enhance regulatory initiatives (e.g., CT.gov, WHO platform)
    - Expand the knowledge base of what exists
- Protect study subjects (participants) and re-identification
    - Improve Informed Consent process and Data Use Agreements
    - Articulate best practices

Appendix 1: Meeting Participants

Last Name:	First Name:	Company:	Affiliation	Group
Adams	Monique	Bristol-Myers Squibb	Industry	Blue
Aldinger	Carmen	MRCT Center at Harvard	Academia	Blue
Barnes	Mark	Ropes and Gray LLP/ MRCT Center at Harvard	Academia	Yellow
Bierer	Barbara	Brigham and Women's/ MRCT Center at Harvard	Academia	Green
Bruening	Johann	Bayer Pharma AG	Industry	Blue
Brumfield	Martha	Critical Path Institute	Nonprofit	Blue
Buck	Stuart	Arnold Foundation	Nonprofit	Green
Butterworth	Trevor	Director, Sense About Science USA	Nonprofit	Yellow
Childers	Karla	Johnson & Johnson	Industry	Orange
Cohen	Theodora	Harvard Clinical research Institute	Nonprofit	Blue
Coulbourne	Kelly	AstraZeneca	Industry	Green
Cullinan	Patrick	Takeda Pharmaceuticals	Industry	Orange
D'Agostino	Ralph	Framingham Heart Study	Academia	Green
Dickersin	Kay	Johns Hopkins Bloomberg School of Public Health	Academia	Yellow
Dogas	Dimitrios	MRCT Center at Harvard	Academia	n/a
Emmett	Andrew	BIO	Industry	
Gavin	Pamela	National Organization for Rare Disorders	Nonprofit	Yellow
Gerlinger	Christoph	Bayer Pharma AG	Industry	Orange
Ghersi	Davina	National Health and Medical Research Council	Government	Blue
Goodhand	Peter	Global Alliance for Genomics and Health	Academia	Blue
Goodman	Steven	METRICS / Stanford	Academia	Yellow
Guo	Elizabeth	Harvard Law School/ MRCT Center at Harvard	Academia	Green
Hill	Nina	Pfizer	Industry	Blue
Hudson	Kathy	NIH	Government	Yellow
Ingelfinger	Julie R.	Mass. Medical Society/NEJM; Harvard Med School	Academia	Green
Kelley	Russell	Burroughs Wellcome Fund	Academia	Blue
Koft	Joanna	Biogen Idec, Inc	Industry	Orange
Krall	Ronald	University of Pittsburgh	Academia	Orange
Kress	Barbara	Merck	Industry	Blue
Krumholz	Harlan	Yale University	Academia	Green
Kush	Rebecca	CDISC	Nonprofit	Yellow
Lauer	Michael	NIH	Government	Orange
Leimer	Hanns-Georg	Boehringer Ingelheim	Industry	Yellow
Letvak	Laurie	Novartis	Industry	Green

Levenstein	Marcia	Pfizer	Industry	Yellow
Li	Rebecca	MRCT Center at Harvard	Academia	Orange
Lingler	Nick	Deloitte	Industry	Yellow
Lo	Bernard	Greenwall Foundation	Academia	Green
Loder	Elizabeth	British Medical Journal	Academia	Orange
Loechel	Michael	Deloitte	Industry	Orange
Lynch	Gerard	Biogen	Industry	Yellow
Lyons	Pete	Deloitte	Industry	Blue
McCarthy	Justin	Pfizer	Industry	Blue
Miller	Jennifer	Harvard	Academia	Blue
Moorthy	Vasee	WHO	Nonprofit	Green
Moreira	Daniel	Mayo Clinic	Academia	Green
Morris	Sandra	Johnson & Johnson	Industry	Yellow
Nolan	Sarah	University of Liverpool	Academia	Blue
O'Callaghan	Jennifer	Wellcome Trust	Nonprofit	Yellow
Okada	Ellie	Boston Cancer Policy Institute	Nonprofit	Orange
Pencina	Michael	Duke Clinical Research Institute	Academia	Orange
Peppard	Thomas	Bill & Melinda Gates Foundation	Academia	Green
Peronace	Laura	EMD Serono	Industry	
Perrin	Nicola	Wellcome Trust	Nonprofit	Orange
Peterson	Eric	Duke Clinical Research Institute	Academia	Blue
Pierce	Heather	Association of American Medical Colleges	Nonprofit	Orange
Rawle	Frances	Medical Research Council, UK	Government	Blue
Rockhold	Frank	GlaxoSmithKline	Industry	Yellow
Ross	Joseph	Yale University	Academia	Blue
Rotz	Ben	Eli Lilly	Industry	Blue
Scott	Jessica	GlaxoSmithKline	Industry	Orange
Shields	David	Takeda Pharmaceuticals International Co.	Industry	
Stockwell	Caroline	Pfizer Inc	Industry	Green
Stroud	Clare	Institute of Medicine	Nonprofit	Yellow
Subbiah	Ponni	Critical Path Institute	Nonprofit	Green
Sudlow	Rebecca	Roche	Industry	Green
Toga	Arthur	University of Southern California	Academia	Yellow
Tucker	Katherine	Roche	Industry	Blue
Tudur Smith	Catrin	University of Liverpool	Academia	Green
Varghese	Adarsh	Harvard Law School/ Harvard MRCT	Academia	Yellow
Williams	Mark	ACI Clinical	Industry	Yellow
Woolston	Crispin	Sanofi	Industry	Orange
Young	Matthew	Deloitte	Industry	Green
Zarin	Deborah	NIH	Government	Green

Appendix 2: Meeting Agenda

Promoting Clinical Trial Data Transparency Conference

When: March 30 & 31, 2015

Where: Harvard Faculty Club, 20 Quincy Street, Cambridge, Massachusetts

Who: Data generators and data users, academic and government agencies, journal leadership, and other stakeholders (by invitation only)

Conference Objectives

- To discuss high-level principles with the explicit goal of developing an approach whereby participant level clinical trial data could be integrated, enabling researchers to access and combine data across the various platforms and sponsors.
- To present and deliberate centrally managed and federated models.
- To discuss implementable solutions to realize harmonization and enable broader data sharing across platforms.

Agenda

Day 1 To introduce current approaches, lessons learned and criteria for harmonizing data sharing platforms

Time	Topic	Speakers
7:30-8:00	Registration and Breakfast	
8:00-8:45	Welcome Remarks Introduction: The Potential Scope of the Data Sharing Issue Summary of pre-conference survey: commonalities, limitations, and gaps of current data sharing approaches	MRCT Center at Harvard: Barbara Bierer, Mark Barnes Deloitte Consulting

Session I: Lessons Learned and Current Needs

Time	Topic	Speakers
8:45- 9:20 (6-7min each)	Session #1A: Experiences & best practices from current initiatives: • Clinical Study Data Request platform • Duke Clinical Research Institute (DCRI) • Merck • Pfizer	 • Frank Rockhold (GSK) • Eric Peterson (Duke) • Barbara Kress (Merck) • Justin McCarthy (Pfizer)

	• Yale University Open Data Access Project (YODA)	• Joseph Ross (Yale)
9:20 – 9:50	Moderated Panel discussion of key stakeholders from Session #1A: What are the key principles and best practices that can be harmonized and agreed upon? Questions from the audience	Speakers from Session #1A Moderators: Barbara Bierer & Mark Barnes (MRCT Center at Harvard)
9:50 – 10:25 (6-7 min each)	Session #1B: Forward looking strategies and initiatives: • CDISC: role of data standards in harmonization • National Institutes of Health (NIH) • Patient-Centered Outcomes Research Institute (PCORI) • University of Pittsburgh • Wellcome Trust	 • Rebecca Kush (CDISC) • Kathy Hudson (NIH) • Steven Goodman (Stanford) • Ronald Krall (Univ. Pittsburgh) • Nicola Perrin (Wellcome Trust)
10:25-10:55	Moderated Panel discussion of key stakeholders from Session #1B: What are the key principles and strategies that can be harmonized and agreed upon? Questions from the audience	Speakers from Session #1B Moderators: Barbara Bierer & Mark Barnes (MRCT Center at Harvard)
10:55-11:15	BREAK	
11:15 – 12:30 (10-15 min each)	Case Studies: history, lessons learned, best practices • Global Alliance for Genomics & Health • Working group on data sharing of the European Federation of Statisticians in the Pharmaceutical Industry • Laboratory of Neuro Imaging (LONI) • The NHLBI BioLINCC Program Questions from the audience	 • Peter Goodhand (Global Alliance) • Christoph Gerlinger (Bayer) • Arthur Toga (University of Southern California) • Michael Lauer National Heart, Lung, and Blood Institute (NIH/NHLBI)

12:30 – 1:15 LUNCH

Session II: The Way Forward

Objective: To devise and agree on common standards, criteria, approaches and harmonization of data platforms for the sharing of clinical trials data.

1:15 – 1:45	Implications of the IOM Report *Sharing Clinical Trial Data*, including proposed sustainability models	Bernard Lo (Greenwall Foundation)
1:45 – 2:00	Introduction to breakout sessions: All groups discuss all topics, facilitator reports back in plenary. Topics: 1. Vision of ideal platform 2. Barriers and incentives 3. Characteristics of platform 4. Getting data into common platform and accessing data from common platform	MRCT Center at Harvard
2:00 – 4:00	Breakout Groups convene	Deloitte and MRCT Center moderators

4:00 – 4:15 BREAK

4:15 – 5:15	Facilitators report back from each group (10-15 min per group)	Deloitte and MRCT Center moderators
5:15 – 5:30	Closing of Day 1	MRCT Center at Harvard
6:00 – 8:30	Conference Dinner	Adolphus Busch Hall, Harvard Art Museums, 27 Kirkland Street, Cambridge

Day 2 **Session III: Commitments and Next Steps**

Objective: To delineate solutions, next steps and commitments

8:00–8:30 Breakfast

Time	Topic	Respondents
8:30-9:20	**Developing a new model for data sharing** Response from key stakeholders to Topic 1 (Vision of ideal platform) • Academic vs commercial • Disease specific • Portal vs repository • Federated vs centralized • Learned Intermediary (Review Board) vs Data Generator vs Open Access Model • Downloadable vs non-downloadable • Sharing with general public vs approved researchers Discussion: Questions from audience Commitments of stakeholders	Deloitte / MRCT Center at Harvard • Joseph Ross (Yale) • Jessica Scott (GSK) • Deborah Zarin (NIH) • Thomas Peppard (Gates) • Julie Ingelfinger (NEJM) • Kay Dickersin (JHU)
9:20-10:10	Response from key stakeholders to Topic 2 (Incentives to address barriers and gaps) • Incentives for expanded data sharing for non-adopters • Harmonization of standards • Privacy concerns • Criteria to use data • Standards for de-identification of data • Accountability toward research participants Discussion: Questions from audience Commitments of stakeholders	Deloitte / MRCT Center at Harvard • Pamela Gavin (NORD) • Benjamin Rotz (Lilly) • Frances Rawle (Medical Research Council, UK) • Heather Pierce (AAMC) • Elizabeth Loder (BMJ) • Stuart Buck (Arnold)
10:10 – 10:30	BREAK	

10:30-11:20	Response from key stakeholders to Topic 3 (Characteristics of platform) • Business model and governance • Involvement of stakeholders: existing vs new group to review and assess standards • Assessing compliance & adherence to standards • Global or regional • Equitable distribution of costs of data sharing across both data generators and users Discussion: Questions from audience Commitments of stakeholders	Deloitte / MRCT Center at Harvard • Daniel Moreira (Mayo) • Jennifer Miller (Harvard) • Laurie Letvak (Novartis) • Hanns-Georg Leimer (BI) • Kathy Hudson (NIH) • Davina Ghersi (Natl Health & Medical Research Council, Australia)
11:20-12:10	Response from key stakeholders to Topic 4 (Getting data into common platform and accessing data from common platform) • Access mechanism • Protection of personally identifiable information • Independent review • Data Use Agreements • Appropriate data availability standard defined • Pre-uploaded vs as-needed Discussion: Questions from audience Commitments of stakeholders	Deloitte / MRCT Center at Harvard • Catrin Tudur Smith (University of Liverpool) • Sandra Morris (J & J) • Caroline Stockwell (Pfizer) • Martha Brumfield (Critical Path Institute)
12:10 – 12:30	Wrap Up & Next Steps: Summary of proposed new model for data sharing and key commitments of stakeholders	MRCT Center at Harvard

12:30-1:30 LUNCH

Appendix 3: Speaker Biographies

Mark Barnes, J.D., LL.M, is the faculty co-chair of the Multi-Regional Clinical Trials Center at Harvard and practices law as a partner at Ropes & Gray LLP, where he represents academics institutions and industry in matters related to research with humans and animals, clinical trials, research grants and contracts, and research fraud. Mark teaches health care law and research law as a faculty member at Harvard Law School and formerly served as the associate provost and senior research officer for Harvard University. Mark has served as executive vice president and chief administrative officer at St. Jude Children's Research Hospital, had held senior appointed positions in the New York State and City departments of health, was the founding executive director of Harvard PEPFAR's AIDS treatment programs in Nigeria, Tanzania, and Botswana, and headed the national HIV/AIDS lobbying effort in 1995-96 as the executive director of AIDS Action Council, the DC-based national AIDS advocacy coalition. He is currently the NIH ethics advisor to the HPTN 071 trial of HIV testing and treatment interventions with one million participants.

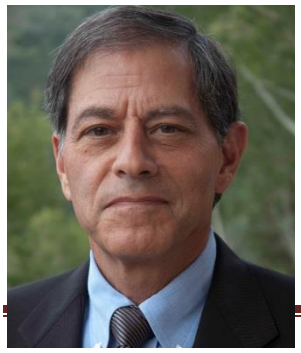
Barbara E. Bierer, M.D., Dr. Barbara Bierer is the faculty co-chair of the Multi-Regional Clinical Trials Center at Harvard University (Harvard MRCT), a Professor of Medicine, Harvard Medical School and Brigham and Women's Hospital, Boston and a hematologist/oncologist. She is the Director of the Regulatory Foundations, Ethics and the Law Program of the Harvard clinical and translational sciences center. Previously she served as senior vice president, research at the Brigham and Women's Hospital for 11 years, and was the institutional official for human subjects and animal research, for biosafety and for research integrity. She initiated the Brigham Research Institute and the Innovation Hub (iHub), a focus for entrepreneurship and innovation. In addition, she was the Founding Director of the Center for Faculty Development and Diversity at the BWH. In addition to her academic responsibilities, she serves on the Board of Directors of Public Responsibility in Medicine and Research (PRIM&R), dedicated to promoting the ethical conduct of biomedical and behavioral research; Management Sciences for Health (MSH), an international organization working in partnership globally to strengthen health care, local capability, and access; and the Edward P Evans Foundation, a foundation supporting biomedical research. Previously she has served as the chair of the Board of Directors of the Association for Accreditation of Human Research Protection Programs (AAHRPP) and as chair of the Secretary's Advisory Committee on Human Research Protections, HHS. She has authored or co-authored over 180 publications and is on the editorial boards of a number of journals including *Current Protocols of Immunology*. Dr. Bierer received a B.S. from Yale University and an M.D. from Harvard Medical School.

Dr. Christoph Gerlinger, is senior director at Bayer Pharma AG in Berlin, Germany. As Bayer's Expert Statistician for Health Technology Assessment he is responsible for method development and for the biostatistical content of re-imbursement applications. Christoph is a member of Bayer HealthCare's data protection panel. His current methodological research focus lies on the development and validation of patient reported outcomes questionnaires and the empirical derivation of minimal clinically important differences and responder definitions. Christoph is a council member and the regulatory chair of the European Federation of Statisticians in the Pharmaceutical Industry (EFSPI). In that role he was an active member of the Clinical Trial Advisory Group "Good Analysis Practice" that advised the European Medicines Agency on their policy on clinical data transparency. Christoph serves also as vice president of the Arbeitsgemeinschaft pharmazeutische Forschung of the German Region of the International Biometric Society. In his spare time he teaches at the Department of Gynecology, Obstetrics, and Reproductive Medicine of the University Medical School of Saarland, Germany, and serves as an Expert Evaluator for the European Commission's Horizon2020 research program. Christoph (co-)authored over 60 papers in peer-reviewed scientific journals. Christoph started his professional career as a statistician in the pharmaceutical industry with Laboratories Fournier SCA in Dijon, France. After 6 years he moved to Schering AG, now Bayer Pharma AG, in Berlin where he held various positions with increasing responsibilities from senior statistician to director statistics. Christoph graduated from Universität Stuttgart, Germany, with an MSc in Mathematics and a minor subject in Business Administration. Before graduating, he spent one year as a graduate exchange student in Statistics at the University of Massachusetts at Amherst, USA. He received his PhD in Public Health from the Technische Universität Berlin, Germany, with a thesis on the statistical evaluation of menstrual bleeding patterns.

Peter Goodhand, is the Executive Director of the Global Alliance for Genomics & Health in Toronto. He is a leader in the global health sector as a senior executive and board member in the health research advancement community. Goodhand played a key role in the creation of the global alliance to accelerate progress in genomic research and medicine and in June 2013, he was appointed Executive Director of the Alliance for its critical development phase. Prior to this role, he was the President and CEO of the Canadian Cancer Society - Canada's largest health charity. Before joining the charitable sector, Goodhand had a 20 year career in the global medical technology industry, including strategic leadership roles with multinational healthcare companies such as American Cyanamid and Johnson & Johnson and as the founding Managing Director of the Health Technology Exchange (HTX). He is currently Chair of the Board of HTX; Chair of the Steering Committee of the Occupational Cancer Research Center; Board Member, MaRS EXCITE (Excellence in Clinical Innovation and Technology Evaluation); and works with the Institute of Corporate Directors to advance governance within the not-for-profit sector. He served as board Chair and President of Canada's Medical Device Industry association (MEDEC), chaired the Government of Canada's Expert working group on the future of medical isotope production, and was a member of the Canadian delegation to the UN summit on non-communicable diseases. Goodhand had a 12-year experience as a patient advocate, caregiver and navigator throughout his family's battle with a rare cancer.

Steven Goodman, M.D., M.H.S., Ph.D., is Associate Dean of Clinical and Translational Research and Professor of Medicine and of Health Research & Policy at Stanford University, directing Stanford's CTSA/Spectrum training programs in medical research methods and serving as chief of the Division of Epidemiology in the department of Health Research and Policy. He is co-director of the Meta-research innovation Center at Stanford (METRICS), a group dedicated to examining and improving the reproducibility and efficiency of scientific research. He has been a senior statistical editor of *Annals of Internal Medicine* since 1987 and from 2004-2013 was Editor of *Clinical Trials: Journal of the Society for Clinical Trials*. He is Vice-chair of the Methodology Committee of the Patient Centered Outcomes Research Institute (PCORI), where he leads their open science and data sharing efforts, and is scientific advisor for the national Blue Cross–Blue Shield Technology Assessment Program. He has served on numerous Institute of Medicine committees, including a committee on sharing data from clinical trials, whose report was released in January, 2015. Most recently, he was appointed to an advisory group to the NIH director on the future of the National Library of Medicine. From 1989-2011, Steve served on the faculties of the Johns Hopkins Schools of Medicine and Public Health. He received an AB from Harvard University, an MD from New York University School of Medicine, and an MHS in Biostatistics and PhD in Epidemiology from Johns Hopkins Bloomberg School of Public Health.

Kathy L Hudson, Ph.D., is the Deputy Director for Science, Outreach, and Policy at the National Institutes of Health (NIH). Dr. Hudson leads the science policy, legislation, and communications efforts of the NIH and serves as a senior advisor to the NIH director. She is responsible for creating major new strategic and scientific initiatives for NIH and was a key architect of the National Center for Advancing Translational Sciences and the NIH BRAIN Initiative. She directs the agency's efforts to advance biomedical science through policy development and innovative projects and partnerships. Dr. Hudson's professional experience includes serving as the Acting Deputy Director of the National Center for Advancing Translational Sciences, NIH; the NIH Chief of Staff; the Assistant Director of the National Human Genome Research Institute, NIH; and the founder and Director of the Genetics and Public Policy Center at John Hopkins University. Also at Hopkins, Dr. Hudson was an Associate Professor in the Berman Institute of Bioethics, Institute of Genetic Medicine, and Department of Pediatrics. Dr. Hudson holds a Ph.D. in Molecular Biology from the University of California at Berkeley, an M.S. in Microbiology from the University of Chicago, and a B.A. in Biology from Carleton College.



Ronald Lee Krall, M.D., is Adjunct Associate Professor of Neurology and Member of the Center for Bioethics and Health Law at the University of Pittsburgh and a collaborator in the Observational Health Data Sciences and Informatics (OHDSI) project. He is a member of the Board of the Foundation of the NIH, of the Scientific Advisory Board of Kala Pharmaceuticals, of the Safety Board of Takeda Pharmaceuticals and consults for a number of healthcare companies. Former Chief

Medical Officer for GlaxoSmithKline (Retired), Dr. Krall worked for four companies (Lorex Pharmaceuticals, Abbott Laboratories, Zeneca/AstraZeneca and GlaxoSmithKline) over 25 years. His areas of expertise include the ethics of human subject experimentation, drug development, regulatory science and the safety of medicines. Dr. Krall was a founding member of the Observational Medical Outcomes Partnership, served on its Executive Committee and led its research subcommittee. Dr. Krall currently makes his home in Steamboat Springs, Colorado, where he and his wife own and operate *Off the Beaten Path*, an independent bookstore, coffeehouse and bakery café (www.steamboatbooks.com) and Dr. Krall serves on the Board of the Yampa Valley Community Foundation and is President of the Timbers Water and Sanitation District.

Barbara Kress, R.N., is Department Head, Trial Disclosure and Data Access, with Merck, Sharp and Dohme. She is responsible for registration and result posting activities for the US and EU as well as the data access process at Merck. She joined Merck 18 years ago as a clinical scientist working on the successful development and submission of several compounds. Prior to joining Merck, Barbara started her career as a critical care and emergency room nurse. She is currently participating in the EFPIA Standards working group, DIA Clinical Trial Disclosure community and the Multi-Regional Clinical Trials (MRCT) Center at Harvard University Return of Results working group.

Rebecca Daniels Kush, Ph.D., is Founder, President and CEO of the Clinical Data Interchange Standards Consortium (CDISC), a non-profit (501c3) standards developing organization (SDO) with a focus on global clinical research standards and their link with healthcare. The CDISC vision is *"Informing patient care and safety through higher quality medical research".* She is also a Director on the Board of the CDISC Europe Foundation, which is a research partner in three consortia of the Innovative Medicines Initiative (IMI). Dr. Kush has over 30 years of experience in the area of clinical research, including positions with academia, the U.S. National Institutes of Health, a global Clinical Research Organization and biopharmaceutical companies in the U.S. and Japan. She earned her doctorate in Physiology and Pharmacology from the University of California San Diego (UCSD) School of Medicine. She leads the Scientific Advisory Committee for the Coalition For Accelerating Standards and Therapies (CFAST), a partnership with the Critical Path Institute in collaboration with the FDA. Dr. Kush currently serves as a member of the IT Workgroup of the U.S. NIH/NCI National Cancer Advisory Board and was appointed in 2011 to represent research on the U.S. Health Information Technology Standards Committee. In 2013, she launched the Essential Standards to Enable Learning (ESTEL) Initiative for the Learning Health Community.

Michael S. Lauer, M.D., has served as Director of the Division of Cardiovascular Sciences at the National Heart, Lung, and Blood Institute since October 14, 2009. Dr. Lauer is a cardiologist and clinical epidemiologist noted for his work on diagnostic testing, clinical manifestations of autonomic nervous system dysfunction, and clinical comparative effectiveness. Dr. Lauer received a BS in biology from the Rensselaer Polytechnic Institute and an MD from Albany Medical College; he also participated in the Program in Clinical Effectiveness at the Harvard School of Public. He received post-graduate training at Massachusetts General Hospital, Boston's Beth Israel Hospital, and the Framingham Heart Study. Prior to coming to NIH, Dr. Lauer was a Professor of Medicine, Epidemiology, and Biostatistics at the Cleveland Clinic Lerner College of Medicine of Case Western Reserve University and a Contributing Editor for JAMA (Journal of the American Medical Association). He is an elected member of the American Society of Clinical Investigation and won the Ancel Keys Award of the American Heart Association in 2008. In 2010 he won the NIH Equal Employment Opportunity (EEO) Award of the Year. In 2010 he won the NIH Equal Employment Opportunity (EEO) Award of the Year. In 2012 he won the prestigious Arthur S. Flemming Award for Exceptional Federal Service in recognition of his efforts to grow a culture of learning and accountability.

Rebecca Li, Ph.D., has over 17 years of experience spanning the entire drug development process with experience in Biotech, Pharma and CRO environments. Dr. Li currently serves as the Executive Director of the Multi-regional Clinical Trial Center at Harvard. The Center was chartered to improve the design, conduct and oversight of multi-regional clinical trials in the developing world and simplifying research through best practices. She is also a Fellow in the Division of Medical Ethics at Harvard Medical School. Prior to joining Harvard, Dr. Li served as the VP of Clinical Research at the New England Research Institutes for 6 years. She also was employed at Wyeth Research as the Associate Director in Translational Clinical Research. She earned her PhD in Chemical and Biomolecular Engineering from Johns Hopkins University.

Nick Lingler, is a Senior Manager of Strategy and Operations for Deloitte's Life Science Consulting Practice. He has over 14 years of industry and consulting expertise and he focuses specifically on development of pharmaceutical, biotech, and medical device products. He has extensive experience in fostering strategic thinking and leading Global and Regional teams through complex business situations to achieve actionable results.

Bernard Lo, M.D., is President of the Greenwall Foundation, whose mission is supporting bioethics research and young researchers in bioethics. He is Professor Emeritus of Medicine and Director Emeritus of the Program in Medical Ethics at the University of California San Francisco (UCSF). A member of the Institute of Medicine (IOM), Dr. Lo served on the IOM Council and chaired the Board on Health Sciences Policy. He chaired IOM committees on clinical trial data sharing, on conflicts of interest in medicine, and on confidentiality in health services research and is currently chairing a committee on data sharing in clinical trials. Dr. Lo serves on the Board of Directors of Association for the Accreditation of Human Research Protection Programs (AAHRPP) and on the Medical Advisory Panel of Blue Cross/Blue Shield. Formerly he was a member of the National Bioethics Advisory Commission, the NIH Recombinant DNA Advisory Committee, and the Ethics Subcommittee of the Centers for Disease Control and Prevention. Dr. Lo and his colleagues have published around 200 peer-reviewed articles on ethical issues concerning decision-making near the end-of-life, oversight of research, the doctor-patient relationship, and conflicts of interest. He is the author of Resolving Ethical Dilemmas: A Guide for Clinicians (5th ed., 2013) and of Ethical Issues in Clinical Research (2010). He continues to care for a panel of primary care internal medicine patients at UCSF.

Pete Lyons, M.B.A., is Principal at Deloitte Consulting LLP and leader of Deloitte's Life Sciences Research & Development practice. Pete focuses on helping Life Sciences R&D organizations address a variety of strategic and operational challenges, with a specific emphasis in Clinical Development. Pete has extensive experience helping Life Sciences companies improve R&D productivity through projects focused on operating model and organization structure design, quality management, process improvement, outsourcing and vendor management, cost reduction, global expansion, and implementation of R&D enabling technologies. Pete has a B.S. in Economics and Information Systems from Boston College and a MBA from the Kellogg School of Management at Northwestern University.

Justin P. McCarthy, J.D., is Senior Vice President, Global Policy & International Public Affairs, at Pfizer Inc., New York. Mr. McCarthy is also a Senior Vice President at Pfizer Inc. and in January 2014 established the Global Policy and International Public Affairs function. In this role, Mr. McCarthy is responsible for defining Pfizer's global policy positions, and driving the advocacy agenda internationally. He leads Pfizer's engagement in international trade and business associations, and serves on the Boards of the Business Council for International Understanding and Acritas and on the harmonization subcommittee of the HHS Secretary's Advisory Committee on Human Research Protections (SACHRP). He also serves as Secretary to the Pfizer Board's Science and Technology Committee. Most recently, Mr. McCarthy was the Chief Counsel for Pfizer's Worldwide Research and Development division. In that role, he coordinated all legal support, advised on regulatory, policy and bioethics matters,

and held responsibility for Pfizer's global Intellectual Property activities. He has extensive experience negotiating novel research collaborations with academia, governments and other biopharmaceutical companies. Mr. McCarthy joined Pfizer in 1993 based at corporate headquarters in New York, where he provided regulatory law support for all Pfizer businesses. In 1998, he relocated to Brussels, where he provided legal support to Pfizer's European Operations. He returned to the US in 2001 to support Pfizer's expanded research and development operations after the merger with Warner-Lambert. Prior to joining Pfizer in 1993, Mr. McCarthy was an associate in the Washington, D.C., law firm of Keller & Heckman, where he focused primarily on food and drug law. He holds a BS in Pharmacy from the University of Rhode Island and a JD from the Catholic University of America.

Nicola Perrin, is Head of Policy in the Strategy Division at the Wellcome Trust, responsible for leading policy development and advocacy work at the Trust. Particular areas of focus include research base funding, uptake of innovation in the NHS, data sharing and the use of patient information in research. She has recently led the development of the Trust's thinking in relation to opening up access to clinical trial data access, and has been involved in establishing the Trust's role running the Independent Review Panel for the clinicalstudydatarequest.com access system. Prior to joining the Trust, Nicola worked at the Nuffield Council on Bioethics as Communications and External Affairs Manager, and before that, she was an exhibition manager at the Science Museum. She has degrees from the University of Oxford, the University of Cambridge, and Imperial College, London.

Dr. Eric Peterson, M.D., M.P.H., is the Fred Cobb Distinguished Professor of Medicine in the Division of Cardiology, a DukeMed Scholar, and the Executive Director of the Duke Clinical Research Institute (DCRI), at Duke University, Durham, NC, USA. Dr Peterson is the Principal Investigator of the National Institute of Health, Lung and Blood Institute (NHLBI) Coordinating Center for its Outcome Research Network, and the Agency for Healthcare Research and Quality (AHRQ) Cardiovascular Center for Education and Research on Therapeutics (CERTs). He is also the Principal Investigator of the Data Coordinating Centers for the Society of Thoracic Surgeons (STS), National Cardiac Surgery Database, the American College of Cardiology National Cardiac Database (ACC-NCDR) and the American Heart Association Get With the Guidelines Database (AHA-GWTG). Dr Peterson participates on multiple national committees, as well as CV guideline and performance measure development groups. Dr Peterson is a member of the American Society for Clinical Investigation (ASCI), the Association of American Physicians (AAP), and the Association of University Cardiologists (AUC). In 2013, Dr Peterson received the American Heart Association Meritorious Achievement Award and was voted one of world's top 400 most influential researchers in biomedicine. This year, he was named by Thomson Reuters as a 'Highly Cited Researcher', ranking him among the top 1% most cited in clinical medicine. He is a Contributing Editor on the *Journal of the American Medical Association* and is a recognised leader in outcomes and quality research with over 700 peer-reviewed publications in the field.

Frank W. Rockhold, Ph.D., is Senior Vice President, Global Clinical Safety and Pharmacovigilance, at GlaxoSmithKline (GSK). He has 40+ years' experience as a leader of global organizations in drug research and development. While he is primarily known for his expertise in biostatistics and clinical trial design, in the last five years he has expanded his reach by running the Cardiovascular Development and Clinical Safety and Pharmacovigilance Departments. Over the span of his forty year career, he has had diverse leadership opportunities and experiences inside and outside the company in clinical trials, data standards, benefit/ risk, clinical research, epidemiology, and most recently pharmacovigilance. For the past twelve years, Frank has been a leader in GSK and the scientific community as a whole in promoting data disclosure and transparency in clinical research. Additionally he has served for 9 years on the board of directors of the non-profit Clinical Data Interchange Standards Consortium, most recently as Chairman and is past president of the Society for Clinical Trials. Frank is also a member of the Patient Centered Outcomes Research Institute (PCORI) Advisory Panel on Clinical Trials. He has over 150 publications and presentations in major scientific journals across a wide variety of topics and has held adjunct faculty appointments at six universities, including his current posts as Adjunct Professor of Biostatistics and Bioinformatics at Duke University School of Medicine and Affiliate Professor of Biostatistics at the Virginia Commonwealth University Medical Center. Frank is a Fellow of both the American Statistical Association and the Society for Clinical Trials and an Accredited Professional Statistician, PStat.

Joseph S. Ross, M.D., M.H.S., is an Assistant Professor in the Section of General Internal Medicine at the Yale University School of Medicine in New Haven, CT. He completed his undergraduate degrees in biological science: neuroscience and psychology at the University of Rochester and his medical degree at the Albert Einstein College of Medicine, Bronx, NY. After completing his post-graduate training in primary care internal medicine at Montefiore Medical Center in Bronx, NY, Dr. Ross was a fellow in the Robert Wood Johnson Foundation Clinical Scholars program at Yale University, earning a Master's degree in health sciences research. Using health services research methods, Dr. Ross's research focuses on examining factors which affect the use or delivery of recommended ambulatory care services for older adults and other vulnerable populations, evaluating the impact of state and federal policies on the delivery of appropriate and higher quality care, and issues related to pharmaceutical and medical device evidence development and post-market surveillance. In addition, he collaborates with a multi-disciplinary team of investigators under contract for the Centers for Medicare and Medicaid Services to develop statistical models that are used to measure and publicly report hospital and ambulatory care clinical outcomes using administrative data.

Arthur Toga, Ph.D., is a professor in the Keck School of Medicine at the University of Southern California, Laboratory of Neuro Imaging Arthur's research focuses on neuroimaging, informatics, mapping brain structure and function, and brain atlasing. He developed multimodal imaging and data aggregation strategies and applied them in a variety of neurological diseases and psychiatric disorders. He holds a PhD in Neuroscience.

Appendix 4: Worksheet for Breakout Groups

Promoting Clinical Trial Data Transparency Conference - Worksheet

March 30-31, 2015, Harvard Faculty Club

Topic 1: Envision an ideal and sustainable data sharing platform

Checkmark	**Platform Elements**		Checkmark
	Academic	Commercial	
	Disease specific	Not disease specific	
	Portal	Repository	
	Federated	Centralized	
	Review Board	Open Access	
	Downloadable	Non-downloadable	
	Sharing with general public	Sharing with approved researchers only	

Rationale for chosen essential elements:

Element	**Rationale**

Topic 2: Discuss incentives to address barriers and gaps to data sharing

Barriers	Essential to address	Nice to address	Not essential
Discoverability/awareness of platform			
Access to relevant existing datasets			
Harmonization of standards			
Concerns about participants' consent for data sharing			
Privacy / concern about personally identifiable information (PII)			
Review of research requests			
Requirements for accessing data: ownership terms of research results; credentials required for data requestors			
Data can only be analysed on data owner's / repository server			
Concerns about providing competitive advantage to others			
Cost to prepare data and to post data to existing platforms			
Lack of consistent method or policy within organization			

Select the three most important barriers or gaps and propose incentives to address them.

Barriers / Gaps	Incentives

Topic 3: Define characteristics of the platform

Key questions:

1. Should there be an independent, central or a federated body to manage the process?

2. What does an independent review process need to look like?

3. What does the business model and governance arrangements need to look like?

4. What stakeholders should be involved?

5. How will it be funded?
   How will costs of data sharing be equitably distributed across both data generators and users?

6. How will stakeholders assess compliance / adherence to standards, and who will review and update standards moving forwards?

7. Should a platform be global or regional? (short / medium / long-term aspirations?)

Report back responses to up to three of the questions above, including a rationale.

Platform Characteristic	**Rationale**

Topic 4: Describe how to get data into the common platform and access data from the common platform

Key questions:

1. What does it mean for data to be available?
2. Should data sets allow for harmonization and for merging with other data sets?
3. Should data from all regions be included?
4. What can data be used for?
5. How are data accessed?
6. What about de-identification? How can one protect PII?

Report back up to three selected principles, including rationale.

Data access / availability principles	**Rationale**