Immune Tolerance Network BRDI Data and Information Challenge Entry

May 15, 2014

Entrant Contact Information:
Adam Asare, PhD
Executive Director, Bioinformatics
Immune Tolerance Network
7500 Old Georgetown Road, Suite 800
Bethesda, MD 20814
240-235-6149
aasare@immunetolerance.org
www.itntrialshare.org
www.immunetolerance.org

ITN TrialShare: Enabling True Clinical Trial Transparency

Introduction

The goal of clinical research is the development of new medicines, products and techniques that prevent and cure disease and improve human health. Clinical trials are a critical step in this development process, being the stage where both effectiveness and safety are established. In the past several years, a series of scandals over the safety of approved drugs (Merck's Vioxx is an example) has led to a growing chorus of researchers and public agencies seeking more complete and transparent disclosure of clinical trial results. The European parliament, for instance, recently voted overwhelmingly to require summary results, positive or negative, to be disclosed within one year of completion. In fact, a number of pharmaceutical companies, academic and governmental organizations have already committed to an even greater level of transparency, by allowing researchers to access participant-level data for their trials. However, the race to full disclosure has revealed technical and operational challenges in the reporting and sharing of highly varied and often complex clinical and translational research data that must be addressed for the true benefits of transparency to be realized.

The Immune Tolerance Network (ITN) is an international clinical research consortium sponsored by NIAID, part of the National Institutes of Health, whose mission is to accelerate the clinical development of immune tolerance therapies. The ITN looks beyond the traditional clinical trial endpoints of safety and efficacy, also investigating biological and immunologic mechanisms of disease and drug action within the traditional clinical trial model; the identification and validation of new biomarkers and surrogate endpoints are a particular emphasis. As such, ITN clinical trial datasets are highly complex, including traditional clinical measures and endpoints, as well as laboratory assay data from high-throughput profiling platforms such as flow cytometry and gene expression profiling, all linked to individual participants and visits. Thus, ITN study results represent a fitting proving ground for the development of robust methods for sharing modern clinical and translational research data.

ITN TrialShare

Developed over the past several years, the ITN's multipurpose clinical trial data platform, ITN TrialShare (www.itntrialshare.org) marks a significant advance in data sharing and transparency. Based on the open source LabKey Server platform (www.labkey.org), the system permits access to participant-level clinical data linked to clinical laboratory and other assay data. It allows investigators to re-run and validate other investigators' statistical analyses and interactively perform their own exploratory analyses within the system, while enabling the publication of fully interactive figures for manuscript submission and review.

The use of TrialShare within the ITN is supported by system workflows and operating procedures that have resulted in gains in operational efficiency, as well as data transparency. It has rapidly become an analysis prototyping tool for collaborative data analysis at the close of a clinical study, as well as means of enhancing manuscript preparation, review and publication(Figure 1). Manuscript figures can be presented in an interactive format linked to the

analysis data and code used in their production. This provides journal editors the ability to perform more detailed examination of analysis data sets, supplemental data, analysis code and documentation, and to verify and re-run analyses during the publication review process. The ITN regards this set of features as a major step forward in transparency, given the importance of high quality peer review in the scientific method. These interactive figures can then also be made available to the general public upon publication with ITN TrialShare links embedded directly in the manuscript (Figure 2). Current ITN policy dictates that published data and analysis methodologies are made publicly available following study closeout and the publication of the manuscript describing the primary outcome. The full study database is made available through ITN TrialShare, where it can be explored using the native data analysis tools or exported for analysis with other tools. Eighteen months after last patient last visit, the ITN also makes public the remaining clinical specimen inventory in the study's biorepository, which researchers may use as the basis for sample requests for follow-up studies. The types of data and information provided to the public are shown in Figure 3.

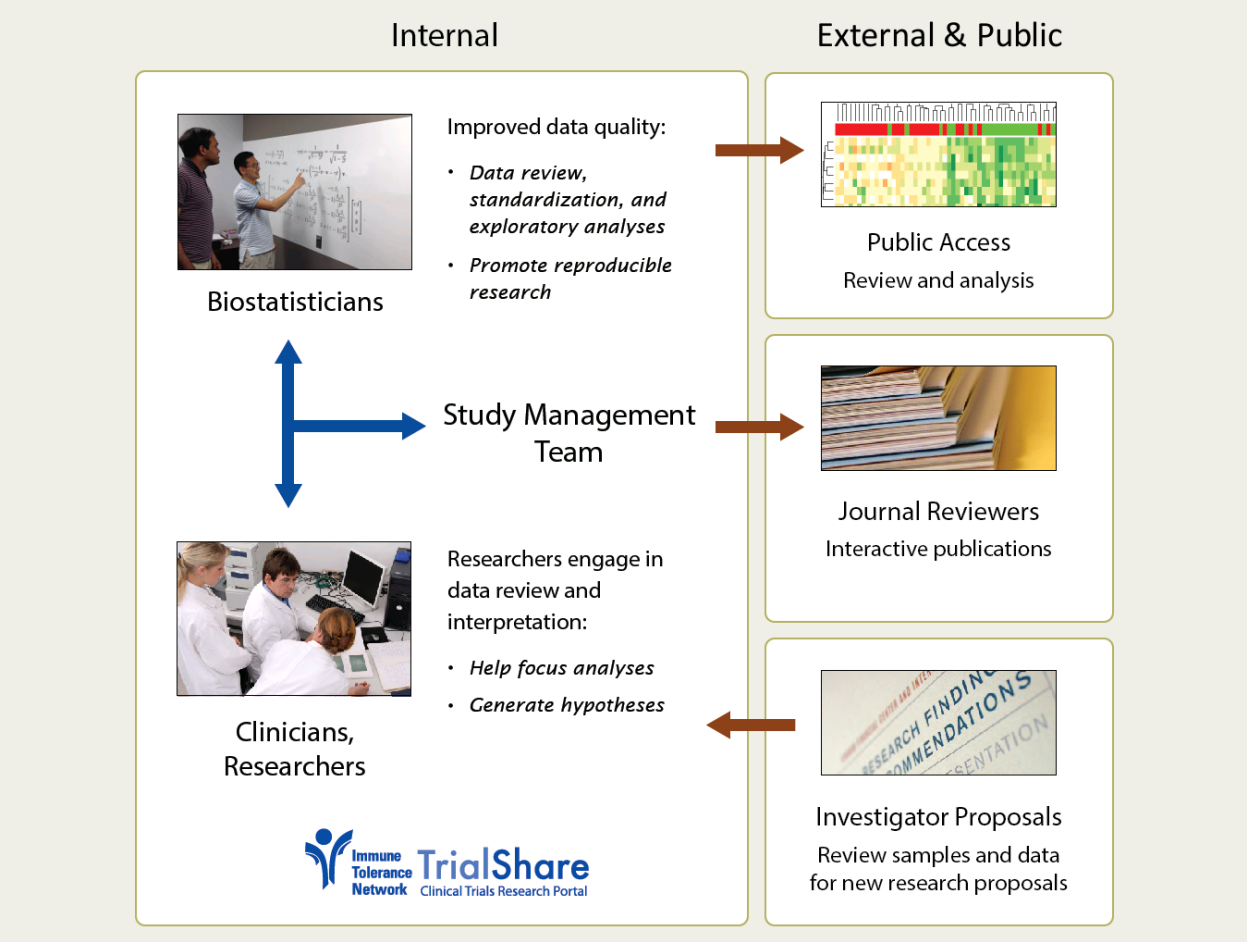

Figure 1: ITN TrialShare's externally facing components address three primary functions: open access data sharing, journal editor review, and specimen repository access to support new research proposals.

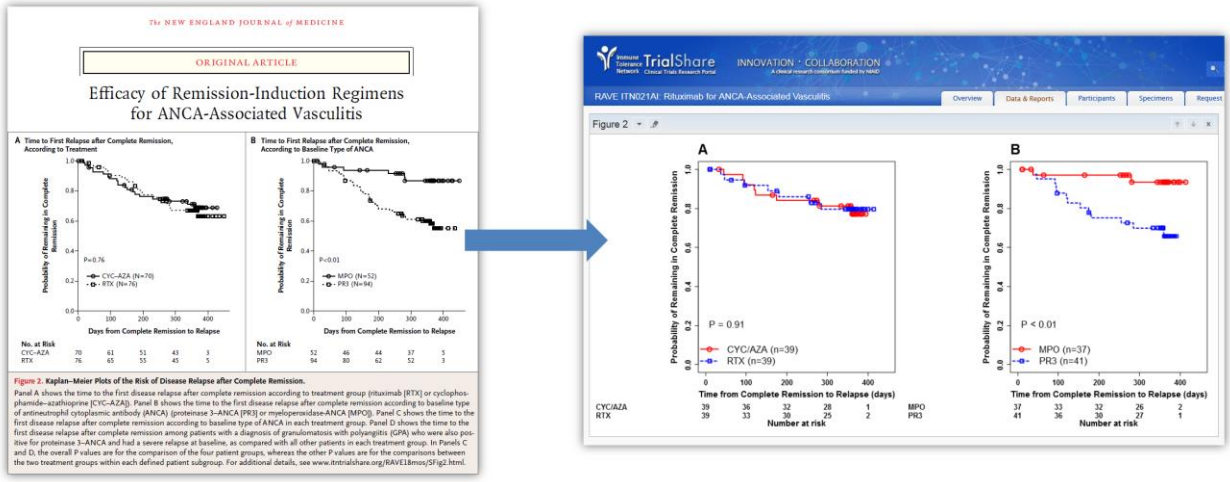

Figure 2: ITN TrialShare links were included in the RAVE ANCA Associated Vasculitis manuscript published in the New England Journal of Medicine, August 1, 2013. Access to published data allows end-users to re-run analyses or view alternative analyses.

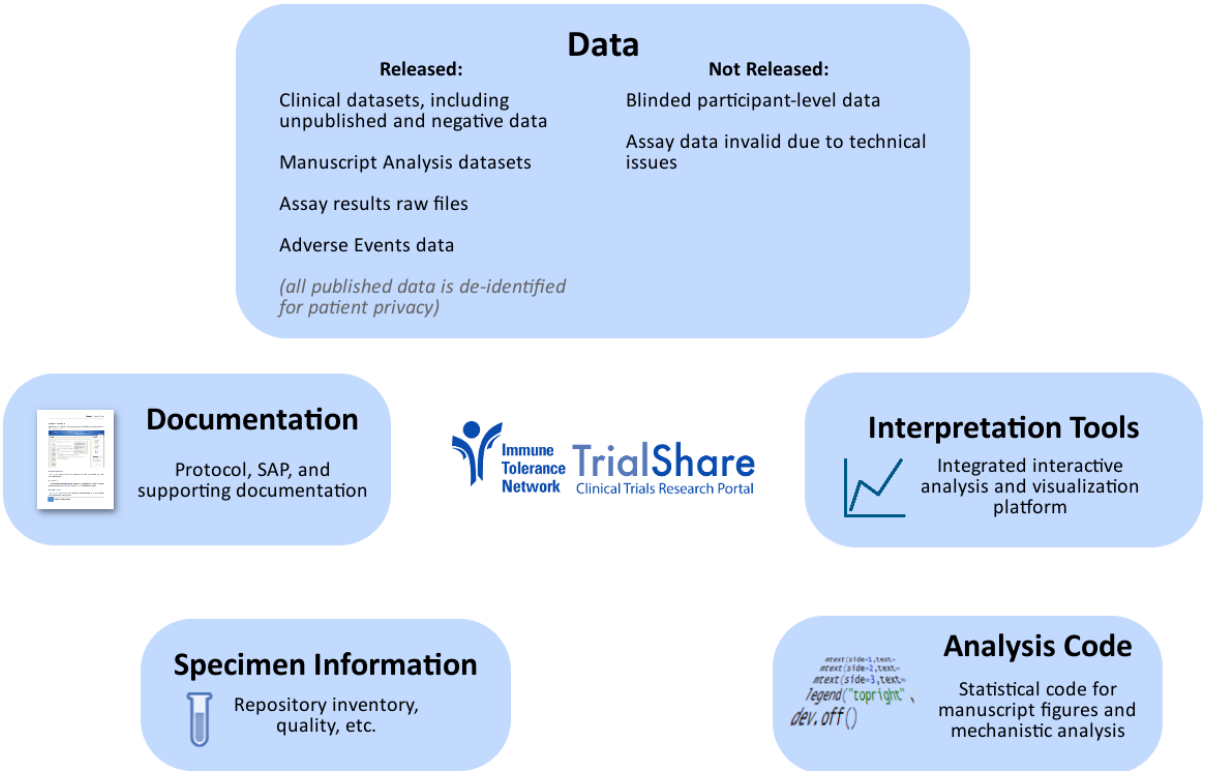

Figure 3: Types of data, documentation and information supporting clinical trials research data released on ITN TrialShare.

At present, the ITN implementation of TrialShare provides comprehensive participant-level data and information for 21 individual clinical trials with access to 607 analysis, clinical and research assay datasets, along with specialty assay data that includes base level raw files such as 83 gene expression CEL files, 2608 flow cytometry FCS files, and 142 sequencing T-cell repertoire files for reanalysis and bioinformatics analysis methodology development. In addition, analysis datasets and code supporting findings are available from seven ITN publications in the New England Journal of Medicine (NEJM), Journal of Clinical Investigation (JCI) and the Journal of the American Medical Association (JAMA), among others (REF). Importantly, the public release of ITN data is not restricted only to studies with positive results - closed studies with negative results are also made available to allow for repurposing of this data for other types of research. This is a dramatic departure from the current standard of only providing data and analyses for published findings.

Using ITN TrialShare

Each clinical study hosted by ITN TrialShare has multiple datasets, reports and visualizations that are conveniently displayed showing the date of posting and author/creator (Figure 4). Analysis figures associated with published manuscripts are provided in an interactive analysis console using the open source R-programming language, displaying the data and analysis code (Figure 5). In addition, sub-group analysis can be performed based on cell surface assay (Figure 6), gene expression, or other specialty assay results (Figure 7) in order to stratify participants or to design follow up experiments using banked samples within the ITN bio-repository. Given its pedigree, ITN TrialShare has a broad focus that incorporates a significant emphasis on integrated analyses of clinical and translational research data. Such data can provide powerful biological insight when analyzed in the context of clinical parameters and

endpoints within the same trial, or when used to compare data from studies of the same drug or similar patient populations across different trials.

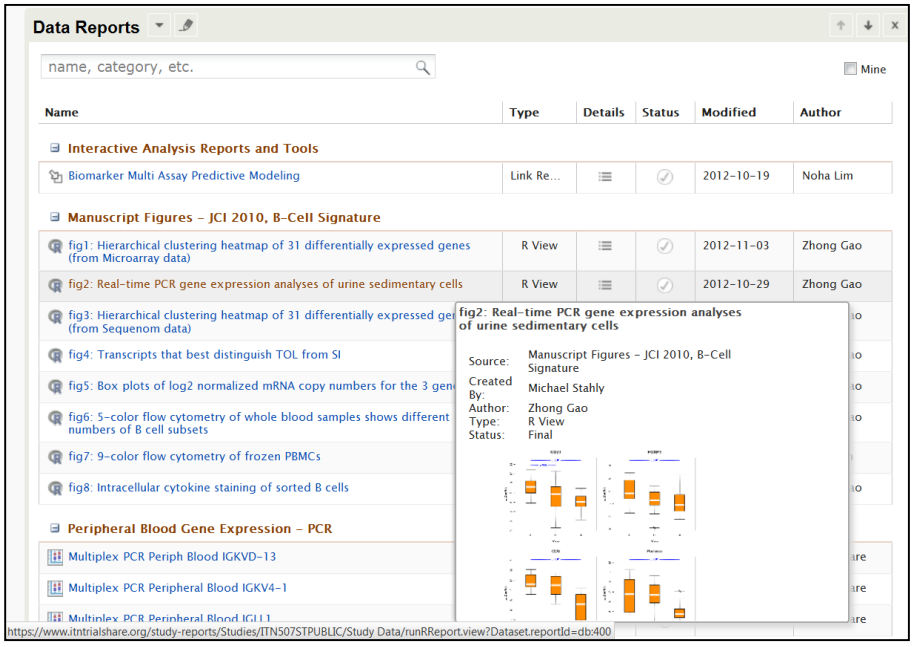

Figure 4: Data sets, interactive analysis plots from published manuscripts, and analysis reports are readily viewable through the “Data and Reports” tab within ITN TrialShare.

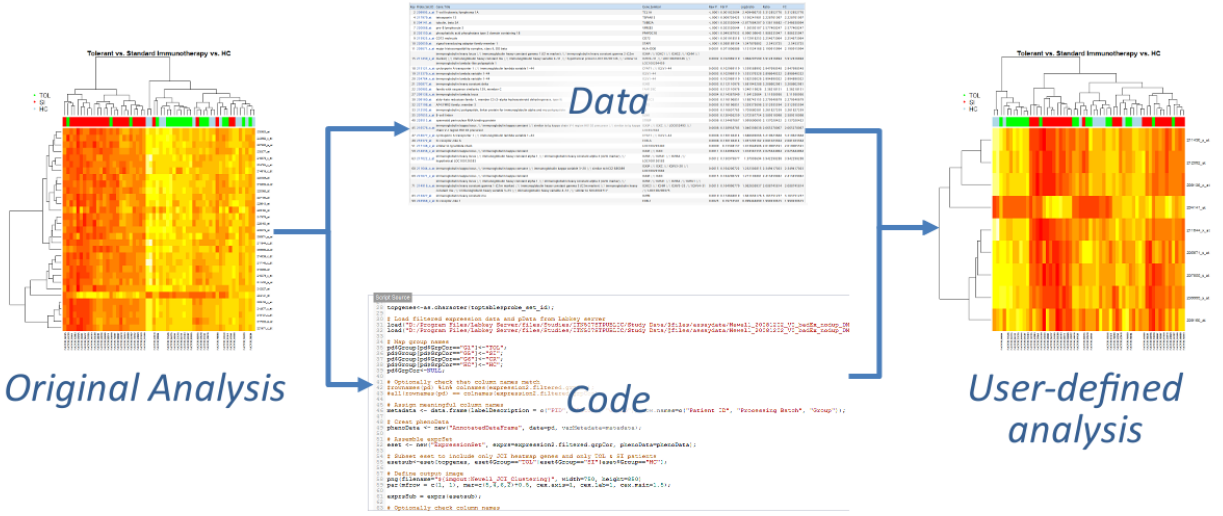

Figure 5: Manuscript figure and data associated within the statistical programming language R interactive console

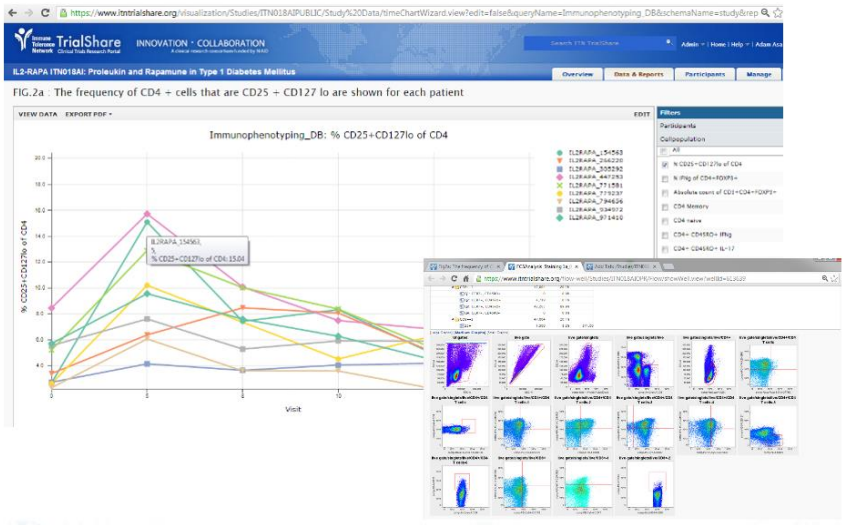

Figure 6: Cell surface immunophenotyping data from the published manuscript is available along with access with a click of a data point to the "gating strategy" used by the domain expert in sub-setting the populations.

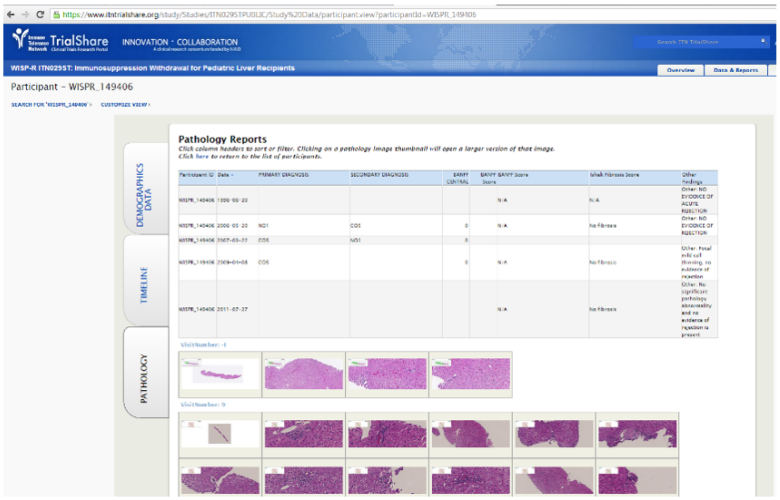

Figure 7: Participant level data access is shown in the "Participants" tab allowing public users to subset data using protocol specified cohorts or user-defined sub-groups for specialty assay data such as liver pathology readouts.

Privacy Considerations

In order to protect study participant privacy, no HIPAA-defined Personally Identifying Information (PII) is stored anywhere on the ITN TrialShare system (including ongoing studies only available internally). In addition, all data made publicly available is carefully de-identified by removing or

masking all HIPAA-defined Protected Health Information (PHI), free text fields, dates, participant study IDs, and clinical site information. All remaining data is then reviewed, and additional fields may be removed or masked if it is determined that they may reasonably be expected to significantly increase the risk of re-identification. For example, if the combination of gender, age and occupation is very rare, one or more of those fields may be removed of grouped into a larger category.

Current Usage

Although only publicly accessible since January 2013, ITN TrialShare already has over 600 registered public users. The system logged upwards of 2,700 page hits and 100 public user accounts created in the first month alone after release of the RAVE NEJM manuscript with embedded ITN TrialShare links in August 2013(See Figure 8). A total of 159 datasets have been downloaded to date across the various trials and therapeutic areas.

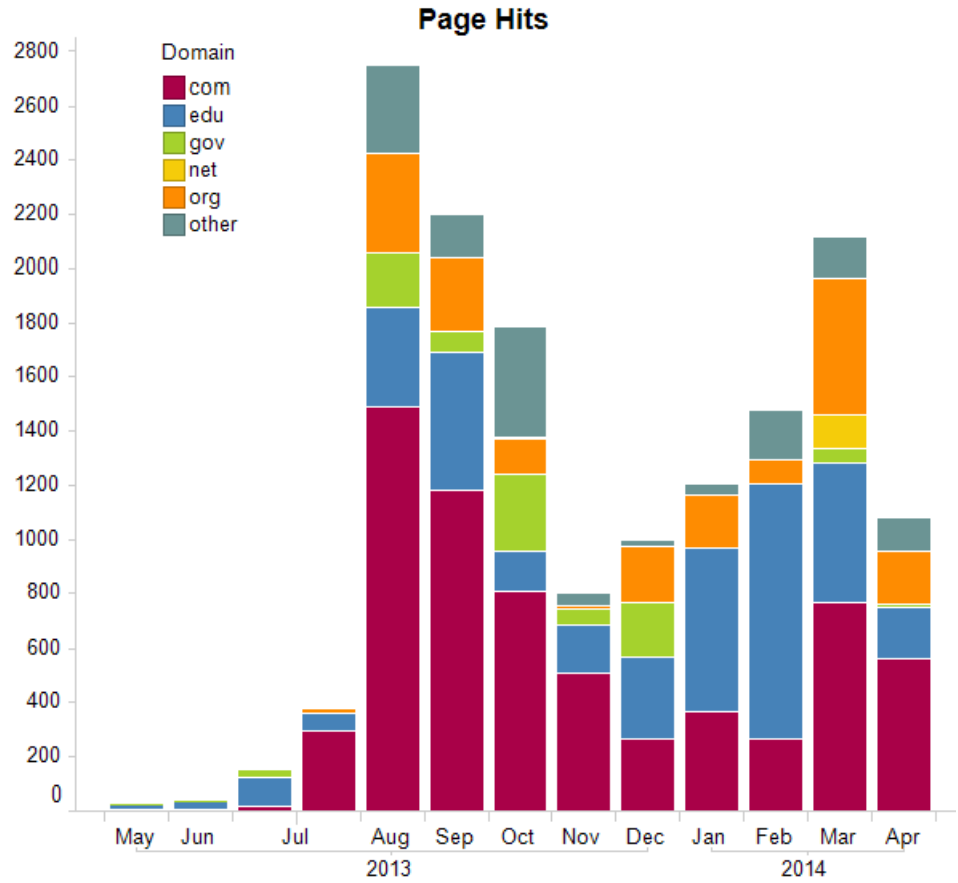

Figure 8: Public access to ITN TrialShare. Prior to the August 1st RAVE NEJM publication that included links to ITN TrialShare, site traffic was nominal. With press pre-release and the publication, public access numbers rose close to 3K per month. Sub-bars show distribution by email domain suffix.

Open source framework and model

ITN TrialShare was developed using the open source LabKey Server platform, which provides a flexible framework for rapidly developing biomedical research portals. Using open source allows investments in the core platform to be leveraged by other research organizations and is a particularly attractive option for publicly funded academic research, as it promotes further feature development by the community and allows best practices for data management to be incorporated within the open source framework. ITN TrialShare has expanded upon the LabKey Server platform with custom design code and in-house developed web-modules.

Models for Data Sharing

A spectrum of data sharing models has begun to emerge with varying levels of restrictiveness, ranging from complete public open access to participant level data, to a “black box” request system in which approved requests for a research analysis are performed by the data holder and only the results are provided back to the requestor. Requirements placed on the data requestor range from open access, as with ITN TrialShare, to requiring submission of an extensive proposal including evidence of sufficient qualifications. The review process for granting requests can range from minimal administrative review to a stringent review by the data provider and/or an independent learned intermediary. Finally, how data is accessed can range from freely downloadable to only available within a restricted system for a limited time (Figure 9).

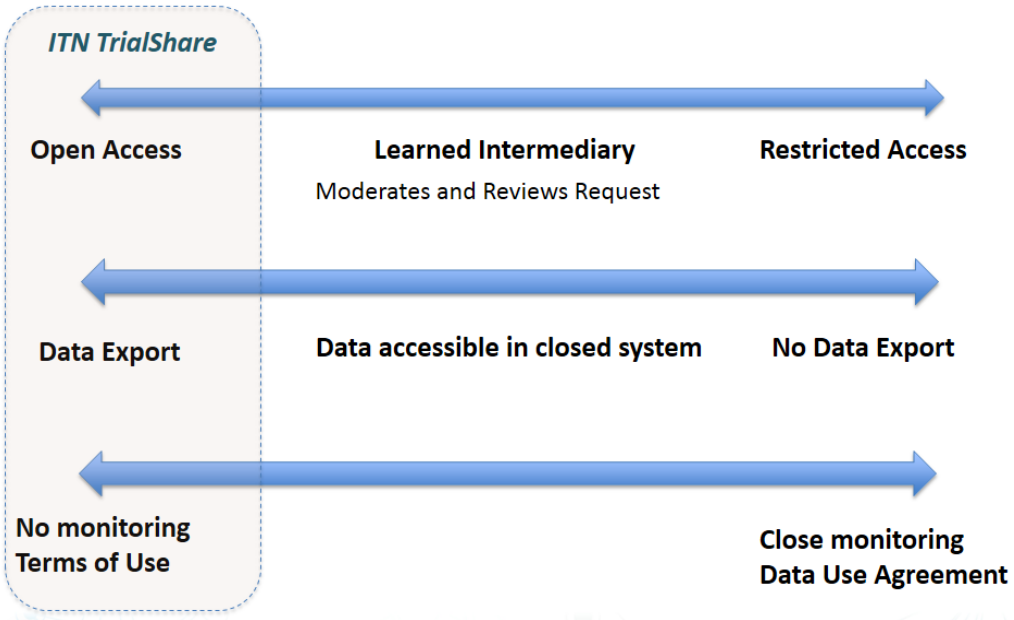

Figure 9: Continuum of data sharing models, functions allowed regarding data, and the level of monitoring

The ITN TrialShare model is one of the least restrictive of data sharing models designed to allow clinical trialists to achieve what has long been one of the tenets of science – the open communication and sharing of data and methods to permit independent validation of the results and exploration of the data. The system allows unlimited open access to downloadable de-identified participant-level data to anyone supplying a valid email address. Registered public users click a check box agreeing to the Terms of Use that require no attempt be made to re-identify participants or otherwise misuse the data in violation of HIPAA regulations or other laws.

Benefits and Metrics of Successful Adoption:

1. Harmonization of workflows for data delivery to biostatisticians, internal sample management staff, external researchers, and the NIAID sponsored Immunology Database and Analysis Portal (ImmPort) repository means that all these groups now see and use the ***same*** data.
2. Higher quality data and faster database lock have resulted from internal validation routines developed by the ITN as data is loaded into TrialShare and refreshed during the course of the study.
3. Facilitation of better-informed decisions during the course of a study regarding experimental design and hypothesis generation.
4. Faster development cycle: by building on an open source platform that has benefited from over $15M in public investment since 2005, the ITN was able to bring TrialShare online for scientific use within a year from the start of its development.
5. Using the open source approach has allowed investments in the platform to benefit other publicly funded research organizations that are adopting the ITN TrialShare approach.

New paradigm for research publishing

At the heart of the philosophy behind ITN TrialShare is a commitment to make participant-level data available to the wider research community, for the purpose of peer review, validation, additional hypothesis generation, etc. The technical challenges inherent in making such data available, especially in the context of associated modern high throughput translational data, are numerous, but not insurmountable.

ITN TrialShare has taken a multipronged approach to the problem, by recognizing that different roles within the research continuum have differing needs with respect to clinical data. Thus, it provides a flexible framework whereby investigators, publishers, readers, reviewers and data analysts may all interact with the data in a manner that meets their individual needs, through well-defined tools embedded within the system. Those seeking the ability to perform more advanced analyses are not forgotten, as they have the ability to download datasets that they define. Finally, TrialShare is built upon an open source platform to maximize the capacity for its application in other clinical trial settings -- by increasing the exposure of the data and code, reliability and extensibility of the work also increases.

References:

Casale TB. Omalizumab pretreatment decreases acute reactions after rush immunotherapy for ragweed-induced seasonal allergic rhinitis J Allergy Clin Immunol. 2006 Jan;117(1):134-40.

Feng S. et. A. Complete immunosuppression withdrawal and subsequent allograft function among pediatric recipients of parental living donor liver transplants. JAMA. 2012 Jan 18;307(3):283-93.

ITN TrialShare: https://www.itntrialshare.org/

Long SA. Rapamycin/IL-2 combination therapy in patients with type 1 diabetes augments Tregs yet transiently impairs β-cell function. Diabetes. 2012 Sep;61(9):2340-8.

Muraro PA. T cell repertoire following autologous stem cell transplantation for multiple sclerosis J Clin Invest. 2014 Mar 3;124(3):1168-72.

Nelson EK et al. An open source platform for scientific data integration, analysis and collaboration. *BMC Bioinformatics* 2011, **12**:71.

Newel KA et al. Identification of a B cell signature associated with renal transplant tolerance in humans. J Clin Invest. 2010 Jun;120(6):1836-47..

Rabesandratana T. European parliament approves bill to increase clinical trial transparency. ScienceInsider http://news.sciencemag.org/europe/2014/04/european-parliament-approves-bill-increase-clinical-trial-transparency 2014 April 3.

Specks, U et al. Efficacy of remission-induction regimens for ANCA-associated vasculitis. N Engl J Med. 2013 Aug 1;369(5):417-27.

Stone J. et al. Rituximab versus cyclophosphamide for ANCA-associated vasculitis N Engl J Med. 2010 Jul 15;363(3):221-32.

Zarin, DA**.** Participant-level data and the new frontier in trial transparency. N Engl J Med. 2013 Aug 1;369(5):468-9.