



# Translational research in the era of heightened patient privacy concerns

Mark A. Collins<sup>1</sup>, Axel Schumacher<sup>2</sup>, & Tamas Rujan<sup>3</sup>

<sup>1</sup>Genedata Inc, Lexington, USA | <sup>2</sup>Genedata GmbH, Munich, Germany | <sup>3</sup>Genedata AG, Basel, Switzerland

*Keywords: Data security, compliance, translational research, precision medicine, patient privacy, protected health information.*

## Introduction

Patient data collected during clinical trials represents a huge resource for translational research. Spanning the divide between highly regulated clinical trials and less regulated discovery research, this data ranges from data about the patient, such as medical history, demographics, and medications, to basic lab tests, biomarker assessments and sophisticated multi-omics analyses (e.g. genomics, proteomics, metabolomics). This data can yield enormously valuable scientific insights:

- **Patient stratification for new trials or new cohorts of existing trials;**
- **Validation and discovery of new biomarkers of disease, drug response or drug toxicity;**
- **Drug repurposing.**

Often this rich data remains largely untapped, as harmonizing this complex, disparate, globally-distributed data and making it actionable while addressing patient privacy and regulatory concerns presents significant challenges. Merging data obtained under the strict regulatory guidelines that govern clinical trials (e.g. Good Clinical Practice: GCP) with data generated in the research environment, which is often much less regulated, raises concerns for pharmaceutical companies regarding protection of patient privacy. These

concerns are further amplified by a constantly changing global regulatory landscape with a myriad of country-specific rules, which leaves organizations knowing that action must be taken, but unsure of the form those actions should take.

In this paper we review the regulatory landscape and its impact on translational research, and define the key attributes of an “ideal” system to overcome the challenges of conducting translational research while respecting patient privacy.

## The regulatory landscape

Balancing legitimate privacy concerns with the need to progress research presents real challenges both within the organization and with external parties, which include patients, health regulatory authorities (HRAs), local internal review boards (IRBs), and drug regulatory authorities such as the European Medicines Agency (EMA) and US Food and Drug Administration (FDA). Recent high-profile rulings by the European Union such as the replacement of US Safe Harbor with Privacy Shield<sup>1,2</sup> and a new EU-wide Data Protection Regulation<sup>3</sup> also raise uncertainties for global organizations that rely on data sharing between sites in the EU and the US. Even if any changes have no direct impact on ongoing activities, time must still be spent on assessing the risk, which may slow critical projects. When drug

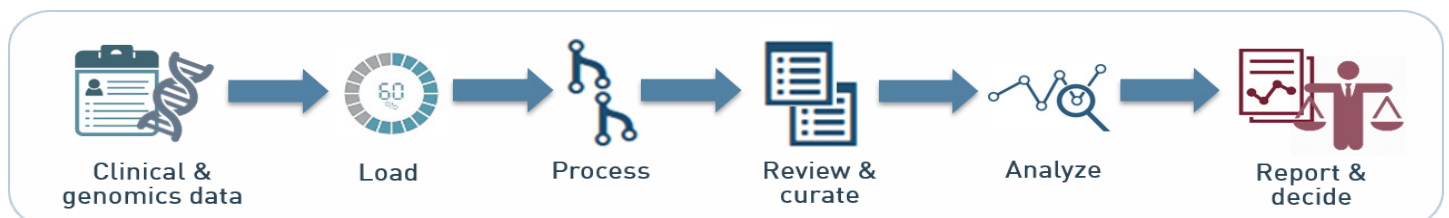


Fig. 1: The complex, interdisciplinary translational research process.

approval processes for similar indications from rival pharmaceutical companies are measured in weeks, this can have significant financial consequences<sup>4</sup>. The key regulatory challenges can be grouped into:

- **Operational**—addressing patient privacy concerns throughout the data lifecycle, from initial trial to any future biomedical research (FBR);
- **Organizational**—building trust amongst various stakeholders in clinical and research environments that patient privacy will be respected;
- **Educational**—keeping up to date with a dynamic global regulatory landscape and its implications.

## The operational challenge

The translational research process (Fig. 1) is complex, multi-step, and interdisciplinary, with a wide variety of data generators, data types, volumes, types and formats, plus a range of data consumers (Fig. 2). This complexity presents a number of operational challenges that have the potential to impact patient security and privacy, as follows:

**Data provenance:** given the complexity of the clinical and research data ecosystem, it is often difficult or even impossible to find out where data has come from—so called “data provenance”. Is the data for example from an academic collaborator, a study site, is it processed data or is it from outside the US and consequently should be handled differently?

**Wide variety of data types and consumers:** this results in multiple handovers of data, e.g. patient phenotype data may be moved from the CRO into a data warehouse, curated, de-identified and then linked with patient genotype data. Further genomic and statistical analyses may then be performed and results written to another data warehouse or used for other research projects. Tracking and auditing this ever more complex data “chain of custody” is a huge challenge which if not solved poses a significant audit risk.

**Regulation of electronic records:** data may be considered to be electronic records, which would require organizations to adhere to US FDA 21 CFR Part 11 regulations regarding changes to data, reasons for change, and audit trails.

**Data may contain PHI (protected health information):** examples include name, age and gender. While most clinical data is de-identified,

lapses and exceptions to this process occur and regulations such as the US Health Insurance Portability and Accountability Act: HIPAA<sup>5,6</sup> permit PHI to be viewed in certain circumstances. In addition, what constitutes PHI changes; for example, the new EU Data Protection Regulation<sup>3</sup> now considers genomic data to be PHI.

**Usage of data:** can the data be worked with at all and by whom? It is critical to know the consent status of the patient and what permissions, if any, the individual gave for their data and specimens. Furthermore, not all stakeholders should be permitted to work with all data, and those who are must be tracked.

Solving this operational challenge is burdensome to organizations, but is key to resolving the related organizational and educational challenges.

## The organizational challenge

The process of clinical trials is subject to a high degree of regulation (GCP and GLP: Good Laboratory Practice) that encompasses trial sponsors, study participants, contract research organizations (CROs), central labs and study sites. In addition the data acquired during the trial is also subject to regulations under HIPAA, Safe Harbor (or its replacement, Privacy Shield) and Binding Corporate Rules, while informed consent processes govern the permitted use of patient specimens and data. Trial sponsors (pharmaceutical companies) invest huge sums of money and time to ensure compliance with all these regulations and to validate the systems and processes that support them.

Leveraging rich, highly curated and well organized clinical trial data is the goal of translational research. However it is often extremely difficult to access data from the clinical side, link it with additional in-house and third party research data, and perform analyses. In particular:

**Clinical researchers** worry that too many scientists in the less regulated research environment may not respect patient privacy due to system limitations or lack of awareness of regulations and risks.

**Translational researchers**, on the other hand, often do not appreciate the risks of non-compliance, so downplay the need to protect patient privacy, lack the systems to manage it anyway and do not want any barriers to achieving their research goals.

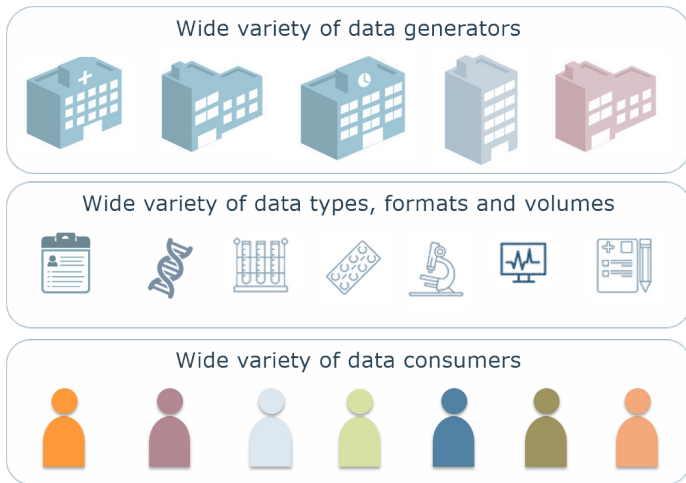


Fig. 2: Distributed data generators and consumers in translational research.

**Data curators** worry about data quality and data applicability. Trial data is collected with the goal of achieving a drug approval, not for answering more open-ended research questions.

Overall these challenges lead to a lack of trust between clinical and research scientists that negatively impacts the translational research process.

## The educational challenge

This challenge probably gets least attention in translational research but plays into the other two in the following ways:

**Lack of knowledge:** often researchers do not know what, if any, regulations they need to adhere to. The drive to meet research goals, poor communication between the discovery, clinical operations and regulatory teams, and the lack of training of researchers often result in regulations simply being ignored. Unfortunately lack of knowledge is not an excuse accepted by regulatory authorities.

**Myriad country-specific regulations:** The global regulatory landscape comprises a staggering variety of geographic, country and local regulations, often further confounded by site-specific IRB rules. This can be extremely frustrating for researchers wishing to share and use the widest variety of data.

**Dynamic regulatory landscape:** Regulations change all the time. Often those who have the responsibility to track changes are in clinical and regulatory operations, concerned with the impact of regulations on trials, not research.

## The ideal infrastructure for secure, compliant translational research

Realizing an effective and efficient translational research process that delivers on scientific goals while protecting patient privacy requires organizations to overcome the often overwhelming operational, organizational and educational challenges detailed above. The operational challenge is considered the most pressing to solve; anecdotally, lack of suitable infrastructure is seen as the root cause. Delivering the infrastructure capable of overcoming the operational challenges is dependent on that infrastructure having the following core attributes:

**Secure, controlled data access:** the infrastructure must have robust role-based access controls (RBACs) that limit what data is seen by whom and when, which is key to compliance with GxP and 21 CFR Part 11 regulations. RBACs need to be flexible and study-centric so that various stakeholders, from data managers through bioinformaticians to heads of translational research, can work while adhering to patient privacy rules. Furthermore, RBAC permissions must extend to all aspects of the data system including file systems, databases and High Performance Computer (HPC) clusters, whether internal, external or public domain.

**Manage data provenance:** the infrastructure must effectively manage the provenance of data so that researchers understand the origin of data and have tools to seamlessly hand data over to others in a controlled manner and with appropriate safeguards. Also important for the management of data provenance is a clear understanding of any permissions related to use of the data, including details of the consent given by the patient.

**Comprehensive chain of custody (CoC):** since data involved in translational research can be considered an “electronic record” for the purposes of GxP compliance, the infrastructure must ensure a comprehensive CoC of data from raw results (such as BAM files and laboratory results) through derived data to analyses and summary reports. Comprehensive CoC is also key to data quality and should extend to the methods (scripts, algorithms and statistics) used to process raw data to results and generate scientifically meaningful conclusions. The CoC must also document in an audit trail any

changes to the data, by whom, when, and for what reason. Furthermore all the CoC records must be secure, tamper-proof and available for inspection.

While these attributes solve the operational part of the regulatory challenge, taken together they form an holistic overall approach that also solves the organizational challenge. By building trust among clinical and research teams while providing capabilities that “bake in” domain knowledge, the ideal infrastructure also ensures that researchers are able to meet the educational challenge.

### How Genedata Profiler™ helps

With nearly two decades of scientific, computational and analytical experience, Genedata is recognized as the trusted partner of pharmaceutical companies, research organizations and research consortia worldwide. Genedata works closely with customers to build collaborative, scalable, and powerful software platforms which optimize the key business processes of life science research and development to empower scientific discovery.

Delivering on the promise of precision medicine through effective translational research is “top of mind” for pharmaceutical organizations. However, as we have discussed, balancing the drive for better therapies with research that increasingly leverages patient data, yet must address privacy concerns, poses significant challenges.

To address those challenges we have proposed the attributes of an “ideal” infrastructure, offering secure controlled data access, management of data provenance, and a comprehensive chain of

custody. Designed to embody these core attributes and developed in collaboration with leading pharmaceutical companies, Genedata Profiler is a new translational research software platform designed to effectively process, manage, and analyze omic and phenotypic data to the highest standards of data quality and regulatory compliance.

Organizations that adopt the Genedata Profiler platform benefit from an effective and efficient translational research process that delivers on scientific goals while protecting patient privacy.

### References

1. Court of Justice of the European Union. *Maximillian Schrems v Data Protection Commissioner (Case C-362/14)*, Oct 6, 2014.
2. European Commission, *Restoring trust in transatlantic data flows through strong safeguards: European Commission presents EU-U.S. Privacy Shield*, Feb 29, 2016.
3. European Commission, *Reform of EU data protection rules*, Mar 23, 2016.
4. John Carroll, *Bristol-Myers' pioneering PD-1 drug Opdivo OK'd by FDA for melanoma*, Dec 22, 2014
5. U.S. Department of Health & Human Services, *The HIPAA Privacy Rule*
6. Dept. Health and Human Services USA *Protecting Personal Health Information in Research: Understanding the HIPAA Privacy Rule*, NIH Publication Number 03-5388
7. Guidance for Industry Part 11, *Electronic Records; Electronic Signatures — Scope and Application*, U.S. Dpt. of Health and Human Services Food and Drug Administration, Aug, 2003.

Please cite this article as:

Collins, M.A., Schumacher, A., & Rujan, T. *Translational research in the era of heightened patient privacy concerns* (2016), doi: 10.13140/RG.2.1.3550.5045



Genedata Profiler™ is part of the Genedata portfolio of advanced software solutions that serve the evolving needs of drug discovery, industrial biotechnology, and other life sciences.

Basel | Boston | Munich | San Francisco | Tokyo  
[www.genedata.com/profiler](http://www.genedata.com/profiler) | [profiler@genedata.com](mailto:profiler@genedata.com)