PhUSE 2014

Paper PP10

Data Transparency Through Metadata Management

Judith Goud, SOA Software, Bennekom, The Netherlands
Julie Smiley, SOA Software, San Antonio, US

ABSTRACT

In response to public pressure for data transparency, regulatory authorities are requiring that clinical trial sponsors make more trial data available, and industry organizations are adopting principles to publish 'all sponsored trials irrespective of the results'.

While publication of clinical trial data is a challenge, data transparency also requires assurance that reported data are accurate and are coming from the official source. Currently however, the data resides in a myriad of systems and formats, making it difficult to maintain the lineage from data collection through analysis and reporting. By managing data about the data or metadata across the enterprise, organizations can provide full data lineage for regulatory compliance and improve business efficiency at the same time.

REGULATIONS

With the general public increasingly calling for more transparency in government security and surveillance, it's not surprising that the pharmaceutical industry has lately been the target for similar calls for greater disclosures surrounding their processes and methods of doing business. After several years of pressure by medical and open access advocates, regulatory authorities (especially in the US and the European Union) began requiring that clinical trial sponsors make more trial data available.

The first requirement was for clinical trial protocols to be published so everyone has access to see all clinical studies (planning to be) performed[1,2]. Secondly, regulators were requesting that clinical trial reports containing the outcome of clinical trials be made available[3], but are now requiring these to be registered publicly[4,5].
Currently publication of anonymized clinical trial data is a hot topic. Legislation is still in progress[6,7], but initial regulation is expected to become effective in the European Union by the end of 2014. In the United States the FDA started publishing compliance and enforcement data as part of their data transparency initiative[8], and is working on final guidance for the publication of masked/de-identified non-summary data[9]. Other regulators like Health Canada and the NHS in the United Kingdom are moving in the same direction[10,11].

In order to publish patient-level clinical trial data with a common understanding of its meaning, regulators and industry need to agree on common data definitions. Also, data transparency requires assurance that reported data are accurate and are coming from the official source.

INITIATIVES

Industry organizations like the European Federation of Pharmaceutical Industries and Associations (EFPIA) and Pharmaceutical Research and Manufacturers of America (PhRMA) have also recognized the need for data transparency and have jointly laid down principles for responsible data sharing[12]. In these principles the pharmaceutical industries agree to the following: "All company-sponsored clinical trials should be considered for publication in the scientific literature irrespective of whether the results of the sponsors' clinical trials are positive or negative. At a minimum, results from all phase 3 clinical trials and any clinical trial results of significant medical importance should be submitted for publication. This commitment also pertains to investigational medicines whose development programs have been discontinued."

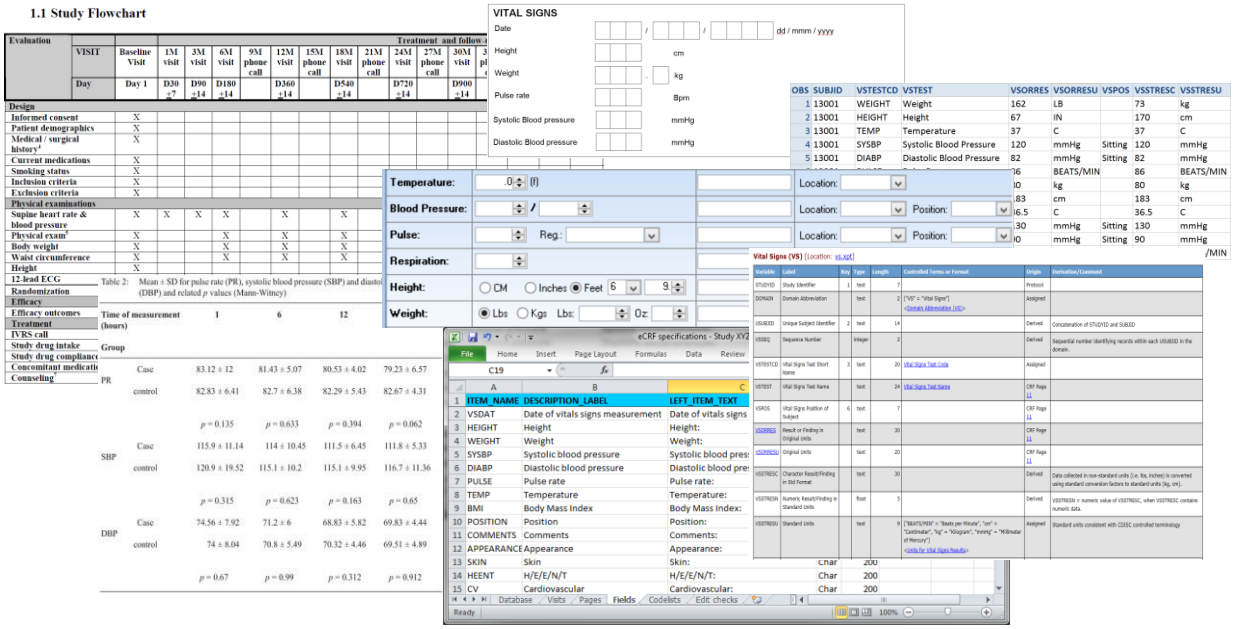
In alignment with these principles many BioPharma companies publish clinical study results on their websites, and some also provide access to their (patient-level) clinical trial data upon legitimate request. Several companies have joined forces to create common databases containing research or patient-level clinical trial data for consultation[13,14]. For population of such common databases it is helpful to use similar (standard) data definitions and formats, in which the Clinical Data Interchange Standards Consortium (CDISC) can play an important role. CDISC was established in 1997 as a standards development organization (SDO) that partners with industry and regulators to develop data standards that increase patient safety, data quality, and transparency. To date CDISC has developed multiple standards across the full clinical data lifecycle, from protocol through to submission[15].
Several regulatory authorities (European Medicines Agency, US Food & Drug Association, Pharmaceuticals and Medical Devices Agency of Japan) have recognized the CDISC data standards initiative and provided (draft) guidance that they will require clinical study data be submitted in CDISC format[16,9,17].

DATA FORMATS

While clinical data standards provide common definitions for data transparency, the data still resides in many different formats and systems: the protocol, paper case report forms (CRF) and analysis results are usually distributed in PDF or MS Word document formats, eCRF and edit check specifications are generally in spreadsheets, data collection databases and datasets for statistical analysis are stored in proprietary software formats, and the submission file in XML format (see figure 1).

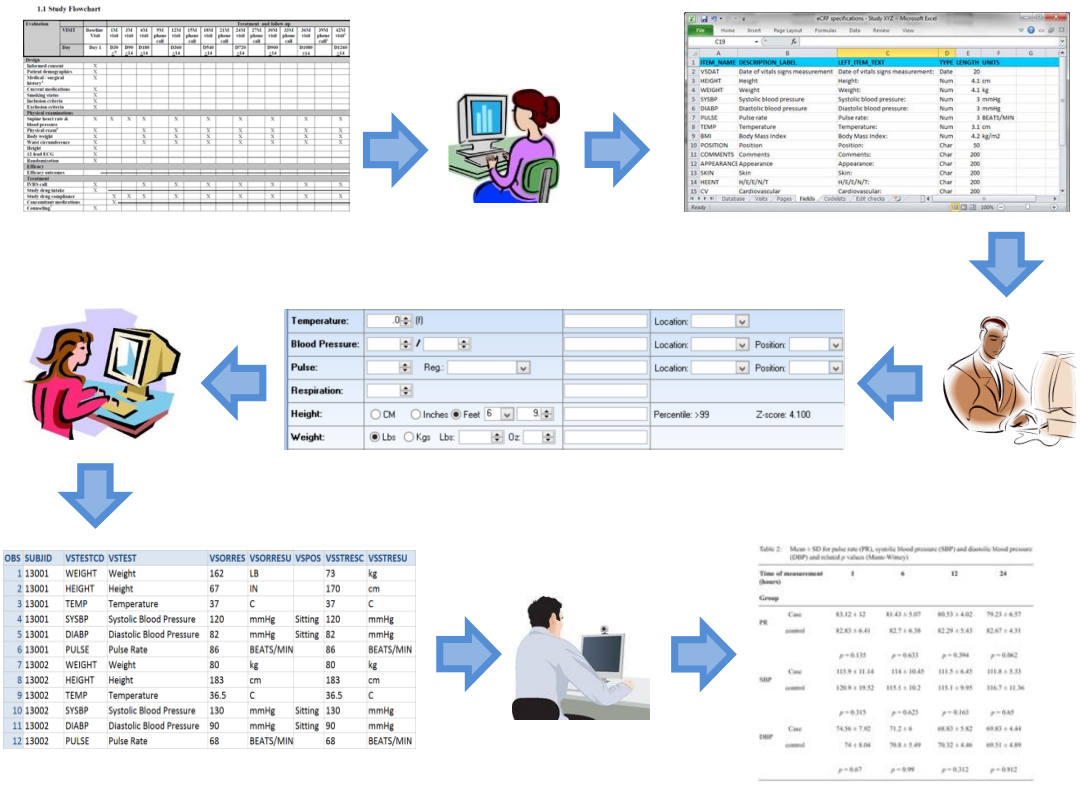
Also these files/artifacts are stored in many different folders across networks and each file/artifact is typically version controlled based on internal functional processes, making it difficult at times to know which version is the latest official version of a file or artifact.

Figure 1 Data Format Examples for Vital Signs



In common practice these files/artifacts are created via human interpretation of the previous artifact and standards, e.g. a Data Manager will interpret the protocol and CDASH standard to create a CRF, a Database Programmer will review the CRF and CDASH/SDTM standards and create a database, etc. (see figure 2). This process is prone to errors and requires significant resources for quality control (peer review or even double programming).

Figure 2 Data Lifecycle



To keep track of a data through all the different systems and formats, information about the data should be managed. These data elements should be defined broadly for use across different clinical lifecycle stages and systems, but managed more granularly by managing data about the data, the 'metadata'[18].

To reach this broader-based definition, data should be defined by industry level and preferably semantically interoperable standards, such as those developed by CDISC and the Biomedical Research Integrated Domain Group (BRIDG). The advantage of semantic interoperability (SI) is the ability of computer systems to exchange data with an understood or common meaning of the data without human intervention or interpretation. This enhances data transparency by preventing human error during data processing. More granular management of clinical content maximizes its reuse, and can significantly improve business process efficiency. It does not necessarily require managing the data itself any differently, but it does necessitate additionally managing data about the data (i.e., metadata).

METADATA

Metadata can be described as the what, when, why and how of a data, and consists of two different concepts: information about the definition and purpose of an element, plus its usage-specific information.

The usage-specific information describes the data characteristics like the name and version, and how the data element is formatted in different systems and artifacts. For example the vital sign heart rate is a text element called 'Pulse Rate' in the protocol document and a numeric data field of 3 digits with unit beats/min in the clinical database. It also contains information about references between systems and artifacts and transformations the data element is subject to during the data lifecycle. This type of metadata provides clear lineage between clinical data lifecycle stages and associated systems.

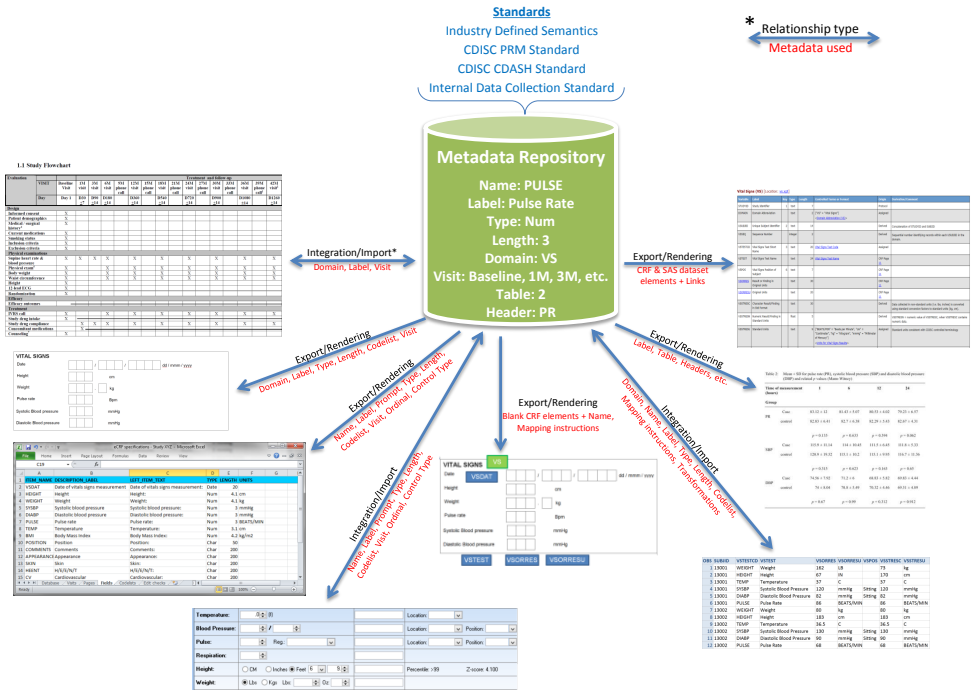
The information about purpose and usage describes at what moment a data element was used and why. This is crucial for true end-to-end data transparency as it shows the relationships from data as defined for a specific collection use case (e.g. clinical trial protocol as the starting point of data generation) through its use during analysis and reporting for a dossier (e.g. describing the drug's safety and efficacy). Further it enables tracking across all usages (e.g. different studies) and impact analysis of changes or new data elements.

METADATA MANAGEMENT

The introduction of CDISC clinical data standards raised the awareness in many companies that data standards need to be managed in a way that facilitates their consistent use. Currently the majority of BioPharma companies manage data definitions using documents and spreadsheets. However these tools are originally designed for other purposes and therefore do not support many requirements for proper content management. This results in things like change control, versioning, business rule implementation, impact analysis of changes to or reuse of elements have to be performed manually through peer review. This outcome is creation of copies (and keeping track of them) and team consultation, rather than automated governance review. This involves a significant amount of manual effort, which can lead to human error as processes need to be set up for review, governance, approval and tracking of impact analysis and where data definition changes need to be implemented.

A metadata repository is better suited for these activities and can serve as the single source or truth within an organization. The repository should hold all information about data element structure and meaning, and where and when a data element is/has been used. Ideally it also enables (automated) reuse of data elements and capture of all references and transformations the data element was subject to during the data lifecycle. Figure 3 shows an example how a centralized repository can overcome the challenges and enable effective data standard implementation.

Figure 3 How Metadata Enables Transparent Data Reuse



CONCLUSION

Data transparency is enabled by managing metadata end-to-end, including:

- Data standards
- The structure and meaning of clinical data
- Transparent lineage across the data lifecycle, as well as specific usages (studies and systems)
- Common definitions

The recommended solution for solving these data transparency challenges is a central metadata repository that:

- Serves as a single source of truth for the structure, definition, source, and use of clinical data
- Maximizes reuse of existing organizational artifacts and data assets
- Enforces consistent creation, maintenance, use and governance of standards
- Automates the impact analysis and inheritance of changes on other people, processes and systems
- Facilitates the electronic exchange of data with minimal human interpretation or intervention
- Increases efficiency across the clinical development lifecycle

REFERENCES

1. Food and Drug Administration Modernization Act, FDA, 21-Nov-1997
2. Article 11 of the Clinical Trial Directive 2001/20/EC, European Parliament, 4-Apr-2001
3. European Medicines Agency policy on access to documents (related to medicinal products for human and veterinary use), 1 December 2010
4. Section 801 of the Food and Drug Administration Amendments Act (FDAAA 801), FDA, 27-Sep-2007
5. Clinical trials: clearer rules, better protection for patients, European Parliament Press release, 2-Apr-2014
6. (Proposed legislation) Availability of Masked and De-Identified Non-Summary Safety and Efficacy Data, FDA, 3-Jun-2013
7. European Medicines Agency agrees policy on publication of clinical trial data with more user-friendly amendments, EMA, 12-Jun-2014
8. Increasing Public Access to FDA's Compliance and Enforcement Data, FDA, 22-Apr-2014
9. Providing Regulatory Submissions in Electronic Format - Standardized Study Data, FDA, Feb-2014
10. Health Canada Launching Clinical Trials Transparency Database, Health Canada, 29-Apr-2013
11. NHS England announces new commitments for increased openness and transparency, NHS, 31-Oct-2013
12. Principles for Responsible Clinical Trial Data Sharing, joint EFPIA/PhRMA statement, 18-Jul-2013
13. Roche and AstraZeneca launch medicinal chemistry data-sharing consortium to further accelerate drug discovery, Press release, 26-Jun-2013
14. Website for requesting patient level data from Bayer, Boehringer Ingelheim, GSK, Lily, Novartis, Roche, Sanofi, Takeda, UCB and/or ViiV Healthcare, see https://clinicalstudydatarequest.com
15. See http://www.cdisc.org/standards-and-implementations
16. Policy 0070: Publication and access to clinical-trial data, EMA, 24-Jun-2013
17. Basic Principles on Electronic Submission of Study Data for New Drug Applications, PDMA, 20-Jun-2014
18. Moving to a Standards-Based, Agile Clinical Development Lifecycle, White paper

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Judith Goud
Sr. Solution Architect – BioPharma
SOA Software
6-9 The Square
Stockley Park, UB11 1FW
United Kingdom
Work Phone: +31 (0)6 1241 7026
Email: judith.goud@soa.com
Web: http://www.soa.com

Julie Smiley
Director of Product Management - BioPharma
SOA Software
12100 Wilshire Blvd., Suite 1800
Los Angeles, CA 90025
United States
Work Phone: +1 210 710 2901
Email: julie.smiley@soa.com
Web: http://www.soa.com