Ahead of the Curve: Leading with Industry Data Requirements

Maria Dalton, GlaxoSmithKline, Collegeville, Pennsylvania
Nancy Haeusser, GlaxoSmithKline, Collegeville, Pennsylvania

ABSTRACT
Most programmers in the pharmaceutical industry are aware of the requirement to submit data in CDISC data format. However, the CDISC mandate is not always clear-cut as there are some nuances and ambiguities to this requirement. Also, there are many other data requirements for the pharmaceutical industry, such as data disclosure regulations, data sharing policies, and dictionary requirements. In addition, clinical trials may now include real world data or electronic health care data which follow new data standards. It can be very challenging for companies to stay “ahead of the curve” in this area. This paper discusses some of the recent developments in industry data requirements.

This paper covers industry data standards and regulatory data requirements, so it is not specific to any programming language. This paper will be helpful to all levels of programmers in the pharmaceutical industry, including lead programmers and managers.

INTRODUCTION
In this paper, we provide an overview of the current landscape of data requirements for the pharmaceutical industry. We do not provide in-depth training on CDISC or any other data standard or requirement. Instead, we explain the breadth of data requirements facing the pharmaceutical industry with particular focus on requirements for clinical data. We also list resources for obtaining more information.

DATA REQUIREMENTS FOR THE PHARMACEUTICAL INDUSTRY
In this section, we describe the different organizations that are driving data requirements for the pharmaceutical industry, and we provide information on the key data requirements governed by or required by each organization.

INTERNATIONAL COUNCIL FOR HARMONISATION (ICH) (www.ich.org)
(formerly the International Conference on Harmonisation)
The International Conference on Harmonisation (ICH) was launched in 1990 with the goal of harmonizing pharmaceutical regulatory requirements across different countries. ICH has created key foundational guidelines and standards that are widely used in the pharmaceutical industry such as:

- ICH Guidelines – a set of key guidelines for the design, conduct, and reporting of clinical trials.
- MedDRA (Medical Dictionary for Regulatory Activities) – a free standardized dictionary for medical terminology (e.g. Adverse Events). It is widely used by regulatory authorities, global pharmaceutical companies, clinical research organisations and health care professionals
- Common Technical Document (CTD) – a common format for electronic submissions that “led” to harmonized electronic submission that, in turn, enabled implementation of good review practices. For industries, it has eliminated the need to reformat the information for submission to the different ICH regulatory authorities.

REGULATORY AGENCIES
Of course, the pharmaceutical regulatory agencies are the main source of data requirements since these agencies define the legal requirements for regulatory submissions. The regulatory agencies have established data requirements in the following areas.

1. Data Standards
2. Dictionary Requirements
3. Requirements for Disclosure of Clinical Trial Results and Sharing of Patient Data
DATA STANDARDS
“The nice thing about standards is that you have so many to choose from.” – Andrew Tanenbaum

REGULATORY DATA STANDARDS CATALOGS
Currently the only regulatory agencies that require the submission of clinical data with drug applications (i.e. CRT packages) are the US Food & Drug Administration (FDA) and the Japan Pharmaceuticals and Medical Devices Agency (PMDA). As noted in the quote from Andrew Tanenbaum above, there is an abundance of data standards for clinical and medical data. The FDA and PMDA specify which data standards and versions should be used in data packages submitted to them in their Data Standards Catalogs. These catalogs list the agency's supported data standards and versions along with deadline/support dates for each version. Here are links to the FDA and PMDA Data Standards Catalogs:

- Link to FDA Data Standards Catalog: http://www.fda.gov/downloads/ForIndustry/DataStandards/StudyDataStandards/UCM340684.xlsx
- Link to PMDA Data Standards Catalog: https://www.pmda.go.jp/files/000216763.zip

REGULATORY STUDY DATA TECHNICAL CONFORMANCE GUIDES
A quick look at the FDA and PMDA Standards Catalog shows that the FDA and PMDA differ in the specific versions they require and their deadline/support dates. They also differ in their rules about how to apply versions of standards to studies: FDA requires that study data comply with the version on the FDA standards catalog at the time of study start, while the PMDA requires that the study data comply with the version on the PMDA Standard Catalog at time of submission. Both agencies list these requirements (along with many other requirements) in their Study Data Technical Conformance Guides. Here are links to the FDA and PMDA Technical Conformance Guides:


STUDY DATA STANDARDISATION PLAN AND FORM 8 DOCUMENT
Both the FDA and the PMDA require or will require a document that specifies which data standards and versions are used for every study included in a submission. The FDA calls this document the Study Data Standardisation Plan, while the PMDA’s document is called the Form 8 document.

STANDARDS FOR INTEGRATED ANALYSES
Another topic addressed in the FDA and PMDA Technical Conformance Guides is the harmonization of data standard versions for integrated analyses of data from multiple studies. The FDA Study Data Technical Conformance states: “Regardless of the specific versions used for individual studies, pooled analyses of coded terms across multiple studies (e.g., for an integrated summary of safety) should be conducted using a single version of a terminology. This will ensure a consistent and coherent comparison of clinical and scientific concepts across multiple studies.” The PMDA Technical Conformance Guide on Electronic Study Data Submissions states: “Datasets of integrated analyses of multiple clinical studies should be created using the same version, even if the version used to create the dataset of each clinical study was different.” Thus, both agencies require that integrated data packages be harmonized to a single version of a data standard. CDISC has stated that they will address CDISC standards for integrated data from multiple studies in future releases of the ADaM Implementation Guide.

REGULATORY REQUIREMENTS FOR CDISC
CDISC has become the industry standard for clinical data, primarily because the FDA now mandates the use of CDISC standards. The FDA mandate requires that all studies starting after 17-Dec-2016 must be submitted in CDISC format. The PMDA does not yet require CDISC, but they have announced that they will require CDISC beginning 01-Oct-2016, with a 3.5 year transitional period. In addition, the China Food and Drug Administration (CFDA) has announced their support of CDISC. It is likely that more countries will support and/or require data in CDISC format. The CDISC organization maintains a very useful guide to Global Regulatory Requirements for CDISC on their website at https://www.cdisc.org/resources/impending-regulatory-requirements.

As part of its implementation of the CDISC requirement, the FDA has implemented a new Technical Rejection Process effective in December 2016. This process requires the submission of an SDTM Trial Summary (TS) dataset for every study in a submission, even for non-CDISC studies. Non-CDISC studies must include the Study Start date in their TS dataset. The FDA will check the study start date in the TS dataset to determine if the study data should be in CDISC format. If the study start date is after 17-Dec-2016, the study data should be in CDISC format and the FDA will proceed to check the study data for compliance to CDISC. The FDA has also clarified that the study start date for clinical trials is the earliest date of informed consent among any enrolled subjects.

So, the CDISC data standard is clearly a mandate for the pharmaceutical industry. However, this mandate is not always clear-cut and straightforward. One complication for companies who are adopting CDISC standards is that the FDA and PMDA have differences in their CDISC requirements. Companies need to carefully consider these differences when designing data packages that will be submitted to both the US and Japan. Some of these
PhUSE 2017

differences are:

- They have different deadlines for supporting and requiring specific versions as recorded in their standards catalogs.
- These agencies have different criteria for rejecting a CDISC data package. Pinnacle21 provided a Webinar in 2016 that includes information about differences in PMDA and FDA validation rules, available at https://www.pinnacle21.com/sites/default/files/blog/2016/01/P21-PMDA-Validation-Rules.pdf.
- The PMDA is stricter about use of System International (SI) units. The PMDA has stated that they will require that SDTM datasets contain SI units, while the FDA allows use of US conventional units.\(^{16,19}\)
- PMDA prefers Analysis Results Metadata to be included in the ADaM Define.xml while the FDA does not state a preference or requirement.\(^{20}\)

To complicate the situation even more – there are cases where the regulatory agencies are more stringent about specific CDISC requirements than the CDISC model. For example, the FDA Technical Conformance Guide states that the FDA wants all applicable SDTM domains to include EPOCH.\(^{21}\) However, EPOCH is a permissible variable in the CDISC SDTM model for most domains.\(^{22}\) Similarly, the SDTM LB.LBLOINC variable is permissible according to CDISC, but the FDA lists LOINC under Terminology Standards in its Data Standards Catalog and expects to receive it.\(^{23}\)

**DICTIONARY REQUIREMENTS**

Regulatory agencies may also require that certain clinical data dictionaries or specific versions of dictionaries be used with clinical data. The PMDA and the EMA both mandate that the MedDRA dictionary for Adverse Events be used for electronic reporting. The FDA does not mandate MedDRA but it is the de facto standard in the US, and the FDA uses it in their database systems.\(^{24}\) Unlike MedDRA, there is no single industry standard for medication coding or reporting, although WHO-DD, a dictionary for medications created by the WHO Programme for International Drug Monitoring\(^ {25}\), is commonly used in the industry. Recently, the PMDA has communicated that it will require use of the WHO Drug dictionary in the future.\(^ {26,27}\)

Clinical trials are often conducted at different times during the drug development cycle, thus the data for these studies will use different versions of dictionaries and controlled terminologies. The FDA expects that Adverse Event data included in an Integrated Summary of Safety be harmonized to use a single version of MedDRA.\(^ {28}\) It is also good practice to harmonize versions of medication dictionaries and other controlled terminologies to a single version in integrated data sets.\(^ {29}\)

**REQUIREMENTS FOR DISCLOSURE OF CLINICAL TRIAL RESULTS AND SHARING OF PATIENT DATA**

Over the past 20 years, the FDA, EMA, and other organizations have expanded the requirements for public disclosure of results and data from clinical trials. Below is a brief history of the evolution of these requirements.

**1997:** The US passed the FDA Modernization Act (FDAMA) that required registration of clinical trials.\(^ {30}\)

**2005:** The International Committee of Medical Journal Editors (ICMJE) passed a requirement that “as a condition of consideration for publication, the prospective registration of certain clinical trials in a public trials registry. Failing to register makes the results of the trial ineligible for publication in the ICMJE member journals.”\(^ {31,32}\)

**2007:** The US passed the FDA Amendments Act (FDAAA). Section 801 of the FDAAA expanded the registration requirement to require “the submission of summary results, including adverse events, for certain trials” into http://www.clinicaltrials.gov.\(^ {33}\)

**2014:** the EU passed laws with new requirements for public disclosure of results and data:

- Clinical Trial Regulation EU No. 536/2014 requires that results from EU clinical trials be publicly available”.\(^ {34}\) To meet this requirement, companies post clinical trial results in the European Clinical Trials Database (EudraCT) at www.clinicaltrialsregister.eu.\(^ {35}\) This regulation also requires that “a layperson summary should accompany the summary of CT results”. This regulation will come into full effect in 2018.\(^ {36,37}\)

- Policy 70: European Medicines Agency policy on publication of clinical data for medicinal products for human use\(^ {38}\) was released in two phases:
  - Phase 1 requires the publication of anonymised information from clinical study reports, including patient narratives.\(^ {39}\) Phase 1 of Policy 70 became effective on January 1, 2015.\(^ {40}\)
  - Phase 2 will require the public sharing of individual patient data. The EMA has stated that Phase will be implemented at a future date\(^ {41}\), but they have not yet given a specific date.

The EMA has published a guidance document for Policy 70 called External guidance on the implementation of the European Medicines Agency policy on the publication of clinical data for medicinal products for human use for industry best practices and requirements for anonymisation of personal data. This guidance contains recommendations on best practices and techniques for anonymisation of personal data. The guidance also recommends selecting an appropriate metric for measuring the risk of re-identification of personal data, and it advises to set the threshold of the risk metric to 0.09, meaning that the risk of re-identification should be < 0.09.\(^ {42}\)
2016: The US issued the Final Rule for the FDAAA act requires that additional types of clinical trials must comply with the registration requirement and requires that additional data elements be entered into the trial registry. The final rule of the FDAAA act took effect on January 18, 2017.43

The end result of these regulations is that pharmaceutical companies must register their clinical trials, disclose and publish results from those trials, and prepare for the upcoming requirement for public sharing of anonymised patient data. In particular, the requirement to share patient data will require careful planning in order to develop robust processes that protect the identity and privacy of patients.

ORGANIZATIONS DEVELOPING OR SUPPORTING DATA STANDARDS
There are a number of organizations that develop or support the development of data standards for clinical and pre-clinical data. Note that these organizations do not mandate the use of their standards - only regulatory agencies can do that.

CLINICAL DATA INTERCHANGE STANDARDS CONSORTIUM (CDISC)
The Clinical Data Interchange Standards Consortium (CDISC) is very well-known in the pharmaceutical industry, and there is a wealth of information and resources available for CDISC standards. Consequently, we will only describe a few highlights of the CDISC organization.

CDISC is a volunteer, multidisciplinary organization that develops data standards for clinical research.44 It has developed standards in the following areas:

- Foundational data standards, e.g., SDTM, CDASH, ADaM, SEND
- Data Standards for Therapeutic Areas
- Controlled Terminology
- Data Exchange Standards (e.g. Define.xml and ODM – Operational Data Model)45

CDISC also maintains the CDISC Shared Health and Research Electronic Library (SHARE) – an electronic metadata repository for CDISC standards.46

International Organization for Standardization (ISO)
ISO is an international non-governmental consortium of 164 different standards organizations.47 ISO primarily creates requirement documents, specifications, and guidelines about standards. They describe their role as "similar to that of a conductor, while the orchestra is made up of independent technical experts."48 Examples of ISO standards that impact the pharmaceutical industry are

- ISO 8601 standard for Date & Time format (incorporated into CDISC)
- ISO Identification of Medicinal Product (IDMP) standards for pharmaceutical products. IDMP covers medical products as well as substances, routes of administration, dosage forms, units of presentation and packaging, and units of measurement.49 CDISC has announced that it will develop a new Clinical Trial Registry 2 (CTR2) standard in the future that will align CDISC and IDMP controlled vocabularies.50

HEALTH LEVEL SEVEN (HL7) GROUP
Founded in 1987, HL7 is a non-profit organization that develops data standards “for the exchange, integration, sharing, and retrieval of electronic health information.”51 HL7’s standards are recognized as the most commonly used in the world for electronic health record data.52 HL7 also provides Implementation Guides, technical specifications, and education. 53

SNOMED INTERNATIONAL
SNOMED International is a non-profit standards developing organization that has created SNOMED CT, a standard for healthcare terminology used in electronic health records.54 The FDA Data Standards Catalog requires the use of SNOMED CT for indication in the SDTM TS Domain.55

FDA SUBSTANCE REGISTRATION SYSTEM - UNIQUE INGREDIENT IDENTIFIER (UNII)
The Substance Registration Board, a joint board of the FDA and the United States Pharmacopeia, have created unique ingredient identifiers (UNII) for substances in drugs, biologics, foods, and devices.56 The FDA Data Standards Catalog requires the use of UNII identifiers in the SDTM and SEND TS domain and in CM.CMDECOD.57

NATIONAL DRUG FILE-REFERENCE TERMINOLOGY (NDF-RT)
The National Drug File Reference Terminology is a standard for Pharmacological Class of drugs developed by the Department of Veterans Affairs/Veterans Health Administration. 58 The FDA Data Standards Catalog requires the use of NDF-RT in the SDTM and SEND TS domain.59
PhUSE 2017

LOINC®
The Regenstrief Institute, a health research institution, has developed the LOINC® standard for laboratory and clinical tests, measurements, and observations. The FDA will require the use of LOINC® codes in the SDTM LB domain starting in 2018.

CONCLUSION
The landscape of pharmaceutical data requirements is complex and constantly changing. Keeping up with all of the rules and standards can be overwhelming. To stay ahead of the curve of pharmaceutical data requirements, companies must commit time, money, and resource. They must build a strong infrastructure of staff and computer systems so they can proactively monitor and comply with data requirements. A deliberate and proactive approach is key to avoiding crises where companies scramble at the last minute to try to comply with a new requirement. From our experience, these last-minute scrambles are resource-intensive and inefficient, and they put quality at risk, since last-minute changes to systems and processes can increase the risk of errors.

REFERENCES


GlaxoSmithKline communication with FDA eData Team, November 2016.


CONTACT INFORMATION

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

Author Name Maria Dalton and Nancy Haeusser
Company GlaxoSmithKline
Email: Nancy.N.Haeusser@GSK.com

Brand and product names are trademarks of their respective companies.