Regulatory Toxicology and Pharmacology 91 (2017) 77–85

Contents lists available at ScienceDirect

Regulatory Toxicology and Pharmacology

journal homepage: www.elsevier.com/locate/yrtph

# Making sense of SEND; the Standard for Exchange of Nonclinical Data

Katharine Briggs

*Lhasa Limited, Granary Wharf House, 2 Canal Wharf, Leeds, LS11 5PS, United Kingdom*

ARTICLE INFO

*Keywords:*
SEND
KNIME
Open access
Data mining
eTOX

ABSTRACT

The Standard for Exchange of Nonclinical Data (SEND) is currently the preferred submission format for nonclinical animal data by the US FDA and became a requirement on the 18th December 2016. Application of these data standards is the first step to being able to perform cross-study querying and is expected to open up opportunities for data mining and meta-analysis by the pharmaceutical industry. This paper reports on our experiences in developing a tool to allow recent SEND formatted studies to be explored alongside historical nonclinical data already gathered as part of the eTOX project. Combining SEND data with historical data will positively impact the power of any analysis performed and increase the likelihood of being able to detect rare effects. It describes the use of KNIME in generating dose group averages and incidences from individual animal level data captured in SEND. There are a number of options for opening and reading SEND files but the benefits of using KNIME are that it is a free, open source data mining framework which allows the data to be viewed in a holistic manner rather than one domain at a time. Additionally it incorporates several nodes useful for aggregating and visualising the data to more easily identify patterns and trends.

## 1. Introduction

The Standard for Exchange of Nonclinical Data (SEND) is intended to be a format for interchange of data between pharmaceutical companies and contract research organisations (CROs) and for submission to regulatory authorities such as the US Food and Drug Administration (FDA). SEND is based on the Study Data Tabulation Model (SDTM), the standard format for electronic submission of clinical data to the FDA. SEND, SDTM, and the associated Controlled Terminology have been developed by the Clinical Data Interchange Standards Consortium (CDISC), a non-profit standards development organization. Work on the SEND standard began in 2002 and the first essentially complete Implementation Guide (SEND-IG 3.0) was issued in 2011 (Keenan & Goodman, 2014). SENDIG v3.0 was designed to support single- and repeat-dose general toxicology and carcinogenicity studies[1].

SEND is currently the preferred submission format for the US FDA and became a requirement on the 18th December 2016[2] (see Table 1). By standardizing the electronic formats for both clinical (SDTM) and nonclinical (SEND) study submissions the FDA aim to reduce the new drug application examination time (Anzai et al., 2015). The Nonclinical Information Management System (NIMS) suite used by the FDA provides tools that are built to use SEND datasets so that they are able to review a submission more efficiently than when they receive only PDF or printed submissions.

Data standardization is the first step to being able to perform cross-study querying and SEND is expected to open up opportunities for data mining of nonclinical data and meta-analysis by the pharmaceutical industry (Brown et al., 2016). For instance comparison of the effectiveness of different study designs, identification of species specific effects, target related effects, reference ranges, Bayesian priors, e.t.c (Kaufman et al., 2016). These benefits have yet to be realised and are dependent either on sufficient time passing to gather the necessary quantity of SEND datasets in an appropriate repository, or by combining SEND data with historical datasets such as has been gathered in the eTOX project.

The eTOX project is a public-private partnership within the framework of the European Innovative Medicines Initiative (Steger-Hartmann & Pognan, 2017). Data from previously unpublished, nonclinical legacy reports from 13 European pharmaceutical companies have been incorporated into a database. The data have been harmonised using ontologies aligned to SEND and INHAND initiatives which were also developed as part of the eTOX project. The legacy data donated to the project to date was not in SEND format and required extraction by a CRO.

To avoid the cost of extracting data in the future, a software tool which could be used to import SEND formatted data directly into the eTOX database schema was identified by the consortium as critical for sustainability. In addition, it would open up possibilities to explore recent SEND formatted studies alongside this valuable dataset of historical data. Combining SEND data with historical data will allow a larger analysis of preclinical data across pharmaceutical organization archives. This will positively impact the power of any analysis

*E-mail address:* Katharine.briggs@lhasalimited.org.

http://dx.doi.org/10.1016/j.yrtph.2017.10.012
Received 7 August 2017; Received in revised form 16 October 2017; Accepted 17 October 2017
Available online 21 October 2017

**Abbreviations**

| | |
|---|---|
| ANDA | Abbreviated new drug application |
| ALT | Alanine aminotransferase |
| BLA | Biological licensing application |
| CDISC | Clinical data interchange standards consortium |
| CRO | Contract research organisations |
| EFPIA | European federation of pharmaceutical industries and associations |
| eTOX | Integrating bioinformatics and chemoinformatics approaches for the development of expert systems allowing the in silico prediction of toxicities |
| eTRANSAFE | Enhancing Translational Safety Assessment through Integrative Knowledge Management |
| FDA | US food and drug administration |
| IBM | International business machines corporation |
| IEEE | Institute of electrical and electronics engineers |
| IMI | Innovative medicines initiative |
| IND | Investigational new drug application |
| INHAND | International harmonization of nomenclature and diagnostic criteria |
| ISO | International organization for standardization |
| KNIME | Konstanz information miner |
| NDA | New drug application |
| NexGETS | Next generation of electronic translational safety |
| NIMS | Nonclinical information management system |
| PDF | Portable document format |
| PhUSE | Pharmaceutical Users Software Exchange |
| SAS | Statistical analysis system |
| SEND | Standard for exchange of nonclinical data |
| SENDIG | SEND implementation guide |
| SDTM | Study data tabulation model |
| ULN | upper limits of normal |
| XPT | SAS V5 XPORT transport files |

**Table 1**
Dates when FDA require submissions to be submitted in SEND format.

| Submission Type | Dates when FDA require SEND format |
|---|---|
| NDA, ANDA, and certain BLA submissions | Studies which start after 2016-12-18 (December 18th, 2016) |
| Commercial INDs and amendments, except for submissions described in section 561 of the Federal Food, Drug, and Cosmetic Act | Studies which start after 2017-12-18 (December 18th, 2017) |

performed and increase the likelihood of being able to detect rare effects.

The tool would need to open, read and summarise the data before import as the data in eTOX is captured at a different level of granularity; SEND data is captured at the individual animal level whereas eTOX captures dose group averages and incidences. The software would also need to prompt for information on chemical structure and pharmacology as this data is not currently captured in SEND but is considered important metadata e.g. for data mining and read-across purposes.

This paper reports on our experiences in developing SEND import functionality, handling differences in example SEND submissions from multiple sources and issues encountered with attempting to summarise it.

## 2. Method

As a first step in developing SEND import functionality we compared the eTOX database schema to the SENDIG v3.0 and performed an initial mapping of the concepts. We then developed a Konstanz Information Miner (KNIME) workflow[3] in order to test out this mapping using real SEND submissions, including 2 public submissions[2,4] and 11 SEND submissions donated by 3 EFPIA partners participating in the eTOX project. This allowed us to produce a revised mapping factoring in lessons learned from the example submissions. The revised mapping was then used as the basis for developing the software tool. Our experiences with converting the data contained in the example SEND submissions also led to improvements in the eTOX database schema. The small number of SEND submissions that were donated reflects the results of a survey of EFPIA partners who were participants in the eTOX project conducted in September 2014 which indicated that the majority were in the early stages of implementation of the SEND standard (Watanabe et al., 2017).

### 2.1. Variation in the domains included in SEND submissions

A first version of the SEND import tool was released to the eTOX consortium in November 2016. This was considered to be a prototype since the number of example SEND submissions used to develop the software was quite low and the diversity encountered with these submissions was relatively high particularly in the case of the clinical signs domain.

The example SEND submissions were all provided as zipped files which when unzipped included several files in .XPT (SAS Transport v5) format representing the different SEND domains. SEND submissions should only contain domains for which data were collected which means that zipped files can vary significantly in content. The PhUSE SEND implementation wiki[2] suggests that TS, TX, TA, TE, SE, DM and EX domains (abbreviations explained in Table 2) can be expected as a minimum. However of the 13 example SEND submissions we received, SE and EX domains were not present in all cases. Table 2 reports on the composition of the example SEND submissions:

### 2.2. Options for reading SEND submissions

There are a number of open source offerings for opening .XPT files; 1) SAS - provides a free universal .XPT file viewer (previously known as the SAS Viewer)[5], 2) R - a free software environment for statistical computing and graphics[6] or 3) pandas - a python data analysis toolkit[7].

However, in order to read the data into KNIME, a free and open source data mining framework, we opted to create a specific 'XPT Reader Node'. Reading the data into KNIME was relatively straight forward apart from the need to convert numeric data from the IBM mainframe double precision floating point number format used in .XPT files to the IEEE standard for floating point arithmetic used by most platforms[8].

### 2.3. Differences in the variables included in SEND domains

The data in a single .XPT is structured around a set of variables (field names, column headings) associated with that domain. These variables can be classified as required (must be present, must be populated), expected (must be present, may or may not be populated) or permissible (may or may not be present, may or may not be populated),