PHUSE US CSS 2020 : PP09

SEND Data Factory

Bob Friedman, Xybion; Kevin Snyder, FDA; Eli Miller, Atorus Research; Daniel Russo, FDA

Abstract

The PHUSE non-clinical scripts team is developing a SEND Data Factory tool which can create example SEND datasets. This will allow industry stakeholders to produce example SEND datasets for use in the testing visualizations, data exchange and tools development.

The program is in a preliminary but working state, written in R-Shiny and shared on the PHUSE Github for group collaboration or downloading and customization. This tools allows for the selection of the study design (number of dose groups and animals per group), desired measurement (body weights, clinical observations, micropathology, etc.) and it then can output .xpt dataset files with synthetic data.

Future work is planned to improve the realism of the data by allowing configuration of species and age specific ranges, simulated growth and dose responses, and the desired distribution of pathological observations.

Methods:

- This tool is written in R-Shiny for GUI and data manipulation.
- Numerical data is synthesized through random data within configurable ranges by specie. Finding observations data is synthesized through random data based upon configurable expected frequency of occurrence.
- A successful dataset will be defined as one that passes a validator with no errors and that has data that is scientifically possible.

Use cases:

- Demonstrating data visualization tools
- Testing systems that receive data from the evolving standards
- Creating large data sets for testing data loading
- As example code to show:
  - R script creation of XPT files
  - R script reading of the CDISC SENDIG standard (currently PDF)
  - R script reading of NCI Control Terminology files (xls version)

Features

- Controlled Terminology version selection
- Study type, species selections
- Number of groups and animals per group
- Recovery and TK groups
- Trial summary data and dosing settings

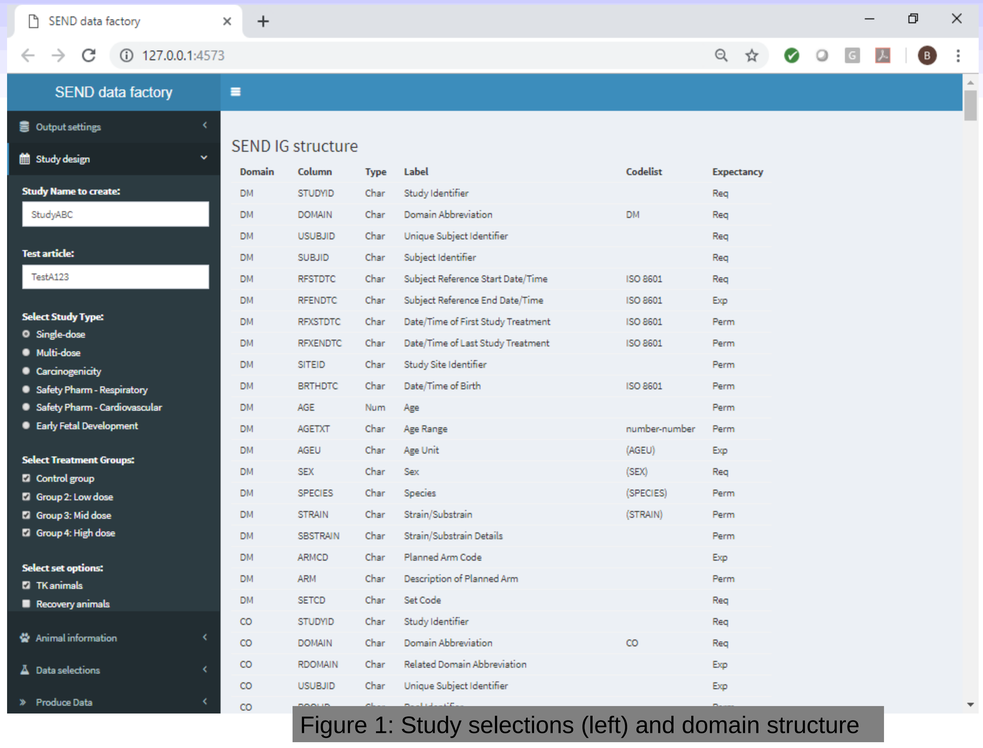

Figure 1: Study selections (left) and domain structure

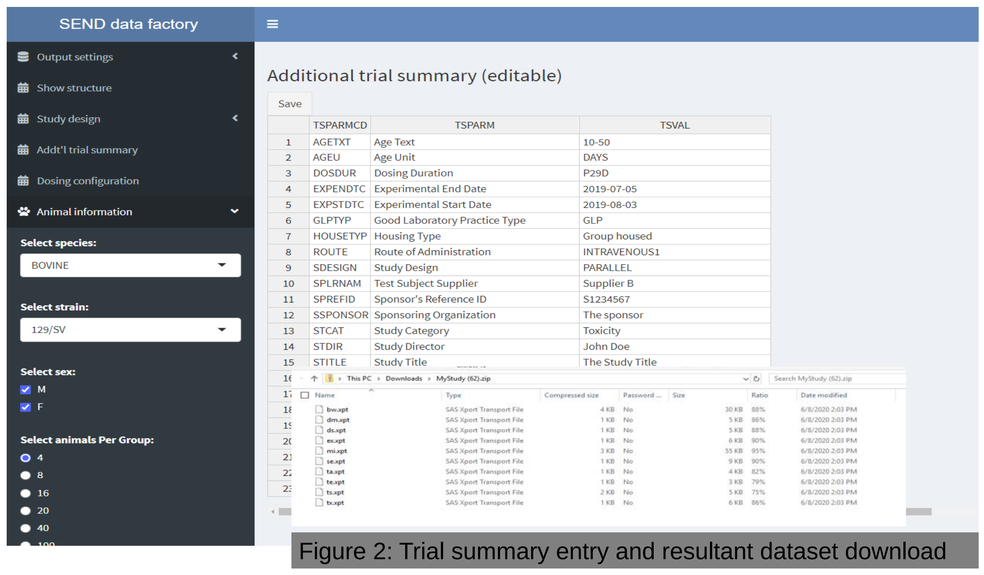

Figure 2: Trial summary entry and resultant dataset download

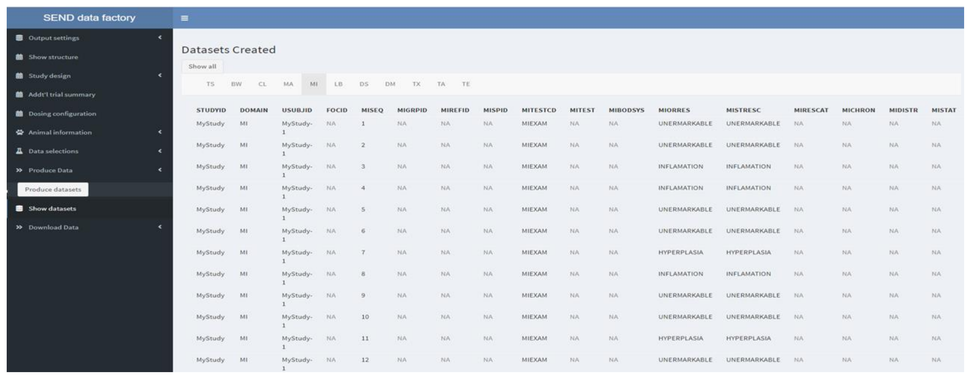

Figure 3: Dataset tabbed view

Next steps:

- Ability to create custom and experimental domains
- Simulated time progress for numerical endpoints
- Simulated dose responses for endpoints
- Create test datasets with validation rule violations

Conclusion:

This tools can be used by an industry group or an organization to create proof-of-concept datasets. The tools can be used in the future to test changes and enhancements to SEND IG by giving stakeholders an application to interact with changes in a practical way.

Note: The opinions expressed in this poster are those of the authors and do not necessarily represent the opinions of their respective organizations.

Figure 4: Nonclinical Script Assessment PhUSE Wiki page

Figure 5: SEND Data Factory Github Project

Figure 6: Public website running SEND Data Factory