Standards and Open-Source Hand-in-Hand: Leveraging Automation to Expedite Drug Market Request Review Process

Angelo Tinazzi, Cytel Inc.
Angelo Tinazzi is Senior Director, Statistical Programming, responsible for Clinical Data Standards and Data Submission at Cytel.

He is a well-published and recognized expert in statistical programming and data standards with over 25 years' experience working with different organizations in Italy, UK, and Switzerland.

Angelo is an authorized CDISC ADaM instructor, stream chair for PHUSE-EU, committee member of the CDISC European Committee, where he also manages the Italian-speaking CDISC User Network.

Since 2018 he maintains a Cytel blog “The Good Data Submission Doctor” https://www.cytel.com/blog/author/angelo-tinazzi
My journey so far
Born in Milan (Italy) sometimes in ’70….

Football  Fashion  History
My journey so far
Born in Milan (Italy) sometimes in '70…. In what is considered today the Milan Chinatown

It is the oldest and largest Chinese community in Italy, with about 21,000 people in 2011

Originally established in via Canonica in the 1920s, used to operate small textile and leather workshops

Today the district is filled with hairdressing salons, fashion boutiques, silk and leather stores, libraries, traveling agencies, medicine centres and massage parlours
My journey so far

Grew-up and studied in Milan (Statistics)

Professionally grew-up at Mario Negri Pharmacological Institute in the nineties
- Discovered Clinical Trials
- Discovered the beauty of Clinical Data and SAS

Fine-tuned Computer science applied to Stats in Cambridge, UK, late ’90

Joined Pharma Industry in 2000 (Pharmacia & Upjohn)

Discovered CDISC in 2003

- Failed some submissions

Joined Cytel in 2012
- Successfully submitted first time in 2013
- Lost the count of how many submissions, mostly successful

Live in France (Ferney Voltaire) since 2010
- Crossing the border to go to work in Geneva everyday

Meanwhile one wife and three kids (21, 12, 10)
Agenda

• 01 • Preview of Data Submission in 2041
  Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore Lorem ipsum dolor sit amet, consectetur adipiscing

• 02 • 2023 targeting 2041

• 03 • Standards and Open-Source Initiatives

• 04 • Final Remarks
Imagine Data Submission in 2041, what do you see?

Summary of stories discussed, technologies hypothesis

From: *AI 2041*, Kai-Fu Lee – Chen Qiufan
Imagine Data Submission in 2041, what do you see?

A Reviewer stepping into your lab through a visor checking your data and results?
An IA doing the submission for you and discussing with the reviewer?
An IA reviewer?
An xD / Multidimensional Database? Well, that’s already possible
Let’s go back to 2023
What are we doing today to target 2041 vision?

NEW DATA STANDARDS

OPEN SOURCES INITIATIVES

NEW TECHNOLOGIES
Let’s go back to 2023  Some attempts
What are we doing today to target 2041 vision?

Automatic Defining ADaM for new Clinical Studies Using Machine Learning
Thomas Rye Olsen,
Student at Department of Computer Science, University of Copenhagen

Henning Pontoppidan Föh,
Statistical Programming Director, Biostatistics, Novo Nordisk A/S
Let’s go back to 2023
Open-Source Initiatives – FDA Statement

Statistical Software Clarifying Statement

FDA does not require use of any specific software for statistical analyses, and statistical software is not explicitly discussed in Title 21 of the Code of Federal Regulations [e.g., in 21CFR part 11]. However, the software package(s) used for statistical analyses should be fully documented in the submission, including version and build identification.
Let’s go back to 2023
Open-Source Initiatives – A change already occurred

PHUSE EU 2016-2023 Presentations showing or discussing SAS, Open-Source or other tools
Let’s go back to 2023
Open-Source Initiatives – A change already occurred

PHARMASUG – CHINA 2023 ….. THIS CONFERENCE
Sponsor aim to get market approval for their products

- A new Device
- A new Drug

*A software, a SAS macro, an R library is not what a sponsor wants to make money from*
2023 → 2041 Obstacles

- ✔ Collaboration
- ✗ Data Format
- ➔ still submitting
- ✗ “Dynamic/Interactive” submission
2041? Or may be earlier “Dynamic/Interactive”

Demonstrate data packages can be created using other software e.g., R

- Ut ut lorem eget justo accumsan eleifend.
- Ut vulputate lorem in est euismod egestas.
- Pellentesque ultrices velit eu lorem porta ultricies.

Sharing R Packages

- Include data visualization
- Provide tool the HA can re-use work with it
- Installing and using Sponsor R packages

Sed id mauris ac ex gravida porttitor

- Sed vitae nulla varius, mattis massa id, gravida quam.
- Donec nec enim quis ex venenatis scelerisque.
- Phasellus eu ipsum convallis leo imperdiet faucibus ut ac ante.
2041? Or may be earlier “Dynamic/Interactive”
A good example “The R Pilot Submission Experience”

FDA can receive submission with R code/packages

- Objective: to test the concept that an R-language-based submission package can meet the needs and the expectations of the FDA reviewers, including assessing code review and analyses reproducibility.
- Evaluating FDA’s acceptance of system/software validation evidence is not in the scope of this pilot.
- All submission materials and communications from this pilot are publicly available. (Link)

Submitted Material

- What R consortium submitted:
  - ADaM datasets (.xpt files)
  - A pdf report with 4 analysis outputs
  - Analysis Data Reviewer’s Guide (ADRG)
  - Analysis output programs (.r files)
  - Sponsor developed R package (.txt file)

Install Sponsor Packages and reproduced

- Receive electronic submission package in eCTD approved formats.
- Reconstruct and load the submitted sponsor-developed R package.
- Install and load open-source packages used in this submission.
- Reproduce the analysis results.
- Share potential improvements for submission deliverables and processes via written communications.

FDA Feedback

- Using R version 4.1.1, FDA was able to run the submitted code and confirm the submitted tables and figures.
- Using FDA developed code, FDA was able to independently generate tables and figures using the submitted data.
- There were minor issues.
  - Rounding issue
  - Important information was not given in the table

https://github.com/RConsortium/submissions-pilot1-to-fda
2041? Or may be earlier “Dynamic/Interactive”
A good example “The R Pilot Submission Experience”

https://rconsortium.shinyapps.io/submissions-pilot2/

Pilot 2 Shiny Application

**Introduction**

This application is intended for a pilot submission to the FDA composing of a Shiny application, as part of the R Submissions Working Group Pilot 2. The data sets and results displayed in the application originate from the Pilot 1 project. Visit the Usage Guide for information on using the application. Below is a brief description of the application components:

**Demographic Table**

In this interface, summary statistics associated with baseline clinical characteristics and other demographic factors is shown.

**KM-Plot for TTDE**

A Kaplan-Meier (KM) plot of the Time to First Dermatologic Event (TTDE) with strata defined by treatment group is displayed along with an informative risk set table across time.

**Primary Table**

A summary table of the primary efficacy analysis is shown for each of the time points of assessment (baseline and week 24) comparing each treatment group. The primary efficacy variable (change from baseline in ADAS Cog (11)) was analyzed using an Analysis of Covariance (ANCOVA) model with treatment and baseline value as covariates, comparing Placebo to Xanomeline High Dose.

**Efficacy Table**

A summary table of an additional efficacy analysis is shown for baseline and week 20. The efficacy variable (Glucose) was analyzing using ANCOVA model with treatment and baseline value as
2041? Or may be earlier “Dynamic/Interactive”
A good example “The R Pilot Submission Experience”
https://rconsortium.shinyapps.io/submissions-pilot2/

Pilot 2 Shiny Application

Important information:
The analyses performed when utilizing subgroups or other subsets of the source data sets are considered exploratory.
- Treatment information variables from the ADTTE data set are excluded from the variable list. Use the treatment variables present in the ADSL set to perform treatment-related filters.
- In rare situations, applying filters with variables from both ADSL and ADTTE that overlap in content could result in an invalid data subset. When possible, select variables with distinct content.

Active Filter Summary

<table>
<thead>
<tr>
<th></th>
<th>Obs</th>
<th>Subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADSL</td>
<td>254/254</td>
<td>254/254</td>
</tr>
<tr>
<td>ADTTE</td>
<td>254/254</td>
<td>254/254</td>
</tr>
</tbody>
</table>

Active Filter Variables

- ADSL
- ADTTE

Add Filter Variables

Add ADSL filter
Select variable to filter

Add ADTTE filter
Select variable to filter
CDISC Initiatives at Glance
Standards for 2041 Vision

Analysis Results Standards

- Use analysis results metadata to drive the **automation of results**
- Support **storage**, **access**, **processing** and **reproducibility** of results
- Improved **navigation** and **reusability** of analyses and results
- **Traceability** to Protocol/SAP and to input ADaM data

With permission from Peter Van Reusel, Chief Standards Officer, CDISC
No More XPT? Piloting New Dataset-JSON For FDA Submissions

By Sam Hume, D.Sc., vice president, data science, CDISC

When submitting study datasets, the FDA requires organizations to use the SAS V5 XPORT (XPT), a format that dates back to 1989. Originally announced in the FDA’s 1999 Guidance for Industry—Providing Regulatory Submissions in Electronic Format: General Considerations, the XPT requirement set the format for representing datasets in the CDISC Foundational Standards. Since then, many aspects of the submission process have improved, yet the XPT requirement has remained. The outdated XPT format imposes restrictions on submission data, including limited data types, no Unicode support, variable name and field size constraints, inefficient use of storage space, lack of extensibility, and insufficient metadata. Moreover, XPT’s binary format limits its utility for use in many modern data exchange scenarios.

CDISC developed Dataset-JSON, a new dataset exchange format, to replace XPT. Working with the FDA and industry participants, CDISC and PHUSE lead a new pilot project to test Dataset-JSON for use in regulatory submissions as well as other dataset exchange...
CDISC Initiatives at Glance

Standards for 2041 Vision

Regulatory
Currently Mandated by Regulatory Agencies

Pre-history
Considered outdated and antiquated

Limitations of XPT v5
• Numeric limitations, antiquated format
• Stores data in its own numeric way
• Character limitations, no UTF-8 encoding
• No support for characters from other languages
• String & Column limitations (variable names > 8, labels > 40, data > 200)
• No metadata extensibility

No more public complaints...
...from my friend Jozef Aertz

Thanks, and Goodbye SAS XPT
Welcome Dataset-JSON
CDISC Initiatives at Glance Standards for 2041 Vision

What is Dataset-JSON?

- A dataset exchange standard for exchanging tabular data leveraging JSON designed to meet the regulatory submission needs and eliminating limitations of legacy formats.
- Based on the JSON standard used worldwide.
- Simple transformation to/from SAS data.
- Open-source and truly human readable.
- Same or smaller file sizes relative to current required format.
- Remove variable naming, width, or format limitations.
- Pilot

Thanks and Goodbye SAS XPT Welcome Dataset-JSON

CDISC Initiatives at Glance Standards for 2041 Vision
CDISC Initiatives at Glance
Standards for 2041 Vision

Standards as a Service

Software Applications Consume Standards Metadata via the API

With permission from Peter Van Reusel, Chief Standards Officer, CDISC
Open-Source for 2041 Vision

Supports and promotes open-source software projects that create tools for implementing or developing CDISC standards to drive innovation in the CDISC community.

With permission from Peter Van Reusel, Chief Standards Officer, CDISC
Open-Source for 2041 Vision
Open-Source for 2041 Vision
Open-Source for 2041 Vision
Open-Source for 2041 Vision
Final Remarks
What is there for you?

We have one fundamental advantage over our peers, 5 years ago we invested heavily in a data lake called, we’re leveraging that data lake to move AI very quickly in the company

Vasant Narasimhan
Useful Resources

- “Integrated Summaries of Effectiveness and Safety: Location Within the Common Technical Document”, FDA Guidance for the Industry
- “ADaM Structures for Integration: A Preview”, W. Zhong, K. Minkalis and D. Bauer, PharmaSUG 2018
- “ADaMIG v1.2 & ADaM Integration”, CDISC Webinar, 2019, https://www.cdisc.org/events/webinar/adamig-v1.2-adam-integration
- “Study Data Standardization Plan” https://phuse.s3.eu-central-1.amazonaws.com/Deliverables/Optimizing+the+Use+of+Data+Standards/SDSP.zip
Thank You

Angelo Tinazzi, Senior Director
Cytel Inc.
Standards, Systems, CDISC Consulting, Statistical Programming
Clinical Research Services
Route de Pré-Bois 20
C.P 1839, 1215 Geneva, SWITZERLAND
angelo.tinazzi@cytel.com

Subscribe to Cytel Blog to receive update on my
Good Data Submission Doctor Blog Series
https://www.cytel.com/blog/topic/cdisc
Abstract

In recent years, the emergence of data standards has significantly accelerated the review process and fostered a sense of common understanding within the pharmaceutical community, now speaking the same language. These standards have also facilitated better communication with regulatory agencies worldwide. This remarkable transformation owes its success to organizations like CDISC and the support of numerous volunteers from diverse organizations across the globe.

As we progress further, a new revolution is underway, driven by the proliferation of open-source initiatives. This movement brings together volunteers from various organizations, united in their collaborative efforts to develop and maintain open-source tools. This, combined with the effective utilization of data standards, plays a crucial role in harnessing automation within our industry.

Through this presentation, I aim to explore these pioneering initiatives and evolving standards, providing a glimpse into the future of data submission.
Backup Slides
Timeline

- **2020**: Lorem ipsum dolor sit amet, consectetur adipiscing elit
- **2021**: Lorem ipsum dolor sit amet, consectetur adipiscing elit
- **2022**: Lorem ipsum dolor sit amet, consectetur adipiscing elit
- **2023**: Lorem ipsum dolor sit amet, consectetur adipiscing elit
- **2024**: Lorem ipsum dolor sit amet, consectetur adipiscing elit
- **2025**: Lorem ipsum dolor sit amet, consectetur adipiscing elit
- **2026**: Lorem ipsum dolor sit amet, consectetur adipiscing elit
- **2027**: Lorem ipsum dolor sit amet, consectetur adipiscing elit
Four parts with laptop

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua sed

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua sed

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua sed

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua sed
Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua sed