Open to Integration:
A Case Study on Open Source Tools for ISS Harmony

Presented by Steve Nicholas, Director, Analytics Engineering, Atorus Research
14Oct2025

Meet the Speaker

Steve Nicholas

Title: Director, Analytics Engineering

Organization: Atorus Research

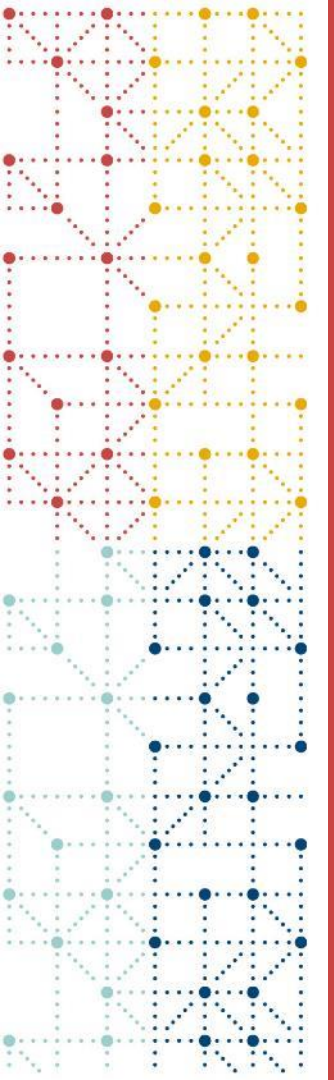
Steve is an experienced leader in statistical programming with expertise in CDISC standards, integrations, regulatory submissions, and the FSP model.

He’s led global, cross-functional teams to deliver high-quality results.

A family man, passionate about outdoor activities and coaching - always striving to balance leadership, technical expertise, and personal fulfillment.

Disclaimer and Disclosures

- *The views and opinions expressed in this presentation are those of the author(s) and do not necessarily reflect the official policy or position of CDISC.*
- *The author has no real or apparent conflicts of interest to report.*



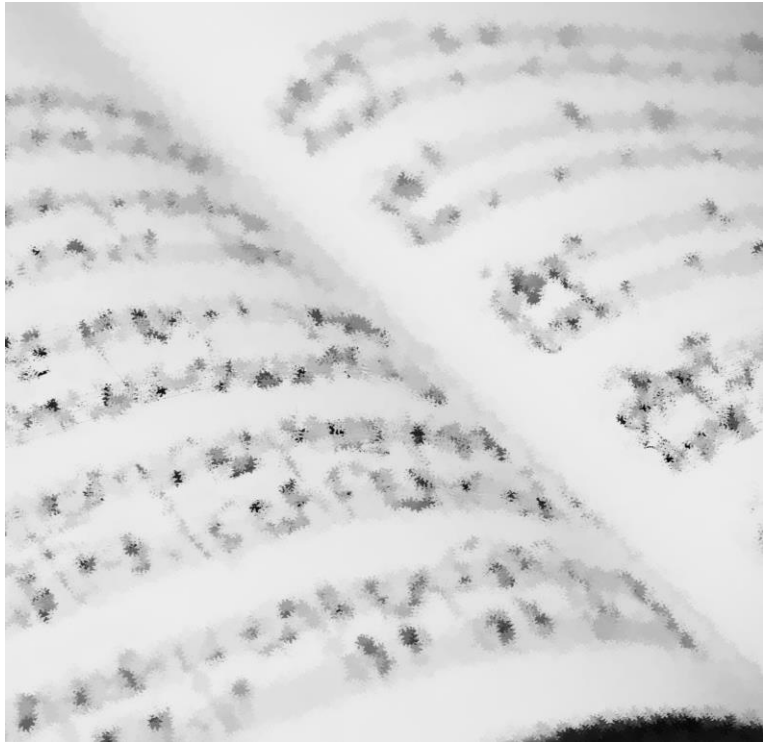
Agenda

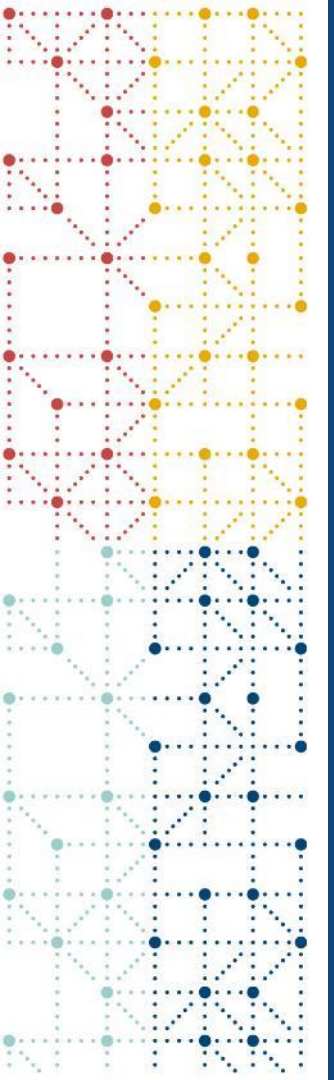
1. ISS Overview & Challenges
2. Open Source Landscape
3. Case Study Walkthrough
4. Challenges, Benefits, & Future Directions

Integrated Summary of Safety: Overview & Challenges

ISS Overview & Challenges

- Why ISS Matters
  - Critical for regulatory submissions
  - Requires integration across many studies
- Core Challenges
  - **Standards diversity:** differing data collection methods, MedDRA versions, derivation logic
  - **Sponsor-specific nuances:** custom specs, varied conventions, legacy data structures
  - **Traceability gaps:** linking study-level datasets to pooled outputs is complex
- Practical Barriers
  - Resource & time intensive
  - No “one-size-fits-all” solution

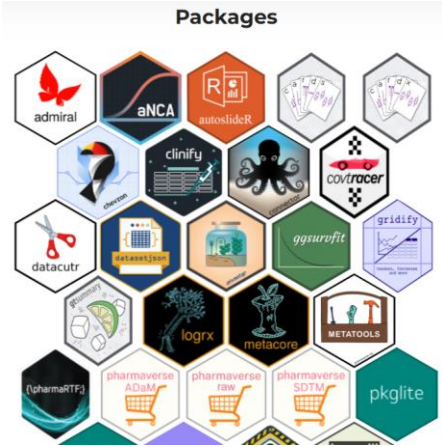


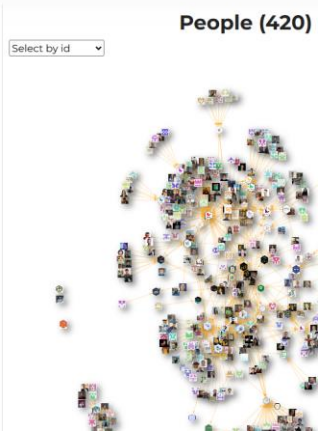


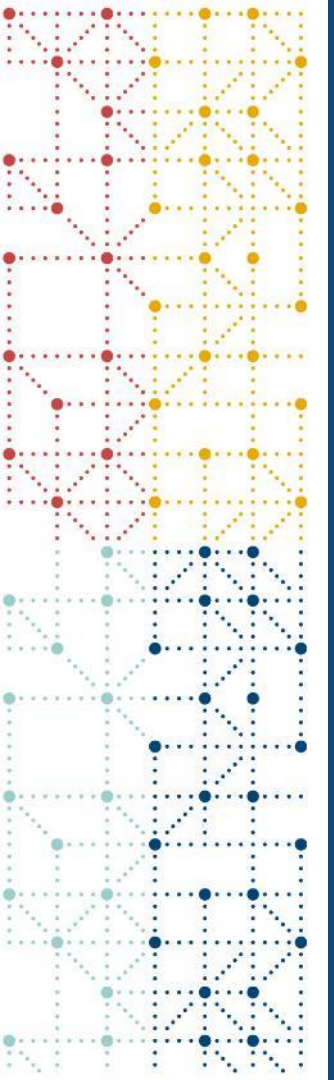
Open-Source Landscape

Open Source Landscape: Collaboration & Community

- People + Packages
  - Multi-lingual
- pharmaverse
  - sdtm.oak
  - metacore & metatools
  - admiral
  - xportr
- Foundational tools
  - tidyverse
  - git/github
  - shiny
- CDISC Open Source Alliance (COSA)

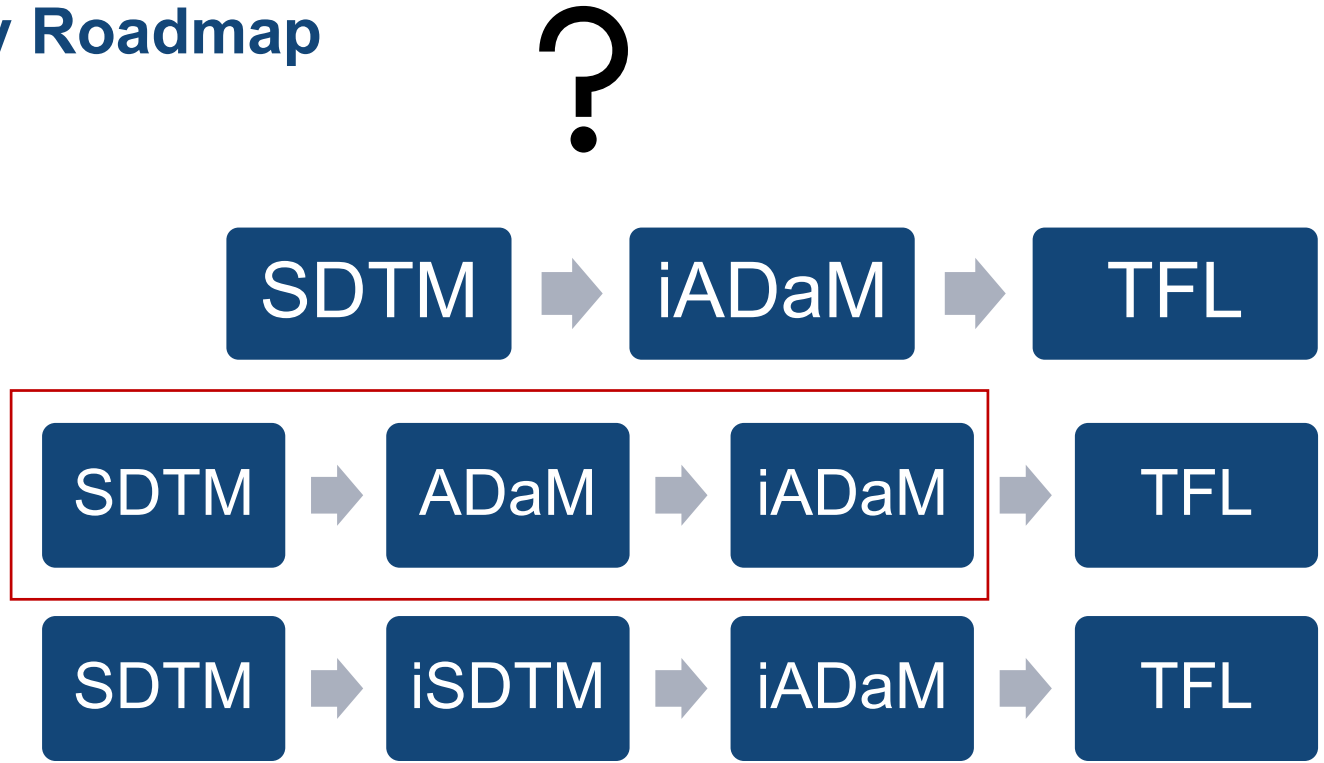






Case Study Walkthrough

Case Study Roadmap



SDTM {sdtm.oak}

- Consistent specifications
- Modular, metadata-driven SDTM builds
- ISS-friendly traceability & controlled terminology
- Framework for automation


```
assign_ct(
  raw_dat = ae_vars,
  raw_var = "AEOUT",
  tgt_var = "AEOUT",
  ct_spec = study_ct,
  ct_clst = "C66768",
  id_vars = oak_id_vars()
)%>%
```

term_synonyms	Name	codelist_code	Data Type	Order	term_value	term_code	collected_value
OUT	Outcome of Event	C66768	text	1	FATAL	C48275	Death Related to Adverse Event
OUT	Outcome of Event	C66768	text	2	NOT RECOVERED/NOT RESOLVED	C49494	Not Recovered or Not Resolved
OUT	Outcome of Event	C66768	text	3	RECOVERED/RESOLVED	C49498	Recovered or Resolved
OUT	Outcome of Event	C66768	text	4	RECOVERED/RESOLVED WITH SEQUELAE	C49495	Recovered or Resolved with Sequelae
OUT	Outcome of Event	C66768	text	5	RECOVERING/RESOLVING	C49496	Recovering or Resolving
OUT	Outcome of Event	C66768	text	6	UNKNOWN	C17998	Unknown

ADaM {admiral}

- Reusable derivation functions (ADSL, BDS, ADAE, etc.)
- Modular, extensible framework for study or pooled ADaM
- Traceable & reproducible across multiple studies
- Extensions: admiralonco, admiralpeds, & more

```
#### Derive TRTEMFL ####
adae <- adae %>%
  derive_var_trtemfl(
    start_date = ASTDT,
    end_date = AENDT,
    trt_start_date = TRTSDT,
    trt_end_date = TRTEDT
  )
```

Metadata {metacore & metatools}

- {metacore} stores variable attributes, derivation rules, and dependencies
- {metatools} uses the provided metadata to build/enhance and check the dataset
- Easy integration & harmonization

```
adae <- adae %>%
  drop_unspec_vars(metacore) %>% # Drop unspecified variables from specs
  check_variables(metacore) %>% # Check all variables specified are present and no more
  check_ct_data(metacore, na_acceptable = TRUE) %>% # Checks all variables with CT only contain values within the CT
  order_cols(metacore) %>% # Orders the columns according to the spec
  sort_by_key(metacore) %>% # Sorts the rows by the sort keys
  xportr_type(metacore, domain = "ADAE") %>% # Coerce variable type to match spec
  xportr_length(metacore, domain = "ADAE") %>% # Assigns SAS length from a variable level metadata
  xportr_label(metacore, domain = "ADAE") %>% # Assigns variable label from metacore specifications
  xportr_df_label(metacore, domain = "ADAE") %>% # Assigns dataset label from metacore specifications
```

Other Instruments – Datasets

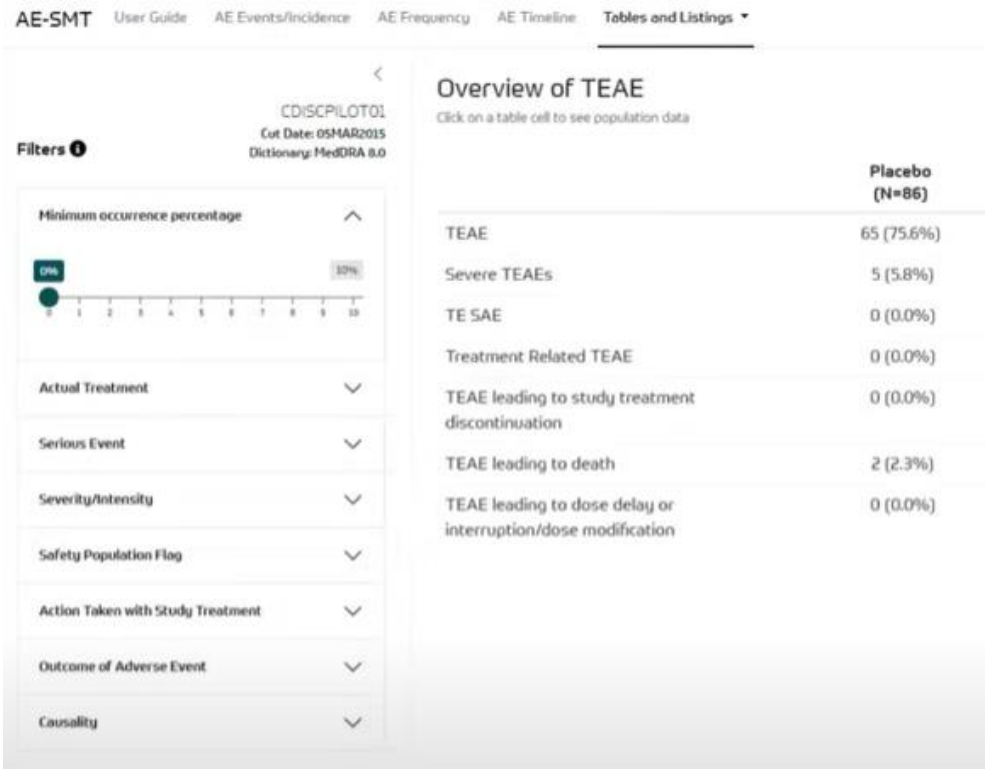
- Foundational helpers: {tidyverse} / {dplyr}
- Input, output & submissions:
  - {haven} – read in SAS datasets
  - {datacutr} - apply datacut to SDTM
  - {xportr} - CDISC-compliant XPT creation & validation
  - {datasetjson} – read & write CDISC Dataset JSON files
  - {logrx} - generate execution log
- QC:
  - {diffdf} - comparisons
  - {sdtmchecks} & CDISC CORE - compliance checks
  - {metacore} & {metatools} - don't forget they QC too!

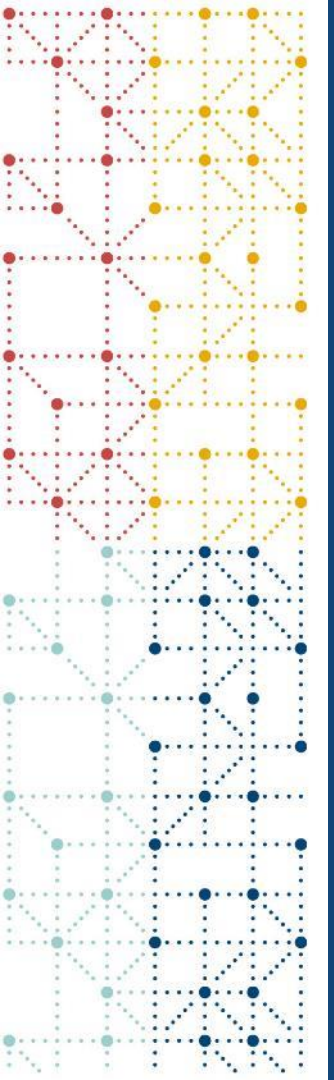
Other Instruments – TFL + more

- TLG-level tools:
  - {Tplyr}, {rtables}, {gtsummary} – tables
  - {cards}, {cardx} - metadata bridge (Analysis Results Dataset)
  - {clinify}, {pharmaRTF} - submission-ready formatting
  - {ggplot2} - produce graphs
  - {rlistings} - produce listings
- Utility/Workflow:
  - {pkglite} - represent & exchange R package source code as text files
  - {envsetup} - support setup of an R environment
  - {renv} - reproducibility
- +more in the pharmaverse!

Dynamic Exploration (Shiny)

- Explore pooled data interactively: drill down by study, treatment, subgroup, etc.
- Dynamic visualizations complement static TFLs
- Easily trace back to the source





Challenges, Benefits, & Future Directions

Challenges & Benefits of Open Source in ISS

Challenges

- Organizational Readiness
- Governance & Version Control
- Skills & Mindset
- Tool Selection & Integration

Benefits

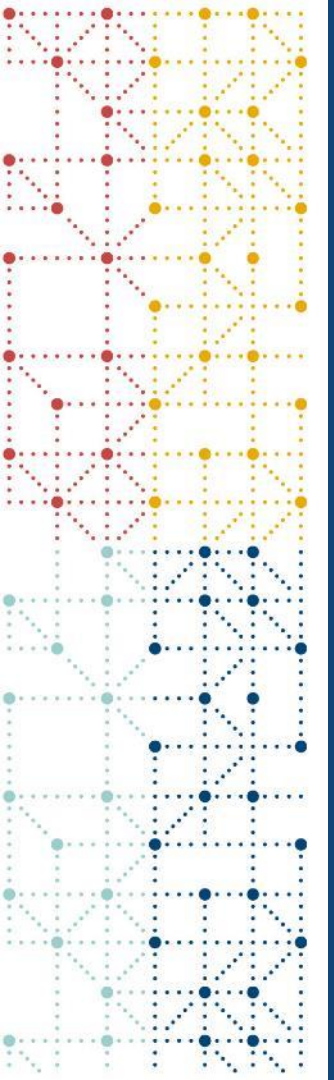
- Consistency & Transparency
- Automation & Reproducibility
- Collaboration
- Scalability
- Flexibility

Considerations

- Hybrid Approach Works
  - You don’t need an all-or-nothing migration; SAS + R can coexist.
- Assess Current State
  - Understand where your organization stands with open-source adoption.
- Governance & Environment Matter
  - Tool selection is important, but setup, compliance, and version control are critical.
- Start with High-Value Use Cases
  - Begin where the benefit is clearest, e.g., ISS dataset integration.
    - Core Dataset Stack – {admiral} + {metacore} + {metatools} provide a standardized foundation.
- Implementation Scenarios Vary
  - Choose the workflow that fits your situation:
    - ISS only in R → import, harmonize, and pool datasets
    - Study & ISS in R → reuse derivations and metadata across studies

Future Directions & Final Takeaways

- Increasing adoption
- Metadata-driven automation
- AI-assisted pipelines
- Clear ISS benefits
  - whether or not individual studies use R
- Engage with industry initiatives (pharmaverse, CDISC, etc.) for continued growth & impact



Thank you!