Paper DS09

CDISC Analysis Results Standard - Update and Progress

Karl Wallendszus, University of Oxford, Oxford, UK
Hansjörg Frenzel, ICON Plc, Mannheim, Germany

ABSTRACT

The CDISC Analysis Results Standard (ARS) Team is working towards creating standards to improve and facilitate the automation, reproducibility, reusability, and traceability of analysis results. The team is working towards this objective by creating an analysis results data model (ARDM) to support reuse of analysis results as well creating an analysis results metadata technical specification (ARM-TS) to support traceability, creation of analysis results, and creation of data displays. This paper provides an update of the team's progress and visualizes the development goals of the ARS team.

INTRODUCTION

Clinical trials generate many analysis results in the form of tables, figures and written reports, but they are rarely output in a form that is machine-readable, and there has been no standard way of describing and organizing these results. This makes it difficult to automate their generation, make them reproducible, trace their origin or enable them to be reused in other outputs. This was recognized by Chris Decker in a 2019 PHUSE paper [1]. He proposed separating the generation of the results themselves from the generation of the displays in which they are presented, and showed that this facilitates other uses of the same results, thus increasing efficiency.
One of us (HF) later described a practical implementation of some of these ideas [2]. That used a custom standard as current CDISC data standards do not cover this aspect of analysis results. A standard way of representing and describing analysis results and associated metadata is therefore desirable. In this paper we describe the progress of this project to date. Note that it is not yet a published standard and so should be considered work in progress.

CURRENT STATE

Standardization of data is widespread in the pharmaceutical industry for CRF data (using CDASH), tabulations (SDTM) and analysis datasets (ADaM), but the results of analyses are generally handled in a largely manual way. Typically, static sets of results are created for Clinical Study Reports (CSRs), publications and other such outputs. These entail a large number, perhaps hundreds, of tables and figures in a PDF document. This usually means they are not machine-readable, and are difficult to navigate even for humans. It is also difficult to trace particular results back to the ADaM data used to generate them, and even more difficult to trace them to further back in the pathway, e.g. to the study protocol or statistical analysis plan (SAP). For these reasons there is little opportunity to re-use a result in multiple outputs, making the results expensive to generate.

ARM 1.0

In 2015 CDISC published *Analysis Results Metadata (ARM) v1.0 for Define-XML v2.0* [3]. This provided a way to include some metadata about analysis results into Define.XML. It is not currently required for regulatory submissions but is highly recommended by the Pharmaceuticals and Medical Devices Agency (PMDA) in Japan. ARM includes high level descriptions of results with reference to the displays they appear in, but is not based on a formal model of analysis or results data. It is only expected to be provided for key analyses, and there are limited regulatory use cases. It is therefore limited in scope and lacks features which could be used to drive automation.

APPROACH

CDISC set up a multi-national team of volunteers drawn from the pharmaceutical industry, academic trial groups, regulators and software providers to develop a new Analysis Results Standard (ARS). Through a number of use cases, principle requirements for the standard were identified as automation, reproducibility, reusability and traceability.
The ARS team's goals were to:

1. Define an Analysis Results Metadata Technical Specification (ARM-TS), to support automation, traceability, and creation of data displays.
2. Define an Analysis Results Data (ARD) structure, to support reuse and reproducibility of results data.
3. Illustrate and exercise ARD and ARM-TS with a set of common data displays.

THE ANALYSIS RESULTS STANDARD

The Analysis Results Standard comprises two components: ARD and ARM-TS

ARD

The ARD model defines data structures to contain the result itself together with a label and a description of the statistical method, and separate elements describing the analysis, comparison and grouping variables, the source dataset, population and other selection criteria. Representative variables are shown in Figure 1.

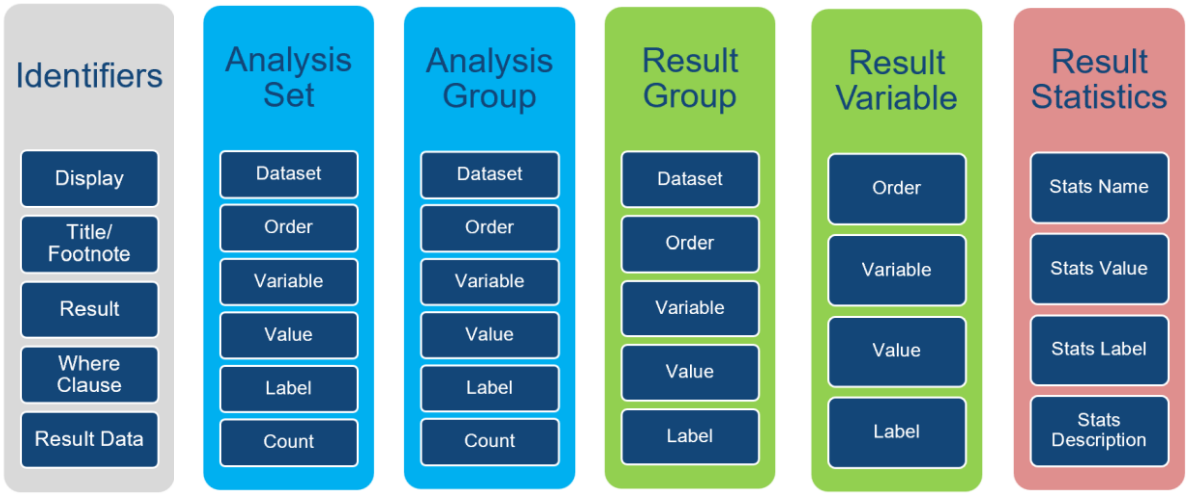

Figure 1: Key elements of the Analysis Results Dataset

ARM-TS

The ARM-TS includes separate, linked elements describing attributes of the ARD elements and displays, such as tables and figures. Because of the focus on automation, it serves a different purpose to the existing ARM in Define-XML. ARM-TS itself is technology- and tool-neutral. The extracts shown in the examples in this paper are represented in XML, but it will also be possible for ARM-TS to be expressed in JSON or other formats. Representative ARM-TS elements are shown in Figure 2.

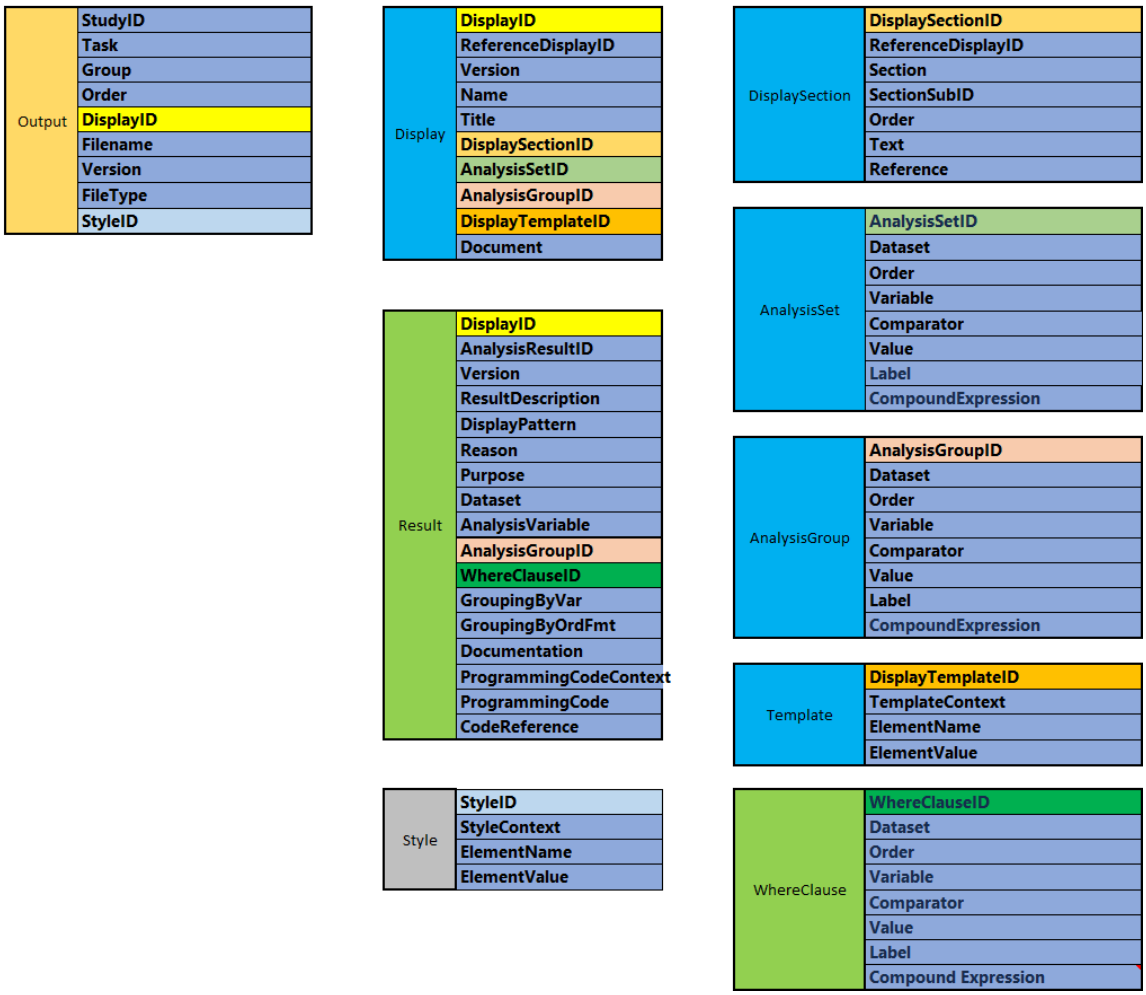

Figure 2: Key elements of the Analysis Results Metadata Technical Specification

EXAMPLES OF USAGE

The team's vision for ARS is for it to facilitate the automation, reproducibility, reusability and traceability of analysis results. The starting point for this is the relationship between ADaM datasets, ARM-TS and ARD. ARM-TS will be sufficiently granular for it to be able to drive at least the most common kinds of analysis through a meta-programming approach. That is, ARM-TS will contain all the information necessary to precisely specify the analyses, so a program can use this to run the relevant analyses, illustrated in figure 3.

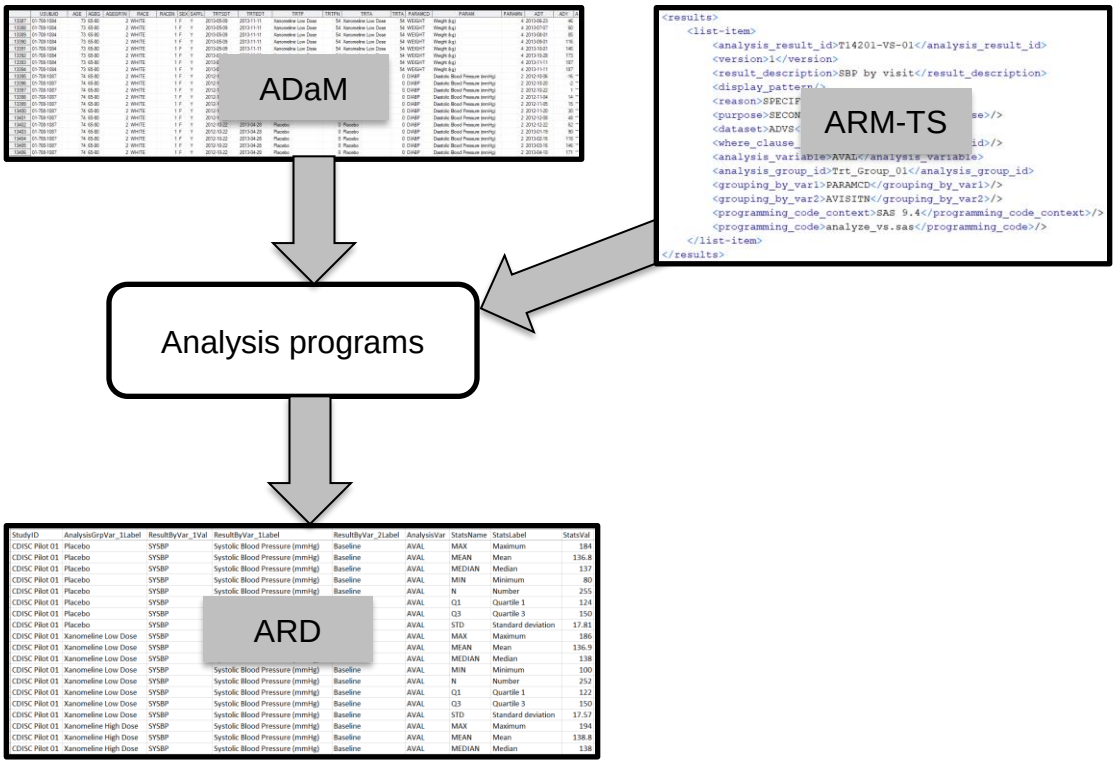

Figure 3: Workflow to produce ARD from ADaM and ARM-TS

AUTOMATION

The workflow described above already includes a substantial degree of automation, but we also envisage the standard enabling the automation of creating displays. ARM-TS includes elements describing attributes of displays which an automation engine can combine with results taken from the ARD to generate tables and figures. Figure 4 illustrates this process, starting with the workflow shown in figure 3..

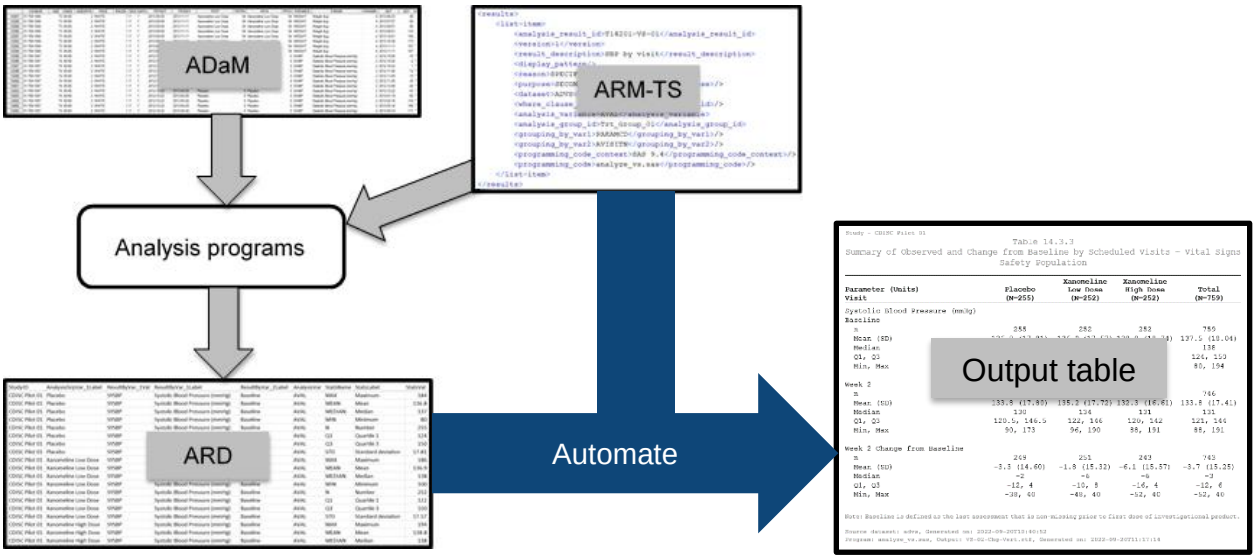

Figure 4: Illustration of how ARM-TS and ARD can be used to automate table generation

TRACEABILITY

The same elements of ARS which enable automatic generation of displays also enable traceability of the results shown, back to the ARD, and from there to ADaM and further back to SDTM and beyond through the use of existing CDISC standards and metadata. Figure 5 illustrates this.

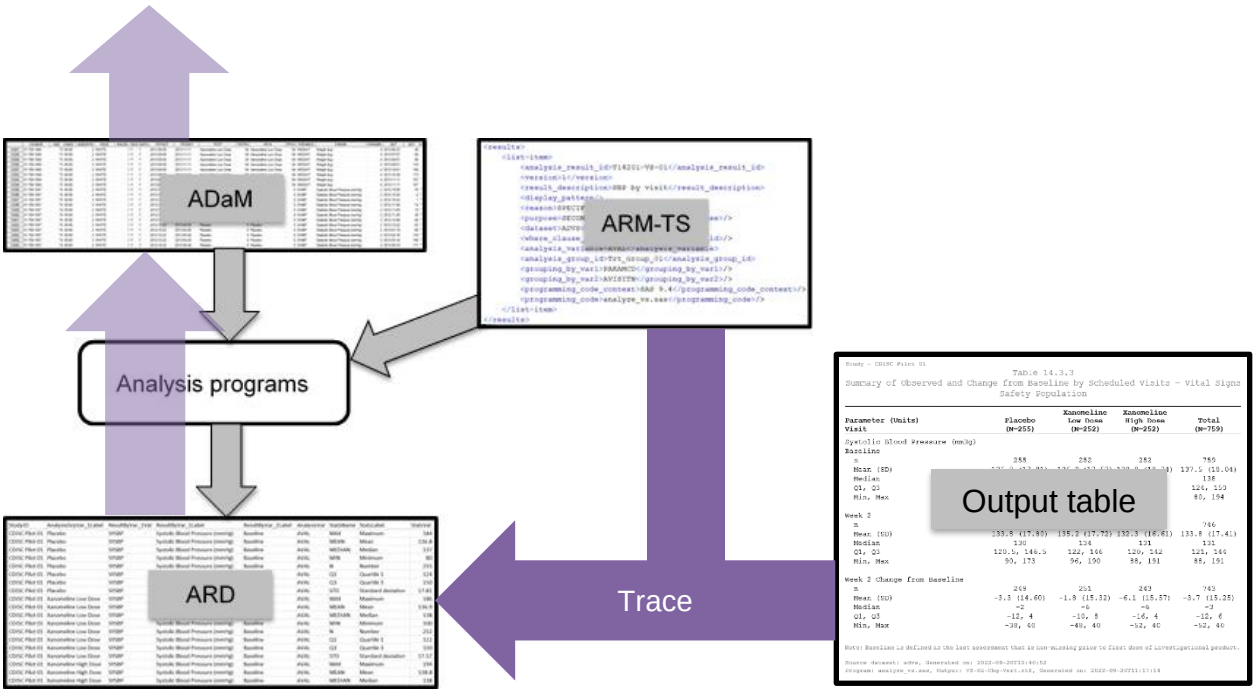

Figure 5: Illustration of how ARM-TS and ARD can be used to trace the origin of table contents

REPRODUCIBILITY

One advantage of using a standard for analysis results is that both inputs to and outputs from an analysis program are in a known, predictable format. So if another programmer writes a program to perform the same analysis using the same ADaM datasets, the resulting ARD should be the same, and because it is machine-readable, comparing outputs in this way can itself be automated, greatly improving the efficiency of validation processes.

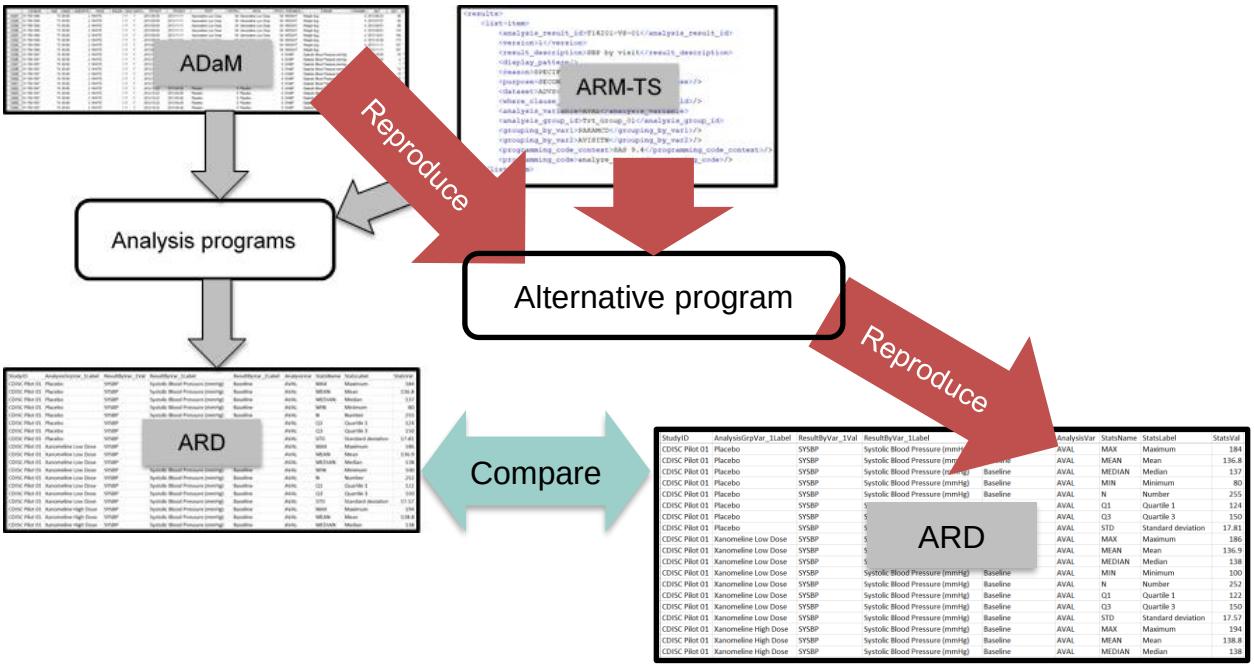

Figure 6: Illustration of how ARM-TS and ARD can be used to reproduce and validate analysis results.

Figure 6 shows how this can be done at the stage of generating analysis results, but similar process could be used to reproduce tables and figures from the ARD.

REUSABILITY

The analysis results in an ARD dataset, together with ARM-TS, are designed to be fully self-describing. This means that a result can be generated by the statistical analysis program once, and then used in multiple displays without any

loss in traceability. For example, the same results might be presented as both a table and a figure as shown in figure 7.

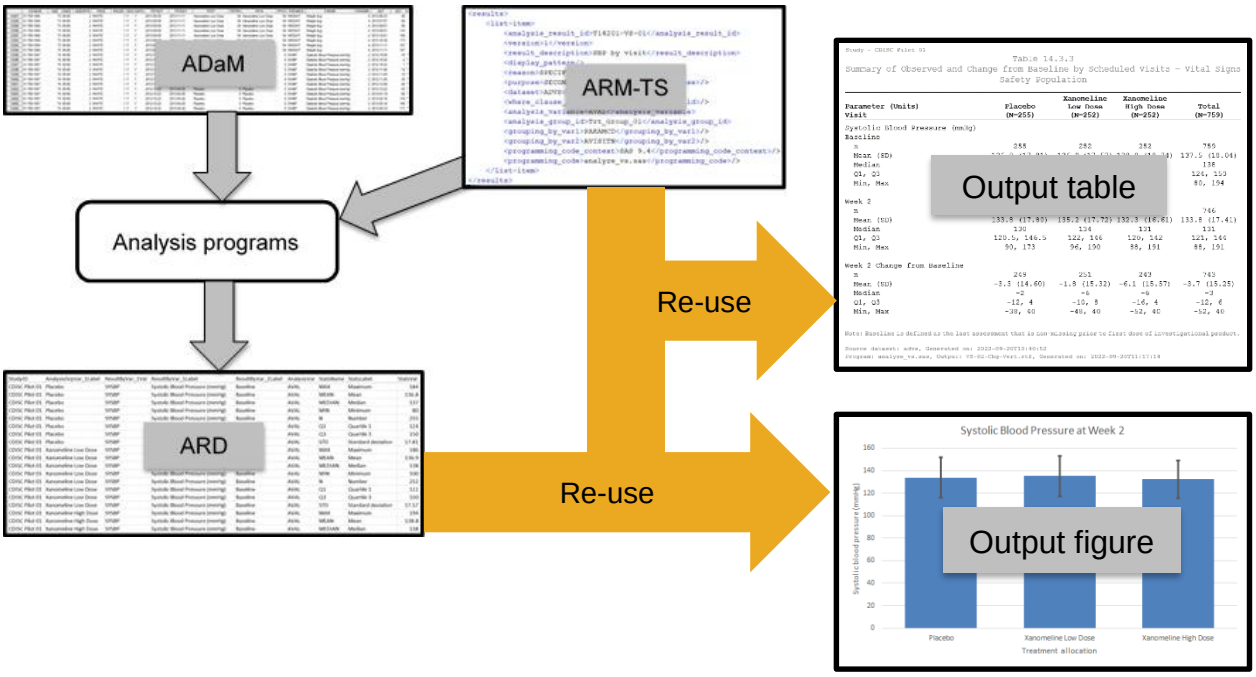

Figure 7: Illustration of how ARM-TS and ARD can be used to re-use results in multiple displays.

NEXT STEPS AND DELIVERABLES

We are currently working on formalizing the model and preparing for CDISC internal and public review processes. Our plan is for version 1.0 of the standard to include specifications of ARM-TS and ARM and an Implementation Guide, along with associated Controlled Terminology and identification of conformance rules. Examples will be provided which apply the standard to some common safety analyses. We anticipate that the first version will be released in 2023.

We hope that publication of the ARS will provide an impetus for software providers to build tools which use the standard in analysis and output generation, and that this might include some open source projects. Indeed, some existing open source tools, such as TFL Designer [4] are well suited to use ARS once it is available.

CONCLUSION

The conclusion summarises your paper and ties together any loose ends. You can use the conclusion to make any final points such as recommendations predictions, or judgments.

REFERENCES

1. Decker, C. 2019. Analysis Results that Save Trees - #KillTFLs. Paper DH12, PHUSE US Connect 2019
2. Frenzel, H. 2021. Untangling the Knot - Implementing Analysis Results Standards (Using SAS®). Paper DS-130, PharmaSUG 2021
3. CDISC. 2015. Analysis Results Metadata (ARM) v1.0 for Define-XML v2.0
4. Busa, B. TFL Designer. https://github.com/bhavinbusa/tfldesigner

ACKNOWLEDGMENTS

This work has only been possible through the mostly voluntary contributions from our fellow members of the CDISC Analysis Results Standard team: Jeff Abolafia, Ramya Ambidi, Nancy Brucken, Bhavin Busa, Sally Cassells, Cedric Davister, Dante Di Tommaso, Pierre Dostie, Wenwen Fang, Nate Freimark, Prafulla Girase, Brian Harris, Xie Kang, Bess LeRoy, Kent Letourneau, Andrew Miskell, Taniya Muliyil, Rudra Sheth, Tatiana Sotingco, Edwin van Stein, Alana St Clair, Nicole Thorne, Wei Wang. We thank them all.

CONTACT INFORMATION

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

Karl Wallendszus
Oxford Population Health, University of Oxford, UK
Email: karl.wallendszus@ndph.ox.ac.uk

Web: https://www.ndph.ox.ac.uk/team/karl-wallendszus