



# Explore in CDISC ARS by a Linked Data lover

Ippei Akiya

ICON Clinical Research GK, Biometrics

---

# Disclaimer

- The content of this presentation includes my personal research, reflections, and opinions, and has no connection whatsoever with the intellectual property rights or business activities of the company I belong to or its clients.
- Furthermore, it does not represent the views or opinions of any academic societies or organizations I am affiliated with.
- The copyrights and intellectual property rights for the software and other products introduced in this presentation belong to their respective owners.

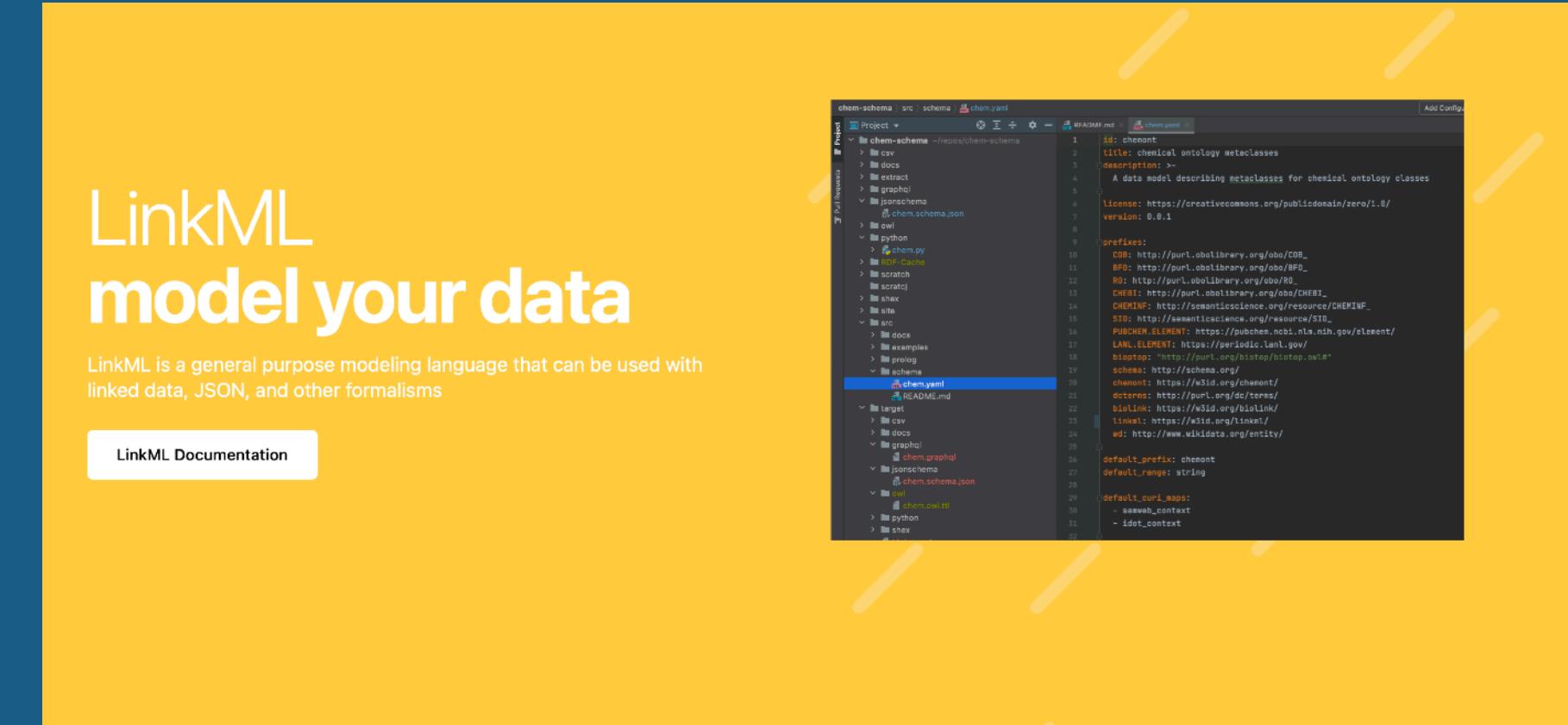
---

# Introduction

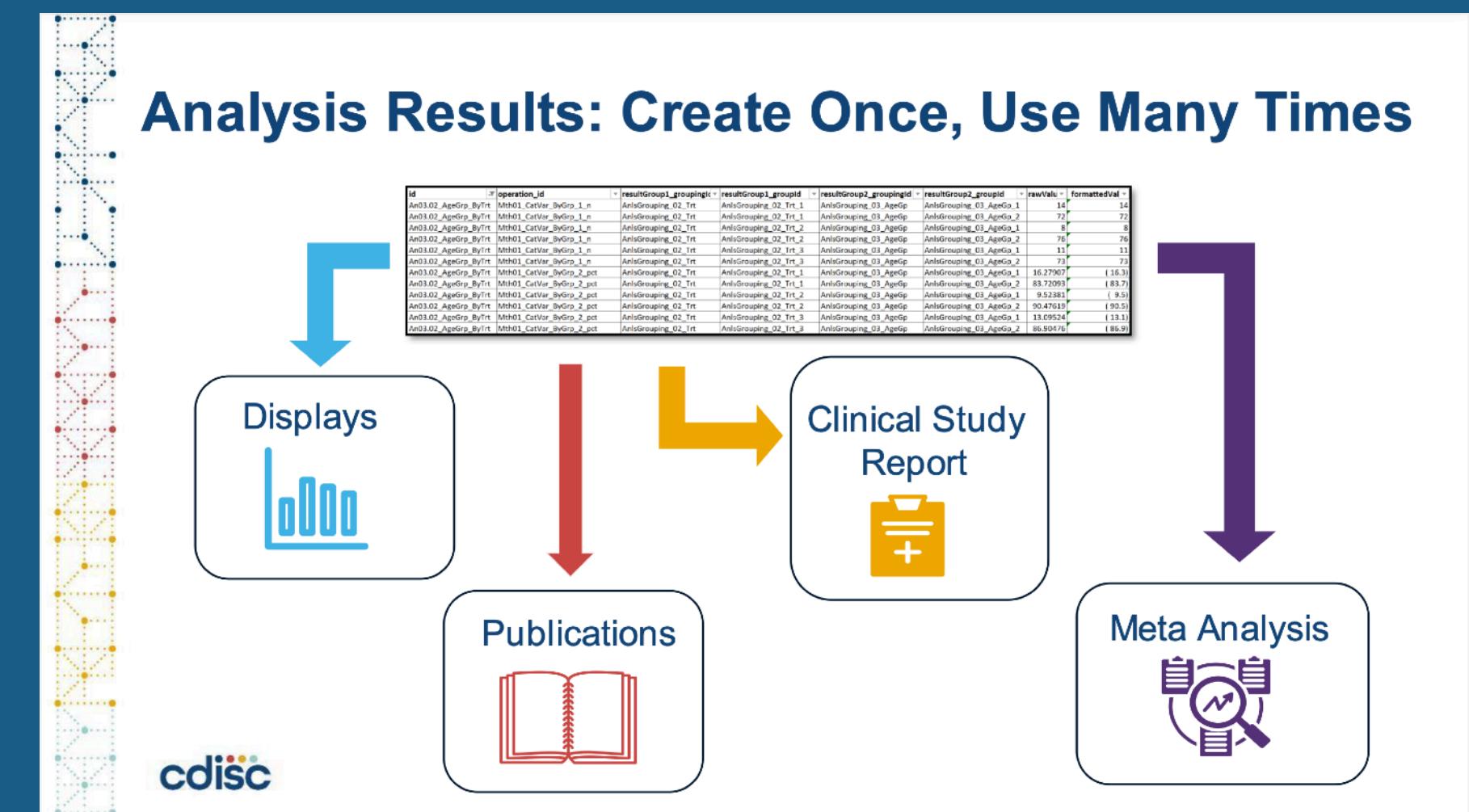
- This is an explanation of CDISC Analysis Results Standard v1.0 (ARS) from the perspective of a Linked Data enthusiast.

- The reason for explaining it from a Linked Data enthusiast's perspective is that ARS uses LinkML, a Linked Data Modeling Language.

- A highly useful guidebook for understanding ARS is Getting Started with the New CDISC Analysis Results Standard by Bess LeRoy and Richard Marshall.



<https://linkml.io/>



[https://www.cdisc.org/sites/default/files/2024-04/CDISC EU 2024 ARS 20240419.pdf](https://www.cdisc.org/sites/default/files/2024-04/CDISC%20EU%202024%20ARS%2020240419.pdf)

---

# Overview of CDISC ARS

# Analysis Results Key Objectives



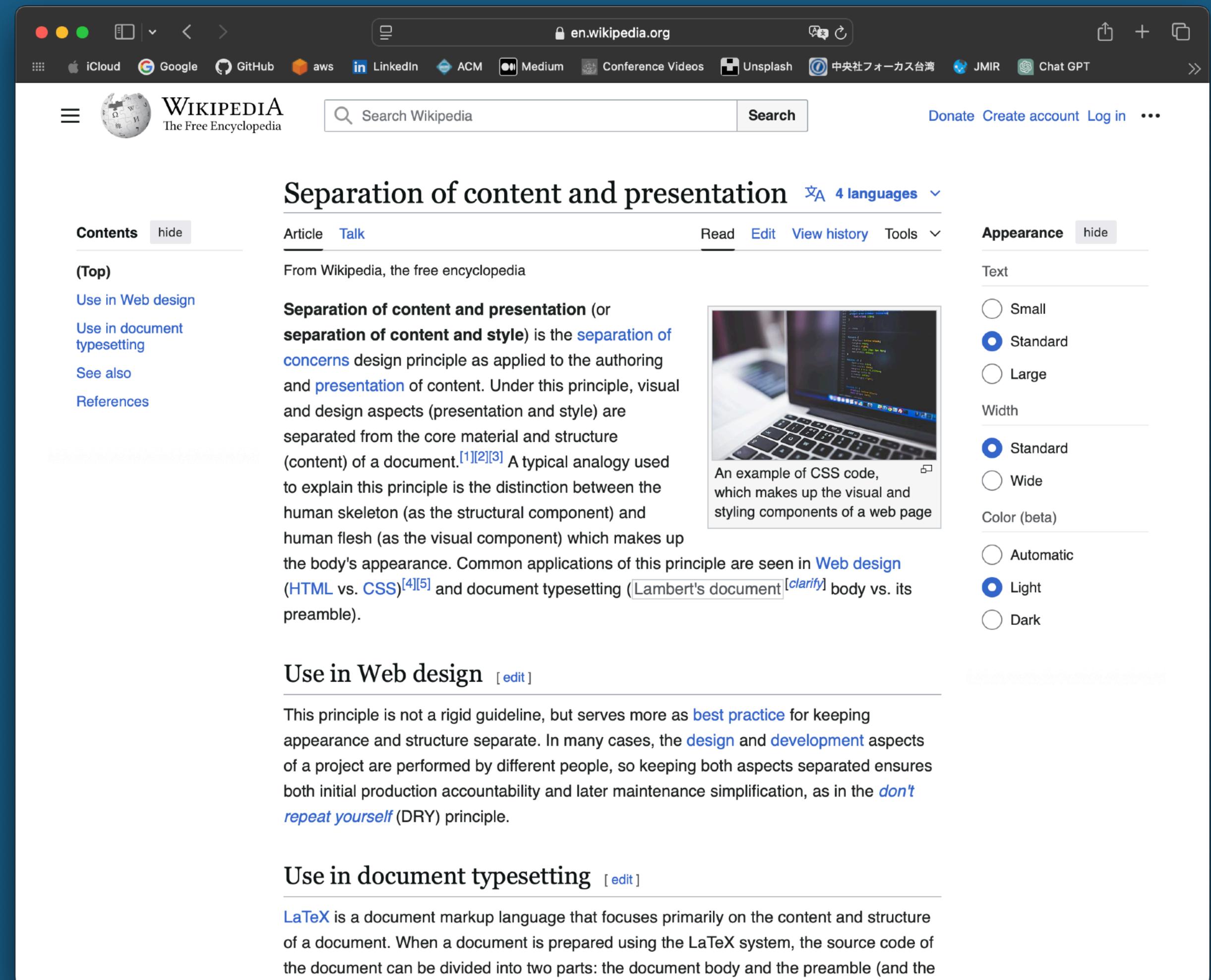
Leverage analysis results metadata to drive the automation of results



Support storage, access, processing, traceability and reproducibility of results

# Why CDISC ARS is developed?

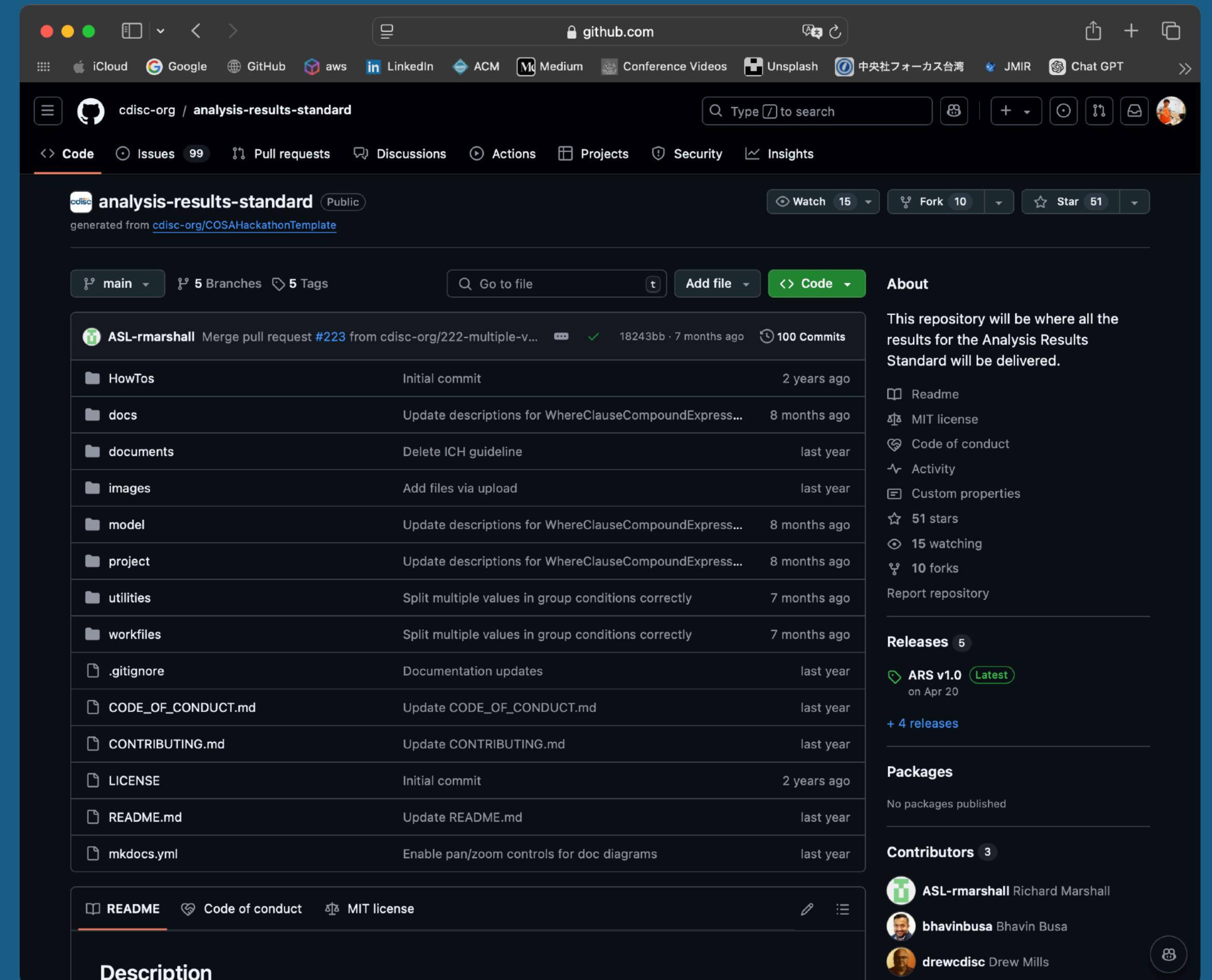
- Currently, the analysis results (TFLs) of clinical trials are output as files, such as RTF or PDF, through analysis programs.
- These files do not separate the results data or content from styles such as fonts and borders, making it difficult to extract the results data as literal values.
- As a result, it becomes challenging to reuse the result data.
- However, Define-XML serves as an example where content and style are completely separated.
  - `define.xml <- Contents`
  - `define.xslt <- Style`



The screenshot shows a web browser displaying the Wikipedia article 'Separation of content and presentation'. The page title is 'Separation of content and presentation' with a '4 languages' link. The main content area includes a sidebar for 'Appearance' (Text, Font Size: Standard, Width: Standard, Color: Light). The main text discusses the principle of separating content from presentation, using the analogy of a human skeleton and flesh. It mentions applications in web design (HTML vs. CSS) and document typesetting (LaTeX). A sidebar on the right shows a laptop displaying code.

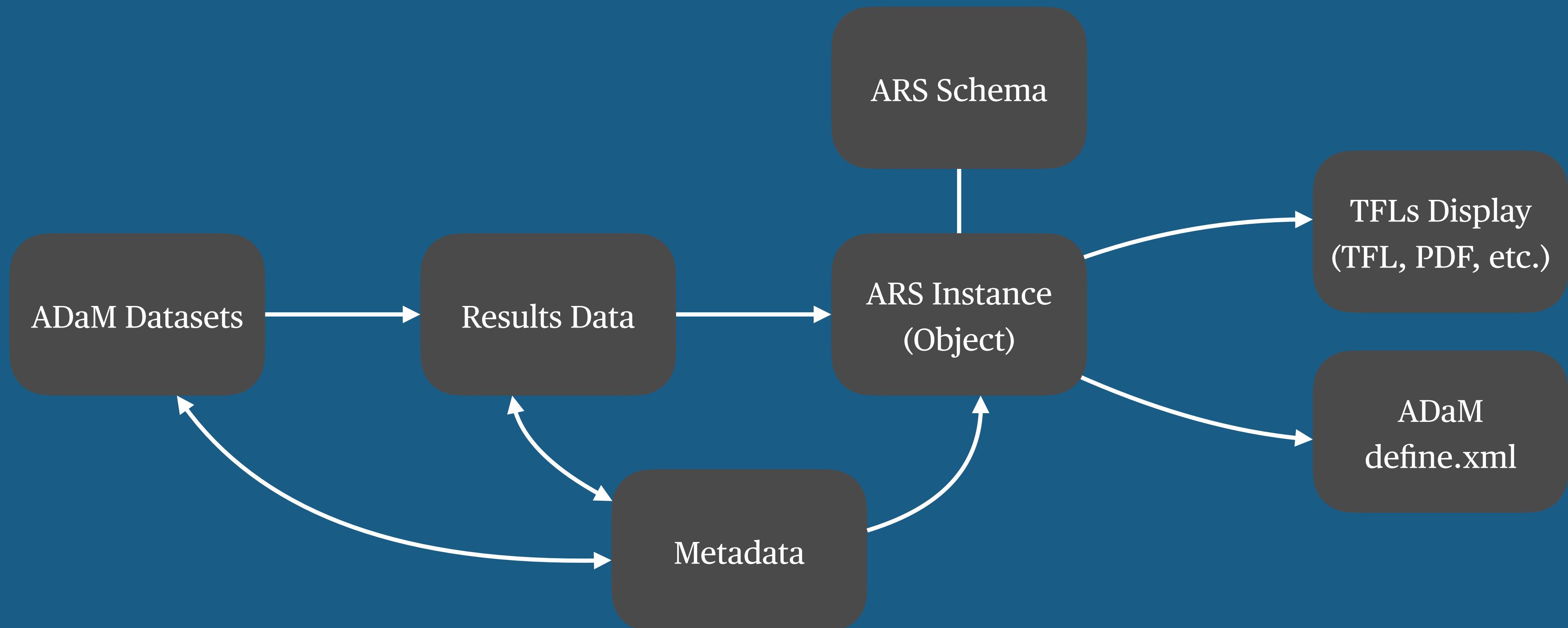
# Why CDISC ARS is developed?

- The following are my assumptions.
  - It becomes easier to extract data or results data for conducting meta-analysis, enabling the accurate selection of target trials.
- Utilization by healthcare authorities during approval reviews.
- As an advanced topic, converting not only results data but also SDTM and ADaM data into Linked Data would make it easier to confirm data traceability. This would enhance the transparency of statistical analyses.



<https://github.com/cdisc-org/analysis-results-standard>

# Workflow



# Mapping Results Data from ARS Instance to TFLs

```

- name: Summary of Age by Treatment
  label: Age
- operationId: Mth02_ContVar_Summ_ByGrp_2_Mean
  resultGroups:
    - groupingId: AnlsGrouping_01_Trt
      groupId: AnlsGrouping_01_Trt_1
      rawValue: '75.2093023'
      formattedValue: '75.2'
    - groupingId: AnlsGrouping_01_Trt
      groupId: AnlsGrouping_01_Trt_2
      rawValue: '75.6666667'
      formattedValue: '75.7'
    - groupingId: AnlsGrouping_01_Trt
      groupId: AnlsGrouping_01_Trt_3
      rawValue: '74.3809524'
      formattedValue: '74.4'
    - operationId: Mth02_ContVar_Summ_ByGrp_3_SD
      resultGroups:
        - groupingId: AnlsGrouping_01_Trt
          groupId: AnlsGrouping_01_Trt_1
          rawValue: '8.5901671'
          formattedValue: '( 8.59)'

```

Study - CDISC 360

Table 14.1.1<sup>2</sup>  
Summary of Demographics ↗  
Safety Population ↗

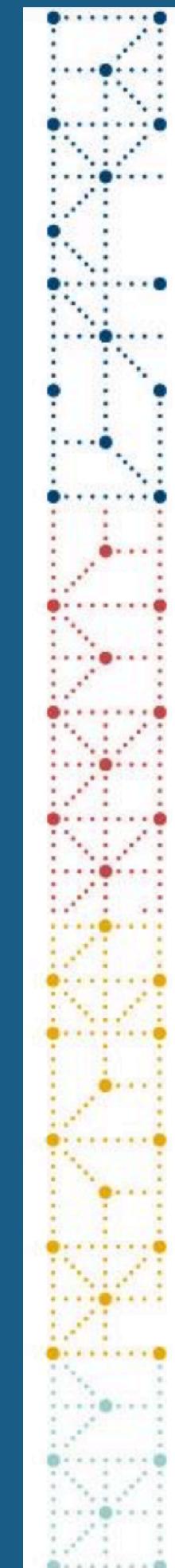
Page x of y ↗

Characteristics	Placebo ↓ (N=XX) ↗	Xanomeline ↓ Low Dose ↓ (N=XX) ↗	Xanomeline ↓ High Dose ↓ (N=XX) ↗	p-value [1] ↗
Age (years) ↗	XX ↗	XX ↗	XX ↗	x.XXXX ↗
n ↗	XX.X (XX.XX) ↗	XX.X (XX.XX) ↗	XX.X (XX.XX) ↗	
Mean (SD) ↗	XX.X, XX.X ↗	XX.X, XX.X ↗	XX.X, XX.X ↗	
Median ↗	XX, XX ↗	XX, XX ↗	XX, XX ↗	
Q1, Q3 ↗	XX.X, XX.X ↗	XX.X, XX.X ↗	XX.X, XX.X ↗	
Min, Max ↗	XX, XX ↗	XX, XX ↗	XX, XX ↗	
Age Group, n (%) ↗	XX ( XX.X ) ↗	XX ( XX.X ) ↗	XX ( XX.X ) ↗	x.XXXX ↗
< 65 years ↗	XX ( XX.X ) ↗	XX ( XX.X ) ↗	XX ( XX.X ) ↗	
≥ 65 years ↗	XX ( XX.X ) ↗	XX ( XX.X ) ↗	XX ( XX.X ) ↗	
Gender, n (%) ↗	XX ( XX.X ) ↗	XX ( XX.X ) ↗	XX ( XX.X ) ↗	x.XXXX ↗
Male ↗	XX ( XX.X ) ↗	XX ( XX.X ) ↗	XX ( XX.X ) ↗	
Female ↗	XX ( XX.X ) ↗	XX ( XX.X ) ↗	XX ( XX.X ) ↗	
Ethnicity, n (%) ↗	XX ( XX.X ) ↗	XX ( XX.X ) ↗	XX ( XX.X ) ↗	x.XXXX ↗
Hispanic or Latino ↗	XX ( XX.X ) ↗	XX ( XX.X ) ↗	XX ( XX.X ) ↗	
Not Hispanic or Latino ↗	XX ( XX.X ) ↗	XX ( XX.X ) ↗	XX ( XX.X ) ↗	

[https://github.com/cdisc-org/analysis-results-standard/blob/main/workfiles/examples/ARS\\_v1/Demog\\_Table\\_Shell.rtf](https://github.com/cdisc-org/analysis-results-standard/blob/main/workfiles/examples/ARS_v1/Demog_Table_Shell.rtf)

# TFL DESIGNER

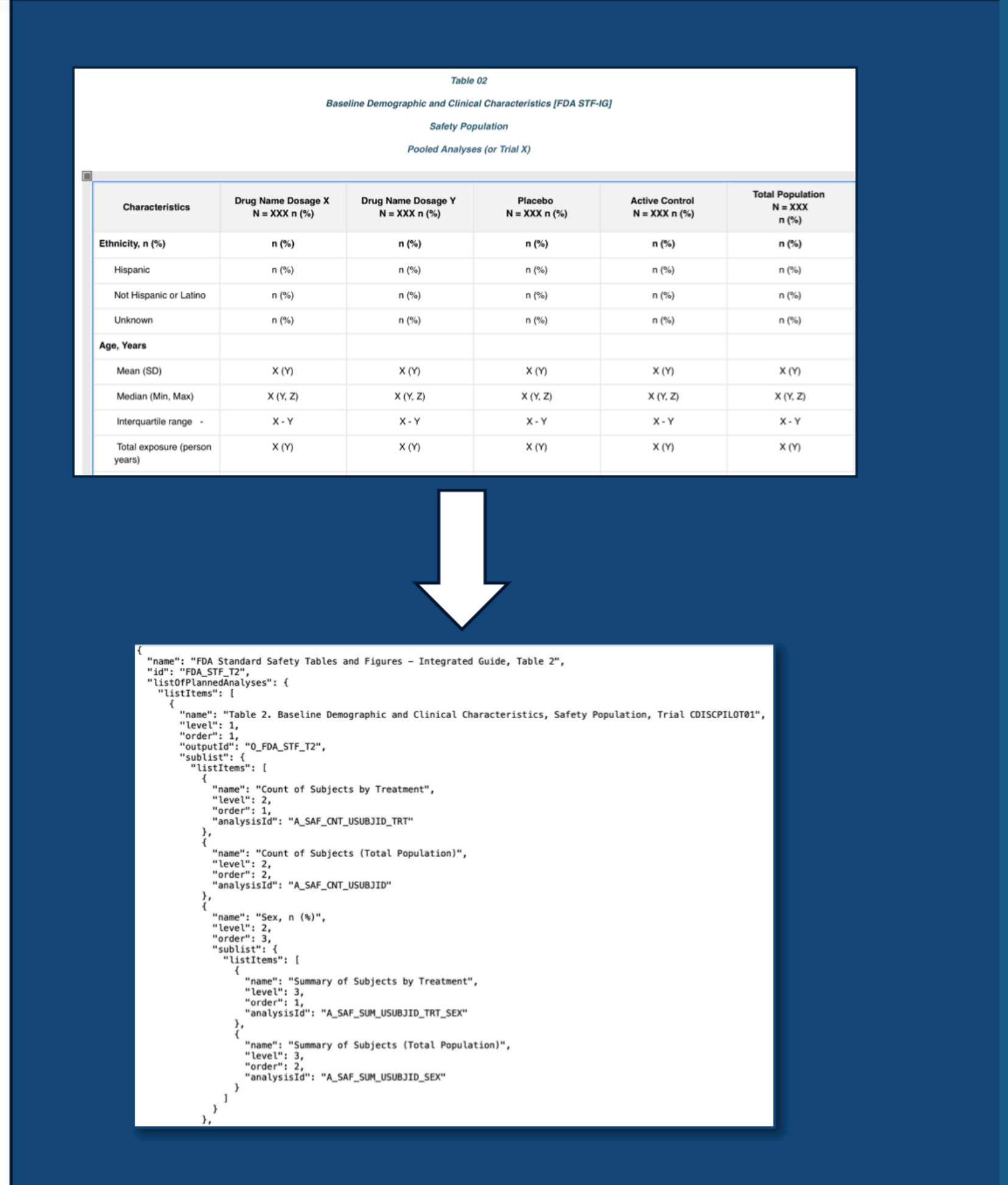
- Defining the design of TFLs in text can be a tedious task, so a visual editor called TFL DESIGNER has been developed to alleviate this burden.



- A community tool to create study specific analysis displays
- Export analysis results metadata per the CDISC ARS model in JSON and Excel formats

cdisc

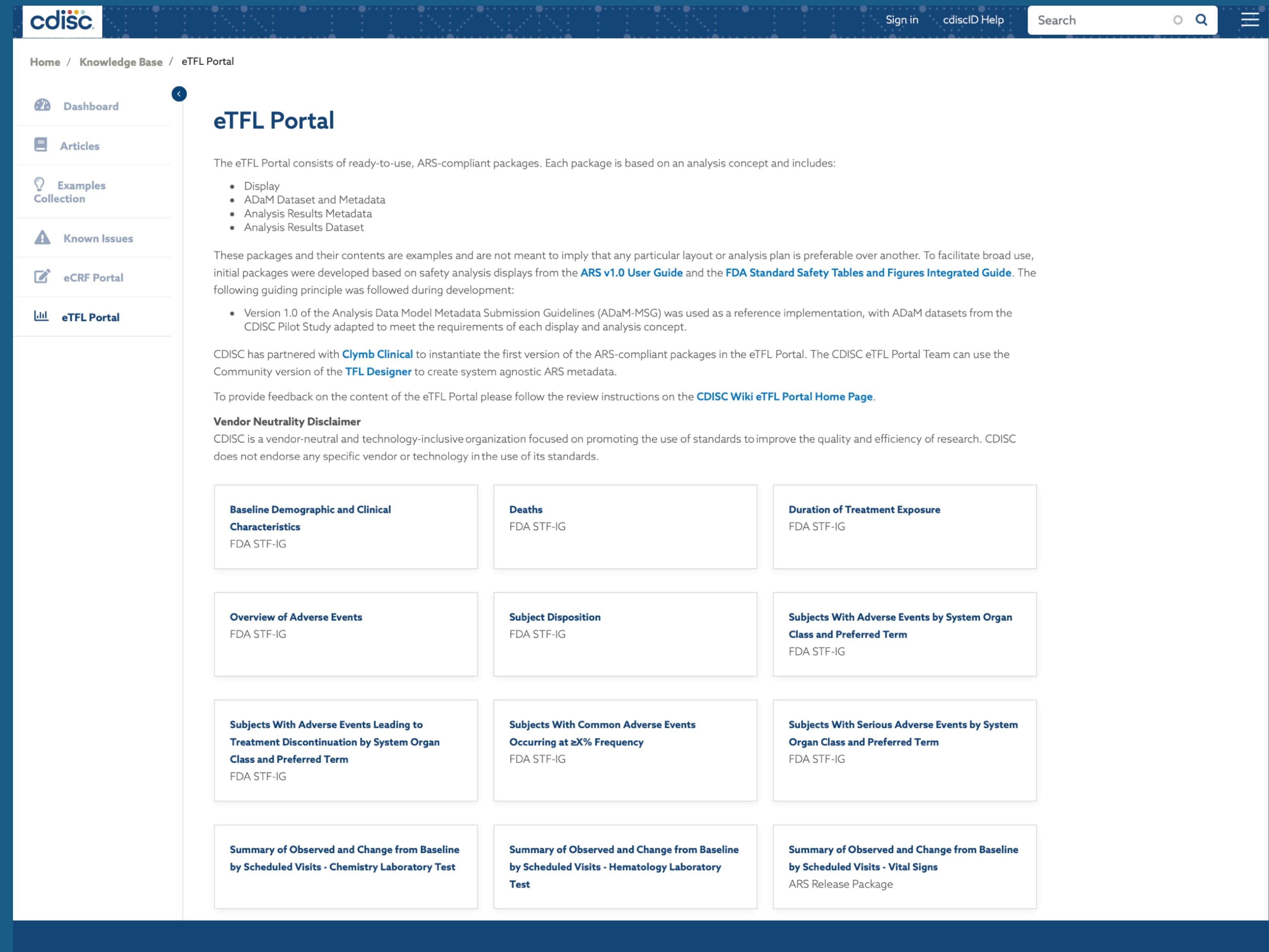
Source: Bhavin Busa, Clymb Clinical;  
<https://tfldesigner.org/>



# eTFL Portal

<https://www.cdisc.org/kb/etfl>

- Various samples of TFLs are available on the CDISC eTFL Portal.
- By referring to these samples, it should also be possible to define ARS-format TFLs in text.
- An AI trained on these samples might be able to generate ARS-format TFLs (Shells) with simple prompt inputs.



The eTFL Portal consists of ready-to-use, ARS-compliant packages. Each package is based on an analysis concept and includes:

- Display
- ADaM Dataset and Metadata
- Analysis Results Metadata
- Analysis Results Dataset

These packages and their contents are examples and are not meant to imply that any particular layout or analysis plan is preferable over another. To facilitate broad use, initial packages were developed based on safety analysis displays from the [ARS v1.0 User Guide](#) and the [FDA Standard Safety Tables and Figures Integrated Guide](#). The following guiding principle was followed during development:

- Version 1.0 of the Analysis Data Model Metadata Submission Guidelines (ADaM-MSG) was used as a reference implementation, with ADaM datasets from the CDISC Pilot Study adapted to meet the requirements of each display and analysis concept.

CDISC has partnered with [Clymb Clinical](#) to instantiate the first version of the ARS-compliant packages in the eTFL Portal. The CDISC eTFL Portal Team can use the Community version of the [TFL Designer](#) to create system agnostic ARS metadata.

To provide feedback on the content of the eTFL Portal please follow the review instructions on the [CDISC Wiki eTFL Portal Home Page](#).

**Vendor Neutrality Disclaimer**  
CDISC is a vendor-neutral and technology-inclusive organization focused on promoting the use of standards to improve the quality and efficiency of research. CDISC does not endorse any specific vendor or technology in the use of its standards.

Baseline Demographic and Clinical Characteristics FDA STF-IG	Deaths FDA STF-IG	Duration of Treatment Exposure FDA STF-IG
Overview of Adverse Events FDA STF-IG	Subject Disposition FDA STF-IG	Subjects With Adverse Events by System Organ Class and Preferred Term FDA STF-IG
Subjects With Adverse Events Leading to Treatment Discontinuation by System Organ Class and Preferred Term FDA STF-IG	Subjects With Common Adverse Events Occurring at $\geq X\%$ Frequency FDA STF-IG	Subjects With Serious Adverse Events by System Organ Class and Preferred Term FDA STF-IG
Summary of Observed and Change from Baseline by Scheduled Visits - Chemistry Laboratory Test	Summary of Observed and Change from Baseline by Scheduled Visits - Hematology Laboratory Test	Summary of Observed and Change from Baseline by Scheduled Visits - Vital Signs ARS Release Package

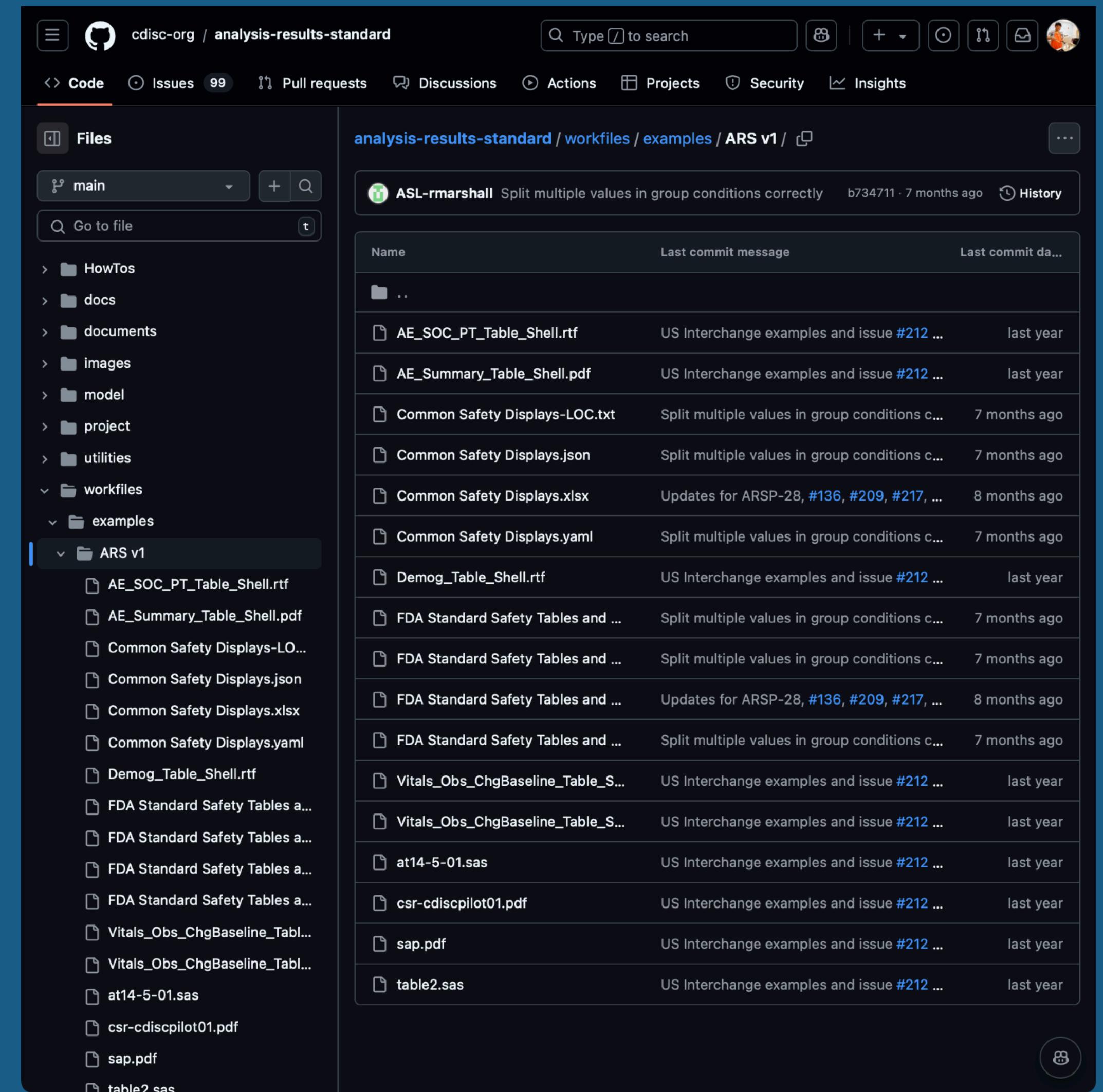
<https://www.cdisc.org/kb/etfl>

---

# How to make TFLs with CDISC ARS?

# CDISC ARS Github Repository

- Examples are stored in the CDISC ARS GitHub Repository.  
<https://github.com/cdisc-org/analysis-results-standard/tree/main/workfiles/examples/ARS v1>

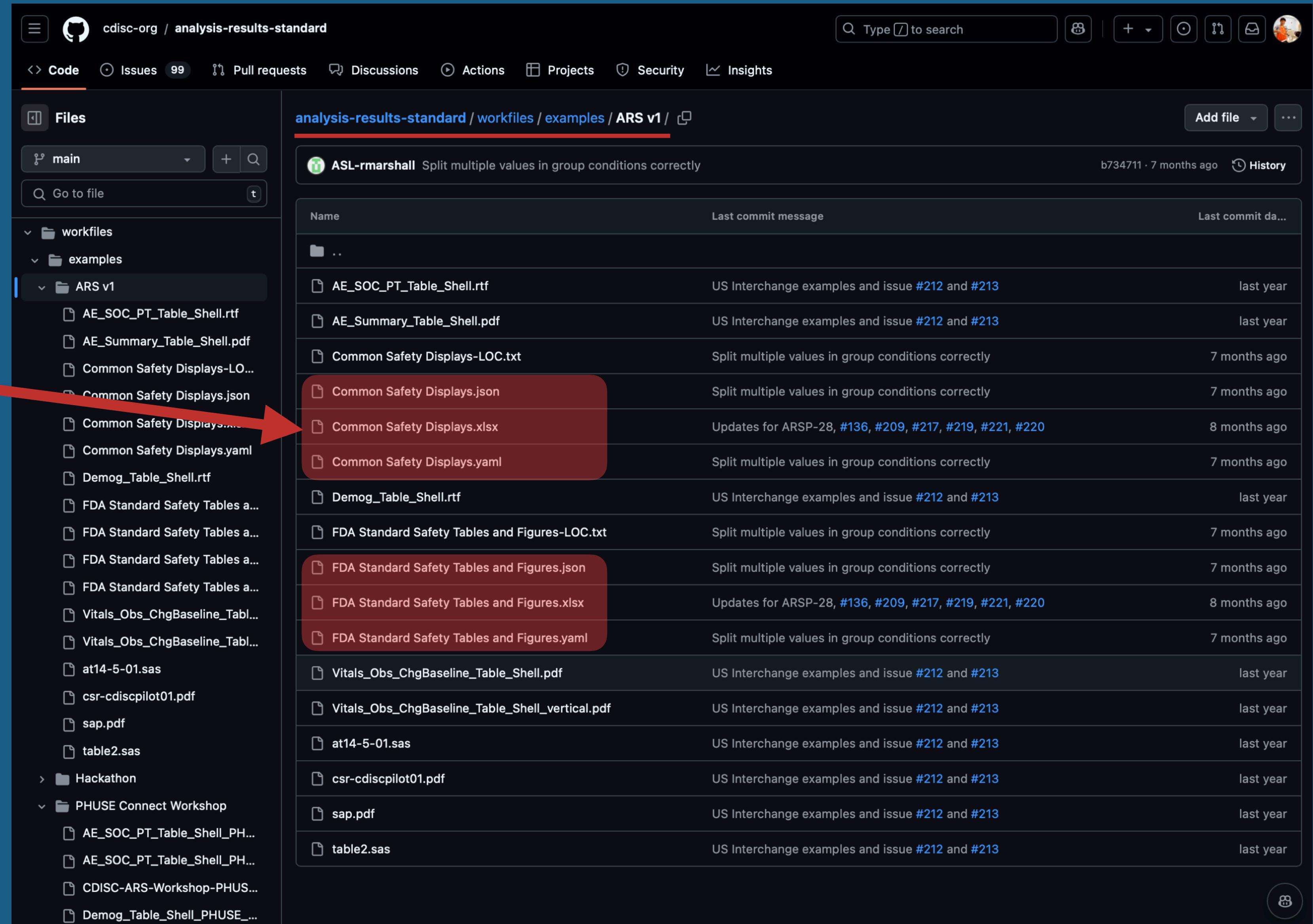


The screenshot shows a GitHub repository page for 'cdisc-org / analysis-results-standard'. The 'Code' tab is selected. On the left, the file structure is shown with a tree view. The 'examples' folder under 'workfiles' is expanded, and its subfolder 'ARS v1' is also expanded, showing various files like 'AE\_SOC\_PT\_Table\_Shell.rtf', 'AE\_Summary\_Table\_Shell.pdf', etc. On the right, a list of files is displayed with their names, last commit messages, and last commit dates. A specific commit by 'ASL-rmarshall' is highlighted, showing a commit message: 'Split multiple values in group conditions correctly' made 7 months ago.

Name	Last commit message	Last commit date
...		
AE_SOC_PT_Table_Shell.rtf	US Interchange examples and issue #212 ...	last year
AE_Summary_Table_Shell.pdf	US Interchange examples and issue #212 ...	last year
Common Safety Displays-LOC.txt	Split multiple values in group conditions c...	7 months ago
Common Safety Displays.json	Split multiple values in group conditions c...	7 months ago
Common Safety Displays.xlsx	Updates for ARSP-28, #136, #209, #217, ...	8 months ago
Common Safety Displays.yaml	Split multiple values in group conditions c...	7 months ago
Demog_Table_Shell.rtf	US Interchange examples and issue #212 ...	last year
FDA Standard Safety Tables and ...	Split multiple values in group conditions c...	7 months ago
FDA Standard Safety Tables and ...	Split multiple values in group conditions c...	7 months ago
FDA Standard Safety Tables and ...	Updates for ARSP-28, #136, #209, #217, ...	8 months ago
FDA Standard Safety Tables and ...	Split multiple values in group conditions c...	7 months ago
Vitals_Obs_ChgBaseline_Table_S...	US Interchange examples and issue #212 ...	last year
Vitals_Obs_ChgBaseline_Table_S...	US Interchange examples and issue #212 ...	last year
at14-5-01.sas	US Interchange examples and issue #212 ...	last year
csr-cdiscpilot01.pdf	US Interchange examples and issue #212 ...	last year
sap.pdf	US Interchange examples and issue #212 ...	last year
table2.sas	US Interchange examples and issue #212 ...	last year

# Work Process in CDISC ARS

- The repository contains materials from workshops held by CDISC and PHUSE.
- One of these files can be created, and TFLs (RTF, PDF, etc.) can be generated from it.
- Programs or software are required to create ARS data files.
- Additionally, programs or software are needed to generate TFLs from ARS data files.



The screenshot shows a GitHub repository interface for 'analysis-results-standard'. The left sidebar shows a tree view of files under 'main', with 'workfiles' and 'examples' expanded. The 'examples' folder contains a 'ARS v1' folder, which is also expanded, showing files like 'AE\_SOC\_PT\_Table\_Shell.rtf', 'AE\_Summary\_Table\_Shell.pdf', 'Common Safety Displays-LOC.txt', 'Common Safety Displays.json', 'Common Safety Displays.xlsx', 'Common Safety Displays.yaml', 'Demog\_Table\_Shell.rtf', 'FDA Standard Safety Tables a...', 'Vitals\_Obs\_ChgBaseline\_Table...', 'Vitals\_Obs\_ChgBaseline\_Table...', 'at14-5-01.sas', 'csr-cdiscpilot01.pdf', 'sap.pdf', 'table2.sas', 'Hackathon', 'PHUSE Connect Workshop', 'AE\_SOC\_PT\_Table\_Shell\_PH...', 'AE\_SOC\_PT\_Table\_Shell\_PH...', 'CDISC-ARS-Workshop-PHUS...', and 'Demog\_Table\_Shell\_PHUSE...'. A red arrow points from the 'Common Safety Displays.json' file in the 'ARS v1' folder to the same file in the list of files on the right. Both files are highlighted with a red box. The right side shows a list of files in the 'analysis-results-standard / workfiles / examples / ARS v1 /' directory, including 'Common Safety Displays.json', 'Common Safety Displays.xlsx', 'Common Safety Displays.yaml', 'Demog\_Table\_Shell.rtf', 'FDA Standard Safety Tables and Figures-LOC.txt', 'FDA Standard Safety Tables and Figures.json', 'FDA Standard Safety Tables and Figures.xlsx', 'FDA Standard Safety Tables and Figures.yaml', 'Vitals\_Obs\_ChgBaseline\_Table\_Shell.pdf', 'Vitals\_Obs\_ChgBaseline\_Table\_Shell\_vertical.pdf', 'at14-5-01.sas', 'csr-cdiscpilot01.pdf', 'sap.pdf', and 'table2.sas'. The 'Common Safety Displays.json' file is the second item in the list. The 'Last commit message' column shows details like 'US Interchange examples and issue #212 and #213' and 'Split multiple values in group conditions correctly'. The 'Last commit date' column shows dates like 'last year' and '7 months ago'.

Name	Last commit message	Last commit date
..		
AE_SOC_PT_Table_Shell.rtf	US Interchange examples and issue #212 and #213	last year
AE_Summary_Table_Shell.pdf	US Interchange examples and issue #212 and #213	last year
Common Safety Displays-LOC.txt	Split multiple values in group conditions correctly	7 months ago
Common Safety Displays.json	Split multiple values in group conditions correctly	7 months ago
Common Safety Displays.xlsx	Updates for ARSP-28, #136, #209, #217, #219, #221, #220	8 months ago
Common Safety Displays.yaml	Split multiple values in group conditions correctly	7 months ago
Demog_Table_Shell.rtf	US Interchange examples and issue #212 and #213	last year
FDA Standard Safety Tables and Figures-LOC.txt	Split multiple values in group conditions correctly	7 months ago
FDA Standard Safety Tables and Figures.json	Split multiple values in group conditions correctly	7 months ago
FDA Standard Safety Tables and Figures.xlsx	Updates for ARSP-28, #136, #209, #217, #219, #221, #220	8 months ago
FDA Standard Safety Tables and Figures.yaml	Split multiple values in group conditions correctly	7 months ago
Vitals_Obs_ChgBaseline_Table_Shell.pdf	US Interchange examples and issue #212 and #213	last year
Vitals_Obs_ChgBaseline_Table_Shell_vertical.pdf	US Interchange examples and issue #212 and #213	last year
at14-5-01.sas	US Interchange examples and issue #212 and #213	last year
csr-cdiscpilot01.pdf	US Interchange examples and issue #212 and #213	last year
sap.pdf	US Interchange examples and issue #212 and #213	last year
table2.sas	US Interchange examples and issue #212 and #213	last year

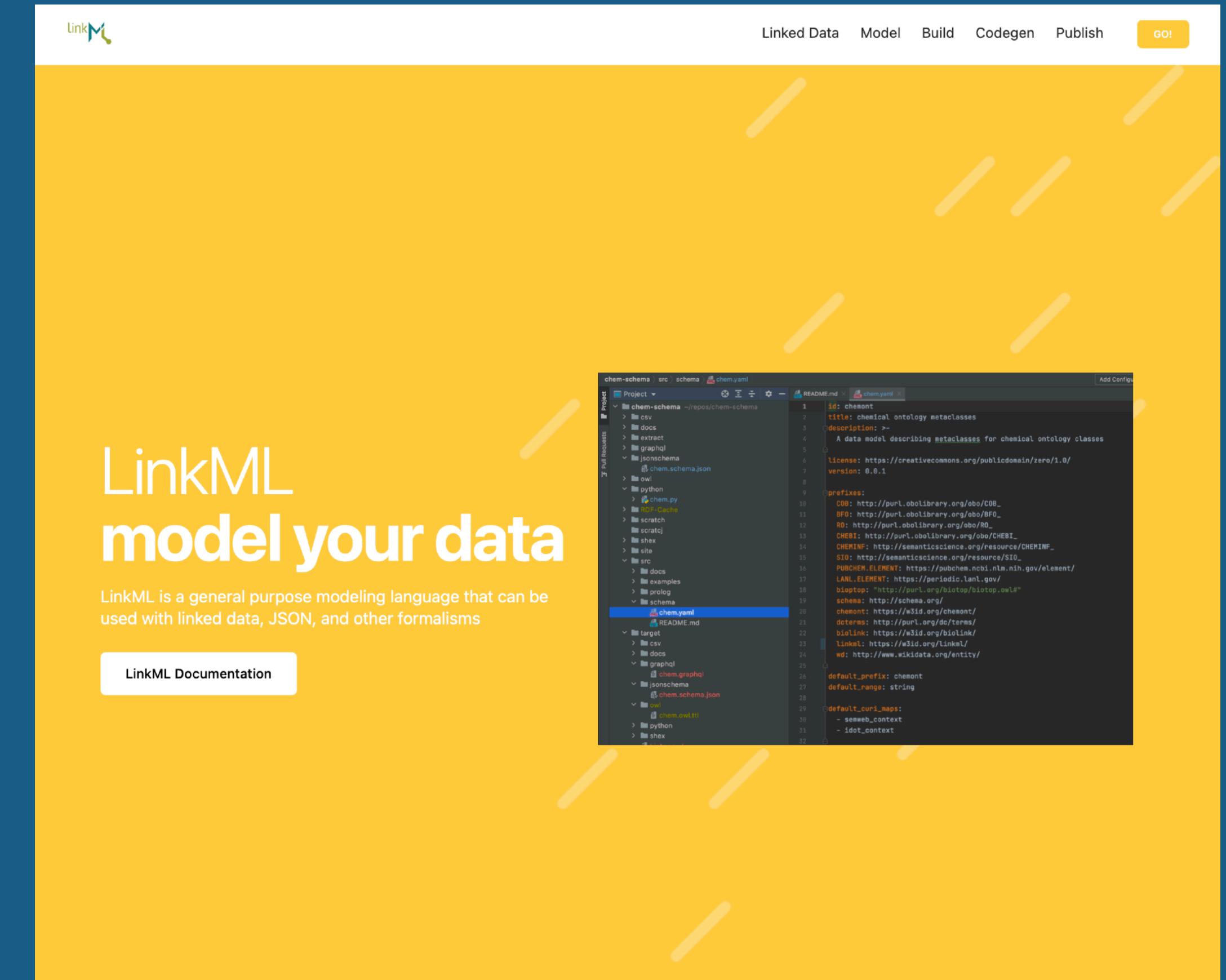
---

# LinkML

---

# LinkML

- Meta Schema
  - Schema generator for ...
  - JSON-Schema
  - RDFS
  - OWL
  - GraphQL
  - SQL DDL
  - Python dataclass



<https://linkml.io/>

# Explanation of Specific Terms

Term	Definition	Synonym	Used in
Schema	A definition of the data structure for a specific domain.	Class	Database, JSON, and etc.
Class	A definition of the data structure for a specific domain. However, in the context of programming languages, it includes not only data but also the definition of behaviors.	Schema	JSON, Python, Property Graph, and etc.
Entity	An entity that includes the data itself.	Object	Property Graph, and etc.
Object	An entity that includes the data itself.	Entity	JSON, Python, and etc.
Literal	Just characters or numbers.		

# A simple example to understand LinkML.

An example of generating a Python dataclass from a LinkML schema

LinkML Schema  
(Meta-Schema)

YAML File: personinfor.yaml

```
id: https://w3id.org/linkml/examples/personinfo
name: personinfo
prefixes:
  linkml: https://w3id.org/linkml/
  personinfo: https://w3id.org/linkml/examples/personinfo
imports:
  - linkml:types
default_range: string
default_prefix: personinfo

classes:
  Person:
    attributes:
      id:
      full_name:
      aliases:
      phone:
      age:
```

Python Dataclass

Python Class: personinfor.py

```
@dataclass(repr=False)
class Person(YAMLRoot):
  _inherited_slots: ClassVar[List[str]] = []

  class_class_uri: ClassVar[URIRef] = PERSONINFO["Person"]
  class_class_curie: ClassVar[str] = "personinfo:Person"
  class_name: ClassVar[str] = "Person"
  class_model_uri: ClassVar[URIRef] = PERSONINFO.Person

  id: Optional[str] = None
  full_name: Optional[str] = None
  aliases: Optional[str] = None
  phone: Optional[str] = None
```

# A simple example to understand LinkML.

An example of generating a JSON schema from a LinkML schema

LinkML Schema  
(Meta-Schema)

YAML File: personinfo.yaml

```
id: https://w3id.org/linkml/examples/personinfo
name: personinfo
prefixes:
  linkml: https://w3id.org/linkml/
  personinfo: https://w3id.org/linkml/examples/personinfo
imports:
  - linkml:types
default_range: string
default_prefix: personinfo

classes:
  Person:
    attributes:
      id:
      full_name:
      aliases:
      phone:
      age:
```

JSON Schema

JSON Schema File: personinfo.schema.json

```
{
  "$defs": {
    "Person": {
      "additionalProperties": false,
      "description": "",
      "properties": {
        "aliases": {
          "type": [
            "string",
            "null"
          ]
        },
        "full_name": {
          "type": [
            "string",
            "null"
          ]
        }
      }
    }
  }
},
```

# A simple example to understand LinkML.

An example of generating a YAML data file from a CSV data file

CSV

CSV File: personinfo\_data.csv

```
id,full_name,phone,age
ORCID:1234,Clark Kent,555-555-5555,32
ORCID:4567,Lois Lane,,33
```

YAML

JSON Schema File: personinfo\_data.yaml

```
persons:
  - id: ORCID:1234
    full_name: Clark Kent
    phone: 555-555-5555
    age: 32
  - id: ORCID:4567
    full_name: Lois Lane
    age: 33
```

# A simple example to understand LinkML.

An example of generating a JSON-LD file from a CSV data file

CSV

CSV File: personinfo\_data.csv

```
id,full_name,phone,age
ORCID:1234,Clark Kent,555-555-5555,32
ORCID:4567,Lois Lane,,33
```

JSON-LD

JSON Schema File: personinfo\_data.jsonld

```
{
  "persons": [
    {
      "id": "ORCID:1234",
      "full_name": "Clark Kent",
      "phone": "555-555-5555",
      "age": 32
    },
    {
      "id": "ORCID:4567",
      "full_name": "Lois Lane",
      "age": 33
    }
  ],
  "@type": "Container",
  "@context": {
    "xsd": "http://www.w3.org/2001/XMLSchema#",
    "linkml": "https://w3id.org/linkml/",
```

# The advantages of using LinkML

- One of the major advantages of using LinkML for data modeling is the ease with which you can convert schemas and data between various file formats.
- The LinkML Schema can also output an Excel template for data input. By entering data into this template, it is possible to convert the data into other formats and handle it as Linked Data.

1	2	3	4	5	6	7	8	9
name	descriptive_label	listItem_label	listItem_name	listItem_description	listItem_label	listItem_label	listItem_analysisId	
1 List of Planned Analyses	LOPA	1	Summary of Demographics					
2 List of Planned Analyses	LOPA	2	Summary of Subjects by Treatment					
3 List of Planned Analyses	LOPA	2	Age					
4 List of Planned Analyses	LOPA	3	Summary by Treatment					
5 List of Planned Analyses	LOPA	3	Comparison by Treatment					
6 List of Planned Analyses	LOPA	2	Age Group					
7 List of Planned Analyses	LOPA	3	Summary of Subjects by Treatment					
8 List of Planned Analyses	LOPA	3	Comparison of Subjects by Treatment					
9 List of Planned Analyses	LOPA	2	Sex					
10 List of Planned Analyses	LOPA	3	Summary of Subjects by Treatment					
11 List of Planned Analyses	LOPA	3	Comparison of Subjects by Treatment					
12 List of Planned Analyses	LOPA	2	Ethnicity					
13 List of Planned Analyses	LOPA	3	Summary of Subjects by Treatment					
14 List of Planned Analyses	LOPA	3	Comparison of Subjects by Treatment					
15 List of Planned Analyses	LOPA	2	Height					
16 List of Planned Analyses	LOPA	3	Summary by Treatment					
17 List of Planned Analyses	LOPA	3	Comparison by Treatment					
18 List of Planned Analyses	LOPA	3	Overall Summary of Treatment-Emergent Adverse Events					
19 List of Planned Analyses	LOPA	2	Summary of Subjects by Treatment					
20 List of Planned Analyses	LOPA	2	Number of subjects with at least one event					
21 List of Planned Analyses	LOPA	3	TEAE					
22 List of Planned Analyses	LOPA	3	Related TEAE					
23 List of Planned Analyses	LOPA	3	Serious TEAE					
24 List of Planned Analyses	LOPA	3	Related Serious TEAE					
25 List of Planned Analyses	LOPA	3	TEAE Leading to Death					
26 List of Planned Analyses	LOPA	3	Related TEAE Leading to Death					
27 List of Planned Analyses	LOPA	3	TEAE Leading to Dose Modification					
28 List of Planned Analyses	LOPA	3	TEAE Leading to Treatment Discontinuation					
29 List of Planned Analyses	LOPA	1	Summary of TEAE by System Organ Class and Preferred Term					
30 List of Planned Analyses	LOPA	2	Summary of Subjects by Treatment					
31 List of Planned Analyses	LOPA							
32 List of Planned Analyses	LOPA							
33 List of Planned Analyses	LOPA							
34 List of Planned Analyses	LOPA							

---

# Power of Linked Data

---

# Machine Readable

- For example, the COVID-19 infection status was published on various countries' websites, but aggregating this data was not an easy task.
- Simply being able to read an Excel file on a computer does not necessarily make it "machine-readable" in the context of structured data.
- For a data value (literal) to be truly "machine-readable," the computer must be able to interpret what type of data it is. This ensures that the data is in a sufficiently machine-readable state.
- Therefore, it is necessary to add extensive metadata to ensure that the computer can interpret and understand the data correctly.

# Linked Data

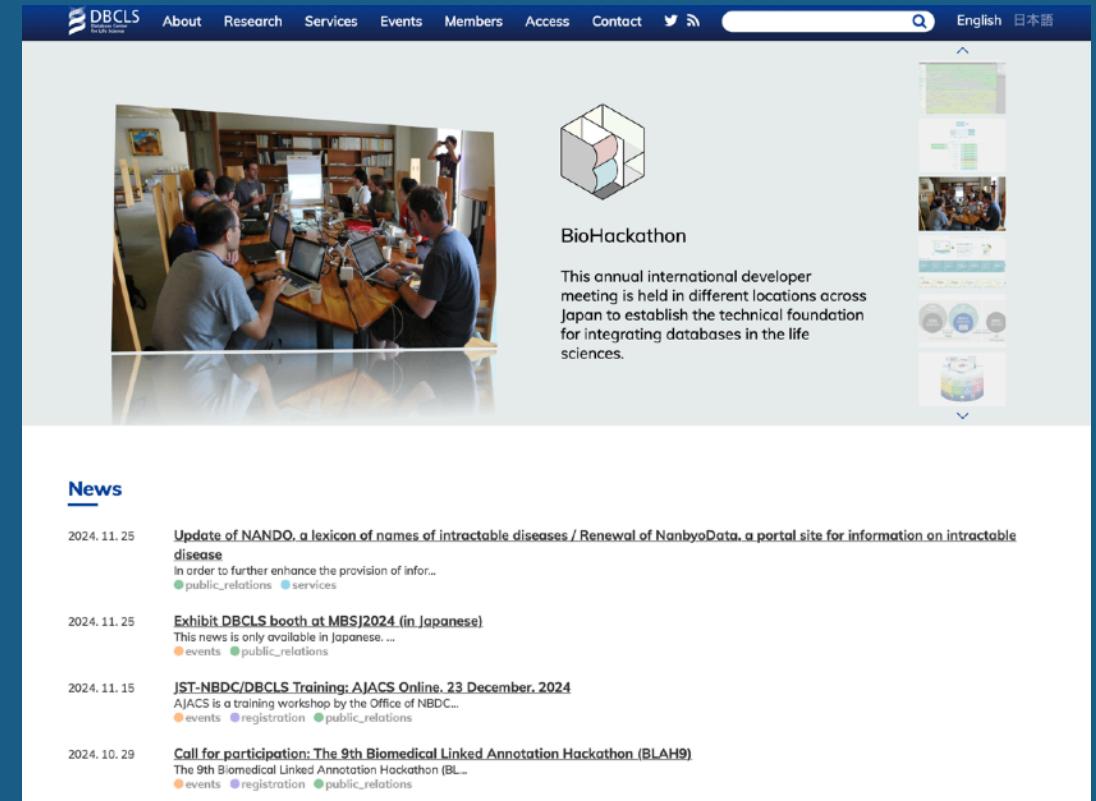
- To solve the various issues mentioned above and achieve the goal, Linked Data was conceived. However, it has not become widely popular...
- A system designed to connect data across the web worldwide.  
[https://en.wikipedia.org/wiki/Semantic\\_Web](https://en.wikipedia.org/wiki/Semantic_Web)
- It is the foundational technology for W3C's Web 3.0. (Note: This is different from Web3, also known as Web Three!)



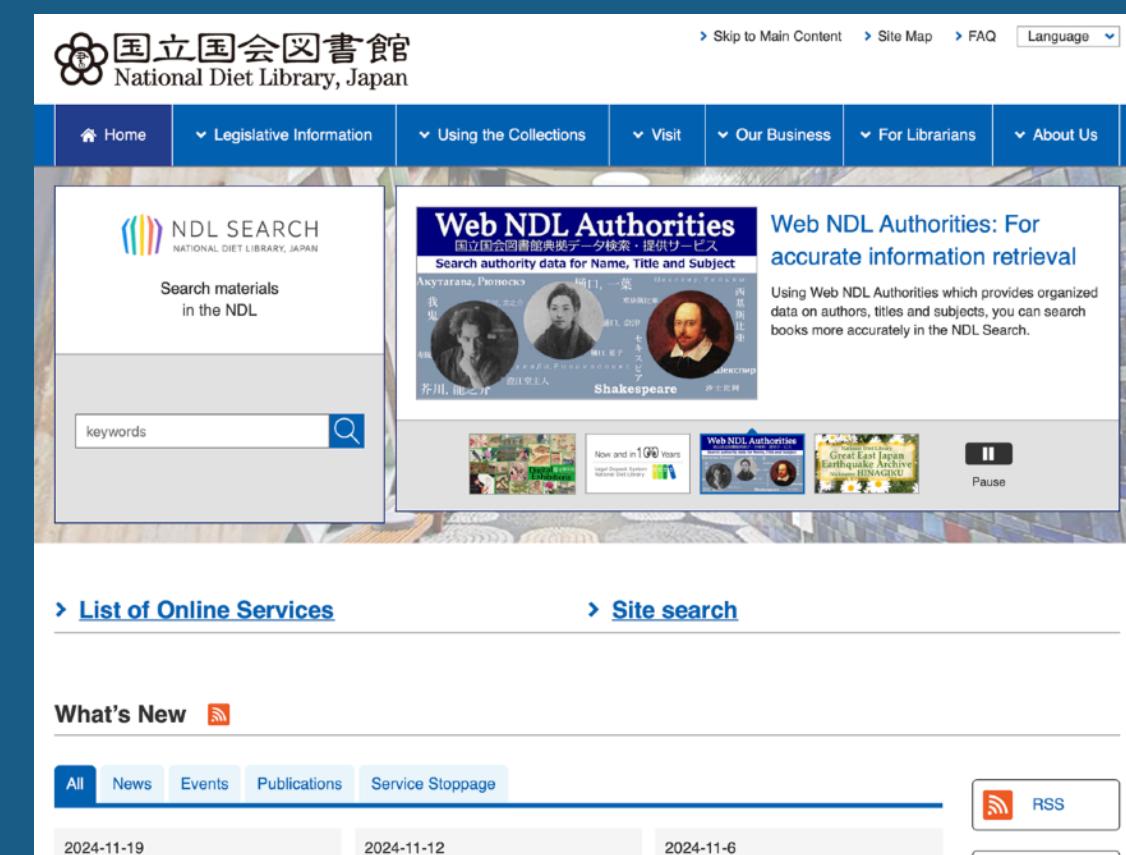
<http://www.stat.go.jp/info/today/095.htm>

# Examples of Linked Data Utilization in Japan

- Database Center for Life Science  
<https://dbcls.rois.ac.jp/>

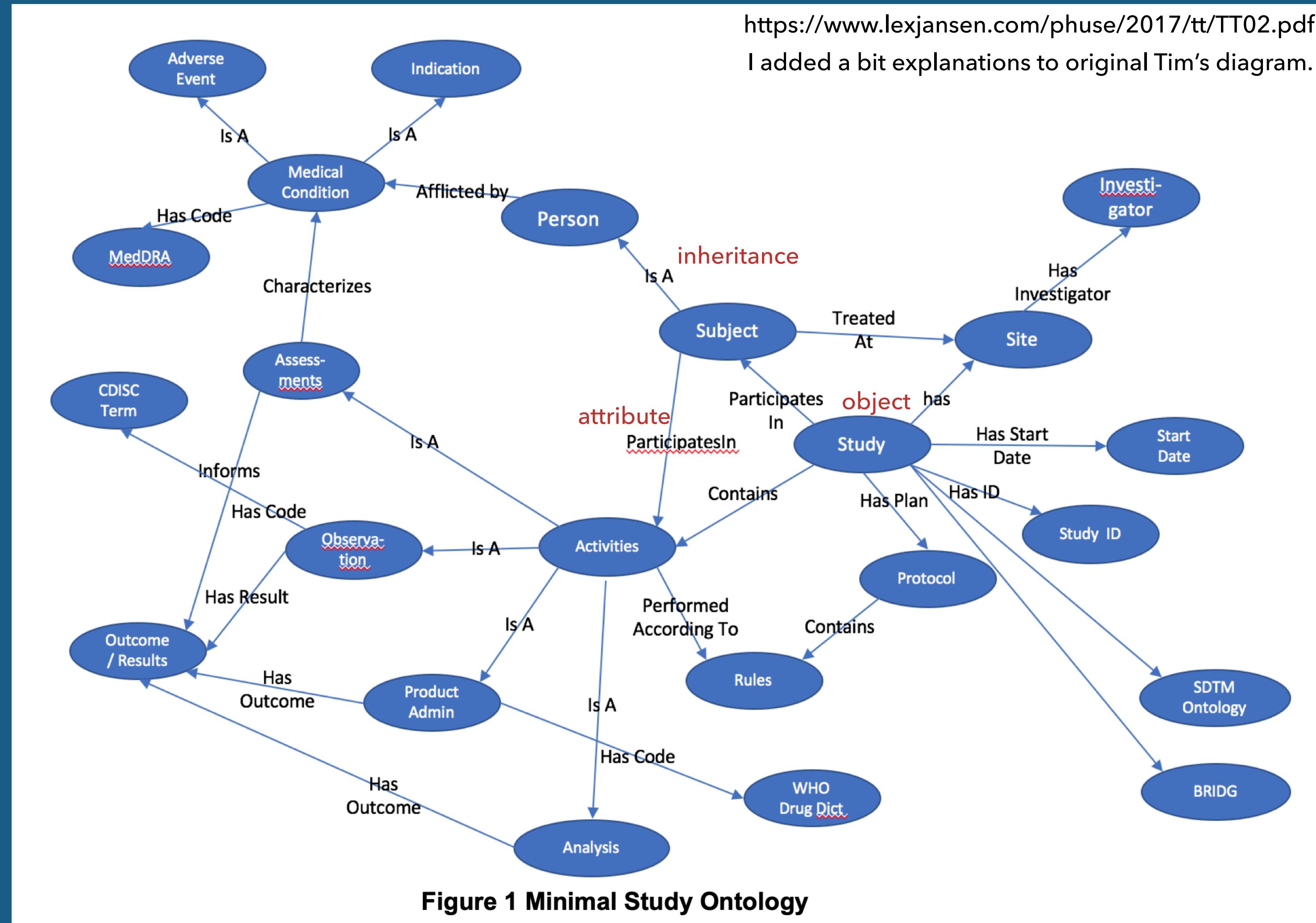


- e-Stat(Statistics of Japan)  
<https://www.e-stat.go.jp/>



- National Diet Library, Japan  
<https://www.ndl.go.jp/jp/dlib/standards/lod/index.html>

# The fundamental of Linked Data is a property graph



- By linking the Results Data from clinical trials conducted worldwide, you can execute a query once to extract all the necessary Results Data from all trials that match the query conditions.

# Advantages and Disadvantages of Using Linked Data

- Advantages
  - It becomes easier to extract the Data and Results Data needed for performing a meta-analysis, allowing for correct selection of the trials to be included.
  - By converting not only Result Data but also SDTM and ADaM data into Linked Data, it becomes easier to verify data traceability. This enhances the transparency of statistical analysis!
  - You can extract the required data from trial data and trial results worldwide with a single query.
  - Utilization by regulatory authorities, such as during approval reviews.
- Disadvantages
  - Conceptually, it's difficult. -> However, AI will eventually handle this, so we may not need to be consciously aware of Linked Data. (AI will create the graph, and the graph will form the AI's RAG).
  - It is necessary to add a lot of metadata.
  - It is necessary to create a Linked Data application.
  - Even without using Linked Data, things are manageable for now, and since AI seems to be handling it well, motivation tends to be low.

I like Linked Data because methods in graph theory and Graph Neural Network are available.

# Is ARS Linked Data?

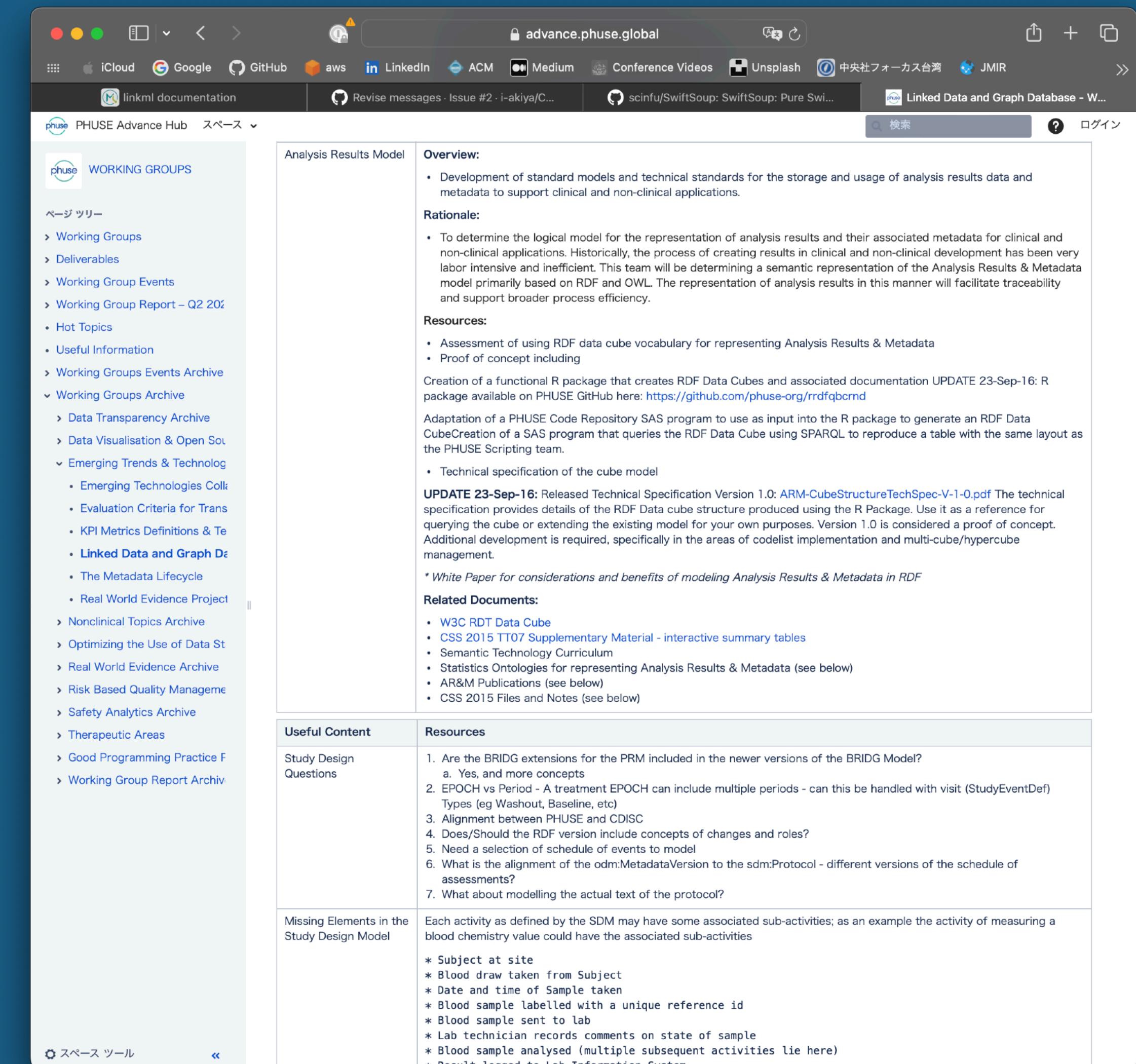
- Yes, but not yet enough to be perfect Linked Data.
  - CDISC ARS is developed with LinkML that is Linked Data modeling language.
  - Currently, the focus seems to be on automating the output of TFLs and define.xml.
  - It seems that there has been little mention of utilizing Linked Data.
- It seems there has been no mention of query languages such as SPARQL or GraphQL for extracting Results Data across multiple items and trials.

# Expectations for CDISC ARS

- I would like to see more diverse examples for eTFL, such as survival analysis.
- If examples of creating Summary Tables for CTD (Common Technical Document) from TFLs of multiple trials were provided, it would help people better understand the benefits of CDISC ARS.
- Currently, there are examples of Safety Tables, but there is a need for data implementation examples for creating various charts.
- By incorporating a standardized set of statistical terms, the statistical analysis methods themselves would also become Machine Readable, making ARS even more useful. By the way, the standardization of statistical terminology (Ontology) is addressed by STATO. <https://stato-ontology.org/>
- I hope that software vendors and open-source developers will become more active in developing software related to ARS.
- There is a desire for the development of software that allows regulatory authorities to review TFLs more easily when using ARS.

# Past work on Linked Data at PHUSE

- Linked Data and Graph Database  
<https://advance.phuse.global/display/WEL/Linked+Data+and+Graph+Database>



The screenshot shows a web browser window with the URL <https://advance.phuse.global/display/WEL/Linked+Data+and+Graph+Database>. The page is titled 'Analysis Results Model' and contains the following content:

- Overview:**
  - Development of standard models and technical standards for the storage and usage of analysis results data and metadata to support clinical and non-clinical applications.
- Rationale:**
  - To determine the logical model for the representation of analysis results and their associated metadata for clinical and non-clinical applications. Historically, the process of creating results in clinical and non-clinical development has been very labor intensive and inefficient. This team will be determining a semantic representation of the Analysis Results & Metadata model primarily based on RDF and OWL. The representation of analysis results in this manner will facilitate traceability and support broader process efficiency.
- Resources:**
  - Assessment of using RDF data cube vocabulary for representing Analysis Results & Metadata
  - Proof of concept including
- Creation of a functional R package that creates RDF Data Cubes and associated documentation** UPDATE 23-Sep-16: R package available on PHUSE GitHub here: <https://github.com/phuse-org/rrdfqbcnrd>
- Adaptation of a PHUSE Code Repository SAS program to use as input into the R package to generate an RDF Data Cube** Creation of a SAS program that queries the RDF Data Cube using SPARQL to reproduce a table with the same layout as the PHUSE Scripting team.
- Technical specification of the cube model**
- UPDATE 23-Sep-16:** Released Technical Specification Version 1.0: [ARM-CubeStructureTechSpec-V-1-0.pdf](#) The technical specification provides details of the RDF Data cube structure produced using the R Package. Use it as a reference for querying the cube or extending the existing model for your own purposes. Version 1.0 is considered a proof of concept. Additional development is required, specifically in the areas of codelist implementation and multi-cube/hypercube management.
- \* White Paper for considerations and benefits of modeling Analysis Results & Metadata in RDF**
- Related Documents:**
  - [W3C RDT Data Cube](#)
  - [CSS 2015 TT07 Supplementary Material - interactive summary tables](#)
  - [Semantic Technology Curriculum](#)
  - [Statistics Ontologies for representing Analysis Results & Metadata \(see below\)](#)
  - [AR&M Publications \(see below\)](#)
  - [CSS 2015 Files and Notes \(see below\)](#)
- Useful Content** **Resources**
- Study Design Questions**
  - 1. Are the BRIDG extensions for the PRM included in the newer versions of the BRIDG Model?
    - a. Yes, and more concepts
  - 2. EPOCH vs Period - A treatment EPOCH can include multiple periods - can this be handled with visit (StudyEventDef) Types (eg Washout, Baseline, etc)
  - 3. Alignment between PHUSE and CDISC
  - 4. Does/Should the RDF version include concepts of changes and roles?
  - 5. Need a selection of schedule of events to model
  - 6. What is the alignment of the odm:MetadataVersion to the sdm:Protocol - different versions of the schedule of assessments?
  - 7. What about modelling the actual text of the protocol?
- Missing Elements in the Study Design Model**

Each activity as defined by the SDM may have some associated sub-activities; as an example the activity of measuring a blood chemistry value could have the associated sub-activities

  - \* Subject at site
  - \* Blood draw taken from Subject
  - \* Date and time of Sample taken
  - \* Blood sample labelled with a unique reference id
  - \* Blood sample sent to lab
  - \* Lab technician records comments on state of sample
  - \* Blood sample analysed (multiple subsequent activities lie here)
  - \* Result logged to Lab Information System

# CDISC's Guide related to Linked Data

- CDISC Standards in RDF Reference Guide
- CDISC Standards in RDF User Guide

<https://www.cdisc.org/standards/data-exchange/rdf>

CDISC Standards in RDF Reference Guide



**CDISC Standards in RDF Reference Guide**  
Version 1.0 Final  
Prepared by  
The PhUSE CS Semantic Technology Working Group

**Notes to Readers**  
This document is the specification of the RDF vocabulary and the RDF data sets that represent the current foundational CDISC standards in a W3C endorsed semantic format.

**Revision History**

Date	Version	Summary of Changes
2014-12-12	Draft 1.0	Version 1.0 for public comment.
2015-06-18	Final 1.0	Version 1.0 Final

© 2015 Clinical Data Interchange Standards Consortium, Inc. All rights reserved  
Version 1.0 Final

Page 1  
June 18, 2015

CDISC Standards in RDF User Guide



**CDISC Standards in RDF User Guide**  
Version 1.0 Final  
Prepared by  
The PhUSE CS Semantic Technology Working Group

**Notes to Readers**  
Based on the Reviewer's Guide that was created to support the public review process, this document provides additional information related to the access to and use of the CDISC Standards in RDF.

**Revision History**

Date	Version	Summary of Changes
2014-12-12	Draft 1.0	Version 1.0 Reviewer Guide for public comment.
2015-07-21	Final 1.0	Version 1.0 Final User Guide

© 2015 Clinical Data Interchange Standards Consortium, Inc. All rights reserved  
Version 1.0 Final

Page 1  
July 21, 2015

---

# Thank you for your Attention

---