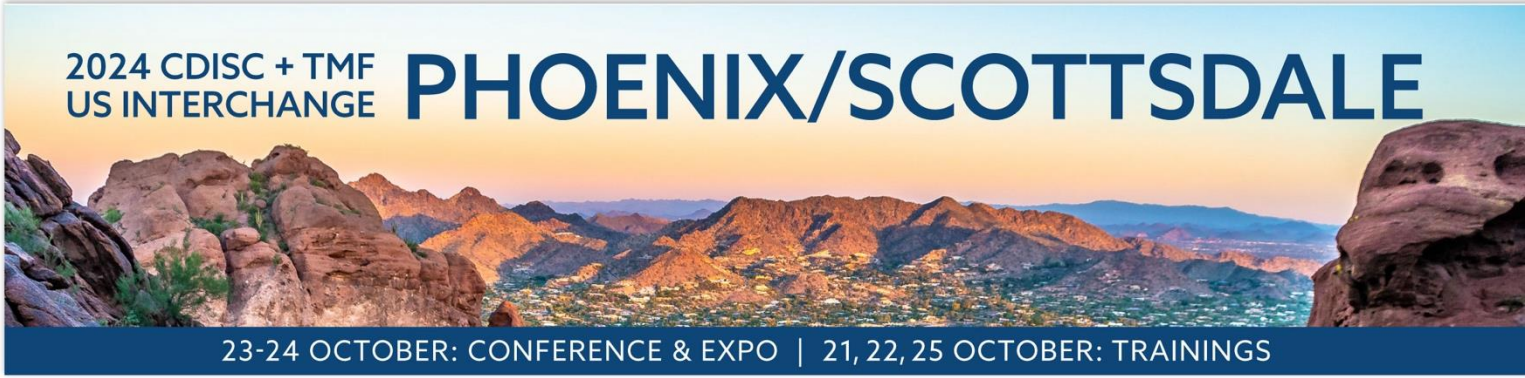


USDM in Action: From Protocol to SDTM

Dave Iberson-Hurst
Data4knowledge ApS

Meet the Speaker

Dave Iberson-Hurst

Title: Partner

Organization: data4knowledge ApS

40+ years' experience across several industries with the last 20 years spent in the pharmaceutical industry combining his technology and software development experience with clinical data standards. During this time, he has served as the CDISC CTO, worked on, and led, several CDISC teams, presented in many forums in Europe, the US, and elsewhere across the globe.

He has worked closely with the FDA, EMA, HL7, ISO, and other standards organizations and was was a member of CDISC's Blue Ribbon commission. He is currently the CDISC Product Owner for the Digital Data Flow project.

He is a partner at data4knoweldge in Copenhagen and is focused on getting greater value and utility from clinical trial data

Meet the Contributors

Kirsten Walther Langendorf

Title: Partner

Organization: data4knowledge ApS

20+ years' experience in the pharmaceutical industry within programming, IT implementation & validation, process improvement, CDISC standards implementation, and statistics.

As partner at data4knowledge in Copenhagen, she has been involved in implementing various e2e metadata driven systems based on linked data technologies.

Johannes Ulander

Title: Partner

Organization: data4knowledge ApS

20+ years' experience in standardizing clinical data and have been involved in implementing CDISC standards from an end-to-end perspective for the last 15 years. For the last 7 years by using linked data and graph databases.

He is a partner at data4knowledge in Umeå and an authorized CDISC SDTM instructor.

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.*
- *Contractor to CDISC as the DDF USDM Product Owner*

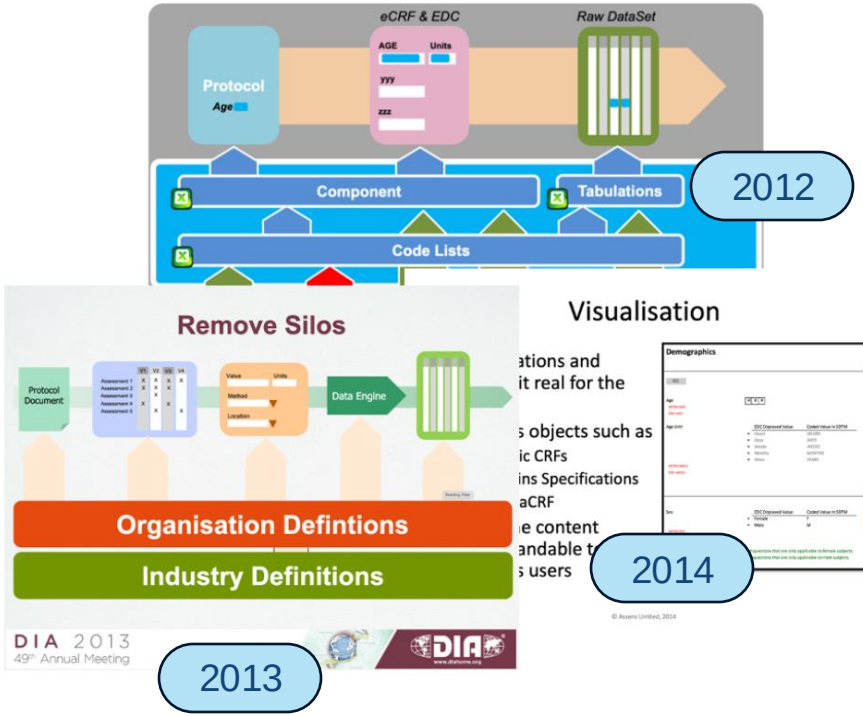
Disclosure: It's All About Hats

This talk is about using the USDM as the foundation for much of what we do and the potential for the removal of silos and automation.

It is about the sharing of ideas.

Old Ideas Whose Time Has Come

- We [industry] have been looking at removing silos for a decade or more
- We have been looking at “eProtocol” for probably two decades or more
- DDF, USDM, ICH M11, precisonFDA … all these initiatives / standards are making it a reality



One Model: USDM at the Centre

data4knowledge

Making better use of data

We help our customers make better use of their primary asset, the clinical data.

Contact

dih@data4knowledge.dk
kwl@data4knowledge.dk

Overview

Over the past few years, the partners at d4k have worked on ways in which the handling of data can be improved, silos removed, and automation increased. But showing these ideas in action requires more than a few slides. To this end, we at d4k have built a demonstrator to illustrate the potential of using the Unified Study Definitions Model (USDM) with Biomedical Concepts (BCs) providing the foundation for the automation of downstream processes that industry has been lacking.

The demonstrator implements USDM version 3, CDISC Biomedical Concepts, CDISC Controlled Terminology, CDISC SDTM, and the emerging ICH M11 CeSHARP eProtocol standards to illustrate how the detailed study definition provides the foundation to link subject data to the study design thus allowing the automated generation of SDTM without the need for any programming. This same design can also be used for the generation of submission artefacts and this poster discusses how an implementation based upon the USDM can serve multiple purposes.

Key Takeaways:

- USDM is the foundation industry has been missing
- USDM opens the door to multiple new innovations and approaches
- The TransCelerate end-to-end Digital Data Flow vision is achievable today

Study Design

The technology demonstrator implements a single linked model. The heart of the model is provided by the USDM.

The USDM provides the ability to define the overall study design and logic and provides the foundation for everything that follows.

Study Detail

Biomedical Concepts are already linked into the USDM and provide the necessary detail and precision not normally found in existing, paper, protocols. BCs define the data to be collected and provide the basis for the data contract, the data needed to meet the study's needs

Operational Data

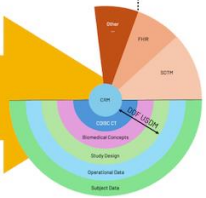
The USDM contains some site information, but this is related to amendments and recruitment. This information is expanded to include more details needed to link subject data into the model.

Subject Data

We now have a full study definition to which subject data can be linked. This then creates a single linked graph containing both the study design plus the data; no more silos!

SDTM & Other Exports

We can now link SDTM to the data. The SDTM is linked to a small model (the CRM) that models the intricacies of observations. This model is also linked to the BCs (in fact BCs should be based on this model) such that there is a link from data to BC to SDTM. A similar approach can be taken with other data exports



Technology Demonstrator

The technology demonstrator implements the model described here and then populates it with data to show the concepts working.

We have leveraged the test data provided by the TransCelerate / CDISC DDF project taking the LZZT protocol and associated raw data and loading these data into a single model. From that we can automatically extract SDTM, aCRF and define.xml. Data capture instruments can also be driven from the metadata

Future work will look at such topics as subject and site burden, study risk and study cost.

Scan the QR code to see a video of the demonstrator in action

Views

The expanded USDM model can serve many purposes. Many of the artefacts required within clinical development are simply extracts, exports or views, of one consistent and linked set of data.

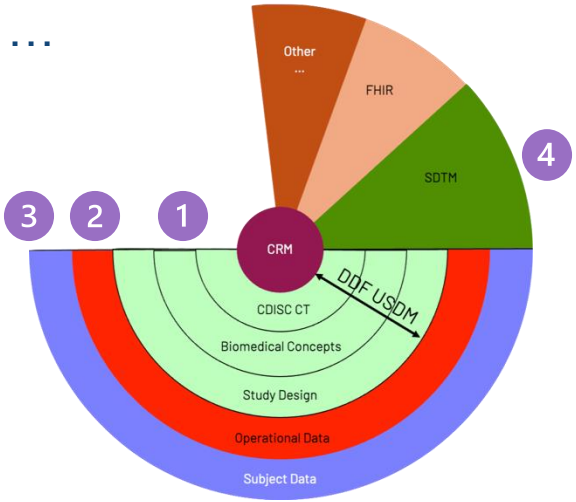
Protocol: The entire protocol can be extracted, either in a sponsor template or the M11 template.
Specific Views: Tailored views for specific roles, e.g. EDC build, data monitoring, can be accommodated.
EDC: Machine and human readable exports for EDC configuration.
CRF & Define: Data capture specifications, aCRFs and Define.xml can be generated from the model prior to a single data point being captured.
SDTM: As already noted the model supports the automated generation of SDTM datasets capable of supporting multiple versions of SDTMIG.

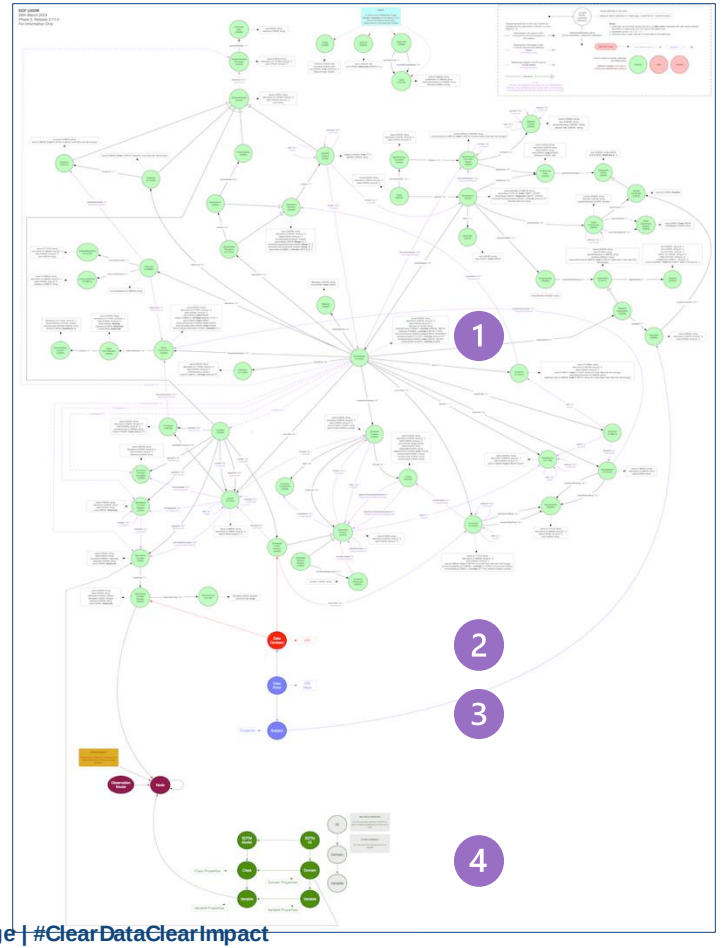
And many more: The expanded USDM model is not limited to the above exports or views. Many other use cases and exports could be envisaged such as subject journey, site risk, TMF and CTMS.

USDM as the Foundation

1. USDM is the foundation
2. Add the “data contract”
3. Attach subjects and their data
4. Link to SDTM

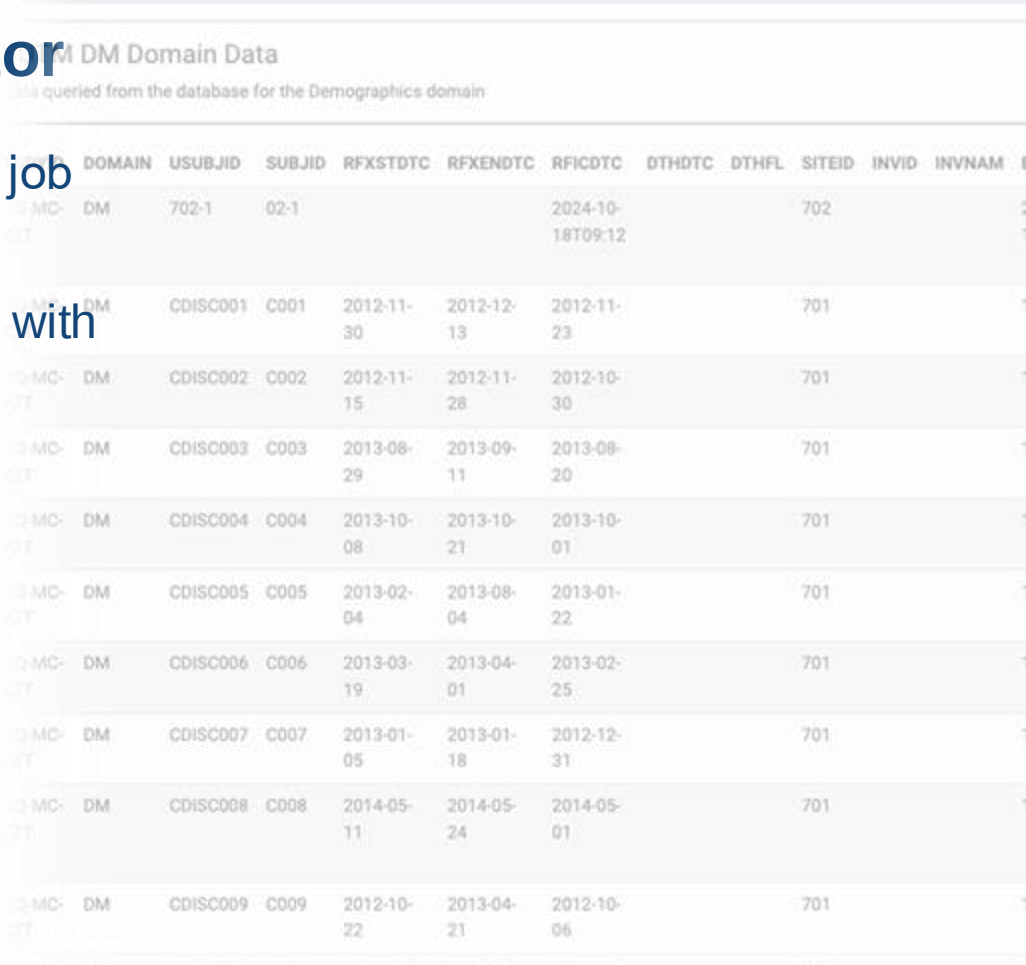
… and use …





Technology Demonstrator

- A Powerpoint just doesn't do the job
- Need to see the ideas in action
- Implemented in Python, FastAPI with a Neo4j database
- Has a basic User Interface (UI)
- We are continuing to work on it



Paper to eProtocol using USDM

1

2. Objectives

2.1. Primary Objectives

The primary objectives of this study are

- To determine if there is a statistically significant relationship (overall Type 1 error rate, α=.05) between the change in both ADAS-Cog (see Attachment LZZT.2) and CIBIC+ (see Attachment LZZT.3) scores, and drug dose (0, 50 cm² [54 mg], and 75 cm² [81 mg]).
- To document the safety profile of the xanomeline TTS.

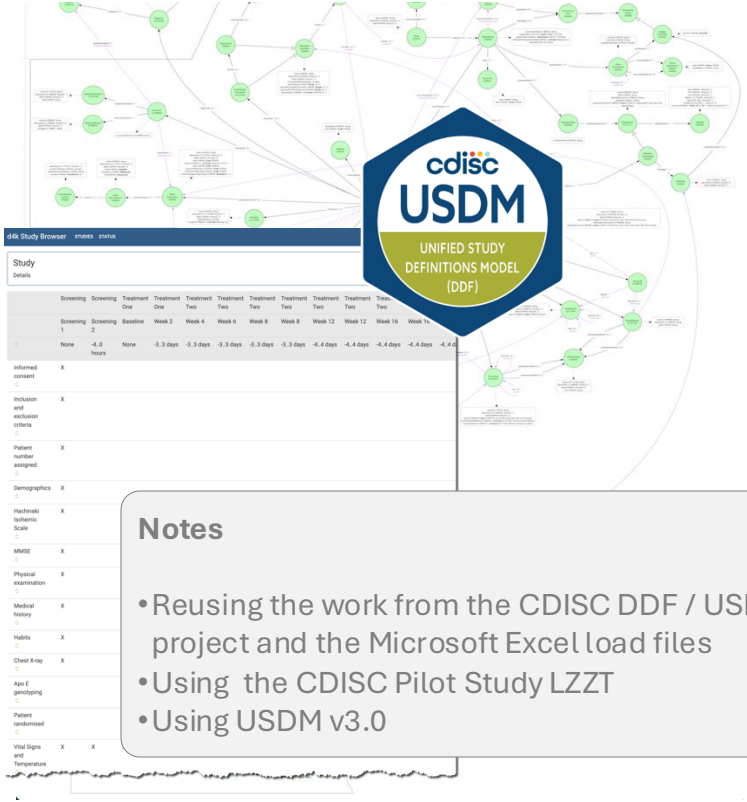
2.2. Secondary Objectives

The secondary objectives of this study are

- To assess the dose-dependent improvement in behavior. Improved scores on the Revised Neuropsychiatric Inventory (NPI-X) will indicate improvement in these areas (see Attachment LZZT.4).
- To assess the dose-dependent improvements in activities of daily living. Improved scores on the Disability Assessment for Dementia (DAD) will

Protocol Attachment LZZT.1
Schedule of Events for Protocol H2Q-MC-LZZT(c)

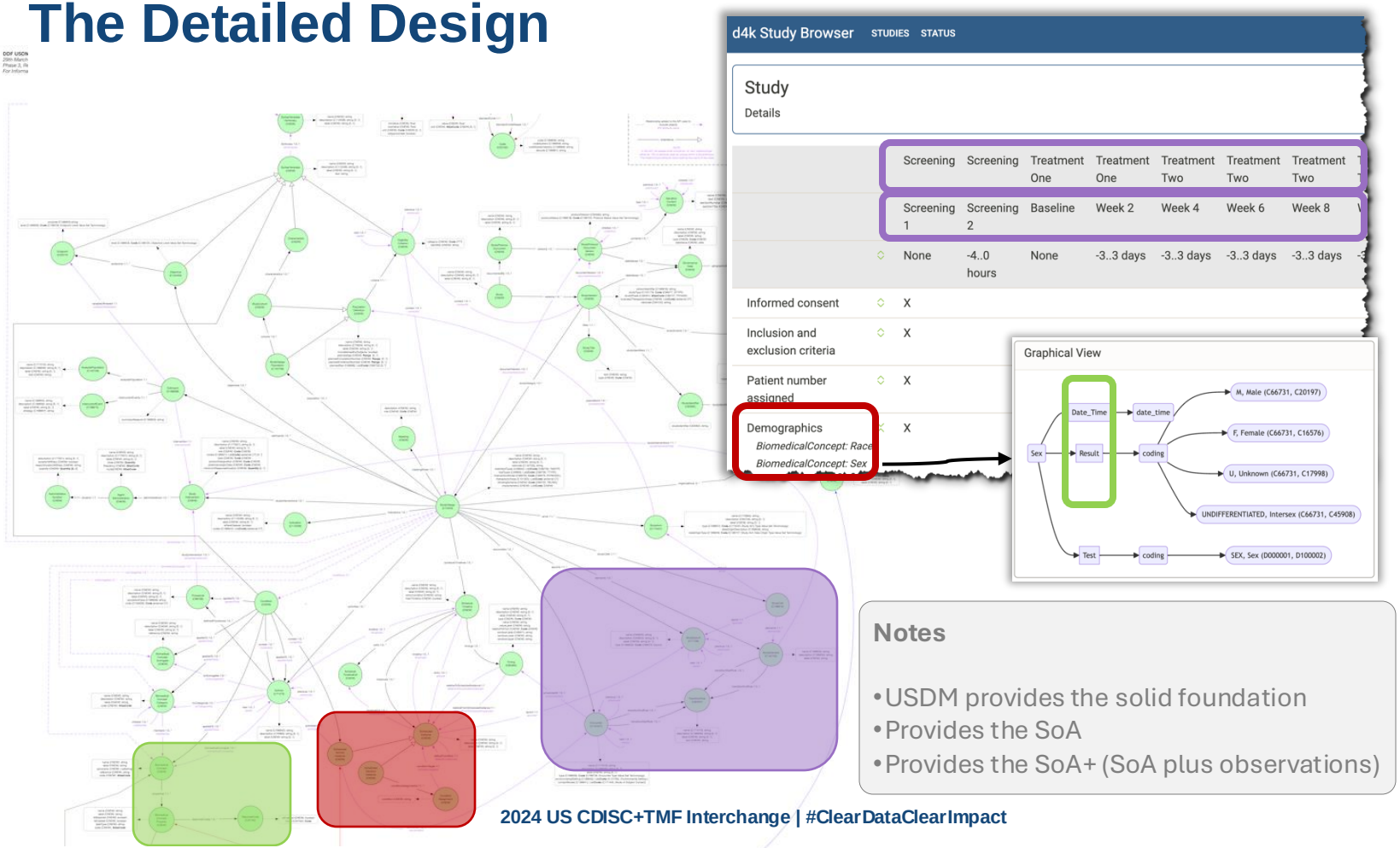
	VISIT	1	2	3	4	5		7	8
ACTIVITY	WEEK	-2	-3	0	2	4		6	8
Informed consent		X							
Patient number assigned		X							
Hachinski ≤4		X							
MMSE 10-23		X							
Physical examination		X							
Medical History		X							
Habits		X							
Chest x-ray		X							
Apo E genotyping					X				
Patient randomized				X					
Vital signs/Temperature		X	X	X	X	X		X	X
Ambulatory ECG placed			X						
Ambulatory ECG removed				X					
ECG		X			X	X		X	X
Placebo TTS test		X							



Notes

- Reusing the work from the CDISC DDF / USDM project and the Microsoft Excel load files
- Using the CDISC Pilot Study LZZT
- Using USDM v3.0

The Detailed Design

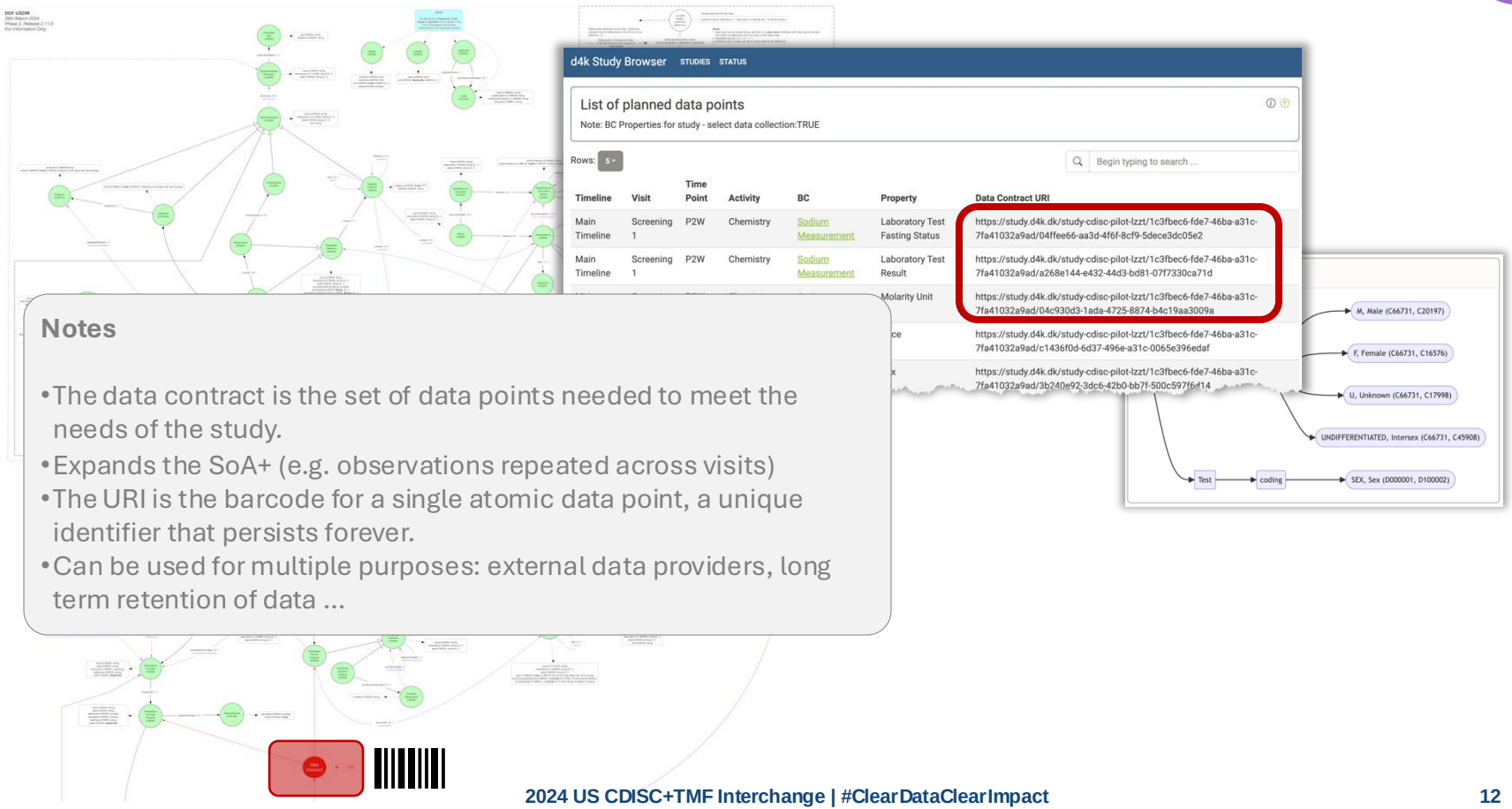


Notes

- USDM provides the solid foundation
- Provides the SoA
- Provides the SoA+ (SoA plus observations)

Add in a “Data Contract”

2

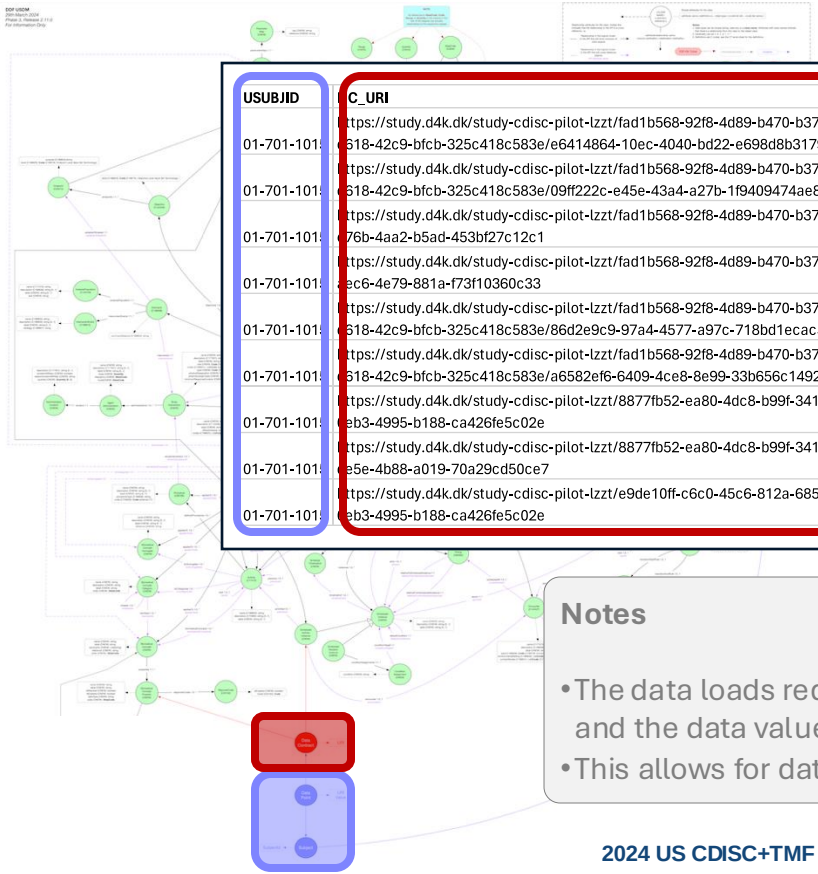


Notes

- The data contract is the set of data points needed to meet the needs of the study.
- Expands the SoA+ (e.g. observations repeated across visits)
- The URI is the barcode for a single atomic data point, a unique identifier that persists forever.
- Can be used for multiple purposes: external data providers, long term retention of data …

Link the Subject Data

3



USUBJID	DC_URI	DATAPOINT_URI	VALUE
01-701-101	https://study.d4k.dk/study-cdisc-pilot-lzzt/fad1b568-92f8-4d89-b470-b37fcb6f4d8d/eed3c040-d618-42c9-bfcb-325c418c583e/e6414864-10ec-4040-bd22-e698d8b31790	https://study.d4k.dk/study-cdisc-pilot-lzzt/fad1b568-92f8-4d89-b470-b37fcb6f4d8d/eed3c040-d618-42c9-bfcb-325c418c583e/e6414864-10ec-4040-bd22-e698d8b31790/01-701-1015	64
01-701-101	https://study.d4k.dk/study-cdisc-pilot-lzzt/fad1b568-92f8-4d89-b470-b37fcb6f4d8d/eed3c040-d618-42c9-bfcb-325c418c583e/09ff222c-e45e-43a4-a27b-1f9409474ae8	https://study.d4k.dk/study-cdisc-pilot-lzzt/fad1b568-92f8-4d89-b470-b37fcb6f4d8d/eed3c040-d618-42c9-bfcb-325c418c583e/09ff222c-e45e-43a4-a27b-1f9409474ae8/01-701-1015	mmHg
01-701-101	https://study.d4k.dk/study-cdisc-pilot-lzzt/fad1b568-92f8-4d89-b470-b37fcb6f4d8d/15ece647-e76b-4aa2-b5ad-453bf27c12c1	https://study.d4k.dk/study-cdisc-pilot-lzzt/fad1b568-92f8-4d89-b470-b37fcb6f4d8d/15ece647-e76b-4aa2-b5ad-453bf27c12c1/01-701-1015	58.00.00
01-701-101	https://study.d4k.dk/study-cdisc-pilot-lzzt/fad1b568-92f8-4d89-b470-b37fcb6f4d8d/955c7c7a-aec6-4e79-881a-f73f10360c33	https://study.d4k.dk/study-cdisc-pilot-lzzt/fad1b568-92f8-4d89-b470-b37fcb6f4d8d/955c7c7a-aec6-4e79-881a-f73f10360c33/01-701-1015	IN
01-701-101	https://study.d4k.dk/study-cdisc-pilot-lzzt/fad1b568-92f8-4d89-b470-b37fcb6f4d8d/eed3c040-d618-42c9-bfcb-325c418c583e/86d2e9c9-97a4-4577-a97c-718bd1ecac5a	https://study.d4k.dk/study-cdisc-pilot-lzzt/fad1b568-92f8-4d89-b470-b37fcb6f4d8d/eed3c040-d618-42c9-bfcb-325c418c583e/86d2e9c9-97a4-4577-a97c-718bd1ecac5a/01-701-1015	57
01-701-101	https://study.d4k.dk/study-cdisc-pilot-lzzt/fad1b568-92f8-4d89-b470-b37fcb6f4d8d/eed3c040-d618-42c9-bfcb-325c418c583e/a6582ef6-6409-4ce8-8e99-33b656c1492b	https://study.d4k.dk/study-cdisc-pilot-lzzt/fad1b568-92f8-4d89-b470-b37fcb6f4d8d/eed3c040-d618-42c9-bfcb-325c418c583e/a6582ef6-6409-4ce8-8e99-33b656c1492b/01-701-1015	BEATS/MIN
01-701-101	https://study.d4k.dk/study-cdisc-pilot-lzzt/8877fb52-ea80-4dc8-b99f-341d70c5a796/4b7708f0-0eb3-4995-b188-ca426fe5c02e	https://study.d4k.dk/study-cdisc-pilot-lzzt/8877fb52-ea80-4dc8-b99f-341d70c5a796/4b7708f0-0eb3-4995-b188-ca426fe5c02e/01-701-1015	41
01-701-101	https://study.d4k.dk/study-cdisc-pilot-lzzt/8877fb52-ea80-4dc8-b99f-341d70c5a796/1b2e1fa0-de5e-4b88-a019-70a29cd50ce7	https://study.d4k.dk/study-cdisc-pilot-lzzt/8877fb52-ea80-4dc8-b99f-341d70c5a796/1b2e1fa0-de5e-4b88-a019-70a29cd50ce7/01-701-1015	U/L
01-701-101	https://study.d4k.dk/study-cdisc-pilot-lzzt/e9de10ff-c6c0-45c6-812a-685d20747ce7/4b7708f0-0eb3-4995-b188-ca426fe5c02e	https://study.d4k.dk/study-cdisc-pilot-lzzt/e9de10ff-c6c0-45c6-812a-685d20747ce7/4b7708f0-0eb3-4995-b188-ca426fe5c02e/01-701-1015	45

Notes

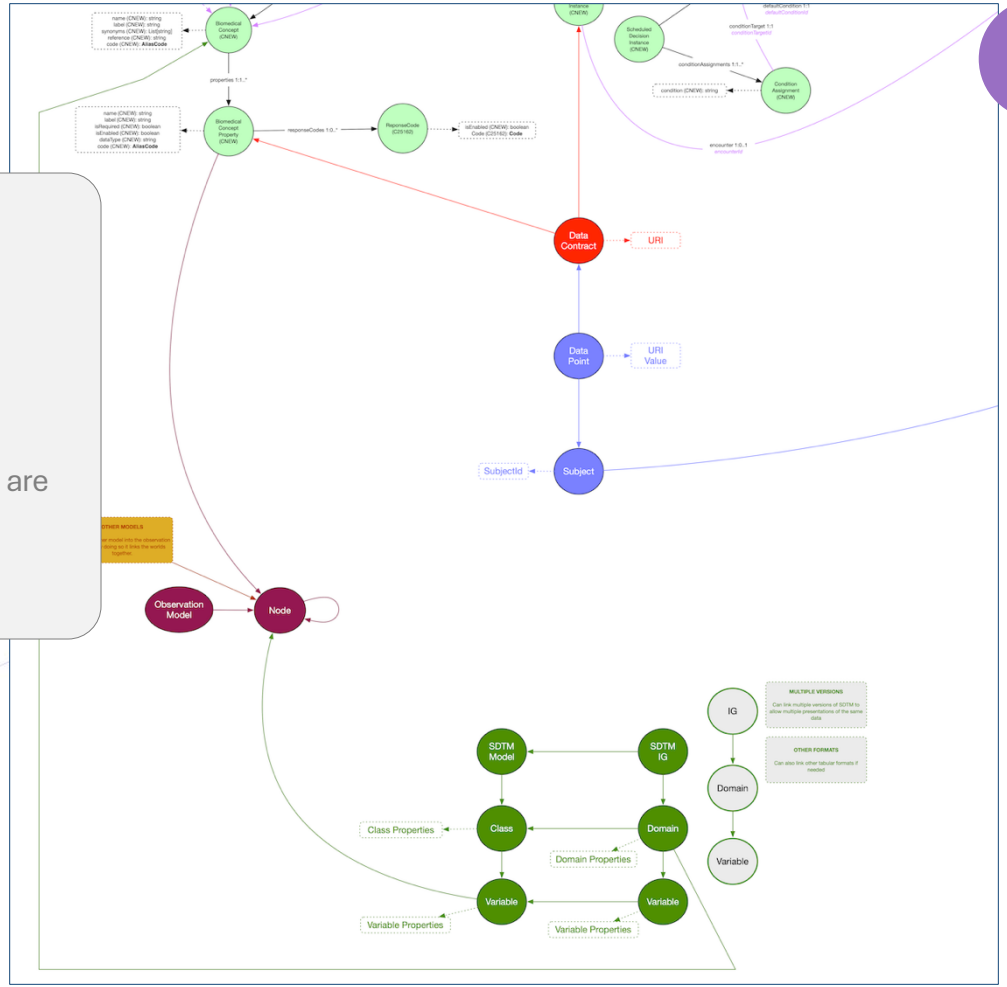
- The data loads require a triple of the subject identifier, the data contract URI and the data value
- This allows for data to be linked into the overall data, in bulk or individually

Link SDTM

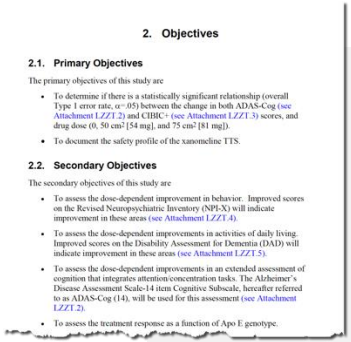
4

Notes

- We have a small model that models what an observation is
- Link the SDTM IG via the SDTM Model to the observation model
- Link BCs to the same observation model
- We take this further in that all the BCs we use are based on that same observation model
- Can link to multiple versions of the IG
- Can link to other models other than SDTM



And so …



2. Objectives

2.1. Primary Objectives

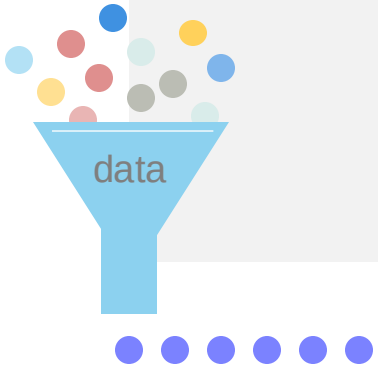
The primary objectives of this study are

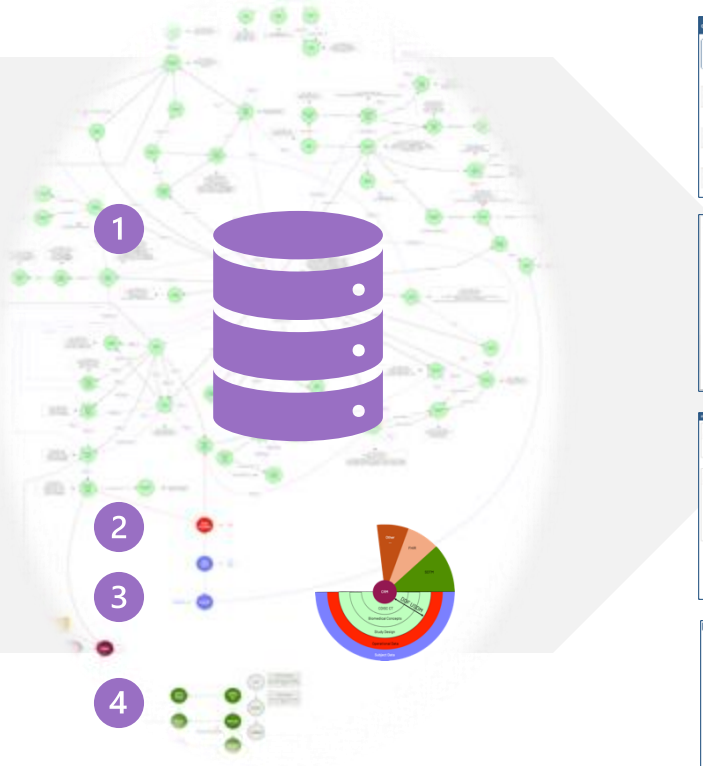
- To determine if there is a statistically significant relationship (overall Type 1 error rate, α=.05) between the change in both ADAS-Cog (see Attachment LZZT.2) and CIBIC+ (see Attachment LZZT.3) scores, and drug dose (0, 50 cm² [54 mg], and 75 cm² [81 mg]).
- To document the safety profile of the xanomeline TTS.

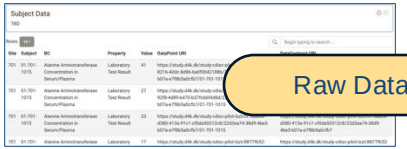
2.2. Secondary Objectives

The secondary objectives of this study are

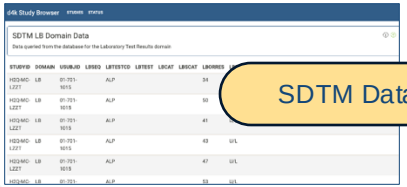
- To assess the dose-dependent improvement in behavior. Improved scores on the Revised Neuropsychiatric Inventory (NPI-X) will indicate improvement in these areas (see Attachment LZZT.4).
- To assess the dose-dependent improvements in activities of daily living. Improved scores on the Disability Assessment for Dementia (DAD) will indicate improvement in these areas (see Attachment LZZT.5).
- To assess the dose-dependent improvements in an extended assessment of cognition that integrates attention/concentration tasks. The Alzheimer's Disease Assessment Scale-14 item Cognitive Subscale, hereafter referred to as ADAS-Cog (14), will be used for this assessment (see Attachment LZZT.2).
- To assess the treatment response as a function of Apo E genotype.



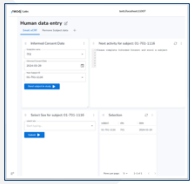


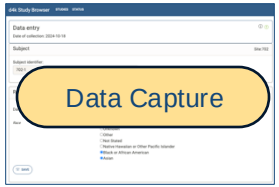


Raw Data

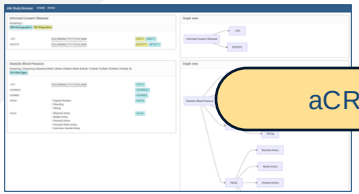


SDTM Data

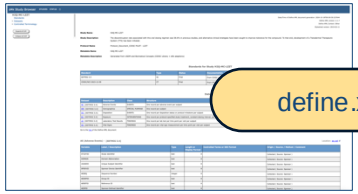




Data Capture

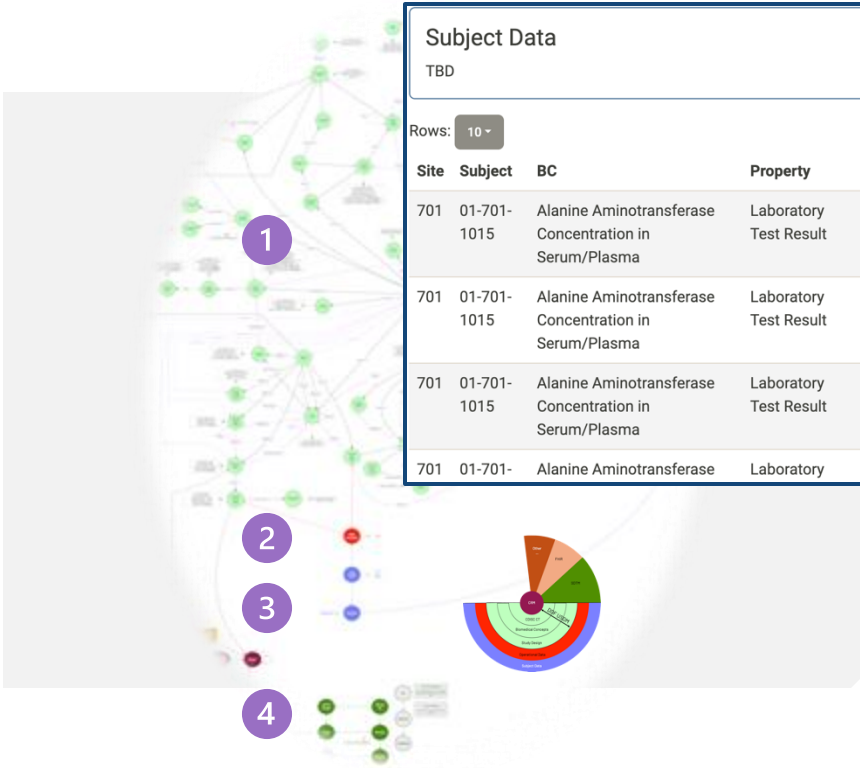


aCRF



define.xml

Present Raw Data



Subject Data

TBD

Rows: 10

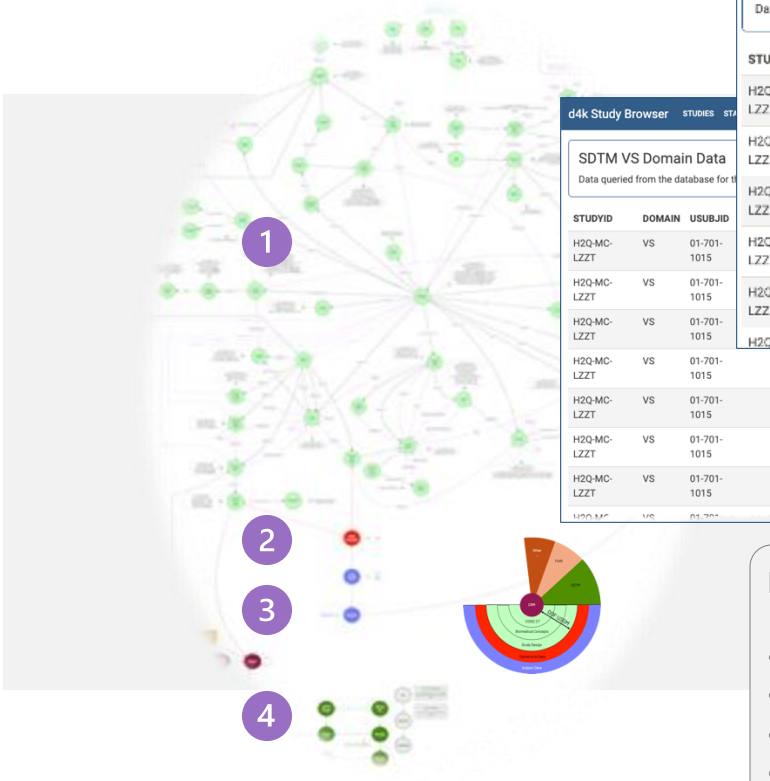
Begin typing to search ...

Site	Subject	BC	Property	Value	DataPoint URI	DataContract URI
701	01-701-1015	Alanine Aminotransferase Concentration in Serum/Plasma	Laboratory Test Result	41	https://study.d4k.dk/study-cdisc-pilot-lzzt/6f26a59c-8216-40dc-8d86-ba0f0042188b/22d2ea74-38d9-4be3-b07a-e7f8b5a0cfb7/01-701-1015	https://study.d4k.dk/study-cdisc-pilot-lzzt/6f26a59c-8216-40dc-8d86-ba0f0042188b/22d2ea74-38d9-4be3-b07a-e7f8b5a0cfb7
701	01-701-1015	Alanine Aminotransferase Concentration in Serum/Plasma	Laboratory Test Result	27	https://study.d4k.dk/study-cdisc-pilot-lzzt/fad1b568-92f8-4d89-b470-b37fcb6f4d8d/22d2ea74-38d9-4be3-b07a-e7f8b5a0cfb7/01-701-1015	https://study.d4k.dk/study-cdisc-pilot-lzzt/fad1b568-92f8-4d89-b470-b37fcb6f4d8d/22d2ea74-38d9-4be3-b07a-e7f8b5a0cfb7
701	01-701-1015	Alanine Aminotransferase Concentration in Serum/Plasma	Laboratory Test Result	23	https://study.d4k.dk/study-cdisc-pilot-lzzt/0f7dba56-d380-413a-91c1-cf0da50312c8/22d2ea74-38d9-4be3-b07a-e7f8b5a0cfb7/01-701-1015	https://study.d4k.dk/study-cdisc-pilot-lzzt/0f7dba56-d380-413a-91c1-cf0da50312c8/22d2ea74-38d9-4be3-b07a-e7f8b5a0cfb7
701	01-701-	Alanine Aminotransferase	Laboratory	17	https://study.d4k.dk/study-cdisc-pilot-lzzt/8877fb52-	https://study.d4k.dk/study-cdisc-pilot-lzzt/8877fb52-

Notes

- The data in its natural form

Present SDTM Data



d4k Study Browser STUDIES STATUS

SDTM LB Domain Data

Data queried from the database for the Laboratory Test Results domain

STUDYID	DOMAIN	USUBJID	LBSEQ	LBTESTCD	LBTEST	LBCAT	LBSCAT	LBORRES	LBORRESU	LBRESSCL	LBRESTYP	LBCOLSRT	LBORNRLO	LBORNRI
H2Q-MC-LZZT	LB	01-701-1015		ALP				34	U/L					
H2Q-MC-LZZT	LB	01-701-1015		ALP				50	U/L					
H2Q-MC-LZZT	LB	01-701-1015		ALP				41	U/L					
H2Q-MC-LZZT	LB	01-701-1015		ALP				43	U/L					
H2Q-MC-LZZT	LB	01-701-1015		ALP				47	U/L					
H2Q-MC-LZZT	LB	01-701-		ALP				53	U/L					

d4k Study Browser STUDIES STA

SDTM VS Domain Data

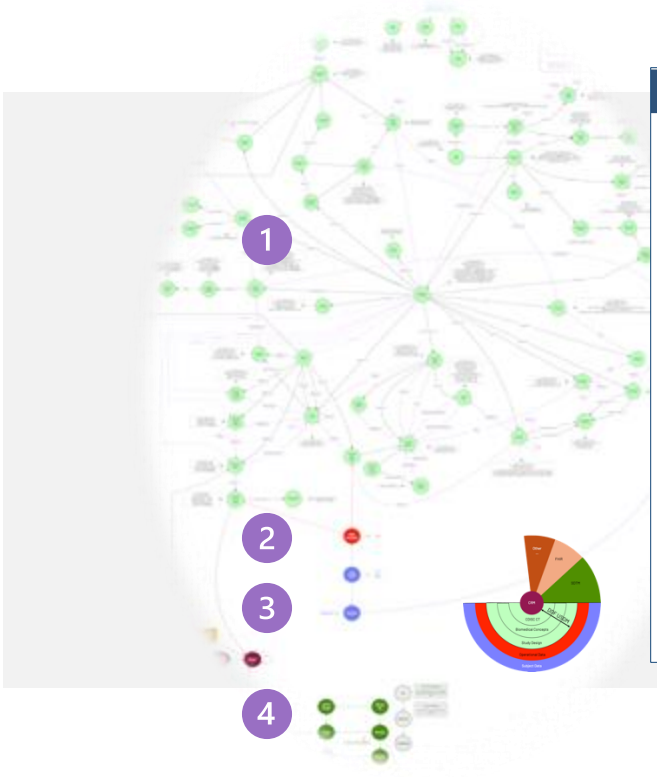
Data queried from the database for t

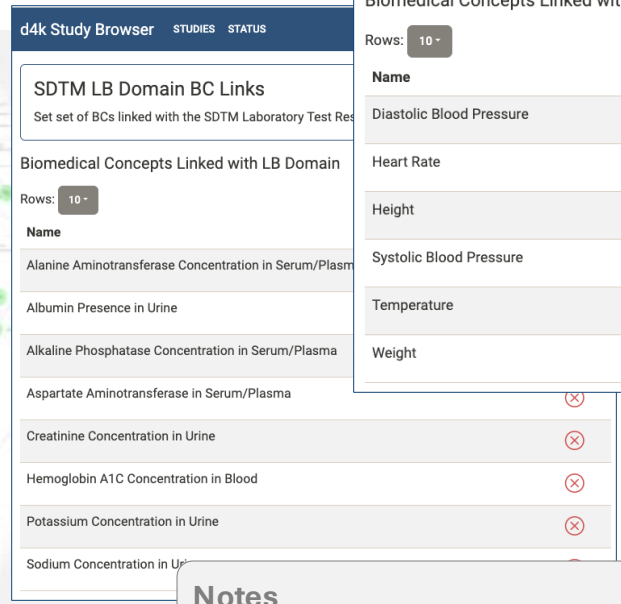
STUDYID	DOMAIN	USUBJID									
H2Q-MC-LZZT	VS	01-701-1015	DIABP	59	mmHg			Screening 2	Screening	PT1M
H2Q-MC-LZZT	VS	01-701-1015	DIABP	71	mmHg			Screening 2	Screening	PT2M
H2Q-MC-LZZT	VS	01-701-1015	DIABP	68	mmHg			Screening 2	Screening	PT5M
H2Q-MC-LZZT	VS	01-701-1015	DIABP	51	mmHg			Baseline	Treatment 1	PT1M

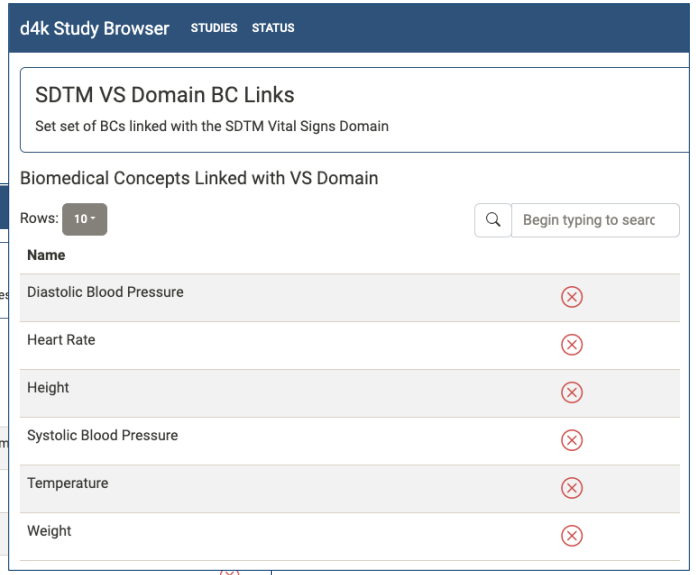
Notes

- Can export the data into SDTM via query
- Derived data is handled
- The solution can handle the “unexpected”
- Think about the data, not how it is presented
- It works in ‘real time’

Configure Domains







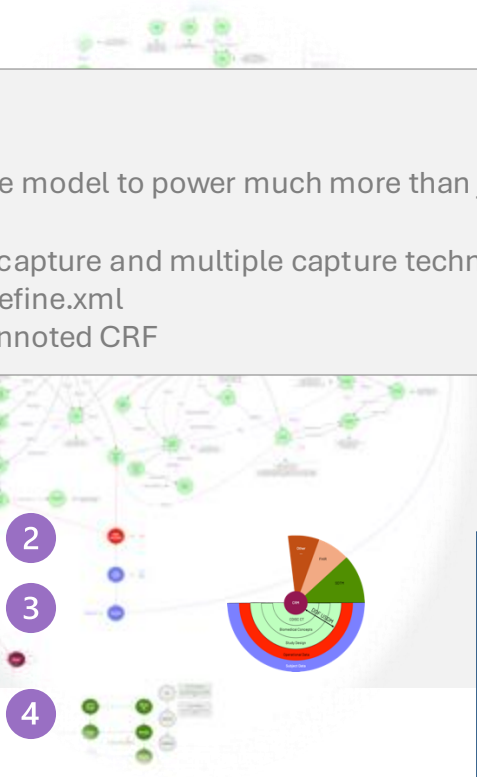
Notes

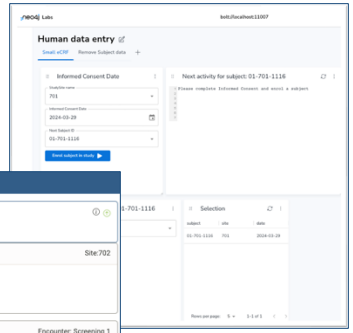
- Allows for observations to be “assigned” to domains
- Allows for custom domains etc
- Allows for data to be repeated across domains

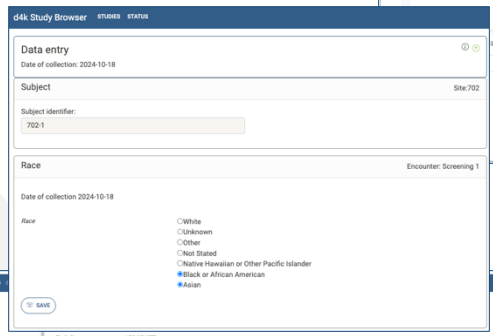
Data Collection, aCRF and Define.xml

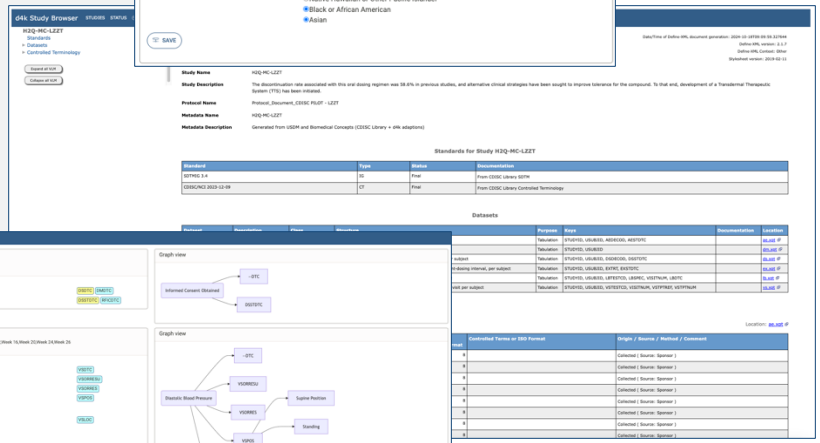
Notes

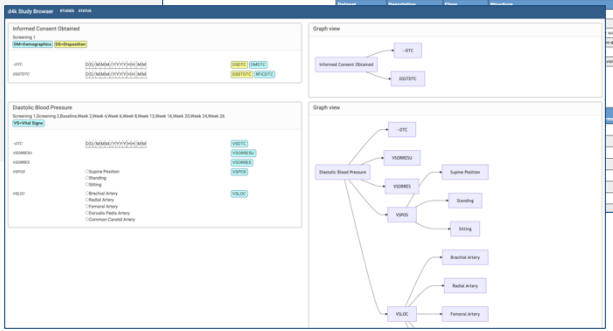
- Can use the model to power much more than just SDTM
- Drive data capture and multiple capture technologies
- Export of define.xml
- Export of annotated CRF



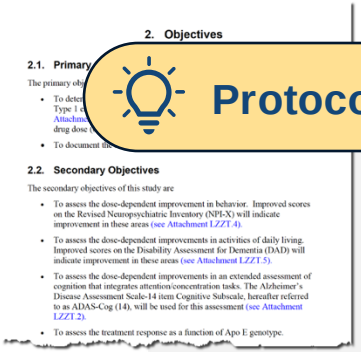




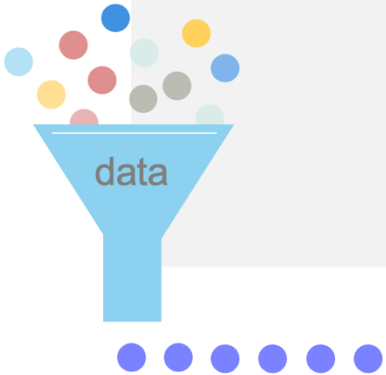


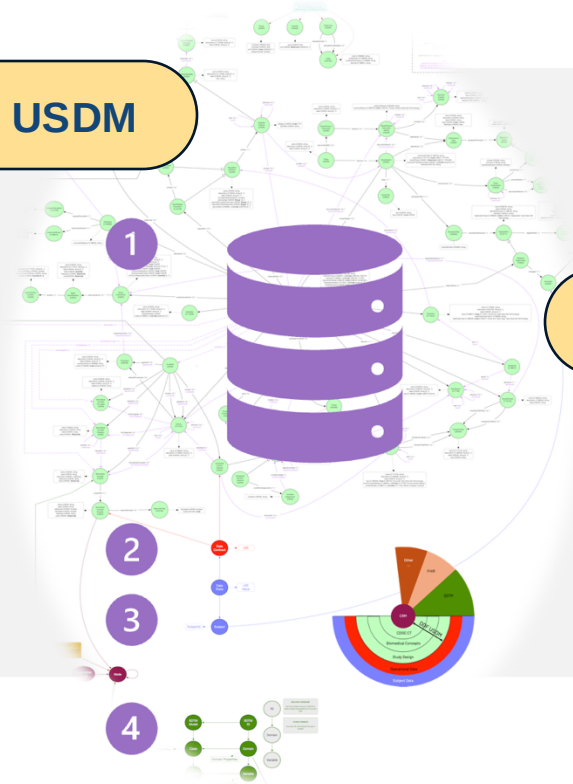


And Next …



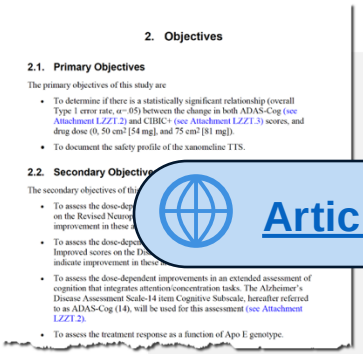
Protocol → USDM

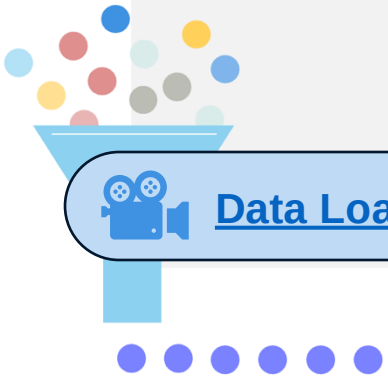


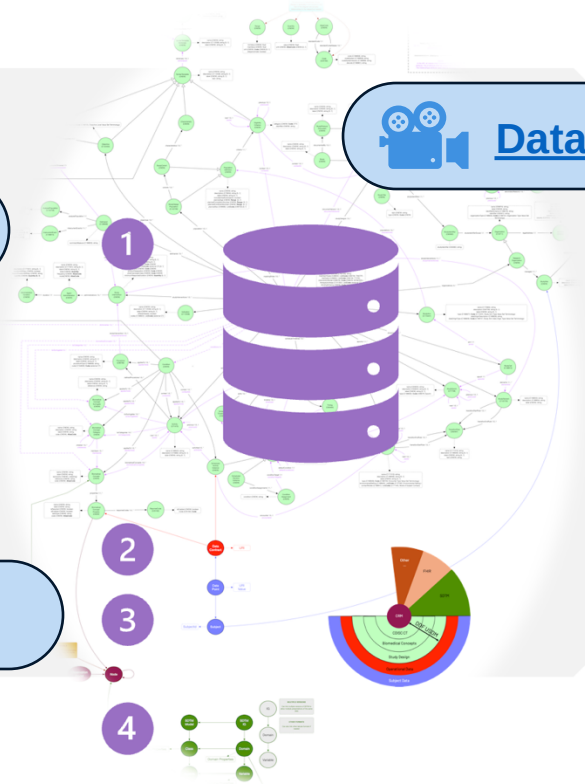


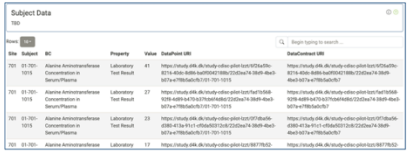
Subject Journey

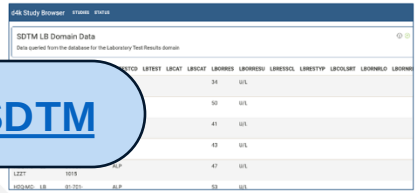
More Info

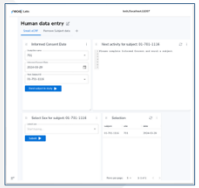


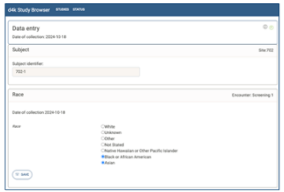












Article

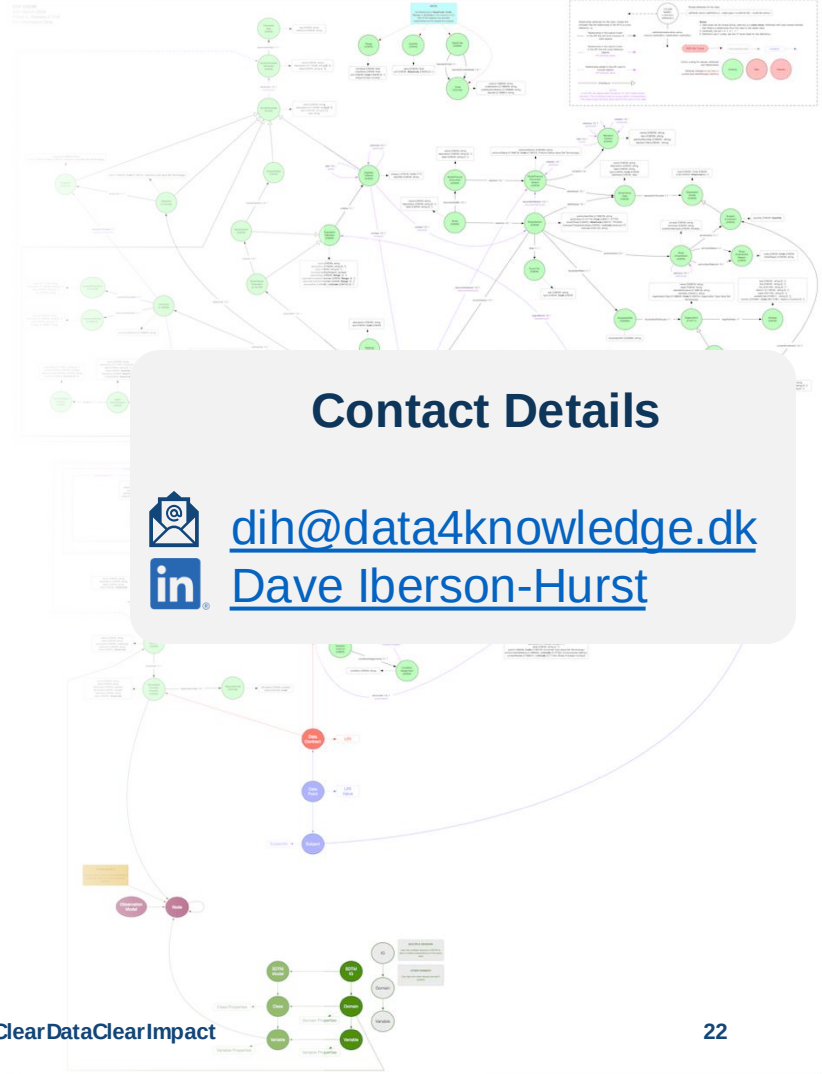
Data and SDTM

Data Load

Forms and Define

Summary

- USDM provides the strong foundation
- We extended USDM …
  - Established the data contract
  - Linked in the subject data
  - Linked in SDTM
  - Allows for data capture
  - Extracted SDTM, aCRF and define.xml
- And more to come …



Contact Details

dih@data4knowledge.dk
Dave Iberson-Hurst