

From Silos To Knowledge

A data centric and technology agnostic approach

TH06, PHUSE EU Connect, Belfast

14th November 2022

Version 2



Dave Iberson-Hurst

Dave has over 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 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 a member of CDISC's Blue Ribbon commission.

He is a partner at data4knowledge in Copenhagen and is focused on getting greater value and utility from clinical trial data.

Abstract

In 2019, Transcelerate ran a hackathon as part of their Digital Data Flow (DDF) initiative. This has subsequently been expanded with a collaboration with CDISC for a new standard, the Unified Study Definitions Model (USDM).

This presentation links to the Transcelerate / CDISC work integrating the emerging study design model with other standards such as Biomedical Concepts (BCs) and CDISC SDTM to form an integrated industry knowledge graph.

This knowledge graph allows for the integration of not only traditional data sources such as CRF data with the study design but also handles other sources such as Real-World Evidence (HL7 FHIR) data, bringing data and the study design together into one coherent knowledge graph. The resulting graph can then be used to generate a variety of outputs: not only the SDTM submission artefacts but any desired view of our data for secondary use; from silos to knowledge.

From Silos to Knowledge

Solid Foundation

A well defined set of definitions sourced from the CDISC library and elsewhere combined with a Biomedical Concept (BC) library. BCs are a unit of knowledge, representing a real-world concept, pre-defined and bound to terminology that are ready to be linked into studies. The BCs provides reusable components that bring data consistency across study builds and data collection and thus subsequent processing. The BCs protect our study data against future capture technology change.

CDISC Library + 1

2 Study Design & Configuration

Study Build

Combining the power of a study design model, e.g. the TransCelerate Digital Data Flow (DDF) Unified Study Design Model (USDM), and BCs allows for complete and well structured machine-readable study designs to be constructed. The combination provides for a precise study data contract that benefits the data collection process. It also allows for study artefacts such as define.xml and study annotated CRFs to be produced in an automated manner prior to a single data point being captured.

And We Gain ...



Other desired outputs such as OMOP can be produced (queried) from the same graph. We achieve a repository of searchable "studies", both the metadata and the data, all linked.

Old studies can be imported and brought back to life to prevent data decay and the model will allow for translations between SDTM versions. Multiple studies can be held together and thus searched and queried. We achieve our goal of better data: better data quality (from the BCs being more precise) and better data utility (from the linking into a single graph and access to the study design).



Automate SDTM

A precise and well structured study design linked to the subject data via BCs provides for a complete study graph from which SDTM can be queried and the datasets produced. Edge cases and supplemental qualifiers need consideration but the BC design allows for such flexibility.

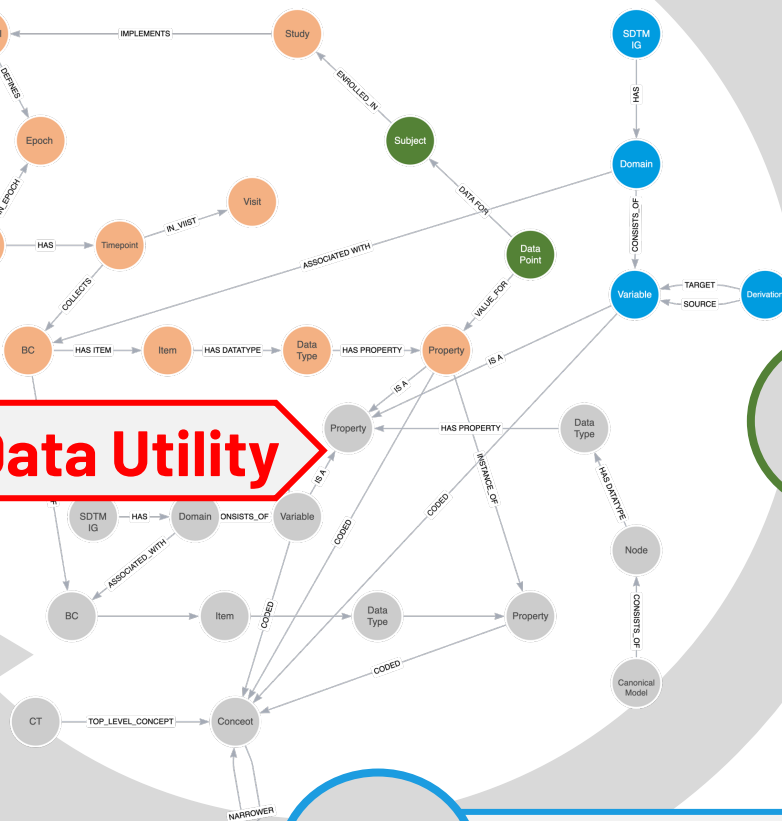
5 Data Utility

3 Data

Import Data

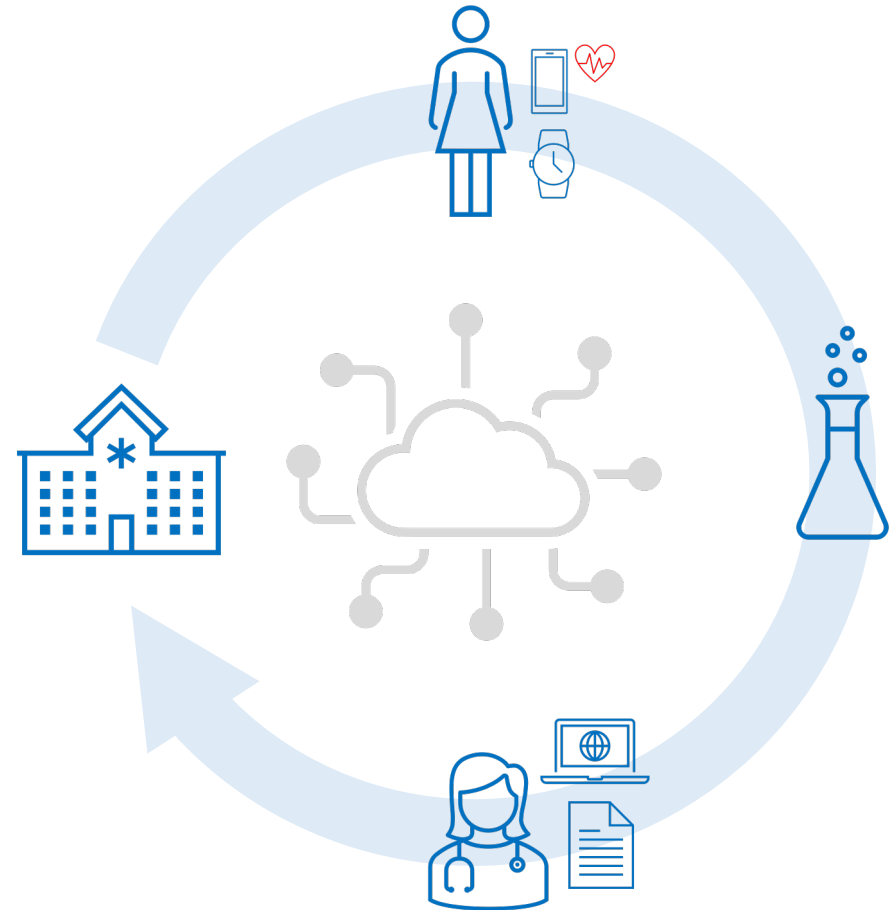
Data can be imported from a variety of sources in a range of formats including tabular (e.g. lab) formats and from other sources such as HL7 FHIR (RWE). The use of a "translation" model (the Clinical Recording Model) allows for a variety of input sources to be accommodated.

4 SDTM Export

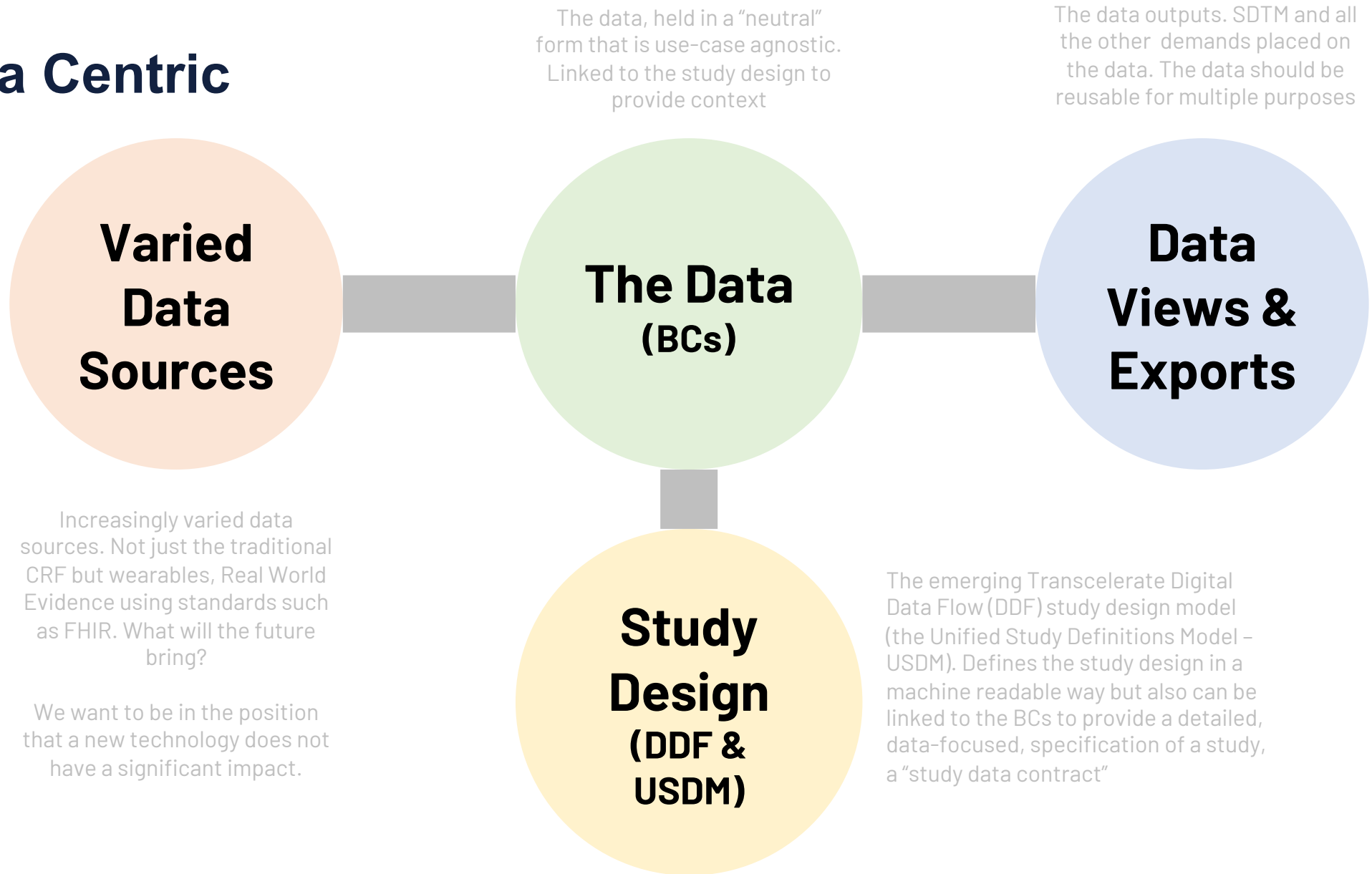


Data Complexity & Technology

- Observations that make up a study are varied, are complex, with complex result types, qualifiers etc.
- Traditional data sources like CRF, Lab are being augmented with a set of increasingly diverse data sources, such as wearables, Real World Evidence (RWE) ...
- Desire to decouple the presentation (e.g. SDTM) from the data's natural form so allow greater reuse (utility) of the data
- And so ...
 - Focus on the "observations" to enable better data handling and technology integration
 - Place data into context, link it to the design



Data Centric



The Work

- The work has been undertaken as a private research project
- Driven by the desire to document the theory behind biomedical concepts and how BCs can be used
- Supported by implementation work
 - 2018 PHUSE paper
 - Paper & Prototype
 - Technical Demonstrator

Biomedical Concepts

A Treatise

Dave Iberson-Hurst

data4knowledge

Draft 0.3
18th May 2022

```
JID SUBJID RFSTDTC RFENDTC RFXSTDTC \
237 1237
236 1236
235 1235
234 1234
239 1239
238 1238

RFENDTC DTHDTC DTHFL SITEID INVID \
S-0002 I-0002
S-0001 I-0001
S-0001 I-0001
S-0001 I-0001
S-0002 I-0002
S-0002 I-0002

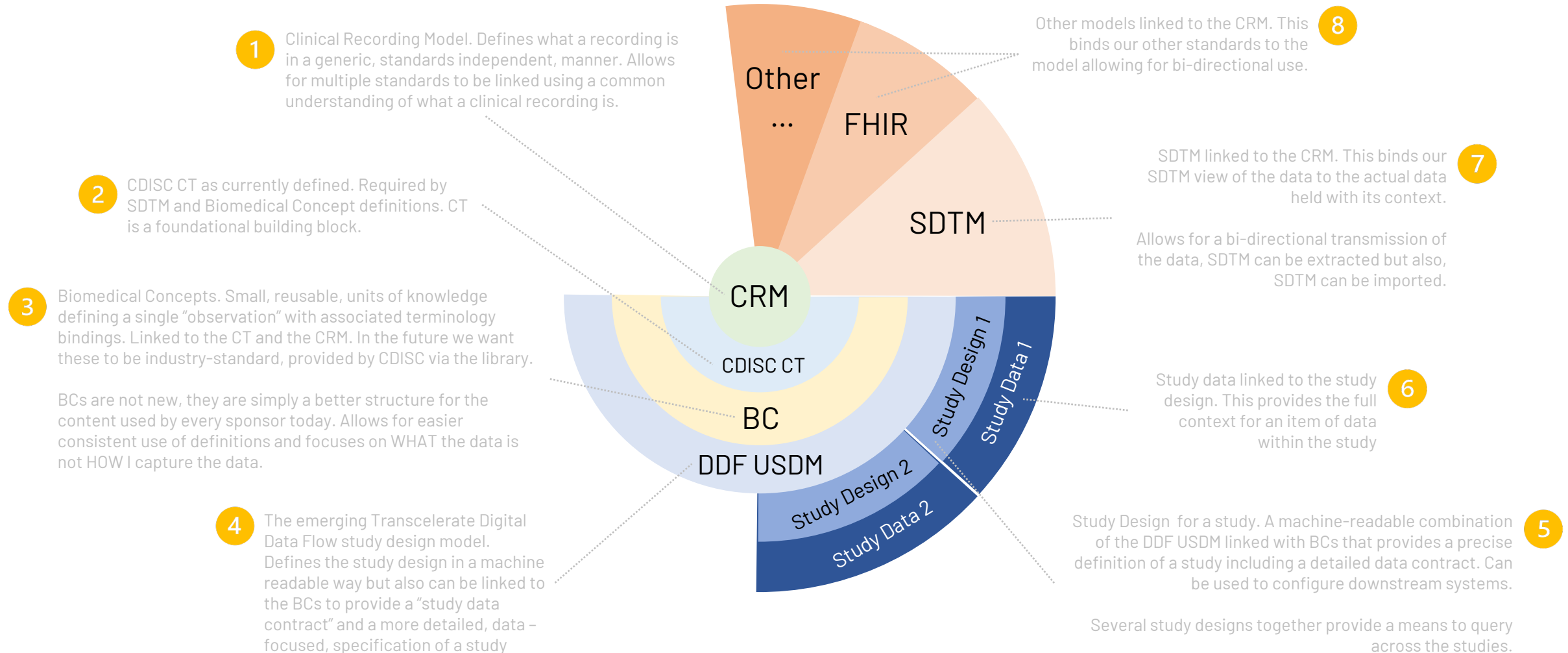
INVTNAM BRTHDTC AGE AGEU SEX RACE ETHNIC ARMCD \
0 Dr Macron 28 YEARS M WHITE NOT HISPANIC OR LATINO
1 Dr Smith 37 YEARS M ASIAN NOT HISPANIC OR LATINO
2 Dr Smith 62 YEARS M ASIAN NOT HISPANIC OR LATINO
3 Dr Smith 22 YEARS M WHITE NOT HISPANIC OR LATINO
4 Dr Macron OTHER
5 Dr Macron MULTIPLE

ARM ACTARMCD ACTARM ARMNRS ACTARMUD COUNTRY DMDTC DMDY \
0 FRA 2022-02-15T11:17:00
1 GBR 2022-02-13T13:27:00
2 GBR 2022-02-12T10:17:00
3 GBR 2022-02-11T09:57:00
4 FRA 2022-02-17T13:17:00
5 FRA 2022-02-16T12:17:00

RACE0TH RACE1 RACE2
0
1
2
3
4 Jedi
5 ASIAN WHITE
```

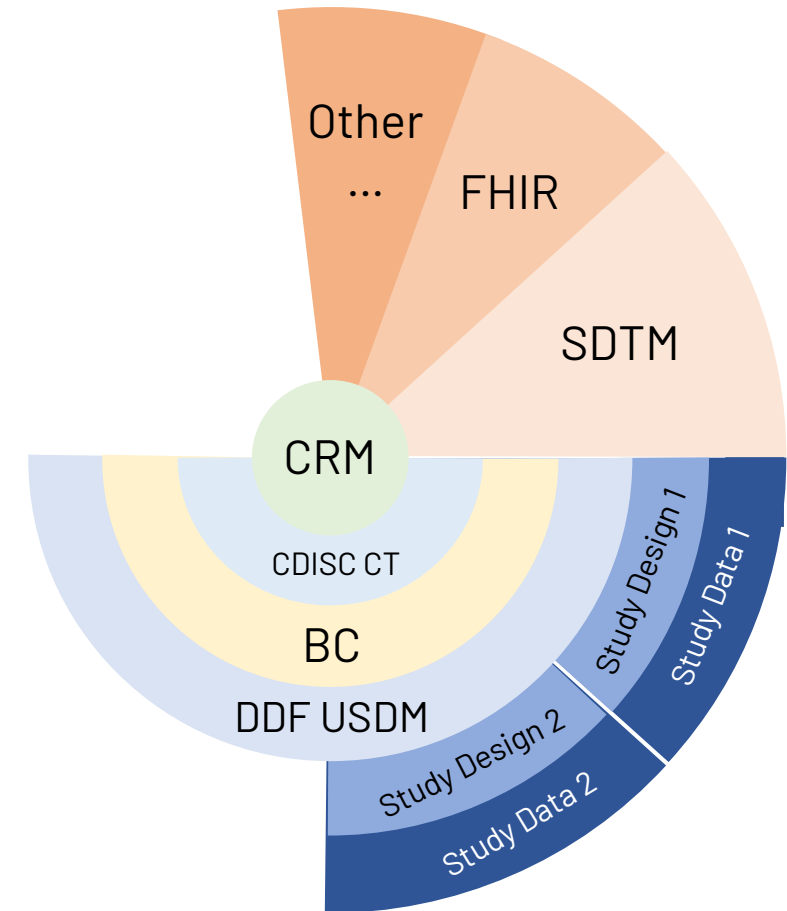
https://www.linkedin.com/posts/daveibersonhurst_biomedical-concepts-activity-6986233355474239488-pMRj/

The Model



Example Use Cases

Data Capture	Automate the setup of data capture systems ,incl. RWE, and capture the data.	★
SoA	Use the study design to build the FHIR SoA message.	★
Data Import	Import data from a variety of sources . Can be re-exported thus allowing for conversion across versions.	
Common Protocol Template (CPT)	Generation of the CPT from a study design.	
Data Decay	Re-import data using the USDM as a framework to rebuild a study design & data using the SDTM Trial Design Domains.	
Scoring	The “scoring” of a study for such purposes as site impact, subject impact, environmental impact etc.	
Feasibility	The use of the design to determine study feasibility including subject recruitment. A study data template.	
CT Registry	The provision of study information to a CT registry.	
FAIR Data	The use of the design to aid Findability, Accessibility, Interoperability, and Reusability.	



CTMS, TMF ...	The provision of protocol information to down stream systems needing “study” information.
Query	Having multiple studies that have a common structure allows for data export and query across the set of studies

PHUSE, FHIR & 2018

A3 FHIR/SDTM Demonstration (dev): Ventoux (v0.4.0) Only User [Read, Edit, System Admin]

Study Form EHR patients Study Graph SDTM Domain

Domain List

Show 10 entries Search:

Identifier	Label		
VS Domain	Vital Signs	Add	Tabulation

Showing 11 to 11 of 11 entries Previous 1 2 Next

Tabulation

STUDYID	DOMAIN	USUBJID	VSSEQ	VSGRPID	VSSPID	VSTESTCD	VSTEST	VSCAT	VSSCAT	VSPOS	VSORRES	VSORR
VENTOUX	VS	VENTOUX-SMART-9995679	1			TEMP	Temperature					
VENTOUX	VS	VENTOUX-SMART-981968	2			TEMP	Temperature				36.05556	C
VENTOUX	VS	VENTOUX-SMART-2354220	3			TEMP	Temperature				36.61111	C
VENTOUX	VS	VENTOUX-SMART-1213208	4			TEMP	Temperature				36.83333	C
VENTOUX	VS	VENTOUX-SMART-1272431	5			TEMP	Temperature				37.55556	C

Prototype system that could ingest a form definition, build a mini study design based in it, populate it with data from an EHR and extract partial SDTM domains.

This did not use the concept of the CRM.

PhUSE EU Connect 2018

Paper SI12

Into the Fire, Linking CDISC & FHIR

D Iberson-Hurst, Assero Limited & A3 Informatics ApS

ABSTRACT

The FDA have announced that “LOINC codes will be required in NDAs, ANDAs, and BLAs for studies that start after March 15, 2020” [1]. Meanwhile talk continues over integrating clinical research with healthcare and interest expressed in Fast Healthcare Interoperability Resources (FHIR), the new Health Level 7’s (HL7) standard.

This paper (and the associated presentation) will report on research encompassing:

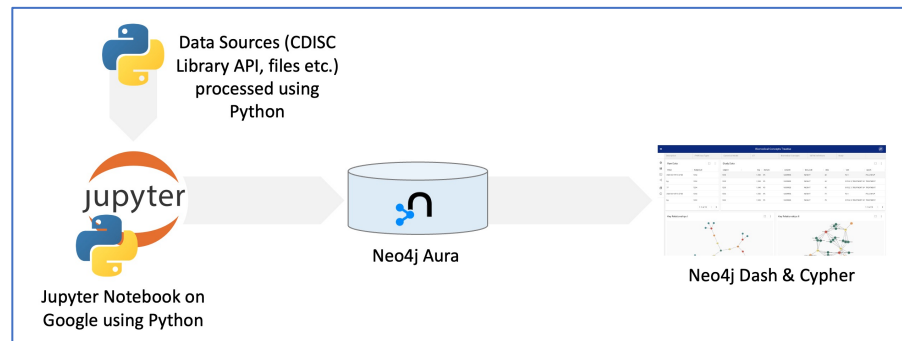
- Mappings from CDISC terminology to LOINC to aid in the population of SDTM datasets.
- Generate SDTM datasets from data held in a database with that data being extracted from an Electronic Health Record (EHR) via FHIR.
- Can the data extracted from EHRs be driven by research artefacts such as forms based on biomedical concepts?

Paper

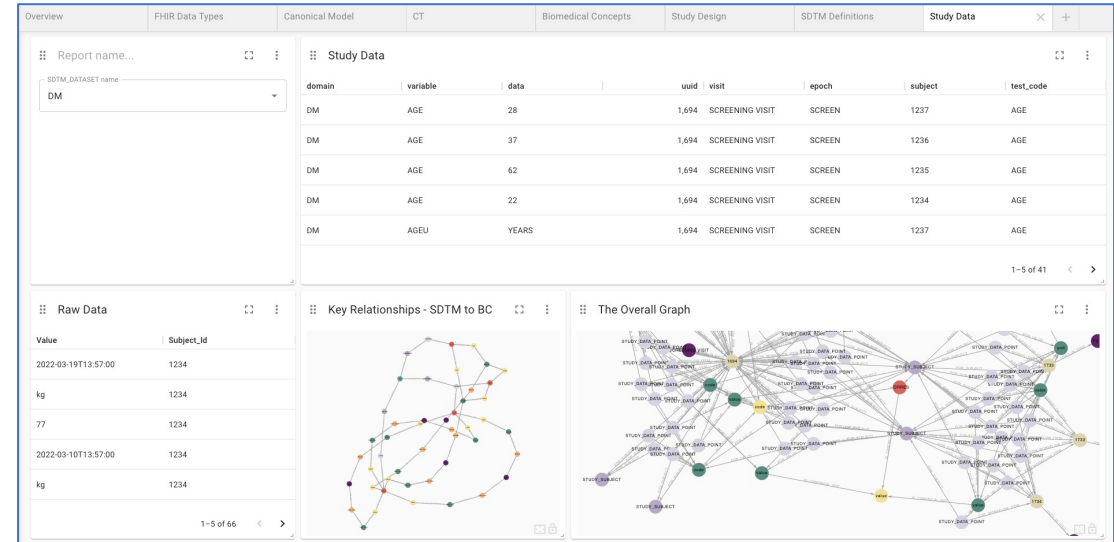
<https://www.lexjansen.com/phuse/2018/si/SI12.pdf>

Initial Proof of Concept

- A combination of:
 - Jupyter notebook
 - Neo4j database
 - Neo4j Dash dashboard
- Build just enough to make it real and prove the concepts
- Generation of DM domain, findings, events and interventions
- Contains examples of derived data, supplemental qualifiers



Neo4j dashboard using the NeoDash tool. A “no code” mechanism for viewing data

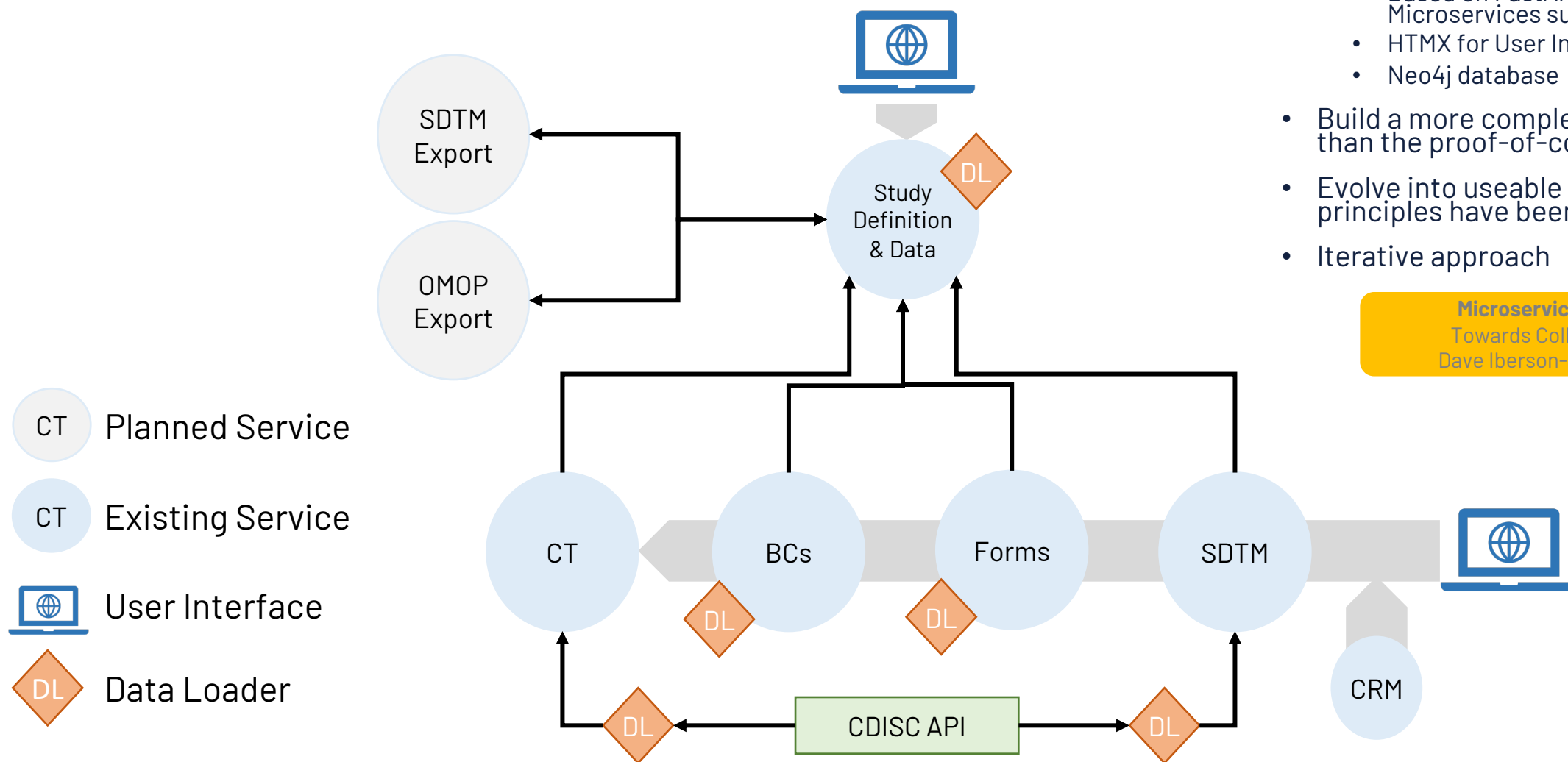


1	Dr Smith	37	YEARS	M	ASIAN	NOT HISPANIC OR LATINO
2	Dr Smith	62	YEARS	M	ASIAN	NOT HISPANIC OR LATINO
3	Dr Smith	22	YEARS	M	WHITE	NOT HISPANIC OR LATINO
4	Dr Macron				OTHER	
5	Dr Macron				MULTIPLE	

ARM	ACTARMCD	ACTARM	ARMNRS	ACTARMUD	COUNTRY	DMDTC	DMDY
0					FRA	2022-02-15T11:17:00	
1					GBR	2022-02-13T13:27:00	
2					GBR	2022-02-12T10:17:00	
3					GBR	2022-02-11T09:57:00	
4					FRA	2022-02-17T13:17:00	
5					FRA	2022-02-16T12:17:00	

RACE0TH	RACE1	RACE2
0		
1		
2		
3		
4	Jedi	
5	ASIAN	WHITE

Microservices Demonstrator



- Technical Aspects:
 - Developed in Python
 - A series of cooperating micro services, each tasked with a focused area of functionality.
 - Based on FastAPI for API / Microservices support
 - HTMX for User Interface work
 - Neo4j database
- Build a more complete demonstrator than the proof-of-concept
- Evolve into useable toolsets once principles have been proved
- Iterative approach

Microservices Paper
Towards Collaboration
Dave Iberson-Hurst, 2021

The FHIR Use Case – 2018

- Focused on
 - Using BC & Form metadata to drive the EHR request (CDISC test code -> LOINC)
 - The CT challenge of results not being in CDISC CT (e.g. UCUM units)
- Used bespoke code to “convert” / “map” FHIR resources to target format



A CRF constructed from BCs. BC metadata used to translate form data requirements into LOINC codes with CT mapping on the return to place the data into a study data warehouse for extraction as SDTM

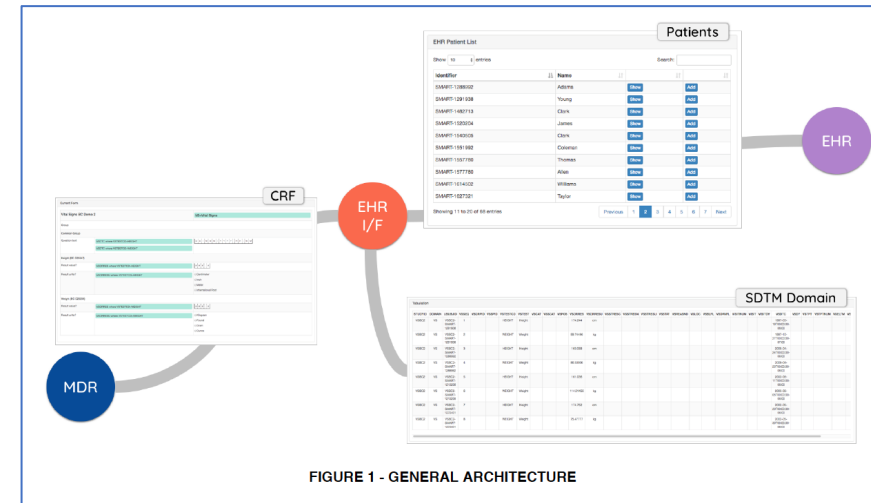


FIGURE 1 - GENERAL ARCHITECTURE

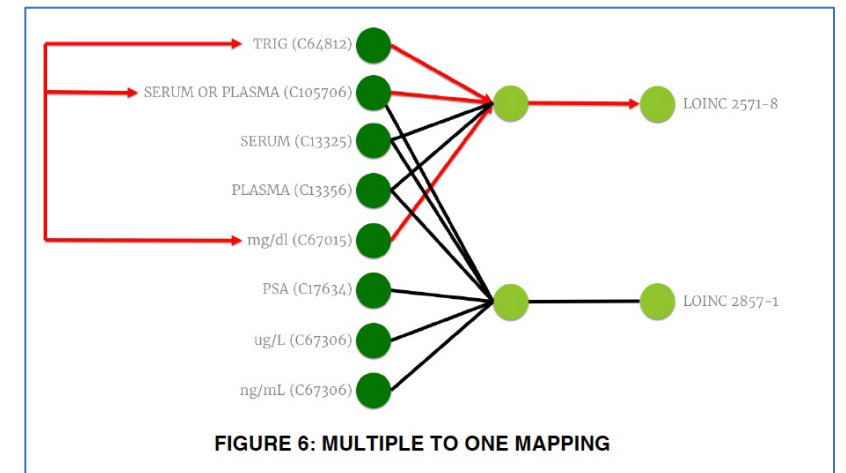


FIGURE 6: MULTIPLE TO ONE MAPPING

The FHIR Use Case – Proof of Concept

- Much more focused on using the model approach. Combination of
 - Clinical Recording Model (CRM)
 - BC model linked to the CRM
 - FHIR resource model linked to the CRM
- Used generic, model driven, approach to “convert” / “map” FHIR resources to the target

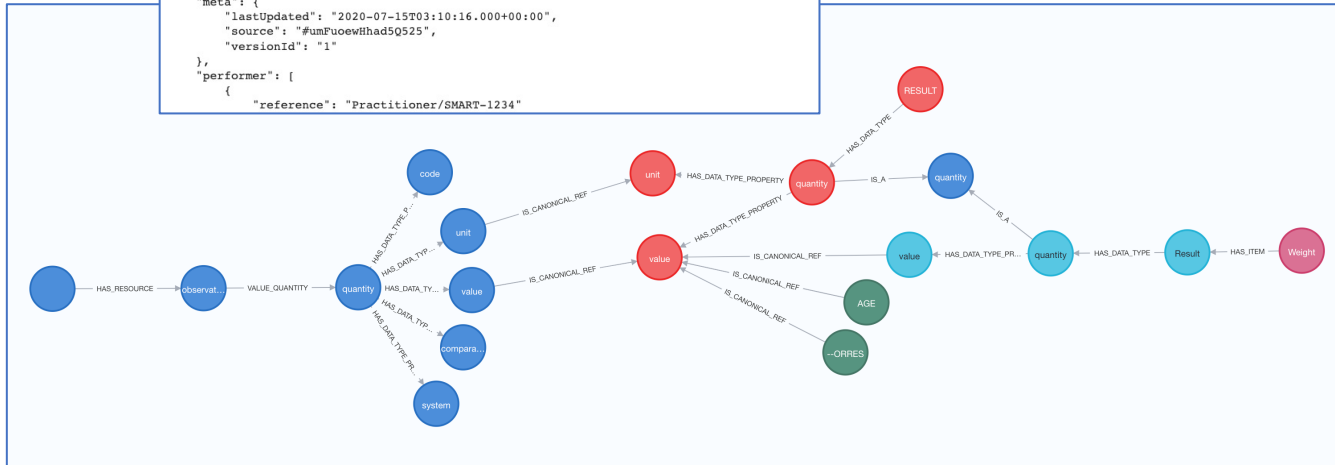


A study design constructed from BCs. BC metadata used to translate form data requirements into LOINC codes. Less attention on the CT aspects

```
EHR Patient ID: smart-981968

FHIR Response
=====

{
  "category": {
    {
      "coding": [
        {
          "code": "vital-signs",
          "display": "Vital Signs",
          "system": "http://terminology.hl7.org/CodeSystem/observation-category"
        }
      ],
      "text": "Vital Signs"
    }
  },
  "code": {
    "coding": [
      {
        "code": "3141-9",
        "display": "weight",
        "system": "http://loinc.org"
      }
    ],
    "text": "weight"
  },
  "effectiveDateTime": "1997-04-25",
  "id": "smart-3095-weight",
  "identifier": [
    {
      "system": "http://www.bmc.nl/xorgportal/identifiers/observations",
      "use": "official",
      "value": "smart-3095-weight"
    }
  ],
  "meta": {
    "lastUpdated": "2020-07-15T03:10:16.000+00:00",
    "source": "#umFuoeWthad5Q525",
    "versionId": "1"
  },
  "performer": [
    {
      "reference": "Practitioner/SMART-1234"
    }
  ]
}
```



The FHIR Use Case – Microservices

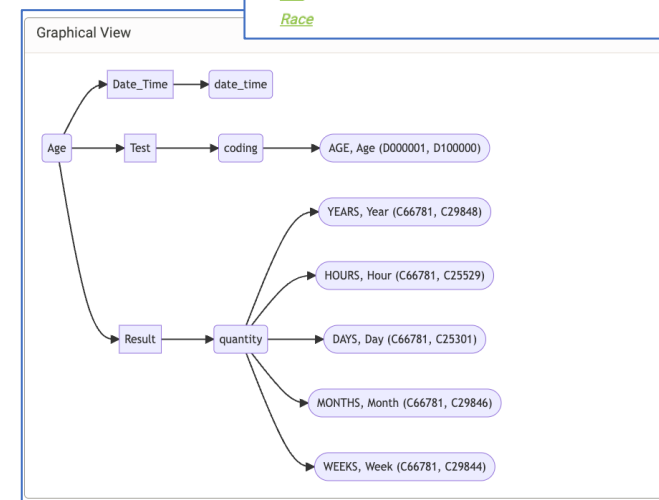
- Focused on use cases
 - Easing (semi-automated) data import of varied input formats (typically Excel) into a study design
 - SDTM generation
- Using the DDF USDM Study Design model linked with BCs



The FHIR work is not the current focus but the generic facility will be added based on the previous two prototypes / proof of concept with improved models. But the system contains a CRM service to provide the CRM support needed. BCs and SDTM are linked to the CRM.

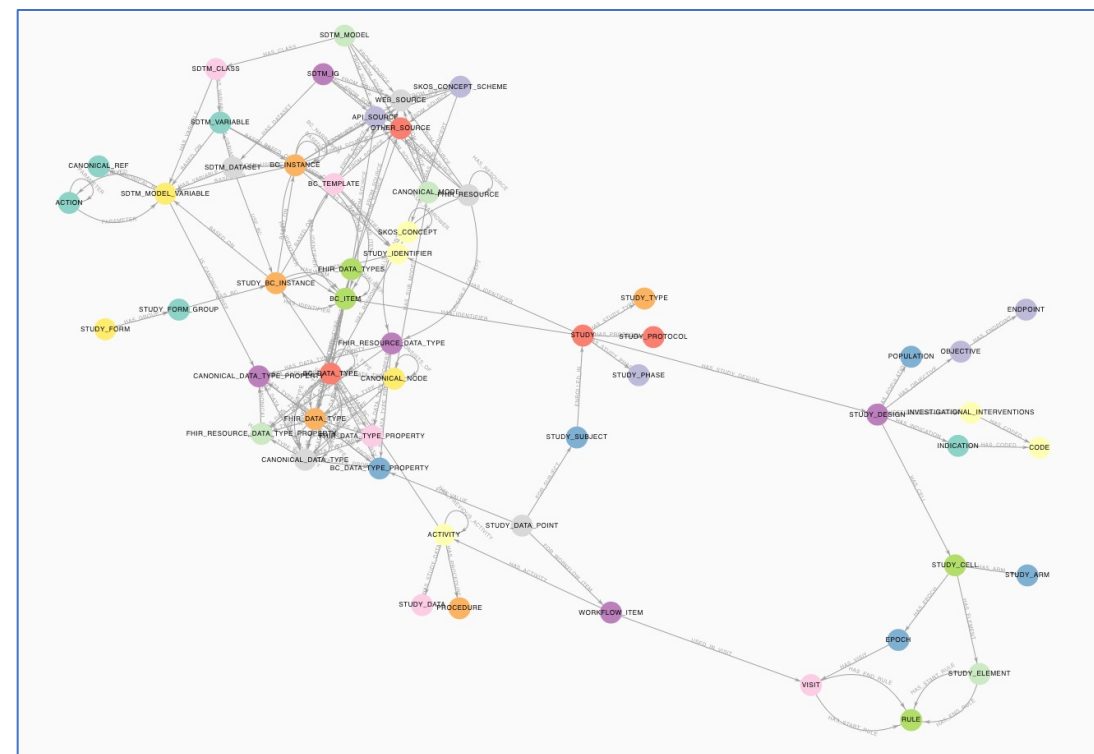
Study Tocilizumab in Patients With Severe COVID-19 Pneumonia - WA42380										
Owner: , Version: 0.1.0										
	Screening	Treatment	Treatment	Treatment	Treatment	Treatment	Treatment	Treatment	Treatment	Treatment
	Screening	Baseline	Day 1	Day 2A	Day 2B	Day 3	Day 4	Day 5	Day 6	Day 7
Informed consent	◇ X									
Inclusion/exclusion criteria	◇ X	X								
Demographics	◇ X									
Randomization	◇	X								
Medical history	◇	X								
Physical examination	◇ X									
Weight	◇	X								
COVID-19 diagnosis	◇ X									

	Screening	Treatment	Treatment	Treatment	Treatment	Treatment	Treatment
	Screening	Baseline	Day 1	Day 2A	Day 2B	Day 3	Day 4
Informed consent	◇ X						
Inclusion/exclusion criteria	◇ X	X					
Demographics	◇ X						
Age							
Sex							
Race							



Next Steps

- Continue with the microservices demonstrator work to complete:
 - Semi automation of data import
 - Automated generation of SDTM domains
- Then ...
 - Import of data from other technologies, incl. FHIR with associated CT Mappings in a technology agnostic manner
 - General Controlled Terminology issues
 - Observational Medical Outcomes Partnership Common Data Model (OMOP) Export
 - Associated Person
 - FAIR data
 - Update the paper!



Summary



di@data4knowledge.dk



<https://www.linkedin.com/in/daveibersonhurst/>

- It is time to look towards automation, including RWE / FHIR, using technology agnostic models
- Pieces are beginning to fall into place, e.g. the DDF work provides the study design piece
- Move to a data-centric approach supported by better models
- As we link up our world we see the complexity
- Controlled Terminology will always be one of the significant challenges

Data Centric

