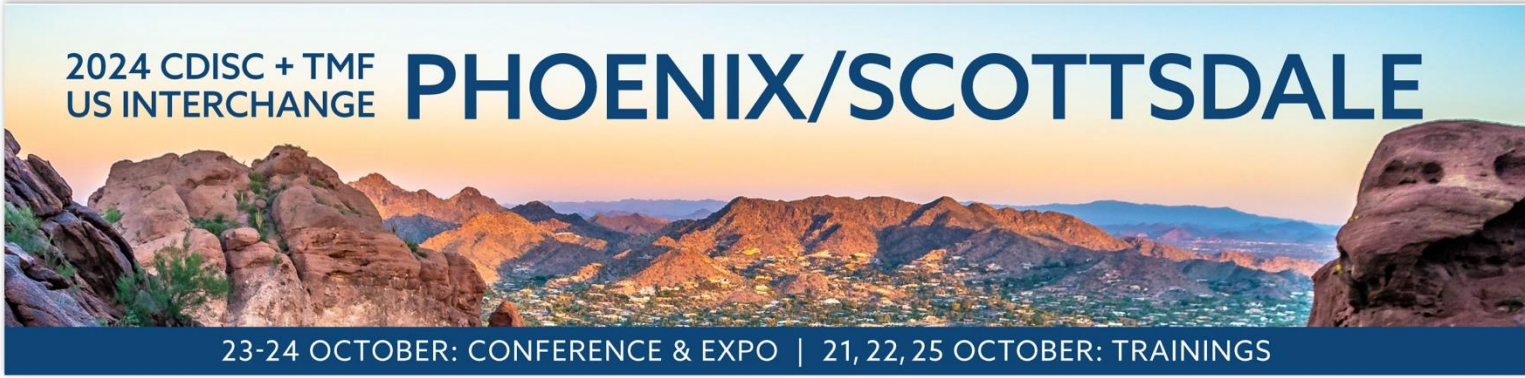


From Source to Submission: Getting the Best of Multiple Standards

Berber Snoeijer, Manager Innovative Process and Solution Design, ClinLine

Meet the Speaker

Berber Snoeijer

Title: Manager Innovative Process and Solution Design

Organization: ClinLine

Berber has a degree in Biomedical Sciences and more than 25 years of experience in clinical research. She owned a Clinical Research Organisation from 2001 to 2008, where she established and managed business processes, created solutions and standards to streamline work, and collaborated with her team on numerous phase 1 to phase 4 trials in various therapeutic areas. Following that, she worked as an R&D manager in a real-world data institute, unlocking the potential of real-world data and designing efficient and reliable tools for this purpose. In 2018, Berber founded ClinLine, which focuses on optimizing the clinical trial data process. Drawing on stakeholder input and requirements, she provides input and designs for data structures, solutions, and process optimization.

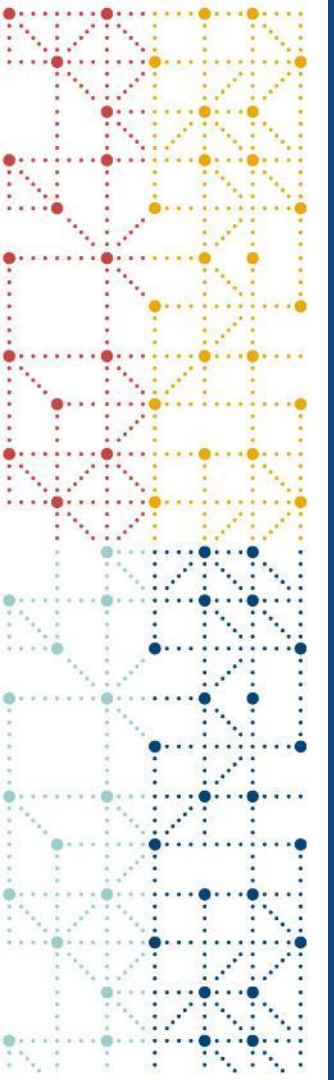
Disclaimer and Disclosures

- *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.*

- *{Please disclose any financial relationship or conflict of interest relevant to this presentation here OR}*
- *The author(s) have no real or apparent conflicts of interest to report.*

Agenda

1. Introduction
2. Data Standards and Approach
3. Standardization Practice
4. Use case and experiences
5. Summary



Introduction

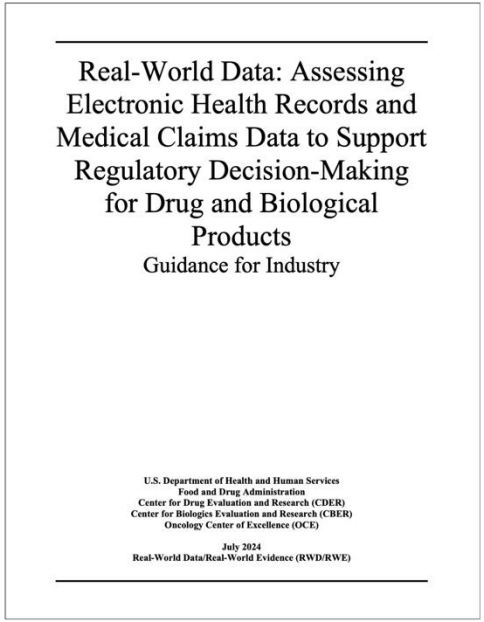
Real-World Data and Standards

FDA's Definition of Real-World Data:

Data that is related to patient health status or the delivery of health care routinely collected from a variety of sources.

Variety of Electronic Health Record Sources:

- General Practitioner
- Hospital Data
- Pharmacy
- Laboratory
- Other


Real-World Data: Assessing Electronic Health Records and Medical Claims Data to Support Regulatory Decision-Making for Drug and Biological Products

Guidance for Industry

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)
Oncology Center of Excellence (OCE)

July 2024
Real-World Data/Real-World Evidence (RWD/RWE)

Electronic Health Records, Claims and Registries

- Electronic Health Records
  - Data collected at the source
  - Collection variety within and between domains
  - Collection variety within and between countries/regions
  - Standardization at different levels
    - During collection
    - After collection at the source (FHIR)
    - In centralized repository
- Claims are based on Electronic Health Records
  - Information needed for reimbursement
  - Records only exist if reimbursed
  - Cross-sectional
- Registries
  - Derived from EHR
  - Focussed information – disease related

Electronic Health Records

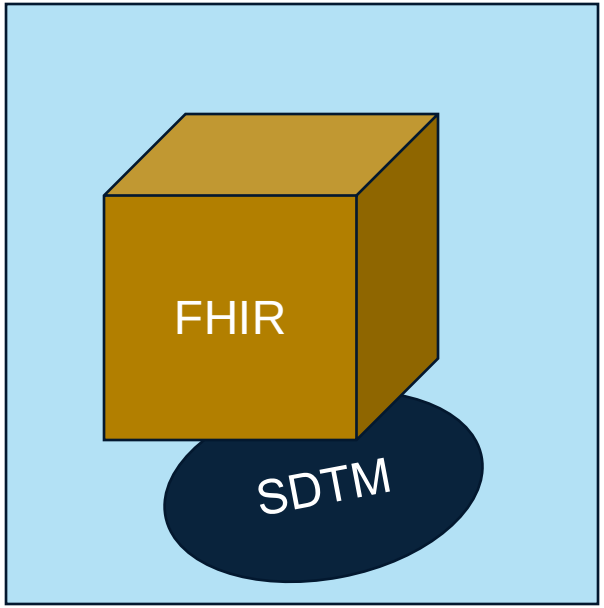
- Collection practice has effect on **reliability** and analysis
  - Data collected and stored in information system according to own specifications.
- What is collected has effect on **relevance** of data
- Standardization has effect on **reliability** and **relevance**
  - Improving interoperability
  - Traceability/lineage requirements
  - Reducing variety of information

FDA guideline: "Data in CDM-driven networks rarely contain all the source information present at the individual health care sites, and the data elements chosen for a given CDM network may not be sufficient for all research purposes or questions."

Data Standards and Approach

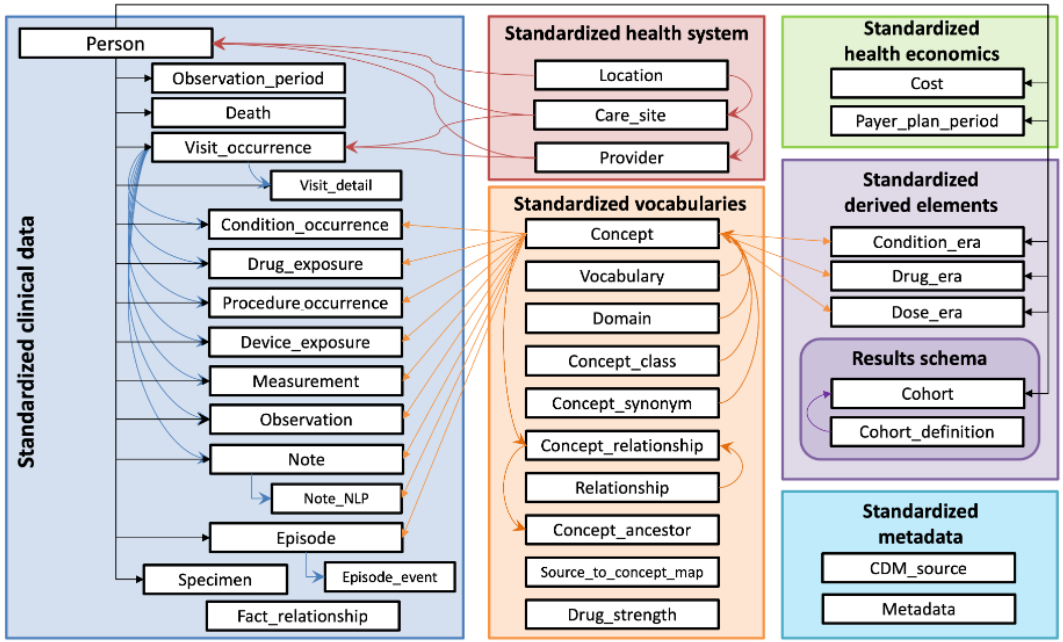
What data standard to use for real-world data?

- FHIR
  - Flexible, flat json format
  - 1 file per patient
  - Focus on routine health care information
  - Standard extraction tools are not sufficient
    - Not full set of (coding) information
    - Still a lot of input specifications needed (given the FHIR flexibility)
- SDTM
  - Clinical trial focussed
  - Not all traceability information can be included
  - Not all information needed for fit for purpose used for submission
  - Not all information that is expected for SDTM is available in EHR source
  - Required for submission
  - Clinical study design is the starting point



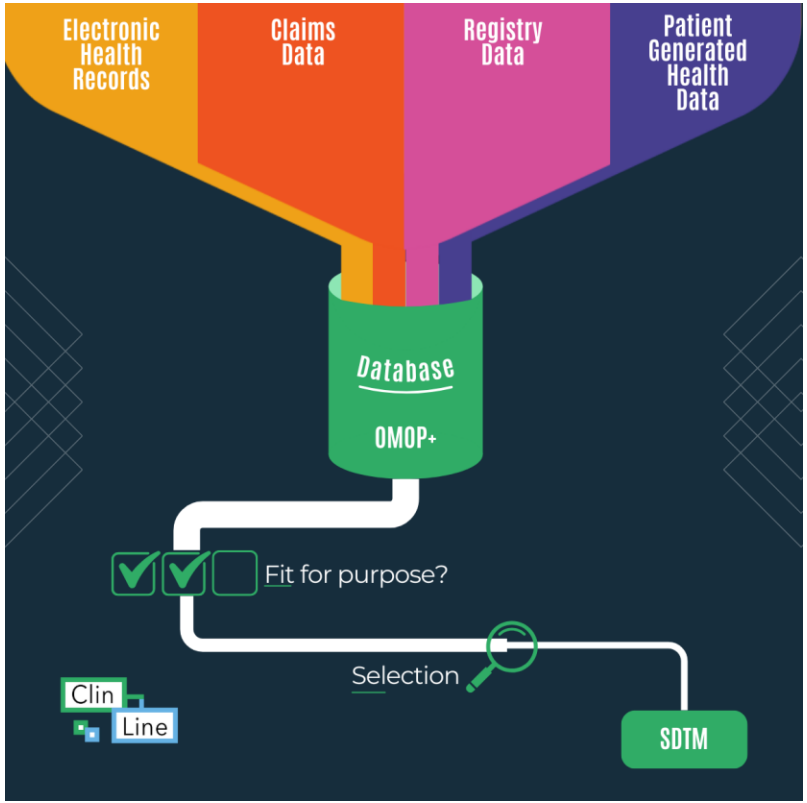
What data standard to use for real-world data?

- OMOP
  - Observational research standard
  - Basic structure - comparable to structures used at data vendor sites
  - Contains source information variables
  - Univocal: No repeated instances of same information
  - Includes mapping capabilities
    - Like SNOMED to MedDra
  - Includes child/ancestor relationships



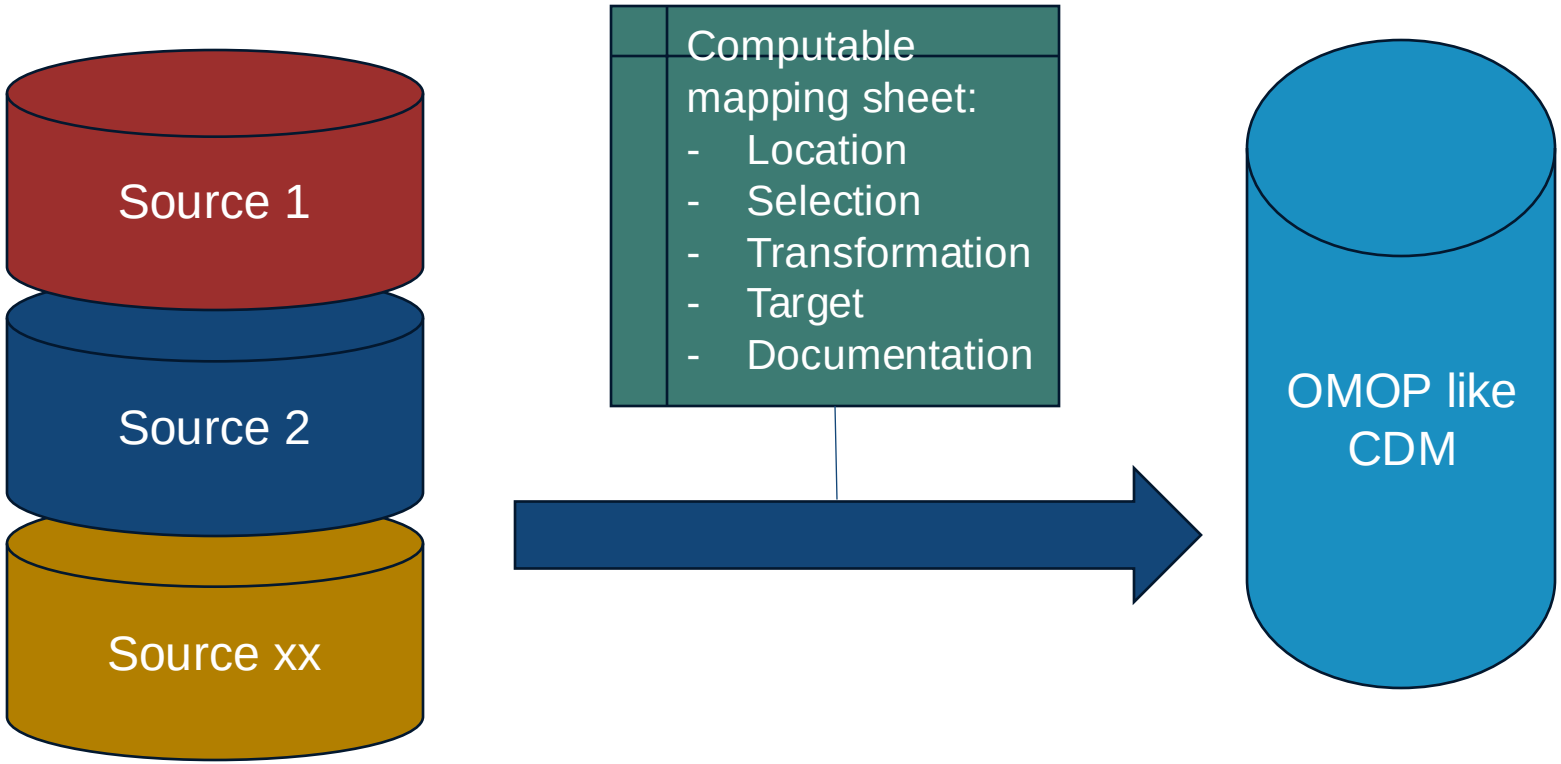
Our approach

1. Standardize from source to OMOP-like datasets
   - Reduced ds if not needed
   - Added FHIR variables if beneficial for traceability and later processes
2. Select cohort
3. Assess fit for purpose
4. Match and impute
5. Transform to SDTM
6. Study specific analysis and reporting



Standardization Practice

Data mapping automation and traceability



Data transformations

- Mapping of dictionary codes
  - Open OMOP relationship table -> Traceable
  - Automated recognition of unmapped codes
  - Text matching / recognition => suggest
  - Review unmapped / suggested items
  - Building custom – manual mapped and verified dictionary

- Calculations
  - Needed if level of detail varies between sources
    - Example: full score available in 1 source while only sub scores available in another source
  - Needed if expected (frequently used) endpoint is not at expected detail level
  - Specified in mapping sheet

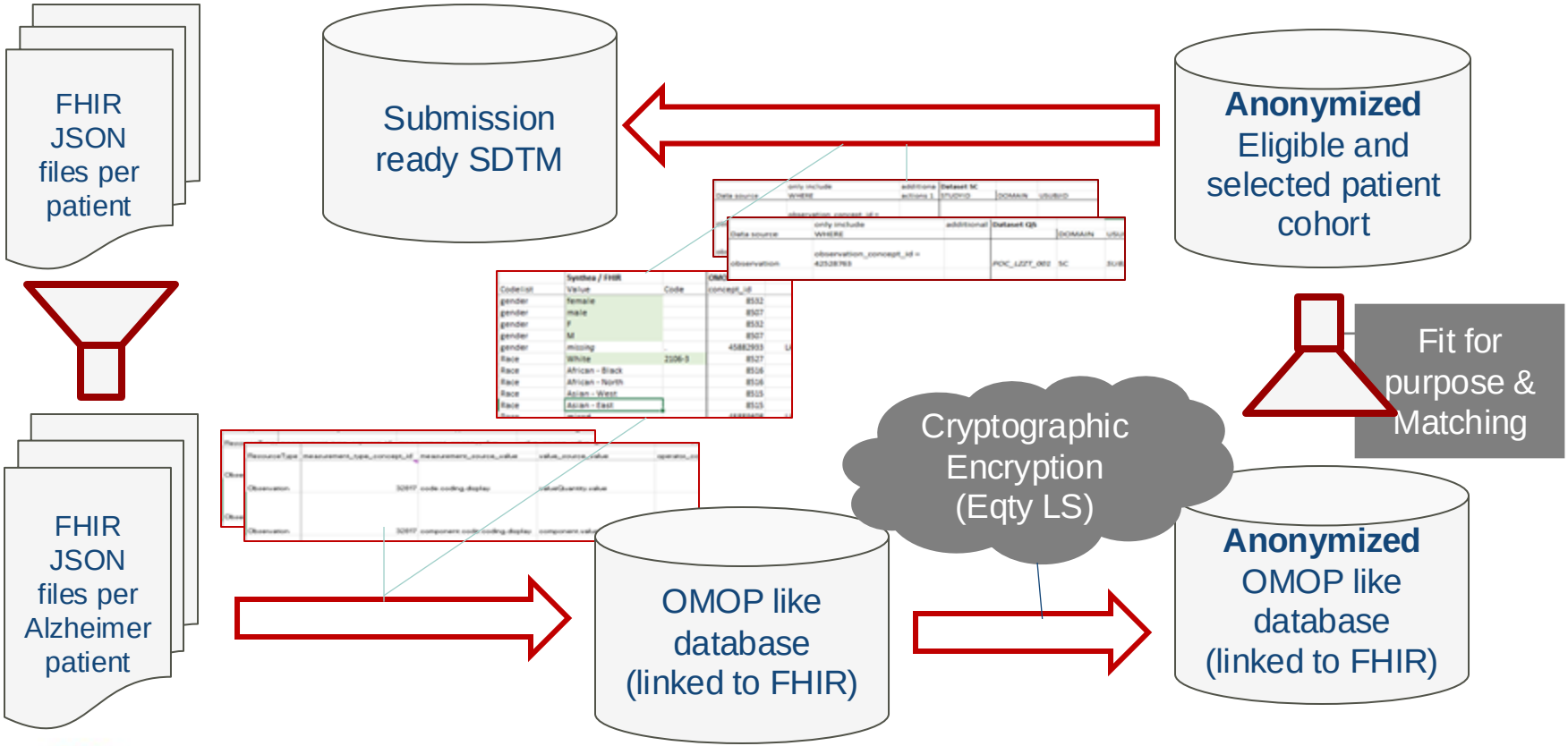
Data transformation

- Custom codelist mapping
  - Automated transformation
  - Traceable
  - Address misalignment between sources
  - Include full coding trail from source to submission
  - Address misalignment in the trail from source to submission

	Source 1		Source 2		OMOP				CDISC		
Codelist	Value	Code	Value	code	concept_id	Code	Name	Vocabulary	Codelist	Codelist code	CT
Route	p.o.	0	PO	1	4132161	26643006	Oral route	SNOMED	C66729	C38288	ORAL
Route	p.r.	1	PR	2	4290759	37161004	Rectal route	SNOMED	C66729	C38295	RECTAL
Route	s.c.	2	SC	3	4142048	34206005	Subcutaneous route	SNOMED	C66729	C38299	SUBCUTANEOUS
Route	i.m.	3	IM	4	4302612	78421000	Intramuscular route	SNOMED	C66729	C28161	INTRAMUSCULAR
Route	i.v.	4	IV	5	4171047	47625008	Intravenous route	SNOMED	C66729	C38276	INTRAVENOUS
Route	nasal	5	NASAL	6	4262914	46713006	nasal route	SNOMED	C66729	C38284	NASAL
Route	td	6	TD	7	4262099	45890007	transdermal route	SNOMED	C66729	C38305	TRANSDERMAL
Route			SL	8	4292110	37839007	sublingual route	SNOMED	C66729	C38300	SUBLINGUAL
Route			INH	9	45956874	9011000001100	Inhalation	SNOMED	C66729	C38216	RESPIRATORY (INHALATION)
Route			OTHER	10	9177	74964007	Other	SNOMED			
Route			NOT AVAILABLE	NA	45884396	LA7338-2	Not Available	LOINC			

Use case and experiences

From FHIR source to SDTM: Use case



Experiences: Standardization to OMOP-like CDM

- FHIR information nested at different levels
  - Blood pressure alone or part of a panel
  - Complexity in some resources like AllergyIntolerance
  - Awareness and separate mapping needed
- Distinction between OMOP measurement or observation data
  - Source concept review needed
- Identification of sensitive / non-proportional information
  - Source concept review needed
- Some information not accounted for in OMOP standards
  - Additional grouping variables from FHIR source.
  - SDTM like visit numbers from registries
  - Status (active/completed)
- Source values already more aligned to SDTM than to OMOP
  - Registries may be partly aligned to SDTM
- The need to map information of associated documents
  - Decode values
  - Assessment details

Experiences: Standardization to OMOP-like CDM

Library MEASUREMENT

Freeze Hide Show... Format Filter... Font... Find Go to row

Table View

	measurement_id	person_id	visit_occurrence_id	measurement_concept_id	measurement_date	measurement_datetime	measurement_source_value	value_source_value	value_as_number	value_as_concept_id	unit_c
12	12	1001	47	3024171	2023-01-12	2023-01-12T06:00:24+...	Respiratory rate	14	14		
13	13	1001	47	3012888	2023-01-12	2023-01-12T06:00:24+...	Diastolic Blood Pressure	79	79		
14	14	1001	47	3004249	2023-01-12	2023-01-12T06:00:24+...	Systolic Blood Pressure	122	122		
15	15	1001	17	3036277	2016-12-08	2016-12-08T06:00:24+...	Body Height	167.7	167.7		
16	16	1001	17	43055141	2016-12-08	2016-12-08T06:00:24+...	Pain severity - 0-10 verbal numeric...	0	0		
17	17	1001	17	3025315	2016-12-08	2016-12-08T06:00:24+...	Body Weight	83.5	83.5		
18	18	1001	17	3038553	2016-12-08	2016-12-08T06:00:24+...	Body mass index (BMI) [Ratio]	29.68	29.68		
19	19	1001	17	3027018	2016-12-08	2016-12-08T06:00:24+...	Heart rate	82	82		
20	20	1001	17	3024171	2016-12-08	2016-12-08T06:00:24+...	Respiratory rate	16	16		
21	21	1001	17	3012888	2016-12-08	2016-12-08T06:00:24+...	Diastolic Blood Pressure	91	91		
22	22	1001	17	3004249	2016-12-08	2016-12-08T06:00:24+...	Systolic Blood Pressure	134	134		

unit_concept_id	unit_source_value	measurement_source_cc	id	category_code	category_system	category_display	panel
8541	/min	9279-1	37f2397f-6f00-7c...	vital-signs	http://terminolog...	Vital signs	Respiratory rate
8876	mm[Hg]	8462-4	9e31b0cb-89ba-...	vital-signs	http://terminolog...	Vital signs	Blood pressure p...
8876	mm[Hg]	8480-6	9e31b0cb-89ba-...	vital-signs	http://terminolog...	Vital signs	Blood pressure p...
8582	cm	8302-2	ff675c97-c01c-8...	vital-signs	http://terminolog...	Vital signs	Body Height
	{score}	72514-3	a379e2acf35d-9...	vital-signs	http://terminolog...	Vital signs	Pain severity - 0-...
9529	kg	29463-7	3153e131-e099-...	vital-signs	http://terminolog...	Vital signs	Body Weight
9531	kg/m2	39156-5	197531ab-3307-...	vital-signs	http://terminolog...	Vital signs	Body mass index ...
8541	/min	8867-4	ae812848-0133-...	vital-signs	http://terminolog...	Vital signs	Heart rate
8541	/min	9279-1	eb9703ea-6e84-...	vital-signs	http://terminolog...	Vital signs	Respiratory rate
8876	mm[Hg]	8462-4	624fb9b5-e7d9-5...	vital-signs	http://terminolog...	Vital signs	Blood pressure p...
8876	mm[Hg]	8480-6	624fb9b5-e7d9-5...	vital-signs	http://terminolog...	Vital signs	Blood pressure p...

Note: this data is synthetic data generated from Synthea databases: synthea.mitre.org

Experiences: Standardization to SDTM

- SV based on index date and windowing – multiple visits within 1 window
- RFSTDTC / RFENDTC definitions needed
- Adverse events versus Medical history
  - Index date related
- Concomitant Medication versus Exposure
  - Study Objective related
- Source to CDISC dictionaries
  - LOINC is OMOP standard so LOINC to CDISC transformation can be automated
  - SNOMED to MEdDRA goes very well (>90% automated)
  - WHODrug mapping not available in OMOP
  - Unit transformations via standardized UCUM – CDISC unit codelist
  - Character results (NORMAL/ABNORMAL etc) via dictionary mapping sheet

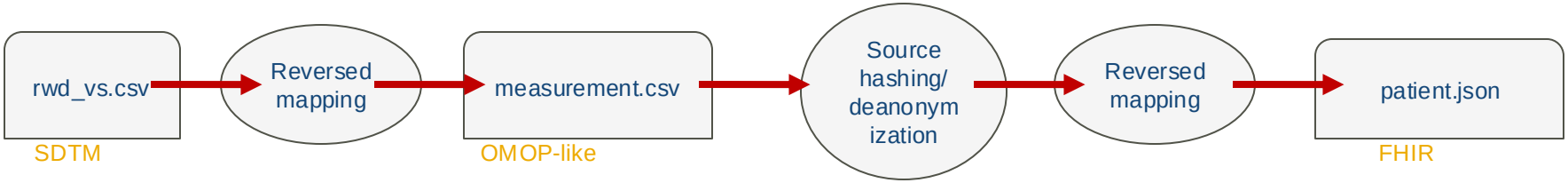
Experiences: Standardization to SDTM

	STUDYID	DOMAIN	USUBJID	VISITNUM	VISIT	VISITDY	SVSTDTC	SVENDTC	SVSTDY	SVENDY
1	POC_LZZT_001	SV	POC_LZZT_001-2003	3	VISIT 3	1	2022-03-30	2022-03-30	1	1
2	POC_LZZT_001	SV	POC_LZZT_001-2003	10	VISIT 10	118	2022-07-25	2022-07-25	118	118
3	POC_LZZT_001	SV	POC_LZZT_001-2003	11	VISIT 11	127	2022-08-03	2022-08-03	127	127
4	POC_LZZT_001	SV	POC_LZZT_001-2003	11.1	VISIT 11.1	134	2022-08-10	2022-08-10	134	134
5	POC_LZZT_001	SV	POC_LZZT_001-2011	3	VISIT 3	1	2014-05-11	2014-05-11	1	1
6	POC_LZZT_001	SV	POC_LZZT_001-2011	4	VISIT 4	8	2014-05-18	2014-05-18	8	8
7	POC_LZZT_001	SV	POC_LZZT_001-2011	10	VISIT 10	113	2014-08-31	2014-08-31	113	113
8	POC_LZZT_001	SV	POC_LZZT_001-2014	3	VISIT 3	1	2025-06-26	2025-06-26	1	1
9	POC_LZZT_001	SV	POC_LZZT_001-2014	4	VISIT 4	8	2025-07-03	2025-07-03	8	8
10	POC_LZZT_001	SV	POC_LZZT_001-2014	4.1	VISIT 4.1	15	2025-07-10	2025-07-10	15	15

	STUDYID	DOMAIN	USUBJID	SCSEQ	SCTESTCD	SCTEST	SCORRES	SCSTRESC	VISITNUM	VISIT	VISITDY	SCDTC
1	POC_LZZT_001	SC	POC_LZZT_001-2003	8	EDULEVEL	Education Level	High school diplo...	High school diplo...	3	VISIT 3	1	2022-03-30T09:14:56
2	POC_LZZT_001	SC	POC_LZZT_001-2003	9	EDULEVEL	Education Level	High school diplo...	High school diplo...	11	VISIT 11	127	2022-08-03T08:08:14
3	POC_LZZT_001	SC	POC_LZZT_001-2003	22	JOBCLAS	Employee Job Class	Full-time work	Full-time work	3	VISIT 3	1	2022-03-30T09:14:56
4	POC_LZZT_001	SC	POC_LZZT_001-2003	23	JOBCLAS	Employee Job Class	Full-time work	Full-time work	11	VISIT 11	127	2022-08-03T08:08:14
5	POC_LZZT_001	SC	POC_LZZT_001-2011	8	EDULEVEL	Education Level	More than high s...	More than high s...	3	VISIT 3	1	2014-05-11T21:11:46
6	POC_LZZT_001	SC	POC_LZZT_001-2011	9	EDULEVEL	Education Level	More than high s...	More than high s...	10	VISIT 10	113	2014-08-31T19:53:38
7	POC_LZZT_001	SC	POC_LZZT_001-2011	21	JOBCLAS	Employee Job Class	Part-time or temp...	Part-time or temp...	3	VISIT 3	1	2014-05-11T21:11:46

Note: this data is synthetic data generated from Synthea databases: synthea.mitre.org

Auditable traceability




```
"resource": {
      "resourceType": "Observation"}
      "category": [ {
        "coding": [ {
          "system":
"http://terminology.hl7.org/CodeSystem/observati
on-category",
          "code": "vital-signs",
          "display": "Vital signs"
        } ]
      } ],
      "code": {
        "coding": [ {
          "system": "http://loinc.org",
          "code": "39156-5",
          "display": "Body mass index (BMI)
[Ratio]"
        } ],
        "text": "Body mass index (BMI) [Ratio]"
```

```
},
"encounter": {
        "reference": "urn:uuid:110a54bd-0ef2-1d33-
9a4d-acf6cd109be1"
},
      "effectiveDateTime": "2021-08-
23T08:19:34+01:00",
      "valueQuantity": {
        "value": 29.36,
        "unit": "kg/m2",
        "system": "http://unitsofmeasure.org",
        "code": "kg/m2"
      }
    },
    "request": {
      "method": "POST",
      "url": "Observation"
    }
  }
```

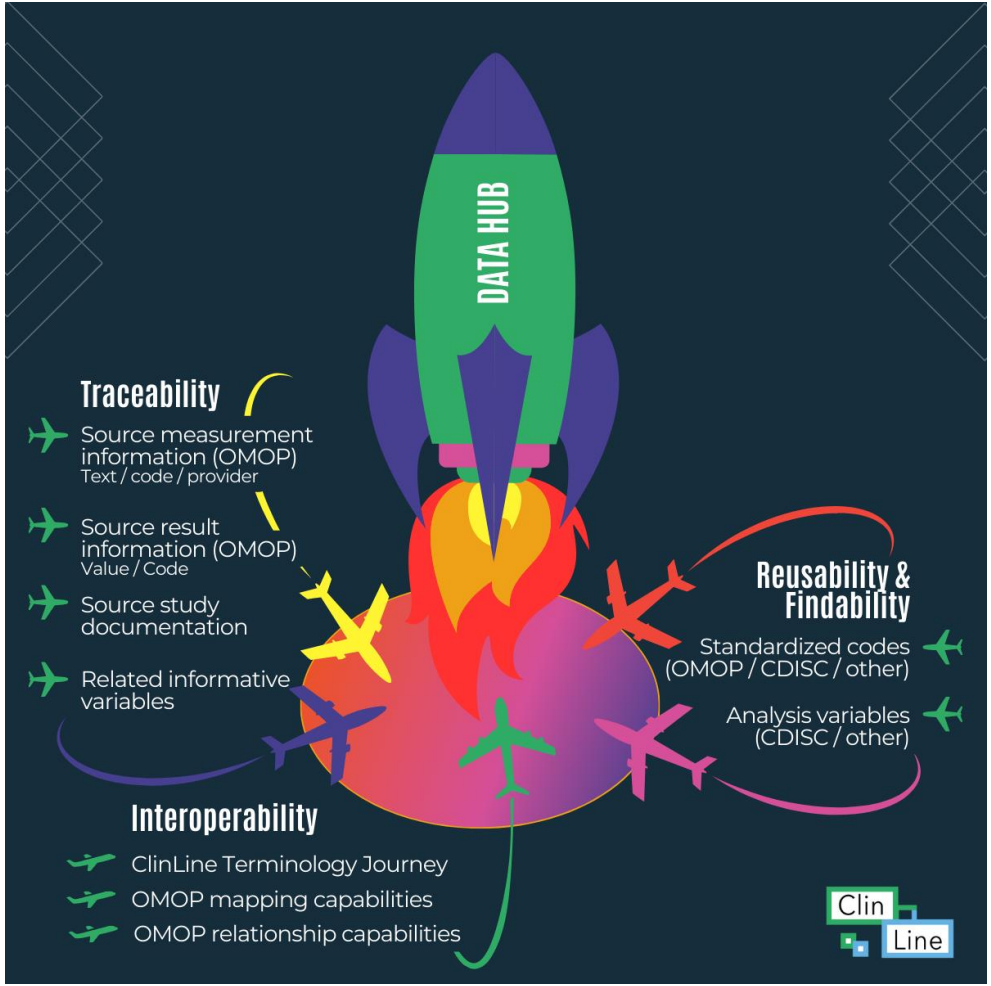
Experiences with process and programmers

- Educate programmers who are used to clinical trials
- Daily alignment and interaction needed
  - Discuss unexpected data structures and issues
  - Discuss availability of information
  - Discuss order of actions and approach
- Missing data is OK at this stage !!
- OMOP-like is not full OMOP
- Defensive programming is needed and important!
- Governance, Documentation and Lineage needed
- Include relevant source information even if only in textual documentation

Summary

Aligning with the regulatory requirements

- Traceability
  - Source documentation
  - Lineage
- Interoperability
  - Terminology Journey
  - Standardization / Mapping
  - OMOP relationships
- Reusability & Findability
  - Standardized information
  - Traceable information and transformations for analysis purposes



Thank You!

b.snoeijer@clinline.eu