Experiences in

Real-World Data Mapping

OCS LIFE SCIENCES

Clin Line

Presenters

- Caro Sluijter
  - Degree in biomedical sciences
  - Proposition lead of training at OCS Life Sciences
  - 6 years of experience in SDTM data conversion
- Berber Snoeijer
  - Degree in biomedical sciences
  - 25 year data management to reporting experience with clinical data (trial and real-world data)
  - Held management positions, business owner of ClinLine
  - Process & solution design

OCS LIFE SCIENCES

Content

- Introduction
- Data Standards and approach
- Mapping experiences
- Mapping practice
- Conclusion

Introduction

- Real-world data sources
  - Electronic Health Records
  - Claims
  - Patient Registries
  - Patient Generated Health Information
- Real-world data usage
  - Feasibility of trials and selection of patients
  - Additional evidence in regulatory trials
  - External control arm
  - Post-study safety/efficacy follow-up

OCS LIFE SCIENCES

Main concerns for using Real-World data

- Can we access the source data?

  => Privacy and Governance

- Do we know how data is collected?

  => Provenance

- Do we have documentation, audit trials and data security in place?

  => Traceability, Consistency, Reliability

- Is our data fit-for-purpose?

  => Representative, Available, Qualitive

OCS LIFE SCIENCES

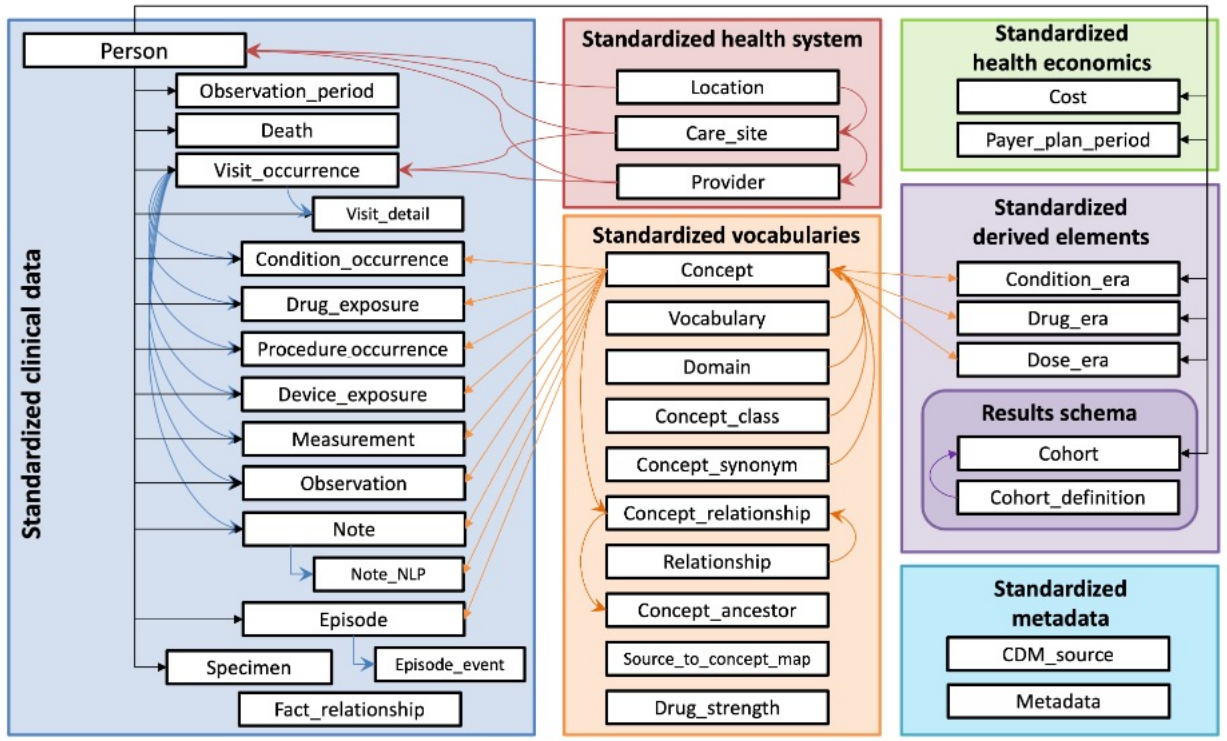
How to assess fit for purpose?

- Know your source!
  - What is information collected and what not?
  - How is information collected?
- Standardize!!
  - Source Documentation
  - Data structure
  - Controlled terminology
- Analyse
  - Representative
  - Availability of Outcomes
  - Baseline characteristics

OCS LIFE SCIENCES

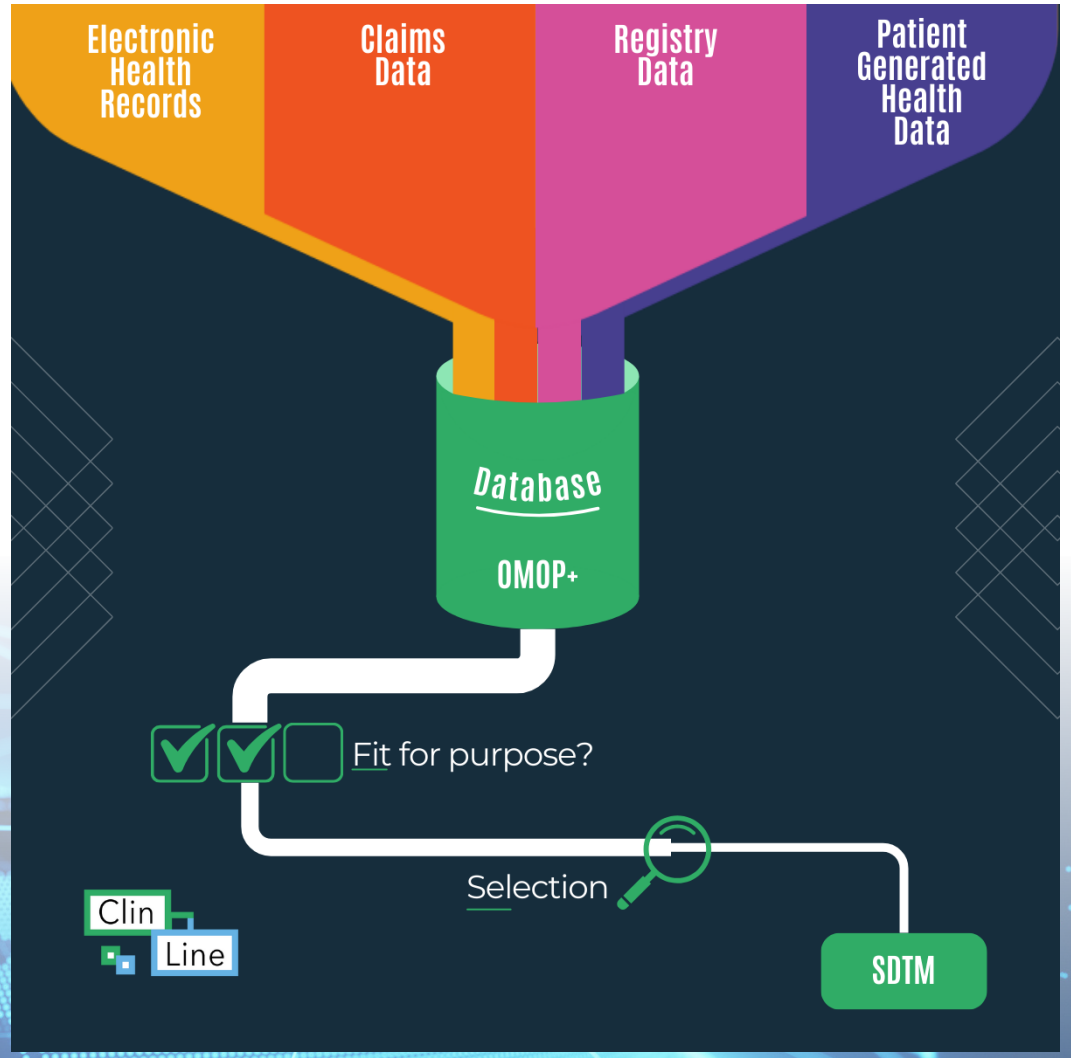
What data standard to use?

- SDTM
  - Transitions are cumbersome
  - Not all expected information available
  - Aligns to interventional data
  - Required for submissions
- FHIR
  - Flexible for exchange of data
  - Focusses on clinical practice
  - 1 JSON file per patient
- OMOP
  - Basic structure
  - Used for observational studies
  - Focusses on traceability to source



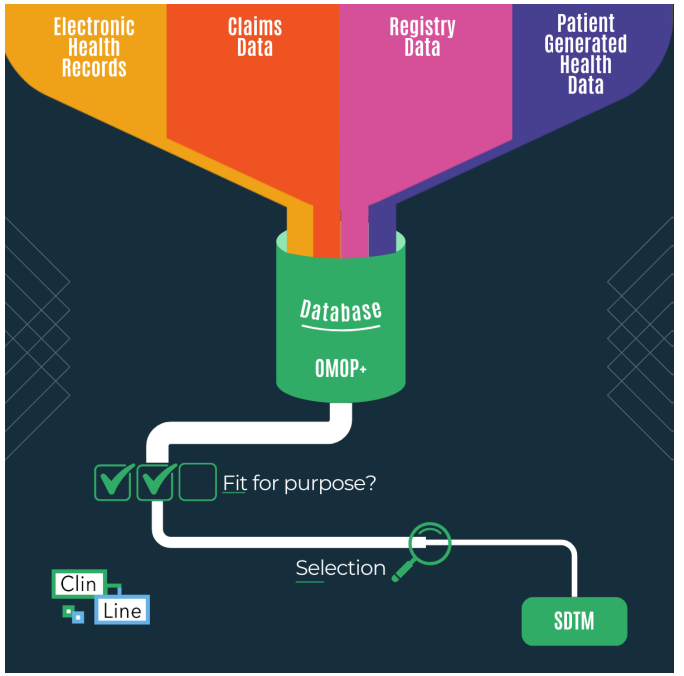
The ClinLine way

- 1: Extract
- 2: Standardize
  - OMOP +
- 3: Assess fit for purpose
- 4: Select data / match
- 5: Transform & Analyse
  - SDTM / ADAM



OCS LIFE SCIENCES

The ClinLine way



- OMOP+
  - Ensure traceability
  - Added variables to further enhance traceability
  - Keep variables needed for next steps
- Mapping process
  - Mapping sheets: automation ready
  - Mapping overview: insight in how diversity of sources map to the same concepts

	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			

Experiences

- Diversity in information captured between sources
- Diversity in documentation between and within sources
- Diversity in codelists between and within sources
- Diversity in standardization and documentation practices between regions
- Getting all the source / registry information to define provenance
- Codes not available for specific assessments
- Codes ambiguous
- Version control in mapping

OCS LIFE SCIENCES

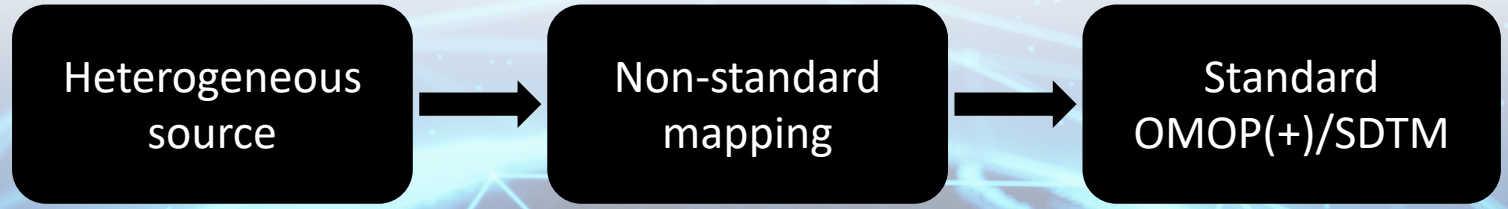
Challenges in mapping real-world data

- Heterogeneous data sources
  - Not collected according to a (global) standard
- Data quality
- Voluminous data
- Windowing
- SDTM compliance

Challenges in mapping real-world data

- Clinical trial data

- Real-world data



OCS LIFE SCIENCES

Solutions

- Pre-analysis of source data

Dataset	Variable	Label	Data Type	Value	Data source 1	Data source 2	Remarks
DEMO	RACE	Race	Char	Asian	Y	Y	
DEMO	RACE	Race	Char	African American		Y	
DEMO	RACE	Race	Char	Black or African American	Y		
DEMO	RACE	Race	Char	Other	Y	Y	
DEMO	BIRTH_YY	Birth year	Char			Y	Values are too diverse to present.
DEMO	BIRTH_YY	Birth year	Num		Y		Values are too diverse to present.

OCS LIFE SCIENCES

Solutions

- Check in e.g. OMOP/SDTM creation program for any new source variables or removal of source variables
  - Output warning in log to indicate that there is a change in source variables
  - Make automatic updates to your mapping specification file

Source_dataset	Source_variable	Target_dataset	Target_variable	Specification
SOURCE.DEMO	RACE	DM	RACE	Recode according to codelist RACE
SOURCE.DEMO	BIRTH_YY	DM	BRTHDTC	Generate ISO-8601 date variable from BIRTH_YY, BIRTH_MM and BIRTH_DD
SOURCE.DEMO	BIRTH_MM	DM	BRTHDTC	See derivation at BIRTH_YY
SOURCE.DEMO	BIRTH_DD	DM	BRTHDTC	See derivation at BIRTH_YY
SOURCE.DEMO	RACEOTH	SUPPDM	RACEOTH	Copy from source variable
SOURCE.DEMO	RACEOT			

		only include	additional	Dataset Observation (OMOP variables)				
Source	dataset	WHERE	actions	observation_source_value	observation_type_concep	value_source_value	value_as_numl	value_as_text
SOURCE 1	BASELINE		x	"Occupation"	32879	occup33	occup33	*Value(Occupation,* occup33*)*
SOURCE 2	VISITS		x	"Occupation"	32879	fcoccupation	fcoccupation	*Value(Occupation,* fcoccupation*)*
SOURCE 2	VISITS		x	"Occupation"	32879	occupatn	occupatn	*Value(Occupation,* occupatn*)*

Solutions

- Check in e.g. OMOP/SDTM creation program for any new source values
  - Use codelists
  - Output warning in log when source value is not found in codelist

Codelist	Type	From	To
RACE	C2C	Asian	ASIAN
RACE	C2C	African American	BLACK OR AFRICAN AMERICAN
RACE	C2C	Black or African American	BLACK OR AFRICAN AMERICAN
RACE	C2C	Other	OTHER
RACE	C2C		

OCS LIFE SCIENCES

Solutions

- Apply defensive programming
  - Make your program robust
  - Expect the worst of your input data (e.g. incomplete data, inconsistent data)
  - Build in checks for unexpected data (e.g. data type, out-of-range values, unpredictable values)

OCS LIFE SCIENCES

SDTM compliance – conformance issues

- Required datasets may not be applicable, e.g. AE, LB, VS, EX, DS, TA, TE, SE
- Required variables may not be applicable, e.g. SITEID, VISIT(NUM), RF-variables, ARM-variables

MH vs AE vs CE

- Retroactively define whether a condition is a:
  - Medical history event
  - Adverse event
  - Clinical event

Conclusions

- Ensure traceability and information readiness
  - Clear mapping overview
  - Documentation
  - Inclusion of traceability information in resulting OMOP+ datasets
- Account for high diversity in source data
- Utilize defensive programming
- Include SDTM logic at early stage if available
  - AE / MH
  - CM / EX
  - STDM CT

OCS LIFE SCIENCES

Questions and contact

- Questions?
- Like to learn more?
  - B.snoeijer@clinline.eu
  - Caro.sluijter@ocs-consulting.com

  - Break & Learn webinars: www.clinline.eu/break-learn/
    - A Dive Into the World of Biomedical Concepts – recorded Jan 2024
    - From Study Design to SDTM Trial Design Datasets,
      Using the USDM to Optimize the Data Flow – recorded 22 May 2024

OCS LIFE SCIENCES