AstraZeneca

Raw Data Standards as Key Element of the Clinical Data Journey

Anna Kamińska, Associate Director Standards Developer, AstraZeneca

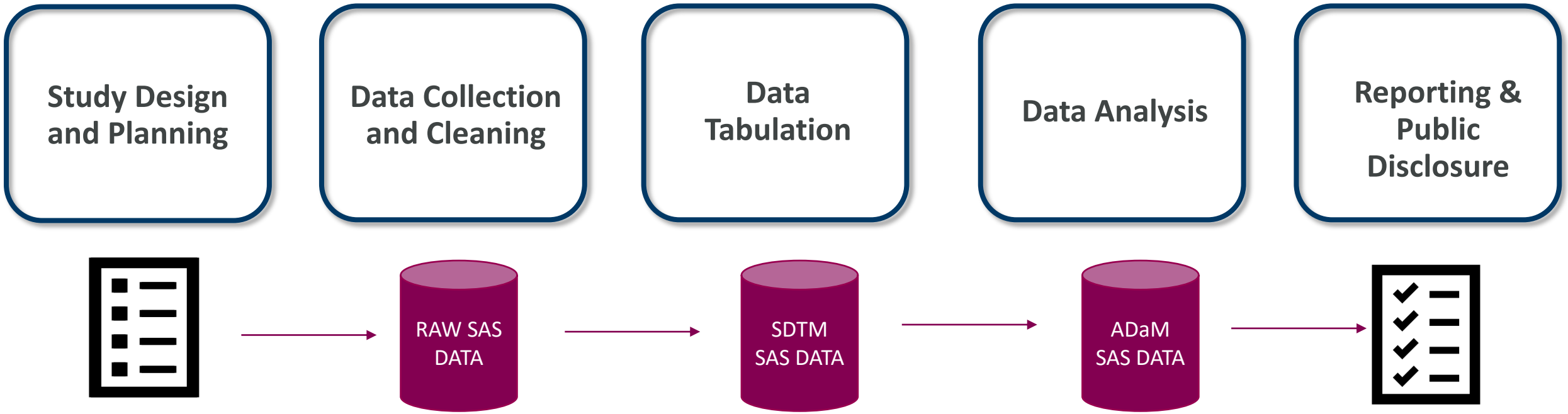
Joanna Trzos, Associate Director Standards Developer, AstraZeneca

PHUSE EVENT 2023 Utrecht, 28-09-2023

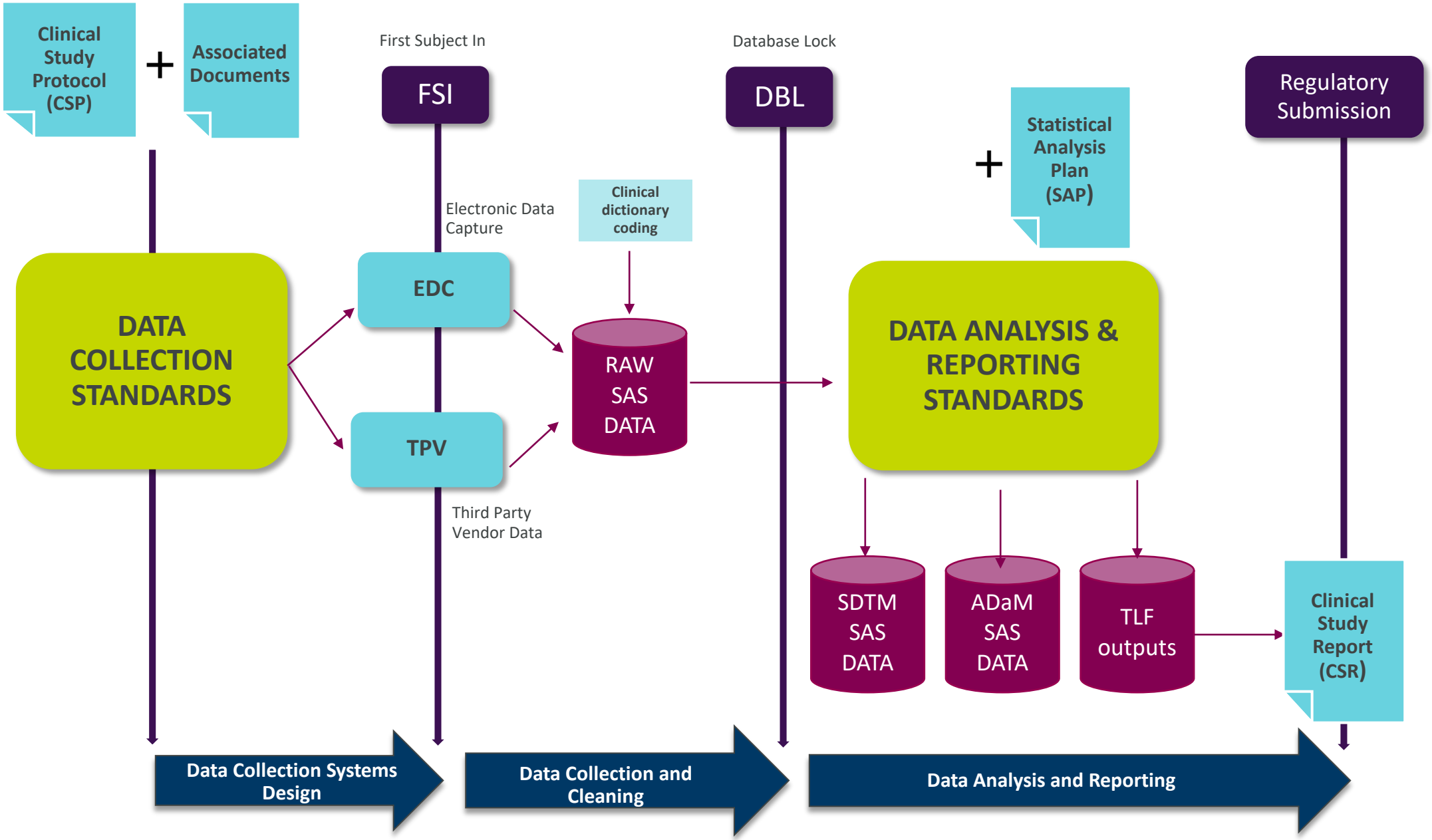
Agenda

1. Clinical Data Flow
2. Clinical Data Standards in Data Flow
3. Clinical Data Standards in AstraZeneca
4. Raw Data Standards Concept
5. End-users of Raw Data
6. Experience in Raw Data Standards Maintenance
7. Benefits
8. Challenges

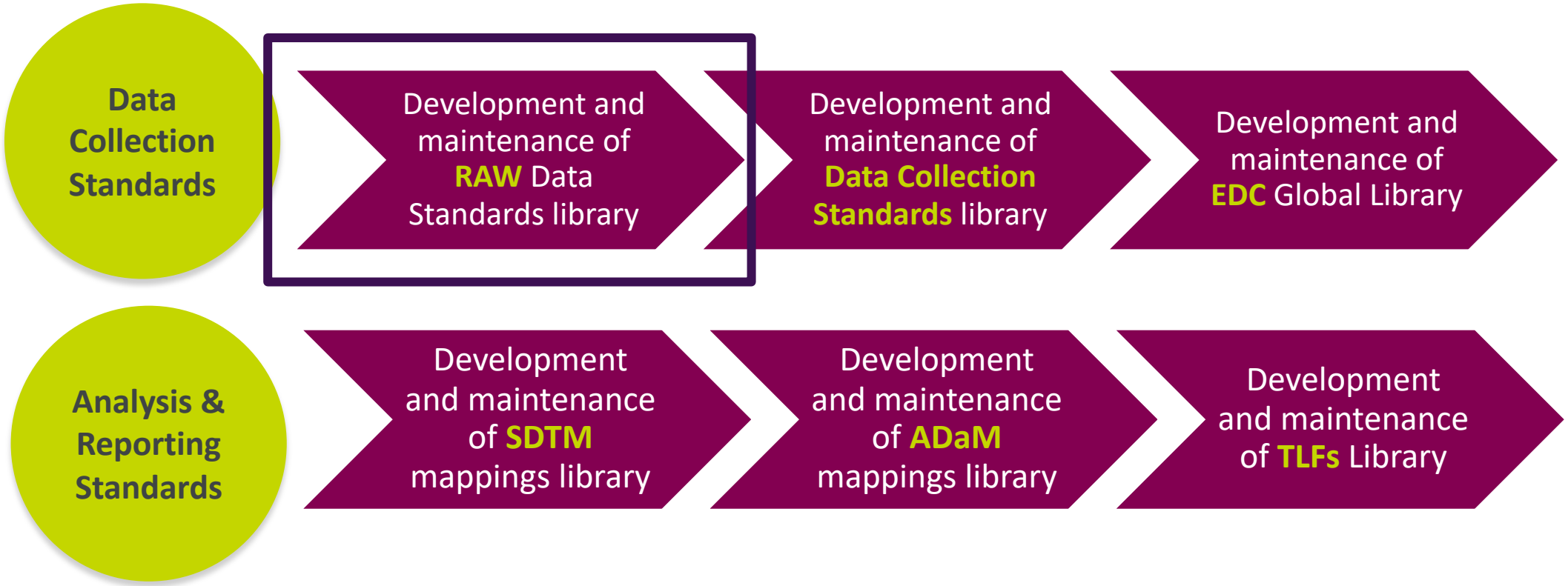
Clinical Data Flow



Clinical Data Standards in Data Flow



Clinical Data Standards in AstraZeneca



*EDC = Electronic Data Capture

* RAW Data Standards = RDS

RAW Data Standards Concept

Raw Data Standards (RDS) describe rules for RAW SAS datasets creation and compromise data collection and data tabulation phases of clinical data journey.

Metadata for entire RAW SAS dataset e.g.: structure, long name, short name

Metadata for individual variables: variable name, SAS labels, length, format, valid values, required attributes

Performing function of **data extraction specification** and standardizing SAS dataset regardless of way of data collection

Following **CDASH** rules and **support data collection standards** creation

Developing with **SDTM** in mind and **enable automations** of SDTM creation

Non-standardized RAW SAS dataset

EDC data

STUDY NUMBER 1: DXXXXC0000X

STUDY	SUBJECT	MODULE	VIS_DAT	LINE	IEYN	IECAT	IETESTCD	DSCSTDAT
DXXXXC0000X	A	IE	2018-02-10	1	C49488			2018-02-10
DXXXXC0000X	A	IE	2018-06-26	2	C49487	C25370	4	2018-06-26
DXXXXC0000X	B	IE	2018-01-03	1	C49487	C25532	2	2018-01-03
DXXXXC0000X	C	IE	2018-07-17	1	C49487	C25370	3	2018-07-17
DXXXXC0000X	D	IE	2018-10-14	1	C49487	C25370	10	2018-10-14
DXXXXC0000X	D	IE	2018-11-14	2	C49488			2018-11-14

STUDY NUMBER 2: DXXXXC1111X

STUDY	SUBJECT	MODULE	VISDAT	MODULE_O	IEYN	IECAT	IETESTCD	IESTDAT
DXXXXC1111X	A	INEX	2020-01-26	1	0			2020-01-26
DXXXXC1111X	A	INEX1	2020-11-14	2	0			2020-11-14
DXXXXC1111X	B	INEX	2020-08-25	1	1	C25532	TEXT	2020-08-25
DXXXXC1111X	C	INEX	2020-01-03	1	1	C25370	TEXT	2020-01-03
DXXXXC1111X	D	INEX	2020-02-04	1	0			2020-02-04

→ Different metadata for dataset: name and structure

→ Different metadata for variables: names, coded options and type

** Above datasets are dummy datasets created for the purpose of this presentation*

*EDC = Electronic Data Capture

Standardized RAW SAS dataset

EDC data

STUDY NUMBER 1: DXXXXC0000X

STUDY	SUBJECT	MODULE	VIS_DAT	LINE	IEYN	IECAT	IETESTCD	DSCSTDAT
DXXXXC0000X	A	IE	2018-02-10	1	C49488			2018-02-10
DXXXXC0000X	A	IE	2018-06-26	2	C49487	C25370	4	2018-06-26
DXXXXC0000X	B	IE	2018-01-03	1	C49487	C25532	2	2018-01-03
DXXXXC0000X	C	IE	2018-07-17	1	C49487	C25370	3	2018-07-17
DXXXXC0000X	D	IE	2018-10-14	1	C49487	C25370	10	2018-10-14
DXXXXC0000X	D	IE	2018-11-14	2	C49488			2018-11-14

→ The same metadata for dataset: name, structure

→ The same metadata for variables: names, coded options and type

STUDY NUMBER 2: DXXXXC1111X

STUDY	SUBJECT	MODULE	VIS_DAT	LINE	IEYN	IECAT	IETESTCD	DSCSTDAT
DXXXXC1111X	A	IE	2020-01-26	1	C49488			2020-01-26
DXXXXC1111X	A	IE	2020-11-14	2	C49488			2020-11-14
DXXXXC1111X	B	IE	2020-08-25	1	C49487	C25532	4	2020-08-25
DXXXXC1111X	C	IE	2020-01-03	1	C49487	C25370	1	2020-01-03
DXXXXC1111X	D	IE	2020-02-04	1	C49488			2020-02-04

** Above datasets are dummy datasets created for the purpose of this presentation*

*EDC = Electronic Data Capture

Non-standardized RAW SAS dataset

ePRO data

STUDY NUMBER 1: DXXXXC0000X

			EQ5D01-Mobility	EQ5D01-Self-Care	EQ5D01-Usual Activities	Evaluator
STUDY	SUBJECT	MODULE	**EQ5D0101**	**EQ5D0102**	**EQ5D0103**	**QSEVAL**
DXXXXC0000X	A	EQ_5D_3L	1	1	3	C41189
DXXXXC0000X	B	EQ_5D_3L	2	1	4	C41189
DXXXXC0000X	C	EQ_5D_3L	2	3	3	C41189
DXXXXC0000X	D	EQ_5D_3L	4	4	4	C41189

→ Dataset follows QRS naming and business rules thus created with SDTM in mind

STUDY NUMBER 2: DXXXXC1111X

↕ Datasets not standardized between each other (different metadata for variables)

			Mobility	Self-Care	Usual Activities	Evaluator
STUDY	SUBJECT	MODULE	**EQ5D1**	**EQ5D2**	**EQ5D3**	**EVAL**
DXXXXC1111X	A	EQ_5D_3L	2	2	2	1
DXXXXC1111X	B	EQ_5D_3L	4	4	4	1
DXXXXC1111X	C	EQ_5D_3L	1	1	1	1

→ Dataset does not follow QRS naming and business rules, more modifications needed on SDTM level

** Above datasets are dummy datasets created for the purpose of this presentation*

*ePRO = Electronic Patient Reported Outcome

Standardized RAW SAS dataset

ePRO data

STUDY NUMBER 1: DXXXXC0000X

			EQ5D01-Mobility	EQ5D01-Self-Care	EQ5D01-Usual Activities	Evaluator
STUDY	SUBJECT	MODULE	**EQ5D0101**	**EQ5D0102**	**EQ5D0103**	**QSEVAL**
DXXXXC0000X	A	EQ_5D_3L	1	1	3	C41189
DXXXXC0000X	B	EQ_5D_3L	2	1	4	C41189
DXXXXC0000X	C	EQ_5D_3L	2	3	3	C41189
DXXXXC0000X	D	EQ_5D_3L	4	4	4	C41189

Both datasets aligned to QRS naming and business rules thus support SDTM creation. Standardized metadata for raw dataset and variables

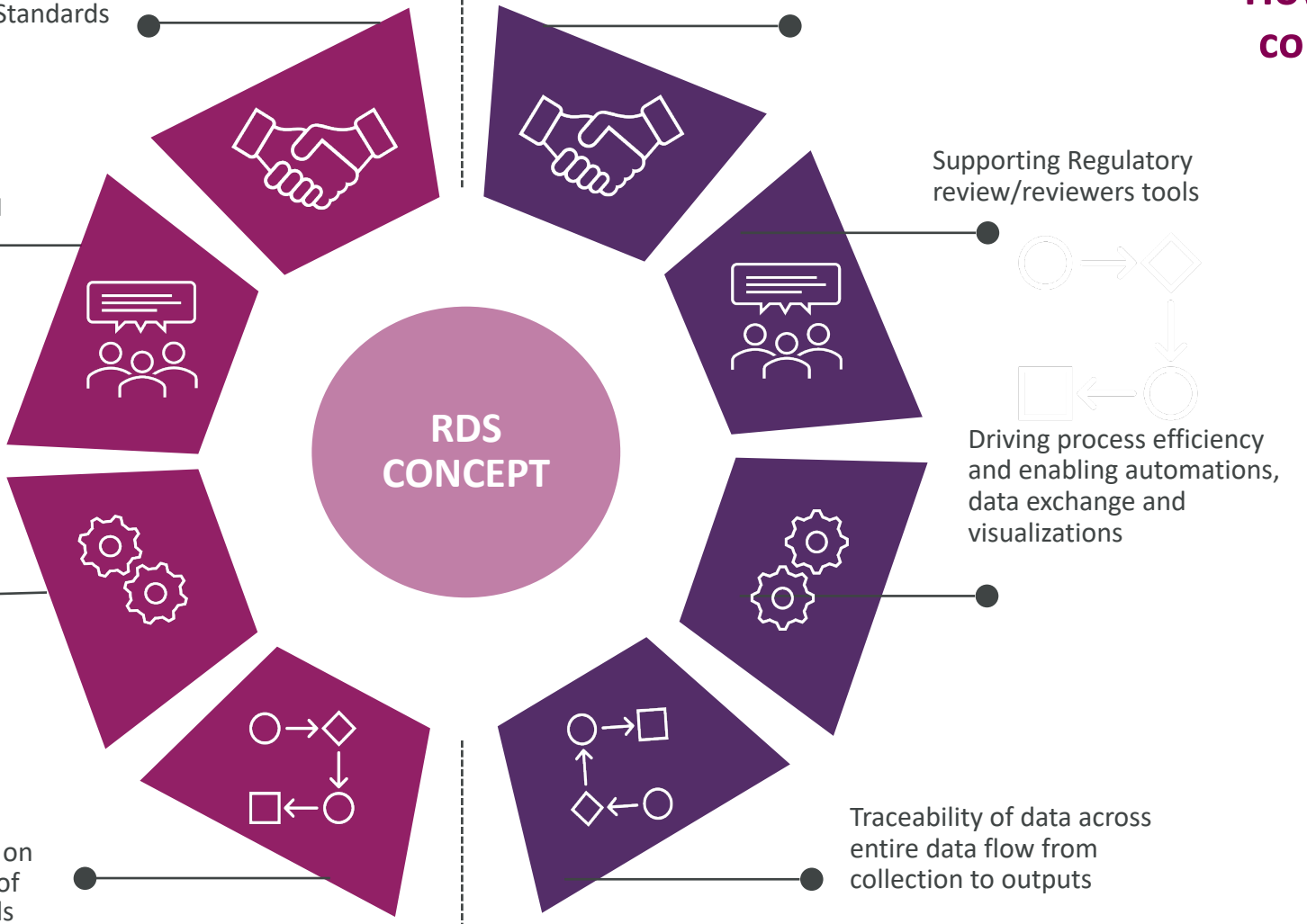
STUDY NUMBER 2: DXXXXC1111X

			EQ5D01-Mobility	EQ5D01-Self-Care	EQ5D01-Usual Activities	Evaluator
STUDY	SUBJECT	MODULE	**EQ5D0101**	**EQ5D0102**	**EQ5D0103**	**QSEVAL**
DXXXXC1111X	A	EQ_5D_3L	2	2	2	C41189
DXXXXC1111X	B	EQ_5D_3L	4	4	4	C41189
DXXXXC1111X	C	EQ_5D_3L	1	1	1	C41189

** Above datasets are dummy datasets created for the purpose of this presentation*

*ePRO = Electronic Patient Reported Outcome

How Raw Data Standard concept fit into Clinical Data Standards assumptions in AstraZeneca?



Following CDASH rules and supporting creation of Data Collection Standards

Alignment with Industry Standards

Developed with SDTM in mind and enabling SDTM automations

Supporting Regulatory review/reviewers tools

Define metadata for entire RAW SAS dataset and metadata for individual variable (possible comparison between different studies on RAW level)

Driving process efficiency and enabling automations, data exchange and visualizations

Standardized RAW data enables standardization on collection level and creation of analysis & reporting standards

Traceability of data across entire data flow from collection to outputs

End-users of RAW data

- RAW data **standardization** enables setting up **reusable tools and dashboards** for raw data end users
- Automatic creation of **reports**, **visualizations** and **patients profiles** for study teams
- Automation of **SDTM creation, EDC and integrations set-up**
- Stakeholders involvement at the **early stage of raw data standards creation** ensures needs of functional groups are met

Medical Team	Review physician's reports, metrics

Centralized Monitoring	Prepare centralized monitoring systems, and dashboards

Programmers	Perform gap analysis for SDTM and ADaM, program data listings, set up data tools

Data Managers	Perform gap analysis for data collection, review data listings, edit checks, visualizations, reports

Database Developers	Perform integrations with different data collectors (e.g. Safety Database, Interactive Response Technology, Sample Tracking System)

Experience in Raw Data Standards Maintenance

MDR

1. Maintaining **Metadata Repository** for standards in **machine readable format** and enabling study instance metadata creation

Standards Library

2. Maintaining **Standards Library** divided into standard types, level and Therapeutic Area

Governance

3. Established **governance process** and **governance team** as cross-functional group responsible for standards content

Core Team

4. Clinical Data Standards Team as a **central team** responsible for standards implementation and maintenance

RDS as key element

5. Raw Data Standards as **main standard component** for defining other standards, building efficient EDC Library, study instance metadata, visualizations and automations

Following Standards

6. Studies are expected to **adhere** to Clinical Data Standards and follow established process

Quality

7. Following standards, having process defined, systems in place, enables **high quality of data** aligned to **ALCOA** and **FAIR** principles

Benefits

- Standardized approach for data collection (e.g. raw dataset metadata, variables metadata)
- Standardization of data collection from different sources (EDC, TPV)
- Possibility of re-usage (streamlining and standardizing the process across studies)
- Automation of integrations set-up between systems (e.g. establishing Interactive Response Technology -EDC integration, EDC-safety database integration etc.)
- Automation of SDTM creation (less manual efforts during data transformation; aligning raw data to be SDTM ready)
- Automation of EDC creation (e.g. using macros for CRF creations)
- Accelerated timelines for study build and cost savings
- Efficiency in end-to-end automation during study set-up
- Enabling visualizations (e.g. establishing tools for raw data visualizations)
- Supporting raw data review by different teams (e.g. medical review and data monitoring)
- Automation of study instance metadata creation (Raw Data Standards maintained in MDR platform)
- Alignment to Industry Standards

Challenges

- Study level standards exceptions defined by Clinical Study Protocol
- Broad scope of portfolio (different phases and indications)
- Technical limitations and different Electronic Data Capture systems
- Compromising data collection and data analysis & reporting needs
- Limit data collection to essential data only
- Keep standards library stable
- Trainings delivery to standards end-users (how to read & use standards)
- Collaboration with different stakeholders

Q&A

Thank you.