

Implementation of Real World Structured/Unstructured Clinical Data in Real World Evidence Trial (RWE) for Regulatory Submissions

Ravi Mokkarala (Ravi.Mokkarala@pfizer.com)

Yan Shen (yan.shen@pfizer.com)

PHUSE US Connect 2022



Breakthroughs that change patients' lives





Contents

- 1. What's RWD, RWE, EHR, NLP and AI**
- 2. Workflow from Case Studies**
- 3. Real World Data Challenges and Solutions**
- 4. References**



What's RWD, RWE, EHR, NLP and AI

Real World Data* (RWD): Data describing health status and / or the delivery of health care routinely collected from variety of sources (Electronic Health Records, Claims and Billing Activities, Product and disease registries, In home use settings, Data from other devices IoT etc.,)

Real World Evidence* (RWE): Evidence of usage and potential benefits or risks of medical product derived from analyses of RWD

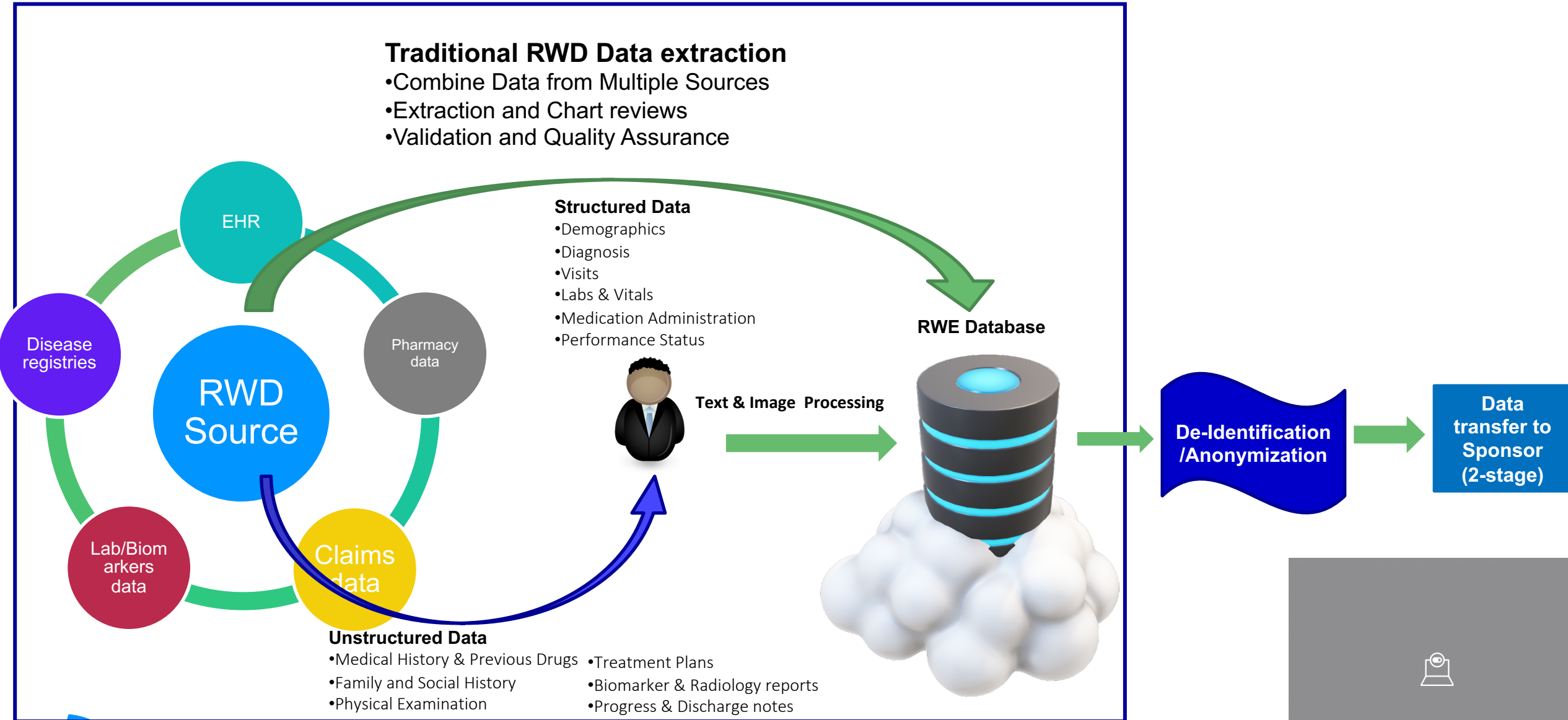
Electronic Health Record (EHR):** is an electronic version of a patient's medical history, that is maintained by the provider over time, and may include all of the key administrative clinical data relevant to that persons care under a particular provider, including demographics, progress notes, problems, medications, vital signs, past medical history, immunizations, laboratory data, and radiology reports.

Natural Language Processing (NLP): Technology used to help computers understand human language. In this case words in medical records such as progress notes, discharge notes etc.,

Artificial Intelligence (AI): Ability of computer or computer controlled robot to perform tasks associated with intelligent beings. Machine Learning (ML) is part of AI and it is the study of computer algorithm that can improve automatically through experience and by use of data

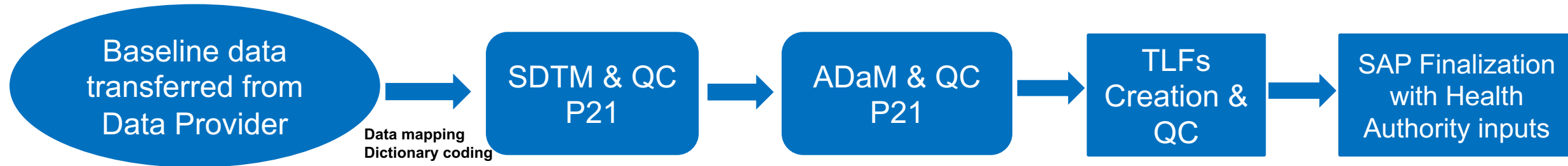


Real World Data Extraction Workflow – Data Provider

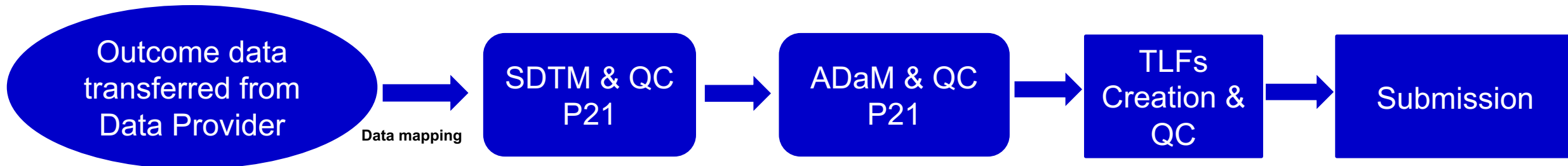


Real World Data Analysis & Reporting Workflow – Sponsor

Stage 1: Baseline Data Analysis & Reporting



Stage 2: Outcome Data Analysis & Reporting



e-Submission Package:

- XPT files
- cSDRG
- Define.xml

- XPT files
- aDRG
- Define.xml
- SAS Code

- CSR
- SAS code



2 - Stage Data Analysis & Reporting

Stage 1 – Baseline Data

- Mapping and Validation
- Covariate selection
- Model building
- Inverse Probability Treatment Weighing (IPTW)
- Propensity Score Stratification
- Propensity Score Matching (PSM)
- Finalize Statistical Analysis Plan (SAP)
- Send SAP to Health Authority for review / guidance on patient selection and analyses

Stage 2 – Outcome Data

- Update / Revision to SAP and Model selection (if required based on Health Authority inputs)
- Mapping and Validation
- Analysis of Outcome data
 - Real World Response Rate (rwRR)
 - Time to Event Analysis
 - Real World Progression Free Survival (rwPFS)
 - Overall Survival (OS)
- Clinical Study Report and eSub to Health Authority

Advantages of Stage Wise Approach

- Keep data blinded before SAP finalization to avoid bias
- Shorten the timeline: allow data provider to build and validate stage 2 data while stage 1 data analysis
- Clinical & Statistical review to improve data quality
- Building Mapping and Analyses codes to save time and improve efficiencies



Real World Data Challenges

- Combining Data from Multiple Sources
- Data Quality, Traceability and validation
- Clinical notes are free-text format, contain acronyms and abbreviations, and may involve spelling and typing errors.
- The lack of a uniform dictionary makes medical abbreviations not easy to handle
- Clinical abstractors need to read texts & images manually and provide their assessment in structured format which is subjective
- Bias and Confounding
- Missing Data
- Challenges in Mapping (CDISC) & Analyses
 - Non-availability of Expected Domains/variables (Ex: AE/LB)
 - Data that doesn't fit into any domain (Ex: Volume of Patients at Site)
 - Data not collected using Control Terminology
 - Some of Trial Design Domains cannot be built (Ex: TE, TS)
 - Conformance / Validation issues



An example of unstructured data with manual extraction - Issues

Unstructured data

- Pathology
- Progress notes
- Discharge notes
- Radiology report
- History of Present illness and Medication

Manual Text & Image Processing



Clinical
Abstractor

Structured / Derived data

- Confirmation of diagnosis and date
- Date of initial diagnosis
- Stage at diagnosis
- Biomarker details
- Prior therapies with dates
- Line of therapy & Regimen
- Response status and date

Potential Issues

- Data Quality
- Traceability
- Validation
- Reproducibility
- Time & Cost

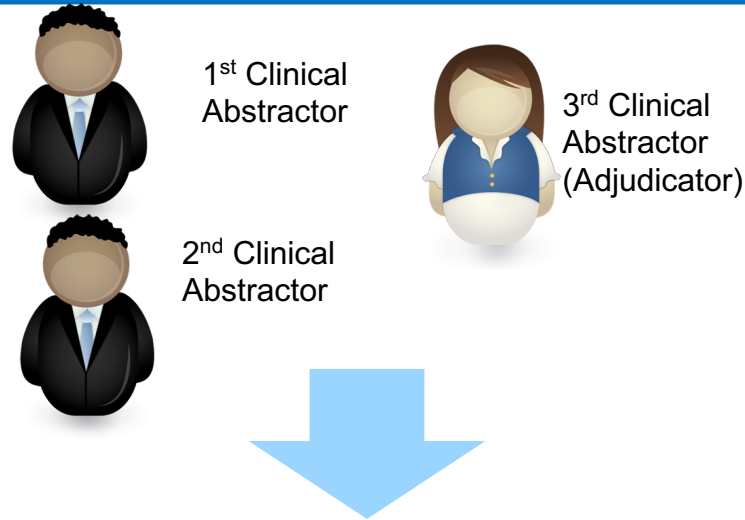


An example of unstructured data with manual extraction – One Proposed Solution

Unstructured data

- Pathology
- Progress notes
- Discharge notes
- Radiology report
- History of Present illness and Medication

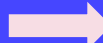
Manual Text & Image Processing with Adjudication



Structured / Derived data

- Confirmation of diagnosis and date
- Date of initial diagnosis
- Stage at diagnosis
- Biomarker details
- Prior therapies with dates
- Line of therapy & Regimen
- Response status and date

Improvement Areas

- Data Quality 
- Traceability 
- Validation 
- Reproducibility 
- Time & Cost 

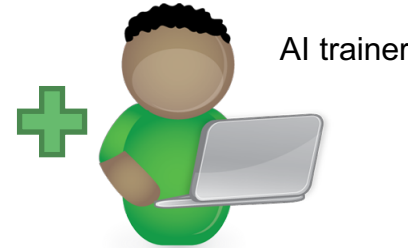
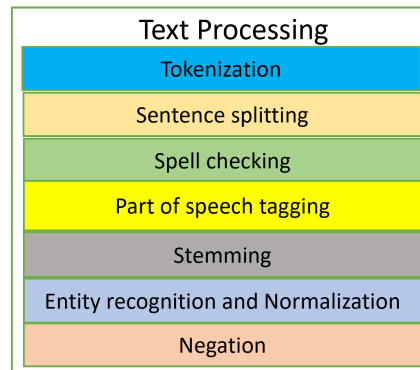


An example of unstructured data extraction with NLP/AI (Proposed Automated Solution)

Unstructured data

- Pathology
- Progress notes
- Discharge notes
- Radiology report
- History of Present illness and Medication

NLP/AI based Processing



Structured / Derived data

- Confirmation of diagnosis and date
- Date of initial diagnosis
- Stage at diagnosis
- Biomarker details
- Prior therapies with dates
- Line of therapy & Regimen
- Response status and date

Improvement Areas

- Data Quality
 - Traceability
 - Validation
 - Reproducibility
 - Time & Cost
-
- Four green arrows pointing upwards, indicating improvement in the listed areas.



References

- [Real-World Evidence | FDA](#)
- [FHIR to CDISC Joint Mapping Implementation Guide v1.0 | CDISC](#)
- [All you need to know about text preprocessing for NLP and Machine Learning](#)
- [Natural Language Processing for Smart Healthcare](#)
- [Artificial intelligence approaches using natural language processing to advance EHR-based clinical research](#)
- Considerations for using CDISC standards in Observational Studies by Jon Neville and Bess LeRoy – PhUSE US Connect 2019





Thank You for your Attention

Any Questions?

