

Project objective

This project focusses on the usability of real-world data (RWD) for regulatory submission purposes. Addressing RWD fit-for-purpose requires an assessment of data relevance including availability¹, representativeness² and timeliness². Secondly, it involves verifying the reliability of the data including data accuracy^{1,2}, precision², plausibility², completeness¹, extensiveness², coherence² and traceability¹.

The white paper under development in this project explores the challenges we face as an industry based on relevant final regulatory guidelines, the data vendor requirements and assessments to comply with these guidelines and the technical requirements for innovative solutions used to ensure the technical reliability of RWD.

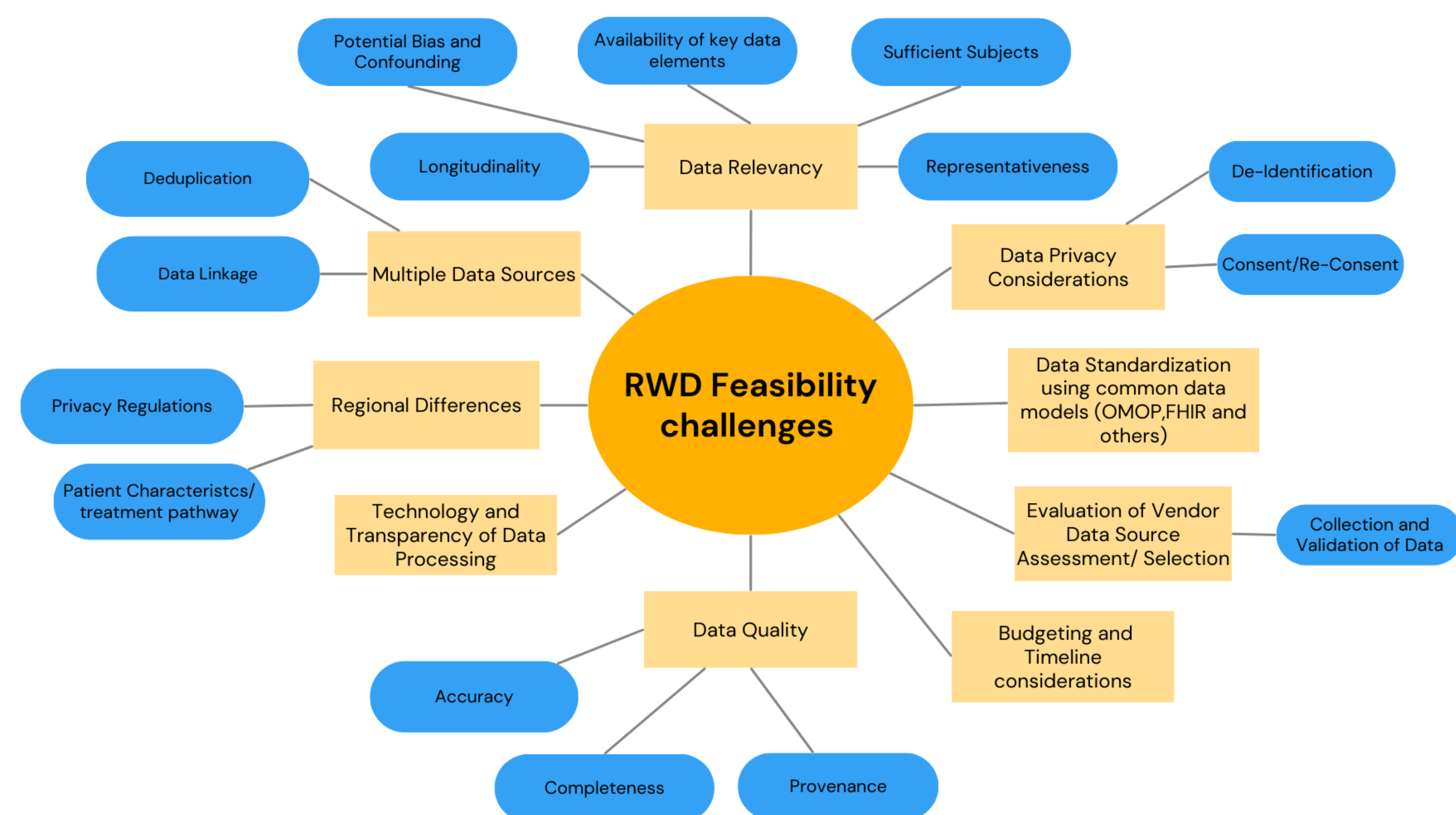
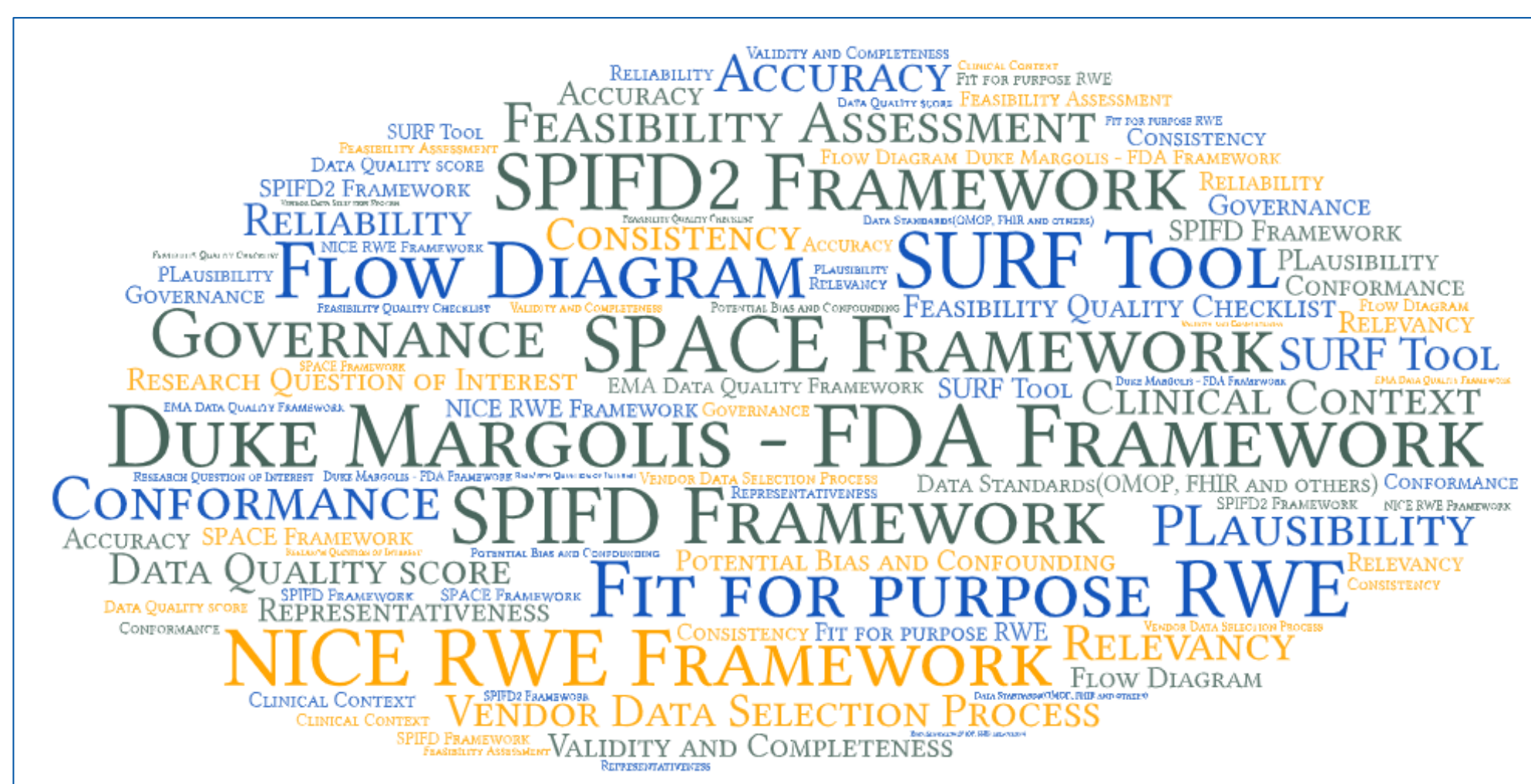
- 1) Real-World Data: Assessing Electronic Health Records and Medical Claims Data to Support Regulatory Decision-Making for Drug and Biological Products, FDA guidance for industry, July 2024
- 2) Data Quality Framework for EU medicines regulation, EMA October 2022

Assessing Feasibility of Real-World Data

One of the most challenging issues is to get data of sufficient quality from data vendors. Therefore, we are addressing the issue in the "Assessing Feasibility" section of our white paper based on review of and alignment with available fit for purpose guides like SPFD2³.

The white paper will include a vendor request for information template to get the necessary information on data availability and data quality from data vendors.

- 3) N. M. Gatto, et al., "A Structured Process to Identify Fit-for-Purpose Study Design & Data to Generate Valid and Transparent Real-World Evidence for Regulatory Uses," Clin. Pharm. Therap. 113, no. 6 (2023): 1235–1239.



Relevant Regulatory Guidelines

Assessing **Registries** to Support Regulatory Decision-Making for Drug and Biological Products (FDA, December 2023)

Considerations for the Use of Real-World Data and Real-World Evidence to Support Regulatory Decision-Making for Drug and Biological Products (FDA, August 2023)

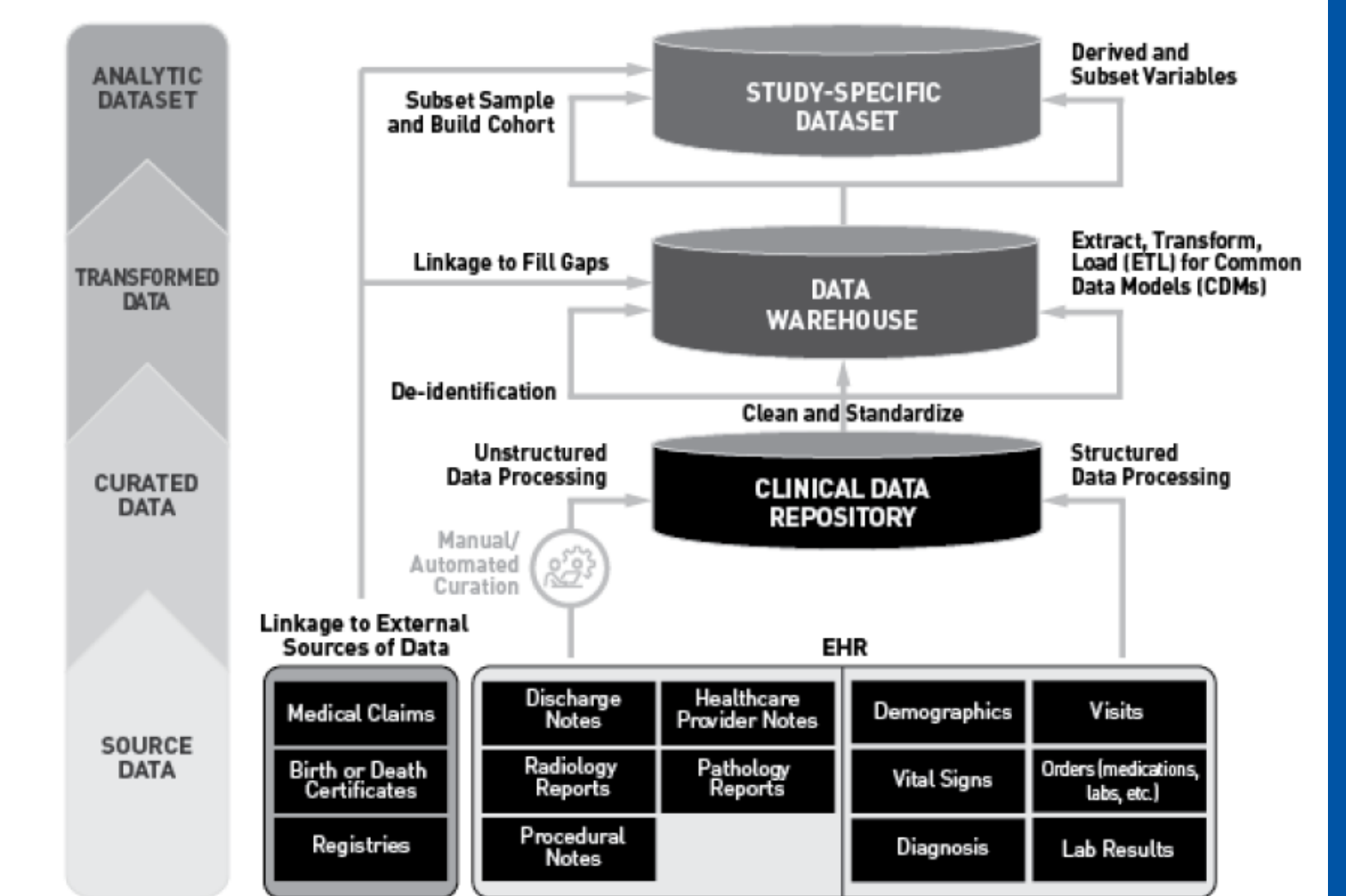
Consistent themes and challenges

- Feasibility evaluation of RWD - **Relevance** and **Reliability**
- Pre-specification of data selections and analysis
 - *Chicken and Egg*
- Understanding data collection processes
 - *Presence and absence of variables*
- Verifiable traceability throughout data life cycle
- Transparency and engagement across stakeholders
 - *Vendor, sponsor and regulator*

Data Quality Framework for EU medicines regulation (EMA, October 2023)

Assessing **Electronic Health Records and Medical Claims Data** to Support Regulatory Decision-Making for Drug and Biological Products (FDA, July 2024)

Figure 1: Illustrative Example of the Life Cycle of EHR Data **FDA, July 2024¹⁾**



Technical Reliability

Reliable data abstraction from unstructured data

AI
ML

NLP
OCR

It is essential to provide regulators explainability whenever AI/ML models are utilized for abstraction or analysis purposes, contemporaneously documenting any impact on data quality in study protocols¹.

Validate data reliability

- Complete Audit trail
- Governance control
- Backwards mapping
- Proof of Integrity

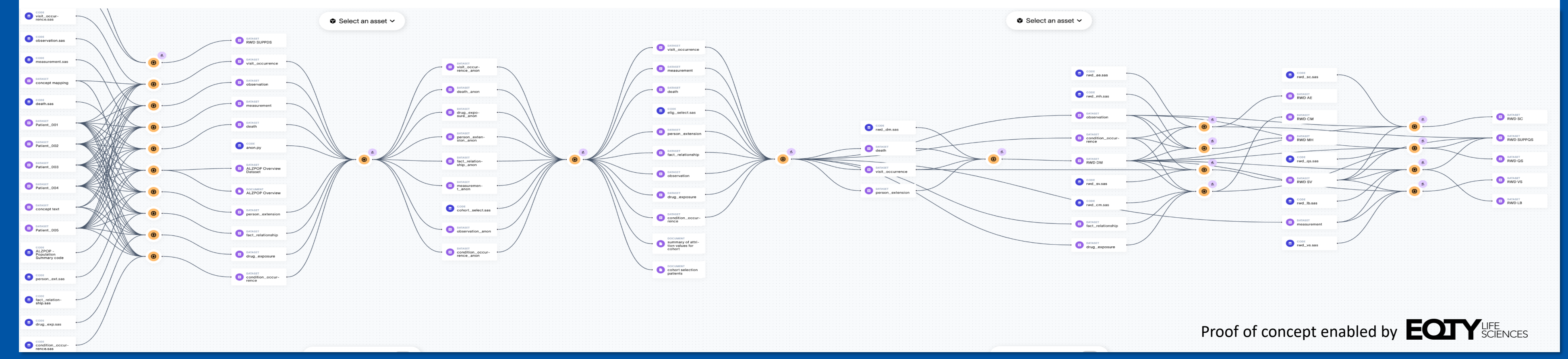
Anonymization methods

- Trusted Third Party
- Hashing
- Data abstraction
- Tokenization
- Data modification

Traceability and data lineage

Verifiable data integrity and unified governance controls across organizations and systems are critical to ensuring the traceability and transparency needed to support RWD reliability. Novel techniques play a pivotal role to create a secure, tamper-evident record of system, code, and codified human data interactions. These techniques ensure that each system serves as a deterministic source of truth, providing real-time documentation of all data transactions that proves data quality.

Cryptographical Data Reliability Over RWE Workflow: To showcase new techniques, a functional proof-of-concept is presented below where a RWD cohort was built as a control arm for a regulatory submission trial. The figure shows the data flow from FHIR JSON source to SDTM datasets.



Exemplifying Verifiable Data Reliability – Full lineage, traceability and governance control over the construction of the White Paper: To illustrate the capability of cryptographic certifications to provide data reliability over a workflow, the construction, inputs, governance controls and outputs of our project white paper will be verifiable by the PHUSE community.



Visit our webpage if you like to learn more about the project and read our blogs about feasibility and the regulatory guidances evaluated for the white paper.