

Making Pain Data FAIR (Findable, Accessible, Interoperable, Reusable) Through Data Standardisation

Erum Masood¹, Gordon Milligan¹, Emily Jefferson^{1,5}, Christopher Hall¹, Philip Appleby¹, Gillian Martin², Jillian Beggs, Antony Chuter, Tom Giles³, Armando M Villalon³, Philip Quinlan³, Christian Cole^{1,4}

¹Health Informatics Centre, University of Dundee, Scotland UK

²Tayside Clinical Trials Unit, School of Medicine, University of Dundee, Scotland UK

³Digital Research Service, University of Nottingham, UK

⁴Population Health and Genomics, School of Medicine, University of Dundee, Scotland UK

⁵Health Data Research UK

Background

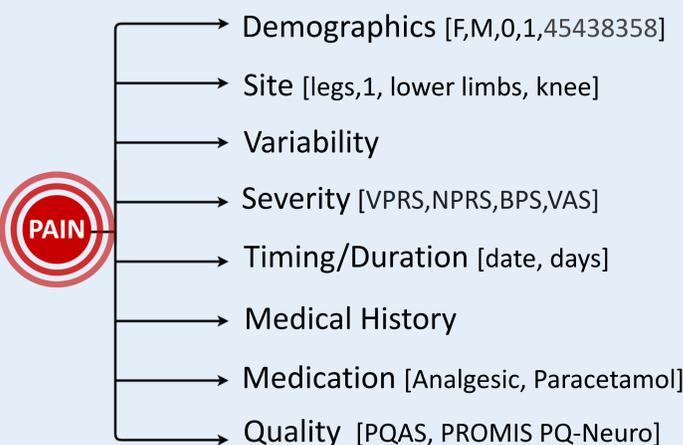
According to National Institute for Health and Care Excellence at least one third of the UK population is affected by chronic pain. It has a detrimental impact on an individual's overall health, quality of life, ability to function and work, family life and even wider society. Pain research helps acquire new knowledge on the mechanisms, pathogenesis, diagnosis, and treatment of pain. In Alleviate Data Hub, the ultimate goal of FAIR is to optimise the reuse of Pain Data.

Problem Statement

Research Pain Datasets are not readily available for reuse due to data silos, poor visibility of meta data and data to validate research cohort suitability, varying data governance procedures and

Lack of Data Standardisation, it has resulted in limited ability to compare and link data between the various research pain cohorts due to varying

- Data schema and storage,
- Formats (Free text, codes, vocabularies)
- Pain parameter units (scales, duration, aggravating and relieving factors)
- Other biometric data (medical conditions and procedures, demographics, and prescribed pain relief medications)

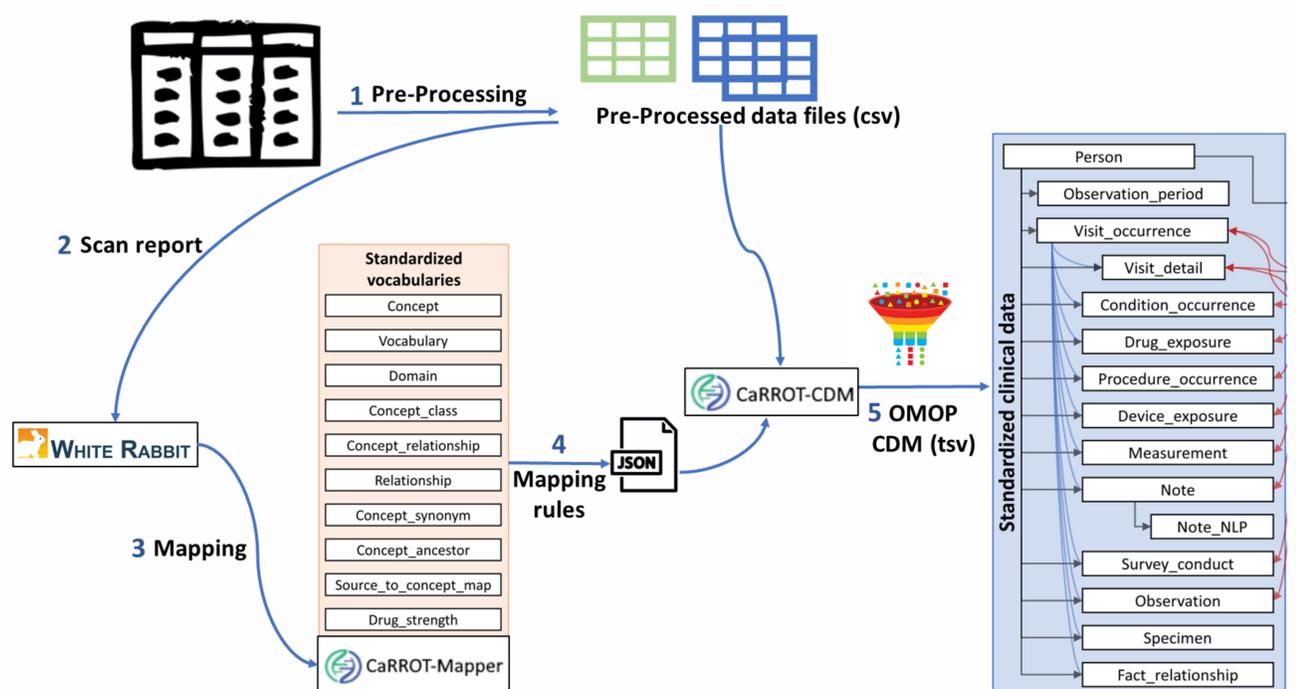


Methods

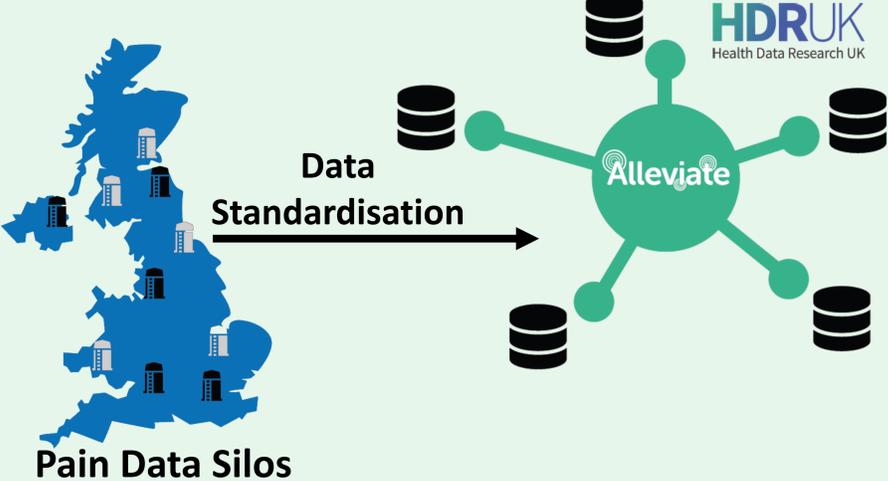
Alleviate Pain Data Hub is aimed at making data Findable, Accessible, Interoperable and Reusable. Data standardisation in Alleviate calls for federated data transformation and reformatting using **OMOP CDM v5.4** approach before it is made available on **HDR Cohort Discovery Portal** (a gateway to run queries on federated data safely without accessing the data source).

Semi-automated data standardisation methodology is adopted using the Alleviate open-source tools.

- **Pre-Processing**, data manipulation to resolve inconsistencies, standardise dates, units and numeric values, remove duplicates, interpolate missing information (where possible) and pseudonymise personal information.
- **CaRROT Mapper**, a webapp that uses the metadata (produced by WhiteRabbit tool) of dataset to produce term and structural mappings from selection of standard vocabularies used in OMOP CDM.
- **CaRROT CDM**, a python tool to perform ETL on the pain data using the mappings generated from CaRROT Mapper.



RESULTS



Data Discovery

- Federated data standardised to OMOP CDM
- Approved researchers can safely and securely perform federated data query by the HDR UK Cohort Discovery tool.
- The tool queries multiple datasets from across the UK at the same time to identify suitable data for their research.

Available Datasets

- *GOAL*: genetics of osteoarthritis and lifestyle
- *Webex*: exercise in patients with osteoarthritis
- *Omega-3*: dietary impact of omega-3 on pain
- 17 others in HDR UK Metadata Catalogue

Trusted Research Environment

- OMOP Data can be ingressed into Alleviate TRE supporting data for research.

Findable

Accessible

Interoperable

Reusable

