

Glossary of Terms

Federated Analysis: State of the Science

Collective Learning Series

Last updated: 10 January 2024

The purpose of this glossary is to establish common understanding of terminologies/concepts within the *Federated Analysis: State of the Science* Collective Learning Series.

Key terms	Working Definition
Common data model	<p>Common data model (CDM) allows for systematic analysis of disparate observational databases across jurisdictions by transforming the data contained within each jurisdiction's source databases into an agreed set of standardized data elements and data tables. The main advantages of CDMs are the short time frame needed to process queries across multiple data centres, and output quality control (1,2).</p> <p>Examples of existing CDMs:</p> <ul style="list-style-type: none"> • Canadian Network for Observational Drug Effect Studies (3) CDM adapted the FDA Sentinel CDM (4) for use in Canada. • Observational Medical Outcomes Partnership CDM (5) has been adopted to varying degrees in many countries, including the US, UK, Australia and 22 European countries.
Distributed data	Data stored across multiple organizations, institutions, or data centres (6).
Federated data	Distributed data that are able to be analyzed together while remaining separate (6,7,8).
Federated analysis	Analysis of data across multiple datasets in a fast and secure manner, where the data are not co-located (i.e., distributed data) (9).
Federated learning	Machine learning applied to data that are decentralized data. Similar to federated analysis, models are applied locally at each participating institution and only model characteristics (e.g., parameters, gradients) are transferred (6,7,8,10).
Horizontal federation	Partitions that include the same features/measures but for different people (e.g. different provinces, commonly defined data for health care use) (11).
Vertical federation	Partitions that include the same people but different features (e.g. different data



	sets, one on health care use and the other a survey) (11). <i>Note: while explicit linkage is not necessarily required, there is an assumption that the populations covered in the vertically federated data sets are the same or very largely the same.</i>
Parallel analysis	An empirical method used to decide how many components in a principal component analysis (PCA) or factors in a common factor analysis (CFA) (12).
Meta-analysis	Statistical technique for combining the findings from independent studies (13).
Pooled analysis	Analysis of individual level data that are combined from multiple locations and/or sources.
Synthetic data	Data generated from computer simulations of algorithms as an alternative to real-world data. Synthetic data can be used for training federated learning (example paper here) (14).
Trusted environments	Trusted environments are highly secure computing environments that provide remote access to health data (15,16).

References

1. [Canadian Network for Observational Drug Effect Studies \(CNODES\) CDM](#)
2. [The Book of OHDSI: Chapter 4 The Common Data Model](#)
3. [CNODES](#)
4. [FDA Sentinel CDM](#)
5. [OMOP](#)
6. [From distributed machine learning to federated learning: A survey.](#)
7. [Federated learning: Challenges, methods, and future directions.](#)
8. [Federated learning for healthcare informatics.](#)
9. [Federated networks for distributed analysis of health data.](#)
10. [The future of digital health with federated learning.](#)
11. [Federated machine learning: Concept and applications.](#)
12. [A rationale and test for the number of factors in factor analysis.](#)
13. [What is meta-analysis?](#)
14. [Synthetic data example](#)
15. [Trusted Research Environments.](#)
16. [Trusted research environments \(TRE\) green paper.](#)

