L’analyse fédérée : glossaire

Dernière mise à jour : 9 octobre 2024

Le présent glossaire vise à établir une compréhension commune des termes et des concepts relatifs à l’analyse fédérée. Les termes n’y sont pas classés par ordre alphabétique; ils sont plutôt classés de sorte qu’ils s’ajoutent les uns aux autres.

|  |  |
| --- | --- |
| **Principaux termes** | **Définition provisoire** |
| **Modèle d’information** | Modèle qui décrit les caractéristiques des éléments d’information pour certaines entités importantes du mode réel et les liens entre eux (p. ex. « hospitalisation ») ([1](#bookmark=id.e0i2vjdmod1h)). Ce modèle ne définit pas la façon dont l’information est stockée ni les technologies utilisées. |
| **Modèle de données** | Modèle qui établit les caractéristiques (p. ex. les types de données) des éléments de données qui décrivent les entités d’un modèle d’information ([2](#bookmark=id.ekj5uf272ww2)). Ce modèle définit la façon dont les éléments de données sont stockés dans un système particulier. Un modèle de données est toujours conforme à un modèle d’information, explicitement ou non. |
| **Fichier de données** | Ensemble d’éléments d’information structurés et organisés selon un modèle d’information. |
| **Source de données** | Ensemble d’éléments de données représentant un fichier de données dans lequel les éléments de données sont structurés et organisés selon un modèle de données. Il s’agit d’une mise en œuvre d’un fichier de données dans le monde réel au moyen d’une technologie déterminée (p. ex. une base de données MS SQL ou des fichiers CSV). |
| **Modèle de données commun** | Modèle de données reconnu par un vaste groupe de chercheurs dans l’intention de l’utiliser pour de nombreuses études menées dans un but commun. Certains modèles fournissent des outils qui facilitent l’extraction, la transformation et le chargement de fichiers de données provenant de différentes sources afin de créer un modèle de données convenu ([3](#bookmark=id.7xg0oasgzvjh)).  Exemples de modèles de données communs existants :   * Modèle de données commun du Canadian Network for Observational Drug Effect Studies ([4](#bookmark=id.c30rlqs600r5)), une adaptation du modèle Sentinel de la Food and Drug Administration ([5](#bookmark=id.qx3znvd42uhd)). * Modèle de données commun du Observational Medical Outcomes Partnership (OMOP) ([5](#bookmark=id.z7kl94gf167r)), adopté à différents degrés dans de nombreux pays, y compris les États-Unis, le Royaume-Uni, l’Australie et 22 pays d’Europe. * Modèle de données commun i2b2 ([7](#bookmark=id.r3l1zl91nubo)), utilisé par plus de 250 établissements à l’échelle mondiale. * Fast Healthcare Interoperability Resources (FHIR) ([8](#bookmark=id.1u4v9kj6mvf)), avec guides de mise en œuvre pour les États-Unis, l’Australie et la Nouvelle-Zélande. |
| **Partition de données** | Sous-ensemble d’une source de données ([9](#bookmark=id.kcui94p79rka)). |
| **Source de données distribuées** | Source de données composée de différentes partitions de données stockées séparément, y compris dans différents lieux physiques, organismes, établissements, centres de données ou serveurs ([10](#bookmark=id.cem8w3ashmlo)) pouvant relever de différentes autorités administratives ou juridiques. |
| **Source de données distribuées horizontalement** | Source de données distribuées dans laquelle les partitions comprennent les mêmes caractéristiques ou mesures, mais pour des sujets différents (p. ex. la même structure de données dans différentes provinces) ([11](#bookmark=id.il94aq165js5)). |
| **Source de données distribuées verticalement** | Source de données distribuées dans laquelle les partitions portent sur les mêmes sujets, mais comprennent des caractéristiques différentes (p. ex. fichiers de données différents, l’un qui porte sur l’utilisation des soins de santé, l’autre qui comprend les données d’un sondage) ([11](#bookmark=id.il94aq165js5)). |
| **Source de données fédérées** | Source de données distribuées dans laquelle les partitions suivent le même modèle de données (et modèle d’information) ([12](#bookmark=id.wivff8nwvokv)). |
| **Analyse distribuée** | Analyse statistique d’une source de données distribuées. Elle vise à protéger la confidentialité des données et ne repose pas sur l’échange de données au niveau individuel hors des sites locaux ([13](#bookmark=id.rpyea11ggpfy)). |
| **Analyse fédérée** | Type d’analyse distribuée réalisée avec des données fédérées. L’utilisation de données fédérées (qui suivent le même modèle de données) simplifie l’analyse, car certains aspects technologiques (système de base de données, types de données, etc.) de l’analyse sont les mêmes, ce qui facilite la production du code analytique requis ([14](#bookmark=id.lmofwt6zqqw2), [15](#bookmark=id.4l88h8ixaaoy)). |
| **Apprentissage fédéré** | Apprentissage automatique appliqué à des données fédérées. Les modèles sont appliqués localement à chaque nœud afin que seules leurs caractéristiques soient transférées ([16](#bookmark=id.xx8pdi8j6z9h), [17](#bookmark=id.aokrm8s7ytmi)). |
| **Analyse groupée** | Analyse statistique où toutes les partitions d’une source de données sont réunies dans le même système. |
| **Méta-analyse** | Méthode statistique qui consiste à regrouper les résultats d’un groupe d’études ([18](#bookmark=id.512s5lwtynkn)). |
| **Données synthétiques** | Données générées par des algorithmes dans le cadre de simulations par ordinateur, utilisées en remplacement de données du monde réel dans le but de reproduire certaines caractéristiques d’une distribution de données liée à un phénomène du monde réel. Les données synthétiques peuvent servir à entraîner l’apprentissage fédéré (exemple d’article [ici](https://ieeexplore.ieee.org/abstract/document/8759317?casa_token=1WLUf7WMvzYAAAAA:EwcJPrQHyBdcrdesOr9gAuPUPxj6LtyeLemxycl7jlyYlH4s8wH0k-Fsx62rx3IcbrPyysaN)) ([19](#bookmark=id.194j70d3wd6l)). |
| **Environnements de confiance** | Environnements informatiques hautement sécurisés qui permettent un accès à distance aux données ([20](#bookmark=id.41q9lit3raay), [21](#bookmark=id.v0sdlz6xk6ix)). |

# Références

1. [Cadre pancanadien de contenu des données sur la santé](https://www.cihi.ca/fr/soins-connectes/cadre-pancanadien-de-contenu-des-donnees-sur-la-sante" \o "Cadre pancanadien de contenu des données sur la santé)

1. [Data Model Considerations for Clinical Effectiveness Researchers](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3824370/" \o "Data Model Considerations for Clinical Effectiveness Researchers)

1. [The Book of OHDSI: Chapter 4 The Common Data Model](https://ohdsi.github.io/TheBookOfOhdsi/CommonDataModel.html" \o "The Book of OHDSI: Chapter 4 The Common Data Model)

1. [CNODES](http://www.cnodes.ca/" \o "CNODES)

1. [FDA Sentinel CDM](https://www.sentinelinitiative.org/methods-data-tools/sentinel-common-data-model" \o "FDA Sentinel CDM)

1. [OMOP](https://www.ohdsi.org/data-standardization/" \o "OMOP)

1. [i2b2 CDM](https://community.i2b2.org/wiki/display/BUN/i2b2+Common+Data+Model+Documentation" \o "i2b2 CDM)

1. [FHIR CDM](https://hl7.org/fhir/summary.html" \o "FHIR CDM)

1. [Determining the optimal redistribution for a given data partition](https://icl.utk.edu/~herault/papers/040%20-%20Determining%20the%20optimal%20redistribution%20for%20a%20given%20data%20partition%20-%20ISPDC%20(2014).pdf" \o "Determining the optimal redistribution for a given data partition)

1. [A Random Sample Partition Data Model for Big Data Analysis](https://arxiv.org/pdf/1712.04146" \o "A Random Sample Partition Data Model for Big Data Analysis)

1. [Federated machine learning: Concept and applications](https://arxiv.org/pdf/1902.04885.pdf)

1. [Federated Data Systems: Balancing Innovation and Trust in the Use of Sensitive Data](https://www3.weforum.org/docs/WEF_Federated_Data_Systems_2019.pdf" \o "Federated Data Systems: Balancing Innovation and Trust in the Use of Sensitive Data)

1. [Wide Area Analytics for Geographically Distributed Datacenters](https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7442496" \o "Wide Area Analytics for Geographically Distributed Datacenters)

1. [Federated Analytics: Opportunities and Challenges](https://ieeexplore.ieee.org/abstract/document/9617564" \o "Federated Analytics: Opportunities and Challenges)

1. [Blueprint for harmonising unstandardised disease registries to allow federated data analysis](https://openres.ersjournals.com/content/erjor/8/4/00168-2022.full.pdf)

1. [Federated learning for healthcare informatics](https://arxiv.org/pdf/1911.06270.pdf" \o "Federated learning for healthcare informatics)

1. [The future of digital health with federated learning](https://www.nature.com/articles/s41746-020-00323-1" \o "The future of digital health with federated learning)

1. [The potential and limitations of meta-analysis](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1060723/" \o "The potential and limitations of meta-analysis)

1. [Synthetic data example](https://ieeexplore.ieee.org/abstract/document/8759317?casa_token=1WLUf7WMvzYAAAAA:EwcJPrQHyBdcrdesOr9gAuPUPxj6LtyeLemxycl7jlyYlH4s8wH0k-Fsx62rx3IcbrPyysaN" \o "Synthetic data example)

1. [Trusted Research Environments](https://www.hdruk.org/access-to-health-data/trusted-research-environments/" \o "Trusted Research Environments)

1. [Trusted Research Environments (TRE) Green Paper](https://zenodo.org/records/4594704)