Health Data Sharing for Secondary Use

A Critical Overview

Alessio Famiani, Ruggero G. Pensa

Privacy-Aware Data Sharing Model for Health Data (PADS4Health)

PRIN 2022 PNRR – PADS4Health | Codice progetto nr. P2022MSMAW | CUP: H53D23010880001

UNIVERSITÀ DI TORINO

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

Funded by the European Union
NextGenerationEU

About the project

- **Title**: *Privacy-Aware Data Sharing Model for Health Data* (PADS4Health).
- **Objectives**:
  - raising **awareness** about the **importance of data privacy**;
  - **enhancing the workflow of scientific research**, improving its times;
  - identifying a scalable **privacy-compliant data management solution** for the **re-use of health data for research purposes.**

PRIN 2022 PNRR – PADS4Health | Codice progetto nr. P2022MSMAW | CUP: H53D23010880001

UNIVERSITÀ DI TORINO

Ministero dell'Università e della Ricerca

Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

Funded by the European Union NextGenerationEU

Advantages of digitising healthcare [1]

Healthcare Delivery

Fast and accurate care, less costs, personalised treatments, remote monitoring, decisions support.

Collaboration

Easy information sharing among stakeholders, better coordination, possibilities of automation.

Scientific Research

Access to large volumes of health data for analysis, faster innovation, improved workflow.

Managing the workflow of this data is challenging due to some barriers...

Technical barriers

Interoperability

Data Fragmentation

Data is scattered across location [1, 2].

Heterogeneity

Each system "speaks its own language" [1, 2].

Security & Privacy

Data protection

Privacy and security for in-transit and at-rest data [3].

Delegated access

Delegation mechanisms [6] *and break-the-glass* [4, 5].

Scientific Research

Data Quality

Accurate and high quality data, free of bias/errors [7].

Secure data query

Safely query data without having access to it.

Data linkage

Gather all data from of a single individual [2].

Private Analysis

Perform privacy-friendly analysis on data.

Legal barriers

Data Subject's Control

Data scattered across various locations hinders data subjects' control over their data and rights [14].

Intellectual Property

Need of protecting intellectual property of research institutions, researchers and/or companies [16].

Safeguarding rights

Adhering to existing regulations and safeguarding people's rights and freedoms.

Data Minimisation

Collecting the same details colludes with data minimisation and once-and-only principles [8, 15].

Multiple Legal Bases

Need of keeping track of different legal bases aside consent [8, 13].

Data sharing schemes

- Developers came up with **various schemes to safely exchange health data** for research within the **fragmented healthcare environment**;
- We analysed existing data sharing schemes in literature in terms of:

Architecture

The overall structure of the system.

Data Flow

How data is shared and handled across/within systems.

Security and privacy

How sensitive information is protected.

Architectural patterns [25]

- **Data enclaves** (single repository): curated repositories hosted by trusted parties that merge data from multiple sources;
- **Distributed Data Analysis**: where parties locally process their data, partial results are then shared and combined across locations;
- **Cryptographic Secure Multi-party Computation Systems**: exchange of data is encrypted and only aggregated statistics can be decrypted;

+ **Blockchain**: decentralized and immutable ledgers that ensure transparency, traceability, and secure management of data access and transactions.

Data Flow patterns

- **Data Harmonisation:** standardising data coming from multiple sources with disparate schemas;
- **Distributed Queries:** queries are computed locally and results are subsequently transmitted back to the queriers;
- **Central Aggregation:** partial results are aggregated to obtain comprehensive global insights;
- **Data to Analytics:** processed data is transferred to other parties for analysis.

Approaches to interoperability

Resources

Using mapping tools and resources to harmonise data from various sources.

e.g. SHRINE [17] *institutions map their data schemes to a unified ontology of medical concepts.*

Principles

Applying some guidelines (e.g. FAIR) during the design of the system.

e.g. PHT [19] *uses FAIR principles to make data and analysis reusable and interoperable.*

Standards

Building on widely adopted standards and software frameworks.

e.g. SHRINE [17] *and MedCo* [18] *are built on top of widely used frameworks for clinical research.*

Security & Privacy patterns

- **Autonomy & Sovereignty**: participants control their data and decisions;
- **Data Ownership & Locality**: raw data stays at its origin;
- **Distributed Trust**: collaboration without a central authority, trust spreads across multiple actors;
- **Cryptography & Access Control**: data is protected with encryption, strong authentication, and shared via secure protocols;
- **Anonymisation & Aggregation**: identifiers are removed, and only aggregated (and sometimes encrypted) statistics/results are shared.

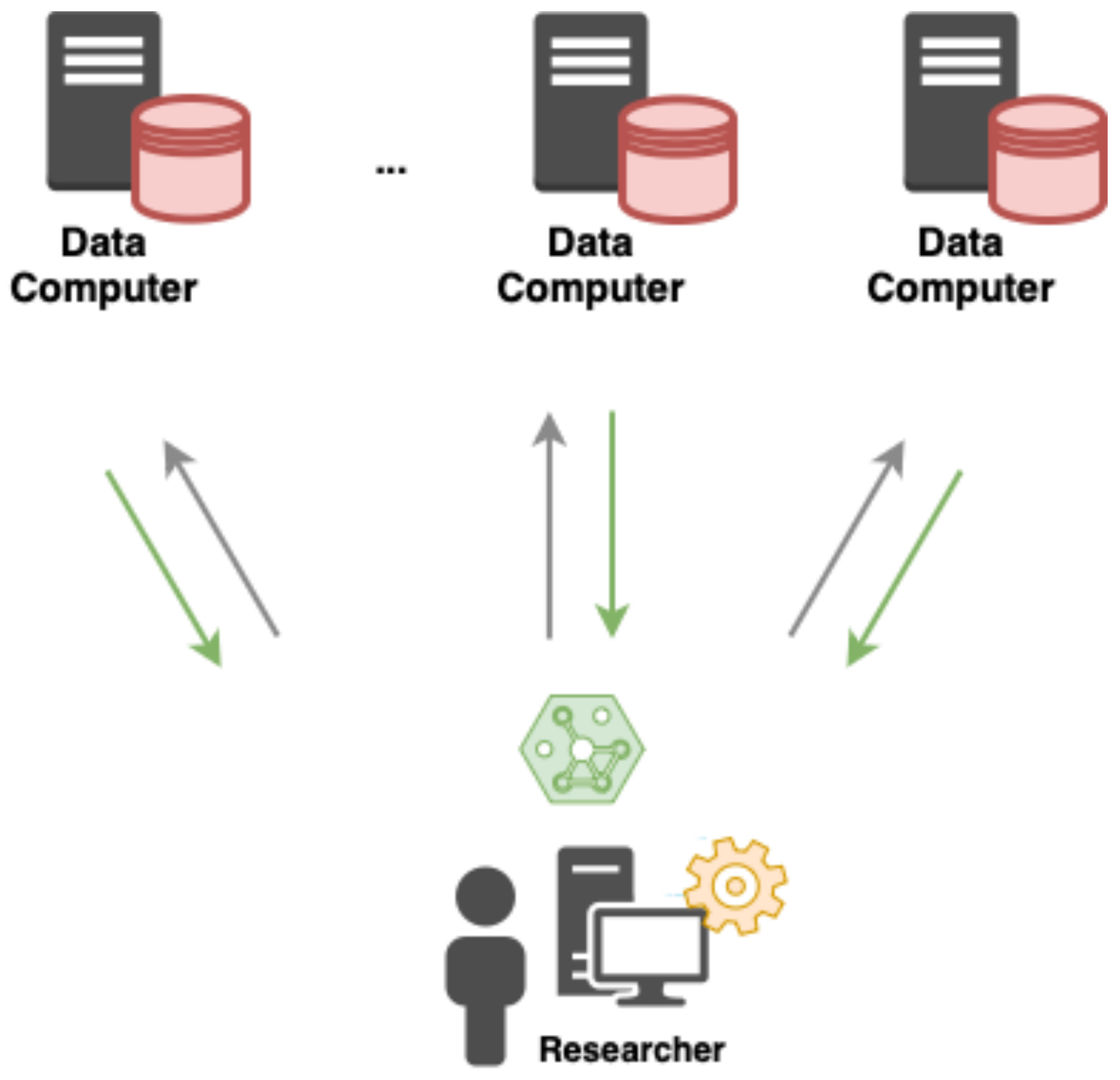
Privacy-Preserving Analysis

- **Federated Learning** (FL): multiple clients collaboratively training a model;
- **Differential Privacy** (DP): injecting random noise to individual-level data;
- **Secure Multi-Party Computation** (MPC): jointly computing a function without disclosing the inputs;
- **Homomorphic Encryption** (HE): performing operations on encrypted data without the need of decrypting it first. Results are encrypted too.

Rise in the adoption of cryptography-based technologies since **encrypted data is anonymised, thus not regulated under the GDPR**.

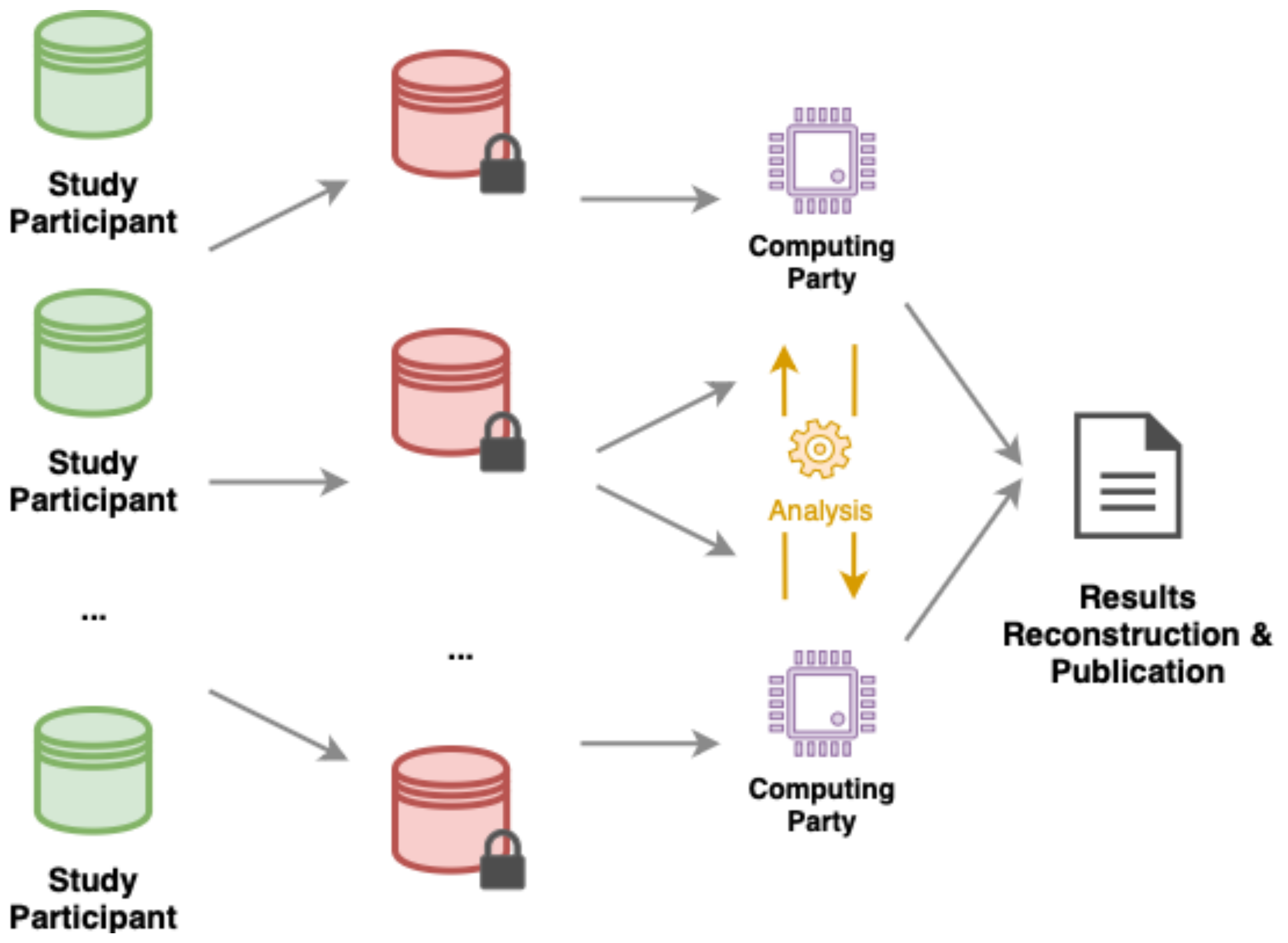
Examples: DataSHIELD [16]

- federated analysis on individual-level data stored on various **databases maintained by a participants**;
- **a central computer issues commands** to several computers where **data is stored and queried locally** in a parallel fashion;
- local **non-identifiable results are sent back** to the central computer;
- results are then **aggregated**;
- the **central computer performs the analysis**.



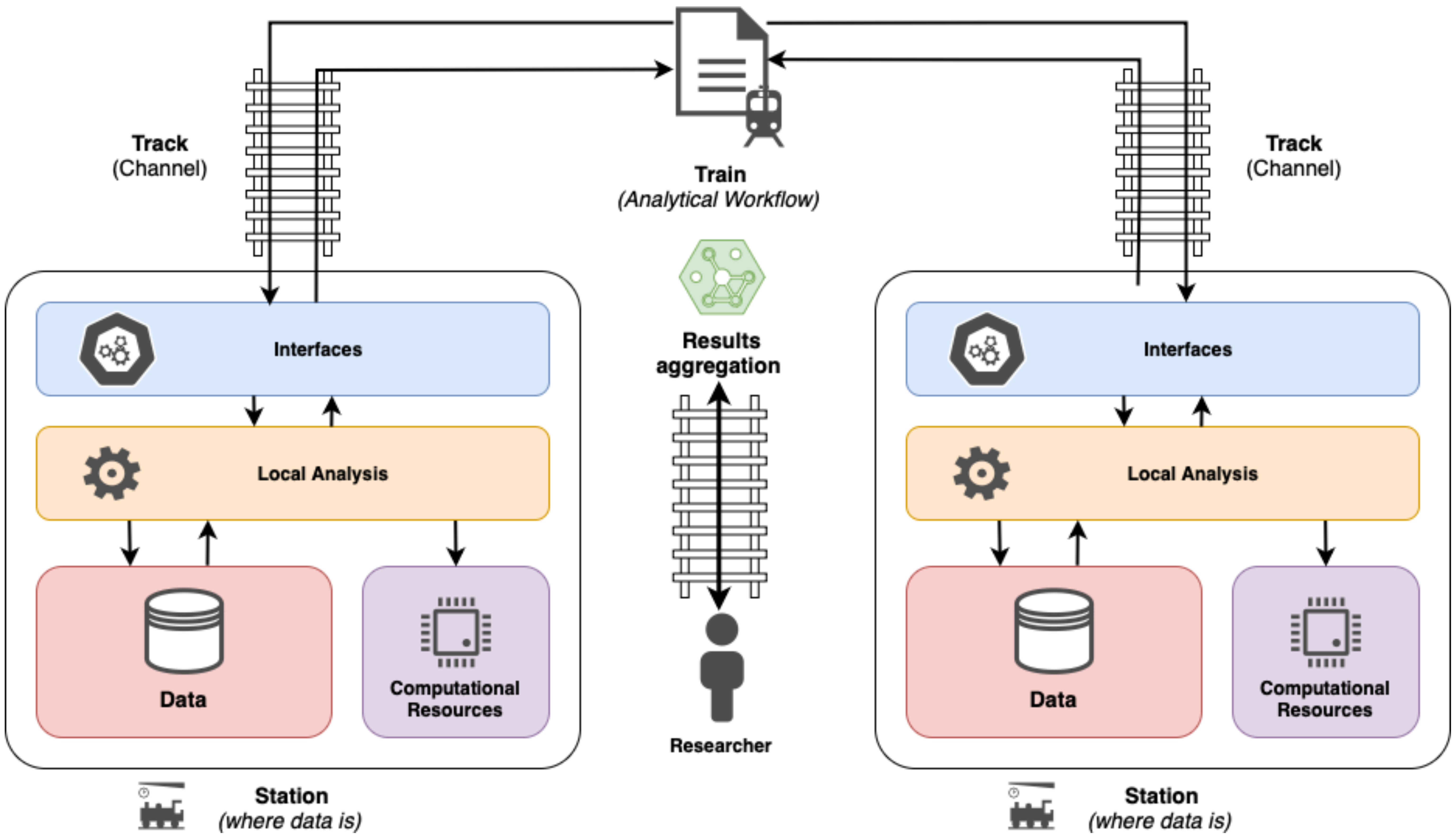
Examples: Cho et al. [23]

- **participants (SPs) own the data** & **computing parties (CPs) provide computational resources**;
- **SPs share data with CPs** using **cryptographic techniques**;
- **CPs jointly execute the required analysis** task over the secrets shares;
- **CPs combine their results together** to obtain the final results.



Examples: PHT [19]

- **stations stores data, offer resources** and interfaces for executing queries;
- **trains contain analytic workflows** (queries + algos);
- **tracks orchestrate the end-to-end workflow** (e.g. results aggregation) and enforce rules;



Requirements

- **Primary use**: *Break-the-glass, assisted health-care decision making;*
- **Secondary use**: *Safe search on data, Privacy-preserving analysis, bias/errors detection and mitigation;*
- **Data Governance**: *Intellectual property protection, Impossibility of downloading data, Respect institutional autonomy, Data portability and Interoperability, Delegation mechanisms, Records linkage, Management of legal basis;*
- **Security & Privacy**: *Security of at-rest and in-transit data, Non-repudiation, Integrity,* ***General Robustness from attacks, Anonymisation & Pseudonymisation, Data minimisation and once-only****, AC, Authorisation, etc...*

UNIVERSITÀ DI TORINO

ITA DATA 2025

Thank you for your time.

References

[1] A. Torab-Miandoab, T. Samad-Soltani, A. Jodati, P. Rezaei-Hachesu, Interoperability of heteroge-neous health information systems: a systematic literature review, BMC medical informatics and decision making 23 (2023) 18.

[2] M. Lehne, J. Sass, A. Essenwanger, J. Schepers, S. Thun, Why digital medicine depends on interoperability, NPJ digital medicine 2 (2019) 79.

[3] I. Keshta, A. Odeh, Security and privacy of electronic health records: Concerns and challenges, Egyptian Informatics Journal 22 (2021) 177–183.

[4] D. Povey, Optimistic security: a new access control paradigm, in: Proceedings of the 1999 workshop on New security paradigms, 1999, pp. 40–45.

[5] A. D. Brucker, H. Petritsch, Extending access control models with break-glass, in: Proceedings of the 14th ACM symposium on Access control models and technologies, 2009, pp. 197–206.

[6] M. Joshi, K. P. Joshi, T. Finin, Delegated authorization framework for ehr services using attribute-based encryption, IEEE Transactions on Services Computing 14 (2019) 1612–1623.

References

[7] D. Cirillo, S. Catuara-Solarz, C. Morey, E. Guney, L. Subirats, S. Mellino, A. Gigante, A. Valencia, M. J. Rementeria, A. S. Chadha, et al., Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare, NPJ digital medicine 3 (2020) 81.

[8] European Parliament, Council of the European Union, Regulation (EU) 2016/679 of the European Parliament and of the Council, 2016. URL: https://data.europa.eu/eli/reg/2016/679/oj.

[9] The 104th United States Congress, Health insurance portability and accountability act, 1996. Public Law 104-191.

[10] European Parliament, Council of the European Union, Directive (eu) 2019/1024 of the European parliament and of the council of 20 june 2019 on open data and the re-use of public sector information (recast), 2019. URL: http://data.europa.eu/eli/dir/2019/1024/oj.

[11] European Parliament, Council of the European Union, Regulation (eu) 2022/868 of the european parliament and of the council of 30 may 2022 on european data governance and amending regulation (eu) 2018/1724 (data governance act) (text with eea relevance), 2022. URL: http://data.europa.eu/eli/reg/2022/868/oj.

References

[12] European Parliament, Council of the European Union, Regulation (eu) 2025/327 of the european parliament and of the council of 11 february 2025 on the european health data space and amending directive 2011/24/eu and regulation (eu) 2024/2847 (text with eea relevance), 2025. URL: http://data.europa.eu/eli/reg/2025/327/oj.

[13] P. of the Italian Republic, Italian personal data protection code - legislative decree no. 196 of 30 june 2003, 2003. URL: https://www.gazzettaufficiale.it/atto/serie_generale/caricaDettaglioAtto/originario atto.dataPubblicazioneGazzetta=2003-07-29&atto.codiceRedazionale=003G0218.

[14] A. Ghorbel, M. Ghorbel, M. Jmaiel, Privacy in cloud computing environments: a survey and research challenges, J. Supercomput. 73 (2017) 2763–2800.

[15] European Parliament, Council of the European Union, Communication from the commission to the european parliament, the council, the european economic and social committee and the committee of the regions eu egovernment action plan 2016-2020 accelerating the digital transformation of government, 2016. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52016DC0179.

References

[16] A. Gaye, Y. Marcon, J. Isaeva, P. LaFlamme, A. Turner, E. M. Jones, J. Minion, A. W. Boyd, C. J. Newby, M.-L. Nuotio, R. Wilson, O. Butters, B. Murtagh, I. Demir, D. Doiron, L. Giepmans, S. E. Wallace, I. Budin-Ljøsne, C. Oliver Schmidt, P. Boffetta, M. Boniol, M. Bota, K. W. Carter, N. deKlerk, C. Dibben, R. W. Francis, T. Hiekkalinna, K. Hveem, K. Kvaløy, S. Millar, I. J. Perry, A. Peters, C. M. Phillips, F. Popham, G. Raab, E. Reischl, N. Sheehan, M. Waldenberger, M. Perola, E. van den Heuvel, J. Macleod, B. M. Knoppers, R. P. Stolk, I. Fortier, J. R. Harris, B. H. Woffenbuttel, M. J. Murtagh, V. Ferretti, P. R. Burton, Datashield: taking the analysis to the data, not the data to the analysis, International Journal of Epidemiology 43 (2014) 1929–1944. URL: https://doi.org/10.1093/ije/dyu188.doi:10.1093/ije/dyu188.

[17] A. J. McMurry, S. N. Murphy, D. MacFadden, G. Weber, W. W. Simons, J. Orechia, J. Bickel, N. Wattanasin, C. Gilbert, P. Trevvett, S. Churchill, I. S. Kohane, SHRINE: Enabling Nationally Scalable Multi-Site Disease Studies, PLOS ONE 8 (2013) 1–11. URL: https://doi.org/10.1371/journal.pone.0055811. doi:10.1371/journal.pone.0055811.

[18] J. L. Raisaro, J. R. Troncoso-Pastoriza, M. Misbach, J. S. Sousa, S. Pradervand, E. Missiaglia, O. Michielin, B. Ford, J. Hubaux, MedCo: Enabling Secure and Privacy-Preserving Exploration of Distributed Clinical and Genomic Data, IEEE ACM Trans. Comput. Biol. Bioinform. 16 (2019) 1328–1341. URL: https://doi.org/10.1109/TCBB.2018.2854776. doi:10.1109/TCBB.2018.2854776.

References

[19] O. Beyan, A. Choudhury, J. van Soest, O. Kohlbacher, L. Zimmermann, H. Stenzhorn, M. R. Karim, M. Dumontier, S. Decker, L. O. B. da Silva Santos, A. Dekker, Distributed analytics on sensitive medical data: The personal health train, Data Intell. 2 (2020) 96–107. URL: https://doi.org/10.1162/dint_a_00032. doi:10.1162/DINT_A_00032.

[20] A. Sadilek, L. Liu, D. Nguyen, M. Kamruzzaman, S. Serghiou, B. Rader, A. Ingerman, S. Mellem, P. Kairouz, E. O. Nsoesie, J. Macfarlane, A. Vullikanti, M. V. Marathe, P. Eastham, J. S. Brownstein, B. A. y Arcas, M. D. Howell, J. Hernandez, Privacy-first health research with federated learning, NPJ Digit. Medicine 4 (2021). URL: https://doi.org/10.1038/s41746-021-00489-2. doi:10.1038/S41746-021-00489-2.

[21] K. Yu, L. Tan, X. Shang, J. Huang, G. Srivastava, P. Chatterjee, Efficient and privacy-preserving medical research support platform against COVID-19: A blockchain-based approach, IEEE Consumer Electron. Mag. 10 (2021) 111–120. URL: https://doi.org/10.1109/MCE.2020.3035520.doi:10.1109/MCE.2020.3035520.

[22] S. Mendelsohn, D. Froelicher, D. Loginov, D. Bernick, B. Berger, H. Cho, sfkit: a web-based toolkit for secure and federated genomic analysis, Nucleic Acids Res. 51 (2023) 535–541. URL:https://doi.org/10.1093/nar/gkad464. doi:10.1093/NAR/GKAD464.

References

[24] D. Froelicher, J. R. Troncoso-Pastoriza, J. L. Raisaro, M. A. Cuendet, J. S. Sousa, H. Cho, B. Berger,J. Fellay, J.-P. Hubaux, Truly privacy-preserving federated analytics for precision medicine with multiparty homomorphic encryption, Nature communications 12 (2021) 5910. URL: https://doi.org/10.1038/s41467-021-25972-y. doi:10.1038/s41467-021-25972-y.

[25] F. N. Wirth, T. Meurers, M. Johns, F. Prasser, Privacy-preserving data sharing infrastructures for medical research: systematization and comparison, BMC Medical Informatics Decis. Mak. 21 (2021) 242. URL: https://doi.org/10.1186/s12911-021-01602-x. doi:10.1186/S12911-021-01602-X.

[26] P. H. Scotland, National Safe Haven (NSH), https://publichealthscotland.scot/resources-and-tools/health-intelligence-and-data-management/electronic-data-research-and-innovation-service-edris/national-safe-haven-nsh/, 2025. [Accessed 28-07-2025].

[27] E. Business, I. Agency, Estonian e-Health Records — e-estonia.com, https://e-estonia.com/solutions/e-health-2/e-health-records/, 2025. [Accessed 24-06-2025].

[28] K. Mäeots, KSI Blockchain Stack: Zero Trust Applications - DigiExpo — digiexpo.e-estonia.com, https://digiexpo.e-estonia.com/cyber-security/ksi-blockchain-stack-zero-trust-applications/, 2025. [Accessed 24-06-2025].

References

[29] E. Business, I. Agency, X-Road - e-Estonia — e-estonia.com, https://e-estonia.com/solutions/interoperability-services-x-road/x-road/, 2025. [Accessed 24-06-2025].

[30] G. Hripcsak, J. D. Duke, N. H. Shah, C. G. Reich, V. Huser, M. J. Schuemie, M. A. Suchard, R. W. Park, I. C. K. Wong, P. R. Rijnbeek, et al., Observational health data sciences and informatics (ohdsi): opportunities for observational researchers, in: MEDINFO 2015: eHealth-enabled Health, IOS Press, 2015, pp. 574–578.

[31] Harvard Catalyst, Development - SHRINE - open.catalyst wiki — open.catalyst.harvard.edu, https://open.catalyst.harvard.edu/wiki/display/SHRINE/Development, 2025. [Accessed 24-06-2025].

[32] U.S. Centers for Disease Control and Prevention, ICD - ICD-9-CM - International Classification of Diseases, Ninth Revision, Clinical Modification — archive.cdc.gov, https://archive.cdc.gov/www_cdc_gov/nchs/icd/icd9cm.htm, 2025. [Accessed 24-06-2025].

[33] NCBO, Ncbo bioportal, https://bioportal.bioontology.org, 2025. [Accessed 24-06-2025].

[34] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, et al., The fair guiding principles for scientific data management and stewardship, Scientific data 3 (2016) 1–9. URL: https://doi.org/10.1038/sdata.

References

[35] Y. Zhuang, L. R. Sheets, Y. Chen, Z. Shae, J. J. P. Tsai, C. Shyu, A patient-centric health information exchange framework using blockchain technology, IEEE J. Biomed. Health Informatics 24 (2020) 2169–2176. URL: https://doi.org/10.1109/JBHI.2020.2993072. doi:10.1109/JBHI.2020.2993072.

[36] K. Kõnd, A. Lilleväli, E-prescription success in estonia: The journey from paper to pharmacoge-nomics., Eurohealth 5 (2019) 18–20.

[37] Research Data Assistance Center, CCW Virtual Research Data Center (VRDC) | ResDAC — res-dac.org, https://resdac.org/cms-virtual-research-data-center-vrdc, 2025. [Accessed 24-06-2025].

[38] Research Data Assistance Center, CCW Virtual Research Data Center (VRDC) FAQs | ResDAC —resdac.org, https://resdac.org/virtual-research-data-center-vrdc-faqs, 2025. [Accessed 24-06-2025].

[39] S. Padmanabhan, L. Carty, E. Cameron, R. E. Ghosh, R. Williams, H. Strongman, Approach to record linkage of primary care data from clinical practice research datalink to other health-related patient data: overview and implications, European journal of epidemiology 34 (2019) 91–99.