IJDC | *Conference Paper*

Trusted Research Environments: Analysis of Characteristics and Data Availability

Martin Weise
TU Wien

Andreas Rauber
TU Wien

Abstract

Trusted Research Environments (TREs) enable the analysis of sensitive data under strict security assertions that protect the data with technical, organizational, and legal measures from (accidentally) being leaked outside the facility. While many TREs exist in Europe, little information is available publicly on the architecture and descriptions of their building blocks and their slight technical variations. To highlight on these problems, an overview of the existing, publicly described TREs and a bibliography linking to the system description are provided. Their technical characteristics, especially in commonalities and variations, are analysed, and insight is provided into their data type characteristics and availability. The literature study shows that 47 TREs worldwide provide access to sensitive data, of which two-thirds provide data predominantly via secure remote access. Statistical offices (SOs) make the majority of sensitive data records included in this study available.

Submitted 9 February 2024 ~ *Accepted* 22 February 2024

Correspondence should be addressed to Martin Weise. Email: martin.weise@tuwien.ac.at

This paper was presented at the International Digital Curation Conference IDCC24, 19-21 February 2024

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: http://www.ijdc.net/

Introduction

Evidence-based research demands access to sensitive data to apply analysis on high-quality data from trusted sources, improving the state of the art in the major fields [31] of: 1) Life Sciences, 2) Physical Sciences, Mathematics, and Engineering, 3) Arts and Humanities, and 4) Social and Behavioral Sciences. Policymakers globally have recognized the value of research with sensitive data in the previous years and allowed the use of this data for scientific purposes in many cases when ensuring public trust in the appropriateness of use, confidential treatment, fair terms for access, and transparency of data usage. Historically, modes of accessing sensitive started with physical visits to Trusted Research Environments (TREs), conducting research on de-identified data in a safe room that is monitored and that heavily restricts allowed tools and materials to be used and the research produced. Alternatively, some TREs may allow working with sensitive data at trusted partner TREs, a modus operandi of Federal Statistical Research Data Centres (FSRDCs) in the United States of America (USA) that allows researchers to work with census data in any of the 33FSRDCs. [1] Protecting and maintaining control over digital sensitive data, confidential data, or data related to intellectual property when striving to give third parties access to the data poses a significant challenge. The TREs have been established in the last decade that, when properly set up and operated, help ease this problem by providing high-security guarantees of a monitored and highly controlled environment.

The umbrella term *sensitive* interchangeably is used with confidential data to signify that the reason why the data becomes sensitive (e.g., containing personal data, commercial value) can be disregarded. Any TRE should maintain control over sensitive data in any case regardless. Similarly, no common nomenclature for the concept of a TRE exists: the term Secure Research Environment (SRE) is predominantly used in the USA, Secure Data Environment (SDE) in Great Britain (GB), Sensitive Data Service (SDS) in Scandinavian Countries, and Secure Processing Environment (SPE) in a pan-European context. The goal of this literature study is to discover existing TREs and analyse their characteristics and data availability to give an overview of available infrastructure for sensitive data research because many European initiatives have been emerging recently. The main contributions of this work toward "transparency and trust in research practices" are:

- A comprehensive list of available TREs and a bibliography.
- Analysis of the TRE operation characteristics.
- Analysis of sensitive data availability and access.

Methodology

In this literature study, existing TREs and available datasets are identified globally using scholarly databases (e.g., Scopus, Web of Science, IEEE Xplore, and Science Direct), a computer science bibliography,[2] Google, and grey literature, which focus on the retrieval of the following materials.

- Peer-reviewed articles where available
- TRE websites
- TRE metadata catalogs.

[1] https://www.census.gov/about/adrm/fsrdc/locations.html
[2] https://dblp.org/

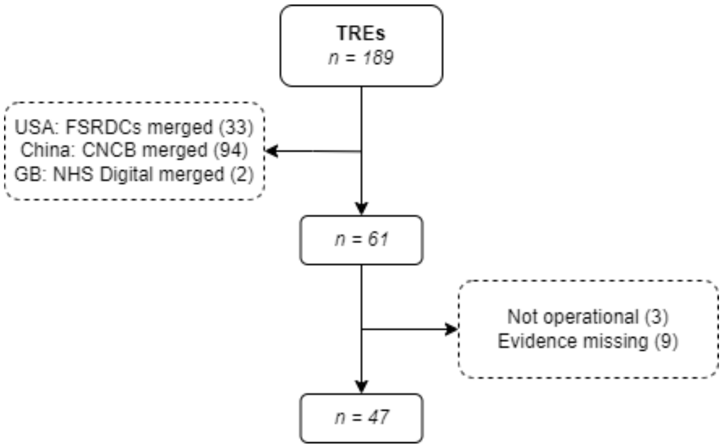

Figure 1. Flow chart for literature selection.

TREs are included (see Figure 1) in the inventory based on the following criteria: the infrastructure must implement safeguards similar to the five safes framework introduced by (Ostriker, Kuh, & Voytuk, 2003), describing a safe centre for sensitive research data in five risk dimensions. This already excludes research data repositories that cannot provide a safe environment to analyse sensitive data. To be included in the inventory, each infrastructure must additionally fulfill the requirements of a TRE defined (Hubbard, Reilly, Varma, & Seymour, 2020), who extend the *safe setting* formulated by (Desai et al., 2016). with safe computing and the possibility to safely map the research results back to, for example, individual clinical care in a *safe return* requirement.

For each infrastructure in the inventory, related infrastructures were merged with a common governance board (e.g., 33 FSRDCs governed by the USA Census Bureau) and copies of the same infrastructure. Currently, TREs were excluded that are not in operation but provide reference models (Weise, Kovacevic, Popper, & Rauber, 2022) and (Ekaputra et al., 2022) or do not have sufficient evidence that describes the infrastructure operation. This especially holds for CSC ePouta and CSC SD Services in Finland, who did not publish a peer-reviewed article but provided sufficient evidence that a TRE is in operation. Eliciting this information via other means, such as interviews or on-site visits, was out of the scope of this initial study to showcase publicly available information. These are being kept in reserve for future investigation.

The number of datasets available in the metadata catalogs was determined as follows. If they were not provided by the TRE website or reports: 1) the public metadata catalog application programming interface (API) was used, 2) de-structured collections (e.g., the HUNT studies of the HUNT Cloud (Krokstad, Sund, Kvaløy, Rangul, & Næss, 2022)), which are available as six collections that partly have iterations (i.e., repetitions) but contain 56 separate studies; therefore, 56 datasets were counted, and 3) scraped public endpoints, for example, the Croatian Bureau of Statistics has an undocumented endpoint available in Croatian and English that accepts thematic abbreviations, producing HTML that can be scraped and filtered.[3] For merged TREs, the number of datasets available was not collected; however, they were not excluded from the literature study to acknowledge their importance.

[3] https://intra.dzs.hr/cat/v2/list/

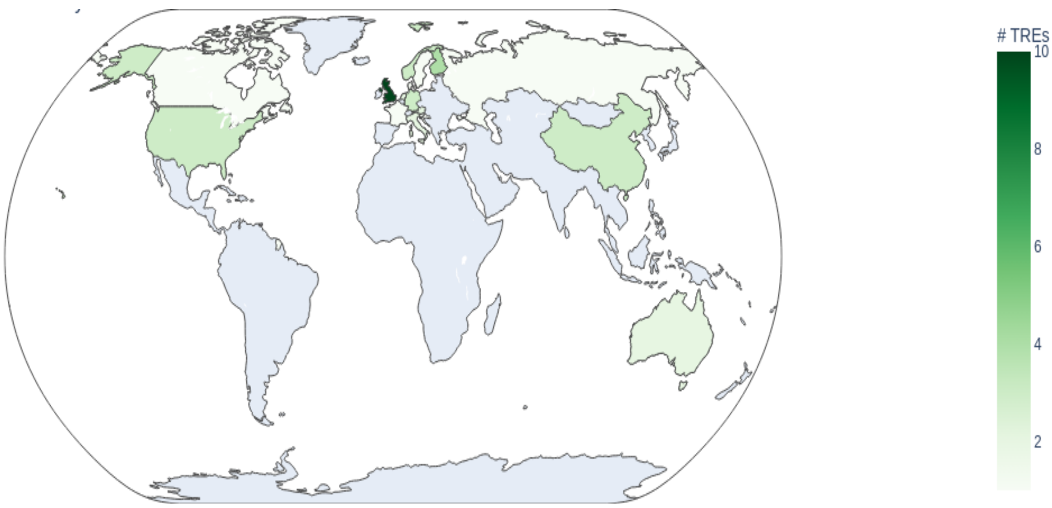

Figure 2. TREs globally.

Results

TRE Operation Characteristics

The TREs are predominantly in Europe (*n* = 39, 83%), followed by Asia (*n* = 3, 6%), North America (*n* = 3, 6%) and Oceania (*n* = 2, 4%). The countries with the most TREs are GB (*n* = 10, 21%) and Finland (n = 4, 9%), followed by China, the USA, Germany, and Norway (n = 3, 6%). The TREs operating within the European Union (n = 24, 51%) benefit from the common legislative norms in place to protect individual privacy under the General Data Protection Regulations. The 47 TREs (Table 1) were compared for architectural similarities and differences. A map of the TREs aggregated by country is shown in Figure 2, excluding pan-European TREs (*n* = 5, 11%). A very balanced mix of TREs was found, which provided raw data (*n* = 23, 49%) and provided a secure platform (*n* = 24, 51%) to link data securely. In general, before any research on sensitive data, a request must be sent to the data owner containing; 1) personal data to identify the analyst, 2) the required data, 3) the required tools to perform the analysis, and 4) the task and research questions that should be answered with the required data (Ritchie & Welpton, 2016).

In total, 31 TREs were identified that have a public metadata catalog. An overwhelming majority (*n* = 27, 87%) provide access to structured data, such as tabular data. Some TREs provide access to unstructured data (*n* = 4, 13%), such as images. The TRE with the most datasets (*n* = 1.2million, 90%) available is Eurostat (Reuter & Museux, 2010), the head statistical office in the European Union (see Figure 3). A metadata catalog [4] exists for human interaction as well as multiple machine-actionable interfaces, such as SDMX,[5] bulk downloads, and API endpoints.

The SeRP (Ford et al., 2009) provides infrastructure solutions to their *tenants*.[6] Each has its technical and governance requirements depending on the data it holds, how it is generated and what it has to do with, who funds it, and the regional or national data landscape in which it operates. Access to data is enabled directly by the tenant's unique access conditions.

[4] https://ec.europa.eu/eurostat/data/database/
[5] https://sdmx.org/
[6] https://serp.ac.uk/tenants/

Table 1. TREs included in this literature study. For space constraints we provide countries in ISO 3166-1 alpha-2 coding.

Name & Publication	Country	Available
Secure Unified Research Environment (O'Keefe, Gould, & Churches, 2014)	AU	2017
Secure eResearch Platform Australia (Jones, Ford, Ellwood-Thompson, & Lyons, 2016)		2011
DEXHELPP (Popper, Endel, Mayer, Bicher, & Glock, 2017)	AT	2017
Austrian Microdata Centre (Fuchs, G¨ollner, Hartmann, & Thomas, 2023)		2022
Population Data British Columbia (Pencarrick Hertzman, Meagher, & McGrail, 2013)	CA	2008
Croatian Bureau of Statistics (Poljicak & Stancic, 2014)	HR	2020
Scientific Data Centre of Chinese Academy of Science (Z. Zhang et al., 2019)	CN	2019
National Genomics Data Centre (Xue et al., 2022)		2022
National Bureau of Statistics China (C. Zhang & Hou, 2020)		2020
de.NBI Cloud (Hoffmann et al., 2023)	DE	2017
UseGalaxy.eu (Jalili et al., 2020)		2016
RemoteNEPS (Barkow et al., 2011)		2011
Statistics Denmark Remote Desktop (Thygesen, 1995)	DK	2008
European Health Data Evidence Network (Bastião Silva, Trifan, & Luís Oliveira, 2018)	EU	2022
EJP Rare Diseases Virtual Platform (Kaliyaperumal et al., 2021)		2023
European Genome-phenome Archive (Lappalainen et al., 2015)		2008
GAIA-X DataLoft (Boll & Meyer, 2022)		2023
Haematology Outcomes Network in Europe (Bardenheuer, Van Speybroeck, Hague, Nikai, & Price, 2022)		2018
CSC ePouta	FI	2020
CSC SD Services		2022
FIONA (Rat für Sozial- und Wirtschaftsdaten, 2019)		2010
SPESiOR (Soini, Hallinen, & Martikainen, 2022)		2022
Health Data Hub (Cuggia & Combes, 2019)	FR	2019
Pedianet Database (Cantarutti & Gaquinto, 2021)	IT	1998
BIRD (Bruno, D'Aurizio, & Tartaglia-Polcini, 2014)		2007
Eurostat (Reuter & Museux, 2010)	LU	2013
National Statistics Office Malta (Zilhão, Baldacci, Bergdahl, & Djerf, 2022)	MT	2021
ODISSEI Secure Supercomputer (de Zeeuw et al., 2021)	NL	2018
CBS Microdata (Schouten & Cigrang, 2003)		2003
HUNT Cloud (Krokstad, Sund, Kvaløy, Rangul, & Næss, 2022)	NO	2020
TSD (Øvrelid, Bygstad, & Thomassen, 2021)		2014
Statistics Norway (Langsrud, 2019)		2019
Federal State Statistics Service (Akatkin & Yasinovskaya, 2019)	RU	2002
Statistical Office of the Republic of Slovenia (Križman, Klanjšček, & Čerk, 2002)	SI	2006
Microdata Online Access (Hjelm, 2005)	SE	2005
Electronic Data Research and Innovation Service (Pavis & Morris, 2015)	UK	2013
SHAIP (Wilde, Anderson, Boyle, Pinder, & Weir, 2022)		2022
EBI Embassy Cloud (Cook et al., 2015)		2013
Data Access Environment (Wood et al., 2021)		2019
Secure Research Service (Ritchie, 2008)		2004
Secure Anonymised Information Linkage Databank (Ford et al., 2009)		2009
National Safe Haven (Gao et al., 2022)		2020
Secure eResearch Platform (Jones, Ford, Ellwood-Thompson, & Lyons, 2016)		2011
Clinical Practice Research Datalink (Herrett et al., 2015)		1993
QResearch (Hippisley-Cox, Stables, & Pringle, 2004)		2004
Federal Statistical Research Data Centres (Habermann, 2010)	US	1982
Secure Research Environment (Rose et al., 2022)		2019
Secure Research Infrastructure (Deumens et al., 2021)		2021

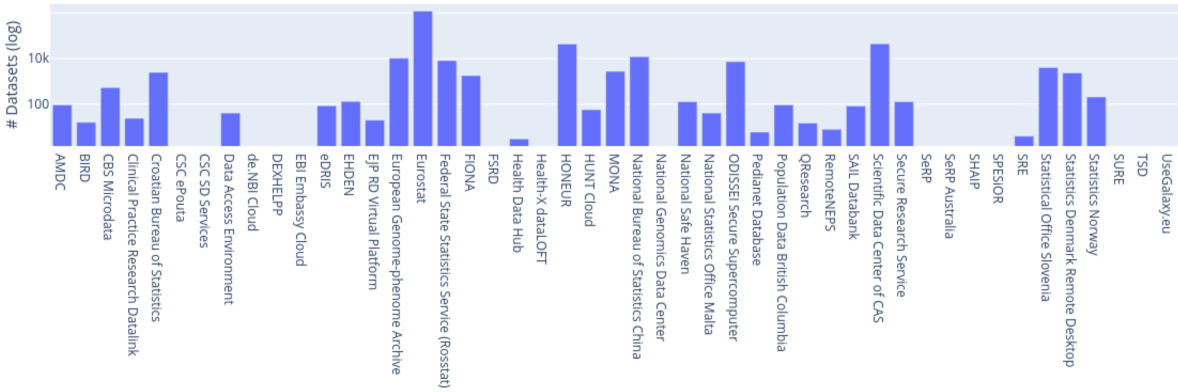

Figure 3. Sensitive data records available in TREs for researchers.

The EJP RD Virtual Platform (Kaliyaperumal et al., 2021) was launched very briefly before writing this literature study in June 2023. It is in the process of onboarding data resources from the rare disease research community at different levels of integration, starting with basic textual descriptions, followed by deeper discovery, and finishing with federated querying and analysis of the data, providing extensive data curation.

The Health-X dataLOFT (Boll & Meyer, 2022) started operating in 2023, enabling the linkage of primary care data (e.g., electronic patient records) with secondary data (collected by fitness tracker devices). The first TRE, Clinical Practice Research Datalink (Herrett et al., 2015), started operating in 1993 and contains 24 datasets of primary care patient data.

Eurostat started in 2010 with measures to provide a remote access infrastructure to non-anonymized but de-identified data for research purposes from their national statistic offices, making the need for researchers to travel to Luxembourg, Belgium, obsolete under legislative regulations coming into effect earlier. Many TREs (n = 16, 34%) started operating during the COVID-19 pandemic (2019–2022), with 50% of the included TREs starting their operation between 2008 and 2020 (first and third quartile). Because physical visits to TREs are impossible, many TREs allow researchers remote access to sensitive data. In addition, some TREs that began operating during the COVID-19 pandemic allow for remote access and others who started operating before mainly allow physical visitation of the data. The available datasets fall into the major fields (c.f. Table) of Life Sciences, Physical Sciences, Mathematics and Engineering, and Social and Behavioral Sciences.

Sensitive Data Availability and Access

The most common mode of access identified in the literature survey was visiting the data remotely with secure technical measures, predominantly (n = 46, 98%) implementing safeguards similar to TREs and the Five Safes Principles (Hubbard, Reilly, Varma, & Seymour, 2020). Physical visits (n = 10, 21%) are mainly supported by statistical office TREs (n = 7, 70%) compared with non-SOs (n = 3, 30%).

Most TREs provide data from the Life Sciences (n = 23); this data is mainly available in TREs that are not SOs (see Table 2). The least supported mode of accessing sensitive data (c.f. Table) are external physical visits, allowing access to sensitive data through trusted or approved TREs that are closer to the researcher than the TRE holding the actual sensitive data. By federating access to the proxy TRE, the researcher can visit the sensitive data without the need to travel long distances. Sensitive data is available in three levels: 1) L1 is identifiable data (i.e., the US Census Bureau allows researchers to visit FSRBC infrastructures to work with identifiable, sensitive data, 2) L2 is de-identified data, such as pseudonymized data, all identifiable information has been removed, e.g., Scientific Use Files (SUFs), and 3) L3 is anonymous data such as Public Use Files (not considered sensitive anymore).

Table 2. Sensitive data available per major research fields in statistical offices (SO) and TREs that are not a statistical office (non-SO)

Major Field	TREs			
	SO		non-SO	
Social and Behavioral Sciences	*n* = 13	62%	*n* = 8	38%
Physical Sciences, Mathematics, and Engineering	*n* = 12	71%	*n* = 5	29%
Life Sciences	*n* = 6	21%	*n* = 23	79%
Arts and Humanities	*n* = 0	0%	*n* = 1	100%

Table 3. Sensitive data availability for researchers at SOs and TREs that are non-SO in three levels of data sensitivity (L1 = identifiable data, L2 = de-identified data, and L3 = anonymized data).

Data access method	TREs (*n*)					
	SO			non-SO		
	L1	L2	L3	L1	L2	L3
Scientific use files	0	7	0	0	2	1
Physical visit	0	7	0	0	3	0
External physical visit	1	1	0	0	1	0
Remote data visit	0	11	1	0	31	3

This survey revealed that SUFs are prevalent in SOs (n = 7, 70%) in Europe (n = 9, 90%).

Strengths and Limitations

This literature study did not use qualitative research methods (i.e., a survey with each TRE infrastructure provider was not performed) but used information available publicly, the conclusions drawn reflect the information that can be gathered with reasonable effort and time. The available data was categorized into four fields using the well-established taxonomy of fields developed by the National Research Committee of the USA.

In some cases, adequate information could not be found regarding the sensitive data records available: 1) National Bureau of Statistics of China: to the best of our knowledge, no definite number of available records is publicly available; therefore, their public endpoint with a bash-script was crawled, and 2) Federal State Statistics Service Russia: during this literature study all non-Russian connect requests were denied, [7] or unavailable.[8] In total, 8.217 records were found using a proxy with a Russian IP address (Akatkin & Yasinovskaya (2019) reported 12.187 data sets).

[7] https://fedstat.ru/, no access September 2023 to January 2024
[8] https://data.gov.ru, no access September 2023 to January 2024

Conclusions and Future Work

High-quality data from trusted sources such as TREs is essential for evidence-based research. Publicly, little information is available on the technical implementation and descriptions of building blocks and their slight variations. By giving researchers access to sensitive data under strict technical security assertions and organizational and legal measurements, researchers can improve the state of the art in many research domains, balancing transparency against privacy, especially in health and sharing of genomic data.

In total, 47 TREs, most of which exist in Europe (n = 39, 83%), followed by Asia (n = 3), North America (n = 3) and Oceania (n = 2), provide access to sensitive data (raw data or the possibility to link with external data) of which two-thirds provide sensitive data themselves. Their technical characteristics were analysed, and a very balanced mix of TREs was found that provided raw data (n = 23, 49%) compared with a platform (n = 24, 51%), for example, to link sensitive data. Remote data visiting was predominant (98%) compared with external physical visits, where researchers visit an authorized external TRE physically instead of accessing data through the statistical office directly, as seen in the case of FSRDCs in the USA. The SOs provide the majority (92%) of available sensitive data in our literature study.

In this initial literature study, we started on a broad exploration of existing TREs using public information. In the future, on-site visits should be conducted for selected TREs, compiling a multi-case study with qualitative research methods and observations of day-to-day activities, which would gain more insight into the operational characteristics of TREs that will (normally) never be published, containing anonymized reports of near accidents, common pitfalls, and procurement, and giving access to sensitive data. Further, taking the initial building blocks for the secure handling of sensitive data from this technical blueprint (Weise, Kovacevic, Popper, & Rauber, 2022), this description could be extended into a machine-understandable context from taxonomies such as Computer Science Ontology.[9]

Supplementary Material

All analysis results and all scripts used to query the metadata catalogues are available as Jupyter Notebook in our code repository.[10]

References

Akatkin, Y., & Yasinovskaya, E. (2019). Data-Centricity as the Key Enabler of Digital Government: Is Russia Ready for Digital Transformation of Public Sector. *Electronic Governance and Open Society: Challenges in Eurasia*, 439–454. doi:10.1007/978-3-030-13283-5_33

Bardenheuer, K., Van Speybroeck, M., Hague, C., Nikai, E., & Price, M. (2022). Haematology Outcomes Network in Europe (HONEUR) - A collaborative, interdisciplinary platform to harness the potential of real-world data in hematology. *European Journal of Haematology*, 109 (2), 138–145. doi:10.1111/ejh.13780

Barkow, I., Leopold, T., Raab, M., Schiller, D., Wenzig, K., Blossfeld, H.-P., & Rittberger, M. (2011). Remote NEPS: Data dissemination in a collaborative workspace. *Zeitschrift für Erziehungswissenschaft*, 14 (S2), 315–325. doi:10.1007/s11618-011-0192-5

[9] https://cso.kmi.open.ac.uk/home

[10] https://gitlab.tuwien.ac.at/martin.weise/tres

Bastião Silva, L., Trifan, A., & Luís Oliveira, J. (2018). MONTRA: An agile architecture for data publishing and discovery. *Computer Methods and Programs in Biomedicine*, 160, 33–42. doi:10.1016/j.cmpb.2018.03.024

Boll, S., & Meyer, J. (2022). Health-X dataLOFT: A Sovereign Federated Cloud for Personalized Health Care Services. *IEEE MultiMedia*, 29 (1), 136–140. doi:10.1109/mmul.2022.3153121

Bruno, G., D'Aurizio, L., & Tartaglia-Polcini, R. (2014). Remote Processing of Business Microdata at the Bank of Italy. *Statistical Methods and Applications from a Historical Perspective*, 239–249. doi:10.1007/978-3-319-05552-7_21

Cantarutti, A., & Gaquinto, C. (2021). Pedianet Database. *Springer Series on Epidemiology and Public Health*, 159–164. doi:10.1007/978-3-030-51455-6_13

Cook, C. E., Bergman, M. T., Finn, R. D., Cochrane, G., Birney, E., & Apweiler, R. (2015). The European Bioinformatics Institute in 2016: Data growth and integration. *Nucleic Acids Research*, 44 (D1), 20–26. doi:10.1093/nar/gkv1352

Cuggia, M., & Combes, S. (2019). The french health data hub and the german medical informatics initiatives: Two national projects to promote data sharing in healthcare. *Yearbook of Medical Informatics*, 28 (01), 195–202. doi:10.1055/s-0039-1677917

Desai, T., Ritchie, F., & Welpton, R. (2016). Five Safes: Designing Data Access for Research. In Economics working paper series.

Deumens, E., Samant, T., Byrd, E., Porter, S., Haenssen, A., Hillegas, C. W. W., … Kopaliani, I. (2021). Secure Research Infrastructure Using tiCrypt. *Practice and Experience in Advanced Research Computing*. doi:10.1145/3437359.3465577

de Zeeuw, E. L., Voort, L., Schoonhoven, R., Nivard, M. G., Emery, T., Hottenga, J.-J., … Boomsma, D. I. (2021). Safe Linkage of Cohort and Population-based Register Data in a Genomewide Association Study on Health Care Expenditure. *Twin Research and Human Genetics*, 24 (2), 103–109. doi:10.1017/thg.2021.18

Ekaputra, F. J., Ekelhart, A., Mayer, R., Miksa, T., Šarčević, T., Tsepelakis, S., & Waltersdorfer, L. (2022). Semantic-enabled Architecture for Auditable Privacy-preserving Data Analysis. *Semantic Web*, 1–34. doi:10.3233/sw-212883

Ford, D. V., Jones, K. H., Verplancke, J.-P., Lyons, R. A., John, G., Brown, G., … Leake, K. (2009). The SAIL Databank: Building a National Architecture for e-health Research and Evaluation. *BMC Health Services Research*, 9 (1). doi:10.1186/1472-6963-9-157

Fuchs, R., Göllner, T., Hartmann, S., & Thomas, T. (2023). Fostering Excellent Research by the Austrian Micro Data Centre. *Jahrbücher für Nationalökonomie und Statistik*. doi:10.1515/jbnst-2023-0043

Gao, C., McGilchrist, M., Mumtaz, S., Hall, C., Anderson, L. A., Zurowski, J., … Jefferson, E. (2022). A National Network of Safe Havens: Scottish Perspective. *Journal of Medical Internet Research*, 24 (3), e31684. doi:10.2196/31684

Habermann, H. (2010). Future of Innovation in the Federal Statistical System. *The Annals of the American Academy of Political and Social Science*, 631 (1), 194–203. doi:10.1177/0002716210374014

Herrett, E., Gallagher, A. M., Bhaskaran, K., Forbes, H., Mathur, R., van Staa, T., & Smeeth, L. (2015). Data Resource Profile: Clinical Practice Research Datalink. *International Journal of Epidemiology*, 44 (3), 827–836. doi:10.1093/ije/dyv098

Hippisley-Cox, J., Stables, D., & Pringle, M. (2004). QRresearch: a new general practice database for research. *Journal of Innovation in Health Informatics*, 12 (1), 49–50. doi:10.14236/jhi.v12i1.108

Hjelm, C.-G. (2005). MONA - Microdata Online Access at Statistics Sweden. In *Monographs of official statistics. work session on statistical data confidentiality* (pp. 21–27). Office for Official Publications of the European Communities

Hoffmann, N., Maus, I., Beier, S., Belmann, P., Krüger, J., Tauch, A., … Sczyrba, A. (2023). Embedding the de.NBI Cloud in the National Research Data Infrastructure Activities. *Proceedings of the Conference on Research Data Infrastructure*, 1. doi:10.52825/cordi.v1i.387

Hubbard, T., Reilly, G., Varma, S., & Seymour, D. (2020). Trusted Research Environments Green Paper. Zenodo. doi:10.5281/zenodo.4594704

Jalili, V., Afgan, E., Gu, Q., Clements, D., Blankenberg, D., Goecks, J., … Nekrutenko, A. (2020). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2020 update. *Nucleic Acids Research*, 48 (W1), W395–W402. doi:10.1093/nar/gkaa434

Jones, K. H., Ford, D. V., Ellwood-Thompson, S., & Lyons, R. A. (2016). The UK Secure eResearch Platform for public health research: A case study. *The Lancet*, 388, S62. doi:10.1016/S0140-6736(16)32298-X

Kaliyaperumal, R., Queralt Rosinach, N., Burger, K., Bonino da Silva Santos, L. O., Hanauer, M., & Roos, M. (2021). Enabling FAIR Discovery of Rare Disease Digital Resources. *Studies in Health Technology and Informatics*. doi:10.3233/shti210101

Križman, I., Klanjšček, M., & Čerk, A. (2002). Management, organisational and policy issues related to electronic data reporting in the Statistical Office of the Republic of Slovenia. *Statistical Journal of the United Nations Economic Commission for Europe*, 19 (3), 141–152. doi:10.3233/sju-2002-19304

Krokstad, S., Sund, E. R., Kvaløy, K., Rangul, V., & Næss, M. (2022). HUNT for better public health. *Scandinavian Journal of Public Health*, 50 (7), 968–971. doi:10.1177/14034948221102309

Langsrud, Ø. (2019). Information preserving regression-based tools for statistical disclosure control. *Statistics and Computing*, 29 (5), 965–976. doi:10.1007/s11222-018-9848-9

Lappalainen, I., Almeida-King, J., Kumanduri, V., Senf, A., Spalding, J. D., ur-Rehman, S., …. Flicek, P. (2015). The European Genome-phenome Archive of Human Data Consented for Biomedical Research. *Nature Genetics*, 47 (7), 692–695. doi:10.1038/ng.3312

O'Keefe, C. M., Gould, P., & Churches, T. (2014). Comparison of Two Remote Access Systems Recently Developed and Implemented in Australia. In *Privacy in statistical databases* (pp. 299–311). doi:10.1007/978-3-319-11257-2_23

Ostriker, J. P., Kuh, C. V., & Voytuk, J. A. (2003). Assessing Research-doctorate Programs: A Methodology Study. doi:10.17226/10859

Øvrelid, E., Bygstad, B., & Thomassen, G. (2021). TSD: A Research Platform for Sensitive Data. *Procedia Computer Science*, 181, 127–134. doi:10.1016/j.procs.2021.01.112

Pavis, S., & Morris, A. D. (2015). Unleashing the power of administrative health data: The scottish model. *Public Health Research & Practice*, 25 (4). doi:10.17061/phrp2541541

Pencarrick Hertzman, C., Meagher, N., & McGrail, K. M. (2013). Privacy by Design at Population Data BC. *Journal of the American Medical Informatics Association*, 20 (1), 25–28. doi:10.1136/amiajnl-2012-001011

Poljicak, M., & Stancic, H. (2014). Statistical Microdata & Production of Safe Datasets, Transparent Presentation of Contents and Advanced Services for Users through Metadata Authorization System. *37th International Convention on Information and Communication Technology, Electronics and Microelectronics*. doi:10.1109/mipro.2014.6859808

Popper, N., Endel, F., Mayer, R., Bicher, M., & Glock, B. (2017). Planning Future Health: Developing Big Data and System Modelling Pipelines for Health System Research. *Simulation Notes Europe*, 27 (4), 203–208. doi:10.11128/sne.27.tn.10396

Rat für Sozial- und Wirtschaftsdaten. (2019). Remote Access zu Daten der amtlichen Statistik und der Sozialversicherungsträger. *Rat für Sozial- und Wirtschaftsdaten Output*, 5. doi:10.17620/02671.42

Reuter, W. H., & Museux, J.-M. (2010). Establishing an Infrastructure for Remote Access to Microdata at Eurostat. *Privacy in Statistical Databases*, 249–257. doi:10.1007/978-3-642-15838-4_22

Ritchie, F. (2008). Secure Access to Confidential Microdata: Four Years of the Virtual Microdata Laboratory. *Economic & Labour Market Review*, 2 (5), 29–34. doi:10.1057/elmr.2008.73

Rose, J., Dong, W., Kim, U., Hnath, J., Statler, A., Saroufim, P., … Koroukian, S. M. (2022). An Informatics Infrastructure to Catalyze Cancer Control Research and Practice. *Cancer Causes & Control*, 33 (6), 899–911. doi:10.1007/s10552-022-01571-0

Schouten, B., & Cigrang, M. (2003). Remote Access Systems for Statistical Analysis of Microdata. *Statistics and Computing*, 13 (4), 381–389. doi:10.1023/a:1025675124850

Soini, E., Hallinen, T., & Martikainen, J. (2022). Comparison of Certified Secure Processing Environments – Novel Platforms for Collection, Augmentation, Structuring, Transfer, Management, Analysis, Sharing, Reporting, and Storage of Sensitive Data? *Value in Health*, 25 (12), S460. doi:10.1016/j.jval.2022.09.2288

Thygesen, L. (1995). The Register-based System of Demographic and Social Statistics in Denmark – An Overview. *Statistical Journal of the United Nations Economic Commission for Europe*, 12 (1), 49–55. doi:10.3233/sju-1995-12106

Weise, M., Kovacevic, F., Popper, N., & Rauber, A. (2022). OSSDIP: Open Source Secure Data Infrastructure and Processes Supporting Data Visiting. *Data Science Journal*, 21. doi:10.5334/dsj-2022-004

Wilde, K., Anderson, L., Boyle, M., Pinder, A., & Weir, A. (2022). Introducing a new Trusted Research Environment - the Safe Haven Artificial Platform. *International Journal of Population Data Science*, 7 (3). doi:10.23889/ijpds.v7i3.2056

Wood, A., Denholm, R., Hollings, S., Cooper, J., Ip, S., Walker, V., … Sudlow, C. (2021). Linked electronic health records for research on a nationwide cohort of more than 54 million people in england: Data resource. *BMJ*, n826. doi:10.1136/bmj.n826

Xue, Y., Bao, Y., Zhang, Z., Zhao, W., Xiao, J., He, S., … Chen, R. (2022). Database Resources of the National Genomics Data Centre, China National Centre for Bioinformation in 2023. *Nucleic Acids Research*, 51 (D1), D18–D28. doi:10.1093/nar/gkac1073

Zhang, C., & Hou, L. (2020). Data Middle Platform Construction: The Strategy and Practice of National Bureau of Statistics of China. *Statistical Journal of the IAOS*, 36 (4), 979–986. doi:10.3233/sji-200754

Zhang, Z., Zhao, W., Xiao, J., Bao, Y., He, S., Zhang, G., … Xue, Y. (2019). Database Resources of the National Genomics Data Centre in 2020. *Nucleic Acids Research*. doi:10.1093/nar/gkz913

Zilhão, M. J., Baldacci, E., Bergdahl, H., & Djerf, K. (2022). Peer Review Report, 1–42. [accessed 2023-09-20]. Retrieved from https://nso.gov.mt/wp-content/uploads/Final-per-review-report-Malta-.pdf