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 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 & their slight technical variations. To shine light on these problems, we give an overview of existing, publicly described TREs and a bibliography linking to the system description. We further analyze their technical characteristics, especially in their commonalities & variations and provide insight on their data type characteristics and availability. Our literature study shows that 47 TREs worldwide provide access to sensitive data of which two-thirds provide data themselves, predominantly via secure remote access. Statistical offices make available a majority of available sensitive data records included in this study.

Submitted October 1st 2023 ~ *Accepted* November 9th 2023

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 (i) Life Sciences (ii) Physical Sciences, Mathematics, and Engineering (iii) Arts and Humanities, and (iv) Social and Behavioral Sciences. Policy makers globally have recognized the value of research with sensitive data in the last years and allowed the use of this data for scientific purposes in many cases while 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 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 to work with sensitive data at trusted partner-TREs, a modus operandi of Federal Statistical Research Data Centers (FSRDCs) in the United States that allows researchers to work with census data in any of the 33[1] FSRDCs. Protection and maintaining control over digital sensitive data, confidential data or data related to intellectual property while also striving to give third parties access to the data poses a significant challenge. Trusted Research Environments (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.

We use the umbrella term *sensitive* interchangeably 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 United States, Secure Data Environment (SDE) in Great Britain, Sensitive Data Service (SDS) in the Scandinavian Countries and Secure Processing Environment (SPE) in a Pan-European context. The goal for this literature study is to discover existing TREs, analyze their characteristics and data availability to give an overview on available infrastructure for sensitive data research as many European initiatives have been emerging in recent months. Our main contributions towards "transparency and trust in research practices" are:

- Comprehensive list of available TREs and bibliography containing system descriptions
- Analysis on TRE operation characteristics
- Analysis on sensitive data availability and access

Methodology

In this literature study, we identified existing TREs and available datasets globally using scholarly databases (Scopus, Web of Science, IEEE Xplore, Science Direct), a computer science bibliography[2], Google and grey literature, focusing on retrieval of the following source material:

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

[1] https://www.census.gov/about/adrm/fsrdc/locations.html, accessed 2024-02-09
[2] https://dblp.org/, accessed 2023-09-25

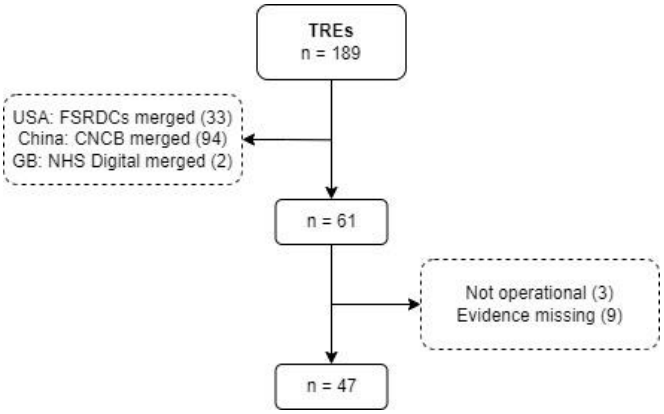

Figure 1: Flow chart of literature selection.

We included TREs (c.f. Figure 1) based into the inventory based on the following criteria: the infrastructure must implement safeguards similar to the five safes framework introduced by Desai et al. [11], describing a safe center for sensitive research data in five risk dimensions. This already excludes research data repositories who cannot provide a safe environment to analyze sensitive data. To be included in the inventory, each infrastructure must additionally fulfil the requirements of a TRE defined by Hubbard et al. [22], who extend the *safe setting* formulated by Desai et al. with safe computing and the possibility to safely map the research results back to e.g. individual clinical care in a *safe return* requirement.

For each infrastructure in the inventory, we merged related infrastructures that have a common governance board (e.g. 33 Federal Statistical Research Data Centers (FSRDCs) governed by the US Census Bureau) and copies of the same infrastructure. We, for the time being, excluded TREs that are not in operation but e.g. provide reference models [43][13] or do not have a sufficient evidence that describes the infrastructure operation. This especially holds for CSC ePouta & CSC SD Services in Finland who did not publish a peer-reviewed article, but provide sufficient evidence that a TRE is in operation nonetheless. Eliciting this information via other means such as interviews or on-site visits, was out of scope for this initial study to showcase publicly available information. These are being kept in a backlog for further investigation.

We determined the number of datasets available in the metadata catalogs as follows: if not provided by the TRE website or reports, we (i) used the public metadata catalog API, (ii) de-structured collections, e.g. the HUNT studies (of the HUNT Cloud [27]), which are available as 6 collections that partly have iterations, i.e. repetitions, but contain in total 56 separate studies, we therefore count 56 datasets, (iii) scraped public endpoints, e.g. the Croatian Bureau of Statistics has an undocumented endpoint[3] available in Croatian and English language that accepts thematic abbreviations, producing HTML that can be scraped and filtered. For merged TREs, we did not collect the number of datasets available but did not exclude them in the literature study to acknowledge their importance.

Results

TRE Operation Characteristics

TREs are predominantly located 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 Great Britain (n = 10, 21%), Finland (n = 4, 9%) followed by China, USA, Germany and Norway (n = 3, 6%). TREs operating within the European Union (n = 24, 51%) benefit from the common

[3] https://intra.dzs.hr/cat/v2/list/, accessed 2023-09-20

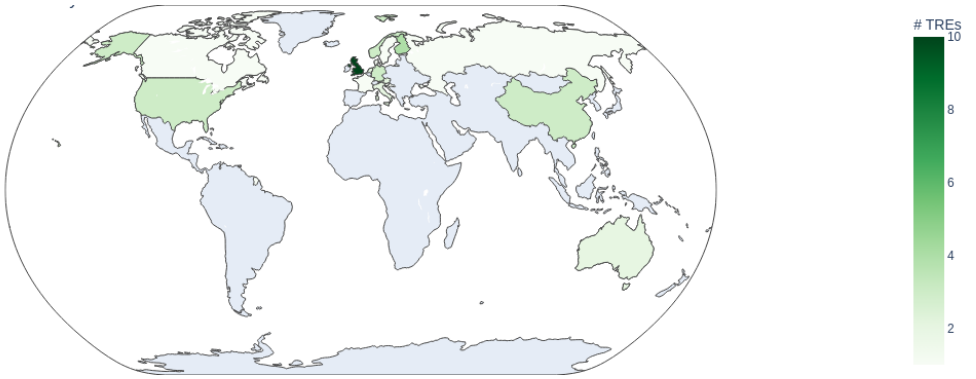

Figure 2: Trusted Research Environments globally.

legislative norms in place to protect individual's privacy under the General Data Protection Regulation. We compared the 47 TREs (c.f. Table 1) for architectural similarities and differences. A map of TREs aggregated by country is shown in Figure 2, excluding Pan-European TREs (n = 5, 11%). We found a very balanced mix of TREs that provide raw data (n = 23, 49%) and provide a secure platform (n = 24, 51%) to e.g. link data securely. In general, prior to any research on sensitive data a request must be sent to the Data Owner containing (i) personal data to identify the Analyst (ii) required data (iii) required tools to perform the analysis (iv) task and research questions that should be answered with the required data [43].

We identified 31 TREs in total that have a public metadata catalog, of which 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 = 1.2M, 90%) available is Eurostat [37], the head statistical office in the European Union (c.f. Figure 3). A metadata catalog[4] exists for human interaction as well as multiple machine-actionable interfaces such as SDMX[5], bulk downloads and REST API endpoints.

SeRP [24] provides infrastructure solutions to their *tenants*[6] who each have their own technical and governance requirements depending on the data they hold, how it is generated and what to do with it, who funds them and regional/national data landscape in which they operate. Access to data is enabled directly by the tenant's unique access conditions.

The EJP RD Virtual Platform [25] was launched only very briefly before writing this literature study in June 2023 and is in the process of onboarding data resources from the rare-disease research community in different levels of integration, starting with basic textual descriptions, followed deeper discovery and finishing with federated querying and analysis on the data, providing extensive data curation.

The Health-X dataLOFT [5] started its operation too in 2023, enabling linkage of primary care data (i.e. electronic patient records) with secondary data (collected by fitness tracker devices). The first TRE, Clinical Practice Research Datalink [18] started operation in 1993 and contains 24 datasets of primary care patient data.

Eurostat started 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 in effect earlier.

Many TREs (n = 16, 34%) started their operation during the COVID-19 pandemic (2019 - 2022) with 50% of included TREs starting their operation between Q1 = 2008 and Q3 = 2020. Since physical visits to TREs are not possible, many TREs allow researchers remote access to the sensitive data. Many TREs who began their operation during the COVID-19 pandemic allow for

[4] https://ec.europa.eu/eurostat/data/database/, accessed 2023-09-21
[5] https://sdmx.org/, accessed 2023-09-21
[6] https://serp.ac.uk/tenants/, accessed 2024-02-08

Name	Publication	Country	First Available
SURE	[30]	AUS	2017
SeRP Australia	[24]	AUS	2011
DEXHELPP	[35]	AUT	2017
AMDC	[15]	AUT	2022
Population Data British Columbia	[33]	CAN	2008
Croatian Bureau of Statistics	[34]	HRV	2020
Scientific Data Center of CAS	[48]	CHN	2019
National Genomics Data Center	[46]	CHN	2022
National Bureau of Statistics China	[47]	CHN	2020
Statistics Denmark Remote Desktop	[42]	DNK	2008
EHDEN	[4]	Europe	2022
EJP RD Virtual Platform	[25]	Europe	2023
European Genome-phenome Archive	[29]	Europe	2008
Health-X dataLOFT	[5]	Europe	2023
HONEUR	[2]	Europe	2018
CSC ePouta		FIN	2020
CSC SD Services		FIN	2022
FIONA	[36]	FIN	2010
SPESiOR	[41]	FIN	2022
de.NBI Cloud	[21]	DEU	2017
UseGalaxy.eu	[23]	DEU	2016
RemoteNEPS	[3]	DEU	2011
Pedianet Database	[7]	ITA	1998
BIRD	[6]	ITA	2007
Eurostat	[37]	LUX	2013
ODISSEI Secure Supercomputer	[10]	NLD	2018
CBS Microdata	[40]	NLD	2003
HUNT Cloud	[27]	NOR	2020
TSD	[50]	NOR	2014
Statistics Norway	[28]	NOR	2019
National Statistics Office Malta	[49]	MLT	2021
Federal State Statistics Service	[1]	RUS	2002
Statistical Office of the Republic of Slovenia	[26]	SVN	2006
MONA	[20]	SWE	2005
eDRIS	[32]	GBR	2013
SHAIP	[44]	GBR	2022
EBI Embassy Cloud	[8]	GBR	2013
Data Access Environment	[45]	GBR	2019
Secure Research Service	[38]	GBR	2004
SAIL Databank	[14]	GBR	2009
National Safe Haven	[16]	GBR	2020
SeRP	[24]	GBR	2011
Clinical Practice Research Datalink	[18]	GBR	1993
QResearch	[19]	GBR	2004
FSRDC	[17]	USA	1982
SRE	[39]	USA	2019
Secure Research Infrastructure	[12]	USA	2021

Table 1: TREs included in this literature study.

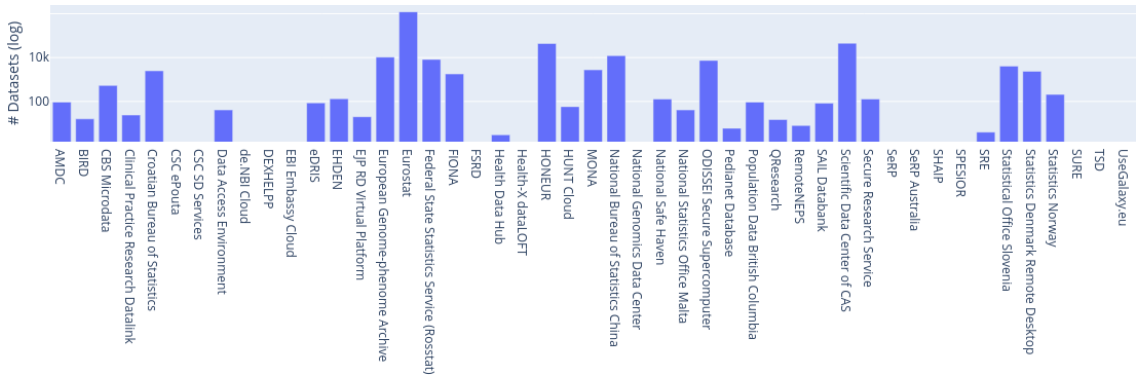

Figure 3: Sensitive data records available in TREs for researchers.

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

remote access, while others who started their operation before mostly only allow physical visitation of the data. The available datasets fall into the major fields (c.f. Introduction) of Life Sciences; Physical Sciences, Mathematics and Engineering & Social and Behavioral Sciences.

Sensitive Data Availability and Access

The most common mode of access identified in our literature survey is visiting the data remotely through secure technical measures, predominantly (n = 46, 98%) implementing safeguards similar to TREs [22] and the Five Safes Principles [11]. Physical visits (n = 10, 21%) are mostly supported by statistic office TREs (n = 7, 70%) in comparison to non-statistical offices (n = 3, 30%).

Most TREs provide data from the Life Sciences (n = 23), this data is mostly available in TREs that are not a statistical office (c.f. Table 2). The least supported mode of accessing sensitive data (c.f. Table 3) are external physical visits, allowing access to sensitive data through trusted/approved TREs in closer proximity 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 far distances.

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

Table 3: Sensitive data availability for researchers at statistical offices (SO) and TREs that are not a statistical office (non-SO) in three levels of data sensitivity (L1=identifiable data, L2=de-identified data, L3=anonymized data).

Sensitive data is available in three levels: (L1) identifiable data, i.e. the US Census Bureau allows researchers to visit FSRBC infrastructures to work with identifiable sensitive data; (L2) de-identified data such as pseudonymized data, all identifiable information has been removed, e.g. Scientific Use Files; (L3) anonymous data such as Public Use Files (not considered sensitive anymore). Our survey reveals that Scientific Use Files (SUFs) are prevalent in statistical offices (n = 7, 70%) in Europe (n = 9, 90%).

Strengths and Limitations

Since this literature study did not make use of qualitative research methods (i.e. we did not perform a survey with each TRE infrastructure provider) but used information available publicly (c.f. Methodology), the conclusions drawn reflect the information that can be gathered with reasonable effort and time. The available data is categorized into four, using the well-established Taxonomy of Fields developed by the National Research Committee of the United States.

In some cases, we were not able to find adequate information regarding the sensitive data records available: (i) National Bureau of Statistics of China: to the best of our knowledge, no definite number of available records is publicly available, so we crawled their public endpoint with a bash-script; (ii) Federal State Statistics Service Russia: during our literature study all non-Russian connect requests were denied[7] or unavailable[8]. We found 8.217 records using a proxy with Russian IP address (note that Atakin & Yasinovskaya [1] reported 12.187 data sets in 2019).

Conclusion & Future Work

High-quality data from trusted sources such as Trusted Research Environments (TREs) is essential for evidence-based research. Little information is available publicly on the technical implementation as well as descriptions of building blocks and their slight variations. By giving researchers access to sensitive data under strict technical security assertions, 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. We found 47 TREs, of which most exist in Europe (n = 39, 83%), followed by Asia (n = 3), North America (n = 3) and Oceania (n= 2), providing access to sensitive data (raw data or possibility to link with external data) of which two-thirds providing sensitive data themselves. We have analyzed their technical characteristics and found that a very balanced mix of TREs that provide raw data themselves (n = 23, 49%) compared to a platform (n = 24, 51%) to e.g. link sensitive data. Remote data visiting is predominant (98%) in contrast to external physical visits where researchers visit an authorized external TRE physically instead of accessing data through the statistical office directly as seen via the case for Federal Statistical Research Data Centers (FSRDCs) in the United States. Statistical offices provide a majority (92%) of available sensitive data in our literature study.

Starting from this initial literature study who focuses on broad exploration of existing TREs using public information. We want to conduct on-site visits of selected TREs soon, compiling a multi-case study through qualitative research methods and observation of day-to-day activities, we want to give more insight into the operational characteristics of TREs who will (normally) never get published, containing anonymized reports of near-accidents, common pitfalls, but also

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

procurement- and giving access to sensitive data. Further, taking our initial building blocks for secure handling of sensitive data from our technical blueprint [43], we want to extend this description with machine-understandable context from taxonomies such as the Computer Science Ontology[9].

Supplementary Material

All analysis results as well as all script used to query the metadata catalogs are available as Jupyter Notebook in our code repository[10].

References

[1] Y. Akatkin and E. Yasinovskaya. 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*, page 439–454, 2019. doi:10.1007/978-3-030-13283-5_33.

[2] K. Bardenheuer, M. Van Speybroeck, et al. 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, 2022. doi:10.1111/ejh.13780.

[3] I. Barkow, T. Leopold, M. Raab, et al. Remote NEPS: Data Dissemination in a Collaborative Workspace. *Zeitschrift für Erziehungswissenschaft*, 14(S2):315–325, 2011. doi:10.1007/s11618-011-0192-5.

[4] L. Bastião Silva, A. Trifan, and J. Luís Oliveira. MONTRA: An agile Architecture for Data Publishing and Discovery. *Computer Methods and Programs in Biomedicine*, 160:33–42, 2018. doi:10.1016/j.cmpb.2018.03.024.

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

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

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

[8] C. E. Cook, M. T. Bergman, R. D. Finn, et al. The European Bioinformatics Institute in 2016: DataGgrowth and Integration. *Nucleic Acids Research*, 44(D1):20–26, 2015. doi:10.1093/nar/gkv1352.

[9] https://cso.kmi.open.ac.uk/home, accessed 2024-02-09
[10] https://gitlab.tuwien.ac.at/martin.weise/tres, accessed 2024-02-08

[9] M. Cuggia and S. Combes. 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, 2019. doi:10.1055/s-0039-1677917.

[10] E. L. de Zeeuw, L. Voort, et al. 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, 2021. doi:10.1017/thg.2021.18.

[11] T. Desai, F. Ritchie, and R. Welpton. Five Safes: Designing Data Access for Research. In *Economics Working Paper Series*. 2016.

[12] E. Deumens, T. Samant, E. Byrd, et al. Secure Research Infrastructure Using tiCrypt. Practice and Experience in *Advanced Research Computing*, 2021. doi:10.1145/3437359.3465577.

[13] F. J. Ekaputra, A. Ekelhart, R. Mayer, et al. Semantic-enabled Architecture for Auditable Privacy-preserving Data Analysis. *Semantic Web*, page 1–34, 2022. doi:10.3233/sw-212883.

[14] D. V. Ford, K. H. Jones, J.-P. Verplancke, et al. The SAIL Databank: Building a National Architecture for e-health Research and Evaluation. *BMC Health Services Research*, 9(1), 2009. doi:10.1186/1472-6963-9-157.

[15] R. Fuchs, T. Göllner, S. Hartmann, et al. Fostering Excellent Research by the Austrian Micro Data Center. *Jahrbücher für Nationalökonomie und Statistik*, 2023. doi:10.1515/jbnst-2023-0043.

[16] C. Gao, M. McGilchrist, S. Mumtaz, et al. A National Network of Safe Havens: Scottish Perspective. *Journal of Medical Internet Research*, 24(3):e31684, 2022. doi:10.2196/31684.

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

[18] E. Herrett, A. M. Gallagher, K. Bhaskaran, et al. Data Resource Profile: Clinical Practice Research Datalink. *International Journal of Epidemiology*, 44(3):827–836, 2015. doi:10.1093/ije/dyv098.

[19] J. Hippisley-Cox, D. Stables, and M. Pringle. QRresearch: A New General Practice Database for Research. *Journal of Innovation in Health Informatics*, 12(1):49–50, 2004. doi:10.14236/jhi.v12i1.108.

[20] C.-G. Hjelm. MONA - Microdata Online Access at Statistics Sweden. Monographs of official statistics. Work session on statistical data confidentiality, page 21–27. 2005.

[21] N. Hoffmann, I. Maus, S. Beier, et al. Embedding the de.NBI Cloud in the National Research Data Infrastructure Activities. *Proceedings of the Conference on Research Data Infrastructure*, 1, 2023. doi:10.52825/cordi.v1i.387.

[22] T. Hubbard, G. Reilly, S. Varma, et al. Trusted Research Environments Green Paper, 2020. doi:10.5281/zenodo.4594704.

[23] V. Jalili, E. Afgan, Q. Gu, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2020 update. *Nucleic Acids Research*, 48(W1):W395–W402, 2020. doi:10.1093/nar/gkaa434.

[24] K. H. Jones, D. V. Ford, S. Ellwood-Thompson, et al. The UK Secure eResearch Platform for public health research: a case study. *The Lancet*, 388:S62, Nov 2016. doi:10.1016/S0140-6736(16)32298-X.

[25] R. Kaliyaperumal, N. Queralt Rosinach, K. Burger, et al. Enabling FAIR Discovery of Rare Disease Digital Resources. *Studies in Health Technology and Informatics*, 2021. doi:10.3233/shti210101.

[26] I. Križman, M. Klanščjek, and A. Čerk. 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, 2002. doi:10.3233/sju-2002-19304.

[27] S. Krokstad, E. R. Sund, K. Kvaløy, et al. HUNT for better Public Health. Scandinavian *Journal of Public Health*, 50(7):968–971, 2022. doi:10.1177/ 14034948221102309.

[28] Ø. Langsrud. Information Preserving Regression-based Tools for Statistical Disclosure Control. *Statistics and Computing*, 29(5):965–976, 2019. doi:10.1007/s11222-018-9848-9.

[29] I. Lappalainen, J. Almeida-King, V. Kumanduri, et al. The European Genome-phenome Archive of Human Data Consented for Biomedical Research. *Nature Genetics*, 47(7):692–695, 2015. doi:10.1038/ng.3312.

[30] C. M. O'Keefe, P. Gould, and T. Churches. Comparison of Two Remote Access Systems Recently Developed and Implemented in Australia. In *Privacy in Statistical Databases*, pages 299–311, 2014. doi:10.1007/978-3-319-11257-2_23.

[31] J. P. Ostriker, C. V. Kuh, and J. A. Voytuk. Assessing Research-doctorate Programs: A Methodology Study. *The National Academies Press*, 2003. doi:10.17226/10859.

[32] S. Pavis and A. D. Morris. Unleashing the Power of Administrative Health Data: the Scottish Model. *Public Health Research & Practice*, 25(4), 2015. doi:10.17061/phrp2541541.

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

[34] M. Poljicak and H. Stancic. 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*, 2014. doi:10.1109/mipro.2014.6859808.

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

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

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

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

[39] J. Rose, W. Dong, U. Kim, et al. An Informatics Infrastructure to Catalyze Cancer Control Research and Practice. *Cancer Causes & Control*, 33(6):899–911, 2022. doi:10.1007/s10552-022-01571-0.

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

[41] E. Soini, T. Hallinen, and J. Martikainen. 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, 2022. doi:10.1016/j.jval.2022.09.2288.

[42] L. Thygesen. 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, 1995. doi:10.3233/sju-1995-12106.

[43] M. Weise, F. Kovacevic, N. Popper, et al. OSSDIP: Open Source Secure Data Infrastructure and Processes Supporting Data Visiting. *Data Science Journal*, 21, 2022. doi:10.5334/dsj-2022-004.

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

[45] A. Wood, R. Denholm, S. Hollings, et al. Linked Electronic Health Records for Research on a Nationwide Cohort of more than 54 Million People in England: Data Resource. *BMJ*, page n826, 2021. doi:10.1136/bmj.n826.

[46] Y. Xue, Y. Bao, Z. Zhang, et al. Database Resources of the National Genomics Data Center, China National Center for Bioinformation in 2023. *Nucleic Acids Research*, 51(D1):D18–D28, 2022. doi:10.1093/nar/gkac1073.

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

[48] Z. Zhang, W. Zhao, J. Xiao, et al. Database Resources of the National Genomics Data Center in 2020. *Nucleic Acids Research*, 2019. doi:10.1093/nar/gkz913xue2022database.

[49] M. J. Zilhão, E. Baldacci, H. Bergdahl, et al. Peer Review Report. pages 1–42, 2022. [accessed 2023-09-20]. URL: https://nso.gov.mt/wp-content/uploads/Final-per-review-report-Malta-.pdf.

[50] E. Øvrelid, B. Bygstad, and G. Thomassen. TSD: A Research Platform for Sensitive Data. *Procedia Computer Science*, 181:127–134, 2021. doi:10.1016/j.procs.2021.01.112.12.