IJDC | *Brief Report*

ASSURED: Assuring Safe Research by Safe People

Deborah Wiltshire
GESIS Leibniz Institute for the Social Sciences

Simon Parker
Deutsches Krebsforschungszentrum

Vanessa González Ribao
Deutsches Krebsforschungszentrum

Abstract

The ASSURED project aims to address the need for standardised training for researchers and professionals working with sensitive data in Trusted Research Environments (TREs) in Germany. As data sharing for research continues to grow, safeguarding sensitive data is critical, particularly as Open Science and FAIR data principles promote wider access. However, ensuring secure data access while minimising risks requires robust safeguards, such as the Five Safes Model. An essential element of this model is the 'Safe People' component, emphasising the importance of well-trained individuals who understand data confidentiality and disclosure risks. Currently, training for researchers and TRE staff in Germany is inconsistent, with few formal systems in place. To remedy this, the ASSURED project has developed an e-learning programme offering flexible, modular training to ensure that researchers and TRE staff meet essential data security standards. The programme includes core modules applicable to all users, with additional role-specific content tailored to various TRE services. By integrating the training with an Authentication and Authorisation Infrastructure (AAI), the programme ensures streamlined tracking of completion and facilitates cross-service access. The ASSURED project aims to enhance data protection and support the European Open Science Cloud initiative, promoting responsible data use across borders.

Submitted 8 May 2025 ~ *Accepted* 12 May 2025

Correspondence should be addressed to Dr. Deborah Wiltshire. Email: deborah.wiltshire@gesis.org

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: Background Context

The volume of data generated for research or made available for research continues to grow. This has great potential to make new discoveries or to develop advancements that benefit society; however, it is not without legal and ethical challenges. Very often, the greatest benefits are realised by processing data that is sufficiently detailed, such that the identification of the data subjects is possible, that contains highly sensitive information, or both. These data present the highest risk to the data subjects, who could face discrimination or harm if their data is linked to them and misused. At the same time, recently, there has been a move towards Open Science and FAIR data principles, with the ambition that research data is as openly available as possible.

When research data are fully anonymised or can be anonymised, open sharing is seen as generally unproblematic. Usually, the anonymisation process requires details to be removed from the data; in some circumstances, data can be anonymised without a significant loss of utility; however, in other cases, the reduction in detail may limit the research potential. While anonymisation can help address the tension between a desire to make data accessible to researchers and minimise the risk of harm to data subjects, the potential limitations to research benefits have prompted the development of alternative methods for the safe sharing of data in a research context.

An important approach to this issue has been to implement safeguards that are appropriate and proportionate to the risk that the data poses to the data subjects. One such safeguard for particularly sensitive data[1] has been the creation and adoption of Secure Data Centres or Trusted Research Environments (TREs) within secure-access facilities (Bishop et al., 2022). These research infrastructures first emerged in the social sciences but have increasingly been adopted in health and genomic research.

The TREs are highly secure and controlled computing environments that allow approved researchers from authorised organisations a safe way to access, store, and analyse sensitive data. Researchers may access the data via an on-site Safe Room or via a Remote Desktop. In addition, TREs are often referred to as Secure Data Centres, Secure Processing Environments, or Data Safe Havens. The TREs typically utilise access controls that prevent researchers from removing data from them, as well as what additional data, software, and codes could be brought in. By reassuring data controllers and producers that data can be shared safely, TREs have become key in facilitating access to data that would not have been shared otherwise.

The TREs face the often-difficult balancing act between offering as much research freedom as possible and the increased potential for risks that comes from that freedom. Therefore, the secure-access community has developed strong security models and frameworks to ensure the safe, legal, and ethical use of sensitive data. One such framework is the Five Safes Model (Figure 1), which was developed by researchers in the UK in 2003 (Desai, Ritchie, & Welpton, 2016), with the ambition to provide an effective decision-making process around the safeguards that secure-access facilities could implement to achieve the safe use of personal data in a research context. The framework is a straightforward but powerful mechanism for thinking about how to build security models.

[1] The term sensitive data is used as a blanket term to describe data that requires special protections and therefore cannot be downloaded by researchers. Such data is made available via services, such as the GESIS Secure Data Center. Data may be deemed sensitive for a number of reasons, for example, very detailed social survey data could include increased risk of a data subject being reidentified; environmental data on bird habitats could lead to the identification of nesting sites for rare species; or business data could lead to commercially sensitive financial information being disclosed about a particular business.

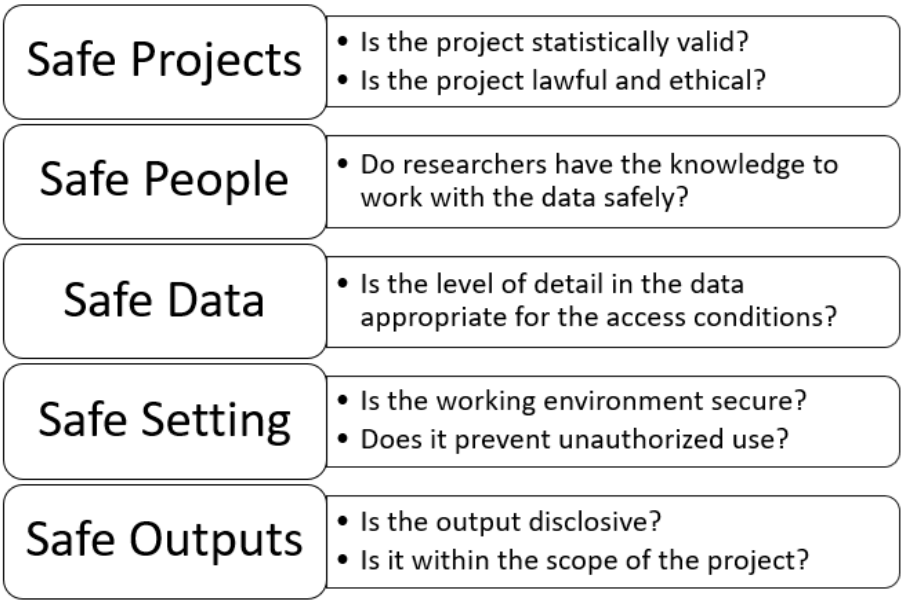

Figure 1. The Five Safes Framework

Perhaps, the most crucial component in the successful implementation of the Five Safes Model is 'Safe People'; they are researchers who have sufficient knowledge and understanding of data confidentiality and disclosure risk to work safely with sensitive or legally controlled data. People are simultaneously the strongest and the weakest link in the Five Safes Model. It is the element that is the hardest to control, and where arguably the greatest potential for unpredictability lies. In a human-based security model, accepting that people are human beings who can, and do, make mistakes is crucial. Conversely, researchers with the right attitude towards data protection and knowledge of their responsibilities will contribute positively to the implementation of the other Safes (Desai & Ritchie, 2010). Therefore, TREs need to consider how to ensure that researchers have the necessary knowledge and understanding of how to work with sensitive or disclosive data safely.

Many secure-access facilities require researchers who access data to undergo mandatory training, including some form of assessment, before they access data, to ensure they are Safe Individuals. Although the exact impact of training on key behaviours, such as submitting non-problematic outputs, has not been formally studied, Desai and Ritchie (2010) argue that training researchers leads to better cooperation and fewer mistakes by them. This is certainly supported anecdotally by TRE staff, who perceive that knowledgeable, well-trained researchers make fewer of the type of mistakes that could lead to harm to data subjects. In addition, it is acknowledged that these researchers require less support and cooperate well with support staff, an important factor in running an effective TRE service.

Based on this experience, it could be stated that the Safe People component of the Five Safes Model has proved to be highly successful. However, it could be argued that a vital part of the equation remains unresolved. It is not only researchers who are involved in achieving the Safe Use of research data; there are also professionals who work in secure-access facilities, responsible for handling and managing the data, as well as training and guiding researchers and the support teams.

Due to the relatively recent widespread adoption of TREs, there is currently a lack of agreed-upon principles regarding those professionals tasked with supporting their use. Staff working in secure-access facilities have diverse backgrounds, as there is no formal career pathway into these roles. Many, but not all, will have some experience in

performing research and statistical analyses; however, these experiences may differ significantly from those of the researchers they are supporting. In addition, there is rarely any formal training for individuals in these roles, a fact that is somewhat surprising, given the legal implications of potential mistakes. Training is usually 'on the job', and knowledge is gained through practice and experience gained over many years. This model is not automatically problematic. When a new secure data access professional joins an established TRE with a dedicated team, they benefit from working with experienced and knowledgeable colleagues. However, many TRE support teams are small, perhaps even a single person, and there may not be experienced and expert colleagues to guide the other professionals.

A lack of training could result in untrained staff and those new to the roles approaching their work with anxiety and low confidence in their ability to adequately make decisions. This lack of confidence could lead to staff being more cautious and restrictive than is necessary, resulting in a less research-friendly service. Confidence and skills are gained through experience, but developing confidence can take time, especially when formal training is not available. The recruitment of experienced TRE staff may be particularly challenging due to the rapid growth in the number of secure-access facilities that require their services.

Existing Training Models: The UK Safe Researcher Training

Therefore, it is clear that training for those using TREs is highly desirable and beneficial; there have been steps to implement a formal, standardized training scheme for researchers. In the UK, an Accredited Researcher scheme[2] was launched as part of the introduction of the Digital Economy Act 2017 (DEA)[3]. The DEA requires that researchers wishing to undertake analyses using sensitive data provided by the Office for National Statistics (ONS) receive training provided by one of a small number of approved services. Upon successful completion of the training, the researcher is granted Accredited Researcher status, which enables them to access ONS services. Following the ONS's lead, a number of other services, including those providing health data, now require researchers to obtain Accredited Researcher status when applying for data. The formalisation of training in this way has several benefits:

- Researchers have a better idea of how to work with sensitive data; therefore, they are less likely to make mistakes that might prove harmful to data subjects;
- Researchers do not need to undergo similar training courses repeatedly, because their trained and accredited status could be carried over to other services;
- It supports the Safe People aspect of the Five Safes Model by providing standardised training materials delivered by experienced services;
- The process of analysing sensitive data in a Research Data Centre (RDC) and publishing results from projects in an RDC will be more efficient for both sides: the researchers and the RDC staff, because they have more knowledge about disclosure and the procedures are known from the beginning of a project.

[2] Becoming an accredited researcher: https://www.ons.gov.uk/aboutus/whatwedo/statistics/requestingstatistics/secureresearchservice/becomeanaccreditedresearcher

[3] Digital Economy Act 2017: https://www.legislation.gov.uk/ukpga/2017/30/contents. Chapter 5, sections 64 and 71 are particular pertinent here.

Drivers for Proposing a Similar Model in Germany

Currently, there is a lack of consistency regarding training for researchers who wish to access sensitive data in Germany. Some researchers will be expected to undergo service-specific training; however, this tends to focus on the key rules of a particular service, delivered when they arrive at the Safe Setting, rather than on the general principles associated with being Safe People. Therefore, the contents and quality of training can vary dramatically, which does not enable researchers to access data at multiple secure-access facilities without retraining. In addition, there is no formal training for those working in the TREs who are tasked with supporting and advising researchers. Professional networks, such as the International Secure Data Facilities Professionals Network (Wiltshire, Lichtwardt, & Bishop, 2024), could help to provide support for TRE staff; however, they are not a substitute for high-quality, careful, targeted training. Therefore, a number of drivers exist for proposing a similar training model in Germany:

- Researchers often have insufficient knowledge and understanding of data confidentiality and disclosure risk to work safely with sensitive or legally controlled data;
- Information could be and is provided prior to the visit; however, experience shows that this is insufficient on its own for several reasons:
    - Researchers do not always read the information;
    - Researchers read them, but skim them and do not fully take on board the information or understand it;
    - Researchers may not fully see the importance of the information and the rules.
- Training for TRE staff is usually 'on the job' and knowledge gained through experience over many years, leading to inconsistent training across services;
- There are often difficulties with recruiting people with existing knowledge and experience into these roles and with staff retention. With no clear career path, these roles can be seen as 'dead-end' positions with limited opportunities for career progression;
- Germany, unlike a lot of other countries, has an RDC Infrastructure with over 40 different centres, many making sensitive data available via a TRE. Currently, no federated, targeted training schema exists.

ASSURED: Building a 'Safe Researcher' Training Programme for the German TRE Community

In response to this need in Germany and building on the experience of implementing the Safe Researcher Training (SRT) in the UK, the ASSURED project has been developed. ASSURED will offer an e-learning training programme and widely recognised accreditation for those wishing to become Safe People in Germany, whether as researchers or data access professionals. Through this, the aim is to give the public confidence that their sensitive data is being used responsibly in a research context. Utilising a modular approach to training

The ASSURED training programme is designed to be a suite of self-study e-learning modules, available via Moodle. Each module will take around 10 minutes to complete and includes different activities with an assessment at completion to test the trainees' understanding. This design allows more flexibility for the TREs and the researchers; for TREs, delivering such training in-person or online is resource-intensive and impractical. Having self-learning training that allows trainees to learn independently and that automates some of the administrative processes means that implementing a training requirement is not unattainable, even for the smallest TREs. In addition, for the trainee, this approach offers the flexibility to fit the training more easily around their schedules. Trainees are not bound to a specific course date; rather, they could work through the modules one at a time when it is convenient.

ASSURED consists of a few core modules that are deemed to be mandatory for all, regardless of the TRE and their role. These modules would then be augmented by additional modules that are service, datatype, or role-specific, allowing TREs to specify a training pathway that is tailored to their needs. For example, a researcher accessing digital behavioural data via the Secure Data Centre (the TRE at GESIS Leibniz Institute for the Social Sciences[4]) would be required to complete additional modules to the core modules that relate to digital behavioural data and how to use the Secure Data Centre effectively. Additional role-specific modules would be available for new recruits joining the TRE team. These modules could help existing team members develop their skills to better support their career development.

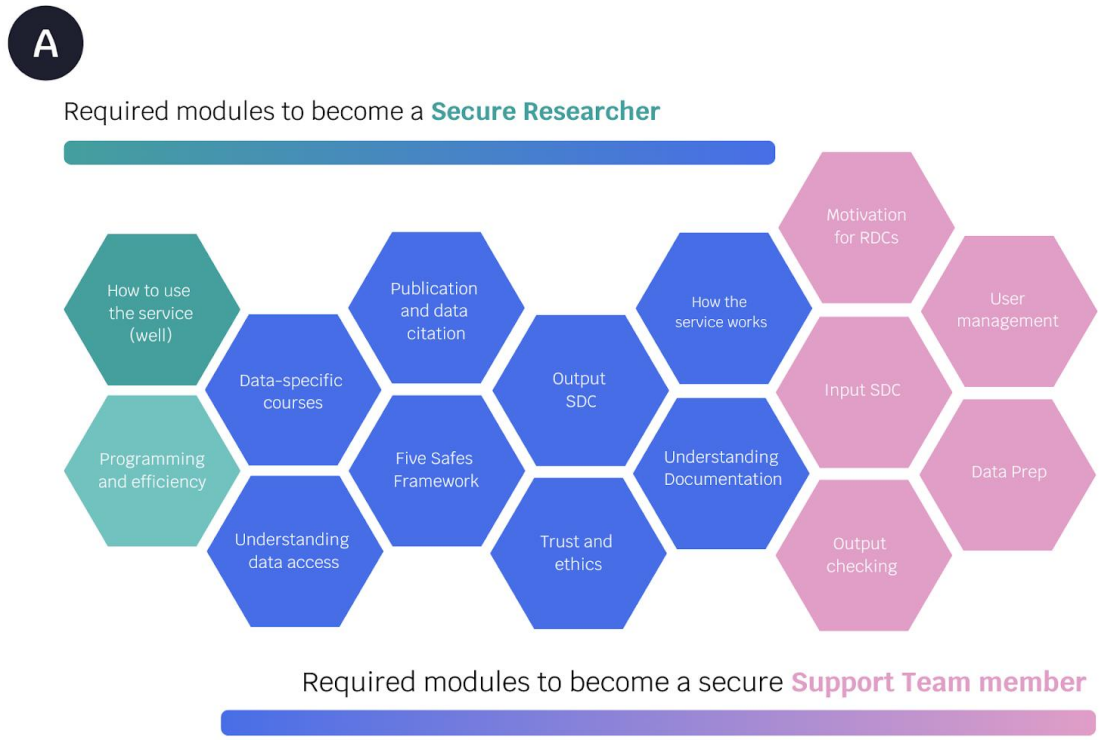

Figure 2. Example modules and programme structure for researchers and TRE teams.

[4] Secure Data Center: https://www.gesis.org/en/services/processing-and-analyzing-data/analysis-of-sensitive-data/secure-data-center-sdc

Figure 2 shows the core modules for both target audiences. These core modules represent the minimum standard expected of researchers and TRE staff working with sensitive data. Completing these modules would classify the researcher as an 'Approved Researcher' who has the basic knowledge and understanding to work with sensitive data. This would reduce the need for researchers to attend the same or similar training offered by different TREs.

Assessment and Validation of Module Completion

For each module, trainees must complete a short assessment quiz to pass that module. To maintain an accurate record of the successful completion of modules, the training status of researchers would be integrated with an Authentication and Authorization Infrastructure (AAI), such as the Life Science Login, which would be directly connected to the training platform.

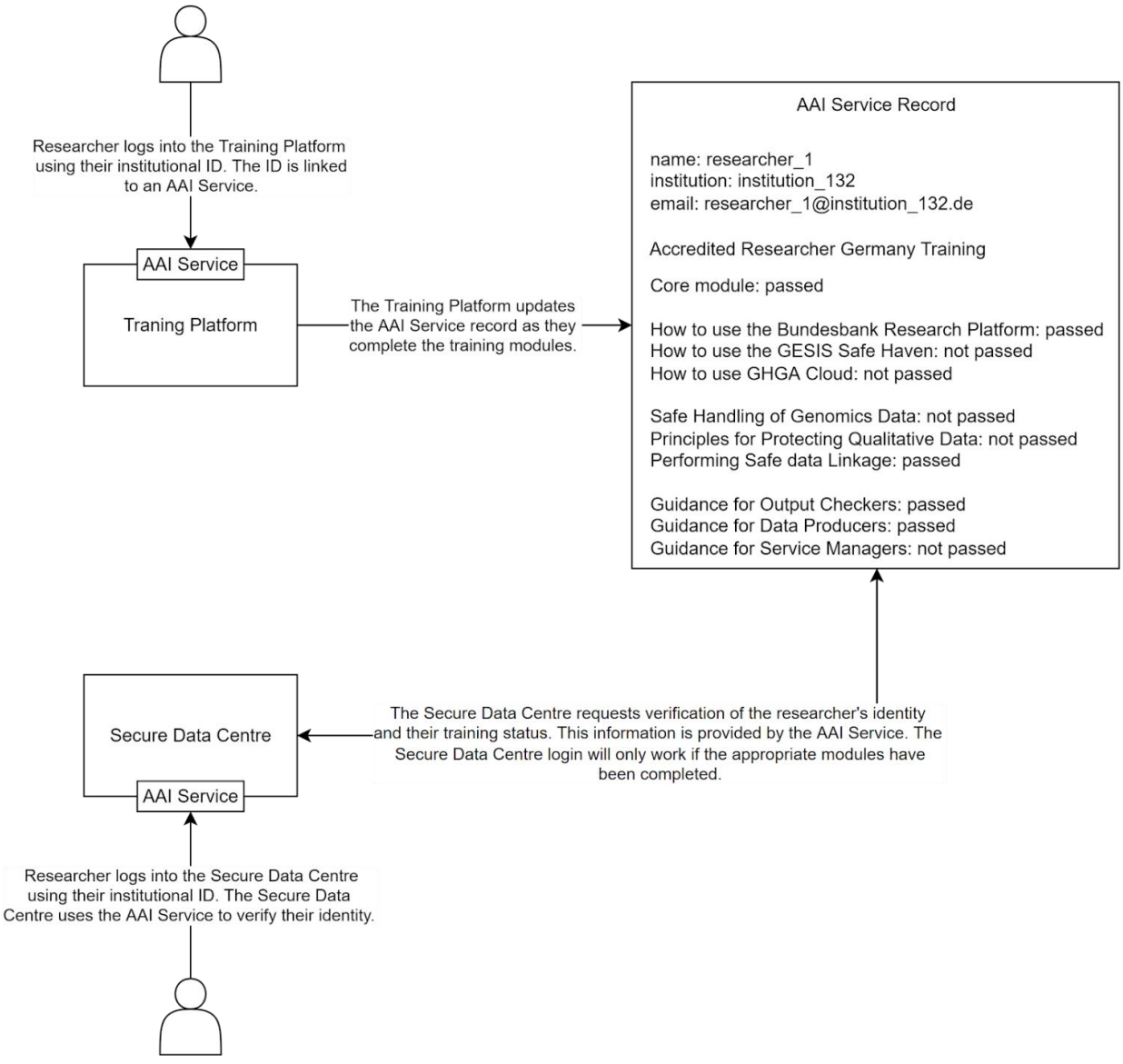

Figure 3. The Proposal for an ASSURED AAI.

Figure 3 shows how such an AAI could work. As an example of how this would work in practice, a researcher applying to access data through TRE A would be invited to complete the core modules, pre-selected by the TRE. They would then log into the ASSURED training platform and complete those modules. The successful completion of those modules would then be recorded in the researcher's training record in the AAI system. Completion of all the pre-selected modules would earn researchers the title of 'accredited researcher' and would enable them to access the data via the TRE. This setup aims to improve on the UK SRT model, where training is currently delivered by TRE staff who then keep a manual record of the training, often in an Excel spreadsheet, meaning there is no central training record, a much more labour-intensive model.

Linking the training platform to an AAI will also enable researchers to take the evidence of their training to other data providers. For example, if the researcher then wished to apply to access data via TRE B, staff there would be able to check their training status via the AAI without the need to contact TRE A to request confirmation of training. In this way, TREs will be able to easily ascertain if someone requesting access to their services has undergone the required training.

Additional improvements could include integrating a 'date of training completion' field into the platform so that regular retraining can be mandated, and automated reminders issued to trainees. Through the use of AAI, ASSURED aims to make the process of training and management of training records much less labour-intensive, so that even smaller TREs and data services with limited resources can benefit from the adoption of high-quality training for their users and staff.

Launching the ASSURED Project

In early 2024, the ASSURED website went live. The website outlines the project's aims and vision, and includes details of the training. In September 2024, with an example module created in Moodle, we held a virtual workshop with colleagues from key TREs across Germany to launch the project. During the workshop, the project's aims were presented, the module was demonstrated, and feedback was gathered from the audience. The feedback was overwhelmingly positive, with all agreeing that the training should meet their needs. Suggestions on how to adequately assess trainees' understanding of the module content were extremely helpful and will inform the further development of modules.

Looking Ahead

Through seed funding, an external company has been engaged to help set up the Moodle platform, which will incorporate the ASSURED corporate design and establish the necessary functionality for the training program. The content for the core modules is complete. Therefore, once the platform set up is complete, these modules will be created in the platform. Then, the next step will be to conduct user testing for user experience and technical or content errors. Many attendees from the launch workshop have volunteered to test the modules, and additional testers will come from the UK TRE community, which has direct experience in delivering SRT-type training. A round of edits will follow based on the feedback, before work moves on to the development of the additional specialist modules.

A funding proposal is currently under review, which, if successful, would fund a full-time position to oversee the further development and daily operation of the ASSURED training for the next 3 years. In addition, this would open the possibility to expand to other target groups and sectors, such as industry and the public sector. As part of this funded post, research will be carried out to try to measure the impact of implementing training, as

this has not previously been performed in a formal way, using the Secure Data Centre at GESIS as our use case.

This programme would initially focus on researchers in Germany; however, it has the potential to be expanded across Europe in alignment with the objectives of the European Open Science Cloud (ESOC) project[5] and, particularly, with the EOSC-ENTRUST[6] project, which aims to create a European network of TREs via the development of a common blueprint for services providing access to data which requires additional protections because of potential concerns around sensitivity, disclosure risk or commercial sensitivity for example. With such projects moving towards networks and federated data access, the flexibility of the ASSURED training means that it could be ideally suited to providing standardised training across scientific disciplines and international borders.

Acknowledgements

The ASSURED team would like to acknowledge the German Human Genome-Phenome Archive and BERD@NFDI for their early financial contributions to the project. We would also like to acknowledge Dr. Wiebke Weber and Markus Herklotz from Ludwig Maximilian University of Munich for their partnership in ASSURED, as well as Professor Felix Ritchie from the University of the West of England and Stefan Bender at the Deutsche Bundesbank for their input and guidance in the initial development of the ASSURED concept.

References

Bishop, L., Broeder, D., van den Heuvel, H., Kleiner, B., Lichtwardt, B., Wiltshire, D., & Voronin, Y. (2022). *D5.10 White paper on remote access to sensitive data in the Social Sciences and Humanities: 2021 and beyond*. Retrieved from https://doi.org/10.5281/zenodo.6719121

Desai, T., & Ritchie, F. (2009, December 2-4). Effective researcher management [Paper presentation]. W*ork session on statistical data confidentiality*. 2009 Joint UNECE/Eurostat work session on statistical data confidentiality, Bilbao, Spain. https://unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.46/2009/wp.15.e.pdf

Desai T., Ritchie F., & Welpton R. (2016). The Five Safes: designing data access for research. *Working papers in Economics,* 1607. Retrieved from https://www.researchgate.net/publication/306579162_Can_a_change_in_attitudes_improve_effective_access_to_administrative_data_for_research [accessed May 08 2025].

Wiltshire, D., Lichtwardt, B., & Bishop, L. (2024). Building human networks to drive forward innovations in international data access: Introducing the International Secure Data Facility Professionals Network (ISDFPN). *IASSIST Quarterly*, *48*(3). Retrieved from https://doi.org/10.29173/iq1097

[5] The Open Science Cloud: https://open-science-cloud.ec.europa.eu/

[6] EOSC-ENTRUST: https://eosc-entrust.eu/