

Research and Innovation Action

Social Sciences & Humanities Open Cloud

Project Number: 823782

Start Date of Project: 01/01/2019

Duration: 40 months

Report on **Milestone 29** Tested connections between partners with live data and researcher projects

Dissemination Level	PU
Due Date of Milestone	15/04/22 (M29)
Actual Achievement Date	05/04/22
Lead Beneficiary/LTP	CESSDA ERIC / GESIS
Work Package	WP5 Innovations in Data Access
Task	Task 5.4 Remote Access to Sensitive Data
Version	V 1.0
Number of Pages	p.1 – p.19

Abstract:

This Milestone assesses the testing of the remote Safe Room connections between the Secure Data Facilities of the GESIS Secure Data Center and the UKDS Secure Lab. These tests were carried out by bona fide researchers who had submitted full research applications to the respective data service and involved the analysis of live data. This report includes a summary of the researcher applications and tests as well as the researcher feedback collected following the test, which will inform further improvements to the service for researchers.

The information in this document reflects only the author's views and the European Community is not liable for any use that may be made of the information contained therein. The information in this document is provided "as is" without guarantee or warranty of any kind, express or implied, including but not limited to the fitness of the information for a particular purpose. The user thereof uses the information at his/ her sole risk and liability. This deliverable is licensed under a Creative Commons Attribution 4.0 International License.





SSHOC, "Social Sciences and Humanities Open Cloud", has received funding from the EU Horizon 2020 Research and Innovation Programme (2014-2020); H2020-INFRAEOSC-04-2018, under the agreement No. 823782





Author List

Organisation	Name	Contact Information
GESIS	Deborah Wiltshire	Deborah.Wiltshire@gesis.org
GESIS	Yevhen Voronin	Yevhen.Voronin@gesis.org
UKDS	Beate Lichtwardt	blicht@essex.ac.uk

History

Version	Date	Reason	Revised by
0.1	21/03/2022	First draft	Deborah Wiltshire
0.2	22/03/2021	First review, editing and formatting corrections	Yevhen Voronin
1.0	04/04/2022	Outcome of researcher tests added, final editing	Deborah Wiltshire, Yevhen Voronin, Beate Lichtwardt

1. Introduction

This Milestone sits within WP5.4 and delivers a bilateral agreement between the Secure Data Facilities at GESIS and the UKDS SecureLab. The primary task is to set up a secure remote connection between the two Secure Data Facilities to allow safe access to controlled/ secure data across international borders.

This remote connection has been established and a number of technical tests carried out to confirm that it functions as it should.



A detailed report of the setting up and technical testing of the remote connections between GESIS Secure Data Center and the UKDS SecureLab is provided in the SSHOC Report D5.11 "ERAN Pilot: Setting up a Secure Remote Connection between two Trusted Research Environments"¹.

Milestone 29 focuses on the next stage, which is to demonstrate fully functional, tested connections between the partner organisations with live data and researcher projects.

2. Description of the Milestone

2.1. Role of the Milestone

Following the successful setting up of the remote connections between the Secure Data Facilities of the GESIS Secure Data Center and the UKDS SecureLab, the next stage of this deliverable is researcher tests whereby bone fide researchers in Germany and the UK apply to access secure data held in the opposite organisations' Secure Data Facility and carries out data analysis for a live research project, this milestone demonstrates the successful completion of these tests with live data and research projects.

2.2. Means of verification

This Milestone can be verified by the research project proposals submitted by two independent social science researchers, and by the feedback provided by both researchers, which is made available as part of this report. Further, verification will be available in the future via the production of research outputs from their analysis. The proposals and details of the research undertaken are discussed in section 3.1.1. and 3.1.2. Some preliminary research has been carried out, and initial research outputs are awaiting Statistical Disclosure Control (SDC) checks. Once these research outputs have been cleared, they can be made available as verification of the Milestone. However, this would not be possible within the timeframe of the SSHOC Project which ends at the End of April 2022.

2.3. Explanation on delay in achieving the Milestone

There have been external factors at play which are unique to the current period and wholly unforeseen at the start of the SSHOC project. The event of a global pandemic with subsequent restrictions and

¹ Beate Lichtwardt, Matthew Woollard, Deborah Wiltshire, & Elizabeth Lea Bishop. (2022). D5.11 ERAN Pilot: Setting up a Secure Remote Connection between two Trusted Research Environments (v1.0) (forthcoming)



lockdowns has meant that both the Safe Rooms at GESIS and the UK Data Service closed at the start of the COVID-19 Pandemic in March 2020. These restrictions prohibited staff from working on-site and also prohibited visits from external researchers.

Both Safe Rooms have remained closed throughout the pandemic, and were only in early Spring 2022, permitted to allow staff and external visitors on-site for the technical and researcher tests. Further delays were caused as earlier test dates had to be cancelled due to covid infections of key people involved in the setting up and execution of the tests.

3. Researcher tests

The first stage of the researcher testing was to work with Germany and UK based researchers to carry out two short pieces of research for the purpose of testing the bi-lateral remote connection.

3.1. The application process

The following sections outline the application process required for both researchers.

3.1.1 Test of the UKDS to GESIS connection:

The test researcher was Silvia Avram, a senior researcher based at the Institute of Social and Economic Research (ISER) at the University of Essex. Silvia has many years of experience carrying out research using secure access data via the UKDS SecureLab.

Silvia submitted a research application to GESIS to access data from the German General Social Survey (ALLBUS) study. The research proposed was as follows:

Title of research project	User testing GESIS Access Point at UKDS/UKDA Safe Room - Wage growth and wage distribution
Data Requested	German General Social Survey (ALLBUS) Cumulation 1980-2018; ALLBUS/ GGSS - Sensitive Regional Data 1980-2018
Description of the research	Exploratory Analysis of the "German General Social Survey (ALLBUS) Cumulation 1980-2018: ALLBUS/GGSS - Sensitive Regional Data 1980- 2018* to establish feasibility of future comparative research regarding wage growths and wage distribution



Description of methodological approach (Research methods)	I will look at wage growth at different points of the hourly earnings distribution and, consequently, the wage inequality. Examined will be the wage growth at the 5th, 10th. 15th, 20th, 30th and 50th percentile of the local wage distribution.
	This project will also form part of the User Experience Testing of the new remote access connection between the UKDS and GESIS, which has been developed as part of the Social Science & Humanities Open Cloud (SSHOC) project. This new connection is an important part of the move to open up access to sensitive data internationally

The completed research application was assessed and approved by the Research Data Center responsible for the ALLBUS study.

To complete the process, Silvia signed a <u>GESIS Data Use Agreement</u> contract which details the terms and conditions of access.

3.1.2. Test of the GESIS to UKDS connection:

The test researcher was Simon Parker, a Data Manager based at Deutsches Krebsforschungszentrum (DKFZ) – German Cancer Research Centre in Heidelberg. Simon has many years of experience carrying out research using secure access data via the UKDS SecureLab and the Secure Research Service at the Office for National Statistics, and in operating Secure Data Facilities both in the UK and Germany.

Simon submitted a research application to the UK Data Service to access data from three UK birth cohort studies. The research proposed was as follows:

Title of research project	The Geographic Propinquity of Vaccine Hesitancy
Data Requested	SN 8658 COVID-19 Survey in Five National Longitudinal Cohort Studies: Millennium Cohort Study, Next Steps, 1970 British Cohort Study and 1958 National Child Development Study, 2020-2021 SN 8658 1970 British Cohort Study: Age 46, Sweep 10 Geographical Identifiers, 2016-2018: Secure Access SN 8085 National Child Development Study: Sweeps 3-9, 1974- 2013, Townsend Index (LSOA) Linked Data: Secure Access



	SN 8757 Millennium Cohort Study: Geographical Identifiers, Seventh Survey: Secure Access
Description of the research	Vaccine hesitancy has gained prominence as a key factor in lower vaccination rates during the recent pandemic. Lower vaccination rates have been linked with higher demand on the health services, and studies suggest that among who require hospital care for COVID infections, a higher proportion are unvaccinated. Thus, vaccine hesitancy is a vital area of research for health policy.
	Many hypotheses exist for why people may be reluctant to receive a vaccine, such as the influence of social media and easy access to anti- vaccination discussions, conspiracy theories and misinformation. It might also be a cohort effect, so those who are aware of the controversies around previous vaccinations such as the MMR might be more reluctant to vaccinate.
	We would like to utilise the three large birth cohort studies in the UK to see if there is evidence of age or cohort effects. But we would like to extend this by investigating whether there are regional variations in vaccine hesitancy and in the reasons given for non-vaccination.
	Regional differences may suggest that other factors are at play, such as proximity to large scale protests or closer contact with COVID restrictions experienced in larger urban areas. For this we require small area geographical variables.
	The second aim of the project is to explore international data access pathways. This project will be an important test case for a new remote access connection between the UK Data Service and GESIS Institute for the Social Sciences in Germany. This connection has been developed as part of the Social Science and Humanities Open Cloud (SSHOC) project, a large EU funded project designed to open up data access across international borders. This project will mark the final stage of making UK cohort data available to researchers in Germany.

In addition, to access controlled UKDS SecureLab data, researchers have to be an accredited researcher, which entails undergoing mandatory training. Simon already had accredited researcher status, having previously used data through the UKDS SecureLab and the Office for National Statistics SRS in the UK. He was also required to submit a <u>researcher form</u>, demonstrating his expertise and prior quantitative research experience. The completed research application was assessed and approved by the Data Access Committee at the Centre for Longitudinal Studies (CLS) who are responsible for all three birth cohort studies.



To complete the process, Simon submitted a signed <u>Secure Access User Agreement</u>, which was countersigned by their institution.

3.2. The researcher tests

Prior to the tests

Prior to the researcher tests, a number of joint technical tests were carried out to test for potential technical problems with the setup of the Thin Clients and the remote connections. All tests were successful and are described in more detail in the deliverable report D5.11 "ERAN Pilot: Setting up a Secure Remote Connection between two Trusted Research Environments". These tests also functioned to give the respective IT teams the opportunity to get familiar with the remote connections so that they are better able to respond to any technical issues that might arise once the connections are live.

The full test schedule is outlined here:

Test	Date carried out
Technical test 1	25 January 2022
Technical test 2	2 February 2022
Researcher tests	1 April 2022

The researcher tests were both scheduled for 1st April 2022. Prior to the tests, both organisations prepared the project areas, ensuring everything was in place for the commencement of the research projects. Also prior to the test, GESIS sent across a few useful files to the UK researcher, including guidance on logging into and working in the GESIS virtual environment, a syntax file to assist with creating outputs and a guide on how to compile outputs for statistical disclosure control. It is intended that all researchers will receive such guidance documents prior to accessing the virtual environment.



On the day of tests

On the day of the tests, staff responsible for the running of the Safe Rooms were present to greet and support the external researchers. A representative from both organisations' IT teams were also present onsite on the day of the tests to ensure that any technical issues could be resolved quickly. The teams met briefly via Microsoft Teams just to exchange any last minute information, such as offline contact details in case of emergencies. The teams remained in the MS Teams meeting until Silvia had successfully logged into the GESIS virtual environment via the UKDS Safe Room. Simon arrived later, and liaising with the UKDS IT team during the test was conducted via telephone.

On the whole both tests went extremely well. Both researchers were able to access their projects in the virtual environment, were able to carry out some initial exploratory analyses, produce a piece of research output and request an output request, which are now subject to Statistical Disclosure Control checks.

There were some small issues that cropped up during the tests, some procedural and some technical, but nothing that impeded the researchers' analysis. Having the support and IT teams from both organisations available for the entire period of the tests meant that all issues were quickly resolved.

3.3. Researcher feedback

The researchers were asked to write a short feedback summarising their experience of accessing data via the remote connections. Prior to the researcher tests, both researchers were provided with a list of areas to think about whilst accessing their project areas and carrying out their analyses to aid them with their feedback task. The list of questions can be reviewed in Appendix 2.

The researcher feedback focuses on 4 main areas: the application/pre-test experience; the physical environment; the virtual environment; and the overall experience. The full feedback reports from both researchers can be viewed² although a summary is discussed here.

The application/pre-test experience:

Both researchers were successfully able to submit a research application. Silvia found the process of applying to GESIS to be smooth and without any difficulties, although she acknowledges that she had

² Full feedback reports from both researchers can be found in Appendix 3 of this report.



assistance from UKDS staff which would not normally be available. Silvia did not request any imports to her account prior to the researcher testing.

In applying to the UKDS, Simon experienced some difficulties. There was some confusion over his accredited researcher status, which was resolved by the UKDS but it did not seem easy to verify this with the ONS. There was also an issue with his import requests and with adding and assigning End User Licence data to his online account. However, the issues were resolved and the data were available on the day of the test.

The physical environment:

Both researchers found the physical Safe Room environment that they accessed to be pleasant, and to have all the necessary equipment for their visit.

In the GESIS Safe Room, Simon found a discrepancy between the keyboard and the setup of the virtual environment. The keyboard in the Safe Room is an ANSI keyboard, whilst the system is set up for ISO keyboards. This results in the λ key being unavailable, unless the researcher manually changes the language settings each time.

The virtual environment:

For the GESIS virtual environment, Silvia found the instructions on logging in and using the virtual environment clear and very helpful. She advised that these could be made available permanently in the UKDS Safe Room for future researchers.

Login credentials were transferred via an encrypted Veracrypt container, which Silvia felt might be difficult for researchers not familiar with Veracrypt. She also advised that the password for the container should be provided earlier.

Her final recommendations are that guidance could be provided on how to request an output and how to request an import of syntax. She noted that the turnaround time for an output to be released to the researcher is very long.

For the UKDS SecureLab virtual environment, Simon had the following comments: The initial log in went smoothly, but there were some issues with locking the screen or logging off and back in. On such occasions, like after lunch, Simon had to restart the system each time in order to get back to the correct log in screen. As the virtual environment is persistent, this did not cause any loss of work. There was also



a pop-up window that should not have been there, but the IT team at the UKDS were on standby and resolved this very quickly.

Although Simon had requested the import of some files, not all were available at the start of the test. However, the Secure Lab Support Team was on standby and this was resolved very quickly.

Overall experience:

Both researchers considered their experiences to be positive overall, providing the following key feedback.

Silvia acknowledged that she is very familiar with working with secure data, but she felt that the process of accessing GESIS data via the UKDS Safe Room should also be easily navigated by less experienced researchers.

Simon also found the experience to be positive overall, and noted that he was impressed with the technical improvements made to the UKDS SecureLab since he last accessed it. He raised one question for the teams to consider before the remote connection goes 'live' - usually when a researcher uses UKDS SecureLab data via the GESIS Safe Room for the first time, they have to undergo a quick, in-person training session. This did not occur during the test, and was not required as Simon is familiar with the environment, but he recommends that how and when such training will be delivered is considered going forward.

4. Conclusions and next steps

The purpose of this stage of the Assessment has been to test whether the new remote Safe Room access between the Secure Data Facilities of the GESIS Secure Data Center and the UKDS SecureLab functions and to assess whether the researcher user experience is satisfactory. The minor issues experienced by Simon in the UKDS SecureLab virtual environment have been fed back to the IT team at the UKDS and have already been resolved.

The researcher user experience tests carried out by the test researchers were largely successful with the feedback from both researchers overwhelmingly positive. Issues experienced during the tests were fairly minor, and easily resolved by the respective teams. A key recommendation from this test for parties looking to set up and test remote connections, is therefore to ensure that support and IT teams from both organisations are available throughout the tests.



Some points to consider were raised by both researchers, some related to the remote connection, others to the service in general. These will be taken forward and addressed by both services to help improve the experience for future researcher visits. One key recommendation would be to ensure that both organisations provide guides to key processes, such as output and import procedures, for researchers prior to access. GESIS provided such guides prior to the test, and it was noted that these were very helpful.

Currently the GESIS and UKDS Safe Rooms remain closed due to COVID restrictions. These restrictions have meant that staff have not been permitted to work routinely on-site, and external visitors have not been permitted. These restrictions are now being eased, and confirmation is expected soon that both Safe Rooms will be permitted to open for researcher visits.

The next step is to finalise the administrative processes that are required to ensure the successful running of the remote connections. Visits cannot be scheduled until researchers have completed the application process, so cooperation and clear processes between the services is required to allow data services to check that this is indeed the case, and that researchers are not booking visits prior to the successful setup of their project areas. A meeting is scheduled for this in early May 2022.

Once the remote connections are 'live' (i.e. ready to accept bookings for researcher visits), the respective services websites will be updated to advertise the new access options to the research community.



5. Appendix 1

Application forms and User Agreements:

GESIS, Secure Data Center. Data Use Agreement: https://www.gesis.org/fileadmin/upload/dienstleistung/daten/secure_data_center/GESIS_Data_Use_Agr eement_SDC_On-Site.pdf.

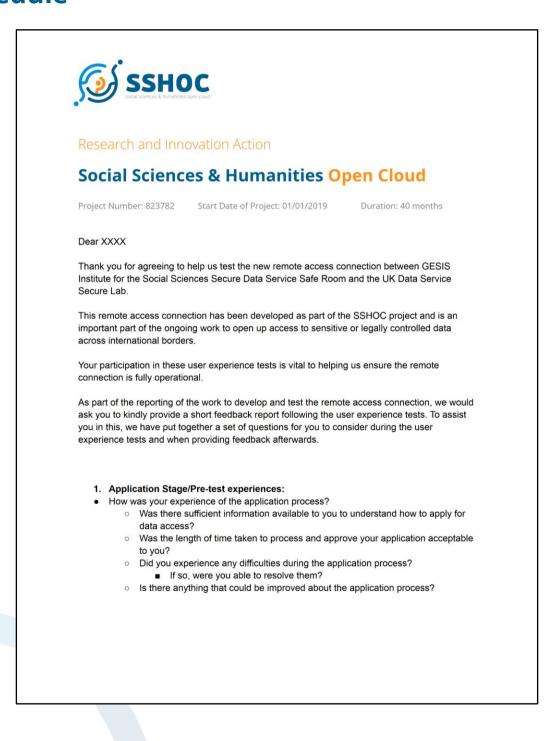
UKDS. Secure Access User Agreement (IDAN): <u>https://drive.google.com/file/d/1UuKk2zxiJH85vTibb2mvKvn935v_wRpZ/view</u>.

UKDS. ESRC accredited researcher and researcher proposal forms:

<u>https://drive.google.com/file/d/1Intlh4NfPUJfJRtiQTdGnapMdWwxQuib/view?usp=sharing</u>
<u>https://drive.google.com/file/d/1BNcgTFM4xaj4GPLZfnYaY8fcabZNV_Md/view?usp=sharing</u>

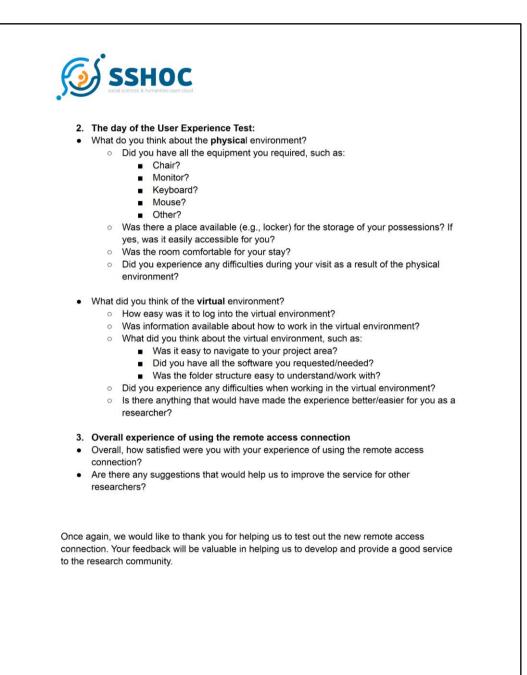


6. Appendix 2: Researcher Feedback questions schedule











7. Appendix 3: Researcher Feedback Reports

Feedback Report from Silvia Avram, access to GESIS data from the UKDS Safe Room





should come prepared with their login details. I have used VeraCrypt before but perhaps for the user who has not, maybe it would be useful to have some instructions on how to retrieve the login details.

Instructions for logging in and out of the virtual environment: These were very clear and everything functioned as expected. I think it would be useful though to have these instructions in the Safe Room. The process is quite intuitive, but I think that especially before a researcher becomes familiarized with the procedure, it would be good to have these instructions in physical format for reference. Same for logging our instructions (although in this case, the instructions can be accessed electronically).

The instructions on the structure of the project area were clear and intuitive. I also found the provided do file very helpful! Both the first phase that generates logs, copies files and especially

The last part that generates the output is very useful. I also found the instructions on putting together the output very good and clear. Although perhaps it could be made clearer that one can submit a log file (this is normally not allowed by the UKDS). I was wondering also whether something else than creating some filed in the Output area is needed to request an output.

Finally, is there a limit on the number of outputs that can be requested from the same project? The suggested three weeks needed for the control seem quite long and indicate that the expectation is output will only be requested once.

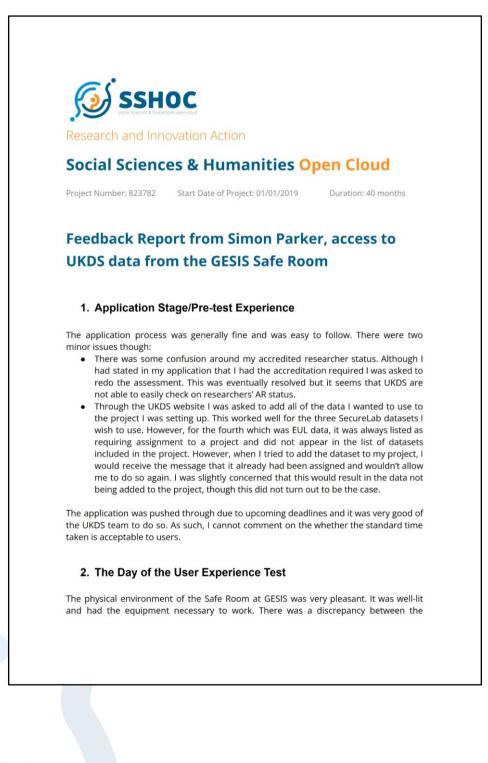
I have only tested Stata in the work environment. This seemed to work very smoothly. One question I have is if I wanted a user written command that is not already installed, how could I have that? Also, suppose that I have my own syntax files that I would like to import in the working environment, would that be possible?

4. Overall Experience

Overall, the experience was very good. I think the only hiccups were around getting login details and logging in but since this is a new procedure, it is to be expected. I should say though that I have worked with restricted data quite a bit before and so perhaps I'm already familiar with some of the procedures. That said, I think it should be fairly easy for a less experienced researcher to get used to the procedures.



Feedback Report from Simon Parker, access to UKDS data from the GESIS Safe Room





supplied keyboard and the keyboard language that the system was using that caused a minor issue; there was no keyboard key that could produce a backslash (\). This is because the key denoted as such on the keyboard was mapped to a different key by the system settings, and the key that would normally be the backslash on a English UK keyboard was absent. It was therefore necessary to look up the keyboard shortcut to produce a backslash (Alt+92) in order to work in R. This may cause confusion for some users.

Lockers for the storage of items were available.

3. Using the Virtual Environment

Overall, I found the virtual environment easy to use.

Logging in was more difficult than it needed to be. It would appear that when the thin client is started, users are automatically logged into UKDS Citrix launcher from which they can enter their username and password and enter the SecureLab. This worked well and the environment booted quickly. However, when left for a little while, it would seem that the thin client logs out of the account that loads the launcher. Users are presented with a standard Windows lock screen, requesting the password for the account 'User'. This doesn't appear to correspond with the Citrix launcher and entering my SecureLab password did not allow me to proceed. It was completely unclear how to get beyond this lock screen. In the end, I had to reboot the thin client so that the automatic login would work so that I could get to the Citrix launcher screen. This also happened after locking my session to go for lunch, though fortunately the session is persistent, so nothing was lost when I rebooted the thin client.

Once logged in I also experienced an odd pop-up when I opened Windows Explorer for the first time. I was asked if I wanted to have permission to access the local files on the thin client with three options available to me. Regardless of the option chosen I was able to see list a C drive and a Z drive alongside the standard M and R drives. If I selected the option to have read and write access, I was able to browse the system files of the thin client and presumably edit them. Whilst this didn't impact the usability of the environment in any way, it is perhaps something of a security and potential stability problem if users are able to access and edit these files.

There were a few small issues with the project though these were most likely the result of the haste with which the project had been set-up. Some of the R packages I required were not available despite having requested them, a script that was needed had not been uploaded, and the data were in Stata and SPSS form although we had applied for the tab-delimited versions. The team were able to quickly resolve the first two issues, and fortunately I was able to use R to work with the data without requiring them to be in tab-delimited format.



