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 28**

Assessment of Existing Platforms

Dissemination Level	PU
Due Date of Milestone	31/08/21 (M32)
Actual Achievement Date	31/08/21
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.7

Abstract:

This Milestone assesses selected, currently operational systems delivering Remote Desktop Access to sensitive data for social sciences and humanities. Prior to selection, criteria were established in three domains: technical, legal, and organizational/administrative. A significant consistency was found across the systems, and this is grounds for cautious optimism that "scaling up" from these systems may be possible. However, several factors contribute to the need to accommodate a solution with substantial flexibility. Recommendations will be provided in SSHOC D5.10 White Paper on Remote Access to Sensitive Data in the Social Sciences and Humanities: 2021 and beyond.

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.

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	Elizabeth Bishop	ElizabethLea.Bishop@gesis.org

1. Introduction

It is now widely recognized that the ideal of "open data" needs to be balanced with the privacy protection of data subjects and other factors that can require moderating access to sensitive data, as reflected in the European Commission's (2016) stance of "as open as possible, as closed as necessary." Developments in the past five years have advanced data access, primarily through connected "safe enclaves" (i.e., secure rooms in which researchers access data). This represents a major improvement for data accessibility but is still limited to a specific physical location. International, comparative, efficient research requires enabling remote access to data from researchers' desktop computers. Established solutions exist (e.g., Leibniz Institute for Educational Trajectories (LifBi), UK Data Archive SecureLab, Inter-university Consortium for Political and Social Research Virtual Data Enclave (ICPSR VDE)), but often these face limitations such as the scope of data that can be made available. More recently, new infrastructures have been created, some spanning several countries (e.g., Nordic Microdata Access Network or NordMAN). These efforts are commendable and represent major improvements. However, limited resources and complex legal variations (national implementations of the General Data Protection Regulation), as well as other factors, have prevented implementation of a broader solution. As countries across Europe look to the Nordic model, it is crucial to address the need for a sustainable European solution.

SSHOC Task 5.4 Remote Access to Sensitive Data will deliver several components to support this goal, including a White Paper with recommendations for such a European solution. These recommendations will be informed by a rigorous assessment of existing secure remote access systems provided in Milestone 28 (MS28) Assessment of Existing Platforms.

2. Description of the Milestone

2.1. Role of the Milestone

Deliverable D5.10 White Paper on Remote Access to Sensitive Data in the Social Sciences and Humanities: 2021 and beyond will review the current landscape and offer recommendations for how EOSC should consider moving forward to enable Secure Remote Access more widely. An essential prerequisite for recommendations is to assess selected existing systems, and this task is accomplished in this MS 28 Assessment of Existing Platforms.

This Milestone assessed selected, currently operational services delivering Remote Desktop Access to sensitive data for social sciences and humanities. Remote Desktop Access means that the data remain in a fixed location, and the user, working in another location, remotely accesses the data via a secure internet connection. To qualify for assessment, services had to meet certain criteria:

- Enable remote desktop access to sensitive data.
- Provide social science or humanities data (or be explicitly extendable to such data).
- Offer essential services needed for data access, i.e., not be only a technology platform.

Thirteen platforms were selected for review. Due to the European nature of SSHOC, primarily European services were chosen.

To conduct the assessment, criteria were established and agreed in three domains: technical, legal, and organizational/administrative. Assessment criteria are not standardized. However, multiple sources were consulted (European Commission, Eurostat, 2021; Rat Für Sozial- Und Wirtschaftsdaten (RatSWD), 2019; Schiller et. al., 2017) to produce the final categories: technical, legal, and organizational/administrative. Further details about the assessment criteria can be found in Appendix 1. Abbreviated versions of some of the criteria are used in the spreadsheet for readability.

2.2. Means of verification

This Milestone can be verified by consulting the Task leader. An internal spreadsheet stored in the SSHOC Document Repository containing the assessment was created and is used for follow-up work.

2.3. Explanation on delay in achieving the Milestone

Change in the delivery date of this milestone that was set in the Grant Agreement was approved within the SSHOC 2nd Amendment.

3. Conclusions and next steps

The purpose of this stage of the Assessment has been to determine if there is an emerging pattern for the features and requirements needed to safely and securely operate a Remote Desktop Access system, and this does seem to be the current situation. For example, systems provide similar user account management capabilities, encryption standards, and legal compliance. Nonetheless, there are also areas where variations exist, e.g., choice of IT platforms, requirements for researcher training, use of automation in disclosure checking, and support for research groups (not only individual researchers).

The consistency found is grounds for cautious optimism that "scaling up" from these largely institutional, or dataset-specific, systems may be possible. However, several factors contribute to the need to accommodate a solution with substantial flexibility: data owners are often risk averse and seek trusted relationships with entities providing data access, customization is often needed, and finally, nationally specific legal protections must be followed.

This Milestone will continue to be refined. One part of the process is to double check with the service provided directly, via calls or email, to confirm that all details are accurate. Next, the spreadsheet and the supporting list of requirements will be added to the D5.10 White Paper. The final step will be to synthesize from this Milestone information needed to provide final recommendations in the White Paper.

4. Bibliography

European Commission. (2016). Guidelines on FAIR Data Management in Horizon 2020. Version 3.0. https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

European Commission, Eurostat. (2021). Remote access to European microdata. Report on legal, organisational, methodological and technical aspects (Unit A-5: Methodology; Innovation in Official Statistics) [Preliminary access].

Rat Für Sozial- Und Wirtschaftsdaten (RatSWD). (2019). Remote Access zu Daten der amtlichen Statistik und der Sozialversicherungsträger. RatSWD Output Paper Series. https://doi.org/10.17620/02671.42

Schiller, D. H., Eberle, J., Fuß, D., Goebel, J., Heining, J., Mika, T., Müller, D., Röder, F., Stegmann, M., & Stephan, K. (2017). Standards des sicheren Datenzugangs in den Sozial- und Wirtschaftswissenschaften—Überblick über verschiedene Remote-Access-Verfahren. In RatSWD Working Papers (No. 261; RatSWD Working Papers). German Data Forum (RatSWD). https://ideas.repec.org/p/rsw/rswwps/rswwps261.html

5. Appendix 1

Technical

- The internet (or private network) provides the remote connection between researcher and data.
- Connection uses standardized technologies for encryption of communication.
- Two-factor authentication is provided.
- Terminal server technical (e.g., CITRIX, VPN).
- Software for input devices to communicate with server (i.e., mouse and keyboard).
- Confidential data does not leave the host repository.
- Restrictions to prevent download, copy, or printing of data.
- Data only accessible for a defined period.
- Enable access to other software needed by user (SPSS, Stata, R, python, tools for tables and graphics etc.).
- Ability to import external data supplied by the user and subject these data to disclosure risk assessment.
- Ability to support data linking between data provided by user and sensitive data (screening of external data for security risks) (optional, not mandatory).
- Personal workspace for end-user (own storage space for result files, code libraries and other user created files).
- Ability for an authorized group of researchers to access the same data.
- Ability to submit results for disclosure review.
- Capacity to scale, as remote access increases demand – servers, support simultaneous users, service processing capacity.
- Logging facility of user activity is implemented:
    - Logon Duration.
    - Failure time.
    - Session state.
    - Session change time.
    - Session type.
    - Associated user Display Name.
    - Application Name.
    - Published application Name.
    - Application start time.
    - Application end time.
- Metadata in standard formats and at least Dublin Core available for datasets in RA facility.
- User needs to have: web browser, internet connection (other needs will vary with the system deployed, e.g. use of thin clients).

Legal

- Have some form of Secure Access User Agreement - legal document specifying:
  - services of parties outlining the tasks and responsibilities of Party A (RA provider) and Party B (RA user),
  - the period of the agreement and options for modification and termination,
  - specifications regarding available research data,
  - the occurrence of fees,
  - application process for Users, (not sure if this is in the User Agreement),
  - data access is based on appropriate legislative framework (GDPR and national laws),
  - a pledge on data secrecy/ a Secure Access Agreement (to be signed by the user and their organization),
  - reporting obligation for user to provider in case of a breach (of information security or procedure),
  - sanctions in case of a breach (of information security or procedure).
- Data handled in compliance with legal and intellectual property rights requirements of data owner.
- Compliance with national, and international legal requirements.
- Ability to prosecute individuals and institutions in case of breach.
- Contract and sanctions sufficient to replace physical security checks of Safe Rooms.
- Description of the obligations of the research entities hosting the access points.

Administrative

Procedures

- Define eligible user base (e.g., students).
- User account management:
  - Account set up.
  - User to sign agreement – need to validate signature.
  - Authenticate user identity.
  - ID user device (identity and location).
- Access request form, specifying user info and affiliation, data collections requested, purpose of research, access duration).
- List of accredited access points.
- Manage appointments.
- Set up working environment, including requested secure data.
- Output checking workflow and staff (ranging from self check to double review).
- Training for users in security requirements (various modes possible and consider recertification).
- List software supported.
- Interface language is English (mandatory; additional local language interface is recommended).

Resource Requirements

- Hardware costs.
- Software licenses.
- Need for admin staffing, support, service level.