Imports and Outputs at UKDS: The Practicalities

Simon Parker
Senior User Support and Training Officer

H.M.R.C.
11th December 2018

UK Data Service

Imports

- Researchers using the Secure Lab are permitted to import a wide range of files into their project areas.
- Users are unable to transfer files between the Secure Lab and their own PCs, and so all imports are carried out by the User Support team.
- Typically the files we are asked to import are data, syntax, and documents.

Type of Import	Number in 2018
Own data	142
Public data	227
Restricted data	39
Syntax	169

UK Data Service

The Data Import Process

- When researchers wish to import data into the Secure Lab they submit an Import Data Request form which is sent via email to the User Support team.

- To complete the form, researchers have to provide information about the source and ownership of the data, as well as a description about any intended linkage. A list of variables included in the data must also be provided.

- If the import was part of the original project proposal, or if it is publicly available aggregate data to be linked with ONS data, then we ask the researcher to securely transfer the data to us using our ZendTo service.

- If the previous conditions are not met, then we obtain permission from the Secure Access data owner for the import.

The Data Import Process Continued

- When the data is received the User Support Team officer assigned to the task is responsible for testing the data.
- A standard anti-virus scan is performed and then the contents of the data are checked.
- The purpose of the check is to ensure that the data files match the request made by the researcher. It is also to ascertain that direct identifiers are not included in the files; if any are found they must be removed before the data can be imported.
- After the officer has completed the checks, the data is added to the Uploads folder of the project area. They will then notify the researcher by email and update JIRA, our query tracking system.

Other Import Processes

- To import documents and syntax researchers do not have to complete an Import Data Request form. They instead email the User Support team the files that they wish to have imported.
- The files are also scanned and checked to ensure that they are safe to be imported. They are then placed in the Uploads folder of the appropriate project area.
- If a researcher requests a lookup table that may be use to other researchers, such as a spreadsheet to match postcodes to LSOAs then these are placed on the References drive. This drive is accessible to all Secure Lab researchers.

Stata ado files and R Packages

- On the References drive there are folders containing Stata ado files and R packages for use within the Secure Lab. If a required package is missing then researchers can email the User Support team requesting that it is added to the libraries.

- The required packages are downloaded by the User Support officer from either Ideas (for Stata ado files) or CRAN/MRAN (for R packages).

- These files are checked prior to being imported to ensure that they do not contain any data or files that have the potential to be problematic within the Secure Lab.

- In the event that a large number of R packages are requested simultaneously the officer will only need to inspect 10-20% of the files.

Outputs

- As researchers cannot remove files from the Secure Lab, they must make an output request to have their work released. They have to complete an Output Request form on our website through which a notification is emailed User Support team. As part of the request they must complete a checklist to confirm that the outputs meets the minimum required standards.

- Researchers may not request the release of data files, PowerPoint presentations, or Excel spreadsheets.

- If the output is in an acceptable format, a member of the User Support team will perform the first check of the output. We use a four-eyes approach with two independent checks of outputs conducted.

- One of stages of the check is use Word’s Inspect Document function which checks for hidden content, embedded objects, and other issues that may not be part of traditional SDC.

UK Data Service

Outputs Continued

- In addition to checking for problematic content, the first output checker will also lock a document to ensure that it is not edited by the researcher between checks.

- Checks are conducted in accordance to the SDC guidelines in the SDAP SDC Handbook. We aim to respond to researchers within 4 working days.

- When they have completed their checks, the first output checker assigns the task in JIRA to the second checker. They will also provide notes about their checks. This is used to highlight aspects which are problematic, but also to explain the justification behind decisions.

- The second checker will then perform their checks and unlock the document.

Releasing Outputs

- If both checkers agree that a document is safe to be released then the second checker is responsible for doing so. The document is attached to the JIRA issue and is emailed to the researcher.

- The document is attached to the JIRA issue so that we have a record of all documents that have been released. This is to protect UKDS from SDC issues caused by document editing after release.

- If both checkers agree that a document cannot be released, then the second checker will notify the researcher as to our concerns. They will then be expected to make the required changes, or provide the necessary information in a new document within the SDC folder.

- We request that the save updates in new documents to make rechecking outputs quicker.

UK Data Service

Rechecking Outputs

- When a researcher has made the changes requested they notify the User Support team via email.
- As the issue is still assigned to the second checker, they are responsible for the first recheck.
- It is possible to compare Word documents using the comparison function, and tex files using Notepad++. This significantly speeds up rechecks.
- The process works in the same way as the first attempt. When the checker has completed the first recheck they assign the issue to the original first checker to conduct the second recheck.
- If the changes have been made in satisfactory way the document is release to the researcher.

UK Data Service

Disagreements

- Output checkers should be in agreement before a document is released or before the researcher is asked to make revisions.

- It is very rare for checkers to disagree, although they may spot different issues or interpret documents differently.

- Most usually the output checkers will discuss their thoughts and will arrive at a consensus of what to do.

- In the very rare event that a decision cannot be agreed upon, a third output checker will be asked to review the document.

- Researchers may formally challenge a decision not to release a document. At which point the User Support Manager will be asked to review the decision. The Director could also be asked to adjudicate.

UK Data Service

Reviewing

- The User Support Manager is responsible for reviewing the decisions taken by output checkers.
- They independently recheck a sample of previously released documents to check that unsafe documents are not being released.
- The inclusion of information about the decisions taken in the JIRA task is invaluable for reviewing releases. Checkers should not simply highlight perceived problems but should, where possible, explain why they believe aspects are safe to be released. This is particularly pertinent in a principles-based SDC approach.
- The reviews and checkers' notes feed into any external audit of the decisions taken by UKDS with regards to releasing outputs.

UK Data Service

Questions

support@ukdataservice.ac.uk

Simon Parker