STASH User Guide

STASHResearch

STASH is a version-controlled, direct data deposition platform. This means that STASH automatically pulls data from the data source specified by the project owner into it's own repository as the data is being accumulated in real-time. Via version control all changes made to the data are automatically tracked and recorded, allowing for the researcher to keep a sharable record of how the research data accumulated during the research project. This sharable history can also serve as a verifiable record that the data is kept in its original raw form as it was collected without any manipulation. Or, if data needs to be edited or deleted during the study (for example due to GDPR compliance) the changes in the data can be transparently commented. Importantly, may or may not be linked with data sharing. The STASH repository's history can still serve as a record of data integrity without data sharing. The project owner has control over who they share data with and which parts of the data they share.

Disclaimer
The first part of this Guide is written in a way that you can set up your own test project in STASH by following it step-by-step. In order to get the demonstration material which we have used in the guide, please visit: https://github.com/equarius93/stashtest/blob/main/video%20demonstration%20covid.csv. Please save this file on your computer and also add it to a Google Drive folder which you are going to be using as a data source. This file contains some COVID-related data about Afghanistan from 2019. Please make sure to transform this file into a Google Sheets document in your Google Drive folder.

I. Profile and Project Registration

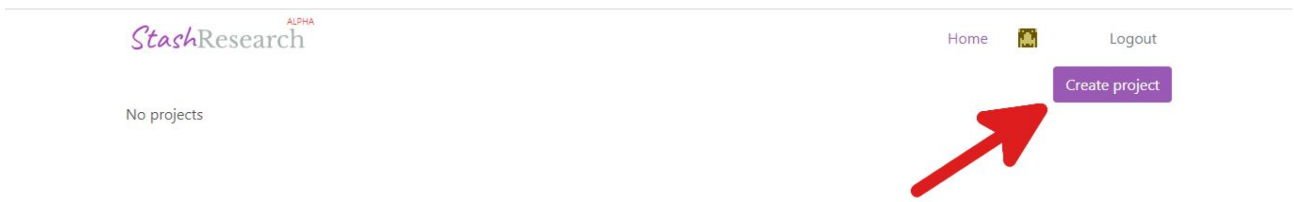
I.1 Profile Registration and Login

In order to use the STASH website, you are going to need a registered account. On the main page of the website, click on „Register".

After filling out the required information, hit submit and you are going to receive a confirmation email to the email you've specified. Now you can log in.

I.2 Project Registration

For each of your projects you are going to need a separate project registration. In order to register a project, head to the front page of the site and hit ,,create project".

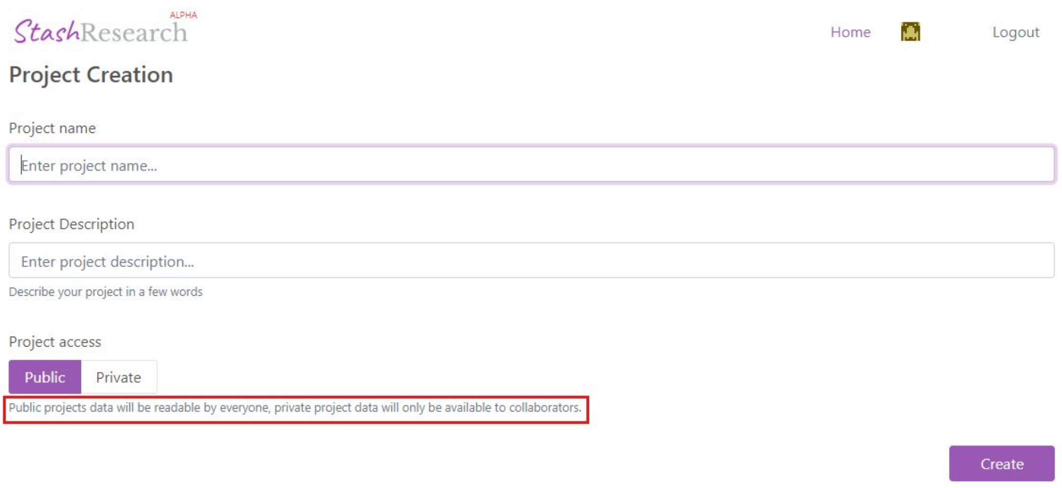


On the following page you will have to give a name and a description for the project. you can fashion the description whatever way you would like to. We advise that you try to summarize your project. you can make the project either public or private. Public projects' data will be readable by everyone, private project data will only be available to collaborators.

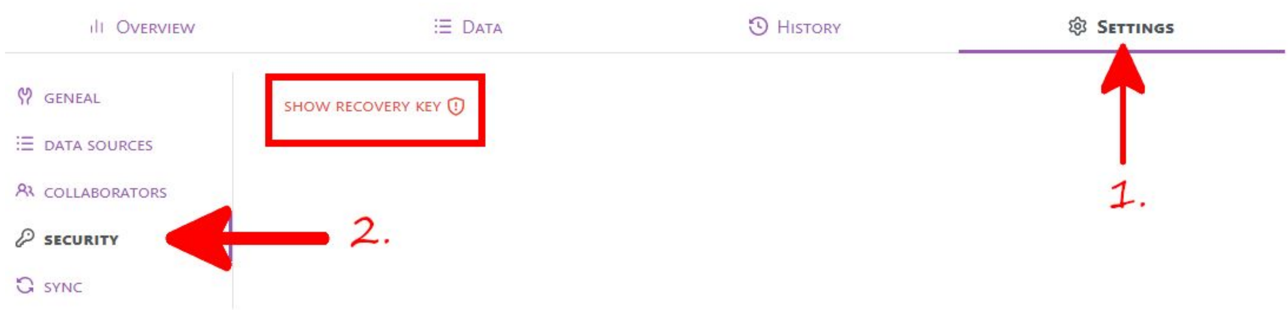
Please note that the name, description and the type (public or private) of the project cannot be changed later in the current version of STASH.

Hit “Create” when you are done. On the next screen you are going to have to give your account’s password to ensure the safety of your data.

We take the security of your data extremely seriously and have multiple features built in to STASH that can guarantee that only you and your specified collaborators (feature upcoming) have access to the data.



After you have confirmed your account password, you are going to be greeted by a data encryption page. Please make sure that you save the recovery key to a private and secure place. You can check your recovery key anytime under project→settings→security.

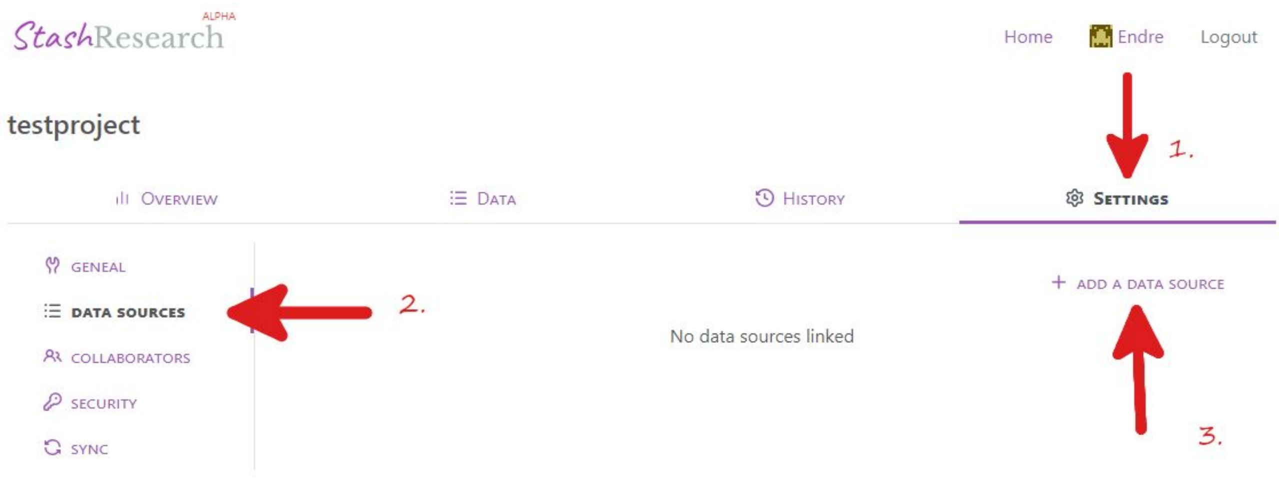


You are going to need this in case you forget or lose the password of your Account. Save your recovery key and hit “submit” to continue.

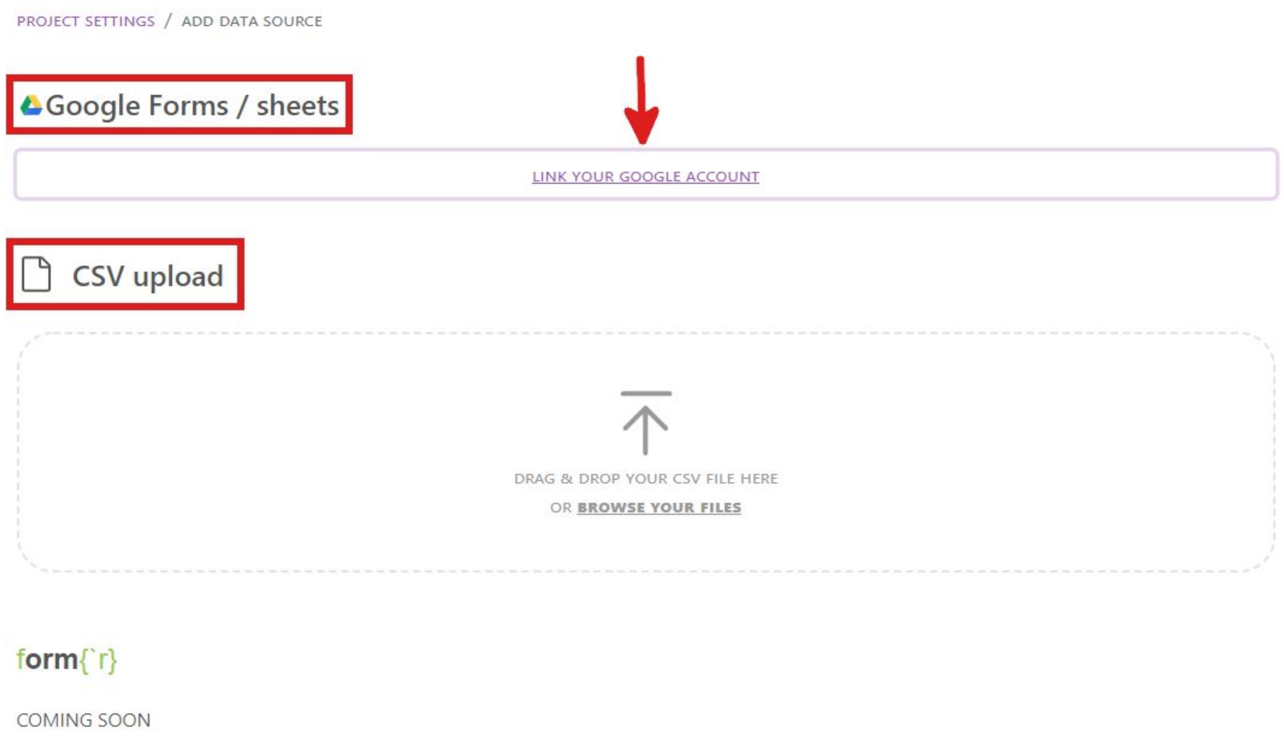
I.2.2 Data Source for the Project

Now we have our project created, time to fill it with content. First we are going to have to add a data source, from which Stash is going to pull the data.

After the data encryption page you are going to be greeted by the overview page of your project. To add a data source, head over to “Settings” and select “Data Sources” on the left pane. As of now this should be empty, select ,,Add a Data Source”. Please note that the name of the data source needs to be unique within a project.



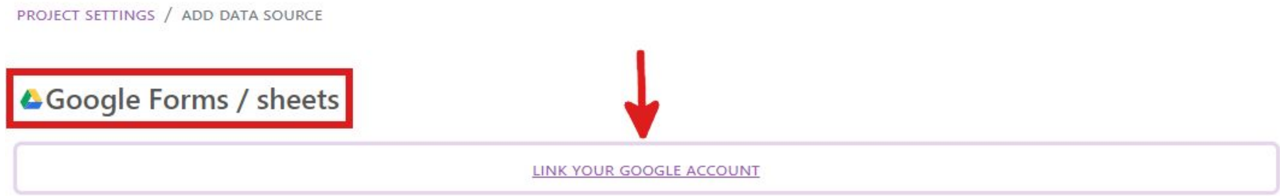
As of now, there are two kinds of data sources in STASH: Google Sheets (this includes the google sheets in which Google Forms collect data) and manual upload of .csv files. FormR is also going to be supported in the future.



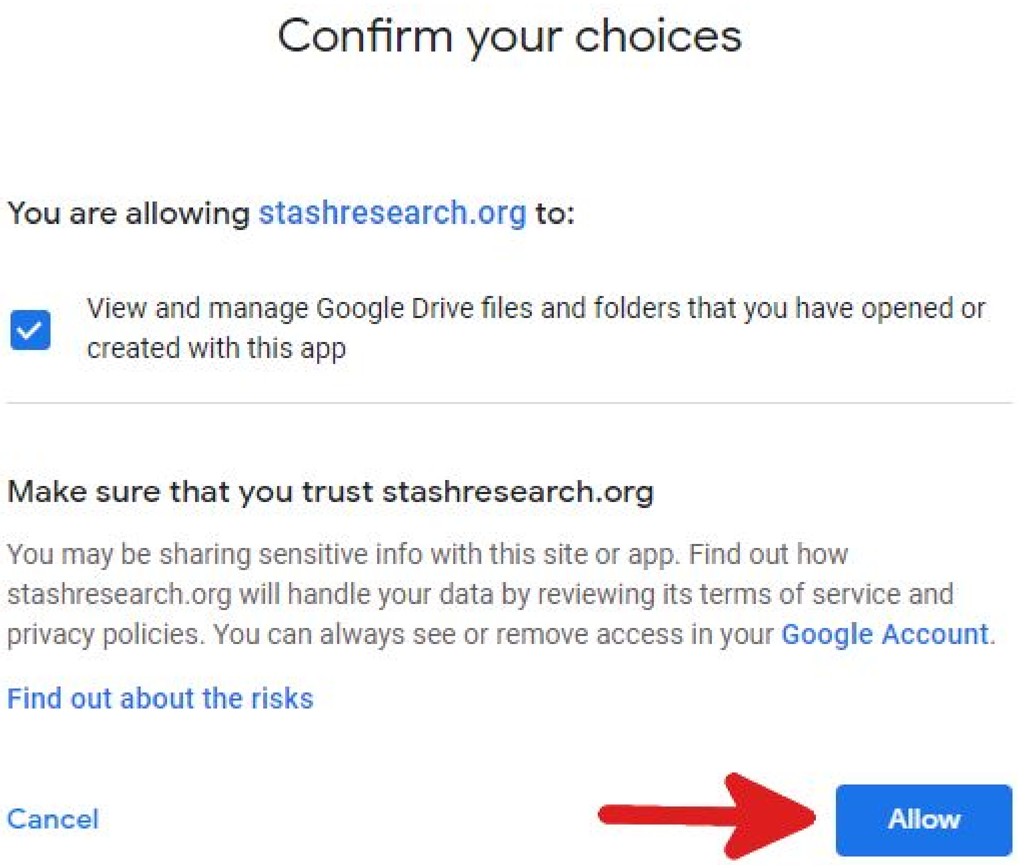
I.2.2.1 Google Sheets/Forms

This section tells you how to link a Google Sheet as a data source to STASH. First of all you are going to need a Google Sheet file. For a demonstration material please follow this link.

Head inside your project, and then (1) Settings→ (2) Data Sources. Hit (3) Add a Data Source.

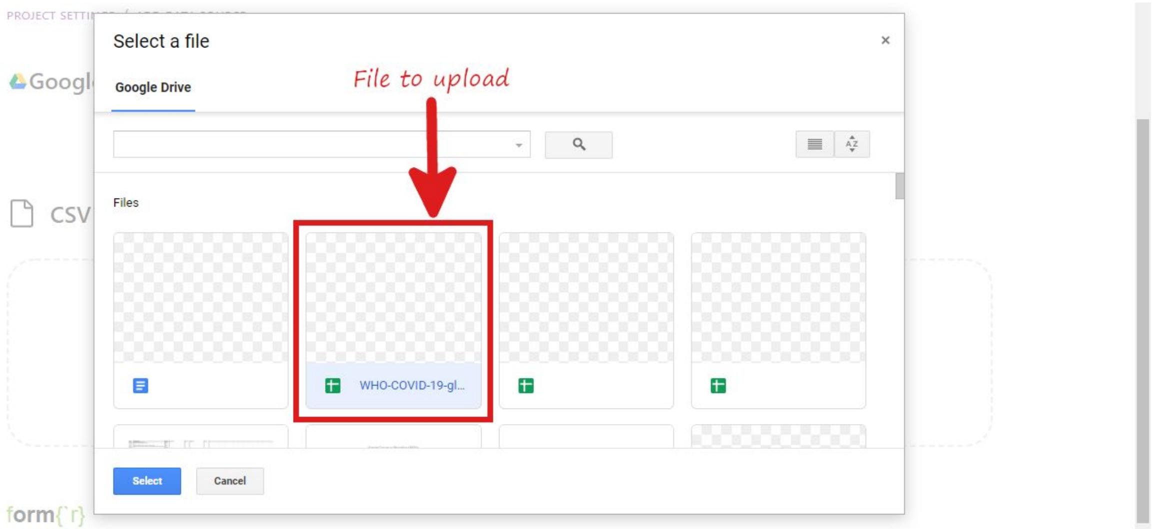


On the following page you are going to see several methods for adding data sources. hit “link your google account”.
A Google login page will pop up. You will have to grant STASH certain permissions, which you can read on the page. Stash will only ask access to the specific file that you have selected, so it won’t have access to any other files or folders on your Google Drive. Make sure to hit “Allow” before you close the window. Please note that currently only one Google account can be linked to a STASH account.



Now that you’ve given all the necessary permissions to RA you should refresh the data source page. As you can see, a “Link new file” option appeared under “Google Forms / Sheets”. Click on that to proceed.

A window is going to pop up where you can select the file which contains your database. Please note that it has to be a Google Sheets file. you can use the search field to look for your file.



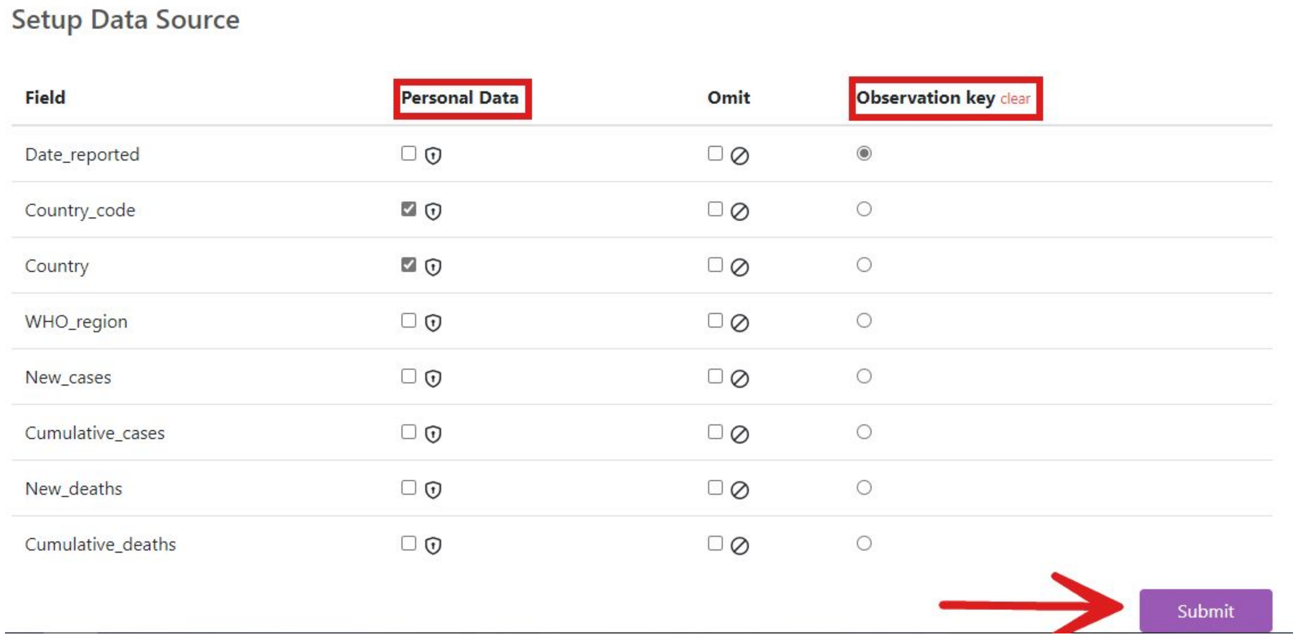
On the following screen you can see the headers of your data. On this page you have the option to mark columns as personal data, omit columns and select an observation key.

you have to specify an Observation Key column. should be a field which is unique to each row in the dataset. This serves as a row identifier for Stash to better be able to determine what was changed in the dataset between different versions of the dataset. This allows us to mark the reordering of the observations without flagging a change in the data. If you don't use any observation key then the number of the rows are going to be used for this purpose. In this case if you reorder the rows in the source dataset, Stash will flag a change in the data change history.

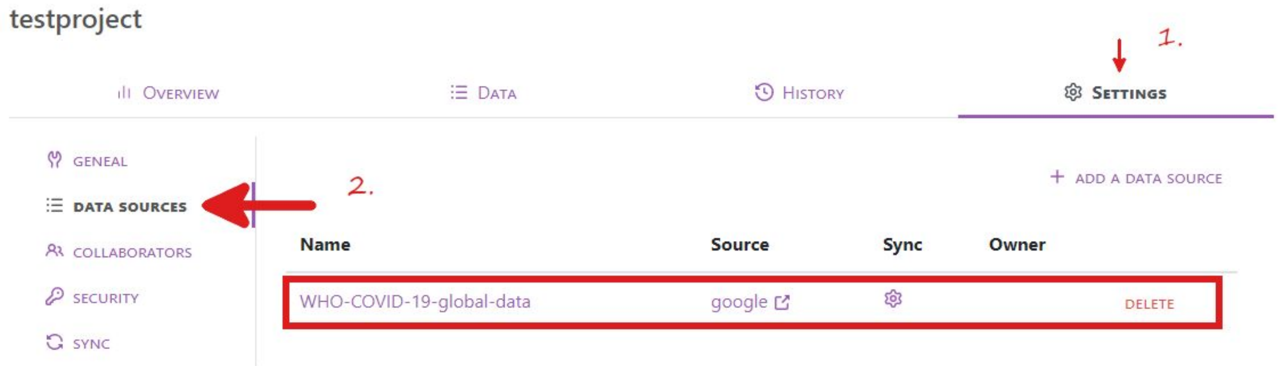
All columns marked as containing Personal Data will be encrypted. The decryption key for this data is linked with your user password so that no one can see these data fields without permission.This has been added as an extra security measure to ensure the safety of any personal or sensitive data. Note that not even the RA team can access data encrypted this way. This means that if you forget your password, the data cannot be recovered by the RA team.

Make sure to save the recovery key to your project securely! You can check your recovery key later on inside your project→settings→Security. If you have to reset or change your password for some reason, the new password will NOT unlock the encrypted data, because the decryption key is linked with the password with which you encrypted the data. The only way to recover the encrypted data is via supplying the password you used then, or by supplying the recovery key that is presented to you at the creation of the project. That is why it is very important to save the recovery key for your projects securely.

Once you have done the data setup there is no way to change the options.



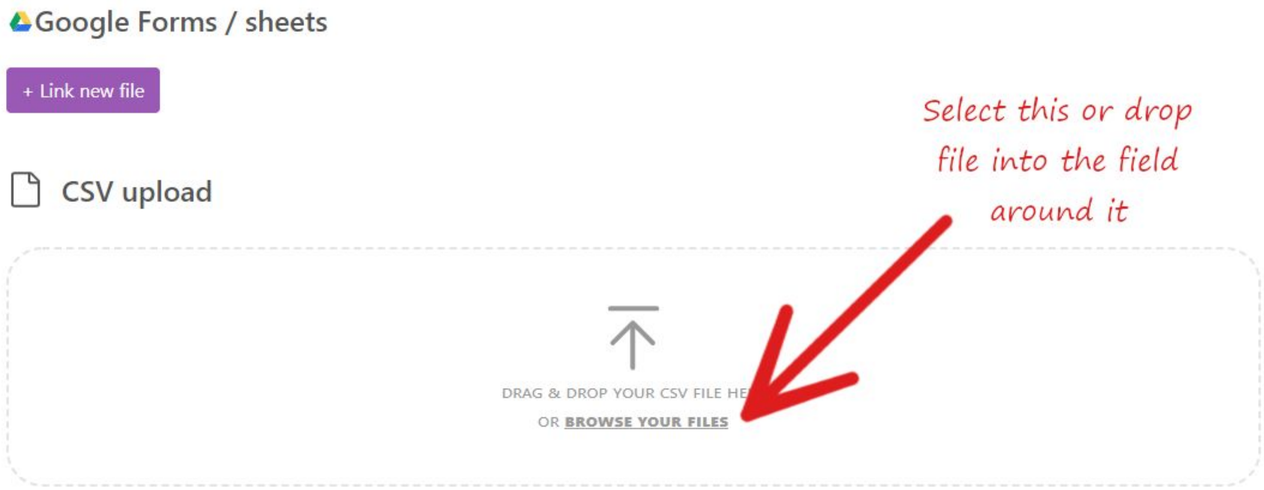
Whenever you add a new data source, you should make sure that it has been added by going over to the “Data Sources” inside “Settings” in your project. If everything went well the new source should be shown and the sync is going to be marked as “continous(started)”.



As you can see, there are several things that you can do to manage your sources here. For a more detailed description, please check chapter II.1.2.2.1 “Data Sources”. Also if you head over to the “Data” page inside your project, you should see the complete dataset. For a more detailed description about the Data page, please check chapter II.1.2.1 “Data”.

I.2.2.3 CSV File

If you would prefer to upload your dataset from a file manually, you have to make sure that it is saved in a .csv file format. Head inside your project, and then Settings/Data Sources/Add a Data Source. Under CSV Upload, hit “Browse your files” or just drag and drop the file there.

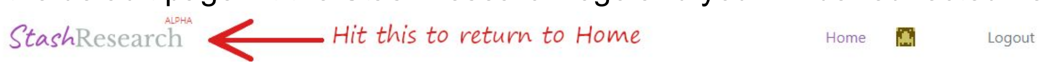


Data Setup page here as well?

II. User Interface

II.1 Dashboard

There are 3 tabs on the main screen: Home, Account Settings and Logout. By default you are going to find yourself on the Home tab after logging in. Whenever you would like to return to the default page hit the StashResearch logo and you will be redirected here.



II.1.1 Home

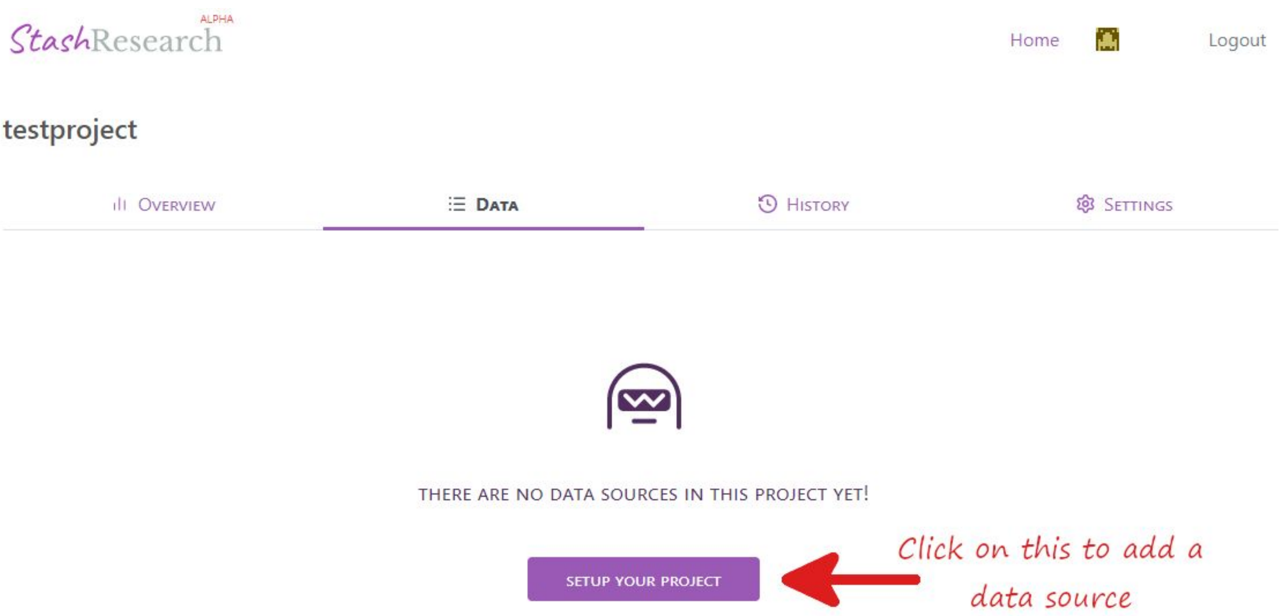
On your Home page you can manage the projects of your Account. If you already have a project for this account, you should see it's name here (1). you can also check when was the last update on this project (2) or delete it (3). Please note that once a project has been deleted, currently there is no way of recovering it. In order to create a new project, select "Create project" under the "Logout" button (4). As always, you can hit the StashResearch logo to return to the Home page (5).

II.1.2 Project page

When you select a project on the "Home" page you will find yourself inside the project's page. Currently only two pages are functional here: "Data" and "Settings".

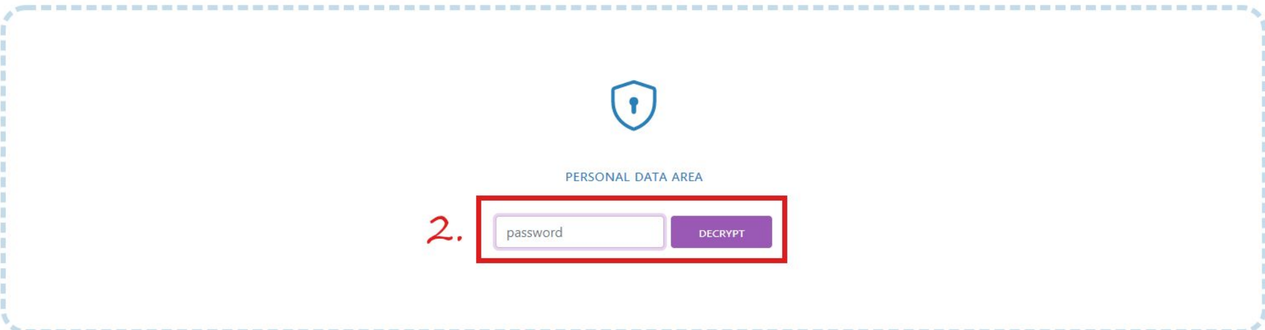
II.1.2.1 Data

This page contains the data for your project. If you haven't added any data source yet, this page is going to suggest that you do. In order to do this, you should hit "Setup your project" which is going to take you inside "Settings" or head over to "Settings" right away.



Once you have successfully added at least one data source to your project, you are going to see them listed here.
When you select a data source the table of your data is going to open. you can see which columns were marked as personal (sensitive) data by looking at the headers. Sensitive data have a shield icon above them and also by default their content is hidden.

you can reveal these hidden data by clicking on “show personal data” and providing your account’s password.

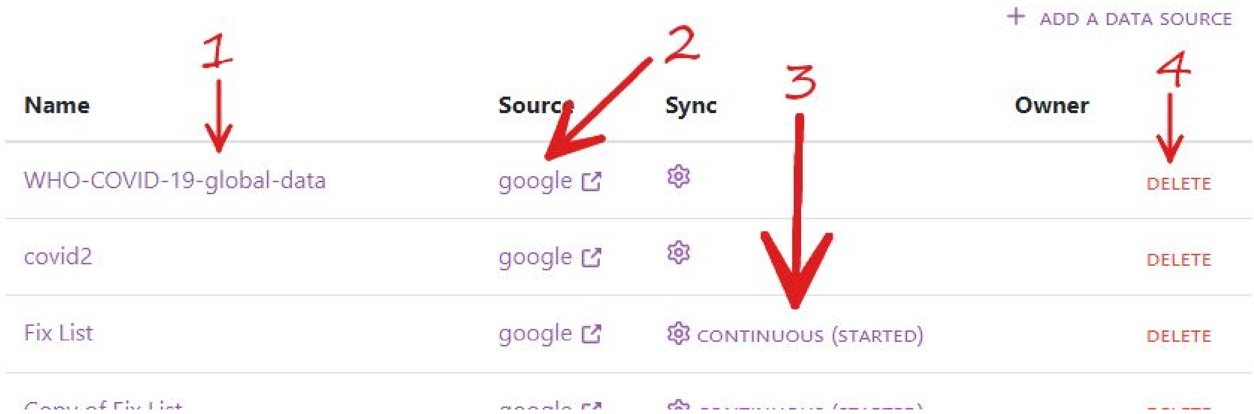


II.1.2.2 Settings

Under your project’s Settings page you can find the most important parameters of your project. These include the managing of data sources, checking the security key of your project and exporting data onto GitHub.

II.1.2.2.1 Data Sources

On this page you can add new data sources to your project or manage those that you have already added. To add a new data source click on “add a data source”. For a detailed guide about adding data sources please check chapter I.2.2 “Data Source for the Project”.



If you take a look at your current data sources, you can see their (1) names, (2) the data source, (3) the status of the synchronization, (4) or you can delete those that you no longer need. Clicking on the name of the data source will take you right to its dataset in the Data page of the project (1). There is also a direct link to the data source (for example if the data source is a Google Sheet, this link will open that Sheet in Google Drive (2).

II.1.2.2.2 Collaborators

To invite collaborators to your project, head inside the settings page (1) of your project and select collaborators (2) on the left pane.

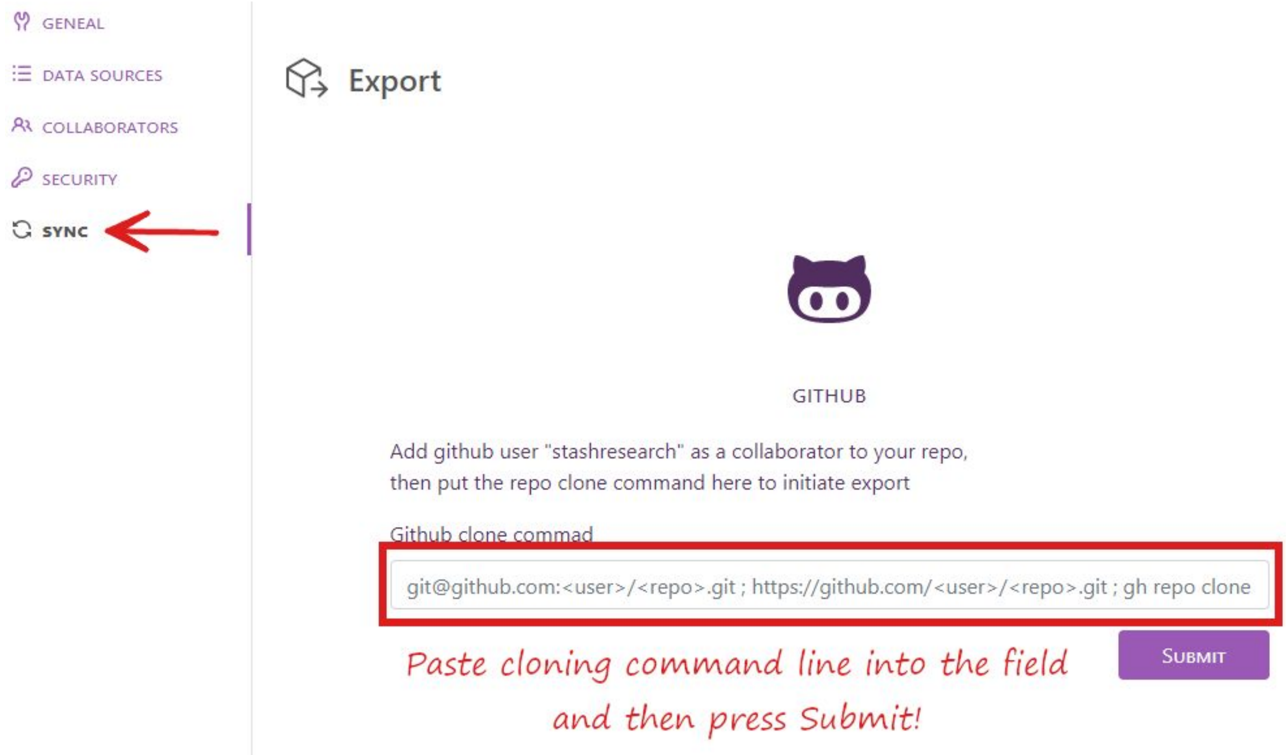
Here you’re going to have to provide the email, the first name and the role of the collaborator.

II.1.2.2.3 Security

On this page you can check the recovery key for your project should you need it. In order to get access to it, click on “show recovery key” and provide your account’s password on the following page.

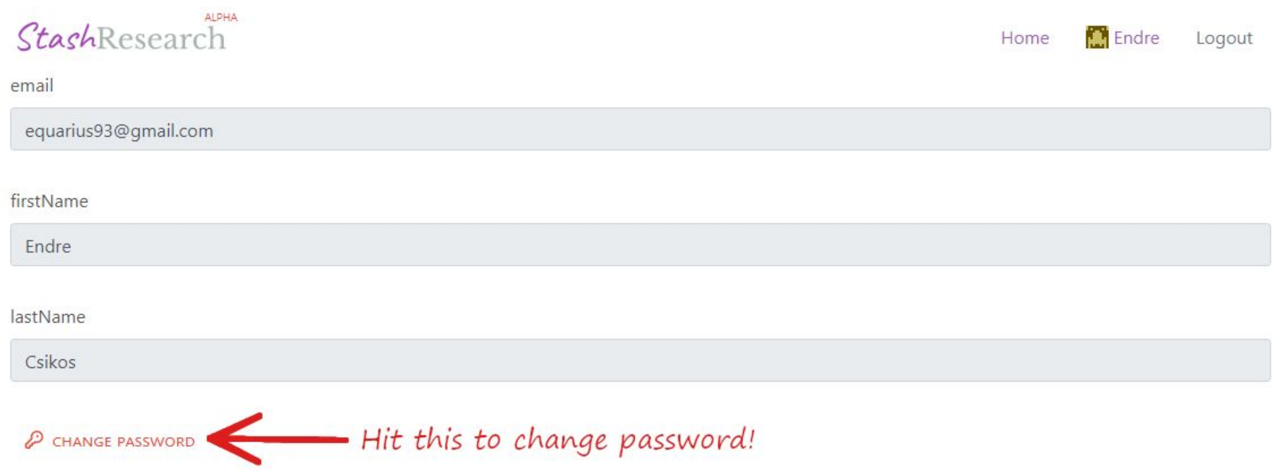
II.1.2.2.4 Sync

On this page you have the option to export the data of your project to a repository on GitHub. you can do this by adding “stashresearch” as a collaborator to your GitHub repo and then pasting the clone command in the Sync page on RA. For a more detailed guide on how to export data in RA please check chapter [V. “Data Sharing & Export”].



II.1.2 Account Settings

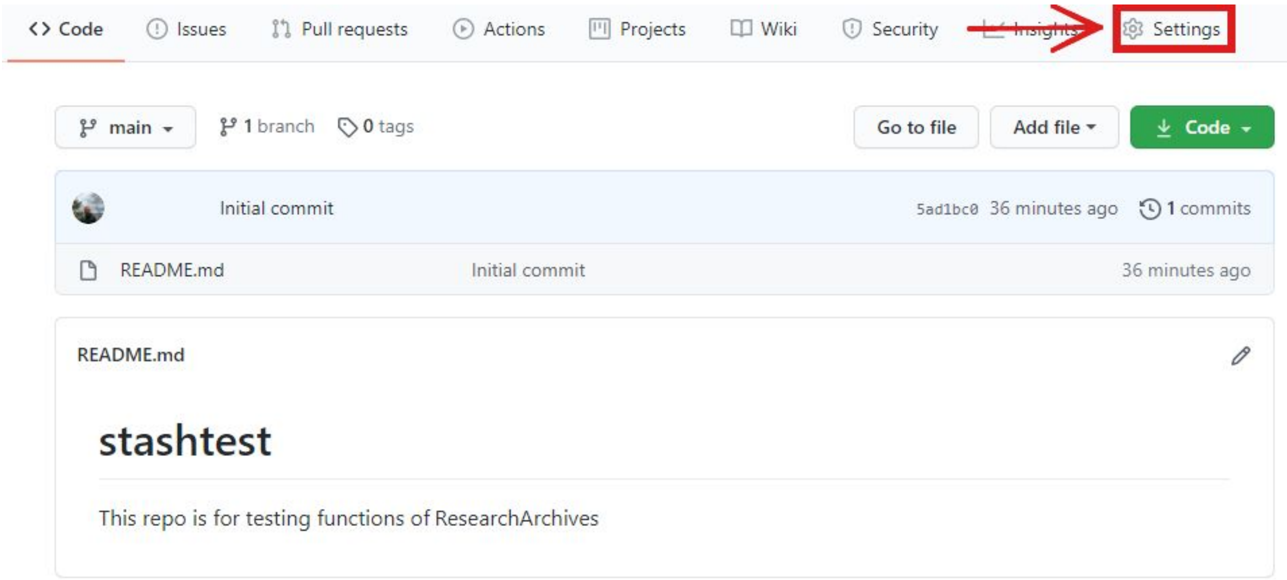
On this page you can see the associated email address, first and last name for the account. you can also change your password if you hit “Change Password”. To do this, RA is going to send you a confirmation mail to the email address which you have given for this account.



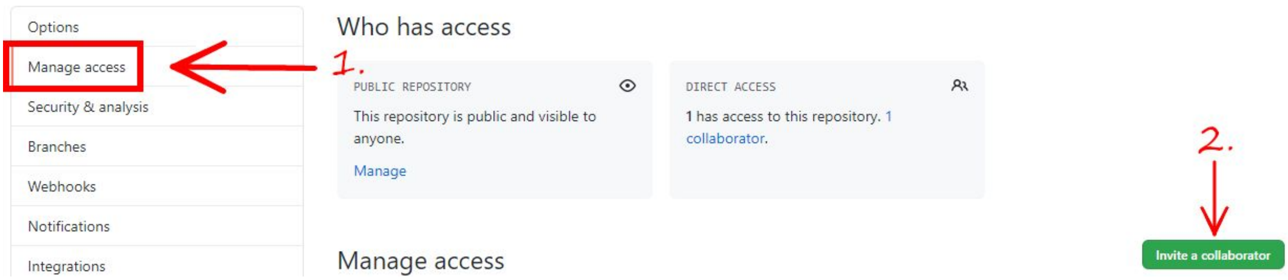
III. Data Sharing & Export

ResearchArchives allows you to export the content of your project to a GitHub repository. To do this, first we are going to create our repository. If you don’t have a GitHub repository yet and are not familiar with it, you can create an account at https://github.com/. you can

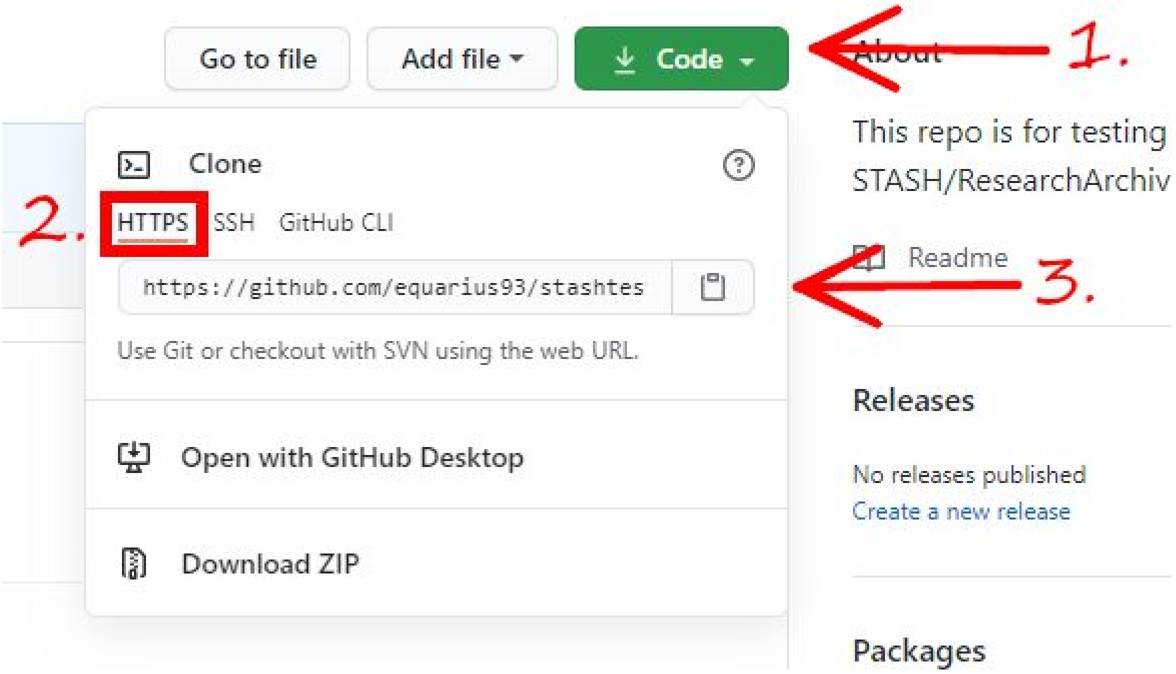
read a brief guide on how to create a repository here (you can stop following the guide once you created a repository.)
Head to the main page of your GitHub repository and select “Settings”.



Once inside settings, look for “Manage access” on the left pane (1). On this page you can manage who has access to your GitHub repository. In order for RA to be able to push the exported data onto your repository, you will have to give the necessary permissions. To do this, we are going to invite RA as a collaborator. Hit “Invite a collaborator” (2) and enter the username "stashresearch". If you have done everything right, "stashresearch" should appear as a pending invitation under “Manage Access”.



Now all we have to do is copy the clone git command and paste it inside Sync on RA. Let’s head back to the overview page of your GitHub repository. Just under the “Settings” you can find a big green button labeled “Code” (1). Click on it - make sure that HTTPS is selected (2) - and hit the copy button (3).



Now head back to the Sync page in RA and paste the code into the field and hit “Submit”. If everything went well, the data should appear on the main page of your GitHub repository.