



seriss

SYNERGIES FOR EUROPE'S
RESEARCH INFRASTRUCTURES
IN THE SOCIAL SCIENCES

Deliverable Number: D4.15

Deliverable Title: Online Library of Harmonisations: Report on workflows for harmonisation submission and community peer-review processes

Work Package: WP4: Interactive tools for cross-national surveys

Deliverable type: Report

Dissemination status: Public

Submitted by: CESSDA (GESIS – Leibniz Institute for the Social Sciences)

Authors:

Dr. Kristi Winters (Task 4.5 Project Manager)

Matthäus Zloch (Task 4.5 Computer Scientist)

Martin Friedrichs (Task 4.5 Deliverables Manager)

Date Submitted: December 2017

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 654221.





www.seriss.eu  @SERISS_EU

SERISS (Synergies for Europe's Research Infrastructures in the Social Sciences) aims to exploit synergies, foster collaboration and develop shared standards between Europe's social science infrastructures in order to better equip these infrastructures to play a major role in addressing Europe's grand societal challenges and ensure that European policymaking is built on a solid base of the highest-quality socio-economic evidence.

The four year project (2015-19) is a collaboration between the three leading European Research Infrastructures in the social sciences – the European Social Survey (ESS ERIC), the Survey of Health Ageing and Retirement in Europe (SHARE ERIC) and the Consortium of European Social Science Data Archives (CESSDA AS) – and organisations representing the Generations and Gender Programme (GGP), European Values Study (EVS) and the WageIndicator Survey.

Work focuses on three key areas: Addressing key challenges for cross-national data collection, breaking down barriers between social science infrastructures and embracing the future of the social sciences.

Please cite this deliverable as: Winters, K., Zloch, M. and Friedrichs, M. (2017) *Online Library of Harmonisations: Report on workflows for harmonisation submission and community peer-review processes*. Deliverable 4.15 of the SERISS project funded under the European Union's Horizon 2020 research and innovation programme GA No: 654221.

Contents

Introduction..... 4

Project development to date 5

Next steps in the project’s development..... 5

Constructing the user interface..... 6

 Figure 1: Place holder content for User Personal Profile page. 7

 Figure 2: Place holder content for User Submissions Overview page. 8

 Figure 3: Place holder content for User Submissions Pending page. 9

 Figure 4: Place holder content for User Invitation to Review page..... 10

The Workflow of the Community Peer-Review Digital Infrastructure..... 11

 Users..... 11

 Background research 12

 Table 1: Core steps identified as the basis of the workflow by user role and rights. 12

 Submission review, acceptance and depositing workflow 12

 Figure 5: Overview of the submission, review, acceptance and depositing workflow. 15

Introduction

Digital tools are the cornerstone of research infrastructures in the social sciences. Despite our digital age, a large amount of work is carried out manually and this results in reduced efficiency, redundancy of effort, and wasted resources. The suite of tools developed by SERISS members will take cross-national survey research to a new level by automating key processes. Our task will provide support for data harmonisation practices with improved online tools

Cross-national and longitudinal data harmonisation and data linking are vital in realising the full reuse value of data. The research community will benefit greatly from data synergies and the simplification and routinisation of ex ante and ex post data harmonisation workflows. Task 4.5 develops a new, online library of harmonisations where digitally documented projects will be archived and accessed. Users will be able to submit their harmonisations work to an accessible user-friendly website interface for peer-review by fellow scientists and, when accepted, the harmonisation project will be published with persistent identifiers for use as a citeable reference.

Online users will be able

- to search for published harmonisation projects (not subject to copyright and therefore open to public use);
- inspect the harmonisation documentation;
- download the published syntax for replication with new harmonisation and/or data integration projects, with only the obligation to give credit to the harmonisation author in everything they publish; and
- the online tool will accept digital harmonisation routine submissions using an XML export from the desktop harmonisation tool and other appropriate structured file formats (i.e. DDI).

This deliverable first provides a summary of the work that has been completed on developing the online library of harmonisations. It then provides more details on the user interface that will be developed for the tool before focusing on defining the workflow for the process of peer review harmonisations submitted to the online library will undergo.

Project development to date

Server setup (prepared) Between March and October 2017, our computer scientist set up the main server. The main server will host the Online Library web application; a database where application data will be stored (e.g. user management data, like user permissions and user roles); a search index database (for quick and easy access of searchable metadata); and further libraries (e.g. CharmStats modules) to be used by the Online Library application.

Extraction of CharmStats modules as reusable libraries (prepared) The CharmStats core application contains functionality that will be adapted for the Online Library application, e.g. the Report Generator, Graph Tool, and further related utilities for harmonization script conversions. To reuse these modules in the Online Library the computer scientist had to extract them from the CharmStats software code and make them be usable as independent libraries. We will use these to build an independent CharmStats microservice (more below) where all CharmStats libraries will be hosted. This is a necessary preparation to ensure submission process workflow functions properly in the Online Library.

Development of peer-review process (done) The Online Library application contains a peer-review process that begins after a user submits a harmonization script. By 'script' we mean the relevant metadata on the value, variable, question and study levels such that an outside researcher would have all the information they require to replicate the work.

A workflow process identifying the necessary users, actions, roles, and permissions was developed within the team and each step in the process was specified. The workflow presented later in this report will guide the construction of the website. As a next step, the peer-review process will be implemented in the Online Library (due mid-2018).

Creation of mock-ups in preparation for implementation (done) As a preparation for the actual implementation of the application we created graphical sketches, so called mock-ups, of the planned application. This is a common practice in modern software development as the mock-ups illustrate how information flows in the application. The Online Library team members agreed upon the preliminary mock-ups, and they will be used as templates for the implementation.

Considerations about technical architecture and software stack (almost done) In a preparation for the actual implementation of the application we have spent time considering the technical architecture and the possible software stack to be used. As a first third-party product trial we set up and evaluated the *dspace* platform for use as document storage on our main server, in addition to using some reusable modules from the CharmStats core application (see above).

Next steps in the project's development

Implementation of the Submission Process The Submission Process describes the process from a user's initial login to the Online Library until the peer-review process starts (see Figure 1). In this process the user submits a script or a CSxD (CharmStats Exchange Document) providing all the necessary metadata for the submission (e.g. variable, question and study level metadata). In case the user uploads a manually generated, non-CSxD script, the manual version of the script along with all its additional metadata will be transformed using the CharmStats library (see preparation above).

In this phase the basic structure of the web application will be implemented. Also, the microservice that stores a submission will be implemented. The work is due at the end of January 2018, and will be deployed as our version v0.2. The goal is to have the basic functionality to submit and store harmonization scripts as well as the metadata in the database.

Implementation of the Search Functionality ‘Search Functionality’ describes how users of the Online Library may search the system for available harmonizations. In this phase the basic structure of the search interface and results page will be implemented, including the setup of the index itself on our server. This functionality is scheduled to be ready by the end of February 2018 and will be deployed as our version v0.3. The goal is to produce a search interface with appropriate results.

Testing with example data In this phase we will import example data provided by our partners at the Heinrich-Heine-University Düsseldorf into our system. Access to and use of their pooled dataset with newly harmonized variables (complete with the relevant metadata on the question and study levels) has already been granted. We will use their completed projects and data to test the website’s data handling systems, the project submission process (e.g. detecting duplicate study level information to connect with an existing study level entry in the dataset), and the search and result interface.

Download of Submissions Uploaded harmonization scripts may be downloaded by other users. In this phase the download functionality will be implemented. Since this includes the transformation of the actual scripts and additional metadata into the CSxD-format the required CharmStats microservice must be set up and running on our server. This functionality is planned to be realized by the end of March 2018 and will be deployed as our version v0.4.

Next steps After the above parts are implemented and are debugged and stable for use, we will continue developing the functionality associated with the peer-review process (starts by **April 2018**).

Constructing the user interface

In this section we review the preliminary structure of the user interface. Given the early stage we are in, we expect these layouts to change significantly while the library is built. This includes, but is not limited to, any SERISS branded logos, font use requirements, color scheme requirements, or links to other SERISS-related websites.

To keep the project running on time, before our computer scientist left on his family leave we requested he specify which elements users would interact with on the website and the options involved with each.

Figures 1 through 4 are the first attempt to assemble and display relevant information to various users; be that studies users have submitted and their study documentation or reviewers who are inspecting the harmonization documentation. Please note that the information provided on each page will be specific to the logged-in user, based on the rights and access granted them (e.g. what those submitting documentation view onscreen versus what a reviewer sees).

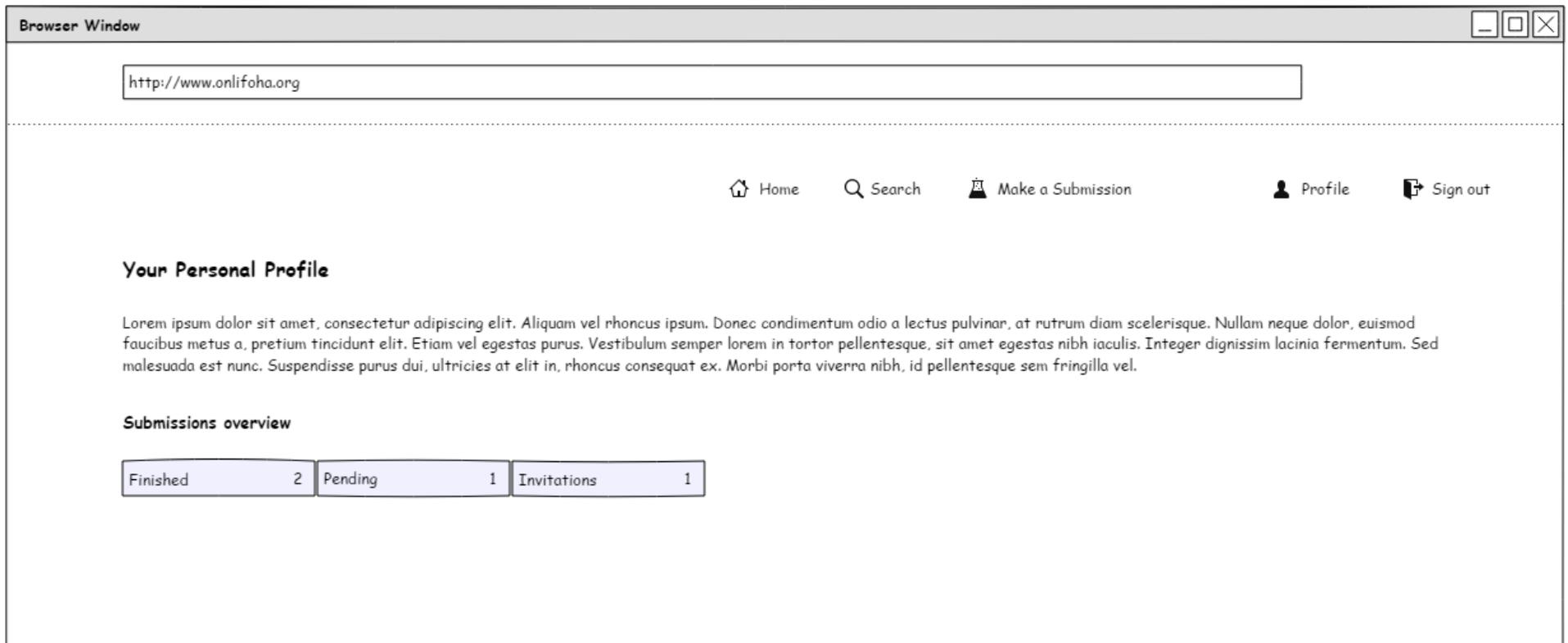


Figure 1: Place holder content for User Personal Profile page.

The “User Profile” view. The user sees an overview of all his/her/their submissions. Users are offered the following information related to their account: to review “Finished” submissions, “Pending” submissions, and any “Invitations” to review they have received.

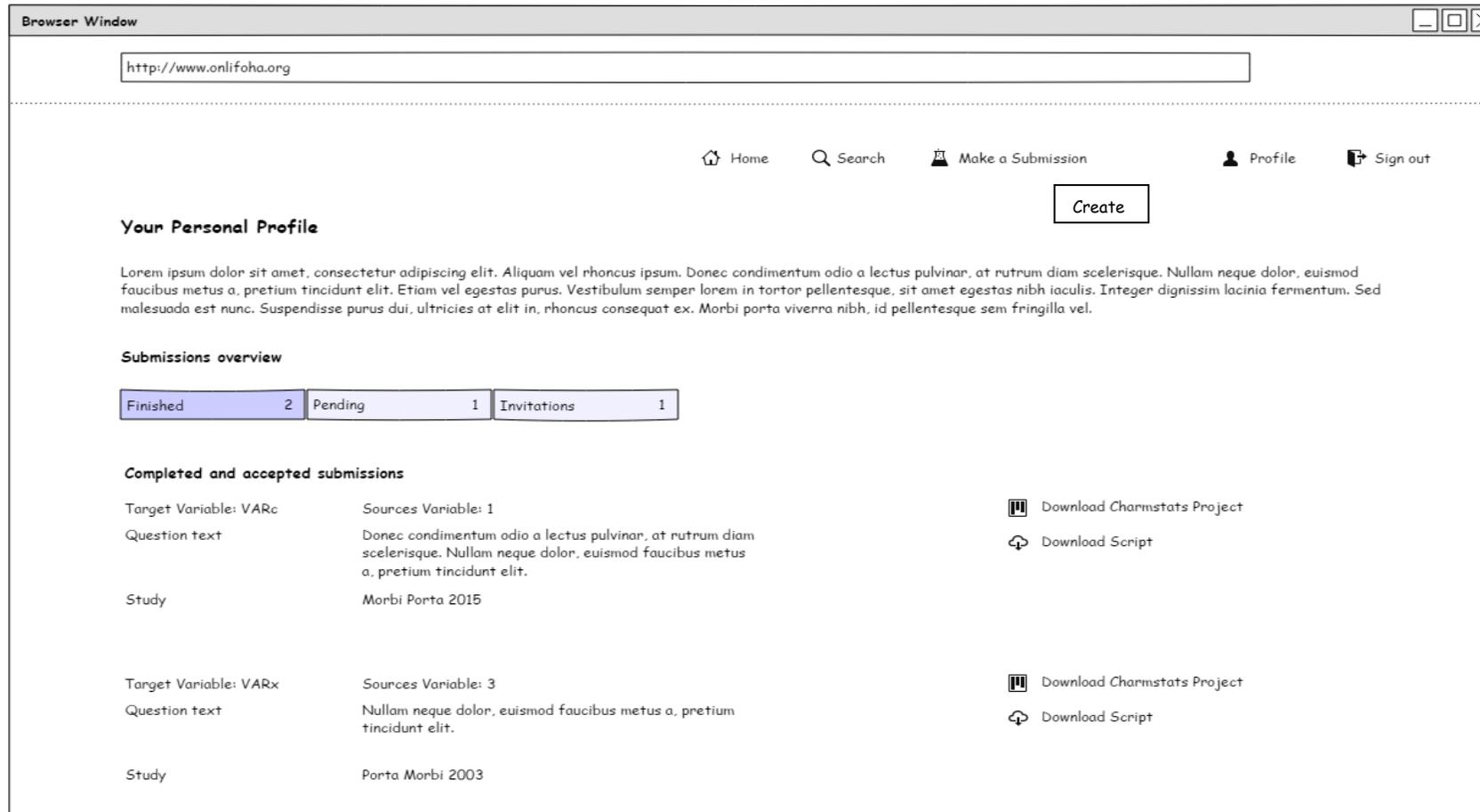


Figure 2: Place holder content for User Submissions Overview page.

A list of completed and accepted submissions is shown (peer review process finalized). Also shown are relevant metadata for the submitted datasets (e.g. target variables, question text, study). The CSP-file and corresponding harmonization script may be downloaded through the links on the right.

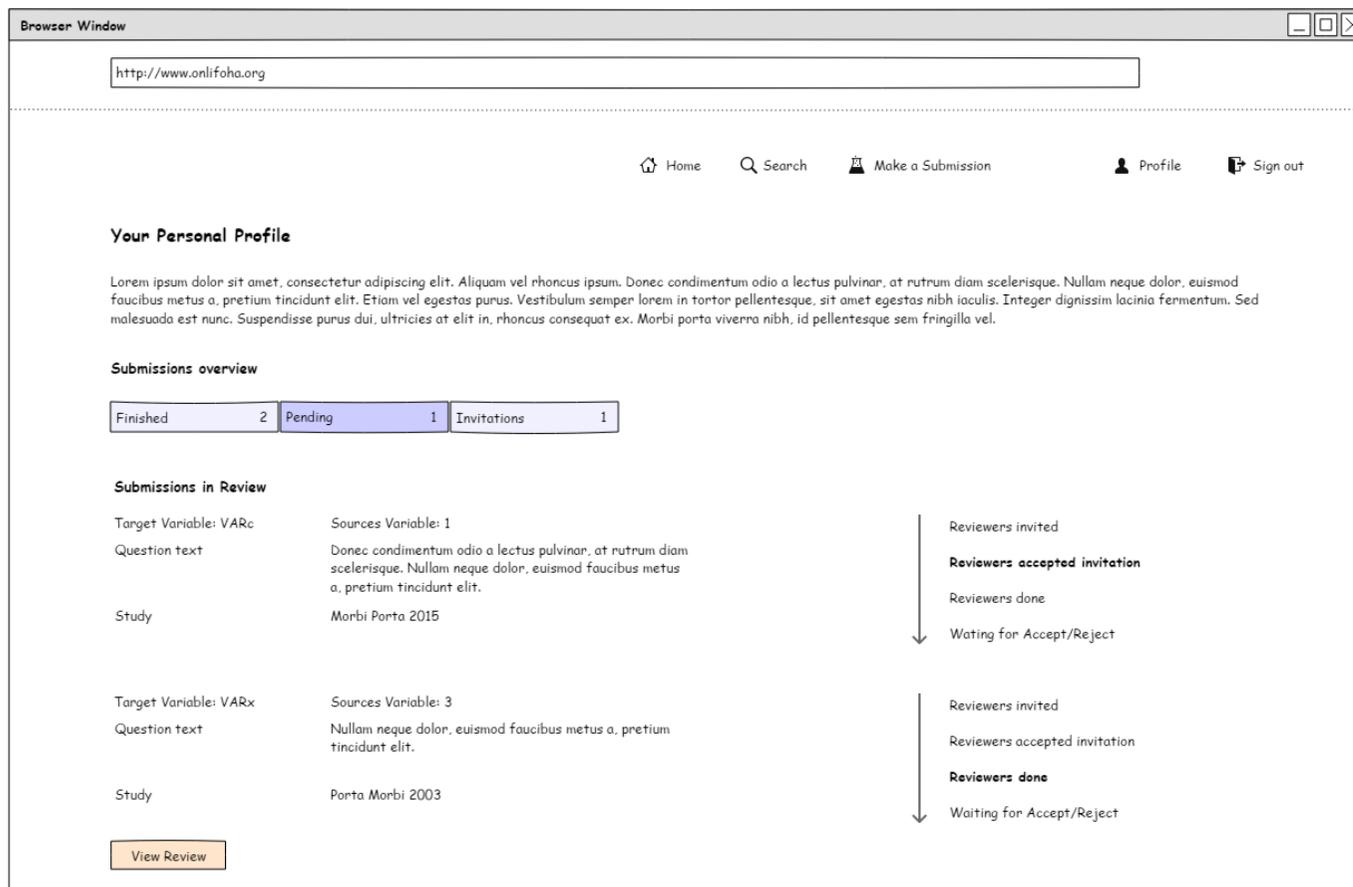


Figure 3: Place holder content for User Submissions Pending page.

This mock-up displays a list of pending submissions, meaning the peer-review process is ongoing. Metadata on the submitted data sets (target variables, question text, study) are provided as well. On the right-hand side, users can learn where their submission is in the peer-review process. The mock-up provides examples of two submissions: the first displays that reviewers were invited and invitations were accepted. The second submission indicates that the peer review process is complete and users (the submitter) may view the comments.

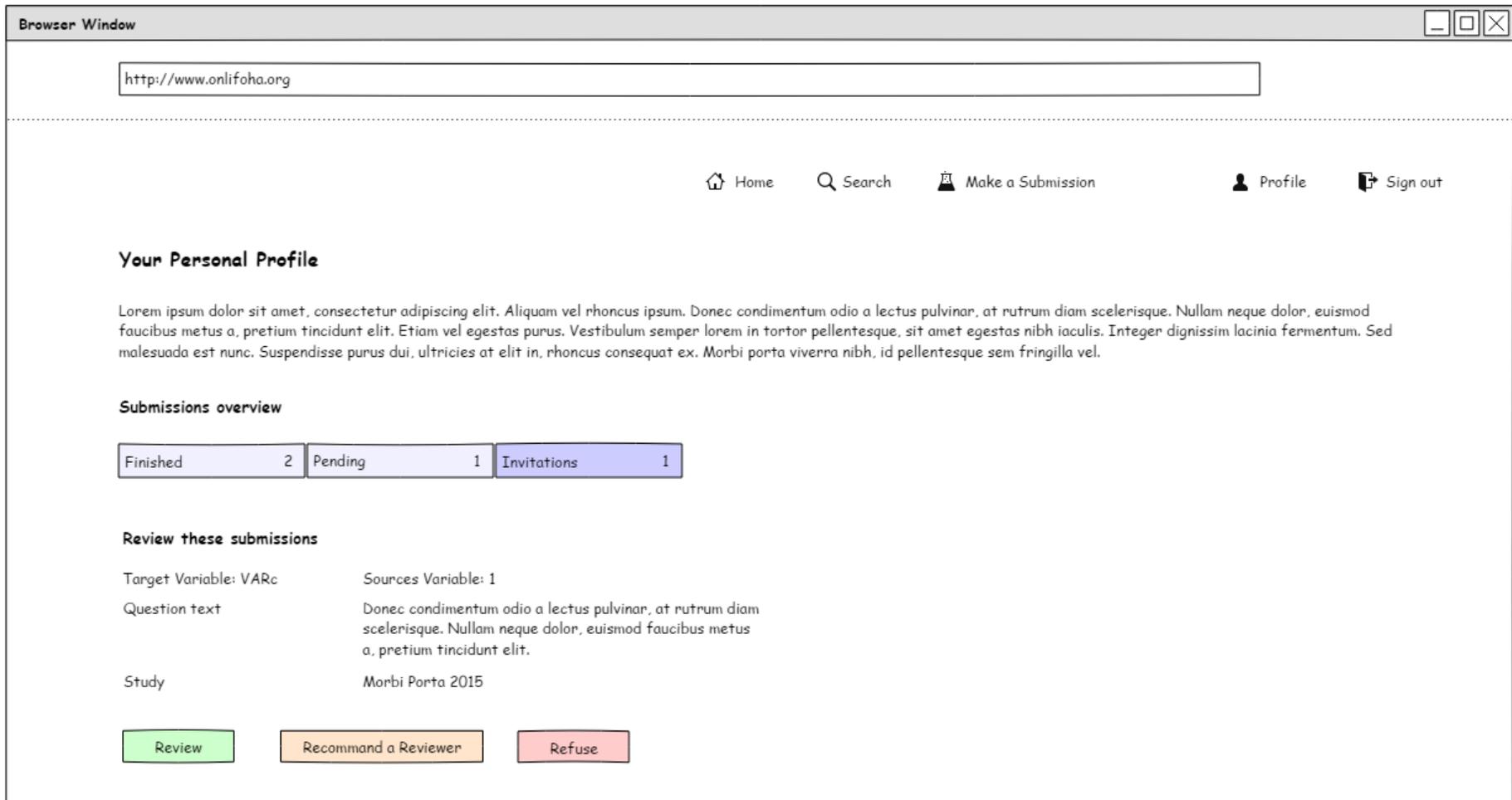


Figure 4: Place holder content for User Invitation to Review page.

This mock-up list the other submissions assigned to a potential reviewer. The user – in their role of reviewer – may now accept or reject the invitation by clicking on “Review” or “Refuse”. They may also recommend someone else.

The Workflow of the Community Peer-Review Digital Infrastructure

Before a variable recoding or harmonization project can be published and made searchable, it has to be validated. The validation of submitted harmonization projects takes place through a peer-review process. The standards for peer review of referencing formats for harmonization projects in bibliographies will be informed by the standards and guidelines in Winters and Netscher (2016). The website interface is designed to guide users through the steps required for a study's harmonization documentation to be added to the library.

Users

Below we list and define the different types of users (roles, actors) that individuals can be assigned in the planned system. These are: depositors, reviewers, and the librarian.

- The **depositor** is the user who submits harmonization data documentation for publication (listed as submitter in the workflow diagram).
- The **reviewer** reviews submissions to evaluate whether they meet the necessary and sufficient conditions for publication (see Winters and Netscher, 2016).
- The **librarian** manages accepted submission and explaining to depositors with rejected submissions how to get their documentation accepted.

One can also consider the **system** as a fourth actor in this process, since the application itself starts and stops the appropriate background processes for each step in the whole process.

The system will notify:

- the librarian when
 - a submission is received
 - a depositor recommends a reviewer
 - an invited reviewer rejects or accepts an invitation
 - when a review has been submitted (accept, R&R or reject)
- the reviewer when
 - the librarian sends out an invitation to review
- the depositor when
 - the review process has ended with advice on its acceptance status
 - an accepted project is marked completed and has been published.

Notifications will be send out via e-mail with appropriate hints on the user's profile (see screenshot above), when an input is needed. For instance, when the reviewer receives an invitation to review a submission, there will be a hint on the user's profile in the "Invitations" section.

Note: we expect most submissions will be reviewed by the project manager of the CharmStats team, however we have designed a system with capacity to bring outside experts into the process.

Background research

As a preliminary step to designing the harmonization documentation submissions workflow, the Task 4.5 team conducted research into existing online publishers of academic journals. The task manager, with experience both as a journal author and as a reviewer, took screenshots of multiple online submission workflows. The team examined the steps each had in common, noted unique features that distinguished them, and decided on which steps would form the core of the submission process.

Once the core steps of the submission process were identified, each user's role in the workflow was specified relative to the other users (see Table 1). In Figure 5, letters A thru Q are located next to points of action in the workflow. Please refer to this in combination with the description of the workflow that follows.

Table 1: Core steps identified as the basis of the workflow by user role and rights.

Actions	1	2	3	4	5	6	7	8
	Submit	Recommend a reviewer	Invite to review	Review	Submission comment	Decision comment	Recommend or reject	Accept / Reject
Depositor	x	x	-	-	-	x	-	-
Reviewer	-	-	-	x	x	-	x	-
Librarian	-	-	x	x	x	x	x	x

Submission review, acceptance and depositing workflow

Below we describe the submission review and acceptance workflow as if there were only one reviewer for the submission (see Figure 5).¹ However, in practice, there may be more than one reviewer involved in the same steps of the peer review process. Each reviewer will follow the same steps; we therefore only refer to the reviewer in the singular throughout.

Following a submission (A) the Depositor may recommend a reviewer who is familiar with or an expert on their project's harmonization process (B). If the user does not have a reviewer to recommend the submission is then sent to the librarian's user account (C1).

Recommended reviewers' details are kept in a dataset for future invitations to review on related harmonization work.

Some depositors may know a reviewer who has the expertise to evaluate their work (C2). Alternatively, large-scale or longitudinal studies already adhere to their own professional standards of reporting. We will allow managers of professional studies to 'certify' their standards of reporting using the structure of the peer review process. This will reduce the time needed to process and publish long-standing standards for reporting variable recoding

¹ The peer-review process diagram (Figure 5) provides an outline or overview of the system workflow, but should not be viewed as finalized. The workflow may be adapted based on the results of beta testing and feedback from user experience once the first version of the system becomes operational.

documentation. For instance, professional studies such as the European Values Survey will be able to self-review and approve their variable recoding and harmonization documentation on the basis that it conforms to the reporting standards and practices of their earlier publications.

If the depositor recommends a reviewer, he/she provides contact information for that reviewer (C2). The recommended reviewer may or may not already be a user in the Online Library of Harmonization's system. In the latter case, the recommended user must fill out the registration form before starting the review.

The selection of one or more experts to review a submission is made by the librarian (D). The librarian will use the website's system to send out an invitation to review the suggested reviewer (B) or assign him/herself to review the submission. If the librarian takes on the reviewer duties by inviting themselves and accepting the invitation, the system will assign them a second role on the project, that of the reviewer for the submission with no other special permissions. Thus the librarian is treated as having two distinct roles for that project from the point of view of the system.

A reviewer may refuse an invitation to review a submission (G), and the system will notify the librarian. If the reviewer accepts an invitation to review (F) they will be provided with the necessary and sufficient conditions to accept each section of a harmonization project and then are made part of the peer review workflow. After accepting an invitation the reviewer may start the review process (I).

Two further possibilities are activated when a reviewer is assigned to a project. The depositor now has the ability to withdraw their submission (J), and although we don't expect many users to do this, starting the peer review process activates this possibility. The second capacity activated is to allow the librarian to 'desk reject' the submission (H). We also don't expect many desk rejections, but the capacity is needed.

The workflow of a rejected proposal could look like this: a depositor submits a form that, for instance, fails to provide basic information needed for publication. The librarian would first inspect the submission and determine it was incomplete. S/he would then assign themselves as the reviewer, and in the role of reviewer recommend rejection and provide instructions on how to meet the requirements. The librarian would then log out of the reviewer role and log back in as the librarian to accept the rejection, and the depositor would be notified by the system of the decision.

When the review process is started, the system will track when the submission is revised by the depositor. This will help track versioning in the peer review process. In the first instance, a tag assignment version 01 (v0.1) will be created for each submission.

The reviewer may comment on the submission on a per field basis (K). Finally, the reviewer may give a recommendation of accept or reject (M). If the harmonization documentation is given a revise and resubmit, the system activates the capacity for the depositor to reply to the reviewers comments on the information fields flagged as not up to standard (L).

Since the online library is not a journal, all rejections are considered 'revise and resubmit' rather than rejection. The peer review comments by field are meant to provide the depositor with instructions that detail precisely what needs to be corrected/added/removed/cited in

order to be published. If a project is recommended for acceptance and the librarian agrees, the project is published to the system (P).

If the project is rejected, the librarian will submit the comments to the depositor with instructions on how to get the project published by responding to the reviewer's comments. The depositor is provided with the reviewer's comments on a per field basis and given the ability to comment back (L). The revisions and supporting comments are returned to the librarian. When the depositor submits the corrections a new version is assigned to the project (e.g. v0.02) and the review process begins again (O).

The peer-review process is completed once the librarian accepts a submission or a depositor withdraws their submission. If the librarian accepts a submission, the system will initiate a (publishing) follow up process where it takes care of the archiving, indexing of the dataset, and the registration of a DOI for that data set in DA|RA². The team will explore whether it is practical to provide landing pages for other forms of metadata (for instance, one for each variable) in the website testing phase. In the case of a withdrawn submission, the project and all connected data, including comments, are deleted.

Bibliography

Winters K, and Netscher S (2016) Proposed Standards for Variable Harmonization Documentation and Referencing: A Case Study Using QuickCharmStats 1.1. PLoS ONE11(2): e0147795. <https://doi.org/10.1371/journal.pone.0147795>

² 'da|ra is the registration agency for social science and economic data jointly run by GESIS and ZBW.' See more at: <https://www.da-ra.de/en/home/>.

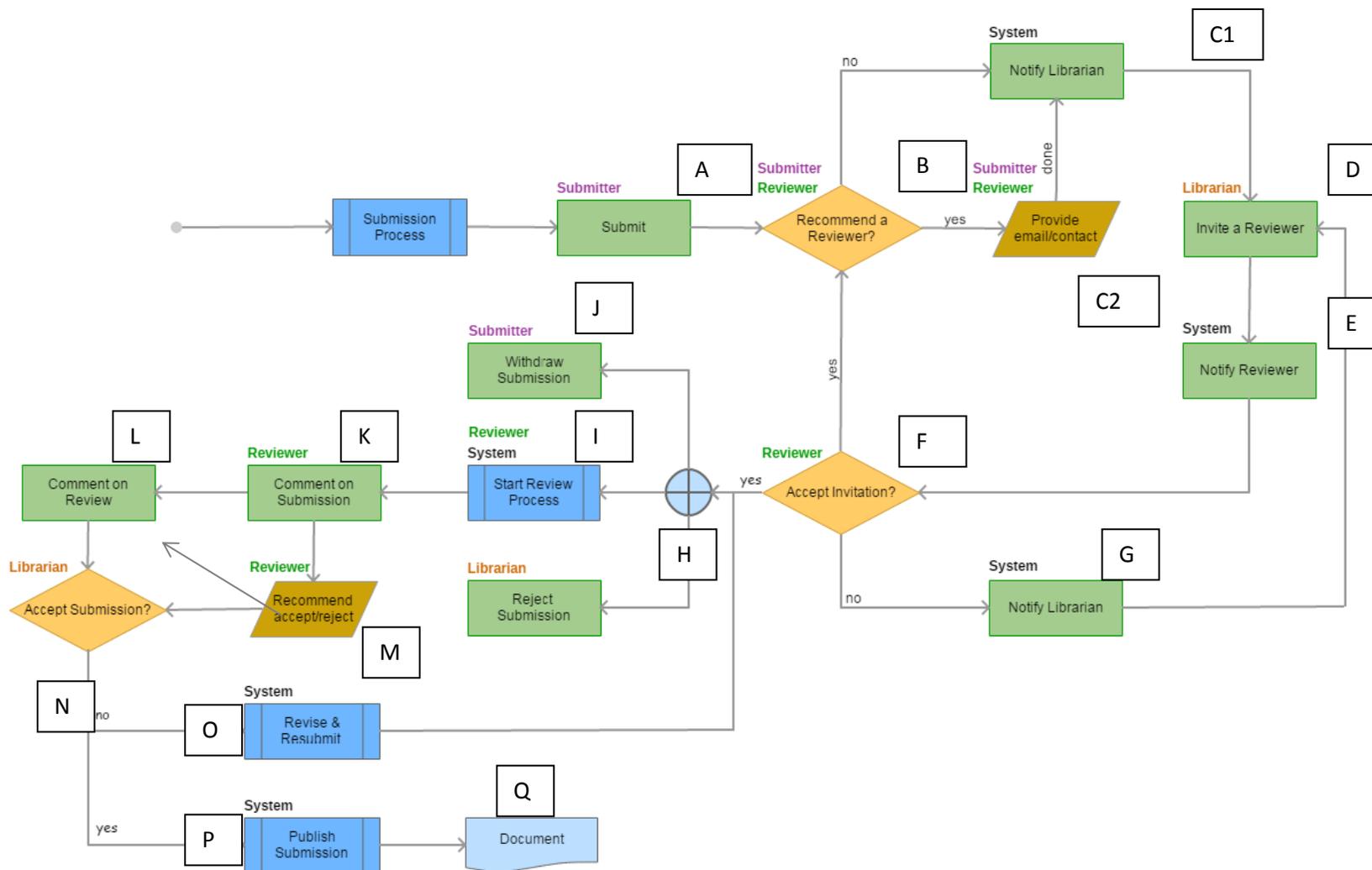


Figure 5: Overview of the submission, review, acceptance and depositing workflow.

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 654221.



