Grant Agreement 270939

ENUMERATE

ENUMERATE Data Platform

Deliverable number	*D3.3*
Dissemination level	*PU*
Delivery date	*October 2012*
Status	*Final*
Author(s)	*Jesús L. Domínguez Muriel (Digibís)*

This project is funded under the
ICT Policy Support Programme part of the
Competitiveness and Innovation Framework Programme.

Contents

1 EXECUTIVE SUMMARY 4
2 THE ENUMERATE DATA PLATFORM 4
3 PLATFORM STRUCTURE AND FUNCTIONAL DESCRIPTION 5
3.1 Repository. 5
3.2 Benchmark Tool 6
3.3 Contact form 6
3.4 Review of Visualization Tools 6
4 PLATFORM SCREENSHOTS 7

List of figures

Figure 1: ENUMERATE Data Platform Home Page 7
Figure 2: File Repository – Reports section 8
Figure 3: File Repository – Datasets section 8
Figure 4: File Repository – Other files 9
Figure 5: File Repository – Search results 9
Figure 6: Benchmark tool – data entry 10
Figure 7: Benchmark report 10
Figure 8: Contact form – Participate in the next survey 11
Figure 9: Contact form - Collaborate 11
Figure 10: Review of visualization software 12

1 Executive summary

This document describes the ENUMERATE Data Platform, a set of web pages developed for the ENUMERATE Thematic Network in order to make available the results of the project.

This deliverable describes the platform, its structure and main functionality. Screenshots from the launch version have been included (as on October 31st 2012).

2 The ENUMERATE Data Platform

The ENUMERATE Data Platform is available at: **http://dataplatform.enumerate.eu**

The Data Platform is a website where the raw data obtained in the ENUMERATE surveys, the reports and the digests can be consulted and downloaded. It also provides an overview of some selected visualisation tools for enhancing the accessibility of the data, and a benchmark tool that enables institutions to see how they are performing compared to other institutions in the same country or of the same type.

The main goals are to:

- Provide a 'one-stop-shop' to easily search, retrieve and download all the statistical products of the project, including reports and the underlying datasets;
- Encourage the reuse of these datasets;
- Provide information, at the European level, of the state of digitisation, preservation and online access to cultural heritage institutions. This is so that they will get involved in the project and to help improve their participation in future surveys.

The Data Platform will be updated continuously during the rest of the project. The partner responsible for the website content is DIGIBÍS. All the other partners can access the website and publish new files. The datasets stored in the Data Platform have been produced according to international open standards on the structuring and processing of statistical data.

3 Platform structure and functional description

The website has been built using the widely used Joomla Content Management System, including several commercial components like JoomDoc and Chronoforms. The Benchmark tool has been specifically developed for this project at DIGIBÍS, using Javascript and some Ruby scripts to process the survey data.

The website is being actively monitored by DIGIBÍS, who will track uptime, use and response time. Usage statistics are being collected using a Google Analytics profile with specific goals defined according to the website objectives.

Apart from the Home page (Figure 1), which includes a short description of the project, and the Data Platform functionality, the website has four main sections: Repository, Benchmark Tool, Review of Visualization Tools and Contact form.

3.1 Repository.

The contents of the file repository have been divided in three subsections:

- **Reports** – Includes all the statistical reports produced by the ENUMERATE project, and digests presenting the main outcomes of the ENUMERATE Surveys (Figure 2)
- **Datasets** – Files containing the data obtained in the Core and Thematic Surveys in a format suitable for statistical analysis (Figure 3). These datasets have been cleaned and normalised, and have been anonymised by removing all the fields such as institution name, and e-mail address, that could allow the identification of the survey participants.

  The full datasets are available in SPSS and Excel format; other files with processed datasets, for example, the dataset used to generate the benchmark tool, will be available in Excel or CSV format. We have also produced files in SDMX format but have found that available open source SDMX software is not currently mature enough to easily handle these files, so they are not currently published.
- **Other files** – Files that are neither reports nor datafiles, but can be useful for institutions interested in the survey, such as the survey questionnaires translated by the National Coordinators and terminology lists (Figure 4).

User profiles and control access

Most of the content of the repository is publicly available, but the website has been designed so that detailed permissions can be established to control access to specific sections or files of the repository. Several profiles have been created to control who can upload new files, and to protect certain datasets for which we would like to track, at least initially, how many people is interested in them. For those datasets, the web shows a message asking interested users to register in the system before downloading the files.

The profiles initially defined are:

- Administrator;
- ENUMERATE Partners;
- National Coordinators;
- Registered users;
- Public.

Data Platform Search

The repository includes a function to quickly search, accessible from a search box in the site menu. It performs a full-text search and shows all the files in the repository accessible to the current user whose description matches the terms entered. The searched terms are highlighted in the search results (Figure 5).

3.2 Benchmark Tool

The Benchmark Tool is an easy to use tool developed specifically for this project. It allows any cultural instititution, participating in the surveys or not, to compare the status of its digitisation activities against the mean and median values found in the core surveys. It produces a simple report showing the results in a visual way.

Data entry

Institutions are asked for the data required to produce the benchmark report (Figure 6): institution name; country; type of institution; and numeric values of the variables to be benchmarked, such as the part of their collections to be digitised or the part already digitised. This data is not stored in any database to avoid privacy issues, only used in the moment to make the comparisons and produce the results.

As all users of the Benchmark Tool are asked for this information, institutions that did not participate in the ENUMERATE surveys can also use the benchmark.

Report

The benchmark report can be seen in Figure 7. The introduced values are compared with the whole set of European responses, with the values obtained in their countries, and with the values obtained from institutions of their same type.

As suggested by one of the ENUMERATE partners, we show not only the mean values of the responses collected in the surveys, but also the median values in order to give a better idea of the shape of the distribution. The number of responses obtained in the selected country and institution type is also shown, so that the benchmark uses can evaluate the representativity of the results they are being compared with – for small countries and/or very specific types the number of survey responses and therefore the representativity of the data is low.

The report includes a button for sharing the results via Twitter, so that institutions can easily share their results with their communities. It is also a way to advertise in a viral way the existence of the tool and promote its use.

Finally, the report includes also links to the ENUMERATE newsletter subscription page and to a simple form where institutions willing to participate in the next ENUMERATE survey can register their contact data so that we can contact them before the survey (Figure 8).

3.3 Contact form

This section of the website contains a simple form where institutions can contact the Data Platform administrator to request an account so that they can access a protected data file, to send a suggestion, or to inform us of some use of our datasets so that we can register and/or link to it. (Figure 9).

The data registered in this form, or in the "Participate in the next survey" form is only used to send an e-mail to the Data Platform administrator and is not recorded in any way, as explained in detail in the Privacy Policy published in the website – linked in each page's footer. In both forms we have used the reCaptcha system to block automatic bots from introducing false data.

3.4 Review of Visualization Tools

In order to facilitate the reuse and exploitation of the ENUMERATE datasets, a separate section of the Data Platform website has been prepared with pointers and brief descriptions of a range of free visualization tools, that can be used for the analysis of the datasets stored in the repository. (Figure 10)

4 Platform screenshots

Figure 1: ENUMERATE Data Platform Home Page

Figure 2: File Repository – Reports section

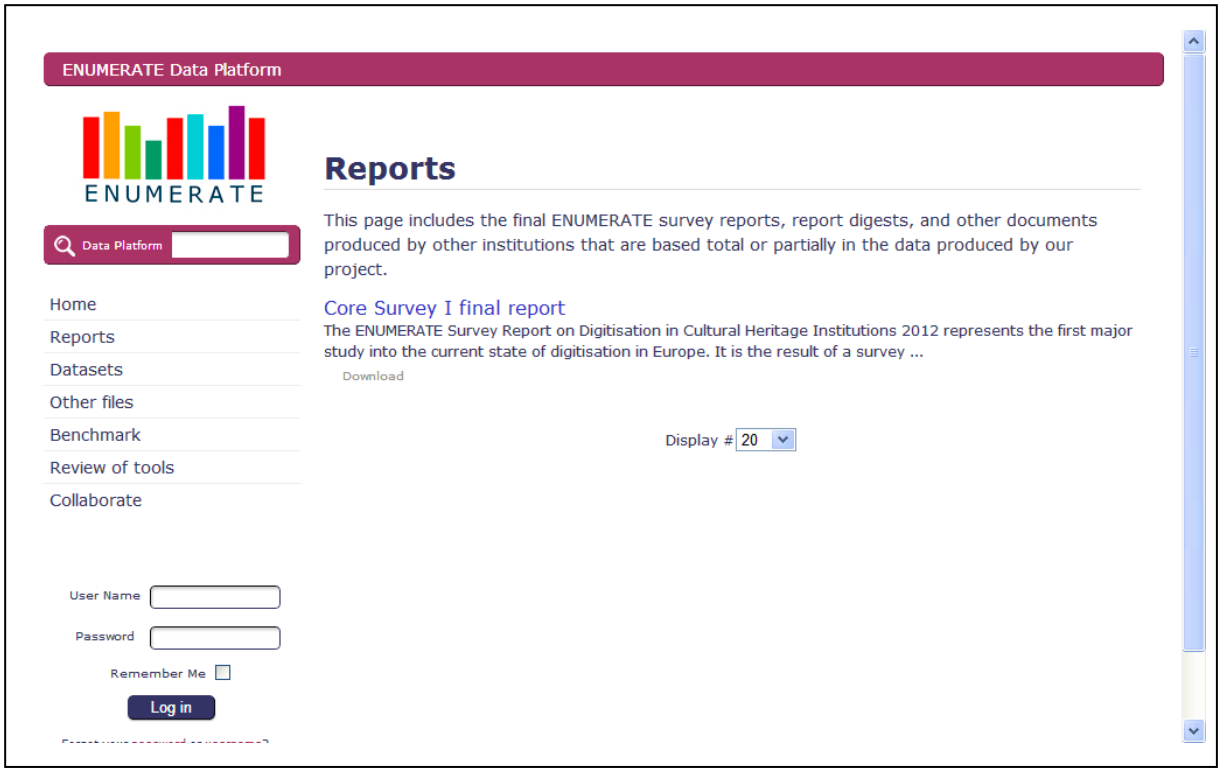

Figure 3: File Repository – Datasets section

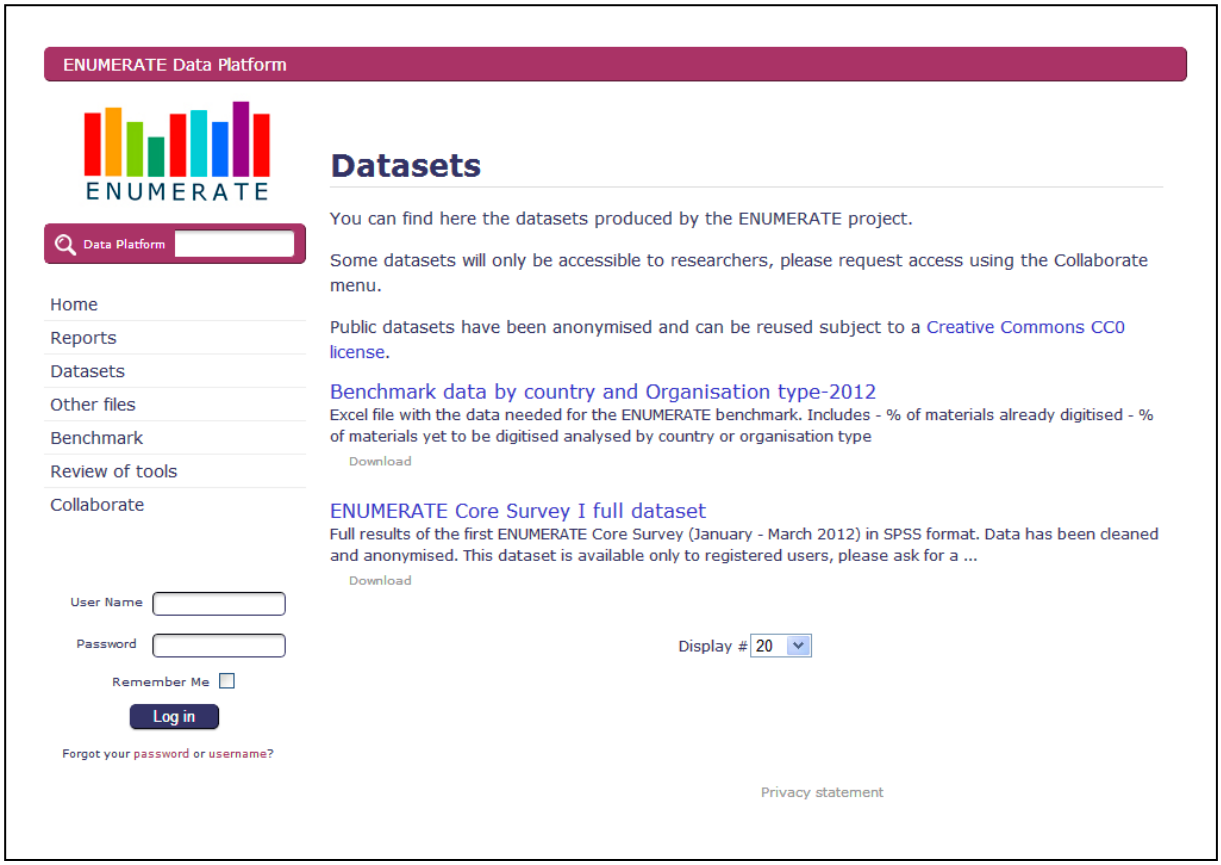

Figure 4: File Repository – Other files

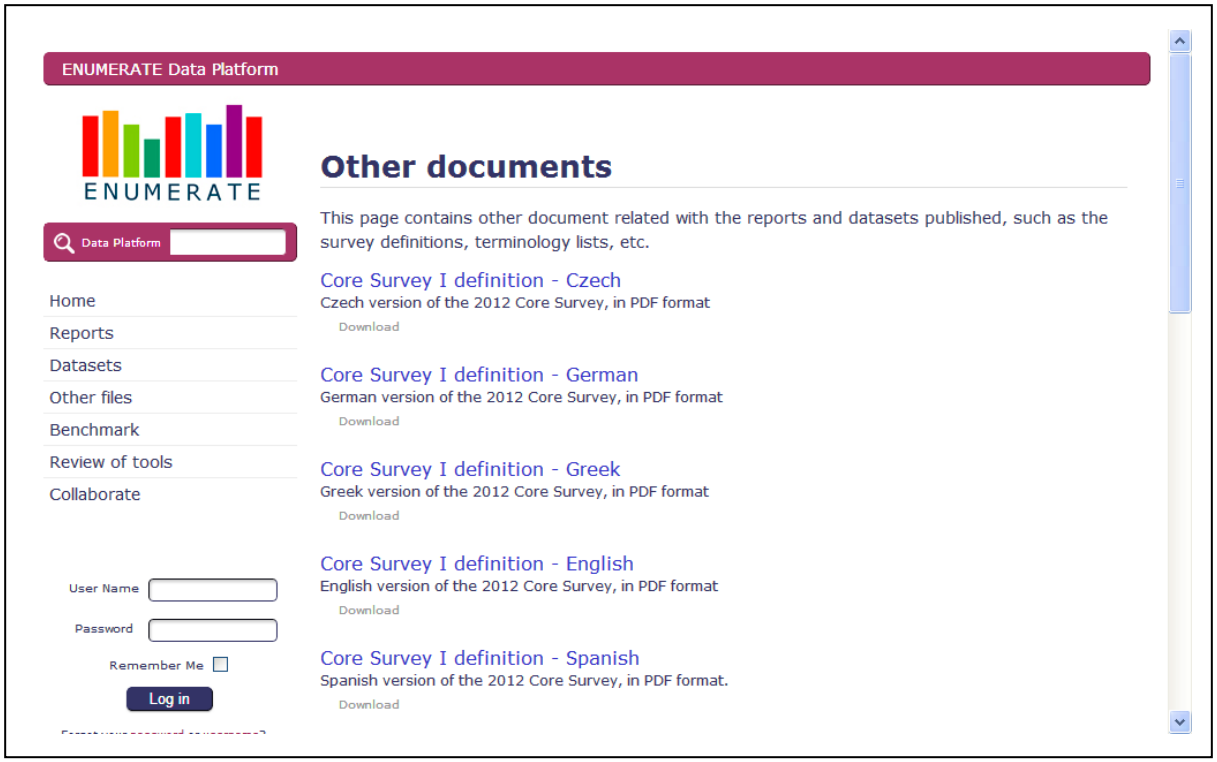

Figure 5: File Repository – Search results

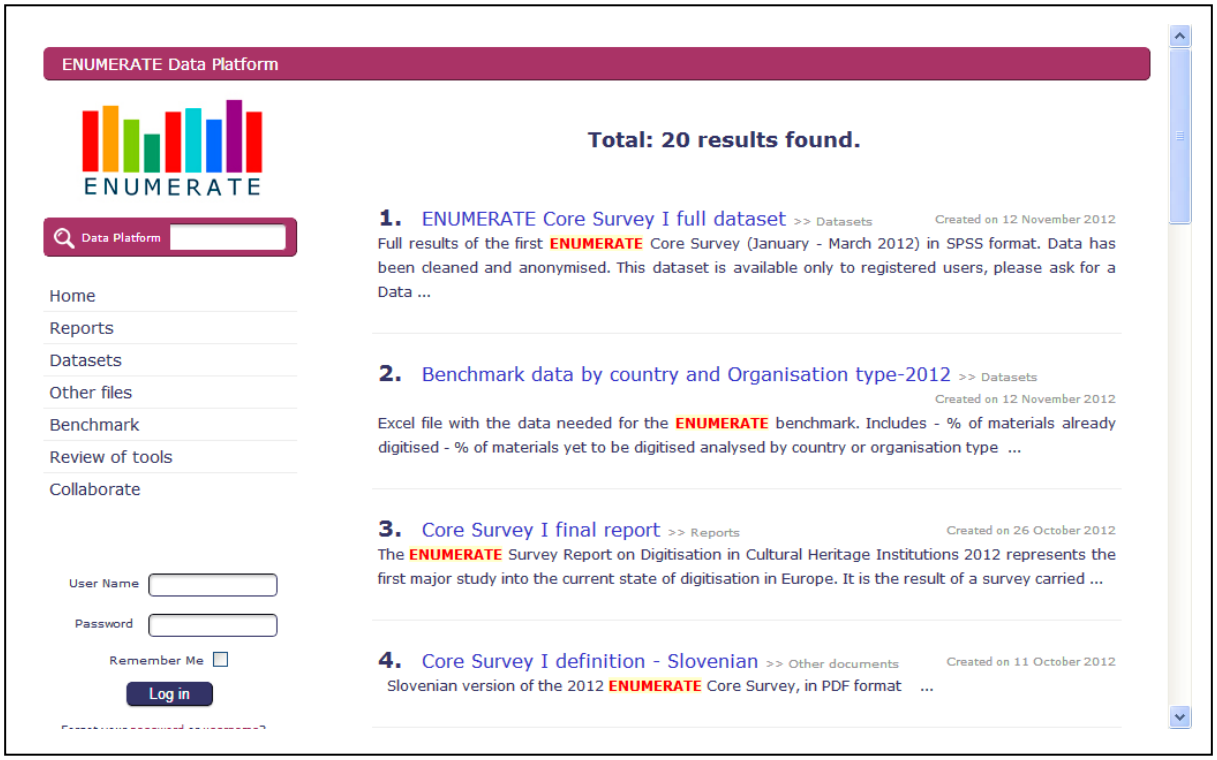

Figure 6: Benchmark tool – data entry

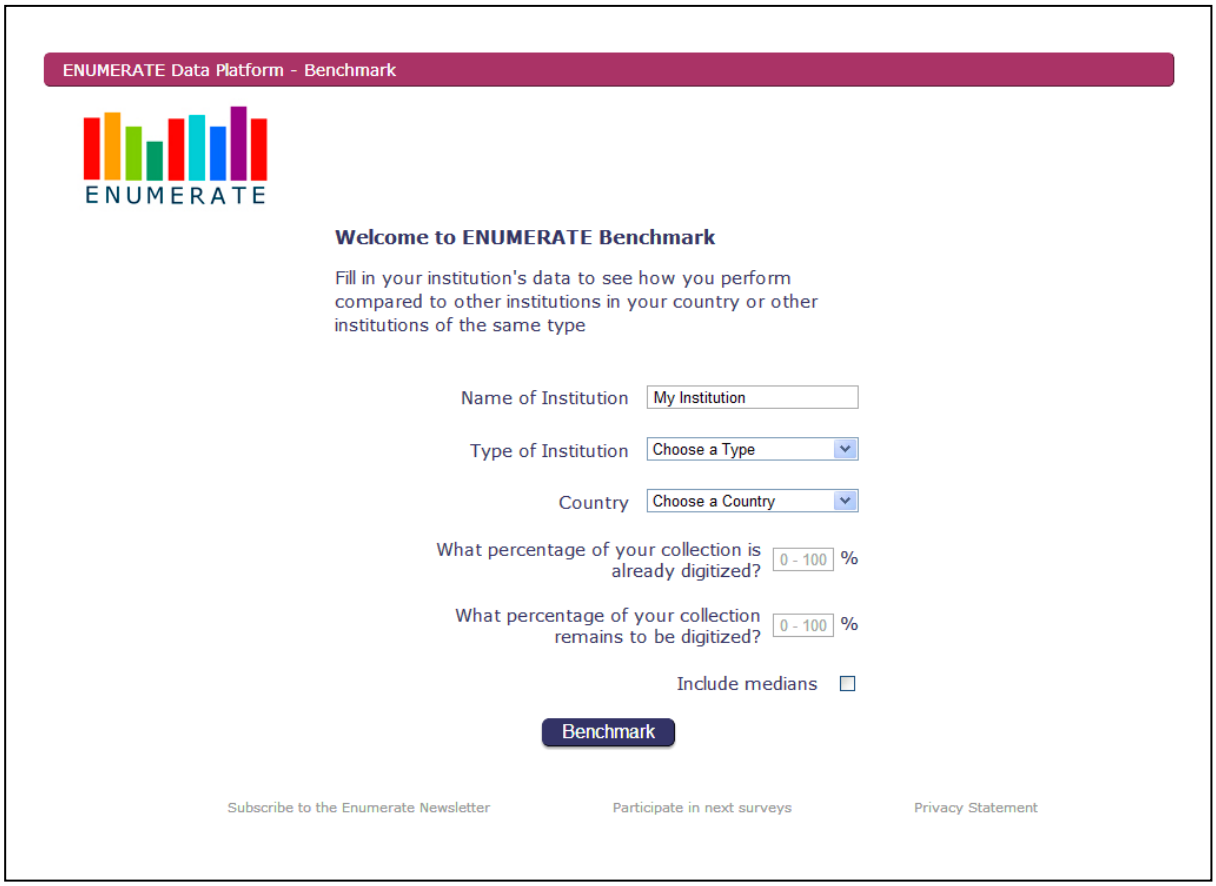

Figure 7: Benchmark report

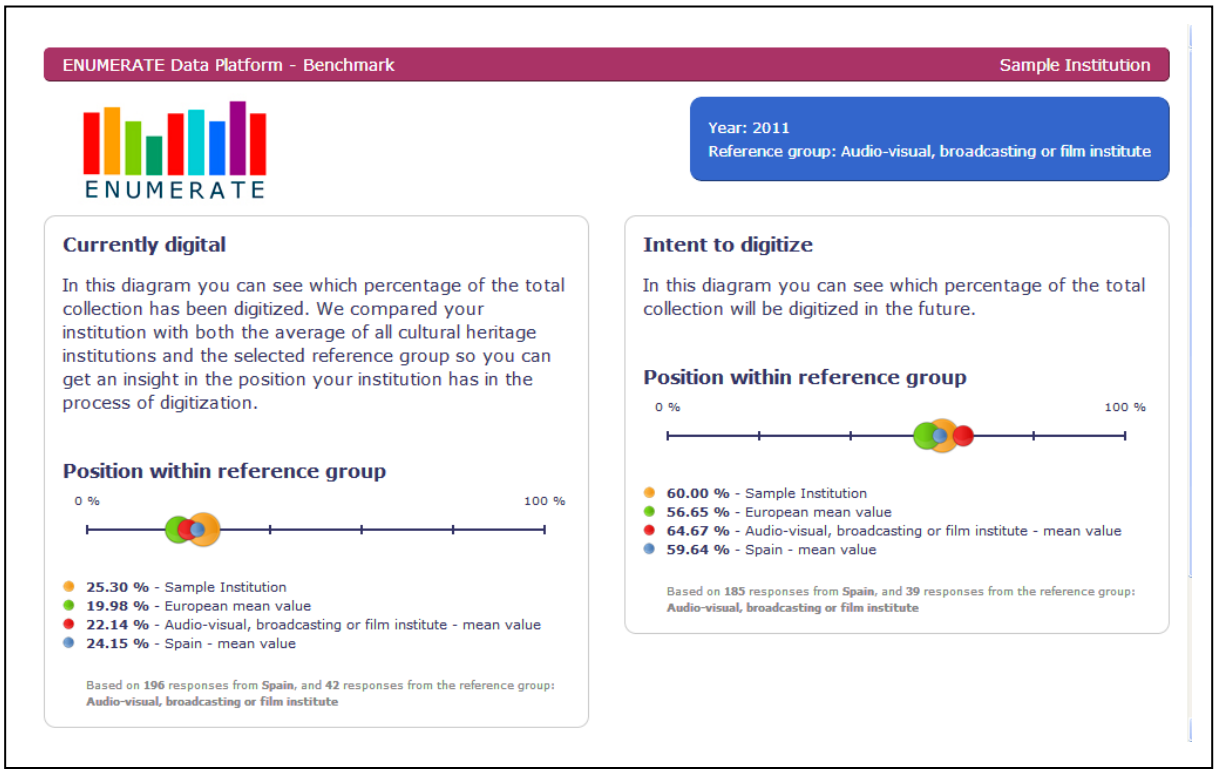

Figure 8: Contact form – Participate in the next survey

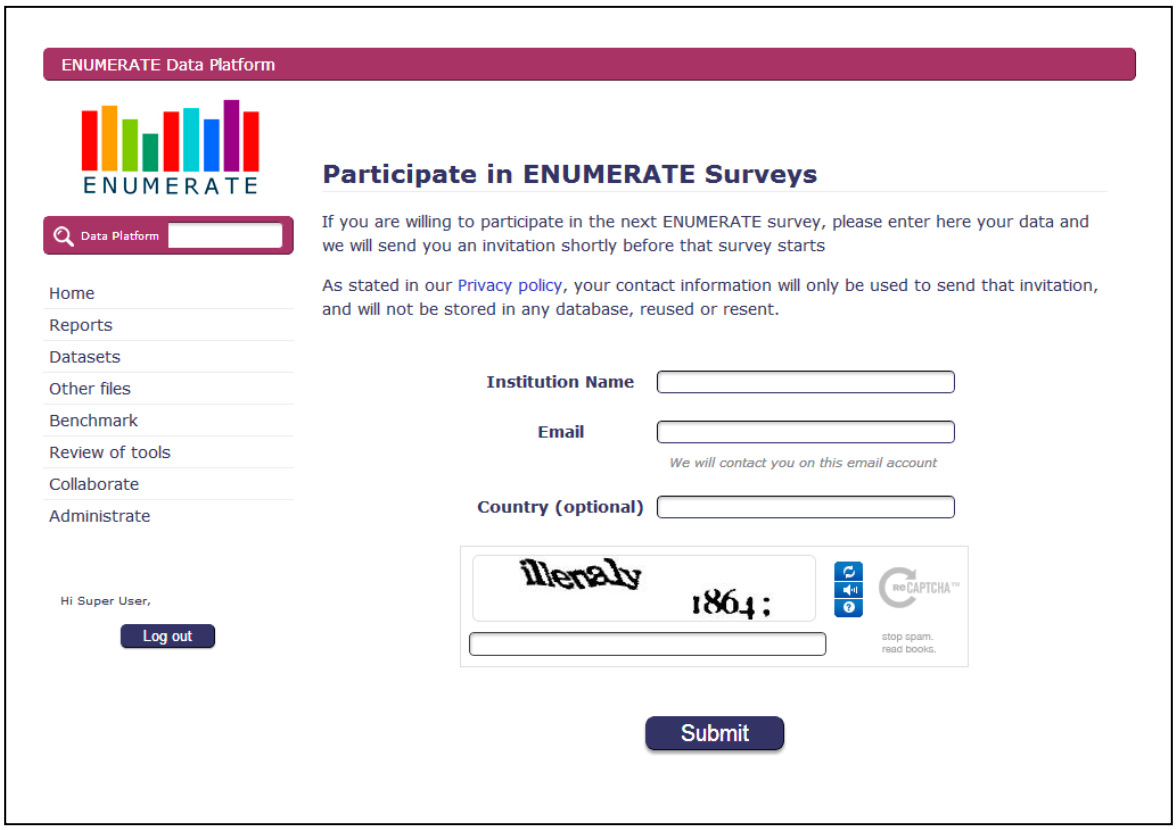

Figure 9: Contact form - Collaborate

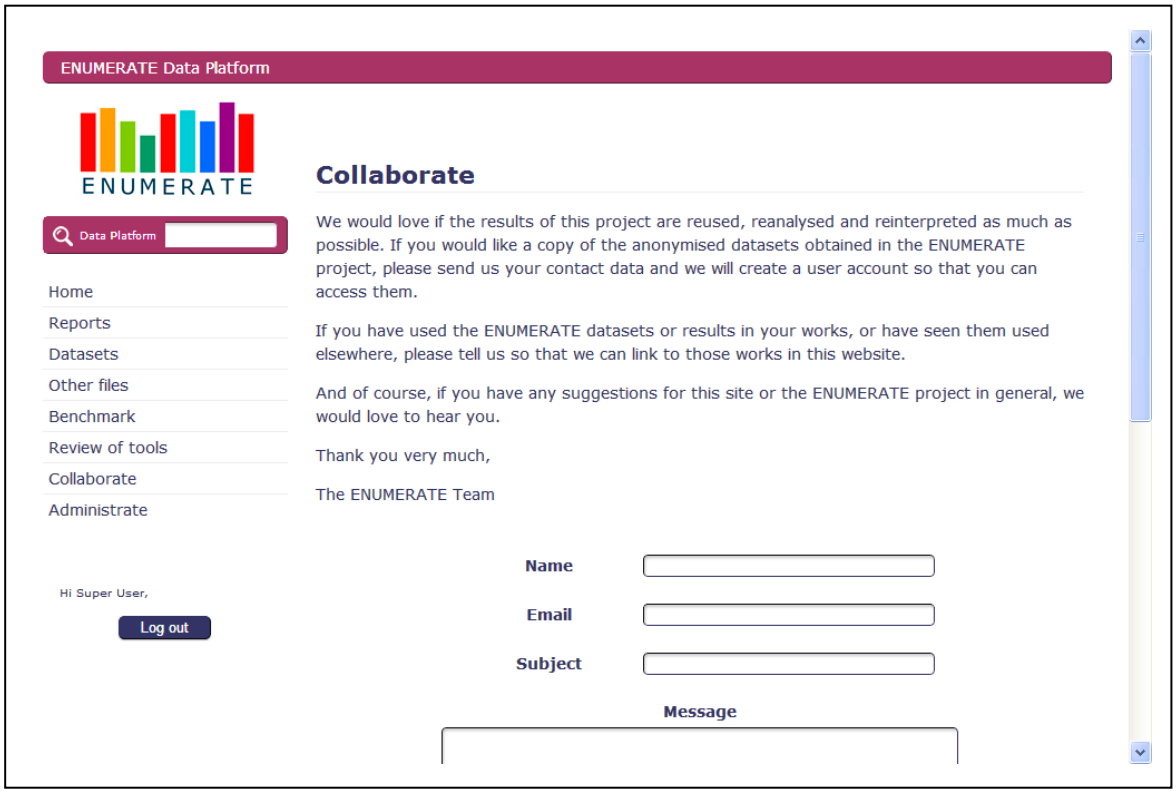
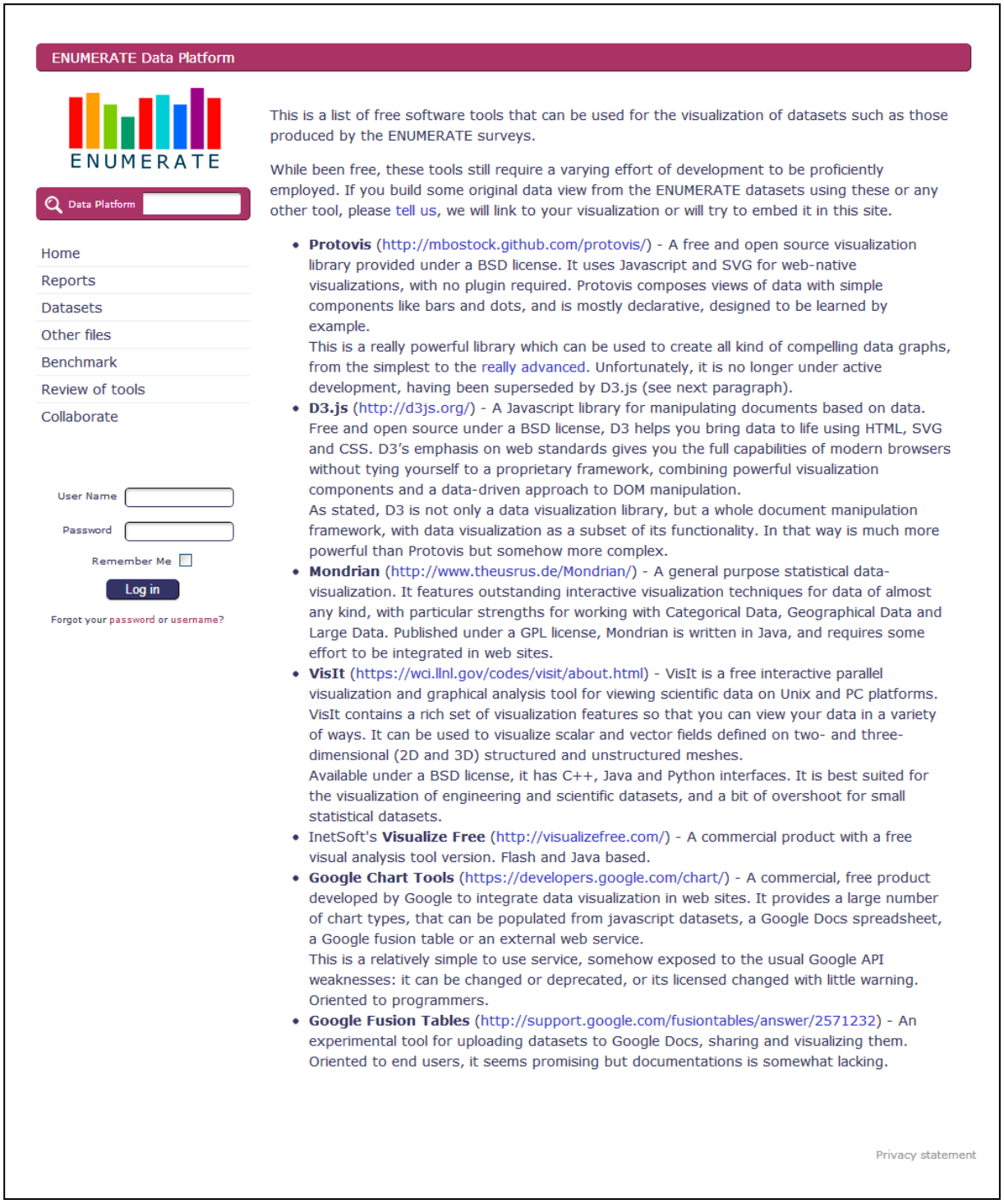
Figure 10: Review of visualization software