



Grant Agreement 270939

ENUMERATE

ENUMERATE Data Platform

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



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	<i>Repository</i>	5
3.2	<i>Benchmark Tool</i>	6
3.3	<i>Contact form</i>	6
3.4	<i>Review of Visualization Tools</i>	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://datapatform.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 institution, 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

ENUMERATE Data Platform



- Home
- Reports
- Datasets
- Other files
- Benchmark
- Review of tools
- Collaborate

Remember Me

Forgot your password or username?

ENUMERATE Data Platform

ENUMERATE is an EC-funded project, led by Collections Trust in the UK. The primary objective of ENUMERATE is to create a reliable baseline of **statistical data about digitization, digital preservation and online access to cultural heritage in Europe**. Currently, statistical data on Europe's digital heritage is tentative and scattered at best. For the European Commission and many of the agencies and actors in the field of culture there is no consistent evidence base for making strategic decisions on investments in digitisation. ENUMERATE will bring about major improvements in the quality and availability of intelligence about digital heritage.

The **ENUMERATE Data Platform** is a repository where the raw datasets, the reports and the digests produced by ENUMERATE can be consulted and downloaded. It also provides a benchmark tool that institutions can use to see how they are performing compared to other institutions in the same country or of the same type.

A consortium of ten partners is at the heart of the ENUMERATE 'Thematic Network'. Together they will initiate a Europe-wide community of practice to share statistical data and knowledge on the progress of digitisation. This will be achieved by a **multi-annual programme of coordinated surveys**. There will be wide-scale harmonized statistical data-gathering and more in-depth surveying of digitization activities by European cultural heritage institutions. All activities of ENUMERATE start from the principle that **heritage institutions will receive useful information** in return for sharing their data. **Results will be published here**. Raw and summary data can be viewed and collated by interested parties.

ENUMERATE EC-funded project partners:

- Collections Trust, UK (project coordinator);
- Digitaal Erfgoed Nederland (DEN), Netherlands;
- Stiftung Preußischer Kulturbesitz, Germany;
- Digibis, Spain;
- FARO Vlaams Steunpunt voor Cultureel Erfgoed, Belgium;
- Ministère de la Culture et de la Communication, France;
- Österreichische Nationalbibliothek, Austria;
- Narodna in univerzitetna knjižnica (National and University Library), Slovenia;
- Országos Széchényi Könyvtár (OSZK; National Széchényi Library), Hungary;
- The European Library (hosted by the Koninklijke Bibliotheek, Netherlands).

ENUMERATE is funded by the ICT Policy Support Programme of the European Commission and runs for three years, starting February 2011.



Privacy statement

Figure 2: File Repository – Reports section

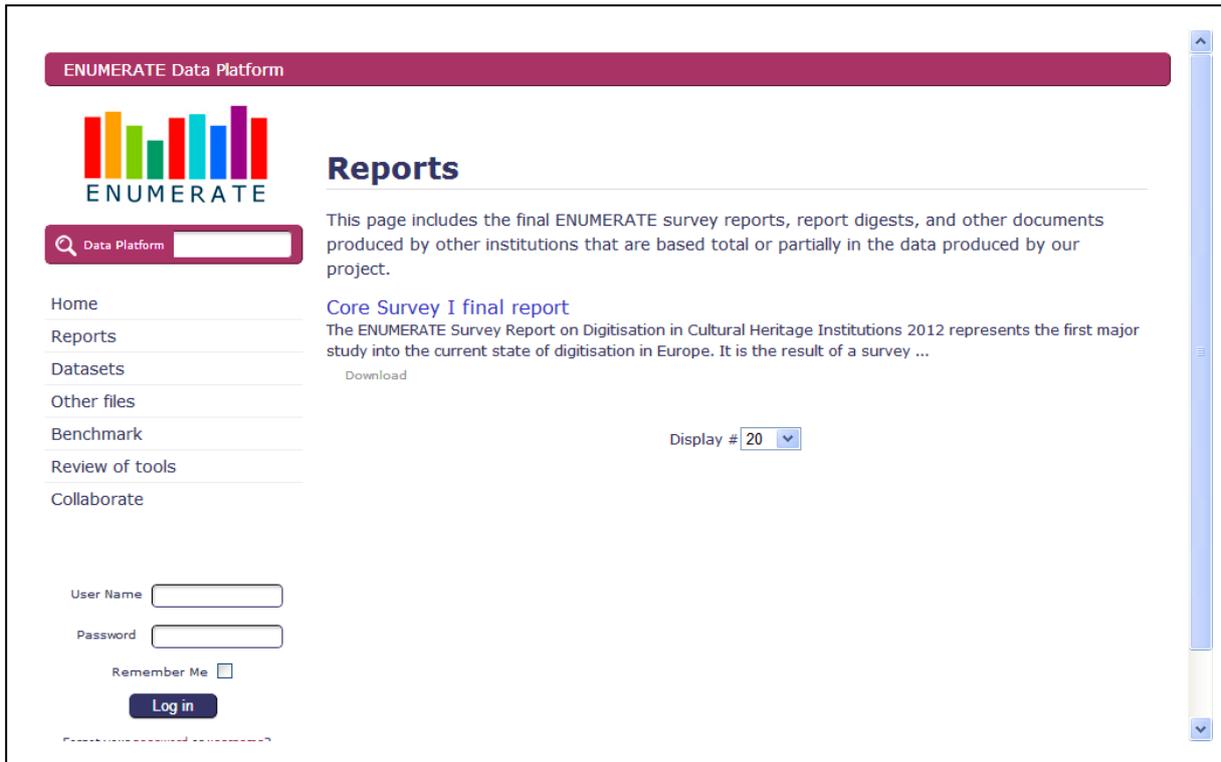


Figure 3: File Repository – Datasets section

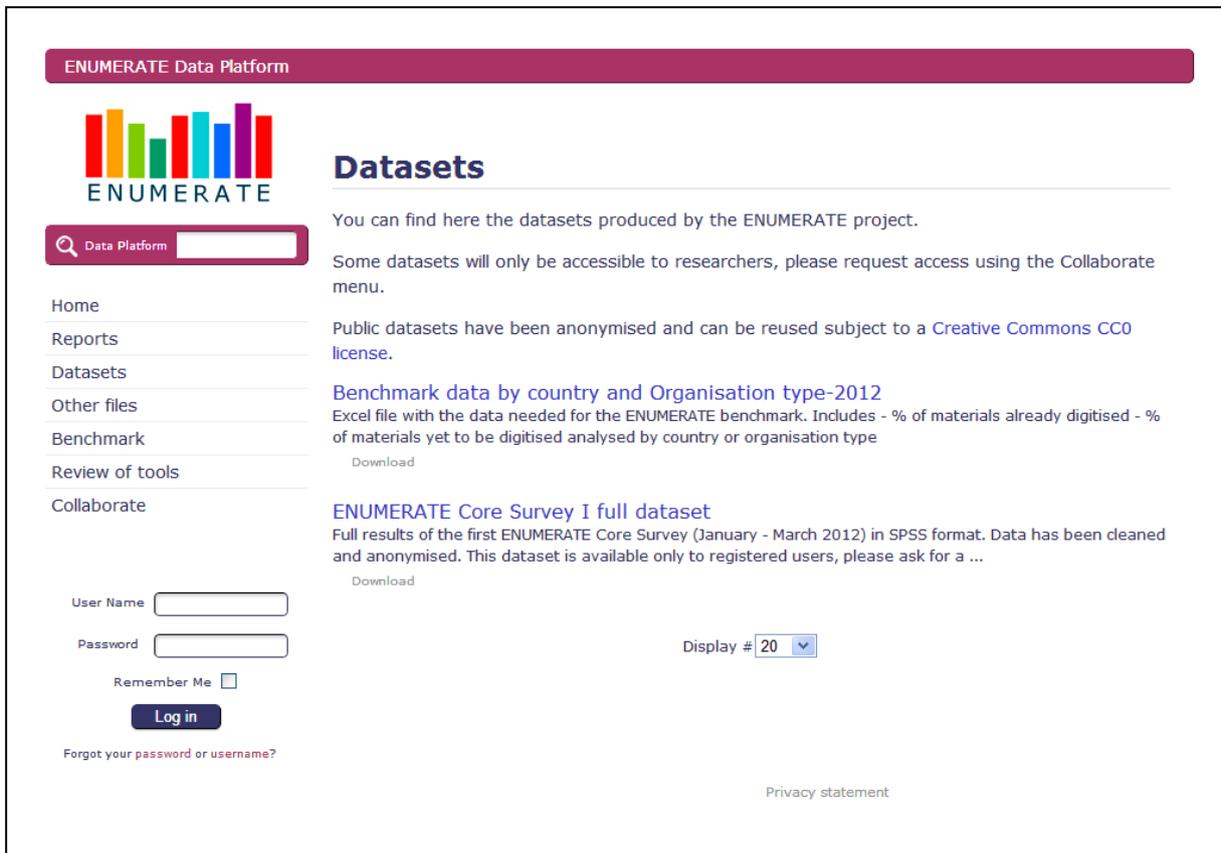


Figure 4: File Repository – Other files

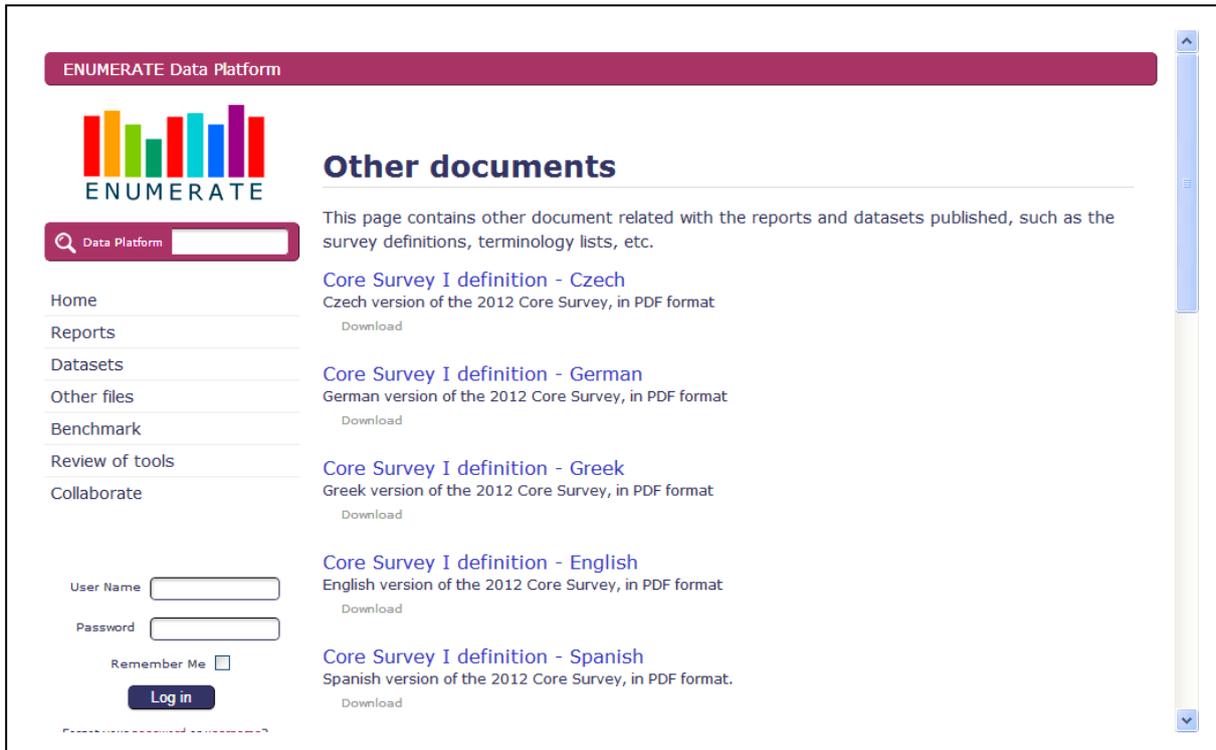


Figure 5: File Repository – Search results

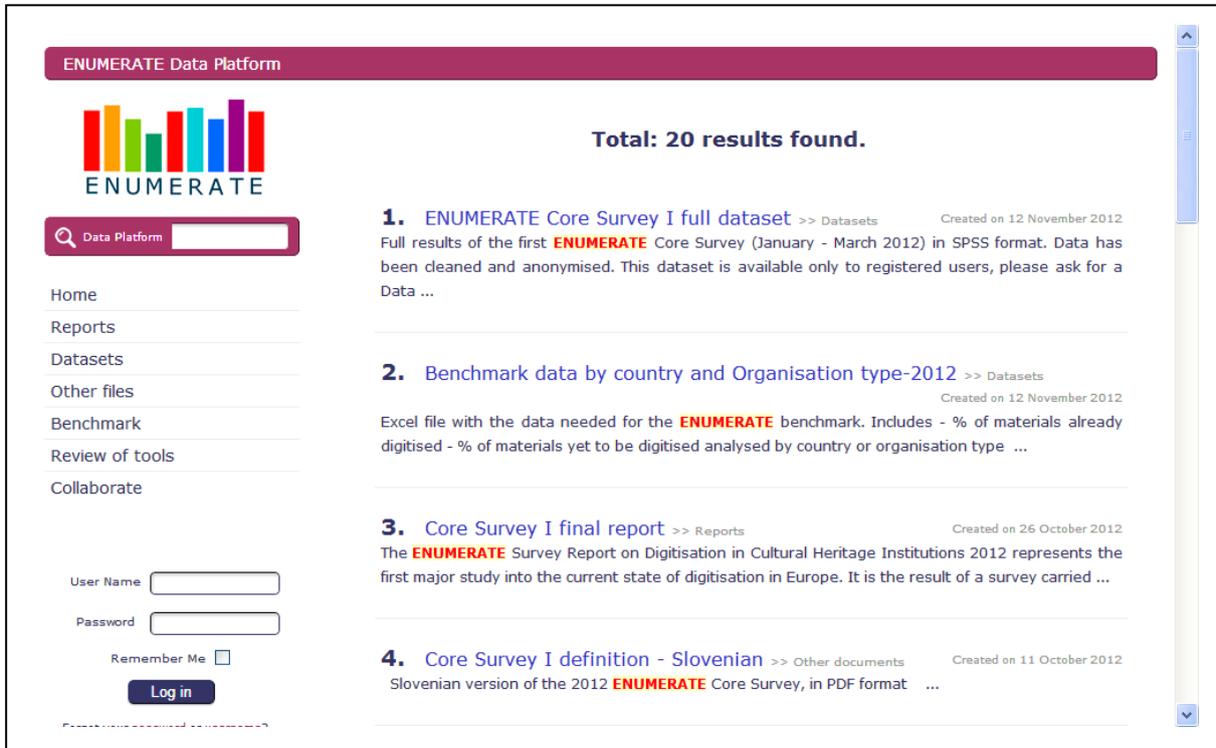


Figure 6: Benchmark tool – data entry

ENUMERATE Data Platform - Benchmark



Welcome to ENUMERATE Benchmark

Fill in your institution's data to see how you perform compared to other institutions in your country or other institutions of the same type

Name of Institution

Type of Institution

Country

What percentage of your collection is already digitized? %

What percentage of your collection remains to be digitized? %

Include medians

[Subscribe to the Enumerate Newsletter](#)
[Participate in next surveys](#)
[Privacy Statement](#)

Figure 7: Benchmark report

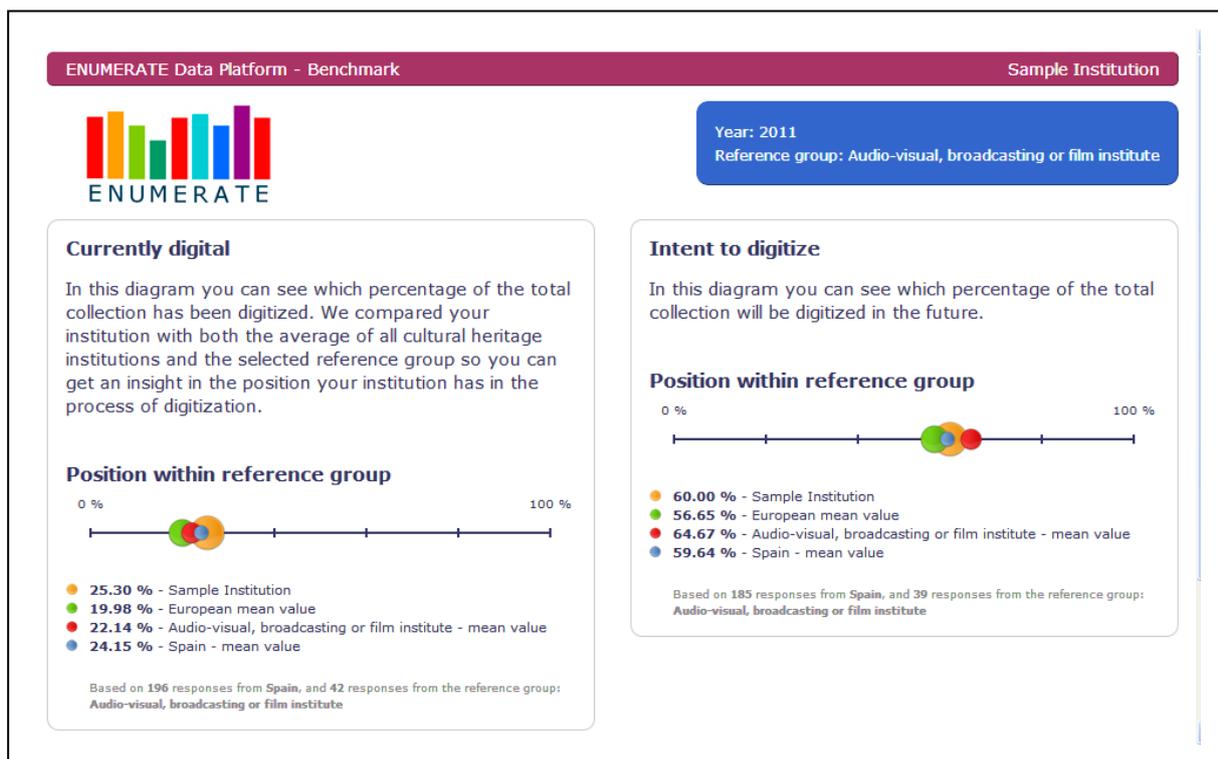


Figure 8: Contact form – Participate in the next survey

ENUMERATE Data Platform

ENUMERATE

Participate in ENUMERATE Surveys

If you are willing to participate in the next ENUMERATE survey, please enter here your data and we will send you an invitation shortly before that survey starts

As stated in our [Privacy policy](#), your contact information will only be used to send that invitation, and will not be stored in any database, reused or resent.

Home
Reports
Datasets
Other files
Benchmark
Review of tools
Collaborate
Administrative

Hi Super User,
[Log out](#)

Institution Name

Email
We will contact you on this email account

Country (optional)

literaly 1864;

reCAPTCHA™
stop spam. read books.

[Submit](#)

Figure 9: Contact form - Collaborate

ENUMERATE Data Platform

ENUMERATE

Collaborate

We would love if the results of this project are reused, reanalysed and reinterpreted as much as possible. If you would like a copy of the anonymised datasets obtained in the ENUMERATE project, please send us your contact data and we will create a user account so that you can access them.

If you have used the ENUMERATE datasets or results in your works, or have seen them used elsewhere, please tell us so that we can link to those works in this website.

And of course, if you have any suggestions for this site or the ENUMERATE project in general, we would love to hear you.

Thank you very much,
The ENUMERATE Team

Home
Reports
Datasets
Other files
Benchmark
Review of tools
Collaborate
Administrative

Hi Super User,
[Log out](#)

Name

Email

Subject

Message

Figure 10: Review of visualization software

ENUMERATE Data Platform



This is a list of free software tools that can be used for the visualization of datasets such as those produced by the ENUMERATE surveys.

While been free, these tools still require a varying effort of development to be proficiently employed. If you build some original data view from the ENUMERATE datasets using these or any other tool, please [tell us](#), we will link to your visualization or will try to embed it in this site.

Q

Home

Reports

Datasets

Other files

Benchmark

Review of tools

Collaborate

- **Protovis** (<http://mbostock.github.com/protovis/>) - A free and open source visualization library provided under a BSD license. It uses Javascript and SVG for web-native visualizations, with no plugin required. Protovis composes views of data with simple components like bars and dots, and is mostly declarative, designed to be learned by example.
This is a really powerful library which can be used to create all kind of compelling data graphs, from the simplest to the [really advanced](#). Unfortunately, it is no longer under active development, having been superseded by D3.js (see next paragraph).
- **D3.js** (<http://d3js.org/>) - A Javascript library for manipulating documents based on data. Free and open source under a BSD license, D3 helps you bring data to life using HTML, SVG and CSS. D3's emphasis on web standards gives you the full capabilities of modern browsers without tying yourself to a proprietary framework, combining powerful visualization components and a data-driven approach to DOM manipulation.
As stated, D3 is not only a data visualization library, but a whole document manipulation framework, with data visualization as a subset of its functionality. In that way is much more powerful than Protovis but somehow more complex.
- **Mondrian** (<http://www.theuser.de/Mondrian/>) - A general purpose statistical data-visualization. It features outstanding interactive visualization techniques for data of almost any kind, with particular strengths for working with Categorical Data, Geographical Data and Large Data. Published under a GPL license, Mondrian is written in Java, and requires some effort to be integrated in web sites.
- **VisIt** (<https://wci.llnl.gov/codes/visit/about.html>) - VisIt is a free interactive parallel visualization and graphical analysis tool for viewing scientific data on Unix and PC platforms. VisIt contains a rich set of visualization features so that you can view your data in a variety of ways. It can be used to visualize scalar and vector fields defined on two- and three-dimensional (2D and 3D) structured and unstructured meshes.
Available under a BSD license, it has C++, Java and Python interfaces. It is best suited for the visualization of engineering and scientific datasets, and a bit of overshoot for small statistical datasets.
- InetSoft's **Visualize Free** (<http://visualizefree.com/>) - A commercial product with a free visual analysis tool version. Flash and Java based.
- **Google Chart Tools** (<https://developers.google.com/chart/>) - A commercial, free product developed by Google to integrate data visualization in web sites. It provides a large number of chart types, that can be populated from javascript datasets, a Google Docs spreadsheet, a Google fusion table or an external web service.
This is a relatively simple to use service, somehow exposed to the usual Google API weaknesses: it can be changed or deprecated, or its licensed changed with little warning. Oriented to programmers.
- **Google Fusion Tables** (<http://support.google.com/fusiontables/answer/2571232>) - An experimental tool for uploading datasets to Google Docs, sharing and visualizing them. Oriented to end users, it seems promising but documentations is somewhat lacking.

User Name

Password

Remember Me

[Forgot your password or username?](#)

[Privacy statement](#)