



# DELIVERABLE

**Project Acronym:** DCH-RP  
**Grant Agreement number:** 312274  
**Project Title:** Digital Cultural Heritage Roadmap for Preservation – Open Science Infrastructure for DCH in 2020

---

## D5.2 Upgraded eCulture Science Gateway

**Revision:** version 1.2

---

**Authors:**

**Roberto Barbera (INFN)**  
**Antonio Calanducci (INFN)**

**Reviewers:**

**Maciej Brzezniak (PSNC)**

Project co-funded by the European Commission within the ICT Policy Support Programme		
Dissemination Level		
P	Public	X
C	Confidential, only for members of the consortium and the Commission Services	

## Revision History

Revision	Date	Author	Organisation	Description
1.0	15.04.2013	Roberto Barbera	INFN	First public released for internal review
1.1	22.04.2013	Roberto Barbera	INFN	Second release taking into account the comments received by the internal reviewers
1.2	23.04.2013	Claudio Prandoni	Promoter	Formal check

### Statement of originality:

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

## TABLE OF CONTENTS

<b>EXECUTIVE SUMMARY .....</b>	<b>4</b>
<b>1 INTRODUCTION .....</b>	<b>5</b>
<b>2 ENABLING TRANSPARENT ACCESS TO DIGITAL ARCHIVES THROUGH SCIENCE GATEWAYS AND IDENTITY FEDERATIONS.....</b>	<b>6</b>
2.1 LINKING THE DCH-RP ECSG TO IDENTITY FEDERATIONS.....	6
2.2 USER AUTHENTICATION ON THE DCH-RP ECSG WITH IDENTITY FEDERATIONS AND IDENTITY PROVIDERS.....	7
2.2.1 <i>Steps to support Identity Federations</i> .....	8
2.2.1.1 <i>Integrating the Service Provider functionality into the eCSG</i> .....	8
2.2.1.2 <i>Installing an Identity Federation</i> .....	8
2.2.1.3 <i>Installing Registering the DCH-RP eCSG as a federated resource</i> .....	8
2.2.2 <i>The Sign-In procedure</i> .....	9
2.3 USER AUTHORISATION ON THE DCH-RP ECSG.....	10
2.4 USER TRACKING AND ACTIVITY LOGGING WITH THE ECSG .....	11
<b>3 IMPLEMENTATION OF THE DCH-RP ECSG USING THE GLIBRARY FRAMEWORK.....</b>	<b>13</b>
3.1 THE DCH-RP ECSG MOBILE .....	15
<b>4 UPLOADING CONTENTS THROUGH THE DCH-RP ECSG .....</b>	<b>18</b>
<b>5 CONCLUSIONS .....</b>	<b>20</b>

## EXECUTIVE SUMMARY

This document shows the architecture and implementation of the DCH-RP e-Culture Science Gateway that is one of the platforms that will be used to implement some selected Proofs of Concept identified by the project. Particular attention has been paid to simplify the access to non-expert users and to define a fine-grained authentication and authorization system that could protect digital cultural contents in an effective way.

We start presenting the reasons and the benefits of adopting the Science Gateway paradigm as a means to provide a simplified access for users to the contents of the digital archives.

In particular, we will describe how we achieved user authentication and authorization integrating Science Gateways single-sign-on mechanism with Identity Federations and user tracking and logging for any Grid transaction.

Then, we present the gLibrary platform that have been used to create the repositories on the storage resources and metadata service of the Grid infrastructure used, and the gLibrary APIs to create a set of “portlets”, deployed into the e-Culture Science Gateway, to provide an easy-to-use front-end for discovering, finding and retrieving assets of the three digital archives implemented.

The present deliverable is organised as follows.

- **Section 2** contains the presentation of Science Gateway principles and their connection with Identity Federations.
- **Section 3** presents the gLibrary framework and how it has been connected to the eCSG.
- **Section 4** reports on some limitations of the EMI1 middleware for Grid Storage Elements (SEs) that currently hamper the development of the upload functionalities promised in the Description of Work.
- Conclusions are then drawn in **Section 5**.

---

<sup>1</sup> <http://www.eu-emi.eu>

## 1 INTRODUCTION

The goal of this deliverable is to present the architecture and the current implementation of the DCH-RP e-Culture Science Gateway (eCSG) that is one of the platforms used to demonstrate some of the Proofs of Concept (PoCs) that will be identified by the project.

Since this deliverable comes before the actual implementation of the first phase of the PoCs, in order to demonstrate its functionalities some of the demonstrative repositories built in the context of the previous INDICATE project<sup>2</sup> have been temporarily included.

The DCH-RP e-Culture Science Gateway aims at proposing a model to enable transparent access to Digital Cultural Heritage contents for as many researchers all around the world as possible. However, the management of authorisation procedures, if implemented in a traditional way, i.e. assigning credentials to each new user and maintaining them during their lifecycle, would imply a significant overhead for the eCSG manager. In the meantime, end users would also get an additional set of credentials to be remembered and kept private, with the usual drawbacks: usage of weak password, re-use of the same password (thus weakening security levels) and risk of identity theft. For these reasons, we have decided to implement the Federated Access to the eCSG. This approach offers a number of advantages:

- The pool of potential users dramatically increases and it is immediately extended to all end users belonging to existing identity federations supported by the eCSG;
- The eCSG manager is exonerated from creating and keeping on its servers the users' credentials, as they are managed by Identity Providers at single Federated organizations which connect to the eCSG;
- End users don't need to obtain, manage and remember a new set of credentials, and use the usual credentials provided by their home organization.

So far, the DCH-RP is integrated in IDEM<sup>3</sup>, the Italian AAI Federation dedicated to Research, Education and Culture, managed by GARR<sup>4</sup>, and in GridP<sup>5</sup>, a "catch-all" federation also managed by GARR. Plans are to register the DCH-RP eCSG also as a Service Provider of the eduGAIN<sup>6</sup> inter-federation.

---

<sup>2</sup> <http://www.indicate-project.eu>

<sup>3</sup> <http://www.idem.garr.it>

<sup>4</sup> <http://www.garr.it>

<sup>5</sup> <http://gridp.garr.it>

<sup>6</sup> <http://www.edugain.org>

## 2 ENABLING TRANSPARENT ACCESS TO DIGITAL ARCHIVES THROUGH SCIENCE GATEWAYS AND IDENTITY FEDERATIONS

One of the main obstacles for non-IT-expert users to exploit e-Infrastructures, such as Grids, is the fact that they are based on complex security mechanisms such as Public Key Infrastructures (PKI) and accessed through low level (command-line based, i.e. non-graphical) user interfaces. The approach used in DCH-RP to solve both the previous problems and make available the PoC repositories to the largest possible number of users, was to deploy them into a “Science Gateway” whose access is regulated by “Identity Federations”.

In the recent past, interesting developments have actually been independently carried out by the Grid community with the Science Gateways and by the National Research and Education Networks with the Identity Federations to ease, from one side, the access and use of Grid infrastructures and, from the other side, to increase the number of users authorised to access network-based services.

*A Science Gateway is a “community-developed set of tools, applications, and data that is integrated via a portal or a suite of applications, usually in a graphical user interface, that is further customized to meet the needs of a specific community (US Teragrid/XSEDE project).”*

An Identity Federation is made of “[...] the agreements, standards, and technologies that make identity and entitlements portable across autonomous domains (Burton Group)”. Identity Federations have the aim of setting up and supporting a common framework for different organisations to manage accesses to on-line resources. They are already established in many countries and currently gather a number of people which is in the order of  $O(10^7)$ .

To make e-Infrastructures easy to use and more accessible, the Italian National Institute of Nuclear Physics is developing since more than two years the so-called Catania Science Gateway Framework<sup>7</sup> to create a new type of Science Gateways that implements an authentication schema based on Identity Federations. The Catania Science Gateway Framework has indeed been adopted to create the DCH-RP e-Culture Science Gateway to deploy on it some of the Proofs of Concept that will be identified by the project.

### 2.1 LINKING THE DCH-RP ECSG TO IDENTITY FEDERATIONS

Identity Federations (IdFs) usually bring together organizations in the field of Education, Research and Culture, namely Universities, Research Institutes, supercomputing centres, medical research centres, National Libraries and Museums, and other cultural institutions.

Organizations subscribing to an IdF link their Identity Provider (IdP) to the Federation. An Identity Provider is a service which enables end users belonging to the organization to use their usual credentials, and more generally their Digital Identity, in order to connect not only to resources provided by their own organization, but also to those offered by other federated organizations. Thanks to the federated approach, once the eCSG links to a specific Federation, all end users belonging to that federation are immediately enabled to be **authenticated** into the eCSG. This does not imply that they are automatically **authorised**. Indeed, unlike “old fashion”, command line based, access to Grid infrastructure, where X.509 digital certificates and their proxies, possibly containing VOMS extensions, are used both to authenticate and authorise users, one of the most interesting features of the Catania Science Gateway Framework is that it decouples the authentication and authorization steps, the first one being demanded to the IdPs

---

<sup>7</sup> <http://www.catania-science-gateways.it>

while the second one remains with resource owners and implements their own access policies. Each user of an IdF will need to be authorised to access a specific resource within the eCSG according to its owner's policies. So, different user groups will access different subsets of resources and may have different rights on them.

## 2.2 USER AUTHENTICATION ON THE DCH-RP ECSG WITH IDENTITY FEDERATIONS AND IDENTITY PROVIDERS

As stated before, one of the strengths of the DCH-RP eCSG, compared with the low-level, command-line use of Grid middleware, is the logical and practical separation of the authentication (AuthN) phase from the authorization (AuthZ) one. In order to access the eCSG, a user must be both authenticated and authorized but we treat the two steps separately and with different technologies. The schema for authentication and authorization is depicted in Fig. 1.

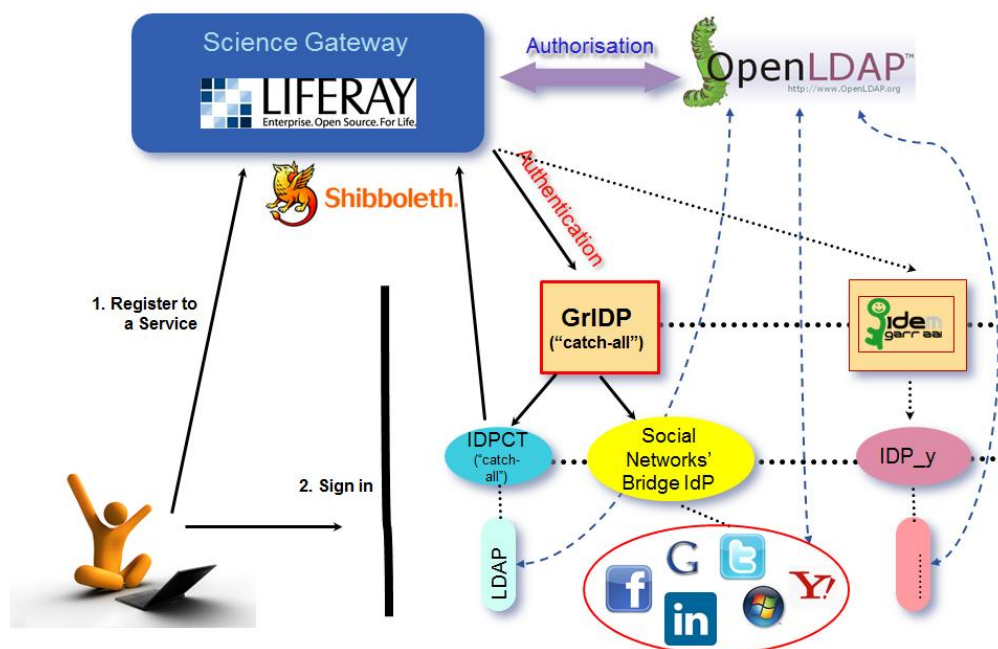


Figure 1: Authentication and Authorisation schema of the eCSG

User authentication relies on IdPs that are members of one or more Identity Federations. We only support federations based on the SAML 2.0 standard specifications and on its implementation done by Shibboleth and SimpleSAMLphp.

We currently support the GARR IDEM Federation and, through it, the eduGAIN inter-federation. We also support all the Identity Providers of the Grid IDentity Pool<sup>8</sup> (GrIDP), a “catch-all” Identity Federation jointly operated by INFN Catania and GARR that we have expressly created to gather all the IdPs that do not already belong to any official federations and all the users of the eCSG who are not (already) registered in any IdPs. This is particularly important and useful in the contexts where it is necessary to authenticate the so-called “citizen scientists” (i.e., people belonging to the general public) and let him/her access the e-Infrastructure for dissemination and self-learning purposes.

Inside the GrIDP Federation, we have also created two special IdPs: IdP Open<sup>9</sup> (a normal Shibboleth-based service) and the “Social Networks’ Bridge Identity Provider”<sup>10</sup>, that allows people to get

<sup>8</sup> <http://gridp.garr.it>

<sup>9</sup> <http://idpopen.garr.it>

authenticated with the same credentials they already have with the most known and populated social networks. Both IdPs have recently been endorsed by GARR and are maintained at GARR premises and their availability is in line with the recommendations contained in the recent TERENA AAA Study<sup>11</sup>.

## 2.2.1 Steps to support Identity Federations

As an example, the following are the technical steps implemented in DCH-RP in order to link the eCSG with the IDEM GARR federation.

### 2.2.1.1 Integrating the Service Provider functionality into the eCSG

Being basically a web portal, the eCSG can be easily integrated with the Relying-party functionality (also known as Service Provider) from the SAML Web Single-Sign-On profile. This can be done by exploiting several software frameworks. In the pilot we chose the Shibboleth Service Provider solution. Once implemented the Service Provider functionalities on the eCSG, the service can be registered as a resource in the Federation.

### 2.2.1.2 Installing an Identity Federation

A discovery service provides a browser-based interface where a user selects his/her Identity Provider. The service provider uses this information to initiate SAML Web Browser SSO.

By activating a dedicated Discovery Service, it is possible to support other federations and single IdPs that are not integrated in any federation. In the DCH-RP, the selected implementation for this component, done in GrIDP, is Shibboleth's Centralized Discovery Service.

### 2.2.1.3 Installing Registering the DCH-RP eCSG as a federated resource

In order to become part of a Federation, an Organization needs to subscribe an agreement with the Federation itself. The agreements imply accepting shared security policies and relating to the offered services and management of identity data shared by federation members. By accepting these policies and security standards, federation participants build a network of trust, which allows Service Providers to accept Digital Identities that are guaranteed by Identity Providers without any need to verify them. This trust also allows Identity Providers to share the users' attributes with Service Providers, who are bound to use them according to the Federation's rules.

In this case, the organization responsible for the DCH-RP eCSG, i.e. the INFN, had to subscribe to IDEM Federation. Once registered the eCSG as an IDEM resource, all IDEM end users are immediately enabled to authenticate in it.

Likewise, it was possible to register the eCSG in the GrIDP "catch-all" federation and will be possible to register it in the eduGAIN inter-federation. eduGAIN is intended to enable the trustworthy exchange of information related to identity, authentication and authorisation among the GÉANT<sup>12</sup> Partners' federations. To this end, eduGAIN coordinates elements of the federations' technical infrastructure and offers a common policy framework controlling the exchange of this information. Its initial goal is to enable Pan-European Web Single Sign On (Web SSO) to both GÉANT services and to those provided by other

---

<sup>10</sup> <http://idpsocial.garr.it>

<sup>11</sup> <https://confluence.terena.org/download/attachments/30474266/2012-AAA-Study-report-final.pdf?version=1&modificationDate=1355503760046&api=v2>

<sup>12</sup> <http://www.geant.net>



communities represented by, or associated with, the GÉANT Partners. Thanks to the participation in eduGAIN, users from the member Federations can be automatically enabled to access the eCSG.

## 2.2.2 The Sign-In procedure

The user opens in his/her web browser the URL <http://ecsg.dch-rp.eu> (see Fig, 2) and chooses the “Sign In” option (on the upper-right part of the webpage).

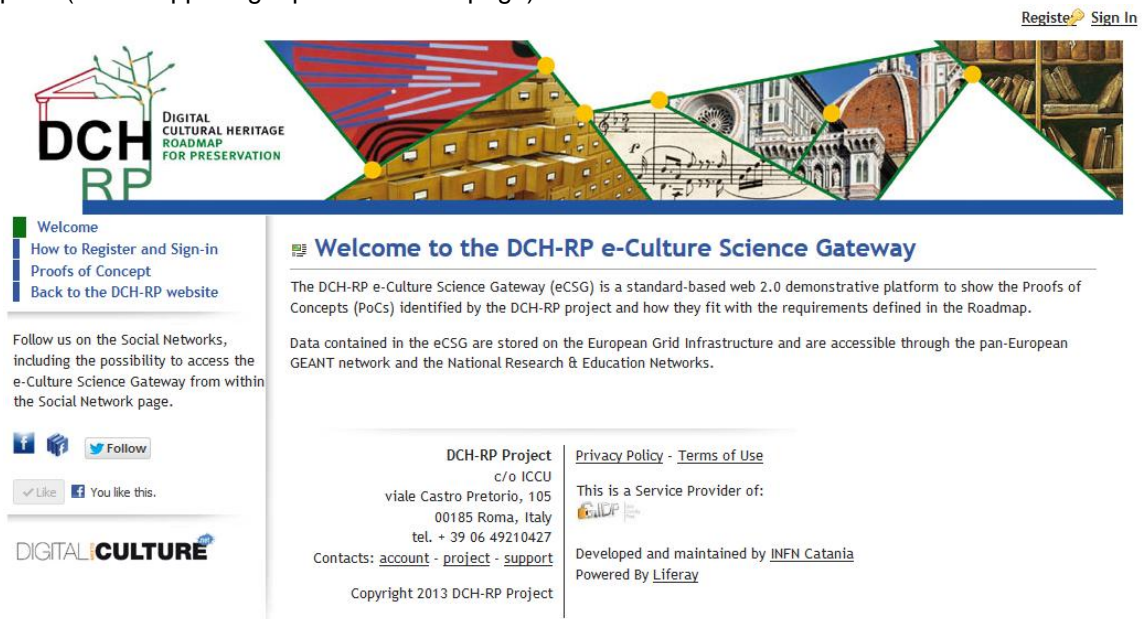


Figure 2: Home page of the DCH-RP eCSG

The user is redirected to the Discovery Service (see Fig. 3), i.e. a webpage where s/he selects his/her federation and home organization (i.e., his/her IdP).

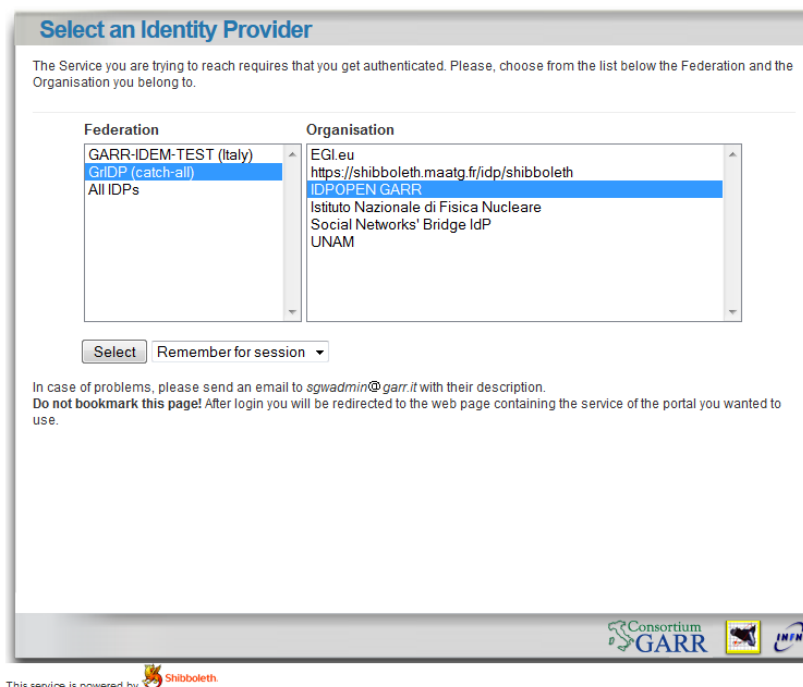
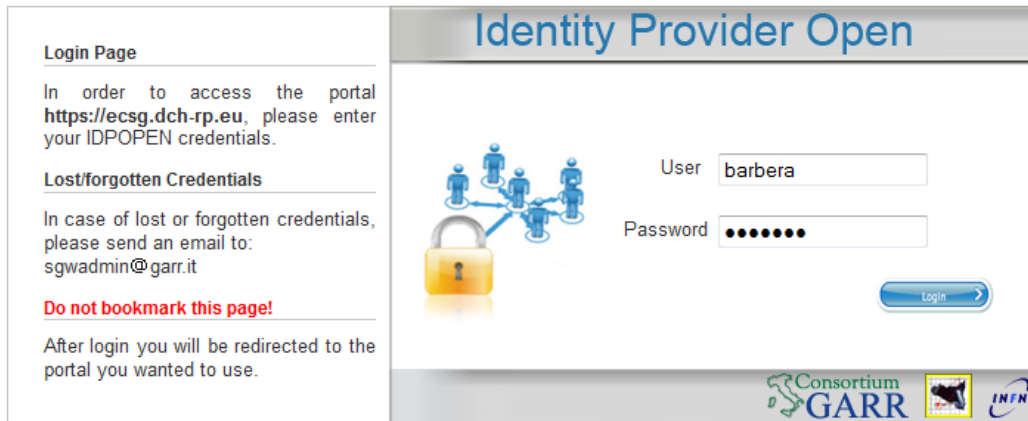


Figure 3: Selection of the Identity Federation and Identity Provider through the Discovery Service

The user is now redirected to his/her usual IdP login webpage (see Fig. 4), where s/he logs in with his/her credentials.




This service is powered by  Shibboleth.

Figure 4: Login page of the selected Identity Provider (IDPOpen GARR in this case)

The user is now signed-in and redirected to the eCSG, where s/he will access only those resources for which s/he got an authorization (see Fig. 5).

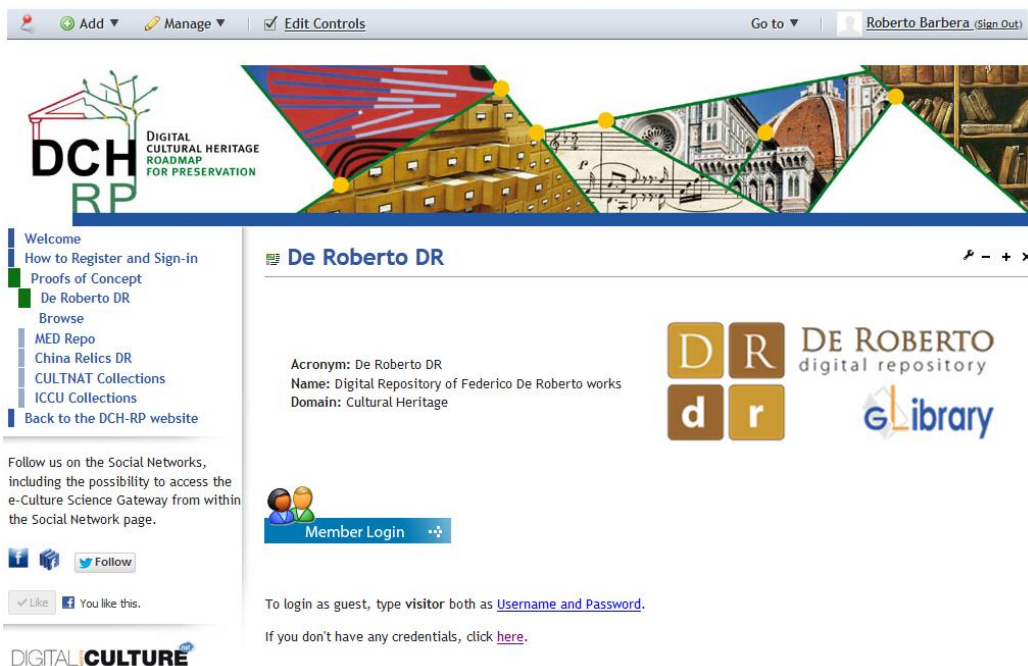
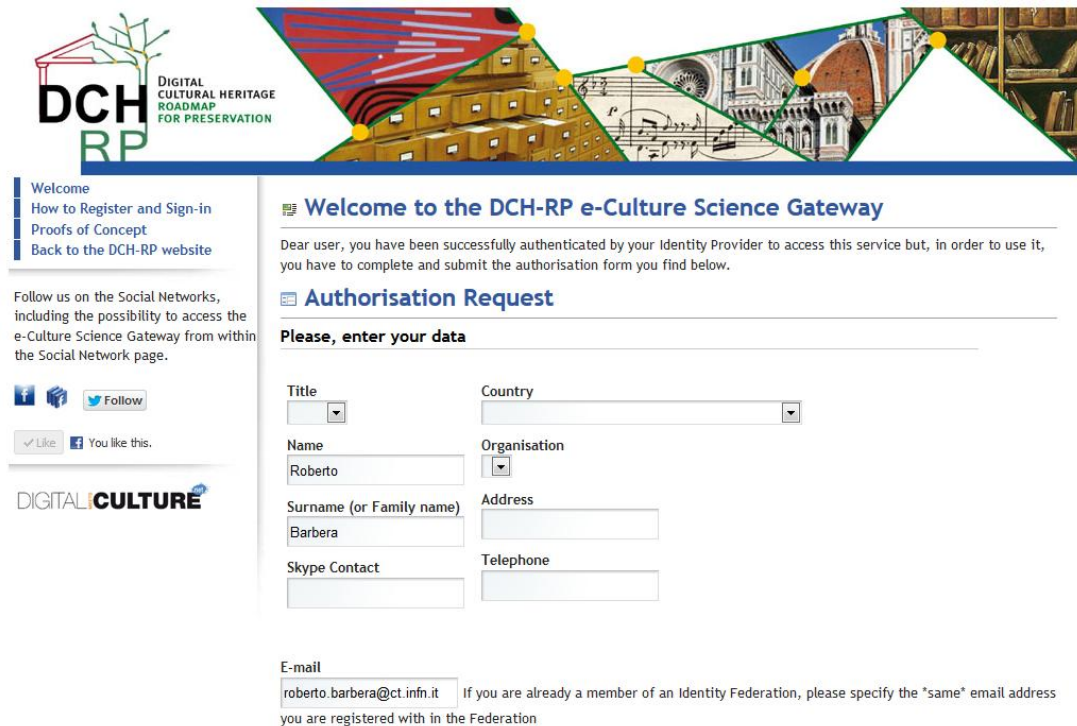


Figure 5: Redirection to the eCSG home page with authentication confirmed and browser functionality visible

## 2.3 USER AUTHORISATION ON THE DCH-RP ECSG

Unlike authentication, user authorization is carried out at the level of the eCSG: users whose request to register (see Fig. 6) is approved by the managers of the eCSG, are stored in a LDAP-based registry together with the roles they have and the privileges they are granted.



Welcome

[How to Register and Sign-in](#)

[Proofs of Concept](#)

[Back to the DCH-RP website](#)

Follow us on the Social Networks, including the possibility to access the e-Culture Science Gateway from within the Social Network page.

[Follow](#)

[Like](#) You like this.

**DIGITAL CULTURE**

---

## Welcome to the DCH-RP e-Culture Science Gateway

Dear user, you have been successfully authenticated by your Identity Provider to access this service but, in order to use it, you have to complete and submit the authorisation form you find below.

### Authorisation Request

Please, enter your data

Title

Country

Name

Organisation

Surname (or Family name)

Address

Skype Contact

Telephone

E-mail  If you are already a member of an Identity Federation, please specify the "same" email address you are registered with in the Federation

Figure 6: Authorization request form

At present, only the group “genericUser” has been created in the LDAP registry for the eCSG. It gathers all authenticated users without any other special authorization grants. Of course, as many other groups as needed can be added to control user access down to the single file containing the single content unit of a given repository.

Once a user has been authorized to access the eCSG, he/she can then sign in and browse the digital repositories he/she is entitled to access within the portal.

## 2.4 USER TRACKING AND ACTIVITY LOGGING WITH THE ECSG

It is worth mentioning that, in order to be compliant with the strict rules of the European Grid Infrastructure VO Portal and Grid Security Traceability and Logging policies, each operation done by the user inside the eCSG is stored on a User Tracking Database that can be inspected at any time by the administrator of the portal. This ensures the non-repudiability of Grid transactions which is one of the most important requirements of the Grid Security Infrastructure.

The authentication and authorization mechanism described above has the big advantage of being based on standards and greatly simplifies the access to e-Infrastructure by non-IT-expert users avoiding the need for them to get and manage personal digital certificates. However, all Grid transactions must be signed with proxies generated by standard X.509 digital certificates so we have implemented in the Catania Science Gateway Framework a mechanism that creates proxies on the fly and on user request. This is done by a service called eToken server. The eToken server generates proxies starting from robot certificates. Robot certificates are special, yet standard, digital certificates stored in USB Smart Card, referred to as etokens. It is possible to bind robot certificates with **digital repositories** and allow people to access them without any personal credentials. According to this schema, when a user is authorized to access the eCSG and wants to access one of the digital archives she is allowed to, the portal retrieves on her behalf a valid proxy for the eToken server. The proxy generated on the fly contains the extensions that specify the role and privileges of the robot certificate inside the VO supported by the Science

Gateway, so different proxies can be created according to the different roles and privileges of the user in the LDAP registry. This ensures a fine grained authorization and provides the portal manager with the complete control of deciding what a given user can see and do. As an example, portal managers can allow all authenticated users to access low-resolution images while restricting the access to high-resolution ones to a sub-group of them.

### 3 IMPLEMENTATION OF THE DCH-RP ECSG USING THE GLIBRARY FRAMEWORK

Besides the Catania Science Gateway Framework, the main software package behind the DCH-RP e-Culture Science Gateway is gLibrary<sup>13</sup>, a framework developed by INFN Catania since several years that allows the creation, organization, browsing and retrieving of digital assets on Grid-enabled digital repositories, hiding the underlying technical details to the end users.

The multi-layer architecture of gLibrary is presented in Fig. 7

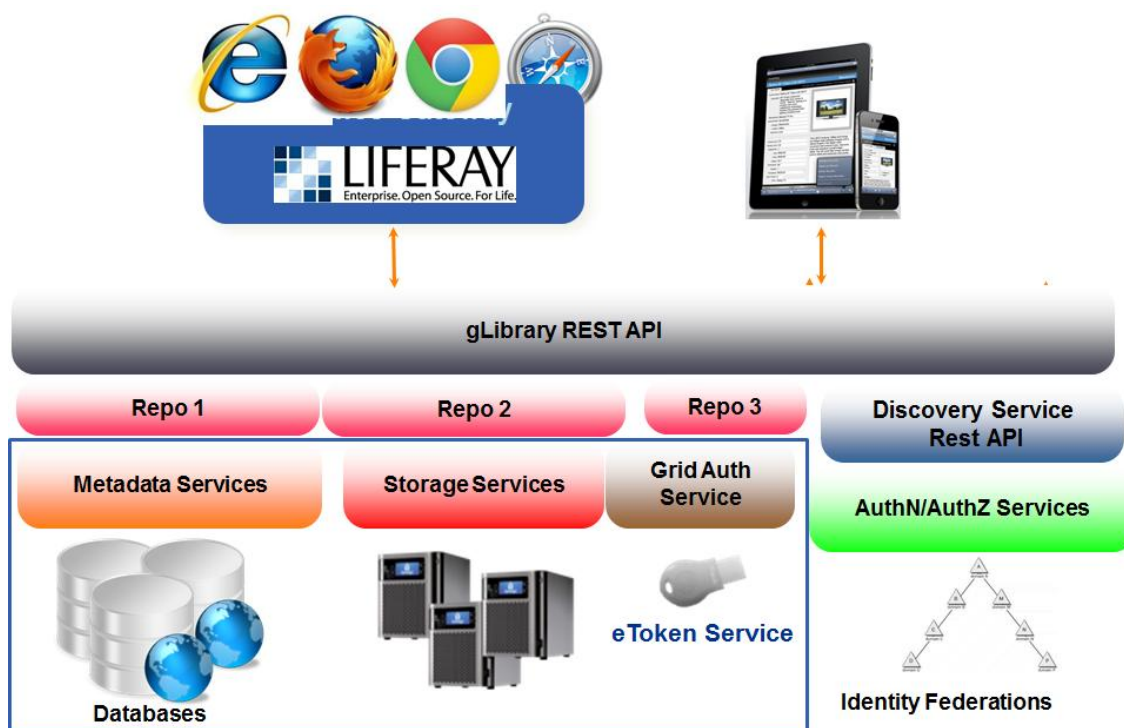


Figure 7: Multi-layer architecture of the gLibrary framework

A RESTful API allow both web browser and mobile appliances to manage federated access to Grid Storage Elements to manage what we call “digital assets”, i.e. the combination of digital data and metadata associated to them.

The browsing system (see Fig. 8) has been designed to quickly retrieve the desired content among thousands of items using an intuitive filtering system on metadata, accessed on the header of each column. Once user has found the object s/he needs, a geo-map (see Fig. 9) of the storage resources that have an available replica of the selected item is shown letting him/her choose the storage element to download the desired content from (see Fig. 10).

<sup>13</sup> <https://glibrary.ct.infn.it>

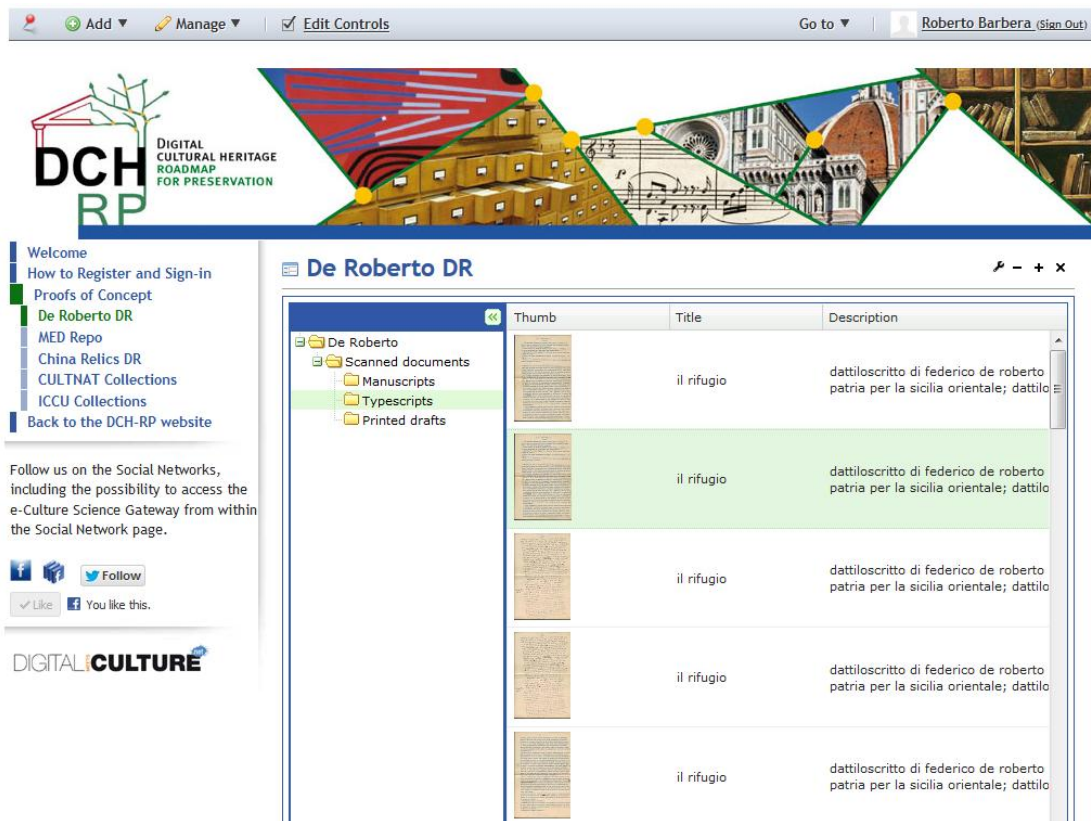


Figure 8: Browsing of a digital repository

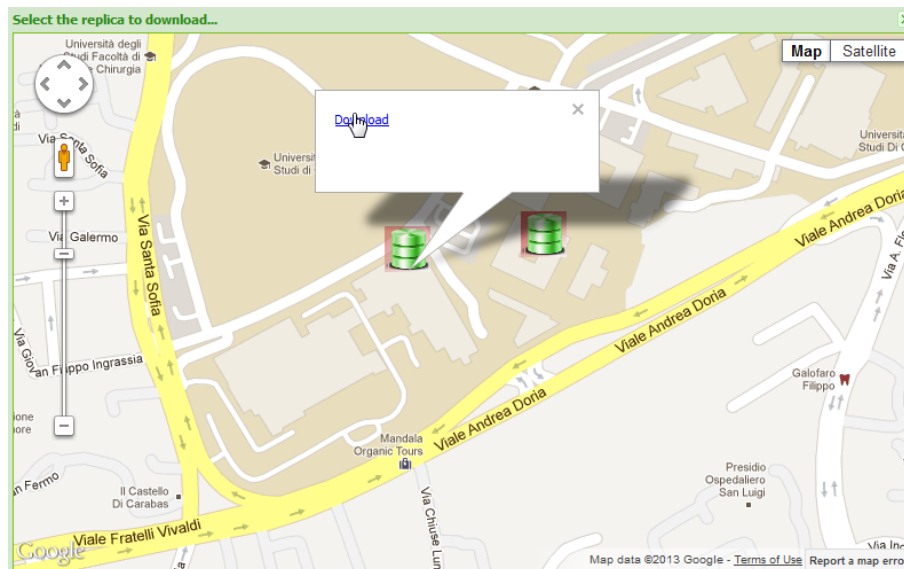


Figure 9: Selection of a replica of one of the contents of a given repository with the option to download it

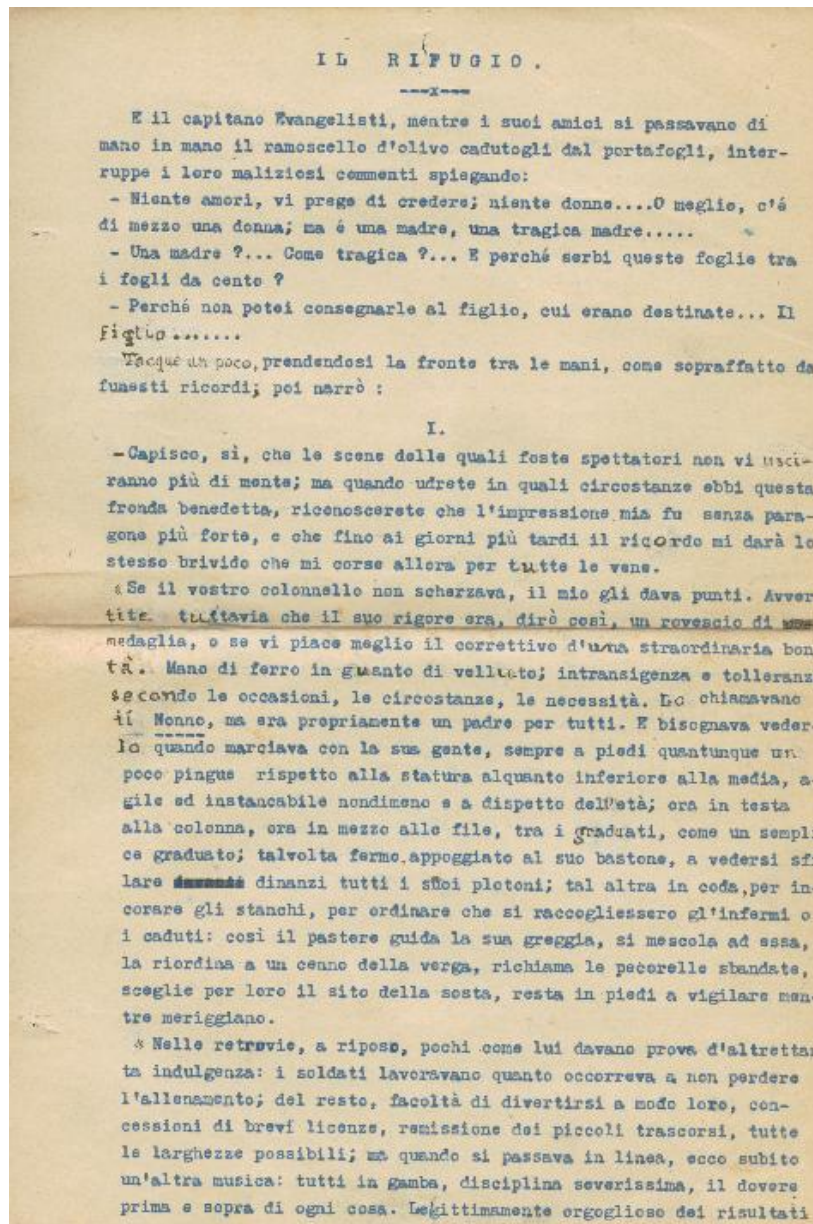


Figure 10: A content downloaded

### 3.1 THE DCH-RP ECSG MOBILE

In order to demonstrate the power of the gLibrary RESTful API, the DCH-RP eCSG Mobile<sup>14</sup> has recently developed as an Android application.

Figures 11 shown the “splashscreen” of the app, while Fig. 12 and 13 show the same functionalities of the web-based eCSG available in the app.

<sup>14</sup> <https://play.google.com/store/apps/details?id=it.inf.n.ct.dchrpSGmobile>



Figure 11: Splash-screen of the DCH-RP eCSG Mobile

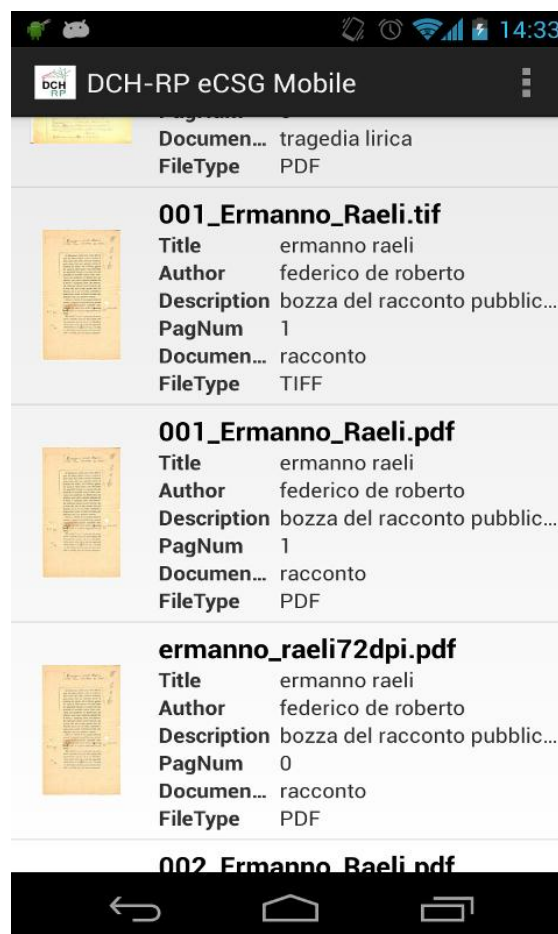


Figure 12: List of contents available in a given repository





Figure 13: Inspection of metadata of a particular element of the repository

## 4 UPLOADING CONTENTS THROUGH THE DCH-RP ECSG

One of the distinctive features of the DCH-RP eCSG, promised in the DoW, was a browser-integrated tool to easily upload contents on PoC repositories and create/modify the corresponding metadata.

In order to perform the upload from the computer of the repository manager to the Grid SEs, there are two possible ways: the indirect upload and the direct upload.

The indirect upload, depicted in Fig. 14, is a two-step process by means of which a content manager first uploads the files on the eCSG and then the eCSG moves them to a Grid Storage Element using the SRM protocol and the gLibrary API.

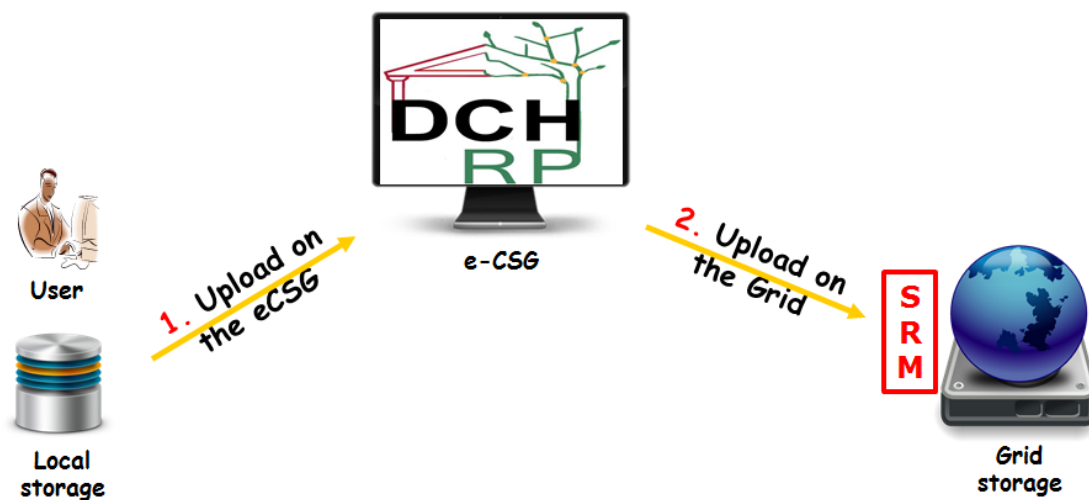


Figure 14: Indirect upload

The advantage of the indirect method is that it is immediately feasible but has the big drawback of not being scalable as the disk of the eCSG and the network bandwidth of the site where it is installed will constitute, sooner or later, severe bottlenecks.

A better alternative is represented by the direct upload, shown in Fig. 15, where (i) the repository manager informs the eCSG that s/he want to upload some files; (ii) the eCSG chooses a Storage Element where the vo.dch-rp.eu Virtual Organisation is authorised and prepare the upload, and (iii) the repository managers uploads the files directly on the SE using its HTTP/HTTPS interface.

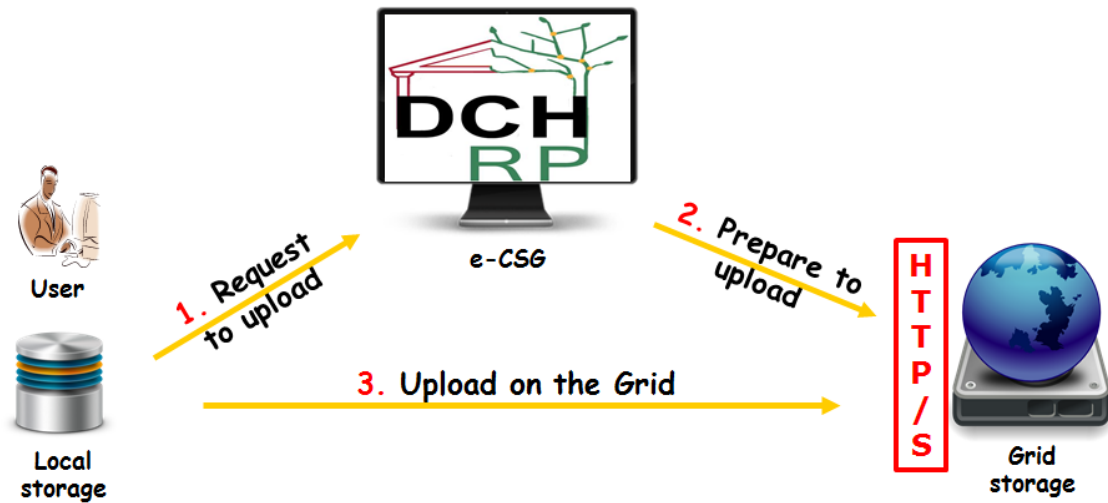


Figure 15: Direct upload

The direct uploads eliminates the bottlenecks of the indirect method but it is unfortunately not immediately feasible as none of the Storage Element software included in the EMI middleware release (i.e., dCache, DPM and StoRM) currently support the POST verb of the HTTP protocol. We have informed the middleware developers of this strong requirement coming from DCH-RP but, at least at the time of writing, there is not fixed deadline for this requirement to be satisfied.

In order to keep the promise done in the DoW, INFN Catania team has developed a patch for DPM to allow the direct upload. The patch has been successfully tested on the Catania SE devoted to DCH-RP and the graphic user interface for the upload, integrated in the eCSG, is being developed. We will include its description in a next-future update of the present deliverable.

## 5 CONCLUSIONS

E-Infrastructures can be very beneficial platforms for the Digital Cultural Heritage (DCH) community, provided they are «easy to use».

The DCH-RP e-Culture Science Gateway is a major step forward towards the uptake of Grid technology by the DCH community. The adopted Catania Science Gateway Framework, supporting Identity Federations and Social Networks, can revolutionize the way Grid infrastructures are used, hugely widening their potential user base, especially non-IT experts and the “citizen scientist”, through the web and on mobility. The adoption of standards, in particular, represents a concrete investment towards sustainability.

The direct upload method will further simplify the ingestion of contents into data repositories managed by the eCSG and we will follow very closely the further developments of the EMI middleware and the possible deployment of the found patch at least on the site belonging to the DCH-RP infrastructure.

In conclusion, it is worth noting the fact that the eCSG, based on the gLibrary framework, is a new-generation web and mobile based front-end but it does not implement any particular vertical solution for long term digital preservation of contents. This will come up in the roadmap that WP3 is expected to devise.