The DIGARV Platform: A collaborative platform for working with cultural heritage data and research data

Johan Åhlfeldt[1] and Arild Matsson[1]

[1] GRIDH, University of Gothenburg, Box 100, 405 30 Gothenburg, Sweden

Abstract

This article covers an easy-to-use research tool for collaborative work. The tool has been adapted for structured data and high-resolution images within four research projects at GRIDH. The platform is especially designed for working with temporal and spatial data. Furthermore, the platform gives researchers access to a relational database system through input forms and access to external cultural heritage data including high-resolution images. This way the platform also aims to utilize external data published as Linked Open Data (LOD) and, at the same time, prepare its own research data for publishing as LOD. Because of the spatial and temporal nature of the data, it is visualized in time and space through maps and timelines to give overview and context during the data management phase.

Keywords

Database management, Linked Open Data, IIIF, SQL, Online publishing, Online collaboration, GIS.

1. Introduction

In 2019 the Centre for Digital Humanities (CDH), now known as the Gothenburg Research Infrastructure in Digital Humanities (GRIDH), was part of two successfully funded projects in the *Digitisation and accessibility of cultural heritage* (DIGARV) call, https://digarv.se/en/. The two projects were *Expansion and Diversity*[1] and *Mapping Lived Religion*[2]. Both projects are characterized by data driven approaches, curated datasets, large projects groups and collaborative work in collecting and managing data from various sources, both digital, printed sources and manuscripts.[3] The responsibility of GRIDH was to facilitate the collection of existing datasets into an information system, one database for each project, and set up a system for collaborative data management among the researchers.

Despite the different project focuses, Medieval history & Art history and Cultural studies respectively, both are methodologically oriented towards visualizing and analyzing data in time and space on maps. The map and timeline visualize the pattern and distribution of the phenomenon under study on the aggregate level (distant reading) and at the same time, give users access to individual cases with all its evidence in full detail (close reading). Both projects maintain a need to import and manage data online.

We decided to build two spatial enabled and relational databases (SQL) from data models that emerged from the needs of each project and their respective research questions. The platform, which we at GRIDH called the *DIGARV Platform*, was built at the same time to accommodate the needs of both projects. We decided to build both projects on the same code base split between a backend and a frontend

---

[1] https://www.gu.se/en/research/expansion-and-diversity-digitally-mapping-and-exploring-independent-performance-in-gothenburg-1965-2000

[2] https://lnu.se/en/research/research-projects/mapping-lived-religion-medieval-cults-of-saints-in-sweden-and-finland/

[3] Nilsson *et al* 2023. Liepe, L., & Ellis Nilsson, S. (2021)

part, as detailed in later sections, the only difference lies in their different configuration files. Each project was assigned a dedicated sub domain on top of the domain for GRIDH which is at https://dh.gu.se, https://expansion.dh.gu.se and https://saints.dh.gu.se respectively. The intention was to build a common platform that could also be used in future projects which was the case for the project *Pehr Strands flöjtur* (Pehr Strand's Pipe Organs) [3] and the internal project at GRIDH *Iconographia Mediaevalis Suetica*[4].

The DIGARV Platform is primarily used as a space for collaborative work. However, both projects decided to keep the platform open to the public from the beginning of its existence, except for the part of editing data. This open approach was important in the *Expansion and Diversity* project to communicate with the informants of the project and, in the *Mapping Lived Religion* project the members of the reference group. In the Mapping Lived Religion project, the platform is also used by a similar project at Tampere university in Finland, *Lived Religion in Medieval Finland*, together more than 10 researchers. During the project period the *Mapping Lived Religion* project have had 16 students editing data as part of their academic training. Users can be added to the system as members to gain edit privileges by the system administrator at GRIDH. The platform has been continuously evaluated and adjusted to better meet the needs of the projects and its members and to create a smooth workflow.

This paper doesn't cover the resulting public interfaces for the two projects which were developed later.

2. User interface

The DIGARV Platform was created with a split frontend–backend architecture. The frontend is implemented in the Vue 2 JavaScript framework[5] and notably uses the OpenLayers map library[6] and the OpenSeaDragon image viewer (see **Section 7**), as well as multiple smaller libraries for interactive components, etc.

The user interface is designed to visualize spatial and temporal data on maps. The setting for the temporal data connected to places on the map is adjusted by a time slider, using Min, Max and step values, see the top left in **Figure 1**. The width of the map and the tab pane can be adjusted seamlessly keeping the center of the map. Multiple layers can be displayed at the same time and different background layers can be selected.

To the right of the map is a set of five tabs, see the top right in **Figure 1**. The *Home* tab is a short introduction to the project in question. The *Layers* tab is where the user selects background layers and available layers as overlays. The *Search* tab is where the user search for entities in the database and selects one of them for display or edit. From the Search tab users can also create new records for each entity. The *Show* tab displays all data connected to one record, including high resolution images and relations to other objects. Finally on the *Edit* tab records can be created or edited.

Navigation between tabs is automatic but users can also shift between the tabs at any time and content is preserved. This way, a multitude of information and settings are displayed in a very compact and easy to access way.

The *Edit* and *New* buttons and the Edit tab are only displayed when a user is logged in, see near the top right in **Figure 1**. When users are logged in, individual contributions are tracked and saved to the database. This is important for responsibility, traceability and later in the public interface, proper attribution.

The default language is English. However, we have prepared for several interface languages, in the case of Mapping Lived Religion, the public interface will also be available in Swedish and Finnish.

---

[4] https://iconographia.dh.gu.se, no English version available.
[5] https://v2.vuejs.org/
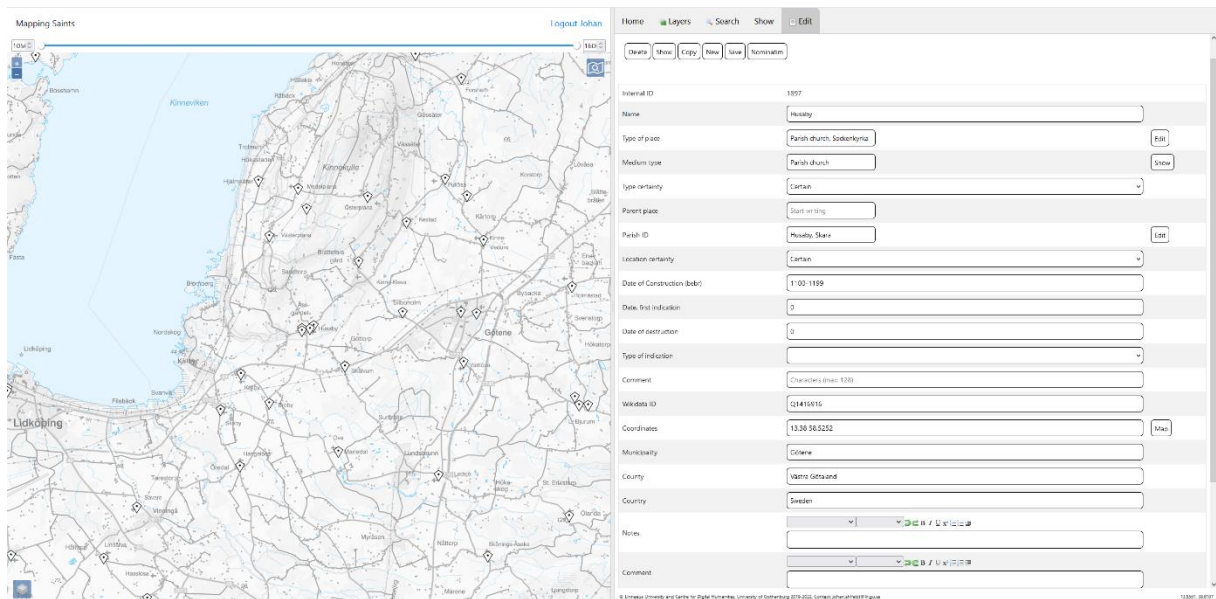[6] https://openlayers.org/

Figure 1: The interface of the *Mapping Saints* application. The map is centered around Husaby parish church, Götene municipality in Sweden. The Edit tab is available when logged in as a member. The terrain map (in gray scale) by Lantmäteriet (Land Survey) is selected as the background layer. The place overlay in this example represents all places available in the database and information about a particular place can be accessed by clicking the marker on the map, or by searching for the place name in the search interface. Map navigation takes place through pan and zoom. A full resolution screenshot is available at https://dh.gu.se/data/figure1_mapping_saints_husaby.png. The *Show* tab of the Husaby parish church is available here, https://saints.dh.gu.se/place/1897 with further public access to all tabs except the Edit tab.

3. Database implementation and population

According to the aim of each project and their data, two different database models were created. The commonality of their structures made it easy to develop a similar code base. The databases are implemented in MySQL, with database tables of two sorts.

The first is for the main entities in the database, e.g., cult, place, parish, persons in the *Mapping Lived Religion* project, and organisation, person, place, production, event tables in the *Expansion and Diversity* project.

The second type of tables are the relation tables. These connect the entities to each other and sometimes with a qualifier property such as certainty or a specific role of an agent.

There are also tables connecting entities in our database with entities in external databases, especially the Swedish Open Cultural Heritage database (SOCH)[7]. This allows us to explore and reuse cultural heritage objects within our own interface in real time and can subsequently be used as evidence in our research (see below **Section 4** on Linked Open Data).

In all tables, an integer field id is used as primary key, i.e., a unique identifier for each row in the database table. If a field is a reference attribute or foreign key to another table, it is always the id field that is referenced. For geographical data we always use the GEOMETRY data type, capable to hold different kinds of geometries, i.e., POINT, LINESTRING or POLYGON. This column is always named *geom*. Many of the tables contains fields like *created*, *updated* and *modified*. They can be changed automatically with the information that a logged in user provides or the current date and time.

Aside from the object and relation tables containing the main project data, each project database also contains a special *Column Labels* table. Each row describes a field in a database table and contains a human-readable label, instructions for the editing form and whether the field should be hidden from

---

[7] https://www.raa.se/in-english/digital-services/about-soch/

display, read-only and searchable. For fields that function as references to other tables, they also indicate the table and field name of the reference target. These values are included when the backend application serves detailed data for a given object, helping the frontend application to use and render the fields appropriately.

In the early phase of the *Expansion and Diversity* project, we received a copy from Göteborgs Stadsmuseum [the City Museum of Gothenburg] of their database Carlotta, regarding data about non-institutional performing art groups as well as related persons and events. There were differences between this data model and our own – for instance, we separate Production entries (e.g. a play or a dance piece) from Event entries (the performance of a Production at a certain time and place). Nevertheless, we were able to use this data firstly for populating parts of our own database and secondly for maintaining identifiers of corresponding data entries in SOCH. When our list of performing art groups was expanded as a result of our research, we made a simple API for the museum to use to collect our persistent identifiers for groups already in their collection and the new ones.

By assigning subdomains to projects and minting persistent identifiers for the main entities in the database, we can reference everything in the database directly by HTTP URIs. These URIs can be used by other datasets to reference and download content from our information system in real time. They can also be used in scholarly literature. For instance, https://saints.dh.gu.se/place/1897 denotes Husaby parish church, Götene municipality, Sweden, and https://saints.dh.gu.se/cult/88 denotes the Altarpiece with five sculpted saints and paintings depicting scenes from St. Lawrence's legend in Adelöv parish church.

4. Linked Open Data (LOD)

The concept of Linked Open data is a compilation of principles and methods for data publishing reusing/sharing of data. When everyone publishes their data online with an open license, makes them machine readable, describes them in a standardized way and links them to other data, the Internet itself can be used as a common database. This enables anyone to build digital services based on data from several providers, for instance, Cultural heritage and Research institutions. In our case, we can integrate data from remote providers in our own platform instead of making copies and storing them on our servers.

Since many of the digital sources we utilize are published in SOCH, it is a good idea to also publish our research data in SOCH. Our contribution, especially in the Mapping Lived Religion project, is to bring scholarly comments and concepts on cultural heritage objects and put them in a historical and analytical context. This is accomplished by referring to pieces of source material by their persistent identifiers. Linked open data is also helping us define our concepts and make them compatible with other online datasets. Besides Wikidata, the Art & Architecture Thesaurus[8] and Iconclass[9] are two examples of such classification systems. Another contribution of the Saints project is the compilation of a list of medieval cult places, with coordinates, variant and historically attested names and with references to online sources, such as Bebyggelseregistret [protected buildings] and Fornsök [archaeological sites] at Riksantikvarieämbetet, the Swedish National Heritage Board.[10]

5. Backend application

The backend of the DIGARV Platform is a lightweight PHP 7-driven server application. Except for some specific functionality, the code is free from dependencies to any specific research project. It features a database query layer and an API (see next section) for listing, describing and editing data.

---

[8] https://www.getty.edu/research/tools/vocabularies/aat/ by the Getty Research Institute.
[9] https://iconclass.org is a classification system for the content of images and maintained by the Henri van de Waal Foundation.
[10] Nilsson et al (2023).

There is a strict separation of concerns between data content and appearance, so the backend accepts data in GET and POST HTTP requests and responds with data in the JSON format.

The generalized code is specified to a certain research project by means of a *Tables File*. This is a PHP file that declares the data model: Object types, relationships between them and what fields to use for labelling objects, searching them, positioning them on a map, et cetera. The file contains a single multi-dimensional array following a documented structure.

To deploy the backend for a certain project, the code is copied to a web-accessible root directory on the server. A config file is added, containing database credentials and a pointer to the Tables File associated with the project.

6. API

The web application programming interface (API) of the backend offers ten endpoints for listing, finding, describing and editing data. Describing three of these may be sufficient to outline the capacity of the service.

- The *Search* endpoint lists items of a given type according to given criteria. As an example, a usage could be to list *Organisations* founded after 1980 and having a page on Wikipedia.
- The *Map* endpoint does something similar but returns the list in the *GeoJSON* format [5], compatible with GIS software such as the map display in the frontend.
- The *Edit* endpoint, given an object type and an identifier, returns the full details of the indicated object. As the name suggests, the returned data is used not only for displaying information, but also to generate an edit form (by a user with the appropriate permissions).

7. Serving high-resolution images with IIIF

Newly added data points in the Mapping Lived Religion project are being connected with previously existing data in the *Medeltidens bildvärld* photographic collection of the Swedish History Museum[11] and the *Iconographic register*[12]as part of the digital archive at the National Heritage Agency. These collections are available by API through SOCH, but high-resolution images could only be obtained offline (on physical storage media, in batches).

The *IIIF* protocol was developed by a consortium of actors in the libraries and academia, founded in 2015, to solve various problems with delivering media on the web. Its Image API specifies a method of saving bandwidth when serving high-resolution images[13]. In essence, it relies on fixed-size slices of each image at different resolutions. Thus, a single low-resolution slice may be enough to show a thumbnail. Conversely, when zooming in on a section of the image, only the few high-resolution slices overlapping with the viewport are necessary.

Among the multiple open-source IIIF implementations, the *IIPImage server*[14] was selected for its speed and installed in the local server infrastructure at GRIDH. As a prerequisite, images need to be in a multi-resolution format. To this end, the *VIPS* image processing utility[15] was executed on the server after uploading the original Medeltidens bildvärld images.

The frontend application employs the *OpenSeaDragon* viewer component[16] to use the IIIF-supported image service. This allows visitors to the Mapping Saints website to explore photographic representations of cult objects in great detail, without much lag or disruption. The images from the

---

[11] Medeltidens bildvärld has its website (with only low-resolution images) at
https://medeltidbild.historiska.se/medeltidbild/default.asp
[12] Liepe, L., & Ellis Nilsson, S. (2021).
[13] Read the IIIF Image API 3.0 at https://iiif.io/api/image/3.0/
[14] https://iipimage.sourceforge.io/documentation/server
[15] https://www.libvips.org/API/current/using-cli.html
[16] https://openseadragon.github.io/

Iconographic register (register cards with photographic images) were much smaller in size and didn't need to be in a multi-resolution format to display quickly in the *OpenSeaDragon* viewer.

8. Legacy

The platform backend code is almost entirely developed internally by research engineers at GRIDH. Unlike most open digital research projects, it does not utilize many external code packages. Consequently, it suffers from some issues including incomplete security, susceptibility to bugs and a high time cost of adding common functionality.

This motivated GRIDH, in later years, to create a new system with similar goals as the DIGARV Platform but relying more on free and open-source software frameworks. Since then, the new system, named Diana, has been the preferred collaboration platform for new research projects.

Diana builds on top of Django[17], a Python library for managing database systems, to create input forms and API endpoints. Django also have a robust authentication and authorisation system down to the database table level for individual users and groups. The admin interface of Django (the input forms) is generated by directives written in Python code. The output are HTML pages which can be customized further for the purpose of displaying source material like documents and images. We have already implemented a display of PDF-documents and a static IIIF-client to display images next to the input forms. There is also a map editing interface which can modify existing geometries and create new ones.

As discussed above, the DIGARV Platform has many of these features, some of them are implemented in a rudimentary way, others are implemented in a more advanced way, highly adapted for the research task at hand. The advantage of the DIGARV Platform at this point is that it implements a visualization of all places and optionally more specific layers important for the editing process. The platform also fully implements a IIIF-client where images can be panned, zoomed and rotated. Much work remains to build a custom platform on top of Django with the same functionality as the DIGARV Platform.

9. Conclusions

The DIGARV Platform is a highly adaptable platform for collaborative research working with structured data and relational databases. It is also a platform for the integration of data from external information system, including image evidence as source material for scholarly work. To researchers and the general public alike, it presents data for distant and close reading, and specializes in spatial-temporal data most notably through an interactive map with an adjustable time window.

Compared to the later Django-based system, the DIGARV Platform has both advantages and drawbacks. While most functionality comes included with Django, certain specific features are easier to develop in the bespoke in-house platform.

The programming code for the entire platform will be published on GRIDH's Github account, https://github.com/gu-gridh

Acknowledgements

---

[17] https://www.djangoproject.com/

References

[1] Ellis Nilsson, S., Zachrisson, T., Fröjmark, A., Liepe, L., & Åhlfeldt, J. (2023). 'Mapping Saints: creating a digital spatial research infrastructure to study medieval lived religion'. In: *Digital Spatial Infrastructures and Worldviews in Pre-Modern Societies*, A. Petrulevich & S. Skovgaard Boeck (eds.), Arc Humanities Press, 2023, p. 33-58

[2] Liepe, L., & Ellis Nilsson, S. (2021). 'Medieval Iconography in the Digital Age: Creating a Database of the Cult of Saints in Medieval Sweden and Finland'. Iconographisk post: *Nordisk tidskrift för ikonografi.* (2). 45-63.

[3] Norrback, Johan (2021). "Pehr Strands flöjtur." Centrum för digital humaniora, Göteborgs universitet. URL: https://strand.dh.gu.se (version 1).

[4] von Rosen, A., Holgersson, H., Strömberg, M., Adler Sandblad, F. & Grehn, S. "Expansion och mångfald - en relationell forskningsdatabas". Expansion och mångfald: Digital kartläggning och analys av den utominstitutionella scenkonsten i Göteborg 1965-2000 – Institutionen för kulturvetenskaper, Göteborgs universitet, 2021.

[5] Butler, H., M. Daly, A. Doyle, Sean Gillies, T. Schaub, and Stefan Hagen. "The GeoJSON Format." Request for Comments. Internet Engineering Task Force, August 2016. https://doi.org/10.17487/RFC7946.