THE RD&I DOCUMENT SEARCH ENGINE OF ETIP-DG

Eugenio Trumpy[1], Gianluca Gola[1], Serena Botteghi[1], Anna Pellizzone[1], Pavel Sorin[2], Philippe Dumas[3], Valentina Pinzuti[3], Ben Laenen[4], Adele Manzella[1]

[1] CNR – IGG, Via Moruzzi 1, 56124 Pisa – Italy

[2] Penrose-cdb, Square de Meeûs 35, 1000 Brussels, Belgium

[3] European Geothermal Energy Council, Place du Champ de Mars 2, Brussels, Belgium

[4] VITO NV, Boeretang 200, 2400 Mol, Belgium

e.trumpy@igg.cnr.it

Keywords: deep geothermal, research, development, innovation, document repository, project deliverables, Europe.

ABSTRACT

An organized catalogue of deliverables of European projects has been created in the frame of the European Technology & Innovation Platform on Deep Geothermal (ETIP-DG) using Zenodo, in order to describe RD&I in the deep geothermal sector also from a historical perspective, and to retrieve all the necessary information for highlighting success stories and gaps.

A platform was then designed to provide a framework for access, retrieve and query his detailed and comprehensive collection of documents from past and actual RD&I projects and activities. The resulting search engine, the European Geothermal RD&I Document Search Engine EGRISE, is embedded in the ETIP-DG website and offers a public access.

Primary users of the EGRISE are the working group leaders and all the members of ETIP-DG working on the drafting of the strategic documents, in particular the Strategic Research Agenda and the Roadmap.

1. INTRODUCTION

With the aim of fostering geothermal energy development in Europe, strategic documents have been envisaged by the European Technology & Innovation Platform on Deep Geothermal (ETIP-DG) (Pinzuti et al, 2019). In order to describe the research, development and innovation (RD&I) in the deep geothermal sector also from a historical perspective, and to retrieve all the necessary information for highlighting success stories and gaps, a detailed and comprehensive collection of documents from past and actual RD&I projects and activities has been designed.

The main information, e.g. the Deliverables of European funded projects available on-line have been collected and an organized catalogue has been then created using Zenodo (Zenodo). Zenodo has been considered an ideal tool for the collection, since it is, a repository for research outputs, created by OpenAIRE (OpenAIRE web site) and CERN and funded by European Commission (EC) to provide a place for researchers to deposit their research products. For ETIP-DG's purpose, all collected documents have been described with metadata and uploaded in Zenodo, in a Community called 'Deep Geothermal'.

The collected public documents are discoverable by a dedicated search engine, which is embedded in the ETIP-DG website (ETIP-DG website). This web-based information has been completed by June 2018 both for the front-end and for the upload of about 350 documents.

The platform uses modern ICT technologies, and provides a framework for access, retrieve and query the documents collected for ETIP Deep Geothermal in Zenodo.

Primary users of the platform are the working group leaders and all the members of ETIP-DG working on the drafting of the strategic documents, in particular the Strategic Research Agenda and the Roadmap. Moreover, the access to the Geothermal Search Engine is guaranteed to general public (ETIP-DG non-members) interested in EU projects report discovery upon requests.

2. DATA COLLECTION

To implement the Geothermal Search Engine some tasks have been performed: 1) search and collection of documents; 2) definition of the rules to prepare documents metadata; 3) metadata creation and document upload. Here below the above mentioned activities are described.

2.1 Collection of documents

During the last years, many geothermal project funded by EU in the frame of different programmes (e.g., H2020, FP7, FP6, IEE, INTERREG, …) produced a large amount of reports. Reports are usually published and made available in the project websites. Search operations for recent geothermal projects were overall easy. This cannot be said for old (e.g. those funded in FP6 or FP5 programmes) and national or local projects

(e.g., those funded in INTERREG or national programmes).

The CORDIS web portal (CORDIS) was very useful since H2020, FP7 and sometimes FP6 project description were found. By using the text field with the 'geothermal' keyword all the geothermal projects were listed. The list was then refined highlighting the deep geothermal projects.

The search of geothermal projects was detailed with the help of KEEP web portal (KEEP). The KEEP portal is the only source of aggregated data regarding project and beneficiaries of European Union cross-border, transnational and interregional cooperation programmes among member states and between member States and neighbouring countries. KEEP portal covers the 2000-2006, 2007-2013 and 2014-2020 periods.

Project information were the mapped and stored in a summary spreadsheet. For each project the program, the acronym, the title, the number of available deliverables, the web site URL, the project lifespan, the starting date, ending date, number of contract, the coordinator, the EC grant and the total budget were registered.

Thanks to the collected project information each project website – if still on-line – was visited and the available project deliverables were downloaded and collected.

2.2 Rules for metadata compilation

The homogeneous uploading and metadating process required to define a way of work to facilitate the operations. The main issue in metadata instruction was the choice of keywords, since they were strategic for the type of analysis performed for ETIP-DG. We established to use at least 4 keywords defined as follows: i) a fixed keyword 'geothermal energy'; ii) to insert at least one keyword chosen among the following categories: a) exploration; b) drilling; c) production; d) surfaces systems/generation; e) non technical; f) environmental; iii) if one keyword is 'exploration' to choose another keyword among the following topics: a) geochemistry, b) geophysics, c) structural geology, d) hydrogeology, e) geochronology, f) volcanology; iv) if one keyword is 'non technical' to choose another keyword among the following topics: a) social aspects, b) codes & definitions, c) energy sector status, d) skills education & training, e) research roadmaps, f) regulatory aspects, g) economics, h) financing; v) to insert a keyword with the name of the location if the document is referred to a particular place; vi) to insert the country if the document is referred to a particular nation. It was possible to assign more than one category to one document at point ii) and more than one topic referred to point iii) and iv).

2.3 Document reusability

During this work, there were concerns about the re-usability of the documents produced by the EU geothermal projects. Although the EC recommends (Open Access manual) the funded research projects to deliver public and reusable results, it is not completely clear how the documents resulting from projects can be collected and stored for different purposes beyond the aim of the single project.

To assure the maximum transparency of ETIP-DG activity, all the coordinators of the projects who produced the collected documents were contacted by mail, informed about the activity and invited to participate to ETIP-DG. They were given the possibility to remove the documents from the Zenodo community if they considered improper to store the deliverables beside the running project websites. Up to now none of them requested such drop.

3. THE EGRISE APPLICATION

The European Geothermal RD&I Document Search Engine (EGRISE) is a web based application that use the Zenodo repository as back-end to store and describe documents with metadata and a dedicated web user interface as front-end to allow users to discover the archived project documents. Back-end-front-end connection is guaranteed by the application program interface (API) provided by Zenodo. By means of the API code the user requests for a document performed from the web user interfaces are sent to the Zenodo back-end, which interprets the API request and send back to the front-end the information on the discovered document (Fig. 1).

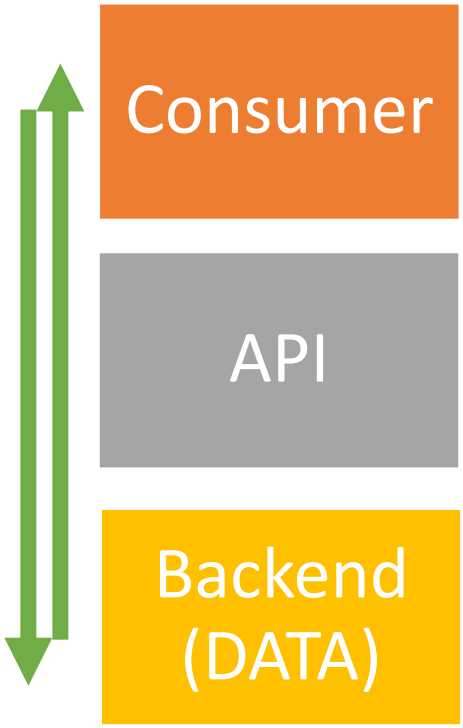

Figure 1: GSE architecture.

3.1 The Zenodo back-end

In Zenodo a digital object identifier (DOI) is automatically assigned to all Zenodo files, which can be then uploaded in any file format. In our case, if the original documents already had a DOI, the DOI assigned before uploading in Zenodo was the original

one; otherwise, all documents were flagged with a new DOI assigned by Zenodo. Data are stored in the CERN cloud infrastructure.

Zenodo is compliant with the open data requirements of Horizon 2020, the EU Research and Innovation funding programme and OpenAIRE, the EC-funded initiative in support of the OA policies of the European Union (Zenodo web site).

Moreover, Zenodo allows to organize own repositories by creating communities and identifies grants to be related to documents on research funded by the EC via OpenAIRE.

Eventually Zenodo is compliant to the DataCite Metadata schema and follow the FAIR principles (i.e., Findable, Accessible, Interoperable, Reusable) (Wilkinson, et al. 2016).

3.2 The EGRISE front-end

The EGRISE web user interface is embedded in the ETIP-DG website. It was developed by Penrose-cdb by using the API provided by Zenodo to allow the interaction with the back-end and by taking into account the style and layout of the ETIP-DG web site to guarantee a common look and feel.

General public can access EGRISE directly from the main menu of the ETIP-DG website. The users are requested to provide some general information (i.e., First name, Family name, email address, country, reason for requesting access) and to declare that the data retrieved will not be misused before the first access. An email informs the applicant that the request has been accepted and shows a link to a page for creating the credentials to log-in and access the Geothermal Search Engine.

The main access to the platform is however designed for ETIP-DG members, in the private area of the ETIP-DG website (on the Members Area Menu). These are not requested for credentials and the access is direct.

3.3 Using the EGRISE

The search engine appears as in Figure 2. Documents can be retrieved using the general search tool (“Enter your search term” in Fig. 2), which searches for words in all fields of the metadata used to describe the documents uploaded in Zenodo (e.g., Authors, document Title, Project name, Funding Program, Abstract and all the abstract content). The search is case insensitive.

With the help of the keywords, documents have also been categorized using the ETIP-DG Working Group Titles (i.e., exploration, drilling, production, surface systems / generation, environmental and non technical), so that they can be easily filtered out by category. Other ways to filter documents is by Funding Programme, Project Title and Documents type.

When a document is chosen, it can be explored by clicking on its title. The exploration brings to the Summary Page of the document (Fig. 3) where the user has access to the main metadata of the document, including its keywords, its abstract and may download the pdf of the complete document.

The Platform is meant to be up and running for the entire lifetime of the ETIP-DG website.

The collection of documents will be updated during the DG-ETIP project, following also the hints from WG activities.

WIDE FIGURE

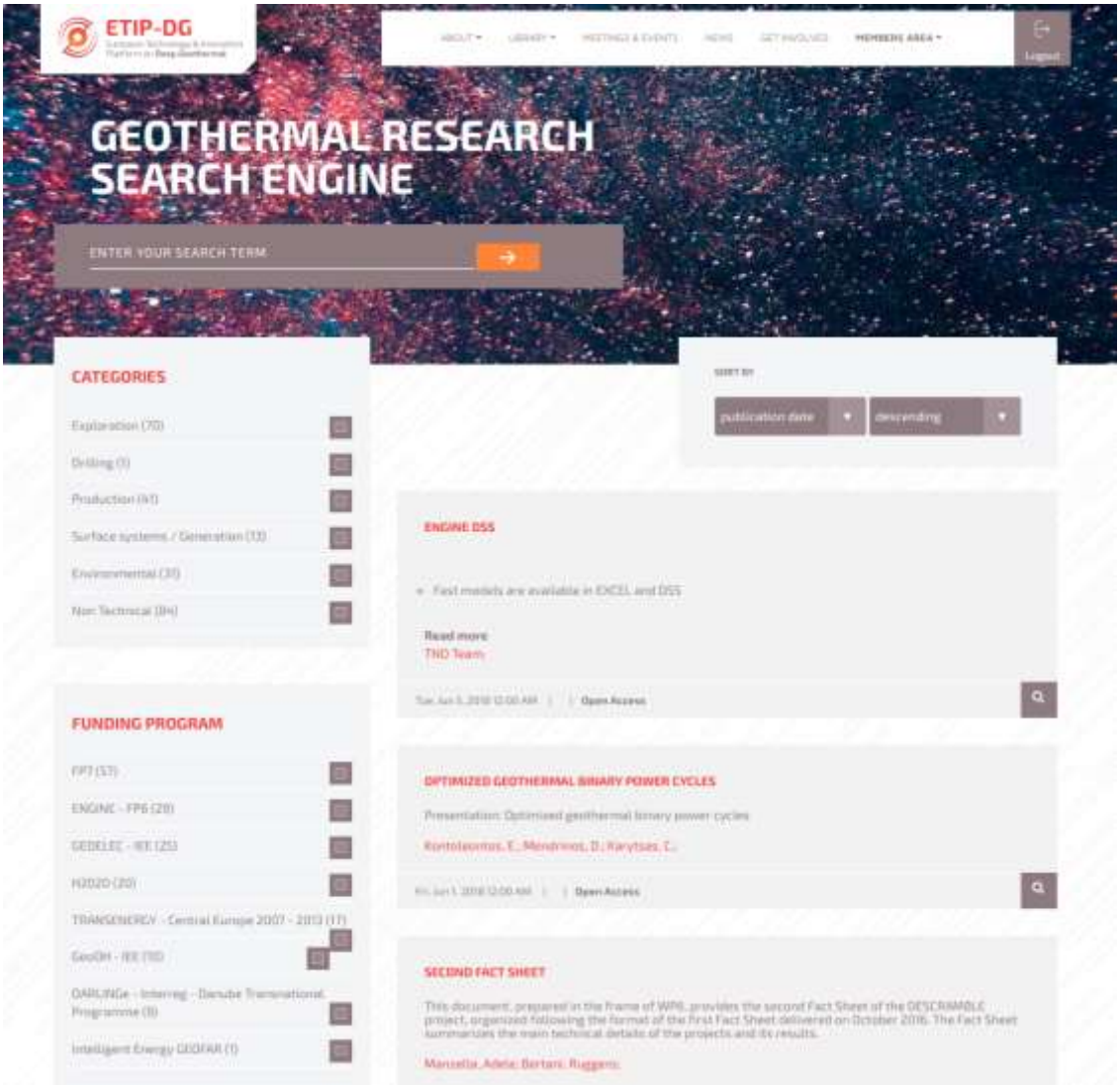

Figure 2: Screenshot of the upper part of the search engine. Categories and funding programmes filters are shown on the left side.

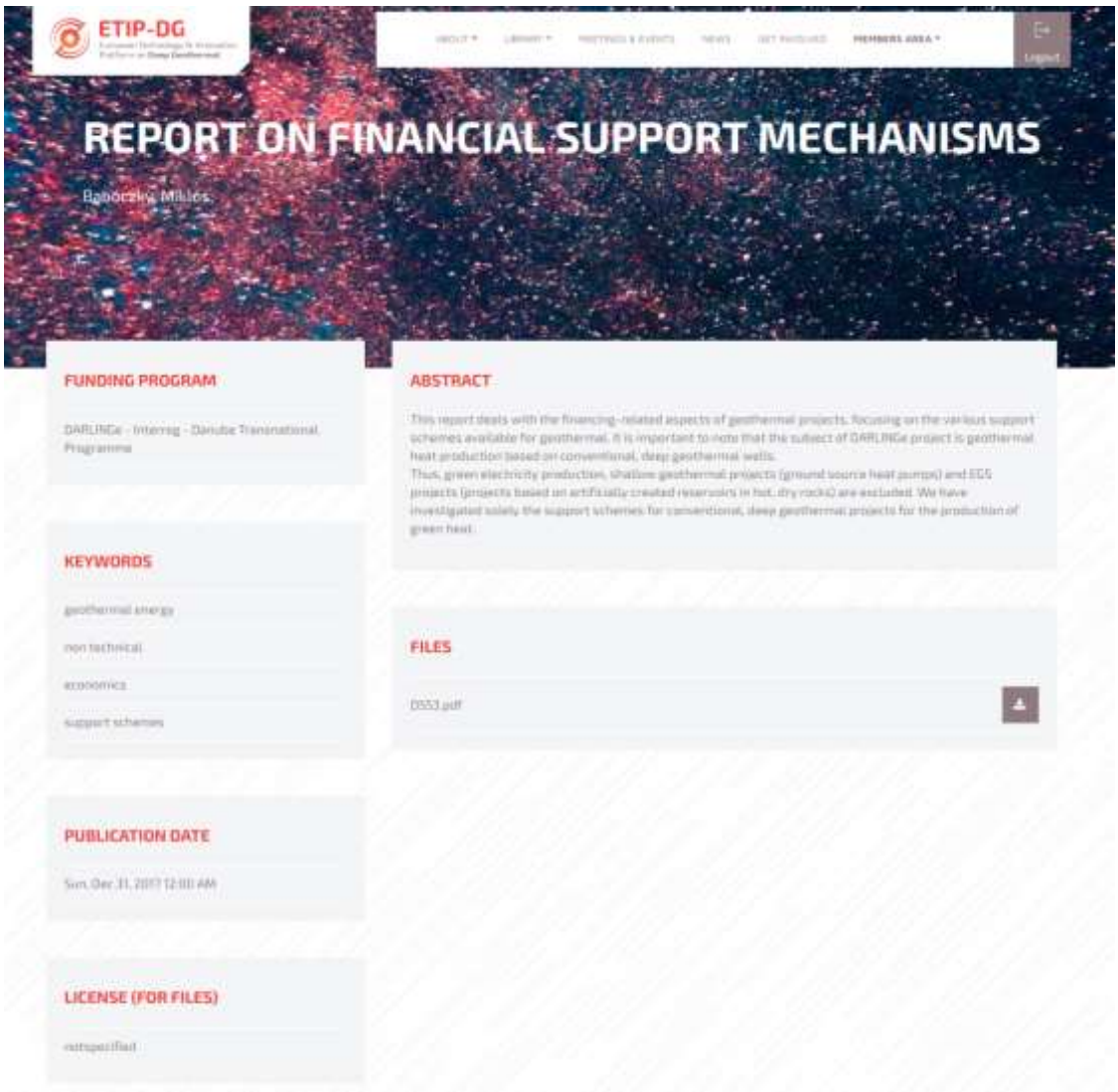

Figure 3: Screenshot of the Summary Page for a document, once a document is selected.

4. CONCLUSIONS

The collected documents in EGRISE are in this way enriched and enhanced with time. The information organised and offered from a single platform can be used for different purpose as drafting strategic documents, prepare project proposal, to assess the geothermal sector's state-of-the-art for different topics in order to foster the research, development and innovation or for scientific papers. EGRISE is the first tool which aims to offer all the accessible geothermal knowledge produced by EU projects for different kind of stakeholders.

Acknowledgements (optional)

This proceeding presents results of the DG-ETIP Project, funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No. 773392.

REFERENCES

CORDIS web portal, https://cordis.europa.eu/projects/en (04 March 2019)

ETIP-DG website, http://www.etip-dg.eu (04 March 2019)

KEEP web portal, http://www.keep.eu (04 March 2019)

Open Access Manual, http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/open-access_en.htm (04 March 2019)

OpenAIRE website, http://www.openaire.eu (04 March 2019)

Pinzuti, V., Dumas, P., Garabetian, T., Manzella, A., Trumpy, E., Laenen, B. and Lagrou, D., European Technology and Innovation Platform on Deep Geothermal, A presentation. EGC 2019, Den Haag, The Netherlands, 11-14 June 2019.

Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 3:160018 doi: 10.1038/sdata.2016.18 (2016)

Zenodo, http://zenodo.org (04 March 2019)

Zenodo API, http://developers.zenodo.org/ (04 March 2019)