Geophysical Research Abstracts
Vol. 20, EGU2018-17425, 2018
EGU General Assembly 2018


# How to deal with old and industrial data? The example of the Data Center for Deep Geothermal Energy (CDGP)

Marc Schaming (1), Alice Fremand (2), Nicolas Cuenot (3), Eleonore Dalmais (3), Jean-Francois Girard (1), Clement Grellier (2), Marc Grunberg (2), and Jean Schmittbuhl (1)

(1) Université de Strasbourg, CNRS, IPGS UMR7516, F-67000 Strasbourg, France (marc.schaming@unistra.fr), (2) Université de Strasbourg, CNRS EOST UMS830, F-67000 Strasbourg, France, (3) ES-Géothermie, 5 rue de Lisbonne F-67300 Schiltigheim, France

Old and industrial data are useful datasets, but because of their specific status, they bring new challenges for their diffusion: recovering them and dealing with Intellectual Property Rights (IPR) are often big difficulties. The Data Center for Deep Geothermal Energy (CDGP, https://cdgp.u-strasbg.fr) set up by the LabEX G-EAU-THERMIE PROFONDE has developed specific procedures to preserve, archive and diffuse old and industrial data acquired on geothermal sites of the Upper Rhine Graben. It is now a local node for the European Plate Observing System (EPOS) Anthropogenic Hazard platform (https://tcs.ah-epos.eu).

Data processed at CDGP mainly consist of seismological and hydraulic data that have been acquired during stimulation or circulation phases at Soultz-sous-Forêts pilot plant. They are gathered into "episodes": time-correlated collections of geophysical, technological and other relevant geo-data over a geothermal area.

Dealing with old data becomes an issue if datasets are not regularly processed. With time, media, formats, software become obsolete and information may be lost. The first episode, now available on the CDGP and EPOS platforms, corresponds to 1993 stimulation. The time-consuming process to recover approximately 70 percent of the dataset included transfer of data from old media to modern ones, conversion of data into standardized formats, comparison to old literature as part of quality-control process, and addition of documentation and metadata. Now new users give interesting feedbacks, contributing to the completeness of the dataset.

Industrial data bring new challenges in terms of IPR. Indeed, even if the share of datasets for research purpose is very interesting for the development of the geothermal industry, information about the geothermal resources and specific expertise in the geothermal field may need certain confidentiality for business reasons. Agreements with the industrial partners allow the CDGP to distribute sensible data to at least the academic community. Therefore, specific terms of use and procedures have been set up: an Authentication, Authorization and Accounting Infrastructure (AAAI) ensures the good distribution of data according to IPR, user's affiliation (i.e. academic, industrial, . . . ) and distribution rules, either automatically or after approval from the data owner.

The lessons learnt in terms of data management are manifold and most of the issues could have been overcome by setting data maintenance and preservation rules (described in a data Management Plan). For instance, defining the status of data, the owners, the embargo period and the distribution rules before the acquisition of the datasets could have saved a lot of time. These defined rules are mandatory to allow industrial partners having confidence in the data center. Moreover, the key to preservation is distribution by keeping datasets alive. Documentation, use of standard formats and data description in metadata are crucial for data re-use and preservation. Procedures to process new datasets at CDGP have been set up following these simple recommendations.