Proceedings World Geothermal Congress 2020+1
Reykjavik, Iceland, April - October 2021

Initiating FAIR geothermal data in Indonesia

Dasapta E. Irawan, Arif Susanto, Suryantini, and Niniek R. Herdianita

Faculty of Earth Sciences and Technology, Institut Teknologi Bandung

r-win@office.itb.ac.id

Keywords: open data, FAIR data, data sharing

ABSTRACT

One of the main keys to scientific development is data availability. Not only is the data easily discovered and downloaded, there's also needs for the data to be easily reused. Geothermal researchers, research institutions and industries are the three main stakeholders to foster data sharing and data reuse. Very expensive deep well datasets as well as advanced logging datasets are very important not only for exploitation purposes but also for the community involved, e.g., for regional planning or common environmental analyses. In data sharing, we have four principles of F.A.I.R data. Principle 1 Findable: data uploaded to open repository with proper data documentations and data schema, Principle 2 Accessible: removed access restrictions such as user id and password for easy downloads. In case of data from commercial entities, embargoed data is permitted with a clear embargo duration and data request procedure, Principle 3 Interoperable: all data must be prepared in a manner for straightforward data exchange between platforms, Principle 4 Reusable: all data must be submitted using common conventional file format, preferably text-based file (e.g., `csv` or `txt`) therefore it can be analyzed using various software and hardware. The fact that geothermal industries are packed with for-profit motivations and capital intensive would give even more reasons to embrace data sharing. It would be a good way for them to share their role in supporting society. The contributions from multiple stakeholders are the most essential part in science development. In the context of the commercial industry, data sharing is a form of corporate social responsibility (CSR). It shouldn't be defined only as giving out funding to support local communities.

1. INTRODUCTION

One of the main keys to scientific development is data availability[1]. Not only is the data easily discovered and downloaded, there's also a need for the data to be easily reused[2]. Geothermal researchers, research institutions and industries are the three main stakeholders to foster data sharing and data reuse[3,4], as subsurface data are very expensive to get, especially deep well datasets. However, such data are not only needed for geothermal exploration and exploitation itself, but they are also important for regional planning or common environmental analyses. Provincial and local level governments are lacking thematic spatial datasets to help them make regional development plans that could lead to disasters, e.g., planning a settlement area which is prone to tsunamis or has potential of mud volcano build ups.

To tackle such problems, a data sharing policy is needed. In this paper, we are proposing a model of data sharing among geothermal stakeholders, to bring more benefit for the society. Table 1 shows examples of data types, funders, and beneficiaries.

Table 1: Examples data types, funders, direct and extended beneficiary of the data

No	Example of data types	Funders	Direct beneficiary	Extended beneficiary
1	Geological maps	Government	related governmental bodies	wider governmental bodies
		Private companies	private companies who give the funding	provincial and/or local level government
				provincial and/or local level government
				educational sector
				common public
2	Drilling data	Private companies	private companies who give the funding	national/provincial/local level government
				other private companies
				educational sector
				common public
3	Water quality data	Government	related governmental bodies	wider governmental bodies

		Private companies	private companies who give the funding	provincial and/or local level government
				provincial and/or local level government
				educational sector
				common public

2. OVERVIEW

2.1 The basics of F.A.I.R data

In data sharing, we have four principles of F.A.I.R data[56]. Principle 1 Findable: data uploaded to open repository with proper data documentations and data schema, Principle 2 Accessible: removed access restrictions such as user ID and password for easy downloads. In case of data from commercial entities, embargoed data is permitted with a clear embargo duration and data request procedure, Principle 3 Interoperable: all data must be prepared in a manner for straightforward data exchange between platforms, Principle 4 Reusable: all data must be submitted using common conventional file format, preferably text-based file (e.g., `csv` or `txt`) therefore it can be read using general software on minimum specification hardware. Those four principles should be converted into technical guidance which is simple and easy to understand and to follow.

Based on our short evaluation on Garuda, Indonesia's Scientific Database created by the Ministry of Ristekbrin, generally there are only a very small portion of documents providing FAIR data. We are still conducting a detailed evaluation (**Figure 1**) to provide numbers.

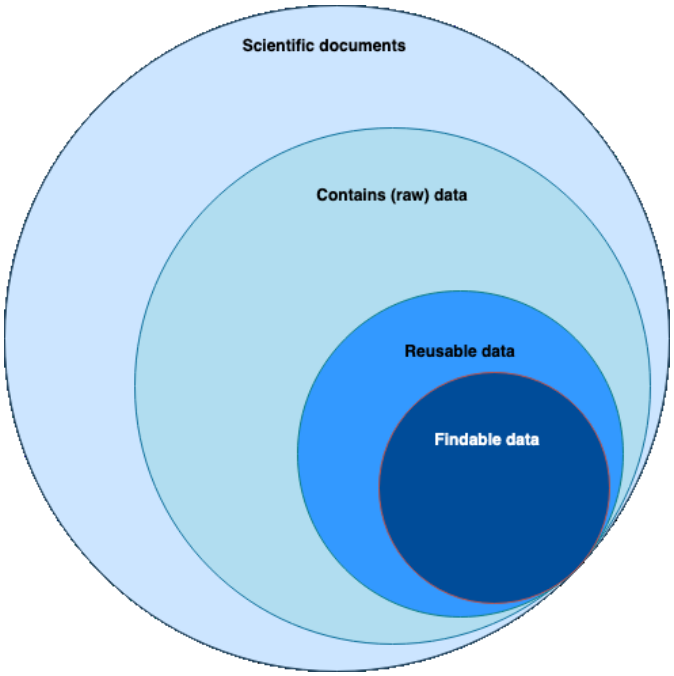

Figure 1: The general situation of data sharing in Indonesia

2.2 Open access repositories

In the field of geoscience there are notably open access repositories, that are usable to store data as well (see **Table 2**). The full list can be visited in several repository aggregators like r3data:
(https://www.re3data.org/search?subjects[]=34%20Geosciences%20(including%20Geography))
and OpenDOAR:
(https://v2.sherpa.ac.uk/cgi/search/repository/advanced?screen=Search&repository_name_merge=ALL&repository_name=&repository_org_name_merge=ALL&repository_org_name=&content_types_merge=ANY&content_subjects=6&content_subjects_merge=ANY&org_country_browse_merge=ALL&org_country_browse=&satisfyall=ALL&order=preferred_name&_action_search=Search). Some of the repositories are general repositories (for any fields of science) and some of them are subject specific. Aside from the following public repositories, please note that geothermal companies usually also have their own closed dataset.

Based on the types of submission and access we can find various combinations depending on the funding sources of a repository. For repositories funded by the government they would be likely to open submission and open access for the public. Variation occurs

when the funding comes from the private sector. Only members of the organization can submit documents. They can then limit the access to their members or open the access for the public (**Figure 2**).

Table 2: Examples of general and subject specific open access repositories

No	Name of repository	Country	Access	General/specific	Link to portal	Lead organization	Organization's website
1	International Earth Data Alliance	USA	OA	Specific	https://www.iedadata.org/	Lamont-Doherty Earth Observatory	https://www.ldeo.columbia.edu/
2	Earth and Space Science Open Archive	USA	OA	Specific	https://www.essoar.org/	American Geophysical Union	https://www.agu.org/
3	Repositori Ilmiah Nasional (RIN LIPI)	Indonesia	OA	Specific	http://rin.lipi.go.id/	LIPI	http://lipi.go.id/
4	Goddard Earth Sciences Data and Information Services Center	USA	OA	Specific	https://disc.gsfc.nasa.gov/	NASA	https://nasa.gov
5	National Oceanography Centre	UK	OA	Specific	https://www.bodc.ac.uk/about/what_is_bodc/	British Oceanic Data Centre	https://www.bodc.ac.uk/
6	Pangea	Germany	OA	Specific	https://www.pangaea.de/	Alfred Wegener Institute, Helmholtz Center for Polar and Marine Research (AWI) and the Center for Marine Environmental Sciences, University of Bremen (MARUM)	https://www.awi.de/
7	4TU.ResearchData	The Netherlands	OA	Specific	https://data.4tu.nl/	4TU.Federation	https://www.4tu.nl/en/
8	Zenodo	Switzerland	OA	GENERAL/specific	https://zenodo.org	CERN	https://home.cern/
99	National Geothermal Data System	USA	OA	Specific	https://data.geothermaldata.org/	US Department of Energy US Geological Survey Southern Methodist University	https://www.geothermaldata.org/
10	Geothermal Data Repository	USA	OA	Specific	https://gdr.openei.org/home	US Department of Energy	https://www.energy.gov/

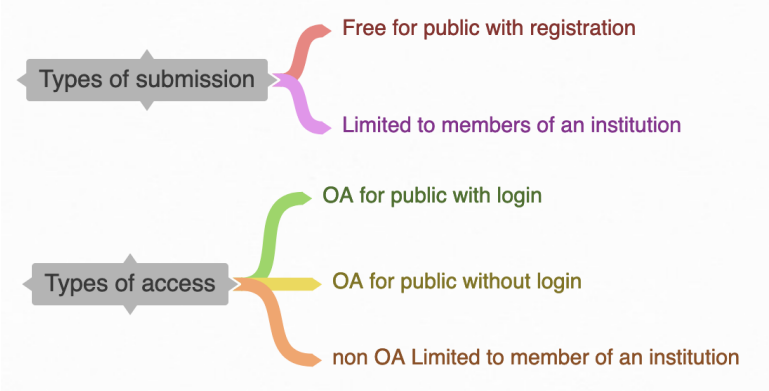

Figure 2: The categories of earth science/geoscience repositories based on submission and access

2.3 Mapping of stakeholders in geothermal research and development in Indonesia

In this paper we only discuss a few stakeholders that are directly involved in research and development of geothermal energy in Indonesia (**Figure 3**). These stakeholders are generally available in most geothermal resource producing countries. It should be noted that every geothermal and power industry maintains an exploration and development department that also produces datasets. Indonesia's government at national and provincial levels also has research institutes that usually collaborate with academic and independent research institutions to work on geothermal exploration and/or exploitation.

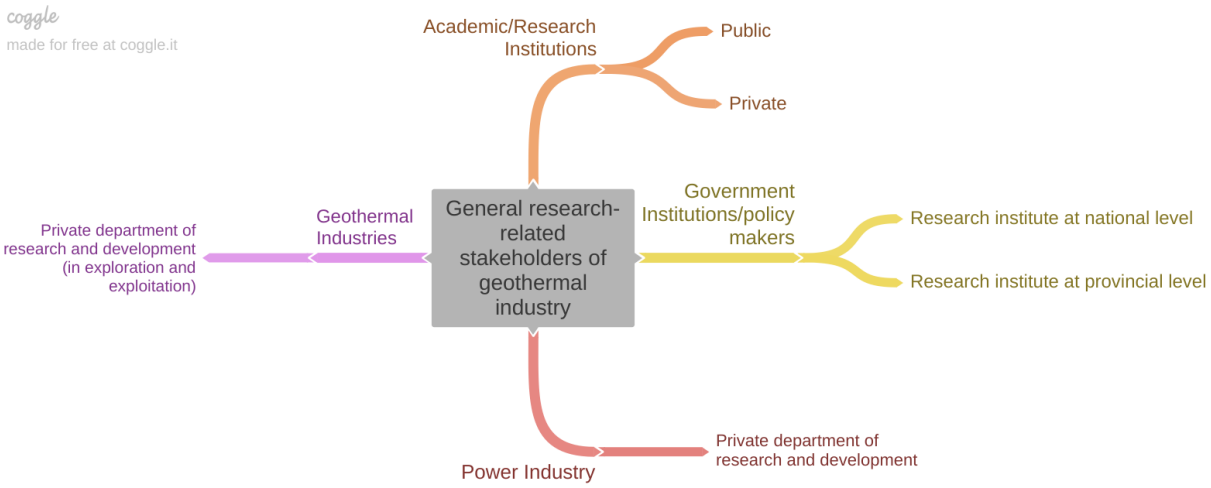

Figure 3: The general stakeholders of geothermal industry

3. PROPOSED IMPLEMENTATION

3.1 Flow of data

We propose the following flow of data:

1. Datasets from governmental bodies:
    a. storage location:
        i. stored in an internal repository or
        ii. stored in a public repository managed by government institution (e.g., Repositori Ilmiah Nasional - RIN which is managed by PDDI-LIPI)
    b. status of the datasets: immediately released to public as open access data
2. Datasets from private sectors:
    a. storage location:
        i. stored in an internal repository or
        ii. stored in a public repository managed by government institution (e.g., Repositori Ilmiah Nasional - RIN which is managed by PDDI-LIPI)
        iii. stored in an internal repository and then sent to public repository after passing the embargo period (see status of datasets)
    b. status of the datasets:
        i. embargoed: the datasets are opened limited to members of the organization and released to public after passing the embargo period (usually in two years period),
        ii. immediately open access: the datasets are immediately released to the public.

Figure 4 summarizes the types of submission and access that we could apply to a repository. This would largely depend on the policy of the data creator, especially the economic implications of the data when it is shared publicly online, for example, in exploration of new potential locations.

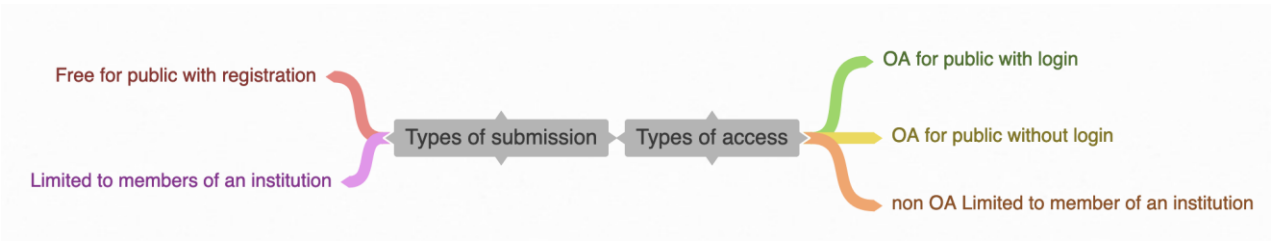

Figure 4: The types of submission and access of data repositories

3.2 Resources needed

There are at least four resources that should be made available upon creating a data repository. The first resource would be a proper Research Data Management Plan/Policy. There are many references, even templates to create this document. However the members of an organisation should agree to the plan and they're able to propose changes in later development (**Figure 5**). The policy would then be converted to the funding allocated in creating the repository. The funding would be important in the determination of storage size[7], selection of software (proprietary or open source)[8], and manpower to manage the data after submission (data stewardship)[9]. However, we would recommend the following principles: to create a simple policy, to use common hardware specifications and open source software, and to maximum dissemination.

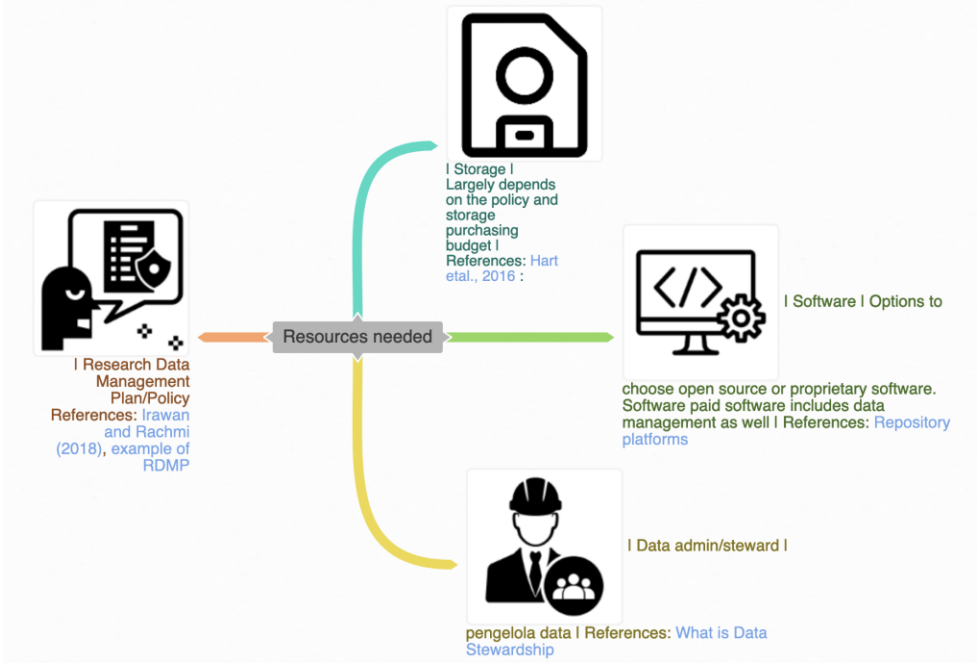

Figure 5: Resources need in creating a data repository

3.3 Potential barriers

Several studies have mentioned the barriers and drivers in data sharing[10] [11] [12] [13] [14], including capturing views from geoscientists[15]. The results barriers are generally the same between different fields of science. Most of scientists are questioning the following aspects: data ownership, long term preservation, legal issues (misuse of data), incompatible data types, data scooping, more to do but lack of incentives, which one to publish first (paper or data), don't know why and how they do it. We lack the potential drivers to overcome the barriers. However, those above-mentioned studies have also mentioned education and training as the most important solution. Such training should introduce and develop new perceptions in the scientific community: **(1)** That data should be published independently from the report or research papers. **(2)** To add some textual documentation, researchers could publish data papers. Papers about data could be presented from ideas, data collecting, or data processing. Such a decision would enrich the context of the company report or published paper itself. **(3)** The publication of research data would increase the value of the data that in the end would add the value of the institution to common society. This would move us closer toward data sharing culture in geoscience[16].

7. CONCLUSIONS

The fact that geothermal industries are packed with for-profit motivations and capital intensive would give even more reasons to embrace data sharing. It would be a good way for them to share their role in supporting society. The contributions from multiple stakeholders are the most essential part in science development. In the context of the commercial industry, data sharing is another impactful form of corporate social responsibility (CSR) in addition to the tradition giving away money; it shouldn't be defined only as giving out funding to support local communities.

REFERENCES

1. "Data sharing in PLOS ONE: An analysis of Data Availability ...." 2 May. 2018, https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0194768. Accessed 5 Jan. 2021.
2. "The FAIR Principles - MIT Press Journals." https://www.mitpressjournals.org/doi/pdf/10.1162/dint_e_00023. Accessed 5 Jan. 2021.
3. "Determining barriers to developing geothermal power ...." https://www.sciencedirect.com/science/article/pii/S0301421513004394. Accessed 5 Jan. 2021.
4. "Geothermal barriers, policies and economics in Chile ...." https://www.sciencedirect.com/science/article/pii/S1364032115006486. Accessed 5 Jan. 2021.

5. "Make scientific data FAIR - Nature." 4 Jun. 2019, https://www.nature.com/articles/d41586-019-01720-7. Accessed 5 Jan. 2021.
6. "Promoting data sharing among Indonesian scientists: A ...." 6 Jul. 2018, https://riojournal.com/article/28163/. Accessed 5 Jan. 2021.
7. "Ten Simple Rules for Digital Data Storage - NCBI - NIH." 20 Oct. 2016, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5072699/. Accessed 9 Jan. 2021.
8. "Repository Platforms - Research Data Network." https://research-data-network.readme.io/docs/repository-platforms. Accessed 9 Jan. 2021.
9. "What is Data Stewardship? - DATAVERSITY." 6 Nov. 2017, https://www.dataversity.net/what-is-data-stewardship/. Accessed 9 Jan. 2021.
10. "4239b46defca4705b5b3b5f39faad144.pdf - Amazon AWS." https://library.oapen.org/bitstream/handle/20.500.12657/28008/1001989.pdf?sequence=1. Accessed 9 Jan. 2021.
11. "Baseline Report on Drivers and Barriers in Data Sharing | EPIC." https://epic.awi.de/30335/. Accessed 9 Jan. 2021.
12. "Sharing data: Practices, barriers, and incentives - Tenopir ...." https://asistdl.onlinelibrary.wiley.com/doi/full/10.1002/meet.2011.14504801026. Accessed 9 Jan. 2021.
13. "The Research Data Alliance: globally co-ordinated action ...." https://onlinelibrary.wiley.com/doi/abs/10.1087/20140503. Accessed 9 Jan. 2021.
14. "Build It, But Will They Come? A Geoscience ...." 11 Jul. 2016, https://datascience.codata.org/articles/10.5334/dsj-2016-008/. Accessed 9 Jan. 2021.
15. "Capturing the views of geoscientists on data sharing: a focus ...." 10 Nov. 2020, https://qjegh.lyellcollection.org/content/early/2020/11/06/qjegh2019-138. Accessed 9 Jan. 2021.
16. "Toward the Geoscience Paper of the Future: Best practices for ...." 26 Jul. 2016, https://agupubs.onlinelibrary.wiley.com/doi/full/10.1002/2015EA000136. Accessed 9 Jan. 2021.