

# Reusable data is the new oil

*Dasapta Erwin Irawan<sup>1\*</sup>, Yuniarti Ulfa<sup>1</sup>, Astyka Pamumpuni<sup>1</sup>, Indra Andra Dinata<sup>1</sup>, Thomas Tri Putranto<sup>3</sup>, and Hari Siswoyo<sup>3</sup>*

<sup>1</sup>Faculty of Earth Sciences and Technology, Institut Teknologi Bandung, Jalan Ganesa No. 10, Bandung

<sup>2</sup>Faculty of Engineering, Universitas Diponegoro, Semarang

<sup>3</sup>Faculty of Engineering, Universitas Brawijaya, Malang

**Abstrak.** New oil data is 21st century jargon. This movement has not been widely echoed in Indonesia. Although some initiatives should be recognized and appreciated, the status of the availability of reusable data in most countries, especially in Indonesia is still low. Most of the data published in Indonesian open access journals are in the form of pdf files that cannot be reused. We advise editors of Indonesian scientific journals to consider adopting FAIR data sharing by encouraging authors to share their data as additional files in a machine readable format, e.g. csv or xls. This effort will also contribute to the principles of transparency and sustainable development in Indonesia's research ecosystem.

## 1. Introduction

Data is new oil [1]. In conducting research, data is the main component. However, data may not be a significant component of published papers [2-4]. Most of the data is shared as non-reusable PDF files along with the descriptive text. The data itself cannot be directly reused for further study. Other researchers have to retype the data to research and develop it. This article explores the importance of reusable data, how to share it, and evaluates the availability of reusable data in scientific papers published in Indonesian and international earth science journals.

## 2. Reviews in FAIR Data

The focus on data sharing has shifted from data tables provided in a scientific article to FAIR data tables. FAIR here is an acronym for Findable, Accessible, Interoperable and Reusable [5].

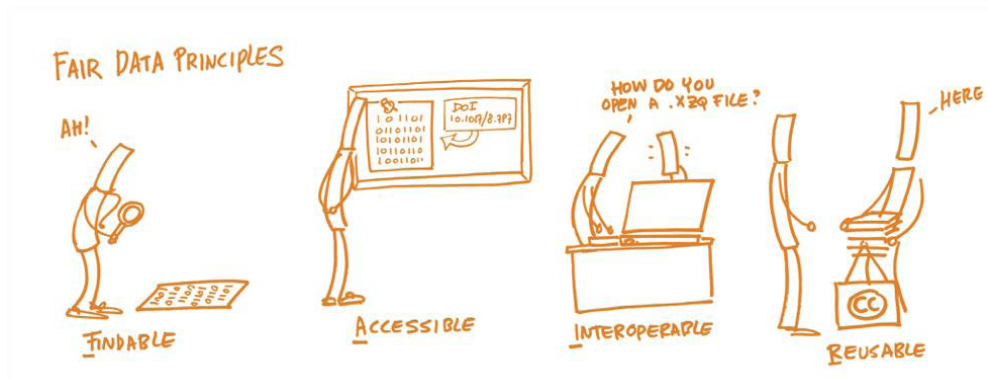


Fig. 1. A cartoon describing FAIR data sharing[4].

## 2.1 Findable

Data needs to be found by search engines, either general web search engines or specific ones looking for scientific material [5]. Standard search engines here, for example, are Google or Bing, while specific search engines, for example, are Google Scholar, Microsoft Academic, or Lens. The main requirement for a document to be found online is complete and openly uploaded metadata.

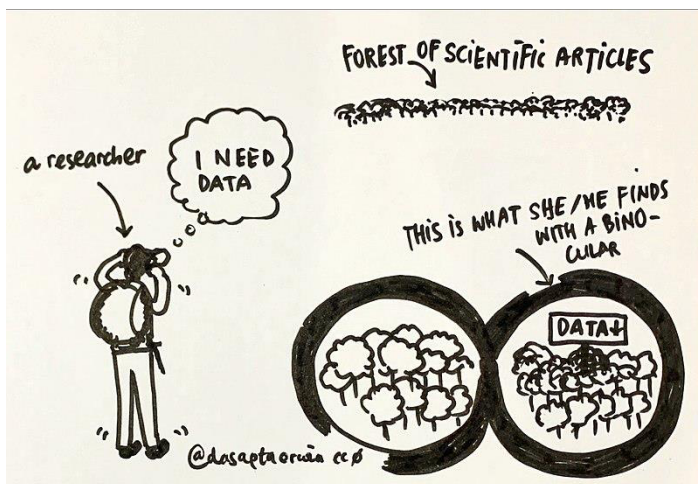


Fig. 2. A cartoon describing the problem in finding earth science data[6].

## 2.2 Accessible

Data needs to be able to be downloaded and opened without unnecessary authorization. Often, we share files via Google Drive links that are password protected. Protection is indeed required for certain types of data, for example, data that does not have a good data usage agreement, as well as sensitive data [5].

## 2.3 Interoperable

Data needs to be interoperable using various applications on various platforms; for example, data needs to be opened by applications running on Linux, Mac, and Windows operating systems. Here the file format should be generic and can be opened by various applications [5]. An example that we can give here is a table format which generally uses the xls or xlsx format. Although Microsoft Excel has become the standard nowadays, keep in mind that binary format files such as xls or xlsx can only be opened using specific applications, which also depends on the application's version. The recommended data format is non-binary or text, so it can be opened using various applications, such as txt or csv formats.

## 2.4 Reusable

Data needs to be shared in a way that makes it easy for others to reuse [5]. The effort expended needs to be minimal; for example: in geoscience, maps shared only in raster format (e.g. jpg and png) will not be reusable by readers. Additional efforts in the form of replotting are needed.

## 3. Materials and Methods

We explored the status of data availability in three high profile national journals in the field of geosciences listed in the SINTA database. Steps: 1) visit the site [sinta2.ristekbrin.go.id](http://sinta2.ristekbrin.go.id), 2) open the "Journals" menu, 3) use the keyword "geo". We got three journals: the Indonesian Journal on Geosciences (IJOG), Geoplanning: Journal of Geomatics and Planning (GGP), and the Indonesian Journal of Geography (IJG).

We compared the results with three international journals in the field of earth sciences that we got from the Scimago Journal Ranking list. Steps: 1) visit [scimagojr.com](http://scimagojr.com), 2) activate the filter subject areas "earth and planetary sciences: and filter the subcategories "earth and planetary sciences (miscellaneous)". List of three journals: 1) Nature Geoscience (NG), 2) Annual Review of Earth and Planetary Sciences (AREPS), 3) Earth System Science Data (ESSD).

The observations are stored as supplementary files in csv and xls formats (Table 1). Please refer to the Data Availability section.

## 4. Results and discussions

### 4.1 Results

The three national journals are high-profile earth science journals entered in the Sinta 1 category. All three have been entered into the DOAJ database, but only IJOG (Indonesian journal) whose papers have been entered into the Scopus database. All national journals are open access journals, published in hybrid, electronically, and in print. Non-profit institutions also publish all three in the form of universities and state institutions. Even though the journals are open access journals, only one journal charges an article processing fee (APC), namely IJG of USD 300; the other two journals do not charge an article processing fee (APC) at all.

In the writing guide, the three national journals do not have any data sharing policy. Based on a random review of several publications, we found no articles with supplementary data. All data is packed in PDF format along with texts.

We compare those national journals with international journals, NG, AREPS, and ESSD. They are all high-profile journals and are recorded in the Scimago database (Quartile 1). Among the three, only NG is managed by commercial companies. It publishes articles in hybrid mode, offering both non-OA and OA routes to authors. The APC for the OA route is more than USD 11,000. AREPS and ESSD are association journals. ESSD is an OA journal with no APC, published by the European Geoscience Union (EGU) in collaboration with the Copernicus publishing company. At the same time, AREPS is an association journal published on a non-OA basis. Readers can read the articles by paying a subscription fee.

European publishers have been exposed more extensively to data sharing, especially with the massive open science movement in the European continent, e.g. with Plan S and Horizon2020.

**Table 1.** Dataset[7]

Journal title	Journal website	P-ISSN	E-ISSN	Publisher	Data sharing guidelines	Data as supplementary materials	Link to author guidelines	Operational	OA	APC (USD)
Indonesian Journal of Geosciences	<a href="https://ijog.geologi.sdm.go.id/index.php/IJOG">https://ijog.geologi.sdm.go.id/index.php/IJOG</a>	2355-9314	2355-9306	Indonesian Geological Agency	None	None	<a href="http://ijog.bgl.esdm.go.id/index.php/IJOG/about/submissions#authorGuidelines">http://ijog.bgl.esdm.go.id/index.php/IJOG/about/submissions#authorGuidelines</a>	not for profit	yes	free
Geoplanning: Journal of Geomatics and Planning	<a href="https://ejournal.undip.ac.id/index.php/geoplanning/">https://ejournal.undip.ac.id/index.php/geoplanning/</a>	None	2355-6544	Diponegoro University	None	None	<a href="https://ejournal.undip.ac.id/index.php/geoplanning/about/submissions#authorGuidelines">https://ejournal.undip.ac.id/index.php/geoplanning/about/submissions#authorGuidelines</a>	not for profit	yes	free
Indonesian Journal of Geography	<a href="https://jurnal.ugm.ac.id/ijg/">https://jurnal.ugm.ac.id/ijg/</a>	0024-9521	2354-9114	Gadjah Mada University	None	None	<a href="https://jurnal.ugm.ac.id/ijg/about/submissions#authorGuidelines">https://jurnal.ugm.ac.id/ijg/about/submissions#authorGuidelines</a>	not for profit	yes	300
Nature Geoscience	<a href="https://www.nature.com/ngeo/">https://www.nature.com/ngeo/</a>	1752-0894	1752-0908	Nature Publishing Group	Yes	Yes	<a href="https://www.nature.com/ngeo/for-authors/preparing-your-submission">https://www.nature.com/ngeo/for-authors/preparing-your-submission</a>	for profit	hybrid	11,275
Annual Review of Earth and Planetary Sciences	<a href="https://www.annualreviews.org/journal/earth">https://www.annualreviews.org/journal/earth</a>	None	0084-6597	Annual Reviews	Yes	Yes	<a href="https://www.annualreviews.org/page/authors/general-information#earth">https://www.annualreviews.org/page/authors/general-information#earth</a>	not for profit	no	None
Earth System Science Data	<a href="https://www.earth-system-science-data.net/">https://www.earth-system-science-data.net/</a>	1866-3508	1866-3516	European Geoscience Union - Copernicus Publishing	Yes	Yes	<a href="https://www.earth-system-science-data.net/submission.html">https://www.earth-system-science-data.net/submission.html</a>	not for profit	no	currently waived

## 4.2 Discussions

### 4.2.1 The importance of data sharing

Data sharing is essential because retrieving data costs money, and for the field of earth sciences, it can be costly; for example, the price of a single drill can reach millions of dollars. Costly prices will be more effective when used for more than one purpose. Another benefit is that the data can be used for different purposes; for example, drilling data from oil and gas companies serve as an exploration tool, but the same data can also

be helpful for local governments to find out the subsurface conditions of their area. Local governments cannot fund expensive subsurface data.

Another benefit is that the data can be appropriately managed. The open science guide is highly recommended to make research data management for each research activity. This step can make data management more planned, in which parts of the data can be shared, with whom, and which parts cannot be shared.

The next benefit is that the data is stored better and can be searched by the public[8,9]. Amid today's connected internet, online search engines are the most effective and efficient reminder engines. When data has been uploaded online, search engines can find it.

The following fundamental benefit is that by sharing research data, the verification process will be carried out by readers. The reader can detect the errors made by the researcher if the data is shared fairly. Furthermore, further researchers can also carry out further analysis or re-analysis of the data. They can even combine old data with new data to form complete time-series data. The analysis will be more in-depth. In the end, science will develop faster [10–12].

Based on our observation, the awareness of the data sharing policy of Indonesian national journals is low. More education to promote good governance of data sharing is needed.

#### 4.2.2 *The earth science is people's science*

Although geoscience is often perceived as a hard science because it deals with deep underground objects, earth science is one of the sciences related to the lives of many people. This situation will not change when scientists only communicate their knowledge with each other. *Earth science* is a science belonging to the community. Therefore, its delivery must be made as friendly as possible with them[13].

Data is one of the connecting points between this science and other sciences understood by the public. By applying the FAIR principle, the public can obtain geological data as quickly as they can get information on the price of essential commodities. With the advancement of the connected digital world, it is straightforward to do. Scientists can upload their data to places recognized by search engines, such as campus or office repositories such as the [National Scientific Repository](https://rin.lipi.go.id) managed by LIPI (<https://rin.lipi.go.id>) or open public repositories ([Zenodo](https://zenodo.org) – <https://zenodo.org>). Social media such as [ResearchGate](https://researchgate.net) (<https://researchgate.net>) may be used, subject to terms and conditions, relating to its status as a business product.

#### 4.2.3 *Data as public goods*

Data is a public good, primarily if the research is publicly funded. Its public nature is limited to data that is not sensitive. Privately funded research data may be excluded, although providing public data can improve the reputation of the private sector[14]. Data not supported by a data usage agreement is also prone to conflict when made public[13,15].

*Public data* is a form that fits perfectly with the goals of open science, which makes data easy to find for purposes of transparency and accountability. This principle is very relevant to the recent conflicts that stem from the lack of data disclosure[14].

Reluctance to disclosing data is generally due to reasons that are no longer relevant, such as fear of data being stolen and misused and fear of research ideas being taken by other researchers. Developing a research data management plan (RDMP) by selecting sensitive data can be overcome those fears[15]. Each field of science may have different characteristics concerning data sharing, data release, and data usage agreement.

Some journals currently require authors to provide the raw dataset in a repository and share the data under a public domain waiver license (CC0). One of the examples is the F1000R journal.

## 5. Conclusions

The description above can add to the argument that FAIR data is the new oil. Data can be fuel for future research. Just like oil, reusable data could spark more ideas that would give more benefits at large. Fear of scooping must not restrain the willingness to share data. Data sharing should be the default mode in research dissemination and more education in sharing sensitive data. It will trigger a more multidisciplinary approach to the same problem.

## Acknowledgment

We thank the Bandung Institute of Technology for financial support through the 2021 P3MI Research Fund. We also thank Mrs. Zulidyana D. Rusnalasari from Indonesian Journal Volunteers and Mr. Hendro Subagyo from PDDI LIPI for the help of valuable information and reviews.

## References

1. The Economist, Economist (2021)
2. L. M. Federer, C. W. Belter, D. J. Joubert, A. Livinski, Y.-L. Lu, L. N. Snyders, and H. Thompson, PLoS One **13**, e0194768 (2018)
3. E. M. Hart, P. Barmby, D. LeBauer, F. Michonneau, S. Mount, P. Mulrooney, T. Poisot, K. H. Woo, N. B. Zimmerman, and J. W. Hollister, PLoS Comput. Biol. **12**, (2016)
4. H. Brinken and T. Ross-Hellauer, *Open Science Training Handbook* (FOSTER Open Science, 2018)
5. A. Jacobsen, R. de Miranda Azevedo, N. Juty, D. Batista, S. Coles, R. Cornet, M. Courtot, M. Crosas, M. Dumontier, C. T. Evelo, C. Goble, G. Guizzardi, K. K. Hansen, A. Hasnain, K. Hettne, J. Heringa, R. W. W. Hooft, M. Imming, K. G. Jeffery, R. Kaliyaperumal, M. G. Kersloot, C. R. Kirkpatrick, T. Kuhn, I. Labastida, B. Magagna, P. McQuilton, N. Meyers, A. Montesanti, M. van Reisen, P. Rocca-Serra, R. Pergl, S.-A. Sansone, L. O. B. da Silva Santos, J. Schneider, G. Strawn, M. Thompson, A. Waagmeester, T. Weigel, M. D. Wilkinson, E. L. Willighagen, P. Wittenburg, M. Roos, B. Mons, and E. Schultes, *Data Intelligence* **2**, 10 (2019)
6. D. E. Irawan, *Data Sharing - Wikimedia Commons* (2021)
7. D. E. Irawan, *Figshare* (2021)
8. K. Baždarić, I. Vrkić, E. Arh, M. Mavrinac, M. G. Marković, L. Bilić-Zulle, J. Stojanovski, and M. Malički, PLoS One **16**, e0244529 (2021)

9. B. Fecher, S. Friesike, and M. Hebing, PLoS One **10**, e0118053 (2015)
10. J. Helly, H. Staudigel, and A. Koppers, *Geochem. Geophys. Geosyst.* **4**, (2003)
11. A. Hubner, *Data Sci. J.* **19**, (2020)
12. L.-Z. Miao, Z.-L. Li, J. Li, and C.-W. Yang, *Journal of Nanjing University of Posts and Telecommunications (Natural Science)* 01 (2012)
13. D. Kinkade and A. Shepherd, *Geosci. Data J.* **n/a**, (2021)
14. H. van Vlijmen, A. Mons, A. Waalkens, W. Franke, A. Baak, G. Ruiter, C. Kirkpatrick, L. O. B. da Silva Santos, B. Meerman, R. Jellema, D. Arts, M. Kersloot, S. Knijnenburg, S. Lusher, R. Verbeeck, and J.-M. Neefs, *Data Intelligence* **2**, 276 (2020)
15. D. E. Irawan, in *EGU General Assembly Conference Abstracts* (2021), pp. EGU21-14438
16. D. E. Irawan, D. Irawan, and C. Rachmi, *Research Ideas and Outcomes* (2018)