

Membuat data geosains lebih dapat dibagi (shareable) dan dirujuk (citeable)

Dasapta Erwin Irawan

Saat ini data merupakan bahan baku berputarnya seluruh kegiatan di dunia. Kegiatan ekonomi, perdagangan, juga dalam lingkup artikel ini adalah kegiatan riset, seluruh komponennya akan berotasi mengelilingi komponen data. Analisis tidak akan dapat dilakukan tanpa data yang valid. Artikel ini bertujuan untuk menambah wawasan tentang maghzaab *open science* (praktek keilmuan yang terbuka dan transparan), *open data* (praktek membuka akses data seluas-luasnya), dan *open access* (praktek menyediakan berbagai sumber daya ilmiah secara luas) yang menjadi *trend* akhir-akhir ini mengikuti *trend open source* pada dekade sebelumnya. Dalam pembahasan,

makalah ini juga memasukkan komponen repositori data sebagai komponen sentral.

Apakah Repositori Data Itu?

Menurut *Google Translate*, definisi "repository" adalah:

- *a place, building, or receptacle where things are or may be stored: "a deep repository for nuclear waste".*
- *a place in which something, especially a natural resource, has accumulated or where it is found in significant quantities: "accessible repositories of water".*
- *in COMPUTING, a central location in which data is stored and managed: "the metadata will be aggregated in a repository".*

Definisi yang paling sesuai menurut saya adalah yang ke-3 bahwa repositori data tidak hanya berkaitan dengan penyimpanan tetapi juga pengelolaannya. Di dalamnya ada praktek penggunaan ulang data (*data re-use*) dan analisis ulang data (*data re-analysis*).

Mengapa Perlu Membuka Akses Data

Ini merupakan pertanyaan penting. Setidaknya ada dua pertimbangan:

- *Pertimbangan substansial*: Bila data tersedia secara bebas dan dapat dianalisis ulang, maka makin banyak orang yang menguji validitas data dan validitas analisis. Bila ini terjadi ada dua kemungkinan dampaknya: hasil analisis kita akan didukung oleh hasil analisis orang lain, yang kedua, mungkin juga hasil pemikiran kita direvisi oleh orang lain. Bila jenis yang kedua terjadi, apakah ini akan mengurangi nilai saintifik makalah kita? Jawabannya "tentu tidak". Karena dunia saintifik adalah dunia yang obyektif, maka dokumen analisis kita tetap (yang telah dipublikasikan) akan tetap terekam dalam khasanah ilmiah, walaupun hasilnya diperbaiki oleh peneliti lainnya.
- *Pertimbangan ekonomis*: Bila orang lain dapat menggunakan ulang (*reuse*) data kita, maka biaya yang dikeluarkan untuk mengambil data tersebut akan memiliki benefit yang berlipat. Kemudian pada akhirnya nilai ekonomisnya akan terus bertambah.
- *Pertimbangan waktu*: waktu yang dihabiskan untuk mencari data akan dapat setara dengan

mencari data baru. Data yang mudah dicari menyebabkan waktu lebih efisien.

Berikut ini adalah simulasi pendek situasi yang kerap dihadapi dalam kasus pemetaan geologi. Seseorang akan memetakan kondisi geologi di suatu daerah dengan tujuan untuk mengetahui potensi sumber air. Maka ia pertama kali akan mencari peta geologi regional. Peta ini telah diterbitkan oleh Badan Geologi untuk seluruh wilayah Indonesia dalam skala 1:100.000. Skala ini tentunya masih terlalu kecil untuk keperluan Sang Peneliti. Ia dapat membelinya di UPT Perpustakaan Kantor Badan Geologi di Bandung. Pertanyaan pertama bila ia tidak tinggal di Bandung bagaimanakah ia dapat memperoleh peta tersebut secara resmi? Di sini masalah mulai timbul. Praktek yang biasa dilakukan adalah menggunakan jejaring di Bandung untuk membelikan peta dimaksud. Tentunya ini bukan praktek yang dapat diformalkan. Belum lagi kendala waktu pemesanan dan pengiriman. Pertanyaan kedua, bila peta-peta tersebut dibuat dengan anggaran negara, apakah peta tersebut perlu dihargai? Bukankah semestinya dokumen-dokumen tersebut bersifat terbuka untuk masyarakat (*public domain*) dan di era *high speed internet* ini apakah tidak sebaiknya tersedia *online*.

Apakah Data Dapat Dirujuk (*Citeable*)

Dari beberapa peneliti luar negeri yang saya tanya, baik secara personal maupun dari fasilitas Q&A *ResearchGate* dan *Quora*, jawabannya adalah "ya, data dapat disitasi", karena data sebagaimana halnya makalah lengkap me-

rupakan hasil olah pikir saintifik. Dalam era *open data* dan *open science*, data harus memiliki aksesibilitas dan visibilitas yang tinggi. Brase (2016) menyampaikan bahwa integrasi data juga merupakan bagian penting kolaborasi saintifik. Bagian ini memberikan peluang untuk verifikasi hasil riset sebagai awal dari kolaborasi. Agar data dapat secara formal dibagikan (*shareable*) dan dapat disitasi (*citeable*) perlu ada beberapa komponen (GESIS 2010), yakni: validasi data, metadata, format file, dan digital object identifier (DOI) sebagai salah satu metoda pemberian label (*labeling*) terhadap dokumen di dunia maya, dan staf kurator data. Tahap berikutnya adalah mendisain platform *data publishing* yang handal (*Research Data Alliance*, 2015). Namun tahap ini telah banyak dibantu dengan tersedianya layanan-layanan *data hosting* gratis (dengan kapasitas terbatas) dan berbayar yang akan dijelaskan pada bagian berikut ini.

Bagaimana Agar Data Dapat Dirujuk

Secara konvensional, apapun yang tersedia online dapat disitasi selama ada alamat rujukan yang jelas. Namun perkembangan saat ini, penggunaan kode digital object identifier (DOI) sudah makin meluas, termasuk untuk dataset. Dengan DOI maka obyek yang kita sitasi, apapun itu, tidak akan tertukar dengan obyek yang lain. Kerena DOI adalah kode unik. Jadi saat ini sudah sangat jamak penggunaan (*assigning*) DOI untuk file data yang telah diunggah, sejak pertama kali digunakan tahun 2000. Saat ini diperkirakan telah 120 juta DOI

digunakan di dunia dengan pertumbuhan sekitar 18% pert tahun (DOI org, 2016).

Untuk kasus data geosains, dalam hal ini peta geologi, komponen yang harus ada adalah: koordinat (mengikuti sistem koordinat yang baku), penulis atau pembuat peta, tahun, organisasi pelaksana, dan sumber dananya. Kelima komponen tersebut harus didefinisikan ke dalam metadata peta.

Layanan *Data Hosting*

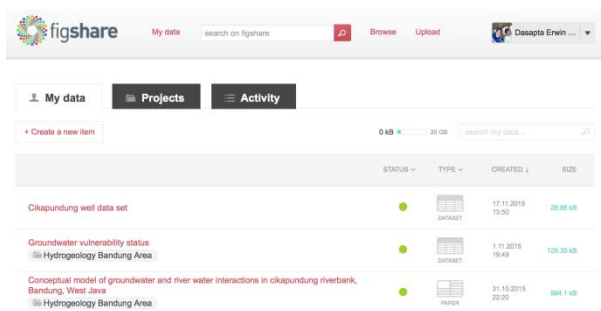
Beberapa contoh perusahaan atau organisasi yang menyelenggarakan *data hosting* bersifat gratis (dengan kapasitas tertentu) dan berbayar, misalnya: *Figshare*, *Zenodo*, dan *Datadryad*. Repositori data dalam bentuk lain, misalnya:

- *Researchgate*: ini sebenarnya adalah media sosial saintifik. Mereka memberikan opsi unggah "*dataset*" selain opsi "*full paper*". Anda dapat secara langsung menawarkan kolaborasi dengan data yang anda unggah.
- *ArXiv*, *BiorXiv*, *PeerJ pre-print repository*: situs ini adalah repositori preprint yaitu makalah yang sedang atau baru akan disubmit ke jurnal atau konferensi. Di dalam dokumen pdf yang kita unggah, kita dapat menyertakan dataset yang berkaitan. Dalam abstrak yang diunggah kita dapat memberikan penjelasan bahwa dalam dokumen juga disertakan kumpulan data mentah.

Seperti halnya *Google Scholar*, *Crossref* atau *Scopus* yang mengindeks karya tulis, ternyata ada pula organisasi yang mengindeks repositori data *online*, yaitu:

- *Datacite*, sebuah organisasi nirlaba yang digagas komunitas *open science* pada tahun 2009.
- *Re3data*, organisasi nirlaba yang mendaftarkan berbagai repositori data di dunia.

Di bawah ini saya sisipkan, tutorial singkat untuk membuat akun di repositori *Figshare*. Prosedurnya sangat baku, yakni diawali dengan registrasi (*Sign up*) dan kita bisa langsung memiliki akun gratis dengan spasi penyimpanan 1 GB. Cukup besar untuk menyimpan *dataset* dalam format *binary xls* atau *ASCII comma separated values (csv)*, serta beberapa video klip pendek bila diperlukan. Setelah mendaftar, maka kita akan masuk ke layar "*home*" seperti ini.



Gambar 1. Layar "home" Figshare

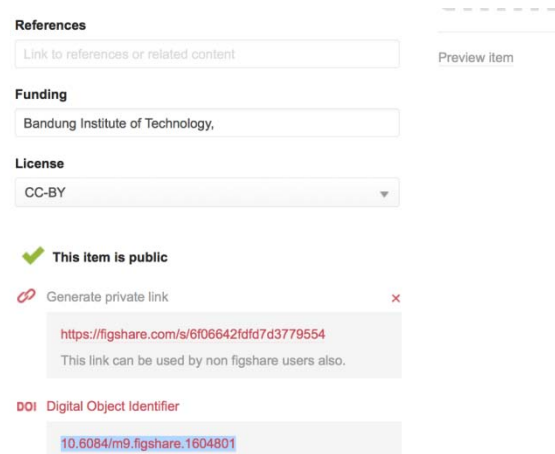
Tautan ini adalah contoh dataset saya yang diunggah di *Figshare*. Bila Anda mengakses <https://figshare.com/s/6f06642fd7d3779554>, akan muncul layar seperti pada Gambar 2.

Memilih Lisensi

Hal paling akhir yang perlu dilakukan adalah memilih lisensi berkas yang kita unggah ke

Figshare atau ke repositori data manapun. Yang paling umum digunakan komunitas *open science* adalah lisensi *Creative Commons (CC)*. Lisensi yang paling umum adalah:

- **CC-BY**: kita hanya boleh merujuk, menggunakan ulang, menganalisis ulang, memodifikasi dokumen kita dengan syarat menyebutkan sumbernya atau,
- **CC-BY-SA**: sama dengan ketentuan di atas, ditambah keharusan dokumen turunan untuk dibagikan dengan lisensi yang sama, *Share Alike (SA)*, dengan dokumen milik kita.



Gambar 2. Layar Dataset Figshare

Kondisi ke Indonesia

Bagaimana dengan kondisi di Indonesia? Bila mengamati situs lembaga pemerintah yang menyediakan data, hingga saat ini yang tersedia adalah situs perpustakaan yang mengindeks berbagai dokumen yang tersimpan, termasuk di dalamnya sebagai contoh adalah peta geologi. Namun demikian daftar tersebut belum terhubung kepada data *softfile* dari peta yang

bersangkutan. Belum pula ditemukan lembaga yang secara format menyediakan fasilitas pemesanan dan pengiriman dokumen peta. Operasi ini selalu ditawarkan via komunikasi langsung (misal melalui komunikasi telepon atau email) dengan staf.

Kondisi lebih maju diberikan oleh berbagai repositori tugas akhir milik perguruan tinggi. Beberapa repositori yang sangat mudah dicari melalui *Google* diantaranya adalah perpustakaan digital (*digital library*) milik ITB, IPB, UGM, yang umumnya menyediakan berkas berjenis pdf. Dokumen tersebut biasanya sebagai syarat kelulusan tugas akhir mahasiswa. Data risetnya dipastikan akan ada tersisip (*embedded*) ke dalam berkas pdf. Dengan kondisi demikian, maka data tidak dapat secara langsung dirujuk dan digunakan. Tahapan konversi dari pdf kembali ke format tabel, misal xls (*Ms. Excel*) atau ods (*open document spreadsheet*). Lampiran berupa peta seringkali tidak disertakan, padahal sangat penting untuk bidang geosains. Berikut ini beberapa *snapshot* layar tipikal situs perpustakaan digital di Indonesia.



Gambar 3. Beberapa *snapshot* layar tipikal situs perpustakaan digital di Indonesia: (a) Perpustakaan digital Badan Geologi (<http://www.bgl.esdm.go.id/>); (b) Perpustakaan digital Universitas Gajah Mada (<http://lib.ugm.ac.id/ind/>); (c) Perpustakaan digital Institut Teknologi Bandung (<http://digilib.itb.ac.id/>)

Penutup

Secara ringkas, penulis merekomendasikan perpustakaan digital milik perguruan tinggi sebagai salah satu prototip dari inisiasi *open data* ini. Kondisi repositori saat ini dinilai telah cukup baik, ditandai dengan server yang handal, mudah dikenali *Google*, dan menyediakan metadata berkas secara sistematis, biasanya terdiri dari: judul, tahun, penulis, pembimbing, universitas, jenis dokumen.

Berdasarkan kondisi tersebut, penulis menawarkan beberapa langkah penyesuaian menuju *open data* yang sederhana:

- metadata dokumen harus dapat diekspor ke dalam format yang dikenal, misalnya: *BibTex*, *txt*, *RIS*, *Mendeley*, atau *EndNote*. Dengan adanya menu “*export citation*” ini, maka pembaca tidak perlu lagi mengetik identitas laporan atau data secara manual ke dalam naskahnya.
- data mentah yang dianalisis dalam tugas akhir harus diunggah terpisah dalam format yang familiar, misalnya: *xls*, *csv*, *txt* dll.



- Perlu aplikasi DOI generator untuk memberikan penanda atau tautan khusus terhadap dokumen-dokumen yang diunggah.

Dari uraian singkat di atas, maka data adalah output riset yang potensial untuk dikembangkan. Syaratnya harus tersedia secara mudah (visibilitas tinggi) dan dapat diunduh dalam format yang paling umum dipakai (bukan pdf). Dengan dibukanya akses data, maka peluang untuk menghasilkan publikasi-publikasi turunan akan terbuka lebar. Kolaborasi dengan peneliti lain (dari dalam dan luar negeri) juga dapat dibangun dengan lebih mudah.

Harapannya, di masa mendatang "laci" bukan lagi repositori yang paling disukai.

Daftar Pustaka

1. *ArXiv*, Pre-print database, url: <http://arxiv.org>, diakses 20 Januari 2016.
2. *Badan Geologi Indonesia*, Perpustakaan digital, url: <http://www.bgl.esdm.go.id/>, diakses 20 Januari 2016.
3. *BiorXiv*, Pre-print database, url: <http://biorxiv.org>, diakses 20 Januari 2016.
4. Brase, J., "Making data citeable", url: <http://book.openingscience.org>, doi: 10.1007/978-3-319-00026-8_26, diakses 20 Januari 2016.
5. *Creative Commons*, Open source license, url: <http://creativecommons.org>, diakses 20 Januari 2016.
6. *Crossref*, Scientific database, url: <http://crossref.org>, diakses 20 Januari 2016.
7. *Datacite*, Data repository database, url: <http://www.datacite.org>, diakses 20 Januari 2016.
8. *Datadryad*, Online research repository, url: <http://datadryad.org>, diakses 20 Januari 2016.
9. *DOI organisation*, url: <http://doi.org>, diakses 20 Januari 2016.
10. *Figshare*, Online research repository, url: <http://figshare.com>, diakses 20 Januari 2016.
11. GESIS, 2010, "European values study (EVS), 4th wave, integrated dataset", GESIS archive, doi: 10.4232/1.10188, url: https://www.researchgate.net/publication/280076717_Making_data_citable, diakses 20 Januari 2016.
12. *Google Scholar*, Scientific database, url: <http://scholar.google.com>, diakses 20 Januari 2016.
13. *Institut Teknologi Bandung*, Perpustakaan digital, url: <http://digilib.itb.ac.id>, diakses 20 Januari 2016.
14. *PeerJ pre-print*, Pre-print database, url: <http://peerj.com/preprints>, diakses 20 Januari 2016.
15. *Re3data*, Data repository database, diakses 20 Januari 2016.
16. *Research Data Alliance*, Data Publishing Workflows, url: <http://rd-alliance.org>, diakses 20 Januari 2016.
17. *Researchgate*, Scientific networking, url: <http://researchgate.net>, diakses 20 Januari 2016.

18. *Universitas Gajah Mada*, Perpustakaan digital, url: <http://lib.ugm.ac.id/ind/>, diakses 20 Januari 2016.
19. *Scopus*, Scientific database, url: <http://scopus.com>, diakses 20 Januari 2016.
20. *Zenodo*, Online research repository, url: <http://zenodo.org>, diakses 20 Januari 2016.

=====

** Dasapta Erwin Irawan adalah dosen di Institut Teknologi Bandung, peneliti di bidang hidrogeologi. Ia juga seorang blogger dan penggiat open science. Karya-karya tulisnya dapat dibaca di blog <http://dasaptaerwin.net/wp> dan repositori http://researchgate.net/profile/dasapta_irawan.*