

5-10-2017

Data Management

Jack O'Gorman

University of Dayton, jogorman1@udayton.edu

Follow this and additional works at: https://ecommons.udayton.edu/roesch_fac_presentations



Part of the [Information Literacy Commons](#), and the [Scholarly Communication Commons](#)

Recommended Citation

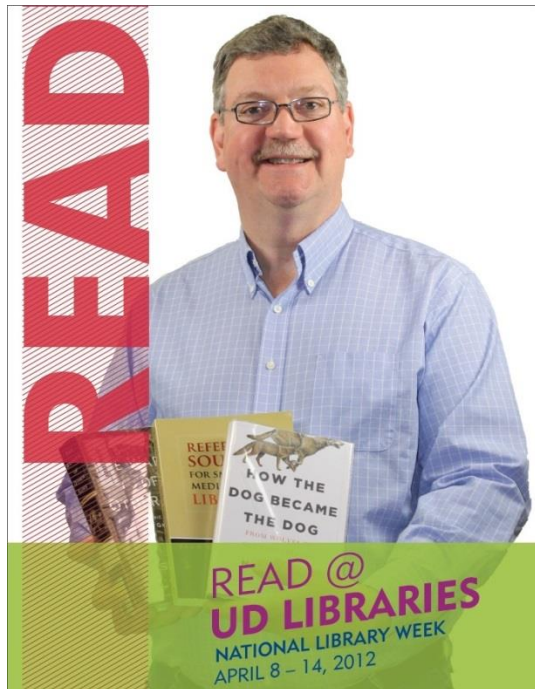
O'Gorman, Jack, "Data Management" (2017). *Roesch Library Faculty Presentations*. 39.
https://ecommons.udayton.edu/roesch_fac_presentations/39

This Presentation is brought to you for free and open access by the Roesch Library at eCommons. It has been accepted for inclusion in Roesch Library Faculty Presentations by an authorized administrator of eCommons. For more information, please contact frice1@udayton.edu, mschlangen1@udayton.edu.

Data Management

SOCHE

May 2017



Jack O'Gorman
Reference and Instruction Librarian/
Associate Professor
Roesch Library
University of Dayton
jogorman1@udayton.edu
937-220-2324

Research Data Defined

“Research data means the recorded factual material commonly accepted in the scientific community as necessary to validate research findings, but not any of the following:

- Preliminary analyses
- Drafts of scientific papers
- Plans for future research
- Peer reviews
- Communication with colleagues”

(Briney p.5 from OMB Circular A-81)

Briney, K. (2015). *Data management for researchers: Organize, maintain and share your data for research success*. Exeter, UK: Pelagic Publishing.

Big Data Defined

“Big Data should be understood as a putatively new mode of knowledge production, based on the global retention of all data, rather than the measure of a specific volume. ”

Poskanzer, D. R. (2015). Big Data. In J. B. Holbrook (Ed.), *Ethics, Science, Technology, and Engineering* (2nd ed., Vol. 1, pp. 210-212). Farmington Hills, MI: Macmillan Reference USA. Retrieved from <http://go.galegroup.com/ps/i.do?id=GALE%7CCX3727600092&v=2.1&u=dayt72472&it=r&p=GVRL&sw=w&asid=78089cdec3fc0d472414b929dd9721fa>

Data Management Defined

“Data Management is the compilation of many small practices that make your data easier to find, easier to understand, less likely to be lost, and more likely to be usable during a project, or ten years later. Data Management is fundamentally about taking care of one of the most important things you create during the research process: your data.”

(Briney p.7)

Briney, K. (2015). *Data management for researchers: Organize, maintain and share your data for research success*. Exeter, UK: Pelagic Publishing.

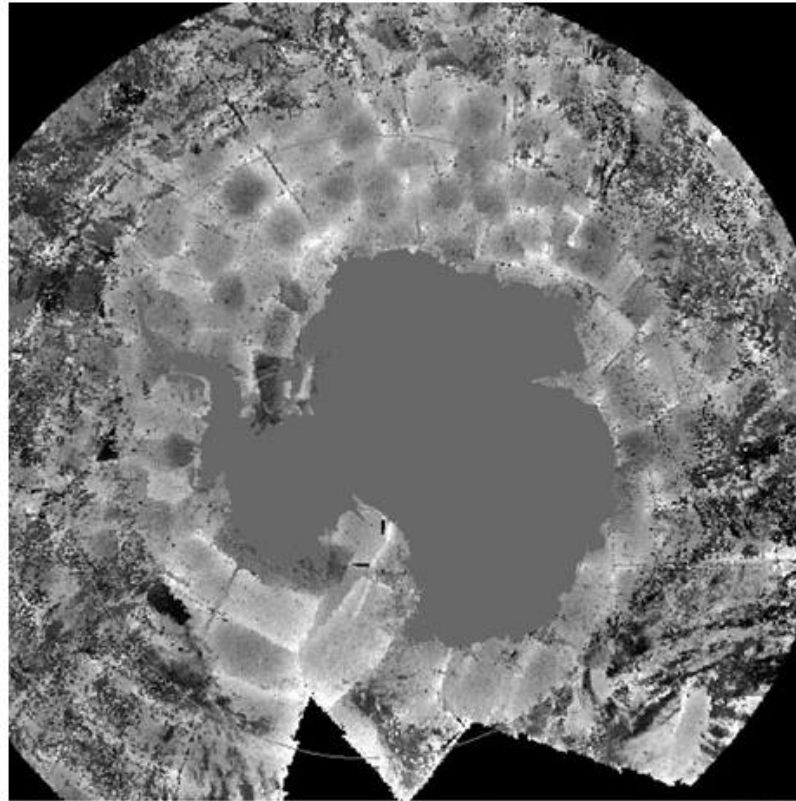
Data Management Plans

“Proposals submitted to NSF must include a supplementary document of no more than two pages labeled "Data Management Plan" (DMP). This supplementary document should describe how the proposal will conform to NSF policy on the dissemination and sharing of research results. Proposals that do not include a DMP will not be able to be submitted.”

Problem : Unreadable Data

Earliest Satellite Images of Antarctica

By Keith Cowing on November 6, 2014 6:45 PM



<http://www.moonviews.com/2014/11/earliest-satellite-images-of-antarctica.html#more>


Problem: Digital Dark Ages

University of Dayton - Por... Library Awards - commun... W Digital dark age - Wikiped... Google Translate

nl.wikipedia.org/wiki/Digital_dark_age

Registreren Niet aangemeld Overlegpagina IP-adres Bijdragen IP-adres Aanmelden

Artikel Overleg Lezen Bewerken Geschiedenis Zoeken


 **WIKIPEDIA**
De vrije encyclopedie


Hoofdpagina
Vind een artikel
Vandaag
Etalage
Categorieën
Recente wijzigingen
Nieuwe artikelen
Willekeurige pagina

Informatie
Gebruikersportaal
Snelcursus
Hulp en contact
Donaties

Hulpmiddelen
Links naar deze pagina
Verwante wijzigingen
Bestand uploaden
Speciale pagina's
Permanente koppeling
Paginagegevens
Wikidata-item
Deze pagina citeren

Afdrukken/exporteren
Boek maken
Downloaden als PDF
Printvriendelijke versie

In andere talen 
العربية
English
Italiano
Occitan
Română
中文

 Koppelingen

Digital dark age

Digital dark age (Engels, vertaald: *digitale donkere tijd*) is een begrip waarmee een mogelijke toekomstige situatie wordt beschreven waarbij het moeilijk of onmogelijk wordt om historische documenten te lezen omdat ze opgeslagen zijn in in onbruik geraakte bestandsformaten of de benodigde hardware en/of software niet meer beschikbaar is. De naam refereert aan de *Dark Ages* (de middeleeuwen) en het daarbij behorende (vermeende) gebrek aan geschreven bronnen.

De huidige tijd waarin veel documenten alleen op **digitale media** met een beperkte houdbaarheid opgeslagen worden (in tegenstelling tot bewezen langdurig houdbare media zoals papier, hout en steen) zou in de toekomst beschouwd kunnen worden als een "zwart gat" in de geschiedenis omdat er (bijna) geen documentatie overgebleven is.


Een oplossing voor het langdurig bewaren van gegevens is de **M-DISC**, een eenmalig beschrijfbaar schijf die volgens de ontwikkelaars duizend jaar mee kan gaan.^[1]

Voorbeelden [bewerken]

Bij latere analyse van de **NASA**-bestanden van de **Vikingmissies** uit 1976 bleek de data onleesbaar omdat deze in een later niet meer bekend bestandsformaat was opgeslagen. Uiteindelijk slaagde men met de nodige moeite om de data te verwerken.

Een ander voorbeeld is het **Domesday Project** van de BBC uit 1986, dat een moderne multimedia-voortzetting was van het **Domesday Book** uit 1086, waarbij meer dan een miljoen personen informatie verzamelden over het dagelijks leven in het Verenigd Koninkrijk in de jaren tachtig. Het Domesday Book is meer dan negenhonderd jaar na dato nog steeds leesbaar, maar vlak na de millenniumwisseling dreigde het Domesday Project ontoegankelijk te worden omdat de voor het project noodzakelijke hardware zeldzaam werd.

IT-wetenschappers wisten het project te converteren naar modernere systemen.



BBC Master-computer met LV-ROM laserdisclezer, waarop het Domesday Project draaide.

Zie ook [bewerken]

- Linkrot

Bronnen, noten en/of referenties

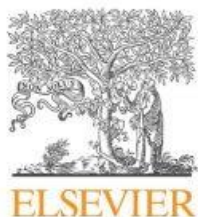
- ^(en) BBC website met de inhoud van het Domesday Project
- ^(en) How to stave off a digital 'dark age', Robert Colvile, Daily Telegraph, 5 juli 2007
- ^(en) Why the Demise of Print Media May Be Bad for Humanity, Tony Bradley, PCWorld, 19 maart 2012

- [↑] ^(en) mDisc Review: A Thousand Years of Storage, Network Computing, 16 augustus 2011

Categorieën: Computer | Archief

https://www.google.com/url?sa=i&rc=j&q=&esrc=s&source=images&cd=&ved=0CAMQjxw&url=http%3A%2F%2Fnl.wikipedia.org%2Fwiki%2FDigital_dark_age&ei=RdK-VOui9etyASPIYlwCA&bvm=by:83829542,d:Aw&psig=AFQj...

http://nl.wikipedia.org/wiki/Digital_dark_age



Contents lists available at ScienceDirect

Environmental Research

journal homepage: www.elsevier.com/locate/envres



Submicrometer particles and their effects on the association between air temperature and mortality in Brisbane, Australia



Lina Wang^a, Shilu Tong^b, Ghasem (Sam) Toloo^b, Weiwei Yu^{b,*}

^a State Environmental Protection Key Laboratory of Environmental Risk Assessment and Control on Chemical Process, East China University of Science and Technology, Shanghai, China

^b School of Public Health and Social Work, Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, QLD, Australia

Problem: Retracted Articles



Retraction notice to “Submicrometer particles and their effects on the association between air temperature and mortality in Brisbane, Australia” [Environ. Res. 128 (2014) 853–860]



Lina Wang^a, Shilu Tong^b, Ghasem (Sam) Toloo^b, Weiwei Yu^b

^a State Environmental Protection Key Laboratory of Environmental Risk Assessment and Control on Chemical Process, East China University of Science and Technology, Shanghai, China

^b School of Public Health and Social Work, Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, QLD, Australia

This article has been retracted: please see Elsevier Policy on Article Withdrawal (<http://www.elsevier.com/locate/withdrawalpolicy>).

This article has been retracted at the request of the Authors following an inquiry from the Queensland University of Technology (QUT). The inquiry found and it was confirmed by the authors that:

The article contained misleading and inaccurate information. Specifically, there was:

- incorrect analysis of the data, partly as a consequence of a failure of communication regarding the nature of the data set and how it was collected.
- incorrect and misleading description of how and when the data set was collected
- failure to correctly attribute the source of the data.
- an incorrect description of the statistical methods and data selection.
- incorrect citation of papers in the description of the data collection.

2. Accuracy of publication:

- The authors of the paper make misleading statements at Sections 2.1 (Data Collection) and 2.2 (Data Analysis). Specifically the authors:
 - imply that one or more of them participated in the collection of the dataset;
 - gave a false impression that the analysis was based on a far greater body than actually existed; and
 - conducted an analysis that was based on incorrect assumptions about the data.
- Conclusions of the paper are flawed.

3. Incorrect citation and author acknowledgement:

- The final sentence in Section 4 (Discussion) of the Paper makes two incorrect citations to Crause et al (2009, 2010)

Retractionwatch.com

Retraction Watch

Tracking retractions as a window into the scientific process

Top 10 most highly cited retracted papers

without comments

Ever curious which retracted papers have been most cited by other scientists? Below, we present the list of the 10 most highly cited retractions. Readers will see some familiar entries, such as the infamous *Lancet* paper by Andrew Wakefield that [originally suggested a link between autism and childhood vaccines](#). You'll note that many papers — including the #1 most cited paper — received more citations after they were retracted, which [research has shown is an ongoing problem](#). As always, we will update the list as more information comes to light.

Article	Year of retraction	Cites before retraction	Cites after retraction	Total cites from journals indexed by Web of Science
1. Visfatin: A protein secreted by visceral fat that mimics the effects of insulin. SCIENCE, JAN 21 2005				
<i>Fukuhara A, Matsuda M, Nishizawa M, Segawa K, Tanaka M, Kishimoto K, Matsuki Y, Murakami M, Ichisaka T, Murakami H, Watanabe E, Takagi T, Akiyoshi M, Ohtsubo T, Kihara S, Yamashita S, Makishima M, Funahashi T, Yamanaka S, Hiramatsu R, Matsuzawa Y, Shimomura I.</i>	2007	247	776	1023
2. Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. LANCET, FEB 28 1998				
<i>Wakefield AJ, Murch SH, Anthony A, Linnell J, Casson DM, Malik M, Berelowitz M, Dhillon AP, Thomson MA, Harvey P, Valentine A, Davies SE, Walker-Smith JA.</i>	2010	675	308	983

Subscribe to Blog via Email

Join 12,056 other subscribers

Pages

[Help us: Here's some of what we're working on](#)

[How you can support Retraction Watch](#)

[Meet the Retraction Watch staff](#)

[About Adam Marcus](#)

[About Ivan Oransky](#)

[The Center For Scientific Integrity](#)

[Board of Directors](#)

[The Retraction Watch FAQ, including comments policy](#)

[The Retraction Watch Transparency Index](#)

[The Retraction Watch Leaderboard](#)

[Top 10 most highly cited retracted papers](#)

[The Retraction Watch Store](#)
[Upcoming Retraction Watch appearances](#)

[What people are saying about Retraction Watch](#)

Problem: Personal Data Shared



“In early August we discovered that sensitive personally identifiable information appeared in some full text documents contained in the ERIC collection. Specifically, social security numbers and other highly sensitive information were found in multiple documents and in a way that could not easily be isolated. For that reason, we had to temporarily disable access to many full text documents.”

“We are seeking to restore access to documents as soon as possible. In order to restore access to ERIC, we have to check every document to see if it contains personally identifiable information.”

“Documents will be returned on a rolling basis and may take several weeks, but we are working as fast as possible.” —email from “ERIC Team” August 2012

Opportunity: Shared Data

Global scientific community commits to sharing data on Zika

10 February 2016



Leading global health bodies including academic journals, NGOs, research funders and institutes, have committed to sharing data and results relevant to the current Zika crisis and future public health emergencies as rapidly and openly as possible.

Organisations including the Bill and Melinda Gates Foundation, Médecins Sans Frontières, the US National Institute of Health and the Wellcome Trust, along with leading academic journals including Nature, Science and the New England Journal of Medicine, have signed a joint declaration and hope that other bodies will come on board in the coming weeks.

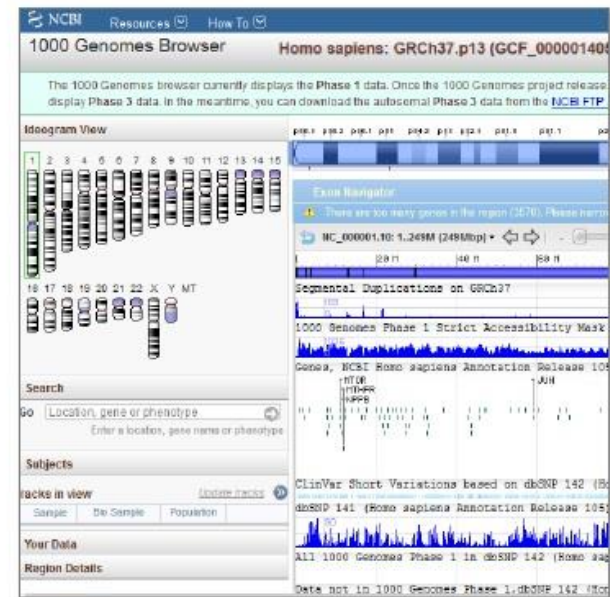
<http://www.wellcome.ac.uk/News/Media-office/Press-releases/2016/WTP060169.htm>

Scholarly examples of “big data”

- First human genome sequence took resources from 20 institutions, 13 years, and \$3 billion to determine the order of 3 Billion nucleotides. Today some labs are sequencing the equivalent of 320 genomes per week.

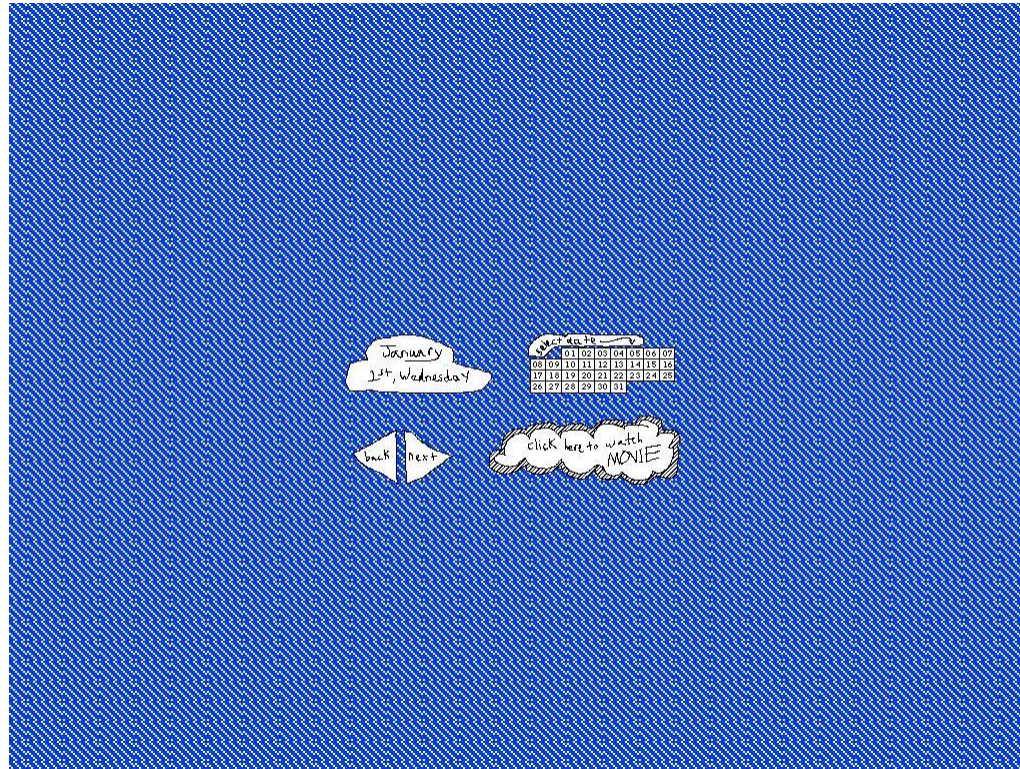
Fukomys damarensis	12/11/2014	10:45:00 PM
Fulmenus glacialis	1/9/2013	2:04:00 PM
Fungi	6/12/2014	12:00:00 AM
GENOME_REPORTS	5/15/2014	12:00:00 AM
Galeoscoptes variegatus	1/8/2015	8:56:00 AM
Gallus gallus	6/28/2013	12:00:00 AM
Gavia stellata	1/8/2015	11:12:00 AM
Geospiza fortis	1/23/2014	12:00:00 AM
Glycine max	11/14/2014	3:31:00 PM
Gorilla gorilla	10/29/2013	12:00:00 AM
HUMAN MICROBIOM	4/19/2012	12:00:00 AM
H. sapiens	7/28/2014	12:00:00 AM
Halaeetus albicilla	1/9/2015	2:04:00 PM
Halaeetus leucoccephalus	12/9/2014	3:12:00 AM
Haplochromis burtoni	10/30/2014	5:18:00 PM
Harpegnathos saltator	1/23/2015	9:38:00 PM
Heterocephalus glaber	10/29/2013	12:00:00 AM
Homo sapiens	12/14/2012	12:00:00 AM
Hydra magnapapillata	3/8/2013	12:00:00 AM
Hydra vulgaris	11/27/2013	12:00:00 AM
IDS	2/6/2015	3:49:00 PM
INFLUENZA	2/9/2015	11:29:00 AM
Ictidomys tridecemlineatus	1/23/2014	12:00:00 AM
Jaculus jaculus	10/29/2013	12:00:00 AM
Leimicichthys crocea	12/23/2014	9:52:00 PM

AAGATTCCGAATGCA
GCGTATAGCTAGCTA
TCTACGATAGTACTG
TGCGACGTACCTAAA



New Form of Scholarly Communication

Data Descriptor





Pantheon 1.0, a manually verified dataset of globally famous biographies

Amy Zhao Yu, Shahar Ronen, Kevin Hu, Tiffany Lu & César A. Hidalgo

[Affiliations](#) | [Contributions](#) | [Corresponding authors](#)

Scientific Data **3**, Article number: 150075 (2016) | doi:10.1038/sdata.2015.75

Received 12 February 2015 | Accepted 13 November 2015 | Published online 05 January 2016

About *Scientific Data*

Scientific Data is an open-access, peer-reviewed journal for descriptions of scientifically valuable datasets. Our primary article-type, the **Data Descriptor**, is designed to make your data more discoverable, interpretable and reusable.

E-alert

RSS

Facebook

Twitter



Submit manuscript ▶

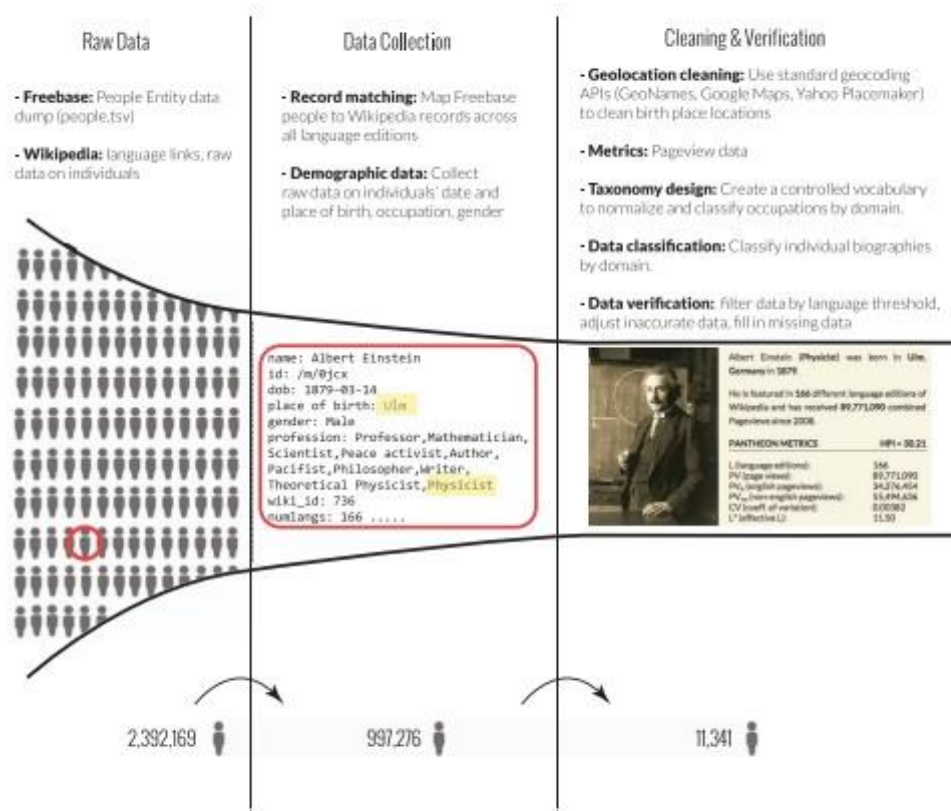


Figure 1. Pantheon Data Workflow. Flow diagram detailing the data collection process for the Pantheon 1.0 ($n = 11,341$). Inset image from pantheon.media.mit.edu.

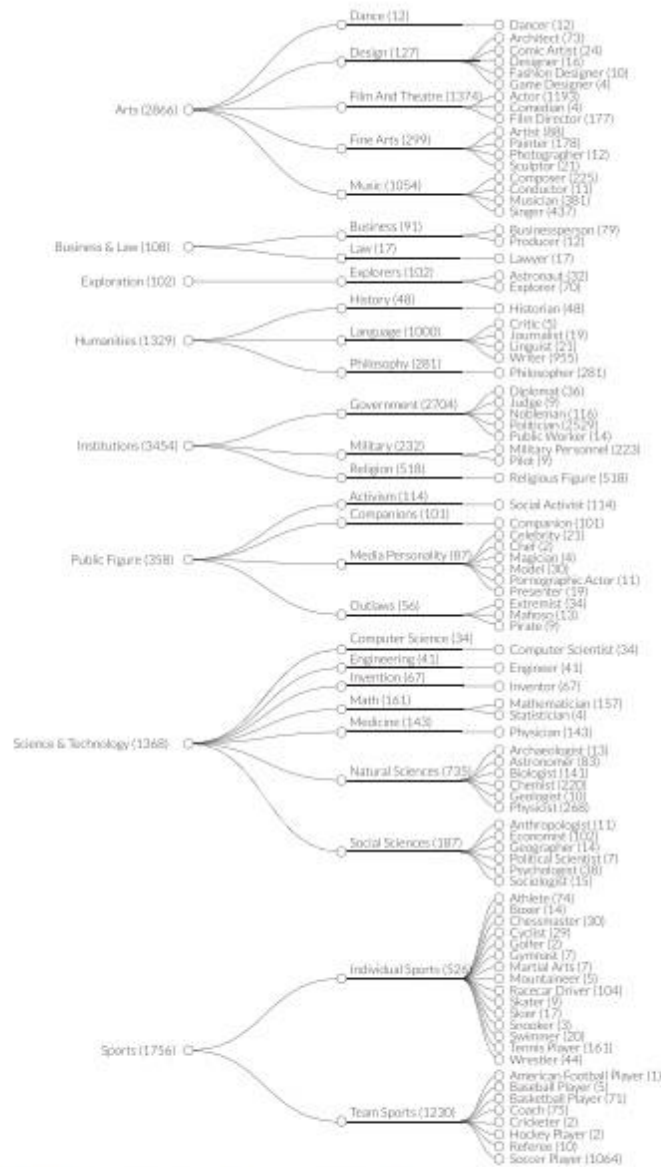


Figure 3. Domain Taxonomy. From left to right: domain (i.e., Sports), industry (i.e., Team Sports) and occupation (i.e., Soccer Player).

Famous People:

- Pirate (9)
- Mathematician (157)
- Soccer Player (1064)

Data Records

The Pantheon dataset is publicly available on the Harvard Dataverse Network and can be accessed directly at: <https://dataverse.harvard.edu/dataverse/pantheon>. The dataset is visualized at <http://pantheon.media.mit.edu>, a data visualization engine that allows users to dynamically explore the dataset through interactive visualizations.

The data consists of three files—`pantheon.tsv`, `wikilangs.tsv`, and `pageviews_2008–2013.tsv` (Data Citation 1).

The first file, `pantheon.tsv`, is a flattened tab-limited table, where each row of the table represents a unique biography. Each row contains the following variable fields:

- **name**—name of the historical character (in English)
- **en_curid**—unique identifier for each individual biography, maps to the pageid from Wikipedia. To map to an individual's biography in Wikipedia, use the `en_curid` field as an input parameter to the following URL: [http://en.wikipedia.org/?curid=\[en_curid\]](http://en.wikipedia.org/?curid=[en_curid]). We use the English curid as the unique identifier in the Pantheon dataset; we confirmed that all biographies with $L > 25$ as of May 2013 had an entry in the English Wikipedia.
- **countryCode**—ISO 3166-1 alpha2 (based on present-day political boundaries)
- **countryCode3**—ISO 3166-1 alpha3 country code (based on present-day political boundaries)
- **countryName**—commonly accepted name of country
- **continentName**—name of continent
- **birthyear**—birthyear of individual
- **birthcity**—given birthcity of individual
- **occupation**—occupation of the individual
- **industry**—category based on an aggregation of related occupations
- **domain**—category based on an aggregation of related industries
- **gender**—male or female
- **TotalPageViews**—total pageviews across all Wikipedia language editions (January 2008 through December 2013)
- **L_star**—adjusted L (see Appendix for calculation)
- **numlangs**—number of Wikipedia language editions that each biography has a presence in (as of May 2013)
- **StdPageViews**—s.d. of pageviews across time (January 2008 through December 2013)
- **PageViewsEnglish**—total pageviews in the English Wikipedia (January 2008 through December 2013)
- **PageViewsNonEnglish**—total pageviews in all Wikipedias except English (January 2008 through December 2013)
- **AverageViews**—Average pageviews per language (January 2008 through December 2013)
- **HPI**—Historical Popularity Index (see equation (4))

References

1. Ronen, S. *et al.* Links that speak: The global language network and its association with global fame. *Proceedings of the National Academy of Sciences* **111**, E5616–E5622 (2014).
2. Wagner, C., Garcia, D., Jadidi, M. & Strohmaier, M. It's a Man's Wikipedia? Assessing Gender Inequality in an Online Encyclopedia. *arXiv* arXiv:1501.06307 [cs.CY] (2015).
3. McLuhan, M. *Understanding Media: The Extensions of Man* (MIT Press, 1964).
4. Eisenstein, E. L. *The printing press as an agent of change* (Cambridge University Press, 1979).
5. Murray, C. *Human Accomplishment* (Harper Collins, 2003).
6. Michel, J.-B. *et al.* Quantitative Analysis of Culture Using Millions of Digitized Books. *Science* **331**, 176–182 (2011).
7. Popescu, A. & Grefenstette, G. Spatiotemporal mapping of Wikipedia concepts. *Proceedings of the 10th annual joint conference on Digital libraries* 129–138 (2010).
8. Skiena, S. & Ward, C. *Who's Bigger? Where Historical Figures Really Rank* (Cambridge University Press, 2013).
9. Schich, M. *et al.* A network framework of cultural history. *Science* **345**, 558–562 (2014).
10. Giles, J. Internet encyclopedias go head to head. *Nature* **438**, 900–901 (2005).
11. Spinellis, D. & Louridas, P. The collaborative organization of knowledge. *Communications of the ACM* **51**, 68–73 (2008).
12. Hedden, H. Taxonomies and controlled vocabularies best practices for metadata. *Journal of Digital Asset Management* **6**, 279–284 (2010).
13. Abel, G. & Sander, N. Quantifying Global International Migration Flows. *Science* **343**, 1520–1522 (2014).
14. Ford, H. in *Critical Point of View: A Wikipedia Reader*, 258–268 (Institute of Network Cultures, 2011).
15. Brown, A. R. Wikipedia as a Data Source for Political Scientists: Accuracy and Completeness of Coverage. *PS: Political Science and Politics* **44**, 339–343 (2011).
16. Royal, C. & Kapila, D. What's on Wikipedia, and What's Not? *Social Science Computer Review* **27**, 138–148 (2009).
17. UNESCO. *The 2009 UNESCO Framework for Cultural Statistics* (UNESCO Institute for Statistics, 2009).

Data Citations

1. Yu, A. Z., Ronen, S., Hu, K., Lu, T. & Hidalgo, C. *Harvard Dataverse* <http://dx.doi.org/10.7910/DVN/28201> (2014).

Data Set Citations

ELSEVIER

Share     

[View in a browser](#)

Scopus[®]

Scopus Newsletter: April 2017

Linking to research datasets

Research dataset links can now be found on the Scopus Document details page. Through collaboration with the Data Literature Interlinking (DLI) service and Scholix, Scopus is able to link articles with datasets when available on the external data repository. Now, when research datasets are available on the external data repository for an article, the Scopus Document details page will include a “Related Research Data” sidebar, located to the right of the article details.

Another Example



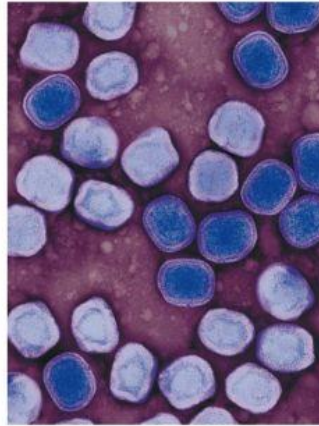
A Pox on History

The smallpox came in the mail, in a little piece of 350-year-old human skin. And the surprise discovery is dramatically rewriting the history of one of the most notorious diseases to plague humankind.

Ana Duggan, an evolutionary geneticist at McMaster University, was collaborating with scientists in Lithuania who were studying desiccated corpses found in the crypt of a church in Vilnius. Out of the more than 500 corpses interred there, most had decomposed, but about 200 were preserved by the cool, dry conditions in the crypt. The sample that startled Duggan came from the leg of a toddler entombed between 1643 and 1665.

Duggan and colleagues were analyzing DNA in the skin sample to learn something about the child's life and death. While probing for evidence of the JC polyomavirus, a common bug they tested the sample for virus DNA—and instead found the variola virus, which causes smallpox.

Known for its signature blisters and gruesome deaths, smallpox killed 300 million people in the 20th century and has long been a scourge of humanity. But how long? Scientists have argued that it has been with us for at least 10,000 years. Researchers have found pockmarks on the face of a 3,000-year-old Egyptian mummy. And historical texts describe epi-



demics of a smallpox-like disease in fourth-century China and second-century Rome.

Still, the 17th-century variola virus DNA detected by Duggan and colleagues is the oldest definitive trace of smallpox that researchers have found. Which is why Duggan and her adviser, Hendrik Poinar, took the next step: After piecing together the genome of the virus in their sample, they compared it with the published genomes of 42 other variola strains collected in the 20th century before 1980, when smallpox was eradicated. As a virus replicates and copies its DNA, errors sneak into

The last natural case of smallpox was recorded in Somalia in 1977.

the genome at a fairly regular rate; the newer the virus strain, the more mutations it will harbor. Looking at the DNA mutations in all those variola virus strains, and assuming a steady mutation rate, the researchers worked backward to create a variola family tree and calculate the age of the strain that gave rise to all the others, including the one in 17th-century Vilnius.

Duggan and Poinar's analysis, published in *Current Biology*, concludes that variola as we know it likely arose in the late 1500s or

A child from 17th-century Europe might have just rewritten the book on a deadly scourge

early 1600s—thousands of years later than researchers currently believe. "We have to go back and rethink it all," says Ann Carmichael, a historian at Indiana University, Bloomington, who studies smallpox epidemics.


If variola virus didn't cause deadly outbreaks until about 500 years ago, what was behind the earlier plagues attributed to smallpox? "That's the million-dollar question," Poinar says. One possibility, researchers say, is another virus with similar symptoms, like chickenpox or measles.

Another puzzle: If smallpox virus wasn't around until the late 1500s or so, how did epidemics of smallpox or a similar disease strike indigenous people in the Americas before then? Researchers think those outbreaks might have been triggered by a less virulent ancestor of variola that Europeans had become immune to before they carried it to the New World, where people were susceptible to it. Meanwhile, in Europe, the virus mutated into something more lethal, causing terrible outbreaks, one of which took the life of that Lithuanian child. —LIZZIE WADE

SCIENCE

were analyzing DNA in the skin sample to learn something about the child's life and death. While probing for evidence of the JC polyomavirus, a common bug, they tested the sample for virus DNA—and instead found the variola virus, which causes smallpox.

Known for its signature blisters and gruesome deaths, smallpox killed 300 million people in the 20th century and has long been a scourge of humanity. But how long? Scientists have argued that it has been with us for at least 10,000 years. Researchers have found pockmarks on the face of a 3,000-year-old Egyptian mummy. And historical texts describe epi-



demics of a smallpox-like disease in fourth-century China and second-century Rome.

Still, the 17th-century variola virus DNA detected by Duggan and colleagues is the oldest definitive trace of smallpox that researchers have found. Which is why Duggan and her adviser, Hendrik Poinar, took the next step: After piecing together the genome of the virus in their sample, they compared it with the published genomes of 42 other variola strains collected in the 20th century before 1980, when smallpox was eradicated. As a virus replicates and copies its DNA, errors sneak into

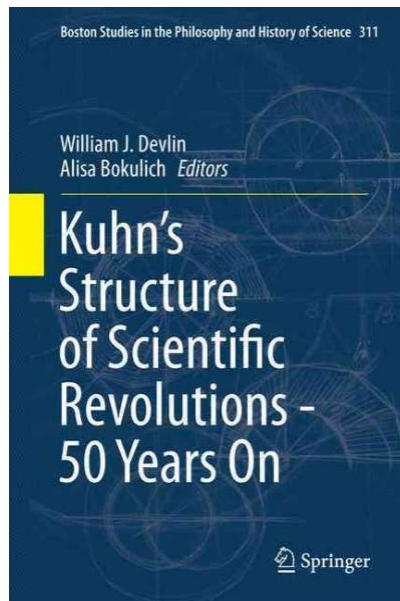
The last natural case of smallpox was recorded in Somalia in 1977.

the genome at a fairly regular rate; the newer the virus strain, the more mutations it will harbor. Looking at the DNA mutations in all those variola virus strains, and assuming a steady mutation rate, the researchers worked backward to create a variola family tree and calculate the age of the strain that gave rise to all the others, including the one in 17th-century Vilnius.

Duggan and Poinar's analysis, published in *Current Biology*, concludes that variola as we know it likely arose in the late 1500s or

Paradigm Shift

- Scholars as users and producers of scholarly resources - including data
- Scholars need to know of the existence and availability of data repositories

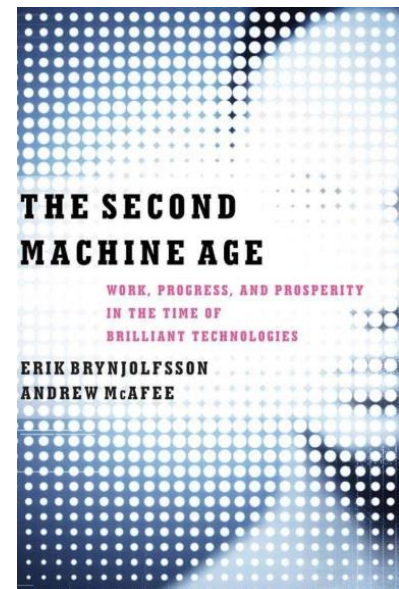


Discoverability of Data

- How does one find data sources?
 - Discipline specific
 - Repositories are generally tied to functional domains
 - Usual literature sources – pay attention to data citations.
 - Data management resource guides

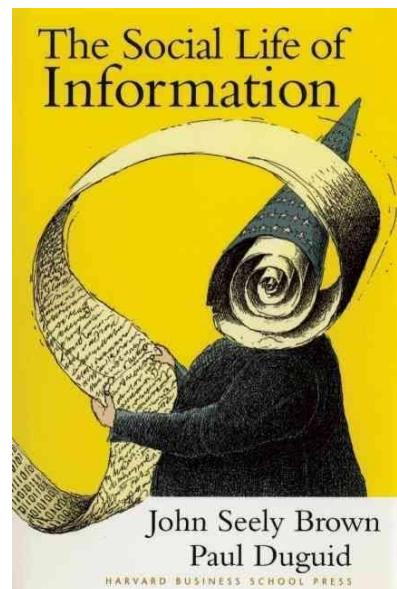
Keeping Up

- But there is the challenge of keeping up...
- Changes described above are coming faster than the pace of producing new scholars.
- Scholars must stay agile and aware of developments.
- Second half of the chessboard.



Networks of Scholars

- Increasing importance of social networks to find these partners and sources
- Knowledge is found in a distributed network of people



Data Clearinghouses: re3data.org

The screenshot shows the re3data.org search results page for the query 'human dna'. The page has a dark navigation bar with links: Home, Search, Browse, Suggest, FAQ, About, Schema, API, Contact, and Legal notice / Impressum. On the left, there is a sidebar with expandable categories: Subjects, Content Types, Countries, AID systems, API, Certificates, Data access, Data access restrictions, Database access, Database licenses, Data licenses, Data upload, Data upload restrictions, Enhanced publication, Institution responsibility type, Institution type, Keywords, Metadata standards, PID systems, Provider types, Quality management, Repository languages, Software, Syndications, Repository types, and Versioning. The main content area shows the search results for 'human dna'. It includes a search bar with the query, a 'Search' button, and a 'Toggle short help' link. Below the search bar, there are pagination controls showing 'Previous', '1' (selected), '2', '3', and 'Next'. A 'Sort by' dropdown is also present. The results section indicates 'Found 56 result(s)'. The first result is 'OAGR Online Ancient Genome Repository'. It includes a subject filter (Humanities and Social Sciences, Ancient Cultures, Prehistory, Biology), a content type filter (Plain text, Raw data, Audiovisual data, Archived data, other), and a country filter (Australia). The description of OAGR states: 'It captures and catalogues ancient human genome and microbiome data, including raw sequence and processed data, along with metadata about its provenance and production. Included datasets are generated from ancient samples studied at the Australian Centre for Ancient DNA, University of Adelaide in collaboration with other research groups. Datasets and collections in OAGR are open data resources made freely available in a reusable form, using open file formats and licensed with minimal restrictions for reuse. Digital object identifiers (DOIs) are minted for included datasets and collections to facilitate persistent identification and citation.' The second result is 'NCBI Reference Sequences RefSeq'. It includes a subject filter (Human Genetics, Microbiology, Virology and Immunology, Medicine, Life Sciences, Biology).

service.re3data.org/search?query=human+dna

Home Search Browse Suggest FAQ About Schema API Contact Legal notice / Impressum

Subjects Content Types Countries AID systems API Certificates Data access Data access restrictions Database access Database licenses Data licenses Data upload Data upload restrictions Enhanced publication Institution responsibility type Institution type Keywords Metadata standards PID systems Provider types Quality management Repository languages Software Syndications Repository types Versioning

human dna Search

Toggle short help

← Previous 1 2 3 Next →

Sort by ▼

Found 56 result(s)

OAGR
Online Ancient Genome Repository

Subject(s) Humanities and Social Sciences Ancient Cultures Prehistory Biology
Humanities Life Sciences

Content type(s) Plain text Raw data Audiovisual data Archived data other

Country Australia

It captures and catalogues ancient human genome and microbiome data, including raw sequence and processed data, along with metadata about its provenance and production. Included datasets are generated from ancient samples studied at the Australian Centre for Ancient DNA, University of Adelaide in collaboration with other research groups. Datasets and collections in OAGR are open data resources made freely available in a reusable form, using open file formats and licensed with minimal restrictions for reuse. Digital object identifiers (DOIs) are minted for included datasets and collections to facilitate persistent identification and citation.

NCBI Reference Sequences
RefSeq

Subject(s) Human Genetics Microbiology, Virology and Immunology Medicine Medicine
Life Sciences Biology

Citing Data Resources

References

1. Ronen, S. *et al.* Links that speak: The global language network and its association with global fame. *Proceedings of the National Academy of Sciences* **111**, E5616–E5622 (2014).
2. Wagner, C., Garcia, D., Jadidi, M. & Strohmaier, M. It's a Man's Wikipedia? Assessing Gender Inequality in an Online Encyclopedia. *arXiv arXiv:1501.06307 [cs.CY]* (2015).
3. McLuhan, M. *Understanding Media: The Extensions of Man* (MIT Press, 1964).
4. Eisenstein, E. L. *The printing press as an agent of change* (Cambridge University Press, 1979).
5. Murray, C. *Human Accomplishment* (Harper Collins, 2003).
6. Michel, J.-B. *et al.* Quantitative Analysis of Culture Using Millions of Digitized Books. *Science* **331**, 176–182 (2011).
7. Popescu, A. & Grefenstette, G. Spatiotemporal mapping of Wikipedia concepts. *Proceedings of the 10th annual joint conference on Digital libraries* 129–138 (2010).
8. Skiena, S. & Ward, C. *Who's Bigger? Where Historical Figures Really Rank* (Cambridge University Press, 2013).
9. Schich, M. *et al.* A network framework of cultural history. *Science* **345**, 558–562 (2014).
10. Giles, J. Internet encyclopedias go head to head. *Nature* **438**, 900–901 (2005).
11. Spinellis, D. & Louridas, P. The collaborative organization of knowledge. *Communications of the ACM* **51**, 68–73 (2008).
12. Hedden, H. Taxonomies and controlled vocabularies best practices for metadata. *Journal of Digital Asset Management* **6**, 279–284 (2010).
13. Abel, G. & Sander, N. Quantifying Global International Migration Flows. *Science* **343**, 1520–1522 (2014).
14. Ford, H. in *Critical Point of View: A Wikipedia Reader*, 258–268 (Institute of Network Cultures, 2011).
15. Brown, A. R. Wikipedia as a Data Source for Political Scientists: Accuracy and Completeness of Coverage. *PS: Political Science and Politics* **44**, 339–343 (2011).
16. Royal, C. & Kapila, D. What's on Wikipedia, and What's Not? *Social Science Computer Review* **27**, 138–148 (2009).
17. UNESCO. *The 2009 UNESCO Framework for Cultural Statistics* (UNESCO Institute for Statistics, 2009).

Data Citations

1. Yu, A. Z., Ronen, S., Hu, K., Lu, T. & Hidalgo, C. *Harvard Dataverse* <http://dx.doi.org/10.7910/DVN/28201> (2014).

DOI

- dx.doi.org/10.7910/DVN/28201
- Digital Object Identifier
- Assumes that items are “born digital”



[Airiti, Inc.](#)



[Crossref](#)



[China National Knowledge
Infrastructure \(CNKI\)](#)



[ISTIC \(The Institute of Scientific and Technical
Information of China\)](#)



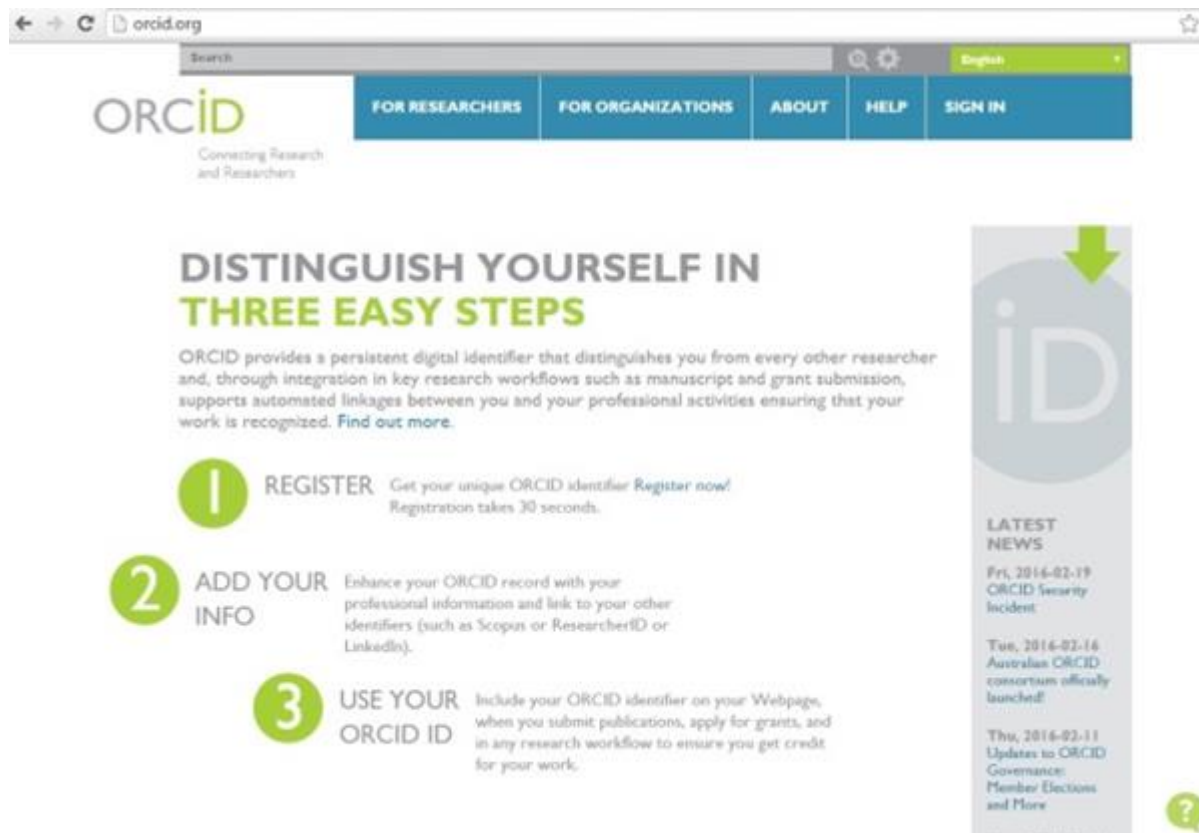
[JaLC \(Japan Link Center\)](#)



[OP \(Publications Office of the European Union\)](#)

Finding Researchers

- Role of disambiguation in researcher networks (ORCID)



The screenshot shows the ORCID.org homepage. At the top, there's a navigation bar with the ORCID logo and the tagline "Connecting Research and Researchers". To the right of the logo are five buttons: "FOR RESEARCHERS", "FOR ORGANIZATIONS", "ABOUT", "HELP", and "SIGN IN". Below the navigation bar, the main heading reads "DISTINGUISH YOURSELF IN THREE EASY STEPS". Under this heading, a paragraph explains that ORCID provides a persistent digital identifier. Below the paragraph are three numbered steps: 1. REGISTER, 2. ADD YOUR INFO, and 3. USE YOUR ORCID ID. Each step has a brief description. On the right side of the page, there's a vertical sidebar titled "LATEST NEWS" with three news items dated from 2016-02-19 to 2016-02-11. A green arrow points down towards the "id" logo in the sidebar.

ORCID
Connecting Research and Researchers

FOR RESEARCHERS FOR ORGANIZATIONS ABOUT HELP SIGN IN

DISTINGUISH YOURSELF IN THREE EASY STEPS

ORCID provides a persistent digital identifier that distinguishes you from every other researcher and, through integration in key research workflows such as manuscript and grant submission, supports automated linkages between you and your professional activities ensuring that your work is recognized. [Find out more.](#)

- 1 REGISTER** Get your unique ORCID identifier *Register now!*
Registration takes 30 seconds.
- 2 ADD YOUR INFO** Enhance your ORCID record with your professional information and link to your other identifiers (such as Scopus or ResearcherID or LinkedIn).
- 3 USE YOUR ORCID ID** Include your ORCID identifier on your Webpage, when you submit publications, apply for grants, and in any research workflow to ensure you get credit for your work.

LATEST NEWS

- Fri, 2016-02-19
ORCID Security Incident
- Tue, 2016-02-16
Australian ORCID consortium officially launched!
- Thu, 2016-02-11
Updates to ORCID Governance: Member Elections and More

ORCID

- ORCID is stands for Open Researcher and Contributor ID <https://orcid.org/content/o-orcid>
- When UD graduate students complete their thesis, the library encourages them to get an ORCID as part of Electronic Theses & Dissertation registration
- You should consider one too



Connecting Research
and Researchers

FOR RESEARCHERS

FOR ORGANIZATIONS

ABOUT

HELP

SIGN OUT

MY ORCID RECORD

INBOX

ACCOUNT SETTINGS

DEVELOPER TOOLS

LEARN MORE

1,974,491 ORCID IDs and counting. [See more...](#)

Jack
O'Gorman

Biography



ORCID ID

orcid.org/0000-0002-4256-0677

[View public version](#)

[Get a QR Code for your iD](#) ⓘ

Also known as

Country

Keywords

Websites

Emails

jogorman1@dayton.edu

▼ Education (2)

+ Add education

⌵ Sort

Saint John's University : Jamaica, NY, United States

1981-09 to 1983-01-31 (Library and Information Sciences)

MLS

Sources: Jack O'Gorman

Created: 2016-02-24



Walsh University: North Canton, OH, United States

1977-09 to 1981-05-15 (Mathematics)

BA

Sources: Jack O'Gorman

Created: 2016-02-24



▼ Employment (1)

+ Add employment

⌵ Sort

University of Dayton: Dayton, OH, United States

1999-01 to present (Roesch Library)

Reference and Instruction Librarian/ Associate Professor

Sources: Jack O'Gorman

Created: 2016-02-24





Build your Data Management Plan

"Many funding agencies require data management plans (DMP) as part of grant proposals. The DMPTool helps institutions & researchers to create high-quality DMPs that meet funder requirements, including those from:"



datacure@googlegroups.com

Data Management for Archival Users

Joanna Thielen <jthielen@oakland.edu>: Apr 24 08:44AM -0700

Hello DataCure!

I'm collaborating with my library's Coordinator of Archives and Special Collections to help researchers who use archival collections implement data management practices ...[more](#)

"Shorish, Yasmeen L - shorisyl" <shorisyl@jmu.edu>: Apr 24 04:00PM

Hi Joanna,

I think Ixchel Faniel's 2015 post on this might have some good leads: <http://hangingtogether.org/?p=5375> and <http://hangingtogether.org/?p=5413> It's repository-heavy from Ixchel's ...[more](#)

lynn yamey <yamey@gmail.com>: Apr 24 10:14AM -0600

Hi Joanna and datacure,

Can you help me understand how 'archival data management' and 'regular' (?) data management are different? I would have thought they are the same before I read your ...[more](#)

Joanna Thielen <jthielen@oakland.edu>: Apr 24 01:45PM -0400

Hi Lynn,

To clarify, we're looking to provide research data management support to researchers who uses archival collections (such as historians). Before embarking on our own study, we want to know ...[more](#)

"Creamer, Andrew" <andrew_creamer@brown.edu>: Apr 24 02:07PM -0400

Hi Joanna,

Sam Simas, a grad student at URI who is interning in our Library's CDS, and I made this attached presentation for 2nd year history students based off our interviews with students ...[more](#)

[Back to top](#)

Institutional data catalog/inventory

Megan O'Donnell <mno@iastate.edu>: Apr 24 09:38AM -0700

Thanks for sharing Ho Jung.

I saw TritonSHARE at RDAP last week and got excited because it's a "proof-of-concept" that I was hoping was possible: the harvesting of metadata for research data ...[more](#)

Library's Role



- Continuation of our traditional role.
 - Acquire information
 - Store information
 - Organize information
 - Make information available

With this new information format, our role has expanded

The Next Big Thing

“Research Data Management is the next big development. As libraries have taken the lead in storing and preserving traditional research outputs via archiving in institutional repositories, so they will also play a part in the development of research data repositories. “

Natalia Madjarevic LSE Research Online

<http://blogs.lse.ac.uk/impactofsocialsciences/2012/10/10/madjarevic-open-access-libraries-respond/>

eCOMMONS

A digital archive of the research, creative works and history of the University of Dayton

[UD HOME](#)

[UD LIBRARIES](#)

[eCOMMONS HOME](#)

[ABOUT](#)

[FAQ](#)

[MY ACCOUNT](#)

Enter search terms:

in this repository ▼

[Advanced Search](#)

[Notify me via email or RSS](#)

CONTRIBUTE YOUR WORK

[Submit Research](#)

[Create a SelectedWorks Researcher Profile](#)

[Copyright for Scholars and Creators](#)

[More about eCommons](#)

[More about SelectedWorks](#)

[Author FAQ](#)

BROWSE

[Collections](#)

[Disciplines](#)

[Authors](#)

CONTACT

[Send email](#)



BROWSE RESEARCH AND SCHOLARSHIP

- [Research unit, center, or department](#)
- [SelectedWorks Gallery](#)
- [Journals and peer-reviewed series](#)
- [Featured Publications](#)
- [Conferences and Events](#)
- [Newsletters and Magazines](#)



ABOUT

[FOLLOW](#)

eCommons, a service of the University Libraries, is a space for content produced by our diverse campus community. It is a permanent multimedia archive of UD's continuous pursuit of transformative education and world-changing research.

AT A GLANCE

Top 10 Downloads


All time

Recent Additions

20 most recent additions

[Activity by year](#)

Data Management

 UNIVERSITY of DAYTON

RESEARCH GUIDES

[UD Libraries](#) / [LibGuides](#) / [Data Management Services](#) / [Overview](#)

Data Management Services: Overview

[Overview](#) | [Writing a Data Management Plan](#) | [Metadata](#) | [Data Repositories](#) | [Best Practices](#) | [Library as Data Partner](#)

What's in the Guide




This guide provides information on creating a data management plan and the library services available to support data management. Use the tabs to navigate through the pages of this guide.

- [Writing a Data Management Plan](#)
- [Metadata](#)
- [Data Repositories](#)
- [Best Practices](#)
- [Library as Data Partner](#)

If you wish to set up a consultation to discuss data management, please contact your [liaison librarian](#).

We acknowledge both Boston College and Massachusetts Institution for Technology for content within this LibGuide.

Contact Us



We've got answers.

Email: knowhub@udayton.edu

Phone: 937-229-4270

Text us: 937-412-0023

Twitter: [@roeschlibrary](#)

Data Management Plan 22 questions

- What type of data will be produced?
- Will it be reproducible?
- What would happen if it became unusable?
- How much data will there be?
- How often will it change?
- How rapidly will it grow?
- Who will use it now?
- Who will use it later?
- Who controls it (the institution, the PI, students?)
- How long should it be retained?
- Is there software required to use the data?
- Privacy requirements?
- Security requirements?
- Sharing requirements? Funding agency sharing policy?
- Other funding agency requirements for data management plans?
- Project and data documentation?
- File naming conventions?
- What file formats, and are those formats long lived?
- Will project and data identifiers be assigned?
- When and where will the data be published?
- Who is an appropriate community of scholars to share the data with?
- Who in the group will be responsible for data management?

What format should you use

Data Type	Preferred	Acceptable	Not Recommended
<i>Text and textual data</i>	.txt, .csv, .xml (with appropriate DTD)	.rtf	.docx, .doc
<i>Images</i>	.jp2, .tif, .png	.jpg, .gif, .dng	.psd, .jpx or .jpf,
<i>Documents</i>	.pdf/a, .epub		
<i>Audio</i>	.wav, .aif	.flac, .ogg,	.aifc, .ram, .wma
<i>Moving Image & Video</i>	.mj2, .mov		

<http://libguides.udayton.edu/c.php?g=15360&p=83498>



<http://crosstech.crossref.org/2014/02/many-metrics-such-data-wow.html>

Further Reading

- Briney, K. (2015). *Data management for researchers: Organize, maintain and share your data for research success*. Exeter, UK: Pelagic Publishing.
- Carlson, J. (2011). Demystifying the data interview Developing a foundation for reference librarians to talk with researchers about their data. *Reference Services Review*, 40(1), 7-23.
- Krier, L., & Strasser, C. A. (2014). *Data management for libraries: A LITA guide*. Chicago : ALA TechSource, 2014
- MacMillan, D. (2014). Data Sharing and Discovery: What Librarians Need to Know. *Journal Of Academic Librarianship*, 40(5), 541-549. doi:10.1016/j.acalib.2014.06.011