



# 研究寄存之 最強可再用 – 美國樹妖

Andrea Wei-Ching Huang  
Institute of Information Science, Academia Sinica, Taiwan.  
2020/06/11

*the Story of RDR- part 1:Dryad*

# 最強可再用 (reusable):

CC0 +  
data package

資料集、人、  
組織皆有國際  
標準PID註冊

後設資料兼顧  
常人&機器需求

- 1) 發佈資料集均屬公眾領域 CC0
- 2) 上傳者(非作者)須具 ORCID，組織基金來源 ROR ID
- 3) 1 dataset = 1 data package = 9-12 DOIs (6 DOIs, min)
- 4) 後設資料: 2014 開始輸出 RDF, Turtle, 與JSON-LD ，  
目前建議透過 *DataCite* 或 *CrossRef* 取用 RDF 或  
Turtle ; 常人則以 data paper 型式下載獨立 PDF
- 5) Geospatial: meta(geolocation) +  
空間範圍檢索 + 地方(地名)過濾





## DRYAD : 12歲的美國樹妖

### Timeline

- 2008 # created by US National Evolutionary Synthesis Center (NESCent)
- 2010 DOIs assigned to data packages & data files.
- 2012 *DataOne* integration
- 2013 \$\$ Data Publishing Charge (DPC)
- 2014 ORCID live
- 2016 \*Initial *GeoBlacklight* app
- 2018 Join *Datacite*
- 2018 # California Digital Library (CDL) & Dryad Partnership
- 2019 Dryad and Zenodo Partnership
- 2019 \*New Dryad | independent datasets submittion
- 2019 ROR ID (Research Organization Registry) & funder vocabulary

V1: DSpace

New: CDL Dash

# Example: a Nature Article

nature.com/articles/473285a

Published: 18 May 2011

## Data archiving is a good investment

Heather A. Piwowar  Todd J. Vision & Michael C. Whitlock

*Nature* 473, 285(2011) | [Cite this article](#)

430 Accesses | 43 Citations | 37 Altmetric | [Metrics](#)

We have found that ongoing financial investment in data-archiving infrastructure yields an impressive scientific return, and believe that it should be whole-heartedly supported by research funding agencies (see, for example, <http://go.nature.com/nzftf3>).

We used Dryad (see <http://datadryad.org>), an international, open, cost-effective data repository for the biological sciences, to estimate the cost of archiving data from more than 10,000 publications. We found that these could be curated and the data preserved at an annual cost of about US\$400,000.

Download PDF 

Sections

Author information

Ethics declarations

Rights and permissions

About this article

Further reading

Comments

[v1.datadryad.org/resource/doi:10.5061/dryad.j1fd7](https://v1.datadryad.org/resource/doi:10.5061/dryad.j1fd7)

 **DRYAD** About ▾ For researchers ▾ For organizations ▾ Contact us Log in Sign up



**Data from: Data archiving is a good investment**

Piwowar HA, Vision TJ, Whitlock MC

Date Published: April 28, 2011

DOI: <https://doi.org/10.5061/dryad.j1fd7>

**Submit data now**

[How and why?](#)

**Search for data**

Enter keyword, DOI, etc.

[Advanced search](#)

**Be part of Dryad**

We encourage organizations to:

[Become a member](#)  
[Sponsor data publishing fees](#)  
[Integrate your journal\(s\)](#), or  
All of the above

**Files in this package**

Content in the Dryad Digital Repository is offered "as is." By downloading files, you agree to the [Dryad Terms of Service](#). To the extent possible under law, the authors have waived all copyright and related or neighboring rights to this data.

Title	PubMed Central reuse of GEO datasets deposited in 2007
Downloaded	366 times
Description	This is the raw data behind the analysis. It contains one row for every mention of a 2007 GEO dataset in PubMed Central. Each row identifies the mentioned GEO dataset, the PubMed Central article that mentions the dataset's accession number, whether the authors of the dataset and the attributing article overlap, and whether this is considered an instance of third-party data reuse.
Download	<a href="#">PMC_reuse_of_2007_GEO_datasets.csv (1.842 Mb)</a>
Download	<a href="#">README.pdf (747.4 Kb)</a>
Details	<a href="#">View File Details</a>

**data package**

# data package citation

v1.datadryad.org/resource/doi:10.5061/dryad.j1fd7

DRYAD About For researchers For organizations Contact us Log in Sign up

Data from: Data archiving is a good investment Submit data now

When using this data, please cite the original publication:

Piwowar HA, Vision TJ, Whitlock MC (2011) Data archiving is a good investment. *Nature* 473: 285. <https://doi.org/10.1038/473285a>

Additionally, please cite the Dryad data package:

Piwowar HA, Vision TJ, Whitlock MC (2011) Data from: Data archiving is a good investment. Dryad Digital Repository. <https://doi.org/10.5061/dryad.j1fd7>

Cite | Share

Download the data package citation in the following formats:

RIS (compatible with EndNote, Reference Manager, ProCite, RefWorks)

BibTex (compatible with BibDesk, LaTeX)

# V1: DSpace

# data package full metadata

The diagram illustrates the transition from a DSpace page to a Dryad page, specifically focusing on the 'Show Full Metadata' feature.

**DSpace Page:** The top section shows a DSpace page with the following details:

- Pageviews: 4208
- Keywords: [data sharing](#), [research funding](#), [data archives](#), [return on investment](#), [data reuse](#)
- Abstract:** Funding agencies are reluctant to support data archiving, even though large research funders such as the National Science Foundation (NSF) and the National Institutes of Health acknowledge its importance for scientific progress. Our quantitative estimates of data reuse indicate that ongoing financial investment in data-archiving infrastructure provides a high scientific return.

A red dashed box highlights the "Show Full Metadata" button, which is connected by a curved arrow to the corresponding button on the Dryad page.

**Dryad Page:** The bottom section shows a Dryad page with the following details:

- Dryad is a nonprofit repository for data underlying the international scientific and medical literature.
- Powered by **DSPACE**
- Latest build Fri, 27 Mar 2020 15:56:23 UTC. Served by `vagrant-dryad`

The "Show Full Metadata" button on the Dryad page is also highlighted with a red dashed box and connected by a curved arrow to the "Show Full Metadata" button on the DSpace page.

**Resulting Package Details:** The "Files in this package" section displays two entries:

Title	Description
PubMed Central reuse of GEO datasets deposited in 2007	This is the raw data behind the analysis. It contains one row for every mention of a 2007 GEO dataset in PubMed Central. Each row identifies the mentioned GEO dataset, the PubMed Central article that mentions the dataset's accession number, whether the authors of the dataset and the attributing article overlap, and whether this is considered an instance of third-party data reuse.
Aggregate Table Data	Aggregate table data behind the figures and results in the README associated with the main dataset. Includes Baseline metrics used for extrapolating PubMed Central (PMC) results to PubMed. Number of mentions of a 2007 GEO dataset by authors who submitted the dataset, and Number of mentions of a dataset by authors who DID NOT submit the dataset across 2007-2010.

**URL:** <https://v1.datadryad.org/resource/doi:10.5061/dryad.j1fd7?show=full>

**DC 15 elements + Files in this package**

 DRYAD

Data from: Data archiving is a good investment

Piwowar, Heather A.  
Vision, Todd J.  
Whitlock, Michael C.  
Publication date: April 28, 2011  
Publisher: Dryad  
<https://doi.org/10.5061/dryad.j1fd7>

Citation

Piwowar, Heather A.; Vision, Todd J.; Whitlock, Michael C. (2011). Data from: Data archiving is a good Investment. Dryad, Dataset, <https://doi.org/10.5061/dryad.j1fd7>

Abstract

Funding agencies are reluctant to support data archiving, even though large research funders such as the National Science Foundation (NSF) and the National Institutes of Health acknowledge its importance for scientific progress. Our quantitative estimates of data reuse indicate that ongoing financial investment in data-archiving infrastructure provides a high scientific return.

Usage Notes

PubMed Central reuse of GEO datasets deposited in 2007  
This is the raw data behind the analysis. It contains one row for every mention of a 2007 GEO dataset in PubMed Central. Each row identifies the mentioned GEO dataset, the PubMed Central article that mentions the dataset's accession number, whether the authors of the dataset and the attributing article overlap, and whether this is considered an instance of third-party data reuse.  
[PMC\\_reuse\\_of\\_2007\\_GEO\\_datasets.csv](#)

Aggregate Table Data  
Aggregate table data behind the figures and results in the README associated with the main dataset. Includes Baseline metrics used for extrapolating PubMed Central (PMC) results to PubMed, Number of mentions of a 2007 GEO dataset by authors who submitted the dataset, and Number of mentions of a dataset by authors who DID NOT submit the dataset across 2007-2010.  
[tables.csv](#)

References

This dataset is supplement to <https://doi.org/10.1088/1742-6596/47/3/032058>

Explore Data | About ▾ | Help ▾ | Login

Search

Download dataset ~ 3 MB

Download Data Publication (PDF)

**Data Files**

April 28, 2011

PMC_reuse_o...atasets.csv	1.84 MB	
README_for_...atasets.pdf	747.42 kB	
tables.csv	1.18 kB	

**Metrics**

1847 views  
215 downloads  
2 citations

**Keywords**

research funding  
entrez  
PubMed Central  
data archives  
return on investment  
Data reuse  
Text mining  
Data sharing  
Gene Expression Omnibus

**License**

This work is licensed under a CC0 1.0 Universal (CC0 1.0) Public Domain Dedication license.

**Dataset Zip**

Download dataset ~ 3 MB

**Data Paper**

Download Data Publication (PDF)

**Data File Display**

**Data Files**

April 28, 2011

PMC_reuse_o...atasets.csv	1.84 MB	
README_for_...atasets.pdf	747.42 kB	
tables.csv	1.18 kB	

# New: CDL Dash

包含所有的  
描述性&管理性  
(descriptive & administrative)  
後設資料  
提供使用者可單一  
下載 PDF 文件

## Data Paper (metadata) in PDF

[https://datadryad.org/stash/data\\_paper/doi:10.5061/dryad.j1fd7](https://datadryad.org/stash/data_paper/doi:10.5061/dryad.j1fd7)

Piwowar, Heather A.; Vision, Todd J.; Whitlock, Michael C. ... doi:10.5061/dryad.j1fd7

**Data from: Data archiving is a good investment**

Piwowar, Heather A.  
Vision, Todd J.  
Whitlock, Michael C.  
Publication date: April 28, 2011  
Publisher: Dryad  
<https://doi.org/10.5061/dryad.j1fd7>

---

**Citation**

Piwowar, Heather A.; Vision, Todd J.; Whitlock, Michael C. (2011), Data from: Data archiving is a good investment, Dataset, <https://doi.org/10.5061/dryad.j1fd7>

---

**Abstract**

Funding agencies are reluctant to support data archiving, even though large research funders such as the National Science Foundation (NSF) and the National Institutes of Health acknowledge its importance for scientific progress. Our quantitative estimates of data reuse indicate that ongoing financial investment in data-archiving infrastructure provides a high scientific return.

---

**Usage Notes**

**PubMed Central reuse of GEO datasets deposited in 2007**  
This is the raw data behind the analysis. It contains one row for every mention of a 2007 GEO dataset in PubMed Central. Each row identifies the mentioned GEO dataset, the PubMed Central article that mentions the dataset's accession number, whether the authors of the dataset and the attributing article overlap, and whether this is considered an instance of third-party data reuse.  
[PMC\\_reuse\\_of\\_2007\\_GEO\\_datasets.csv](#)

---

**Aggregate Table Data**  
Aggregate table data behind the figures and results in the README associated with the main file. Includes Baseline metrics used for extrapolating PubMed Central (PMC) results to

Piwowar, Heather A.; Vision, Todd J.; Whitlock, Michael C. ... doi:10.5061/dryad.j1fd7

PubMed, Number of mentions of a 2007 GEO dataset by authors who submitted the data, and Number of mentions of a dataset by authors who DID NOT submit the dataset across 2007-2010.  
[tables.csv](#)

**References**

This dataset is supplement to <https://doi.org/10.1038/473285a>

**Keywords**

research funding, entrez, PubMed Central, data archives, return on investment, Data reuse, Text mining, Data sharing, Gene Expression Omnibus

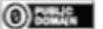
**Files**

3 files for this dataset

<a href="#">PMC_reuse_of_2007_GEO_datasets.csv</a>	1.84 MB	text/csv
<a href="#">README_for_PMC_reuse_GEO_datasets.pdf</a>	747.42 kB	application/pdf
<a href="#">tables.csv</a>	1.18 kB	text/csv

**License**

This work is licensed under a [CC0 1.0 Universal \(CC0 1.0\) Public Domain Dedication license](#).

 This releases your work to the public domain for any use.

metadata publication  
= data publication  
= data paper publication

Page 1 of 2

Dryad

Page 2 of 2

has(had) Getty TGN  
for Place Name

```
#/components/schemas/geolocation
geolocation <-
  {
    place <-
      string
      a place name.

    box <-
      swLongitude
      swLatitude
      neLongitude
      neLatitude

    point <-
      geolocationPoint <-
        latitude
        longitude
  }
```

# 空間範圍檢索

## Dryad - Blacklight Search Results

You searched for: Bounding Box > -123.068848 28.459033 -113.840332 37.300275

Start Over

Sort by relevance ▾ 10 per page ▾

« Previous | 1 - 10 of 76 | Next »

- 1. Aerial Measurements from Outdoor 2.4GHz 802.15...
- 2. Anaheim, CA Digital Elevation Models
- 3. Annual Survey of Orange County 1982
- 4. Annual Survey of Orange County 1983
- 5. Annual Survey of Orange County 1984
- 6. Annual Survey of Orange County 1985
- 7. Annual Survey of Orange County 1986
- 8. Annual Survey of Orange County 1987
- 9. Annual Survey of Orange County 1988
- 10. Annual Survey of Orange County 1989

Search when I move the map

Leaflet | © OpenStreetMap contributors, © CartoDB

# Schema.org JSON-LD exported from DataCite

```
{  
  "@context": "http://schema.org",  
  "@type": "Dataset",  
  "@id": "https://doi.org/10.5061/dryad.j1fd7",  
  "identifier": {  
    "@type": "PropertyValue",  
    "propertyID": "DOI",  
    "value": "https://doi.org/10.5061/dryad.j1fd7"  
  },  
  "url":  
    "http://datadryad.org/stash/dataset/doi:10.5061/dryad.j1fd7",  
  "additionalType": "dataset",  
  "name": "Data from: Data archiving is a good investment",  
  "author": [  
    {  
      "name": "Heather A. Piwowar",  
      "givenName": "Heather A.",  
      "familyName": "Piwowar",  
      "@type": "Person"  
    },  
    ...  
  ],  
  "description": "Funding agencies are reluctant to support data  
archiving, ...provides a high scientific return.",  
  ...  
}
```

- <http://hdl.handle.net/10255/dryad.33537>



```
"license":  
  "https://creativecommons.org/publicdomain/zero/  
  1.0",  
  "version": "1",  
  "keywords": "research funding, Data reuse, ...",  
  "inLanguage": "en",  
  "contentSize": "2590611 bytes",  
  "datePublished": "2011",  
  "@reverse": {  
    "isBasedOn": {  
      "@id": "https://doi.org/10.1038/473285a",  
      "@type": "ScholarlyArticle"  
    }  
  },  
  "publisher": {  
    "@type": "Organization",  
    "name": "Dryad"  
  },  
  "provider": {  
    "@type": "Organization",  
    "name": "DataCite"  
  }  
}
```

*dc.relation.isreferencedby*

- package/file relation:  
*dc.relation.haspart*

- domain description ex.  
*dwc:scientificName*



**DRYAD**

CC0 +  
data package

國際標準PID:  
資料集 + 人 +  
基金來源組織

常人&機器的  
後設資料

# 美國樹妖 7 大參考網址

1. Dryad Repository: <https://datadryad.org/>
2. Dryad (DSpace): <https://api.datadryad.org/>
3. Dryad REST APIs: <https://datadryad.org/api/v2/docs/>
4. Dryad GitHub: <https://github.com/CDL-Dryad>
5. Dryad wiki: <http://wiki.datadryad.org/>
6. Dryad blog: <http://blog.datadryad.org/>
7. DryadLab (for education):  
<https://api.datadryad.org/pages/dryadlab>

*Some Overlapping*  
*Promotion & Education*

*Many Thanks For Your Attentions.*