

# RESEARCH DATA SERVICES

Offering expert data assistance at every stage of the research process.

## 1: PLANNING

We can assist you with developing a data management plan and designing your planned data analysis, including:

- Implementing plans, using tools, and creating workflows for managing research data
- Advising on study design, power analysis, and choice of statistical methods
- Helping to meet increasingly stringent criteria from funding agencies

## 2: FINDING & COLLECTING

We have access to thousands of sources of data and experts who will help you:

- Locate, evaluate and format data
- Create metadata and data documentation protocols for new data collection
- Capture data using best practices and appropriate technology

## 3: ANALYZING

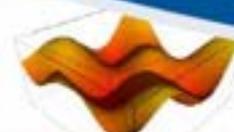
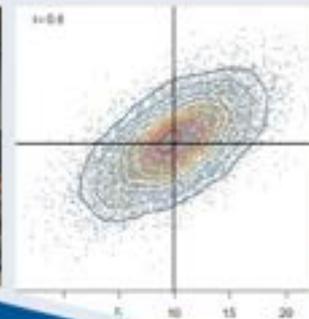
Get expert assistance from statistical, spatial, or media specialists to analyze your data and present your research:

- Learn to use cutting-edge tools and methods
- Experiment with high-resolution visualization technologies
- Develop graphical representations that bring impact to your analysis

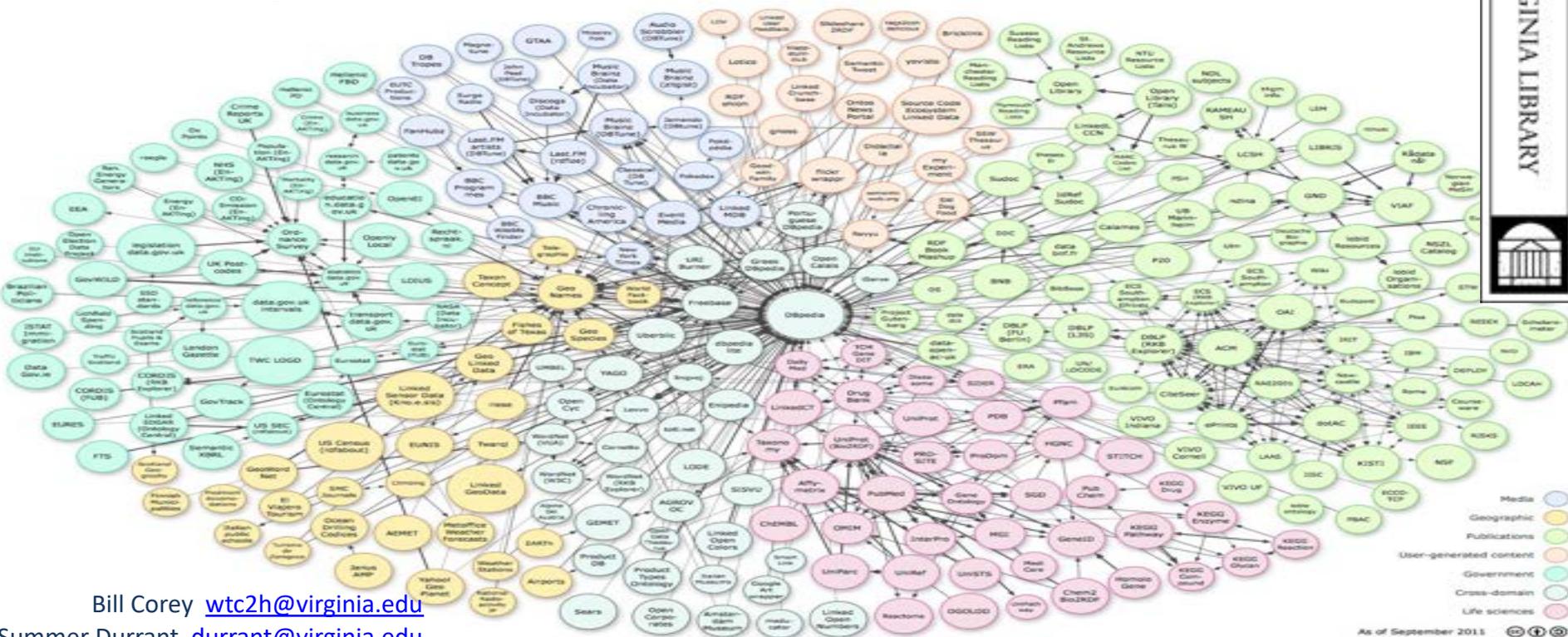
## 4: SHARING & ARCHIVING

We can consult with you on strategies to help others discover or access your research by:

- Adhering to data sharing policies and norms
- Selecting a data-sharing repository
- Making your data easier to discover and reuse



# Finding and Acquiring Data: Discovering and Obtaining Data from Library & Non-library Sources



Bill Corey [wtc2h@virginia.edu](mailto:wtc2h@virginia.edu)

Summer Durrant [durrant@virginia.edu](mailto:durrant@virginia.edu)

Research Data Services

University of Virginia Library

Linking Open Data cloud diagram, by Richard Cyganiak and Anja Jentzsch. <http://lod-cloud.net/>

# Workshop Goals

**You're looking for data for your project: where do you find it and how do you acquire it?**

You'll learn how to :

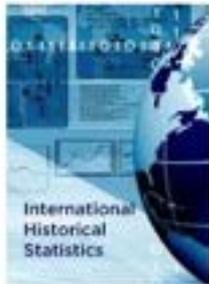
- identify potential sources of data
- locate the data you need
- gain access to it

We'll explore the library's collections and subscriptions, in addition to data residing in repositories that can be identified by data citations and article references.

# Data Collections

<http://guides.lib.virginia.edu/datastats>

## Featured Resource



- International Historical Statistics, 1750-2010  
Extensive collection of statistical data covering a broad range of socio-economic topics.

## Recent Acquisitions

- Dave Leip Election Data 

## Welcome

Welcome to the **Data and Statistics Research Guide**. This guide contains information on important resources for locating current and historical data from a variety of sources, including national governments, international organizations, researchers, and research firms.

- **Databases A-Z** - see a list of our licensed databases
- **Purchased Datasets** - access datasets acquired by the library
- **United States** - statistical compendia, census, economic indicators
- **International** - IMF, UN, OECD
- **Research Data Repositories** - ICPSR, Roper, Odum

## Research Assistance

Need help finding data for your research project? **Contact me** to schedule a research consultation. Here are some examples of the ways I can assist you:

- guidance on potential data sources
- accessing datasets
- understanding data documentation / codebooks
- acquiring data in desired format
- citing data

## Data Librarian



Summer Durrant

### Contact Info

Data Librarian  
University of Virginia  
durrant@virginia.edu  
[Send Email](#)

### Links:

[Profile & Guides](#)

## Research Data Services

- [March 2014 Newsletter](#)

# Planning Phase

Gather information about the data set you hope to find:

## Known Dataset

- Name of study or data set
- Principle investigator(s)
- Data producer or vendor
- Edition or version

## Variable Search

- Variable(s)
- Geographic coverage
- Time period
- Frequency

# Questions

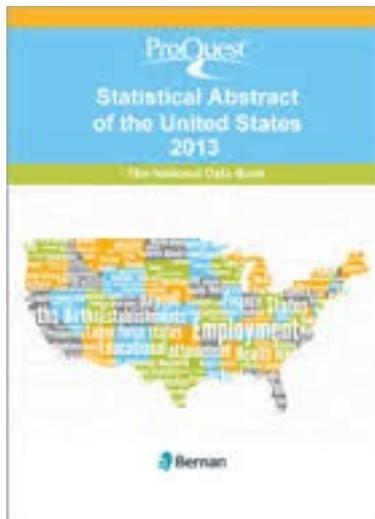
## Who is likely to collect data on this topic?

- government agencies
- academic researchers
- IGOs and NGOs
- non-profit organizations
- private sector

## How are data likely to be collected?

- survey (census, sampling)
- administrative records

# Search Strategy #1: Start with a Statistical Publication



**Table 1259. Performing Arts—Selected Data: 1990 to 2011**

[Sales, receipts, and expenditures in millions of dollars (262 represents \$262,000,000). For season ending in year shown, except as indicated]

Item	1990	1995	2000	2005	2006	2007	2008	2009	2010	2011
<b>Legitimate theater:</b> <sup>1</sup>										
<b>Broadway shows:</b>										
New productions	40	33	37	39	39	35	36	43	39	42
Attendance (mil.)	8.0	9.0	11.4	11.5	12.0	12.3	12.3	12.1	11.9	12.5
Playing weeks <sup>2,3</sup>	1,070	1,120	1,464	1,494	1,501	1,509	1,560	1,548	1,464	1,588
Gross ticket sales	282	406	603	769	862	939	938	943	1,020	1,060
<b>Broadway road tours:</b> <sup>4</sup>										
Attendance (mil.)	11.1	15.6	11.7	18.2	17.1	16.7	15.3	14.3	15.9	13.1
Playing weeks	944	1,242	888	1,389	1,377	1,400	1,138	1,112	1,242	1,003
Gross ticket sales	367	548	572	707	831	833	955	881	947	803
<b>Nonprofit professional theatres:</b> <sup>5</sup>										
<b>Companies reporting:</b> <sup>6</sup>	185	215	262	1,490	1,893	1,910	1,919	1,825	1,807	(NA)
Gross income	308	444	791	1,647	1,791	1,881	1,884	1,779	1,913	(NA)
Earned income	188	281	496	845	923	962	955	811	964	(NA)
Contributed income	119	163	325	802	868	919	929	968	948	(NA)
Gross expenses	306	445	708	1,530	1,667	1,742	1,860	1,892	1,870	(NA)
Productions	2,265	2,646	3,241	12,000	14,000	17,000	15,000	17,000	16,000	(NA)
Performances	46,131	56,608	66,123	169,000	172,000	197,000	202,000	187,000	163,000	(NA)
Total attendance (mil.)	15.2	18.6	22.0	32.5	30.5	31.0	32.0	30.0	31.0	(NA)
<b>OPERA America professional member companies:</b>										
Number of companies reporting <sup>7</sup>	98	88	98	83	94	97	85	84	97	83
Expenses <sup>8</sup>	321	435	637	742	762	872	826	816	883	688
Performances <sup>9</sup>	2,336	2,120	1,768	1,893	1,851	1,961	1,753	1,744	1,677	1,407
Total attendance (mil.) <sup>10</sup>	7.5	4.1	6.2	5	5.3	5.3	5.1	4.3	6.8	6.6
Main season attendance (mil.) <sup>11</sup>	4.1	3.9	3.8	3.3	3.4	3.6	3.1	2.9	2.2	1.8
<b>Symphony orchestras:</b> <sup>11</sup>										
Concerts	18,931	29,328	33,154	37,196	36,731	37,109	33,029	32,813	35,348	35,968
Attendance (mil.)	24.7	30.9	31.7	26.5	29.1	28.8	28.7	25.4	26	26
Gross revenue	378	536	734	812	945	1,052	992	969	1,030	1,038
Operating expenses	622	859	1,126	1,513	1,603	1,808	1,862	1,864	1,849	1,822
Support	258	351	521	626	713	721	785	726	695	820

NA Not Available <sup>1</sup> Source: The Broadway League, New York, NY. For season ending in year shown, <sup>2</sup> All shows (new productions and holdovers from previous seasons). <sup>3</sup> Eight performances constitute one playing week. <sup>4</sup> North American Tours include U.S. and Canadian companies. <sup>5</sup> Source: Theatre Communications Group, New York, NY. For years ending on or prior to August 31. <sup>6</sup> Beginning in 2002, nonprofit theatre data is based on survey responses and extrapolated data from IRS Form 990. <sup>7</sup> Source: OPERA America, © 2004, 2005, 2011 Cultural Data Project. © 2001, 2009 OPERA America. For years ending on or prior to August 31. <sup>8</sup> U.S. and Canadian companies for 1999 and 2009-10; U.S. companies only for all other years. <sup>9</sup> Includes educational performances, outreach, etc. <sup>10</sup> For paid performances. Data for the Metropolitan Opera are not included. <sup>11</sup> Source: League of American Orchestras, New York, NY. For years ending August 31. Prior to 1995, represents 254 U.S. orchestras; beginning 1995, represents all U.S. orchestras, excluding college/university and youth orchestras. Beginning 1995, data based on 1,200 orchestras.

Source: Compiled from sources listed in footnotes. See also <<http://www.livebroadway.com/>>; <<http://www.tcg.org/>>; <<http://www.operaamerica.org/>>; and <<http://www.americanorchestras.org/>>

# Search Strategy #2: Use the Methodology/Data Section of Articles

## I. Methodology and Data

It is important to use a consistent theoretical framework to define income and consumption. The most comprehensive concept of income and consumption is drawn from the writings of Haig and Simons. Haig (1921) stated that income was “the money value of the net accretion to one’s economic power between two points of time” (p. 7) and Simons (1938) defined personal income as “the algebraic sum of (1) the market value of rights exercised in consumption and (2) the change in the value of the store of property rights between the beginning and end of the period in question” (p. 50). No current household survey has the necessary variables to create a full measure of Haig-Simons income.<sup>3</sup>

Our goal is to have measures of disposable income and consumption that are accurate and as closely linked as possible (given the data limitations) to compare their distributions. While there may be reasons to exclude durable goods, education, or medical care from the measurement of consumption, removing these items from con-

equivalence of homeownership, and the value of federal government rental assistance.<sup>4</sup>

To match the income and consumption for each household and obtain annual measures of consumption, we use only those consumer units who participated in the survey for all four quarters. In this manner, we obtain the income and consumption for the same 12-month period. We do not restrict our sample by age, place of residence, or income reporting status as previous research has done. Previous papers also restricted their samples to “complete income reporters” as defined by the CE Survey.<sup>5</sup>

The CE Survey began imputing income in 2004 but did not impute previous years.<sup>6</sup> We replicate the BLS methodology as closely as possible and impute income for 1985–2010. For income taxes we use the National Bureau of Economic Research’s TAXSIM program (see Feenberg and Coutts 1993) to estimate federal, state, and FICA taxes, and tax credits such as the Earned Income Tax Credit.<sup>7</sup> All values are equivalized using the square root of household size, and the weights are adjusted to account for using only four-quarter consumer units and to reflect person weights.

# Search Strategy #3: Search in Data Repositories

The screenshot displays the ICPSR website's search interface. At the top, a navigation bar includes links for 'Log In/Create Account', 'Find & Analyze Data', 'Membership In ICPSR', 'Deposit Data', 'ICPSR Summer Program', 'Resources for Instructors', and 'Data Management & Curation'. Below this is the 'ICPSR Find & Analyze Data' header. A secondary navigation bar offers 'Find Data', 'Search/Compare Variables', 'Find Publications', 'Resources for Students', and 'Get Help'. The main content area is divided into three columns: 'Our Mission' (describing ICPSR's role in social and behavioral research), 'Announcements' (with links to guides and releases), and 'Find Data' (featuring a search bar and 'More Search Options' like 'Browse by topic', 'Browse by series', 'Browse by geography', 'Browse by investigator', 'Browse international data', and 'View all studies'). A 'Restricted Data' section explains confidentiality measures. On the right, a 'Search Tips' box provides advice on effective searching and syntax, such as using Boolean operators, stemming, and quotes.

# Discovering & obtaining data from non-library sources

What do you do if you have researched your potential data source and it is not available through one of the databases that the library has a license to?

You may want to contact the researcher directly if they have mentioned a dataset in an article that your citation research identified.

You can widen your search to include resources that are available on the internet, via Google or Bing. One of the easiest ways to do this is by **searching in a directory of data repositories.**

# Uncovering data on the internet

It is second nature to search the internet for everything and anything.

- But how do we know if the resources we find are 'good' resources?
- How do we know if we can use them for our own research?

Materials on the web were put there by somebody. It is always important to remember that anything you find online belongs to someone. To use it appropriately, you need to identify the owner and obtain permission.

# Data Registries

Finding the source of your ideal dataset can be intimidating, confusing, and very time consuming.

There are 2 international registries of data repositories that can speed up that search (Databib & Re3data merged in 2015):

**Databib** <http://databib.org/>

**Re3data** <http://www.re3data.org/>

There is also a list of data repositories by subject at Simmons College.

[http://oad.simmons.edu/oadwiki/Data\\_repositories](http://oad.simmons.edu/oadwiki/Data_repositories)

# Advantages of Data Registries

## They provide information to narrow your search:

- Persistent Identifiers – unique, citable, and searchable
- Access controls
- Terms of Use & Licenses
- Data preservation -- migrating to new formats or emulating old formats
- Professional backup & documentation
- Repository Standards ensure commitment and quality



<http://www.plosone.org/article/info:doi/10.1371/journal.pone.0078080>

# Finding data using a data registry

**Databib and re3data can be searched at multiple levels.**

The registries are similar in their content:

- **re3data** has more European & Asian repositories
- **Databib** has more from the America's.

Both organizations have recently agreed to merge their resources into a single registry in 2015, under the **DataCite** umbrella. **DataCite** provides DOI's for datasets: Digital Object Identifiers, which allow datasets to be cited just like an article. A dataset with a DOI can be found more easily by other researchers.

# Finding data in a data repository

Search for **History**.

Databib has 2 offerings, while re3data has 17.

Look at the information in a particular repository record.

You've probably noticed the icons to the right of the repository name in re3data's listings. Those icons provide a lot of useful info up front, so you don't have to dig for it. These icons will be the standard in the combined registry.

Let's look at those icons, and then look at Sandrart.net.

# re3data.org icons

re3data.org  
REGISTRY OF RESEARCH DATA REPOSITORIES

<http://re3data.org>

## What do the icons mean?



[http://www.plosone.org/article/fetchObject.action?uri=info:doi/10.1371/journal.pone.0078080.g001&representation=PNG\\_M](http://www.plosone.org/article/fetchObject.action?uri=info:doi/10.1371/journal.pone.0078080.g001&representation=PNG_M)

# The re3data record for Sandrart.net

## Repository details

### Sandrart.net



[← Back to results](#)

General

Institutions

Terms

Standards

## General information

Name of repository	<b>Sandrart.net</b>
Repository URL	<a href="http://ta.sandrart.net/en/">http://ta.sandrart.net/en/</a>
Subjects	<a href="#">Q Art History</a> <a href="#">Q Fine Arts, Music, Theatre and Media Studies</a> <a href="#">Q Humanities</a> <a href="#">Q Humanities and Social Sciences</a>
Description	Sandrart.net: A net-based research platform on the history of art and culture in the 17th century. The project's main goal was an annotated, enriched and web-based edition of Joachim von Sandrart's Teutscher Academie der Edlen Bau, Bild- und Mahlerey-Künste (1675–80), one of the most important source texts of the early modern period. Having lived and worked in a number of places throughout Europe, Sandrart's biographical background makes his writings (with first-hand narrations on art, artists and art collections) a work of European dimension.
Content types	<a href="#">Q Archived data</a> <a href="#">Q Images</a> <a href="#">Q Plain text</a> <a href="#">Q other</a>
Keywords	<a href="#">Q Joachim von Sandrart</a> <a href="#">Q ancient architecture</a> <a href="#">Q painting</a> <a href="#">Q curriculum vitae</a> <a href="#">Q 17th century</a> <a href="#">Q art writer</a>
Repository type	disciplinary
Research data repository language(s)	deu eng
Data and/or service provider	dataProvider

Both registries show the same entry text for the Sandrart Repository. Both provide a URL to the repository, both tell you it is a data provider (not accepting deposits), and that it is a disciplinary repository.

# The Databib record for Sandrart.net

**Title:** Sandrart.net

**URL:** <http://ta.sandrart.net/en/>

**Authority:**

**Subjects:** History

## Description:

Sandrart.net: A net-based research platform on the history of art and culture in the 17th century. The project's main goal was an annotated, enriched and web-based edition of Joachim von Sandrart's *Teutscher Academie der Edlen Bau, Bild- und Mahlerey-Künste* (1675–80), one of the most important source texts of the early modern period. Having lived and worked in a number of places throughout Europe, Sandrart's biographical background makes his writings (with first-hand narrations on art, artists and art collections) a work of European dimension.

**Access:** open

**Start Date:** 2007

**Reuse:** restricted; ODC - <http://opendefinition.org/licenses/odc-odbl/>

**Deposit:** closed

**Type:** disciplinary



## Annotations

Both registries show the same entry text for the Sandrart Repository. Both provide a URL to the repository, both tell you it is a data provider (not accepting deposits), and that it is a disciplinary repository.

# The re3data record for Sandrart.net

## Repository details

**Sandrart.net**



[← Back to results](#)

[General](#) [Institutions](#) [Terms](#) [Standards](#)

### Terms (1)

Policy name	Nutzungsbedingungen
URL	<a href="http://ta.sandrart.net/de/info/nutzungsbedingungen/">http://ta.sandrart.net/de/info/nutzungsbedingungen/</a>

### Legal aspects

#### Database access

Type of access to research data repository	open
--	------

#### Data access

Type of access to data	open restricted
Data access restriction type(s)	registration

#### Data licences (1)

License name	ODC
License URL	<a href="http://opendefinition.org/licenses/odc-odbl/">http://opendefinition.org/licenses/odc-odbl/</a>

#### Data upload

Type of data upload	closed
---------------------	--------

But re3data provides additional information on the other tabs: **Institutions** responsible for the repository, **Terms** of use and **Standards**. You might also notice that Databib says it is a “restricted” repository, while re3data says it is open. In reality, it is both.

# Sandrart.net

Deutsche Version

**Sandrart**.NET

Edition Information  
Project Information  
Editors  
Contact/Imprint

## Teutsche Academie der Bau-, Bild- und Mahlerey-Künste

Joachim von Sandrart, Nürnberg 1675/1679/1680

Scholarly annotated online edition,

ed. by Thomas Kirchner, Alessandro Nova, Carsten Blüm, Anna Schreurs and Thorsten Wübbena, 2008–2012

### *At a glance:*

*"Sandrart.net in a Nutshell"*

Full text of the "Teutsche Academie"  
Structure of the "Teutsche Academie"  
Translations

Search  
People  
Places  
Works of Art  
Bibliography

The Sandrart.net site itself also provides additional citation information, as well as navigation tools, a French translation, a PURL for every page and image, and TEI, REST-Service and PND-Beacon feeds for the basic files.

Citation guidelines

# openICPSR

Warning! This is a test site. Any material you upload, create, publish, or share will be periodically removed from our systems. Use at your own risk. All feedback is welcome, though, and can be submitted via the question mark at the bottom of the page.

Three easy steps: 1. Name your project 2. Upload and describe files 3. Purchase a curation option



Watch our videos >



Share your social and behavioral science research data

Get started now >

Maximize Access

Be recognized and cited

Store Safely

Store your data with confidence

ICPSR is creating a new repository that they will call openICPSR. It will be separate from the institutionally-funded ICPSR data repository, will require a fee to deposit the data, and will preserve for 10 years. In the Self-Deposit Package, the data depositor is responsible for the documentation & anonymization of the data.

# Data Archives & Repositories

## Examples



# What have we learned about locating data?

**There is a LOT of data out there that you can reuse.**

- re3data.org currently lists 586 repositories
- Databib lists 971 repositories

**There are many sources of data that aren't in a repository.**

- Disciplines may have an awareness of and access to data repositories that they don't share with others.
- Colleagues may be working with datasets that are still being created. Ask them!
- Authors may cite data in an article, but provide no information about it. Contact them!

# Best Practices for identifying and obtaining datasets

When identifying and obtaining data from non-library sources, it is important to remember a few best practices.

- someone created the data, so who owns it?
- do I have permission to do what I want with the data?
- are there restrictions on reuse?
- does the dataset come with documentation that makes it understandable and useful?
- is the data freely accessible, or is it restricted access?
- do I need to purchase the data?
- do I need to register with the repository to obtain the data?



DVN hosts multiple, individually-branded Dataverses

- Researchers control the design, content, dissemination of their Dataverse, and can embed it in their own webpage
- DVN assigns handles (persistent id) and Universal Numerical Fingerprint (data fixity/verification)
- Extracts metadata for discovery, imports/exports metadata in multiple XML formats (DDI, Dublin Core, FGDC); data across DVNs searchable within one DVN
- Accepts data in multiple formats (Stata, SPSS, CSV), converts to preservation format
- Data can be subset, recoded, analyzed online
- Data sharing on DVN:

[http://thedata.org/files/thedata\\_new2/files/gettingstartedguidefinal.pdf](http://thedata.org/files/thedata_new2/files/gettingstartedguidefinal.pdf)

# Code & Multi-discipline Data Archives & Repositories

**Open Science Framework:** <https://openscienceframework.org/> Scientists can use OSF for free to archive, share, find, and register research materials and data.

**figshare:** <http://figshare.com/> “figshare is a repository where users can make all of their research outputs available in a citable, shareable and discoverable manner.”

**GitHub:** <https://github.com/> Share and archive code and collaborate with anyone.

**runmycode:** <http://www.runmycode.org/> Online repository allowing people to share and download computer code and data associated with scientific publications.

**ResearchGate:** <http://www.researchgate.net/> Online research social networking. Collaborate, share publications and data.

# Data Sources JISC

**JISC content**

Home Find a collection Explore themes Inspire me

Search for collections

All collections Filter by available

## Digital collections and archives for learning, teaching and research

This website provides an introduction to digital collections designed for education. They are mainly aimed at university students, researchers and librarians but many of the entries address an area as broad as anyone. The collections cover areas such as: [Arts](#), [Business](#), [Education](#), [Health](#), [History](#), [Languages](#), [Literature](#), [Mathematics](#), [Music](#), [Science](#), [Social Sciences](#), [Sports](#), [The Arts](#), [The Environment](#), [The Humanities](#), [The Sciences](#), [The Social Sciences](#), [The Arts](#), [The Environment](#), [The Humanities](#), [The Sciences](#), [The Social Sciences](#).

**1818 1883**

**A Vision of Britain Through Time**

Two centuries of rapid, wide-reaching change charted across the British population.

**How to get started...**

- Quickly find a collection using a search engine.
- Explore the collection theme of interest.
- See how the collections can help learning, teaching and research.
- Create an account and save your favourites.

**Resource types**

Statistics, Databases, Lists, Maps, Catalogue, 3D6 image, Text

**Time period**

18th Century (1700 - 1799)  
19th Century (1800 - 1899)  
20th Century (1900 - 1999)  
21st Century (2000 - present)

**Publisher**

University of Portsmouth

**JISC content**

Home Find a collection Explore themes Inspire me

Search for collections

All collections Filter by available

## Find a collection

Quick and easy ways to find a collection

This section includes tools for our collections of digital archived materials and resources that can help with teaching, research and study. To find a collection that meets your interests simply filter the list by choosing a subject area, searching the A-Z menu or combining a quick keyword search.

**Collection by subject**

Find a collection by subject area

Properties: All collections Search

Arts and Humanities  
Architecture and Planning  
Communication and Media Studies  
Creative and Performing Arts  
Education  
Health and Life Sciences  
Humanities  
Law  
Life and Physical Sciences  
Mathematics  
Medicine and Allied Health  
Natural Sciences  
Physical Sciences  
Social Sciences  
The Arts  
The Environment  
The Humanities  
The Sciences  
The Social Sciences

**A-Z search**

Find a collection using the A-Z menu

ABCDEFGHIJKLMNOPQRSTUVWXYZ

0 1 2 3 4 5 6 7 8 9

**Resource types**

Statistics, Databases, Lists, Maps, Catalogue, 3D6 image, Text

**Subject area**

Arts and Humanities, Architecture and Planning, Communication and Media Studies, Creative and Performing Arts, Education, Health and Life Sciences, Human Sciences, Law, Life and Physical Sciences, Mathematics, Medicine and Allied Health, Natural Sciences, Physical Sciences, Social Sciences, The Arts, The Environment, The Humanities, The Sciences, The Social Sciences

**A vision of Britain through time**

Home Places Statistical data Historical maps Census reports Travel writing Learning zone Help/Contact

Search for places

Find a place

Search by geographical area

Historical maps

Travel writing

Census reports

Election results

Old Maps Online

**Historical maps**

Thousands of maps from the 18th and 19th centuries

**Travel writing**

This is the largest collection of historical travel writing on the web. Each author gives a different perspective on the land and people they see.

**Census reports**

More than 100 years of census data available in a searchable format.

**Election results**

See how people from across the country voted in the 19th and 20th centuries.

**Old Maps Online**

A searchable online map of the United Kingdom and Ireland, showing the country as it was in the past.

# Questions?

Thanks for attending today's workshop: **Finding and Acquiring Data: Discovering and Obtaining Data from Library & Non-library Sources.**

Please contact us if you have any questions, or need assistance in locating and obtaining data for your research activities.

## **Summer Durrant**

Data & Geographical Information Librarian  
Research Data Services  
University of Virginia Library  
[durrant@virginia.edu](mailto:durrant@virginia.edu)

## **Bill Corey**

Data Consultant  
Research Data Services  
University of Virginia Library  
[wtc2h@virginia.edu](mailto:wtc2h@virginia.edu)