



JSTOR

Data for Research Service (DfR)

Michael Gallagher  
JSTOR Training & Education  
Princeton, New Jersey 08540  
Tel: 609 986-2263

[michael.gallagher@jstor.org](mailto:michael.gallagher@jstor.org)

# What is DfR?

- A self-serve tool for obtaining research data from the JSTOR archive
- A researcher-oriented exploration tool complementing the search and browse capabilities offered by the JSTOR main site
- A lot of fun

# What is JSTOR?

- 5.4 million journal articles
  - ~14 billion words
- 33 million pages of text
- 50 disciplines
- 350+ years of academic thought

# What value does DfR provide for researchers?

- Spot trends in academic language
- Determine currency of academic language
- Isolate key terms for whole disciplines
- Powerful faceted search interface
- Download large data sets

# Data for Research – Explore Tool

mgallagher1 | [Edit Account](#) | [Log out](#)



## Data for Research Beta

[Explore](#)

[My Data Requests](#)

[Help](#)

[About](#)

[Contact Us](#)

Results: 5,124,080

[Summary](#) | [Results List](#) | [Key Terms](#) | [References Profile](#) | [Submit Data Request](#)

**Search** ?

  
  
Field: [Anywhere](#) ▼

Narrow results by:

▼ **Year of Publication**



▸ Resource Type

▸ Discipline Group

▸ Discipline

▸ Journal

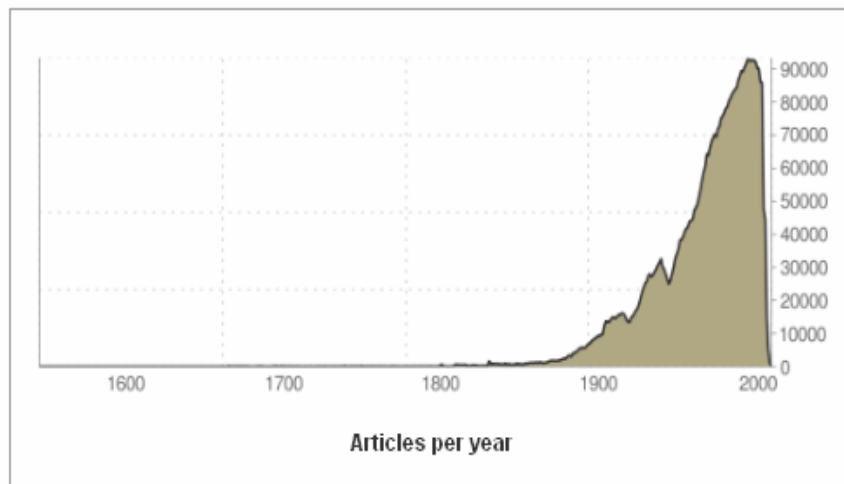
▸ Publisher

▸ Author

▸ Article Type

▸ Language

▸ Images



[Download Chart Data](#) ▼



## Research-oriented data views of JSTOR content

- Word frequencies
- N-grams (bigrams, trigrams, quadgrams)
- Key terms
- Reference citations and statistics

## Data retrieval options

- Online viewing
- Bulk downloading API access
- API

## Data Visualizations

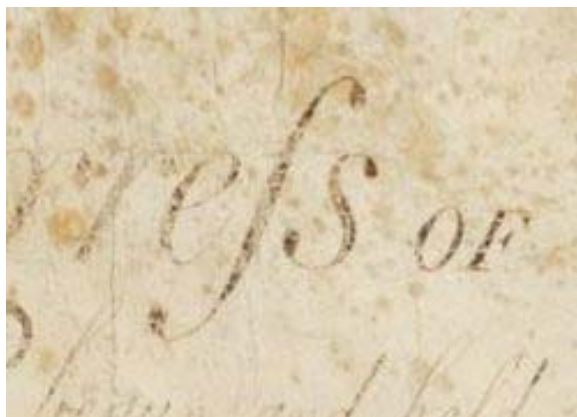
- Documents by year
- Documents by discipline group
- Reference Profiles
  - Average references per document (by year)
  - Average age of references per document (by year)
- Keyword tag cloud

## Future Plans

- More visualizations
  - Ngram cloud
  - Citation networks
  - Side-by-side comparisons
- Auto-extracted key phrases
- Search for similar or related documents
- Expand API
  - Support for faceting

# Data for Research Demo

- Use of the long 's' - The long 's' is a form of the minuscule letter 's' formerly used where 's' occurred in the middle or at the beginning of a word.



- Since the long s is generally misinterpreted by OCR engines as the character 'f' it's easy to see when its use began to wane.
  - [Occurrences of the word 'thefe' in OCR text](#)

# Data for Research - Demo

- Introduction of terms
  - ["Climate change"](#)
  - ["Financial derivatives"](#)
  - ["Swine flu"](#)
- Changing citation patterns
  - [Corpus wide](#)
  - By field/discipline:
    - [Sciences](#)
    - [Humanities](#)
- Tools for citation analysis
  - Search of articles citing the [New York Times](#)



Results: 5,124,080

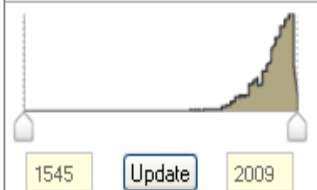
[Summary](#) [Results List](#) [Key Terms](#) [References Profile](#) [Submit Data Request](#)

**Search** ?

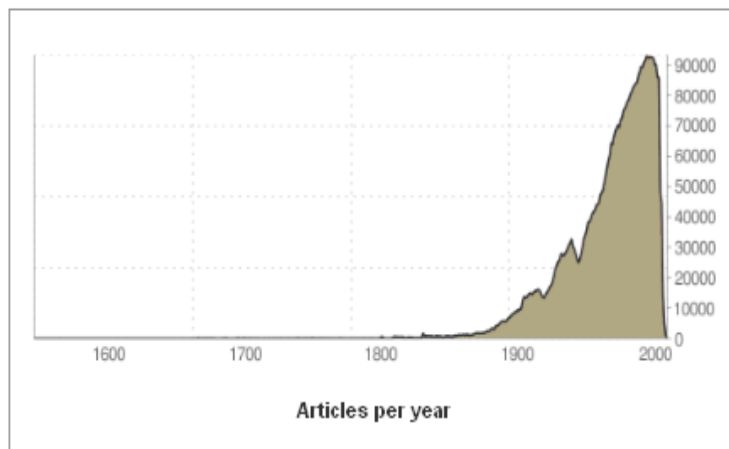
  
  
Field: [Anywhere](#) ▼

**Narrow results by:**

▼ **Year of Publication**



- ▶ Resource Type
- ▶ Discipline Group
- ▶ Discipline
- ▶ Collection
- ▶ Journal
- ▶ Publisher
- ▶ Author
- ▶ Article Type
- ▶ Language
- ▶ Images



[Download Chart Data](#)



[Download Chart Data](#)





[Explore](#)

[My Data Requests](#)

[Help](#)

[About](#)

[Contact Us](#)

Results: 17,007

[Summary](#)

[Results List](#)

[Key Terms](#)

[References Profile](#)

[Submit Data Request](#)

Search



Go

Field: [Anywhere](#)

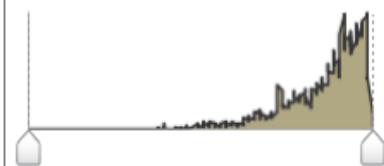
Selection Criteria

[Clear All](#)

Full Text:  
swine

Narrow results by:

▼ Year of Publication



1664

Update

2008

► Resource Type

► Discipline Group

► Discipline

Results: 1 - 10 of 17,007

Sort By: [Relevance](#)

Go

Page 1 of 1,701

« Previous

1

2

3

...

1,701

Next »

1. **Serologic Studies of Swine Influenza in Hawaii**

R. M. Nakamura

*The Journal of Infectious Diseases*, Vol. 126, No. 2 (Aug., 1972), pp. 210-211

Extracted key terms: swine, influenza, hawaii, virus, titer, serologic, antibody, farm, hemagglutination, detect, disease, test, inhibition, respiratory, wisconsin, antigen, clinical, severe, positive, sample, serum, erythrocyte, midwestern, state

[Word Counts](#) | [Bigrams](#) | [Trigrams](#) | [Quadgrams](#) | [Key Terms](#) | [References](#)

2. **The Prevalance of Trichiniasis in Swine in the United States, 1960-70**

W. J. Zimmermann, D. E. Zinter

*HSMHA Health Reports*, Vol. 86, No. 10 (Oct., 1971), pp. 937-945

Extracted key terms: swine, garbage, percent, infect, raise, prevalence, trichinosis, butcher, breeder, spirali, carcass, sample, diaphragm, market, program, state, study, examine, eradication, trichinella, disease, originate, scrap

[Word Counts](#) | [Bigrams](#) | [Trigrams](#) | [Quadgrams](#) | [Key Terms](#) | [References](#)

3. **Swine Influenza Virus Infections in Humans**

Walter R. Dowdle, Michael A. W. Hattwick

*The Journal of Infectious Diseases*, Vol. 136 (Dec., 1977), pp. S386-S389

Extracted key terms: swine, influenza, virus, human, antibody, Shope, disease, serologic, infection, contact, wisconsin, isolate, jersey, pandemic, pair, evidence, transmission, center, surveillance, titer, myalgia

[Word Counts](#) | [Bigrams](#) | [Trigrams](#) | [Quadgrams](#) | [Key Terms](#) | [References](#)

4. **Antibody Responses of Swine to Type A Influenza Viruses during the Past Ten Years in Japan**

H. Goto, Y. Ogawa, T. Hirano, Y. Miwa, F. Z. Piao, M. Takai, S. Noro, N. Sakurada

*Epidemiology and Infection*, Vol. 100, No. 3 (Jun., 1988), pp. 523-526

Extracted key terms: swine, influenza, virus, antibody, human, japan, sugimura, strain, hokkaido, veterinary, japanese, jersey, ishida, epidemic, shimizu, yamane, year, serological, journal, prevalence, virology, period

[Word Counts](#) | [Bigrams](#) | [Trigrams](#) | [Quadgrams](#) | [Key Terms](#) | [References](#)





Results: 118,628

Search



Go

Field: Anywhere

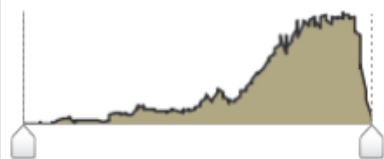
Selection Criteria

Clear All

Discipline:  
Anthropology

Narrow results by:

Year of Publication



1848

Update

2007

Resource Type

Discipline Group

Discipline

Summary

Results List

Key Terms

References Profile

Submit Data Request

### Extracted Keyterms

[Download Keyterm Counts](#)

aboriginal account africa african america american analysis ancient animal anthropological anthropologist anthropology archaeological archaeologist archaeology archeological archeology arizona article artifact author behavior belief bibliography black british brother burial california century ceramic ceremony change chapter chief child chinese class cloth coast collection colonial community concept country cultural culture custom dance department development editor ein english eskimo essay ethnic ethnographic ethnography ethnology european evidence evolution excavation family father female field fieldwork folklore form gender guinea health historical house household human hunting identity index india indian indigenou individual informant information interest island issue japanese journal kinship language lied linguistic local london manuscript marriage material meaning medicine meeting member mexican MEXICO model mother museum music musical name narrative native nicht north northern note number object organization origin page paper pattern peasant people period person plain plate political population pottery power practice prehistoric prehistory present press primitive problem production professor publication publish pueblo reader record recording region relation relationship religion religious report review ritual river rural school science settlement site skull society song south spanish spirit state stone story structure student study style system table tale text theoretical theory tradition traditional tribal tribe type university urban valley village washington water werden western white woman world  
MLT



# Request Form



Results: 118,628

[Summary](#) [Results List](#) [Key Terms](#) [References Profile](#) [Submit Data Request](#)

Results: 118,628

### Download Options

Some time will be needed to process and assemble your dataset after you submit your request. You will be emailed when this is available to download.

Please be aware that you are only allowed **1,000 articles** per dataset request and your search contains **118,628 articles**.

If you would like to request access to more articles, please contact us at [dfc@jstor.org](mailto:dfc@jstor.org). In your email provide a description of your research project and the number of articles desired.

Please provide a title for your search. This title will help you identify this search in a list of other pending or existing searches.

#### Search

Field: [Anywhere](#)

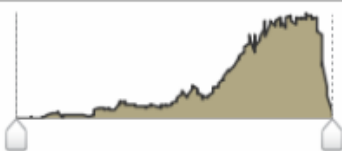
#### Selection Criteria

[Clear All](#)

Discipline:  
[Anthropology](#)

#### Narrow results by:

##### Year of Publication



1848  2007

▸ Resource Type

▸ Discipline Group

##### Discipline

- [Anthropology](#) (118,628)
- [Archaeology](#) (14,443)
- [Language & Literature](#)

#### Data Type:

- Citations Only (all requests come with citations by default)
- Word Counts
- Bigrams
- Trigrams
- Quadgrams
- Key Terms
- References

#### Output Format:

- XML
- CSV

All data files in one directory

Job Title:



# References Profile



## Data for Research

Beta

mgallagher1 | [Edit Account](#) | [Log out](#)

[Explore](#)

[My Data Requests](#)

[Help](#)

[About](#)

[Contact Us](#)

Results: 20,369

[Summary](#)

[Results List](#)

[Key Terms](#)

[References Profile](#)

[Submit Data Request](#)

Search



Go

Field: [Anywhere](#)

Selection Criteria

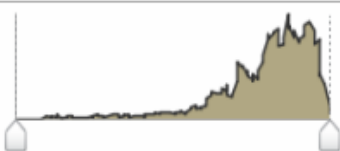
[Clear All](#)

Discipline:

Anthropology

Narrow results by:

Year of Publication



1848

Update

2007

Resource Type

Discipline Group

Discipline

Anthropology (20,369)

Archaeology (5,272)

Linguistics (2,112)

Results Set

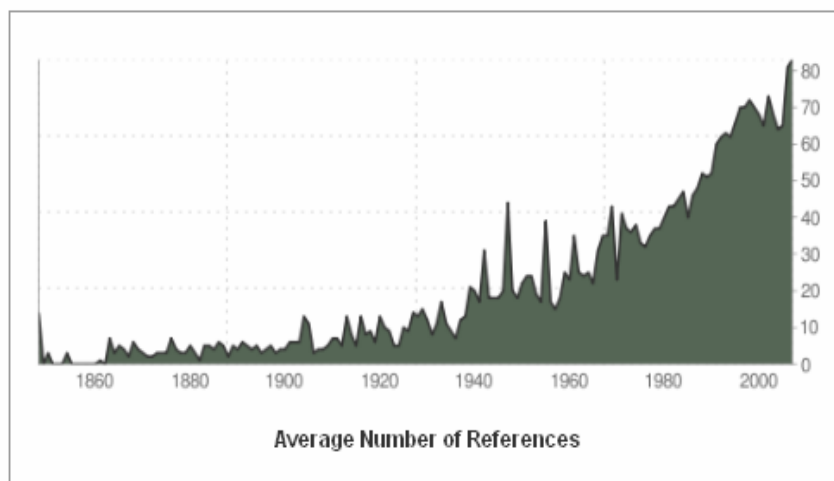
JSTOR Corpus

Number of documents with references: 20,369 (100.0%) 1,304,243 (25.5%)

Total references: 869,330 36,220,914

Mean references per document: 42.7 27.8

Mean reference age: 17.6 16.8



[Download Chart Data](#)

