

REUTERS/Nicky Loh

THE DATA CITATION INDEX & DATACITE

NIGEL ROBINSON

26 AUGUST 2014



THOMSON REUTERS

OVERVIEW

- What is the Data Citation Index
- Collaboration with DataCite
- Requirements to participate



DATA CITATION INDEX

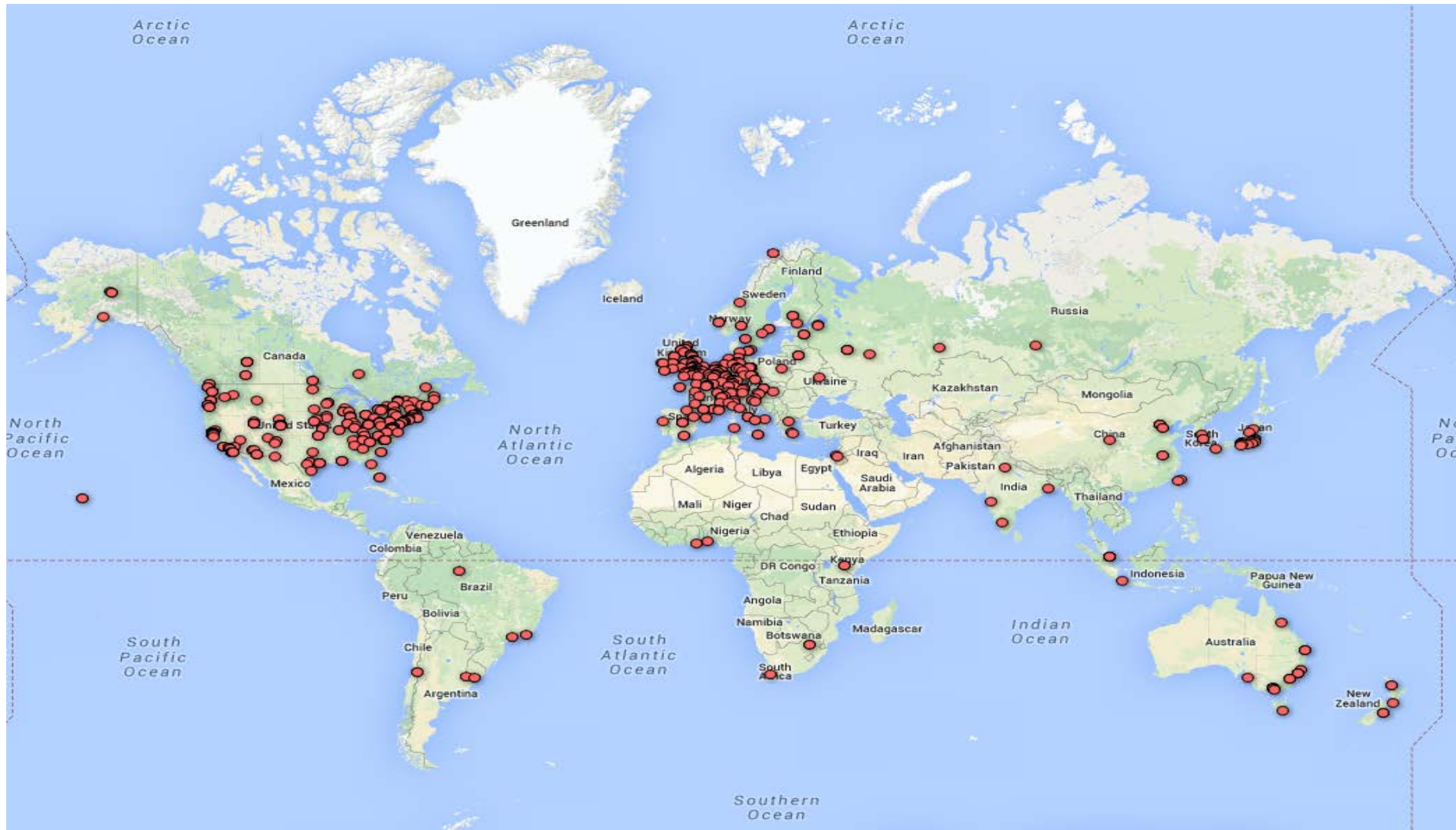
- Enable the discovery of data repositories, data studies and data sets in the context of traditional literature
- Link data to research publications
- Help researchers find data sets and studies and track the full impact of their research output
- Provide expanded measurement of researcher and institutional research output and assessment
- Facilitate more accurate and comprehensive bibliometric analyses



Launched October 2012
4M data records

DATA REPOSITORIES

- Over 1100 repositories identified



TYPES OF DATA BY DISCIPLINE

ART & HUMANITIES

CULTURAL
HERITAGE

LANGUAGE CORPUS

IMAGE
COLLECTIONS

RECORDINGS

SOCIAL SCIENCES

POLL DATA

ECONOMIC
STATISTICS

LONGITUDINAL DATA

NATIONAL CENSUS

PUBLIC OPINION
SURVEYS

SCIENCE & TECHNOLOGY

MAPS

ALGORITHMS

GENOMICS

SKY SURVEYS

ASTROPHYSICS

REMOTE SENSING

MUSEUM SPECIMENS

METADATA PROCESSING

Repository provides metadata feed

- Collaboration on metadata handling

Normalisation and enhancement of metadata

- Controlled vocabularies
- Indexing

Loading to DCI as data object records

- Citations from repository
- Citations from literature

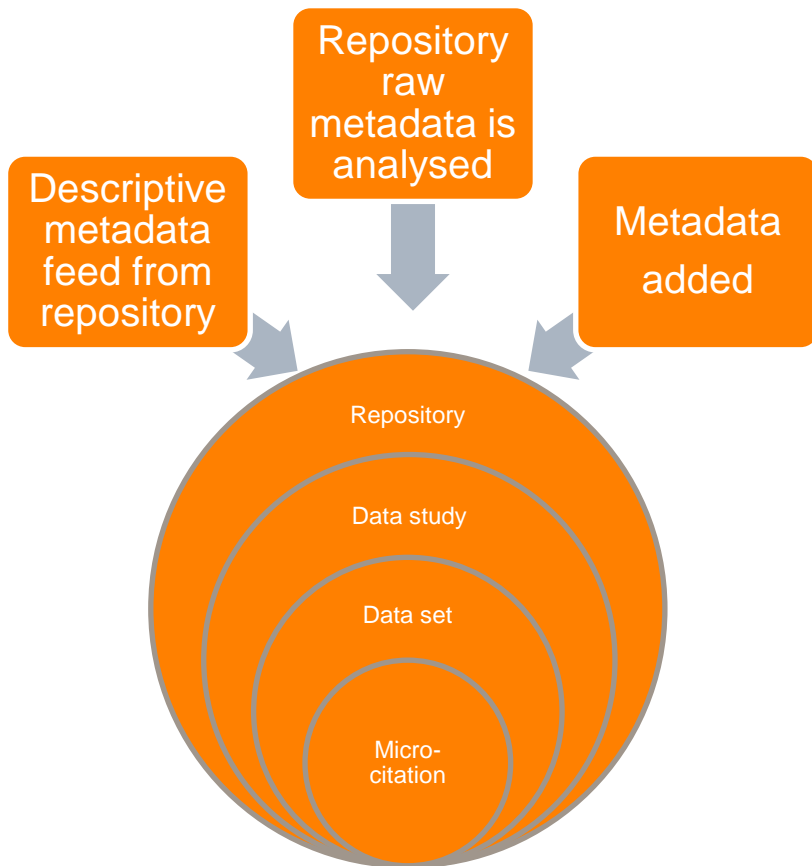
Metrics

- Citation counts

INDEXING A DATA REPOSITORY ON WEB OF SCIENCE

Record Types

- **Repository/Source:** Comprises data studies, data sets and/or microcitations. Stores and provides access to the raw data.
- **Data Study:** Descriptions of studies or experiments with associated data which have been used in the data study. Includes serial or longitudinal studies over time.
- **Data Set:** A single or coherent set of data or a data file provided by the repository, as part of a collection, data study or experiment.
- **Microcitation:** (nanopublication) An assertion about concepts that have been found to be linked by scientific enquiry, and can be uniquely identified and attributed to its author. Made up of three separate parts: a subject, a predicate and an object.



Search Results within the Data Citation Index present the powerful Web of Science options for exploring a body of information. Data becomes discoverable alongside literature



Results: 11,658

You searched for:
TOPIC: (child* health*) ...More

Create Alert

Refine Results

Search within results for...

Web of Science Categories

Subject Areas

Document Types

- DATA SET (8,776)
- DATA STUDY (2,881)
- REPOSITORY (1)

more options / values...

Refine

Authors

Group Authors

Editors

Source Titles

Data Types

- APPLICATION MSWORD (683)
- RNA (538)
- SURVEY DATA (524)
- NUMERIC DATA INDIVIDUAL MICRO LEVEL (309)
- GENOMIC (216)

more options / values...

For advanced options, use

Analyze Results

Sort by: Year Published -- newest to oldest

Page 1 of 1,166

Select Page



Save to EndNote online

Add to Marked List

- Countdown to 2015 for Maternal, Newborn and Child Survival and the Commission on Information and Accountability for Women's and Children's Health.** Times Cited: 1 (from All Databases)
By: Newby, Holly; Requejo, Jennifer Harris; Bryce, Jennifer Figshare
DOI: <http://dx.doi.org/10.1371/journal.pmed.1001416.t001>
Document Type: Data set
- Coverage Indicators for Global Monitoring of RMNCH: Millennium Development Goal Framework, Countdown to 2015 for Maternal, Newborn and Child Survival, and the Commission for Information and Accountability for Women's and Children's Health, 2012.** Times Cited: 1 (from All Databases)
By: Newby, Holly; Requejo, Jennifer Harris; Bryce, Jennifer Figshare
DOI: <http://dx.doi.org/10.1371/journal.pmed.1001416.t002>
Document Type: Data set
- Prevalence of Childhood Obesity, Borough.** Times Cited: 0 (from All Databases)
Group Author(s): Information Centre for Health and Social Care
London Datastore
Source URL: <http://data.london.gov.uk/datastore/package/prevalence-childhood-obesity-borough>
Document Type: Data set
[View Abstract](#)
- Cebu Longitudinal Health and Nutrition Survey GWAS.** Times Cited: 0 (from All Databases)
Group Author(s): NIDDK
European Nucleotide Archive
Source URL: <http://www.ebi.ac.uk/ena/data/view/PRJNA171954>
Document Type: Data study
[View Abstract](#)
- Sequencing of exons to locate the mutation in dearisch mice.** Times Cited: 0 (from All Databases)
Group Author(s): The Wellcome Trust Sanger Institute
European Nucleotide Archive
Source URL: <http://www.ebi.ac.uk/ena/data/view/PRJEB2587>
Document Type: Data study
[View Abstract](#)
- Human gut microbiome viewed across age and geography.** Times Cited: 0 (from All Databases)
Group Author(s): Colorado Center for Microbial Ecology, University of Colorado at Boulder
European Nucleotide Archive
Source URL: <http://www.ebi.ac.uk/ena/data/view/PRJEB3079>
Document Type: Data study
[View Abstract](#)
- Gut microbiomes of Malawian twin pairs discordant for kwashiorkor.** Times Cited: 0 (from All Databases)
Group Author(s): Colorado Center for Microbial Ecology, University of Colorado at Boulder
European Nucleotide Archive

Back to Search

My Tools Search History Marked List

Save to EndNote online Add to Marked List Back to List 1 of 1

Roper Center for Public Opinion Research

Edited by: Roper Center

Roper Center for Public Opinion Research

Source URL: <http://www.ropercenter.uconn.edu>

Viewed Date: 07 Dec 2011

Published: 1947

Abstract

Founded in 1947, the Roper Center for Public Opinion Research is the most intelligent, responsible and imaginative group of national research. Through the maintenance of the Roper Center strives to improve the practice of public opinion research in Connecticut. All Rights Reserved

Author Information

Addresses:

1. Roper Center, University of Connecticut, 369 Fairfield Way, Unit 2164, Storrs, CT 06269-2164, USA

Categories / Classification

Research Areas: Social Sciences - Other Topics

Web of Science Category: Social Sciences, Interdisciplinary

Document Information

Document Type: Repository

Language: English

Accession Number: DRCI:DATA2012001000656741

Other Information

Cited References in Data Citation Index: 0

Associated Data: 12,064 (from Data Citation Index)

Roper Social and Political Trends Data, 1973-1994.	Data study	Link to External Source
NORC Survey:.	Data study	Link to External Source
General Electric Quarterly Survey 70-01 (NO CODEBOOK).	Data study	Link to External Source
Demographic Survey of the Jewish Community of Richmond.	Data study	Link to External Source

[View All Associated Data](#)

WEB OF SCIENCE™

Back to Search

My Tools Search History Marked List

Associated Records: 12,064

Sort by: Year Published -- newest to oldest

Page 1 of 1,207

You searched for: Roper Center, Roper Center for Public Opinion Research. ...More

Create Alert

Refine Results

Search within results for...

Web of Science Categories

Select Page Save to EndNote online Add to Marked List

- Roper Social and Political Trends Data, 1973-1994.**
Group Author(s): The Roper Organization
Roper Center for Public Opinion Research
Source URL: http://www.ropercenter.uconn.edu/cgi-bin/hsrun.exe/roperweb/Catalog40/Catalog40.htm?start=summary_Link?Archno=USRoper1994-Trends
Document Type: Data study
[View Abstract](#)

Times Cited: 0 (from All Databases)

- NORC Survey:.**
Group Author(s): National Opinion Research Center [NORC]
Roper Center for Public Opinion Research
Source URL: http://www.ropercenter.uconn.edu/cgi-bin/hsrun.exe/roperweb/Catalog40/Catalog40.htm?start=summary_Link?Archno=USNORC1950-0274
Document Type: Data study

Times Cited: 0 (from All Databases)

0 in ScELO Citation Index

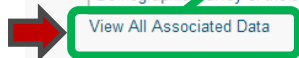
This record is from: Data Citation Index SM

[How to cite this Resource](#)

Suggest a correction

If you would like to improve the quality of the data in this record, please suggest a correction.

Data deposition makes it possible to show related data from the repository



Associated Records: 6

For: Effectiveness of influenza vaccination in working-age adults with diabetes: a population-based cohort study ...More

1. Hospital Abstracts; Hospital Discharge Abstracts; Hospital Separation Abstracts	Times Cited: 92 (from All Databases)
2. Manitoba Health Insurance Registry	Times Cited: 73 (from All Databases)
3. Canada Census	Times Cited: 62 (from All Databases)
4. Medical Services; Medical Claims	Times Cited: 78 (from All Databases)

based

(Johnson,

adults (<65
vaccination

in 2000 to
hza-like
analyses were

Citation Network

0 Times Cited
35 Cited References
View Related Records
View Citation Map
Create Citation Alert

(data from Web of Science™ Core Collection)

All Times Cited Counts

0 in All Databases
0 in Web of Science Core Collection
0 in BIOSIS Citation Index
0 in Chinese Science Citation Database
0 in Data Citation Index
0 in SciELO Citation Index

Associated Data: 6

View the data associated with this record (from Data Citation Index™)

Results We included 543 367 person-years of follow-up, during which 223 920 ILI, 5422 PI and 94 988 ALL occurred. The majority (58%) of adults with diabetes were working age. In this group, influenza vaccination was associated with relative reductions in PI (43%, 95% CI 28% to 54%) and ALL (28%, 95% CI 24% to 32%) but not ILI (-1%, 95% CI -3% to 1%). VE was similar in elderly adults for ALL (33-34%) and PI (45-55%), although not ILI (12-13%). However, similar estimates of effectiveness were also observed for all three groups during non-influenza control periods.

Other Information

IDS Number: 164VH
Cited References in Web of Science Core Collection: 35
Times Cited in Web of Science Core Collection: 0

Associated Data: 6 (from Data Citation Index)

Hospital Abstracts; Hospital Discharge Abstracts; Hospital Separation Abstracts	Data study	Link to External Source
Manitoba Health Insurance Registry	Data study	Link to External Source
Canada Census	Data study	Link to External Source
Medical Services; Medical Claims	Data study	Link to External Source

View All Associated Data

Because data are accessible and able to be cited, they can be linked to publications describing research which uses them



Save to EndNote online

Add to Marked List

Back to List < 1 of 1 >

Roper Center

Edited by: Roper

Roper Center for... Source URL: htt... Viewed Date: 0... Published: 1947

Abstract

Founded in 1947, the intelligent, res... increasingly com... national research... Roper Center stri... Connecticut. All F

Author Inform

Addresses:

1. Roper Center,

Categories / C

Research Areas: Social Sciences - Other Topics

Web of Science Category: Social Sciences, Interdisciplinary

Document Information

Document Type: Repository

Language: English

Accession Number: DRCI:DATA2012001000656741

Other Information

Cited References in Data Citation Index: 0

Associated Data: 12,064 (from Data Citation Index)

Roper Social and Political Trends Data, 1973-1994.	Data study	Link to External Source
NORC Survey.	Data study	Link to External Source
General Electric Quarterly Survey 70-01 (NO CODEBOOK).	Data study	Link to External Source
Demographic Survey of the Jewish Community of Richmond.	Data study	Link to External Source

View All Associated Data



Welcome

RECOMMEND TO LIBRARY Sign In | iPOLL | Datasets | Support

NEW SEARCH

Dataset Abstract

DOWNLOAD DOCUMENTATION

Survey questionnaires and documentation are available for free download.

- Word file (157kb)
- PDF file (127kb)

DOWNLOAD DATASET

Datasets are available for this survey.

If you are a member with a personalized account, please Sign In, otherwise contact Data Services at 860.486.4440 or rcweb@ropercenter.uconn.edu.

Not yet a member of the Roper Center?

Become a member of the world's largest and oldest opinion archives.

Title:

Roper Social and Political Trends Data, 1973-1994 [USRoper1994-Trends]

Survey Firm:

The Roper Organization

Study Date:

1973-1994

Sample:

National Adult

Sample Size:

410116

Survey Notes:

This data set is a compilation of several surveys. Please see the codebook for details.

Variables:

Abstract:

Roper Social and Political Trends Data, 1973-1994 from RoperASW serves as one of three primary datasets used extensively in the research reported in Bowling Alone: Collapse and Revival of American Community by Robert D. Putnam, Harvard University. In an effort to explore the current state of social capital in the United States, Putnam has tapped the rich collection of time-series data gathered by RoperASW (now know as gfkamerica) from the fall of 1973 through 1994. This unique resource has now been contributed to the Roper Center at the University of Connecticut in an effort to further the study of social capital and scientific social inquiry. This dataset was compiled and cleaned through the efforts of a joint research team, headed by Professor Henry Brady of the University of California-Berkeley and Professor Putnam. The dataset includes selected items from 207 public opinion surveys conducted as part of the Roper

How to cite this Resource

Link out directly to the original item, in this case a Data Study.

[Back to Search](#)[My Tools ▾](#)[Search History](#)[Marked List](#)[Save to EndNote online](#)[Add to Marked List](#)[Back to List](#) ◀ 2 of 2 ▶

UniProt Knowledgebase.

Web of Science™ InCites® Journal Citation Reports® Essential Science Indicators™ EndNote® Sign In ▾ Help English ▾

WEB OF SCIENCE™ THOMSON REUTERS™

[Back to Search](#) [My Tools ▾](#) [Search History](#) [Marked List](#)

[Full Text](#) [Look up full-text](#) [Save to EndNote online](#) [Add to Marked List](#) [Back to List](#) ◀ 1 of 11 ▶

From protein sequences to 3D-structures and beyond: the example of the UniProt Knowledgebase

By: Hinz, U (Hinz, Ursula)^[1]
 Group Author(s): UniProt Consortium

CELLULAR AND MOLECULAR LIFE SCIENCES
 Volume: 67 Issue: 7 Pages: 1049-1064
 DOI: 10.1007/s00018-009-0229-6
 Published: APR 2010
[View Journal Information](#)

Abstract

With the dramatic increase in the volume of experimental results in every domain of life sciences, assembling pertinent data and combining information from different fields has become a challenge. Information is dispersed over numerous specialized databases and is presented in many different formats. Rapid access to experiment-based information about well-characterized proteins helps predict the function of uncharacterized proteins identified by large-scale sequencing. In this context, universal knowledgebases play essential roles in providing access to data from complementary types of experiments and serving as hubs with cross-references to many specialized databases. This review outlines how the value of experimental data is optimized by combining high-quality protein sequences with complementary experimental results, including information derived from protein 3D-structures, using as an example the UniProt knowledgebase (UniProtKB) and the tools and links provided on its website (<http://www.uniprot.org/>). It also evokes precautions that are necessary for successful predictions and extrapolations.

Keywords

Author Keywords: Data flood; Annotation; Swiss-Prot; Knowledgebase; UniProtKB; Proteomics; Structural genomics; Protein 3D-structure
KeyWords Plus: DEPENDENT K+ CHANNEL; X-RAY-STRUCTURE; STRUCTURAL GENOMICS; CRYSTAL-STRUCTURE; UNIPROTKB/SWISS-PROT; DISORDERED PROTEINS; INTERACTOME NETWORK; ATOMIC STRUCTURES; ESCHERICHIA-COLI; DATA-BANK

Author Information

Reprint Address: Hinz, U (reprint author)
 ✉ Swiss Inst Bioinformat, Swiss Prot Grp, 1 Rue Michel Servet, CH-1211 Geneva, Switzerland.

Addresses:
 ✉ [1] Swiss Inst Bioinformat, Swiss Prot Grp, CH-1211 Geneva, Switzerland
 [2] Swiss Inst Bioinformat, Vital IT Grp, CH-1015 Lausanne, Switzerland

Citation Network

11 Times Cited

0 Cited References

[Create Citation Alert](#)*(data from Web of Science™ Core Collection)*

All Times Cited Counts

11 in All Databases

11 in Web of Science Core Collection

10 in BIOSIS Citation Index

0 in Chinese Science Citation Database

0 in Data Citation Index

Start to build citation maps associated with data through the association of data and literature

[View All](#)This record is from:
Data Citation Index™

Citation Network

9 Times Cited

106 Cited References

[View Related Records](#) [View Citation Map](#) [Create Citation Alert](#)*(data from Web of Science™ Core Collection)*

All Times Cited Counts

9 in All Databases

9 in Web of Science Core Collection

8 in BIOSIS Citation Index

0 in Chinese Science Citation Database

0 in Data Citation Index

0 in SciELO Citation Index

Most Recent Citation

Maraite, Andy. Lipase from *Pseudomonas stutzeri*: Purification, homology modelling and rational explanation of the substrate binding mode. JOURNAL OF MOLECULAR CATALYSIS B-ENZYMATIC, MAR 2013.[View All](#)

Associated Data: 1

[View the data associated with this record \(from Data Citation Index™\)](#)

[Back to Search](#)[My Tools](#)[Search History](#)[Marked List](#)[Save to EndNote online](#)[Add to Marked List](#)[Back to List](#) ◀ 1 of 1 ▶

GSE12195: Mutations of multiple genes deregulate the NF-kB pathway in diffuse large B cell lymphoma.

From Repository: [Gene Expression Omnibus](#)
By: Pasqualucci, Laura; Basso, Katia

Gene Expression Omnibus

Source URL: <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE12195>

Viewed Date: 12 Dec 2011

Published: 2009

Thomson Reuters recommends citing this resource as:
Pasqualucci, Laura; Basso, Katia (2009): GSE12195: Mutations of multiple genes deregulate the NF-kB pathway in diffuse large B cell lymphoma. Gene Expression Omnibus.
<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE12195>

Abstract

Diffuse large B-cell lymphoma (DLBCL), the most common form of lymphoma in adulthood, comprises multiple biologically and clinically distinct subtypes including germinal center B cell-like (GCB) and activated B cell like (ABC) DLBCL. Gene expression profile studies have shown that its most aggressive subtype, ABC-DLBCL, is associated with constitutive activation of the NF-kB transcription complex. However, except for a small fraction of cases, it remains unclear whether NF-kB activation in these tumors represents an intrinsic program of the tumor cell of origin or a pathogenetic event. Here we show that >50% of ABC-DLBCL and a smaller fraction of GCB-DLBCL carry somatic mutations at multiple genes, including negative (TNF, NIP3/A20) and positive (CARD11, TRAF2, TRAF5, MAP3K7/TAK1 and TNFRSF11A/RANK) regulators of NF-kB. Of these, the A20 gene, which encodes for a ubiquitin-modifying enzyme involved in termination of NF-kB responses, is the most commonly affected one, with ~30% of the patients displaying biallelic inactivation by mutations and/or deletions, suggesting a tumor suppressor role. Less frequently, missense mutations of TRAF2 and CARD11 produce molecules with significantly enhanced ability to activate NF-kB. Thus, our results demonstrate that NF-kB activation in DLBCL is caused by genetic lesions affecting multiple genes, whose loss or activation may promote lymphomagenesis by leading to abnormally prolonged NF-kB responses. We show that most ABC-DLBCL and a smaller fraction of GCB-DLBCL display genetic lesions affecting multiple NFkB pathway genes, with A20 representing the most frequently mutated gene.

Keywords

Author Keywords: Phenotypic characterization of human DLBCL

Author Information

Addresses:

1. Columbia University, Institute for Cancer Genetics, 1130 St Nicholas Ave, New York, 10032, USA

2. Columbia University, Institute for Cancer Genetics, 1150 St. Nicholas Avenue, New York, NY 10032, USA

E-mail Addresses: lp171@columbia.edu; kb451@columbia.edu

Categories / Classification

Research Areas: Biochemistry & Molecular Biology; Genetics & Heredity

Web of Science Category: Biochemistry & Molecular Biology; Genetics & Heredity

Taxonomic Data:

Citation Network

4 Times Cited

0 Cited References

[Create Citation Alert](#)

(data from Web of Science™ Core Collection)

All Times Cited Counts

4 in All Databases

4 in Web of Science Core Collection

4 in BIOSIS Citation Index

0 in Chinese Science Citation Database

0 in Data Citation Index

0 in SciELO Citation Index

Most Recent Citation

Pasqualucci, Laura. Inactivating mutations of acetyltransferase genes in B-cell lymphoma. NATURE, MAR 10 2011.

[View All](#)

This record is from:
Data Citation Index™

[How to cite this Resource](#)

Suggest a correction

If you would like to improve the quality

Provide assistance in how to associate data and literature through citation

RESEARCHER PROBLEMS

- Access & discovery
- Citation standards
- Lack of willingness to deposit and cite
- Lack of recognition / credit



Data sharing leads to more science & more knowledge



DEFINITIONS

Data repository

- An online resource where data are deposited and stored for preservation and access

Data

- Facts collected for reference or analysis.
- Non traditional scholarly output of scientific research often analysed in traditional research publications. May include numerical, textual, image, video or software information

REPOSITORY SELECTION & EVALUATION



REUTERS/ Mohan Raza

REPOSITORY EVALUATION, SELECTION, AND COVERAGE POLICIES

FOR THE DATA CITATION INDEX™ WITHIN THOMSON REUTERS WEB OF KNOWLEDGE™

As we evaluate repositories for inclusion, some of the things we consider are:

- Editorial Content - ensuring that material is desirable to the research community.
- Persistence and stability of the repository, with a steady flow of new information.
- Thoroughness and detail of descriptive information.
- Links from data to research literature.



THOMSON REUTERS

DATA REPOSITORIES

Data deposit

- Repository must hold “data”
- Repository must provide access to data

Active

- Material added/updated
- Provide statistics on deposited data
- Actively curate data in the archive

Persistence

- Persistent IDs, DOIs or other permanent ID
- Contacts available for confirmation of interpretation
- Indication of intention to preserve data or provide access over the long term
 - Contingency if repository was to cease to operate
- Make data accessible (or state licensing terms)
- Sustainable
 - Funding information available for repository and deposited data

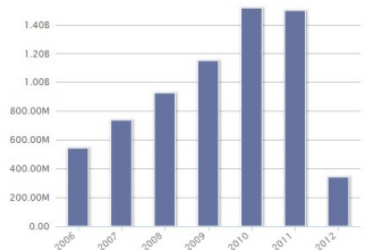
Data reuse

- Links to literature
- Citation in literature databases



CHALLENGES

- Metadata
 - Resources
 - Expertise
- Citable data source
- Metadata quality
 - Unique & persistent identifiers
 - Consistency
- Data repositories are not static
 - How is version control handled?
- Partnerships



```

:GAGCATTGTGCACAATCGTGGTTATAGCGGCCGGTGGCGCCGTC
:GTACTACGAAAGCTTTGGCGTCCAGACGCCAGCTGGCTTACTG
:GAGGTGCCCTTCTTAATAAGGCTATCGGGGAGGCGCCCGAGGA
:ACCCGGTGCATCTACTCAGAGTCTTACCACCCTACAACCTCTG
:TTCCATCGGGCAACTCCAACCCCGCTGTGTGCACGACGTCA
:GACAATCTGGTAGCCATGGGCTTACAGTCTCCACAGGGAGGG
    
```

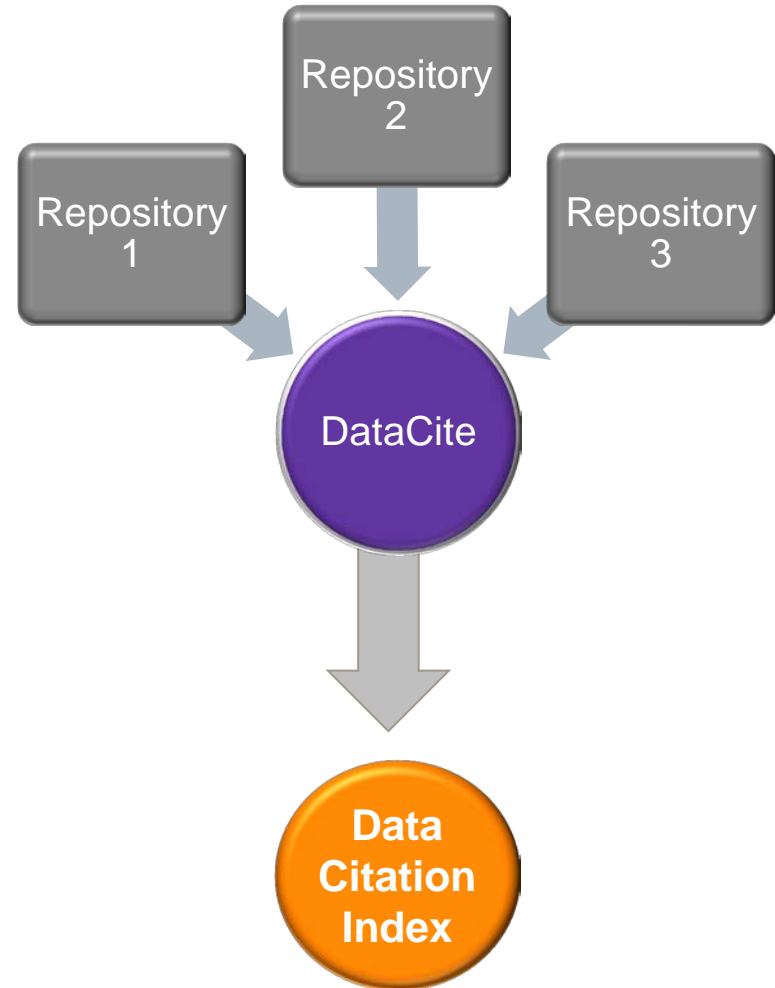
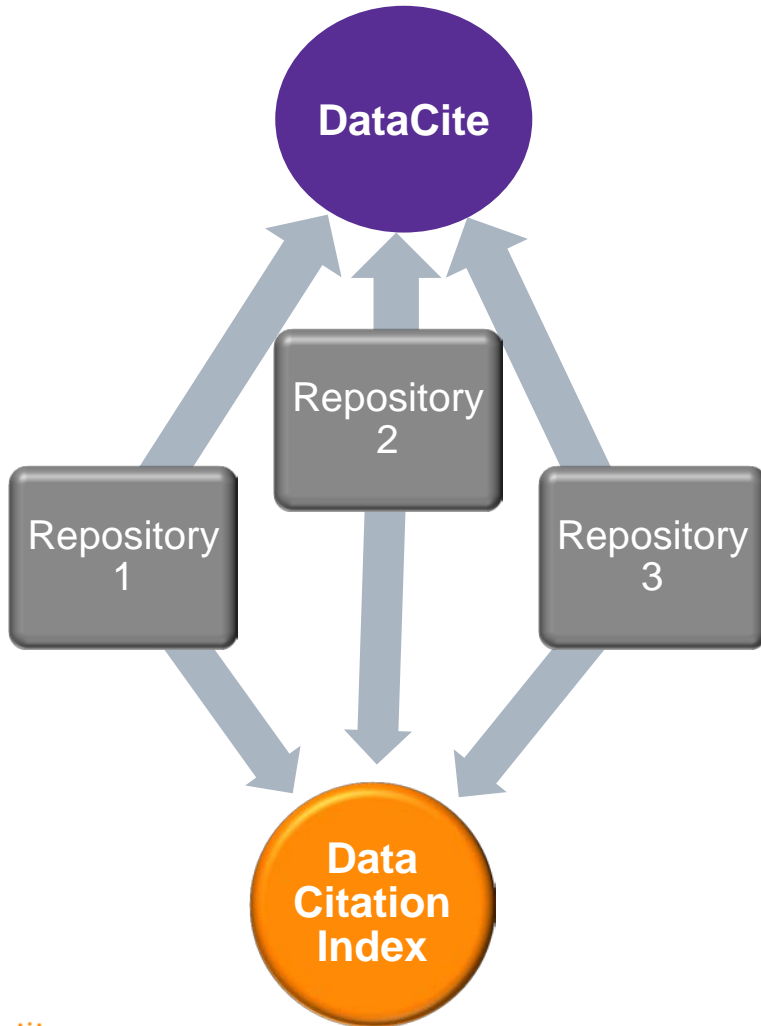
Partner	2007	2008	2009	2010	2011	Avg Ann growth
World	\$1,162,708.3	\$1,300,135.6	\$1,656,816.3	\$1,277,388.4	\$1,479,492.0	6.8
Europe and Eurasia	\$286,484.8	\$328,286.9	\$258,646.0	\$286,023.6	\$328,873.4	3.7
European Union	\$246,356.9	\$272,951.8	\$219,880.2	\$238,913.9	\$267,463.7	2.1
–Austria	\$3,172.0	\$2,649.3	\$2,538.5	\$2,427.6	\$3,886.6	-2.2
–Belgium	\$25,291.9	\$29,026.4	\$21,629.7	\$25,551.1	\$29,876.9	4.5
–Cyprus	\$169.0	\$217.4	\$160.8	\$134.3	\$0.1	-10.6
–Czech Republic	\$1,262.4	\$1,378.3	\$969.7	\$1,415.6	\$1,680.6	8.3
–Denmark	\$2,927.2	\$2,712.1	\$2,058.5	\$2,125.3	\$2,244.7	-5.8
–Estonia	\$242.3	\$225.6	\$183.2	\$188.0	\$311.0	10.2
–Finland	\$3,133.2	\$3,761.6	\$1,665.6	\$2,181.2	\$3,159.0	0.2
–France	\$27,407.1	\$29,186.9	\$36,523.3	\$27,010.1	\$27,844.2	0.4
–Germany	\$49,652.0	\$54,732.3	\$43,296.6	\$48,201.2	\$49,134.2	-0.3
–Greece	\$2,111.0	\$1,931.8	\$2,475.7	\$1,106.8	\$1,083.4	-12.2
–Hungary	\$1,291.7	\$1,431.2	\$1,231.7	\$1,290.2	\$1,473.2	3.5
–Ireland	\$9,010.7	\$8,652.9	\$7,516.4	\$7,272.0	\$7,607.7	-3.9
–Italy	\$14,141.3	\$15,478.6	\$12,252.6	\$14,191.4	\$15,991.3	3.3

COLLABORATION BETWEEN DATACITE & THOMSON REUTERS



- Increasing visibility of DOI
- Synergies
- Support for data citation principles

DATA CITATION INDEX PARTNERSHIPS

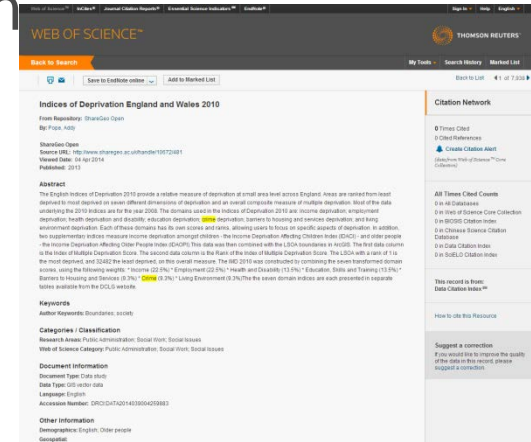


REQUIRED METADATA

- Unique ID in repository
 - Date provided
 - Author
 - Repository
 - URL/DOI
 - Title
 - Year Published
- Allows creation of a data citation using DataCite guidelines
 - Compliance with DataCite Metadata schema v3
 - Allows matching of data citations encountered to known data records

PARTNERSHIP BENEFITS

- Access to DCI to review implementation



- Badge for website



- API to enable citation counts

DATAcite PARTNER REPOSITORIES

- 68 repositories eligible for evaluation, including:
 - Archaeology Data Service
 - Chemotion
 - Collaborative Research in Computational Neuroscience (CRCNS)
 - eyeMoviePedia
 - FLOSSmole
 - German Center for Gerontology
 - GigaDB
 - MatDB
 - Movebank Network for Earthquake Engineering Simulation (NEES)
 - Swedish National Data Service
 - UNAVCO
 - University of Southampton
 - World Data Centre For Climate
 - Zenodo



REASONS FOR NON SELECTION

- Not meeting selection criteria
 - Not “data”
 - No data type
- Poor quality or inconsistent metadata
- Defective DOIs
- More complete metadata from elsewhere
 - Crossover with other aggregation services
 - Australian National Data Service
 - Repository

DATA CITATION TRACKING



- Infrastructure in place



- Formal citations



- Data citation matching process



- Capture of informal citations

DATA CITATION

Current citation style (in full text of article as informal citations)

III. Data Description

Our first data set is the Bureau of Justice Statistics "Murder Cases in 33 Large Urban Counties." This is a random sample of homicide cases drawn from prosecutors' files. The data set includes information on offender characteristics, victim characteristics, and trial outcomes for 2800 murders. The 75 largest counties account for more than half of the murders in the U.S. each year. This data set brings together information on the crime, the offender, the victim, and the sentence. Such information cannot all be linked in other larger data sets such as the Uniform Crime Reporting (UCR) Data or the National Crime Victimization Survey (NCVS). Most crime



Desired/future citation style (as formally cited references)

U.S. Dept. of Justice, Bureau of Justice Statistics (1996): MURDER CASES IN 33 LARGE URBAN COUNTIES IN THE UNITED STATES, 1988. Version 1. Inter-university Consortium for Political and Social Research.
<http://dx.doi.org/10.3886/ICPSR09907.v1>

Supplementary Material

Supplemental Data
[Click here to view.](#)

Acknowledgments

We thank Junghwa Seo, Lei Cho, and Jongmin Kim for technical assistance and Hyunjung Lim for consultation on image handling.

This work was supported, in whole or in part, by National Institutes of Health Grants AG5131 and AG18440. This work was also supported by the Disease Network Research Program (Grant 20090084180) from the National Research Foundation of Korea funded by the Ministry of Education, Science and Technology, Republic of Korea and by the Korea Science and Engineering Foundation funded by the Korea government (Grant 20090083737).

The microarray data reported in this paper have been deposited to the Gene Expression Omnibus (GEO) data base under accession number [GSE11574](#).

S The on-line version of this article (available at <http://www.jco.org>) contains supplemental [Figs. S1-S8](#) and [Tables 1-4](#).

²A. Jang, H.-J. Lee, J.-E. Suk, J.-W. Jung, K.-P. Kim, and S.-J. Lee, submitted for publication.



Lee, Seung-Jae; Lee, He-Jin; Cho, Ji-Hoon; Rho, Sangchul; Hwang, Daehee (2008): GSE11574: The responses of astrocytes stimulated by extracellular a-synuclein. Gene Expression Omnibus.
<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE11574>

DATA CITATION

Our first data set is the Bureau of Justice Statistics "Murder Cases in 33 Large Urban Counties."

This is a random sample of homicide cases drawn from prosecutors' files. The data set includes information on offender characteristics, victim characteristics and trial outcomes for 2800 murders. The 75 largest counties account for more than half of the murders in the U.S. each year. This data set brings together information on the crime, the offender, the victim, and the sentence. Such information cannot all be linked in other larger data sets such as the Uniform Crime Reporting (UCR) Data or the National Crime Victimization Survey (NCVS). Most crime

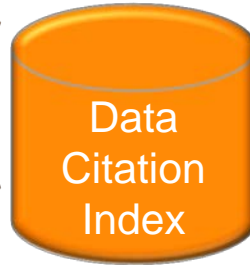
Lee, Seung-Jae; Lee, He-Jin; Cho, Ji-Hoon; Rho, Sangchul; Hwang, Daehee (2008): GSE11574: The responses of astrocytes stimulated by extracellular a-synuclein. Gene Expression Omnibus. <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE11574>

Factor	2007	2008	2009	2010	2011	Avg Ann growth
World	\$2,145,788.3	\$2,306,122.8	\$2,478,838.0	\$2,771,388.4	\$3,176,480.0	8.1
Europe and Eastern	\$206,484.0	\$228,386.0	\$253,848.0	\$286,023.0	\$328,714.0	8.7
European Union	\$206,484.0	\$228,386.0	\$253,848.0	\$286,023.0	\$328,714.0	8.7
Australia	\$17,720.0	\$18,680.0	\$19,580.0	\$20,420.0	\$21,280.0	2.2
Belgium	\$27,200.0	\$28,000.0	\$28,800.0	\$29,600.0	\$30,400.0	4.7
Cyprus	\$100.0	\$100.0	\$100.0	\$100.0	\$100.0	0.0
Czech Republic	\$1,200.0	\$1,200.0	\$1,200.0	\$1,200.0	\$1,200.0	0.0
Denmark	\$2,000.0	\$2,000.0	\$2,000.0	\$2,000.0	\$2,000.0	0.0
France	\$100.0	\$100.0	\$100.0	\$100.0	\$100.0	0.0
Germany	\$100.0	\$100.0	\$100.0	\$100.0	\$100.0	0.0
Italy	\$100.0	\$100.0	\$100.0	\$100.0	\$100.0	0.0
Japan	\$100.0	\$100.0	\$100.0	\$100.0	\$100.0	0.0
South Korea	\$100.0	\$100.0	\$100.0	\$100.0	\$100.0	0.0
Spain	\$100.0	\$100.0	\$100.0	\$100.0	\$100.0	0.0
Sweden	\$100.0	\$100.0	\$100.0	\$100.0	\$100.0	0.0
Switzerland	\$100.0	\$100.0	\$100.0	\$100.0	\$100.0	0.0
United Kingdom	\$100.0	\$100.0	\$100.0	\$100.0	\$100.0	0.0
USA	\$1,000,000.0	\$1,000,000.0	\$1,000,000.0	\$1,000,000.0	\$1,000,000.0	0.0

Published data sets



Scientific literature



New data metrics

DATA CITATION INDEX

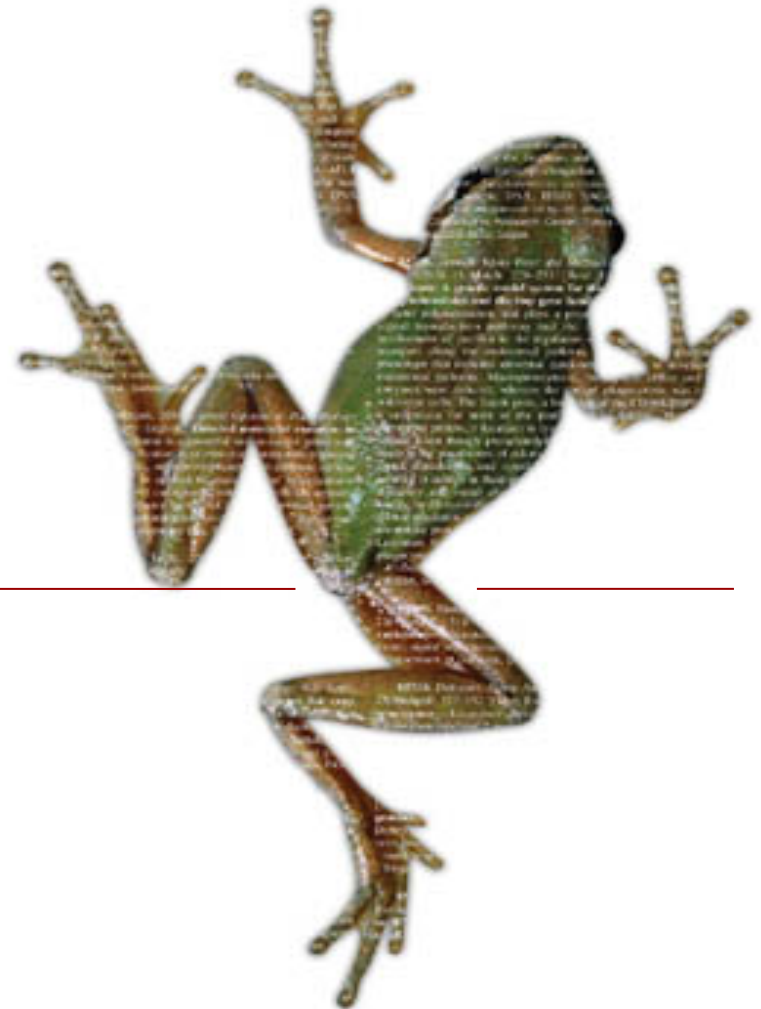
- Discovery of data most important to scholarly research
- Data linked to published research literature
- Measures of data citation, use and reuse with attribution assisted by identifiers
- New metrics for digital scholarship



THANK YOU

Nigel Robinson

nigel.robinson@thomsonreuters.com

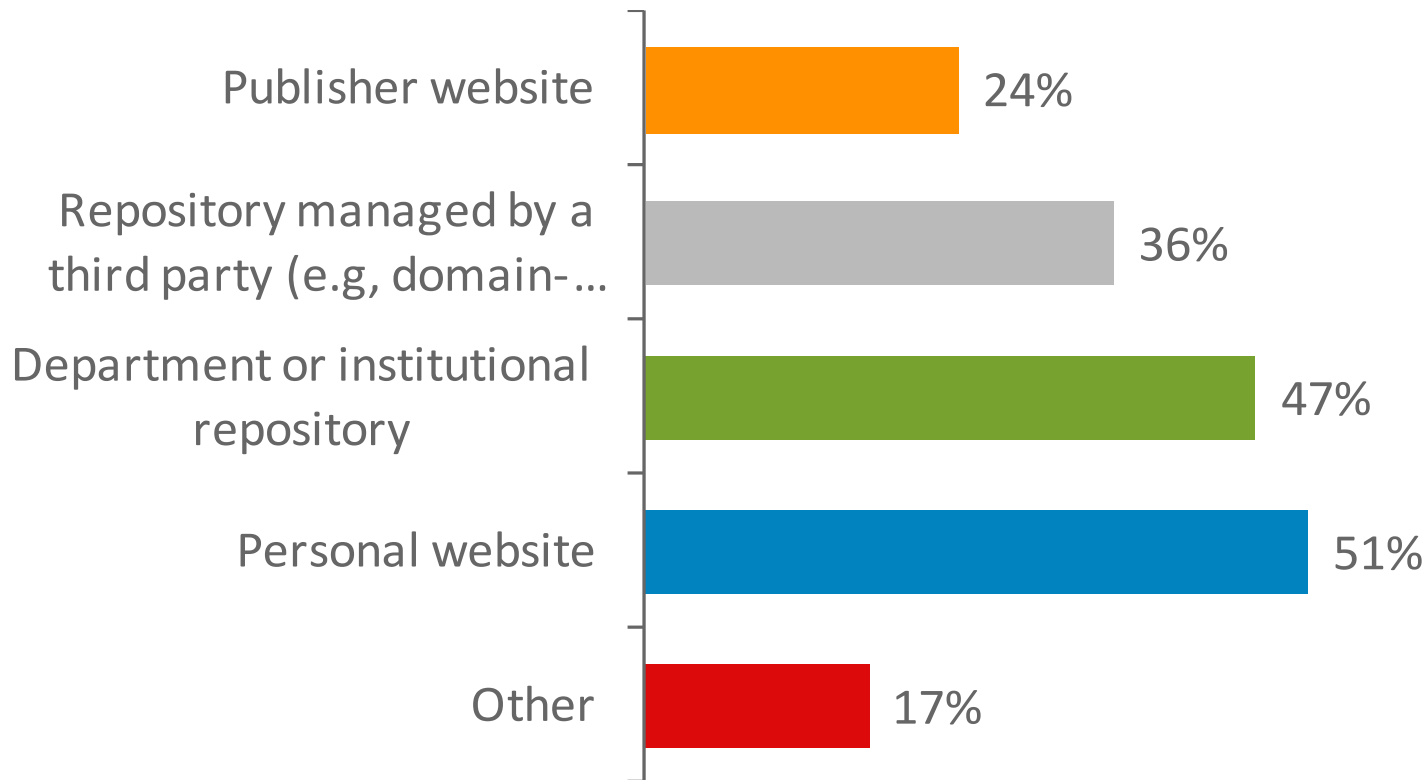




ADDITIONAL SLIDES



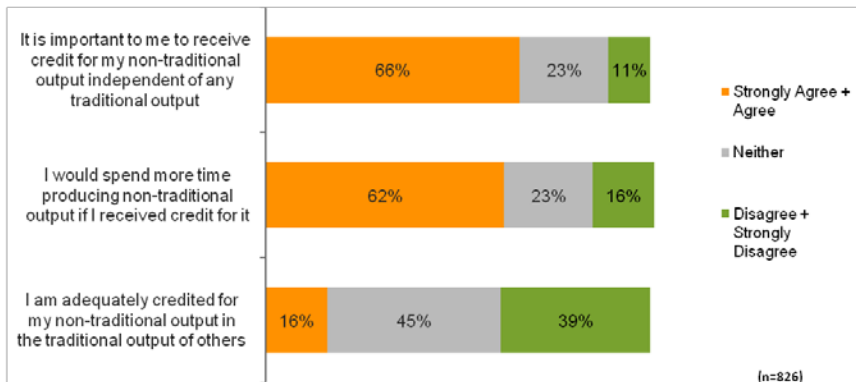
DEPOSITION OF DATA BY RESEARCHERS



RESEARCHERS NOT RECEIVING CREDIT

► Researchers are not receiving adequate credit for digital scholarship

- Researchers would spend more time producing output if they received credit
- Few are currently receiving adequate credit



► Institutional policies and the lack of standards for sharing digital output is an impediment both to research and sharing

“Lack of knowledge about standards for citation and of proper scholarly recognition and/or evaluation of such materials.” (Researchers, Canada)

Barriers to creating and sharing data:

- Researchers are hesitant to spend time and effort to create and share data because they don't feel the work is **adequately exposed** or **accredited**
- Researchers find it difficult to expose data they have produced because **data repositories do not have clear standards or mechanisms** in place for doing so

17