

# SPIRES



National Institutes of Health  
*Office of the Director*

# PROJECT ABSTRACT

## **Abstract Title**

- Scientific Publication Information Retrieval and Evaluation System (SPIRES)

## **Abstract Authors**

- Paul Jordan, FO Finch III, Ying Gao

## **Abstract Affiliations**

- OD, OER, Office of Data Analysis Tools and Systems

## **Abstract Category**

- Tools and Tool Development

## **Abstract Text**

The Scientific Publication Information Retrieval and Evaluation System (SPIRES) is a database that automatically maps NLM PubMed publication records to NIH project numbers on a daily basis. The majority of these publications are produced by extramural investigators, but SPIRES also includes publications resulting from intramural projects via the NIH Intramural Database, and incorporates publications reported through the NIH Manuscript System (NIHMS) and NCBI MyBibliography. The resulting database is used in several other enterprise and extension systems, including QVR, RePORTER, and others. Recently, SPIRES has also added data on the number of times NIH-supported publications have been cited by other articles, and this data is made available to analysts through SPIRES, QVR, and iRePORT.

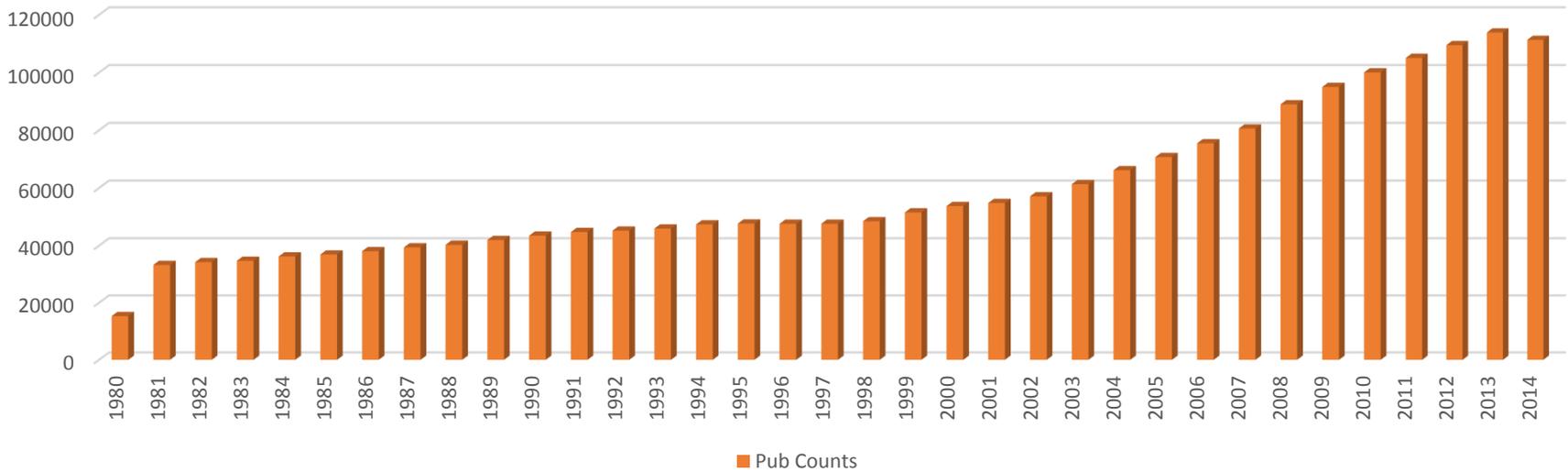
# SPIRES DATA: Where Does It Come From?

- SPIRES project began at NIEHS in 2001, expanded to include all NIH ICs in 2006.
- Every day, SPIRES retrieves all NEW publications from the NLM PubMed database:
  - *Where the publication year is  $\geq 1980$ .*
  - *Where the GRANT tag contains an NIH organization code (2 letter phs\_org\_code or full abbreviation [ie-CA or NCI, HL or NHLBI, etc.]).*
  - *Where the AFFILIATION tag contains references to the NIH or an NIH organization.*
- Extramural and Intramural publications are included.
  - *Intramural publications are updated daily and annually (from NIDB).*
- SPIRES also searches for changes in existing publication records:
  - *If the PubMed record has been revised, the SPIRES record is replaced with the newer version.*
- SPIRES retrieves publication data from the NLM MyBibliography system (MyBib).
  - *Publication/Grant number associations created/edited by PIs/Authors are updated daily.*

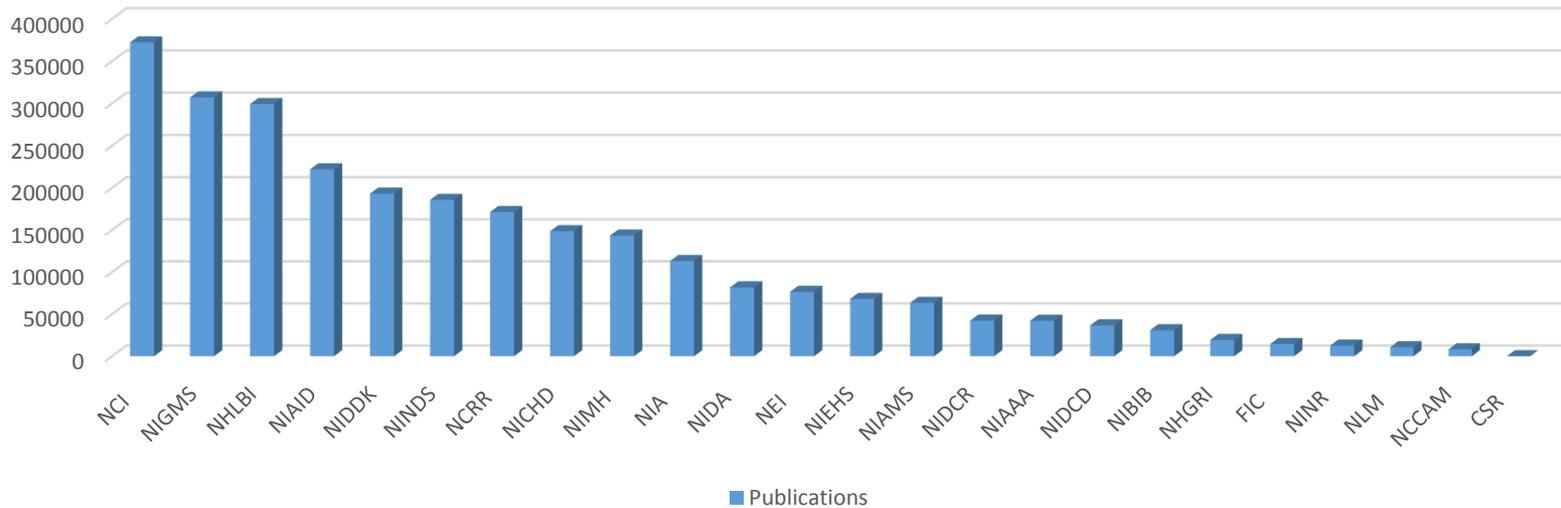
# SPIRES DATABASE: Summary Statistics

- NIH supported publication data 1980 – Present
- 2,116,373 Total Publications (Unique)
- 4,790,658 Total Grant to Publication Matches
- 340,015 Unique Base Project Numbers Matched
- 14,912 Journals Represented
- Publication data sources:
  - NLM PubMed/PubMed Central
  - NLM My Bibliography system (MyBib)
  - NIH Intramural Database (NIDB)

## Publication Counts by Year



## Publication Counts by IC (all years)



# SPIRES Data Quality: Match Case Scores

- The Match Case Score System is your assurance of the quality of a SPIRES project number match. Here's how it works:
  - SPIRES parses the grant number strings from publications into their component parts of activity code, NIH IC Code and grant serial number.
  - SPIRES assembles those components into the correct grant number format and attempts to validate that number against the database of awarded NIH projects.
  - The results of that attempt to match the components of a grant number are reflected in the Match Case Score.
  - Based on long term tracking of the percentage of Match Case Scores in the SPIRES database, SPIRES data quality is improving over time. 84% of SPIRES project number matches are absolute matches against a single, valid project number (Match Cases 4 and 5).

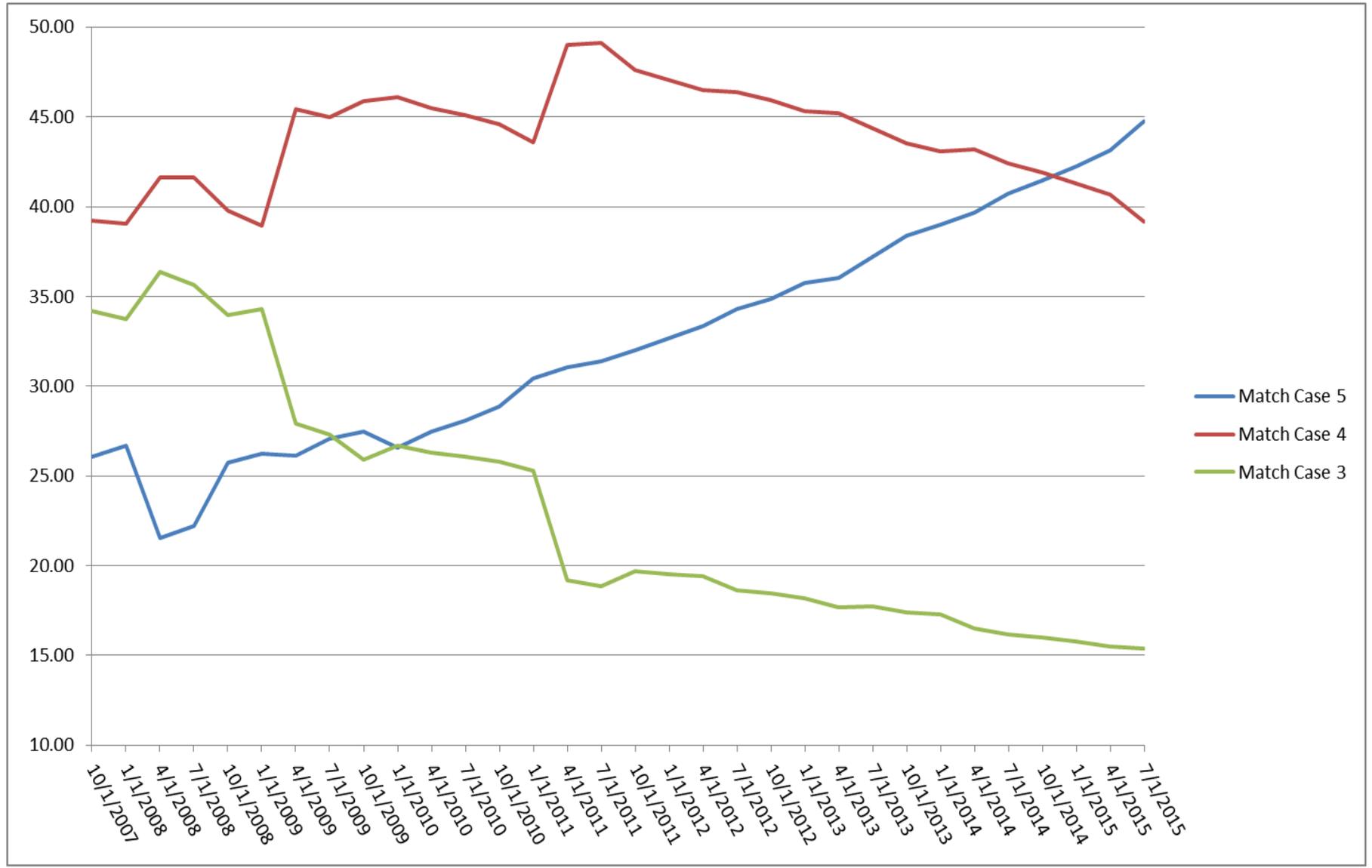
## SPIRES "Match Case" Score System

*Rates the "match quality" for each project number*  
*(Database statistics as of July 14, 2015)*

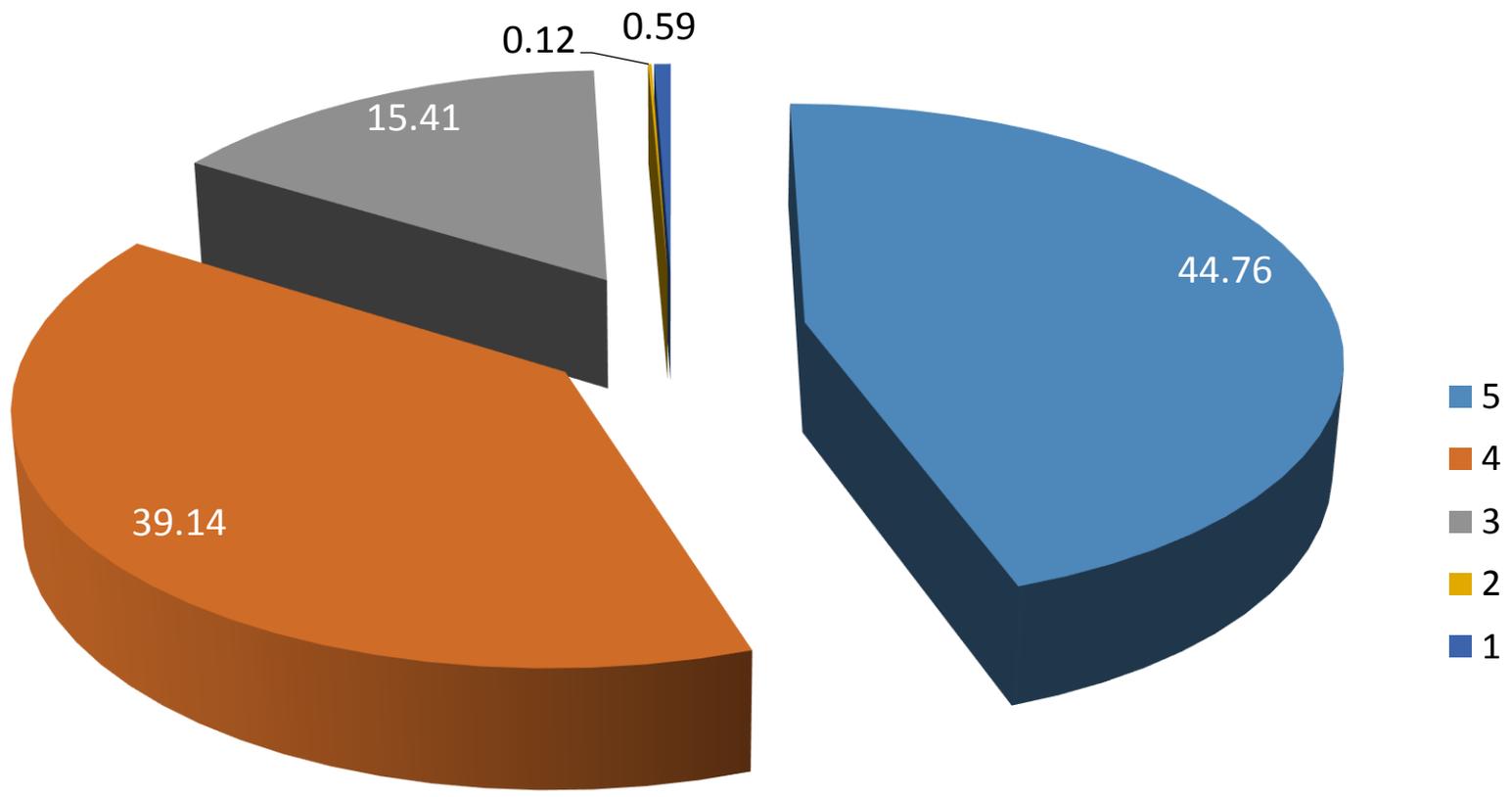
Case	Activity Code	PHS Org	Serial Number	Rows Matched	% Matched
5	YES	YES	YES	One ONLY	44.76%
4	N/A	YES	YES	One ONLY	39.14%
3	N/A	YES	YES	More than 1	15.41%
2	YES	YES	N/A	No match	0.12%
1	YES	N/A	YES	No match	0.57%
0	N/A	N/A	N/A	No match - No raw data	0.00%

- *% matched = Total count of project number rows for each match case score divided by total number of rows in project number match table.*
- *Spires matches using BASE project number (Activity Code + PHS Org Code + Serial Number)*
- *Combining "match case" with other query parameters improves the quality of search results.*
- *Grant numbers reported by author/pi via NIHMS and MyBib automatically receive a match score of 5.*

# Match Case Distribution Over Time



# Match Case Score Report



# Things Learned from SPIRES about NIH Research

- The fraction of articles reflecting synergies between multiple NIH institutes is increasing over time. 29% of current articles acknowledge grants from multiple Institutes.
- The median time lag to publication for a new grant is 3 years. Most grants that will produce an article will do so within 5 years.
- Each grant contributes to roughly 1.7 articles per year, averaged over all activity types.

Downloaded from [jamia.bmj.com](http://jamia.bmj.com) on May 5, 2011 - Published by [group.bmj.com](http://group.bmj.com)

## Research and applications

### Metrics associated with NIH funding: a high-level view

Kevin W Boyack,<sup>1</sup> Paul Jordan<sup>2</sup>

<sup>1</sup>SciTech Strategies,  
Albuquerque, New Mexico, USA  
<sup>2</sup>National Institutes of Health,  
Office of Research Information  
Systems (OD OER ORIS),  
Research Triangle Park, North  
Carolina, USA

**Correspondence to**  
Dr Kevin W Boyack, SciTech  
Strategies, 8421 Manuel Cia Pl.  
NE, Albuquerque, NM 87122,  
USA;  
[kboyack@mapofscience.com](mailto:kboyack@mapofscience.com)

Received 24 February 2011  
Accepted 2 April 2011

#### ABSTRACT

**Objective** To introduce the availability of grant-to-article linkage data associated with National Institutes of Health (NIH) grants and to perform a high-level analysis of the publication outputs and impacts associated with those grants.

**Design** Articles were linked to the grants they acknowledge using the grant acknowledgment strings in PubMed using a parsing and matching process as embodied in the NIH Scientific Publication Information Retrieval & Evaluation System system. Additional data from PubMed and citation counts from Scopus were added to the linkage data. The data comprise 2 572 576 records from 1980 to 2009.

**Results** The data show that synergies between NIH institutes are increasing over time; 29% of current articles acknowledge grants from multiple institutes. The

mined, standardized, and placed in a publicly available database for bulk download until just recently. The National Institutes of Health (NIH) Research Portfolio Online Reporting Tool Expenditures and Results (RePORTER) website was made available in mid-2009. This site allows one to query a wide variety of NIH funding and publication data. The ExPORTER extension ([http://project-reporter.nih.gov/exporter/ExPORTER\\_Catalog.aspx](http://project-reporter.nih.gov/exporter/ExPORTER_Catalog.aspx)) to make selected data available for bulk download was introduced several months later. Data have been added to the ExPORTER site at intervals over the past year—data back to 1985 are currently available, including tables for projects, links (articles to projects), articles, and patents.

Prior to the introduction of RePORTER and ExPORTER, it was possible to obtain funding and

Source: *Metrics Associated with NIH Funding: A High Level View*, Kevin W. Boyack, Paul Jordan, JAMIA, April 2011

# The SPIRES Application: Tools for Mining Publications

The SPIRES application lets the user approach their search for publications TWO WAYS:

- Query using projects search criteria – See related publications
- Query using publication search criteria – See related projects
- Manage your data and queries:
  - Create saved queries
  - Create saved “shopping carts” for publications or projects
    - Use shopping carts as part of your saved queries
  - Import shopping carts from QVR
    - Appl ID Carts
    - Person Carts
  - Upload project numbers to a cart
- Create custom reports:
  - From search results hit lists, create custom reports (custom download) for
    - Publication data
    - Project data
  - Download only the data items you select

**SPIRES**  
Scientific Publication Information Retrieval and Evaluation System

Publication Years in SPIRES: 1980 to Present  
 Total Publications in SPIRES: 1,724,327  
 Data Last Refreshed On: Apr 10, 2012 2:38 am  
About 19 hours ago

[Home](#)
[Search Publications](#)
[Search Projects](#)
[Reports](#)
[Utilities](#)
[Logout](#)

[Contact SPIRES Administrators](#)

### SPIRES

SPIRES is an eRA database, updated daily, that automatically maps NLM PubMed publication citations to NIH Grant Numbers. SPIRES includes publications resulting from intramural projects (NIDB), and incorporates data from the NIH Manuscript System (NIHMS). SPIRES citations include the PubMed Central ID number, when available.

SPIRES serves as a data provider to other NIH enterprise and extension systems, including QVR, RePORTER, eSPA, NDAR and others.

Advanced search forms are available

The next planned outage for SPIRES

For user support issues, please con

Try the new SPIRES Mobile app  
<http://spires.era.nih.gov/spires/mobil>

Read more about SPIRES:

- [SPIRES Release Notes, Nov](#)
- [SPIRES System Overview](#)
- [Journal Articles about SPIRE](#)
- [Match Case Scores](#)
- [NIH Institutes and Centers](#)
- [SPIRES Database Document](#)

**SPIRES**  
Scientific Publication Information Retrieval and Evaluation System

Publication Years in SPIRES: 1980 to Present  
 Total Publications in SPIRES: 1,724,327  
 Data Last Refreshed On: Apr 10, 2012 2:38 am  
About 19 hours ago

[Home](#)
[Search Publications](#)
[Search Projects](#)
[Reports](#)
[Utilities](#)
[Logout](#)

[Contact SPIRES Administrators](#)

[Search Projects](#)
[Default Criteria](#)
[Clear Criteria](#)

[Scroll to Section »](#)

Basic Search Options

Output Option:

Saved Query:

IC's:

AA	▲
AG	▢
AI	▢
AM	▢
AR	▢
AT	▼

Click item in left box to select  
Click item in right box to remove

Shopping Carts

Saved Project Carts:    Include projects in any of these carts:  
                                  Exclude projects in any of these carts:  
                                  Intersection of projects in all of these carts:

Saved Publication Carts: Include publications in any of these carts:

### Search Projects

### Search Publications

**SPIRES**  
Scientific Publication Information Retrieval and Evaluation System

Publication Years in SPIRES: 1980 to Present  
 Total Publications in SPIRES: 1,724,327  
 Data Last Refreshed On: Apr 10, 2012 2:38 am  
About 19 hours ago

[Home](#)
[Search Publications](#)
[Search Projects](#)
[Reports](#)
[Utilities](#)
[Logout](#)

[Contact SPIRES Administrators](#)

[Search Publications](#)
[Default Criteria](#)
[Clear Criteria](#)

[Scroll to Section »](#)

Basic Search Options

Output Option:

Saved Query:

IC's:

AA	▲
AG	▢
AI	▢
AM	▢
AR	▢
AT	▼

Click item in left box to select  
Click item in right box to remove

ALL
-----

Shopping Carts

Saved Project Carts:    Include projects in any of these carts:   
                                  Exclude projects in any of these carts:   
                                  Intersection of projects in all of these carts:   
 Saved Publication Carts: Include publications in any of these carts:

## Automated Shopping Carts: Let SPIRES Build Your Publication Collection For You

What are Automated Shopping Carts? - *“Automated Publication Carts” are used to automatically create collections of publications over time based on query parameters and refresh intervals specified by the user.*

- Automated carts are empty containers when created. They are populated based on one or more of your saved queries.
- You may specify their data refresh schedule OR you may run them on demand.
- Each time they are refreshed, automated carts add new data to the cart contents. Previously existing data remains in the cart until you explicitly remove it.
- Automated cart contents can be searched or filtered by the cart refresh date.
- Automated carts are color coded in the cart hit list: Yellow indicates an empty cart. Green indicates a cart with content.

Use automated carts to track the outputs from specific research programs or initiatives over time.

# SPIRES Data in QVR

- Display a publication list for each project.
- Run standard Bibliography reports.
- QVR “Like” feature is driven by SPIRES and “MBR” publication data fingerprints.
  - Publication fingerprints now include all pubs in PubMed from 2000 to the present – over 10 million publications!
- QVR Bibliography Reports are available in multiple formats:
  - Public Access Compliancy Check
  - Project Summary
  - Classic Bibliography
  - Raw data only

# REPARS: Reference Parsing and Retrieval Service

## Abstract

As federal programs are held more accountable for their research investments, The National Institute of Environmental Health Sciences (NIEHS) has developed a new method to quantify the impact of our funded research on the scientific and broader communities. In this article we review traditional bibliometric analyses, address challenges associated with these measures, and describe a new bibliometric analysis method, the Automated Research Impact Assessment (ARIA). ARIA taps into a resource that has only rarely been used for bibliometric analyses: references cited in “important” research artifacts, such as policies, regulations, clinical guidelines, and expert panel reports. The assessment includes new statistics that science managers can use to benchmark contributions to research by funding source. This new method provides the ability to conduct automated impact analyses of federal research that can be incorporated in program evaluations. We apply this method to several case studies to examine the impact of NIEHS funded research.

Source:

### **Automated Research Impact Assessment: A New Bibliometrics Approach**

Authors: Christina H. Drew<sup>1</sup>, Kristianna G. Pettibone<sup>1</sup>, Fallis Owen Finch, III<sup>2</sup>, Douglas Giles<sup>2</sup>, Paul Jordan<sup>3</sup>

<sup>1</sup> National Institute of Environmental Health Sciences, Program Analysis Branch

<sup>2</sup> Open Intelligence, Inc.

<sup>3</sup> National Institutes of Health, National Institutes of Health, Office of Research Information Systems (OD OER ORIS)

*Corresponding author: Christina H. Drew, Chief, Program Analysis Branch, National Institute of Environmental Health Sciences, Research Triangle Park, NC 27709 EMAIL: drewc@niehs.nih.gov  
TEL: 919-541-3319 FAX: 301-480-3142*

