

# The Scholarly Database (SDB)

---

Nianli Ma, Russell Duhon, Katy Borner

for the Digital Library Brown Bag

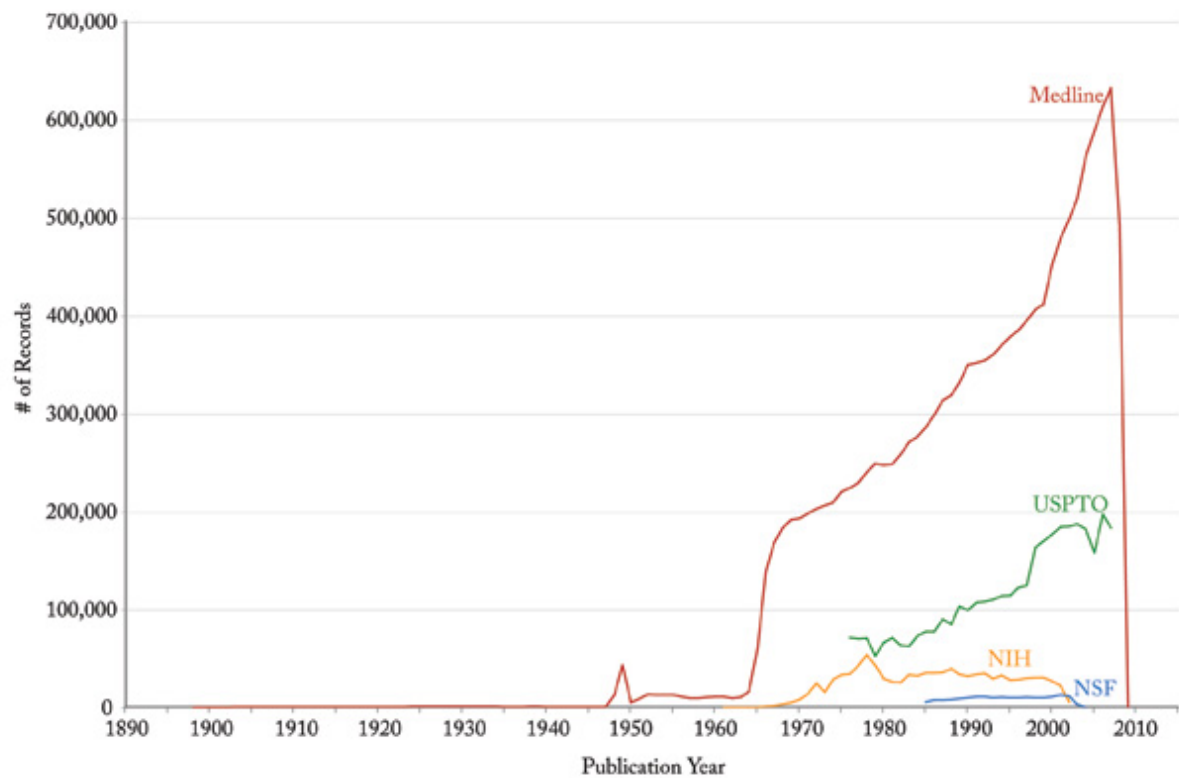
Nov 5<sup>th</sup>, 2008

# Data Sets

<b>Dataset</b>	<b>#Records</b>	<b>Years Coverage</b>	<b>updated</b>	<b>Restricted Access</b>
Medline	16,053,495	1898-2008	Yes	
PhysRev	398,005	1893-2006		Yes
PNAS	16,167	1997-2002		Yes
JCR	59,078	1974,1979,1984, 1989,1994-2004		Yes
USPTO	3,710,952	1976-2007	Yes	
NSF	174,835	1985-2003	Yes	
NIH	1,043,804	1972-2002	Yes	
Total	21,456,336	1893-2008	4	3

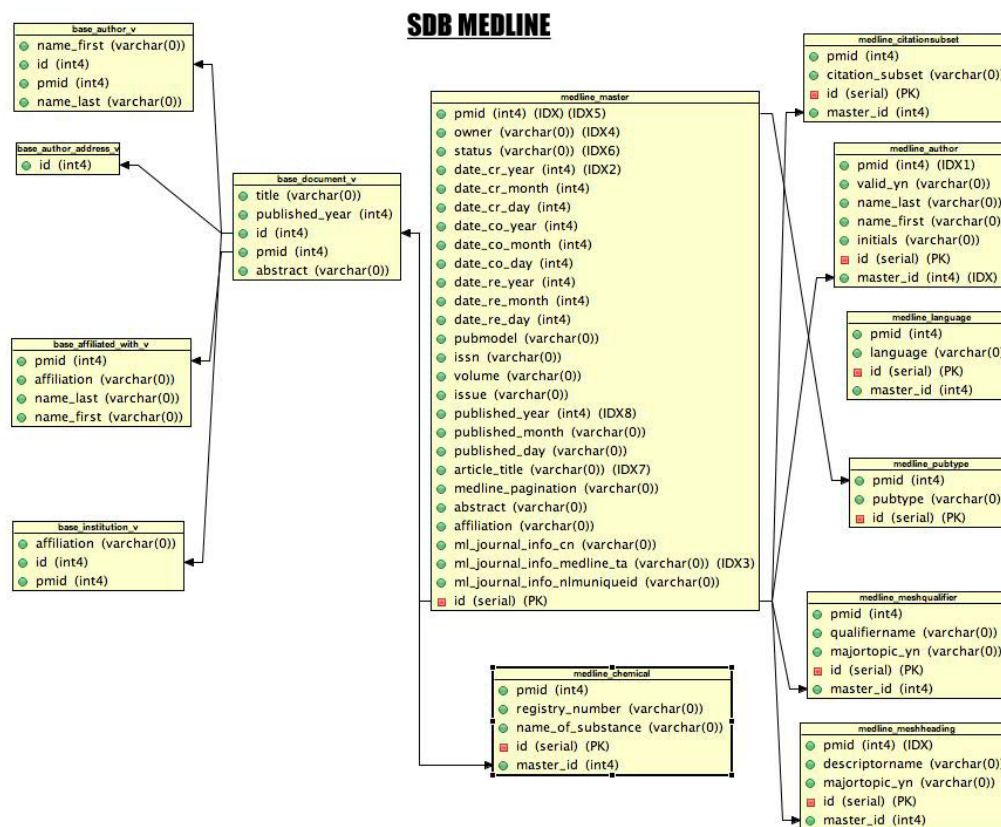
# Records Per year

---

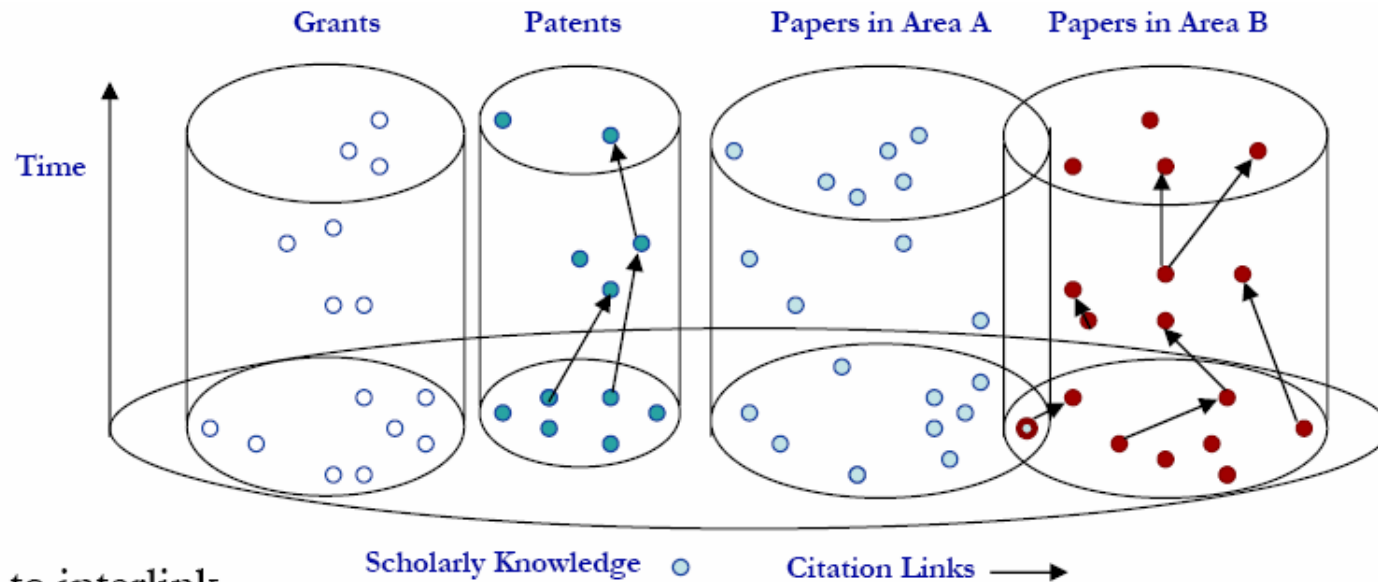


# Detailed Information

<https://nwb.slis.indiana.edu/community/?n=ScientometricsDatasets.HomePage>



# Challenges



## Need to interlink

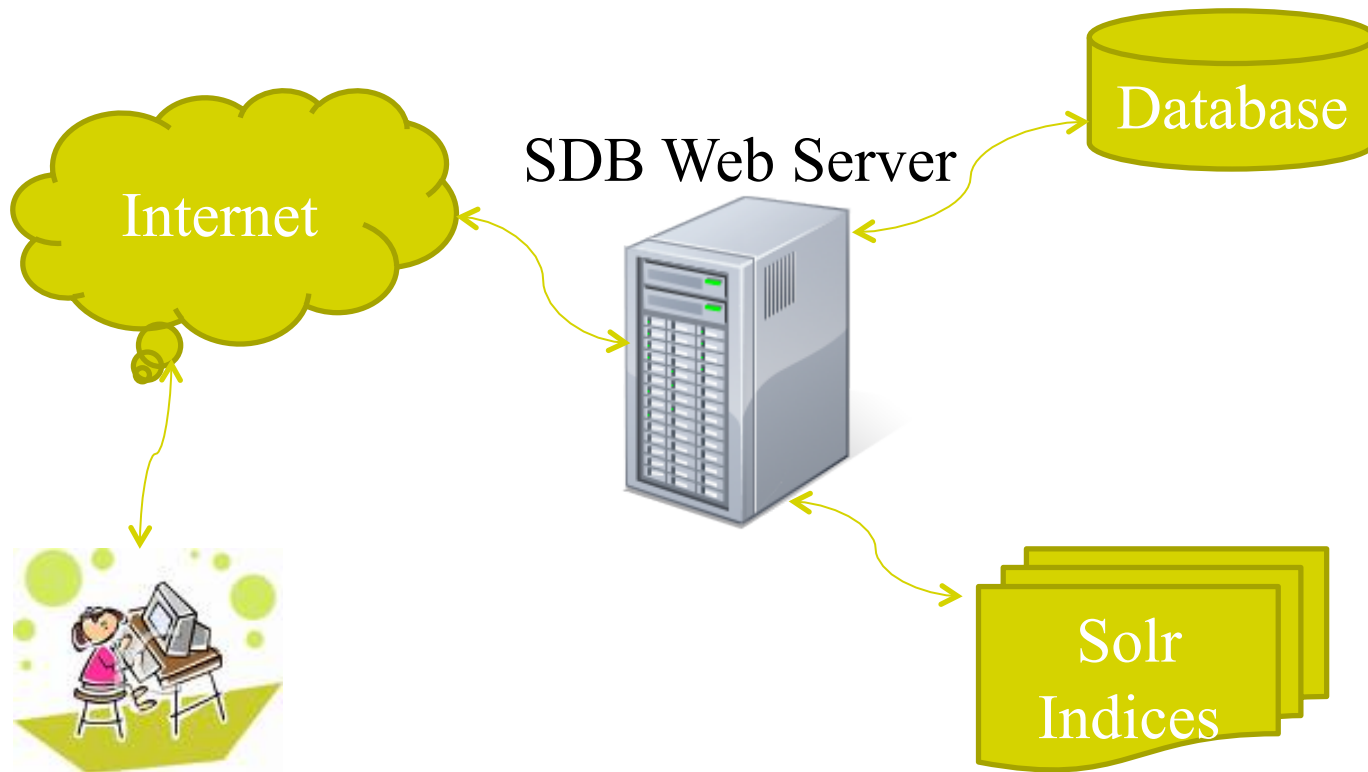
- Grants and papers/patents.
- Grants/papers/patents and their PIs/authors/inventors, etc.

## Use resulting networks to

- Count #papers, #citations, etc.
- Determine strength of co-PI/author/inventor relations, etc.

# SDB Architecture

---






# SDB Architecture (continued)

---

- Solr: full-text search server
  - <http://lucene.apache.org/solr/>
  - Open source
  - Uses the Lucene search library
- Interface developed in Django
  - <http://www.djangoproject.com/>
  - Open source
  - Particularly suited for content-focused web applications

# SDB Interface <http://sdb.slis.indiana.edu>



## SCHOLARLY DATABASE

Cyberinfrastructure for Network Science Center, SLIS, Indiana University, Bloomington

### IU User

IU Users must login using the Central Authentication Service (CAS), the standard IU authentication system. Please click the button below to proceed to the IU login page.

[Go to IU Login](#)

### Non-IU User

Email:

Password:

[Login](#)

**Not Registered Yet?**

[Register as an IU User](#)

[Register as a Non IU User](#)

**In the News**

Whitfield, John. 2008. *Group Theory*. Nature, 455, 9: 720-723.




**Please Cite As**

La Rove, Gavin, Ambre, Sumeet, Burgoon, John, Ke, Weimao and Börner, Katy. (2007) The Scholarly Database and Its Utility for Scientometrics Research. In Proceedings of the 11th International Conference on Scientometrics and Informetrics, Madrid, Spain, June 25-27, 2007, pp. 457-462.  
<http://ella.slis.indiana.edu/~katy/paper/07-issi-sdb.pdf>

**Acknowledgements**


The Scholarly Database is funded by the School of Library and Information Science and the Cyberinfrastructure for Network Science center at Indiana University, the National Science Foundation under Grants No. IIS-0238261 and IIS-0513650, and a James S. McDonnell Foundation grant in area Studying Complex Systems.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.



James S. McDonnell Foundation

Last Modified October 6th, 2008 | Graphic Design by Elisha Hardy



## SCHOLARLY DATABASE

Cyberinfrastructure for Network Science Center, SLIS, Indiana University, Bloomington

[Search](#) | [Edit Profile](#) | [Admin](#) | [About](#) | [Logout](#)

### Search

Creators:

Title:

Abstract:

Full Text:

First Year:

Last Year:

Medline (1898 - 2008)

NIH (1961 - 2002)

NSF (1985 - 2004)

USPTO (1976 - 2007)

[Search](#)

If multiple terms are entered in a field, they are automatically combined using 'OR'. So, 'breast cancer' matches any record with 'breast' or 'cancer' in that field.

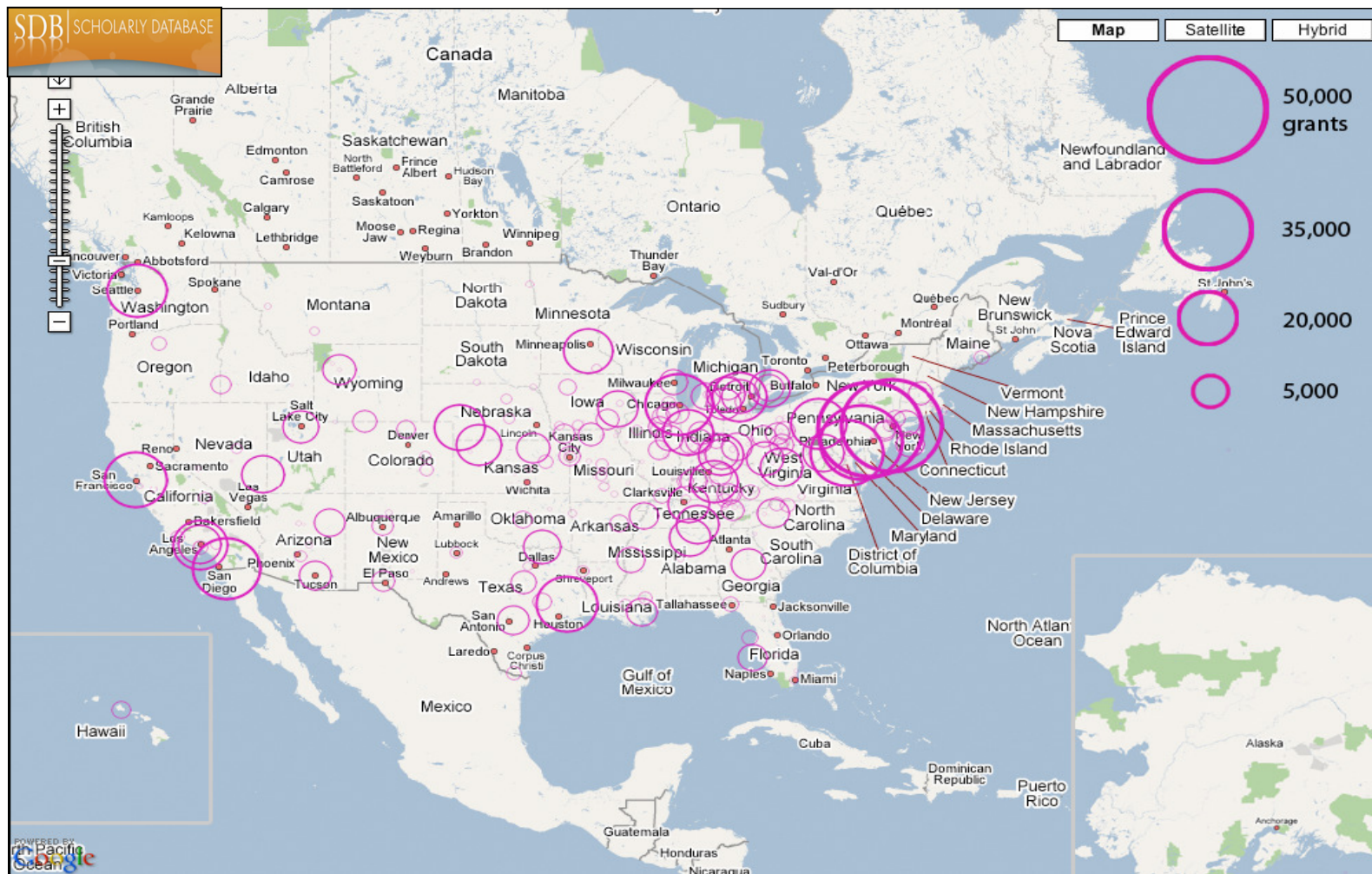
You can put AND between terms to combine with 'AND'. Thus 'breast AND cancer' would only match records that contain both terms.

Double quotation can be used to match compound terms, e.g., "breast cancer" retrieves records with the phrase 'breast cancer', and not records where 'breast' and 'cancer' are both present, but not the exact phrase.

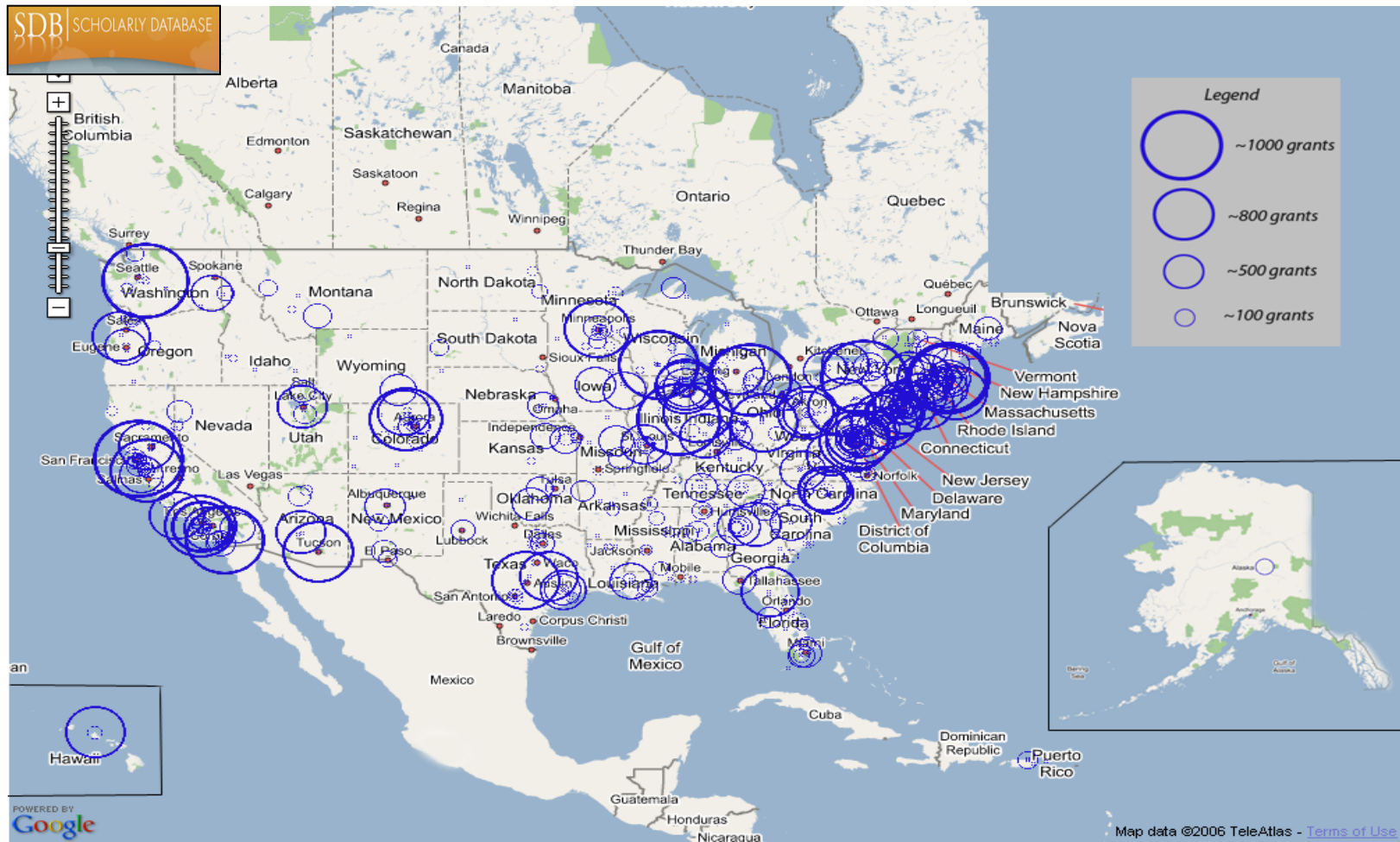
The importance of a particular term in a query can be increased by putting a ^ and a number after the term. For instance, 'breast cancer^10' would increase the importance of matching the term 'cancer' by ten compared to matching the term 'breast'.



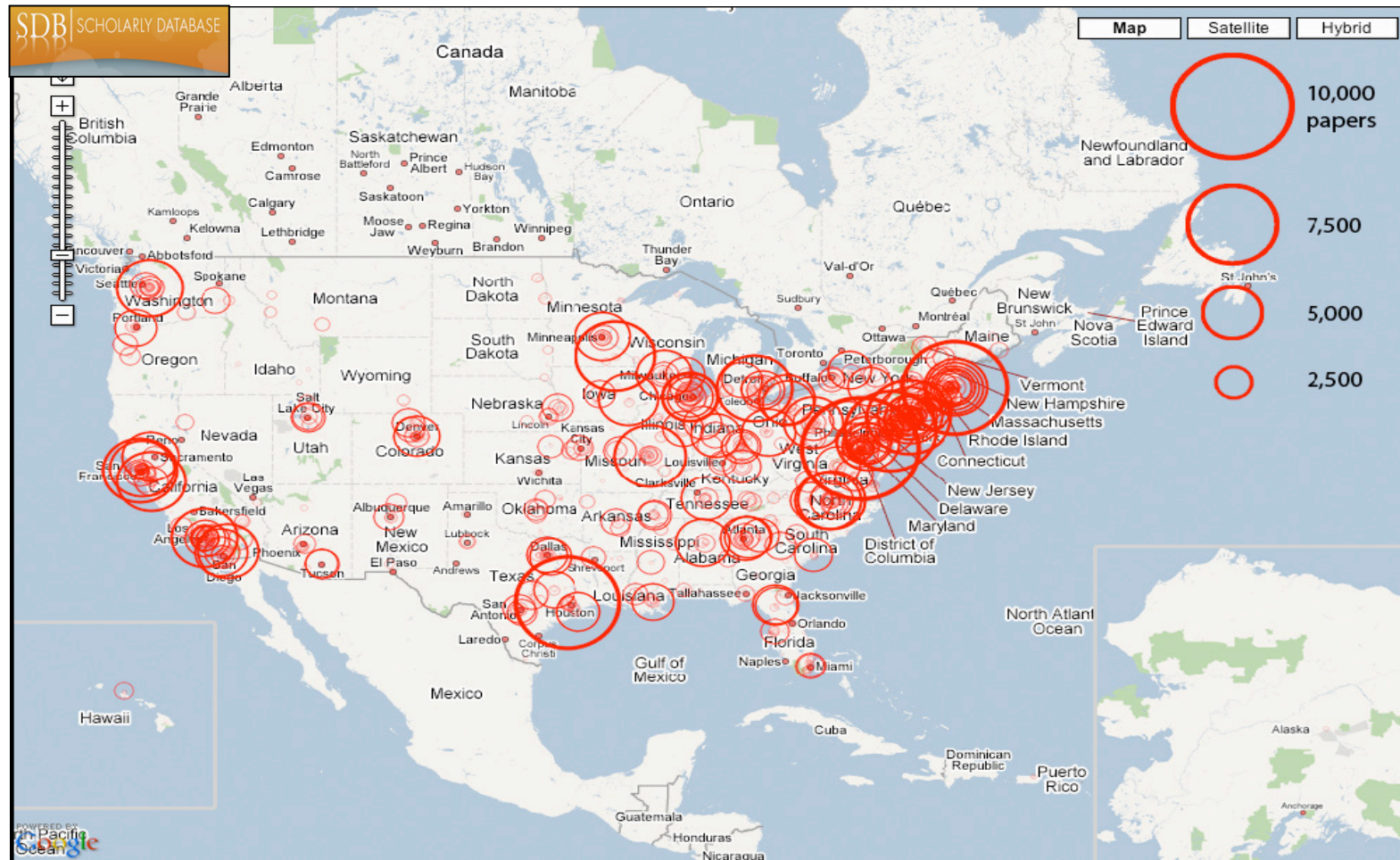
# NIH Awards by Geo-location (U.S. only)



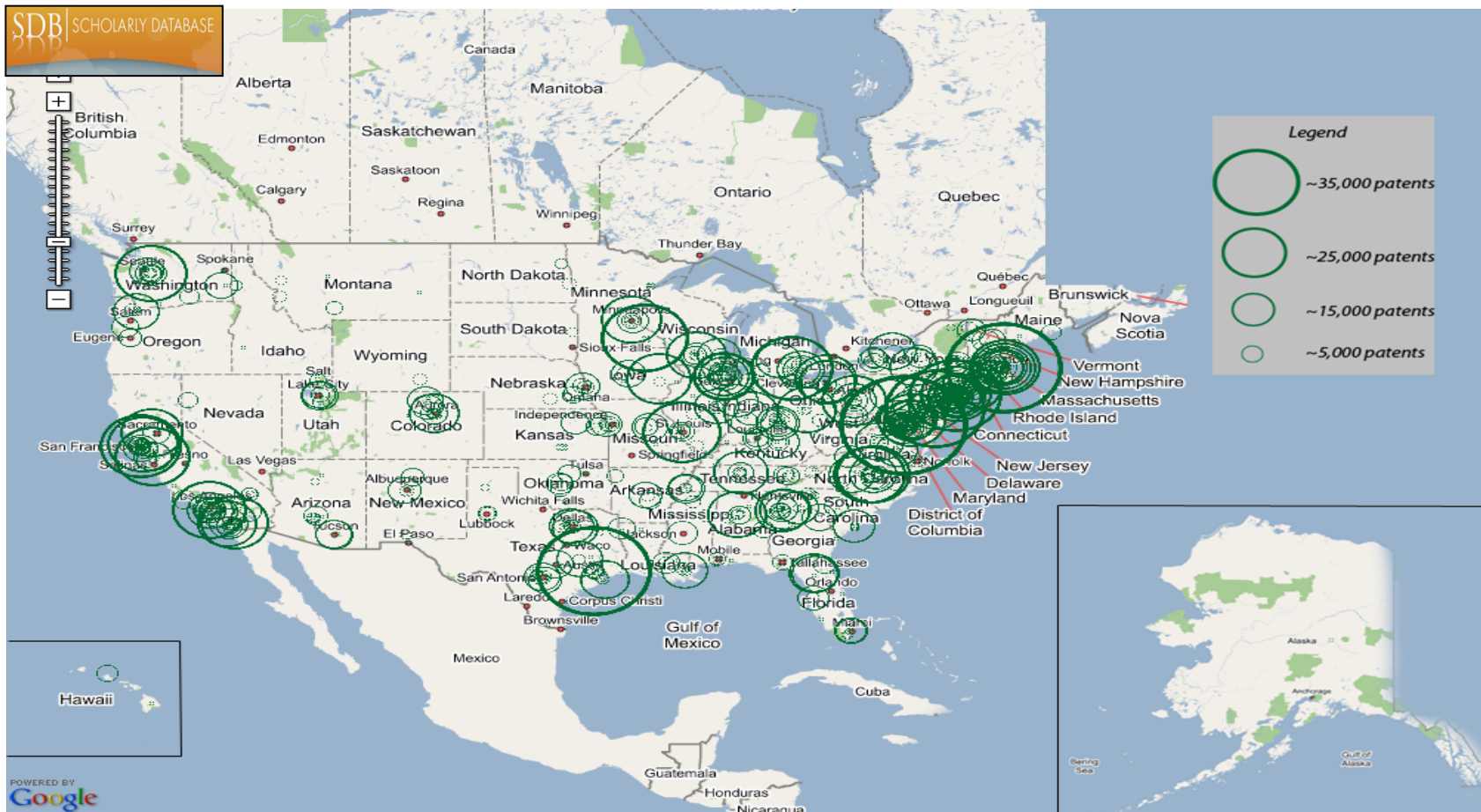
# NSF Grants by Geo-location (U.S. only)



# Number of Medline Publications



# US Patent by Geo-location (U.S. only)





# Future Work

---

- Provide full-text search on science news
- Add JCR impact factor to publications
- Discover user interests through what they search and download



# References

---

- La Rowe, Gavin, Ambre, Sumeet, Burgoon, John, Ke, Weimao and Börner, Katy. (2007) The Scholarly Database and Its Utility for Scientometrics Research. In Proceedings of the 11th International Conference on Scientometrics and Informetrics, Madrid, Spain, June 25-27, 2007, pp. 457-462.  
<http://ella.slis.indiana.edu/~katy/paper/07-issi-sdb.pdf>
- Whitfield, John. 2008. [Group Theory](#). Nature, 455, 9: 720-723.



# Thanks

---

- You !
- National Science Foundation under grants nos. IIS-0238261 and IIS-0513650
- James S. McDonnell Foundation grant in the area of Studying Complex Systems