Compelling evidence: New tools and methods for aligning collections with the research mission

Presenter: Joelen Pastva

Project Team: Bart Davis, Karen Gutzman, Stacy Konkiel, Ramune Kubilius, Aaron Sorensen

NASIG 2019, Pittsburgh, PA
Profile:
Galter Health Sciences Library & Learning Center
Northwestern University
Feinberg School of Medicine
Chicago, Illinois
Our community:
Feinberg School of Medicine (FSM)
- 3351 students, residents, & fellows
- 4180 faculty
- Staff
- FSM generates 69% of all research money at Northwestern

Northwestern University – Main Library
- Administratively separate w/shared systems, cost-sharing of subscriptions
- Standalone subscriptions for medical collections
- Resource sharing w/affiliated hospitals
NUCATS provides scientists with consultative resources and expertise in order to accelerate how quickly transformative scientific discoveries make their way to patients and the community. It is our goal to continually increase the quality, safety, efficiency and speed of innovative clinical and translational research.”
Northwestern University Clinical & Translational Science (NUCATS)

Outside of traditional scholarly communication, how can Galter Library best support the research needs of NUCATS and the FSM community?

Translational research ➔ from observation to invention

Common outputs:
• Clinical trials
• Patents
Opportunity for collaboration

Dimensions

Re-imagining discovery and access to research

History:
- 2017 citation analysis project
- Symplectic Elements customer
- Development partner for Dimensions (though not yet a subscriber)
Enriched and interlinked – a modern approach to research information management

- Dimensions data
  - Policy docs
  - Patents
  - Data sets
  - Altmetric
  - Clinical trials
  - Publications
  - Grants

- Enrichment
  - Institution identification
  - Categorization
  - Concept extraction
  - Researcher disambiguation
  - Reference extraction

- Enriched metadata
  - Concepts
  - Researchers
  - Organizations
  - Classifications
  - References
  - Metrics
The data and links driving Dimensions...

- **1.0bn Publications**
  - Improved metadata of 60m
  - 12m links

- **4.5m Grants**
  - $1.3+ trillion in funding
  - 370k to funders

- **700k Links**

- **460k Clinical trials**
  - 399k links

- **38m Patents**
  - 155k links

- **700k Links**

- **84m Altmetric data points**
  - 9.9m links

- **318m Links**

- **100m Publications**
  - 1.0bn links

- **422k Policy documents**
  - 16m to funders

- **865k Links**

- **10m Links**

- **202k to funders**
A version for everyone

<table>
<thead>
<tr>
<th>Dimensions Free</th>
<th>Dimensions Plus</th>
<th>Dimensions Analytics</th>
<th>Dimensions API</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Free access for individual researchers</td>
<td>- Additional fully integrated databases for grants, patents and clinical trials</td>
<td>- Workflow support for funders / publishers</td>
<td>- The Dimensions Search Language - Easy to use even with little technical knowledge</td>
</tr>
<tr>
<td>- More than 99M publications</td>
<td>- Download 5,000 records</td>
<td>- Download 50,000 records</td>
<td>- Extract data for reports</td>
</tr>
<tr>
<td>- More than 1B citations</td>
<td>- Single Sign On (SSO)</td>
<td>- Additional analytical views</td>
<td>- Make use of aggregated metrics, clustering, and sorting features</td>
</tr>
<tr>
<td>- Download 500 records</td>
<td>- Full Altmetric badges/ aggregated metrics</td>
<td>- Additional filters</td>
<td></td>
</tr>
<tr>
<td>- Article-level Altmetric data</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Dimensions Profiles: A custom solution designed to support your organization
Dimensions Plus

The most complete view of the research landscape

- Simultaneously search across all content types
- Subject areas consistently applied to all content types, as well as funder and research organizations
- Extended data get the full picture of past, current, and future research
The Dimensions API

Powerful API - designed to allow flexible use of the enriched data

- Use without constraints for internal purposes
- Use data outside of the web-app; e.g. in admin systems or analytical software
- Querying language made and documented specifically for Dimensions
Notes on the data – Clinical Trials

Defined:
Experiments or observations in clinical research designed to answer specific questions about interventions, such as drugs or medical devices, with a goal of collecting safety and efficacy data.

Registration of trials may be required depending on the country of the trial. ClinicalTrials.gov is the largest and most widely used registry.\(^2\)

Citations or publications related to study results or background may be provided by the investigator or sponsor.
Notes on the data – Clinical Trials

@FSM

● Hundreds of trials ongoing daily, though may not be registered in ClinicalTrials.gov³
● 4571 clinical trials conducted in 2017-2018

@Dimensions

Data from 10 registries from across the world, including ClinicalTrials.gov⁴
Notes on the data – Patents

Defined:
A grant of a property right giving its holder exclusive access to an invention for a limited period of time.\textsuperscript{5}

Prior art – documents (other patents or non-patent literature) demonstrating what already exists at the time a patent is filed in order to justify its novelty. Non-patent literature frequently includes scientific journal articles, though the amount of prior art required varies by patent region.
Notes on the data – Patents

@FSM
Grants and programs to accelerate commercialization and increase entrepreneurship

@Dimensions
Data from 10 patent offices from across the world, including the US Patent and Trademark Office (USPTO) and the European Patent Office (EPO)
Methodology

Data criteria (Galter)

Clinical trials (pilot)
- Northwestern affiliated
- Subject/department – Dermatology (trials containing keyword “skin”)

Patents
- Northwestern University as assignee
- 2008-2017 (filed date)

Data gathering (Dimensions)
- Queries constructed and run using the Dimensions API user interface
Results

Clinical trials pilot:
- List of journals cited with counts of number of times cited

Patents:
- Spreadsheet of patent-level description joined with cited reference metadata
  - Filtered based on presence of Dimensions-applied Research, Condition, and Disease Categorization (RCDC) code or medical topic in title to focus on FSM-related activities
  - Analysis using Excel and Python (NumPy, Pandas via Jupyter notebook)
  - Visualizations using Excel and Tableau
Pilot – Clinical Trials

Cited Reference Analysis
of clinical trials affiliated with Northwestern in the field of dermatology

730
journals were cited

Please note: Trial data (including cited references) were exported from Dimensions in 12/2018. Data analysis and visualization of the cited references were completed using Microsoft Excel.

Top 20 Most Cited Journals

<table>
<thead>
<tr>
<th>JOURNAL NAME</th>
<th>NUMBER OF CITED REFERENCES FROM JOURNAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>JAMA DERMATOLOGY</td>
<td>40</td>
</tr>
<tr>
<td>JOURNAL OF THE AMERICAN ACADEMY OF DERMATOLOGY</td>
<td>27</td>
</tr>
<tr>
<td>ANESTHESIA &amp; ANALGESIA</td>
<td>22</td>
</tr>
<tr>
<td>QUALITY OF LIFE RESEARCH</td>
<td>18</td>
</tr>
<tr>
<td>CIRCULATION</td>
<td>15</td>
</tr>
<tr>
<td>JOURNAL OF THE AMERICAN COLLEGE OF CARDIOLOGY</td>
<td>13</td>
</tr>
<tr>
<td>Atherosclerosis</td>
<td>10</td>
</tr>
<tr>
<td>NEW ENGLAND JOURNAL OF MEDICINE</td>
<td>10</td>
</tr>
<tr>
<td>DIABETES CARE</td>
<td>10</td>
</tr>
<tr>
<td>THE JOURNAL OF IMMUNOLOGY</td>
<td>9</td>
</tr>
<tr>
<td>AMERICAN JOURNAL OF EPIDEMIOLOGY</td>
<td>9</td>
</tr>
<tr>
<td>BRITISH JOURNAL OF DERMATOLOGY</td>
<td>9</td>
</tr>
<tr>
<td>AMERICAN JOURNAL OF CLINICAL NUTRITION</td>
<td>9</td>
</tr>
<tr>
<td>TRANSPLANTATION JOURNAL</td>
<td>9</td>
</tr>
<tr>
<td>BRITISH JOURNAL OF ANAESTHESIA</td>
<td>8</td>
</tr>
<tr>
<td>THE LANCET</td>
<td>8</td>
</tr>
<tr>
<td>JAMA</td>
<td>7</td>
</tr>
<tr>
<td>JOURNAL OF NEUROSURGERY</td>
<td>7</td>
</tr>
<tr>
<td>ANESTHESIOLOGY</td>
<td>7</td>
</tr>
<tr>
<td>JOURNAL OF CLINICAL EPIDEMIOLOGY</td>
<td>6</td>
</tr>
</tbody>
</table>
Cited Reference Analysis
of patents granted to Northwestern-affiliated faculty from 2008 - 2017

1,163 journals were cited from 2008 - 2017

43% of articles cited are open access
Though it is not known if they were so at the time the patents were filed...

Please note: Patent data (including cited references) were exported from Dimensions in 5/2019. Data analysis and visualization of the cited references were completed using Microsoft Excel.

Top 20 Most Cited Journals

<table>
<thead>
<tr>
<th>JOURNAL NAME</th>
<th>NUMBER OF CITED REFERENCES FROM JOURNAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES*</td>
<td>249</td>
</tr>
<tr>
<td>JOURNAL OF THE AMERICAN CHEMICAL SOCIETY</td>
<td>201</td>
</tr>
<tr>
<td>SCIENCE*</td>
<td>180</td>
</tr>
<tr>
<td>NATURE*</td>
<td>159</td>
</tr>
<tr>
<td>JOURNAL OF BIOLOGICAL CHEMISTRY*</td>
<td>124</td>
</tr>
<tr>
<td>NANO LETTERS</td>
<td>105</td>
</tr>
<tr>
<td>LANGMUIR</td>
<td>82</td>
</tr>
<tr>
<td>BIOMATERIALS</td>
<td>75</td>
</tr>
<tr>
<td>JOURNAL OF MEDICINAL CHEMISTRY</td>
<td>64</td>
</tr>
<tr>
<td>ADVANCED MATERIALS</td>
<td>64</td>
</tr>
<tr>
<td>JOURNAL OF NEUROSCIENCE</td>
<td>59</td>
</tr>
<tr>
<td>APPLIED PHYSICS LETTERS</td>
<td>59</td>
</tr>
<tr>
<td>THE JOURNAL OF PHYSICAL CHEMISTRY B</td>
<td>50</td>
</tr>
<tr>
<td>ANGEWANDTE CHEMIE INTERNATIONAL EDITION</td>
<td>50</td>
</tr>
<tr>
<td>THE JOURNAL OF CELL BIOLOGY</td>
<td>48</td>
</tr>
<tr>
<td>BIOCHEMISTRY</td>
<td>48</td>
</tr>
<tr>
<td>AMERICAN JOURNAL OF CANCER</td>
<td>42</td>
</tr>
<tr>
<td>NUCLEIC ACIDS RESEARCH</td>
<td>36</td>
</tr>
<tr>
<td>CELL</td>
<td>36</td>
</tr>
<tr>
<td>THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA</td>
<td>35</td>
</tr>
</tbody>
</table>

80% of the citations (4,488 of the 5,611 citations) were to the top 30% of the journals (338 of 1,163 journals)
* Also in top 20 most highly cited journals in 2017 article analysis
Patents – top 20 RCDC codes

- Heart Disease
- Peripheral Neuropathy Cardiovascular
- Pain Conditions - Chronic
- Rare Diseases
- Prevention
- Brain Disorders
- Assistive Technology
- Neurosciences
- Cancer
- Urologic Diseases
- ALS
- Bioengineering
- Nanotechnology
- Stem Cell Research - Nonembryonic - Non-Human
- Neurodegenerative
- Biotechnology
- Acquired Cognitive Impairment
- Regenerative Medicine
- Dementia
- Prostate Cancer
- Aging
- Alzheimer's Disease
- Genetics
Cited Reference Analysis
of patents granted to Northwestern-affiliated faculty from 2008 - 2017

Please note: Patent data (including cited references) were exported from Dimensions in 5/2019. Data analysis and visualization of the cited references were completed using Microsoft Excel.
Cited Reference Analysis of patents granted to Northwestern-affiliated faculty from 2008 - 2017

Number of Citations for Top 20 Titles, Per Journal Per Year Patent Filed

Year Patent Filed


Number of Citations to Journal

- Advanced Materials
- Biochemistry
- Angewandte Chemie International Edition
- American Journal of Cancer
- Angewandte Chemie
- Applied Physics Letters
- Biochemistry
- Biomaterials
- Cell
- Journal of Biological Chemistry
- Journal of Biomedical Materials Research
- Journal of Medicinal Chemistry
- Journal of the American Chemical Society
- Langmuir
- Journal of Neuroscience
- Nature
- Nucleic Acids Research
- Nano Letters
- Proceedings of the National Academy of Sciences
- Science
- The Journal of Cell Biology
- The Journal of Physical Chemistry B
- The Journal of the Acoustical Society of America
Cited Reference Analysis
of patents granted to Northwestern-affiliated faculty from 2008 - 2017

Average Number of Cited References, Per Patent Per Year

Please note: Patent data (including cited references) were exported from Dimensions in 5/2019. Data analysis and visualization of the cited references were completed using Microsoft Excel.
Cited Reference Analysis of patents granted to Northwestern-affiliated faculty from 2008 - 2017

Please note: Patent data (including cited references) were exported from Dimensions in 5/2019. Data analysis and visualization of the cited references were completed using Microsoft Excel.
Cited Reference Analysis of dermatology-related patents granted to Northwestern-affiliated faculty from 2008 - 2017

101 journals were cited from 2008 - 2017

**Top 5 Most Cited Journals**

<table>
<thead>
<tr>
<th>JOURNAL NAME</th>
<th>NUMBER OF CITED REFERENCES FROM JOURNAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>JOURNAL OF BIOLOGICAL CHEMISTRY</td>
<td>11</td>
</tr>
<tr>
<td>JOURNAL OF INVESTIGATIVE DERMATOLOGY</td>
<td>5</td>
</tr>
<tr>
<td>NATURE BIOTECHNOLOGY</td>
<td>5</td>
</tr>
<tr>
<td>PLASTIC &amp; RECONSTRUCTIVE SURGERY</td>
<td>4</td>
</tr>
<tr>
<td>PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES</td>
<td>4</td>
</tr>
</tbody>
</table>

Please note: Patent data (including cited references) were exported from Dimensions in 5/2019. Data analysis and visualization of the cited references were completed using Microsoft Excel.
Cited Reference Analysis of dermatology-related patents granted to Northwestern-affiliated faculty from 2008 - 2017

Please note: Patent data (including cited references) were exported from Dimensions in 5/2019. Data analysis and visualization of the cited references were completed using Microsoft Excel.
Vs traditional metrics

Impact factor
“measure of the frequency with which the “average article” in a journal has been cited in a particular year or period.”⁶

COUNTER usage statistics:
JR1 – Successful full-text article requests by journal
● Still an antiquated process at Galter, complicated by shared purchase landscape at NU
● Tends to correspond to other usage metrics

Usage R-value = 0.8137
(strong correlation)
IF R-value = 0.1635
(weak correlation)

Usage R-value = 0.6068 (moderate correlation)
IF R-value = 0.301 (weak correlation)

Removing usage outliers Nature and Science make for R-value of 0.8398 (strong correlation)
Discussion

- Data caveats:
  - Clinical trials – pilot run; waiting on improved API functionality
  - Patents
    - Filed year vs. publication year
    - Patent process muddies origin of citations
    - Overall smaller number of citations to journal literature compared to journal articles
  - Smaller dataset compared to traditional journal article citations
    - Field-specific citation patterns have larger impact on overall picture
Discussion

- Collection development applications:
  - No gaps identified
  - Usage vs citation shows some variation, but overall a strong positive correlation
  - Older articles maintain significance as prior art for patents
  - Different “core” set of journals in patent universe
  - Strong OA showing despite seeming reluctance to embrace in STEM
    - Perhaps research funding plays a role
Conclusion

Where we started –

- Attempt to replicate traditional citation analysis study using Dimensions

Where we ended –

- Investigating new resource types and new data fields for potential further research (i.e. patent-patent, OA status, article metrics, RCDC and other classification systems)
Conclusion

Clinical trials & patents are fruitful resources for institutions involved in these activities, but should be combined with other metrics given the smaller, specialized sample pool.

Still so much more to investigate with this data!

Dimensions is a great tool (with great people supporting it) for getting started. New features are being added regularly.
1. **Facts and figures**. Retrieved from https://www.feinberg.northwestern.edu/about/facts/index.html

2. **Learn about clinical studies**. Retrieved from https://clinicaltrials.gov/ct2/about-studies/learn#ClinicalTrials

3. **Clinical trials**. Retrieved from https://www.feinberg.northwestern.edu/research/clinical-trials/trials.html


Thank you!

Joelen Pastva
joelen.pastva@northwestern.edu
@jolophon

Stacy Konkiel
stacy@altmetric.com