

Working with the Dimensions and Altmetric APIs to search, visualize, and integrate research information across library systems



Stacy Konkiel
ER&L 2019, Austin, TX
stacy@altmetric.com

@skonkiel #dsmetrics

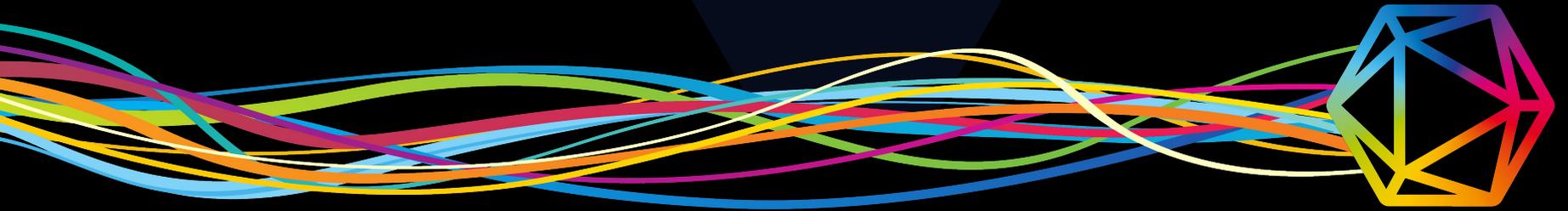
Today's agenda

- Introductions + intention setting (10m)
- Introduction to Dimensions + Altmetric data (20m)
- Working with JSON APIs using Python and Jupyter notebooks (20m)
- Exercise: Retrieving and developing metrics for reporting (10m)
- Exercise: Visualizing research information, citations, and altmetrics (10m)
- Exercise: Porting research information to a website (10m)
- Wrap up (10m)

Together: Your name, role,
institution

Individually: What you hope to be
able to do with the skills you learn
today

Dimensions data overview



The general approach to data in Dimensions

Inclusive approach

We do not decide what is relevant research output

Dimensions is open for data integration (within reason; e.g. no fraudulent journals)

The user should decide what is relevant use-case by use-case

We take the responsibility to enable that by providing the right tools

Building the publications backbone: Two steps

STEP 01

Getting a metadata backbone in place

Aggregate, deduplicate, homogenise,
and more - 97M publication records*

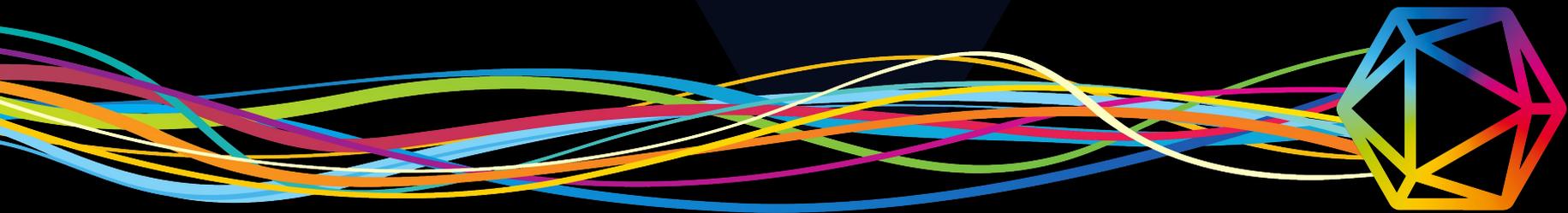
STEP 02

Increasing discoverability

Process full text; so far 60M -
enrich records, count citations,
build search index, classify at
publication record level

* CrossRef, Pubmed, PubMed Central, jstage, arXiv, bioRxiv, Europe PubMed Central, OpenAire, and more

Indicators



Indicators available for analysis

Times cited	Number of times a publication or patent was cited by other publications/patents	Article-level Aggregated
Altmetric Attention Score	Number of mentions in blogs, news outlets, policy documents, social media, etc	Article-level Aggregated
RCR	Relative Citation Ratio quantifies the influence of a research article by using its co-citation network to field-normalize the number of citations it has received	Article-level Aggregated
FCR	Field Citation Ratio compares an article's citations against articles of similar age and in a similar subject area	Article-level Aggregated
Citation recency	Shows the number of citations received by an article in the last two years	Article-level Aggregated
Counts	For publications, patents, grants, policy documents, clinical trials	Aggregated
Funding	Amount awarded per grant	Grant-level Aggregated

Data to Innovate

- Invent your own metrics

Example: PGI (publications per active grant)



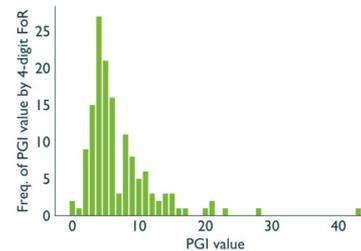
A Composite Indicator for Publications per Grant



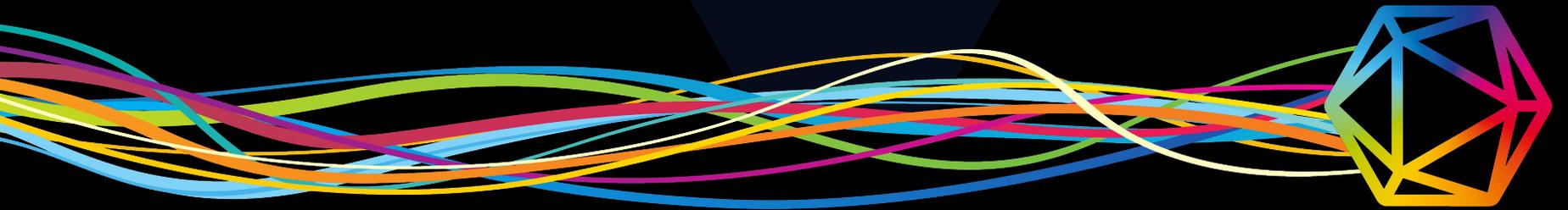
The Dimensions platform provides access to global data for both grants and publications and makes possible the construction of links between different aspects of research. The dataset will enable increasingly sophisticated composite indicators to be created as needed by the research community. We explored a simple composite indicator: the ratio of publications to active grants. This is an imperfect representation of research inputs and outputs, but it offers real value for exploratory purposes. For example, a high ratio of publications to grants (relative to a subject average) might indicate: research areas that do not attract grant funding (e.g. Complementary Medicine); strong industry funding (e.g. Pharmaceutical companies); or internally collaborative research.

To create a Publications-to-Grants Indicator (PGI), we took articles published after 2015 and compared this to a count of active grants in the period 2015–17. Active grant count (as opposed to the funding amount) addresses varying award lengths and funding systems in different countries and is more meaningful for cross-disciplinary comparisons. Indicators were calculated for data grouped in categories denoted by common FoR codes for grant and publication classification.

Dimensions will enable increasingly sophisticated composite indicators to be created by the research community



Outputs

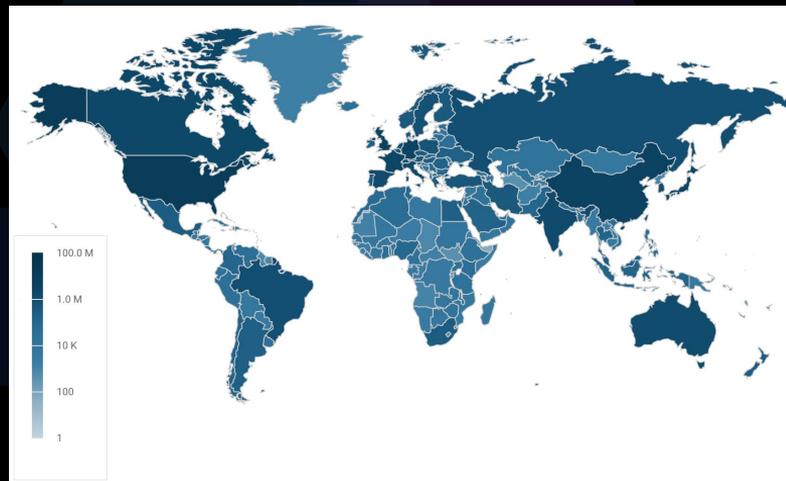


Publications data

00

PUBLICATIONS

- 98M research outputs
- Related information:
 - OA status
 - Authors, affiliations, locations
 - Supporting grants
 - Subject areas
 - Publisher
 - Journal (where applicable)
 - References
 - Citing documents (via other outputs)

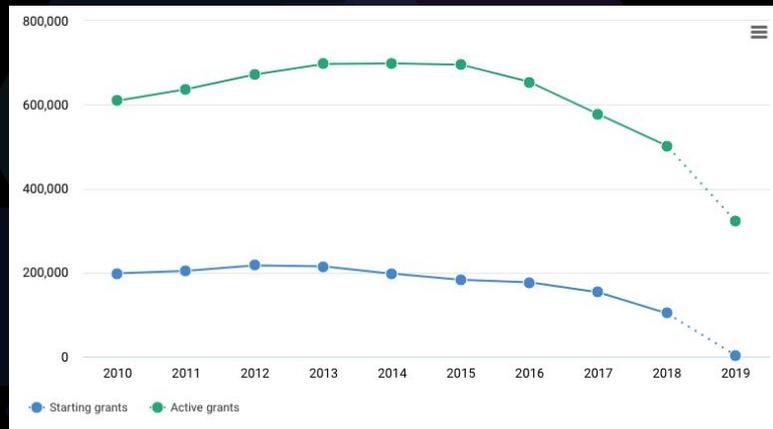


Grants data



GRANTS

- 4.3M grants
- ~300 funders globally
- Related information:
 - Funding amounts
 - Abstracts
 - Location
 - Dates
 - Related publications
 - Subject area



Patent data



PATENTS

- 37M patents
- Related information:
 - Countries
 - Abstracts
 - Subjects
 - Funders
 - Full-text

	0912 Materials Engineering	0801 Artificial Intelligence and Im...	0903 Biomedical Engineering	0906 Electrical and Electronic Eng...
Samsung (South Korea)	56,446	55,939	15,855	48,878
BM (United States)	42,819	59,902	8,572	18,897
Siemens (Germany)	37,830	27,140	26,178	40,492
Panasonic (Japan)	45,047	33,343	12,727	44,866
Robert Bosch (Germany)	25,719	18,576	26,985	36,361
LG Corporation (South Korea)	29,859	21,327	7,618	19,762
Canon (Japan)	25,984	61,900	10,886	17,408
Sony (Japan)	19,045	48,374	6,832	31,739
Philips (Netherlands)	25,313	31,717	15,875	32,945
Hitachi (Japan)	34,563	22,371	9,176	25,523
General Electric (United States)	32,116	16,192	7,309	19,805
Toshiba (Japan)	37,635	22,979	6,197	26,893
Qualcomm (United States)	3,805	19,788	3,022	23,729
NEC Corporation (Japan)	17,632	14,097	3,123	26,512
Fujitsu (Japan)	17,530	15,672	3,982	22,416
Microsoft (United States)	1,260	44,241	2,426	3,772
Mitsubishi Electric (Japan)	20,394	12,754	6,842	26,269
Ericsson (Sweden)	1,831	6,494	1,463	14,131
Intel (United States)	9,775	18,850	4,237	16,104
Toyota Motor Corporation (Japan)	12,618	6,065	8,739	19,702

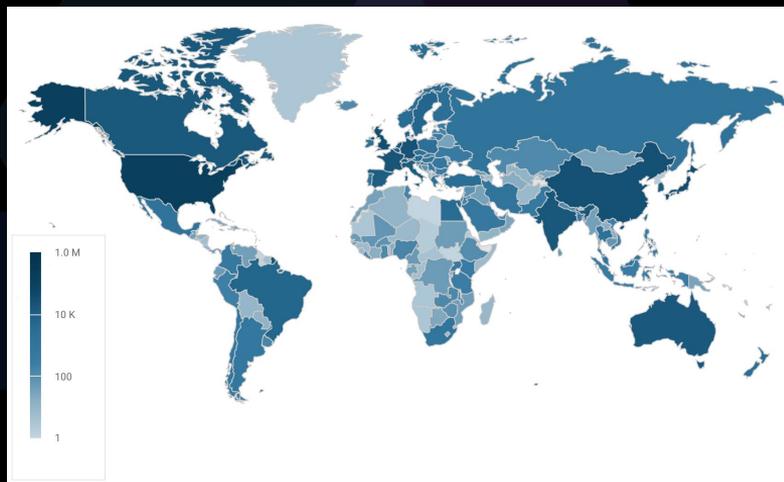
Source: <https://app.dimensions.ai>
Exported: January 23, 2019
Criteria: none.

Clinical Trial data



CLINICAL TRIALS

- More than 440k trials
- Related information:
 - Abstracts
 - Related grants
 - Countries
 - Conditions
 - Subject areas
 - Sponsors/collaborators

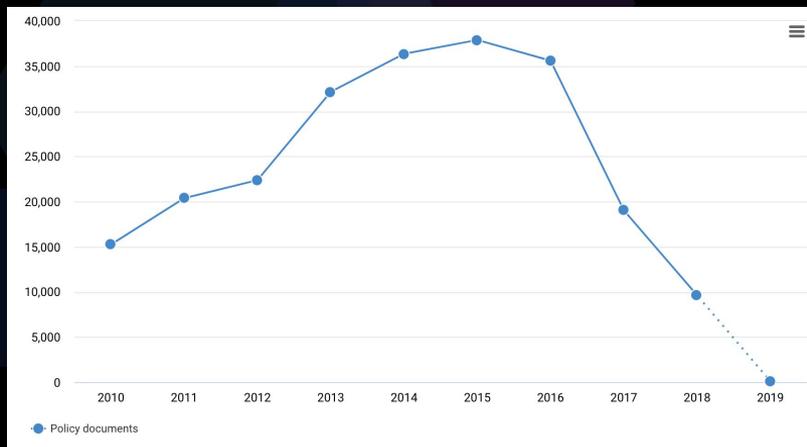


Policy Documents data

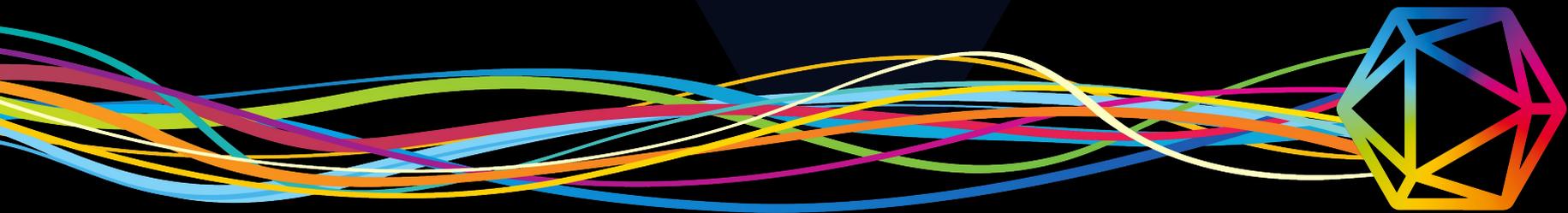


POLICY DOCUMENTS

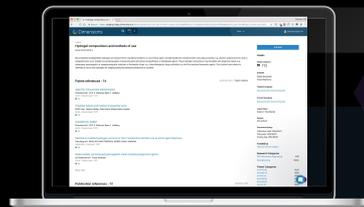
- Over 384,000 policy document records, linked to publication records
- Related information:
 - Subject area
 - Organization (GRID ID)
 - Country
 - Full-text
 - References to publications



Relationships

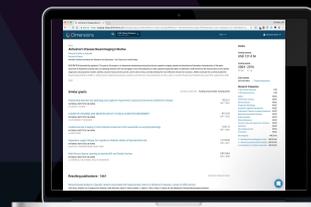


Links between the different content sources



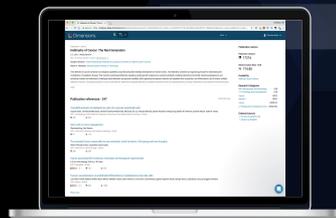
Patents

- Patent references
- Publication references
- Supporting grants
- Patent citations



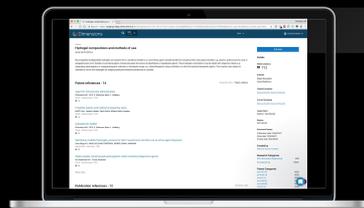
Grants

- Resulting publications
- Resulting patents
- Resulting clinical trials



Publications

- Publication references
- Publication citations
- Supporting grants
- Patent citations
- Linked clinical trials



Policy Documents

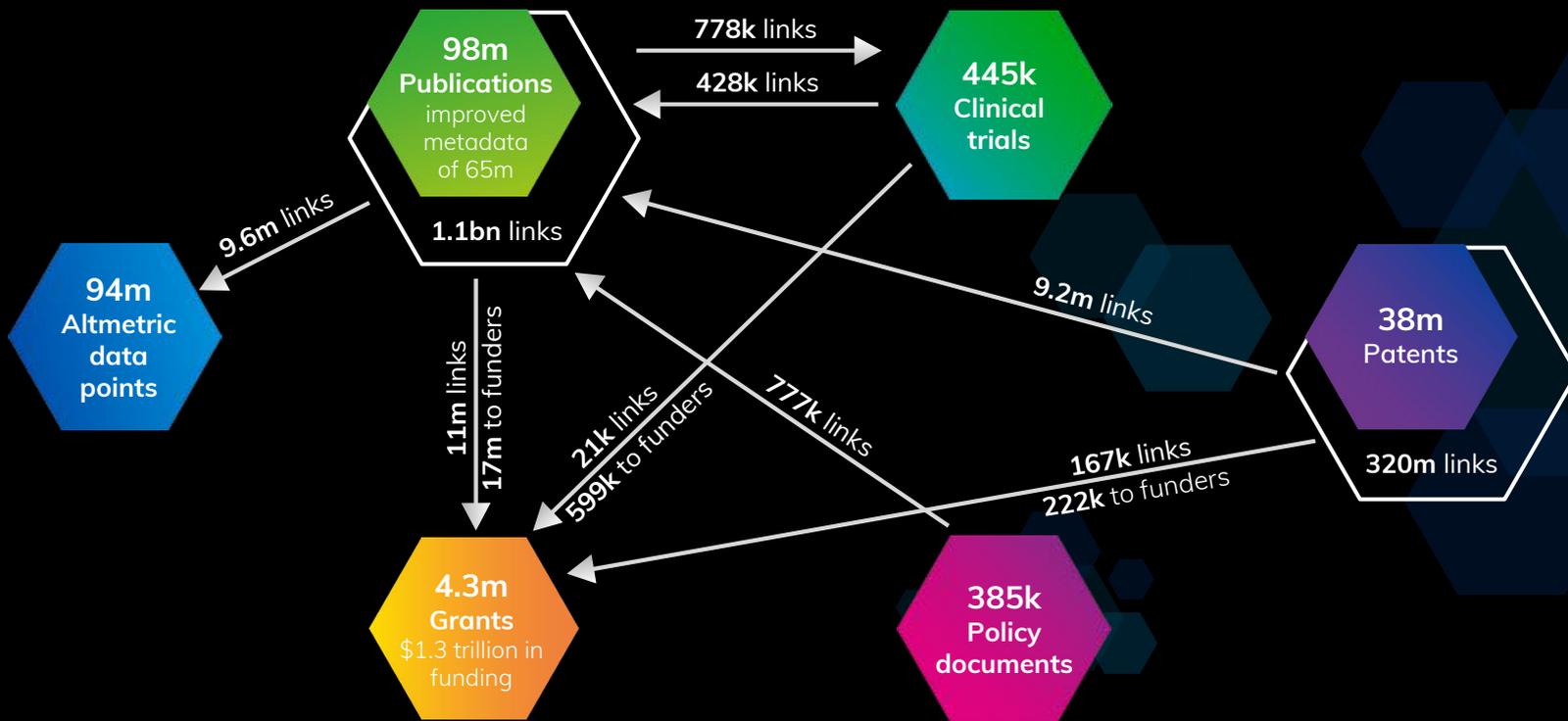
- Publication references



Clinical trials

- Linked publications
- Supporting grants

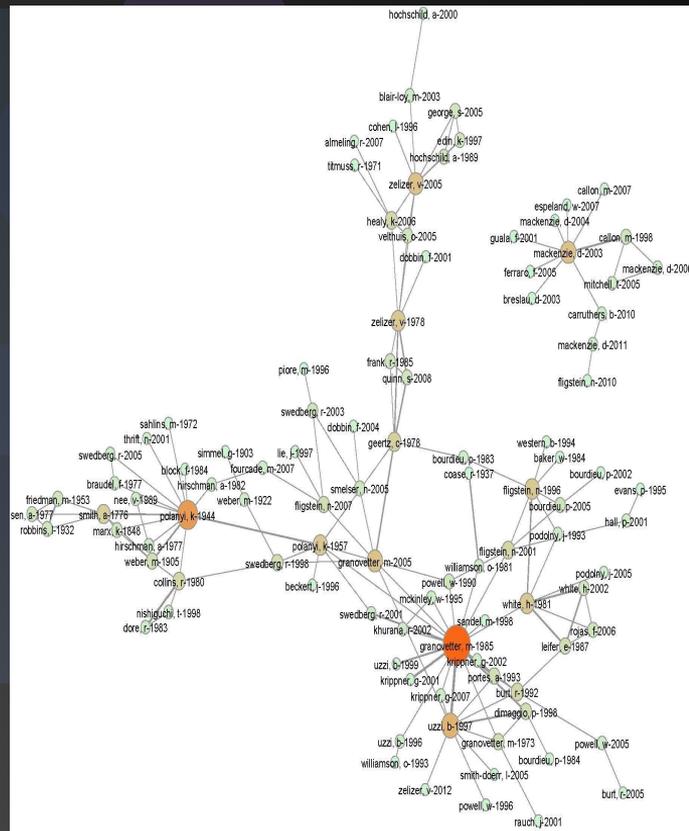
The data and links driving Dimensions...



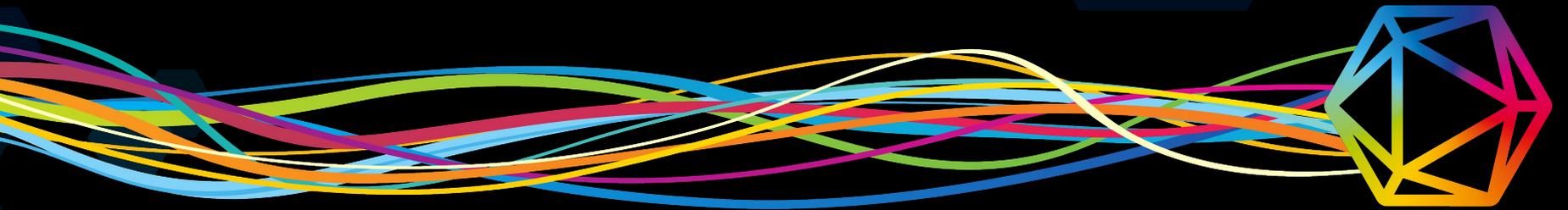
Resulting citation graph

1+ billion references resolved

- We are still working to resolve more references
- The planned addition of new content will lead to even higher references numbers



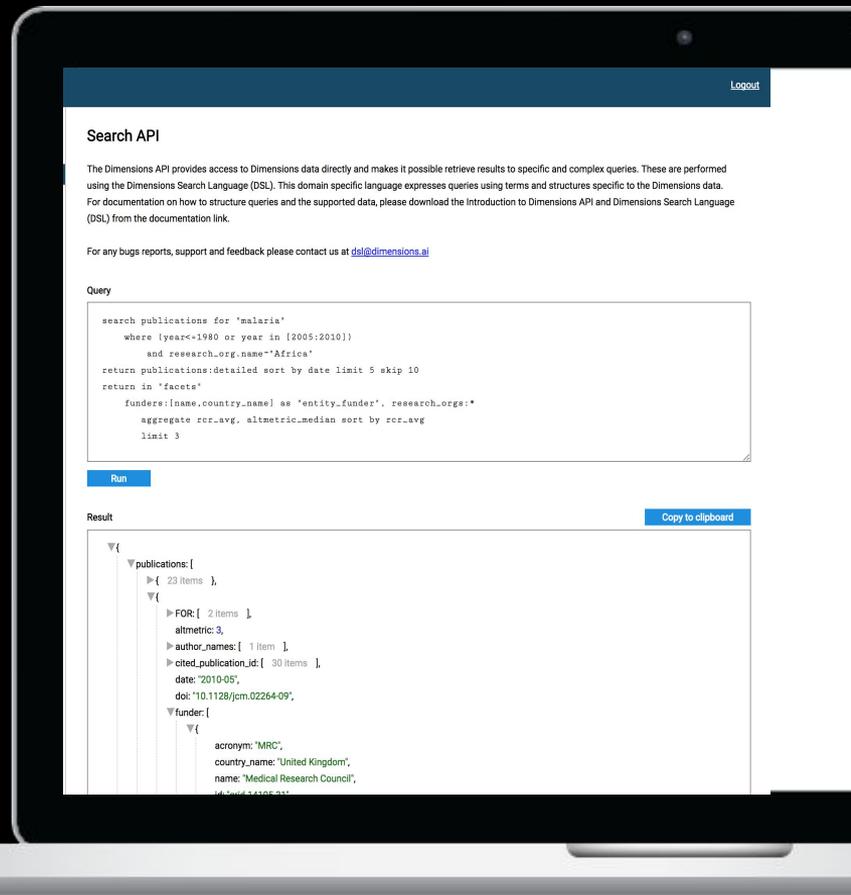
Using the API for research



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



Dimensions Search Language (DSL)

Powerful query language developed around simple syntax, allowing users with various levels of technical skills to use.

- Basic query based around two phrases
- More complexity can be included by adding filters
- Returned data can be specified using custom fieldsets

```
search publications
return publications
```

Dimensions Search Language (DSL)

Powerful query language developed around simple syntax, allowing users with various levels of technical skills to use.

- Basic query based around two phrases
- More complexity can be included by adding filters
- Returned data can be specified using custom fieldsets

```
search publications for "leukemia"  
where year in [2015:2017]  
return publications
```

Dimensions Search Language (DSL)

Powerful query language developed around simple syntax, allowing users with various levels of technical skills to use.

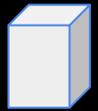
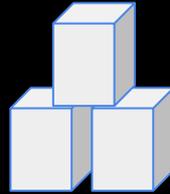
- Basic query based around two phrases
- More complexity can be included by adding filters
- Returned data can be specified using custom fieldsets

```
search publications for "leukemia"  
where year in [2015:2017]  
return publications [title+doi]
```

Dimensions API takes an analytical approach to a data service

Standard Approach

API call



Dimensions API takes an analytical approach to a data service

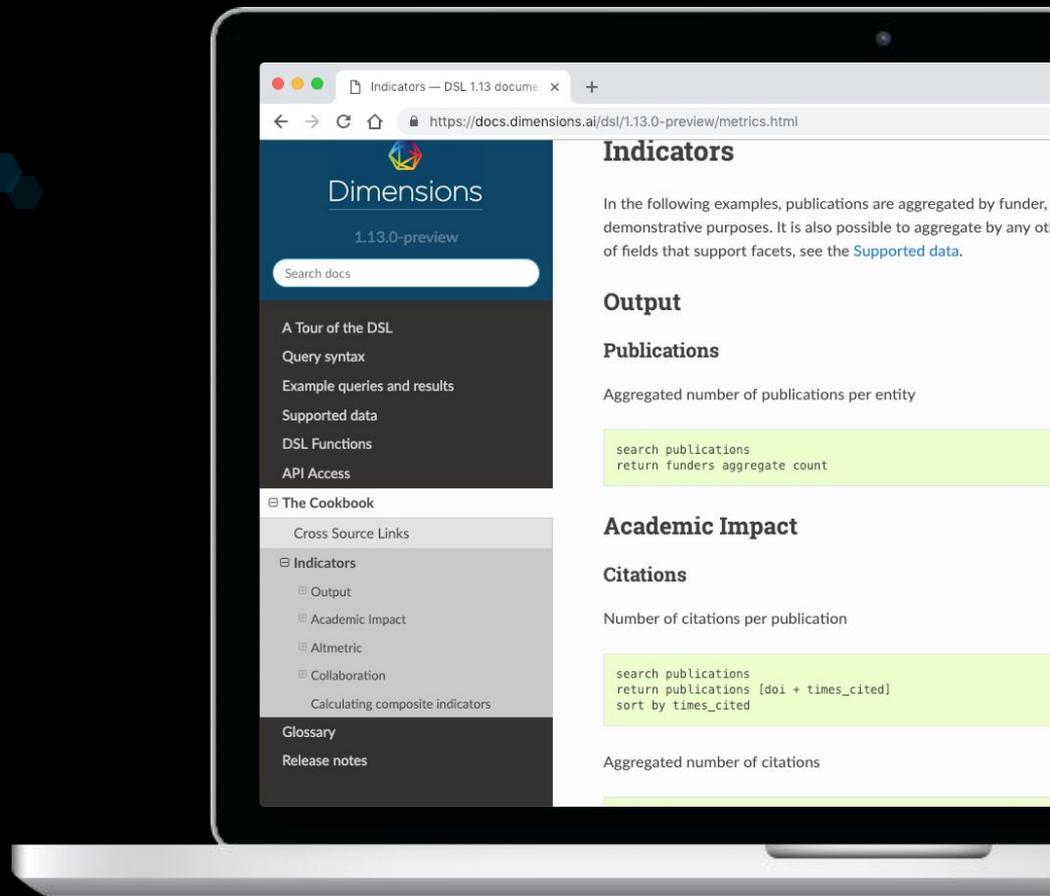
Analytical Approach



API documentation

The Dimensions API has extensive documentation, containing examples, tutorials and supported data tables

<https://docs.dimensions.ai/ds/latest/>



Altmetric data overview



Huge amounts of data!

12 million+
outputs w/ attention

1.6 million+
Wikipedia
mentions

1.4 million
policy
references

9.7
million
patent
citations

4.6 million
news mentions

97 million+
mentions since
2011



Scope of our data

What we track

Online discussions
around research
outputs of all kinds!

We do not track

Social media metrics
Usage statistics
Concepts*
Author names*



For research to be tracked, we need...



A research output
with a...



Persistent identifier
that is...



Mentioned in a
source we track



Research outputs



Articles & preprints

Books

Book chapters

Datasets

Clinical trial records

News stories

*...and countless other research formats
if a persistent identifier has been assigned!*



Video data on figshare: [10.6084/m9.figshare.5721088.v1](https://doi.org/10.6084/m9.figshare.5721088.v1)



Identifiers



- DOIs
- PubMed IDs
- ISBNs
- Handles
- arXiv IDs
- ADS IDs
- URNs
- SSRN IDs
- RePEC IDs
- ClinicalTrials.gov records
- URLs
- URIs
- ORCID identifiers



Data sources



17 types of platforms, thousands of sites indexed

- Policy documents
- News
- Blogs
- Twitter
- Post-publication peer-reviews
- Facebook
- Sina Weibo
- Syllabi
- Wikipedia
- Google+
- LinkedIn
- Reddit
- Faculty1000
- Q&A (Stack Overflow)
- Youtube
- Pinterest
- Patents



Altmetric attention score



Score based on:

- Volume
- Sources
- Authors



Follow a list of domains representing thousands of websites



Search for links to those domains in attention sources.



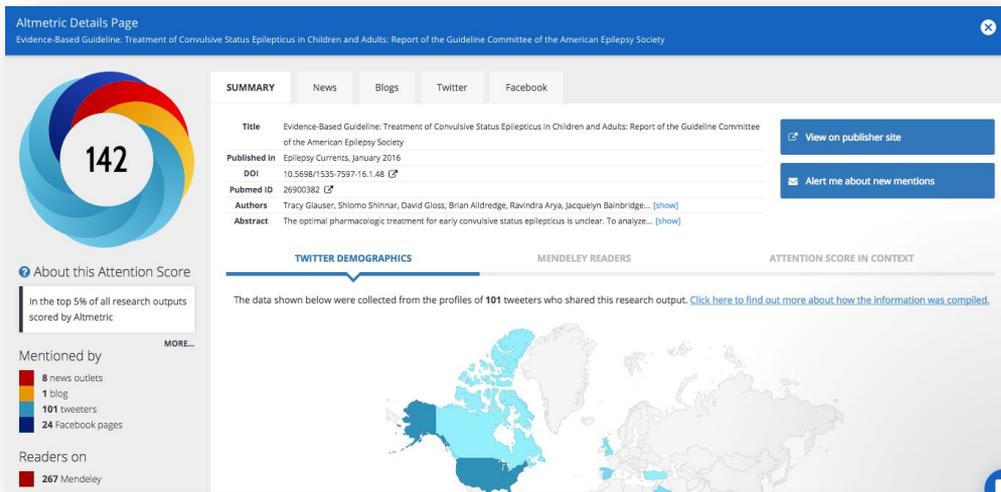
Collate attention.



EXCEPT news, policy docs, and syllabi



Display data in Altmetric Details Pages.



Details Page API - Counts only

```
{  
  "title": "Protein hormone boosts memory",  
  "doi": "10.1038/news.2011.49",  
  "nlmid": "0410462",  
  "journal": "Nature News",  
  "altmetric id": 218594,  
  "cited by fbwalls count": 1,  
  "cited by posts count": 2,  
  "cited by rdts_count": 1,  
  "score": 0.5,  
  "history": {  
    "1d": 0,  
    ...  
    "1y": 0.5,  
    "at": 0.5  
  },  
  "url": "http://www.nature.com/news/2011/110126/full/  
news.2011.49.html",  
  "subjects": [  
    "science"  
  ],  
  "readers": {  
    "citeulike": "0",  
    "mendeley": "1",  
  }  
}
```

Unlimited no. of calls

- **Timeframe** (fixed, e.g. previous 24h, one year, all time)
- "Mentioned in": Reddit, news, etc.
- Output identifier - 1/call

Continuously updated



Working with JSON APIs

Using Python and Jupyter notebooks

Python

Libraries

```
import carRepair
```

Modules

```
from carRepair import diagnostics
```

Variables

```
oil = 'Valvoline 28'
```

Functions

```
def changeOil {  
    if oilAge > 6:  
        do something here  
}
```

JSON

key-value pairs



```
{  
  "name": "Tina Turner",  
  "age": 8,  
  "location": "Minneapolis",  
  "cute": True  
}
```

JSON: complex data types

```
nested objects {  
  "Publication":  
  {  
    "Title": "Altmetrics manifesto",  
    "Authors": "J. Priem, D. Taraborelli, P. Groth, C. Neylon",  
    "Publication date": "October 2010"  
  }  
}
```

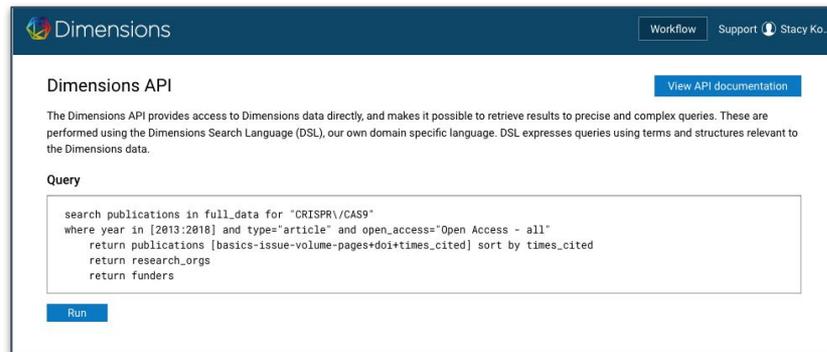
JSON: complex data types

nested arrays

```
{“Title”: “Altmetrics manifesto”,  
  “Authors”:  
  [{ “name”: “. Priem”,  
     “Org”: “Impactstory”  
  },  
  { “name”: “. Taraborelli”,  
     “Org”: “Wikimedia Fondation”  
  }],  
  “Publication date”: “October 2010”}
```

Dimensions JSON API

Query the API using DSL
Returns data in JSON



<https://docs.dimensions.ai/dsl/index.html>

Check out “Supported Data” and Cookbook/cross-source links

Altmetric JSON API

Query the API repeatedly using Python, by DOI
Returns [data](#) in JSON

http://api.altmetric.com/docs/call_citations.html

Check out “Response object” to understand what you get back

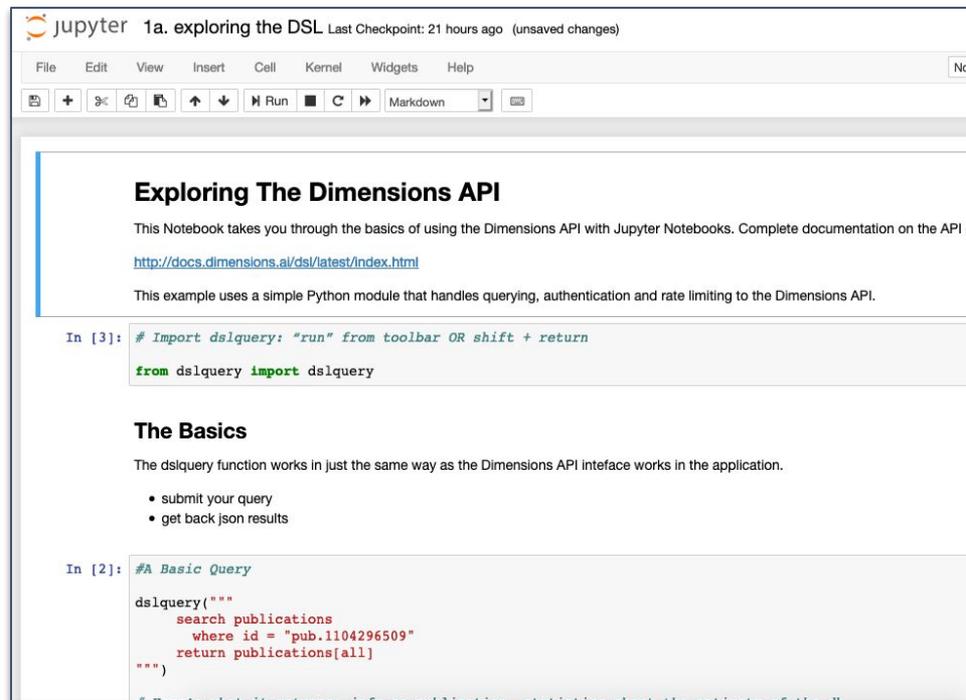
Jupyter notebooks

Exploring the DSL API

Exploring the Altmetric API

Manipulating results from
Dimensions

Querying grants, publications,
clinical trials, patents



jupyter 1a. exploring the DSL Last Checkpoint: 21 hours ago (unsaved changes)

File Edit View Insert Cell Kernel Widgets Help

Run C Markdown

Exploring The Dimensions API

This Notebook takes you through the basics of using the Dimensions API with Jupyter Notebooks. Complete documentation on the API <http://docs.dimensions.ai/dsl/latest/index.html>

This example uses a simple Python module that handles querying, authentication and rate limiting to the Dimensions API.

```
In [3]: # Import dslquery: "run" from toolbar OR shift + return
from dslquery import dslquery
```

The Basics

The dslquery function works in just the same way as the Dimensions API interface works in the application.

- submit your query
- get back json results

```
In [2]: #A Basic Query
dslquery("""
    search publications
    where id = "pub.1104296509"
    return publications[all]
""")
```

Retrieving and developing metrics for reporting

Guided Exercise

Visualizing research information, citations, and altmetrics

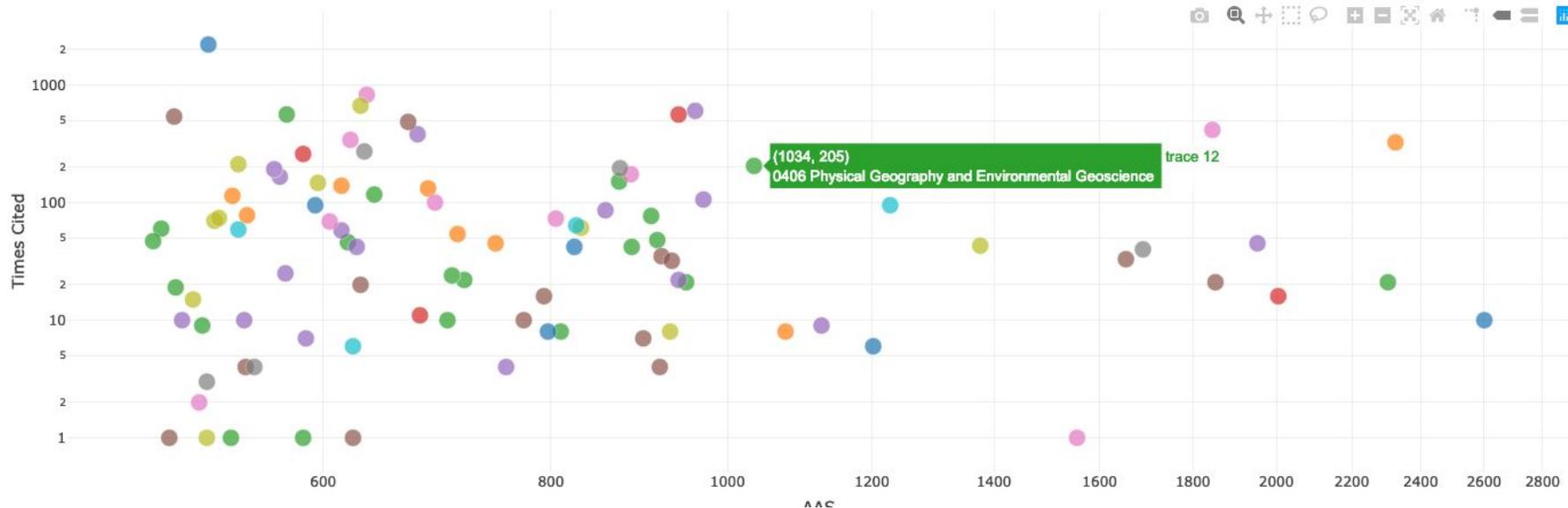
Guided Exercise

Porting research information to a website

Guided Exercise

UT Austin: 2009-2019

Here is a log normalized scatterplot showcasing UT Austin publications with the highest attention and citations, by subject area.





Thanks



Keep in touch
stacy@altmetric.com

Jupyter notebooks: <https://figshare.com/s/6b087b2b3bc3155308b6>