AAUDE Annual Meeting 2021

Technology to Uncover the Academic Story

A presentation of emerging technology for institutional analytics

DIGITAL science

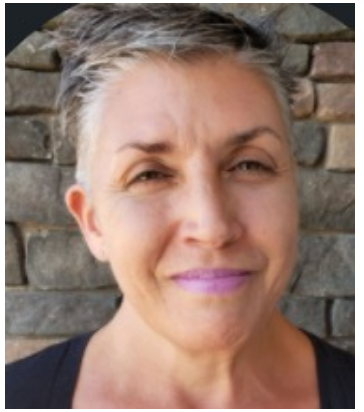
Agenda

- Welcome and introductions
- Building a progressive data landscape at a Research Institution
- How Dimensions provides innovative options
  The Data You Need, The Way You Want It
- Q&A

DIGITAL science

Colleagues Joining

Kelsey Rosell
Sr Vice President and Commercial Director, Dimensions
k.rosell@digital-science.com



Liz Tomich
Sr Solutions Account Manager, Digital Science
l.tomich@digital-science.com

Isabel Thompson
Head of Data Platform, Dimensions
i.thompson@digital-science.com

Sunset at the Pacific, San Diego, CA

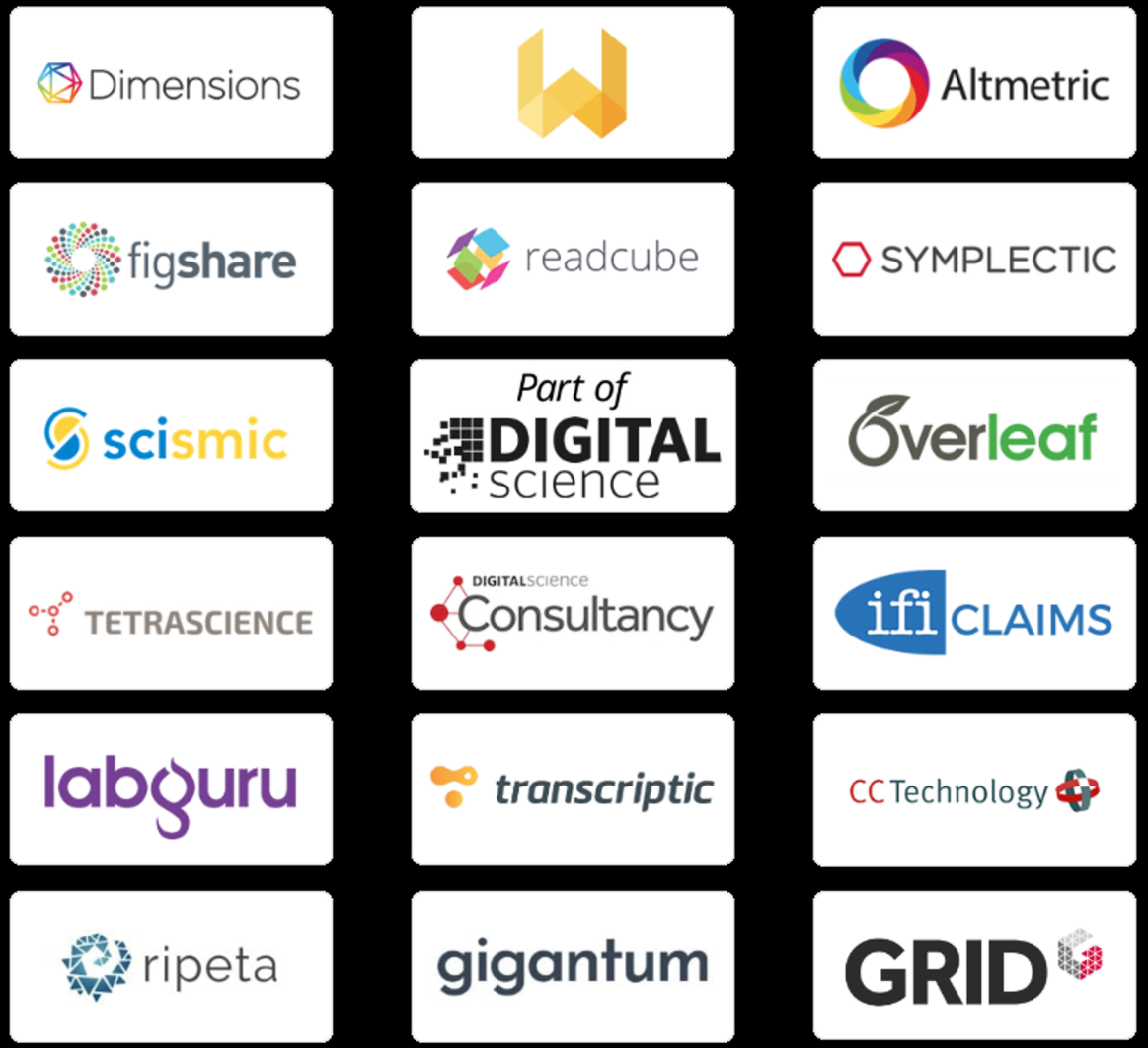
Dimensions is part of Digital Science

Digital Science **invests**, **supports** and **nurtures** small innovative software companies.

Rooted in research, most founders have an academic research background and started a company around solving a self-encountered problem.

Since 2009, Digital Science has invested in and/or started numerous companies with nearly 300 colleagues to call on from across the Digital Science portfolio.

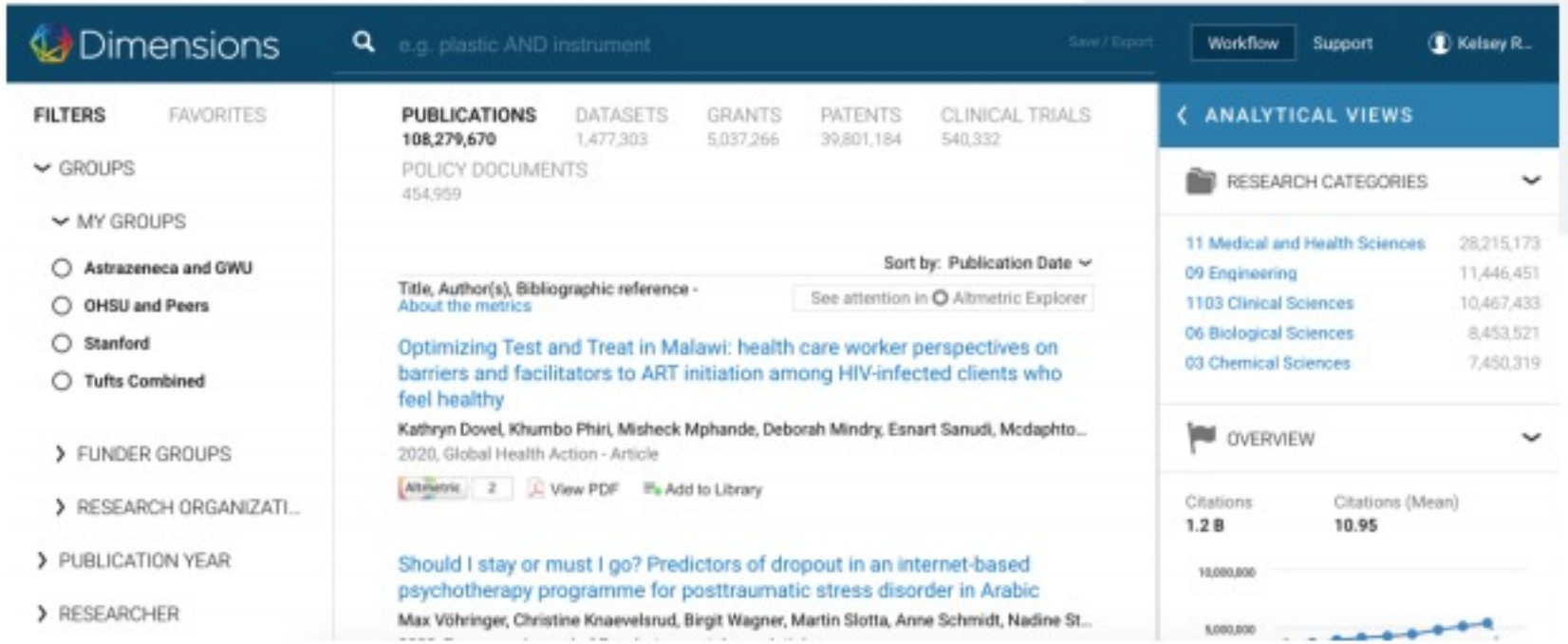
Making the research process more open, efficient and effective…



What is Dimensions?

A linked, categorized and disambiguated database of:

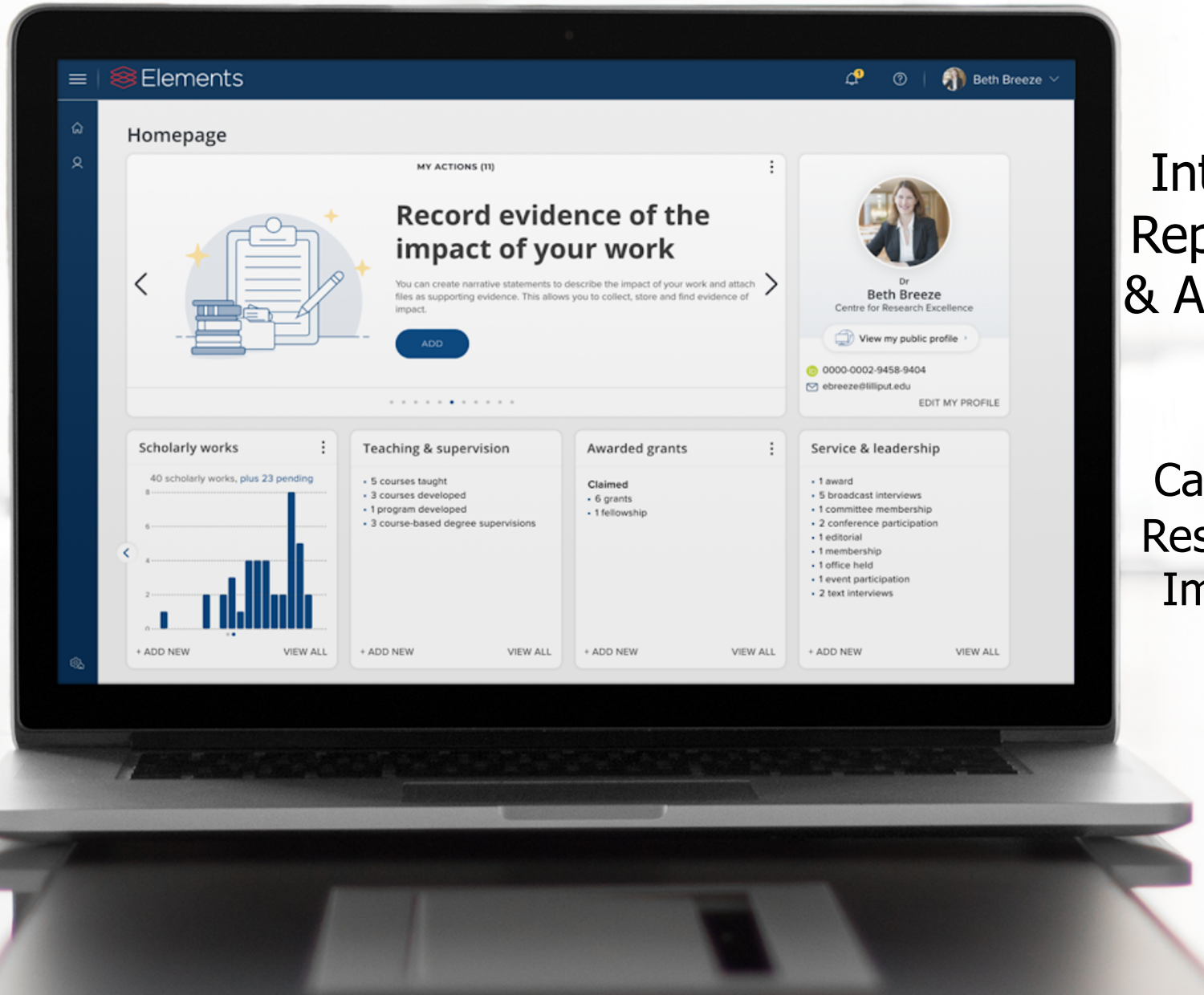
- Publications
- Awarded Grants
- Patents
- Clinical Trials
- Policy Documents
- Data sets



Dimensions

Part of DIGITALscience

Symplectic Elements



Internal Reporting & Analysis

Record Research Outputs

Populate Public Profiles

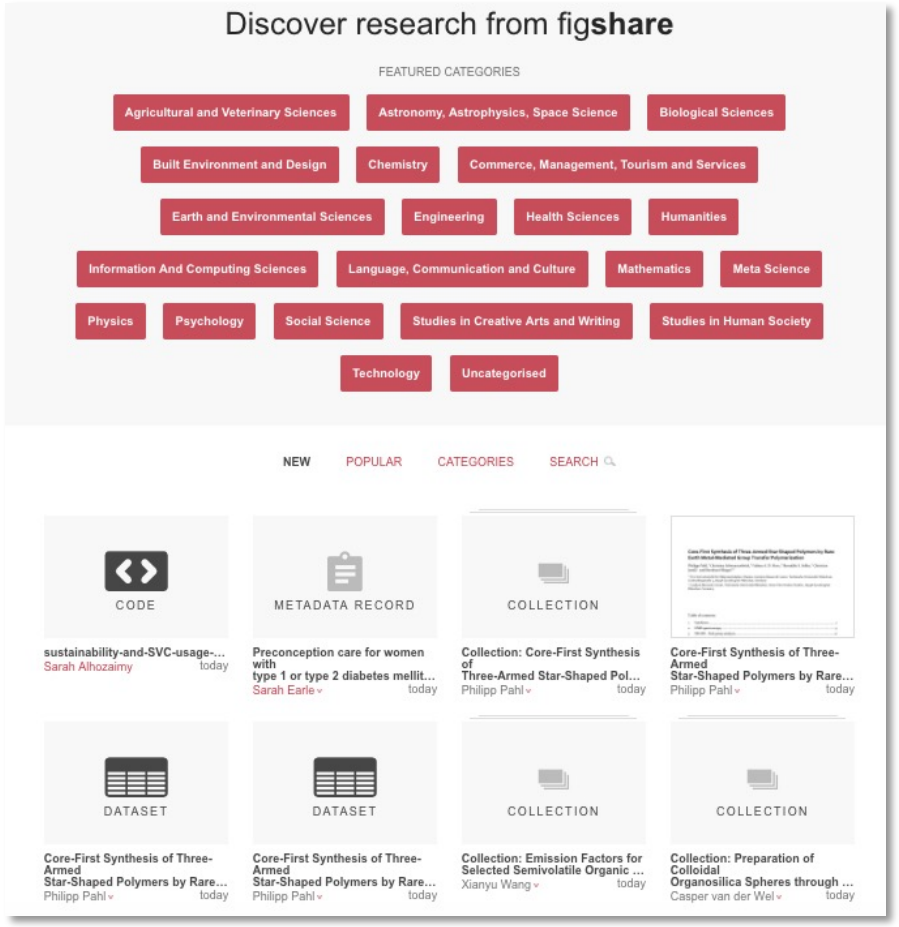
Capture Research Impact

Support Open Access

Annual Reviews & Assessment

(e.g. REF/Annual Faculty Review/PBRF/ ERA)

figshare: store, share, discover research



Who we are and what we do:

Figshare is an online digital repository where researchers can preserve and share their research outputs in any format, including (but not limited to) figures, tabular data, code, and preprints. Data can be published under licenses that promote reuse and remixing of published content, with every public item on the platform receiving a Digital Object Identifier (DOI) to aid in the discover citation of published content.

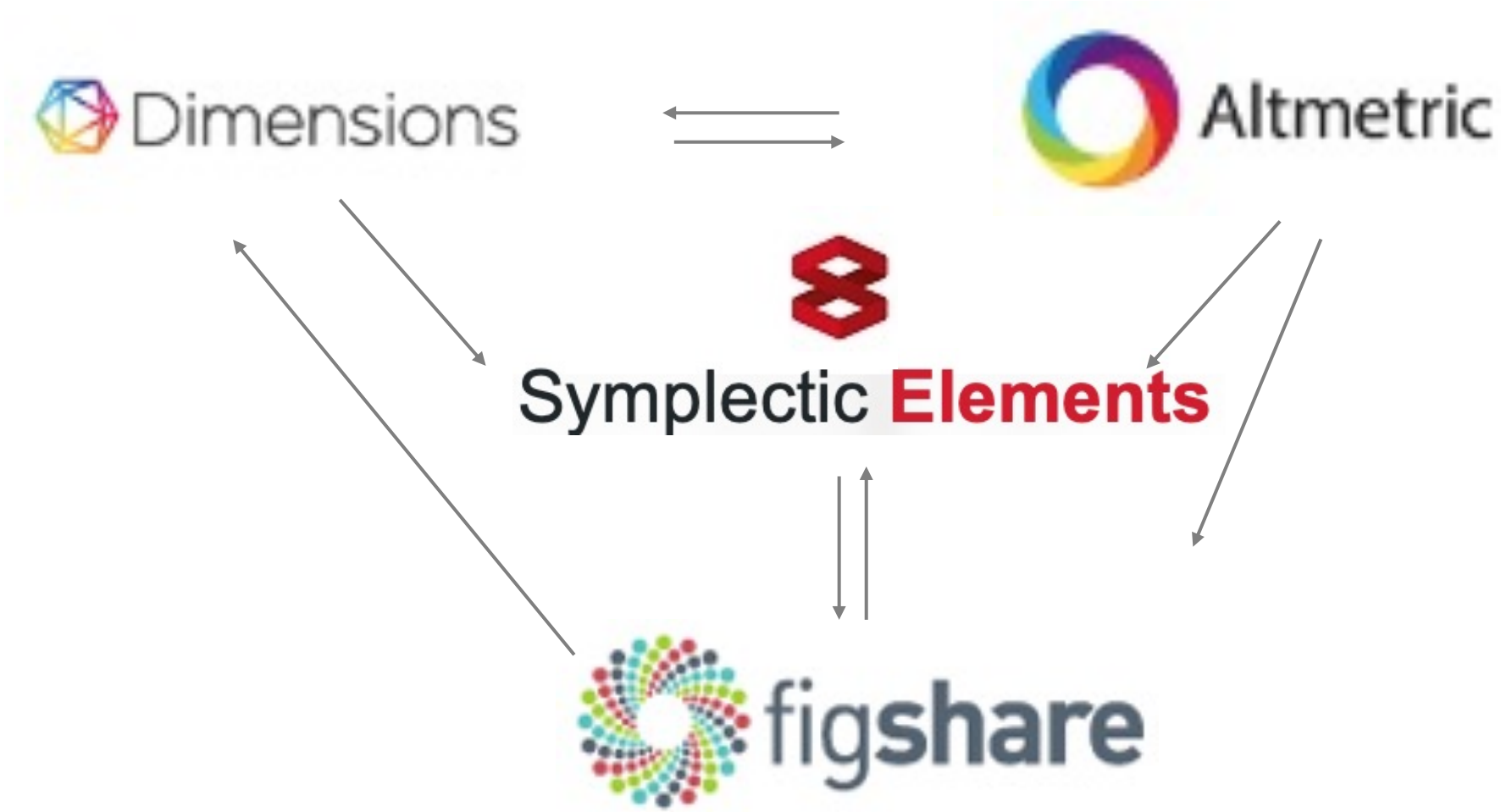
Altmetric Attention Sources

The Colors of the Donut

- 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

17 categories representing thousands of websites

Even better together…



AAUDE Annual Meeting 2021

Progressive Data Solutions at CU Boulder

Liz Tomich, Sr Solutions Account Manager, Digital Science
Supporting Academia in the US and Canada

About CU Boulder

University of Colorado **Boulder**

Partners at CU Boulder

- **Academic Affairs**
- **VC for Research**
- **Libraries**
- **Institutional Research**
- **Strategic Relations and Communicators**
- **Research Institutes and other early adopters**
- **Campus Power Users (units, chairs, staff)**
- **FIS Technical Team**

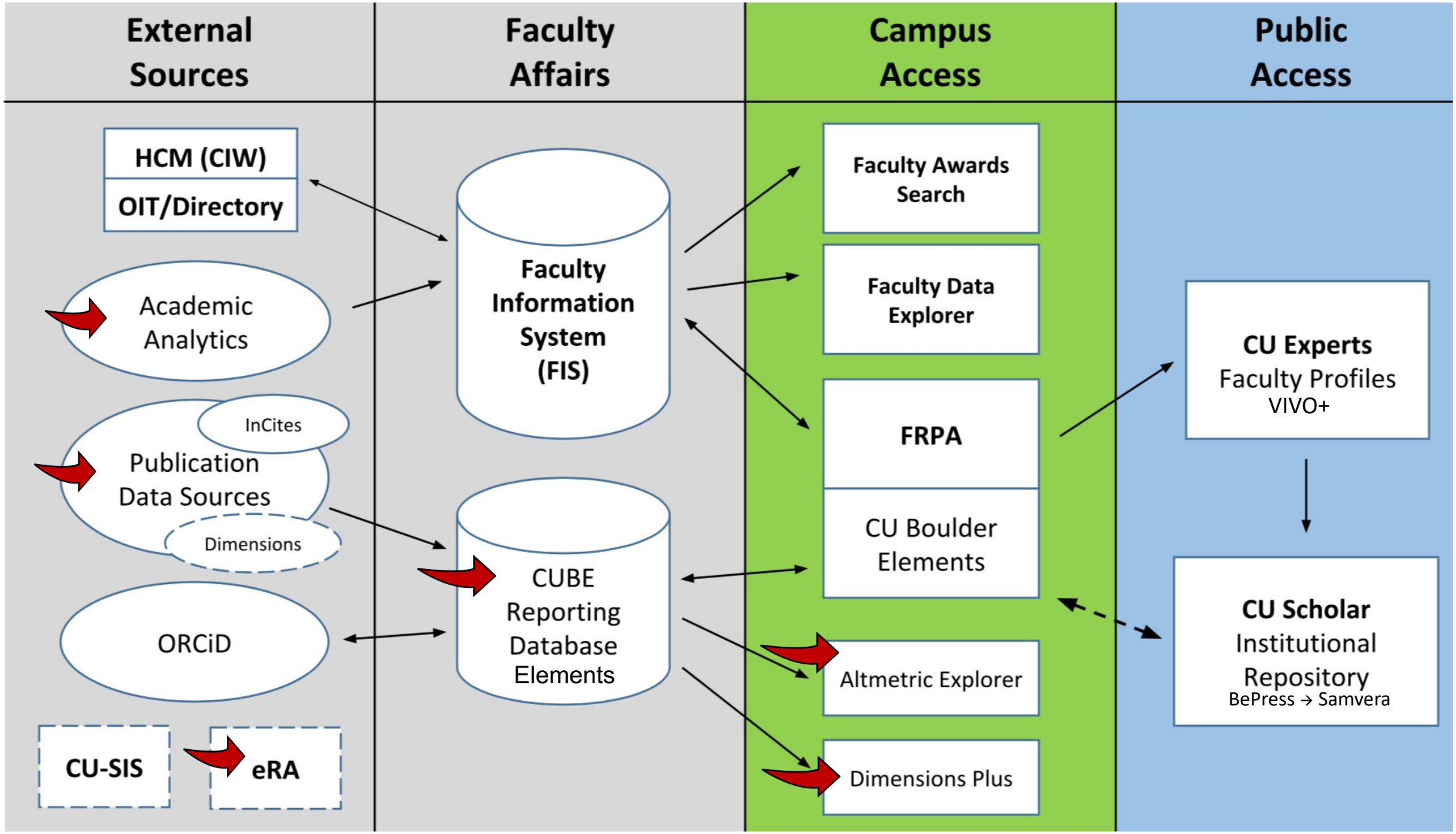
Campus Values - What droves decisions?

- 1. All faculty disciplines – Science *and* Creative Works
- 2. Diversity
- 3. Global connections – International partners and recruiting
- 4. Research compliance
- 5. Creative solutions for faculty – to save time
- 6. Faculty reputation and campus prestige

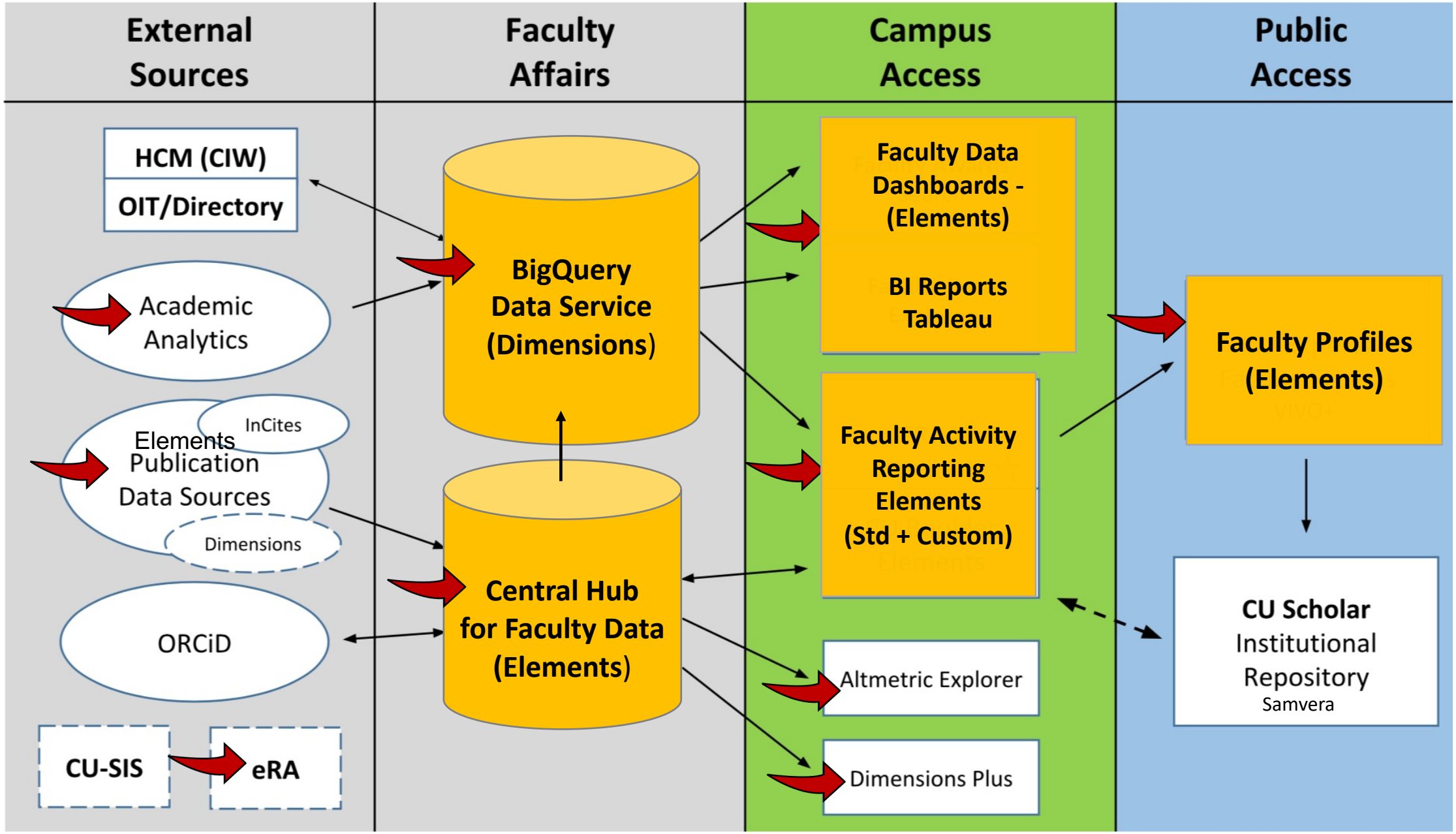
Campus Values - What drove decisions?

- 7. Measuring impact of scholarly work for the public good
- 8. Attracting superstar faculty and keeping them
- 9. Being innovative
- 10. Compliance - funding laws and Regent policy
- 11. Creating new revenue streams
- 12. Donors and fundraising
- 13. Who does what research?
- 14. Strategic Imperatives

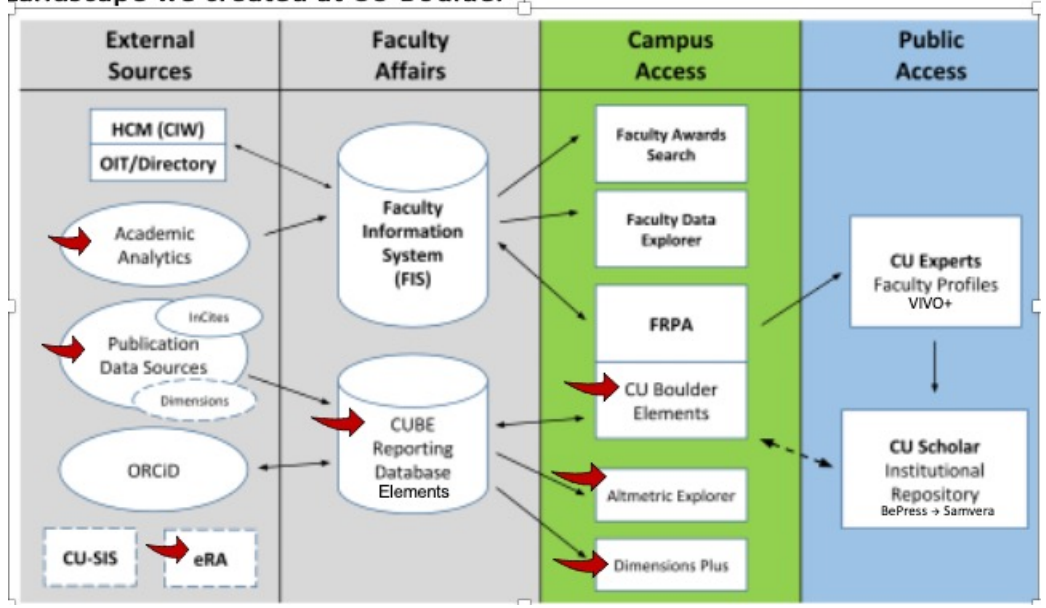
Landscape we created at CU Boulder



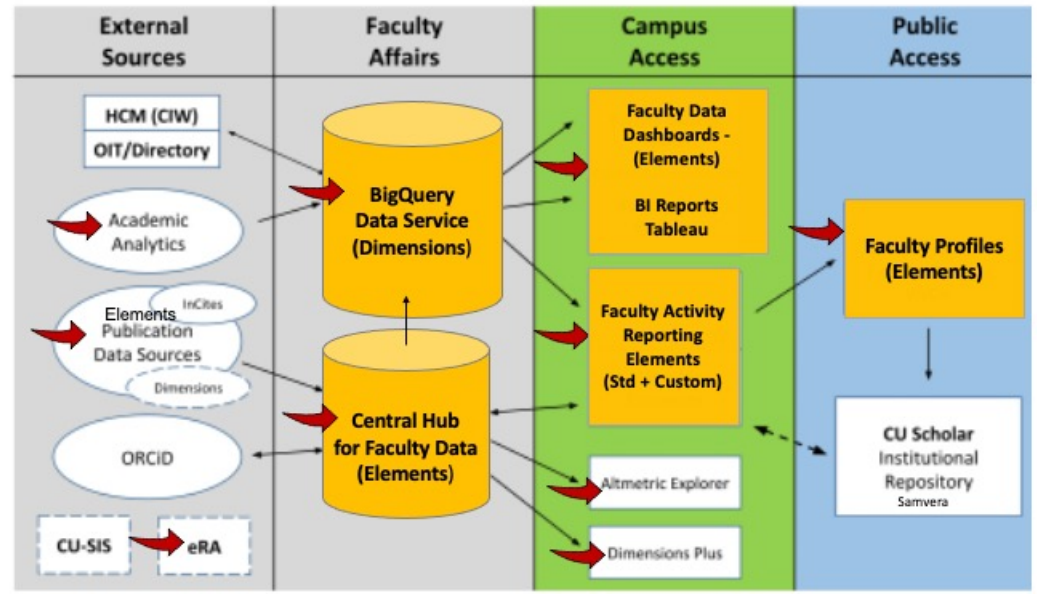
What if starting today? New Possibilities



Landscape we created at CU Boulder



What if starting today? New Possibilities



Dimensions meets needs for the entire campus

Chancellery, Deans & Planning

Strategic planning & competitive intelligence
- Peer group analysis & comparisons
- Internal benchmarking
- Cost-effectiveness of research
- Horizon scanning

Advocacy
- Impact, value of institutions & research

Talent planning & retention

Networks & context
- Industry collaboration & revenue generation

Research Office

Rankings & assessment
- National assessment
- International ranking
- Custom benchmarking

Impact assessment and narrative

Funding & grants
- Strategy
- Sources (incl. industry)
- Trends
- ROI assessment

Talent
- Recruitment & capability building
- Career tracking

Collaboration networks

Library

Collection Development
- Holdings & ROI assessment
- Publisher negotiations & transformative agreements

Open access
- Trends
- Citation and impact analysis
- Compliance tracking

Publication strategy
- Support researchers on where to publish

Information hub
- High value data asset for training

Faculty & Research

Strategy
- Funding trends & opportunities
- Faculty output tracking
- Find partners and recruit talent
- Emerging research trends
- Patent analysis

Bibliometrics
- Co-citation analysis

Use for research
- e.g. Computer Science, Business, Economics, Social Science
- Dataset for analysis and visualisation

And Beyond

Commercialization of research
- Science-->Innovation
- Tech transfer
- Support university spinouts & deep tech
- Patent analysis
- Find Key Opinion Leaders
- Industry partnerships

Marketing & Communications
- Market segmentation & targeting
- Monitor & grow wider impact & attention, publicity for research

Check these links out for more info…

How to get started with Dimensions on Google BigQuery in 5 minutes

Upcoming webinar: May 18 Texas A&M discusses using Dimensions to track postgraduate success

Dimensions webinars on demand - various topics

CU Boulder's legacy of innovation - 2 minutes Scroll down to trailer.

Dimensions

Dimensions

The data you need, delivered the way you want it

Presented to AAUDE Members

Kelsey Rosell, Sr. Vice President, Dimensions Commercial Director
Liz Tomich, Sr. Solutions Sales Manager, Digital Science
Isabel Thompson, Head of Data Platform

May 5, 2021

Part of **DIGITAL** science

Today’s world

The research sector is more complex and sophisticated than ever before

- Impact agenda
- Rapid publishing
- Knowledge translation & tech transfer
- Global research challenges, SDGs
- Compliance

Priorities are:

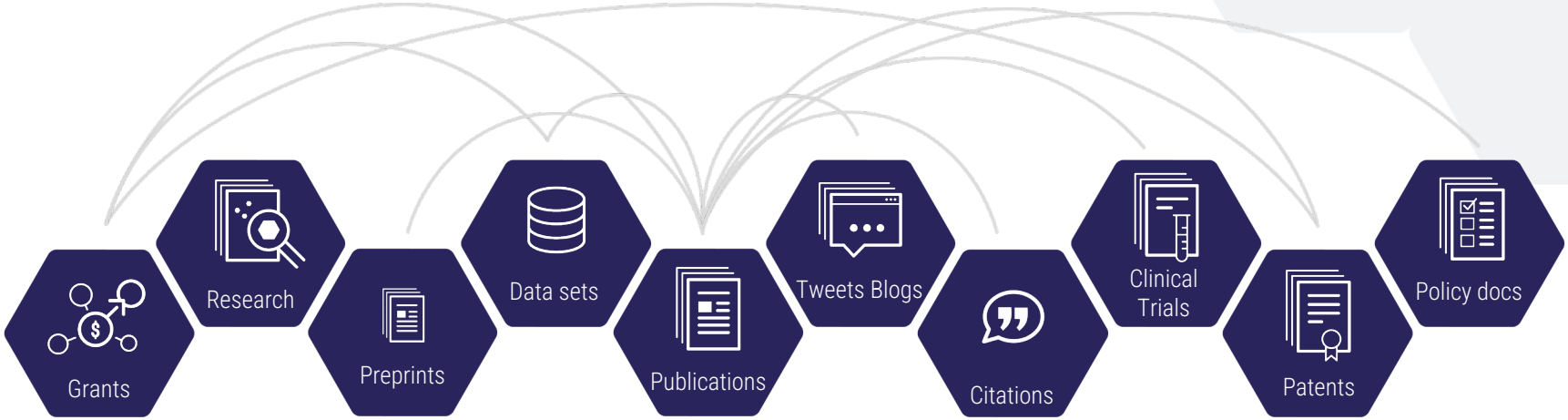
- Context around research decisions
- Value of research

Dimensions

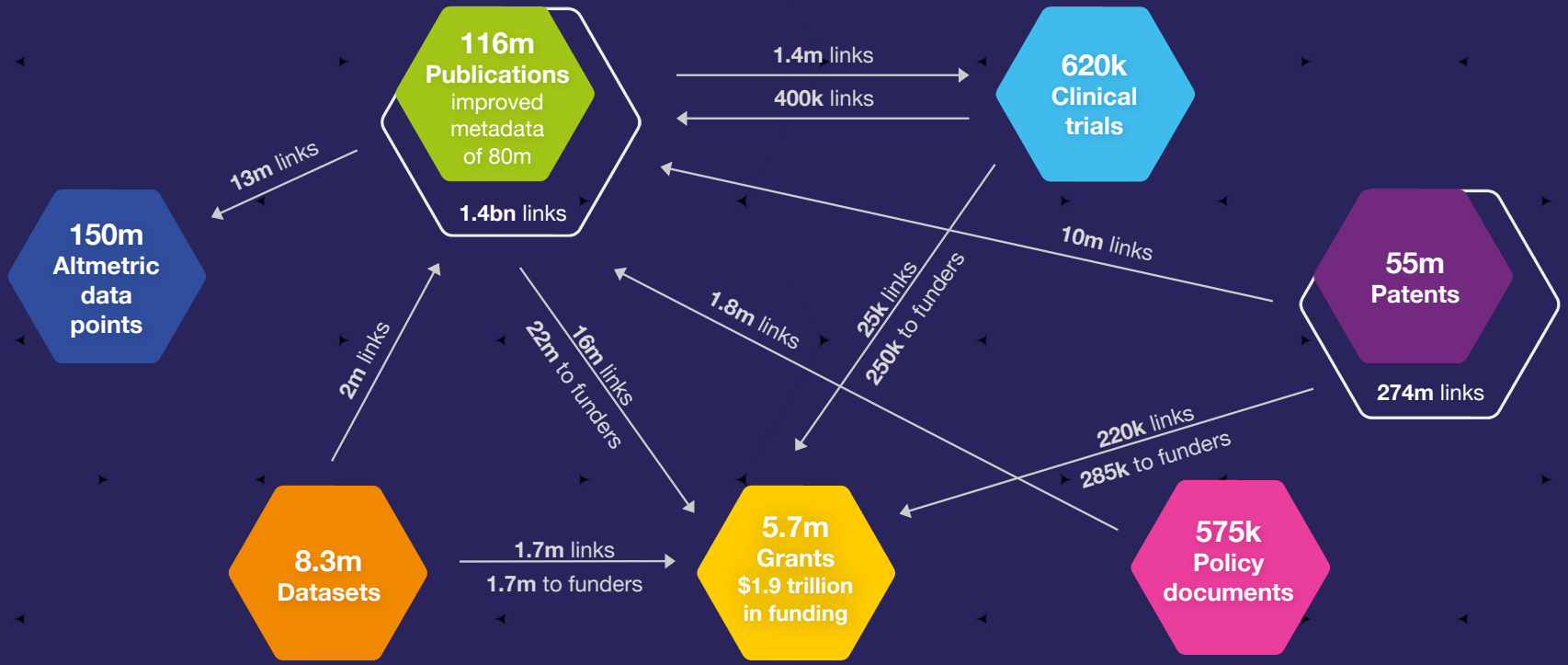
Today’s agenda

- Summary of Dimensions
- Data curation approach
- Navigating the new normal
- How you can engage with the data
- Discussion

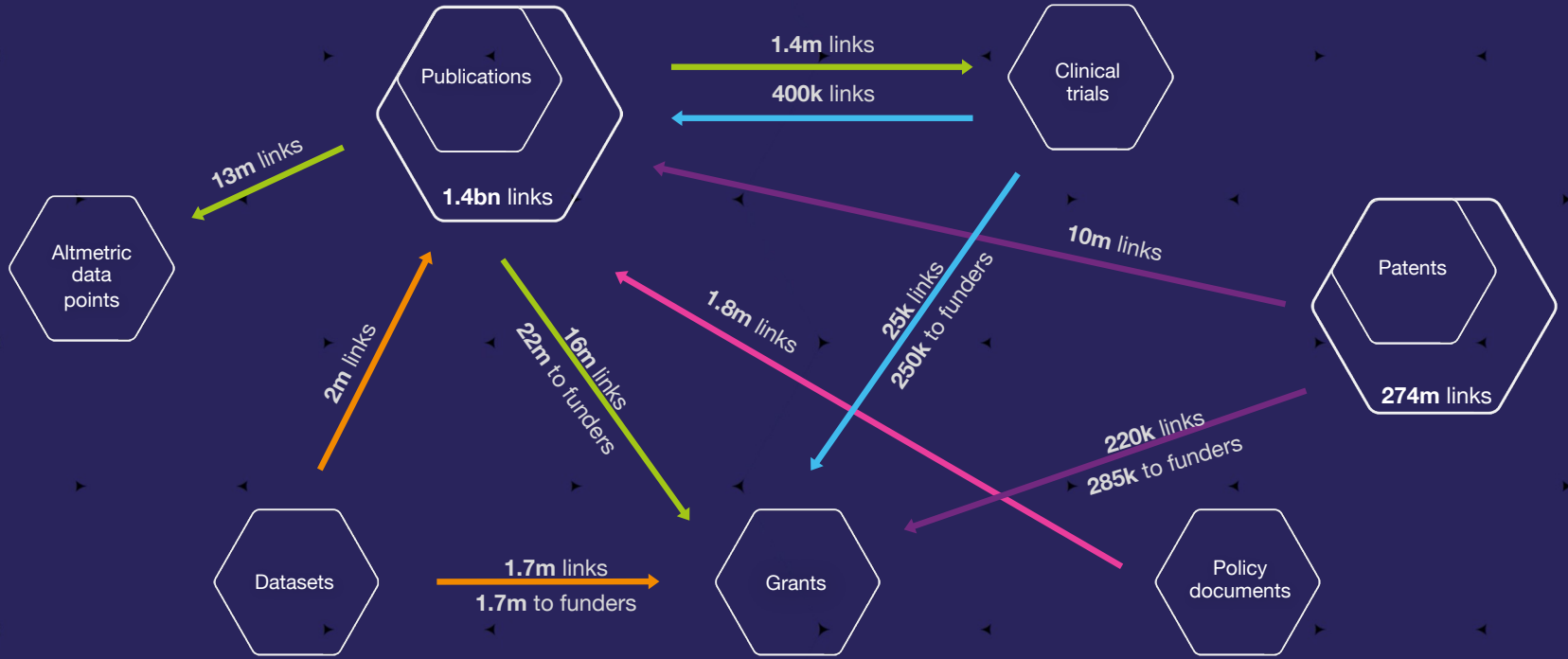
More context along the entire trajectory of research



Dimensions: A joined-up perspective on research

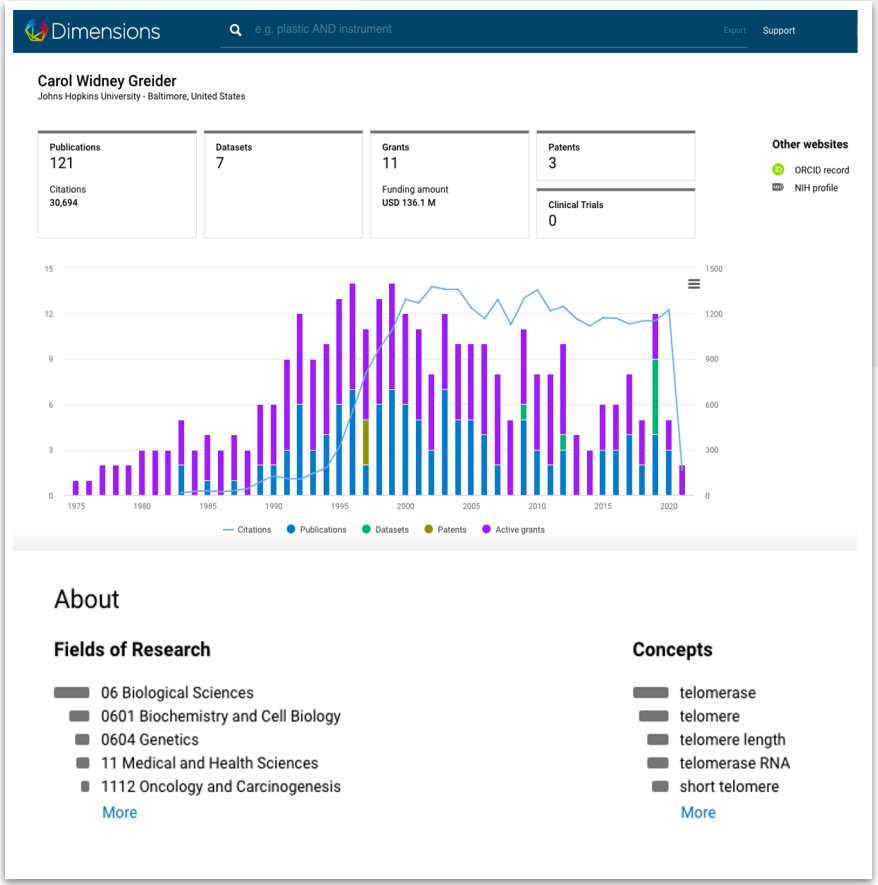


Links enable you to explore relationships



Practical, informative views - always adding more

- Dashboard: Researcher profile
- Content Type views
- Classification Category: Sustainable Development Goals
- Advanced Search incl Concept Search
- Download limits to 50k
- Journal list
- Open Access
- Funders etc.



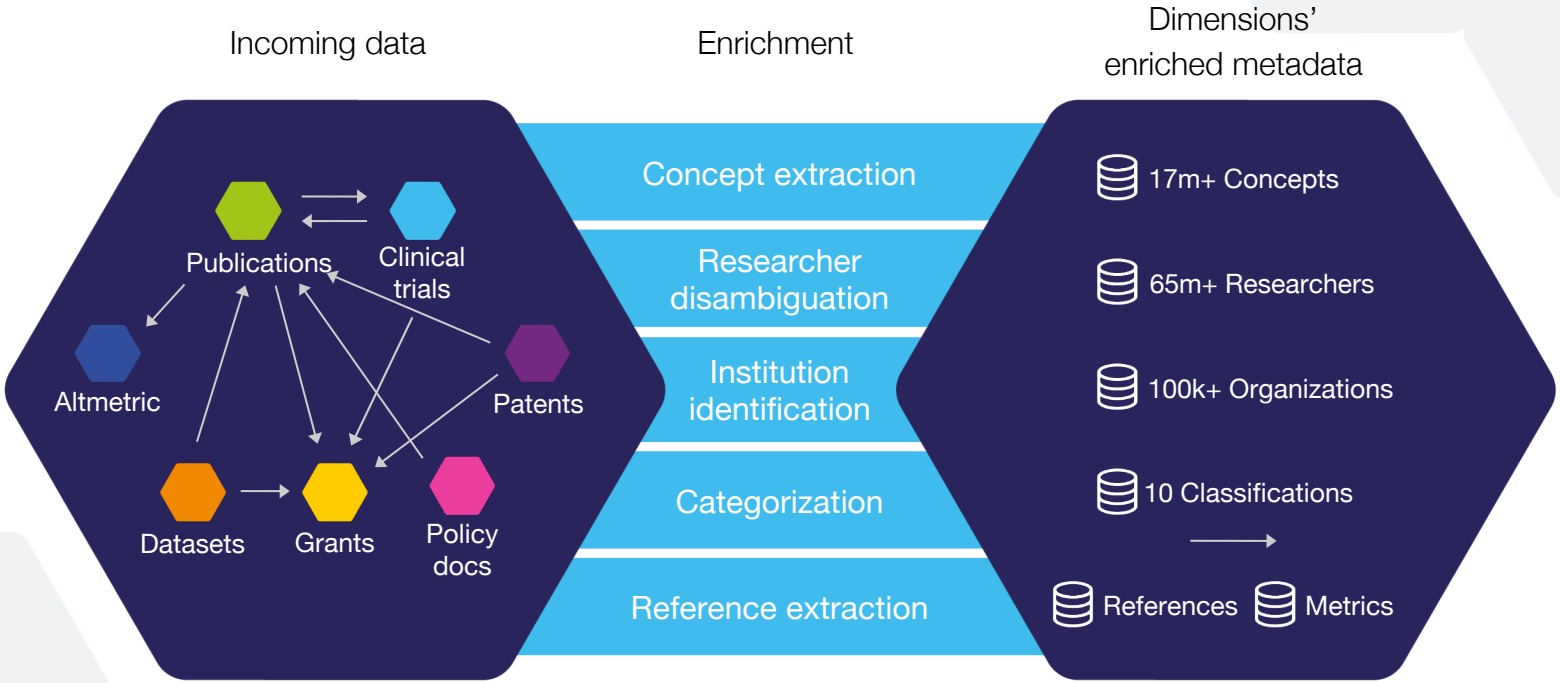
https://app.dimensions.ai/details/entities/publication/author/ur.0703533536.40

Dimensions

Data curation Approach

Part of DIGITAL science

A modern approach to data curation



Our approach to classifications in Dimensions

- Dimensions uses a machine learning approach, which enables us to implement multiple classification systems efficiently.
- The classification is based on analysis of the content of the documents, not journal attributes, which we don't believe are precise enough.
- This enables:
    - Article level classification
    - Consistent application across all content types (grants, patents, clinical trials, datasets)

Classifications in Dimensions are driven by the *content of the documents*

Dimensions

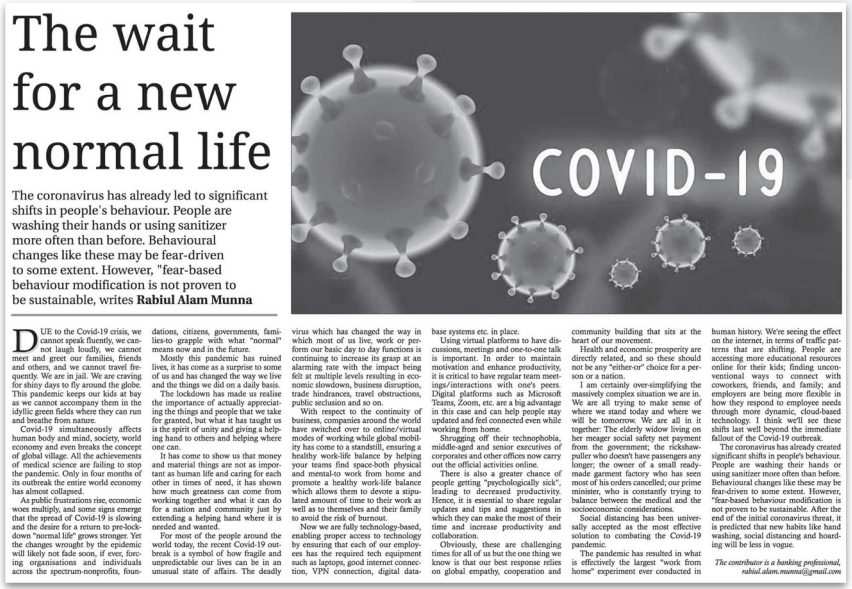
Navigating the New Normal

And why the approach matters

Part of DIGITAL science

What do we need to navigate The New Normal?

- Covid has changed the sector worldwide,
- Now more than ever, universities are looking to:
  - Find new revenue streams
  - Reduce costs
  - Better utilise existing resources
- More need for data-led approaches.
- Universities have access to lots of powerful and potentially informative data - but it's often siloed or difficult to work with.
- We need data that is ready-to-analyze and designed for answering questions.



The wait for a new normal life

The coronavirus has already led to significant shifts in people's behaviour. People are washing their hands or using sanitizer more often than before. Behavioural changes like these may be fear-driven to some extent. However, "fear-based behaviour modification is not proven to be sustainable, writes **Rabiul Alam Munna**

DUE to the Covid-19 crisis, we cannot speak fluently, we cannot laugh loudly, we cannot meet and greet our families, friends and others, and we cannot travel frequently. We are in jail. We are craving for shiny days to fly around the globe. This pandemic keeps our kids at bay as we cannot accompany them in the idyllic green fields where they can run and breathe from nature.

Covid-19 simultaneously affects human body and mind, society, world economy and even breaks the concept of global village. All the achievements of medical science are failing to stop the pandemic. Only in four months of its outbreak the entire world economy has almost collapsed.

As public frustrations rise, economic woes multiply, and some signs emerge that the spread of Covid-19 is slowing and the desire for a return to pre-lockdown "normal life" grows stronger. Yet the changes wrought by the epidemic will likely not fade soon, if ever, forcing organisations and individuals across the spectrum-nonprofits, foundations, citizens, governments, families-to grapple with what "normal" means now and in the future.

Mostly this pandemic has ruined lives, it has come as a surprise to some of us and has changed the way we live and the things we did on a daily basis.

The lockdown has made us realise the importance of actually appreciating the things and people that we take for granted, but what it has taught us is the spirit of unity and giving a helping hand to others and helping where one can.

It has come to show us that money and material things are not as important as human life and caring for each other in times of need, it has shown how much greatness can come from working together and what it can do for a nation and community just by extending a helping hand where it is needed and wanted.

For most of the people around the world today, the recent Covid-19 outbreak is a symbol of how fragile and unpredictable our lives can be in an unusual state of affairs. The deadly virus which has changed the way in which most of us live, work or perform our basic day to day functions is continuing to increase its grasp at an alarming rate with the impact being felt at multiple levels resulting in economic slowdown, business disruption, trade hindrances, travel obstructions, public seclusion and so on.

With respect to the continuity of business, companies around the world have switched over to online/virtual modes of working while global mobility has come to a standstill, ensuring a healthy work-life balance by helping your teams find space-both physical and mental-to work from home and promote a healthy work-life balance which allows them to devote a stipulated amount of time to their work as well as to themselves and their family to avoid the risk of burnout.

Now we are fully technology-based, enabling proper access to technology by ensuring that each of our employees has the required tech equipment such as laptops, good internet connection, VPN connection, digital database systems etc. in place.

Using virtual platforms to have discussions, meetings and one-to-one talk is important. In order to maintain motivation and enhance productivity, it is critical to have regular team meetings/interactions with one's peers. Digital platforms such as Microsoft Teams, Zoom, etc. are a big advantage in this case and can help people stay updated and feel connected even while working from home.

Shrugging off their technophobia, middle-aged and senior executives of corporates and other offices now carry out the official activities online.

There is also a greater chance of people getting "psychologically sick", leading to decreased productivity. Hence, it is essential to share regular updates and tips and suggestions in which they can make the most of their time and increase productivity and collaboration.

Obviously, these are challenging times for all of us but the one thing we know is that our best response relies on global empathy, cooperation and community building that sits at the heart of our movement.

Health and economic prosperity are directly related, and so these should not be any "either-or" choice for a person or a nation.

I am certainly over-simplifying the massively complex situation we are in. We are all trying to make sense of where we stand today and where we will be tomorrow. We are all in it together: The elderly widow living on her meager social safety net payment from the government; the rickshaw-puller who doesn't have passengers any longer; the owner of a small ready-made garment factory who has seen most of his orders cancelled; our prime minister, who is constantly trying to balance between the medical and the socioeconomic considerations.

Social distancing has been universally accepted as the most effective solution to combating the Covid-19 pandemic.

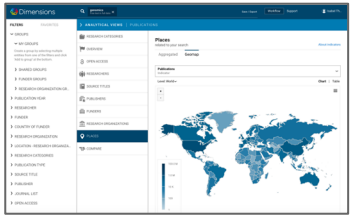
The pandemic has resulted in what is effectively the largest "work from home" experiment ever conducted in human history. We're seeing the effect on the internet, in terms of traffic patterns that are shifting. People are accessing more educational resources online for their kids; finding unconventional ways to connect with coworkers, friends, and family; and employers are being more flexible in how they respond to employee needs through more dynamic, cloud-based technology. I think we'll see these shifts last well beyond the immediate fallout of the Covid-19 outbreak.

The coronavirus has already created significant shifts in people's behaviour. People are washing their hands or using sanitizer more often than before. Behavioural changes like these may be fear-driven to some extent. However, "fear-based behaviour modification is not proven to be sustainable. After the end of the initial coronavirus threat, it is predicted that new habits like hand washing, social distancing and hoarding will be less in vogue.

The contributor is a banking professional, rabiul.alam.munna@gmail.com

One amazing database – multiple ways to access it

Web App



Search & discovery; top analytical use cases

Dedicated UI; inbuilt visualizations

In the browser, no specialized knowledge required

For everyone

API

Ad hoc analyses
& topic modelling

Full-text search & special functions
e.g. affiliation extraction

Product integrations e.g. CRIS

For API users
+ data & analytics teams

Google BigQuery

Fast, large scale analyses; dynamic dashboards; automated reporting

Join private & public data, access previously unsurfaced links

Direct integration with BI tools e.g. Tableau, Qlik, PowerBI

For data & analytics teams
+ dashboards for everyone

Why does this matter so much?

- **Committed to DORA**
  - Community classifications and metrics.
  - Available, auditable data.
  - Specific, contextualised assessment - appropriate to a particular institution, field, individual.
  - Creation of new metrics.
  - Real-time bibliometrics, reactive to changing environment.

- **Opportunity to expand horizons beyond narrow metrics**
  - We all know that metrics impact behaviour.
  - But “research excellence” is richer: a product of many threads coming together - people, funding, collaborations, culture.
  - Understanding and fostering research excellence can only be done fairly with access to an analyzable global dataset.

What do we need to deliver on this?

We're being aspirational - there are many pieces

- Good understanding of sector, the data and the business questions.
- Flexible, direct access to underlying data.
- Strong community of practice.
- Reusable examples and code.
- **Ultimately, we're trying to support fast execution and short time to value**.

Dimensions

Putting it into practice

What direct access to the data means

Part of DIGITAL science

\+

Direct access to <u>all</u> the underlying data

Ready to go

Google Cloud Platform ds-data-solutions-gbq Search products and resources

FEATURES & INFO SHORTCUT HIDE PREVIEW FEATURES

Explorer + ADD DATA

Type to search

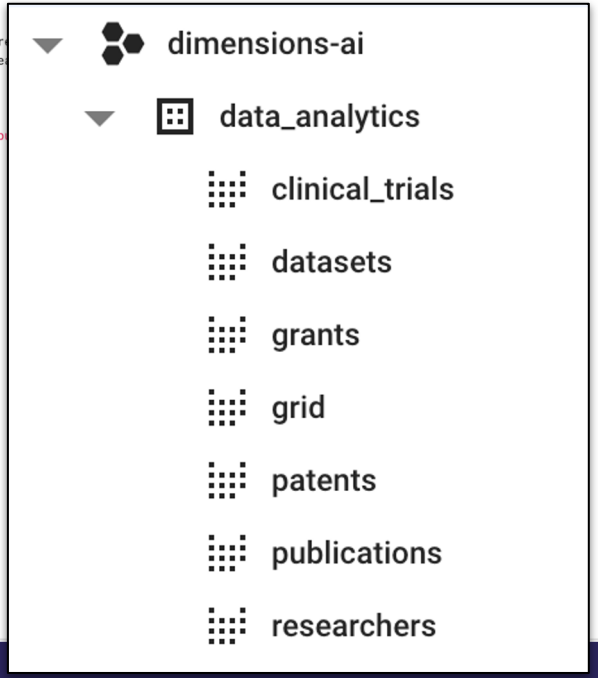
Viewing pinned projects.

- ds-data-solutions-gbq
- accuweather-com
- bigquery-public-data
- covid-19-dimensions-ai
- dimensions-ai
- dowjones-com
- ds-publisher-gbq
- fh-bigquery
- grid-ac
- patents-public-data
- test-bigquery-bob
- the-psf

*A COUNT... COMPOSE NEW QUERY

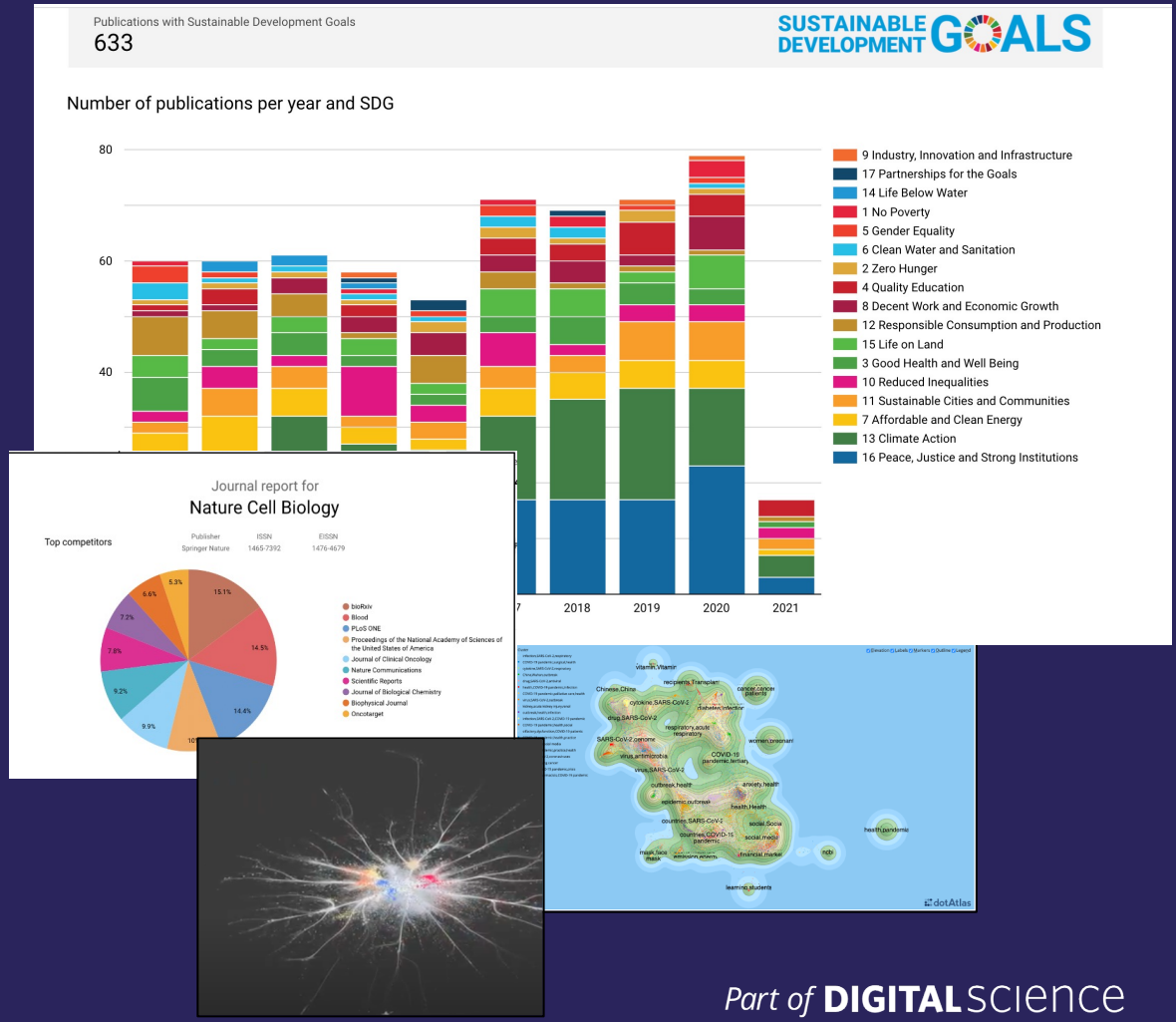
RUN SAVE SCHEDULE MORE This query will process 35.4 GiB when run.

```
1  --- Which publications, countries, organizations, and publishers is New Zealand citing? Which FoRs? How much of what NZ researchers are citing is Open Access?
2  WITH refids AS
3  (
4    --- select all of the publication IDs that are referneced in publications that contain a research org from New Zealand
5    SELECT DISTINCT ref
6    FROM
7      `dimensions-ai.data_analytics.publications` p, UNNEST(reference_ids) r
8    WHERE  year >=2010 and type = "article" and "New Zealand" IN UNNEST(rese
9  ),
10 refpubs AS
11 (
12   --- from the publications, find the publications from the above set of p
13   SELECT
14     p.id as reference_id,
15     publisher.name as publisher,
16     journal.title as journal,
17     year,
18     p.metrics.field_citation_ratio as fcr,
19     research_orgs as orgs,
20     ARRAY_LENGTH(research_orgs) as corg,
21     research_org_country_names as countries,
22     p.category_for.first_level.full as cat_for,
23     open_access_categories as oa
24   FROM `dimensions-ai.data_analytics.publications` p
25     INNER JOIN refids r ON r.ref = p.id
26 )
27 /*
28 SELECT
29   *
30 FROM refpubs r
31 WHERE r.corg >= 1
32 ORDER BY r.corg desc
33 */
34 SELECT
35   * EXCEPT(cat_for),
36   CONCAT(cat_for.code," ", cat_for.name) as cat_for
37 FROM refpubs r, UNNEST(cat_for) cat_for
38 WHERE r.corg >= 1
```



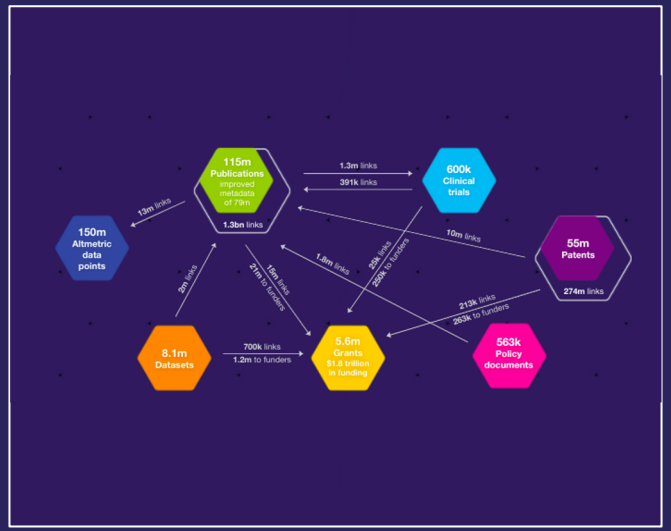
View and analyse Dimensions data however you want

- Slice, dice, ask whatever questions you want.
- Build custom dashboards that provide answers on a specific topic.



Break down barriers between your data and our data

Dimensions world on Google BigQuery

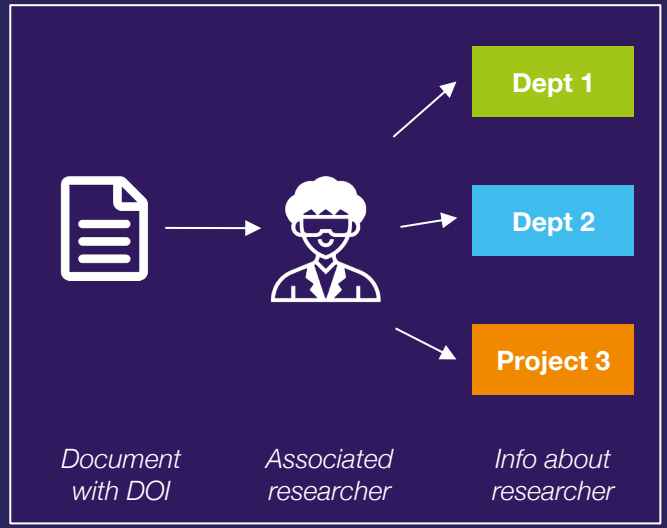


Representing the entire research landscape globally

DOI

Persistent identifier creates the link between both worlds e.g. DOI, GRID, researcher IDs

Your world in your CRIS or internal systems

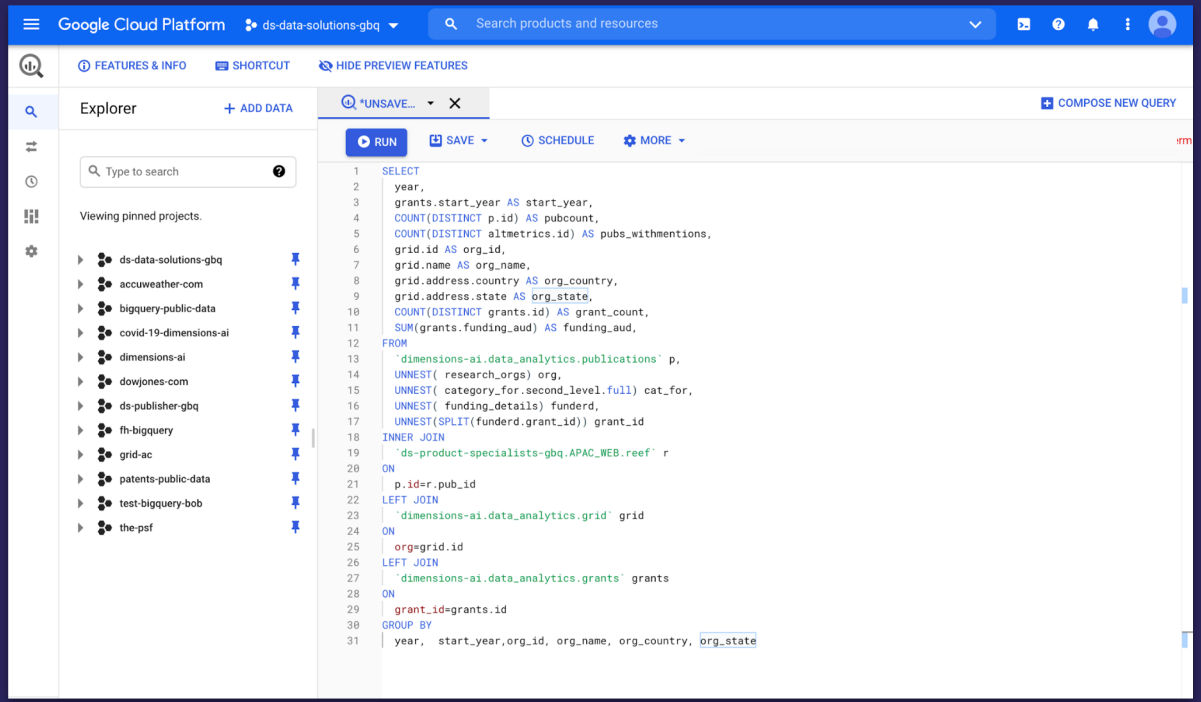


Carefully curated university-specific, private information

Dimensions

Part of DIGITAL science

One query. 31 Lines of code

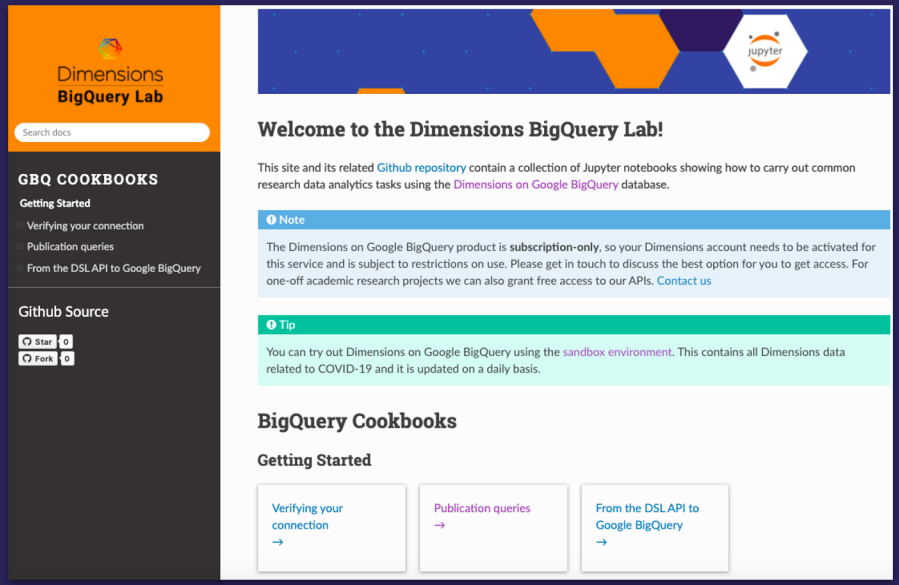


Dimensions

Part of **DIGITAL** science

Making it easy to work with the data

- **The BigQuery Lab**
  Example queries and notebooks
- **Bi-weekly analytics training**
- **Interactive Covid-19 dashboard**
- **Walkthroughs & step-by-step guides**
  - How to connect to Tableau, PowerBI
  - Build your own Covid dashboard
  - Machine Learning topic modelling



Dimensions

Who is it for

Part of DIGITAL science

Now you can create dashboards, reports and custom analyses for:

Chancellery, Deans & Planning

Strategic planning & competitive intelligence
- Peer group analysis & comparisons
- Internal benchmarking
- Cost-effectiveness of research
- Horizon scanning

Advocacy
- Impact, value of institutions & research

Talent planning & retention

Networks & context
- Industry collaboration & revenue generation

Research Office

Rankings & assessment
- National assessment
- International ranking
- Custom benchmarking

Impact assessment and narrative

Funding & grants
- Strategy
- Sources (incl. industry)
- Trends
- ROI assessment

Talent
- Recruitment & capability building
- Career tracking

Collaboration networks

Library

Collection Development
- Holdings & ROI assessment
- Publisher negotiations & transformative agreements

Open access
- Trends
- Citation and impact analysis
- Compliance tracking

Publication strategy
- Support researchers on where to publish

Information hub
- High value data asset for training

Faculty & Research

Strategy
- Funding trends & opportunities
- Faculty output tracking
- Find partners and recruit talent
- Emerging research trends
- Patent analysis

Bibliometrics
- Co-citation analysis

Use for research
- e.g. Computer Science, Business, Economics, Social Science
- Dataset for analysis and visualisation

And Beyond

Commercialization of research
- Science-->Innovation
- Tech transfer
- Support university spinouts & deep tech
- Patent analysis
- Find Key Opinion Leaders
- Industry partnerships

Marketing & Communications
- Market segmentation & targeting
- Monitor & grow wider impact & attention, publicity for research

Competitive intelligence

Flexible benchmarking

Research culture and environment

Compliance & foreign influence monitoring

Academic <-> Industry partnerships and collaboration, Technology foresight

Emerging research trends & horizon scanning

Impact analysis

Organization, topic, or journal overview

Creation of new metrics & indicators with transparency

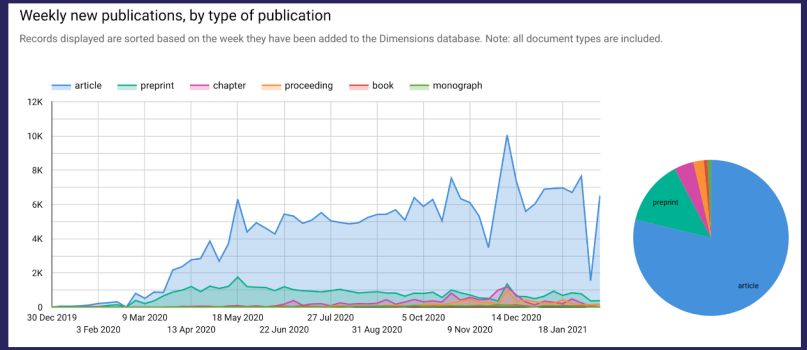
Rising stars, KOLs, recruitment and capability building

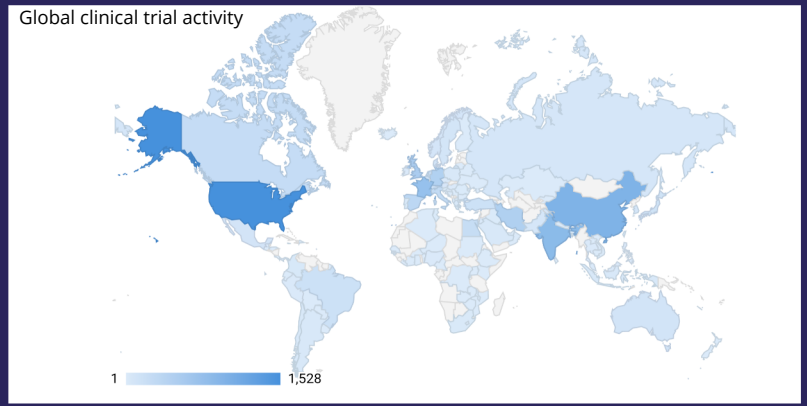
Open access behaviour shifts

Example of Dimensions Data on Google Big Query

Try it out! Our Covid-19 dataset is openly available on BigQuery

- Click here to get started: https://docs.dimensions.ai/bigquery/sandbox.html
- Explore our **interactive dashboard** using the Google BigQuery + Data Studio technology stack: https://reports.dimensions.ai/covid-19/
- More inspiration & code for analyses here: https://bigquery-lab.dimensions.ai/





Dimensions

Part of **DIGITAL** science

Discussion

Ways you can work with us

- **Across your organization:** Organizational subscription for the analytical Web Application
- **Direct data access** for your analytics team
- **Project licences**
- **Discuss with our consultancy team**: They can work with you on custom projects or on formulating these types of questions

Kelsey Rosell

k.rosell@digital-science.com

Liz.tomich

l.tomich@digital-science.com

Thank you! Questions?

Get in touch
info@dimensions.ai

or via

dimensions.ai/bigquery

Limited only by your imagination

(not our technology)

Appendix Slides

Publications (1/2) - Publication metadata backbone

PUBLICATIONS

- Journal articles, pre-prints and books/book chapters
- 100M + records based on metadata
- Metadata and citations derived from multiple available databases
- OA tagging
- Rule-based document type identification

JOURNALS / BOOKS

Crossref

PubMed

Europe PMC

J-STAGE

Japan Link Center

PRE-PRINT / OA

bioRxiv
THE PREPRINT SERVER FOR BIOLOGY

medRxiv

SSRN
Tomorrow's Research Today

PeerJ

ChemRxiv™

arXiv.org

Research Square

ESSOAr
Earth and Space Science Open Archive

SOC ARXIV

PsyArXiv Preprints

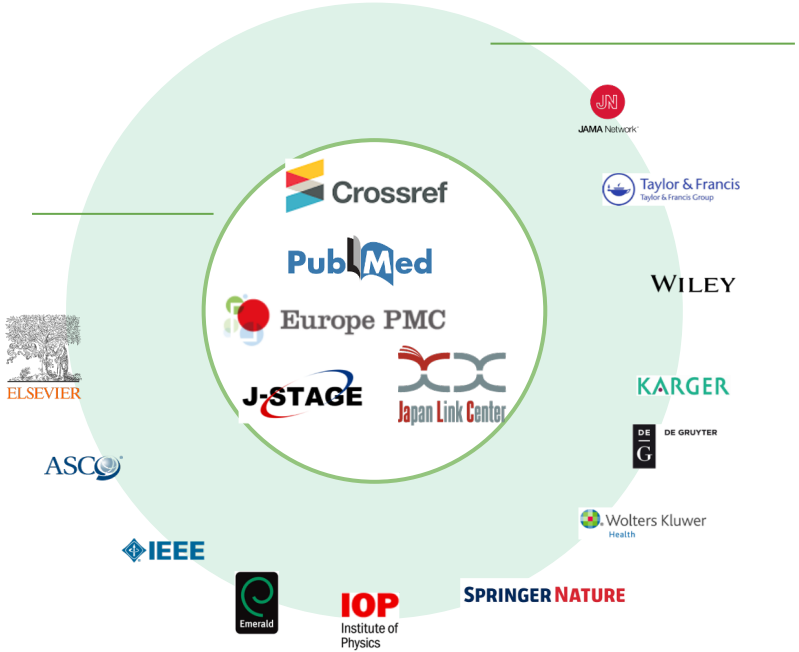
EarthArXiv

OSF PREPRINTS

...and more!

Publication metadata is enriched based on indexed full text data from publishers

- 112M records based on metadata
- ‘Backbone’ for Dimensions
- OA tagging
- Rule-based document type identification



- Full text for 78,620,438 publications (direct relationships with >130 publishers)
- Improved representation compared to the ‘backbone’ record
  - Additional metadata
  - Fill gaps
  - Deep indexing

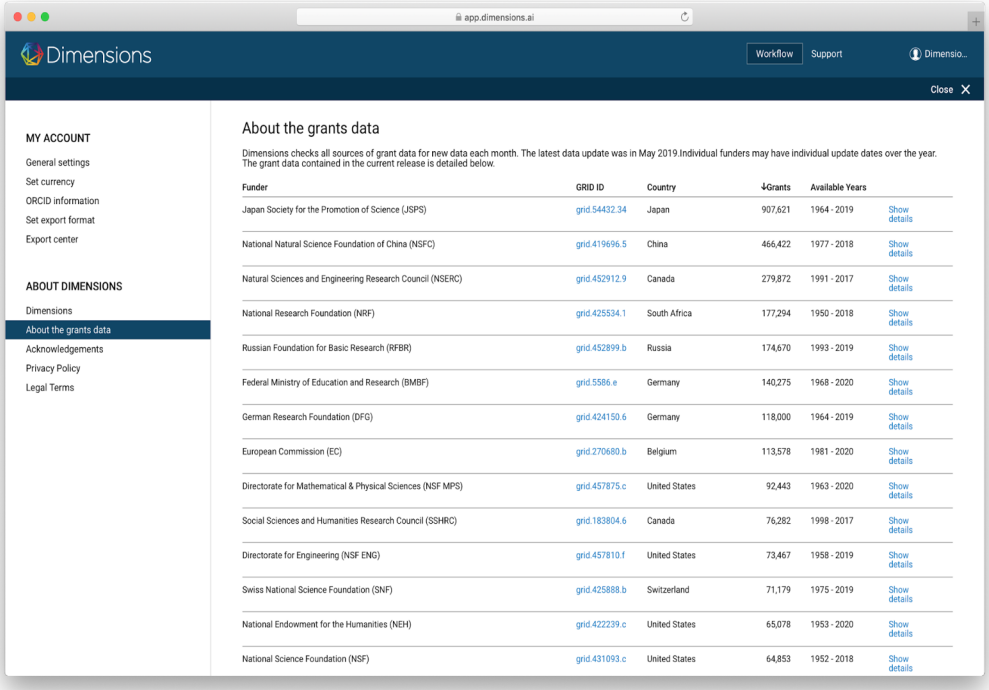
Open access data

- More than 32M open access publications
- Based on Unpaywall data integration and additional improvements by Dimensions
- All publications categorized as
  - Closed
  - Gold (Pure gold, Hybrid, Bronze)
  - Green (published, accepted, submitted)

Grants data

GRANTS

- Project funding
- Over 5M grants from 600+ funders globally
- $1.8 trillion of funding
- Sourcing
  - Direct relationships with funders
  - Data available via APIs
  - Data available via websites which we crawl

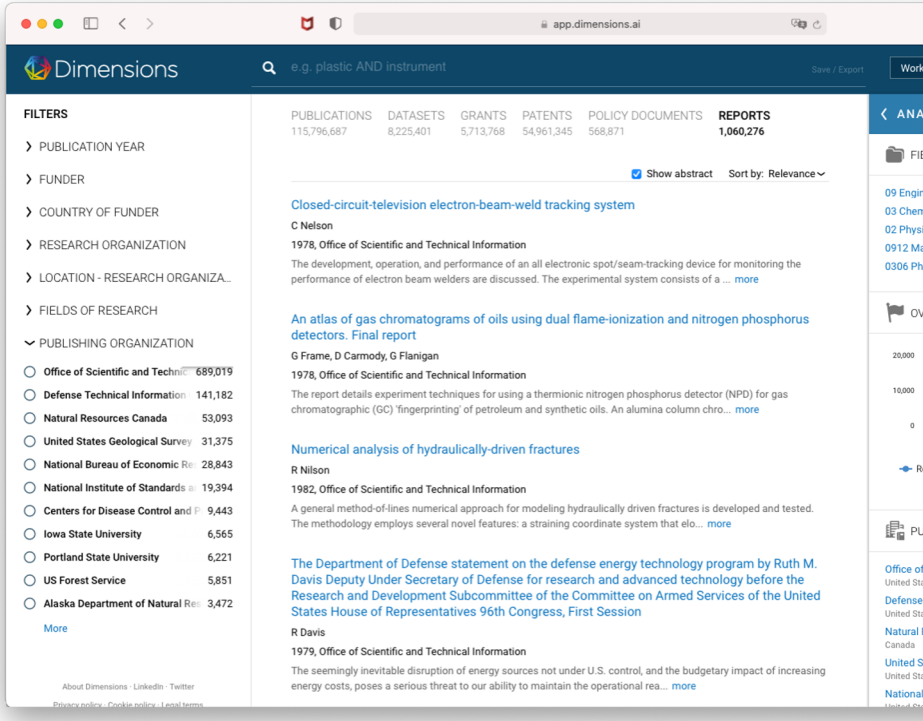


Technical reports data

TECHNICAL REPORTS

More than 1 million technical reports - including but not limited to:

- Office of Scientific and Technical Information (OSTI)
- Defense Technical Information Center (DTIC)
- Natural Resources Canada (NRCan)
- United States Geological Survey (USGS)
- National Bureau of Economic Research (NBER)



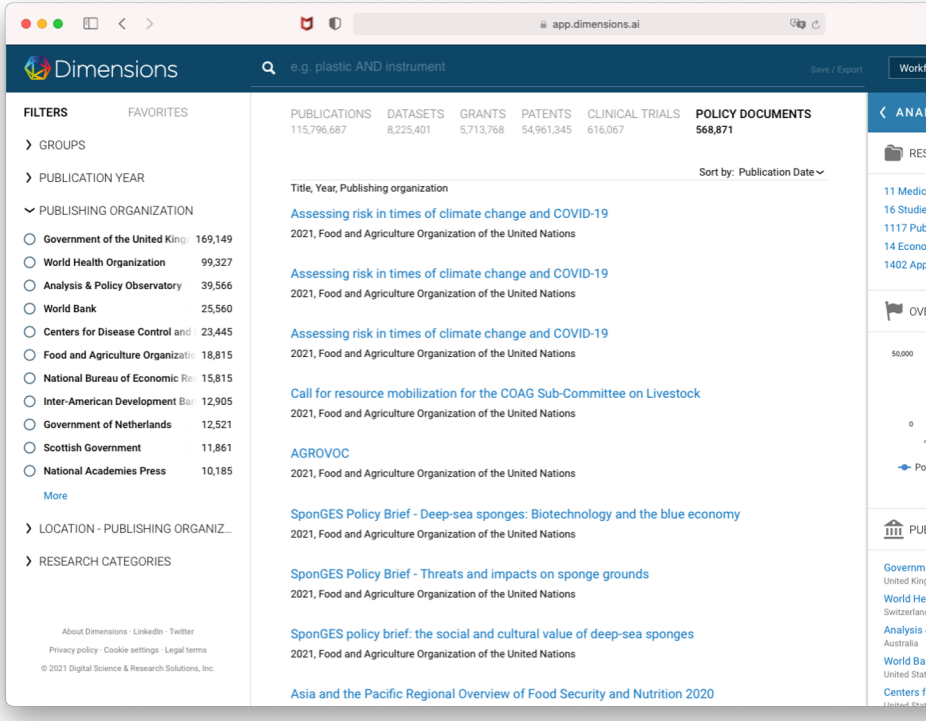
Policy documents data

POLICY DOCUMENTS

Over 500,000 policy document records, linked to publications

Including but not limited to:

- World Health Organization
- World Bank
- Centers for Disease Control & Prevention
- Government of the United Kingdom
- National Bureau of Economic Research

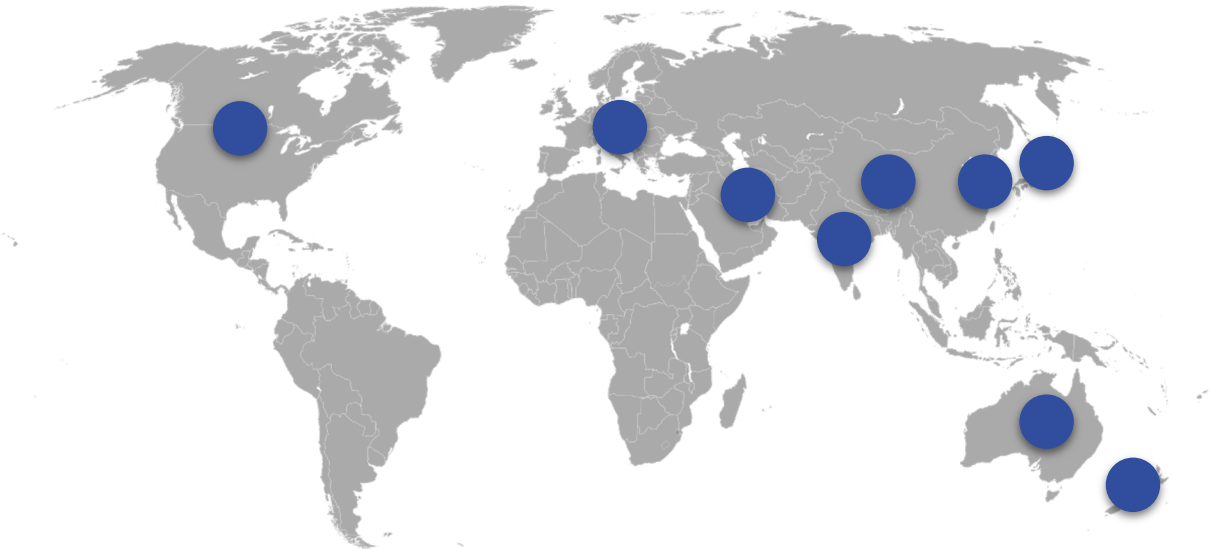


Clinical trials data

CLINICAL TRIALS

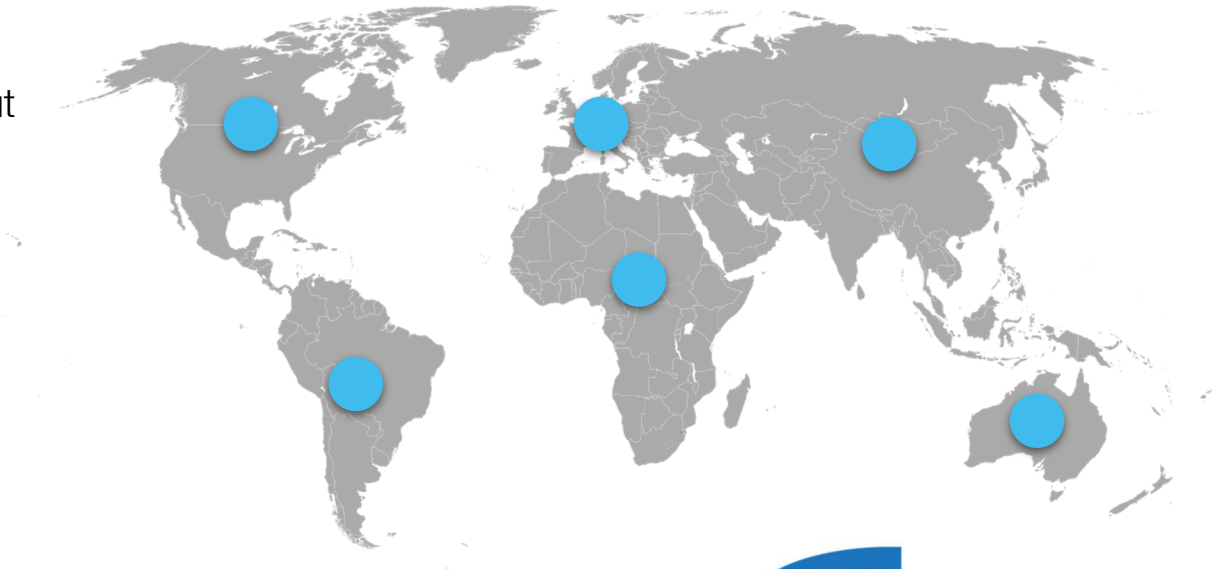
- ClinicalTrials.gov
- EU-CTR
- UMIN-CTR
- ISRCTN
- ANZCTR
- CHICTR
- ENCePP
- NTR
- GCTR
- CTRI
- CRIS
- IRCT

… and more are coming



Patents data

- 134 million+ patent documents
- Global coverage
- 100+ jurisdictions, including but not limited to:
  - China
  - Japan
  - United States
  - Germany
  - European Union
  - South Korea



ifi CLAIMS

Dimensions

See also: https://www.ificlaims.com/docs/Claims+Global+Data+Coverage.htm

Part of **DIGITAL** science

Books in Dimensions

- ~800,000 monographs, 296,891 edited books (10m chapters)
- Relevant research outputs to represent disciplines in a fair way
- From the 795k monographs, 242k are Open Access