Dimensions

Leveraging AI and Machine Learning for REF Preparations

Re-imagining discovery and access to research:

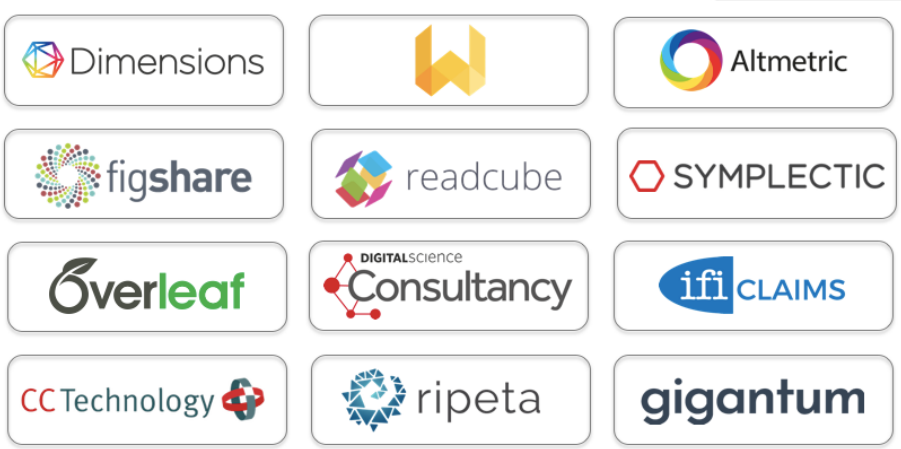
Grants, publications, citations, clinical trials and patents in one place

Juergen Wastl
Director Academic Relations and Consultancy

Part of DIGITAL science

Digital Science – a portfolio of innovative companies

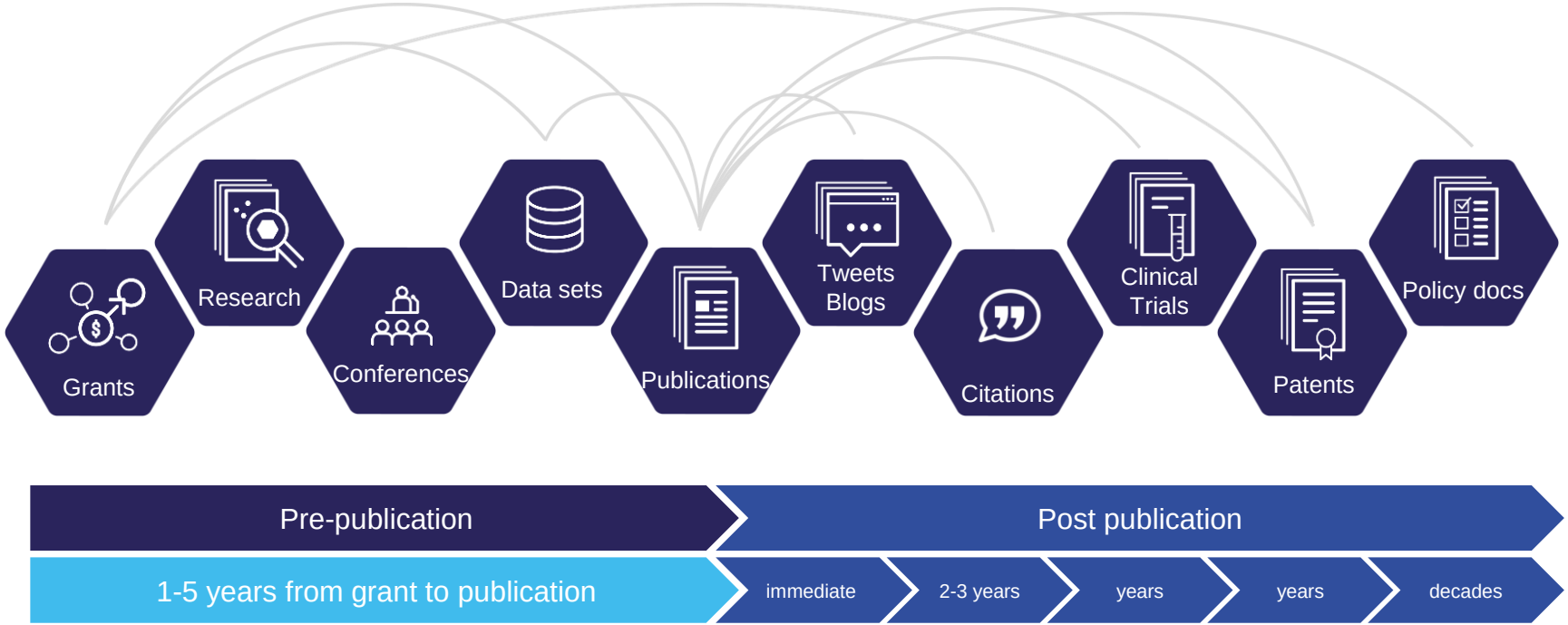
- Digital Science invests, supports and nurtures small innovative software companies
- Rooted in research, most founders come from an academic background, started a company around solving a self encountered problem
- Key principle: provide latitude, enable entrepreneurial spirit and activity, mix of ‘leaving alone’, ‘subtle guiding’, ‘having their backs’ and ‘open arguments’ like in every healthy family
- Started in 2009, invested in and started 12 companies, to date about 300 colleagues



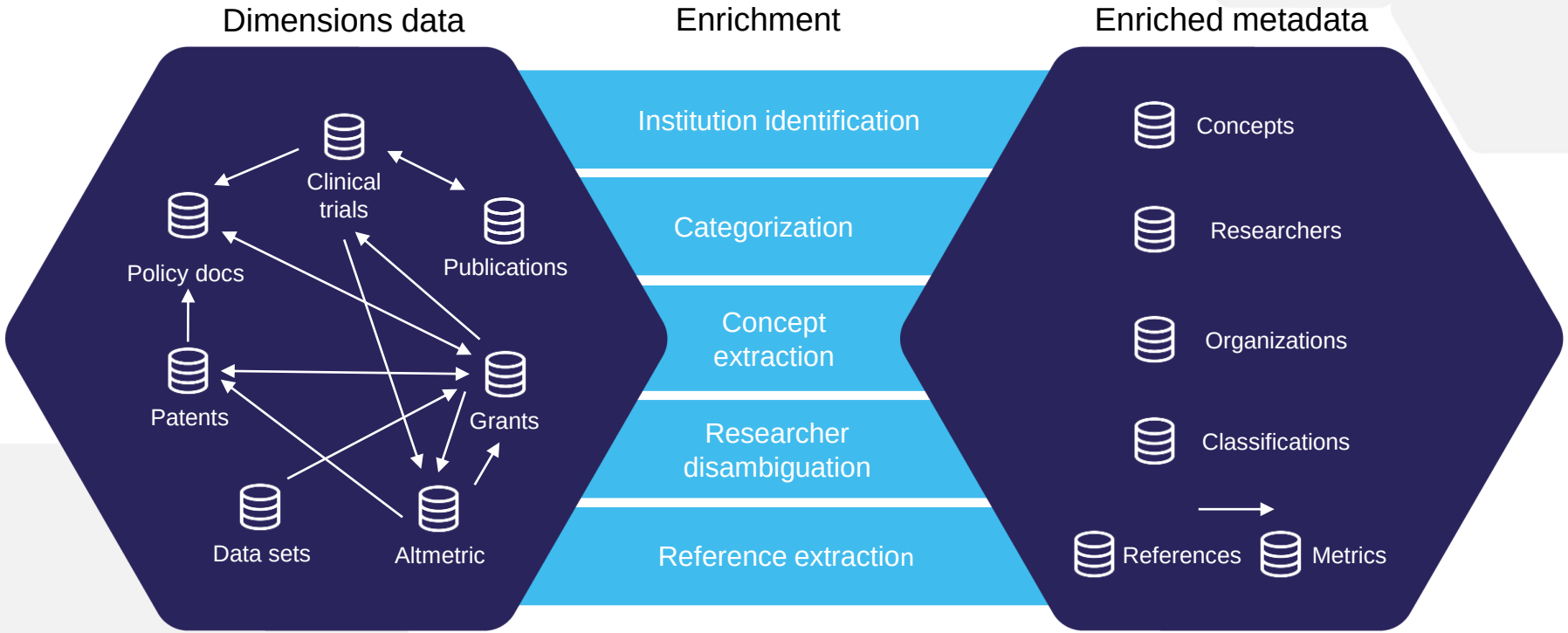
Why and how we built Dimensions

Six Digital Science companies worked together to find a solution to these challenges

Data integrated in multiple dimensions - pun intended



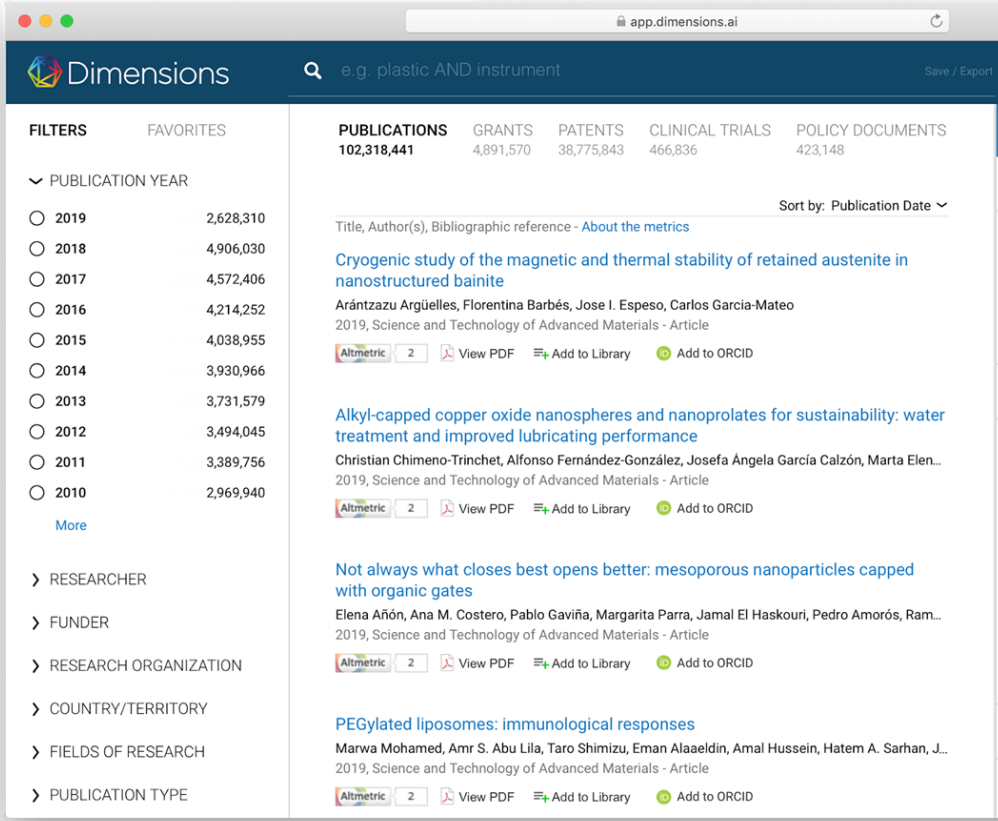
Enriched and interlinked – a modern approach to research data management



Dimensions Plus

Explore additional content types and discover their connections for the most complete view of the research landscape:

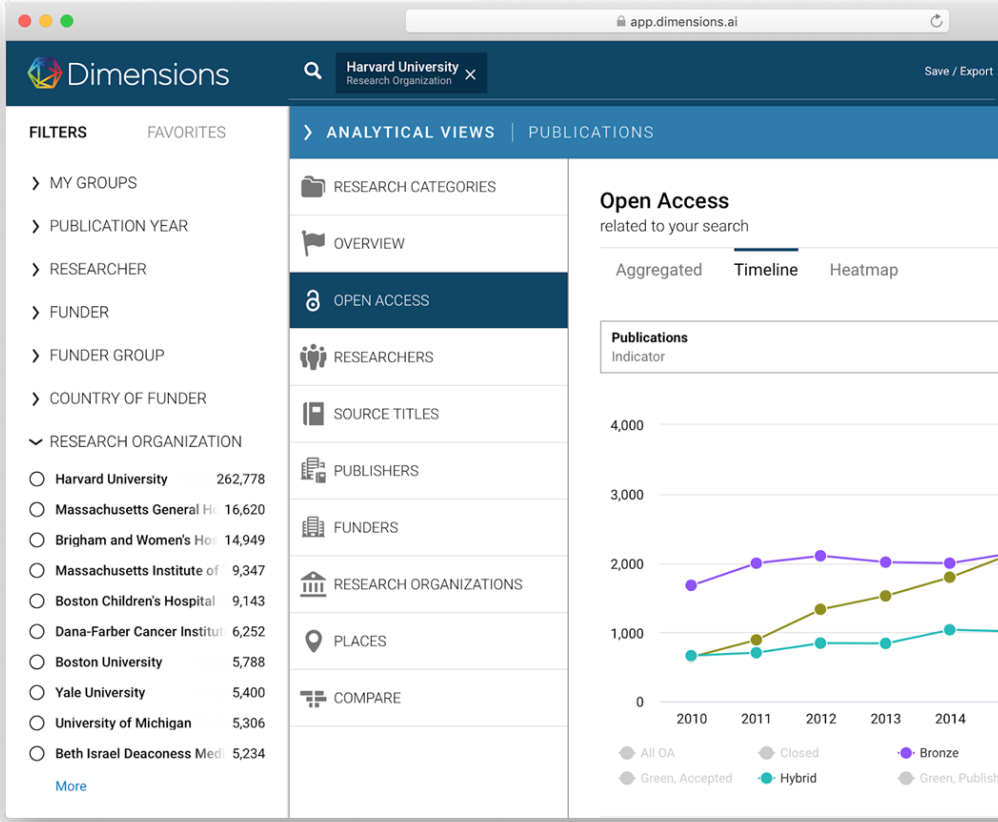
- Simultaneously search across all content types
- Article-level research categories consistently applied to all content types, as well as funder and research organizations
- Extended data aggregations to get the full picture of past, current, and future research



Dimensions Analytics

The complete data plus tools for in-depth data analytics and specific use cases of academic institutions, research funders, publishers and R&D:

- additional research classification systems based on machine learning and AI
- custom implementations and integrations with client system
- increased download limits of 50,000



Dimensions

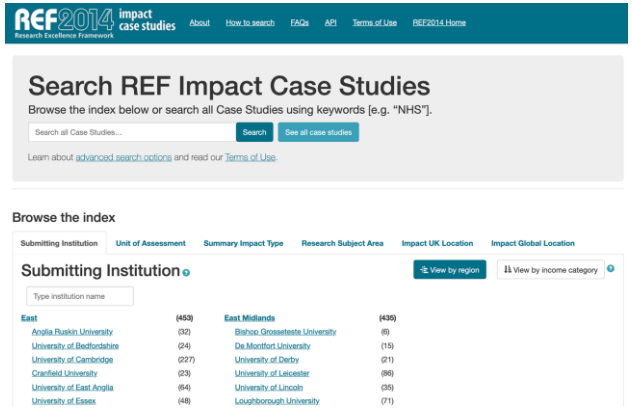
Digital Science and REF 2021

Part of DIGITAL science

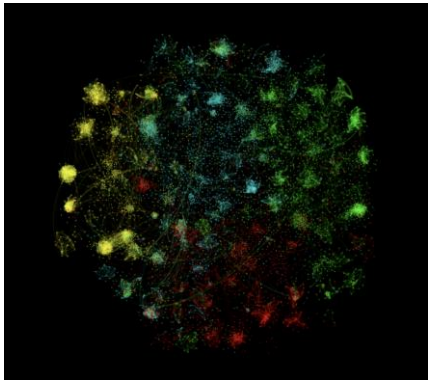
DIGITALscience Consultancy

REF TRACK RECORD:

source



https://impact.ref.ac.uk/casestudies/



https://www.digital-science.com/visualizations/ref-case-study-similarity-network/

Dimensions

Part of DIGITALscience

Portfolio’s contribution to REF preparations

	Symplectic	Dimensions	Altmetric
REF2 Research Outputs	Elements’ dedicated assessment module enables you to manage your UoA submission and comply with all mandatory REF data and justification statements. This includes identification, proposal, review, selection, attribution, and Open Access compliance	Compare your Institutionally held data with Dimensions’ world view on your staff’s outputs and research in context to retrieve a fuller picture of your output portfolio, including OA status and bibliometrics	From social media to policy citations, Altmetric’s insights can help you discover additional context in order to help inform your REF2 selections
REF3 Impact case studies	Elements’ dedicated Impact module enables you to gather impact narratives, evidence, links, and metadata to help inform case study selection and submission.	Dimensions enables you to gain insights into your UoA’s research portfolio, in context: grants, patents, and clinical trial links show you your unit’s research impact beyond citation counts to give you a fuller picture	Identify your research outputs’ links to policy documents, patents and online activity such as news and blogs to help evidence and corroborate impact case studies
REF5 Research Environment	Collect, collate and report on professional activities in your UoA to inform your submission strategy and populate your REF5a and REF5b Research Environment statements	Gain insights into the context of your UoA’s research (including patents, citations, grants, funders, collaborations, Open Access) to inform your submission strategy and populate your REF5a and REF5b Research Environment statements	Offering a comprehensive overview of research impact indicators and links to policy documents, patents and online media, Almetric’s insights can help enrich and underpin your REF5a and REF5b Research Environment statements

https://www.digital-science.com/ref2021/

Institutional REF preparations

The external change

REF from staff submission to publication submission

Pressure for improved selection of outputs higher than ever

The Research Funding Landscape (B-word, looming Spending review)

Degrees of uncertainty

Do you know your Institution?

What is your University's strategic goal for REF 2021?

Huge process, machinery

Usually a mixture of PVC/Provost office, Research Office/Services and Library

The Library now even more engaged (OA, repositories, Scholarly Comms, Research Impact)

Degrees of complexity

How can we help you prepare?

Categorising Dimensions data according to REF Units of Assessment

The REF 2021 UoA classification

Dimensions

e.g. plastic AND instrument

Workflow Support Jürgen W...

Close

BROWSE CATEGORIES for

Publications

Fields of Research

Research, Condition, and Disease ...

Health Category (HRCS)

Research Activity Codes (HRCS)

Broad Research Areas

Health Research Areas

ICRP Cancer Types

ICRP Common Scientific Outline (...

Units of Assessment

About Research Classifications

Units of Assessment

The Units of Assessment (UoA) is a classification scheme used by the Research Excellence Framework 2021 (REF) for assessing the quality of research in UK Higher Education Institutions. We have implemented this system using automated allocation of UoA codes to documents in Dimensions based on category definitions defined by machine learning.

A
- A01 Clinical Medicine
- A02 Public Health, Health Services and Primary...
- A03 Allied Health Professions, Dentistry, Nursin...
- A04 Psychology, Psychiatry and Neuroscience
- A05 Biological Sciences
- A06 Agriculture, Veterinary and Food Science

B
- B07 Earth Systems and Environmental Sciences
- B08 Chemistry
- B09 Physics
- B10 Mathematical Sciences
- B11 Computer Science and Informatics
- B12 Engineering

C
- C13 Architecture, Built Environment and Planning
- C14 Geography and Environmental Studies
- C15 Archaeology
- C16 Economics and Econometrics
- C17 Business and Management Studies
- C18 Law
- C19 Politics and International Studies
- C20 Social Work and Social Policy
- C21 Sociology
- C22 Anthropology and Development Studies
- C23 Education
- C24 Sport and Exercise Sciences, Leisure and T...

D
- D25 Area Studies
- D26 Modern Languages and Linguistics
- D27 English Language and Literature
- D28 History
- D29 Classics
- D30 Philosophy
- D31 Theology and Religious Studies
- D32 Art and Design: History, Practice and Theory
- D33 Music, Drama, Dance, Performing Arts, Fil...
- D34 Communication, Cultural and Media Studie...

Potential users and use cases

<u>UoA view on:</u>

Individual (or groups of individuals)

Institution

UK-wide

Potential users and use cases

<u>UoA view on:</u>

Individual (or groups of individuals)

Institution

UK-wide

<u>UoA view on:</u>

Publications

Grants

Impact

(any text)

Table: comparison UoA with FOR classification scheme

	UoA	FOR
Units	34 (organised in 4 Main Panels)	22 (two digit) & 157 (4 digit)
Training set	Research outputs REF 2014	Grants ANZSRC (FOR, 2008)
Assigned by	Selection to UoA	Manually coded

132k training set

Machine Learning/NLP

Individual: UoA vs FOR (example Stephen Hawking)

Stephen W Hawking
Researcher

Save / Export | Workflow | Support | Juergen ...

ANALYTICAL VIEWS | PUBLICATIONS

- RESEARCH CATEGORIES
- OVERVIEW
- OPEN ACCESS
- RESEARCHERS
- SOURCE TITLES
- PUBLISHERS
- FUNDERS
- RESEARCH ORGANIZATIONS
- PLACES
- COMPARE

Research Categories

related to your search

About indicators

Aggregated | Bar Chart | Timeline | Heatmap

Classification: Units of Assessment

Export table

Name	↓ Publications Relevant	FCR Mean	RCR Mean
Mathematical Sciences	84	0.00	1.06
Physics	23	0.00	0.43
Engineering	1	0.00	0.00
Philosophy	1	0.00	0.00

W Hawking

Save / Export | Workflow | Support | Juergen ...

CAL VIEWS | PUBLICATIONS

H CATEGORIES

CESS

HERS

TITLES

ERS

H ORGANIZATIONS

Research Categories

related to your search

About indicators

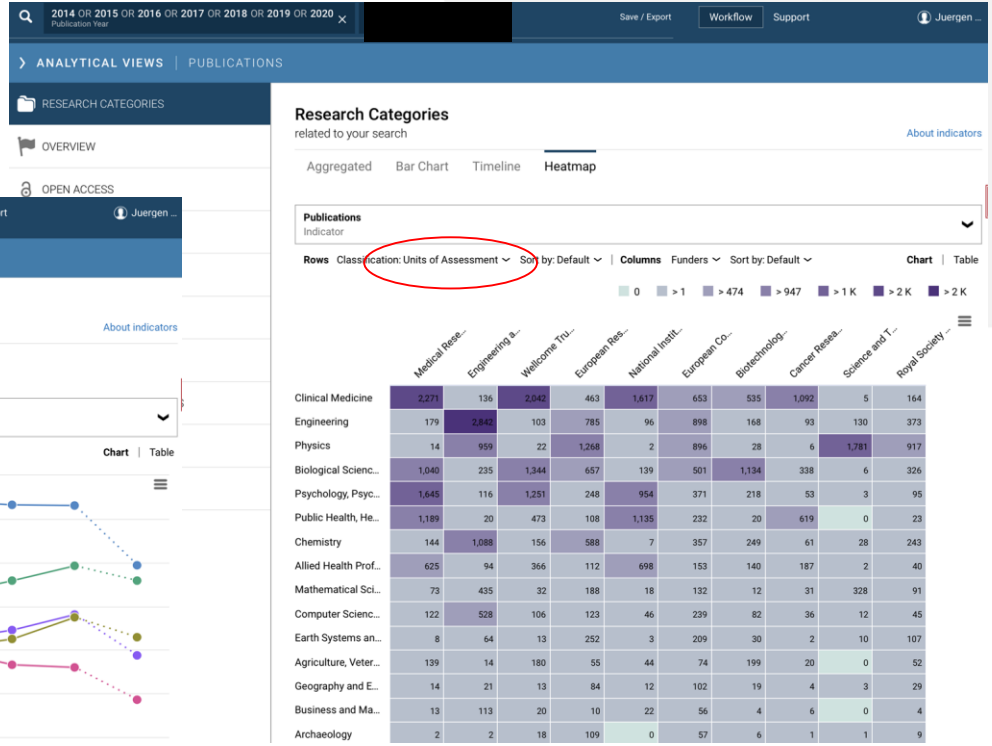
Aggregated | Bar Chart | Timeline | Heatmap

Classification: Fields of Research

Export table

Name Fields of Research code	↓ Publications Relevant	FCR Mean	RCR Mean
Physical Sciences 02	101	0.00	1.02
Astronomical and Space Sciences 0201	55	0.00	0.91
Atomic, Molecular, Nuclear, Particle and Plasma Physics 0202	36	0.00	0.96
Quantum Physics 0206	36	0.00	0.83
Mathematical Sciences 01	34	0.00	0.00
Pure Mathematics 0101	16	0.00	0.00
Mathematical Physics 0105	14	0.00	0.00
Other Physical Sciences 0299	6	0.00	0.00
Engineering 09	1	0.00	0.00

Main trends: UoA view in REF 2021 time frame



Dimensions

Part of DIGITALscience

UoA comparison with peer Institutions

2020 OR 2019 OR 2018 OR 2017 OR 2016 OR 2015 OR 2014
Publication Year

Save / Export | Workflow | Support | Jürgen W...

ANALYTICAL VIEWS | PUBLICATIONS

- RESEARCH CATEGORIES
- OVERVIEW
- OPEN ACCESS
- RESEARCHERS
- SOURCE TITLES
- PUBLISHERS
- FUNDERS
- RESEARCH ORGANIZATIONS
- PLACES
- COMPARE

Compare
related to your search

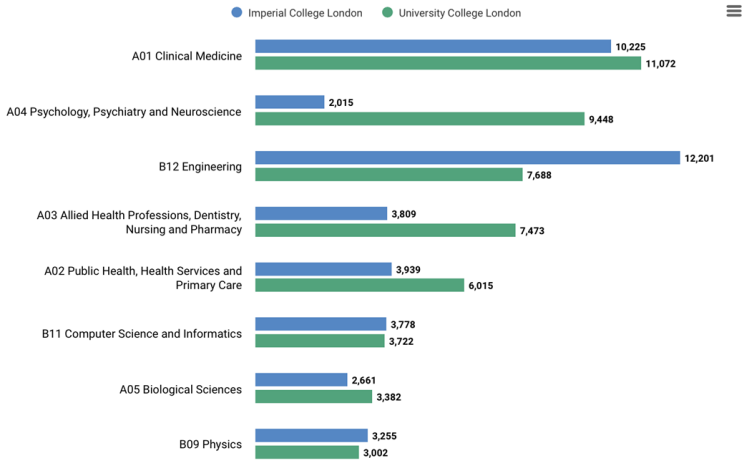
About indicators

Publications
Indicator

Research Organization | Funder

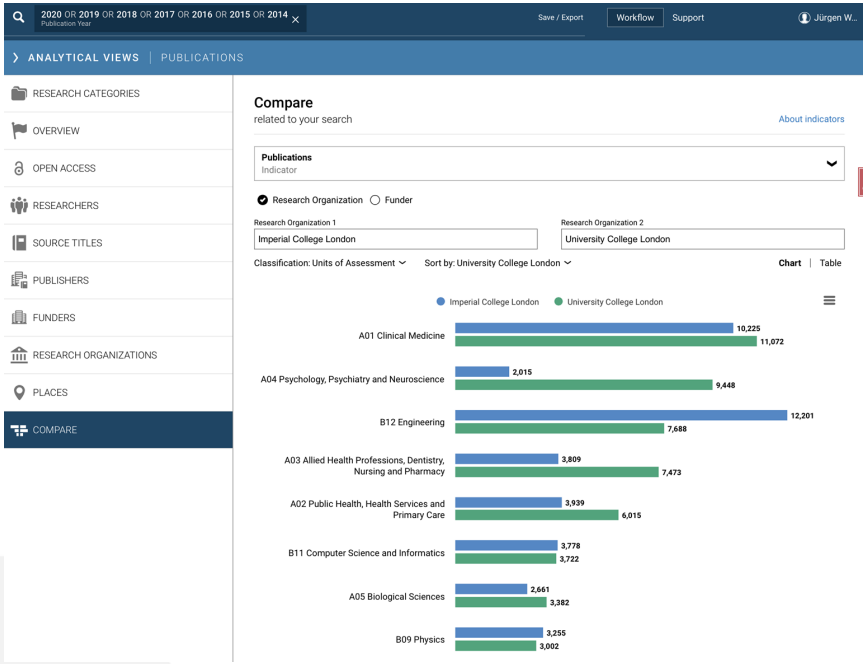
Research Organization 1: Imperial College London
Research Organization 2: University College London

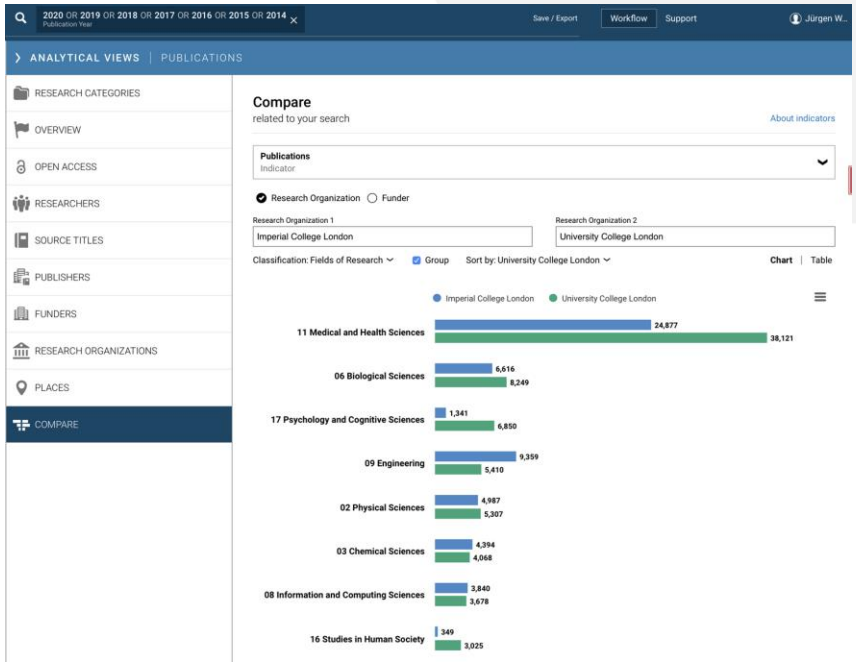
Classification: Units of Assessment | Sort by: University College London | Chart | Table



UoA comparison with peer Institutions

(FOR)





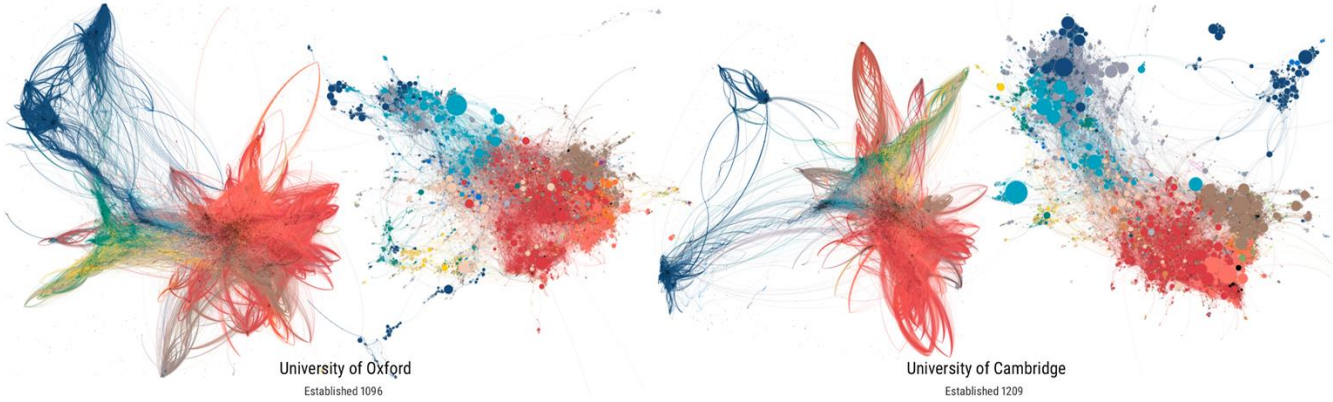
Visualisation of UoA Networks

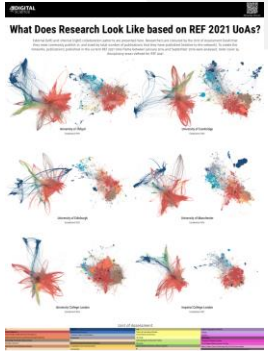
DIGITAL science

Interactive Version

What Does Research Look Like based on REF 2021 UoAs?

External (left) and internal (right) collaboration patterns are presented here. Researchers are coloured by the Unit of Assessment (UoA) that they most commonly publish in, and sized by total number of publications that they have published (relative to the network). To create the networks, publications published in the current REF 2021 time frame between January 2014 and September 2019 were analysed. UoAs cover 34 disciplinary areas defined for REF 2021





https://doi.org/10.6084/m9.figshare.9926492

Why use UoA classification?

For the REF:

- inform submission strategy
- identify outputs you've not considered
- identify & (in)form research groupings

… and wider use beyond this REF (and UK):

- Ability to see your research in an ongoing basis, classified by UoA - rather than just this snapshot every 7 years
- Possibility to ingrain UoAs more completely in your internal analyses- like a 'neutral Unit'
  - not departmentalised (which I found always to be problematic for reporting etc)

UoA classification: One step further beyond publication

Country comparison

UoA sector wide comparison (incl citation analysis)

Grant comparison (funder analysis in a new light?)

Impact

UoA Classification of Impact

Impact summary (100)

Underpinning Research Statement (500)

UoA classification scheme

Details of the Impact (750)

Underpinning research (doi)

May be worth having another retrospective look at your Impact case studies

REF2014 Case studies Cambridge

Count of Case Study Id	UoA predicted (Underpinning Research)																																						
UoA submitted	A 01 (	A 02 I	A 03 /	A 04 I	A 05 I	A 06 /	B 07 I	B 08 (	B 09 I	B 10 I	B 11 (	B 12 I	C 13 /	C 14 (	C 15 /	C 16 I	C 17 I	C 18 I	C 19 I	C 20 S	C 21 S	C 22 /	C 23 I	D 25 /	D 26 I	D 27 I	D 28 I	D 29 (	D 30 I	D 31 '	D 32 /	D 33 I	D 34 (	(blani	Grand Total	
A 01 Clinical Medicine	18	1																																	19	
A 02 Public Health, Health Services and Primary Care	1	5		1																															7	
A 03																																				
A 04 Psychology, Psychiatry and Neuroscience				6							1	1						1																	9	
A 05 Biological Sciences	3				7	2	3	2			1	1		1																					20	
A 06 Agriculture, Veterinary and Food Science						5																													5	
B 07 Earth Systems and Environmental Sciences							4					1																							5	
B 08 Chemistry			1				1	2			1																								5	
B 09 Physics								1	6		3	3																							13	
B 10 Mathematical Sciences		1					2		3	1	1	3	1																				1		13	
B 11 Computer Science and Informatics											5									1															6	
B 12 Aeronautical, Mechanical, Chemical and Manufacturing Engineering				1			1					2																							4	
B 13 Electrical and Electronic Engineering, Metallurgy and Materials												4																							4	
B 15 General Engineering											2	14					2														1				19	
C 16 Architecture, Built Environment and Planning													3					1																	4	
C 17 A Geography, Environmental Studies and Archaeology							1							5																					6	
C 17 B Geography, Environmental Studies and Archaeology															5																				5	
C 18 Economics and Econometrics																1	3																		4	
C 19 Business and Management Studies							1										3																		4	
C 20 Law																				3															3	
C 21 Politics and International Studies																	1	1	2																4	
C 22																																				
C 23 Sociology																	1			1	1														3	
C 24 Anthropology and Development Studies																				1		3	1												5	
C 25 Education																							4												4	
D 27 Area Studies																			1			1		2											4	
D 28 Modern Languages and Linguistics															1										8		1				1				11	
D 29 English Language and Literature														1									1			3	1				1	1			8	
D 30 History																					1	1	2				7				1		1		13	
D 31 Classics													1		2										1			1							5	
D 32 Philosophy															1				1							1	1		3		1				8	
D 33 Theology and Religious Studies																											1			2					3	
D 34																																				
D 35 Music, Drama, Dance and Performing Arts																							1									2			3	
D 36																																				
(blank)																																				
Grand Total	22	7	1	8	7	7	13	5	9	1	14	29	5	7	9	1	10	3	4	6	2	5	9	2	9	4	11	1	3	2	5	3	2	226		134
																																				0.59

Dimensions

Part of DIGITAL science

REF2014 Case studies Cambridge

Count of Case Study Id	UoA predicted (Underpinning Research)																																				
UoA submitted	A 01 (	A 02 f	A 03 /	A 04 f	A 05 f	A 06 /	B 07 f	B 08 (	B 09 f	B 10 f	B 11 (	B 12 f	C 13 /	C 14 (	C 15 /	C 16 f	C 17 f	C 18 l	C 19 f	C 20 S	C 21 S	C 22 /	C 23 f	D 25 /	D 26 f	D 27 f	D 28 f	D 29 (	D 30 f	D 31 '	D 32 /	D 33 f	D 34 (	(blani	Grand Total	
A 01 Clinical Medicine	18	1																																	19	
A 02 Public Health, Health Services and Primary Care	1	5		1																															7	
A 03																																				
A 04 Psychology, Psychiatry and Neuroscience				6							1	1						1																	9	
A 05 Biological Sciences	3				7	2	3	2			1	1		1																					20	
A 06 Agriculture, Veterinary and Food Science						5																													5	
B 07 Earth Systems and Environmental Sciences							4					1																							5	
B 08 Chemistry			1				1	2			1																								5	
B 09 Physics								1	6		3	3																							13	
B 10 Mathematical Sciences		1					2		3	1	1	3	1																				1		13	
B 11 Computer Science and Informatics											5									1															6	
B 12 Aeronautical, Mechanical, Chemical and Manufacturing Engineering				1			1					2																							4	
B 13 Electrical and Electronic Engineering, Metallurgy and Materials												4																							4	
B 15 General Engineering											2	14					2														1				19	
C 16 Architecture, Built Environment and Planning													3					1																	4	
C 17 A Geography, Environmental Studies and Archaeology							1							5																					6	
C 17 B Geography, Environmental Studies and Archaeology															5																				5	
C 18 Economics and Econometrics																1	3																		4	
C 19 Business and Management Studies							1										3																		4	
C 20 Law																				3															3	
C 21 Politics and International Studies																	1	1	2																4	
C 22																																				
C 23 Sociology																	1			1	1														3	
C 24 Anthropology and Development Studies																				1		3	1												5	
C 25 Education																							4												4	
D 27 Area Studies																			1			1		2											4	
D 28 Modern Languages and Linguistics															1										8		1				1				11	
D 29 English Language and Literature														1									1			3	1				1	1			8	
D 30 History																					1	1	2				7				1		1		13	
D 31 Classics													1		2										1			1							5	
D 32 Philosophy															1				1							1	1		3		1				8	
D 33 Theology and Religious Studies																											1			2					3	
D 34																																				
D 35 Music, Drama, Dance and Performing Arts																							1									2			3	
D 36																																				
(blank)																																				
Grand Total	22	7	1	8	7	7	13	5	9	1	14	29	5	7	9	1	10	3	4	6	2	5	9	2	9	4	11	1	3	2	5	3	2	226		134
																																				0.59

Dimensions

Part of DIGITAL science

REF2014 Case studies Cambridge

Count of Case Study Id	UoA predicted (Underpinning Research)																																						
UoA submitted	A 01 (	A 02 I	A 03 /	A 04 I	A 05 I	A 06 /	B 07 I	B 08 (	B 09 I	B 10 I	B 11 (	B 12 I	C 13 /	C 14 (	C 15 /	C 16 I	C 17 I	C 18 I	C 19 I	C 20 S	C 21 S	C 22 /	C 23 I	D 25 /	D 26 I	D 27 I	D 28 I	D 29 (	D 30 I	D 31 '	D 32 /	D 33 I	D 34 (	(blani	Grand Total	
A 01 Clinical Medicine	18	1																																	19	
A 02 Public Health, Health Services and Primary Care	1	5		1																															7	
A 03																																				
A 04 Psychology, Psychiatry and Neuroscience				6							1	1						1																	9	
A 05 Biological Sciences	3				7	2	3	2			1	1		1																					20	
A 06 Agriculture, Veterinary and Food Science						5																													5	
B 07 Earth Systems and Environmental Sciences							4					1																							5	
B 08 Chemistry			1				1	2			1																								5	
B 09 Physics								1	6		3	3																							13	
B 10 Mathematical Sciences		1					2		3	1	1	3	1																				1		13	
B 11 Computer Science and Informatics											5									1															6	
B 12 Aeronautical, Mechanical, Chemical and Manufacturing Engineering				1			1					2																							4	
B 13 Electrical and Electronic Engineering, Metallurgy and Materials												4																							4	
B 15 General Engineering											2	14					2														1				19	
C 16 Architecture, Built Environment and Planning													3					1																	4	
C 17 A Geography, Environmental Studies and Archaeology							1							5																					6	
C 17 B Geography, Environmental Studies and Archaeology															5																				5	
C 18 Economics and Econometrics																1	3																		4	
C 19 Business and Management Studies							1										3																		4	
C 20 Law																				3															3	
C 21 Politics and International Studies																	1	1	2																4	
C 22																																				
C 23 Sociology																	1			1	1														3	
C 24 Anthropology and Development Studies																				1		3	1												5	
C 25 Education																							4												4	
D 27 Area Studies																			1			1		2											4	
D 28 Modern Languages and Linguistics															1										8		1				1				11	
D 29 English Language and Literature														1									1			3	1				1	1			8	
D 30 History																					1	1	2				7				1		1		13	
D 31 Classics													1		2										1			1							5	
D 32 Philosophy															1				1							1	1		3		1				8	
D 33 Theology and Religious Studies																											1			2					3	
D 34																																				
D 35 Music, Drama, Dance and Performing Arts																							1									2			3	
D 36																																				
(blank)																																				
Grand Total	22	7	1	8	7	7	13	5	9	1	14	29	5	7	9	1	10	3	4	6	2	5	9	2	9	4	11	1	3	2	5	3	2	226		134
																																				0.59

Dimensions

Part of DIGITALscience

REF2014 Case studies (

Low-energy buildings

Summary of the impact

Research undertaken at the University of Cambridge Department of Applied Mathematics and Theoretical Physics (DAMTP) was the first to demonstrate that low-energy systems could be modelled in the laboratory and that the complex ventilation flows within buildings could be represented accurately by simple algorithms. These algorithms were

Read More

Submitting Institution	Unit of Assessment	Summary Impact Type
University of Cambridge	Mathematical Sciences	Technological

Research Subject Area(s)

Engineering: Interdisciplinary Engineering

Built Environment and Design: Building, Other Built Environment and Design

| Count of Case Study Id | UoA predicted (Underpinning Research) | | |

Impact summary classification vs Underpinning research classification

| Count of Case Study Id based on underpinning research narrative in ICS REF 2014 | UoA predicted (with REF2021 UoA order) | | |

Thank you very much for your attention

Accuracy

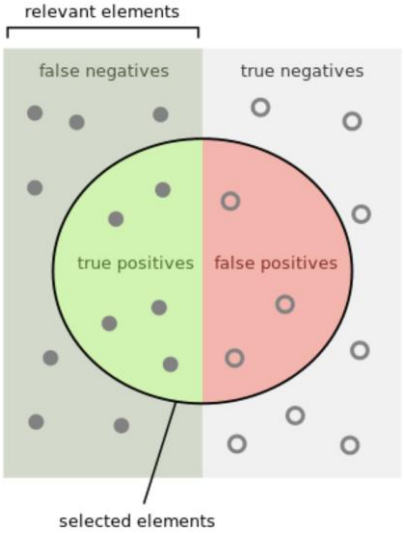
Category	precision	recall	size	relative size
(B, 10, Mathematical Sciences)	0.779993	0.784258	6035	0.045564

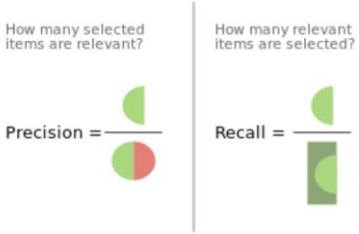
Precision: how useful are the search results?

Recall: how complete are the results?

132k training set

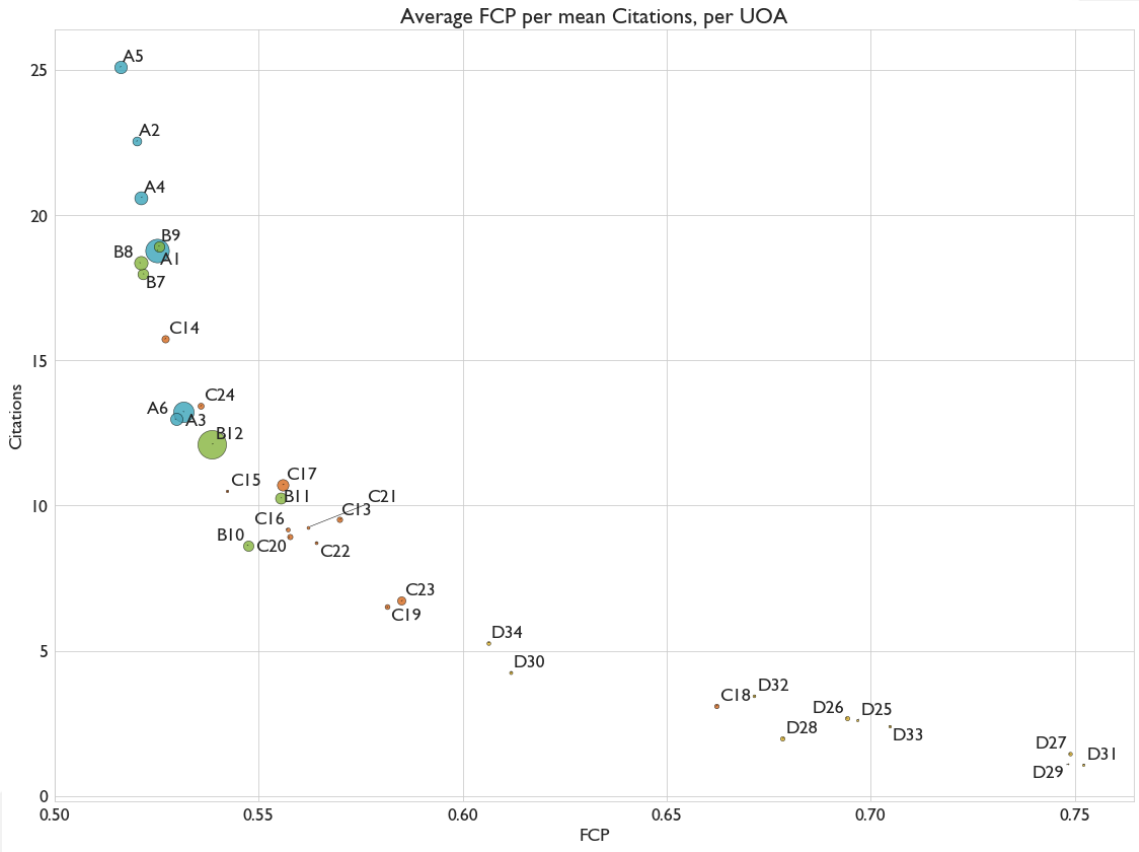
Machine Learning/NLP





Source

Citations within UoAs



Dimensions

Part of DIGITAL science