Service and Research at the Information Retrieval Service of the CPT Section (IVS-CPT)

Robin Haunschild

04/21/2023

Service at IVS-CPT

Main services

- Bibliometric analyses (e.g., reports about persons or MPIs)
- Special scientific literature searches (e.g., exhaustive publication set regarding a topic or a compound)
- Patent searches (e.g., prior art or FTO searches)
- . . .

Main databases

- Web of Science
- Scopus
- STN (CAS, Inspec, . . .); Derwent via Science IP
- OpenAlex (successor of Microsoft Academic Graph)
- . . .

Research at IVS-CPT

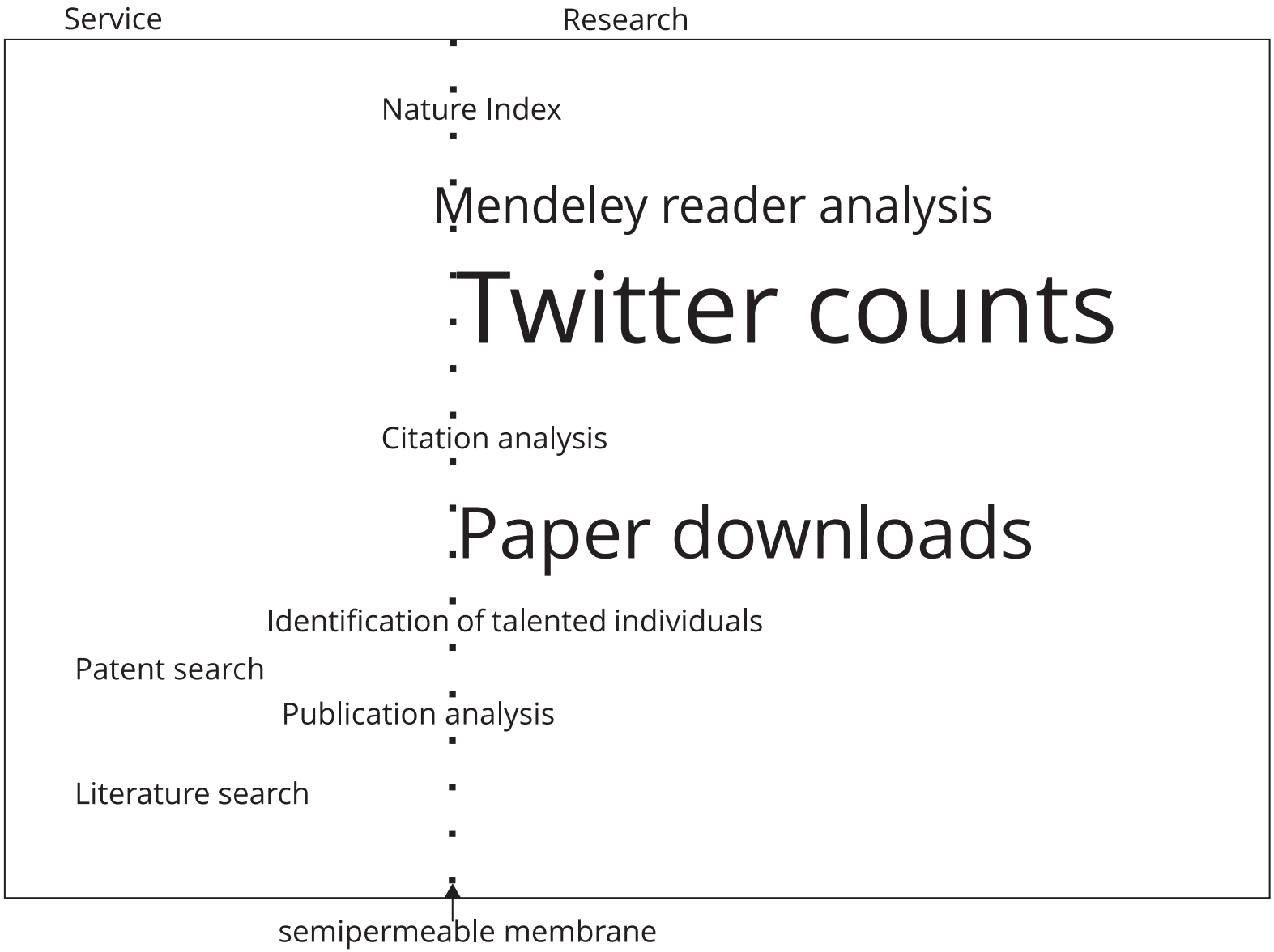
Main research activities

- Scientometrics, bibliometrics (e.g., analysis of research fields and testing/validating or proposing new indicators and methods)
- Chemical bibliometrics (e.g., counting chemical compounds or element occurrences instead of citations)
- Altmetrics (e.g., analysis of data from Twitter or Mendeley)
- . . .

Main databases

- Web of Science
- Scopus
- STN (CAS, Inspec, . . .)
- OpenAlex (successor of Microsoft Academic Graph)
- Dimensions
- Altmetric.com
- . . .

IVS-CPT cell

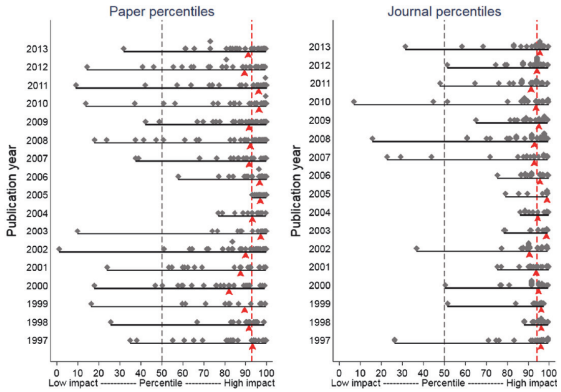

Service
Research
Nature Index
Mendeley reader analysis
Twitter counts
Citation analysis
Paper downloads
Identification of talented individuals
Patent search
Publication analysis
Literature search
semipermeable membrane

News from the Chemical Abstracts Service, CAS

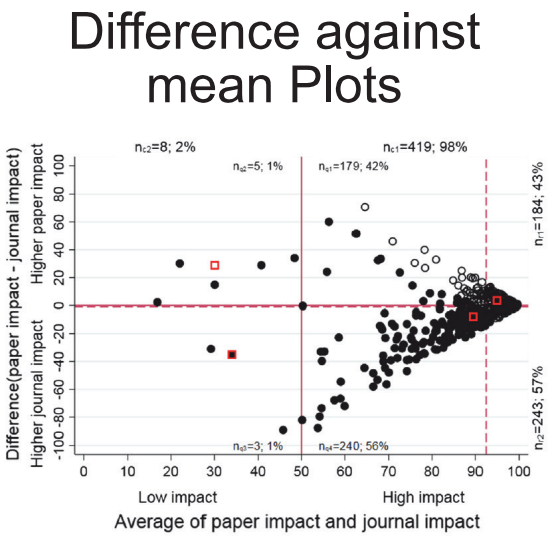
SciFinder vs. SciFinder^n

- The MPG has switched over to SciFinder^n last year.
- Next online trainings are probably next month. Topics will probably be:
  - structure search in CAS databases
  - retrosynthesis
  - polymers and materials
  - biosequences and bioterms
  - artificial intelligence in chemistry

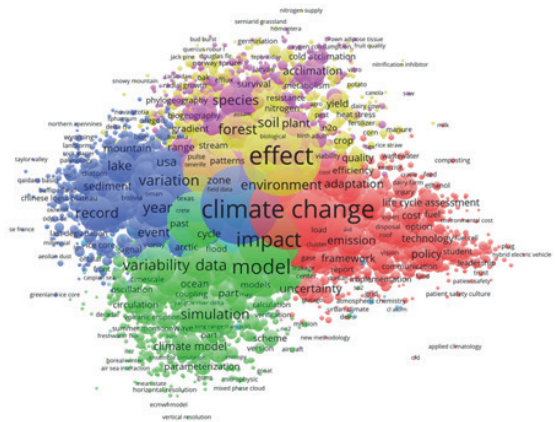
IVS-CPT research impressions



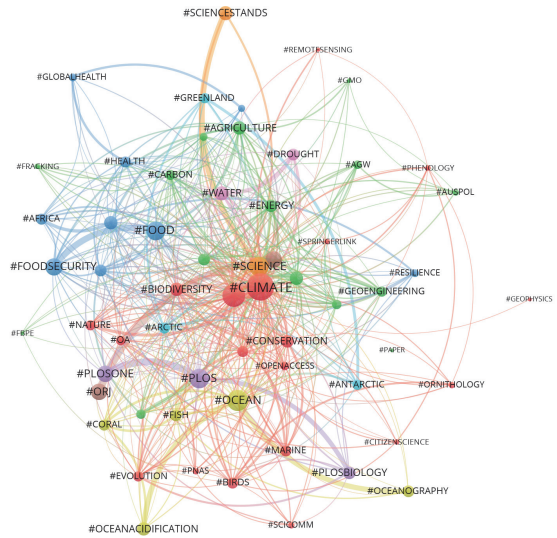
Beam Plots

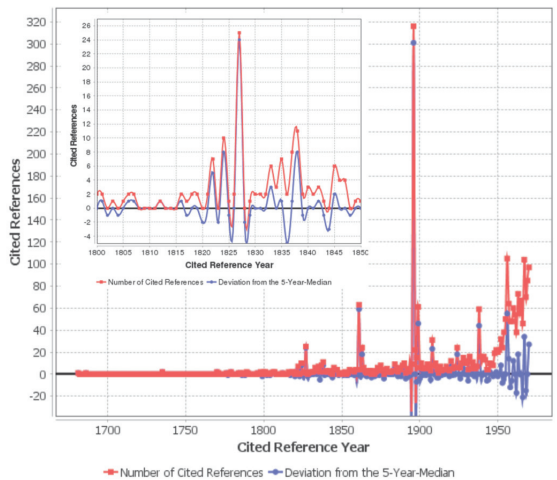


Chemical Bibliometrics



Word clouds and semantic maps





RPYS

How to identify young talented individuals in academia?

Possible search strategies

- Visit conferences
- Remember authors of influencial recent papers
- Ask colleagues
- …
- How about bibliometrics?
  - Citations need time to accrue
  - Need for indicators that depend only indirectly or not at all on citations
  - Restricted to areas that are well assessable by bibliometrics (e.g., natural sciences)

Our proposal

Identification of young talented individuals in the natural and life sciences using bibliometric data

Lutz Bornmann*,** and Robin Haunschild*

*R.Haunschild@fkf.mpg.de; l.bornmann@fkf.mpg.de
Max Planck Institute for Solid State Research, Heisenbergstr. 1, 70569 Stuttgart (Germany)

**bornmann@gv.mpg.de
Science Policy and Strategy Department, Administrative Headquarters of the Max Planck Society, Hofgartenstr. 8, 80539 Munich, (Germany)

Journal of Informetrics 17 (2023) 101394

Presented at an international conference (STI 2022) and published in JoI in an extended version.

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

Identification of potential young talented individuals in the natural and life sciences: A bibliometric approach

Robin Haunschild[a,*], Lutz Bornmann[a,b]

[a] Max Planck Institute for Solid State Research, Heisenbergstr. 1, Stuttgart 70569 Germany
[b] Science Policy and Strategy Department, Administrative Headquarters of the Max Planck Society, Hofgartenstr. 8, Munich 80539 Germany

High publication frequency

Sinatra, Wang, Deville, Song, and Barabási (2016)

examined publication records of scientists in physical sciences (2,887 physicists across at least 20 years) and 24,630 Google Scholar career profiles from multiple disciplines. The study points out the great relevance of publication productivity: **Growth of productivity is more pronounced for high-impact scientists and is modest for low-impact scientists.**

Sinatra, R., Wang, D., Deville, P., Song, C., & Barabási, A.-L. (2016). Quantifying the evolution of individual scientific impact. Science (New York, N.Y.), 354(6312), aaf5239. 10.1126/science.aaf5239.

High publication frequency

Zuckerman (1977) and Li, Yin, Fortunato, and Wang (2020)

investigated career paths of Nobel laureates and found that **“Nobel laureates were energetic producers from the outset, publishing almost twice as many papers as scientists in our comparison group”.**

Li, J., Yin, Y., Fortunato, S., & Wang, D. (2020). Scientific elite revisited: Patterns of productivity, collaboration, authorship and impact. Journal of The Royal Society Interface, 17(165), Article 20200135. 10.1098/rsif.2020.0135.
Zuckerman, H. (1977). Scientific elite. Nobel laureates in the united states. New York, NY, USA: Free Press.

Publishing in top journals

Lindahl (2018)

investigated career paths of 406 mathematicians from the field of number theory. For the author, the results of the study point out “the high relative importance of publishing in top journals and the indication that those who publish many articles in top journals, which implicitly require a high publication count, have a higher probability of attaining excellence. **These results suggest that publishing in top journals is very important in the process of attaining excellence in the early career in addition to publication volume**” (Lindahl, 2018, pp. 530-531).

Lindahl, J. (2018). Predicting research excellence at the individual level: The importance of publication rate, top journal publications, and top 10% publications in the case of early career mathematicians. Journal of Informetrics, 12(2), 518-533. 10.1016/j.joi.2018.04.002.

Publishing as corresponding author

van Dijk, Mannor, and Carey (2014), von Bartheld, Houmanfar, and Candido (2015), and Sánchez-Jiménez, Guerrero-Bote, and de Moya-Anegón (2017)

analyzed the author bylines and evaluated the later success of the authors (e.g., becoming a principle investigator). **They demonstrated the importance of being the main actor among co-authors.**

Sánchez-Jiménez, R., Guerrero-Bote, V. P., & de Moya-Anegón, F. (2017). The role of guarantor in scientific collaboration: The neighbourhood matters. Journal of Informetrics, 11(1), 103-116. 10.1016/j.joi.2016.11.004.

van Dijk, D., Manor, O., & Carey, L. B. (2014). Publication metrics and success on the academic job market. Current Biology, 24(11), R516-R517. 10.1016/j.cub.2014.04.039.

von Bartheld, C. S., Houmanfar, R., & Candido, A. (2015). Prediction of junior faculty success in biomedical research: Comparison of metrics and effects of mentoring programs. PeerJ, 3. 10.7717/peerj.1262.

Set indicators

Basic indicators

- O: output, number of papers
- Q1: number of papers in top quartile journals
- C: number of papers as corresponding author

Combinations of indicators

- OxQ1
- OxC
- Q1xC
- OxQ1xC

Talent and control groups

Talent group(s)

If a scientist surpasses the top 1% threshold for one of the indicators or their combinations in a specific field, this scientist belongs to the talent group of this specific field.

Control group(s)

If a scientist surpasses the top 10% threshold but does not surpass the top 5% threshold for one of the indicators or their combinations in a specific field, this scientist belongs to the control group.

Dataset

Publication data

Scopus snapshot from April 2021
Publication years: 1996-2020
Document types: article, review, and proceedings paper
45,709,395 publications
Exclusion of Scopus fields that belong to Arts and Humanities or Social Sciences or are multidisciplinary
Identification of authors via Scopus author ID

Grant data

Dimensions publications snapshot from January 2022
Dimensions grants snapshot from March 2022

Methodology

Indicator validation

- Calculation of basic and combined indicators for a cohort with their first publication between 1999 and 2003 for the first ten years of their publication activity
- Assignment of authors to talent and control groups
- Calculation of median Hazen percentiles based on citations for the time period ten years after their first publication until 2018
- Calculation of the number of publications that are linked to a grant for US-based authors for the time period ten years after their first publication until 2018
- Determine the best performing indicator (combination)

Identification of young talented individuals

- Identify the authors who surpass the top 1% threshold according to the best performing indicator (combination) for the cohort with their first publication between 2007 and 2011

Validation with Hazen percentiles

Authors with first publication between 1999 and 2003

Statistical values of the Hazen percentiles of papers for each group and indicator combination since ten years after the scientists' first publication until 2018.

Group	Indicator combination	1st quartile	Median	Mean	3rd quartile
Potentially talented individuals	O	43.41	57.59	56.45	70.80
	Q1	55.96	68.32	66.45	79.20
	C	40.39	55.29	53.87	68.89
	OxQ1	56.07	67.84	66.27	78.46
	OxC	43.67	57.53	56.04	70.17
	Q1xC	55.56	67.15	65.31	77.15
	OxQ1xC	55.87	67.13	65.47	77.06
Control group	O	36.69	53.10	52.20	68.29
	Q1	44.58	59.27	57.94	72.78
	C	35.18	51.65	50.81	66.84
	OxQ1	44.14	57.96	56.64	70.54
	OxC	37.37	52.66	51.95	67.19
	Q1xC	43.78	57.92	56.59	70.60
	OxQ1xC	44.38	58.04	56.61	70.47
Differences between potentially talented individuals and control group	O	6.72	4.49	4.24	2.51
	Q1	11.37	9.05	8.51	6.41
	C	5.21	3.65	3.06	2.05
	OxQ1	11.93	9.88	9.64	7.92
	OxC	6.31	4.87	4.09	2.98
	Q1xC	11.78	9.23	8.73	6.55
	OxQ1xC	11.49	9.08	8.86	6.59

Validation with grant data

Authors with first publication between 1999 and 2003

Statistical values of the number of papers that are linked to a grant from US-based author for each group and indicator combination since ten years after the scientists' first publication until 2018.

Group	Indicator combination	1st quartile	Median	Mean	3rd quartile
Potentially talented individuals	O	3.00	9.00	14.98	20.00
	Q1	3.00	8.00	13.83	18.00
	C	3.00	9.00	14.54	19.00
	OxQ1	4.00	10.00	16.89	23.00
	OxC	4.00	10.00	16.97	23.00
	Q1xC	4.00	10.00	17.26	23.00
	OxQ1xC	4.00	11.00	18.89	25.00
Control group	O	2.00	5.00	9.73	12.00
	Q1	2.00	6.00	10.73	14.00
	C	2.00	6.00	10.35	13.00
	OxQ1	2.00	6.00	10.56	14.00
	OxC	2.00	6.00	10.87	13.00
	Q1xC	2.00	7.00	10.62	14.00
	OxQ1xC	2.00	7.00	10.75	14.50
Differences between potentially talented individuals and control group	O	1.00	4.00	5.25	8.00
	Q1	1.00	2.00	3.11	4.00
	C	1.00	3.00	4.19	6.00
	OxQ1	2.00	4.00	6.33	9.00
	OxC	2.00	4.00	6.09	10.00
	Q1xC	2.00	3.00	6.64	9.00
	OxQ1xC	2.00	4.00	8.14	10.50

Dataset with young potentially talented individuals

First publication between 2007 and 2011

46,200 young potentially talented individuals that surpass the top 1% threshold of OxQ1:

- 8,026 with their first paper in 2007
- 8,854 with their first paper in 2008
- 9,380 with their first paper in 2009
- 9,529 with their first paper in 2010
- 10,411 with their first paper in 2011

https://doi.org/10.17617/1.98

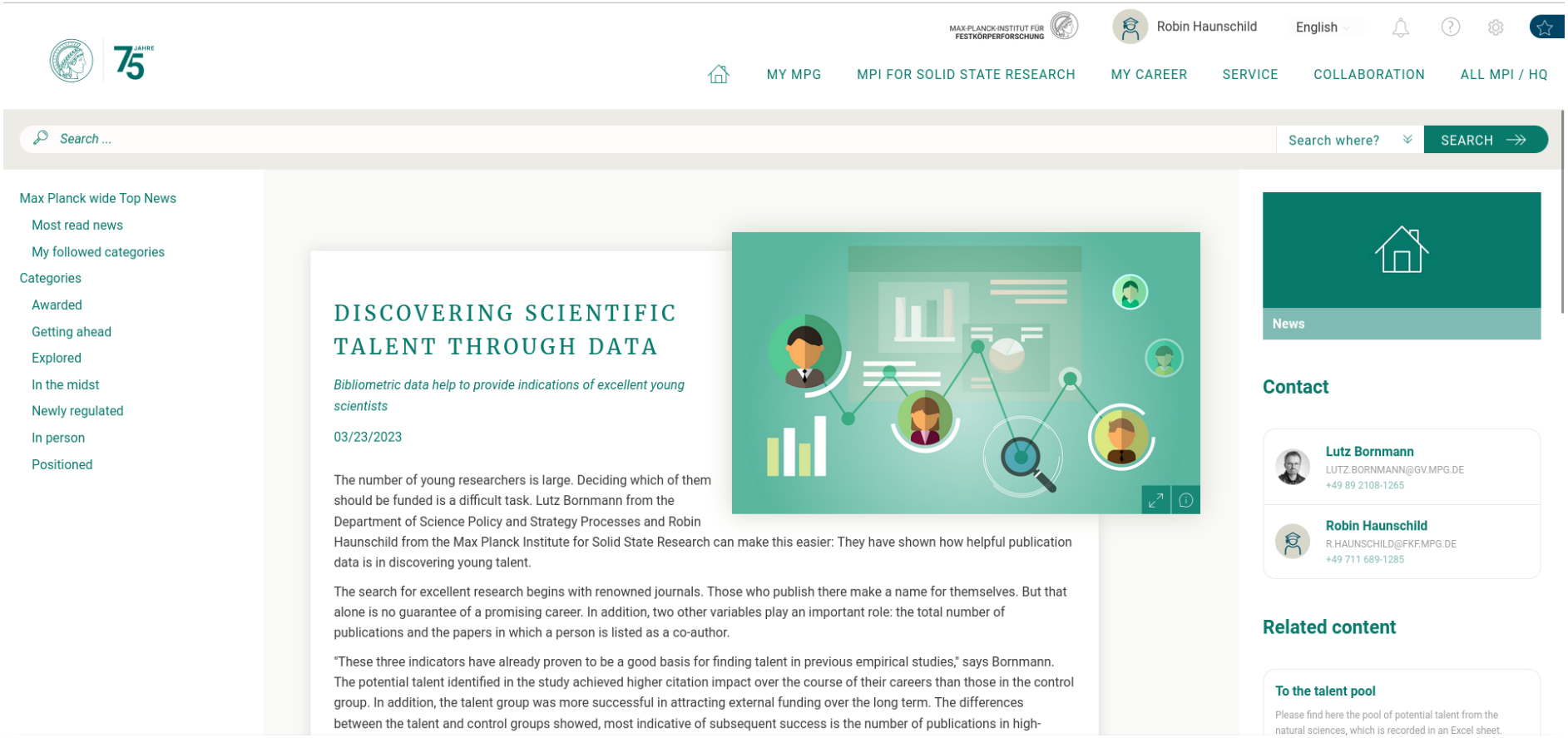
Dataset with additional information requested by the MPG scouting officers is available upon request for internal purposes.

Statistics about the talent dataset

Number of author IDs for each ASJC field in the potential talent database in alphabetical order of the ASJC name.

ASJC code	ASJC name	Number of author IDs	Percentage of author IDs
11	Agricultural and Biological Sciences	2,516	0.31
13	Biochemistry, Genetics and Molecular Biology	3,864	0.29
15	Chemical Engineering	2,072	0.38
16	Chemistry	2,977	0.37
17	Computer Science	4,716	0.46
35	Dentistry	213	0.34
19	Earth and Planetary Sciences	1,566	0.40
21	Energy	1,561	0.38
22	Engineering	6,192	0.36
23	Environmental Science	2,598	0.44
36	Health Professions	418	0.24
24	Immunology and Microbiology	809	0.19
25	Materials Science	3,800	0.43
26	Mathematics	1,398	0.25
27	Medicine	9,343	0.43
28	Neuroscience	756	0.25
29	Nursing	638	0.29
30	Pharmacology, Toxicology and Pharmaceutics	1,176	0.24
31	Physics and Astronomy	6,124	0.67
34	Veterinary	219	0.18
	Sum (including talented individuals overlapping multiple ASJCs)	52,737	
	Sum (unique talented individuals)	46,200	

Coverage in MAX by the MPG Press Office



https://max.mpg.de/News/Pages/Discovering-scientific-talent-through-data.aspx

Summary

Take-home messages

- IVS-CPT offers a broad range of services and performs research with the aim of improving services
- The MPG has switched from SciFinder to SciFinder^{n}
- In collaboration with the administrative headquarters, we have proposed a method for identification of young potentially talented individuals
- There will be young talented individuals outside our database
- Some authors in our database will not qualify as young talented individuals upon closer inspection
- Our methodology and the resulting dataset are intended to be one of many sources of information when identifying young potentially talented individuals

R.Haunschild@fkf.mpg.de @rhaunschild
@rhaunschild@social.mpdl.mpg.de
https://www.fkf.mpg.de/ivs