Service & Research at the IVS-CPT

Thomas Scheidsteger

December 7, 2018 – Directors' Board Meeting

Overview of Services & Research

Service

- Literature, Data & Patent search – **beyond Google**
- Bibliometric reports – **beyond Hirsch Index**

Research

- Bibliometrics – **beyond Web of Science**
- Altmetrics

Service : Search Tasks

Literature and Numerical Data:

- *Complex search*, which can NOT be done by the scientist themselves nor by librarians
- *Translation of the scientist's question* into retrieval procedure and matching with available data sources
- *Adaptation and translation of results* to the scientist's context

Patents:

- explore *state of the art* in research and technology
- *exhaustive search for publications or patents* that would *prohibit* the application for a patent
- collaboration with *Max Planck Innovation*

Service: Search Tools

General Tools for MPS:

- Web of Science (WoS) Core Collection, Scopus *Basic Search*
- *SciFinder* of Chemical Abstract Service (CAS)
- *Google Patents*
- *WoS: Derwent Innovations Index*

Advanced Tools for IVS:

- WoS, Scopus **Advanced Search Language**
- **STN** databases with **Elaborate Search Language**:
  - **CAS REGISTRY, CAplus, MARPAT** - beyond *SciFinder*
  - **INSPEC, . . .**
  - **DWPI** and **Patent Offices' Fulltext Databases**

Example Search
NOT feasible with *SciFinder*, but with STN

Search for alloys of four 3d transition metals

- *SciFinder*:
  - Structure editor: substructure search with rest groups restricted to alloys $\rightarrow$ more than 1 M hits
  - Exclusion of other elements by complement search not feasible, because of restriction of saving search to 20k item
- **STN**:
  - building a search strategy in some minutes $\rightarrow$ 7061 correct alloys found

Bibliometric reports: Approach

Professional approach acc. to Leiden manifesto :

Use of *field-normalized scores* in order to reflect impact of papers against the backdrop of their reference set - *papers published in the same field and at the same time*.

Normalized Citation Score (most used)

$$NCS = \frac{c_i}{e_i}$$

- c_i: citation count of a focal paper,
- e_i: corresponding average citation count in the scientific field and publication year *dependent on field definitions*

Hicks et al. (2015). The Leiden Manifesto for research metrics. Nature, 520(7548), 429.

Bibliometric reports: Portfolio & Tools

Quant. evaluation of output and impact of research units

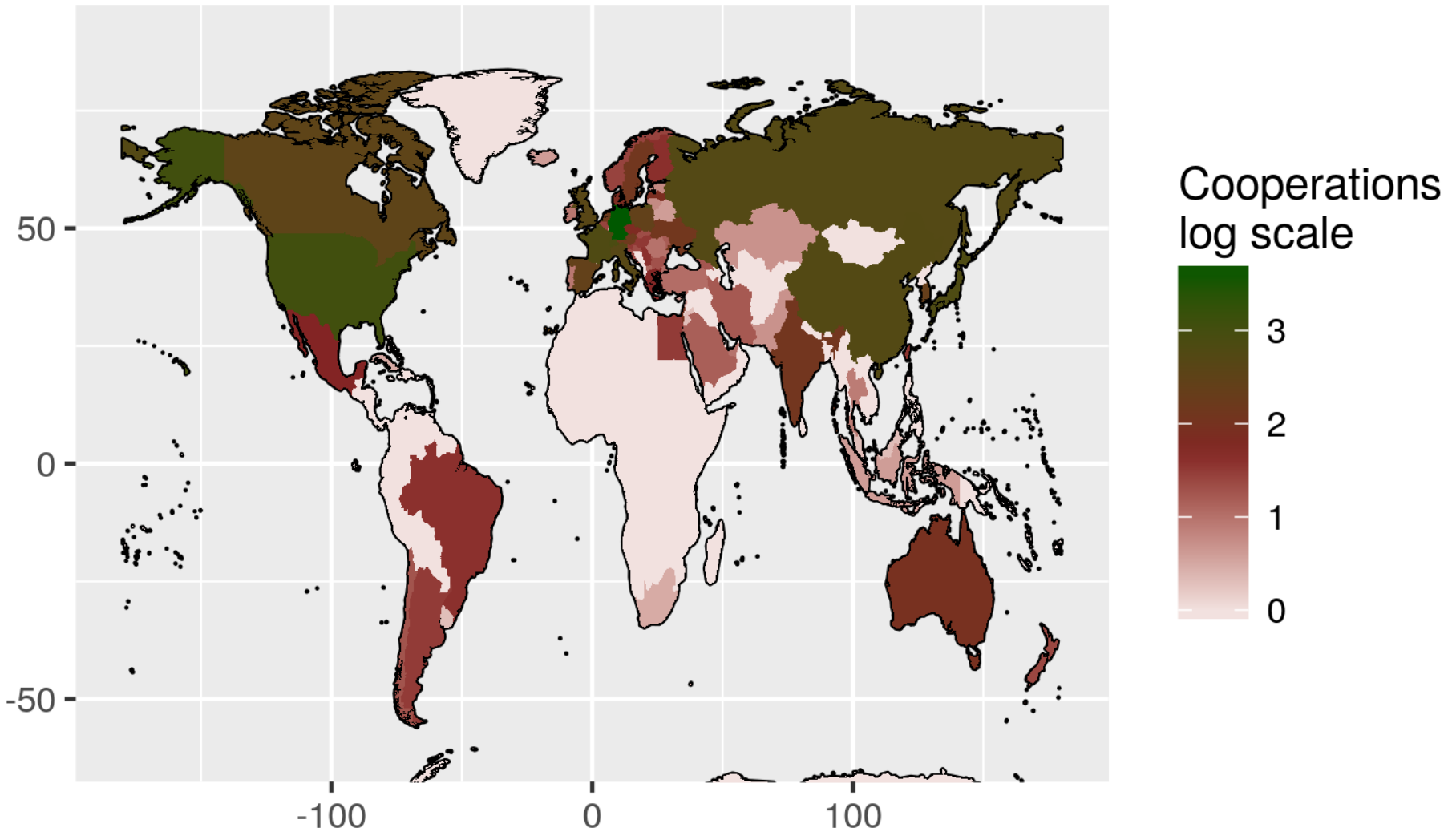
- single persons ($\rightarrow$ promotions, prizes, . . .)
- working groups | departments | centres
- whole institutes ($\rightarrow$ SAB)
- several institutes | whole CPT section | whole MPS
- countries

Tools:

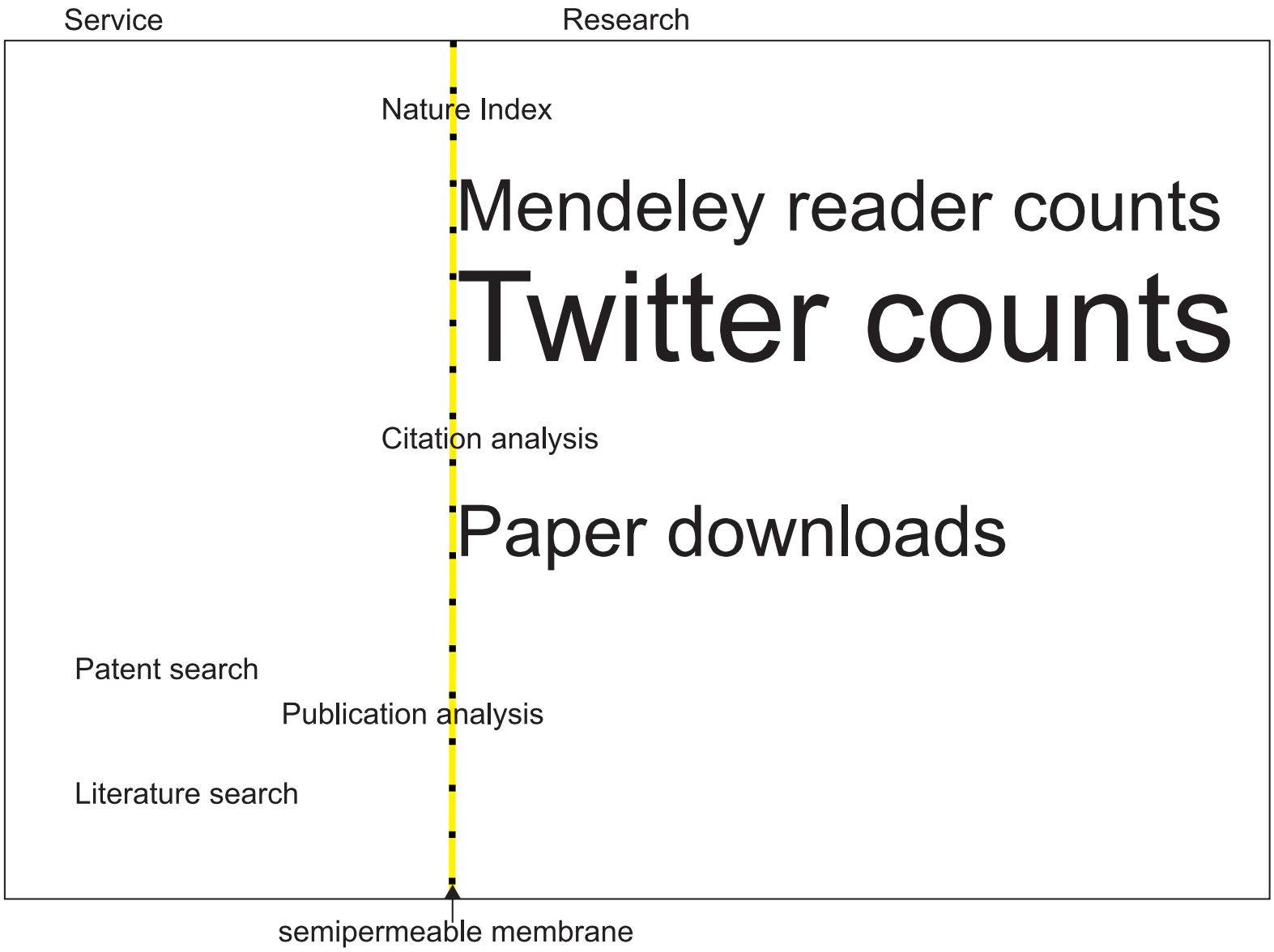
- WoS, Scopus **Advanced Search** with elaborate address queries
- **in-house database** developed and maintained in cooperation with the MPDL incl. **field-normalized indicators derived from WoS**

Example of output analysis: MPI's cooperation analysis

Collaborating countries



Which indicators from research are mature enough for service?



Research

Aims

- Bibliometrics:
  - development of new and better indicators
  - chemical bibliometrics
  - **comparison of databases & classification schemes**
- Altmetrics = alternative metrics:
  - Twitter counts
  - Mendeley reader counts
  - Paper downloads from publishers
  - ...

Databases

- Established: Web of Science, Scopus, STN (CAS, ...)
- New: Microsoft Academics , Dimensions (Digital Science)

Definition of fields - I

Intellectual approach

- Journal based
  - Pooling of Journal Sets acc. to Subject Classification
  - **WoS**, Scopus, ...
  - NCS(JS) as standard indicator
  - Problem: with multi-disciplinary journals
- Paper based
  - Assignment by authors or experts acc. to mono-disciplinary classification schemes
  - **Chemical Abstract^R Sections (CA)**, Physics and Astronomy Classification Scheme (PACS), ...
  - NCS(CA)
  - Problem: borders of disciplines

Definition of fields - II

Algorithmic Approach

- direct citation relations
  - multi-disciplinary classification scheme $\rightarrow$ unnamed clusters
  - NCS(CR) used for Leiden Ranking, ...
- paper based (**NEW**)
  - (unpublished) AI algorithm, partially based on semantic relatedness $\rightarrow$ hierarchical classification scheme
  - **Microsoft Academic (MA)**
  - NCS(MA)

Definition of fields - II

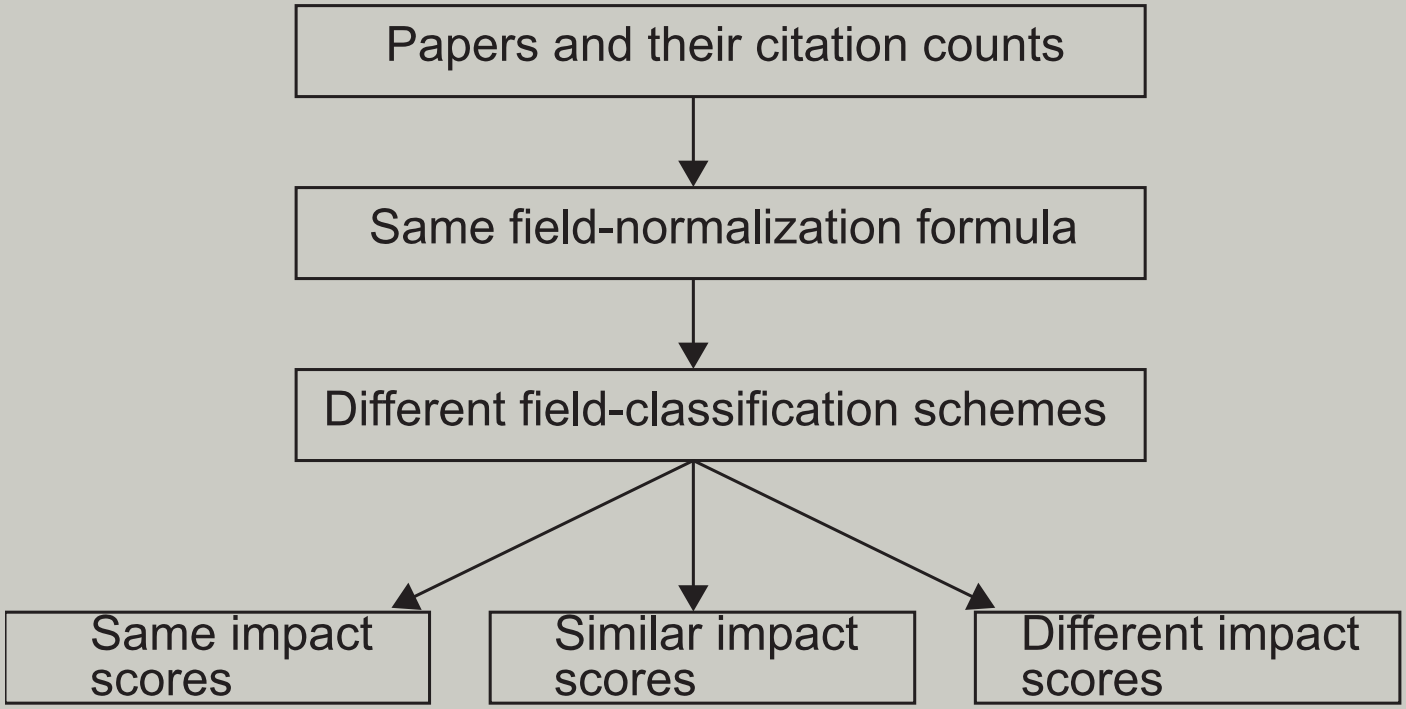
Algorithmic Approach

- direct citation relations
  - multi-disciplinary classification scheme $\rightarrow$ unnamed clusters
  - NCS(CR) used for Leiden Ranking, ...
- paper based (**NEW**)
  - (unpublished) AI algorithm, partially based on semantic relatedness $\rightarrow$ hierarchical classification scheme
  - **Microsoft Academic (MA)**
  - NCS(MA)

Research Question of two papers at STI 2018

Relationship between field-normalized indicators calculated with different approaches of field-categorization

How much do values of field-normalized indicators differ when different approaches of field-categorization are used and the formula for the field-normalized indicators is held constant?



Haunschild et al.: Focus on Chemistry

Definition of paper set to calculate NCS

- Matching via UT and DOI $\rightarrow$ approx. **2.7 M papers**
- statistical analysis, incl. correlation coefficients and assignment of **4 impact classes**:
  - poorly cited
  - fairly cited
  - remarkably cited
  - outstandingly cited

Results

Evaluation	impact classes
NCS(CA) v. NCS(JS)	**82.2%**
NCS(JS) v. NCS(CR)	74.7%
NCS(CA) v. NCS(CR)	73.3%

CA similar to JS, but other pairings significantly worse

Scheidsteger et al.: Focus on Computer Science

Wos vs. M

- 65 M papers in WoS vs. 165 M in MA
- 252 WoSSC vs. 292 Level 1 Fields of Study

Definition of paper set to calculate NCS

- 2157 papers of a German CS Institute (2005 - 2010)
- 20% matched via DOI in both databases

Results

Evaluation	Impact Classes
NCS(MA) v. NCS(JS)	**81.0%**
NCS(CA) v. NCS(JS)	82.2%

Information professionals

- Robin Haunschild, 2M11,

- Thomas Scheidsteger, 2M9

- Offices in the Library - **Drop by!**
- ivs-cpt@fkf.mpg.de

References

- Cumming, G. (2012). *Understanding the new statistics: effect sizes, confidence intervals, and meta-analysis.* London, UK: Routledge.
- Glanzel, W., Debackere, K., & Thijs, B. (2016). *Citation classes: a novel indicator base to classify scientific output.* doi: 10.1007/s11192-017-2535-3
- Hicks et al. (2015). The Leiden Manifesto for research metrics. Nature, 520(7548), 429.
- **Haunschild et al. (2018). Relationship between field-normalized indicators calculated with different approaches of field-categorization, STI 2018 Conference Proceedings. https://openaccess.leidenuniv.nl/handle/1887/65267**
- Lin, L. I. (1989). *A concordance correlation-coefficient to evaluate reproducibility.* Biometrics, 45(1), 255-268. doi: 10.2307/2532051
- Koch, R., & Sporl, E. (2007). *Statistical methods for comparison of two measuring procedures and for calibration: Analysis of concordance, correlation and regression in the case of measuring intraocular pressure.* Klinische Monatsblatter Fur Augenheilkunde, 224(1), 52-57. doi: 10.1055/s-2006-927278.
- **Scheidsteger et al. (2018). The concordance of field-normalized scores based on Web of Science and Microsoft Academic data: A case study in computer sciences. STI 2018 Conference Proceedings. https://openaccess.leidenuniv.nl/handle/1887/65358**


53 23 16
IVS