

# DEKiF: Case Study 4

## Berlin Metropolregion

Aliakbar Akbaritabar (Ali)

Akbaritabar@DZHW.eu

17 October, 2019



German Centre for Higher Education Research  
and Science Studies ■

# Two draft reports are prepared

- Look at Faust folder, AP9-4

Aliakbar Akbaratabar

DEKIF Project

Draft Report of Case Study 4:  
Berlin Metropolregion

Report to DEKIF Berlin Team  
June 2019



Aliakbar Akbaratabar

DEKIF Project

Disambiguating and Geocoding WOS  
Organizations with Wikidata and GRID

Case Study 4 on Berlin Metropolregion  
September 2019



# Brief on data and methods

- ① A preliminary view to Berlin region's scientific output
- ② All WOS/Scopus 1990-2017 (b\_2018 KB) pubs
- ③ Article, Review and Conference proceeding as document types
- ④ With at least one author/institution from Berlin, Germany
- ⑤ Number of publications, fractional count, 3 years citations
- ⑥ Disambiguation & geocoding report using WOS 2018, AR/RE
- ⑦ Wikidata 27<sup>th</sup> March 2019, GRID 17<sup>th</sup> February 2019

# Brief on data and methods

- ① A preliminary view to Berlin region's scientific output
- ② All WOS/Scopus **1990-2017** (b\_2018 KB) pubs
- ③ *Article, Review and Conference proceeding* as document types
- ④ With at least one author/institution from **Berlin, Germany**
- ⑤ Number of publications, fractional count, 3 years citations
- ⑥ Disambiguation & geocoding report using WOS 2018, AR/RE
- ⑦ Wikidata 27<sup>th</sup> March 2019, GRID 17<sup>th</sup> February 2019

# Brief on data and methods

- ① A preliminary view to Berlin region's scientific output
- ② All WOS/Scopus **1990-2017** (b\_2018 KB) pubs
- ③ *Article, Review and Conference proceeding* as **document types**
- ④ With at least one author/institution from **Berlin, Germany**
- ⑤ Number of publications, fractional count, 3 years citations
- ⑥ Disambiguation & geocoding report using WOS 2018, AR/RE
- ⑦ Wikidata 27<sup>th</sup> March 2019, GRID 17<sup>th</sup> February 2019

# Brief on data and methods

- ① A preliminary view to Berlin region's scientific output
- ② All WOS/Scopus **1990-2017** (b\_2018 KB) pubs
- ③ *Article, Review and Conference proceeding* as **document types**
- ④ With at least one author/institution from **Berlin, Germany**
- ⑤ Number of publications, fractional count, 3 years citations
- ⑥ Disambiguation & geocoding report using WOS 2018, AR/RE
- ⑦ Wikidata 27<sup>th</sup> March 2019, GRID 17<sup>th</sup> February 2019

# Brief on data and methods

- ① A preliminary view to Berlin region's scientific output
- ② All WOS/Scopus **1990-2017** (b\_2018 KB) pubs
- ③ *Article, Review and Conference proceeding* as **document types**
- ④ With at least one author/institution from **Berlin, Germany**
- ⑤ Number of publications, fractional count, 3 years citations
- ⑥ Disambiguation & geocoding report using WOS 2018, AR/RE
- ⑦ Wikidata 27<sup>th</sup> March 2019, GRID 17<sup>th</sup> February 2019

# Brief on data and methods

- ① A preliminary view to Berlin region's scientific output
- ② All WOS/Scopus **1990-2017** (b\_2018 KB) pubs
- ③ *Article, Review and Conference proceeding* as **document types**
- ④ With at least one author/institution from **Berlin, Germany**
- ⑤ Number of publications, fractional count, 3 years citations
- ⑥ Disambiguation & geocoding report using WOS 2018, AR/RE
- ⑦ Wikidata 27<sup>th</sup> March 2019, GRID 17<sup>th</sup> February 2019

# Brief on data and methods

- ① A preliminary view to Berlin region's scientific output
- ② All WOS/Scopus **1990-2017** (b\_2018 KB) pubs
- ③ *Article, Review and Conference proceeding* as **document types**
- ④ With at least one author/institution from **Berlin, Germany**
- ⑤ Number of publications, fractional count, 3 years citations
- ⑥ Disambiguation & geocoding report using WOS 2018, AR/RE
- ⑦ **Wikidata** 27<sup>th</sup> March 2019, **GRID** 17<sup>th</sup> February 2019

# Unique organizations (problematic?!)

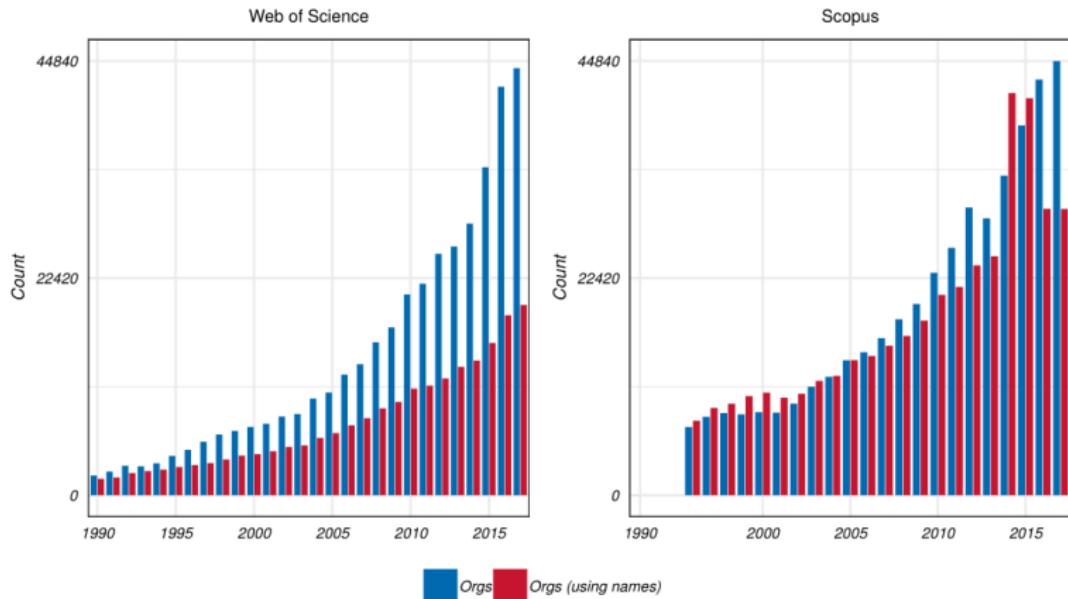
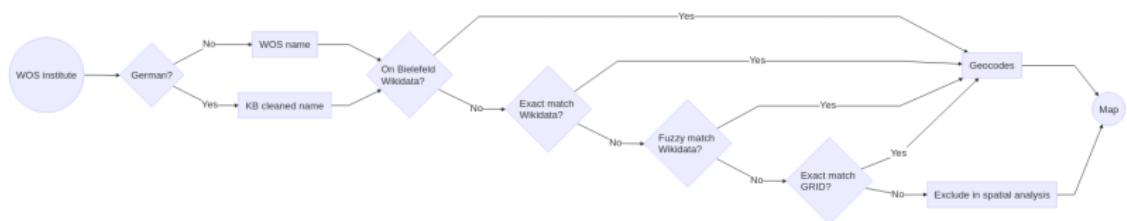


Figure 10: Unique organizations with which Berlin region institutes and universities have collaborated in Articles, Reviews and Conference proceedings in WOS and Scopus in 1990 - 2017

# Matching logic

- First, disambiguation & geocoding!



# Exact and fuzzy matching with GRID and Wikidata

DZHW

Table 1: Descriptive metrics on Berlin metropolitan region organizations, countries and cities (WOS, only article and reviews and where organization was not previously disambiguated by Rimmert et al's project)

Metric	Value
# FK_INSTITUTIONS	287087
# with unique KB name	1028
# Without KB name	286060
# unique WOS country	201
# unique WOS city	11299
# unique German orgs	21726
# unique orgs located in Berlin	10554
# unique WOS ORGANIZATION1	72486
# Normalized name (baseline)	72609
# unique WOS orgs exact match with Wikidata (%)	42665 (58.76%)
# unique Wikidata organizations (exact match disambiguation result)	4444
# unique WOS orgs exact match with GRID (%)	45409 (62.54%)
# unique Wikidata organizations (GRID match disambiguation result)	4614
# unique WOS orgs FUZZY match with Wikidata (%)	60841 (83.79%)
# unique Wikidata organizations (FUZZY match disambiguation result)	7004

# International organization example (1/2)



PK_INSTITUTIONS	PK_KB_INST	KB_NAME	ORGANIZATION1	fuzzy_match_wiki_id	fuzzy_match_wiki_name	fuzzy_jw_level
All	All	All	aalborg univ	All	All	All
2 29009365	NA	NA	aalborg univ	Q601956	aalborg university	0.85 <  JW < 1
3 1077601	NA	NA	aalborg univ	Q601956	aalborg university	0.85 <  JW < 1
4 30422015	NA	NA	aalborg univ	Q601956	aalborg university	0.85 <  JW < 1
6 8497964	NA	NA	aalborg univ	Q601956	aalborg university	0.85 <  JW < 1
7 3823950	NA	NA	aalborg univ	Q601956	aalborg university	0.85 <  JW < 1
8 5559997	NA	NA	aalborg univ	Q601956	aalborg university	0.85 <  JW < 1
9 20140789	NA	NA	aalborg univ	Q601956	aalborg university	0.85 <  JW < 1
10 2583771	NA	NA	aalborg univ	Q601956	aalborg university	0.85 <  JW < 1
11 21284558	NA	NA	aalborg univ	Q601956	aalborg university	0.85 <  JW < 1
12 27026789	NA	NA	aalborg univ	Q601956	aalborg university	0.85 <  JW < 1
14 34366196	NA	NA	aalborg univ	Q601956	aalborg university	0.85 <  JW < 1
15 23317559	NA	NA	aalborg univ	Q601956	aalborg university	0.85 <  JW < 1
16 24242911	NA	NA	aalborg univ	Q601956	aalborg university	0.85 <  JW < 1
17 28620165	NA	NA	aalborg univ	Q601956	aalborg university	0.85 <  JW < 1
18 4170458	NA	NA	aalborg univ	Q601956	aalborg university	0.85 <  JW < 1
19 6898547	NA	NA	aalborg univ	Q601956	aalborg university	0.85 <  JW < 1
20 33059701	NA	NA	aalborg univ	Q601956	aalborg university	0.85 <  JW < 1
22 24567558	NA	NA	aalborg univ	Q601956	aalborg university	0.85 <  JW < 1
23 10351477	NA	NA	aalborg univ	Q601956	aalborg university	0.85 <  JW < 1
24 17626795	NA	NA	aalborg univ	Q601956	aalborg university	0.85 <  JW < 1
25 22866279	NA	NA	aalborg univ	Q601956	aalborg university	0.85 <  JW < 1
26 7812723	NA	NA	aalborg univ	Q601956	aalborg university	0.85 <  JW < 1
27 10269492	NA	NA	aalborg univ	Q601956	aalborg university	0.85 <  JW < 1
28 14067026	NA	NA	aalborg univ	Q601956	aalborg university	0.85 <  JW < 1
30 27769743	NA	NA	aalborg univ	Q601956	aalborg university	0.85 <  JW < 1
31 18178514	NA	NA	aalborg univ	Q601956	aalborg university	0.85 <  JW < 1
33 4883551	NA	NA	aalborg univ	Q601956	aalborg university	0.85 <  JW < 1
34 16148970	NA	NA	aalborg univ	Q601956	aalborg university	0.85 <  JW < 1
35 26011466	NA	NA	aalborg univ	Q601956	aalborg university	0.85 <  JW < 1
36 23019925	NA	NA	aalborg univ	Q601956	aalborg university	0.85 <  JW < 1

## German organization example (2/2)

	FK_INSTITUTIONS	PK_KB_INST	KB_NAME	ORGANIZATION1	CITY	COUNTRYCODE	POSTALCODE
1	24966247	NA	NA	alexander von humboldt inst internet & gesell	berlin	deu	D-10117
2	26263851	NA	NA	alexander von humboldt inst internet & gesell	berlin	deu	NA
3	25284785	NA	NA	alexander von humboldt inst internet & soc	berlin	deu	NA
4	19041909	NA	NA	alexander von humboldt inst internet & soc	berlin	deu	D-10117
5	23459193	NA	NA	alexander von humboldt inst internet & soc	berlin	deu	NA
6	9814790	NA	NA	alexander von humboldt inst internet & soc	berlin	deu	D-10117
7	5548014	NA	NA	alexander von humboldt inst internet & soc	berlin	deu	NA
8	32465471	NA	NA	alexander von humboldt inst internet & soc	berlin	deu	D-10117
9	6357212	NA	NA	alexander von humboldt inst internet & soc	berlin	deu	NA
10	2595255	NA	NA	alexander von humboldt inst internet & soc hiig	berlin	deu	NA



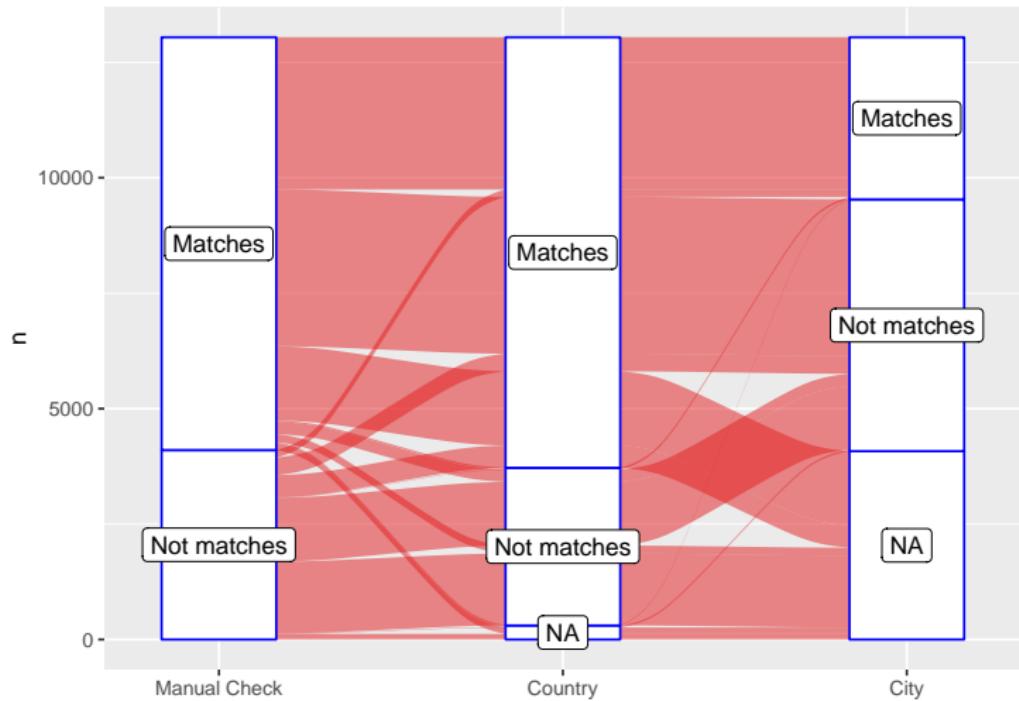
fuzzy_match_wiki_id	fuzzy_match_wiki_name	fuzzy_jw_level	fuzzy_city_status
NA	NA	NA	NA
NA	NA	Only matched with English Wikipedia, "gesell"	NA
Q30261359	alexander von humboldt institute for internet and society	0.85 < JW < 1	Matches
Q30261359	alexander von humboldt institute for internet and society	0.85 < JW < 1	Matches
Q30261359	alexander von humboldt institute for internet and society	0.85 < JW < 1	Matches
Q30261359	alexander von humboldt institute for internet and society	0.85 < JW < 1	Matches
Q30261359	alexander von humboldt institute for internet and society	0.85 < JW < 1	Matches
Q30261359	alexander von humboldt institute for internet and society	0.85 < JW < 1	Matches

# Comparative view to match procedure performances

Table 2: Comparative view to exact, Fuzzy and GRID match performances in case of each WOS institution

Exact Match	GRID Match	Fuzzy Match	Count
Not match found	Not match found	Best match	32,076
Not match found	Best match	Not match found	2,292
Not match found	Best match	Best match	3,688
Not match found	Best match with $0.85 < JW$	Not match found	329
Not match found	Best match with $0.85 < JW$	Best match	800
Best match	Not match found	Not match found	1,568
Best match	Not match found	Best match	1,554
Best match	Best match	Not match found	9,839
Best match	Best match	Best match	16,541
Best match	Best match with $0.85 < JW$	Not match found	385
Best match	Best match with $0.85 < JW$	Best match	130
Best match with $0.85 < JW$	Not match found	Not match found	6
Best match with $0.85 < JW$	Not match found	Best match	1,237
Best match with $0.85 < JW$	Best match	Not match found	595
Best match with $0.85 < JW$	Best match	Best match	1,370
Best match with $0.85 < JW$	Best match with $0.85 < JW$	Not match found	5,995
Best match with $0.85 < JW$	Best match with $0.85 < JW$	Best match	3,445

# Manual check of match procedure performances



# Countries match performances

**Table 3: Comparative view to exact, Fuzzy and GRID match performances in case of 20 countries with highest number of unique WOS institutions (limited to Berlin collaborations sample)**

COUNTRYCODE	FK_INSTITUTIONS	ORGANIZATION1	Normalized name	WOS vs Wikidata (%)*	Wikidata orgs	WOS vs GRID (%)*	GRID orgs	WOS FUZZY vs Wikidata (%)*	Wikidata orgs (FUZZY)
DEU	31,788	21,631	21,726	1204 (5.54%)	706	1300 (5.98%)	765	2100 (9.67%)	1,052
USA	44,958	7,648	7,648	12011 (157.05%)	896	12570 (164.36%)	1,015	15662 (204.79%)	1,587
FRA	23,871	4,915	4,915	2600 (52.9%)	315	2304 (46.88%)	328	4364 (88.79%)	448
ITA	17,007	3,935	3,935	1410 (35.83%)	177	1349 (34.28%)	169	3045 (77.38%)	320
GBR	22,147	3,930	3,930	3990 (101.53%)	471	4104 (104.43%)	499	5891 (149.9%)	754
ESP	11,705	2,779	2,779	1164 (41.89%)	123	479 (17.24%)	134	1192 (42.89%)	230
CHE	10,675	2,212	2,212	992 (44.85%)	132	1109 (50.14%)	150	1369 (61.89%)	207
NLD	11,676	2,068	2,068	2420 (117.02%)	165	2172 (105.03%)	185	2473 (119.58%)	227
AUT	7,657	1,657	1,657	743 (44.84%)	74	792 (47.8%)	85	620 (37.42%)	136
RUS	5,747	1,582	1,582	571 (36.09%)	82	535 (33.82%)	78	787 (49.75%)	146
BEL	6,004	1,308	1,308	371 (28.36%)	99	347 (26.53%)	103	1157 (88.46%)	141
CHN	6,345	1,235	1,235	2309 (186.96%)	213	2332 (188.83%)	218	2210 (178.95%)	274
JPN	6,973	1,234	1,234	2433 (197.16%)	196	2836 (229.82%)	208	1805 (146.27%)	245
CAN	6,643	1,042	1,042	1167 (112%)	92	1493 (143.28%)	90	2016 (193.47%)	192
POL	4,390	978	978	223 (22.8%)	46	540 (55.21%)	40	468 (47.85%)	60
AUS	5,944	945	945	1353 (143.17%)	121	1418 (150.05%)	126	2180 (230.69%)	215
BRA	3,604	925	925	65 (7.03%)	33	87 (9.41%)	38	289 (31.24%)	69
IND	2,096	894	894	349 (39.04%)	120	350 (39.15%)	119	430 (48.1%)	167
SWE	5,896	886	886	1134 (127.99%)	64	2248 (253.72%)	73	2083 (235.1%)	106
DNK	4,454	655	655	358 (54.66%)	45	883 (134.81%)	57	810 (123.66%)	87

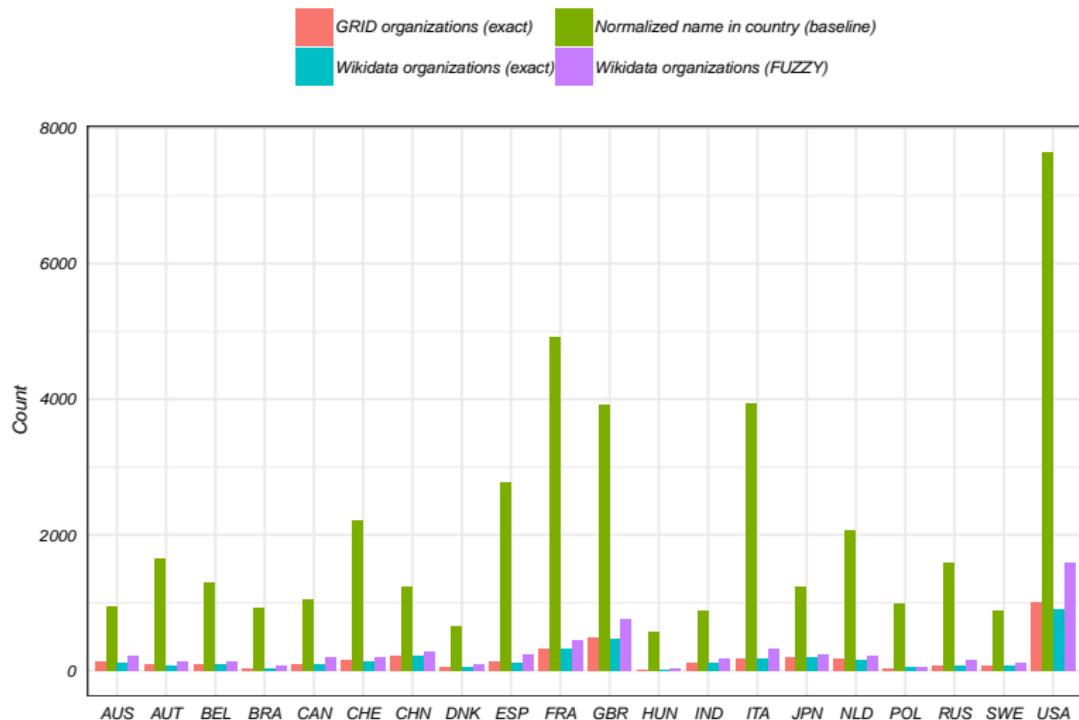
Note:

\* These are the number of matches detected in Wikidata or GRID for the unique WOS organizations.

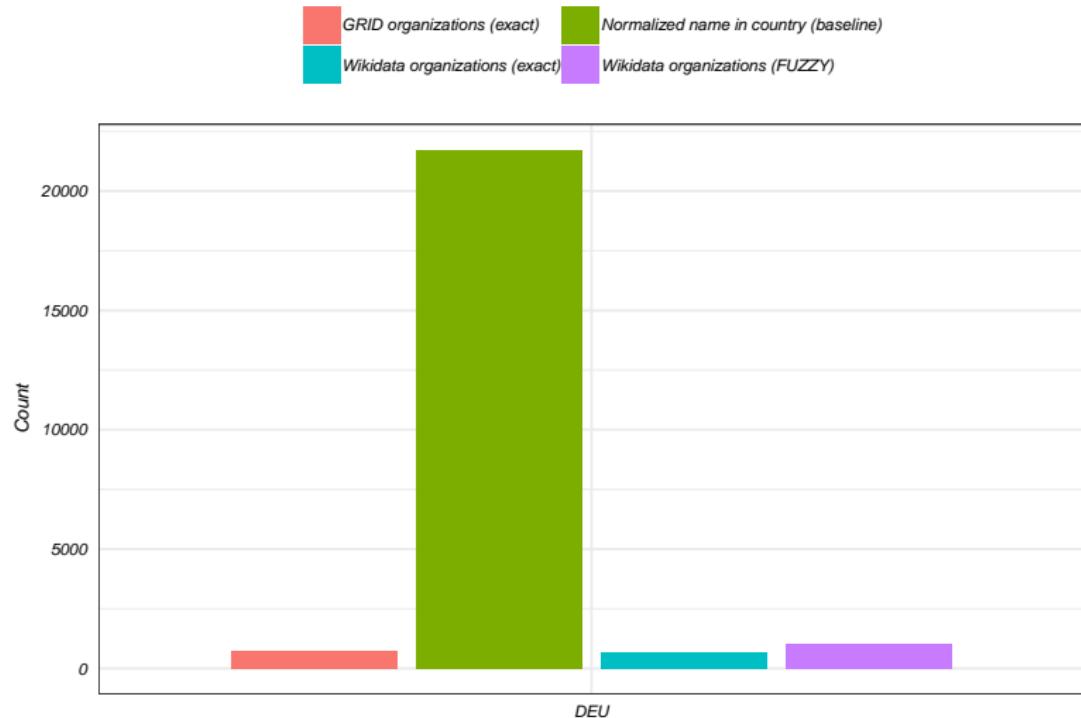
\* There could be more than one match detected for any organization.

\* Therefore counts and percentages can be higher than unique WOS organizations.

# Countries match performances



# How about match performance for Germany?

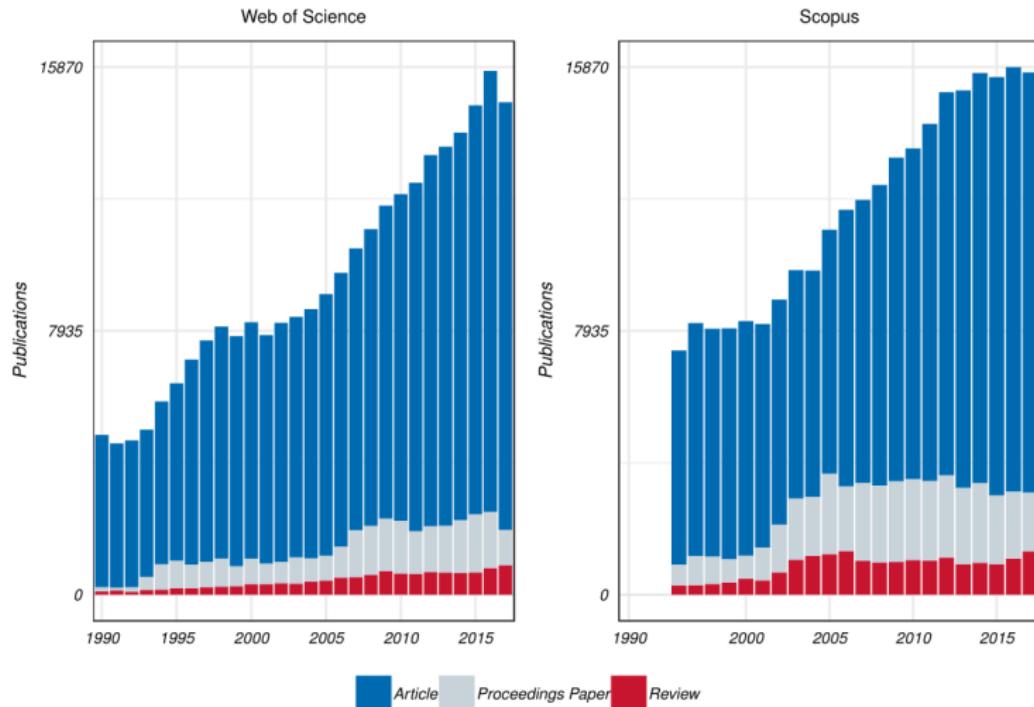


## Back to AP9-4: Berlin sample description

Table 1: Descriptive metrics on Berlin metropolitan region articles, organizations, countries and cities (WOS and Scopus from 1990-2017)

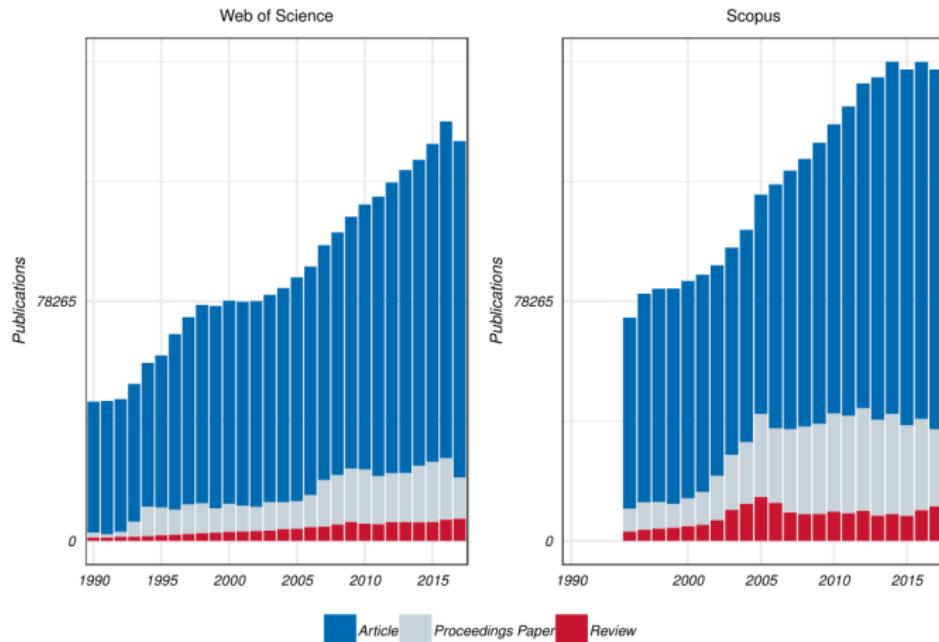
Metric	Value
Articles and Reviews and proceedings (WOS)	265,004
Articles and Reviews and proceedings (Scopus)	256,909
Organisations (WOS)	283,745
Organisations (Scopus)	356,918
Countries (WOS)	198
Countries (Scopus)	204
Cities (WOS)	14,313
Cities (Scopus)	72,053

# Berlin sample publications



# German publications (baseline)

- Berlin follows/leads the general trend



# Berlin sample publications (fractional countries)

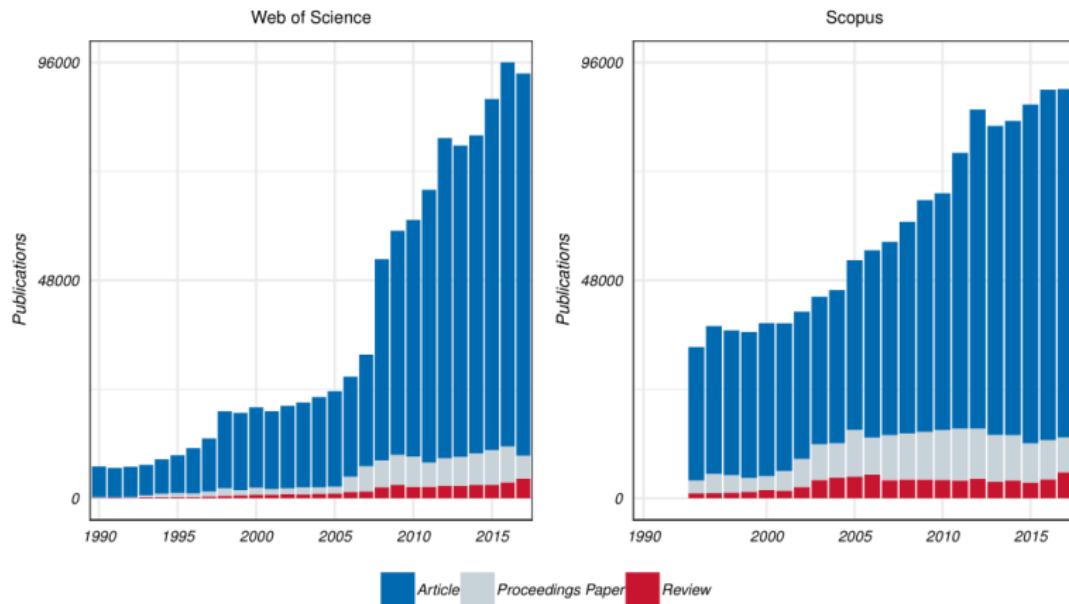


Figure 4: Articles, Reviews and Conference proceedings published by Berlin region institutes and universities in WOS and Scopus in 1990 - 2017 (Fractional count based on countries)

# Berlin sample publications (fractional orgs)

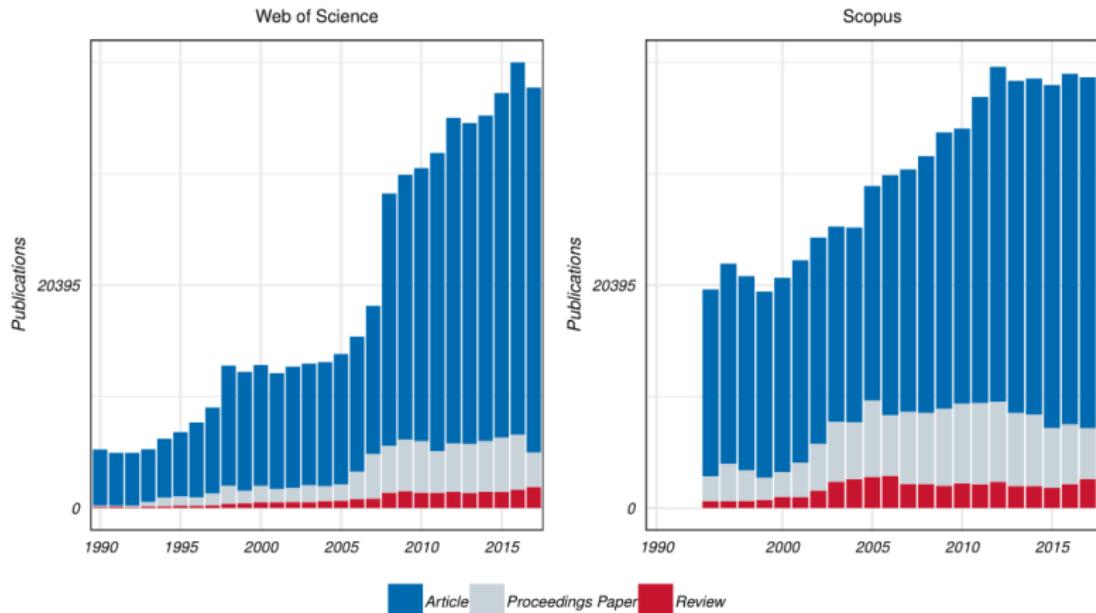


Figure 5: Articles, Reviews and Conference proceedings published by Berlin region institutes and universities in WOS and Scopus in 1990–2017 (Fractional count based on institutes)

# Disciplines with most pubs (WOS/Scopus)

- Hard sciences leading

Subject Classification	Number of pubs
Biochemistry & Molecular Biology	19,576
Physics, Applied	16,375
Chemistry, Physical	15,908
Physics, Condensed Matter	15,483
Materials Science, Multidisciplinary	13,938
Engineering, Electrical & Electronic	13,430
Chemistry, Multidisciplinary	12,934
Oncology	11,914
Neurosciences	11,813
Pharmacology & Pharmacy	11,232
Optics	10,869
Surgery	9,811
Cardiac & Cardiovascular Systems	9,719
Immunology	9,704
Cell Biology	9,037
Physics, Atomic, Molecular & Chemical	8,666
Clinical Neurology	8,590
Medicine, General & Internal	8,487
Physics, Multidisciplinary	7,924
Multidisciplinary Sciences	7,240

Subject Classification	Number of pubs
Condensed Matter Physics	22,140
Electronic, Optical and Magnetic Materials	15,957
Electrical and Electronic Engineering	15,385
General Medicine	12,858
General Physics and Astronomy	11,839
Biochemistry	11,415
General Chemistry	10,601
Molecular Biology	9,938
Physical and Theoretical Chemistry	9,326
Atomic and Molecular Physics, and Optics	9,140
General Materials Science	8,386
Genetics	7,285
Immunology	7,245
General Biochemistry, Genetics and Molecular Biology	7,145
Cardiology and Cardiovascular Medicine	7,119
Computer Science Applications	6,919
Applied Mathematics	6,789
Surgery	6,713
Cell Biology	6,708
Materials Chemistry	6,613

# Journals with most pubs (WOS/Scopus)

- Plos & hard science journals leading

Journal Name	Number of pubs
PLOS ONE	2,226
PHYSICAL REVIEW B	2,132
PHYSICAL REVIEW LETTERS	2,015
APPLIED PHYSICS LETTERS	1,591
JOURNAL OF BIOLOGICAL CHEMISTRY	1,146
JOURNAL OF CHEMICAL PHYSICS	1,113
JOURNAL OF APPLIED PHYSICS	1,010
ANGEWANDTE CHEMIE-INTERNATIONAL EDITION	978
SURFACE SCIENCE	976
PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA	930
PHYSICAL REVIEW B	899
JOURNAL OF CRYSTAL GROWTH	854
PHYSICAL CHEMISTRY CHEMICAL PHYSICS	771
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY	769
CHEMICAL PHYSICS LETTERS	725
DEUTSCHE MEDIZINISCHE WOCHENSCHRIFT	716
PHYSICAL REVIEW A	652
SCIENTIFIC REPORTS	645
PHYSICS LETTERS B	640
JOURNAL OF PHYSICS-CONDENSED MATTER	596

Journal Name	Number of pubs
Physical Review B - Condensed Matter and Materials Physics	2,569
Proceedings of SPIE - The International Society for Optical Engineering	2,545
Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)	2,390
PLoS ONE	2,275
Physical Review Letters	1,905
Applied Physics Letters	1,464
Optics InfoBase Conference Papers	1,153
Journal of Biological Chemistry	1,059
Angewandte Chemie - International Edition	961
Journal of Chemical Physics	948
Proceedings of the National Academy of Sciences of the United States of America	917
Journal of Applied Physics	890
Physical Review A - Atomic, Molecular, and Optical Physics	787
Physical Chemistry Chemical Physics	756
Surface Science	742
Physica B: Condensed Matter	709
Journal of the American Chemical Society	685
Journal of Crystal Growth	657
Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz	657
Scientific Reports	617

# Organizations with most pubs (WOS/Scopus)

- Charite & Freie leading

Table 2: 50 Most prolific organizations located in Berlin from Web of Science in 1990 - 2017

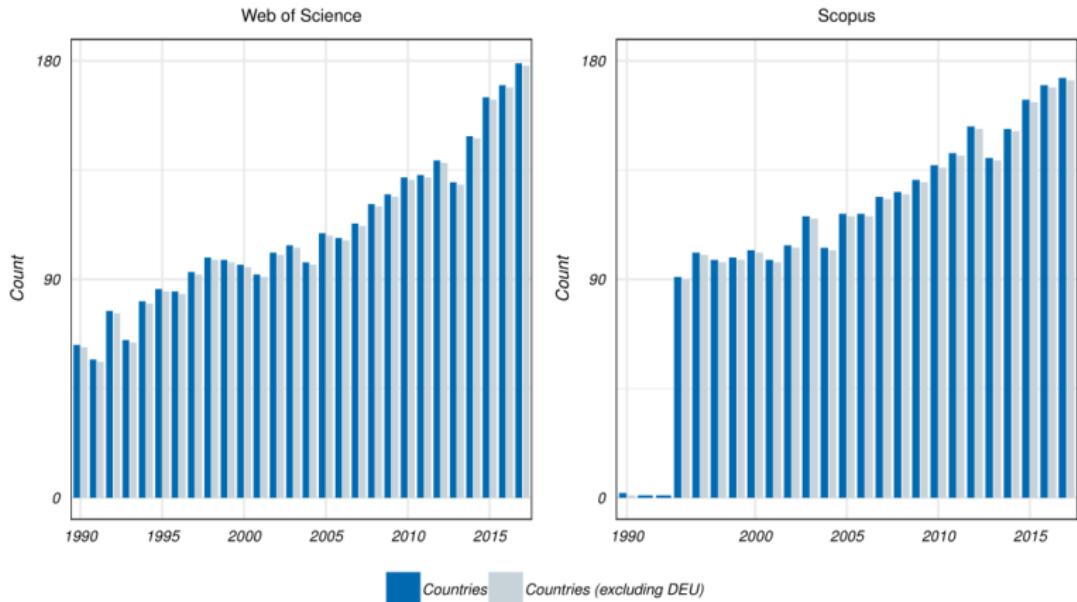
Name	Number of pubs
Charite - Universitätsmedizin Berlin	61,864
Freie Universität Berlin	40,650
Technische Universität Berlin	35,568
Humboldt-Universität zu Berlin	32,997
Helmholtz-Zentrum Berlin für Materialien und Energie	13,288
Max-Delbrück-Centrum für Molekulare Medizin (MDC) Berlin-Buch	8,202
Fritz-Haber-Institut der Max-Planck-Gesellschaft	6,512
Bundesanstalt für Materialforschung und -prüfung	6,448
Robert Koch-Institut	4,439
Max-Planck-Institut für molekulare Genetik (MPIMG)	3,849
Max-Born-Institut für Nichtlineare Optik und Kurzzeitspektroskopie (MBI)	3,680
Bayer HealthCare AG	3,393
Sammelkategorie Kliniken	3,191
Fraunhofer-Institut für Nachrichtentechnik Heinrich-Hertz-Institut	2,781
Deutsches Zentrum für Luft- und Raumfahrt e.V. (DLR)	2,699
Paul-Drude-Institut für Festkörperelektronik (PDI)	2,585
HELIOS Kliniken GmbH	2,498
Leibniz-Institut für Gewässerökologie und Binnenfischerei	2,456
Weierstraß-Institut für Angewandte Analysis und Stochastik Leibniz-Institut im Forschungsverbund Berlin e. V. (WIAS)	2,433
Physikalisch-Technische Bundesanstalt	2,370
Deutsches Herzzentrum Berlin	2,275
Vivantes - Netzwerk für Gesundheit GmbH	2,202
Max-Planck-Institut für Bildungsforschung	2,139
Museum für Naturkunde Leibniz-Institut für Evolutions- und Biodiversitätsforschung (MfN)	2,057
Leibniz-Institut für Molekulare Pharmakologie, Berlin (FMP)	1,775

Table 3: 50 Most prolific organizations located in Berlin from Scopus in 1990 - 2017

Name	Number of pubs
Charité - Universitätsmedizin Berlin	57,615
Freie Universität Berlin	32,448
Technische Universität Berlin	32,438
Humboldt-Universität zu Berlin	31,577
Helmholtz-Zentrum Berlin für Materialien und Energie	12,379
Max-Delbrück-Centrum für Molekulare Medizin (MDC) Berlin-Buch	7,490
Bundesanstalt für Materialforschung und -prüfung	6,237
Fritz-Haber-Institut der Max-Planck-Gesellschaft	6,101
Max-Born-Institut für Nichtlineare Optik und Kurzzeitspektroskopie (MBI)	4,685
Robert Koch-Institut	4,481
Sammelkategorie Kliniken	4,201
Max-Planck-Institut für molekulare Genetik (MPIMG)	3,703
Fraunhofer-Institut für Nachrichtentechnik Heinrich-Hertz-Institut	3,366
Deutsches Zentrum für Luft- und Raumfahrt e.V. (DLR)	3,108
Weierstraß-Institut für Angewandte Analysis und Stochastik Leibniz-Institut im Forschungsverbund Berlin e. V. (WIAS)	2,720
Leibniz-Institut für Gewässerökologie und Binnenfischerei	2,532
Paul-Drude-Institut für Festkörperelektronik (PDI)	2,442
	2,436

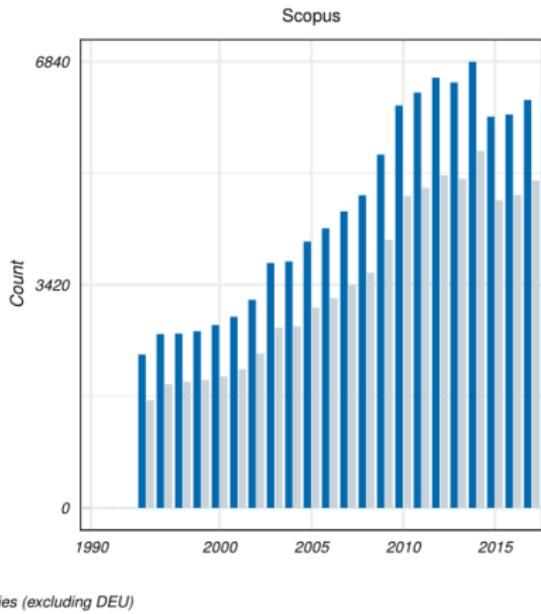
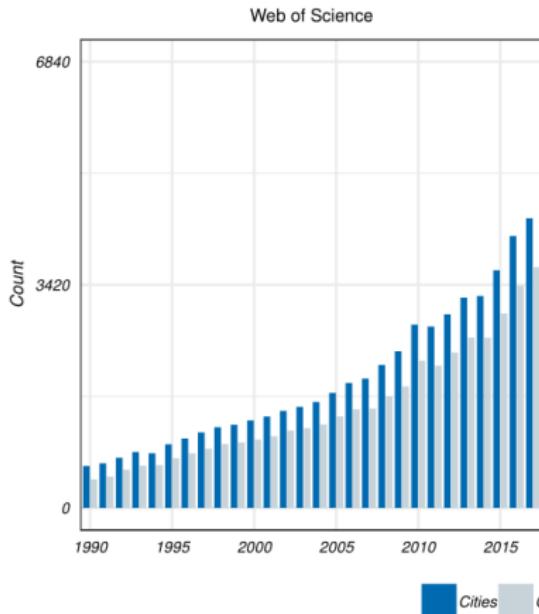
# Unique countries

- Close to 180 unique countries



# Unique Cities

- More than half *international* cities in all years



# Main Cities

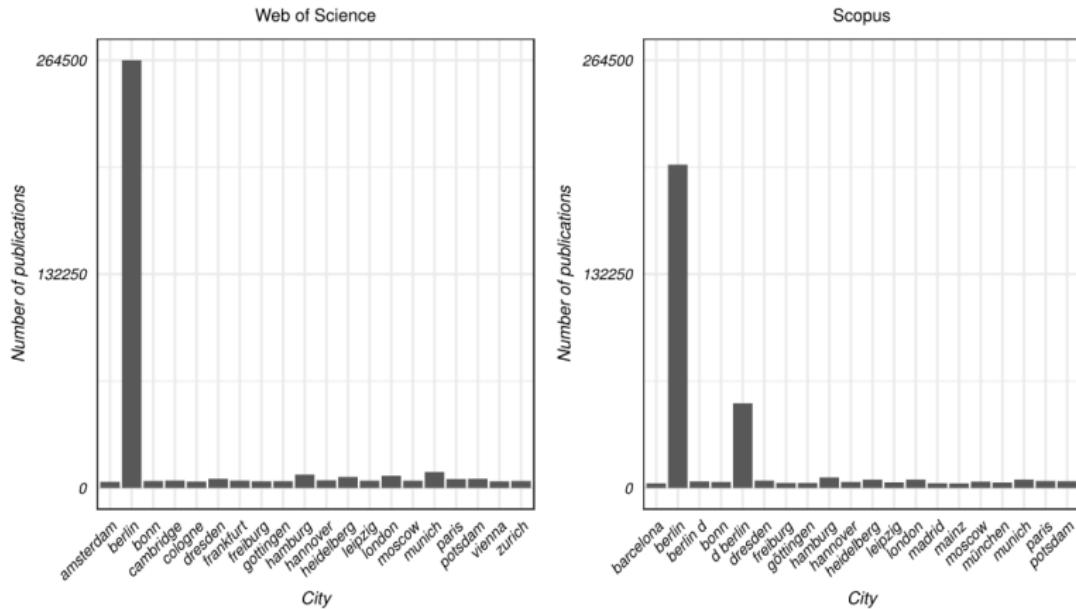


Figure 6: Twenty cities with highest occurrence in affiliation addresses in Articles, Reviews and Conference proceedings published by Berlin region institutes and universities in WOS and Scopus in 1990 - 2017

# Main Cities (excluding Berlin)

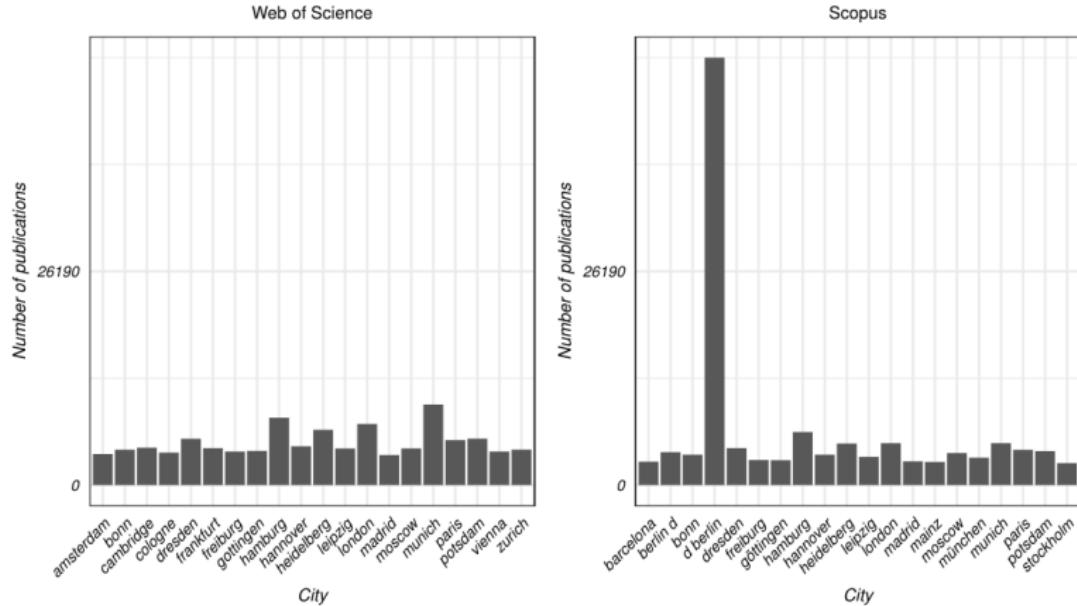


Figure 7: Twenty cities (Excluding Berlin itself) with highest occurrence in affiliation addresses in Articles, Reviews and Conference proceedings published by Berlin region institutes and universities

# Unique organizations

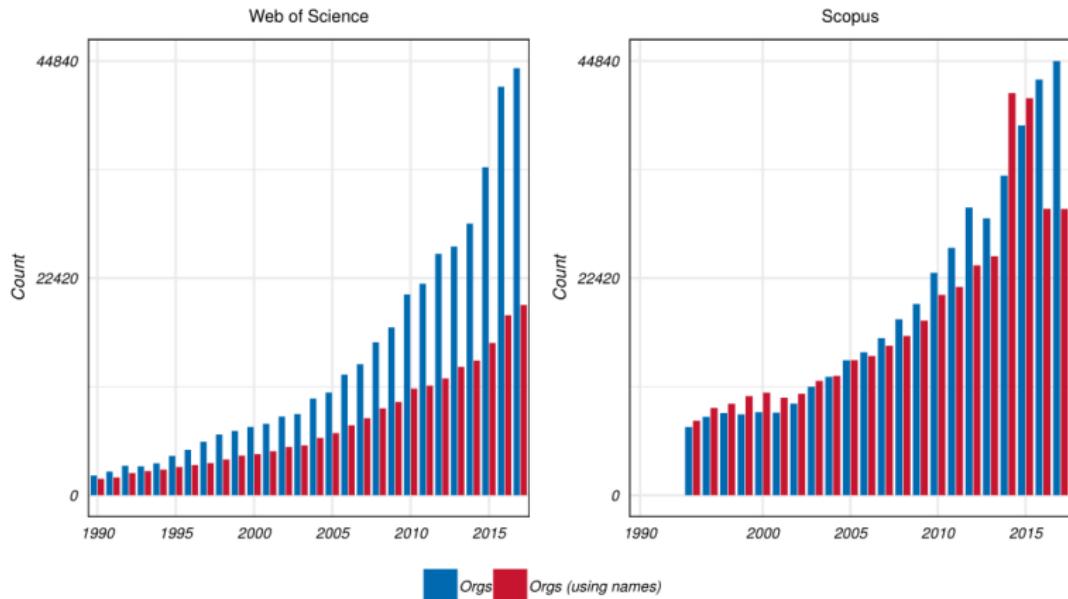


Figure 10: Unique organizations with which Berlin region institutes and universities have collaborated in Articles, Reviews and Conference proceedings in WOS and Scopus in 1990 - 2017

## Preliminary conclusions

- ① Need for lengthy & time consuming disambiguation
- ② It is a must as 1 in 8 unique organization IDs proved reliable
- ③ Network analysis view to collaborations, composition & temporal evolution

## Preliminary conclusions

- ① Need for lengthy & time consuming disambiguation
- ② It is a must as 1 in 8 unique organization IDs proved reliable
- ③ Network analysis view to collaborations, composition & temporal evolution

## Preliminary conclusions

- ① Need for lengthy & time consuming disambiguation
- ② It is a must as 1 in 8 unique organization IDs proved reliable
- ③ Network analysis view to collaborations, composition & temporal evolution

## DEKiF: Bibliometric Study

Aliakbar Akbaritabar (Ali)

Akbaritabar@DZHW.eu

17 October, 2019



German Centre for Higher Education Research  
and Science Studies ■

# A draft report is prepared

- Look at Faust folder, AP8

Aliakbar Akbaritabar

DEKiF Project

Draft Report of Bibliometric Case Study

Report to DEKiF Berlin Team  
June 2019

DZHW  
Deutsche Zentrale für Hochschulbibliotheken und Wissenschaftsinformationen

# Brief on data and methods

- ① A broad/descriptive view to inform next steps
- ② 8 Future lab funded institutions
- ③ 32 other cases, including TU9 who didn't apply for funding
- ④ 3 institution types, *Full*, *Applied* and *Technical*
- ⑤ Historical info, n.o students, staff, research staff (Destatis, Wikidata)
- ⑥ Social aspect, *gender* composition of research staff
- ⑦ WOS/Scopus 1990-2017 (b\_2018 KB)
- ⑧ Article, Review and Conference proceeding as document types
- ⑨ Number of publications, fractional count, 3 years citations
- ⑩ Identifying Future lab projects' [exclusive] publications

# Brief on data and methods

- ① A broad/descriptive view to inform next steps
- ② **8 Future lab funded institutions**
- ③ 32 other cases, including *TU9* who didn't apply for funding
- ④ 3 institution types, *Full*, *Applied* and *Technical*
- ⑤ Historical info, n.o students, staff, research staff (Destatis, Wikidata)
- ⑥ Social aspect, *gender* composition of research staff
- ⑦ WOS/Scopus **1990-2017** (b\_2018 KB)
- ⑧ Article, Review and Conference proceeding as **document types**
- ⑨ Number of publications, fractional count, 3 years citations
- ⑩ Identifying Future lab projects' [exclusive] publications

# Brief on data and methods

- ① A broad/descriptive view to inform next steps
- ② **8** Future lab funded institutions
- ③ **32** other cases, including *TU9* who didn't apply for funding
- ④ 3 institution types, *Full*, *Applied* and *Technical*
- ⑤ Historical info, n.o students, staff, research staff (Destatis, Wikidata)
- ⑥ Social aspect, *gender* composition of research staff
- ⑦ WOS/Scopus **1990-2017** (b\_2018 KB)
- ⑧ Article, Review and Conference proceeding as **document types**
- ⑨ Number of publications, fractional count, 3 years citations
- ⑩ Identifying Future lab projects' [exclusive] publications

# Brief on data and methods

- ① A broad/descriptive view to inform next steps
- ② **8** Future lab funded institutions
- ③ **32** other cases, including *TU9* who didn't apply for funding
- ④ **3** institution types, *Full*, *Applied* and *Technical*
- ⑤ Historical info, n.o students, staff, research staff (Destatis, Wikidata)
- ⑥ Social aspect, *gender* composition of research staff
- ⑦ WOS/Scopus **1990-2017** (b\_2018 KB)
- ⑧ Article, Review and Conference proceeding as **document types**
- ⑨ Number of publications, fractional count, 3 years citations
- ⑩ Identifying Future lab projects' [exclusive] publications

# Brief on data and methods

- ① A broad/descriptive view to inform next steps
- ② **8** Future lab funded institutions
- ③ **32** other cases, including *TU9* who didn't apply for funding
- ④ **3** institution types, *Full*, *Applied* and *Technical*
- ⑤ Historical info, n.o students, staff, research staff (Destatis, Wikidata)
- ⑥ Social aspect, *gender* composition of research staff
- ⑦ WOS/Scopus **1990-2017** (b\_2018 KB)
- ⑧ Article, Review and Conference proceeding as **document types**
- ⑨ Number of publications, fractional count, 3 years citations
- ⑩ Identifying Future lab projects' [exclusive] publications

# Brief on data and methods

- ① A broad/descriptive view to inform next steps
- ② **8** Future lab funded institutions
- ③ **32** other cases, including *TU9* who didn't apply for funding
- ④ **3** institution types, *Full*, *Applied* and *Technical*
- ⑤ Historical info, n.o students, staff, research staff (Destatis, Wikidata)
- ⑥ Social aspect, *gender* composition of research staff
- ⑦ WOS/Scopus **1990-2017** (b\_2018 KB)
- ⑧ Article, Review and Conference proceeding as **document types**
- ⑨ Number of publications, fractional count, 3 years citations
- ⑩ Identifying Future lab projects' [exclusive] publications

# Brief on data and methods

- ① A broad/descriptive view to inform next steps
- ② **8** Future lab funded institutions
- ③ **32** other cases, including *TU9* who didn't apply for funding
- ④ **3** institution types, *Full*, *Applied* and *Technical*
- ⑤ Historical info, n.o students, staff, research staff (Destatis, Wikidata)
- ⑥ Social aspect, *gender* composition of research staff
- ⑦ WOS/Scopus **1990-2017** (b\_2018 KB)
- ⑧ Article, Review and Conference proceeding as **document types**
- ⑨ Number of publications, fractional count, 3 years citations
- ⑩ Identifying Future lab projects' [exclusive] publications

# Brief on data and methods

- ① A broad/descriptive view to inform next steps
- ② **8** Future lab funded institutions
- ③ **32** other cases, including *TU9* who didn't apply for funding
- ④ **3** institution types, *Full*, *Applied* and *Technical*
- ⑤ Historical info, n.o students, staff, research staff (Destatis, Wikidata)
- ⑥ Social aspect, *gender* composition of research staff
- ⑦ WOS/Scopus **1990-2017** (b\_2018 KB)
- ⑧ Article, Review and Conference proceeding as **document types**
- ⑨ Number of publications, fractional count, 3 years citations
- ⑩ Identifying Future lab projects' [exclusive] publications

# Brief on data and methods

- ① A broad/descriptive view to inform next steps
- ② **8** Future lab funded institutions
- ③ **32** other cases, including *TU9* who didn't apply for funding
- ④ **3** institution types, *Full*, *Applied* and *Technical*
- ⑤ Historical info, n.o students, staff, research staff (Destatis, Wikidata)
- ⑥ Social aspect, *gender* composition of research staff
- ⑦ WOS/Scopus **1990-2017** (b\_2018 KB)
- ⑧ Article, Review and Conference proceeding as **document types**
- ⑨ Number of publications, fractional count, 3 years citations
- ⑩ Identifying Future lab projects' [exclusive] publications

# Brief on data and methods

- ① A broad/descriptive view to inform next steps
- ② **8** Future lab funded institutions
- ③ **32** other cases, including *TU9* who didn't apply for funding
- ④ **3** institution types, *Full*, *Applied* and *Technical*
- ⑤ Historical info, n.o students, staff, research staff (Destatis, Wikidata)
- ⑥ Social aspect, *gender* composition of research staff
- ⑦ WOS/Scopus **1990-2017** (b\_2018 KB)
- ⑧ Article, Review and Conference proceeding as **document types**
- ⑨ Number of publications, fractional count, 3 years citations
- ⑩ Identifying Future lab projects' [exclusive] publications

# 40 institutions for the first descriptive view (1/2)

Table 1: 40 institutions chosen for bibliometric profiling and their funding status in Future Lab

Row number	Name	Is funded by Future Lab?	Institute Type
1	HTW Berlin	Did not apply	Applied
2	HS Bochum	Funded	Applied
3	HS München	Funded	Applied
4	HAWK Hildesheim/Holzminden/Göttingen	Not Funded	Applied
5	HS Harz	Not Funded	Applied
6	Deutsche Universität für Verwaltungswissenschaften Speyer	Not Funded	Applied
7	HAW Hamburg	Not Funded	Applied
8	Hochschule der Medien	Not Funded	Applied
9	HS Aschaffenburg	Not Funded	Applied
10	HS Offenburg	Not Funded	Applied
11	HS Osnabrück	Not Funded	Applied
12	HS für Wirtschaft und Umwelt Nürtingen-Geislingen	Not Funded	Applied
13	FH Münster	Not Funded	Applied
14	Humboldt University of Berlin	Did not apply	Full University
15	Free University of Berlin	Did not apply	Full University
16	Charité	Did not apply	Full University
17	Universität Potsdam	Funded	Full University
18	Universität Leipzig	Funded	Full University
19	Universität Marburg	Funded	Full University
20	Universität Kassel	Funded	Full University
21	Universität Erlangen-Nürnberg	Not Funded	Full University
22	Universität Koblenz-Landau	Not Funded	Full University
23	Zeppelin Universität	Not Funded	Full University
24	Universität der Künste	Not Funded	Full University
25	Universität Bremen	Not Funded	Full University
26	Universität zu Kiel	Not Funded	Full University
27	Universität Magdeburg	Not Funded	Full University
28	Medizinische Hochschule Hannover	Not Funded	Full University
29	TU Berlin	Did not apply	Technical
30	KIT	Did not apply	Technical
31	TU Dortmund	Did not apply	Technical
32	RWTH Aachen	Did not apply	Technical
33	TU Braunschweig	Did not apply	Technical
34	TU Darmstadt	Did not apply	Technical
35	Leibniz Universität Hannover	Did not apply	Technical
36	Universität Stuttgart	Did not apply	Technical
37	TU Dresden	Funded	Technical
38	TU München	Funded	Technical
39	TH Brandenburg	Not Funded	Technical
40	TH Nürnberg	Not Funded	Technical

# 40 institutions for the first descriptive view (2/2)

- Blue = Full university, Red = Applied university, Green = Technical university, Euro sign = Funded, Question mark = Did not apply, Cross mark = Not funded



# Example of a funded project: Leipzig University

- Blue = Full university, **Euro sign** = Funded
- *Name variants; Future Lab*: Universität Leipzig, **DE**: Universität Leipzig, **EN**: Leipzig University
- Wikidata ID: Q154804, **GRID ID**: grid.9647.c,
- KB cleaned name: Universität Leipzig, **KB code**: 111
- Project tag name: *Deutsches Zentrum für integrative Biodiversitätsforschung (iDiv)*

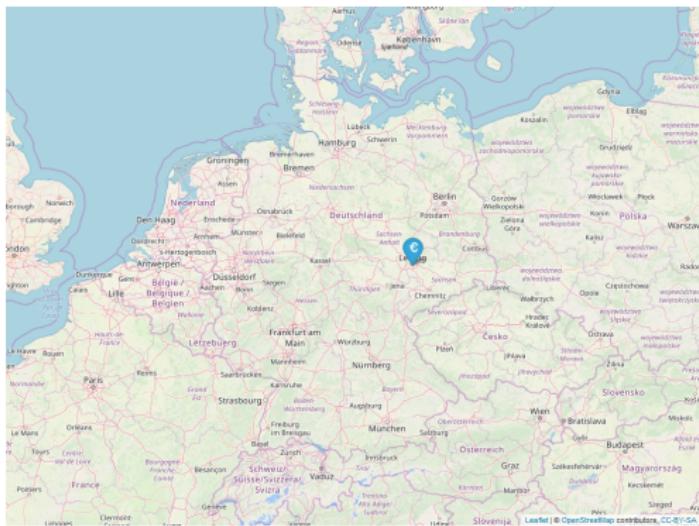
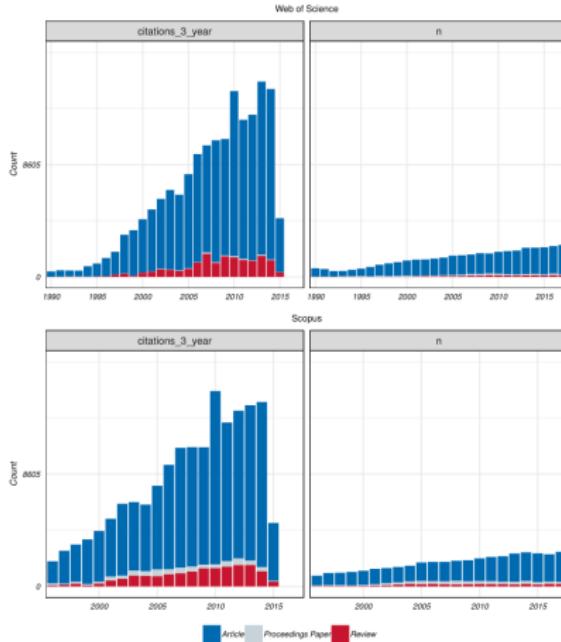


Table 19: Description of Universität Leipzig

Metric	Value
Type of institute	Full University
Funded by Future Lab?	Funded
Establishment year (from Wikidata)	1409
Number of students (from Wikidata)	30226
First publication year (WOS)	1990
First publication year (Scopus)	1996
Number of employees (from Destatis)	10614
Research employees	5092
Research employees (male)	2701
Research employees (female)	2391
Administration employees	5522
% Research employees	48
% Research employees (female)	47

# Leipzig University publications



# Leipzig University publications (fractional)

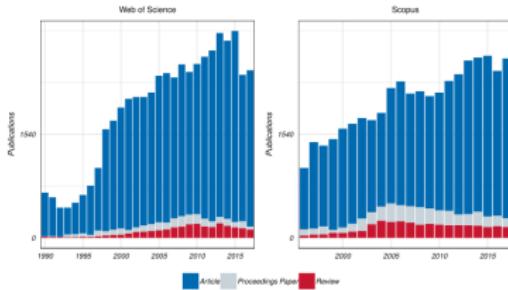
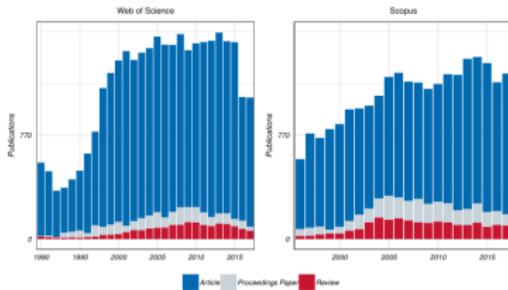


Figure 108: Articles, Reviews and Conference proceedings published by Universität Leipzig in WOS and Scopus in 1990 - 2017 (Fractional count based on countries)



# Leipzig University countries, cities, orgs

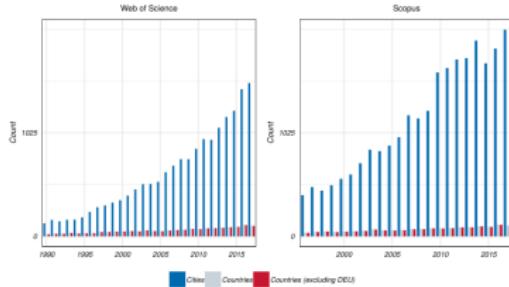
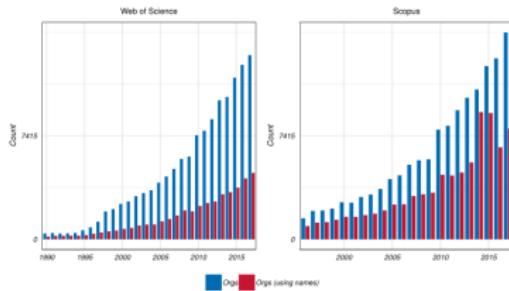


Figure 110: Unique countries and cities with which Universität Leipzig has collaborated in publications



# Publications explicitly mentioning funded project

- Deutsches Zentrum für integrative Biodiversitätsforschung (*iDiv*)
- Different combination of name parts searched in `address_full`

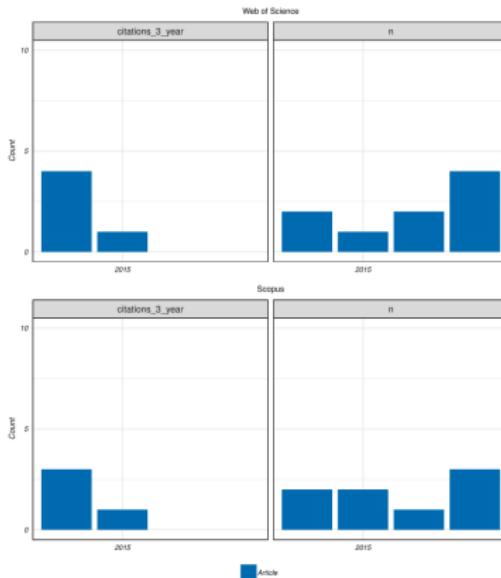


Figure 112: Articles, Reviews and Conference proceedings published and citations accrued in the first three years by sub units of Universität Leipzig which have applied for Future Lab funding indexed in WOS (top) and Scopus (bottom) in 1990 - 2017 (Keywords used to find the sub unit: *iDiv* )

## Next step, further probe into this project

- Deutsches Zentrum für integrative Biodiversitätsforschung; iDiv; sDiv; yDiv
- **Website:** <https://www.idiv.de/de/forschung/publikationen.html>
- Consortium of **3** universities and **8** non-university research institutions
- Universität Leipzig; Friedrich-Schiller-Universität Jena; Martin-Luther-Universität Halle-Wittenberg; Helmholtz-Zentrum für Umweltforschung GmbH – UFZ; Max Planck-Institut (MPI) EVA; MPI BGC; MPI CE; Leibniz DSMZ; Leibniz SMNG; Leibniz IPB; Leibniz IPK;
- **Disambiguation** of collaborating organizations in the project
- Presenting the bibliometrics results to project members for **feedback**
- Expand the scope of profiling to other type of scientific output **not** reflected in databases

## Preliminary conclusions

- ① A separate case study for each Future Lab funded project
- ② Highest level of organizational structure as context/baseline
- ③ Network analysis view to collaborations, composition & temporal evolution
- ④ More *sociological* view to each project's profile

## Preliminary conclusions

- ① A separate case study for each Future Lab funded project
- ② Highest level of organizational structure as context/baseline
- ③ Network analysis view to collaborations, composition & temporal evolution
- ④ More *sociological* view to each project's profile

## Preliminary conclusions

- ① A separate case study for each Future Lab funded project
- ② Highest level of organizational structure as context/baseline
- ③ Network analysis view to collaborations, composition & temporal evolution
- ④ More *sociological* view to each project's profile

## Preliminary conclusions

- ① A separate case study for each Future Lab funded project
- ② Highest level of organizational structure as context/baseline
- ③ Network analysis view to collaborations, composition & temporal evolution
- ④ More *sociological* view to each project's profile

Thanks for your attention!