CWTS Meaningful metrics

BIBLIOMETRIC REPORT

Bibliometric evaluation of the SystemsX.ch initiative: performance, benchmark and collaboration analysis

March 2017

Universiteit Leiden

date
CWTS B.V.
Centre for Science and
Technology Studies,
Leiden University

Bibliometric evaluation of the SystemsX.ch initiative: performance, benchmark and collaboration analysis

Report for SystemsX.ch, the Swiss Initiative in Systems Biology

Project team

Ed Noyons, *Project leader*

CWTS B.V.
P.O. Box 905
2300 AX Leiden, The Netherlands
Tel. +31 71 527 3909
Fax +31 71 527 3911
E-mail info@cwts.nl

Table of contents

Executive summary

An extensive bibliometrics study was executed to monitor the potential effect of SystemsX.ch funding. SystemsX.ch is a funding initiative in the area of systems biology. The report focuses on the entire initiative as well as on 5 specific projects within. In the report we provide statistics on production (publication output), impact as measured by citations, interdisciplinarity and collaboration as measured by co-authorships. In such analysis we use the world as benchmark. In this case we also defined ad hoc benchmarks to position in more detail the impact.

In total we analyzed more than 1200 publications acknowledging StystemsX.ch from 2008 onwards. These publications have a high citation impact normalized by field and year, not only as compared to the world but also much higher than the defined benchmarks (UK and Germany). Moreover, SystemsX.ch researchers managed to get their work published in high impact journals. This means that at least that SystemsX.ch was able to select high impact researchers. We cannot claim that this impact is *due* to the funding. Regarding collaboration, however, we do see a positive effect of the funding. The funded researchers not only publish more as from 2008 onwards, but also their mutual collaboration increased, as measured by co-authorships. Moreover, we found that the increased collaboration does not only involve the funded publications. They increased their collaboration measured in all their output. This indicates an impact of SystemsX.ch beyond the directly funded research. Finally, we did not find any evidence of an increased interdisciplinary character of research funded by SystemsX. It seems that this remains at a similar level throughout 2000-2015.

1. Introduction

SystemsX.ch is a public research initiative in Switzerland focusing on a broad topical area of basic research[1]. The initiative advances systems biology in Switzerland with the aim of positioning Switzerland among the world leaders in this area of research.

The work of the initiative as a whole, as well as that of the individual research projects, is monitored by the Swiss National Science Foundation (SNSF). This report entails a robust bibliometric analysis to measure its impact.

The two key statements which need to be answered in the impact analysis are:

1. SystemsX.ch has funded interdisciplinary research projects that had a high impact output
    - Interdisciplinarity of research consortia should be shown
    - Impact of research/publications to be shown rather than pure absolute numbers of publications
2. SystemsX.ch systematically initiated and funded interdisciplinary, inter-institutional research collaborations between the partner institutions
    - New collaborations between partner institutions/research groups
    - Interdisciplinarity of research consortia

This report provides the results of a sophisticated bibliometric analysis that supports the two key statements.

This study uses standard bibliometric performance analyses to measure output and impact of research funded by the SystemsX.ch initiative. As the positioning of this research as compared to similar initiatives is difficult, if not impossible, we developed a way to benchmark the performance.

[1] Information on the initiative is available at systemsx.ch.

2. Data and Method

Data

The evaluation study regards the output of SystemsX.ch as a whole and of 5 projects within: InfectX/ TargetInfectX, LipidX 1&2, NeuroChoice, PhosphoNetX/ PhosphoNetPPM and PlantGrowth 1&2.

In the bibliometric evaluation we analyze publications within the context of Web of Science (WoS, the core collection). CWTS has developed an enhanced version of WoS for bibliometric purposes.

Publications for SystemsX.ch and the 5 projects were collected in 2 steps:

1. A list of publications was compiled by SystemsX.ch for the 5 projects and sent to CWTS. These publications were already linked to Web of Science data through a publication identifier.
2. CWTS collected publications by using the acknowledgement information in WoS. If possible publications were assigned to one of the five projects. If only a reference to SystemsX.ch was mentioned, the publication was added to the overall set.

The process yielded the following numbers of publications, as listed in the table below (Table 1).

Table 1 Numbers of publications collected for SystemsX.ch and 5 projects

Actor	# WoS pubs
SystemsX.ch	*1254*
InfectX/TargetiInfectX	54
LipidX 1&2	175
NeuroChoice	82
PhosphoNetX/PhosphoNetPPM	58
PlantGrowth 1&2	100

The distribution of number for the entire SystemsX.ch is in Table 2. The overview shows that in 2008, which was the starting year of the initiative, the number is rather low. This is caused by the early stage of the initiative but also by the fact that in WoS only since 2009 acknowledgements were processed. For that year we were not able to collect data besides the ones provided by SystemsX.ch. Therefor it should be noted that the analysis may not be covering the entire output of SystemsX.ch. The number in 2016 is not complete as we had to limit the results to the ones processed until the third quarter WoS update. For the performance analysis as well as for the benchmark analysis we do not consider the 2016 publications at all. For citation analysis we require a full year of citations in order to be included. For the benchmark analysis we used the publication classification (See Annex D) which covers 2000-2015.

For the performance (citation) analysis we consider articles, letters and reviews only, c.f., citable publications. For SystemsX.ch we analyzed 1143 publications (2008-2015).

Table 2 Distribution of publications of years identified for SystemsX.ch

Pub Year	# WoS pubs
2008	30
2009	82
2010	125
2011	176
2012	210
2013	184
2014	159
2015	190
2016	98

Methods

In this section we discuss the methods used for each type of analysis: performance measurement, benchmarking and collaboration networks

Performance

The first analysis involves a standard performance analysis of (citable) publications. In this analysis we measure output and impact using the CWTS standard indicators (See Annex A). We thus characterize output in terms of volume as well as collaboration type (proportion involving (inter)national collaboration and collaboration involving Industry). Moreover we characterize the output by means of the impact of the journal (MNJS). Furthermore, we measure impact of output by means of absolute numbers of citations as well as by citations normalized by field. More details about the method in Annex A. In addition we performed for all actors a trend analysis of their impact (MNCS) accompanied by stability intervals. The stability intervals are calculated using a bootstrap method ([ref]) and provide some idea of the stability we found for the measures in the publication sets at stake. In general, if the number of publications is low the stability tends be low and the other way around. The results of this trend stability analysis is Annex C.

Research profile

In order to characterize the performance in more detail, we create a research profile of SystemsX.ch. Publications are distributed over the 250 WoS subject categories, through the journals in which they are published. For each subject category we calculated the average normalized impact (MNCS, PP[top10]) and added that information to the overview of largest subject categories. If a journal is assigned to more than one category, a SystemsX.ch paper (and impact) is fractionalized over the categories accordingly.

Interdisciplinarity

Measuring interdisciplinarity is a challenge because there are probably as many definitions as there are disciplines in science. Not only is it difficult to properly define and measure interdisciplinarity, there are no world standards to define above average interdiscplinarity or the like. Therefore, we confine ourselves to measure the development of the interdisciplinarity of SystemsX.ch funded research output.

The definition of interdisciplinarity is based on the assumption that references at the end of a publication represents the content and knowledge base. We define interdisciplinarity as the measure in which research output refers to (cites) other fields of science. Moreover, we will take the cognitive distance between the publication and the cited fields into account. Citing a field that is cognitively distant shows more interdisciplinarity than citing a field nearby. We operationalized fields and their cognitive distance at two levels of the CWTS classification system (c.f., Annex D; top level of 27 clusters and intermediate level of around 800 clusters) and measured the interdisciplinarity of the individual years from 2008 to 2015.

Benchmarks

The data collection for the benchmark analysis is more complicated and arbitrary choice had to be made. However during the testing of thresholds to apply, we monitored the results on sensitivity.

The method to benchmark the SystemsX.ch results comprised the following steps. The output of SystemsX.ch was distributed over the more than 4000 publication clusters of the CWTS publication classification. The output of SystemsX.ch appeared in 274 clusters. The best populated clusters were used to define the field to which (the output of) SystemsX.ch belongs. In the field defined as such, we collected the output of the UK, Germany and Switzerland. The output with at least one affiliation from the UK and from Germany was used as benchmarks.

The selection of clusters to be used to define the 'SystemsX.ch field is challenging. The distribution of SystemsX.ch publications over clusters is such that using all 274 clusters would be far too broad, whereas using only the top cluster (including 118 SystemsX.ch publications) would be far too narrow and representing only 10% of the SystemsX.ch output. In order to study the effect of adding clusters to the field definition we applied the following indicators:

- Proportion of SystemsX.ch output represented
- Ratio SystemsX.ch output and output in Switzerland (in field definition)
- Ratio SystemsX.ch output and Germany output (in field)
- Ratio SystemsX.ch output and UK output (in field)

An overview of the indicators is depicted in the diagram below (Figure 1).

The chart shows the development of ratios if the threshold of SystemsX.ch publications populating a cluster from the classification is increased or decreased. For example, if the threshold is set to 10, over 50% of the total SystemsX.ch output is represented (Purple line), and the output of SystemsX.ch is 13% of the total Swiss output (Blue line) and 4% of the UK (Green)) or German (Red) output. If the threshold is set to 5, the representation of SystemsX.ch increases to almost 70%, at the expense of representation of 8% Swiss output, and only 3% of the German or UK output.

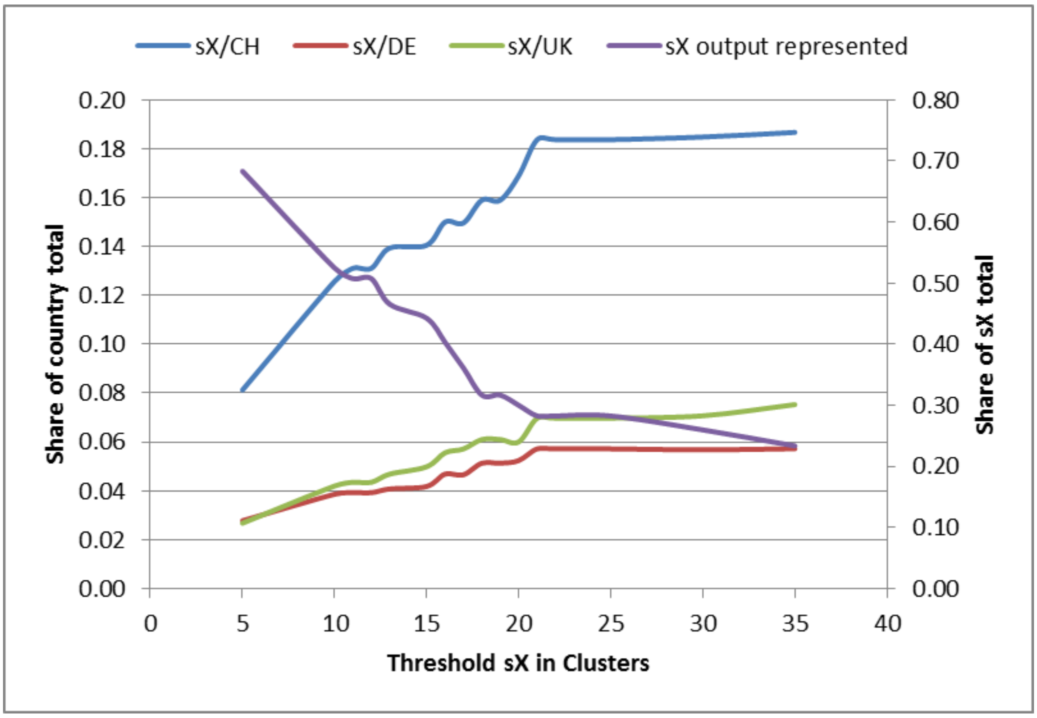


Figure 1 Effects of thresholds to define 'SystemsX.ch field'

The three indicators were used in the following way to decide upon a threshold. First of all a substantial part of SystemsX.ch output should be represented. We chose a threshold that would represent at least 50%. Subsequently we look at the ratio SystemsX.ch output and the output of Switzerland, Germany and the UK. If we increase the threshold, the 'home advantage' of SystemsX.ch will be bigger. In order to make a fair comparison the field should not be too biased. Compared to the overall output of Switzerland, the share is over 70% if the threshold is higher than 20. We noticed that if we the threshold is higher than 20, the ration SystemsX.ch as compared to UK, Germany and Switzerland countries increases substantially. We consider this a sign of data becoming more biased towards SystemsX.ch. All these findings considering, we chose a threshold of 12. This means that 24 clusters will define the SystemsX.ch field covering almost 135,000 publications in entire Switzerland in 2008-2015. More details about the selected clusters can be found in Annex D.

Within the definition of the field, we collected publications from Germany and the UK to serve as benchmarks.

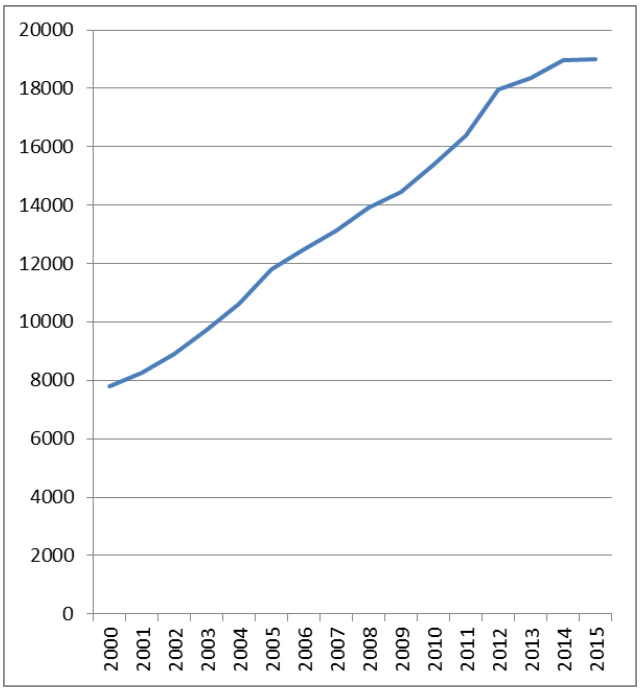
Collaboration networks

Co-author Network analyses provide insight into the collaboration patterns of actors at stake. The challenge with such analysis is the interpretation of results. There are no actual benchmarks or references to assess the network. However, we do have a list of actors (organizations, institutions) involved in SystemsX.ch and hence are able to compare the results since the initiative started with the period before. The problem is that the actors involved are in our data only available at main organization level, in other words: the universities and research organization at the top level, not at the faculty or institutional level. For that reason, a 'SystemsX.ch field' delineation is necessary to monitor co-authoring before 2008, taking only relevant publications into account. The definition of this field is the same as for identifying benchmarks. The number of publications considered in 2000-2015 is around 217,000.

In addition to the analysis on the level of organizations we conducted an analysis on a much lower level. For a selection of principle investigators (PI's) with SystemsX.ch funding (in the mentioned 5 projects) we investigated collaboration in terms of co-authorships. For these PI's we collected their full oeuvre 2000-2015 and analyzed their network over time.

Table 3 Output per year for the 'SystemsX.ch field', 2000-2015

Pub year	# pubs
2000	7,801
2001	8,279
2002	8,898
2003	9,737
2004	10,633
2005	11,813
2006	12,472
2007	13,127
2008	13,919
2009	14,452
2010	15,404
2011	16,405
2012	17,977
2013	18,364
2014	18,964
2015	18,995



We detected a steady growth throughout the entire period. In absolute numbers the collaboration is expected to increase at a similar pace. We will use the total numbers in 2000-2007 and in 2008-2015 to normalize the statistics on collaboration. The networks were visualized with VOSviewer (www.vosviewer.com).

Results

In this section we will describe the results for each of the three analyses and relate them to each other if possible or appropriate.

Performance

The research performance of SystemsX.ch and its five projects are in the table below.

Table 4 Performance statistics of SystemsX.ch and 5 projects

Unit	P	MNJS	TCS	MCS	MNCS	PP (top10)	PP UIC	PP collab	PP Intl collab
SystemsX.ch	1,143	2.14	24,432	21.37	2.35	0.32	0.04	0.75	0.55
InfectX	54	2.39	2,342	43.37	4.53	0.37	0.09	0.76	0.46
LipidX	165	2.28	3,865	23.42	2.38	0.34	0.01	0.65	0.53
NeuroChoice	81	2.24	2,420	29.88	2.74	0.43	0.00	0.80	0.78
PhosphoNetX	52	4.29	2,959	56.63	5.07	0.63	0.08	0.87	0.66
PlantGrowth	93	1.70	1,736	18.67	1.89	0.28	0.02	0.82	0.61

The output analysis shows that in all projects and in SystemsX.ch in general researchers get their results published in high impact journals. The MNJS ranges from high (70% above world average) to very high (more than four times world average). Particularly the MNJS of SystemsX.ch overall is worth mentioning because even with more than 1100 publications the MNJS is more than two times world average. The impact of journals in which authors with SystemsX.ch funding managed to get their research published (MNJS) shows al slight decrease until 2011 (stabilizing at 2, Figure 2) which is often observed when the amount of output increases.

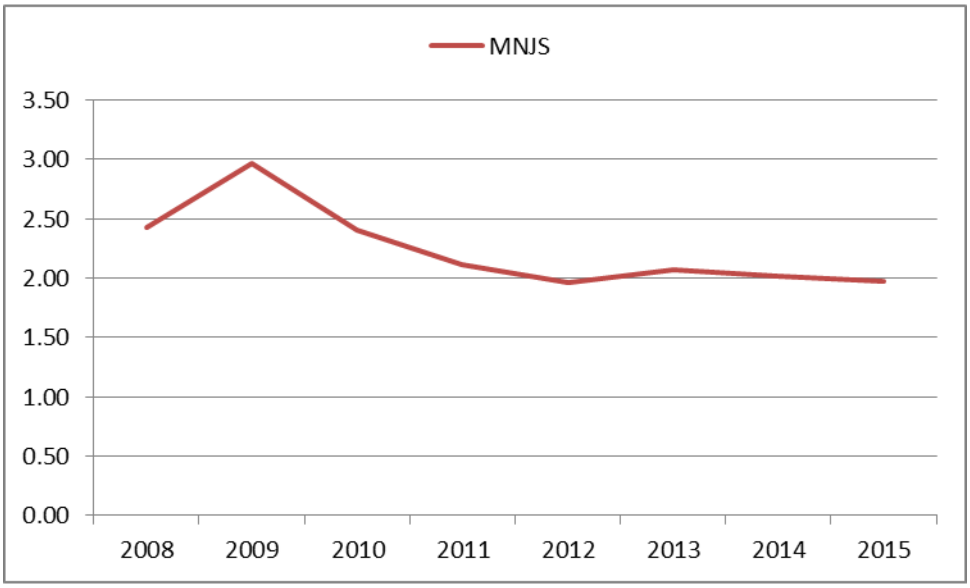


Figure 2 Development of impact of journals published in (MNJS, 2008-2015)

Furthermore, 65%-87% of the publications involve collaboration, while 45% to even 78% involves international collaboration.

The impact of the studied units is high. In all 6 case the impact as measured by MNCS is at least 90% above world average. And in two cases even four and five times world average. In the case of the entire SystemsX.ch it is more than two times world average. The proportion of top 10% most highly cited publications indicator (PPtop10), which is not sensitive for outliers, is an alternative measure for impact and shows basically the same picture. SystemsX.ch has 32% of its publications in the top 10% which is even more than three time the expected 10%.

Finally we listed the ten academic organizations citing most often SystemsX.ch funded publications over the entire period. The result is in the table below.

Table 5 Top ten organizations citing SystemsX.ch funded publications

Organization	# Cits
Harvard University	879
Max Planck Society	834
Centre National de la Recherche Scientifique	603
National Institutes of Health	582
Chinese Academy of Sciences	582
University of Toronto	431
University College London	411
University of Cambridge	409
Massachusetts Institute of Technology	397
ETH Zurich	377

Research profile

To characterize the output and impact in more detail, we calculated them by WoS subject category. The results are in the table below.

Table 6 Research profile SystemsX.ch (output & impact 2008-2015)

Category	P	MNCS	PP[top10]
MULTIDISCIPL SC	162	2.98	0.40
BIOCHEM&MOL BIOL	156	3.25	0.39
CELL BIOLOGY	134	2.67	0.43
BIOCHEM RES METH	108	2.48	0.31
NEUROSCIENCES	58	2.89	0.45
PLANT SCIENCES	49	1.75	0.18
GENETICS&HEREDIT	46	2.17	0.30
BIOTECH&APPL MIC	41	1.79	0.25
MATH&COMPUT BIOL	38	1.10	0.12
MICROBIOLOGY	37	3.02	0.39
DEVELOPMENT BIOL	32	1.62	0.29

The top 11 largest categories show the most relevant areas for SystemsX.ch funded research with their impact (all 11 have more than 30 publications in 2008-2015). **The largest category is multidisciplinary sciences, which means that most papers are published in journals such as PLOS, PNAS and Nature.** These are obviously also the journals in which high impact is achieved. Actually in all 4 largest categories the MNCS is very high (2-3 times world average). In addition, we observe Neurosciences and microbiology as high impact categories but in these categories the output is much lower than in the top 4.

Interdisciplinarity

The measure of interdisciplinarity of SystemsX.ch funded output is measured from year to year. We applied two structures of science as a kind of sensitivity analysis. We observe a rather stable measure over the years in both ways of measurement. We may see a slight increase from 2009 onwards, particularly if we use the high level structure (Blue line). However, the differences between years are very small.

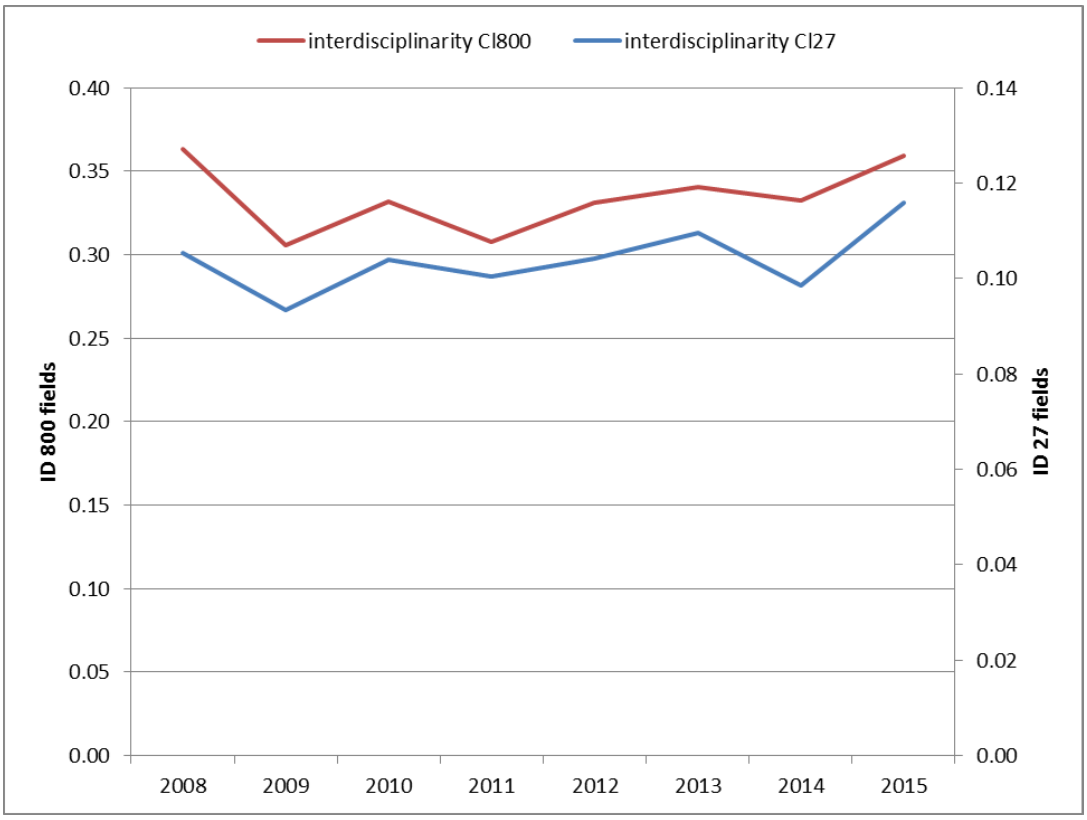


Figure 3 Development interdisciplinarity of SystemsX.ch funded output (2008-2015)

Benchmarks

Within the research field defined as 'SystemsX.ch field', we collected data from UK, Germany as well as for Switzerland and SystemsX.ch. This means that in this analysis we do not consider the entire SystemsX.ch output but only a representative subset. The results of this analysis is in the table below.

Table 7 Performance (output & impact) for SystemsX.ch and benchmarks

Unit	P	MNJS	TCS	MCS	MNCS	PP (top10)	PP collab	PP Intl collab
SystemsX.ch (BM selection)	580	2.13	14,293	24.63	2.48	0.31	0.77	0.59
Germany	14,781	1.15	183,562	12.42	1.20	0.13	0.74	0.54
United Kingdom	13,338	1.27	193,706	14.52	1.38	0.15	0.72	0.61
Switzerland	4,424	1.52	76,519	17.29	1.76	0.20	0.80	0.68

First of all, we see that the subset of SystemsX.ch output, covering around 50%, is representative for the entire set. The MNJS, MNCS as well as PPtop10 are almost equal as measured by the entire SystemsX.ch set. This is also the case for the proportion of publications involving national and international collaboration.

Then, if we compare these results to the selected benchmarks Germany and the UK, we see large differences of output of course. This should be noted interpreting the results. Higher impact scores are 'more easily' reached with smaller number of publications. Still it is clear that the impact of SystemsX.ch (MNCS and PPtop10) is much higher, even twice the measure. Meanwhile the proportion of publications involving collaboration and international collaboration are a similar level. **Furthermore, the impact of journals (MNJS) in which SystemsX.ch publishes is substantially higher.**

To estimate the impact scores considering the large differences in output, we took Switzerland as a whole into consideration. The output is then still not as large as the UK and Germany but eight times the output of SystemsX.ch. Still the impact is considerably higher. We assume that the impact of SystemsX.ch is for a great deal responsible.

Collaboration networks

Main organizations

As indicated in the method section, the numbers of publications increased at a steady pace from around 8,000 in 2000 to almost 19,000 in 2015. For the network analysis we selected a set of author affiliations (institutes/ organizations) to be included. These institutes should at least have 8 publications in the entire period within the fields as defined by SystemsX.ch research.

The proportion of publications involving collaboration by these institutes is stable over the 15 years. The set we analyzed for collaboration patterns covers around 25% of the number of publications in each year. The number of institutes involved in the network analysis is stable at around 50 over the years. In that sense we cannot say that the network has become bigger over the years. **However, if we look at the number of co-authorships and the number of connections, we can conclude that the network has become more dense.**

We compared the period 2000-2007 with 2008-2015. In the first period the set consists of just over 20,000 publications. In the second period, the set consists of almost 35,000 publications. The amount of actors is 50 in period 1 and 52 in period 2. The main statistics are in the table below.

Table 8 Network statistics field SystemsX.ch for selected affiliations (2000-2007 and 2008-2015)

	2000-2007	2008-2015
# connections	703	1,008
# nodes (affiliations)	50	52
# possible connections	1,225	1,326
Density	0.57	0.76
# pubs	20,256	34,150
sum weighted connections	3,792	13,558
weighted per connection	5.39	13.45
divided by #pubs	0.266	0.394

The number of connections increases from around 700 to 1000. This means that there are more institutes collaborating since SystemsX.ch started. If we divide this number by the number of possible connections (between 50 in the first and 52 in the second period) we see that in the first period 57% of the possible connections has actually happened, while 76% in the most recent period (Density).

Still, as the number of publications increased as well, we assumed we should look at the weight of the connections (i.e., the number of co-authorships between the pairs). The sum of these weights increased from less than 4,000 to over 13,000. If we normalize this sums by the number of connections in each period, we see an increase from 5.39 to 13.45. And even if we divide that score by the number of publications, we see an increase from 2000-2007 to 2008-2015. In other words: there is a more dense collaboration between SystemsX.ch funded partners since the start of the initiative.

Finally, we visualized the networks. The co-authorships between the affiliations involved in the two periods were depicted using VOSviewer. In the figures below we present screenshots of these networks.

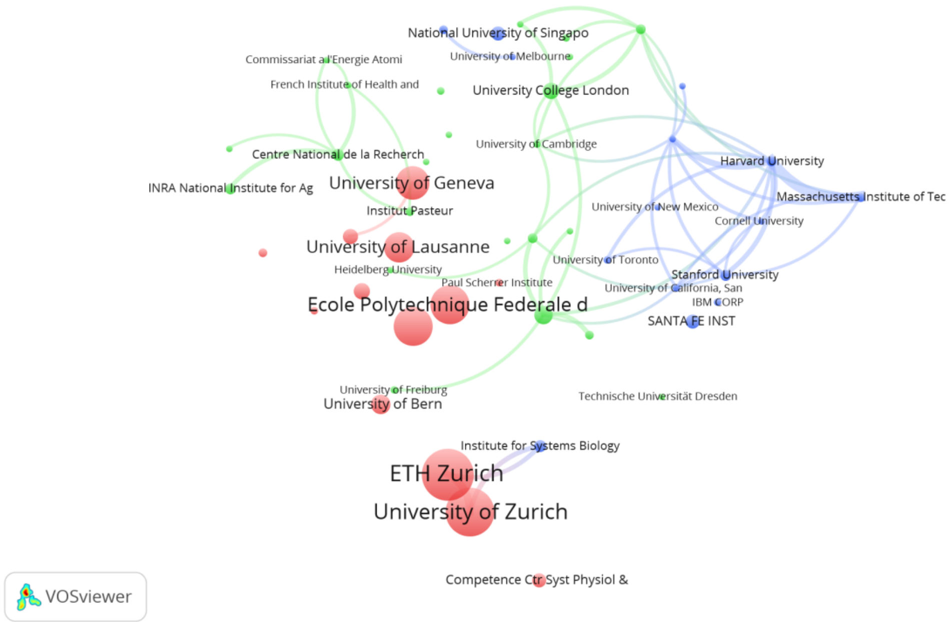


Figure 4 Collaboration network in SystemsX.ch field (2000-2007)

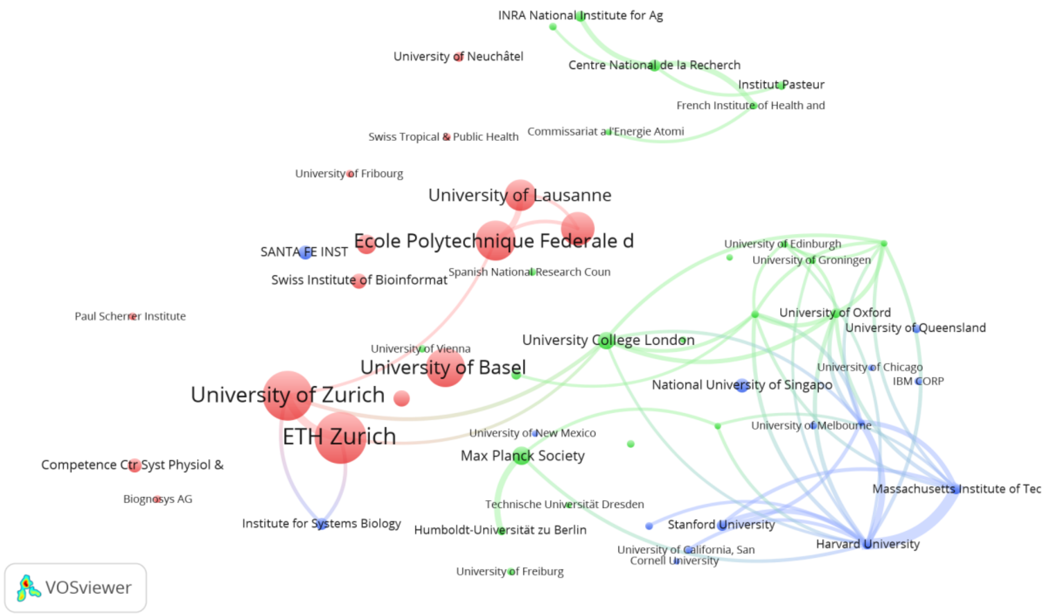


Figure 5 Collaboration network in SystemsX.ch field (2008-2015)

In these networks we depict all 50 and 52 affiliations as well as the 50 most prominent linkages. The size of a circle indicates the number of publications by an affiliation. The color represents the geographical location (Red: Switzerland, Green: Europe and Blue: rest of the world).

We detect in these graphs the usual tendency to collaborate with partners nearby. The Swiss institutions prefer to collaborate with other Swiss institutions, then with Europeans and ultimately outside Europe. This we observe in both periods. Furthermore, we visually don't detect any clear difference between the two networks at all. It is the numbers in Table 8 that indicate a more dense network of collaboration that may be stimulated by the SystemsX.ch initiative.

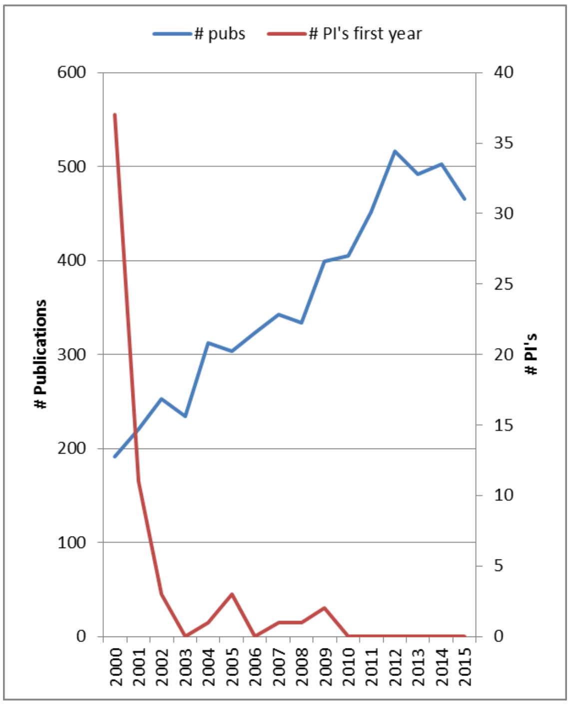
Collaboration network at the level of individual PI's

In more detail we investigated the **collaboration patterns between selected principle investigators** in the 5 programs. In order to get more robust data and results, we aggregated the data for the 5 programs but kept the information of the program the PI's belong to. In some cases PI's were connected to more than one program. In total there were 70 combinations of 64 PI's with 5 programs. For 59 of them we were able to collect publications in the period studied.

In the results below, we integrated the oeuvres of all 59 PI's and created a network of collaboration in the period 2000-2007 and 2008-2015. Four of the programs started in 2008, while of started in 2010. Taking into account the overall objective and the complexity of splitting oeuvres of individuals, we chose for this split of periods. By comparing results in the two periods we will get a useful insight into the effect of the funding. Overall we collected almost 6000 publications of 59 PI's in the 16 years of our analysis. The amount of publications increases over the years with 7% on average. In the third column we included the number of PI's that had their first publication (included in the analysis). This should shed light on the influence of new PI's to the total amount of publications per year. As 51 out of the 59 are represented in the data, we consider this effect almost zero. In other words, a great majority of PI's are represented over the entire period of our analyses. This means that we can 'safely' make a comparison of the periods 2000-2007 and 2008-2015.

Table 9 Number of publications of the 64 PI's (2000-2015)

pub_year	# pubs	# PI's first year
2000	191	37
2001	221	11
2002	253	3
2003	234	0
2004	312	1
2005	304	3
2006	323	0
2007	343	1
2008	334	1
2009	399	2
2010	405	0
2011	452	0
2012	516	0
2013	492	0
2014	503	0
2015	466	0



In the analysis of the 59 PI's we created a collaboration network of the entire period and of the two period separately. Each PI is characterized with a color to indicate the program they belong to. Those active in more than one program, we colored dark Grey.

The network of the entire period was used to position each PI in a two dimensional space. Subsequently we drew lines between the PI's indicating their mutual collaboration, in terms of co-authorships.

There are a few important observations to be made, comparing the two networks. First of all the overall structure matches the assignments of PI's to programs. This means that the collaboration between PI's is (as expected) according to the grouping by programs. **Secondly, we see a clear acceleration of connections since 2008. In the period until 2007 there are 17 connections while in the period since 2008 the amount of connections is 144.**

Finally we observed the following. The map of 2008-2015 (Figure 7) includes all collaborations between these PI's, regardless if these collaborations were funded by SystemsX.ch. If we would confine the map to only SystemsX.ch publications, the map looks mainly the same. Out of the 144 connections, 116 would remain (Figure 8).

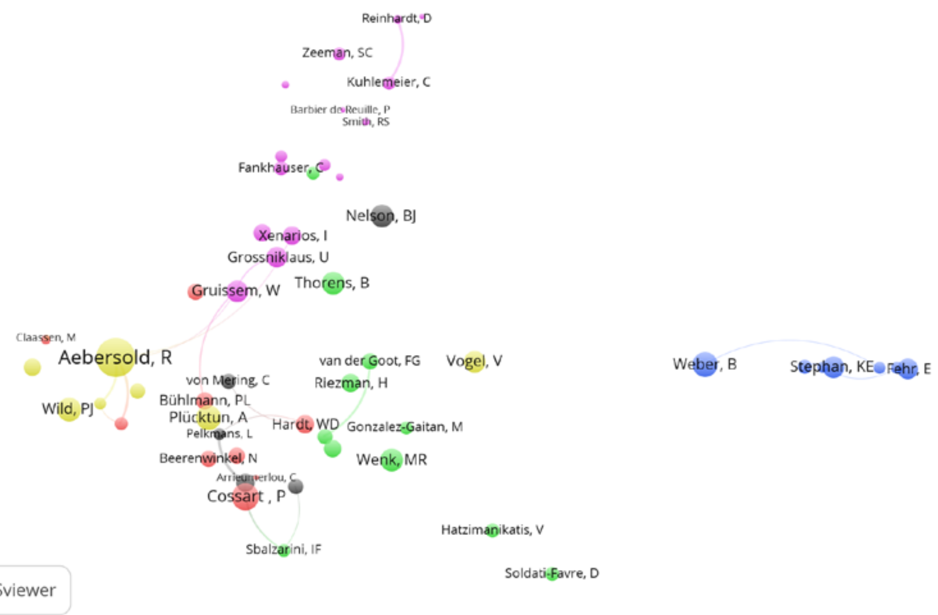


Figure 6 Collaboration map of PI's from 5 SystemsX.ch programs (2000-2007)

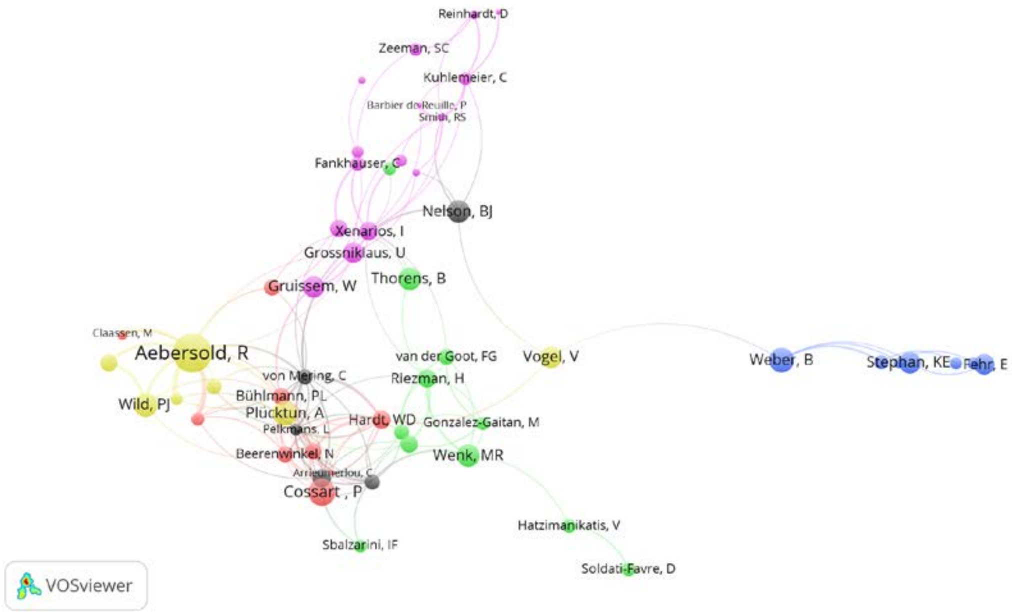


Figure 7 Collaboration map of PI's from 5 SystemsX.ch programs (2008-2015)

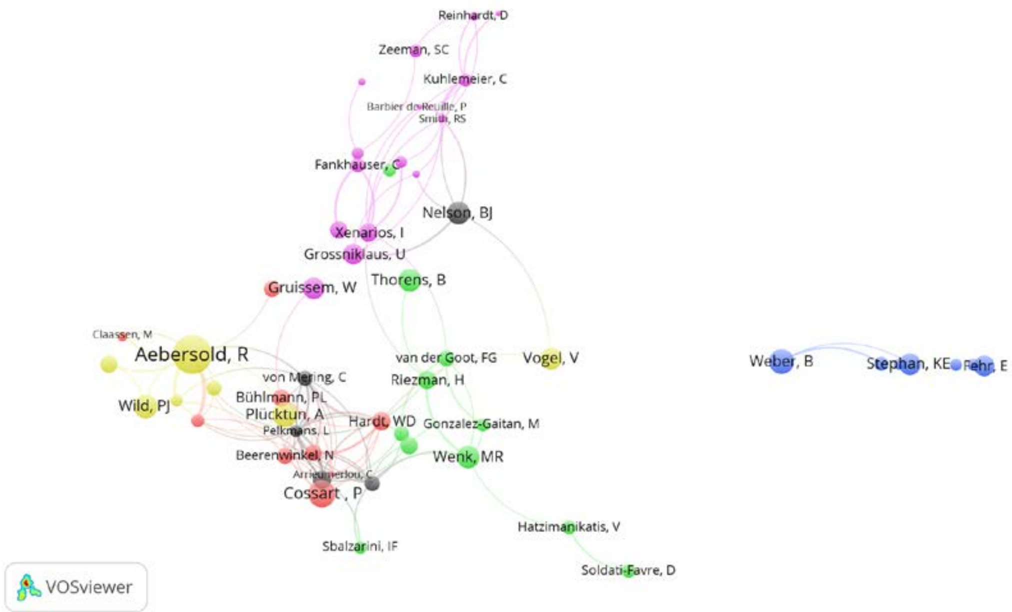


Figure 8 Collaboration map including SystemsX.ch funded publications only (2008-2015)

In each of the above networks PI's are in the same position. Only the number of connections (lines) between them differs from one network to the other.

Finally, we visualized two largest networks of SystemsX funded projects (PlantGrowth and InfectX) by zooming in and only including the PI's involved.

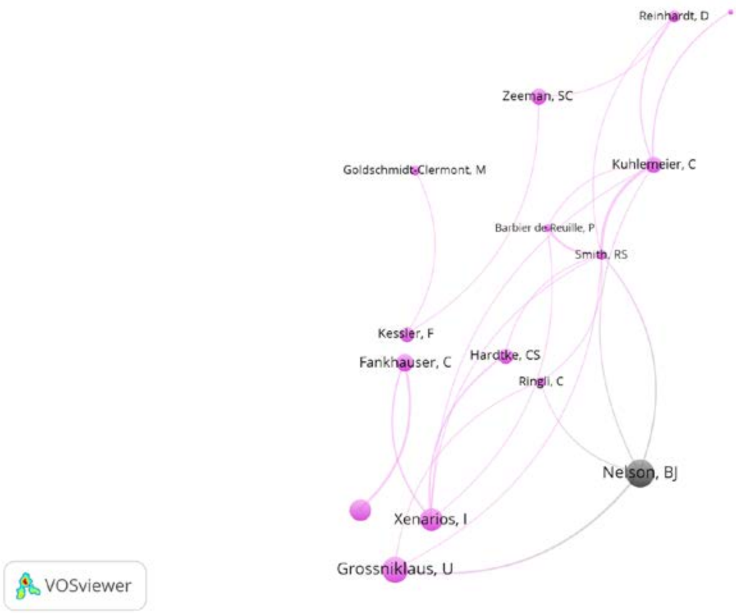


Figure 9 Collaboration network of SystemsX.ch PlantGrowth project (2008-2015)

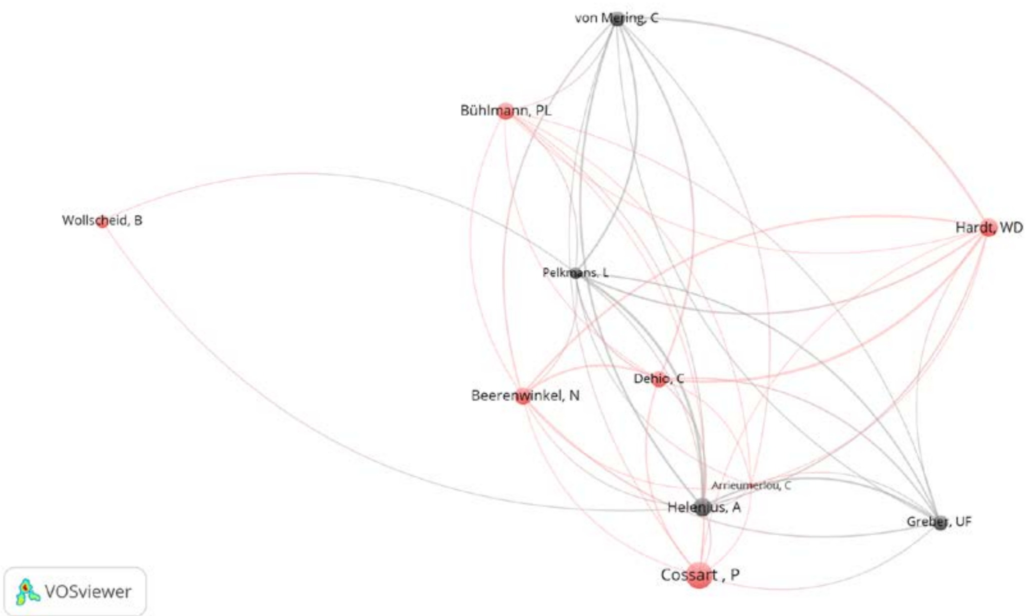


Figure 10 Collaboration network of SystemsX.ch InfectX project (2008-2015)

We used the same color-coding for PI's as in the overall Systems.ch network. Hence we can see that particularly in the InfectX project many PI's (Grey: 4 out of 11) are also involved in other projects. In PlantGrowth there is only one PI involved in another project as well.

Bibliography

Waltman, L., Calero-Medina, C., Kosten, J., Noyons, E.C.M., Tijssen, R.J.W., Van Eck, N.J., Van Leeuwen, T.N., Van Raan, A.F.J., Visser, M.S., & Wouters, P. (2012). The Leiden Ranking 2011/2012: Data collection, indicators, and interpretation. Journal of the American Society for Information Science and Technology, 63(12), 2419–2432.

Waltman, L., & Van Eck, N.J. (2012). A new methodology for constructing a publication-level classification system of science. *Journal of the American Society for Information Science and Technology*, *63*(12), 2378–2392.

Waltman, L., N.J. van Eck, Th.N. van Leeuwen, M.S. Visser, and A.F.J. van Raan, Towards a new crown indicator: Some theoretical considerations, *Journal of Informetrics*, 2011a, 5 (1), 37-47

Waltman, L., N.J. van Eck, Th.N. van Leeuwen, M.S. Visser, and A.F.J. van Raan, Towards a new crown indicator: An empirical analysis. *Scientometrics*, 2011b, 87 (3), 467-481

Waltman, L., & Van Eck, N.J. (2013a). Source normalized indicators of citation impact: An overview of different approaches and an empirical comparison. Scientometrics, 96(3), 699–716.

Waltman, L., & Van Eck, N.J. (2013b). A systematic empirical comparison of different approaches for normalizing citation impact indicators. Journal of Informetrics, 7(4), 833–849.

Ruiz-Castillo, J., & Waltman, L. (2015). Field-normalized citation impact indicators using algorithmically constructed classification systems of science. Journal of Informetrics, 9(1), 102–117.

Annex A

Indicators and parameters used for all performance analyses

```
Parameters:

Database:              All publications in Web of Science
Classification system: Publication-level classification system (about 4000 fields)
Publication window:    2008-2015
Citation window:       Fixed length of 4 year(s)
Letters:               Included (weight 0.25)
Counting method:       Full counting
Self citations:        Excluded
Top indicators:        top 10%
```

Annex B

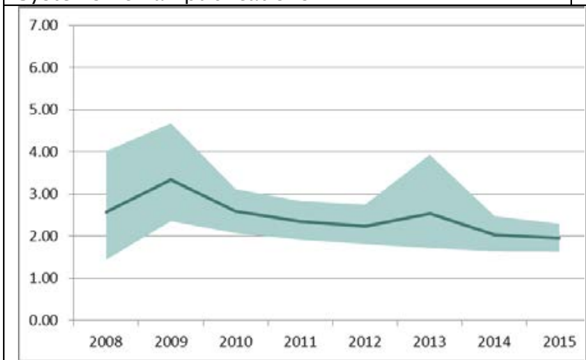
CWTS Standard performance indicators

- Number of publications (**P**) in international journals of the unit of analysis in the period.
- Total and average number of citations received in four years after publication (**TCS** and **MCS**). For these and all other citation impact measures, self-citations
- The mean field normalized citation score (**MNCS**); the actual number of citations (without self-citations) is divided by the expected number of citations on a paper basis. Here, the expected number of citations is based on the world-wide average citation score without self-citations of all similar papers belonging to the same field (cluster of the publication-based classification) and in the same year. In this way, a field normalized score is calculated for each paper. Next, the MNCS indicator is computed for each unit of analysis, by taking the average of these field normalized citation scores for individual papers. A value above 1 indicates that the mean impact for the unit is above world average whereas a value below 1 indicates the opposite.
- The mean normalized journal score (**MNJS**) indicates the average citation impact of the journals in which the papers appeared that were published by the unit of analysis. The indicator is calculated based on the same principles as the MNCS, i.e., normalized by field and publication year. It shows to what extent the analyzed unit was able to get its papers published in the higher impact journals.
- Proportion of highly cited publications (**PP[top10]**) in international journals of the unit of analysis in the period. Normalized by field and year of publication.
- Number and proportion of publications involving international collaboration (**PP IntCollab**).
- Number and proportion of publications co-authored by a university and industry (**PPui**).
-

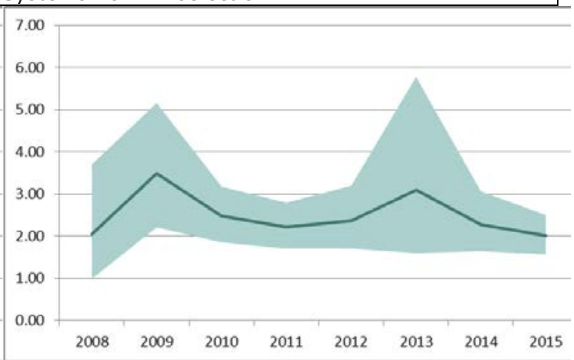
Annex C

Trend and stability analysis of SystemsX.ch, Switzerland and benchmark sets

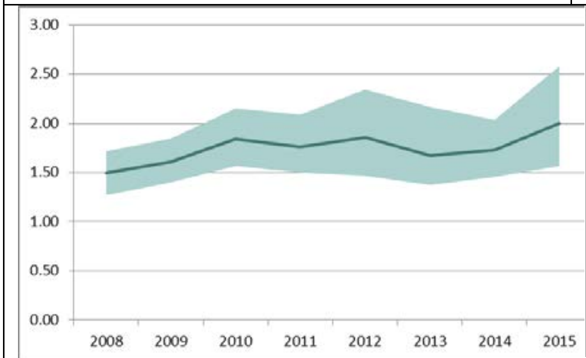
SystemsX.ch all publications



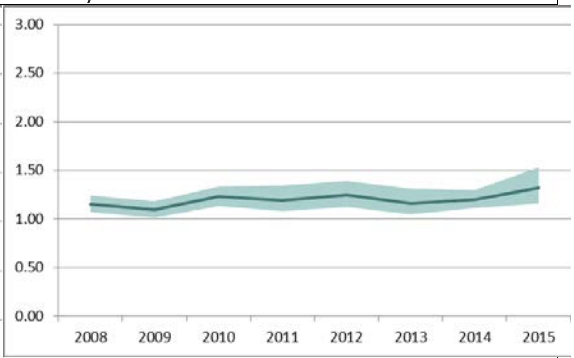
SystemsX.ch BM selection



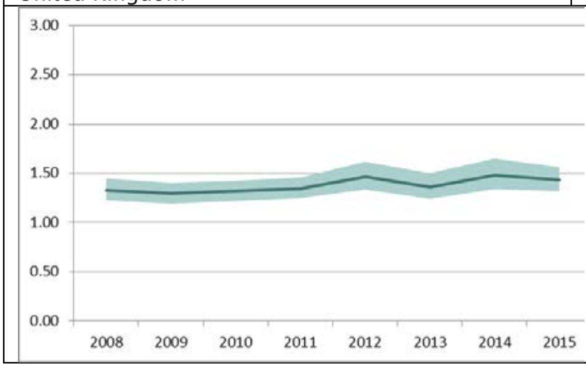
Switzerland



Germany



United Kingdom



Annex D

The CWTS publication classification system

The CWTS citation database is a bibliometric version of Web of Science (WoS). One of the special features of this database is the publication based classification. This classification is an alternative to the WoS journal classification, the WoS subject categories. The reason to have this publication based classification is the problems we encounter using the journal classification for particular purposes. We discern the following as most prominent ones.

1. *Journal scope (including multi-disciplinary journals)*

A journal classification introduces sets of journals to represents a class, in this case a subject category. This implies that journals have a similar scope. The don't need to be comparable with regard to volume (number of articles per year) but they should represent a similar specialization. This is not the case, of course. Journals represent a very broad spectrum. There a very specialized journals (e.g., *Scientometrics*) and very general ones (e.g., *Nature* or *Science* but also *British Medical Journal*). The classification scheme can therefore not be very specialized. In WoS a subject category Multi-disciplinary hosts the very general ones so that a bibliometric analysis of, for instance, the *Social Sciences* or *Nanotechnology*, using this classification, will not take papers in Nature into consideration.

2. *Granularity of the WoS subject categories*

The WoS journal classification scheme contains 250 elements. As such it is a stable system. In many cases however, it appears that these 250 subject categories are insufficient to be used for proper field analyses. The problem, however, is that the granularity of the system looks somewhat arbitrary. 'Biochemistry & Molecular biology' on the one hand and 'Ornithology' on the other, for instance, represent rather different aggregates of research. This is illustrated by the number of journals in each of them. Where the category 'Biochemistry & Molecular biology' contains almost 500 journals, 'Ornithology' has only 27. We acknowledge that there is no perfect granularity but we argue that in the WoS subject categories the differences are really too big. A classification based on more objective grounds does not solve this problem but at least is transparent.

3. *Multiple assignment of journals to categories*

In journal classifications from multi-disciplinary databases, journals are assigned to more than one category. Journals often have broader scopes than the categories 'allow'. Also here there are large differences between categories. In the example we used before, 'Biochemistry & Molecular biology,' journals are on average assigned to almost 2 categories. This means that (on average) each journal in this category is also assigned to one other category. For the more specialized category of 'Ornithology' the average is 1. This means that in this category all journals are assigned to this one only. If publications in journals with a multiple assignment would always cover the categories at stake, this should not necessarily be a problem. However, mostly it means that such journals contains structurally publications form the different categories. Therefore, publications may be assigned to two categories although they belong to just one of them.

The CWTS publication based classification scheme

An advanced alternative for the Web of Science journal classification has been developed at CWTS. It counters three major issues:

1. Journal scope (including multi-disciplinary journals)
2. Granularity of the WoS subject categories
3. Multiple assignment of journals to categories

The CWTS publication based classification is developed as described in Waltman & Van Eck (2012) . Since the first version there have been yearly updates of the system. The main characteristics of the classification are as follows.

Publication to publication citation clustering

Clusters of publications are created on the basis of citations from one publication to another. Almost 18 Million of publications are processed. The clusters contain publications from multiple years (2000-2015). Each publication is assigned to one cluster only at each level. A cluster is considered and in many cases validated as representative for disciplines, research areas, fields or sub-fields. For each cluster, we can calculate growth indices pointing at changing research foci over time.

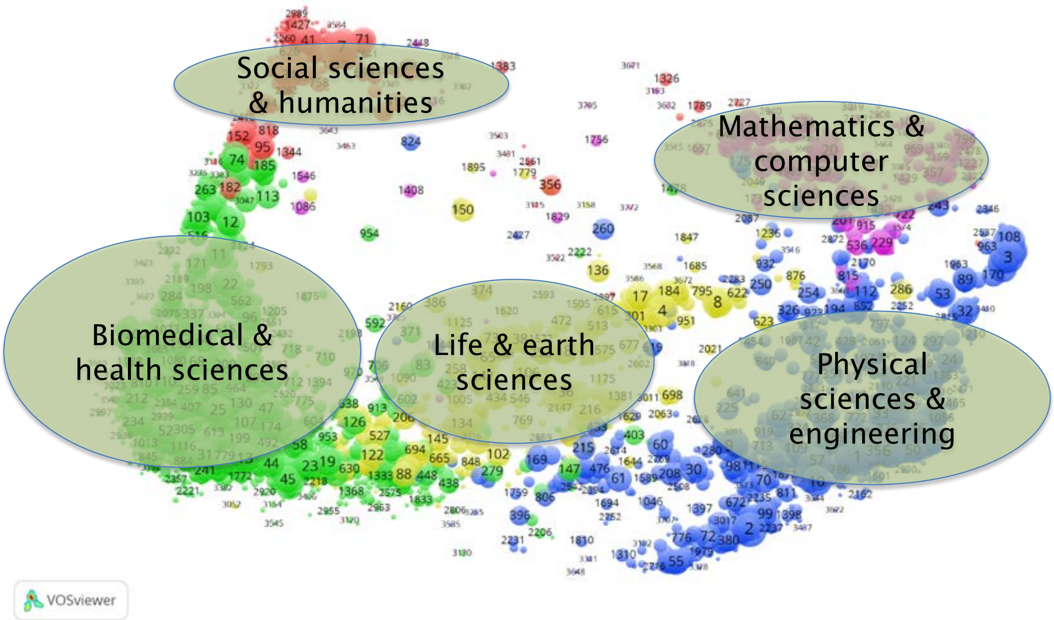
Multi-level clustering

The classification scheme has at present three different levels. The clusters are hierarchically organized. Currently we discern the following levels.

1. A top level of 27 clusters (areas)
2. A second level of 817 clusters (fields)
3. A third level of 4,113 clusters (sub-fields)

Labels

In a 'self-organized' classification scheme like ours, the labeling of clusters is the biggest challenge. As such, our clusters have no name. Still there is sufficient information available for each cluster to characterize them by suggested labels. These suggestions are based on journal categories, journal names, keywords, publication titles and key authors. An impression of our classification scheme is depicted in the VOSviewer map below. In this map the citation relations between the clusters on the second level are used to position the hundreds of clusters in a two dimensional space. The VOS mapping technique places clusters that have a strong citation traffic in each other vicinity while clusters with a weak relation are distant from each other.



Map of all sciences based on WoS publication classification (817 clusters at intermediate level)

Selection of Clusters to define the research field of SystemsX.ch and its benchmarks

The field defining the context of SystemsX.ch funded research is delimited by 24 of the 4113 clusters in which SystemsX.ch publications occurred the most. These 24 clusters are depicted in the map of all sciences (Red colored in te figure below).

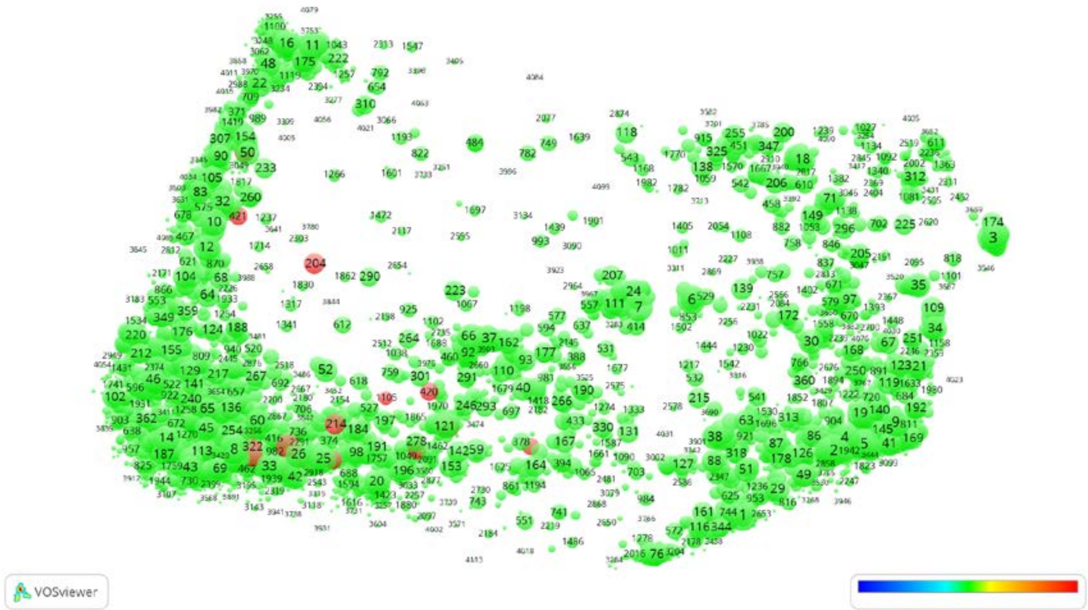


Figure 11 Positioning of SystemsX.ch research in the map of all sciences

The clusters are self-organized and as such don't have names or labels. We characterize the clusters by the most discriminative keywords derived from titles and abstracts. In the table below we describe the 24 clusters using keywords together with added the amount of publications worldwide and funded by SystemsX.ch. The map and list indicate that the fields covers a broad variety of research in the biomedical & health, life & earth as well as physical sciences.

Table 10 Clusters defining the context of SystemsX.ch funded research

id	P world 2000-2015	p systemsx.ch	Keywords to describe the cluster
53	16244	81	proteomic analysis; dimensional gel electrophoresis; quantitative proteomic; proteome; phosphopeptide
128	13795	17	ezh2; lysine; saccharomyces cerevisiae; histone; nucleosome
173	12752	53	gravitropism; microtubule; cytokinesis; endoreduplication; pollen tube
204	12229	45	layer; spike timing dependent plasticity; neuron; spiking neuron; pyramidal neuron
214	12060	13	protein interaction network; biological network; network; gene ontology; protein complex
320	10686	15	iowa gambling task; performance monitoring; error processing; negativity; decision
322	10663	30	hermansky pudlak syndrome; golgi; trans golgi network; escrt; endosome
367	10242	12	typhoid fever; salmonella pathogenicity island; serovar typhi; enteric fever; salmonella infection
393	10034	26	fission yeast; schizosaccharomyces; saccharomyces cerevisiae; kluyveromyces lactis; cytokinesis
408	9874	12	dendritic cell; expression; purification; macrophage; biosynthesis
420	9790	50	noise; gene regulatory network; delay; boolean network; bistability
421	9787	20	functional mri; fmri data; arterial spin labeling; functional magnetic resonance imaging; fmri
437	9670	39	corynebacterium glutamicum; production; metabolic engineering; glycerol; propanediol
470	9439	15	atomic force microscopy; substrate stiffness; mechanic; nanotopography; fabrication
550	8765	17	fluorescence correlation spectroscopy; green fluorescent protein; fret; gfp; diffusion
555	8722	19	sphingosine; phosphate; lysophosphatidic acid; ceramide; fty720
564	8653	12	reductase; photoinhibition; ferredoxin nadp; f complex; cyclic electron flow
681	7945	12	toxoplasma gondii; neospora caninum; congenital toxoplasmosis; seroprevalence; pregnant woman
894	6681	16	mads box gene; circadian clock; phytochrome; flc; constan
1105	5831	16	epistasis; positive selection; experimental evolution; adaptation; approximate bayesian computation
1338	5057	13	transcription factor binding site; site; chip seq; enhancer; motif discovery
1591	4315	16	morphogen gradient; hox gene; hox; gastrulation; branchio oto renal syndrome
2635	2023	17	cat scratch disease; bartonella henselae; mycoplasma; bacillary angiomatosis; neuroretinitis
2646	2004	15	laser scanning cytometry; flow cytometry; cd4; lymphocyte subset; high content screening