A Systematic Mapping of the Research Literature on System-of-Systems Engineering

Jakob Axelsson
Software and Systems Engineering Laboratory
Swedish Institute of Computer Science (SICS)
Kista, Sweden
jakob.axelsson@sics.se

Abstract – *The research area systems-of-systems engineering has increased rapidly over the last decade and now contains a substantial body of literature. To get an overview of the field, a systematic mapping of the literature has been done, covering over 3000 papers. It revealed a field massively dominated by US researchers, with an emphasis on military and space systems. A large number of people are involved, but few researchers focus on the area, and citations are fairly low compared to other fields. Important research topics include architecture, modeling and simulation, integration and interoperability, communication, sustainability, and safety and security. There are signs of immaturity within the research area, and it is recommended that existing venues are complemented with an international scientific event with very high standards for submissions.*

Keywords: System-of-systems, systematic literature mapping.

1 Introduction

Systems-of-systems (SoS) consist of operationally and managerially independent systems that cooperate to achieve a purpose through the emergent behavior of the SoS [1]. They have gained in practical importance and attention over the years, and it can be expected that this will continue as we move into an even more rapid digitization of society [2]. Being able to successfully engineer those SoS is simply a prerequisite for the society we are moving into.

SoS Engineering has also been an area of active research during a few decades, and a substantial body of literature has emerged. Given this development, we find it interesting to stop for a minute, and reflect on where we are, what has been accomplished, and what is meaningful to do research on to meet future challenges. What do we really know about how to engineer SoS? How can research better support practitioners in dealing with the future challenges?

To provide at least a partial answer to these questions, we have conducted a study of the research literature in the SoS field with the objective of providing an overview of the research area. This included which topics have been researched, and who is engaged in the research community.

Based on the review results, some observations were made about the topics, such as which were missing, and what should be the focus of future research, and on how the community could develop to be able to advance the state-of-the-art even more rapidly and improve research quality.

The remainder of the paper is structured as follows. In the next section, the methodology called systematic mapping is introduced, and we explain how it was applied in this case. Then, in Section 3, the results from analyzing collected data are presented. These results are then discussed further in Section 4, and the main conclusions from the paper are summarized in the final section.

2 Methodology

To bring light to the research questions of this paper, we need to dig into the research literature, and for this two established methodologies exist. The first, and most common one, is a systematic literature review [3], where the papers relevant to a specific research question are identified and read in detail, resulting in a summary of the results. The second approach, called systematic literature mapping [4], is shallower, and aims at giving an overview and structure to a broader area. Given the nature of this work's objectives, the second method was used.

In the next subsection, this generic methodology is explained more in detail, followed by a summary of how it was applied in this work, in terms of specific research questions, data collection, and data analysis.

2.1 Systematic mapping

A systematic mapping study, according to [4], aims at building a classification scheme and structure of a research field. The analysis focuses on frequencies of publications for categories within the scheme, leading to a picture of the coverage of the research field. The methodology, in summary, consists of the following steps:

1. Define the research scope, and the detailed research questions.
2. Conduct literature search for primary studies.
3. Screen the resulting set of papers to identify relevant ones.
4. Build a classification scheme, by identifying keywords in the abstracts of the papers.

5. Extract data from the papers, and classify them according to the scheme, resulting in a map of the area.

We followed the above steps in this work, and in the following subsections, the details will be explained.

2.2 Scope and detailed research questions

The literature of interest in this study deals with engineering of SoS, and the overall research objective is related to the overall structure of this research area, as described in Section 1. In more detail, the following research questions are emphasized:

1. How has the field SoS developed over time? The metric we use is number of publications per year.
2. What is the geographical distribution of SoS research? The metric is the number of publications per author country.
3. Who are the leading researchers in the field? Metrics are the number of publications, the current number of citations, and the h-index, which is a combination of the two.
4. Which are the key papers in the area? The metric is the current number of citations per paper.
5. Which are the publication sources (journals, conferences, etc.)? The metric is the number of publications per venue.
6. Which are the main application areas? The metric is the frequency of papers being classified in each application area, as defined in the scheme.
7. Which are the most important research topics? The metric is the frequency of terms from the classification scheme.

2.3 Data collection

The data for this study was extracted from the database Scopus, which is provided by Elsevier but also contains data from many other publishers. It claims to be the largest research database in the world, and an initial sampling confirmed that it did indeed contain data from many of the sources where we had prior knowledge of SoS literature appearing. A drawback of the database is its focus on recent literature, primarily after 1995, but since we expected the bulk of work in SoS to be after this date, it was considered a minor problem. Scopus has previously been shown to give results of higher quality compared to open databases such as Google Scholar [5], which was also a factor in the decision on which database to use.

The search string used was simply “system-of-systems”, appearing in the title, abstract, or keywords fields of the database, for any publication year. The database was intelligent enough to also include results using other ways of writing the term, such as “system of systems”.

The results were exported to a text file, containing all available data fields in the database. This file formed the basis for the further analysis. The data collection was performed on January 8, 2015, and resulted in 3274 papers.

2.4 Screening

A screening was performed on the data to check for relevance. It turned out that 112 of the papers did not have any authors listed, and a closer inspection revealed that these were references to entire proceeding volumes. Those records were thus removed. A further limitation was to remove papers written in other languages than English, removing a further 140 papers (125 of them were in Chinese). The number of papers carried forward to analysis was thus 3022.

2.5 Classification

Our previous experiences with systematic literature reviews and mappings, is that they become very time consuming if done too rigorously, and there is always a risk of digging too deeply into each paper. In this work, we chose to use an iterative approach for the analysis part (step 4 and 5), where a random sample of papers was selected in each iteration. In this way, it was not necessary to manually go through all identified papers, but instead the classification scheme was built iteratively, and when the frequencies of different categories converged (which was tested using basic statistic methods such as confidence intervals for binomial distributions), the analysis stopped. In total, 116 papers were analyzed in this way. At this point, a saturation level had been reached in the number of categories in the classification scheme.

The analysis was done semi-automatically, using a script written in the programming language Python. By encoding all steps of the analysis in this script, even the manual ones, the analysis becomes completely repeatable, and will always yield the same result given the same data file. Many steps, such as extracting data about publication years, authors, citations, etc., were completely automated. The output of the script was a textual and graphical report of the detailed analysis results.

In some cases, the automation needed some manual assistance. This included the classification of publication sources, since the same source appeared with slightly different names. For example, the IEEE System of Systems Engineering Conference appeared under about 10 different names, due to various abbreviations, and sometimes inclusions of the conference subtitle. This could only be resolved manually. In the same way, keywords are not written in a standardized way, so keywords such as “system-of-systems” appear in many different variants that need manual grouping.

The classification scheme was built iteratively using a tree structure, by reading the abstracts and identifying key terms. Prior to this, a tentative first level structure was put in place. It contained the high-level areas Process, Technology, Application area, Property (capturing different non-functional characteristics and quality attributes), Tools, Methodology, Business, and Organization. Most of these areas turned out to be meaningful, and subareas to them were added as the iterative analysis proceeded.

During the manual classification, some papers were also classified as irrelevant if the abstract revealed that they were out of scope. A typical example of this was when they mentioned SoS as a potential application area, but without really contributing to SoS engineering. Very few of these papers were totally unrelated, and it was sometimes a borderline decision.

3 Results

In this section, the results of the systematic mapping are presented. It is based on the 3022 papers identified after screening, with a detailed classification of 116 randomly selected papers, as described above. In the sample, 16 papers were classified as irrelevant, corresponding to 14% of the total literature base. This is a reasonably low number, indicating that the literature search yielded a reliable result.

3.1 Time and geography

The first research question was how the SoS field has developed over time, and this is illustrated in Figure 1. Since the Scopus database is focusing on recent papers, the graph starts at 1995, but when compared to other sources, papers before that date are sporadic. Since the data was extracted very early in 2015, it is not likely that the data for 2014 was completed yet, so the ending dip does not reflect a trend and if the study is redone later, this number is likely to increase. As can be seen from the graph, activity started to increased sharply around 2003-4, and then reached the current level about five years later.

The second research question concerned the country of origin. Here, the metric used was the number of publications with at least one author from that country. The result is summarized in Figure 2, where the first (blue) bar shows the overall count for the 15 most common nations.

As the graph indicates, US domination is almost overwhelming, with an order of magnitude more papers than each of the following nations. Among the other top 15 countries, 13 are from Europe. It is worth commenting on China, and recalling that the initial screening actually removed 125 papers in Chinese. If they were included,

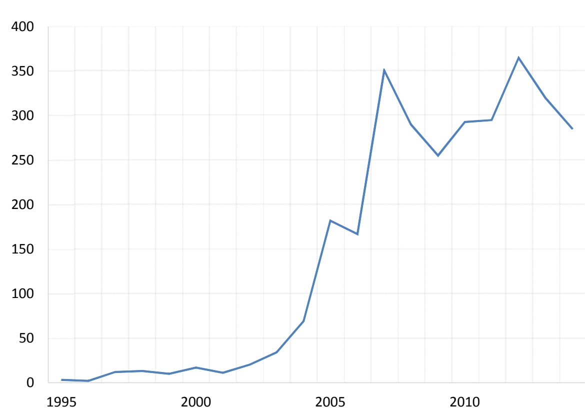

Figure 1. Number of papers per year.

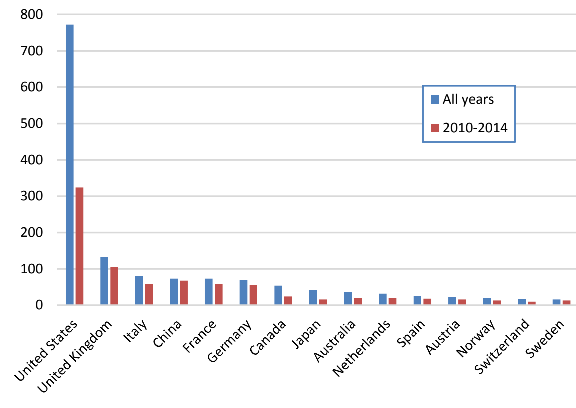

Figure 2. Number of papers per country.

China would be in a clear second place, and there is thus substantial research activity in this country. However, a lot of it is not connected to the international research community, due to the use of language.

The second (red) bar in the figure shows the count for the last five years, i.e. 2010-2014. The top 15 countries are still the same, but the US dominance is less pronounced, and many of the other countries actually have the majority of their publications in this period.

3.2 Authorship, papers, and citations

In the data set, there are 5632 individual authors (with some reservation for variations in spelling that could lead to the same author being counted several times). On average, there were approximately 3.0 authors per paper. Around 56% of the authors were affiliated with academia, and the rest with industry, government agencies, etc.

Research question 3 concerned who the leading researchers in the field are, and this is something which is a bit hard to measure. Common metrics include the number of publications, but also the number of citations. Nowadays, it is customary to also combine those two metrics into an *h*-index, where *h* is the largest number such that a given author has published at least *h* papers that each has at least *h* citations.

Some authors, such as Boehm, have been highly productive outside the area of SoS. Here, we have only included citations to papers within SoS, but not citations to papers in other fields by the same author, since we are not evaluating the total contribution of these authors, but rather their contributions to the SoS field. (For the citing papers, however, no discrimination was made regarding their topic.) The same principle applies for calculating the *h*-index.

The 10 most productive authors in the SoS field are listed in Table 1, showing their total number of papers. The 10 most cited authors are shown in Table 2. As can be seen, DeLaurentis, Keating, Sauser, and Boardman appear in both tables.

The results for *h*-index are fairly low. Keating, Sauser, and Boardman each reach $h = 7$, with Jamshidi and

Table 1. Number of papers per author.

Author	No. papers
DeLaurentis, D	67
Mavris, D	53
Jamshidi, M	38
Keating, C	32
Sauser, B	28
Lane, J	27
Dagli, C	26
Michael, J	26
Boardman, J	23
Gorod, A	22

Table 2. Number of citations per author.

Author	No. citations
Keating, C	307
Jackson, M	288
Keys, P	223
Sauser, B	210
Sousa-poza, A	210
Boardman, J	198
DeLaurentis, D	181
Rabadi, G	166
Boehm, B	165
Unal, R	162

DeLaurentis at $h = 6$. The total set of papers has $h = 32$.

Research question 4 was about the key papers in the area. The 10 most cited papers are listed in Table 3 (see reference list for complete details). It is worth noting that about 2/3 of the papers have never been cited.

3.3 Publication sources

Research question 5 concerned publication sources. In the data set of over 3000 papers resulting from the search and screening, the vast majority (74%) were conference publications, followed by journals (21%). The remaining papers included book chapters, reviews, editorials, etc. A more detailed view of the publication sources is provided in Figure 3 that indicates the number of publications covering the 15 most frequently used sources.

As could be expected, the arenas provided by the IEEE and INCOSE are the most frequented ones, but it is also interesting to see a large amount of research published by the Society of Optical Engineering. It should be mentioned that there is a very long tail in this distribution, with a total of close to 1000 source titles (about 1/3 with only one publication).

3.4 Application areas

Research question 6 concerned the most common application areas, and the metric was based on the manual classification of the 100 papers that remained in the sample after removing irrelevant papers. The results here were very

Table 3. Most cited papers.

Title	Year	No. cit.
Towards a system of systems methodologies [6]	1984	223
System of systems engineering [7]	2003	160
Integration of quality and environmental management systems [8]	1998	98
System of systems (SoS) enterprise systems engineering for information-intensive organizations [9]	2001	92
A system-of-systems perspective for public policy decisions [10]	2004	84
Some future trends and implications for systems and software engineering processes [11]	2006	74
System-of-systems engineering management: A review of modern history and a path forward [12]	2008	70
Critical infrastructures at risk: A need for a new conceptual approach and extended analytical tools [13]	2008	69
A theory of enterprise transformation [14]	2005	66
The emerging joint system of systems: A systems engineering challenge and opportunity for APL [15]	1992	57

clear: 21% of the papers were from the Military domain, followed by 11% from the Space area. However, there were also 18 other application areas mentioned, each with between 1-5%, including Health care, Disaster management, Aircrafts, Robotics, Power systems, etc. 67% of the papers related to at least one application area.

An area worth special comments is earth observation. Many papers addressed the Global Earth Observation System of Systems (GEOSS). A number of these papers were however classified as irrelevant, since they did not relate to the SoS aspects of GEOSS, but rather to some component, algorithm, etc. to be used in that SoS.

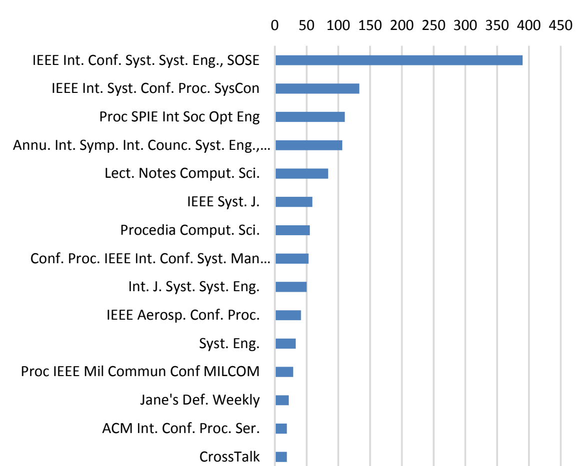

Figure 3. Number of papers per publication source.

3.5 Research topics

The final research question was about the most important research topics, as deduced by the manual classification of the 100 papers. The first area we studied was Processes, where we tried to relate the papers to different life-cycle processes. Of the papers, 56% related to processes of some kind. The ones that stood out were Modeling (22%) and Integration (17%). At a somewhat lower level, Risk management (9%), and Requirements and Design (each 7%) were significant.

A second area was Properties, and not surprisingly, a wide range of properties (30 in total) were mentioned, and many papers (43%) made reference to at least one property. The ones that were mentioned more than just a few times were: Sustainability (8%); Interoperability (7%); Cost, Effectiveness, Safety, and Security (each 6%); and Efficiency and Reliability (each 5%). Although the sample does not allow us to determine the exact ranking of these properties in the total set of papers, they are all likely to be important characteristics of the SoS research area.

The third area concerned Technology and design, to which 49% of the papers made reference. The dominating theme in this area is Architecture (23%), with Communication (14%) in a clear second place. Further behind, Sensors and Interconnections (each 7%) are also worth mentioning.

The fourth area was Tools (mentioned in 20% of the papers), and the only significant finding here was Simulation (12%).

In the high-level areas Business and Methodology, no clear patterns emerged from the literature, and fairly few papers (2% and 13%, respectively) touched these topics.

4 Discussion

Based on the results from the previous section, we will now discuss certain aspects of the field. This discussion is more subjective, and based on interpretations and extrapolations from the data. The focus is on how SoS engineering research can advance further in the future.

4.1 Overall status of the research field

When looking at the research field on a high level, it is interesting to notice that this is an area that attracts a large number of people. However, half of them have contributed to only 5 papers or less. We see two possible explanations. One is that researchers from other fields occasionally touch upon the SoS subject, but do not see it as their primary concern. The other is that there are many practitioners who sometimes contribute, but their main occupation is to build real systems, and not write research publications.

Another striking observation is the citation practices. The number of citations to the most cited papers in the collection is very low, compared to other fields, and the same goes for citation count and h-index for the leading individuals. Although there are certainly excellent exceptions to this, it appears that much of the research is not systematically building on previous research, in a way that is otherwise characteristic of a mature and well-functioning research area.

One must also comment of the unusual dominance of one country, the USA. Although it appears that a shift towards a more even balance has already begun, it is vital for the field to even further reach out to the global community of researchers. In particular, links with Chinese researchers who primarily publish in their own language today, could inject new energy into the field.

4.2 Broader application areas

Another unbalance in the publications is the preoccupation with military and space applications. Although these are important, they are surrounded by very particular conditions, such as advanced acquisition models, very small production volumes, trained users, etc. At the same time, there is vibrant development in other parts of society, with research and development in cyber-physical systems, Internet of Things, and software ecosystems, that all relate to SoS. Getting closer ties with these communities would provide a broader range of example applications and building blocks, and hence lead to an improved understanding of SoS in general. This knowledge would certainly spill back on military and space applications as well. It would also be valuable to study the business models for commercial SoS usage.

4.3 Systematic use of empirical data

One of the strengths of the SoS area is that there is a strong interaction between practitioners and researchers, a fact upon which other communities would look with envy. However, it does not appear that researchers take the maximum scientific advantage of this. Most of the papers reporting real applications appear not to be based on any systematic empirical research methods, such as case studies or experiments, but are mainly providing anecdotal evidence. This makes it difficult to draw sharp conclusions, to compare different studies to one another, and to connect the data to theories.

SoS researchers should seek inspiration from the Software Engineering domain. When that domain was at a similar stage as SoS today, leading researchers advocated the use of systematic methods and basing research on empirical data. This has greatly contributed to the development of the area and are now accepted as mandatory practices for submissions to top conferences and journals.

4.4 Focus on core principles of SoS

Once systematically collected empirical data is available, it becomes possible to study and formulate sound theories that can be validated against further data. Many of the topics that stand out in the study, such as architecture, modeling and simulation, integration and interoperability, communication, sustainability, and safety and security are likely to deserve continued attention. However, researchers need to even clearer formulate in what ways these topics have to be handled differently for SoS, than for other systems. There is thus room for improving the understanding of the core principles of SoS, and connecting that to various properties, technologies, etc.

4.5 Validity of the study

As with any other empirical study, there are many threats to validity of the findings in this paper as well. The selection of literature database, the search string used, the manual and automatic processing, and the use of sampling can all lead to less confidence in the results. For this reason, triangulations have been used, for instance to compare the Scopus results with searches in Google Scholar, to compare the results of the manual classification with the keywords selected by authors and librarians, and to use confidence intervals for the sample statistics. Our impression is that the overall results in Section 3 are likely to be similar even if the method would be changed in different ways.

5 Conclusions

In this paper, we have presented a systematic literature mapping of the SoS area, including over 3000 research papers. In summary, the main findings are that the field is dominated by US researchers, and that there is a very strong focus on military and space applications. A large number of people are involved, but few persons focus on the area, and citations are fairly low compared to other fields. Some of the key research topics include architecture, modeling and simulation, integration and interoperability, communication, sustainability, and safety and security.

In our view, there are signs of immaturity within the research area, with only limited use of systematic empirical methods that are common in other domains, and also that new research results are not building systematically on previous research.

To improve this situation, we would recommend the creation of an international scientific event with very high standards for submissions. This would give the leading researchers in the field an opportunity to build a community and to focus on the scientific study of SoS engineering. To build on one of the strengths of the area, namely the good connection to practitioners, the event should ideally be co-located with an existing event, such as the IEEE System of Systems Engineering Conference.

Acknowledgements

This work has been supported by Vinnova, the Swedish Agency for Innovation Systems, grant no. 2014-05923.

References

[1] M. Maier. Architecting Principles for Systems-of-Systems. Proc. INCOSE Intl. Symposium, pp. 565-573, 1996.

[2] E. Brynjolfsson, A. McAfee. *The second machine age*. W.W. Norton, 2014.

[3] B. Kitchenham, S. Charters. Guidelines for performing systematic literature reviews in software engineering. Tech. Report EBSE-2007-01, School of Comp. Science and Mathematics, Keele University, 2007.

[4] K. Petersen, R. Feldt, S. Mujtaba, M. Mattsson. Systematic Mapping Studies in Software Engineering. In Proc. 12th Intl. Conf. on Evaluation and Assessment in Software Engineering, pp. 68-77, 2008.

[5] M. E. Falagas, E. I. Pitsouni, G. A. Malietzis, G. Pappas. Comparison of PubMed, Scopus, Web of Science, and Google Scholar: Strengths and weaknesses. FASEB Journal 22 (2): 338–42, 2008.

[6] M. C. Jackson, P. Keys. Towards a system of systems methodologies. Journal of the operational research society, pp. 473-486, 1984.

[7] C. Keating, R. Rogers, R. Unal, D. Dryer, A. Sousa-Poza, R. Safford, G. Rabadi. System of Systems Engineering. Eng. Management Journal, 15(3), 2003.

[8] S. Karapetrovic, W. Willborn. Integration of quality and environmental management systems. The TQM Magazine, 10(3), 204-213, 1998.

[9] P. G. Carlock, R. E. Fenton. System of Systems (SoS) enterprise systems engineering for information-intensive organizations. Systems Engineering, 4(4), 242-261, 2001.

[10] D. DeLaurentis, R. K. Callaway. A System-of-Systems Perspective for Public Policy Decisions. Review of Policy Research, 21: 829–837, 2004.

[11] B. Boehm. Some future trends and implications for systems and software engineering processes. Systems Engineering, 9(1), 1-19, 2006.

[12] A. Gorod, B. J. Sauser, J. T. Boardman. System-of-Systems Engineering Management: A Review of Modern History and a Path Forward. IEEE Systems Journal, 2(4), 484-499, 2008.

[13] W. Kröger. Critical infrastructures at risk: A need for a new conceptual approach and extended analytical tools. Reliability Engineering & System Safety, 93(12), 1781-1787, 2008.

[14] W. B. Rouse. A theory of enterprise transformation. Systems Engineering, 8(4), 279-295, 2005.

[15] W. H. Manthorpe. The emerging joint system of systems: A systems engineering challenge and opportunity for APL. Johns Hopkins APL Tech. Digest, 17(3), 1996.