Systematic and open exploration of FaaS and Serverless Computing research

Mohammed Al-Ameen[1,2], Josef Spillner[2]
[1]University of Sharjah, OpenUAE research group, United Arab Emirates
[2]Zurich University of Applied Sciences,
School of Engineering, Service Prototyping Lab (blog.zhaw.ch/splab), Switzerland
{alam,josef.spillner}@zhaw.ch

1 Introduction

Research interest in Function-as-a-Service (FaaS) development, execution and ecosystems is growing. Consequently, an increasing body of literature focusing on FaaS and cloud services is evolving. While the field is still young, we propose a community-maintained and curated open dataset which uniquely and umambiguously references relevant articles in order to derive comparable bibliometric data and statistics. The dataset supports the generation of knowledge about the evolving history, research trends and significance. This survey enablement paper introduces the 60-article dataset, explains the governance model and benefits, and shows first insights derived by a literature analysis. We argue that along with accelerating technological trends, fresh research method flavours assist in faster and more comprehensive knowledge exploration and dissemination.

2 Background Information

For many years, software technologies evolved independently on two main axes: cloud computing to deliver application services from software packaged in long-running virtual machines and containers, and discrete event processing to process data on demand while remaining idle otherwise. In 2014, the axes converged as cloud-hosted, elastically scalable and usage-billed event processors emerged in the form of short-lived, memory-constrained and almost-stateless cloud functions. The resulting *Function-as-a-Service (FaaS)* offering was adopted by major cloud platform providers. While the developer community quickly took note of this promising execution model and saw the introduction of the related term *serverless computing*, it took researchers until May 2016 to publish the first articles comparing FaaS services, exploring FaaS runtimes or analysing the FaaS cost models [VGO+16]. Owing to the popularity of the term serverless computing, but also due to increasing discussions of appropriate architectures and support services beyond FaaS, we decided to name the dataset accordingly.

This work is not a survey, but rather an enablement for future surveys and systematic literature reviews (SLRs) with anticipated high quality, consistency and comparability. Its extensible tree structure will not be sufficient for all use cases, including graph analysis which would require directed graphs with annotations or ontology representations, but it does also not preclude the production of such enhanced representations. Moreover, beside the literature view, it will allow for different views including one on FaaS technologies which will raise interest in derivative works with software and cloud engineers in industry.

3 Serverless Literature Dataset

As with most areas of research, the body of literature on FaaS, serverless computing and the serverless ecosystem is growing quickly. Traditionally, survey articles would filter and consolidate the important contributions from selected articles. But often, the selection process and the source of presented metrics cannot be traced or reconstructed, leading to non-reproducible research works. While we think that the larger share of publications on FaaS is still ahead, we claim that we need to produce an early framework for enabling data-driven reproducible surveys, and contribute a community-curated dataset along with diverse maintenance and statistics assembly code. At the time of writing, the public dataset contains 60 articles covering the years 2016, 2017 and the first half of 2018 (January to July) while the queue of new articles to be still included for the second half of 2018 already encompasses a known set of around 20 articles.

Governance. Our dataset is published [Spi18] and semi-regularly or upon request updated with versioned Digital Object Identifier (DOI) in the *Serverless Computing* community at Zenodo, an open science tool for scholarly processes and research outputs in the form of digital artefacts. Any researcher can suggest a new version and, as long as the changes are only additive or corrective in nature, the upload will be accepted by the dataset maintainer. Any contributor qualifies as co-maintainer in order to ensure the long-term distribution of maintenance tasks. The precise extent of governance remains unknown for now due to a lack of comparable datasets, but upon being informed about this model, for instance at ESSCA 2018, many researchers in the field have signalled interest and support.

Population. To ensure quality publications with comparable results, it is mandatory that publications appear indexed in the DBLP computer science bibliography in order to qualify. Moreover, the publication must be found with a DBLP title keyword search for the terms `serverless application`, `serverless computing`, `serverless` in general, `function-as-a-service`, `lambda` or `cloud function`, or despite absence of a title match evolve closely around these topics. Terms in risk of overgeneralisation, such as `serverless` and even more so `lambda`, similarly require brief title or even article reading to decide on the eligibility for inclusion in a manual and potentially error-prone post-filtering step.

Further indexation services and keywords can be agreed on by the community dataset maintainers as the technology evolves, not just to capture more works, but also to subdivide sets of works into specialised subsets. Similarly, manual overrides are possible when works or entire collections of works are inadvertently missing from or misrepresented in DBLP, which is known to still occur sometimes despite best efforts to prevent mistakes [RH11].

The dataset is populated in a structured way, starting with manually assigned unique and consecutively increasing identifiers and associated unique DOIs, if available, captured in a first file. A script then fetches bibliographic details and amends the metadata in a second file. For preprints without assigned DOI, a manual addition is possible. Further metadata is added manually to two additional files, one matching the publications structure and one orthogonally capturing technology aspects. The resulting data structure with its metadata attributes appears as follows:

```
id / DOI
→ {title,author,journal,year}
  (automatically populated)
→ {countries,institutions,...}
  (manually annotated)
↪ {technologies,open source,...}
  (manually annotated)
```

Most manual curation steps could potentially be automated or semi-automated by querying indexation services and performing semantic NLP on the full-text works. However, at the time of writing and due to the still small number of works, a manual process has been chosen to keep the initial effort low and stimulate early community involvement. Furthermore, automation is non-trivial due to the need to disambiguate in a context-aware way the pure mention of terms from their detailed study.

Representation. All files are represented as four structured and extensible JSON files which are either manually or in the case of the bibliographic information file automatically ordered through the maintenance scripts. Automatic ordering eases maintenance by suppressing diff noise, but is not always easy to implement due to lexicographic versus numeric ordering (e.g. 1, 10, 2). The files encompass around 1330 lines containing 970 key-value assignments.

In order to gain insight into all works, the dataset contains the provision to store the PDF files of all publications in a subfolder. We have assembled this *companion dataset* and will use it in this article exemplarily but for copyright and licencing reasons are not able to distribute it publicly. The companion dataset combines 560 pages of research communication on FaaS-related topics and has a cumulative size of 43 MB.

Verification. In order to ensure a high quality and validity of the dataset, we run the included consistency checking scripts and we manually verify the completeness with external samples. Concretely, we perform a cross-check with the previous proceedings of the International Workshop on Serverless Computing (WoSC) in conjunction with a DBLP countercheck and with Google Scholar to detect whether all relevant publications have been included in the dataset. As a result, we found an additional paper not matching any search terms, one more matching `lambda` which slipped through the manual post-filtering, as well as three more papers which were not originally available on DBLP

Table 1: FaaS-related publications per year

Year	Publications	Notes
2016	7	
2017	26	
2018	27	partial; interpolated: ca. 45

but at cross-check time they had already been indexed. All five papers are already included with the dataset. More importantly, we found proceedings of one of the WoSC editions which are distributed via the ACM Digital Library but, for unknown reasons, not yet available in DBLP. We have marked these papers as potential addition. Finally, a trivial search on Google Scholar revealed no additional articles. In summary, our approach to find credible quality articles about FaaS systematically is working but assumes a timely indexing into DBLP and still causes occasional omissions.

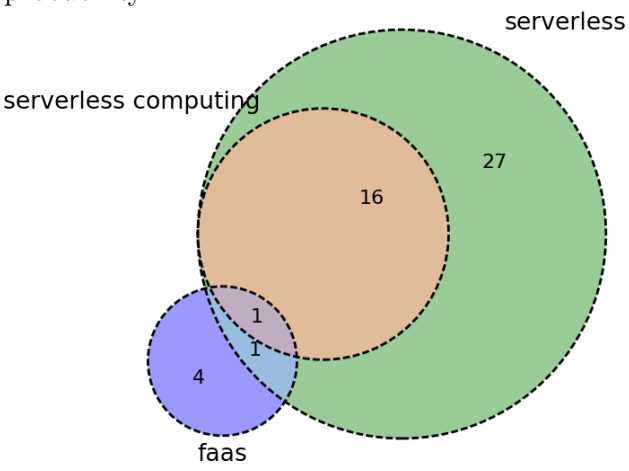
Exploitation. By not only sharing bibliometric and content data, but also scripts to produce associated figures, our reusable dataset leads to standardised visuals which allow for comparison across published works. All figures in this article have been produced by the scripts contained in the dataset package with some manual additions based on generated numbers and we suggest future publications on the same topic do the same. Among possible exploitation routes are state-of-the-art sections in research proposals and papers, as well as detailed surveys and systematic literature reviews.

4 Bibliometric Analysis

Key Metrics. Covering the years 2016, 2017 and most of 2018, the serverless literature dataset contains a total of 60 articles. Of those, 45 have a DOI assigned but 15 have not. The publications per year are shown in Table 1. The growth is remarkable; while focused researchers have been able to maintain an unassisted overview for the first two years, a systematic collection has become indispensable for any further holistic view on the field.

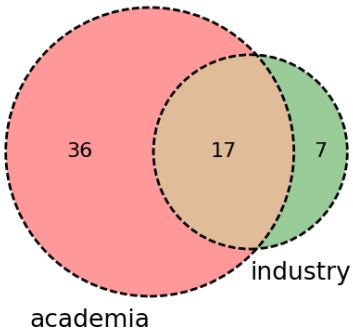
The most successful and growing search term is `serverless` with 27 occurrences of which two thirds are complemented with `computing`. On two occasions, the corresponding title also mentions `faas` which also appears in four other titles. All other search terms only occur individually or in single pairings. A Venn diagram showing the matching keyword relations is shown in Fig. 1.

Figure 1: Overlap of search terms with major match probability



The ratio of academic to pure industrial to mixed academic-industrial research is 36 : 7 : 17, with 87% of works involving academic institutions. On average, each publication involves authors from 1.8 institutions, with 68 institutions (subsuming all sub-units such as different research groups or labs) being involved in total. Fig. 2 contains the visual overview about the institution types.

Figure 2: Overlap of type of author institution



The most active countries in absolute terms of institutions publishing are the US (26), Switzerland (7) and Canada (4) followed by Germany, Spain, Colombia and Austria (all 3). In total, research on FaaS is documented to happen in 21 countries across six continents. Fig. 3 gives a geographical overview about the countries with publishing activity.

Finally, the selection of publication paths shows

Figure 3: Distribution of FaaS-related publications across countries of institution

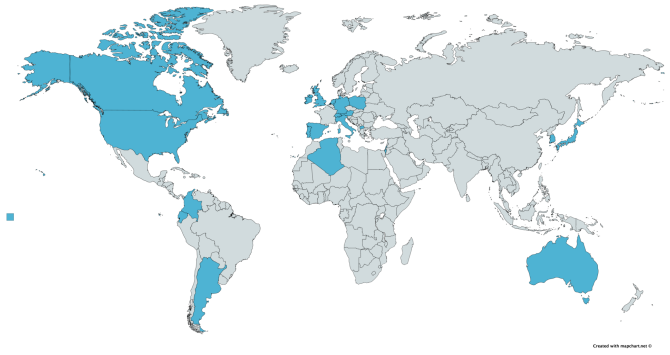


Table 2: FaaS-related publications per publisher

Publishing path	Type	Publications
IEEE	P	20
ACM	P	11
USENIX	O	9
arXiv	O	8
(other)	–	7
Springer	C	4
Elsevier	C	1

some interesting characteristics in Table 2. There is a mix of publishing through professional societies (P; 31 or 52%), open proceedings via arXiv or USENIX (O; 17 or 28%) and commercial publishers (C; 5 or 8%).

5 Content Analysis

This analysis builds on the private companion dataset containing all PDF representations of the articles. A script then converts these PDFs into text files and concatenates the output into one large file (2.2 MB) which can be uploaded to a word cloud service to turn the list into sets with specified cardinality of occurrence. The set of words with 50 occurrences or more is contained in the dataset to allow for tracking of trends, encompassing 693 top words. Of interest is that unambiguous subject-specific words such as `cold start`/`coldstart`, `handler(s)` and `stateless` do appear in the top words list but with less than 200 occurrences each, i.e. below the top 20% of the list.

The word cloud services furthermore produce a visual representation of the term frequency. Fig. 4 shows an exemplary unfiltered result over all top words. It is evident that terms such as `function(s)` (2936 times), `serverless` (1517 times), `Lambda` (1081 times), `time` (853 times) and `FaaS` (409 times) appear very often, but so do stopwords such as `can` or `use` which at the time being are not yet filtered automatically and will also be subject to increased automation in future versions of the curation scripts. Nevertheless, the word cloud reveals researcher concerns about specific technologies and characteristics, including also considerations of data, messages, containers, execution and requests. For example, on the first half of documents, `Lambda` (328 times) occurred more often than `serverless` (310 times), signalling a decline in prominence. Nevertheless, in terms of concrete implementations or services mentioned, `Lambda` still leads ahead of `Functions` (402 times), which in the capitalised form presumably refers to Google, Microsoft and IBM offerings, `OpenWhisk` (285 times) and `OpenLambda` (88 times).

Figure 4: Word cloud with occurrence-proportional font size highlighting key terms in FaaS-related publications

6 Technology Analysis

In the dataset, we aggregated information about the technologies prominently referenced in the studies and experiments. In total, 22 different technologies including FaaS runtimes, tools and commercial services

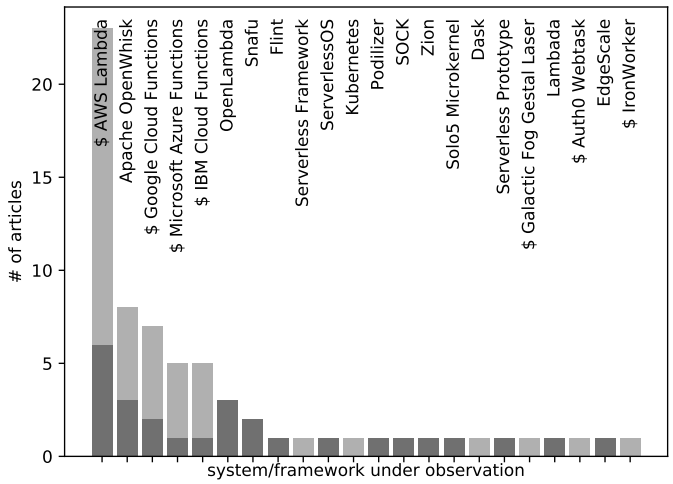


Figure 5: Serverless technologies referenced in academic publishing (dark grey: authors are among primary developers of technology)

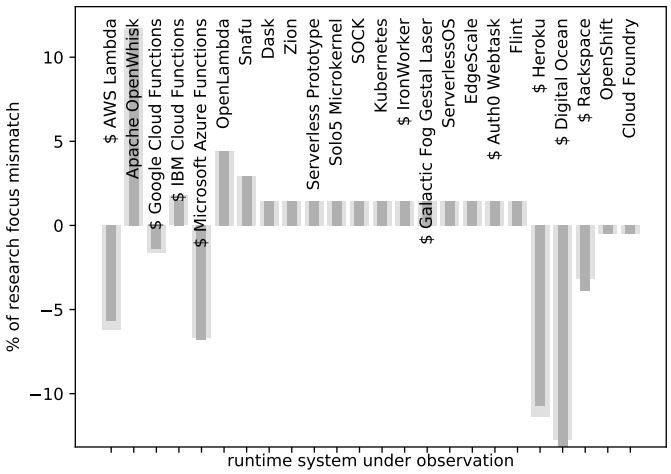


Figure 6: Mismatch between academic publishing and developer focus (dark grey: between academic publishing and self-identified serverless expert developers)

could be identified of which 13 are available under open source and free software licences. Among them, AWS Lambda (23), OpenWhisk (8), Google Cloud Functions (7), Azure Functions and IBM Cloud Functions (both 5) are the technologies most reported on, followed by a long tail of others which raise less interest with researchers.

Fig. 5 shows their distribution across the covered years. Interestingly, one of the early works by Lynn et al. states that AWS is by far the dominating research platform [LRLE17]. While AWS still dominates, the field is now more diverse, although many of the new contenders are research prototypes not offered for commercial service. All commercial offerings are marked with $ in the figure, and Lambda now accounts for slightly more than half of their coverage. Hence, Lynn's statement is still correct for research on public FaaS when considering a relative majority.

New FaaS technologies appear with high frequency and often with widely disseminated public announcements. This includes recent additions such as KNative and Qinling in 2018. This raises the question of relevance for researchers: Apart from clearly technology-independent research, sometimes technical aspects require the focus on one particular implementation. Which one to choose, then?

In Fig. 6 we show how the current focus on technologies in research publications on serverless computing and FaaS-related topics mismatches the apparent needs of developers, based on a systematic developer survey conducted in 2017/2018 [LWSH18]. The brighter bars correspond to all participants, including experienced and prospective serverless developers, whereas the darker bars only include developers who have used serverless offerings in the past; the differences between both are insignificant. AWS Lambda and Microsoft Azure Function dominate in production and hence are currently about 6% and 7% underreported on, respectively. In contrast, Apache OpenWhisk received a lot of attention but is not used a lot in practice, leading to around 12% overreporting. However, the validity of these numbers may be limited by a number of factors, including the recent rebranding of IBM Bluemix OpenWhisk to IBM Cloud Functions, and the inclusion of general cloud providers with strong PaaS but no dedicated FaaS offering such as Heroku and Digital Ocean. Still, we believe that it is valuable to continue tracking the (mis)match over time.

In contrast to the literature dataset, the survey is a one-time snapshot with a decreasing applicability to the evolving field of serverless computing and applications. Therefore, to maintain the insights into the mismatch, a recurring mixed-methods study or at least a recurring survey regarding the technologies will have to be conducted in the future, while some metrics can be derived from recurring industry surveys such as the one from CNCF which in its recent edition mentions among the installable platforms Kubeless with 42%, OpenWhisk with 25% and OpenFaaS with 20% [Bar18].

7 Conclusion

We have assembled an evolvable dataset to track the publicly available research communications on FaaS-related topics [Spi18]. The numerical and visual analysis of this dataset, assisted by associated scripts, gives insight into research actors, topics, trends and mismatches. We invite all experts on FaaS topics to collaboratively maintain future revisions of the dataset which will serve as substantial foundation for future surveys and comparison articles. By increasingly applying data analytics methods, we expect to gain more insight over time as the dataset increases, including a mapping of technology popularity over a multi-year timeframe. Moreover, we envision value-added services exploiting the dataset such as a FaaS solution recommender service to appear as prototype or even as commercial solution on the cloud market.

References

[Bar18] Kaitlyn Barnard. CNCF Survey: Use of Cloud Native Technologies in Production Has Grown Over 200%. online: `https://www.cncf.io/blog/2018/08/29/cncf-survey-use-of-cloud-native\-technologies-in-production-has-grown\-over-200-percent/`, August 2018.

[LRLE17] Theo Lynn, Pierangelo Rosati, Arnaud Lejeune, and Vincent C. Emeakaroha. A preliminary review of enterprise serverless cloud computing (function-as-a-service) platforms. In *IEEE International Conference on Cloud Computing Technology and Science, CloudCom 2017, Hong Kong, December 11-14, 2017*, pages 162–169, 2017.

[LWSH18] Philipp Leitner, Erik Wittern, Josef Spillner, and Waldemar Hummer. A mixed-method empirical study of Function-as-a-Service software development in industrial practice. PeerJ Preprints 6:e27005v1 (`https://doi.org/10.7287/peerj.preprints.27005v1`), June 2018.

[RH11] Florian Reitz and Oliver Hoffmann. Did they notice? - A case-study on the community contribution to data quality in DBLP. In *Research and Advanced Technology for Digital Libraries - International Conference on Theory and Practice of Digital Libraries, TPDL 2011, Berlin, Germany, September 26-28, 2011. Proceedings*, pages 204–215, 2011.

[Spi18] Josef Spillner. Serverless Literature Dataset. Zenodo (10.5281/zenodo.1175423), February 2018.

[VGO+16] Mario Villamizar, Oscar Garces, Lina Ochoa, Harold E. Castro, Lorena Salamanca, Mauricio Verano, Rubby Casallas, Santiago Gil, Carlos Valencia, Angee Zambrano, and Mery Lang. Infrastructure Cost Comparison of Running Web Applications in the Cloud Using AWS Lambda and Monolithic and Microservice Architectures. In *IEEE/ACM 16th International Symposium on Cluster, Cloud and Grid Computing, CCGrid 2016, Cartagena, Colombia, May 16-19, 2016*, pages 179–182, 2016.