On the Nature of Software Sub-Ecosystems and their Health

Yannick Mijsters, Amr Mustafa, Ionut Mihai, Slinger Jansen
Utrecht University
Utrecht
{g.y.mijsters,a.m.a.a.ahmed,i.mihai}@students.uu.nl,slinger.jansen@uu.nl

ABSTRACT

Background. The concept of sub-ecosystems, widely used in natural ecosystem theory, has never been introduced in software ecosystem analysis. It provides a perspective on software ecosystems that can be used to create better understanding of them and more effective ecosystem health analysis. **Objectives.** The objective of this research is to introduce the concept of sub-ecosystems to the field of software ecosystems. An extension on the Open Source Ecosystem Health Operationalization for measuring the health of a sub-ecosystem is created and evaluated with three small case studies. **Method.** A literature review of both software and natural ecosystem research is used for the definitions of key concepts. Design Science is used for the extension of the Open Source Ecosystem Health Operationalization. Finally, for the case studies, data is gathered using several data repositories and analyzed to show how the concept of sub-ecosystems is used. **Results** The concept of software sub-ecosystems is defined. Next to that an extension to the Open Source Ecosystem Health Operationalization (OSEHO) framework is introduced for considering sub-ecosystems in health assessments. **Conclusion** The subject of sub-ecosystems provides a promising new perspective on software ecosystems that improves the understanding of this research field for both researchers and practitioners. Additionally, the extended OSEHO framework can be used to more accurately measure the health of an ecosystem by looking at both larger and smaller ecosystems around it.

KEYWORDS

Sub-ecosystems, OSEHO, Super-ecosystems, Virtual Assistant, Siri, Google Assistant, Amazon Alexa, Ecosystems, Design Science

ACM Reference Format:
Yannick Mijsters, Amr Mustafa, Ionut Mihai, Slinger Jansen. 2018. On the Nature of Software Sub-Ecosystems and their Health. In *SoHeal'18: SoHeal'18:IEEE/ACM 1st International Workshop on Software Health , May 27-June 3 2018, Gothenburg, Sweden.* , 8 pages.

1 INTRODUCTION

An increasing number of organizations opens up their software platforms to third-party developers and enable them to extend their platforms. Software ecosystems provide a new perspective on groups of actors that collaboratively create software around projects and platforms. Software ecosystems are defined as [9]: *"a set of actors functioning as a unit and interacting with a shared market for software and services, together with the relationships among them. These relationships are frequently underpinned by a common technological platform or market and operate through the exchange of information, resources and artifacts."*

One of the most instrumental ways of looking at Software Ecosystems has proven to be the health perspective on software ecosystems [11]. Similar as in biology, the health of an ecosystem indicates how the ecosystem is doing and whether it is growing. The health of a software ecosystem is defined, by Lucassen et al. in the following way: "*longevity and a propensity for growth*"[10]. Nevertheless, several models and frameworks have been introduced on this subject, one of which is the Open Source Ecosystem Health Operationalization (OSEHO) model by Jansen [7] and the business ecosystem health model by Den Hartigh [4], as will be discussed in more detail in the next section.

The concept of sub-ecosystems is widely used in natural ecosystem literature, however, has only been mentioned very briefly in software ecosystem research. In natural ecosystem research, the concept is often used to divide a large ecosystem into several smaller parts that have a common dividing property, for example, different geographic locations[6]. Additionally, in this context, the overall ecosystem, in which the sub-ecosystems are based, is referred to as the super-ecosystem. An example of this concept is the ecosystem around the Google Assistant application, which is inside the larger Android "super-ecosystem". All Google Assistant applications have to be Android applications, however, not all Android applications use Google Assistant, therefore, the Google Assistant ecosystem is a subset, or a sub-ecosystem, of the whole Android ecosystem.

This research focuses on the subject of sub-ecosystems. We define a software sub-ecosystem as **an ecosystem that exist in a super- ecosystem**. A software sub-ecosystem has all the properties of a software ecosystem, however, it exists inside one or more software super-ecosystems. The actors in the software sub-ecosystem are underpinned by a common technological platform or market, different from, and typically related to, that of its super-ecosystem.

The case studies in this research illustrate the sub-ecosystems Virtual Assistants and their super-ecosystems. Since the release of Siri by Apple inc. in 2011, technology companies started to give more attention to AI based assistants. In a few years, the technology improved significantly and several applications have been released competing with Siri, most notably the Google Assistant and Amazon Alexa. Additionally, most of these applications now allow for integration of the technology in third-party applications. This has given a rise to an entire ecosystem of applications around these applications. The case studies discover the health of the ecosystems around these applications by looking not only at the ecosystem

SoHeal'18, May 27-June 3 2018, Gothenburg, Sweden

ACM ISBN 978-1-4503-5730-2/18/05…$15.00
https://doi.org/10.1145/3194124.3194127

itself but also the super-ecosystem using the proposed extension to the OSEHO framework.

In the next section, we introduce a background to this research. In the third section, we describe the research approach and methods used for this research. Thereafter, in the fourth and fifth section, the concepts of sub-ecosystem and sub-ecosystem health respectively are defined and discussed. The sixth section discusses the selection process for the proposed extension of OSEHO. In the seventh section, this proposed extension is explained. The eighth section discusses the case studies executed. Finally, the ninth and tenth sections discuss the results and conclude the findings respectively.

2 BACKGROUND

One of the frameworks used in this paper is the Open Source Ecosystem Health Operationalization (OSEHO) developed by Jansen [7]. This model provides a matrix of indicators to measure the health of open source ecosystems. These metrics are categorized into two layers, namely, the Network Level and Project Level. While the network level consists of metrics relevant for the ecosystem, the project level consists of metrics based on the projects in the ecosystem. On the other dimension of the matrix, three main themes are described, namely, Productivity, Robustness and Niche Creations. Productivity generally refers to how an ecosystem is productive and how projects contribute to the whole ecosystem. Robustness refers to how an ecosystem can handle rapid change in the ecosystem and how fast can it recover. Finally, niche creations refer to the variety of projects, opportunities and room for creating new projects.

Additionally, another model for measuring the health of business ecosystems is the Business Ecosystem Health model by den Hartigh, Visscher, Tol and Salas [4]. While the pillars are the same as OSEHO, den Hartigh et al. has different levels, namely Partner Level and System Level. The approach followed in this paper of evaluating metrics was applied in our research. This research has four main factors to evaluate their metrics three of which inspired our research, Understandability which refer to how easy it is to understand what does the metric imply. Availability refers to the possibility of collecting the data of this metric. Finally, Long Term Usage is a factor where data can be measured and compared at multiple moments in time.

Additionally, while the use of sub-ecosystems in software ecosystems' literature is limited, there is related literature in the field of natural ecosystems. Franco Bedoya et al. discuss the connection and a taxonomy of subjects that connects natural and software ecosystems [3]. Moreover, in natural ecosystem' theory, sub-ecosystems can be used to determine the health of an ecosystem. Xu et al. introduce a metric for determining the health of the ecosystem inside a lake by looking at a number of metrics for sub-ecosystems inside that lake resulting in the Ecosystem Health Index Methodology (EHIM) [13].

3 RESEARCH APPROACH

The method used in this research is Design Science. Hevner defined three main cycles for a design science research, a rigor cycle, a relevancy cycle and design cycle [5]. In the rigor cycle, the literature was reviewed aiming to understand the state of the art in the field of Software Ecosystem health. Besides, the concept of sub-ecosystems in nature was also reviewed to identify any possible overlap or general intersections. The following step, the relevance cycle, was to identify the problem and how this research and the artifact developed are going to influence the environment of applications. At the heart of the Design Science method comes the design cycle. In this core cycle, the solution of the problem is developed and evaluated. In this research, we proposed a model solution to measure the health of sub-ecosystems within super ecosystems based on current models.

Additionally, the new OSEHO is evaluated by applying case studies on the health of the ecosystems of Virtual Assistants and their super-ecosystems. Yin [14] developed a framework of case studies. In this research, multiple case studies were used where a specific set of information is studied in each case. Specifically, the case studies were about the health of the ecosystems around Google Assistant, Siri and Alexa. They were compared to their super-ecosystems, namely Android, iOS Amazon Web Services. The health of the ecosystems was finally compared based on the proposed metrics.

4 SUB-ECOSYSTEM AND SUPER-ECOSYSTEM DEFINITION

The concept of sub-ecosystems and super-ecosystems in the field of software ecosystems has been mentioned in literature, but is not operationalized as it has been in the domain of natural ecosystems [13]. In natural ecosystems, sub-ecosystems are used regularly, for example for the assessment of the health of a large natural ecosystem by looking at the health of geographically divided sub-ecosystems inside that natural super-ecosystem [12]. How natural sub-ecosystems are divided is dependent on the perspective taken by a researcher. As per the example, an ecosystem can be divided geographically, however, different altitudes or other differentiating factors can be used to define natural sub-ecosystems as well.

This definition can be transferred to the concept of Software Ecosystems. Considering the definition of Software Ecosystems, as discussed in Section 1, any set of businesses functioning as a unit interacting in the same market for software and services, together with their relationships can be considered an ecosystem. This would make any subsets of businesses in a software ecosystem a sub-ecosystem. However, the second part of the definition, specifically: *relationships are frequently underpinned by a common technological platform or market and operate through the exchange of information, resources and artifacts*, separates the more interesting sub-ecosystems. These are the sub-ecosystems that have their own common technological platform or market inside a super-ecosystem. Therefore, the following definition for sub-ecosystems is proposed. *A software sub-ecosystem is an ecosystem that exist in a super- ecosystem.* As will be discussed in more detail later, this definition does not require a sub-ecosystem to be fully contained by one super-ecosystem, as sub-ecosystem could branch out to allow businesses inside other super-ecosystems to join them. An interesting question is whether every software ecosystem ever studied is not a sub-ecosystem of the worldwide software ecosystem, i.e., all actors that produce software in relation to other actors. However, although it may be true, the perspective is more relevant for the analysis of smaller ecosystems and their sub-ecosystems.

These sub-ecosystems are identifiable in many software ecosystems. Anytime an application or company inside an ecosystem opens up some technological platform or market, where companies operate in a shared market, a new sub-ecosystem is created. Additionally, as such a sub-ecosystem grows there is the potential that one of the companies in it creates another under-pinning technological platform or market creating another level of sub-ecosystems. In some of the major ecosystems this can go down four levels. Next to that, sub-ecosystems can be as small as any software ecosystem, however, according the original software ecosystem definition [8], they are to be underpinned by a technological platform or market different from that of the super-ecosystem. In theory, three companies working together in a technological platform, different from that of the super-ecosystem they are both part of, have created their own sub-ecosystem. Figure 1 demonstrates the concept of super and sub ecosystems where three main entities are defined. The Keystone represents the orchestrator of super-ecosystem. An actor, in general, is a member of an ecosystem. In the case of sub-ecosystems, an entity can be a *member* of the super-ecosystem and an *orchestrator* of the sub-ecosystem.

Another property of software sub-ecosystems is that they can overlap. In natural ecosystems, a large ecosystem is often divided into sub-ecosystems that add up to the large ecosystem [13]. For example in natural ecosystem health assessments, the health of a super-ecosystem can be measured by calculating a weighted summation of its sub-ecosystems. For this calculation, the sub-ecosystems are often geographically divided and do not overlap [12]. However, in software ecosystems, there is often overlap between sub-ecosystems, with businesses being part of multiple sub-ecosystems. This key difference makes the identification of sub-ecosystems in software ecosystems a challenging task.

Finally, software sub-ecosystems are not always fully contained by one super-ecosystem. When an ecosystem or sub-ecosystem grows large enough it is possible to branch out to other super-ecosystems. When this happens, however, it would directly interfere with the idea that a sub-ecosystem is only part of one larger super-ecosystem. Nevertheless, this type of ecosystem is not on the same level as its original super-ecosystem as the key technological platform or market and its partners are still underpinned by at least one of the super-ecosystems. These sub-ecosystems that have branched out to other ecosystems, or at least outside of their original super-ecosystem we will now refer to as *intersecting sub-ecosystems*, as similar to set theory they intersect several larger ecosystems. Figure 2 exemplifies such an intersecting sub-ecosystem. Additionally, sub-ecosystems that are purely contained by the super-ecosystem will be referred to as *proper sub-ecosystems*, as, similar to the definition in set theory, they are a subset of the super-ecosystem that is not equal to the super-ecosystem in its entirety. Figure 1 exemplifies a sub-ecosystem.

5 SUB-ECOSYSTEM AND SUPER-ECOSYSTEM HEALTH

As discussed in the introduction, a regularly used metric in research into software ecosystems, for both practitioners and researchers, is the health of an ecosystem. In literature several models and frameworks for identifying the health of software ecosystems have been

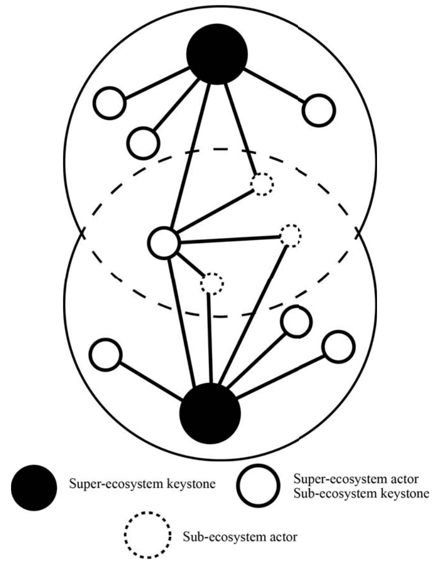

Figure 1: Super/Sub Ecosystems

proposed, specifically the OSEHO framework, as introduced by Jansen [7] and the business ecosystem health model as introduced by Den Hartigh will be used in extensively in this research[4]. However, none of these have considered sub-ecosystems in the creation of these frameworks. For an initial review of the health of sub-ecosystems, the existing models for software ecosystem health can be used for a good indication. They can quite accurately quantify the health and they can help indicate certain weaknesses in the ecosystems that either the keystones or third parties can work on. However, the health of a sub-ecosystem is not only influenced by itself and its current status. As it is part of one or more larger super-ecosystems, it is influenced by changes in the super-ecosystem as well. This can both be in a positive and negative way. For example, a sudden extension of the larger ecosystem to another market can bring many potential users and developers to the sub-ecosystem. However, a sudden change in policy in the super-ecosystem can negatively impact the sub-ecosystem and can hurt its current users or developers. Additionally, the health of sub-ecosystems can impact the super-ecosystem, as sub-ecosystems influence many projects in the super-ecosystem. Therefore, a more specific health measurement model for ecosystem can significantly increase the accuracy of such an assessment. Additionally, it can help companies considering to start a sub-ecosystem decide which super-ecosystems are more open to sub-ecosystems before they start building a technological platform.

Receptiveness - The first factor of the super-ecosystem that impacts the health of the sub-ecosystem is the **receptiveness** of the super-ecosystem for sub-ecosystems. The receptiveness of a super-ecosystem can be defined as the behavior of the keystone towards the creation and growth of other ecosystems inside its ecosystem. This category has much similarity with the openness of an ecosystem as discussed, for example in this research by Anvaari and Jansen [1], however, they can also create certain limitations, for example, by prohibiting any third-party app stores in an ecosystem.

An example of how an keystone can promote sub-ecosystems is by showcasing them actively in their marketplace. The receptiveness of a super-ecosystem can also impact the health of a super-ecosystem itself, as a sub-ecosystem can provide more variety and more projects to an ecosystem they often are beneficial for the health of super-ecosystems. Additionally, super-ecosystems that lack receptiveness open themselves up to competing ecosystems that do support sub-ecosystems.

Additionally, there are several indirect ways in which a company can impact their receptiveness to sub-ecosystems. The reputation of a company can have an influence on companies that potentially could start a sub-ecosystem [2]. If a company is known for cracking down on sub-ecosystems as they grow to a certain size then this will reduce the number of sub-ecosystem. Even one incident can impact the other existing sub-ecosystems, as they may lose trust. This can also happen when the keystone copies ideas of sub-ecosystems in their own platform or start their own competing service. However, a good reputation can also invite more sub-ecosystems for example when the keystone has the tendency to reward larger sub-ecosystems for their contribution to the ecosystem, either monetary or with attention. They may, for example, even acquire keystones of sub-ecosystems when they become popular, to secure their place in the ecosystem.

Relative size - Relative size is another factor that impacts interaction in software ecosystems. The relative size of the two ecosystems considers the difference in size between the super- and sub-ecosystem. The impact of the differences in size can result in different outcomes, discussed in two scenarios. The first scenario is when the difference in size between the two ecosystems is large. Therefore, the relative size of the super-ecosystem is large and of the sub-ecosystem is small. In this case, there is a relatively small impact that the ecosystems have on each other in terms of health. The sub-ecosystem is, at this point, too small to have a serious impact on the health of the super-ecosystem. Additionally, changes in the health of the super-ecosystem will not directly impact the sub-ecosystem as the pool of potential developers or users in the super-ecosystem is large enough that even significant changes in this pool will not directly be noticed by a growing sub-ecosystem. The second scenario is when the difference in size between the two ecosystems is relatively small. In this scenario the ecosystems can impact each other's health heavily. First of all, the sub-ecosystem has a significant percentage of users of the super-ecosystem, making it that any change to its health can directly impact the super-ecosystem. A negative change in the health of the sub-ecosystem may cause a significant number of users to leave the super-ecosystem as a whole, while an increase in the health of the sub-ecosystem will likely draw in new developers or users from outside the super-ecosystem. Finally, if the health of the super-ecosystem changes then this will directly impact the health of the sub-ecosystem in similar ways.

Super-ecosystem properties - There are many properties for the health of the super-ecosystem that will impact the health of the sub-ecosystem too. An example of this is the documentation of a super-ecosystem that is built around a technological platform [7]. When this documentation improves, it not only positively impacts the health of the super-ecosystem, but also the health of the sub-ecosystem, as this helps the developers in the sub-ecosystem as well. In contrast to the previously discussed categories, however, for many of these factors, the effect on the health is only "one way". A similar change to a property of the sub-ecosystem will often not directly impact the health of the super-ecosystem. In the previously mentioned example, if it was the sub-ecosystem improving the documentation, then only developers in the sub-ecosystem are affected by this improvement and this does not directly impact the health of the super-ecosystem, or only very slightly. Additionally, a property change that positively affects the health of the super-ecosystem may not always affect the sub-ecosystem in the same way. An introduction of a new functionality to the super-ecosystem that is also a key function in the sub-ecosystem may increase the health of the super-ecosystem, but it will hurt the health of the sub-ecosystem. Many of the factors in a super-ecosystem that impact the health of a sub-ecosystem will be discussed in section 7 in the proposed SEHO framework. Finally, the effect of these three categories is not independent of each other. For example, changes to properties of the super-ecosystem, as discussed in the third category, may have a different effect on sub-ecosystems of different sizes.

6 CRITERIA AND SELECTION FOR OPERATIONALIZATION

The goal of this research is to introduce an operationalization of metrics that impact the health of sub-ecosystems and that indicate how sub-ecosystems influence the super-ecosystems. As discussed in the previous section, there are many factors that impact the health of a sub-ecosystem and the interaction between these factors depends on other variables. In this framework, these effects are operationalized for an overall assessment of the health of either sub- or super-ecosystems.

The first step towards extending the OSEHO framework consists of creating an inventory of metrics that can potentially affect the health of a sub-ecosystem. As discussed, the metrics from the OSEHO and Business Ecosystem health model are selected as a starting point. First of all, they are selected as metrics applicable on either the level of a sub-ecosystem or on the network level in the original framework. As no individual projects are actually researched on the sub-ecosystem level, no new metrics were considered for this layer. Thereafter, for some metrics, basic translations are made to the perspective of sub-ecosystems instead of super-ecosystems. Examples of this are the changes from *added knowledge about the ecosystem* to *added knowledge about the sub-ecosystem* and *new related projects* to *new related sub-ecosystems*. The impact of these factors on the health of the sub-ecosystem is similar to their effect on the health of a software ecosystem in the original OSEHO framework [7].

Finally, all remaining indicators are judged according to criteria specified by Jansen [7] during the creation of the OSEHO framework. Namely that the metric should *contribute positively to at least sub-ecosystem health*. Each of the metrics has to increase the 'longevity and propensity for growth' of at least the sub-ecosystem and preferably the super-ecosystem. As in some rare cases a positive change to the health of a sub-ecosystem can negatively impact the super-ecosystem, the focus on sub-ecosystem health is considered for the sub-ecosystem layer. Secondly, the metric should be *operationalizable into a measurable entity*. This is relevant for users of the framework, as they have to be able to gather the data for the selected metrics. The final criterion is that metrics should be *generalizable to multiple projects and sub-ecosystems.* For a measurement of the health of a super-ecosystem, all sub-ecosystems and projects in the super-ecosystem should be considered.

7 THE EXTENDED OSEHO

The new metrics in the model is discussed. Further explanations about the existing metrics can be found in the original work by Jansen [7]. Table 1 shows the metrics included in the OSEHO model along with the new metrics represented in gray which are added to measure the sub-ecosystem health. Additionally, as many of the metrics are based on metrics in the original OSEHO framework paper, a description of how these metrics can be measured can be found, either directly explained or referenced, in that work. However, explanations of metrics that are incomplete for the extended framework are updated in this section.

Next to the metrics added to the new layer, several metrics have been added to the original network layer. These metrics can be considered both when research into a super-ecosystem as well as into a sub-ecosystem is done. For super-ecosystem health, they give an indication of the state of a super-ecosystem, as sub-ecosystems can be beneficial to the health of the super-ecosystem, for example by bringing more connectivity. Additionally, as discussed, sub-ecosystem health can be impacted significantly by the super-ecosystem, therefore, the proposed metrics should be considered when researching the health of a sub-ecosystem.

Network layer metrics - On the network level, there are several indicators added. First of all, in the category of productivity, there is the *new sub-ecosystems* metric. This metric is an indicator for both super- and sub-ecosystem health. For super-ecosystems, this metric is an indicator of how productive the ecosystem is in terms of creating new sub-ecosystems, which can help grow a super-ecosystem. For sub-ecosystem health, this is an indicator of how receptive a super-ecosystem is to sub-ecosystems, if this number is very small for a particular super-ecosystem, then they are likely not receptive to sub-ecosystems and that will make it hard to create a healthy sub-ecosystem in that super-ecosystem. This metric can be measured by identifying the sub-ecosystems in the super-ecosystem and see how many have been created in the past time period. This metric is similar to the original *new related projects* metric [7].

In the productivity column, three metrics are added. The first metrics is *number of intersecting sub-ecosystems*, this is another indicator for the connectedness to other super-ecosystems similar to the number of outbound links to other SECO. Similar as for that metric, intersecting ecosystems, which, as discussed in section 5, are sub-ecosystems that exist in multiple super-ecosystems, can better withstand negative events in one super-ecosystem because they can get revenue or users from other super-ecosystems [7]. This can also help the effected super-ecosystem, as these sub-ecosystems are a more stable factor for that super-ecosystem to rely on. This metric can be measured through an analysis of the sub-ecosystems and the super-ecosystems they are available in. Additionally, there is *average size of sub-ecosystems* and *number of active sub-ecosystems*, which for sub-ecosystem health further indicate the receptiveness of the ecosystem to new sub-ecosystems. The total number gives an overview of the receptiveness and the average size is an indicator of how well sub-ecosystems can grow inside the super-ecosystem and if the super-ecosystem is receptive to larger sub-ecosystems. Moreover, the super-ecosystem health is positively affected by larger sub-ecosystems, as they are likely to bring more users from outside to the super-ecosystem.

Thereafter, in the niche creation column, the metric for *variety in sub-ecosystems* is added. This metric is very similar to the metric *variety in projects* as introduced in the OSEHO framework [7], however, this variety metric can help differentiate between types of projects more easily, as sub-ecosystems could be built around one specific type of project for the variety in project metric. Additionally, this metric can show the ability to create niche sub-ecosystems inside a super-ecosystem, which impacts the health of sub-ecosystems as well.

Sub-ecosystem level metrics - On the sub-ecosystem level, indicators for single sub-ecosystems that impact the health of an ecosystem are identified. These metrics both impact the health of the sub-ecosystem as well as the health of the super-ecosystem.

In the productivity column there are the *new related projects* and the *added knowledge about the sub-ecosystem.* Both metrics are inspired by similar metrics in the network level [7], however, they are now applied on the sub-ecosystem. They are, therefore, also relevant to the health of the sub-ecosystem in the same way they are relevant to the health of the super-ecosystem on the network level. These indicators for the productivity in the sub-ecosystem also impact the health of the super-ecosystem, because productive sub-ecosystems bring more projects to the super-ecosystem as well.

In the robustness column, there is *total number of active projects.* This metric, similar to the metrics in the productivity column, is an indicator that is also present on the network level and is applicable on the sub-ecosystem in the same way the network layer metric is applicable to the super-ecosystem. Moreover, there is the *number of outbound links to other sub-ecosystems* indicator. From the perspective of the super-ecosystem this is a metric similar to the *project connectedness* metric as this indicates how connected the different sub-ecosystems are [7]. From the perspective of the sub-ecosystem, this metric is similar to the *outbound links to other SECO* in the network level and impacts the health of similar reasons.

Finally, in niche creation *variety in projects* metric is present. This metric has a similar effect on the sub-ecosystem, as the same metric has on the super-ecosystem [7]. Additionally, a more varied sub-ecosystem will also create more variety in the super-ecosystem impacting the health of the super-ecosystem in a similarly positive way.

Some of the newly proposed metrics create an overlap in the metrics present in the framework, especially on the network level. The *total number of active sub-ecosystems* together with the *average size of sub-ecosystems* give a result similar to the total number of active projects. Additionally, the *number of intersecting sub-ecosystems* can be quite similar to the number of *outbound links to other SECO* [7]. However, for both metrics it should be considered that there still may be many projects that are not in any sub-ecosystem and these would not be considered in the sub-ecosystem specific metrics. Next to that, a comparison between these similar metrics can give an indication of how widely sub-ecosystems are used by projects in the super-ecosystem, which can be another good indicator of

Table 1: The Extended OSEHO, based on Jansen (2014)

Level	Productivity	Robustness	Niche Creation
Network Level		Total number of active projects	
	New related projects	Project connectedness/cohesion	
	Downloads of new projects	Core network consistency	
	Added knowledge about ecosystem	Outbound links to other SECOs	Variety in projects
	Events	Switching costs to other SECOs	
		Number of intersecting sub-ecosystems	
	New sub-ecosystems	Total number of active sub-ecosystems	Variety in sub-ecosystems
Sub-Ecosystem Level	New related projects	Total number of active projects	Variety in projects
	Added knowledge about sub-ecosystem	Number of outbound links to other sub-Ecosystems	
Project Level		Partnerships and embeddedness	
	KLOC/time period added	Organizational maturity	
	New tickets	Commercial patronage	
	New downloads	Capital contributions and donations	Variation in contributor type
	Knowledge and artifact creation	Contributor satisfaction	Variation in project applications
	Mailing list responsiveness	Active contributors	Supported natural languages
	Bug-fix time	Contributor ratings and reputation	Variety in supported technologies
	Spin-offs and forks	Multi-homers	Variety in development technologies
	New partnerships	Contributor connectedness	Multiple markets
	New patents	Interest: Page views, search statistics	
	Usage	Market share	
		Switching costs to alternatives	
		User loyalty and usage	
		User satisfaction or ratings	
		Artifact quality	

sub-ecosystem health. For these reasons, no original metrics were removed.

For the application of this framework by either researchers or practitioners, it can be useful to consider the *OSEHO Analysis Method* as discussed by Jansen [7]. Additionally, when applying the metric for an inquiry into a specific sub-ecosystem, the metrics in the super-ecosystem that impact the health of the sub-ecosystem should be considered, as discussed in this section.

8 CASE STUDIES

Three main applications are examined in the case studies: Apple's Siri, Amazon's Alexa and Google's Assistant. These applications can perform tasks on mobile voice based on voice commands, not requiring any physical interaction between a device and its user. All three applications allow third party developers to create applications that can be interacted with through these applications. Additionally, all three of these companies have released stand-alone hardware for these applications, however, the focus of this case study is on the mobile phone integration of the application.

Data Collection - A mixture of manual and public API data mining was used. For the public API data mining, a python library was created. The library was used in combination with a Jupyther Notebook to analyze the data. The library and the data is public on Github and an access link is provided[1].The Github project can be used to evaluate the data gathered, draw conclusions from it and extend it in future related studies.

The APIs that were used to retrieve data are the iTunes search API and StackExchange API. The iTunes search API is provided by Apple for those who want to integrate iTunes search abilities in their applications or websites. For each voice assistant, we use a keyword that defines it like "Alexa","Google Assistant" and "Siri". The iTunes API returns the data on the applications that have the keyword in their title or description. The StackExchange API was used to retrieve data on the questions asked and answered on the website StackOverflow through quires tag. In the case of manual data mining certain websites that provide data about the metrics were accessed and the information there was analyzed and noted down.

Not all the metrics were analyzed due to several reasons. The main reason was the time constraint. Since software ecosystem health it is at an incipient phase, there are no automatic tools or libraries to gather data about health metrics. OSEHO is meant to be used for open source software (OSS). The case studies subjects aren't OSS products, as such data on several metrics isn't available.

[1] https://github.com/Mihai-Ionut-Aurel/VAHAF

Table 2: Case Studies Results

Super-ecosystem Level	Android	IOS	Amazon
Added knowledge about ecosystem	>3000	>3000	1950
Total number of active projects	2.800.000	2.200.000	-
Interest	78	79	13
New patents	1134	231	7
Supported Natural Languages	693	42	10
Variety in supported technologies	3	3	>
Variety in development technologies	8	2	8
Variety in sub-ecosystems	5	4	19
Sub-ecosystem Level	Google Assistant	Siri	Alexa
Added knowledge about ecosystem	47	5	78
Total number of active projects	2590	557	25000
Variety in projects	18	22	23
New Projects	47	12	>10000

The data for the case studies was gathered from the same source were possible. For several metrics, the data was gathered from similar sources. The sources selection was done based on their popularity. By selecting similar and popular data sources, the results are comparable and insights can be drawn from them.

Case Studies Results - The results of the case studies can be found in table 2. For the *Added knowledge about the ecosystem* metric, data on the questions on Stack Overflow in the last two months of 2017 were compared. For questions to be considered, it was necessary for them to have at least one answer. The data gathering API has a limit on the number of results it returns. Therefore, the Android and IOS results with the sign > mean that the questions count is greater than 3000. It can be deduced that, while Android and iOS heavily outweigh Amazon in this aspect, on the sub-ecosystem this seems to be the other way around as most interest seems to be centered around Alexa. Amazon can support almost any technology because it is designed to be used by Internet of Things products. For the *total number of active projects* on the sub-ecosystem level, the number of applications available to users was considered. The defined framework was used to acquire data on the number of applications that mention Siri in their title or description. Google Assistant has its own store now where a total of 2564 applications where identified. For Alexa, a skill was considered as an active project, of which 25000 are available. A number of Google Assistant and Alexa applications were found on the iTunes store and were taken into consideration. In terms of the *variety of projects*, the applications categories were taken into account. At the sub-ecosystem level, we can find a similar variety in all of them. This is because the applications stores categories are common.

Amazon has a high variety in sub-ecosystems while Android and iOS have a significantly lower variety. While Google Assistant and Siri have fewer competitors, they do not have better results than Alexa. As such, it seems that the variety in sub-ecosystems does not necessarily correlate in unhealthy sub-ecosystems.

In conclusion, the differences between iOS and Android are limited and, as Amazon has another customer type these results are hard to compare. However, on the sub-ecosystem level the differences are significant. Alexa seems to have attracted more interest than Google Assistant and Siri, especially by developers. Siri seems to perform the worst in most aspects, however, it should be considered that Siri only relatively recently opened its ecosystem up to third-party developers, as it was for a long time only available for hand-picked partners. These results could be considered surprising, as traditionally Google and Apple are more technology-based companies, where Amazon is more an eCommerce based company, however, they seem to have outperformed the tech companies in this aspect.

9 DISCUSSION

The goal of this research was to extend the OSEHO framework with a health measurement operationalization for sub-ecosystems and their effect on the health of super-ecosystems. This has resulted in the extended OSEHO framework. The framework has been evaluated with case studies into the virtual assistants while taking into account their respective main super-ecosystem.

Results showed how a sub-ecosystem can affect the health of the super-ecosystem from which it started on. In the case of Google Assistant and Alexa, the sub-ecosystems have brought a number of new applications and new partners for their super-ecosystem. They even branched out of their starting super-ecosystem into other super-ecosystems. New sub-ecosystems can revitalize a super-ecosystem health, keep the interest in it and bring new innovations to it. These are factors for the long run of a software ecosystem health as technology evolves at a faster pace.

Because a sub-ecosystem can branch out of their super-ecosystem it can be stated that at some point they have a better health than the super-ecosystem that they started from. This is possible as they tap in the resources of several super-ecosystems. With a bigger pool of resources, the sub-ecosystem health and growth is accelerated. For example, Google Assistant has branched out of the Android super-ecosystem and has its own App Store. Because of this, its pool of users is drawn not only from people who depend on the Android ecosystem but also on those from the iOS ecosystem.

Finally, the suggested framework requires validation. The evaluation executed in this case study is the first step towards a validated framework. However, as it is only applied to one type of ecosystem, more research is needed for the generalization of the framework. In future research, the OSEHO should be taken as a starting point for the evaluation of a sub- or super-ecosystem and not only validating the framework but also creating additional metrics or improving the interpretation of the framework can be relevant contributions to the framework.

10 CONCLUSION

In this research, we have introduced the concept of sub-ecosystems and super-ecosystems to the field of software ecosystems. A definition has been formed for software sub-ecosystems. The concept is compared with natural sub-ecosystems and several similarities and differences are identified. This concept introduces a new way of looking at a software ecosystem by not only considering the projects in the ecosystem or at the complete ecosystem, but also by considering the smaller ecosystems with a common boundary as an important part of the super-ecosystem. Two types of sub-ecosystems are identified, namely proper sub-ecosystems and intersecting sub-ecosystems.

To measure the health of sub-ecosystems, three factors were identified that impact the health of sub-ecosystems, next to the factors that in earlier research into the health of software ecosystems have been identified. These factors are the receptiveness of the super-ecosystem, the relative size of the sub- and super-ecosystem and the properties of the super-ecosystem. Thereafter, the extension of the OSEHO framework, considering the health of sub-ecosystems, was introduced. This framework was created by considering the metrics used to identify the health of software ecosystems in current literature and evaluating their applicability to the problem of measuring sub-ecosystem health and considering their impact on the super-ecosystem. The resulting change to the OSEHO consists of an additional layer on the sub-ecosystem level and several added metrics on the network level.

Then case studies were executed to test and begin the validation process of the extended OSEHO framework. The case studies were done on the sub-ecosystems around virtual assistants, which exist in the layer software ecosystems of mobile phone operating systems. Data was collected on the ecosystems around Google Assistant, Siri and Amazon Alexa and their super-ecosystem Android, iOS and Amazon Web Services respectively. The framework was used to create a representation of the health of these ecosystems. These case studies showed several limitations to the created model. Most notably are limitations with regards to interpretation of the results, as the complete overview of the ecosystem is hard to summarize in a concrete result.

Future Research and Impact on Practitioners - The subject of sub-ecosystems can be further defined. In this research several interesting concepts in the domain of sub-ecosystem have been discussed, however, not all of them are actively researched in this work. Examples of a research topic on this could be a comparison between proper and intersecting ecosystems and what the key differences are in terms of operating or joining such an ecosystem as a company. Additionally, this new perspective on software ecosystems can be used to create new ways of modeling or evaluating ecosystems. For example, by researching the power dynamic between sub-ecosystems and between sub-ecosystems and their super-ecosystem new types of power models can be created that can give further inside into the workings of software ecosystems.

Additionally, the subject of the health of both sub- and super-ecosystems can be researched more. Some possible interactions between the two are introduced in this research. However, there may be more ways the two influence each other and the effect of the already identified interactions can be researched in more detail. A possible subject could be looking into how the health of both super- and sub-ecosystems are impacted by their relative size.

The theory of sub-ecosystems allow keystones in ecosystems to look at their ecosystems from a different perspective. Software ecosystems are about cooperation and joined innovation and sub-ecosystems allow different for closer collaboration between related actors in ecosystems. Therefore, the perspective of sub-ecosystems can be used to look at the level of cooperation in ecosystems. Additionally, the view of sub-ecosystems can be used by actors and potential future actors in an ecosystem to identify all opportunities in the ecosystem. Sub-ecosystems can be used to identify the health of an ecosystem and can be especially useful to identify the niches in an ecosystem. This information can be used by actors to identify any business opportunities inside an ecosystem.

REFERENCES

[1] Mohsen Anvaari and Slinger Jansen. 2010. Evaluating architectural openness in mobile software platforms. In *Proceedings of the Fourth European Conference on Software Architecture: Companion Volume*. ACM, 85–92.

[2] Africa Ariño, José De la Torre, and Peter Smith Ring. 2001. Relational quality: managing trust in corporate alliances. *California management review* 44, 1 (2001), 109–131.

[3] Oscar Franco-Bedoya, David Ameller, Dolors Costal, and Xavier Franch. 2017. Open source software ecosystems: A Systematic mapping. *Information and Software Technology* 91 (2017), 160 – 185. https://doi.org/10.1016/j.infsof.2017.07.007

[4] Erik den Hartigh, Wouter Visscher, Michiel Tol, and Adolfo Jiménez Salas. 2013. Measuring the health of a business ecosystem. In *Software Ecosystems: Analyzing and Managing Business Networks in the Software Industry*. Edward Elgar Publishing Ltd.

[5] Alan R Hevner. 2007. A three cycle view of design science research. *Scandinavian journal of information systems* 19, 2 (2007), 4.

[6] Thomas Hoepner. 1999. A procedure for environmental impact assessments (EIA) for seawater desalination plants. *Desalination* 124, 1 (1999), 1 – 12. https://doi.org/10.1016/S0011-9164(99)00083-1 European Conference on Desalination and the Environment.

[7] Slinger Jansen. 2014. Measuring the health of open source software ecosystems: Beyond the scope of project health. *Information and Software Technology* 56, 11 (2014), 1508 – 1519. https://doi.org/10.1016/j.infsof.2014.04.006 Special issue on Software Ecosystems.

[8] Slinger Jansen, Sjaak Brinkkemper, and Anthony Finkelstein. 2009. Business Network Management as a Survival Strategy: A Tale of Two Software Ecosystems. In *Proceedings of the First International Workshop on Software Ecosystems (IWSECO-2009), co-located with the 11th International Conference on Software Reuse , Falls Church, Virginia, USA, September 27, 2009.* http://ceur-ws.org/Vol-505/iwseco09-5JansenBrinkkemperFinkelstein.pdf

[9] S. Jansen, A. Finkelstein, and S. Brinkkemper. 2009. A sense of community: A research agenda for software ecosystems. In *2009 31st International Conference on Software Engineering - Companion Volume*. 187–190. https://doi.org/10.1109/ICSE-COMPANION.2009.5070978

[10] Garm Lucassen, Kevin Van Rooij, and Slinger Jansen. 2013. Ecosystem Health of Cloud PaaS Providers. 150 (06 2013).

[11] Konstantinos Manikas and Klaus Marius Hansen. 2013. Software ecosystems: A systematic literature review. *Journal of Systems and Software* 86, 5 (2013), 1294 – 1306. https://doi.org/10.1016/j.jss.2012.12.026

[12] An-Ning Suo, You-Cai Xiong, Tian-Ming Wang, Dong-Xia Yue, and Jian-Ping Ge. 2008. Ecosystem Health Assessment of the Jinghe River Watershed on the Huangtu Plateau. *EcoHealth* 5, 2 (01 Jun 2008), 127–136. https://doi.org/10.1007/s10393-008-0167-z

[13] Fu-Liu Xu, Zhen-Yan Zhao, Wei Zhan, Shan-Shan Zhao, R.W. Dawson, and Shu Tao. 2005. An ecosystem health index methodology (EHIM) for lake ecosystem health assessment. *Ecological Modelling* 188, 2 (2005), 327 – 339. https://doi.org/10.1016/j.ecolmodel.2005.01.058

[14] Robert K Yin. 2011. *Applications of case study research*. Sage.