Association for Information Systems

AIS Electronic Library (AISeL)

Wirtschaftsinformatik 2021 Proceedings

Track 9: Data Science & Business Analytics

Knowledge Sharing in Digital Platform Ecosystems – A Textual Analysis of SAP's Developer Community

Martin Kauschinger
Technische Universität München

Maximilian Schreieck
Technische Universität München

Markus Böhm
Technische Universität München

Helmut Krcmar
Technische Universität München

Follow this and additional works at: https://aisel.aisnet.org/wi2021

Knowledge Sharing in Digital Platform Ecosystems – A Textual Analysis of SAP's Developer Community

Martin Kauschinger[1], Maximilian Schreieck[1], Markus Boehm[1], and Helmut Krcmar[1]

[1] Technical University of Munich, Department of Informatics, Munich, Germany
{martin.kauschinger,maximilian.schreieck,markus.boehm,helmut.krcmar}@tum.de

Abstract: Research on digital platform ecosystems is growing rapidly. While the relevance of third-party applications is commonly known, scholars have made only minor attempts to analyze knowledge sharing between platform owners and third-party developers. We find that third-party application development is a knowledge intensive task that requires knowledge to cross organizational boundaries. In this paper, we use computational analytic methods to analyze knowledge sharing in a digital platform ecosystem. We collected trace data about a third-party developer ecosystem with frequent knowledge exchange between the platform owner and third-party developers. We developed a web scraper and retrieved all 4866 pages of SAP's developer community that were tagged 'SAP Cloud Platform'. Next, we used text mining to render a topic model. Based on the latent dirichlet allocation algorithm, we extracted 25 topics that were frequently discussed in the community. We clustered the topics into the following six meta-topics: User Accounts and Authentication, Connectivity, Cloud Database, Specific Technologies, SAP Resources, and Installation. Platform owners can use our approach to (1) identify frequently discussed topics, (2) generate meta-knowledge in these topics and (3) use the meta-knowledge to improve their platform core and its boundary resources.

Keywords: Platform Ecosystem, Enterprise Software, Knowledge Sharing, Application Development, Text Mining

1   Introduction

Digital platforms have risen to such prominence in the global economy that they have stimulated a rapidly growing body of scholarly research [1, 2]. By deriving their economic power from the conglomerate of external actors, digital platforms outperform traditional companies across various industries. For example, apple revolutionized the mobile operating systems market by opening application development for third-parties. So far, digital platforms have not only outperformed traditional companies, they have also disrupted several industries by changing the provisioning and consummation of digital services [3]. Economic indicators reveal a similar picture about the economic power of digital platforms. In 2020, according to market capitalization, seven of the top ten public companies used platform business models [4].

On an abstract level, platforms describe the notion of providing a technological system that acts as a foundation upon which other firms can develop complementary products, technologies or services [5]. Tiwana, Konsynski and Bush [6, p.675] adapted the notion of platforms to the software context and define a digital platform 'as the extensible codebase of a software-based system that provides core functionality shared by the modules that interoperate with it and the interfaces through which they interoperate'.

As platforms bring together a variety of actors, they depend on so-called network effects [7]. These effects are best described as the increase of utility that a user derives from the consumption of a good or service for every other person consuming the same good or service [8]. In other words, network effects imply that a technology's usefulness increases as the number of user increases [7]. This coherency is often illustrated by using the telephone as an example. The first telephone did not have any value for its owner because other callable telephones did not exist. However, this changes as soon as other telephones enter the network. Then, the value increases for every new telephone [9]. As of now, researchers produced a plethora of scholarly articles that stress the importance of network effects for curating platform ecosystems [e.g. 10, 11]. For example, a big proportion of Amazon's retailing success originates from the large network of independent retailers that sell their products on Amazon's marketplace. Another example are mobile operating systems. The large variety of apps being available on Android's Playstore or Apple's Appstore were significant drivers for their success. On the contrary, Microsoft's Windows Phone failed miserably due to missing third-party applications.

We adopt the view of Hein et al. [3] that 'a digital platform ecosystem comprises a platform owner that implements governance mechanisms to facilitate value-creating mechanisms on a digital platform between the platform owner and an ecosystem of autonomous complementors and consumers'. Besides the widespread success of digital platforms in consumer markets, more and more companies adopt platform-based business models in business-to-business markets. Thereby, the enterprise software market is no exception. In recent years, traditional ERP vendors such as Oracle and SAP have gotten into fierce competition with cloud-native companies such as Salesforce or ServiceNow. While the latter pursued a cloud platform strategy from their beginning, the former transition from on-premises system to cloud-based solutions. Generally speaking, the concept of app stores being implemented in enterprise software platforms (e.g. the SAP Cloud Platform or the Now Platform) is very similar to the ones that are known from mobile operating systems. The central element is the provisioning of a base system that comes with a marketplace that can be used to install and deploy new applications. Furthermore, the majority of applications are developed and maintained by third-parties [12, 13].

The widespread uptake of platform strategies creates several challenges for vendors of enterprise software [14-16]. On the one hand, vendors who previously sold on-premises systems have to cope with the increasingly complex information systems landscape of their customers. Nowadays, many enterprises use a mixture of on-premises and cloud solutions that result in difficulties with respect to technical integration [17]. Usually, the on-premises systems have gone through a long series of

update cycles and are inherently associated with legacy issues. Such legacy systems require additional integration tools to be compatible with modern cloud solutions. On the other hand, the uptake of platform strategies turns once product-based vendors of enterprise software into ecosystem curators. This shift requires platform owners to collaboratively develop and commercialize a shared technology with customers, consultants and third-party developers [14]. Consequently, those vendors have to shift their focus from product development to governing partnerships and complementary products [18-20].

To enable the development of third-party applications, platform owners have to share development related knowledge with third-party developers. As of yet, the scholarly discourse on knowledge sharing between platform owners and third-party developers was limited to the concept of boundary resources. Prior research identified three types of boundary resources: Software development kits (SDKs), application programming interfaces (APIs), and technical documentation [3, 19, 21]. These boundary resources ease third-party development by providing information about the platform's functionalities [21, 22]. We identified that platform owners use several additional resources to share knowledge with third-party developers. These resources consist but are not limited to tutorials, code snippets, online communities, trainings, and blogs [14].

In this paper, we study the role of sponsored online communities for knowledge sharing in digital platform ecosystems. In particular, we investigate how platform owners can use online community data to generate insights into their platform ecosystem. We find that this relatively unexplored area is worth investigating for several reasons. First, empirical evidence suggests that developers get a vast amount of knowledge from online communities [23, 24]. In that regard, third-party application development is no different. Second, we explored online communities of leading enterprise software vendors and discovered that these communities accumulated a vast amount of peer reviewed knowledge. In fact, Oracle's developer community consists of 3.7 million users, 2.2 million discussions and 7.8 million comments [25]. Salesforce's developer community features 264,000 discussions, without considering Salesforce's Trailblazer community or questions asked on StackExchange [26]. SAP's online community comprises 2.5 million questions of which 1.0 million have been answered. Additionally, the community has 2.8 million users and 123,000 blog posts [27]. Third, we argue that online communities have decent scaling potentials for platform owners. In such communities, third-party developers can share knowledge among one another with minimum moderation effort required by the platform owner.

For our study, we retrieved all 4866 pages of SAP's developer community that were tagged 'SAP Cloud Platform'. Next, we used text mining and rendered a topic model [28, 29]. Based on the latent dirichlet allocation algorithm [30], the model extracted 25 coherent topics that we clustered into the following six meta-topics: User Accounts and Authentication, Connectivity, Cloud Database, Specific Technologies, SAP Resources, and Installation.

With our findings we contribute to the discourse on digital platform ecosystem by providing researchers and practitioners with an analytic lens to study knowledge sharing between platform owners and third-party developers. Furthermore, we help

platform owners to generate insights into their platform ecosystem by analyzing digital trace data. Platform owners can use these insights to improve the platform's attractiveness for third-party developers [22].

The remainder of the paper is structured as follows. In the next chapter, we clarify the theoretical background of the paper. This includes concepts such as modularization, boundary resources, knowledge sharing, and online communities. We conclude this chapter with our research question. In the third chapter, we report our study design. We also cover our case company, our dataset, and our research model and its parameters. In the fourth chapter, we report our results before interpreting them in chapter five. Finally, we summarize the contribution of our paper before we point out limitations and avenues for future research.

2   Theoretical Background

The software industry is currently experiencing several changes that go hand in hand with a concept called software 'platformization'. This concept characterizes the process in which a platform owner creates access and interaction opportunities around the platform core [31]. Besides, software is getting more and more embedded into our daily personal and professional routines [9]. This embeddedness requires software systems which can execute services in a flexible and independent way. The majority of software vendors coped with these changes by transforming their former monolithic architectures into modular designs [16]. Baldwin and Clark [32, p. 1117] describe a system as modular 'if its parts can be designed independently but will work together to support the whole system'.

Platform ecosystem are built upon such modular designs to enable the development and execution of third-party applications. It is through their modularity that they leverage outside innovation and spur ecosystem growth [33]. Compared to traditional business models, this concept offers innovative ways for joint value creation between platform owners and external developers. On the one hand, platform owners can expand their service portfolio by integrating a new group of stakeholders into the value creation process [34]. Furthermore, third-party developers follow a solution-driven development approach, which is often unmatched by large and hierarchical organizations. On the other hand, third-party developers can use the platform's marketplace to distribute and sell their applications to a high number of potential customers [22]. By this means, developers can amortize their development costs significantly faster than by establishing own distribution channels.

Scholars and practitioners stress the relevance of boundary resources for cultivating platform ecosystems through third-party development [14, 35]. In a broad sense, platform boundary resources are any resources that help external developers in their development work [36]. In a more narrow sense, boundary resources can be defined as 'the software tools and regulations that serve as the interface for the arm's-length relationship between the platform owner and the application developer' [21, p. 176]. The boundary resource model by Ghazawneh and Henfridsson [21] describes the usage of boundary resources and the associated interplay between platform owners and third-party developers. According to this model, platform owners craft boundary resources

and provide them in a space accessibly by third-party developers. Then, third-party developers use these boundary resources to build complementary applications. Researchers also theorize that platform owners can use boundary resources to govern third-party application development [2].

Prior research was mostly limited to three types of boundary resources: SKDs, APIs [21], and technical documentation [2, 37]. We argue that these resources fall short in explaining knowledge sharing in digital platform ecosystems to its full extent. Although third-party developers acquire a profound amount of knowledge through technical documentation, we find that platform owners maintain a plethora of additional resources to address knowledge boundaries within their ecosystem. Examples for such additional resources are blog posts [19], information portals, online communities, and sample code [14]. All of these examples are designed as self-services. Through this design, third-party developers can use the resources mostly independently. Foerderer, Kude, Schuetz and Heinzl [14] also describe the above-mentioned examples broadcasting approaches because they are accessible by third-party developers without having to interact with the platform owner. Consequently, such resources have efficient scaling potentials compared to helpdesks or account managers.

In this paper, we follow these more recent approaches and investigate the role of online communities for knowledge sharing at the boundary between platform owners and third-party developers. [23] points out, that software companies invest heavily in creating official documentation for millions of topics concerning their APIs. Thereby, writing technical documentation comes inevitable with the problem that very few experts compose documentation for a large and heterogenous crowd of developers. By doing so, these companies neglect how developers integrate information from the web into their development work. On the contrary, [23] and Parnin and Treude [24] describe a process called crowd documentation, which characterizes that developers produce a huge amount of indirect documentation by publishing and reading blog posts and question and answer forums [23, 24]. Furthermore, [23] found that developers get as much as 50% of their code from online communities like StackOverflow. Additionally, developers visit online communities up to then ten times more often than the official documentation [23].

Against this background, we explore how platform owners can profit from crowd documentation posted in online communities. Our subsequent argumentation is built upon the work of Fisher [38], who reasons that firms derive competitive advantage when engaging with online communities. More precisely, Fisher [38] claims that firms can profit from three types of benefits: Information benefits, influence benefits, and solidarity benefits. Information benefits arise because members of a firm will most likely be exposed to valuable, novel, and insightful messages that are shared among community members. Examples for information benefits are market insights or user innovations. Influence benefits describe that firms may be able to utilize a sense of obligation and reciprocity when engaging with an online community. Lastly, solidarity benefits characterize loyalty and willingness to do things for one another, without an expectation of getting something in return. In other words, by building rapport, community members might be turned into evangelists for the firm's products and services [38].

In this paper, we focus on information benefits and conceptualize online communities as a key boundary resource for third-party developers. Furthermore, we define online communities as 'open collectives of dispersed individuals with members who share common interests, and these communities attend both their individual and their collective welfare' [39, p. 1224]. Emerging from technology-enabled forums, they facilitate communication and exchange among individuals and entities with shared interest [40]. However, in the information systems field, the role of online communities has mostly been discussed with respect to open source communities being a functional form of organization [41]. Some researchers investigated online communities as a means for knowledge sharing [42] and drivers for user contributions [39, 43-45]. We differentiate between autonomous and sponsored online communities [46, 47]. Whereas autonomous communities are acting mostly independent, sponsored online communities have at least one corporate entity that governs its activities. Due to our focus on digital platform ecosystems, we solely focus on online communities that are sponsored by a platform owner. Examples for such communities are SAP's Developer Community, Salesforce's Trailblazer Community or the Now Community.

While these communities have not received much attention in the platform ecosystem literature yet, they offer the potential to generate significant insights into the work and problems of third-party developers. For example, platform owners can engage in moderating behavior and thereby build relationship and trust with external developers. Some companies even use online communities as a social customer relationship tool (e.g. the Microsoft Office Support Forum) [48]. However, not only platform owners benefit from online communities. As mentioned above, online communities are strongly embedded into the work of software developers. For example, third-party developers can share development related problems and ask for solutions to be provided by the community. Also, when searching the web for potential solutions, online communities provide a vast amount of peer-reviewed knowledge articles. Prior research has shown that platform owner's engagement in sponsored communities has a significant positive effect on member's knowledge contributions [49]. In this paper, we seek to explore the information benefits that platform owners derive when engaging in sponsored communities. Thus, we formulate the following research question:

> *Research question: How can platform owners generate information benefits when engaging in sponsored online communities?*

3 Dataset and Research Method

To answer our research question, we conducted a single case study with SAP being the focal firm of our study [50]. We chose SAP for several reasons. First, SAP has a long history of collaborating with external partners to develop extensions for their ERP system. In other words, the modularity of their systems existed several years before the platform ecosystem literature emerged. For that reason, SAP managed to establish a large and dynamic ecosystem of partners and consulting firms around their technology. Second, we chose the context of enterprise software because the adoption of complex digital platforms requires complementary and specialized knowledge to unlock their

productivity [14]. Consequently, frequent knowledge exchange between the platform owner and third-parties is necessary to establish a successful ecosystem. The extensibility of SAP's system has been further increased by the introduction of the SAP Cloud Platform [16]. Third, due to the idiosyncratic and specific needs of customers, SAP's products require customization to fit specific business practices. Therefore, we assume an accumulation of expert knowledge by third-parties. Fourth and most significant, SAP is hosting the SAP Community Network since 2003. Back then, the community was a major knowledge hub for developers of SAP's partner firms. Over the years, the community evolved into a knowledge repository for several other stakeholders such as SAP users, technical architects, consultants and system integrators. Today, the community comprises several areas: A question and answer forum, expert blogs, a technical library, a code-sharing gallery, e-learning catalogues, and wikis [15, 49]. Eight years after its introduction, the SAP community network had more than 2.5 million monthly active users [51].

We developed a web scrapper to extract data from the SAP community network. In particular, we crawled the question and answer forum of the community. In this forum, community members post questions that are answered by SAP employees or by other community members. Once a question has been posted, other members can either answer or comment on the question. Members can use 'likes' to upvote contributions of others. The thread initiator can mark an answer as 'accepted' to indicate that the answer solved his problem. With more than 2.5 million questions, the forum contains a vast amount of knowledge related to SAP's technology. Due to the scope of the paper, we limited ourselves to the topic 'cloud platform', crawling only pages that had the tag 'SAP Cloud Platform'. We collected the data in October 2019 and retrieved a dataset of 4866 pages. For our analysis, we used four data points per page. First, we excerpted the title of each page. Usually, the title describes the respective question in a short sentence (e.g. *'On-Premise connectivity without using cloud connector'*). Second, we extracted the question asked by the thread initiator. Third and fourth, we collected all corresponding answers and comments (see [52] for an example).

Next, we used a text mining approach [53] to analyze the huge amount of digital trace data [54-56]. Text mining is a method for analyzing big chunks of textual data like blog posts, social media data, or online discussion forums [29, 57]. Due to its automated, computationally intensive approach, it is an adequate method for analyzing large data sets such as SAP's developer community. Furthermore, it enables researchers to analyze text collections that are too large to code by hand [54]. Researchers have used several approaches for text mining, for example latent semantic analysis [58], probabilistic latent semantic analysis [59], latent dirichlet allocation [30] and sentiment analysis [60]. We used the latent dirichlet allocation (LDA) algorithm [30] of the python package 'Gensim' [61], because it enables the discovery of latent structures in textual data. With more than 28,000 citations, the LDA algorithm is one of the most frequently used algorithm for text mining [30]. Studies using the LDA algorithm have been published in leading IS journals, such as MIS Quarterly [62].

More precisely, we used topic modeling – an approach that uses the LDA algorithm [29]. Topic models rely on statistical associations of words in a text to generate latent topics. Such models search for clusters of co-occurring words that represent higher

order constructs [29, 63]. Compared to traditional research methods like interviews or surveys, topic models provide a computational lens into the structure of large text collections [64]. A disadvantage of the LDA algorithm is that it does not consider how topics are related to one another. We addressed this issue by in-depth sensemaking and content analysis of the topics. Additionally, we clustered semantically related topics into meta-topics.

Before we transformed our data into the required estimation form, we cleaned it from remaining HTML-tags. Then, we followed the steps as outlined in the literature [29, 61, 65]. We started with lowercasing our documents before we tokenized them by splitting them into single words. Thereafter, we lemmatized our tokens by transforming them into their dictionary form. The removal of irrelevant stop words such as 'this' or 'it' was done with the list of stop words from the python package 'nltk'. Where necessary, we manually added stop words during the first iterations of our model estimation. We added bi-grams and tri-grams for tokens that appeared more than 5 times. In the end, the data consist of 35729 unique tokens that we derived from 17058 documents.

We specified our model parameters as follows. First, we had to determine the number of topics to extract. Therefore, we used the number of unique tags as a proxy for the amount of topics [66]. Consequently, we evaluated all 'SAP Cloud Platform' sub-tags (e.g. 'SAP Cloud Platform Integration Suite') and merged similar sub-tags into one topic. For example, the tags 'SAP Cloud Platform Big Data Services' and 'SAP Cloud Platform Big Data Services Tools' were synthesized into a single topic. Once we evaluated the coherence of all tags, we decided to extract 25 topics from the data. Then, we set the chunk size to 17058 to process all our documents at once. Passes specify how often we train the model. We checked when additional passes added only marginal improvements. Consequently, we set this value to 25. Finally, we decided to loop over each document for 100 iterations to reach proper document convergence. **Table 1** summarizes the parameters.

Table 1. Model parameters

Parameter	Value
Number of topics	25
Chunk size	17058
Passes	25
Iterations	100
Number of tokens	35729
Number of documents	17058

We trained the model and received 25 topics respectively. More precisely, the model provided us with word combinations that co-occur frequently within the documents. Similar to Shi et al. [62], we focused on the top ten words per topic. Then, we applied qualitative sensemaking as outlined by Lindberg et al. [56] and analyzed the word-topic combinations in-depth. This analysis started with gathering and investigating examples in which the word combinations occur. We followed up with an iterative process of labeling the topics and stopped once we reached saturation. Background research was carried out where necessary. Once we had a clear concept of the topics, we started to developed topic descriptions. Based on the examples gathered in the previous steps, we searched for illustrative examples of the topics. Since the LDA algorithm does not consider relationships between the topics, we clustered our topics into six meta topics to further improve the structure and clarity of our results. Due to space constraints, we only report meta-topic names, topic names, topic descriptions and examples. A list of word-topic combinations, as well as a list of example pages per topic is available from the authors upon request.

4 Results

We report the results of our analysis in **Table 2**. Based on our findings, we developed the following six meta-topics: User Accounts and Authentication, Connectivity, Cloud Database, Specific Technologies, SAP Resources and Installation. In the following, we describe these meta-topics by using illustrative topic excerpts.

An exemplary topic from the meta topic **User Accounts and Authentication** is Trial Account Privileges. While drilling into the details of the topic, we found that SAP is providing free trial accounts for the SAP Cloud Platform. However, these accounts come with inferior account privileges that result in several unexpected errors. One user reports the following issue: *"Everything goes fine except when doing create table statement, an insufficient privilege error appear[s]"*. The meta-topic **Connectivity** comprises all topics regarding backend connectivity. An exemplary topic is Cloud Connector. This connector was developed by SAP to connect existing on-premises systems with the SAP Cloud Platform. Such integration is – of course – not done on the fly. Another user reveals: *"And after that, whenever I try to open the Cloud Connector, it says 'Could Not open Service manager'"*. The meta topic **Cloud Database** contains all topics related to managing cloud databases and their associated tools. For the topic Database Administration, a user reports: *"The error message is Existence of database user/schema for schema Idf2c could not be checked in the underlying DBMS system due to an error. Contact HCP support if the error persists"*. The meta topic **Specific Technologies** comprises several topics that focus on a single technology such as OSGi, OData or the SAP Document Service. OData is an open communication standard for REST APIs. It is part of the SAP Gateway. An example for this topic is a user who states: *"My team has set up a[n] odata provisioning in Neo environment [...]. However, we have a new requirement to reuse the odata provisioning destination in [the CloudFoundry environment] [...]"*.

Table 2. Model results

Topic	Description	Example
Meta-Topic: User Accounts and Authentication		
User Authentication	Issues related to accessing restricted applications or systems.	*"Need help to register an user for an application in Cloud IDP when I enable the option in User Application Access: Private (Only users registered by the application can log on)"*
Trial Account Privileges	SAP provides free trial accounts for their cloud. Issues due to missing privileges of trial accounts or users.	*"I'm creating a JAVA app to create a table and access data within HDI in SCP with trail account. Everything goes fine except when doing create table statement, an insufficient privilege error appear."*
Cloud Platform Account	Issues related to accessing or creating cloud platform (trial) accounts.	*"I've verified this behavior with different developers. It's the same and all of them now blocked from accessing their SCP trial accounts."*
Anonymous User and Client	Anonymous logins provide access to SAP systems without any form of authentication. Issues related to such logins.	*"I setup the anonymous login according to note: "1828575 - Anonymous login not supported while calling AS2 adapter."*
Meta-Topic: Connectivity		
Cloud Connector	The SAP Cloud Connector connects the cloud platform with on-premises systems.	*"I installed SAP Cloud Connector 2.0. It installed without any problem. And after that, whenever I try to open the Cloud Connector, it says 'Could Not open Service manager'."*
Connecting to Apache Server	Connection issues associated with Apache servers.	*"sometimes, while deploying a .war file to HCP, the execution fails with an 'internal server error'"*
Apache Catalina	Issues regarding the Catalina Services of Apache Servers.	*"java.lang.NullPointerException: Cannot invoke org.apache.catalina.Context.getServletContext() Anyone experienced this error after the server startups?"*
Connecting from SAP Tools for Eclipse	Issues with respect to establishing a connection between SAP Developer Tools for Eclipse and a Back-End system.	*"It seems to me that you are trying to add a HANA system on your Eclipse IDE. The error: "Connection to host 'hanatrial.ondemand.com' failed" tells me that you have tried to add your Hana Trial Account and while Eclipse tried to connect to that account and retrieve the available schema IDs it failed"*
Accessing Cloud Repository	Issues regarding connections to GIT or ABAP repositories.	*"git clone [url of repository in SAP Cloud Platform Git service], I get an error of fatal: Authentication failed for [url of repository in SAP Cloud Platform Git service]'."*
Meta-Topic: Cloud Database		
Database Administration	Issues regarding the management and administration of databases.	*"The error message is Existence of database user/schema for schema Id f2c could not be checked in the underlying DBMS system due to an error. Contact HCP support if the error persists."*
Tables and Database Schemas	Issues regarding tables and database schemas of SAP Cloud systems.	*"HANA on SCP Neo: How can I create a HANA schema with JPA/Eclipselink?"*
SAP HANA Cockpit	The SAP HANA cockpit provides several tools for administration and monitoring of HANA databases.	*"I Tried to access SAP HANA Cockpit (administration tool) after creating MDC database in SAP Cloud trial, I am getting 404 error."*
Mapping Issues	Issues regarding the mapping of values and tables.	*"I have requirement in message mapping. I need to map the value dynamically based on following. for example: EN -> ENGLISH"*
Cloud Instances	Issues related to instantiating processes or services in the cloud.	*"A week ago I created a HANA Cloud instance on the Cloud Foundry Trial environment. This SAP HANA Instance stops after a certain time of inactivity. However, today I was not able to start the instance at all. It gives a message 'Stopping Failed'."*

Meta-Topic: Specific Technologies		
OSGi	OSGi is a framework for developing and deploying modular java-based applications. It is part of SAP's technology stack.	*"I cannot find any good samples showing how to create and deploy a WAB (web application bundle) to HCP Java EE 6 Web Profile Server along with deploying the osgi bundles it requires."*
OData	OData (Open Data Protocol) is a communication standard for REST APIs. It is part of the SAP Gateway.	*"My team has set up a odata provisioning in Neo environment and the UI5 app is able to query data from it when deploying to Neo. The authentication type is AppToAppSSO. However, we have a new requirement to reuse the odata provisioning destination in CF and build a new UI5 app using that destination which deployed to CF."*
CMIS and SAP Document Service	CMIS is an open standard that allows different content management systems to interoperate. The SAP document service is an implementation of the CMIS standard.	*"In the openSAP course we showed the following scenario: The Document Service implements the CMIS protocol but is available only from apps running inside HCP. The CMIS protocol can however be proxied, such a proxy is already preimplemented and you only need to configure & deploy it, see Document Service: Access from External Applications"*
Email Integration	All issues related to Email integration of the SAP Cloud Platform (e.g. for sending notifications).	*"I want to deploy a spring boot application in SAP Cloud Platform Neo environment. It has a endpoint /sendmail which sends a mail to a particular user when called. For now, I have hard coded the credentials in application.properties file and it works."*
Kepler IDE	Kepler is a version of the Eclipse IDE. All issues associated with the Kepler IDE.	*"I am getting the following error while installing HANA tools in eclipse: Unable to read repository https://tools.hana.ondemand.com/kepler"*
Mobile Services	All issues related to SAP mobile cards and services.	*"My approach: to develop a nodejs app based on SAP Approuter. In my scenario, I was using Mobile Services on Cloud Foundry and we have Application Runtime service quotas, so I decided to build a CAP nodejs app with approuter"*
NetWeaver Technology	SAP NetWeaver is the software stack for many SAP applications. All issues related to the NetWeaver technology.	*"There exists a free Gateway Demo system provided by SAP. It provides different example services. Maybe they are useful for your needs. All details are described in post SAP Netweaver Gateway Demo System and the posts linked in that post (e..g what services are provided, how you get access ...)"*
Application Runtimes	All issues related to runtime environments and deployment of applications.	*"I am trying to create a Full Stack Application for Cloud Foundry in WEB IDE Full Stack. Project Structure/modules consists of java cds hdb. Required Project settings done. Not able to find the root cause for the same or not able to debug what could be the issue."*
Interoperability	All issues related to the interoperability of technologies used by the SAP Cloud Platform.	*"The error message indicates that you have an issue with your Java truststore. What (Open?)JDK version do you have installed? Do you have the cacert file installed in the Java folder under /lib/security?"*
Meta-Topic: SAP Resources		
SAP Help Portal	Issues related to the SAP Help Portal. This portal is a major information resource for SAP's Partners.	*"For more information on managing entitlements, see: https://help.sap.com/viewer/65de2977205c403bbc107264b8eccf4b/Cloud/enUS/c8248745dde24afb91479361de336111.html"*
Meta-Topic: Installation		
Installing SAP Tools for Eclipse	All issues related to installing SAP Developer Tools for Eclipse.	*"I had no problems installing the SAP HANA Cloud Platform Tools, but I cannot install SAP HANA Tools. (I am using Eclipse Java EE IDE for Web Developers. Version: Mars.2 Release (4.5.2) Build id: 20160218-0600 with Java Web Server)"*

The remaining two meta topics consist of one topic each. The meta topic **SAP Resources** covers the topic SAP Help Portal. This portal is a central information hub of SAP's partner and comprises content such as product hierarchies or learning journeys. The contributions from this topic usually reference some parts of the portal. In **Table 2** we provide an example in which a SAP employee answers a question by referencing an article on entitlement: *"For more information on managing entitlements, see: [...]"*. The meta topic **Installation** consists of the topic Installing SAP tools for Eclipse. A member reports: *"I had no problems installing the SAP HANA Cloud Platform Tools, but I cannot install SAP HANA Tools. (I am using Eclipse Java EE IDE for Web Developer)"*. In the next step, we discuss how SAP can generate information benefits from the ongoing discourse in the community.

5   Interpretation

The aim of this paper was to explore the information benefit that platform owners derive from analyzing activities in online communities. We showed that platform owners can use topic modeling to extract latent topics that are frequently discussed in the community. Thereby, we provide them with a means to generate information benefits from digital trace data. Furthermore, by clustering the topics into higher order meta-topics, we added semantic relationships between the topics.

Platform owners can use the aforementioned information benefit in two ways. On the one hand, they can use it to improve the tuning of existing boundary resources. On the other hand, they can use feedback from third-party developers to refine the platform core (e.g. through bug fixes). We structured the remaining discussion along these two avenues. With **Figure 1**, we provide a model that illustrates this idea.
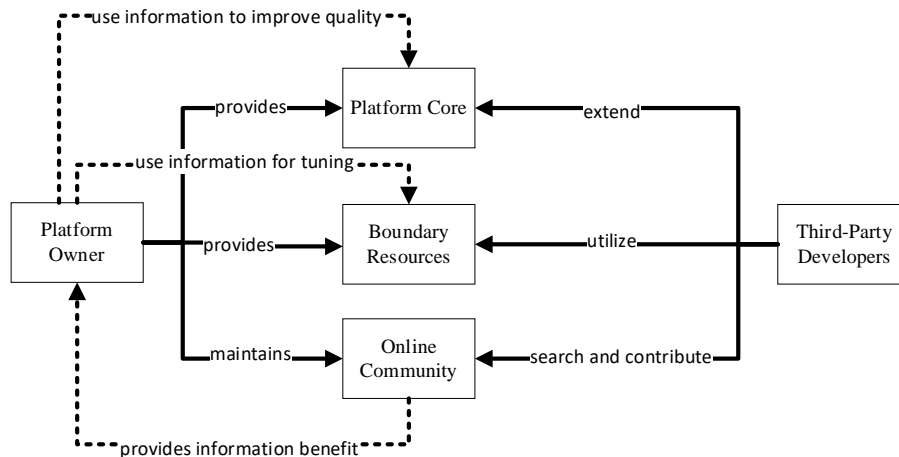

Figure 1. The role of online communities in third-party application development

Regarding the improvement of the platform core, platform owners can use the topics and meta-topics to prioritize questions and issues of the community. For example, consider the following two topics: 'Trial Account Privileges' and 'Tables and Database Schemas'. The usage of trial accounts is free of charge for community members. Issues from this topic will probably not affect any running systems. However, issues from the topic 'Tables and Database Schemas' might affect a variety of SAP's customers. Consequently, it can directly affect SAP's value delivery in a harmful way. Such prioritization is also relevant because platform owners can adjust the allocation of time and resources to the most relevant topics. Platform owners can drill into more details by comparing total and relative statistics of the topics, for example by using 'term frequency – inverse document frequency' (commonly known as TF-IDF) measures [67]. Additionally, our data-based topic extraction helps platform owners to structure areas without tags (e.g. the topic 'Accessing Cloud Repository').

Furthermore, platform owners generate significant insights into bugs that third-party developers experience. In this regard, we differentiate between actual bugs and errors that arouse from incorrect handling of the technology (e.g. wrong connection settings). For the former, we identified a prime example in the topic 'Cloud Instances'. Thereby, an issue reported by a community member led to a bug fix. After the issue was reported, an SAP employee replied: *"Update: the problem was identified and has been fixed. You should be able to start/stop your instances again. Please let us know if you still encounter problems"*. We identified another fix in the topic 'SAP HANA Cockpit'. After several community members reported an issue regarding the Admin Cockpit, an SAP employee opened a ticket. After the fix he stated: *"Hi All, [i]t should work now. I will close this ticket. If you have another issue then please open a new ticket"*.

We classify the feedback on errors that arose from incorrect handling of the technology as an information benefit. Such errors are indicators for missing, misleading, or outdated information in technical documentation or tutorials. For example, in the topic Mobile Services a user reports: *"I'm trying to follow the tutorial 'Implement Your First Screen in an iOS APP' and at step 5, when the following code has to be added [...] there is an error saying [...]"*. For this particular example, the destination of a controller was not set correct. SAP can use this feedback to update the tutorial. Another user reported: *"I am starting to play with HCP IoT Services and I am hitting a problem following the Starter Kit for SAP HCP IoT Services tutorial/instructions [...]. The problem appears when trying to simulate sending data from a device using the python script provided in the starter kit"*. Besides an information benefit regarding tutorials, we found similar issues regarding the SAP Help Portal. One user reports: *"We have build [a] Proxy Bridge for Document Service. Based on the Help guide [...] https://help.sap.com/[...]. But while [we] access the Url for testing the service [...]. [We get the] error message [...]."* All three cases demonstrate how SAP generates an information benefit regarding their boundary resources. In the topic application runtimes, we managed to identify a top contributor because his name appeared as the fifth most frequent word. Consequently, platform owners can identify experts and reward them with badges or titles.

6   Conclusion and Limitations

In this paper, we built upon a new and broader understanding of boundary resources in digital platform ecosystems. More precisely, we emphasized the necessity for extending the threefold differentiation of APIs, SDKs, and technical documentation. By considering new types of boundary resources (e.g. blogs and online communities), we contribute and expand the current discourse on knowledge sharing in digital platform ecosystems. Furthermore, we pursued on investigating how platform owners can generate an information benefit when engaging in sponsored online communities. Based on the LDA-algorithm, we presented a data-driven and text-mining based approach for generating information benefits from online community data. Furthermore, we theorize and show how platform owners can transform the results into competitive advantage.

As any other research, our paper is not without limitations. Firstly, we conducted a single case study [50, 68] with SAP being the focal firm of our study. Therefore, we acknowledge that our results are specific to our case company [69]. Although single case studies are limited with regards to drawing causations and generalizability [69], we see no issues in repeating our study with any other case company. Second, due to using data from a question and answer forum of a third-party developer ecosystem, we approve an overrepresentation of negative feedback about the platform ecosystem due to errors and issues reported in the community. We are currently digging deeper into how platform owners can use information benefits from online communities by conducting interviews.

We suggest that future research addresses the following three areas. First, whilst we focused on sponsored online communities of enterprise software platforms, future research should also investigate the role of autonomous communities for digital platform ecosystem. Therefore, other researchers should shed light onto the role of StackOverflow for digital platform ecosystem. For example, on StackOverflow, more than 1.3 million questions are tagged 'Android'. Second, whilst the research on digital platform ecosystems is still growing, the area of platform evolution remains largely untouched. We identified that gathering longitudinal case data is a major challenge for conducting research on platform evolution. By using trace data from an online community, we can use the evolvement of topics over time as a proxy for platform evolution. Third, as outlined by Gaskin et al. [70], we suggest the analysis of sociomaterial routines in third-party developer communities. Based on a typology of questions, we might derive activity-routine combinations that help platform owners in moderating their community.

References

1. Gawer, A.: What Drives Shifts in Platform Boundaries: An Organizational Perspective. In: DRUID. (2015)
2. Schreieck, M., Wiesche, M., Krcmar, H.: Design and Governance of Platform Ecosystems - Key Concepts and Issues for Future Research. In: Twenty-Fourth European Conference on Information Systems (ECIS). (2016)
3. Hein, A., Schreieck, M., Riasanow, T., Soto-Setzke, D., Wiesche, M., Böhm, M., Krcmar, H.: Digital Platform Ecosystems. Electronic Markets 30, 87-98 (2020)

4. Cusumano, M.A., Yoffie, D.B., Gawer, A.: The Future of Platforms. MITSloan Management Review 61, 45-54 (2020)
5. Yoo, Y., Boland, R.J., Lyytinen, K., Majchrzak, A.: Organizing for Innovation in the Digitized World. Organization Science 23, 1398-1408 (2012)
6. Tiwana, A., Konsynski, B., Bush, A.A.: Platform Evolution: Coevolution of Platform Architecture, Governance, and Environmental Dynamics. Information Systems Research 21, 675-687 (2010)
7. de Reuver, M., Sorensen, C., Basole, R.C.: The Digital Platform: A Research Agenda. Journal of Information Technology 33, 124-135 (2017)
8. Katz, M.L., Shapiro, C.: Network Externalities, Competition, and Compatibility. The American Economic Review 75, 424-440 (1985)
9. Tiwana, A.: Platform Ecosystems - Aligning Architecture, Governance, and Strategy. Morgan Kaufmann, Walham (2014)
10. Song, P., Xue, L., Rai, A., Zhang, C.: The Ecosystem of Software Platform: A Study of Asymmetric Cross-Side Network Effects and Platform Governance. MIS Quarterly 42, 121-142 (2018)
11. Weiss, N., Schreieck, M., Wiesche, M., Krcmar, H.: Setting Up a Platform Ecosystem - How to Integrate App Developer Experience. In: 2018 IEEE International Conference on Engineering, Technology and Innovation. (2018)
12. SAP SE, https://store.sap.com/en/ (Accessed: 22.12.2020)
13. ServiceNow Inc., https://store.servicenow.com/sn_appstore_store.do#!/store/home (Accessed: 22/12/2020))
14. Foerderer, J., Kude, T., Schuetz, S.W., Heinzl, A.: Knowledge Boundaries in Enterprise Software Platform Development: Antecedents and Consquences for Platform Governance. Information Systems Journal 29, 1-26 (2019)
15. Huang, P., Ceccagnoli, M., Forman, C., Wu, D.J.: When Do ISVs Join a Platform Ecosystem? Evidence from the Enterprise Software Industry. In: Proceedings of the Thirtieth International Conference on Information Systems. (2009)
16. Schreieck, M., Wiesche, M., Kude, T., Krcmar, H.: Shifting to the Cloud - How SAP's Partners Cope with the Change. In: Proceedings of the 52nd Hawaii International Conference on System Sciences. (2019)
17. Amazon Web Services Inc., https://aws.amazon.com/de/blogs/awsforsap/aws-and-sap-announce-iot-interoperability-solution/ (Accessed: 22.12.2019)
18. Ceccagnoli, M., Forman, C., Huang, P., Wu, D.J.: Cocreation of Value in a Platform Ecosystem: The Case of Enterprise Software. MIS Quarterly 36, 263-290 (2012)
19. Eaton, B., Elaluf-Calderwood, S., Sorensen, C.: Distributed Tuning of Boundary Resources: The Case of Apple's iOS Service System. MIS Quarterly 39, 217-243 (2015)
20. Sarker, S., Sarker, S., Sahaym, A., Bjorn-Andersen, N.: Exploring Value Cocreation in Relationships Between an ERP Vendor and its Partners: A Revelatory Case Study. MIS Quarterly 36, 317-338 (2012)
21. Ghazawneh, A., Henfridsson, O.: Balancing Platform Control and External Contribution in Third-Party Development: The Boundary Resources Model. Information Systems Journal 23, 173-192 (2013)
22. Goldbach, T., Benlian, A.: How Informal Control Modes Affect Developers' Trust in a Platform Vendor and Platform Stickiness. In: Proceedings der 12. Internationalen Tagung Wirtschaftsinformatik. (2015)
23. Ninlabs Research, http://blog.ninlabs.com/2013/03/api-documentation/ (Accessed: 22.12.2020)
24. Parnin, C., Treude, C.: Measuring API Documentation on the Web. In: Web2SE'11, pp. 25-30. (2011)
25. Oracle Corporation, https://community.oracle.com/tech/developers/categories/ (Accessed: 22.12.2020)

26. Salesforce.com Inc., https://developer.salesforce.com/forums/ForumsCategories (Accessed: 22.12.2020)
27. SAP SE, https://community.sap.com/ (Accessed: 22.12.2020)
28. Blei, D.M.: Probabilistic Topic Models. Communications of the ACM 55, 77-84 (2012)
29. Hannigan, T.R., Haans, R.F.J., Vakili, K., Tchalian, H., Glaser, V.L., Wang, M.S., Kaplan, S., Jennings, P.D.: Topic Modeling in Management Research: Rendering New Theory from Textual Data. Academy of Management Annals 13, 586-632 (2019)
30. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent Dirichlet Allocation. The Journal of Machine Learning Research 3, 993-1022 (2003)
31. Benlian, A., Kettinger, W.K., Sunyaev, A., Winkler, T.J.: Special Section: The Transformative Value of Cloud Computing: A Decoupling, Platformization, and Recombination Theoretical Framework. Journal of Management Information Systems 35, 719-739 (2018)
32. Baldwin, C.Y., Clark, K.B.: The Architecture of Participation: Does Code Architecture Mitigate Free Riding in the Open Source Development Model? Management Science 52, 1116-1127 (2006)
33. Boudreau, K., Lakhani, K.R.: How to Manage Outside Innovation. MIT Sloan Management Review 50, 69-76 (2009)
34. Schreieck, M., Wiesche, M.: How Established Companies Leverage IT Platform for Value Co-Creation - Insights from Banking. In: Proceedings of the Twenty-Fifth European Conference on Information Systems., pp. 1726-1741. (2017)
35. West, J., Mace, M.: Browsing as the Killer App: Explaining the Rapid Success of Apple's iPhone. Telecommunications Policy 34, 270-286 (2010)
36. Pruegl, R., Schreier, M.: Learning from Leading-Edge Customers at The Sims: Opening the Innovation Process Using Toolkits. R&D Management 36, 237-250 (2006)
37. Dal Bianco, V., Myllärniemi, V., Komssi, M., Raatikainen, M.: The Role of Platform Boundary Resources in Software Ecosystems: A Case Study. 11th Working IEEE/IFIP Conference on Software Architecture (WICSA), (2014)
38. Fisher, G.: Online Communities and Firm Advantages. Academy of Management Review 44, 279-298 (2019)
39. Faraj, S., Jarvenpaa, S.L., Majchrzak, A.: Knowledge Collaboration in Online Communities. Organization Science 22, 1224-1239 (2011)
40. Autio, E., Dahlander, L., Frederiksen, L.: Information Exposure, Opportunity Evaluation, and Entrepreneurial Action: An Investigation of an Online User Community. Academy of Management Journal 56, 1348-1371 (2013)
41. Lindberg, A., Berente, N., Gaskin, J., Lyytinen, K.: Coordinating Interdependencies in Online Communities: A Study of an Open Source Software Project. Information Systems Research 27, (2016)
42. Faraj, S., Johnson, S.L.: Network Exchange Patterns in Online Communities. Organization Science 22, 1464-1480 (2011)
43. Ma, M., Agarwal, R.: Through a Glass Darkly: Information Technology Design, Identity Verification, and Knowledge Contribution in Online Communities. Information Systems Research 18, 42-67 (2007)
44. Ray, S., Kim, S.S., Morris, J.G.: The Central Role of Engagement in Online Communities. Information Systems Research 25, 528-546 (2014)
45. Dahlander, L., Frederiksen, L.: The Core and Cosmopolitans: A Relational View of Innovation in User Communities. Organization Science 23, 998-1007 (2012)
46. Blohm, I., Kahl, V., Leimeister, J.M., Krcmar, H.: Enhancing Absorptive Capacity in Open Innovation Communities. In: Leimeister, J.M., Rajagopalan, B. (eds.) Virtual Communities. M.E. Sharpe Publisher, Armonk, New York (2014)
47. West, J., O'Mahony, S.: The Role of Participation Architecture in Growing Sponsored Open Source Communities. Industry and Innovation 15, 145-168 (2008)

48. Lu, Y., Singh, P.V., Sun, B.: Is a Core-Periphery Network Good for Knowledge Sharing? A Structural Model of Endogenous Network Formation on a Crowdsourced Customer Support Forum. MIS Quarterly 41, 607-628 (2017)
49. Huang, P., Tafti, A., Mithas, S.: Platform Sponsor Investments and User Contributions in Knowledge Communities: The Role of Knowledge Seeding. MIS Quarterly 42, 213-240 (2018)
50. Yin, R.K.: The Case Study as a Serious Research Strategy. Knowldge: Creation, Diffusion, Utilization 3, 97-114 (1981)
51. Hinchcliffe, D., Kim, P.: Social Business By Design: Transformative Social Media Strategies for the Connected Company. Jossey-Bass, San Francisco (2012)
52. SAP SE, https://answers.sap.com/questions/12992903/on-premise-connectivity-without-using-cloud-connec.html (Accessed: 22.12.2020)
53. Bissantz, N., Hagedorn, J.: Data Mining. Business & Information Systems Engineering 1, 118-122 (2009)
54. Berente, N., Seidel, S., Safadi, H.: Research Commentary - Data-Driver Computationally Intensive Theory Development. Information Systems Research 30, 50-64 (2018)
55. Cecez-Kecmanovic, D., Davison, R.M., Fernandez, W., Finnegan, P., Pan, S.L., Sarker, S.: Advancing Qualitative IS Research Methodologies: Expanding Horizons and Seeking New Paths. Journal of the Association for Information Systems 20, 246-263 (2020)
56. Lindberg, A.: Developing Theory Through Integrating Human and Machine Pattern Recognition. Journal of the Association for Information Systems 21, 90-116 (2020)
57. Cogburn, D.L., Hine, M.J., Peladeau, N., Yoong, V.Y.: Text Mining in Big Data Analytics. In: Proceedings of the 51st Hawaii International Conference on System Sciences. (2018)
58. Deerwester, S., Dumais, S.T., Furnas, G.W., Landauer, T.K., Harshman, R.: Indexing by Latent Semantic Analysis. Journal of the American Society for Information Science 41, 391-407 (1990)
59. Hofmann, T.: Probabilistic Latent Semantic Analysis. In: Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence (UAI1999). (1999)
60. Pang, B., Lee, L.: Opinion Mining and Sentiment Analysis. Foundations and Trends in Information Retrieval 2, 1-135 (2008)
61. Rehurek, R., Sojka, P.: Software Framework for Topic Modelling with Large Corpora. In: Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks. (2010)
62. Shi, Z., Lee, G.M., Whinston, A.B.: Toward a Better Measure of Business Proximity: Topic Modeling for Industry Intelligence. MIS Quarterly 40, 1035-1056 (2016)
63. DiMaggio, P., Nag, M., Blei, D.: Exploiting Affinities between Topic Modeling and the Sociological Perspective on Culture: Application to Newspaper Coverage of U.S. Government Arts Funding. Poetics 41, 570-606 (2013)
64. DiMaggio, P.: Adapting Computational Text Analysis to Social Science (and Vice Versa). Big Data & Society July-December, 1-5 (2015)
65. Bird, S., Klein, E., Loper, E.: Natural Language Processing with Python - Analyzing Text with the Natural Language Toolkit. O'Reilly Media (2009)
66. Safadi, H., Johnson, S.L., Faraj, S.: Core-Periphery Tension in Online Innovation Communities. Organization Science (Forthcoming)
67. Aizawa, A.: An Information-Theoretic Perspective of TF-IDF Measures. Information Processing & Management January 2003, 45-65 (2003)
68. Yin, R.K.: Case Study Research - Design and Methods. Sage, Thousand Oaks (2009)
69. Ruddin, P.L.: You Can Generalize Stupid! Social Scientists, Bent Flyvbjerg, and Case Study Methodology. Qualitative Inquiry 12, 797-812 (2006)
70. Gaskin, J., Berente, N., Lyytinen, K., Yoo, Y.: Toward Generalizable Sociomaterial Inquiry. MIS Quarterly 38, 849-872 (2014)