Towards a personal data vault society: an interplay between technological and business perspectives

Sofie Verbrugge
IDLab
Ghent University-imec
Ghent, Belgium
sofie.verbrugge@ugent.be

Frederic Vannieuwenborg
IDLab
Ghent University-imec
Ghent, Belgium
frederic.vannieuwenborg@ugent.be

Marlies Van der Wee
IDLab
Ghent University-imec
Ghent, Belgium
marlies.vanderwee@ugent.be

Didier Colle
IDLab
Ghent University-imec
Ghent, Belgium
didier.colle@ugent.be

Ruben Taelman
IDLab
Ghent University-imec
Ghent, Belgium
ruben.taelman@ugent.be

Ruben Verborgh
IDLab
Ghent University-imec
Ghent, Belgium
ruben.verborgh@ugent.be

Abstract—**The Web is evolving more and more to a small set of walled gardens where a very small number of platforms determine the way that people get access to the Web (i.e. "log in via Facebook"). As a pushback against this ongoing centralization of data and services on the Web, decentralization efforts are taking place that move data into personal data vaults. This positioning paper discusses a potential personal data vault society and the required research steps in order to get there. Emphasis is given on the needed interplay between technological and business research perspectives. The focus is on the situation of Flanders, based on a recent announcement of the Flemish government to set up a Data Utility Company. The concepts as well as the suggested path for adoption can easily be extended, however, to other situations/regions as well.**

Keywords—personal data vaults, digital ecosystems, data pods, business model, digital platform ecosystem, Solid

I. THE PROBLEM WE ARE FACING: THE WEB HAS EVOLVED INTO A SET OF WALLED GARDENS

In its origin the Web was based on freedom and independence [1][2], the idea was that everyone could get access to the Web, independent of the actual operating system or device used. Through this principle of universality, the Web allowed for *permissionless innovation*: anyone can launch their idea without depending on specific companies.

Gradually, however, the Web evolved into a set of walled gardens where a very small number of platforms (Facebook, Google, etc.) determine the way that people get access to the Web (i.e. "log in via Facebook"). Clearly, this situation has important advantages in terms of user experience (technical and user interface), but it has issues in terms of innovation potential and privacy. Users are not required to re-enter their personal data several times, application developers can rely on strong technical interfaces (SDKs, APIs, etc.) to build their applications. However, the bigger the platform grows and in absence of easy ways of switching to competing (smaller) platforms, both end users and application providers experience a very strong platform lock-in. An important observation is that the user data becomes centralized in a handful of Web platforms and that Web developers currently largely depend on these centralized platforms for data integrity and end users are indeed locked-in.

What plays here is what economists call (cross-side) network effects on top of a multi-sided platform connecting users and application providers: the more users connect to a certain platform, the more valuable it becomes for the application providers; the more application providers build applications for a specific platform, the more valuable it becomes for the users. As a result, it is this network of producers and consumers that forms the main asset of the platform, think about Google, Uber and Apple for example. Phrased differently, value creation within platform ecosystems has become cyclic (consumers generate value for producers and producers generate value for consumers), as opposed to linear value creation in traditional industries that are built on value chain dynamics (where suppliers generate value for consumers), for example in energy and automobile sectors.

II. SUGGESTED TECHNICAL SOLUTION: PROVIDE USERS WITH A PERSONAL DATA VAULT
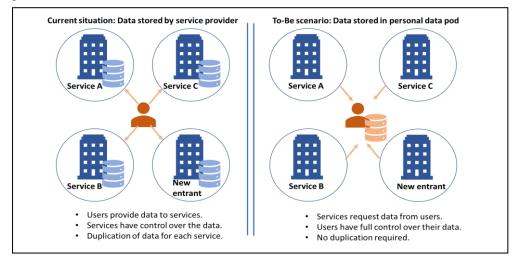
We need to shed a new light on the interplay of businesses, people, and their data. An alternative under active exploration is an ecosystem of data spaces of personal data, in which every person has their own personal data vault, which we also refer to as a data pod. Think of a data pod as a personal online data store to which you as an individual hold the keys. Your data pod can then store and safeguard every single piece of data that you produce as well as every single piece of data that companies and organizations produce about you.

As the controller of your data pod, you can decide for every piece of data which parties you want to share it with. This way, companies can get access to specific data without writing or collecting it themselves, while respecting every individual's preferences. Acquiring data through harvesting (what big platforms are doing today) is expensive and leads to suboptimal access

to data. When individuals are in control, relevant and up-to-date data can be readily available. An ecosystem based on personal data vaults brings back the choice that enables people to pick apps they need.

Several technical implementations can support an ecosystem based on personal data vaults. In our research, we base all technical assumptions on the Solid technology [3][4][5]. This technical implementation leads to the development of decentralized Web applications, based on personalized data pods. Technically, a Solid server is a regular Web server with support for access control and support for data, it is application-agnostic. Solid clients are browser apps or native apps that read/write to personal data pods (when the user gives permission). Note that this solution also has important advantages for end users in terms of their privacy and feel of control.

It is important to stress that Solid in essence provides a set of publicly available standards, that allow for multiple data pod providers and can form the basis of an open platform ecosystem. That way, any party can build servers, clients, and other infrastructure for Solid. In contrast to other pod-solutions like Doccle [6], Itsme [7], etc, Solid is more open. Rather than acting as the single pod provider in the platform, it provides a set of standards that any party can implement in order to create a pod provider.

III. WHERE TO START BUILDING A PERSONAL DATA VAULT BASED ECOSYSTEM?

Providing all end users with personal data vaults, based on Solid technology, has the potential to break the current existing platform lock-in leading to the walled gardens described above. Thanks to the fact that the Solid technology enables or even expects the co-existence of multiple data pod providers, the resulting ecosystem has the potential to grow into an alternative, open platform ecosystem where innovation is enabled or even encouraged and where end user privacy is the default. In this way, we envision a true personal data vault based society. However, the different steps in the direction of this future scenario, remain unclear for the time being. It is exactly the uncertainty around how to move towards this data vault-based society that still forms a major challenge. A sustainable business model for all actors in the ecosystem is definitely an important condition. Future research will aim at identifying suitable pathways in this direction.

As soon as a substantial number of end users dispose of a personal data vault, implemented e.g. as a Solid data pod, application providers will be motivated to create applications that are using those pods. The other way around, as soon as there is a substantial amount of applications using Solid data pods, end users will be motivated to create their own data pod and provide their data therein. This way, the cross-side network effects can start doing their work, leading to a cyclic value creating dynamic, but without incurring the lock-in effects of the walled gardens experienced today. However, as described in platform literature [8][9], it is exactly this chicken-or-egg dilemma that is one of the essential challenges in platform strategy. How to attract end users in the absence of a lot of application providers? And how to attract application providers in the absence of a lot of end users? As formulated by Cusumano, Yoffie and Gawer [10] "all platform companies face the same four business challenges. They must choose the key 'sides' of the platform (that is, identify which market participants they want to bring together, such as buyers and sellers, or users and innovators). They must solve a chicken-or-egg problem to jump-start the network effects on which they depend. They must design a business model capable of generating revenues that exceed their costs. And finally, they must establish rules for using (and not abusing) the platform, as well as cultivating and governing the all-important ecosystem."

A potential path of adoption for the Solid technology and resulting growth in the personal data vault based ecosystem, can start from an offer based on personal data pods by local government and utility companies (e.g. to store your driver's license and marriage certificate). The government can give incentives to citizens to start using individual data pods and to utility companies to start building an application based on them. Within Flanders (Belgium), the launch of a government-initiated Data Utility Company was recently announced [11][12], so that the Flemish Government will act as a Solid-based data pod provider. Once the data pods are filled with data used by public service providers and citizen get acquainted with the Solid technology, it will become more attractive for other parties to engage in this ecosystem. In a next step, commercial application and service providers in the local context can engage in the ecosystem followed by more internationally oriented application and service providers in a later step. Several types of private companies can can be faster or slower adopters of Solid-based technology based on their specific situation. The size of the company (and more importantly the budget available for innovation) can be of importance, as

well as the level to which they are bounded in the currently data silos already. Innovative SMEs without legacy may be very eager to migrate into the personal data vault based society. Bigger companies with R&D departments may be on the forefront even so. Big global platforms may follow their own logic at their own pace. Google already provides links to other non-Google platforms such as Doodle (via OAuth protocol), but the data remains in their control. However, intentions and efforts concerning personal data sharing continuously grow since they have participated in the Data Transfer project (together with other companies such as Apple, Twitter and Microsoft) [13].

A. Potential success factors and remaining challenges

A precondition for the successful setup of a personal data vault based ecosystem is to find one or more actors willing to take up the role of data pod provider, a role that is not existing in the current ecosystem organized around walled gardens. Based on the initiative of the Flemish Government, the Data Utility Company will definitely be a first candidate for taking up the role of data pod provider in Flanders. In a later stage, we expect also commercial cloud providers to appear in the ecosystem (e.g. Amazon, Google or more local companies such as Combell). Since the technology is standards-based, data vaults can be moved between providers at any point. These public and commercial data pod providers will (jointly) act as platform providers for the set-up of the personal data vault based ecosystem. Based on the interest of both public and commercial data pod providers, we can be confident that the precondition for the existence of a personal data vault based ecosystem will indeed be fulfilled. Therefore, the major challenge for a successful ecosystem remains attracting both application providers and end users in sufficient amounts (solving the chicken-and-egg problem mentioned above) while ensuring a sustainable business model for the resulting (platform) ecosystem. As a consequence, future research in this domain should focus on the position of both end users and application service providers in the ecosystem based on personal data vaults.

The Solid technology can act as a stable software-based platform with boundary resources, such as Solid-specific tools (Solid Filemanager: to interact with Solid Pod Sessions, Solid Local Pod Manager: to serve parts of a local filesystem as local pod, etc.), and Solid-specific libraries for Authentication/Access control (ACL Check), Linked data (rdflib, GraphQL-LD for Solid), etc. [14]. This allows application and service providers (the complementors mentioned above) to develop innovative Solid-based applications. As stated in [15] and [9] the strength of such technology-based platforms stems from the fact that complementors can build applications without extensive knowledge of the underlying architecture. As the Solid technology provides common technological building blocks (for Authentication, Linked Data, Resource management and Interfacing [14]) that can be used to build complementary products and services, the Solid platform can be considered an example of an innovation platform [10]. As opposed to the better-known transaction platforms that perform pure matchmaking between e.g. sellers and buyers, innovation platforms allow complementors like application service providers to develop complementary products or services, using boundary resources provided by the platform owner, like software development kits [9]. Customers, on the other side, are the beneficiaries of these products or services. The link between technical properties of the digital platform and the value-creating mechanisms in the resulting platform ecosystem remains to be better understood [9][16][17]. What form should the boundary resources take in order to facilitate application development and resulting value creation by the complementors? In order to get to a better understanding of this matter, insights from the technical and the business perspective will need to be combined.

B. Literature in different research disciplines points to a leap of knowledge

Given the assessment of the personal data vault based ecosystem as a multi-sided platform above, the open challenges related to multi-sided platforms also apply here. Several authors have pointed to open research questions in digital platform ecosystems that are interdisciplinary in nature. Rietveld and Schilling [18] bring together insights from the fields of management and organization, information systems, economics and marketing and call for future interdisciplinary research. De Reuver et al. [17] set forward a research agenda for digital platform research in information systems and recommend scholars to study platforms on different architectural levels and in different industry settings while a.o. providing clear definitions that specify the unit of analysis, degree of digitality and the sociotechnical nature of digital platforms. Schreieck, Wiesche and Krcmar [19] provide a research review on the design and governance of platform ecosystems and "suggest future research to (1) integrate the market- and technology-oriented perspective when discussing phenomena on platform ecosystems, (2) integrate complementors and end-users into the analysis in addition to the platform owner, and (3) study data as boundary resource in more detail".

As such, one of the major challenges in setting up a successful digital platform relates to the platform governance, i.e. the way that the platform is controlled and operated and how decisions are made. Van Alstyne and Parker [20] state that, because of this difficult governance issue, for every successful platform, there are many more that struggle or simply don't make it. Hagiu [6] identifies platform governance rules as one of the four major strategy challenges for so-called multisided platforms. This includes deciding setting up strategies in order to regulate access to the platform, i.e. decide who is allowed to join the platform and organize the interactions on the platform, i.e. what are the users and suppliers allowed to do and how. Concerning the use of personal data vaults some authors explicitly point to issues in the set-up of the ecosystem, the definition of a sustainable business model for the resulting digital platform as well as an appropriate way to orchestrate the activities between the different stakeholders involved [21].

IV. RESEARCH CHALLENGES AHEAD

In order to evaluate the feasibility of digital platform ecosystems in general and the market potential of personal data vault based ecosystems in particular, some interdisciplinary research questions remain to be answered.

More specifically, these link to the three points that were made explicit by Schreieck, Wiesche and Krcmar [19], see also above. First, access to data, and more specifically personal data, has been identified as the key boundary condition for the set-up of a new successful digital ecosystem, away from the existing walled gardens. The Solid technology provides the means for

decoupling data from applications, avoiding the lock-in that is omni-present today. The impact of direct access to personal data shouldbe studied in detail. Second, next to the technical development in the domain of Solid, major insights have already been gathered in the platform business ecosystems and the potential impact of the markets in different application domains. Based on interaction between technical researchers and business modelling/ techno-economic researchers, market and technology perspectives need to be integrated. Third, given the fact that data pods providers start arising (the Flemish Data Utility Company as well as commercial pod providers), who act as platform owners, the first analysis should focus on integrating application providers (the complementors in this platform) and end users.

Based on the observations above, two major outstanding research challenges have been identified. A first research objective is to understand the relationship between technical requirements and the value-creating potential in a digital ecosystem based on personal data vaults. A detailed insight in the current technical possibilities of the Solid solution will allow to highlight the value-creating potential by application and service providers (complementors to the existing personal data vault based ecosystem) towards end users (and vice-versa). On the other hand, a detailed study of the business potential of some realistic use cases based on personal data vaults may allow to identify additional requirements or boundary conditions for further technical development.

The second research objective is to get insights in the trade-offs that both end users and application and service providers are facing when deciding to engage in a personal data vault based digital ecosystem. End users will decide whether or not to store their personal data in a personalized data pod based on a utility-based benefit-cost trade-off on an individual level. This trade-off should not be seen as purely monetary. The use of the personalized data pod may lead to easier switching between offers by different application and service providers and allow for access to a wider range of offers than the ones restricted to the walled garden ecosystem they are currently situated in. Next to that also privacy issues are reduced. However, these benefits come at the cost of migrating towards this new ecosystem. For the application and service providers, the situation is even more complex, because the benefits offered by easier access to new customers comes at the cost of losing control over the personal data of all customers.

Tackling these two major research challenges will allow to formulate tangible recommendations towards different actors in the digital ecosystem based on personalized data vaults. Besides end users and application and service providers, recommendation and guidelines can also be formulated towards data pod providers, the government, and Solid technology developers.

V. USE CASES TO INDICATE THE IMPACT OF THE TRANSITION TO A PERSONAL DATA VAULT BASED ECOSYSTEM

In order to have tangible outcomes for actual stakeholders and to fully understand the potential as well as the existing barriers of the envisioned personal data vault based society, the research should be closely linked to real-life implementations and applications.

A. *Personal data used and processed by the retail sector: The case of a supermarket.*

A first potential use case stems from the retail sector where companies have always been very eager to collect customer data (since these data provide valuable insights in consumption patterns and fuel their operational processes). This data is used for marketing efforts, operations incl. demand forecasts and logistics planning, customer binding and finances/accounting. When purchasing a product in a physical store, purchase data gets linked at the counter. A transaction ID is coupled with product IDs, a timestamp, in addition maybe an ID of the cashier and an ID of the payment type or credit/debit card is added. Processing this data can be done in an anonymized way and insights in this data are extremely valuable for both the retail shop owner and the product supplier. However, from the moment a customer signs up as a member in e.g. the retail chain loyalty program, via an app or on paper, additional personal info can also be linked to this transaction. Further processing of this data typically results in a personalized service experience in terms of specifically directed marketing efforts (e.g. incentive systems, health points, etc.). Today this is done based on retailer specific applications.

There are clear benefits for the business actors in the situation as-is where all retail companies work with siloed inhouse applications. Retail companies benefit from customer loyalty via lock-in (both technically and economically). Product suppliers benefit from the fact that they can have insights into anonymized product consumption data. However, these actors also face some challenges. Retail companies are responsible for storing and maintaining all the data, which is especially hard for non-frequent clients (outdated data) according to the GDPR policies. Product suppliers typically do not have access to customer purchases except if they buy it from the retail companies. Similarly, third party service providers which are interested in processing these data for other purposes do not get easy access. The biggest challenges are for the end customers as they do not have real control on the use of their data. They experience high transaction costs (in terms data portability to other retail companies) as well as effort duplication (for each retailer a specific app).

In the situation to-be, we can envision the existence of in-house applications of the different retail companies that make data requests back and forth with the personal data pods of the clients. Several benefits can be expected in this situation. There will be better control of the personal related data usage and processing. For the retail company, data storage and data maintenance can be reduced since personal data is no longer stored in their system. When having access to specific parts of the personal data pods of the customers themselves, data quality will improve (outdated data will no longer be an issue). Further service customization is possible when retail companies also receive access to different types of personal data. Product suppliers as well as third parties can request data access permission to the customers directly instead of buying data reports on purchase behavior from the retail companies. However, there are also challenges, specifically for the application and service provider that experience a weaker market position as they no longer dispose of the customer data, so that their link to the end customers

reduces. In extremis, the latter could result in increased customer churn, which is the opposite of the intention of customer loyalty programs.

B. Other interesting use cases: Smart meter data and personal profiles made by employment agencies

In Flanders, as it is the case for other European Member States, grid operators have been deploying smart energy meters during the last years. Via this meter, energy consumption and production data of every household is transmitted every 15 mins. Because the energy consumption pattern can tell a lot about the lifestyle of the client, this data must be seen as personal data. On the other hand, this data is also key for safeguarding a proper energy supply and distribution in the era of intermittent energy generators such as solar and wind power (for energy consumers and suppliers, distribution, and transmission operators). The data platform that will digest, process, manage and share all this raw data of the smart meters with the relevant stakeholders (e.g. Energy providers, Demand Response aggregators, etc.) is designed as actor-neutral clearing house in the case of Flanders. These different actors need these data for billing and operational purposes, but in addition can use it to synthesize user profiles from it. Although there are vivid ongoing discussions on who is the actual owner of this data, the reality is that end customers typically do not have control over these profiles and data.

A transition to a personal data vault-based future would turn the current modus operandi upside down, since it would mean that all generated data would be stored in a personal pod and that the different energy actors should request access to that data. In the case these actors use the data to model personal energy profiles, also this data should be stored in the personal data vaults. Although this approach would give full control to the end users and this data would be crucial to allow new energy actors with new services on the energy market, without the need of redoing all the user profiling, still many technical and economic barriers exist. For instance; How can it be assured that the raw meter data is not tampered with by the end users? What is the incentive for energy actors to share user profile data with the end users themselves?

A last example of a potential use case to indicate upcoming challenges is focusing on employment agencies. As a matchmaker, employment agencies want to have a correct and complete overview of the profile of a job seeker to be able to select the perfect candidate for a specific job position. Although the quality of the linking process (and therefore also the profile of a candidate) is a determinant of the company value, being able to share personal profiles between competing employment agencies could also be beneficial. Benefits of the Solid scenario for employment agencies would be: less costs and effort to store the data and keep it up to date, more resources to focus on the matchmaking process itself, easier access to new personal data points to improve the matchmaking process. At the same time this poses challenges from an economic perspective; what are the incentives and effects for an employment agency to share profiling data with their clients and allow them to share it with competition? What assurance do employment agencies have that these profiles are original and not adjusted by the job seekers themselves?

VI. SUMMARY

Within this paper we sketched the issues with the existing implementation of digital platforms, that have basically reduced the interaction of individual customers with companies over the Internet to a set of walled gardens. We shortly present the technical potential of a personal data vault based implementation, as proposed by the Solid technology.

Putting the envisioned personal data vault based ecosystem in the perspective of the multisided platform literature, we have highlighted the importance of attracting enough users at all sides (individual customers as well as application and service providers) in order to generate cross-side network effects that ultimately lead to actual value for the platform ecosystem. An essential condition for the potential existence of such an ecosystem, is the presence of data pod providers. Given the fact that governments have already shown interest to take up this role (launch of a Data Utility Company that has been announced by the Flemish Government) and the commercial pod providers are expected to follow soon after, this precondition seems to be met. Focus can then be on the adoption of the personal data vault by the other actors in the ecosystem, mainly application and service providers as well as end users. A step-wise approach is foreseen. Once the data pods are filled with data used by public service providers, it becomes more attractive for other parties to engage in this ecosystem. Private application and service providers are expected to follow after the public companies have paved to way. We have made a call for future research on two fronts. A first research objective should be to understand the relationship between technical requirements and the value creating potential in a digital ecosystem based on personal data vaults. The second research objective relates to getting insights in the trade-offs that both end users and application and service providers are facing when deciding to engage in a personal data vault based digital ecosystem.

Finally, we have identified that research should be linked to or based on realistic use cases of personal data vault based solution in different application domains.

REFERENCES

[1] Berners-Lee, T. J. (1989). Information management: A proposal (No. CERN-DD-89-001-OC).

[2] Berners-Lee, T. (n.d.). Long Live the Web: A Call for Continued Open Standards and Neutrality. Scientific American. Retrieved March 29, 2021, from https://www.scientificamerican.com/article/long-live-the-web/

[3] Mansour, E., Sambra, A. V., Hawke, S., Zereba, M., Capadisli, S., Ghanem, A., Aboulnaga, A., & Berners-Lee, T. (2016). A demonstration of the solid platform for social web applications. 223–226.

[4] Werbrouck, J., Pauwels, P., Beetz, J., & van Berlo, L. (2019). Towards a decentralised common data environment using linked building data and the solid ecosystem. In 36th CIB W78 2019 Conference (pp. 113-123).

[5] Capadisli, S., Guy, A., Verborgh, R., Lange, C., Auer, S. and Berners-Lee, T. (2017), "Decentralised Authoring, Annotations and Notifications for a Read–Write Web with dokieli", in Proceedings of the 17th International Conference on Web Engineering, pp. 469–481, available at: https://csarven.ca/dokieli-rww

[6] Doccle, https://doccle.be/en/how/

[7] Itsme, https://www.itsme.be/en

[8] Hagiu, A. (2015). Strategic decisions for multisided platforms. MIT Sloan Management Review, Top 10 Lessons on Strategy, 4-13.

[9] Hein, A., Schreieck, M., Riasanow, T., Setzke, D. S., Wiesche, M., Böhm, M., & Krcmar, H. (2019). Digital platform ecosystems. Electronic Markets, 1-12.

[10] Cusumano, M. A., Yoffie, D. B., & Gawer, A. (2020). The future of platforms. MIT Sloan Management Review, 61(3), 46-54.

[11] Vlaamse Regering, Relanceplan Vlaamse Regering – Vlaamse Veerkracht, Beleidsdocument • september 2020, https://www.vlaanderen.be/publicaties/relanceplan-vlaamse-regering-vlaamse-veerkracht

[12] Van Leemputten P. (2020). Vlaamse regering wil data-nutsbedrijf oprichten. Datanews, September 2020.

[13] Willard, B., & Fair, G. (2018, July 20). Introducing Data Transfer Project: An open source platform promoting universal data portability. Google Open Source Blog. https://opensource.googleblog.com/2018/07/introducing-data-transfer-project.html

[14] Tools and libraries overview · Solid. (n.d.). Retrieved March 30, 2021, from https://solidproject.org/developers/tools

[15] Tiwana, A., Konsynski, B., & Bush, A. A. (2010). Platform evolution: Coevolution of platform architecture, governance, and environmental dynamics. Information Systems Research, 21(4), 675–687.

[16] Tilson, D., Lyytinen, K.,&Sørensen, C. (2010). Research commentary— Digital infrastructures: The missing is research agenda. Information Systems Research, 21(4), 748–759.

[17] de Reuver, M., Sørensen, C., & Basole, R. C. (2018). The digital platform: a research agenda. Journal of Information Technology, 33(2), 124-135.

[18] Rietveld, J., & Schilling, M. A. (2020). Platform competition: A systematic and interdisciplinary review of the literature. Journal of Management, 0149206320969791.

[19] Schreieck, M., Wiesche, M., & Krcmar, H. (2016). Design and governance of platform ecosystems–key concepts and issues for future research.

[20] Van Alstyne, M., & Parker, G. (2017). Platform business: from resources to relationships. NIM Marketing Intelligence Review, 9(1), 24-29.

[21] Vojíř, S., Smutný, Z., & Kučera, J. (n.d.). SOCIAL AND TECHNICAL ASPECTS OF RE-DECENTRALIZED WEB. Proceedings of the 28th Interdisciplinary Information Management Talks, 107–116.Comment platform vs standard or ecosystem