Challenging Dropbox: The Adoption and Usage of New Cloud Storage Service "sciebo" at German Universities

Dominik Rudolph¹, Raimund Vogl¹, Holger Angenent¹, Anne Thoring³, Andreas Wilmer¹, Christian Schild³

¹Westfälische Wilhelms-Universität Münster, ZIV, Röntgenstraße 7-13, 48149 Münster/Germany, {d.rudolph|r.vogl|h.angenent|a.thoring|a.wilmer|schild}@uni-muenster.de

Keywords
diffusion of innovations · adoption · cloud-storage · project report

1. ABSTRACT
As reported at EUNIS 2015, a large academic cloud storage service was launched in the beginning of 2015 by the majority of the public research and applied science universities in the German state of North Rhine-Westphalia (NRW) under the brand name “sciebo”. One year after the start, we will examine if the predictions made in the preparatory project phase on system performance and on service adoption based on the well-known diffusion model by Rogers apply to reality. This is the first study about the adoption of a specific cloud service at several universities. We identify two factors affecting the speed of diffusion: share of technophiles and the use of marketing measures. Organization size does not seem to influence the speed of diffusion. Regarding system performance and availability, one year of operation shows that the agreed on targets were met.

2. INTRODUCTION
Cloud storage services like Dropbox or Google Drive became quite popular in the course of the last half decade, not least in the academic context among students and researchers, making it possible to easily share documents with others and to synchronize data across multiple devices. Those commercial services are very comfortable in use, but security concerns about their data utilization arise, especially after the Snowden disclosures. In 2013, as a consequence, the majority of the public research and applied science universities in the German state of North Rhine-Westphalia (NRW) formed a consortium to start a jointly operated private cloud service for the academic community. This sync and share storage platform should be free of charge, easy to use and, most importantly, it should be hosted on premise at several university data centers to be fully compliant with German data protection regulations (Vogl et al., 2013). With respect to the software functionality and the required hardware setup for potentially 500.000 users, the system design was grounded on empirical user studies.

A first exploratory survey on the demand for a university operated alternative to Dropbox etc. was conducted among potential users at Münster University in 2012 and extended to a multi-site survey with more than 10.000 participants from three major universities in late 2013 (Stieglitz, Meske, Vogl, & Rudolph, 2014). Both surveys focused on the participants’ intention to use such a university operated cloud service, their demand for storage space and client platforms, the type of content (file types) they intended to store, and the communities they wanted to collaborate with using the service’s file sharing functionalities. The procurement of the software solution as well as the sizing of the hardware platform were based on the adoption and usage estimates derived from these surveys.

In February 2015, after extensive preparatory work done for the funding proposal, the procurement process, and the system setup and commissioning, the sync and share cloud storage service was launched under the brand name “sciebo - theCampuscloud” (sciebo being short for science box) with three university data centers (Bonn, Duisburg-Essen and Münster) hosting the system platforms on premise. Almost exactly one year after the start, it is now the right time to review how the initial expectations on service adoption and usage as well as system performance and availability correspond with reality. After a year of operation (as of Feb 02 2016), exactly 40.000 users from 24 universities...
and one public research center have signed up for sciebo through the self-enrollment web portal.

The case of sciebo is unique because it allows us to observe the diffusion of a technical innovation from the beginning in a well-controlled setting. There is plenty of literature about the adoption of cloud systems in organizations like SMEs (Alshamaila, Papagiannidis, & Li, 2013; Tehrani & Shirazi, 2014; Trigueros-Preciado, Pérez-González, & Solana-González, 2013) or special industries (Cegielski, Allison Jones-Farmer, Wu, & Hazen, 2012; Gupta, Seetharaman, & Raj, 2013; Ivanova & Ivanov, 2010; Khanagha, Volberda, Sidhu, & Oshri, 2013; Lian, Yen, & Wang, 2014; Low, Chen, & Wu, 2011; Morison & Moeser, 2016; Oliveira, Thomas, & Espadanal, 2014), but only little is known about the adoption behavior of end-users who can decide freely if they want to use a new cloud service or not (Shin, Jo, Lee, & Lee, 2014). Universities are a special case: On the one hand, they are organizations with a quite uniform population and a manageable size. On the other hand, because of the principle of freedom of research and teaching held high in Germany, there is no possibility to command the use of a system, so users have to be convinced.

3. PREDICTIONS

In preparing the sciebo project and applying for substantial funding, reliable predictions on the user adoption of this new service were crucial for the system design and amount of hardware to procure. For the university data centers volunteering to host the platform, estimates of the internet bandwidth to be dedicated to sciebo was important, and for the universities that had to decide if they wanted to join the sciebo project consortium it was necessary to know what quality of service, especially with respect to system availability, they could expect. Thus, based on empirical research, predictions were made on the required storage volume for the sciebo system platform and the required internet bandwidth - both directly connected to the adoption of the new service by its eligible users. System availability scores were estimated based on the analysis of three years of well documented operations incidents at the University of Münster.

According to Rogers (Rogers, 2003), innovativeness, i.e. the readiness and the degree to which a person or an organization adopts an innovation (i.e. a new product) compared with the other members of his population, follows the Gaussian distribution. He identifies five adopter categories (innovators, early adopters, early majority, late majority, laggards; see Fig. 1) who have different characteristics referring to their innovativeness. If you accumulate the adoption decisions of all those adopters over time, you will get an S-shaped curve, the diffusion curve. The faster the innovation is adopted the more steeply this curve will rise. The speed of diffusion depends on the characteristics of the innovation, in particular its relative advantage compared to other existing products, compatibility with existing values and practices, simplicity and ease of use, trial ability and observable results.

As our data from a large user survey conducted in 2013 (Stieglitz et al., 2014) show, Dropbox is the most used storage service among members of the universities with about 80 percent of market share.
This value recurs in another survey conducted in 2015 at the same universities (unpublished work), so we conclude that Dropbox obviously has reached the saturation of demand five years after its inception in 2007 and two years after the release of the first stable version 1.0 in 2010. Examining Dropbox’s worldwide diffusion (see Fig. 2), a flat growth is visible in the first two years and a take-off in the third year.

![Fig. 2. Diffusion of Dropbox (in mio. users) (Dropbox, 2015)](image)

For sciebo, we predicted an even faster diffusion because the technology is already known from Dropbox. Moreover, sciebo’s high security standards and bigger free storage space seem to be significant relative advantages as stated by the participants of a survey in 2013 (Stieglitz et al., 2014). According to Diffusion Theory, market potential is not the total of all potential users (i.e. all members of the participating universities), but the total of all those persons who will realistically use a new service (Kleinyltenkamp & Plinke, 1999). In the survey, 92.5 percent of the participants stated that they wanted to use sciebo. Being informed that their usage authorization would be revoked when leaving university, the count dropped to 65 percent. Thus, 65 percent of all members of the participating universities - that is about 252,000 individuals - constitute the estimated market potential of sciebo.

Based on the distribution of per user storage demands from the survey, we could refine the initial assumption that each user would utilize the planned 30 GB quota to the max and were able to predict an average storage volume of 8 GB (pessimistic scenario) to 16 GB (optimistic scenario) per user and could ascertain that a maximum storage space of 30 GB should fit most users. Assuming that users would switch their academic data from another platform to sciebo in the first days after the registration, we expected a quite linear growth with a 30 percent basis synchronization at the beginning and just small gain of 3 percent a month.

Considering the predictions on service adoption and storage demand, different scenarios were derived to estimate the size of storage systems to be procured and the internet bandwidth required. The total storage volume required for the operation of sciebo in the long term was estimated at 1.7 PB (pessimistic scenario) to 5 PB (optimistic scenario), and the internet connection bandwidth requirement for service operation was estimated at 3 Gbps in the optimistic scenario.

The predictions on system availability (based on three years of incident logs at Münster) resulted in an agreement amongst the sciebo consortium partners that the availability scores have to be 99.5 percent per year for each of the sites with a minimum of 98 percent per month.
4. FINDINGS

4.1. User Diffusion

Nearly one year after the official launch, sciebo hit another milestone with now 40,000 users – this means an actual market share of 17.3 percent. In terms of the Diffusion Theory, this implies that sciebo’s diffusion has already reached the early majority phase.

However, diffusion speed varies significantly at the different universities. Figure 3 shows the state of diffusion at the 14 universities that started sciebo in Feb 2015. The spectrum ranges from 6.7 percent at the University of Paderborn to 33.9 percent at RWTH Aachen. University size might serve as one explanation, as information should flow very fast at a small campus with a manageable number of departments. As stated by Rogers (2003), diffusion can be seen as a communication process: In smaller and spatially closer populations, communication between the members is much more likely and easier than in a complex university with lots of different departments distributed over the whole city.

Though in theory size suggests itself as a reason for the different diffusion speeds, it does not seem to be a good explanation in our case: Comparing same-sized universities - e.g. the Universities of Münster, Duisburg-Essen and RWTH Aachen with about 44,000 to 49,000 members each (see Fig. 4) - the differences in market share are still evident. Results show a remarkable variance of 24.5 percent between RWTH Aachen (33.9%) and the University of Duisburg-Essen (9.2%), with the University of Münster (23.0%) ranking mid.

Taking all universities into account, RWTH Aachen appears to be an outlier with its high market share. Both, the University of Münster (23.0%) and the University of Duisburg-Essen (9.2%), rank much closer to the overall average. One possible explanation for RWTH’s high performance is that, unlike the Universities of Münster and Duisburg-Essen, RWTH is a technical university with many technophiles. They resemble the innovators described by Rogers and are the first to adopt new technologies. Logically, a technical innovation like sciebo diffuses faster in a technophilic environment than in other populations.

The low performance of the University of Duisburg-Essen, compared with the same-sized universities and the overall average diffusion, is similarly interesting. A closer look reveals that the universities’ commitment in terms of marketing activities might be another decisive factor. While RWTH Aachen
and the University of Münster, in particular, performed a variety of marketing activities (i.e. direct mailings to all members), the University of Duisburg-Essen did not to that extent. Therefore it is likely that only innovators and early adopters who are interested in innovations and actively search for information on their own account for their share of sciebo users. Further monitoring will show if an early majority can be reached with no marketing and just word of mouth, or if the number of users will be stagnating. According to some authors there is a gap between the early adopters and the early majority which has to be bridged by marketing activities (Christensen, 2011; Moore, 2014), while Rogers (2003) considers both groups as a continuum.

Examining the diffusion curves of the different universities (see Fig. 4), deviations from the ideal S-curve of the diffusion model are clearly visible. Usually, they are caused by special events. The first boost in February 2015 is the official launch of the service. In the run-up we realized a large Facebook campaign with posts in over 400 user groups related to the participating universities. Also, test users were now added to the statistics. The second and largest user increase in April, at the start of the summer term in Germany, is triggered by direct mailing, that most participating universities did send to their members. The diffusion curves of those universities passing up this opportunity show no such steep rise. In October most universities welcome their largest share of new students for the winter term, explaining the next boost. In December, some universities used direct mailing to promote an online survey related to sciebo, again gaining attention and an additional boost for new users for sciebo.

As regards storage space, we initially expected 9 GB (30% of the intended per user quota limit of 30 GB) right after registration and a monthly growth of 3 percent (until the quota limit is reached). Currently, the average volume needed by an active user (i.e. a user who uploaded some data) is 3.3 GB, amounting to a total of 99.8 TB storage space used in sciebo. Those universities which grant access to sciebo only to their staff have a substantially higher storage demand per active user (e.g. Düsseldorf
University with 7.4 GB) than most other universities where usually three out of four active users are students (e.g. Münster University with 4.7 GB).

4.2. Data Storage

In Figure 5 we analyzed the storage load on an individual user basis. In particular, we looked at the dependency between the consumed storage space of a single account and its age. Shown is the mean used disk storage for user accounts in dependency of the account lifetime (solid black line), the 0.05-quantile (lower grey line) and the 0.95-quantile (upper grey line) in a logarithmic plot. The broken black line represent the expected and the dotted black line the observed linear model of the user behavior.

Altogether 6,581 user accounts were analyzed on a day-wise basis. The statistical values were computed across an ensemble of user accounts for a specific account age. In addition, two datasets from different time points are independent from each other because accounts were not tracked over time. The analysis was restricted to active accounts with a used storage capacity of more than 10 MB. Thus, inactive accounts from seasonal side effects, such as beginning of a new semester, are excluded from analysis. In addition, a moving average with a window size of 7 days was used to accumulate the number of user accounts for statistical analysis, i.e. in average $N=225\pm92$ accounts were analyzed for each day.

We observed two main findings: First, we predicted that on average an account initially requires 30 percent of its full 30 GB quota and grows in a linear fashion with 3 percent of its quota per month. One can rewrite this assumption to a linear equation of the form $f(x)=A+Bx$ with the function $f$ describing the disk usage in dependency of the time $x$ and the coefficient $A$ as the initial off-set, $B$ as the slope of the function, i.e. we expected $A_{exp}=9000$ [MB] and $B_{exp}=29.6$ [MB/Day]. However, we observed an offset $a_{obs}=2077.1 \pm 239.3$ [MB] and a slope $b_{obs}=33.5 \pm 2.0$ [MB/Day] with a linear Least-Squares Fit ($p<.001$ and adjusted R-Squared 0.578). The observed results show that on average a user
synchronizes less data directly after the subscription than initially expected, but fairly consistent with 30% of the average storage space per user of 8 GB in the pessimistic scenario deduced from the survey findings. However, the growth of the data synchronized is higher than expected.

Second, sciebo has to handle a variety of usage scenarios. In Figure 5 the close distance of the average to the 0.95-quantile indicates a positively skewed underlying distribution, which is caused by an extensive disk usage of some few accounts. This indicates on the one hand that usage scenarios will differ in strong fashion between users and, on the other hand, that sciebo is capable to deal with a wide variety of use cases.

4.3. Bandwidth

The initial estimates of bandwidth requirements were essential to make sure that the internet connection bandwidth of the three university data centers hosting the sciebo platform was not entirely consumed by the new sciebo service. Based on simply models of service utilization (up- and downloads) an overall limit of 3 Gbps sustained for the whole sciebo system, thus approx. 1 Gbps for each of the datacenter sites, was predicted as being sufficient.

One year after the start of operation, this sustained data rate has not been reached by far, but temporary bandwidth peaks at each of the three sites are in the 800 Mbps range (see Fig. 6). With continuous growth of the sciebo user base and storage volume, bandwidth demands will necessarily grow, but negative effects on the internet connectivity of the hosting universities (each currently has a 10 Gbps internet link) are, as initially predicted, not to be expected, especially since traffic policies limiting the bandwidth allocated to individual connection could still be imposed. The mutual data backups between the three sites are schedules in the 12am to 6am timeframe where service utilization is low and thus do not negatively impact the bandwidth budget.

4.4. System Availability

To ensure high system availability for sciebo according to the agreed on availability scores, a set of measures was taken with respect to resilient system design and has proven very effective in the first year of operation. Availability monitoring of the complete sciebo service stack through periodic automated ownCloud file access operations from probes at all three sites checking the respective other two sites with event correlation using NAGIOS and Check_MK show good monthly availability scores even for the initial three months period of operation and for the whole 11 months for which availability data has been recorded up to now (Table 1).

<table>
<thead>
<tr>
<th></th>
<th>Site A</th>
<th>Site B</th>
<th>Site C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mar 2015</td>
<td>98.50</td>
<td>99.98</td>
<td>100.00</td>
</tr>
<tr>
<td>Apr 2015</td>
<td>99.69</td>
<td>100.00</td>
<td>99.93</td>
</tr>
<tr>
<td>May 2015</td>
<td>99.92</td>
<td>99.87</td>
<td>99.21</td>
</tr>
<tr>
<td>Mar 15 - Jan 16</td>
<td>99.64</td>
<td>99.81</td>
<td>99.80</td>
</tr>
</tbody>
</table>

Table 1. Availability scores (in %) for the three sites hosting sciebo (anonymized)

They are well in line with the availability scores previously agreed on amongst the consortium partners of 99.5% per year for each of the sites and a minimum of 98% per month. These results are also comparable with numbers publicly disclosed on Dropbox, where 99.63% to 99.85% availability (where the lower number includes unconfirmed downtimes) for the service was monitored by independent sources for the July to December 2012 period.
4.5. Additional Findings

Apart from those findings related to our predictions, some additional outcomes are worth mentioning. The first finding broaches the issue of user activity: 38 percent of the registered users are inactive, i.e. they have not uploaded any data to sciebo yet. Based on Rogers’ Diffusion Theory (2003), this inactivity of a substantial user fraction could be interpreted as either a prolonged phase of decision making or as discontinuance (without having used the service apart from signing up) (Black, 1983; Parthasarathy & Bhattacherjee, 1998). This finding needs further research.

The second finding focuses on the key collaboration feature of sciebo - sharing data with other sciebo users or externals (share via hyperlink). With an overall average of 2.4 shares per active user, this feature is not used very strongly yet. Folders (66.5%) are shared more often than files (33.5%). Approximately 50 percent of all shares are performed via link (primarily intended for external exchange),
contrary to expectations from the survey (Stieglitz et al., 2014), where 65 percent of the participants intended to share within their university and only 21 percent intended to share with externals.

5. CONCLUSION

These first results show that the predictions on service adoption and system availability made in the design phase of the sciebo service do well conform to the reality of one year of operation. Especially the prognoses on required system platform parameters phrased in the aftermath of the 2013 survey (Stieglitz et al., 2014) are – up to now – in line with the service’s adoption, and, moreover, Rogers’ diffusion theory (Rogers, 2003) has proved to be an adequate model. We could identify two factors influencing the speed of diffusion of the sciebo cloud-service:

1. Share of technophiles in the organization
2. Use of marketing measures

Both findings are supported by the diffusion model. As known from the diffusion literature, an innovation is more likely to be adopted if it is not too complex and consistent with known products. Consequently, technophiles who understand a technical innovation much better and usually find it less complex than other people, will be more likely to adopt an innovation quickly. As noted by some authors, there might be a gap - in terms of missing peer-to-peer connections - between innovators and early adopters on the one hand and the early majority on the other hand, because of the significant differences between those groups (Moore, 2014). Marketing measures like direct mailings, Facebook posts, YouTube videos etc. can bridge this gap by informing the early majority about a new service, and thus speed up the diffusion process. According to our data, organization size does not influence the diffusion speed.

Finally, the universities’ heterogeneous rate of adoption and the high fraction of inactive users leave a wide field for further research. In the upcoming months, analyzing the reasons for discontinuance of use will be a key focus.

6. REFERENCES


7. AUTHORS' BIOGRAPHIES

**Dominik Rudolph** is a research assistant and executive secretary of the Zentrum für Informationsverarbeitung (the computing center) of the University of Münster (Germany). He received his Ph.D. from the University of Münster, where he also studied communication sciences, economics and modern history. His research focuses on the diffusion of media innovations from the user perspective. Related to the sciebo project he is responsible for marketing and evaluation of the service.

**Raimund Vogl** holds a Ph.D. in elementary particle physics from the University of Innsbruck (Austria). After completing his Ph.D. studies in 1995, he joined Innsbruck University Hospital as IT manager for medical image data solutions and moved on to be deputy head of IT. He served as a lecturer in medical informatics at UMIT (Hall, Austria) and as managing director for a medical image data management software company (icoserve, Innsbruck) and for a center of excellence in medical informatics (HITT, Innsbruck). Since 2007 he has been director of the Zentrum für Informationsverarbeitung (the university computing center) of the University of Münster (Germany). His research interests focus on management of complex information systems and information infrastructures.

**Holger Angenent** has studied physics at the Technical University of Ilmenau and the University of Münster (Germany). He worked four years as a research assistant at the institute of Theoretical Physics in Münster. Since 2010 he is a research assistant at the Zentrum für Informationsverarbeitung (the university computing center) of the University of Münster and responsible for high performance computing systems and cloud services. In the sciebo project he is responsible for the technical operations.

**Anne Thoring** is a research assistant for public relations and marketing at the Zentrum für Informationsverarbeitung (the computing center) of the University of Münster (Germany). She graduated from University College London (UK) and the University of Münster with degrees in communication sciences, publishing and strategic communication. Her research focuses on strategies and tools of corporate communications.
Andreas Wilmer is a research assistant at the Zentrum für Informationsverarbeitung (the computing center) of the University of Münster (Germany). He received his Ph.D. from the University of Münster, where he also studied physics and applied mathematics. Since 2014 he is part of the sciebo team where he implements new software tools.

C. Schild is a research assistant at the Zentrum für Informationsverarbeitung (the university computing center) of the University of Münster (Germany) and responsible for authentication systems. He studied physics at University of Münster. He was a leading member of the JOIN project at the university of Münster which participated in several european cooperative projects (some ipv6 task forces, 6WIN, 6NET) to develop and propagate the IPv6 protocol.