Turbulence in the Clouds:
Challenges of Cloud-Based Information Work

Amy Voida, Judith S. Olson & Gary M. Olson
Donald Bren School of Information and Computer Sciences
University of California, Irvine
{amyvoida, jsolson, gary.olson}@uci.edu

ABSTRACT
We report on a qualitative study of the user experience of cloud-based information work. We characterize the information work practices and challenges that exist largely at the different intersections of three constructs—cloud-based services, collaborations, and digital identifiers. We also demonstrate how the misalignment of these three constructs is experienced as a "losing battle" that has led to miscommunication among collaborators, the abandonment of cloud-based services, and the irreparable blurring of digital identities.

Author Keywords
Cloud computing; digital identifiers; information work; information management

ACM Classification Keywords
H.3.5 **[Information Storage and Retrieval]:** Online Information Services—Web-Based Services

INTRODUCTION
> When I try to wrap my head around all of my different virtual documents…. It kind of makes my head hurt to think about it ("Rebecca"/F2[1]).

One participant in this research, "Rebecca," carries out a great deal of her work using cloud-based services. She participates in numerous collaborations, with both family members and work colleagues, using both Google Docs and Dropbox. She also manages four different email accounts, which she tries to use very distinctly to correspond with different audiences: a university email account for official correspondence with students, a departmental email for other work-related communication, a Yahoo email for commercial interactions, and a personal Gmail address for everything else. Nevertheless, information from different collaborations, work-related and otherwise, have all become aggregated under a single account of a cloud-based service. Many work colleagues, for example, have her personal Gmail address just so they can collaborate with her in

Google Docs. But this also means that when she opens Google Docs in work meetings, her personal documents are intermingled with her work documents and the distinctions that she was trying to maintain between audiences using email are blurred:

> I would love to keep my documents and files more separate, but they're just a disaster…. If I try to pull up my Google Docs in front of students or whatever and the top thing may be, I don't know, our mortgage repayment doc where I'm tracking that stuff or whatever because [my husband] happened to be looking at it and now it's a whole thing. So that's kind of frustrating ("Rebecca"/F2).

As an increasing number of activities move into the cloud, this scenario may become even more commonplace. A 2010 Pew/Internet survey of technology experts suggested that by 2020, most people could be expected to have transitioned (if they had not done so already) to working primarily via cloud-based services instead of using local desktop applications [1]. Indeed, work is *already* increasingly being carried out in the cloud, where information is backed up and archived, uploaded so it can be accessed from multiple devices and by multiple collaborators, and transformed synchronously and asynchronously by collaborators using applications that are also hosted in the cloud.

But to talk about information work in "the cloud" is also fundamentally misleading, because there isn't just one cloud. There are many, each owned, generally speaking, by a different company—Google's Docs or Drive, Amazon's Cloud Drive, Apple's iCloud, Microsoft's SkyDrive—many of which also provide the email addresses and other digital identifiers that individuals use to organize different facets of their lives online. So carrying out information work in the cloud isn't simply a matter of overcoming the already significant hurdles of information management and collaboration in an increasingly fragmented space; it is also a matter of figuring out how to work with and around the accounts associated with these cloud-based services— accounts that tie information and collaborative working

---

[1] We refer anonymously to participants in this research. Faculty are denoted by F*n*; graduate students, by G*n*.

relationships to some facet of one's digital identity via digital identifiers.

In this paper, we report on a qualitative study of the user experience of cloud-based information work, examining the following exploratory research question: What challenges emerge as more and more information work is carried out across an ecosystem of cloud-based services? We characterize the information work practices and challenges that exist largely at the different intersections of three constructs—cloud-based services, collaborations, and digital identifiers. Finally, we demonstrate how the misalignment of these three constructs is experienced as a "losing battle" that has led to miscommunication among collaborators, the abandonment of cloud-based services, and the irreparable blurring of digital identities.

RELATED WORK

The User Experience of Cloud Computing

Most research in the area of cloud computing has focused on improving the underlying infrastructure to support a diversity of types of services (e.g., [2]). Various forms of infrastructure, then, provide different constituent services (often owned by companies with different vested interests) that, together, help to define the capabilities, limitations, information use policies, and degree of interoperability among the more familiar, user-facing applications that people typically think of when talking about the cloud (e.g., DropBox or Google Docs).

Marshall and Tang's study of the user experience of this deep and complex stack of cloud-based services suggests that users have constructed a variety of mental models but that these models primarily (if not exclusively) relate to the way that cloud-based services are encapsulated as web-based applications [17]. Marshall and Tang identified five distinct conceptual models that users have of cloud-based services, varying in their degree of sophistication from the relatively simplistic model of a personal cloud repository to the more complex synchronization mechanism supporting replicated collaboration. At best, users who maintain less sophisticated conceptual models are unable to take full advantage of cloud-based services; at worst, these users suffer from misconceptions that lead to breakdowns in collaboration or the loss of information.

Additionally, Marshall and Tang found that a preponderance of their participants used multiple cloud-based services [17]. Our research builds on this finding, in particular, examining the challenges that emerge as people try to juggle the use of multiple services in a growing cloud-based sociotechnical ecosystem.

Information Management in the Cloud

Marshall and Tang's conceptual models suggest that cloud-based services are used for both personal and group information management [17]. Differences in the sophistication of these models influence whether individuals and groups use cloud storage more for file

transfer (e.g., in lieu of using email attachments) or as a more dynamic, synchronized work environment.

Neither the personal nor group information management literature has set its unit of analysis on "the cloud," in a broad sense. Both (overlapping) bodies of research typically explore information management in the context of individual media, some of which are hosted in the cloud—particularly in the group information management literature [8, 16]. One particularly relevant thread of group information management research has explored file sharing, both through public venues like Facebook [9] and through private or enterprise file sharing services (e.g., [3, 19, 20, 21]). One recurring tension exposed by this research has been the extent to which members of groups have different organizational schemes, management strategies, and naming conventions. Researchers have posited different information management 'personality types' that seem to exist (e.g., savers and deleters), helping to explain some of the additional coordination overhead required within collaborative groups (e.g., [3]).

We extend this body of research by examining information work practices across the ecology of systems and services that individuals use when collaborating with a wide range of friends, family, and colleagues. Taking a view that extends across multiple systems and potentially involves multiple digital identifiers also allows us to draw insights about additional information management 'personality types'—*segregators* and *aggregators*.

Digital Identifiers

People have multifaceted identities (e.g., [4, 14]) and take on multiple roles in their lives (e.g., [11]). The multifaceted nature of identity also holds true in the digital world and manifests itself through practices that rely on both *segmentation* and *aggregation*.

In some instances, individuals *segment* their digital identity, creating multiple digital identifiers—strings such as email addresses or usernames that are unique for a particular domain—and use them in ways that align with different roles or facets of their identities (e.g., having both professional and personal email, instant messaging or social media accounts) [12, 13, 23, 24]. This multiplicity of identifiers also supports a more explicitly segmented presentation of self, for example, through the use of a prestigious alumni email address or through the use of an anonymous Twitter account that allows an individual to discuss personal political opinions without attributing those opinions to his professional persona [12, 23].

In other instances, individuals cannot or choose not to segment their identity across multiple digital identifiers, managing a more *aggregated* presentation of self for multiple audiences, such as when posting for both friends and colleagues on Facebook [6, 15, 22] or blogs [7]. When individuals manage a more aggregated presentation of self across diverse audiences, research has found that some

users limit the amount of information that they disclose [6, 15, 22]. Stutzman notes that in the context of social media that relies on peer production, this withholding of content is highly problematic [23]. Somewhat ironically, it is often sites like these that constrain individuals into aggregating their identity online. Farnham and Churchill argue that a required aggregation of digital identity is a "problematic trend" that runs counter both to social science research and to the extended reach of social media, which now pulls from diverse audiences across many sectors of society [9].

Notably, strategies for managing a multifaceted identity through segmentation and aggregation are not wholly distinct and may be better understood as endpoints of a continuum [23]. For example, someone who maintains multiple Twitter accounts to segment tweets topically may publicize her ownership of the multiple accounts and may create explicit links between them by retweeting content.

Most human-centered research related to digital identifiers has focused on their relationship to communication services [12, 13, 23]. This emphasis makes a good deal of sense since a 2006 study affirmed that digital identifiers associated with communication services were valued most because the costs for switching were highest [18]. With the increasing prevalence of cloud services, what happens when identifiers are linked not just to communication services but to information management services, as well—supporting or thwarting both personal information management and collaboration?

METHOD AND RESEARCH CONTEXT
We conducted semi-structured interviews with 24 individuals about their cloud-based information work. Participants included faculty and graduate students at a university in the United States that was in the process of adopting Google Apps for Education. The Google Apps rollout provided a timely inflection point for participants to reflect very concretely on a migration decision that stood to impact not merely their digital identity and communication practices ("Should I migrate my university email to a university-branded Gmail account?") but also their information management practices ("Do I want to have a Google Apps account associated with my university email address?"). We used the occasion strategically to ground our interviews about cloud-based information work and the relationship between digital identifiers and cloud-based information management practices.

At the university where our study took place, most faculty members and graduate students were given the option of migrating their university email accounts to university-branded Gmail[2]. With the migration of their email, these

---

[2] Individuals associated with some schools, such as the School of Medicine, were ineligible due to strict confidentiality requirements that were not supported by the University's contract with Google.

individuals would also gain access to Google Apps for Education—including Docs, Sites, and Talk—from their university Gmail accounts.

Participants
We interviewed 15 faculty members (5 assistant, 5 associate, and 5 full professors) and 9 graduate students. Participants represented 10 of the 14 schools on campus and included 11 females and 13 males.

Participation in the Google Apps rollout was not a prerequisite for participation in the study; in fact, none of the faculty and only 3 of the graduate students interviewed had actually migrated their accounts to the university-branded Gmail account. Instead, most participants opted to manage separate university email and Gmail accounts.

Because we wanted to understand the factors that individuals actively considered when deciding whether or not to migrate their email accounts, we did not interview other populations at the university who had less autonomy in deciding what technologies to use. Many of the undergraduate students at this university, for example, were migrated to university-branded Gmail automatically.

Because we elected to interview members of an academic community, our findings may not generalize across all populations. The autonomy of decision-making and the flexibility in adoption afforded to our participants (in contrast to a prescriptive and/or firewalled corporate environment, for example) may provide a greater number of opportunities to blur boundaries among digital identifiers. Additionally, the challenges to work–life balance often experienced by academics may also lead participants to adopt more extreme measures in negotiating those boundaries. Overall, however, we believe that because of the flexibility that academics have in choosing if and how to juggle various permutations of collaborations, cloud services, and identifiers, our data is likely to show more breadth and/or a superset of practices than we would be likely to see in other populations.

Data Collection
We conducted semi-structured interviews, asking participants whether they had chosen to migrate their university email to university-branded Gmail and to reflect on both the communication and information management implications of that decision. We also asked participants to describe each of the digital identifiers that they used for information management and what they used each identifier for. Finally, we asked participants to describe any cloud-based information management activities in which they engaged and, where applicable, to which digital identifier this information management was tied. In response to themes that emerged during the first few interviews, we extended our interview protocol to ask subsequent participants to describe any strategies they might have for either aggregating or segmenting information and communication across their various accounts. We

interleaved data collection and analysis throughout the research and stopped collecting data once we had achieved data saturation across our participant population.

Data Analysis

We audio recorded and transcribed each interview for inductive analysis [5]. Our first phase of data analysis called our attention to the varied practices of segmentation and aggregation in the management of both identity and information as well as to a range of challenges associated with personal information management and collaboration practices. Our second phase of data analysis helped us to understand the relationship among these themes and the importance of alignment among them.

RESULTS

We identified three primary constructs that shape the user experience of information management in the cloud:

1. Individuals often maintain multiple *digital identifiers*, each associated with a different facet of one's real-world identity;
2. Individuals frequently use multiple different *cloud-based services*, each with different affordances for personal and group information management; and
3. Individuals participate in multiple different *collaborations*, each with different work practices.

The relationships among these constructs are key to understanding information management in the cloud (Figure 1). Different *digital identifiers* are often associated with or are able to access the features and functionality of different *cloud-based services*; different *digital identifiers* are also sometimes associated with different accounts in the same *cloud-based service*. Different *collaborations* frequently make use of different *cloud-based services*. And many individuals selectively associate particular *digital identifiers* with particular *collaborations* in order to help manage the complexity of information work.

In the remainder of this section, we characterize the relationships among these constructs and describe some of the challenges of carrying out information work in the cloud. Throughout, our analysis highlights the significant influence of practices related to both the aggregation and segmentation of both identity and information.

Digital Identifiers + Collaborations

All but one participant in this research segmented his or her digital identity to some degree. Participants described having an average of 5 different email accounts, a number that is in line with averages found in previous research [13], although the degree of segmentation varied from 1 to 10 accounts. A majority of participants also managed at least one social media account, although not all social media accounts (e.g., Google+) were associated with *additional* digital identifiers.

Participants most commonly distinguished among work identifiers for professional information and communication,
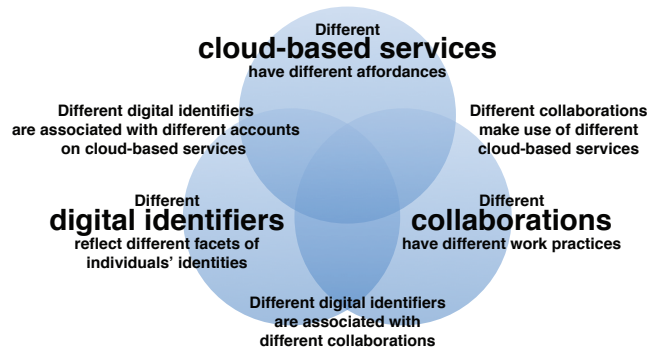

Figure 1. The intersections of three constructs that shape the user experience of information management in the cloud.

personal identifiers for use with friends and family, and another identifier for use with commercial organizations (often referred to as their "spam" account). A majority of participants also managed identifiers that represented one or more historical affiliations with organizations (e.g., a university they attended previously); they maintained these identifiers so that people who knew them during a particular phase of their lives would still be able to get in contact with them and so that they could continue to receive information from organizational mailing lists.

Critically, then, individuals typically segmented their digital identity for different audiences, but at a much coarser granularity than for particular collaborations. Individuals participated in many more collaborations than they had identifiers and typically managed multiple collaborations under the auspices of a single, audience-focused digital identity. Over time, the diversity of these collaborations and the ever-increasing number of files associated with them made information management in the cloud challenging: "Opening my Google Docs now is a little bit upsetting. There's just too much in there" (F2).

Motivations for Aggregating and Segmenting

When participants described practices of aggregation, they often cited the ease of having everything in one place as a primary motivating factor:

> I like to have everything in one place, and that's why I don't have separate accounts for personal and my work…. It's kind of a waste of time for me to switch back and forth, checking emails (G3).

> It just feels a little safer to have everything… in one place where I'm not going to forget to look at it; where I'm not going to be confused in any way; where it's just… all easily accessible (G6).

Participants described a greater diversity of motivations for segmenting their information. One graduate student (G8), for example, managed four different digital identifiers that were each associated with a different Google Docs account; two of those Google Docs accounts were also associated with different university email addresses. She preferred to

segment her information to avoid confusion between similar files created in different contexts:

> I prefer that way because I guess I like to keep things separate, so that it doesn't confuse me. And then different institutions… I want to keep them separate, too, so that a class that I have been taking with a similar name would not confuse me from another institution (G8).

Another participant reported segmenting information in email so that he could apply different information management practices to different accounts:

> My [personal] Gmail account, I don't clean up on a regular basis. It just keeps growing and growing. I delete a few, but I mainly just leave—let stuff accumulate there. But on my [university] one I like to keep it really clean, so I have subfolders, and I put stuff into subfolders, and I like to keep my main inbox as short as possible, so that I'm—I like to be most responsive to work-related email, so I want to know exactly what's going on… (F12).

Other participants reported additional motivations similar to those reported in previous research such as wanting to segment incoming email to minimize work interruptions when at home and vice versa (see also [12, 13]).

Variance in Aggregation and Segmentation Practices
In contrast to previous research, however, individuals' practices did not vary simply along a spectrum between segmentation and aggregation. The same individual sometimes varied his or her segmentation and aggregation practices by identifier, information type, and/or device type as these practices were sometimes fundamentally influenced by the affordances of the cloud.

Most participants aggregated a subset of their identifiers and kept a different subset of identifiers segmented. Participants would often describe two of their identifiers as "primary," indicating a primary work and primary personal email address. Often, these two email addresses were aggregated, either in a single mail reader or by forwarding one to the other. Other identifiers, such as email addresses used for commercial communication and email addresses affiliated with alumni institutions or specific projects were often kept separate.

A few participants varied in their tendency to aggregate or segment based on information type. Email, in a few cases, was managed differently than documents, with incoming email, for example, aggregated across identifiers, while documents were segmented:

> What happens with the mail is that it's a constant flow of information… It doesn't really matter where it's coming from. And because it's a constant feed… the distinction really doesn't make any difference to me. The documents, on the other hand… they don't change that much, and I guess they were a little more manageable. But then, if you mix them up—I

somehow feel it's much more difficult for me to manage the documents, especially differentiating between my work-related documents and my personal documents (G4).

Additionally, participants varied in their information management practices by device type. Most commonly, participants aggregated information for access on laptop and desktop machines while segmenting information on handheld devices. This strategy tended to help participants be more responsive to a subset of the information that they received while mobile:

> When I'm on my laptop I can monitor all of them, but… I really only monitor the [departmental email account] on my mobile…. That's one of the reasons why I'm worried about the segregation is because I want people to send me stuff that's important that I get to within four hours to the right email address that I know to monitor (F3).

The "Losing Battle" of Segmentation
Regardless of whether their practices reflected strategies of aggregation and/or segmentation, participants often characterized segmentation as being a losing battle, explaining that aggregation was sometimes a pragmatic response to the difficulty of trying to segment information and maintain segmentation over time:

> I've surrendered. Given up. There's no sense in trying to separate these worlds…. You know, it's a losing battle. There are just way too many emails to keep up with all of them…. So, I've just kind of surrendered, and I don't—I'm not bothered when things cross the boundary anymore (F9).

> I like to separate work and personal stuff. But the reality is, the moment you go into Dropbox, you can't do that very easily anymore…. Everything that I do is in Dropbox. Every single thing. Right? And it syncs with my home computers and my work computers. And so, all of a sudden I've lost the ability to separate files. But that's the reality of the solution. Now, technically, I guess I could spend a lot of time separating folders out. But it's just easier to have everything in one place (F11).

> Some part of me that has some sense of some ideal would like to separate things. But I clearly don't have a penchant for it. It's more about in the moment of I'm lazy; I'm busy; I don't want to go one extra place; I'm not going to bother with separating things (F1).

One participant had recently created a new work email address in an effort to re-segment her work email from her personal email, but this attempt was deemed mostly unsuccessful and suggests just how difficult the process of re-segmenting digital identity may be:

> So, I had this idea that [setting up an additional work email] was going to help me separate those things a little bit…. Of course, since I've been using my Gmail

address for work stuff for years, I still get all those emails there. And so, now they're in two places instead of one. So, I'm not sure it's done me any good at all…. I could, I guess, set up a different address that I use for personal stuff, because I could probably convince those people to use it. But that seems like a pain in the ass (G2).

Digital Identifiers + Cloud-Based Services

Individuals associated different digital identifiers with different cloud-based services. In some instances, individuals used different digital identifiers to create multiple accounts on the same cloud-based service in order to more explicitly segment their information.

Information Aggregation and the Blurring of Digital Identities
For almost all participants in this study, the primary consideration in deciding what digital identifier should be associated with information in the cloud was the perceived longevity of the digital identifier—not the identifier associated with or known to the most relevant collaborators. Participants wanted to ensure that they would always have access to their information:

And then my other issue with [the university email], that's preventing me from using this before, is that… I'm going to lose [it] when I leave here…. It also makes me not want to use that one for anything where I'm going to be potentially wanting to still communicate with this person after I leave [this university]. Because I don't want them to then not be able to get in touch with me (G2).

I wish in Google Docs I could have everything in one account. And I wish in Google+ I could have everything in one account, too. But the administration of the accounts is different and that makes it a problem because if for some reason I stop paying for my family email account, then I lose everything that's associated with the family account. So it's not just an issue of the information that's associated with the account, it's the administration of that account, too. So if I think about stuff that I want available longer than maybe my employment at the university, I've got to make sure that's in the account that's administered differently (F3).

I guess I trust [this university] is not going to go out of business. At least not for 20 years when I die. So I figure that if I stick with this I'm sticking with something that I know is going to be around. Google appears to be doing really, really well, but I have no idea. And what if they went bankrupt tomorrow? What would happen? (F4).

Whether participants banked on their relationship with the university or another organization providing digital identifiers, the end result was almost always the same: information that would typically be associated with one digital identifier was managed under an account associated with another digital identifier with greater perceived longevity, blurring the distinctions between facets of identity that the identifiers were originally trying to manage. Most graduate students and many pre-tenure faculty members, for example, managed work information under a personal digital identifier because they wanted to ensure that they would still have access to the information when or if they left the university and lost access to their university-based identifier.

A similar blurring was also caused by the adoption of cloud-based services that were owned by companies who provided digital identifiers previously appropriated for use with a different facet of one's identity. Most participants managed multiple email addresses including, for example, a university email address that they used for work-related communication and a Gmail address that they used for personal communication. When those individuals wanted to use Google Apps for work collaborations, however, they typically used the Gmail address that they had only wanted to use for personal reasons. This blurring caused some frustration when both work and personal documents were visible in work meetings, for example.

Blurring across digital identities also occurred when participants received emails at an email address that was primarily used for other purposes (e.g., receiving an email about work at a personal account or vice versa). Problems ranged from the minor sense that receiving the email at the "wrong place" was "jarring" (F5) to miscommunication among colleagues caused by information that was missing from the signature of a work email because the response defaulted to being sent from the personal email address at which it was received (G6). Participants reported additional problems related to the different timescales at which they checked various email accounts and the responsiveness that they wanted to project through some of their digital identifiers:

Sometimes people will send me an email to the wrong email address and I actually notice that. So someone from my startup world will send me an email to my commercial family Gmail and I really noticed that. That sort of… makes me uncomfortable because I feel like in their address book I'm fracturing. And that makes me nervous… because I've had this experience of having four email addresses for someone and not knowing how they've structured their life and so not sure which one they monitor and which one's the one that they never check and that sort of thing. So I get concerned…. I want them to be sending to the semantically correct one so that if I go on vacation and I make sure I'm monitoring the work email, I'm actually monitoring work email (F3).

Identity Segmentation and the Fragmentation of Digital Information
When individuals were able to segment their digital information across accounts associated with different

identifiers, the end result was that their information was fragmented across a greater number of services. In some instances, this fragmentation was viewed positively. One graduate student (G8), for example, explained that her highly fragmented information helped her avoid confusion among similar information generated in different contexts. In other instances, this fragmentation was problematic and resulted, instead, in confusion over which of the numerous accounts contained the information one was looking for and in problems related to redundancy across accounts. One faculty member (F3), for example, described 7 different Gmail addresses that he managed, most of which were also associated with Google Docs accounts and two of which were associated with Google+:

> I log in with both my family [Gmail] and my conference [Gmail] in Google Docs. And that's a mess, too, because I have no way of really thinking about what I do in one and what I do in the other. So I will often find that I can't remember where the Doc is I'm looking for and I have to go look in both of them…. The family Gmail, I do Google+. The conference email, I do Google+. And that's become a mess because now some people are friends on one and not the other and it's outside of the whole circle infrastructure, so that's just a disaster…. There are duplicates. There are overlaps…. And I tried to think "how do I want to manage this?" and I just don't have enough time in my life to manage it. So I was just like I give up on Google+ (F3).

The confusion about what digital identifier is associated with any given piece of information in the cloud may also be caused, in part, by the influence of browser cookies. One graduate student (G2) noted an instance in which a professor shared a Google Doc with her university-branded Gmail address. At the time she clicked on the link in the invitation email, however, she was logged into her personal Gmail account, and the shared document ended up being associated with her personal Google Docs account, instead.

Cloud-Based Services + Collaborations
Information work in the cloud was also influenced by one's collaborations, where external influences over information management were the most significant challenge for many participants.

Cloud-based services differ in the degree to which they integrate into the desktop computing environment, that is, the degree to which they rely on local applications for file editing and management (e.g., Dropbox) versus offering a web-based document-editing experience where both tools and documents are part of the cloud (e.g., Google Docs). These different models of cloud-based services also provide different kinds of visibility about the work of collaborators. Participants found both models to be challenging—one providing too much visibility about collaborations and the other, not enough. Many participants, for example, were frustrated by the web-based working environment provided by Google Docs, in which any update by any collaborator influenced one's view of the virtual filesystem:

> When I open up my Google Docs, it's the exact opposite of what I want when I open a work environment. It's not controlled by me. It's controlled by whoever most recently shared stuff with me…. I want to open my work environment and see what's important and critical for my day—what I've set out to be the work that I want to focus on (F10).

Other participants were equally frustrated by the lack of visibility of collaborators' activities, particularly during synchronous work in Dropbox, which has maintained the metaphor used in the operating system's filesystem—hierarchical files and folders—while adding collaborative capabilities:

> We get a lot of change conflicts on Dropbox when we're editing things at the same time (G2).

Yet, cloud-based services afford a variety of compelling models for collaboration [17], and nearly all participants used at least one form of cloud-based service for collaborative work.

Choosing Cloud-Based Collaboration Technologies
Cloud-based services were not always selected based on the features and functionality of the service. Often, they were selected because a critical mass of collaborators had access to or knew how to use a particular service: "I think it depends on… What do people have already? What are the current stuff that they have access to and they know how to use?" (G1).

If not all collaborators used a particular cloud-based service, had a digital identifier to give them access to a cloud-based service, or if the "wrong" digital identifier was connected to that cloud-based service, then the work of the collaboration could fragment across multiple media. This fragmentation could result in some combination of data redundancy and inconsistency. One participant described a collaborative ride-sharing activity that was coordinated primarily over email until one colleague created a shared Google Doc:

> But then, a bunch of people, because that invite went out to whatever email address people had used to sign up for the retreat, because they just replied to all…. So, a bunch of people were all like, "I can't log in because I don't have—this is not a Gmail account that you invited me with. Blah-blah-blah." So, anyway, [the Google Doc is] partly being used, and it's partly, there are still emails going around (G2).

Aggregating Information Across Multiple Collaborations
Participants—particularly those who aggregated information under a single digital identifier—found themselves, over time, with documents from an intractable number of collaborative activities managed under a single digital identifier. The negative experience of Google Docs' dynamic feed view was amplified when documents

reappeared at the top of the feed from collaborative activities that had long ago been set aside:

> Something that I haven't looked at in two years, suddenly someone else is doing something with it and it pops up to the top and I can't figure things out…. There's too many people doing too many different things (F2).

Additionally, different collaborations had different practices with respect to the pace of document creation and editing, so collaborations that involved larger numbers of documents or more a more bursty pace of editing monopolized the dynamic feed and served to mask work taking pace in other, potentially more relevant collaborations:

> So the undergrads have shared their set of [project] documents with me and they're creating hundreds of documents every day because they're wacky. And it completely blots out everything that I feel like I need to be accountable for…. But it's really just a disaster…. You need some other system to remind yourself that you're working on a document because you can't just turn on Google Docs and be like, oh that's right, I'm working on this because there's just a ton of documents there (F3).

Different collaborators also frequently maintained different naming and organizational schemes for their information. Other researchers have suggested that this group information management challenge is largely due to differing granularities among collaborators' mental models of collaborative activities [25]. When the granularities of those mental models diverge, participants reported frustrations with cloud-based information management. Those frustrations were magnified when participants were trying to manage information across multiple collaborative activities under the auspices of a single digital identifier:

> And Google Docs is a wreck for me. It's kind of a nightmare…. [Students will] send something that's like [conference] paper and I'm like, great, there's five of you sending me a Google Doc called [conference] paper. And Google doesn't have a very nice way of adjusting that for me. And it's not clear who sent it to me (F2).

Digital Identifiers + Cloud-Based Services + Collaborations
The relationship among all three constructs is more complex than that between any two of those constructs. Each construct interacts with the other constructs, and tensions or breakdowns where two constructs intersect seem to increase the likelihood that tensions or breakdowns will emerge in relation to the third construct, as well.

For example, while the facet of one's identity associated with a digital identifier remained relatively constant, the cloud-based services associated with that digital identifier sometimes changed depending on which companies collaborated with or were bought out by the company that provided the digital identifier. But as these cloud-based services came and went, they matched more or less well with the needs of the group of collaborators that participants associated with that facet of their digital identity. When these services, instead, matched with the needs of collaborators that had been associated with another digital identifier, tensions emerged throughout the sociotechnical system.

One concrete instance of this phenomenon was reported by multiple participants who all had digital identifiers provided by Yahoo that were associated with their commercial activities (typically referred to as "spam" identifiers). These participants also reported using cloud-based services like Evite or Evernote and being prompted to register for them using their Yahoo email addresses. These participants reported that their Yahoo "spam" email address ended up being shared with work collaborators in conjunction with work coordinated via either Evite or Evernote. As a result, they all received notifications about work to their "spam" email address and had ended up missing many of those notifications because they were sent to an email address that they never checked:

> And then I have a Yahoo email address that I use for spam. The weird thing about that has been I also use it for things like it's associated with my Evernote account or it's associated with those kinds of things. Which has only become a problem recently because a couple of students have started keeping research notebooks in Evernote and they want to share them with me. And so then I have to be like, okay, well— what they first do is they try to share it with my [departmental] account and then it doesn't link up. And so I have to be like, no, share it with Yahoo…. So then, now this email account that was all spam and crap, sometimes I get real stuff in there, which is kind of annoying because I have to go find it. And I don't really know how to deal with that… (F2).

THE (MIS-)ALIGNMENT OF DIGITAL IDENTIFIERS, CLOUD-BASED SERVICES, AND COLLABORATIONS
When the particular digital identifier that one wants to use with a particular group of collaborators also enables that individual to access the cloud-based services that the collaborators have decided should be used to support their work, then participants generally reported that the collaboration and their information management proceeded smoothly. However, when these constructs do not align, participants reported a variety of breakdowns:

- When participants reported choosing a *cloud-based service* to match the needs of their *collaboration* but not all colleagues had *digital identifiers* that gave them access, the work fragmented across multiple media.
- When participants' *digital identifiers* with access to a particular *cloud-based service* were not the same digital identifiers that they preferred to use with particular *collaborations*, those participants experienced a blurring

of identity across identifiers. This blurring sometimes caused collaborators to send email to participants at the wrong email address, resulting in missed emails and miscommunication; at other times, this blurring caused embarrassment when work colleagues glimpsed personal information stored in cloud-based repositories.

- When *collaborations* successfully connected with the preferred assortment of *digital identifiers*, but choose a *cloud-based service* based on what those identifiers had access to or what individual collaborators had experience with rather than choosing a service based on the match between its affordances and the nature of the collaborative work, then participants experienced editing conflicts or challenges managing files and maintaining an awareness of what work one was accountable for in the collaboration.

Once participants experienced misalignment among digital identifiers, cloud-based services, and/or collaborations, they reported that setting things straight again was often difficult and sometimes intractable. Once information that was most relevant to one facet of a participant's identity was stored in a cloud-based service linked to the digital identifier associated with a different facet of that participant's identity, those accounts became blurred from that point forward. In those cases, some participants reported being unsure about where to find information, while others reported giving up on the accounts or the cloud-based service altogether. Once collaborators had gained access to a participant's "wrong" digital identifier because of the shared use of a particular cloud-based service, from that point forward, those collaborators had multiple email addresses for the participant and were no longer guaranteed to send email to them at the "correct" address, making it difficult for the participant to know that they were always receiving or being appropriately responsive to those collaborators' emails.

Some aggregators expressed a desire to be more segmented in their information management, but felt that it was either too late—the blurring among identities had already happened and it was a losing battle to try and get them back again—or that it would just be too much work to manage additional identifiers. Some aggregators, then, were aggregators because they wanted to be. Other aggregators started off as segmenters, would have preferred to be segmenters, but found the challenges associated with managing multiple identifiers in a world that is moving increasingly into the cloud to be just too intractable.

Participants who had resolved these challenges to their satisfaction generally did so by employing one of two strategies: (1) they segmented digital identifiers, information, and collaborations in such an extreme way that it prevented bleeding in the first place, or (2) they aggregated most of their digital identifiers, information, and collaborations so that they aligned by default.

Extreme segmenters seemed to have resolved these challenges by segmenting more rigidly and/or to an even greater extent. One faculty member (F10), for example, used five different accounts for five different cloud-based services to manage each of his different collaborations. One graduate student (G8) created four different Gmail accounts, some of which were created solely to mirror other university email accounts. Each of these accounts represented a different facet of her identity, and each was associated with a unique Google Docs account that contained a distinct set of files, shared with a distinct set of collaborators.

More extreme aggregators, in contrast, tended to have fewer digital identifiers and fewer frustrations related to the blurring of digital identities. Because they were dealing with a greater diversity of information under the auspices of a given identifier, however, their frustrations more typically related to the management of that diversity of information.

Each potential resolution to the challenges of maintaining alignment among digital identifiers, information, and collaborations has its own tradeoffs. Maintaining a greater number of digital identifiers requires more work managing their separation and more work remembering what information is stored under the accounts associated with each digital identifier. However, this approach requires less work in managing the information stored under the auspices of each identifier and less complicated facework, as one doesn't have to contend with interacting with multiple audiences from a single identifier. Maintaining a more aggregated digital identity requires less work managing the identifiers and less work remembering what information is stored under the account for which particular identifier. However, aggregation does require more work to contend with the proliferation of information all managed under a single identifier.

CONCLUSION

In this study of information management carried out across an ecosystem of cloud-based services, we have made the following contributions:

- Our framework of the alignment and misalignment among collaborations, digital identity, and cloud-based services provides significant explanatory power about the breakdowns in cloud-based information work. We also contribute a descriptive account of specific breakdowns at each intersection of the three constructs that can be used to improve the design of cloud-based services.

- We found that the 'personality types' of *aggregators* and *segregators* applies not just to the management of communication but to information management, as well. Our research adds nuance to the applicability of these 'personality types,' which often do not apply wholesale to any one individual's practices. Individuals often mix and match practices of both aggregation and segmentation and sometimes do so differentially across identifiers, information type and/or device.

Most critically, then, as information management moves increasingly into an ecosystem of cloud-based services, it is imperative that the design of cloud-based services better reflect the highly interdependent relationships between information and identity.

REFERENCES
1. Anderson, J.Q. & Rainie, L. (2010). The future of cloud computing. Online: http://www.pewinternet.org/Reports/2010/The-Future-of-cloud-computing.pdf

2. Armbrust, M., Fox, A., Griffith, R., Joseph, A.D., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I., & Zaharia, M. (2010). A view of cloud computing. *Communications of the ACM,* 53(4), 50–58.

3. Berlin, L.M., Jeffries, R., O'Day, V.L., Paepcke, A., & Wharton, C. (1993). Where did you put it? Issues in the design and use of a group memory. In *Proc. CHI 1993*. New York: ACM Press, pp. 23–30.

4. Brewer, M.B. & Gardner, W. (1996). Who is this "we"? Levels of collective identity and self representations. *Journal of Personality and Social Psychology,* 71(1), 83–93.

5. Corbin, J. & Strauss, A. (2008). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (3rd ed.). Los Angeles: Sage Publications.

6. DiMicco, J.M., & Millen, D.R. (2007). Identity management: Multiple presentations of self in Facebook. In *Proc. Group 2007*. New York: ACM Press, pp. 383–386.

7. Efimova, L. & Grudin, J. (2007). Crossing boundaries: A case study of employee blogging. In *Proc. HICSS 40*. Los Alamitos, CA: IEEE Computer Society Press, pp. 86a.

8. Erickson, T. (2006). From PIM to GIM: Personal information management in group contexts. *Communications of the ACM*, 49(1), 74–75.

9. Farnham, S.D. & Churchill, E.F. (2011). Faceted identity, faceted lives: Social and technical issues with being yourself online. In *Proc. CSCW 2011*. New York: ACM Press, pp. 359–368.

10. Farnham, S.D., Turski, A., & Halai, S. (2012). Docs.com: Social file sharing in Facebook. In *Proc. ICWSM 2012*. Cambridge, MA: AAAI Press.

11. Goffman, E. (1961). *Encounters: Two studies in the sociology of interaction.* Indianapolis, IN: Bobbs-Merrill.

12. Gross, B.M. (2009). Names of our lives. In *Proc. CSE*. Los Alamitos, CA: IEEE Computer Society Press, pp. 747–752.

13. Gross, B.M. & Churchill, E.F. (2007). Addressing constraints: Multiple usernames, task spillage and notions of identity. In *Ext. Abstracts CHI 2007*. New York: ACM Press, pp. 2393–2398.

14. Howard, J. (2000). Social psychology of identities. *Annual Review of Sociology,* 26, 367–393.

15. Lampinen, A., Tamminen, S., & Oulasvirta, A. (2009). All my people right here, right now: Management of group co-presence on a social networking site. In *Proc. GROUP 2009*. New York: ACM Press, pp. 281–290.

16. Lutters, W.G., Ackerman, M.S., & Zhou, X. (2007). Group Information Management. In W. Jones & J. Teevan (Eds.), *Personal Information Management* (pp. 236–248), Seattle: University of Washington Press.

17. Marshall, C. & Tang, J.C. (2012). That syncing feeling: Early user experiences with the cloud. In *Proc. DIS 2012*. New York: ACM Press, pp. 544–553.

18. Mueller, M.L., Park, Y. Lee, J., & Kim, T. (2006). Digital identity: How users value the attributes of online identifiers. *Information Economics and Policy*, 18, 405–422.

19. Muller, M., Millen, D.R., & Feinberg, J. (2010). Patterns of usage in an enterprise file-sharing service: Publicizing, discovering, and telling the news. In *Proc. CHI 2010.* New York: ACM Press, pp. 763–766.

20. Rader, E. (2009). Yours, mine, and (not) ours: Social influences on group information repositories. In *Proc. CHI 2009*. New York: ACM Press, pp. 2095–2098.

21. Shami, N.S., Muller, M., & Millen, D. (2011). Browse and discover: Social file sharing in the enterprise. In *Proc. CSCW 2011*. New York: ACM Press, pp. 295–304.

22. Skeels, M.M., & Grudin, J. (2009). When social networks cross boundaries: A case study of workplace use of Facebook and LinkedIn. In *Proc. GROUP 2009*. New York: ACM Press, pp. 95–104.

23. Stutzman, F. & Hartzog, W. (2012). Boundary regulation in social media. In *Proc. CSCW 2012*. New York: ACM Press, pp. 769–778

24. Voida, A., Newstetter, W.C., & Mynatt, E.D. (2002). When conventions collide: The tensions of instant messaging attributed. In *Proc. CHI 2002*. New York: ACM Press, pp. 187–194.

25. Voida, S., Mynatt, E.D., & MacIntyre, B. (2007). Supporting activity in desktop and ubiquitous computing. In Kaptelinin, V. & Czerwinski, M. (Eds.), *Beyond the Desktop Metaphor: Designing Integrated Digital Work Environments* (pp. 195–222). Cambridge, MA: MIT Press.