Most research libraries typically acquire personal archives from creators at the end of careers, or lives, or even after their death, which effectively means that material can be acquired decades after its creation. Because of technological obsolescence and how fluid our personal uses of technology are, it’s logical to conclude that creators need to become more active in curation activities, and libraries need to engage them much earlier in the creation process, else we risk losing valuable cultural material down the digital black hole.

But I think there are some challenges to compressing this distance between creation and acquisition. For starters, do creators want to be bothered with our curatorial recommendations? In a report on born-digital literary manuscripts, Matthew Kirschenbaum and others concluded, based on interviews with working authors, that suggesting to them certain technological behaviors might impinge on their creative process, and hence be unwelcome advice. In our grant's initial survey less than half of respondents answered affirmatively to the question of whether librarians should offer them personal archiving and digital preservation assistance. Several respondents even explicitly stated their opposition to being told how to preserve their own material in a free text comment field.

Of course, we shouldn't just assume that creators are absent from the preservation process to begin with. For example, 86% of respondents to our initial survey indicated they regularly back up important material. I know redundancy does not constitute preservation, but it was encouraging to see that most of the respondents back up with a frequency falling between “continually and monthly.”

My favorite answer, and the funniest one, to this question about backup frequency was a single word: “episodically”. I actually think this answer could be instructive to archivists and librarians: for purposes of preservation we may be inclined to isolate material into neat little piles based on documentary types or by file type, as discreet items on a file system, or even as atomized chunks of data, but creators themselves may instead view them as events in a continuum of research – they progress from one to the other and together form a whole, a
whole that is not entirely separable from other aspects of their appointments, including teaching, dissertation advising, etc.

And it's not entirely separable from the physical spaces in which they work either. In Penn State Special Collections we have a physical reconstruction of the author John O'Hara’s study – his writing desk, typewriter, books shelves, even his fireplace and mantle. Now... projects like the Rushdie one out of Emory are starting to make us think of the author's study as a virtual place, his reconstructed computer desktop screen replacing the physical reconstruction of a creative space. But in our first in-person interview with a Penn State faculty member in the Communications department, we got an intimate view into how these spaces are really wrapped up with one another in complex ways we do not yet fully understand. This researcher commented on the ways in which his forty-plus years in the same office and his pending move have had an impact on the way he works. At a much higher level, being with one institution for so long influenced his relationship to technology over the years. In the early 80s he learned of a mainframe on campus and successfully petitioned to have terminal access to it, which he used to draft a book in the SCRIPT markup language. As we proceed with the interviews, we’ll be documenting both the physical and virtual desktop environments graphically, and I think this is an area of the grant work that can be especially enlightening and fruitful.

In another Mellon funded digital archives grant, the AIMS project, conducted by Stanford, Yale, UVa, and Hull, they concluded that institutions need to spend more pre-acquisition time collaborating with potential donors and getting to know the details of their digital ecologies in order to make informed decisions about ingest, management, and preservation. The AIMS project produced a comprehensive model donor survey that can and should be used by archivists, and one that helped guide some of my recommendations to Ellysa on this grant. It contains questions about general computing habits, tools, use of mobile devices, social networking and even privacy. I think if we could always get this level of detail from potential donors, then a grant like ours wouldn’t be necessary. But donors of personal papers are busy, distracted, outgoing people who lead complicated and active lives just like the rest of us. In my own very limited experience with donors, I have not always had great luck teasing the most basic responses out of them, and I think it will be hard to have the kind of in-depth discussions about their digital desktops that we as archivists desire.
With this in mind, I think what we need are more case studies about different classes of creators to draw on for insight. There are two common methods for deciding what to acquire in our profession – one suggests that we appraise a collection based on the value of the content, and another (called macro-appraisal) suggests an emphasis on the role played by the records creator and how that person fits within an established and documented institutional collecting area. The latter has become increasingly common, but it presents significant challenges when the material to be acquired is digital, distributed, duplicated (in both digital formats and analog), and dependent on specific hardware or software.

In the absence of a collaborative relationship with potential donors, a condition which will likely persist, I think archives and libraries need a community-produced set of data about different types of creators to fall back on. This kind of appraisal might fall somewhere between micro and macro. Let’s call it meso-appraisal: the idea being that we start to document, quantitatively, the tools and habits used and exhibited by creators working in different fields. We've seen this kind of work already coalescing around literary archives. And in a way, I think this approach might succeed in aligning our appraisal efforts with our preservation efforts, which focus on developing preservation strategies at an intermediate level, usually for different formats. Not individual items, and not whole collections of disparate items.

For instance, heading into this grant I was curious to see how many unusual or non-standard file formats we might encounter. With the exception of data sets, the survey results have showed a surprising uniformity in the common types of digital material people have. There is persistent use of documentary types familiar to us all: word docs, spreadsheets, email, image files, and already some indication that certain formats, like PDF, are rather ubiquitous. And only 5% of respondents checked the "other" option on the question related to formats.

I think we can also produce some useful information about how people are using certain tools. Email is already turning out to be an interesting case study. People use it for everything, and not just communication. They use it for sharing documents, obviously, but also rudimentary backup, and even version control.

I think for digital library professionals, some of these findings can help inform our approach to the development of tools, especially related to repository services,
which is something I've been thinking about a lot since Penn State just released its Hydra based repository, ScholarSphere.

If there's anything our initial findings are demonstrating, it's that researchers on campus are not shy about going outside the academy to get the tools they need. The communications professor we interviewed uses Dropbox because it's easier, and provides more space than his network space at the university, even though it adds a financial burden. He uses Gmail because it's more user friendly than the email system his department offers. In general, researchers appear to use such tools interchangeably for a variety of purposes, including sharing, versioning, and redundancy. But they don't know where to go for things like data recovery, and have simply accepted certain levels of data loss as part of doing business. I think the data is already trending in a way that demands that our repository efforts reduce as many barriers to adoption as possible -- that our repository be as easy to use as Dropbox or email, or even be interoperable with these tools. Most important is that any repository services we develop fall comfortably within their existing workflows. And if we can at the same time sell the benefits that they cannot get from other tools -- greater accessibility, data integrity, etc. -- then I think we'll be able to spur much greater adoption.