

## Lightning rounds: Challenges facing preservation of born-digital news applications

Katherine Boss, Meredith Broussard and Eva Reveal, New York University | Oct. 14, 2016 | Charles E. Young Research Library, UCLA | Dodging the Memory Hole 2016: Saving online news

---

**KATY BOSS: [00:08]** Hi everybody. You guys probably remember me from yesterday, but I'll do quick introductions again. I'm Katy Boss. I'm at the New York University Libraries, librarian for journalism, media culture and communication. My two co-authors on this project are Eva Reveal and Meredith Broussard, who you guys heard from earlier in the news apps panel; both are at the Arthur L. Carter Journalism Institute. And, we're talking about news applications and the challenges around archiving, preserving, and — also very important for libraries — making these things discoverable for our users.

**BOSS: [00:47]** I think I'm preaching to the choir a little bit here, but, who is creating these things and why would we care to save them? Data journalism projects and news app creation is exploding. And these projects represent some of the most creative, innovative projects. As a subject selector in libraries, I can confidently say that these are things we need to save. If we're going to save anything, we need to save these. So, what are some of the issues facing this?

**BOSS: [01:21]** We also heard about this a little bit on the panel, but, one of the fundamental issues is that these are not static, digital objects. They are dynamic, digital objects. So, there is currently no way to save these things. The Internet Archive can't do it; nobody's figured out how to do it. And yet, we have years of these projects that are not being saved anywhere except on random hard drives and servers. We have to figure out a solution. One of the big problems related to this is something that's known as "dependency hell." These things are built on complex software stacks and every aspect of these things has dependencies and libraries that they need to run. So, if any one of these parts in the engine that runs these news applications is missing, it's not going to work.

**BOSS: [02:18]** A lot of the talks that we've heard from today are everyone scrambling to migrate things to a more current form. Migrating things from microform to PDFs. The New York Times having to migrate things from different forms, get it all in one form so that it's easier to batch-process things later — we're all doing migration. The digital archiving community has come to one consensus on dynamic digital objects like video games, or news applications; things that are interactive, that rely on a database. And that is that migration is not really a solution because there are too many things to migrate. You start thinking about it, all of these different software tools and libraries and frameworks: Django, Rails, etc. Ben was mentioning this. They go on and on. Are we going to migrate all of those? The digital archiving community has said, "That's crazy, and it's not going to work." It's already not working with analog objects. So, emulation has been reached as the consensus. Instead of trying to migrate these things, we need to emulate them.

**BOSS: [03:34]** How are we going to emulate things? This is the research that we're working on right now. How can we create a virtual machine that we could use to preserve this, and then emulate the entire environment on a computer, or whatever it's going to be 100 years from now. Long-term preservation. We're not just looking at 10 years, we're looking at 100 years. So, the first step to answering this question — and I'm going to try and really speed through so Eva can talk — is, that we needed to figure out, "How are these news apps being built?" In order to figure out a tool that we can properly use to save them, we need to really be able to describe, "What are they?" So, we started a survey tool and Eva's going to talk about that for one minute.

**EVA REVEAR: [04:19]** Hi. So, I've been working on gathering the data that we've asked for in the survey. We have data on about 50 apps regarding the type of software frameworks they're built with, who's building

them, where they're stored. We've been able to gather some initial findings, such as the most popular frameworks, which, I think, were already mentioned like Django and Rails. We've also been able to see the different processes by which they are stored and served, primarily S3, like plot file-type storage. We've also been able to look a lot at the legal framework and copyright. There's a lot of open source, but also a lot of proprietary.

**BOSS:** [\[05:06\]](#) We wanted to do a shout out to Ben Welsh because your django-bakery is apparently being used by everyone to build these things, that's what we're discovering. So, this is a group effort, saving these things. We need everybody involved and on board. Thank you so much for your time.