SPECIAL FEATURE

Zoe LeBlanc
Andrew Janco
Alex Wermer-Colan
Quinn Dombrowsi
Ann Kijas
Sebastian Majstorovic
Dena Strong
Erica Peaslee

*Saving Ukrainian
Cultural Heritage Online*

A Conversation with the Organizers of Saving Ukrainian Cultural Heritage Online (SUCHO)

After the Russian invasion of Ukraine this February, a global community of volunteers has endeavored to help preserve Ukraine's online cultural heritage. While this community comprises over thirteen hundred volunteers, many of them work as librarians or in cultural preservation, including Quinn Dombrowski and Anna Kijas, two leaders of this group. Dombrowski and Kijas, along with Sebastian Majstorovic, have been instrumental in coordinating this community of experts across time zones and also spearheading what the *Washington Post* described as "a lifeline for cultural officials in Ukraine" (Verma 2022). To capture both their experiences, as well as how librarianship has informed SUCHO, we convened a roundtable with the organizers, as well as two active volunteers—Dena Strong and Erica Peaslee—who also work in galleries, libraries, archives, and museums (GLAM).

While this conversation represents only a brief window into the labor and work of SUCHO, we believe it will be particularly relevant for librarians; in particular, this work offers ways to fight against the destruction of libraries, museums, and galleries in Ukraine. Web archiving efforts cannot bring back the people who sought shelter in the Mariupol theater or restore burning churches from afar. But it is possible to preserve digital images and collections before servers go offline.

To help demystify how one even gets involved in a project like SUCHO, we asked our five panelists a series of questions over Zoom in April 2022. Each of our panelists joined the project at different times. Quinn Dombrowski, Academic Technology Specialist in the Division of Literatures, Cultures, and Languages, and in the Library, at Stanford University, and Anna Kijas, Head of Lilly Music Library at Tufts University. Anna organized an initial archiving event focused on music collections and joined forces with Quinn and their efforts to support displaced Ukrainian scholars.

Figure 1. Rodzik, Olga. Odessa Opera and Ballet Theater. Digital Illustration. Source: Olga Rodzik.

The effort quickly expanded as Sebastian Majstorovic introduced Browsertrix, a brand-new web archiving tool by Ilya Kreymer. Unlike the Internet Archive's Wayback Machine, which stores static images of web pages, Browsertrix stores a complete and interactive version of a site. This new tool initially required specialized knowledge and a virtualization tool called Docker, but the introduction of Browsertrix Cloud enables any SUCHO volunteer to archive complex sites without additional training.

Dena Strong joined the project early on and has been the heart and soul of communications in this online community, helping onboard new volunteers to the wide range of technologies and workflows. Erica Peaslee found her way to the group in hopes of helping with running Browsertrix, but soon combined her experience in museums and emergency management and began coordination monitoring efforts and prioritizing projects in response to events on the ground.

The Discussion

To begin our roundtable, we asked the panelists how they would describe or define SUCHO.

DENA STRONG: I would describe it as a worldwide all-volunteer data rescue and backup team, trying to preserve the things that are already digital in hopes that they can be useful for reconstruction.
SEBASTIAN MAJSTOROVIC: I've seen it in several articles, now that people have used this phrase that we are kind of a "global group of twenty-first-century digital monuments [men]."
DENA: If you give me just two words, Digital Dunkirk.

While SUCHO started in late February, it quickly grew at an exponential rate. We asked the organizers what had changed since those initial first days.

DENA: The first two weeks things were changing drastically enough that I was redoing pretty significant parts of the training. I would learn something one day, and I would be teaching it the next day. And then the day after that it would be changing again; from Browsertrix to Browsertrix Cloud, going from the DSpace Python scripts to Colab[1].

It's been fantastic. We are building this stuff on the fly as we go, like the Wallace and Gromit thing, where the little dog is putting out the train tracks a split second before the train gets there. That's us.
QUINN DOMBROWSKI: Yeah, the first four or five weeks it felt like this completely exhausting whirlwind. I feel like I don't remember March. Right? March must have happened, but somehow we went from "oh sh**, a war" to early April and things in the meantime... yeah, it was just a complete blur.

I'm fortunate enough that my job has been really supportive. I did literally nothing else for an entire month. After about four to five weeks the work hit a rhythm and I wasn't completely exhausted at the end of every day.
ANNA E KIJAS: Yeah, it's like iteration on steroids. We're just iterating, iterating, iterating, and refining as we go. It was really intense, and I mean there was a

> « ...We are kind of a "global group of twenty-first-century digital monuments [men]." »

1 Browsertrix is a tool developed by Ilya Kreymer that allows you to programmatically "crawl" a website and capture webpages. Browsertrix Cloud is a version of the tool that works in the cloud. DSpace is a popular software for cultural heritage websites that present certain challenges for crawlers like Browsterix, which is why the SUCHO team developed customized scripts, which can now be run in Google Colab notebooks. For more on these tools and workflows, see the SUCHO Tech site: https://www.sucho.org/tech.

need for it, because we were just figuring all this out, we were putting together tutorials and documentation and training and so forth, and doing the work.

DENA: I think on the process front, on day one I was like "I don't know what's going on here," I gotta learn what's going on here so that I can then write down what's going on here.

So I was making sure to build signposts, you know, it was not "do things this way," it was "look at this editable document for what the latest information is, there's one of those pins at the top of each [Slack] channel." Or you know "here is the pattern that I'm establishing, for how people can document the troubleshooting in this world-editable Google Doc."

Because so many people will come into the channel and say "Hi, I'm lost and I don't have the time to reread the past two months' worth of Slack so here's my question." And if they've got a table of contents sitting up there in a one-click Google Doc, it becomes a lot easier for people to make references than to attempt to scroll through that channel, and then the other channel two channels over, and then the other channel to find where the answers were.

QUINN: It was a really interesting shift going from the original three co-organizers to sort of working with sub-admins like Dena.

Everything that we did in the first round of things, we did what we would do for ourselves or people who think the way that we think. That works for a certain subset of the volunteers, but not everybody, so by broadening the pool, we had more people step up and write their own documentation and document workflows in different ways.

I think that ended up making it much more accessible to a bigger group of people than if it was just the three of us trying to run the whole show.

While SUCHO is an exceptional initiative, it was also not the first attempt at preserving cultural heritage in the face of crisis. In 2016, Data Rescue was an active community that responded to fears that the Trump Administration would delete or neglect federal data related to climate change (Wiggin 2017). Anna Kijas has been very involved with Data Rescue, so we asked her and the rest of the team, how that model informed their work and also how SUCHO's more spontaneous and unstructured model of collaboration differs from these earlier efforts.

QUINN: I remember looking at the giant spreadsheet from Data Rescue that like you guys had cloned and thinking to myself "yeah, that's not gonna work here." It was so library, it was like...

ANNA: The most library of.

QUINN: ...library spreadsheets I had ever seen, and bless them, obviously it worked for that and made sense for that project.

But doing everything, everywhere, all at the same time, across time zones, with websites going up and down, you cannot be that methodical. Now granted, we have created our own completely ridiculous spreadsheet. But it is a monster of our own creation and responds to our own needs and the things that we have to do in this situation.

There is precedent for this kind of work with Data Rescue, but there's a huge difference when you have two months to methodically archive things in a thoughtful manner across servers that you can more or less count on being up for the whole time. You can go about that systematically. But when you're in a war situation and things are changing day-to-day, or hour-to-hour, and websites could go down at any time, you basically have to embrace some of the chaos.

" But when you're in a war situation and things are changing day-to-day, or hour-to-hour, and websites could go down at any time, you basically have to embrace some of the chaos. "

Otherwise, you'll just get stuck.

DENA: For my two cents, I told people regularly, "what I'm doing here I would never in a million years do in an enterprise production." At the same time, I can't go in and clean up the data and retroactively change things, because people have been trained on stuff we recorded at the beginning of March. That needs to bear enough resemblance to what they see today because we haven't been able to redo the training. I would love to clean up the workflow. I would love to fold things together and standardize things. I can't do that right now because there are so many people depending on the thing being the way it is and I've got to be very careful with even, you know, phrasing changes to the status columns. Fifteen hundred people out there are not going to re-watch the training. You know, because they have already got the thing they're doing.

SEBASTIAN: Yeah, I think it's important to emphasize that this is an emergency response.

One thing I will credit us with is creating spaces where things can happen with a certain free flow. We never could have imagined that. We weren't even sure if we wanted to make a Slack organization. But that was one of these little decisions. Should we open up a channel for that? Ok. And then things start happening in these channels and people start doing things with the spreadsheet that I didn't know was even possible.

I think that maybe a combination of humanitarian crisis and new digital spaces/tools contributed to us working in such a creative manner.

DENA: I also love how people's standards of success have continued to escalate as we've gotten more and more tools online. People are like, "I don't know if I can get all five hundred thousand pages. What happens if my five hundred thousand-page Browsertrix Cloud runs out?" I'm like, "that's five hundred thousand pages more than they would have had otherwise!" Put in a line for somebody else to redo it more extensively later if we have the time, but keep moving.

Given the unprecedented scale of SUCHO, we asked the organizers what they thought was unique about their processes and operations that enabled volunteers to collaborate, to teach, and learn from one another. When a volunteer joins SUCHO, their first task is joining the Slack team, which to date has 31 channels and 1,371 members. Volunteers are encouraged to introduce themselves and then depending on their skillsets are directed to certain channels. We wondered how much this workflow was by design or an emergent property since it has helped many volunteers find their niche quickly and feel confident contributing to the project.

DENA: I would love to say it was intentional, but I have a feeling it's a happy accident. I don't know how you build that from scratch on a scale of fifteen hundred people, aside from being in the chat channels as regularly as you can.

And I feel bad about how much I've not been able to be there lately. I can't do twelve hours a day when I've also got my day job and my life. But when I could, for a couple of weeks in there, it was really helpful to be able to be the cheerleader scanning to see when a person was confused, and saying "hey if you don't like this, maybe this other thing would work better."

ANNA: I agree with you that it was somewhat of a "happy accident" that this turned into a community of people able to find their own niche and excel.

When we started this work, our goal was to create a community that would be able to do this work. We would facilitate the work, but not necessarily be, you know, responsible for all aspects. We wanted people to take the lead. And

they are. All across the different tasks and channels, they're coming up with brilliant ideas and brilliant solutions.

It's just this amazing, really great energy that exists across everyone here. People are focused because we have this shared goal. We're trying to archive and save all this content. So I think it's been very motivating for people.

You know, I think most of us are working regular jobs or have other responsibilities, in addition to this, but we're still so eager to jump into this and help in any way that we can. It's just really been amazing to see how people come together.

Another challenge beyond simply coordinating all these volunteers is also the fact that this type of initiative might attract bad actors, from trolls to more organized efforts. We were curious about how librarianship and information sciences informed SUCHO's approach to user experience and data privacy for volunteers, and how much they had to refine their processes along the way. *DENA*: That is one of the things that changed over the course of March because on day one I didn't perceive that being as much of a consideration as it was two to three weeks later when we started locking down certain channels and separating out certain websites.

Operational security things, like to not talk about where these things have been moved to. That's one of the places I think a diversity of perspectives is really, really handy. I used to work for a start-up company that was doing Department of Defense secure operating systems, but that was securing the bits, not securing the physical objects.

QUINN: We tried really hard in the beginning, and for a long time, to stay out of the physical world. We really didn't want to touch anything but websites. When you're just dealing with websites, there are fewer security concerns.
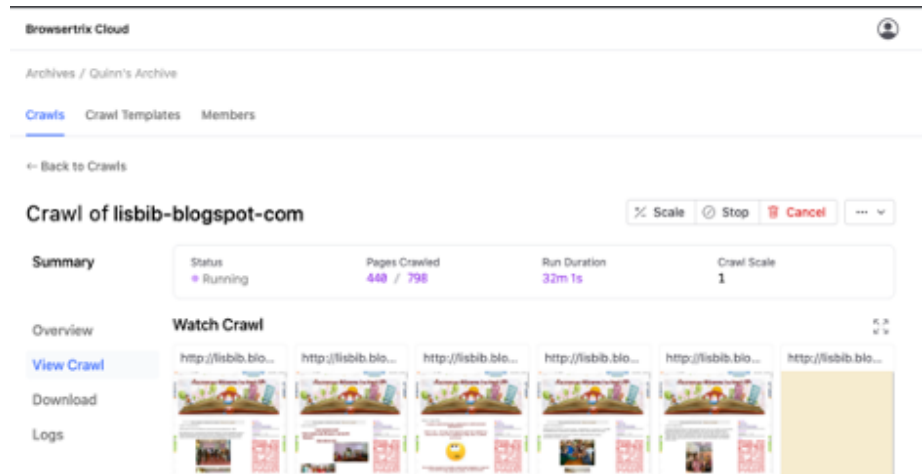

Figure 1. Screenshot of Browsertrix Cloud

But it turns out that it is actually impossible to cleanly divide the digital from the physical because people need digitization equipment and, for a while, we were trying to pass it off to other people. But it eventually became clear that we couldn't sustainably maintain those boundaries. Things started getting porous, and as the physical world dimension crept in, then complications followed. *DENA:* There is a very sizable portion of our volunteer base who would never have been able to install Browsertrix on their own computer. For them, Browsertrix Cloud is transformational. I'm one of them. I could never have installed Docker on my one and only machine, when Chris Nelson told me he turned three machines into bricks trying to get Docker running on Windows

machines.[2] I'm like "yep I cannot do this," this is my window on the world and I can't roll those dice. But with Browsertrix Cloud I can do it. Even a six-year-old can do it and I love it!

QUINN: So Sam [Quinn's eight-year old] woke up at 6:18 this morning with a nightmare, and crawled into my bed. What he wanted to do to cheer himself up was use Browsertrix Cloud to archive some Ukrainian websites. I had to break it to him that we can't do that right now because we're migrating the infrastructure, so instead we looked at Ukrainian memes.

But, I mean, being able to run a web archiving event at an elementary school is phenomenal. It's just kind of hard to imagine SUCHO without that.

While we were fascinated to learn about the day-to-day operations of SUCHO, we were also curious about the lessons learned so far and what the organizers would hope to share with other similar efforts in the future.

SEBASTIAN: Well, we are in talks with different people, associations, and organizations to transfer the lessons that we are learning in SUCHO to develop preventive infrastructure and web archiving.

This needs to happen before a crisis of any kind, not just wars, but natural disasters as well, like floods or the recent fire of the national museum in Brazil[3]. A couple of years ago in my hometown, in Cologne, I saw the city archive collapse in 2009 because of construction of a subway tunnel. Cultural heritage can always be lost. We are talking to Europeana and some other places, and the goal is to try to implement mutual web archiving so that we don't need to do SUCHO.

But beyond that, for me personally, I've been very interested in the last couple of years in how to bring people together online. I've done that only in a really small kind of framework. Where would you meet online to have a group discussion, a simple group discussion? There's really only Facebook groups. There's no other place where you can go and just have a group, not even a Slack with channels and so on, which is also a commercial product, by the way. So I'm really thinking, apart from the cultural heritage and archiving and so on, I'm really trying to think about how to bring tools together as a sort of platform for activism, something that you can replicate.

We ended our discussion asking the panelists if they had any advice to offer to librarians' reading that are hoping to get involved in doing this kind of work, even if they don't have any prior experience with web archiving or humanitarian crises.

QUINN: People aren't going to prepare for a war in advance. Maybe what's more realistic is to go find out who does web archiving at your library. Go find out who, if anyone, does social media archiving. Figure out where that happens in your organization and go get coffee with those people. Because that way, when the day comes that you need to do web archiving in a hurry, you won't be cold-emailing them. Go make some friends over in that part of the library if that's not a part that you already hang out in, and those connections will pay off.

ERICA: Talk to someone to make an emergency plan. Larger museums have them, some larger libraries probably have them as well. That even includes

---

2 For those unfamiliar, 'bricking a machine' is a colloquialism that refers to a software or hardware update that breaks a machine, essentially turning it into a brick.

3 In 2018, a fire gutted the National Museum in Rio de Janeiro, and more recently in 2020, another fire destroyed the Natural History Museum and Botanical Garden at the Federal University of Minas Gerais (UFMG) in Belo Horizonte.

what to do during a fire drill, like there's that one thing you've got to get out, there's going to be a list of what to do.

What we do in emergency management is what we call "exercises." It could be anything from a tabletop exercise like they implement in cybersecurity, to actual exercises where you get up, you go do what you need to do in the event of an emergency, and if you have to walk something out of the building, you actually pick it up and you walk it outside of the building.

If you're in an institution that doesn't advocate for that, I know the Smithsonian offers a model, and there are a lot of companies and nonprofits like Blue Shield and UNESCO. They all have guidelines, and American Alliance of Museums has kind of like "inspirational guidelines" that you can use to implement an emergency plan if you don't have the budget to go out and hire someone who does it for a living.

While these answers are only selections from our roundtable, we believe they provide a crucial window into SUCHO and thank each of our panelists for their time and thoughts. To learn more about SUCHO, you can visit their website https://www.sucho.org/ and the group also has plans to publish a monograph on their experiences in the upcoming year. Most importantly though, we hope that this roundtable has underscored the importance of collective action and librarianship in times of crisis, and also encouraged librarians to advocate for this sort of preventive and emergency web archiving as part of your job description and/or library services.

References

Verma, Pranshu. 2022. "Meet the 1,300 librarians racing to back up Ukraine's digital archives." Washington Post. April 8, 2022. https://www.washingtonpost.com/technology/2022/04/08/ukraine-digital-history/

Wiggin, Bethany. 2017. "Data Rescue Philly Builds Data Refuge | Penn Program in Environmental Humanities." Penn Program in Environmental Humanities. January 15, 2017. https://ppeh.sas.upenn.edu/field-notes/data-rescue-philly-builds-data-refuge.