This document is intended to be a faithful synthesis of the presentations and discussions that took place at *Engaging the Public: Best Practices for Crowdsourcing Across the Disciplines*, a workshop held by the Crowdsourcing Consortium on May 6-8, 2015. It is meant to serve as a resource for those who attended, for the Consortium, for funding sources, and for the field at large. It does not necessarily reflect the views of the Consortium, of individual meeting participants, or of the organizations they represent.

Participant comments may be paraphrased and the sequence of remarks reorganized. These are not exact quotes, rather they are an attempt to capture the content and meaning of the ideas presented.

#crowdcon
@crowdconsortium
crowdconsortium.org

NATIONAL ENDOWMENT FOR THE
Humanities

INSTITUTE *of*
Museum and **Library**
SERVICES

Alfred P. Sloan
FOUNDATION

About this workshop . . .

This workshop was sponsored by Crowdsourcing Consortium, a new national organization. The aim of the workshop was to culminate and then broaden the conversation begun in the regional meetings and webinars taking place through the auspices of Dartmouth's 2014 IMLS-funded *National Forum in Crowdsourcing for Libraries and Archives: Creating a Crowdsourcing Consortium* (CCLA). Through this capstone event, the intent was to consolidate the earlier work of CCLA and advance a national agenda.

CCLA aims to support crowdsourcing efforts among diverse institutions and research communities and to forge a collective consortium. Throughout the workshop, the central concern was on the question of how institutions might best adopt and employ crowdsourcing strategies for increased public engagement, integrating data into existing collections and increasing knowledge in the humanities and related domains. Support for the workshop from three different funders, each with their own distinct communities to bring into the conversation, helped to ensure a rich cross-disciplinary dialogue, sent a very public signal about the importance of these emerging practices, and increased the overall impact of the workshop.

Cover photo: Project working groups in action
Left: Active tweeting throughout the conference

Participants

Nick Adams
University of California, Berkeley

Zaven Arzoumanian
Wild Me Conservation

Liz Barry
Public Lab

Jeff Bigham
Human-Computer Interaction Institute,
Carnegie Mellon University

Tom Blake
Boston Public Library

Brett Bobley
National Endowment for the Humanities (NEH)

Ben Brumfield
Collaborative Manuscript Transcription,
FromThePage

Eva Caldera
National Endowment for the Humanities (NEH)

Peter Carini
Dartmouth College

Darlene Cavalier
SciStarter

Perry Collins
National Endowment for the Humanities (NEH)

Jeremy Dean
Hypothes.is

Katherine Doyle
Pulitzer Center

Meghan Ferriter
Smithsonian Transcription Center

Mary Flanagan
Tiltfactor, Dartmouth College

Neil Fraistat
Maryland Institute for Technology in the
Humanities (MITH), University of Maryland

Rachel Frick
Digital Public Library of America

Michael Haley Goldman
United States Holocaust Museum

Ashwin Gopi
New York University

Jen Hammock
Encyclopedia of Life & Smithsonian Institution

Bob Horton
Smithsonian Institution,
National Museum of American History

Kirk Jalbert
Rensselaer Polytechnic Institute

Katie King
University of Washington, Seattle

Edith Law
University of Waterloo

Sharon Leon
George Mason University

Chris Lintott
Zooniverse, Oxford University

Liz MacDonald
NASA, Aurorasaurus

Peter Mangiafico
Stanford University

Project Co-Directors

Mary Flanagan,
Sherman Fairchild Distinguished Professor in
Digital Humanities and Director of Tiltfactor Lab
at Dartmouth College

Neil Fraistat
Professor of English and Director of the Maryland
Institute for Technology in the Humanities (MITH),
University of Maryland

Andrea Wiggins
Assistant Professor in the College of Information
Studies, University of Maryland

> **Andrea Wiggins** @AndreaWiggins · May 6
> And we're off! #crowdcon is underway with an all-star group of participants from more disciplines than you can shake a digital stick at!
> ↩ ⟲ 1 ★ 4 ...
>
> **Courtney Young** @librarycourtney · May 6
> First time I've received a name tag with my Twitter handle on it. Pretty cool. #crowdcon
>
> crowd
> CONSORTIUM
> **Courtney Young**
> Pennsylvania State University,
> ↩ ⟲ 1 ★ 8 ... View photo

Facilitation

Jake Dunagan • Matthew Manos • Sheena Yoon
verynice.co

Documentation

Catherine McEver
The Bureau of Common Sense

Pre-conference prepping and mingling

Austin Mast
Florida State University, iDigBio

Dominic McDevitt-Parks
Wikipedia,
National Archives and Records Administration

David Miller
National Institutes of Health (NIH)

Ben Miller
Georgia State University

Hector Mongi
University of Dodoma

Carsten Oesterlund
Syracuse University

Tim Olsen
Gonzaga University

Trevor Owens
Institute of Museum and Library Services (IMLS)

Amy Patterson
Wisconsin Technical College

Tina Phillips
Cornell Lab of Ornithology, Cornell University

Lieke Ploeger
OpenGLAM

Daniel Powell
King's College London (University of Victoria)

Jenny Preece
University of Maryland

Trish Rose-Sandler
Missouri Botanical Garden,
Biodiversity Heritage Library (BHL)

Alexis Rossi
Internet Archive

Lacy Schutz
Museum of the City of New York

Sara Sikes
Massachusetts Historical Society

Lauren Tilton
Yale University

Victoria Van Hyning
Zooniverse, Oxford University

Ben Vershbow
NYPL Labs, New York Public Library

Raffaele Viglianti
Maryland Institute for Technology in the
Humanities (MITH), University of Maryland

Amanda Visconti
Maryland Institute for Technology in the
Humanities (MITH), University of Maryland

Jon Voss
Shift/HistoryPin

Andrea Wiggins
University of Maryland

Kim Christen Withey
Mukurtu, Washington State University

Pamela Wright
National Archives and Records Administration

Jeremy York
HathiTrust

Courtney Young
Pennsylvania State University,
American Library Association

Jessica Zelt
Patuxent Wildlife Research Center,
North American Bird Phenology Project

Contents

CCLA retweeted

Tom Cramer @tcramer · May 6
Would love to see @crowdconsortium legitimatize crowd sourcing as massive lever for LAMs. Still thought of as "edge" toys. #crowdcon

↩ ↻ 5 ★ 5 ···

Project prototype design group

CCLA retweeted
Jacqueline Wernimont @profwernimont · May 6
@djp2025 I'll be interested to hear how the #crowdcon conversation grapples with the demographic skewing of crowdsourcing and impacts

↩ ⟲ 1 ★ 2 •••

Tweeting project prototype pitches

CCLA @crowdconsortium · May 8
#crowdcon

View photo

↩ ⟲ 1 ★ 2 •••

I. Workshop Opening

WELCOME

Neil Fraistat (@fraistat)
Maryland Institute for Technology in the
Humanities (MITH), University of Maryland

I am one of three co-organizers of *Engaging the Public: Best Practices for Crowdsourcing Across the Disciplines*. On behalf of my two co-organizers, Mary Flanagan of Dartmouth College and Andrea Wiggins of the University of Maryland, I would like to welcome you here and gratefully acknowledge the support of the National Endowment for the Humanities (NEH), the Institute of Museum and Library Services (IMLS), and the Alfred P. Sloan Foundation, who have made this event possible.

Neil Fraistat and Mary Flanagan

We are proud of the interdisciplinary breadth of our attendees, and to kick off the proceedings with a welcome, we are delighted to have our campus guru for research interdisciplinarity, Patrick O'Shea, Vice President and Chief Research Officer of the University of Maryland.

OPENING REMARKS

Professor Patrick O'Shea (@eau_chez)
Vice President and Chief Research Officer,
University of Maryland

My name is @eau_chez. On my job I am the coach of the research team. I don't control research, and if you asked Andrea Wiggins or others if I do, they would say no. My job is to facilitate the research activities at this campus because we are a major institution that is dedicated not just to dissemination of knowledge but to its creation and application for the good of all people, for the crowd.

What are the hot topics we are working on, you might ask? It is very simple and hasn't changed in a thousand years. If you asked the people in Oxford, Britain a thousand years ago at the foundation of their great university what is important, they would have said it is figuring out better ways to house and heal and feed and fuel people in an advanced society that is safe, secure and free. All of those things are transdisciplinary activities that major research

Live Tweeting

We encourage you to live tweet our sessions. We are @crowdconsortium and our hashtag is #crowdcon. • Neil Fraistat

Citizen Science Podcast

Andrea Wiggins, one of my co-organizers, got things under way brilliantly yesterday when she was a guest on NPR on Diane Rehm's show talking about citizen science. You should listen to the podcast because it was a fabulous show. • Neil Fraistat

"The Environmental Outlook: Citizen Science," *The Diane Rehm Show*, May 5, 2015

http://thedianerehmshow.org/shows/2015-05-05/the-environmental-outlook-citizen-scientists

Andrea Wiggins

Neil Fraistat @fraistat · May 5
Excellent post by @AndreaWiggins on Citizen Science data quality as a design problem: andreawiggins.com/tag/citizen-sc.... She is Co-directing #crowdcon.
↩ ⇄ 4   ★ 3   •••

http://andreawiggins.com/tag/citizen-science/

universities like this and your institutions are very well set up to handle.

I have to work at the intersection of the humanities, social sciences, the sciences, and so forth, and MITH is a perfect example of that intersection between science, technology, humanities, and the arts. They have been into this for 16 years now, before anybody thought this was important.

Crowdsourcing and crowdfunding is by the people, for the people, and if you think about research right now, it is crowdfunded because your taxes are what pays for the research. It is filtered in a complicated way from your pocketbook, your paycheck, through the government or some foundation and eventually sort of staggers back into research. Wouldn't it be great, with modern technology, if we could figure out ways of short-circuiting that so that people could actively engage in directly funding research and also actively contribute? Most people don't realize they are funding research and may not fully realize that they are getting benefits, and they have to get benefits. If you read the United Nation's *Universal Declaration of Human Rights*, it actually says that it is a basic human right that all people should benefit from the fruits of research. It is not just a good idea, it is actually considered a human right.

So the crowd, the people, should play a more active role in today's society in all aspects of that, but because we follow the funding mechanisms that we have in place there is quite an enormous disconnect. I think there is a tremendous opportunity for this group to figure out completely new paradigms for doing business. You get the data and the ideas from people, and the funding—the funding is as important. The first duty of a researcher, beyond having good ideas, is to get money. If you don't have any money you can't do anything. Just as the first duty of a politician is to get elected. If you don't get elected you can't do anything.

So figuring out ways to do things that are from the people, by the people, for the people, for the crowd, is absolutely critical to the new paradigm of citizen science, citizen scholarship, and citizen funding, as is citizens then directly seeing the fruits of your work coming back into their lives. I look forward to being here and talking to you, and it looks like it is a great time to have this workshop. The opportunities are boundless.

Patrick O'Shea

KEYNOTE:
BUILDING A CROWDCON

Introduction

Andrea Wiggins (@AndreaWiggins)
University of Maryland

I am so excited that this event is finally happening after much planning and discussion. I did get to talk on NPR yesterday about citizen science [see sidebar, page 8] and at the very end a nice question was posed about digital humanities as well, so that was a nice plug in for this event also.

My duty at the moment is to introduce my co-organizer Mary Flanagan. Mary is an artist, an author, an educator and a designer, or in other words, a Renaissance woman. She is the inaugural chair-holder of the Sherman Fairchild Distinguished Professorship in the Digital Humanities at Dartmouth College, and the Director of the Tiltfactor studio, an innovative game research lab. Her academic book, *Values at Play in Digital Games*, with philosopher Helen Nissenbaum, was just released by MIT Press [http://mitpress.mit.edu/books/values-play-digital-games], and you should run out and buy a copy.

In the context of crowdsourcing, Mary is best known for her work on Metadata Games, which encourage the public to interact with historic photographs and help title them with descriptive metadata. Metadata Games are not just being used at Dartmouth, they have been adopted by 40 collections at nine institutions. Some of the recent partners include the American Antiquarian Society and the British Library. So without further ado, Mary Flanagan.

Keynote Address

Mary Flanagan (@criticalplay)
Tiltfactor, Dartmouth College

Thank you all for coming. I am so grateful you were able to attend this historic interactive workshop for an engaging discussion around crowdsourcing. On behalf of Dartmouth College and the University of Maryland, and with the support of the National Endowment of the Humanities, the Institute for Museum and Library Services, and the Alfred P. Sloan Foundation, we are pleased to host you at *Engaging the Public: Best Practices for Crowdsourcing Across the Disciplines*. We are pleased to have you here and I know that our funding agencies are pleased to be sponsoring this interdisciplinary conversation.

It is fortunate timing right now for us to be gathered together, especially in light of last week's IMLS meeting focusing on the future of libraries. There are a lot of interesting sets of conversations going on, and I am hoping we can bring that together here and really move it forward. I think we have an amazing shot at that.

The aim of this workshop is to culminate and then broaden the conversations from a series of regional meetings and webinars taking

BUILDING A #CROWDCON

http://www.tiltfactor.org
http://www.metadatagames.org
http://www.maryflanagan.com
@tiltfactor, @criticalplay

Mary Flanagan

place through the auspices of Dartmouth's 2014-2015, IMLS-funded, National Forum in Crowdsourcing for Libraries and Archives: Creating a Crowdsourcing Consortium (CCLA), to help advance a truly cross-disciplinary agenda. I will tell you a little more about that while I am chatting here with you.

We have about 60 participants here, key leaders and decision makers, and if you don't know each other yet you will by the end of the next two days. This capstone event is something of a historic moment. Not only are we bringing together amazing scholars and practitioners from across the humanities, sciences, and social sciences, we are bringing together those from institutions public and private, from museums and archives, from library and web-only memory institutions.

I'm going to begin by giving you a little background on how I got here and introduce the project that Andrea mentioned. In 2011 we received a National Endowment for the Humanities Digital Humanities start-up grant for Metadata Games, a software platform that uses games to engage the public to contribute metadata to images and videos. To date, the project has served 45 collections, including those at the British Library, the Holocaust Memorial Museum, and more, generating hundreds of thousands of tags. The system works with customizable plug-ins as portals for various kinds of data gathering and data correction purposes

The project has been working really well and

I have talked to many of you who have said, "Gosh, we really need to talk more with each other about sharing stuff, about who is sharing what source code and who is doing what." Each time someone starts one of these projects they have to ask the same set of questions and they have to go through the same set of things, and we need to advance the dialog. We are all jumping through the same set of hoops and I don't think we need those hoops, it is just an inefficiency. We will actually be able to advance our scientific inquiry and our critical inquiry and our humanistic inquiry further if we can make something work together. So Metadata Games instigated this conversation.

CCLA retweeted

Tom Cramer @tcramer · May 6
@criticalplay produced Metadata Games. One of many entrants in @crowdconsortium field. How can we reuse our tools and methods more?

http://www.metadatagames.org/#games

HISTORY:
METADATA GAMES

Stupid Robot
Tag Accumulation / Verification Game

NexTag
Image & Video Tag Accumulation Game

Pyramid Tag
Tag Accumulation/ Verification Game

METADATA GAMES Server Instance
Saves Tags, Allows Editing, JSON Export & Verification

HISTORY:
PARTNERSHIP Biodiversity Heritage Library

Beanstalk
"Meditative" transcription verification between 2 different OCR manuscripts

Smorball
"Gamey" transcription verification between 2 different OCR manuscripts

We are also working in a similar vein with the Biodiversity Heritage Library and Trish Rose-Sandler, who is a collaborator on that project, is here. That is a set of things to normalize discrepancies over different sets of OCR data on plant manuscripts. We have developed two games that are currently launching in the coming weeks that work between different OCR

systems, and using crowdsourcing to correct those.

So from transcripts, to gathering data tags, to all kinds of different metadata, I am really interested in how we can move beyond that into new systems. In our work on Metadata Games I began to sense a real need for a national conversation. Who is sharing tools? What is open source and what's not? Who is interested?

NEED:
NATIONAL CONVERSATION, INCLUDING
-LIBRARIES
-ARCHIVES
-MUSEUMS
-RESEARCHERS

… SCIENTISTS!
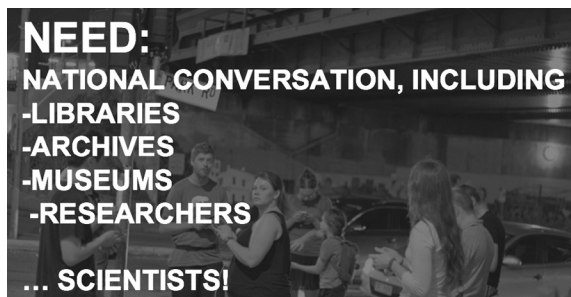
I wanted to help the IMLS pursue what I perceived to be an escalating national need for dialogue about crowdsourcing and about dialoguing across the disciplines. Many of the review panels at ACLS, NEH, and IMLS receive requests for one-off, unconnected crowdsourcing projects or mini-archival projects that seem somewhat unsustainable and could benefit from connectivity and public engagement. There is especially a need to unite this work with that of scientists, who have been working with citizen science ideas for many more years than the 'citizen archivist' or 'citizen scholar' approach has been around.

I began talking with Bob Horton, who was at

the IMLS, and who invited me to give a keynote about Metadata Games and crowdsourcing at last year's IMLS *WebWise* conference. He asked if I could help them create some kind of national conversation about crowdsourcing. I said, "Well, I don't know of a lot of people doing this, but I know some." In late summer of last year we began work on an IMLS Forum project: Developing a Crowdsourcing Consortium for Libraries and Archives. It just so happened that Neil Fraistat and Andrea Wiggins were also asking about having this kind of national meeting with Brett Bobley at the NEH. The three of us kind of triangulated because of the program officers, and I couldn't be happier because it's a great team to start this conversation.

Now called "Crowd Consortium," we started a website [crowdconsortium.org] and pursued a number of avenues to inform a national conversation, and this is kind of the culminating conversation. Even though all of us haven't been engaged yet, we are all engaged in crowdsourcing in a deep way and have things to offer.

We conducted webinars with OCLC and some leaders in the field. The first, "Crowdsourcing 101 with Mia Ridge," who was our host, featured the NYPL and Zooniverse talking about their projects. That webinar had 187 registrants from 42 states and several other countries. Registrants came from organizations that ranged from public, private, academic, and federal libraries and archives, as well as

CCLA:
CROWDSOURCING CONSORTIUM FOR LIBRARIES &ARCHIVES

* Two intro Webinars (Crowdsourcing 101 w/NYPL, Zooniverse hosted by Mia Ridge, 187 registrants; Scoping & Funding w/ NEH & IMLS hosted by Sharon Leon, 144 registrants)

* Two regional in-person meetings (Boston Public Library and UCLA, appx 45 meetings)

* People from over 400 institutions involved

* This meeting – National Workshop

* Results from two National Surveys . .

CCLA retweeted
Perry Collins @perrycollins · May 6
Wow, reps from over 400 institutions involved in @crowdconsortium webinars/regional mtgs in past year. #crowdcon
2   1   ...

CCLA:
WHAT WE HAVE LEARNED

* There is a keen interest in crowdsourcing (CS) (hundreds of respondents; thoughtful answers)

* Survey 1: 83% of 354 respondents want info about existing projects; 70% cited lack of technical expertise, not interest, for lack of involvement

* Survey 2 focused on barriers: only 10% of 157 respondents' institutions use LOD; barriers include lack of internal opportunities, lack of hiring strategies to support CS, and funding allocations that do not prioritize CS

Daniel Powell @djp2025 · May 6
Curious that 70% think technical expertise is a major issue in crowdsourcing. IMO social momentum is key, not infrastructure. #crowdcon

Daniel Powell @djp2025 · May 6
Metadata games (metadatagames.org/#games) prompting conversation btwn interoperability, partnership, moving beyond specialists. #crowdcon

crowd
CONSORTIUM

http://www.crowdconsortium.org

medical centers, government agencies, commercial vendors, and inquisitive citizens.

The second webinar, "Scoping and Funding Crowdsourcing Projects with IMLS AND NEH Program Officers," hosted by Sharon Leon, had 144 registrants and 14 states represented. OCLC does a lot of assessment of their programs, and afterwards found that according to attendees' self reporting regarding the degree to which they increased their skills, 37.5% increased, 21.9% significantly increased, and 21.9% somewhat increased.

We also, as part of our charge, did two national environmental scans aimed at libraries and archives to get some metrics on the lay of the land because we may think we know certain things but don't actually have a lot of evidence. The first survey of over 350 people found that 70% of respondents thought that a lack of technical expertise was a barrier to things like crowdsourcing in their institutions. Less than 10% reported that their institution had incorporated linked open data (LOD), and a really low number of respondents were actually planning to incorporate LOD. So these are some interesting stats when we are trying to think about the future and what are conductive systems that in five, ten, twenty years really help us shift that bar from less than 10% using linked open data. That has a lot to do with our interoperability plan.

What is great is that people from over 400 institutions have already been involved, either in the CCLA meetings, the webinars, or the mailing list. That makes me really happy. And we have got international participants, not only in this room but also online. So this is where we have come thus far, and then we can count this meeting. Why did we come here? I want to have a moment of reflection.

WHY?

Perhaps you are interested in the power of so many voices coming together, the idea that we all have a say in an open democracy. Maybe that is what is motivating us. Perhaps you are seeking to solve a problem that you yourself or your research team could just not accomplish, can't get to, can't afford to do. Perhaps you are just interested in the possibilities offered by new technologies as they emerge; you're an early adopter, you want to see what happens.

I want to take a moment to stop and ask each of us to consider *why* we are all drawn to this research and this practice in the first place. Why do we want to engage the public to help address these massive needs or problems? I want us to reflect on that and to keep that question alive. I would put forward that, at least for me, it is because fundamentally, *we*

are the public too. We see the value in crowd-sourcing approaches. I see the excitement in sharing this mission to democratize technological systems with countless others like us, and unlike us, in engaging in a national or even international quest toward knowledge, meaningful knowledge. I am really excited about that from a human values perspective, from a human rights perspective: that we own our own histories and we should be able to embellish and enrich what we know about our own history. That is what is motivating me, and we all have some motivation that is meaningful to us.


CURRENT:
NATIONAL DIGITAL PLATFORM

Provide everyone with:
Content, Connectivity, Community & Services

The national digital platform is the combination of software, social and technical infrastructure, shared digital services, system, infrastructure, and preservation

I would like to bring up a few key questions that emerged from last week's national meeting in D.C., sponsored by the IMLS, about a "national digital platform" or what was called a "national digital ecosystem" or a "portfolio." What is meant in discussions about a national digital platform? What came out of the meeting is this idea that the goal of the national digital platform is to provide everyone with content, connectivity, community and services. This could be in the library space, but we could also think of it as a research space, and even

throughout the education space as well. The national digital platform is the combination of software, social and technical infrastructure, shared digital services, systems, infrastructure, and even preservation of what we know and digital content. And this is a national and international kind of idea.

CHALLENGES:
FROM THE IMLS FUTURE OF THE LIBRARY
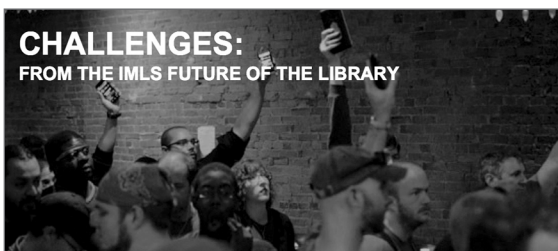
Now this platform obviously can't be a monolithic thing. As Trevor Owens from the IMLS said last week, "It isn't a piece of software or a website. The platform would be all of those things and how they add up."

In this room we have the foundations of an ecosystem. We have an ecosystem of a "crowd-sourcing diaspora" (I think that is Neil's term). We have this national digital platform that already exists comprised of our individual projects, but that platform is largely a diffuse set of disconnected components. What accumulates from the tools, skills, and assets? Where can we go?

There are four questions I want to share from last week's conference because they are so relevant to us. One of the challenges that emerged from this conversation was: What are the top 10 gaps that must be addressed to

CCLA retweeted
Tom Cramer @tcramer · May 6
@imls and @brettbobley at @NEHgov both interested in more organized crowd space. (Moving from mob to crowd?) genesis of @crowdconsortium

CHALLENGES:
FROM THE IMLS FUTURE OF THE LIBRARY

1) What are the top 10 gaps that must be addressed for engaging the public through crowdsourcing?

2) How do these effect the future of the library, museum, archive? Science, inquiry, learning, citizenship?

3) What can we recommend for: researchers, institutions, publics? *This year, 5 years out, 15 years from now?*

4) What does 'scale' look like: national digital ecosystem? *(think Youtube, World of Warcraft, Reddit; 1b views? What radical phase shift can change the relevance of learning & memory institutions?)*

CHALLENGES:
FROM THE IMLS FUTURE OF THE LIBRARY

1) What are the top 10 gaps that must be addressed for engaging the public through crowdsourcing?

2) How do these effect the future of the library, museum, archive? Science, inquiry, learning, citizenship?

3) What can we recommend for: researchers, institutions, publics? *This year, 5 years out, 15 years from now?*

4) What does 'scale' look like: national digital ecosystem? *(think Youtube, World of Warcraft, Reddit; 1b views? What radical phase shift can change the relevance of learning & memory institutions?)*

PLAN

imagine the future of the library. I have said on the slide, "What are the top 10 gaps that must be addressed for engaging the public through crowdsourcing?" What are these gaps? Can we just name 10 and prioritize them and move forward?

The second is: How do these affect the future of the library, museum, and archive? Science, inquiry, learning, and citizenship.

Third: What can we recommend for researchers, institutions, publics? This year, five years out, 15 years from now?

A high-level person at the Smithsonian asked at this conference last week, "I'm a bureaucrat, tell me what to do to change my institution. Tell me what we need to do to get there. Do you need money? Different people? Tell me." As a community, that is a great invitation and we should be able to say, "Do this…" Maybe we can get to these answers.

The fourth question is: What does 'scale' look like? What is a radical phase shift that can change the relevance of cultural institutions and make our products and website services reach 10 million or 10 billion people? Can we make the library into Reddit (or "Redditish")? What really needs to happen? This is a great rich ground for our thinking during these next two days.

These questions are on my mind and they came up last week. I know that each of us brings other questions, and those are equally important to share during the next two days. Over

those two days we will enjoy a mix of short formal and informal presentations with active facilitated discussion. We will focus on questions around how researchers and institutions might best leverage crowdsourcing strategies for increasing public engagement, integrating data into existing collections, and improving knowledge production in a variety of domains. You are here because we believe that you can make an important contribution to this conversation.

The disputed Venn diagram appears in my talk. Right now, as I alluded, crowdsourcing in our
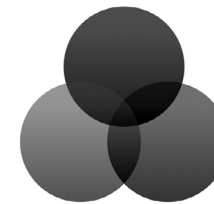
communities is a diffuse set of largely disconnected components. Accumulating what we know really needs to happen. Where do we accumulate tools, skills, and assets? Where can we go?

Right now we might be able to move these useful energies to more federated or combined or coordinated ways of working together to really take advantage of what we've learned over the last 15 years. This conversation opens the gates across fields to ensure we have a true cross-disciplinary group working together.

It is our belief that a consortium of some kind is needed, precisely to move these ecologies forward in a connected way. I am talking about organizing and coordinating. I know that in talking about an organization to coordinate, there are potential challenges. Needs might be very

different, publics might be very different. We don't need "one ring to rule them all," but how could a consortium foster a series of connected crowdsourcing ecologies—that is one big question for us here—a consortium to help us tackle more complex problems and dig deeper into those problems? We need to be finding, identifying, and developing the overlaps.

As we work together, we are working towards more interoperability, joint services, more tools, evaluation techniques, federating our data in ways that will scale and reach more people. We can share what we have learned, build on each other's knowledge, and there are lots of different groups coming at crowdsourcing. We haven't yet crystallized around one set of tools and best practices, and it is time we at least move forward the conversation on best practices. So that is our task together in the next few days, to move the conversation forward. Over the next two days, we will enjoy all of this stuff and move forward as we focus on questions around how researchers and institutions might best leverage crowdsourcing strategies for increasing public engagement, integrating data into existing collections, and improving knowledge production in a variety of domains.

Finally, I want to talk for a moment about permission. We all have jobs that in some ways shape how we think and what we are allowed to do and not allowed to do. I ask that over the next two days we give ourselves permission to dream big, to try new things outside

YES! AND

of our discipline or our comfort zone (you are in a very interdisciplinary group for a reason), and move to "Yes, and" approaches to idea generation. Anything at this point is possible. Maybe it's not probable, but we can make it probable if we all work towards it. As thought leaders working together, we are crafting the foundations of national ways, and even international ways, we can advance knowledge and engagement of fellow citizens, not only with our particular specialties but with inquiry and curiosity overall.

As a dedicated group of technologists, designers, and strategists, and as leaders engaged

CCLA @crowdconsortium · May 6
.@criticalplay "We don't need one ring to rule them all". Supporting related, overlapping crowdsourcing efforts: interoperability #crowdcon

THANKS

I have a few thanks to make at this point. I would first like to thank my co-organizers, Neil Fraistat, Director of the Maryland Institute for Technology in the Humanities, and Andrea Wiggins, Information Scientist and faculty at University of Maryland. We would like to thank IMLS, NEH, and Sloan for their support and enthusiasm for such an ambitious gathering. And I would like to thank Stephanie Sapienza and Danielle Taylor, who expertly took charge of our logistics and your travel and deserve our gratitude.

We have here at the meeting a team of expert facilitators, verynice.co, who will help us reach our goals.

We also have an active note taker and observer who will be transforming what we talk about into a proceedings document. Thanks to Cathy McEver in advance for her careful documentation.

We have a webcasting team to thank, as well as students and helpful organizers here at the University of Maryland. Thank you.

with creating practical solutions to pressing needs as we move into an unprecedented era of connectivity, speed, and simultaneity, we can shape how our engagement with the public can benefit all. My hope is that this little talk has provoked some ideas for us, that our conversations over the next two days are a chance for us to transcend our everyday operations and our constraints of what is and what has to be and why, and instead consider what can be, what could be, how could it be.

So thank you all for coming and making this conversation happen. Take crowdsourcing as we have known it to the next level, nationally and internationally.

Questions, Answers, Discussion

MOVING FORWARD, FUNDING, INTEROPERABILITY

• Can you talk a little about what you think might happen at the conclusion of this effort with this survey and now this national capstone event? • Tom Cramer, Stanford (visitor)

• It is a capstone event for the IMLS proposal, but actually it is the beginning of a conversation in other ways. We want to keep the Crowd Consortium as an entity if it is useful. It is only useful if we as a community think it is useful. So far it has proven useful, and it seems that there is support from a variety of agencies to keep that conversation going. I can't speak for our program officers, but

there is interest in adding voices and trying to move forward.

Many people are very interested in trying to understand how funding proposals should work, what kind of calls make sense given this ecosystem, how particular funding programs can support this kind of work that may be difficult to define, or works across projects. There is a lot of need to retool the way we actually are supported as well as how we are making things.

One of the reasons interoperability is in such a crisis is, in one way, because of the way projects are funded. We are all working on our own projects. If we are actually allowed to move and create across projects, that is one way that agencies, for example, could encourage us. In a way we are shifting the paradigm a little bit about research. • Mary Flanagan, Tiltfactor, Dartmouth College

DATA MOVEMENT, SCULPTING, CURATING

• Once you can crowdsource, what do you do then? How do you bring that conversation and get citizen archivists to communicate with other people who are collaborating, and how can both parties learn from this? • Kyle Bickoff (one of Neil Fraistat's students)

• That's a great question and I think it's part of our charge. A certain subset of us are looking at data, getting new data, putting the data back. It's like there's a data trail. Maybe you are talking about interpretation of the data

DEFINING "CROWD"

• I was wondering if in this effort at some point there is a definition or parameters around "crowd." Does it have to be someone who is outside of your immediate community? Of course, people talk about crowdsourcing within libraries and so forth. I wondered if there is some kind of definition out there. • Jeremy York, HathiTrust

• We can define that here. I have defined it really loosely as problem solving, information gathering, or changing certain kinds of things with large numbers of people. That has been kind of my operating definition of crowdsourcing. But again, we can define that more clearly if that's needed and that is what we want to do as a community. • Mary Flanagan, Tiltfactor, Dartmouth College

and sculpting the data and curating the data, which is another way to see these tracks of data moving from an archive out to the public and back to an archive. I think you are talking about how data movement can happen and how we can add to it in educational settings or other settings. It is very important and I think it's part of this conversation but again, I don't think any of us would claim to have an answer to that. If you do, please address this question.

I think what we are up against is that we don't necessarily know where each other are coming from, but we have to try to bridge these silos between our projects and our disciplines. It's a tall order, but we are all very passionate about it and you were all brought here for a reason. I think that conversation can happen here in really dynamic ways.
• Mary Flanagan, Tiltfactor, Dartmouth College

THE ANALOGY WITH CITIZEN SCIENCE

• The analogy with citizen science is an interesting one, and it's something we think a lot about at Zooniverse because we cross both worlds. But citizen science has always been a bottom-up effort, and we haven't had this meeting, we haven't had this sort of funding to look at what citizen science should be. So the analogy is interesting, but we are coming at this from completely opposite directions. No one sat down and decided that we should do citizen science, it was people stumbling into it. Whereas here there seems

to be more of a desire to do this sort of thing for the other reasons that you listed. • Chris Lintott, Zooniverse, Oxford University

• I think you'll find that there are a lot of projects that came out of grassroots efforts. There are a lot of projects here that are not citizen science that are from a bottom-up effort. I don't think that it's quite as clear a divide as that. But now that the projects in this room are all active, I think that's why we are having this cross-conversation. It's not about originating projects so much as it is about connecting them and moving everything forward. But we will find out through our conversations if perhaps there are as many differences as you perceive, or if people are actually in the same place. • Mary Flanagan, Tiltfactor, Dartmouth College

FOSTERING COMMUNITY

• I am also interested in not just how it pleases the crowd in terms of what we are doing with them but also, when that crowd gathers, how can we promote that and have an online community and develop that and take it to that next level so that they please each other and they work with our stuff together? Right now that is super-siloed. I am very interested in that and I think that the student who spoke also mentioned that. How do you get that community online and build that? I am very excited to talk with folks here about that. • Pamela Wright, National Archives Records Administration

CROWD ENJOYMENT, SATISFACTION, LABOR RIGHTS

• I want to follow up on that citizen science question. I am somewhat of an interloper here. I am not a librarian, I worked at Rap Genius|Genius for two-and-a-half years doing a crowdsourcing project. My thought about the crowd there was the experience of the crowd and how people will enjoy this project or want to do this project and get involved, as opposed to the data, the result of such activity. I appreciate the nuance between grassroots or bottom-up to top-down efforts, but thinking about what the crowd gets out of this is something that concerns me as opposed to just the data that results from that activity. • Jeremy Dean, Hypothes.is

• There has been a lot of conversation in New York at the New School and many, many years of conference about labor. There are real questions of labor that should be asked and talked about and what is in it for participants. • Mary Flanagan, Tiltfactor, Dartmouth College

• Yes, where the pleasure is. • Jeremy Dean

SAVVY PARTNERS, NOT REINVENTING THE WHEEL

• In response to that citizen science comment, as an institution that is very interested in crowdsourcing, it still seems like it is more art than science. I would love partners who are further ahead, so I don't have to reinvent their wheels. So I don't need to force crowdsourcing on people, I just don't want to reinvent what others have already done. • Tom Cramer, Stanford (visitor)

• That's right, and we don't really have the funding to do that either. It's money, it's time, it's creativity, and it's also a lost opportunity to build up, or over, or however we want to visualize that. We can't go deeper if we keep starting over. • Mary Flanagan, Tiltfactor, Dartmouth College

• When I was saying that there is a group of us who really want to track data and there is a group of us who really want to curate the data, there is also a big group of us who want to engage with people either where they're at or where they want to be. That is why we actually called this "Engaging the Public." It is not about a superficial engagement, it is about what is genuinely useful and interesting. What would people be passionate about and how do we help that happen if that desire is there? • Mary Flanagan, Tiltfactor, Dartmouth College

CONTROL AND TRUST ISSUES

• When I was tweeting, Twitter was autofilling "crowdcon" as "crowdcontrol." Where is the control? How much control is too much control? How much are those of us who are building crowdsourcing tools just giving up in order to foster new thoughts? • Neil Fraistat, MITH, University of Maryland

• These are really rich areas. It goes along with the digital labor question and it also goes along with the other kinds of fears that I've run into at certain institutions about, "Oh, if we bring in knowledge from the crowd how do we know if it's correct?" So there is that kind of crowd control need or real worry about the validity of information. And then there is the completely opposite side: "I want my experience to be what I think it is and not be coopted." From a values perspective there is a lot of rich language we could use about trust and about responsible systems, responsible design. I think we should engage in these discussions. • Mary Flanagan, Tiltfactor, Dartmouth College

Tom Cramer @tcramer · May 6
@criticalplay nails it: "we can't go deeper if we keep starting over." #crowdcon
↩ ⟳ 2   ★ 4   •••

II. Surveying the Future: Key Trends, Opportunities, and Challenges

WORKSHOP GAME PLAN

Introduction

Neil Fraistat
MITH, University of Maryland

In her remarks Mary Flanagan eloquently sounded some of the key themes for this event, the most important of which is to see how we might gather together a "crowdsourcing diaspora" in some coordinated way to foster collaboration, knowledge, best practices, sustainability, and mutual support. We have named the notional emerging entity "Crowd Consortium," or "crowdcon," and will be giving thought during the next two days to how that might best develop.

Much of our thinking will also revolve around the biggest challenges facing crowdsourcing efforts across the disciplines and how best to meet them. Our facilitators, whom we will introduce in a moment, will be talking shortly in more detail about our goals for this event. But for me, it will be a success even if we only succeed in identifying the low-hanging fruit that different parts of this group are then able to pursue, and we are able to make the crowdsourcing movement across the disciplines more visible to itself as a community.

We are looking forward to your creative thought and your energy and believe that we have assembled a veritable dream team of crowdsourcers to work together on our common challenges. Here's wishing you speedy, thoughtful thinking.

Mary Flanagan (@criticalplay)
Tiltfactor, Dartmouth College

I want to thank all of you again for coming and for your engagement. There is a lot of really great energy in the room and there are ideas popping up all over the place. All we have to do is look at Twitter and see them all coming out. So please keep it up, and keep up the Twitter conversation. Again, think "Yes, and." We will be trying something new and getting out of our comfort zones.

About the Facilitators

Andrea Wiggins, University of Maryland

We have a lovely group of people from verynice.co. Matt, Jake, and Sheena will be taking over for us and leading us so that we can play with all of you and be part of the conversation too.

The verynice.co team (from left): Sheena Yoon, Matthew Manos, Jake Dunagan

Putting the Big Ideas into Action

Matthew Manos, verynice.co

You are mainly going to be hearing from Sheena and me tomorrow. We are the people who are going to bring these big ideas into actual plans with you all, so we are excited to mainly observe today and make tomorrow as useful as possible.

Facilitation Overview

Jake Dunagan, verynice.co

We are from verynice, a design, strategy, and foresight firm based in LA, Austin, and New York. We are excited to lead you through a couple of days of exploration, of practical application. We are going to learn together. In that spirit, we are asking you to come together. You have a lot of different experiences and knowledge bases and languages being spoken. We would like for you to have a spirit of intellectual hospitality and openness. I'm going to paraphrase what Mary said earlier: You have permission to transcend everyday constraints in the pursuit of advancing knowledge. So what a mandate we have! Let's do that together.

We are going to do the best we can to facilitate and serve as guides. When we have a structured process we want to follow that, but we also want to have a chance to improvise and go down different directions if we need to. We will allow things to float as we see them going, and we might have to constrain to keep things moving forward. We are going to be open but forward-moving at all times, that is our goal.

There are a few ground rules I would like to lay out. I mentioned this idea of mutual respect and intellectual hospitality, and I think we are going to have that. We want you to have strong opinions but weakly held, so have a passionate point of view, but also be open to changing that. Observe other people's ideas and open yourselves to changing your opinions. There are 60 people here and if we cut you off, just know that it is in the spirit of keeping the process moving forward. We also want to stress punctuality and efficiency as we move in and out of small groups.

Matt and Sheena are the designers and will be working with you tomorrow. I'll be leading you today, and today is about surveying, it's about exploration. So for those of you who have a tendency to want the practical and to put things into action, save those tendencies for tomorrow. For those of you who like to take a step back and explore, today is your day. Tomorrow you need to be a little bit more focused. So we are going outward today and coming back together tomorrow to push for some real outcomes.

ISSUES, CONCERNS, PRIORITIES

Introduction

Jake Dunagan, verynice.co

We are breaking into six groups and in those groups we want you to introduce yourself and share a "signal from the future." A signal from the future is something that is sort of edgy, it's weird, it's new, it's different, but it could be the seed of something big in five years or ten years. Five or ten years ago some guys were messing around with dot matrix printers and figured out how to do 3D printing, and that was just weird hobbyist activities that nobody really cared about. Then within a few years there was Makerbot selling for $600 million, and now the U.S. Postal Service is putting 3D printers in their offices. So we want those early things, the strange hobbyists, things that aren't on everyone's radar.

So share those and at the end of this session I am going to go around and ask for one or two insights from the conversation or the most interesting signals that came up.

Group Report-Outs

GROUP A

Reporting: Daniel Powell, King's College London

Our group talked about the end goals of crowdsourcing, including remediating materials and gathering information for environmental purposes. We also talked about what crowdsourcing means in terms of voluntary participation in information gathering versus involuntary participation in information gathering, which happens more and more with the Internet. And then, how do you scale up and tap into existing communities and networks that are interested in these kinds of things? What came up especially in our group was pedagogical and in the classroom. How do we integrate pedagogy in producing research with large scale participation?

Process Note

While the original charge was for these groups to share "signals from the future" as described in the introduction, most participants instead focused on issues, concerns, and priorities important to them as they head into this workshop and tackle the challenge of advancing crowdsourcing to the next level. Report-outs from those groups are offered here, along with a sampling from the Twitter stream during this section of the conference.

CCLA retweeted
Daniel Powell @djp2025 · May 7
#crowdcon Signal: 3D data, capturing thick information, expressing itself in multiple dimensions.

CCLA retweeted
Daniel Powell @djp2025 · May 7
#crowdcon Working in our small groups to pick up some signals from varied communities. So far: integrating into pedagogy, classrooms.

chrislintott @chrislintott · May 7
.@andreawiggins Citizen science is exciting to me precisely because of the potential to change people's relationship to science. #crowdcon

Raffaele Viglianti @raffazizzi · May 7
Someone mentioned this handwriting OCR challenge using @TranscriBentham crowdsourced transcriptions transcriptorium.eu/~htrcontest/ #crowdcon

http://transcriptorium.eu/~htrcontest

Photos in this section: groups discussing issues, concerns, priorities

CCLA retweeted
Dr Meghan Ferriter @MeghaninMotion · May 7
My signal: Empathy, or the HOW of crowd/COP connecting goals across digital space - care for those there with you in projects #crowdcon

CCLA retweeted
Daniel Powell @djp2025 · May 7
#crowdcon How do you tap into existing networks, existing communities for knowledge work? Teaching skills and prototyping as part of rsrch

CCLA retweeted
Victoria Van Hyning @VanHyningV · May 7
Some great signals from the future from #crowdcon: how to incorporate crowdsourcing in the classroom and beyond, e.g. apprenticeship-like

CCLA retweeted
Amy Patterson @amycep · May 7
#crowdcon How do you scale up? How to integrate pedagogy with large-scale research?

CCLA retweeted
Jon Voss @jonvoss · May 7
Insights from our #crowdcon signals group: crowds to community, anonymous to organized.

Reporting: Alexis Rossi, Internet Archive

Courtney Young crowdsourced her signal from the future, which we all thought was pretty amusing. One thing that was brought up was using technology to enforce anonymity, and obviously we were also talking about people collecting data without our knowledge, but also opting into things and whether you want to.

GROUP B

Reporting: Andrea Wiggins, University of Maryland

We talked primarily about the learning outcomes that can be achieved for various constituencies by participating in crowdsourcing in all kinds of different contexts, and were particularly thinking about:

• Annotation as an everyday activity.

• Humanities being promoted like science and STEM are right now.

• Using digital storytelling and thinking about ways that participation in these projects could move people toward real professional skill acquisition. As sort of a complement to the academic track, we've got the apprenticeship track and actually doing real work in meaningful projects that could then be used to help those people move forward professionally.

• Narratives as resources across a lot of these projects.

• Crowdsourcing becoming a fundamental component of our learning practices.

Reporting:
Neil Fraistat, MITH, University of Maryland

One of our signals from the future is that the Crowd Consortium celebrates its 50th anniversary. Another was that social media itself becomes a subject of crowdsourcing. I think increasingly it will be easier to collect social media archives, and then they would be subject to annotation and additional metadata analysis. It's an important content area that I think crowdsourcing is going to move into.

Reporting: Trevor Owens, IMLS

Jeremy had a great point, which was a signal from the past about what the web was supposed to be able to do: have an annotation layer across the whole space and obvious things sort of fitting together. I think that also fit with a set of other early ideas about the web as a collective intelligence or a global knowledge base. And there was the idea of the possibility for learning as individuals that exists in that space through this meaningful participation in different projects. I think we had a great example in the Zooniverse Folger project, where you are potentially actually learning a lot about manuscripts and interpreting them. The idea is trying to come up with ways to help capture that, so there is individual learning that is in some ways much more meaningful because it's legitimate participation in the production of knowledge globally.

So there is this interesting element about meaningful engagement in projects furthering knowledge globally and also personally and individually.

GROUP C

Reporting: Jen Hammock,
Encyclopedia of Life & Smithsonian Institution

My group covered a big area, but one thing that came up several times was crowdsourcing as an in-person or social occurrence activity. A lot of this does take place online, but that doesn't mean you can't have a social element if the technology is there and several people have already had events at their institutions, where people show up in groups and hang out and chat and transcribe.

Reporting: Jenny Preece, University of Maryland

We also talked about including people of various backgrounds in crowdsourcing activities.

GROUP D

Reporting: Kim Christen Withey,
Mukurtu, Washington State University

One thing that Chris said that was important for all of us was that open sourcing our projects, our code, is, if not a best practice, certainly a good practice. It is also not sustainability. It doesn't equal sustainability. Thinking that we're going to have large amounts of developers out there to contribute is a fallacy. And while funders may want that and it is good to have in there, we may need to rethink it.

Another point was the difference between platforms and projects. If we are building platforms and people are doing projects on them, how do we sustain both of those and the different types of crowds within that?

I will throw in my own nod to make sure we talk about multiple publics and not just "the public" at different scales.

CCLA retweeted
Victoria Van Hyning @VanHyningV · May 7
Digital storytelling as a form of empowerment. Volunteers collect and analyze data, but also need mechanisms to tell their stories #crowdcon
6   8

CCLA retweeted
Amy Patterson @amycep · May 7
Open sourcing code doesn't equal sustainability #crowdcon
1   2

CCLA retweeted
Daniel Powell @djp2025 · May 7
#crowdcon Signals: in person crowdsourcing activities are vital; inclusion of diverse groups in such efforts.
1

CCLA retweeted
Amy Patterson @amycep · May 7
Group considers crowdsourcing as in-person or social occurrence, event -- not just online #crowdcon
1

critical play @criticalplay · May 7
The crowd is dead. The community lives! #crowdcon
1   7

CCLA @crowdconsortium · May 7
Shift from "crowds" to "communities"; no abstract crowd, but specific people; how do we get individuals to be motivated #crowdcon
4   6

Trevor Owens @tjowens · May 7
Brought up @plevy's notion of Collective Intelligence and participation in a global knoweldge base in #crowdcon bit.ly/1F0D88Y
3   3

http://bit.ly/1F0D88Y

http://bit.ly/Kyz7ta

> **CCLA retweeted**
> **Trevor Owens** @tjowens · May 7
> Important point raised. Crowd is a misnomer. It's about engaging communities. #crowdcon connects to -> bit.ly/Kyz7ta
> ↩ ⟲ 5 ★ 8 ⋯   View summary

> **CCLA retweeted**
> **Victoria Van Hyning** @VanHyningV · May 7
> Shift from crowds to communities. How do we involve people in 'big picture'; ask their own questions; pursue their own research #crowdcon
> ↩ ⟲ 3 ★ 8 ⋯

> **CCLA retweeted**
> **Daniel Powell** @djp2025 · May 7
> #crowdcon Happy to hear one group talk about labour rights (such as reliance on Mechanical Turk)
> ↩ ⟲ 2 ★ 3 ⋯

> **Dr. Dean** @dr_jdean · May 7
> One response to labor issues in crowdsourcing: recognize [and perhaps enable] the delight in the work -- @MeghaninMotion at #crowdcon
> ↩ ⟲ 6 ★ 6 ⋯

> **Brett Bobley** @brettbobley · May 7
> Crowdsourcing runs a gamut from "crowd workers" (strictly in it for money) to those contributing to a project for fun/pleasure. #crowdcon
> ↩ ⟲ 4 ★ 5 ⋯

Reporting: Ben Brumfield, Collaborative Manuscript Transcription, FromThePage

There was the question of scaling down projects and turning them off. What is the best practice for doing that? These projects aren't around forever, they're not funded forever, and again, open sourcing the platform doesn't mean someone is going to be around to maintain it in five years. We don't have good answers, but it's a really good question.

GROUP E

Reporting: Dominic McDevitt-Parks, Wikipedia, National Archives and Records Administration

One discussion was process versus product and how we value data and whether the participants are also getting something of benefit out of the process, whether that's exposure to the collection or learning in some other way.

We also had a good conversation about platforms and communities, whether it's aggregating communities, but also the value of having different communities doing different things for different reasons.

Reporting: Brett Bobley, National Endowment for the Humanities

I put Rachel Frick from Digital Public Library of America (DPLA) on the spot and asked, when I'm on DPLA and looking at a collection, for example, do I know if that collection has a crowdsourcing option? Is there someone who is aggregating all of these cool crowdsourcing projects that are going on and figuring out ways there could be a recommender system, like in Facebook? For example: "If you're interested in the eBird project, you might also be interested in this other great citizen science project." Is there a way to incorporate that kind of function?

GROUP F

Reporting: Ashwin Gopi, New York University

One of the most important things that kept coming up is this shift in perspective from crowds to communities. I think most of us don't believe that there is some abstract crowd that exists somewhere, but rather communities with specific interests, specific beliefs, and specific assumptions about either the process or the issue, and this has a lot of implications.

One of the implications is: How do we get individuals to be more motivated in these projects when they are performing micro-tasks? How do we get them to understand what the big picture is and get them really involved in the big picture?

Another one was: How do we get individuals to ask their own questions, even in citizen science, for example, rather than having a

hierarchy of scientists versus citizen scientists? How do we get people to be curious and ask their own questions and find the answers for those questions?

And finally, there is the implication that if we assume there is a community of actual people rather than a crowd, there is the issue of labor rights. That keeps coming up because in the future if we all become cloud workers, there must be some sort of self-organization to protect us.

Reporting: Liz MacDonald, NASA, Aurorasaurus

We had a lot of discussion around historical signals from the past and the future and how we are currently documenting our lives, how that fits into crowdsourcing, and how that was done in the past and is also important for understanding location-based history and other cultural and scientific interests.

Reporting: Nick Adams,
University of California, Berkeley

Anyone who is building a project is facing the challenge of designing tasks so that they are easy enough for crowd workers or citizen scientists to do, but hard enough and interesting enough that they want to continue doing them.
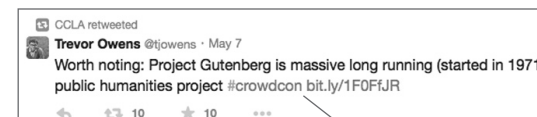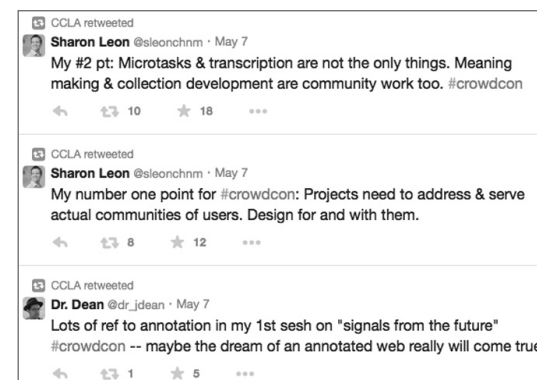
Summary Thoughts

Jake Dunagan, verynice.co

A couple of things that stood out to me are definitions and words, and obviously words have power: "the public" versus "multiple publics" changes the way we see it. Or "community" versus "crowd." So these words come with values and meaning and expectations tied to them. Today is a great day for exploring those intricacies of meaning and what word we are using strategically for what purpose, and attending to those things very carefully.

CCLA retweeted
Sharon Leon @sleonchnm · May 7
My #2 pt: Microtasks & transcription are not the only things. Meaning making & collection development are community work too. #crowdcon
↩ ↻ 10 ★ 18 ...

CCLA retweeted
Sharon Leon @sleonchnm · May 7
My number one point for #crowdcon: Projects need to address & serve actual communities of users. Design for and with them.
↩ ↻ 8 ★ 12 ...

CCLA retweeted
Dr. Dean @dr_jdean · May 7
Lots of ref to annotation in my 1st sesh on "signals from the future" #crowdcon -- maybe the dream of an annotated web really will come true
↩ ↻ 1 ★ 5 ...

CCLA retweeted
Trevor Owens @tjowens · May 7
Worth noting: Project Gutenberg is massive long running (started in 1971) public humanities project #crowdcon bit.ly/1F0FfJR
↩ ↻ 10 ★ 10 ...

http://bit.ly/1F0FfJR

Trevor Owens @tjowens · May 7
.@chrislintott learning often isn't the users objective, it comes as an outcome of participating in meaningful work/activity #crowdcon
↩ ↻ 12 ★ 10 ...

Pamela Wright @PamelaSW · May 7
#crowdcon we seem to be talking about crowdsourcing on 3 levels. 1technology, 2 content, 3 communities.
↩ ↻ 5 ★ 8 ...

CCLA retweeted
DominicMcDevittParks @Dominic_MP · May 7
Seems to be general agreement at #crowdcon so far that amorphous term "crowd" obscures the real communities behind participatory projects.
↩ ↻ 3 ★ 7 ...

Roundtable Participants

Moderator:

• Brett Bobley,
National Endowment for the Humanities

Panelists:

• Neil Fraistat,
MITH, University of Maryland,

• Liz MacDonald, NASA, Aurorasaurus

• Jessica Zelt,
Patuxent Wildlife Research Center,
North American Bird Phenology Project

• Ashwin Gopi, New York University

• Trish Rose-Sandler, Missouri Botanical
Garden, Biodiversity Heritage Library

Jessica Zelt and Neil Fraistat

ROUNDTABLE: DISPATCHES FROM THE FIELD

Introduction

Moderator: Brett Bobley,
National Endowment for the Humanities

I am very pleased to be your moderator for the first panel where we are going to be talking about both emerging and established projects. I am going to ask each of the panelists in turn to briefly introduce themselves and offer their spiel, and then we will open it up to conversation and discussion.

Shelley-Godwin Archive

http://shelleygodwinarchive.org

Neil Fraistat, MITH, University of Maryland

I'm here because of a project I direct called the Shelley-Godwin Archive, which consists of the digitized manuscripts of Percy and Mary Shelley, and Mary Shelley's parents, William Godwin and Mary Wollstonecraft. We received a grant from the NEH to start the project up and the project partners are the New York Public Library, the Bodleian Library, the British Library, the Huntington Library, and Harvard's Houghton Library. Among them, we have 90% of all the known manuscripts. At this point in the project we actually have all of that digitized.

We started off with the idea that we were doing this mostly as a form of access. Once we started going and saw what we had and what might be done, we started to think about this as a participatory archive. We are very concerned and interested to get the material we have curated both into the classroom and outside to the public. We have manuscript images, sometimes we have transcriptions that have been fully checked, sometimes we have no transcriptions, sometimes we have complete markup in TEI, sometimes we don't. So the kind of capacities we were looking for included transcription, text markup, annotation.

Regarding the challenges or problems we face, there is nothing out-of-the-box to use. You have to do your own research to figure out what the affordances are of each of the kinds of tools out there and what they might allow you to do. This isn't easy work, and it's not easy work getting those things to do what you want to do.

We originally started just thinking about the technical problems and how data would come and pass through, and what we learned was that maybe an even larger problem was community design. We are interested in the people who participate actually feeling they are learning and that they are contributing to a larger site about learning, so community design becomes really important. It is not so easy to find best practices for developing and enriching communities like the ones we want to create.

We are midway in thinking through these issues and our discussions here will be really useful in helping us think about the best way to go forward.

Aurorasaurus

http://www.aurorasaurus.org

Liz MacDonald, NASA, Aurorasaurus

I am the founder of a project called Aurorasaurus. I am a scientist, a physicist, and this citizen science project was designed to allow people to get a better idea of when they could see the northern lights. The reason we needed this was because there are no products designed for the public to do this right now, and the public is very interested in seeing the aurora, they're tweeting about it. We wanted to put those observations on a map and do better at informing more people about when this visibility was available to them. It is a rare event that happens in a relatively short time period, but it is also a really large event that happens globally.

We have built a platform for doing that, and we have interdisciplinary projects with Goddard, with a nonprofit called the New Mexico Consortium, and with Penn State University, where Andrea Tapia is interested in how this rare event that we are forecasting and the notifications that we give people might help them build better early warning systems for other types of rare events, disasters that you

can't simulate. So this might be a closed-loop kind of example.

The participants sign up to get a location-based alert near them when people actually see something; then they can report that on the website or our apps. We are also getting observations off of Twitter, and you can upload or download these tweets which are related to our topic and find the needle in the haystack that way in terms of finding when somebody has just tweeted, "I just saw the Aurora," versus, "I just want to see the Aurora." That is why we need people to help us do that. We also have some really basic gamification with points—you get points for interacting with the website.

This is definitely an emerging project. We have been live for under a year. There was a really large space event of aurora being visible very far south in the Northern Hemisphere on St. Patrick's Day, so we saw 100% increase in the number of users and collected a lot of data that can be useful as ground truth for the activity of the aurora right now. It's useful for

Panel (from left): Liz McDonald, Trish Rose-Sandler, Jessica Zelt, Neil Fraistat, Ashwin Gopi

CCLA retweeted
Victoria Van Hyning @VanHyningV · May 7
Great question/musing from Jessica Zelt #crowdcon: 'What makes a superuser?'. Check out this article about her work! bit.ly/1dPLXZa
↩ ⟲ 6 ★ 8 •••

Jessica Zelt

scientists and for people to help them prove the forecasts that they get.

We resemble the Venn diagram Mary Flanagan showed during her presentation. We have space science, informal science education and human-centered computing, and here we are in the middle. One disappointment, I think, is that you think of the Venn diagram as being equally distributed, and there aren't that many people in the middle. We are trying to encourage and find people who are interested and get the word out about this and at least in my field, as a space scientist, people are not on Twitter at all. They are not seeing this as a real data source, so we are changing the perception by showing them the data that we have. That is a challenge, but a good challenge.

And as Neil said, there is not a lot of out-of-the-box for this now. I am very interested in learning more about what people here have been doing. We have a small team and this was a totally different project for us, so the best practices in terms of building software or an app are also a challenge.

There have been a couple of unexpected outcomes. There is certainly advocacy for this very small niche field, and this is a good way to engage with the public. We have also made connections with people around the globe who are already doing this. They are hunting the aurora and photographing it and they are very, very good at it and interested in being connected to others who are doing that as well.

I'll stop there. I'm here to learn and eager to learn more.

North American Bird Phenology Project

https://www.pwrc.usgs.gov/bpp

Jessica Zelt, Patuxent Wildlife Research Center, North American Bird Phenology Project

I work for the US Geological Survey at the Patuxent Wildlife Research Center and help to manage a citizen science program called the North American Bird Phenology Project. This is a program that was actually one of the first citizen science projects in the United States back in 1881. A teacher in the Mississippi Valley was really interested in bird migration and started getting his friends to start recording that information. That grew into a network throughout North America. At that point they were looking at bird distribution and migration of all migratory birds we see in the United States.

We have taken that dataset that was in storage for about forty years—the original program closed in 1970—and repurposed it to look at phenology and how climate change affects arrival and departure dates of migratory birds. To do that we have scanned all of our records and we have a large network of volunteers who participate with us to transcribe those records. Then the records go through a validation process and are put back out to the public so that anyone from the scientific community can use those records as well as the public themselves.

One of the things that I think is important about this project is that it takes a pretty daunting subject area like climate change and boils it down to something that is very simple and optimistic. We have created a way that volunteers can get involved and see the outcome and impact of their work. When they transcribe they get to see the work that they've done and all the work that the other volunteers are doing. They compare and are pretty competitive with each other in trying to get their work done, and they feel like they are invested in the outcome. All of the data, when it's used in scientific research, which is what we are doing now, is then put back out to the public to share what it is that we're doing with all of the work they put into it.

OpenIDEO and Women for Human Rights

https://openideo.com
http://whr.org.np

Ashwin Gopi, New York University

I'm here to talk to you about a project that existed three weeks ago, but it has completely changed. A couple of years ago we were on this online platform called OpenIDEO, which is an open innovations platform where the members try to tackle social issues. There was a question asked by the United Kingdom's International Development Fund and the question was very simple: How might we empower women in low-income neighborhoods to be able to make their own decisions to become independent, and how do we improve their safety overall? We came up with this idea to create informal networks that are self-sustaining that help women support each other in terms of education and skill sharing in order to make them economically independent.

Initially the idea itself was generated on this online platform by people who came together from all over the world to create this open-source toolkit that anyone in the field could take and implement. They evaluated it and refined it, but still we could not find anyone who could actually implement it on the ground, so we reached out to an online community of not-for-profits that are focused on women's issues.

Last year we had someone at a not-for-profit organization called Women for Human Rights in Kathmandu, Nepal who was very interested in taking this project forward. We reached out to them and have been working with them, and we have been building these informal networks, both online and offline. They have been very helpful in creating connections and acting as our researchers and our middle men to help us connect to women in rural Nepal who don't have Internet connections. The volunteers online have been mentoring them in different skills and teaching them how to teach so that they can go out to these places and teach these women basic skills, and also connecting them to broader international markets to help them sell their handicraft products.

This community and this network existed

Ashwin Gopi

CCLA retweeted
Andrea Wiggins @AndreaWiggins · May 7
Gopi's concept: develop self-sustaining informal networks for
empowering women in developing countries. #crowdcon
↩ ⟲ 1 ★ 1 ⋯

until three weeks ago when, as most of you know, there was a massive earthquake that shook the very foundations of everyday life in Kathmandu, and suddenly our project didn't make sense any more. I was supposed to go to Kathmandu three days from now, but that has been cancelled because it doesn't make sense now. However, what remained was still this network of people. Now, even though the tasks they were initially invited to participate in have changed, even though the purpose of the project has completely changed, still the network of people remains.

What happened is that in working with these people over a year, the people started getting names and faces and stories, and it is very hard to walk away from a project when you have that kind of emotional commitment. And these are people who have never met and will probably never meet in their whole lives. So now the volunteers that are on the ground are still part of an informal network that is still there, and they are helping each other in their relief efforts in distributing food and materials. And there are still people online helping them coordinate their efforts online and doing inventory management and helping them with their social media efforts.

So what initially had started as a crowdsourcing project has now turned into a network of community members trying to help each other in times of trouble. I think that's a very good example of how it is important for us to understand that the abstract crowd as we

think about it is slowly changing. Maybe it's something to do with technology, maybe it's something to do with society itself, but we are starting to become more emotionally involved in these projects.

Biodiversity Heritage Library

http://www.biodiversitylibrary.org

Trish Rose-Sandler, Missouri Botanical Garden, Biodiversity Heritage Library

I work at the Missouri Botanical Garden in St. Louis and was hired there to work on the Biodiversity Heritage Library (BHL). BHL is a consortium of natural history museums and botanical garden libraries that come together and digitize their collections and put them up online for open access as part of a global "biodiversity commons." Today at our portal we have about 46 million pages of text that we have digitized about plants and animals. So we have this critical mass of data and wanted to do things with it, and crowdsourcing was a way to do that.

We have been engaged in crowdsourcing activities for several years. As part of our day-to-day digitization activities, we have folks who give us feedback on the types of content they would like us to scan as well as helping us to correct bibliographic metadata. We also use crowdsourcing to tackle particular data access challenges through grant-funded projects. We currently have three and I will give you a brief synopsis of those. We have two from IMLS and

one from NEH. One of the IMLS-funded grants is called Purposeful Gaming and BHL, and in that project we are testing the effectiveness of gaming for crowdsourcing OCR text correction. Tiltfactor is designing the games for us and those games are going to go live in late May or early June. The other IMLS grant is called Mining Biodiversity and this was part of the international Digging into Data Challenge. That project is using crowds to verify the accuracy of semantic markup of text that was done by automated algorithms.

And then we have a third project funded by NEH called Art of Life, which just ended last month. This project came about because in addition to these millions of books and journals, we have within them beautiful natural history illustrations which are not findable right now because we don't have any metadata about them. The Art of Life goal was to develop algorithms to find out which pages had illustrations, and then we crowdsourced the classification and description of them.

The two key points that we have learned from all of our different crowdsourcing activities have been that crowdsourcing is a really effective way to accomplish a task that you otherwise just couldn't do with the limited resources that you have, and it's a great way to engage people in dialog about your content. But you also need to be aware that crowdsourcing puts a strain on your staff time. There is the need to either go out and find a tool that exists and adopt it or build your own. There is

the need to figure out how to use a tool with your local system and how to get data in there and how to get data out of there, which can take a lot of time. Then, of course, there is bringing that data back into your local system and figuring out how to blend it with the data you have. Finally, when you open things up to users you have to be responsive and you have to account for the time you have to spend answering their questions and addressing their problems.

Questions, Answers, Discussion

CROWD OR COMMUNITY: TARGETS, REALITIES, FUTURE GOALS

- I'm going to kick off the Q & A with my own question. Earlier we were discussing the notion that we are not necessarily trying to appeal to the crowd, but rather different communities, so my question is threefold. Before you launched your respective projects, who did you imagine the crowd would be? Number two, who did it end up being and how did that surprise you? And three, what other crowds or communities would you like to see better engaged in the future? So that is my three-part question for each of you: Who did you imagine, who actually showed up, who would you like to get in the future? • Brett Bobley, NEH

Communities from Kathmandu to Tanzania Reaching Out to the Project

- Initially when we started this project the

Brett Bobley, moderator

Attributes and Motivation of Long-Term Volunteers

In terms of who we are trying to look at for the future, I'm still figuring that out. I think we se that it takes a certain personality who is going to be a long-term volunteer. A lot of people are interested in the subject matter and are willing to contribute their time, but there is a core group of people who are willing to stick with this, transcribing hours a day for years. We have people who have been with the program since 2009. I am really interested now in what is it that motivates them to stay with it for that amount of time. • Jessica Zelt, Patuxent Wildlife Research Center, North American Bird Phenology Project

people we were trying to reach out to were people in education, and people and groups specifically interested in and focused on women's issues locally. The reason that we started this project was to tie with each local community and to work with low-income neighborhoods in Brooklyn. However, because of the digital nature of this platform it developed, spreading out from this. We never planned to work with anyone in Nepal in the first place; our goal was to work with someone in our own neighborhood, so it was a big surprise when someone from Nepal reached out to us. We had no idea of how to go about implementing it. What was really interesting was that soon we started getting a lot of local volunteers in Kathmandu who were willing to work on this project and joining this platform to provide us with more contacts and information.

Up to a point, even though we were in New York, we were able to connect NGOs together. For example, there were students in a town near Kathmandu who focus on women's health and sanitation issues. We were able to connect them to this group in Kathmandu and they are now working together. Surprisingly even now, after the earthquake, they are still working together. So that was a big surprise to us, to find that this project was growing. It is way bigger than the initial group.

Now there is a group in Tanzania that is willing to take this project forward and implement it in their community. We also have a lot of people here in the United States who are interested in taking on the learning and who help to refine this open source toolkit to make it context-independent. Then there are now people in Tanzania who are very interested in joining the project in order to convert this context-free idea into something that will work on the ground. And the funny thing is, we never reached out to them, they reached out to us. I think that is the one surprise to me, the fact that people are willing to reach out and the motivation is intrinsic, it comes from within them. The call is not the same as it used to be. Now we have people reaching out to us, so it has flipped on us. • Ashwin Gopi, New York University

Multigenerational Birders & History Buffs

• I think when we were first scanning our records we were imagining the core group of birders—people who are really involved in their local bird group or people who are already involved in some of the other online bird collection databases—so that was the community that we targeted initially. We ended up with a lot of people we hadn't targeted, people who were really more interested in the historical aspect of our records than they were the birds, so that was a big surprise. Also, birding passes from generation to generation, people teach their children how to bird. People were finding family members who had contributed to the

original records, and they became a core group of volunteers for us as well. • Jessica Zelt, Patuxent Wildlife Research Center, North American Bird Phenology Project

**Tracking and Attracting
Volunteers on Various Platforms**

• I'll talk about our Art of Life project, which is the one that is finished, so we have audiences I can talk about in relation to that. What we did for the classifying description was to put it out to different platforms, including Flickr and Zooniverse, and we did a little bit in Wikimedia Commons. We primarily ended up using Flickr and Zooniverse as our platforms.

Flickr is interesting because you don't really have ways to interact with your users and it is somewhat anonymous. You don't necessarily know where the tags are from unless folks have added some comments to the images as well. I was looking for a particular species in our Flickr stream the other day, and I noticed there was one user who had been doing a whole lot of tagging. Obviously that person was really interested in that species and was spending her time doing exclusively that species, and that's great.

Now with Zooniverse it is a whole different ball game. Zooniverse is great because it has a talk functionality, which means that people can make comments on particular illustrations, they can ask questions, they can pass judgements, so they can aggregate together
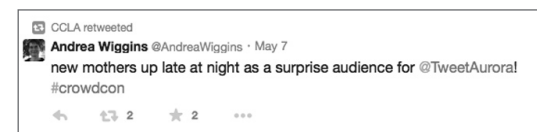
on a certain topic. There is a whole lot of dialog that goes on in Zooniverse, which has been really amazing, and we get to find out what motivates these users and what they are interested in.
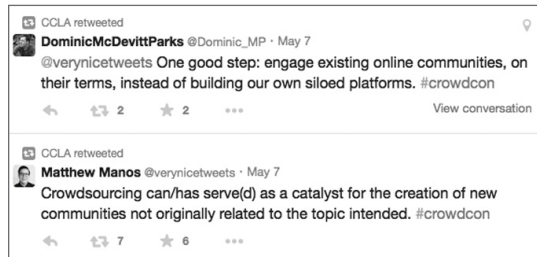
With our Zooniverse project, which is called Science Gossip, we partnered with another group which was focused on 19th century periodicals. We took a selection of BHL content, just 19th century periodicals, so obviously we drew crowds that had that specific interest, and that's where we were getting the word out about it when we launched it live. And then the BHL community was coming there as well. What was interesting was that we also got people from other Zooniverse projects. That is what is great about Zooniverse. You have these multiple communities coming together on one platform, so you just broadcast your project out to all of these people and the reaction is, "That sounds really cool, I'm going to check that out," and you draw them into your project. So there is a lot of cross-pollination that goes on with Zooniverse.
• Trish Rose-Sandler, Missouri Botanical Garden, Biodiversity Heritage Library

**Aurora Community:
Experts, Photographers, New Mothers**

• Initially we didn't know who would be into auroras. There are people who see it a lot, in Alaska for instance, and there are these rare events when it can be seen farther south and there are people who are really good at see-

CCLA retweeted
Andrea Wiggins @AndreaWiggins · May 7
new mothers up late at night as a surprise audience for @TweetAurora! #crowdcon

CCLA retweeted
DominicMcDevittParks @Dominic_MP · May 7
@verynicetweets One good step: engage existing online communities, on their terms, instead of building our own siloed platforms. #crowdcon
↩ ⟲ 2 ★ 2   ···   View conversation

CCLA retweeted
Matthew Manos @verynicetweets · May 7
Crowdsourcing can/has serve(d) as a catalyst for the creation of new communities not originally related to the topic intended. #crowdcon
↩ ⟲ 7 ★ 6   ···

CCLA retweeted
Daniel Powell @djp2025 · May 7
@sleonchnm Co creation is a good one. I like commons based peer production (word, but accurate for many circumstances). #crowdcon
↩ ⟲ 1 ★ 3   ···   View conversation

CCLA retweeted
Matthew Manos @verynicetweets · May 7
Could crowdsourcing become a platform for "family game night" in the future? Bonding/connecting through crowdsourcing #crowdcon
↩ ⟲ 4 ★ 4   ···

ing it there. We have been circling with how to market and target this most effectively because it is inclined more towards rural areas where you don't have as much light pollution. Beforehand there had been a number of informal groups on Facebook, people who are photographers who are sharing lots of photos informally. Those people are from very diverse backgrounds. One of the groups was started by new mothers who were up in the middle of the night and got in touch with each other about going to see the aurora in Alaska. There are certainly people in those communities who have not heard of citizen science—they want to share their photo on Facebook.

Our thought was that we want to be able to help people be able to see the aurora. We looked at Twitter and there are a lot of tweets with wonderful photographs of the aurora, and lots tweets about people's bucket lists. But I think first the experts are helping us and each other, so we are building community from the ground up, and that's great. It's an evolving process. • Liz MacDonald, NASA, Aurorasaurus

Envisioning the Online Volunteer Shelley-Godwin Archive Community

• We don't yet have the participatory part of the archive up, but the community we imagine includes people in the classroom studying Romantic Era works, and people in the public. The interest in Percy Shelley,

and especially in Mary Shelley is astonishing. When we put up the *Frankenstein* manuscripts, within 24 hours we had over 60,000 unique visitors to see it, and it was all over social media. The surprise I had just from looking at Google Analytics on that was that a large percentage of people who came to see the *Frankenstein* manuscripts were from Latin America and from Eastern Europe. That automatically told me something about the reception of *Frankenstein* that I didn't know, and in turn raised some interesting questions.

So I can't tell you yet who will actually come, but I promise I will come back and tell you once people are participating. • Neil Fraistat, MITH, University of Maryland,

COMMUNITY MOTIVATION TO LEARN

• We have talked about communities learning from these crowdsourcing projects. I wonder if you have any sense that your communities actually want to learn when they get into these projects. We certainly see that if we introduce too much content early, on the way in, it puts people off. I'd love to hear about your experiences with this. • Chris Lintott, Zooniverse, Oxford University

Shelley-Godwin Archive Pilot & TEI

• We did some piloting of participation using University of Maryland and University of Virginia graduate students. They checked transcription and encoded in TEI 100 pages

of *Frankenstein* manuscript. The kind of feedback they gave about their experience was astonishing. Because we are interested in the genetic development of the work and page, they have to be really, really careful. And that gave them understanding about the manuscript that they would never have seen with the novel if they just went in and looked at the printed edition. The by-product of that was that we trained about 20 students to have a pretty significant knowledge of how to use TEI. That is a very small pilot, but that is the kind of knowledge that is a payoff. • Neil Fraistat, MITH, University of Maryland

• But my point was, did those people go in thinking that they wanted to learn that? Learning occurred, but what if you said to them, "Do you want to do an exciting project and learn TEI?" The answer may have been very different. • Chris Lintott, Zooniverse, Oxford University

• We sort of did, but we didn't force anybody to participate, so they did choose to be in the project. But you're right, this is not a more open-ended, who-chooses-to-come approach. • Neil Fraistat, MITH, University of Maryland

Covering the Broccoli with Cheese: Social Forums and Interaction Make Education Palatable

• That is a very good question and we are still trying to figure that out in terms of surveying what people know and what they are looking for. I will say anecdotally that you are kind of covering the broccoli in cheese. You want slip in some educational content and create that in a very engaging way, and I would love to see best practices for doing that across a variety of disciplines. People seem to be more receptive to that in a social forum kind of way, and when the questions are generated by the group. A question will come up about a specific kind of aurora and they are very interested in responding to that and discussing that. It is when they are really getting interactions with a group of scientists that seems to be working. • Liz MacDonald, NASA, Aurorasaurus

BHL: Giving the Crowd the Opportunity to Wade Through Full Periodicals

• We had this interesting conversation working with Zooniverse before we launched about whether we should put up all pages of a periodical or just the pages that had illustrations, because the point was we wanted to know what those pages were. From some information that we got from Zooniverse, they felt the crowd would be much better at picking those out, so let's put up everything. We had mixed reactions from users. Some were, "Why do I have to look through all of these pages of text to get to the really cool illustrations?" Other people seemed to really enjoy the text because it gives you context for the illustration, and some people loved that. We are talking about filtering

Bird Phenology: Balancing Worthwhile Contribution with Personal Interest

• I've found that a lot of our participants have come because of the overall goal that our program is trying to achieve. They are not so interested in the records themselves. They want to know that the records are being used and how they are being used, but really they want a summary of that. They are interested in what birds people have seen and in what locations; they're interested in the original observers and what their lives were about; they are interested in some of the historic locations that are no longer current locations. I find that a lot of people take the time to go off on their own and Google those kinds of things and they want to share that. I get a lot of emails saying, "Look at this funny art," or "Look what this person did."

They are interested in the records and want to know it is happening, but they want the summary. They want to know who used it and at what university and what was the finding, and that's enough for them. They can then say, "Okay, the information is being used, so it's worthwhile for me to do." So on a smaller level, what keeps them involved on a daily basis is certainly the records. • Jessica Zelt, Patuxent Wildlife Research Center, North American Bird Phenology Project

out some of that text now to make it easier and more efficient for people, but we do wonder whether some people will be upset if we take that text away. • Trish Rose-Sandler, Missouri Botanical Garden, Biodiversity Heritage Library (BHL)

Water Access: Learning from Communities with Aligned Goals

• I want to talk about a specific instance. There is a school in the Philippines working on creating a water filtration system with ceramic pots. They reached out to us because they wanted to specifically learn something about the context of rural Nepal in terms of access to water and access to clean water. They provided a lot of very interesting insights as to how their project was addressing the problem in the Philippines, but they came to us because they wanted to expand. We were able to get people in the field in Nepal to ask very specific questions regarding water so that we could provide them some value. So in some cases people come if they have goals that are aligned. They can come to the platform with very specific goals of learning. • Ashwin Gopi, New York University

SUSTAINING COMMUNITY CONVERSATION

• My question is about sustaining communities in conversation. The lifeblood of any community is an ongoing conversation and interaction. I would be interested in hearing how you have thought about that, how you would develop that community conversation

and interaction. Did you plan for it? Develop it? Did it take additional people (though I guess it depends on the scale of the project)? And then how successful were you, and were you surprised by anything along the way? • Katie King, University of Washington, Seattle

Picking the Right Platform; Communicating with Visuals

• Initially we started out interacting only on the online platform on which the idea began. However, the platform is not very conducive to that kind of collaboration. It is good for editing text together, but no way in terms of coordinating efforts, so we decided to have the conversations outside of the platform. We decided to have a Google Hangout meeting. We didn't expect anyone to join, but we just put the link there and knew that we would change the link once a week. Slowly people started joining and it grew very organically and people started to join the platform.

There is one thing we did that was weird. Normally on online platforms it is very hard to connect directly to the purpose; it is very hard to see the direct impact of your work. But, randomly, we had this Polish filmmaker who was backpacking through Kathmandu at that time and he wanted to volunteer. I have no idea how he found out about us. He actually went to the villages and took pictures and videos. He interviewed the women and he put up these interviews on YouTube.

CCLA retweeted
Trevor Owens @tjowens · May 7
.@chrislintott learning often isn't the users objective, it comes as an outcome of participating in meaningful work/activity #crowdcon
↩ ♺ 12 ★ 10 ····

CCLA retweeted
Matthew Manos @verynicetweets · May 7
So we have all of this "stuff" from crowdsourcing - now what do we do with it? How can we keep track of what it inspires? #crowdcon
↩ ♺ 1 ★ ····

That helped to kind of humanize the efforts that people were doing. It is easier to share something that is so visual, and it's easier to identify with this project once you see the direct impacts of the project. That was great for getting participation, and after that the amount of participation increased. There was a palpable difference. • Ashwin Gopi, New York University

Budgeting for Volunteer Coordinators; Looking to the eBird Model

• I want to credit Andrea Wiggins for helping us figure out how to budget for and plan for volunteer coordinators for our project. We also looked at some of the other citizen science programs, eBird in particular is one that I'm thinking of, and how they did that in terms of groups of super users and expert

DEFINING SUCCESS

• I am really interested in your view of success and how you've defined it for your projects. I imagine it could be quite quantitative in some cases, but not in others. And also, how did you measure it? • Austin Mast, Florida State University, iDigBio

Useful Data, Published Results

• There are many, many different measures of success, but for me, for our projects, showing that the data we are collecting can be useful for both scientists and the community members is one of those goals. I guess we want to measure that through publishing those results. • Liz MacDonald, NASA, Aurorasaurus

Tags, Exposing New Audiences, New Discoveries

• I don't really know if we define success, other than one tag is better than no tags. As many tags as we can get would make us happy. But one of our goals was to expose content to new audiences, and I certainly think it was successful in that. We are pushing out to platforms outside the BHL community where people were

seeing the cool stuff we have and getting really excited about it. Also, when you push out to new audiences they have different interpretations from people within your own domain, so those tags that they add are very valuable for people outside your community who would probably look at things differently from people inside your community.
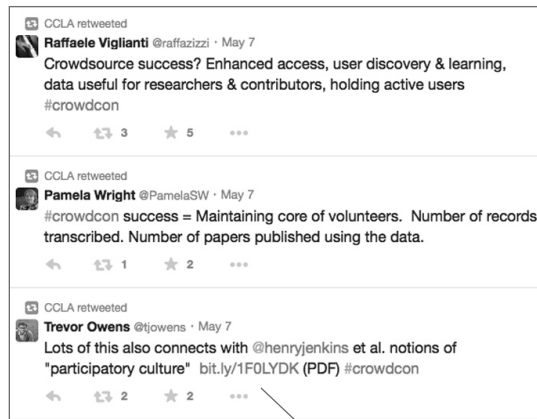
One thing that was interesting in terms of successes was unearthing new discoveries. Sometimes people would come across a letter or an illustration that turned out to be kind of seminal in the field. For those people who have more time to dig into this at a deeper level than you do, supporting them by digitizing en masse and working at a much higher level, people working at a case level or an image level just brings a whole new context to it. • Trish Rose-Sandler, Missouri Botanical Garden, Biodiversity Heritage Library

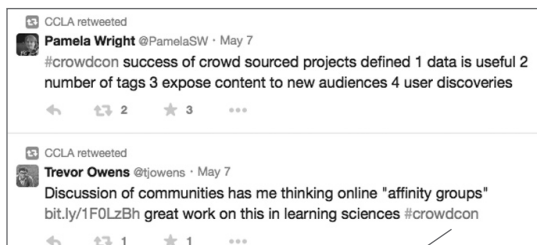**Numbers: Volunteers, Records Transcribed, Research Articles**

• For the Bird Phenology Project, I've been gauging success through engaging the public, and also the

scientific data that is produced for the public and for research. My goal for the volunteers was to maintain a consistent group of volunteers who would stick with the program. There are the volunteers who will transcribe every day, and then there is a different group of volunteers whom we actively are trying to engage all the time. There are people who are coming in and transcribing for a few days or a few months and then stopping, so I am always trying to either get them back or get new people involved in the program.

We have 1.2 million records online, and all of them have now been transcribed twice, and we have about 370,000 records that are available to the public, so we gauge a lot of success based on those numbers of records and what stage they have gone through. And also the scientific research. We use the data now, and it takes a lot of time to work with the data once it is transcribed, but we have published about five articles using that data. • Jessica Zelt, Patuxent Wildlife Research Center, North American Bird Phenology Project

> **CCLA** retweeted
> **Raffaele Viglianti** @raffazizzi · May 7
> Crowdsource success? Enhanced access, user discovery & learning, data useful for researchers & contributors, holding active users #crowdcon
> ↩ ⇄ 3 ★ 5 •••

> **CCLA** retweeted
> **Pamela Wright** @PamelaSW · May 7
> #crowdcon success = Maintaining core of volunteers. Number of records transcribed. Number of papers published using the data.
> ↩ ⇄ 1 ★ 2 •••

> **CCLA** retweeted
> **Trevor Owens** @tjowens · May 7
> Lots of this also connects with @henryjenkins et al. notions of "participatory culture" bit.ly/1F0LYDK (PDF) #crowdcon
> ↩ ⇄ 2 ★ 2 •••

http://bit.ly/1F0LYDK

> **CCLA** retweeted
> **Andrea Wiggins** @AndreaWiggins · May 7
> measuring success is challenging—some tools for evaluating learning, evaluating value of services & outputs rendered much harder #crowdcon
> ↩ ⇄ 2 ★ 1 •••

> **CCLA** retweeted
> **Pamela Wright** @PamelaSW · May 7
> #crowdcon success of crowd sourced projects defined 1 data is useful 2 number of tags 3 expose content to new audiences 4 user discoveries
> ↩ ⇄ 2 ★ 3 •••

> **CCLA** retweeted
> **Trevor Owens** @tjowens · May 7
> Discussion of communities has me thinking online "affinity groups" bit.ly/1F0LzBh great work on this in learning sciences #crowdcon
> ↩ ⇄ 1 ★ 1 •••

http://bit.ly/1F0LzBh

birders and how that might translate. In our case one thing that has worked well (and this is a work in progress) is that we have a network of scientists who are across the country volunteering to translate our field into interactions for the public and our users. This is an obscure field and there aren't that many outlets if you study auroras, and people were really interested in getting the opportunity to do that. I would love to scale that up with the volunteers as well, who are very capable and involved, and figure out how to do that.
• Liz MacDonald, NASA, Aurorasaurus

LONGEVITY: WHAT HAPPENS WHEN THE GRANT RUNS OUT?

• My question is about the longevity of these projects. A lot of them seem to be grant-funded, so they only have a budget for a specific period of time. Once the grant funding runs out, what kind of planning do you do to make sure that you can continue the project or at least continue the availability of the data that came out? • Alexis Rossi, Internet Archive

Self-Sustaining via the Members

• You're right, actually. Our project had initial funding for only six months. Afterwards we had to find our own source of funding. What we did was try to create a self-sustaining network, and this is in terms of the off-line networks that we have created. With these women we were working with, the idea was to create a co-op model, but also part of their income would go back and sustain the project. We were also trying to get people more engaged so that the ones who benefited from this project would also then go out and volunteer themselves to go and perpetuate the system. So the idea is for the members of the community to sustain the project, first of all because we don't want to be trying to get grants all the time, and also they are not the easiest people to work with. We would rather be completely independent, so the goal is to get members to take responsibility. • Ashwin Gopi, New York University

Graceful Degradation; Ongoing Availability of Significant Data

• I'm going to take this question from another end. It's something that emerged in our "Signals from the Future" discussion, which is that we also need to think about graceful degradation of our projects. Not all of them will live forever, but the ones that have significant data, that's something to think about. What happens? What kinds of processes can we bring into play so they wind up in a graceful way and still have information available to people who want to visit them somewhere? • Neil Fraistat, MITH, University of Maryland

Diversifying the Ongoing Funding; Evolving Technology

• Our project is funded by the NSF INSPIRE program, which is for interdisciplinary, in-

novative ideas, and I highly recommend it to any of you who might be interested. Your question is a very good one, and in this case we are very fortunate that multiple aspects of the project are funded. Citizen science projects tend to be large and going in many different directions. In the future, there aren't many funding sources that would encompass all of those directions, so we are probably going to look for the educational and citizen science aspects in one area, and the space science research in another area, and so on. But I also think that this current technology solution is ephemeral and everything will need to evolve. Now that we have this we can move on with the next technology and be ready for the next solar mass ejection when everything will probably be completely different a couple of years from now. • Liz MacDonald, NASA, Aurorasaurus

Integrating External Tools with Internal Operations and Architecture

• I'm thinking again in terms of our Art of Life project and the image tagging. We don't have an end date right now for the Zooniverse project, but we will need to at some point. With Flickr, we can just keep stuff on there forever as long as Flickr exists, and we will always digitize new images and put more stuff up there. But we know we need to also get stuff out of there. The original goal was to increase access within our own local system, so we want to pull that data out of Flickr and Zooniverse and make it searchable there. Part of it is figuring out how you build these tools that are external into your day-to-day operations and your architecture. • Trish Rose-Sandler, Missouri Botanical Garden, Biodiversity Heritage Library (BHL)

Crowd Consortium Goal: Addressing Sustainability Issues

• I'm hoping that any kind of Crowd Consortium that gets built will be useful in thinking through sustainability issues and coming up with strategies for sustainability. • Neil Fraistat, MITH, University of Maryland

CCLA retweeted
Victoria Van Hyning @VanHyningV · May 7
Great Q @alexisrossi: how do we sustain projects? So many are short-term and/or grant funded. Lots of interesting answers. :) #crowdcon
↩ ⇄ 1 ★ 2 •••

CCLA retweeted
Raffaele Viglianti @raffazizzi · May 7
@fraistat brings up issue of graceful degradation: what happens once a crowdsourcing project is over? Also #DH projects in general #crowdcon
↩ ⇄ 1 ★ •••

"We need to…"

"We can't…"

"How do you…"

Photos in this section:
"Big Challenges" discussion groups in session

IDENTIFYING THE BIG CHALLENGES

Introduction

Jake Dunagan, verynice.co

For our next session we are going to look at the big challenges: the barriers, the needs, the things you are facing. Try to express them as clearly as you can in your working groups, recording them on sticky notes. Come up with a minimum of ten and rank them in order of importance in order to distill what are the most important, urgent challenges you are facing. You are going to report on the top three in plenary. Use the suggested question template [at left] or you can come up with your own. Begin by being open and brainstorming. Today is about expansiveness, about coverage, more than it is about precision and getting exactly the right thing. So just get the big ideas down and then as a group decide on the most important ones.

Group Report-Outs

Ed. Note: Group numbers below are used only to reflect sequence in which presentations were made.

GROUP ONE

Reporting: Alexis Rossi, Internet Archive

• How do we create a critical mass of engaged and trusted participants for our projects?

• How do we preserve and access the 20th century, meaning how do we get over copyright concerns in order to do our job?

• How do we include new communities, meaning people who might be underrepresented both in the collections of things that we're looking at and in the communities of people who might be participating?

GROUP TWO

Reporting: Mary Flanagan, Tiltfactor, Dartmouth College

• Move crowdsourcing from the edge to the core in the day-to-day workflow: This is about how, in institutions, we can legitimize crowdsourcing not as a special project or unique experiment but one that is part of the future of how we might want to continue to work and integrate the public.

• Connect education to engagement really thoughtfully: how systems connect in different levels of both higher education and K through 12 and help further what we know and also develop new audiences for source materials, etc.

• Training, education and documentation: best practices and case studies. So many people want to know about what's out there; reading a website doesn't really help us understand the challenges, even how to install something. Getting over some of these technological hurdles would really be great to do as a kind of short-term strategy, as
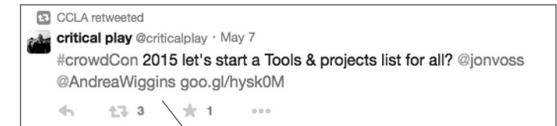
well as a longer-term strategy about how we create packages of tools and technologies for people.

GROUP THREE

Reporting: Chris Lintott,
Zooniverse, Oxford University

• Lots of people are worried about the rise of mobile for transcription-type projects because they are restricted to the smaller screen, but also because of the cost. You're not only developing this for the web, but also for apps and so on, which is much harder to do, and in a space where technology is changing.

• Recognizing failure and talking about it, as opposed to saying that all of our projects are wonderful. Linked with that is the idea of in-ternally framing our projects as experiments rather than as complete systems. So instead of just saying, "We are going to produce this data," or "We are going to produce this information to add to our collection," or "We are going to enable this audience," we say, "We will also learn about how best to build these projects." We are not yet at the point where we know what best practices are, and yet we spend a lot of time trying to convince funders and others that we do.

• Working harder to get buy-in across institutions, particularly in museums, libraries and archives, getting all parts of an institution to buy in. For example, in a museum, collections and education are different departments, and yet you have projects that span both, which could equally be developed. That has very different needs from,

CCLA retweeted

critical play @criticalplay · May 7
#crowdCon 2015 let's start a Tools & projects list for all? @jonvoss @AndreaWiggins goo.gl/hysk0M
↩ ♺ 3 ★ 1 ⋯

http://goo.gl/hysk0M

say, the archives and so on. It is finding ways to frame these projects so that everyone can get behind them and support them.

GROUP FOUR

Reporting: Tom Blake, Boston Public Library

- We thought we were being clever because we thought everybody was going to talk about defining success, so we said, "Let's talk about failure." That was just brought up, and I think we have a little more to add to that. It's not just recognizing that failure is a challenge, it is really defining what that means. Is it catastrophic failure? Is it, "Oh, this project didn't work, we can't do anything with the data"? That's a failure. Or maybe that's not a failure, maybe that's just a zero-sum game.

  We connected that to: How do we understand and navigate the true risks associated with the release of certain types of data? What is truly detrimental? We think this is all fun and maybe there will never be any harm, but maybe there will be, and you have to be aware of that to survive as a community.

- The second challenge is similar. How do we align our practices with the eth-

ics of labor and research? This would include, just for clarification, things like attribution and provenance of the data, giving credit where credit is due. We talked about making sure that our communities actually get something out of this as well, rather than just us harvesting the fruits of the crowd.

- How do we develop and evolve our communities into peer producers? Does the crowd always have to stay the crowd? Can we understand and develop different levels of expertise? Can we create projects that manage themselves? A lot of us are project managers. Is it possible to have a crowd manage itself?

GROUP FIVE

Reporting: Michael Haley Goldman, United States Holocaust Museum; Andrea Wiggins, University of Maryland

- Crowdsourcing communities need to be sustainable, engaging, social and multidimensional. One sub-part of that is about balancing traditional expertise.

- We need standards, protocols and interoperability. One of the examples of that is about integrating data into a collection.

- We need modular tools for common CMS to accomplish basic functions. The basic idea is that there are a lot of different platforms that are in use. Rather than rebuilding all of the basic functionality that every open source content management system already has and we all need, why not just build some

plug-ins for crowdsourcing projects, and toolkits and guides for how to use them?

GROUP SIX

Reporting: Nick Adams, University of California, Berkeley

- One very actionable item would be coming up with some shared templates for IRBs, for privacy and terms of use, so we don't all have to write those individually.

- Another is creating tasks and platform models that appeal to user communities with multiple motivations. Some people want to focus on a broad, contextual path to do the work, others just enjoy doing little microtasks, and somehow they should be able to choose that so that they're happy while they're working.

- Another challenge is effectively verifying data using algorithms, somehow combining digital humanities and computer science in a way that works for both communities.

- I brought this up in my group and it didn't get a ton of uptake, so maybe it's not a big deal. I am concerned with getting the data used and I've heard a few other people say that. We can produce a lot of stuff that nobody then uses. • **Jen Hammock, Encyclopedia of Life & Smithsonian Institution**

- This didn't quite make the cut, but I do think it's really important to focus on legal rights and responsibilities of data collection in working with communities from a global perspective, not just a U.S. perspective. • Katie King, University of Washington, Seattle

- In our group we also talked about accessibility in terms of visual impairment issues and things like that, as well as reaching audiences who are not usually online and how we can be successful that way. • Amy Patterson, Wisconsin Technical College

Ongoing Process

- In later working groups we will be digging into these challenges to define them and to explore the territory. We want to map the landscape with a little bit more detail. Think about which of these challenges you are most passionate about and want to work on. • Jake Dunagan, verynice.co

Additional Input

- I'm going to open the floor for one or two additional comments. Are there any glaring omissions that you see? Is there a domain that wasn't discussed at all that should be on the board? • **Jake Dunagan, verynice.co**

Additional Challenges

These additional challenges were generated by the working groups but didn't make the cut when groups prioritized challenges and picked the top three for report-out. They are offered here to show the full array of participant thinking and to serve as fodder for ongoing Crowdsourcing Consortium development. They are in no particular order and are taken directly from the original sticky notes.

- New models for platform
- Reaching niche communities
- Data round trip
- Data ambiguity
- Verification
- Bridging DH + computer science/Kaggle
- Context or not
- Multiple motivations
- Trust
- Privacy - transparency
- Saturation
- Accessibility (visual impairment, etc.) + online/offline
- Build trust in external code, expertise, etc., etc.
- Authoritative vs. non-authoritative data (institutional vs. community)
- Shutting process down
- Beyond microtasking
- Cross-linking
- Sharing data — related data
- Data output (to right people, right format)
- Long-term preservation
- Reaching audiences who are not (easily) online
- Getting to where people are (e.g., museum queue)
- Finding out how to get started (as a project builder)
- Crowdfunding is complicated (& not the same)
- Sharing data — user data
- Should we limit participation? (24,000 images a day)
- Readability of cursive
- Build community expert support through continuing ed. & training process

- Community engagement
- Community engagement managers are seen as "optional"
- Connect! Advertise crowdsource projects
- Are you really sharing?
- Better tech documentation and instructional videos
- Micro-course on crowdsourcing - cont. ed. - cross disciplinarity
- Document — case studies — YouTube video
- Document best practice for use of crowdsourcing in education
- Engage specialist/technologist around crowdsource project
- Better tech documentation & training on crowdsourcing tools/platforms
- Knowledge transfer of tools from one domain to another
- Interoperability of crowdsource platform
- How do we enable interoperability and still program for serendipity?
- Using crowdsourcing for students — targeted market - "Makin' it Real"
- Shared infrastructure platform
- Who owns it?
- Re-use, recycle crowdsourcing tools and platforms to other projects
- Tell the story — leads to layered education
- Changing culture of institutions to accept "outside" information
- Connecting the data produced with data consumers
- How do we demonstrate net effects/impacts from crowdsourcing?
- What are our success benchmarks for common tasks? Helps to determine success & failure and helps planning.

- How do we establish benchmarks of best practices without "over-organizing" into a movement that is no longer agile or innovative?
- Building effective & dynamic (yet low cost) data visualization tools
- Balancing the tension between the scientific/ program goals & volunteer needs & expectations
- How do we attract & work with coders?
- Sustainability: long-term involvement of participants
- Crowdsourcing as a dirty word, digital labor
- Help the crowd fund & manage resources relevant to their project. Introduce librarians to the crowd?
- Evaluating return on investment
- How to plan research with an uncertain timeline for data collection
- Think about impact of data we are releasing
- Ethical/legal limits for crowdsourcing sensitive materials
- Inclusion in projects that matter to a wide range of topics — e.g., cross-language, cross-culture, cross-educational ability, income inequality, technology inequality
- Engaging diverse communities (really engaging them)
- Work with communities before launch/from the start ... not assume "the public"
- Not a linear trajectory from a "crowd" to a "community"
- How do we grow online/in-person communities?
- How to survey the crowd without alienating them?
- We need best practices to wind down projects

- We need to create best practices for many things, including interface design, user engagement strategies, data cleaning/normalizing, etc.
- We need to establish protocols, standards, for interoperability across projects to give projects a chance to leverage web services, etc., across community
- Human side of machine learning loop: How do humans learn in this loop?
- Educating people about their attribution
- How do we ID gaps in tools & survey landscape?
- How do we align our needs with those of participants?
- How can we understand different definitions/types of "crowdsourcing"?
- How do we give credit for meaningful contribution?
- How can we protect anonymity on sensitive topics?
- How can we foster unexpected communities (e.g., historians in citizen science project)?
- Need to understand legal limits & responsibility
- How do we fill gaps in digital access?
- How do we let people just fix issues in OPAC records? (Decentralized correction of data across the web)
- We need new types of metrics (not just papers)—individual benefit, improvement to performance, actual use of data/collection
- How do we ensure/secure sustainability?

- How do we help participants produce quality data? (Access to tools, instruments)
- Data deposit wizard for *individuals* with pingback tracking. Whose data gets used? Acquisition? — quality, motivation, skills/knowledge
- What are realistic expectations for project outputs? At what intervals?
- Encourage use of quantitative methods in humanities

Meghan Ferriter

ROUNDTABLE:
BIG CHALLENGES AND
HOW THEY MIGHT BE MET

Introduction

Moderator: Trevor Owens,
Institute of Museum and Library Services

I am glad to have the opportunity to kick off this panel conversation, continuing some of our earlier activities related to big challenges and how they might be met. We will begin by getting a reaction from each of the panelists about what the big challenges are, whether those are ones we just talked about or are the ones that are on your mind that weren't part of our conversation.

Smithsonian Transcription Center

https://transcription.si.edu

Meghan Ferriter, Smithsonian Transcription Center

Thanks to everyone for being here and for sharing and voting on the big challenges. I continue to hear issues related to sharing information and that is the one that stands out the most to me, whether that is sharing information on how to do what we are doing already really well, or how to improve, how to accept and receive feedback from our peer community here and elsewhere: "You're doing that okay so far, but here's a way to tweak that even better." One thing I would like to continue to do as well is share data that is

being generated through these different processes.

The three challenges that stand out to me with the Smithsonian Transcription Center are: trust, workflow, and acknowledgment. I believe people come to experiences carrying their cumulative experiences, their baggage, with them. Sometimes that can be really helpful in expanding the experience, sometimes it takes a little bit longer to work out exactly what is happening in a collaborative or crowdsourcing project. But sometimes it is like Mary Poppins' carpetbag, with a depth and tools inside that we didn't even know existed. By trusting in the potential and the well-meaning nature of the community, we can tap into some of those hidden reservoirs of skill and knowledge that lie within a community.

We have also seen the results. I think people in this room have subscribed to this idea, to varying degrees: we want to trust the crowd. Now we need to think about ways that we can make a cultural shift toward working with that trust, rather than repeated asking whether or not we should trust the crowd. Trust is risky. Working with trust changes design elements, moves us to more effective tools within our system, and trust helps me communicate with our volunteers in the Transcription Center. It generates and supports the rhetorical approach that I use when giving instructions, improving the words and images and workflows of participation in the project. I also think trust makes authentic communication possible, mostly by becom-

ing vulnerable in that moment. Trust is a vital component of successful public engagement and it also remains challenging.

The reality of crowdsourcing is a call for help—asking people to relieve some of the constraints that projects are experiencing; whether with expertise or in scale. I find workflows and allocating resources at the fore-ground of managing projects. It is a frequent concern heard around preparing projects for the Smithsonian Transcription Center, and something I think about every day. We can con-tinue to break down the steps of these types of projects to consider ways to improve and insert volunteers into the workflow in different ways; not just in the crowdsourcing of information but also in the activity currently performed by staff members as they are working in the data curation element. With this approach we can see an experience that is both small and man-ageable, and connected and coherent overall.

And finally, I think the last challenge is ac-knowledgement, which is intentionally considering the best ways to acknowledge contributions and create many ways for partici-pants, organizations and voices to join. One of those approaches through which I'm constantly thinking: how can we honor the organizational side of the bargain and articulate that through our actions. We are doing our part, you are helping us and we are helping you. It is more than just walking around high-fiving people virtually or celebrating them, which is very important, don't get me wrong! I'd never want

to be quoted as saying that's not important because it is one of the main things that I do! Yet, I think it is also building that acknowl-edgment into the system; finding ways, for example, of noting the crowdsourcing efforts in newly created collection records, as we do in the Collection Service Center.

We also have to come to terms with some of the challenging pieces of acknowledgment, such as respecting the community's new knowl-edge and providing them opportunities to use the knowledge they've generated and created in new ways, and find ways to make products available in less restricted ways - opening the entire process with interoperability in mind and other related challenges.

Finally, I think that acknowledgment requires us being there, listening, thoughtful develop-ment, open communication, and a willingness to go a few extra steps to demonstrate that crowdsourcing is important to us from an inter-nal organizational perspective.

HathiTrust

http://www.hathitrust.org

Jeremy York, HathiTrust

I have just prepared to talk about some of the challenges that we have faced in HathiTrust, and I will just go ahead and do that. I am re-ally glad to be a part of this discussion. To me it is one that is very deep and multifaceted and very exciting and challenging. Some of

CCLA retweeted
Victoria Van Hyning @VanHyningV · May 7
We need flexible UIs, multiple paths into #crowdsourcing and broader institutional engagement. I paraphrase @MeghaninMotion #crowdcon
3 2

CCLA retweeted
Raffaele Viglianti @raffazizzi · May 7
Crowdsourcing as a call for help #crowdcon
1 2

CCLA retweeted
Victoria Van Hyning @VanHyningV · May 7
.@MeghaninMotion identifies 3 challenges in #crowdsourcing: engendering trust between institutions and crowds/individuals #crowdcon 1/3
2 1

CCLA retweeted
Amy Patterson @amycep · May 7
Trust, workflow and acknowledgement: creating trust to tap into hidden reservoirs of skills and knowledge @MeghaninMotion #crowdcon

Jeremy York and Meghan Ferriter
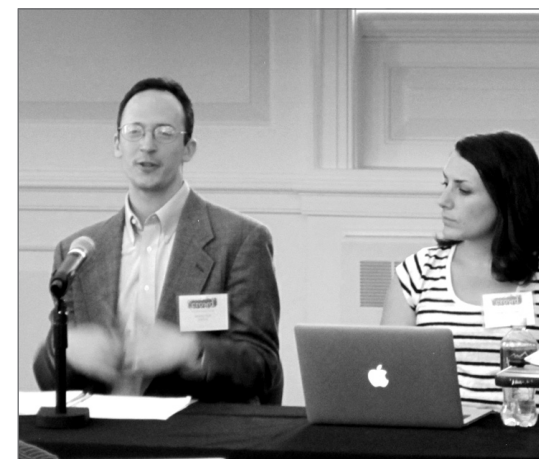
the challenges that we face have to do with incorporating changes that people may make, challenges with liability and authority, and also just with adequate resources.

Here are some quick examples. You may know that a lot of the corpus material that's in HathiTrust was digitized by Google, and they have certain restrictions on the data that have to do with our ability to distribute the data. So a key challenge that often comes up with regard to OCR correction, which everybody would like, is that if we were to take OCR corrections from the community, they would get out of sync with the materials from Google that we actually re-download as they are constantly improving. We have a kind of classic update problem where we might get improved images, but the OCR is off, and all kinds of things happen there. That is a real challenge for us.

Rachel Frick

Another challenge is that, as far as I'm aware, a lot of the use cases for improving HathiTrust have to do not with just one text or one book but with a body of works, often for computational research. And it is often a real project to do that. It's not just something that a person can go and do, it's often a team of people working on it. And they may end up cleaning the data in one way or another as part of the project, so we have another challenge: that people clean the data in different ways based on what they want to do with it.

And I will offer just a quick example of liability. We have bibliographic metadata that is submitted with all of the digital books that

we get, and we make rights determinations based on that bibliographic metadata. When institutions partner with the trust they actually sign an agreement saying that they take responsibility for their bibliographic metadata, meaning if something is opened inappropriately because it was listed as 1917 instead of 1971, they take responsibility for that. It's a real challenge for community-wide engagement and correction that the institution has the liability. They don't want anyone to correct their bibliographic metadata.

And then regarding the resources, a few years ago we were working with Google and it's a real challenge for them that libraries have such varied metadata. We came up with a whole scheme to have people put in their enumeration chronology information (e.g., "Volume One, 1973"). It was beautiful, it was great, but what institution has the resources to undertake that? Those are some real challenges that we've faced.

Digital Public Library of America

http://dp.la

Rachel Frick, Digital Public Library of America

There were some really great conversations this morning and I'm taking a pause to just acknowledge the great ideas that were put up there. There is a question someone brought up, and it is something I was thinking about before coming here: Who owns what? Who owns what part of the process? Who owns what

part of the product? And how do we communicate in a way that facilitates uptake and reuse? I don't think we express very transparently what pieces or parts are reusable, and if we did find something that was proprietary or there is something that is locked down, why? If we do express that, we tend to express it in human ways, and we don't do things that are for machines. An example at DPLA is that we are going to announce tomorrow some white papers around rights. As boring as that sounds, the biggest hurdle for collection release for cultural heritage organizations is being able to express rights in other ways besides licenses because cultural heritage organizations don't feel they have the right to express a license because they might not own that intellectual property on the items in their collections. They might own the item, but not the intellectual property.

We are recommending rights statements that are machine-actionable. For example, if you wanted to come to the DPLA and run a computational query on 10 million records and find everything that is in the public domain, you can't. We did an analysis of eight million records, and of those eight million records we have over 78,000 unique expressions of rights. Right now in our metadata analysis, over 25% of the content in our metadata records describe rights, access and reuse rights, and it is not consistent. That is a big hurdle for reuse and it is not glamorous, it's that functional and operational thing. So are we building up things that we think we are sharing, but if somebody

from another domain comes in and tries to use your system and your tools, there might be things that you've programmed in that are causing some friction. We need to really enable flow.
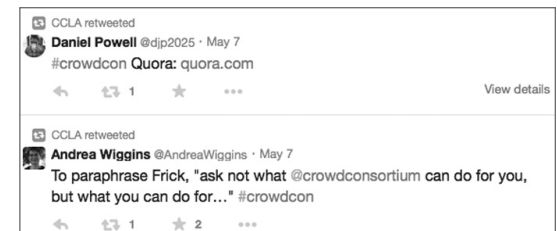
The last piece I have is: How can we check for ego? I think a lot of times we have organizations or groups that want to own the whole problem and the whole solution instead of saying, "I do step two really awesomely." Or, "I built this part of the widget that is excellent, but step one and step three are a little shaky. I saw somebody else's step one that was really great." And we need to document this really clearly. How are we building for the best interoperability of the best pieces of the best projects?

Human-Computer Interaction Institute, Carnegie Mellon University

https://www.hcii.cmu.edu

Jeff Bigham, HCII, Carnegie Mellon University

I come from the perspective of having used crowds as part of building systems for the past few years. We have built system that use crowds and computation to drive robots, to answer tens of thousands of visual questions for blind people, to convert speech to text in three or four seconds for deaf and hard of hearing people. We can code hours of digital video in about five minutes. Most recently we were writing an academic paper from a watch,

CCLA @crowdconsortium · May 7
.@rlfrick: How can we check for ego? Groups wanting to own the whole problem/solution instead of recognizing both success+failure? #crowdcon

CCLA retweeted
Daniel Powell @djp2025 · May 7
#crowdcon Quora: quora.com

CCLA retweeted
Andrea Wiggins @AndreaWiggins · May 7
To paraphrase Frick, "ask not what @crowdconsortium can do for you, but what you can do for..." #crowdcon

Jeff Bigham and Rachel Frick

CCLA retweeted
Andrea Wiggins @AndreaWiggins · May 7
@chrislintott @jim_salmons @tjowens Agree, notion that basics are "figured out" is seeing current art as a science, not there yet. #crowdcon
↩ 1   ↻ 1   ★ 1   •••   View conversation

CCLA retweeted
Andrea Wiggins @AndreaWiggins · May 7
impressed that so many of the "big challenges" raised by panelists focus on respect and benefits for participants! #crowdcon
↩   ↻ 3   ★ 3   •••

CCLA retweeted
Andrea Wiggins @AndreaWiggins · May 7
"What would I be proud of my daughter doing as crowdwork?" - great question to ask to ensure that you're respecting volunteers #crowdcon
↩   ↻ 2   ★ 2   •••

Jeff Bigham

orchestrating a crowd of technical writers behind the scenes.

While I come from this perspective of building stuff, it has led me to talk to a whole bunch of great people who work in this space, so what I did was to tweet out the question: "What are some of the big challenges in crowdsourcing?" I got a lot of great responses and tried to categorize them. There were three big things that came up.

First, how can we teach computers to be as good as experts? While we are all very excited about getting people to contribute to our projects I think that many of us, especially those of us in computer science, are thinking about how we can gradually turn over a lot of that work to computers. I think a lot of the ability to do that comes from how we structure the work that is done by the crowd. It is very valuable, and it is a big first step to collect data that we can use to train machine learning, for instance, but what is the next step? The next big challenges involve figuring out how we can get the crowd to structure a problem so that it is more amenable to machine-learning approaches. How do we get the crowd to figure out what is important about expertise in a particular domain so that the relatively rare experts don't have to define all of that? This is about cost, but it's also about speed; it's potentially about privacy and confidentiality. Once you can get a computer to do it, it's not some person who is seeing you as data, which you may not want.

Second, what I have been really proud about

in my own community and what is really taking off is this idea of how we can build platforms that actually encourage a brighter future for crowd work. The platforms that we build embody certain expectations, certain assumptions about both the crowd workers and also the task that we will have them do. Are those tasks meaningless, menial jobs, or is there something brighter? Is there a path from these little tasks that we ask people to do towards something bigger? You can imagine these kind of crazy ideas like, "Well, I started out as a crowd worker and then I became an expert in the field." What kind of platforms could you build that could encourage that kind of transition?

My colleague, Aniket Kittur [co-author of "The Future of Crowd Work" available at http://dl.acm.org/citation.cfm?id=2441923], had this concrete, nice way of saying it: "What would make me proud of my daughter being a crowd worker?" I think it's a really interesting, compelling question to ask. It turns out Aniket and I both have daughters about the same age, so we are both asking that question.

I teach a class on crowd programming and the first assignment is to have my students, who are some of the top future scientists in the world, make $10 on Mechanical Turk, and it is amazing how much trouble they have doing that. These are the people who are going to go right out of undergrad and make $150,000 a year, or whatever it is. So something seems to be missing there.

This is kind of related to my final challenge,

which is: How can we protect the fundamental humanity of crowd workers while simultaneously benefiting from treating crowd workers as computational units? There is this tension that is fundamental for us to resolve going forward. The reason crowdsourcing is so nice is because we don't have to necessarily always think about each individual as this person who needs to be on-boarded separately, or someone with whom we need to develop social rapport. It's nice to be able to call people via APIs.

That isn't great long-term. This has huge consequences for the people who are engaging in our crowd platforms. If we want to make sure this is a long-term thing, something that we can keep doing in the future and really reap all of the benefits from, somehow we need to address this tension. People are both crowd workers and they are individuals, and I think our platforms and the tasks we design need to recognize that.

NYPL Labs, New York Public Library

http://www.nypl.org/collecitons/labs
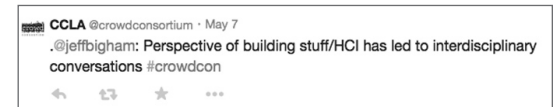
Ben Vershbow, New York Public Library

First I would like to thank the conveners. At conferences there is usually a panel on this, and it is like that panel exploded into two-and-a-half days, which is fantastic, so thank you for doing this.

I had trouble distilling this. My challenges have

kind of churned and mixed into the conversations we had earlier. I have tried to group them into a few categories and the challenges are around the why. Why are we doing this? There are many answers to that question depending on the project, on the discipline, on the sector that you are working with, on the communities that you are trying to engage or eventually serve. But this can come down to brass tacks: What are we doing with the data produced through these projects?

Coming from primarily a cultural heritage, digital humanities, library kind of space, but very deliberately experimenting in adjacent zones like citizen science and even with journalism, I am thinking about the different reasons for doing this. Are you generating a free-standing research data set? Are you trying to return data back to a corpus who will enrich it and make it more discoverable—that round-trip question of the data? I think we are past that first, "Will this even work?" phase or "Will anyone even do this?" I think we need to start asking some very big questions about why we are doing this and how to open up a path to integrating them or getting the data out for research and other purposes.

There are questions around the how: How are we doing this? I think this gets to questions about the team and staffing needs. What are the professional work categories that need to be created to actually support this work beyond, "Let's try a trial phase"? I think that is something we are very much doing with NYPL,

CCLA @crowdconsortium · May 7
.@jeffbigham: Perspective of building stuff/HCI has led to interdisciplinary conversations #crowdcon

Ben Vershbow

http://buildinginspector.nypl.org
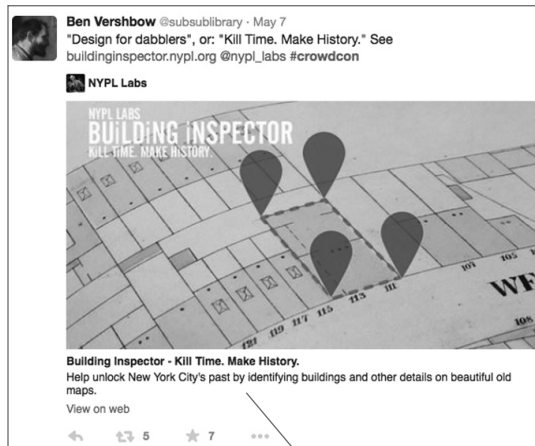
Panel (from left): Ben Vershbow, Jeff Bigham, Rachel Frick, Jeremy York, Meghan Ferriter

where we have demonstrated some models, we have been collaborating with various partners working in this space, and I think we have proven that there is a lot of potential here. But we are still kind of funding it and resourcing it as an additive. There isn't a hard examination of: Wow, maybe this is what cataloging looks like, or a part of what cataloging looks like. Maybe this is what metadata teams need to do, they need to be figuring out a kind of data life cycle that goes through various publicly engaged iterations and then returns back for the verification, validation, and quality control paces that it needs to be put through.

Then you start to expand to: How can we leverage computational methods? How can we explore computer-human collaboration? How can we do that not just in the now but in training for more advanced computational purposes beyond this work eventually? There are just so many questions.

I think you could ask questions about new kinds of librarianship, new kinds of curatorial work, new kinds of archival work, and those need to

be addressed and not just seen as additive or tacked on to largely unreconstructed institutional structures.

And very much drawing from both of those is the bigger story. How do we go from a very nascent, "Let's try things out, and see what kind of larger paradigm this points to," which may be in very faint outlines right now. Language has been thrown around about what does a national crowdsourcing platform look like, and there are platforms being built. Zooniverse already has a platform and is now building an even more robust technical platform where you can spin out projects without any technical assistance, and other platforms have emerged. When we talk about platforms from a cultural heritage space, I think there needs to be a narrative around that. I think with a lot of digital projects we've tried at NYPL, we've tried to create really compelling narratives that gesture at that why, that explain enough of the how for people to see where they're fitting in.

I think on a project basis, a collection basis, we have seen good examples of those narratives being well done, even narratives that have just been reported. For example, when the social layer around the Shelley-Godwin archive is built they are already, as a sector, thinking of the national, integrated, interlinked corpora that we are trying to build. What is the big narrative around that? I think we need to figure that out. I sometimes think of this as a kind of generational process of migrating data forward, migrating knowledge to a

new medium. That may be a little abstract and may not appeal to everyone, but it is a place to start.

What are we all collectively doing? What is this big public works project that we are all engaging in? Some of those bigger metanarratives might help us both think more broadly in a longer range way about what we are building together and how to link all of those efforts together, but also how to engage the public.

I was talking to Sharon Leon earlier today about the survey data that comes in in response to: "Why are you participating?" A lot of people articulate something very broad: "I see it as my civic duty." That's really interesting. Maybe we can tap into that one.

Questions, Answers, Discussion

VALUES, PUBLIC MISSION

• Those were some interesting points, and to bring together some of the comments from the last two panelists, there were some interesting values questions here, and that may be part of the story, right? I think when you talk about click labor or something like Mechanical Turk, which is another space for that, when you come back to libraries and archives (the "L" and "A" in GLAM), we have this very strong public mission. It is public memory, the collections are the people's collections, and the collections themselves have been built often through contributions, and that is part of the volunteerism story.

I am curious to hear from folks on the panel, what is the narrative we have, and how is this similar or different from crowdsourcing in other spaces? Jeff, in your case you were zeroing in on some of the labor issues, and I think Ben was getting into some of the points about when we are serving the public good in some way. For example, the Zooniverse projects recruiting people to help scientists solve a problem and make the world a better place. • Trevor Owens, Institute of Museum and Library Services

OFFERING THE REAL...WHAT?

• I just ventured one maybe far-from-perfect example but yes, even the Zooniverse tagline. It wasn't real science online, it's a little more drawn out maybe, but it is about essentially saying that this is a way for you to participate in actual scientific research, you are entering the research life cycle and there will be results, there will be publications, there will be research done with this. I was thinking in terms of the cultural heritage or the GLAM sector, this is participating in real what? • Ben Vershbow, NYPL Labs, New York Public Library

PUBLIC PARTICIPATION/OWNERSHIP IN THE STORIES WE TELL

• In a cab ride on the way down to D.C., there was an advertisement on WDEL radio, which is the Baltimore talk radio station. There was a message that said something like, "This

The Personal Discovery and Pleasure Payoff

• We are in the right place to focus on digital labor and, as organizations and project leaders, think through how we want to launch or create these kinds of projects. And still, the feedback that I hear most from our volunteers relates to delight and discovery and leisure and pleasure in participating in these projects. Even though some of our most actively discussing volunteers are describing their activities with gravity—they understand what they're doing is serious, it's important, it's powerful—they also love it. They're having a wonderful time, they are finding things and making connections to their own existing knowledge and experiences. For example, "This clock is in my home town," or some personal discovery relating collections to their local lives. They are also describing the experience as the opposite of work. So we may be analyzing with the lens: "This is work for them, how are we going to find a compelling narrative?" But that's not always how they perceive it, so that is maybe something to pull into our debates, as well. • Meghan Ferriter, Smithsonian Transcription Center

**User Stories Highlighting
the Social Good of Crowdsourcing**

• Regarding this whole idea about joy and delight and civic duty, I want to shout out to Jon Voss and the Shift organization. They have a really good way of talking about decreasing senior isolation and increasing conversations between generations as part of their activity. My wish list for the crowd consortium: How can we collect these user stories? I like "user stories" instead of "case studies" because in my mind it's more captivating. Could we have in one place a five-minute snippet, a YouTube video narrating and pointing out and collecting these user stories of the social good of crowdsourcing? • Rachel Frick, Digital Public Library of America

> **CCLA** retweeted
> **Amy Patterson** @amycep · May 7
> Participating memory, user stories: what would Humans of Crowdsourcing look like? #crowdcon
> ↩   ⇄ 1   ★   •••

http://stackoverflow.com

> **CCLA** retweeted
> **Daniel Powell** @djp2025 · May 7
> Great examples of robust user communities (although not without issues): Stack Overflow: stackoverflow.com #crowdcon
> ↩   ⇄ 1   ★ 1   •••   View summary
>
> **CCLA** retweeted
> **Andrea Wiggins** @AndreaWiggins · May 7
> Frick: how do we capture these narratives that show the benefits and social good from crowdsourcing? #crowdcon
> ↩   ⇄ 2   ★ 2   •••

is a message from the Maryland Historical Society. We are making an all-call for any pictures, images, sound or video involving the recent activities in Baltimore. Professionals, amateurs, cell phone photos are welcome. Send it to this address. It will become part of the historical record at the National Archives and Records Administration. It just went through the whole thing of where your collections will be.

I thought to myself, maybe libraries and archives are relevant to the 21st century. That was just such a crack into my thinking. Here was this ad in a cab on the radio and it was really powerful. I think that's a powerful narrative around crowdsourcing with cultural heritage institutions. It is building our collections, it is connecting people and current activity and that story we are going to tell about what happened yesterday. • Rachel Frick, Digital Public Library of America

• In that vein, the September 11th Archive is in that space, and StoryCorps is another example of exactly that kind of collecting activity. • Trevor Owens, Institute of Museum and Library Services

• The idea of the narrative is really compelling to me. Putting it in stark terms, in my mind I think the narrative that we've had often as cultural heritage institutions has all been about authority and having the authoritative records, the authoritative story, that kind of thing. And a large fear

is having people corrupt that story. I think that what we are seeing here, and a very compelling thing to me, is the acknowledgment that the authority is distributed. That is, all these perspectives are what make up the authoritative story. That to me is a very powerful narrative in terms of empowering people. And I think why there is the delight and excitement is because my piece, vis-à-vis what's happening, is part of that larger story and that is incredibly meaningful. That makes me want to contribute, if what I say is part of the whole thing. That just strikes me as a very engaging way to view it. • Jeremy York, HathiTrust

• I would like to briefly respond to the idea of authority as a narrative and now authority is distributed. I think that's wonderful, it's beautiful, and I think that's part of it. Also there is what Rachel said with the radio analogy. And Trevor, not to curry favor with the moderator, but you have written about this and you just mentioned the memory aspect, participating in active memory as a society or as it pops up in sub-communities of society. I think that is the narrative of memory and moving that into the present day, not only in response to contemporary events but also in terms of stewardship and bringing memories forward and not forgetting.

We are in the early stages of sketching out a framework at NYPL for basically sharing a digitization candidate list with profiles of

possible digitization projects, kind of in the mold of participatory budgeting, where we say, "Look, we want to do all of these things, they all could be great, help us prioritize where to start, where to focus our effort and develop some meaningful mechanisms." Part of that effort, that digitizing, is part of the memory process—remembering it onto the Internet in some way. It's a trope I've been playing around with a lot, memory as a narrative. • Ben Vershbow, NYPL Labs, New York Public Library

CONTRIBUTING TO MACHINE LEARNING

• Jeff brought up the machine learning element in this, and I think that's another interesting component in valuing people's time because the work is actually leveraged and sort of turned forward. I'm curious whether we have cultural heritage crowdsourcing projects that have actually better informed machine-learning activities, or have we just gone in one direction with this. Specifically, it seems like OCR data from Trove has been being corrected for an extensive period of time. Do we have better OCR algorithms because we have that corrected data? How do we get to there? • Trevor Owens, Institute of Museum and Library Services

• We have an archive of television. We have a volunteer-run project. We archive television from Philadelphia during an election cycle. We have volunteers identify political ads in

that. This is all done by hand and crowd-sourced. Then we used that as a training set to evaluate different methodologies for automatically identifying commercials within a larger set of television. We've had a few iterations and have managed to get it above 90% accuracy at this point. Now, when we do this again in the future, we will be able to have a human say, "This is an ad," and then automatically go and find every other instance of the ad throughout the entire corpus. • Alexis Rossi, Internet Archive

• That is exactly the kind of example I was asking for. • Trevor Owens, Institute of Museum and Library Services

• I can't speak in great detail about it, but I know that UCL London and a number of partners around Europe have been trying to work on manuscript-recognition technology, handwriting-recognition technology, using manuscripts that were crowdsourced transcriptions by humans. They are trying to train machines to be able to read handwriting. It's in an early stage, but I think that's really exciting and we should be thinking about those opportunities too. • Victoria Van Hyning, Zooniverse, Oxford University

• One thing we talked about in our working group that I think is really important is how vital it is to have a bidirectional communications channel between the people who are collecting this data and the people

http://www.18thconnect.org/typewright
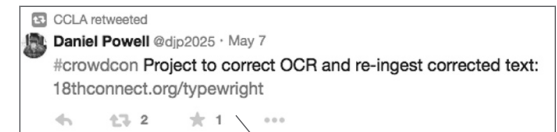
Scale and Exploitative Potential

• One thing I brought up in our earlier working group discussion was the idea of scale. As this sort of approach becomes more and more popular, as more and more different projects use this labor force, I hope that they continue to find this inherently valuable and they are doing it for altruistic reasons. I don't think we should be lulled into this false sense of security thinking it doesn't matter exactly what we do because people are here for these reasons and are interested in science and so on, because you can still be exploitative even in that sort of environment. And even if it's not for money, if it's for points or other things, it is still something we want to take seriously. • Jeff Bigham, Human-Computer Interaction Institute, Carnegie Mellon University

Frameworks for Serving Social Good

• There is this question of getting labor out of workers, but there is also the potential for creating projects that unquestionably are serving specific social goods as well, and that is actually the opposite. Some of those activities may look similar, but it is a different sort of framework. • Trevor Owens, Institute of Museum and Library Services

CCLA @crowdconsortium · May 7
.@jeffbigham: Using computers instead of volunteers might be better for individuals' privacy in terms of data being worked with #crowdcon

CCLA @crowdconsortium · May 7
.@jeffbigham: How do we get crowd to express what's important about expertise in particular domain? #crowdcon

CCLA @crowdconsortium · May 7
.@jeffbigham: How can we teach computers to be as good as experts? Considering what human work computers could take over in future #crowdcon
★ 1

CCLA retweeted
Victoria Van Hyning @VanHyningV · May 7
But authority is distributed, and all views are part of our collective societal and institutional stories. @hathitrust rep #crowdcon 2/2
↩ ♺ 2 ★

CCLA retweeted
Victoria Van Hyning @VanHyningV · May 7
There are institutional narratives of authority underpinned by fears that expert authority will be 'corrupted' by 'amateurs'. #crowdcon 1/2
↩ ♺ 3 ★ 3

CCLA retweeted
Andrea Wiggins @AndreaWiggins · May 7
Zombies: "oh no, if I don't help out, world will end" (climate change) vs Unicorns: "the universe is so beautiful!" (@Zooniverse) #crowdcon
↩ ♺ 3 ★ 1

CCLA retweeted
Sharon Leon @sleonchnm · May 7
These are 2 different things, scale of work. Transcription v. creative commitment, investment, expression from the people #crowdcon
↩ ♺ 2 ★ 3

CCLA retweeted
Brett Bobley @brettbobley · May 7
Crowdsourcing runs a gamut from "crowd workers" (strictly in it for money) to those contributing to a project for fun/pleasure. #crowdcon
↩ ♺ 4 ★ 5

CCLA retweeted
Andrea Wiggins @AndreaWiggins · May 7
.@MeghaninMotion reminds me of potential point of divergence: do GLAM & DH have zombies as well as unicorns like #CitizenScience? #crowdcon

who are potentially working on the machine-learning algorithms to be able to automate some of it. And there are two reasons for that. One is, from the data collector's side, if it is not packaged up in a way that makes sense to a computer scientist then it is much less likely that they will actually work on it. From the other direction, many of the problems that come up in these projects are problems that, from the computer science research point of view may have already been solved in some way. It would be really nice to understand when there is something new that comes up, something that hasn't been seen before, that could both satisfy the needs of the project and also allow the researcher to make progress. That all points to the need to have this communication, not just putting data up and thinking maybe someone will come along and figure out how to automate it, or vice versa. • Jeff Bigham, Human-Computer Interaction Institute, Carnegie Mellon University

VIABILITY OF A PLATFORM SCALE

• I think this would be a good argument, one of many, for more of a platform scale for this activity. If everyone is doing their own data export and data packaging we are not going to get as far as fast. If you have a larger aggregated cultural heritage crowdsourcing effort with the DPLA and others, there could be an amazing sort of research center, a HathiTrust research center in a sense, where there could be training data and this stuff could be in an aggregate. And it also speaks to these different audiences and different ways that we need to serve data to different audiences.

We are trying to move towards building an interoperable metadata structure through the web with other institutions, but that is too heavy for a web developer who just wants visualizations to start playing around, so you also want CSVs and JSON experts. You want to package it up, offering that new suite of distribution channels and formats. I think that is something that we need to be experimenting with and we can do better together. • Ben Vershbow, NYPL Labs, New York Public Library

• When you said to come to the DPLA API and download 10 million metadata records to use as a test of the data to run a crowdsourcing thing, it never would have crossed my mind. But it is an awesome potential use for the data and I'm putting it on my to-do list. • Rachel Frick, Digital Public Library of America

INTERNAL ORGANIZATION AND STAFFING

• I have a question directed to Rachel. How important do you think it is for us to figure out who does this? Does this need a department? In my experience, I think a lot of us came here because someone asked us to. Does digital services do crowdsourcing? Does technical services do Crowdsourcing? I'm

from a library, so my structure is different. Do you think it's important for us to have something like that, or is this distributed across an institution? • Tom Blake, Boston Public Library

• I would rather build the environment and the framework that allows people to focus to the best of their ability on the portion of the problem that they can speak to best. I have found when you try to centralize who is responsible, then the community starts decreasing their feeling of ownership and responsibility towards the shared activity. So how do we set up a framework? And instead of saying who is responsible and trying to point externally to one set organization, how do we communicate openly and clearly and document how we all fit together or connect, and those points of connection and confluence? How can we turn that thing around and say, "What can I do to support this?" "Where can I lead to move us forward?" How do I maybe suppress my ego and hand it on to a friend over here? • Rachel Frick, Digital Public Library of America

• I'll jump in here too. Meghan's job exists, and that's why a lot of these activities happen. There aren't a lot of organizations that have jobs like yours. This opens a question of the organizational home for this kind of work. If we are redrawing our organizational charts and asking where are the people who engage in community management around crowdsourcing projects, there is a lot of potential for this to fit into 21st century librarianship. • Trevor Owens, Institute of Museum and Library Services

• I think it's about making the physical infrastructure of the space to be flexible for multiple kinds of contributions, activities, or ways in which people want to participate. Also, trying to encourage and be comfortable with dealing with things that come back that are not what you asked for, but finding ways to connect that to value from an institutional perspective. A number of our participating groups from the Smithsonian internally have different goals and objectives, yet each has been able to benefit from the success of their peers by having projects in the same space. • Meghan Ferriter, Smithsonian Transcription Center

• The org. charts are tricky and I was just thinking through them. I think certain kinds of roles are important, such as a point for coordinating the collaborative activities, and probably required. But I would say it is probably more now a part of many different people's work. It needs to be understood and it needs to be quantified to the extent that it's not just additive—"Here's this extra thing we expect you to do." • Ben Vershbow, NYPL Labs, New York Public Library

CCLA retweeted
Neil Fraistat @fraistat · May 7
What is the future of librarianship in the crowdsourcing space? #crowdcon
3   2

Moderator Trevor Owens

CCLA @crowdconsortium · May 7
.@subsublibrary: We're still funding crowdsourcing projects as an additive, not as a core piece like cataloguing #crowdcon
↩   ↻ 1   ★   •••

CCLA @crowdconsortium · May 7
.@subsublibrary: What are team staffing needs for crowd projects? #crowdcon
↩   ↻   ★   •••

Session Feedback:
Memories and Voices of Users

• There were some phrases in this session that stuck out to me. The idea of participating in memory is such an evocative phrase. And in terms of user stories, I was thinking of that great series you see on the web, *Humans of New York*, but instead *Humans of Crowdsourcing*, the real faces and stories of people working in crowdsourcing.   • Jake Dunagan, verynice.co

CROSS-INSTITUTIONAL COLLABORATIVE STRUCTURES

• I'm also thinking about how we can think about this beyond institutional bounds. We have big New York geography and history projects that are coming out of data mining maps. We just got a grant to try to weave our Building Inspector project together in a historical research framework. One of the points is to aggregate materials through time and place from all of these institutions, so I am starting to think about how we can work together as a network of institutions in New York and distribute that work across our groups.

Likewise, we have a theatrical playbill transcription project which is in the early stages, something we are hoping to put more focus on soon. I think NYPL could be a great leader there. We have a whole performing arts research center, a whole theater division, but there are amazing performing arts archives all over the city, amazingly engaged cataloguers and archivists. What if a little of their time was devoted to this? So I'm thinking about projects that need to extend across institutional bounds around disciplinary centers of gravity, or a collection that is physically distributed in different archives, and therefore a collective stewardship model is needed, like the Shelley-Godwin for example.   • Ben Vershbow, NYPL Labs, New York Public Library

TREND MAPPING AND GAP ANALYSIS

Introduction

Facilitator: Jake Dunagan, verynice.co

We are going to do this next session together as a group, and it is more of a rapid-fire filling up of the "bestiary of possibilities" that we are working with today. I have distilled a few of the big stories and trends that have come out of the previous two regional workshops and will present those to you. What I want to do is fill out this landscape of trends and emerging issues that are going to serve as material for our thinking. This is another effort at filling in a space with new ideas and is basically a brainstorming session. As you call out ideas Matt Manos is going to record and categorize them on the board.

Big Stories and Trends from Regional Workshops

Facilitator: Jake Dunagan, verynice.co

These are things that came out of the previous two workshops that were noted as trends, things to look at, and emerging issues, and some of them have already been mentioned again during this workshop.

Big Stories and Trends

- **Machine Learning**
  Obviously cognitive computing and algorithms are potentially going to be a big part of the technological infrastructure.

- **Internet of Things**
  Billions and billions of actual objects connected and networked together, talking to each other and being coordinated in that way.

- **Globalization**
  Obviously this is a meta-issue, things like the expectation that multilingual and multicultural communication and intelligibility needs to be taken care of and addressed.

- **Class Gap and Inequality**
  This has been mentioned a couple of times in this workshop. This is creating a workforce of people who are desperate, in need, and maybe interested and have free time on the other end, so it is the idea of increasing inequality.

- **Data Overload**
  This is kind of a simple one. I don't know what the factor is now, but we are producing more data every year than we have in human history. I don't know what it's up to, but it's accelerating. There are massive amounts of data: How do we deal with that?

- **Science of Persuasion**
  One of the really interesting things is around the science of persuasion and engagement and motivation, so we are learning why humans do things with more precision and how to turn those knobs, for better or worse.

- **Transparency and Openness**
  Whether you want to be open or not, there is this almost default expectation that if it is being produced and shared it is going to get out there (ask Sony Pictures).

- **Freelance/Microtasker**
  Similar to the idea of the class gap, there is this freelance microworker, microtasker class consciousness, maybe also known as the rise of the proletariat. It is this idea of the people who are doing fragmented, piecemeal work having a political voice and the new labor movement that is built around that. The Uber strikes, the Mechanical Turk protests, those kinds of things are all on the rise and it's something to think about as a big trend.

- **Crisis of Higher Education**
  Can the tuition and infrastructure costs of running these institutions keep going up? Are there going to be threats from being top-heavy, but also threats from new kinds of educational opportunities, such as Kahn Academy and others, which are filling in different gaps?

- **End/Transformation of Expertise**
  Maybe the "end of expertise" is too strong, but the transformation of expertise, the idea that groups can be experts or have the function of expertise as opposed to the sort of priesthood model we've had.

- **Global Connectivity/Mobile Internet**
  You have four or five billion people connected to the network together, so that's an access issue in many cases.

- **More Data – Less Persistence**
  We have all this data, but you may not be able to find it or it just fades from memory. How do we capture that data and keep it alive, as it were, and useful?

Ed. Note: Attribution
There is no attribution in this brainstorming session other than a few instances in which participants mentioned specific projects.

Categorization/Headers
The categories into which the brainstorming ideas have been placed here are those which were used by the facilitators during the session, with some exceptions. Headers have been expanded to better reflect category content.

Brainstorming

Facilitator: Jake Dunagan, verynice.co

I'm now going to throw it over to you and again, this is really a flow session. We want you to just throw out ideas, things that you think are important, things that you've seen. We've had a discussion about signals from the future and maybe you want to add some of those that you think are important. You can also think about adding to the list of big ideas from the regional sessions, and you can build on or comment on those ideas. Or you could point out a counter-trend or ask a question. I want to push this out a little further than just crowdsourcing to factors that affect the process of crowdsourcing, who the labor market might be, how you connect, the technological infrastructure, all of the pieces of the puzzle and things that are happening in those spaces. We are trying to just understand the landscape. We are not looking for solutions to these problems, we are looking for trends and issues that we should be paying attention to.

MACHINE LEARNING/HUMAN LEARNING

- We have brought up machine learning. The question I have is, what is the human learning part of the machine learning? If humans are part of the machine learning, how can you design that so that humans learn and have meaningful tasks as part of that?

- Yes, humans plus machines tends to be better than humans only or machines only, so humans and machines working in concert together is amplified above what machines can do right now, so that interface is really important.

- Also part of that is how does it lead to tasks that humans find meaningful? So not just the combined power, but how does this work remain meaningful and inspiring for humans?

- Ideally, as we automate more of this work, humans would be doing what humans do best, so creative exercises, connection, caring, or empathy. Can we design a system where we offload machine work to things that machines do best and leave the good stuff for humans? Not the bad stuff, the stuff that we don't want to do.

machine learning

INTERNET OF THINGS

how are humans part of this?
human + (creating new meaning)

GLOBALIZATION/DIVERSITY

- I was thinking of issues of scale. We are trying to think big, and usually when we talk about the idea of scale these things are bigger, our platforms are bigger. But we may also think about the issue of scale in terms of the cultural, the political, and diversity, so the scale might not be more people, but people with a different historical perspective or different political perspective or a different background, so that scale doesn't always just mean numbers or the platform support.

- So scaling as targeting different communities effectively.

SILOED IDENTITIES AND DATA

- Logistically speaking, in the landscape of things that are going on that humans can get involved in there is siloing. I think it's an old problem that data produced is siloed and not always accessible from everywhere, but there's also an identity siloing problem. We have contributors in my work who contribute photos through Flickr and information through Wikipedia and a variety of other reporting platforms, and it's hard to tell who is the same person. For issues like people having credentials and people knowing what you're good at and what you have done, your body of work, it is difficult to get a united identity for a contributor.

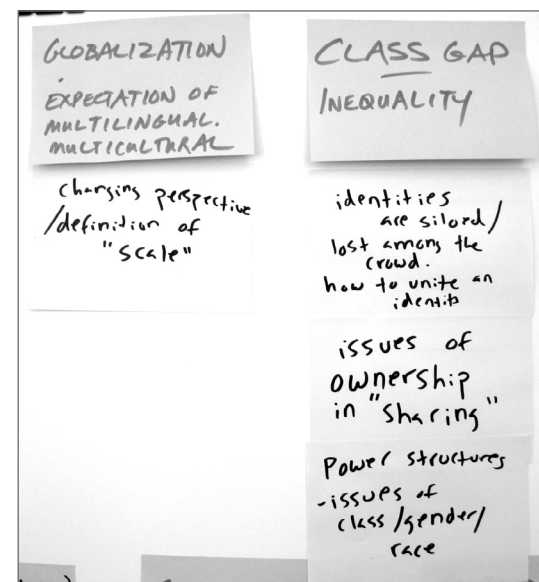- So as a following space in this trend mapping there is identity siloing, being able to connect these different platform and domain activities
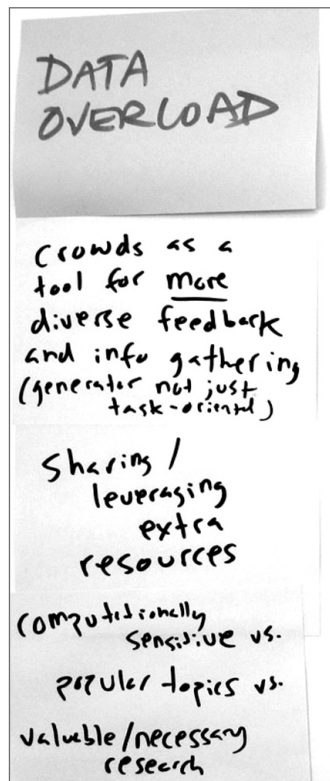
CLASS GAP AND INEQUALITY/ NEW MODELS OF OWNERSHIP

Ownership Issues

- We have to also put the ownership structure on the table. We are talking about Airbnb and Uber, and yes we've dropped the transaction costs a little based on all the sharing we're doing, but the company owns that and the profit is being extracted from the network. So a collaborative ownership model would be appropriate if we really want to live up to the promise of a "sharing economy." I actually deny the use of that phrase to Airbnb and Uber.

- So new ownership models is a category—emerging types of ownership models is a great trend to have up there.

- When we talk about the crowd owning this and going beyond microtasks to do more complex things, I think we underestimate how scary it often is asking our crowds to do that. We understand that there is value in many different perspectives, but that is not what the education system is for those outside the academy. You may understand this is valuable, but you need to convince people. If we want to do those more complex things we need to build systems that are going to start that, different routes.

GLOBALIZATION
EXPECTATION OF MULTILINGUAL. MULTICULTURAL

Changing perspective /definition of "Scale"

CLASS GAP — INEQUALITY

identities are silored/ lost among the crowd. how to unite an identity

issues of ownership in "sharing"

power structures -issues of class/gender/race

CCLA retweeted
Daniel Powell @djp2025 · May 7
From @mukurtu : Scaling can mean depth, no just breadth, of participation. #crowdcon

CCLA retweeted
Liz MacDonald @spaceyliz · May 7
.@tweetaurora software for crowdsourcing real-time social media events on a map is opensource! What else could it be used for #crowdcon?

CCLA retweeted
Jon Voss @jonvoss · May 7
new ownership models of crowdsourcing projects, and deconstructing "sharing economy", aw yeah. #crowdcon

CCLA retweeted
Andrea Wiggins @AndreaWiggins · May 7
.@lizbarry suggests cooperative ownership models as a way forward; @chrislintott sees such opportunities as scary for some. #crowdcon

> **CCLA** retweeted
> **Victoria Van Hyning** @VanHyningV · May 7
> We must try to not recapitulate problematic power structures that pervade society. I paraphrase a portion of our discussion. #crowdcon
>
> ↩   ♺ 3   ★ 1   •••

> **Andrea Wiggins** @AndreaWiggins · May 7
> agree w/ both points: cooperative ownership makes sense, but don't overwhelm ppl who aren't ready for it. Context matters, a lot. #crowdcon
>
> ↩   ♺ 2   ★   •••

> **Lacy Schutz** @LacySchutz · May 7
> Agree - Telling the crowd their perspective is valuable is in direct contravention of the message from academia @chrislintott #crowdcon

Class/Gender/Race Issues

- We shouldn't underestimate the roles of class, race and gender in the first place in access to institutions and the power structures at stake in that, and that we can replicate problematic power structures in the ways that we do this. This relates to working from the crowd-up or working down when those dynamics are in major play, and it is a critical test to be applied, particularly with many speaking from positions of institutional power with the resources related to those positions.

- So we may be recapitulating power structures that denigrate traditionally under-represented groups and need to be aware of that when designing projects.

DATA OVERLOAD

Crowdsourcing/Filtering the News & Info

- I think we talk about crowdsourcing a lot as getting people to do a particular task for us, but personally what I use crowds for a lot is getting more information, gathering information. For example, when there is an ongoing news event in a place where people have access to technology, I might read a news story to get some context, but I basically don't believe whatever they're telling me in the news because of whatever their perspective is or whatever their motivation is in selling newspapers or ads or whatever. What I instead do is use the crowd. I go on Twitter to actually gather information and then I choose to filter, however I decide to do that. It is the idea of using the crowd as the generating mechanism for data as opposed to just task-oriented workers.

Sharing/Leveraging Extra Resources

- I'm not sure how this relates, but we're talking broad landscape: the idea of efficiency and carrying the fallow surplus into something productive and worthwhile. Server virtualization—Amazon Cloud services didn't happen because Amazon decided to have a cloud service, they had extra servers that they needed to turn to something productive. Annheuser Busch has breweries all over the place and can't always use all of those, and rents them out to small microbrewers. That part of the landscape has something to do with what we're doing here. You buy stuff or you have a big resource that you don't always need all of. We are trying to figure out ways to harness that so it is turned into something productive.

- It's the whole idea of an overlap, collaborative consumption and a sharing economy where you're using your extra resources and capacity, whether it's Airbnb or whatever. The idea of collaborative use or collaborative consumption and now having the tools to connect those things. If I have the tools for doing that very efficiently, now all of a sudden new potential opens up.

Categorizing Domains of Crowdsourced Material

- Managing the discrepancy between projects that are dealing with computationally sensible material versus the ones that are on popular topics and the ones that are on valuable and necessary research. There are three domains to keep in mind. The three may align but they may not. We could end up with lots of judgments on material that is not of value.

SCIENCE OF PERSUASION/ATTRACTING AND CULLING SUPER USERS

Targeting and Sustaining "Super Users"

- There should be something about the illusion of the crowd. A lot of times we see these projects and they have these massive numbers of people they report, but then if you dig in or you talk to the people who followed it, there are like maybe a handful of people who are actually doing the contributions that matter.

- That's the old pyramid of participation. You've got the one percent who do 90% of the work. There is rewarding them and rewarding multiple levels of participation. So you reward those who just come in and click "yes," or "like"; you reward those who may spend a few hours; and then you also want to reward the people who are there every day, for days and weeks and months at a time.

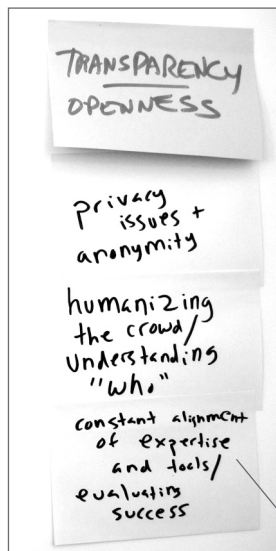- It plays into some of the last comments,

which is that it seems a lot harder to put some of these projects out than maybe it is because you see a project that had tens of thousands of people, but really it's just a filtering process. Can I get enough people here so I can filter down to the people who will really contribute to my project? And that's like five people.

- So 100,000 gets you down to 10 useful ones, so it's an oligarchy. It's a small number of people who are actually contributing, not this giant crowd. It's not that hard if you focus in on that 10, once you've thrown that blanket out there.

- I want to crowdsource the room and ask if there are any examples where people have gone after their super users to successfully catch them. Or have you targeted the people who really want these tools, built them the tools, and so on?

- So once you have that small super group, can you sustain them as a group for other projects as a coherent unit?

Elevating Super User Roles

- Tina Phillips: There is the eBird example. A lot of their super users became state or county reviewers, so they actually helped to verify data points which may seem out of whack. And then they constantly get feedback from those users about what is the next tool that would help us. As Chris said, eBird really just turned it around. They're collect-

Daniel Powell @djp2025 · May 7
#crowdcon Is participatory creation able to be scaled to complex projects?
2

critical play @criticalplay · May 7
move beyond the crowd for tasks & see the crowd(s) for knowledge, community & information #crowdcon
6 3

SCIENCE OF PERSUASION & ENGAGEMENT

How to target and sustain a group of "super users"

nomination-based leadership committees that share and create resources

elevating roles among "crowd" members

CCLA retweeted
Andrea Wiggins @AndreaWiggins · May 7
@mhaklay @mukurtu I see communities of practice/interest becoming increasingly blurred from functional standpoint in these spaces. #crowdcon
1 View conversation

CCLA retweeted
critical play @criticalplay · May 7
expect the pyramid of participation--are all of the levels offering meaningful exchange? #crowdcon
1 1

Aligning Expertise with Goals

• We work in varying disciplines and it is important that we evaluate the relevance of whatever tools we use in supporting activities which align to groups.

• So aligning the expertise that's out there with the goals of the project.

ing all the data, but they want a need, and they've created a whole bunch of ancillary tools that the community finds very valuable.

• So those users have moved on to become trusted verifiers.

• Tina Phillips: Yes. People want to know, "If I went to Georgia in May, what are the birds I'm going to see that I haven't already seen?" So there are hotspot maps, and you can get a checklist printed out in 30 seconds of all the birds that you'll see in May in Georgia.

• But that is a case where there are a few doing most of the work. Are there any examples of more equal distribution?

Super Users & Merit-Based Leadership Roles

• Liz Barry: That is a new and interesting question. The long tail is a cliche in citizen science online participation. I don't have the stats ready to respond to that, but I think it's interesting for us to think about. To the earlier question about reaching out to super users, in the Public Lab community, there is actually a leadership structure people who are very active can join, including those who are very active as developers or doing field research or using data for advocacy, whatever they are doing. It is nomination based, something called the Public Lab Organizers, which is a voting body that kind of acts like faculty in a way. We share resources like press kits and slide decks and connections to journalists, grant writing templates.

• So as part of the best practice conversation: How do you reward heavy effort? And there is some kind of meritocracy involved, allowing people to have more responsibility and trust as they contribute more.

• Darlene Cavalier: I'm from SciStarter and we have metadata about projects and we are over 1,000 now. We were hearing a lot from teachers who said, "There aren't enough projects aligned to science and engineering practices." Or, "We don't have education materials." So there's a team of 60 teachers that will align 500 projects to their needs. They are the experts and we just kind of create the matchmaking and the platform for them to be able to do that. But they have elevated themselves to being reviewers of projects now.

TRANSPARENCY AND OPENNESS

Privacy, Security and Anonymity Issues

• There are issues around privacy and security of information. If you're contributing to a project you choose whether you are publicly identified, and that is where that single identity issue is good in some cases and possibly not good in other cases if you want to remain anonymous. Who controls the contributions, and do you cede control of that contribution to the project or company?

FREELANCE/MICROTASKER AND HIGHER LEVEL, MORE COMPLEX ALTERNATIVES

"Flash Teams": Tapping Varying Expertise

• A lot of the crowd tasks we are seeing now are rather simple classification tasks or tagging something. I am really curious about what we can do with highly complex tasks that require different levels of expertise, like different projects require different levels of expertise, but really organized in a way that it is being done efficiently with experts, non-experts, machine learning, all of it together. That kind of thing is very complex and it is very project-specific, but there are a lot of overlapping points.

• So how high can the crowds go in the level of complexity of the project? Jeff Bigham did a pretty complete thing, using the crowds or at least the platforms to put together a paper.

• Jeff Bigham: I was writing a paper and I was the expert, but I was orchestrating the crowd.

• Jake Dunagan: My colleagues at Institute for the Future have been talking about this for a year and they just completed a project where they wrote a report for a *Fortune 50* company, but it started with signals from Mechanical Turk about a certain topic. Next they used Nexus oDesk to organize those into a paragraph, then they shipped it out to Elance to edit and put it into final form. They had to be there to write the rules for each step, but then they push a button and
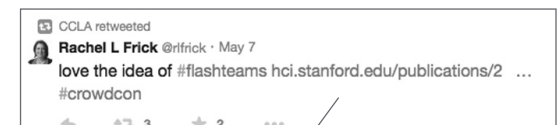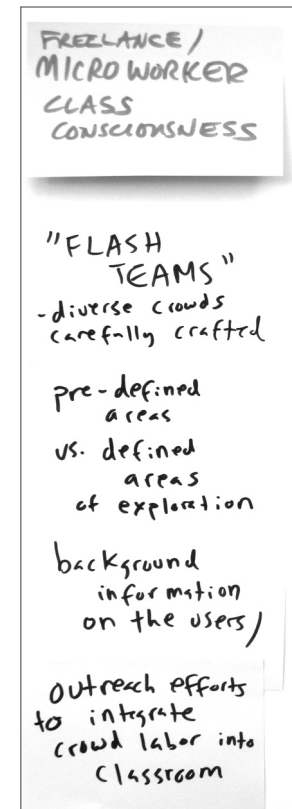
once those algorithms are in place that's it. They call it iCEO, so they can run a think tank without any managers. As a prototype it was pretty successful.

• Jeff Bigham: There is a fascinating project out of Stanford, Michael Bernstein et al., called Flash Team. Basically they were recruiting workers with varied expertise from oDesk (now Upwork) to do things like put together a team and in a couple of hours make a workable prototype. Or they also made an animated movie that was very high quality and they brought together the animator, they brought together the story designer, even a professional singer, all from this platform, thrown together in a few hours to do something really complicated, drawing on each of their expertise, with the computation sort of managing and planning their activities so that they could stage each step.
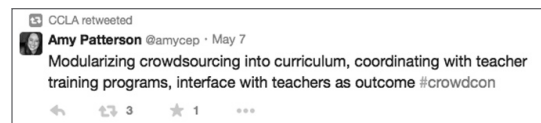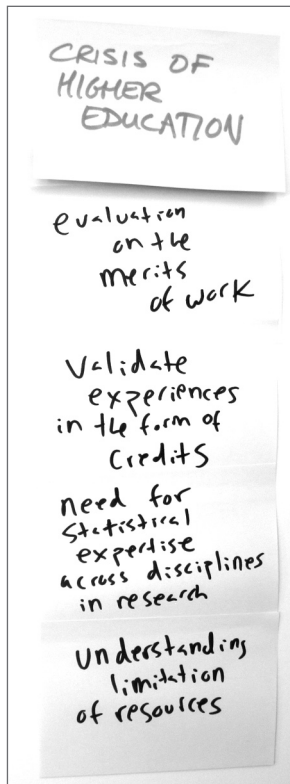
So I think as far as proof of concepts for that higher level complexity, I think they're out there. I don't know if they're marketing it right yet, but they're getting close.

Predefined vs. Crowd-Defined Tasks

• I would like to add a distinction between achieving predefined complex tasks and working in a problem space that is not yet defined, because the modes of organization vary widely based on that. If it is about coordinating the team structure in a crowdsourcing way that's one thing, but when we cross into actually figuring out what are we

FREELANCE / MICRO WORKER CLASS CONSCIOUSNESS

"FLASH TEAMS"
- diverse crowds carefully crafted

pre-defined areas
vs. defined areas of exploration

background information on the users)

outreach efforts to integrate crowd labor into classroom

CCLA retweeted
Rachel L Frick @rlfrick · May 7
love the idea of #flashteams hci.stanford.edu/publications/2 …
#crowdcon

http://stanfordhci.github.io/flash-teams/

CRISIS OF HIGHER EDUCATION

evaluation on the merits of work

Validate experiences in the form of credits

need for statistical expertise across disciplines in research

Understanding limitation of resources

doing and tackling, and more innovation is required of the organization, we enter into a different mode, primarily referred to as peer production.

- So the rules and the outcome and the purpose that have emerged organically from the crowd as opposed to predetermined tasks.

Outreach Efforts
Integrating Crowd Labor into Classrooms

- I wonder if there is not some outreach effort that needs to be done directly to high schools, community colleges, college classrooms. What if we design this stuff in a way that it can be in-class assignments or homework assignments so that it's actually getting picked up?

- Jeff's idea of making your students earn $10 on Mechanical Turk is a great example, and how hard that is. We could modularize that into some curriculum they could use and try out.

- At a higher level, coordinating so that a community like this has some liaison or point of intersection with teacher training programs so that student teachers are learning about this as a way to do the teaching that they're going to have to do anyway, and think about it as a basic part of the toolbox.

- That's a wonderful preface for tomorrow, thinking about that as an outcome, some sort of interface with high school and college teachers, or younger.

CRISIS OF HIGHER EDUCATION/RESEARCH & EDUCATION CROWDSOURCING ISSUES

Legitimizing Crowdsourcing/ Merit-based Evaluation

- There are issues of legitimizing the products of these alternative approaches in our own peer communities. We need to teach our grant reviewers and peer reviewers to evaluate on the merits of the work and not assumptions.

Translating Crowdsourcing into Academic Credit

- Connected to education, it's a matter of validating the experiences that people have in a way that can be translated into community college credits or some kind of career path. These alternative activities need to be translatable into credits, credentials, etc., or there will be a disincentive to participate.

Need for Statistical Expertise

- This is not universal, but I would imagine on the research side across disciplines there is a lack of basic statistical expertise, which would make it difficult to incorporate or even understand the asymmetries that would show up. There is a need for statistical expertise, being able to model populations and things like that.

Resource Issues

- I think there's a very practical problem of not knowing what resources it takes to get

to a successful project outcome. Our 800-pound gorilla projects in citizen science had a lot more money poured into getting to where they are today than the projects currently starting out are requiring to get to the same point. And yet nobody knows what the bottom line cost is for start-up and maintenance.

Translating a Task into Crowdsourcing

• To elaborate on that, we need to learn a lot more about what it takes to translate a task that is now being done by individuals or small groups into the crowd. What translations are needed to make this more successful so an individual's task can be shared by a group?

END/TRANSFORMATION OF EXPERTISE; MULTIPLE VOICES

Preserving Expertise/Spectrum of Authority

• It seems to me crowdsourcing or peer sourcing or community development is about tapping into expertise, but it might not be in front of your face. It doesn't mean that there's less expertise.

• I view it as "troubling" authority, challenging what we know we're asking for already and questioning our practices and changing those sets of practices that are dominant now.

• There is also the preserving of authority and expertise. In certain situations there is too much noise, and expertise is flattened. I'm all for valuing the crowd and troubling expertise but in certain situations, surfacing expertise in some way or surfacing some form of it is needed.
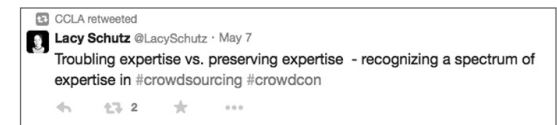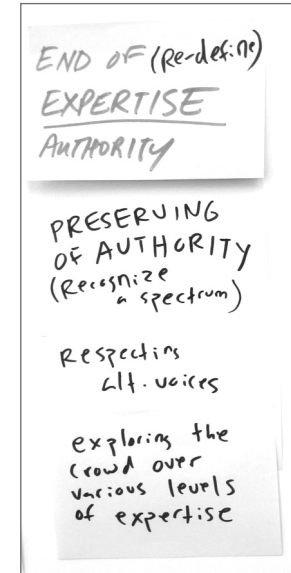
• We can value it in a different way, kind of spread it out, decentralize it, and see how certain kinds of functions may become clearer.
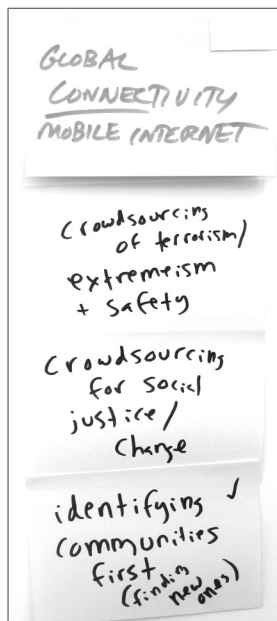
• It might be expressed as a spectrum of expertise.

Respecting Multiple Perspectives/Voices

• In that vein too there is just perspective. One of the tricks is that a lot of this seems to be about objective tasks, and when we get more and more into elements in the humanities, there is a clear idea that people should have different perspectives, or will, and that respecting that and laying that out is giving people a voice, which I think is a different tack than trying to get the right answer.

Honoring/Elevating Crowd Expertise

• I like this example of public governance and being elevated to the directors' circle. This answers some of the other questions about how we overcome barriers of class and race, how we overcome issues of redistributing expertise and authority. What you are doing when you put people in the position of being in a leadership circle is you are allowing them to steer resources, to answer questions that are important to them and not necessarily to the people who established the project. I think that's the kind of relin-

END OF (Re-define)
EXPERTISE
AUTHORITY

PRESERVING
OF AUTHORITY
(Recognize
a spectrum)

Respecting
dift. voices

exploring the
crowd over
various levels
of expertise

CCLA retweeted
Lacy Schutz @LacySchutz · May 7
Troubling expertise vs. preserving expertise - recognizing a spectrum of expertise in #crowdsourcing #crowdcon
2

quishment of control that you need in these projects in order to really address some of these other things like transparency and openness in government, in science, in social structures.

- So an openness and willingness to go where the group takes it will get over some of the biases and prejudices that are built in in other more structural ways.

GLOBAL CONNECTIVITY/MOBILE INTERNET/ IDENTIFYING COMMUNITIES

Terrorism, Extremism & Safety

- If we're thinking broadly, I don't think we can ignore extremism and terrorism and at the same time, because I think it's all related, calls for justice that are there. We are seeing the crowdsourcing of terrorism, and it's a factor.

- Something mentioned in a previous workshop was the idea not of public good but public evil, people who actually want to poison the well of what we're doing who might be out there for whatever reason, whether they have a political agenda or not. The fact that those agents and actors might be there is something to contend with as well.

Social Justice

- There is also the use of crowdsourcing to promote social justice and social change.

Identifying and Attracting Communities

- The popularity point is an interesting one. Who are the people who are going to want to do this? When someone says it's an astronomy project, you know there is an amateur astronomer community that is out there. I think it's interesting that you can map out communities like that: people into trains, people into ships, genealogy. There are a lot of these communities and groups, and in some ways it may make more sense for us to be thinking backwards from where we know communities are instead of forwards from what we think is our most interesting material.

- So you know there are these communities out there, but as we heard in this morning's discussion there are other communities that find their way to it as well, so there's this balance between targeting communities and being open to those who may find their way from a different community.

- Chris Lintott: Just a point of clarification on that. Our astronomy projects don't have many amateur astronomers in them, and most participants have never looked through a telescope. Those amateur astronomers have a different way to engage in astronomy and are excited by data in a different way and would need a different kind of platform. So actually we've found it easier in most cases to look for new people.

- That's an interesting point. These communi-

GLOBAL
CONNECTIVITY
MOBILE INTERNET

Crowdsourcing
of terrorism/
extremism
+ safety

Crowdsourcing
for social
justice/
Change

identifying ✓
communities
first in
(finding new ones)

ties already have rules of engagement going, so they don't need you.

- Chris Lintott: eBird served their community by building tools that the community wanted. That's very different than building a project designed to do something else.

- So it may be easier to call into being a community as opposed to targeting one that already exists.

Defining "Community"

- I think we're also talking about communities in two different ways. There is "community" as people who have similar interests, such as astronomy, and then there are communities that have historical cultural, linguistic, or social ties that identify them as a human community. I think those are two different things. You might not be able to target astronomers, but you would absolutely be able to target a collection of historical documents related to those people. I think those are two different things. They are not necessarily exclusive, but we should just be aware of them.

- So affinity groups that have something in common, or a community, a neighborhood, a demographic group, young mothers. There are lots of ways to slice it. And that same person may be in several of those.
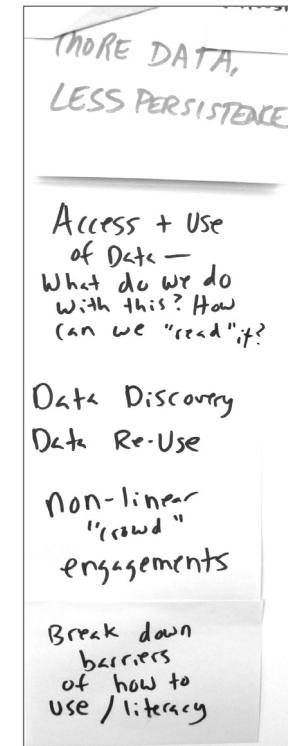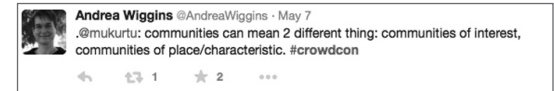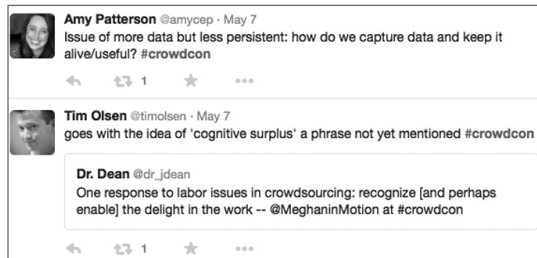
Data Access and Use/Data Discovery

- There is something to do with the persistence of data—how you deal with access and use as much as anything else. So the idea is, we have all this data. Are there ways we can funnel it to appropriate interfaces or formats? It answers the question of what comes next. After I do this project, can I use it in the context of other collections, or can I visualize it in some way? Or where do I go to find what I need or somewhere to put a stamp of approval on this as a legitimate data set? So it's just the idea of finding ways of using the data that we have so that there is a lot of stake in keeping it around.

- Can we call that "data discovery"?

- I think so, and reuse.

Non-linear Crowd Engagement/Dabblers

- Getting back to this notion of designing projects for super users or maybe engaging a larger group, there is a paper done about a Zooniverse project called Old Weather, designed for dabblers who will come and do a little bit and then go away. A lot of projects require more time from people and large tasks, so one approach is to think about how to break those tasks down into smaller units. On the one hand it might seem like it is breaking it down too much, but it is good to have different ways for people to engage.

Andrea Wiggins @AndreaWiggins · May 7
.@mukurtu: communities can mean 2 different thing: communities of interest, communities of place/characteristic. #crowdcon

THORE DATA, LESS PERSISTENCE

Access + Use of Data — What do we do with this? How can we "read" it?

Data Discovery
Data Re-Use

Non-linear "crowd" engagements

Break down barriers of how to use / literacy

CCLA retweeted
Neil Fraistat @fraistat · May 7
@perrycollins stresses importance of data discovery, access, and reuse. #crowdcon

Amy Patterson @amycep · May 7
Issue of more data but less persistent: how do we capture data and keep it alive/useful? #crowdcon

↩ ♺ 1 ★ •••

Tim Olsen @timolsen · May 7
goes with the idea of 'cognitive surplus' a phrase not yet mentioned #crowdcon

Dr. Dean @dr_jdean
One response to labor issues in crowdsourcing: recognize [and perhaps enable] the delight in the work -- @MeghaninMotion at #crowdcon

↩ ♺ 1 ★ •••

CCLA retweeted
Daniel Powell @djp2025 · May 7
Which is not to say thinking about public facing/integration of knowledge work is not vital, but would we feel the same pressure? #crowdcon

↩ ♺ 2 ★ 1 •••

CCLA retweeted
Daniel Powell @djp2025 · May 7
I honestly think we wouldn't be so interested in crowdsourcing without the gradual collapse in public funding for heritage & ed. #crowdcon

↩ ♺ 2 ★ 3 •••

It's also important to have educational materials.

- I like that concept, design for dabblers. Even heavy users may have a minute while waiting at a bus stop, or five minutes. In that little bit of time, give them something to do that you can stitch back together in a useful way.

- And you can get a lot of effort out of that.

Break Down Barriers to Use/Literacy

- I see a real skills gap in humanities, a broad kind of inability or unwillingness to work with the kind of data you might get out of crowdsourcing, the kind of large, quantitative approaches. I think we want to help humanities types use this data, but also more broadly to transfer these skills to a broader population.

- So fill those gaps that may hinder someone from trying it out.

- From using the results of this. It's not enough that someone wants all of the transcriptions, it is: What are you going to do with that?

- Related to that, there is the issue of literacy and the changing meaning of what it means to be literate.

- You mean alternate language literacy like coding or statistics?

- I mean even literacy in social media, literacy in any of these things that might empower a person or might not and being conversant in those worlds. So humanities folks being literate about data.

- Howard Rheingold calls them "21st century literacies," those kinds of things that we need to understand. One of my favorites is "crap detection," being able to understand what is good data, what is junk, what are the filters and processes that we need to have in place to be able to distinguish good information from bad.

ROUNDTABLE: A VIEW FROM THE FUNDERS' PERSPECTIVE

Introduction

Moderator: Sharon Leon, George Mason University, Rosenzweig Center for History and New Media

Neil Fraistat kindly asked me to facilitate the conversation with the funders. I am by no means a funder. Most of the time I am asking these folks for funding. I would be interested to hear what everyone has to say about funding priorities for communitysourcing projects, the way you think that these priorities fit into your larger portfolios, and the ways that we can think about how to amplify the work that we do with communitysourcing projects across agencies. As someone who works in history and the humanities, to have someone from NIH here who thinks about science and science implementation is a really good perspective for me to hear, and I'm looking forward to this conversation. We will just go down the line and start with Perry Collins from the National Endowment for the Humanities.

National Endowment for the Humanities

http://www.neh.gov/divisions/odh

Perry Collins, NEH, Office of Digital Humanities

I am from the Office of Digital Humanities (ODH) at the National Endowment for the Humanities (NEH). ODH is actually the smallest of NEH's grantmaking divisions. We are small but mighty, and we have funded most of the representatives and crowdsourcing projects that are here. I was looking back, and while we had previously funded participatory archives and public history projects, the very first project we funded that self-identified as a crowdsourcing project and called itself crowdsourcing was Mary Flanagan's planning grant for Metadata Games in 2009, which was followed very quickly by the Scripto project that Sharon Leon ran, which I think began the next year.

So we have been funding this for a few years and I wouldn't say that it has been a flood—the gates didn't open and suddenly everything is crowdsourcing all the time—but it has been a healthier trickle of projects of increasing complexity, building on what has come before, and it has been interesting to watch that.

Very briefly, before we get into questions, I will talk in terms of what we would like to see at NEH. On the one hand we would like to see how crowdsourcing or communitysourcing, or however we want to define it, can engage new and diverse publics in accessing and contributing to some of the same primary sources and collections that academic researchers may be more privileged to access more deeply and more often. This might be projects based in collections we are familiar with that range from work with historical menus, to manuscripts, to correcting newspaper OCR. How can
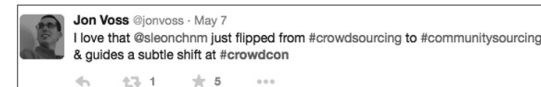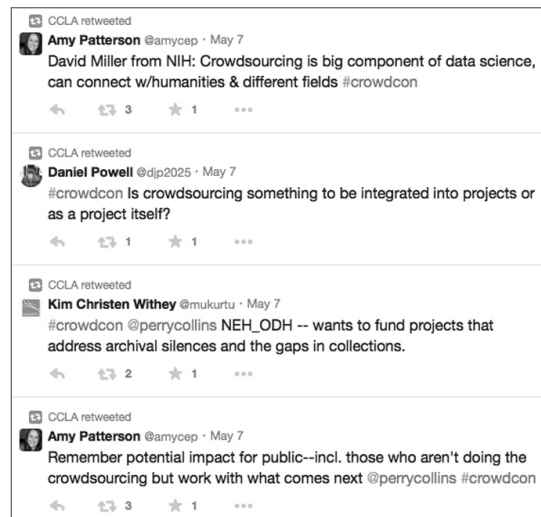
Roundtable Participants

Moderator:

• Sharon Leon,
George Mason University, Rosenzweig Center for History and New Media

Panelists:

• Trevor Owens, Institute of Museum and Library Services (IMLS)

• David Miller,
National Institutes of Health (NIH)

• Perry Collins, Office of Digital Humanities (ODH), National Endowment for the Humanities (NEH)

Jon Voss @jonvoss · May 7
I love that @sleonchnm just flipped from #crowdsourcing to #communitysourcing & guides a subtle shift at #crowdcon
↩ ⟲ 1 ★ 5 •••

CCLA retweeted
Lauren T @nolauren · May 7
Excited about idea of community sourcing projects that address historical and archival silences and gaps #CrowdCon
↩ ⟲ 1 ★ •••

CCLA retweeted
Amy Patterson @amycep · May 7
.@perrycollins from @NEH_ODH: seeing more projects with increasing complexity, building on what's come before #crowdcon

CCLA retweeted
Amy Patterson @amycep · May 7
David Miller from NIH: Crowdsourcing is big component of data science, can connect w/humanities & different fields #crowdcon
↩ ⟲ 3   ★ 1   ...

CCLA retweeted
Daniel Powell @djp2025 · May 7
#crowdcon Is crowdsourcing something to be integrated into projects or as a project itself?
↩ ⟲ 1   ★ 1   ...

CCLA retweeted
Kim Christen Withey @mukurtu · May 7
#crowdcon @perrycollins NEH_ODH -- wants to fund projects that address archival silences and the gaps in collections.
↩ ⟲ 2   ★ 1   ...

CCLA retweeted
Amy Patterson @amycep · May 7
Remember potential impact for public--incl. those who aren't doing the crowdsourcing but work with what comes next @perrycollins #crowdcon
↩ ⟲ 3   ★ 1   ...

Ben Miller @Intransitive · May 7
#crowdcon @NEH_ODH "Small but mighty" funder of humanities projects with crowdsourcing elements that enhance and create collections.
↩ ⟲ 1   ★   ...

Joe Carrano @joecar25 · May 7
Watching the #crowdcon stream from 3 floors below where it's happening. Sounds like a good time from what I can hear: original.livestream.com/engagingthepub...
↩ ⟲ 1   ★ 1   ...

Lauren T @nolauren · May 7
Excited about idea of community sourcing projects that address historical and archival silences and gaps #CrowdCon
↩ ⟲ 1   ★   ...

we ask the public to contribute their unique abilities to develop those projects?

But we are also thinking about where are the archival silences, where are the gaps in collections, and how can we also ask communities to develop entirely new data corpora of digital media or histories or photographs? The last panel talked about some of these digital-memory-based projects that we've developed. How can we make sure that we are both enhancing what we already have and actually creating whole new collections?

On the other side, especially from the perspective of the Office of Digital Humanities, we are also really keen to help all of the institutions running crowdsourcing efforts to feed all of this back into something else, to somehow push this back into their own collections—maybe pushing it down into smaller projects that they're doing, maybe pushing it up into bigger aggregations of data, something like DPLA—and making it easier to discover, to analyze, and to interpret this cultural heritage, both at the level of an individual item but also at scale in the aggregate. How can we help both researchers and the public make sense of all of this once we have it?

And then we are actually encouraging the preservation of our cultural heritage by doing this. This is important enough that we want to keep working on it, and it's not really just a static digital collection that is going to sit there but something we are creating a commu-

nity around. As much as we talk about public engagement and participation in the act of crowdsourcing itself, it is important that we don't forget how much potential impact there is on the other side too—that there is actually a lot of potential for the public, people who aren't doing the crowdsourcing, to work with all that comes after that with all of that data. That is my overview.

National Institutes of Health

http://www.nih.gov

David Miller, NIH

In addition to my program director duties at the NIH, I am also an AAAS Science & Technology Policy Fellow, so I have an interest not just in the technological components of crowdsourcing but also the policy, things that can be done. Some of the questions I ask about crowdsourcing are: How can we improve the ability of program folks within NIH to issue funding opportunities for the communities that want to do crowdsourcing type projects? And do they need to be specific, or are there opportunities within existing funding opportunities that perhaps aren't so clear and could be improved by giving additional notice or just getting the word out that this funding is available? And I will get to those existing funding opportunities in a moment.

NIH views crowdsourcing as a large umbrella that covers partnerships between the traditional primary investigator/researcher funded

by NIH and the crowd or community that is being engaged with a project. So it doesn't have to be this top-down view that, for whatever reasons, may be perceived by some agencies. I should also say that this is not something new for NIH, we have been involved for this for quite some time, funding these projects.

I think what keeps some people back, particularly from the humanities, is that you hear, "Well, NIH's mission is funding biomedical health initiatives." But that can be quite broad. There is health-relevant information within social media streams, with behavioral and socioeconomic data—a wide swath of expertise. To me, and I don't know if other people share this, crowdsourcing is a component of data science, or vice versa, and since it is inherently a multidisciplinary field you are going to be having these sort of partnerships between fields. Stepping back for a moment, anything that a programmatic person at an agency can do to improve or increase the number of these types of community partnerships will improve the general space for what you folks are trying to do here.

Institute of Museum and Library Services

http://www.imls.gov

Trevor Owens, IMLS

I am in the Office of Library Services at the Institute of Museum and Library Services. Our office is focused on the work of libraries and

archives across the country. We are focused on work that will help libraries and archives achieve their missions, meet the needs of their publics, and amplify their abilities to be the most compelling and competitive. There are two grant programs we run which I think are likely most relevant to folks here. We have the National Leadership Grants for Libraries Program, which supports a range of projects that have the potential to make a broad national impact, and the Laura Bush 21st Century Librarian Program, focused on the future of librarianship and education and training for librarians and archivists.

I think there are a lot of opportunities in both of those programs for work in this space. Earlier in our conversations here we explicitly brought up the "What is the future of librarianship in this?" question. In that vein, work that could help to equip librarians with the skills and abilities to make this work happen in libraries and archives is clearly relevant to the Laura Bush 21st Century Librarian Program. In particular, a huge part of that education and training is going to need to be ongoing professional development because in our organizations we are not going to just hire our way into the future workforce. We have amazing librarians who have deep and abiding skills and expertise and we need to figure out what we can do to

Daniel Powell @djp2025
The Wellcome Trust might be of interest to those looking @ NIH affiliated/interdisciplinary projects. wellcomelibrary.org/what-we-do/dig...

http://wellcomelibrary.ort/what-we-do/digitisation

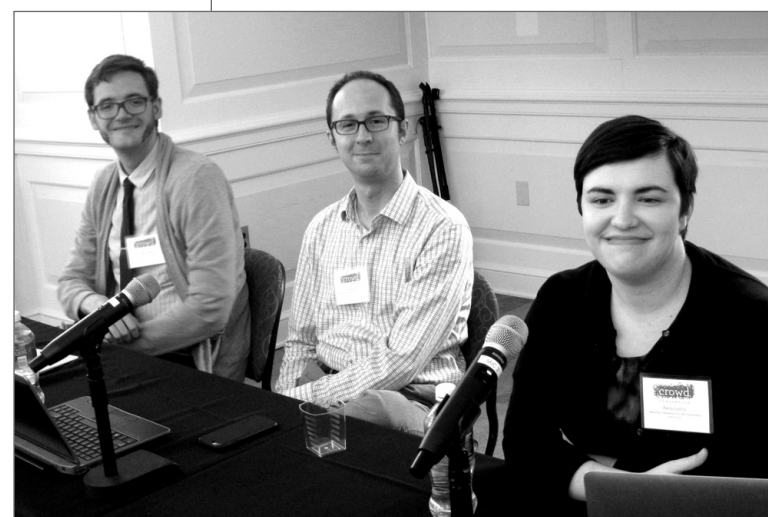CCLA retweeted
critical play @criticalplay · May 7
resource from David Miller - Data Science program at NIH ow.ly/MFhWU #crowdcon
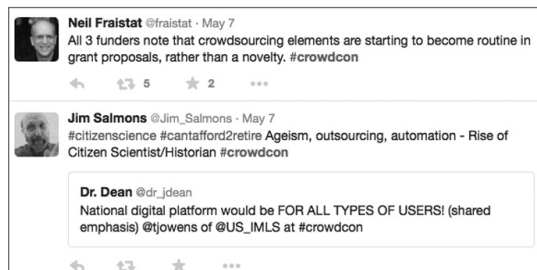
http://ow.ly/MFhWU (BD2K site)

Panelists (from left):
Trevor Owens, David Miller, Perry Collins

help them make use of tools and techniques for crowdsourcing. That is a huge thing.

At IMLS, my particular area of focus is the National Digital Platform priority, which spans across both the National Leadership Grants Program and the Laura Bush 21st Century Librarian Program. I'll briefly talk a bit about that as some context on this priority. The National Digital Platform is the combination of software, social-technical infrastructure and staff expertise that provides a wide array of content and services to all users in the United States. Let me underscore "all users"—every type of person, every area of interest. All kinds of libraries should be underscored also, including a wide range of academic, public, state, tribal, rural, you name it. These are libraries serving communities in the United States, and the extent to which tools and services can be deployed to help them serve their individual organizations' missions and meet the needs of their constituencies is the kind of thing that is going to resonate most with the folks who review our proposals.

Questions, Answers, Discussion

CURRENT NUMBER OF CROWDSOURCING PROPOSALS AND GRANTS

• Of the kind of proposals that you field and read and support, can you give us a sense of how many of them are directed towards the kinds of things we have been talking about today in that broad spectrum, as David Miller suggested, of what crowdsourcing, communitysourcing, participatory work can be?

• Sharon Leon, George Mason University

NEH: From Experimental to Refinement of Tools, Models, Best Practices

• I think overall we've seen a big uptick in terms of projects. Even if they don't define themselves as crowdsourcing *per se*, they are trying to engage new kinds of communities in doing the work of the humanities, and I think they are engaging with a lot of the big challenges and questions that we are asking here today. But when it comes more specifically to projects that did talk about crowdsourcing, I think it was Mary who talked earlier about moving from the edge to the core. We are finally, after a few years, starting to see some of these earlier crowdsourcing projects become a little bit run-of-the-mill in a good way. Our reviewers are saying, "This isn't really experimental anymore. There is a tool to do this." Obviously it is still a challenge, you can't just assume it will happen and work, but maybe this is something that can be attached to another kind of program. And in fact, other NEH divisions have started to fund crowdsourcing as part of broader digitization and digital collection projects. It is one more tool in the box for doing that.

And I am actually very happy to see that because now, in Digital Humanities, we can move on and say, "Okay, can we fund things like this?" Can we establish best practices?

Can we fund the training programs? Can we identify gaps in the tools that still need to be developed? What are the models that still need to be worked on? And we have started talking about what are the more complex kinds of questions that we can ask, while still having the funding for the simpler kinds of work that we may know how to do at this point. • Perry Collins, NEH

NIH: Initiatives Targeting Crowdsourcing

• I can actually give a couple of specific examples. A big chunk of my time is spent on a data science project called the Big Data to Knowledge initiative (BD2K) [http://datascience.nih.gov/bd2k]. What is interesting about this is that NIH doesn't define what is considered "big data." So you have the typical assumptions—they are very, very large data sets, they're heterogeneous data sets, they're all high throughput—but this is really much more about capture issues of dealing with collections of digital objects that, for whatever reasons, don't work well with each other. So it can be a data integration problem, it can be any number of things. What we define as "big data" for this project doesn't have to be the standard definition. It is really what is big for your community.

There are some examples from biomedical health. Genomics data can be absolutely massive compared to, say, protein structure data. But for folks working in protein structure the data is still very, very large.

Comparing the two is a bit apples and oranges and NIH understands that. So your data set doesn't have to fit a pre-assumed mold of what big data is.

There are initiatives such as the Biomedical Information Science and Technology Initiative (BISTI) [http://bisti.nih.gov] that specifically calls out crowdsourcing and collaborative online environments and other additional crowdsourcing related tools as part of its software development initiative. And there are four different versions of that and they are all pretty similar. There is early stage, later stage, some for small businesses, and Small Business Technology Transfer (STTR), which is a partnership between the scientist PI, a business and an academic institution.
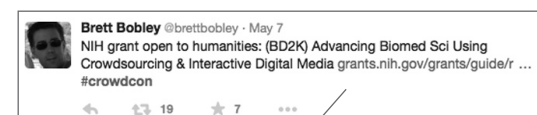
I think the reason that I am here is that I was at the Game Developers Conference a few months ago and ran into Mary Flanagan and was talking to her about an initiative that we have out now incorporating the use of interactive digital media, video games, etc., within a crowdsourcing context for solving biomedical data analysis problems. We don't say what your specific problem needs to be, but it is going to be covering any component of health-related information problem sets because again, this is a trans-NIH initiative. This is a cooperative effort of all 27 institutes and centers. If you just Google "digital media NIH" you will find information about that opportunity [http://grants.nih.gov/grants/guide/rfa-files/RFA-CA-15-006.html].

http://bisti.nih.gov

> **Amy Patterson** @amycep · May 7
> David Miller from NIH: BISTI calls out crowdsourcing, collaboration bisti.nih.gov #crowdcon
> ↩ ⟲ 1 ★ 2 ⋯

> **Daniel Powell** @djp2025 · May 7
> Medical Humanities project in Britain (The Casebooks Project, dig med records from 16/17th centuries): magicandmedicine.hps.cam.ac.uk #crowdcon
> ↩ ⟲ 2 ★ 2 ⋯

> **Daniel Powell** @djp2025 · May 7
> Big Data to Knowledge: datascience.nih.gov/bd2k #crowdcon
> ↩ ⟲ 3 ★ 6 ⋯

http://datascience.nih.gov/bd2k

http://magicandmedicine.hps.cam.ac.uk

> **Brett Bobley** @brettbobley · May 7
> NIH grant open to humanities: (BD2K) Advancing Biomed Sci Using Crowdsourcing & Interactive Digital Media grants.nih.gov/grants/guide/r ... #crowdcon
> ↩ ⟲ 19 ★ 7 ⋯

http://grants.nih.gov/grants/guide/rfa-files/RFA-CA-15-006.html
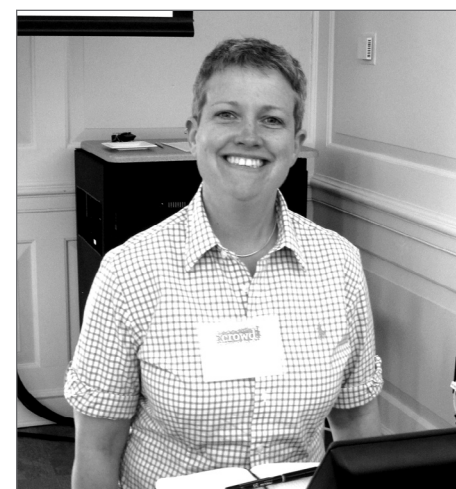
Moderator Sharon Leon

If I can switch gears a little bit, going over to training, there has been a big, big push recently with the concept of training new data scientists at all levels of academia, not just post-graduate. We are trying to create Ph.D.s in library sciences and data curators and such, but also at earlier stages as well. Under BD2K there are a number of funding opportunities [https://datascience.nih.gov/bd2k/announcements], and one of them is even specific to short courses and MOOCs and bootcamps and such. And as crowdsourcing is a data science field, it is possible to apply to that source of funds if, for example, you want to run some sort of summer bootcamp on best practices for crowdsourcing. It could be the technology underpinning that, it could be the crowdsourcing itself, it could be anything within that. NIH leaves it up to you to define how you submit these things.

So there is certainly no lack of opportunities at NIH for funding these types of things. It may seem like I am leaving out behavioral health sciences and others; I am not, these are all inclusive within these data science projects. This is one of the goals of the BD2K and its leadership, Phil Bourne, who is a bit of a rebel whom NIH pulled from the West Coast to lead NIH's new data science initiatives. He is very open to these out-of-the-box, integrated initiatives. • David Miller, NIH

IMLS: Building Capacity, Creating Capabilities, Moving the Needle More Broadly

• Our agency is seeing more and more crowdsourcing projects or projects that have some crowdsourcing element added to them. That is part of what made this workshop so compelling. As opposed to everyone adding it to some project they proposed where they were now going to do crowdsourcing or something like that, there is potential for common tools, practices and approaches and for clear ways to evaluate them. To that end there is a need for the field to clarify and answer questions like: What does success look like? How does scale work in those cases? I think that gets into questions about what is going to make it more competitive going forward because in all of our cases we have run competitive grant programs. It's not about what could be funded, it's more about what would be the most competitive.

In that vein, particularly in keeping with the priorities that we're seeing, at the end of the day we can't fund everyone to do one little thing that runs alongside of a collection or something like that. The things that I think are ultimately going to be the most competitive are ones that build capacity and create the capabilities for this to become more and more a part of the work of our organizations. In that case, I think it's probably best to think about all of this as fuel to catalyze it and look for things that are going to move the needle more broadly. That is the

challenge to our applicants. As there is more work and knowledge in this space it ups the need for demonstrating and making those more competitive. • Trevor Owens, IMLS

MAINTENANCE AND SUPPORT

• I have a question about ongoing maintenance and support that I think ties in with a number of things that were just said. They are not sexy but very necessary things, especially as we think about potentially creating platforms or code bases that we hope people will recycle and reuse. Do your grant bodies have any interest in that, or do you see it more as an institutional problem? • Victoria Van Hyning, Zooniverse, Oxford University

Required in Applications and Part of Early Advisory Board Conversations

• I think it's both. We are certainly very interested and worried about it too. We can't fund these projects in perpetuity, and not many projects even necessarily get more than one grant to do something. I think part of why these kinds of events are helpful is that as funders we can start to coordinate a bit. What are the different pieces of projects that are appropriate? How can we encourage our applicants to be more savvy in how they apply for funding so that you might apply to NIH to do one thing and you might apply to IMLS to do another thing?

But at the end of the day, part of the reason why we ask our applicants to develop data management and sustainability plans is that we want this to be part of the conversation from the very first advisory board meeting. We don't have the answer for you either. We're funding often 10% of the projects that we get any way. At the very least we want to see that we are starting to have this conversation. I don't really have a good answer. Maybe one of the others could answer this question. • Perry Collins, NEH

Fitting It into General Management and Interoperability Grants

• I would echo a bunch of that. Within programmatic talks, we do have conversations about maintenance and how to support that. Are you going to see a specific, full-out funding opportunity for maintenance? Maybe, maybe not, but with your ability to submit to sort of general data management interoperability grants and things like that, you can incorporate some sustainability with expansion of your mission. You can fit things in. I agree that it can be a challenge on both sides because this issue of "How sexy is that?" does come up when attaining support for maintenance versus new development. • David Miller, NIH

Opportunities & Models for Emerging Innovative Networks; Legacy of Expertise
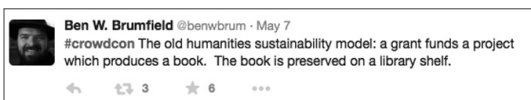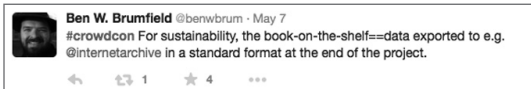
• It is important to recognize what project-based grant funding is and what that means. There is a call for proposals, so let's go after

Graceful Degradation - Preservation Priorities

• This also goes back to Neil Fraistat's point earlier about "graceful degradation." From the beginning of projects, or at least very early in projects, can we start prioritizing about what we really need to preserve rather than saying we want to keep everything and keep maintaining it and having it forever. It won't be possible with everything, so at least try to establish some priorities for different projects. Also, document how you go about it, deciding what is the most important thing to send off to the library for long-term preservation.

Or are there ways to put this in your budget? At NEH we make a point of saying you can put fees for data management into your budget. Obviously that doesn't get to all of the aspects of sustainability in terms of administrative commitment and all of that, but in terms of preserving parts of the project I think we can do that. And we can do a better job of making sure applicants know that can be part of these projects too. • Perry Collins, NEH

Ben W. Brumfield @benwbrum · May 7
#crowdcon For sustainability, the book-on-the-shelf==data exported to e.g. @internetarchive in a standard format at the end of the project.
↰   ⇄ 1   ★ 4   •••

CCLA retweeted
Amy Patterson @amycep · May 7
.@perrycollins: From the beginning of projects, can we start prioritizing what we really need to preserve? Can't grow forever. #crowdcon
↰   ⇄ 1   ★ 1   •••

Ben W. Brumfield @benwbrum · May 7
#crowdcon The old humanities sustainability model: a grant funds a project which produces a book. The book is preserved on a library shelf.
↰   ⇄ 3   ★ 6   •••

that, that sounds great, but at the same time, in the sustainability section you are supposed to make a case. If the case is, "We'll be back in 18 months for continuing operations," that is not going to be a compelling case. If you look at library budgets across the United States and where that money is, IMLS is not the lion's share by any stretch of the imagination. At the end of the day the bulk of the library and archive resources are local.

And ideally, folks are coming up with ways to catalyze and use our project-based funds to help establish and develop ways of sustaining projects, programs and organizations. So if you look at things like DuraSpace that have emerged to carry on the work, they become membership organizations that have demonstrated that value, they get the buy-in when they come and do project-based work in the future, they've got the sustainability capability. I think in the same way the Center for History and New Media (CHNM) is in a position where they've got a lot of different sorts of strands of funding that are associated with projects. In many ways another element with this is that you see organizations with connections with professional associations, you see things falling into the core operations of some organizations, or you see long-term commitments that emerge in collaborations.

But it's a challenge to everybody. For example, humanities scholars' sustainability

plan for a book is that libraries keep them on their shelves. When someone is making interactive media, that is more complicated. We need as much innovation in that space as we do in the space for creating tools and services. That is, what are the organizations and communities that can take on this work and sustain it?

But there is the education and training program that we have, the Laura Bush 21st Century Librarian Program. I think the sustainability for some of this becomes the expertise and developing professionals who can participate in them. You look at some of these projects and you see very active communities drawing around them all sorts of elements that have to do with professional practice, and this is becoming part of the work that we do for the bread and butter parts of our organizations.

So it is complicated, but there is definitely a lot of opportunity in there for emergence and innovation around ways to sustain these activities.   • Trevor Owens, IMLS

CO-FUNDING WITH CROWDFUNDING

• Patrick O'Shea made the point yesterday that each of your agencies are really doing crowdfunding. That is, they are spending tax dollars. Recognizing that, I'd like to ask a question about crowdfunding in the way we typically use the term. A lot of these projects are ideally positioned to do crowdfunding

perhaps, and I'm wondering if your agencies have any kind of either officially stated policy about co-funding with crowdfunding platforms or unstated positions. Maybe they're not in policies but rather the sort of thing you talk about at the water cooler.
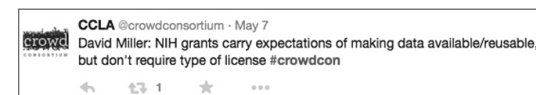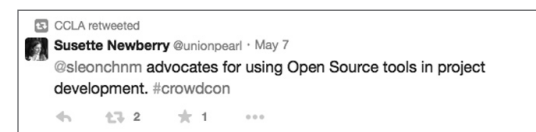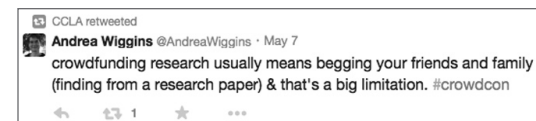• Austin Mast, Florida State University, iDigBio

Pro Co-Funding in General

• I think certainly we want to encourage co-funding in any way. It might be through crowdfunding, it might be through foundation funding, it might be through another federal agency. It is deciding when projects garner enough interest to want to do crowdfunding, but of course that has all of the challenges of doing a crowdsourcing project in the first place, of actually getting people motivated and committed enough to donate to something that they think is important. So I wouldn't necessarily specifically say we have a position on crowdfunding. It's one more way that these projects can keep going. • Perry Collins, NEH

Crowdsourcing Input as Well as Tax Dollars

• I'm not aware of any limitations at all on how you attain additional funds and as you say, all of the agencies are doing broadly defined crowdfunding. In essence, there are going to be a lot of investigators that are co-funded, that is, funded by NSF and also by NIH. NIH's perspective is you can't be more than 100% funded by NIH, but that doesn't extend to other sources of funding.

And I should say that we do more than just crowdfunding, we also do crowdsourcing of information that directs how we spend our funds. We put out requests for information, we do public engagement in various forms, not just with the agencies themselves but also sometimes through broader means. The Networking and Information Technology Research and Development (NITRD) program [https://www.nitrd.gov/] is a higher level group set up by the Office of Science and Technology Policy that lays out the landscape and strategy for spending federal funds across all the agencies, usually within some specific technology, whether it's big data or online security or whatnot. They and OSTP also ask for public input in this process, and that's open to everyone from university experts to high school groups.

CCLA retweeted
Andrea Wiggins @AndreaWiggins · May 7
crowdfunding research usually means begging your friends and family (finding from a research paper) & that's a big limitation. #crowdcon

CCLA retweeted
Susette Newberry @unionpearl · May 7
@sleonchnm advocates for using Open Source tools in project development. #crowdcon

CCLA @crowdconsortium · May 7
David Miller: NIH grants carry expectations of making data available/reusable, but don't require type of license #crowdcon

SANCTIONING SUPPORT OF KEY OPEN SOURCE TOOLS/INFRASTRUCTURE

• I have a question about those of us building critical pieces of infrastructure that other projects are building on. Do you have the opportunity to suggest that people who use open source tools may want to kick back some open source money for support and maintenance of those tools since they are using those critical pieces of the infrastructure? • Sharon Leon, George Mason University

• I don't know that we are empowered to direct how people spend their money once they've developed things. I would say that within grants that are funded by the NIH, there are expectations of data sharing and expectations of software sharing and dissemination. Some of them have more teeth than others depending on the program, but this is public funding going to public projects and you are expected to make your data available as widely as possible. We don't usually go to the fine-grain point of saying which specific license; rather, the sharing and dissemination language within a particular funding opportunity will lay out the expectations NIH has for funded applicants to release their project's software and source code. • David Miller, NIH

Slim Odds

- I want to make a public service announcement. Crowdfunding is really just fundraising and you do it on Indiegogo. And only in an infinitesimally small percentage of cases is your stuff going to go viral and people you've never met before are going to chip in money.
  • Nick Adams, University of California, Berkeley

And if you think you have a special case with your crowdfunding project, you can always reach out to whatever funding office or agency and inquire. • David Miller, NIH

Tax Liabilities, Institutional Restrictions, and Rhizome Example

- The only other thing I'd add is that diversifying the resources that support our work is always going to further demonstrate the value that you're providing. I remember seeing someone on Kickstarter saying, "I'm going to raise matching funds for my NEH grant on Kickstarter," and I think the guy probably just got a 1099 and paid taxes out of it and just showed up with money. In many cases though, there are restrictions that exist around those sorts of things, likely more in the Office of Sponsored Projects area for institutions because obviously, how organizations do fundraising is tied up in some places within their own system. Different parts of organizations want to get their cut and want to make sure all of the policies and requirements for working with funds are being accounted for.

So it's something exciting for people to be exploring, but it's also something that people need to go into realizing that while it might be easy to just jump in and start a Kickstarter project, there are all kinds of other implications about what would happen in terms of tax and a variety of other things. Rhizome recently did a Kickstarter project

that funded putting up a set of games online, an emulator that they provide access to. That sort of funding does demonstrate the vitality and the fact that there's a community that really wants to see the work happen.
• Trevor Owens, IMLS

BEST "BIG ASK" APPROACH TO SINGLE AGENCIES

- Is there a preference to have a small project that is substantively interesting and self-contained to advance knowledge in one particular area versus a project that may be large and creating some generalizable tool but costs more money? More specifically, I wonder if you'd give guidance on how someone approaches funders when they know they have a big ask that might look expensive and they don't necessarily want to go through the rigmarole of finding seven different agencies and having this uncertainty across all of these agencies. • Nick Adams, University of California, Berkeley

From Starting a Conversation and Community to Strong Use Cases

- In humanities you're probably going to have to go through the rigmarole. We don't give out a lot of multi-million-dollar awards. In terms of small versus big generalizable projects, it can go both ways. We fund a lot of small, focused projects that take particular research questions or problems. The most successful ones make a good case for why

these are generalizable questions or might lead to something bigger. Our reviewers are savvy enough now to see through the: "We'll be a model for everyone always, forever." That's not going to work. But I think if you can make a real effort to say you are starting a conversation—I was just reading Bethany Nowviskie's talk from the IMLS last week about starting a community around these kinds of questions—then that's good. But on the other side, if you are going for a bigger grant to do something more generalizable it goes down the same way. Do you have good case studies? Do you have solid examples and people who come to it with their own perspectives and their own projects? Those big generalizable platforms tend to fall flat unless you have strong use cases and people who are already excited about doing it. • Perry Collins, NEH

Tailor a Legitimate Request, Don't Over-Ask

• NIH does do $2-million-a-year grants for things like Centers of Excellence, but for that you definitely have to go through the rigmarole. I will say, though, that you should strategize and you shouldn't try to overly-game your project. It's best to start from a perspective of yourself and what you want to do and keep in mind that at all of these agencies your proposals are being reviewed by your peers as carefully as possible. If what you're trying to do requires a big project versus something small, then ask for that. If not, then ask for a small exploratory grant.

Again, the reviewers are going to be very savvy and can see that you're just trying to add this fourth thing because you want to have four aims. They will notice and sometimes you will be docked for that, sometimes they will just suggest that you remove that aim and take a sufficient reduction in costs. Again though, when it's 10% or 20% of things getting funded, competition is going to be fierce. Any attempt to get around things is not going to work. • David Miller, NIH

Peer Review, ROI, and Coherent Pieces

• I was going to say a similar thing about peer review. The folks proposing projects are often in a really good position to evaluate and ask, "Would I be compelled by this?" In that vein too, I would add that when the peer reviewers are looking at it there is an element of return on investment. I think that's a similar point as well. There are small projects that make a great case for how they're going to make an impact, and that impact is commensurate with the budgets, and there are really big projects that do that. It's about the pieces fitting together and being coherent. • Trevor Owens, IMLS

> **CCLA** @crowdconsortium · May 7
> .@perrycollins: Reviewers are savvy to vague claims about your project being a generalizable model, so make detailed effort #crowdcon
> ↩ ⇄ 2   ★ 1   •••

Tap Program Manager Expertise

• Also, the rules can be different at different agencies and funding institutions. If you have questions, call the program managers. That's what we're here for and that's why we have this Letter of Intent due date so that someone at the agency can discuss responsiveness or other issues with you and you can make changes. They will help you hone your project. Things that go for review often benefit by having early conversations with an agency's program staff. If you've pre-contacted the person they'll know you, they'll understand your project better, and they'll be in a better position to present your project in its best light. • David Miller, NIH

Challenges Left on the Table

Challenges below were not selected by participants but are included here for post-conference consideration.

- *Risks of release of certain types of data (IRB template)*
- *Utilize mobile devices*
- *Developing standards and protocols*
- *Institutional buy-in, especially outside libraries*
- *Aligning practices with ethics of labor & research*
- *Can we create projects that manage themselves?*

REDEFINING THE CHALLENGES

Introduction

Jake Dunagan, verynice.co

We are now going to pivot a bit in preparation for tomorrow, the design day. To get there, what we want to do is flesh out our big challenges. We are going to use a modified version of Sohail Inayatullah's Futures Triangle. This is a way of mapping the different dynamics of change for the future, and I thought it would be good to use this for our challenges. It systematizes the way we look at different conditions.

It is basically a triangle that consists of three different sections or angles. One is the push of the present, things like new technologies or cultural trends that are pushing forward towards the future. Next is the weight of the past. These are things that are either holding back change or potential barriers to change. Then the pull of the future. These are images, compelling ideas, visions, values, that are pulling us toward some other kind of future.

We are going to map out the big challenges using this structure and break up into our groups. We have rewritten the challenges from our previous session and are going to have you choose which one you want to address. You will then divide into these groups and afterwards we will hear report-outs from each group.
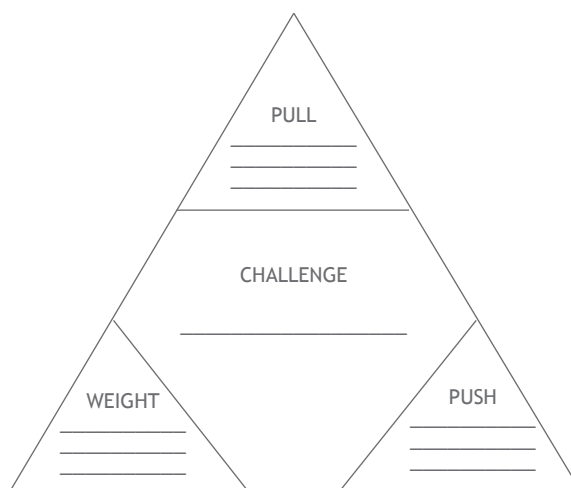
Weight:
What trends/developments/conditions (technological, demographic, cultural, economic, legal, etc.) are holding you back, in the way, barriers to overcoming your challenge?

Push:
What trends/developments/conditions (technological, demographic, cultural, economic, legal, etc) are moving in a direction that will help you overcome your challenge?
In what ways might you leverage them to overcome your challenge?

Pull:
What is the purpose for overcoming this challenge? Who are potential partners, collaborators, communities? Express a vision or description of your world when the challenge is overcome.

The Groups & Challenges

The groups below and the challenges they address carry through to subsequent design tasks.

How do we include multiple communities and stakeholders?

- Kim Christen Withey,
  Mukurtu, Washington State University
- Sara Sikes, Massachusetts Historical Society
- Sharon Leon, George Mason University
- Jon Voss, Shift/HistoryPin
- Lauren Tilton, Yale University
- Ben Vershbow, NYPL Labs,
  New York Public Library

Creating tasks, models that appeal to user communities with multiple motivations

- Meghan Ferriter,
  Smithsonian Transcription Center
- Lieke Ploeger, Open GLAM
- Ashwin Gopi, New York University
- Carsten Oesterlund, Syracuse University
- Katherine Doyle, Pulitzer Center
- Hector Mongi, University of Dodoma, Tanzania

Balance traditional and non-traditional expertise

- Nick Adams, University of California, Berkeley
- Courtney Young, Pennsylvania State University,
  American Library Association
- Jeff Bigham, Human-Computer Interaction
  Institute, Carnegie Mellon University
- Perry Collins,
  National Endowment for the Humanities
- Kirk Jalbert, Rensselaer Polytechnic Institute
- Jeremy York, HathiTrust

Skills training to become effective crowdsourcers

- Katie King, University of Washington, Seattle
- Bob Horton, Smithsonian Institution,
  National Museum of American History
- Brett Bobley, National Endowment for the
  Humanities
- Eva Caldera,
  National Endowment for the Humanities
- Darlene Cavalier, SciStarter
- Tim Olsen, Gonzaga University

Move crowdsourcing from edge to core workflow

- Daniel Powell, King's College London
  (University of Victoria)
- Victoria Van Hyning, Zooniverse, Oxford University
- Tom Blake, Boston Public Library
- Trish Rose-Sandler, Missouri Botanical Garden,
  Biodiversity Heritage Library
- Rachel Frick, Digital Public Library of America
- Jessica Zelt, Patuxent Wildlife Research Center,
  North American Bird Phenology Project
- Jeremy York, HathiTrust

Verifying data using algorithms

- Zaven Arzoumanian, Wild Me Conservation
- David Miller, National Institutes of Health
- Peter Mangiafico, Stanford University
- Alexis Rossi, Internet Archive
- Ben Brumfield,
  Collaborative Manuscript Transcription
- Ben Miller, Georgia State University
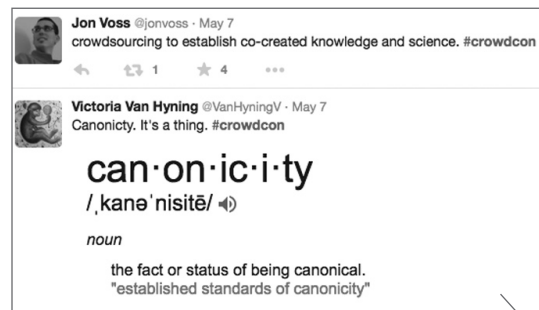- Mary Flanagan, Tiltfactor, Dartmouth College

Dealing with failure and framing it as an experiment

- Lacy Schutz, Museum of the City of New York
- Liz MacDonald, NASA, Aurorasaurus
- Austin Mast, Florida State University, iDigBio
- Michael Haley Goldman,
  United States Holocaust Museum
- Peter Carini, Dartmouth College
- Andrea Wiggins, University of Maryland

Connect education to engagement

- Amy Patterson, Wisconsin Technical College
- Chris Lintott, Zooniverse, Oxford University
- Jeremy Dean, Hypothes.is
- Jen Hammock, Encyclopedia of Life (EOL)
  & Smithsonian Institution
- Tina Phillips, Cornell Lab of Ornithology,
  Cornell University
- Neil Fraistat, MITH, University of Maryland
- Edith Law, University of Waterloo
- Jenny Preece, University of Maryland

* Names shown here are those who signed the group
  roster cards. Names of those who spontaneously
  joined groups may be missing.

Jon Voss @jonvoss · May 7
crowdsourcing to establish co-created knowledge and science. #crowdcon

Victoria Van Hyning @VanHyningV · May 7
Canonicty. It's a thing. #crowdcon

can·on·ic·i·ty
/ˌkanəˈnisitē/ ◀))

noun

the fact or status of being canonical.
"established standards of canonicity"

*Edge to Core Workflow group (left)
and Skills Training group (right)*

Group Report-Outs

Jake Dunagan, verynice.co

For these report-outs we want you to just tell us your challenge and then the three sections: the weight, the push, the pull. In this section we are trying to just get a feel for the conversation, the headine-level ideas that you're dealing with.

CHALLENGE:
MOVE CROWDSOURCING FROM EDGE TO CORE WORKFLOW

Weight

• Infrastructure: We just don't necessarily have all of the platforms available all of the time that we want to.

• Accountability.

• Canonicity.

Push

• Connectivity.

• Economic feasibility/necessity.

• Valuing diversity.

Pull

• Shared ownership of knowledge-ecosystem (that's one word).

• Distributed authority.

• Empowerment.

CHALLENGE:
SKILLS TRAINING TO BECOME EFFECTIVE CROWDSOURCERS

We quickly learned we had two different definitions of this challenge. One was the skills and training that institutions would need in order to be able to sponsor crowdsourcing projects, and then the skills and training crowdsourcers themselves would have to have in order to

be able to participate in and undertake those projects.

Weight

- Inertia: There are a whole variety of forces keeping people and institutions from participating.

- Technical capacity: Institutions that don't think they can do crowdsourcing projects; people who don't have access to broadband.

- Awareness: They just don't know what's out there on any level, or if it's there in the system.

Push

- Abundance: There is an increasing abundance of opportunities, so there are more and more crowdsourcing projects and ways to get engaged.

- Community Engagement: There is an increasing emphasis in cultural heritage organizations from the management side on promoting community engagement, which will start to challenge and balance the inertia. It is saying, "You have to change, this is a priority of the institution."

- CCLA: There are organizations like this, which we hope will help things move forward.

Pull

- Awards.

- Broadband access increasing.

- Audience ownership is an ideal.

CHALLENGE:
BALANCE TRADITIONAL AND NON-TRADITIONAL EXPERTISE

First we defined this at the individual as well as the institutional level. Expertise can be associated with either.

Weight

- Pigeon-holed: How tasks traditionally, and in the way that crowdsourcing has been defined, are pigeon-holed tasks associated with different levels of expertise.

- Antiquated technologies that have permeated from that past into the present.

- Legitimacy: Who has the supposed right to make knowledge claims and how those have also existed in the past and into the present.

Push

- Higher expectations: A future model is one in which we have much higher expectations about what collaborative science can do.

- New interfaces: We are starting to see, as evidenced from today, many of the new infrastructures and many of the new interfaces that are coming into play that will allow for multiple points of expertise.

- Community narratives: We are seeing increased value of community narratives as

Jacqueline Wernimont @profwernimont · May 7
moving into the v. important discussion of credit/acknowled when thinking about broad contributor base #digdiv2015 relevant to #crowdcon

Siobhan @SiobhanLeachman · May 7
@CitSciChat - might be of interest to crowdsourcing bods. In fact the whole twitter feed from #Crowdcon is great!

Traditional & Non-traditional Expertise group

Verifying Data group

counter-narratives to traditional expertise.

Pull

- Purpose: The purpose of this is to create new knowledge ecosystems and networks (which is a nice echo to the Workflow group).

- Partners: We have to think about different kinds of expertise—situated knowledge of amateurs as well as traditional knowledge institutions and how they can co-recognize each other.

- Vision: Eventually our vision is to create co-created knowledge.

CHALLENGE:
## VERIFYING DATA USING ALGORITHMS

Weight

- Hiring: Finding people who are qualified to do the work and how much money they cost us when we do find them.

- Tools: That might be a lack of tools or a lack of open tools, or it might be a lack of awareness of what tools are out there and available to people across disciplines.

- Problem definition: How do you define your problem? That has a lot to with what your outcomes are, so getting the problem definition right is critical.

Push

- Conversations: Having conversations like this one. This is a good example of people from different disciplines being able to have conversations about how to get things done.

- Openness: The openness that is being encouraged both in our community and by funders, so this means to open up your software, open up your data.

- Cloud Computing: The availability of cloud computing, so people can do things on a larger scale now than they could previously.

Pull

- Precision: We think that being able to do things on a larger scale will increase the precision of the outputs of our projects.

- Scale: Being able to take a small amount of human input and really explode that to a much larger scale of impact.

- Opportunities: Unknown opportunities that these types of technologies will make available to us.
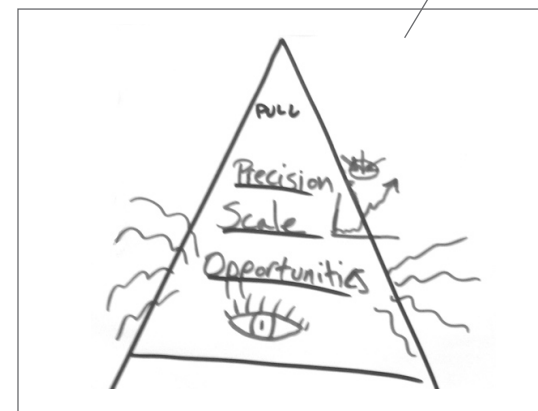
CHALLENGE:
CREATING TASKS, MODELS THAT APPEAL TO USER COMMUNITIES WITH MULTIPLE MOTIVATIONS

Weight

- Ambiguity: What are the motivations and strategies? We don't know yet.

- Engagement: How do we compete with Netflix and videogames for people's attention to participate in crowdsourcing?

- Legal aspects: Both in the launching of a project and participation in projects

Push

- Flexibility: The increasing flexibility of our platforms and design of platforms.

- Emotional engagement: People getting more and more emotionally attached to either the purpose or the topic or the community as they participate more.

- Activism: In terms of digital labor rights.

Pull

- Learning: One of the outcomes that we want to have in the future is that people are interested in learning and feel that they have individual interest. More people are curious and willing to share, and there will be STEM learning opportunity and interest.

- Ownership: Hopefully the main people who are going to be involved in this are going to be participants who take ownership of the projects and platforms they are participating in.

- Knowledge production: Scalable participation in the knowledge-production process. How can we grow engagement up to a point where even if there is a huge community people stay engaged?

CHALLENGE:
DEALING WITH FAILURE AND FRAMING IT AS AN EXPERIMENT

Weight

- Stigma: The stigma and punishment for failing, all through our entire lives in every single way.

- Measurement/benchmarks: The lack of usable measurements and benchmarks—we don't have much basis for comparison.

- Blinded by sexiness: We are all blinded by the sexiness, which basically translates to unrealistic expectations.

Push

- Visibility: The increasing visibility of everything that's happening in these spaces. It's much easier to see successes and failures these days.

- Celebrate insights: Another push would be celebrating insights and the development of the peer communities around organizing these kinds of

Victoria Van Hyning @VanHyningV · May 7
For those who have a moment at #crowdcon pls check out/RT an opportunity to try DIY crowdsourcing w/ me in Oxford! bit.ly/1K8r1XZ

5   1

http://bit.ly/1K8r1XZ

Multiple Motivations group

projects, like the Citizen Science Association, the Crowd Consortium, etc.

Pull

- Reframing: Reframing failure as progress toward a goal and not an end in itself.
- Best practices: Which we are doing here.
- Failing differently: Quit doing the same things that we keep failing at in the same ways and fail in different ways.

*Failure as an Experiment group (above);
Connect Education to Engagement group (far right)*

CHALLENGE:
**CONNECT EDUCATION
TO ENGAGEMENT**

We struggled at the beginning about whether we are talking about formal education or informal education and decided to talk about both as much as possible.

Weight

- Belief/motivation: This is expressed nicely in the literature with the phrase, "threshold fear." Do people feel they can have the skills to participate in these kinds of projects?
- Access/literacy: This captures a lot of different things, from digital literacy to language barriers, juggling URL issues, so it is sort of a catch-all and a lot of important things go here.
- Professional/curricular structures: Specifically in formal education, is there room in the classroom to incorporate these kinds of projects? Are the teachers skilled or aware enough to integrate them into their curriculum?

Push

- Curriculum: Curriculum also appears here because obviously there is a lot of push for digital humanities projects and all kinds of projects involving technology.

- Digital/community engagement: That's a broad push that includes big data from the White House. And there are other kinds of community engagement projects. We've got something going for us in that there is kind of a turn going on here.

- Gaming/credentialing: Gaming and credentialing are powerful motivating factors for getting people involved in these kinds of projects from an educational standpoint.

Pull

The pull category was hard to boil down into phrases and words.

- Citizen-scholarship: There are people who are really passionate and also skilled in humanities and sciences who are disconnected from academic research projects. The kind of projects that we're talking about and processes that we are talking about can connect them to these projects and make them intimately involved.

- Mobile/social technology: One thing we are thinking of as a specific example here is geolocation technologies and the way that our new tools allow us to participate in stuff outside the classroom as we wander around this crazy world.

- "Slipping in learning": This is a phrase we borrowed from somebody on one of the panels. The idea is student-centered pedagogies where students are driving their own learning, but also where ordinary citizens

are driving their own learning through places where there are lines of inquiry about to be pursued.

CHALLENGE:
HOW DO WE INCLUDE MULTIPLE COMMUNITIES AND STAKEHOLDERS?

Weight

- Identification: One weight we focused on was actually identifying those communities and outreach to stakeholders.

- Trust: Establishing and maintaining trust with those communities.

- Access: By this we mean technical and platform access for those communities to actually participate.

Push

- Open access: The growing availability of access to materials and software.

- Shared authority: There is a growing comfort level with shared authority from the institutional perspective but also from a community perspective.

- Mobile: The increasing ubiquity of mobile computing, which helps mitigate that question about access to technical platforms and those sorts of things for participants.
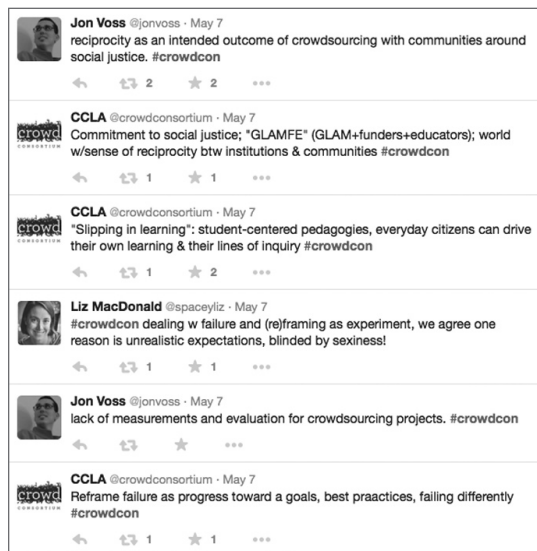
Pull

- Social Justice: We decided that our purpose in doing this work is a commitment to social justice

Jennifer Serventi @JenServenti · May 7
#crowdcon I like this new group from @sleonchnm and colleagues- GLAMFE (Galleries, Libraries, Archives, Museums, Funders, Educators)
2   3

Andrea Wiggins @AndreaWiggins · May 7
I'm still voting for GLAM STEAM, maybe a bit broader than GLAMFE? #crowdcon
1   2

Multiple Communities and Stakeholders group

Amanda Visconti @Literature_Geek · May 7
Really impressed w/ #crowdcon event. Lots of thinking around labor, attribution, empowerment issues; real volunteers instead of "the crowd".
↩   ♺ 1   ★ 3   •••

critical play @criticalplay · May 7
.@AndreaWiggins re: GLAM STEAM, how about "ALMAGEST"? i.e. resurrect the knowledge of the ancients #zomg #crowdcon

Jon Voss @jonvoss · May 7
reciprocity as an intended outcome of crowdsourcing with communities around social justice. #crowdcon
↩   ♺ 2   ★ 2   •••

CCLA @crowdconsortium · May 7
Commitment to social justice; "GLAMFE" (GLAM+funders+educators); world w/sense of reciprocity btw institutions & communities #crowdcon
↩   ♺ 1   ★ 1   •••

CCLA @crowdconsortium · May 7
"Slipping in learning": student-centered pedagogies, everyday citizens can drive their own learning & their lines of inquiry #crowdcon
↩   ♺ 1   ★ 2   •••

Liz MacDonald @spaceyliz · May 7
#crowdcon dealing w failure and (re)framing as experiment, we agree one reason is unrealistic expectations, blinded by sexiness!
↩   ♺ 1   ★ 1   •••

Jon Voss @jonvoss · May 7
lack of measurements and evaluation for crowdsourcing projects. #crowdcon
↩   ♺   ★   •••

CCLA @crowdconsortium · May 7
Reframe failure as progress toward a goals, best praactices, failing differently #crowdcon
↩   ♺ 1   ★ 1   •••

- GLAMFE: Who are our potential partners and collaborators in this work? We have coined a new term, "GLAMFE," which is GLAM plus funders and educators.

- Reciprocity: The vision of the world that we would like to see when we overcome these obstacles is a world where there is a sense of reciprocity amongst these communities and the institutions that they work with.

Conclusion & Next Steps

Jake Dunagan, verynice.co

I don't know if reciprocity was mentioned before, but that is a great word for this conversation to end on. Give yourselves a hand for this session and for the whole day. It has been fantastic. We have laid the groundwork, we have mapped a wide swath of territory and great concepts are in play now, so I think we are set up very well for tomorrow and your prototype implementation exercises. Come tomorrow with a lot of energy because you are going to be making and thinking and putting things together and actively working. We are moving from the futurist to the designer, from surveying the future to thinking about and implementing designs, so I will pass this over to Matt.

Matthew Manos, verynice.co

Jake did a good job of introducing the goals for tomorrow. One thing to keep in mind is that we are going to keep you in these groups, so you are going to be further exploring these themes, especially because they are ones you've gravitated towards naturally already. As Jake mentioned, it is going to be a day of making and attempting to create solutions around some of these emerging issues that we've been discussing. We are going to do that in a few different ways, through further building out what the personas of this audience might be and what those people might actually want. We are going to do something pretty exciting called a "project canvas," which is a unique iteration on a business model canvas we created specifically for this event. It is supposed to help you create a road map for action plans for this product or service or whatever comes out of this. We are going to end the day tomorrow with pitches. It doesn't have to be as corny as *Shark Tank*, but we are referencing *Shark Tank*.

III. Implementing Insights

ROUNDTABLE: BEST PRACTICES

Introduction

Moderator: Andrea Wiggins, University of Maryland

We have a great panel of all-stars and experienced folks from the field. Each person has been prompted to prepare their thoughts on three best practices. They will give just a short overview of those and then we will have a little more discussion with the panelists and open it up to the room to pitch in ideas and have some discussion.

OpenGLAM

http://openglam.org

Lieke Ploeger, OpenGLAM

I am with the Open Knowledge Foundation, which is an organization working on promoting open data in a variety of fields. I work for one of the working groups for Open Knowledge which is called OpenGLAM, which is what I will be talking about mainly for best practices. Open Knowledge set up these working groups as a way of enabling communities in various fields around open data to work together and contribute to the specific field of open data further.

I work for the OpenGLAM working group as Community Manager. OpenGLAM is working on promoting free and open access to cultural data. GLAM, as you know, stands for galleries, libraries, archives and museums. "Open" we define as Open Definition-based. Open Definition says that everything is open which has a license conformant with the Open Definition's principles, for example CC0 (public domain) or Creative Commons CC-BY or CC-BY-SA. These licenses state that you should attribute the organization that the content is coming from or "share-alike," share it in the same way.

We run a big community around open cultural data through the OpenGLAM working group. We started this in around 2011-2012 and it has grown into quite a large, active community. It is a topic that many people feel very passionate about and we have established a structure we feel is quite successful in running this working group in an efficient way. It has been running now for about three years and growing consistently and also spreading out to various countries where there are now OpenGLAM working groups as well, as a way we have of spreading out.

Our four channels are a mailing list, where we share discussion around OpenGLAM issues and ask for feedback on specific things we are


Roundtable Participants

Moderator:

- Andrea Wiggins, University of Maryland

Panelists:

- Ben Brumfield,
  Collaborative Manuscript Transcription
- Chris Lintott, Zooniverse,
  Oxford University
- Tina Phillips, Cornell Lab of Ornithology,
  Cornell University
- Lieke Ploeger, OpenGLAM
- Kim Christen Withey, Mukurtu,
  Washington State University

Moderator Andrea Wiggins
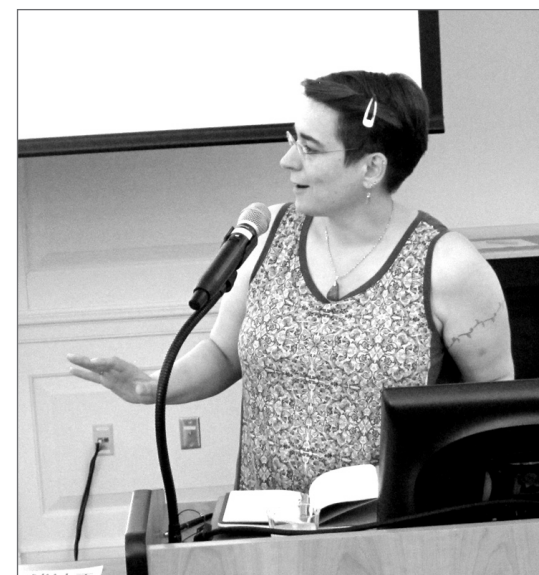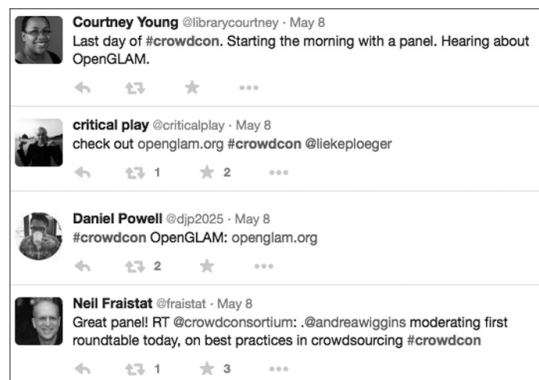
From left: Lieke Ploeger, Ben Brumfield, Chris Lintott

working on from the OpenGLAM community. We have quite an active Twitter account. Then we run a smaller, more focused core working group of 17 people who are the most active members of the OpenGLAM community. They meet every month and are all from various countries and they are OpenGLAM ambassadors in their countries. They share every month what is going on, we share thoughts, we plan events, we discuss how to take OpenGLAM forward. Then we also have an Advisory Board. This consists of high-level thought leaders in the field such as people from DPLA, Europeana, and the Internet Archive. We ask these people for feedback on specific key strategic issues moving forward with OpenGLAM.

One of the interesting developments in recent years is that we have established several local groups of OpenGLAM. We now have local groups in four countries, which are more specific, local communities focused on OpenGLAM in their own country. We also work on specific tasks for that country, so there could be different issues in different countries and they do this in their local language, which is probably more of a European thing. We found it works much better than having everything always in English. These local groups also report back to the main OpenGLAM working group and we co-organize events with them.

We also started up one of the most recent local groups as a crowdsourced effort. What we did was have a meetup at the Open Knowledge Festival last year and invited everybody

interested in starting up such a local group in Germany for people involved in German cultural heritage institutions. We invited everyone to discuss in an open manner why they think such a group should even be started, what kind of tasks they should focus on, and what kind of structure they would like. They then defined several key areas that they wanted to work on, and on the basis of that they started up a group and started organizing their first event for next year focused on outreach to smaller institutions in the GLAM sector.

We have also been working on organizing sprints. Sometimes we have some things we want to do as a GLAM community. For example, in the beginning we wanted to define some key principles that we think every institution should follow to really be an OpenGLAM institution, and we co-wrote these principles together, asking input from mailing lists first of all, and then getting a core group of people together to write up these principles in more detail. Of course, we asked for feedback from our Advisory Board before really putting these principles out there and making them final.

Recently we've been working on updates for our Open Collections page. We list a number of collections out there with open content and we implemented some new software for this recently. We asked our community to help us first of all in defining exactly what the page should look like, how the software would be best usable for everyone, and then to help us update the metadata on each collection or

to put extra metadata up there through tags. First we asked the mailing list who would like to help us out, and then we organized some short sprints with key participants to update this and go through it together. So we always try and involve our users throughout the process in whatever we do with OpenGLAM.

I would also like to mention that as one of our core principles we always like to ask institutions if they involve their audience in new ways. Of course, one of the recommendations is that they try to pursue crowdsourcing efforts where possible.

Collaborative Manuscript Transcription

http://manuscripttranscription.blogspot.com
http://beta.fromthepage.com

Ben Brumfield,
Collaborative Manuscript Transcription

I'd like to draw in pretty deeply on one best practice. The best practice I would like to talk about today is turning the product of the crowdsourcing effort back to the volunteer. Now what do I mean by "product"? I'm not talking about the final product, the item-level finding aids, the published research papers, and the scholarly press. I'm talking about the raw product, the actual contribution of individual volunteers and their fellow volunteers: the corrected text of a newspaper article, the transcribed letter, the comments, the identifications that they have made personally or that
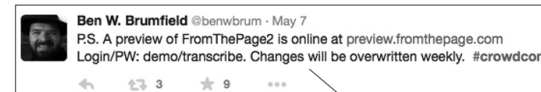
people within their community have made.

Why should we do this? The first reason is, it's the right thing to do. Yesterday we heard a lot about reciprocity, about social justice, but this is a pretty old concept. Deuteronomy says, "Thou shalt not muzzle the ox when he treadeth out the grain." Many, many platforms for transcription support this immediately. If you're using a wiki-like system like FromThePage, like WikiSource, and I believe Scripto, that's just a side effect of the platform. With others, users can't get at any of their contributions. I would like to call out the Smithsonian Institution Transcription Center. They add the ability for the users to download PDFs of all of their contributions and all of the community contributions. They added this feature on purpose to their tools in order to support this because they believed it was the right thing to do.
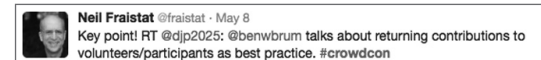
Now that we're done quoting the Bible, what are the instrumental reasons to do this? For one thing, if you expose the raw data early on we can align our projects with the things that incentivize our volunteers. Our volunteers often do not care about our institutions. They are not passionate about clean metadata in the catalog system, that's not what motivates them. The reason they are doing this is because they are immersing themselves in the subject matter. They are sitting with a bird watcher in 1918. They are marching alongside a soldier in the Civil War. Exposing the things that have immersed them allows them to share

For a slightly different version of Ben Brumfield's presentation, see his blog:

http://manuscripttranscription.blogspot.com/2015/05/best-practices-at-engaging-public-at.html

Ben W. Brumfield @benwbrum · May 7
P.S. A preview of FromThePage2 is online at preview.fromthepage.com Login/PW: demo/transcribe. Changes will be overwritten weekly. #crowdcon
↩ ⟲ 3 ★ 9 •••

http://preview.fromthepage.com

Neil Fraistat @fraistat · May 8
Key point! RT @djp2025: @benwbrum talks about returning contributions to volunteers/participants as best practice. #crowdcon

this and show others what excites them about it.

On one of the first projects that I worked on we got a super-volunteer early on who just blazed through all of our material. Afterwards I talked with him on the phone and asked, "What can we do to thank you for this?" He said the thing that he valued most, the thing that would be most important, would be if he could print out and publish and have a bound copy of the Mexican-American War diary that he had transcribed because the way that the heritage organization to which he belongs rewards and advances its members requires those members to publish a book. Well, the thing that he had done matched the contributions of any of his fellow amateur members of the Sons of the Republic of Texas, but without actually delivering that to him in a useful format, he wasn't going to be able to get that recognition. So there is this idea of that extrinsic reward.

Another reason to expose the work in process, these raw contributions, is to enhance recruitment. I have often told the story of one of the super-volunteers on a number of projects on FromThePage who found the site by doing a vanity search. He Googled his own name, and the top response turned out to be all of the entries in the Julia Brumfield Diaries which mentioned a man who had the same name he did. It was the diarist's postman. He recognized that this was his great-uncle, jumped in, transcribed an entire diary on his own, and then moved on to use his previous experience

analyzing ichthyology records and field books to transcribe scientific field books on the same platform.

If we had not exposed the work in progress from a previous volunteer, he would not have gotten there. It's also possible that if we had exposed it in some other format, some official site, he still wouldn't have got there. He found the transcript within the crowdsourcing project and he knew how to contribute immediately.

The last reason I'm advocating this is for engagement and productivity. For the last three years I have been involved in a nonprofit in the UK called Free UK Genealogy. For 15 years now they have had volunteers from around the world who transcribe genealogical records of interest—census records, the civil registers of births, marriages and deaths—all using offline tools including spreadsheets, putting CD ROMs in the Royal Post, and getting things online to an online database that they can publish.

I was brought in to revise their transcription system, overhaul it completely and bring it into the modern century. And I encountered incredible resistance. The volunteers had a system that worked for them. Why change it? What was this going to give them? After a few months we switched course and we focused instead on the delivery system, a searchable website that people can go to and see. So we replaced the oldest of the systems, one called FreeREG, with a new one we built completely from scratch. What happened is that once we launched that, once we replaced the old

delivery system, the old product of the volunteer effort, with a new one, for one thing it was the most positive project launch I have been involved with in two decades of software development. The people loved it. For another thing, it re-engaged the volunteers. Since we've gone live with this we have seen contributions using the old spreadsheet-based online system go up. We just passed 32 million records.

In conclusion, expose volunteer efforts within your crowdsourcing system as they're produced. It's the right thing to do.

Zooniverse

https://www.zooniverse.org

Chris Lintott, Zooniverse, Oxford University

Thank you for the chance to speak. I think I'm here because we've run 44 Zooniverse projects, so you'd think I'd have some sense of what best practice is, but the truth is we are still in a phase where I can point you at stuff that doesn't work and tell you what we are trying to do to fix that. We are not at a point where I can tell you how to build your definitive crowdsourcing project, which is fine, and exciting.

But there are some things in common across all 44 projects, so let's see what we can get out of figuring out what's in common. The first thing in common is that every project we've ever launched has run in two phases. There's an enormous spike of excitement in interest and effort at the beginning, and then that drops back down and we have a long-term, stable set of volunteers that carry on. People come and go, but there are these two distinct phases.

So the first best practice is an obvious one, which is that you have got to design for both of these phases. That means you need a system that scales, that can stand up to the traffic at the beginning, but you also need something that is going to engage people for the long term. If the cardinal rule of crowdsourcing is not to waste anyone's time, you need to be doing that when you're busy and when things are quiet as well.

It's worth thinking about the fact that the audience changes over time as well. At the beginning, literally no one is invested in your project. I know there are communities we think we are targeting with these projects. We thought we were targeting our first project, Galaxy Zoo, at amateur astronomers. But that kind of astronomy is different from an academic understanding of astrophysics, and so even amateur astronomers coming into Galaxy Zoo are neutral when they start regarding whether they care about this project. So you need to be able to give experience to people who haven't yet decided whether they want to learn this stuff. I think we often lose sight of that. We need to make sure that we convince the people who are coming in before we do anything else.

That is really difficult because those of us in

Trevor Owens @tjowens · May 8
.@benwbrum's focus on motivations of volunteers & the value of exposing their work to them in forms that matter to them is great #crowdcon

Amy Patterson @amycep · May 8
#crowdcon @chrislintott: design for both phases of crowdsourcing--scale for early phases of excitement + keep engagement when calms down

Victoria Van Hyning @VanHyningV · May 8
Golden rule of #crowdsourcing: don't waste anyone's time. @chrislintott #crowdcon

Neil Fraistat @fraistat · May 8
@benwbrum: Expose volunteer contributions to your crowdsourced projects as they are produced. #crowdcon

Matthew Manos @verynicetweets · May 8
Panel highlights the importance of exposing volunteer contributions throughout the crowdsourcing process #crowdcon

galleries and archives and libraries and muse- ums sit in a reserved and specialized space. There is this concept I fell in love with from the museum world in the museum literature called "threshold fear," which is the idea that you don't go into a gallery unless you think you can get something out of looking at a paint- ing. No matter what that gallery does inside its building to broaden its audience or to provide engagement, it won't work unless you can get people up the steps.

And we have that in our projects. People browse onto our Zooniverse projects and as soon as they realize that something is real they have a visceral reaction to the fact that they are doing science, which they were terrible at in school. My guess is that this applies as much to the humanities as it does to the sciences because the kind of reading that you're asking people to do or the kind of engagement that you're asking people to have with the text is not natural to people, and it is something they haven't grown up thinking of themselves as be- ing able to do.

Dealing with threshold fear is important, and you have these two kinds of people. You have the dabblers, where you have to overcome that fear, and the people who are on board and committed to the subject. Both of those need to have a transformative experience. If you don't have a long tail, if you don't have these people who are going to stick around for months and work for the bulk of our effort, you need something that is going to transform

people, the cautious dabblers who haven't gone over that threshold yet, into committed people. And I think we know what that is.

My second best practice is that what you do is give people a real experience that they believe in. That means convincing them that they're going to do something useful. That they have done something useful is better. It's much bet- ter to turn around and say, "Look, you did this and 37 other people agree with you." Or, "You transcribed this text and now I know this." It is showing both the large-scale results—people like you have helped us discover planets or find text—but also, "You mentioned this place in this text; there are 33 other people who have mentioned that place and look, here is how your point connects to theirs." It is critical in transforming people from dabblers to commit- ted volunteers.

I think that means that you end up building in two modes as well. You build a microtask that is immediately convincing and easy to grasp, and then you allow free exploration and discus- sion. You need both. I don't think you can build a project that is just being done with things like comments on Flickr Commons. It's not good enough to just say to people, "Come en- gage with our collection," because they don't believe that they can.

I have this hatred of the fashion in museums and galleries where, when you get to the end of an exhibition which has been beautifully curated, there are little bits of paper with the question, "What did you think?" And you

Victoria Van Hyning @VanHyningV · May 8
Some nice echoes between @benwbrum and @chrislintott, both encouraging us to give real feedback and design for genuine engagement. #crowdcon

Matthew Manos @verynicetweets · May 8
.@the_zooniverse reminds us that it is critical to illustrate the impact each "cog" in the crowdsourcing "machine" makes. #crowdcon

Courtney Young @librarycourtney · May 8
Crowdsourcing: Important to give people an experience to believe in. #crowdcon

Amy Patterson @amycep · May 8
.@chrislintott: Give people a real experience they can believe in, showing large-scale results and connections to transform #crowdcon

Dr Meghan Ferriter @MeghaninMotion · May 8
YES! @chrislintott reflects on @zooniverse experiences: must design flexible/scalable space for new AND continuing participation #crowdcon

Amy Patterson @amycep · May 8
Threshold fear: you don't go into a gallery unless you think you can get something out of it @chrislintott #crowdcon

pin it up and in some way you are supposed to engage with the process of curation. That feels like tokenism, and we keep building (at Zooniverse as well as everywhere else) the digital versions of that in which we say to people, "No, your comment is fine, do anything with it." We need to do better than that. We need the microtask to convince and then you need the discussion and exploration.

That leads to my third and final best practice, which is that I don't think we should be planning for our platforms and our projects to convince people of content. People shouldn't go to Galaxy Zoo to learn about galaxies, they shouldn't go to Operation War Diary to learn about the First World War. On those sites people are constantly engaged in learning as they go from a simple interaction to more complex interactions. They are learning how to use your features, how to use your discussion forum, how the community behaves, how to deal with the results that you're giving them, how to interpret what they're seeing.

What these projects do, the good ones, is they act as engines of motivation. We can show you that good projects convince people to go out and learn more elsewhere, to act much more like the engaged researchers that we want them to become. So you should not try to provide the platform for the full-scale engaged research experience. What we need is a tight, focused, believable, real, authentic experience that inspires people to go out and explore for themselves.

Cornell Lab of Ornithology

http://www.birds.cornell.edu

Tina Phillips,
Cornell Lab of Ornithology, Cornell University

The Cornell Lab of Ornithology is probably best known for creating and implementing the traditional form of citizen science where people go and collect lots of data out in the field and share it with scientists who are interested in answering some questions. I have been there for about 16 years and in that time I have been a practitioner running projects, I have been an evaluator, and I have been a researcher. I am going to give you a best practice from each of those different hats that I wear.

The first is from my researcher hat. A best practice would be to really know and understand your audience. As I said we have done mostly the traditional form of citizen science, but in our one real effort to do crowdsourcing in the online digital mode we assumed we knew something about our audience. We said, we are going to put all of these images of birds on the website and we are going to get gamers because gamers love to do this kind of stuff, right? Well, we were really wrong. We didn't get any gamers, we got people who knew about the Lab, and the aspects of what we thought we knew about that were not really fitting well with the design of the project.

We should have really taken the time to understand their needs, their wants, and what their expectations for the project were. That

Chris Lintott and Tina Phillips
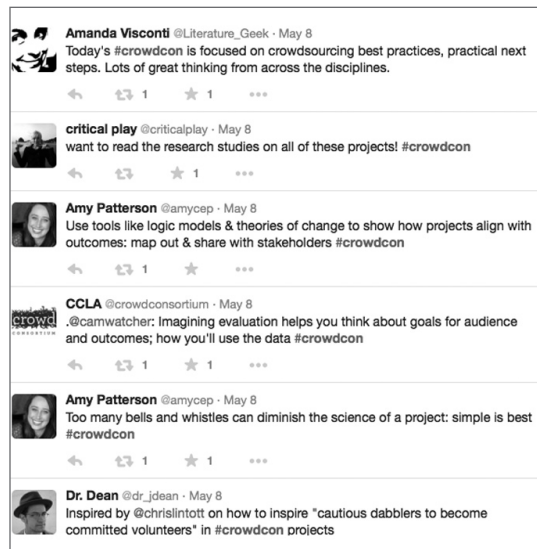
project, which was called CamClickr, was in some ways a failure right from the get-go because we just didn't understand the audience. Doing that will obviously help keep your audience retained over the long term, and we weren't able to keep that retention after the initial spike that Chris talked about. In contrast to that, eBird is a project that the Lab is successful with and that is because we really understand our audience. We have people who are project leaders, who have face-to-face contact with a lot of the people who make up that group. So there is a vast difference in comparison of how successful a project can be just by understanding the audience.

For my practitioner hat I am going to use the CamClickr example again. We over-designed that project. We had a wonderful designer who was very good at what she did, but we added all sorts of bells and whistles and things that actually made the navigation of the project very clunky. We built it in JAVA as opposed to having it on a web-based platform, so it was not simple, it was over-designed. That is a detriment.

There has to be a point at which practitioners say to designers, "Okay, we've done enough here. We don't need to add any more bells and whistles." Because I think the end result of that is to actually diminish the science. On the one hand we were trying to make a very scientific project, but on the other hand we were trying to make it a game, and we had leaderboards which people did not like at all, we had

sound and video effects in the middle of trying to play this game—it was just overly designed. So I would encourage people: Simple is best.

For the last point I will wear my evaluator hat. Even if you don't have the funds to do an evaluation, I think it's a really important process to imagine or pretend that you are going to evaluate your projects. What that does is make you really think about the plan. It makes you immerse yourself in thinking about what your intended outcomes are for the program and for the participants. You can use simple tools like logic models or theory of change to put on a graphical representation of what you're providing the audience and the activities that they're doing, and the things that you'll get out of it, and how you think all of those align with outcomes.

We often develop projects with these preconceived notions of how they will need to be. Until you actually map it out and share it with other stakeholders, other people who are involved, and realize that you have a lot of assumptions about how your project will achieve those outcomes, it is really easy to carry on this process of developing something without having real alignment between those goals and what the people are actually doing.

So even if you don't have the money to do an evaluation, just make pretend. Pretend that you actually are going to evaluate it so that you can start delving into these planning tools early in your process. And there are lots and lots of resources out there to help people think

about how to build their project with that evaluative framework in mind.

Mukurtu

http://www.mukurtu.org

Kim Christen Withey,
Mukurtu, Washington State University

I'm going to talk about three best practices: ethical exchange, collaborative curation, and sustained support. I like alliterations obviously. These grow out of my role as project director of Mukurtu, which is a free and open source content management system built with the needs of indigenous communities in mind, but I think these are just three best practices in general. That was the charge so I'm sticking to it.

The first is ethical exchange. Whatever your content is—collections, metadata, whatever it is—the sharing should be based on a clear set of understandings of the stakes of all of the parties involved in the project. And these should be spelled out, not assumed. We make a lot of assumptions. There should be ample time set aside at the beginning to define what the stakeholder concerns are, from the very practical—where will the content live, what will happen to it afterwards, ownership versus stewardship—to the cultural. How does the content express community values, or how are community values embedded? We shouldn't default openness, it should not be assumed that this is what we want at the end. Or are there

social mechanisms in place already for how this content accumulates knowledge, etc.? So we have the practical, we have the cultural, and now we have the process.

So how do we think about the process in these sorts of ethical ways? We have to create workflows that honor local systems of knowledge exchange. I would suggest that workflows are really deeply ethical practices because they ask us all to engage with information, materials, and knowledge in deeply personal ways. This is not just a flowchart. These are things from the beginning: the name, the creator, how things move around, where they are shared, how we talk about them. So the workflow is something that should also be examined from an ethical standpoint, not just technical.

The second best practice is collaborative curation. I see curation as sets of practices that are formed around the act of reciprocity, this notion of giving and receiving. Here, in this case, it allows us to not just think about bringing people together, the collaborative part, but to also understand how many stakeholders can work together without expectations of expertise weighing down the process. That is really key, when we bring stakeholders together in a collaboration and try to level the playing field in one sense, to just look at what everybody brings to the table without that notion or expectation of expertise. So the goal here is giving and receiving.

The second part of that is, if curation is generally understood within archives and libraries as

Kim Christen Withey
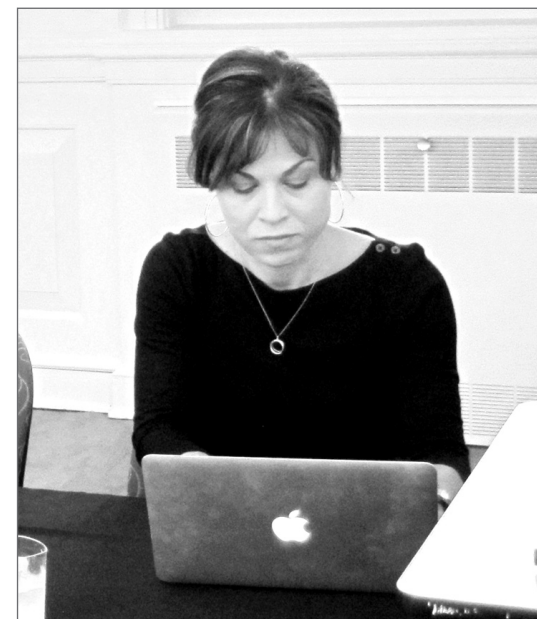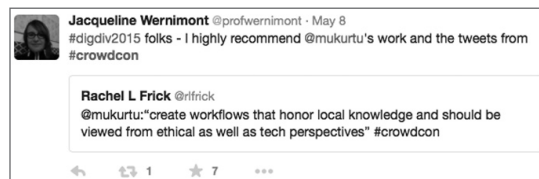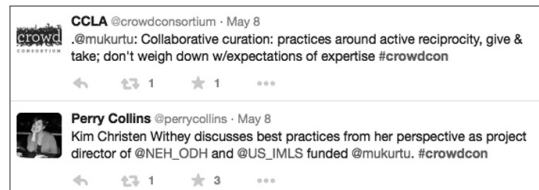
adding value to collections, I see collaborative curation as providing locally-based context in really meaningful ways. That meaning can be made historically, linguistically, socially and culturally, all of which then enrich the records. Importantly, they also enrich the communities by facilitating those connections with collections in embodied and tactile ways. So actually engaging with the collections, being with them, just being in the space, is really an important process. Collaborative curation at its best is a material cultural practice that brings content to life.

The third best practice is sustained support. This is really about trust building. Support is both establishing and maintaining relationships of respect. Respect is the baseline: respect for diverse sets of values and understandings about knowledge circulation, management

and preservation; respecting both the how and the why of a community's value system as it relates to knowledge production. So how do communities themselves see knowledge production? Why is it important? What is actually happening? Set aside a lot of time to get at this in a really ethnographic micro-sense.

These issues of respect, then, will provide the foundation for support. Not just technical support, and I'm not going to dive into technical support because I think we all know about that. That is a critical factor of course. But just as fundamental, and sometimes overlooked in digital projects, is ongoing community engagement as support. It extends the tool, the project, the data, and loops back into maintaining those relationships in sometimes unexpected ways. What do people want to do with that data once the project is over? "Oh, now we are going to engage school kids in a classroom." So how can we support extending the project? That is a support that only happens if you think about establishing and maintaining relationships for the long term.

All of these three for me come back to the mantra that we are first and foremost creating relationships, not records.

Questions, Answers, Discussion

GLEANING THE BEST FROM WORST PRACTICES

• One of the things that was brought out,

PANEL RECAP

Andrea Wiggins, University of Maryland

So this is a perfect example of the value of the crowd. I could not have come up with this myself. That was excellent, folks. What I heard from Lieke was that democratizing and structuring interactions was really valuable in the community and that respecting the local needs, especially around language and across cultural contexts was really important for them. Ben focused on returning the product and the nuances and values of that for various purposes. Chris gave us several very nice nuggets: designing for multiple phases of the project, designing for multiple roles of individuals, and not focusing on content as much as the engagement. Tina told us about understanding audiences, being wary about over-designing and focusing on simplicity and accomplishing the main task at hand, and thinking about evaluation as a way to drive things forward even if it's not a formalized process. And Kim gave us some thoughts on ethics, the collaborative nature of curation in communities, and the nature of sustained support not just in the sense of financial sustainability but in a community sustainability sense. That is something that often gets overlooked, so I'm glad you were able to raise it.

by Chris in particular, was that we know what doesn't work more than we know what works, and we are trying things out. We had a chat about this in several side conversations between last night and this morning. The sense is that the term "best practice" as we are using it in this space at the moment is a reaction to a worst practice. It is the experiment to solve a problem. I was hoping we could go through each of our panelists and get a sense of whether they have a worst practice from which some of their best practices have evolved. • Andrea Wiggins, University of Maryland

Cultural Values and Reworking Content Management Standardization

• Maybe I don't understand the question, but I'll just go for it. I don't consider myself someone who really does crowdsourcing, so when I got the invitation I wasn't sure you wrote to the right person. Now I see that it makes a lot of sense. One of the things that we did really specifically within Mukurtu when building a content management system was build off of the reason that content management systems did not work for the communities that we worked in. They didn't work for very specific reasons, and one of those reasons is the standardization.

I know everyone cares about standardization and I've bought into the notion that standardization is really important. Standardization is also a value, a value that we

hold in a very particular cultural way. And that's fine, and we can think about that, but those standardized fields are also oppressive in certain situations. They deny agency to the communities for whom these are not records, these are memories, these are communities, these are our lives, this is our knowledge, this is how we get from A to B, this is our language, those types of things.

So it is really recreating those fields and seeing what they do. I am not saying it is a worst practice, but it was understanding the importance of the data and standardization and at the same time seeing the standardization and metadata that we have in traditional content management systems needed to be reworked. So we have two different sets of records that happen simultaneously as well as allowing for tribal knowledge, cultural narratives, all to be embedded in one record. That happens at the level of text, audio and video. A song might be the best way to express cultural knowledge about a piece of content. • Kim Christen Withey, Mukurtu, Washington State University

Tapping Outside Expertise to Address User Expectations & Satisfaction

• I'll use the same project I did before as an example. With the CamClickr project, one of the things that we did was really focus solely on the scientific outcome that we wanted to achieve. We had eight million images of birds

Neil Fraistat @fraistat · May 8
Standardization is also a value, but traditional standardized fields can also be oppressive in terms of community needs. #crowdcon

Amy Patterson @amycep · May 8
.@AndreaWiggins: What are worst practices that best practices have evolved from? We know what doesn't work more than what works! #crowdcon

critical play @criticalplay · May 8
what are our #worstpractices? #crowdcon

CCLA @crowdconsortium · May 8
.@mukurtu: How do we think about crowdsourcing process in ethical ways? Honor local systems of knowledge exchange #crowdcon

Pamela Wright @PamelaSW · May 8
#crowdcon ethical exchange. Define what the stakeholder concerns are. Don't assume, spell it out. Practical, cultural, process.

Victoria Van Hyning @VanHyningV · May 8
Tina Phillips @camwatcher reflects of the failures of gamification when it isn't aligned with the audience who shows up. Simplify! #crowdcon

Dr. Dean @dr_jdean · May 8
Powerful reminder of the real goal here at #crowdcon: says @mukurtu, "we are first and foremost creating relationships not records"
2   2

Rachel L Frick @rlfrick · May 8
"what does the user get out of this tagging project, besides carpal tunnel?" Crowdsource participants deserve more. #crowdcon
2   5

Daniel Powell @djp2025 · May 8
Bad practice: not thinking through how knowledge work is not external to participants, but active part of culture/society/life. #crowdcon
4   4

Courtney Young @librarycourtney · May 8
The right tool for the right job. #crowdcon
1

Jacqueline Wernimont @profwernimont · May 8
@djp2025 @mukurtu -- is crowdsourcing the same as community-based and community-servicing approach of Mukurtu? #crowdcon
1

Pamela Wright @PamelaSW · May 8
#crowdcon successful projects need community. If people behind the project aren't engaged at the top, it won't be sustained.
3   3

Daniel Powell @djp2025 · May 8
#crowdcon Bad practice: crowdsourcing projects will not succeed without active community engagement from all participants.
1   1

Trevor Owens @tjowens · May 8
Great point on involving social scientists & educational researchers who can explore/think through the socio-cultural issues #crowdcon

and wanted help tagging those images. What we didn't account for was what the user was going to get out of this other than carpal tunnel. We actually did get emails from our poor users and they were having a terrible time. In hindsight, were we to do that again, we didn't necessarily have the expertise in-house to think about what those participant expectations or outcomes might be. One of the things we do now is work with social sciences quite a bit. We engage a social scientist in our work, especially for those projects that are led by a natural science person who really doesn't feel comfortable in the realm of working with people or understand that their time is precious and they deserve to get something out of this as well.

That's what I would say: Even if you don't have the expertise yourself, try to find an educator, a social scientist, an evaluator, who can help you navigate the water between the sciences and the humans.

I think in this group that probably isn't the case because we have a lot of digital humanities people here, but in the world that I live in that is not often the case.

• Tina Phillips, Cornell Lab of Ornithology, Cornell University

Bottom Line = Community: Engaged Researchers, Top-Down, Starting with a Problem You Want the Crowd to Solve

• One thing we know doesn't work is a project without community. We tried a cancer research project called Cell Slider and were worried because of issues of sensitive discussions and topics, plus the images, stains of pathological samples that all looked like something from standard cell biology textbooks (apologies to any cell biologists in the room). We thought, what are people going to talk about? That project is successful, but it has a huge turnover of people because there's no community.

We also know that if the scientists or the researchers don't engage at the beginning, the community dies. We have data that shows that. If the people behind the work aren't visible in the community within the first week or two, then the community will evaporate.

So engagement is important and that means, I think, that the big lesson we've learned is that if you want to develop a crowdsourcing project, top-down doesn't work. We have found this repeatedly. We keep having people who say they would love to do a project. For example, I really think planetary geology is the way to go, so we're going to find some planetary geologists. And I'm good about getting people in the room and getting enthusiastic about an idea and getting them to sign up. But when it comes to engagement with the community six months later when you've built the project, they're not there, and those projects really struggle. So you need to have that authentic engagement that comes from having a problem that you

want the crowd to solve, rather than starting with a design to do something with a crowd and then go searching for a problem. • Chris Lintott, Zooniverse, Oxford University

The Right Tool for the Job

• The worst practice that I've learned to mention is using the wrong tool for the job. This may seem self-serving because I'm a tool maker, but the fact of the matter is that this is a really tricky problem. Mary talked about not building things from scratch, not starting over. Every project is not a special snowflake. I feel like we in this room need to give people guidance. It is hard because if you have manuscripts and you're using Flickr, if you are transcribing those manuscripts it is a terrible platform, but if you are asking people to identify fragments of manuscripts it is a great platform. This is tricky, and people don't have the guidance and they don't have the tools. • Ben Brumfield, Collaborative Manuscript Transcription

Flexible Response to Community Needs

• Chris said that the top-down approach doesn't really work. When we started OpenGLAM we never envisioned so many local groups because we thought we would have one big community and everybody would work together like that. Actually, what we found going forward was that there was really a need for these local groups. I think it's important to be flexible enough to recognize the needs of the community

and work with that. At the moment there is a lot more visibility to the local groups and they are very successful. We just found a way to connect them through the wider OpenGLAM group and in that way it works really well. If we had stayed trying to control everything through the core OpenGLAM group it wouldn't have worked so well. So it's important to be flexible to whatever the community needs. • Lieke Ploeger, OpenGLAM

USER INPUT IN PLATFORM DESIGN

• Thank you all for a really super panel. I hope you can help me with an issue that we have. We are using a smartphone app, but first of all I will give you the context. It is informal learning within a local nature preserve, working with the naturalists there and the communities that come in. They tend to be around for a week or two weeks, like a national park in a way, and the platform we

CCLA @crowdconsortium · May 8
(Reference to the Ransom Center Flickr project? blog.hrc.utexas.edu/2012/07/26/man...) #crowdcon

http://blog.hrc.utexas.edu/2012/07/26/manuscript-waste/

*Best Practices Panel (from left):
Lieke Ploeger, Ben Brumfield, Chris Lintott,
Tina Phillips, Kim Christen Withey*
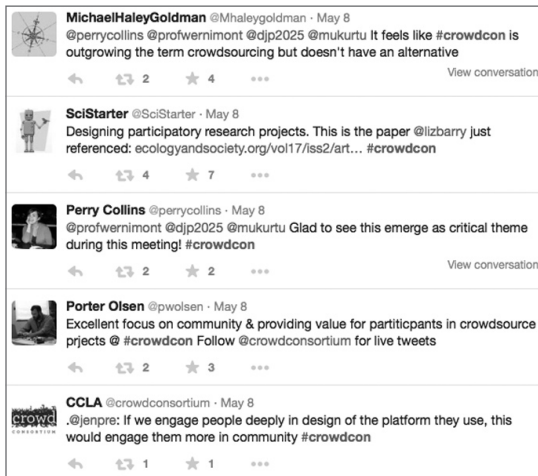
developed puts together a smartphone app that you can use as you go out and take pictures of the things that you see and upload them. We have it quite nicely scaffolded for educational activities within the nature preserve. Those are automatically geocoded and time-stamped, and they appear on the tabletop and also on the website so that people form a community around those three integrated pieces of the technology platform.

One of our research questions was not just the scaffolding of activities at the nature data collection site, but also the belief that if we engaged people deeply in the design of this platform that they use, this would engage them more as community members. We had a lot of difficulty scaffolding for that. As an HCI person I know it's hard to get design ideas from people but within this distributed set-up, which is very different from a typical participatory co-located design activity, it is hard. Since you have a lot more experience, I wonder if any of you have good suggestions to get the parts of the project to work much better, or at least to explore what are the useful parts of the project. • Jenny Preece, University of Maryland

Using Volunteers from a Different Crowdsourcing Project for Your Platform Design

• We are all looking horrified. It's a very hard problem and it's something we struggle

with. I talk about the difficulty of getting people to the point where they think they can participate in a project. There's another huge jump to get to the point where you get people to think they can help you design it. By that point I think one of the problems is that you are dealing with an audience that is designing for itself. Certainly we have a lot of pushback from our volunteers who want us: a) to stop dabbling with these other problems and concentrate on galaxies or whatever they're involved in; and b) build more tools "for people like me." By that stage people are confident and excited and we get a lot of feature-creep when we ask people for input.

It has been done in our group by this project called SpaceWarps [http://www.spacewarps.org], which is another astronomy project. They identified 20 people from a different project and flew them to a workshop like this and spent three days with them, and that worked really well. I think about half of the people came. What worked was going to a different crowdsourcing project and asking for help there to design a new one. So it is going to a different community, and maybe this points to us sharing our communities and being open to that. • Chris Lintott, Zooniverse, Oxford University

User Testing/Input at Prototype Phase, Not Design Phase

• I don't fully understand the context or the

content of the project, but I think there's a danger in bringing in too many minds in the design phase. I think once you've got something that is a prototype, that's the point when you want to bring in a group of people, and I think 12 to 20 is probably the right number of people to test a prototype. You can get about 90% of all the things that people would want, and then it's up to you to prioritize what are those things that we can reasonably do to avoid that scope creep.

• Tina Phillips, Cornell Lab of Ornithology, Cornell University

• We are actually doing a lot of that. The research question is: Can we engage people more deeply in contributing data by involving them much more deeply in the design.

• Jenny Preece, University of Maryland

• That's an interesting question. • Tina Phillips

INTEGRATING CROWDSOURCING WITH DISCOVERY (OR NOT)

• I wanted to follow up on Chris's comment about the idea that crowdsourcing doesn't need to provide the full research experience, it's just this launching-off point. It made me think about a new project coming out of the U.K. that has mid-Victorian illustrations. They bill it as an archive site and it is primarily a discovery site, so you can go there and browse, search, create exhibits, and there is information about how to take

this into the classroom. But there is also a crowdsourcing tool there, so when you go in and search an image it brings it up and it brings up the metadata. On the right side it also says, "What kind of image is this?" And then once you classify it, "Give us some tags that tell us more about it."

I'm wondering if that is where we should be thinking about crowdsourcing tools as we go forward. Do we want to integrate them with discovery systems, or is it better to keep those two separate because they are different audiences and they have two different needs? • Trish Rose-Sandler, Missouri Botanical Garden, Biodiversity Heritage Library

Crowdsourcing First with Links to Exploration

• My sense is that it's the opposite, that you want to get your crowdsourcing site and then provide easy access to discovery tools so that the way in, this transformative experience, comes from participating. So you want to start there. With most Zooniverse sites it's one or two clicks and a very short tutorial that probably doesn't tell you everything you need to know, but just enough to get you started, and then points sensibly to other places. I think it's that model that works for me anyway. That way you get people participating and then moving on to exploring. I think people who are already exploring aren't that interested in stopping along the way to help you with your
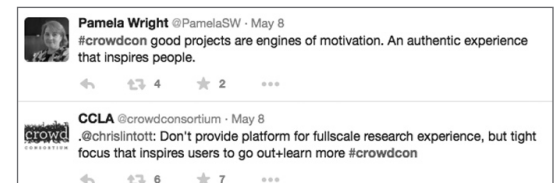

Resource:
Deep Engagement in the Science

"No Windows. One Exit. Free Drinks:
Casino-Driven Design for Crowdsourcing,"
Al Shaw, *ProPublica,* March 20, 2013

https://www.propublica.org/nerds/item/
casino-driven-design

Recommended by: Ben Brumfield,
Collaborative Manuscript Transcription

Trevor Owens @tjowens · May 8
The "Casino-Driven Design for Crowdsourcing" piece instantly begs Qs on manipulation/user motivation bit.ly/1JTMlmx **#crowdcon**

critical play @criticalplay · May 8
@jonvoss we need to take care with gambling and #gamification references! **#crowdcon**
View conversation
↩   ⇄   ★ 2   •••

Victoria Van Hyning @VanHyningV · May 8
Good shout @benwbrum: 'No Windows. One Exit. Free Drinks: Casino-Driven Design for Crowdsourcing' propublica.org/nerds/item/cas… **#crowdcon**
View summary
↩   ⇄ 2   ★ 7   •••

---

crowdsourcing. • Chris Lintott, Zooniverse, Oxford University

The Zone of Producing

• I would like to recommend a paper. The goal is to maximize productivity. You want to get your volunteers in the zone of producing and you don't want to constantly be saying, "Oh hey, look here, look over here." • Ben Brumfield, Collaborative Manuscript Transcription

Pointers to Exploration When They're About to Stop Classifying

• But for the broader goal we haven't come to a decision on this yet, but we'll tell you when we do. We can predict when somebody's about to stop classifying on one of our projects, so we've been debating what we do at that point. One of the options

is that we can probably interrupt them and say, "Now you can go and explore." You can start thinking about complex interactions like that. • Chris Lintott, Zooniverse, Oxford University

ATTRACTING/IDENTIFYING VOLUNTEERS; UNDERSTANDING AUDIENCE TYPES

Interest-Driven Collaboration (à la Public Lab)

• This pathway that we are identifying for people who are browsing, looking for what might get them interested, looking for an early window to hook them on participation, that is a very strong pathway. In addition there is another very strong pathway, which is connecting to people's interests that they already have. Not everyone is a blank slate looking to be caught or have their interest piqued. Many people already have their interests about the world.

If I can share an example from an open source community, Public Lab [http://publiclab.org], people are able to connect with the community starting with what question they have about the world. What people are then collaborating on (there is much less crowdsourcing in this case) is figuring out what tools and methods are needed to go out and research their question. And then, having those tools in hand, to go do research which is maybe not novel scientific discovery but the type of

---

Volunteer Productivity as a Measure: Context-Driven & Dependent on Project Goals

• I want to complicate it a little bit. I don't disagree with what Chris said, but I also think the notion of getting as much as we can in productivity can also be off-putting and it works for a certain type of audience or community but will dissuade others. So I think it is going to be very context-driven. In the instance of Mukurtu, which was the Plateau People's Web Portal funded by NEH, in the first year we digitized less than a hundred items. Now obviously we could have digitized those items in a day, but getting the community input for each one of

those when you're dealing with images of boarding schools where children were taken away, you sit with people for days for them to talk about it. So trying to get people in and out really fast works for some things but not for others. I just don't think we can make a blanket statement about productivity without again thinking about what are the goals for each project. I'm not disagreeing, I'm just complicating it. • Kim Christen Withey, Mukurtu, Washington State University

science that is monitoring and surveillance, and getting some real-world outcomes in policy change. There is evidence, holding industrial polluters accountable in courts, and actual advocacy wins.

Zooming out a little bit, I want to say something about the type of engagement. I'm hearing a lot of discussion where people are saying you start from a small task and then you scale into this interactive community. I want to put out that maybe that path works, but that seems like a hard way to go. If there are a diversity of tasks to begin with, and your primary point of engagement is what people are actually interested in, and you provide the place for people to connect with other people with similar interests, you've started a community, which then identifies the tasks.

• Liz Barry, Public Lab

Attracting the Unconfident and Science Averse

• I agree with 99.9% of that. Everything you've said is true, but I think my ambition is broader than that. I am after the community who wouldn't even consider themselves able to frame a task. I'm after the community who run away, who say, "I hated science when I was at school." For people who are confident and when there is a world of tools, you can start with their interests. But I'm also after people who not only aren't interested in astronomy but think they can't be. So I think we need both regimes and then

people can move between those powered by their own interests. But we shouldn't give up on people just because they don't yet feel ready to think in scientific terms or think in research terms. • Chris Lintott, Zooniverse, Oxford University
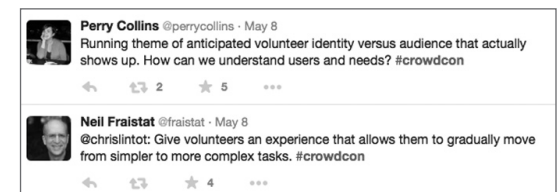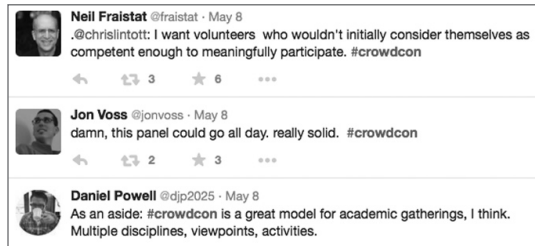
Concern-Driven Motivation

• The people I think Liz might be talking about are not saying they're interested in science. They're sometimes coming at it from a concern that their drinking water isn't safe—a community that has been affected by fracking, for example. So they are concerned, but they don't necessarily know how to articulate the research questions, how to access the right tools, or interpret the data in a way that is useful so that they can then make a difference in policy. If we approach them saying, We have a citizen science project where you express all of those concerns that you have," they wouldn't become engaged at that level. But what Liz is talking about is a very important factor and it is different, but it is a form of crowdsourcing, and some of those considerations can probably be used in a forum that is more digital in structure.

• Darlene Cavalier, SciStarter

Many Publics, Diverse Motivations

• I think these comments underscore what was said earlier about many publics. There are many, many different kinds of people who come to our projects, many different

Perry Collins @perrycollins · May 8
Running theme of anticipated volunteer identity versus audience that actually shows up. How can we understand users and needs? #crowdcon
2   5

Neil Fraistat @fraistat · May 8
@chrislintot: Give volunteers an experience that allows them to gradually move from simpler to more complex tasks. #crowdcon
4

Jon Voss @jonvoss · May 8
tension of co-designing projects w communities vs. designing questions w community interest in mind. #crowdcon
3   6

critical play @criticalplay · May 8
the way in, the transformative experience, comes from participating @chrislintott #crowdcon

> **Neil Fraistat** @fraistat · May 8
> .@chrislintott: I want volunteers who wouldn't initially consider themselves as competent enough to meaningfully participate. #crowdcon
>
> ↩   ↻ 3   ★ 6   •••
>
> **Jon Voss** @jonvoss · May 8
> damn, this panel could go all day. really solid. #crowdcon
>
> ↩   ↻ 2   ★ 3   •••
>
> **Daniel Powell** @djp2025 · May 8
> As an aside: #crowdcon is a great model for academic gatherings, I think. Multiple disciplines, viewpoints, activities.

motivations. They can be diverse and complex, ranging from, "There are some health issues in my backyard that I want to get politically or civically involved in." Or there are people who have said, "You know what? I just need my science niche, and just looking at these bird images gives me that 10 minutes of time so that I can still feel like a scientist even though I work in an administrative office and I don't do any science at all, all day long." There is just this incredible diversity of motivations for why people engage. I think it underscores that "many publics" idea. • Tina Phillips, Cornell Lab of Ornithology, Cornell University

**Defining Commonalities,
Defining the Audience**

• I want to follow up on this great discussion and what Tina said about knowing and understanding the audience and what Chris said about no one being invested at the beginning in this brand new thing, which really makes sense. I have a thought on this. I'm wondering if there are any commonalities between the participants who are partly interested or might be interested because any time you try to build something you have to define your audience to some extent, otherwise you can't build the right tools. So maybe there are commonalities we could identify, or maybe some of you have best practices for engaging and then defining the audience to the extent that is necessary for these new projects. • Liz MacDonald, NASA, Aurorasaurus

BEST PRACTICES BRAINSTORMING

Introduction

Matthew Manos, verynice.co

We just had a great panel around best practices (and worst practices to an extent). What we want you to do in this next session is reflect on these practices that were presented but also reflect on your own best practices that you've seen either through these workshop discussions or through your work out in the field and in your institutions.

One thing I want to remind you of is that best practices very often are attributed to success, but I've found more often than not they actually are the result of failure. So if you can use that as a way to help you investigate and identify best practices, it can help open up some doors for you in this brainstorm process.

Facilitators Matthew Manos and Sheila Yoon

What we want each of the original challenge groups to do [*see page 83*] is come up with eight to ten best practices and write those down on sticky notes. Then we are going to have you select your top two and post them on the board.

Brainstorming Report-Outs: Best Practices

1. AB TESTING (ITERATIVE DEVELOPMENT):

We first identified a term called "AB testing," which we identified as the iterative development of a site design and tasks associated with it. We associate that with understanding audience below.

2. UNDERSTANDING AUDIENCE

What we think we gleaned from the panel discussions is that the lessons about a community from one project didn't necessarily tell you anything and wouldn't necessarily translate or apply to your new project community. So the need is for specific understanding of your audience and the difference between online and offline activities, what makes them unique, what the motivations are, etc. The AB testing would be the way to explore that in the design and implementation of your site to make sure that you are on the right path and moving towards the right product.

Daniel Powell @djp2025 · May 8
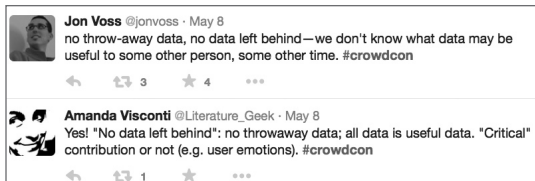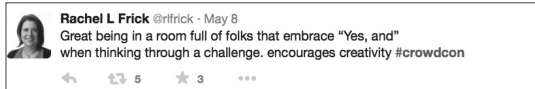#crowdcon Best practices: new tech not always answer, but connecting tools + communities. No data left behind.
2   1

Amanda Visconti @Literature_Geek · May 8
Let's hear it for iterative user testing! #crowdcon
1

Best Practices Ideas Attribution:

Best practice numbers and the challenge groups that generated those practices are listed below. For membership in the challenge groups listed below see page 83.

1 & 2:   Skills training to become effective crowdsourcers

3 & 4:   Balance traditional & non-traditional expertise

5 & 6:   How do we include multiple communities and stakeholders?

7, 8 & 9: Connect education to engagement

10 & 11:   Dealing with failure and framing it as an experiment

12 & 13:   Creating tasks, models that appeal to user communities with multiple stakeholders

14 & 15:   Verifying data using algorithms

16 & 17:   Moving crowdsourcing from edge to core workflow

Rachel L Frick @rlfrick · May 8
Great being in a room full of folks that embrace "Yes, and" when thinking through a challenge. encourages creativity #crowdcon
5   3   ...

Jon Voss @jonvoss · May 8
no throw-away data, no data left behind—we don't know what data may be useful to some other person, some other time. #crowdcon
3   4   ...

Amanda Visconti @Literature_Geek · May 8
Yes! "No data left behind": no throwaway data; all data is useful data. "Critical" contribution or not (e.g. user emotions). #crowdcon
1   ...

Brainstorming in progress

3. NOVEL QUESTIONS WITH SIMPLE TOOLS

In many cases it doesn't require novel technologies to answer the questions that you are trying to answer with your community. I'm thinking about Public Lab, I'm thinking about SciStarter and about other projects where, in some cases, other people have the tools and you have the community. There's no point in redeveloping the tools. Sometimes very simple tools can answer these questions, but that's not necessarily an attractive proposition for funders, right? They seek novel things. That's a challenge, but I think it's also a best practice: to be comfortable as a community in working across these partnerships where people have already developed successful systems.

4. NO THROW-AWAY DATA

No throw-away data, no data left behind. It's easy in some cases for us to think, well, here's the data that we need and so, therefore, this is the data that we are going to collect. But all data is useful data, right? I was thinking about the U.S. Geological Survey that is going back into hundreds of years of data that they never even realized would be useful. A related story is that of the photographer who broke the Monica Lewinsky scandal. He was the only photographer still photographing with film, and all of the digital photographers had ditched all of their archives except for the ones they needed for that story, and he just happened to have the film that contained this useful frame. So the best practice idea is retaining and creating metadata for retention, long-term access and use.

5. CLARITY ABOUT DEMANDS OF PROJECT TYPE & PROJECT GOALS

Our attention was focused on Liz's comment and Chris's comment about the fact that there have been so many different projects and each is going to have a different purpose depending on the particular goals, what you are trying to produce, and how you are trying to work with your users and your audience. So clarity about the demands of the particular project and project goals led us to a kind of action item, a rough taxonomy of project types that could serve as a decision tree regarding some of the project planning steps and design stages you need to take. That could be useful thing to bounce off of to start planning a project. So clarity about the demands of a particular project type and project goals.

6. FOCUS ON RELATIONSHIPS, NOT RECORDS

Focusing on relationships, not records, means focusing on the audience first, even in cases where you might not know who the audience is going to be. You may have data or collections

that you are bringing to the surface and may find new audiences that you didn't expect, or there are projects where the audience is very primary in mind and that is the relationship that you are trying to establish before you can even begin design. So relationships are the focus either way, even if they're relationships you don't know about yet and are trying to attract or relationships you know you can build upon. Focus on that and not just on the records or the metadata.

7. SCAFFOLD: EARLY PRESENCE OF RESEARCHER, DIFFERENT LEVELS OF INTERACTION, EMPOWER THE CROWD

One best practice is this scaffolding idea. It really useful to have the early presence of the researcher in the project, and we emphasize early because that's when we need them. But from that point on you want to enable more complex interactions for the people who are really deeply interested in the project and move them into the leadership roles. There is almost this kind of onion model of interaction where you start of as a novice being guided by the researcher, to becoming an expert and a leader.

8. FEEDBACK: DIRECTED & SPECIFIC (INDIVIDUAL), DATA USE (GROUP), GAME/GOAL SETTING

Another best practice involves feedback, and there are different types of feedback, including directed and specific feedback for individuals, giving them very specific feedback about their contribution. Then there is aggregated feedback to the crowd, so you can talk about how data is being used and the research. Another type of feedback involves gamelike and goal-setting types of things: How many contributions until everything is finished? How many coins have you gotten since last month? Those are quite useful.

9. ENABLING DIFFERENT TYPES OF INTERACTION

Enabling different types of interaction relates to this difference between crowd and community, where a crowd may be individual people who don't talk to each other, they do their work and then they leave; a community is people who actually build relationships and talk with each other in forums and so on. You can support different types simultaneously. Some people might want a forum, and we can provide those and moderate the interaction but also have a different design for people who just want to show up and participate.

10. REFLECT & REVISE

This is a "yes and" best practice. We also talked about AB testing and iterative design pieces, but we wanted to take
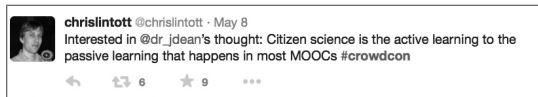
Neil Fraistat @fraistat · May 8
Important question: What are the differences between "crowd" and "community"? #crowdcon

Jon Voss @jonvoss · May 8
crowd vs. community, providing different types of interactions simultaneously to support both. #crowdcon

chrislintott @chrislintott · May 8
@crowdconsortium Excellent argument about gamification to be had. Read Eveleigh's for an argument against : discovery.ucl.ac.uk/1412171/ #crowdcon

http://discovery.ucl.ac.uk/1412171

Trevor Owens @tjowens · May 8
I'd love to see a whole library of these kinds of papers of #crowdcon research!

Trevor Owens @tjowens
""I want to be a Captain!": Gamification in the Old Weather Citizen Science Project" bit.ly/1EUX0sN via @chrislintott #crowdcon

Brainstorming in progress

that a little bit broader and talk about general reflecting and revising across the project. It's not always adding more stuff, it's often really thinking about and focusing on what needs to be revised.

Amy Patterson @amycep · May 8
Onion model of interaction: start as novice, guided by research to become expert and leader. Feedback as best practice. **#crowdcon**

↩ 2   ⇄ 4   ★ 4   •••

Daniel Powell @djp2025 · May 8
#crowdcon Best practices: focusing on relationships, not records, being attuned to audience and tone.

↩   ⇄ 1   ★ 2   •••

chrislintott @chrislintott · May 8
Interested in @dr_jdean's thought: Citizen science is the active learning to the passive learning that happens in most MOOCs **#crowdcon**

↩   ⇄ 6   ★ 9   •••

Brainstorming in progress

11. IDENTIFYING NECESSARY ROLES/ THOUGHTFUL STAFFING

Another best practice involves the roles and kinds of expertise you have on projects and really thinking about identifying the necessary roles and thoughtful staffing. Who are the people with the expertise that is needed on your project? You want to make sure they are included in some way.

12. AUTOMATICALLY ACKNOWLEDGE ENGAGEMENT WITH FEEDBACK

This is about acknowledging engagement with feedback. Many volunteers, particularly newcomers, may be thinking, "Am I on the right track?" There are ways to incorporate feedback for them. For example, there are Zooniverse projects that offer feedback like, "Well done. Maybe you missed this one, but..." It helps to let them know if they are on track through the type of feedback or data appropriate at your particular site.

13. DESIGN FOR PROGRESSIVE PARTICIPATION

This involves building useful roles you can take on the project into the design. You might have core tasks, but you need to think about what other types of participation there might be later on. Maybe they are not all contained in the same project, maybe it is stretching out to resources that lead to outside of your immediate project.

14. LOW BARRIER FOR ENTRY

Low barrier for entry means that your user should be able to engage quickly. In other words, don't give them a registration form, allow them to just get started or to start almost immediately and then capture that information later when you need it, or if you need it. That was another discussion: Make sure you're not capturing more than you actually need and making it hard for your users to get going.

15. RIGHT TOOL/PLATFORM FOR PURPOSE

This best practice is choosing the right tool or platform for the purpose. For example, if you're doing image contributions and you need tagging and licensing, Flickr may be your choice because people are already there and they have APIs, etc. But if you want to do metadata enhancement, don't use Flickr as your institutional repository because they may not be there forever. But don't discount it just because it's not good for one purpose.

16. SMART PROJECT MANAGEMENT

We started out talking about an implementation plan and having a clear closing date for things and decided on encapsulating that in terms of choosing a SMART project. Having SMART goals would be something that would help define the tasks and things to be done.

17. CREDIT, RECOGNITION, ATTRIBUTION & PROVENANCE

This involves, at a high level, tracking the provenance of the contributions, and at a lower level providing the credit and recognition to communities for their contributions.

Brainstorming in progress

Additional Best Practices

The best practices below were generated as part of the brainstorming process but not selected by groups when choosing which best practices to prioritize and present.

- Crowd vs. community — moderation of community
- Construct meaningful experiences
- Matching community & collections
- Cultural context matters!
- Audience analysis
- Be mindful of project goals and <u>audiences</u>
- Flexibility - dialogue - revision
- Be prepared for unexpected audiences & outcomes
- Respect the possibility of many audiences
- Balance stakeholders' goals/authenticity
- Be flexible with best practices, standardization
- Collaborate with diverse networks
- Consider ethics in forming workflow
- Engage professional designer
- Plan to evaluate even if you don't

- Social rewards — same mindset, distributed people connecting
- Two leaderboards: objective — fastest growing
- Synthetic data for tutorials
- Unexpected rewards and external resources
- How data will be used — who uses data
- Go to where users are
- Does data meet need? What format? What is end use?
- Benevolence — support users
- Guided exploration
- Be willing to end/shut down a project
- Pilot (culture and technology)
- Clear messaging about project goals
- Enabling many paths to engagement/ways in to tasks — variety of motivation
- Foster institutional buy-in

PROCESS DEVELOPMENT: ROADMAPS AND AUDIENCES
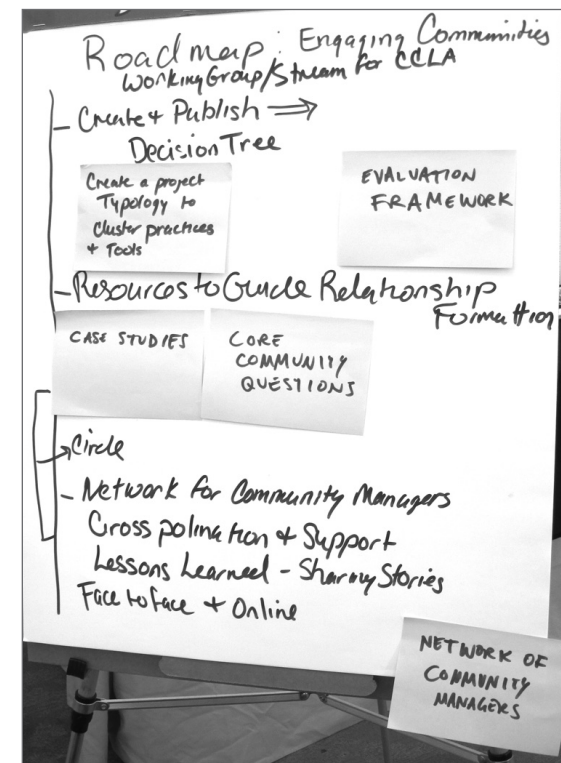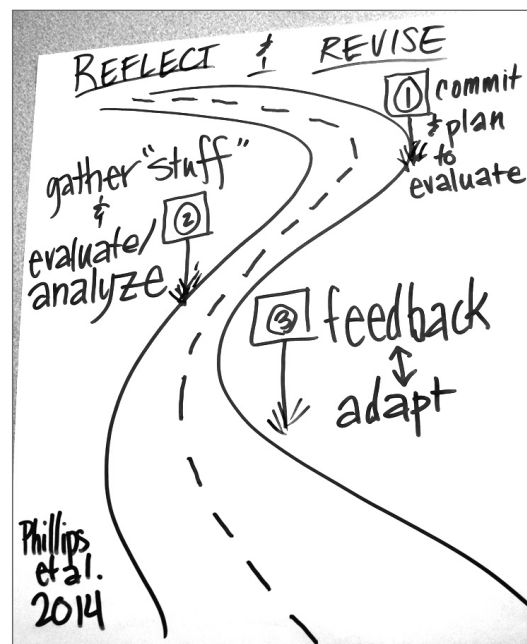
Road Maps

Sheena Yoon, verynice.co

We are going to have you choose one of the best practices from the previous session that you find applicable for your audience based on your original challenge triangles. The exercise is to develop a roadmap with a time frame, a year for example, with milestones you want to achieve regarding your challenge. See how you can integrate and leverage these best practices for these challenges. And extra points if you can also include resources into that timeline.

Personas

Matthew Manos, verynice.co

We have heard a lot of the panels talking about the public, the key audience, the community, or multiple publics. What we want to do is try to get a little more resolution about that and start to identify who are those people. Obviously there could be endless options for that, but we want to start putting some of

Sample roadmaps from Framing Failure as an Experiment group (left) and Engaging Multiple Communities and Stakeholders (right)

those down on paper. We want you to create four personas related to your challenge. These could be people you need to engage to overcome or begin to overcome this challenge, people who would be engaged in your crowdsourcing service or a project related to your challenge. With education, for example, these could be students, teachers, administration, parents. For some of the groups it's going to get more abstract. For example, for Edge to Core Workflow this could need to be organization-to-organization, and so on. It's going to be very different for each group. Sheena developed a framework for doing this, which I'll let her introduce.

Sheena Yoon, verynice.co

You don't have to follow all of this, these are just some ideas and guidelines. A lot of you have already brought up these kinds of questions: What motivates them? What are their fears? What are their values?

Personas of Your Key Audiences

"**Who** is your audience, **What** are they trying to accomplish/gain, **What** goals drive their behavior, **What** gets them excited, **How** do they interact with you, **Why** are they or should they continue to engage with your organization…"

Name:
Age (range):
Gender:
Location:
Education:
Profession:

Level of Tech Literacy: (attention, participation, collaboration, critical consumption)

Interests/Concerns:

Primary Communication Channel: (social media, website, physical venue, email, mail, phone, app)

Primary Service/Resource Utilized from Your Org.:

Primary/Favorite Platforms (websites, apps):
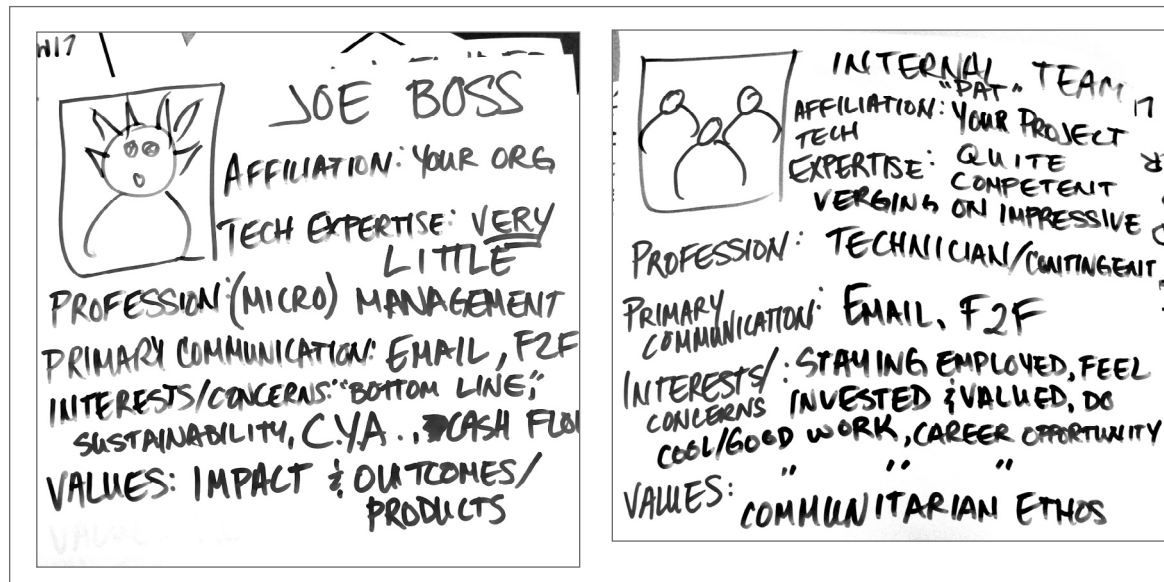Social Media:
News Source:
Mobile Apps:
Entertainment:
Browsing:

Commonalities with Your Org.:

Other Unique Characteristics: (Values, Fears)

Sample personas from Framing Failure as an Experiment group

Process Questions/Feedback:

• When you design personas for real projects in the world you can't do it without doing research. The idea about personas being useful is that they're based on interviews with real people in the world and then they are composites. • Sharon Leon, George Mason University.

• This is just a collective creative exercise, imagining who these might be. • Sheena Yoon, verynice.co

• Okay, but I just want to say out loud, don't ever use anything that we do in this exercise because it's based on what lives in our heads and not real people in the world. • Sharon Leon
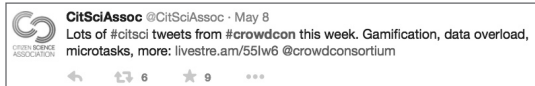
**Process Note:
New Group Formation**

While challenge groups were focused on micro-elements of the topic at large, the facilitators and organizers felt there was a need for one more group that would zoom out and serve as more of an umbrella, focusing on the future of this consortium as a topic. Five participants shifted from their old groups and appear on the new group roster (see page 134)

PROJECT DESIGN PROTOTYPES

Retrospective and Overview

Jake Dunagan, verynice.co

Before giving instructions for this final session I want to give a big picture view of the two days and the level we've set ourselves for this last push. This is drawing towards a design-oriented workshop, so we've been pushing toward the active creation of ideas and responses to the challenges that we've created. I just want to lay out where we've been.

We went over some major concepts and frameworks for crowdsourcing. We've looked at driving trends and burning issues that are relevant to this space. We have heard a lot of examples from practitioners on the ground who are actually doing this work, and we heard about best practices and worst practices. We have had a good base of examples to pull from regarding how to approach these issues. Then we were able to identify some core challenges by asking good questions. I think all of those things allow us to understand the context, the situation, what we're dealing with, and the core challenges that are facing us in a useful, robust way.

Now we are moving into the more design-focused phase. We have created user personas, we have tried to take a more fine-grained look at the people who will be involved in this,

real human beings, even though we imagine them. I totally echo what Sharon was saying in practice. We are in a workshop, so we can't go out there and interview people, but the standard of practice is to go out and actually talk to these people and get some flesh and blood on these assumptions that we are making.

Finally, we are now going to be generating responses to these challenges, whether a prototype or service, some kind of tangible responses that are plausible, robust, and can be actionable after we leave. Our success metric is not whether the things that come out of this room specifically get funded, but that we have learned to stitch together all of these pieces that we have worked through over the past two days, and do that coherently so they are sensible now. Now if some of these actually do turn into big projects that would be fantastic. I think we are heading in that direction and there might be some of those.

So that is what we are going toward. We want to come up with a model process that you can recreate and take out there, so you know what you're doing and have a shared language when you do want to put something together, whether that is in the form of something we come up with in our last pitch, some variation of that, or something new. You understand the process intimately, step by step. That is what we have been trying to do, that is what we are pushing for, and I encourage you to focus in on this last session. Come up with something good that

you would be proud to show anybody outside of this room. We are going to now move into instructions for our final session.

Design Instructions

Matthew Manos, verynice.co

What we are really aiming towards is this final pitch session. We are using the term "pitch" loosely and not necessarily talking about a traditional pitch to some venture capital firm. It is in line with what Jake was saying. We are trying to weave together this big picture with all of these things that we've been doing.

In the interest of getting more specific with this presentation or pitch, we want you to be answering these three questions:

- *What is the concept?*
- *What is the experience or how does this thing work?*
- *How can it be executed or funded?*

We want you to present that in three minutes, and then we are going to have a group discussion as a whole for another three minutes after each of those sessions.

The way we are going to get there is by using a new method called a Project Design Canvas. It is based on the Business Model Canvas, which some of you might be familiar with, but we have tailored it more toward projects that are in this zone. Each group will have a big poster with five different sections on it. This is a framework to help you work towards creating this presentation. This is something that is not necessarily set in stone, but it is something that can pave the way for you.

The first is impact and outcomes. What is the intended impact of this deliverable? How will that impact be measured? Then there is funding and sustainability: How will the deliverable be funded? How could it be self-sustained? For the execution plan, how will the project happen? What does that team look like to actually create this? Then partners: What external partners can take part in the deliverable to help you make this happen? And finally, products and services. How does this actually work? What is this thing?

We would like for you to create a crude prototype. We don't have that much time to actually make something incredible, but we want some way of making these ideas tangible and bringing them to life.

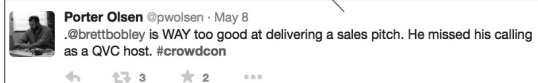| Impact & Outcomes	Execution Plan
	Partners
Funding & Sustainability	
	Products & Services

Project Design Canvas Worksheet

Group Members

- Katie King,
  University of Washington, Seattle
- Bob Horton, Smithsonian Institution,
  National Museum of American History
- Brett Bobley,
  National Endowment for the Humanities
- Eva Caldera,
  National Endowment for the Humanities
- Darlene Cavalier, SciStarter
- Tim Olsen, Gonzaga University

Dr. Dean @dr_jdean · May 8
. @brettbobley drops the mic at **#crowdcon** in pitch for a crowdsourcing bootcamp--baddest CIO in the game!

↩   ⬆ 1   ★ 2   •••

Andrea Wiggins @AndreaWiggins · May 8
Here's my Brett meme: memegenerator.net/instance/62029... **#crowdcon**

↩   ⬆ 2   ★ 2   •••

One of many Brett Bobley memes as he pitches CCLA Bootcamp

Porter Olsen @pwolsen · May 8
.@brettbobley is WAY too good at delivering a sales pitch. He missed his calling as a QVC host. **#crowdcon**

↩   ⬆ 3   ★ 2   •••

Design Pitches

CHALLENGE:

SKILLS TRAINING TO BECOME EFFECTIVE CROWDSOURCERS

The Pitch

Presenting: Brett Bobley, NEH

Let's just say for the sake of discussion that you don't know where a librarian is or a public library and you say to yourself, "Boy, I have a terrific collection of local resources. I'd love to do a crowdsourcing project, but I just don't know where to begin or where to start." Or maybe you're a scientist and you're collecting data and you're thinking, it would be great to get the crowd involved to help me collect my data, but I've never done a crowdsourcing project before and I don't really know where to begin.

Well I can tell you where to begin: CCLA Bootcamp! CCLA Bootcamp is the place to go. Our professionals, who have literally done millions and millions of hours of crowdsourcing, can teach you everything you need to know about the latest platforms, how to install them, how to configure them, how to engage your audience, and ways you can train your end users to make them effective crowdsourcers who will not only be bringing you terrific data but enjoying engagement with your collections in a great way.

Skills Training to Become Effective Crowdsourcers: CCLA Bootcamp

Impact & Outcomes

- Better projects
  - User training
  - Better creators
- More projects
- Better products/data/research outcomes

Funding & Sustainability

- LB 21st IMLS grant
- Promote standards (for re-use)

Execution Plan

- Write the grant
- Analyze platforms/tools/design
- Analyze audience
- Write use cases
- Do research
- Conduct training and evaluation

Partners

- Professional associations (ALA, CCLA, etc.)
- Advisory board
- Service providers
  - Technology
  - Curriculum

Products & Services

- Research product
- Educational products
- Academic publications
- Practitioner publications
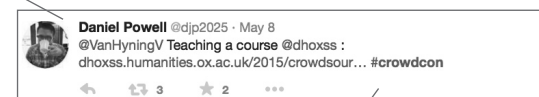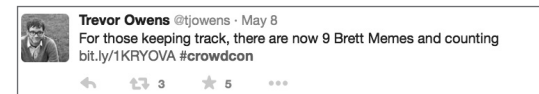- Workshop—online tools, materials
- Great curriculum

CCLA Bootcamp, by the way, is of course sponsored by CCLA as well as several other professional societies with wide expertise in crowdsourcing. So you might ask, who is behind this CCLA bootcamp and how do we get a role in it? What we did was apply for a grant at the IMLS via the Laura Bush 21st Century Librarian program, which is already an established program that creates training sessions and opportunities. We got the grant from them and then did a careful research survey to learn everything we could about crowdsourcing and all the latest platforms, and we brought in some of the top instructors. Our founder is Mary Flanagan, and you will be trained by people like Mary and others on everything you need to know to become a crowdsourcing expert.

And after taking our crowdsourcing training and workshop, if you are not a crowdsourcing expert, Mary will give you your money back.

Money back guarantee
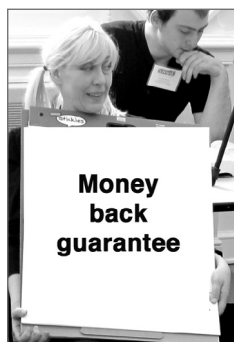
Questions, Answers, Input

• I know it's shameless and I'm sorry, but I am running a little session at Oxford this summer at the Digital Humanities Summer School and you can do this—set up your own project, try out the metadata, get people to crowd-source. So if you want to come and try that, come and talk to me. • Victoria Van Hyning, Zooniverse, Oxford University

Trevor Owens @tjowens · May 8
For those keeping track, there are now 9 Brett Memes and counting bit.ly/1KRYOVA #crowdcon
↩ 🔁 3 ⭐ 5 ···

Daniel Powell @djp2025 · May 8
@VanHyningV Teaching a course @dhoxss :
dhoxss.humanities.ox.ac.uk/2015/crowdsour... #crowdcon
↩ 🔁 3 ⭐ 2 ···

http://dhoxss.humanities.ox.ac.uk/2015/crowdsourcing.html

Skills Training group

Formerly: Balance Traditional and Non-Traditional Expertise

Group Members

- Nick Adams,
  University of California, Berkeley
- Courtney Young, Pennsylvania State University, American Library Association
- Jeff Bigham, Human-Computer Interaction Institute, Carnegie Mellon University
- Perry Collins,
  National Endowment for the Humanities
- Kirk Jalbert,
  Rensselaer Polytechnic Institute
- Jeremy York, HathiTrust

Nick Adams and Perry Collins

CHALLENGE:
CAPITALIZING ON DIVERSE EXPERTISE

The Pitch

Presenting: Perry Collins,
National Endowment for the Humanities

So far we have done kind of a poor job or only a so-so job of capitalizing on a diverse range of expertise. Earlier today we thought the issue was that we have done a bad job of balancing traditional and non-traditional expertise, but we decided it is diversity we are talking about—a range or even multiple ranges of expertise and not a binary split between traditional and non-traditional.

What we are proposing is the "Crowd Co-Creation Cookbook." What we would like to see is something that allows us to legitimize all expertise—not a particular kind or just the gaps, but all expertise—and define flexible roles for the many different stakeholders who might come into these projects. Through documentation and some kind of toolkit, cookbook, whatever you want to call it, we want to give people somewhere to go to find resources on this. And again this is not resources on crowdsourcing as a whole or all of the things you might do, it is specifically a resource to work on co-creation and on bringing new kinds of stakeholders into the picture.

We are thinking of things like a workflow design and user manual; a series of use cases designed by different disciplines from different kinds of communities, ranging from people who are in the academy to people who are maybe even casual users to talk about their experiences with crowdsourcing projects; and ways for us to talk about documenting contributions to what is essentially a series of recommenda-

---

Capitalizing on Diverse Expertise: Crowd Co-Creation Cookbook

Impact & Outcomes

- Legit all expertise
- Defining flexible roles
- Career path

Funding & Sustainability

- Planning/dissemination with coordinator
- Rotating editor
- Crowdcon
- Snapshot

Execution Plan

- Planning workshop
- Surveying existing projects
- IDing reps of different expertise

Partners

- Crowdsource experience
- Community organizers
- Reps of diverse constituencies

Products & Services

- Crowd co-creation cookbook
- Workflow/user manual
- Use cases
- Tool/recommendations
- Sustaining/sunsetting
- Documenting contribution (GitHub?)

tions and other kinds of tools for all different kinds of people who might come into these projects.

In terms of execution we would like to do some kind of planning workshop, a survey of existing projects, and ideally ID all of the representatives of different expertise who might be part of this conversation to fit them into the 21st century planning workshop. The people we would like to see are people who have had a lot of experience, and many of the people in this room have lots of experience, but there are also people who are tackling bigger problems of community organizing and participation who maybe don't even know what crowdsourcing is in the sense that we do. And also representatives of diverse constituencies, people who come in and actually feel empowered to ask their communities for feedback, not speaking for them but trying to translate some of their perspectives.

In terms of funding we are going to ask for some sort of planning grant, possibly from the NEH, to have this workshop but also to fund an actual person, which I think is the most important thing, someone to facilitate and coordinate this. And I recommend that it happen in perpetuity, so in the long term we are talking about a rotating editor, someone who is either volunteering or, ideally, someone who has at least a tiny bit of institutional funding, maybe coming from volunteer contributions to something like the CCLA, some kind of pot from which people can pull small amounts of funding. The very baseline is to at least have this preserved and have an institutional home for it, so we are taking a snapshot of this cookbook even if at some point it isn't necessarily added to.

Capitalizing on Diverse Experience group

CHALLENGE:

HOW DO WE ENGAGE MULTIPLE COMMUNITIES AND STAKEHOLDERS IN PARTICIPATORY PROJECTS?

The Pitch

Presenting: Jon Voss, Shift/HistoryPin

Our group was a little frustrated with the whole "crowdsourcing" thing, so we kind of represent "communitysourcing" in a way. Starting with our impacts and outcomes, this is what we want the world to look like as an outcome of this: Projects have stronger community partnerships built into them; there is an evaluation for community engagement also built in; projects and funders support a methodology of co-creation or as Kim calls it, "community agile methods"; and we can create, design and implement projects with reciprocity and trust at the very foundation. So we are tilting on some of the really core concepts of what these co-created projects could be.

Our execution plan starts with a survey of existing successful case studies, a needs and audience analysis, and a draft of evaluation frameworks. We are really looking at things that people can use in their projects. We would also draft case studies, draft project typology to cluster practices and tools, and we think a feedback loop at all stages is important.

Some of the partners we are looking at, aside from the GLAMFEs (galleries, libraries, archives, museums, funders, educators) include local community groups. They are obviously also key to this, getting an understanding of what their needs and interests are. Partners would also include the personas we looked at: community managers, collection stewards, technologists, and project managers. And of course Ben's mother-in-law, who is in a Santa Rosa retirement community, is one of our key partners, which represents getting outside of our regular groups and getting on the ground with communities.

This leads to products and services as outputs, which are toolkits that include: resources to guide relationships; model evaluation frameworks; questions to guide project formation,

---

How Do We Engage Multiple Communities and Stakeholders in Participatory Projects?

Impact & Outcomes

- Projects have stronger community partnerships
- Evaluation for community engagement built in
- Projects & funders support co-creation, "community agile methods"
- Create, design & implement projects that have reciprocity & trust at the foundation

Funding & Sustainability

- IMLS NLG Research Grant
  Community Engagement Framework
- Lightweight infrastructure
- Institutional home, advisory board
- Crowdsourcing Consortium or working group

Execution Plan

- Survey of existing successful case studies
- Needs/audience analysis
- Draft evaluation frameworks
- Draft case studies with specific format
- Draft project typology to cluster practices/tools
- Feedback loop at all stages

Partners

- Local community groups
- GLAMFE (galleries, libraries, archives, museums, funders, educators)
- Our personas (community managers, collection stewards, technologists, PMs)
- Dorit's mom in Santa Rosa retirement community

Products & Services

- Toolkits
  - Resources to guide relationships
  - Model evaluation framework
  - Questions to guide project formation
  - Case studies
  - Core principles document
  - Sample workflows
- Network of community managers
  - Online/face-to-face

the questions that you ask yourself when you are thinking about engaging communities that you really want to take into account; core principles documents; and sample workflows. The other key output is a network of community managers, which could be online or face-to-face. The Crowdsourcing Consortium could certainly be a working group of this.

For funding and sustainability we plan on having an IMLS grant around community engagement, so that is the first step of this, an NLG Research Grant. And all of my panelists are now on the advisory board of this, as well as anyone else in this room who wants to join the advisory board. We would also have a lightweight infrastructure, which helps us in terms of sustainability. It could be something as simple as what we've done for lodlam.net, where resources that are posted are shared with the Crowdsourcing Consortium. And finally, finding an institutional home and an ongoing advisory panel, and again, the Crowdsourcing Consortium is an obvious home for this.
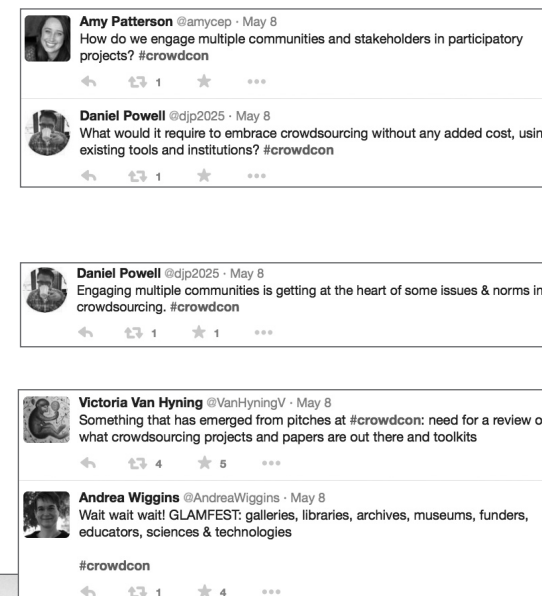
Questions, Answers, Input

• A joke: How does this differ from what we have just presented? • Nick Adams, University of California, Berkeley

• There is a lot of overlap so far between the presentations. • Anon.

• We [*Move Crowdsourcing from Edge To Core Workflow group*] kind of came up with the

same thing too, so this suggests a shared need and we are all on the same page.
• Victoria Van Hyning, Zooniverse, Oxford University

• We will be getting to yours, but was yours focused on the community element as well?
• Jon Voss, Shift/HistoryPin

• It was more from the institutional perspective, but there was the idea of the review of papers and projects and things that are out there, aggregating that information and having people share resources. But you have more of a community focus and we have more of an institutional focus. • Victoria Van Hyning

Amy Patterson @amycep · May 8
How do we engage multiple communities and stakeholders in participatory projects? **#crowdcon**

Daniel Powell @djp2025 · May 8
What would it require to embrace crowdsourcing without any added cost, using existing tools and institutions? **#crowdcon**

Daniel Powell @djp2025 · May 8
Engaging multiple communities is getting at the heart of some issues & norms in crowdsourcing. **#crowdcon**

Victoria Van Hyning @VanHyningV · May 8
Something that has emerged from pitches at **#crowdcon**: need for a review of what crowdsourcing projects and papers are out there and toolkits

Andrea Wiggins @AndreaWiggins · May 8
Wait wait wait! GLAMFEST: galleries, libraries, archives, museums, funders, educators, sciences & technologies

#crowdcon

*Engage Stakeholders
in Participatory Projects group*

Group Members

- Amy Patterson,
  Wisconsin Technical College
- Chris Lintott,
  Zooniverse, Oxford University
- Jeremy Dean, Hypothes.is
- Jen Hammock, Encyclopedia of Life (EOL)
  & Smithsonian Institution
- Tina Phillips, Cornell Lab of Ornithology,
  Cornell University
- Edith Law, University of Waterloo
- Jenny Preece, University of Maryland

CHALLENGE:
CONNECT EDUCATION TO ENGAGEMENT

The Pitch

Presenting: Jeremy Dean, Hypothes.is

My name is Dr. Jeremy Dean from the CCLA project on education. Before I begin I would like to thank the Gates Foundation for all of the funding that has made our specific project possible at CCLA. Our broad mandate is to help crowdsourcing projects get integrated into educational spaces and get adopted by educational institutions. We see education as a place for exponential growth of crowdsourcing communities and are also excited to involve students from the beginning of crowdsourcing projects, not just as consumers at the end. This has also been a great space for fundraising because it has expanded CCLA's ability to apply for funding more broadly.

The specific project that I am here to talk about today is our Ambassador Program. We have a suite of training materials, we have training webinars, we have training videos, modules, things we can put in the mail, things you can find online, that will help you build an Ambassador Program that will support your community's education users. For example, we can help train your user base to go on the ground into local schools to activate classrooms in their communities to become participants in crowdsourcing projects. This would increase participation. It is also about empowering the user community and giving them leadership roles, which we think is important for sustainability.

Regarding the execution plan, the really easy way to do this is to just do it. If you don't want to use our materials, we can just inspire you to go into classrooms and get your users to go into classrooms and get it going.

Connect Education to Engagement: The Ambassador Program

Impact & Outcomes

1. Increased participation
2. Empowering users (ambassadors)
3. Crowd to community
4. Awareness & learning
5. Institutional adoption
6. Nucleates local communities
7. Enhances project fundability
8. Feedback for projects

Funding & Sustainability

YES!!!!!

Execution Plan

1. Probe for interests
   - Community
   - Funders
2. Co-design pilot program
3. Iterative test and design

4. Secure partners & funding

* Just Do It

Partners

1. speakers4schools.org
2. Projects (e.g., eBird, Celebrate Urban Birds)
3. Schools (co-op, extracurricular credit)
4. Volunteer organizations
5. Venues (festivals, schools, GLAM)
6. School admin

Products & Services

1. Centralized system for keeping track (portal) of ambassadors & their activities (dashboard)
2. Training and support for ambassadors
3. Resource library (standards, decks, templates)
4. Recognition (credentialing)
5. Swag!

Connect Education to Engagement group

CHALLENGE:
## DEALING WITH FAILURE
AND FRAMING IT AS AN EXPERIMENT

The Pitch

Presenting: Andrea Wiggins, University of Maryland

We actually did a wireframe of the Crowd Consortium site. Look at that—it's amazing! We were thinking about reframing failure by focusing on the future and on solutions. Rather than what went wrong, we focused on what we need to fix things.

Our best impact and outcome would be that our effort dries up and withers away because we have taken care of the issue (not that that is realistic). The real goal is accumulating a body of knowledge that is shared, that is publicly accessible, and that helps us shift our practices and perspectives around failure and increase risk tolerance, especially when it comes to trying out some new innovative approaches and funding projects that are on the edge.

Regarding funding and sustainability, the way we were thinking of this it would be done with very little money. Just enough to keep the Crowd Consortium going and the people invested would probably be what it would take. We would want this to be something with a limited lifespan, with the products remaining persistent and available as a reference point for people, sort of like a resource library but not.
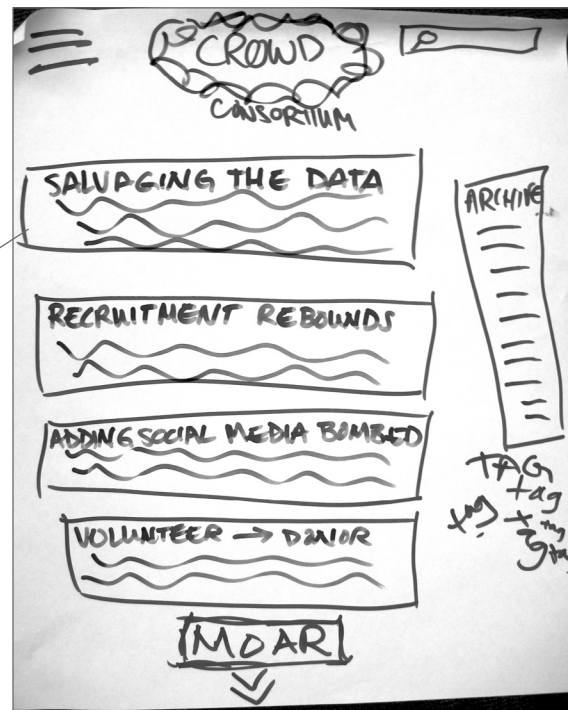
Group Members

- Lacy Schutz,
  Museum of the City of New York
- Liz MacDonald, NASA, Aurorasaurus
- Austin Mast,
  Florida State University, iDigBio
- Michael Haley Goldman,
  United States Holocaust Museum
- Peter Carini, Dartmouth College
- Andrea Wiggins, University of Maryland

Dealing with Failure and Reframing It as an Experiment

Impact & Outcomes

- Obsolescence: Lessons Learned
- Accumulate body of knowledge to inform future work
- Shift in practices & perspectives
- Multi-stage funding opportunities
- Greater risk tolerance

Funding & Sustainability

- Crowdfund & crowdsource crowd consortium
- RCN (+CRI?)
- Limited lifespan with persistent "products"

Execution Plan

- Implement multiparty blog
- * Recruit core <<animateurs>> (*Editor in Chief)
  - Recruit initial & ongoing contributors
  - Elicit stories & self-authored accounts

- Advertise
- Get funders to recruit/prompt community
  - Insert into reporting
- Ingest from external sources

Partners

- Professional orgs/networks
- Funders
- Students
- Crowd Consortium
- Known leaders in research & practice

Products & Services

- Blog, bibliography - with annotation, tags, etc.
- Faceted searching
- Push posts to related communities

The biggest issue with our execution plan is that it is pretty easy to turn on a blog function, but recruiting your core editors and other core people who would drive activity would be the hard part. We need people who would own it and push it forward. If we did recruit some initial contributors they could write their own stories. We could have students in qualitative methods classes help document the stories from any experts who don't want to take the time to do the writing themselves. Then we would need to advertise the heck out of this stuff, and we thought maybe we could even get funders to recruit and prompt the community to contribute to this and maybe insert it as part of the reporting. We would even ingest related material from external sources.

Our partners would basically be the community, funders, students, known leaders in research. We would need to seed heavily with known names and projects to begin with and could then start expanding the scope once there is enough mass to legitimize this.

The main product or service would be a blog, a bibliography, annotation tags, and faceted searches, so it would be easy to find stuff as more accumulates. What would be really great would be to be able to push some of the material out to related communities. For instance, the Citizen Science Association could potentially pull in posts tagged with "citizen science" that would then lead people back to the Crowd Consortium, so you get a nice virtual cycle going.

Dealing with Failure and Framing It as an Experiment group

CHALLENGE:
CREATING TASKS, MODELS THAT APPEAL TO USER COMMUNITIES WITH MULTIPLE MOTIVATIONS

The Pitch

Presenting: Various group members in relay [*see sidebar for members*]

There are multiple publics with multiple motivations that research is bringing to crowdsourcing sites. And not only that, those motivations and interests change as people get more engaged, as they learn new things. So the big question is: How do we create sites that allow for people's motivations to change over time and for the tasks that they are engaged in to change over time? And maybe how can the researchers' questions evolve as the project evolves?

That is why we are creating fulfilling engagement for multiple motivations through customizable, plug-in-driven environments. This interface meets and expands the expectations of both researchers and participants. Let us show you how.

This is the CCLA Repository of Plug-Ins (we couldn't come up with a more creative name, something catchier). We are creating buckets of plug-ins and each bucket corresponds to a certain type of motivation. Different people are motivated by different things. They could be interested in the task, they could be interested in studying an aspect of the topic, or by the extraction and analysis of data.

So what does a crowdsourcing website or platform look like if they have flexible plug-ins?

Group Members

- Meghan Ferriter, Smithsonian Transcription Center
- Lieke Ploeger, Open GLAM
- Ashwin Gopi, New York University
- Carsten Oesterlund, Syracuse University
- Katherine Doyle, Pulitzer Center
- Hector Mongi, University of Dodoma, Tanzania

Submit now: Type, Size, Upload

Note: Sticky note "plug-ins" shown on this and following page were written in ballpoint pen on colored paper and photograph poorly. As a result, the text of the notes is typed underneath.

Creating Tasks, Models that Appeal to User Communities with Multiple Motivations

Impact & Outcomes

- Engagement for multiple motivations
- Agency for determining your own experience
- Design for flexibility in terms of motivation
- Taking ownership and iterating the features
- Take what you want/can out of it
- Positive recommendations—driving new users
- Requests for new plug-ins
- Amount of feedback acquired by CCLA
- * Number of changes made
- Number of new users per plug-in/sets of plug-ins

Funding & Sustainability

- Economic feedback loop
- Grant funded (research)
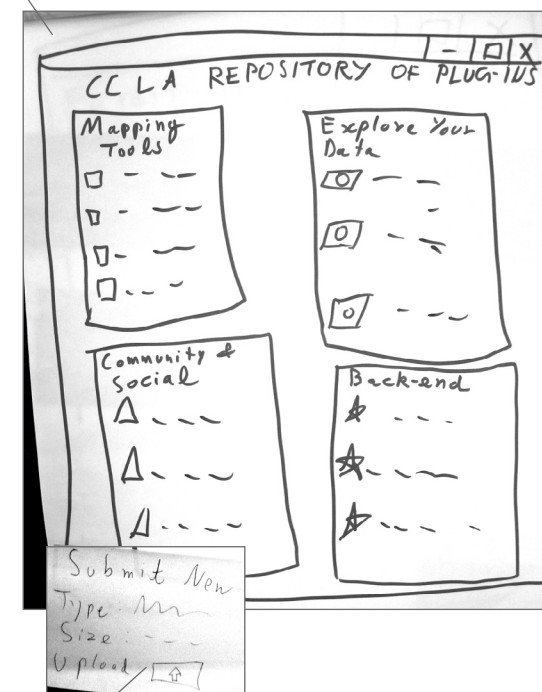- Community ownership

Execution Plan

- Get a grant from CCLA for researcher
- Design for simplicity & ease of use & learning
- User feedback refines the process/task
- Feedback is used to refine plug-ins

Partners

- CCLA & universities
- Individual researchers
- Project's community
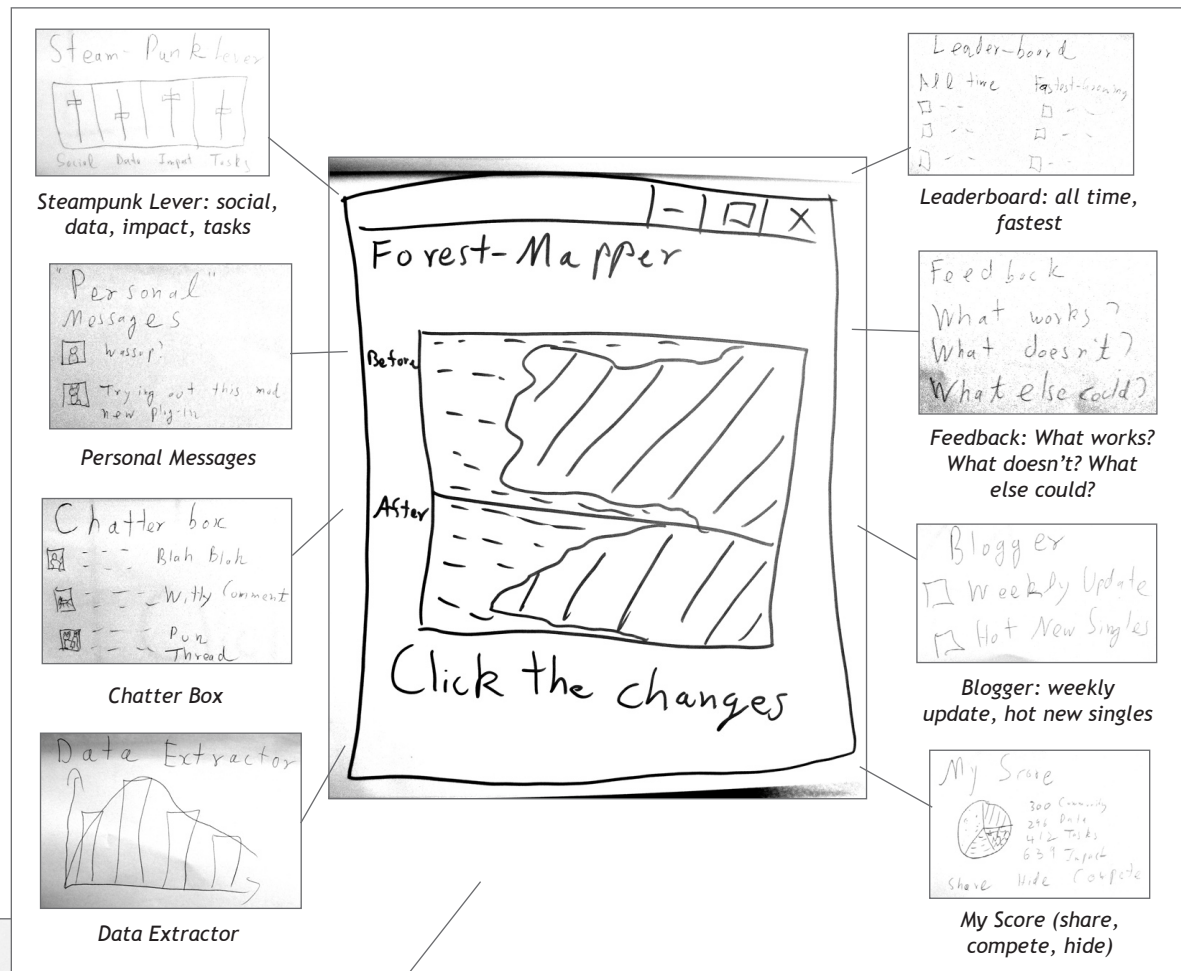- Emerging leaders in community
- Open Innovation community

Products & Services

 [*see prototype graphics*]

Steampunk Lever: social, data, impact, tasks

Personal Messages

Chatter Box

Data Extractor

Leaderboard: all time, fastest

Feedback: What works? What doesn't? What else could?

Blogger: weekly update, hot new singles

My Score (share, compete, hide)

User Communities with Multiple Motivations group

Traditional crowdsourcing platforms have a main task. This is a mapping platform. You are supposed to click the differences and figure out how far the forest has receded. However, what we have added here is a Steampunk-style lever where you can use various sliders. One is for the level of social interactivity you want. You may think, oh, I don't like interacting with humans, so you can use a slider for the level of social activity that you want. And since you don't like interacting with humans you can decide not to see other people's comments on this. Or you may decide you do or do not want to see your score. If you're particularly interested in a dating approach, for example, and you have a huge ego and you want to see your picture up

on the date board you can add that. Or you can add personal messaging if you want to hit on someone.

The main goal here is to get people to take ownership of and shape their own experience, and to have more authority. They can give feedback on the platform to help developers figure out what is the best kind of platform for the users.

Our platform provides support to researchers, community members, and to funders such as CCLA through our databases and our platform. Over the next two years we will develop four plug-ins which will involve community feedback as well as open-source development. We will be able to spend more time with the community making the experience meaningful for everyone in a co-created space.

Questions, Answers, Input

• Who are your users and will they be able to use all of these tools? Could you describe how they're going to be able to do that? Some of them might not know what a plug-in is. • Katie King, University of Washington, Seattle

• The idea is that as people become more comfortable with it or have experience with an item, then they can opt in and begin to alter an item. There will be guidance from community moderators and researchers at the beginning of the experience to say, "It seems like you would be really interested in sharing your experience with this person. Why don't you try pushing this button, which integrates this plug-in for discussion." • Various Multiple Motivation group members

• Do you guys have any plans to add annotation for those supporting your system? • Jeremy Dean, Hypothes.is

• Sure, why not? We can work together to develop that plug-in for you. Or if you have a plug-in that is ready to plug into our database, we welcome it and will integrate it into our projects. • Various Multiple Motivation group members

Ashwin Gopi demonstrating prototype

Former title: Verifying Data Using Algorithms

Group Members

- Zaven Arzoumanian, Wild Me Conservation
- David Miller, National Institutes of Health
- Peter Mangiafico, Stanford University
- Ben Brumfield, Collaborative Manuscript Transcription
- Ben Miller, Georgia State University

Brett Bobley @brettbobley · May 8
Baby whales & crowdsourcing. **#crowdcon**
↩ ♺ 1 ★ 3 •••

Ben Miller @Intransitive · May 8
#crowdcon crowdsourcing supported by ML algorithms for entity resolution of whale sharks & zebras. Future project of Whaleshark.org
↩ ♺ 2 ★ 1 •••

Pamela Wright @PamelaSW · May 8
#crowdcon taking advantage of computational capabilities. Doing science with contributions from YouTube videos...
↩ ♺ 2 ★ 1 •••

CHALLENGE:

EMPLOY/EXPLOIT COMPUTER VISION & MACHINE LEARNING ALGORITHMS TO IMPROVE CROWDSOURCING IN CITIZEN SCIENCE PROJECTS

The Pitch

Presenting: Zaven Arzoumanian, Wild Me Conservation

Our challenge is to better employ and exploit computer vision and machine learning algorithms to improve crowdsourcing in citizen science projects by mining data, verifying and validating data, and improving workflows for human participants. Computational capabilities are exploding and we need to take advantage of those capabilities.

Our specific project is a real-life, long-running, wildlife conservation research and education tool called Wildbook for Whale Sharks [http:www.whaleshark.org]. We want to expand its reach to data that's available on YouTube, Flickr, Twitter and other social media. Most of this is posted by people who have no idea that they could be doing science with their YouTube videos, with the GoPro that they took into the water with them on a diving trip, and we want to reach out to those folks. We want to put targeted ads on their YouTube postings, we

---

Employ/Exploit Computer Vision & Machine Learning Algorithms to Improve Crowdsourcing in Citizen Science Projects: Wildbook for Whale Sharks

Impact & Outcomes

Impact:
- Order of magnitude - more wildlife population data
- Improved understanding of demographics, migrations, life histories, etc.

Outcomes:
- Mining (automated) of wildlife data from social media
- Higher data quality
- Engaged public and more conservation awareness

Funding & Sustainability

- Seek traditional NSF, etc. grants
- Partner with academic, corporate researchers
- Exploit partner resources (e.g., university libraries, CS departments)
- Ongoing community recruitment by tapping already engaged social media participants
- Large grant-making foundations

Execution Plan

- Grant writing, develop pitches
- Targeted ads for data taggers
- Identify research partners

Partners

- Corporate (Google, HP, Xerox, etc.)
- Academic (CS, machine learning)
- Conservation organizations

Products & Services

- Service to researchers: more, better data, research tool, low-cost/free data management solution
- Increased public awareness, support for wildlife conservation & environment that supports it

want them to be aware that they can contribute to good science.

So we need tools to take those videos, find the appropriate frames, turn them into still photos that we can feed into our automated pattern-matching algorithms, and say this spot pattern on this shark or that stripe pattern on that zebra tells us that zebra is the same as the zebra seen by another person at this other time and place. This allows us to track the migration of these animals and figure out how many there are. Are their populations declining or growing? Is their environment healthy in sustaining that population?

The need is to engage researchers, whether they are in academic, commercial, government, or foundation environments, to get their expertise in machine learning and computer vision capabilities to work with these crowdsourcing conservation projects where the crowd can be the people who are unwittingly creating all of this data and posting it to social media. The crowd can also be active participants in helping us filter that data to get the science out of it, or they can be the ones who are validating the computerized matching algorithm output at the end of the pipeline.

This reinforces sustainability, the fact that we have the capability to reach out to a crowd that is already invested. They have traveled half-way around the world, they got on a boat, they went on a dive, and they took pictures and video and posted it for the world to see, so they are already engaged. They are already interested, and we want to tap into that potential.

It improves access to resources if we tap into partners at universities where there are computing capabilities and long-term data storage capabilities, and obviously the built-in avenue for ongoing recruitment that I just mentioned.

The outcomes of all of this are more and higher quality data for the research community, greater awareness of conservation issues for the public and better environmental management to ultimately help your planet. Imagine being able to "friend" a wild animal on Facebook and follow its activities. If you learned that your favorite humpback whale that you saw when you were out on a boat off of Seattle has been spotted again a year later by someone else, only this time it has a baby swimming along with it, wouldn't you care more about how much garbage goes into the ocean?

[*Audience: "Awwwww."*]

Computer Vision & Machine Learning Algorithms group

Group Members

- Daniel Powell, King's College London (University of Victoria)
- Victoria Van Hyning, Zooniverse, Oxford University
- Tom Blake, Boston Public Library
- Trish Rose-Sandler, Missouri Botanical Garden, Biodiversity Heritage Library
- Rachel Frick, Digital Public Library of America
- Jessica Zelt, Patuxent Wildlife Research Center, North American Bird Phenology Project
- Jeremy York, HathiTrust

Pamela Wright @PamelaSW · May 8
#crowdcon next up. Communities as a core part of the organization.

View photo

CHALLENGE:
MOVE CROWDSOURCING FROM EDGE TO CORE WORKFLOW

The Pitch

Presenting: Jeremy York, HathiTrust

Our concept is that we are an institution that has been collecting materials for research initiatives and the educational enrichment of our community for centuries, yet many in the community do not know what is in our collections, they can't find the useful materials, we are not being as effective as we could be in fulfilling our vision. We are losing interest from the community and stakeholders are questioning our value. We believe we can turn all of this around by engaging our community more deeply in key aspects of our institution's operations, from collection development, to cataloging, to delivery of services like user support and document delivery. And we believe that by making communities and potential communities a core part of our organization, we can increase the ownership and investment of the community and we can increase our relevance and impact.

We don't just think this on a whim, we've had a period of about a year of investigation into other crowdsourcing projects and have identi-

---

Move Crowdsourcing from Edge to Core Workflow

Impact & Outcomes

Increase:
- Access to value of collections
- Institutional support for crowdsourcing long-term
- Technological responsiveness
- Visibility
- Relevance
- Community contributions and investment
- A 50% increase in volunteer contribution

Decrease:
- Cost of metadata [insert x] production

Link
- Consumption—production

Funding & Sustainability

- Grant for investigation
- Grant for pilot
- Institutional buy-in
- Interinstitutional collaboration
- Community manager (interact with volunteers)
- New staffing/changed staffing

Execution Plan

- Sharing results of investigation—education
- Establish consulting service/comm
- Investigation (see attached)
- Pilot/example
- Plan for staged implementation

Partners

- Collaborating institutions
- Domain orgs, community groups, volunteers
- Users of data
- Finance, IT, administration, middle management, staff

Products & Services

- Service to enable collaborative editing
- Training
- Consultative service
- OCR correction, pilot, report, discovery, document delivery, etc.

fied very real impacts that come from crowd-sourcing in the visibility and relevance that they have, the access to the collections, and also the willingness of stakeholders to invest in increased collaborative opportunities.

What we are proposing to do is identify three other institutions in our area and we are going to engage in a collaborative project that leverages a small crowdsourcing project we've had on the side, a small cataloging project. We are going to expand that to go across all of our collections in this area, and we hope that by generating success or showing some success in this, we can reach a stage of development whereby we integrate the community's input into all of the aspects of the operation that we deem are important, which are a significant number of them.

We have identified some organizations, both regionally and nationally, that we think will be able to encourage some of their members to participate in this. We think they will be interested to form a core group. We are going to be

hiring a community manager. We are initially using grant funding, but over time we believe and hope that if we demonstrate success this will become a core part of what our staff do that is integrated into their training and their activities.

Edge to Core Workflow group

New "metachallenge" group, formed on day three

Group Members
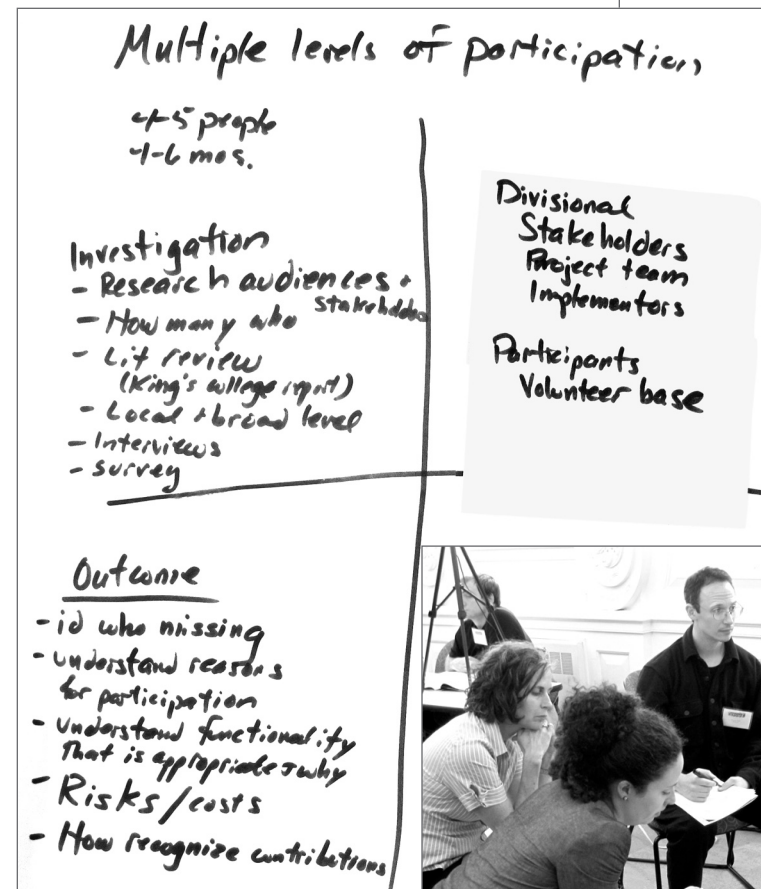
- Neil Fraistat,
  MITH, University of Maryland
- Alexis Rossi, Internet Archive
- Mary Flanagan,
  Tiltfactor, Dartmouth College
- Trevor Owens, Institute of
  Museum and Library Services
- Pamela Wright, National Archives
  and Records Administration
- Amanda Visconti,
  MITH, University of Maryland

CHALLENGE:
A USEFUL, THRIVING CONSORTIUM

The Pitch

Presenting:

Mary Flanagan, Tiltfactor, Dartmouth College

Our challenge was to create a useful, thriving consortium and many of you have actually done the work for us, but some of the things that we talked about were really different. Our pitch here has to do with creating a lightweight, nimble community as a consortium of organizations from which members promote collaboration, raise visibility of projects, host meetings, organize educational resources, and share tools and best and worst practices. So we have cookbooks, we have failure case studies, we have bootcamps. We have studies and evaluation materials, and even have central funding proposals. These are all ideas that came out of those post-it notes that I was

A Useful, Thriving Consortium

Impact & Outcomes

- Create useful, thriving consortium
- 500+ members, 3+ meetings/year
- Knowledge sharing through Google groups, etc.
- Find collaborators
- Get more users for projects

Funding & Sustainability

- DH - startup
- Laura Bush 21st Century
- BS Plan dev
- Tiers $ for Fed Orgs
- NEH - Institute
- Preservation + Access
- FEDLINK - sell stuff to libraries

Execution Plan
- See roadmap

Partners

- Folks/orgs here and that have participants

Products & Service

- Project profiles & Publications
- Agreed-upon set of best practices
- List of successful projects (including funded proposals)
- List of crowdsource projects (User-facing? Internal?)
- Asks & offers
  - Student project possibilities
  - Institutions want help
- Documentation of prior projects
- Have: website, mail lists, announcements, studies/results
- Tool list
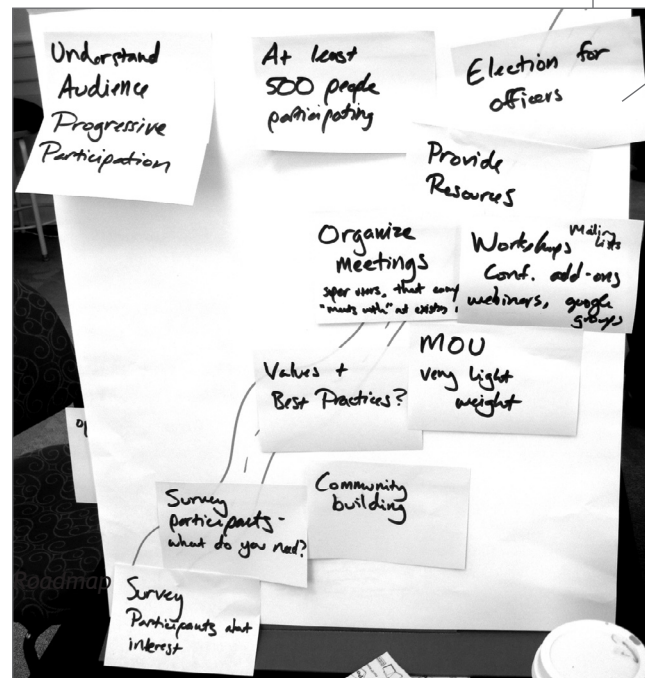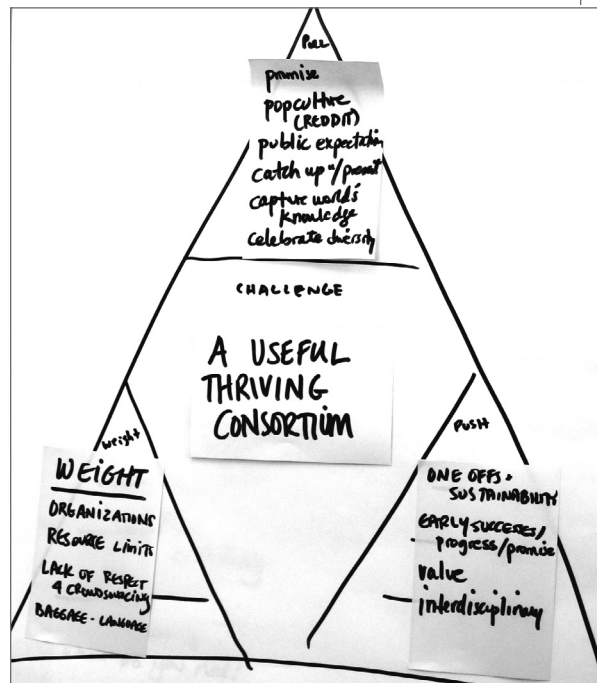
Triangle (far left), road map (left)

sending around to elicit your input about the consortium. We even have material developed by a new group in this room and beyond the GLAMFEs. We need you to come forward, however, and chip in.

Here are the principles of this consortium, which relate to Neil's expression, "joined at the hip." There is this ethos that when we are going to write letters for an existing proposal or are going to go in and say, "Hey, we could do this thing with Crowd Consortium," we involve more than one person from the consortium in our application.

Neil Fraistat, MITH, University of Maryland

So if you want to leverage the Crowd Consortium in a grant, you would have to include at least one other member of the consortium within the grant project.

Mary Flanagan, Tiltfactor, Dartmouth College

It is an ethos that we are actually building a community and working together. It was pointed out by a program officer in the room that this would strengthen proposals.

Other input includes: respecting audiences, decentralized leadership, low barrier to access, and making a lightweight memorandum of understanding. This came from some of our larger organizations who feel the need to have an MOU statement for institutional buy-ins that let them spend some time or some other kinds of commitment. So people would have some

kind of MOU on the website and you could log in and get a quick piece of paper. The idea is to keep it nimble, keep it lightweight. Another founding principle of this group is to promote diversity and inclusion.

So how does this work? We talked a little about organization. The first year would be a start-up board, almost like when you buy a new condo—it comes with a board and then there are elections. During that time people are going after projects and finding ways to share some support for the organization. After a year everybody could get a vote and we set up a diverse slate of leadership possibilities, and those folks act as a guide. Usually within the Crowd Consortium the proposals that we wind up putting forward are bootcamp ideas or hosting a meeting. That way we can keep contributing in some way to welcoming everybody getting back together again because that was one of the most important interactions, physically talking to each other here. This was the first time many of us have been in a room with people who do this kind of work of this size, and that is really exciting. Keeping the conversation going is one of our big tasks.

The website has already been designed and is a destination for dissemination. It is a mechanism to help your work

Useful, Thriving Consortium group

Importance of/Need for Consortium

• If something like this doesn't come into being, all of these lovely ideas by all of the groups probably won't come into being either. If there's a united push behind them there's a much better chance that things like this can happen. **• Neil Fraistat, MITH, University of Maryland**

be more sustainable and have more impact. Our services are member-generated and disseminated by the group: workshops, documentation, lists of projects and tools.

One phrase was "asks and offers." Let's say there's a group and they don't know who could do research with them, and maybe there's a graduate student somewhere who is really interested in crowdsourcing projects, so there would be this kind of exchange.

Neil Fraistat, MITH, University of Maryland

Another idea was that there could be add-ons

in the form of workshops to already established conferences.

Mary Flanagan, Tiltfactor, Dartmouth College

We also talked about the benefits of having more conversations moving us forward. We have had a great couple of days and speaking for myself, and I hope for you as well, there have been many moments of transformation, seeing my own projects differently, seeing structures differently, seeing the kind of values that people are talking about very differently. It was great and moving, thinking about how to further our work and better meet our commitments to our people.

We plan on applying for a start-up through the Laura Bush 21st Century Librarian Program to try to move this forward, but there are also other funding possibilities, such as an NEH Institute. We are also thinking about Preservation and Access because we think that's an untapped field for some of the folks in the room. And we talked about FEDLINK, which I know very little about though some folks in the room do, and coming up with a sustainable funding model that doesn't just rely on grants forever. How do we do this, and how do we also do it through time? No one wants to join a consortium where there is no kind of tiered access. We have to figure out how it's really available first, figure out parts of budgets, and how we can all contribute some of these resources, videos, trainings, etc.

Securing the Required Resources

• I am putting a question mark by one statement Mary made: "Nobody wants to join a consortium that doesn't have tiered levels of access." **• Alexis Rossi, Internet Archive**

• What I meant by that is I don't think we should say, "Here, pay $4,000 everybody." There is no membership fee. **• Mary Flanagan, Tiltfactor, Dartmouth College**

• But for a million dollars we'll name the consortium after you. **• Neil Fraistat, MITH, University of Maryland**

• But even if there is not a membership fee, I'm still thinking about how we could have resources. Not necessarily, "You owe $4,000 to be part of this community," but, "Will you pay for development of content over here, and you can do this other piece over here." But somebody at the center has to organize what those contributions are coming into it, whether it's money or people or something else. • Perry Collins, National Endowment for the Humanities

• Exactly. I think the hardest thing for us is that we are so busy in our own areas and we also have other commitments. Generally all of us have at least two jobs probably, if not three. And the idea that we can set aside funding from a desperate project and put it into this consortium is really challenging. So we have to figure out how we can co-brand, collaborate, and then possibly offer value to libraries and institutions that may be willing to collectively support this work. We hope they are willing to collectively support this work. **• Mary Flanagan**

Prioritizing the Proposed Design Prototypes

TALLY PROCESS

Matthew Manos, verynice.co

For the next step in this process, you will not be held to this, but if there was something you heard today that you felt inclined to be a part of or wanted to support in some way, put your name up on the board underneath the related challenge. You each will have two votes. Your first choice is with a black pen, second choice is with a red pen. Again, you won't be held to this. We are trying to gauge interest in these different conversations that we have been having.
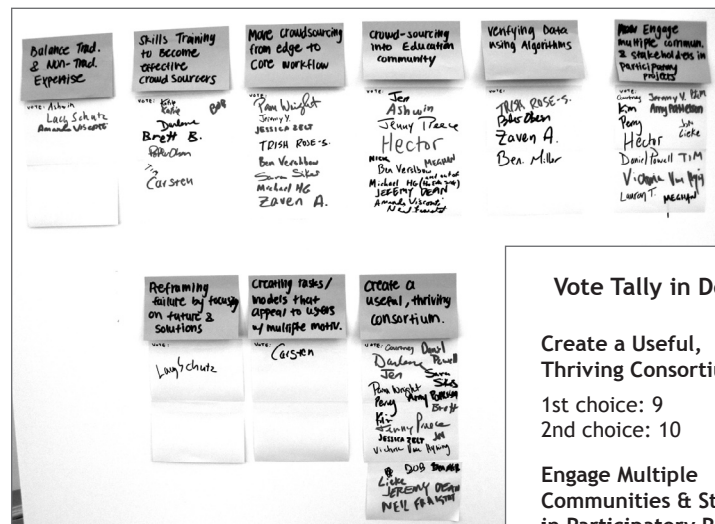
The vote tally: raw (above), descending priorities (right)

Vote Tally in Descending Order of First Choice Votes

Create a Useful, Thriving Consortium

1st choice: 9
2nd choice: 10

Engage Multiple Communities & Stakeholders in Participatory Projects

1st choice: 8
2nd choice: 4

Crowdsourcing into Education Community

1st choice: 8
2nd choice: 3

Skills Training to Become Effective Crowdsourcers

1st choice: 4
2nd choice: 4

Verifying Data Using Algorithms

1st choice: 3
2nd choice: 1

Move Crowdsourcing from Edge to Core Workflow

1st choice: 2
2nd choice: 6

Balance Traditional & Non-traditional Expertise

1st choice: 1
2nd choice: 2

Reframing Failure by Focusing on Future & Solutions

1st choice: 1

Creating Tasks/Models that Appeal to Users with Multiple Motivations

1st choice: 1

Ben W. Brumfield @benwbrum · May 7
I'll bet we can finish the @crowdconsortium white paper on crowdsourcing tools at #crowdcon, but I need your help! docs.google.com/document/d/1QL...

Google Docs

http://tinyurl.com/pt7acwb

Daniel Powell @djp2025 · May 8
Does #crowdcon think that pushing for collaborative authorship, participatory scholarship, within orgs would help overall mission?

Courtney Young @librarycourtney · May 8
Great to meet @jonvoss in person. #crowdcon

Daniel Powell @djp2025 · May 8
So, from #crowdcon : awareness of crowdsourcing platforms & skills; large org for oversight, sharing, coord; engaging multiple publics

Jim Salmons @Jim_Salmons · May 8
@changemakers #Play2Learn - Gathering about over, but search #crowdcon for lots of useful crowdsourcing ideas! :D

SciStarter @SciStarter · May 8
Scistarter (PPSR core source) is aggregating tools, events, community spaces + crowdsourced projects. @tsuomela #crowdcon

Todd Suomela @tsuomela · May 8
#crowdCon 2015-Tools, Platforms, & Projects Directory - Google Sheets bit.ly/1AJ7bwF

http://bit.ly/1AJ7bwF

Thanks & a Call for Ongoing Involvement

• I want to thank Bob Horton, who was formerly at the IMLS and provoked this whole meeting in the first place on the IMLS side. I want to then thank Brett Bobley and Perry Collins, who provoked the NEH involvement and funding support on their side. Thank you all for coming and taking your time. It was fun to work and play with you. If you are interested in putting in more energy than your average citizen scientist and humanist behind this effort, we welcome your help to get something rolling really fast. • **Mary Flanagan, Tiltfactor, Dartmouth College**

• And I would like to thank Josh Greenberg, one of our program officers, who had an internal meeting he had to go to at the Sloan Foundation and couldn't be here with us. • **Neil Fraistat, MITH, University of Maryland**

Dr Meghan Ferriter @MeghaninMotion · May 8
Thank you @crowdconsortium team + @criticalplay @fraistat @AndreaWiggins for leading catalyzing & connected thought this week! #crowdcon
↩ ♺ 2   ★ 10   •••

Amanda Visconti @Literature_Geek · May 7
#CrowdCon has started a spreadsheet of crowdsourcing tools. Add yours: docs.google.com/spreadsheets/d...
↩ ♺ 7   ★ 9   •••

http://tinyurl.com/paaqa85

FINAL THOUGHTS, NEXT STEPS

From Priorities to a Collectively Built Crowd Consortium

Neil Fraistat, MITH, University of Maryland

Maybe we should start by looking at the prioritized board [*page 137*] because there are some things that clearly have a lot of support. I think what we are seeing is "Creating a Useful, Thriving Consortium" is an overwhelming priority, but we are also seeing "Move Crowdsourcing from Edge to Core," we are seeing "Crowdsourcing into Education." There are actually a lot of cramped signatures at the bottom, so you may not see how many names are there, but that is considerable and, for me, really encouraging. "Verifying Data Using Algorithms" got some uptake, "Engaging Multiple Communities & Stakeholders in Participatory Projects" got more. And then "Skills Training to Become Effective Crowdsourcers."

That seems to be a pretty coherent curriculum, one that can be driven by a consortium, and one for which there is a lot of buy-in from this community. Now one thing that we know for sure is that this project that we are doing will have an end. It has been funded and we are going to have a wonderful detailed report of our findings collectively to share with you and the rest of the world. But there will come a point at which there is no funding support

until we get new funding support. That will mean people who believe that these things are important, and now we know who you are and you can expect to get tapped: Do you want to go in on a grant for this or that? And especially at the beginning it may mean that the Crowd Consortium exists only or mainly in the actions of the individuals who are bringing multiple members into some kind of project that they want to do. So there is organizing to do and it will depend on you and others who care about these things to do it.

Potential Gatherings and Next Steps

Mary Flanagan, Tiltfactor, Dartmouth College

Thank you so much for hosting. I like Neil's energy here. What is great is that we have something to take seriously and we have some work to do. I am just really heartened that people in the room will help follow through on those things and I know who to call for certain things and you know who to call now as well.

I think the next steps are for us to think about what is next, besides the web resources and repositories and all of the things we've talked about today, which are really rich and interesting. When do we three meet again? In what guise? Maybe it goes into different professional meetings or what have you, but I have had several conversations about sending technical people and having just a technical meeting where people are building stuff, hacking stuff,

doing code reviews. I'm not exactly sure what that would look like, but maybe that is the next in-person gathering, and I could certainly use that, it would be great. If that is a priority for us as a group it would be great to work on that.

If you have other priorities please tweet them out. We are still listening. The Crowd Consortium website will still be up, it's not going away. I am planning on maintaining it for the foreseeable future until we figure out how that works. Maybe there are people who could help draft an MOU, for example. That's not on the big picture list, but it's something I don't know how to do and could use some help.

Neil Fraistat, MITH, University of Maryland

We also might be asking this group for people who want to be involved in trying to coordinate at the larger scale. There is no reason why the three of us have to be the only ones, or even *the* ones. If you are interested in that, please make your interests known.

Outcome: Collaborations

Andrea Wiggins, University of Maryland

This has been about generating ideas for concrete outcomes, rallying the crowd, "I think we can do this," but there are other kinds of concrete outcomes from a meeting like this that take a little more time to show up and are a little harder to detect. I would like to see a show of hands from anyone who believes that they will be collaborating with someone in this room on a new project in the future. [*Most of the people in the room raise their hands*.] That is what I wanted to hear. Excellent, thank you.

Kudos for Conference Support

• I want to thank Everett White for video recording our proceedings, Amanda Visconti for tweeting so well and so often, and Porter Olsen for video streaming. • Neil Fraistat, MITH, University of Maryland

> **Jennifer Serventi** @JenServenti · May 8
> Yes, yes, thank you for the #crowdcon media team. It was really helpful to be able follow along from afar. *clap, clap, clap*

> **Daniel Powell** @djp2025 · May 8
> #crowdcon @SSHRC_CRSH is also quite interested in multinational partnerships, especially in digital humanities work.

> **Amanda Visconti** @Literature_Geek · May 8
> .@NEH_ODH & @chnm One Week | One Tool (#owot) a good model for future #crowdcon tech work sprint/barn-raising?

> **Daniel Powell** @djp2025 · May 8
> #crowdcon What can we take away now, to our home institutions, to encourage participatory knowledge creation?

> **CCLA** @crowdconsortium · May 8
> Next steps: technical experts meeting, hack/building/code review meeting? #crowdcon

Funding Potential: International and Federal

• There are people from different countries here and you've talked a little bit about funding. I wonder if anybody has thought about cross-country funding. We've also got people from funding agencies here that may need some nudges that this would be something we need to deal with. • Jenny Preece, University of Maryland

• I work at the NEH and I have agreements with many other grant funders from around the world, so if you want to shoot me an email and let me know the name of your humanities funder in your country, I will let you know if I know them. I'd be happy to give them a call, give them a heads-up. I can't promise they'll give you a grant, but I'd be happy to help. • Brett Bobley, National Endowment for the Humanities

• One other thing is that we heard yesterday that NIH is interested in funding even people from the humanities domain who are working in computational ways with these problems. It would be great if one of more of you think that you might have something that would work for them. I would encourage you all to look at their calls because we are being specifically solicited, and why let that opportunity go to waste if you are doing work that really fits? • Neil Fraistat, MITH, University of Maryland

• I think the deadline is June 3, but if you're ready to have a panic attack, I think it's worth it. • Mary Flanagan, Tiltfactor, Dartmouth College