



Arts Access Datathon Final Report and Toolkit

[Scott Reed](#), Datathon Archivist

Executive Summary

On Saturday, April 22, 2017, nearly 150 participants, speakers, volunteers, and observers assembled at The Reef, the Downtown Los Angeles campus of Sotheby's Institute of Art at Claremont Graduate University, to spend the day exploring how data can be used to improve access to the arts for all ten million residents of Los Angeles County. The Arts Access Datathon was co-hosted by the City of Los Angeles Department of Cultural Affairs and the Los Angeles County Arts Commission, with major support from Sotheby's Institute of Art at Claremont Graduate University and Arts for LA.

The Arts Access Datathon brought together artists, arts administrators, educators, students, community advocates, and researchers, as well as professionals in culture, museums, and urban planning to not only learn about what datasets exist, but also to begin to manipulate, visualize, and interpret the data. While some attendees possessed skills in working with data, this was not universally true. The structure of the event and related resources had been designed with this in mind, in order to accommodate the wide range of skills and knowledge expected.

Event organizers assembled participants into eleven groups that balanced knowledge and skill sets so that each group would have at least one artist, an arts administrator, a data-savvy person, and a technologist.

A series of presentations from subject matter experts and technologists oriented participants to the datasets and tools. Groups were given opportunities after each set of talks to explore and select resources that they wanted to use to address issues revolving around access to the arts in Los Angeles.

The groups spent three hours working on their individual projects collaboratively. Because of the limited time and varied skill levels working with data, the intention was to present clear and insightful uses of pre-existing data or to propose new projects based on critical engagement with the data. The panel of judges represented a range of arts- and data-related expertise. To select the winners, the judges scored projects based on 1) the likelihood they would improve access to the arts, 2) their use of data, 3) feasibility, and 4) creativity and innovation.

Two groups tied for first place. Each designed prototypes of apps/services built on available data. One prototype would make literary arts more widely available. The second would help public schools find skilled teaching artists in their area. The group that came in third place analyzed multiple datasets in an effort to build a better model for understanding arts nonprofits in light of neighborhood change.

The goal of the Arts Access Datathon was to bring together a diverse group of stakeholders and practitioners to explore how data could be used to improve access to the arts in Los Angeles. Survey feedback reflected the exploratory design of the Arts Access Datathon. One participant noted that, “learning about public datasets was something that [they] had no idea the public had access to,” while another attendee declared their newfound familiarity with datasets would allow them to “leverage existing data to be [a] better storyteller.”

Participants also offered suggestions for improvement. In particular, several expressed a desire for a more focused or pre-defined problem, as well as additional time for groups to flesh out their proposals. One respondent suggested narrowing “the scope of the ambition, perhaps giving some suggested topics to explore.”

What follows is a detailed report of what happened throughout the day at the Arts Access Datathon. Also included is a toolkit outlining the planning process the Los Angeles County Arts Commission and the City of Los Angeles Department of Cultural Affairs undertook, including several key decisions along the way. Our hope is that the toolkit will be useful for organizations or individuals interested in holding a datathon as a way to improve both the knowledge and use of data in their fields.

Arts Access Datathon Final Report

[Scott Reed](#), Datathon Archivist



Laura Guerrero-Nieto (left) and Brandon Turner (right) check in participants for the Arts Access Datathon at The Reef, April 22, 2017.

Introduction

On Saturday, April 22, 2017, nearly 150 participants, speakers, volunteers, and observers attended the Arts Access Datathon hosted by [Los Angeles County Arts Commission](#) and the [City of Los Angeles Department of Cultural Affairs](#). The event took place at [Sotheby's Institute of Art at Claremont Graduate University](#) in the [Reef Building](#). This report summarizes the proceedings of the event and is followed by a toolkit for others who might be interested in organizing a datathon of their own.

The Arts Access Datathon brought together artists, arts administrators, educators, students, community advocates, and researchers, as well as professionals in culture, museums, and urban planning to address how to use data to improve access to the arts for all 10 million residents across LA County, including the 88 cities and

unincorporated areas within it. The event sought to not only educate about what available datasets existed, but also how to access them and what tools were available to manipulate, visualize, and handle data.

A datathon is similar to a hackathon, typically understood as a competition in which groups solve a problem or build technological tools in a limited time period, usually over the course of a weekend or several days. The final product of a hackathon is often a software application prototype judged by a panel. Hackathon teams are interdisciplinary and possess software scripting ability to “hack” or develop new apps. This datathon was different in that it presumed participants would have a wide range of experience and engagement with data, from significant to none, and it did not assume the technical ability to produce new tech tools or in-depth analysis. One goal of the Arts Access Datathon was to bring together a diverse group of stakeholders and practitioners to engage with data as a way to explore the potential for addressing issues of access to the arts in LA. Final projects for the datathon did not have to be software applications, but could be proposals for projects, campaigns, reports, or policies based on available data. Similar datathons have been organized for [archivists](#) and the [environmental science community](#).

A core goal of the Arts Access Datathon was to educate stakeholders about open data related to the arts in Los Angeles - what exists, how to use it, and what tools and questions might be applied to the data. While some attendees possessed skills in working with and manipulating data, this was not universally true. The structure of the event and related resources did not assume expertise.



Group 1 members (left to right): Patrick Varon, Nisa Karnsomport, Mark Seldis (back turned), Cynthia Luján, and Tahnee Cadrez.

Working Groups

In order to facilitate access to data and point towards useful tools to work with the data, the [Arts Datathon website](#) served as an educational portal to explore curated datasets related to the arts community from a variety of local and national sources. After the event, the website continues to serve this informational purpose. It includes downloadable datasets, a list of dashboards, maps and databases that visualize or make data interactive.

Datasets on the website were carefully chosen. Contextual information and metadata were provided to help participants understand the context of each dataset and how to use them. Information such as periodicity, geography, format, owner/author, and license were provided.

Examples of arts related data available on the site include [Grantees of LA County's Organizational Grants Program](#), [Bookstores in LA County](#) and [Free Concerts in LA County Public Sites](#) compiled by the LA County Arts Commission. Other datasets were collected from the City of LA's Department of Cultural Affairs ([Community Art](#)

[Centers, Theaters, and Historic Sites; and Organizational Grantees from the City's Cultural Grants Program](#)), Los Angeles Unified School District ([The Arts Equity Index](#)), Western States Arts Federation (WESTAF) ([Creative Occupations in LA County](#)), DataArts ([Event Attendance for Arts Nonprofits in the US](#)) and Curate LA. Dashboards and interactive visualization tools are also listed, including the [National Center for the Arts KIPi Dashboard](#) and the [California Arts Education Data Project](#) hosted by CreateCA. Many of the available data were sourced from the [LA County Open Data Portal](#) and the [City of LA's Open Data Portal](#), which serve as data clearinghouses for a wide variety of County and City departments.

All the information resources were compiled into interactive and downloadable tables hosted by AirTable, a web-based data platform. These interactive tables also contained metadata useful for sorting and categorizing the various datasets. The [Data Tools](#) page on the website outlined options for analyzing datasets using a variety of methodologies including visualization and infographics, mapping, scripting, and textual analysis.

Finally, a [Github repository](#) served as the “statement of record” for data used during the event. That means it is a stable source for understanding the data as they were presented to participants, a “snapshot in time” capturing data versions at the time of the event even as the datasets themselves will have changed over time through their source locations on the web.



Participants Sylvia Aguinaga (left) and Susan Chun (right).

Datasets

After checking in, networking, and eating breakfast, participants were assigned to numbered tables. These would be their working groups for the rest of the day. Before the group work began, the day kicked off with a series of welcome addresses from leaders in the arts ecology: Jonathan T. D. Neil (Director, [Sotheby's Institute of Art at CGU](#)), Danielle Brazell (General Manager, [City of Los Angeles Department of Cultural Affairs](#)) and Laura Zucker (Executive Director, [Los Angeles County Arts Commission](#)).

Bronwyn Mauldin, Arts Access Datathon co-organizer and Director of Research and Evaluation at the LA County Arts Commission, introduced the format and structure of the day. Sunil Iyengar, Director of Office of Research and Analysis at the [National Endowment for the Arts](#), launched the Arts Access Datathon with a keynote address.

Artist and performer Luke Kanter led groups through team building exercises that included movement and story creation as ice-breakers. Kanter was the first of two artists who bookended the event, serving as a reminder of the purpose of the Arts Access Datathon as well as to infuse the day with creative energy.



Ruby Rios participates in Luke Kanter's warm up exercise.

Participants were inspired to consider pre-existing datasets in their inquiries for later in the afternoon. To help facilitate this, a group of data experts presented on the current work and opportunities for engaging with data:

1. Michelle Higgins, Community Engagement Manager, DataArts [[Slides](#)]
2. Matt Agustin, Research Coordinator, Arts Education, LA County Arts Commission, "Data About Arts Education" [[Slides](#)]
3. Yvonne Lee, Datathon Co-Organizer and Registrar, LA County Arts Commission, "Public Art Collections" [[Slides](#)]
4. Zannie Voss, Director, National Center for Arts Research [[Dashboard](#)]
5. Wendy Hsu, Datathon Co-Organizer and Digital Strategist for the City of LA Department of Cultural Affairs, "Social Media as Arts Data" [[Slides](#)]

These experts were available throughout the day to assist the participant groups in utilizing datasets and tools to develop their proposals.



Michael Manalo demonstrates features on Socrata's Open Data platform.

Data Tools

It's one thing to know your data, but what can you do with it? The next series of presentations addressed accessible tools for creating visualizations and analyzing datasets. Starting with perhaps the most common (and some argue most powerful tool), Susannah Laramie Kidd, Research Analyst and Mellon/ACLS Public Fellow with the LA County Arts Commission presented on data visualization in Excel.

[“A Guide to Spreadsheets for the Spreadsheet Phobic”](#), a zine created by Bronwyn Mauldin, was also distributed, focusing on the core elements of spreadsheets and their power for working with data. The approach was particularly suited for artists. The zine encouraged the reading and manipulation of spreadsheets like the creation of a story, expanding the concept of rows, columns, and cells into the raw materials for compelling data analysis.

Mike Manalo presented on [Socrata](#), the platform that both the City of Los Angeles and the Los Angeles County governments use for their Open Data portals to publish datasets and provide in-browser visualizing features that generate maps and charts based on user-defined parameters. Katja Krivoruchko of Esri presented on the functions of [ArcGIS online](#). These experts were also on-hand throughout the day.



Katja Krivoruchko of Esri (right) assists Participant Reina Imagawa (left).

Data Projects: Creation

The eleven groups spent roughly three hours working together on their individual projects, pooling their collective experience, expertise, interests, familiarity, and comfort with data along with their newly acquired skills in the world of available datasets and tools presented in the morning. The groups' final products took one of many forms. Because of the limited time and varied skill sets working with data, the groups were encouraged to present clear and insightful uses of pre-existing data or to propose new projects based on critical engagement with the data.

Val Zavala of [KCET television](#) emcee'd the final group presentations. The judges panel represented a variety of arts and data related expertise and was comprised of Sophie Fanelli ([Stuart Foundation](#)), Nina Kin (County of Los Angeles and [Hack for LA](#)), Mike Manalo ([Socrata](#)), Jill Moniz (independent curator and educator), and Zannie Voss ([National Center for Arts Research](#)).



(Left to right) Susannah Laramée Kidd and Bronwyn Mauldin (standing) with judges Nina Kin, Jill Moniz, Michael Manalo, Sophie Fanelli, and Zannie Voss.

To select the winners, the judges scored proposals using the following criteria:

- Likelihood of improving access to the arts
- Use of data
- Feasibility
- Creativity and innovation

The three winning groups represented diverse approaches and methods for engaging with data. Two of the groups designed apps/services built on existing data, proposing new information pipelines useful for locating arts services in Los Angeles. Another group analyzed multiple datasets in an effort to create a better model for understanding arts nonprofits in light of neighborhood change and proposed new research strategies based on lessons learned during their group work. Visuals from the group presentations are available to view online.

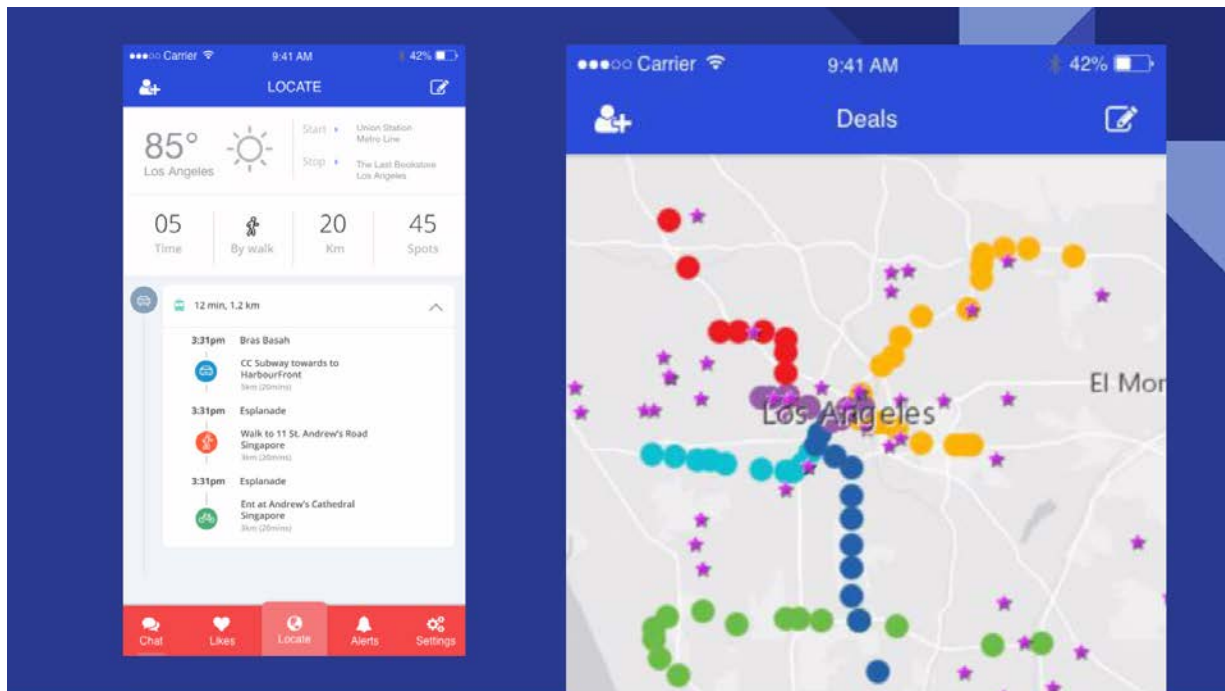


Image: Prototype of the “Book Line” application, among the winners of the Arts Access Datathon competition.

Data Projects: Presentations

Group 1: “It’s All About the Purple” (tied for 1st place)

Group 1 developed “Book Line,” a mobile application design proposal for mapping bookstores along Metro lines. With the expansion of Metro through the passing of Measure M and the abundant but perhaps not well-known supply of bookstores in Los Angeles, the app would help create hubs of interaction and commerce around books. Beyond just connecting consumers and bookstores, “Book Line” could also promote literary focused events and public interactions, including mini libraries in and around Metro stations and public transportation hubs. The design relied on the [Bookstores in LA County](#) dataset, [Metro Transit Project Lines data](#) hosted by the City of LA’s Geohub. The group built a visual prototype of the app using Adobe XD. [[View presentation](#)]

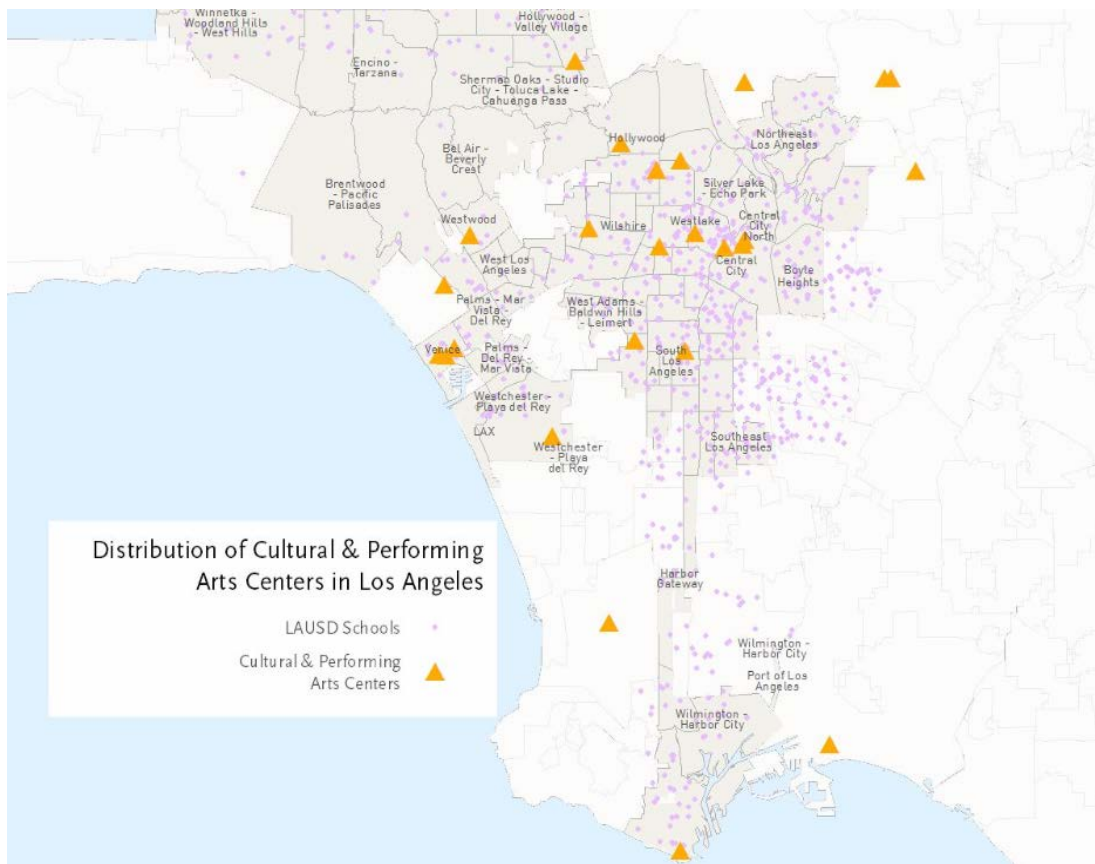


Image: The “Bazaar” app is based on this spatial analysis of local needs of schools for arts education.

Group 11: “Bazaar” (tied for 1st place)

Group 11 proposed “Bazaar,” an online platform connecting schools and teaching artists using data related supply-and-demand for arts services. Bazaar can serve a “two-sided” market that helps schools find organizations and individuals to teach art to their students. Group 11’s proposal utilized data on [Cultural Centers](#) in the City of LA, [Community Arts Partners serving LA County Public Schools 2012](#), demographic data, and the [Arts Equity Index](#) from LA Unified School District. [[View presentation](#)]

	zip code	nonprofit orgs	nonprofit attendance	wage average	2015 jobs	dca 14-15	dca 10-11	Change index	dca change
zip code	1.0000								
nonprofit orgs	-0.3333	1.0000							
nonprofit attendance	-0.1975	0.8258	1.0000						
wage average	-0.2392	0.1887	0.1665	1.0000					
2015 jobs	-0.1549	0.3183	0.2706	0.4246	1.0000				
dca 14-15	-0.2136	0.9035	0.8719	0.0941	0.2068	1.0000			
dca 10-11	-0.2507	0.7341	0.6955	0.1418	0.3198	0.7470	1.0000		
Change index	-0.3646	0.4283	0.1854	0.0182	0.0314	0.2929	0.2776	1.0000	
dca change	-0.1538	0.8184	0.7964	0.0516	0.1091	0.9396	0.4742	0.2450	1.0000

Image: Visualization by Group 7 exploring correlations between neighborhood change and nonprofit arts organizations.

Group 7: Nonprofits and neighborhood change

Do the available data prove the prevalent story of the relationship between artists and gentrification? While the group did find positive correlations between arts grants and neighborhood change, the correlation was not as strong as they expected. Moving forward, the group proposed investigating outliers, i.e. neighborhoods that experienced significant change without arts and those with arts that have not experienced similar change. With these additional analytics and modeling, the group proposed that better and more responsible arts policy might be developed to bring arts resources to neighborhoods while not contributing to displacement or other symptoms of gentrification. The group utilized data from the [Los Angeles Indices of Neighborhood Change](#), as well as [arts nonprofit event attendance data](#) accessible from DataArts, [Organizational Grantees of City of Los Angeles Cultural Grants Program](#), and creative industry employment data from the [Creative Vitality Suite](#). [\[View presentation\]](#)

Remaining eight group presentations

While only three groups were selected as prize winners, all of the groups demonstrated critical engagements with the data, achieving the goals of the Datathon in the short amount of time allowed. Projects, proposals, and questions presented by the groups included:

- An asset-based community development project focused on Monterey Park.
- Proposal for future research based on a preliminary spatial analysis of arts resources across LA County, addressing the needs for integrating lived experience of the arts in local communities and doing multi-lingual analysis of informal conversations about the arts on social media.
- Analysis of the barriers to entry for public art.
- Analysis of arts participation based on the availability of art in specific geographic regions/neighborhoods in LA County.
- Investigation of access to arts outside of formal institutions by analyzing available data and considering what is not currently being captured and categorized (example: DIY art communities).
- A study of how neighborhoods vulnerable to gentrification are using the arts to address solutions.
- A study of how early arts education impacts participation and leadership in the arts and promotes overall diversity and cultural relevance of arts programming.
- Analysis of two Los Angeles schools similar in demographics, one with limited access to arts education and one that is an arts magnet school. They found the school with more arts education graduated more students that were college ready, suggesting new policy and programs to engage students early on and track/understand their future participation in the arts.

Organization Grants For Arts Services

discipline | Music-Choral/Opera

grantee name | Long Beach Chorale & Chamber Orchestra Inc.

grant award | 16600

application year | 2015-16



Sound artist Inouk Demers created a site- and event-specific installation utilizing some of the arts datasets participants engaged with

Closing

As participants exited the Reef, they passed through sound artist Inouk Demer's [event-specific installation](#). Demers sourced datasets used in the datathon to create map visualizations and electronic music. In his [own words](#): "Datasets were produced by researchers, arts agencies, educators, and community advocates in order to improve access to the arts within LA County. The content of these datasets (GPS coordinates, data on museums, ethnicity, gender, events) was extracted during the performance and has a real-time effect on the 3D video map and the sonic parameters (distance, filter, bit depth, etc.) of the ambient music itself."

Feedback

A survey was sent to all participants asking for feedback. Their comments were generally positive, while also offering suggestions for improvement. Many comments reaffirmed that an educational perspective geared towards data novices helped create an empowered community of practice. One participant noted, “...learning about public datasets was something that I had no idea the public had access to. I think promoting those datasets is important to maintain a free and informed society.”

The educational focus was appreciated by many: “The presentations held in the first half of the event were very helpful in seeing the variety of projects creative professionals are developing through different datasets.” Another attendee noted that the newfound familiarity with datasets would allow them to “leverage existing data to be a better storyteller.”

Participants also offered suggestions for improvement. This Arts Access Datathon was organized around the goal of improving access to the arts for everyone, and as such allowed for broad participation and flexibility in the specific topics addressed by participants. Additional time for groups to flesh out their proposals and a more focused or pre-defined problem to be tackled by the groups might be considered for a future iteration.

In addition, organizing the event around pre-assigned small groups contributed to building diverse project teams. One reviewer noted they “enjoyed meeting diverse members of the arts field.” Another commented they “felt very inspired and full of ideas,” but would have liked “a chance to meet others who were there, not just those at [their] table.”

This report and the toolkit that follows are presented as tools that may be useful for anyone interested in holding a datathon as a way to improve both the knowledge and use of data within their fields.

Acknowledgements

Photos courtesy of Alexia Lewis, www.krafted.la

Notes on Presentations provided by Susannah Laramée Kidd, PhD, Research Analyst and Mellon-ACLS Public Fellow 2015-17, LA County Arts Commission

Licensing/Copyright

This document, corresponding photos, and the Arts Datathon website contents are licensed under Creative Commons Attribution-NonCommercial 4.0 International ([CC BY-NC 4.0](https://creativecommons.org/licenses/by-nc/4.0/)) License. Licensing for each dataset listed on the Datathon website is specified in the [Airtable](#).

Appendix/Resources

1. Datathon Website

Datasets: <http://artsdatathon.org/data/datasets/>

Data Tools: <http://artsdatathon.org/data/data-tools/>

2. Github: <https://github.com/dcadigital/Arts-Datathon>

3. Photographs, courtesy of Alexia Lewis (www.krafted.la) <https://app.box.com/s/kscjzf29yphntrv7qrpmw4ix7piqnm90>

4. If your organization is interested in hosting a datathon, please take a look at our own toolkit as well as these other toolkits we referenced while planning our event:

- Ouishare Fest's Open Source Toolkit for festival or conference organization: <https://www.gitbook.com/book/ouishare/ouishare-fest-toolkit/details>
- The Signal (Library of Congress blog) post on co-hosting a datathon: <https://blogs.loc.gov/thesignal/2016/07/co-hosting-a-datathon-at-the-library-of-congress/>
- Archives Unleashed, a website documenting a series of datathons including the event co-hosted by the Library of Congress: <http://archivesunleashed.com/>
- "Reflections of a data hack judge" by Andrew Lewis: <https://lab.sciencemuseum.org.uk/reflections-of-a-data-hack-judge-555d1e3a2825>

Arts Access Datathon

Datathon Toolkit

[Scott Reed](#), Datathon Archivist

This toolkit is intended to provide resources, recommendations, and communication templates for organizing a similar datathon, based on our experience with the Arts Access Datathon that took place on April 22, 2017. It covers the following topics:

1. Planning and managing the event
2. Compiling datasets
3. Designing the contest
4. Archiving and preserving event materials, media, and datasets

1. Planning and Managing the Event

The Arts Access Datathon leadership group began planning for the event approximately a year in advance. The core leadership team consisted of three (later expanded to four) members from the two co-hosting organizations. An informal memorandum of understanding (see a generic [MOU template here](#)) was drafted that outlined each organization's roles and responsibilities. We used a combination of Google Docs and Sheets to coordinate and document our planning process.

Steering Committee

The leadership team brought together a diverse group of stakeholders to serve as a steering committee, as well as a larger advisory committee. Diversity included race and ethnicity, gender, age, seniority in their field, LGBTQIA identity, and geography. In addition, we were careful to include people with different levels of knowledge and perspectives on the arts, access, and data. We intentionally included people with very limited formal experience working with data. We also made sure to include artists. This diversity of viewpoints and experiential backgrounds at times slowed our planning process as we built a shared understanding of what we sought to achieve, but the benefits made it worthwhile. It deepened our work and helped us to ensure an event that would be relevant and welcoming to a broad range of people.

Steering committee members met on a monthly basis, providing leadership and resources for the event. The advisory committee was loosely assembled via email and their attendance was not mandatory at any point. Steering committee members were primarily engaged in planning as well as securing in-kind resources and volunteers.

Registration Fee

After much discussion among the steering committee, the steering committee decided to charge a nominal fee to attend. Significant financial support for the Arts Access Datathon came from Sotheby's Art Institute at Claremont Graduate University, which also provided space for the event at no charge. Arts for LA provided support for registration and labor. While we secured other donations of funds, goods, and services, we did not seek the kind of major corporate underwriting that often covers costs for hackathons. The registration fee helped cover a fraction of the event costs. However, our biggest reason for charging a registration fee was to reduce the number of no-shows and to increase the likelihood that people would stay for the full day. We created a tiered pricing system:

- Students and Artists - \$10
- Early Career Professional- \$20
- Professional - \$40
- Executive - \$80

We allowed participants to select the category that was right for them and did not ask for proof of the category they chose. We made it clear in all of our materials that no one would be turned away for lack of funds. Everyone who requested a scholarship was invited to pay the student/artists rate if they could afford it and given the option to register for free.

Volunteers

During the planning process, we developed a list of volunteer tasks needed. These tasks ranged from set up and break down, staffing registration table, tech support, and time-keeping for the speakers and presentations. This list helped determine how many volunteers were needed. Many volunteers were staff from host and donor organizations. Some steering committee members were volunteers. Other volunteers were sourced from a general call to the arts community. In total, 20 people, in addition to the core leadership team and speakers, volunteered that day.

Other Key Decision Points

Other key decision points in the planning process included:

- Selecting the topic and articulating the goals of the event
- Identifying the intended participants
- Program
 - Length of time
 - Ensuring people are engaged throughout

- Event date
- Speakers
- Location
- Budget and donors
- Datasets (readily available and wish list)
- Communications and marketing plan
 - Rollout
 - Logos and messaging
- Prizes
- Technical aspects
 - Website
 - Registration site
 - Dataset repository
 - On-site Internet connectivity (connectivity speed of over 100MPS recommended)

2. Compiling Datasets

Both the [County](#) and [City of Los Angeles](#) utilize an existing infrastructure to publish and share datasets through Socrata's Open Data platform. Other local governments and agencies are starting to publish raw data beyond what is presented in flat reports, but this is not always the case. No matter the topic of your datathon or data-driven event, think broadly about the issues that affect the ecology in which your topic exists. For the issue of arts access, we sought to be inclusive in how we defined arts and access. Therefore, we included datasets about library circulation, bookstores in LA County, public art, arts education, funding, demographics, and arts organizations.

The next step was to reach out to the authors or keepers of the datasets identified as relevant to our topic. This was a great opportunity to talk about the upcoming event and gauge interest and expertise on related fields. Many of our advisory committee members and some speakers were discovered during this process of compiling datasets. We discovered some nonprofits and private companies were willing to share aggregated datasets with us, as part of their outreach and support to the field.

Once we secured the right person to grant access or permission to the data, we discussed use. Often, the dataset we wanted was behind a paywall or needed some degree of cleaning and anonymization. We made the decision early on to only include datasets that the owners were willing to make available for free indefinitely,

so others who were not able to participate in the live event would be able to replicate or retrace the work done during the Arts Access Datathon. While some of the datasets were already licensed, we recommended the Creative Commons Attribution-NonCommercial 4.0 International ([CC BY-NC 4.0](#)) License to all the owners of data previously unpublished.

We created several ways of presenting the data for all levels of participants. For the data novices, we created a [narrative-based introduction to the datasets on the website](#). In this narrative we included a description of what the dataset contained and what kind of analysis each one might be useful for with its limitations. For those familiar with data in tabular (spreadsheet) form, we embedded an [Airtable](#) into a web page. With sorting and filtering features, the Airtable contains data download links and metadata fields to indicate the geography, periodicity, licensing, owner, related topics, and description of each dataset. Finally, for programmers and developers, we created a [GitHub repository](#) for select datasets (uploaded as csv files).

3. Designing the Contest

When thinking about the goals of the datathon, we discussed what possible deliverables could be produced by the project groups in the course of an eight-hour event. Since we expected the audience of the first Arts Access Datathon to primarily be made up of non-data people working in the arts, we veered from the usual hackathon format of awarding big (cash) prizes. Instead our goal was to inspire people to think about arts data, i.e. what they are, where they are, how they are collected, and how they can be used. We also reminded ourselves regularly that while the datathon was our tool, the primary focus of the day would always be on improving access to the arts.

Datathon participants were briefed on datasets and data tools through a series of five to ten minute talks and encouraged to spend time going through each of the resources together in their groups. In all, the participants had three hours to develop a project proposal that they would then present to a panel of judges. The panel of judges was carefully selected to mirror the diversity of the arts ecology. It included artists, arts administrators, and arts funders alongside technologists and data experts from outside the arts ecology.

To help the judges with their task, we provided them with a [simple scoring rubric](#). They were asked to score each proposal on four key factors (the likelihood of improving access to the arts, use of data, feasibility, and creativity and innovation), with five points per factor. The judges did express some concern that they could not go back and change their scores after watching the first few presentations, but this

did not turn out to be an issue. There was general consensus from many after the event that the three winners were the best proposals.

There was some discussion within the leadership team about whether to award prizes or not. We decided that since our prizes were of nominal value, awarding them would not create undue incentives that might remove the focus from improving access to the arts. The prizes consisted of arts experiences (including public art tours, screenings, performance tickets), technical books, and gift certificates. Local arts organizations and tech book publishers donated most of the prizes.

4. Archiving and Preserving Event Materials, Media, and Datasets

Many data-rich resources, documents, and presentation files will be created during the planning and execution of any datathon. These may have historical value or use-value outside of the event itself. For example, the Arts Access Datathon website and presentations may be of interest to a variety of practitioners, researchers, and students interested in learning about or using open data. The goal of most archives is to provide access to users in perpetuity. In order to achieve this goal with digital files it is important to organize, contextualize, and securely store files as soon as possible.

Digital Archiving for Your Event

The most important aspect to remember when preserving and archiving digital resources is that they are not more stable or safer than physical resources (such as paper documents), and sometimes the opposite may be true. Technology changes at a rapid pace, and content is often lost from the internet or corrupted while stored on a hard drive or removable storage device. If your event generates media attention, social media activity, or maintains a significant web presence, it is also important to capture this while it still remains on the web. The following are some high-level considerations for digitally archiving a datathon event.

● File Formats

Avoid proprietary file formats when possible, and choose formats that can be accessed by a variety of operating systems. This means that if a file requires a specific piece of software to open, you may be jeopardizing your ability to access the file at a later date. For example, while a particular piece of software may be favored for accessing or manipulating a file by most users today, this software might become obsolete in a few years. In addition, users with a different operating system may be unable to open it today.

Recommended Formats statement from the Library of Congress:

<https://www.loc.gov/preservation/resources/rfs/TOC.html>

Resources for determining File Types:

<https://www.archives.gov/preservation/products/definitions/filetypes.html>

● Naming Files

Consistent and detailed file naming can provide necessary context and organization, particularly when local knowledge and memory of the event and its corresponding materials have been lost or forgotten. Never use spaces or special characters in file names, and include a date. While some file metadata might keep track of “Last Opened” or “Created” date, these can change often or might not reflect the preferred creation date (for example when migrating to a different file format).

Example of a bad file name:

Photo 145.jpg

Example of a good file name:

0145_Judges_Datathon_04222017.jpg

For the event, uploaded files to the GitHub repository utilized the following file naming convention: [Author/Owner-SubjectMatter-YYMMDDpulled].[file format]. For example, if the uploaded dataset is Los Angeles County Arts Commission's civic art collections data, the uploaded file was named: LACAC-CivicArtCollctn-170410.csv. As for abbreviations in the file name, these were left open so long as the abbreviation was explained in the Additional Documentation section.

Additional tips for file naming are available from Stanford Research Library:

<https://library.stanford.edu/research/data-management-services/data-best-practices/best-practices-file-naming>

● LOCKSS

An important concept to keep in mind when preserving digital materials is “Lots of Copies Keeps Stuff Safe” (LOCKSS). The acronym originated from [a peer-to-peer network](#) of the same name. You might keep your files highly organized in the most stable formats, but if they exist solely on an aging external hard drive you might be out of luck sooner than you think. Making copies of files and keeping them in geographically dispersed places (including

external drives, local repositories, and in the “cloud”) can help ensure long term access to files when one storage site is compromised or fails.

"The cloud" refers to digital storage accessed remotely through the web. Common cloud storage platforms include Box, Dropbox, Google Drive, and iCloud. Each service maintains physical servers in different geographic locations and will have varying storage and security policies, as well as terms of use. No matter what service you use, it is important to keep local copies whenever possible and refrain from storing documents with confidential or private information on the cloud.

● **Web Archiving**

The most well-known web archive is the Internet Archive and its [Wayback Machine](#). The Internet Archive uses automatic crawlers that create shallow “snapshots” of websites and then make accessible archived versions available online. Even if a website is taken offline, you can still browse archived versions of the site by inputting the URL in the Wayback Machine and selecting a capture date.

Not all websites are captured by the Wayback Machine automatically. As of May 2017, you can input an individual URL to be captured and included in the Wayback Machine using the “Save Page Now” feature.

Other tools exist to create web archives that can be downloaded and preserved along with other digital assets from your event. The most robust and accessible tool is [WebRecorder.io](#), a project of Rhizome/New Museum. Using their browser-based tool, you can record a website as you browse through it, and download a WARC file of the archive. WARC is the ISO standard for web archiving. Their [Webrecorder Player](#) can load WARC files from your computer offline to ensure the completeness of the archive.

Web archiving is a not a backup of a website, and cannot be used to rebuild a website if it disappears. In addition, if a website utilizes complex Javascript or other dynamic features, it is possible that not all functionality will be captured.

● **Datasets**

Datasets that your participants use will come from a variety of sources and are unlikely to be of the same file type. Perhaps the simplest and most stable format is CSV (comma separated values), which is ideal for moving data between tools. CSV files can be easily opened and manipulated by using

common spreadsheet software such as Excel, as well as reformatted into more functional formats such as GeoJSON.

Spreadsheets accessible online through such services as Google Docs and Airtable should also be downloaded as a CSV and stored offline.

The Arts Datathon website linked to various datasets hosted elsewhere, such as the [County of Los Angeles Data Portal](#). In addition to these links, we also maintained a GitHub repository that stored data in spreadsheet form. This served as a snapshot of the data used at the event. In addition, it can serve as yet another copy of resources from the event for preservation purposes.

Datasets will come in a variety of formats in addition to CSV. To learn about the various formats, the Format Descriptions for Dataset Formats from the Library of Congress is an excellent resource:

https://www.loc.gov/preservation/digital/formats/fdd/dataset_fdd.shtml

● **Create a Finding Aid or Data Manifest**

Archivists will often create a [Finding Aid](#) that indexes or inventories objects (in this case, datasets) to assist researchers in accessing materials in a collection. While a formal Finding Aid may not be necessary, a manifest of the various items/files you are preserving along with minimal metadata including creators, copyright, and descriptions, can help maintain organization and institutional memory around an event.

Automated Manifest creators can simplify the process when documenting hundreds or thousands of files. [Karen's Directory Printer](#) is a favorite among digital archivists.

For questions about the Arts Access Datathon, visit the website at artsdatathon.org, or drop us a line:

Bronwyn Mauldin
LA County Arts Commission
bmauldin@arts.lacounty.gov

Wendy Hsu, PhD
City of LA Dept. of Cultural Affairs
wendy.hsu@lacity.org