|  |  |  |  |
| --- | --- | --- | --- |
| C:\Users\bjaco\AppData\Local\Microsoft\Windows\INetCache\Content.Word\SLS-Teaching-Toolkit-Logo_Stacked-Initials.jpg | Civic Data Guides: Thinking Critically  about Digital Public Records | | |
| **Discipline:** All | **Type:** Take-home assignment; Project | **Time Commitment:** One week per module. Up to four. | **Category**: Civic Data |
| **Big Ideas:** [Social, Cultural, and Environmental Context](https://serve-learn-sustain.gatech.edu/big-idea/social-cultural-environmental-context) ; [Information Visualization](https://serve-learn-sustain.gatech.edu/big-idea/information-visualization) | | | |
| **OVERVIEW:**  Over the past decade public institutions have put considerable resources towards improving the quality and availability of civic data, such as budget and expenditure information, building permits, air quality readings, police incidents, and property ownership records. The agencies behind these efforts claim that data sets alone are sufficient to create transparency, increase civic engagement, foster innovation, and ultimately make our communities more sustainable. However, making civic data accessible does not necessarily make them valuable or actionable. To take effective and ethical action, we need contextual information about the processes involved in creating, managing, and interpreting civic data. In this modular, multi-day tool, you will create an accessible, practical guide to an existing civic data set. Working through the modules below can help you, and subsequently others, engage with civic data in productive ways.  This tool was contributed by Yanni Loukissas, with assistance from Catherine D’Ignazio and Bob Gradeck. | | | |
| **INSTRUCTIONS:**   1. Ask students to acquire existing civic data sets on their own or from a list of examples assembled for the purposes of your course. 2. Have students work individually or in teams to create their own civic data guides using the modules described below. Students can complete up to four modules, but should start with module A. Each module should take no more than one week to complete. 3. Encourage students to exchange their guides and evaluate the intelligibility of each other’s work in class or through a subsequent project. 4. See the end of this tool if you would like to incorporate community-based learning opportunities. | | | |
| **SLS STUDENT LEARNING OUTCOMES & ASSESSMENT:**  The Serve-Learn-Sustain toolkit teaching tools are designed to help students achieve not only SLS student learning outcomes (SLOs), but the unique learning outcomes for your own courses. Reflection, concept maps, rubrics, and other assessment methods are shown to improve student learning. For resources on how to assess your students’ work, please review our [Assessment Tools](http://serve-learn-sustain.gatech.edu/tool-category/assessment).  **This tool achieves SLO 1. See the end of this tool for further details.** | | | |

**Want Help?**

Yanni Loukissas is the contact for this tool. You can reach him at yanni.loukissas@lmc.gatech.edu

Civic Data Guides: Thinking Critically About Digital Public Records

Working with civic data—digital records created and made available for the public good—requires more than a spreadsheet. All data are entangled with people, places, technologies, and organizations that fundamentally shape their significance and use. Without proper context, users of civic data are ill-equipped to identify quality constraints, uncertainties, limitations, biases, and potential privacy and ethical issues that can cloud their analysis. However, there is no widely accepted formula for collecting this context. Furthermore, it’s not entirely clear that there should be, because data standards and their uses vary widely. Using this tool, you will explore what *you* think should be contained in such a guide. Your guide can be short and to the point or it can have many sections that explore the complexity of the data from different perspectives.

The modules below are meant to suggest a variety of perspectives from which you might look at data and cumulatively assemble an accessible and actionable civic data guide. Begin by completing **Module A: Unpacking the Data Setting**. Subsequently, you can integrate any (or all) of the other modules, depending on what is relevant for your goals. You might think of the primary audience for your guide as undergraduate students, but your guide can also serve a community partner (see the section on community-based learning opportunities).

**Example Civic Data Guides**

Helpful precedents abound, if you know where to look for them: Librarians have long used “readme” files in order to convey important context for existing data sets. More recently, data intermediary organizations like the Western Pennsylvania Regional Data Center and others throughout the National Neighborhood Indicators Partnership network have begun to create new forms of documentation for civic data specifically. The Regional Data Center internally developed [guides to over 40 civic datasets](http://www.wprdc.org/data-user-guides/). Other examples of such guides include “[A Guide for Home Mortgage Disclosure Act Data](https://www.urban.org/research/publication/guide-home-mortgage-disclosure-act-data)” by Kathy Pettit at the Urban Institute, and “[Crime and Punishment in Chicago](http://crime-punishment.smartchicagoapps.org/)” a directory of criminal justice data developed by the Smart Chicago Collaborative.

In addition, please see these sample Civic Data guides, created by Georgia Tech students:

* [Zillow Field Guide, Tzu-Chieh Kurt Hong](http://sls.gatech.edu/sites/default/files/documents/Toolkit-Docs/hong_fieldguide_zillow.pdf)
* [ACS Field Guide, Qing Tian](http://sls.gatech.edu/sites/default/files/documents/Toolkit-Docs/tian_fieldguide_acs.pdf)
* [Open Street Map Field Guide, Udaya Lakshmi](http://sls.gatech.edu/sites/default/files/documents/Toolkit-Docs/tattamangalam_fieldguide_osm.pdf)
* [Property Tax Assessment Field Guide, Jemma Yang](http://sls.gatech.edu/sites/default/files/documents/Toolkit-Docs/yang_fieldguide_propertytax.pdf)

**Module A. Unpacking the Data Setting**

Select a data set that interests you. Next, acquire a copy of the data in a standardized format (i.e. .csv, .xml, .json). Read it through as best you can. Don’t clean the data (just yet). Compose a short overview of the data setting: the context(s) in which the data are made and used. Use the following prompts:

1. First, do some online investigating: How would you characterize the setting or settings where the data are made? Where are those settings? Who works there? When were the data made? Is there a clear purpose or audience for the data? If you can’t answer these questions right away, simply explain what is unclear to you and why.
2. Next, look closely at the data. Start with the standards. What is the data format? Who created this format and how has it evolved in the recent past? What is the current structure (i.e. rows and columns or a hierarchical tree)? Then turn to the specifics. What is an example of a typical data point? What is an example of a data point that you find surprising or confusing? Delving further, you will notice that each data point has a set of fields (or categories) that hold values. To the best of your ability, explain what you think each of the fields means. If you don’t know, make notes on that. What range of values does each field hold? If there are existing metadata or data dictionaries, those may hold the answers to these questions.
3. Find an example of similar data created in another time or place. What stands out as being different about your chosen data set?
4. Find existing uses of your chosen data in journalism, scholarship, or on government and commercial websites. Who is using the data? What interpretations are these uses producing? What things or people are the data being used to examine?
5. Has your inquiry raised any ethical issues thus far? Who do the data represent (directly or indirectly)? What categories are used to represent them? Are there implicit values and assumptions in the use of these categories? Who might they leave out? Does the data raise privacy or security issues?

**Resources for Further Reading**

Kitchin, Rob. *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. Thousand Oaks: SAGE Publications, 2014.

Loukissas, Yanni Alexander. *All Data Are Local: Thinking Critically in a Data-Driven Society.* Cambridge: MIT Press, 2019.

**Module B. Conducting a Contextual Interview**

In this module, you will further examine the social context of your data. The preferred mode of learning about your data from a social perspective is talking to people who are involved in creating, managing or using those data. In-person interviews are most valuable, but phone or online interviews can be productive as well. An email exchange is not an interview, but is acceptable if the subject declines a live exchange.

The interview should be semi‐structured. This means that you should approach the interview with a set of prepared questions, but not be limited by those questions. Semi‐structured interviews allow the subject to jointly‐direct the discussion and raise questions that the interviewer might have not thought to ask. Ask your subject questions that will help you understand how they see the context of the data, its importance and its limits. Try to challenge the assumptions that you have made about the data set in the previous module. Here are some general questions to get you started:

* What is your role and/or relationship to the data?
* What training or experience helps you interpret the data?
* What would you like people to understand about the data and how to use it?
* What kinds of errors can people expect to find in the data?
* Why do errors appear and how can we compensate for them?
* What important information do the data obscure or leave out? Who is most likely to be affected by those omissions?
* What is the pre‐history of the data set? What led to its collection?
* How is it used by the organization that created it?
* How was similar data collected in the past?
* How are such data collected differently in other places?

The interview should last at least 40 minutes. With the permission of your subject, make an audio recording of the interview using a digital device of your choice (i.e. a purpose‐built digital recorder, smartphone, or laptop computer) Soon after the interview, compose a field note that recounts the most salient points of the conversation and uses selective quotes (pulled from the recording) in order to illustrate these points. If your subject wishes to be anonymized, do not include her/his/their name in the writing. However, do include the pertinent details of their identity. If you need some tips on how to conduct an interview, see the Resources for Further Reading section below.

Note: if you plan on using the results of your interview in a publication (any form of documentation that is publicly available), you will need to acquire approval from the Georgia Tech internal review board (IRB). For more information on how to submit an IRB proposal, follow this [link](http://researchintegrity.gatech.edu/institutional-review-board).

**Resources for Further Reading**

Bernard, Russell H. *Research Methods in Anthropology: Qualitative and Quantitative Approaches*. Thousand Oaks: Sage Publications, 1995.

Bowker, Geoffrey, and Susan Leigh Star. *Sorting Things Out: Classification and Its Consequences*. Cambridge: MIT Press, 1999.

**Module C. Diagraming the Data Life Cycle**

In this module, you will diagram the "data life cycle" of your chosen data set. This means learning more about processes of collection, normalization, and analysis used on the data. As a starting point, read through the "Workflow" section the [City of Pittsburgh’s 311 Data User Guide](https://data.wprdc.org/dataset/311-data/resource/5601f50b-05c5-4e73-a1a7-bb5cbb54ae3c). Here is an excerpt:

*Today, the 311 system accepts service requests over the phone, through the internal 311 call center, Web interface, twitter, occasional email requests, or letters in the mail. Other departments are also able to submit requests as well. A mobile application has also been launched by the City on March 11, 2016.*

This example recounts the sources of the data, but nothing about how the data are processed, stored and eventually retrieved. Your life cycle analysis should take this broader process into account. Make sure to include the role of any algorithms in the life cycle. You might start by writing about the processes your data go through, but ultimately, your task is to produce a visual representation of the life of the data from their origins to their obsolescence.

Draw your data life cycle diagram in a systematic manner. You can make use of digital tools like Adobe Illustrator, or create the diagram by hand. Either way, think creatively about the structure of your diagram. It might take the form of a traditional flow chart. It might also be a map, a timeline, a network, or even a comic. Consider sharing your diagram with your interview subject (from Module B) and revising it based on their feedback.

**Resources for Further Reading**

Edwards, Paul N. *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming*. Cambridge: MIT Press, 2010.

Kurgan, Laura. *Close Up at a Distance: Mapping, Technology, and Politics*. New York: Zone Books, 2013.

**Module D. Creating Comparative Visualizations**

*Visualizations are visual representations of data. They are used to help people make sense of data or to allow people to explore data. They take the form of graphs, charts and other more complex or less familiar diagrams. (Helen Kennedy)*

In this module, you will create two complimentary visualizations of your data. Creating more than one visualization is useful for understanding your data independently of any one way of representing them. Don’t attempt to answer any specific questions with these visualizations. Instead, try to make general patterns or anomalies visible. These visualizations are meant to reveal the large-scale structures within your data and serve as a starting point for further exploration.

If you are new to data visualization, visit “Developing Visualization Literacy” on the [Seeing Data](http://seeingdata.org/developing-visualisation-literacy) website and consider using one of the following accessible web visualization platforms:

* [Databasic.io](https://databasic.io/en)
* [Rawgraphs](https://rawgraphs.io/).io
* [Datawrapper.de](https://www.datawrapper.de/)

Your visualizations can be plots, graphs, timelines, networks, or maps. The choice of visualization *type* is up to you. Each platform above has certain standard approaches or templates that you can use. You can also consult the [visualization zoo](http://queue.acm.org/detail.cfm?id=1805128). As you create your visualization(s), consider the following questions:

* Why did you choose these different visual forms?
* What kinds of patterns, if any, do they reveal in the data?
* What questions do they raise? Share your visualizations with your interview subject from Module B if you can and incorporate their feedback.

**Resources for Further Reading**

Fry, Ben. *Visualizing Data: Exploring and Explaining Data with the Processing Environment*. Sebastopol: O’Reilly Media, 2008.

Hall, Peter. “Critical Visualization.” In *Design and the Elastic Mind*, edited by Paola Antonelli, 120–31. The Museum of Modern Art, 2008.

**Suggestions for Community-Based Learning Opportunities**

There are a number of ways you might get a community partner involved with students in the creation of their civic data guides. This will make it easier for students to find subjects for their interviews and/or imagine potential audiences for their guides other than themselves. Moreover, the community partner might wish to make practical use of the civic data guides after the course is over. Here are a few suggestions about who you might ask to form such a partnership:

1. A provider of civic data (i.e. [Trees Atlanta](https://treesatlanta.org/)) can offer important details about the data setting and perhaps put the students’ guides to good use with their own constituents.
2. A group of civic data users (i.e. [Housing Justice League](https://www.housingjusticeleague.org/)) can offer missing context on the needs of audiences or even the subjects of data. Additionally, such civic data users might find the guides helpful for their own work, which might involve advocacy, planning, or evaluation.
3. A data intermediary, such as a librarian, planner, journalist (i.e. an [AJC](https://www.ajc.com/) reporter), or data visualization designer has skills in exploring and unpacking unfamiliar data sets. They have their own audiences that they might wish to share the student guides with.

If you are interested in working with a community partner, Serve-Learn-Sustain is eager to help you. Contact [Ruth Yow](mailto:ruth.yow@gatech.edu), our SLS Service Learning and Partnerships Specialist, for ideas and assistance, and check out our [Service Learning and Community Engagement Nuts & Bolts](http://serve-learn-sustain.gatech.edu/slce-nuts-bolts).

SLS Student Learning Outcomes

1. Identify relationships among ecological, social, and economic systems.
2. Demonstrate skills needed to work effectively in different types of communities.
3. Evaluate how decisions impact the sustainability of communities.
4. Describe how to use their discipline to make communities more sustainable.\*

\* *Note:* SLO 4 is intended to be used by upper division, project-based courses such as Capstone.