

Beyond the Numbers 2022 Schedule

November 8, 2022

- 1:00 – 5:30 PM [Pre-Conference Workshop](#) (requires separate registration)
Data Carpentry Lessons:
- [Data Organization in Spreadsheets for Social Scientists](#)
 - [OpenRefine for Social Science Data](#)
- 5:30 – 8:00 PM Opening Reception & Welcoming Remarks
[Economy Museum](#)

November 9, 2022

- 8:00 – 8:45 AM Breakfast
- 8:45 – 9:45 AM **Keynote Address by Stacey Vanek Smith**
- 10:00 – 10:40 AM [Breakout Sessions A](#)
- 10:50 – 11:35 AM [Plenary Session, U.S. Census Bureau](#)
- 11:35 – 12:45 PM Lunch & Lightning Talks
- 12:45 – 1:25 PM [Breakout Sessions B](#)
- 1:35 – 2:20 PM Data Provider Q&A Panel
- 2:20 – 2:45 PM Break - visit with exhibitors
- 2:45 – 3:30 PM [Plenary Session, Bank of Canada](#)
- 3:45 – 4:15 PM [Breakout Sessions C](#)
- 4:25 – 5:10 PM Plenary Session, Federal Reserve Bank of St. Louis

November 10, 2022

- 8:00 – 8:45 AM Breakfast
- 8:45 – 9:45 AM **Keynote Address by Erica Groshen**
- 10:00 – 11:40 AM [Breakout Sessions D](#)
- 11:40 – 1:00 PM [Lunch & Lightning Talks](#)
- 1:00 – 1:40 PM [Breakout Sessions E](#)
- 1:50 – 2:35 PM Plenary Session, Federal Reserve Bank of St. Louis
- 2:45 – 5:00 PM Optional Networking Events

Abstracts:

Pre-Conference Workshop:

Data Carpentry Lessons: Data Organization in Spreadsheets and OpenRefine

Presenter(s): Jennifer Stubbs and Edward Junhao Lim

Data Organization in Spreadsheets for Social Scientists

Good data organization is the foundation of any research project. Most researchers have data in spreadsheets, so it's the place that many research projects start. In this 2-hour and 15-minute lesson you will learn how to think about data organization and some practices for more effective data wrangling. In this lesson, however, you will not learn about data analysis with spreadsheets. Find more lesson details on the [Data Carpentry website](#).

OpenRefine for Social Science Data

OpenRefine (formerly Google Refine) is a powerful free and open-source tool for working with messy data: cleaning it and transforming it from one format into another. This 2-hour and 15-minute lesson will teach you to use OpenRefine to effectively clean and format data and automatically track any changes that you make. Many people comment that this tool saves them literally months of work trying to make these edits by hand. Find more lesson details on the [Data Carpentry website](#).

This workshop will require a separate paid registration from the free conference. Registration includes both topics, each about two hours and fifteen minutes. There will be a 30-minute break between lessons.

Participants will need to bring a personal laptop with admin privileges to install software.

- Data Organization Lesson [Setup Instructions](#)
- OpenRefine Lesson [Setup Instructions](#)

Breakout Sessions A

Safeguarding Research Code and Data: What Economics Journals' Policies Don't Get About Preservation

Presenter(s): Kira Lillard, Federal Reserve Bank of Kansas City

Economics journals have published policies encouraging researchers to share code and data files for many years, for the purposes of replication. Yet the journals are not preserving the files according to archival standards. Without the application of preservation techniques, the files likely are not findable or accessible in the future, rendering them un-replicable. Using content analysis to review journal policies from 184 economics journals, this presentation will discuss those code and data policies, analyze the policy language which emphasizes sharing but stops short of preservation, and argue that the long-term preservation of code and data files is foundational to a data sharing policy in support of replication. By understanding journals' publicly stated preservation practices, knowledge workers can better advocate for long-term care of replication materials.

Beyond 2D data visualization: Animated, interactive 3D and virtual reality visualizations

Presenter(s): Bill Chau and Brittany Chen, University of Toronto

This session will review some of the latest techniques to enhance visual data exploration for researchers and create more engaging storytelling visualizations for users. Today's data visualizations are presented digitally and are no longer limited to 2-dimensional static graphs. The session will first present the application of data animation, a form of temporal visualization. Animation can be an effective presentation option when multiple categories of time series data exist. Data change animates through time, highlighting the changing relationships between two variables in a 2D graph. They're followed by the

discussion of 3-dimensional visualization, which has been frowned upon for many different reasons. 3D visualizations can create value if used correctly. The presentations will cover the best practices of creating 3D graphs, followed by a walk-through coding sample using Plotly. Plotly is a Python library that allows users to create interactive 3D visualizations easily. Lastly, we will also explore the concerns and opportunities in data visualization with virtual reality.

Plenary Session

Economic Innovation at the Census Bureau

Presenter(s): [Stephanie Studds and Adeline Tran, U.S. Census Bureau](#)

The Economic Directorate of the Census Bureau has been working to transform several programs to gain operational efficiency. We will present some highlights of that effort:

- Business Trends and Outlook Survey (BTOS) is the by-product of the Small Business Pulse Survey, which started during the pandemic to track the effect of changing business conditions.
- Business Formation Statistics measures business initiation activity and the cycle from initiation to realized business formation. Data from business formations has been crucial to tracking the effects of the pandemic on the economy.
- Re-engineering the construction indicators to use modern technology, like satellite imagery and adopting the use of third-party data, to scale back the reliance on traditional survey methods. The goal is to reduce survey expenses while providing more granular data.
- Consolidating multiple annual surveys into the Annual Integrated Economic Survey to improve efficiency across programs.
- Transitioning multiple research efforts from the Center of Economic Studies into standard data products such as the Longitudinal Employer-Household Dynamics and Post-Secondary Employment Outcomes programs.

Breakout Sessions B

Central Bank Survey Database from the Atlanta Fed

Presenter(s): [Ernie Evangelista, Federal Reserve Bank of Atlanta](#)

The Federal Reserve Bank of Atlanta's Economic Survey Research Center created a one-stop-shop for central bank surveys. This brief presentation will give an overview of that tool with a deeper dive into these Atlanta Fed surveys with national coverage: Business Inflation Expectations, Survey of Business Uncertainty, and The CFO Survey.

Bridging the gap: A case study on growing economic data services at Oregon State University

Presenter(s): [Diana Castillo, Oregon State University](#)

Academic libraries have become hubs for supporting data needs on university campuses, from helping researchers navigate data management requirements to housing institutional repositories. This presentation will focus on efforts at Oregon State University, which is a STEM-focused research university, to assess its data services and outreach towards its economics faculty and graduate students and implement new strategies. The case study comes from a needs assessment directed at the School of Public Policy to identify areas of strength and potential growth in the library's liaison work with the school. This session will examine the conditions that led to the needs assessment survey, the initial findings, and how the assessment has influenced data services moving forward. It will also include reflections on how the assessment fits within the larger goals for research data services, what next steps might be, and what could be replicated at similar institutions.

Plenary Session

From Acquisition to Reproducibility – Data Management at the Bank of Canada

Presenter(s): Michele Sura and Jennifer Trower, Bank of Canada

We will provide an overview of the role of the Knowledge and Information Services Team's work in data acquisitions, data licensing and metadata management for data assets at the Bank of Canada. In the past 3 years, significant progress has been made in our ability to support compliant access, sharing and metadata management of the Bank's data assets. We will describe the implementation and functionality of our Data Catalogue along with details on the metadata schema which supports it. We will explain how the Data Catalogue is used to provide internal access to research data and research reproducibility packages, and how these activities support our goals to continually evolve our alignment with the FAIR principles. We will discuss some of the key successes and challenges, and next steps relating to preservation and open access.

Breakout Sessions C

International Survey Data: An Introduction to Sources and Discovery

Presenter(s): Jim Church, University of California, Berkely

This presentation will serve as an overview of sources and strategies for finding and working with international survey data from foreign national governments, international organizations, data archives, catalogs, repositories, and more. A brief introduction to some of the "greatest hits" in the global survey data universe will be followed up by tips and strategies for procuring the data once discovered, as well as potential permissions needed, obstacles that may be encountered, and managing student and researcher expectations. Among the sources reviewed (in brief) will be IPUMS International, the World Bank Data Catalog, the International Household Survey Network, and the UK Data Archive, as well as selected guides and repositories such as the Harvard Dataverse and Princeton University's Data and Statistical Services (DSS).

Business and Finance Data Quality Problems: What Economic Researchers Need to Know

Presenter(s): Grace Liu, West Chester University, FHG Library

The data quality of commercial business and financial databases can greatly affect their research quality and reliability. The presence of data quality problems can not only distort research results, but also seriously damage decisions based upon such research. Based on an extensive literature review of data quality problems, this presentation will highlight 11 categories of common data quality problems of business and finance data, which include missing values, data errors, discrepancies, biases, inconsistencies, static header data, and more. These data problems exist in the most frequently used databases such as the Center for Research in Security Prices (CRSP), Compustat, S&P Capital IQ, I/B/E/S, Datastream, Worldscope, Securities Data Company (SDC) Platinum, and Bureau van Dijk (BvD) Orbis. After addressing these data quality problems with some examples, the presenter will share practical advice for economic researchers on what they need to be aware of when using business and finance data.

Upgrading Data Management at BLS

Presenter(s): Dan Gillman and Clayton Waring, U.S. Bureau of Labor Statistics

The U.S. Bureau of Labor Statistics is undertaking several initiatives to improve the way it manages its data and metadata systems. Two examples include planning for the replacement of its public facing LABSTAT data query system and efforts within its Office of Productivity and Technology to combine multiple production systems within a single cross-divisional database platform. Within these projects, BLS views its time-series data as a combination of three elemental conceptual components: measure element; person, place, and thing element; and time element. The authors turned this basic approach into

a more formal conceptual model represented in UML. The UML model describes multi-dimensional data, of which time-series are a kind, and is very flexible in that it supports any kind of query into the data. The taxonomy described in previous Beyond the Numbers conferences fits naturally into this framework.

Breakout Sessions D

90-minute Workshop:

Utilizing U.S. Census' "On the Map" spatial analysis tool for Labor Economic Data Literacy

Presenter(s): Whitney Kramer, Cornell University and Charissa Jefferson, Princeton University

This workshop leads participants through the U.S. Census' "On the Map" tool, which has potential to answer a variety of economic development research questions by analyzing local phenomena of industry groups and characteristics of labor market flows. This workshop is geared towards librarians who work with economic data, but all are welcome.

This workshop will begin with an evaluative overview of the data sources behind the tool, including the available methodologies, variables, and levels of geography. **Participants are encouraged to bring their own laptops** as the bulk of the workshop will be a scaffolded hands-on demonstration of the tool's six built-in analysis capabilities to create maps, charts, and graphs. After participants have their hands-on experience using the tool, round table or small-group discussions will commence. Participants will be able to share ideas about "opportunities for agitation", a phrase coined by critical information literacy and pedagogies practitioners, as we collectively explore applications of spatial and data literacy skills broadly to socio-economic issues.

Learning outcomes include:

- Participants will be able to describe an authoritative public data resource that can be applied to visual, data, and spatial literacy instruction in practice.
- Participants will have an increased confidence in using a freely available mapping tool.
- Participants will expand their arsenal of information sources for consulting on labor economics topics related to workforce analysis.
- Participants will expand their knowledge of tools that can be used to analyze local communities.

Supporting ESG Research Questions: The Current ESG Data Climate

Presenter(s): Michael Deike, University of Notre Dame Libraries

Recently, the Thomas Mahaffey Jr. Business Library at the University of Notre Dame has seen a notable increase in research questions concerning Environmental, Social, Governance (ESG) data from faculty and students. ESG offers a framework by which to score the performance of companies on a variety of social issues in contrast to traditional financial performance metrics, and it is a growing area of research interest across multiple disciplines including economics, business, and political science. This session will introduce the concepts driving ESG and provide an inside look into tools and resources that librarians can use to support ESG research. During the session, we will compare the usability and accessibility of several popular ESG data sources on the market. We will also discuss the challenges and implications of ESG data in an academic and social context.

No need to argue: Making sense of ESG data sources

Presenter(s): Edward Junhao Lim, University of Connecticut

There has been a wave of developments in environmental, social, & governance (ESG) reporting and standards, with climate and sustainability thrown into the mix. There is so much unpacking to do in this session as we focus on informing our academic researchers on making sense of ESG data sources. This panel discussion will help you see beyond the smoke, and focus on what's available for researchers on ESG, including data sources covering private companies, and companies from non-Western countries.

Ideally, researchers should be able to disaggregate ratings and decide their weighting on indicators as companies shift towards publicly disclosing such data.

Panelists include: Jiebei Luo, Financial & Economic Data Analysis Librarian at New York University, Carolyn Klotzbach-Russell, Social Sciences Librarian at the University of Buffalo, and Reece Steinberg, Business Liaison Librarian at Toronto Metropolitan University

Data from PDFs: Mission Possible?

Presenter(s): [Christine Murray, Bates College](#)

"Wow, that's a great dataset, but I can only find it trapped in a PDF." Maybe you have been in this situation where some great economic information was scanned from print volumes or created only as a PDF. Is this a job for human typists, or can machines do the work for us? This presentation is a whirlwind tour of those tools that purport to turn your human-readable PDFs into machine-readable data. For each tool, attendees will learn the pros and cons based on testing with real examples.

Lightning Talks

The Role of the Business Librarian in Data Ethics: Serving and Preserving Individual Academic and Institutional Reputations and Scholarly Integrity

Presenter(s): [Chelsea Jacobs, West Virginia University Libraries](#)

While often not as widely publicized as misconduct, fabrication of data, or plagiarism of text, hypotheses, or methods or other forms of academic fraud in related disciplines, unethically produced content is also present in business and economic scholarship. Librarians have a role in preventing these types of individual fraud and in preserving institutional reputations.

Not your mother's pyramid scheme: multi-level marketing, direct sales, and financial literacy in the influencer era

Presenter(s): [Margaret Mahoney, Penn State University Libraries](#)

This lightning talk will introduce participants to the nuances between multi-level marketing, influencer marketing, and pyramid schemes. Attendees will also learn introductory strategies for researching influence in consumer purchasing decisions.

openICPSR's Self-publishing Workflow for the Economics

Presenter(s): [Kyrani Reneau](#)

The presenter will give a demonstration of openICPSR --- the Inter-university Consortium for Political and Social Research (ICPSR's) self-publishing repository. openICPSR hosts repositories for several journals and is heavily used by the Economics profession. The presentation will also highlight the customized workflow created for the American Economic Association (AEA) by which data and code are reviewed by editorial staff prior to becoming available to secondary users.

Breakout Sessions E

A Case Study of a North Carolina Mill Community: Using Archival Collections to Understand Economic Decline and Renewal

Presenter(s): [Jesse Akman and Betty Garrison, Elon University](#)

Alamance County, North Carolina is currently experiencing extensive redevelopment of its historic industrial neighborhoods as it grows into a bedroom community for neighboring urban centers. But where did those neighborhoods come from? Why are those buildings located where they are? Why are they empty now? We offer a case study in using archival collections to help undergraduates situate current

economic conditions in their historical contexts. Using archival materials, we trace the history of the regional textile industry to its origins in chattel slavery, explosion in the interwar synthetic fabrics boom, and finally its apogee and decline in postwar America. We then synthesize these ideas with the current real estate industry, showing students that economies and infrastructure are never created in a vacuum.

The Shifting Ground of NAICS: Retail, Information, and other Subsectors have changed

Presenter(s): [Jennifer Boettcher, Georgetown University](#)

North American Industry Classification System (NAICS) is the way companies define competitors and peers. Governments use it to regulate procurement and the environment. NAICS was created to count establishments. How do the completing roles of NAICS affect researchers? To understand this open taxonomy of the standard industry definition we will answer who creates and approves new industries and how does NAICS become a norm every five years? For some, these new NAICS codes mean a simple crosswalk for the 2017 definitions to 2022. It will create breaks in a lot of datasets. What are some of the best practices by governments, publishers, and researchers to reflect historical data? Will this be the time for the popularity of the North American Product Classification System (NAPCS) to come into its own?