Padua 2017 Abstract Submission

I want to submit an abstract for:
Conference Presentation

Corresponding Author
Amy Azzarito

E-Mail
azzarito@ucdavis.edu

Affiliation
University of California, Davis

Keywords
Wine Labels, Retail Wine Catalogs, Ephemera, Data Modeling

Research Question
How can you transform historic economic ephemera into data sets for scholars to use and explore?

Methods
Using methods of crowdsourcing the text transcription, we will extract tabular data from historic retail wine catalogs.

Results
The resulting database will be used to create an interface that allows researchers to do a variety of data modeling.

Abstract
The wine collection at the University of California, Davis, has been called the "best in the world." For the past year and half, we've been begun to create an equivalent experience online. In July 2016, the Library at launched a crowdsourced wine label transcription project, "Label This," as the first step in creating digital assets to create a larger platform through which wine-lovers, researchers and historians around the world can understand the story of wine. Combining the wine label information with data from contemporaneous sources requires the development of methods for extracting structured data from scanned data, as well as a comprehensive ontology for wines and the wine industry.

This presentation will discuss lessons learned in the development of crowdsourcing projects that draw on ephemera collections; how to organize these data; and how to develop interfaces that allow historians and scholars to shape new narratives. The presentation will examine both the process and the potential of utilizing ephemera collections for scholarly communication.

Robert Mondavi once told The New York Times about University of California, Davis professor Maynard Amerine, "If you did what he told you, you couldn't help but make outstanding wine. He was my mentor, and I wouldn't be where I am today if it weren't for him." In 1935, Amerine had been hired by the fledgling University of California, Davis' newly created Department of Viticulture and Enology to help establish a solid technical foundation for the California wine industry. But Amerine wasn't just interested in training students -- he was also driven to rectify the damage wrought by Prohibition to the wine industry in California.

As he traveled California and the world, educating and counseling winemakers on how to improve the quality of their wines, he collected over 5,000 wine labels. In an era before computers, Amerine pasted the labels into black spiral-bound notebooks, loosely organized by region, wine color, and in some cases further divided into sub-regions or 'appellations.'

Before Label This, the only way to search Amerine's collection of historic wine labels was to flip through the hundreds of spiral-bound notebooks housed in Special Collections at the UC Davis Library. Our original goal was to provide an application to bring that collection to the public.

Starting with the open-source Scribe Framework built by NYPL Labs and Zooniverse, we developed tool that enabled the general public to help decipher what's on each label. When complete, scholars will be able to search by vintage, varietal, region, or even the type of image on a label. The tool allows for the Library to foster engagement and interest in collection, and also provides the foundation on which to build a larger discovery platform. For researchers around the world — historians, sommeliers, oenologists, even novelists looking for period details — this online resource will fill a gap in their understanding of wine history.

In a companion project, the Library wanted to bring into life historical wine pricing as a way for historians to map changing palettes as well as to create a means for economic historians to meaningfully access this data. Founded in 1934, Sherry-Lehmann has been selling wine in New York City for 82 years. The Library has a run of the wine catalog from 1955 through 1970. Each catalog has articles by food luminaries such as James Beard, but more importantly, lists wines for sale and their prices. For example, the November 1959 catalog has a bottle of Volnay, Pousse D'OR, 1949 listed for $4.99. For the past several decades, these catalogs have housed in the UC Davis Library where they remained relatively unexplored and underutilized.

The initial step is highlighting this collection was making it digital. Even after running the digital files through Optical Character Recognition, we're only left with the ability to do a rough keyword search. Our goal is to move the data from the page into a database of prices. Page layout and transcription inconsistencies make price data extraction a more challenging problem.

While investigating better methods of machine extraction, we leveraged our experience with Label This, creating a crowdsourcing application that asks users to help us identify the wine and its listed price. We've used the automated OCR pathway to jump start the process, but for the crowdsourcing piece, we're looking for very specific answers to very specific questions.

Once we've exacted the pricing data into tabular form, we'll have a search interface that allows users to query the database by region, type, vintage, year sold and bottle size. This ultimately enables a tool with which wine industry professionals can answer questions like: From what countries or states in the world is the wine coming from? How has the proportion between California and Europe changed? Or, where is the most expensive/least expensive wine coming from (country, region)? This type of geographically-based query can be visualized as an overlay map, which would include an interface for filtering through various facets, like price range, countries, etc.

This direction is a reflection of the importance of storytelling initiatives as a way of eliciting interest among wine professionals as well as a general public. As librarians, we need to do more than create tools, but also create avenues for shaping narratives. It's through that storytelling narratives that we help shape new avenues for access and discovery. One of the key things for us is to enable users to try to select data subsets and create visualizations on the fly so they can understand what happened to wine in California - How California wine became such an important economic and social force. By moving past the crowdsourcing applications into more interface and visualization, we are seeking a new way of tying together disparate data, developing ways for people to understand the implications of agriculture throughout history.