# History of Women in Printing Pre-1800: Leveraging Wikidata As a Discovery Tool – Transcript of audio

Please stand by for realtime captions.

---

Hello, everyone. My name is Don Sensabaugh from GPO. On with my colleague Ashley Dahlen who is doing tech support. Today's webinar is the history of women in printing Pre-1800 , leveraging with the data as a discovery tool. Our presenters are Daniela Rovida and Cindy Tian. Cindy Tian is the metadata services librarian for the Kresge Law Library at the University of Notre Dame. With the introduction I will let Daniela take it away.

---

Good afternoon. Thank you for the kind introduction and for inviting us to presented a. It's an honor to be here. We are delighted to share our project which aims to shed light on the significant contributions of Women Printers during the hand press era. As you may know, -- has a long and resonating history and women have played a vital role in this industry since its inception. However, achievement of many female printers have been overlooked or forgotten overtime. Our project seeks to resurface their stories. Through our research we hope to honor this remarkable women and inspire future generations to celebrate their legacy. This project began in 2022 one we joined forces to shine the light on the often overlooked contribution of Women Printers to the global printing history. Will recognize the potential of wiki data and data platform and decided to leverage its power along with tools like open refined and quick statements to achieve our goal. It's worth noting that anyone can undertake similar projects with just a basic knowledge of wiki data many resources are available to help you get started. Numerous institutions are experimenting with these tools and can provide support. By working with linked data, it becomes easier to link and share information making data more open and accessible on the semantic web. This is an emerging area of interest and other libraries and cultural heritage organizations are finding innovative ways to implement similar projects. Our passion for this topic and opportunity to work together, share ideas, and combine our skills and knowledge help us move forward with this project. We were willing to experiment and saw this activity as a valuable learning opportunity. Today, we aim to introduce you to the power of Wikidata as a link data tool that libraries can use to make data more accessible on the web. We are willing to introduce a few remarkable printers brick we also highlight the important role that women have played in the printing industry throughout history, both in economic and cultural --. We hope this presentation will inspire librarians and other information professionals to consider similar projects that promote the discoverability of traditionally siloed resources within Library special collections and cultural heritage organizations. Finally, we will delve into the specific steps and processes we use in our project. Discussing the challenges we faced and the lessons we learned along the way. We will -- our expenses will be helpful to anyone who want to embark on a similar endeavor. This is Martina plantain, the daughter of a French printer who was involved in her father's printing business from the age of five. The family moved to Belgium and continue to work in the printing industry. There printshop was the largest printing business in the world at the time. After the death of her father and husband, Martina ran the family printing business for 40 years. Cindy and I are both employed in technical services in two separate libraries at the University of Notre Dame. Despite having different roles in our libraries, we share a passion for cataloging materials. We stumbled upon this story of draw anime, an Italian printer who worked in Peru's general during the 16th century. Girolama's name appeared on several books she printed. Yet, there have been attempts to erase her existence and deny her contribution to the field of printing. In one instance, and Italian author and bibliographer transcribed the name of the printer from one of Girolama's books and changed Girolama the Latin for Girolama female name to the Latin for Girolama a male name. Women Printers like

Girolama have been largely ignored and underresearched for various reasons. They often did not find their names to the books they produced or designated themselves as the air or wife of the master printer. Additionally, our cover documents related to their work is difficult to locate. Despite these challenges, Women Printers can be considered well-established businesswomen in a field and occupation where male presence was dominance. The professional printing requires a certain level of literacy and women who took over the businesses of master printers had to establish networks of relationships with authors, papermakers, booksellers, binders, and publishers. It is important to recognize and acknowledge the contribution made by these Women Printers to the printing industry and to the wider culture and economic spheres. We selected Wikidata as our discovery platform due to its benefits as a central storage for structured data on the web. Usability of data, Wikidata supports multiple languages along with editors to work with the same data despite language differences. As more data is added all the time, many libraries and institutions are contributing to making library resources more discoverable to wiki projects. By using queries, data can be accessed, retreated, reused and displayed in meaningful ways. We will demonstrate this later. Moreover, data can be added and edited either manually or through batch loading processes. For those who may not be familiar with Wikidata structure, facts are scored stored and structured within entities. Each of which is indexed by a unique identifier called a Q number. Comprises three parts that connect an item and property and value are stored within each entity. Wikidata data power relies in relationships between statements. We will show you a couple of examples to demonstrate how the user can search facts. There are interconnected. Wikidata pages also provide links to Wikipedia articles in multiple languages. Wikipedia articles are available in different languages , Wikidata editors can work with the same data across languages. Essentially, Q numbers, properties, and their values constitute the basic building blocks of link data. On the screen you can see a portion of the Wikidata item for Mary Katherine Goddard. Un-American printer whose name appeared on the --, aversion of the independence printed in Baltimore in 1777. I'm going to share my screen. I will share my screen to show you all the statements and identifier associated with this printer. One emerging use of the Wikidata is I -- institutional identifier. Many maintain unique identifiers for people, subjects and works and Wikidata pulls these identifiers together into one location. You can see here multiple identifiers associated with this printer. This page is also linked to Mary's brothers page. He was also a printer. From here we can access his Wikidata page. Which includes a portrait. While many male printers can be discovered and Wikidata, the number of female printers is much smaller and it is rare to find a portrait or image of them. This is another example of how Wikidata makes it possible to record the relationship between entities such as the link between Mary Goddard and the work she printed which was the first version of the Declaration of Independence to lease the name of the sinus. Here is a screenshot of the portion of the document where we find evidence that it was printed by Mary Goddard. The Wikidata page for this document includes information about the printer. This is the document. This is the Wikidata page for the Goddard -- and one of the statements tells us that it was printed by Mary Goddard. By clicking on her name we can open the printers Wikidata page . I opened her page in Wikidata . I'm going to stop sharing. Another interesting fact about the history of printing is that the first person to own a printing press in America was also a woman, Elizabeth Glover in Cambridge, Massachusetts. After her death, the printing press was donated to Harvard and is now Harvard University press. As a team, one of our first test in this project was to identify sources of names and information about printers and we have focused on the first two sources on this list. This is not a complete list, however. I would like to highlight the women's print history project. A bibliographical database of women's contributions to print during the 18th century. It is a source of data about female authors, publishers, editors, and also printers. Another promising source is the book, notes on Women Printers in colonial America and the United States. It provides information on women who contributed to the printing industry in America from the 1700 -- 17th to the 20th century. We try to

incorporate this and other sources to further expand our database. Now, I will pass it on to Cindy who will provide more details on our project.

---

Thank you, Daniela. Our next step is to build a data model with Wikidata robberies and values that would best describe these Women Printers. There are many existing data models for describing people and the one we referenced was the data models constructed by the Stanford libraries wiki project. They are presented at three levels. Basic description, which is the bare minimum to create an item such as the person's name as the label. Court description, which are descriptive elements and help identify the person, such as aliases, occupation and local -- work location. Finally, extended description, which include robberies that are optional to add to the item if certain information is known about that person. Our workflow is illustrated in this graph. Are still at the preliminary stage of the project and currently focus on two data sources that helped us collect 203 names of Women Printers before 1800. Within in just these names into OpenRefine which is a data editing and reconciliation tool that helps us reconcile these names against Wikidata through automated process. We don't have to search by hand for each one of them. Although, it does require some manual verification of the results, especially when they are duplicated names in Wikidata. At this point we had 114 names matched which means these names are already established items in Wikidata . 67 names unmatched, which means we needed to create new Wikidata items to surface these things and make them discoverable. We then used QuickStatements, a tool that can edit and create Wikidata items in batch based on text commands. For the matched names we added the statement occupation equals printer, if not already present, for query purposes, which I will explain later. For the unmatched links, we created new Wikidata items in batch based on the data model we established earlier. We also had 22 names that were only known history as -- without given names. When we create Wikidata items for these names, do we want to label them as someone or can we assume their given names were the same as their late husband's names? What would be the best practice here? These are the questions that we need to answer. Before we get to the results, I would like to reflect X Lane how queries work and how practical data applications is a power tool that is built on SPARQL language. I'm going to share my screen to show you an example here. This is an example of a Wikidata query to retrieve all the name entries in Wikidata that is a female and whose occupation is a printer and with a birth date before 1800. I run the query and we can see that there are currently 284 names in Wikidata that matches these conditions. Each with a link to that person's Wikidata item page . Depending on the type of data that is available, there is also data visualization options besides table like a met view or chart view or timeline view, which we will take a look at later. Finally, I want to point out there is a community of users and resources that can help us get started with learnings SPARQL or try to build queries that are little bit complicated. There are examples here that we can copy and modify so we don't have to start from scratch. There is also this query builder, which is a visual interface to build simple queries, which is ideal for users with little or no experience in SPARQL. Next, we want to show a few outcomes of our preliminary work. The first as we saw earlier, the number of Women Printers before 1800 and Wikidata increased to 284 after our batch processing. As our project continues to collect more data from different sources, we hope that number will continue to grow. From the same query, because we designated it's work locations of these Women Printers, Wikidata can pull the coordinates and display the locations on the map. I'm going to share my screen again. I'm going to show you the map view. You can see that most of the Women Printers in the hand press era worked in different parts of Europe. A few of them worked on the East Coast in America. Lastly, we fixed some data errors in wiki data as we encounter them. One example as shown here is that we found some items mistakenly French names as English. Lacked the real name of the female printer, only designated her as the widow of. As mentioned in the beginning of the presentation, women were usually designated the heirs of the male master printers and sometimes were better known in history as those names. One of our goals of this project was to make about their real names, if known from certain reliable sources, as

well as their better-known names searchable on the web to facilitate related research. In this example, we made the change to the label from where Daniel Horthy mulls that was mistakenly placed as an English label to Marianne atelier the printers fully name while living the other name as the French level so it remains searchable on the web as well. One of the challenges we encountered was the name variations. For example, Katerina Del Silvestro was recorded on page 270 of migration -- one of our reference books with a letter a were it is noted as Katerina Del Silvestro with a D and other resources to cases like this require more research and manual fix to avoid creating duplicate entries in wiki data. Another thing is, in the query that we did earlier, we were restricting it to find the items with a statement of which the occupation has the value printer. If an item happens to lack such statements, the specific query will not successfully pull up that item for us. Upon further examination, we found some statements have printer -- bookseller which is a class a printer as the value for occupation. This could be due to a different data model used from other Wikidata contributors. You can see that they -- the inconsistencies can create challenges. We want to be able to catch everything with one simple query. We also found that when adding geographical names for work locations we need to pay attention to the contemporary constraint for using that property. For example, France, which is Q1 42 in Wikidata has a inception date of 1792 and therefore people or activities predating 1792 will be flagged when attempting to add Q 142 as a value. One lesson we learned was that one working on projects that involve historical figures and places in Wikidata , we need to choose the entity that matches the persons time. We recognize that at this preliminary stage we are working with a small sample of names, so one of the next steps is to continue to collect more women printer names from different sources. We also plan to further refine the query views such as replacing the data -- date of birth before 1800 with -- before 1800 as a search parameter. We can see if we can get more accurate results since many dates of birth are unknown for these names. Following that is more enhancement work. We want to figure out the best practices of handling incomplete names known as widow or daughter of. We will also work on back linking the Women Printers names to their spouses or fathers Wikidata page. We will build that symmetric relationship and highlight the women's inheritance and continuing management of the printing business. We also hope to add more identifiers such as VF I.D., the women's print history project person I.D. and library of Congress authority I.D. to these names. As Daniela mentioned, Wikidata can essentially be an entity management system that groups different identifiers together for one entity. Other future work enclosed more research on the corporate body names such as monasteries and convince were nuns contributed to the printing of text. This picture is a screenshot of the woodcut printers device used by the nuns of -- and Augustinian monastery in Italy were the volumes printed were mostly Latin and vernacular devotional works. Several were written by the nuns themselves. We realize that even though the nuns names were probably lost in history, we could still document their contributions by presenting the printing office that they worked in. We also hope to get to the works that the early Women Printers had produced and interlink the personal names and works just like the linkage between Mary Katherine Goddard and the Goddard Declaration of Independence that was explained earlier. Lastly, as Daniela mentioned we like to step out of our project a little bit and encouraged fellow librarians participation in Wikidata. It's a good way to get our feet wet with editing link data and make resources more accessible and discoverable on the semantic web. There are a lot of Wikidata pilot projects undertaken by the library community, especially in the special collections and cultural heritage area. We started with a very basic knowledge of Wikidata and the beauty of it is, it is open to everyone to contribute and there are resources online on how to get involved and use different tools. You can start small with your library resources are and see where these might take you. Here is a page of our references. We are happy to take questions. Thank you.

Yes, I'm going to give everybody an opportunity to put some questions in the chat. I have two questions. At the first question I'm going to lead off with his, where are you looking in order to get more sources for Women Printers? Do you have any places that you have identified that you are looking at?

I can answer this question. Yes, we are identifying some articles that sometime, some of them are not English but talk about printers, particularly one article was very interesting about Women Printers in Italy. We had to use our language skills to identify names and Women Printers. Also, other sources we identified some printing material books and a database that focuses on one century but it's also a valuable source of information. And then we are looking for more sources. I'm sure there is more than that out there so that is one of the steps, our future steps is to identify as we make progress in our project, to identify more sources. The other interesting thing about Wikidata is, like Cindy was explaining, it is a community. One we started our project and we created a page for our wiki project and we were contacted by a group of Wikidata group in Italy that is doing similar work on Italian printers, their scope is a little --. The only focus on Italian printers but they reached out to us and they asked that they would like to contribute and maybe connect and work together. That is also the beauty of Wikidata pick it's open and everybody can contribute and can see what other libraries are doing or other is to tuitions. Increase opportunity to work together with other institutions.

We do have a question in the chat. I'm going to break up this first question. Have you had any challenges with Wikidata's requirements when getting items for these printers?

I think with Wikipedia there might be some kind of requirements but Wikidata is essentially open to everyone to contribute. Everyone with a Wikidata account -- actually, I don't even think you need an account to create a new item in Wikidata . Personally, I'm not aware of any requirements for Wikidata . Everyone can contribute and create entries and, yeah. You might wonder, in that case, what if there are data errors or unreliable information that was added there? The answer to that would be, anything that is entered in Wikidata can be reverted , so I think there was a discussion about the concern of vandalism in Wikidata , but I think so far it is just people are not motivated enough to do that.

It looks like there are some people who have also responded to that question with some information that they have about Wikipedia and Wikidata. Thank you for that. The next part of this question was, do you find the Wikidata community requires side link, for example to Wikipedia or wiki source for your new items to be accepted?

Like I mentioned, Wikidata will accept any data that you contribute, any item you want to create, so there is really no requirements for your items to be accepted. Of course you can link to other sources. Actually, there is --. Each Wikidata item, there is a link . I think for every Wikipedia article, there will be a Wikidata item but not vice versa.

Next question that we have is in reference to the slides he presented, did you make a decision on printer versus winter bookmaker? Are either of those a subcategory of the other?

Yeah. Printer -- bookmaker or bookseller is a subclass of printer. When we designed the data model, we decided to go with printer just because it is a broader term to use, but then we realized some items only have that subclass as the value, which makes it a little bit tricky using that query to be able to catch everything. So, moving forward we will further refine the query to help us find the existing items in Wikidata , but when we do the enhancement for creation of the item, we will still use the value printer.

Let me add a little bit to that. It's possible to add more than one occupation, so if you encounter a person that is listed as of printer bookseller, you can also add the value as just printer, or a lot of times these printers were also involved with binding or papermaking, so it is possible there occupation was more than just printers. It will help for when you need to retrieve information. If you do a query and you are looking for a certain category, as long as the statement is listed on their page, you can retrieve the list of women involved with a specific occupation.

Our next question Ritz, I joined the webinar late so I apologize if it has been mentioned. Did you add links between the Wikidata you are creating and the resources are in addition has so that your resources are more discoverable?

That's one of our future plans. Speaking -- Daniela and I work in different libraries at the University and my library, which is the law library, has a saw small collection of rare books and we have yet to identify the works that we own that have Women Printers name on it. But if we do, in the future, we think it would be a meaningful way to link the person's name and the works they produce.

I identified --. When I was researching the -- the printer I talked about in the presentation, I identified two rented books that were in her name on the title page. It would be nice to be able to link her work to this -- to her Wikidata page . There is also, there is a way you can make your library or archive of resources more discoverable in Wikidata. One thing, one way is to link the printer to the work they printed, but I'm also involving another project that might be interesting to some of you. We actually try to --. A lot of archival material, so we added a statement. We created Wikidata page four this individuals for which we have in our archival material. We added a statement that says archives at with a link to our finding aid. If anybody is searching a poet and they can see where the material is archived and other libraries are doing the same thing, so, they basically -- it is a way to let users know where they can find archival material for this particular individual. It's also a way to make your resources more visible and to bring together anything that is specific to an author or individual. So, there is a lot of different projects out there that are attempting to make resources more visible.

I see a comment in the chat about SPARQL queries that are easier to do. I have a question about the wiki pages. I think it was on Mary Katherine Goddard. The identifiers or the tabs, the data has a link to references or at least it will say there is zero references or one reference for a certain entry and I'm curious where those references come from, if you might be able to go back to those pages.

I'm sorry, were you talking about the data source pages slide?

I probably am.

Actually, I might be talking about --. If you go back to the Mary Katherine Goddard were you screen shared. If you can screen share again and show that page.

Sure.

I was interested in the references, actually. See how for human it says one reference and I'm curious, is it linking to other Wikipedia pages?

Yes, it could be. Any reference URLs can be put in here as a reference. It just happens that these were pulled from the Wikipedia articles. For her date of birth you can see there are different references listed here. These were either entered manually or it was harvested by bots I think from the external sites.

That's interesting because I'm more familiar with Wikipedia articles and using the sources of the very bottom of the page where you will reference. I wasn't aware of how the references in Wikidata worked. It looks like we might have more questions. Let me see. Have you added your project to the wiki project section in Wikidata so others can join you or learn from this? Your model is excellent and it would be a great benefit for those who want to do the same thing or similar projects.

Yes, we have created a page, a Wikidata item for our wiki project but we have not added this project to the wiki project section on Wikidata , which we need to do. Thank you for the suggestion.

Where do you find data models?

I was going to do the screen share again. I was going to show you that this is some data models that were devised by Stanford University library. Like I said, their data models were presented at three different levels, which we think makes sense for describing people, especially like, for example this is for the academic faculty. They have these three different levels of descriptions and we just take that model and modify the properties and values that we thought was suitable for our person. I hope that makes sense.

There is a lot of wiki projects and they usually use data models and the data models very if it is for an individual, person, corporate body or book, so there is a lot of resources available. You don't need to start from scratch. You can look at what other libraries have done and shared. You can develop your own model.

I'm putting in the chat our webinar satisfaction survey. I'm not releasing any more questions in the chat so I'm going to thank Daniela Rovida and Cindy Tian for the webinar. For those of you in the chat, make sure to check out our upcoming webinars. We have one in the FCC's affordable connectivity on Monday. And USGS publications warehouse on May 11. Please fill out the webinar survey if you have the opportunity. Thank you so much and have a great day.

[ Event concluded ]