

About

With approximately 45,000 menus dating from the 1840s to the present, The New York Public Library's restaurant menu collection is one of the largest in the world, used by historians, chefs, novelists and everyday food enthusiasts. Trouble is, the menus are very difficult to search for the greatest treasures they contain: specific information about dishes, prices, the organization of meals, and all the stories these things tell us about the history of food and culture.

To solve this, we're working to improve the collection by transcribing the menus, dish by dish. Doing this will allow us to dramatically expand the ways in which the collection can be researched and accessed, opening the door to new kinds of discoveries. We've built a simple tool that makes the transcribing pretty easy to do, but it's a big job, so we need your help. [Feeling hungry?](#)

Questions? Comments? Want to stay in touch as the project develops? Contact us at menus@nypl.org.

The menus

The New York Public Library's menu collection, housed in the Rare Book Division, originated through the energetic efforts of Miss Frank E. Buttolph (1850-1924), who, in 1900, began to collect menus on the Library's behalf. Miss Buttolph added more than 25,000 menus to the collection, before leaving the Library in 1924. The collection has continued to grow through additional gifts of graphic, gastronomic, topical, or sociological interest, especially but not exclusively New York-related. The collection now contains approximately 45,000 items, about quarter of which have so far been digitized and made available in the [NYPL Digital Gallery](#).

The Rare Books Division of The New York Public Library houses approximately 200,000 titles, covering five centuries of printing—from the 1490s to the present—and representing Continental Europe, England, and the Americas.

FAQ

What are the goals for this project?

When we launched in late April, 2011, our sights were set on the approximately 9,000 menus photographed several years ago for inclusion in the [NYPL Digital Gallery](#). Volunteers transcribed those in about three months! Since then, we've been steadily scanning additional items from the collection and loading them into the transcription queue. The ultimate goal is to get the whole collection transcribed and to make the data available for exploration and use by researchers, educators, chefs and other interested folks. Along the way, we may add other user-solicited tasks such as geolocation or various remediation and linking of the data. We're also looking into ways of expanding the scope of the data set through partnerships with other libraries and archives with significant menu collections.

What exactly will transcribing accomplish?

Right now, the only information that is digitally searchable in our menus is the descriptive data created for each item when they were cataloged. This includes useful things like the name of the restaurant, its geographical location, date etc. But the actual menu contents — all the dishes and wines once upon a time offered to the customer as they pondered the options for their meal — is only accessible through good old-fashioned sifting.

How will this information be used?

Researchers who use the collection — be they historians, chefs, nutritional scientists, or novelists looking for a juicy period detail — often have very specific questions they're trying to answer. Where were oysters served in 19th century New York and how did their varieties and cost change over time? When did apple pie first appear on the Library's menus? What about pizza? What was the price of a cup of coffee in 1907? To find out these sorts of things more easily, we need to extract all the delicious data frozen as pixels inside these digital menu photos. The best way to do this is transcription.

Update: A great example of scholarly investigation into the still-evolving data set is the [Curating Menus](#) project of Katie Rawson and Trevor Muñoz.

So just transcribe, and presto?

Well, the data will need some additional cleanup in order for our search engine to handle synonyms, spelling variants, faceting, all that good stuff, but hopefully you'll start to get a palpable sense right away of what you're helping to build. Every transcribed item instantly becomes part of a searchable index, which allows you to much more nimbly trace dishes, ingredients and prices across the collection. We'll be occasionally [blogging](#); and [tweeting](#); about interesting discoveries that come up along the way. We also have begun to offer some fun visualizations of the data (like [this](#)).

Why transcribe? Why not just OCR?

Good question! There are a few reasons. First, while we could get decent OCR output from some of the clearer printed menus, many others are handwritten, or use fanciful typography and have idiosyncratic layouts that will result in little more than alphabet soup if we use mechanical translation methods. A more compelling reason is that we're interested in unpacking some specific types of information that are highly relevant to researchers: dishes and prices (and eventually menu sections, geographical locations and perhaps other data). Even with a crystal clear OCR text, a human being will still need to go through and identify each individual dish, price, section (appetizers, entrees, wines etc.), and so on. We're not just scooping out text from pages, we're building a database of dishes!

Plus, as a library we know that the more that people use a collection, the more we collectively learn about it. Our hunch is that there is a lot to be gained by inviting the public to help us go through these fascinating artifacts with careful attention, menu by menu, dish by dish. We also hope that by doing so, we'll stoke people's appetite (so to speak) to explore the collection further.

Wait, what's OCR?

Optical Character Recognition. Basically, it's the process by which we extract usable, searchable text from scanned pages. It's how Google Books and Hathi Trust do their search. Wikipedia has a [good explanation](#).

Need help with anything other than transcription?

Right now, we're focused primarily on two tasks: 1) transcribing dishes and prices; and 2) [geotagging menus](#). We've also considered other kinds of work down the road that will help us add even more value to the database. Possibilities include: menu sectioning (appetizers, entrees, wines etc.). We're also interested to see what the public does with the data and what new research questions they propose.

Can I create an account?

Our current policy is to keep things as open as possible, but we intend eventually to tie into a NYPL-wide user account system that's currently in the works. We'll always preserve the option of participating without a login, but providing a way for more intensive contributors to identify themselves will allow for a community to develop and the possibility of more complex tasks. We're grateful for the time/effort you've devoted to this project so far, and hope to be able to recognize some of our top contributors down the road.

Will you provide an API, data exports etc.?

We already have! All of that can be found [here](#).

Has the Library ever done something like this before?

We have an active project where we're collaborating with the public around our historic maps. [Take a look!](#)

Team

Rebecca Federman, Project Curator

Rebecca is The New York Public Library's Culinary Collections Librarian and Electronic Resources Coordinator. She is also the co-curator of the New York Public Library's latest exhibition [Lunch Hour NYC](#). Rebecca writes about the Library's culinary collections on her blog [Cooked Books](#) and is a visiting professor at Pratt Institute's School of Information and Library Science.

Michael Inman, Project Curator

Michael is Curator of Rare Books for The New York Public Library, administering a number of the institution's collections and departments, including the Rare Book Division, George Arents Collection, and Historic Children's Book Collection, among others. He holds both a B.A. and a M.A. in English from the University of North Texas and a M.L.S. from Pratt Institute's School of Information and Library Science. Michael also serves as a visiting professor at Pratt Institute, where he teaches courses on printing history and special collections librarianship.

Ben Vershbow, Project Director

Ben is the Director of NYPL's Digital Library + Labs department, overseeing metadata, digitization, permissions/reproductions, and the explorations of the [NYPL Labs](#) team. Before joining the Library in 2008, he was Editorial Director of the Institute for the Future of the Book, a small Brooklyn-based think tank exploring the future of reading, writing and publishing.

Mauricio Giraldo, Designer/Developer

Mauricio spent the last twelve years designing and developing interaction design projects for a wide range of commercial, academic, private and public institutions. Mauricio is an Industrial Designer from Universidad de los Andes in Bogotá, Colombia where he also lectured for six years. He also holds a Master in Human-Computer Interaction from Carnegie Mellon University.

Kristopher Kelly, Application Developer/Data Analyst

Kris has a BA in English from Harvard University and a Masters in Science and Information Studies from the University of Texas in Austin. Currently, he serves as a Senior Applications Developer for NYPL's IT group, where he works on a range of internal applications.

David Riordan, Product Manager

David is the product manager of NYPL Labs. When not at work, he's biking, reading, and involved in policy advocacy.

Emeriti

- Rebecca Austin, Collection Assistant
- Amy Azzarito, Project Co-Founder
- Edith Bellinghausen, Intern
- Barbara Bleck, Intern
- Ben Chartoff, Intern
- Amanda Glassman, Community Manager
- Jayme Hall, Intern
- Leslie Harker, Intern
- Katie Krause, Intern
- Zeeshan Lakhani, Application Developer
- Michael Lasearides, Project Advisor/Prototyper
- Meredith Mann, Intern

