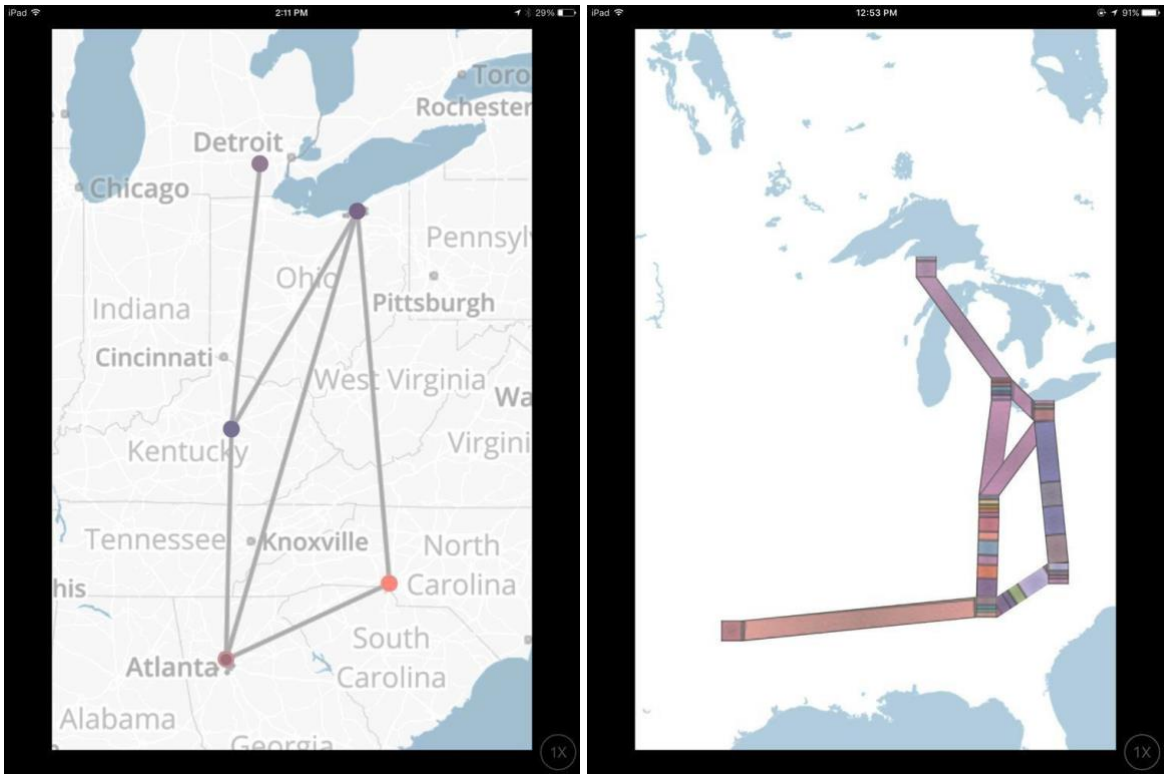
Week 4: Vocabularies and Metadata

Metadata

In his monograph on the subject, Jeffrey Pomerantz (2015) says that he believes any course of study in Information Science should begin with a class on metadata, as all of the rest of what we do builds off of it. While we didn't start with metadata in this class, we're spending a good amount of time on it, as Pomerantz is correct about its importance to our field.

We work with metadata every day. It makes what we do possible. For example, tagging data artifacts allows us to design faceted search solutions. In addition to the metadata created for use with our content, we also make products that result in user-produced metadata.

Artist Laurie Frick creates room-sized installments based on the personal data collected by the devices and applications she uses. She also produced a now-defunct iPhone app called FRICKbits that allowed users to visualize their own location data in the style of her artworks. The two screencaps below show examples of how the app worked.



The first screencap shows the initial view of data when users first opened the app. It drew lines that showed where the user had been (the view could be changed down to the neighborhood level up to the whole world). The app would then remove data like borders and place names

from the map and replace the initial dots & lines with ones evoking Frick's style in her work, as shown in the second screencap.

The app was fun, but it also served the purpose of showing users the metadata their lives produced in a clear way. In addition to the visualizations, the app had the option to export GPS coordinates into a spreadsheet. Entering those coordinates into something like Google Maps would show each individual place registered by the app as a location where I had been, including my home and place of work. It was sobering to see how much data my phone produced when tracking my location.

We deal with people's personal data every day in ways that may not seem immediately obvious. Most times, we do this with the best goals in mind. We want people to share their locations so that we can give them contextual information like taking the weather into account. We gather information passively so that we can personalize in a way that doesn't feel invasive. Many times, the best compliment for a product is that "it just works!" and metadata is an important part of getting to that feeling. It is our job as information professionals to remember the "personal" in personal data and to advocate for its responsible and informed use.

A recent example of this kind of metadata making the news is the company Niantic with their location-based games such as *Pokémon Go*. While not the phenomenon it was when it first came out, it's still a popular game that gets people outdoors exploring (I'm one of those people—I played it a lot as a motivator to take walks while staying at home during the pandemic). It also collects a lot of personal data that results in articles with titles like, "The Creators of *Pokémon Go* Mapped the World. Now They're Mapping You" (2019), in which the authors discuss the things they could infer about players only from their location metadata. The authors of that article were able to get access to that data because players in Europe requested it using the General Data Privacy Regulation, the European Union's rules covering personal data (GDPR).

With requirements like GDPR and the California Consumer Privacy Act (CCPA), how we deal with personal metadata moves beyond the philosophical into a practical reality in that the companies we work for open themselves up to legal issues if they don't comply. And directly or indirectly, if you work for a company that does business in a state or country with these kinds of privacy protections, you'll have to understand them. This can be rewarding. One of my favorite projects I worked on at IHG Hotels & Resorts was working with the Legal department on forms for CCPA compliance data requests.

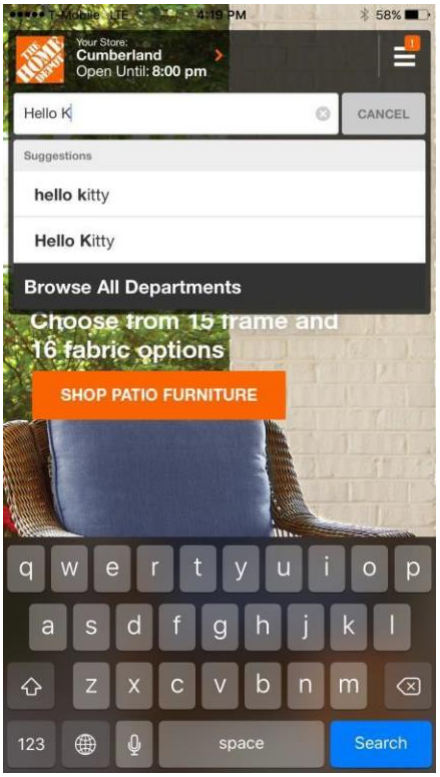
Vocabularies

You will also work with vocabularies on a regular basis, though there's a good chance that you won't refer to this kind of data as a "vocabulary" or "thesaurus" unless you're tasked with producing an original one. In my experience, you'll likely work with an existing thesaurus. If you work in eCommerce, for example, a catalog of products will probably already exist. Even when producing a new vocabulary, your starting point may be a current vocabulary that you then augment and adapt for your project.

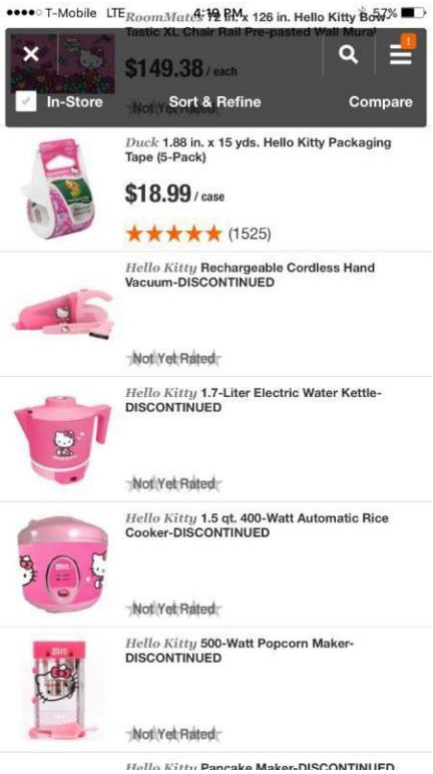
Much of my work has revolved around the interfaces to access existing vocabularies. I've rarely worked with changing the terms that people can search for, but I've often worked with how to display those searches and make sure that needed data is communicated.

One example from when I worked for The Home Depot involved trying to make it clear to customers when a product in our catalog was no longer available because it had been discontinued by the manufacturer. While we were working on a solution, my team and I used "Hello Kitty" as our search term because we knew there would always be discontinued products in the results.

The screencap below shows that the catalog has two synonymous entries for "Hello Kitty": one uses title case, one uses all lowercase, and both show as type ahead options in the search form.



Once a customer has selected a "Hello Kitty" search option, products with that label show as results. The system prioritized showing available products before showing discontinued ones, so all we had to do was scroll to the bottom of the results to find a discontinued example.



The screencap above shows the results before the solution we came up with went into development. There are many differences in presentation between available products and discontinued ones, but customers were still getting confused. Our approach involved changing the visual design to more clearly mark a product as discontinued, but there are other ways in which this could have been approached. How would you have approached it? Stay in the habit of asking yourself that question as you continue to develop your IA skills.

References

California Consumer Privacy Act (CCPA) (n.d.). *State of California Department of Justice*. Retrieved from https://www.oag.ca.gov/privacy/ccpa

D'Anastasio, C. & Mehrotra, D. (Oct 16, 2019). "The Creators of *Pokémon Go* Mapped The World. Now They're Mapping You." *Kotaku*. Retrieved from https://kotaku.com/the-creators-of-pokemon-go-mapped-the-world-now-theyre-1838974714/

GDPR – User-Friendly Guide to General Data Protection Regulation. (n.d.). *GDPR*. Retrieved from https://www.gdpreu.org/

Pomerantz, J. (2015). *Metadata*. Cambridge, MA: MIT Press.