

Ten Quick Tips for Making Things Findable

Sarah Lin, MSLIS^{1,*}

¹ RStudio, Inc.

* sarah.lin@rstudio.com

Abstract

Library science principles offer a framework and methodology for working through the maze of information generated in professional life. Information ecosystems are composed of users, context, and content, and all three areas need to be addressed to make information findable. Scientists should define their users and needs and then leverage the structural elements of the software they use to create, store, and access information. Only then can refine their content and how they present it via naming and subject metadata. Ultimately, improving findability promotes scientists' work through contribution, iteration, and dissemination of information and ideas.

Author summary

Sarah Lin is the Information Architect and Digital Librarian at RStudio, PBC.

Introduction

Researchers have always had to manage information of various kinds, but the “exponential growth of electronic data increasingly calls for new means of organizing and accessing information” [1]. The resulting glut of information, “has also led to the creation of new technologies to help people organize, find, and make better use of information” [2]. Trying to find a particular bit of information when needed can waste precious time and cause frustration. On top of this, information often resides in many formats and is shared with varied audiences, sometimes in multiple versions, which makes it even harder to separate what's useful from what's merely intriguing.

Luckily, librarians make a career out of helping people find information and use information management technologies. Library science principles offer ways to work through the maze of information generated in professional life, and librarians' skills can be applied by any researcher who feels overwhelmed. The ten quick tips in this paper are simple to understand, easy to apply, and effective. They are based on the fact that information ecosystems have three parts: users, context, and content [2]. Solving the challenges of information retrieval at the point of need therefore requires addressing the people who use your information, the context within which it is created, and the actual content of the information.

1. Know Your Users

The first step to making information findable is to establish who will be finding information. In a professional context, your users are people who have a “viable and

legitimate interest in the work you're doing" [3]. Through contributions, they can help you iterate your own work or expand upon it in their work. They are also transmitters of the information you share through their own work and networks, and serving their information needs ultimately benefits you as well as them.

Depending on your content and the nature of your role, you might have only a small, expert audience. If your work is publicly available, however, accept the idea that complete novices will find it, thereby making your actual user base much larger than you might realize.

Novice users should be able to understand and navigate your content just as well as subject experts. This much broader range of users means that you would expect a wide spectrum of domain knowledge as well as facility navigating your discipline's routine practices.

The information you wish to convey and the way it is currently organized probably makes perfect sense to you. Its meaning for your users, however, is determined by *their* experience. That knowledge is "*whatever a user interprets* from the arrangement or sequence of things they encounter" [3]. This is another way of saying that the organizational structure(s) you employ are a communication channel in their own right, in addition to the domain knowledge you share. In addition to knowing who your users are and how important they are to the impact of your work, remember that any information you transmit to them is shaped by the mechanism(s) you use.

2. Define Your Findability Problem

Before you can design a solution, you must first establish both the extent of the problem as well as the end state you seek. Those two elements together are the *why* that drives this process. Given how easy it is to create information in a digital environment and the plethora of software and formats you may employ personally and professionally, it is quite likely that your information is a bit of a mess. "The first step to taming any mess is to shine a light on it so you can outline its edges and depths" [3]. This means taking a look at both *what* information you have and *where* it is located. Once you know the full scope of the problem you have, you can move towards greater findability.

Creating a picture of the end state you will achieve through information organization can be personally motivating, yet in this context your users' needs should be paramount. You must figure out what *they* will consider a good outcome, and "[e]very decision you make should support what you've defined as good" [3]. This may not completely overlap with what you currently define as good for your own work, but the entire point of organizing information is to make it findable to others as well as ourselves, so you must balance your present workflow needs with the needs of your users, rather than assuming that their needs perfectly match your own. And remember: your future self is also one of your users, while your past self is effectively a novice, so anything you do for others will likely pay off for yourself eventually.

Remember too that the question of how to find something can have several different meanings. Users may need to find your work on the web, find a specific item within a website, and/or find a particular piece of information within a specific file or webpage. Depending on your content, you may have challenges in all three areas. There are fine-tuning techniques appropriate for each; the remaining tips related to context and content will help you address the full scope of your findability problem.

Take, for example, the faculty website of a researcher coming up for tenure.

- The researcher created the website as a place to dump her vitae content and make it easier to fill out grant applications by having professional activities listed in one place.

- A librarian might access that site looking for journal articles to include in the university's institutional repository, whereas a student might come looking for course information.
- Tenure committee members might be interested in that laundry list of accomplishments, as a way to determine the impact factor of the researcher's published work. Three of these users need the researcher's list of professional publications, but three different 'versions' of that information.
- The researcher might only provide a plain text citation which meets her needs, but the librarian wouldn't get a link to the full text (or a PDF download) and the university administrator would not find each publication's impact factor.
- The student needs are not considered at all.

Much of the content you create through your work has multiple audiences, so figuring out the extent of *who* needs to find *what* is essential before moving forward.

3. Use Textual Structure

Your information's context is one aspect of structural communication. Findability issues at the document, post, or article level can be addressed by taking advantage of the elements provided by each program you use to improve organization and navigational support as well as retrieval support for your users [1].

For example, a key component of web searching is scanning text in search of target information. Textual structure helps that process [4]. Examples of this are using formatted headers (rather than just enlarged text) and bulleted or numbered lists, as well as **highlighting** terms that are important. Depending on the software you are using, headings and table of contents can be hyperlinked, which supports both scanning and navigation.

Textual content is created and aggregated in so many forms and using so many different programs that specifying strategies beyond headings, lists, and highlighting is difficult. While you want to exploit the features of the textual program (your word processing software, data aggregator, or other information-processing-tool) to the greatest degree possible, you must first investigate each software for ways to do so and think about each program as a way to notate information in addition to creating, manipulating, or storing it. Examples include:

- GitHub offers tags that can be added to committ messages and are searchable.
- Electronic lab notebooks of any kind can make use of an XML schema, like Darwin Core, EML, or FITS [5].
- Using specific Google Docs heading levels creates a table of contents in real time, visible when the file is open.
- Since CSV and spreadsheet files contain data, use a README file or create a data dictionary to provide information about the data and store it with raw data files.

4. Add Metadata

Just like people who end up with piles of photographs with nothing written on the back, we all have digital mounds of files and content with no format or subject information in the Properties (feature used by most modern operating systems). Metadata, such as

those contained in a file's Properties, adds extra access points for information retrieval, which gives users more avenues to find what they need. However, what metadata you are able to add depends on the software used to create, store, and access your information, and on the file formats that information is stored in.

Many file storage, word processing, and website construction programs have built-in metadata capabilities, though they may be hard to find and harder to understand how to leverage. The most difficult part may be getting into the habit of adding metadata to your content after creation. You'll need to have done some thinking about format and subject, but if your website software has any metadata functionality (like tagging; see tip 9), you will want to take advantage of the additional discoverability. If you ever have a choice of software, one that makes use of metadata and structural labels enhances search tremendously. It is also useful to examine how metadata can be transferred from an old system to a new one if you have the luxury of multiple software options (or have had a change forced on you). For example:

- The tagging implemented on a WordPress website doesn't correspond to a taxonomy field in a .yml header file.
- Keywords added to a journal article submission are not automatically added to the file metadata in PDF format, and pulling a citation from an article database doesn't give you the keywords as well.

Unfortunately, metadata is often software-specific and it is worth thinking through what portability options you might have. XML is an established standard that would port metadata, if the software you're using accomodates it. Exporting metadata in a portable file format, like plain text or XML, would have more flexibility in migrating software formats.

5. Use Search *and* Browsing

Research into information-seeking behavior shows that people use a combination of searching *and* browsing when they're trying to find information they need. As they browse a website or document or file, they put together a mental map of the content they could possibly find and search based on that map. "In the process, they modify their information requests as they learn more about what they need and what information is available from the system" [2]. They refine their search further as they see results, often navigating into a result or two to further their understanding. You have probably seen or done something similar with a print book, trying to determine if it's one you want by looking at the table of contents and the back cover. These two functions work together because search allows users to find information they know they need, whereas browsing allows users to find information they don't know that they need [6].

Given this behavior, it's important to make information accessible both ways and to "look for ways to support moving easily from search to browse and back again" [2]. Tags and other metadata help with searching, while browsing a list of tags, a table of contents, or even a file directory tells users a lot about the content contained in the information they are looking at.

Textual signifiers—elements within a text that convey meaning outside of the words themselves—and software structures also help users navigate between searching and browsing by making the information framework visible to the users. That communication, "enables the answers to users' questions to 'rise' to the surface" and answer questions like, Where am I? What's here? Where can I go from here?" [2].

These are essential wayfinding questions on the path to findability and interact closely with the navigational choices you make for your information.

A system of headings marked in a textual file is a great example of intradocument wayfinding. As a user scans the text, the headings act as topical markers, so they can both summarize what information is contained in the document, but also refine what they would search for based on the terms used in those headings. Navigation bars on websites function in a similar way. If the user knows exactly what they are looking for, they can scan the menu and select the option that matches their need.

In a situation where they don't know exactly what they need, the terms in the menu help them understand both the vocabulary used in this domain as well as the boundaries for what is included (i.e., terms are listed) or excluded (i.e., no menu terms exist). In this way, search and browse work together to enhance findability. Though search might be thought of as a universal solution, without knowing what it is possible to find and the words used to find it, search is not a stand-alone solution.

6. Mimic Real World Directions

Even in a digital environment, our language mirrors that used for physical directions. We “visit” or “go to” a website without actually changing our physical location. Using the navigational metaphor completely is very helpful to ensuring that you assist your users as much as possible, because “the words you use in the navigation systems and headings of [your content] help you find what you're looking *for*, but they also help you understand what you're looking *at*” [7]. Navigation is yet another structural element that provides context to your users and communicates important information about your work and how they can find what they need, so give your users a virtual map with sign posts and directions so they can figure out where they need to go.

As a content creator, you might think that since your audience is educated, they will have either an innate understanding of how you have structured your content or unlimited patience to persist until they get what they need. In reality users scan but don't read: they click on the first close thing they see and give up very, very quickly. Menus, headings, file paths, breadcrumb trails on websites, and tables of contents give users a sense of where the information resides as part of the larger information ecosystem they are currently in as well as which new paths they can take [4]. Your directions should be consistent throughout your information ecosystem just as highway signs are with regards to appearance, style, and type of information,

7. Name for Finding

File and site naming affect findability both in terms of how they interact with your search application and how human-readable they are once they are retrieved as search results or through browsing. Naming can be a challenge because of the need to balance a human-readable name that conveys the ‘aboutness’ of the item as well as a pattern or schema that is also machine-readable. For file directories with many files, leveraging machine-readable components like sorting with a particular date format can automatically place files in an order that enhances findability. The utility of one choice over another hearkens back to what you defined as good for your users.

There are many ways to develop a naming schema, largely related to the nature of the information you create. At the most basic level, “you should use consistent names for the same reason that you use good file organization: so you can easily find and use data later. Additionally, good naming helps you avoid duplicate information” [5]. Researchers with multiple research projects or significant complexity in their data

sources might want to establish a unified system of abbreviations for those projects or sources (these would be best summarized in a data dictionary or README file). Consistency is key: choosing your preferred case, preferred date format (YYYYMMDD or YYYY-MM-DD will both sort chronologically), and any project or source codes related to your research and/or data types are all areas where standardization will benefit anyone searching for information.

If you have things to name that are not files, such as projects, web pages, or document headings, remember that the more generic the name the harder it is to understand afterward, they can be nearly impossible to find things later if their names do not reference their actual contents. Try to ensure that your names immediately and uniquely identify what they contain or reference. When naming, you will want to imagine that someone will enter the name into a search box at some point in the future. Will it lead them back to this particular item or content? It should. Also, think about nicknames or shortened versions of your names and make sure they are present in text or tags so that the content can be discovered by a search engine and a user. Naming your raw data file “raw” or a downloaded file “download” makes finding the information they contain nearly impossible. Renaming files can seem like a waste of precious time, but since the research cycle doesn’t end with publication [5], there is a very high likelihood that someone will need to reuse your data and will have to try to figure out what files corresponded to what part of your research.

8. Use Tags

Software metadata choices are part of the structural context of your information, but what you choose to put in metadata fields related to content subject and formats depends on your source material. Additionally, digital information can handle multiple tags, which aides searching. This is a key feature, because it means that you can now ‘file’ information in multiple locations— something that was not possible in the era of print-only information. Multiple tags also assist users from varied backgrounds because the terms can be customized to each type of user your information has.

You need to be consistent in your depth of topical term assignment (how specific your terms are) and selection of terms for subject and format (the number of terms you use to describe each subject and format). Exactly which terms you choose may not matter much for your discipline, beyond a need for consistency, but in some disciplines the particular terms chosen matters a great deal. However, if your discipline has an established thesaurus, it can be great source for standardized subject terms. Thesauri have built-in subject hierarchies that can help you create navigational structure. Examples include the Astronomy Thesaurus, the Getty Art & Architecture Thesaurus, the Education Resources Information Center (ERIC) Thesaurus, the NASA Thesaurus, the UNESCO Thesaurus, the GBA Thesaurus of Geosciences, and the Medical Subject Headings (MeSH) [8].

These types of term lists are controlled vocabularies. They have a defined list of “official” terms that is created and maintained by experts. There may or may not be relationships built between terms, such as equivalencies (“CA” for “California”), broader/narrower terms (United States/California), and/or replacement (weed USE marijuana). One benefit of using established subject terms is that they that will match article databases and library catalogs that you and your users might already be familiar with.

Alternatively, you can create your own taxonomy of subject keywords, which is called a *folksonomy*. Folksonomies are what you see with tags on Flickr and Unsplash: the content creators assign terms as they see fit. Additionally, users can tag with whatever terms they like and the list of possible terms is both uncontrolled (no one is

watching for synonyms, misspellings, or inconsistencies) and crowd-sourced (created by non-experts). The good news is that there is probably a place you can find tags in use that are a similar topic or content to yours that will inspire the tag structure and/or terms you'll want to use. If you create your own list of content terms, beware of plural/singular discrepancies in your terms and be sure to review that list on a regular basis for errors and misspellings.

9. Understand the Difference Between Form and Substance

When you create metadata tags—whether controlled like a thesaurus or uncontrolled like a folksonomy—you need to address the distinction between format and subject, which adds another layer of conceptual analysis to your tagging process. Format describes what your content *is*, while subject describes what it is *about* [9]. About-ness is the most common content analysis, but you will likely have issues of is-ness that affect your users and will want to add another layer of metadata to address it, where appropriate. A simple example of this is a blog post on a website. The post is *about* a subject, but it *is* a blog post. If it were your website, you would want to tag that file both for its subject matter but also that it's a blog post. This enables users to search or filter by topic, format, or both.

Going back to your users, what subjects are important to them? And, do those topics carry over or change between differences in format? This is basically a question of combined terms: are your format terms uniquely matched to topical terms (blog posts are always about news) or do you have multiple topics in each format (blog posts and tutorials on the same subject)? When you think of file formats, sometimes the software chooses the format which in turn signals what the content is, for example, data stored in CSV files or a presentation saved as a Keynote or PowerPoint file.

In these cases, you know a CSV file will contain some type of a data to be analyzed, whereas Keynote and PowerPoint will have slides that are viewed unidirectionally. However, dissemination sometimes changes a file's format (e.g., printing slides to a PDF), in which case careful naming and metadata conventions would recommend putting the format into the name and/or file metadata. Once you are aware of format and subject analysis, you will likely see more opportunities to distinguish information sources for your users. You might create a folder specifically for conference presentations that could include both .ppt and .pdf file extensions, where each presentation's files have the exact same name (with different filetypes). Likewise, journal articles you store will need a naming convention that distinguishes between articles you have written and those relied on for research.

10. Do Not Abbrevt

As an expert in your field, you “can no longer imagine what it's like to *not* see the world that way” [10], and have to work to ensure that the information you share with your users is accessible to them. Spelling out acronyms and abbreviations (or hyperlinking to definitions) breaks down the exclusivity of language used by a select group and makes newcomers feel welcome.

Remember that acronyms are often repurposed by different professions or disciplines: what seems obvious to you is probably not obvious to a number of other people. Given that your discipline likely has many common abbreviations, you can write out all abbreviations and acronyms in their full form, write them out the first time they appear, or create or point to a term dictionary.

0.1 Conclusion

Changing work habits is hard, and you must decide whether to start new work with a clean slate of new information principles or if you must go back and alter—perhaps rename—existing work products [5]. Remember that while “perfection isn’t possible, progress is” [3]. Creating and structuring good information for your users is your goal, but these tips and the “ways you enforce your way of doing things changes how users think about the place[s] you made and perhaps ultimately, how they think about you” [3]. Circling back to the network of users you are situated among, making your stuff findable is a truly impactful way to disseminate your work to them that will positively change your network and your career.

References

1. Hedden H. The Accidental Taxonomist. Information Today; 2016.
2. Louis Rosenfeld JA Peter Morville. Information Architecture. O’Reilly Media Inc.; 2015.
3. Covert A. How to Make Sense of Any Mess. CreateSpace Independent Publishing Platform; 2014.
4. Krug S. Don’t Make Me Think, Revisited. New Riders; 2014.
5. Briney K. Data Management for Researchers. Pelagic Publishing; 2015.
6. Bates MJ. Toward an integrated model of information seeking and searching. New Review of Information Behaviour Research. 2002;3:1–15.
7. Arango J. Living in Information. Two Waves Books; 2018.
8. of Indexers AS. Online Thesauri and Authority Files; 2020. <https://www.asindexing.org/about-indexing/thesauri/online-thesauri-and-authority-files/>.
9. Daniel N Joudrey DPM Arlene G Taylor. Introduction to Cataloging and Classification. Libraries Unlimited; 2015.
10. Wilson J. Teaching Tech Together. CRC Press; 2020.