

Aligning Collections With Emerging Needs in Research Informatics

Heidi J. Tebbe
North Carolina State University Libraries, hjtebbe@ncsu.edu

Darby Orcutt
North Carolina State University Libraries, dcorcutt@ncsu.edu

Follow this and additional works at: <https://docs.lib.purdue.edu/charleston>



Part of the [Collection Development and Management Commons](#), and the [Scholarly Communication Commons](#)

An indexed, print copy of the Proceedings is also available for purchase at:

<http://www.thepress.purdue.edu/series/charleston>.

You may also be interested in the new series, Charleston Insights in Library, Archival, and Information Sciences. Find out more at: <http://www.thepress.purdue.edu/series/charleston-insights-library-archival-and-information-sciences>.

Heidi J. Tebbe and Darby Orcutt, "Aligning Collections With Emerging Needs in Research Informatics" (2015). *Proceedings of the Charleston Library Conference*.
<http://dx.doi.org/10.5703/1288284316298>

Aligning Collections With Emerging Needs in Research Informatics

Heidi J. Tebbe, *Collection Manager for Engineering and Data Science, North Carolina State University Libraries*

Darby Orcutt, *Assistant Head of Collection Management, North Carolina State University Libraries*

Abstract

Some of the North Carolina State University (NCSU) Libraries' largest investments are in collections, digital library development, and technology-rich collaborative spaces. The goal of the NCSU Libraries Fellows Program initiative, "Aligning Collections with Emerging Needs in Research Informatics," is to ensure these areas leverage one another to the benefit of our users in support of emerging research informatics needs through licensing and acquisition of new data sources, as well as leveraging the capabilities of new high-tech library spaces. Over its two years, this initiative seeks to address and mainstream subject specialists' and selectors' consideration of high-tech research informatics needs of users.

Early accomplishments of the initiative include content mining agreements, increased awareness of scholarly APIs, and an ontology to describe research informatics. Ongoing work includes an investigation of relevant collections, licensing terms, and the landscape of the current marketplace; an environmental scan of NCSU research and teaching contexts that would benefit from greater availability of content as data for computational purposes; how-to documentation and training for more technologically sophisticated uses of existing resources; negotiations of select licenses to allow for more flexibility of content use; and revision of our website to promote the research informatics capacities of the Libraries' collections to our users.

Background

Research informatics involves the use of technology to solve complex problems in the humanities, arts, and the social, natural, and applied sciences, through creating, storing, finding, manipulating, and sharing information. Discipline-specific areas of informatics include health informatics, social informatics, and digital humanities.

The NCSU Libraries Fellows Program (<http://www.lib.ncsu.edu/fellows>) offers new librarians a two-year appointment during which they develop expertise in a functional area and contribute to an initiative of strategic importance. Projects outside of their home department support strategic initiatives while providing concrete and innovative opportunities for fellows to interact with librarians from other departments and achieve at a professional level.

Some of the NCSU Libraries' largest investments are in collections, digital library development, and technology-rich collaborative spaces. The goal of

the NCSU Libraries Fellows Program initiative, "aligning collections with emerging needs in research informatics," is to ensure these areas leverage one another to the benefit of our users in support of emerging research informatics needs through licensing and acquisition of new data sources, as well as leveraging the capabilities of new high-tech library spaces.

Over its two years, this initiative seeks to address and mainstream subject specialists' and selectors' consideration of high-tech research informatics needs of users.

Accomplishments

Content Mining Agreements

In June 2014, the NCSU Libraries announced the first of many successful negotiations with content providers for blanket content mining access for North Carolina State researchers, when we became the first institution in the world to offer this model of access to primary historical content from a major for-profit vendor (Gale). These

accomplishments have been documented in several press releases, with more to come soon:

Gale: <http://www.infodocket.com/wp-content/uploads/2014/11/final-Gale-data-mining-press-release-1103142.pdf>

Unlimited Priorities/Accessible Archives: <http://www.accessible-archives.com/2015/03/unlimited-priorities-and-ncsu-libraries-partner-to-create-model-data-mining-agreement/>

Adam Matthew: <http://news.lib.ncsu.edu/blog/2015/08/14/ncsu-libraries-adam-matthew-digital-strike-groundbreaking-content-mining-agreement/>

Let's Eat Lunch and Talk About APIs!

Application programming interfaces, or APIs, are used to build software applications and may be used by scholarly publishers and others to allow users to retrieve data in a greater variety of ways. As such, they are important tools for librarians and researchers to know how to wield.

In April 2015, Heidi Tebbe led a brown bag discussion for subject specialists to increase awareness of APIs. The discussion included topics such as how APIs are used externally (Google Maps, Twitter) and internally (CatalogWS, FacetBrowser 2.0, scholarly publications repository, Suma, GIS Lookup), API vocabulary, potential issues an API user might experience, and things an API user should look for in the documentation (limitations, key, base URL, available resource types, response type).

Web Page for Users Seeking Data Sources

We have also created a list of scholarly resources that offer APIs or other online tools to access or manipulate data. Some of the resources allow large data downloads for research purposes, while others are intended for smaller scale research activities such as experimentation with visualization and development of tools to query journal and citation databases. More information is available at this link:

<http://go.ncsu.edu/textmining>

Herding Research Informatics CATSS

Two essential questions for the initiative are:

1. How does a library evaluate their research informatics offerings?
2. What base knowledge and areas of expertise should library staff acquire?

“Informatics” and “research informatics” are terms that are too broad for effective investigation of these questions, and search results often get overtaken by the more active areas of discipline-specific informatics, such as health informatics. An informatics-specific vocabulary or taxonomy was needed.

A list of “things” related to research informatics was compiled by Tebbe, including visualization, text mining, APIs, and XML. Based on this list, she created a taxonomy with five categories, centered around functionality of the “things”:

- Concepts: ideas that are relevant to research informatics
- Activities: things researchers do
- Tools: things researchers use
- Software: formally packaged tools
- Sources: things that provide researchers with data

This hierarchical representation was a good start, but what really was needed was a description of the domain of knowledge related to research informatics. An ontology would provide semantic relationships. A consideration of other ways “things” related to research informatics could be described led Tebbe to TaDiRAH (Taxonomy of Digital Research Activities in the Humanities). While TaDiRAH focuses on tools and methods for digital humanities, the research activities category is broad enough to describe research across disciplines. Research activities is broken down into the following subclasses: capture, creation, enrichment, analysis, interpretation, storage, and dissemination. For the purposes of this ontology, analysis and interpretation are considered to be the same activity.

We also needed a way to make connections with how we present library services. On the Libraries

Research Support portal, we currently have the following “bins”: collections, citation management, data and GIS, grants and funding, data management, publishing and copyright, visualization, and measuring research impact.

Combining these ways of describing research informatics, the ontology (<https://github.com/NCSU-Libraries/ontology-for-research-informatics>) now has the following relationships:

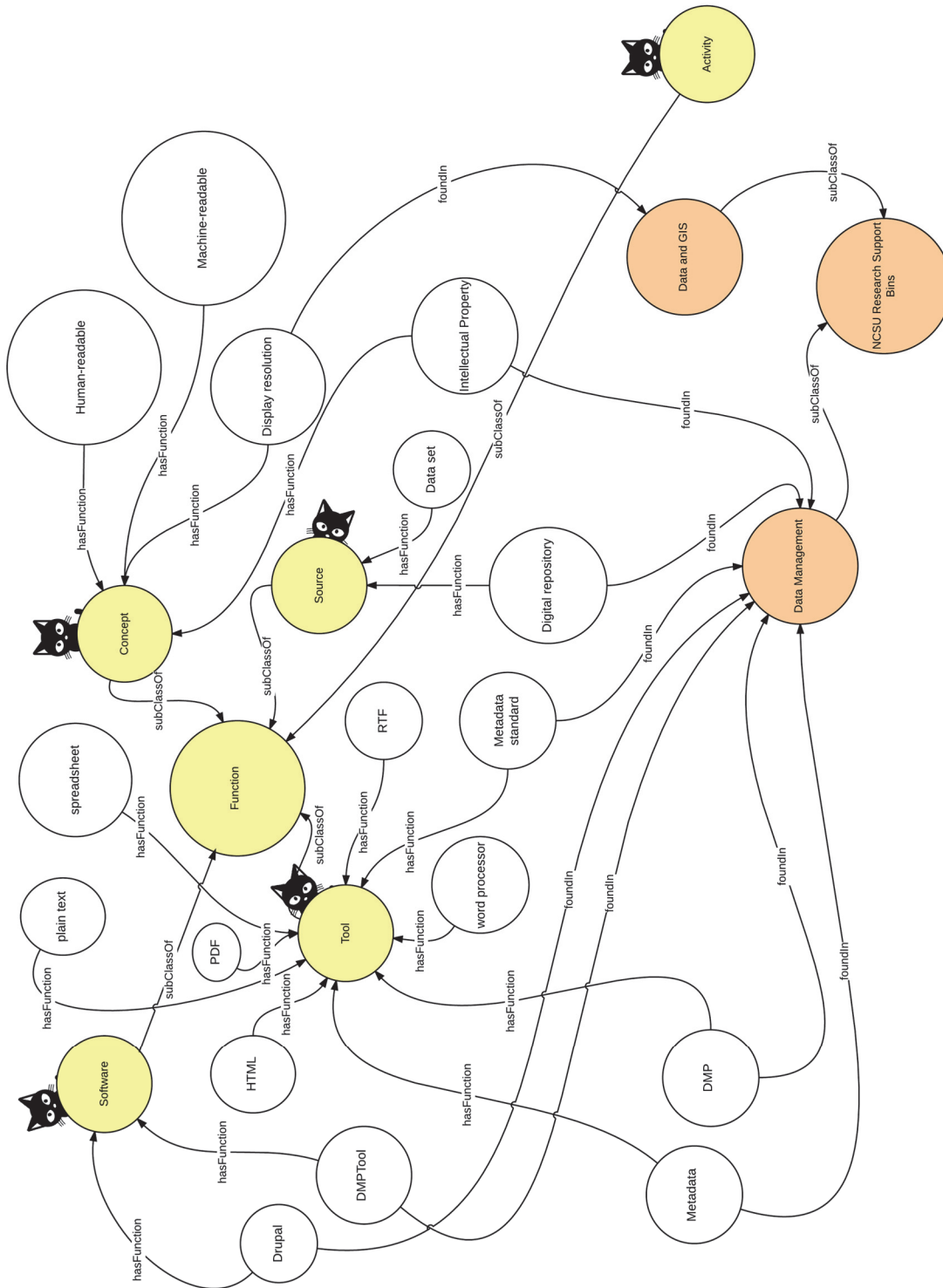


Figure 1. In June 2015, the Libraries' Data Management pages referenced one Concept, three Tools, two types of Software and one Source.

- Thing hasFunction Function
- Thing [is] usedDuring TaDiRAH Research Activity
- Thing [is] foundIn Research Support bin

Additionally, each thing has a definition, a definition source, and Libraries page references, which are Libraries pages or collection items that refer to the thing in question. A thing can only have one function, but can be used during multiple research activities and be found in multiple research support bins.

The value of an ontology like this is directly related to the objects described by it. In other words, almost 150 research informatics things have been described by the ontology, but this is by no means an exhaustive list and that will limit any investigations using the ontology. While the ontology can't provide definitive answers to our questions, it can point to areas that need more investigation and can serve as a starting point for exploration.

Ways to use the ontology include:

- Compare research informatics "things" to website search analytics
- Find the intersection of library departments who mention research informatics things the most with the research activity with the largest number of things to find things with which some library staff need to have some base knowledge
- Find what the Libraries are promoting (or not promoting) on the website

For example, looking at the NCSU Libraries data management "bin," we see one "concept":

intellectual property, three "tools": Mmetadata, DMP, and metadata standard, two examples of "software": Drupal and DMPTool, and one "source": digital repository. We could also look at "things" used during the "storage" activity (shown in Figure 1), one of the TaDiRAH research activities closely related to data management, and see what we aren't talking about (namely, PDF, spreadsheet, HTML, word processor, plain text, RTF, dataset, display resolution, machine-readable, and human-readable). Display resolution is mentioned in the data and GIS bin, machine-readable is mentioned on a nonresearch support page, and a few of the things are found in the library collections, but what about the rest? Do they need to be mentioned?

Ongoing Work

Other planned work associated with the initiative include:

- An investigation of relevant collections, licensing terms, and the landscape of the current marketplace
- An environmental scan of NCSU research and teaching contexts that would benefit from greater availability of content as data for computational purposes
- How-to documentation and training for more technologically sophisticated uses of existing resources
- Negotiations of select licenses to allow for more flexibility of content use
- Revision of our web site to promote the research informatics capacities of the Libraries' collections to our users