



# Wikipedia and Wikidata Help Search Engines Understand Your Organization

Authors: Kenning Arlitsch, & Justin D. Shanks

This is a preprint of a book chapter that originally appeared in the book, [Leveraging Wikipedia: Connecting Communities of Knowledge](#) in 2018.

Arlitsch, K., & Shanks, J. (2018). Wikipedia and Wikidata Help Search Engines Understand Your Organization: Using Semantic Web Identity to Improve Recognition and Drive Traffic. In M. Proffitt (Ed.), *Leveraging Wikipedia: Connecting Communities of Knowledge*. Chicago, IL: ALA Editions.

Made available through Montana State University's [ScholarWorks](#)  
[scholarworks.montana.edu](https://scholarworks.montana.edu)

# Wikipedia and Wikidata Help Search Engines Understand Your Organization

Using Semantic Web Identity to improve recognition and drive traffic

KENNING ARLITSCH  
DEAN OF THE LIBRARY  
MONTANA STATE UNIVERSITY

JUSTIN SHANKS  
DIGITAL SCHOLARSHIP LIBRARIAN  
MONTANA STATE UNIVERSITY

## Introduction

Semantic Web Identity (SWI) is the condition in which search engines formally recognize entities and their relationships. Entities can be people, places, organizations, landmarks, or other “things” but in this chapter, entities will be defined as academic organizations: libraries, but also other academic units such as colleges, departments, centers, and institutes. An entity can be said to have achieved SWI if a formal display known as a Knowledge Graph Card (KC)<sup>1</sup> appears for it in search engine results pages (SERP). The KC offers information about the entity directly in the search engine window, including such elements as address, phone number, hours of operation, description, link to the website, user reviews, etc. More importantly, the KC is an indicator that the search engine has achieved a machine-based comprehension of the existence and nature of the entity. With this understanding, the search engine can be more precise in its referrals and can hand off information about the entity to other semantic technologies. Far from being an end in itself, the display of an accurate and robust KC should simply be considered a positive indicator of SWI. Unfortunately, most academic organizations have not achieved SWI at this time.

A search engine displays a KC when it has gathered enough verifiable facts about an entity. Search engines gather some facts organically while indexing website documents. But verifiable facts are more likely to be harvested from proprietary knowledge bases such as Google My Business, and from knowledge bases on the

---

<sup>1</sup> A Knowledge Graph Card may also be known as a Knowledge Card, Knowledge Panel, or Information Panel.

Linked Open Data (LOD) cloud, such as Wikipedia and Wikidata. Academic organizations have the best chance of controlling their SWI by proactively creating and curating records in these knowledge bases.

This chapter will: (1) explain the significance of SWI; (2) describe a new library service developed at Montana State University that helps campus organizations implement SWI; and (3) demonstrate how SWI was successfully achieved in three case studies.

## Why is SWI Important?

The achievement of Semantic Web Identity implies that a search engine understands certain things about an entity, such as the nature of the enterprise, its relationships, and its location. “Understanding,” in this case, implies a machine-based comprehension of facts, which search engines can use to provide more accurate search results and help surface otherwise undiscovered relationships that may lead users to organizations they would not have thought relevant to their queries. That comprehension also helps search engines transfer that information to semantic technologies, such as mapping applications and intelligent personal assistants. A type of database known as a knowledge graph is used by search engines to “define, structure, and link hundreds of millions of entities...to improve the answers of their artificial personal assistants...” (Bernstein, Hendler, & Noy, 2016). Apple’s Siri, Amazon’s Alexa, Microsoft’s Cortana, or Google Now are examples of personal assistants that become more competent at providing accurate directions to an organization if they can tap into the verified information that search engines have assembled in their knowledge graphs.

### Higher education

Information-seeking behavior in higher education almost always begins (and often ends) with Internet search engines (Connaway & Dickey, 2010), but the effect of SWI in the discovery process is only gradually becoming understood. In academia, SWI tends to be most robust at the top levels of the institution; searching for the name of a university will usually yield search engine results that include the presence of a robust KC. However, that comprehension quickly diminishes as the search moves deeper into the institution to the level of college, department, research institute and center. KC for those lower-level organizations are often missing altogether, display a paucity of information, or even show an organization that is different from the one searched.

A lack of search engine comprehension of organizations can logically be assumed to result in fewer referrals. If the search engine doesn't understand the academic organization, then it is less likely to send its users to the physical or digital location, particularly if the search engine has other options in its knowledge graph. The diminished level of referral may negatively affect the attraction of research funding, faculty talent, and students. For instance, funding agencies that seek evidence that a university is engaged in specific research may fail to find a credible connection if the relevant research institute hasn't addressed its SWI. Students seeking a match for their study interests may likewise not be referred if a particular university department hasn't addressed its SWI, leaving search engines to rely on potentially inaccurate information that other sources may supply (DePianto, 2016). These interrelated factors could even negatively affect university rankings and reputation if research funding doesn't find its way to the institution, if student enrollment declines, or if faculty can't be recruited because they are not attracted by the university's stated research interests.

## SEO to SWI

### The Origins of Semantic Web Identity

SWI can trace its roots to the practice of search engine optimization (SEO), which developed in response to the competitive nature of commerce on the Web. SEO practices have proved necessary to assure use and justify the investment in locally-developed library digital repositories (Arlitsch & OBrien, 2013).

The search engine landscape was crowded in the late 1990's, but by the early twenty-first century Google had cemented its dominance and many search engines fell by the wayside ("Timeline of web search engines," 2016; Wall, 2017). Google continues to control two-thirds of the "explicit core search engine market" in the U.S. (comScore, Inc., 2016) and estimates range as high as 90% for Google's search engine market share in Europe (Meyer, 2015). Although Microsoft's Bing is a competitor and should not be discounted, Google's dominance is the driver behind most SEO and SWI efforts. SWI is an offshoot of SEO in that its aim is to optimize the way search engines interact with a digital presence, but SWI is specific to entities (people, places, things) in the environment of the Semantic Web.

### Semantic Web

Producing search results in previous generations of the World-Wide-Web was limited to algorithms that weighed a series of "signals" against strings of text. The latest generation of the Web is known as the Semantic Web, and it is commonly

understood as an entity-based environment in which machines interact with data records to better understand concepts, identities, and relationships.

Tim Berners-Lee and his colleagues formally introduced the concept of the Semantic Web in 2001 as an evolved version of the Web, where data and information could be processed automatically by computers that have access to structured collections of information (Berners-Lee, Hendler, & Lassila, 2001). These data records are stored in numerous knowledge bases that comprise a Linked Open Data (LOD) cloud intended “to enrich the Web with structured data” (Thalhammer & Rettinger, 2014). In this environment, search engine practices have evolved from the previous limitation of matching search queries to textual strings, to a new process in which queries are matched with verified facts about entities and their relationships. The Semantic Web promises greater accuracy of search results, often providing answers to search queries rather than a list of websites: “Search engines no longer only return documents — they now aim to return direct answers” (Vaish, Wyngarden, Chen, Cheung, & Bernstein, 2014). The phrase “strings to things” has come to serve as a metaphor for the transition to the Semantic Web (Singhal, 2012).

### Knowledge Graphs

Graph databases have existed for decades, but they have seen a resurgence in use as the Semantic Web has grown. The nodes, properties, and edges that characterize graph databases are more adept at organizing and quickly retrieving Semantic Web data (and data relationships) than the hierarchical system of tables in relational databases. Google’s Knowledge Graph and Bing’s Satori are examples of graph databases recently built by search engine companies. The graph databases are populated with facts about entities and their relationships, drawn from “the publication of interlinked datasets on the Web, in a form that enables people and computer programs to use these datasets for navigation, integration, and web-scale reasoning” (Bouquet, Stoermer, & Vignolo, 2012).

Search engines populate their knowledge graphs with entities that are defined and verified in knowledge bases the search engines trust. Some of these knowledge bases are proprietary (ex., Google My Business), while others are publicly available on the LOD cloud<sup>2</sup>. Over 1,000 knowledge bases currently comprise the LOD cloud, and many of these contain information about entities from which search engines can learn (Schmachtenberg, Bizer, & Paulheim, 2014). The knowledge bases considered to be the most significant are represented near the center of the LOD cloud, and

---

<sup>2</sup> Linking Open Data cloud diagram 2017, by Andrejs Abele, John P. McCrae, Paul Buitelaar, Anja Jentzsch and Richard Cyganiak. <http://lod-cloud.net/>

include DBpedia (Auer et al., 2007), Wikipedia (Lih, 2009), YAGO (Suchanek, Kasneci, & Weikum, 2007), the CIA World Fact Book (Central Intelligence Agency, 2015), and Wikidata (Erxleben, Günther, Kröttsch, Mendez, & Vrandečić, 2014). Freebase was once a significant source of information for Google's Knowledge Graph, but Google began shutting it down in 2014 (Butzbach, 2014) and facilitated a migration project that transferred Freebase data into Wikidata (Tanon, Vrandečić, Schaffert, Steiner, & Pintscher, 2016).

Google has acknowledged that its Knowledge Graph draws information from some LOD knowledge bases (Singhal, 2012; Sullivan, 2012), but it is difficult to know exactly which sources Google and Bing tap for their respective knowledge graphs. For instance, while Google acknowledges that its Knowledge Graph uses "public sources such as Freebase, Wikipedia and the CIA World Factbook" (Singhal, 2012), there is no evidence from Google that it draws data directly from DBpedia, which seems odd since DBpedia publishes rich structured linked data records extracted from Wikipedia entries. Google will only characterize its relationship to DBpedia as one of "transitivity," meaning that it is indirect and established only insofar as it draws from sources that do have direct relationships with DBpedia (Mendes & Jakob, 2012). Microsoft also only hints at its use of LOD knowledge bases, such as this quote from its patent application for a process it developed for entity detection and disambiguation: "...the entity-based search system recognizes particular content sources as authoritative sources for discovering entity information. For example, the system may identify Wikipedia as having particularly strong and trustworthy entity information and may recognize various pages at that site as describing entities" (Li, Li, Zhou, Lv, & Cao, 2013).

### Knowledge Graph Cards as Indicators of SWI

Google's Knowledge Graph and Bing's Satori database are invisible to the public, but the Knowledge Graph Card (KC) is one visible manifestation that Google began to display in 2012 (Singhal, 2012) in its search results. Bing followed a short time later with its own version of the KC. Aside from providing quick and easy "answers" about an organization to users, the display of a KC serves as an indicator that the search engine has discovered sufficient verified facts about the organization to establish it as an entity in its knowledge graph. The authors characterize this condition as Semantic Web Identity (SWI). Conversely, when a KC fails to appear for an organization, or it displays few or inaccurate facts, the condition may be characterized as lacking SWI, or poor SWI, respectively.

## Knowledge bases that help establish SWI

### Google My Business (GMB)

GMB is a proprietary knowledge base owned by Google, Inc. Registering or “claiming” a business or organization with GMB helps promote that organization and facilitates interactions with customers (Shenoy & Prabhu, 2016). Registering with GMB also has the wide-ranging effect of integrating with multiple Google properties. Registering with GMB may be the single most effective step an organization can take to begin achieving SWI. Registration consists of a two-step “claim and verify” process that helps Google assure the veracity of the claim and the location of the organization (Google, Inc., 2017).

### Wikipedia

Wikipedia is most useful to the Google Knowledge Graph as a source of descriptive text that is used to populate the “Description” element in the KC. Google appears to use the first sentence of the Wikipedia article verbatim for the description in the KC, so academic organizations would do well to craft that sentence very carefully. While Wikipedia does not explicitly forbid editors from writing about organizations or companies with which they are affiliated, it does publish community-generated guidelines and behavioral norms. Editors are encouraged to identify any conflicts of interest. Provided that an editor is able to maintain a neutral point-of-view, utilizes reliable and published sources, and demonstrates transparency about connections with subjects, it is possible for an organizational employee to write and edit Wikipedia content pertaining to that same organization.<sup>3</sup> Indeed, the iterative processes of researching, writing, and editing can provide opportunities for critical self-reflection which encourage editor(s) and organization to think thoughtfully about organizational history and strategic goals. Nevertheless, editing an article about one’s own organization is an activity that should be approached with caution.

### Wikidata

Wikidata is a sister project to Wikipedia (both are administered by the Wikimedia Foundation) and is fast becoming a significant source of structured data records. Launched in October 2012, it uses a crowdsourcing model to create and edit structured data records while reconciling data from various Wikipedia language versions. It currently contains over 27 million items and has over 2.9 million registered contributors, of which more than 17,000 are active (Wikimedia Foundation, Inc., 2016), and the data are exposed in machine readable formats such as JSON, XML and RDF (Vrandečić & Krötzsch, 2014). Wikidata administrators

---

<sup>3</sup> More information about Conflict of Interest editing in Wikipedia is available at [https://en.wikipedia.org/wiki/Wikipedia:Conflict\\_of\\_interest](https://en.wikipedia.org/wiki/Wikipedia:Conflict_of_interest).

acknowledge its role in helping to populate Google’s Knowledge Graph, but they are quick to point out that it does not replace Freebase. “Whereas Freebase was the open core of the Knowledge Graph, this is not true for Wikidata. Wikidata is one source of the Knowledge Graph among many, but does not have the same standing as Freebase had” (Wikidata, 2015). Some consultants maintain that Wikidata does influence KC results, but stress that there are no guarantees (Edward, 2015).

## SWI Environmental Scan

The 125 member organizations<sup>4</sup> of the Association of Research Libraries (ARL) were studied in 2015-16 to determine the condition of their SWI and to determine whether there was a correlation of SWI with the existence of records in certain knowledge bases (Arlitsch, 2017), (Arlitsch, 2016). In addition to searching Google to determine whether a KC was visible for each organization, searches were conducted for the presence of records for the organizations in five knowledge bases: Google My Business; Google+; Wikipedia; DBpedia; and Wikidata.

### Results

Data collected for this study demonstrate that there is room for significant improvement in the current state of SWI among ARL libraries, as measured by the presence or lack of accurate KC., and that most libraries have not created or maintained records for their organizations in Semantic Web knowledge bases.

The study results bring into focus the semantic difference between the concept of an entity and the name by which that entity must be located in a search engine. The names of ARL member libraries became an unexpected, but impactful part of this study. ARL libraries voluntarily submit a primary (official) name to ARL for inclusion in the ARL membership directory<sup>5</sup>. But 94 of the 125 member libraries also use alternate names<sup>6</sup>, and they use each of these names inconsistently and usually don’t explain the relationship of the names to search engines.

Google searches for only the primary names of library organizations yielded accurate KC just 46% of the time, while searches for only alternate names showed accurate KC 79% of the time. Combined, only 60% of the 219 primary and alternate names of the libraries displayed an accurate KC during Google searches. Clearly, there is a discrepancy in the primary names ARL libraries provide for lists like the

---

<sup>4</sup> The number of ARL member libraries declined to 124 in 2016, after the data collection had been completed.

<sup>5</sup> ARL membership directory: <http://www.arl.org/membership/list-of-arl-members>

<sup>6</sup> Alternate names are other names by which libraries refer to themselves. Often these are building names rather than organization names.

ARL membership directory and how they represent themselves in other fora on the Web.

The analysis becomes much more interesting with the introduction of data that recorded the “same as” relationship when two KC displayed for a given ARL member library. It can be stated that 82% of the ARL library organizations displayed an accurate KC when the results of primary or alternate names are taken into consideration. If the search engine understands that an organization has two different names, then a single KC will be displayed, regardless of whether the primary or the alternate name was being searched. However, the data show that only 37% of ARL libraries enjoy this “same as” status. When accurate KC display for both the primary and alternate name of a member library the two KC are usually not the same. In addition, the two different KC often display different facts about the same organization.

Approximately 11% of KC that displayed during searches were for the wrong organization. In most cases, this inaccuracy was recorded because a KC displayed for a branch library on campus rather than the main library or the library umbrella organization that was being searched. For example, one would expect the name “Boston College Libraries,” to be the umbrella name for all the libraries at Boston College, but using that search term in Google resulted in the display of a KC for the “Babst Art Library.” Similarly, a search for “Yale University Library” displayed a KC for Yale’s “Divinity School Library.”

The data also show a low presence of records in Semantic Web knowledge bases, which correlate to the lack of KC (see Table 1).

<b>Knowledge Base</b>	<b>Primary (% of 125)</b>	<b>Alternate (% of 94)</b>	<b>Total (% of 219)</b>
Google My Business	28 (22%)	40 (43%)	68 (31%)
Google Plus (unverified)	25 (20%)	17 (18%)	42 (19%)
Google Plus (verified)	22 (18%)	19 (20%)	41 (19%)
Wikipedia (w/o infobox)	10 (8%)	16 (17%)	26 (12%)
Wikipedia (w/infobox)	30 (24%)	26 (28%)	56 (26%)
DBpedia	30 (24%)	39 (41%)	69 (32%)
Wikidata	26 (21%)	37 (39%)	63 (29%)

Table 1: Number and percentage of records that exist in knowledge bases for primary and alternate names of ARL Libraries

## SWI as a Service

Surveys conducted by our research team as early as 2012 identified shortcomings in Montana State University Library's representation on the Semantic Web, as well as numerous other organizations on campus. In August 2015, we launched a new library-based service that establishes robust SWI from the top of the institution (colleges and departments) down to centers, institutes, and laboratories. The goal of the SWI service is to overcome ambiguity, incomplete description, and partial understanding of academic entities.

Methods developed during our research not only hold great promise for improving the SWI of academic entities, but also blaze an exciting new path of service-oriented research for academic libraries and librarians. In addition to describing the development and implementation of these new services, this section addresses reasons why academic libraries are ideally situated and academic librarians are uniquely qualified to engage in this "new knowledge work".

## Goals of an SWI Service

1. **Disambiguation** – Accurate and authoritative content in LOD sources decreases ambiguity. For example, Montana State University hopes that an open web query for "MSU College of Business" will provide search results featuring information about Jake Jabs College of Business and Entrepreneurship (JJCBE). However, this query is beset by ambiguity. Query language is incomplete,<sup>7</sup> the "MSU" abbreviation is used by various institutions,<sup>8</sup> a colloquial name is used, and "College of Business" is a commonly used name. SWI service improves machine comprehension by translating the query "MSU College of Business" from a string of terms into a specific thing in a particular place with unique relationships with other entities.
2. **Curation** – Creating and curating content allows machines to ingest information about an entity, including its name, location, type, function, characteristics, and its relationship to other entities. While many LOD sources are aided by community

---

<sup>7</sup> Montana State University does not have an academic entity with the official or colloquial name "College of Business." Instead, accounting, finance, marketing, and other business courses and degree programs are housed within the academic entity known as *Jake Jabs College of Business and Entrepreneurship*. Nevertheless, "College of Business" is common query language.

<sup>8</sup> In addition to Montana State University, the MSU abbreviation is commonly used by Manonmaniam Sundaranar University, Michigan State University, Mississippi State University, Missouri State University, Moldova State University, Morgan State University, and Murray State University, among others.

contributions, monitoring the content over time ensures accuracy and minimizes staleness.

3. **Search Results** – Robust, reliable, and contextual search results are a strong proxy indicator that an entity is well described within LOD sources and has achieved some degree of SWI. When entities are understood within contextual relationships, search results and machine-generated recommendations can be contextual and adaptive.
4. **Reputation** – SWI service also intends to credibly enhance the reputation of an academic entity. Increased discoverability and confirmed information about the organization increase credibility and trust, creating a virtuous cycle of increasing reputation.

## Demonstrating Value

Demonstrating value associated with an SWI service can be tailored to the specific interests of stakeholders. To maintain partner involvement and to clearly demonstrate the value of SWI service, MSU Library provides bi-monthly analytics briefs (i.e., a jargon-free slide deck that includes quantitative data, screenshots, and proposed next steps). Additionally, library staff routinely meet with stakeholders, share updates with university administration, present at conferences, publish research articles, etc. While modes, methods, and frequency of communication will vary as needed, there are two primary metrics that provide universal applicability for demonstrating value of SWI service: visibility and reputation.

### Visibility Measured via Web Analytics

Comprehensive web analytics data are used to demonstrate efficacy and impact of the SWI service. MSU Library collects and analyzes data from Google Analytics, Google Search Console, and Google My Business Insights.

**Google Analytics** – Widely used in academic libraries and various other industries, Google Analytics<sup>9</sup> is a powerful web analytics platform that tracks and reports numerous website traffic variables. The platform collects copious amounts of data including anonymized user demographics, geolocation, route to website, behaviors within website, etc.

---

<sup>9</sup> <https://analytics.google.com>

**Google Search Console (GSC)** – Data gathered from GSC indicate how Google search engine crawlers interact with the website, which in turn affects how the website appears in search results. Understanding what search engines know (or don't know) about the website, and what problems they encounter while crawling enables optimization of website performance in search results. GSC provides information about *search appearance* – presence and functionality of structured data, reports errors, suggests HTML improvements; *search traffic* – search terms, clicks, impressions, click through rate (CTR), position; Google search engine indexing, sitemapping, etc. It is important to note that GSC only provides a sliding 90-day window from which data can be accessed.

**Google My Business Insights (GMB)** – GMB Insights<sup>10</sup> provides data about how users find the entity<sup>11</sup> on the web and actions users take with provided information. After an entity has been claimed and verified, GMB Insights provides information about *how* (i.e., searching directly for entity name or searching for category, product, or service) and *where* (i.e., search or maps) users search for the entity. Additionally, GMB Insights indicates actions that users take on the GMB listing (i.e., visiting the website, requesting driving directions, making a phone call, viewing photos).

Consistent review of analytics data provides various quantitative measures of change in visibility of an entity. Moreover, thoughtful consideration and comparison of analytics data over time and among entities provides a data-rich guide for engaging in curation activities to further refine techniques and content.

## PROCESS

MSU Library has developed a systematic and iterative process for collaborating with academic units to create and curate content required by LOD sources. The process developed at MSU Library focuses on building partner relationships (*Outreach*), identifying any existing components of SWI (*Baseline*), populating (*Creation*) and maintaining (*Curation*) trusted LOD sources, and monitoring changes in entity visibility and reputation (*Analysis*). As depicted in Figure 1, data gathered during *Analysis* is likely to lead to the creation of additional content or guide the curation of existing content.

---

<sup>10</sup> <https://business.google.com/manage/>

<sup>11</sup> Given its commerce-focus, Google My Business refers to users as “customers” and entities as “business”. While Google My Business Insights provide an array of useful and informative data, application of such data for Semantic Web Identity services does require a certain degree of translation of variables.

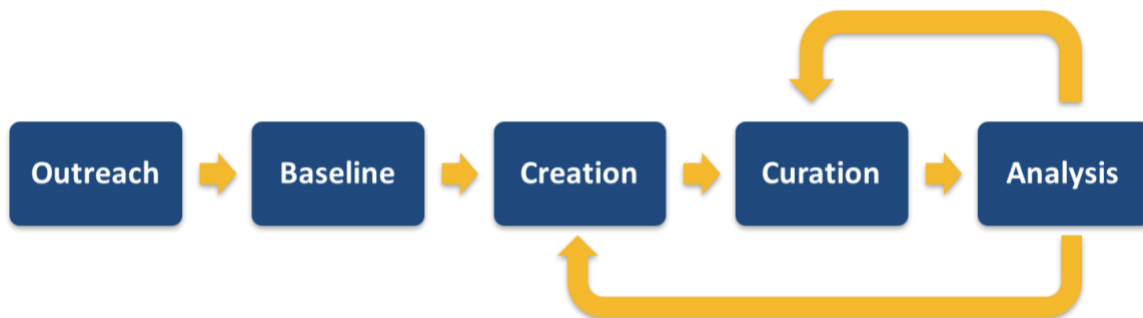


Figure 1: Iterative Process of Establishing and Curating Semantic Web Identity as Developed by Montana State University Library

To date, MSU Library has used this process to establish successful partnerships with 22 academic entities across campus (see Table 2). After successfully enhancing the SWI of MSU Library, our research team conducted a pilot with our first on-campus partner, Center for Biofilm Engineering<sup>12</sup> (CBE). Lessons learned from the CBE pilot led to further refinements to the SWI service process.

Pilots	Wave 1 Fall 2015	Wave 2 Spring 2016	Wave 3 Summer 2016	Wave 4 Fall 2016	Wave 5 Spring 2017
MSU Library	Campus Planning, Design and Construction	College of Agriculture	Office of Admissions	College of Arts and Architecture	Gallatin College
Center for Biofilm Engineering	Honors College	College of Education, Health and	Office of International Programs	School of Architecture	College of Nursing
	Jake Jobs College of Business and Entrepreneurship	College of Engineering	College of Letters and Science	School of Art	Western Transportation Institute
	Office of the Provost	MSU Extension		School of Film and Photography	
		Graduate School		School of Music	
		Food and Health Lab			

Table 2: List of Montana State University entities currently participating in MSU Library’s Semantic Web Identity Service

<sup>12</sup> [www.biofilm.montana.edu](http://www.biofilm.montana.edu)

## Outreach

Outreach is a critical first step in building a successful partnership and must continue throughout the duration of partnership. Maintaining open communication and providing partners with frequent updates helps to maintain partner involvement and clearly demonstrates the value of SWI service.

## Baseline

Before engaging in efforts to enhance the SWI of an academic entity, it is important to document what search engines and LOD sources already know, don't know, or think they know about the entity. Documenting SWI baseline is most easily accomplished via data download (when available) and screenshots (most common method). Baseline data and screenshots are used routinely during analysis, and data should be well-described, organized, and preserved.

## Creation

When machines better understand an entity, search engines are more likely to present relevant and actionable information to a search user. SWI service focuses on creating verifiable content in three trusted LOD sources.

## Google My Business

Creating a GMB listing is arguably the single most impactful step an academic organization can take to establish robust SWI. Since GMB is a proprietary database owned by Google, creation of a GMB property includes the wide-ranging effect of integrating with multiple Google platforms. For example, claiming and verifying a GMB property generates a verified Google+ profile.

Two paths can lead to the creation of a GMB listing: claiming an existing property or creating a new property. *Claim* – Begin by conducting an open web search for the entity. If search results display an organic KC, then clicking the “Own this business?” link will initiate the process of claiming the entity. *Create* – If search results do not display a KC, then it is necessary to navigate to GMB, search for the entity, and follow steps for creating a new GMB listing.

Verification of a new GMB listing is required for information to appear in other Google platforms. When claiming or creating an entity within GMB, a user is required to provide a physical address. GMB then mails a postcard containing a unique verification code to the entity's physical address. Entering the entity address and requesting a verification postcard is relatively straightforward. GMB expects most postcards to arrive within five days.

Academic addresses pose problems for timely delivery of GMB verification postcards. Academic addresses might be expressed as room number and building name, and a single academic building likely contains multiple entities. Campus mail systems regularly utilize campus-specific post office boxes featuring a distinct postal code.

Many campus mailing addresses operate independent from the physical street address of an entity. For these reasons, postcard delivery may require more than the five days expected by GMB.

When postcard delivery become significantly delayed, GMB is willing to provide alternate methods for verifying the authenticity of a listing. Having requested, but not received postcards for some entities, MSU Library has been able to verify GMB listings via phone conversations, emailed photographs of the entity, and Google Hangouts (i.e., livestreaming video). Certain entities may be eligible for phone, email, or instant<sup>13</sup> verification. GMB Help<sup>14</sup> can clarify verification processes or request alternative verification methods.

### Wikipedia

Creating an encyclopedic Wikipedia entry is an important component to establishing a robust SWI for an academic entity, but it may also be the most time-consuming component of the creation phase of SWI service, as it involves the iterative process of researching, writing, and editing. However, creating a stub entry may be enough to get experts in the community to help research, write, and polish the article, thereby leveraging the power of crowdsourcing.

At its most collaborative, creating Wikipedia content involves equal investment by the library and the academic partner. Equitable distribution of the workload tends to enable more efficient creation of encyclopedic content. Conversely, in a more centralized model, the library is predominantly responsible for the bulk of researching, writing, editing, and publishing activities with limited input from the academic partner.

No matter the degree of collaboration, it is the responsibility of the library (as SWI service provider) to expertly guide research, writing, editing, and publication of encyclopedic Wikipedia content. Representatives from the academic entity are most familiar with that entity and are therefore well positioned to identify source materials and provide tacit knowledge. However, creating encyclopedic content for Wikipedia differs in tone, purpose, and process from marketing and promotional content traditionally produced by academic entities. Wikipedia should not be used as a soapbox or means of promotion. Articles that do not maintain a neutral point-of-view or lack reliable sources are likely to be revised or deleted by the community of editors. Offering SWI service requires the library to provide expertise regarding Wikipedia's purpose, formatting, organization, and other norms. In short, the academic entity provides content knowledge and the library offers process and context expertise.

---

<sup>13</sup> GMB provides instant verification if the entity's website has previously been verified with Google Search Console. Given that academic entities tend to operate as website sub-domains within a centralized structure of IT and web administration, this verification method is generally unavailable.

<sup>14</sup> <https://support.google.com/business>

The researching and writing process tends to work most smoothly when the library uses its Wikipedia expertise to develop a richly descriptive template for the academic partner. With preliminary headings and suggested references, the template provides a framework to initiate the collaborative research and writing process. Research regarding Wikipedia best practices suggests that any entries about academic entities (see Figure 2) should, at minimum, include an infobox<sup>15</sup>, lead (i.e., abstract), and table of contents. Additional encyclopedic information (e.g., historical background, location and institutional affiliation, description of academic programs, distinguished faculty, notable alumni, and non-academic partnerships) should be appropriately situated among various sections of the entry.

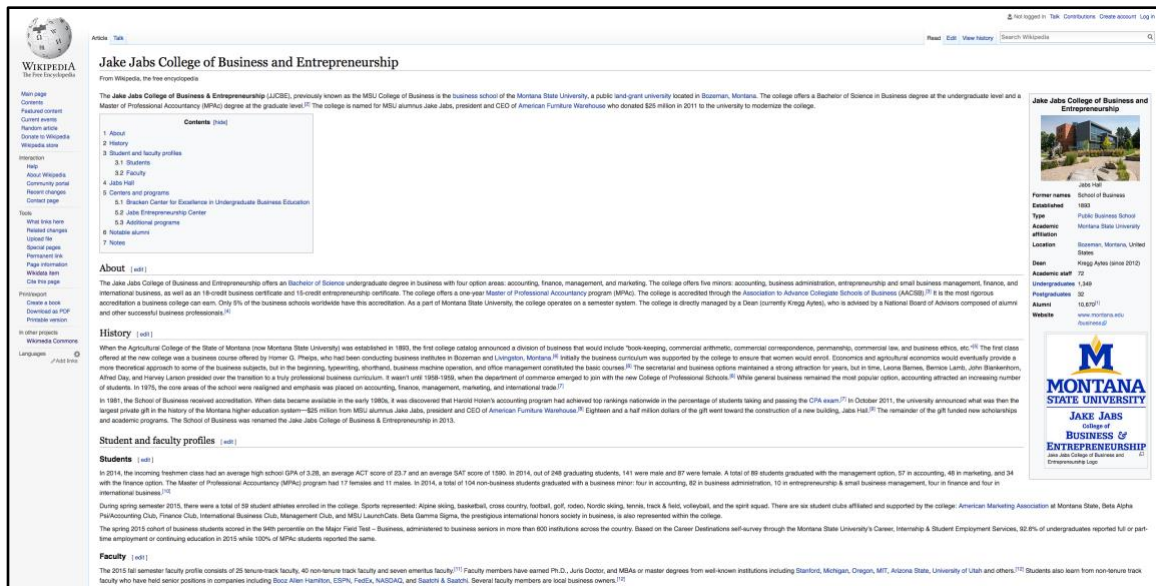


Figure 2: When creating a Wikipedia entry, it is important to abide by Wikipedia standards and select appropriate templates. As shown in the Wikipedia entry for “Jake Jabs College of Business and Entrepreneurship”<sup>16</sup>, important features include “Lead”, “Table of Contents”, and “Infobox” with important (and oft-used by LOD applications) structured statements and image/logo from Wikimedia Commons. This March 2017 screenshot displays a portion of the entry, but does not display all information.

The degree of collaboration will influence assignment of specific responsibilities during the creation of Wikipedia content. The SWI service recipient (i.e., academic partner) and SWI service provider (i.e., library) will collaboratively conduct research and co-write a draft. Thereafter, the library will provide editorial feedback based upon expert knowledge of Wikipedia principles and SWI best practices. Editorial feedback will result in revisions, additions, and redactions. As indicated in Figure 3, the creation of Wikipedia content is inherently iterative and necessarily connected with subsequent and ongoing efforts to curate content.

<sup>15</sup> There are various infobox templates that can be used when developing a Wikipedia entry. Research conducted by the SWI group at MSU Library has suggests that “Template:Infobox academic division” is most appropriate for presenting structured data about academic entities. More information is available at [https://en.wikipedia.org/wiki/Template:Infobox\\_academic\\_division](https://en.wikipedia.org/wiki/Template:Infobox_academic_division).

<sup>16</sup> [en.wikipedia.org/wiki/Jake\\_Jabs\\_College\\_of\\_Business\\_and\\_Entrepreneurship](https://en.wikipedia.org/wiki/Jake_Jabs_College_of_Business_and_Entrepreneurship)



Figure 3: Iterative Process of creating and curating Wikipedia content to enhance Semantic Web Identity of academic entities.

Those unfamiliar with writing or editing content for Wikipedia are encouraged to familiarize themselves with the encyclopedia’s fundamental principles.<sup>17</sup> Reviewing how-to guides, such as “Wikipedia:Your first article”<sup>18</sup> is an easy method for increasing your Wikipedia fluency. In general, all Wikipedia content must adhere to three core principles:

*Neutral Point of View* – fair, proportionate, and non-bias representation of significant views that have been published on a topic;

*Verifiability* – readers can confirm that information is drawn from reliable, published sources;

*No Original Research* – articles should not present new analysis, content must be attributable to reliable, published sources.

In addition to reviewing Wikipedia documentation, it is tremendously helpful to connect with the institution’s Wikipedia Visiting Scholar<sup>19</sup> (if available) and/or contact veteran Wikipedians for insight. These individuals will be able to provide expert feedback about scope, tone, and references for intended entries as well as guidance regarding Wikipedia norms. Also, be sure to investigate how the parent institution is represented within Wikipedia. Contact university communications if there is no Wikipedia entry for the institution or if the existing entry is insufficient. If available, be sure to utilize existing the Wikipedia template and/or category.

<sup>17</sup> [https://en.wikipedia.org/wiki/Wikipedia:Five\\_pillars](https://en.wikipedia.org/wiki/Wikipedia:Five_pillars)

<sup>18</sup> Available at [https://en.wikipedia.org/wiki/Wikipedia:Your\\_first\\_article](https://en.wikipedia.org/wiki/Wikipedia:Your_first_article), “Wikipedia:Your first article” provides a thorough overview of Wikipedia concepts, guidelines, processes, and tools.

<sup>19</sup> [https://en.wikipedia.org/wiki/Wikipedia:Visiting\\_Scholars](https://en.wikipedia.org/wiki/Wikipedia:Visiting_Scholars)

As a final note, it is important to acknowledge that Wikipedia is an ongoing, open, and collaborative project. Content is editable, by anyone and those contributing to the project should surrender the notion of control and embrace the idea of collaboration with the community of editors. While it is undeniably important to create the best possible (i.e., neutral, verifiable, authoritative, notable, descriptive, etc.) entry, it is equally critical to recognize that upon publication the entry will require recurring maintenance (see “Curation”). There is much that can be written about an academic entity. Resisting compulsions to write as much as possible is not only in the best interest of time and personnel resources, but also follows Wikipedia best practices. Wikipedia entries should contain enough information to demonstrate encyclopedic depth but avoid unnecessary detail that results in an information deluge.

### Wikidata

After successful publication of an encyclopedic Wikipedia entry, attention is next turned to the free and open knowledge base, Wikidata. Either by machine or Wikimedia community member, at some point after the publication of a Wikipedia entry, a connected Wikidata record will be created.

It is tempting to follow a *laissez-faire* approach to record creation and allow a Wikidata record to simply appear. However, there are two discernable benefits associated with proactively creating a Wikidata record immediately after successful publication of a Wikipedia entry. First, the proactive approach to Wikidata establishes a structured data record immediately after publication of a Wikipedia entry, providing a second source of verified information for search engine knowledge graphs. Second, the *laissez-faire* approach most often results in skeletal Wikidata records – containing only the most basic information. Although these records can eventually be enriched, it is more expedient to immediately create a highly descriptive and well-sourced Wikidata record via SWI service.

Creating a new Wikidata record requires three primary components. This information provides the foundation for structured data and appears in a table atop every Wikidata record (see Figure 4):

*Label* – most common name by which the entity will be known; only one label is allowed per record; it is recommended (but not required) that the Wikidata label be identical to the connected Wikipedia entry title

*Description* – non-biased information that briefly describes and disambiguates a label

*Aliases* – alternative names for the entity; an entity can have as many aliases for an item as necessary; this is a crucial field for establishing “same as” relationships for alternate names as discussed in the *SWI Environmental Scan* section earlier in this chapter.

Item Discussion

## Jake Jabs College of Business and Entrepreneurship (Q23303133)

Jake Jabs College of Business & Entrepreneurship is the business school of Montana State University in Bozeman, MT. Named for alumnus [Jake Jabs](#), the college offers Bachelor of Science in Business & Master of Professional Accountancy.

Montana State University College of Business | Montana State University Jake Jabs College of Business and Entrepreneurship | MSU Business School | Montana State University Business School | MSU College of Business | MSU Jake Jabs College of Business and Entrepreneurship

[In more languages](#) *Configure*

Language	Label	Description	Also known as
English	Jake Jabs College of Business and Entrepreneurship	Jake Jabs College of Business & Entrepreneurship is the business school of Montana State University in Bozeman, MT. Named for alumnus Jake Jabs, the college offers Bachelor of Science in Business & Master of Professional Accountancy.	Montana State University Colleg... Montana State University Jake J... MSU Business School Montana State University Busine... MSU College of Business MSU Jake Jabs College of Busin...
Spanish	No label defined	No description defined	
Traditional Chinese	No label defined	No description defined	
Chinese	No label defined	No description defined	

**Statements**

instance of [business school](#) [inception](#) 1893 [edit](#)

+ 1 reference [+ add](#)

country [United States of America](#) [edit](#)

+ 1 reference [+ add](#)

location [Bozeman](#) [edit](#)

located at street address [302 Jabs Hall Montana State University Bozeman, MT 59717-3040](#) [edit](#)

+ 1 reference

[Gallatin County](#) [edit](#)

- 0 references [+ add reference](#)

[Montana](#) [edit](#)

- 0 references [+ add reference](#)

[United States of America](#) [edit](#)

- 0 references [+ add reference](#)

[+ add](#)

Wikipedia (1 entry) [edit](#)

en [Jake Jabs College of Business and Entrepreneurship](#)

Wikibooks (0 entries) [edit](#)

Wikinews (0 entries) [edit](#)

Wikiquote (0 entries) [edit](#)

Wikisource (0 entries) [edit](#)

Wikiversity (0 entries) [edit](#)

Wikivoyage (0 entries) [edit](#)

Other sites (0 entries) [edit](#)

Figure 4: As depicted in this record for “Jake Jabs College of Business and Entrepreneurship”<sup>20</sup>, Wikidata records are most often associated with Wikipedia entries (but can also exist without a corresponding Wikipedia entry), present a short textual description, contain alternative names, and display structured data statements about an academic entity. This March 2017 screenshot shows a partial Wikidata record, but does not display all “Statements”, “Identifiers”, or “References” for the entity.

In addition to the three primary components, a robust Wikidata record consists of various statements expressed as property-value pairs. Each property is populated with predefined fields. A singular property can be paired with multiple values. Statements can be further enriched with additional details (i.e., qualifiers) and supported by references. When creating a Wikidata record for an academic entity, some of the more common statements include: “instance of”, “location”, “parent organization”, “official website”, “Commons Category”, as well as various “Identifiers.”<sup>21</sup>

<sup>20</sup> [www.wikidata.org/wiki/Q23303133](http://www.wikidata.org/wiki/Q23303133)

<sup>21</sup> Wikidata allows users to input statements that correspond to various types of external authority control. Wikidata automatically creates a separate section (titled “Identifiers”) that groups all authority control statements. Some common external identifiers for academic entities include: Facebook ID, Google+ ID, ISNI,

## Curation

Populating trusted LOD sources with accurate, authoritative, and consistent information is a significant step toward the establishment of robust SWI for an academic entity. An academic entity's SWI can only be as accurate, authoritative, consistent, and useful as the information upon which it is built. Given the ever-evolving nature of information, it is necessary to attentively and recurrently curate LOD sources to ensure ongoing accuracy of information.

Curation of LOD sources can be grouped into two categories:

**Respond** – It is necessary for SWI service team to either confirm or correct changes suggested by other users within Google My Business, Wikipedia, or Wikidata. To assist with responsive curation, each of the three LOD sources mentioned above provides various notification methods. The GMB dashboard provides an “Account Summary” that indicates status of listings as well as a “To-do” list that identifies problem areas or suggested changes for all listings. Both Wikipedia and Wikidata allows users to add pages to their “Watchlist”, which tracks recent changes made to watched pages. Wikipedia and Wikidata users can opt-in to receive email notifications when any watched page is changed.

**Update** – Academic entities morph, merge, and otherwise evolve. An academic entity's SWI can only be as robust as the underlying information. As new information becomes available about an academic entity and its relationships, it is important for the SWI service to update appropriate LOD sources. Although the library can carefully watch for updates, changes, or news from academic partners, it is more effective to receive updates directly from SWI service recipients. An academic entity is in the best position to know about any changes pertinent to its description and/or relationships.

It is possible for the library and academic partner to share some curation responsibilities. An academic partner may have resources to curate its GMB listing – updating special hours or adding new photos. The library is best suited to maintain structured data in Wikidata. Specifics of the collaboration curation agreement will vary with each partnership. SWI service tends to operate most efficiently and generate the greatest impact when leveraging the combined strengths of a fully collaborative partnership.

When developing the SWI service model, MSU Library made the strategic decision to locate SWI activities with our Department of Resource Description and Metadata Services (RDMS). There is no doubt that our SWI service leverages skills, knowledge, and personnel from other areas of the organization. However, as information curators, database managers, and controlled vocabulary experts, RDMS possesses

---

Instagram username, Ringgold identifier, Twitter username. A full list of Wikidata properties with data type “external identifier” is available at <https://www.wikidata.org/wiki/Special:ListProperties/external-id>.

expertise that ideally aligns with description and curation of information, which aid in the discovery of entities.

## CASE STUDIES

The following case studies demonstrate how MSU's SWI service helped achieve SWI for three academic entities. The first case study addresses the internal application of SWI service to MSU Library. The subsequent case studies examine the Jake Jabs College of Business and Entrepreneurship (JJCBE) and the Honors College, which were two of the earliest adopters among SWI service recipients (see Table 1 for a full list of academic partners at MSU). JJCBE and Honors College at MSU differ in their academic disciplines, organizational history, and student bodies, but both entities saw the strategic benefits associated with SWI service.

### Montana State University Library

#### *About*

Montana State University Library serves the students, faculty, and surrounding community of Montana State University, a land-grant research university whose flagship campus is located in Bozeman, MT, USA.

#### *Baseline*

In November 2012, a Google search for "Montana State University Library" revealed a surprising entry in Google's Knowledge Graph display. Instead of displaying the flagship library of the Montana State University system, located in Bozeman, the search results page included a Knowledge Card for a branch campus academic library located on another MSU campus in Billings, MT (see Figure 5). In early 2013, several MSU Library staff and faculty initiated the first formal steps that would eventually lead to dramatically improved SWI for the MSU Library and eventually give rise to our SWI service model.

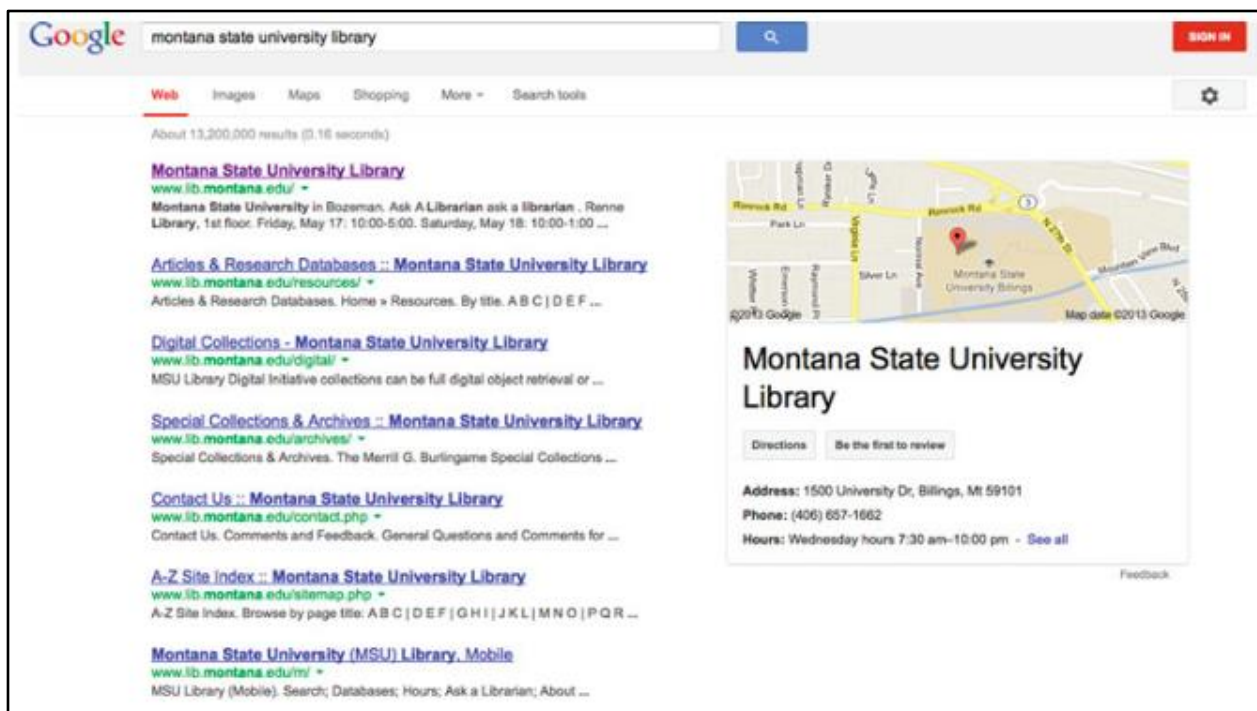


Figure 5: November 2012 screenshot of Google search results page for “Montana State University Library”

From the perspective of the Google search engine, the MSU Library in Bozeman simply was a text string and did not exist as a “thing” (entity). With the exception of Freebase, no records existed for the organization in knowledge bases (Table 3). The information being ingested by the Knowledge Graph led Google to incorrectly conclude the MSU library was a building in Billings, Montana.

Google Knowledge Graph Card	Displayed incorrect organization and incorrect location
Google My Business	Neither claimed nor verified
Google+	Two profiles existed. Neither profile verified.
Wikipedia	No entry
DBpedia	No record
Wikidata	No record <sup>22</sup>
Freebase	A record for Renne Library (the alternate name for the MSU Library) was created on March 10, 2012 by someone unknown to the MSU Library. There was no evidence of a record for Montana State University Library.

Table 3: Summary of SWI Conditions – Montana State University Library, January 2013

### Process

The process to establish SWI at the MSU Library began in early 2013. Actions to establish an authoritative and visible SWI are summarized in Table 4.

<sup>22</sup> Launched in October 2012, Wikidata was initially very limited in its scope (Vrandečić and Krötzsch 2014). Given the timing and initial aim of Wikidata, there is little reason to expect that Wikidata would contain a record for MSU Library in January 2013.

Google My Business	Business claimed and the record improved by August 2014.
Google+	Officially verified in August 2014. Duplicate Google+ profile successfully deleted.
Wikipedia	Research and writing began in the spring of 2013 under the guidance of experienced Wikipedia editors. The article was published on September 5, 2013.
DBpedia	Record appeared in April 2014 data release.
Wikidata	Record generated from Wikipedia by a bot on November 29, 2013.
Freebase <sup>23</sup>	A new Freebase record was auto generated on September 10, 2013, five days after the publication of the Wikipedia article. The record was titled <i>Montana State University Library</i> and was generated by a bot called "wikirecon_bot." MSU Library faculty added a "same as" declaration to the <i>Renne Library</i> Freebase record that linked it to the <i>Montana State University Library</i> Freebase record.
Google Knowledge Graph Card	An accurate KC began to appear for the MSU Library in September 2013, and gradually evolved to become much more robust as other knowledge bases were populated.

Table 4: Summary of SWI Actions and Results – Montana State University Library

### Current Conditions

Populating trusted LOD sources with authoritative information about MSU Library has resulted in a search results page (Figure 6) markedly improved from the November 2012 screenshot (Figure 5).

In addition to the improved website and social media content found in the left column, the right column of the search results page now presents a very robust KC for MSU Library. Based upon information added to the knowledge bases, the search engine is able to determine that the query "montana state university library" refers to the academic entity located within the building known as Renne Library on the campus of Montana State University. Within the KC, search users are presented information about MSU Library, including: physical location, contact information, type of entity, short description (ingested from Wikipedia), map location, verified social media profiles, etc. Information presented within the KC is actionable (e.g., visit website, get directions, call phone number), indicates a high degree of interoperability (pulling data from GMB, Wikipedia, Wikidata, social media, etc.), and seeks to answer likely questions (e.g., How do I contact MSU Library? Is MSU Library currently busy? What do other users think about MSU Library?). Creation of authoritative content within trusted LOD sources facilitated machine learning, which in turn presents more robust, authoritative, and useful information to human search users. In short, the query "montana state university library" has evolved from a search of a string of

<sup>23</sup> Freebase was acquired by Google in 2010. In 2014 Google announced that it would cease building Freebase in favor of Wikidata (Butzbach, 2014). Therefore, MSU Library is the only instance of SWI service that incorporates Freebase. All subsequent partnerships have removed Freebase from the SWI service process. Migration of Freebase records to Wikidata was completed in 2016 (Tanon, Vrandecic, Schaffert, Steiner, & Pintscher, 2016). Freebase is now available in "read-only" mode at <https://developers.google.com/freebase/>.

words to a search for a thing in a specific place with particular relationships to other things.

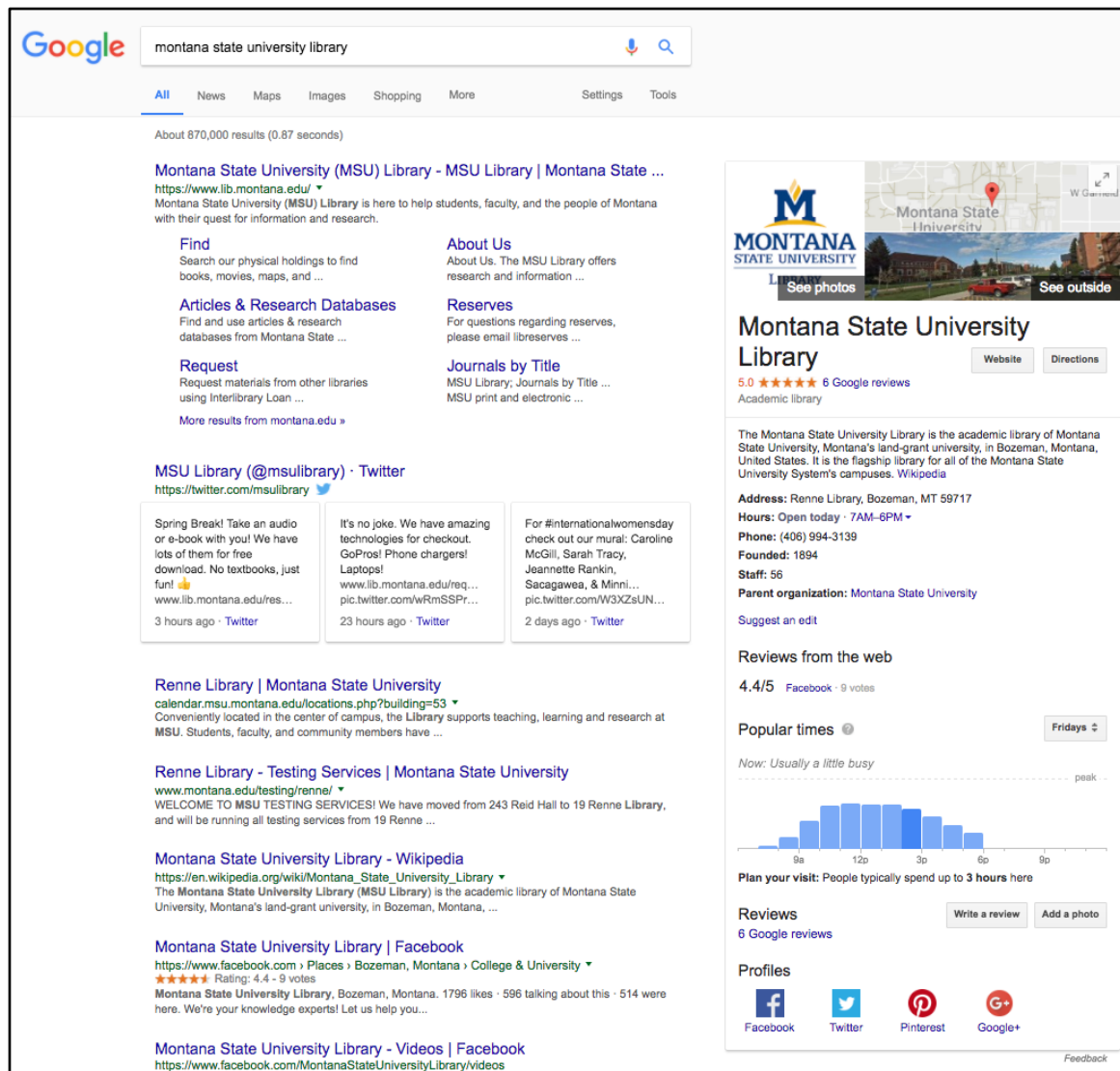


Figure 6: March 2017 screenshot of Google search results page for "montana state university library"

## Jake Jobs College of Business and Entrepreneurship

### About

Previously known as the MSU College of Business, the Jake Jobs College of Business and Entrepreneurship (JJCBE) is the business school of Montana State University.

### Baseline

Prior to SWI service, JJCBE was understood by search engines as a string of search terms, but not as an entity. Due to this lack of understanding, an open web search for "montana state university college of business" returned sparse, ambiguous, and

loosely related results. As displayed in Figure 7, the Google search results page features discrete bits of data functioning within standalone systems that do not readily communicate. There is some useful information, including a nicely site-mapped entry for JJCBE. However, search users are presented with unauthoritative information, which leads to ambiguity and uncertainty. After the initial result, the search user is left to wonder about the relevance of information about “Jabs Hall” or “Montana State University: College of Nursing”. These results are influenced by absent and incomplete knowledge base records, as summarized in Table 5.

Google Knowledge Graph Card	None
Google My Business	None
Google+	None
Wikipedia	No entry
Wikidata	No record
DBpedia	No record
Freebase	No record

*Table 5: Summary of SWI Conditions – Jake Jabs College of Business and Entrepreneurship, August 2015*

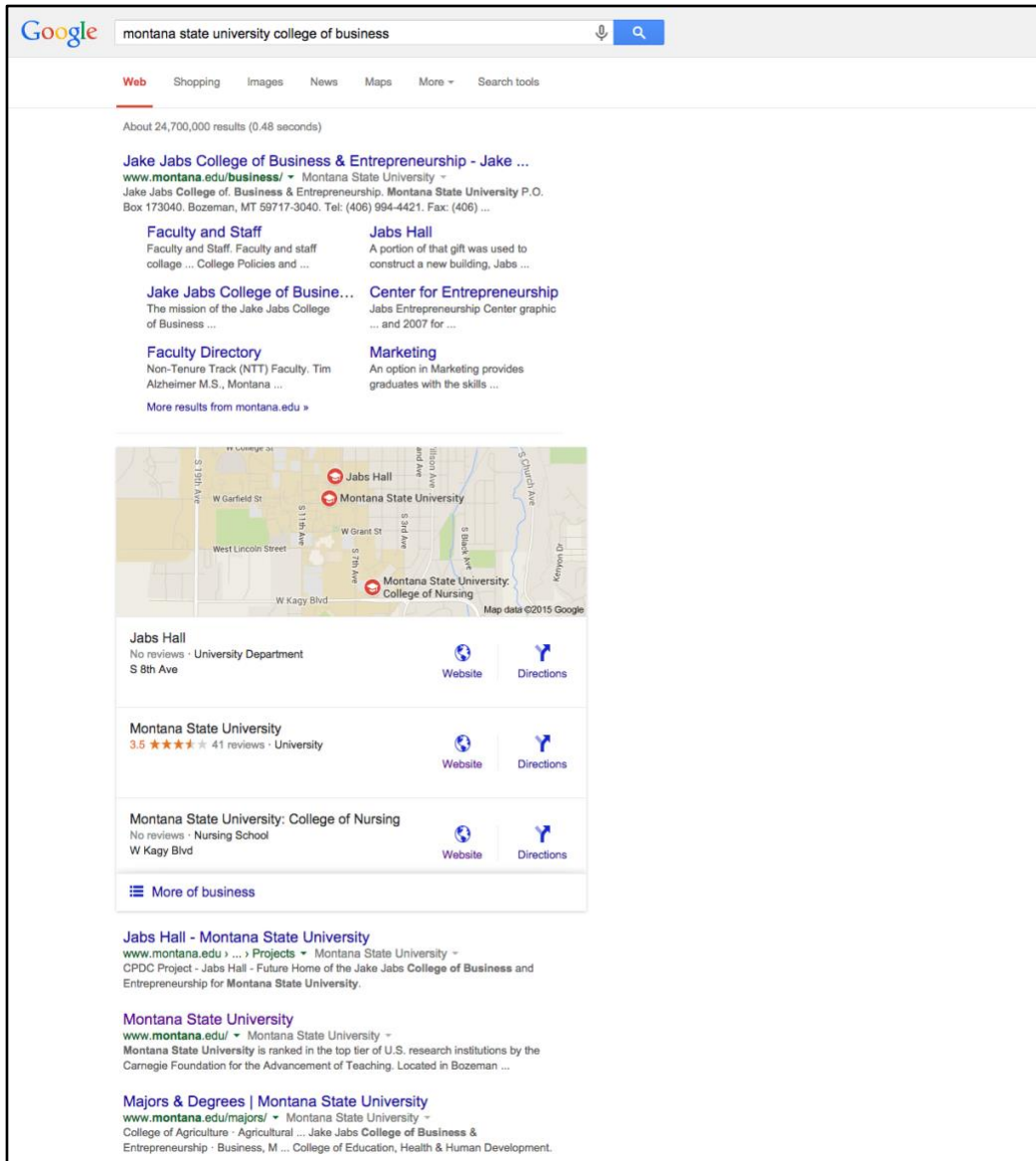


Figure 7: August 2015 screenshot of Google search results page for “montana state university college of business”

### Process

The process to establish SWI for JJCBE began in August 2015. JJCBE was the first academic partner to receive SWI service based upon the iterative process explained earlier (Figure 1). Actions to establish an authoritative and visible SWI are summarized in Table 6.

<p>Google My Business</p>	<p>Business claimed in late August 2015. Due to delays in receiving GMB postcard, the listing was not verified until November 2015. After two failed attempts to receive a verification postcard, an alternate verification method was arranged via Google My Business Help. JJCBE GMB listing was</p>
---------------------------	--

	verified through a combination of phone calls and emailed photos of the entity's physical location.
Google+	Officially verified in September 2015.
Wikipedia	Research and writing began in the August 2015 under the guidance of MSU Library's SWI Researcher and with assistance from MSU Library's Wikipedia Visting Scholar. A staff member from JJCBE worked in close collaboration with MSU Library to research, write, and revise the initial content. The article was published on February 16, 2016.
DBpedia	Record appeared in April 2016 dataset release.
Wikidata	Record created by user:Danrok on March 20, 2016.
Google Knowledge Graph Card	An accurate KC began to appear for JJCBE in November 2015 (after the GMB listing was verified). The JJCBE KC gradually evolved to become much more robust as other LOD sources were populated with authoritative information.

*Table 6: Summary of SWI Actions and Results – Jake Jabs College of Business and Entrepreneurship*

### *Current Conditions*

As displayed in Figure 8, the post-SWI service search results page looks considerably different than the pre-SWI service screenshot (Figure 7). The initial result item (a site-mapped website entry) is retained, but is now complimented by other authoritative, robust, and actionable information. The left column contains links to contextually associated websites, popular news articles, Wikipedia content.

The right column of the search results page now presents a robust KC for JJCBE, including: physical location, contact information, type of entity, short description (ingested from Wikipedia), map location, and verified social media profiles. Information presented within the KC is actionable (e.g., visit website, get directions, call phone number) and indicates a high degree of interoperability (pulling data from GMB, Wikipedia, Wikidata, etc.). In the event that the search user mistyped the query or was looking for another entity but could not quite recall the name, the JJCBE KC also presents some related entities that might be of interest.

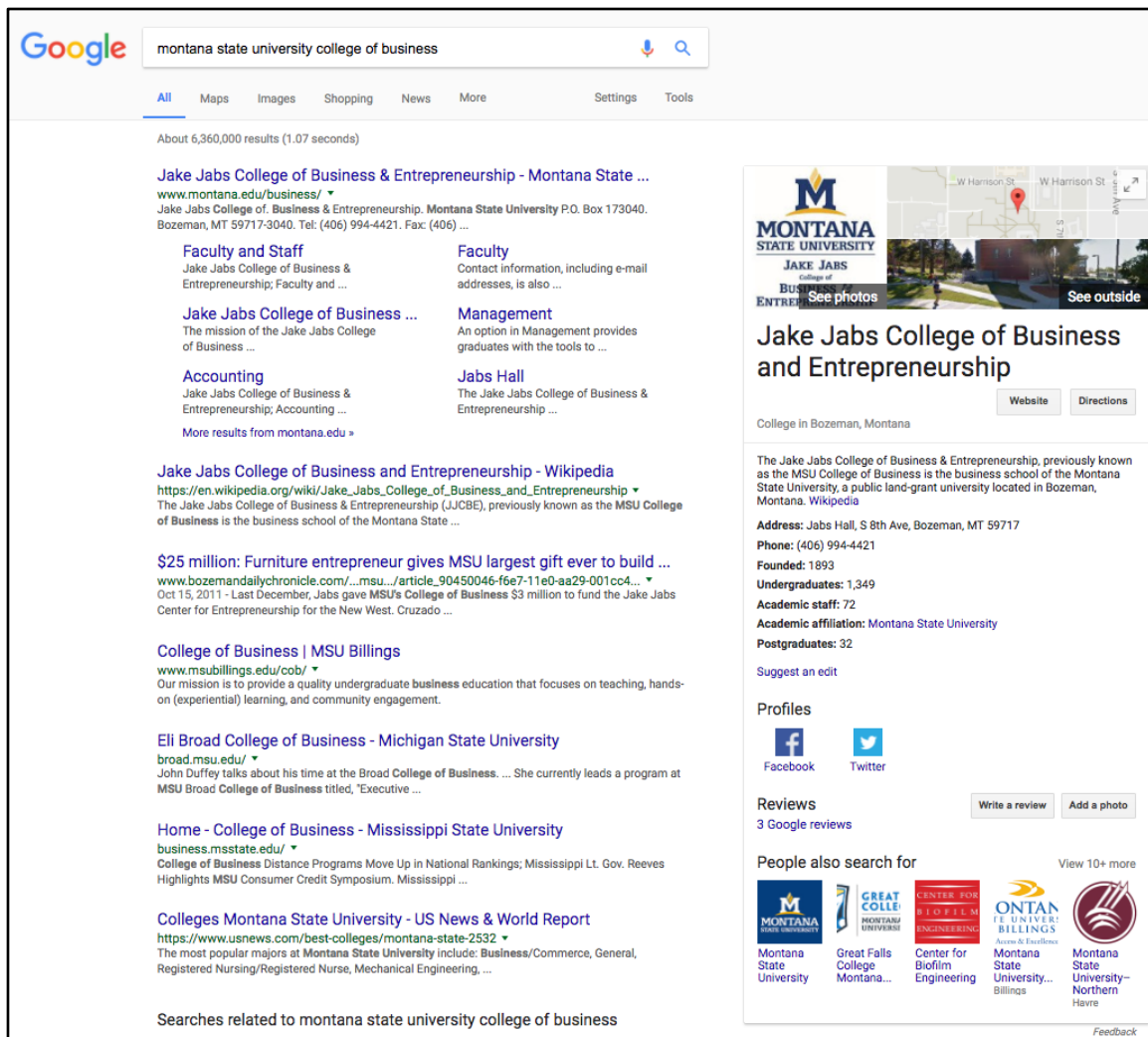


Figure 8: March 2017 screenshot of Google search results page for "Montana State University College of Business"

## Honors College at Montana State University

### About

Originally established in 1964 as an honors program and elevated to college status in 2013, the Honors College at Montana State University now provides enriched academic opportunities for more than 1,300 MSU students annually.

### Baseline

As displayed in Figure 9, the Google search results page featured some useful information, yet much more information remained buried and unavailable to search users. Aside from clicking one of the provided links and then searching a different website for the desired bits of information, search users are presented with few options. These results are influenced by absent and incomplete records in knowledge bases trusted by the search engine as summarized in Table 7.

Google Knowledge Graph Card	None
Google My Business	None
Google+	None
Wikipedia	No entry
Wikidata	No record
DBpedia	No record

Table 7: Summary of SWI Conditions – Honors College at Montana State University, August 2015

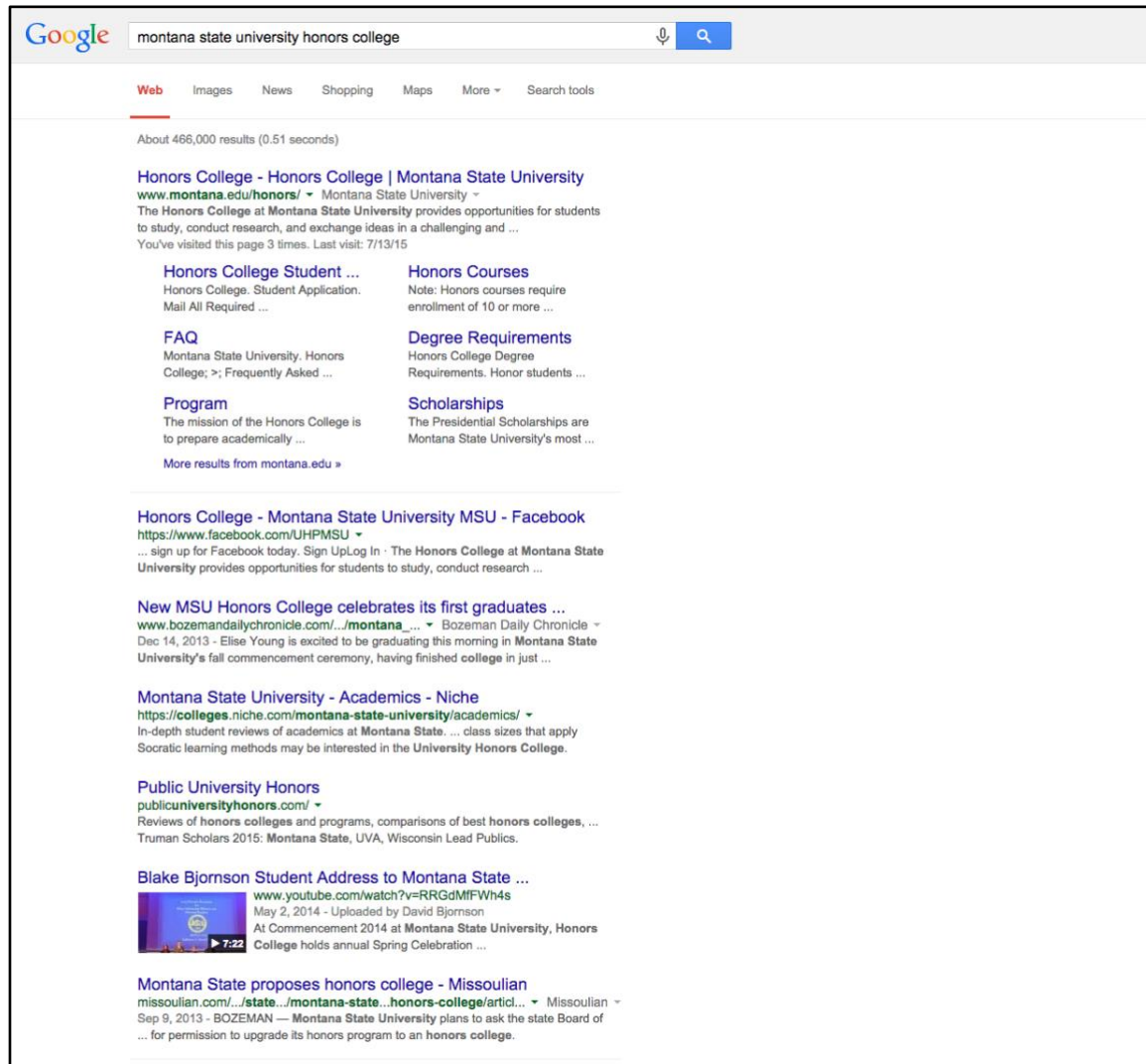


Figure 9: August 2015 screenshot of Google search results page for “Montana State University Honors College”

## Process

The process to establish SWI for Honors College at MSU began in August 2015. The Honors College at MSU was among the first cohort of academic partners to receive SWI service. Actions to establish an authoritative and visible SWI are summarized in Table 8.

Google My Business	Business claimed in September 2015. Due to delays in receiving the GMB postcard, the listing was not verified until November 2015. After two failed attempts to receive a verification postcard, an alternate verification method was arranged via Google My Business Help. The Honors College at MSU GMB listing was verified through a combination of phone calls and emailed photos of the entity's physical location.
Google+	Officially verified in October 2015.
Wikipedia	Research and writing began in the August 2015 under the guidance of MSU Library's SWI Researcher. Due to time and personnel constraints, the Honors College at MSU could only provide editorial comments on the initial Wikipedia content. The article was published on July 11, 2016. Since publication of the Wikipedia article, staff and faculty from the Honors College at MSU have provided MSU Library with information updates to aid in curation of Wikipedia content.
DBpedia	The most recent DBpedia dataset release occurred in April 2016. This dataset release contains Wikipedia data dating from March/April 2016. Since the Honors College at MSU Wikipedia article was not published until July 2016, there is no record available in any existing DBpedia dataset releases.
Wikidata	Record created by user:Jdshanks (MSU Library's SWI Researcher) on July 19, 2016.
Google Knowledge Graph Card	An accurate KC began to appear for Honors College at MSU in November 2015 (within 24 hours of verifying the GMB listing). The Honors College at MSU KC gradually evolved to become much more robust as other LOD sources were populated with authoritative information.

*Table 8: Summary of SWI Actions and Results – Honors College at Montana State University*

### Current Conditions

After application of the SWI service to the Honors College at MSU, an open web search for “montana state university honors college” now provides search users with a rich assortment of authoritative and actionable information.

As displayed in Figure 10, the post-SWI service search results page looks considerably different than the pre-SWI service screenshot (Figure 9). The initial result item (a site-mapped website entry) is retained, but is now complimented by other authoritative, robust, and actionable information. The left column contains links to contextually associated websites, popular news articles, Wikipedia content.

In addition to the website links found in the left column, the right column of the search results page now presents a robust KC for the Honors College at MSU, with similar information elements as the MSU Library and the JJCBE.

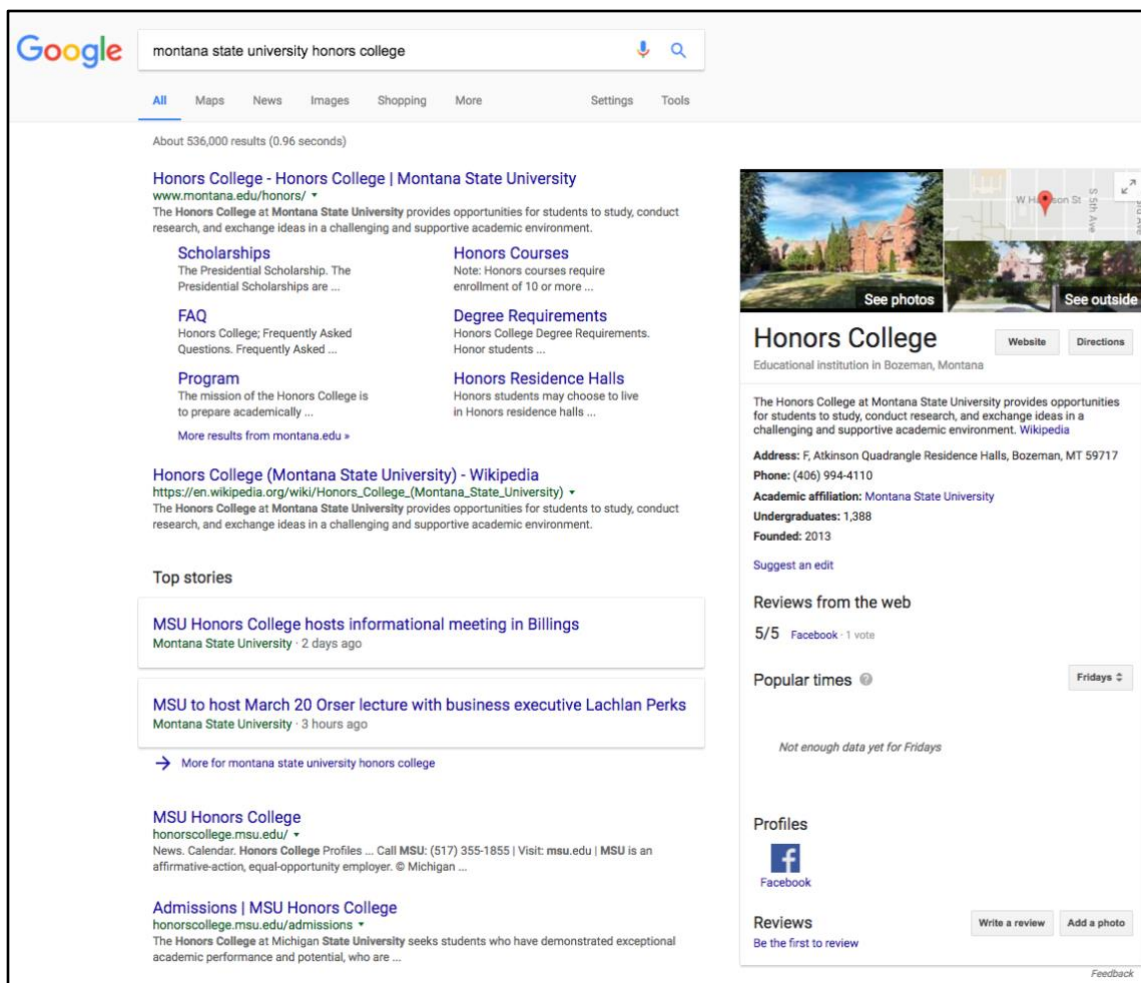


Figure 10: March 2017 screenshot of Google search results page for "Montana State University Honors College"

## Summary of case studies

In each of the three cases presented above, the iterative process outlined in Figure 1 was applied. Each case resulted in clear improvements to existing KC, or creation of new KC, with accurate information elements and interoperable features. Nearly twenty other organizations at Montana State University are participating in the SWI service. Each instance has yielded similar results, clearly demonstrating that the approach of creating or improving data records in Google My Business, Wikipedia, and Wikidata, works.

## Why is the library best suited to provide an SWI service?

At first glance, components of the SWI service may appear to overlap with the aims and efforts of other divisions commonly operating within most contemporary universities. Some readers may wonder why a university's information technology (IT) office does not monitor DBpedia records or Wikidata records. Or perhaps why a university's communications or marketing personnel are not engaging with Google+,

Wikipedia, or GMB. Indeed, it is common for a university's communications and/or marketing personnel to manage a university-level Wikipedia entry, GMB listing, and/or Google+ profile. However, such efforts are focused at top-level identity of an academic institution and tend to have little trickledown benefit for colleges, departments, centers, institutes, or laboratories. Moreover, ensuring visibility of all academic entities on campus requires careful creation and consistent curation of LOD content. Working in isolation, IT or communications/marketing units are unlikely to have the resources or knowledge to successfully and consistently deploy SWI service for all academic entities across campus.

Given its role as a centralized, multi-disciplinary information hub on college and university campuses, the academic library is well positioned to liaise and collaborate with other academic units to implement an SWI service. Moreover, with expertise as information scientists, data curators, and digital application developers, academic librarians are well positioned to offer services that carefully populate and consistently curate structured data to ensure a robust, contextual, and authoritative SWI across their institution. Creating and curating structured metadata records is the business of libraries, even (or perhaps we should say "especially"), on the Semantic Web.

## CONCLUSION

Semantic Web Identity is an exciting opportunity to develop new library services that are expected to become highly valued by the campus community. Driving more relevant user traffic to products and services is important to any organization trying to distinguish itself and increase its value proposition. Increased research funding, student enrollment, faculty recruitment, and institutional reputation and ranking are goals of great importance to higher education administrators. Libraries can provide technical solutions that address these strategic aims, and the SWI process established by the MSU Library has proven that libraries can be successful in this area.

Establishing SWI has boosted the visibility of MSU, its academic entities, and its scholarly products, and this visibility is expected to enhance MSU's reputation and positively influence student enrollment and grant funding. In addition to data collected from web analytics, MSU Library is currently developing assessment methods to measure influence of the SWI service on metrics of significance for our university partners. Combining web analytics with data collected by MSU partners (e.g., Office of Admissions, Alumni Foundation, Office of Sponsored Programs, etc.) will demonstrate how the SWI service increases actionable events such as student applications, donations, and grant funding.

While solving the technical problems associated with this issue can be interesting, viewing SWI only from a technical standpoint misses the greater picture. The incredible advances in computational and network power have led to the Internet of Things, where machines converse with each other's data records to surface robust and actionable information. Those machine-based interactions are dependent on accurate and verified structured data records, which librarians have traditionally curated in library systems and can do so again on the Semantic Web in trusted LOD sources like Google My Business, Wikipedia, and Wikidata.

## References

- Arlitsch, K. (2016, November 19). Data set supporting the dissertation “Semantic Web Identity in Academic Organizations: Search engine entity recognition and the sources that influence Knowledge Graph Cards in search results.” Montana State University ScholarWorks. Retrieved from <https://doi.org/10.15788/M2F590>
- Arlitsch, K. (2017, January 11). *Semantic Web Identity of academic organizations – search engine entity recognition and the sources that influence Knowledge Graph Cards in search results* (Dissertation). Humboldt-Universität zu Berlin, Berlin, Germany. Retrieved from <http://edoc.hu-berlin.de/docviews/abstract.php?lang=ger&id=43177>
- Arlitsch, K., & O'Brien, P. S. (2013). *Improving the visibility and use of digital repositories through SEO*. American Library Association.
- Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., & Ives, Z. (2007). DBpedia: A Nucleus for a Web of Open Data. In *The Semantic Web* (Vol. 4825, pp. 722–735). Berlin, Heidelberg: Springer Berlin Heidelberg. Retrieved from [http://link.springer.com/10.1007/978-3-540-76298-0\\_52](http://link.springer.com/10.1007/978-3-540-76298-0_52)
- Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web. *Scientific American*, 284(5), 34–43. <https://doi.org/10.1038/scientificamerican0501-34>
- Bernstein, A., Hendler, J., & Noy, N. (2016). A new look at the semantic web. *Communications of the ACM*, 59(9), 35–37. <https://doi.org/10.1145/2890489>
- Bouquet, P., Stoermer, H., & Vignolo, M. (2012). Web of Data and Web of Entities: Identity and Reference in Interlinked Data in the Semantic Web. *Philosophy & Technology*, 25(1), 5–26. <https://doi.org/10.1007/s13347-010-0011-6>
- Butzbach, A. (2014, December 9). Freebase is shutting down - What does it mean for the Knowledge Graph and SEO? Retrieved from <http://www.brafton.com/news/freebase-shutting-mean-knowledge-graph-seo/>
- Central Intelligence Agency. (2015). CIA World Factbook [Government]. Retrieved from <https://www.cia.gov/library/publications/the-world-factbook/>
- comScore, Inc. (2016, March 16). comScore releases February 2016 U.S. desktop search engine rankings [Commercial]. Retrieved from <https://www.comscore.com/Insights/Rankings/comScore-Releases-February-2016-US-Desktop-Search-Engine-Rankings>
- Connaway, L. S., & Dickey, T. J. (2010). *The digital information seeker: report of findings from selected OCLC, RIN, and JISC user behavior projects* (p. 61). Dublin, Ohio: OCLC Research. Retrieved from

- <http://www.jisc.ac.uk/media/documents/publications/reports/2010/digitalinformationseekerreport.pdf>
- DePianto, S. (2016, September 30). Helping prospective students make decisions about their future. Retrieved from <https://blog.google/topics/education/helping-prospective-students-make-decisions-about-their-future/>
- Edward, T. (2015, May 1). Leveraging Wikidata to gain a Google Knowledge Graph result. Retrieved May 23, 2016, from <http://searchengineland.com/leveraging-wikidata-gain-google-knowledge-graph-result-219706>
- Erxleben, F., Günther, M., Krötzsch, M., Mendez, J., & Vrandečić, D. (2014). Introducing Wikidata to the Linked Data Web. In P. Mika, T. Tudorache, A. Bernstein, C. Welty, C. Knoblock, D. Vrandečić, ... C. Goble (Eds.), *The Semantic Web – ISWC 2014* (Vol. 8796, pp. 50–65). Cham: Springer International Publishing. Retrieved from [http://link.springer.com/10.1007/978-3-319-11964-9\\_4](http://link.springer.com/10.1007/978-3-319-11964-9_4)
- Google, Inc. (2017). Get your free business listing on Google [Commercial]. Retrieved from <https://www.google.com/business/>
- Li, K., Li, Y., Zhou, Y., Lv, Z., & Cao, Y. (2013, July). Knowledge-based entity detection and disambiguation. Retrieved from <https://www.google.com/patents/US20130173604>
- Lih, A. (2009). *The Wikipedia revolution: how a bunch of nobodies created the world's greatest encyclopedia* (1st ed). New York: Hyperion.
- Mendes, P., & Jakob, M. (2012, August 24). Who's New in Google Summer of Code 2012: Part 3. Retrieved July 11, 2015, from [http://google-opensource.blogspot.co.uk/2012/08/whos-new-in-google-summer-of-code-2012\\_24.html](http://google-opensource.blogspot.co.uk/2012/08/whos-new-in-google-summer-of-code-2012_24.html)
- Meyer, R. (2015, April 15). Europeans use Google Way, Way More than Americans Do: Google's huge market share is part of what strengthens the EU's antitrust case. *The Atlantic*. Retrieved from <http://www.theatlantic.com/technology/archive/2015/04/europeans-use-google-way-way-more-than-americans-do/390612/>
- Schmachtenberg, M., Bizer, C., & Paulheim, H. (2014, August 30). State of the LOD Cloud 2014. University of Mannheim. Retrieved from <http://linkeddatacatalog.dws.informatik.uni-mannheim.de/state/>
- Shenoy, A., & Prabhu, A. (2016). Introducing the Google Tools Suite. In A. Shenoy & A. Prabhu, *Introducing SEO* (pp. 37–55). Berkeley, CA: Apress. [https://doi.org/10.1007/978-1-4842-1854-9\\_4](https://doi.org/10.1007/978-1-4842-1854-9_4)
- Singhal, A. (2012, May 15). Introducing the Knowledge Graph: things, not strings [Corporate]. Retrieved June 20, 2015, from

- <http://googleblog.blogspot.com/2012/05/introducing-knowledge-graph-things-not.html>
- Suchanek, F. M., Kasneci, G., & Weikum, G. (2007). Yago: a core of semantic knowledge. In *Proceedings of the 16th international conference on World Wide Web* (pp. 697–706). Banff, Alberta, Canada: ACM Press.  
<https://doi.org/10.1145/1242572.1242667>
- Sullivan, D. (2012, May 16). Google Launches Knowledge Graph to Provide Answers, Not Just Links. Retrieved from <http://searchengineland.com/google-launches-knowledge-graph-121585>
- Tanon, T. P., Vrandečić, D., Schaffert, S., Steiner, T., & Pintscher, L. (2016). From Freebase to Wikidata: the great migration. In *Proceedings of the 125th International Conference on World Wide Web* (pp. 1419–1428). Montreal: IW3C2. <https://doi.org/10.1145/2872427.2874809>
- Thalhammer, A., & Rettinger, A. (2014). Browsing DBpedia Entities with Summaries. In V. Presutti, E. Blomqvist, R. Troncy, H. Sack, I. Papadakis, & A. Tordai (Eds.), *The Semantic Web: ESWC 2014 Satellite Events* (Vol. 8798, pp. 511–515). Cham: Springer International Publishing. Retrieved from [http://link.springer.com/10.1007/978-3-319-11955-7\\_76](http://link.springer.com/10.1007/978-3-319-11955-7_76)
- Timeline of web search engines. (2016, September 14). In *Wikipedia: The Free Encyclopedia*. Wikimedia Foundation, Inc. Retrieved from [https://en.wikipedia.org/wiki/Timeline\\_of\\_web\\_search\\_engines](https://en.wikipedia.org/wiki/Timeline_of_web_search_engines)
- Vaish, R., Wyngarden, K., Chen, J., Cheung, B., & Bernstein, M. S. (2014). Twitch crowdsourcing: crowd contributions in short bursts of time. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 3645–3654). Toronto, ON, Canada: ACM Press.  
<https://doi.org/10.1145/2556288.2556996>
- Vrandečić, D., & Krötzsch, M. (2014). Wikidata: a free collaborative knowledgebase. *Communications of the ACM*, 57(10), 78–85.  
<https://doi.org/10.1145/2629489>
- Wall, A. (2017). History of search engines: from 1945 to Google today. Retrieved April 1, 2017, from <http://www.searchenginehistory.com>
- Wikidata. (2015, June 18). By adding to Wikidata, I have a free ticket into Google's Knowledge Graph, right? Retrieved from <https://www.wikidata.org/wiki/Help:FAQ/Freebase>
- Wikimedia Foundation, Inc. (2016, October 9). Wikidata:Statistics [Non-profit]. Retrieved July 2, 2017, from <https://www.wikidata.org/wiki/Wikidata:Statistics>