Every month, millions of reference questions—ranging from bar bets to homework queries, from television watchers curious about the next episode to doctors getting up to speed on obscure diseases—are answered by consulting Wikipedia (http://wikipedia.org), the free online encyclopedia. Just over half a billion people access Wikipedia’s 31 million articles in over 200 languages every month, making it by far the most-used reference work in the world. Given this extraordinary reach, what should the relationship between librarians and Wikipedia be? What role do we have in ensuring that our patrons and students use it appropriately and well, and how can libraries and librarians help make the site better? And what lessons can we draw from evaluating Wikipedia to apply to other user-generated reference resources?

This chapter will explore the relationship between librarians and the world’s largest reference work. It will give a brief overview of how Wikipedia works and how librarians can contribute. It will then examine other user-generated content sites and the benefits and challenges of such projects.

**HOW WIKIPEDIA WORKS**

Wikipedia is always changing. Each day, thousands of edits are made to thousands of articles, in a collective process of writing and editing by largely anonymous contributors. Some edits add whole new paragraphs, references, and ideas; some simply fix a typo, or add a category or a picture. Most edits make articles incrementally better. Some edits are not construc-
tive; however, these can be quickly “reverted” by other editors who simply change the article back to the previous version.

All of this is possible because Wikipedia is based on a wiki, a type of website invented in 1995 by Ward Cunningham, that allows for immediate changes to be made by anyone accessing it. While Wikipedia has become the best-known wiki since it was begun in 2001, the free software that it runs on, MediaWiki (http://mediawiki.org), is available for download and use by anyone and is widely used by other sites.

Wikipedia is run by a nonprofit foundation, the Wikimedia Foundation (http://wikimediafoundation.org), which provides the infrastructure to run Wikipedia and has the overall mission of collecting and disseminating freely, licensed educational content. The Wikimedia Foundation is based in San Francisco and is financed entirely by private donations, which mostly come from Wikipedia readers and average under $30. As of 2014, there are around 200 paid staff at the Wikimedia Foundation, who work with a large global community of volunteers to support the behind-the-scenes work of the project, such as keeping Wikipedia online, improving the site’s code, legal support, and event organizing. Additionally, there is a network of 40 Wikimedia chapters around the world that support activities in their local geographies, as well as many informal volunteer groups.

The Wikimedia Foundation also runs ten “sister projects” in addition to Wikipedia. Reference librarians should in particular be familiar with Wikimedia Commons (http://commons.wikimedia.org), which is an enormous repository of meticulously categorized, freely licensed photos and other media; Wiktionary (http://wiktionary.org), a multilingual dictionary project; Wikisource (http://wikisource.org), a repository for public domain source texts; and Wikivoyage (http://wikivoyage.org), a global travel guide. A complete list of projects is available at http://wikimediafoundation.org/wiki/Our_projects.

Each of these projects is written by volunteer editors. Editors do many jobs on the Wikimedia projects, including copyediting and adding references to articles, welcoming newcomers, organizing WikiProjects, and deleting articles when needed. None of this work is centrally organized. Editors work collaboratively but independently, operating within editorial guidelines and keeping an eye on each other’s edits, but without top-down oversight.
While no one group is responsible for editorial decisions, there is an extensive network of policies and guidelines that have been developed over the years by contributors. These core editorial policies include:

- **Neutral point of view (or NPOV):** This is the concept that Wikipedia articles should not take a particular side, or point of view, but rather should reflect what reputable published sources and scholarly consensus has to say. Articles should reflect all major points of view in disputed topics (but not overweight minority views).

- **No original research:** Wikipedia is a tertiary source; it is not a forum for publishing original discoveries or research. Original work should be published elsewhere first, where it can be subject to traditional peer review.

- **Verifiability:** Information in Wikipedia should be factual, but also verifiable via reliable outside published sources.

- **Free content:** Wikipedia’s entries, photographs, and media are all free—in cost, but also freely licensed so that others may use and reuse them (this is described later in more depth). This means, for instance, that text copied from other places (such as organization websites) will be deleted unless it can be proved to be compatibly licensed.

- **Wikipedia is an encyclopedia:** This guideline helps shape what Wikipedia is and how topics are covered. It also helps define what Wikipedia is not; for example, the site is not meant as a business directory.

There are also many behavioral policies that shape how contributors interact with each other. These include the ideas of maintaining civility toward fellow contributors, not using Wikipedia to promote yourself or your organization, and being bold in making needed changes and edits. One of the most important concepts is that edits speak for themselves; regardless of whether someone is an expert in a subject, that person is expected to contribute in accordance with Wikipedia’s editorial and behavioral norms. Simply claiming to be an authority on a topic is not enough; edits still need to be backed up by citations to sources that can be checked by others (though lack of access to subscription resources can sometimes hinder editors in this endeavor).

In general, editorial and other disputes are solved by contributors talking things through on article and project discussion pages. If that does not work, other editors may step in, and in extreme cases, contributors may be temporarily blocked from editing.
Wikipedia in Other Languages

Wikipedia editions exist in 287 languages, and around 220 of those have more than 1,000 articles. These Wikipedias range from the very large, with over a million articles (English, German, Spanish), to the quite small, with just a few thousand entries (Tibetan, Somali). These editions do not typically rely on translations; instead, editors working in these languages usually write articles from scratch. Multilingual contributors are always needed, as there is often a great deal to do (and a much smaller community to do it) in the smaller language editions.

If a corresponding article exists in another language in Wikipedia, a link will appear to it on the left-hand sidebar. For reference librarians, these links can also provide a quick way to find translations of terms or names. These “interlanguage links” are maintained, along with other data such as traditional authority file identifiers, in the centralized database Wikidata (http://wikidata.org), which can be updated as articles are added.

EVALUATING ARTICLES

Wikipedia differs from most traditional reference sources not just in the method of production and in its huge breadth of coverage, but also in the consistency of articles. Each article is a work in progress; articles are started at various times, worked on by different people, and are in different stages of being shaped by Wikipedia guidelines. Thus, some articles are comprehensive and accurate, with 20 or more page references; others are short, unreferenced stubs that are poorly written and are missing major areas.

Assessing the difference is a core task both for Wikipedia editors and for reference librarians guiding patrons. When evaluating Wikipedia articles, beyond looking at the article itself, you can also look at the article talk page and editing history to get a sense of how the page was produced. Specific elements to look at include:

- The text of the article: Is it well written? Are there gaps in the coverage of a subject (e.g., in a biography, is there no mention of a person’s early life)?
- The references: Are facts and any potentially questionable statements in the article footnoted to outside sources? Are these references to reliable, scholarly publications when possible? Is there a bibliography of further reading?
• Warning messages: Are there any warning boxes (such as “this article needs additional citations”) at the top of the page? These messages are meant to identify articles needing work for both readers and editors, and they provide a major part of how Wikipedia maintenance is done. Any editor can leave them on an article when they think there is a problem, and any editor can remove them once they fix the issue.

• The article talk page: The “talk” or discussion page is a separate wiki page that you can get to by clicking on the “talk” tab at the top of the article. Every article has an associated talk page that is meant for discussion of the article by editors, as well as providing a place for readers to leave comments. Looking at the talk page can give you a quick indication of whether there have been any major disputes associated with writing the article.

• The article history: You can get to the complete production history of any article by clicking the “view history” tab at the top of the article. This shows you, edit by edit, all of the changes that have been made to the article since it was created, who made them, and when. Clicking on the date of each edit shows you the article as it was at that time. Edit histories can be long and cumbersome, but quickly browsing through the history can give you a sense of whether lots of people have worked on the article or only a few; whether there have been any dramatic changes (such as sections being removed or vandalism); and whether the article has been updated recently. In general, the more people who have worked on the article the better it tends to be; however, lots of vandalism or disputes over the text may render the article incomplete or disjointed.

For more materials on how to evaluate articles, including handouts and presentations, see Wikimedia Outreach at http://outreach.wikimedia.org/wiki/Evaluating_articles.

CONTRIBUTING TO WIKIPEDIA

While volunteers contribute millions of hours of work to make Wikipedia what it is—a recent study by Geiger and Halfaker (2013) estimated over 100 million hours of work had gone into Wikipedia in total, and 41 million hours in the English Wikipedia—the number of active editors (around 80,000 total, and 30,000 on the English Wikipedia) is surprisingly small considering the size of the project, and more help is always needed. Wikipedia also
suffers from an imbalance in contributors. As of 2013, only about 1 in 5 active editors on the English Wikipedia is female, and editors are concentrated in Europe and North America.

Many librarians edit Wikipedia as individual contributors. Whether fact-checking, building bibliographies, polishing prose, or documenting little-known corners of the world, the work of Wikipedia fits with librarians’ professional skills. And there is always plenty to do—articles need to be rewritten to be more clear language; tags need to be replaced with sources; references need to be formatted or improved; and outdated information needs to be updated.

To get started as an editor, the steps are simple: create an account, find a topic that interests you, and start by copyediting or adding sources. The help pages (linked from the left-hand sidebar on Wikipedia) give several excellent tutorials for starting out, and the Wikipedia Teahouse (https://en.wikipedia.org/wiki/WP:TEAHOUSE) provides a space for newcomers. For those who are interested in writing about libraries and library science topics, WikiProject Libraries on the English Wikipedia (https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Libraries) is an effort to bring contributors together around the subject.

**Asking Questions on Wikipedia**

One little-known area of Wikipedia is the reference desk, where editors and readers can ask each other questions (and answer them) on any subject. “Leave a question here,” the page promises, “and we’ll get back to you.” Dozens of questions are posted every day, sometimes with quick factual answers and sometimes leading to extended, opinionated discussion. Access the reference desk (and help add answers) at https://en.wikipedia.org/wiki/WP:RD. There are also separate help desks for people with questions about Wikipedia itself.

**Contributing Institutionally**

As reference professionals, there are also several ways to participate institutionally. This includes acting as educators in how Wikipedia works. For such a commonly used reference source, few people understand the nuts and bolts of Wikipedia. Libraries can also support local Wikipedia volunteers by hosting and promoting events. Finally, libraries with public domain collections can add these to Wikimedia projects.
Wikipedia Loves Libraries

Wikipedia Loves Libraries (http://en.wikipedia.org/wiki/Wikipedia:Wikipedia_Loves_Libraries) is a grassroots effort, begun by Wikimedians in New York City in 2011, to hold Wikipedia events in libraries everywhere in October and November (coinciding with Open Access Week). Depending on the type of library and institution, events might include a backstage tour for Wikipedians (for instance, if a library has special collections to show off); an editor community meetup or edit-a-thon (which just requires the library provide the community with a meeting room and Internet access); or hosting a training workshop on how to edit. A list of some different library projects and ideas is available at http://outreach.wikimedia.org/wiki/Libraries. Hundreds of libraries have participated from around the United States and the world.

Sharing Collections via Wikimedia Commons

For libraries that have special collections and archives that are freely licensed or in the public domain, making those collections available online through Wikimedia Commons is an option. There are many major collections from archives, museums, and libraries now uploaded to Wikimedia Commons, including image collections from the National Archives and Records Administration (NARA), the British Museum and British Library, the Library of Congress, the New York Public Library, and many more. These rich collections of public domain images are a treasure trove for researchers and Wikipedians alike. Adding these images to Wikimedia Commons means that it is possible for editors to use them in Wikipedia articles, and it also means that the images will get deeply categorized by volunteers. Texts and manuscripts can go on Wikisource, where they will be categorized and transcribed. This type of project supports the mission of the contributing institution as well. As David Ferriero, the archivist of the United States, said in 2012 about NARA’s experiences contributing to Wikipedia: “Our work with Wikipedia is . . . great for us because it takes our goals of transparency, public participation, and collaboration to a new level” (Ferriero, 2012).

GLAM and Wikipedians-in-Residence

In 2010, a young Australian historian, Liam Wyatt, offered his services as a volunteer Wikipedian to the British Museum for a summer. He helped train
museum staff in how to edit Wikipedia, assisted Wikipedians with gaining access to the museum’s curatorial knowledge, and helped the museum share its treasures by adding photos of its collections to Wikimedia Commons. He also coordinated Wikipedians to improve articles about several of the museum’s important artifacts and collections, such as the Hoxne Hoard and the Rosetta Stone. Wyatt coined the term GLAM, standing for “Galleries, Libraries, Archives, and Museums,” a term that stuck in the Wikimedia world for projects relating to working with cultural institutions.

The project sparked interest by many other cultural institutions, and positions like this, called “Wikipedians-in-Residence,” have subsequently been adopted by dozens of cultural institutions around the world. Wikipedians-in-Residence focus on increasing public exposure to and appreciation of the institution’s collections, while also educating staff and volunteers, and making the Wikimedia projects more accurate and comprehensive in their coverage of cultural heritage materials. Find out more about Wikipedians-in-Residence projects at http://glamwiki.org.

Projects to contribute to Wikipedia like the ones detailed above help point the way toward a future where libraries can share their collections and institutional knowledge, and librarians can share their reference skills, with a much larger audience than has ever previously been possible. Improving Wikipedia can be seen as a way of meeting users where they are, and allows librarians to answer current and future reference questions from a global audience who might not otherwise have the opportunity to access a library of any kind.

**USER-GENERATED AND OPEN CONTENT**

Wikipedia is an example of “user-generated content”—work that is written by a large number of people who are also the work’s users or readers, rather than being written and edited solely by a selected team of writers and editors. Wikipedia is also open content, a term that describes work that is freely licensed and thus can be freely reused and remixed. Though Wikipedia is one of the most successful and widely used examples of both user-generated content and open content, there are many other reference websites and projects that are produced by their users. “User-generated content” is a broad term, and these sites come in many types, though all share the characteristic of being written by large groups of (generally volunteer) contributors.
The following will categorize some types of user-generated content sites that are particularly useful for reference librarians, and outline some broad ways to think about analyzing current and future user-generated works.

Reference Services and Sites
User-generated content sites that are meant explicitly as reference works aim to be reliable, and sometimes scholarly, resources. This category includes several specialty dictionary and encyclopedia projects, which may review user contributions before accepting them—a process that helps ensure consistency but adds an administrative burden for the site. One example is the Encyclopedia of Life (EOL) (http://eol.org/), which collects data about all life on Earth—animals, plants, and bacteria—from many other open collections on the web (including Wikipedia and Wikimedia Commons) and aims to consolidate information about each species. While readers can contribute, they must be signed up first and contributions are reviewed.

Another example of a reference project is OpenStreetMap (OSM) (http://www.openstreetmap.org/), which aims to build a map of the world with open, reusable data. It is based on map data from governments and includes user contributions, including point-based edits (such as additions of historical landmarks) and GPS traces. Because the OSM data is open, other open content projects (such as LocalWiki, http://localwiki.org, which is an effort to build wikis for community information) can use it in mapping applications.

Curated Collections
User-generated collections of materials often bring to mind photo sites like the nonprofit Wikimedia Commons or commercial Flickr (http://flickr.com), but user-built collections can also include bibliographic databases and archives, such as the shared database of references built by Mendeley users (http://mendeley.com), the collection of pre- and post-prints at the physics and math arXiv (http://arxiv.org), or the lists of books that are created by Goodreads (http://goodreads.com) community participants.

These collections may be explicitly curated (where submissions are chosen or reviewed before acceptance, such as the EOL’s datasets), lightly curated (like Wikimedia Commons, where volunteers review images for copyright status), or essentially uncurated (like Mendeley’s database).
Encouraging broad participation and having clear inclusion standards are key for these sites—the larger the collection and the more authoritative and detailed the metadata, the more useful the collection.

**Question-Answering, Education, and How-To Sites**

Anyone who has ever searched online for the solution to an odd computer problem is familiar with this category of sites and their potential usefulness in answering questions. Hundreds of thousands of reference questions a day are answered online by other users, often in forums meant for conversation on a topic (childcare, cooking, a particular type of camera), sometimes in sites largely meant for publishing other content, such as blogs that have built a community in the comments section, and sometimes in dedicated question and answer (Q&A) sites, like Quora (http://www.quora.com) or StackExchange, which runs 114 specialty Q&A sites on various topics (http://stackexchange.com/).

The quality and value of Q&A and review sites comes from the aggregate of individual answers, which give the information seeker a variety of perspectives to triangulate her own experience against. Amazon (http://amazon.com), which depends heavily on user-provided reviews to sell products, knows this and exploits their tremendous data resources to provide features like “most helpful favorable review” and “most helpful negative review.”

Some user-generated sites exist explicitly to teach skills and share educational information. For example, the site WikiHow (http://wikihow.com) uses an open wiki community to write how-to articles. WikiHow provides a reference resource for those looking for instructions on building, making, or doing things, ranging from “How to ask someone on a date” to “How to breed goldfish.”

**Discussion and Sharing Sites: Social Media**

What makes media “social?” The term is loosely used to encompass anything that is meant as a place for people to connect to one another through sharing (often personal) information and content. Many sites have social features, even if that is not their main purpose. (It can also be a fraught term; many long-term Wikipedia contributors resist any description of Wikipedia as social media, despite its active community and discussion aspects, arguing that they are there to write an encyclopedia, not for idle chatter). Face-
book, Twitter, and YouTube are the giants of this sphere, each providing platforms for publishing and sharing. Blogging sites and aggregators also provide publishing platforms, some with large communities. One example is Global Voices (http://globalvoicesonline.org/), which features citizen reporting and writing from around the world, and also serves as a community of people interested in international news and little-covered stories. Another example is the long-running MetaFilter (http://metafilter.com), whose community curates and discusses interesting stories from around the web.

For these sites, as with other types of user-generated content sites, having a vibrant community where members find enjoyment and value in participating is critical to attracting new members and long-term success.

**EVALUATING USER-GENERATED CONTENT**

All user-generated sites and projects can be analyzed along a few dimensions, including who contributes and why, site guidelines and mechanisms for assuring authority, and site purpose. Specific questions to ask include:

- Are submissions moderated or peer reviewed, or displayed directly as received?
- Who can sign up as a contributor, and are there standards for who can contribute? Are potential motivations for contributors clear (that is, is there a clear mission or purpose that contributors work toward)?
- How quickly is the site updated (or submissions accepted)? If the site includes quickly changing topics, are these current? If the site depends on user-provided review and curation (such as Wikipedia’s editors checking new edits), does this review actually happen?
- If it is a site that depends on a variety of individual perspectives (such as a Q&A site or review site), are there a diversity of replies? Are questions answered, and is there a respectful community around discussions?
- What are the motivations of the site publisher, and are they for-profit or nonprofit? Is contributed content simply a means for the site owners to make money by getting more clicks from search results? Who is the intended audience of the site?
- If the site uses material from many sources (like the EOL), are those sources (and their licenses) clearly indicated? If the site is meant for academic or reference purposes, does it also point to other relevant resources?
- Is the content freely licensed or under copyright, and is the content’s license (or copyright) clearly indicated? (See the next section for more on this.)
Why Does Openness Matter?

Who owns the content posted on a user-generated site? Is reusing something you find on a user-generated site a copyright violation—and if so, against whom? The answers to these questions reveal what an end user can do with the site’s content, but may also have implications for contributor motivations.

Wikipedia and the other Wikimedia projects are licensed with the Creative Commons Attribution-ShareAlike license (abbreviated as CC-BY-SA). There are several Creative Commons licenses offering various combinations of requiring author attribution, granting the ability to commercially reuse material, and requiring that remixed materials also be shared under a Creative Commons license. Read more about the Creative Commons licenses at http://creativecommons.org/licenses/. The fact that Wikipedia is licensed under Creative Commons means that only material that is freely licensed or in the public domain can be added to Wikipedia, and that while Wikipedia editors do retain their copyright, they also agree that their contributions will be licensed CC-BY-SA.

The license specifies that anyone can freely reuse, share, and adapt the material on Wikipedia as long as they agree to the terms of the license—that the source of material must be attributed, and if you alter or build upon it, you must license your resulting work under CC-BY-SA as well. This means that as long as you follow these terms, you can (for instance) translate a Wikipedia article into another language, adapt it for a class, or use a picture from Wikimedia Commons in a presentation without getting explicit permission first.

An open license means that contributors are explicitly making their work available for the public good. This can be an important, motivating factor for Wikipedia contributors and others who choose to freely license their work.

Understanding open licensing is a critical task for all librarians. Many other user-generated sites are also licensed under Creative Commons licenses, but certainly not all. If it is not specified, the kind of reuse that an open license makes possible is not allowed.

Open licensing also means that content that originated on one site may (quite legitimately) turn up on another. For instance, photos from Flickr that are Creative Commons licensed are routinely harvested for use in Wikimedia Commons, which aims to collect all useful open-licensed media. In another case of reuse, in recent years a few publishers have been collect-
ing Wikipedia articles into books and selling these online. If the publisher clearly cites the source and respects the licensing terms, this is a legitimate reuse, though many of these publishers do not do this. Sometimes reuse is not obvious. As of 2014, data from Wikipedia helps fuel Google’s “knowledge graph,” the box that displays quick information about a search subject on the right-hand side of search results. Though there is a link to the source, it is also not clear that Wikipedia can be edited if the information’s wrong.

LOOKING FORWARD: USER EXPECTATIONS AND THE FUTURE OF WIKIPEDIA, USER-GENERATED CONTENT, AND REFERENCE SOURCES

It seems clear that along with increased global Internet access, the need and desire to access Wikipedia is not ceasing anytime soon. The user-generated model that Wikipedia helped pioneer enables extraordinary resources that would be very difficult to produce with a traditional authorship model—for instance, a vast collection of photos of important monuments from around the world (from the Wiki Loves Monuments contest, http://www.wikilovesmonuments.org/) or important articles about health translated into dozens of underserved languages, as led by Wiki Project Med on Wikipedia (http://meta.wikimedia.org/wiki/Wiki_Project_Med).

But for the many millions of Wikipedia users, information is available and easy to search for, but it is not necessarily complete or polished. Does the fact that people use it anyway indicate a change in user expectations, or is it simply a reflection of the various needs of information seekers—sometimes an incomplete or brief answer is better than none at all? For everyone who uses Q&A sites and forums to find the answer to questions, answers do not need to necessarily come from a credentialed expert (and there might not even be an easy way to find an expert to answer some questions). And for many—perhaps the vast majority—of those readers who use Wikipedia and other user-generated sites, they may not know or consider who has produced the information they are using.

Wikipedia continues to grow in usage, but it also faces a worrying trend. For the past several years, there has been a drop-off in the numbers of active editors who are contributing to the project. This matters because without a sustained active contributor base, the project is unsustainable, as there is a large amount of ongoing work (such as triaging and reviewing
new contributions, updating entries to match new developments, and ac-
culturating new contributors) that has to be done simply to keep the project
operational and maintain quality.

All reference works, whether traditionally published or created by
a community of users, face similar concerns. Successful reference works
must provide high-quality information, develop a sustainable publishing
and business model, make sure the work is accessible over the long term,
and gather an audience of curious readers. But user-generated sites face ad-
ditional concerns, including keeping an active community of contributors
satisfied, attracting new contributors, converting readers into contributors, and making sure that the site is not taken over by spam or excess bureaucracy that might hinder contribution. The unique challenge of user-generated works is that contributions are not guaranteed; simply because it is possible to review an entry, add information, or answer a question does not mean that it will happen. Thinking about contributor motivations and project sustainability is an important part of critical analysis of these reference works, as well as something that each project must tackle in order to survive. More than ever before, these reference sources do not exist in a vacuum, but rely on their community of readers and contributors.

The benefits of a user-generated and open model that encourages participation, sharing, critical thinking, and reuse, continue to be demonstrated by Wikipedia and many other projects. But the future of reference publishing is not assured by either a traditional or user-generated model. Each faces challenges.

REFERENCES

Ferriero, D. (2012, July 14). Prepared remarks of archivist of the United States David S. Ferriero at the closing plenary at Wikimania 2012, Lisner Auditorium, George Washington University, Washington, DC. Retrieved from http://www.archives.gov/about/speeches/2012/7-14-2012.html

Geiger, R. S., & Halfaker, A. (2013). Using edit sessions to measure participation in Wikipedia. Proceedings of the 2013 ACM Conference on Computer Supported Cooperative Work (pp. 861–870). New York: ACM. http://dx.doi.org/10.1145/2441776.2441873
