Stepping Beyond Libraries: The Changing Orientation in Global GLAM-Wiki

Alexander D. Stinson\(^{(a)}\), Sandra Fauconnier\(^{(b)}\), Liam Wyatt\(^{(c)}\)

\(^{(a)}\) Wikimedia Foundation; \(^{(b)}\) Wikimedia Foundation; \(^{(c)}\) Europeana

Contact: Alexander Stinson, astinson@wikimedia.org; Sandra Fauconnier, sfauconnier@wikimedia.org; Liam Wyatt, liam.wyatt@europeana.eu

Received: 21 April 2018; Accepted: 21 June 2018; First Published: 15 September 2018

ABSTRACT
Wikipedia and its community has seen an increasingly close relationship between library communities, with both communities sharing overlapping values and practices related to public access to knowledge, a desire for openness, defence of freedom of speech, representing marginalized communities, and broad shared interest in reliable factual information and citations. This is best in evidence from the IFLA Wikipedia and Libraries Opportunity Papers and the substantial growth and ubiquity of the #1lib1ref campaign. However, the relationships between cultural heritage organizations (known as GLAMs – Galleries, Libraries, Archives and Museums) and the Wikimedia communities working on Wikipedia, Wikimedia Commons and other Wikimedia projects, began in its relations to Museums and Archives: partnerships like that with the British Museum in 2010 led to a collective effort to encourage GLAM organizations to contribute to and participate in Wikimedia Communities.

Though early partnerships in the GLAM-Wiki space focused on batch uploads of digital content to Wikimedia Commons, embedding Wikimedia-designated experts called Wikipedians in Residence, and editing activities, like editathons, which write expert advised content into Wikipedia, in the last 6 years, the landscape in which partnership with cultural heritage institutions has shifted radically. Two major trends have developed in the Wikimedia community: a shift towards facilitating linked open data with Wikidata and the expansion of GLAM-Wiki projects to support institutions that not only have large digital capacity and funding, but also institutions with limited resources, collections focused on marginalized knowledge, and collections in parts of the world with limited digital expertise. In this transition, Wikimedia communities have become change agents in bringing both linked open data and open digital practices to institutions around the world.

In this paper, we will explore how GLAM-Wiki tactics, opportunities and collaboration are changing the GLAM use of Wikimedia projects from being viewed as just a platform for exposing collection to a broader public audience, into a growing part of the heritage professional toolkit.

KEYWORDS
GLAM-Wiki; Wikimedia; Wikipedia; Wikidata.

CITATION
The early internet aspired to be an open and free environment for sharing and creating knowledge. Though today the internet appears largely controlled by corporations, alternative idealist efforts of scholars, volunteers, and institutions continues to maintain that open environment. Wikimedia’s volunteer community is an example of this: it creates and maintains Wikipedia and its sister projects,¹ and hopes to build a central “hub of hubs”, making the “sum of all knowledge” available to each person in their own language. In this aspirational effort, Wikimedia projects generate 15 billion page views a month, making it a top 6 website, and through its emphasis on verifiable information, acts as a top referrer to other reliable websites.

To realize its goal of sharing knowledge with the world, a 2017 strategy process led the Wikimedia movement to identify two key strategic priorities: first, to support the development of “Knowledge as a Service”, where Wikimedia knowledge platforms can be used in many technical and social contexts, allowing for dissemination of knowledge in many different formats; and second, “Knowledge Equity”, where a diversity of communities from throughout the world see their knowledge reflected in Wikimedia projects (“Wikimedia Movement 2017 Strategic Direction” 2017).

Building platforms, technologies and services to advance knowledge for all people in their own contexts, requires partnership with many different types of organizations. Wikimedia communities frequently partner with institutions who historically control knowledge: from educational institutions, to health research institutes, to governments and even for-profit publishers. Since at least 2009, some of the best allies in this work have been GLAMs (Galleries, Libraries, Archives, and Museums) who share the same principal objective as the Wikimedia community: to ensure that accurate and well supported knowledge is accessible.

The collaboration between the GLAM-Wiki and the professional GLAM communities is not only mutually beneficial, but potentially transformative, changing the way that each works and represents knowledge. Wikimedia communities are able to work towards better sourced, and more equitable, free knowledge about the world, learning from GLAM professionals’ expertise and rigor in mediating, structuring and safeguarding knowledge. But GLAM-Wiki partnerships also bring professional change in the GLAM field. When working with Wikimedia projects and communities, staff gain experience in the principles of openness and open collaboration, and discover the (often vastly) increased impact for collections and domain-specific knowledge when it’s shared beyond the institutions’ own walls.

Where does GLAM-Wiki come from?

GLAM-Wiki has a long history of practice within the Wikimedia community. Early Wikimedia communities pursued a number of limited collaborations with local GLAMs. However, the beginnings of Wikimedia’s international GLAM-Wiki community emerged during the 2008 annual conference of the Wikimedia community, Wikimania, which was hosted at the Library of Alexandria in Egypt that year. As part of the conference, Australian Wikimedian Liam Wyatt, alongside a small

¹ For the full list of sister projects see: https://www.wikimedia.org/.
group of other influential Wikimedians, took backstage tours of the Library of Alexandria, and the ensuing discussion cemented an idea: that libraries, and other heritage organizations, really were doing much the same work as Wikimedians, but with an in-person approach to “sharing heritage with the world”.

Inspired by these conversations, Wyatt and Wikimedia Australia developed a local conference with the community of heritage institutions in New Zealand and Australia, using an acronym used to highlight cross-disciplinary collaboration in the region: GLAM. This 2009 “GLAM-Wiki Conference” facilitated a two way dialogue towards finding common ground between how cultural institutions do their work of preserving and curating access to heritage materials, and how the Wikimedia community does share knowledge (“GLAM-WIKI 2009 Report” 2009). Often summarised as “we are doing the same thing, for the same reason, for the same people, in the same medium. Let’s do it together”. The conference delivered a recommendations document that solidified the opportunity for skill and resource exchange among the communities, making the case for open and collaborative processes at institutions that bring collections to a broader public (“GLAM-WIKI Recommendations” 2009). Though previous upload activities and other partnership strategies had happened in other parts of the Wikimedia movement, such as uploads to Commons from the Bundesarchiv in Germany in late 2008 and from the Tropenmuseum in the Netherlands in 2009, this conference cemented “GLAM-Wiki” as the term describing the relationship between GLAMs and the Wikimedia movement. It also reset the terms of the relationship between the cultural sector and wikimedians, from one of defensiveness to collaboration – the event was set amid the tensions of a very public copyright dispute between the UK’s National Portrait Gallery and Wikimedians that had erupted only one month prior.

Wyatt then furthered the concept and brand of GLAM-Wiki during a 5 week volunteer “Wikipedian in Residence” position at the British Museum in summer 2009. Create a rotating position of “Wikimedian in Residence” was one of the items within the conference recommendations and, having asked to volunteer at various institutions, the British Museum permitted the experiment. There, Wyatt demonstrated how increasing the quality of the Wikipedia content related to the museum’s collections increased the pageviews of those items on Wikipedia, leading to a direct increase in the quantity of visitors the museum’s own website, leading to an increase in broad public awareness about the collection (“Wikipedia:GLAM/British Museums” 2010). For example, Wyatt and the museum discovered that information about the Rosetta Stone got 5 times more visits on Wikimedia sites than on the museum’s website (Lih 2015). The novelty of a world-renowned institution sharing content on a crowdsourcing platform caught on in the English language press, including venues like the New York Times (Cohen 2010), and led to a cascade of international interest amongst other major institutions and Wikimedia communities around the world.

---

2 Also known as “the heritage sector” (i.e. UNESCO), or LAM (as in LODLAM, “Linked Open Data, Libraries, Archives and Museums”), MLA, or CHIs (“Cultural Heritage Institutions”), “memory institutions” and sometimes “collecting institutions”. There is no agreed term for the sector as a whole, and debate continues about the edge cases such as botanical gardens, zoos, and broadcasters. “Galleries” in the Commonwealth context does not necessarily connote a commercial in the North American and some European contexts.
The British Museum Residency offered a proof of concept for many other strategies and tactics: harnessing the multilingual Wikipedia community; giving focused attention one topic with all the expertise of the institution; to provide a ‘matchmaking’ service between experts and interested volunteers, and various social and training events aimed at breaking down barriers (“Wikipedia: GLAM/British Museum” 2010). The specific goal of that project was to demonstrate the viability of the model of bringing a Wikipedian ‘in house’ since it was considered a fairly heretical act by both communities in 2010: museum professionals often considered Wikipedians to be anti-expertise, while Wikipedians were concerned about potentials of conflict of interest and undermining the amateur spirit of the community. Neither side, at the time, could imagine a paid, long lasting project. Much has changed since. The stated measure of success for such a short project was to make the concept of a resident Wikipedian redundant – for collaboration to continue naturally without the need for a dedicated intermediary. Or, if not that, then as implied by the 2009 conference recommendation that an ongoing but ‘rotating’ position be created to give many people the chance to gain skills and experience in the sector.

In the year following Wyatt’s Residency, a growing group of “Wikipedians in Residence” emerged in Catalonia, France, the United States, the United Kingdom, and Israel (“Wikipedian in Residence” 2018). Recognizing this trend, the Wikimedia Foundation created a GLAM-Wiki position held by Wyatt as part of a one-year fellowship program. In that capacity, Wyatt convened a “GLAM-Wiki Camp” in 2010 to document the best practices and needs of that emerging community. With Wyatt’s support, core pieces of infrastructure for the GLAM-Wiki community were developed, including a newsletter, documentation, and mailing lists (Bartholomew 2011).

In addition to the communication tools Wyatt concluded his Residency by recommending, among other things, a one-year fellowship focused on capacity development for the United States, to support burgeoning regional interest in GLAM-Wiki in that country where, unlike other nations with mature editing communities, no national association existed to take a coordinating role (Wyatt 2011). The Wikimedia Foundation hired Lori Byrd-Phillips, who had been the Wikipedian in Residence at the Indianapolis Children’s Museum, as the U.S. Cultural Partnerships Coordinator for 2012 (Bartov 2011). The 2010-2013 period, bookended by Wyatt’s Residency and Byrd-Phillips’s foundation position, left GLAM-Wiki with a relatively tight-knit core of practitioners sharing practices in the cultural-partners mailing list and the “This Month in GLAM” newsletter, alongside local-language working groups in Dutch, French and several other European languages. Following British Museum success and other residencies and partnerships, three tactics became the principle tactics associated with GLAM-Wiki:

- a **Wikipedian in Residence** to build capacity at the institution by providing staff training, and smoothing any interactions between the two communities of practice;
- **batch uploads to Wikimedia Commons** for institutional buy-in through pageview metrics and a clear benefit to open content;
- **editing events** to engage the local volunteer community, institutional staff and other allies in that initiative (at various points and with different activities these editing events have taken the form of “meetups”, “backstage pass tours”, “editathons”, and “editing workshops”).
In many cases, batch or “mass” uploads to Commons showed the most promise in the eyes of administrators advocating for collaboration: institutions could take advantage of pageview numbers on Wikipedia at several magnitudes larger than their institutional website. Take for example the United States National Archives, three out of four viewers of the Archives’ digital documents are on Wikimedia sites (McDevitt-Parks 2014). Or the National Library of Wales in the United Kingdom, which reports 320 million pageviews of 15,000 images (Evans 2017). The impressive statistics allow for institutions to point towards a broader public impact from sharing their collections. However, these projects favour institutions with large and unique digital collections, such as ethnographic photography or institutionally photographed 3D objects. Indeed, one of the earliest documented GLAM partnerships, in 2008, was a batch upload of 80,000 digitised images from the German Federal Archive for which limited metadata existed – requesting the Wikimedia community provide categorisation and fact-checking on the captions (“Commons:Bundesarchiv” 2016).

Moreover, the process for uploading to Commons was painfully complex – requiring institutional expertise, Wikimedia expertise, and a developer able to tie both websites together. Partially instigated by Wyatt during his Residency in 2011, a coalition of European Wikimedia Chapters and Europeana to seek grant funding from the Wikimedia Foundation to develop the “GLAM-Wiki toolset”. The toolset was a batch uploader designed to take XML files of institutional metadata alongside hosted media files, and upload that content into the appropriate structure on Wikimedia Commons. Though the toolset ended up being a less than “foolproof” solution, not the least because it required that the batch of metadata be “mapped” to the unique structures of Wikimedia Commons and the website hosting the files needed to be manually “whitelisted” by Wikimedia administrators, it smoothed the opportunity for batch uploads to Commons with a consistent and publicly documented, if complex, workflow.

In response to growing complexity and demand for GLAM partnerships, Wikimedia communities have found innovative and context-appropriate ways to coordinate their activities. Though both Wyatt and Byrd-Philips’ short term positions developed a community of practice, the Wikimedia Foundation internally de-prioritized GLAM in 2013, leaving affiliates and other volunteer community groups to develop GLAM projects without central support (Gardner 2012). Through this decentralization, in those countries where a professionalised affiliates existed, a community of GLAM coordinators have been hired by Wikimedia affiliates to manage these relationships. In locations where there was no organised capacity, the Wikimedian generally had to fend for themself. For example, the third Residency took place at the palace of Versailles in a formal partnership between the museum and Wikimedia France, while the fourth Residency was hosted at the Museu Picasso, Barcelona where no recognised Wikimedia affiliate existed at the time. Crucially, and as a direct result of the Wikimedia Foundation’s absenting itself from the conversation during this time, “GLAM

---

3 Generated for the most part with this tool: https://tools.wmflabs.org/glamtools/baglama2/

4 The project encountered a number of technical hurdles because the batch uploader was built within Common’s MediaWiki software, which required WMF code review and development by a developer familiar with MediaWiki. Since then, the greater flexibility and availability of OAuth for user authentication and the MediaWiki and Commons APIs has allowed for simpler app-like tools to be developed outside of Commons’ core code. For more recent research on how developers can make improvements to this process, see: https://meta.wikimedia.org/wiki/Research:Supporting_Commons_contribution_by/GLAM_institutions.
"Work" frequently became the vector by which loose local communities coalesced into more coordinated groups as the relationships with local museums forced longer-term thinking and relationship building on a scale that required larger and more professional operations. Many European Wikimedia affiliate organisations were founded in this period, and to this day GLAM activities form a core component of their annual plans. The increased demand for formal partnership programs from these external organisations obliged the community to try to create systems to coordinate the relationship on a longer term, and therefore more professionalised, basis.

During 2014-2016, the three tactics of editathons, batch uploads and residencies traveled widely through the movement, in part due to their visibility in the documentation of the GLAM-Wiki community and the WMF Grants program (“GLAM Newsletter” 2018). Alongside these reinforced tactics were a number of innovative new structures for partnership with GLAM institutions, such as digitization projects, work with institutional metadata on Wikidata, ethnographic photography expeditions, and transcription on Wikisource (“GLAM Model Projects” 2018). This spreading led the community of practice to grow increasingly decentralized, with local volunteer communities developing highly specialized and diverse tactics for working with local GLAMs. Notable examples include Catalonia’s work with its public library network (“GLAM Case Studies: Catalonia’s Network of Public Libraries” 2015); Wikimedia Deutschland’s “GLAM on Tour” (Iliev 2016) and “Coding Da Vinci” (Fischer 2016); Wikimedia Argentina’s Wikipedian in Residence who supports digitization throughout a number of local institutions (“GLAM Case Studies: Wikimedia Argentina Digitization Project” 2016); or Wikimedia GLAM Macedonia’s student clubs at local museums (“GLAM Case Studies: Wiki Club in Macedonia: From Idea to Award” 2017). In turn, different subcommunities have formed around these tactics at movement events, like Wikimania 2016 and 2017, and in new communication channels (most notably, Facebook groups such as Wikidata + GLAM and Wikipedia + Libraries).

**Wikipedian in Residence positions**

The spread of projects is uneven, as some areas of the world have a far greater density of both Wikimedia volunteers and digitally-minded cultural heritage sector. This uneven spread, has led to very limited implementation of projects in emerging regions like Sub-Saharan Africa and South East Asia. However, in recent years these gaps have begun to narrow. Wikipedian in Residence examples: for example, the Philippines community announced in April 2018 that it was creating the first Wikipedian in Residence program in East/South-East Asia – hosted at a memorial museum ‘Bantayog ng mga Bayani’. Separately, but operating at the same time, the first Residency in Nigeria was taking place – supporting a culture TV program ‘Gode Africa’ (“Wikipedian in Residence” 2018).

Clearly, there is no “one correct way” for formal relationship between a cultural institution and a Wikimedian to occur, and there is tension between trying to create consistent best-practices but also allow for innovation and adapting to local context. The particularly intense projects of Wikipedian in Residence positions, tend to fall into different clusters, based on three factors by which Residency projects can be grouped: the time allocated (from short-term part time, through permanent full time); the method of ‘reward’ (from volunteer to class-credit through to stipend or salary); and the method
of recruitment (from cases where the GLAM will publicly advertise to cases where an individual Wikipedian will make the initial approach) (“Wikipedian in Residence” 2018). Combinations of all of these have occurred, and still others which don’t necessarily fit the format either – including ‘online residencies’ (where no physical presence is required) and projects with a network or consortium of cultural institutes across a city. The main improve connections between the two communities (and their content, audience and expertise) but not the editing of Wikipedia articles as a core goal.

On several occasions there have been calls to formalise and in some way provide ‘international accreditation’ for a Wikipedian in Residence program. However, due to the diverse ways in which successful programs have been operated, the risks of adding a control-layer on top of a vibrant community may reduce the innovative community. Nonetheless, as the number and variety of projects using the moniker “Wikipedian in Residence” have increase, there are cases where the general principle that the Resident is not a “paid editor”, causing reputational damage to GLAM Wiki as a whole. For example, in 2012-13 the Wikimedia Foundation itself advertised and supported a paid Residency which had limited interaction with the community and strong ties to an ongoing donor relationship – a project it later publicly apologised for (“Assessment of Belfer Center Wikipedian in Residence Program” 2014).

At the first Residency, at the British Museum, both the museum and the English Wikipedia community were agreed that it should only be a purely volunteer position. From the Wikipedians’ perspective this was to avoid conflict of interest (as determined by a debate held at the conflict of interest noticeboard: COIN) and from the museums’ perspective, in a period of budget cuts, this helped avoid the internal perception that it had funds available to allocate to new projects. After undertaking a lengthy risk-assessment of the reputational damage that could occur if seen to be embracing Wikipedia in 2010, the museum invited Wyatt to volunteer. Much has changed since then. The acceptance of the need for in-house expertise in Wikipedia and its ecosystem has tracked a similar trend to previous technology shifts before – of social media, and of the internet in general. In 2013 the United States National Archives became the first to make the position of Wikipedian in Residence a full time and permanent role (AOTUS 2013).

The effect of supply and demand means that as more cultural organisations wish to build a relationship with Wikimedia (or, at the very least, improve the quality of the representation about “their” topic on Wikipedia), the proportion of paid Residency positions has increased. Sometimes funded through grants within the Wikimedia ecosystem, sometimes funded by the institution directly (either on a stipend or salary role), and sometimes as a 50:50 split. This has both positive and negative effects: valuing the expertise of the community members willing to take on this role, but also increasing the potential (real, and imagined) that it encourages editing by people with a financial conflict of interest. Nevertheless, creating roles within and without the institutions themselves has been a crucial part of growing the potency of GLAM projects in the movement – using Wikipedians

---

The definition of what constitutes “undisclosed paid editing” are complex, especially as pertains to financially compensated partnership with like minded cultural organisations (or employed cultural sector professionals). The term “undisclosed paid advocacy” might be more appropriate – as it focuses on the attempt to promote a point-of-view rather than the focusing on the issue of payment itself: https://meta.wikimedia.org/wiki/Terms_of_use/FAQ_on_paid_contributions_without_disclosure.
Recent trends

By the 2015 GLAM Conference, three trends were emerging that indicated the changing nature of the community and its core activities: the importance of libraries in creating sustainable partnerships; the exciting opportunities that Wikidata might offer (best exhibited by the volunteer-led project universal catalogue raisonné project “The Sum of All Paintings”); and the growing number of “emerging” Wikimedia communities (“Defining Emerging Communities” 2017) trying to replicate the “big three” tactics, but encountering complications or challenges not previously seen in the European and North American context.

This fracturing developed, because of a lack of central facilitation and strategy after the Wikimedia Foundation withdrew direct support. But more importantly, the growing diversity reflected a reality of the work: GLAM-Wiki always included a mixed set of tactics (uploads to Commons, editing events and campaigns, Wikipedians in Residence, etc.) in a diversity of cultural regions in partnership with many kinds of institutional and professions (national libraries, local public libraries, massive natural history museums, small local history museums, institutional archives and independent archives, zoos, scientific research centers, etc). This was in direct contrast to the highly replicable models in Wikimedia Education Program: Higher Education partnerships, where assignment modules could be repeated in different universities and software “dashboards” for managing class assignments were built to track and score students’ progress.

Even after increased GLAM support at the Wikimedia Foundation since 2016, it has been challenging to find the shared knowledge and skill-sets holding this community of practice together. To this end, most WMF energy during 2016-17 focused on understanding and communicating change in the GLAM-Wiki community (“GLAM Model Projects” 2018) and addressing capacity-debt in the GLAM Wiki Community around the main tactical areas. This work includes advising on the development of Wikipedian in Residence positions and other capacity in emerging or unsupported communities, developing a best practices guide for editathons (“Editathons Training” 2017), and advocating for the Alfred P. Sloan-funded project Structured Data on Commons (“Structured Data on Wikimedia Commons” 2017).

Tension between existing tactics, emerging tactics, and diverse communities of practice drive current innovations in GLAM-Wiki. Beyond working with libraries, discussed elsewhere in this JLIS Issue, the strategic opportunity for change in GLAM-Wiki partnerships focuses on two major areas: the integration of structured data, and creating equity by expanding support for marginalized language and knowledge communities. These two strategic opportunities closely parallel the two major threads of the Wikimedia 2030 Strategic Direction adopted during the 2017 movement strategy process.
Wikidata, bridging cultural memory and approaching the dream of the semantic web

One of the main focuses of the Wikimedia Movement Strategic Direction was that the Wikimedia community needs to think of “Knowledge as a service”:

To serve our users, we will become a platform that serves open knowledge to the world across interfaces and communities. We will build tools for allies and partners to organize and exchange free knowledge beyond Wikimedia. Our infrastructure will enable us and others to collect and use different forms of free, trusted knowledge (“Wikimedia Movement 2017 Strategic Direction” 2017).

Within the Wikimedia Movement, the GLAM-Wiki community contains significant experience creating partnerships to create infrastructure and public access to institutional knowledge. To continue working with GLAMs requires finding ways to share the collections and expertise beyond the walls of the institution, into Wikimedia’s ecosystem of knowledge platforms through data. Data embraces the expertise and investment of institutions in creating that data, while connecting collections with the nearly two decades of contextual knowledge created by the Wikimedia community.

The principal platform for sharing data in the Wikimedia community is, since its launch in late 2012, Wikidata. A language-independent, linked, open, structured database, Wikidata started as a platform for identifying the relationship between articles in different language Wikipedias that are about the same topic. The project uses semantic three part statements (object, property, subject – known as “triples”) to describe the relationship between these concepts. In turn, these semantic statements create a web of relationships between knowledge both within and outside Wikipedia language editions and other Wikimedia projects. Unlike many other open-data platforms, Wikidata supports labels and description in any language; this in turn allows different Wikimedia communities to work together to describe the world. Crucially to its data-model is the ability for the system to accommodate contesting, or even contradictory, factual statements and references for each. Just as professional lexicographers are at pains to emphasise their publications are prescriptive, not descriptive, Wikidata is built to embrace the inherent messiness behind any attempt to place order upon data.

As of this writing, Wikidata includes more than 45 million concepts, described on average with nearly nine statements. Over 50% of the more than 4500 properties link to external authorities (such as national library authority controls and professional vocabularies) or other unique identifiers (such as Twitter and Facebook handles, Quora topics, etc). Wikidata describes Wikimedia’s knowledge in relation to other Wikimedia projects. Moreover, it can act as a hub connecting many unique concepts across many different projects – it has the potential to offer a “universal crosswalk” between different language and professional vocabularies. Professional vocabularies rely on heavy technical and editorial practices to create authoritative information. On the other hand, Wikidata has low editorial overhead and a universal scope, offering the opportunity to allow metadata created by experts to reach well beyond its initial niche.

Wikidata’s breadth creates a sweeping landscape of opportunities for a large and flexible linked data hub, for describing collections or the knowledge contained within them. Wikidata can offer memory and heritage institutions a range of opportunities, such as:
- improving and enriching tagging within the institutional catalogues in order to enrich concepts already used;
- contributing to Wikidata in order to build greater representation of objects within the collection;
- developing tools, software and visualizations that create unexpected discovery and analysis opportunities beyond the original intention of catalogers.

Institutions can take advantage of Wikidata to augment the contextual information within collections. Matching professional vocabularies to Wikidata concepts allows for the institution to depend on a reliable source for authority control while connecting that authority to other datasets. As Europeana describes in a 2017 blog, the matching of different professional vocabularies through a central hub allows aggregators and other technical reusers of these collections to consistently link entities and objects across collections (“Why Data Partners Should Link Their Vocabulary to Wikidata: A New Case Study” 2017; “Get Your Vocabularies in Wikidata… so Europeana and Others Can Get Them” 2017).

In turn, once a collection’s thesaurus has been matched to Wikidata, developers can employ a range of both practical or highly complex computing activities to enrich collection experiences. In the Laurentian University library catalogue, systems librarian Dan Scott created snapshot cards of information in his catalogue drawn from Wikidata. This produces an experience similar to the information you would find in a Wikipedia article infobox or Google’s Knowledge Graph. Now when users search for their favourite album in the library catalogue, they can find valuable connective information, such as discographical information from Musicbrainz or the Twitter username for the artist (Scott 2017). The National Library of the Netherlands (KB) applied matched authorities data to a discovery challenge: creating arbitrary questions across unknown sets of entities. By matching the Dutch authority file against Wikidata items, KB could run queries to search its newspaper database, to successful get results from searches such as, “Give me all newspaper articles, that are tagged with entities that were politicians in the Dutch parliament, but born outside of the Netherlands” (Veen 2017; “KB Research Portal” 2018). Wikidata allows existing metadata to be leveraged for more than its original intention, answering arbitrary questions well beyond the scope of the institution’s dataset.

The National Library of Wales connected data from a collection of Welsh landscape prints to Wikidata concepts. Previously the collection only had English language descriptions with controlled strings. By matching the metadata to Wikidata concepts, the library can now represent its collection in Welsh or any other language using translations from the Wikimedia community. Additionally, the collection can be browsed from other perspectives, such as mapping the coordinates of the locations depicted in the collection, or organizing the collection by topic (Evans 2017).

---

Wikidata takes advantage of the relationships between different types of knowledge, and bypasses the need for institutions to individually partake in herculean efforts required to research all of the context of every facet of their collection.

Matching existing institutional data vocabularies against Wikidata delivers utility from the institutional perspective, but we are also beginning to see value in ingesting data about the objects themselves into Wikidata proper. The most extensive such effort is WikiProject Sum of All Paintings (SoaP). SoaP is a volunteer-led initiative seeking to create a universal catalogue raisonné of paintings. Using openly-licensed data sets provided by museums around the world, the projects has collected nearly 250,000 paintings at time of writing (of which nearly half are paired with openly-licensed digital images of the paintings) (“Wikidata: WikiProject Sum of All Paintings” 2014).

However, ingesting entire institutional collections leads to ongoing questions: how wide and inclusive can we make the scope of Wikidata? When does a dataset or concept become significant, flexible, or utilitarian enough to Wikidata? When should these databases be built into a separate instance of Wikibase, the software behind Wikidata, and drawing on the properties and items on Wikidata? From pottery shards to records of individual slaves, the academic community increasingly finds opportunities for applying linked data to large domains; however, maintaining these data sets in Wikidata may be too cumbersome. Equally, an institution may wish to be connected with Wikidata, but retain sole authorship and scoping rights to the content of their project. Building a federation of Wikibases may offer a powerful solution.

Wikidata could become a “hub-of-hubs” for heritage data allowing us empowering communities to document the world’s heritage. Wikimedia volunteers have been building activities on top of heritage data sets for many years, starting with the global Wiki Loves Monuments campaign. Now the largest photography contest in the world, Wiki Loves Monuments became the first serious attempt to aggregate all heritage listings. Because of work done by Wikimedia Sweden’s Connected Open Heritage project, much of the country’s monuments can be found on Wikidata, fostering the photography and documentation of these monuments. The project has also been expanded to serve additional purposes: creating global maps of that heritage, or analyzing classes of heritage, their relationship to major architectural/heritage movements, etc.

Wikidata projects can help systematically address gaps in Wikimedia’s coverage as well, helping advance the movement’s strategic goal of knowledge equity. In 2016, Wyatt, now Wikimedia community liaison manager at Europeana, organized a Wikidata translation and Wikipedia article writing campaign focused on 280 artworks representative of the cultural heritage of each country in the European Union, 10 per member country. By encouraging volunteers to translate artwork metadata in Wikidata and write articles in each of the official languages of the European Union, the Europeana community could develop a more equitable representation of cultures in many languages (“Wikidata: Europeana Art History Challenge” 2015). In another case, the volunteer-led WikiProject Women in Red uses Wikidata lists to encourage writing biographies of women to help address the content gender gap on English Wikipedia (“Gender Gap” 2011). Many of these lists are populated by Wikidata items described with, and often sourced from, professional vocabularies and in professions of interest to heritage communities (“Wikipedia: WikiProject Women in Red” 2015).
While Wikidata can empower institutional collections to have value and connection beyond their walls, the Wikimedia and professional communities will need to address a number of challenges:

- How do we ensure a consistent set of tools for monitoring and correcting data in Wikidata, when it’s used for increasingly more and more important functions within institutional collections?
- How do we empower communities of GLAM and other heritage-interested professionals to participate sustainably in Wikidata and other Wikimedia projects? Many of the batch upload or vocabulary matching projects so far have relied heavily on existing Wikimedia volunteers for monitoring and support – though good for the institutions who can find that support, this does not create a sustainable practice.
- Documentation and professional knowledge of Wikidata for professional communities is currently very scattered and limited. How can Wikimedia ensure that these changes find their way into professional practices?
- How do we integrate the use of Wikidata alongside the authorities and controlled vocabularies matched in Wikidata into professional practice and software? Though library communities regularly use machine-readable data, museums and archives don’t have widespread, consistent adoption across the profession globally.

As a five-year-old project just beginning to be adopted by the cultural heritage community, Wikidata offers a wide range of opportunities for professional communities to run experiments, document use cases and develop new applications. Metadata attached to digital media files offers the other big strategic opportunity. Historically, batch uploads to Wikimedia Commons started partnerships. This tactic followed a fairly simple logic: institutions spend large amounts of money digitizing collections; the mission of institutions is to get their knowledge before a public; Wikimedia projects, especially Wikipedia, have huge international audiences; sharing files on Commons and embedding them in Wikipedia can expose your collections to that audience.

However, high-impact institutional collection uploads ran into major technical complications: Wikimedia Commons uses the same platform as Wikipedia – a platform designed to support text. Thus uploads required institutional collections to go from a fairly structured, and often very consistent metadata format at the institution, into a highly unstructured and radically inconsistent free-text wiki environment. Wikimedia Commons partnerships lost quality of data, diminishing the value of institutional investment in that data, and getting collections into an unstructured environment required complex technical and social understanding of the Commons community and its practices.

In recognition of this, the Wikimedia Foundation is changing that Commons infrastructure with the Alfred P. Sloan Foundation-funded “Structured Data on Commons” project. Structured Data on Commons allows us to rethink the often burdensome and hard to navigate workflow of uploading large batches of content by integrating Wikibase’s machine-readable fields into Wikimedia Commons (Morgan 2018). The project seeks to provide better support for the needs of GLAM uploaders. With the same infrastructure as Wikidata, and the same connection to authoritative vocabularies, Commons can become a powerful platform for not only ingesting content from institutions, but sharing it with the world through consistent APIs, translation of content, and more effective crowdsourcing.
Ensuring that the diverse corpus enters a global memory: working with marginalized communities

By integrating existing institutional data from established digital collections, we work to build a more broadly accessible global heritage. Historically, however, the GLAMs who work with Wikimedia tend to also be the organizations with money and highly educated staff in Europe and North America – in turn perpetuating the dominance of colonialism-tainted collections and professional practices in retelling history. To help the Wikimedia movement achieve its aspirations for Knowledge Equity for all people and in all languages, the GLAM-Wiki community will need to make a concerted effort to look beyond the collections to which we have ready access – to imagine how to collect the knowledge not already represented, or systematically marginalized, in digital spaces.

Fortunately, the Wikimedia Movement Direction provides a framework for thinking about how to do equity work from a strategic level:

As a social movement, we will focus our efforts on the knowledge and communities that have been left out by structures of power and privilege. We will welcome people from every background to build strong and diverse communities. We will break down the social, political, and technical barriers preventing people from accessing and contributing to free knowledge (“Wikimedia Movement 2017 Strategic Direction” 2017).

With a focus on breaking down the barriers to participation in Wikimedia projects, GLAM-Wiki frequently becomes the vehicle for testing diversity-focused initiatives. Wikimedia communities throughout the world, from Argentina to Indonesia, India to Ghana, Ukraine to the United States, are beginning to model best practices for working with collections that have been underrepresented on the international scene up till now.

However, communities working with marginalized knowledge face challenges very different than those many of the experienced GLAM-Wiki affiliates in Western Europe encounter, whether because of the size of the language or cultural community that controls that knowledge, access to that knowledge bringing a history of colonization, or because of lack of technical skill or knowledge by the GLAM partners or local Wikimedia communities.

These challenges sometimes reflect the practical realities of working with GLAMs in less wealthy or less technologically supported contexts. Take for example how tactics like batch uploads to Commons and Wikipedian in Residence positions frequently change their character and practice in communities digitizing hard-to-access knowledge.

Wikimedia Indonesia adopted digitization as a strategy for working with GLAMs. Because much of Indonesia’s documentary history and culture is not readily available on the internet, the Chapter decided to focus some of its efforts on digitizations. However, this introduced a number of challenges – for example, transporting the digitization equipment between islands and managing a complex digitization workflow – which is hard or almost impossible to do at scale with volunteers.
Wikimedia Ghana is working with the National Archive to digitize collections relevant to the archive. Getting to the process of digitization required first negotiating access to the collection; then, after a transition in government leadership, the Wikimedians had to renegotiate the project with the new leadership. Because Wikimedia projects are not designed to be stable repositories for cultural heritage (“anyone can edit” also means that content may be changed or deleted), the local organizers had to help the archive work with a local LIS professional to get a stable “item of record” repository set up for the digitization process. What in another region or context might have been a multi-month negotiation ultimately required a multi-year collaboration (Nartey and Berchie 2015).

Wikimedia Argentina decided that if they wanted access to digital GLAM collections, their partners needed increased digitization capacity. Instead of hiring contractors, like in Indonesia, Wikimedia Argentina focused on training GLAM staff to do their own digitization. To set up GLAMs with appropriate capacity Wikimedia Argentina hired a Wikipedian in Residence who trains and supports a group of institutions in creating digitization workflows and uploading those batches of digitized content to Wikimedia Commons. The role takes on a role that most other residencies haven’t: generating knowledge before it reaches Wikimedia projects (“GLAM Case Studies: Wikimedia Argentina Digitization Project” 2016).

In Serbia, Wikipedian in Residence positions rarely focus on the development of capacity within the GLAM institution: because of limited Wikimedia Serbia resources and their partner institution capacity, Wikipedian in Residence positions focus on creating content and digitizing materials. Though not focused on expanding institutional open knowledge practices, they share Serbian heritage in venues unreachable otherwise (“GLAM Newsletter October 2017: Serbia Report” 2017; “GLAM Newsletter November 2017: Serbia Report” 2017).

Each of these communities developed a new definition of what it means to do batch uploads and Wikipedian in Residence positions. When developing GLAM-Wiki models and guidance, it is increasingly important for the more experienced communities (especially those in Europe) to recognize that their practices and models of activity change due to social context. However, at the same time, communities working in these less supported contexts need to learn the skills, practices and capacities acquired by more experienced GLAM communities. If we are going to act as an international community, GLAM-Wiki practitioners need to learn how to support these affordances as part of new contexts and models for GLAM.

However, some more sophisticated challenges extend beyond logistics. Effectively creating knowledge equity requires deep conversations between local Wikimedia communities and the knowledge-holding communities to develop sensitivity and awareness of how to share these cultures respectfully.

For example, the non-profit “Whose Knowledge?” campaign,9 which works with marginalized communities, identified a number of non-obvious challenges with sharing marginalized knowledge on Wikimedia projects. Through work with a Dalit community, a queer archive in Bosnia Herzegovina and a indigenous community in California and Baja, Mexico, the Whose Knowledge? team developed...

a conscientious series of diagnostic questions to ensure that they adequately understand and anticipate the concerns and priorities of marginalized communities when working within Wikimedia projects (Sengupta 2017).

In one of the more interesting museum partnerships in the movement, Wikimedia Poland worked with the National Ethnographic Museum to document the cultures of the Carpathian region. The project required deep training and support to volunteers and the careful supervision of museum staff to ensure effective and accurate collection of knowledge about the communities they were documenting (Szafran-Kozakowska, Pędzich, and Moraczewska 2016).

There are also strategic opportunities for existing, well-established GLAM-Wiki practitioners and communities to strategically invest their time and energy towards projects which elevate previously underrepresented topics:

The AfroCrowd and Black Lunch Table Initiatives (“Afrocrowd” 2018; “Black Lunch Table” 2018), centering around Afro-descent peoples and black artists respectively, have been very successful at developing programs based loosely on the success of the “editathon”, in part through the support of local Wikimedia community organizers and contributors in the New York area. These projects have also found strong allies among libraries, archives and other GLAM organizations, as hosts, communications support, and source material providers.

In Canada, two initiatives have emerged as part of the increased awareness within Canada of the erasure of Indigenous history and culture. First, working in direct response to the Truth and Reconciliation Commission’s recommendation that archives reevaluate their collections with Indigenous communities, York University in Toronto is exploring how to use Wikidata to develop metadata that better represents these communities (Allison-Cassin 2017). Additionally Wikimedia Canada is working with heritage organizations in Indigenous communities to write Wikipedia articles in their own languages (“Aboriginal Communities Outreach” 2017).

Wikimedia Netherlands runs a project called Netherland and the World, to collect requests from communities outside of the Netherlands for materials from GLAM institutions that fill strategic gaps, especially around materials collected as part of Dutch colonial history. This project allowed thousands of strategic uploads to Commons for a number of targeted communities working in other languages (“Wikimedia Nederland: The Netherlands and the World” 2016).

Collaboration with GLAM partners provides the potential for addressing Knowledge Equity throughout the Wikimedia movement. To do this effectively we need to anticipate and build capacity to work within the affordances of that knowledge context. As the 2017 Diversity conference highlighted, the movement must invest in infrastructure, access, capacity and people who can facilitate work in these new areas (“Wikimedia Diversity Conference 2017: Diversity Conversation” 2017).

One of the greatest underinvested components of GLAM-Wiki with these marginalized knowledge areas is Wikisource. A sister project of Commons, Wikipedia and Wikidata, and most akin to the more widely known “Project Gutenberg”, Wikisource publishes fulltext primary source materials and facilitates their transcription. Both digitally well-supported language communities, such as French and Italian, and digitally underrepresented languages, including various Indian language communities, have found Wikisource to be a powerful tool for sharing hard to find source materials
(“Wikisource: BAnQ” 2015; Panigrahi 2015). However, Wikisource also has technical challenges similar to those experienced by Commons: high technical and community barriers to participation. Wikisource could offer another platform for Wikimedia communities are serious about working with GLAMs to include underrepresented knowledge. With an improved Wikisource, GLAMs could start at the source materials contained within textual collections, and ensure that they get brought into digitally accessible formats, in order to eventually become sources for Wikimedia projects, including Wikipedia.

Learning together, towards a knowledge commons

As the GLAM-Wiki community embraces the 2030 movement direction, and collectively invests in the opportunities and open questions created by structured data and knowledge equity, both the Wikimedia movement and the larger collective heritage ecosystem will be healthier. Wikimedia and GLAMs share a common mission: to bring reliable knowledge to the public. Together we can build that aspirational knowledge commons on the internet: we can empower communities and institutions around the world to use Wikimedia projects as platforms, as infrastructure and collectively as portals to the most reliable knowledge on the web.

References


