

Allison M. Kittinger. ARMYPEDIA: A “Special Chronicle” of Multilingual Digital Crowdsourcing. A Master’s Paper for the M.S. in L.S degree. January, 2020. 37 pages. Advisor: Ryan Shaw

This case study investigates how ARMYPEDIA, a contributory digital archive, reflects the global nature of the BTS K-pop fandom. The research provides an overview of fan interaction with ARMYPEDIA by randomly sampling ARMYPEDIA posts and cluster sampling tweets about ARMYPEDIA, coding the content of the posts and tweets, and analyzing the codes for themes in languages used, text content, and multimedia contributions. The analysis of this case has potential implications for digital library spaces with regard to crowdsourcing, multilingual contributions, and engagement.

Headings:

Crowdsourcing

Fandom

Digital Libraries

Social Media

ARMYPEDIA: A “SPECIAL CHRONICLE” OF MULTILINGUAL DIGITAL
CROWDSOURCING

by
Allison M. Kittinger

A Master’s paper submitted to the faculty
of the School of Information and Library Science
of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements
for the degree of Master of Science in
Library Science.

Chapel Hill, North Carolina

January 2020

Approved by

Ryan Shaw

Table of Contents

Introduction.....	2
Literature Review.....	4
Multilingualism and Digital Archives	4
Multilingualism and K-pop Internet fandom	7
Contributory Digital Fan Spaces in K-pop	9
BTS and ARMYPEDIA.....	11
Methodology	14
Positionality / Researcher Role.....	14
Sample.....	15
Data Collection Methods	15
Data Analysis Methods	17
Ethical Considerations	17
Results.....	19
Discussion.....	26
Conclusion	30
References.....	32
Appendix A. Codebook	35

Introduction

In February 2019, Big Hit Entertainment launched ARMYPEDIA (<https://www.armypedia.net/>). ARMYPEDIA is a crowdsourced digital archive project created for Korean pop (K-pop) group BTS by the group's managing company Big Hit Entertainment and designed to document every day of BTS since the group's debut. The event lasted for one month, and over five hundred thousand BTS fans from around the world participated in unlocking and contributing to ARMYPEDIA (Benjamin, 2019). Fans posted on ARMYPEDIA in a variety of languages, resulting in a multilingual, transnational collection of fan memories and sentiments about the group.

Fans also discussed ARMYPEDIA extensively on social media, with Twitter being a major platform. BTS already had a very large following on Twitter; the BTS Twitter account holds a record for highest engagements and has several of the most-liked tweets on the platform. Big Hit Entertainment also launched a Twitter account specifically for ARMYPEDIA. Currently, it has over one million followers.

This case study sought to investigate how the ARMYPEDIA project reflects the global nature of the BTS K-pop fandom, ARMY. How was multilingualism present and/or supported within ARMYPEDIA? How did fans of different nationalities participate in multilingual interactions or acts of translation? Through a content analysis of tweets about and posts on ARMYPEDIA as well as a collection of background information about the project, this study explores answers to these questions. The results

of this analysis of ARMYEDIA have implications for digital library spaces with regard to crowdsourcing, multilingual contributions, and engagement.

Literature Review

Multilingualism and Digital Archives

Situating ARMYPEDIA as a multilingual crowdsourced digital archive warrants looking at the literature on multilingual crowdsourced digital archives. Borgman (1997) and Budzise-Weaver, Chen, and Mitchell (2012) both emphasize the importance of supporting multilingualism in digital library environments, highlighting the fact that the internet allows for highly multilingual and multimedia communication. In order to preserve data and make it accessible, digital libraries need to be prepared to provide content and search functions in different languages. Nichols et al. (2007) add that digital libraries are important means of information dissemination in developing countries.

There are some challenges in implementing multilingual functionality and access in digital library spaces. The first large, multilingual digital libraries were developed in the 1960s (Borgman, 1997), but the standards in these libraries were not developed with the future Internet in mind. As technology advances, standards have had to be revised to continue to support multilingualism in digital library spaces (Borgman, 1997). Current technical challenges involve issues of translation, designing for global systems, character sets, and interoperability (Borgman, 1997). In addition, Budzise-Weaver, Chen, and Mitchell (2012) noted that all of the four multilingual digital libraries they studied used human translators rather than machine translation techniques, and none of them employed cross-language information retrieval. This is both an issue of resources and translation quality for digital libraries.

Overall, crowdsourcing is important for addressing these challenges.

Technological advances have lowered the barrier to participation in collecting and producing information within digital libraries, opening up “new opportunities for minority cultures and developing countries to participate actively in our information society, rather than observing it from outside” (Nichols et al., 2007). Some of the libraries studied by Budzise-Weaver, Chen, and Mitchell (2012) employed volunteers to translate metadata and search functions, while others set up collaborative relationships to provide translations. Budzise-Weaver, Chen, and Mitchell (2012) concluded that collaboration and crowdsourcing are key to multilingual digital library efforts, as all four of the libraries they studied were products of collaborations between both organizations and individuals.

Crowdsourcing has long been a tradition in digital archives as well, especially in communities whose histories have been historically underrepresented due to “absences and biases in mainstream heritage narratives” (Flinn, 2010). Digital crowdsourcing and participatory technologies have given independent heritage and community groups increased ability to share information while bypassing spatial concerns, and “formal” cultural heritage and archival institutions are also looking to these technologies to increase diversity in their collections. (Flinn, 2010). Severson and Sauvé (2019) note that with regard to digital collections, “bringing in more perspectives and knowledge when describing or tagging collections, a process that was previously typically controlled by experts, can bring in a wider diversity of languages and allow us a deeper understanding of our descriptive practices” (p. 3).

Crowdsourced archive “formality” is well documented and presents challenges to acceptance in scholarly communities. Jones (2015) notes that crowdsourced archives “exist on a continuum of professional involvement, from the community-led archives with no professional archivists monitoring contributions and annotations, to the institution-led archives, where archivists, historians and other professionals or specialists guide the process in great detail” (p. 14). Within institution-led efforts, Oomen & Arroyo (2011) distinguish between contributory projects, collaborative projects, and co-created projects, with involvement of the public in the actual design and creation of the project increasing in each model (p. 139). While some researchers and scholars question the reliability and trustworthiness of community-led contributory archives (Flinn, 2010), Nichols et al. (2007) note that “although it is undeniably useful to disseminate information collections built in the developed world, as present digital libraries tend to do, a better strategy for sustained long-term development is to disseminate the capability to create information collections rather than the collections themselves,” allowing communities to preserve and share their own histories and cultures.

A challenge to institution-led crowdsourced archives is “finding sufficient knowledgeable and loyal users” to achieve a critical mass of contributions, especially in more involved projects that require more skills and knowledge to participate. Conversely, projects that are merely contributory and not co-designed or -created will attract a larger base (Oomen & Arroyo, 2011, p. 138).

There is little LIS literature, however, documenting multilingual crowdsourced digital archives created by non-GLAM institutions. This is an area that could be developed by future research in the field. How do these types of archives support

multilingualism? Do the tools and resources they use differ from those used in libraries and museums? Do the contents of these archives meet traditional archival quality standards? Do the motivations of corporations and other entities for creating these types of archives differ from the motivations of cultural heritage institutions? Such comparisons could be mutually beneficial to both GLAM institutions and non-GLAM institutions in terms of innovation, idea sharing, and supporting multilingualism.

Multilingualism and K-pop Internet fandom

The online K-pop fan community exhibits a strong tradition of multilingualism. The transnational spread of K-pop is often cited as an example of globalization that is not a process of Americanization or homogenization, demonstrating that “lots of new societies and cultures have the potential to become indigenized in various metropolises” (Riedel, 2020). Different languages coexist in the same fan space and often in the same message as many K-pop fans who do not speak Korean fluently incorporate romanized Korean phrases or construct simple sentences in hangul (the Korean writing system) to construct a “fan group identity” regardless of their native languages (Lee, 2018).

As a result of K-pop fans’ desire to both signal belonging in the fandom as well as communicate with Korean artists, fan boards and social media platforms like Twitter have become “informal learning environment to learn bits and pieces of the Korean language” (Riedel, 2020). Lee (2018) notes that the Korean wave is a main motivation for Korean language learners, and these learners rely not only on traditional language learning methods but also “K-pop and Korean dramas and movies, especially for learning colloquialisms.” Some company-mediated fan spaces have actually created their own Korean language learning series based on K-pop, such as the “Learn Korean with BTS”

series that was released on Weverse (<https://www.weverse.io/>), a platform for fans to communicate with artists under Big Hit Entertainment. In Korean drama boards as well, “English was a means for learning Korean and practicing multilingual identities” as fans tried to pick up Korean words and phrases from the shows they were watching (Kim, 2016). A popular Korean drama platform called Viki (<https://www.viki.com/>) has a tool called “learn mode” that presents Korean subtitles and subtitles in a preferred language simultaneously. To anecdotally corroborate this “informal learning” phenomenon, the author of this study notes that she taught herself hangul by practicing dictation of K-pop lyrics.

Because of this, translation is a strong tradition in online K-pop fan spaces. Before entertainment companies began widely translating Korean media for a global audience, fans played a prominent role in creating and circulating translations through social networks and video streaming sites (Lee, 2018). On Twitter, fan-run translation accounts publish translations of official content related to a particular group or company posted in Korean, from tweets to snippets of variety shows to entire speeches (Riedel, 2020). Fan subtitling (known as fansubbing) of K-pop music videos has taken place in YouTube comments and on VLIVE (<https://www.vlive.tv/>), a Korean celebrity livestreaming service that created a fansubbing tool known as V Fansubs specifically for this purpose (Aisyah & Nam, 2017). When an official Twitter account for a group does not translate something, one can see the process of translation happening in real time in the replies as well (Aisyah, 2017).

Acts of Korean language use and translation among international K-pop fans serve as connections to members of the fandom as well as sites of identity exploration

and language learning. The strong and widespread traditions of K-pop fan translation and informal language learning media have been officialized by mainstream Korean video platforms and entertainment companies with built-in tools for fansubbing and language learning, an implicit recognition of the work of these fans as integral to the continued globalization of Korean popular media.

Contributory Digital Fan Spaces in K-pop

Most of the spaces for fan engagement with K-pop artists would fall under Oomen & Arroyo's (2011) category of contributory projects; websites and fan spaces for K-pop stars are "created and maintained by well-established management companies," so fans are not participating in developing these spaces but are contributing comments and media to them (Lee, 2018). These spaces are highly mediated, with management companies moderating the spaces and removing hate comments (Lee, 2018). This provides a degree of quality assurance to these spaces. Because K-pop entertainment companies are big corporate entities, they have the resources to perform this moderation at a large scale.

The beginnings of these spaces were fanclubs and fancafes. The most widely used fancafe platform is DaumCafe (<http://top.cafe.daum.net/>), which is a forum-based space. There are official DaumCafes for artists set up by their entertainment companies on this site that allow artists to post and fans to respond in the forums. DaumCafes are open to all and free to join (Kakao Corp., n.d.). Fanclubs, on the other hand, are usually run as a part of an artist's official website. They often require paid membership, with different levels offering various benefits. Fanclubs are more exclusive, often only opening for recruitment a few times a year. In addition, many fanclubs do not allow international

fans, requiring signup with a Korean identification number (Park, 2018). However, due to the increasingly transnational nature of K-pop fandom, many fanclubs have opened to international fans, such as JYP Fan's (<https://fans.jype.com/>), the official fanclub platform for JYP Entertainment artists (n.d.). There are also video-based fanclubs, such as VLIVE Fanship, a paid membership platform for exclusive video content and other benefits. Like DaumCafe, VLIVE Fanship is a third-party platform on which entertainment companies can create Fanships for their artists. Spaces open to international fans usually have built-in translation functions, although some are limited in their language offerings.

A new and increasingly popular fan space offering is the company-created app. One example is Lysn (<http://www.lysn.com/>), created by SM Entertainment (n.d.). Lysn has two components: a free discussion space and messenger function as well as paid fanclub membership spaces. Another example is Weverse, an app with a similar discussion and messaging function to Lysn. However, artists under Big Hit Entertainment communicate directly with fans on Weverse without paywalls, although there is a merchandise shop within Weverse that requires in-app purchase (Big Hit Entertainment, n.d.). Both of these apps have translation functions for the facilitation of multilingual fan interactions with artists and each other.

ARMYPEDIA is another mediated contributory digital fan space from Big Hit Entertainment designed only for BTS, but it differs from the previously described types of fan spaces in important ways. First of all, ARMYPEDIA consisted of both an event and a digital archive product. Second, it is the only fan space among these to make use of crowdsourcing. Third, its primary function was not to facilitate fan communication with

BTS and each other but to engage fans in documenting and preserving the memory of the journey of BTS as a group. ARMYPEDIA is discussed in depth in the next section of this review.

The simultaneous evolution of the makeup of the international K-pop fandom, Korean entertainment companies, and technology has made way for increasingly innovative and transnational spaces for fans to interact with K-pop. The international popularity of K-pop shows no sign of slowing down, and the continued debut of new K-pop artists continues to expand the landscape of contributory digital fan spaces in the K-pop industry.

BTS and ARMYPEDIA

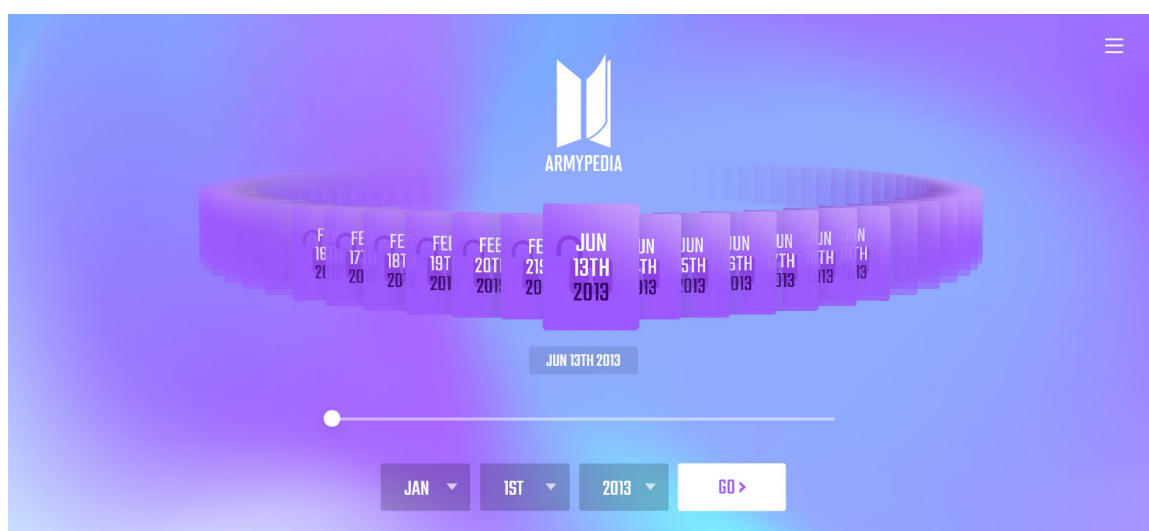
The fanbase of BTS, known by their fan name Adorable Representative MC for Youth (ARMY), is particularly engaged and invested in creating international fandom. One reason for this is that many fans relate to the message that the group has come to represent in their eyes. Many BTS songs contain relevant social commentary, and the group is known as a UNICEF antibullying ambassador as part of the “Love Myself” campaign. BTS is also seen as authentic, because BTS members participate in the songwriting and producing process as well as interact directly with fans on their official social media accounts. The relatability of the group’s message of positivity and self-love and the level of connection the group maintains with fans has cultivated a highly engaged and passionate fanbase (McLaren & Jin, 2020).

ARMYPEDIA was presented as an event celebrating fans’ connections to BTS. On the archive’s about page, ARMYPEDIA is described as “a special chronicle of the memories of ARMY’s journey together with BTS over each and every one of those

unforgettable days” from debut to the current day, which was February 2019 at the time, totaling 2080 days (Big Hit Entertainment, n.d.). The site also describes the original ARMYPEDIA event and how fans participated. Fans made an account on the ARMYPEDIA site and worked to uncover 2080 “puzzle pieces,” one for each day represented in the archive, scattered across the world in digital and physical formats. These puzzle pieces had QR codes that fans could scan, and to unlock the day fans had to correctly answer a question about BTS. After the day was unlocked, BTS fans could contribute memories in the form of posts related to that day and BTS. Posts could be liked in the form of a heart like Twitter, and the most-liked post on a particular day became the “top memory” of that day. Once all the dates were unlocked and filled, ARMYPEDIA was made available to the public and rewards were given to fans according to participation level in the form of Big Hit Shop points (Leelo, 2019). There were also fan events and livestreams related to ARMYPEDIA conducted throughout the month-long event (Benjamin 2019).

Figure 1

ARMYPEDIA Homepage



Note. ARMYPEDIA homepage. Retrieved from <https://www.armypedia.net/>

When ARMYPEDIA was made public, contributions were closed, and it now stands as a completed archive. The archive is organized by date, with a search interface. The site can be viewed in Korean, English, Japanese, and Chinese. Clicking on a date brings the user to the collection of posts fans contributed for that date. At the bottom of each page users can see the total number of posts and hearts on each date. Users can also see the number of hearts on each post, although posts can no longer be liked. Each post in the archive has a drop-down menu of languages that appears to use Google Translation API to provide automatic translations. Fans were able to post in any language they chose and could translate posts to read them in their native languages.

During the ARMYPEDIA event, fans were participating in conversations about it on Twitter. As BTS and BTS fans are particularly active on Twitter, when the ARMYPEDIA event began it became a part of fan conversations on the platform. ARMYPEDIA created the hashtag #ARMYPEDIA in English and Korean (complete with a special ARMYPEDIA logo emoticon) and translated tweets about ARMYPEDIA's launch and important events mainly into English and Korean. The research in this study will serve to explore how BTS fan Twitter conversations about ARMYPEDIA track with characterizations of online multilingual K-pop fan interactions and activities outlined in this review.

Methodology

For this case study, I reviewed announcements and fan-created descriptions of ARMYPEDIA and analyzed the content of ARMYPEDIA posts and tweets. I had also intended to interview the creators of ARMYPEDIA, but I never received a response after contacting the email address for the company listed on the ARMYPEDIA page. Because the nature of this research is exploratory, the data the study generated is qualitative, and the study examines a single phenomenon and its implications, the case study approach was determined to be the most suitable.

Positionality / Researcher Role

As the sole researcher of this study, I was responsible for all its parts: the literature review, data collection and analysis, and interpretation.

As for positionality, I am an insider at the intersection of the topics represented in this paper. My LIS track is in digital libraries, and I have a strong interest in digital scholarship. I also actively listen to K-pop and have an account on Weverse. Although I did not participate in ARMYPEDIA, I followed it closely as it unfolded. I am intrigued by nonlibrary online spaces and how they mirror and/or inform the field of library science, which is what led me to my research topic. ARMYPEDIA describes itself as a digital archive, so I wanted to explore what that means both for ARMYPEDIA and the field of library science.

Sample

For this study, one source of data is background information about ARMYPEDIA. Other sources of data are 327 posts on ARMYPEDIA sampled randomly, 384 Tweets about ARMYPEDIA sampled in clusters, and 10 posts on ARMYPEDIA sampled purposively, which make up the content analysis.

Data Collection Methods

Because ARMYPEDIA posts are generally organized by the number of hearts they receive, I used a simple random sample. I randomly selected dates and posts in the simple random sample. Then, I conducted a purposive sample of ARMYPEDIA dates that are important to fans. I also collected tweets containing the hashtags #ARMYPEDIA (English) and #아미피디아 (Korean) via clustered sampling.

A simple random sample of ARMYPEDIA posts and tweets will not include all languages used or all fan interactions regarding ARMYPEDIA. However, I had the ability to sample all posts made on ARMYPEDIA, and I chose to sample on Twitter rather than other platforms because BTS fans are very active on Twitter.

I first determined the sample size needed for both ARMYPEDIA posts and Twitter posts. For a 95% confidence interval when generalizing to the entire population of ARMYPEDIA posts, I determined that I should sample 109 dates for a total of 327 posts (three posts per date). For a 95% confidence interval when generalizing to the entire population of tweets, I have determined that I should sample 384 tweets. I split this number up between the two official ARMYPEDIA Twitter hashtags that will be included in the sample, #ARMYPEDIA and #아미피디아, for 192 tweets per hashtag. I

determined that the short post length on both ARMYPEDIA and Twitter makes these sample sizes feasible.

I originally intended to collect Twitter data manually with a random number generator, but I soon found that given the large number of posts and the fact that ARMYPEDIA and Twitter auto-load on scrolling, automating some of the data collection and initial number assignment to posts would be more efficient and accurate.

For the Twitter clustered sampling, I used the twint Python Twitter scraping tool (<https://github.com/twintproject/twint>) to scrape tweets containing “#ARMYPEDIA” and “#아미피디아.” I used twint because I needed to scrape tweets beyond the 7-day window of the Twitter API. I only scraped tweets that were written during the event, between February 22 and March 24, and I exported them to separate CSV files for the English and Korean hashtags. I then used Caltech Library data tools (<https://caltechlibrary.github.io/datatools/>) to select the 192 tweets for my sample in each file.

The structure and multimedia nature of the ARMYPEDIA website made it difficult to scrape, so I collected posts manually. For the ARMYPEDIA simple random sampling, I randomly sampled the range of dates (2080 days total) using a random date picker and then sampled three posts within each date by assigning them numbers and selecting them with a random number generator. For the purposive ARMYPEDIA sample, I searched for important dates in the history of BTS that I anticipated to be high-volume, which I defined as more than 1,000 posts. For recurring dates such as birthdays, I chose the dates with the most posts. Then, I sampled the most popular “top memory” post from each date. I collected the posts by taking screenshots.

The benefit of collecting ARMYPEDIA posts and tweets is that it gives a clear, rich picture of fan engagement with the archive, but I will also be relying on translation in these spaces to some extent, so my coding and results will only be as good as my translation abilities.

Data Analysis Methods

I first developed a preliminary codebook adapted from the UNC SILS master's paper by Jones (2015). I refined the codebook through my first pass at hand-coding the Twitter and ARMYPEDIA data. I then made a second pass at coding the Twitter and ARMYPEDIA data, ensuring the accuracy of codes and noting trends in the data. Finally, I generated charts in Excel to illustrate some of the trends in the data.

Limitations of this approach include the reduced validity of coding done by an individual researcher and the inability of content analysis to describe potential causes or reasoning underlying observations. Overall limitations include the language barriers and acts of translation involved in collecting and analyzing multilingual data.

Ethical Considerations

With regards to ethical concerns, publishing social media data runs the risk of identifiability for the authors. Even if the names of users are omitted, tweets can be searched by text on Twitter, leading to the original tweet and the profile of the user who wrote it. In addition, I run the risk of introducing bias in my study because I am the sole researcher and coder.

To mitigate these ethical concerns, I have not included whole tweets or posts verbatim or language that is easily attributable to one user, choosing to focus instead on wider trends and only quoting common sentiments. I understand that in doing this I

sacrifice some trustworthiness as my full sample is not included in my study, but I chose to prioritize the privacy and safety of the users.

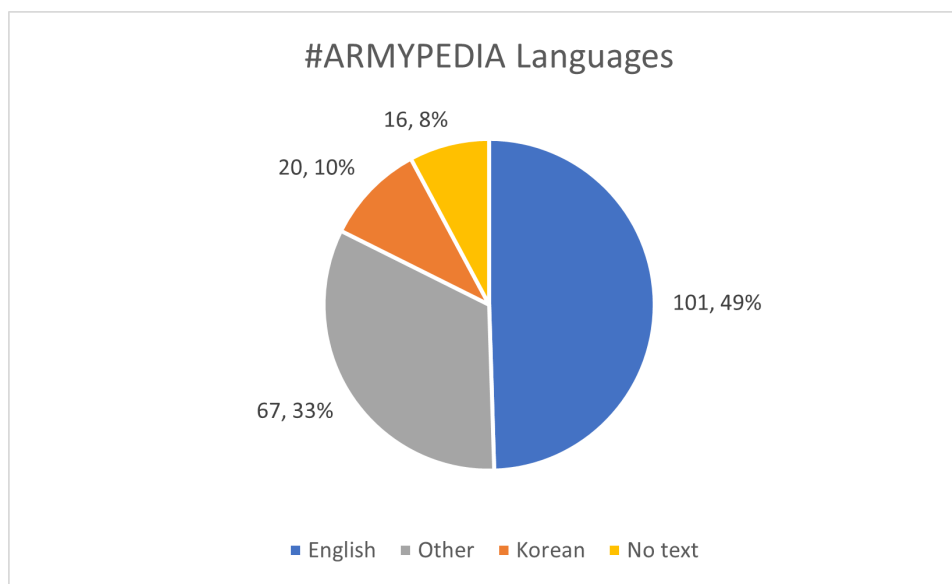
Results

Overall Twitter Trends

There are two significant trends present across tweets containing #ARMYPEDIA and #아미피디아. ARMYPEDIA tweets about the event itself made up the majority of both samples, at 88% of the #ARMYPEDIA tweets and 90% of the #아미피디아 tweets. In addition, the official BTS Twitter account was by far the most tagged account in tweets across both hashtags. This can most likely be attributed to the fact that BTS members have significant control over their social media accounts, inviting direct interaction from fans. The pattern of fans interacting with each other and the online presence of BTS to participate in ARMYPEDIA and its related events is evident in these trends, continuing the patterns of fan interaction found in previous studies. An interesting but minor trend is the additional use of the hashtags to share memories, mirroring the function of the ARMYPEDIA website.

#ARMYPEDIA

The #ARMYPEDIA tweets leaned towards being written in English and languages other than English and Korean, with 49% of tweets being written in English and 33% of tweets being written in languages other than English and Korean. Bilingual tweets were coded as containing multiple languages, so these percentages represent a count of languages that is larger than the original sample.

Figure 2*#ARMYPEDIA Language Distribution*

Note. Visualization of #ARMYPEDIA tweet text language distribution.

Within the #ARMYPEDIA tweets, the most significant source of interaction between fans was around the QR codes needed to unlock dates within ARMYPEDIA. Fans exchanged codes, listed codes and dates they unlocked, and provided answers to quiz questions. Some larger accounts aggregated these findings and kept running lists of found QR codes. This type of interaction continues the fan tradition of online information exchange within the fandom. The QR codes were visual and identified by number, making them amenable to exchange across languages.

The data also show that BTS fans were thinking about the languages and countries in which ARMYPEDIA was offered during the event. Although the ARMYPEDIA posts themselves are supported by the Google Translate API, the site infrastructure is only available in English, Korean, Japanese, and Chinese. This led to some acts of translation: two Twitter threads from BTS fan accounts in the data translated

the ARMYPEDIA instructions into different languages. In addition, nine tweets complained about either the limited availability of the ARMYPEDIA website in other languages or the lack of physical ARMYPEDIA codes in the users' countries.

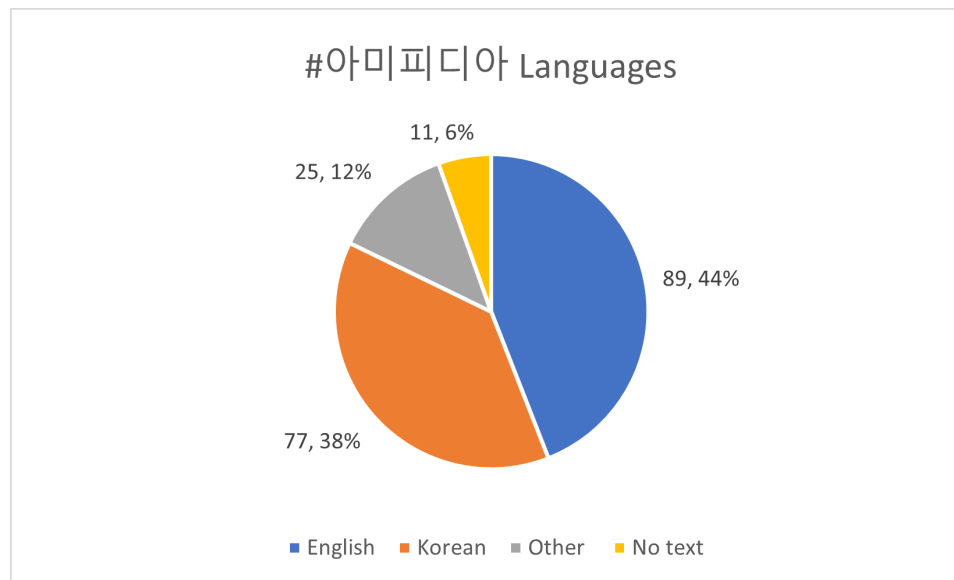
Illustratively, one tweet (quoted here in translation) notes that the location of most of the physical ARMYPEDIA codes in big cities across the world shows “how Big Hit sees the world.”

#아미피디아

Tweets containing #아미피디아 were more likely than tweets containing #ARMYPEDIA to be written in Korean and less likely to be written in languages other than Korean and English. A total of 44% of tweets were written in English and 38% in Korean, almost reversing the percentages of tweets written in Korean and tweets written in languages other than Korean and English in the #ARMYPEDIA tweets. Again, these percentages count individual languages in tweets containing multiple languages.

Figure 3

#아미피디아 Language Distribution



Note. Visualization of #아미피디아 tweet text language distribution.

One potential reason for this distribution is that these tweets were much more localized to Korea due to ARMYEDIA live events that happened in Seoul called Run ARMY in Action and ARMY United in Seoul. Many of the Korean-language fan interactions in these tweets focused on exchanging tickets, asking for tickets, or sharing their experiences related to these events. QR codes remained another part of fan interaction in this data, especially when it came to physical ARMYEDIA codes in Korea.

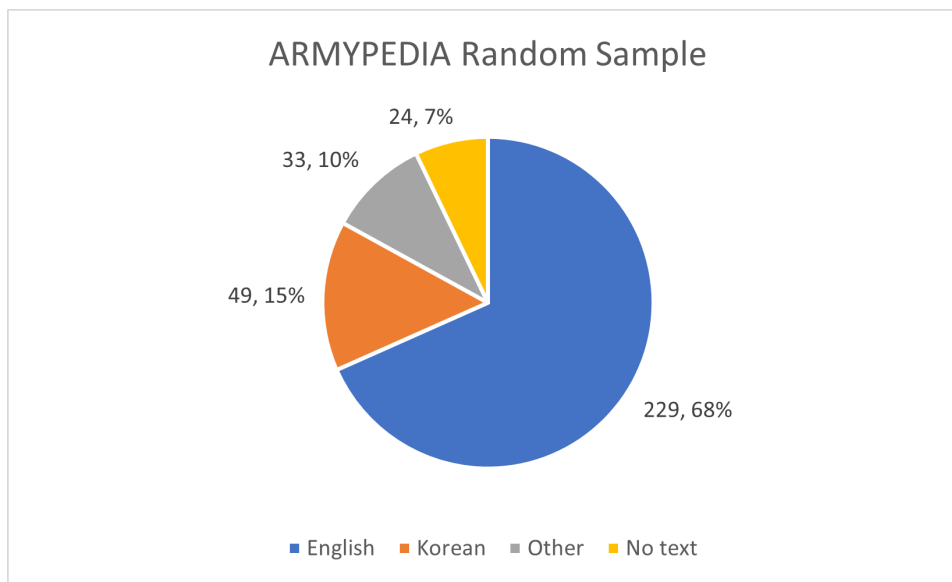
There were fewer acts of or thoughts on translation in the #아미피디아 tweets, with one tweet written in a language other than Korean and English that was posted early in the event wondering how the ARMYEDIA website would support translation and multilingual fan contributions to the archive.

Overall ARMYPEDIA trends

Perhaps not surprisingly, far fewer ARMYPEDIA posts than tweets were related to the event itself – only three in the random sample data. The majority of posts instead focused on memories of and feelings about BTS, with some others recounting personal events. In agreement with the findings of Lee (2018) discussed in the literature review, Korean was used to signal a “fan group identity” in ARMYPEDIA posts. Four posts in the random sample and one in the purposive sample written in hangul or romanized Korean were identified as being definitively written by non-native speakers. This trend was not evident in the Twitter data, either due to its lack of existence or a lack of definitively non-native Korean-language tweets. There may also be more posts in ARMYPEDIA written in Korean that could not be definitively attributed to non-native speakers. Despite these observations, the ARMYPEDIA post data is overall less multilingual than the ARMYPEDIA Twitter data. This is discussed further within the contexts of the two sets of ARMYPEDIA post data below.

ARMYPEDIA Random Sample

Out of the ARMYPEDIA posts in the random sample, 68% were written in English, 15% were written in Korean, and 10% were written in languages other than English and Korean. As in the Twitter data, these percentages count individual languages in tweets containing multiple languages separately.

Figure 4*ARMYPEDIA Post Language Distribution*

Note. Visualization of ARMYPEDIA post text language distribution.

English is more prevalent and languages other than English and Korean are less prevalent in the ARMYPEDIA posts when compared with the Twitter data. Potential explanations for this phenomenon include:

- 1) The language barrier of the website itself that some fans pointed out on Twitter
- 2) The fact that many fans addressed BTS directly through ARMYPEDIA posts. Fans who knew English but not Korean may have wanted to write to BTS in a language they knew at least some of the members would understand (the leader of BTS is fluent in English, and other members have been vocal about actively learning). This would presumably include native and non-native speakers of English, but these could not be reliably distinguished.

More research would need to be done – ideally involving ARMYPEDIA participants – to define the explanations behind the differences in language distribution between the ARMYPEDIA random sample data and Twitter data.

ARMYPEDIA Purposive Sample

Within the purposive ARMYPEDIA sample of top memories on high-volume dates, all of the posts were written either in English (three) or Korean (seven), with one post written in hangul by a self-identified non-native speaker. The purposively sampled posts were more likely to contain media than the randomly sampled posts; nine out of 10 (90%) of the purposively sampled posts contained either an image or video, whereas only 89 out of the 327 (27%) randomly sampled posts contained the same.

The same percentage of purposively sampled and randomly sampled posts (80%) contained text related to BTS rather than text about the ARMYPEDIA event or personal events. However, the purposively sampled posts all contained actual memories, whereas the randomly sampled ARMYPEDIA posts often contained text other than memories such as links, declarations of love, and simply the word “BTS.” Time posted did not seem to be a significant factor in the accumulation of hearts; some top memories were posted in March and April.

Given these results, it looks as though ARMYPEDIA users favored posts that were written in English or Korean and contained media as well as memories – whether or not they were of BTS. This makes sense in that the framing of ARMYPEDIA as a collaborative effort in memory creation makes such posts more relevant to the purpose of the archive.

Discussion

The Twitter and ARMPYEDIA data analyzed in this study clearly show two sides to the project: the archive itself and one of the digital spaces through which contributors were participating in and talking about it. The activities of unlocking, translating, attending events, and providing feedback all happened externally to the archive. Many fans also shared their ARMPYEDIA posts back to Twitter, signaling their participation and adding to the sort of external archive that the ARMPYEDIA hashtags became on the platform.

This demonstrates the importance of meeting potential crowdsourcing project contributors and users where they are and giving them an external interaction space. Big Hit and BTS are aware that their fans interact in great numbers on Twitter, so the company situated ARMPYEDIA's activity hub there with its own account and customized hashtags. Some GLAM institutions may also have large Twitter followings that they can direct to crowdsourcing initiatives, or perhaps the most effective outreach strategy would be through a newsletter, or even word of mouth. GLAM institutions are (or should be) more localized and tapped into their communities than a globally oriented company like Big Hit Entertainment, and collaborating with motivated potential participants in the community could be helpful in shedding light on where to locate a crowdsourcing project's activity hub to maximize participation if no obvious platform has emerged through observation. This also benefits GLAM institutions because they can

readily gather feedback about the project by looking at what users and contributors are saying about it on their platform of choice.

That being said, agility in the face of such feedback is crucial, especially when it comes to digital crowdsourcing projects in GLAM and cultural heritage institutions. Online interaction moves quickly, and efforts to improve the experience of the project for users must keep pace. ARMYPEDIA lost potential participants because the website was not available in their languages, and never will be. In creating the archive, Big Hit Entertainment focused more on providing translation methods for the user-submitted content than the site infrastructure itself, and there was perceived bias towards big cities regarding the physical locations of ARMYPEDIA QR codes. Of course, using the automated Google Translation API on the posts is simpler than manually translating website infrastructure in terms of both time and money, and the company may have been merely attempting to disseminate physical QR codes in high-population areas where they felt the codes would be more likely to be found by fans. Regardless, the reasoning behind these choices around languages and locations were not acknowledged or explained and led some BTS fans to feel that their communities were left out.

This reiterates the importance of multilingual digital infrastructure outlined in Weaver, Chen, and Mitchell (2012). When the translatability of content is not supported by translated infrastructure, the true potential range of multilingual submissions will not be accommodated or reflected in the submissions. When undertaking a digital crowdsourcing initiative, GLAM institutions should consider prioritizing the translation of the digital infrastructure present such as instructions, search functions, etc. over functions to translate content. This, as noted in the literature review and evidenced by

ARMYPEDIA, is challenging and can take time and resources that some institutions may not have in house. There are additional functions and archival standards to consider that are not present in ARMYPEDIA, such as searchability and discoverability. Translation efforts are where collaborative projects as defined by Oomen & Arroyo (2011) can prove fruitful both in the translation of content and community connections.

Big Hit Entertainment also chose very intentionally to make ARMYPEDIA a space for fan contributions. It should be noted that although criticism of ARMYPEDIA from fans was evident on Twitter, none of these criticisms were found in the ARMYPEDIA post sample. Whether this is due to fans censoring themselves or imposed censorship (or both) is unclear, but it does highlight the way in which a population and its beliefs can influence an archive's content. ARMYPEDIA is named for ARMY, for BTS fans, so there is an implicit rule in this naming that "true" fans will only make complementary or positive contributions to the archive, when in fact many fans criticize out of love or a desire to improve the reputation of the object of their fandom. Ideally GLAM archives should be very transparent about their positionality in relation to the subject matter as well as any potential biases found in the archival content or the contribution program to maximize contributor comfort with sharing their true thoughts and minimize the overrepresentation of a single perspective.

Of course, quality control and content moderation are still important in contributory archives. GLAM archives may deem some contributions unsuitable due to profanity, irrelevance, or any number of things. There is always this risk when opening up contributions to the community, but there are some ways it can be combated as well, such as making contribution removal guidelines clear to contributors. Big Hit

Entertainment was explicit about removing profane and spam posts from ARMYPEDIA, but they also mentioned removing posts that “violate the intention” of ARMYPEDIA. This “intention” is not stated, and what this means in terms of post content is not elaborated on. This is most likely intentional to give ARMYPEDIA administrators some leeway in removing things like criticism and disagreeable opinions, but it may have also led to increasingly biased post content due to contributor uncertainty about what would be removed. What ARMYPEDIA should probably have done (and GLAM institutions should do) if they truly wanted to represent a range of opinions from contributors is demystify the intention of the archive and clearly state the line between what counts as hate and what counts as constructive criticism, and between what counts as misinformation and what counts as opinion. These lines can and will be different depending on, again, the intention of the archive; what is important is making any intentions and lines clear to the community.

A final important consideration in digital crowdsourcing that ARMYPEDIA illustrates is transparency and communication. Especially in more collaborative or community co-created projects, it is crucial to communicate and explain—or even invite feedback on—decisions about the project. Showing the work is intentional and created with its users in at least some capacity will make the project more trustworthy and inclusive.

Conclusion

Although contributory digital archives remain a niche study in GLAM institutions, taking the full picture of how this method of archival crowdsourcing is being used across contexts and disciplines is imperative in understanding community needs and motivations for participating in such a project. ARMYPEDIA is one such context, and the needs and motivations of its contributors are explored in depth here.

Limitations of this study include the subjectivity of coding done by an individual researcher and the inability of content analysis to describe potential causes or reasoning underlying observations. Limitations to the scope of my research include the platforms ARMYPEDIA and Twitter and the language barriers and acts of translation involved in collecting, analyzing, and translating multilingual data.

These limitations, though, could lead to more questions: how does ARMYPEDIA compare to other digital spaces for BTS fans in terms of interaction and multilingualism? What about Twitter and other social media platforms? Are these results reproducible? Future studies could carry this work forward by investigating these questions, applications for this research, or comparisons between ARMYPEDIA and other multilingual digital archives in library or institutional contexts.

This study has implications for multilingualism in fandom studies as well as multilingual participation in contributory digital archives. Most of the current literature published about ARMYPEDIA emphasizes its fandom context and implications therein rather than its function as a digital archive and what that might mean for other

digital archives. It is my hope that this analysis of how the transnational BTS fandom interacted with ARMPYEDIA can inform future applications and strategies for multilingual participation in digital archives.

References

- Aisyah, A. (2017). Korean-English Language Translational Action of K-Pop Social Media Content: A Case Study on Bangtan Sonyeondan's (BTS) Official Twitter. *The Southeast Asian Journal of English Language Studies*, 23(3), 67-80.
- Aisyah, A. & Nam, Y. J. (2017). K-Pop V Fansubs, V LIVE and NAVER Dictionary: Fansubbers' Synergy in Minimising Language Barriers. *The Southeast Asian Journal of English Language Studies*, 23(4), 112-127.
- Benjamin, J. (2019). How BTS Gathered Over Half A Million Fans For Digital Record Project ARMYEDIA. Forbes.
<https://www.forbes.com/sites/jeffbenjamin/2019/04/03/bts-armyedia-interview-big-hit-entertainment-half-a-million-army/#334d0602d9a9>
- Borgman, C. L. (1997). Multi-Media, Multi-Cultural, and Multi-Lingual Digital Libraries. *D-Lib Magazine*. Retrieved from
<http://www.dlib.org/dlib/june97/06borgman.html>
- Budzise-Weaver, T., Chen, J. & Mitchell, M. (2012). Collaboration and Crowdsourcing. *The Electronic Library*, 30(2), 220-232.
- Flinn, A. (2010). Independent Community Archives and Community-Generated Content: 'Writing, Saving and Sharing our Histories.' *Convergence*, 16(1), 39-51.
- Jones, J. E. (2015). We share what we are: User contributions and annotations in digital contributory archives [master's thesis, University of North Carolina at Chapel

Hill]. Retrieved from

https://cdr.lib.unc.edu/concern/masters_papers/hm50tw497?locale=en

Kim, G. M. (2016). Practicing Multilingual Identities: Online Interactions in a Korean

Dramas Forum. *International Multilingual Research Journal*, 10(4), 254-272.

Lee, J.S. (2018). The Korean Wave, K-Pop Fandom, and Multilingual Microblogging.

Multilingual Youth Practices in Computer Mediated Communication (C. Cutler,

Ed.). Cambridge University Press.

Leelo, J. (2019). What Is ARMYEDIA? BTS' New Website Is Enlisting ARMY To

Write Their History. *Elite Daily*. [https://www.elitedaily.com/p/what-is-](https://www.elitedaily.com/p/what-is-armyedia-bts-new-website-is-enlisting-army-to-write-their-history-15988384)

[armyedia-bts-new-website-is-enlisting-army-to-write-their-history-15988384](https://www.elitedaily.com/p/what-is-armyedia-bts-new-website-is-enlisting-army-to-write-their-history-15988384)

McLaren, C. & Jin, D. Y. (2020). “You Can’t Help But Love Them”: BTS, Transcultural

Fandom, and Affective Identities. *Korea Journal*, 60(1), 100-127.

Nichols, D. M., Witten, I. H., Dewsnip, M., Bainbridge, D. & Keegan, T. T. (2007).

Digital libraries and minority languages. *New Review of Hypermedia and*

Multimedia, 11(2), 139-155.

Oomen, J. & Arroyo, L. (2011). Crowdsourcing in the Cultural Heritage Domain:

Opportunities and Challenges. In M. Foth (ed.), *C&T '11: Proceedings of the 5th*

International Conference on Communities and Technologies (pp. 138-149). New

York: Association for Computing Machinery.

Park, A. (2018). *BLACKPINK Opens 1st Recruitment for Official Korean BLINK Fan*

Club. Blackpink Update. [https://blackpinkupdate.com/blackpink-opens-1st-](https://blackpinkupdate.com/blackpink-opens-1st-recruitment-for-official-korean-blink-fan-club/)

[recruitment-for-official-korean-blink-fan-club/](https://blackpinkupdate.com/blackpink-opens-1st-recruitment-for-official-korean-blink-fan-club/)

- Riedel, L. -A. (2020, March 9). K-Pop as a linguistic phenomenon. *diggit magazine*. <https://www.diggitmagazine.com/articles/k-pop-linguistic-phenomenon>
- Severson, S. & Sauvé, J. -S. (2019). Crowding the Library: How and why Libraries are using Crowdsourcing to engage the Public. *Partnership: The Canadian Journal of Library and Information Practice and Research*, 14(1).
<https://doi.org/10.21083/partnership.v14i1.4632>

Appendix A. Codebook

Category	Description
Image - person	Image of a person or people
Image - place	Image of a place
Image - thing	Image of an item or artifact
Text - event	Text related to the ARMYPEDIA event
Text - BTS	Text related to BTS
Text - personal	Text related to a personal event
Video	Embedded video clip
Audio (not video)	Embedded audio clip
Link	Contains a link to another page (including tags)
Emoticon	Emoticons used in the post or tweet
Other hashtags	Hashtags besides #ARMYPEDIA/#아미피디아
Language - English	Text written in English
Language - Korean	Text written in Korean
Language - other	Text written in languages besides English and Korean
Translation	A translation of text