Tag analysis as a tool for investigating information behaviour: comparing fan-tagging on Tumblr, Archive of Our Own and Etsy.

Journal:	*Journal of Documentation*
Manuscript ID	JD-05-2020-0089.R1
Manuscript Type:	Article
Keywords:	tag analysis, information behaviour, folksonomy, Tumblr, Etsy, Archive of Our Own, tagging, fandom

Title: Tag analysis as a tool for investigating information behaviour: comparing fan-tagging on Tumblr, Archive of Our Own and Etsy.

Abstract

Purpose: This article describes the third part of a three-stage study investigating the information behaviour of fans and fan communities, the first stage of which is described in Price and Robinson (2017).

Design/methodology/approach: Using tag analysis as a method, a comparative case study was undertaken to explore three aspects of fan information behaviour: information gatekeeping; classifying and tagging; and entrepreneurship and economic activity. The case studies took place on three sites used by fans – Tumblr, Archive of Our Own (AO3), and Etsy. Supplementary semi-structured interviews with site users were used to augment the findings with qualitative data.

Findings: These showed that fans used tags in a variety of ways quite apart from classification purposes. These included tags being used on Tumblr as meta-commentary and a means of dialogue between users, as well as expressors of emotion and affect towards posts. On AO3 in particular, fans had developed a practice called 'tag wrangling' to mitigate the inherent 'messiness' of tagging. Evidence was also found of a 'hybrid market economy' on Etsy fan stores. From the study findings, a taxonomy of fan-related tags was developed.

Limitations/implications: Findings are limited to the tagging practices on only three sites used by fans during Spring 2016, and further research on other similar sites are recommended. Longitudinal studies of these sites would be beneficial in understanding how or whether tagging practices change over time. Testing of the fan-tag taxonomy developed in this paper is also recommended.

Originality/value: This research develops a method for using tag analysis to describe information behaviour. It also develops a fan-tag taxonomy, which may be used in future research on the tagging practices of fans, which heretofore have been a little-studied section of serious leisure information users.

Keywords

tag analysis; fandom; information behaviour; folksonomy; Tumblr; Etsy; Archive of Our Own; tagging

Introduction

The purpose of this paper is to describe the final part of a three-stage study into the information behaviour of fans and fan communities. The first two parts comprised a literature analysis and a Serious Leisure Delphi study, and are described in Price and Robinson (2017). The Delphi study identified three aspects for further investigation: information gatekeeping, classifying and tagging; and entrepreneurship and economic activity. The final stage took the form of a comparative case study.

Since tagging was a topic of some contention in the Delphi study, it was decided to investigate these three aspects (i.e. information gatekeeping; classifying and tagging; and entrepreneurship and economic activity) through the practice of fan-tagging on 3 different sites used extensively by fans – Tumblr, Archive of Our Own (AO3) and Etsy. Fan-tagging is essentially synonymous with tagging, but is used to denote tagging practices in the context of fan activities. Tag analysis was the method used

to investigate fan-tagging practices on the three sites. Upon completion of the tag analysis, a comparison was made of the findings from each site, supplemented by additional interviews with fans who were users of the websites. The results of the case study then helped to develop a fan-tag taxonomy.

Each site under investigation serves the fan community in different ways, and their information architecture is very different. Tumblr is a social media blogging site which allows traditional long-form text-based posting, although the site is more heavily skewed towards audiovisual material than most blogging platforms, with animated gifs, memes, and video clips circulating widely. Website links, chat transcripts and quotes can also be uploaded (Attu & Terras, 2017). Since Tumblr does not have a traditional commenting functionality, unless an entire post is reblogged by another user, the tagging function "has been creatively exploited by some users beyond its folksonomic use" for "discourse functions such as expressing an opinion, a reaction, or including asides" (Bourlai, 2018, p. 46).

AO3 is a fanfiction repository which stores mostly text, although other formats such as images, audio and video may be embedded into a post using HTML. Despite this, the vast majority of works are text-based. Users may tag their posts freely, or use pre-defined and/or canonical tags which have been already been approved by the site's volunteer tag wranglers, "volunteers who look through tags and match non-canonical tags with canonical tags to ensure some level of control over terminology" (Johnson & Forsythe, 2019, p. 17). This system is "a unique combination of author-generated metadata and backstage volunteer federation" (Gursoy, Wickett & Feinberg, 2018, p. 492), similar to Hidderley & Rafferty's (1997) concept of democratic indexing. Tag wranglers must exercise their judgement and expertise in standardising tags, but it is not yet clear how accurate their judgement might be, whether their wrangling choices accurately reflect the intentions of the original tagger, or whether they inadvertently practising a form of information gatekeeping by choosing to merge tags that are not synonymous to the original tagger.

Lastly, Etsy holds a very different function as an online marketplace for small businesses dealing in craft, vintage or handmade goods. Custom fandom merchandise is produced by many Etsy stores, although these are unlicensed (King & Ridgway, 2019). Similar to AO3, users may freely tag their posts, or use pre-defined tags. Thus far, investigation of tagging on Etsy has been limited, with Blanchflower & Hodges (2014) being the only research of significance on this topic; however, this study does give intriguing insight into how strategic tagging could bolster the success of selling an item.

Each platform, with their individual, unique functions, was chosen to address each of the three specific aspects under investigation. Tumblr, with its "discourse functions" of tagging (Bourlai, 2018), was used to investigate how tags are used to communicate between fans online; AO3, with its tag wranglers, to investigate information gatekeeping on the site; and Etsy, as an online marketplace, to investigate the practices and attitudes of fan entrepreneurs online.

As stated in Price and Robinson (2017), the aim of the study is to examine the information behaviour of fans from a library and information science (LIS) perspective. While serious leisure and fun life contexts have been a subject of research within LIS in recent years (Ocepek et al., 2018), fandom itself has rarely been so. It is therefore a domain that is deserving of deeper consideration, especially when considering the rich, complex and diverse information practices as already examined in Price and Robinson (2017). Understanding this domain would benefit not only LIS, but also other disciplines, such as the media industry, which may wish to understand how intellectual properties are consumed and remediated; publishers and policymakers may find fan publishing paradigms and

attitudes to copyright of interest; and educators teaching information and digital literacy may benefit from an understanding of fan practices. Future exploration may include the domains of mental healthcare, and the positive effects fan activities may have on both fan creators and consumers (c.f. Anderson, 2019).

Specifically, the case studies reported here aim to address the following points:

- Understanding fan-tagging practices may help to improve information architectures used in fan contexts;
- Fan-tagging practices will give insight into how fans share, disseminate, promote and organise their work;
- Fan-tagging practices may give insight into the non-classificatory ways in which fans use tags, such as on Tumblr.
- Investigating the practice of 'tag wrangling' on AO3 may tell us if and how fans act as information gatekeepers;
- Investigating fan sellers on Etsy may have wider implications for domains outside of LIS, i.e. the media industry, copyright law, the sharing economy (Cohen & Sundararajan, 2015) etc.

A more in-depth discussion of these case studies, including thorough details of the tag analysis process, can be found in chapter 5 of Price (2017).

Scope and limitations

The three case studies focus on 3 online platforms – Tumblr, Archive of Our Own (AO3), and Etsy. All have very different functions, but are widely used by fans. Tumblr is mainly an image-based social media site; AO3 is a fanfiction repository; Etsy is a marketplace for small businesses primarily selling handmade, vintage, or arts and crafts goods. As we have discussed, these sites were specifically chosen to investigate and address the objectives outlined above.

The case studies focus on one particular fandom, or sub-fandom, specifically called 'Romy'. Romy fans follow the 'ship', or relationship, between two characters called Rogue and Gambit, who are mutant superheroes from Marvel's *X-Men* franchise. The term 'Romy' itself is a portmanteau of the two character's names, i.e. Rogue + Remy = 'Romy'. Both this term and its synonyms are widely used in online contexts to tag fanworks (i.e. fan-made creative works) related to these two characters. The reason for choosing this fandom is twofold. Firstly, it is a relatively small fandom and easier to investigate than more popular ones such as Harry Potter and Star Wars. Secondly, fan nomenclature tends to be rather impervious and difficult for outsiders to understand. Some Marvel Universe-specific fan terms include: 'ironshield' (to denote a pairing between the characters Iron Man and Captain America); 'fem!clint' (to denote a gender-swapped Hawkeye character); and 'MPreg!Charles' (to denote male pregnancy in the character Charles Xavier). Many Marvel tags are references to obscure characters, story arcs, plot elements and locations that are also inscrutable to outsiders. To analyse the 'Romy' tag and all its co-occurring tags, therefore, would require either substantial knowledge of the Marvel Universe, or substantial research in the area to familiarise

oneself with it. As the author has been a long-time fan of the Marvel Universe and the Romy fandom in particular, an analysis of the 'Romy' tag and its synonyms was chosen for the case studies, as the author's expertise would greatly reduce the time needed researching tag meanings, and also improve coding accuracy.

There are, of course, limitations to the presented case study. Case (2012) has pointed to the fact that case studies can only show a snapshot in time, and this is especially pertinent in the case of tagging, where tags may reflect passing trends, fads, or events. The quantitative data collected in this study reflects a snapshot of time in 2016, and of course, changes will have taken place on all these platforms, particularly on Tumblr, whose user base was drastically reduced after its December 2018 adult content ban (Bronstein, 2020). Additionally, as this case study focuses solely on the Romy fandom, results should not necessarily be generalised to the wider fan community.

Also, manually coding tags is of course very subjective and there are many instances where a tag can likely fit into more than one category. For example, the term 'X-Men' may refer to the Marvel superhero team, or it may refer to the fandom itself. To put this in simpler terms, using Smith's (2008a) tag categories (see Table 1), the term 'my art' might refer to a resource, or it might be self-reference. In such cases, it is necessary for the researcher to make an informed judgement as to the category that best describes the tag, which naturally leaves the potential for erroneous judgement. Such biases can be mitigated using coding teams, but unfortunately the limited time, resources and expertise available for this study did not allow for this.

Literature review and rationale

Tag analysis, wherein the network properties of tags are analysed, has its roots in hyperlink network analysis, or link analysis (Thelwall 2004), which in turn has its roots in social network analysis (Park and Thelwall 2003), and has seen growing use in the LIS field, and is popular in analysing the use of Twitter hashtags. For examples of research using network tag analysis, see Ahmed & Lugovic, 2019; Buarki & Alkhateeb, 2018; Ewbank, 2015. For other examples of LIS studies looking at tagging in wider contexts, Navarro Bullock, Hotho & Stumme (2018) discuss how effective tags are when accessing information; Samanta & Rath (2020) compare the usefulness of tags and controlled subject headings on LibraryThing; and Szuban (2018) investigates a videogame wiki to appraise how the use of a folksonomy can complement the top-down work of a hierarchical structure. Sarkar & Bhattacharya (2020) give an excellent overview of comparative studies of folksonomies and controlled vocabularies within LIS.

It is also worth mentioning here that research into tagging in fan contexts has been recently conducted. Messina (2019) evaluates tag usage in *The Legend of Korra* fanfiction from a genre studies perspective; Hoch's (2018) paper uses content analysis to examine the posts using the fan-tags "zoethian" and "sjips" on Tumblr, in order to highlight fan conduct norms; Gursoy, Wickett & Feinberg (2018) investigate the different uses of tags within fanfiction repositories; and Navar-Gill & Stanfill (2018) looks at tagging from the lens of media studies, discussing queer fan-tag campaigns as interventions in the creation of media products.

Attu and Terras (2017) give a summary of the academic research that has been done on Tumblr thus far. Their overview shows that "blogging, education, libraries, identity, and fandom are the most frequent topics in Tumblr-focused research" (2017, p. 544). No mention, however, is made of any research on Tumblr tagging metrics, in stark contrast to Twitter. This suggests that the kind of research reported in this paper is novel and should present some new insights.

Attu and Terras' work only covered papers up to 2015; more recent research includes Misailidou (2017), McGuire (2017), Bury (2016), Gonzalez (2016), Petersen (2017), Bourlai (2018), and Hoch (2018). Of these studies, only Bourlai (2018) and Hoch (2018) look specifically at Tumblr tagging. However, the emphasis appears to be on the discourse functions and fannish cultural norms observed in fan-tagging, rather than on information behaviour, the network properties of fan-tagging, or the development of a fan-tagging taxonomy. This further verifies the novelty of the research presented here.

Some research has been done on AO3's tagging system. As seen above, Messina (2019) looks at AO3 tags from a writing analytics and genre perspective; Gursoy, Wickett & Feinberg (2018) recommend creating tagging systems that allow for expressivity, according to the findings of their research, which concluded that fans use tags creatively, as well as for information organisation purposes. Prior to this Johnson (2014) gave an overview of how tagging affected the findability of fanfiction on AO3; and Dalton's (2012) Masters thesis surveyed AO3 users about their perceptions of the site's tagging structure.

Research into Etsy's tagging system is sparser. Blanchflower & Hodges (2014) mention tagging briefly as a strategy for improving traffic to products on the site; and Close (2016) briefly touches on similar findings of strategic tagging. As yet, this appears to be the extent of research on this topic.

The above studies are useful in that they shed some light on tag users' behaviour in several ways – however, most are *not* conducted from a LIS perspective, nor is their intention to investigate information behaviour per se. Where the research presented here differs from these studies is that it employs tag network analysis as a quantitative method to investigate human information behaviour, undertaken firmly within an LIS context. As far as can be ascertained, LIS has not used either tag or network analysis in this way before i.e. as a method to investigate human information behaviour. Since information behaviour has been defined as "the totality of human behavior in relation to sources and channels of information, including both active and passive information seeking, and information use" (Wilson 2000), tag analysis can offer some unique insights into how humans relate to information, in the way they seek, classify, organise and manage it. Not only this, but it may also shed light on the thoughts and motivations behind human information behaviour.

First, and most obviously, using network theory measurements such as betweenness centrality can illustrate how effective a tag is as an information or content carrier; a densely clustered group of tags will imply that they are highly influential within the network. But the type of tags used, as well as their frequency, can provide important information on how humans interact with information. They can tell us how users organise information and communicate on different platforms (e.g. Flickr, Twitter, Instagram, Tumblr, Bibsonomy etc.), and whether those methods change in different environments. Tag types can also suggest the motivations behind tagging, and the community and/or cultural norms behind tagging practices (such as Wargo's #socialjusticewarrior users on Tumblr). For example, the use of the emoticon tag type (as seen in Table 2) can indicate how an individual tagger affectively annotates a resource. Looking at the use of the emoticon tag type within an entire Tumblr dataset, and comparing its use within an entire Instagram dataset, can tell us whether affective annotations are more important on Tumblr or Instagram. It can also tell us about the motivations for tag usage – are users more concerned with using tags to organise resources, or to signify social identity? Are they more interested in using tags to disseminate resources, or to annotate them?

This research builds on the premise that tags can provide a new and innovative method for investigating human information behaviour. Since this is a novel application for this method within

LIS, its use here is exploratory, but it is hoped that this can be refined and developed in further research.

Ethics

The research received ethical approval from the relevant City, University of London, ethics committee.

The tag analysis required the use of web crawling, which brings with it particular ethical concerns. Thelwall and Stuart's (2006, p. 1777) policy of ethical webcrawling was used, namely:

- Email webmasters of large sites that are about to be crawled to allow them to make an informed decision to opt out.
- Obey the robots.txt convention.
- Follow the robots guidelines.

Additionally, all data obtained using webcrawling methods was anonymised, with any data which might lead to the identity of those included on the dataset (such as blog URL) removed.

Methods

The comparative case study was split into two sections. The first was a social media data analysis, specifically using tag analysis as a research method. The second comprised semi-structured interviews. For the first section, several programmes were used to aid in the social media data analysis. Tags were scraped from all three sites using different methods:

- Tumblr: An open source Python script, get_tagstats, available on GitHub, was used to access the Tumblr API and scrape tags used on the site. This necessitated the installation of Python 2 to run the script. A Tumblr account was needed to use the API.
- AO3: The free web crawler, SocSciBot, developed by Mike Thelwall and the Statistical Cybermetrics Research Group at the University of Wolverhampton, was used to scrape tags. Pajek, an open source programme for analysing and visualising networks, was used to a render the resulting dataset in a format that could be imported into NodeXL.
- Etsy: A free online tool, EtsyRank, was used to scrape tags from Etsy. Since EtsyRank is connected to the Etsy API, an Etsy account is needed to use it.

All resulting datasets were cleaned in Microsoft Excel before being imported into the NodeXL network analysis tool (Smith, Shneiderman, Milic-Frayling et al. 2009).

In the second section of the study, semi-structured interviews were conducted with taggers and tag-wranglers, and the data analysed thematically in Nivo.

The sites

A brief description of the tagging systems used in each of the 3 sites investigated in these case studies is given here, to better facilitate an understanding of how these systems affect the social media data analysis.

Tumblr

Tumblr uses a hybrid tagging system. Firstly, it is a self-tagging system, i.e. a system where a user can only tag the resources they have created (Zollers, 2007). Secondly, it incorporates an automanual function, where Tumblr itself will suggest recommended tags based upon the first few characters typed in, previous tags the tagger has used, and the most popular tags used on the site (Warner, 2011; Smith, 2008b). Lastly, the site allows the reblogging of other users' resources, and when reblogging, a user has the option of retagging the resource with their own terms. Of note is the fact that only the first 20 tags are indexed by the site (Unwrapping Tumblr, 2014).

[Take in Figure 1]

Figure 1 shows a screencap of a Tumblr post by Participant B. The resource in this picture is an image (fanart) created by Participant B of the Marvel characters Rogue and Gambit. Users may also add descriptions to their post; here Participant B has added a short caption to her artwork, followed, in parentheses, by a communique to another user. Below the description can be seen the tags. These follow the hashtag format popularised by Twitter; however, unlike Twitter, spaces are allowed between words; there are no restrictions to formatting (e.g. numbers and special characters are allowed), nor are there any restrictions on the amount of tags that can be used. Descriptive tagging is used ("#romy", "#rogue", "#gambit"); but another form of tagging, unique to Tumblr, is also used. Three tags are used to express the affective impressions of the tagger, while also enriching the description of the resource itself ("#still pissed about uncanny avengers 5", "#this is how i vent", "#well this and alcohol"). These three tags are separated into phrases, but together form a coherent sentence or thought. Not only do they express affect, but they also communicate Participant B's fan identity (i.e. her disappointment with a certain comic book and its depictions of her favourite characters) to other fans on the site. The final tag used in this post is an emoji ("#0:)"), which represents a smiley face with a halo above it, and lends a playful emphasis to Participant B's affective tags.

Archive of Our Own (AO3)

As with Tumblr, AO3 implements a combined self-tagging and automanual system. The site suggests pre-defined tags when filling them in, although taggers are free to choose whatever terms they wish. Tags do not take a hashtag format, and there are no restrictions on spaces, length or characters. Non-Roman script is also allowed.

Figure 2 shows a random search result of works archived under the 'Remy LeBeau/Rogue' tag. The tags are displayed after the archive warning (in this case, the author chose not to employ any archive warnings, e.g. violence, non-consensual sex, etc.).

[Take in Figure 2]

Greyed out tags are **Ship** tags (i.e. denoting a romantic pairing of two characters), or **Friendship** tags. **Character** tags follow, and then descriptive tags. Normally, pre-defined tags start with a capital (e.g. "Mental Health Issues", "True Love", "School"). There are several tags that the author herself has applied freely to the fanfic. These express themes ("normal people suck", "cute and crazy") and story elements ("AU jumping from the aftermath of the battle with Apocolypse", "Logan is a good daddy"). These tags are typical examples of how users choose to tag their works on AO3, i.e. using a mixture of pre-defined tags and free ones. If a free tag becomes popular enough, it will be merged by a tag wrangler with a standard, pre-defined one; or a new tag will be entered into the system's taxonomy to accommodate it. For example, in Figure 2, the free **Character** tag, "and many many more" has been made equivalent to the standard tag "Other(s)" – if the tag is clicked on, it will lead

to all works on the site that have used the tag "Other(s)" or their equivalents. Therefore, tag equivalencies, once determined by a tag wrangler, are saved automatically in the site's database.

Etsy

Etsy's tagging mechanism is also a self-tagging and automanual hybrid; a maximum of 13 tags are allowed. Others, such as Smith (2008b) and Warner (2011), have already given excellent accounts of how Etsy's tagging works. The automanual function suggests pre-defined tags for the resource. As Smith describes, "Etsy's pre-defined tags form the top-level category navigation on the website. The suggested tags are actually sub-categories for each of the main categories" (Smith, 2008b, p. 16). These can be seen in at the bottom of Figure 3, where four of the tags ("Craft Supplies & Tools", "Frames, Hoops & Stands", "Picture Frames" and "Frames") are hierarchical tags pre-defined by Etsy. The other tags are free tags entered by the user.

[Take in Figure 3]

Tag analysis

This section details the process for the tag analysis on each of the three online platforms investigated.

Tumblr

Tumblr does not allow crawling of the site; therefore, a Python script, get_tagstats.py, was used to access Tumblr tags, via the Tumblr API, and retrieve data on posts using that tag.

Since the fandom, and hence extent of social media activity, being studied was relatively small, the script's search parameters were set to the last 2000 Tumblr posts that used the Romy tag. The minimum incidence of a co-occurring tag was set to 1, so that all tags with at least one other co-occurring tag would be represented. After data cleaning, the dataset comprised 684 posts by 186 users.

The data was analysed using NodeXL. It was necessary to manually create multiple edges through linking vertices into node pairs, since in NodeXL an edge can only connect two vertices, a source and a target, while in this study one source tag often linked to multiple target tags.

Further normalisation was required to remove case sensitivity, concatenated names, and obvious errors. However, synonyms were not merged, so as to preserve the idiosyncrasies of the Romy folksonomy intact, they were preserved in their entirety. Lastly, duplicate vertex pairs were merged together using NodeXL's merge tool. This resulted in a total 2338 vertex pairs or edges.

The remaining tags were then coded by type. The typology used is shown in Table 1; it is due to Smith (2008a), in turn an amalgamation of the typologies developed by Golder and Huberman (2006) and Marlow, Naaman, boyd et al (2006). While coding, it became clear that the fine granularity of the tags in the dataset could not be adequately described by Smith's very broad outlines. Inductive coding was used to create a more detailed typology, adding new categories in a nested format under Smith's main categories. New categories were added as and when a new tag type was encountered in the data.

[Take in Table 1]

This process of inductive coding continued throughout the social media data analysis. From this iterative process, a final taxonomy was developed by the conclusion of the social media data analysis, which is presented in Table 2.

Two datasets were created from the Tumblr data. One of the top tag users in this set was Participant A, whose account functioned as an aggregator of fanworks related to the character Rogue, unlike the other accounts in the dataset. Therefore, in order to moderate the skewing affect Participant A's tagging might have on the set, a second set was created where Participant A was removed and the data thus normalised. This would enable a comparison to be made between the two datasets, and thus ascertain whether there was any appreciable difference between the original and normalised data.

Archive of Our Own (AO3)

Archive of Our Own (AO3) is a fanfiction archive which was developed and run by fans for fans. It allows users to organise and categorise their work using tags. However, in a system dubbed a 'curated folksonomy' by Bullard (2014), volunteers called 'tag wranglers' filter these tags by associating them with established synonyms. This flexible system allows for both individual idiosyncrasies in user tagging behaviour to remain intact, while also enabling efficient search retrieval. It does however require considerable effort on the part of the volunteers to combine user tags with established synonyms.

AO3 tags are based on their own 'home page' that has its own URL, and can only be retrieved via a static URL address. On AO3, 'Romy' is equivalent to the standardised tag 'Remy LeBeau/Rogue' (http://archiveofourown.org/tags/Remy%20LeBeau*s*Rogue/). Figure 4 shows the 'Remy LeBeau/Rogue' tag homepage.

[Take in Figure 4]

A crawl of the Romy tag was made using SocSciBot (Thelwall 2004), and the results imported into NodeXL, giving a total of 8182 individual tags, with a total of 4368 tag names. A separate dataset was created by merging synonyms according to their standardised 'tag wrangled' version.o Thus all incidences of 'Romy' were merged with 'Remy LeBeau/Rogue', 'ultimate x-men' was merged with 'x-men (ultimateverse)', 'logan – oc' was merged with 'logan (x-men)/original character', and so on. Merging synonyms left all tags that were not popular enough to have been 'tag wrangled' yet. These comprised the 'long tail' of the dataset. The wrangled dataset came to a total of 4946 individual tags, with a total of 2752 tag names. This indicates that 63% of tags that co-occur with 'Romy' have been wrangled by tag wranglers. The two datasets, that comprising pre-wrangled tags, and that comprising wrangled tags, were compared, by manually coded each tag to a tag type, using the same iterative, inductive process as used with the Tumblr dataset. Five new tag subtypes (types 1.7-1.11 in Table 2) were added during this stage of the process.

Etsy

Etsy is an online marketplace that caters specifically for handmade and vintage goods; most of its users are individual sellers or small businesses. It is also home to a growing number of fandom-related stores that offer handmade items created for fans. These include postcards, prints, badges and jewellery featuring favoured characters, or even items of clothing for cosplay. Items can be tagged by the seller to enable more efficient search retrieval; at the time of the study, a maximum of 13 tags could be used per item. As with AO3, the 'Romy' tag is not commonly used on Etsy. The

synonyms ‘Rogue and Gambit’ or ‘Gambit and Rogue’ are more often used by Etsy sellers, and so the tag ‘Rogue and Gambit’ was the base tag for Etsy tag analysis.

As with Tumblr, Etsy does not allow crawling of the site. Tag data was gathered using EtsyRank, a free service developed using the Etsy API. This service includes a Keyword Tool, which allows any Etsy tag to be searched; the tool will retrieve all posts that use the tag, along with other details such as item views, likes, shop or seller name, and also co-occurring tags.

EtsyRank was able to harvest data on all currently active postings that used the ‘Rogue and Gambit’ tag. This returned a total of 173 individual posts. Since this dataset was comparatively small, the data was manually imported into an Excel spreadsheet, which was then brought into NodeXL; each tag was then assigned to a tag pair, as outlined above. After this process, the total number of individual tags came to 4603, with only a total of 438 tag names.

Interviews

To supplement the results of the social media data analysis, semi-structured interviews were conducted. Six participants were interviewed, two from each of the platforms investigated in the social media data analysis. To facilitate anonymity, these participants are coded A-F, and are identified thus throughout this paper.

In keeping with the Delphi study that was used in the authors’ previous study into fan information behaviour (Price & Robinson, 2017), it was decided that participants for these interviews should be ‘experts’ or highly influential users of their respective platforms. Whilst this can skew results (Liu, Morstatter, Tang et al (2016), give an excellent account of social media mining bias), this method was chosen as these participants would likely have far greater insight into their tagging strategies and information behaviour. The use of expert participants in LIS studies is not unprecedented. Recent research includes Lund (2020), who gives a thorough overview of the use of experts as Delphi participants within LIS; Howard (2018) discusses educational requirements Australian information professionals need to operate across “blurred cultural heritage boundaries”; and Andresen, Huvila & Stokstad (2020) made a case study of information professionals on their attitudes towards the crowdsourcing activities of their users. Of course, the experts in this study are not experts in a professional and academic sense, but they are highly conversant in the usage of their respective platforms, which gives them valuable insight into how those platforms work.

Tumblr interviewees were selected from the most popular users of the 'Romy' tag, according to the number of 'likes' and 'reblogs', and for Etsy from the most popular sellers of items using the ‘Rogue’ and ‘Gambit’ tags were ascertained according to the number of ‘hearts’ their items had received. These participants would be likely to have greater post (and thus tag) visibility, be more conversant in their tagging practices, and have a better understanding of how their tagging affected both their own information behaviour and that of their audiences.

AO3 interviewees were recruited from the tag wranglers, who would be expected to have the best knowledge of the tagging system.

All interviews were conducted via email. Participants' responses were coded thematically in NVivo. When responses highlighted concepts that required further investigation, a follow-up email was sent with further questions. Once replies were received, they were also coded in NVivo.

[Take in Table 2]

Analysis

The final taxonomy resulting from the inductive coding and tag analysis is shown in Table 2. While most tag types speak for themselves, it is worth further clarifying the **Opinion** and **Play & Performance** types. The **Opinion** type refers to any tags that are the author's performative judgement of their own work; or, in the case of Tumblr (where tags may be implemented by users on the works of others) users' judgements of other peoples' works. **Play & Performance** refers to tags that are used on posts that are part of fannish events, such as contests, competitions, themed months, celebrations, and other (semi-)formal fan performance.

The data was visualised as co-occurrence graphs, with tags clustered according to tag type, and laid out using the Fruchterman-Reingold algorithm, which attempts to produce a layout that clusters tightly connected nodes together (Fruchterman and Reingold, 1991, Smith et al 2009). Compared to alternative algorithms investigated, Fruchterman-Reingold was the most visually coherent and most clearly presented nodes and the relationship between clusters. Edge opacity is denoted by the edge weight (how many times a tag pair co-occurs); tag type by colour and vertex shape; vertex size by betweenness centrality (how often a node appears on the *shortest path* between two other nodes). The higher the betweenness centrality, the higher the importance of that node in exchanging a resource within the network. This in turn indicates how effective that node is as an information or content carrier. In this case, each node represents a tag or keyword, and thus "[t]he larger the number of betweenness centrality, the higher the influence of one specific keyword […and h]igher betweenness centrality suggesting more importance of one keyword plays the role of a "connector" to keywords in different topics […] therefore indicating an important role in the transmission of data in the network" (Cheng et. al, 2018, p. 642). The importance of betweenness centrality in this transmission of data is also highlighted by Zhang, Luo & Boncella (2020), where it "essentially reveals how important each node is in providing a "bridge" between different parts of the network" (p. 15).

Bar charts (via Excel) also visualised the overall usage of tag (sub-)types. These can be seen throughout this section.

Tumblr

In total there were 624 named tags that co-occurred with 'Romy' in the Tumblr dataset. After merging tags, there was a total of 535 named tags. The final individual tag count was 4582. The entire network is laid out in Figure 5. Figure 6 depicts a similar network, but only those tags with a betweenness centrality of 1 or above: this indicates tags that are on a pathway between two other tags and are therefore more significant carriers of information content. Fifty-six (10.5%) tags reached this value – most of these tags were represented by the **Descriptive** tag type, specifically **Ship**, **Character**, **Fandom** and **Fanwork** sub-types, thus highlighting the importance that these types have in mediating information within the Romy fandom. As can be seen in Figure 6, the tags that have the highest betweenness centrality are those that have the largest nodes. Since 'Romy' is the base tag, it is not surprising that it has the highest betweenness centrality, and is therefore the highest carrier of information content between tags; this was followed by **Character** type tags – 'Rogue' and 'Anna Marie' (synonymous), and 'Gambit' and 'Remy LeBeau' (synonymous); and **Resource** type tags – 'fanart', 'fanfic' and 'fanfiction' (synonymous), and 'Romy fanfic' also had high betweenness centrality.

[Take in Figure 5]

[Take in Figure 6]

Figure 7, which shows the total tag count arranged by tag type, demonstrates that the bulk of all Tumblr tags used were classified as **Descriptive** tags. When arranged by sub-type, Figure 8 shows

that the **Character** and **Ship** sub-types (both **Descriptive** sub-types) were the most popular respectively. This was followed by the **Creator/Source** sub-type (an **Ownership** sub-type), which is used to denote the author of a fanwork. This type was heavily used by Participant A's account, which, as an aggregator of Rogue-related fanworks, concentrated on reblogging other peoples' work, and used the tag function to indicate the original author of the post.

Participant A's blog was the top user of the 'Romy' tag on Tumblr, but this account - a community account, aggregating fanworks and serving as a hub for the Rogue fandom on Tumblr - was very different from the other, personal, accounts in the dataset. A second, normalised, dataset from the Tumblr results, with A's data removed, was therefore created, to mitigate any skewing effects.

[Take in Figure 7]

[Take in Figure 8]

Figure 9 shows the results of this normalisation. Contrary to expectations, there was little overall difference to the original dataset., showing that A's tag usage did not inordinately skew the results of the dataset.

Figure 10 shows the same comparison, this time broken down by tag sub-type. Again, there was little significant difference between the normalised and pre-normalised datasets. The most notable difference was in the **Character** sub-type. Usage of this tag sub-type was markedly lower after normalisation (23% versus 16% respectively). This suggests that A's account uses the **Character** sub-type far more while tagging their posts. Possibly this is because A's is a Rogue community account, rather than a Romy community account, and therefore focuses on other romantic pairings involving Rogue, e.g. Rogneto (Rogue/Magneto) and Rogan (Rogue/Logan), so that naming characters becomes more important in classifying the different ships.

Figure 11 shows a comparison of the tag usage of participants A and B, which is very similar, except in two respects. Participant A has a higher occurrence of tags in the **Ownership** type, specifically the **Creator/Source** sub-type. As an aggregator of others' posts, A is careful to cite or credit the source of the original work as well as add relevant descriptors, as mentioned in their follow-up interview:

> I'm thinking of the audience I have in mind when I tag things, so I put in character names, relationship names, and artists/creator.

[Take in Figure 9]

[Take in Figure 10]

In comparison Participant B has a higher occurrence of **Opinion** type tags, specifically in the **Communication** sub-type. These are idiosyncratic tags, sometimes formatted as sentences over multiple tags, which the user may employ to communicate thoughts, feelings and emotions. Some examples of these tag-(sub)types from Participant B are:

> its been 6 months since Ive actually drawn a full picture];
>
> Rogue and Gambit have gotten me thru a lot of dark times;
>
> Rawrrr comics are supposed to make a person happy Damnit;
>
> READ THIS NOW;

> i have no idea how to tag people on tumblr dear;
>
> :)

These examples express a wide range of affective and communicative impulses, and indeed B was the heaviest user of the **Opinion** and **Communication** (sub)-type tags in the dataset:

> More often than not I use tags to voice personal opinions so I don't ruin the aesthetic of the original post.

Participant B also mentions other forms of communication via Tumblr tagging during reblogging, such as "personal opinions or replies to the tags in the original post. My personal favorite thing is when followers say nice or supportive things in the tags when they repost my stuff."

Analysis of the Tumblr tag usage shows that, in the Romy fandom, descriptive tagging is of importance, with heavy usage of the **Character** and **Ship** tag sub-types. This indicates that fan identity is paramount, i.e. signalling to other fans the basic classification of ones' own fan identity – in other words, *this is the ship I support, these are the characters I support*. Content description is a form of communication, a way of 'speaking' to other fans, of making sure that an author's post reaches the right audience – a fellow fan with the same basic fan identity. This kind of phenomenon has been seen in contexts other than fandom – for example, Wargo (2017) discusses how the #donttagyourhate tag on Tumblr is a way to signal (or curate, as Wargo puts it) one's own identity as a #socialjusticewarrior. In a similar vein, tags such as 'Romy' can be used to curate one's own fan identity, as well as the content of one's post or blog.

[Take in Figure 11]

Also of significance is the importance placed on ownership. While crediting sources might be considered basic netiquette, Tumblr Romy fans, particularly participant A, made a point of naming sources. This is especially important on Tumblr, as one of the site's main functions is to easily 'reblog' other people's posts, where adding an original **Creator/Source** tag would be appropriate.

In some cases, the tags were used in a manner which appears to be unique to Tumblr. Multiple hashtags could be employed by a user to express thoughts or emotions that were especially complex. This phenomenon might be considered similar to footnotes, meta-commentary, or indeed, paratext, qualifying or framing the main content of the post. These tag types are however more personal in nature, and not all users chose to employ them. Accounts like A's, whose function in the Tumblr Romy community is as an aggregator, used **Communication** or **Opinion** type tags much less frequently.

Archive of Our Own (AO3)

The 'wrangled' dataset showed higher levels of homogeneity and much lower tag name counts than the 'pre-wrangled' set: 4946 individual tags as against 8182, and 2752 tag names as against 4638. 63% of the tags in the set had been processed by tag-wranglers. In contrast to the Tumblr findings, the Romy fandom on AO3 exhibited a much wider scope of co-occurrence with other tags used on the site. While on Tumblr the 'Romy' tag seemed to co-occur mostly with other tags related to the Romy and X-Men fandom, on AO3 its use was distributed amongst other fandoms, mostly from the wider Marvel Universe (e.g. the Avengers), but also amongst non-Marvel fandoms such as Harry

Potter. This suggests a much higher degree of connectedness and crossover between fandoms on the AO3 platform.

Figure 12 shows a directed co-occurrence graph for the 'Romy' tag on AO3. Figure 13 depicts all tags with a betweenness centrality of 1 or above (that is, tags that appear to be more efficient carriers of information content within the network). Three hundred and nineteen (11.6%) tags reached this value, which was similar to that found on Tumblr (10.5%). Again, and in common with Tumblr, most of these tags were represented by the **Descriptive** tag type, specifically **Ship**, **Character** and **Fandom** sub-types – these figures were similar in both 'pre-wrangled' and 'wrangled' datasets.

The similarity between the results of both the 'pre-wrangled' and 'wrangled' datasets is notable, as it suggests no evidence for gatekeeping or bibliographical control. Figures 14 and 15 compare the number of tag names in both pre-wrangled and wrangled datasets – they show very similar tag type and sub-type usage patterns, despite the standardisation of the wrangled set.

[Take in Figure 12]

[Take in Figure 13]

This suggests that tag wrangling is not a form of gatekeeping of the vernacular – or indeed, the taxonomy – used in the Romy or wider Marvel fandoms, nor of the bibliographical control of fanworks. This is borne out by the interviews with the tag wranglers, who confirmed that they do not see themselves as gatekeepers, and that they try to follow the original tagger's meaning and intent as closely as possible. Interview statements show considerable expertise in their chosen area, and tag wranglers pride themselves on being able to recognise the obscure references in certain tags, preserve them in the wrangling process, and standardise them if warranted:

> What I do see myself as providing is a chance to make too many years reading a lot of comic books useful. Marvel has a very, shall we say, dense, history. But if you think there aren't users out there who will tag for characters who appeared in one issue of Fantastic Four back in 1973, I want to assure you: you are wrong (Participant D).

[Take in Figure 14]

[Take in Figure 15]

This knowledge capital could equate to some prestige in the fan community, but tag wranglers do not benefit from this, as they are unseen and anonymous.

Unlike Tumblr, **Ownership** type tags were seldom employed as ownership was inherent in the post itself (i.e. the author of the story is the poster of the content). The **Descriptive** type was highest, particularly the **Character** sub-type. **Communication** was the second highest tag type employed, and surprisingly this type showed a marked percentage rise post-wrangling. This may be because **Communication** (sub)-type tags are idiosyncratic, and unlikely to be used more than once. Since their usage is so low, this means that they are rarely tag wrangled and merged with synonymous tags. Therefore, their percentage of the total tag count tends to remain static, while other tag types, particularly **Descriptive** and **Resource** tags, tend to be readily merged with already-existing synonyms.

As with **Ownership**, there was negligible use of **Self-reference** or **Task** organising tags (less than 1%) in both datasets. There was some slight use of the **Play and Performance** tag type, used in very

specific instances, e.g. stories written as part of events, competitions, contests, challenges or games, often hosted on other social media sites. Examples are: 'i accepted a few prompts', 'community: xmen15', 'secret mutant ficathon 2014' and 'x-men big bang challenge'.

Figures 16 and 17 show a comparison of the total tag count used in the 'pre-wrangled' and 'wrangled' datasets. Figure 16 demonstrates that on the level of tag type, there is little difference between the two sets. Figure 17, however, shows some significant disparities at the level of tag sub-type. These are at the following sub-types: **Ship**, **Character**, **Friendship**, **Citation** and **Explanatory Communication**.

[Take in Figure 16]

This is where the 'long tail' of tags, not popular enough to have been wrangled, manifests itself most clearly. Some obscure tags are 'shoehorned' into a standardised tag that did not encapsulate the specificity of its original meaning. For example, in the **Citation** sub-type, the tag 'Star-Lord and Kitty Pryde', which is the title of a comic series, and of low popularity, has been tag-wrangled into a synonym of the much broader Fandom sub-type, 'Marvel'. Likewise, the **Ship** tag 'loroki' (denoting the Loki/Storm pairing) has also been made a synonym of 'Marvel'. Not only has the precise meaning of the original tags has been lost, but both have also been classified under an inaccurate sub-type. While examples of this are not common, they are still frequent enough to suggest that, at the tag sub-type level, tag wrangling is slightly less successful than it is at the tag type level. Such funnelling of less popular tags into inaccurate tag sub-types might account for the unusual spikes in the **Friendship**, **Citation** and **Explanatory** sub-types in the 'wrangled' dataset, and in the **Character** sub-type in the 'pre-wrangled' dataset.

[Take in Figure 17]

Etsy

In this small data set with 438 named and 4603 individual tags, merging tags was not required. While this tag count was similar to Tumblr's, the number of named tags used was lower, suggesting that Etsy has a smaller core taxonomy for the Romy fandom than Tumblr.

Figure 18 depicts a directed co-occurrence graph for the 'Gambit and Rogue' tag on Etsy. Figure 19 depicts all tags with a betweenness centrality value of 1 or above, i.e. those most effective as information carriers. Most are the **Descriptive** tag type, specifically the **Character** sub-type. Other tags with high usage in the **Descriptive** and **Resourc**e types were generic rather than fandom-specific, e.g. 'blue', 'drawing' and 'vintage'. Resource tags such as 'comic books', 'illustration' and 'art' were also popular. These generic type tags saw much less use on Tumblr and AO3. This suggests that Etsy taggers assign more importance to the nature of the item being sold, to attract appropriate buyers.

[Take in Figure 18]

[Take in Figure 19]

The **Ship** tag sub-type, so popular on Tumblr and AO3, makes little appearance on Etsy, being reflected only in the 'Gambit and Rogue' tag. The **Ship** element of this dataset co-occurs with other more generic terms such as 'couples', 'love' and 'wedding'. This suggests that **Ship** sub-type tags might be used to sell fanworks targeted at couples.

Figure 20 shows how tag usage is heavily skewed to **Descriptive** and **Resource** type tags, although **Resource** type tags are used at a far lower frequency on Tumblr and AO3, again suggesting that Etsy

tags are used to describe the material attributes of individual items on sale. Olbrich and Holsing (2011) have shown that tag usage on e-marketplaces increases buyer click-throughs. This would account for the mundanity of many keywords in the Etsy dataset which describe formal characteristics, such as item type ("fabric blocks", "postcard"), material used ("pencil", "vinyl"), colour ("brown", "black and white") and size ("3 5", indicating 3.5 inches).

[Take in Figure 20]

[Take in Figure 21]

Figure 21 shows the lack of granularity seen in the Etsy tags, with very little use of tag sub-types, apart from **Character** and **Creator/Source**. The latter was utilised on items such as art prints and individual comic issues, where potential buyers might be looking for the work of certain creators, but, unlike Tumblr and AO3, does not appear to fulfil the function of acknowledging or crediting the original creator per se.

23 out of the 174 posts were assigned tags that did not refer to the item on offer, but described characters or fandoms peripherally related to the actual characters or fandoms represented in the item; Figure 22 shows an example where only two tags accurately describe the item – 'sticker' and 'costume'. None of the other tags were relevant, and may be an attempt to draw traffic by adding tags irrelevant to the product, but relevant to a popular film just released at the time.

The complete lack of **Communication** and **Play and Performance** tag types in the Etsy dataset further serves to validate the idea that tagging on Etsy is not intended for social or community purposes, as might be seen on Tumblr and AO3. Zollers (2007), in a study of tagging on Amazon, noted that:

> The consumption centric system creates a different dynamic between the users and the system [compared to social media sites], and perhaps elicits new motivations for tagging since the users are no longer tagging the content that they produce, but rather predefined content already available on the site (n.p.).

[Take in Figure 22]

Etsy, while also an online marketplace, implements a different tagging system in that sellers may tag their own work. It is also a 'consumption-centric' tagging system, and the taggers, while fans on the one hand, are also concerned with selling their wares. Their tagging practices are motivated not only by their fan identities, but also by their primary purpose in being on the Etsy site, which is to sell their fanworks.

This is borne out by the interviews with two Etsy sellers. For example, referring to her tagging practices, Participant E (who sells cross stitch patterns of characters from many fandoms) said:

> I use [tags] with every listing and try to have some which are broad – like "cross stitch" and "geeky" – and some which are more specific – like my shop name and the names of the characters in the pattern – in hopes of capturing the attention of as many customers as possible.

While her tagging appears to be sales-driven, E nevertheless considers herself a fan:

> I've always considered our patterns as fanart – a way for us to celebrate the movies, shows, books, and comics that we love so much. We're working in a medium which certainly isn't traditional in the fanart community, but that doesn't make it any less relevant.

and:

> Yes, I am a fangirl […] Being a fan has certainly influenced the business from both directions though – it's not just our own passions that dictate what we design, but those of our customers too – we've had so many requests for shows and movies and comics that we ourselves weren't initially familiar with, that we just had to get to know those franchises too.

This suggests that a clear dichotomy is present in these sellers' tagging motivations – the languages of business and fandom are *both* present. Moreover, there does not appear to be any conflict between their fan and business identities. Both identities seem to be equally strong. Participant E even embraces the participatory aspects of fandom while selling items for her business – she takes requests (a notable characteristic of fandom's gift economy; see Jones, 2014), and even joins new fandoms in order to learn enough about them to do her products justice.

Participant F, who owns a store that sells corsets patterned with fan motifs, showed a similar approach to tagging her products: "I try to think about what is most culturally relevant to my product and market on what I think my buyers are looking for". As with E, she felt that being a fan influenced her identity as a seller:

> I feel like being a fan of the properties that I'm making inspired pieces by helps me to know what another fan would want to buy. It provides more passion and inspiration of the works if I'm a part of the fandoms like I'm pulling from.

In these cases, fan identity is not abandoned for business identity. These sellers see themselves as fans providing fanworks and collectibles to other fans, and pride themselves on knowing their audiences and what fans want to see, because they are fans themselves. This does seem to influence tagging behaviours, since sellers who are fans share enough 'subcultural literacy' (Zollers, 2007) with their intended buyers to effectively target fan-tags at them. Yet simultaneously much of their tagging behaviour is also commercially-driven, the tagging strategies employed being influenced by the need to promote the consumption of products.

Discussion

When comparing the three datasets, several points of similarity are seen. On all three sites, **Descriptive** tag types are the most popular, reiterating the primary function of tags as descriptors of a resource's content. Figure 23 shows this, the **Descriptive** type having the highest usage in all sites and datasets. Of the **Descriptive** type, the **Character** sub-type was the most popular; see Figure 24), presumably as character is the main criterion in determining the interest of the audience in a resource.

Likewise, all three platforms showed lowest usage in the **Self-Reference**, **Task Organising** and **Play & Performance** types. Of all the tag types, these three also present least granularity, implying that, in

the Romy fandom at least, these types serve more specialised, and comparatively little-used, functions.

[Take in Figure 23]

[Take in Figure 24]

There were, however, significant differences between the three sites studied. These are detailed below.

Tumblr

Tumblr shows the highest incidence of **Ownership** and **Opinion** tag types. While Participant A, as an aggregator of Rogue-related fanworks, used the **Ownership** tag type particularly heavily to credit sources, there was heavy use of this type by other users, suggesting a strong culture of crediting other peoples' work. Indeed, A mentioned in their interview that,

> [s]ome people are inconsiderate by not posting the creator's name when it comes to fan works, but over time, most posters learn that it's a Tumblr faux-pas.

Participant B noted:

> I know of one poster who makes a point of sharing the art/post directly from the original poster (as a sign of respect). As far as a lack of respect, I've also frequently seen works reposted with no link back to the original artist/writer. There are two people I follow who become incensed whenever this happens to their own work.

Since Tumblr thrives on the reposting of the works of others, a culture has grown within many of its communities that makes it best practice to credit the source of the original post. AO3 and Etsy have no reposting function; the post's author is an inherently visible attribute of that post, and therefore there is no need for **Ownership/Source** type tags to be applied on these platforms.

The use of the **Communication** tag sub-type, rarely used on other platforms, highlights Tumblr as a site for the novel use of tagging, that is, as a form of creative expression in itself, as a form of engagement with other users in its own right, or as a means of communicating with other fans. Only interviewees from Tumblr talked explicitly about using this type of tag:

> More often than not I use the tags to voice personal opinions so I don't ruin the aesthetic of the original post (Participant B).

> I occasionally make a smart-ass remark in the tags, instead of having as [sic] permanent commentary (Participant A).

As an aggregator of Romy posts, Participant A felt that her purpose was not to write commentary on the resource, but merely to describe it or credit its creator; only sparingly did she pass a "smart-ass remark". However, her account showed good evidence of the **Conversational & Enunciative** tag sub-type (meaning that the tag flagged a resource as a two-way dialogue between two users). Part of the account's function is to run a question-and-answer service – anyone can ask a question about the

character Rogue, and A will answer it, tagging it with '#questionsanswered'. An example of this type of post is in Figure 25. Other questions may be encyclopaedic in nature, and in such cases A will use this format to share her extensive knowledge of Rogue and her backstory with other fans.

Tumblr's fan-tagging is notable for the discursive and playful ways in which tags are used. Monnin (2009) highlighted the ludic aspects of tagging, using Flickr as an example of how Web 2.0's technological affordances have blurred the line between work and play. Sites like Flickr are spaces where "a user's browsing trails are leveraged to generate a singular gaming experience by transforming any casual task, any toilsome effort – or rather the data they leave behind, into playful actions" (Monnin, 2009, p. 2). Monnin calls Flickr and other similar sites that use collaborative tagging "ludic computer-human interfaces… transforming senseless tasks… into entertaining and enjoyable meaningful actions while operating at a completely different level to perform a desired task" (p.3).

[Take in Figure 25]

This is the sort of phenomenon witnessed on sites such as Tumblr and AO3, where fan-tagging is consistently used in playful and creative ways. As well as being descriptive, tags can be expressive and dialogic, or symbolic and visual, reminding us, as Monnin says, that "tags are not, and never were, mere words" (p. 4).

Both interviewees believed the tagging system could be improved; most of their objections centred on problems with searching and homonymous terms, as there is currently no function to filter tags:

> Take the tag "Rogue," for instance. While I may be looking for an X-Men character, many other things come up: an anime character, crossfit gear, roleplaying character class, and non-English versions of Harry Potter […] The aforementioned issues when searching through common terms could be aided by having a way to exclude other words. There have also been issues with spam, where completely incorrect tags were applied by spambots and they brought up disturbing results. Manually blocking all of those spambots was irritating (Participant A).
>
> I would like to be able to see the collected tags from all posters in the same place instead of clicking on each repost and reading their tags. (Participant B)

These responses give some idea of the trade-off Tumblr users must pay for the free use of tags on the site. While users can use tags in uninhibited and creative ways, the lack of formal, bibliographic and taxonomic control makes information retrieval challenging. There is no way to filter results; there is a great deal of redundancy (reblogs of a single post can be reduplicated in a search), and some tags are not even indexed by the system. This contrasts with both AO3 and Etsy, where various controls have been put into place.

Archive of Our Own (AO3)

AO3 had the highest tag usage of the three platforms, taking nearly half (47%) of the total combined individual tag count. AO3 tagging practices show very high density and granularity, as authors attempt to convey the details of their fandom, as well as the plots of their stories. Fans are known to be particular about the types of fanworks they will engage with (Driscoll, 2006), showing preference according to characters, ships, genres and 'kinks' (i.e. the sexual predilections depicted in fanfic), and these elements are important both for audience and creator. Hence, tagging on AO3 in particular, becomes an important finding aid.

While there is the unique practice of 'tag wrangling' on AO3, this did not seem to affect the overall meaning or sense of the original tags used. Tag wranglers in Marvel fandom appeared to have an in-depth expertise in their area, and, judging by the marked similarity between the pre- and post-wrangled AO3 datasets, they were for the most part correctly able to interpret and maintain the sense of the original tags. Indeed, the tag wranglers appeared to be very dedicated to this mission, and to preserving the authenticity of the original tags, as a form of fanwork in itself:

> The Ao3 Terms and Conditions and the Wrangling First Principles both strictly prevent us from being gatekeepery. We can't change tags, we can't tell users how to tag in any official capacity ("describe not proscribe"). Our goal is to organize tags in a way that fans will be able to find what they're looking for. To do that, we have to speak their language and use the words they use. (Participant C)

> One of the most important principles of tag wrangling is that we don't alter a user's tags. The beauty of the AO3's system is that everyone can tag for whatever they want, in exactly the format they want. As well, most large fandoms have multiple wranglers assigned to them, and that means that there has to be a general consensus on how to handle any given tag that is for some reason challenging, or requires a judgement call of some kind. (Participant D)

> Tag wrangling is a way I can contribute to a community that I love. I like this kind of work and, with the decline of livejournal, I felt less connected to the community and less like I was pulling my own weight. Wrangling both lets me meet people from across fandom and help out (Participant C).

> I consume a great many fanworks in my day to day life, but I don't really create that many. Tag wrangling is a way that I can feel as though I give something back to the community that has brought me so much joy (Participant D).

Both interviewees rejected the idea that they were gatekeepers within their fandom. Participant C even went so far as to say "I don't think of myself as a gatekeeper, mostly because I hate that word". Despite this, it seems that tag wranglers are information gatekeepers in the sense that they are "shaping, emphasizing, or withholding" information, or the flow of information (Case, 2012, p. 339). Rather than actively or intentionally withholding or emphasizing certain aspects of information within their fan community, they are shaping it by streamlining its flow and facilitating greater access. Many more co-occurring tags in this dataset had a higher betweenness centrality, and hence a greater role in information exchange, when compared to the other sites. AO3's tag wranglers, and by extension Bullard's (2014) 'curated folksonomy', are effective in mitigating the unpredictable effects of online tagging.

While users of Tumblr felt that there was improvement to be made on the site in terms of information retrieval, the AO3 tag wranglers interviewed had positive views on the tagging system. Participant C felt that the filtering system could be better streamlined in order to increase retrieval accuracy. Participant D was largely satisfied with the system, but felt the "most changes that could be proposed would have more to do with changes in policy":

> For example, there are an unfortunate number of tags floating about that can't be wrangled because users entered them in the wrong field, but if you put "Tony Stark" in the Fandom field, we can't make it a synonym of Tony Stark the character tag. Changing the type of a given tag is changing what a user entered in a way that we don't do as a matter of policy, and it's a policy I have to agree with.

This is indicative of the trade-off between 'messiness' and control that is so often seen in folksonomic systems (Smith, 2008a). While AO3 suffers in some ways from maintaining this balancing act, it seems to be managing this successfully.

Etsy

Etsy tagging appears closer to that of online marketplaces, rather than social information sharing sites. For example, Amazon shows minimal tagging activity in comparison to sites such as LibraryThing, because "the tagging feature is not given much prominence […]. Additionally, the main purpose of the site is commercial and not organizational, so users might not be as motivated to tag content" (Zollers, 2007, n.p.). Etsy's percentage of the tag count across the platforms studied was similar to Tumblr, at 27% and 26% respectively, rather than AO3's 47%. This indicates that despite both the low prominence of tags on Etsy, at the bottom of a post's page, and the limit to the number of tags per resource, users still make as much use of tags as possible. One of the interviewees thought that a higher cap on tags was the only improvement that Etsy could make:

> It would be nice to have a few more tags - sometimes 13 just doesn't seem enough. 20 tags per listing would be wonderful! (Participant E)

While Etsy, like Tumblr and AO3, shows heavy usage of the **Description** tag type, there is less evidence of fandom-related sub-types. Instead there is a heavier reliance on generic **Description** and **Resource** tags. Of fandom-related sub-types the most-used is **Character**. There is some use of **Ownership** tag types, particularly in references to artists/creators, but also in indicating the Etsy store name itself. Other tag type usage is negligible.

Overall, tag usage on Etsy points towards both functional and commercially-driven use. While fan-tagging does exist, its primary function seems to be to draw in fans as buyers. This tagging is always supplemented by generic tags describing the characteristics of a resource, and there is significant use of tags that do not describe the fandom in the resource, but fandoms peripheral to it. This could be a strategy to maximise the potential audience and/or buyer pool: as Blanchflower and Hodges (2014) remind us, "strategically tagging an item and giving it a clear title, along with providing clear product images, are the factors that are required for a seller's shop to succeed" (p. 819).

Despite this, Participant E felt a connection to the fan audience:

> […] it's not just our own passions that dictate what we design, but those of our customers too - we've had so many requests for shows and movies and comics that we ourselves weren't initially familiar with, that we just had to get to know those franchises too. Without customers urging us to check out shows like Supernatural and Sherlock, we never would have designed those patterns and become fans of those shows too.

This suggests a kind of symbiosis between the seller and buyer, and that the gift economy that is seen in many fan communities has been transferred in part to the online marketplace. This hybridisation should not surprise us, since fan culture itself is driven by commercial culture (the need

to consume and collect official fan goods and artefacts) as it is by transformative culture (the need to create and share fanworks based on those goods). As Jones (2014) opines:

> […] it is not simply the binary of commodity culture and gift economy that work with (or against) each other […F]an culture itself is influenced by two opposing sets of values that dominate the cultural field and that fans take positions in line with: cultural value based on the profit motive, and cultural production for its own sake—that is, the gift economy (2.6).

These two opposing forces were not a personal source of tension for the two interviewees. This is supported by Cherry's (2016) study of fan fabric artists who also sell their work:

> As an independent dyer, [the seller] can incorporate her fan interests into her business, but as a fan she can "geek out" with other fans […] The sense in which the two sides are both part of a shared fan community […] can lead to the dyer and the customers feeling as though they share the bond of the fan community rather that it straightforwardly being a commercial transaction (p. 171).

and:

> Such shared love of the text between the dyer and her customers is indicative of [a] hybrid market environment in which there is no clear distinction between social and commercial economies […] Furthermore, the social network influences both production and consumption within the fan-knitting community (p. 171).

The interviewees showed an ambivalent attitude towards copyright and licensing, despite having an awareness of copyright issues:

> There's actually a little known loophole when it comes to using licensed fabrics. Since you've already paid the licensing fee by purchasing the fabric, the company has already received a licensing fee for that product if it's a pre-printed fabric. There has actually been a lawsuit over this and the licensor lost, as they'd already received revenue from the product. I don't recall the name of the case, though, but it caused a splash in our community. (Participant F)

Cherry (2016, p. 176) cites two similar cases, one where sellers on Etsy and eBay received cease and desist notices from the BBC after selling knitted Doctor Who adipose dolls, and the other from 20th Century Fox to knitters selling replicas of Jayne Cobb's hat from Firefly.

Despite this, E did not show much concern for copyright, saying "I've always considered our patterns as fanart", as if proclaiming her work as fanart automatically precludes any accusation of copyright infringement. This is an erroneous assumption (Johnson, 2016), but it is one that many producers of fanworks share, led perhaps by the notion that they are part of a moral economy, respecting the original creator while assuming a right to appropriation (Postigo, 2008), a perception that "their endeavours [are] important social work" (Lantagne, 2015, p. 302), a sense of 'safety in numbers', or "a lack of knowledge" or "confusion about copyright law" (Fiesler, Feuston and Bruckman, 2015):

> In my experience I don't think fans worry too much about the copyright laws. They're definitely more concerned that the fanworks are true to the characters and the franchise. If the fanwork is honest and a fair representation of the story/characters that's really what's important. (Participant E)

> I don't think fans care, because the fans want to see more products of their fandom out there. And licensing is hard to come by and so expensive that your average creator and crafters on Etsy don't have the ability to license 100,000 pieces from the intellectual property. Licenses only work when mass-producing, and don't work as well for small runs or custom creations. (Participant F)

Intellectual property, as it currently stands, does not seem to adequately serve small businesses, or indeed fans, or communities that consider themselves a part of the Maker Movement. Perhaps a new paradigm on intellectual property would be beneficial in making licensing and copyright a relevant issue to these communities.

Conclusions

The tag analysis presented in this study, bolstered by interviews with tag users and 'tag wranglers', gives insight into the information behaviour of fans, which has not previously been available. When combined with the literature analysis and Delphi study previously reported, this analysis gives a clear picture of the sophisticated information practices in this form of serious leisure.

Generally, fan-tagging, when used as a classification or organisation tool, exhibits high granularity, which better facilitates findability and sharing. However, it is also used in different ways – as a method of expressing affective and emotional reactions to a resource; to engage in a dialogue with other users; to visually embody affect through emojis; and to give opinions on or explain a resource's content. As can be seen from the examples shown above, particularly on Tumblr and AO3, these motivations can result in rich tagging practices, that evidence the fan community's desire to share and engage widely with one another, as well as to accurately and usefully organise and classify their works. The exception to this is Etsy, where tagging practices take on a far more traditional form. Indeed, tagging on Etsy is far more strategic and commercially-driven, despite (as is clear from the interviews) fan sellers on Etsy emphasising their fan identities, and seeing themselves as providing for, and inspired by, other fans. Additionally, the interviews also suggest that there is an ambivalent view among fans towards copyright and intellectual property issues.

Future studies would be encouraged, to further test the use of tag analysis as a method of investigating information behaviour. This may take place in a longitudinal study, performing the same analysis as was undertaken in this study, in order to observe any changes over time. This would be especially appropriate, since Tumblr's December 2018 adult content ban "cost the platform its user base" (Bronstein, 2020, p. 240), driving many users to other platforms such as Twitter, and which may have subsequently affected user behaviour on the site. Such a longitudinal study may be performed on other fan communities, to see whether there are any differences in observed tagging behaviour between fans. It may also be used on other, non-fandom, communities (e.g. photographers on Instagram), to assess whether different information behaviours can be observed between different types of users, through the prism of their tagging practices. For future study, it would also be useful to test the fan-tag taxonomy in other fandom contexts.

Acknowledgements

This research was supported by a doctoral studentship awarded to Ludi Price from the School of Informatics, City, University London.

A portion of this study was previously published in Price, L. (2019), "Fandom, folksonomies & creativity: the case of the *Archive of Our Own*", in: Haynes, D. & Vernau, J. (Eds.), *The Human Position in an Artificial World: Creativity, Ethics and AI in Knowledge Organization: ISKO UK Sixth Biennial Conference, London, 15-16th July 2019*, pp. 11-37, Ergon Verlag, Baden-Baden.

References

Ahmed, W., and Lugovic, S. (2019), "Social media analytics: analysis and visualisation of news diffusion using NodeXL", *Online Information Review*, Vol. 43 No. 1, pp. 149-160, available at: https://doi.org/10.1108/OIR-03-2018-0093 (accessed 13 November 2020).

Anderson, C. (2019), *Women in online science fiction fandoms: perceived impact on psychological well-being,* Ph.D., Capella University, available at: https://pqdtopen.proquest.com/pubnum/13857793.html (accessed 3 January 2020).

Andresen, H., Huvila, I., and Stokstad, S. (2020), "Perceptions and Implications of User Participation and Engagement in Libraries, Archives and Museums", in Audunson, R. et al. (Eds.), *Libraries, Archives and Museums as Democratic Spaces in a Digital Age*, de Gruyter, Berlin, pp. 185-206.

Attu, R., and Terras, M. (2017), "What people study when they study Tumblr: classifying Tumblr-related academic research", *Journal of Documentation,* Vol. 73 No. 3, pp. 528-554, available at: http://dx.doi.org/10.1108/JD-08-2016-0101 (accessed 21 July 2019).

Batagelj, V., and Mrvar, A. (2014), "Pajek", in Alhajj, R. and Rokne J. (Eds.), *Encyclopedia of Social Network Analysis and Mining*, Springer, New York, pp. 1245-1256.

Blanchflower, T., and Hodges, N. (2014), "Understanding Etsy: social media and marketing within a community of sellers", in: *Ideas in Marketing: Finding the New and Polishing the Old. Proceedings of the 2013 Academy of Marketing Science (AMS) Annual Conference in Monterey, CA, USA, 14-18 May 2013*, pp. 818-821, available at: http://dx.doi.org/10.1007/978-3-319-10951-0_297 (accessed 25 December 2019).

Bourlai, E. E. (2018), "'Comments in Tags, Please!': Tagging practices on Tumblr", *Discourse, Context & Media*, Vol. 22, pp. 46-56, available at: https://doi.org/10.1016/j.dcm.2017.08.003 (accessed 24 November 2018).

Bronstein, C. (2020), "Pornography, trans visibility, and the demise of Tumblr", *Transgender Studies Quarterly*, Vol. 7 No. 2, pp. 240-254, available at: https://doi.org/10.1215/23289252-8143407 (accessed 25 October 2020).

Buarkhi, H., and Alkhateeb, B. (2018), "Use of hashtags to retrieve information on the web", *Electronic Library*, Vol. 36, No. 2, pp. 286-304, available at: https://doi.org/10.1108/EL-01-2017-0011 (accessed 19 October 2020).

Bullard, J. (2014), "Values and negotiation in classification work", in *CSCW Companion '14: Proceedings of the companion publication of the 17th ACM conference on Computer supported cooperative work & social computing in Baltimore, USA, 15-19 February 2014*,

ACM, New York, pp. 45-48, available at: https://doi.org/10.1145/2556420.2556820 (accessed 3 January 2020).

Bury, R. (2016), "Technology, fandom and community in the second media age", *Convergence: The International Journal of Research into New Media Technologies*, Vol. 23 No. 6, pp. 627-642, available at: http://dx.doi.org/10.1177/1354856516648084 (accessed 3 January 2020).

Case, D. O. (2012), 3rd ed. *Looking for information: a survey of research on information seeking, needs, and behaviour.* Emerald Group, Bingley.

Cheng, F., Huang, Y., Yu, H., and Wu, C. (2018), "Mapping knowledge structure by keyword co-occurrence and social network analysis: Evidence from Library Hi Tech between 2006 and 2017", *Library Hi Tech*, Vol. 36 No. 4, pp. 636-650, available at: https://doi.org/10.1108/LHT-01-2018-0004 (accessed 19 October 2020).

Cherry, B. (2016), *Cult media, fandom, and textiles: handicrafting as fan art.* Bloomsbury, Bloomsbury.

Close, S. (2016), "The political economy of creative entrepreneurship on digital platforms: case study of Etsy.com", in *HICSS 2016 49th Hawaii International Conference on System Sciences*, Koloa, HI, USA, 5-8 January, 2016, pp. 1901-1908, available at: https://doi.org/10.1109/HICSS.2016.241 (Accessed 16 November 2020).

Cohen, M., and Sundararajan, A. (2015), "Self-regulation and innovation in the peer-to-peer sharing economy", *The University of Chicago Law Review Online,* Vol. 82 No. 1, pp. 116-133, available at: https://chicagounbound.uchicago.edu/uclrev_online/vol82/iss1/8/, (accessed 3 January 2020).

Dalton, K. L. (2012), *Searching the Archive of Our Own: the usefulness of the tagging structure*, Masters, University of Wisconsin-Milwaukee, available at: http://dc.uwm.edu/etd/26 (Accessed 29 August 2014).

Driscoll, C. (2006), "One true pairing: the romance of pornography and the pornography of romance", in Hellekson, K. and Busse K. (Eds.), *Fan fiction and fan communities in the age of the internet.* McFarland, Jefferson, pp. 79-96.

Ewbank, A. D. (2015), "Library advocacy through Twitter: a social media analysis of #Savelibraries and #getESEAright", *School Libraries Worldwide,* Vol. 21 No. 2, available at: https://www.questia.com/library/journal/1P3-4037785131/library-advocacy-through-twitter-a-social-media-analysis (accessed 3 January 2020).

Fiesler, C., Feuston, J., and Bruckman, A. (2015), "Understanding copyright law in online creative communities", in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing – CSCW '15 in New York, NY, USA, 14-18 March 2015*, available at: http://dx.doi.org/10.1145/2675133.2675234 (accessed 23 January 2019).

Fruchterman, M. J., and Reingold, E. M. (1991), "Graph drawing by force-directed placement", *Software: Practice and Experience*, Vol. 21 No. 11, pp. 1129-1164, available at: http://dx.doi.org/10.1002/spe.4380211102 (accessed 3 January 2020).

Golder, S. A., and Huberman, B. A. (2006), "Usage patterns of collaborative tagging systems", *Journal of Information Science*, Vol. 32 No. 2, pp. 198-208, available at: http://dx.doi.org/10.1177/0165551506062337 (accessed 3 January 2020).

Gonzalez, V. M. (2016), "Swan Queen, shipping, and boundary regulation in fandom", *Transformative Works and Cultures,* Vol. 22, available at: http://dx.doi.org/10.3983/twc.2016.0669 (accessed 20 September 2019).

Gursoy, A., Wickett, K., and Feiberg, M. (2018), "Understanding tag functions in a moderated, user-generated metadata ecosystem", *Journal of Documentation*, Vol. 74 No. 3, pp. 490-508, available at: https://doi.org/10.1108/JD-09-2017-0134 (accessed 13 November 2020).

Hidderley, R., and Rafferty, P. (1997), "Democratic indexing: an approach to the retrieval of fiction", *Information Services and Use,* Vol. 17 No. 2/3, available at: http://dx.doi.org/10.3233/ISU-1997-172-304 (Accessed 22 October 2020).

Hoch, I. N. (2018), "Content, conduct, and apologies in Tumblr fandom tags", *Transformative Works and Cultures,* Vol. 27, available at: http://dx.doi.org/10.3983/twc.2018.1198 (accessed 3 January 2020).

Howard, K. J. (2018), "Emergence of a new method: The grounded Delphi method", *Library and Information Research*, Vol. 42 No. 126, pp. 5-31, available at: https://doi.org/10.29173/lirg746 (accessed 13 November 2020).

Johnson, B. (2016), "Live long and prosper: how the persistent and increasing popularity of fan fiction requires a new solution in copyright law", *Minnesota Law Review,* Vol.100 No. 4, pp. 1645-1687, available at: http://www.minnesotalawreview.org/wp-content/uploads/2016/04/Johnson_ONLINEPDF.pdf (accessed 3 January 2020).

Johnson, M., and Forsythe, C. (2019), "Disability and Accessibility Language in Subject Headings and Social Tags", *Catalogue & Index*, No. 197., pp. 16-26, available at: https://ir.lib.uwo.ca/fimspub/339 (accessed 16 November 2020).

Johnson, S. F. (2014), "Fan fiction metadata creation and utilization within fan fiction archives: three primary methods", *Transformative Works and Cultures,* 17, available at: http://dx.doi.org/10.3983/twc.2014.0578 (Accessed 16 November 2020).

Jones, B. (2014), "Fifty shades of exploitation: fan labor and *Fifty Shades of Grey*", *Transformative Works and Cultures,* Vol. 15, available at: http://dx.doi.org/10.3983/twc.2014.0501 (accessed 27 July 2019).

King, M., and Ridgway, J. (2019), "The female fan goes shopping: Satisfaction, involvement and utilitarian value when shopping for women's *Star Wars* merchandise", *Journal of Fandom Studies*, Vol. 7 No. 3, pp. 229-244, available at: https://doi.org/10.1386/jfs_00002_1 (accessed 16 November 2020).

Lantagne, S. M. (2015), "Sherlock Holmes and the case of the lucrative fandom: recognizing the economic power of fanworks and reimagining fair use in copyright", *Michigan Telecommunications and Technology Law Review,* Vol. 21 No. 2, pp. 263-315, available at: https://repository.law.umich.edu/mttlr/vol21/iss2/2 (accessed 3 January 2020).

Liu, H., Morstatter, F., Tang, J., et al. (2016), "The good, the bad, and the ugly: uncovering novel research opportunities in social media mining", *International Journal of Data Science and Analytics,* Vol. 1 No. 3-4, pp. 137-143, available at: http://dx.doi.org/10.1007/s41060-016-0023-0 (accessed 3 January 2020).

Lund, B. (2020), "Review of the Delphi method in library and information science research", *Journal of Documentation,* Vol. 76 No.4, pp. 929-960, available at: https://doi.org/10.1108/JD-09-2019-0178 (accessed 8 September 2020).

Marlow, C., Naaman, M., boyd, d., et al. (2006), "HT06, tagging paper, taxonomy, Flickr, academic article, to read", in *Proceedings of the seventeenth conference on Hypertext and hypermedia - HYPERTEXT '06 in Odense, Denmark, 22-25 August 2006*, Association for Computing

Machinery, New York, NY, pp. 31-40, available at: http://dx.doi.org/10.1145/1149941.1149949 (accessed 3 January 2020).

McGuire, M. (2017), "Reblogging as writing: the role of Tumblr in the writing classroom", in Bryant, K. N. (Ed.), *Engaging 21st Century Writers with Social Media,* IGI Global, Hershey, PA, pp. 116-131.

Messina, C. M. (2019), "Tracing Fan Uptakes: Tagging, Language, and Ideological Practices in *The Legend of Korra* Fanfictions", *The Journal of Writing Analytics*, Vol. 3, pp. 151-182, available at: https://wac.colostate.edu/docs/jwa/vol3/messina.pdf (accessed 4 January 2020).

Misailidou, E. (2017), "Social media fandom: The construction of identity in the cases of "The 100" and "Once Upon A Time" Tumblr communities", in *Wireless Telecommunications Symposium (WTS), 2017* in Chicago, IL, USA, 26-28 April 2017*,* pp. 1-11, available at: http://dx.doi.org/10.1109/WTS.2017.7943539 (accessed 3 January 2020).

Monnin, A. (2009), "From Game Neverending to Flickr. Tagging systems as ludic systems and their consequences", in *Proceedings of the WebSci'09: Society On-Line in Athens, Greece, 18-20 March 2009,* pp. 1-5, available at: https://www.academia.edu/586952/From_Game_Neverending_to_Flickr._Tagging_systems_as_ludic_systems_and_their_consequences_short_paper_ (accessed 3 January 2020).

Navar-Gill, A., and Stanfill, M. (2018), ""We Shouldn't Have to Trend to Make You Listen": Queer Fan Hashtag Campaigns as Production Interventions", *Journal of Film and Video*, Vol. 70, No. 3-4, pp. 85-100, available at: https://doi.org/10.5406/jfilmvideo.70.3-4.0085 (accessed 10 November 2020).

Navarro Bullock, B., Hotho, A., and Stumme, G. (2018), "Accessing Information with Tags: Search and Ranking", in Brusilovsky P., He D. (Eds.), *Social Information Access,* Springer, Cham, pp. 310-343.

Ocepek, M. G., et al. (2018), "Fandom, food, and folksonomies: the methodological realities of studying fun life-contexts", in *Proceedings of ASIS&T Annual Meeting 2018*, *in Vancouver, Canada, 10-14 November 2018,* pp. 712-715, available at: https://doi.org/10.1002/pra2.2018.14505501089 (accessed 3 January 2020).

Olbrich, R. and Holsing, C. (2011), "Modeling consumer purchasing behavior in social shopping communities with clickstream data", *International Journal of Electronic Commerce,* Vol. 16 No. 2, pp. 15-40, available at: https://doi.org/10.2753/JEC1086-4415160202 (accessed 3 January 2020).

Petersen, L. N. (2017), ""The florals": female fans over 50 in the *Sherlock* fandom", *Transformative Works and Cultures*, Vol. 23, available at: http://dx.doi.org/10.3983/twc.2017.0956 (accessed 21 July 2019).

Postigo, H. (2008), "Video game appropriation through modifications: attitudes concerning intellectual property among modders and fans", *Convergence,* Vol. 14 No. 1, pp. 59-74, available at: https://doi.org/10.1177/1354856507084419 (accessed 3 January 2020).

Price, L. (2017) *Serious leisure in the digital world: exploring the information behaviour of fan communities*, Ph.D., City, University of London, available at: http://openaccess.city.ac.uk/19090/ (accessed 24 November 2018).

Price, L., and Robinson, L. (2017), ""Being in a knowledge space": information behaviour of cult media fan communities", *Journal of Information Science,* Vol. 43 No. 5, pp. 649-664, available at: http://dx.doi.org/10.1177/0165551516658821 (accessed 3 January 2020).

Samanta, K., and Rath, D. (2020), "User-generated social tags versus librarian-generated subject headings: A comparative study in the domain of history", *DESIDOC Journal of Library and Information Technology*, Vol. 40 No. 3, pp. 176-184, available at: https://doi.org/10.14429/djlit.40.03.15413 (accessed 20 June 2020).

Sarkar, A., and Bhattacharya, U. (2020). "Controlled Vocabularies versus Social Tags: A Brief Literature Review", *Library Philosophy and Practice,* 3658, available at: https://digitalcommons.unl.edu/libphilprac/3658/ (accessed 9 February 2020).

Smith, G. (2008a), *Tagging: people-powered metadata for the social web*. New Riders, Berkeley, CA.

Smith, G. (2008b), "Information architecture: Tagging: emerging trends", *Bulletin of the American Society for Information Science and Technology*, Vol. 34 No. 6, pp. 14-17, available at: https://doi.org/10.1002/bult.2008.1720340607 (accessed 3 January 2020).

Smith, M., Shneiderman, B., Milic-Frayling, N., et al. (2009), "Analyzing (social media) networks with NodeXL", in *Proceedings of the Fourth International Conference on Communities and Technologies – C&T '09 in University City, Pennsylvania, 25-27 June 2009*, pp. 255-263, available in: https://doi.org/10.1145/1556460.1556497 (accessed 3 January 2020).

Szuban, P. (2018), "Reconstituting Vocabularies: User Generated Databases, Social Tagging, and Folksonomies in Giantbomb's Videogame Wiki Database", *The iJournal*, Vol. 4 No. 1, pp. 41-49, available at: https://theijournal.ca/index.php/ijournal/article/view/32139 (accessed 26 January 2019).

Thelwall, M. (2004), *Link analysis: an information science approach.* Elsevier Academic Press, London.

Thelwall, M., and Stuart, D. (2006), "Web crawling ethics revisited: cost, privacy and denial of service", *Journal of the American Society for Information Science and Technology*, Vol. 57 No. 13, pp. 1771-1779, available at: http://dx.doi.org/10.1002/asi.20388 (accessed 3 January 2020).

Unwrapping Tumblr (2014), "Tumblr tags limit 20 per post", 15 November, available at: https://unwrapping.tumblr.com/post/102742630432/tumblr-tags-limit (accessed 3 January 2020).

Wargo, J. M. (2017), "#donttagyourhate: reading, collecting and curating as genres of participation in LGBT youth activism on Tumblr", *Digital Culture & Education,* Vol. 9 No. 1, pp. 14-30, available at: https://www.digitalcultureandeducation.com/s/wargo-February-2017.pdf (accessed 3 January 2020).

Warner, I., (2011), *Will social classification replace traditional forms of organising information on the web? Will organizations structure content differently for their clients to access?* Masters of Information Studies, University of Canberra, available at: https://www.academia.edu/1044388/Will_social_classification_replace_traditional_forms_of_organising_information_on_the_web (accessed 3 January 2020).

Wilson, T. D. (2000), "Human information behaviour", *Informing Science,* Vol. 3 No. 2, pp. 49-55, available at: http://inform.nu/Articles/Vol3/v3n2p49-56.pdf (accessed 17 October 2020).

Zhang, L., Luo, M., and Boncella, R. (2020), "Product information diffusion in a social network", *Electronic Commerce Research*, Vol. 20, No. 1, pp. 3-19, available at: https://doi.org/10.1007/s10660-018-9316-9 (accessed 9 November 2020).

Zollers, A. (2007), “Emerging motivations for tagging: expression, performance, and activism”, in *WWW 2007: Proceedings of the 16th International World Wide Web Conference*, in Banff, Canada, 8-12 May 2007, available at: http://www2007.org/workshops/paper_55.pdf (Accessed 15 December 2019).

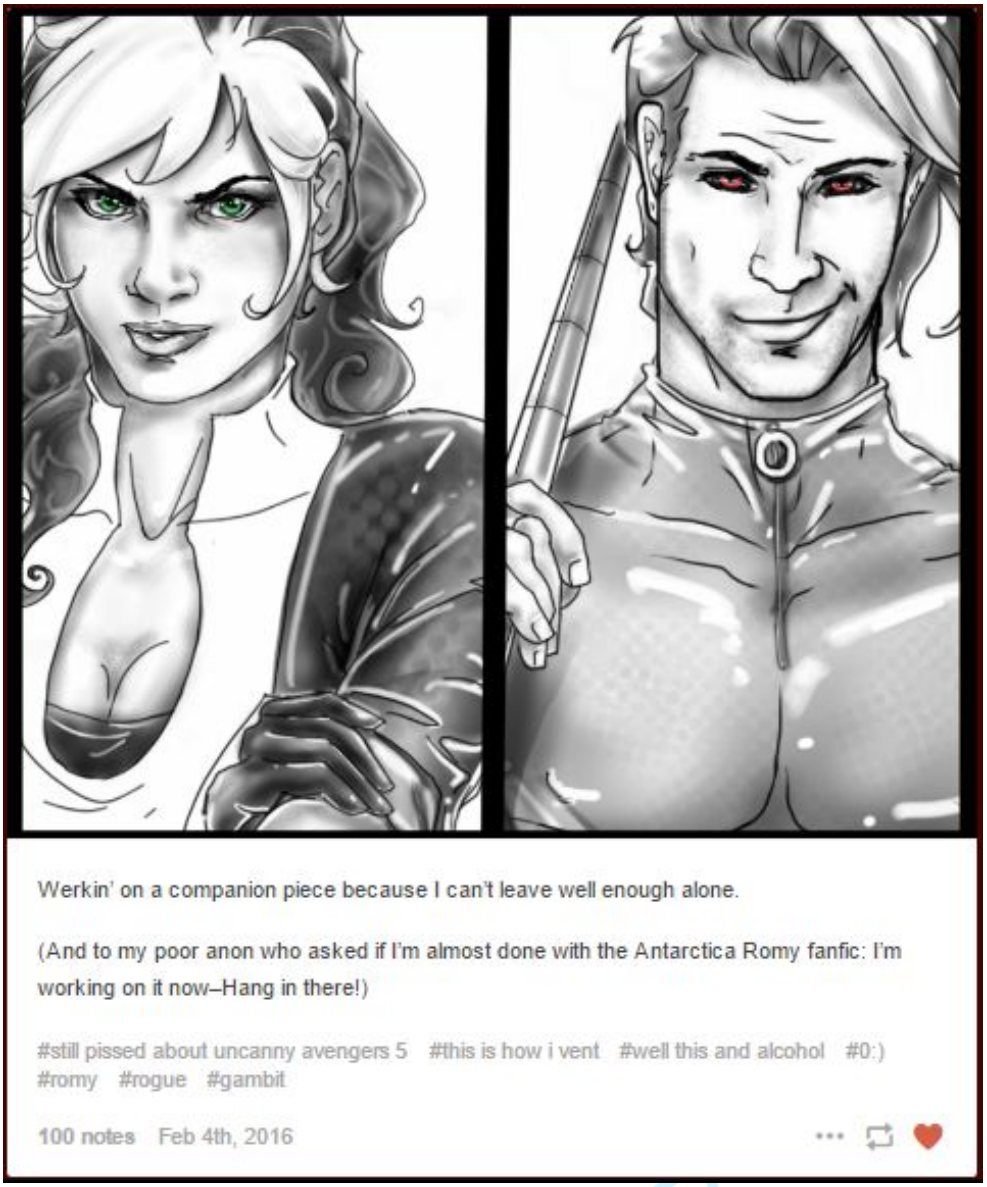

Figure 1. Tumblr post by Participant B, 4th Feb 2016 (reproduced with permission).

They Never Trained Me For This by IronRaven

14 Oct 2016

X-Men Evolution

Creator Chose Not To Use Archive Warnings, laura/alex summers, Kurt/kitty, rogue/remy, Logan/Ororo Munroe, Kitty Pryde & Kurt Wagner, Rogue & Kurt Wagner, rogue & kitty pryde, Laura/X-23, Kitty, Rogue, Logan, Ororo Munroe, Kurt Wagner, Remy LeBeau, Elizabeth Braddock, Alex Summers, Charles Xavier, And many many more, AU jumping from the the aftermath of the battle with Apocolypse., Xavier Institute., cute and crazy, Friendship, True Love, Violence, Revenge, School, normal people suck, Conspiracies, Mental Health Issues, Logan is a good daddy, Laura hates it when people call him her dad, Originally Posted on FanFiction.Net

X-23 joins the Xavier Institute. For most schools, madness and terror would ensue. For this one, she's a little more odd than most, but she's pretty normal person here. Normal is something she's never been. A person is something she's never been. She was trained to track, to kill, to survive and to adapt. Not to have family, friends, people she can count on.

Untrained, she will be thrown into the deep end. The end where they catch you.

Language: English Words: 12,387 Chapters: 7/? Collections: 1 Comments: 4 Kudos: 33 Bookmarks: 5 Hits: 858

Figure 2. Archive entry for an X-Men Evolution fanfiction on AO3, "They Never Trained Me For This", 14th October 2016. Source: http://archiveofourown.org/ (Retrieved 5 January 2020).

300x144mm (72 x 72 DPI)

Tag type	Examples
Descriptive	css, webdesign, ajax, Minnesota, drama, gardening, zen, microfinance, music, halo3, networks, sushi, hibiscus
Resource	blog, book, video, photo
Ownership/Source	nytimes, genesmith (author), newriders
Opinion	cool, funny, *****, lame, beautiful, crap, defective by design
Self-reference	mystuff, mine, me
Task Organizing	toread, todo, work
Play and Performance	squaredcircle, seenlive, aka vogon poetry

Table 1. Tag types (Smith 2008a).

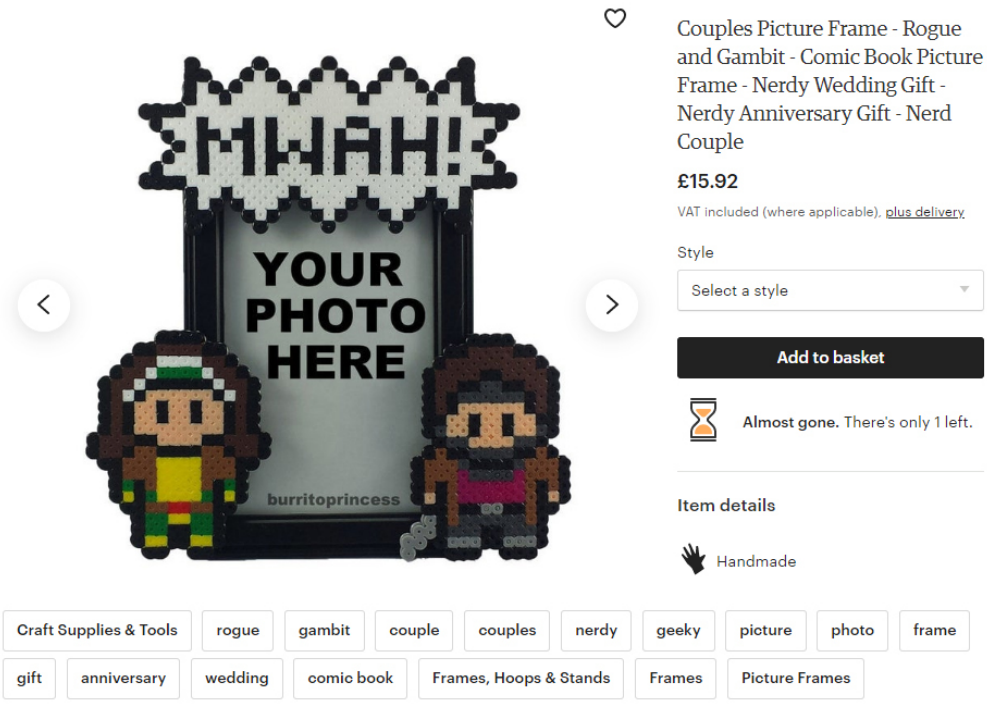

Figure 3. Couples picture frame – Rogue and Gambit. Tags are displayed at the bottom, in the box titled 'Related to this item'. The item details, usually under the picture, have been cropped here to save space. Source: https://www.etsy.com/ (Retrieved 5 January 2020)

324x228mm (72 x 72 DPI)

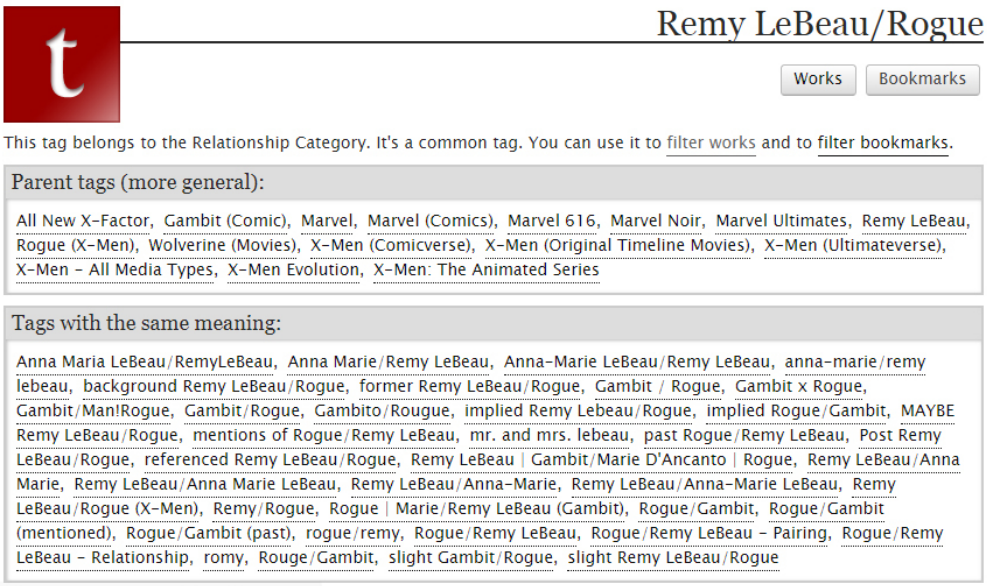

Figure 4. AO3's 'Romy' tag homepage. Note that the standard form for 'Romy' on AO3 is 'Remy LeBeau/Rogue'. User tags that have been designated by tag wranglers as synonyms are shown under 'Tags with the same meaning' (retrieved 5 January 2020).

302x179mm (72 x 72 DPI)

CODE	TAG TYPE/SUB-TYPE	DEFINITION	EXAMPLES
1	**Descriptive**	Describes content	vintage; commission; black and white; regram
1.1	Fandom	Describes fandom	X-Men; Marvel; Avengers; Harry Potter
1.2	Ship	Describes characters in a romantic relationship	Romy; Erik Lehnsherr/Charles Xavier; loroki
1.3	Character	Describes characters	Gambit; Rogue; Thor; Wade Wilson
1.4	Genre	Describes genre of resource	drabble; fluff; angst; slash; steampunk
1.5	Event	Describes a 'real world' event	Christmas; Valentines Day; dragoncon
1.6	Person	Describes a 'real world' person	Channing Tatum
1.7	Friendship	Describes characters in a friendship	kitty pryde & kurt wagner; darcy and logan
1.8	Organisation/Team/Group	Describes a group of people	witches; Hydra; X-Men; Illuminati
1.9	Location	Describes a location or setting	Alkali Lake; Xavier Institute; Wakanda
1.10	Plot	Describes a fictional story element	M-Day; Crimson Gem of Cyttorak
1.11	Warning	Describes sensitive content	spoilers; swearing; rape/non-con
2	**Resource**	Type of resource	comics; drawing; photo; video
2.1	Fanwork	Type of fan resource	fanfic; fanart; cosplay; fanfic rec list
2.2	Title of fanwork	Title of fan resource	In Between; Loki and the Loon
2.3	Citation	Citation of fan resource	Episode: Shadowed Past; X-Men Legacy 272
3	**Ownership**	Ownership of tagger	mike draws
3.1	Creator/source	Name of fan resource creator	Jim Lee; toyscomics; bbrae; ishandahalf
3.2	Recipient	Name of intended recipient of fan resource	txpeppa
4	**Opinion**	Opinion on resource	sexy; geeky; quirky; badass; epic
4.1	Communication	Communicates thoughts	I blame Tumblr; I need this shirt; great gift idea
4.2	Explanatory	Explains resource content	this is how I vent; iron fist is shameless
4.3	Affective	Explains emotional reaction	poor Pietro; ineedhelp; theyre so cute omg
4.4	Conversational & enunciative	Instigates or responds to a dialogue	why?; ask me stuff; leah shut up; askbox
4.5	Emoticon	Visual communication	XD; :D; 0:)
5	**Self-reference**	Reference to tagger/self	personal post; my art; self; my life
6	**Task organising**	Personal organisation of resource	work in progress; other character tags to be added; queueballs
7	**Play & performance**	Resource is part of an event, or (semi-) formal performance	prompt fill; fangirl challenge; frostiron month

Table 1. Fan-tag taxonomy as developed during analysis of the 'Romy' tag. Tags in bold denote Smith's (2008a) original categories. Non-bold tags denote those developed through inductive coding of the datasets. Examples given are taken from the datasets.

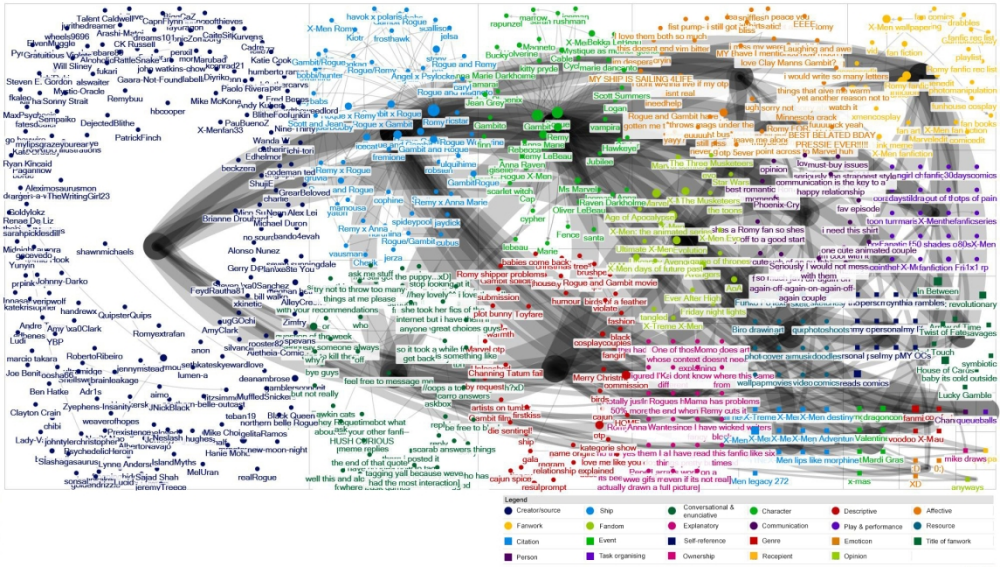

Figure 5. Co-occurrence graph for the 'Romy' tag on Tumblr.

529x301mm (96 x 96 DPI)

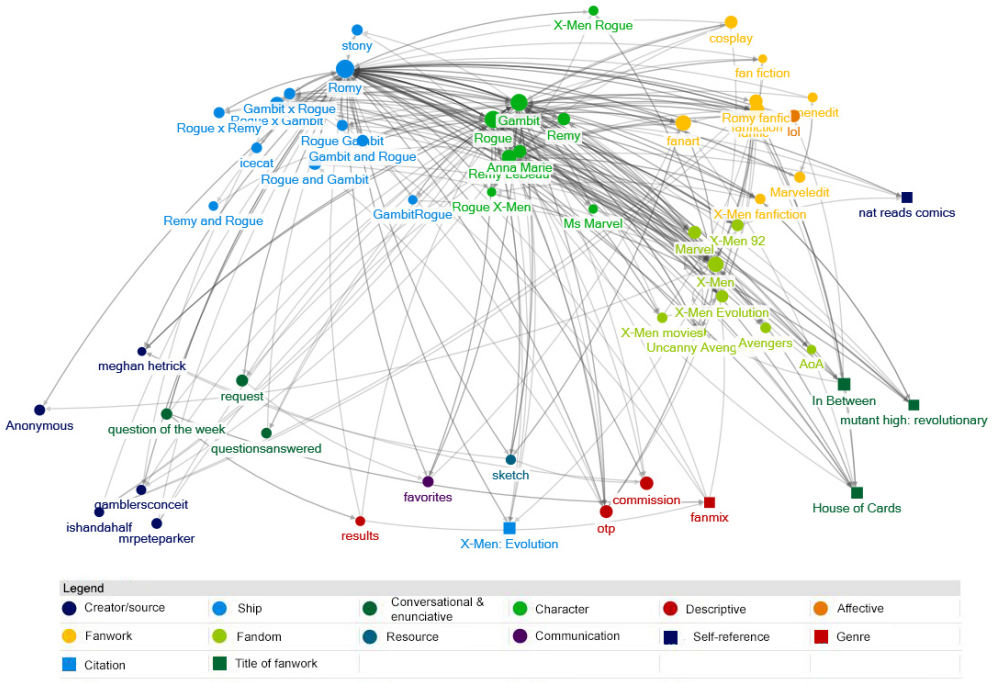

Figure 6. Co-occurrence graph for the 'Romy' tag on Tumblr. All displayed tags have a betweeness centrality of 1+.

384x262mm (72 x 72 DPI)

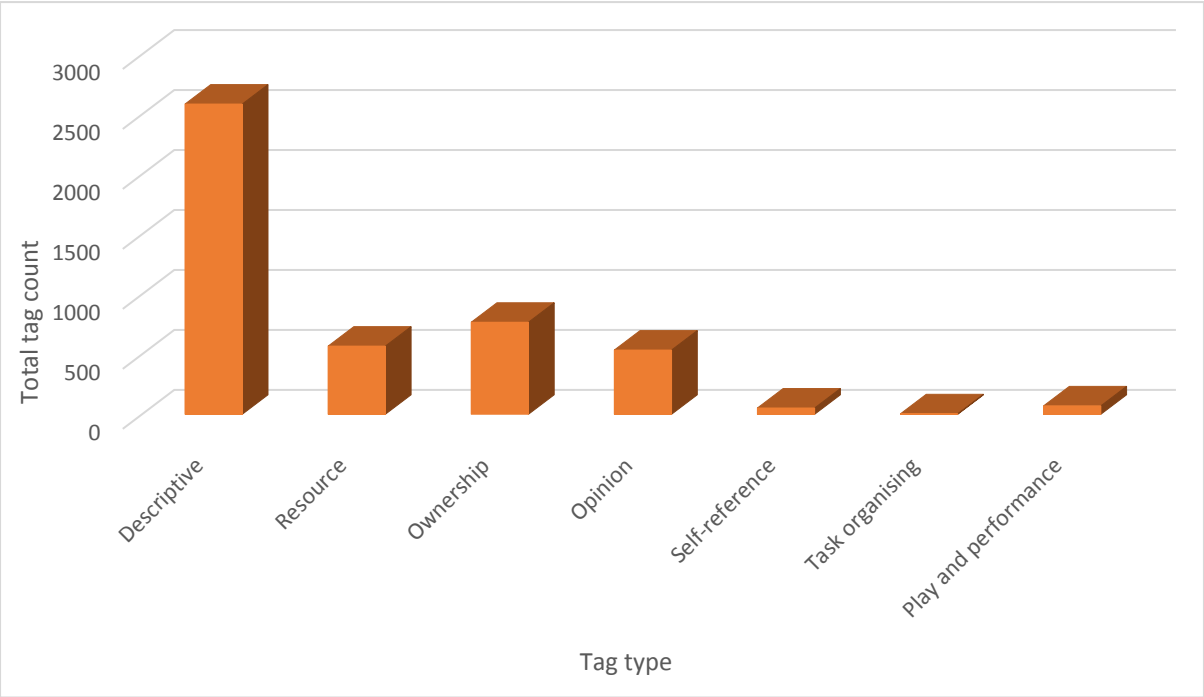

Figure 7. Total Tumblr tag count by type (see Table 2 for tag types).

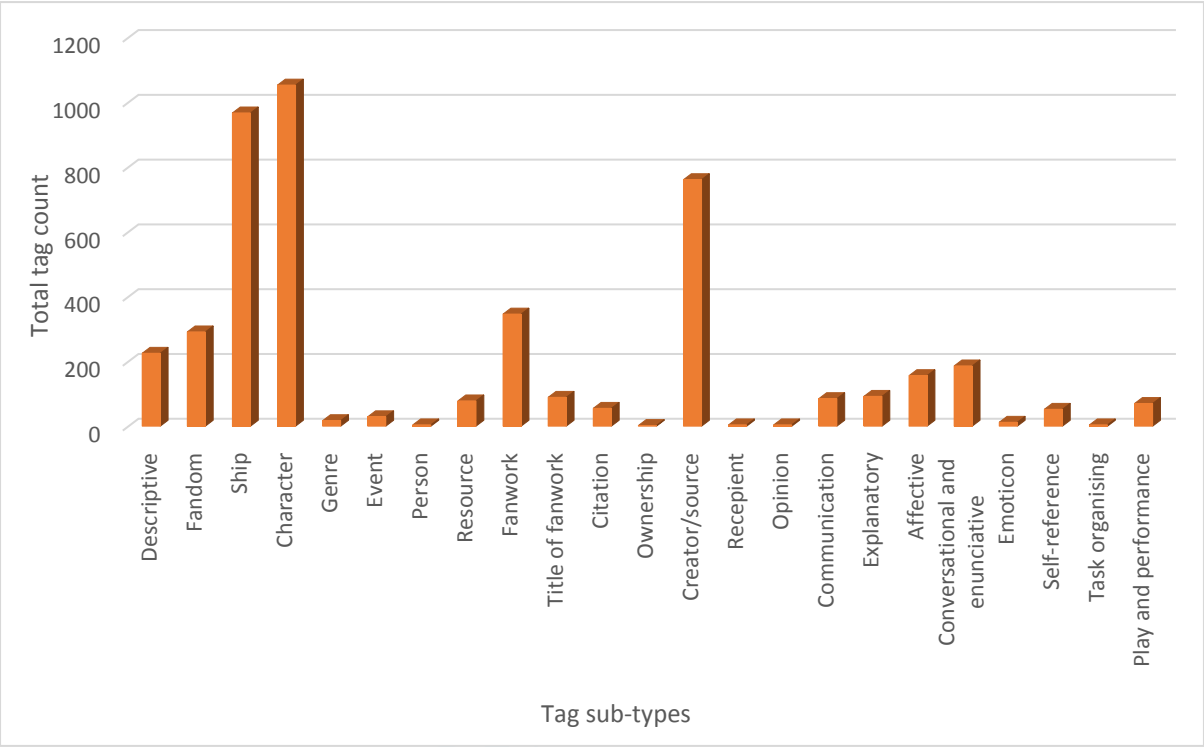

Figure 8. Total Tumblr tag count by sub-type (see Table 2 for tag types). The **Character** type (a Descriptive sub-type) was the most frequently used tag, closely followed by **Ship** (also a Descriptive sub-type).

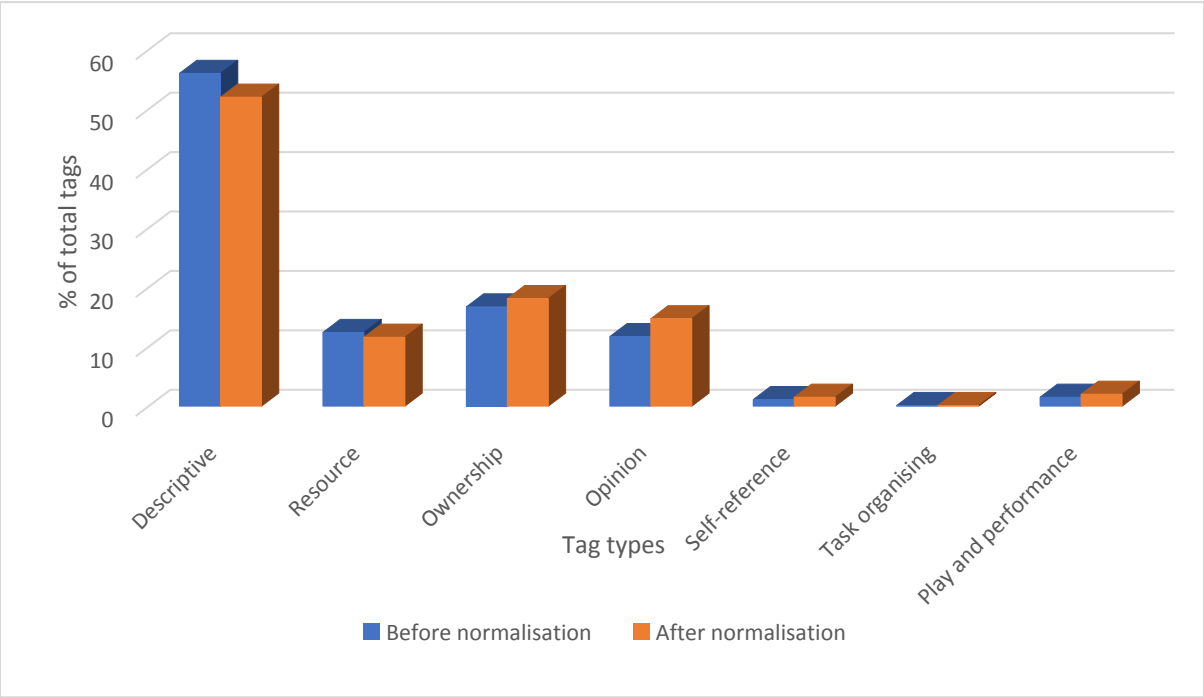

Figure 9. A comparison of Tumblr tag counts by type, before and after normalisation (see Table 2 for tag types). There is no marked difference between datasets before and after normalisation.

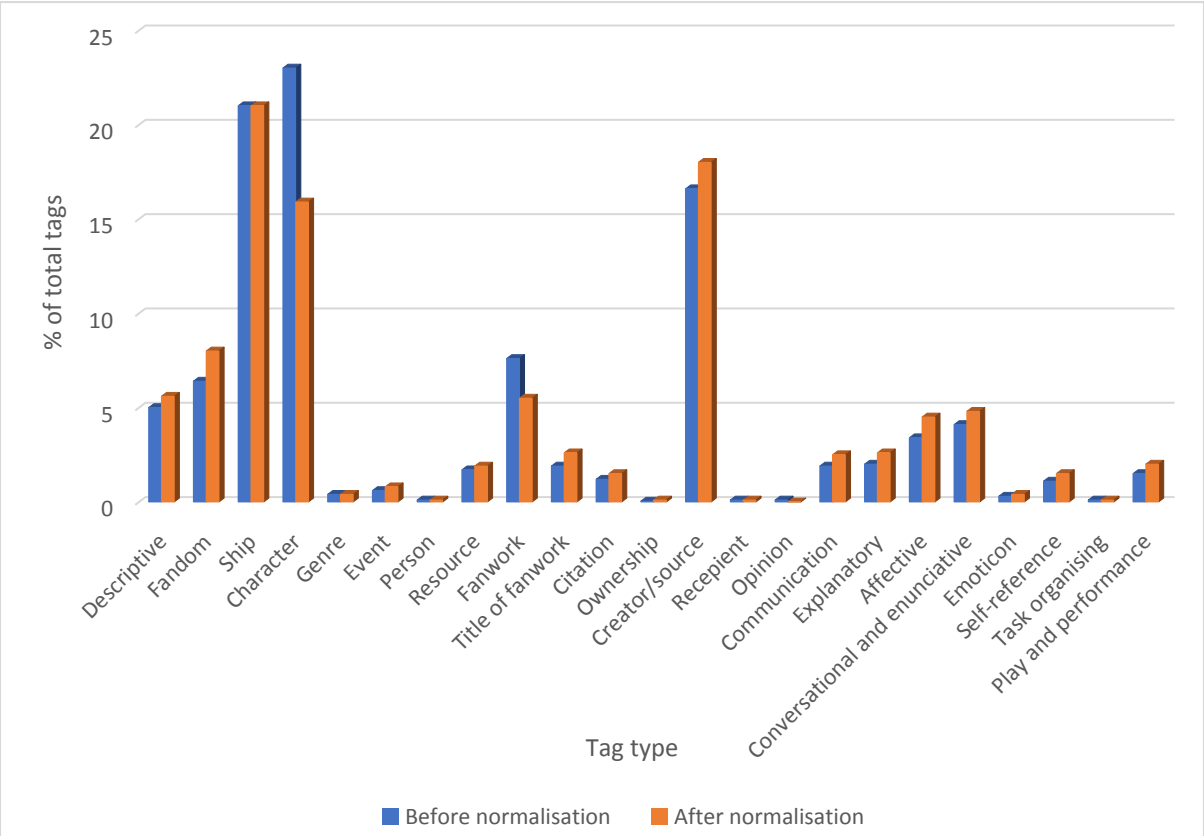

Figure 10. A comparison of Tumblr tag counts by sub-type, before and after normalisation (see Table 2 for tag types). The most marked difference is in **Character** sub-type usage.

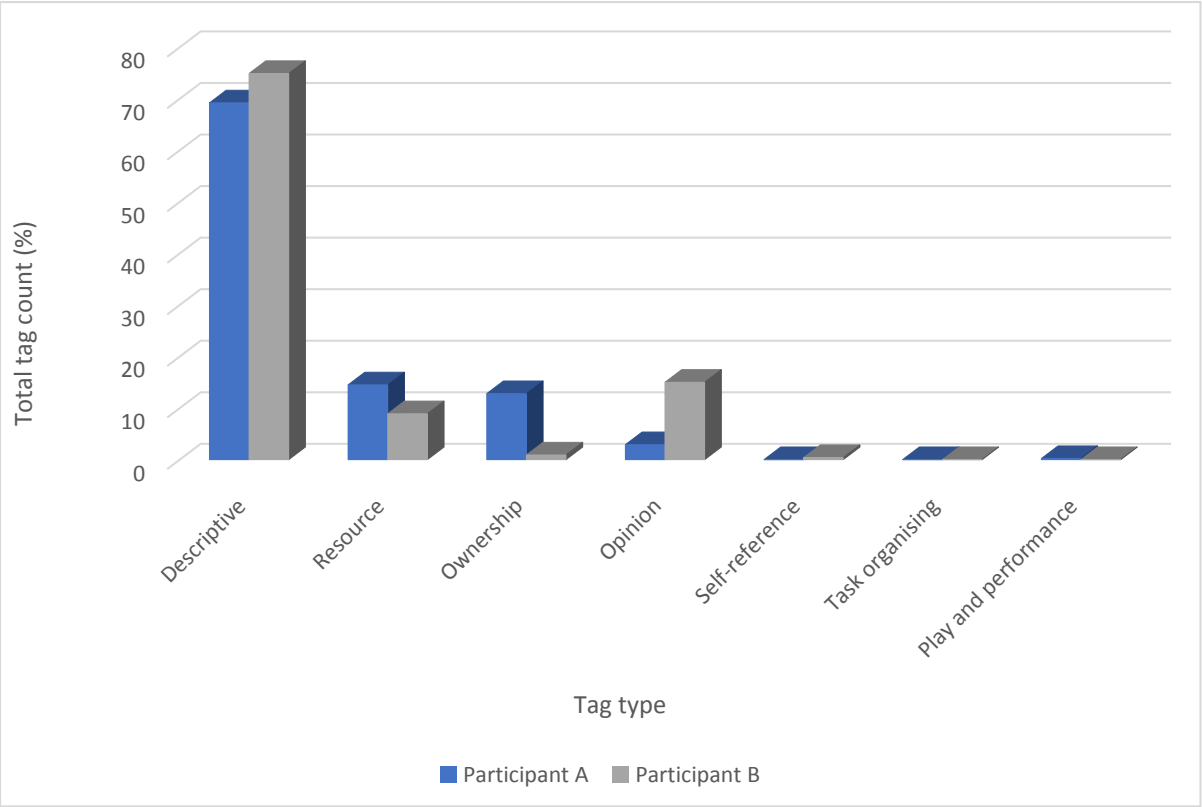

Figure 11. Comparison of tag type usage by Participant A and Participant B, by percentage (see Table 2 for tag types). Participant B showed marked use of the **Opinion** tag type.

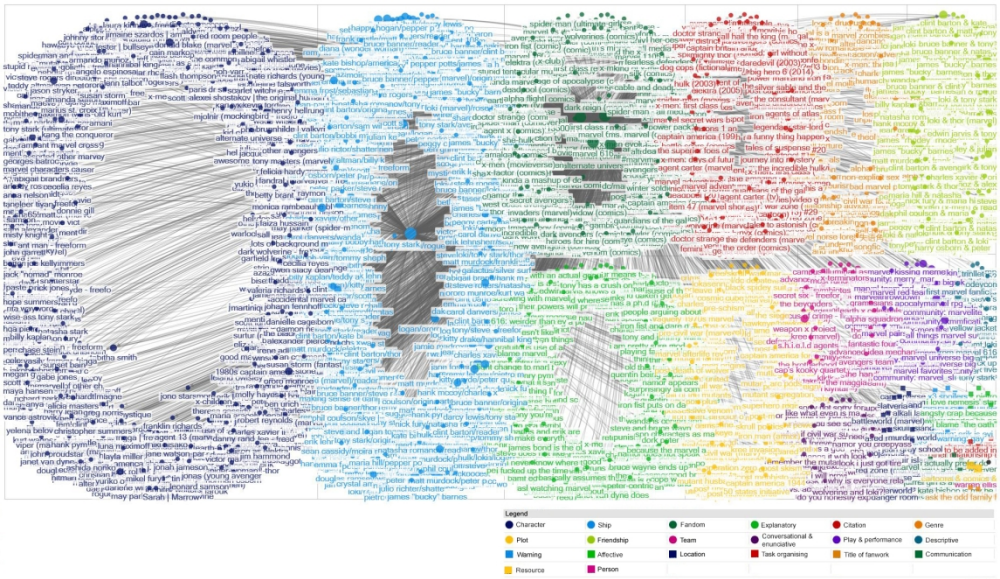

Figure 12. Co-occurrence graph for the 'Rogue/Remy LeBeau' tag on AO3 (from the wrangled dataset).

529x306mm (96 x 96 DPI)

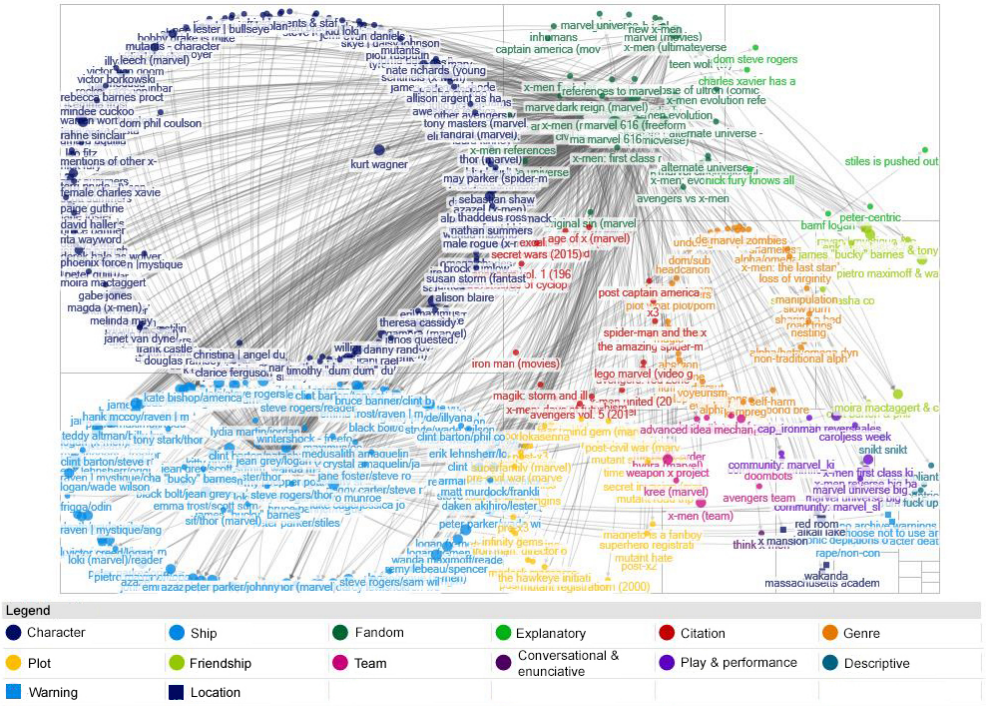

Figure 13. Co-occurrence graph for the 'Rogue/Remy LeBeau' tag on AO3 (from the wrangled dataset). All displayed tags have a betweeness centrality of 1+.

264x188mm (96 x 96 DPI)

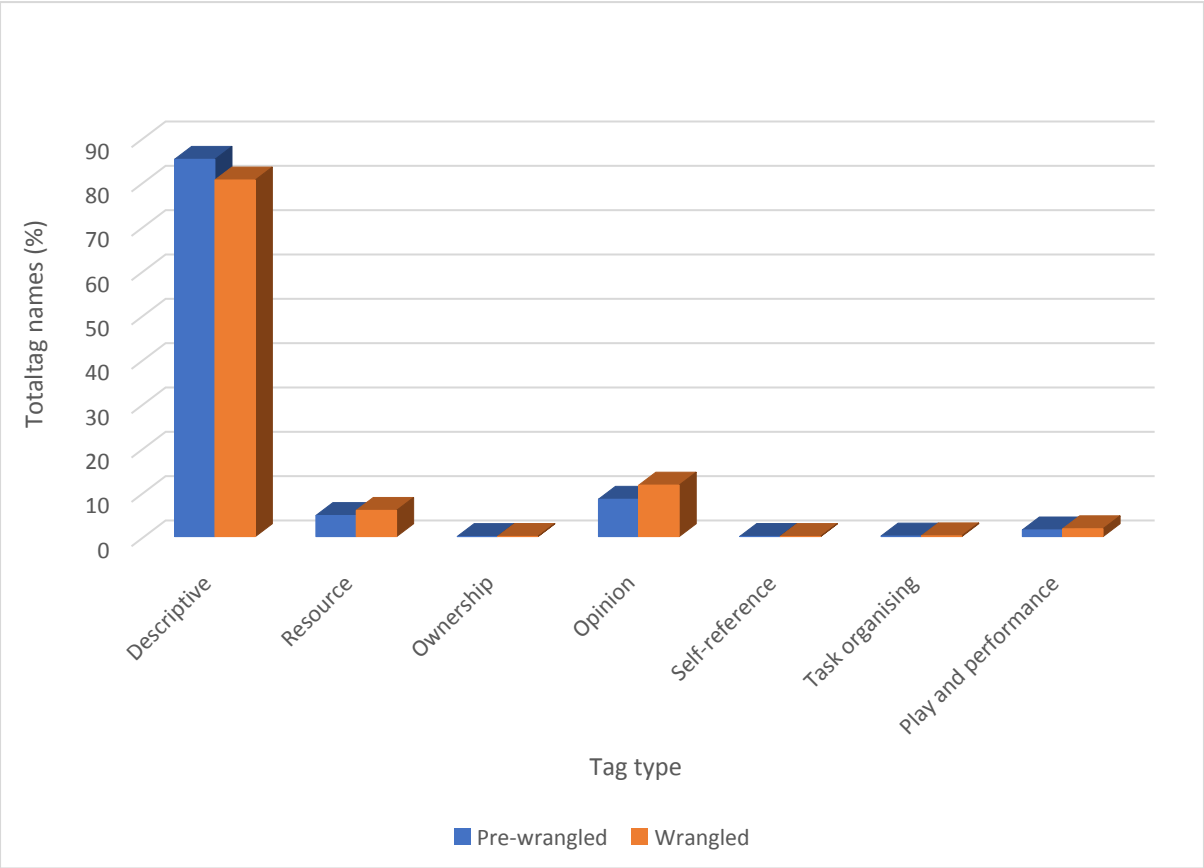

Figure 14. Comparison of pre-wrangled and post-wrangled tag names, by type (see Table 2 for tag types). There is little appreciable difference.

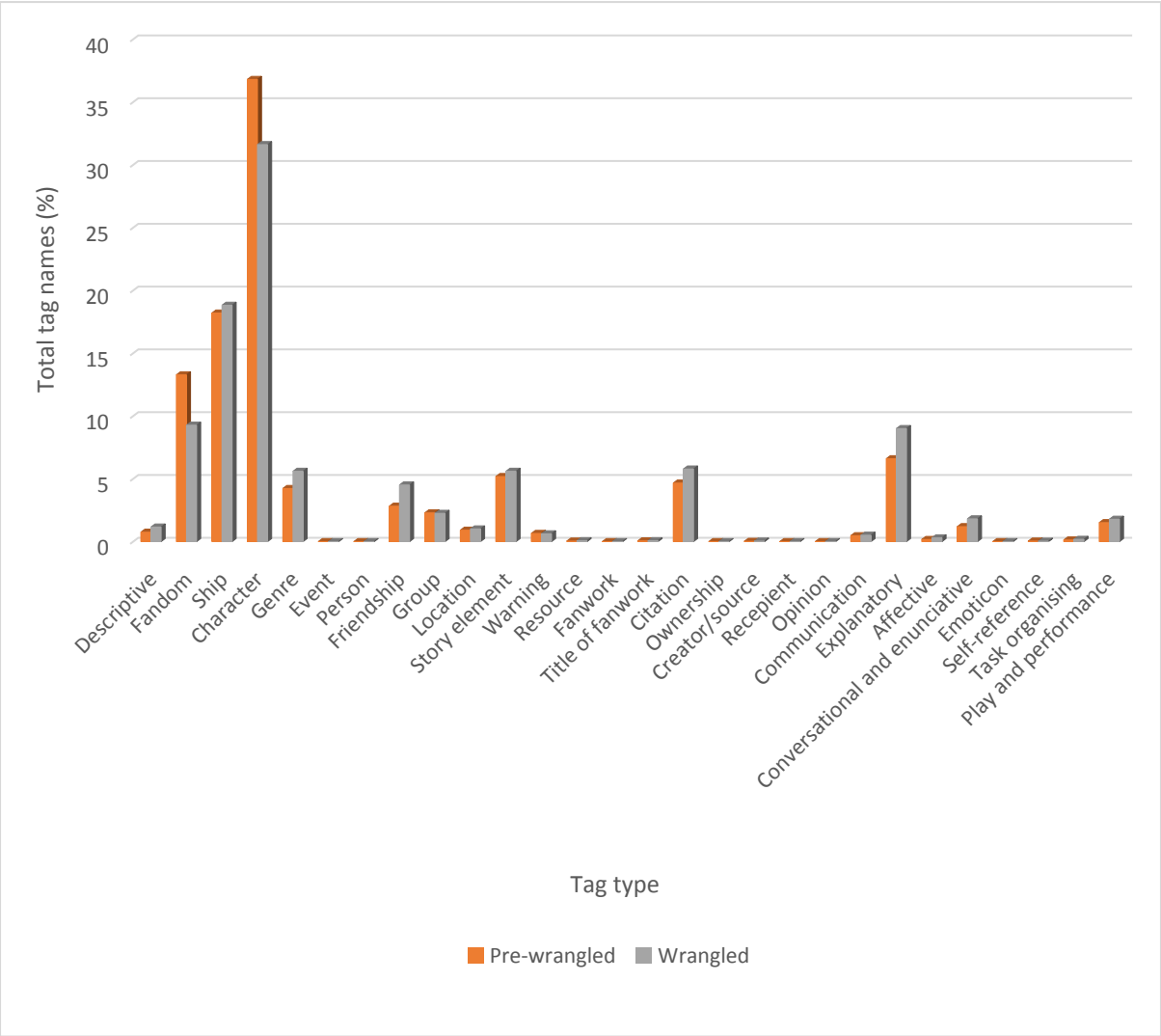

Figure 15. Comparison of pre-wrangled and post-wrangled tag names, by sub-type (see Table 2 for tag types).

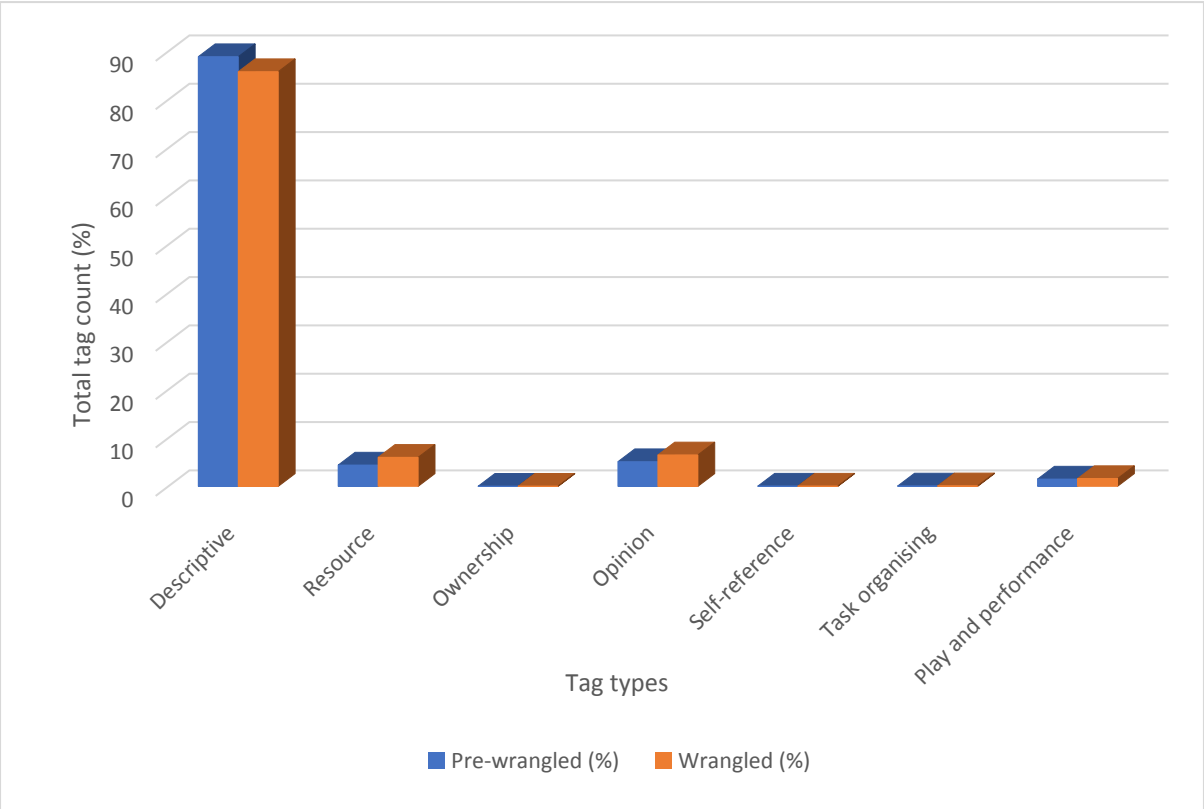

Figure 16. Comparison of the percentage of the total tag count in the pre- and post-wrangled datasets, arranged by tag type (see Table 2 for tag types).

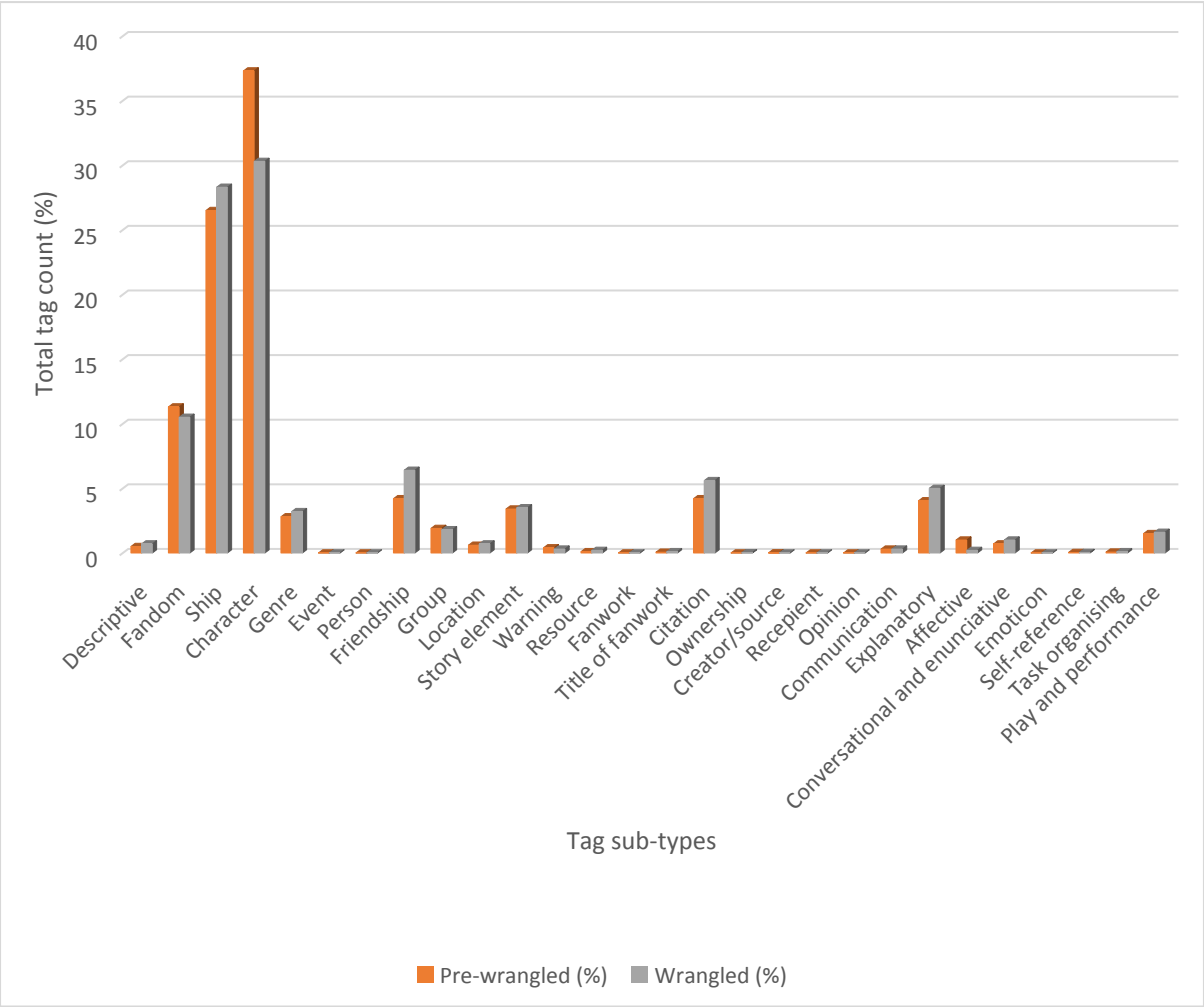

Figure 17: Comparison of the percentage of the total tag count in the pre- and post-wrangled datasets, arranged by tag sub-type (see Table 2 for tag types). **Ship**, **Character**, **Friendship**, **Citation** and **Explanatory** sub-types show the most marked difference.

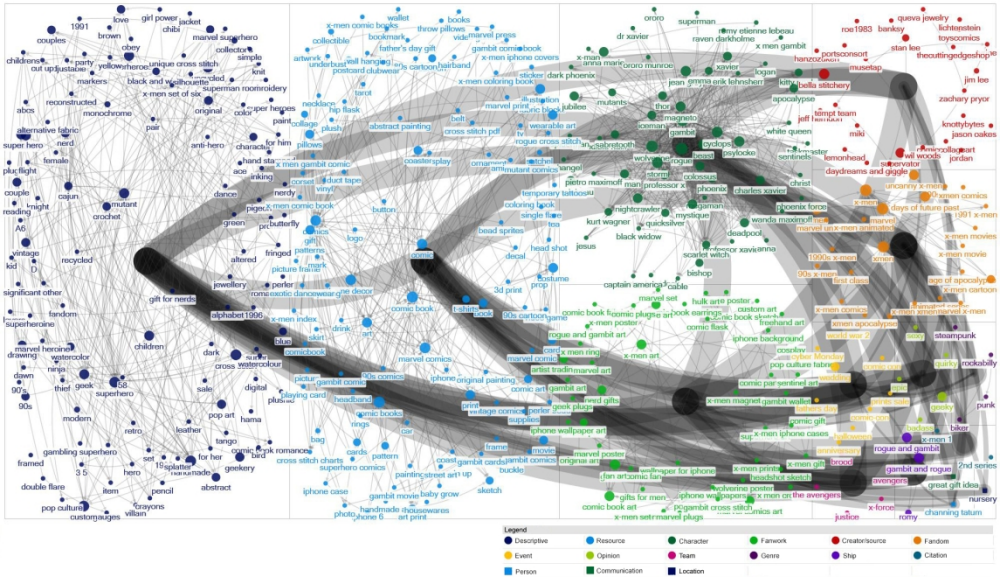

Figure 18. Co-occurrence graph for the 'Rogue and Gambit' tag on Etsy.

529x306mm (96 x 96 DPI)

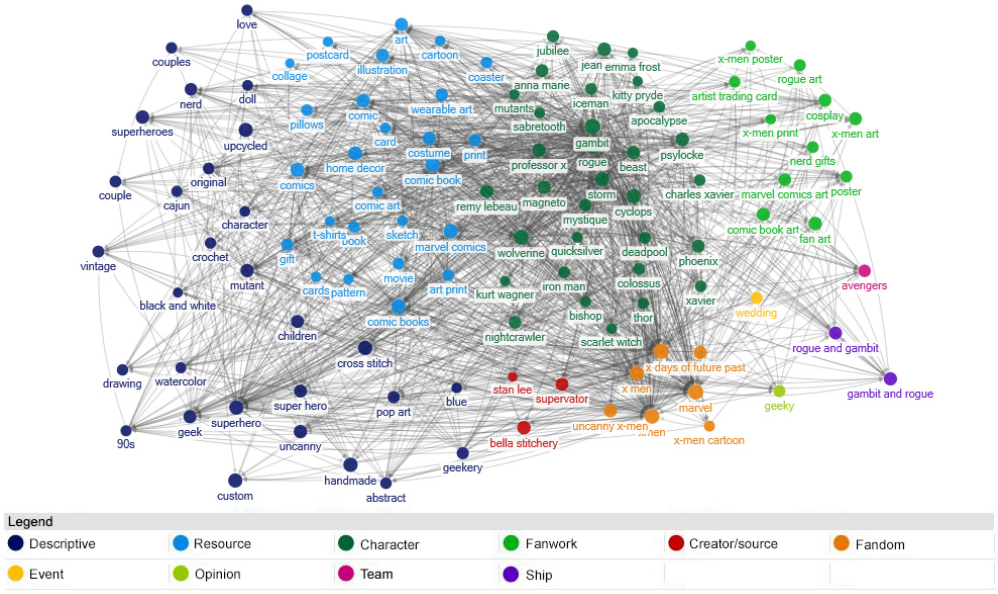

Figure 19. Co-occurrence graph for the 'Rogue and Gambit' tag on Etsy. All displayed tags have a betweenness centrality of 1+.

349x214mm (72 x 72 DPI)

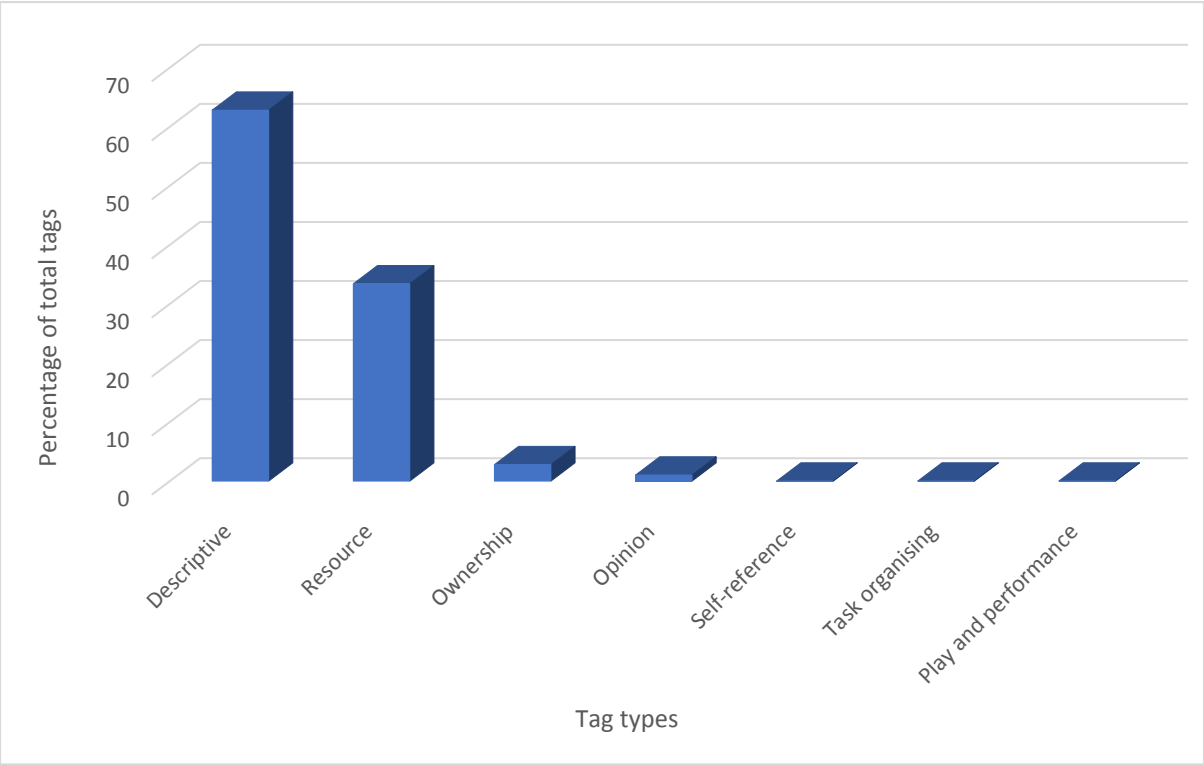

Figure 20. Percentage of Etsy tag usage by type (see Table 2 for tag types). **Descriptive** and **Resource** tag types dominate this set.

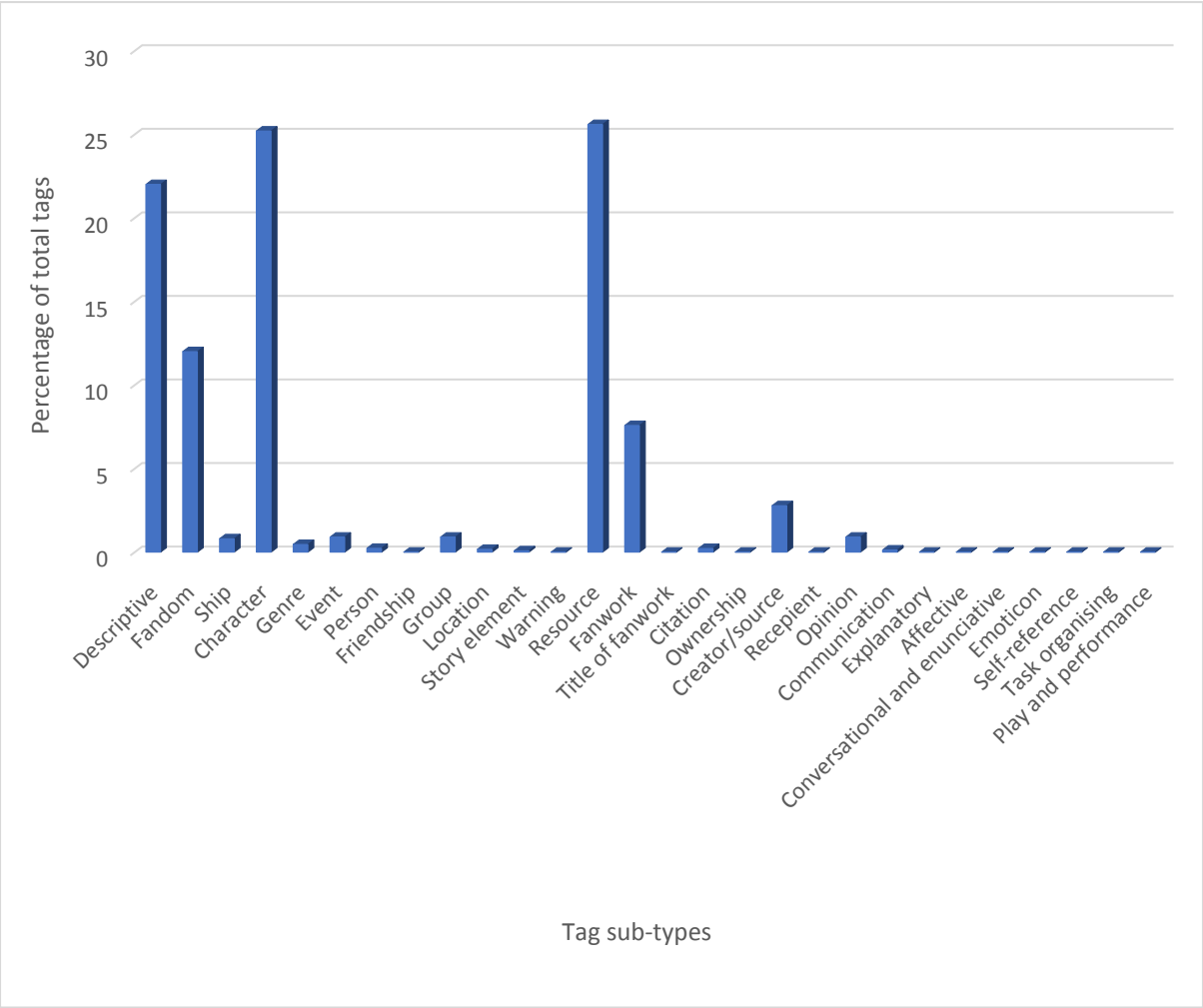

Figure 21. Percentage of Etsy tag usage by sub-type (see Table 2 for tag types). The bulk of **Descriptive** tag sub-type usage is the **Character** sub-type.

Figure 22. Phoenix symbol curved outline decal, on sale on Etsy. Source: https://www.etsy.com (Retrieved 5 January 2020).

295x189mm (72 x 72 DPI)

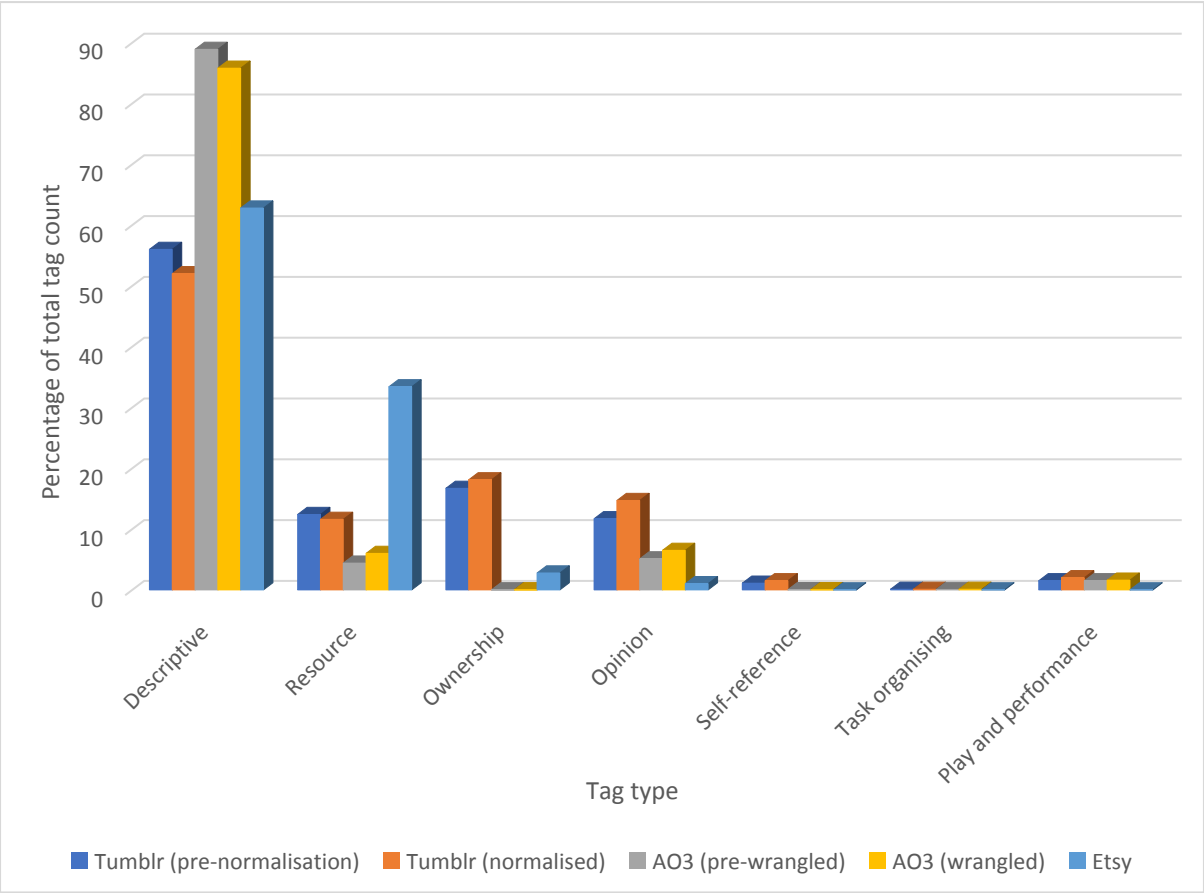

Figure 23. Comparison of the total tag count by type, across the three platforms scraped (see Table 2 for tag types). **Descriptive** tag types were the most frequently used across all datasets.

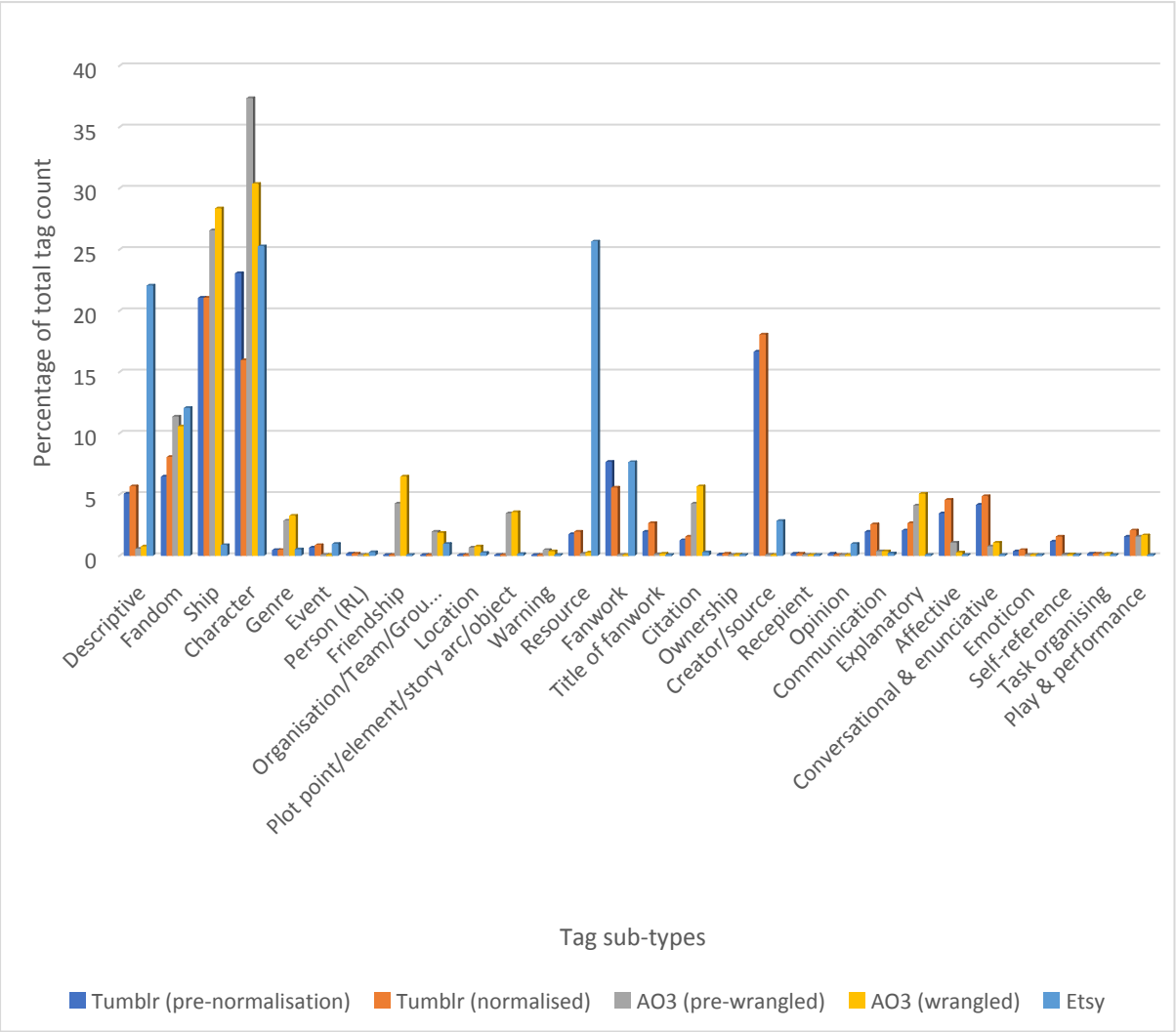

Figure 24. Comparison of the total tag count by sub-type, across the three platforms scraped (see Table 2 for tag types). The **Character** sub-type (a **Descriptive** tag sub-type) was most frequently used in all datasets, apart from Etsy, where the **Resource** tag type saw slightly more usage than the **Character** sub-type.

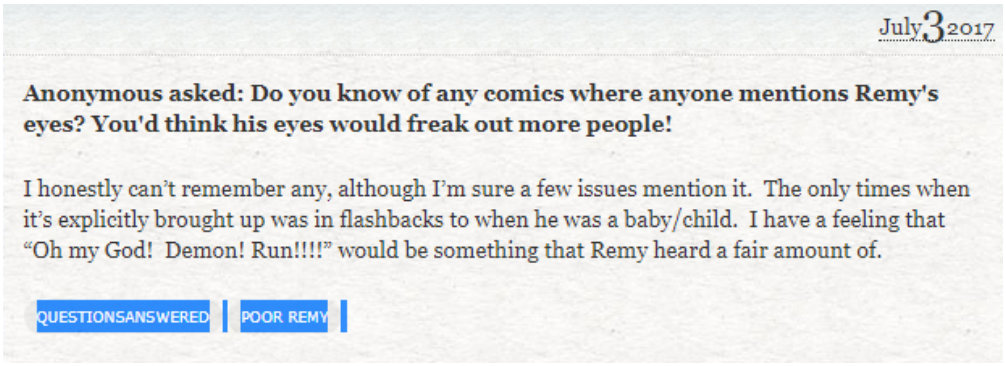

Figure 25. A 'Questions Answered' Tumblr post by Participant A. The tags used are "#questionsanswered" and "#Romy".

218x78mm (72 x 72 DPI)