Open Research Online

Collecting and preserving digital comics

Jen Aggleton
7th September 2017

Collecting and preserving digital comics

Contents

1. Introduction

This report contains a summary of the work conducted for a PhD placement between 12th June 2017 and 8th September 2017. The placement remit was to investigate the collection and preservation of digital comics published in the UK and Ireland, in line with The Legal Deposit Libraries (Non-Print Works) Regulations 2013 (Department for Culture, Media and Sport, 2013). The placement also involved the establishment of a special collection of webcomics as part of the British Library Web Archive. The placement was funded by the Economic and Social Research Council.

2. Types of digital comic

Following a survey of UK and Ireland digital comics publications, conducted through web searches, the following types of digital comic were identified:

2.1 Webcomics

These are digital-born comics published directly onto the internet. They are largely free, though some subscription sites exist (e.g. Aces Weekly, http://www.acesweekly.co.uk/). Some are supported through the Patreon crowd-funding platform, whilst others are non-commercial. Many webcomics have dedicated sites, though some publish through comics hosting sites such as Webtoons, and others publish on social media sites, predominantly Tumblr. Webcomics generally fall into four categories which lead to different preservation requirements:

- Print-style webcomics making no explicit use of digital features
- Game or motion comics which use Flash
- Hyperlink comics
- Comics making use of animation through scrolling features

2.2 E-reader files

These are generally print comics which have been additionally published as digital files in order to be read on an e-reader. Common file formats include PDF, EPUB/EPUB3, CBZ, CBT and CBR. Less common file formats include DJVU and TAR. Currently, CBZ appears to be considered the most effective file format for e-reader versions of comics (http://chunkyreader.com/). These e-reader comics

files are usually identical in content and format to the print versions of the comics.

2.3 Platforms for comic apps

Several platforms exist which host comic book apps. Apps are designed specifically for these platforms, and many of the platforms publish their own apps as well as hosting publications from other creators and publishers. The major platforms used by UK and Ireland based comics publishers are:

2.3.1 ComiXology

(https://www.comixology.co.uk/)

ComiXology holds by far the largest market share for the publication of digital comics apps (Johnston, 2017). The comiXology app mediates the reading experience through its 'Guided View' system, presenting the reader with one panel at a time, rather than showing the whole page (or double page). Whilst this is an example of digital mediation which is different from the experience of reading print, the mediation is conducted by the platform, rather than the individual publication, and all publications are mediated in largely the same way. There are some examples of apps hosted by comiXology which do contain additional digital features, however the majority are largely digitised versions of print equivalents, or are constructed directly for the app as though they had print equivalents, without making use of specific digital features.

2.3.2 Issuu

(https://issuu.com/)

Issuu provides a publishing service for a wide variety of digital content, including comics. It functions in many ways as a self-publishing tool for digital content and has both free and premium options for publication. The Issuu reader provides a close digital equivalent to the reading experience of a print item, with full double page views and a page turn effect, allowing the reader to see the item in the same way they would see a print item. Many of the digital comics published on Issuu have print equivalents, and Issuu does not allow for digital features such as hyperlinks or audio.

2.3.3 Madefire

(http://www.madefire.com/)

Madefire provide a platform for both reading and creating 'motion books' or 'motion comics': digital comics which incorporate multimodal aspects such as animation and sound. Madefire publish their own comics (whilst the founders are mostly British, the company is based in America, so they do not fall under legal deposit) which can be read through the Madefire reader (available on iOS, Android, and for Windows 10). Madefire also provide a platform through their social media page on deviantart (https://madefire.deviantart.com/) for anyone to publish motion books using the Madefire tool for free (these publications include works by UK and Ireland based creators). Occasionally motion books published by users on deviantart are added to the Madefire reader store, if the publication team believes them to be of sufficient quality. Whilst the company has adapted several print comics titles into motion book format, the motion book versions are significantly different from the print versions due to the way they incorporate digital features.

2.3.4 Sequential

(http://www.sequential.cc/)

Sequential is an iOS only comics reading platform app which publishes digital graphic novels designed specifically for the app. They work with a number of publishers including UK and Ireland based publishing houses such as SelfMadeHero. Sequential titles are heavily curated and tend to include comics more towards the 'literature' than 'popular' end of the spectrum. The app allows for reading both in panel by panel mode (similar to comiXology's 'guided view') and full/double page mode. The app also incorporates basic digital features such as zoom options and an index view. There is the option for additional digital material to be included, such as audio commentaries, though not all publications include these additional materials. As with comiXology, many of the publications are straightforward digitisations of print publications, with the platform being the significant factor in the reading experience rather than the publications themselves.

2.3.5 2000 AD

(http://2000ad.com/)

2000 AD is the digital comics app platform of the long-running British comics publisher 2000 AD (owned by Rebellion). The app is available on both iOS and Android, and has the distinction amongst digital comics platform apps of providing DRM-free access to digital versions of its publications through the Rebellion ID system, including options to download different file formats (such as PDF and CBZ) from the 2000 AD web store (https://shop.2000ad.com/). The publications are digitised versions of print comics, though there are some subscription options which are only available digitally.

2.3.6 Other comics platform apps

Other platforms for reading digital comics apps exist but have not been included in this list due to either exclusively hosting non-UK and Ireland regional specific publications (e.g. Marvel Unlimited), or because they function as file readers and do not provide direct access to publications (e.g. Chunky).

2.4 Comic book apps

These are titles produced as an app rather than being hosted through a larger app platform such as comiXology. They may be published as one-off titles, or through a comics publisher such as ROK Comics. These apps vary in the amount of digital interactivity they provide. Due to the difficulty of searching app stores by location, it is very hard to identify one-off comics apps which fall under legal deposit.

2.5 Virtual Reality

Virtual reality (VR) comics are just beginning to emerge onto the market, with Madefire collaborating with Oculus to create the first VR comics (https://www.oculus.com/experiences/gear-vr/1214584621926197/). At present no publications are available, however this is likely to be a growing field and should be taken into consideration when looking to preserve digital comics.

3. Multiple versions of comics

Many digital comics are adaptations of print-born comics, a lot of which are identical to the print original except in physical manifestation. For app platform-hosted adaptations of print-born comics the difference between the print and digital version is often based within the mediation of the app platform rather

than the publication itself. For these items it may not be worth collecting both the print and digital equivalents.

Other digital adaptations of print comics do contain elements which are significantly different from the print equivalent, such as the incorporation of hyperlinks, animation, or sound. As such, they represent a 'new expression', or a 'new work' under FRBR guidelines (Tillett, 2003), and in these cases it may be worth collecting both versions.

Family of Works
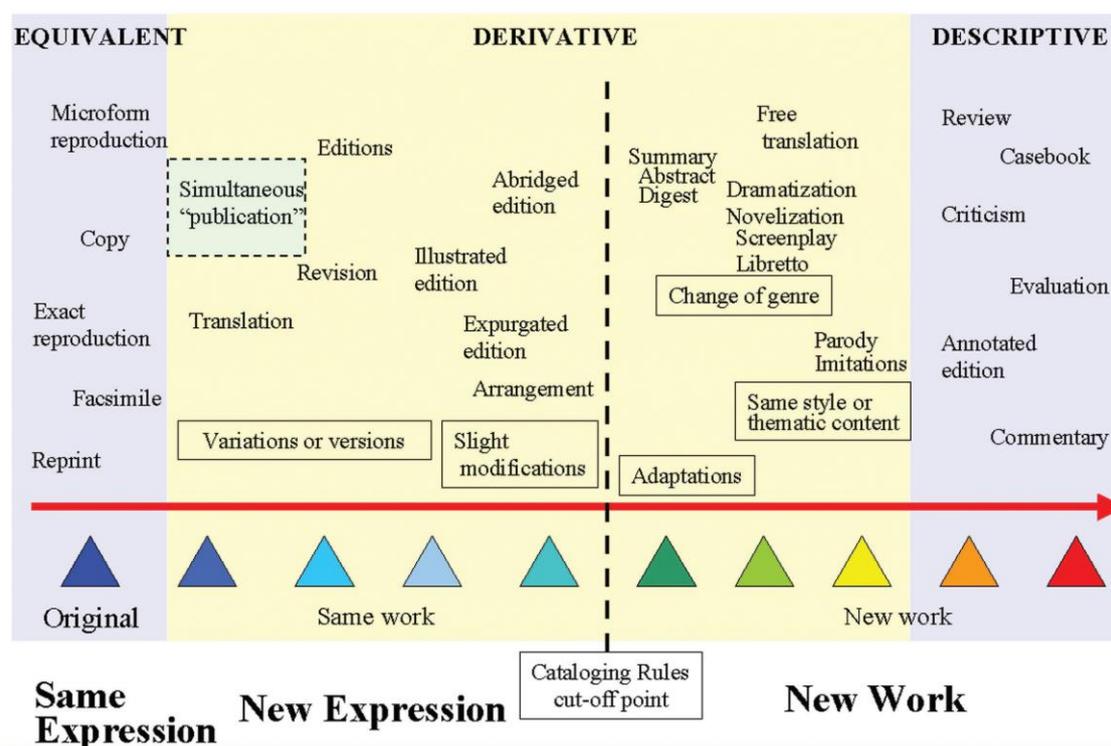
Figure 1: Tillett, B. (2001). "Bibliographic Relationships" in C.A. Bean and R. Green (Eds.) *Relationships in the Organization of Knowledge*, Boston: Kluwer Academic Publishers, p.23

For digital-born content where there is a print equivalent, such as is the case with many webcomics, the websites generally contain considerable amounts of paratext, often including additional information from the author and frequently including feedback and conversations from readers, as in this example:

Figure 2: Padua, S. (2009). [webcomic series]. *Lovelace and Babbage vs The Economy*, http://sydneypadua.com/2dgoggles/lovelace-and-babbage-vs-the-economy/, reproduced with permission of the author

This paratext is an integral part of the reader's experience of the comic. In these cases the print version of the comic is generally less comprehensive than the digital version, and as such it may be preferable to either collect both versions, or to only collect the digital version.

For both print-born and digital-born content, scholars may be interested not only in the content of the item but also its originally intended production and publication modes (Brienza, 2010), and therefore it is recommended that the original publication format is collected in all instances, regardless of whether another format is also collected.

4. Preservation tools and requirements

4.1 Webcomics

The preservation requirements for webcomics vary based on the extent of dynamic content on the web site. Two different web archiving tools were tested: Heritrix and Web Recorder.

4.1.1 Heritrix

For webcomics with low amounts of dynamic content, the Heritrix web crawler is able to effectively capture the sites. Advantages of Heritrix include the ability to set regular automated crawl schedules, so that archived content can be updated without the need for additional staff time. Heritrix is also already integrated into the British Library web archiving system making it the preferable tool where it remains effective.

Heritrix can be less effective at capturing sites which use Flash, incorporate video, animation, audio, or have a large number of nodes. For these sites Heritrix can usually capture the text around the non-captured content, and a limited number of supplementary pages through hyperlinks.

4.1.2 Web Recorder

Web recorder is able to effectively capture most websites, including those which contain dynamic content and use Flash (up to a certain age, a test of hyperlink comics published on e-merl.com showed that web recorder was effective on sites published from 2007 onwards, but not those published earlier). As long as the archivist has clicked on every single hyperlink whilst running web recorder, the playback interface works in the same way as the live site, so readers can gain an accurate idea of how the site both looks and works.

The drawbacks of web recorder are that it requires considerably more staff time than Heritrix, and cannot be automated to update the capture on a regular basis. At present it also cannot be integrated into the web archive. However, if a way could be found to integrate web recorder into the archive then web recorder would be a useful tool to utilise on sites which Heritrix fails to capture accurately. Because of the additional staff time required when using web recorder, it may be most useful to use it as a secondary tool when the crawls done on Heritrix do not meet quality assurance standards.

4.2 Apps

During the course of the placement a satisfactory method of collecting and preserving comic book apps was not identified, and it may be the case that a method of collecting apps needs to be specifically developed. The collection and preservation of apps was explored through the examination of a number of apps

as well as several professional conversations with app publishers and developers. Through these investigations the following issues were identified:

4.2.1 Platform

Most comic book apps are hosted on iOS, with some also being hosted on Android and occasionally Windows. This is due to iOS having greater capability for apps, as well as more rigorous standards, meaning that if an app passes the iOS standards it will also pass the Android standards. For collection purposes, decisions will need to be made as to which platform version(s) are collected. In order to run the apps, a platform will also be required, either through preserving the original platform as well as the app, or by creating an emulator or software programme which can run the apps.

4.2.2 Platform Version

iOS and Android platforms are updated frequently, and when new versions are released comic book apps are either updated or lost. This has two major implications:

1) Preservation choice of whether to collect a specific version of an app or all versions of an app
2) Apps can have a very short availability and many are being lost altogether due to the current inability to collect and preserve them

4.2.3 Environment

As well as requiring a platform to run on, some apps are linked into the hardware environment of the device they are running on, utilising features such as the camera. In preserving these apps an assessment will need to be made of how important these environmental requirements are to the collection and preservation of the apps. It may be possible to create a hardware repository to address this problem, but it would depend on issues such as battery life and storage. Creating a link with an institution which already preserves hardware, such as the Centre for Computing History in Cambridge, may provide an alternative to creating an in-house repository. Alternatively, it may be possible to find an alternative solution to these aspects of the apps through software emulation.

4.2.4 DRM

Many apps are subject to DRM (digital rights management), which means that when apps are downloaded they provide the reader with access to, rather than ownership of, the file. Many files are stored on the cloud, and are not transferable between devices. File ownership for the purposes of collection would therefore need to be negotiated directly with app publishers where DRM exists.

4.2.5 Code

As well as collecting the apps themselves, it may be desirable to collect the code for the apps. Whilst this code is not a publication and would therefore not fall under legal deposit, it may be of great interest to current and future researchers. Additionally, if no adequate way of collecting and preserving apps is found, collecting the code may allow the apps to be re-engineered in the future.

4.2.6 Filming app use

The possibility of filming a user interacting with an app as a form of preservation was discussed. Whilst this may give a basic idea of how the app functions, and may be preferable to not preserving apps at all, there are a number of drawbacks to the approach. The primary one is that only one pathway (out of possibly thousands) through the app will be shown, giving a very limited view of the app's capabilities. Additionally, it can be very difficult to tell from a film exactly how an interface is working. This method would also require a great deal of staff time. It would also be very hard to isolate the audio track of the app from any comments or verbal reactions made by the user, and these user responses may present a particular type of response as typical, when in fact users are likely to respond to the same app in many different ways. As such, it is recommended that an alternative method of preservation is sought.

4.3 Virtual Reality

As no VR comics publications have yet been released, it has not been possible to assess the collection requirements for VR comics. However, there are likely to be similar issues with preserving VR comics as there are with comics apps. Additionally, the playback facilities for collected VR comics may be complicated by the need not only for the VR device but also for enough space to use the device safely.

5. Webcomic Archive

5.1 Collection policy

In order to effectively collect and preserve digital comics it was necessary to outline the scope of the collection by creating a working definition of what constitutes a digital comic. This was a problematic task as there is no popular or academic consensus as to what exactly is or is not a comic. After extensive research into current views on both print and digital comics, as well as analysis of various types of digital comics, the following collection policy was developed:

The scope of this collection covers items with the following characteristics:

- The collection item must be published in a digital format
- The collection item must contain a single panel image or series of interdependent images
- The collection item must have a semi-guided reading pathway[1]

In addition, the collection item is likely to contain the following:

- Visible frames
- Iconic symbols such as word balloons
- Hand-written style lettering which may use its visual form to communicate additional meaning

The item must not be:
- Purely moving image
- Purely audio

For contested items, where an item meets these categories but still does not seem to be a comic, it will be judged to be a comic if it self-identifies as such (e.g. a digital picturebook may meet all of these criteria, but self-identifies as a picturebook, not a comic).

Where the item is an adaptation of a print born comic, it must be a new expression of the original, not merely a different manifestation, according to FRBR guidelines: http://www.loc.gov/cds/FRBR.html.

[1] Definition of a semi-guided reading pathway: The reader has autonomy over the time they spend reading any particular aspect of the item, and some agency over the order in which they read the item, especially the visual elements. However reading is also guided in the progression through any language elements, and likely to be guided in the order of movement from one image to another, though this pathway may not always be clear. This excludes items that are purely pictures, as well as items which are purely animation.

These guidelines were chosen in order to allow for the very varied nature of comics, and particularly digital comics which may have multimedia features, whilst still recognising the iconic aspects which many people associate with comics. In addition they provide a way to identify those items which are more difficult to categorise, such as motion comics or game comics, in order to allow for the inclusion of innovative uses of the medium without misrepresenting items as comics which could more accurately be classified as animation, games, or pictures.

5.2 Acquisition methodology

Items for the webcomic collection were identified through a combination of methods. Primarily, known UK webcomics such as *Scary Go Round* by John Allison were added to the collection. This was followed by extensive web searching for comics on the UK domain. A call for nominations was sent to the prominent UK digital comics website *Pipedream Comics* (http://pipedreamcomics.co.uk/), where it was picked up and reblogged by *Down the Tubes* (http://downthetubes.net/), another prominent UK comics website. A blog post describing the placement and calling for submissions was published on the British Library's Web Archive blog, and submission calls were also posted on social media through Facebook and Twitter. Through these methods an initial 42 sites were identified and added to the web archive.

Whilst there are likely to be large numbers of UK and Ireland based comics creators publishing on *Webtoons* (http://www.webtoons.com/en/) and Tumblr (https://www.tumblr.com/), these very large sites have no systematic way of searching for content by location. Identifying content on these sites which falls under the scope of legal deposit was therefore not considered to be an efficient use of time given the limited period of the three-month placement.

5.3   Sociocultural considerations

5.3.1 Issue of authority

Creating a webcomic archive for the British Library has important sociological considerations, especially given the lack of consensus in both academic and popular circles as to what constitutes a comic. Archives and libraries are non-neutral spaces, and the choices made about what is and isn't included in a

collection have the power to construct people's views about any number of subjects (Schwartz & Cook, 2002; Stead, 2013; Steedman, 2001). This is particularly true of collections at the British Library, which, due to its nature as a national institution, is likely to be seen by many as an authority. The very act of collecting digital comics is therefore inherently an act of definition, as if the British Library chooses to include an item in its digital comics collection, users of that collection are likely to accept that the item is a digital comic. This is particularly problematic when considering the long-term nature of the British Library collections. Whilst most library collections are aiming to provide for the immediate needs of the community they serve (Futas, 1995), archives, and the British Library in particular, not only wish to serve the needs of current users, but also of future generations. It is therefore important to consider both what future users might possibly think of as digital comics so as not to exclude items at risk of loss, and also to attempt to accurately represent what contemporaries consider to be digital comics (or not). These factors are compounded by the sociocultural nature of reading. The context in which an item is read is likely to influence the way that it is read (Gee, 2000; Snow & Sweet, 2003), and therefore a comic in the British Library may be more likely to be read as a literary item, whilst a comic on a gallery wall may be more likely to be read as art.

5.3.2 Potential solutions

- Explicit collection guidelines

  One method of addressing the definition-forming potential of the webcomic collection is to be transparent as to what the guidelines for the collection are, and to acknowledge that there is no consensus at to what constitutes a comic. In order to effect this, the description of the webcomics special collection will be set up to include links to a simple outline of the collection guidelines and acquisition methodology, hosted on a British Library blog, and to an academic article which goes into greater depth about the reasoning behind the collection guidelines.

- User indexing

  Another potential solution would be to incorporate user indexing (or tagging) into the collection. This would allow users of the collection to tag items with

particular keywords or phrases (for example: not comic – animation) which would highlight the debate around these contested items, and improve discoverability by providing searchable index terms. User tagging is currently available (though not well flagged and therefore underused) on the main library catalogue, but is not presently integrated into the web archive. The extension of user indexing to the web archive would not only highlight the debate around the collection items, but could also improve the usability of the collection by providing indexing which archivists may not have the time nor expertise to apply.

6. E-reader recommendations

Possible e-readers for digital comics files held by the British Library were investigated. Due to the extremely large number of e-readers available, the selection was narrowed by reading user reviews. Both tablet and PC viewers have been included to allow for a breadth of options, and key features of the readers have been highlighted.

6.1 Issuu

(https://issuu.com/)

As well as acting as a publisher, Issuu can provide an Issuu reader which is able to display a large number of e-reader formats, from books to magazines to comics. The Issuu reader provides a magazine-style reading interface, with double page spreads and page turning animations, which closely resembles the experience of reading a print publication. A video demonstrating how the Issuu reader works can be found here:

https://www.youtube.com/watch?v=Ru6wgHksuQM

The advantage of an Issuu reader is that it could potentially be utilised for reading other digital texts as well as comics, limiting the number of different e-readers required by the library. However, whilst often used for digital comics publications, Issuu does not currently support comics-specific file types such as CBZ, CBT and CBR. Additionally, Issuu is currently set up as a publishing tool rather than a discrete e-reader, and therefore negotiations may have to be undertaken with the developers in order to use it purely as an e-reader.

14

6.2 Chunky Comic Reader

(http://chunkyreader.com/)

Chunky Comic Reader is an iOS only e-reader app designed specifically for comic books. It can read CBZ, CBR, CBT and PDF files. The reader uses auto contrast and upscaling to improve the quality of images. It offers both left-to-right and right-to-left reading options making it suitable for manga as well as comics, and offers both single and double page views.

At present, Chunky reader is only available as an iOS app, and as it is published by an independent developer it may be difficult to negotiate alternative platform versions. Additionally, Chunky Comic Reader detects and crops margins, which alters the view of the files when reading.

6.3 Perfect Viewer

(https://sites.google.com/site/rookiestudio/)

Perfect Viewer is an Android only e-reader app which can be used for novels, pictures or comics. It supports a wide variety of file types including EPUB, PDF, CBR and CBZ. It has a wide range of personalisation options for reading which include reading direction, single or double page images, and the ability to colourise black and white images. A demonstration of Perfect Viewer can be found here:

https://www.youtube.com/watch?v=T2okd2f3t84

Like Issuu, Perfect Viewer is able to host a number of different types of digital texts rather than being comics specific, but has the added advantage of having additional configuration options for comics reading. Perfect Viewer is free to use even within a commercial organisation, but at present is only available on Android platforms, so adaptation to alternative platforms would have to be negotiated with the developer.

6.4 Evince

(https://wiki.gnome.org/Apps/Evince)

Evince is a document viewer for multiple document formats developed for the GNOME desktop. It is available for both Windows and Linux systems. Evince supports multiple file types including comics specific files such as CBR and CBZ. Evince is open source software, and can therefore be altered to suit the library's purposes provided staff time and expertise are available to work on the code.

Being part of the GNOME project, there are also bug reporting facilities and limited user support available.

6.5 ComicRack

(http://comicrack.cyolito.com/)

ComicRack is a comics reader and library management system available for iOS, Android and Windows. It supports a wide variety of file formats including CBZ, CBR, and CBT, and allows for customisable reading options including double page view and choice of reading direction. Additionally, ComicRack provides a wide range of library management options, with the ability to add metadata and indexing tags. ComicRack is free on Windows and has a very small cost for Android and iOs, but as a result the support is limited to user forums.

7. Priorities going forwards

7.1 Webcomic special collection quality assurance

Due to an issue with the web archiving system during the term of the placement, whilst websites were added to the web archive, crawling did not begin until after the placement had finished. As such, it was not possible to undertake QA on these sites during the placement. Due to the difficulties in capturing some of the content of webcomics, it is recommended that staff time is allocated to undertaking this QA if at all possible.

7.2 Preservation of apps and the upcoming iOS update

With the release of iOS 11 in autumn 2017, many iOS apps are at immediate risk of loss. It is therefore recommended that a temporary preservation solution be considered. Possible temporary solutions include:

- Downloading apps onto an iPad running iOS 10 and then disconnecting that iPad from the internet to prevent automatic updating of the operating system
- Contacting app publishers to ask for any documentation around the apps which may allow at least a partial understanding of the apps' content and function

Additionally, it is recommended that further research is undertaken into the preservation of digital comics apps in order to avoid future loss of published material.

8. References

Brienza, C. (2010). Producing comics culture: a sociological approach to the study of comics. *Journal of Graphic Novels and Comics*, *1*(2), 105–119. https://doi.org/10.1080/21504857.2010.528638

Department for Culture, Media and Sport. (2013). The Legal Deposit Libraries (Non-Print Works) Regulations 2013. Retrieved 28 August 2017, from http://www.legislation.gov.uk/uksi/2013/777/contents/made

Futas, E. (Ed.). (1995). *Collection Development Policies and Procedures* (3rd Revised edition edition). Phoenix, Ariz: Greenwood Press.

Gee, J. (2000). Discourse and Sociocultural Studies in Reading. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of Reading Research* (Vol. 3). New York: Routledge.

Johnston, H. (2017). ComiXology and the Future of the Digital Comic Book. *The iJournal: Graduate Student Journal of the Faculty of Information*, *2*(2). Retrieved from http://theijournal.ca/index.php/ijournal/article/download/28126

Schwartz, J. M., & Cook, T. (2002). Archives, records, and power: The making of modern memory. *Archival Science*, *2*(1–2), 1–19. https://doi.org/10.1007/BF02435628

Snow, C. E., & Sweet, A. P. (Eds.). (2003). *Rethinking Reading Comprehension*. New York: Guilford Press.

Stead, L. (2013). Introduction. In C. Smith & L. Stead (Eds.), *The boundaries of the literary archive: reclamation and representation* (pp. 1–12). Surrey: Ashgate Publishing Ltd.

Steedman, C. (2001). *Dust*. Manchester: Manchester University Press.

Tillett, B. (2003). What is FRBR? A conceptual model for the bibliographic universe. Library of Congress. Retrieved from https://www.loc.gov/cds/downloads/FRBR.PDF

9. List of figures

Figure 1: Tillett, B. (2001). "Bibliographic Relationships" in C.A. Bean and R. Green (Eds.) *Relationships in the Organization of Knowledge*, Boston: Kluwer Academic Publishers, p.23

Figure 2: Padua, S. (2009). [webcomic series]. *Lovelace and Babbage vs The Economy*, http://sydneypadua.com/2dgoggles/lovelace-and-babbage-vs-the-economy/, reproduced with permission of the author

10. List of websites

2000 AD, http://2000ad.com/

Aces Weekly, http://www.acesweekly.co.uk/

Chunky Reader, http://chunkyreader.com/

ComicRack, http://comicrack.cyolito.com/

ComiXology, https://www.comixology.co.uk/

Down the Tubes, http://downthetubes.net/

Evince, https://wiki.gnome.org/Apps/Evince

Issuu, https://issuu.com/

Madefire, http://www.madefire.com/

Oculus, https://www.oculus.com/experiences/gear-vr/1214584621926197/

Perfect Viewer, https://sites.google.com/site/rookiestudio/

Pipedream Comics, http://pipedreamcomics.co.uk/

Sequential, http://www.sequential.cc/

Tumblr, https://www.tumblr.com/

Webtoons, http://www.webtoons.com/en/