	Digital Preservation Team	**Project Name:** WP2: File Format Assessment	**Date:** 26/03/2015
		Document Title: TIFF Format Preservation Assessment	**Version:** 1.2

TIFF Format Preservation Assessment

Document History

Date	Version	Author(s)	Circulation
10/12/2014	1.1	Paul Wheatley, Peter May, Maureen Pennock	External
26/03/2015	1.2	Peter May, Akiko Kimura	External

British Library Digital Preservation Team
digitalpreservation@bl.uk

1. Introduction

This document provides a high level, non-collection specific assessment of the TIFF file format with regard to preservation risks and the practicalities of preserving data in this format.

This format assessment is one of a series of assessments carried out by the British Library's Digital Preservation Team. An explanation of criteria used in this assessment is provided in italics below each heading.

1.1. Scope

This document will focus on all versions of TIFF, although with emphasis on the current version 6.0. Issues of both preserving deposited TIFFs and preserving TIFFs created as part of digitisation activities will be considered.

Note that this assessment considers format issues only, and does not explore other factors essential to a preservation planning exercise, such as collection specific characteristics that should always be considered before implementing preservation actions.

1.1 TIFF Summary

TIFF is a bitmap image format developed originally as a more universal alternative to the large numbers of proprietary formats tied to document scanner hardware. It evolved from what was initially a bi-tonal format to increasing colour depths and other functionality in order to support advancing scanner technology. Strong software support across different platforms (and scanning hardware) led to widespread adoption as a master form for storing digitised images. The Wikipedia TIFF page [1] provides a detailed introduction to and description of the format, and a summary of salient links can be found on Just Solve [2].

2. Assessment

2.1 Development Status

A summary of the development history of the format and an indication of its current status

The current copyright owners of the format, Adobe, appear to have focused on Photoshop related formats rather than developments related to TIFF. The FAQ on the libtiff mail list site notes that "while there have been rumours of a new TIFF specification under development within Adobe over the years, it does not appear to have a high priority. Nor is there any public process for specification revision. It is unclear whether a new specification version will be released by Adobe or not" [3]. Regardless of this speculation, the specification itself has not seen a major revision since 1992 (revision 6) [4]. More recent changes have mostly focused on additional extensions for use in predominantly niche functions such as TIFF for Internet Fax [5]. Some, effectively proprietary, extensions have been created by application developers and in some cases are supported only by the related application (see below).

2.2 Adoption and Usage

An impression of how widely used the file format is, with reference to use in other memory organisations and their practical experiences of working with the format

TIFF is widely acknowledged as the de facto standard for master files for digitised still images, where it is used almost universally in Libraries and Archives for this purpose. The Succeed Project assessed existing digitisation recommendations and performed a survey of memory organisations [6]. Both revealed common requirements for usage of TIFF as a master format: "It is the most popular format both in the context of existing recommendations (94% of them indicate TIFF) and Succeed survey results (87% of respondents indicated TIFF)". The widely referenced FADGI "Technical Guidelines for Digitizing Cultural Heritage Materials: Creation of

Raster Image Master Files" [7] describes TIFF as the "'De facto' raster image format used for master files". The American Memory [8] project is an example of large scale digitisation utilising various versions of TIFF as the master format [9].

2.3 Software Support

2.3.1 Rendering Software Support

An overall impression of software support for rendering the format with reference to: typical desktop software; and current support on British Library reading room PCs

Support for TIFF in web browsers is patchy [10]. However support in typical graphics manipulation software such as ImageMagick, GIMP, Photoshop or IrfanView is comprehensive. Open source implementations, such as LibTIFF, are widely used and typically still actively developed, for example LibTIFF has a small but consistent number of commits to its codebase over the last year [11].

Support in application software for some of the more unusual tags introduced beyond "baseline" TIFF 6.0 can be patchy. Wikipedia notes that "when TIFF was introduced its extensibility provoked compatibility problems. The flexibility in encoding gave rise to the joke that TIFF stands for Thousands of Incompatible File Formats". However it also states "TIFF files that strictly use the basic 'tag sets' as defined in TIFF 6.0 along with restricting the compression technology to the methods identified in TIFF 6.0 and are adequately tested and verified by multiple sources for all documents being created can be used for storing documents" [1].

Due to the use of 32bit offsets, TIFF files are limited to a 4GB file size. Although an unofficial standard, the BigTIFF file format specification [12] does provide support for files in excess of this limitation. Wikipedia notes it "was implemented in 2007 in development releases of LibTIFF version 4.0, which was finally released as stable in December 2011". It goes on to say that "support for BigTIFF file formats by applications is limited" [1], however the LoC's BigTIFF format planning page lists a number of libraries and applications which support it, including the MATLAB programming language and environment [13].

Issues

Wikipedia goes on to note that "commonly seen issues encountered in the content and document management industry associated with the use of TIFF files arise when the structures contain proprietary headers, are not properly documented, and/or contain 'wrappers' or other containers around the TIFF datasets, and/or include improper compression technologies, or those compression technologies are not properly implemented". This latter point is illustrated by the number of compression schemes that are cited to have less than "common" support [1].

Discussions on the LibTIFF mail list in 2008 hint at the reasons behind this, despite improving software support over time. One question [14] asks whether it is reasonable to "conclude that most recent libtiff equipped systems (e.g. systems deployed or updated during the last five years) will support tiff/zip?" to which one response was "you can't really conclude anything at all. Yesterday I fixed a bug in my software related to libtiff 3.6.1[1]. People are still using these old versions. Some commercial OSs don't automatically update to the latest version with each major release because they operate on the 'If it ain't broke ...' principle".

As a result of these concerns, memory organisations typically require some constraints on TIFF files generated in digitisation activities, avoiding more unusual tags and compression schemes beyond TIFF v6.0. For example the JISC Digital Media guidelines state "When we mention or specify TIFF, it is important to realise that the TIFF file format comes in a range of

---

[1] For reference, LibTIFF v3.6.1 is pre-2007.

types, supporting different functionality, such as multipages and even a choice of compressions including JPEG. So when we specify TIFF for archival purposes we always mean an uncompressed Baseline TIFF v6 with Intel byte order" [15]. The requirement for "uncompressed" is common but not universal and seems to relate back to LZW patent concerns (see section 2.8 - Legal Issues) and lack of support for some of the more unusual compression schemes.

Wikipedia notes "TIFF readers must be prepared for multiple/multi-page images (subfiles) per TIFF file although they are not required to do anything with images after the first one"; this can be the cause of problems for some users trying to use software lacking rendering support beyond the first page [16; 17]. Worse still, some software applications reportedly have trouble simply viewing a multipage TIFF at all [18; 19]. The multi-page functionality is therefore best avoided when creating a TIFF.

2.3.2 Preservation Software Support

An impression of the availability and effectiveness of software for managing and preserving instances of the file format

Format identification

JHOVE identifies the format version based on the presence of tags that relate to versions 4, 5 or 6 [20]. Unfortunately, identification achieved through validating a file can often fail due to validation errors. JHOVE2 separated out the format identification and characterisation aspects of JHOVE, relying on signature-based format identification using DROID instead. DROID however, does not distinguish between versions of TIFF; previously distinct PRONOM PUIDs for 3, 4, 5 and 6 were deprecated in favour of a single TIFF puid. This attracted some controversy [21] as it seemed to reflect the situation from a DROID perspective (where the file format magic is identical between each version) rather than the reality where distinguishing between versions requires more in depth file parsing (a function that DROID does not perform).

Validation and Detecting Preservation Risks

JHOVE provides validation support for TIFF files, but there is little in the way of published evidence as to the value of this validation, nor examples of problems that this validation might catch. In fact, contrary to this, SPRUCE Mashup attendees identified a number of preservation cases where broken or badly formed TIFF files were either not picked up by JHOVE or were present in TIFF flavours not supported by JHOVE [22]. In particular, this included a case of valid and well-formed TIFF's that rendered incorrectly (or, for some images, not at all) [23]. These were subsequently identified to be malformed TIFFs with mismatching resolution and image data, as described in the solution page [24]. JHOVE2 also provides a TIFF validation module for versions 4, 5 and 6 [25].

Conformance Checking

Although JHOVE distinguishes between TIFF versions 4, 5 and 6, within the v6.0 space it does not (reasonably from the point of view of JHOVE) distinguish between conservative use of TIFF tags - as digitisation guidelines typically require - and some of the more exotic tags that preservers may wish to avoid. The National Digital Newspaper Program (NDNP) Technical Guidelines [26] documented the use of simple validation of digitised content and conformance to a simple TIFF profile, with the aid of a custom validation library "that 'wraps' JHOVE and extends JHOVE's existing TIFF, PDF, and JPEG2000 modules with the NDNP-specific validation rules" enabling the NDNP extension to validate more than just conformance to the TIFF specification, for example validating that a "TIFF file is uncompressed, 8-bit grayscale, and contains the microfilm reel number in tag 269" [27]. The work utilised Schematron as a rule checker[2].

---

[2] Schematron was also used successfully with Jpylyzer for checking JP2 conformance to a policy driven profile [28].

KOST-Val [29] supports conformance checking of TIFF to a policy driven profile. It also wraps JHOVE, whose results are compared against a profile for compliance. This functionality was added to KOST-Val after KOST (coordination centre for long-term archiving of electronic documents) [30] discovered TIFF files that did not render in many current graphics applications.

Metadata Extraction

Apache Tika [31] and Exiftool [32] provide support for extracting image metadata. As is the case with many bitmap graphics formats, Exiftool [33] typically performs useful functions to support trouble shooting with TIFF files [24]. AsTiffTagViewer [34] is a simple windows based viewer, revealing which tags are used by a particular TIFF file.

Migration

Graphics manipulation software typically supports format migration to and from TIFF, as do libraries such as LibTIFF.

2.4 Documentation and Guidance

An indication of the availability of practical documentation or guidance with specific reference to the facilitation of any recommended actions.

Documentation for TIFF includes the published format specification [4] (plus two supplements [35; 36]) and an array of web pages that describe the structure of the format, usually with respect to implementation of software to read or write TIFFs. A comprehensive listing of references can be found on the LoC's TIFF page [9].

2.5 Complexity

An impression of the complexity of the format with respect to the impact this is likely to have on the British Library managing or working with content in this format. What level of expertise in the format is required to have confidence in management and preservation?

Despite seeing the creation of an array of extensions that were added post v6.0 baseline, TIFF remains a reasonably straightforward format to work with. Certainly in relative terms to JPEG2000[3], which requires in depth understanding of the complex wavelet-based image compression schemes, TIFF is considerably easier to work with. Digitisation units within the British Library (such as the Digitisation Studios and the International Dunhuang Project) have in depth and valuable experience of working with TIFF.

2.6 Embedded or Attached Content

The potential for embedding or attaching files of similar or different formats, and the likely implications of this

Extensions beyond baseline TIFF v6.0 effectively provide the potential to embed other formats, for example as ZIP compressed images or JP2 (implemented in Leadtools [37]). Software support for these features (as noted above) can be patchy. While avoiding these issues in digitised files is straightforward, a deposited collection could potentially include them.

2.7 External Dependencies

An indication of the possibility of content external to an instance of the file format that is complimentary or even essential to the intellectual content of the instance.

None known.

---

[3] The Succeed Project describes JP2 as the main alternative to TIFF for the storage of digitised masters and calls it an "emerging format rather than a well-established one".

2.8 Legal Issues

Legal impediments to the use, management or preservation of instances of the file format

LZW compression was covered by patents issued in the 1980s. Unisys began to enforce licensing for LZW in the GIF file format in the 1990s, but received criticism and subsequently changed tack [38]. LZW patents in the US and Europe subsequently expired in 2003 and 2004 respectively [39], removing the need for restrictions on the use of LZW compression in digitisation guidelines (at least for these legal reasons).

2.9 Technical Protection Mechanisms

Encryption, Digital Rights Management and any other technical mechanisms that might restrict usage, management or preservation of instances of the file format

Not applicable.

2.10 Other Preservation Risks

Other evidence based preservation risks, noting that many known preservation risks are format specific and do not easily fit under any of the sustainability factors above.

None known.

2.11 Preservation Risk Summary

A summary of preservation risks and recommended actions (where possible)

TIFF remains reasonably well suited to the simple task of storing digitised masters despite lacking features that have developed with other raster formats to enable enhanced graphics applications (such as progressive transmission and multiple resolutions offered by JPEG2000 files). Although there are preservation concerns with less well supported features, such as some compression schemes, the baseline tags are well supported by software and well tested by many users both within and beyond the GLAM[4] sector.

Risks:

- **Lack of application software support for some TIFF tags**
    - o Deposited files may contain less common TIFF tags that could cause rendering issues. Identifying these tags could be useful to avoid future rendering issues although these are expected to be uncommon.

- **Multi-page TIFFs may cause problems with some software applications**
    - o Software may have problems rendering a multipage TIFF, or (in accordance with baseline TIFF v6.0) only render the first image of a multipage TIFF.

- **Lack of software for detecting corrupted/broken files**
    - o Examples of broken TIFF files seen at SPRUCE events were typically not detectable by JHOVE despite some files being classed as valid and well-formed.

- **TIFFs are limited to 4GB file sizes**
    - o The default TIFF specification is unsuitable for images which produce files larger than 4GB. BigTIFF is a variant on the TIFF specification which overcomes this limitation.

---

[4]   Galleries, Libraries, Archives and Museums.

3. Recommendations for Action

Recommended actions in usage and handling of the format. Recommend actions in the support or development of software applications that provide, or have the potential to provide, significant risk mitigation for the format. Note that these recommendations do not take into account other requirements such as those driven by specific British Library collections, or non-preservation issues such as resourcing.

Implementation of a TIFF parser/validator/profile conformance checker of a similar form to Jpylyzer [40] would be useful in performing assessments of trial runs in new digitisation projects, and allow automated checking of subsequent production runs to the same standards. Detection of poorly supported TIFF extensions would also enable problem content in deposited collections to be identified. Further investigation and/or collaboration with Wellcome Trust (interested in developing a "TIFFylyzer"), the developers of KOST-Val (see above) and the Open Preservation Foundation (new stewards of JHove [41]) should be explored. The PREFORMA Project [42] is also developing TIFF validation software and may provide a useful solution.

Handling Recommendations

- Consider placing restrictions on the TIFF profile, such as:
  o TIFF v6.0 "baseline"
  o 1 image per file
  o LZW compression5

- For deposited collections:
  o Identify format
  o (If a suitable software solution becomes available) Validate to:
    ▪ Identify damaged/badly formed files
    ▪ Identify potentially risky tags
  o Manually check effective rendering of appropriate sample

- For digitised collections:
  o Perform pixel-wise comparison with source image using appropriate software
  o Verify colour table and resolution metadata has been retained
  o (If a suitable software solution becomes available) Validate to:
    ▪ Identify damaged/badly formed files
    ▪ Check conformance to selected profile (e.g. TIFF v6.0 "baseline", 1 image per file, LZW compression)
  o Manually check effective rendering of appropriate sample

Knowledge Recommendations

- To help avoid rendering issues caused by less common TIFF tags:
  o Improve understanding of which extensions are likely to cause rendering issues, with specific reference to extensions that are not supported by LibTIFF.
  o Build up a knowledge base of example files and test cases

Software Recommendations

- Evaluate KOST-Val
- Contribute to development and maintenance of JHove
- Develop a TIFF validator/conformance checker.
  o KOST-Val and/or the JHOVE TIFF module should be considered as a starting point.
- Monitor progress of work from the PREFORMA Project.

---

5 Pressures of storage costs could necessitate migrating uncompressed images to compressed formats. LZW compression could therefore be seen as desirable in meeting this need without incurring any preservation risks as LZW patents have expired and application software support is common.

Monitoring Recommendations

The preservation risks faced by TIFF are unlikely to change rapidly and so review of this document should be considered a low priority. However, awareness of new software developments that may provide useful TIFF validation mechanisms should be maintained.

4. References

1. Tagged Image File Format. *Wikipedia.* [Online] [Cited: 25 February 2015.] http://en.wikipedia.org/wiki/Tagged_Image_File_Format .

2. TIFF. *Just Solve the File Format Problem.* [Online] [Cited: 25 February 2015.] http://fileformats.archiveteam.org/wiki/TIFF.

3. TIFF, Tag Image File Format, FAQ (Question 5). *Aware Systems.* [Online] [Cited: 25 February 2015.] http://www.awaresystems.be/imaging/tiff/faq.html.

4. TIFF Revision 6.0. [Online] 3 June 1992. https://partners.adobe.com/public/developer/en/tiff/TIFF6.pdf.

5. RFC 3949: File Format for Internet Fax. [Online] February 2005. http://tools.ietf.org/html/rfc3949.

6. D4.1 Recommendations for metadata and data formats for online availability and long-term preservation. [Online] 2014. http://www.succeed-project.eu/sites/default/files/deliverables/Succeed_600555_WP4_D4.1_RecommendationsOn FormatsAndStandards_v1.1.pdf.

7. Technical Guidelines for Digitizing Cultural Heritage Materials: Creation of Raster Image Master Files. *The Federal Agencies Digitization Guidelines Initiatives.* [Online] 2010. http://www.digitizationguidelines.gov/guidelines/FADGI_Still_Image-Tech_Guidelines_2010-08-24.pdf.

8. About American Memory. *The Library of Congress.* [Online] [Cited: 25 February 2015.] http://memory.loc.gov/ammem/about/index.html.

9. Sustainability of Digital Formats, TIFF Revision 6.0. *The Library of Congress.* [Online] [Cited: 25 February 2015.] http://www.digitalpreservation.gov/formats/fdd/fdd000022.shtml#local.

10. Comparison of web browsers: Image format support. *Wikipedia.* [Online] [Cited: 25 February 2015.] http://en.wikipedia.org/wiki/Comparison_of_web_browsers#Image_format_support.

11. LibTIFF: Project summary. *Black Duck Open Hub.* [Online] [Cited: 25 February 2015.] https://www.openhub.net/p/8684.

12. The BigTIFF File Format Proposal. *Aware Systems.* [Online] [Cited: 25 March 2015.] http://www.awaresystems.be/imaging/tiff/bigtiff.html.

13. BigTIFF. *Sustainability of Digital Formats Planning for Library of Congress Collections.* [Online] Library of Congress, 25 January 2015. [Cited: 25 March 2015.] http://www.digitalpreservation.gov/formats/fdd/fdd000328.shtml.

14. TIFF and LibTIFF Mailing List Archive. *AWARE [Systems].* [Online] 6 February 2008. [Cited: 20 March 2015.] http://www.asmail.be/msg0054850786.html.

15. Choosing a File Format for Digital Still Images. *JISC Digital Media.* [Online] [Cited: 20 March 2015.] http://web.archive.org/web/20150112161927/http://www.jiscdigitalmedia.ac.uk/guide/choosing-a-file-format-for-digital-still-images/.

16. How to open multi-page tiff in Photoshop CS. *Adobe Communities.* [Online] 13 September 2011. [Cited: 23 March 2015.] https://forums.adobe.com/thread/902232?tstart=0.

17. Can't OPEN Multi Page TIF. *OpenOffice Forum.* [Online] 2 February 2008. [Cited: 25 February 2015.] https://forum.openoffice.org/en/forum/viewtopic.php?f=11&t=5350.

18. TIFF file display issues. *Microsoft Support.* [Online] [Cited: 25 February 2015.] http://support.microsoft.com/kb/2635500.

19. Probles with multi-page .tiff files on XP using IE7. *Microsoft Community.* [Online] 18 March 2011. [Cited: 25 February 2015.] http://answers.microsoft.com/en-us/windows/forum/windows_xp-pictures/problems-with-multi-page-tiff-files-on-xp-using/6c29f8e3-8c51-e011-8dfc-68b599b31bf5.

20. JHOVE TIFF-hul Module: Profiles. [Online] [Cited: 25 February 2015.] http://jhove.sourceforge.net/tiff-hul.html#profiles.

21. **Gattuso, Jay.** FMT 7,8,9,10. *OPF Blog.* [Online] 28 August 2011. http://openpreservation.org/knowledge/blogs/2011/08/28/fmt-78910/.

22. Digital Preservation and Data Curation Requirements and Solutions. *OPF Wiki.* [Online] [Cited: 26 March 2015.] http://wiki.opf-labs.org/display/REQ/Digital+Preservation+and+Data+Curation+Requirements+and+Solutions.

23. **Amato, Maurice de Rooij and Andrew.** SPRUCE Senario: Valid and Well Formed TIFFs with scanline corruption. *OPF Wiki.* [Online] September 2012. http://wiki.opf-labs.org/display/SPR/Valid+and+well-formed+TIFF%27s+with+scanline+corruption.

24. **Amato, Andrew.** SPUCE Solution: Solving TIFF malfunction with Exitrool. *OPF Wiki.* [Online] September 2012. http://wiki.opf-labs.org/display/SPR/Solving+TIFF+malformation+using+exiftool.

25. JHOVE2. *Community Owned digital Preservation Tool Registry (COPTR) .* [Online] [Cited: 25 February 2015.] http://coptr.digipres.org/JHOVE2.

26. National Digital Newspaper Program. *The Library of Congress.* [Online] [Cited: 20 March 2015.] http://www.loc.gov/ndnp/guidelines/NDNP_201517TechNotes.pdf.

27. **Littman, Justin.** A Technical Approach and Distributed Model for Validation of Digital Objects. *D-Lib Magazine, Volume 12, Number 5.* [Online] May 2006. http://www.dlib.org/dlib/may06/littman/05littman.html.

28. **Knijff, Johan van der.** Automated assessment of JP2 against a technical profile. *OPF Blog.* [Online] 4 September 2012. [Cited: 20 March 2015.] http://openpreservation.org/knowledge/blogs/2012/09/04/automated-assessment-jp2-against-technical-profile/.

29. KOST-val. *Community Owned digital Preservation Tool Registry (COPTR).* [Online] [Cited: 25 February 2015.] http://coptr.digipres.org/KOST-Val.

30. KOST-ceco. *Community Owned digital Preservation Tool Registry (COPTR).* [Online] [Cited: 25 February 2015.] http://kost-ceco.ch/cms/.

31. XMP Exif TIFF schema. *Apache Tika.* [Online] [Cited: 25 February 2015.] http://tika.apache.org/1.2/api/org/apache/tika/metadata/TIFF.html.

32. *ExifTool by Phil Harvey.* [Online] [Cited: 26 March 2015.] http://www.sno.phy.queensu.ca/~phil/exiftool/.

33. ExifTool. *Wikipedia.* [Online] [Cited: 25 February 2015.] http://en.wikipedia.org/wiki/ExifTool.

34. AsTiffTagViewer. *Community Owned digital Preservation Tool Registry (COPTR).* [Online] [Cited: 25 February 2015.] http://coptr.digipres.org/AsTiffTagViewer.

35. Adobe PageMaker(R) 6.0 TIFF Technical Notes. *Adobe.* [Online] 14 September 1995. [Cited: 26 March 2015.] http://partners.adobe.com/public/developer/en/tiff/TIFFPM6.pdf.

36. Adobe Photoshop(R) TIFF Technical Notes. *Adobe.* [Online] 22 March 2002. [Cited: 26 March 2015.] http://partners.adobe.com/public/developer/en/tiff/TIFFphotoshop.pdf.

37. TIFF Format SDK. *Leadtools.* [Online] [Cited: 25 February 2015.] https://www.leadtools.com/sdk/formats/tiff.htm.

38. GIF. *Wikipedia.* [Online] [Cited: 26 March 2015.] http://en.wikipedia.org/wiki/GIF#Unisys_and_LZW_patent_enforcement.

39. LZW Patent Information. *Unisys.* [Online] 02 June 2009. [Cited: 26 March 2015.] https://web.archive.org/web/20090602212118/http://www.unisys.com/about__unisys/lzw.

40. *jpylyzer.* [Online] [Cited: 25 February 2015.] http://jpylyzer.openpreservation.org/.

41. JHOVE. *Open Preservation Foundation.* [Online] 2015. [Cited: 26 March 2015.] http://openpreservation.org/technology/products/jhove/.

42. *PREFORMA Project.* [Online] [Cited: 25 March 2015.] http://www.preforma-project.eu/.