

ISO/IEC JTC 1/SC 34/JWG 7

Joint JTC 1/SC 34-TC 46/SC 4-IEC/TC 100/TA 10 WG: EPUB

Convenorship: KATS (Korea, Republic of)

Document type:	Disposition of Comments Report
Title:	Disposition of comments for SC34 N2482 and N2484 Summary of Voting on ISO/IEC PDTS 22424-1 EPUB 3 Preservation - Part 1: Principles
Status:	This document is a disposition of comments for mainly two summaries of voting came from JTC 1/SC 34 and ISO TC 46/SC 4. All comments are combined into single document for efficient resolution process.
Date of document:	2018-05-21
Source:	Project editor (Juha Hakala)
Expected action:	INFO
No. of pages:	10
Email of convenor:	samoh21@gmail.com , zzosang@gmail.com
Committee URL:	https://isotc.iso.org/livelink/livelink/open/jtc1sc34jwg7

0	1	2	3	4	5	6	7	8	9
#	NB ¹	Line number (e.g. 17)	Clause/ Subclause (e.g. 3.1)	Paragraph/ Figure/ Table/ (e.g. Table 1)	Type of comment ²	Comment (justification for change) by the NB	Proposed change by the NB	Proposed Disposition of Comments	Disposition of Comments
1	GB 01 (SC34)				Ge	The scope of the Draft Technical Specification is insufficiently clear in the drafts as presented. The scope statements are insufficiently aligned with the contents of each document.		The scope statement has been edited to make it clearer.	Accepted. As amended. <i>Note: see proposed disposition of comments.</i>
2	GB 02 (SC34)				Ge	The drafts as presented do not make a clear distinction between EPUB profile requirements and packaging needs.	The Technical Specification should provide separate profiles for EPUB and METS, preferably in separate Parts.	Future version of this specification may be divided as suggested. Compared with e.g. PDF/A, this document is much less technical, and it is possible that the EPUB/A specification, if and when created by the EPUB community, may be more detailed than this standard.	Accepted. As amended. <i>Note: see proposed disposition of comments.</i>
3	GB 03 (SC34)				Ge	The Draft Technical Specification contains too much discussion of general digital preservation issues.	The text should focus on the issues and challenges of EPUB preservation.	The working group preparing these documents prefers to provide some introductory information about the topic. However, some of this discussion has been deleted from Part 2.	Acknowledged. No action. <i>Note: see proposed disposition of comments.</i>
4	GB 04 (SC34)				Ge	The Draft Technical Specification relies too heavily on theoretical arguments that do not always align well with the		The OAIS producer does not need to be the producer / publisher of the	Acknowledged. No action.

1 **MB** = Member body / **NC** = National Committee (enter the ISO 3166 two-letter country code, e.g. CN for China; comments from the ISO/CS editing unit are identified by **)

2 **Type of comment:** **ge** = general **te** = technical **ed** = editorial

0	1	2	3	4	5	6	7	8	9
#	NB ¹	Line number (e.g. 17)	Clause/ Subclause (e.g. 3.1)	Paragraph/ Figure/ Table/ (e.g. Table 1)	Type of comment ²	Comment (justification for change) by the NB	Proposed change by the NB	Proposed Disposition of Comments	Disposition of Comments
						experience of practitioners of EPUB preservation. As an example of this, the text makes false assumptions about who is responsible for preparing METS SIP packages: this is not always the producer of the e-book, but frequently the preserving institution.		ebook. In practice, it can be the preserving institution (OAIS archive). The terminology of the draft TS has been clarified in this respect. If the producer of the e-book does not take any responsibility of preservation by e.g. calculating checksums, there is no guarantee that preservation will succeed (since the document may be corrupted during transfer to the archive).	<i>Note: see proposed disposition of comments.</i>
5	GB 05 (SC34)				Ge	The proposed approach to the encoding of “unarchivable” data is probably out of scope, as it depends upon institutional policies.		When EPUB is used, it is possible that SIP contains foreign resources in unarchivable format. The specification has to explain how such situation can be handled during ingest. If there is no archivable fallback, the ingest process will fail.	Acknowledged. No action. <i>Note: see proposed disposition of comments.</i>
6	GB 06 (SC34)				Ge	The attribution of issues documented in Annex A of Part 1 is unclear: material that originally appeared in a British	All sources should be clearly attributed.	OK.	Accepted. As amended.

1 **MB** = Member body / **NC** = National Committee (enter the ISO 3166 two-letter country code, e.g. CN for China; comments from the ISO/CS editing unit are identified by **)

2 **Type of comment:** **ge** = general **te** = technical **ed** = editorial

0	1	2	3	4	5	6	7	8	9
#	NB ¹	Line number (e.g. 17)	Clause/ Subclause (e.g. 3.1)	Paragraph/ Figure/ Table/ (e.g. Table 1)	Type of commen ²	Comment (justification for change) by the NB	Proposed change by the NB	Proposed Disposition of Comments	Disposition of Comments
						Library report is interspersed with comments from other sources.			
7	GB 07 (SC34)				Ge	The text needs significant editorial work to remove repetitions, unhelpful elaborations, broken links, ungrammatical phrases etc.		Broken links and grammar issues can be fixed. Unhelpful elaborations, unless specified, cannot be located, since text regarded as unhelpful may be seen as useful by other readers.	Accepted. As amended. <i>Note: see proposed disposition of comments.</i>
8	US 01 (TC46)				Ge	EPUB 3.1 is referenced, but to date Epubcheck does not support EPUB 3.1. There are also some issues that the EPUB 3 maintenance Community Group in the W3C are considering that would make EPUB 3.1 inappropriate for archiving. The most important item is if the Community Group decides to withdraw EPUB 3.1 and issue a revision to EPUB 3.0.1, which eliminates some issues that make EPUB 3.1 difficult to accept broadly, namely that this spec, when adopted, makes earlier EPUB publications invalid. The considerations in the Community Group would correct this problem.		The text has been modified in such a way that it applies to EPUB versions 3.x.	Accepted. As amended. <i>Note: see proposed disposition of comments.</i>
9	US 02 (TC46)				Ge	EPUB 3.0.1 does not reference the latest version of HTML, which should be corrected in the next version of EPUB, probably EPUB 3.0.2, by the Community Group. These changes should also be beneficial to long term preservation.		OK	Acknowledged. No action.
10	US 03				Ge	The metadata that is referenced by an	We recommend the metadata be internal	The specification recommends that	Accepted. As

1 **MB** = Member body / **NC** = National Committee (enter the ISO 3166 two-letter country code, e.g. CN for China; comments from the ISO/CS editing unit are identified by **)

2 **Type of comment:** **ge** = general **te** = technical **ed** = editorial

0	1	2	3	4	5	6	7	8	9
#	NB ¹	Line number (e.g. 17)	Clause/ Subclause (e.g. 3.1)	Paragraph/ Figure/ Table/ (e.g. Table 1)	Type of commen ²	Comment (justification for change) by the NB	Proposed change by the NB	Proposed Disposition of Comments	Disposition of Comments
	(TC46)					EPUB can be internal or externally referenced. While this is mentioned in the specification, It may be considered by the working group to require that the metadata should be internal to the publication, or that the referenced metadata also be archived. It would be a problem if the EPUB was archived, but the metadata it referenced was lost due to it being externally referenced.	to the publication.	metadata should be embedded. If it is not, the archive must have access to external metadata so that it can be retrieved during ingest and included in AIP.	amended. <i>Note: see proposed disposition of comments.</i>
11	US 04 (TC46)				Ge	We suggest that the EPUB Accessibility Conformance and Discovery Specification 1.0 (which would also change if EPUB 3.1 is retracted), should be metadata that should be included.		This requirement should be discussed in the WG. Accessibility as such does not guarantee that a document can be preserved in long term.	Acknowledged. No action. <i>Note: see proposed disposition of comments.</i>
12	US 05 (TC46)				Ge	Any modern archiving system should require accessibility metadata. Not only is this metadata important to determine if persons with disabilities can effectively use the document, but publications that are accessible are probably better as an archiving format for preservation. For example, the EPUB Accessibility Specification requires that the content that is textual be marked up properly, e.g. the text is not presented as an image. It is possible to create EPUB, which would pass epubcheck validation that, is only images of pages. However, this type of EPUB would not be accessible and the metadata should make that clear.		See above.	Resolved per US04.
13	US				Ge	To assist in archiving, we would also	We recommend including two checkers	OK. Ace is	Accepted. As

1 **MB** = Member body / **NC** = National Committee (enter the ISO 3166 two-letter country code, e.g. CN for China; comments from the ISO/CS editing unit are identified by **)

2 **Type of comment:** **ge** = general **te** = technical **ed** = editorial

0	1	2	3	4	5	6	7	8	9
#	NB ¹	Line number (e.g. 17)	Clause/ Subclause (e.g. 3.1)	Paragraph/ Figure/ Table/ (e.g. Table 1)	Type of commen ²	Comment (justification for change) by the NB	Proposed change by the NB	Proposed Disposition of Comments	Disposition of Comments
	06 (TC46)					suggest that two checkers be used in the archiving process. Epubcheck has already been mentioned, but also Ace by DAISY is available now as beta and will be released in January 2018 in a first release. This is open source and provides valuable information in the evaluation of an EPUB for accessibility purposes. Information about the software referenced can be found at http://www.inclusivepublishing.org	in the archiving process. Add Ace by DAISY.	mentioned in chapter 7.4 of Part 2 and a link is provided to the Ace homepage in Github.	amended. <i>Note: see proposed disposition of comments.</i>
14	CA 01 (TC46)		02	Normative References	Ge	Consider whether ISO 14721 should be listed as a normative reference. A significant portion of this technical specification arises from the OAIS standard ISO 14721:2012.	Add ISO 14721:2012 as a normative reference.	OK	Accepted. As proposed.
15	CA 02 (TC46)		03	Terms and Definitions	Ge	It is recognized that terms and definitions from ISO 14721 have not been added to the ISO OBP. This makes it necessary to add citations to each specific term and definition. This assures the reader that the working group was working with the current definitions from the current version of 14721.	Add a citation to ISO 14721:2012 to all of the terms and definitions copied from the published standard. Doing this will facilitate harmonization efforts within the ISO family in the future.	There are citations to all terms which do not originate from ISO 14721, and a comment that unless otherwise mentioned, term and its definition are from ISO 14721.	Accepted. As amended.
16	US 09 (TC46)		05	3 (pg 10)	Ed	"Both the data (documents) and metadata have to conform to the standards and specifications the producer and the archive have agreed upon. Only if the SIPs are compliant will the archive be able to ingest the data successfully."	There is a move afoot within the preservation community - a move that has long been taken by government preservation agencies who must take all content that comes to them, regardless of format and MD - towards minimal ingest. With minimal ingest, even "broken" content that does not conform to agreed upon specifications is preserved with the intent to correct or improve the content (either the archive alone or with the producer's assistance)	The draft already allows SIPs which contain unarchivable data. Such data should be encoded in such a way that it can be bypassed during Ingest. It should be noted that this TS does not discuss the validation issues	Accepted. As amended. <i>Note: see proposed disposition of comments.</i>

1 **MB** = Member body / **NC** = National Committee (enter the ISO 3166 two-letter country code, e.g. CN for China; comments from the ISO/CS editing unit are identified by **)

2 **Type of comment:** **ge** = general **te** = technical **ed** = editorial

0	1	2	3	4	5	6	7	8	9
#	NB ¹	Line number (e.g. 17)	Clause/ Subclause (e.g. 3.1)	Paragraph/ Figure/ Table/ (e.g. Table 1)	Type of commen ²	Comment (justification for change) by the NB	Proposed change by the NB	Proposed Disposition of Comments	Disposition of Comments
							over time. In general, producers often have a difficult time consistently hitting specifications. Between 15% and 20% of the content Portico receives does not meet agreed upon specifications. This document may want to consider a more flexible approach, as it is better to have preserved less than perfect content than to not preserve any content at all.	(such as unreliable validation tools) since they are beyond the scope of the document.	
17	US 10 (TC46)	06.01.1	Bullet 'i'		Ge	“Preview publications SHOULD NOT be submitted, since they are by definition not complete.” We may want to consider the statement. It is very common in today's world, for publications to be published on an ongoing basis. When publishing on the internet first came to be, it did tend to model the print world more with publications being finite and complete at some moment in time. In today's world, however, it is more and more common for publications to be "open" for a considerable amount of time – they are amended, supplemented, changed, etc. At Portico, we encourage our publishers to send us early publications (ahead of print or ahead of being assigned to a vol/iss, etc.), so long as there is an identifier in the content. Then, we encourage our publishers to send us updates as they are made. So long as the items we receive at different times share identifiers, we will update the originally preserved item with the newer version. We suspect this requirement, quoted above, will rapidly make this standard out-of-date, as publishing today		OK. This requirement has been modified so that preview publications may be submitted if the submission agreement allows it; the assumption is that the final document will be delivered later.	Accepted. As amended. <i>Note: see proposed disposition of comments..</i>

1 **MB** = Member body / **NC** = National Committee (enter the ISO 3166 two-letter country code, e.g. CN for China; comments from the ISO/CS editing unit are identified by **)

2 **Type of comment:** **ge** = general **te** = technical **ed** = editorial

0	1	2	3	4	5	6	7	8	9
#	NB ¹	Line number (e.g. 17)	Clause/Subclause (e.g. 3.1)	Paragraph/Figure/Table/ (e.g. Table 1)	Type of commen ²	Comment (justification for change) by the NB	Proposed change by the NB	Proposed Disposition of Comments	Disposition of Comments
						is dynamic and "live", and will only become more so.			
18	US 11 (TC46)	06.01.2	Bullet 3		Ge	"Submission agreements SHOULD specify metadata formats and file formats approved for ingest and/or archival." --> In general, we have had little luck at specifying the use of specific metadata or file formats. Instead, we try to emphasize with our publishers that we would like them to be consistent. We have been far more successful in requesting that they send us content in the same format every time content is delivered (whatever format they prefer), than in requiring they hit any given specification. An archive that requires producers to hit a specific format and is receiving content at any sort of substantial scale is likely to find it difficult to succeed, due to the amount of content that does not successfully make it into the archive and the amount of communication they must maintain with each producer.		Submission agreement does not need to be a bottleneck from the producers' point of view. It may contain also the metadata formats and/or file formats the producer prefers to use, although they were not ideal from preservation point of view. The important thing is that the producer does not send material that has not been agreed upon in advance, since there is a high risk that the ingest of such SIPs would fail.	Acknowledged. No action. <i>Note: see proposed disposition of comments.</i>
19	US 12 (TC46)	06.02.2			Ed	Pg 17 "international standards such as ISBNs... SHOULD be used as EPUB unique identifiers" == Part 2 explicitly says (pg 13) ISBN should not be used. In addition, at Portico we like to encourage our publishers to provide all identifiers. Often times the intellectual content of a book has multiple ISBN (one for each format/rendition) and our libraries often only care that the intellectual content be preserved, not that any given format is preserved.		ISBN should be used as the identifier of the EPUB e-book, but it should not be used as the identifier of SIPs or other information packages.	Acknowledged. No action. <i>Note: see proposed disposition of comments.</i>
20	US 13	06.04.1	Bullet 1		Te	"METS SHALL be used as the container standard". METS is great, but it is also	Make this to be less strict, perhaps something like "METS or another agreed	OK. The specification does	Accepted. As

1 MB = Member body / NC = National Committee (enter the ISO 3166 two-letter country code, e.g. CN for China; comments from the ISO/CS editing unit are identified by **)

2 Type of comment: ge = general te = technical ed = editorial

0	1	2	3	4	5	6	7	8	9
#	NB ¹	Line number (e.g. 17)	Clause/ Subclause (e.g. 3.1)	Paragraph/ Figure/ Table/ (e.g. Table 1)	Type of commen ²	Comment (justification for change) by the NB	Proposed change by the NB	Proposed Disposition of Comments	Disposition of Comments
	(TC46)					complicated and not human readable and in Portico’s experience introducing barriers to the producer (such as you must hit XYZ standard) delays or prevents preservation.	upon container format” - - something that lets thee producer have some choice, is likely to be more successful.	now allow multiple container format, and the submission agreement must specify which one or which ones are to be used. Part 2 will only cover usage of METS. The TS will not discuss the issue of how to create SIPs in any detail, but producers should have tools to help them. One such tool set has been developed and used successfully in the Finnish National Digital Library project.	amended. <i>Note: see proposed disposition of comments.</i>
21	US 14 (TC46)	06.04.1	Bullet 10		Ed	"Text metadata should be provided in Romanized form"	Is this necessary in the world of Unicode? Text metadata is not always received in Romanized form and it is not a particular problem with modern operating systems, software languages, and web browsers.	The need for this requirement will be discussed in the WG.	Accepted. As amended. <i>Note: see proposed disposition of comments.</i>
22	US 15 (TC46)	06.04.1	Bullet 3		Te	"the minimum required metadata for EPUB publications are title, identifier, and language". A publication date (e.g., one could have multiple books all titled "The Silence" and publication date could be a very important element in	Allow for multiple publication dates.	OK	Accepted. As amended.

1 **MB** = Member body / **NC** = National Committee (enter the ISO 3166 two-letter country code, e.g. CN for China; comments from the ISO/CS editing unit are identified by **)
2 **Type of comment:** **ge** = general **te** = technical **ed** = editorial

0	1	2	3	4	5	6	7	8	9
#	NB ¹	Line number (e.g. 17)	Clause/ Subclause (e.g. 3.1)	Paragraph/ Figure/ Table/ (e.g. Table 1)	Type of commen ²	Comment (justification for change) by the NB	Proposed change by the NB	Proposed Disposition of Comments	Disposition of Comments
						distinguishing between them) is often required. In addition, not all publishers have “language” metadata easy enough to hand to put it into the descriptive metadata for a book.			
23	US 16 (TC46)	06.04.1	Bullet 7		Ed	“how metadata ire stored in SIPs”	Change ‘ire’ to ‘are’	OK	Accepted. As amended.
24	US 17 (TC46)	06.04.2	Bullet 2		Te	"File format identification and validation metadata (created with e.g. EpubCheck) SHOULD be included in the SIP."	Most producers do not validate their content at the point of export. You may want to soften this statement, so that it is not off-putting to producers/publishers who do not and will not ever have this capability.	This has been changed to MAY, pending approval from the WG.	Accepted. As amended. <i>Note: see proposed disposition of comments.</i>
25	US 18 (TC46)	Title			Ed	The title suggests that it is specifically about the EPUB specification, whereas it is really about best practices for preserving content that happens to be in EPUB format.	We recommend the title be revised.	OK. the new title is Preserving content in EPUB Format. French translation needs to be provided separately.	Accepted. As amended. <i>Note: see proposed disposition of comments.</i>
26	W3C 01				Ge	Someone has already pointed out the issue that this refers to 3.1 exclusively. Aside from the point that we are looking at changing 3.1 to 3.0.2, it is worth pointing out that 3.0.1 is in very wide use and will likely continue that way for some time (at least until epubcheck is updated – which will not happen unless we get funding). Even though this document refers to 3.1, all the reference material is for 3.0 or 3.0.1. In most cases, this does not create problems, but it could be misleading. For example, I assume we’d prefer a reference to http://kb.daisy.org/publishing/ over https://i.dpf.github.io/a11y-guidelines/ .		The specification will be edited in such a way that it applies to 3.x.	Accepted. As amended. <i>Note: see proposed disposition of comments.</i>

1 **MB** = Member body / **NC** = National Committee (enter the ISO 3166 two-letter country code, e.g. CN for China; comments from the ISO/CS editing unit are identified by **)

2 **Type of comment:** **ge** = general **te** = technical **ed** = editorial

Date: 2018-01-25	Document: SC34 N_____	Project: ISO/IEC PDTS 22424-1
------------------	-----------------------	----------------------------------

- 1 **MB** = Member body / **NC** = National Committee (enter the ISO 3166 two-letter country code, e.g. CN for China; comments from the ISO/CS editing unit are identified by **)
- 2 **Type of comment:** **ge** = general **te** = technical **ed** = editorial