

What is a 'simple' ebook?

Defining an ebook is not easy - and indeed an important part of the COPIM and OBF projects has been to push the boundaries of any such definition through the Experimental Publications work packages ([link to the related book in Copim Compas](#)).

For the sake of this toolkit we consider content that might reasonably be expected to be published and distributed as an ebook in PDF or EPUB format. Adapting slightly the 'simple' ebook structure developed by (Stewart et al [ref](#) and [link to linkrot report](#)), we identify four main components of such a publication -

1. the core text - primarily text and equations, including their layouts such as in tables and poems;
2. embedded (non-textual) material - such as images, audio and video files;
3. outbound links to external material - which may include additional resources or datasets associated with the publication, as well as citations or other third party content that may be important to the understanding of the work;
4. the metadata associated with the work.

Figure 1 provides a representation of the main components of the type of 'simple' ebook publication we consider.

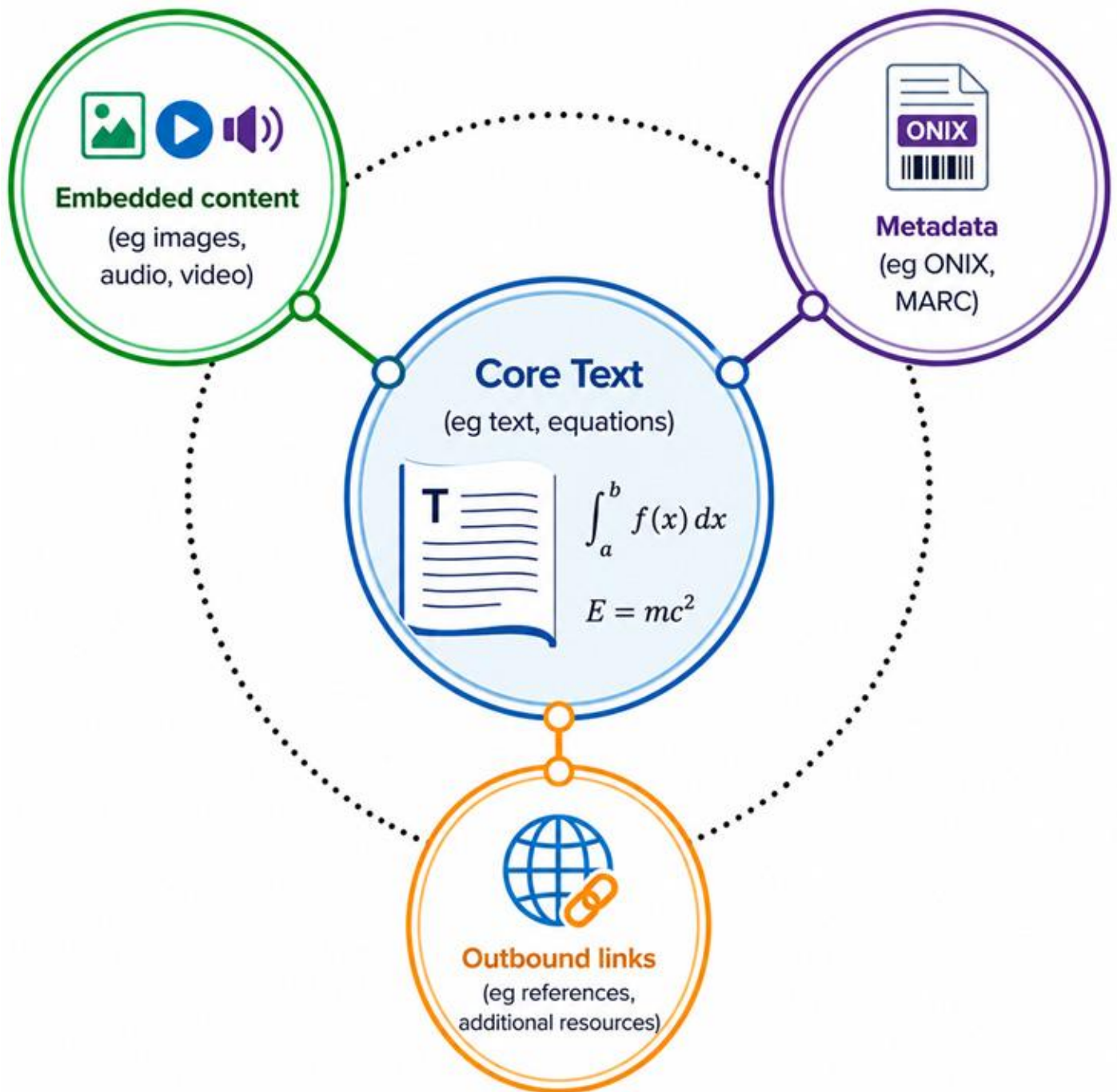


Figure 1: 'Simple' eBook - main components.

As (Stewart et al ... link) notes: "The word 'simple' is used very loosely here(!), as evidently many layers of complexity could permeate in any of this model's components. For example, metadata may be hosted externally in a separate system and linked to the eBook, embedded AV may be playable on an external web platform etc. There are many, many unusual combinations of doing things on the Web, which would be impossible to advise on in every instance."

Furthermore, not all published formats of a single work will necessarily contain precisely the same content - the quantity and format of embedded material in the PDF and the EPUB editions of a work may vary, as may any embedded metadata.

It should also be noted that the content identified as important for archiving may differ from that included in any specific format used to publish the work. While it seems likely that the Core Text should be included in any archiving strategy, the other components may be considered more or less important for the understanding of the work and so be of greater or lesser importance for archiving purposes. Consequently, an effective archiving strategy is likely to differ from a strategy of archiving a single format of the published work.

Revision #1

Created 2026-04-28 14:31:29 UTC by Rupert Gatti

Updated 2026-04-30 09:54:44 UTC by Rupert Gatti