**Audiobook Specification Explainer**

**What is a specification and why are we doing this?**

Audiobooks have grown massively in popularity in the last few years despite being a segment of the industry for far longer. The main reason for this shift is due to several technologies:

* Mobile devices (phones and tablets)
* Improved mobile networks (3G, 4G LTE)
* Improved mobile storage capacity
* Widespread use of wifi

When previously audiobook users would need several CDs to listen to their audiobook, they can now stream, download, and listen to their chosen content anywhere and time they choose. Audiobooks, like ebooks, are now in their pockets on demand.

The main difference between ebooks and audiobooks currently is their level of specification. EPUB has existed in one form or another for almost 20 years, but audiobooks still do not have a common specification. A mature ebooks industry has co-opted audiobooks to make them work within the same contexts, but the lack of a specification has made this process arduous for user agents and by extension, users.

An audiobooks spec would also help with B2B relationships. Today publishers have to deal with multiple non-standards, preparing files and metadata differently for each audiobook distributor. This results in duplicated effort, increases the chances of errors, and makes the supply chain less efficient.

It is in the interest of users, creators, and distributors that we push to create a single specification for audiobooks.

But what do we mean when we talk about specifications? A specification is a set of documented requirements to be satisfied by a product or service. In this case, we are developing a standard specification for the production of audiobooks on the web ([Web Publications](https://w3c.github.io/wpub)) with an extension for packaging those files for distribution ([Packaged Web Publications](https://w3c.github.io/pwpub/)).

This specification outlines the format for audiobook manifests, including metadata, reading (listening) order, additional resources, and the table of contents. We are also working towards a packaging format that will wrap audiobook files into a single package, simplifying their movement in the supply chain.

The specification will not address synchronized media or digital rights management (DRM). Synchronized media is being handled by another community group at the W3C, and we work closely with them to ensure that anything we implement for audio does not interfere in their efforts, but we aren’t considering it for the audiobooks specification. DRM is outside of the scope of our charter, our specification will not actively prevent it’s use, but we do not consider DRM in any of our decisions.

**Goals**

* Create a specification for the audiobook format that is usable on both the web and in packaged contexts
* Create a specification that supports all of the major use cases of audio
* Create a specification that is accessible
* Create a specification that meets the needs and requirements of publishers, user agents, and users

**Out of Scope**

* Synced media — though the needs of sync media in a metadata/structure context will be taken into account
* DRM (as outlined by our charter)

**Key Use Cases**

The Audiobooks Task Force has identified the definition of an essential audiobook experience as:

* A format that can be listened to from beginning to end without user input (moving forwards/backwards through the reading order without manual input)
* A format that can be paused, played, moved forwards and backwards
* A format where the listening position is retained for the next listening session
* A format where the user can access the table of contents at any time
* A format where the user can always find their position
* A format where the user knows how long is remaining in the chapter/section/audiobook
* A format that can be streamed, offlined, and downloaded
* A format with clear metadata support for user libraries (title, author, narrator, cover, duration, etc.)

**The Business Case**

**State of Formats today:**

This table compares the proposed specification to current standards used in some businesses today.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Proposed Specification (W3C)** | **Current Business Case 1** | **Current Business Case 2** | **Current Business Case 3** |
| **Package** | zip with cover, files, manifest | cover and files  sent individually to unique FTP subfolders | cover and files  sent individually to unique FTP subfolders | cover, files, product metadata  sent individually to unique FTP subfolders |
| **Metadata** | Internal and any existing vocabularies (ONIX, Dublin Core, etc.) | ONIX | ONIX | XML |
| **Manifest** | JSON | Text | JSON or Excel | Text |
| **Table of Contents** | Simplified HTML, with JSON being considered as an alternative | not currently sent | included in manifest | sent as chapter level metadata |
| **Track File Format** | Spec will not restrict format | mp3 | MP3 or M4A | mp3 or WAV or other |
| **Track Naming Convention** | Spec does not impose naming conventions | ISBN\_sequence\_MP3.mp3 | isbn\_sequence\_filename | track\_sequence.mp3 |
| **Delivery Of Tracks** | individual tracks | individual tracks | individual tracks | tracks stitched together |
| **Track File Size** | No restrictions on file size | typically one file per chapter  files split if >75 minutes | files < 200mb products < 800mb | < 24 hours |
| **Covers** | WP format for covers in any file format (included as resources) | 1:1 aspect ratio jpeg | 1:1 aspect ratio jpeg | 1:1 aspect ratio jpeg |
| **Previews** | To be discussed | first 5 minutes  custom samples | first 5 minutes  custom samples | first 5 minutes  custom samples |
| **Enhancements** | Supplemental content may be included in any format | sent as PDF | not part of current release still tbd | sent as PDF |

**How It Works**

The basic idea is simple: put all the audio files, a cover image, any supplemental material, and some metadata in a folder, and zip the folder.

The tricky part is the metadata. We use the term “manifest” to describe the list of audio files, in order, that make up the book. The manifest also includes bibliographic metadata and identifies special resources like a cover image or a table of contents. To make all that information easily machine-readable, it’s written in a format called JSON. We expect that there will be simple tools available to create this format.

The manifest format comes from the W3C’s [Web Publication Manifest](https://w3c.github.io/wpub/) spec, which we hope will be used for many different types of publications.

**Considered Alternatives**

There were several considered alternatives:

**DAISY Talking Books**

[DAISY Talking Books](http://www.daisy.org/daisypedia/daisy-digital-talking-book)

**HTML Custom Elements**

HTML does have an audio element. It would be possible to have the Primary Entry Page of the Web Publication be an HTML page that contains audio elements for each constituent resource. Someone wrote a custom element called [x-playlist](https://github.com/heff/x-playlist) that wrapped a sequence of audio elements, auto-playing them in order with no intervention. A member of the working group created a [sample audio book](https://dauwhe.github.io/html-first/flatland-custom-element/) using this approach.

**EPUB 3 Audio Books**

The existing EPUB 3 specification could be changed to allow audio core media types as spine items. We could then take advantage of EPUB’s existing metadata and packaging features.

**M4B**

M4B is an audiobook format based on the [MPEG-4 Part 14](https://en.wikipedia.org/wiki/MPEG-4_Part_14) container specification. They can contain multiple audio files, a cover image, and metadata. Audio book providers such as Librevox often supply M4B to end users. They can be played in iTunes.

**References**

Please note that these experiments are not completely up to date, and the additions to the spec will have more complete examples once the draft is published.

**Experiments:**

**Audiobook Web Publication:**

<https://github.com/w3c/wpub/tree/master/experiments/audiobook> (code)

<https://w3c.github.io/wpub/experiments/audiobook/index.html> (sample)

Sample Manifest:

{

"@context": ["https://schema.org", "https://www.w3.org/ns/wp-context"],

"type": "Audiobook",

"id": "https://librivox.org/flatland-a-romance-of-many-dimensions-by-edwin-abbott-abbott/",

"url": "https://w3c.github.io/wpub/experiments/audiobook/",

"name": "Flatland: A Romance of Many Dimensions",

"author": "Edwin Abbott Abbott",

"readBy": "Ruth Golding",

"publisher": "Librivox",

"inLanguage": "en",

"dateModified": "2018-06-14T19:32:18Z",

"datePublished": "2008-10-12",

"license": "https://creativecommons.org/publicdomain/zero/1.0/",

"resources": [

{"rel": "cover", "url": "http://ia800704.us.archive.org/9/items/LibrivoxCdCoverArt12/Flatland\_1109.jpg", "encodingFormat": "image/jpeg"},

{"rel": "contents", "url": "toc.html", "encodingFormat": "text/html"}

],

"readingOrder": [

{"url": "http://www.archive.org/download/flatland\_rg\_librivox/flatland\_1\_abbott.mp3", "encodingFormat": "audio/mpeg", "name": "Part 1, Sections 1 - 3"},

{"url": "http://www.archive.org/download/flatland\_rg\_librivox/flatland\_2\_abbott.mp3", "encodingFormat": "audio/mpeg", "name": "Part 1, Sections 4 - 5"},

{"url": "http://www.archive.org/download/flatland\_rg\_librivox/flatland\_3\_abbott.mp3", "encodingFormat": "audio/mpeg", "name": "Part 1, Sections 6 - 7"},

{"url": "http://www.archive.org/download/flatland\_rg\_librivox/flatland\_4\_abbott.mp3", "encodingFormat": "audio/mpeg", "name": "Part 1, Sections 8 - 10"},

{"url": "http://www.archive.org/download/flatland\_rg\_librivox/flatland\_5\_abbott.mp3", "encodingFormat": "audio/mpeg", "name": "Part 1, Sections 11 - 12"},

{"url": "http://www.archive.org/download/flatland\_rg\_librivox/flatland\_6\_abbott.mp3", "encodingFormat": "audio/mpeg", "name": "Part 2, Sections 13 - 14"},

{"url": "http://www.archive.org/download/flatland\_rg\_librivox/flatland\_7\_abbott.mp3", "encodingFormat": "audio/mpeg", "name": "Part 2, Sections 15 - 17"},

{"url": "http://www.archive.org/download/flatland\_rg\_librivox/flatland\_8\_abbott.mp3", "encodingFormat": "audio/mpeg", "name": "Part 2, Sections 18 - 20"},

{"url": "http://www.archive.org/download/flatland\_rg\_librivox/flatland\_9\_abbott.mp3", "encodingFormat": "audio/mpeg", "name": "Part 2, Sections 21 - 22"}

]

}

Entry Page:

<!DOCTYPE html>

<html lang="en">

<head>

<title>Audiobook Player</title>

<meta charset="utf-8">

<meta name="viewport" content="width=device-width, initial-scale=1">

<link href="flatland.json" rel="publication">

</head>

<body>

<main>

<img id="cover" style="display:block; margin:auto;">

<audio id="audio-element" controls controlsList="nodownload" style="display:block; margin:auto;">

<source id="audio-source" type="audio/mpeg">

</audio>

</main>

<nav style="text-align: center; margin-top: 10px;">

<a id="previous" rel="prev">&lt;&nbsp;Previous Track</a>

&nbsp;&nbsp;

<a id="next" rel="next">Next Track&nbsp;&gt;</a>

<br />

<a href="toc.html">Table of Contents</a>

</nav>

<script src="polyfills/fetch.js"></script>

<script src="polyfills/urlsearchparams.js"></script>

<script src="player.js" async></script>

</body>

</html>

HTML TOC Sample:

<!DOCTYPE html>

<html>

<head>

<title>Table of Contents for Flatland</title>

</head>

<body>

<nav role="doc-toc">

<ul>

<li>

<a href="http://www.archive.org/download/flatland\_rg\_librivox/flatland\_1\_abbott.mp3#t=71">Part 1 - This World</a>

<ul>

<li>

<a href="http://www.archive.org/download/flatland\_rg\_librivox/flatland\_1\_abbott.mp3#t=80">Section 1 - Of the Nature of Flatland</a>

</li>

<li>

<a href="http://www.archive.org/download/flatland\_rg\_librivox/flatland\_1\_abbott.mp3#t=415">Section 2 - Of the Climate and Houses in Flatland</a>

</li>

<li>

<a href="http://www.archive.org/download/flatland\_rg\_librivox/flatland\_1\_abbott.mp3#t=789">Section 3 - Concerning the Inhabitants of Flatland</a>

</li>

<li>

<a href="http://www.archive.org/download/flatland\_rg\_librivox/flatland\_2\_abbott.mp3#t=18">Section 4 - Concerning the Women</a>

</li>...

JSON TOC Example:

"toc": [

{

"url": "http://www.archive.org/download/flatland\_rg\_librivox/flatland\_1\_abbott.mp3#t=71",

"name": "Part 1 - This World",

"children": [

{

"url": "http://www.archive.org/download/flatland\_rg\_librivox/flatland\_1\_abbott.mp3#t=80",

"name": "Section 1 - Of the Nature of Flatland"

},

{

"url": "http://www.archive.org/download/flatland\_rg\_librivox/flatland\_1\_abbott.mp3#t=415",

"name": "Section 2 - Of the Climate and Houses in Flatland"

},

{

"url": "http://www.archive.org/download/flatland\_rg\_librivox/flatland\_1\_abbott.mp3#t=789",

"name": "Section 3 - Concerning the Inhabitants of Flatland"

},...

**Packaged Audiobook:**

<https://www.w3.org/2018/audiobook_examples/flatland.audiopub>