Web Publication Metadata Specification (Proposal)

*Web Publications are not required to provide **any** metadata.*

Much like HTML itself, Web Publications as a format may have stricter requirements *as a whole* while not making those requirements in the metadata sections themselves.

Example: HTML files must have a title but do not rely on the meta element to provide it.

If a Web Publication manifest requires specific information or properties to be valid, those properties may or may not be provided by the metadata section. This outline for a specification for the metadata section does not affect that or take a stance on those issues in any way.

For optimal interoperability and reliability, the Web Publication metadata section *should* provide:

- Some form of persistent identifier.

- Title (may be required by other sections, this section does not take a stance on that issue).

- Creator(s) with some method for indicating a more specific role.

- Language of the manifest (may be required by other sections)

- Metadata values in general may be qualified by language to allow multiple character sets for the same information

- Publication date.

- Modification date.

- Accessibility metadata.

Identifiers and modification dates are an important aid in the distribution and storage of portable formats (past, current, and present) so authors are urged to include this information whenever they can.

The Web Publication metadata section *may* provide:

- A subtitle.

- Licensing information.

- Intended audience.

- General subject.

- Links to established metadata files. Including, but not limited to, ONIX, MARC, BIBFRAME, and schema.org.

No matter the format of the metadata sections manifest authors *should not* include complex, specialised, and industry-specific metadata in the manifest itself and *should* limit the metadata included in the manifest to the items above. Instead they *should* link to whichever format is most commonly accepted as the authoritative metadata format for their intended audiences. They can include multiple metadata links, one for each intended audience, if needed.

Examples:

- A web publication that is intended primarily for web consumption links to a schema.org JSON-LD file.

- A book intended for general trade links to an ONIX xml file.

- A journal intended mostly for libraries and academia links to a MARC or BIBFRAME file.

User Agents *may* decide to support metadata extensions, other metadata schemes, or additional metadata properties if they so choose. User Agents *must* ignore metadata properties that they do not recognise.

Rationale:

Specifying metadata is an extremely complex task that is already being handled by multiple different, often industry-specific, organisations. Our options are:

1. Adopt one of those pre-existing formats at the risk of reducing the format's general usefulness and support for other industries.

2. Specify our own from scratch to handle every use case.

3. Specify a very limited subset—selected to maximise interoperability—and delegate all specialised requirements to more mature specifications.

The third option is the de facto best practice in both ePub (clients support a very small number of metadata items universally while complex metadata is handled externally) and HTML (even when embedded, formats like schema.org are, as commonly used, external to the markup itself, included in script tags).

Given that this practice is the likely outcome of our efforts—even if we choose one of the other two options—explicitly codifying it in the specification would reduce ambiguity and aid interoperability.