ONIX for Books release 3.1 revision 1

This document summarizes the small number of updates to ONIX 3.1 that form revision 1 (*ie* ONIX 3.1.1). These changes were circulated to the ONIX national user groups and to members of the ONIX international steering committee for discussion. The steering committee ratified the changes at its meeting during the London Book Fair (March 2024).

For the avoidance of any doubt, these additions to the standard are all optional and fully backwards-compatible, so all correctly constructed ONIX 3.1 messages are also valid ONIX 3.1.1 files. In addition, all ONIX messages that were valid under previous iterations of ONIX 3.0 and that avoid the use of deprecated elements and the <Gender> tag remain valid according to the latest 3.1.1 Specification and the latest XML validation tools.

This revision comes only 12 months after the publication of release 3.1. EDItEUR normally aims for a two year cadence to revisions of ONIX, but in this instance, the new functionality was requested by a group of ONIX users and national groups to meet an urgent need to express an 'EU TDM opt-out' on marketing collateral.

Background

In various legal jurisdictions, text and data mining (TDM) of copyright text has been covered by a limited copyright exception. In the UK, for example, non-commercial TDM is subject to an exception while commercial (for profit) TDM is not. This can enable *inter alia* the training of large language models that drive generative AI, as well as other innovative and transformative uses of the material.

In the European Union, TDM of copyright text is similarly subject to a limited copyright exception (meaning the presumption is that TDM is allowed). However, TDM for purposes *other than scientific research* is also subject to an opt-out from the copyright exception: when a rightsholder opts out, the copyright exception is nullified, thus the material cannot legally be used in TDM for non-research purposes. Meanwhile, TDM for research by scientific and cultural heritage institutions remains subject to the copyright exception.

This opt-out raises the possibility of commercial licensing of TDM rights for commercial (non-research) applications in the EU, through direct or collective licensing.

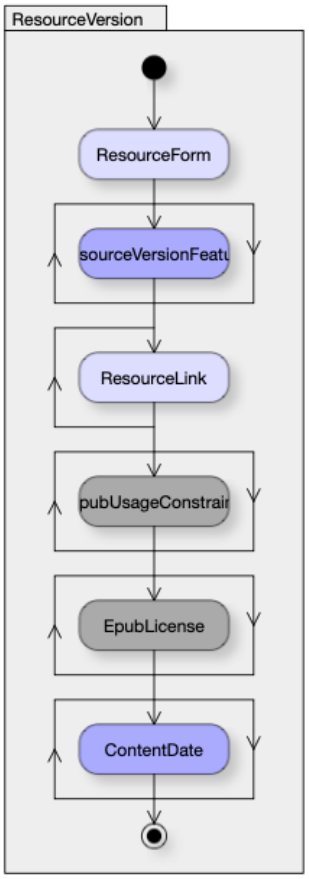
ONIX 3.0 and 3.1 can both express an EU-style opt-out that covers the content of a product, using <EpubUsageConstraint> with Usage type code 11 in Block 1 of ONIX. In 3.1, such constraints can also be applied on a chapter-by-chapter basis in Block 3.

The addition in ONIX 3.1.1 is that such opt-out expressions can *also* be applied to collateral material, including free-of-charge sample chapters of text or audio, cover images and sample spreads, reading guides and even long textual descriptions of the product, all of which are carried in or linked from Block 2. Note that there is no particular need to add constraints to <CitedContent> or <Prize>.

Applying license constraints to collateral materials

ONIX 3.1.1 adds the <EpubUsageConstraint> composite within <TextContent> and within <SupportingResource> (diagram below), in Block 2 and where the two composites occur in Blocks 3 and 7. Thus an opt-out relating to the content of a reading guide could be expressed like this:

```
<SupportingResource>
    . . .
    <ResourceVersion>
        . . .
        <ResourceLink>https://www.mypublisher.com/readingguides/
            .../guide.pdf</ResourceLink>
        <EpubUsageConstraint>
            <EpubUsageType>11</EpubUsageType>     <!-- TDM -->
            <EpubUsageStatus>03</EpubUsageStatus> <!-- is prohibited -->
        </EpubUsageConstraint>
    </ResourceVersion>
    . . .
</SupportingResource>
```



Note this prohibition should be interpreted as 'prohibited to the full extent allowed by the law or otherwise expressly reserved by the rightsholder', since with EU-style opt-outs, the rightsholder cannot opt out of the copyright exception for TDM for research purposes. The usage constraint can be applied to individual versions of the resource (*eg* an audio sample provided both as mp3 and aac files, or an image provided in high and low resolution versions).

Not all possible usage constraints make sense when applied to collateral material. For example, it is possible to limit the usage of the product content prior to purchase (*eg* 'you can preview 5% of the book before buying', using Usage type code 01 with an appropriate usage status and limit) – but this type of usage constraint should not be applied to a supporting resource which is freely available. Equally, opt-outs may not make sense for all possible collateral material types – for example, for review snippets that themselves are included in an ONIX message and used for marketing the book under fair use or fair dealing terms, the creator of the ONIX is unlikely to have legal standing to express an opt-out.

Note also that whenever possible, a supporting resource that is able to carry embedded metadata (*ie* within the resource itself) should express the opt-out in that embedded metadata as well as in the ONIX data that refers to it. For example, a reading guide available as an EPUB file should embed the TDM Reservation Protocol developed by the W3C [1]. A supporting resource that cannot carry the relevant embedded

[1] see https://www.w3.org/community/reports/tdmrep/CG-FINAL-tdmrep-20240202/. Note that this is not a W3C standard, but is a specification developed by a W3C Community Group and widely supported by organisations in the publishing industry. It may become a W3C recommendation in the future.

metadata may express an opt-out using the W3C's TDM Reservation Protocol in the HTTP Response Header sent by the server when the resource is requested from the given <ResourceLink>, or via a JSON-LD file in the .well-known folder on the server.

Finally, a third party – for example a retailer in receipt of a publisher's ONIX file and making use of descriptive text or supporting resources from that ONIX file on its website – must also ensure it makes clear the TDM rights are reserved.

Specifying TDM Reservation Protocol 'policies'

While text and data mining of content – or as above, of collateral material – may be prohibited, a separate license to do so may nonetheless be available from the rightsholder or via collective licensing. ONIX 3.1.1 also adds <EpubLicense> within <TextContent> and <SupportingResource> to enable 'advertising' this separate license.

In ONIX 3.0 and 3.1, <EpubLicense> is used in Blocks 1, 3 or 6 to express the license terms under which the digital product is available. Most commonly, it is used to indicate the product is 'open access', for example by linking to a Creative Commons license – though of course it may be used for closed licenses too.

As part of the W3C's TDM Reservation Protocol, a TDM 'policy' [2] can also be declared. This can be a simple re-statement that TDM is allowed or prohibited – analogous to <EpubUsageStatus>. However, if TDM is prohibited (other than for non-commercial or research purposes, depending on jurisdiction), a policy may *also* express an 'offer' that includes details of an *additional* license that *allows* TDM for commercial or non-research purposes. <EpubLicense> makes it possible to express TDM 'policies' (including the additional licenses) for e-publication content.

```
<EpubLicense>
    <EpubLicenseName>MyPublisher TDM License Policy</EpubLicenseName>
    <EpubLicenseExpression>
        <EpubLicenseExpressionType>20</EpubLicenseExpressionType>
        <EpubLicenseExpressionLink>https://www.mypublisher.com/.policies/
            mypolicy.json</EpubLicenseExpressionLink>
    </EpubLicenseExpression>
</EpubLicense>
```

In ONIX 3.1.1, the ability to express policies and offers using <EpubLicense> is extended to Blocks 2, 3 and 7 to cover collateral material within <TextContent> or linked from <SupportingResource>.

In the absence of concerns about TDM, there is no particular need to attach constraints, a formal license or a policy to content or to collateral material. Marketing material included or referenced in the ONIX is subject to either an *implied license* [3] that it may be used for the purposes of cataloguing, promoting and trading of the product, or a formal license such as the terms and conditions covering metadata use or redistribution by a metadata aggregator.

[2] defined as a 'description of the *kind* of TDM licenses an organisation may obtain from a rightsholder'

[3] See the *ONIX Implementation and Best Practice Guide* at https://www.editeur.org/93/Release-3.0-and-3.1-Downloads/#BestPractices for an explanation of the implied license

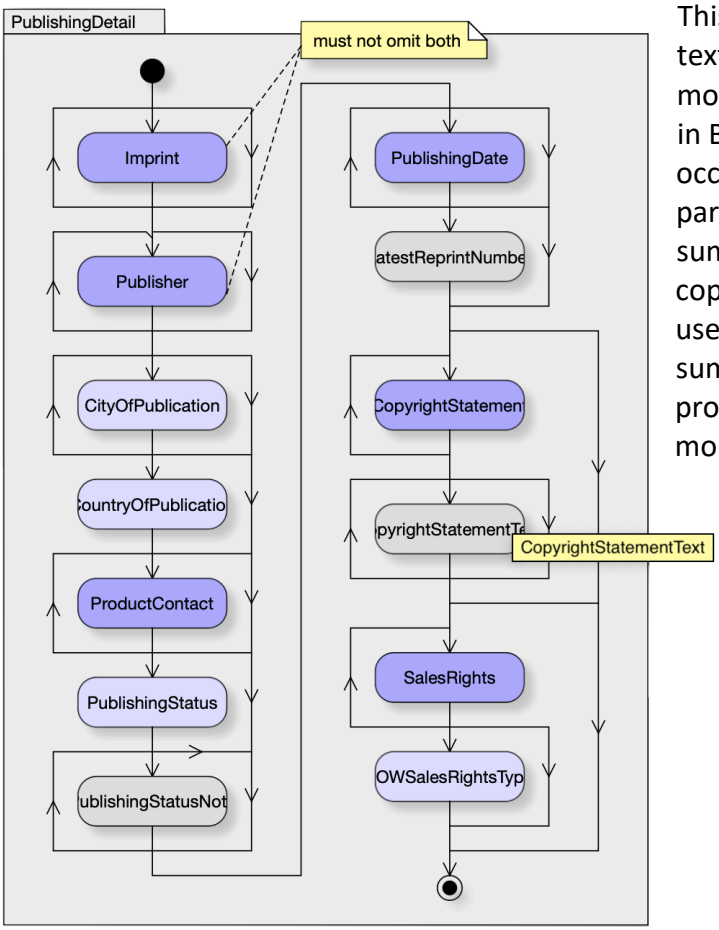
TDM policies and offers are expressed in JSON-LD format using the ODRL license expression vocabulary, and there are examples given in the TDM Reservation Protocol documentation. The 'policy' file would be at the URL given in <EpubLicenseExpressionLink>:

It is vital to understand that an 'offer' in a policy is not an intrinsic part of the e-publication license covering the product content, or the implied license to use the product metadata and collateral material for the purposes of cataloguing, promoting and trading the product. It is a separately-available *additional* license, either a standalone license from the rightsholder or a collective license offered by an RRO. Contact details for obtaining the license from the rightsholder may be given in the 'policy'. In general, these contact details should *also* be expressed in <ProductContact> (using Contact role code 09).

Care must be taken to ensure no inconsistency between <EpubUsageConstraint> and the policy. Expression of an opt-out in the policy linked from <EpubLicense> without also expressing it in <EpubUsageConstraint>, or *vice versa*, or provision of different contact details, is likely to cause confusion.

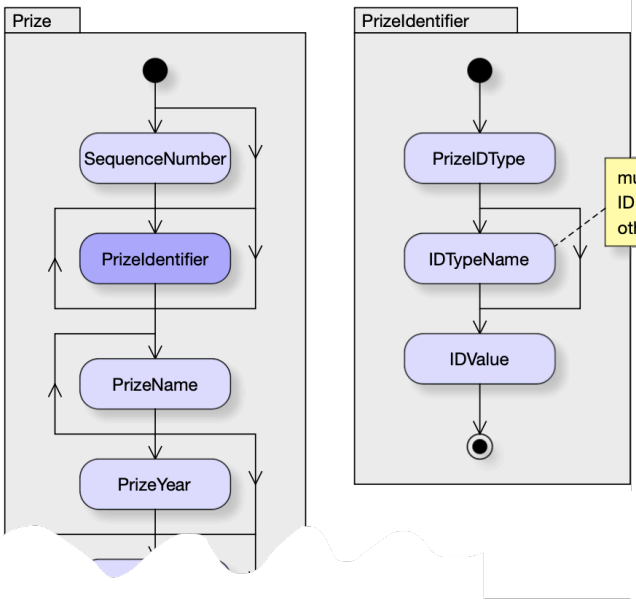
Other minor additions included in ONIX 3.1.1

Addition of <CopyrightStatementText>



This is an additional and optional free text field that immediately follows one or more instances of <CopyrightStatement> in Block 4 (and in Block 3 where it also occurs). It is intended to contain a line or paragraph of simple-to-display text summarizing the various structured copyright statements, analogous to the use of <ContributorStatement> summarizing the authorship of the product, and always *in addition to* one or more <CopyrightStatement> composites.

Addition of <PrizeIdentifier>



This addition aims to support those who wish to provide identifiers for literary and other prizes. There is no international standard for this (nor is one likely) but proprietary identifiers may be useful to disambiguate literary prizes despite changes occasioned by, for example, changes in the prize sponsorship.

Some library authority files, including OCLC's FAST and DNB's GND already contain identifiers for some prizes, and a new codelist, List 263, will allow these Prize identifier types to be added in future. Alternatively, an organization such as a publishers association could create and disseminate proprietary identifiers for use within a particular supply chain or community of practice.

Graham Bell
EDItEUR
19th March 2024