

ONIX for Books

Product Information Message

Application Note: Multilingual metadata in
ONIX 3.0

One of the defining features of ONIX 3.0 is the ability to deliver descriptive book metadata in multiple languages at the same time.

ONIX has always delivered a limited level of language independence – the nature of the ONIX codelists means that while the data contains language-independent codes (*eg* BB for hardback, BC for paperback), the codelists are available in several different languages. The concept represented by the code never varies, but the labels applied to those concepts can be in any language. Thus BB can be labelled ‘hardcover’ (instead of ‘hardback’) in American English, «rilegato» in Italian, ‘ciltli kitap’ in Turkish, or 《精装书》 (jīngzhuāng shū) in Chinese.

However, language independence breaks down with descriptive text – book titles, contributor biographies, book abstracts and marketing descriptions, where text is inevitably ‘in’ a specific language.

Use cases

There is no requirement for multilingual metadata for most books – the books are written in a particular language, the descriptive metadata is in that same language, and the book is sold to end customers who speak and read the language.

But for a small number of books – for example bilingual editions, books sold into second language or multi-language markets (*eg* many European countries where there is a significant trade in books in English, or the market for Spanish books in the USA), or textbooks where teaching (particularly university-level teaching) may be done in a ‘foreign’ language – making descriptive metadata available in more than a single language can be valuable.

To meet this requirement, most data elements in ONIX 3.0 that contain descriptive text are repeatable, like this:

```
<BiographicalNote language="eng" textformat="05"><p>Umberto Eco, professor of semiotics at the University of Bologna, and author of <cite>The Name of the Rose</cite> and <cite>Foucault's Pendulum</cite>, is one of the world's bestselling novelists.</p></BiographicalNote>
<BiographicalNote language="ita" textformat="05"><p>Umberto Eco, professore di semiotica all'Università di Bologna e autore di <cite>Il nome della rosa</cite> e <cite>Il pendolo di Foucault</cite>, è uno dei romanzieri più venduto al mondo.</p></BiographicalNote>
```

1. The text in the two languages is *parallel* text – it says essentially the same thing in each language;
2. <BiographicalNote> occurs inside the <Contributor> composite. The <BiographicalNote> *field* is repeated for each language, not the whole composite;
3. The *language* attribute must be added, to make it clear to the recipient which language is which;
4. In this particular example, of course, the *textformat* attribute is also required because the data contains XHTML markup;
5. The ONIX tags themselves are (at least for ‘long tags’) always mnemonic English words – the second tag is *not* ‘<NotaBiografica lingua="ita" formatoditesto="05">’.

The ONIX tags that can be repeated like this – with multiple languages, the *language* attribute and (X)HTML markup – are:

<AncillaryContentDescription>	<MarketPublishingStatusNote>
<AudienceDescription>	<PrizeJury>
<BiographicalNote>	<PrizeStatement>
<BookClubAdoption>	<PromotionCampaign>
<CitationNote>	<PublishingStatusNote>
<CopiesSold>	<ReligiousTextFeatureDescription>
<ContributorDescription>	<ReprintDetail>
<ContributorStatement>	<SalesRestrictionNote>
<EditionStatement>	<Text>
<EventDescription>	<TextSourceDescription>
<FeatureNote>	<TitleStatement>
<IllustrationsNote>	<VenueNote>
<InitialPrintRun>	<WebsiteDescription>

The most commonly-used of these are probably <Text>, <BiographicalNote>, <TitleStatement> and <ContributorStatement>.

There are also a similar number of fields that can be repeated with multiple languages – but which *cannot* contain (X)HTML markup ¹.

<DeletionText>	<PriceTypeDescription>
<EpubLicenseName>	<PrizeJury>
<EventAcronym>	<ProductFormDescription>
<EventPlace>	<ProductFormFeatureDescription>
<EventName>	<ProfessionalPosition>
<EventTheme>	<RatingUnits>
<ListName>	<ReturnsNote>
<LocationName>	<SourceTitle>
<MessageNote>	<SubjectHeadingText>
<PricePartDescription>	<Websitelink>

Here's an example showing how multilingual Product form feature descriptions can be delivered. Safety warnings like these are an important use case for multilingual metadata:

```
<ProductFormFeature>
  <ProductFormFeatureType>13</ProductFormFeatureType>      <!-- EU toy safety -->
  <ProductFormFeatureValue>01</ProductFormFeatureValue>    <!-- carries CE logo -->
</ProductFormFeature>
<ProductFormFeature>
  <ProductFormFeatureType>13</ProductFormFeatureType>      <!-- EU toy safety -->
  <ProductFormFeatureValue>03</ProductFormFeatureValue>    <!-- carries '0-3' logo and
                                                             hazard warning text -->
  <ProductFormFeatureDescription language="eng">Not          <!-- publisher's exact warning
    suitable for children under 36 months, due              text in English... -->
    to small parts</ProductFormFeatureDescription>
  <ProductFormFeatureDescription language="fre">Ne          <!-- and in French -->
    convient pas aux enfants de moins de 3 ans,
    contient de petites pièces susceptibles d'être
    ingérées</ProductFormFeatureDescription>
</ProductFormFeature>
```

¹ There are in fact three further tags, <ConferenceTheme>, <PromotionContact> and <ReissueDescription>, but these are deprecated and should not be used at all.

The key marketing text fields used by retailers – the short and/or long descriptions, promotional headlines, maybe even the reviews – are also candidates for multilingual text:

```
<TextContent>
  <TextType>01</TextType>
  <ContentAudience>00</ContentAudience>
  <Text language="eng" textformat="05"><p>The year is 1327. Franciscans in a wealthy
    Italian abbey are suspected of heresy, and Brother William of Baskerville arrives
    to investigate. When his delicate mission is abruptly overshadowed by seven
    bizarre deaths, Brother William turns detective.</Text>
  <Text language="ita" textformat="05">Corre l'anno 1327. I francescani di una ricca
    abbazia italiana sono sospettati di eresia, e frate Guglielmo di Baskerville arriva
    per indagare. Quando la sua delicata missione viene improvvisamente messa in ombra
    da sette bizzarre morti, il fratello William diventa detective.</Text>
</TextContent>
```

A more rarely considered use for multilingual metadata that uses another of the repeatable ONIX tags is to provide different lists of keywords in different languages: this could be used by retailers or librarians to provide multiple different discovery methods for the book, for consumers or patrons who prefer different languages:

```
<Subject>
  <SubjectSchemeIdentifier>20</SubjectSchemeIdentifier>
  <SubjectHeadingText language="eng">Mediaeval mystery; Monastery; Abbey; Secret library;
    William of Baskerville; Adso of Melk; Franciscan monk; Dominican; Ecclesiastical;
    doctrinal conflict; Inquisition</SubjectHeadingText>
  <SubjectHeadingText language="ita">Mistero medievale; Monastero; Abbazia; Biblioteca
    segreta; Guglielmo da Baskerville; Adso da Melk; monaco francescano; domenicano;
    Ecclesiastico; conflitto dottrinale; Inquisizione</SubjectHeadingText>
</Subject>
```

Note again that it is the `<SubjectHeadingText>` field that is repeated, not the entire `<Subject>` composite, and if `<SubjectHeadingText>` is repeated, the *language* attribute is mandatory.

Multi-language titles

It's notable that the list of 'repeatable' tags above does not include the text of the title (typically in `<TitlePrefix>` and `<TitleWithoutPrefix>`, `<Subtitle>` and so on). Sometimes what's needed is an alternative title in another language, or perhaps the original title of a work in translation – and in this case, the entire `<TitleDetail>` composite can be repeated. An English-language edition of *The Name of the Rose* could have this title:

```
<TitleDetail>
  <TitleType>01</TitleType>                                <!-- title of the book itself -->
  <TitleElement>
    <TitleElementLevel>01</TitleElementLevel>
    <TitlePrefix language="eng">The</TitlePrefix>
    <TitleWithoutPrefix textcase="02" language="eng">Name of the
      Rose</TitleWithoutPrefix>
  </TitleElement>
</TitleDetail>
```

and this could be supplemented with the following:

```
<TitleDetail>
  <TitleType>03</TitleType>                                <!-- title in original language -->
  <TitleElement>
    <TitleElementLevel>01</TitleElementLevel>
    <TitlePrefix language="ita">Il</TitlePrefix>
    <TitleWithoutPrefix textcase="01" language="eng">nome della
      rosa</TitleWithoutPrefix>
  </TitleElement>
</TitleDetail>
```

to indicate the original title of the book prior to its translation into English. Note that this is the title of *another* book – *Il nome della rosa* is not on the translated book itself, and the product in question does not contain Italian text.

However, this is not the way to deal with *bilingual* titles. Here, the ‘other’ language is not the title of *another* book – it’s another title of *this* book, so the method is different. Here, <TitleElement> is repeated, as there are *two* titles, in different languages:

```
<TitleDetail>
  <TitleType>01</TitleType>
  <TitleElement>
    <TitleElementLevel>01</TitleElementLevel>
    <TitlePrefix language="eng" textcase="02">The</TitlePrefix>
    <TitleWithoutPrefix language="eng" textcase="02">Day of the
      Dead</TitleWithoutPrefix>
    <Subtitle>a bilingual celebration</Subtitle>
  </TitleElement>
  <TitleElement>
    <TitleElementLevel>01</TitleElementLevel>
    <TitlePrefix>El</TitlePrefix>
    <TitleWithoutPrefix language="spa" textcase="01">dia de los
      muertos</TitleWithoutPrefix>
  </TitleElement>
  <TitleStatement>El dia de los Muertos / The Day of the Dead: a
    bilingual celebration</TitleStatement>
</TitleDetail>
```

You’ll notice that both titles – English and Spanish – are at ‘product level’, Title element level 01. Each element has a *language* attribute, and in this case the *textcase* attribute also differs because of the different capitalisation conventions typically used in each language. The final <TitleStatement> suggests the publisher’s recommended way of displaying this title. ²

This book should also have an Edition type code BLL and two repeats of the <Language> composite with code 01 (language of the text in the book):

² There’s an oddity with this real-world book: only one of the languages has a subtitle.

```

<EditionType>BLL</EditionType>           <!-- Bilingual edition -->
<Language>
  <LanguageRole>01</LanguageRole>         <!-- language of the text in the book -->
  <LanguageCode>eng</LanguageCode>
</Language>
<Language>
  <LanguageRole>01</LanguageRole>         <!-- language of the text in the book -->
  <LanguageCode>spa</LanguageCode>
  <CountryCode>MX</CountryCode>          <!-- Mexican Spanish -->
</Language>

```

In contrast, an English edition of *The Name of the Rose* should have a *single* <Language> composite with code 01 because the product is the English translation, not a bilingual parallel text edition), but it might also list the original Italian title using code 02:

```

<Language>
  <LanguageRole>01</LanguageRole>         <!-- language of the text in the book -->
  <LanguageCode>eng</LanguageCode>
</Language>
<Language>
  <LanguageRole>02</LanguageRole>         <!-- original language of translated text -->
  <LanguageCode>ita</LanguageCode>
</Language>

```

Graham Bell
 EDItEUR
 25th June 2021