

## BIC Metadata Futures Group

### Discussion Paper: Metadata description of assets, products, asset/product collections or groupings

17 June 2011

Francis Cave, BIC consultant

#### Introduction

This document considers which metadata descriptive elements could be used in a description of:

- A specific product intended for sale (main current use of ONIX for Books).
- A specific asset used (or intended for use) in making products. This might be anything from an individual illustration or small chunk of text to the entire text of a product.
- A specific asset/product collection or grouping of significance in publishing terms that needs description. This might, for example, be:
  - a set or series
  - all the assets and products associated with a particular title or publishing project
  - all the assets associated with a particular contributor or source
  - all the assets to be used in building a related family of products (e.g. the assets contained in a “release”, used to build ebooks in specific formats);or it could be any other collection or grouping for which there is a need or advantage in managing the assets, or at least some aspects their description, as a single entity.

Many such collections or groupings will only be significant to the publisher in question, and therefore will only require identification and description within internal systems. Others will be of significance in trading relationships at various stages in the publication process, such as on acquisition of the assets, or in the manufacturing processes, or in the supply of saleable products.

The relationships between assets, products and significant collections/groupings are varied and in many cases can be quite complex. The simplest relationships are direct, hierarchical, one-to-one “parent-child-sibling” relationships, such as exist between editions and formats of a single title that share the same content. At the opposite extreme are assemblies of content from a variety of sources, in which there is likely to be a network of indirect and non-hierarchical relationships between the various assets involved.

The extent to which the description of an entity (asset, product or collection/grouping) might depend upon the description of another entity will vary according to the relationship between the entities: a high level of dependency can be expected in the case of simple hierarchical relationships, while a much lower level of dependency can be expected in more complex and more indirect relationships.

In this document we propose in detail only a single example of such a relationship: a simple hierarchical relationship between a title<sup>1</sup>, a “release”<sup>2</sup> and product. In this case description of the title is at a more generic or higher level, description of the release is at an intermediate level, dependent only upon the title-level description, and description of the product is at the lowest level, being dependent upon both the title and release level descriptions.

Intentional differences between descriptions at the higher and lower levels arise for two reasons:

1. The intermediate and lower levels may include descriptive elements that apply to that level of description and possibly also to lower levels, but not to higher levels. Thus the description of a release may include elements that are only relevant to description at that level, plus other elements that are also descriptive of products within the release. Similarly, the description of a product may include elements that are only relevant to description of that specific product.
2. Certain elements of a description may have different values at different levels. This may be a reflection of the fact that the relationship between lower and higher level is not a simple parent-child relationship (e.g. a product contains content not accounted for in description at the title or release level; or a product is given a different description to that specified at the title or release level, possibly for marketing or technical reasons).

In the following table descriptive elements are based upon ONIX for Books Release 3.0 and are presented in their ONIX "blocks" and "groups", in the order that they are described in the ONIX documentation. Each element is categorised according to the highest level of description to which it relates, with an 'X' in one of the three right-hand columns. Only where an element of description at a lower level might differ from the description at a higher level is this indicated with an additional 'X' to the right of the first 'X'. Thus a title element is normally part of the description at the title level, but may differ at the release or product level or both. Numbers (e.g. [1]) in the body of the table refer to notes below the table.

The use of ONIX for Books elements in this table is not intended to imply that ONIX for Books provides a *sufficient* set of elements for description other than at the product level, but to indicate the inter-relationships between product-level description and title-level or release-level description. There will be other elements at the title and release level that are not relevant at other levels and therefore are not inherited or altered. For example, there would need to be some elements for relating titles, releases and products, for which neither <RelatedProduct> nor <RelatedWork> would be entirely suitable.

### Proposed grouping of elements for describing related title, release and product entities

Block	Group	Element	Title	Release	Product
	P.2 Product numbers	<ProductIdentifier>		X[1]	X
		<Barcode>			X
<b>1: Descriptive detail</b>	P.3 Product form	<ProductComposition>		[2]	X
		<ProductForm>		[3]	X
		<ProductFormDetail>		[4]	X
		<ProductFormFeature>		[5]	X
		<ProductPackaging>			X
		<ProductFormDescription>		[6]	X
		<TradeCategory> [7]		X	X
		<PrimaryContentType>	X	X	

<sup>1</sup> A "title" will vary in its definition according to the way in which a publisher conceives a publishing project and how that relates on the one hand to contracts with rights holders and on the other hand to the publications that are produced.

<sup>2</sup> A “release” is taken to be an asset (for example an EPUB file) or bundle of assets released into the supply chain that is not itself intended to be saleable, but from which several saleable products in different formats are to be derived.

Block	Group	Element	Title	Release	Product
		<ProductContentType>	X	X	
		<Measure>			X[8]
		<CountryOfManufacture>			X
		<EpubTechnicalProtection>			X
		<EpubUsageConstraint>		X	X[9]
		<MapScale>[10]	X	X	X
		<ProductClassification>		X[11]	X
	P.4 Product part	<ProductPart>[12]			
	P.5 Collection	<Collection>[13]	X	X	X
	P.6 Title	<TitleDetail>[14]	X	X	X
		<ThesisType>	X		
		<ThesisPresentedTo>	X		
		<ThesisYear>	X		
	P.7 Authorship	<Contributor>	X	X[15]	X[15]
	P.8 Conference	<Conference>	X	X[16]	X[16]
	P.9 Edition	<EditionType>	X	X	
		<EditionNumber>	X		
		<EditionVersionNumber>		X	
		<EditionStatement>	X		
		<ReligiousText>	X	X	
	P.10 Language	<Language>	X		
	P.11 Extents and other content	<Extent>	X[17]	X	X
		<Illustrated>	X	X[18]	
		<NumberOfIllustrations>	X	X[18]	
		<IllustrationsNote>	X	X[19]	X[19]
		<AncillaryContent>	X	X[20]	X[20]
	P.12 Subject	<Subject>	X	X[21]	X[21]
		<NameAsSubject>	X	X[21]	X[21]
	P.13 Audience	<AudienceCode	[22]	X	X
		<Audience>	[22]	X	X
		<AudienceRange>	[22]	X	X
		<Complexity>[23]			
<b>2: Collateral detail</b>	P.14 Descriptions and other text	<TextContent>[24]	X	X	X
	P.15 Cited content	<CitedContent>	X	X	X
	P.16 Links to supporting resources	<SupportingResource>[25]	X	X	X
	P.17 Prizes	<Prize>[26]	X	X	X
<b>3: Content detail</b>	P.18 Content items	<ContentItem>[27]	X	X	X
<b>4: Publishing detail</b>	P.19 Publisher	<Imprint>		X	[28]
		<Publisher>		X	[28]
		<CityOfPublication>		X	[28]
		<CountryOfPublication>		X	[28]
	P.20 Global publishing status	<PublishingStatus>			X
		<PublishingStatusNote>			X
		<PublishingDate>			X
		<LatestReprintNumber>			X
		<CopyrightStatement>			X
	P.21 Territorial rights / sales restrictions	<SalesRights>[29]	X	X	X
		<SalesRestriction>			X
<b>5: Related material</b>	P.22 Related works	<RelatedWork>	X	X	X
		<RelatedProduct>	X	X	X
<b>6: Product supply</b>	P.24 Market	<Market>		X[30]	X
	P.25 Market publishing detail	<MarketPublishingDetail>		X[30]	X
	P.26 Supply detail	<SupplyDetail>		X[30]	X

## NOTES

1. Since it is not a tradable item, a release may be identified in a different way to a product, i.e. using a different identification scheme.
2. It may be appropriate to include an element equivalent to <ProductComposition> in the description of a release, e.g. <ReleaseComposition>. It would not be appropriate to use List 2 as the code list for this element.

3. It may be appropriate to include an element equivalent to <ProductForm> in the description of a release, e.g. <ReleaseForm>. It would not be appropriate to use List 150 as the code list for this element.
4. Is an equivalent element to <ProductFormDetail> needed for releases? If so, probably using a subset of List 175.
5. Is an equivalent element to <ProductFormFeature> needed for releases, possibly using a new code list? Only one value in List 79 is currently relevant to a digital release: 03 (text font).
6. An equivalent element to <ProductFormDescription> may be needed for releases.
7. Code list 12 probably does not currently contain all the values that would be useful in describing the trade category of a digital release or product.
8. This element is only usable in association with physical products.
9. <EpubUsageConstraint> would normally be applicable to an entire release, but may need to be overridden for specific products built from the release files.
10. <MapScale> may be fixed by the artwork/data provided, or it may be adjustable to satisfy presentation constraints for a particular product.
11. <ProductClassification> is technically only intended to apply to saleable products, but it may be appropriate to assign a default product classification to the entire release. It is not entirely clear how this element would be used, if at all, for digital products.
12. <ProductPart> contains most of the preceding elements and applies to parts of a multi-item product. These will therefore apply at the title, release or product level, according to which characteristic is being described.
13. A <Collection> is what in earlier versions of ONIX was either a <Set> or a <Series>. These could be defined at any level, since one might be describing a author-defined series (title level) or a publisher-defined set or series (release or product level). In many cases such information would not be specific to this simple title-release-product hierarchy, but would be inherited from a separate description of the set or series itself.
14. A title would normally be defined at the title level, but as already mentioned, titles are sometimes varied for marketing or design reasons, so there may be a need to override at the release or product level.
15. While most <Contributor> elements would be defined at the title level, certain details might vary at the release or product level. For example, names of contributors, and the sequence in which they are presented, may need to vary for marketing or design reasons. An audio book would need at least one additional contributor: the narrator.
16. A <Conference> would normally relate to the group. However, there may be occasions when a specific product (e.g. an e-book) within a title grouping is published in association with a conference.
17. Illustrations might be specified at the release level, depending upon whether a publisher considers differently illustrated and un-illustrated versions of the same title to be the same title or not.
18. Mostly <Extent> is not relevant at the title level. The exception is that it can be used to express the number of words, which would quite often be an invariant property of the underlying work.
19. The element <IllustrationsNote> may be at the title level, but there may be occasions when notes about resolution or colour reproduction are specific to a release or product.
20. <AncillaryContent> allows detailed description of illustrations and other non-text content. It may be at the title level, but as with <IllustrationsNote>, some information may need to be added at the release or product level.
21. Both <Subject> and <NameAsSubject> would normally be at the title level, but some subject identification/coding schemes may be specific to certain kinds of release or product, in which case they should probably not be associated with the group but with the appropriate releases or products.

22. Audience description might be relevant at the title level, but in practice would probably only be specified for a release or product.
23. Use of the <Complexity> element is deprecated.
24. Depending upon the code value used for <TextType> in <TextContent>, this element could be descriptive at the title, release or product level.
25. Depending upon the code value used for <ResourceContentType> in <SupportingResource>, this element could be descriptive at the title, release or product level.
26. A prize may relate to any level, depending upon the nature of the prize: to whom and for what it is awarded.
27. Content items may be specified with or without page references. Page references are only applicable to product-level description. Some content may be specific to a release, depending upon how the publisher defines a title.
28. A "global" release may include products that don't all share the same imprint, publisher, city and country of publication.
29. Sales rights are, at least implicitly, specified at the title level, being depending upon the contracts with the rights holders. If varied at the release or product level, sales rights would always restrict what is specified at the higher level.
30. Market and supply details may be specified (explicitly or implicitly) for a release, but ultimately they apply most directly at the product level.