The International Standard Text Code (ISTC):
A Work in Progress

A Supply Chain Perspective

**Prepared for
Book Industry Communication
and the Book Industry Study Group
by Michael Holdsworth**

March 2010

BIC

Contents

1. Introduction

The International Standard Text Code (ISTC) is a new identifier for the book industry. Whereas International Standard Book Numbers (ISBNs) provide a unique identity in the supply chain for each individual product format (either physical products—hardback, paperback; or digital products and formats—downloadable audiobook, PDF e-book, etc.), an ISTC identifies the underlying creative textual work.

Within ISTC, textual works are precisely defined. They are the outputs of creative or intellectual effort expressed as text (a unique string of words in a particular order), intended for publication. Thus *Hamlet* and *Pride and Prejudice* are textual works—in this context entirely *abstract* entities, independent of the physical or digital bindings/containers/wrappers in which they may be presented. A unique ISTC identifier therefore is applied to all editions and versions of the same core content. The use of ISTC thus identifies and groups together all formats or manifestations of the underlying work.

ISTC is a sophisticated and flexible standard, but its potentially multilayered capabilities may appear at first sight to be daunting. Impressive claims have been made regarding ISTC's potential value to the book industry. This paper goes behind the formal documentation to examine these claims. It provides context and background; and it offers examples to assist publishers and other stakeholders as they take their first steps in engaging with the Standard.

Successful adoption of ISTC by the industry will depend on early commitment from publishers and the Registration Agencies jointly to build the relationships between the underlying creative work and published products on solid foundations of consensus, rigor and consistent methodology. At the same time, open and flexible elements within the coding structure do allow for choices; and some of the claims made on behalf of some of the beneficiary groups require careful testing and validation.

This paper, commissioned by Book Industry Communication (BIC) in the U.K. and the Book Industry Study Group (BISG) in the USA, is intended to generate better understanding of the ISTC and its potential benefits. As a first step in this direction, it will simply "write things down" in order to stimulate discussion. It will attempt to answer some of the questions being asked by publishers and others about how the rules codified in the Standard ought best to be interpreted and leveraged. It offers a first stab at a perspective on best practice. Above all, however, it is grounded in a BIC–BISG concern for improving the supply chain; it is pragmatic and realistic—prioritizing simplicity over complexity in the interests of effecting productive change.

2. ISTC — how it works

ISTC is an international ISO Standard (21047) that was published in March 2009. The Standard is governed, in a manner similar to ISBN, by a new International ISTC Agency, which in turn licenses Registration Agencies to manage the allocation of the identifiers. However, none of these Registration Agencies has territorial exclusivity. Nielsen Book and Bowker operate Registration Agencies in the U.K. and the USA respectively.

MVB operates a Registration Agency in Germany, and numerous other organizations in Europe, Asia and Latin America have expressed interest in establishing Agency operations.

In addition to the published ISO Standard, the International ISTC Agency has issued documentation to support implementation: namely, a User Manual; and, with EDItEUR, a specification for an ONIX for ISTC Registration (hereafter ONIX-for-ISTC) suite of XML messages, to be used by registrants that submit requests for ISTC allocation electronically to a new global database—the Standard Text Registration System (STRS)—shared by all the Registration Agencies. Although electronic submission is preferred, it is not required; other input methods may also be available to registrants.

ISTCs identify the discrete results of creative/intellectual effort expressed as text—"creative works" such as prose, poetry, lyrics, original screenplays, audio and stage scripts. These works can include illustrations, of course, but works that consist predominantly of illustrations, such as books of photographs with minimal text, are not eligible for ISTC. Graphic versions of novels, comics, cartoons and manga are eligible, but the ISTC in these cases identifies only the text element, not the pictures.

Anyone (author, publisher, agent, library, rights registration organization, collecting society) can register a new work with any of the Registration Agencies. It is important only that the information about a work should be loaded by someone knowledgeable and responsible. The chosen Registration Agency simply applies an algorithmic computer test to the description of the work that has been submitted to it to establish whether the proposed new work has already been registered either by itself or by any of the other Registration Agencies. All ISTCs are held in a single global database—the Standard Text Registration System (STRS).

The ISTC identifier is a 16-character string consisting of four components, separated by hyphens. For example:

ISTC A02-2010-31F4CB2C-B

The first and third elements are "hexadecimal," allowing the numbers 0–9 plus the letters A–F to be used. The first element identifies the issuing Registration Agency. The second element is the year of registration. The third element is the actual unique identifier for the work; and the final element is a check-digit. This hexadecimal numbering allows for 4,096 Registration Agencies and for 4,294,967,296 discrete works to be registered per Registration Agency per year.

While an ISTC provides a mechanism for clustering products containing the same content (or, in some cases, different content with the same source or origin—see Section 5 below), it is important to recognize that the ISTC does not of itself have any element to describe its registrant. No component expresses any form of ownership. A classic work, *Pride and Prejudice*, for example, could be published in any number of formats or languages, by any number of publishers, but could be first registered in the database only as one original creative work.

Within the book industry supply chain at least (and quoting the User Manual), "the value of an ISTC number is . . . realized when applied as an attribute of a textual manifestation"—of the precise and unique work that it is identifying. This manifestation needs to be a textual manifestation, and it will

probably be a published format (hardback, paperback, audiobook, e-book, etc.), which in some or many cases will have an ISBN assigned as well as a product data record (in some cases ONIX-for-Books).

Thus the bibliographic principle is that an ISTC is an attribute within the descriptive product data of one or more ISBNs. The ISBNs that link to the same ISTC are not discoverable within the ISTC record. ISTC is not a supply-chain identifier for products you can buy. Rather, when utilized within cataloging environments, content management systems (CMS) and other book-discovery and e-commerce platforms, ISTCs can make the discovery process more precise. Instead of needing to know all the different titles under which the same work has been published, the user need only find one edition, from which the ISTC number can be used to pinpoint and link all alternative editions of the work during their search queries, particularly when a title is used during the query.

Accordingly, on one creative textual work with a unique ISTC (for example, *Stephenie Meyer: Twilight),* we can hang any number of textual manifestations (formats), each with its own ISBN. And the single ISTC should form part of the ONIX-for-Books product data for each of these (in this case ten) ISBNs:

ISBN 1	*hardback*
ISBN 2	*paperback*
ISBN 3	*film tie-in paperb*ack
ISBN 4	*large print*
ISBN 5	*library binding*
ISBN 6	*downloadable PDF e-book*
ISBN 7	*Kindle e-book*
ISBN 8	*Sony e-book*
ISBN 9	*unabridged downloadable audiobook*
ISBN 10	*unabridged CD audiobook*

It should be recognized that all these manifestations are of exactly the same text—the same sequence of words, characters and spaces, simply presented in different ways. It is not important, within an ISTC context, provided the words are the same, that a consumer might not consider an audio version and a print hardback as "equal;" or whether the author or an actor is doing the reading.

Through ISTC, any of these ten formats can be linked together as being "the same thing;" and they can be grouped together (= "collocated," clustered) for various purposes. Conversely, and just as usefully, ISTC can distinguish these from other products, perhaps with the same or similar titles or product descriptions, that appear to be the same but do not contain the same underlying textual work, and that accordingly have a different ISTC.

It should be noted here that the ISTC structure allows for manifestations to contain more than one ISTC if the product contains more than one discretely identifiable textual work element. The ISTC protocol also requires that any work registered must be designated as either an original work or a derived work. Within the ISTC metadata, any derived work must reference the source ISTC from which it is derived; and in addition it must state the nature of the derivation. Derived works accordingly link together two works (themselves and their parent source) and are discussed later in this paper.

It should also be noted that any supply-chain benefits of ISTC depend, as does so much of the book industry, on rigorous and accurate descriptive product and work-level metadata (catalog information and other attributes)—and metadata in the prescribed industry-standard ONIX-for-Books format. Full benefit depends on publishers providing all product formats with well-formed ONIX-for-Books product metadata, and for this product metadata to contain the appropriate ISTCs. The new version of ONIX- for-Books (v. 3.0) has been specifically revised to accommodate ISTC functionality; ISTCs can be loaded as repeatable fields. Maximum benefit further depends on publishers or others providing well-formed ONIX-for-ISTC metadata when applying to register new works.

Some Registration Agencies may provide other, less technically demanding, nonelectronic means for submitting registration requests and loading the associated metadata. Despite this, dependency on ONIX messages—both ONIX-for-Books and ONIX-for-ISTC—may tend to disenfranchise smaller publishers not yet equipped to manage ONIX data.

Nielsen and Bowker have undertaken to provide a public Internet search utility to access the global STRS database. This database, of course, only contains ISTCs—the identifiers for the individual works, and their descriptive metadata. Users from the book industry, authors, librarians and consumers will be able to search the database to discover a work's correct ISTC, using, for example, title or author as search terms. Registrants of prospective new products will be able to establish whether the underlying work slated for publication has or has not been previously registered.

What these users won't be able to do within the STRS database is to search for, or get results for, ISBN-based product records. Users will not, for example, be able to search for *Twilight* (see above) to get a listing of all the products (ten manifestations in this example) that contain this work.

The database record may still contain some references to manifestation ISBNs. There are optional fields into which registrants can load structured ISBN data; and there is an additional free-form field that registrants can use in any way they like. The main purpose of these, however, is to help the Registration Agency clarify the uniqueness of the new submission at the point of initial registration.

The pared-down metadata set held within the STRS against an ISTC has been set at the absolute minimum required to avoid duplicate registrations. It would simply not be possible, once the ISTC record had been created, to maintain references to all the linked formats within the same database. The originating registrant might at any time be adding new products; and others—agents, other publishers, or licensees, perhaps—might also be adding products (for example, translations) of which the originating registrant was unaware. Mechanisms will exist for the dynamic correction (where necessary) and enhancement of ISTC records and additional linkages. But, simply put, once

the record is created, no one "owns" the ISTC, and nothing in the ISTC record implies "ownership" or rights either.

The management, in any comprehensive way, of the link from ISTC to ISBN—and vice-versa, from ISBN to ISTC—will be principally in the control of the bibliographic services, of the licensing and collecting societies, and of the publishers themselves (though restricted only to the products they themselves control).

3.   Beneficiaries and benefits

Potential benefits from ISTC have been claimed for a comprehensive range of stakeholders within the book industry: authors, agents and authors' representation bodies; publishers and rights-holders; resellers (retail, wholesale, distributors); libraries; and other services—bibliographic and product information, licensing, rights registries.

The benefits can be summarized under the following five headings, each of which will be discussed in more detail within the next section.

A. Collocation and differentiation for search and discovery

As an identifier applied to all manifestations (products, formats, editions) of the same underlying text, ISTC enables all these to be grouped together (collocated, clustered) within a search result (see the *Twilight* example above). All the user needs to do is find one product that is the one they need, then to check out the linked ISTC in that product's descriptive metadata in order to be able to find all the other formats that contain the identical textual work—while excluding any products that may have the same or similar descriptions, but whose text is different.

Collocation based on ISTC also enables discovery of other products with identical content that may for some reason have quite different descriptive product metadata, or even a different title or name.

B. Collocation for sales and library loans reporting and analysis

New technologies and opportunities allow publishers to make their content available in an increasing number of ways—physical (in different paperback formats, library editions, large-print, film tie-in); for audio (in any number of formats, downloadable or on CD); and for digital (online and offline, to e-book readers, mobile phones and an infinity of device-dependent formats, with any number of DRM packages restricting functionality or sharing).

Libraries have expressed a strong need for aggregated data to help them separate multiple formats for lending. Further, they may wish to see their loans differentiated to the finest possible degree of granularity (for example, differentiating between models of handheld e-book–reading device or mobile phone, rather than lumping these together). They, and the compilers of trade bestseller lists, journalists and commentators, may also need the sort of data, aggregated under the heading of the same underlying work content that ISTC can provide.

C. Differentiation and rights management

In an increasingly global Internet marketplace, retailers can often inadvertently infringe territorial rights agreements by listing and selling products (editions) into markets where sales are not permitted. Comprehensive ISTC registration might enable resellers serving international markets to establish which were the available territory-specific editions of any particular work in order to select only those with appropriate market rights. Or this might involve retailers using ISTC to identify infringing products if they had an assertion of exclusive rights from a party that they trusted.

ISTC registration might permit the correct identification of derived works such as translations, that might otherwise not be discoverable by searches on title or other original-language descriptive data.

ISTC collocation might support other rights-based activities. The rights-registration organizations and authors' and artists' collection societies all have an imperative to cluster products together accurately for the purposes of analysis and for the distribution of revenue.

The activities of the new rights registries, and in particular the Book Rights Registry (if established) in the USA and ARROW in Europe (Accessible Registries of Rights Information and Orphan Works towards Europeana), will be of particular importance.

Rights subsist in works, not manifestations, and both these schemes (and others involved in similar rights-management exercises) will need to identify whether any manifestation of a work remains commercially available. Rights registries will need to design clustering algorithms to meet their own specific rights-management requirements. An important requirement will be the need to communicate with other rights-management organizations and third parties. There can be no doubt that a standard work identifier of some kind will arise to meet this need. The question that the new rights registries are not yet in a position to answer is whether ISTC and its metadata set can fully meet their requirements.

D. Cataloging and library collection management

Uniform title control—a mechanism either to establish which products (books, for example) contain essentially identical content or to identify revisions or different editions of earlier publications—has always been a feature of library catalog management. After all, libraries need, from the descriptive metadata available to them, to avoid duplicate orders of works they have already obtained. This might be at the simple level of a hardback and its paperback edition, published subsequently, perhaps under a different imprint or by a different publisher; or it might involve more complex cases of multiple digital versioning. Uniform title control is managed at a comprehensive level by central cataloging from the U.S. Library of Congress, the British Library and library services providers such as OCLC. While there are a number of accepted standards within the library community, all are essentially secondary, and their precise treatment of works, manifestations and versions is variable, sometimes inconsistent and often necessarily subjective. Determination of the correct work/product hierarchy at source, by the actual creator or publisher of the work, as provided by ISTC (with its rigorous rules and taxonomy) will in principle always provide a more accurate and validated result.

For libraries, a parallel concern with authority control is author-name control, particularly in cases where authors have common names, or employ variant names or pseudonyms. Author-name control for libraries is also currently managed, largely manually, through central cataloging mechanisms (LOC, etc., as above). More rigorous control of work and product relationships, enabled by ISTC, will undoubtedly support improved author-name control. In the longer term it is hoped that ISNI—the International Standard Name Identifier (draft ISO Standard ISO 27729)—will better meet these requirements through registration at the source of creation. The ONIX-for-ISTC metadata set allows for an assigned ISNI for authors and contributors.

E. Internal systems (principally publishers, authors' agents, etc.)

As an ISO standard, ISTC is mainly focused on providing public benefit—for public discovery, and for improving product knowledge within the supply chain between trading partners. It has also been suggested that, over time, ISTC may also provide some benefit within publishers' own internal financial and administrative systems wherever there is a requirement to aggregate products.

Multiple products already sit under a single author's contract for royalty accounting purposes; or are combined within a purchase ledger for production budgeting; or are rolled up for sales reporting and analysis, or for the payment of sales commission.

Publishers also, for their own marketing purposes, need to be able to collocate (cluster) product information into "title"-based (= work-based) sets for marketing, whether on their proprietary website catalog pages, for order forms or advance information sheets, or for other leaflets or print/email publicity.

In this connection, publishers have been concerned with the metadata "bloat" consequent on the requirement to allocate separate ISBNs to each individual tradable product format, including in many cases a proliferation of e-book and audiobook formats. The need to maintain almost identical descriptive metadata sets at ISBN level in ONIX-for-Books (which at present can operate only at ISBN level) clearly makes for considerable redundancy and duplication, and adds a significant data-management overhead.

To resolve these issues, publishers are already using their own work or project identifiers to establish clustered relationships (via metadata) within their ISBN records. Some use of ISTC, either on a mapping basis, or directly in place of these proprietary identifiers, may provide opportunities for improved efficiencies. If a common standard is deployed, new system modules, whether homegrown or off-the-shelf, become easier to integrate with preexisting data structures. Systems and interfaces become interoperable, and data can be more readily shared and reused.

4.   Bibliographic functionality

Examples and use cases help explain in greater detail the thinking behind ISTC.

Most bibliographic discovery starts with search, within a search engine environment or on a retail or publisher bookstore platform. ISTC offers the potential for better search, which is simply defined as

getting the results you want, without getting results you don't want. Thus a sixth beneficiary group in the previous section might equally have been the consumer (user, book buyer, researcher, student, librarian).

Most searches start with an attempt to search on title or author, and in many cases this will provide results that at least include all the products the user may be interested in. However, the content of products with matching title and/or author may not in fact be precisely the same (that is, substitutable, differing *only* in format). Such matching author/title combinations may represent something different (for example, an adapted screenplay, or an abridgement). All the matches may not be in the format required or accessible (through the e-book device or phone the user has; or because of DRM/rights/territorial restrictions). And, of course, not infrequently, the same content may be available under a completely different title (perhaps as part of a collection) that won't be found in a full-text title-author search.

Take the classic *Moby-Dick* example. A search for "Moby-Dick" on Amazon finds 19,626 results (products/formats in "Books"), presented in a prioritized order known only to Amazon.

Some of these will be *sort of what you wanted*, that is, if you just wanted the classic novel:

ISBN 1	Moby-Dick Penguin Popular Classics
ISBN 2	Moby-Dick Wordsworth Classics
ISBN 3	Moby-Dick Norton Critical Editions
ISBN 4	Moby-Dick Oxford World's Classics
ISBN 5	Moby-Dick Barnes & Noble Classics
ISBN 6	Moby-Dick Dover Giant Thrift Editions
ISBN 7	Moby-Dick Penguin Classics Deluxe
ISBN 8	Moby-Dick Longman Critical Editions
ISBN 9	Moby-Dick Modern Library Classics
ISBN 10	Moby-Dick Easy Read Large Print Edition

But at the same time you'll get good matches that (worthy as they no doubt are) *aren't quite what you wanted*:

ISBN 11	Moby-Dick. Cliff's Notes
ISBN 12	A Pop-up Moby-Dick
ISBN 13	Moby-Dick. Naxos Audiobook
ISBN 14	Moby-Dick. Shmoop Study Guide
ISBN 15	Moby-Dick The Screenplay

ISBN 16	Moby-Dick for the Kindle
ISBN 17	Moby-Dick: The Good Parts
ISBN 18	Moby-Dick: In Half the Time
ISBN 19	Marvel Illustrated Presents Moby-Dick
ISBN 20	Classic Comics: Moby-Dick

And some equally good matches (on the same text, "Moby-Dick") that *aren't what you wanted at all*:

ISBN 21	New Essays on Moby-Dick
ISBN 22	Melville's Moby-Dick: An American Nekyia—A Jungian Commentary
ISBN 23	In Search of Moby Dick: The Quest for the White Whale
ISBN 24	Two Guys Read Moby-Dick
ISBN 25	Moby-Dick or The Catfish—A Trailer Park Classic
ISBN 26	Whaleship Essex: The True Story of Moby-Dick
ISBN 27	Moby-Dick and the Whaling Industry in the Nineteenth Century
ISBN 28	Frank Stella's Moby-Dick: Words and Shape
ISBN 29	The Errant Art of Moby-Dick: The Canon, the Cold War and the
ISBN 30	Ungraspable Phantom: Essays on Moby-Dick
ISBN 31	Commodore Perry as White Phantom: Moby-Dick in the Context of…

Amazon will not, of course, have presented these results in as orderly a fashion as in the three sections above. Required and not-required items will be sequenced almost randomly. And the above selection still leaves some 19,500 "Moby-Dick" results to go….

Once the core classic text *Moby-Dick* has been registered as an ISTC (let's call it ISTC A), however, this ISTC will be contained only within the ONIX-for-Books records of those products (mostly in the first section in the above example), which contain the full text of the core classic work. This might eventually open up some possibilities for targeted Advanced Search—limiting the found selection to include only products containing ISTC A.

In any consideration of ISTC, the critical point to bear in mind is that its principal value is its inherent accuracy, validated in most cases by the fact that registration will have been made at source, upstream at "creation"—by the author, the author's agent or a publisher.

Amazon, as shown above, other retailers, and library service providers have worked up sophisticated algorithmic techniques to provide collocation and uniform title control. Google Books, on the other hand, retains traditional search on full-text and metadata; its results on the term *Moby-Dick* appear every bit as undifferentiated and randomly sequenced as Amazon's. Library service providers and

the bibliographic companies principally provide services based on a traditional ISBN-based product catalog, rarely with any comprehensive collocation by work or "title."

The ISTC User Manual has a section that describes in detail the process by which registration requests to the Standard Text Registration System (STRS) are compared with records for works already registered to establish whether a new ISTC should be allocated. This process performs a number of comparison tests of elements (title, author, etc.) to see how many exact or fuzzy matches are present.

Bowker and Nielsen Book, in addition to their roles as Registration Agencies, also provide comprehensive bibliographic services based on the receipt of ISBN-based records from publishers. They are therefore in a position to reverse-engineer groupings of titles by manually enhancing any ISTC-specific metadata and using the STRS comparison methodology to back-create registration records for ISTCs. Both Bowker and Nielsen are actively engaged in publisher pilots to validate requirements and workflow assumptions. ISTCs have been assigned to sample sections of these pilot publishers' lists, and to some public domain works. Such pump-priming of the STRS database is essential. Bowker has also announced the intention to preassign ISTCs to some commercially available works from U.S. publishers. This involves using proprietary title-linking algorithms to experiment with the automated creation of ONIX-for-ISTC metadata out of ONIX-for-Books files. A very significant amount of additional manual work is required to clean up the data and deal with missing fields.

It is also likely that rights organizations and the new rights registries will also engage, perhaps on a very significant scale, in the systematic registration of ISTCs for backlist and orphan works.

While this necessary developmental pilot work is valuable, there is certainly a risk of its compromising the fundamental principle that ISTC must be knowledge-based and creator-managed to provide its full benefits. Certainly there will be inaccuracies, and some refinements (for example, the relationships between works, derivations, etc.) may be sacrificed if all works (even if potentially derived) are more easily registered as new. But this may not be too high a price to pay if it encourages the development of interest and documented use cases from all these parties—retailers, rights registries, data licensees who are the ISTC's potential "consumers."

And the Registration Agencies would, however, quite correctly maintain that no works have so far been registered without the Registration Agencies having total confidence that they have brought to bear a degree of precision at least as great as the majority of publishers put into their own product metadata.

5.   Derived works and substitutability

In the preamble to this paper, ISTC was described as a sophisticated and flexible standard with multilayered capabilities. The designers of the standard have allowed for a complex hierarchy of relationships between works; in fact, they have probably correctly allowed for all possible interrelationships.

A careful review of the *Moby-Dick* example above shows that there are three categories of result. The first of these, mainly in the first grouped list of ISBNs, is "the right result" for anyone searching for a full-text copy of *Moby-Dick* to read. These will be more or less identical texts, with the same characters and spaces in the same order, and are thus substitutable means of accessing the core work.

The second category, mainly in the third grouped list of ISBNs, clearly contains mis-hits. These are books referencing or "about" *Moby-Dick* the novel, but they do not contain the text, and they provide no substitute for the real thing. Since the core work *Moby-Dick* is not directly available as content within these, each of these products will have an individual and different ISTC as works in their own right. The ISTC protocol does not extend to secondary or critically referencing works, so it won't be possible within ISTC, to make the "*Moby-Dick* connection." Other than, of course, by the fact that the string = *Moby-Dick* is in the product's title.

A third category, however, mainly grouped in the second list, may well be of interest to someone searching for a version of *Moby-Dick* to read.

For example:

ISBN 12	ISTC B	*A Pop-up Moby-Dick*
ISBN 15	ISTC C	*Moby-Dick The Screenplay*
ISBN 17	ISTC D	*Moby-Dick: The Good Parts*
ISBN 18	ISTC E	*Moby-Dick: In Half the Time*
ISBN 19	ISTC F	*Marvel Illustrated Presents Moby-Dick*
ISBN 20	ISTC G	*Classic Comics: Moby-Dick*

These clearly won't provide the reader with 1,000 pages of unadulterated Melville, but they are nonetheless versions of the core text, without being directly substitutable manifestations of it. And they are clearly creative works in their own right, each entitled to its own ISTC.

The ISTC protocol therefore requires that any work registered must be designated by the registrant as either an *original* work or a *derived* work. Within the ISTC metadata, any derived work must reference the source ISTC from which it is derived; and in addition, it must state the nature of the derivation. The following derivation types are allowed:

Abridged

Annotated

Compilation

Critical

Excerpt

Expurgated

Nontext added or revised

Revised

Translated

Adaptation

Unspecified

Thus any search on *Moby-Dick* might wish to be restricted just to the core classic text (ISTC As—see above), or additionally to take in derived versions (ISTCs B–G). And, of course, for any of these ISTCs (for example, the Screenplay ISTC C), there might be any number of different manifestations—ISBN-based products—hardback, paperback, e-book, etc.

Since ISTC is restricted to textual content, it may also be helpful in selecting out products without text that may have ISBNs (possibly incorrectly) or SKUs within bookselling environments.

Thus (again using "ISTC A" to define the underlying original textual work):

ISTC A   Meyer: Twilight

One core textual work, thus one ISTC.

273 results for = Meyer + Twilight in Books on Amazon

Many manifestations ("formats" at Amazon), all with just the one ISTC A (see also a screen grab shown later in this paper):

ISBN 1   *hardback*

ISBN 2   *paperback*

ISBN 3   *film tie-in paperbac*k

ISBN 4   *large print*

ISBN 5   *library binding*

ISBN 6   *downloadable PDF e-book*

ISBN 7   *Kindle e-book*

ISBN 8   *Sony e-book*

ISBN 9   *downloadable audiobook (unabridged)*

ISBN 10   *CD audiobook (unabridged)*

Filters out unconnected and underived works, each with many ISBNs/manifestations:

    ISTC B   *Meyer: Eclipse (Twilight Saga 3)*

    ISTC C   *Meyer: New Moon (Twilight Saga 2)*

    ISTC D   *Twilight Illustrated Movie Companion*

    ISTC E   *Twilight and Philosophy: Vampires and Vegetarians*

    ISTC F   *Stephenie Meyer: Author of the Twilight Saga*

Removes other products, some of which may have ISBNs:

    NO ISTC   *Twilight Movie DVD*

    NO ISTC   *Twilight Board Game*

    NO ISTC   *Twilight Poster*

    NO ISTC   *Twilight Bella Swan toy/mug/key-ring/choker*

In fact, for key titles such as Stephenie Meyer's *Twilight*, Amazon will have worked hard to provide as complete and "uniform" a catalog page as it can. But this will almost certainly be the result of manual intervention. It is likely that most of the formats offered below in this screen dump are genuinely and substitutably the same. However, ISTC, managed by the publisher Little, Brown, would necessarily have provided a more validated picture of the different manifestation options. In this case, it is at least likely that some of the audio formats listed do not represent a complete and unabridged text. With ISTC functionality, an abridged audiobook would be a separate derived ISTC, linked to the source/parent ISTC A.

Stephenie Meyer; Twilight at Amazon 12 January 2010

Formats	Amazon Price	New from	Used from
Kindle Edition, March 22, 2007	$5.12	--	--
Hardcover, October 4, 2005	$11.69	$9.50	$6.21
Hardcover, August 31, 2006	$15.48	$14.59	$14.95
Library Binding, August 31, 2006	$17.24	$17.24	$14.65
Library Binding, April 17, 2008	$19.99	$19.99	$19.99
Hardcover, Large Print	$20.76	$20.76	$12.20
Hardcover, November 17, 2008	--	$19.50	$9.75
Hardcover, September 30, 2005	--	$29.83	--
Hardcover, December 31, 2004	--	$29.99	$7.75
Hardcover, Import	--	--	$35.76
Paperback, September 5, 2006	$5.50	$4.50	$2.00
Paperback, October 27, 2008	$7.91	$2.70	$0.09
Paperback, Large Print, November 30, 2008	$9.43	$8.96	$8.94
Paperback, December 31, 2004	--	$3.99	$1.50
Paperback, Import, April 29, 2007	--	--	$4.92
Paperback, December 31, 2007	--	--	$10.00
Paperback, Import, December 31, 2005	--	--	$12.59
Paperback	--	--	$15.95
Paperback, Large Print, May 3, 2009	--	--	$97.17
Mass Market Paperback, October 31, 2008	$7.99	$1.69	$0.15
Audio, CD, Audiobook, Unabridged	$19.79	$15.98	$15.97
Audio, CD, November 30, 2005	--	$22.50	$20.68
Audio, Cassette, November 30, 2005	--	--	--
Unknown Binding, September 30, 2008	--	$5.00	$80.86
Audio, Download	$25.46 or less with new Audible membership		

6.   Translations and rights

As mentioned above, comprehensive ISTC registration would enable resellers serving international markets to establish the available territory-specific editions of any particular work, in order to select only those with appropriate market rights. Thus an ISTC search on products with ISTC A = *Stephenie Meyer; Twilight*, and format = "paperback" might find two manifestations with identical content.

> ISBN 1 Little, Brown (Megan Tinley) USA and Canada paperback edition (ISTC A)

> ISBN 2 Little, Brown (ATOM) U.K. and Europe paperback edition (ISTC A)

In this case, while ISTC has correctly brought together two products with the same core textual work content, these two ISBNs should have their specific and different territorial restrictions defined within their ONIX-for-Books metadata. Thus a retailer, finding two manifestations of "the same thing," can determine which it is legally permitted to sell.

Such refined use of ISTC depends, however, on the generator of the ISBN-based ONIX-for-Books record provided to the bibliographic services having correctly populated the ONIX-for-Books territorial rights elements, and for the bibliographic services to have distributed, and the retailer to be have been able to process, this distinction. These are big dependencies.

It has been suggested that future rights deals might see the originating (licensing) publisher or agent insisting on a "best endeavors" clause. Such a clause might require the licensee to use an ISTC specified by the licensor in all its ONIX-for-Books and other bibliographic statements and messages.

An alternative scenario might have retailers relying systematically on an assertion of exclusive rights from only one or another of the parties (provided only that it was a trusted party), without needing all the other rights statements to be present.

7.   Foreign-language editions

The ISTC protocol differs from some other work/manifestation systems (particularly within the library sector) by treating foreign-language translations as derived works from an original source/parent ISTC, even though the actual full text (in terms of characters and spaces) is clearly not the same or directly substitutable.

This functionality might be useful, as in most cases the translated title gives no precise clue as to which core work has been translated. A case in point might again be the *Twilight* series, where foreign-language versions of the first three books have been given unrelated titles, which are not direct translations of the English (and note the German puns). In each of these cases, the foreign-language translation has its own derived ISTC, but it will reference the source ISTC of the original English-language work.

Thus:

ISTC A (1st in series)	*Twilight*	*EN*
ISTC A + ISTC B	*Fascination*	*FR*
ISTC A + ISTC C	*Bis(s) zum Morgengrauen*	*DE*
ISTC A + ISTC D	*Crepusculo*	*ES*
ISTC E (2nd in series)	*New Moon*	*EN*
ISTC E + ISTC F	*Tentation*	*FR*
ISTC E + ISTC G	*Bis(s) zum Mittagsstunde*	*DE*
ISTC E + ISTC H	*Luna Nueva*	*ES*
ISTC I (3rd in series)	*Eclipse*	*EN*
ISTC I + ISTC J	*Hesitation*	*FR*
ISTC I + ISTC K	*Bis(s) zum Abendrot*	*DE*
ISTC I + ISTC L	*Eclipse*	*ES*

Within the system, each of these translated new derived works will have any number of hardback/paperback/e-book ISBN manifestations. And some may already have spawned further derivations (abridgements, comic-strip versions) of their own, with new ISTCs, which in turn would reference the two higher levels of parent and grandparent works.

This is a fine example of ISTC functionality providing a complex and robust solution that anticipates a metadata environment which might seem, at present at least, somewhat fancifully remote. Only the creators/publishers of such complex derivations would be in a position to register these new ISTCs, and it is not clear what would motivate them to do so. Would they derive sufficient benefit from referrals from (or linkages to) their originating "ancestors?" One might simply argue that the ISTC framework will be fit and ready for such purposes should the felt need arise.

(The *Eclipse* example above additionally demonstrates the potential—without ISTC identification—for "identical title"–based confusion between the original and the Spanish edition of Book 3 on sites, such as those in the USA, that carry both language versions.)

8.   How many layers of ISTC?

It will be evident from the two previous sections why publishers might appear uncertain about the level at which they should be engaging with ISTC. Publishers' concerns focus not only on what and how much to do first in terms of ISTC registration, but also on uncertainty about levels of granularity and the complexity of possible hierarchies.

The principle of functional granularity is frequently advanced in discussions of book industry standards and identifiers. Loosely applied, this principle dictates that objects need be identified only down to a level at which they practically and usefully need to be and can be identified. An example of this thinking follows, relating to the treatment of classic texts, using *Moby-Dick* once again. In some respects, this is the debate about lumping versus splitting familiar to zoologists defining species.

This next section is designed simply to demonstrate potential complexity. The examples are entirely hypothetical.

There are various versions of full-text *Moby-Dick* in the first section:

ISBN 1	*Moby-Dick Penguin Popular Classics*
ISBN 2	*Moby-Dick Wordsworth Classics*
ISBN 3	*Moby-Dick Norton Critical Editions*
ISBN 4	*Moby-Dick Oxford World's Classics*
ISBN 5	*Moby-Dick Barnes & Noble Classics*
ISBN 6	*Moby-Dick Dover Giant Thrift Editions*
ISBN 7	*Moby-Dick Penguin Classics Deluxe*
ISBN 8	*Moby-Dick Longman Critical Editions*
ISBN 9	*Moby-Dick Modern Library Classics*
ISBN 10	*Moby-Dick Easy Read Large Print Edition*

Alternative 1

These are clearly not all identical, and they are only to some degree substitutable. Some will have validated and critically researched texts, based on the best and most recent scholarship; others will be public domain or Project Gutenberg versions with less scholarly authority. Basically, these versions will not have all the same words in the same order. Some will provide just a raw text; others will have any amount of additional apparatus (notes, bibliographies, chronologies, introductions, etc.). Nonetheless, all this difference could perhaps be ignored in favor of a simple lumping. All the above are basically the original core work, so Alternative 1 codes them all as *Moby-Dick* ISTC A (continuing this shorthand designation).

Alternative 2

A Melville scholar might argue, by contrast, that all these are modified and dissimilar derivations of an original parent *Moby-Dick* as first published (actually as *The Whale*) in various editions in England and the USA in 1851. It doesn't matter if that particular *Ur*-parent has no present-day precise product manifestations.

ISTC A   *The original source [parent] work Moby-Dick*

Derived works

ISTC B   *Moby-Dick Penguin Popular Classics*

ISTC C   *Moby-Dick Wordsworth Classics*

ISTC D   *Moby-Dick Norton Critical Editions*

ISTC E   *Moby-Dick Oxford World's Classics*

ISTC F   *Moby-Dick Barnes & Noble Classics*

ISTC G   *Moby-Dick Dover Giant Thrift Editions*

Of course, all these derived editions would reference ISTC A as their source.

Alternative 3

Meanwhile, a librarian might argue that some are the same text and some are different. The structure below takes the hypothetical view that Penguin has licensed the Oxford text for its paperbacks; that Wordsworth, Dover, and Barnes & Noble have used Gutenberg; and that Norton has developed its own text.

ISTC A   *The original source [parent] work Moby-Dick*

Derived works

ISTC B   *Moby-Dick Penguin Popular Classics*

ISTC B   *Moby-Dick Oxford World's Classics*

ISTC C   *Moby-Dick Wordsworth Classics*

ISTC C   *Moby-Dick Barnes & Noble Classics*

ISTC C   *Moby-Dick Dover Giant Thrift Editions*

ISTC D   *Moby-Dick Norton Critical Editions*

Alternative 4

Meanwhile, another more pedantic librarian, dancing angels on the head of a pin, might maintain that this wasn't anywhere near complicated enough:

ISTC A   The original source [parent] work Moby-Dick

First derived work: Penguin Popular Classics

ISTC B   ISTC for Compilation of the three elements below (derived work)

ISTC C	Component: OUP text (derived work)
ISTC D	Component: Introduction by A Scribbler (original work)
ISTC E	Component: Notes by B F Rubbishwriter (original work)
Second derived work:	Oxford World's Classics
ISTC F	ISTC for Compilation of the five elements below (derived work)
ISTC C	Component: OUP text (derived work)
ISTC G	Component: Introduction by B Spears (original work)
ISTC H	Component: Biographical Note by H M Biographer (original work)
ISTC I	Component: ''Whales and Whaling'' by J Starbuck (original work)

And so on . . . At an even finer level of granularity, ISTCs could additionally be applied to each chapter or paragraph of the underlying textual work.

One hundred percent in compliance with all the ISTC rules, but . . .

Any recommendation from this paper should focus primarily on the utility of a best alternative. What then would be most helpful to most users of the functionality? In this particular case, there can be no question that it is Alternative 2. These individual *Moby-Dick* products will all have a generic reference to a source ISTC A—the parent original. So searching for that will find them all. But at the same time, a user looking for different manifestations of just the Penguin Classics version (with all its proprietary apparatus), in paperback or in different audio or e-book formats, could collocate these by applying ISTC B.

Similar issues will arise over nonfiction new editions and successive revisions. Publishers and other registrants will need to take a practical and commercial view of what they want ISTC to achieve for them, in each particular case, rather than feel tied to any particular purist methodology. In some cases, new revised editions (of a college text, for example) are more usefully deemed to be "the same;" in others (less directly substitutable) they may be sufficiently different (and worth differentiating) to require a new derived-work ISTC, referencing an original. This process resembles the current subjective determination as to whether a new edition product requires a new ISBN.

9.   Practical aspects of ISTC implementation

"The pressing need for a standard tool for linking different printed and digital manifestations to a common title makes the ISTC potentially the most important standard for our industry since the ISBN."

—BIC/BISG WHITE PAPER ON IDENTIFIERS, 2008

ISTC holds significant potential to revolutionize search and discovery in the increasingly multiproduct, fragmented, complex new world of digital publishing. In addition, ISTC can provide

collateral benefit within a wide range of other sectors within the book industry—sales analysis, rights, royalties, authors' contracts, license management.

This paper has concentrated on the supply-chain perspective. To quote the User Manual: "The most reliable method for accurately relating works to their manifestations is for these linkages to be created in a bibliographic (that is, manifestation) record by the publisher each time a new manifestation is defined. This recognizes that the publisher is the most authoritative source of information on what content is in a particular publication."

Publishers should be in no doubt that there is a strongly felt need by leading retailers, both online and bricks-and-mortar, to see ISTCs on the product records they receive from publishers. Enlightened retailers have always known that better bibliographic data sells more books.

Rights organizations and the Registration Agencies themselves will work together on the registration of classic, public domain, backlist, and orphan works.

If the ISTC is seen to meet their needs, adoption by the new rights registries (and in particular the BRR and ARROW) will be critically important. These registries will have a pressing need for a standard work identifier and will have the capability to register very large numbers of works.

But it will be publishers who must accept the challenge of taking forward the implementation of ISTC for new titles. They will do so first by registering their creative textual works to obtain an ISTC number; and then by loading ISTCs appropriately and accurately within the ISBN-based ONIX-for-Books data feeds they provide, via the bibliographic service companies, to the wider book trade community.

There are at present no open-and-shut proven ROI or financial benefit statements available to persuade publishers to invest in ISTC implementation. But with an eye to future benefits, there is equally no reason why publishers should not take a simple, pared-down, pragmatic, gradualist, and tentative approach to ISTC implementation, which might not imply much in the way of investment.

A critical inflection point for the industry will be whether it can wait for a purist front-loaded adoption of ISTC to build up to critical mass; or whether wholesale automated retrospective registration of backlist, however flawed, might provide greater benefit.

Given the complexity of the ISTC framework, and its dependence on electronic messaging, it is unlikely that authors or their agents will register works for ISTC. Even though some Registration Agencies may provide other, less technically demanding, means for submitting registration requests and loading the associated metadata, authors and agents will perceive registration as the publisher's job. Even if authors and agents were to do so, the critical complementary task of ISTC-to-ISBN loading and data transmission can best be fulfilled by the publisher as product manifestations are prepared for publication.

It is unlikely that many retailers will be in any position to exploit ISTC until the bibliographic-services businesses can provide ISTC-based (= collocated work–based) data services to them. But there is no doubt that the Registration Agencies see ISTC as enabling a range of enhanced services and data

functionality that will better meet the needs of their customers— retail, library and publisher— across a range of disparate products, such as improved discovery for retailers and libraries, better sales analysis for publishers, better bestseller charts and so on. Plans for such enhancements certainly exist, and the Registration Agencies continue to invest upfront in ISTC development. The critical issue for any payback must remain critical mass. That is why Registration Agencies have announced that they will not make any charge for the registration of ISTCs at least in the immediate future.

In the early days of ISBN, some powerful retailers were in a position to insist upon mandatory compliance in the interests of clear-cut distribution and supply-chain utility. This coercion massively accelerated take-up by publishers. Such a situation is unlikely to be repeated with ISTC today. Some vendors may be reluctant to give up the perceived competitive advantage of their present albeit makeshift proprietary collocation mechanisms. On the other hand, other retailers may see this as a new challenge. It has also been suggested that ISTC adoption might be brought into consideration for industry bibliographic accreditation schemes.

It is unlikely that ISTC will replace other keys as an effective high-level internal identifier within publishers' finance and royalty departments for contracts. In fact, this would carry some risk. Publishers should be clear that ISTC is not in and of itself a rights identifier. Products (ISBNs) requiring collocation under a particular publishing agreement with an author will very often have quite different textual content. For example, the terms of one contract may cover a multiplicity, not just of manifestations, but of derivatives—abridgements, workbooks, ancillary Web material, licenses, collections. Nonetheless, contract handling would self-evidently be improved if it was based on works rather than on products/manifestations.

It is equally unlikely that ISTC will replace other overarching keys as an internal collocator of spending on a production ledger, or on revenues on a sales ledger. Publishers should be clear that

ISTC is not a "job," "title" or "project" identifier. The decision to apply a common and aggregated budget or ledger line to a group of products does not imply that their textual content is identical. Very often, it will not be.

10. Recommendations

- **Publishers** should aim first at deploying ISTC to improve discovery and provide more accurate uniform title authority and collocation within the supply chain.

    Publishers should perceive ISTC registration as a component of their product-based (ONIX-for-Books) metadata processes, not as may have been suggested (in a purer sense) upstream at the author's contract creation stage.

    It makes the best sense for publishers to build ISTC allocation—as a modest additional routine—into their advanced marketing and biblio-feed activities. Determination of the appropriate ISTC hierarchy should be a responsibility of publishers' marketing. It should be based on a sympathetically pragmatic approach as to how consumers/users approach search

and discovery. And this ISTC granularity should be managed at the lowest and most practical book-only based level.

- **Booksellers** and bookseller associations should be urged to articulate and document use and benefit cases for ISTC adoption in the retail environment.

  Half of the consortium which designed the ISTC represents **rights and collection societies**. They have invested almost as heavily as the new Registration Agencies. Their motives, and their aspirations for ISTC, fall somewhat outside the scope of this paper, but they, and the new rights registries (BRR, if established, and ARROW), should be encouraged to articulate and communicate to publishers and rights owners their view of the role ISTC should play within their universe.

- **Vendors of bibliographic and marketing software solutions** to publishers should be encouraged to integrate ISTC alongside their existing mechanisms for providing keys to aggregate products/ISBNs under "titles" within bibliographic, royalty, production and sales analysis modules. Such experimentation may establish over time the extent to which ISTC can take over some of this functionality, if, as is likely, most systems have fewer such keys than potential ISTCs. Keys might over time become aggregations of ISTCs instead.

- **Registration Agencies** should recognize that any wholesale registration of ISTCs for the commercially available Books-in-Print catalog might prove a disincentive for publishers to take ownership of the process, at least for new titles. Instead they should concentrate on working with selected publishers (particularly trade houses) on selected parts of their publishing programs.

  In addition to advising publishers on how to get started on frontlist registration, Registration Agencies might be advised to encourage publishers to register at least their bestselling backlist. It has also been suggested that ISTC linkage might most usefully be prioritized for such new titles as demonstrate strong initial e-book sales. For such titles, both retailers and publishers have an interest in aggregating multiple print and digital products in order better to understand the new dynamics of E vs. P cannibalization, and the implications of digital substitution for print runs and physical retail inventory. There will be other examples.

Resources

ISTC
http://www.istc-international.org/

ISTC User Manual
http://www.istc-international.org/multimedia/pdfs/ISTC_User_Manual_2009v1.0.pdf

ONIX for ISTC Registration
http://www.editeur.org/files/ONIX%20for%20ISTC/ONIX-ISTC%20overview%20v1.0.pdf

ONIX for ISTC Registration XML Schema
http://resource.bowker.com/istc/ONIX_ISTC_Metadata.1.0.xsd

ONIX for Books 3.0
http://www.editeur.org/93/Release-3.0-Downloads/

BIC/BISG Paper on Identifiers
http://www.bic.org.uk/files/pdfs/identification-digibook.pdf

ISNI
http://www.isni.org/