ISO/IEC JTC 1/SC 32 N 0949

Date: 2003-01-21

REPLACES: --

ISO/IEC JTC 1/SC 32

Data Management and Interchange

Secretariat: United States of America (ANSI)
Administered by Pacific Northwest National Laboratory on behalf of ANSI

DOCUMENT TYPE	Other Document (Open)
TITLE	(Draft) Applying an ISO/IEC 18022 Model Approach to Dublin Core as an IT-enabled Coded Domain with Multilingual Equivalency: A Business Case Study - Canadian Federal Government
SOURCE	Dr. Jake V.Th. Knoppers and M. Janice Pereira (Canada) Tel: +1-613-234-3244; E-mail: <mpereira@istar.ca>
PROJECT NUMBER	
STATUS	Expert contribution to Standards Development Work of ISO/IEC 18022 "IT-enablement for Widely Used Coded Domains". For information for SC32/WG1 and SC32/WG2 when discussing ISO/IEC 18022 development work at their meetings, Santa Fé, NM, USA 20-31 January, 2003
REFERENCES	
ACTION ID.	FYI
REQUESTED ACTION	
DUE DATE	
Number of Pages	26
LANGUAGE USED	English
DISTRIBUTION	P & L Members SC Chair WG Conveners and Secretaries

Douglas Mann, Secretariat, ISO/IEC JTC 1/SC 32

Pacific Northwest National Laboratory *, 13600 Angelica Court, Chantilly, VA, 20151-3360, United States of America

Telephone: +1 202-566-2126; Facsimile; +1 202-566-1639; E-mail: MannD@battelle.org

available from the JTC 1/SC 32 WebSite http://www.jtc1sc32.org/

*Pacific Northwest National Laboratory (PNL) administers the ISO/IEC JTC 1/SC 32 Secretariat on behalf of ANSI

ISO/IEC 18022 POSSIBLE MODEL APPROACH USING "CODES REPRESENTING DUBLIN CORE METADATA INITIATIVE" IN SUPPORT OF IT-ENABLED CODED DOMAINS WITH BILINGUAL (AND MULTILINGUAL) CAPABILITY

Section **Table of Contents** Page

1.0 INTRODUCTION AND PURPOSE OF CONTRIBUTION

[Note: SC 32/WG1 & SC32/WG2 members are invited to provide comments, questions, suggestions, etc. on this draft document. When revised/finalized, it is intended to serve as a document to be referenced in the bibliography to ISO/IEC 18022.]

The Dublin Core metadata element "coverage" is a free text (variable length field) which provides information on "the extent or scope of the content of the resource". Nevertheless, Dublin core is currently mandated for use by all Canadian government entitities for the description of federal government websites. This Case Study focuses on applying (under development) ISO/IEC 18022 methodology and approach to (1) ensuring a bilingual/multilingual capability in the underlying architectural design; (2) maximizing use of IT-enabled coded domains for implementing and supporting Canadian federal government "CLF" requirements for using Dublin Core.

Development of the ISO/IEC 18022 standard is taking place in parallel with a number of business case studies based on representative user requirements which are of a generic, horizontal nature. Some of these are in the public domain. Others are not. An example of a business case study in the public domain is a report prepared by the authors of this contribution for the USA's National Institute of Standards and Technology (NIST) and funded by the USA Environmental Protection Agency (EPA), i.e., *"Report on Multiple USA FIP Standards for "Codes Representing Administrative Subdivisions of the USA": Analysis and Recommendations"*, February, 2002.

This contribution is an integration of various projects conducted for different ministries of the Canadian federal government in the context of:

(1) the bilingual English/French language requirements of Canada's Official Languages Act as well as overall multilingual metadata needs in Canada (and elsewhere);

(2) various e-government initiatives; and,

(3) an overall Canadian federal government policy with respect to ensuring a "Common Look and Feel" (CLF) for Canadian citizens in accessing federal government websites and any "clients" utilizing e-government products and services. [Note: In this context, taxpayers are "clients"]

Within the Canadian federal government, the responsible entity is the Treasury Board Secretariat (TBS) [akin to the USA's Office of Management and Budget (OMB) as well as part of the role of NIST]. The TBS issues "Treasury Board Information Technology Standards (TBITS). The "Common Look and Feel" (CLF) policy is stated in "TBITS-37".[1]

[1] See further <http://www..cio-dpi.gc.ca/clf-upe/6/6_e.asp>

The federal goverment's Treasury Board Secretariat (TBS) "Common Look and Feel" (CLF) policy references TBITS-39 which in turn references and utilizes the Dublin Core Metadata Element Set (DCMES) to ensure the development and availability of standard metadata elements as descriptive information on each "document" posted to the WWW site of each Canadian federal government entity. At this point in time it is not clear whether and how the Canadian federal government's CLF policy applies to "data sets", "databases", digital mapping systems, "transaction processing" (and related e-government forms), etc. Use of a specified number of Dublin Core metadata elements is mandatory. Use of others is optional.

From the perspective of various government agencies, the use of Dublin Core metadata elements are inadequate and represent a "dumb sizing" of available information in various government document respositories. At the same time, the challenge and difficulties in achieving a government-wide consensus on common "descriptive elements", i.e., metadata tags, for any and all government "documents" for over hunderds of government ministries, agencies, "boards", "commissions", "authorities", and similar entities, cannot (and must not) be underestimated.

2.0 DUBLIN CORE METADATA INITIATIVE - OVERVIEW

The Dublin Core Metadata Initiative (DCMI) owes its origin to the second international World Wide Web (now known as W3C) conference held in 1994 during meetings on scholarly publishing on the web and the delivery of web-based library services. In March 1995 a follow-up metadata workshop was held in Dublin, Ohio sponsored by the National Center for Supercomputing Applications (NCSA) and the Online Computer Library Center Inc. (OCLC). The focus was on developing a core set of descriptive semantics, i.e., in the form of metadata elements, for web-based resources to facilitate access and discovery. (OCLC is a non profit membership organization serving over 41,000 libraries in 82 countries and territories around the world).

The DCMI is *"an open forum engaged in the development of interoperable online metadata standards that support a broad range of purposes and business models"*.

DCMI operates through consensus-driven working groups and annual "Workshops" which actually function as international plenary meetings/conferences. Nine such global workshops have been held so far in England, Australia, Finland, Canada, Japan, the United States, etc.

The initial approach to Dublin Core was the combination of (1) a "AACR-2" and "MARC-21" light (they are library standards for cataloguing); and, (2) use of the ISO/IEC 11179-3:1994 standard for the basic attributes of each data element. From its library-base, Dublin Core expanded to include the audio/visual community, photography, etc., resulting in a more generalized approach, (e.g., the original "author" is now "creator"). There is also a linkage of Dublin Core to GILS (Government Information Locator Service) based on the ANSI Z39.50 standard.

Dublin Core and DCMI documents, i.e., the Reference Description (v.1.1), have already been translated into Arabic, Catalan, Chinese, Danish, Finnish, French, Japanese, Korean, Polish, Russian, and Swedish with those in additional languages, (e.g., German, Norwegian, Portuguese) in the works. However, the quality of the translation leaves much to be desired. {See further below}. In addition, the definitions in other languages appear not to define (or "capture) the same concept as in the "original" English version.

3.0 APPROACH TO ANALYSIS AND RESULTS

3.1 APPROACH

In order to provide an example of a possible model approach in support of bilingual (and multilingual) requirements in support of the CLF policy, we have used the Dublin Core Version 1.1 Reference Descriptions which are already available in twelve (12) different languages (with more in the process of being made available).

The basic architectural design aspect of the ISO/IEC 18022 based model approach is:

(1) to separate the "IT-Interface" information from its possible multiple human interface equivalent linguistic representations; and,
(2) to structure the already agreed upon and thus predefined list of "members" of a coded domain into matrix form using "codes," as identifiers, to identify both the IDs of (a) the "schema"; and, (b) the permitted instances of entities as members of the schema.

[Note: This is a simplified view reflecting the limited and focused scope of this contribution. It does not include the rule base governing the schema, change management, etc. These and other related metadata standardization issues are being addressed internationally as part of overall ISO/IEC 18022 development work].

The two key elements of any IT-Interface are:

(1) the ID of the schema.

Here for the Dublin Core Attributes we have used "DCA" as the Schema ID. And for the Dublin Core we have used "DCMES" as the Schema ID, i.e., based on Dublin Core Metadata Element Set. We note that the use of the international syntax standard ISO/IEC 15445 (a.k.a., as HTML) is a sub-set of ISO/IEC 8879 Standard Generalized Markup Language, the use of the Dublin Core Metadata Element Set is identified by the 2-alpha lower case code of "dc.".

(2) the IDs of the codes representing the "members" of the schema.

3.2 RESULTS

The development of the Dublin Core specifications into matrix form as a "coded domain" and associated analysis yielded many results. They are summarized as follows:

(1) The basis of Dublin Core is that of a "bibliographic" description of a "publication", a "document", etc., irrespective of its form or format.

The focus of Dublin Core is its "standard" fifteen (15) metadata elements as the "core" descriptive data elements for the contents of any "document" irrespective of document types, (e.g., book, article, audio, photo, film, video, images, etc., in short both traditional and new media or "multimedia".

(2) Both the Dublin Core Metadata Element Set (DCMES) and the Dublin Core Attributes (DCA) lend themselves very well to being cast as a bilingual (and multilingual) coded domain.

Sections 4.0 and 5.0 below provide examples using the model approach.

(3) This matrix-based approach, i.e., as a bilingual/multilingual coded domain, has identified significant variations in definitions for the same DCMES element among the different language versions. At the start of this work, we did not expect these data integrity and quality control issues to arise given the widespread user acceptance and use in many countries and application of the Dublin Core specification.

For example, we do not "trust" the French translation provided by Dublin Core. It is (at a minimum/first glance): (1) missing accents, (2) uses variant spellings of words (or different words), (e.g., clefs for clés), and, (3) uses an English word where a French is available, (e.g., Copyright = Droits d'auteur).

A much more professional, quality French language equivalent is required. More than one French language version exists. {See further Section 4.0 below}

For example, the Dutch version not only restricts the use of Dublin Core to Internet-based resources (which Dublin Core itself does not) and but also, many of the definitions in the Dutch of the DCMES elements vary significantly from that of the "original" description of the Dublin Core Elements. The Dutch version states/acknowledges this.
{See <http://www.kb.nl/coop/donor/rapporten/DCsimpleformat.html> [2002-01-13]}

(4) The Dublin Core Metadata Element Set (DCMES) has no hierarchical structure nor any associations, relationships, dependencies, etc., i.e., "cardinalities".

(5) All the DCMES elements are presented as "physical data elements", i.e., element = "value". DCMES contains no "logical data elements" although these are implicit, i.e., as "categories".

3.3 DUBLIN CORE AND SYNTAXES

The Dublin Core documentation provides several examples of the use of different syntaxes[2]. They include:

(1) "HTML"-based, i.e., ISO/IEC 15455:2000.

HTML is widely used on the web.

(2) RDF/XML-based

"RDF" is the "Resource Description Framework" specification of W3C. It allows multiple metadata schemes to be read by humans as well as parsed by machines. Here "XML" is the notation assigned to the W3C user convention for utilizing HTML.

(3) "Metadata Contained in a Resource".

Quoting the "Using Dublin Core" paper:

"Some implementation using Dublin Core have chosen to embed their metadata within the resource itself. This approach is taken most often with documents encoded using HTML, but is also sometimes possible with other kinds of documents. Simple tools have been developed to make provision of Dublin Core metadata within HTML encoded pages fairly easy. One such too, DC.dot, extracts metadata information from an HTML document, and formats it so that it can be edited, then cut and pasted back into the HTML header of the original document".

The Canadian federal government's "Common Look and Feel" (CLF) policy is supported by a common "Toolbox" including a Dublin Core HTML generator. {See further Section 3.4 below}.

(4) Generic Form, i.e, "Stand Alone Metadata"

Again, quoting and emphasizing the following statement form the "Using Dublin Core" paper:

"Stand-alone metadata can exist in any kind of database, and generally provides a link to the described resource. This approach is likely to be practical for many non-textual resources, and is increasingly used for text as well, primarily to support easier maintenance and sharing of metadata".

[2] {See the paper by Diane Hillman titled "Using Dublin Core" dated 2001-04-12, at <http://dublincore.org/documents/2001/04/12/usageguide/>}

We note this in **bold** for four reasons. First of all, many government resources are of a non-textual nature. Secondly, the future trends in "e-standardization" are in support of business processes, commitment exchange, business transactions, etc., i.e., not only search, discovery and access of "documents". Thirdly, one can always utilize a "stand-alone" generic data element approach and then utilize various syntaxes, but the reverse is not true. Fourthly and finally, the more "intelligence" about each "data element", i.e., as "intelligent data", the easier it is to ensure data integrity and quality, semantic unambiguity and interoperability. **Consequently, we recommend that a standards based systematic approach for bilingual (and multilingual) capability for metadata focus on the "stand-alone" data element approach.**

If one wanted to "categorize" the fifteen (15) "Core" Dublin Core Elements , the following "categories " are used in the Dublin Core Documentation. [Note: Use of "DCMES" as schema ID and "0100", "1400", etc., as codes identifying members of this coded domain are created by us].

Table 1: "Categorization of Dublin Core "Core" Metadata Elements"	
Dublin Core IT Interface ID	**Dublin Core Categories [LOGICALS] and Physical Data Elements**
DCMES:	**[CONTENTS]**
DCMES:0100	TITLE
DCMES:0300	SUBJECT
DCMES:0400	DESCRIPTION
DCMES:1100	SOURCE
DCMES:1200	LANGUAGE
DCMES:1300	RELATION
DCMES:1400	COVERAGE
DCMES:	**[INTELLECTUAL PROPERTY]**
DCMES:0200	CREATOR
DCMES:0500	PUBLISHER
DCMES:0200	CONTRIBUTOR
DCMES:1500	RIGHTS

Table 1: "Categorization of Dublin Core "Core" Metadata Elements"	
Dublin Core IT Interface ID	**Dublin Core Categories [LOGICALS] and Physical Data Elements**
DCMES:	**[INSTANTIATION]**
DCMES:0700	DATE
DCMES:0800	TYPE
DCMES:0900	FORMAT
DCMES:1000	IDENTIFIER

3.4 DUBLIN CORE "COMMON LOOK AND FEEL": MANDATORY METADATA ELEMENTS

The Dublin Core has been adopted as part of the federal government's Treasury Board Information Technology Standards (TBITS), i.e., as part of TBITS-39: Treasury Board Information Management Standard, Part 1: Government On-Line Metadata Standard.

This TBITS-39 in turn forms the basis of the federal government's "Common Look and Feel (CLF)" policy, i.e., in the context of federal government information itself being available "on-line", i.e., via the Internet, as an "information resource". Such federal government "information resources" may be (1) end-products in themselves, i.e., as publications, "documents":, etc, or, (2) a set of recorded information which may in turn serve as a description of actual databases, business processes, etc.

The CLF Navigation and Format Section contains a section on metadata tags and we quote (from <http://www..cio-dpi.gc.ca/clf-upe/6/6_e.asp> [2002-03-20].

> *"Standard 6.3*
>
> *All GoC Web sites must adopt the following five metatags as a metadata standard for description of Web resources: Title, Originator, Language of Resource, Date and Controlled Subject".*

and,

> *"These five metadata elements are part of the 15 element Dublin Core Metadata Element Set the leading international metadata standard for on-line resource discovery".*

and,

Rationale

Metadata is a key tool in describing and managing information assets. It is particularly important to have an effective identification system for information assets since many are invisible, hidden in web sites or databases, until a user initiates a search to find the assets relevant to a current need. Whether we in the public service or our clients in the Canadian public are the searchers, we need an effective way to use the labels on our information assets to find them when we need them. You can see that the chair you are sitting on is a chair without looking at the bar code label, but an electronic document is invisible until its label or its text is found by a search tool.

The benefits of using a systematic way of assigning and structuring metadata include:

- *Relevance: providing information that search engines can use to find relevant documents in large collections such as web sites or document databases where text search alone brings up many irrelevant documents or lists of documents too long for users to look at.*
- *Identity: providing descriptive information so that users can tell how old a document is, who wrote it, or how to get additional information. Most documents on government web sites now cannot tell the user whether they are 5 days old or 5 years old. Sometimes the user wants one, sometimes the other. Metadata helps a user know if the information is reliable and current.*
- *Inventory: a list of what information the Government holds so that the information can be managed, tracked, updated, analyzed and used efficiently.*
- *Consistency: The Dublin core, an international metadata standard provides the framework and many of the rules for use so that metadata can be applied consistently in large and diverse organizations such as the Government of Canada. This creates an environment in which users can search for and find information without needing to know which department produced it or to which program it relates.*
- *Interoperability: an international metadata standard such as Dublin Core provides a way for information resources in electronic form to communicate their existence and their nature to other electronic applications (e.g. via HTML or XML) or search tools and to permit migration of information between applications or search systems.*
- *Policy compliance: a critical component of meeting the Management of Government Information Holdings (MGIH) policy requirement to know and be able to find the information Government holds.*

3.5 OPPORTUNITY FOR AN IT-ENABLED DUBLIN CORE AS A CODED DOMAIN

Different French language versions currently exist for the Dublin Core Metadata Element Set (DCMES) Version 1.1. For example, going/navigating via the "Translations" page lands on the French language translation prepared by Anne-Marie Vercoustre {See further <http://www-rocq.inria.fr/~vercoust/METADATA/DC-fr.1.1.html> [2002-03-21]} Navigating via the "Users' Guide for Dublin Core, we find a French language version prepared by Diane Hillman and translated by Guy Teasdale from Université Laval. {See further <http://www.bibl.ulaval.ca/DublinCore/usagegujide-20000716fr.htm> [2002-03-21]>}.

It suffices to note that these two French language versions are not the same. It is also beyond the scope and objectives of this contribution to assess/evaluate differences in the use of the French language in France and in Canada, (or Quebec, or some other province) or other jurisdictions where French is an official language.

Similarly, it is beyond the scope and objectives of this contribution to assess/evaluate differences in the use of the English language (including spelling conventions, choice of terms/words, etc.), among uses of the English languages in Canada, the United Kingdom, USA, Australia, New Zealand, etc.

To ensure that there is no confusion in the use of the Dublin Core Metadata Element Set (DCMES) Version 1.1 from a Canada-wide perspective (including that of the federal government's Official Language Act requirements), it is recommended that:

(1) Canada prepare an Official "English/French (Canada)" translation of DCMES Version 1.1;

(2) progress the same as a National Standard of Canada;

(3) prepare a matrix-based "coded domain version" (including supporting a common "coded domain" or tables); and,

(4) make the same publicly available across Canada as well as ensuring that such an "English/French (Canada)" bilingual version be posted to the Dublin Core website.

In addition, from a multilingual perspective, we recommended that the "Canadian" English/French version of DCMES also include a representative example of a multilingual version language and its two writing systems, i.e., Inuktitut.

4.0 TEN ATTRIBUTES FOR EACH DUBLIN CORE ELEMENT

As stated in the Dublin Core Version 1.1 Reference Descriptions (English and French):

"Each Dublin Core element is defined using a set of ten attributes from the ISO/IEC 11179 standard for the description of data elements".

«Chaque élement du Dublin Core est définit par un ensemble de dix attributs provenant de la description standard ISO 11179 pour des éléments de donnée».

Integrating these ten attributes and their "definitions" into a single bilingual (E/F) table yields the following.

[Note: In Table 2, we have used "DCA" as a (temporary) stakeholder schema ID for "Dublin Core Attributes"]

Table 2: IT Interface and Human Interface Linguistic Equivalents (E/F) for Dublin Core Attributes

IT Interface			Human Interface Linguistic Equivalents				
			English		French		Spare
Schema ID	**Code ID**	**Common Y/N**	**Attribute Name**	**Definition**	**Nom d'attribut**	**Définition**	
(01)	(02)	(03)	(52)	(53)	(62)	(63)	(nn)
DCA	01	2 = N	Name	The label assigned to the data element	Nom	L'étiquette attachée à l'élément de donnné	
DCA	02	2 = N	Identifier	The unique identifier assigned to the data element	Identifiant	L'identifiant unique de l'élément de donnné	
DCA	03	1 = Y	Version	The version of the data element	Version	La version de l'élément	
DCA	04	1 = Y	Registration Authority	The entity authorised to register the data element	Autorité	L'autorité autorisée à enregistrer l'élément	
DCA	05	1 = Y	Language	The language in which the data element is specified	Langue	La langue dans lequel l'élément est défini	
DCA	06	2 = N	Definition	A statement that clearly represents the concept and essential nature of the data element	Définition	Une phrase qui définit clairement quelle est la principale nature de l'élément et à quel concept il correspond	
DCA	07	1 = Y	Obligation	Indicates if the data element is	Obligation	Indique si la présence d l'élément	

Table 2: IT Interface and Human Interface Linguistic Equivalents (E/F) for Dublin Core Attributes							
IT Interface			**Human Interface Linguistic Equivalents**				
			English		**French**		**Spare**
Schema ID	**Code ID**	**Common Y/N**	**Attribute Name**	**Definition**	**Nom d'attribut**	**Définition**	
(01)	(02)	(03)	(52)	(53)	(62)	(63)	(nn)
				required to always or sometimes be present (contain a value)		est obligatoire ou optionnelle. (C'est à dire s'il contient toujours une valeur ou non)	
DCA	08	1 = Y	Datatype	Indicates the type of data that can be represented in the value of the data element	Type	Indique le type de données qui peut être représenté dans la valeur d l'élément.	
DCA	09	1 = Y	Maximum Occurrence	Indicates any limit to the repeatability of the data element	Occurrence	Indique une limite éventuelle sur le nombre maximum de fois que l'éléments peut être répété.	
DCA	10	2 = N	Comment	A remark concerning the application of the data element	Commentaire	Une remarque sur l'utilisation de cet élément de donnée	

5.0 DUBLIN CORE METADATA ELEMENT SET

5.1 INTRODUCTION TO MODEL APPROACH USING DUBLIN CORE EXAMPLE

The model approach is that based on: (a) existing relevant Canadian and international standards; (b) that of advocating Chapter 3.2 "Open Networking Standards" of The Canadian Electronic Commerce Strategy, (c) that of separating (1) unique, unambiguous and linguistically neutral identifiers for "re-useable" objects/resources (as advocated in international standards) from (2) the possible multiple human interface equivalents, representations, etc., of the same, and (d) related requirements for both interoperable cross-sectorial, of semantic components in business transactions, i.e., in support of

"commitment exchange" {See ISO/IEC 15944-1:2002}, and of metadata in support of information sharing among humans (requiring little or no computation of the information with data contents being shared).

The model approach (as being developed in the new international standard ISO/IEC 18022 "IT-enablement for Widely Used Coded Domains" through ISO/IEC JTC1/SC32 "Metadata" has four basic components; namely:

(1) the IT Interface Component;
(2) the Change Management Component;
(3) the Application Syntax Component;
(4) the Human Interface Equivalent Component.

The model approach is directed at recorded information for which 80%+ of the possible contents, i.e., permissible values, can be predefined and structured as well as being rule-based. The model approach is equally applicable to "metadata elements", i.e., metadata as "data about data", and the "data itself".

The IT Interface Component focuses on the use of unique, unambiguous and linguistically neutral identifiers to convey the semantics, i.e., meaning of the contents being interchanged. The identifiers must:

- be parsable according to predefined and IT-enabled rules capturing the business operational views;

- represent the intelligence and business operational rules pertaining to each of the parsable parts, which in turn must be rule-based with the rules explicitly stated from a human understanding perspective as well as capable of being stated from an IT perspective using formal description techniques (FDTs). [Note: The most common FDT in use at present is "UML", i.e., the ISO/IEC DIS 19501-1 *"Information technology - Unified Modeling Language (UML) - Part 1: Specification"/«Technologies de l'information - Langage de modélisation unifié (UML) - Partie 1: Spécification»* standard].

The two basic sub-components of the IT interface identifer are: (1) the schema ID; and , (2) the Code ID within that schema. Together these two sub-components, when standards-based, ensure a "global unambiguous unique ID". Popular and widely-used standards globally of this nature include those for product identification, the ubiquitous use of product IDs and associated bar codes on almost every product bought world-wide, telephone numbers, credit/debit/smart, etc, cards, passports, etc. {See further for details on these and similar international standards, Annex D titled "Existing Standards for the Unambiguous Identification of Persons in Business Transactions (Organizations and Individuals) and Some Common Policy and Implementation Considerations" in ISO/IEC 15944-1:2002 *Information Technology - Business Agreement Semantic Descriptive Techniques - Part 1: Operational Aspects of Open-edi for Implementation*}.

The "Change Management Component" addresses addition and deletion of codes, code splits, temporal aspects, etc.

The "Application Syntax Component" deals with the fact that multiple different syntaxes exist (standards-based and non-standards based) which can be utilized for interchange purposes.

The "Human Interface Equivalent Component" focuses on the representation of the semantics of contents itself, i.e., from a human interface needs perspective. The content can be of a linguistic or non-linguistic nature, (e.g., an image).

Two examples (non-technical and preliminary) in the form of matrices are provided namely:

(1) one focusing on IT Interface plus various associated human interface linguistic equivalents, i.e., human interface semantic elements associated with each unique combination "Schema ID + Code ID" in a bilingual (E/F) context. {See 5.4 below}

(2) one presenting a combination of IT Interface, Application Syntax and Human Interface components doing so in a multilingual context. {See Section 5.5 below}

5.2 IT INTERFACE

5.2.1 Schema ID (= Table ID)

For the purposes of this project, we have assigned "DCMES" as the schema ID identifier for the "Dublin Core Metadata Element Set, Version 1.1: Reference Description".

[Note: A shorter ID could be the "DC". The "dc" lower case is used in the Dublin Core HTML tags, i.e., as HTML syntax use schema identifiers in conjunction with the English name "Title" in lower case. We have used "DCMES" as a syntax independent schema ID].

5.2.2 Code IDs

There are 15 core DCMES elements. We have assigned them Code IDs as follows "0100", "0200" to "1500" rather than "01" through "15". The rationale for using this approach rather than "01" through "15" is because many applications have more detailed and granular data elements for each of these DCMES elements, i.e., as "sub-elements". This approach facilitates such users/applications being able to consolidate their internal more granular data element requirements into consolidated DCMES data elements for export and sharing purposes. For example, for "0300" "Subject and Keywords", one often uses one data element for "subject(s)" based on a thesaurus (or a controlled vocabulary), and one for keywords. Or for Creator, one often distinguishes between "individuals" and "organizations", etc.

The assignment of DCMES code IDs is as follows:

Table 3: Draft Code IDs for Dublin Core Metadata Elements		
Code ID	**Name (English)**	**Name (French)**
0100	Title	titre
0200	Creator	créateur
0300	Subject and Keywords	sujet et mot-clés
0400	Description	description
0500	Publisher	éditeur
0600	Contributor	contributeur
0700	Date	date
0800	Resource Type	type de la ressource
0900	Format	format
1000	Resource Identifier	identifiant de la ressource
1100	Source	source
1200	Language	langue
1300	Relation	relation
1400	Coverage	couverture
1500	Rights Management	gestion des droits

A primary reason for this set of draft Code IDs pertains to the use of DCMES with non-HTML syntax-based environments.

5.3 HUMAN INTERFACE LINGUISTIC EQUIVALENTS AND BILINGUAL/MULTILINGUAL EXPANDABILITY

The approach taken here is that of a repeatable set of columns representing a specified human interface equivalent. Currently the need for four "standard columns" as repeatable sets has already been identified; namely:

(1) Identifier
(2) Name
(3) Definition
(4) Comment

[Note: Frequently, users of Dublin Core view the "Identifier" and "Name" as being the same. Where this is not the case the "Name" element is "longer" , i.e., provides additional semantic information, than the "Identifier" which in reality appears to have the function of a "short name"].

5.4 DUBLIN CORE: TOWARDS AN IT-ENABLED AND BILINGUAL (MULTILINGUAL) VERSION

In this section, we provide an illustrative example in matrix form of the start of the business operational view of an IT-enabled bilingual (E/F) version of some essential Dublin Core information. The key construct here is to separate the IT interface requirements from their human interface equivalents, i.e., in this case linguistic and bilingual (= English/French).

The "IT interface" provides a set of unique, unambiguous and linguistically neutral identifiers for each of the members (as permitted instances) within Dublin Core as an "object class". These IT Interface Identifers in turn are composed of the schema ID and the Code ID (within the schema), i.e., they are "composite identifiers".

It must be noted that the Dublin Core currently has no internationally recognized "Schema ID", although within a HTML context, the convention is to use "dc." (which is also used in the TBS Common Look and Feel policy).

The assignment of the "Code IDs" is arbitrary and simple. In a nutshell, each of the Dublin Core metadata elements, has been assigned a numeric ID based on their order of presentation in the English language in Dublin Core documentation.

We have also chosen the schema ID = "DCMES" for two reasons; namely:

(1) to be independent of any syntax, (e.g., "dc." is used as part of the HTML Identifier in CLF); and.

(2) to support the move towards a syntax independent identifier.

Table 4: Illustrative Example of Dublin Core Separating the IT Interface from its Human Interface "E/F" Linguistic Equivalents

IT Interface	Human Interface Linguistic Equivalents	

Spare	Schema / Table ID	Code ID	ISO English				ISO French				Spare
			Identifier	Name	Definition	Comments	Identifiant	Nom	Définition	Commentaire	
(00)	(01)	(02)	(51)	(52)	(53)	(54)	(61)	(62)	(63)	(64)	(nn)
	DCMES	0100	Title	Title	A name given to the resource	Typically, a Title will be a name by which the resource is formally known	title	titre	Le nom donné à la ressource.	Typiquement, un titre sera le nom par lequel la ressource est officiellement connue.	
	DCMES	0200	Creator	Creator	An entity primarily responsible for making the content of the resource	Examples of a Creator include a person, an organization, or a service. Typically, the name of a Creator should be used to indicate the entity	Creator	créateur	L'entité principalement responsable de la création du contenu de la ressource.	Exemples de Créateur incluent une personne, une organisation, ou un service. Typiquement, un nom du Créateur devrait être utilisé pour désigner cette entité.	
	DCMES	0300	Subject	Subject and Keywords	The topic of the content of the resource	Typically, a Subject will be expressed as keywords, key phrases or classification codes that describe a topic by the resource. Recommended best practice is to select a value from a controlled vocabulary or formal classification scheme.	subject	sujet et mots-clés	Le sujet du contenu de la ressource.	Typiquement, le sujet sera décrit par un ensemble de mots-clés ou de phrases ou un code de classification qui précisent le sujet de la ressource. L'utilisation de vocabularies contrôlés et de schémas formels de classification est encouragée.	
	DCMES	0400	Description	Description	An account of the content of the resource.	Description may include but is not limited to: an abstract, table of contents, reference to a graphical representation of content or a free-text account of the content.	description	description	Une description du contenu de la ressource.	Une Description peut contenir, mais ce n'est pas limitatif: un résumé, une table des matières, une référence à une représentation graphique du contenu, ou un texte libre sur le contenu.	
	DCMES	0500	Publisher	Publisher	An entity responsible for making the resource available.	Examples of a Publisher include a person, an organization, or a service. Typically, the name of a	publisher	éditeur	L'entité responsable de la diffusion de la ressource, dans sa forme actuelle, tels, un département	Exemples d'Éditeur incluent une personne, une organisation, ou un service. Typiquement, le nom d'une maison d'édition	

Table 4: Illustrative Example of Dublin Core Separating the IT Interface from its Human Interface "E/F" Linguistic Equivalents

IT Interface			Human Interface Linguistic Equivalents								
			ISO English				ISO French				
Spare	Schema / Table ID	Code ID	Identifier	Name	Definition	Comments	Identifiant	Nom	Définition	Commentaire	Spare
(00)	(01)	(02)	(51)	(52)	(53)	(54)	(61)	(62)	(63)	(64)	(nn)
						Publisher should be used to indicate the entity			universitaire, une entreprise.	devrait être utilisé ici.	
	DCMES	0600	Contributor	Contributor	An entity responsible for making contributions to the content of the resource	Examples of a Contributor include a person, an organization, or a service. Typically, the name of a Contributor should be used to indicate the entity.	contributor	contributeur	Une entité qui a contribué à la création du contenu de la ressource.	Exemples de Contributeur incluent une personne, une organisation, ou un service. Typiquement, le nom d'un contributeur devrait être utilisé ici pour désigner l'entité.	
	DCMES	0700	Date	Date	A date associated with an event in the life cycle of the resource.	Typically, Date will be associated with the creation or availability of the resource. Recommended best practice for encoding the date value is defined in a profile of ISO 8601 and follows the YYYY-MM-DD format.	date	date	Une date associée avec un événement dans le cycle de vie de la ressource.	Typiquement, une date sera associée à la création ou à la publication d'une ressource. It est fortement recommandé d'encoder la valeur de la date en utilisant le format défini par l'ISO 8601 sous la forme AAAA-MM-JJ.	
	DCMES	0800	Type	Resource Type	The nature or genre of the content of the resource	Type includes terms describing general categories, functions, genres, or aggregation levels for content. Recommended best practice is to select a value from a controlled vocabulary (for example, the working draft list of Dublin Core Types). To describe the physical or digital	type	type de la ressource	La nature ou le genre du contenu de la ressource.	Type inclut des termes décrivant des catégories, fonctions ou genres généraux pour le contenu, ou des niveaux d'agrégation. Il est recommandé de choisir la valeur de type dans une liste de vocabulaire contrôlé (par exemple, la liste provisoire de Types du Dublin Core). Pour	

Table 4: Illustrative Example of Dublin Core Separating the IT Interface from its Human Interface "E/F" Linguistic Equivalents

IT Interface			Human Interface Linguistic Equivalents								
			ISO English				ISO French				
Spare	Schema / Table ID	Code ID	Identifier	Name	Definition	Comments	Identifiant	Nom	Définition	Commentaire	Spare
(00)	(01)	(02)	(51)	(52)	(53)	(54)	(61)	(62)	(63)	(64)	(nn)
						manifestation of the resource, use the FORMAT element.				décrire la matérialisation physique ou digitale de la ressource, il faut utiliser l'élément Format.	
	DCMES	0900	Format	Format	The physical or digital manifestation of the resource.	Typically, Format may include the media-type or dimensions of the resource. Format may be used to determine the software, hardware or other equipment needed to display or operate the resource. Examples of dimensions include size and duration. Recommended best practice is to select a value from a controlled vocabulary (for example, the list of Internet Media Types defining computer media formats).	format	format	La matérialisation physique ou digitale de la ressource.	Typiquement, Format peut inclure le media ou les dimensions de la ressource. Format peut être utilisé pour préciser le logiciel, le matériel ou autre équipement nécessaire pour afficher ou faire fonctionner la ressource. Exemples de dimensions incluent la taille et la durée. It est recommandé de choisir la valeur du format dans une liste de vocabulaire contrôlé (par exemple, la liste des types de media définis sur Internet).	
	DCMES	1000	Identifier	Resource Identifier	An unambiguous reference to the resource within a given context.	Recommended best practice is to identify the resource by means of a string or number conforming to a formal identification system. Example formal identification systems include the Uniform Resource Identifier (URI)	identifier	identifiant de la ressource	Une référence non ambiguë à la ressource dans un contexte donnée.	Il est recommandé d'identifier la ressource par une chaîne de caractère ou un nombre conforme à un système formel d'identification. Exemples de systèmes formels d'identification incluent le "Uniform Resource Identifier"	

Table 4: Illustrative Example of Dublin Core Separating the IT Interface from its Human Interface "E/F" Linguistic Equivalents

IT Interface			Human Interface Linguistic Equivalents								
			ISO English				ISO French				
Spare	Schema / Table ID	Code ID	Identifier	Name	Definition	Comments	Identifiant	Nom	Définition	Commentaire	Spare
(00)	(01)	(02)	(51)	(52)	(53)	(54)	(61)	(62)	(63)	(64)	(nn)
						(including the Uniform Resource Locator (URL)), the Digital Object Identifier (DOI) and the International Standard Book Number (ISBN)				(URI) (qui include le "Uniform Resource Locator" (URL)), le "Digital Object Identifier" (DOI) et le "International Standard Book Number" (ISBN).	
	DCMES	1100	Source	Source	A Reference to a resource from which the present resource is derived.	The present resource may be derived from the Source resource in whole or in part. Recommended best practice is to reference the resource by means of a string or number conforming to a formal identification system.	source	source	Une référence à une ressource à partir de laquelle la ressource actuelle a été dérivée.	La ressource actuelle peut avoir été dérivée d'une autre ressource source, en totalité ou en partie. Il est recommandé de référencer cette source par une chaîne de caractère ou un nombre conforme à un système formel d'identification.	
	DCMES	1200	Language	Language	A language of the intellectual content of the resource.	Recommended best practice for the values of the Language element is defined by RFC 1766 which includes a two-letter Language Code (taken from the ISO 639 standard), followed optionally by a two-letter Country Code (taken from ISO 3166 standard). For example, "en" for English, "fr" for French, or "en-uk" for English used in the United Kingdom.	language	langue	La langue du contenu intellectuel de la ressource.	Il est recommandé d'utiliser comme valeur de l'élément Langue celles définies par la RFC 1766 qui comprend un code de langage à deux caractères (venant du standard ISO 639), éventuellement suivi d'un code à deux lettres pour le pays (venant du standard ISO 3166) ou en français. Par exemple, "en" pour l'anglais, "fr" pour le français, ou "en-uk" pour l'anglais utilisé au	

Table 4: Illustrative Example of Dublin Core Separating the IT Interface from its Human Interface "E/F" Linguistic Equivalents

IT Interface			Human Interface Linguistic Equivalents								
			ISO English				ISO French				
Spare	Schema / Table ID	Code ID	Identifier	Name	Definition	Comments	Identifiant	Nom	Définition	Commentaire	Spare
(00)	(01)	(02)	(51)	(52)	(53)	(54)	(61)	(62)	(63)	(64)	(nn)
										Royaume-Uni.	
	DCMES	1300	Relation	Relation	A reference to a related resource.	Recommended best practice is to reference the resource by means of a string or number conforming to a formal identification system.	relation	relation	Une référence à une autre ressource qui a un rapport avec cette ressource.	Il est recommandé de référencer cette ressource par une chaîne de caractères ou un numéro conforme à un système formel d'identification.	
	DCMES	1400	Coverage	Coverage	The extent or scope of the content of the resource.	Coverage will typically include spatial location (a place name or geographic coordinates), temporal period (a period label, date, or date range) or jurisdiction (such as a named administrative entity). Recommended best practice is to select a value from a controlled vocabulary (for example, the Thesaurus of Geographic Names (TGN) and that, where appropriate, named places or time periods be used in preference to numeric identifiers such as sets of coordinates or date ranges.	coverage	couverture	La portée ou la couverture spatio-temporelle de la ressource.	La couverture typiquement inclut une position géographique (le nom d'un lieu ou ses coordonnées), une période de temps (un nom de période, une date, ou un intervalle de temps) ou une juridiction (telle que le nom d'une entité administrative). Il est recommandé de choisir la valeur de Couverture dans un vocabulaire contrôlé (par exemple, un thésaurus de noms géographiques, comme (TGN) et, quand cela est approprié, des noms de lieux ou de périodes plutôt que des identifiants numériques tels que des coordonnées ou des intervalles de dates.	

Table 4: Illustrative Example of Dublin Core Separating the IT Interface from its Human Interface "E/F" Linguistic Equivalents

IT Interface			Human Interface Linguistic Equivalents								
			ISO English				ISO French				
Spare	Schema / Table ID	Code ID	Identifier	Name	Definition	Comments	Identifiant	Nom	Définition	Commentaire	Spare
(00)	(01)	(02)	(51)	(52)	(53)	(54)	(61)	(62)	(63)	(64)	(nn)
	DCMES	1500	Rights	Rights Management	Information about rights held in and over the resource.	Typically, a Rights element will contain a rights management statement for the resource, or reference a service provided such information. Rights information often encompass Intellectual Property Rights (IPR), Copyright, and various Property Rights. If the Rights element is absent, no assumptions can be made about the status of these and other rights with respect to the resource.	rights	gestion des droits	Information sur les droits sur et au sujet de la ressource.	Typiquement, un élément Droits contiendra un état du droit à gérer une ressource, ou la référence à un service fournissant cette information. Ces droits souvent couvrent les droits de propriété intellectuelle (IPR), Copyright, et divers droits de propriété. Si l'élément Droits est absent, aucune hypothèse ne peut être faite sur l'état de ces droits, ou de tout autre, par rapport à la ressource.	

5.5 INTEGRATION OF IT-INTERFACE, APPLICATION SYNTAX AND HUMAN INTERFACE LINGUISTIC EQUIVALENTS IN A MULTILINGUAL CONTEXT

The purpose of this section is:

(1) to present in an illustrative manner a multilingual approach towards an IT-enabled version of Dublin Core; and,

(2) to introduce the need for being able to utilize various syntaxes as "Application Interface Syntax Equivalents", i.e., the possiblity and requirements to interchange and share the same metadata (and/or data) utilizing different "Application Syntaxes".

Included here are the "HTML Syntax Identifiers" as found in both Dublin Core and the federal government's TBITS-39 "CLF Requirements".

Provision is made for the support of other application syntaxes, (e.g., ISO/IEC 8824 ASN.1, ISO 9735 EDIFACT, ISO/IEC 8879 "SGML", ISO/IEC 15445 "HTML", etc.). {See further above Section 3.2}

Table 5: Integration of IT-Interface, Application Syntax Interface and Human Interface Linguistic Equivalents (Multilingual)

IT Interface			**Application Interface Syntax Equivalents**		**Human Interface Linguistic Equivalents**							
			HTLM Syntax	**Other Syntax**	**English (eng)**	**French (fra)**	**Spanish (esp) (Based on DC ver. 1.0)**	**German (deu) (Based on DC ver. 1.0)**	**Dutch (nld)**	**Finnish (fin)**	**Italian (ita)**	
Spare	**Schema / Table ID**	**Code ID**	**Identifiers**	**Identifiers**	**Name**	**Nom**	**Etiqueta**	**Name**	**Label**	**Name**	**Nome**	
(00)	(01)	(02)	(nn)	(nn)	(52)	(62)	(72)	(82)	(92)	(102)	(112)	(nn)
	DCMES	0100	dc.title		Title	titre	Titulo	Titel	Titel	Nimeke	Titolo	
	DCMES	0200	dc.creator		Creator	créateur	Autor o Creador	Verfasser oder Urheber	Auteur of maker	Tekijä	Creatore	
	DCMES	0300	dc.subject		Subject and Keywords	sujet et mots-clés	Claves	Thema und Stichwörter	Onderwerp en trefwoorden	Aihe	Soggetto e Parole chiave	
	DCMES	0400	dc.description		Description	description	Descripción	Inhaltliche Beschreibung	Omschrijving	Kuvaus	Descrizione	

Table 5: Integration of IT-Interface, Application Syntax Interface and Human Interface Linguistic Equivalents (Multilingual)												
IT Interface			**Application Interface Syntax Equivalents**		**Human Interface Linguistic Equivalents**							
			HTLM Syntax	**Other Syntax**	**English (eng)**	**French (fra)**	**Spanish (esp) (Based on DC ver. 1.0)**	**German (deu) (Based on DC ver. 1.0)**	**Dutch (nld)**	**Finnish (fin)**	**Italian (ita)**	
Spare	**Schema / Table ID**	**Code ID**	**Identifiers**	**Identifiers**	**Name**	**Nom**	**Etiqueta**	**Name**	**Label**	**Name**	**Nome**	
(00)	(01)	(02)	(nn)	(nn)	(52)	(62)	(72)	(82)	(92)	(102)	(112)	(nn)
	DCMES	0500	dc.publisher		Publisher	éditeur	Editor	Verleger bzw. Herausgeber[3]	Uitgever	Julkaisija	Editore	
	DCMES	0600	dc.contributor		Contributor	contributeur	Otros Colaboradores	Weitere beteiligten Personen und Körperschaften	Andere medewerkers	Muu tekijä	Autore di contributo subordinato	
	DCMES	0700	dc.date		Date	date	Fecha	Datum	Datum	Päivämäärä	Data	
	DCMES	0800	dc.type		Resource Type	type de la ressource	Tipo del Recurso	Ressourcenart	Bestands type	Laji	Tipo di risorsa	
	DCMES	0900	dc.format		Format	format	Formato	Format	Format	Formaatti	Formato	
	DCMES	1000	dc.identifier		Resource Identifier	identifiant de la ressource	Indenticador del Recurso	Ressourcen-Identifikation	Bestandsidentificatie	Identifikaatiotunnus	Identificatore della risorsa	
	DCMES	1100	dc.source		Source	source	Fuente	Quelle	Bron	Lähde	Fonte	
	DCMES	1200	dc.language		Language	langue	Lengua	Sprache	Taal	Kieli	Lingua	
	DCMES	1300	dc.relation		Relation	relation	Relación	Beziehung zu anderen Ressourcen	Relatie	Suhde	Relazione	
	DCMES	1400	dc.coverage		Coverage	couverture	Cobertura	Räumliche und zeitliche Maßangaben	Dekking	Kate	Copertura	

[3] German HTML label = DC.PUBLISHER

Table 5: Integration of IT-Interface, Application Syntax Interface and Human Interface Linguistic Equivalents (Multilingual)												
IT Interface			**Application Interface Syntax Equivalents**		**Human Interface Linguistic Equivalents**							
			HTLM Syntax	**Other Syntax**	**English (eng)**	**French (fra)**	**Spanish (esp) (Based on DC ver. 1.0)**	**German (deu) (Based on DC ver. 1.0)**	**Dutch (nld)**	**Finnish (fin)**	**Italian (ita)**	
Spare	**Schema / Table ID**	**Code ID**	**Identifiers**	**Identifiers**	**Name**	**Nom**	**Etiqueta**	**Name**	**Label**	**Name**	**Nome**	
(00)	(01)	(02)	(nn)	(nn)	(52)	(62)	(72)	(82)	(92)	(102)	(112)	(nn)
	DCMES	1500	dc.rights		Rights Management	gestion des droits	Derechos	Rechtliche Bedingungen	Copyright	Tekijänoikeudet	Gestione dei diritti	