

Dublin Core Metadata Element Set について – 現在の状況と利用例

杉本重雄

図書館情報大学

〒305-8550 茨城県つくば市春日 1-2

sugimoto@ulis.ac.jp

概要

Dublin Core Metadata Element Set (Dublin Core) はインターネット上での情報資源の発見 (Resource Discovery) を目的として提案されたメタデータである。Dublin Core は多様な分野の様々な情報資源に対応するため基本的な 15 要素を定めており、現在ネットワーク上での情報資源を表すためのコアメタデータとして注目されている。Dublin Core は 1995 年春に開かれたメタデータに関するワークショップで提案され、現在まで 6 回のワークショップとメーリングリストでの議論を通じて開発が進められてきた。本稿では、はじめに Dublin Core の概要を述べ、次に Dublin Core の現在の状況を簡単に解説する。また、筆者自身が携わった Dublin Core を用いた例について示す。

キーワード :

Dublin Core Metadata Element Set (Dublin Core), Resource Description Framework (RDF), コアメタデータ, 情報資源の発見, インターネット, デジタル図書館

Dublin Core Metadata Element Set – Current Status and Case Studies

Shigeo Sugimoto

University of Library and Information Science

Tsukuba, Ibaraki 305-8550, Japan

sugimoto@ulis.ac.jp

abstract

Dublin Core Metadata Element Set (Dublin Core) is a metadata description scheme aimed to resource discovery in the Internet. Dublin Core defines 15 core elements to describe metadata of various information resources. It is gaining acceptance as a core metadata scheme for networked information resources. Dublin Core was first proposed at the Metadata Workshop held at Dublin, Ohio in March 1995, and has been developed based on discussions at the series of workshops and on the mailing list. This article firstly shows the overview of Dublin Core, secondly discusses the present status, and lastly shows a few case studies.

Keywords :

Dublin Core Metadata Element Set (Dublin Core), Resource Description Framework (RDF), Core Metadata, Resource Discovery, Internet, Digital Library

1. はじめに

メタデータは「データに関する（構造化された）データ」と定義される[1]。現在、インターネット上には非常に多数の情報資源があり、それら情報資源へのアクセス効率を高めるには情報資源に適したメタデータが必要である。図書館や博物館・美術館では所蔵資料の目録や索引といったメタデータが作られ、資料の管理やアクセスに利用されてきた。しかしながら、インターネット上の資源はあまりにも多様であるため従来の詳細に決められたメタデータの規則をそのまま適用することは困難である。Dublin Core Metadata Element Set（一般に Dublin Core と呼ばれる）はインターネット上の様々な情報資源の発見を目的として提案されたメタデータである[2]。

Dublin Core は一連のワークショップと公開のメーリングリストでの議論を基礎にして形作られてきた、いわば草の根メタデータである。Dublin Core を作り上げてきたコミュニティは図書館や博物館・美術館の人たちが中心となっているが、WWW 上でのメタデータという目的ということから WWW 分野を始め多彩な背景の人たちが Dublin Core の開発に貢献してきた。

Dublin Core はインターネット上の文書のメタデータや Subject Gateway の開発のプロジェクトで実際に利用されている。また、インターネット上で直接提供されている資源に限らずより広い範囲の資源を、インターネットを利用して見つけ出すためのメタデータとして利用されようとしている。

以下、本稿では Dublin Core の概要、1998 年 11 月にワシントンの議会図書館で開催された第 6 回ワークショップ等最近の話題、さらに筆者自身が携わっている Dublin Core に基づくメタデータの作成と利用の例について述べる。また、付録として Dublin Core の参照記述を添えた。

2. Dublin Core の概要

2.1 背景

Dublin Core は 1994 年の WWW に関する国際会議での立ち話がきっかけとしてはじまったそうである。その後、1995 年 3 月に NCSA と OCLC が主催し米国オハイオ州ダブリン (Dublin, Ohio) で開催された第 1 回の Metadata Workshop で 13 項目の要素からなるメタデータ (Core Metadata Element Set) が提案され、これが Dublin Core の名前の由来となっている。その後、下に示すように 6 回のワークショップが開かれ議論が進められてきた[3]。

- ・ 第 2 回：1996 年 4 月、イギリスのウォーリック (Warwick)。複数のメタデータ規則に基づく記述のための基本概念である Warwick Framework が提案された[4]。また、SGML や HTML による記述の提案もなされた。
- ・ 第 3 回：1996 年 9 月、ダブリン。イメージデータに関する議論と 15 項目のエレメントの提案。
- ・ 第 4 回：1997 年 3 月、オーストラリアのキャンベラ (Canberra)。サブエレメントおよび Qualifier の議論、ならびに HTML での記述方式に関する議論。
- ・ 第 5 回：1997 年 10 月、フィンランドのヘルシンキ (Helsinki)。15 項目の基本エレメントの合意 (DC Simple)。このワークショップの結果 Dublin Core の標準化への動きが進められた。
- ・ 第 6 回：1998 年 11 月、アメリカのワシントン (Washington DC)。DC Simple (DC1.0) の標準化をさらに進めること、Qualifier 付き Dublin Core (DCQ1.0) の議論をさらに進めること、さらにこれまでの経験と他のメタデータ規則との関連を考慮した次バージョン (DC2.0) に関する議論を始めることに関する合意。

なお、Dublin Core の標準化に関しては、現在インターネット上での標準化に関して IETF で

の RFC2413 が出されている。また、NISO(アメリカ)、CEN(ヨーロッパ)および ISO での標準化が進められることになっている。

2.2 基本 15 エレメント

下に 15 エレメントの概略を示す。より詳しい定義は付録を参照されたい。Dublin Core の記述形式については 2.4 節で述べる。Dublin Core の記述において下の 15 エレメントの各々について、省略が可能であり、かつ繰り返しが可能である。たとえば、複数の著者がいる場合、エレメントを著者の数だけ繰り返して記述することができる。また、「ラベル」はエレメントの識別のために用いられる言語にはよらない表現であり、コンピュータによるエレメントの識別にも利用される。HTML や XML による記述ではエレメントの名前の指示にはここで定義した名前を用いる必要がある。(ラベルはトークン(token)と呼ばれることもある。)

表 1 Dublin Core の 15 エレメント

要素名	ラベル	説明
(1)タイトル	Title	オブジェクトの名前
(2)著者あるいは作者	Creator	情報資源の内容に関して責任を持つ人または組織
(3)主題およびキーワード	Subject	情報資源に述べられたトピック
(4)内容記述	Description	アブストラクトやイメージデータの説明など内容に関する記述
(5)公開者（出版者）	Publisher	情報資源を現在の形態にしたもの（出版社、大学など）
(6)寄与者	Contributor	著者ではないが文書の内容の作成に関わった人または組織（編集者や翻訳者等）
(7)日付	Date	現在の形で利用できるようになった日付
(8)資源タイプ	Type	ホームページ、小説、詩、辞書といった情報資源の型
(9)形式	Format	Postscript ファイルや Windows 実行形式といった、情報資源のデータ形式
(10)資源識別子	Identifier	情報資源を一意に識別するための番号あるいは名前
(11)情報源（出处）	Source	情報資源の出所となった情報資源を一意に示す番号もしくは文字列
(12)言語	Language	情報資源の内容を記述している言語
(13)関係	Relation	他の情報資源との関連づけ
(14)対象範囲（空間的・時間的）	Coverage	地理的場所や時間的な内容に関する情報資源の特性
(15)権利管理	Rights	著作権記述などの権利に関する記述や利用条件に関する記述へのリンク(URL もしくは何らかの URI)

2.3 いくつかのキーポイント

Dublin Core の位置付け、特色を理解するためにいくつかのキーポイントを示す。

(1)インターネット情報資源のためのコアメタデータの必要性

Dublin Core はインターネット上で提供される様々な文書（ないしは文書様オブジェクト、Document Like Object (DLO)）のためのメタデータを記述し、DLO をネットワーク上で発見するために提案されたものである。Dublin Core は、いわば WWW 上の DLO 用の目録規則と見ることでもある。ところが、従来の目録とは下記のような点で異なっている。（注：インターネット上には「Document（文書）」とは言いづらい様々な資料が提供されている。そのため、最近ではあまり DLO ということばは用いられず、「Resource（ここでは情報資源と訳す）」が用いられている。次の段落以降では情報資源、もしくは単に資源と記す。）

- ・ DLO の場合、出版や図書館による所蔵という概念がはっきりしないので、記述対象が明確でない。
- ・ CIP (Cataloging in Publication)のように DLO の中にある程度の書誌情報を入れられることが望ましいが、それを書くのは著者や編者、すなわち目録に関する素人である。
- ・ WWW 上には非常に多様な DLO があるので、どんな種類の DLO にでも適用できる詳細に決められた規則を作ろうとすると、非常に複雑なものになってしまう。
- ・ インターネットという巨大な分散環境で利用するので、相互利用性(Interoperability)が強く要求される。

Dublin Core は、素人にでも書け、多様な分野の情報資源の記述に適用できるという観点から少数の基本的要素(Core Elements)のみによるメタデータ記述を目指してきた。Dublin Core のは単にシンプルなメタデータとして理解すべきではなく、様々な分野に共通な要素によるメタデータ記述、すなわち様々な分野に（できるだけ）共通の概念としてみとめられた属性の要素の集まりとして定義したコアメタデータとして提案されたものであることを忘れてはならない。

(2) Semantic Interoperability

Semantic Interoperability は Dublin Core における忘れてはならないキーワードである。インターネットという巨大な分散環境で利用することを前提にしているので、メタデータは異なる環境で作られ、かつ使われる。そのため、メタデータの相互利用性は非常に重要である。すなわち、メタデータのエレメントの意味（セマンティクス）定義が明確に与えられなければならない、かつ様々な分野で適用できなければならない。これまでの Dublin Core での議論がもたらしたエレメントの意味定義に費やされてきているのはそのためである。また、メタデータは人間が読んで理解できるものでなければならない、かつコンピュータによっても利用できるものでなければならない。（たとえば、索引付けのためにタイトルや主題といったエレメントの意味が利用できなければならない。）

(3) Dublin Core の記述形式（シンタックス）

記述形式(シンタックス)に関する定義を与えないと実際の記述を行うことができない。ところが、シンタックスを厳密に決めることは利用者環境の不均一さや利用者の好みなどのために容易ではない。また、シンタックスの定義がセマンティクスに影響することを完全に避けることは困難である。一方、WWW の進化の速さを考えるとシンタックスを固定することは得策とは考えにくい。そのため、Dublin Core では HTML での記述方法に関する推奨や WWW の HTML のグループへの要望等を行ってきたが、Dublin Core 自身として記述形式を定めることはほとんどしてこなかった。現在では、WWW 上でのメタデータ記述として Resource Description Framework(RDF)が与えられているので、基本的に将来の Dublin Core 記述は RDF に基づくものになっていく。

(4) Simple Dublin Core と Qualified Dublin Core

Dublin Core による記述には大きく分けて二つの考え方がある。ひとつは基本要素を更に細

かく分けることはしないというもので、もう一方は基本要素をさらに細かく分けて記述するというものである。後者は前者を含むという風にも理解できるが、コアメタデータとしてのより細かな要素（基本要素の部分要素）の定義は一般に難しくなる。前者の考え方によって定義されたものを Simple Dublin Core（あるいは DC Simple (DCS)）、後者によるものを Qualified Dublin Core（あるいは DCQ）と呼ばれる。現在までに標準化の作業に取り掛かられているものは Simple Dublin Core である。（注：Qualified DC ではエレメントを細分化・詳細化するための Qualifier を用いてメタデータを表す。より詳細化・細分化されたエレメントをサブエレメントと呼ぶ。また、サブエレメントはサブストラクチャと呼ばれることもある。）

たとえば、著者の名前、所属、住所、電子メールアドレスといった項目を書く場合、DC Simple にしたがうとこれらを一まとめにして書いてしまうか、あるいはデータの中に属性を埋め込んで区別するしかない。一方、Qualified DC に従うとこれらを適切なサブエレメントに分けて定義することができる。

DC Simple で定義した例（HTML の META タグを用いたもの）

```
<META name="DC.Creator" content="Sugimoto, Shigeo, University of Library and  
Information Science, Tsukuba, Ibaraki, Japan, sugimoto@ulis.ac.jp">
```

Qualified DC で定義した例

```
<META name="DC.Creator.Name" content="Sugimoto, Shigeo">
```

```
<META name="DC.Creator.Affiliation" content="University of Library and Information  
Science">
```

```
<META name="DC.Creator.Address" content="Tsukuba, Ibaraki, Japan">
```

```
<META name="DC.Creator.Email" content="sugimoto@ulis.ac.jp">
```

こうして見るとサブエレメントを用いて記述した方が記述の意味が明確であることは明らかである。しかしながら、様々な分野の情報資源が提供されるネットワーク上でのメタデータの流通を考えた場合、共通のサブエレメントの定義をどのようにするかが難しい問題であることも容易に理解できる。

(5) 1:1 原則 (One to One Principle)

実際にメタデータを記述する際、どういう単位を対象に記述するのが問題である。冊子体の資料の場合であれば、記述対象が比較的明確であるのに対し、デジタル資料の場合「ひとまとまり」の資料が何をさすのかが必ずしも明確でない。たとえば、ひとつの論文が章毎に分けられ複数のファイルで構成されている場合、論文にひとつのメタデータを与えるのか、それとも各章毎（すなわちファイル毎、いわば URL を与えられている実体毎）にメタデータを与えるのが問題である。これまで Dublin Core では原則として対象資源とメタデータは 1 対 1 で与えるとしている。筆者は、上の論文の例であれば各章毎に与えることになるかと理解している。また、ひとつの論文が何枚ものページイメージでできている場合、論文全体とページイメージ毎のメタデータがそれぞれ作られることになる。

デジタルコンテンツの場合、物理的な資料をデジタル化して作られることも多い。物理的な対象物をデジタル化して作った資料のメタデータを書く場合、対象資源として「物理的な対象物」と「デジタルイメージ」のどちらをとらえるのかといった問題がある。たとえば、彫刻を写真にとってデジタル化したものの場合、Creator は彫刻の作者なのか、それとも写真家なのか問題になる。Creator は「知的内容の想像に関わった人」である。単なるデジタル化のための撮影であれば Creator は彫刻家である。そうでなく写真としての創造性が認められる場合には写真家が Creator であろう。その場合には彫刻、写真それぞれのメタデータが必要になり、またそれらの間の関係を表す必要が生じる。「原作（古典）の現代語版の翻訳を演

じた劇のビデオをデジタル化したもの」といった例を考えると段々混乱してくるようである。

(6) 多言語 (Multilinguality)

インターネット上の情報資源のメタデータを表す上で多言語の問題は重要な問題である。Dublin Core のワーキンググループの一つに Multilinguality グループがあり、この問題を検討している[5]。Simple Dublin Core の場合、エレメントの意味は言語とは無関係に、世界共通のものとして与えられている。また、記述方法に関しても計算機が読むための形式には統一して与えられたラベルを用いてエレメントを識別すること、値の記述言語を指示するために lang 属性を利用することが了解されている。現在まで多言語グループでは Simple Dublin Core の参照記述(Reference Description)の翻訳を進めてきており、12 言語への翻訳が終わり、他に 6 言語への翻訳が進められている。各言語による参照記述には要素名の翻訳も含められている。翻訳された要素名は基本的に人間が読むためのものであり、ユーザインタフェースなどに用いられる。一方、計算機は統一したラベルを検索や索引付けに利用する。

言語間での Semantic Interoperability を失わないためには英語で作られる標準定義の翻訳による意味の食い違いが生じないようにすることが重要である。また、Dublin Core を世界中で利用するには言語（あるいは文化）固有の情報の記述を導入せざるを得ない。たとえば、日本語の場合、「よみ」は重要な情報である。利用が広がるに連れて言語固有の情報がメタデータの中に記述されることになるであろう。そうした場合でも semantic interoperability を損なわないようにすることが求められる。

2.4 記述に関して

先に述べたように Dublin Core の活動の中ではシンタックスの定義はそれほど熱心にはなされてきていない。WWW 上での XML の利用が広がり RDF に基づく記述が一般化するまでは、WWW 文書の中への埋め込みを目的とした記述は HTML でせざるを得ない。しかしながら、HTML の META タグを用いた記述にはグルーピングの問題（後述）がある。メタデータとしてもとの資源とは別に記述し蓄積する場合、筆者自身は SGML（もしくは XML）を用いて適切な DTD を決めて Dublin Core メタデータ記述をしておけばよいと考えている。これは Dublin Core の参照記述が定義する範囲での記述同士の間では変換を容易に行えるためである。

参考のため、HTML による記述例と独自の DTD に基づき SGML で記述した例を示す。

・ HTML による記述例

```
<meta name="DC.Title" content="図書館情報大学卒業論文目次 平成9年度" lang="ja">
<meta name="DC.Publisher" content="University of Library and Information Science">
<meta name="DC.Publisher" content="図書館情報大学" lang="ja">
<meta name="DC.Contributor" content="Sugimoto, Shigeo">
<meta name="DC.Contributor" content="杉本 重雄" lang="ja">
<meta name="DC.Date" content="1998-6-6">
```

・ SGML に基づく記述例（図書館雑誌 no.11 をデジタル化して作ったデータに対して与えたメタデータである。一部にサブエレメントを用いている。）

```
<title lang=ja>図書館発達の内的動力を論じて本会会員諸君に望む
  <transcription lang=ja>トショカン ノ ナイテキドウリョク ヲ ロンジテ
  ホンカイ カイイン ショクン ニ ノゾム</transcription>
</title>
<creator>
  <personalName lang=ja>和田万吉</personalName>
</creator>
```

```

<subject lang=ja>図書館員</subject>
<subject lang=ja>図書館</subject>
<subject lang=ja>図書館員養成</subject>
<subject lang=ja>日本図書館協会</subject>
<publisher>
  <corporateName lang=ja>日本図書館協会</corporateName>
</publisher>
<date><creationOIC scheme=ANSIX3.30>1911-04-07</creationOIC></date>
<type>Image</type>
<format>gif</format>
<identifier>tosyokanzasshi11/00000002.gif</identifier>
<source>図書館雑誌第十一号</source>
<language scheme=ISO639>ja</language>

```

HTML の META タグを用いた記述の場合、ひとつのタグには一つの値しか記述しないことを推奨している。(ただし、一つのタグの中に複数の値を繰り返して書くことも誤りではない。) META タグを用いた記述では構造が明確に表現できないので、サブエレメントを含むエレメントのグルーピングが陽にできないという問題がある。たとえば、下の例のように二人の著者がいる場合、著者の名前を日英両方で記述した場合、4 つタグでの記述となり、その間の対応関係が機械的には取りづらくなる。こうした曖昧性が生じる場合、何らかの慣習的規則を決め記述の統一性を保つようにすることが必要である。

・グルーピングの問題を含む例

例 1 日本語で書いたタグと英語で書いたタグとの対応が取れない

```

<META name="DC.Creator" content="杉本重雄" lang="ja">
<META name="DC.Creator" content="Sugimoto, Shigeo">
<META name="DC.Creator" content="田畑孝一" lang="ja">
<META name="DC.Creator" content="Tabata, Koichi">

```

例 2 同一人物の Name と Email サブエレメントを結び付けることができない

```

<META name="DC.Creator.Name" content="Smith, John">
<META name="DC.Creator.Email" content="jsmith@foo">
<META name="DC.Creator.Name" content="Suzuki, Taro">
<META name="DC.Creator.Email" content="tsuzuki@bar">

```

Dublin Core では基本的にフリーテキストによる記述を基本としている。しかしながら、まったく自由な形式の記述のみにしてしまうとデータの解析が困難になる。そこである程度の記述形式の基準を与える必要がある。Dublin Core の記述には各エレメントに関して、エレメントの値の他、エレメントの値を記述するための基準(scheme)と値の記述言語(lang)を与えることになっている。たとえば、キーワードを LCSH に基づいて与えれば scheme は LCSH である。また、もしエレメント値を日本語で書いてあれば記述言語を表す lang は ja である (注)。このほか、各エレメントには記述のための基本的な指針が与えられており、メタデータの品質管理が可能のように考えられている。たとえば日付の記述形式の場合 ISO で決められた yyyy-mm-dd (年月日) や yyyy (年) といった形式を用いることを推奨している。エレメントによっては、たとえば形式(フォーマット)エレメントのように、あらかじめ用意された値の中から値を選択することを推奨しているものもある。こうした記述の推奨方式については各エレメント

に関するワーキンググループ等で議論が進められており、現時点での推奨については User Guide グループのワーキング・ドラフトで知ることができる[6]。

(注：対象資源の記述言語を表す Language エlementとElementの記述言語を表す lang 属性とを混同しないように注意が必要である。たとえば、<META name="DC.Language" content="英語" lang="ja">と<META name="DC.Language" content="en">あるいは<META name="DC.Language" content="English">はいずれも対象資源が英語で記述されていることを意味する。ただし、HTML によるメタデータの記述の場合、Elementの記述言語は英語 (lang=en)がデフォルト値である。)

2.5 Dublin Core を利用したプロジェクト

Dublin Core のホームページ[2]から数多くのプロジェクトを見つけることができる。また、第 6 回 Dublin Core ワークショップでもいくつかのプロジェクトの報告があった。たとえば、いくつかの博物館が共同して進めている CIMI(Consortium for the Computer Interchange of Museum Information)の報告では Dublin Core に基づくメタデータの多くの部分を機械的な変換によって生成したそうである。オーストラリアの政府情報を扱う AGLS(Australian Government Information Locator Service)では 15 Elementのほかにサービス(service)や機能(function)に関するElementを付加しているなどの報告があった。このほか、Z39.50 に Dublin Core の 15 Elementの定義が加えられていることなどの報告があった。

図書館情報大学附属図書館では Dublin Core に基づいて図書館情報学や図書館に関するネットワーク情報資源のメタデータの作成を進めることになっている。新たなメタデータはコンピュータの助けを借りた人手による記述が中心になる。一方、既存のメタデータ(たとえば OPAC)との統一的な検索インタフェースを作る上で Dublin Core を基礎とすることになっている。

以上のように、Dublin Core に基づくメタデータは必ずしも人手で一から作るというものばかりではなく、機械的な変換によって作り出したデータの利用、検索のための統一的なインタフェースのための利用が考えられる。これからは個々のデータベースへアクセスして検索するというよりは、ネットワークやデジタル図書館のような大きな情報空間の中から資源を探すことが多くなると考えられるので、そこでのコアメタデータの重要性は疑えない。

3. Dublin Core の現在の状況

3.1 第 6 回 Dublin Core ワークショップから

去る 11 月 2 日から 3 日間に渡ってワシントンの議会図書館で第 6 回の Dublin Core ワークショップが開かれ、筆者も参加した。参加者は約 100 名、16 カ国(日本からは 5 名)であった。ワークショップに先立ち 11 月 1 日に Dublin Core のグループが作っている Technical Advisory Committee (TAC)と Policy Advisory Committee (PAC)のミーティング (TAC-PAC ミーティング)が開かれ、そこで Dublin Core の開発活動の今後についての議論がなされた。それに加えて知的財産権管理のためのメタデータ規則の開発を進めている INDECS グループにおける考え方などに関して Godfrey Rus 氏の講演があった。(同氏は文献[7]で知的財産権管理のためのメタデータの観点から Dublin Core に対する批判的な記事を書いている。なお、同氏はワークショップでも講演した。)

筆者が理解した範囲で今回のワークショップを振り返ってみたい。今回のワークショップでは現在までに定義されたサブElementを持たない Dublin Core (Simple Dublin Core や DC Simple と呼ばれる)に関しては今後基本的な定義変更をすることはせず、定義の明確化や利用

経験に基づく修正にとどめることにし、現在の DC Simple(RFC 2413 で定義されているもの)をバージョン 1.0 (DC1.0)とした。サブエレメントを持つ Dublin Core の検討を引き続き進め qualifier 付き Dublin Core バージョン 1.0 (DCQ1.0 と呼ばれる)の定義を行うこと、さらに David Bearman からのエレメントの見直しに関する提案(注)や Rights Management のためのメタデータを検討している INDECS グループなどとも協調しながらバージョン 2.0(DC2.0)の検討を始めることを了解した。

(注: David Bearman 氏は Creator, Publisher, Contributor といった人あるいは組織が情報資源の発信に関して果たした役割によって分けられているエレメントを実際に記述する際、その区別が必ずしも明快にはできないことなどの理由からこれらをひとまとめにして Agent とでも呼んではどうかという提案をワークショップ前にしていた。ワークショップでは Rust 氏の講演内容なども踏まえエレメントの性質による再分類について講演した。なお、Dublin Core と Rights Management の両方の観点からの論文が Bearman, Rust ほかにによって発表されている[8]。)

筆者は以下の点に注意しなければならないと思っている。

- DC2.0 の検討が始まることによって DC1.0 の存在価値がなくなるもので決してなく、DC1.0 は現在実際に存在する唯一の標準として今後も安定的に利用できること。
- DC2.0 によって情報資源の発見という Dublin Core の役割と権利管理のメタデータ等が RDF という統一的な環境の下で統合的に利用できるようになる可能性があること。また、DC2.0 には DC1.0 との間で Semantic Compatibility (別の表現をすると機械的な変換可能性を有すること)を持つことが要求されること。
- 現在までの実現例の中でも qualifier はすでに多く用いられているが、DCQ は qualifier の利用方法に関する標準を定義する上で必要であること。

Dublin Core の役割は情報資源の発見であり、知的財産権やアクセス制限は Dublin Core の直接的な応用分野ではない。15 エレメントの中に権利管理のためのエレメントがあるとは言うものの、実際の知的財産権管理やアクセス制限に使うというよりは、そうした情報へのリンクを持つという位置づけに近い。知的財産権を扱うメタデータのグループとの協調が進むことは、情報資源の発見と発見された資源の利用のためのメタデータが RDF という共通基盤の上に作り上げられる可能性を意味する。その意味で今回のワークショップは大きな転換点であったように思える。

以上のほか、Dublin Core の発展に伴って Dublin Core を維持管理するための組織に関する議論も行われた。これまではメーリングリストとワークショップを基本とする草の根組織によって進めてきたが、さらに標準化と利用を勧めていく上ではよりしっかりした体制が必要であるためどのような体制で今後進めていくかに関する提案がなされた。(具体的な方法については今後出されるであろうワークショップのレポートを待つことにしたい。)

3.2 Resource Description Framework

Dublin Core に基づくメタデータの記述方法としてもっとも重要なものは WWW コンソーシアム(W3C)で進められている Resource Description Framework (RDF)である。RDF は WWW 上でいろいろなメタデータ規則に基づいてメタデータを記述するための形式と意味を定義している。また、RDF は第 2 回ワークショップで提案された Warwick Framework の影響を受けている。したがって、将来の Dublin Core メタデータ記述は RDF に基づいて記述されることになる。RDF の詳細は W3C のメタデータのページ[9]から選られる。また RDF のモデルとシンタックスに関する Proposed Recommendation[10]や文献[11], [12], [13]等を参照していただきたい。

4. 利用例

4.1 図書館情報大学デジタル図書館システム

図書館情報大学附属図書館が 1999 年 2 月から開始したデジタル図書館システムの主要な機能は図書館および図書館情報学に関するサブジェクトゲートウェイである。この機能を構築するために現在 WWW 上に提供される当該分野の情報資源を収集し、それらのメタデータの蓄積を進めている。収集は機械的に行い、メタデータの作成は人手によっている。また、メタデータ作成と蓄積のためにデザインしたユーザインタフェースとデータベースを用意している。なお、本システムに関する詳しい説明は別の機会に譲りたい。

このシステムでは Simple Dublin Core の 15 エレメントを基礎にして下記の項目を加えている。

1. 文字コード
2. 情報資源の発信国
3. 各エレメントのデータの「よみ」

また、下に示す例のように蓄積するデータは SGML のタグ付きテキストである。

```
<Title>ULIS Library HomePage New</Title>
<Creator lang=ja>図書館情報大学附属図書館</Creator>
<Subject lang=ja>図書館情報大学</Subject>
<Subject lang=ja>大学図書館</Subject>
<Subject lang=ja>図書館情報大学附属図書館</Subject>
<Description lang=ja>図書館情報大学附属図書館のトップページ。以下の…</Description>
<Date>1998-11-06</Date>
<Type>text</Type>
<Identifier Scheme=URL>http://www.ulis.ac.jp/library/</Identifier>
<Language>ja</Language>
<Format>text/html</Format>
<Charcode>x-euc-jp</Charcode>
<Country Scheme=ISO3166>jp</Country>
```

4.2 Dublin Core を利用した横断検索機能を持つ雑誌記事検索閲覧システム

筆者の研究室では Dublin Core を利用して雑誌記事の検索閲覧を行うシステムを作成している。このシステムでは、複数の雑誌（学術雑誌および国際会議論文集）の記事単位に作成したメタデータを利用して雑誌をまたがった検索を行うことができる。雑誌記事のメタデータは Dublin Core に基づくものと NACSIS-ELS の書誌データを参考にして作成したもの(ELS 形式と呼ぶ)がある。Dublin Core に基づくものは第 1 回ワークショップで提案された 13 項目定義に基づくものと現在の 15 項目定義に基づくものがある。本システムでは Dublin Core を基本のメタデータとし、他のメタデータとの間での要素の対応関係を定義した表を準備して検索に利用している。下にメタデータ記述の例を示す。（なお、Dublin Core のデータの DTD は本研究で独自に決めたものである。また、必ずしも最適の形式であるとは言えない。）

15 エレメントの Dublin Core

2.4 節で示した例を参照。

13 エレメントの Dublin Core

```
<dublinCo>
<subject>Digital library</subject>
<subject>電子図書館</subject>
<subject>library</subject>
<subject>図書館</subject>
<subject>library collections</subject>
<subject>蔵書</subject>
<subject>information resources</subject>
<subject>情報資料</subject>
<subject>intellectual</subject>
<subject>知的</subject>
<subject>Incommensurate data</subject>
<subject>比較できないデータ</subject>
<subject>work(intellectual entity)</subject>
<subject>知的作業</subject>
<title>Intellectual Realities and the Digital Library</title>
<author>Francis Miksa</author>
<author>Philip Doty</author>
<objectTy scheme=AACR2>computer file</objectTy>
<form scheme=IMT>text/html</form>
<identifi scheme=URL>http://www.csdl.tamu.edu/DL94/paper/miksa.html</identifi>
<relation type=ContainedIn scheme=URL>http://www.csdl.tamu.edu/DL94/</relation>
<source scheme=ISSN></source>
<language>English</language>
<coverage></coverage>
</dublinCo>
```

ELS 形式のメタデータ (図書館界 vol.47, no.1 より)

```
<article journal = "TKK" vol = "047" num = "01" page = "1" page2 = "1" date = "1
9950501" txtl = "jpn">
<atl><nihongo.title>大震災から 75 日</nihongo.title></atl>
<author>
  <g.name><kanji>正美</kanji><furigana>マサミ</furigana>
  <english>Masami</english></g.name>
  <f.name><kanji>柴田</kanji><furigana>シバタ</furigana>
  <english>Shibata</english></f.name>
</author>
<body><sec>
  <fg><fgart file = "TKK04701/00000005.tif"></fgart></fg>
</sec></body>
</article>
```

この例では Dublin Core 同士でも Creator と Author というように名前の対応が必要となっている。Dublin Core (15 要素) の Title と Creator に対して ELS の atl と author を対応づけている。なお、データの変換は行わず検索時に適切なタグ名を表から得て検索している。ま

た、本システムでは同様な対応表を用いることで ULIS OPAC のデータとの横断検索も実験した。また、本システムではタグ内の属性値の検索は行っていないため、ELS データの日付(Date に対応)や巻・号・ページのデータ (Source に対応) は照合対象としていない。

5. おわりに

メタデータはインターネット上での情報資源の利用やデジタル図書館においてキーとなる技術の一つである。たとえば、NSF/NASA/ARPA が進めた Digital Library Initiative (Phase 1) の中でも中心的な話題の一つであった[14]。NSF と EU が共同で進めたデジタル図書館に関する研究戦略を議論する 5 つのワーキンググループの中にメタデータを議論するワーキンググループが含まれており、中長期に渡る研究項目に関する提案をしている[15][16]。

Dublin Core はネットワーク上での情報資源の発見を目的としたメタデータである。現在まで、いくつものプロジェクトで利用されてきており、これからも利用は広がっていくと考えられる。ネットワーク上での利用を目的としてメタデータには、Dublin Core 以外にも知的財産権や Rating, プライバシーに関するメタデータの開発も進められている。将来、インターネット上での情報資源の流通には様々なメタデータを相互に利用することが必要になると考えられる。

参考文献

- [1] Dempsey, L., Heery, R., Metadata : A Current View of Practice and Issues, Journal of Documentation, Vol.54, No.2, pp.145-172, 1998.3
- [2] Dublin Core のホームページ, <http://purl.org/dc/>
- [3] Workshop Series (一連の Dublin Core ワークショップ), <http://purl.org/dc/about/workshop.htm>
- [4] Lagoze, C., The Warwick Framework, A Container Architecture for Diverse Sets of Metadata, D-Lib Magazine, 1996.7/8, <http://www.dlib.org/dlib/july96/lagoze/07lagoze.html>
- [5] Baker, T., Language for Dublin Core, D-Lib Magazine, 1998.12, <http://www.dlib.org/dlib/december98/12baker.html>
- [6] Hilman, D. (ed), User Guide Working Group, User Guide Working Draft 1998-07-31, http://purl.org/dc/documents/working_drafts/wd-guide-current.htm
- [7] Rust, G., Metadata: The Right Approach, D-lib Magazine, July/August 1998, <http://www.dlib.org/dlib/july98/rust/07rust.html>
- [8] Bearman, D. et al., A Common Model to Support Interoperable Metadata - Progress report on reconciling metadata requirements from the Dublin Core and INDECS/DOI Communities, D-Lib Magazine, 1999.1, <http://www.dlib.org/dlib/january99/bearman/01bearman.html>
- [9] W3C メタデータグループのホームページ, <http://w3c.org/Metadata/>
- [10] Lassila, O., Swick, R.R. (eds), Resource Description Framework (RDF), Model and Syntax Specification, W3C Proposed Recommendation 05 January 1999, PR-rdf-syntax-19990105, <http://www.w3.org/TR/PR-rdf-syntax/>
- [11] 浦本直彦, 武田浩一, インターネットでの情報の記述と交換方式の最近の動向, 人工知能学会誌, Vol.13, No.4, pp.519-527, 1998.7
- [12] 門間敦仁, XML とメタデータ, 情報の科学と技術, Vol.49, No.1, pp.16-22, 1999.1
- [13] 杉本重雄, メタデータについて-Dublin Core を中心として-, 情報の科学と技術, Vol.49, No.1, pp.3-10, 1999.1

- [14] NSF/DARPA/NASA Digital Libraries Initiative Projects
http://www.cise.nsf.gov/iis/dli_home.html
- [15] Schäuble, P. and Smeaton, A. F. (eds), An International Research Agenda for Digital Libraries: Summary Report of the Series of Joint NSF-EU Working Groups on Future Directions for Digital Libraries Research, 1998.10,
<http://www.iei.pi.cnr.it/DELOS/NSF/Brussrep.htm>
- [16] Peters, C., DELOS Workshop on Emerging Technologies in the Digital Libraries Domain, ERCIM News No.36, 1999.1,
http://www.ercim.org/publication/Ercim_News/enw36/peters.html

付録：Dublin Core エレメントの記述

(この文書は Dublin Core の参考記述より翻訳したものであり、<http://www.DL.ulis.ac.jp/DC/>に置かれている。)

下記は Dublin Core Metadata Element Set の定義の基準となる参照記述(reference definition)である。定義済みの限定子(qualifier)を含め、最新の参照記述は Dublin Core のホームページ(<http://purl.org/dc/>)に置かれる。

下記のエレメントの定義の記述において、各エレメントは、メタデータ記述に用いるためのエレメントの構文的仕様定義をより簡明にするための1語で表された形式的ラベルに加えて、各エレメントの意味を理解しやすい形で表すための記述的な名前(descriptive name)を持っている。

HTML のように大文字と小文字を区別しない環境で利用する場合もあるが、XML (Extensible Markup Language, <http://www.w3.org/TR/PR-xml>)のような大文字と小文字を区別する環境で利用するためにメタデータが抽出あるいは変換して利用される場合もあるので、下記に示したエレメントラベルの慣用に従って常に大文字と小文字を使い分けることを強く推奨する。

各エレメントは省略可能であり、かつ繰り返し可能である。さらにメタデータエレメントはいかなる順序で現れてもかまわず、順序は意味を持たない。

グローバルな相互利用性を向上するため、エレメントの値の記述の際に一定の統制された語彙の中から値を選ぶことが推奨されているものが多い。また、この統制語彙とは別に、ある地域、あるいは分野の中での相互利用性を目的として作り上げられた統制語彙を用いることがあってもかまわない。

情報資源の中にエレメントが埋め込まれている場合であっても、あるいは埋め込まれていない場合でも、メタデータエレメントの意味は不変である。

メタデータエレメントはそれが表す情報のクラスないしは範囲によって次の3グループに分けることができる。(1)主として情報資源の内容に関するエレメント、(2)主として知的財産として情報資源を見た場合に関するエレメント、(3)主として情報資源の具現化に関するエレメント。

内容	知的財産	具現化
Title	Creator	Date
Subject	Publisher	Type
Description	Contributor	Format
Source	Rights	Identifier
Language		
Relation		
Coverage		

1. タイトル

ラベル：Title

当該情報資源に与えられた名前。一般には作者もしくは公開者によって与えられる。

2. 著者あるいは作者

ラベル：Creator

情報資源の知的内容の創造に主たる責任を持つ人あるいは組織。たとえば、著述された文書の場合の著者、視覚的資料の場合の画家や写真家、イラストレータ。

3. 主題およびキーワード

ラベル : **Subject**

情報資源のトピック。典型的には、情報資源の主題あるいは内容を説明するキーワードや句。統制語彙や正式な分類体系に基づいて記述することが推奨される。

4. 内容記述

ラベル : **Description**

情報資源の内容に関する説明記述。文書の場合の抄録、視覚的資料の場合の内容記述など。

5. 公開者（出版者）

ラベル : **Publisher**

たとえば、出版社、大学の学科、企業体など、情報資源を現在の形態で利用可能にしたことに責任を持つ実体。

6. 寄与者（他の関与者）

ラベル : **Contributor**

Creator エレメントには示されたものではない人あるいは組織で、当該情報資源を作り出すに当たって知的に重要な寄与をしたもの。Creator エレメントに示された人あるいは組織に次いで大きな寄与をしたもの。（たとえば、編集者、翻訳者、イラストレータ）

7. 日付

ラベル : **Date**

当該情報資源が作成された、あるいは有効になった日付。この日付は Coverage エレメントに書かれる日付と混同してはならない。Coverage エレメントに書かれるものは当該情報資源の知的内容に何らかの関係を持つ日付である。YYYY および YYYY-MM-DD の形式で書く ISO 8601 [W3C テクニカルノート <http://www.w3.org/TR/Note-datetime> の (ISO8601 に基づく) 日付と時刻] が定義する形式に基づいて記述することが強く推奨される。たとえば、この形式では 1994 年 11 月 5 日は 1994-11-05 と表される。

8. 資源タイプ

ラベル : **Type**

情報資源の種類。たとえば、ホームページ、小説、詩、ワーキングペーパー、テクニカルレポート、エッセー、辞書・事典など。相互利用性を保つために、一連のワークショップで現在作成を進めている用語のリストの中から選ぶようにしなければならない。

9. 形式（フォーマット）

ラベル : **Format**

情報資源のデータフォーマット。情報資源を表示したり動作させたりするのに必要なソフトウェアや場合によってはハードウェアを識別するために利用できる情報を記述する。相互利用性を保つために、一連のワークショップで現在作成を進めている用語のリストの中から選ぶことが強く推奨される。

10. 資源識別子

ラベル : **Identifier**

当該情報資源を一意に識別するための文字列もしくは番号。URL や(実現された際には)URN はネットワーク上の情報資源に関する識別子の例である。国際標準図書番号(ISBN)や他の標準化された名前のように全世界的に一意に定まる識別子もこのエレメントの値として適切なものである。

11. 情報源（出处）

ラベル : **Source**

当該情報資源を作り出す元になった別の情報資源に関する情報。一般に、エレメントには当該情報資源に関する情報のみを記述することが推奨されているが、本エレメントには当該情報資源を見つけ出すために有用である別の情報資源に関する日付、作者、形式、識別子あるいは他のメタデータを書くことができる。実際の経験からは本エレメントの代わりに別の情報資源と

の関係を Relation エレメントを用いて表すことが推奨される。たとえば、1996 年に映画化されたシェークスピア劇に関する記述の中で 1603 年という値を Source エレメントに書くことができるが、この場合当該情報資源の Relation エレメントの中では “IsBasedOn” 関係を用いて 1603 年という記述を含む情報資源を参照する方が望ましい。当該情報資源が元の形式である場合には情報源エレメントは適用できない。

12. 言語

ラベル : Language

情報資源の知的内容を記述するために用いられている言語。実際的に利用するには、このエレメントの記述は、たとえば en, de, es, ja, th や zh といった RFC1766 [言語識別のためのタグ, <http://ds.internic.net/rfc/rfc1766.txt>] に適合している要求される。

13. 関係

ラベル : Relation

別の情報資源の識別子および当該情報資源とその情報資源との間の関係。このエレメントには関連する情報資源間のリンクや指示すべき情報資源記述を書くことができる。たとえば、作品の版(IsVersionOf), 作品の翻訳(IsBasedOn), 本の章(IsPartOf), データセットからイメージへの機械的変換(IsFormatOf)がある。相互利用性を得るために情報資源間の関係を表す値については、一連のワークショップにおいて現在定義が進められている値のリストから選択して与えることが推奨される。

14. 対象範囲 (空間的・時間的)

ラベル : Coverage

当該情報資源の知的内容に関する空間的(地理的)あるいは時間的特性。空間的範囲は物理的な範囲(たとえば天球の一部)を表す。この場合、座標(たとえば、経度と緯度)、統制語リストの中から選ばれたあるいは完全な名前を表された地名を用いる。時間的範囲は当該情報資源が表している内容に関する時間的情報を表すものであり、情報資源の作成や公開に関する日付ではない。(後者は Date エレメントで記述すべきものである。)この場合、次の形式で表すこと。Date エレメントと同じ日付と時間(期間である場合が多い)に関する形式 [(ISO8601 に基づく)日付と時間の形式, W3C テクニカルノート, <http://www.w3.org/TR/NOTE-datetime>], 統制語リストから選んだ時間区間記述, あるいは時間区間の完全な記述。

15. 権利管理

ラベル : Rights

権利管理に関する声明文、権利管理に関する声明文へのリンクを表す識別子、あるいは当該情報資源の権利管理に関する情報を提供するサービスへのリンクを表す識別子。