

EML Best Practices for LTER Site

原文 http://im.lternet.edu/sites/im.lternet.edu/files/emlbestpractices-2.0-FINAL-20110801_0.pdf

翻訳 真板英一@国環研 (maita.eiichi@nies.go.jp)

2011 / 10 / 06 版

目次

I	イントロダクション	3
I.1	EML Best Practices Version 1 (2004) からの変更点	3
I.2	EML 管理	4
I.2.1	データセットの作成	5
I.2.2	「属性一値」データモデル	6
I.2.3	GIS から作られた EML	7
II	要素と属性に関する推奨事項 (詳細)	7
II.1	全般的な推奨事項	7
II.2	要素ごとの推奨事項	8
	ルート要素: <eml:eml>	8
	@packageId(XML 属性)	9
	id, system, scope (XML 属性グループ)	9
	access	10
	dataset	11
	alternateIdentifier	11
	title (dataset)	12
	人物と組織 (団体)	12
	creator	12
	metadataProvider	14
	associatedParty	14

contact	15
publisher	15
pubDate	15
abstract	16
keywordSet と keyword	16
intellectualRights	17
distribution	18
coverage	19
geographicCoverage	19
temporalCoverage	21
taxonomicCoverage	22
maintenance	24
methods	24
project	27
[entity] = dataTable, spatialRaster, spatialVector, storedProcedure, view, other- Entity	28
attributeList	32
constraint	40
additionalMetadata	42
III 外部アプリケーションとの互換性に対する推奨事項	43
III.1 Metacat	43
III.2 LTERMapS	44
III.3 EML における由来追跡	44
III.4 EML 適合検査	46
IV この文書と共に提供される EML のサンプルファイルの説明	47

V 他の資源	47
--------	----

VI この文書に貢献してくれた人々	48
-------------------	----

I イントロダクション

2003年、Ecological Metadata Language (EML) が、LTER ネットワークに提供されるメタデータの交換用フォーマットとして採用された。このように、EML は LTER サイトの情報管理システムの一部である。この文書は、EML を LTER ネットワークで使用するときにはこうすると最善だ、と考えられる EML 記述法について、現時点でのお勧めの意見が含まれている。また、技術的なことが得意でない人たちのために、EML の概要的な文書（規範となる文書）を増やしたいという意図が込められている。実作業における注意点が適切に記されている。この文書は LTER サイトが利用できる幾つかの虎の巻 ("Best Practice") のひとつであり、この文書に関連する文書・資料は第 V 節の追加資料にリストアップしておいた。この文書で推奨されていることは、下記の特定のゴールを達成するためのものである、

- a) EML をデータセットに対して実際に使用する場面での指針と明快な説明を提供する
- b) EML データセット用のソフトウェアが簡単に作成・再利用できるように、LTER の EML 文書の不均一性を最小化する
- c) データの統合を促進するために LTER の EML 文書の相互運用性を最大化する

この文書を引用する場合は下記のようにして欲しい、

Long Term Ecological Research Network 2011. EML Best Practices for LTER Sites.

I.1 EML Best Practices Version 1 (2004) からの変更点

EML Best Practice の推奨事項は長い年月を通して発展してきたものだ。これらは LTER の情報管理担当者たちとネットワーク事務局の代表者たちの数度の作業部会が積み重ねた成果なのだ。各グループは、その時点において LTER サイトが提供可能な能力に合わせてこの推奨事項を修正してきた。2011年までに、EML は数年にわたって専用のアプリケーションによって広く使用されてきたし、また EML の項目について多数の組み合わせが議論されて結論が出るのを見る機会があった。以下に、この文書と第一版 (2004) との大きな相違点を記す。

EML のバージョン: 最新の仕様は EML 2.1 であり、すべての例文・推奨事項はこれを参照している。しかし、EML の要素についての推奨事項は EML 2.0.1 にも適用可能である。

文書の編成と EML のメタデータレベル: この文書の第一版では、「EML の完全度」（これは 2004 年に NISAC によって明確にされた機能上の階層に調和している）の節において特定の要素を議

論した。第一版では、EML を徐々に利用することができるように、完全度に応じて段々とデータ資源の説明が包括的になっていくようにした。この文書（第二版）では、推奨事項の細部がよりすっきりと編成され（第 II 節）、また EML 文書の順番を入れ替えた。他の編成方針も考慮された（たとえば要素を人間可読”human-readable” と機械可読”machine-readable” に区別する）。しかし、現時点では、EML 中のすべての情報は人間によって利用される一方で、アプリケーションが使用する要素はほんの少ししかない（たとえばアクセス制御や地理的範囲、データ実体の構造についての情報など）。<abstract> や <methodStep> などのテキスト型の項目を意味論的な用途に用いることができるアプリケーションはほとんど無いが、これらは機械的に検索可能である。現時点で暫定的に単純なテキスト型になっている項目もあるが（たとえば <attributeName>, <costumUnit>）、標準化の努力がネットワークとして実行されているところであるし、近い将来には機械可読なものになるだろう。

要素の説明: 幾つかの重要な EML 要素はデータセットの中の複数の位置に置くことができる（例えば coverage や methods）。配置可能な位置を示すために、それぞれの要素に対して配置が許される XPath をリストにした。XPath は文書の階層構造を表現するもので、ファイルサーバのディレクトリパスによく似ている。個々の要素に対して、その要素が最初に出現する時に推奨事項を一緒にまとめて示した（たいていは dataset 階層である）。これは、それらのツリーがその階層にのみ属するということを含意するものではない。とは言え実際的には、最も粗い階層にツリーを配置することは依然として有益であり推奨されるのだが。

データの可用性: LTER ネットワークの可用性ポリシーは、データ生産物は一般的に利用可能であるべきだ、としている。それゆえに、データ実体が一般の人々に対して利用不可能になるように記述する方法については、ここには限定的な事項しか書かれていない。データへのアクセスについての情報は、EML の規範文書の中に書かれている。

データ型: 第 I.2 節に、ある種のデータ型、または LTER のデータセットの中に共通して見られる状況（とりわけ、空間メタデータを EML に変換することや、数珠つなぎ”string-of-pearls” もしくは属性-値”attribute - value” データモジュールについて）に対する推奨事項を追加した。

外部アプリケーション: 特定の外部アプリケーションで使用されている EML に対する推奨事項を新しい節（第 III 節）として追加した。これらアプリケーションとして Metacat、LTERMaPs プロジェクト、PASTA（とりわけ、EML 適合性検査ツールや、それを LTER データセットの妥当性検査に使用する方法、そして派生したデータセットの由来を調べる方法）が含まれる。

I.2 EML 管理

「データセット」と「データパッケージ」という語はある程度相互に交換可能である。「データパッケージ」という語はここでは、データとメタデータを一緒にした公表上の単位、という意味で使用している。「データセット」は EML 文書においては特別な意味を持っている、なぜならそれはデータオブジェクトの最上位のコンテナだからである（つまり、<dataset> は、他の最上位要素の <citation>, <software>, <protocol> と対置されるものである）。しかし、論理的な「データセット」は少々違っているかも知れない。たとえば、進行中のコレクションを管理する必要から、データセットを徐々に公表することを指示することもあるだろうし、「進行中の気候データセット」

の全体が、それぞれを個別の EML 「データセット」にできるような多数の提供物から成り立っている場合もあるだろう。以下のデータアクセスアプリケーションではデータセットの語がまた違う意味で用いられているかも知れない。

I.2.1 データセットの作成

データセットやデータパッケージの作成について、幾つかの手法が LTER 内で出現したが、そのすべてが妥当なものだった。概して、この文書ではどれかひとつのやり方を推奨するようなことはしない。以下に幾つか例を示す：

(1) はっきりと期間を区切って収集されたデータは、補助的なデータをひとつの EML ファイルの中に（すなわち <dataset> の中に）まとめて記述し、それと合わせて論理的なひとまとまりとして公開する。EML はひとつのメタデータ文書の中に複数のデータ実体の説明を収納することができる。例えば、データテーブルに、サンプリング地点を示した KML ファイルや shape ファイル¹を添付することができる。

(2) 大規模なサンプリング活動によって幾らか独立のパラメータが測定された場合、ひとつの EML データセットに複数のデータ実体を組み合わせることもできるし、または幾つのかデータセットに分割して、そのすべてに共通部分のメタデータを添付し、またそれぞれに個々のデータに固有のメタデータを添付することもできる。どちらのやり方も妥当である。そしてどちらのやり方が自分のデータと状況にとって最良なのかを決めることはデータセットのデザイナーに任されている。データの利用者のことを配慮したり、あるデータ実体群を一緒にするのが自然かどうかにか気をつけるのは賢いことだ。

(3) 時系列データに対しては（それが備え付け機器によるものかどうかにかかわらず）、測定方法や他の補助的な構成要素が一定のままの時、幾らかのデザイナーは既存の EML メタデータを更新することを選択する。他のデザイナーは個々の論理的なまとまりごとに（たとえば「夏季のサンプリング」）新規にデータセットを作成することを選ぶかも知れない。どちらの手法にも利点と欠点がある。前者はデータの厳密なバージョン管理ができないし、一方で後者はデータの利用者に別々に分離されたデータ実体をダウンロードして結合する手間を強要する。

(4) データセットのデザイナーは、どの情報は EML の中に直接書き込むのが良くて、どの情報は補助ファイルに分離するのが良いのか、を決定するための理にかなった手法をよく考えるべきである。たとえば、データテーブルの列で使用されている略号は、EML に直接書き込むこともできるし、あるいは「ルックアップテーブル」の中に格納し（このテーブル自体は別の実体として EML 中に記述しておく）、そして <constraint>（外部キー制約）要素を通じてリンクさせるか、もしくは enumeratedDomain/EntityCodeList/ ツリーを用いてその属性の中に直接記述することができる。²

他の例としては、サイトの位置が挙げられる。これは EML の中に <methods> <sampling>

¹訳註・GIS ソフト用のデータフォーマット

²訳註・ここで言う略号（原文 code）というのは、名義尺度や順位尺度で用いられる語句のことである。あるデータテーブルのある列で使用可能な略号が事前に決まっている場合—たとえば県名—、使用可能な略号のリストを定義する必要がある。その定義を EML に直接記述するのか、それとも略号定義専用の表を別途用意して略号を使用している列に関連付けるのか、という選択の問題を論じているのである。

<spatialSamplingUnits>として書き込むこともできるし、KML や shape ファイルとして供給して別の実体として記述することもできる。この2つの選択肢を選ぶことは主に EML に書き込む分量に依存している。たとえば、略号が500個もあったら別のデータ実体に分離する方がいいし、「略号とその定義」の組が10個だったら EML に直接記述するべきだ。

I.2.2 「属性-値」データモデル

属性と値、または数珠つなぎ(“string of pearls”)データモデルは、ある種の観測データ(より型にはまった行列型のモデルを用いると多数の空白が生じてしまうようなデータ)に対して広く用いられている。このデータモデルは、個々の点観測を、位置・時間・変数(属性)名・値・様々なフラグ(観測方法、データの品質など)を含んだ単一のレコードとして扱う。生態学者にとっては、このデータモデルは(1)生物多様性や、そうでなければ(2)行列形式では非常に幅の大きいテーブルが必要とされるかも知れなかったり、多くの空白セルが生じてしまったりする観測データ(たとえば、生物種が普段は観察されない場合や、センサーの利用法が一貫していない場合)に対して特に有益である。

「属性と値」データモデルの利点は柔軟性と効率性である。しかし、このフォーマットで作られたデータセットは、最新版の EML では簡単に記述できない。基本的に問題となる点は、値欄に記述される値が(たとえば数値の型、収集方法、単位、精度などの)同じ属性を共有する必要がないところである。そこで差し当たり、下記のやり方を最善として推奨する。

生物多様性データ: もしあるテーブル中のすべての値が同じ単位(たとえば、存在/不在、観察数、優占率)を共有しているなら、そのテーブルはこの「属性と値」フォーマットで表現して EML の中に正確に記述されるべきである。もしそうでなければ、そのデータは行列形式で表現されるべきだし、そのテーブルが不合理なまでに横長になってしまう場合には幾つかの行列に分けて表現するべきかも知れない。

センサーデータ: 「属性と値」形式は特にセンサーデータに対して効果的だし、このデータモデルの変種は大量のセンサーデータを扱っているコミュニティによって開発されてきた(たとえばCUAHSIのODM、Observation Data Model)。単独のサイトにとってはこのデータモデルは、センサーが頻繁に変更されたり再配置されたりする時や、あるいはセンサーの配置が何種類かある場合に有用である。EMLを正確に記述するためには、個々のデータテーブルには単一のセンサーに対する「属性と値」形式の値のみが納められることが推奨される。もし単一のメタデータ文書の中に複数のセンサーから得られたデータが含まれる場合は、それぞれのセンサーから得られた値は、ひとつのデータセットの中に入れられた別々の実体(データテーブル)として見えるべきである。(これらのデータモデルの大半は、単独の「データストリーム」という考え方を説明しており、この推奨手法を動的に達成するのに使うことができる)

I.2.3 GIS から作られた EML

空間データセットを文書化するための、確立された基準が複数存在する。最もよく知られている形式は Federal Geographic Data Committee(FGDC) の地理空間基準、National Biological Information Infrastructure(NBII) の生物学的プロファイル、International Standards Organization(ISO) の標準 19115、そして ArcGIS のメタデータ形式である。<http://www.fgdc.gov/metadata/geospatial-metadata-tools> で FGDC に準拠したメタデータを準備するためのツールが利用できる。そしてそのメタデータは、XSL 変換スタイルシートを用いて EML に変換することができる。LTER ネットワークは以下のスタイルシートを提供している。

esri2eml.xsl: ArcCatalog メタデータエディタ (ARCGIS 9.3.1 まで) から EML 文書に変換する。
<http://im.lternet.edu/project/Esri2Eml/docs>

bdp2eml.xsl: FGDC および NBII Biological Profile メタデータを EML に変換する。
<http://im.lternet.edu/project/bdp2eml>

FGDC,ISO, ArcGIS 準拠のメタデータを作成する時に以下のガイドラインに従えば、esri2em.xsl を使って妥当な EML を作成することができるだろう。

- a) 個人の名前: 姓, 名 (つまり、姓と名をカンマを使って区切る)、そして名前と連絡先の要素 (電話、ファクス、電子メール) を複数書く場合には繰り返し記入できる欄 (repeating fields) を用いる。
- b) 属性の定義: すべての属性値には適切な値の定義が必要である。もし値の定義が利用出来ない場合は、値の定義を "unknown", "none", "null" のように記録する。

II 要素と属性に関する推奨事項 (詳細)

II.1 全般的な推奨事項

以下に、EML データセットのメタデータを作成するための最善のやり方 (best practices) の概略を記す。

メタデータの配布: データセットのメタデータとして不正確な情報を含む EML 文書を公に流通させないこと (すなわち、メタデータの内容の可用性に問題が生じるのを回避するため、あるいは EML の妥当性の要求に合わせるため)。草案やデモンストレーションとして、またはテストを目的として作成された EML は、はっきりとそういう目的のものであると明確にされるべきであり、公開のメタデータ倉庫に提供されるべきではない。

テキスト要素: 可能な場合は常に EML テキスト書式タグ (たとえば <section>, <para>, <orderedlist> など) を使用するようにせよ。<literalLayout> は、テキスト型を使う事ができず、書式指定のために HTML を含めねばならない時にのみ使用するようにせよ。注意点として、

テキスト型要素は docbook から引用したため、EML で採用されているような "camelCase" 式の記法を使っていない。³

バージョン管理: メタデータとデータセットのバージョン管理はサイトレベルで扱うことを推奨する。たとえば、進行中の長期データセットに対しては、年に 1 度、データ実体とそのメタデータを（正確な最終日時とバージョンを記述した EML と一緒に）取りまとめることができる。LTER-NIS では {scope}.{identifier}.{revision} という書式のパッケージ ID が使用されている。scope は "knb-lter-fls" という書式になっている。データやメタデータに対する更新や改訂では、identifier は同じものが使われるべきだが、revision は前よりも大きな数字にするべきである。

EML の "id" 属性: 多くの EML 要素では "id" 属性を持つことが許されている。EML 2.1.0 においては、文書の中にあるすべての id は重複してはならない。id 属性を使用して EML の内容を参照したり再利用する時には、重複に気をつけるべきである。リレーショナルデータベースから動的に EML を生成する時は 潜在的な id 衝突の危険を避けるために id を付けずに内容を複製する方がいいかも知れない。id 属性については更に詳しい情報を後述する。

優先度が高い要素: データセットを時空間的に、分類学的に位置づけるのを助けるため、メタデータはできるだけ多くの情報を、3 種類の <coverage> 要素、<temporalCoverage> (何時か)、<taxonomicCoverage> (何か)、<geographicCoverage> (何処か) のところに提供すべきである。

メタデータは、<project>,<methods>,<protocols> 要素に詳しい説明を記述して、潜在的な利用者がこのデータパッケージと自分の研究・プロジェクトとの関連性を評価することができるようにするべきである。

変更履歴は <maintenance> 要素に記録されるべきである。

II.2 要素ごとの推奨事項

ルート要素: <eml:eml>

この要素はすべての EML 文書におけるルート要素である。XPath 記法では:
</eml:eml

すべての EML 文書は、XML 宣言の後、ひとつのルート要素 (<eml:eml>) を持たなければならない。オンライン資源を指し示すシェーマ宣言を入れておくことと有益である。ルート要素に以下のシェーマ宣言を入れることを推奨する。

```
xsi:schemaLocation="eml://ecoinformatics.org/eml-2.1.0
http://nis.lternet.edu/schemas/EML/eml-2.1.0/eml.xsd"
```

すべての EML データセットは、ルート要素 (<eml:eml>) の下に最大で 3 つの要素を持つ。
<access>

³訳註・<タグ名> のアルファベットをすべて小文字にするか大文字を混ぜるか、というスタイルの違いを論じている。

<dataset>
<additionalMetadata>

@packageId (XML 属性)

この属性はすべての EML 2.1 文書において必須である。XPath 上の位置は /eml:eml/@packageId

他の節で概要を記したように（上述のバージョン管理、また後述する @id と <alternateIdentifier> を見よ）、各サイトはローカルレベルで重複の無いように識別子とバージョン番号を管理するべきである（後述の @system についての議論を見よ）。LTER ネットワークは現在、EML リポジトリとして Metacat を使用している。ほとんどのサイトは @packageId を、ハーバスタ上や文書アップロード時に指定される Metacat の "docid" に一致させている。しかし、必ずしも一致させねばならないわけではない。Metacat とそのハーバスタプログラムは文書の管理と同期のために数字のデータセット ID とリビジョン番号に頼っている。このことは、数字以外の ID を使っていたり、データセットのバージョン管理をしていないサイトに対して回避策を講じることを余儀なくさせる。考えられる解決策には、Metacat に最適化した EML を特異的に生成する方法（EML を動的に生成する能力のあるサイト向け）や XSLT 変換する方法がある。

現在、@packageId 属性は LTER データカタログ上での検索においてサイトの EML 文書を識別するのに使われている。KNB の Metacat に提出された EML において、サイトの @packageId 属性は以下のように標準化されているし、また別の書式がサイトと LNO データカタログ管理者によって合意されている。

knb-lter-[site].[dataset number].[revision], たとえば knb-lter-fls.187.4

Metacat 内の EML 文書に関しては第 III 節に他の情報が示されている。

id, system, scope (XML 属性グループ)

この属性グループは以下の EML 要素において使用できる。

<access>
<dataset>
<creator>
<associatedParty>
<contact>
<metadataProvider>
<publisher>
<coverage>
<geographicCoverage>
<temporalCoverage>
<taxonomicCoverage>
<distribution>
<software>
<citation>
<protocol>
<project>
<dataTable>
<otherEntity>
<spatialRaster>
<spatialReference>
<spatialVector>
<storedProcedure>

<view>
<attribute>
<constraint>

これら 3 属性は、@id 属性が参照として使用されない限り、ひとつのグループであり、またたいてい必須ではない。各@id はひとつの EML 文書の中で重複してはいけない、すなわち <creator> は <dataTable> とは異なる ID を持たなければならない。もし同じ人がデータセット制作者と測定手順定義者あるいはプロジェクト開始者に挙げられているなら、同じ @id を繰り返すことはできない。だから@id の内容を変更するか、参照を使って繰り返しであることを示すか、どちらかにしなければならない。現在のこの制限は、(人物データベースのようなところから) ID と一緒に内容を取り出す時に問題を生じさせる。このことは EML 開発者によって熟慮されているところである。理論上はこの 3 属性は一緒に機能する。@scope 属性の値は”system”と”document”のどちらかである。scope が”system”に設定された時には、system 属性が ID 体系を定義し、@id 属性の内容はそのシステムから導かれるという形が望ましい。

現在のところ、合理的で一般的な解法は、<eml:eml> 要素に system を定義し、system をサイトとする (ただし他のレベルでは system 属性を設定しない) 方法か、あるいは <eml:eml> 以外の要素中で scope に”document”を設定するか、どちらかである。

例: 属性 packageId, id, system , scope

```
<?xml version="1.0" encoding="UTF-8"?>
<eml:eml xmlns:ds="eml://ecoinformatics.org/dataset-2.1.0"
  xmlns:xs="http://www.w3.org/2001/XMLSchema"
  xmlns:eml="eml://ecoinformatics.org/eml-2.1.0"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns:stmml="http://www.xml-cml.org/schema/stmml-1.1"
  xsi:schemaLocation="eml://ecoinformatics.org/eml-2.1.0
  http://nis.lternet.edu/schemas/EML/eml-2.1.0/eml.xsd"
  packageId="knb-lter-fls.21.3"
  system="FLS"
  scope="system">
```

4

access

access 要素が現れる XPath 上の位置は

```
/eml:eml/access
/eml:eml/[entityType]/physical/distribution/access
```

<access> はこのファイルとデータ実体に対する許可ルールの一覧を含むものである。<access> の値はデータが格納されているシステムが扱えるものでなければならない。Metacat のアクセス制御書式は、LDAP での個人に対する識別名 (“uid=FLS,o=LTER,dc=ecoinformatics,dc=org” のような) と同じである。⁵

EML 2.1.0 時点で、<access> ツリーが使用出来る場所は 2 カ所ある。ひとつは <eml:eml> ルート要素の最初の子要素 (つまり <dataset> と同じ階層) で、文書全体に対するアクセス制御になる。もうひとつは physical/distribution ツリーの中で、このリソース URL に対するアクセス制

⁴ 訳註・Morpho で EML を作成するとルート要素の system は自動的に”knb”になる。現在のところ Morpho で system を変更する簡便な機能はない。

⁵ 訳註・JaLTER データベースでは EML メタデータの記述に基づいたアクセス制御を行っていない (2011 年 10 月現在)。

御になる。KNB の Metacat に提供された文書の access 要素は下記の例に応じた形式であるべきである。ある種の要注意情報を例外として、メタデータは誰でもアクセス可能であるべきである。<access> 要素は必須ではなく、もし省略されると、おそらくデータセット登録者のみがアクセス可能である。

例: access

```
<access authSystem="knb" order="allowFirst" scope="document">
  <allow>
    <principal>uid=FLS,o=lter,dc=ecoinformatics,dc=org</principal>
    <permission>all</permission>
  </allow>
  <allow>
    <principal>public</principal>
    <permission>read</permission>
  </allow>
</access>
```

dataset

この要素が出現する XPath 上の位置は
/eml:eml/dataset

<dataset> の下で以下の要素が使用可能である。幾つかは必須ではないが、もし使用するなら、この順番がシェーマによって強制されている。全体的にはこの順番で推奨事項を述べるが、人物と組織・団体に関係した要素は、他の要素と用途の違いがはっきりしているのも例外的にまとめて述べる。EML ファイルの中で異なる階層に出現できる要素については、最初に出て来た時か、最も上位の階層で議論をする (I.1 節も見よ)。

```
<alternateIdentifier>
<shortName>
<title>
<creator>
<metadataProvider>
<associatedParty>
<pubDate>
<language>
<series>
<abstract>
<keywordSet>
<additionalInfo>
<intellectualRights>
<distribution>
<coverage>
<purpose >
<maintenance>
<contact>
<publisher>
<pubPlace>
<project>
```

それから、これらの要素の後に、データ実体のための要素が続く。その要素は以下から選んで指定する。

[dataTable | spatialRaster | spatialVector | storedProcedure | view | otherEntity]

alternateIdentifier

この要素が現れる XPath 上の位置は
/eml:eml/dataset/alternateIdentifier
/eml:eml/dataset/[entity]/alternateIdentifier

サイトのデータセット ID は EML の <alternateIdentifier> としてリストされるべきである。とりわけそれが <eml:eml> 要素の "packageId" 属性とは異なる場合は。また <alternateIdentifier> は、データセットが複数の LTER サイトに帰属していることを記載するために、各サイトの ID を別々の <alternateIdentifier> タグに書き込むという風にするべきである。entity 階層では、<alternateIdentifier> はデータテーブル（あるいは他の実体）それ自身の別名を含むべきである。（後述の、データ実体についての追加コメントを見よ）

title (dataset)

title 要素が現れる XPath 上の位置は
/eml:eml/dataset/title
/eml:eml/method/methodStep/protocol/title
/eml:eml/project/title

データセットの <title> は、集められたデータ、地理的な情報や調査サイト（「何」と「どこで」）について記述されるべきである。もしそのデータが将来的に更新されないのなら、時間枠（「いつ」）についても。

例: dataset, alternateIdentifier, shortName, title

```
<dataset id="FLS-1" system="FLS" scope = "system">  
  <alternateIdentifier>FLS-1</alternateIdentifier>  
  <shortName>Arthropods</shortName>  
  <title>Long-term Ground Arthropod Monitoring Dataset at Ficity,  
    USA from 1998 to 2003</title>
```

人物と組織（団体）

人物と組織はすべて "ResponsibleParty" 要素群をもちいて記述される。これが出現する XPath 上の位置は

/eml:eml/dataset/creator
/eml:eml/dataset/contact
/eml:eml/dataset/metadataProvider
/eml:eml/dataset/associatedParty
/eml:eml/dataset/publisher
/eml:eml/dataset/project/creator
/eml:eml/dataset/method/methodStep/protocol/creator

全般的な推奨事項: EML 文書の内部のどこでも、<individualName> が使用される時には、人名を頭の中でアルファベット順に並べてデータを構築するべきである⁶。多くのサイトは、すべてのデータ作成者について完全な連絡先情報を管理することは実際的ではないと気付いている。しかし幾つかの重要な連絡先情報については最新の状態に保つべきである（下記を見よ）。もし人名に接尾辞が含まれているなら、<surName> 要素の中に、姓に続けて書くべきである⁷。

地位が人物とは独立であるような恒久的な役職に対しては、完全な連絡先情報を入れることを推奨する。たとえば、情報管理者やサイトの連絡先に対しては、電話番号に注意を払い、また他の人に移転できるように電子メールの別名機能を使うこと。（後述の <contact> を見よ）

⁶ 訳註・記入漏れを防ぐためか？

⁷ 訳註・ "Jr." などのことを言っているのだろう。

creator

この要素が現れる XPath 上の位置は
/eml:eml/dataset/creator

creator はそのデータセットの著者であると看做される⁸。つまり、制作物に入力された知的情報に対する責任を持つ人物である。<surName> と <givenName> 要素はデータセットの文献情報を作るのに用いられる。そのためこれらの要素は、帰属先が理解出来るようにするために、十分に完全な情報であるべきである。長期にわたるデータセットに対しては、(<organizationName> を使って) LTER サイトの名前を入れたり、(<positionName> を使って) サイトの PI の役職名を入れるようにせよ。気をつけるべき点として、サイトごとに違った手法を用いることは、長期データセットをなるべく効率よく検索しようという時に混乱を引き起こすし、検索する人はしばしば無条件に PI の姓を使って検索する。それゆえより多くの creator を記入することは実戦上合理的である。その結果として長期データセットに対する帰属があいまいになってしまうとしても。

例: creator

```
<creator id="org-1" system="FLS" scope="system">
  <organizationName>Fictitious LTER Site</organizationName>
  <address>
    <deliveryPoint>Department for Ecology</deliveryPoint>
    <deliveryPoint>Fictitious State University</deliveryPoint>
    <deliveryPoint>PO Box 111111</deliveryPoint>
    <city>Ficity</city>
    <administrativeArea>FI</administrativeArea>
    <postalCode>11111-1111</postalCode>
  </address>
  <phone phonetype="voice">(999) 999-9999</phone>
  <electronicMailAddress>fsu.contact@fi.univ.edu</electronicMailAddress>
  <onlineUrl>http://www.fsu.edu</onlineUrl>
</creator>
<creator id="pos-1" system="FLS" scope="system">
  <positionName>FLS Lead PI</positionName>
  <address>
    <deliveryPoint>Department for Ecology</deliveryPoint>
    <deliveryPoint>Fictitious State University</deliveryPoint>
    <deliveryPoint>PO Box 111111</deliveryPoint>
    <city>Ficity</city>
    <administrativeArea>FI</administrativeArea>
    <postalCode>11111-1111</postalCode>
  </address>
  <phone phonetype="voice">(999) 999-9999</phone>
  <electronicMailAddress>fsu.leadPI@fi.univ.edu</electronicMailAddress>
  <onlineUrl>http://www.fsu.edu</onlineUrl>
</creator>
<creator id="pers-1" system="FLS" scope="system">
  <individualName>
    <salutation>Dr.</salutation>
    <givenName>Joe</givenName>
    <givenName>T.</givenName>
    <surName>Ecologist Jr.</surName>
  </individualName>
  <organizationName>FSL LTER</organizationName>
  <address>
    <deliveryPoint>Department for Ecology</deliveryPoint>
    <deliveryPoint>Fictitious State University</deliveryPoint>
    <deliveryPoint>PO Box 111111</deliveryPoint>
    <city>Ficity</city>
    <administrativeArea>FI</administrativeArea>
    <postalCode>11111-1111</postalCode>
  </address>
  <phone phonetype="voice">(999) 999-9999</phone>
```

⁸ 訳註・Morpho の入力ウィザード画面では「データセット所有者 (dataset owner)」と表示される。EML と Morpho で用語が一致していないので注意。

```
<electronicMailAddress>jecologist@fi.univ.edu</electronicMailAddress>
<onlineUrl>http://www.fsu.edu/~jecologist</onlineUrl>
</creator>
```

metadataProvider

この要素が出現する位置 (XPath) は
/eml:eml/dataset/metadataProvider

<metadataProvider> 要素は、メタデータの内容を制作して提供することに対して責任のある人物・組織のリストである。LTER サイトによって生成された一次データセットに対して、通常は、<organizationName> 要素を用いて <metadataProvider> にその LTER サイトを挙げるべきである。後から獲得されたデータセット (<creator> や <associatedParty> がメタデータの中身を制作した人物とは一致しない場合) に対しては、実際にメタデータの内容を提供した人物がリストされるべきである (下記の例を見よ)。それぞれの <metadataProvider> 要素に対して <address>, <phone>, <electronicMailAddress>, <onlineURL> 要素を漏れなく記入せよ。

例: metadataProvider

```
<metadataProvider>
  <organizationName>Fictitious LTER Site</organizationName>
  <address>
    <deliveryPoint>Department of Ecology</deliveryPoint>
    <deliveryPoint>Fictitious State University</deliveryPoint>
    <deliveryPoint>PO Box 111111</deliveryPoint>
    <city>Ficity</city>
    <administrativeArea>FI</administrativeArea>
    <postalCode>11111-1111</postalCode>
  </address>
  <phone phonetype="voice">(999) 999-9999</phone>
  <electronicMailAddress>fsu@fi.univ.edu</electronicMailAddress>
  <onlineUrl>http://www.fsu.edu/</onlineUrl>
</metadataProvider>
```

associatedParty

この要素が出現する位置 (XPath) は
/eml:eml/dataset/associatedParty

なんらかの方法でそのデータに関わった他の人々 (現場の技官や学生アシスタントなど) を <associatedParty> としてリストせよ。すべての <associatedParty> ツリーは <role> 要素を必須とする。所属する大学・研究所・機関を適宜、"owner" の <role> を用いて <associatedParty> としてリストすることもできる。

例: associatedParty

```
<associatedParty id="12010" system="FLS" scope="system">
  <individualName>
    <givenName>Ima</givenName>
    <surName>Testuser</surName>
  </individualName>
  <organizationName>FSL LTER</organizationName>
  <address>
    <deliveryPoint>Department for Ecology</deliveryPoint>
    <deliveryPoint>Fictitious State University</deliveryPoint>
    <deliveryPoint>PO Box 111111</deliveryPoint>
    <city>Ficity</city>
    <administrativeArea>FI</administrativeArea>
    <postalCode>11111-1111</postalCode>
  </address>
</associatedParty>
```

```
</address>
<phone phonetype="voice">(999) 999-9999</phone>
<electronicMailAddress>itestuser@lternet.edu</electronicMailAddress>
<onlineUrl>http://search.lternet.edu/directory_view.php?personid=12010&
  query=itestuser</onlineUrl>
<role>Technician</role>
</associatedParty>
```

contact

この要素が出現する位置 (XPath) は
`/eml:eml/dataset/contact`

<contact> 要素はすべての EML データセットにおいて必須である。データ管理者の役職や他の指定の連絡先について十分な連絡先情報が含まれるべきであり、最新の情報に保つべきであり、人員の変更から独立であるべきである。もし幾つかの連絡先がある場合は（たとえばデータの管理者とサイトの管理者の両方）、すべてを最新に保つべきである。作業を実施した技官は <contact> よりも <associatedParty> に属する。<contact> 要素に対して <address>, <phone>, <electronicMailAddress>, <onlineURL> を漏れなく記入せよ。

例: contact

```
<contact id="pos-4">
  <positionName> Information Manager</positionName>
  <address>
    <deliveryPoint>Department for Ecology</deliveryPoint>
    <deliveryPoint>Fictitious State University</deliveryPoint>
    <deliveryPoint>PO Box 111111</deliveryPoint>
    <city>Ficity</city>
    <administrativeArea>FI</administrativeArea>
    <postalCode>11111-1111</postalCode>
  </address>
  <phone phonetype="voice">(999) 999-9999</phone>
  <electronicMailAddress>fsu.data@fi.univ.edu</electronicMailAddress>
  <onlineUrl>http://www.fsu.edu/</onlineUrl>
</contact>
```

publisher

この要素が出現する位置 (XPath) は
`/eml:eml/dataset/publisher`

LTER サイトはデータセットの <publisher> として記載されるべきである。LTER サイト名は、<organizationName> 要素の中に、省略せずにリストせよ。各 publisher 要素に対して、<address>, <phone>, <electronicMailAddress>, <onlineURL> を漏れなく記入せよ。

LTER-EML のウェブ表示向けの推奨事項: 組織情報に対して <publisher> を使え。

例: 上述の creator に対する参照を用いた publisher

```
<publisher>
  <references system="FLS">org-1</references>
</publisher>
```

pubDate

この要素が出現する位置 (XPath) は
`/eml:eml/dataset/pubDate`

データがオンラインで一般公開された年は、<pubDate> 要素として記載されるべきである。

<pubDate> は、データやメタデータが更新されたり再公開された時には更新されるべきである。書式は西暦 4 桁もしくは ISO の日付書式 (yyyy-mm-dd) のどちらかにすることができる。

abstract

この要素が出現する位置 (XPath) は
/eml:eml/dataset/abstract
/eml:eml/dataset/project/abstract

データセットにおいて、abstract 要素はリソース階層もしくはプロジェクト階層に現れることができる。<abstract> 要素は全文検索に用いられることになる。だからここには説明的な文章をたくさん書くべきである。説明するべきものは「何を」「いつ」「どこで」という情報、幾分かの分類学的情報、またデータセットが進行中か完了したものか、である。全般的な測定方法を説明することは適切であるし、測定項目も記載されるべきだ。測定項目が多数の場合は、すべての項目を記載する代わりにカテゴリを使え (たとえば、窒素、燐、カルシウム等々の代わりに「栄養塩」という語を使う)、ただしそれに併せて、検索にもっとも適当だと思われる項目はそのまま記述すること。

keywordSet と keyword

この要素が出現する位置 (XPath) は
/eml:eml/dataset/keywordSet
/eml:eml/dataset/project/keywordSet

キーワードを意味のあるまとまりに分けてそれぞれを <keywordSet> タグの中に格納することを推奨する。たとえば、ある <keywordSet> は LTER サイトを識別するためのキーワードに使い、またあるものは LTER の統制語彙から取られたキーワードに対して使い、またあるものは LTER の中核分野キーワードに対して使い、などなど。現在のところ各 <keywordSet> は特定のシソーラスの名前を (補助的タグ <keywordThesaurus> の中に) 記入することができる。記入されるキーワードは可能な限り LTER 統制語彙を用いるべきである、少なくともひとつの LTER 中核分野 (もし適切であれば)、3 文字のサイト頭文字、幾らかの意味のある地理的な地名 (州名、市名、地区名)、ネットワークの頭文字 (LTER, ILTER など)、所属組織、資金源 (すなわち他の資金源との共同基金、非 LTER 資金など) に関しては。具体的なキーワードに加えて、関連する概念的キーワードも記入されるべきである。

例: pubDate, abstract, keywordSet, keyword

```
<pubDate>2000</pubDate>
<abstract>
  <para>Ground arthropods communities are monitored in different habitats
    in a rapidly changing environment. The arthropods are collected in traps
    four times a year in ten locations and determined as far as possible
    to family, genus or species.</para>
</abstract>
<keywordSet>
  <keyword keywordType="place">City</keyword>
  <keyword keywordType="place">State</keyword>
  <keyword keywordType="place">Region</keyword>
  <keyword keywordType="place">County</keyword>
  <keyword keywordType="theme">FLS</keyword>
  <keyword keywordType="theme">Fictitious LTER Site</keyword>
  <keyword keywordType="theme">LTER</keyword>
  <keyword keywordType="theme">Arthropods</keyword>
  <keyword keywordType="theme">Richness</keyword>
  <keywordThesaurus>FLS site thesaurus</keywordThesaurus>
</keywordSet>
<keywordSet>
  <keyword keywordType="theme">ecology</keyword>
```



```

<keyword keywordType="theme">biodiversity</keyword>
<keyword keywordType="theme">population dynamics</keyword>
<keyword keywordType="theme">terrestrial</keyword>
<keyword keywordType="theme">arthropods</keyword>
<keyword keywordType="theme">pitfall trap</keyword>
<keyword keywordType="theme">monitoring</keyword>
<keyword keywordType="theme">abundance</keyword>
<keywordThesaurus>LTER controlled vocabulary</keywordThesaurus>
</keywordSet>
<keywordSet>
  <keyword keywordType="theme">populations</keyword>
  <keywordThesaurus>LTER core research areas</keywordThesaurus>
</keywordSet>

```

intellectualRights

この要素が出現する位置 (XPath) は
 /eml:eml/dataset/intellectualRights

<intellectualRights> は、サイトのデータ公開方針と、この特定のデータセット（制限付き公開データセット）に対して適用される個別の規定の説明が含まれるべきである。公表に向けての時間枠についてもよく記入されるべきである。たとえば、LTER ネットワーク規模のデータは2-3年以内にオンラインで公表されるべきであり、もしそうでない場合は、メタデータの中にその理由を記す必要がある。（LTER ネットワークのデータ公開方針も見よ。http://www.lternet.edu/data/netpolicy.html）

例: intellectualRights

```

<intellectualRights>
  <section>
    <title>Copyright Notice</title>
    <para> Copyright Board of Regents, Fictitious State University.
      This dataset is released to the public and may be used for academic,
      educational, or commercial purposes subject to the following
      restrictions:</para>
    <para>
      <itemizedlist>
        <listitem>
          <para>While FLS LTER will make every effort possible to control
            and document the quality of the data it publishes, the data
            are made available "as is".</para>
        </listitem>
        <listitem>
          <para>FLS LTER cannot assume responsibility for damages resulting
            from mis-use or mis-interpretation of datasets or from errors
            or omissions that may exist in the data.</para>
        </listitem>
        <listitem>
          <para>It is considered a matter of professional ethics to acknowledge
            the work of other scientists that has resulted in data used
            in subsequent research.</para>
        </listitem>
        <listitem>
          <para>FLS LTER expects that any use of data from this server will
            be accompanied with the appropriate citations and
            acknowledgments. </para>
        </listitem>
        <listitem>
          <para>FLS LTER encourages users to contact the original
            investigator responsible for the data that they are accessing.
            Where appropriate, researchers whose projects are integrally
            dependent on FLS LTER data are encouraged to consider
            collaboration and/or co-authorship with original
            investigators. </para>
        </listitem>
        <listitem>
          <para>FLS LTER requests that users submit to FLS LTER one copy
            of any publication resulting from the use of data obtained

```

```

    from this site. </para>
  </listitem>
</listitem>
  <para>FLS LTER requests that users not redistribute data obtained
    from this site. However, links or references to this site
    may be freely posted.</para>
</listitem>
</itemizedlist>
</para>
</section>
</intellectualRights>

```

distribution

この要素が出現する位置 (XPath) は
 /eml:eml/dataset/distribution
 /eml:eml/dataset/[entity]/physical/distribution

<distribution> 要素は dataset 階層と entity 階層に現れ、EML 文書内に書かれているデータがどのようにアクセス可能かという情報を含んでいる。<distribution> 要素は資源の位置を説明するために 3 つの子要素のうちのひとつを持つ。それらは <online>、<offline>、<inline> である。

Offline データ: アクセス制限しているデータやオンラインで利用できないデータを説明するには <offline> 要素を使い。<offline> 要素を用いる場合は <mediumName> タグは最小限記入すべきである。

Inline データ: <inline> 要素は EML 文書の中に直接格納されたデータを含む。文章や文字列として記入されたデータは XML として構文解析される。もしデータが構文解析されるべきでない場合は、それらを "<![CDATA[" と "]">" で包んで、"CDATA 節" として符号化せよ。

Online データ: <online> 要素は <url> と <onlineDescription> (任意) の 2 つの副要素を持つ。<url> タグは function という非必須の属性を持つことができ、これは "download" か "information" のどちらかに設定することができる。もしその URL が対象のダウンロードについての情報のみを提供し、直接にはデータの中身を返さない場合は、function 属性は "information" に設定されるべきである。もしその URL へのアクセスが直接にデータの内容を返すのであれば、function 属性は "download" に設定されるべきである。もし function 属性が省略された場合は、"download" であると暗黙に看做される。

EML データセットは少なくともひとつの URL を含むべきである。最低でも、<dataset> 階層 (XPath: /eml:eml/dataset/distribution/url) にあるべきであり、アプリケーションあるいはウェブサイトを指すことができる。この <url> の function 属性は "information" か "download" のどちらかに設定することができる。しかし、entity 階層の URL (たとえば dataTable、/eml:eml/dataset/dataTable/physical/distribution/url) は、要求してきたアプリケーションに対してデータの内容を返すべきであり、また function 属性には "download" が記入されるべきである。言い換えると、entity 階層では、URL は直接データへと導くべきであり、データカタログや使用目的ページへは導くべきでない。URL 接続についてより詳しい情報は、EML についての文書を見ること。

entity 階層で使用する時には、<url> に対して別のタグ <connection> を使う事が出来る。この要素は後述のデータ実体のところで議論する。

EML 2.1 時点で、entity 階層の <distribution> ツリーにおいて <access> という非推奨の要素もある (/eml:eml/dataset/[entity]/physical/distribution/access)。この要素はデータ実体そのものに対するアクセスの制御を特に意図している。<access> に関しては、上述の、アクセスに関する全体的な議論を見よ。

データアクセスを記録に残すことは、LTER のデータアクセスサーバ (DAS) とその URL プロキシシステムを使って実施できる。DAS とその使用法についてはオンラインの文書を読むこと。
http://im.lternet.edu/im_practices/metadata/das

例: distribution

```
<distribution>
  <online>
    <onlineDescription>fls-1 Data Web Page</onlineDescription>
    <url function="information">http://www.fsu.edu/lter/data/fls-1.htm</url>
  </online>
</distribution>
<dataTable>
  <physical>
    ...
    <distribution>
      <online>
        <onlineDescription>fls-1 Data Web Page</onlineDescription>
        <url function="download">http://www.fsu.edu/lter/data/fls-1.csv</url>
      </online>
    </distribution>
  </physical>
</dataTable>
```

coverage

この要素が出現する位置 (XPath) は

```
/eml:eml/dataset/coverage
/eml:eml/dataset/methods/sampling/studyExtent/coverage
/eml:eml/dataset/methods/sampling/spatialSamplingUnits/coverage
/eml:eml/dataset/[entity]/coverage
/eml:eml/dataset/[entity]/methods/sampling/studyExtent/coverage
/eml:eml/dataset/[entity]/methods/sampling/spatialSamplingUnits/coverage
/eml:eml/dataset/[entity]/attributeList/attribute/coverage
/eml:eml/dataset/[entity]/attributeList/attribute/methods/sampling/studyExtent/coverage
/eml:eml/dataset/[entity]/attributeList/attribute/methods/sampling/spatialSamplingUnits/coverage
/eml:eml/dataset/project/studyAreaDescription/coverage
```

<coverage> 要素は dataset, methods, entity, attribute 階層に出現することができ、データセットの範囲について空間、生物分類、時間の観点から説明する 3 要素 <geographicCoverage>, <taxonomicCoverage>, <temporalCoverage> を含む。これらの要素を推奨通りに入れておくことはより高度な利用を考慮してのことである。

geographicCoverage

全体的な情報: <geographicCoverage> 要素は、データセットに関連した調査サイトや地域の地理的な場所を説明するのに用いる。異なる種類の情報に対してこの要素を異なる階層で用いることを推奨する。明らかにこの要素の使い道にはいろいろあり、この文書では、LTER の内部でより一貫性のあるように、また LTERMapS のようなアプリケーションにおいて使いやすくするなるように EML を作成するひとつの手法を推奨する。

dataset 階層 (/eml:eml/dataset/coverage) においては 1 個の <geographicCoverage> 要素が記入されるべきであり、その中の <boudingCoordinates> がデータの範囲を説明する。初期状

態でそうなっているように、この矩形の領域は SiteDB に入っている LTER サイトの全体を記述することができる。しかし、もっと正確な観測範囲が提供されることが強く推奨される。つまり、データの最大範囲を、東西南北それぞれに対して。複数の研究サイトの間に着しい距離がある場合には、追加の `<geographicCoverage>` 要素を入れることができるし、それらをひとつの矩形領域に入れてグループ化すると混乱が生じるだろう。たとえば、サイト横断研究で、ひとつの矩形領域で北アメリカを表現する方法もあるが、より良くは、2 個の矩形領域を使って各サイトを示すべきである。

例: dataset 階層における geographicCoverage

```
<coverage>
  <geographicCoverage>
    <geographicDescription>Ficity, FI metropolitan area,
    USA</geographicDescription>
    <boundingCoordinates>
      <westBoundingCoordinate>-112.373614</westBoundingCoordinate>
      <eastBoundingCoordinate>-111.612936</eastBoundingCoordinate>
      <northBoundingCoordinate>33.708829</northBoundingCoordinate>
      <southBoundingCoordinate>33.298975</southBoundingCoordinate>
    <boundingAltitudes>
      <altitudeMinimum>300</altitudeMinimum>
      <altitudeMaximum>600</altitudeMaximum>
      <altitudeUnits>meter</altitudeUnits>
    </boundingAltitudes>
  </boundingCoordinates>
</geographicCoverage>
</coverage>
```

もしサンプリングが非連続の地点で行われているなら、それらの地点は `/eml:eml/dataset/methods/sampling/spatialSamplingUnits/coverage` に現れるべきである。個々のサンプリング地点のリストは `<spatialSamplingUnit>` の下に入れられ、各地点が別々の `coverage` 要素に入れられるべきである。サンプリング地点についてより詳しくは `<methods>` のところで見る事ができる。

例: 単独のサンプリング地点に対する geographicCoverage

```
<spatialSamplingUnits>
  <coverage>
    <geographicDescription>sitenumber 1</geographicDescription>
    <boundingCoordinates>
      <westBoundingCoordinate>-112.2</westBoundingCoordinate>
      <eastBoundingCoordinate>-112.2</eastBoundingCoordinate>
      <northBoundingCoordinate>33.5</northBoundingCoordinate>
      <southBoundingCoordinate>33.5</southBoundingCoordinate>
    </boundingCoordinates>
  </coverage>
  <coverage>
    <geographicDescription>sitenumber 2</geographicDescription>
    <boundingCoordinates>
      <westBoundingCoordinate>-111.7</westBoundingCoordinate>
      <eastBoundingCoordinate>-111.7</eastBoundingCoordinate>
      <northBoundingCoordinate>33.6</northBoundingCoordinate>
      <southBoundingCoordinate>33.6</southBoundingCoordinate>
    </boundingCoordinates>
  </coverage>
  <coverage>
    <geographicDescription>sitenumber 3</geographicDescription>
    <boundingCoordinates>
      <westBoundingCoordinate>-112.1</westBoundingCoordinate>
      <eastBoundingCoordinate>-112.1</eastBoundingCoordinate>
      <northBoundingCoordinate>33.7</northBoundingCoordinate>
      <southBoundingCoordinate>33.7</southBoundingCoordinate>
    </boundingCoordinates>
  </coverage>
</spatialSamplingUnits>
```

```
</coverage>
</spatialSamplingUnits>
```

緯度・経度は広く使われている同じ測地系に従うべきである（たとえばすべての値を WGS84 あるいは NAD83 で表記する）。また少なくとも小数点以下 6 桁まで表記するべきである（EML 2.1 のシェーマは小数点表記を強制している）。国際的な協定により、本初子午線の東の経度および赤道の北の緯度はプラス記号を頭に付け、西経と南緯にはマイナス記号を付ける。上記の例文を見よ。より多くの情報および他の例については EML の仕様書を見よ。

<geographicDescription> これは文字列で記述する。これは包括的であるべきであり、また国名、州名、地区名・地方名、市名、一般的な地形名、陸標、河川、そして他の関連する情報を含むべきである。<boudingCoordinates>, <boudingAltitudes>, 座標系、測地系などを決定する方法は <geographicDescription> に記入されるべきである。この情報は <geographicCoverage> の下では他のやり方で符号化されないからである。

<datasetGPolygon> 要素は、その研究の対象地が矩形の境界では十分に記述できない時に使用してよい。たとえば、研究対象地を記述するのに不規則なポリゴンが必要である場合や、矩形の境界の内側の一部領域を除外したい場合に。この要素は任意であり、また 2 つの副要素を持つ。

<datasetGPolygonOuterGRing>: これは、範囲をすっぽり包むようなポリゴン形の外周部分である。これは gRing（点のリスト）もしくは 4 個以上の gRingPoint によって生成することができる。FGDC G-Ring に対する文書では、ポリゴンを定義するのに 4 個の点が必要であり、また始点と終点は別であるべきだと明言している。しかしながら、XML シェーマにおいてはこれは強制されない。だから EML においてはポリゴン定義には最低 3 個の <gRingPoint> が必要とされる。そしてこれは始点と終点が結合して閉鎖しているポリゴンであると看做することができる。

<datasetGPolygonExclusionGRing> は、除外領域の、閉鎖していてかつ交差しない境界である。これは datasetGPolygon を用いて生成されたドーナツ形の中心の穴みたいなものだろう。これは gRing（点のリスト）または 1 個以上の gRingPoint によって生成することができる。これは外周ポリゴンから内部ポリゴンを除くべき場合に使用することができる。たとえば、湖を包むような地理的範囲から湖を除く場合に。

EML に位置情報を含めるための別の方法がある。特にそれを外部アプリケーションで使いたい時には。その方法とは GIS shape ファイル、Keyhole Markup Language (KML または KMZ)、または EML の spatial モジュールである。より詳しくは第三節「外部アプリケーションとの互換性についての推奨事項」を見よ。

temporalCoverage

dataset の <temporalCoverage> 要素はデータが収集された期間を表現する。回顧的なデータや歴史的なデータの場合、その研究がまとめられた年ではないので注意⁹。一般的には、<singleDate>¹⁰または <rangeOfDates> 要素が使用される。時には、「現在から何年前」といったような <alternatieTimeScale> が適切である。たとえば、長期間の年輪年代学にとっては（訳註・年月日という日付で <temporalCoverage> を表示するよりも）「何百年前」が適切である。日付の表

⁹訳註・たとえば年輪解析の場合、過去 100 年間分の年輪解析を 2010 年に実施したとすると、<temporalCoverage> は 1910 年から 2010 年までとなる。

¹⁰訳註・singleDate というタグは EML 規格には存在しない。singleDateTime の誤りであろう。

示には 2 種類の書式を使う事が出来る。4 桁の西暦年、または ISO 書式 YYYY-MM-DD である。

時には、データセットが「進行中」（データが定期的に追加される状態）である場合もあるだろう。現在のところ、EML において `<endDate>` タグを空欄のままにしておくことは規格違反である。この種のデータセットに対する一番簡単な解決策は、`<endDate>` 要素にはデータの最新年の終わり¹¹を記入しておき、メタデータを毎年更新することである。しかしながら理想的には、`<endDate>` タグは既にデータセット内に取り込まれているデータのみを反映するべきである。データをまだ利用することができない期間が含まれるような未来の終点を指定するよりも、その日付までの期間はデータが存在していることを保障する終点を、より多くのデータが利用可能になるだろうという情報と一緒に明言するほうが良い。`<maintenance>` タグ（後述）を、更新頻度の記述に使うようにせよ。データコレクションが進行中のものであることを説明するために `methods` や `sampling` ツリーを使うべきである。

例: temporalCoverage

```
<temporalCoverage>
  <rangeOfDates>
    <beginDate>
      <calendarDate>1998-11-12</calendarDate>
    </beginDate>
    <endDate>
      <calendarDate>2003-12-31</calendarDate>
    </endDate>
  </rangeOfDates>
</temporalCoverage>
```

taxonomic Coverage

`<taxonomicCoverage>` 要素は、その研究に関連するすべての生物に対する分類学的情報を文書化するのに使われるべきである。データを最低限の利用可能な水準にするために、種の二名式学名および一般名はなるべく常に記入されるべきである。しかしより幅広い分類学的な検索を支援するため、可能であるならより高レベルの分類群情報を記入するべきである。`<taxonomicClassification>` 要素のブロックは、同じレベルにおいて繰り返し使用するよりも、単一の `<taxonomicCoverage>` 要素の内部で階層的に入れ子にするべきである。`<generalTaxonomicCoverage>` 要素は a) 分類をどのように決定したのか、その全体的な手続きについての説明、b) その研究におけるすべてのフローラ、ファウナについての全体的な説明文 c) どのくらい細かく分類するのか（たとえば「科まで」あるいは「属と種まで」）の記述を含むことができる。

注意点として、分類学的「ツリー」を作るために、`<taxonRankName>` 欄などの下に、複数の要素を組み合わせて階層構造を作ることにはできる（例示はしない）が、このやり方は、複数の文書から `<taxonomicClassification>` 情報を取って来て組み合わせたり再利用することを妨げるだろうから、データセットの文書としては一般に推奨されない。

任意の `taxonomicCoverage/taxonomicSystem` ツリーは、分類学的同定資源の使用について詳しく述べたり¹²、また同様に同定作業の過程についての情報を提供するのに使えるだろう。

`<classificationSystem>` は、分類学的同定に使用した権威ある分類学データベース（ITIS, IPNI, NCBI, Index Fungorum, USDA Plants のような）や分類体系をリストするのに使用すべきである。使用した権威ある資料に関する説明文や関連文献（それらの出典を示す URL を含む）は

¹¹ 訳註・12月31日を意味しているのか？

¹² 訳註・同定作業にどのような人間を当てたのか、とか、そういう情報のことか？


```

<taxonRankValue>Animalia</taxonRankValue>
<taxonomicClassification>
  <taxonRankName>Phylum</taxonRankName>
  <taxonRankValue>Mollusca</taxonRankValue>
  <taxonomicClassification>
    <taxonRankName>Class</taxonRankName>
    <taxonRankValue>Bivalvia</taxonRankValue>
    <taxonomicClassification>
      <taxonRankName>Order</taxonRankName>
      <taxonRankValue>Filibranchia</taxonRankValue>
      <taxonomicClassification>
        <taxonRankName>Genus</taxonRankName>
        <taxonRankValue>Geukensia</taxonRankValue>
        <taxonomicClassification>
          <taxonRankName>Species</taxonRankName>
          <taxonRankValue>Geukensia demissa</taxonRankValue>
          <commonName>Ribbed Mussel</commonName>
        </taxonomicClassification>
      </taxonomicClassification>
    </taxonomicClassification>
  </taxonomicClassification>
</taxonomicClassification>
</taxonomicCoverage>

```

maintenance

この要素が現れる位置は (XPath では)
 /eml:eml/dataset/maintenance

dataset/maintenance/description 要素はデータテーブルやメタデータに対する変更について (更新頻度を含む) 文書化するのに使うべきである。この変更履歴は静的な文書における変更を記述するのにも使える。description 要素 (テキスト型) は、書式付きと書式無しのどちらのテキストブロックも含むことができる。

例: maintenance

```

<maintenance>
  <description>
    <para>Data are updated annually at the end of the calendar year.</para>
  </description>
</maintenance>

```

methods

この要素が現れる位置は (XPath では)
 /eml:eml/dataset/methods
 /eml:eml/dataset/[entity]/methods
 /eml:eml/dataset/[entity]/attributeList/attribute/methods

注意: 以前のバージョンの EML では、<method> と <methods> という要素が両方あったが、これは混乱を引き起こした。これらの要素は今では <methods> に標準化された。

一般的な情報: <methods> ツリーは dataset, entity, attribute 階層に出現する。そして内容は全体的には (機械可読でなく) 人間に読める形式であると看做される。大雑把に言えば、methods は 記述的であり、protocols は規範的である。つまり methods はデータを収集する時に何をしたかを記述するのであり、protocols は手続きや規定された行動の一セットである。方法はしばしば特定のプロトコルを含んだり、または特定のプロトコルに従ったりする。最低でも、dataset 階層において外部のプロトコルへの参照が与えられなければならない。しかしながら、この階層において詳細な方法が記述されることが望ましい。もしさらなる改善が必要であるなら、個々のデータ実

体や個々の <attribute> ¹³に対して必要に応じて方法を定義することができる。定義された方法の適用範囲はそれが適用された EML 文書階層に適合するべきである。たとえば、<dataset> 階層の方法では研究について記述するし、<dataTable> に対しては方法は前処理・後処理の手順を含むだろうし、また <attribute> 階層ではクオリティコントロールを含むだろう。方法についての改善を使用するかどうかはサイトによってさまざまであり、ユーザの中にはすべての方法をひとつの場所に入れることを好む人もいる。方法は主に人間が読むためのものであるので、dataset 階層にすべての手順についての詳しい説明をひとまとまりで置いておくことはとても重要であり、またユーザに大してより親切である。¹⁴

方法の記述には <methodStep>, <sampling>, <qualityControl> 要素が含まれる。

<methods> には最低1個の <methodStep> が必須である。またそれぞれの step は方法の論理的な一部分である。たとえば、対象地、実験室、統計。文章による方法の記述はすべて、<description> およびテキスト型のタグを用いてここに置かれる。

最低限、外部の文書を記述するために2種類のタグが用いられる。出版された文書や論文を参照するには <citation> を用い、それ以外には <protocol> を用いる。最低限、<protocol> は <title>, <creator>, <distribution> タグを要求する。ここで <distribution> ツリーはオンラインの文書を参照するのに使用出来る。このツリーについての上述の推奨事項を見よ。あるいは、プロトコル全体を EML の protocol/methodStep の中に書くこともできる。

<instrumentation> タグは使用した装置についての十分な記述を含むべきである。その情報には製造者、型番、キャリブレーションの期間や精度が含まれる。装置の交換や交換日時は <description> の中でなるべく早く言及されるべきである。

<dataSource> タグは EML データセットを入れ子にするためのものである。そのデータセットは、被説明データの <methodStep> への入力であるかも知れない。たとえば、装置のキャリブレーション情報や計算モデルの入力パラメータである。この要素はまた PASTA 由来の追跡システムによって、派生生産物が作成されて EML で記述される時にその元となるデータを記録するために使用されるだろう。より多くの情報は、第 III 節、外部アプリケーションとの互換性についての推奨事項を見よ。

任意の <sampling> ツリーは、研究サイトや範囲・頻度について、他の階層で挙げられた情報に追加して、重要でかつ詳細な情報を含むことができる。

<studyExtent> は、研究サイトの地理的・時間的・分類学的範囲に加えて、その研究の時間的・地理的範囲について具体的な情報を提供する（たとえば、どの領域に関心があるのか、のような）。これは <project> 内の <studyAreaDescription> の代理になり得る。<studyExtent> の記述は、<description> を用いて単純なテキストにすることもできるし、詳細な時間的・地理的 <coverage> 要素を複数記入することにより、サンプリングした期間を離散的に記述したり、<dataset> 階層において記述された地理的矩形境界全体のうち、サンプリングが実施された小範囲を記述したりもできる。下記の例では、coverage ツリーを用いて、LTER サイトのサンプリング地域全体の内部の、3カ所の昆虫トラップの位置が示されている。単一のサンプリングサイトの

¹³ 訳註・attribute とはデータテーブルの列のことである。

¹⁴ 訳註・Morpho の入力ウィザードの「方法」画面で入力した情報はすべて dataset 階層に配置される。

場所は <spatialSamplingUnits> の中にリストされるべきであることに注意すること。

<samplingDescription> は (サンプリング情報を) 文章で表現したものであり、学術論文における「サンプリング方法」節と同類である。

<spatialSamplingUnits> このツリーには個別のサイトもしくは点の位置が含まれるべきである。下記の例では、その市内のステーションのリストが、studyExtent ツリーの coverage 要素の構造の中に格納されている。LTERMapS プロジェクトはこのツリーの内容あるいは内部参照を利用することを計画している。この用途のためにメタデータを作成する時には第 III 節を見てほしい。

<methods> 内の他のツリーと同様、<qualityControl> は dataset, entity, attribute 階層において、適切であればいつでも使用可能である。最低限、<description> 要素にテキストで説明を加えること。また <citation> あるいは <protocol> タグも使用可能である。

例: methods

```
<methods>
  <methodStep>
    <description>
      <section>
        <title>Pitfall trap sampling for ground arthropod biodiversity
        monitoring</title>
        <para>Supplies used: pitfall traps (P-16 plastic Solo cups with lids)
        metal spades and large bulb planters (to dig holes in which to put
        traps) 70% ethanol (to preserve specimens) Qorpak glass jars with
        lids from the VWR Corporation, 120ml (4oz), cap size 58-400
        (comes included), Qorpak no. 7743C, VWR catalog no. 16195-703.</para>
        <para>Between 10 and 21 traps are placed at each site in suitable
        location. </para>
        <para>All trapped taxa counted and measured (body length), most taxa
        identified to Family, ants to Genus</para>
      </section>
    </description>
    <instrumentation>SBE MicroCAT 37-SM (S/N 1790); manufacturer: Sea-Bird
    Electronics (model: 37-SM MicroCAT); parameter: Conductivity
    (accuracy: 0.0003 S/m, readability: 0.00001 S/m, range: 0 to 7 S/m);
    last calibration: Feb 28, 2001</instrumentation>
    <instrumentation>SBE MicroCAT 37-SM (S/N 1790); manufacturer: Sea-Bird
    Electronics (model: 37-SM MicroCAT); parameter: Pressure (water)
    (accuracy: 0.2m, readability: 0.0004m, range: 0 to 20m); last
    calibration: Feb 28, 2001</instrumentation>
    <instrumentation>SBE MicroCAT 37-SM (S/N 1790); manufacturer: Sea-Bird
    Electronics (model: 37-SM MicroCAT); parameter: Temperature (water)
    (accuracy: 0.002 ° C, readability: 0.0001 ° C, range: -5 to 35 ° C); last
    calibration: Feb 28, 2001</instrumentation>
  </methodStep>
  <sampling>
    <studyExtent>
      <description>
        <para> Arthropod pit fall traps are placed in three different
        locations four times a year</para>
      </description>
    </studyExtent>
    <samplingDescription>
      <para>Six traps were set in a transect at each location.</para>
    </samplingDescription>
    <spatialSamplingUnits>
      <coverage>
        <geographicDescription>site number 1</geographicDescription>
        <boundingCoordinates>
          <westBoundingCoordinate>-112.234566</westBoundingCoordinate>
          <eastBoundingCoordinate>-112.234566</eastBoundingCoordinate>
          <northBoundingCoordinate>33.534566</northBoundingCoordinate>
          <southBoundingCoordinate>33.534566</southBoundingCoordinate>
        </boundingCoordinates>
      </coverage>
    </spatialSamplingUnits>
  </sampling>
</methods>
```

```

</coverage>
<coverage>
  <geographicDescription>site number 2</geographicDescription>
  <boundingCoordinates>
    <westBoundingCoordinate>-111.745677</westBoundingCoordinate>
    <eastBoundingCoordinate>-111.745677</eastBoundingCoordinate>
    <northBoundingCoordinate>33.64577</northBoundingCoordinate>
    <southBoundingCoordinate>33.64577</southBoundingCoordinate>
  </boundingCoordinates>
</coverage>
<coverage>
  <geographicDescription>site number 3</geographicDescription>
  <boundingCoordinates>
    <westBoundingCoordinate>-112.167899</westBoundingCoordinate>
    <eastBoundingCoordinate>-112.16799</eastBoundingCoordinate>
    <northBoundingCoordinate>33.76799</northBoundingCoordinate>
    <southBoundingCoordinate>33.76799</southBoundingCoordinate>
  </boundingCoordinates>
</coverage>
</spatialSamplingUnits>
</sampling>
<qualityControl>
  <description>
    <para>All specimens are archived for future reference. Quality
    control during data entry is achieved with standard database
    techniques of pull-downs that prevent typos and constraints.
    Scientists inspect standard data summary statistics after
    data entry.</para>
  </description>
</qualityControl>
</methods>

```

project

この要素が現れる位置は (XPath では)
 /eml:eml/dataset/project

元々は、<project> ツリーには、そのデータパッケージを支援しているプロジェクト（すなわちこの LTER サイト）についてとても全般的な説明を記入するように推奨していた。しかしながら、実際には、もしすべてのデータセットの EML が同じプロジェクトの記述を含んでいると、そのプロジェクトツリーの検索結果は非効率的で、不適切な結果でいっぱいになってしまう、という事態になってしまった。それに代わって、現在推奨されているのは、個々の LTER サイトで実施されている個別のプロジェクトについてのツリーを記入することである。そして、より小さい副プロジェクトについては <relatedProject> を用いて入れ子にする。

註: 研究プロジェクトの管理は LTERProject シェーマを用いて行うこともできる。このシェーマに対する補足の XSL スタイルシートには EML2.1 の <project> ツリーをエクスポートしたものが含まれている。より詳しくはオンライン文書を見ること。

<http://im.lternet.edu/project/LTERProjectDatabase>

最低でも、LTER サイトレベルのプロジェクトの説明には、<title>, <personnel>, <abstract>, 研究対象地域の説明と研究の目的とするところが含まれていなければならない。<distribution> ツリーはプロジェクトのホームページにリンクしているべきであり、そうでないなら代わりにそのプロジェクトを説明している他の出版物にリンクすることができる。これらの要素はすべて dataset 階層で使われる。具体的な推奨事項についてはそれら個々の要素の項目を見ること。

<project> 要素にしかない独特の要素が 2 つある。

<fundingSource> は機関と交付番号を含むべきである。

<studyAreaDiscription> ツリーと、それに随伴する <citation> ツリーは任意であり、研究対象地の (coverage 以外の) 特性を記述するのに使えるだろう (たとえば気候、地質、攪乱や、あるいは生物学的・地球物理学的な分類体系 (Bailey の生態地域区分や Holdridge の生物地帯など)。studyAreaDescription ツリーは複数の <coverage> 要素にも対応しており、それを使って大きな領域内にある個別の研究対象地の地理的範囲を記述することができる。それらは studyExtent/spatialSamplingUnits/referencedEntityId によって参照されることも可能である。同じ階層の <descriptor> タグはそのサイトについての説明文に対して使う事が出来る。

例: project

```
<project>
  <title>FSL basic monitoring program</title>
  <personnel id="pers-30" system="FLS">
    <individualName>
      <salutation>Dr.</salutation>
      <givenName>Eva</givenName>
      <givenName>M.</givenName>
      <surName>Scientist</surName>
    </individualName>
    <address>
      <deliveryPoint>Department of Ecology</deliveryPoint>
      <deliveryPoint>Fictitious State University</deliveryPoint>
      <deliveryPoint>PO Box 111111</deliveryPoint>
      <city>Ficity</city>
      <administrativeArea>FI</administrativeArea>
      <postalCode>11111-1111</postalCode>
    </address>
    <role>principalInvestigator</role>
  </personnel>
  <personnel id="pers-130" system="FLS">
    <individualName>
      <givenName>Monica</givenName>
      <givenName>D.</givenName>
      <surName>Techy</surName>
    </individualName>
    <address>
      <deliveryPoint>Department for Ecology</deliveryPoint>
      <deliveryPoint>Fictitious State University</deliveryPoint>
      <deliveryPoint>PO Box 111111</deliveryPoint>
      <city>Ficity</city>
      <administrativeArea>FI</administrativeArea>
      <postalCode>11111-1111</postalCode>
    </address>
    <role> principalInvestigator</role>
  </personnel>
  <abstract>
    <para>The FLS basic monitoring program consists of monitoring of
      arthropod populations, plant net primary productivity, and bird
      populations. Monitoring takes place at 3 locations, 4 times a year.
      Climate parameters a continuously measured at all stations.</para>
  </abstract>
  <funding>
    <para>National Science Foundation under Grant No. DEB 9999999,
      Ficticous Long-Term Ecological Research (FLS LTER)</para>
  </funding>
</project>
```

[entity] = dataTable, spatialRaster, spatialVector, storedProcedure, view, otherEntity

この要素が現れる位置は (XPath では)
/eml:eml/dataset/dataTable
/eml:eml/dataset/spatialRaster

/eml:eml/dataset/spatialVector
/eml:eml/dataset/storedProcedure
/eml:eml/dataset/view
/eml:eml/dataset/otherEntity

<dataTable>,<otherEntity>,<view> という実体型は、ASCII データテーブル、KML、画像といったような、一般的によく使われる多くのデータ構造を扱う。他の選択肢 (<spatialRaster>,<spatialVector>,<storedProcedure>) についてはこの文書の将来のバージョンの中でより深く述べられることだろう。しかしながら、空間データの実体 (<spatialRaster>,<spatialVector>) は、ESRI もしくは FGDC のメタデータが利用可能な時にはスタイルシート変換を通じて作成することができる (I.2.3 節を見よ)。そして例を後述する。FGDC もしくは ESRI 準拠のメタデータ記録を持っていない空間データセットは他の仕組みを用いて作成する必要がある (たとえば手作業で)。ベクトルデータの空間オブジェクト (線、点、ポリゴン) は、<spatialVector> ツリーを用いてベクトルのジオメトリ型、カウント、そして位相レベルについての情報を記入するべきである。グリッドあるいはラスタセル型のデータは <spatialRaster> ツリーを用いて記述されるべきであり、そしてラスタセルの空間構造、セルのデータ値、そしてもし画像センサからの派生物ならばその画像および個々のバンドの特性についての情報を含むべきである。

<otherEntity> 要素は、画像や KML ファイルに対して使用されるべきである。また KML ファイルは <spatialVector> 型に変換することもできる (後述)。

すべての EML データ実体はある共通の要素を持つ。それは EntityGroup ツリーと呼ばれるもので、どんなデータ資源についても一般的な情報を記述するものである。他の要素はそれぞれの実体型に対して独特のものである。EntityGroup 内の要素は最初に現れ、そしてその要素とは

<alternateIdentifier>
<entityName>
<entityDescription>
<physical> (任意の <access> を含む)
<coverage>
<methods>
<additionalInfo>

<alternateIdentifier> (任意) : 第一の識別子は entityName¹⁵ の id 属性 (たとえば <dataTable id="xxx">) にあるのが相応しいが、このタグは、異なるデータ管理システムから使用されるかもしれない追加の識別子を格納することができる。これは dataset 階層の <alternateIdentifier> 要素 (上述) と同じ様に使われる。

<entityName> (必須) : テーブル、ファイル、またはデータベースのテーブルの名前。しばしばこれは元々の ascii 表記のファイル名である。

<entityDescription> これは比較的長文で、より記述的な、データ実体についての説明文である。これはその実体中に含まれているデータを説明するべきである。すべての説明文と同様に、これは人間に読める形式であり、またデータの利用者が自分の用途に使えるかどうかを判断するのに

¹⁵ 訳註・例示と一致しない。entity の間違いか。

十分な情報を持つべきである。

<physical> ツリー (/eml:eml/dataset/[entity]/physical) はデータの物理フォーマットについてより深く説明する。

<objectName> は、ダウンロードした時、またはデータベースからテキストとしてエクスポートした時のファイル名であるべきである。<objectName> はしばしばファイルシステム上のファイルの名前であり、またはネットワーク上でアクセス可能な名前である。そして <entityName> と同じであるかも知れない。しかしながら、物理オブジェクト（つまりファイル）は幾つかの実体を持つ事ができる。たとえば Excel のワークブックは複数のシートを持つ。もし Excel ワークブックを公表する場合には（おすすめのやり方ではないが）、ワークシートの名前を <entityName> に記入し、ファイル名は <objectName> に入れなさい。

<externallyDefinedFormat> データ実体が規定の形式（NetCDF, KML, Excel）で作成されている場合は、ここにその形式の名前を記すこと。可能であれば、形式名に続けて mime 型（たとえば "image/jpeg"）を記すことを推奨する。ソフトウェア固有の記述には、製造元、プログラム、バージョンが含まれるべきである（たとえば "Microsoft Excel 2003"）。サンプリング地の KML ファイルはこの欄で "KML" あるいは "KMZ" として宣言することができる。

重要: 可能であればいつでも、サイトは時代遅れで独占的なバイナリ形式（MS-Excel ワークブックのような）でデータを公表しないようにするべきである。

<distribution> は、その資源がどのように配布されるのかについての情報を提供する。そしてこのツリーの内容は一般に <dataset> 階層に含まれる。しかしながら、いくつかのことについてはここで（訳註・entity 階層で）再度記述されるであろう。

entity 階層における <url> 要素の内容はデータを渡すべきであり、他のアプリケーションや使用方法ページを指し示すべきではない。例外は、データアクセスサーバを用いてデータダウンロードの記録を取るために LTER-NIS DAS プロキシ URL がここで使用されるかも知れないことである。DAS およびその使用方法についてはオンライン文書を読むこと。

http://im.lternet.edu/im_practices/metadata/das

<url> の属性である function は、値として "download" を持つべきである。function 属性が省略された場合は暗黙にその値となる。

EML 2.1 時点で、entity 階層の <distribution> ツリーにおいて任意の <access> 要素も存在する。この要素は、メタデータと切り離してデータ実体に対するアクセスを制御するために特に意図されている。<access> ツリーの使用方法についてより詳しくは、access についての全般的な議論（先述）を参照のこと。

推奨事項: 取り扱いに注意すべき情報（たとえば、Human Subjects Act ¹⁶によって取り扱われるデータ）を例外として、データ実体に対する読み取りアクセスは制限されるべきではない。

<coverage> はこの [entity] で使用される地理的、空間的、時間的範囲についての情報を提供する。より詳しい情報は dataset 階層における議論を見よ。

¹⁶ 訳註・人間を対象にした活動に関する法律。日本だと被験者保護法か。

<methods> はこの [entity] において情報を収集するのに使用した具体的な方法についての情報を提供する。より詳しい情報は dataset 階層の議論を見よ。

<additionalInfo> は、そのデータ型について、他の要素では記述することができない情報を何でも記すためのテキスト欄である。

例: EntityGroup における要素、ここでは <dataTable> 実体を見せる。

```
<dataTable>
  <entityName>invertebrates</entityName>
  <entityDescription> invertebrate sampling data</entityDescription>
  <physical>
    <objectName>fls-1.csv</objectName>
    <size unit="rows">561</size>
    <characterEncoding>ASCII</characterEncoding>
    <dataFormat>
      <textFormat>
        <numHeaderLines>1</numHeaderLines>
        <numFooterLines>0</numFooterLines>
        <recordDelimiter>\r</recordDelimiter>
        <numPhysicalLinesPerRecord>1</numPhysicalLinesPerRecord>
        <attributeOrientation>column</attributeOrientation>
        <simpleDelimited>
          <fieldDelimiter>,</fieldDelimiter>
        </simpleDelimited>
      </textFormat>
    </dataFormat>
    <distribution>
      <online>
        <onlineDescription>fls-1 Data File</onlineDescription>
        <url function="download">http://www.fsu.edu/lter/data/fls-1.csv</url>
      </online>
    </distribution>
  </physical>
```

それぞれのデータ型は、共通の要素に続いて、独特の要素セットを持つ。表 1 は entity 型の要約であり、典型的な使用法とその型に対して用いられる要素も示した。

Entity Type	典型的な使用法	EntityGroup に後続する要素
<dataTable>	静的な ASCII 文字の表 データベースからのエクスポート	<attributeList> <constraint> <caseSensitivity> <numberOfRecords>
<view>	データベースクエリから返されたデータ	<attributeList> <constraint> <queryStatement>
<storedProcedure>	データベース内のストアードプロシージャから返されたデータ	<attributeList> <constraint> <parameter>
<otherEntity>	画像、KML, 地図	<attributeList> <constraint> <entityType>
<spatialRaster>	リモセン画像	<attributeList> <constraint> <spatialReference> <georeferenceInfo> <horizontalAccuracy> <verticalAccuracy> <cellSizeYDirection> <numberOfBands> <rasterOrigin> <rows> <columns> <verticals> <cellGeometry> <toneGradation> <scaleFactor> <offset> <imageDescription>
<spatialVector>	ESRI shapfiles	<attributeList> <constraint> <geometry> <geometricObjectCount> <topolgyLevel> <spatialReference> <horizontalAccuracy> <vericalAccuracy>

attributeList

この要素が現れる位置は (XPath では)
/eml:eml/dataset/dataTable/attributeList

/eml:eml/dataset/view/attributeList
/eml:eml/dataset/storedProcedure/attributeList
/eml:eml/dataset/spatialRaster/attributeList
/eml:eml/dataset/spatialVector/attributeList
/eml:eml/dataset/otherEntity/attributeList

<attributeList> ツリーは <otherEntity> を除くすべてのデータ型に対して必須である。これはデータ実体中のすべての変数を個別の <attribute> 要素の中に記述する。その記述には個々の属性の名前と定義、分野、符号化された値の定義、そして他の関連情報が含まれる。

<attributeName> は属性の正式名であり、典型的にはデータテーブルの項目名である。これはしばしば短くて簡潔である。

<attributeLabel> (任意) は <attributeName> で与えられているものよりも曖昧さの少ない別の説明を提供するのに使われる。これは HTML 表示の中で列や行のヘッダとして使用されるかも知れない。

<attributeDefinition> は説明しようとしている属性の精度と完全な定義を与える。それは、データの利用者がその属性について正確に解釈できるように、属性の内容を十分に説明する。

<storageType> は、RDBMS に関しては、システム固有であろう。すなわち、Microsoft SQL の varchar、あるいは Oracle の datetime。この欄は、データを処理するシステムに対して、システムやプログラミング言語においてどのようにその属性を表現したらいいのかについての「ヒント」を表現する。しかしそれはその属性の変域の実際の表現とは明確に異なる¹⁷。システム固有でない値には、float, integer, string が含まれる。

<measurementScale> は、その属性についての値を得るのに用いる尺度の種類を示す。5種類の尺度のうちの一つを用いなければならない。すなわち、名義尺度、順序尺度、間隔尺度、比例尺度、そして日時型。用法は以下に従うこと。

<nominal> (名義) 尺度は名義カテゴリを表現するのに使用する。値は他の観測値から区別できるように割り当てられる。これには符号化された値のリスト (たとえば 1=男性、2=女性) や、平文の説明が含まれるだろう。文字列や単純な文章を含む列は名義的である。たとえば、plot1, plot2, plot3。

<ordinal> (順序的な) 値は、互いに論理的または順序的な関係性を持っているカテゴリであるが、値同士の違いの大きさは定義されないか、意味がない。例: 低、中、高

註・名義尺度および順序尺度はどちらも、テキストであっても列挙型のリストであっても、

<nonNumericDomain> 型と看做される。<enumeratedDomain> は符号化された値に適用され、<codeDefinition> または参照された実体はその符号の説明を持つ。<textDomain> に対して任意のパターンでそのテキストを記述することができる。たとえば、合衆国の電話番号は “\d\d\d-\d\d\d-\d\d\d” という書式で記述することができる。

<interval> この測定尺度は順序的である。しかしそれに加えて値と値の間に同じ大きさの単位を用いる。開始点は恣意的であり、そのためゼロ値には意味がない。例: セ氏温度尺度は等間隔の

¹⁷ 訳註・変数の型とその変数の実際の中身は一致しないこともある、ということか？

「度」を用いるが、このゼロは「絶対零度」（つまり分子運動が停止する温度）を表現しない。そしてセ氏 20 度は 10 度よりも「2 倍熱い」わけではない。

<ratio> 測定尺度は意味のあるゼロ点をもち、値と値の比率を比較することが理にかなっている。たとえば、ケルビン尺度は物質の運動エネルギーの総量を反映している（すなわち、ゼロは、物質がまったく熱エネルギーを伝達しない点である）。そしてそれゆえ、ケルビン単位で計測された温度は比例尺度の測定値である。濃度もまた比例尺度の測定値である。なぜなら、10m mol/l の溶液は 5m mol/l の溶液よりも 2 倍の物質を持っているからである。

註・<interval> と <ratio> 尺度は <unit>, <numericDomain>, <precision> を記述する追加のタグを要求する。

<unit> 単位は、正確な物理的単位によって記述されるべきである。データを説明しているが単位ではない用語は、<attributeDefinition> の中で使用するべきである。たとえば、「1 平方メートルあたりの炭素のミリグラム」と説明されているデータの場合、「炭素」は <attributeDefinition> に属し、一方 <unit> は “milligramPerMeterSquared” である。

<standardUnit> と <customUnit>: 単位名は、(EML に含まれている単位辞書から取られた) <standardUnit> か、または <customUnit> と <additionalMetadata> で定義されたもののどちらかでなければならない。LTER の EML 文書に対しては、単位の登録およびデータベースが利用可能である (<http://unit.lternet.edu/unitregistry/>)。LTER の文書については、すべての単位に対して <customUnit> 要素を使用し、Unit Registry から取り出した内容をそれに添えることが推奨される（たとえ単位が標準の単位辞書にリストされている場合でも）。LTER の文書における単位は Unit Registry に追加されるべきである。

一般的な目的のため、以下のガイドライン（ISO の推奨事項）を <customUnit> に適用する。単位は省略せずに全部書き出すべきである。単位修飾子（“squared”のような）は修飾しようとする単位に後続させるべきである。たとえば、meterSquared は好ましく、squareMeter は相応しくない。単位は単数形（“meter”のような）であるべきであり、（“meters”のような）複数形にするべきではない。

<numericDomain> このタグは、<numberType> および数値の属性に許容される最大値・最小値を特定する要素を含む。<numberType> は EML ハンドブックに説明されている通り real, natural, whole, integer として定義されるべきである。<bounds> は理論的または許容される最大値・最小値（規範的）である一データセットの中の、実際に観測された値の範囲（記述的）であるよりも。<bounds> ツリーは任意である。

<precision> はその属性の小数点以下の桁数を記述する。現在のところ、EML はひとつの列に対して複数の precision 値を認めていない。たとえば、魚の体長を格納する列は、ある種（たとえば大型魚）については 0.01 メートルの精度に合わせ、その他の種については 0.001 の精度に合わせるかも知れない。しかしその「魚の体長」についてのすべてのデータはひとつの属性に集められて、それらの適切な精度の値を用いて表示される。これらの場合に対しては、精度を省略することができる。しかし変数の精度情報は method/methodStep に詳しく記述するべきである。それとともに、<numericDomain> と <precision> の情報はその属性を表現するのに適切なシステム固有のデータ型を決定するのに有効である。たとえば、値の範囲が 0-50000 で精度が 1 の数値の

属性は C 言語においては "long" 値を用いて表現することができるが、精度が 0.5 に変化すると "float" 型が必要になるだろう。

<measurementType> 要素¹⁸である <dateTime> は、グレオリオ暦の日時値であり、ISO 8601 に従う書式で表記することを推奨される。許容される ISO 日時型の例は "YYYY-MM-DD" であり、実例を挙げると 2004-06-25 である。もしくは、より完全にすれば、"YYYY-MM-DDThh:mm:ssTZD" (たとえば 1997-07-16T19:20:30.45Z)。ISO 標準形式は日付の構成要素の構造についてかなり厳格である。過去のデータは標準化されていない日付を含んでおり、また既存の装備 (たとえばセンサー) はいまだに非標準の日付を生み出している可能性があるため、EML の作者たちは追加の許容可能な書式を提供した。EML の説明書に完全なリストがあるので見ること (第 V 節)。重要な注意点として、dateTime 欄は持続時間を記録するのに使うべきではない。その場合は、seconds, nominalMinute, nominalDay のような <standardUnit> を使うか、または <customUnit> を用いてその単位が SI 単位系の秒とどう関係するのかを定義すること。

註・この測定値型は以前の EML では <datetime> と呼ばれていた。内部的な一貫性のために全部小文字 (<datetime>) から "camel case" (<dateTime>) へと変更された。

<missingValueCode> は任意であるが、データセットの中に存在するどんな欠測値コードをも記述するため、記入されるべきである。欠測値は明示的に符号化されるべきであり、単なる空欄にするべきではない。欠測値コードは文字列であり、値ではない。どういう意味かと言うと、正確な解釈のために、この欄の内容はデータ値の場所に現れるものと正確に一致しなければならない。たとえば、もしデータが 0.01 の精度で出力されていて、かつ欠測値が "-9999.00" という書式であった場合、<missingValueCode> 要素の内容は "-9999" ではなくて "-9999.00" でなければならない。

次の例は 2 つの attribute ツリーを示している。ひとつめは SQL システムから生成されたもので、ストレージの型が定義されている。ふたつめの <attributeList> は <customUnit> に対するタグが含まれており、<additionalMetadata> ツリーにその単位の定義がある。

例: attributeList/attribute dataTable

```
<attributeList>
  <attribute id="att.1">
    <attributeName>Site</attributeName>
    <attributeDefinition>FLS-LTER sampling site</attributeDefinition>
    <storageType>integer</storageType>
    <measurementScale>
      <ordinal>
        <nonNumericDomain>
          <enumeratedDomain>
            <codeDefinition>
              <code>1</code>
              <definition>FLS study site 1</definition>
            </codeDefinition>
            <codeDefinition>
              <code>2</code>
              <definition>FLS study site 2</definition>
            </codeDefinition>
            <codeDefinition>
              <code>3</code>
              <definition>FLS study site 3</definition>
            </codeDefinition>
          </enumeratedDomain>
        </nonNumericDomain>
      </ordinal>
    </measurementScale>
  </attribute>
</attributeList>
```

¹⁸ 訳註・ EML 規格には measurementType という要素は存在しない。measurementScale の誤りだろう。

```

        <code>4</code>
        <definition>FLS study site 4</definition>
      </codeDefinition>
    </enumeratedDomain>
  </nonNumericDomain>
</ordinal>
</measurementScale>
</attribute>
<attribute id="att.2">
  <attributeName>Year</attributeName>
  <attributeDefinition>Calendar year of the observation from years 1990 -
    2010</attributeDefinition>
  <storageType>integer</storageType>
  <measurementScale>
    <dateTime>
      <formatString>YYYY</formatString>
      <dateTimePrecision>1</dateTimePrecision>
      <dateTimeDomain>
        <bounds>
          <minimum exclusive="false">1993</minimum>
          <maximum exclusive="false">2003</maximum>
        </bounds>
      </dateTimeDomain>
    </dateTime>
  </measurementScale>
</attribute>
<attribute id="att.3">
  <attributeName>Month</attributeName>
  <attributeDefinition>Calendar month of the observation</attributeDefinition>
  <storageType>integer</storageType>
  <measurementScale>
    <dateTime>
      <formatString>MM</formatString>
      <dateTimePrecision>1</dateTimePrecision>
      <dateTimeDomain>
        <bounds>
          <minimum exclusive="false">1</minimum>
          <maximum exclusive="false">12</maximum>
        </bounds>
      </dateTimeDomain>
    </dateTime>
  </measurementScale>
</attribute>
<attribute id="att.4">
  <attributeName>Day</attributeName>
  <attributeDefinition>Calendar day of the observation</attributeDefinition>
  <storageType>integer</storageType>
  <measurementScale>
    <dateTime>
      <formatString>DD</formatString>
      <dateTimePrecision>1</dateTimePrecision>
      <dateTimeDomain>
        <bounds>
          <minimum exclusive="false">1</minimum>
          <maximum exclusive="false">31</maximum>
        </bounds>
      </dateTimeDomain>
    </dateTime>
  </measurementScale>
</attribute>
<attribute id="att.5">
  <attributeName>Transect</attributeName>
  <attributeDefinition>Transect number (randomly placed within the study
    site)</attributeDefinition>
  <storageType>integer</storageType>
  <measurementScale>
    <nominal>
      <nonNumericDomain>
        <textDomain>
          <definition>Transect number (randomly placed)</definition>
        </textDomain>
      </nonNumericDomain>
    </nominal>
  </measurementScale>
</attribute>

```

```

    </nominal>
  </measurementScale>
</attribute>
<attribute id="att.6">
  <attributeName>Species_Code</attributeName>
  <attributeDefinition>Coded species name of the insect
    observed</attributeDefinition>
  <storageType>string</storageType>
  <measurementScale>
    <nominal>
      <nonNumericDomain>
        <enumeratedDomain>
          <codeDefinition>
            <code>G0</code>
            <definition>Other grasshoppers</definition>
          </codeDefinition>
          <codeDefinition>
            <code>G1</code>
            <definition>Orchelimum spp.</definition>
          </codeDefinition>
          <codeDefinition>
            <code>G2</code>
            <definition>Paroxya clavuliger</definition>
          </codeDefinition>
          <codeDefinition>
            <code>G3</code>
            <definition>Orphulella pelidna</definition>
          </codeDefinition>
          <codeDefinition>
            <code>G4</code>
            <definition>Mermiria intertexta</definition>
          </codeDefinition>
          <codeDefinition>
            <code>G5</code>
            <definition>Schistocerca obscura</definition>
          </codeDefinition>
          <codeDefinition>
            <code>G6</code>
            <definition>Romalea guttatus</definition>
          </codeDefinition>
        </enumeratedDomain>
      </nonNumericDomain>
    </nominal>
  </measurementScale>
</attribute>
<attribute id="att.7">
  <attributeName>Count</attributeName>
  <attributeDefinition>Number of individuals observed</attributeDefinition>
  <storageType>integer</storageType>
  <measurementScale>
    <interval>
      <unit>
        <standardUnit>number</standardUnit>
      </unit>
      <precision>1</precision>
      <numericDomain>
        <numberType>whole</numberType>
        <bounds>
          <minimum exclusive="false">0</minimum>
        </bounds>
      </numericDomain>
    </interval>
  </measurementScale>
  <missingValueCode>
    <code>NaN</code>
    <codeExplanation>value not recorded or invalid</codeExplanation>
  </missingValueCode>
</attribute>
<attribute id="att.8">
  <attributeName>cond</attributeName>
  <attributeLabel>Conductivity</attributeLabel>
  <attributeDefinition>measured with SeaBird Elecronics

```

```

    CTD-911</attributeDefinition>
<storageType>float</storageType>
<measurementScale>
  <ratio>
    <unit>
      <customUnit>siemensPerMeter</customUnit>
    </unit>
    <precision>0.0001</precision>
    <numericDomain>
      <numberType>real</numberType>
      <bounds>
        <minimum exclusive="false">0</minimum>
        <maximum exclusive="false">40</maximum>
      </bounds>
    </numericDomain>
  </ratio>
</measurementScale>
</attribute>
</attributeList>

```

以下の例では、ESRI メタデータ形式から XSLT (スタイルシート) 経由で変換した `<spatialVector>` と `<spatialRaster>` についての完全な entity ツリーを示している。より詳しくは I.2.3 節を見よ。

例: spatialVector に対する実体および属性情報

```

<spatialVector id="Landuse for Ficity in 1955">
  <entityName>Landuse for Ficity in 1955</entityName>
  <entityDescription>This GIS layer represents a reconstructed generalized
  landuse map for the area of current Ficity around the time period of
  1955.</entityDescription>
  <physical>
    <objectName>fls-20.zip</objectName>
    <dataFormat>
      <externallyDefinedFormat>
        <formatName>Shapefile</formatName>
      </externallyDefinedFormat>
    </dataFormat>
    <distribution>
      <online>
        <onlineDescription>fls-20 Zipped Shapefile File</onlineDescription>
        <url function="download"> http://www.fsu.edu/lter/data/fls-20.zip</url>
      </online>
    </distribution>
  </physical>
  <attributeList id="Landuse for Ficity in 1955.attributeList">
    <attribute id="Landuse for Ficity in 1955.FID">
      <attributeName>FID</attributeName>
      <attributeDefinition>Internal feature number.</attributeDefinition>
      <storageType typeSystem="http://www.esri.com/metadata/esriprof80.html">
        OID</storageType>
      <measurementScale>
        <nominal>
          <nonNumericDomain>
            <textDomain>
              <definition>Sequential unique whole numbers that are
              automatically generated.</definition>
            </textDomain>
          </nonNumericDomain>
        </nominal>
      </measurementScale>
    </attribute>
    <attribute id="Landuse for Ficity in 1955.Shape">
      <attributeName>Shape</attributeName>
      <attributeDefinition>Feature geometry.</attributeDefinition>
      <storageType typeSystem="http://www.esri.com/metadata/esriprof80.html">
        Geometry</storageType>
      <measurementScale>
        <nominal>
          <nonNumericDomain>
            <textDomain>

```

```

        <definition>Coordinates defining the features.</definition>
      </textDomain>
    </nonNumericDomain>
  </nominal>
</measurementScale>
</attribute>
<attribute id="Landuse for Ficity in 1955.Z955">
  <attributeName>Z955</attributeName>
  <attributeDefinition>This field signifies the landuse value for each
  polygon.</attributeDefinition>
  <storageType typeSystem="http://www.w3.org/2001/XMLSchema-datatypes">
  string</storageType>
  <measurementScale>
    <nominal>
      <nonNumericDomain>
        <enumeratedDomain>
          <codeDefinition>
            <code>Agriculture</code>
            <definition>Agricultural land use</definition>
          </codeDefinition>
          <codeDefinition>
            <code>Urban</code>
            <definition>Urbanized area</definition>
          </codeDefinition>
          <codeDefinition>
            <code>Desert</code>
            <definition>Unmodified area</definition>
          </codeDefinition>
          <codeDefinition>
            <code>Recreation</code>
            <definition>Recreational land use</definition>
          </codeDefinition>
        </enumeratedDomain>
      </nonNumericDomain>
    </nominal>
  </measurementScale>
</attribute>
</attributeList>
<geometry>Polygon</geometry>
<geometricObjectCount>78</geometricObjectCount>
<spatialReference>
  <horizCoordSysName>NAD_1927_UTM_Zone_12N</horizCoordSysName>
</spatialReference>
</spatialVector>

```

例: spatialRaster に対する実体・属性情報

```

<spatialRaster id="fi_24k">
  <entityName>fi_24k</entityName>
  <entityDescription>Fictitious State 7.5 Minute Digital Elevation
  Model</entityDescription>
  <physical>
    <objectName>fls-30.zip</objectName>
    <dataFormat>
      <externallyDefinedFormat>
        <formatName>Esri Grid</formatName>
      </externallyDefinedFormat>
    </dataFormat>
    <distribution>
      <online>
        <onlineDescription>fls-30 zipped raster data File</onlineDescription>
        <url function="download"> http://www.fsu.edu/lter/data/fls-30.zip</url>
      </online>
    </distribution>
  </physical>
  <attributeList id="fi_24k.attributeList">
    <attribute id="fi_24k.ObjectID">
      <attributeName>ObjectID</attributeName>
      <attributeDefinition>Internal feature number.</attributeDefinition>
      <storageType typeSystem="http://www.esri.com/metadata/esriprof80.html">
      OID</storageType>
    </attribute>
  </attributeList>
</spatialRaster>

```

```

    <measurementScale>
      <nominal>
        <nonNumericDomain>
          <textDomain>
            <definition>Sequential unique whole numbers that are automatically
              generated.</definition>
          </textDomain>
        </nonNumericDomain>
      </nominal>
    </measurementScale>
  </attribute>
  <attribute id="fi_24k.Cell Value">
    <attributeName>Cell Value</attributeName>
    <attributeDefinition>Elevation Value</attributeDefinition>
    <storageType typeSystem="http://www.esri.com/metadata/esriprof80.html">
      Integer</storageType>
    <measurementScale>
      <ratio>
        <unit>
          <standardUnit>meter</standardUnit>
        </unit>
        <precision>1</precision>
        <numericDomain>
          <numberType>integer</numberType>
          <bounds>
            <minimum exclusive="true">-5193</minimum>
            <maximum exclusive="true">14785</maximum>
          </bounds>
        </numericDomain>
      </ratio>
    </measurementScale>
  </attribute>
  <attribute id="fi_24k.Count">
    <attributeName>Count</attributeName>
    <attributeDefinition>Count</attributeDefinition>
    <storageType typeSystem="http://www.esri.com/metadata/esriprof80.html">
      Integer</storageType>
    <measurementScale>
      <ratio>
        <unit>
          <standardUnit>number</standardUnit>
        </unit>
        <precision>1</precision>
        <numericDomain>
          <numberType>whole</numberType>
        </numericDomain>
      </ratio>
    </measurementScale>
  </attribute>
</attributeList>
<spatialReference>
  <horizCoordSysName>NAD_1927_UTM_Zone_12N</horizCoordSysName>
</spatialReference>
<horizontalAccuracy>
  <accuracyReport>not available</accuracyReport>
</horizontalAccuracy>
<verticalAccuracy>
  <accuracyReport>not available</accuracyReport>
</verticalAccuracy>
<cellSizeXDirection>30.0</cellSizeXDirection>
<cellSizeYDirection>30.0</cellSizeYDirection>
<numberOfBands>1</numberOfBands>
<rasterOrigin>Upper Left</rasterOrigin>
<rows>21092</rows>
<columns>18136</columns>
<verticals>1</verticals>
<cellGeometry>matrix</cellGeometry>
</spatialRaster>

```

constraint

この要素が現れる位置は (XPath では)

/eml:eml/dataset/dataTable/constraint
/eml:eml/dataset/view/constraint
/eml:eml/dataset/spatialRaster/constraint
/eml:eml/dataset/spatialVector/constraint
/eml:eml/dataset/storedProcedure/constraint

<constraint> ツリーは実体同士（たとえばテーブル）の間の整合性制約を記述するためのものである、それらがリレーショナルマネジメントシステムにおいて管理されているかのように。

<constraint> ツリーの使用は、データ要素がリレーショナルデータベースからの整合性制約を含んでいる時に奨励される。複数の列を巻き込むような制約は下記の例のように記述されるべきである。

<constraintType> はリレーショナルデータベースの参照整合性制約を記述する。これにはプライマリキー、外部キー、ユニークキーなどが含まれる。<constraint> ツリーが使用される時には、その制約（外部キー）において参照される可能性のあるすべての実体は同じパッケージの中でアクセス可能であるべきである。6種類の制約のうちのひとつを使用するべきである。primaryKey, uniqueKey, checkConstraint, foreignKey, joinCondition, nonNullConstraint。

<primaryKey> は、その制約定義に関係している実体におけるプライマリキーを宣言する要素である。

<uniqueKey> は、参照された実体の内部のユニークキーを表現する要素である。これはプライマリキーとは異なり、他の実体に対する暗黙の外部キー関係をまったく形成しない。しかしながらこれはその実体の内部において重複しないことが要求される。

<nonNullConstraint> は、この実体の中のある属性について null 値が存在するべきでないということを示す制約を定義する。

註・<primaryKey>,<uniqueKey>,<nonNullConstraint> は、この制約を適用するべき属性の組を定義するための追加の <key> タグを要求する。

<checkConstraint> は実体内で条件節を検査する制約を定義する。

<foreignKey> は、CHECK 制約のための SQL 文もしくはその条件を他の言語で実装したものを定義する。一般的に、これは実体内および実体間で値を制約するための手段を提供する。またこれは、符号の説明のためのリンクテーブルに対して有意義な方法で手段を提供する。

<joinCondition> は、この実体と、他の実体のプライマリキーを関連付けるような、実体間の外部キー関係を定義する。

註・すべての <ConstraintType> 要素は追加の <constraintName> と <attributeReference> タグを要求する。

例: 制約（単純なもの、およびプライマリキーと外部キーを混合したもの）

```
<constraint id="soil_chemistry.PRIMARY">  
  <primaryKey>  
    <constraintName>PRIMARY</constraintName>  
    <key>  
      <attributeReference>soil_chemistry.ID</attributeReference>
```

```

    </key>
  </primaryKey>
</constraint>
<constraint id="invertebrate_pk">
  <primaryKey>
    <constraintName>invertebrate_pk</constraintName>
    <key>
      <attributeReference>att.1</attributeReference>
      <attributeReference>att.2</attributeReference>
      <attributeReference>att.3</attributeReference>
      <attributeReference>att.4</attributeReference>
    </key>
  </primaryKey>
</constraint>
<constraint id="soil_chemistry.FK_soil_chemistry_sites">
  <foreignKey>
    <constraintName>FK_soil_chemistry_sites</constraintName>
    <key>
      <attributeReference>soil_chemistry.site_id</attributeReference>
    </key>
    <entityReference>sites</entityReference>
  </foreignKey>
</constraint>

```

additionalMetadata

この要素が現れる位置は (XPath では)
 /eml:eml/additionalMetadata

<additionalMetadata> は、記述対象の資源に関係する他のどんなメタデータでも記入するための融通がきく欄である。その内容は妥当な XML でなければならない。attribute の単位が <customUnit> である場合はその単位をこのツリーの中に記述しなければならない (下記参照)。

<describes> (任意) は、ある EML 要素の id 属性へのポインタである (他の場所で記述された id)。このポインタは、それが指し示す属性の内容と一致しなければならない。それによって自動的処理によって <additionalMetadata> とその記述された属性が関連付けられるのである。<describes> 要素が省略された場合、<additionalMetadata> の内容はその EML 文書の全体に適用されると看做される。

<metadata> 要素は追加のメタデータのためのコンテナである。その内容は妥当な XML であればどんなものでもよい。この要素は EML を拡張するために使用されるべきである、EML 規格の他の部分で既に利用可能になっているのではないメタデータを記入するためであったり、中核的なメタデータを越える部分で必要とされるサイト固有あるいはシステム独自の拡張を記入するために。この欄に格納される追加のメタデータは、これに先行する <describes> 要素において参照されている要素を説明する。<describes> が使用されていない場合は、<metadata> は <additionalMetadata> とそれが記述している属性との関係を定義するための情報を十分に含まなければならない。

その内容が <customUnit> の定義のためのものである場合は、それは <unitList> で始まる妥当な STMMML マークアップであるべきである。<unitList> は、<additionalMetadata> と特定の EML 要素を関連づけるための <describes> タグの使用に対するひとつの例外である。EML パーサは <customUnit> の説明が <unitList> の <unit> 要素の id 属性 (すなわち /eml:eml/additionalMetadata/metadata/unitList/unit) の中に見付かることを期待する。たとえば、“stmmml:unit id=”siemenPerMeter”” は <customUnit> の “siemenPerMeter” を指し示す。<customUnit> を記述する <unit> のための STMMML 形式の内容は Unit Registry

(<http://unit.lternet.edu/unitregistry/>) から STMML 形式で取り出すべきである。

例: additionalMetadata の独自単位

```
<additionalMetadata>
  <metadata>
    <stmml:unitList>
      <stmml:unit id="siemensPerMeter" name="siemensPerMeter" abbreviation="S/m"
        unitType="conductance" parentSI="siemen" multiplierToSI="1"
        constantToSI="0">
        <stmml:description>conductivity unit</stmml:description>
      </stmml:unit>
    </stmml:unitList>
  </metadata>
</additionalMetadata>
```

III 外部アプリケーションとの互換性に対する推奨事項

III.1 Metacat

KNB Metacat サーバおよび関連の Morpho アプリケーションは、全般的に、EML シェーマに従う XML 文書のすべてと互換性がある。しかしながら、以下の推奨事項により、それらのツールを用いた EML 文書の管理や表示が強化され、また LNO Metacat ハーベスタアプリケーションを用いて文書を自動的にアップロードできるようになるだろう。

a) 文書 id とリビジョン番号

KNB Metacat に提供された EML に対する packageId 属性は、以下の形式になっているべきである。

knb-lter-[site].[dataset number].[revision], たとえば knb-lter-fls.187.4

Metacat (および拡張として Metacat ハーベスタ) は、文書の管理と同期のために数値のデータセット id とリビジョン番号に頼っている。これは、非数値の id を使用していたり、データセットのバージョン管理をしていないサイトに対して回避策を余儀なくさせる。実行可能な解決法としては、Metacat および XSLT 変換用に最適化された EML を特異的に生成する方法がある。現在のところ、**packageId** 属性は、LTER のデータカタログ上での検索に対してサイトの EML 文書を同定するのに使用される。KNB Metacat に提供された EML に対するサイトの **packageId** 属性は、上記のような形式でか、またはそのサイトと LNO データカタログ管理者によって合意された他の形式で標準化されるべきである。

b) Metacat は EML 文書で指定されたアクセス制御を完全に扱う。¹⁹EML 文書のアクセス制御について、より詳しくは II.2 節を見よ。

c) LTER の Metacat は、高速な検索のためにあるパスをインデックスするように設定されている。EML の作者は必ず以下のノードに適切な内容を記入するようにするべきである。

title

¹⁹訳註・JaLTER データベースではアクセス制御を行っていない。

keyword
abstract
abstract/path
abstract/section/path
organizationName, surName or givenName (under associatedParty or creator)
url
taxonRankValue
boundingCoordinates (under geographicCoverage)
geographicDescription
entityName

III.2 LTERMapS

LTERMapS アプリケーション: LTERMapS は LTER サイトの情報に対するインターネット地図アプリケーションであり、ユーザがサイト情報や空間データ・表データ（ネットワーク上のデータ、サイトレベルのデータどちらでも）を可視化、検索、ダウンロード、探索できるようにする。
(<http://www.lternet.edu/map/>).

EML データセットを LTERMapS アプリケーションに接続するためには、dataset 階層の **<geographicCoverage>** ツリーを記入しなければならない。

加えて、詳細なサンプリング地点を2つの方法でリストすることができる。どちらの方法も LTER MapS プロジェクトへの参加が許容される。

<spatialSamplingUnits> の中身: **<spatialSamplingUnits>** の coverage 要素の中身としてステーションの記述を含める方法は、EML をその場で ('on the fly') 生成するサイトにとって一番簡単な方法だろう。

データ実体への参照: 分離したデータ実体にサンプリングステーションをリストすることができる。それは `methods/*/spatialSamplingUnits/referencedEntityId` によって参照されるべきである。これはステーションの長いリストに対する最良の選択肢だろう。個別の **<geographicCoverage>** 要素によってリストを作ったら表示がとて長くなるだろうから。

これらのデータ実体は以下の形式のいずれかであるべきである。ESRI shape file, KML/KMZ (**<otherEntity>** を用いて)、カンマ区切りのテキスト (**<dataTable>** を用いて)。ステーションをカンマ区切りの表にリストした場合、メタデータの検索エンジンがそれらの位置を見付けることができないことに気をつけるべきである。研究サイトの位置についての shapefile, KML/KMZ ファイルの作成方法については、説明書が利用可能になっている。

http://im.lternet.edu/im_practices/gis/locations

III.3 EML における由来追跡

Wiktionary は "provenance" を、ある物についての起源の場所や出典、と定義している。(June 2010, <http://en.wiktionary.org/wiki/provenance>) EML は派生物（たとえばデータセット）の由来

を管理する能力がある。ただし、元になった物も EML によって記述されていてかつネットワークプロトコルを通して（たとえば Metacat のようなメタデータカタログを通じて）参照可能であるなら。

派生物についての EML 文書は 3 つの重要な構成要素を持つ。

- 1) 派生物のメタデータ
- 2) 派生物を作成するために元資料に適用するアプリケーション処理（たとえば単位変換）について、適切であればどんなものでも、の説明
- 3) 元資料のメタデータ

派生物のメタデータは、ここで提供されているのと同じ「最善のやり方」(“best practice”)に従うべきである。次の 2 つ（アプリケーション処理と元資料のメタデータ）は明確に派生物の由来に関係している。EML の `<methods>` サブツリーは由来情報を切り取るのに理想的なコンテナである、その目的が派生についての記述的な情報を文書化することであるから。派生のそれぞれの手順は `<methodStep>` 要素に格納されるだろう。

新しい EML 文書における `<methods>` ツリーの位置は、派生物がどこで記述されるかに依存するだろう。たとえば、`<dataset>` 階層の `<methods>` ツリーは、派生物全体についての由来情報を記述するのに使用できる。

例: 由来

```
<methods>
  <methodstep>
    <description>
      <section>
        <para>We utilize NPP data collected from 1906 to 2006 from the ONL
          LTER site. The ONL NPP data unit definition is kg/m^2/yr. This unit
          does not require conversion.</para>
      </section>
    </description>
    <datasource>
      <title>npp data from onl 1906 to 2006</title>
      <creator>
        <organizationname>ONL LTER</organizationname>
      </creator>
      <distribution>
        <online>
          <url>http://metacat.lternet.edu/knb/metacat/knb-lter-onl.23.1</url>
        </online>
      </distribution>
      <contact>
        <organizationName>ONL LTER</organizationName>
        <positionName>ONL Information Manager</positionName>
        <electronicMailAddress>im@onl.lternet.edu</electronicMailAddress>
      </contact>
    </datasource>
  </methodstep>
</methods>
```

III.4 EML 適合検査

註・この節は、現在 LNO で開発中の EML 適合検査プログラムの特徴を先取りしている。この節は今後もっと発展していくだろう。

EML 適合検査プログラム (EML Congruence Checker, ECC) は、EML ファイルと関連するデータファイルの品質を改善し、それらのファイルが LTER ネットワーク情報システムへの提出および PASTA による収集に向けて調整されていることを確実にするために企画された。以下の検査が実行される

EML の決定的な特徴: EML ファイルを下記の観点から検査する。

- (1) well-formed である²⁰
- (2) XML シェーマ²¹に従っている
- (3) ID と参照についての EML の規則に従っている

これらの機能は現在オンラインの EML ID and Reference Parser (<http://knb.ecoinformatics.org/emlparser/>) によって提供されている。EML ファイルがこれらのテストのいずれかに失敗した場合、ECC プログラムは失敗の原因を示すエラーメッセージを表示して停止する。

EML の推奨される特徴: EML ファイルが、この文書から選択された「最善のやり方」とどのくらい一致しているか検査する。ECC プログラムは以下のテストに対する詳細な情報を返す。

- (1) 完全性。ECC プログラムは、EML ファイルの中に推奨された要素が存在するかどうかを検査し、EML ファイル中に存在しない推奨要素のリストを返す。
- (2) URL。ECC プログラムは EML ファイル中のすべての URL について、それらが有効かどうかを検査し、切れている URL のリストを返す。
- (3) キーワード。ECC プログラムは EML ファイル中のキーワードを最新の LTER 統制語彙のリストと比較して、どのキーワードがその統制語彙に存在し、どれが存在しないかを示すリストを返す。
- (4) 単位。ECC プログラムは EML ファイル中の単位を最新の LTER 単位辞書と比較して、どの単位が標準単位で、どれが独自定義単位で、どれが LTER 単位辞書に存在しないかを示すリストを返す。

EML とデータ的一致: EML ファイル中のメタデータの記述が、データファイルを PASTA に取り入れる助けになるのに十分に正確で完全かどうかを検査する。データテーブルに対しては、ECC プログラムは以下のテストについて詳細な情報を返す。

- (1) ECC プログラムは、`<physical><distribution>` ツリーの情報がデータファイルを取り出

²⁰ 訳註・XML の文法に適合しているということ。

²¹ 訳註・EML 規格を定義した仕様書

すのに充分であるかどうか照合する。

(2) ECC プログラムは、`<physical><dataFormat>` ツリーの情報がデータファイルの構造を正確に記述しているかどうか照合する（ファイルサイズ、圧縮とエンコード方式、ヘッダ・フッタの行数、レコードの区切り文字、フィールドの区切り文字）。列の数は `<attributeList>` ツリーの属性の数と一致するべきである。

(3) 個々の属性について、ECC プログラムは、`<measurementScale>` ツリーの情報がデータファイル中の属性の値を正確に記述しているかどうかを照合する。たとえば、すべての `<enumeratedDomain>` の値は符号化された値の特定のセットに含まれているべきである。`<numericDomain>` の値は指定された `<numberType>` に適合するべきである。また `<bounds>` 要素が存在する場合は、すべての `<numericDomain>` の値は `<minimum>` と `<maximum>` によって与えられる範囲内に入っているべきである。`<dateTime>` の値は指定された `<formatString>` に一致するべきである。欠測値コード（もし存在するならば）は指定された `<missingValueCode><code>` に適合するべきである。

同様のテストは他のデータ型 (spatialRaster, spatialVector, otherEntity) に対しても実行される。

IV この文書と共に提供される EML のサンプルファイルの説明

例 1 この文書中の例の元となった EML の完全版²²

例 2: 由来関係を示すための EML の例

V 他の資源

年代順に示す。アプリケーションはタイトルのアルファベット順。

文書:

EML 2.1.0 シェーマ、および規範文書 (2008): <http://knb.ecoinformatics.org/software/eml/>

EML-2.1.0 FAQ (2008): <http://knb.ecoinformatics.org/software/eml/eml-2.1.0/eml-faq.html>

ESRI 形式 から EML に変換するスタイルシートについての文書および推奨事項 (2008):
<http://im.lternet.edu/project/Esri2Eml>

LTER サイトのための EML Best Practices バージョン 1 (2004):
http://im.lternet.edu/sites/im.lternet.edu/files/emlbestpractices_oct2004_final.pdf

EML ハンドブック (2003): <http://im.lternet.edu/sites/im.lternet.edu/files/emlHandbook.pdf>

²² 訳註・ <http://im.lternet.edu/node/910> から入手可能。例 2 も同様。

Report from the 2003 EML implementation workshop at SEV (2003):

アプリケーションとウェブサイト:

LNO データアクセスサーバ: http://im.lternet.edu/im_practices/metadata/das

LNO Metacat ハーベスタ情報: <http://im.lternet.edu/project/LTERProjectDatabase>

LTER 単位辞書: <http://unit.lternet.edu>

Metacat と Morpho: <http://ecoinformatics.org/tools.html>

CUASHI 水文情報システムで使用されている 観測データモデル (ODM): <http://his.cuahsi.org/index.html>

VI この文書に貢献してくれた人々

EML Best Practices Workshop, LTER Network Office, Albuquerque, NM (June 21-23, 2010)

Best Practices Working Group:

- Dan Bahauddin (CDR)
- Emery Boose (HFR)
- Corinna Gries (NTL)
- Margaret O' Brien (SBC)
- Ken Ramsey (JRN)
- Mark Servilla (LNO)
- Theresa Valentine (AND)
- Jonathan Walsh (BES)
- M. Gastil Gastil-Buhl (MCR)

Co-participants:

- James Brunt (LNO)
- Inigo San Gil (LNO)
- Philip Tarrant (CAP)
- Yang Xia (LNO)

EML Best Practices Workshop, LTER Network Office, Albuquerque, NM (May 19-20, 2004)

Best Practices Working Group:

- James Brunt (LNO)
- Corinna Gries (CAP)
- Jeanine McGann (LNO)
- Margaret O' Brien (SBC)

- Ken Ramsey (JRN)
- Wade Sheldon (GCE)

Co-participants:

- Duane Costa (LNO)
- Mark Servilla (LNO)

LTER EML Implementation Workshop, Sevilleta Research Station, Sevilleta National Wildlife Refuge, NM (June 9-10, 2003)

Best Practices Working Group:

- Barbara Benson (NTL)
- James Brunt (LNO)
- Don Henshaw (AND)
- John Vande Castle (LNO)
- Kristin Vanderbilt (SEV)

Working Group Support:

- Jeanine McGann (LNO)

索引

- abstract, 4, 11, 16, 27, 44
- access, 8–11, 19, 29, 30
- additionalInfo, 11, 29, 31
- additionalMetadata, 9, 35, 42
- address, 14, 15
- additionalMetadata, 34
- alternateIdentifier, 11
- alternateIdentifier, 9, 11, 12, 29
- alternateTimeScale, 21
- associatedParty, 9, 11, 14, 44
- associatedParty, 15
- attribute, 10, 25, 33
- attributeDefinition, 33, 34
- attributeLabel, 33
- attributeList, 32, 33, 35, 47
- attributeName, 4, 33
- attributeReference, 41

- boundingAltitudes, 21
- boundingCoordinates, 19, 21, 44
- bounds, 34, 47

- caseSensitivity, 32
- cellGeometry, 32
- cellSizeYDirection, 32
- checkConstraint, 41
- citation, 4, 9, 25, 26, 28
- classificationSystem, 22
- classificationSystemCitation, 23
- classificationSystemModification, 23
- code, 47
- codeDefinition, 33
- columns, 32
- connection, 18
- constraint, 5, 10, 32, 40, 41
- constraintName, 41
- ConstraintType, 41
- constraintType, 41
- contact, 9, 11, 12, 15
- costumUnit, 4
- coverage, 8, 9, 11, 19, 25, 28–30

- creator, 9–12, 14, 25, 44
- customUnit, 34, 35, 42

- dataFormat, 47
- dataset, 4, 5, 9–12, 18, 25, 30, 45
- datasetGPolygon, 21
- datasetGPolygonExclusionGRing, 21
- datasetGPolygonOuterGRing, 21
- dataSource, 25
- dataTable, 9, 10, 25, 28, 29, 31, 32, 44
- dateTime, 35, 47
- datetime, 35
- describes, 42
- description, 25, 26
- descriptor, 28
- distribuion, 30
- distribution, 9, 11, 18, 19, 25, 27, 30, 46

- electronicMailAddress, 14, 15
- eml:eml, 8, 10, 12
- endDate, 22
- entityDescription, 29
- entityName, 29, 30, 44
- entityType, 32
- enumeratedDomain, 33, 47
- externallyDefinedFormat, 30

- foreignKey, 41
- formatString, 47
- fundingSource, 28

- generalTaxonomicCoverage, 22
- geographicCoverage, 8, 9, 19–21, 44
- geographicDescription, 21, 44
- geometricObjectCount, 32
- geometry, 32
- georeferenceInfo, 32
- givenName, 13, 44
- gRingPoint, 21

- horizontalAccuracy, 32

id, 8–10
 identifierName, 23
 imageDescription, 32
 individualName, 12
 inline, 18
 instrumentation, 25
 intellectualRights, 11, 17
 interval, 33, 34

 joinCondition, 41

 key, 41
 keyword, 16, 44
 keywordSet, 11, 16
 keywordThesaurus, 16

 language, 11
 literalLayout, 7

 maintenance, 8, 11, 22, 24
 maximam, 47
 measurementScale, 33, 47
 measurementType, 35
 mediumName, 18
 metadata, 42
 metadataProvider, 9, 11, 14
 metadataProvier, 14
 method, 24
 methods, 5, 8, 20, 24–26, 29, 31, 45
 methodStep, 4, 25, 45
 minimum, 47
 missingValueCode, 35, 47

 nominal, 33
 nonNullConstraint, 41
 nonNumericDomain, 33
 numberOfBands, 32
 numberOfRecords, 32
 numberType, 34, 47
 numericDomain, 34, 47
 nunNullConstraint, 41

 objectName, 30
 offline, 18
 offset, 32

 online, 18
 onlineDescription, 18
 onlineURL, 14, 15
 orderedlist, 7
 ordinal, 33
 organizationName, 13–15, 44
 otherEntity, 9, 28, 29, 32, 33, 44, 47

 packageId, 9, 10, 43
 para, 7
 parameter, 32
 personnel, 27
 phone, 14, 15
 physical, 29, 30, 46, 47
 positionName, 13
 precision, 34
 primaryKey, 41
 project, 8, 9, 11, 25, 27
 protocol, 4, 9, 25, 26
 protocols, 8, 24
 pubDate, 11, 15, 16
 publisher, 9, 11, 15
 pubPlace, 11
 purpose , 11

 qualityControl, 25, 26
 queryStatement, 32

 rangeOfDates, 21
 rasterOrigin, 32
 ratio, 34
 relatedProject, 27
 role, 14
 rows, 32

 sampling, 5, 25
 samplingDescription, 26
 scaleFactor, 32
 scope, 9, 10
 section, 7
 series, 11
 shortName, 11, 12
 singleDate, 21
 software, 4, 9

spatialRaster, 9, 28, 29, 32, 38, 47
spatialReference, 9, 32
spatialSamplingUnit, 20
spatialSamplingUnits, 6, 26, 44
spatialVector, 9, 28, 29, 32, 38, 47
standardUnit, 34, 35
storageType, 33
storedProcedure, 9, 28, 29, 32
studyAreaDescription, 25
studyAreaDiscription, 28
studyExtent, 25, 26
surName, 12, 13, 44
system, 9, 10

taxonomicClassification, 22
taxonomicClassification, 22
taxonomicCompleteness, 23
taxonomicCoverage, 8, 9, 19, 22
taxonomicProcedures, 23
taxonRankName, 22
taxonRankValue, 44
temporalCoverage, 8, 19, 21
temporalCoverage,, 9
textDomain, 33
title, 11, 12, 25, 27, 43
toneGradation, 32
topolgyLevel, 32

uniqueKey, 41
unit, 34, 42
unitList, 42
url, 18, 30, 44

vericalAccuracy, 32
verticalAccuracy, 32
verticals, 32
view, 10, 28, 29, 32

タグ名, 8