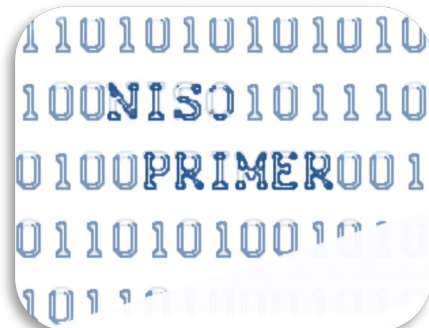


メタデータを 理解する

メタデータとは何か
なぜ必要か

ジェン・ライリー 著



米国情報標準化機構 (NISO) による入門書
訳：国立国会図書館電子情報部電子情報流通課

この『メタデータを理解する：メタデータとは何か、なぜ必要か』は、米国情報標準化機構（National Information Standards Organization：NISO）が発行した次の資料の全訳である。

Understanding Metadata: What is Metadata, and What is it for?
By Jenn Riley

National Information Standards Organization (NISO), 2017.
(NISO Primer Series)

ISBN: 978-1-937522-72-8

URL: <https://www.niso.org/publications/understanding-metadata-2017>

編集・翻訳 国立国会図書館電子情報部電子情報流通課
発行 国立国会図書館
〒100-8924 東京都千代田区永田町 1-10-1
TEL 03-3581-2331（代表）

令和2年3月17日 発行

URL <https://dl.ndl.go.jp/info:ndljp/pid/11471993>

DOI 10.11501/11471993

米国情報標準化機構 (National Information Standards Organization: NISO) 入門書シリーズについて

NISO 入門書シリーズは、データの取扱者を対象とした、全四部の入門ガイドです (他の三部は研究データを対象としています)。本入門書では、2017 年時点におけるメタデータの最新情報、新しいツール、ベストプラクティス及び現在利用できる情報資源について解説します。

本刊行物についての最新情報は、NISO の事務局に問い合わせるか、NISO のウェブサイト (www.niso.org) をご覧ください。

発行元

National Information Standards Organization (NISO)
3600 Clipper Mill Road
Suite 302
Baltimore, MD 21211
www.niso.org

本刊行物の著作権は NISO が保有しています。

© National Information Standards Organization (NISO), 2017.

本著作はクリエイティブ・コモンズ 表示-非営利 4.0 国際ライセンスの下に、その利用が NISO によりコミュニティに許諾されており、以下を自由に行うことができます。

- 共有一のようなメディアやフォーマットでも資料を複製したり、再配布できます。
- 翻案—資料をリミックスしたり、改変したり、別の作品のベースにしたりできます。

その際、以下の条件を満たす必要があります。

- 表示—あなたは適切なクレジットを表示し、ライセンスへのリンクを提供し、変更があったらその旨を示さなければなりません。あなたはこれらを合理的などのような方法で行っても構いませんが、許諾者があなたやあなたの利用行為を支持していると示唆するような方法は除きます。
- 非営利—あなたは営利目的でこの資料を利用してはなりません。



完全なライセンス条項は以下をご覧ください：
<http://creativecommons.org/licenses/by-nc/4.0/legalcode>

著作権は国際・全米著作権協定 (International and Pan-American Copyright Conventions) の下で保護されています。本刊行物は非営利目的においてのみ、正確に再現され、本文書の発行元が明示され、NISO の著作権の状態が確認されている場合に、発行者による書面での事前の許可なしに、いかなる形態あるいはいかなる方法でも複製及び送信することが可能です。他言語への翻訳、商用複製または再配布については、すべて以下にお問い合わせください。

NISO, 3600 Clipper Mill Road, Suite 302, Baltimore, MD 21211.
ISBN: 978-1-937522-72-8

Contents

はじめに.....	1
日常生活におけるメタデータ	2
文化遺産の世界におけるメタデータ	5
メタデータのタイプ	6
メタデータの格納と共有の方法	8
リレーショナルデータベース	8
eXtensible Markup Language (XML)	8
Linked Data と Resource Description Framework (RDF)	9
メタデータの標準化	16
統制語彙	17
コンテンツ標準	17
主要なメタデータ言語：広範囲に使用されるものの例	19
Schema.org	19
Web Ontology Language (OWL).....	21
Simple Knowledge Organization System (SKOS).....	22
ダブリンコア (DC)	23
Friend of a Friend (FOAF)	25
ONline Information eXchange (ONIX).....	26
EXchangeable Image File Format (Exif)	26
主要なメタデータ言語：文化遺産分野の例	27
MAchine Readable Cataloging (MARC).....	27
Bibliographic Framework Initiative (BIBFRAME).....	28
Metadata Object Description Schema (MODS).....	30
CIDOC Conceptual Reference Model (CIDOC CRM)	33
Categories for the Description of Works of Art (CDWA).....	33
Visual Resources Association Core (VRA Core).....	34
Encoded Archival Description (EAD)	34
主要なメタデータ言語：その他の例	35
Data Documentation Initiative (DDI).....	35
PREservation Metadata: Implementation Strategies (PREMIS)	35
Text Encoding Initiative (TEI).....	35
Music Encoding Initiative (MEI).....	36
メタデータはどのように生成されるのか	37
将来の方向性.....	39
付録：資料	41

はじめに

小売業者が商品や顧客の情報をどのように保管しているかをご存じですか。雇用者は従業員とその業務内容の情報、団体は主催する行事の情報、研究機関は各分野の動向と注目すべき人物の情報、図書館や文書館や博物館は保存管理している資料の情報、政府は国民や同盟国や敵対国の情報をどのように保管しているのでしょうか。これらの情報は、すべてメタデータです。メタデータは、事物を言い表すために私たちが作成、保管、共有する情報です。メタデータのおかげで、私たちは事物と関わり合い、必要な知識を得ることができます。「メタデータ」という言葉は古来、その語源が文字通りに表すように「データに関するデータ」という意味を持っています。このような定義の広さから、メタデータは遍在しているのではないかと考える人もいるかもしれませんが、事実そのとおりです。米国では、2013年に米国国家安全保障局（National Security Agency）が、国内の電話通信について発信位置や発信時刻、通話時間、連絡先番号等の情報を収集していたことがメディアで大々的に報じられて以来、「メタデータ」は多くの人にとって身近な言葉になりました。

日常生活におけるメタデータ

メタデータは情報システムのいたるところに様々な形式で存在します。私たちが普段使っているほとんどのソフトウェアパッケージの主要な機能は、メタデータによって実現しています。私たちは、**Spotify** で音楽を聴き、**Instagram** に写真を投稿し、**YouTube** で動画を探し、**Quicken** で資金管理をし、電子メールやテキストメッセージやソーシャルメディアで他の人々と繋がり、モバイル端末に長い連絡先リストを保存します。これらのすべてのコンテンツにメタデータ—すなわち対象となるコンテンツの作成に関する情報、名称、トピック、特徴等が付随しています。メタデータはコンテンツを保持するシステムがうまく機能するための要であり、メタデータのおかげで私たちは興味対象のコンテンツを検索し、それらに関する重要な情報を記録し、その情報を他の人々と共有することができるのです。

大抵の Web ページにはメタデータが埋め込まれています。ある Web ページから他のページへ張られているリンクや、検索結果の中からどのページを選んで閲覧したか等のユーザー行動の記録もメタデータに含まれます。Web サーチエンジンは、ページ内のテキストとそれに付随するメタデータを用いて膨大な索引を構築し、関連する検索結果をユーザーに提供します。Google はさらにその先を行っていて、2012 年には 35 億個の「ファクト」で構成される「**Knowledge Graph**」を発表しました。ファクトとは、約 5 億の人、場所、事物及びそれらの関係についての、メタデータです^{1,2}。Knowledge Graph やその他 Google が保管している構造化されたメタデータは、検索結果を向上させるとともに、付加価値の高い情報提供に使われています。例えば、スポーツの試合結果を表示したり、検索結果と地図の情報を連携させたり、注目すべき人や場所の詳細を検索結果画面に表示する Knowledge Card 機能等がこれにあたります。

無料で、誰でも自由に参加できるクラウドソーシングによるオンライン百科事典の Wikipedia は、メタデータの利用と生成の両方を行います。Wikimedia 財団の Wikidata プロジェクトは、Google の Knowledge Graph に類似した、オープンで共同編集型のナレッジベースです。ここには、特定のトピックに関するファクト情報が、構造化された形で格納され、Wikipedia の記事やその他の情報システムに追加することが可能になっています。DBpedia プロジェクトは、これとは逆に、Wikipedia のインフォボックス、カテゴリ、画像、地理空間情報及びリンクからメタデータを取り出し、様々な方法で再利用できるオープンな情報資源として、構造化されたメタデータを生成します。

メタデータは業務処理にも極めて重要です。小売業者は、取り扱っている商品の価格、入手元、在庫数量、説明情報等、多くの詳細事項を追跡する必要があります。これは、オンライン事業者にとっては一層不可欠です。オンラインショッピングのユーザーは商品を手にとって見ることができないため、キーワードや対象物の種類等の条件で検索したり、ファセットを利用して幅広い商品を把握しやすい数に絞り込む等といったことができることを期待しています。事業者は、検索や取引に関するメタデータを日常的に蓄積し、それに基づいて売上高の推移の分析、将来需要の予測、政府への売上税の支払い等を行います。また、同じメタデータを使用して、購入履歴、複数の出荷地の住所録、商品の推薦情報等の機能を付加し、よりパーソナライズされたショッピング体験をユーザーへ提供することが可能になります。製

¹ <https://googleblog.blogspot.ca/2012/05/introducing-knowledge-graph-things-not.html>

² <http://searchengineland.com/google-launches-knowledge-graph-121585>

造業者はメタデータを使用して、設計、部品及び材料を記録し、研究プログラムの管理を行います。旅行産業も同様に、乗客、得意客、予約に関するメタデータや、フライトやホテルの部屋等の情報資源に関するメタデータに依存しています。ニュースメディアはメタデータを使用して、出来事、報道及び公開コンテンツを追跡します。また、あらゆる事業者は、雇用、給与、業績管理等の人事業務にメタデータを使用しています。

活用されているメタデータ：Amazonとその提携企業

Amazon は多種多様な商品を取り扱い、世界的規模で事業を展開しているオンライン小売業者です。メタデータは Amazon の業務運用の推進に大きな役割を果たしています。Amazon が利用するメタデータは、書籍の出版者またはその他商品の供給業者自身が管理する在庫情報を源泉としています。これらの供給業者はメタデータを Amazon に送信します。Amazon は、このメタデータと、他の何千もの供給業者から送られてくる類似の情報とを統合して Amazon のウェブサイトを構築し、ユーザーに商品を販売します。Amazon は販売に関するメタデータを収集し、さらにそのデータを使用して顧客へお勧め商品の情報を提供し、供給業者との関係の最適化を行います。また、仲介する商品に関するメタデータを、Amazon のサービスに基づいて自らのサービスを構築するアフィリエイトサイトに対して利用可能とすることで、Amazon を通した売上高の拡大を図り、商品の供給業者の事業をも促進しています。

メタデータは、ソーシャルメディアプラットフォームの中核でもあります。Facebook ユーザーは、友達リストを管理したり、近況を投稿したり、写真や動画にコメントを投稿したり、友達の近況に「いいね」を押したり、投稿されたコンテンツを共有したり、自ら写真や動画を投稿する際にメタデータを作成しています。これらのアクティビティを追跡することにより、Facebook は流行りの話題を分析し、Facebook にとり利益を生み出すスポンサー付きの投稿のプロモーションを図ります。Pinterest のユーザーは、関心のある事柄を分類して説明を載せるボードを作成します。これらの分類と説明はメタデータの役割を持ち、事柄同士のつながりに価値が付加されます。Pinterest は、このソーシャルプラットフォームで作成されたメタデータを使用して検索用のインデックスと、ユーザーの興味対象に沿った推奨コンテンツを構築します。Instagram のユーザーは、アップロード及び共有する画像へキャプションを付与し、他のユーザーや企業のアカウントをフォローします。Instagram は、こうしたやり取りのデータを使用して、表示する広告を改善します。Twitter のユーザーは、フォローしている人をリストで整理し、テキストやメディアを投稿し、ハッシュタグを使用してツイートにコメントしたり他のツイートと関連づけたりし、他のユーザーのコンテンツにコメントを付け、あるいは付けずにリツイートしたり、「いいね」を押したりします。これらの動向は Twitter のトレンドリスト等に反映されます。2010 年には、Twitter 内のコンテンツに含まれる社会に関するデータとそのメタデータが重要であるため、米国議会図書館がこの価値ある資料を研究のためにアーカイブするという合意を結びました³。

これらの例は、メタデータと、メタデータが記述する情報との間の、やや曖昧な境界を示しています。両者の区別は、多くの場合において問題になりません。なぜなら、メタデータはあたかもそれがデータであるかのように作成・保管され、扱われるからです。実際、メタデータとデータとの区別は、現実には単に意味論上の違いに過ぎません。

前述の例における特徴の一つとして、すべてのメタデータがある程度構造化され

³ <https://blogs.loc.gov/loc/2010/04/how-tweet-it-is-library-acquires-entire-twitter-archive/>

ていることが挙げられます。メタデータは有益な目的を達成するために収集され、既知のカテゴリに分類されます。この構造の概念により、未加工の情報が実用的なメタデータに変化するのです。特定の要素が収集され、説明ラベルと連動して管理用または一般ユーザー向けインターフェースのいずれかに表示されるような方法で保存されます。このようなラベルは一般的に「プロパティ」や「要素」と呼ばれますが、それぞれの名前はユーザーコミュニティによって異なります。

文化遺産の世界におけるメタデータ

文化遺産の世界（図書館、文書館、博物館）では、堅牢で構造化されたメタデータの作成と共有が長年にわたって行われています。図書館では、これは蔵書目録の形で幾世紀も経て進化してきました。初期の蔵書目録は大部の管理台帳に過ぎませんでした。このような管理台帳は、その後、引き出しに収納できる大きさの目録カードに置き換えられました。コンピュータ化が進むと、図書館はまず検索専用の端末を導入し、インターネットの時代に入ると、現在のウェブベースの情報資源ディスカバリーシステムへ移行しました。図書館はメタデータに対して「書誌的」アプローチを採っていますが、これは、図書館が伝統的に図書を記述することに強みを持っていることに根差しています。書誌的メタデータでは、ユーザーが目的の資料を発見・識別できるよう、個別の資料（アイテム）について詳細に記述することに重点が置かれます。文書館では、「ファインディングエイド」を作成します。「ファインディングエイド」とは、コレクションについての記述的な目録で、資料の理解に必要な歴史的情報とともに提供されます。ここでは、ユーザーはメタデータを使用して、関連する一塊の資料グループを発見することができます。つまり、ある個人または組織の通常業務から生じる資料のグループ（前者は「書類 (papers)」、後者は「記録 (records)」と呼ばれる）、またそのグループ化において最も適切に解釈が可能になる資料を含む資料グループを発見することができます。博物館は、収集、展示及び貸出の履歴を詳細に記録します。博物館の学芸員はコレクションを解釈するための方法の一つとしてメタデータを使用し、来館者に作品の歴史的及び社会的な意義を伝えるとともに作品間の関係を説明します。

文化遺産のメタデータでは、記述情報が重要視されます。図書の場合は、印刷媒体であろうと電子媒体であろうと、タイトル、著者、出版事項及び主題の詳細が主要な情報となります。音楽、映像及び美術作品の場合は通常、タイトル、作者、ジャンル、上演等に関する情報が記録されます。文書館における書類や記録の場合は、それらの作成に関する詳細情報と資料間の相互関係が最も重要です。文化遺産機関では、作品の著作者とその生没年に関する情報もまた、メタデータとして記録されるのが一般的です。

メタデータのタイプ

前述したようなタイプのメタデータ、すなわち検索や理解に役立つ、情報資源の内容に関する情報は「記述メタデータ」と言われます。文化遺産の分野では、記述メタデータを他のタイプのメタデータと区別しています。「管理メタデータ」は、資料の管理に必要な情報や資料の作成に関する情報の総称です。管理メタデータには、デジタルファイルのデコードやレンダリングに必要な情報である「技術メタデータ」（例：ファイルタイプ）、デジタルファイルの長期管理と将来的なマイグレーションやエミュレーションをサポートする「保存メタデータ」（例：チェックサム、ハッシュ）、コンテンツに付随する知的財産権を詳述する「権利メタデータ」（例：クリエイティブ・コモンズ・ライセンス）が含まれます。さらに、記述メタデータとも管理メタデータとも異なるものとして、「構造メタデータ」があり、これは情報資源の各部の相互関係を記述しています。構造メタデータの例としては、一連の順序付けされたページ、マイルストーンとなる章や節の先頭にポイントが合うよう設定された目次、異なる解像度やビット深度により表現された同一コンテンツを関連付けること等が挙げられます。

メタデータのタイプ

記述メタデータ	情報資源の発見や理解を目的とするメタデータ
管理メタデータ	<ul style="list-style-type: none"> - 技術メタデータ：ファイルのデコードやレンダリングを目的とするメタデータ - 保存メタデータ：ファイルの長期的な管理を目的とするメタデータ - 権利メタデータ：コンテンツに付随する知的財産権に関するメタデータ
構造メタデータ	情報資源の各部の相互関係に関するメタデータ
マークアップ言語	メタデータとコンテンツ内の他の構造的または意味的な特徴に対するタグを統合するもの

メタデータの最後のカテゴリは「マークアップ言語」です。マークアップ言語はメタデータとコンテンツを共に扱いますが、これは他のタイプのメタデータではあまり行われません。コンテンツに挿入されたタグは特筆すべき事項を示します。それは、情報資源がテキストの場合には、パラグラフ等の構造的要素をマークアップすること、単語に対して地名であるとか演説の特定の部分であるといったセマンティックな情報を与えること、あるいはイタリック体等の形式情報を与えること等を意味します。

これら様々なタイプのメタデータが、情報システムにおける多様なユースケースをサポートしています。情報の発見はおそらく最も一般的なユースケースでしょう。構造化されたメタデータを使って、ユーザーは興味のある情報資源を検索したり閲覧したりできます。メタデータによって表される属性の多くは、ユーザーに情報を表示し、情報資源を識別または理解する助けになります。異なるシステム間でコンテンツを効果的にやり取りし、相互運用性を確保するには、コンテンツを記述するメタデータに依存しており、関係するシステムが入力されたデータのプロファイリングと各システムの内部構造へのマッチングを有効に行えるようにコンテンツが記述されている必要があります。メタデータは、デジタルオブジェクトの適切なレンダリング、またはユーザーのニーズに適合するバージョンの配信に必要な情報を提供することにより、デジタルオブジェクトの管理をサポートしています。デジタル

オブジェクトの保存のためには、データ移送等の重要なポイントにおいてコンテンツの整合性を検証できるメタデータを作成すること、また、例えばフォーマット変換（マイグレーション）や整合性チェック等保存のための作業を実施する必要がある場合にシグナルを出すことが必要です。さらに、メタデータは、あるページやセクションから次のページやセクション等アイテム内の移動や、解像度が異なる写真画像等バージョンが異なるオブジェクト間のナビゲーションを可能にしています。

メタデータタイプ	プロパティの例	主要な用途
記述メタデータ	タイトル 著者 主題 ジャンル 発行日	検索 表示 相互運用
技術メタデータ	ファイルタイプ ファイルサイズ 作成日時 圧縮方式	相互運用 デジタルオブジェクト管理 保存
保存メタデータ	チェックサム 保存イベント	相互運用 デジタルオブジェクト管理 保存
権利メタデータ	著作権の状態 ライセンス条項 権利者	相互運用 デジタルオブジェクト管理
構造メタデータ	シーケンス 階層内の位置	ナビゲーション
マークアップ言語	パラグラフ 見出し リスト 名前 日付	ナビゲーション 相互運用

メタデータの格納と共有の方法

リレーショナルデータベース

メタデータの形式と符号化方式は多岐にわたります。従来の情報システム的设计では、メタデータは、リレーショナルデータベースのテーブルにフィールドとして格納されてきました。この場合のメタデータの 1 セットは「レコード」と呼ばれます。このモデルの設計は、クエリパフォーマンスの最適化と調和させてストレージ効率を最大化したデータベーステーブルを適切に正規化することで効果的なものにできます。ほとんどのタイプのメタデータはこの方法で格納可能です。この場合には、メタデータはカスタムプロセスによるバッチ処理で読み込まれるか、専用のユーザーインターフェース上で手入力され、いずれの場合もカスタムプログラミングで制御されるのが一般的です。今日ではこのメタデータモデルを使用するソフトウェアシステムが、他者とメタデータを共有する場合、アプリケーションプログラミングインターフェース (API) を使用するのが一般的です。その際、外部のソフトウェア開発者が、システムに対してクエリを実行し、任意のメタデータを取得するためのツールを構築できるように、仕様書を公開しています。

eXtensible Markup Language (XML)

2000 年代に現れた eXtensible Markup Language (XML) は、メタデータの符号化、転送によく使われるほか、システム内部にメタデータを保管するメカニズムに使われることもあります。XML 形式のメタデータは、XML ドキュメントと呼ばれるファイルのセットとして存在します。XML は、内部の値がある特定の意味を持つことを表す要素を定義します。要素は、その内部に他の要素を含むこともあり、そのため XML ドキュメントは構造を持っています。XML ドキュメントは、単一のルート要素から始まるツリーです。この元のルートから他の要素や値が分岐し、ドキュメント内のメタデータの値を意味のあるものにする入れ子構造を構築します。XML 要素は属性を持つことができますが、各属性もまたそれぞれの値を持っています。XML の属性とその値により、それらが含まれる要素の意味が絞り込まれます。XML は、要素の値が表されている言語を示すために前もって定義された属性を指定することで、メタデータの多言語対応をサポートします。リレーショナルデータベースと同様に、ある特定の事物を記述している一つの XML ドキュメントは、メタデータレコードと呼ばれます。そのようなレコードの例を以下に示します。

例：シンプルな XML レコード

```
<?xml version="1.0" encoding="UTF-8"?>
<work type="play">
  <workName>The Tempest</workName>
  <writtenBy>
    <playwright>
      <playwrightName>William Shakespeare</playwrightName>
      <bornInPlace>Stratford Upon Avon</bornInPlace>
    </playwright>
  </writtenBy>
</work>
```

<work>、<playwright>及び<bornInPlace>は XML 要素の例。
「Stratford Upon Avon」は要素の値の例。

「type="play"」は属性の名称(type)と値(play)の例。

XML ドキュメントの変換やクエリに特化したプログラミング言語（よく知られているのは、XPath、XSLT、XQuery）が存在します。また、主要なプログラミング言語のための XML 処理ツールキットもあります。効果的な XML 設計には、十分なクエリパフォーマンスが得られるように、要素と属性をバランスよく使用し、ドキュメントのサイズに気を配ることが肝要です。システムに XML として格納されたメタデータは、外部システムから直接読み込まれたものであることが多いですが、ソフトウェアのユーザーインターフェースを通して生成されたり、他のデータソースから一括してマッピングされたりすることもあります。XML データは、システムに取り込まれた後に格納や索引作成のために他の形式に変換される場合が多いですが、XML に基づくデータベースも存在します。XML の用途は記述メタデータに限定されず、他の多くのタイプのメタデータを XML ドキュメントに格納することが可能です。

文化遺産機関は長年にわたりメタデータを共有してきており、その始まりは米国議会図書館が公共図書館へ目録カード（主に図書用）を配付していたことに遡ります。2000 年代初頭、文化遺産コミュニティは Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) を通して XML ベースのメタデータの共有を開始し、新たな連携の局面を迎えました。OAI-PMH が使用されるようになって、図書館、文書館及び博物館がメタデータを定期的に共有する対象となる資料の種類は大幅に増え、写真コレクションや大学で作成された研究論文のプレプリント（査読前原稿）等の資料をも含むようになりました。一時は、Google がその Sitemaps プロトコルの一部として OAI-PMH の使用をサポートして促進させていましたが、このサポートは 2008 年に停止されました。OAI-PMH は、DSpace 等の一般的なソフトウェアパッケージに実装されているため、機関レポジトリやデジタルコレクションのコミュニティで現在でも使用されていますが、その限界もよく知られています。XML Sitemaps の仕様で動作する後継プロトコルの ResourceSync が、ある程度勢いを得ていますが、近年では、今後のメタデータ共有方法として、Linked Data が非常に有望なものとなっています。

Linked Data と Resource Description Framework (RDF)

Linked Data の概念は、World Wide Web の発明者として知られるティム・バーナーズ=リー氏により、2006 年に提唱されました。この概念を実装するには、各機関は、構造化されたデータをウェブ上で公開する際に、そのデータに含まれる実体を明示的に名付けて他者による参照を可能にし、一方で自身のデータを他者のデータとリンクすることで、世界規模の情報ネットワークの構築を目指す必要があります。この取組みは、機械処理可能なデータの世界的ネットワークのビジョンである「セマンティックウェブ」の実装に向けた最も効果的で現実的な手段となりました。近年、World Wide Web Consortium (W3C) は、セマンティックウェブの当初のビジョンを拡張した「データのウェブの構築」においてリーダー的役割を担っており、Linked Data はその計画の重要な構成要素です。

Linked Data — 設計上の課題

ティム・バーナーズ=リー 2006 年

<http://www.w3.org/DesignIssues/LinkedData.html>

1. 事柄の名前として URI を用いること。
2. これらの名前を参照できるように、HTTP URI を用いること。

3. URIを参照したときに、RDFやSPARQLのような標準技術を用いて、有用な情報を提供できるようにすること。
4. さらに多くの事物を発見できるように、他のURIへのリンクを含むこと。

現行の Linked Data は Resource Description Framework (RDF) 標準に大きく依存しています。RDF はセマンティックウェブ上のメタデータ用に設計された、W3C の一連の仕様です。XML においては情報がツリー状にモデル化されるのに対し、RDF では細分化されたそれぞれの情報が、他の細分化された情報と関連付けられたグラフの形でモデル化され、グラフ内のすべての実体やデータの重要性に差はなく、情報のネットワークは、どのポイントからも平等にアクセスできます。そのため、リレーショナルデータベースや XML で使用されるようなメタデータレコードという概念、すなわち単一の実体についての情報の集合や、一体として扱われるよう定義されたデータ要素の集合といった概念は、RDF モデルにはうまく適合しません。RDF グラフは全体として見るか、ある特定の目的のための特定のコンテキストで使用される簡略化されたサブセットとして見るのが適しています。また、グラフは主語、述語及び目的語の三要素から成るトリプルの集合で構成されます。トリプルでは、主語（トリプルの主題となる実体）が目的語（主語に関係する実体）に、述語（関係の記述子）によって結び付けられていて、例えば次のように表現されます：「テンペスト」（主語）は「ウィリアム・シェイクスピア」（目的語）「によって書かれた」（述語）。

RDF では、すべての主語といくつかの目的語は情報資源のタイプを表す「クラス」としてモデル化されています。RDF の表記規則には、クラスの名前はそれぞれ大文字で始めること、各クラスのラベルに含まれる二番目以降の単語も大文字で始めること、単語間にスペースは入れないことと定められています。RDF のクラスは、例えば Person、Book、Painting、Building、Event、PhilosophicalIdea 等のように、無限に存在し得ます。RDF の述語はプロパティとしてモデル化され、関係を記述します。実際のところどのような関係の概念でも RDF プロパティとして表現できますが、記述メタデータを例示するような関係が最も一般的です。プロパティの名前は小文字で始まります。各プロパティのラベルに含まれる二番目以降の単語は大文字で始まり、単語間にスペースは入れません。例を次に示します: `createdBy`、`memberOf`、`successorTo`、`occurredAtTime`、`sameAs`、`familyName`

RDF のプロパティは定義域を定義することができ、それによってトリプルの主語が指定されたクラスのメンバーであることが示されます。同様に、プロパティの値域を定義することにより、トリプルの目的語が、指定されたクラスのメンバーであることを示します。定義域と値域には二つの目的があります。一つ目は、ある特定のプロパティがどのように使われるべきかを実装者に示すこと、そして二つ目は、処理ツールによって、これらの定義が非明示的に意味する繋がりから新しい RDF の関係を導き出せるようにすることです。例えば、`createdBy` プロパティについて、定義域が `Book`、値域が `Person` として定義されていた場合は、たとえそのような宣言を明示的に行う RDF が見当たらなくても、システムはこのトリプルについて、主語が `Book` というクラスで、目的語が `Person` というクラスに属するものであると推定することができます。

RDF 内で、主語、述語及び場合によっては目的語も統一資源識別子 (URI: Uniform Resource Identifiers) または International Resource Identifiers (IRI。非 ASCII 文字の使用を許可する URI のこと。) で表されます。このようにクラスとプロパティが構造化されて命名されることによって、他のトリプルからの参照が可能となります。トリプルを付加することで、情報資源に関する付加的な事実を記述することができます。例えば、ある一つのトリプルで著者が割り当てられたこの図書は、この年書か

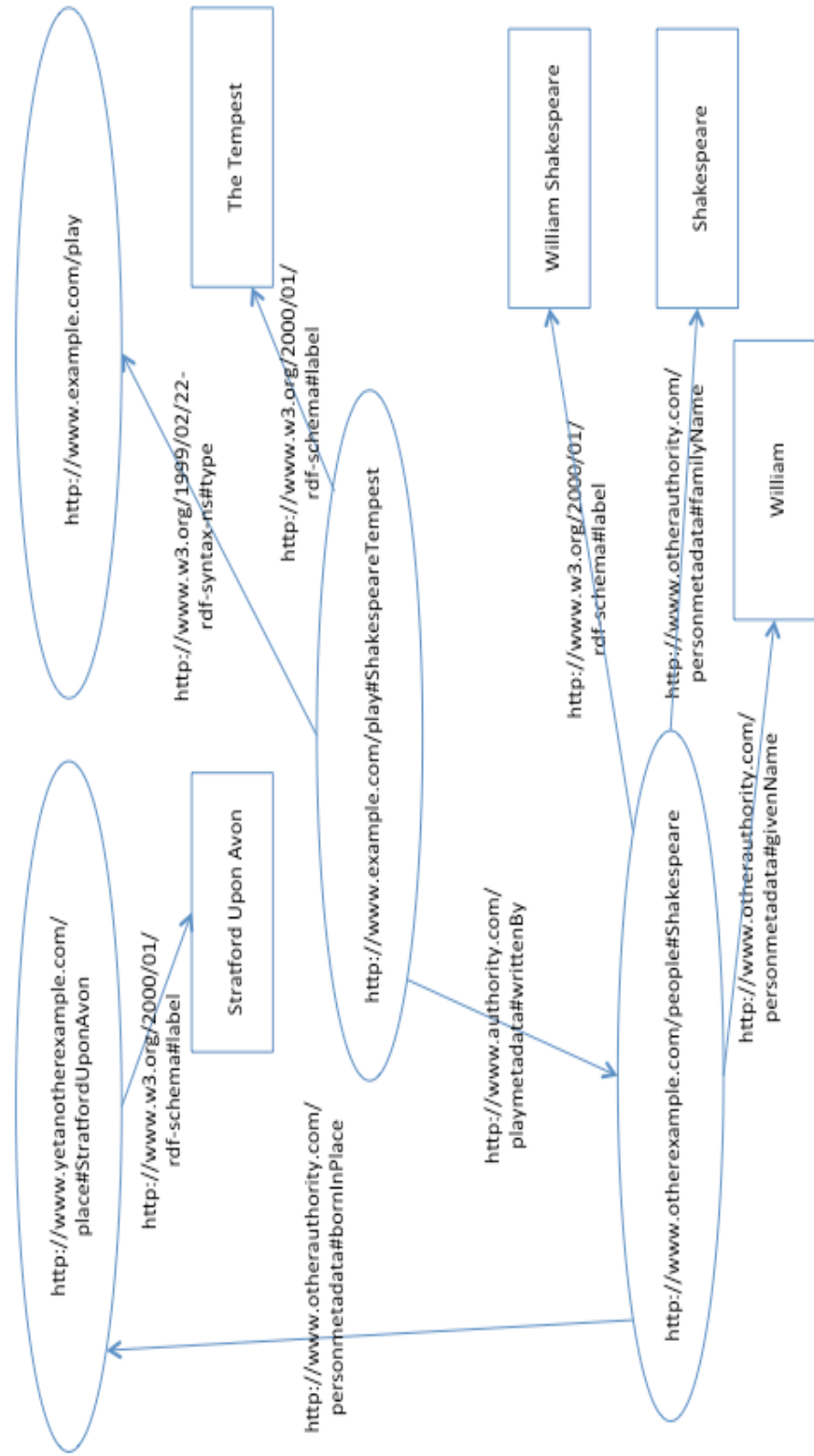
れた、あるいは「フィクション」に分類される、といった記述等です。この付加的な記述ができる構造は、クラスやプロパティを他のクラスやプロパティに関連付ける場合にも使用されます。例えば、「Book というクラスは CreativeWork クラスのサブクラスである」あるいは「ある機関や団体により作成された、作品の著作者の概念を定義するプロパティは、他の機関や団体により作成された同じ目的の他のプロパティと意味的に同一である」等です。RDF では、一つのトリプル内のすべての主語、述語及び目的語が、同一のコミュニティにより定義される必要はありません。Linked Data の力は実に、出自の異なるデータや語彙を結びつけることにあります。このような特徴のため、RDF グラフは世界中からのインプットを得て成長しています。目的語は、自由記述の文字列であってもよく、これは RDF では「リテラル」と呼ばれます。RDF 標準では、リテラルが、ある特定の言語であることや予め定義されたデータタイプに一致していることを明示することができます。

RDF スキーマ (RDFS) は、RDF 言語の作成に使用される主要な RDF 技術です。RDFS は、クラスとプロパティ、目的語のデータタイプ、プロパティの定義域と値域及びクラスとサブクラス間またはプロパティとサブプロパティ間の階層関係を定式的に定義するために使用されます。

Linked Data のベストプラクティスとは、URI が「参照解決可能」であることです。すなわち、RDF で符号化された概念に HTTP URI でアクセスする場合に、人間と機械の両方が、ラベル、定義及び他の情報資源との関係等の有用な情報を参照できるよう、URI が HTTP 経由で機械処理可能であるべきだということです。コンテンツネゴシエーションのプロセスは、同一の URI にアクセスした際、人間のユーザーにはウェブページを、ソフトウェアアプリケーションには生データを提供するために使用されます。これにより、RDF を認識するソフトウェアアプリケーションは、未知のクラスやプロパティを使用できるようになり、RDF で符号化された Linked Data が、複数の情報源からの情報をつなぐための強力なツールとなります。

RDF グラフの例

A Simple Graph for Shakespeare's *The Tempest*



RDF データは様々な方法で共有可能です。特に、`itemprop` 属性を使用して HTML や XHTML の内部に構造化されたメタデータを埋め込むマイクロフォーマットの手法と相性が良く、しばしば一緒に使用されています。それにより、ウェブサイトに含まれるデータを処理するサーチエンジン等のツールも RDF データを利用できるようになります。そのほか、大規模な RDF トリプルストア (RDF データの保存と取り出し用に最適化されたソフトウェア) がエンドポイントを提供し、RDF クエリ言語の SPARQL を使用して API 経由で、リモート検索を可能にしている場合も多く見られます。そうすることで、これらのトリプルストアからのデータを、第三者のアプリケーションが使用できます。

様々な RDF 構文が、Linked Data として広く利用されています。RDF/XML は、結び付けられた RDF グラフを XML の構造を使用してシリアル化したものです。RDF/XML は、RDF の符号化として特に冗長であり、人間にとっては可読性が低いです。RDF/XML はシリアル化された他の RDF よりも古くからありますが、ソフトウェア開発者の間ではあまり人気がありません。Resource Description Framework in Attributes (RDFa) は、マイクロフォーマットと同様に、HTML 属性を使用して、HTML 内に RDF を埋め込む方法です。ここで用いられる HTML 属性は、ウェブブラウザによるページのレンダリングには影響を与えず、ページを読み込むシステムにコンテンツに関するメタデータを提供する役割をします。Terse RDF Triple Language (Turtle) は、非常にコンパクトな RDF シリアル化で、SPARQL クエリ言語の構造の一部を再利用します。Turtle では、RDF グラフをテキストとして表現し、同一の主語が持つ複数の述語を簡潔に表現することができます。N-Triples は、もうひとつ別のテキストベースの RDF シリアル化で、これは Turtle のサブセットであり、より構造的で予測可能な構文に限定されます。JavaScript Object Notation for Linked Data (JSON-LD) は、ソフトウェア開発者が利用しやすい JSON フォーマットをベースとした RDF シリアル化であり、簡単なキー値のペアを使用して情報を記録します。これは一般的に、クエリに応じてデータの短いスニペットを返す簡単なウェブサービスで使用されます。

RDF グラフ: RDF/XML によるシリアル化の例

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:ex="http://www.example.com/"
  xmlns:play="http://www.authority.com/playmetadata#"
  xmlns:person="http://www.otherauthority.com/personmetadata#">
  <ex:play
    rdf:about="http://www.example.com/play#ShakespeareTempest">
    <rdfs:label>The Tempest</rdfs:label>
    <play:writtenBy>
      <rdf:Description
        rdf:about="http://www.otherexample.com/people#Shakespeare">
          <rdfs:label>William Shakespeare</rdfs:label>
          <person:givenName>William</person:givenName>
          <person:familyName>Shakespeare</person:familyName>
          <person:bornInPlace>
            <rdf:Description
              rdf:about="http://www.yetanotherexample.com/place#StratfordUponAvon">
```

```
<rdfs:label>Stratford Upon Avon</rdfs:label>
</rdf:Description>
</person:bornInPlace>
</rdf:Description>
</play:writtenBy>
</ex:play>
</rdf:RDF>
```

RDF グラフ: Turtle によるシリアル化の例

```
@prefix person:
<http://www.otherauthority.com/personmetadata#> .
@prefix play: <http://www.authority.com/playmetadata#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .

<http://www.example.com/play#ShakespeareTempest> a
<http://www.example.com/play> ;
  rdfs:label "The Tempest" ;
  play:writtenBy
<http://www.otherexample.com/people#Shakespeare> .

<http://www.otherexample.com/people#Shakespeare> rdfs:label
"William Shakespeare" ;
  person:bornInPlace
<http://www.yetanotherexample.com/place#StratfordUponAvon> ;
  person:familyName "Shakespeare" ;
  person:givenName "William" .

<http://www.yetanotherexample.com/place#StratfordUponAvon>
rdfs:label "Stratford Upon Avon" .
```

RDF グラフ - N-Triples によるシリアル化の例

```
<http://www.example.com/play#ShakespeareTempest>
<http://www.w3.org/2000/01/rdf-schema#label> "The Tempest" .
<http://www.example.com/play#ShakespeareTempest>
<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
<http://www.example.com/play> .
<http://www.otherexample.com/people#Shakespeare>
<http://www.otherauthority.com/personmetadata#familyName>
"Shakespeare" .
<http://www.otherexample.com/people#Shakespeare>
<http://www.otherauthority.com/personmetadata#bornInPlace>
<http://www.yetanotherexample.com/place#StratfordUponAvon> .
<http://www.yetanotherexample.com/place#StratfordUponAvon>
<http://www.w3.org/2000/01/rdf-schema#label> "Stratford Upon
Avon" .
<http://www.otherexample.com/people#Shakespeare>
<http://www.otherauthority.com/personmetadata#givenName>
"William" .
<http://www.otherexample.com/people#Shakespeare>
<http://www.w3.org/2000/01/rdf-schema#label> "William
Shakespeare" .
```

```
<http://www.example.com/play#ShakespeareTempest>
<http://www.authority.com/playmetadata#writtenBy>
<http://www.otherexample.com/people#Shakespeare> .
```

RDF グラフ – JSON-LD によるシリアル化の例

```
{
  "@context": {
    "person": "http://www.otherauthority.com/personmetadata#",
    "play": "http://www.authority.com/playmetadata#",
    "rdfs": "http://www.w3.org/2000/01/rdf-schema#",
  },
  "@graph": [
    {
      "@id":
"http://www.yetanotherexample.com/place#StratfordUponAvon",
      "rdfs:label": "Stratford Upon Avon"
    },
    {
      "@id": "http://www.otherexample.com/people#Shakespeare",
      "person:bornInPlace": {
        "@id":
"http://www.yetanotherexample.com/place#StratfordUponAvon"
      },
      "person:familyName": "Shakespeare",
      "person:givenName": "William",
      "rdfs:label": "William Shakespeare"
    },
    {
      "@id": "http://www.example.com/play#ShakespeareTempest",
      "@type": "http://www.example.com/play",
      "play:writtenBy": {
        "@id":
"http://www.otherexample.com/people#Shakespeare"
      },
      "rdfs:label": "The Tempest"
    }
  ]
}
```

メタデータの格納と共有を行うその他の方法として、データをデジタルファイル自体に埋め込む方法があります。ほとんどすべてのファイル形式の仕様にはメタデータに関する部分が含まれています。それはそのファイルに関する技術メタデータであることが多く、デコードとレンダリングに使用されます。デジタルカメラ内のソフトウェア等、メタデータを含むファイルを作成するソフトウェアは、ファイルの内部にメタデータを生成して埋め込みます。さらに、JPEG、WAV、PDF、Microsoft Office ファイル等、多くのよく知られている画像、メディア及びドキュメントのファイル形式は、記述メタデータの記録を行います。これらのファイルの作成や編集を行うことができるツールでは、付随する記述メタデータを編集することも可能です。埋め込まれたメタデータは、インデックス作成のために外部システムへの抽出が必要ですが、メタデータをそれが記述するファイル内に保持すると、新しい環境の中でもそのファイルを確実に理解できるという利点があります。

メタデータの標準化

メタデータは、ソフトウェアアプリケーションや、メタデータを使用する人に理解できなければ役に立ちません。理解に役立つように、各機関は特定のニーズを満たすためのメタデータセットを予め定義し、システム設計者（エンドユーザーの場合もある）が参照できるように公開するのが一般的です。XML メタデータで用いられる語彙は、「スキーマ」、「要素セット」または「フォーマット」とも呼ばれます。XML スキーマはそのフォーマットにおいて有効なドキュメントを構成する要素を定義するとともに、各要素がとり得る属性、出現順及びそれらの出現回数を定義します。XML スキーマは、国際標準化機構（ISO）、NISO、または W3C 等の機関を通じて正式に標準化されます。米国議会図書館のような、業界、あるいはコミュニティを先導する機関は、XML ベースのメタデータ標準の策定機関あるいは管理機関としての役割を担うことも多く、それぞれの対象コミュニティにおいて標準が使用されるよう支援しています。XML 文書型定義（DTD）は、XML スキーマよりも古い技術であり、現在その使用は限定的です。

RDF メタデータの仕様のドキュメンテーションと共有をとりまく状況は、XML メタデータの場合と幾分異なっています。RDF 言語は一般的に「語彙」と呼ばれ、クラスとプロパティを定義します。RDF 語彙について正式な標準化が行われることはほとんどありません。コミュニティごとに有用な語彙を構築することが多く、ドキュメンテーションやそれらの語彙を使用するデータのオープンな共有といった取組みを通じて、構築した語彙の使用を促進します。

XML と RDF のいずれも「名前空間」の概念を使用し、要素、属性、クラスまたはプロパティに用いられている所定の語が、いずれの語彙に由来するのかわを示しています。XML と RDF の両方において、名前空間の接頭辞は、メタデータの構文を簡素化するために URI または IRI の代わりに使用されます。メタデータを完全に処理するには、名前空間の接頭辞が完全な URI または IRI に拡張され、使用されている特定の要素、クラス、またはプロパティに追加される必要があります。

RDF における名前空間の展開

名前空間接頭辞 rdf:は...	...以下の URI の省略表現である。
rdf:	http://www.w3.org/1999/02/22-rdf-syntax-ns#
そのため、rdf:に続く語は...	...そのデータの処理時に、上記 URI の後に追加されて処理される。
例えば、rdf:type は...	http://www.w3.org/1999/02/22-rdf-syntax-ns#type

XML では、一つのスキーマは、そのスキーマが定義する要素や属性のための「ホームベース」となるデフォルトの名前空間を持ちます。XML 文書内の明示的に宣言された名前空間を持たない要素はすべて、このデフォルトの名前空間にあるとされます。他の名前空間からの要素や属性は必要に応じて XML スキーマに正式に取り込むことができるため、そのスキーマにより定義される XML 言語の一部となります。XML 文書には他の名前空間からの任意の要素を使用できるメカニズムが備わっていますが、プロセッサが未知の語彙の意味を理解するためのメカニズムが XML 技術にはほとんど備わっていないため、この機能はあまり使用されません。

RDF では、名前空間はアーキテクチャを構成する基本要素の一つです。XML ではデフォルトの名前空間が仮定され、他の名前空間からの要素が例外であるのに対し、RDF においては、どの名前空間からのクラスまたはプロパティも任意のポイントで

使用できるということが前提です。いずれの名前空間も優先されることはありません。実際、あるトリプルの主語、述語及び目的語が、別々の名前空間からの URI になることは珍しいことではありません。まさに、このような RDF の設計こそが、グラフ内における語彙の混合と対応付けを促進しているのであり、Linked Data の目標は可能な限り異なる情報源からの可能な限り多くのデータポイントを連結させることです。主語、目的語または述語に関する付加的な情報を発見するために URI または IRI が参照解決されているベストプラクティスにおいては、それまでに知られていなかった名前空間でソフトウェアアプリケーションがデータの処理を行うのに役立つように設計されています。

統制語彙

メタデータの設計においては、構文の標準化に加え、実際に使用される値を統制することによりメタデータを標準化することもしばしば企図されます。これを実現する方法の一つは、統制語彙を使用することです。統制語彙とは、ある特定のトピックまたはタイプに関する、あらかじめ選定された一連の語のリストを指します。統制語彙では、ある概念に対して最も適切な語やフレーズが通常一つ指定され、同一概念の他の語が、この指定された語にマッピングされることもあります。また、語彙間の関係が定義されることも多く、その関係はしばしば階層型となっています。

統制語彙が統制の範囲とするのは通常単一の言語ですが、大規模なコミュニティでは異なる言語における既存の統制語彙をリンクしたり、異なる言語から語彙を集めて、新しい統制語彙を定義する場合があります。統制語彙には、十数個の語彙のみを使用する非常に簡単なものもあれば、何万語も含むはるかに堅牢なものもあります。統制語彙の例として、インターネットの MIME タイプ、Spotify のジャンル、Book Industry Standards and Communications (BISAC) の語彙、米国議会図書館件名標目表 (LCSH) 等が挙げられます。LCSH 等、文化遺産コミュニティが開発及び維持を手掛ける統制語彙は、一般的に使用されているものの中で最も堅牢な統制語彙であると考えられます。

XML の実装では、統制語彙から選択される語はある要素の値となり、この場合には、選択された語が由来する統制語彙を示す属性が示されるのが普通です。RDF では、この統制された語は、トリプルの目的語として、テキスト文字列ではなく、URI または IRI を用いてメタデータ内で参照されます。そして、この URI または IRI は、参照解決され、被参照語、その語を含む語彙及び他の語との関係に関するさらに詳しい情報を提供します。

コンテンツ標準

メタデータに出現する値を標準化する二つ目の方法は、コンテンツ標準を使用することです。コンテンツ標準は、メタデータ内のテキスト値がどのように構造化されるべきかを記述したガイドラインのセットです。このような標準は、文化遺産コミュニティにおいては公式のガイドライン文書として整備されるのが一般的です。他のコミュニティでは、これらはスタイルガイドとして知られており、より短くまとめられ、非公式である場合が多いです。

コンテンツ標準は一般的に、採録対象の情報が記載されている場所、句読点、大文字と小文字の区別及び略記法が使われるべきか、採録対象の情報を決定する方法等を定めています。小規模の統制語彙の使用に関する定義と記述が含まれることもあります。コンテンツ標準の例としては、Wikipedia Manual of Style guidelines for Infoboxes、アーカイブズ資料のための Describing Archives: A Content Standard (DACS)

及び Rules for Archival Description (RAD)、また図書館コミュニティの Anglo-American Cataloging Rules, second edition (AACR2) とその後継の Resource Description and Access (RDA) 等が挙げられます。XML ベースのメタデータの場合、コンテンツ標準は XML 要素に入力される値を制御します。RDF では、コンテンツ標準はトリプルの目的語のリテラル値に適用されます。

主要なメタデータ言語：広範囲に使用されるものの例

Schema.org

オープンなウェブでは、Schema.org は非常によく知られたメタデータ語彙の一つです。Schema.org は 2011 年に主要なサーチエンジンにより立ち上げられた RDF 語彙であり、ウェブクリエイターはこれを使用してページ内のテキストをセマンティックにマークアップすることができ、各種システムがコンテンツを用いて何か面白いことができるよう促します。この語彙はコミュニティ・ガバナンスを通じて管理されています。2016年時点で、1千万件以上のウェブサイトでこの語彙が使用されていると Schema.org のホームページに記載されています。

Schema.org は、約 600 個の「タイプ」（RDF のクラスとして定義されている）及び 800 個以上のプロパティを定義しています。また、特定分野のコミュニティが使用する語も語彙に含まれるよう、拡張語彙の使用を促進しています。最も使用される Schema.org のタイプは、クリエイティブな作品、埋め込みオブジェクト、イベント、組織、人、場所及び事業者、製品と取引、レビューと評価、そして活動に関連するデータのためのタイプです。Schema.org は、主に記述メタデータに使用されます。各上位カテゴリの下に、階層的に詳細化されたサブカテゴリが連なります。以下に例を示します。

CreativeWork->Article->ScholarlyArticle

または

Place->CivicStructure->GovernmentBuilding->CityHall.

Schema.org の各クラスは、多くの定義済みプロパティを持ちます。例えば、「Article」クラスは「wordCount」プロパティと関連付けられ、「CivicStructure」クラスには、「openingHours」プロパティが使用されます。

Schema.org の使用例: ノースカロライナ州立図書館のホームページにおいて HTML に埋め込まれたマイクロフォーマットの構文

```
<div id="f-address">
  <h2>Contact</h2>
  <h3 itemscope itemtype="http://schema.org/Library"
itemprop="name"><link itemprop="logo" h
ref="//www.lib.ncsu.edu/website/images/logo_small.png"><a
href="//www.lib.ncsu.edu/" id="f-contact-hill"
itemprop="url">D. H. Hill Library</a></h3>
  <p>
    <span itemprop="address" itemscope
itemtype="http://schema.org/PostalAddress">
      <span itemprop="streetAddress">2 Broughton
Drive</span> <br>
      <span itemprop="postOfficeBoxNumber">Campus
Box 7111</span> <br>
      <span
itemprop="addressLocality">Raleigh</span>,>
```

```

        <span itemprop="addressRegion">NC</span>
        <span itemprop="postalCode">27695-7111</span>
<br>
        </span>
        <span itemprop="telephone"><a href="tel:9195153364"
id="f-hill-phone" data-attr="phone">(919) 515-3364</a></span>
    </p>
</div>

```

Schema.org の使用例: Online Computer Library Center (OCLC) WorldCat Linked Data から Nate Silver の著作“The Signal and the Noise”に関するデータ (一部) の Turtle 構文

```

@prefix schema: <http://schema.org/> .
@prefix void: <http://rdfs.org/ns/void#> .

<http://experiment.worldcat.org/entity/work/data/1172581839#Person/silver_nate_1978>
    a
        schema:Person ;
    schema:birthDate "1978" ;
    schema:name "Silver, Nate, 1978-" .

<http://worldcat.org/entity/work/id/1172581839>
    a
        schema:CreativeWork ,
schema:Book ;
    void:inDataset
<http://purl.oclc.org/dataset/xwc> ;
    schema:about
<http://experiment.worldcat.org/entity/work/data/1172581839#Topic/knowledge_theory_of> ,
<http://experiment.worldcat.org/entity/work/data/1172581839#Topic/bayesian_statistical_decision_theory> ,
<http://experiment.worldcat.org/entity/work/data/1172581839#Topic/forecasting_history> ,
<http://id.loc.gov/authorities/subjects/sh85050485> ,
<http://id.loc.gov/authorities/subjects/sh85072732> ,
<http://id.worldcat.org/fast/988194> ;
    schema:creator
<http://viaf.org/viaf/256089470> ,
<http://experiment.worldcat.org/entity/work/data/1172581839#Person/silver_nate_1978> ,
<http://experiment.worldcat.org/entity/work/data/1172581839#Person/silver_nate> ,
<http://id.loc.gov/authorities/names/n2012043409> ;
    schema:description "Silver built an innovative
system for predicting baseball performance, predicted the 2008
election within a hair's breadth, and became a national
sensation as a blogger. Drawing on his own groundbreaking
work, Silver examines the world of prediction."@en ;
    schema:genre "History" , "Nonfiction" ;
    schema:name "The signal and the noise : why
so many predictions fail-- but some don't"@en ;

```

Schema.org がウェブページ内のテキストのセマンティクスに重点を置くことで、

インターネット上のメタデータはより高度なレベルで処理できるようになりました。インターネットの黎明期には、HTML マークアップによって、ページのタイトル、著作者及び主題に関する基本情報がサーチエンジンに提供され、ユーザーのクエリに応じてページを取得することが可能になっていました。近年では、Google の Knowledge Graph のイニシアチブからも明らかのように、サーチエンジンは単にページの索引付けだけではなく、知識の蓄積に重点を置いています。Schema.org はウェブページ内の小さくても極めて重要な知識の符号化を推進することで、このユースケースを促進させるために設計されています。例えば、Schema.org の記述には、ある特定の建物がある特定の地理座標上にあるといった注釈を含めることができます。インターネットを動かしている既存の技術を使って、人間の知識の構成要素を構造化し、機械が読み取ることができる形に符号化しているのです。

Schema.org の語彙は当初、ウェブページ内のセマンティックコンテンツをマークアップするために設計されました。しかし、現在では、例えばバルクダウンロードや、トリプルストアに保存されたのち SPARQL により外部システムからクエリを介して利用可能になった情報等、セマンティックウェブ上で様々な方法で共有されるメタデータの基幹技術として広く使用されています。Schema.org を用いて定義されているクラスやプロパティは、多くの Linked Data アプリケーションに見られます。注目すべき例の一つに、図書館のデータのオープンなウェブへの統合を推進している、非営利の図書館共同組織である OCLC があります。

OCLC の戦略は、長年にわたり図書館が作成してきた価値のあるメタデータを、Linked Data としてウェブ上に公開することにあります。OCLC による Linked Data 実装事業では、2014 年 4 月に図書及び他のクリエイティブな作品についての 1 億 9700 万のオープンな書誌記述を公開するにあたり、中核的な語彙として Schema.org が使用されました^{4,5}。OCLC の共有メタデータでは、Schema.org の語彙に、一般的に使用されている他の RDF 語彙から選択されたクラスやプロパティ、あるいは図書館のニーズを満たすために OCLC が新規に定義したいくつかのクラスやプロパティが追加されています。ここでも、メタデータは、Linked Data のグラフを広げるデータ同士のつながりを構築するように設計されています。例えば、OCLC は、新聞や楽譜等、図書館資料に見られる情報資源のタイプのための新しいクラスを新設しましたが、その定義には、これらのクラスを Schema.org の CreativeWork (クリエイティブな作品) のサブクラスとして定義するトリプルが含まれます。こうすることで、メタデータの提供者によって関係性が明示されない場合でも、Linked Data グラフを参照するソフトウェアは、新聞や楽譜等の情報資源がクリエイティブな作品として理解され得ると推測できます。

Web Ontology Language (OWL)

Web Ontology Language (OWL: ウェブ・オントロジー言語) はセマンティックウェブの初期から存在する、基本的なツールです。OWL には二種類の形式が存在します。一つは RDF/XML 型構文で、OWL の新しいバージョンである OWL2 のすべてのツールにサポートが義務付けられています。もう一方は関数型構文です。語彙の設計者が RDF のクラスやプロパティを正式にドキュメント化する方法はいくつか存在しますが、OWL は最も多用される方法の一つです。OWL は機械可読な方法で正式なセマンティクスを表現し、共有された RDF データに関する機械による推論を促すように設計されています。そのため、OWL 言語は、例えば個人をあるクラスのメンバーとして定義したり、二つのクラスまたは個人を同等のものであると記述したり、二

⁴ <https://www.oclc.org/en-US/news/releases/2014/201414dublin.html>

⁵ <https://www.oclc.org/worldcat/data-strategy.en.html>

つのクラスの相互作用または相互補完の方法を記述したり、あるクラスのメンバーに含まれる場合に別のクラスのメンバーから除外される（「非統合状態」）ケースを示す等、セマンティックウェブの主要なタスクを実行するために用いられます。

OWLには、複雑さが異なるいくつかの形式が存在します。その一つがOWL Liteであり、ほとんどの実装者のニーズはこれで満たされます。RDF 語彙の設計者はしばしば、オントロジーそのものの基本的な記述と管理情報を提供するために、`owl:Ontology` を使い、RDF/XML 形式で完全な OWL オントロジーのドキュメントを作成します。これには、オントロジーの名前、設計者、他のオントロジーに対する関係についての情報等が含まれます。そしてこのドキュメントは、RDF や RDFS からのメカニズムと共に、OWL 言語を使用して、クラス、プロパティ及びそれらの関係を定義します。

OWL の使用例: オントロジーソフトウェア *Protege* を用いて演習用に作られた Pizza Ontology からの抜粋 (RDF/XML 構文)

```
<owl:Class rdf:about="#Pizza">
  <rdfs:label xml:lang="en">Pizza</rdfs:label>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#hasBase"/>
      <owl:someValuesFrom rdf:resource="#PizzaBase"/>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf rdf:resource="#Food"/>
  <owl:disjointWith rdf:resource="#PizzaTopping"/>
</owl:Class>
<owl:ObjectProperty rdf:about="#hasTopping">
  <rdf:type
rdf:resource="http://www.w3.org/2002/07/owl#InverseFunctionalP
roperty"/>
  <rdfs:comment xml:lang="en">
    Note that hasTopping is inverse functional because
    isToppingOf is functional
  </rdfs:comment>
  <rdfs:domain rdf:resource="#Pizza"/>
  <rdfs:subPropertyOf rdf:resource="#hasIngredient"/>
  <rdfs:range rdf:resource="#PizzaTopping"/>
  <owl:inverseOf rdf:resource="#isToppingOf"/>
</owl:ObjectProperty>
<owl:Class rdf:about="#PizzaBase">
  <rdfs:label xml:lang="pt">BaseDaPizza</rdfs:label>
  <rdfs:subClassOf rdf:resource="#Food"/>
  <owl:disjointWith rdf:resource="#PizzaTopping"/>
  <owl:disjointWith rdf:resource="#Pizza"/>
</owl:Class>
```

Simple Knowledge Organization System (SKOS)

Simple Knowledge Organization System (SKOS) もまた、早い時期からある基本的なセマンティックウェブの RDF 語彙です。この語彙の目的は、分類（タクソノミー）、類語集（シソーラス）、件名標目表、分類体系等の知識組織化体系の形式を符号化することです。SKOS 語彙の中心は「Concept」クラスであり、このクラスは表現される知識組織化体系内のエントリー項目に使用されます。Concept は複数のグループ

や体系に組織化されます。それらには優先ラベルまたは代替ラベルとして表されるラベルが付与され、また表記法 (notation) が通例、コードで付与されます。SKOS は、テキストによる定義、概念の範囲に関する注釈及び例を含むことができます。また、より広い、より狭い、あるいは関連する等の語彙間の関係に対するプロパティや、編集注記、知識組織化体系における語彙のエントリー項目の変遷、変更履歴等、概念に関する管理情報も提供します。SKOS のほとんどは、正式な知識組織化体系の符号化という限られたユースケースで主に役立ちますが、関連する概念間の連携に使用されるいくつかのプロパティが、他の Linked Data アプリケーションで広く使われています。SKOS プロパティの中でも特に、`broadMatch`、`closeMatch` 及び `exactMatch` がよく使用されています。

SKOS の使用例: 米国議会図書館分類における “Symphonies” 項目からの抜粋 (N-Triple 構文)

```
<http://id.loc.gov/authorities/classification/M1001>
<http://www.w3.org/2004/02/skos/core#prefLabel> "Symphonies" .
<http://id.loc.gov/authorities/classification/M1001>
<http://www.w3.org/2000/01/rdf-schema#label> "Music and Books
on Music--Music--Instrumental music--Orchestra--Original
compositions--Symphonies" .
<http://id.loc.gov/authorities/classification/M1001>
<http://www.w3.org/2004/02/skos/core#altLabel> "Symphonies" .
<http://id.loc.gov/authorities/classification/M1001>
<http://www.w3.org/2004/02/skos/core#broader>
<http://id.loc.gov/authorities/classification/M1001-M1049> .
<http://id.loc.gov/authorities/classification/M1001>
<http://www.w3.org/2004/02/skos/core#notation> "M1001" .
<http://id.loc.gov/authorities/classification/M1001>
<http://www.w3.org/2004/02/skos/core#inScheme>
<http://id.loc.gov/authorities/classification> .
```

ダブリンコア (DC)

ダブリンコアメタデータ基本記述要素集合 (DCMES) は 1995 年にオハイオ州ダブリンで開催された、ネットワーク化された電子情報のためのメタデータについての検討会議の結果、作成されました。出席者は、ほとんどの種類のデジタル情報に共通する中心的 (コア) な要素を特定するという課題に取り組みました。この最初の会議では、13 個の要素が定義され、その後すぐに、現在 DCMES として知られている 15 個の要素 (寄与者 (contributor)、時空間範囲 (coverage)、作成者 (creator)、日付 (date)、内容記述 (description)、記録形式 (format)、資源識別子 (identifier)、言語 (language)、公開者 (publisher)、関係 (relation)、権利管理 (rights)、出処 (source)、キーワード (subject)、タイトル (title)、資源タイプ (type)) に拡大しました。「シンプルダブリンコア」またはダブリンコア (DC) としても知られるこれらのセットは ISO 15836 及び ANSI/NISO Z39.85 として標準化され、いずれも「ダブリンコアメタデータ基本記述要素集合」と呼ばれています。

DC は意図的に簡潔にされ、その初期から比較的広く採用されました。DC の要素はすぐにウェブページに埋め込まれ、初期のサーチエンジンによる索引付けに多用されました。DC は、OAI-PMH プロトコル経由で共有される記述に必須のメタデータ形式として採用されました。それでもなお、DC 以上の詳細さが多くのメタデータ実装者により求められたため、語彙の維持管理を行っているダブリンコアメタデータイニシアチブ (DCMI) は、「限定子」を使用してシンプルな DC を拡張し、コア

要素に改良を加えました。この拡張版は、現在 DCTERMS として知られています。

DC と DCTERMS は、XML と RDF の両方の語彙として定義されています。ただし、DCTERMS は、RDF としてより厳密に定義され、クラスに対して正式に定義される定義域や値等、RDF の特長を活用しています。ダブリンコアは他のメタデータ語彙以上に、XML と RDF 両方のコミュニティに役立つよう試みることで、両者の橋渡しをしようとしています。しかし、本イニシアチブは、明らかに RDF の方向に向かっています。初期の DC XML 要素と同様に、DCTERMS RDF のクラスとプロパティは現在、実際の世界の **Linked Data** で広く使われていますが、中でも最も広く実装されているのは、オリジナルの 15 個のダブリンコアプロパティを表す DCTERMS のプロパティです。

DCMI は、メタデータ語彙の定義だけでなく、多数の技術仕様をダブリンコアの名のもとに開発してきました。DCMI 抽象モデルは、レコードベースの規範的な XML の手法と、グラフベースのオープンな RDF の手法との妥協点を見出そうとしています。DCMI 抽象モデルは、情報資源を RDF グラフよりも若干規範的なモデルとして記述しています。そこでは、情報資源の特徴についての個別の記述を束ねる有界の「記述セット」を規定し、情報資源の記述のための統制語彙の役割と構文ベースの規則を正式に定義しています。DCMI 抽象モデルは RDF を強化する一方でそれ自体を参照しています。このモデルは現在 DCMI コミュニティ以外ではあまり参照されておらず、RDF モデルに構造を付加することがメタデータのユーザーにとって価値があるかどうかはまだ不明です。

同様に、アプリケーションプロファイルのためのシンガポールフレームワーク（単に「シンガポールフレームワーク」とも言われる。）は、特定の目的のためのメタデータの使用について文書化する際の仕様を定めています。この仕様には、目的と、そのプロファイルに一致する記述で表現され得る情報資源のタイプと、使用可能なメタデータ語彙のリストと、準拠したメタデータを記述するための符号化構文の指定が含まれます。DCMI 抽象モデルと同様に、DCMI コミュニティはシンガポールフレームワークを使用して、他のモデルよりも複雑なモデルを導入しました。この複雑さが有用であるかどうかは時間が経てば分かるでしょう。

DC の使用例: ノーステキサス大学から Portal to Texas History に提供された画像についてのメタデータレコードからの抜粋 (XML 構文)

```
<?xml version="1.0" encoding="UTF-8"?>
<oai_dc:dc
xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:dc="http://purl.org/dc/elements/1.1/"
xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc
/ http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
  <dc:title>
    Maxine Walker Perini as a Child with Pet Dog and
    Doll in Doll Buggy
  </dc:title>
  <dc:description>Copy negative of young Maxine Walker
  Perini wearing a coat and bow, holding a Boston terrier, and
  pushing a doll in buggy. She is on the sidewalk in front of
  the stairs of a house.</dc:description>
  <dc:subject>People - Individuals</dc:subject>
  <dc:subject>Social Life and Customs - Pets -
  Dogs</dc:subject>
```

```

<dc:subject>dolls</dc:subject>
<dc:subject>toys</dc:subject>
<dc:subject>Perini, Maxine Walker</dc:subject>
<dc:subject>pets</dc:subject>
<dc:subject>children</dc:subject>
<dc:coverage>United States</dc:coverage>
<dc:coverage>
    New South, Populism, Progressivism, and the Great
    Depression, 1877-1939
</dc:coverage>
<dc:type>Photograph</dc:type>
<dc:format>1 photograph : negative, b&w ; 4 x 5
in.</dc:format>
<dc:format>Image</dc:format>
<dc:identifier>local-cont-no: 81-00423-7</dc:identifier>
<dc:identifier>
    http://texashistory.unt.edu/ark:/67531/metaph50844/
</dc:identifier>
<dc:identifier>ark:
ark:/67531/metaph50844</dc:identifier>
</oai_dc:dc>

```

Friend of a Friend (FOAF)

オープンなウェブのための初期の RDF 語彙としてもう一つ、Friend of a Friend (FOAF) があります。この語彙は、人や組織に関する記述メタデータを、それぞれの属性及び関係と共に提供します。FOAF のクラスには、Person、Organization、Group、Project 等があり、これらは name、title、member 等のプロパティを利用します。FOAF Core は、改良や精緻化の基礎となるように設計された、少数の主要なクラスとプロパティを定めています。また、FOAF には、PersonalProfileDocument や accountName 等の、人や組織とソーシャルウェブとの相互作用の方法に関連する、クラスやプロパティのリストが含まれます。

FOAF は主に、人や組織を識別して、それらについての基本情報を提供するためにシステムで使用されます。しかし、その語彙には、個人がある特定のグループに属していることを示したり、個々人の趣味や関心事を記載したりする等、より高度な特徴を表すものも含まれます。

FOAF の使用例: ノーベル医学賞を受賞した Peter C. Doherty の ORCID データからの抜粋 (Turtle 構文)

```

<http://orcid.org/0000-0002-5028-3489>
  a      foaf:Person , prov:Person ;
  rdfs:label "Peter Charles Doherty" ;
  foaf:account <http://orcid.org/0000-0002-5028-3489/> ;
  foaf:based_near
    [ a      gn:Feature ;
      gn:countryCode "AU" ;
      gn:parentCountry
<http://sws.geonames.org/2077456/>
    ] ;
  foaf:familyName "Doherty" ;
  foaf:givenName "Peter Charles" ;

```

```
foaf:publications <http://orcid.org/0000-0002-5028-3489/> .
```

ONline Information eXchange (ONIX)

e コマース業界における高度なメタデータのフォーマットとしては、出版業界の ONline Information eXchange (ONIX) XML スキーマがあります。ONIX は三つのフォーマットで構成され (ONIX for Books、ONIX for Serials、ONIX for Publications Licenses)、ONIX for Books が最も広く使用されています。ONIX は、紙、電子書籍及び雑誌分野の国際標準団体である EDItEUR と米国と英国の出版者団体により維持されています。2009 年にリリースされた ONIX 3.0 では、電子書籍へのサポートが強化されました。出版社は ONIX for Books を使用して、取り扱う商品に関する詳細なメタデータを小売業者に提供することができます。フィードには、ISBN やバーコード等の識別子、サイズ、堅牢に符号化された複数の形式のタイトルの、複数の形式による著者や他の貢献者の名前やその所属、図書の表紙に書かれたテキスト等の宣伝材料等の詳細情報が含まれます。

EXchangeable Image File Format (Exif)

The EXchangeable Image File Format (Exif) は適切な名称ではないかも知れません。なぜなら、これはファイルフォーマットではなく、デジタル画像ファイルに埋め込まれたメタデータのためのタグ構造であるからです。これは日本のデジタルカメラ業界から生じ、現在ではほぼすべてのデジタルカメラやスマートフォンの製造業者により使用されています。Photoshop や Flickr 等の、画像の編集や共有を行う様々なソフトウェアパッケージが Exif をサポートしています。しかし、Instagram や Facebook 等の多くのソーシャルメディアサイトは、共有された画像から Exif メタデータを取り除きます。一方でこうしたサイトでは Exif メタデータを取り除く前に地理位置情報を読み込んで保存することもあります。TIFF 及び JPEG ファイルフォーマットは埋め込まれた Exif をサポートしていますが、JPEG2000、PNG 及び GIF はサポートしていません。Exif には主に画像に関する技術メタデータが保存されますが、埋め込まれた音声に関するメタデータのための補足的なセクションも存在します。Exif の仕様には、画素の大きさ、撮影日時、ISO 感度、レンズ口径、ホワイトバランス、使用レンズの情報等のメタデータ要素が含まれます。

主要なメタデータ言語：文化遺産分野の例

MAchine Readable Cataloging (MARC)

図書館コミュニティで最も広く使用されているメタデータ言語は、XML 及び RDF の技術や、現在使用されているメタデータフォーマット一般よりもはるか以前に生まれました。MAchine Readable Cataloging (MARC) は、1968年に、目録カードの情報を機械可読形式で各図書館に配布する実験を行うための、米国議会図書館におけるパイロットプロジェクトから始まりました。それ以来、MARC は、オンライン図書館目録の基礎となるメタデータフォーマットとして、また、図書館間で書誌レコードを共有する方法として定着しています。OCLC の WorldCat データベースは、図書館間におけるレコードの共有と複数機関の所蔵資料の横断検索に使用され、2016年7月現在、約3億8000万のMARC書誌レコードを保持しています⁶。

MARC は、ANSI/NISO Z39.2「情報交換フォーマット」及びISO 2709「情報とドキュメンテーション—情報交換用フォーマット」として標準化されています。ISO 2709 フォーマットは情報保管の効率が最大化するよう設計されています。このレコードの大部分は任意の長さの可変長フィールドで構成されています。レコードの先頭にあるディレクトリは、各フィールドの先頭をプロセッサに示し、空ビットが保存されないようにしています。フィールド名は、言語ベースのラベルではなく、英数字の文字列であり、タグと呼ばれています。ISO 2709 に準拠した MARC の一つのフィールドには、そのフィールドの値に関するメタ情報を示す一文字から成る指標が二個含まれます。これらのフィールドは、さらに英数字のサブフィールドに細分化され、フィールド内のあらゆる情報の符号化に使用されます。

MARC は実際には一種のフォーマット群であり、ISO 2709 は様々な国で多様な実装が行われています。最も知られているものは、米国議会図書館により維持管理されている MARC21 フォーマットであり、これは米国、カナダ及び多くの英語圏の世界で使用されています。MARC21 は、MARC21 書誌、MARC21 典拠、MARC21 所蔵、MARC21 分類表、MARC21 コミュニティインフォメーションの五種類のフォーマットから成ります。最初の二つは広く使用され、三つ目は特定のシステムのみで使用されます。最後の二つの実装はさらに限定されます。MARC21 は数値のフィールドコードのみを使用するため、その母体である ISO 2709 フォーマットよりも若干制限的です。

MARC21 書誌フォーマットは、図書館が所蔵する個別の資料（アイテム）を記述するために使用されます。このフォーマットは数百のフィールドから構成されていますが、頻繁に使用されるのは、ごく少数のコアセットです。例えば、各種タイトル、著者にあたる人または団体、版及び出版に関する事項、形態に関する記述、シリーズ、注記、主題やジャンルを記述するフィールド等です。MARC21 典拠フォーマットは、人、団体、著作のタイトル、主題、ジャンルに関する統制語彙のドキュメンテーションのために使われます。これらの統制語彙は、MARC21 書誌レコード内の適切なフィールドへの入力値として使用され、書誌レコードの整合性を維持し、検索を容易にします。MARC 典拠フォーマットには統制された標目を符号化するためのフィールドが含まれ、ここには、例えば人物の職業、人物または団体の住所、

⁶ <https://www.oclc.org/worldcat/inside-worldcat.en.html>

音楽作品の調性等、実体に関する付加的なメタデータが含まれることが多いです。さらに、その実体の名前の別名を符号化したり、統制された標目の特定の形式が選択された理由を示す注記を付加したりするためのフィールドも含まれます。

MARC の使用例: 英国図書館オンライン目録の Nigella Lawson の“*How to Eat*”の書誌レコード

```
FMT      BK
LDR          am a2200217ua 4500
001      011981326
008      981130s1998    enka  ||    001 ||eng
015      |a GB98Z0319 |2 bnb
020      |a 0701165766 : |c £25.00
040      |a StDuBDS |d Uk
08204    |a 641.5 |2 21
1001     |a Lawson, Nigella, |d 1960-
24510    |a How to eat : |b the pleasures and principles of
          good food / |c Nigella Lawson.
260      |a London : |b Chatto & Windus, |c 1998.
300      |a xviii,526p. : |b ill. (some col.) ; |c 24cm.
336      |a text |2 rdacontent
337      |a unmediated |2 rdamedia
338      |a volume |2 rdacarrier
500      |a Includes index.
650 0    |a Cooking.
85241    |a British Library |b HMNTS |j YK.1998.b.9105
SYS      011981326
```

Bibliographic Framework Initiative (BIBFRAME)

Bibliographic Framework Initiative (BIBFRAME) は、書誌情報の符号化と共有のための新しいモデルを設計することを目指して、米国議会図書館が中心となって実施しているプロジェクトです。このモデルは、図書館のデータが 21 世紀の情報環境下でより有効に用いられ、台頭する「データのウェブ」の一部となることができるよう、Linked Data の原則に従って構造化されています。BIBFRAME は正式な RDF 語彙です。最終的には MARC21 の後継として、MARC21 のセマンティクスの多くを保持することが意図されており、既存データのほとんどを移行することが可能となります。大規模で包括的な BIBFRAME モデルでは、モデルが対象とするスコープにおいて重要と見なされるすべての特徴について、BIBFRAME 自身の名前空間の中にクラスとプロパティを定義します。ただし、最上位レベルのいくつかのクラスは、他のコミュニティにより定義された良く知られている実体のサブクラスとして定義されます。BIBFRAME 2.0 は 2016 年 4 月にリリースされましたが、設計上の多くの問題が議論されているため、このモデルは完全に安定しているとは言えません。

BIBFRAME 2.0 モデルでは、著作（情報資源の概念的な本質）、インスタンス（著作を個別にモノとして具体化したもの）、アイテム（インスタンスの実際の物理的または電子的な複製）、エージェント（著作に関連する個人や団体）及びイベント（著作に記録される出来事）の実体を定義しています。これらはそれぞれ RDF クラスとしてモデル化され、カテゴリ内のより詳細な概念を定義するサブクラスも付随します。

BIBFRAME においては、従来から図書館によって記録され MARC フォーマットで符号化されてきた書誌及び典拠データのタイプに対し、付加的なクラスとプロパティが定義されます。ある特定の図書館に存在する、タイトル、作成者、主題、ジャ

ンル、形式、言語、作成に関する事項、版、物理的特性、識別子、注釈、他の情報資源との関係及び所蔵情報に関するメタデータはすべて、BIBFRAME に含まれます。

BIBFRAME の使用例: Suzanne Pickett の “Hot Dogs for Thanksgiving” についての BIBFRAME Website 上のデータからの抜粋 (RDF/XML 構文)

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:bf="http://bibframe.org/vocab/">
  <bf:Work
    rdf:about="http://bibframe.org/resources/sample-lc-2/731515">
    <rdf:type
rdf:resource="http://bibframe.org/vocab/Text"/>
    <bf:authorizedAccessPoint>Pickett, Suzanne. Hot dogs
for Thanksgiving / by Suzanne Pickett.Hot dogs for
Thanksgiving</bf:authorizedAccessPoint>
    <bf:workTitle
rdf:resource="http://bibframe.org/resources/sample-lc-2/731515title29"/>
    <bf:creator
rdf:resource="http://bibframe.org/resources/sample-lc-2/731515person30"/>
    <bf:language
rdf:resource="http://id.loc.gov/vocabulary/languages/eng"/>
    <bf:subject
rdf:resource="http://bibframe.org/resources/sample-lc-2/731515person32"/>
    <bf:subject
rdf:resource="http://bibframe.org/resources/sample-lc-2/731515topic33"/>
    <bf:subject
rdf:resource="http://bibframe.org/resources/sample-lc-2/731515topic34"/>
    <bf:subject
rdf:resource="http://id.loc.gov/vocabulary/geographicAreas/n-us-al"/>
    <bf:classificationLcc
rdf:resource="http://id.loc.gov/authorities/classification/CT275.P628"/>
    <bf:authorizedAccessPoint xml:lang="x-bf-hash">pickettsuzannehotdogsforthanksgivingengworktext</bf:authorizedAccessPoint>
  </bf:Work>
  <bf:Instance>
    <rdf:type
rdf:resource="http://bibframe.org/vocab/Monograph"/>
    <bf:instanceTitle
rdf:resource="http://bibframe.org/resources/sample-lc-2/731515title44"/>
    <bf:isbn13
rdf:resource="http://isbn.example.org/9781881320760"/>
    <bf:publication>
    <bf:Provider>
```

```

        <bf:providerName>
          <bf:Organization>
            <bf:label>Black Belt Press</bf:label>
          </bf:Organization>
        </bf:providerName>
        <bf:providerPlace>
          <bf:Place>
            <bf:label>Montgomery, AL </bf:label>
          </bf:Place>
        </bf:providerPlace>
        <bf:copyrightDate>c1998.</bf:copyrightDate>
      </bf:Provider>
    </bf:publication>
    <bf:extent>190 p. ;</bf:extent>
    <bf:dimensions>23 cm.</bf:dimensions>
    <bf:instanceOf
rdf:resource="http://bibframe.org/resources/sample-lc-
2/731515"/>
    </bf:Instance>
    <bf:Annotation
      rdf:about="http://bibframe.org/resources/sample-lc-
2/731515annotation40">
      <bf:derivedFrom
rdf:resource="http://bibframe.org/resources/sample-lc-
2/731515.marcxml.xml"/>
      <bf:descriptionSource
rdf:resource="http://id.loc.gov/vocabulary/organizations/dlc"/
>
      <bf:descriptionConventions
rdf:resource="http://id.loc.gov/vocabulary/descriptionConventi
ons/aacr2"/>
      <bf:generationProcess>DLC transform-tool:2015-07-23-
T17:01:00</bf:generationProcess>
      <bf:changeDate>1998-07-21T12:52</bf:changeDate>
      <bf:annotates
rdf:resource="http://bibframe.org/resources/sample-lc-
2/731515"/>
    </bf:Annotation>
  </rdf:RDF>

```

Metadata Object Description Schema (MODS)

米国議会図書館の Metadata Object Description Schema (MODS) は、図書館にとって特に有益な書誌情報のための XML スキーマです。その要素と属性は英語のタグですが、MARC フォーマットのタグに大きな影響を受けています。実際 MODS は、MARC のほとんどのセマンティクスを表現し得る、XML ベースのアプリケーションと環境で使用しやすいスキーマと考えられていました。

MODS XML スキーマは、書誌記述の関連部分をグループ化し、20 個の最上位レベルの要素 (titleInfo、name、typeOfResource、genre、originInfo、language、physicalDescription、abstract、tableOfContents、targetAudience、note、subject、classification、relatedItem、identifier、location、accessCondition、part、extension、recordInfo) に整理しています。MODS の特徴として拡張要素があり、どのような XML の名前空間内の付加的なメタデータも、MODS レコードに埋め込むことができ

ます。これにより拡張性がある程度高まります。すなわち、付加的な要素が必要な場合、それらの要素を同一の記述に含めることが可能となります。もし MODS レコードを処理するシステムが拡張要素内で使用される要素の名前空間を理解した場合は、そのデータの索引付けと処理が可能となります。

MODS の使用例: コロンビア大学の Academic Commons Repository に登録された学術論文のメタデータから抜粋

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<mods xmlns="http://www.loc.gov/mods/v3"
xmlns:xlink="http://www.w3.org/1999/xlink"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.loc.gov/mods/v3
http://www.loc.gov/standards/mods/v3/mods-3-4.xsd">
  <titleInfo>
    <title>You Can&#8217;t Think and Hit at the Same Time:
Neural Correlates of Baseball Pitch Classification</title>
  </titleInfo>
  <name type="personal" ID="jss2212">
    <namePart type="family">Sherwin</namePart>
    <namePart type="given">Jason Samuel</namePart>
    <role>
      <roleTerm type="text">author</roleTerm>
    </role>
    <affiliation>Columbia University. Biomedical
Engineering</affiliation>
  </name>
  <name type="personal" ID="jsm2112">
    <namePart type="family">Muraskin</namePart>
    <namePart type="given">Jordan Scott</namePart>
    <role>
      <roleTerm type="text">author</roleTerm>
    </role>
    <affiliation>Columbia University. Biomedical
Engineering</affiliation>
  </name>
  <name type="personal" ID="ps629">
    <namePart type="family">Sajda</namePart>
    <namePart type="given">Paul</namePart>
    <role>
      <roleTerm type="text">author</roleTerm>
    </role>
    <affiliation>Columbia University. Biomedical
Engineering</affiliation>
  </name>
  <name type="corporate">
    <namePart>Columbia University. Biomedical
Engineering</namePart>
    <role>
      <roleTerm type="text">originator</roleTerm>
    </role>
  </name>
  <extension
xmlns:rioxxterms="http://docs.riox.net/schema/v1.0/rioxxterms
/" xmlns="http://www.loc.gov/mods/v3"
```

```
xsi:schemaLocation="http://docs.riox.net/schema/v1.0/rioxxterms/ http://docs.riox.net/schema/v1.0/rioxxterms.xsd">
  <rioxxterms:funder>National Institutes of Health
(U.S.)</rioxxterms:funder>
  <rioxxterms:projectid>R01-
MH085092</rioxxterms:projectid>
</extension>
<typeOfResource>text</typeOfResource>
<genre>Articles</genre>
<originInfo>
  <dateIssued encoding="w3cdtf"
keyDate="yes">2012</dateIssued>
</originInfo>
<language>
  <languageTerm type="text">English</languageTerm>
</language>
<abstract>Hitting a baseball is often described as the most
difficult thing to do in sports. A key aptitude of a good
hitter is the ability to determine which pitch is coming. This
rapid decision requires the batter to make a judgment in a
fraction of a second based largely on the trajectory and spin
of the ball. When does this decision occur relative to the
ball's trajectory and is it possible to identify neural
correlates that represent how the decision evolves over a
split second? Using single-trial analysis of
electroencephalography (EEG) we address this question within
the context of subjects discriminating three types of pitches
(fastball, curveball, slider) based on pitch
trajectories...</abstract>
<subject>
  <topic>Biomedical engineering</topic>
</subject>
<subject>
  <topic>Neurosciences</topic>
</subject>
<subject>
  <topic>Bioinformatics</topic>
</subject>
<relatedItem type="host">
  <titleInfo>
    <title>Frontiers in Neuroscience</title>
  </titleInfo>
  <part>
    <detail type="volume">
      <number>6</number>
    </detail>
    <detail type="issue">
      <number>177</number>
    </detail>
    <extent unit="page">
      </extent>
    <date>2012-12-19</date>
  </part>
  <identifier
type="doi">http://dx.doi.org/10.3389/fnins.2012.00177</identif
ier>
```

```
<identifier type="issn">1662-453X</identifier>
</relatedItem>
<identifier type="CDRS
doi">http://dx.doi.org/10.7916/D8GQ6VVF</identifier>
<location>
  <physicalLocation
authority="marcorg">NNC</physicalLocation>
</location>
</mods>
```

CIDOC Conceptual Reference Model (CIDOC CRM)

博物館コミュニティもまた、精力的に正式なメタデータ標準を開発し、使用してきました。国際博物館会議のドキュメンテーション委員会（CIDOC）は、博物館のドキュメンテーション及び文化遺産のメタデータで必要とされる概念の基本オントロジーとなる、概念参照モデル（CRM）を開発しました。このモデルは、ISO 21127「情報とドキュメンテーション—文化遺産情報のための参照用オントロジー」として標準化されています。

CIDOC CRMは、テキストを用いながらも形式的な参照モデルであり、その派生として RDFS によっても表すことができます。そのため、このモデルは主に、複数のコミュニティ間及び異なるメタデータフォーマット間で概念を検討するにあたっての、語彙のコアセットとしての使用を想定されています。CIDOC CRM に対応したシステムは、多数の RDF 語彙または XML スキーマを使用してデータを符号化している可能性があります。

CIDOC CRM の際立った特徴は、文化遺産オブジェクトを記述するだけでなく、それらの製作や寿命に関連する行動や出来事（イベント）にも重点を置いていることです。このオントロジーにより定義される実体はそのため多岐にわたり、例えば、場所、期間、イベント、行為者（関連する行為の主体）、物理的な事物、情報オブジェクト等といった特性に対するクラスも含まれます。

Categories for the Description of Works of Art (CDWA)

Categories for the Description of Works of Art (CDWA) は、芸術作品に関する情報の概念フレームワークです。CDWA は芸術作品についての情報を扱うシステムを設計するための基礎となるものとして考案され、主に美術館コミュニティで使用されています。CDWA は約 540 個のデータ要素とそれら要素間の関係を定義していますが、そのうちのいくつかは、作品の記述として意義のあるものを作るために最低限必要な要素として定義されています。CDWA は概念フレームワークであり、情報の符号化方式についての具体的規定はしていません。

CDWA の設計は、質が高く機械処理可能な情報の作成に重点を置いています。多数のカテゴリについて表示用と索引作成用に別々のデータを使用することを推奨することで、より効果的な検索を可能にすると同時に、ユーザーに対してまとまった情報を表示することができます。また、CDWA の記述に記録された内容に対して信頼できる典拠を使用することも推奨されており、ある特定のカテゴリで情報を蓄積する際に、特定の典拠の使用を示唆することさえあります。

CDWA は、タイトルや作者等の基本的な事実、スタイルや制作時期、材料や技法、主題等の作品内容、銘やサイズ等の物理的属性、現在の所在地、収集歴や展示及び貸与履歴等の管理に関する詳細情報、他の芸術作品との関係や参考文献等のカテゴリを定義しています。

Visual Resources Association Core (VRA Core)

視覚情報資源コミュニティは、芸術作品や建築物の複製物の発見と利用に依拠する、美術史等の学問領域をサポートしています。このコミュニティの専門職の団体である視覚資料協会（Visual Resources Association (VRA)）は、芸術作品とそれらの表象（画像）に関する情報を記録するための VRA Core メタデータ語彙を開発しました。VRA Core は、二種類の XML スキーマとして表されます。一つは、多くの VRA Core XML 要素の「type」属性の値として事前に設定した値のみを使うよう厳密に定めた制限的スキーマであり、他方は、「type」属性の値に任意の文字列を入力することが可能な非制限的スキーマです。VRA Core は視覚情報資源コミュニティにより維持管理されていますが、正式な XML スキーマとドキュメンテーションは米国議会図書館が提供しています。

VRA Core の際立った特徴は、芸術作品自体に関するメタデータと、それら作品の画像に関するメタデータが分離していることです。同一の VRA Core 要素を使用して、作品または画像のどちらも記述することはできますが、これらの要素の値は異なることが多分に予想されます。例えば、作成日付や作成者は概して異なる値となるでしょう。VRA Core に含まれる作品や画像をグループ化することもできます。CDWA と同様に、VRA Core には、表示用の値のほかに、索引作成用の値を記録することができます。芸術作品に焦点を当てたメタデータスキーマとして、VRA には、エージェント（作品の制作者）、タイトル、日付、権利等の文化的資料のためのより一般的な要素に加え、材料、技法、銘、寸法、文化的コンテキスト、スタイル・制作年代、作品のタイプといった要素が含まれます。

Encoded Archival Description (EAD)

Encoded Archival Description (EAD) は、英語圏の文書館コミュニティで使用されている主要なメタデータ語彙です。VRA Core と同様に、EAD はこの分野の専門職—この場合は米国アーキビスト協会—により開発と維持管理が行われており、EAD の公開文書は、米国議会図書館が提供するウェブサイトで閲覧できます。文書館は主に、フォンド（同出所記録資料群）と呼ばれるコレクション単位で資料を扱うため、紙であれデジタルであれ、通常は個別資料の記述をしません。その代わりに、階層的な記述を行います。すなわち、共通の出所を持つ情報資源の集合が 1 個の全体として記述され、さらに（任意で）細分化されたサブセットがより詳細に記述されます。これらの階層化された記述は、ファインディングエイドと呼ばれるドキュメント中に、コレクションに関する履歴情報、作成者及びコンテキスト情報と併せて表示されます。

EAD は、純粋な記述メタデータ標準ではなく、ファインディングエイドのための XML マークアップ言語です。しかし、そのタグはフォーマットを重視したものではなく、セマンティックな性質が強いものです。EAD 要素の多くは、アーカイブに特有の概念のためのものであり、たとえば、一連の記録の作成者の経歴や記述対象の資料の範囲や注釈、シリーズ、サブシリーズ、またはファイル等のアーカイブコンポーネント、またはボックスやフォルダ等の保管容器等についての要素等です。これらの各要素は、複数の階層を持つ記述のうち、どの階層でも使用することができます。また、EAD は、リストやパラグラフ等の文章構成を示す要素や、必要に応じて、名前や日付等の注目すべき重要な単語やフレーズをマークアップする要素を提供しています。

主要なメタデータ言語：その他の例

Data Documentation Initiative (DDI)

Data Documentation Initiative (DDI) メタデータ標準は、社会科学、行動科学及び経済学のデータを記述するために設計された大規模な要素集合です。DDI は、研究の構想、データ収集、データの正規化と分析、共有、アーカイブ等の、研究のライフサイクルのすべての段階において有用な要素で構成されています。DDI は対象とする学問領域において主要な標準となっており、これらの研究コミュニティにおいて、研究データの管理計画及びオープンな共有と長期保存の意義が増すにつれて、DDI の重要性もさらに高まっています。

DDI は、複数のモジュールに分けられた多数の XML スキーマとして、標準的に定義されています。DDI には、調査研究、データ収集方法、質問と回答、変数、処理前または処理済みのデータセットへのリンク及び研究とデータセットとの間の関係に関するメタデータが含まれます。XML 版の DDI は、社会科学のデータを取り扱う多くのソフトウェアパッケージに実装されています。データセットとそれらのメタデータを Linked Data として共有するための、DDI の RDF 表現も開発されています⁷。

また、DDI の要素集合を管理する組織である DDI 協会 (DDI Alliance) は、DDI のメタデータで使用するための統制語彙もいくつか定めています。これらの語彙は、分析単位、データタイプ、収集方法等、社会科学にとって特に重要性を持つ領域を対象としています。

PREservation Metadata: Implementation Strategies (PREMIS)

デジタルアーカイブの分野における主要な保存メタデータ標準は、PREservation Metadata: Implementation Strategies (PREMIS) です。PREMIS は、デジタル情報保存のプロセスや、実施された保存のための行為の及び行為の責任者に関する情報の記録に必要なデジタルコンテンツのプロパティを記述するためのものです。PREMIS は米国議会図書館で、コミュニティからのインプットを広く受けつつ有志の編集委員会により維持管理されています。PREMIS は XML スキーマとして定義されていますが、OWL による RDF 定義も公表されています。

PREMIS は、オブジェクト (Objects)、環境 (Environments)、イベント (Events)、エージェント (Agents) 及び権利 (Rights) の 5 つの実体を定義しています。これらの実体はデジタル情報保存システムの中で記述や行為の対象となります。オブジェクトは保存行為の対象となる情報の単位であり、相互に関係が定義された知的実体、表現、ファイル、ビットストリーム等を含みます。環境は、これらのデジタルコンテンツが属するハードウェアまたはソフトウェアを表します。イベントはデジタル情報保存環境において特記すべき行為のことです。エージェントは人、組織、またはソフトウェアであり、権利はコンテンツに関する何らかの許諾を記述したものです。これらの実体はそれぞれの目的に応じた様々なプロパティを持ちます。例えば、オブジェクトには、objectIdentifier、preservationLevel、fixity、size、format 等のプロパティがあり、イベントには eventIdentifier、eventType、eventDateTime 等のプロパティがあります。

Text Encoding Initiative (TEI)

Text Encoding Initiative (TEI) は、散文、韻文、テキスト、口述筆記、辞書、写本及びその他の一次資料等、あらゆる種類の機械可読テキストのためのマークアップ

⁷ 訳注：2019年に公開されている。<https://ddialliance.org/Specification/RDF> 参照。

言語です。TEIは非常に大規模なマークアップ言語であるため、どのような実装においても全体を使用するという事は想定されていません。そうではなく、個別のプロジェクトにおいて使用するモジュール（ある特定の目的に関連する要素のグループ）を選択することで、元となるXMLの定義を、XML DTD、XMLスキーマ、またはRELAX NGスキーマとしてレンダリングすることができます。さらに実装の際に、モジュール内の特定要素の除外や名前の変更、要素または要素の属性の追加を行うこともできます。こうした実装レベルのカスタマイズは、「One Document Does it All」(ODD)と呼ばれる、それ自体が一種のTEIであるドキュメンテーションフォーマットで記録されます。

TEIのモジュールはテキストを構成する様々な要素に適用可能です。テキストの基本的な構造を示すタグが広く使用され、段落、見出し、韻文における句、台詞、話者名、ト書き、引用等をマークアップすることができます。名前、数字、日付等の意味的な要素も同様にマークアップ可能です。テキストに関するメタデータと同様に、表や埋め込み画像等の付加的コンテンツや非テキストコンテンツについてのタグも含むことができます。TEIの多言語サポートは強力であり、テキストの任意の箇所で使用される言語を指定すること、ドキュメントの一節を異なる言語で記載すること、Unicodeサポートを通して複数の文字セットの使用を許可すること、ドキュメント固有の字体を定義すること等が可能です。TEIは、テキストが持つ情報のマークアップを可能にする以外に、「校本」モジュールを通して電子版の校訂テキスト(electronic scholarly editions)をサポートしています。このモジュールの要素と属性を使用して、複数の原典間の異同や、あるテキスト（特に写本）についての複数の解釈の符号化をサポートしています。

Music Encoding Initiative (MEI)

Music Encoding Initiative (MEI) は、ほぼ TEI に則った、楽譜のための符号化フォーマットです。これは XML 言語であり、モジュール構成、特定のプロジェクト用のカスタマイズ方法及び ODD の使用を TEI コミュニティから取り入れています。

MEI は、一般的な西洋の記譜法（ほとんどの読者が慣れ親しんでいるであろう形式）、定量記譜法とネウマ記譜法、ギターとリュートのためのタブラチュア等の、広く使用されているいくつかの記譜法の形式をサポートしています。MEI には、対象の楽譜に関するメタデータを記述するためのヘッダーに加えて、総譜、パート、五線譜、調号、音部記号、小節、小節線、音符、和音等、サポートしている各記譜法に必要なすべての記号のためのタグが含まれています。TEI と同様に、MEI は分析や校訂に係る構造の符号化もサポートしています。

メタデータはどのように生成されるのか

文化遺産分野では、ユーザーによる図書、ジャーナル、手稿及び文化財の検索を支援するために、主に記述メタデータが作成されてきました。その作成は、従来、人手による以外にほとんど方法がありませんでした。タイトル、著者、発行日等の特徴は通常、現物の記載を確認したのち、手書きで目録に転記されていました。著者に関する背景情報や上演、出版履歴等のその他のメタデータについては、専門家が範囲を絞った調査を実施し、その調査結果を記録しました。要約や主題等といった付加価値のある解釈を含む情報は、同様に専門家により作成、提供されました。

技術の進歩により、各機関はこの手作業で転記したメタデータを共有し、作業の重複を削減することができるようになりました。次に、用途別のメタデータ入力システムが現れ、その後、スプレッドシート等の汎用ツールによって、メタデータ作成プロセスは加速し始めました。さらに最近では、メタデータ作成用のインターフェースはより高機能化され、ベースとなっているデータ構造をそのまま再現したようなデータ入力画面を必ずしも表示させることのない、使い勝手のよいデザインが採り入れられました。今日では、メタデータは一般的に、XMLやRDFを直接使用するのではなく、間接的なステップを通じて作成されます。この原則の例外の一つは、TEIにおけるテキストの符号化等、メタデータを使用したコンテンツのマークアップです。

また、情報がデジタル形式でのみ創出されることが一般的になるにつれて、技術の進歩によりメタデータの自動生成の精度が向上してきました。特に技術メタデータはこれに当てはまります。ほとんどのファイルフォーマットには何等かの技術的情報が必ず埋め込まれ、ソフトウェアによるコンテンツの解釈を支援しています。また、ファイルが作成された時点での日付やシステムへログインしていたユーザーのID等、システムレベルの情報を使用して付加的な管理情報をデジタルファイルに追加する、ソフトウェアの履歴もメタデータ自動生成の事例の一つです。

ウェブと、様々なソフトウェアシステム間の大規模な統合もまた、メタデータの効率的な共有を可能にしました。組織とユーザーとの間のデジタルコンテンツの送受信が増加する中、メタデータの相互運用性は極めて重要です。メタデータが効率的に共有されると、作業の重複が削減されるため、すべての人にメリットがあります。Amazonは商品に関するメタデータを収集していますが、ONIXのサプライチェーン経由で図書のタイトルや著者等を取得することは、既存の技術を活用し、下流工程でのメタデータ作成の必要性を削減させられるという、統合によってもたらされるメリットを顕著に表した例です。

近年、デジタルコンテンツを解析し、そのメタデータを自動生成する仕組みが生み出されています。音声及び動画から話し言葉を自動的に書き起こす技術は比較的成熟しており、特に専用のサウンドシステムを備えるコントロールされた環境で収録された音声記録の場合は効果的です。動画及び静止画の顔認識技術は急速に進化しています。テキスト情報については、潜在意味解析とトピックモデリングにより、解析対象のテキストに関連するトピックの半教師付き生成が可能となっています。品詞及び固有表現の認識技術は、研究環境で多用されています。写真の中の対象物を特定するアルゴリズムを使用して、自動的に画像に注釈を付ける仕組みは、急成長している研究分野です。音楽情報検索と呼ばれる類似の研究コミュニティは、音

声ファイルの信号処理に注目しています。この研究は、オンラインの音楽ストリーミングサービスにおけるプレイリストの作成にとって不可欠なものです。音楽のジャンルの自動分類等も手掛けようとしています。プログラム化された手段による高品質なデータ生成の実現への進歩と可能性は高まってきています。メタデータの自動生成におけるこのような成果により、コミュニティ間の連携が強まり始め、それぞれの保有するデータをより有効に共有することが可能となり、他のコミュニティで使用されているメタデータ作成方法に対して各コミュニティがよりオープンになりつつあります。

将来の方向性

Linked Data へと向かう動きは、商業、情報及び文化遺産の分野におけるメタデータの使われ方に多大な影響を及ぼしました。Linked Data システムを動かす技術により、情報共有が容易になり、それによってこの動向の根底を成すデータのオープン化と相関性の確保への取組が促進されました。

Linked Data の考え方は、メタデータとそれを使用するシステムに対する見方にも影響を与えました。Linked Data では、出自ごとに定義された境界のあるデータセットではなく、全体としてのグラフが重要視され、データセット間の差異を維持することよりも結び付きに価値を置くことが主流となりつつあります。セマンティックウェブが前提とするオープンな世界においては、情報の欠如は不確実な結論につながるかと仮定しています（一方、閉じた世界においては、確実性は増しますが、すべて既知の関連情報に依存した結論が導かれます）。この仮定からの帰結として、出自の異なるデータをより密接に連携させ、その結び付きから生じる価値や新たな知識の構築を望む声が出てきました。この文脈におけるメタデータの作成において重視されるのは、データ入力フォームを埋めるのではなく、既存の事物間のリンクを構築することです。今後も、関連付けの深化と知識グラフの大規模化の傾向が続いていくことが予想されます。

21 世紀現在、技術の進展とともに、膨大な量の情報が生成、処理、転送され、メタデータを自動的に作成及び維持管理する手法への依存が高まっています。システムについていえば、情報の生成と処理において純粋に高機能化しています。ある書籍の著者名称の優先形の検索や、テキスト形式で書かれた日付の標準的な形式への正規化等、かつては人手により行われていた作業が、今ではシステムにより自動化でき、貴重な人間の労力を、本当に必要なことに向けることができるようになっていきます。ソフトウェアツールには、ユーザーが見るべきものをあらかじめ決めるのではなく、より多くの情報を表示するとともに、どの情報に価値があり役立つかをユーザー自身が選べるようにすることが期待されるでしょう。ユーザーとの相互作用が必要となるかもしれないデータ量の著しい増加に対処するため、相反する情報が存在する場所を明示し、表示される情報の由来について透明性を確保するようにユーザーインターフェースの設計が改良されると考えられます。

最後に、メタデータのオープン化と相互連携の文化の出現により、「信頼できる」または「良い」メタデータの定義が再定義されています。インターネットによって、かつては社会から重要視されていなかった意見を持つ人々がそれぞれの知識を共有することができる新しい機会を得ることができるようになりました。あらゆるトピックに関して、情報に通じた愛好家のコミュニティがオンライン上に存在し、これらの個人は、コンテンツ管理を担ってはいるものの主題の専門知識を持たない組織や団体よりもはるかに有用なメタデータを提供できることが多いです。このプロセスは、「クラウドソーシング」や「trusting the wisdom of the crowd（和訳：集合知への信頼）」としても知られます。高度なシステムは、ユーザーにとって有用となるような方法で、ユーザーが生成したメタデータと従来どおりの情報源からのメタデータとを組み合わせることができます。メタデータの緊密な統合を促進するために、コミュニティがすでに取り組んでいる活動を基に、専門家ではない人々のためにシステムのインターフェースが非常に合理的で使い勝手が良くなるよう設計されることが期待されます。また、ユーザーの作業に対価を払う、より進化した報酬構造が現れることも予想されます。

メタデータがオープンでつながったモデルに移行するには時間がかかることが予想されます。あるシステムが、統制された、従来からの信頼できる方法でメタデータへアクセスすることや、組織が人間による重要な情報資源の管理を望むことには、実際的で正当な理由があります。これらは今後も、全体的なメタデータの趨勢の一部であり続けるでしょう。それでもなお、メタデータを共有し、他の情報源からの関連語彙を採り入れることの利点は無視することはできません。メタデータのオープン化により、異なるコミュニティが相互に学び合い、連携を深め合うような魅力的な未来が訪れることを願っています。

付録：資料

メタデータ標準と語彙

下記以外のメタデータ標準及びそれらに関する仕様とその解析、次の URL で参照できます：<http://jenriley.com/metadatamap/>

- BIBFRAME <http://www.loc.gov/bibframe/>
- CDWA http://www.getty.edu/research/publications/electronic_publications/cdwa/
- CIDOC CRM <http://www.cidoc-crm.org/>
- DCMI 抽象モデル <http://dublincore.org/documents/abstract-model/>
- DDI <http://www.ddialliance.org/>
- ダブリンコア <http://dublincore.org/>
- EAD <http://www.loc.gov/ead/>
- Exif http://www.cipa.jp/std/documents/e/DC-008-2012_E.pdf
- FOAF <http://xmlns.com/foaf/spec/>
- MARC <http://www.loc.gov/marc/>
- MEI <http://music-encoding.org/>
- MODS <http://www.loc.gov/standards/mods/>
- ONIX for Books <http://www.editeur.org/93/Release-3.0-Downloads/>
- OWL <http://www.w3.org/TR/owl-features/>; <http://www.w3.org/TR/2012/REC-owl2-primer-20121211/>
- PREMIS <http://www.loc.gov/standards/premis/>
- Schema.org <http://schema.org/>
- アプリケーションプロファイルのためのシンガポールフレームワーク <http://dublincore.org/documents/2008/01/14/singapore-framework/>
- SKOS <http://www.w3.org/2004/02/skos/>
- TEI <http://www.tei-c.org/>
- VRA Core <http://www.loc.gov/standards/vracore/>

XML 標準及びXML に関する情報

- OAI-PMH <https://www.openarchives.org/pmh/>
- ResourceSync <http://www.openarchives.org/rs/1.0/resourcesync>
- W3C XML Activity <http://www.w3.org/XML/>
- XPath <http://www.w3.org/TR/xpath/>
- XSLT <http://www.w3.org/TR/xslt20/>
- XQuery <http://www.w3.org/TR/xquery/>

RDF 標準及びRDF に関する情報

- RDF Concepts <http://www.w3.org/TR/rdf11-concepts/>

- RDF Primer <http://www.w3.org/TR/rdf11-primer/>
- RDFS <http://www.w3.org/TR/rdf-schema/>
- SPARQL <http://www.w3.org/TR/sparql11-query/>
- W3C RDF Activity <http://www.w3.org/RDF/>

RDF シリアライゼーション

- JSON-LD <http://www.w3.org/TR/json-ld/>
- Microformats <http://microformats.org/>
- N-Triples <http://www.w3.org/TR/n-triples/>
- RDF/XML <http://www.w3.org/TR/rdf-syntax-grammar/>
- RDFa <http://www.w3.org/TR/xhtml1-rdfa-primer/>
- Turtle <http://www.w3.org/TR/turtle/>

Linked Data に関する情報

- Datahub (Linked Data 情報を多数含むデータレジストリ) <http://datahub.io/>
- Linked Data クラウド (Linked Data を共有する際との視覚化) <http://lod-cloud.net/>
- Linked Data—設計上の課題 (ティム・バーナーズ=リー) <http://www.w3.org/DesignIssues/LinkedData.html>
- Linked Data : グローバルデータ空間のウェブの進化 (Linked Data 実装のためのベストプラクティスのハンドブック) <http://linkeddatatobook.com/book>
- LOD Stats (約 10,000 個の Linked Data データセットの検索サイト) <http://stats.lod2.eu/>
- OCLC Linked Data 戦略 <http://www.oclc.org/data.en.html>
- OCLC WorldCat Linked Data 語彙 <https://www.oclc.org/developer/develop/LinkedData/worldcat-vocabulary.en.html>
- W3C Data Activity <http://www.w3.org/2013/data/>

注目すべきプロジェクト

- BIBFRAME <http://www.loc.gov/bibframe/>
図書館コミュニティで使用するために、書誌及び典拠の情報に焦点を当てた、Linked Data に適した語彙。米国議会図書館により、開発と維持管理が行われています。
- DBpedia <http://www.dbpedia.org>
Wikipedia 記事内の非構造化テキストから、構造化されたデータを抽出し、他の用途のために提供するコミュニティベースのプロジェクト。
- Digital Public Library of America <http://dp.la>
米国の図書館、文書館及び博物館からのコンテンツを検索するポータルサイトであり、そのコンテンツに関するデータのオープンな再利用を可能にするプラットフォーム。
- Europeana <http://www.europeana.eu>
欧州の図書館、文書館及び博物館からのコンテンツを検索するポータルサイ

トであり、そのコンテンツに関するデータのオープンな再利用を可能にするプラットフォーム。

- **LC Linked Data Service** <http://id.loc.gov>
米国議会図書館が維持管理している多数の統制語彙の **Linked Data** へのアクセスを提供する同図書館のサービス。このサービスは、人間のユーザーに対してはウェブ上のユーザーインターフェース経由で、ソフトウェアアプリケーションに対してはコンテンツネゴシエーションにより利用可能です。
- **Linked Data for Libraries (LD4L), LD4L Labs, and Linked Data for Production (LD4P)** <http://www.ld4l.org/>
コーネル大学図書館、Harvard Library Innovation Lab、スタンフォード大学図書館等の一連の共同プロジェクトであり、図書館技術サービスのワークフローの進化として、既存のオントロジー及びソフトウェアを使用し、複数の図書館の情報源のデータを統合する実用的な **Linked Data** 実装を作成しています。
- **RDF Data Shapes Working Group** <https://www.w3.org/2014/data-shapes/charter>
RDF グラフの、構造的な制約の定義方法と検証方法を定義する W3C ワーキンググループ。
- **Schema Bib Extend Community Group (SchemaBibEx)**
<https://www.w3.org/community/schemabibex/>
書誌情報を使用する schema.org の拡張を進めている、World Wide Web Consortium のコミュニティ。
- **Wikidata** <http://www.wikidata.org>
物事、概念及び出来事に関する構造化されたデータを記録するウィキメディア財団から発足したプロジェクト。Wikidata には誰でも参加することができます。Wikidata は、Wikipedia 等他のウィキメディアプロジェクトへ未加工データを提供し、それらのデータを外部のアプリケーションで利用できるようにしています。