慶應義塾大学学術情報リポジトリ
Keio Associated Repository of Academic resouces

Title	書誌学者の立場から見たボーダレスなデータ利活用のための情報組織化
Sub Title	
Author	安形, 麻理(Agata, Mari)
Publisher	慶應義塾大学デジタルメディア・コンテンツ統合研究センター
Publication year	2019
Jtitle	慶應義塾大学DMC紀要 (DMC review Keio University). Vol.6, No.1 (2019. 3) ,p.47- 52
JaLC DOI	
Abstract	
Notes	特集 DMC研究センターシンポジウム第8回「デジタル知の文化的普及と深化に向けて」メタデータ再考 開催日時 : 2018年11月20日 (火) 14:00 ~ 17:30 開催場所 : 慶應義塾大学日吉キャンパス西別館1 話題提供
Genre	Departmental Bulletin Paper
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=KO32002001-00000006-0047

慶應義塾大学学術情報リポジトリ(KOARA)に掲載されているコンテンツの著作権は、それぞれの著作者、学会または出版社/発行者に帰属し、その権利は著作権法によって保護されています。引用にあたっては、著作権法を遵守してご利用ください。

The copyrights of content available on the KeiO Associated Repository of Academic resources (KOARA) belong to the respective authors, academic societies, or publishers/issuers, and these rights are protected by the Japanese Copyright Act. When quoting the content, please follow the Japanese copyright act.

話題提供
「書誌学者の立場から見たボーダレスなデータ利活用のための情報組織化」

安形 麻理
（慶應義塾大学文学部准教授）

ご紹介いただきました安形と申します。メタデータということでしたら、うちの専攻のもっとふさわしい先生も会場にいらしているのですが、私は本日は書誌学者としての立場からということで、先ほど冒頭にありました原先生のお話とだいぶ重なるところもありますが、研究に関わってきた中でのメタデータについて思うところを話題提供という形でお話をさせていただきます。

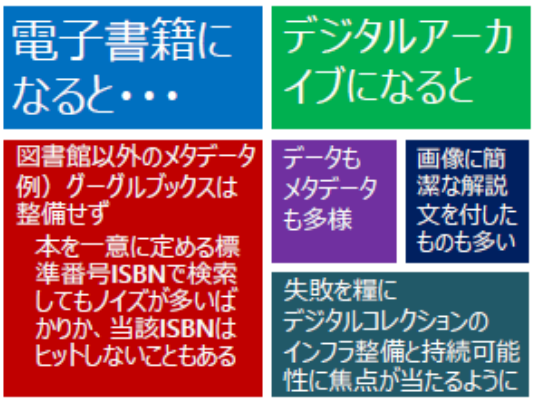
最初に図書館におけるメタデータを考えみると、人類の知的な文化財を扱う機関としては、原先生のお話にもありましたように、非常に早くから標準化が進んできた分野であるといえます。基本的には近現代の刊行物、複製品がたくさんある本を扱ってきたため、目録の標準化が非常にしやすかったということもあって進んできたわけです。

最近はインターネット情報資源など、図書館が物として所有していない情報資源についても記述しなくてはいけないという状況に対応するために、また一方では博物館や文書館とも、デジタルなデータであれば、お互いのメタデータの共有が進むだろうという期待から、ダブリン・コアといった共通の枠組が出てきているわけです。

ただし、一般のいわゆる複製物としての印刷本でしたら何がメタデータかは明確ですが、私が研究対象にしているような 15 世紀の印刷物（incunabula、インキュナブラ）、あるいは会場後方にも 15 世紀の活版印刷物やその前に作られた手書きの写本の

リーフが展示されていますけれど、ああいう資料についてのメタデータをとる際には、メタデータなのかデータなのかという境界はかなり曖昧だといえます。つまり、その物が何であるかということを同定して記述すること自体が研究であるという意味で、メタデータとデータの境界が必ずしもはっきりしないところがあるかと思います。ある場合にはメタデータ、ある場合にはデータともなりうるわけです。

欧米の国立図書館などでは、書誌データを Linked Open Data（LOD）の形で提供するということが行われていて、私も昨年から共同研究者と一緒にインキュナブラの LOD モデルを考えてみたらどうだろうということで取り組んでいます。そうする中でやはりメタデータをどうとるかというのが非常に重要な課題であると認識しました。そちらはまだ検討している途中ですので、今日はちょっと違ったお話をしたいと思います。

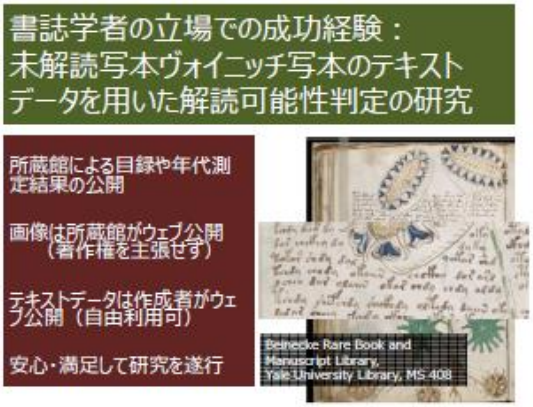


目録は非常に標準化が進んできたと申し上げたのですが、本が電子書籍化されたり、デジタル化された場合になると、図書館以外のアクターもいろいろ関わってきます。たとえばグーグルブックスなどでもたくさんの本がデジタル化された形で手に入るわけですが、実はグーグルブックスでは、紙の本を一意に定めるはずの ISBN がそのデジタル版ときちんと対応付けられておりません。検索で一つの ISBN を入れても非常にたくさんのリストが出てきます。なおか

つ、その中には該当する ISBN が 1 件も含まれていないということもあります。グーグルブックスではそういった形でのメタデータを整備しようという意識はないということが見てとれます。これは先月行われました第 66 回日本図書館情報学会研究大会の「米国における電子書籍化の現状」というポスター発表で知ったことなのですが、そういった状況です。

また、デジタルアーカイブも様々な機関が様々な方針の下で、それぞれのメタデータを付けているわけです。データそのものも、そこについているメタデータも多様で、目録と違い標準化は進んでいません。メタ

データとして書誌・所蔵情報やフォリオ(ページ）番号、撮影の詳細などを付したものもあれば、本一冊分の画像に簡潔な解説文を付してメタデータにあたるものとして提供されているものも多く見られます。



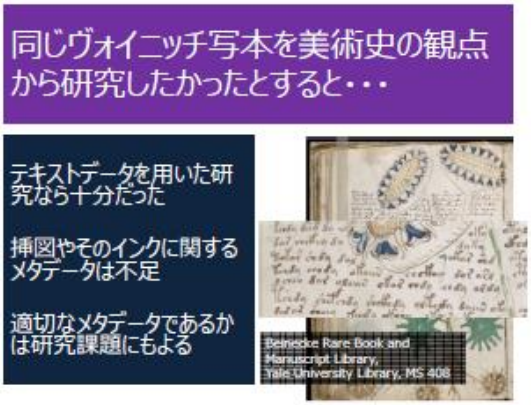
ここで、メタデータと研究の関わりの例を紹介します。スライド 4 に挙げましたのは、ヴォイニッチ写本という、発見から 100 年程経っているのにまだ誰も読むことができていない、オカルト界隈でも大人気の謎の写本です。読めそうなのに読めない文字で書かれているのです。私と共同研究者がこの写本を真面目に研究してみようと取り組んだ 10 年程前の時点では、16 世紀の写本もしくは 20 世紀の偽物というようなメタデータがついておりましたが、その後年代測定が行われて 15 世紀ということは間違いないということになり、メタデータも変わったりしています。

我々はテキストデータをページごとにクラスタリングした結果と、ヴォイニッチ写本はほとんどのページに挿絵がついているので、その挿絵の種類が一致するかどうかを見ることで、文書が一貫した構造をもち内容に意味がありそうかどうかを判定しようという研究を行いました。

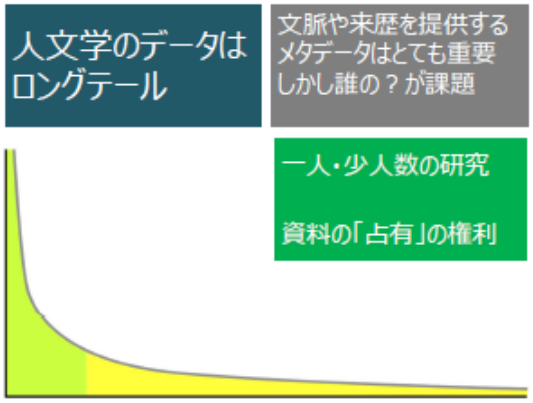
この研究では、比較的詳しい目録データが所蔵館であるイェール大学のバイネッケ図書館の OPAC から入手でき、画像は所蔵館のウェブサイトから公開されていて、なおかつそこにはもう著作権保護期間が満了している資料なので、イェール大学は何の権利も主張しないし、許諾も拒否もしないということが明示されています。権利情報に関するメタデータもきちんと示されています。なおかつ翻刻されたテキストデータも、個人的に翻刻された研究者がウェブで自由利用可として公開されていました。

この写本については、権利関係も非常にクリアということで、安心してスムーズに研究を行うことができました。やはり研究者が知財の法律の専門家である必要があってはならないと思うわけですが、そういった意味でのメタデータがわかりやすく整備されていました。

しかし、この同じヴォイニッチ写本をたとえば美術の観点から研究したかったとしますと、ヴォイニッチ写本の各ページに示されている挿図やそこに使われている顔料、インクといったものに関するメタデータはありません。

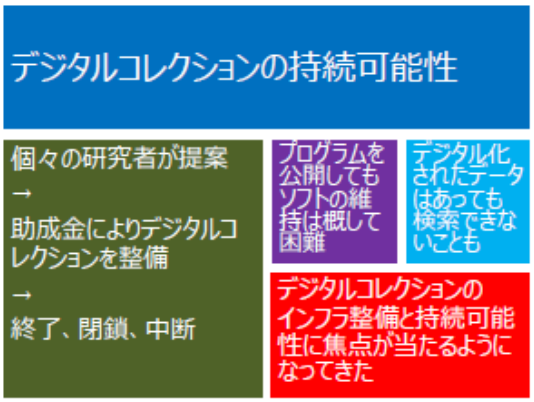


本日の会場の展示コーナーにはたくさんの顔料が並んでいるわけですが、ああいった情報は付与されていません。同じデータとメタデータのセットであっても、研究者が持っている課題によっては適切な十分なメタデータであったり、そうでなかったりするといえるでしょう。



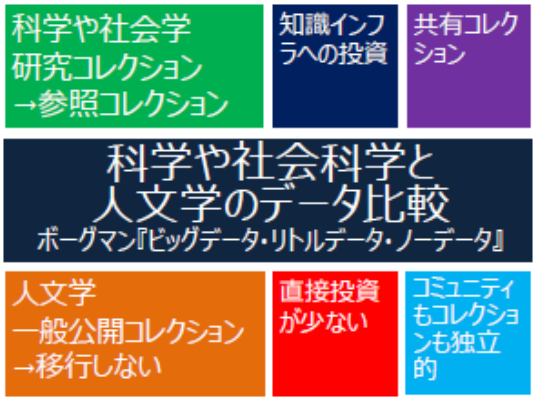
そういう意味でいいますと、人文学のデータ自体がいわゆるロングテールにあたるかと思います。宇宙科学やタンパク質の構造といったビッグデータと関連が深い研究領域もあるわけですが、人文学の研究というのは、しばしば１人、あるいは少人数で行われたり、あるいはその資料を持っている、アクセスできるということ自体が価値であったりすることが珍しくなかったわけです。

それぞれの研究者が様々な課題をもって研究を遂行していくため、そこで必要とされるメタデータが非常に独立的で多様であるというところが一つの特徴であるといえます。



また先ほど原先生が最後の方でおっしゃっていらしたように、いろいろなデジタルコレクション、特に人文学のデジタルコレクションを公開する際には、どちらかというと研究者の側でそのデジタルコレクションの構築がいかに大事かという提案書を書き、助成金を獲得して実現させる形が主流です。待っていてもお金は下りてきません。助成金でデジタルコレクションが整備されても、それがやがて中断、閉鎖、あるいは終了してしまうと、大規模なプロジェクトであっても、データだけは残っているがもはや検索はできないなど、どう使ってよいのか分からないようなデータというものが残されてしまっているわけです。

あるいはプログラムは公開されているものの、そのプログラムを実際にデータに適用し動かしていくようなところでの維持は非常に困難だということも出てきます。欧米の大規模なプロジェクトでもそうした例が多く出てきていることを受けて、最近では、助成機関側がデジタルコレクションのインフラ整備と持続可能性にも焦点を当てるようになってきたというのは、明るい兆候であるといえるでしょう。



最後に、人文学の特徴として『ビッグデータ・リトルデータ・ノーデータ』(勁草書房, 2017）という本にあった比較をスライドに挙げました。これは全体としてビッグデータにしろ、人文学のようなリトルデータにしろ、きちんとしたメタデータやそれを提供する知識インフラが整備されていないかぎりノーデータになってしまう、つまり、使えない、アクセスできない、誰も知らないデータになってしまうということを述べているものです。

人文学のコレクションは、個別の課題、個別の研究課題というものに密接に対応するために、なかなかそのコミュニティ全体で参照して使うような参照コレクションに移行しにくいという特徴があります。直接投資が少なく、個別の研究課題があるからこそです。そうしたことを考えた場合に、人文学のメタデータのあり方を一口にまとめるのは難しいといえます。本当に話題提供という形で、結論も出ていないのですが、私の話は以上です。どうもありがとうございました。

安形 麻理（あがた まり）

慶應義塾大学文学部教授（書誌学、デジタル書物学、図書館・情報学)。学部生時代から貴重書デジタル化プロジェクトに様々な立場で関わってきた。主著に『デジタル書物学事始め：グーテンベルク聖書とその周辺』(勉誠出版, 2010 年)、「利用者による特殊コレクション資料の撮影の許可：北米の研究図書館における動向」『図書館は市民と本・情報をむすぶ』(勁草書房, 2015, p.

52-60)、「科学は贋作を判定できるか」『書物學』no. 14, 2018, p. 34-41.等がある。