

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	<a href="https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=KO32002001-00000006-0047">https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=KO32002001-00000006-0047</a>

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

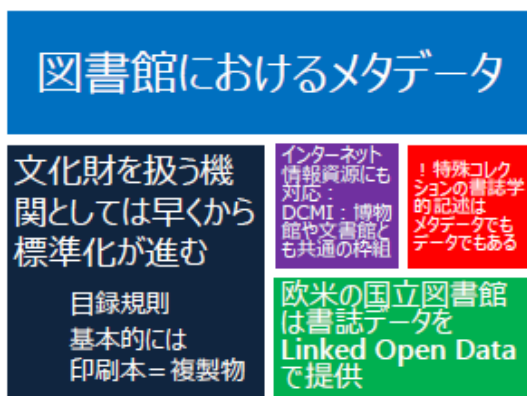
## 「書誌学者の立場から見たボーダレスな データ利活用のための情報組織化」

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なら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と同定して記述  
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。

電子書籍に なると・・・	デジタルアーカイブになると
図書館以外のメタデータ 例) グーグルブックスは 整備せず 本を一意に定める標準 番号ISBNで検索 してもノイズが多いば かりか、当該ISBNは ヒットしないこともある	データも メタデータ も多様 画像に簡 潔な解説 文を付した ものも多い 失敗を糧に デジタルコレクションの インフラ整備と持続可能 性に焦点が当たると

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、

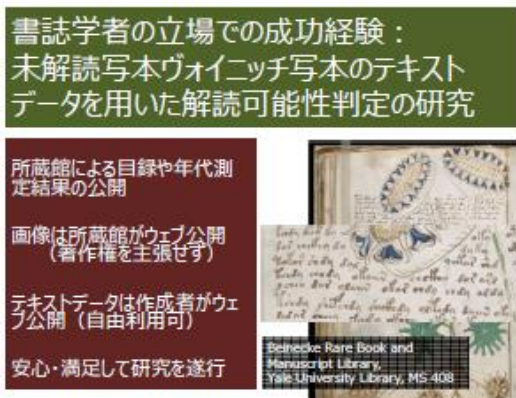
デジタル化された場合になると、図書館以  
外のアクターもいろいろ関わってきます。  
たとえばグーグルブックスなどでもたく  
さんの本がデジタル化された形で手に入る  
わけですが、実はグーグルブックスでは、紙  
の本を一意に定めるはずの **ISBN** がそのデ  
ジタル版ときちんと対応付けられており  
ません。検索で一つの **ISBN** を入れても非常  
にたくさんのリストが出てきます。なおか



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、オカルト界限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エール大学のバイネック図書館の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、イエ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なく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。

## 同じヴォイニッチ写本を美術史の観点から研究したかったとすると・・・

テキストデータを用いた研究なら十分だった

挿図やそのインクに関するメタデータは不足

適切なメタデータであるかは研究課題にもよる



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。

## 人文学のデータはロングテール

文脈や来歴を提供するメタデータはとても重要しかし誰の？が課題



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。

## デジタルコレクションの持続可能性

個々の研究者が提案  
→ 助成金によりデジタルコレクションを整備

プログラムを公開してもソフトの維持は概して困難

デジタル化されたデータはあっても検索できないことも

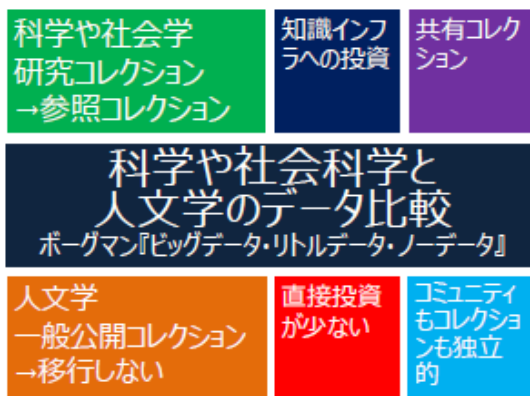
→ 終了、閉鎖、中断

デジタルコレクションのインフラ整備と持続可能性に焦点が当たるようになってきた

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る、人文学のようなリトルデータにしる、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。