

人文学資料マイクロコンテンツの実世界との双方向結合とデータポータル「edomi」

鈴木 親彦・北本 朝展 (ROIS-DS 人文学オープンデータ共同利用/国立情報学研究所)

概要：本論では、デジタル化された人文学資料から抽出した「人文学資料マイクロコンテンツ」と、実世界の実体に関する情報を双方向に参照可能とする方法を提案する。まず文書に関する情報と実世界の実体に関する情報を表す「文書空間識別子」と「実体空間識別子」を定義し、識別子間の結合を行うことで文書と実体の関係を記述する。そして、江戸に関する様々な情報を提供するデータポータルである「edomi」を事例として、観光・買物に関するマイクロコンテンツを公開する試みを示す。
キーワード：マイクロコンテンツ, IIF Curation Platform, 文書空間, 実体空間, データポータル

Two-way linking between micro-contents from humanities documents and the real world to create a data portal “edomi”

Chikahiko Suzuki / Asanobu Kitamoto (ROIS-DS Center for Open Data in the Humanities / National Institute of Informatics)

Abstract: This paper proposes a method for creating bidirectional links between "micro-contents from humanities documents" and "entities in the real world." We first define a "document-space identifier" and an "entity-space identifier" to represent the information of documents and entities and describe the relationship between documents and entities by "two-way linking". We then show a case study to publish micro-contents about tourism and shopping using two-way linking on the data portal "edomi" to deliver various information about Edo.

Keywords: Micro-contents, IIF Curation Platform, Document-space, Entity-space, Data portal

1. はじめに

本論では、IIF Curation Platform を活用してデジタル化された人文学資料の一部を抽出・整理した「人文学資料マイクロコンテンツ」を、実世界の実体に関する情報と紐づけて活用するための方法論を検討する。その際に鍵となるのは、実世界に由来する実体の情報と資料から取り出された情報とを結合し、実世界側を起点にマイクロコンテンツを発見できる仕組みを作ることにある。この問題は、資料由来のメタデータに対応する「文書空間識別子」と、実世界の実体に対応する「実体空間識別子」を設定し、双方向に参照可能な形で紐づけることにより解決可能となる。具体的な例として、我々がマイクロコンテンツとして公開してきた「江戸買物案内」「江戸観光案内」[1][2]と実世界の識別子を結合し、江戸に関する様々な情報を提供するデータポータル「edomi」[3]で公開するまでの実践を示す。

2. 人文学資料マイクロコンテンツと IIF Curation Platform

「人文学資料マイクロコンテンツ」(以下、マイクロコンテンツ)とは、電子出版の普及に伴って登場した出版物の一部を自由に利用するため

の枠組みである「マイクロコンテンツ」を発想源とし、それをデジタル化された人文学資料に応用した概念である。国立国会図書館や国文学研究資料館が進める大規模なデジタル化をはじめとして、多くの組織が人文学資料のデジタル環境での公開を進めている。また、ジャパンサーチの様に、横断的に検索を可能にする環境も整ってきた。このようにデジタル活用が可能となった人文学資料の利点を活かし、特定の目的で一部の情報を取り出し、資料を横断して情報の活用を可能にしたものが、人文学資料マイクロコンテンツである。

我々は、特に絵画や挿絵などの非文字情報に注目し、マイクロコンテンツ化の方法を検討してきた。特に IIF (International Image Interoperability Framework) 形式の公開資料に対しては、IIF Curation Platform (ICP) [4][5]を用いて画像を収集し、メタデータを付与して整理する方法でマイクロコンテンツ化を進めてきた。その成果として、江戸在住商人の情報を集めた『江戸買物独案内』から商人単位で版面を抽出した「江戸買物案内」や、江戸時代に出版された『江戸名所図会』『江戸名所記』をはじめとする名所記・名所案内から挿絵を抽出した「江戸観光案内」などを公開済みである[6]。

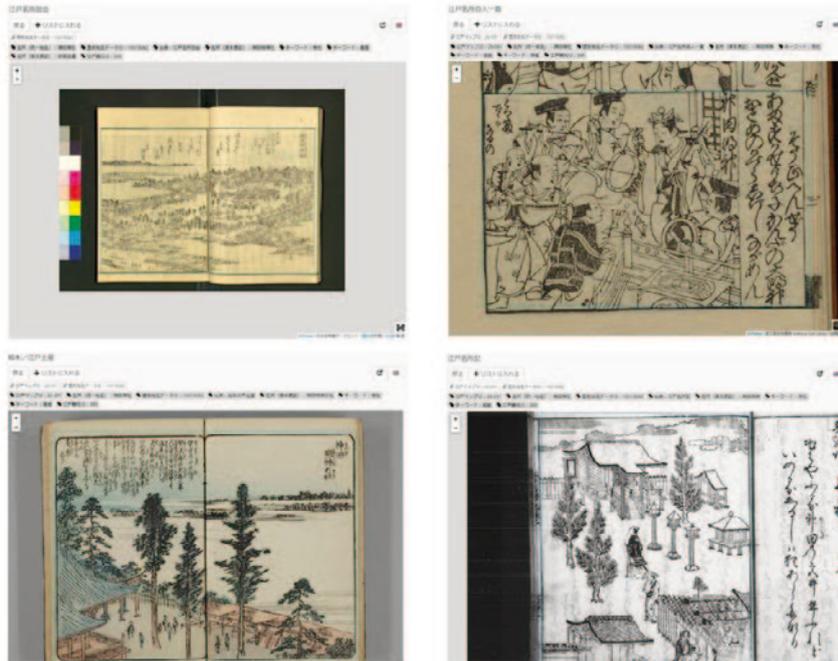


図 1: 「江戸観光案内」の神田神社と挿絵に付されている名称例
 左上『江戸名所図会』「神田明神社」、右上『江戸名所百人一首』「神田明神」
 左下『絵本江戸土産』「神田明神の社」、右下『江戸名所記』「神田明神」(挿絵中には名称無し、
 解説文から確認可能)

これまで進めてきた上記のマイクロコンテンツ化では、非文字情報としての挿絵や版面画像の抽出、画像から取り出せる様々な情報を扱ってきた。つまり非文字資料に内在する情報のメタデータとしての整理と、それらの資料横断的な検索・活用が主な目的であった。しかし次の段階として、マイクロコンテンツを実世界の分析や活動につなげていくためには、マイクロコンテンツを実世界と紐づける仕組みが必要なることに気づく。

例えば、「江戸観光案内」に出てくる名所は現代のどの場所に対応するのか、あるいは「江戸買物案内」に出てくる業種は現在の「日本標準産業分類」のどの業種に対応するのか。マイクロコンテンツを資料の外側に存在する実体や体系に紐づけることで、現代の世界と過去の情報を連続的に活用することが可能となる。そこで、資料の内外を接続する手法である「文書空間と実体空間の双方向結合」について第3章で、方法の実践段階での課題を第4章・第5章で述べる。さらに「文書空間と実体空間の双方向結合」を利用して、実際にマイクロコンテンツを公開した成果であるデータポータル「edomi」についても第6章で紹介する。

3. 文書空間と実体空間の双方向結合

「江戸買物案内」や「江戸観光案内」では、IIIF形式で公開された人文学資料に対して、IIIF Curation Viewer上で関心領域を指定し、メタデー

タとして翻刻テキストや分類語彙などを付与する作業を進めてきた。こうして生成したCuration形式のマイクロコンテンツには、IIIF形式で公開されている資料のManifest URI、そして各Manifestに含まれるCanvasを単位とするCanvas IDが識別子として埋め込まれている。収集した個々の画像領域の座標情報を含むCanvas IDに対し、画像領域を記述するメタデータを付与する構造となっている。このようなCanvas IDは、マイクロコンテンツに対応する文書上の位置や領域を一意に示すものであることから、これを「文書空間識別子」と呼ぶ。

一方、そのマイクロコンテンツが指し示す実世界中の「実体」については、また別の識別子が必要である。例えば「江戸観光案内」には、『江戸名所図会』記載の「神田明神社」[7]、『絵本江戸土産』記載の「神田明神の社」[8]、『江戸名所記』記載の「神田明神」の挿絵が登録されている[9](図1)。資料によって表記には揺れがあるものの、それらが指し示すのは、いずれも東京都千代田区外神田に実在する「神田神社」だと判断できる。これが実世界におけるそれら名所の「実体」となる。このような、実世界に存在する「実体」に対しても識別子をあらかじめ付与しておけば、様々な資料への出現に「神田神社」識別子を付与することで、すべての出現を統一的に扱うことができる。これを「実体空間識別子」と呼ぶ。

このように、文書に関する集合を「文書空間」、実体に関する集合を「実体空間」と呼び、それぞれの空間を構成する要素に識別子を付与することとする。文書空間識別子とは、どの文書のどこの部分かを一意に特定する識別子であり、その中に何が書いてあるかを記述する構造化データを持つ。一方、実体空間識別子とは、時間、空間、人物、イベントなどの実体を特定する識別子であり、それらの実体の性質に関する構造化データを持つ。そして、文書空間識別子と実体空間識別子を結合することで、「ある文書が実世界のある実体に言及している」という関係を表現できると考えられる。ここで両者の結合方法として、文書側の識別子から実体側の識別子を紐づける、あるいは実体側の識別子から文書側の識別子を紐づける、という2つの方向が考えられる。前者をエンティティリンキング、後者をエビデンスハンティングと呼ぶ。なお、前者は自然言語処理や知識表現などの分野で一般的に使われる用語であるが、後者は我々による造語である。

まずエンティティリンキングとは、文書中に出現する実体に着目し、文書空間側から実体空間側に対するリンクを定義する場合を指す。例えば、江戸観光案内において、ある旅行記の挿絵に名所の地名が書いてあれば、それを根拠としてメタデータに実体空間識別子を入力する。これにより、文書空間と実体空間を識別子でリンクすることができる。次にエビデンスハンティングとは、実体に関する何らかの証拠を文書空間中で探索し、実体空間側から文書空間側に対するリンクを定義する場合を指す。例えばある名所に関する記述が様々な本に記述されており、それらを集約して名所を復元したい場合、実体に関するメタデータとして文書空間識別子を入力する。この方法でも、文書空間と実体空間を識別子でリンクできる。

このように、文書空間側からリンクする場合と、実体空間側からリンクする場合の、いずれの方法であっても文書と実世界の紐づけは可能であり、これが文書空間と実体空間の双方向結合に相当する。このような双方向結合を行うためには、両方の空間における識別子を定める必要がある。以下では特に問題となる実体空間側の識別子の定義とそれにまつわる曖昧性解消の問題を論じる。

4. 実体空間識別子と曖昧性解消

実体空間識別子を定義するには、実世界に存在するエンティティ(実体)を特定し、それに識別子とメタデータを付与する必要がある。エンティティを実際にどう定義するかには唯一の解があるわけではないが、実世界で区別することに価値がある情報単位を特定し、その粒度を意識しながら

リスト化していくことになる。そこで「江戸観光案内」では名所の現在における位置を特定する江戸観光ID、「江戸買物案内」では商人および店の位置を特定する江戸商人IDという実体空間識別子を定義した。

ここで重要となるのが曖昧性解消の問題、すなわち文書空間中に出現する表現の曖昧性を、実体空間との結合によって解消する問題である。第一の問題は表記揺れである。「江戸観光案内」では、「神田神社」は「神田明神社」「神田明神の社」「神田明神」などの様々な表記で言及されるが、これらは同じ実体を指している。また「江戸買物案内」では、業種ごとに商人がまとめて記載されるため、多角経営を行う商人はそれぞれの業種ごとに複数回出現するが、実体としては同じ商人である。つまり、一つの実体が様々な表記で言及される場合に、それらを単一の識別子と紐づけることが作業の課題となる。こうした作業は、一般に名寄せと言われる。名寄せは複数の資料に出現する人名や組織名、文献名などを統一する作業として多くの分野で使われており、本研究の課題もそれらと基本的には同様である。

第二の問題は、複数の実体が同じ表記で重なる問題の解消である。「江戸観光案内」では、「氷川明神」のように、表記としては同一であっても実際には複数の実体に対応することがある。また「江戸買物案内」でも、同姓(同屋号)同名だが実体としては異なる商人が複数人存在することがある。このような場合に実体を区別するには、表記以外の手がかりを基準として区別する必要がある。例えば、実体に関する表記の周辺に付加的な情報があれば、それをういて曖昧性を解消することができる。特に有用なのが地理情報である。例えば同名の氷川明神であっても、元資料の画像や説明文に「麻布」や「大宮」といった地名が掲載されていれば、それを基に異同を特定できる。また商人の場合も居所が異なれば別人であると判断できる。このように地理情報があれば、それは曖昧性解消にとって有用な情報となる。逆にそのような明確な手がかりがない場合は、表記が出現する文脈を読み込んで実体を特定することで、曖昧性解消を進めることになる。こうした作業は、名寄せに対して名分けと言われることがあるが、こちらも多くの分野と共通した課題である。

こうした作業を進めることで特定できた実体のリストに対して、実体空間識別子を付与するとともに、実体の呼称(代表的な名称)など各種のメタデータを加え、実体空間のデータベースを整

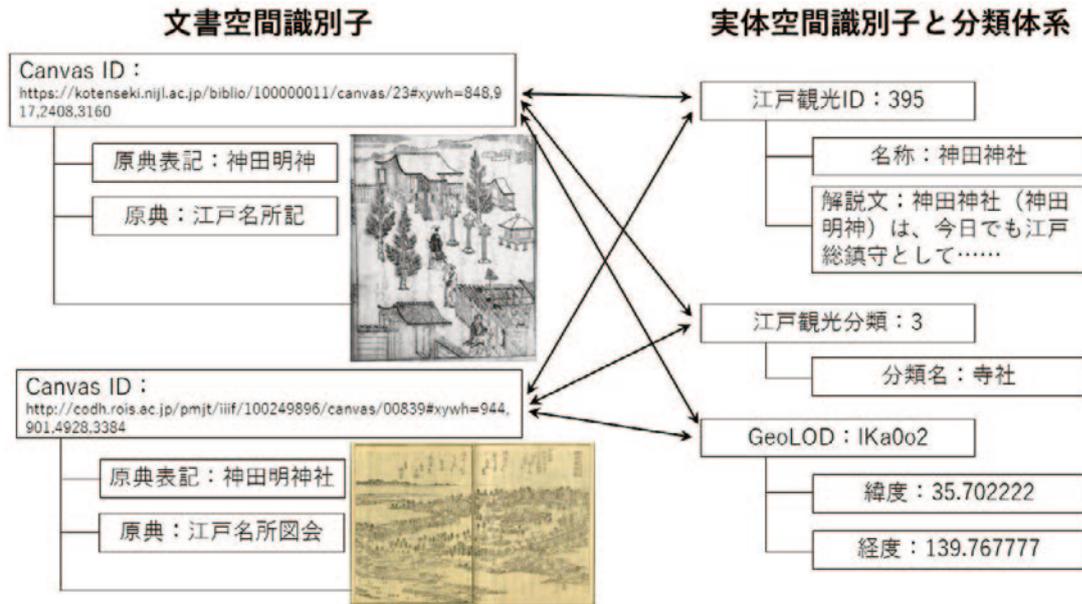


図 2：文書空間識別子と実体空間識別子の双方向結合
ICP で切り取った画像部分を示す文書空間識別子である Canvas ID と江戸名所の実態をリスト化した江戸観光 ID・分類体系とを紐づけている

備した。その成果として、江戸観光 ID、江戸商人 ID のデータベースが構築できた。

5. 文書空間と実体空間の結合

このように構築した実体空間識別子を活用し、文書空間と実体空間の結合を進めた。まず ICP を活用してマイクロコンテンツを作成すると、文書空間識別子として Canvas ID を自動的に付与することができる。そこでこの識別子に対して実体空間識別子を付与するエンティティリンキングの考え方で双方向結合を進める。

IIIF Curation Viewer では、指定した画像領域にメタデータを付与することができるため、実体空間識別子を付与するためのメタデータ項目を定義した。そして、資料の内容を解釈しながら曖昧性を解消し、一意に特定した実体空間識別子を入力した。入力する識別子としては、江戸観光 ID と江戸商人 ID だけでなく、GeoLOD[10][11]が提供する地名識別子 GeoLOD ID も対象とした。GeoLOD には江戸マップβ版[12]の地名も登録されているため、GeoLOD を介して古地図の地名など多様な空間情報と結合することが可能である。

またここで、厳密には実体空間に含まれないが同様の扱いを行う「分類体系 (オントロジー)」にも言及しておきたい。分類体系とは、エンティティのある種の属性に着目してグルーピングするためのものであり、分類項目にはコードと呼ばれる識別子が割り当てられる。こうした分類体系も実世界側のニーズに基づいて定義されている

ことから、文書空間識別子を分類体系と紐づけることも有用である。したがってメタデータには分類体系識別子を付与する項目を定め、そこに識別子を入力していった。具体的には総務省による「日本標準産業分類」を「江戸買物案内」のメタデータに入力することで、資料内の分類体系を現代の分類体系と接続し、過去の商人の業種を現在の視点から把握できるようにした。

このように ICP を用いて作成したマイクロコンテンツ「江戸買物案内」や「江戸観光案内」にメタデータを付与することで、文書空間識別子と実体空間識別子を結合した。その結合実践を図 2 に示す。

さて、この成果を文書空間側から閲覧し検索するための仕組みが IIIF Curation Finder である。IIIF Curation Finder は文書の構造を保持しつつ、そこに付与されたメタデータを閲覧する仕組みであり、これまでの作業で付与した実体空間識別子も文字列としては一覧できる。しかし IIIF Curation Finder はあくまで文書が中心であり、実体を中心とした構造にはなっていない。そこでデータ構造を文書中心から実体中心に反転させ、実体空間側から閲覧し検索するための新たな仕組みが必要である。そのために構築したのが、データポータル「edomi」である。

カテゴリ一覧



図3: edomi カテゴリ一覧画面 (2021年12月段階では公開予定のものも含む)



図4: 同じ実体空間識別子を与えられたマイクロコンテンツの切り替え表示
『江戸名所図会』と『絵本江戸土産』に掲載されている挿絵を、上部の●をクリックすることで切り替えている

6. データポータル edomi

edomi は、都市としての江戸/時代としての江戸に関する各種のデータを一望するデータポータルである。現代のある場所は過去にどんな風景だったのか、江戸には現代で言えばどんな業種の商店が多かったのかなど、現代に生きる人々に馴染みの方法を用いて、現在の視点を起点として情報を探すことを目的とする。その中心的なコンセプトは、文書空間と実体空間の双方向結合に対し

て、実体空間側からのアクセスを提供することにある。そのために実体空間識別子を基準としたデータの組織化を行う。すなわち、ある名所に関する資料、ある商店に関する資料というように、実体に関連する資料やメタデータを、資料横断的に集約するデータ構造となっている。

また実体空間は、現在の実世界にも存在する実体を多く含むことから、現在から過去への連続性を重視することも中心的なコンセプトとなる。そ

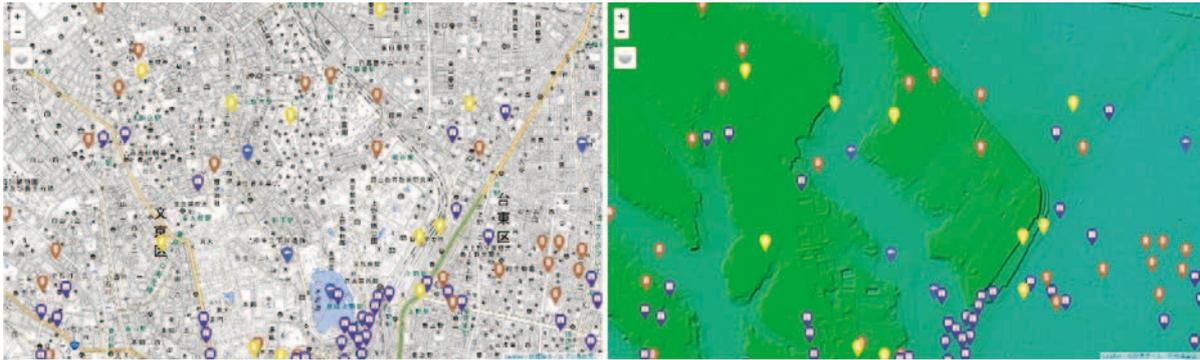


図 5: 「マップ」で上野地域を表示
現在の地名が分かる「地理院タイル単色地図」(左)と、台地との関係が分かる「地理院タイル色別標高図」(右)を切り替え

分類	件数	説明
6032医薬品小売業	205	主として一般用医薬品及び医療用品を小売する事業所をいう。
5512売物卸売業	175	主として売物を卸売する事業所をいう。
5139その他の身の回り品卸売業	130	主としてその他の身の回り品、装身具(貴金属製を除く)を卸売する事業所をいう。
5861菓子小売業(製造小売)	116	主として各種の菓子類、あめ類を製造してその場所で小売する事業所をいう。
7621日本料理店	102	主として特定の日本料理(そば、うどん、すしを除く)をその場所で小売する事業所をいう。
5523化粧品卸売業	92	主として化粧品、石けん、香水、おしろいなどを卸売する事業所をいう。
5594醸造用品・かん具卸売業	88	主として醸造用品、かん具を卸売する事業所をいう。
5211米麦卸売業	83	主として米及び麦を卸売する事業所をいう。
5521医薬品卸売業	71	主として医薬品を卸売する事業所をいう。
5505食品卸売業	60	主として食品を卸売する事業所をいう。

図 6: 総務省「日本標準産業分類」からの検索
「日本標準産業分類」それぞれに件数が表示され(左)、分類ごとに商人を検索可能(右)

ここで、現代のポータルサイトである Yahoo! [13]などを参考に、図 3 のように「トラベル」や「ショッピング」などのカテゴリを設け、文書中心ではなく実体中心の視点から情報を組織化した。また地理情報や分類体系なども現代の実体空間を基準として過去の情報をリンクすることにより、現代でのデータ利活用を促進する助けとした。具体的には以下の方針で edomi を構築した。

1. 「トラベル」カテゴリには、「江戸観光案内」のマイクロコンテンツをまとめ、「江戸観光 ID」単位で検索結果を表示する。
2. 「ショッピング」カテゴリには「江戸買物案内」のマイクロコンテンツをまとめ、「江戸買物 ID」単位で検索結果を表示する。
3. 同一の実体空間識別子が付与されたマイクロコンテンツは、識別子ごとのランディングページで資料横断的に表示を切り替えできる(図 4)。
4. 「マップ」カテゴリには、GeoLOD の緯度経度情報をもとに、マイクロコンテンツが記述する場所を現在の地図上に表示する。

IIIF Curation Finder のように文書を中心とした

仕組みと比べると、edomi では地理情報などの情報をわかりやすく表示できるようになった。IIIF Curation Finder でも同じ地名に関する画像を探すことはできたが、あくまでメタデータの文字情報による検索が限界であった。一方、edomi の「マップ」カテゴリでは、GeoLOD を活用して現代の地図上にマイクロコンテンツを公開できるため、江戸時代の名所・商店を現実世界の位置を基準に、視覚的に確認できるようになった。

なお、江戸に関わるマイクロコンテンツについては、以前の論文[6]でも ICP を活用して「江戸買物案内」の版面を切り抜いた分析を紹介した。そして掲載されている版面サイズや、文字として記載されている居所(住所)から、江戸時代の商人の状況を情報空間に整理する段階までは分析を進めることができていた。しかし本論において、双方向結合の利用と edomi への展開までを進めたことで、文書そのものに含まれる情報にとどまらず、外部の様々な情報と組み合わせることにより、より解像度の高い分析が行えるようになった。

例えば図 5 は「マップ」カテゴリのマイクロコ



図7：文書空間と実体空間の双方向結合と歴史的記録検索エンジン「れきろく」構想

コンテンツ地点を、国土地理院「地理院タイル」を背景地図として表示したものである。背景地図を切り替えることで、現在の道路地図と比較するだけでなく、標高との関係を調べることもできる。一例として、上野の寺社が台地の縁に沿うように点在し、そうした寺社への参拝客を相手にする店舗が、台地の入り口にまとまって存在したことが良くわかる。ICPのみを活用した段階では、こうした位置関係についてはあくまでも同一の住所の集合として捉えるにとどまり、名所の周りには商店も多いという一般的な発見にとどまっていたが、GeoLODとの結合によって外部の地理情報と組み合わせる分析が可能になった。

さらに図6には、「江戸買物案内」に対して各商人が扱っている商品を基準に総務省「日本標準産業分類」[14]を紐づけ、現在の分類から検索可能とした例を示している。現在の分類から見て、江戸にはどのような産業がどのような地理的分布を示していたかが可視化できるようになり、ICPを活用した文書空間の分析とは別の切り口から江戸商人の状況を調べることも可能となる。このような情報の組織化により、歴史資料から取り出したマイクロコンテンツを、経済史の視点で活用する可能性も開けてくる。

現在の edomi のカテゴリはトラベル、ショッピング、マップのみであるが、将来的にはここに江戸に関するデータを集約することで、江戸を多角的に調べるデータポータルとして発展させる計画である。例えば、「グルメ」カテゴリには江戸時代のレシピデータ、「政治」カテゴリには武鑑など江戸時代の政治に関するデータ、「教育」カテゴリにはくずし字に関するデータなどを掲載する予定である。その際に重要となるのは、文書空間識別子と実体空間識別子の結合という部分を可能な限り共通化することである。メタデータ

やそれを edomi 上で組織化する方法は分野ごとに異なる。ゆえに、現代の人々にとって利用しやすいデータの構造化、組織化、可視化を実現するために、基盤となる文書空間識別子と実体空間識別子の共通化を進めていきたいと考えている。

7. 歴史ビッグデータ基盤に向けて

本論で提案した文書空間と実体空間の双方向結合という概念は、edomi データポータルに限定されたものではなく、より長期的には歴史ビッグデータ基盤の基本的なデータモデルとすることを想定したものである。歴史ビッグデータ研究では、過去の人間が書き記した実世界に関する記録を構造化したデータを「歴史的記録」と定義する。現代のニュースでは情報の基本単位は 5W1H とされるが、過去の世界についても同様に 5W1H などの形式で過去の事実に関する基本単位の構造化データを収集し、それをエビデンスとして活用することで、過去の世界を復元し分析することを考える。このように過去のビッグデータを収集して過去の世界を復元するという構想は、欧州におけるタイムマシンプロジェクト[15]など、様々な地域で注目が高まっている。

そこで重要となるのが、文書を単独で理解するだけでは、過去の世界を復元するには十分ではないという認識である。従来の歴史文書を対象とした構造化では、主に文書をマークアップする構造化手法などに意識が向いていた。しかし、文書から構築した半構造化データが文書視点の記述であるのに対し、過去の世界を探求するためのエビデンスとして欲しいのは実世界視点の記述である。そして、文書視点の記述と実世界視点の記述の間に生じる不整合により、両者がうまく接続しない場合があるという問題がこれまで見過ごされてきた。この問題を「構造化ギャップ」と呼ぶ。

この問題を解決するには、文書を出発点とする半構造化データと、実体を出発点とする構造化データを、双方向から接続する方法を確立する必要がある。これが本論で提案する「文書空間と実体空間との双方向結合」の長期的な狙いである。

この構想を実現するための要となるのが、文書空間と実体空間とを結合した「歴史的記録」を蓄積する検索エンジン「れきろく」である。図7に示すように、文書空間の構造化と実体空間の構造化を「れきろく」で結合し、歴史ビッグデータを構築するというのが歴史ビッグデータ構想の概要である。この構想は、本論で扱ったマイクロコンテンツだけでなく、人文学資料一般にも適用できると考えている。一方、この構想を実現するには、資料や分野の特性に応じて、データ構造化を促進するツールを用意する必要もある。そのようなツール群のエコシステムを徐々に広げていくことが、歴史ビッグデータ基盤の構築に向けて重要な課題となる。

8. おわりに

本論は「文書空間と実体空間の双方向結合」という考え方に基づき、従来は文書空間における横断的検索と活用にとどまっていたマイクロコンテンツを、実体空間側から検索して可視化するデータポータル *edomi* を実現した。

データポータル *edomi* の実現は、今後の歴史ビッグデータ基盤構築に向けた第一歩でもある。本研究が対象とする実体の概念は非常に広く、実空間に存在する何らかの「モノ」としては、地球上の位置を示す座標や地名、過去から未来までの時間を特定する時間情報、そして世界に生きた人物に関する情報なども対象に含まれる。そして実体が空間を移動した記録や、そこで見聞きした記述などが歴史的記録の対象となる。これらが幅広い分野にわたって蓄積されていけば、過去の世界の状況を探求するための有用なツールになることが期待できる。

これまで我々は、江戸に関する出版物の内容を情報空間に整備するという自身の研究関心を軸としてマイクロコンテンツの整備を行ってきた。しかし双方向結合によって、他の研究プロジェクトや社会活動に活用可能な形でもマイクロコンテンツが提供できることが分かった。経済史などの分野で既存のマイクロコンテンツを新たな視点で活用できる研究者との共同研究の可能性があるだけでなく、歴史資料を活用した観光などの活動とも協力可能である[16]。

マイクロコンテンツをより充実させていくとともに、マイクロコンテンツ制作者以外にも様々な目的でこれを活用する道を開いていきたいと考

えている。これはマイクロコンテンツ構築に活用してきた ICP が提示している「利用者主導型」という理念をさらに発展させることにもつながるものとなる。

参考文献

- [1] “江戸買物案内”. <http://codh.rois.ac.jp/edo-shps/>, (参照 2021-11-01)
- [2] “江戸観光案内”. <http://codh.rois.ac.jp/edo-spots/>, (参照 2021-11-01).
- [3] “edomi”. <http://codh.rois.ac.jp/edomi/>, (参照 2021-11-01)
- [4] “IIF Curation Platform”. <http://codh.rois.ac.jp/icp/>, (参照 2021-11-01)
- [5] 北本 朝展, 本間 淳, Tarek SAIER, IIF Curation Platform : 利用者主導の画像共有を支援するオープンな次世代 IIF 基盤. 人文科学とコンピュータシンポジウム じんもんこん 2018 論文集. 情報処理学会, pp. 327-334, 2018.
- [6] 鈴木親彦, 北本朝展. IIF Curation Platform による『江戸買物独案内』のマイクロコンテンツ化：非文字情報を軸に. 人文科学とコンピュータシンポジウム じんもんこん 2019 論文集. 情報処理学会, 2019 p.11-18.
- [7] 江戸名所図会 (国文学研究資料館撮影/味の素食の文化センター蔵). 日本古典籍データセット. DOI: 10.20730/100249896
- [8] 絵本江戸土産 (国文学研究資料館蔵). 日本古典籍データセット. DOI: 10.20730/200013687
- [9] 江戸名所記 (北海道大学付属図書館蔵). 新日本古典籍総合データベース. DOI: 10.20730/10000011
- [10] “GeoLOD”. <https://geolod.ex.nii.ac.jp/>, (参照 2021-11-01).
- [11] 北本 朝展, 鈴木 親彦, 寺尾 承子, 堀井 美里, 堀井 洋. 地理的史料を対象とした歴史地名の構造化と統合に基づく江戸ビッグデータの構築. 人文科学とコンピュータシンポジウム じんもんこん 2020 論文集, pp. 171-178, 2020.
- [12] 北本 朝展, 村田 健史. 歴史的行政区域データセットβ版をはじめとする地名情報基盤の構築と歴史ビッグデータへの活用. 情報処理学会技術報告, Vol. 2020-CH-124, No. 1, pp. 1-8, 2020.
- [13] “Yahoo! JAPAN” <https://www.yahoo.co.jp/>, (参照 2021-10-26)
- [14] “総務省日本標準産業分類” (平成 25 年 10 月改定) https://www.soumu.go.jp/toukei_toukatsu/index/seido/sangyo/02tukatsu01_03000023.html, (参照 2021-10-26)
- [15] 北本 朝展. 「過去のビッグデータ」を探る欧州タイムマシン研究計画, カレントアウェアネス. No. E2248, 2020
- [16] 鈴木親彦. 江戸資料のデジタル化：情報空間からメタ観光への展開. メタ観光推進機構オンライン公開勉強会. <https://metatourism.jp/blog/2021/06/25/20210624lecture/>, (参照 2021-10-26)