

史資料からの地理空間情報の収集と管理に関する考察

山田 太造^{1,a)}

概要: 前近代における古文書・古記録や現代における対象地域の観察記録であるフィールドノートなどにおいて、少なからず地理空間に関する情報が記載されている。本研究では、史資料に出現する地理空間に関わる多様なデータを収集し、それらを蓄積するために構築している地理情報基盤システムについて紹介する。またトピックモデルを用いることで地理空間情報における空間的特徴の検出手法および利活用のためのデータベースシステムについて紹介する。

キーワード: 地理空間情報, 史資料, フィールドノート, データベース, トピックモデル

A Study of Collecting and Managing Method for Geospatial Information from Historical Materials

TAIZO YAMADA^{1,a)}

Abstract: Materials (such as an old document and an old diary in the premodern in Japan, and a field note in which a subject of regional observation is written) have description about geographical spaces. In the study, we introduce a collecting method for collecting a variety of data related to the geographic spaces which appear in the materials, and an information system in which the information can be stored. And we introduce a detection method of spatial features with topic models and a database system for utilization of them.

Keywords: geospatial information, historical material, field note, database, topic model

1. はじめに

前近代における古文書・古記録や現代における対象地域の観察記録であるフィールドノートなどにおいて、少なからず地理空間に関する情報が記載されている。

本研究では、記録類のような史資料に出現する地理空間情報の詳細、その蓄積と利用方法、および、ある地域に関する記録をもとに地域の特徴を分析していく手法についても紹介していく。本稿はこれ以降、2節では、フィールドノートをもとにした地域の特徴を分析していく手法を紹介する。3節では、前近代日本史に関する古記録を例に、記述内で出現する地名について紹介する。4節では、史資料にて登場した地理空間情報の蓄積とそれを蓄積するシステム

ム、およびそれを利活用するためのシステムについて紹介する。

2. フィールドノートの例

フィールドで得られる観察記録やメモ、スケッチ、撮影した写真などは、現地のことを知る重要な研究資料 [1] である。調査者が個人的に観察した地域を自らの視点で把握した内容を記すため、調査者以外が利用することは困難であるが、再整理を行うことにより、他者にも利用可能になると考えられる。

高谷好一著『地域研究アーカイブズ フィールドノート集成』(京都大学地域研究統合情報センター CIAS Discussion Paper Series) は、高谷氏の協力のもと、フィールドノートのテキスト化を行い、本人によりイラスト・写真の整理、各種資源の検証を行うことで作成された。これまでに

¹ 東京大学史料編纂所
Historiographical Institute The University of Tokyo

^{a)} t_yamada@hi.u-tokyo.ac.jp

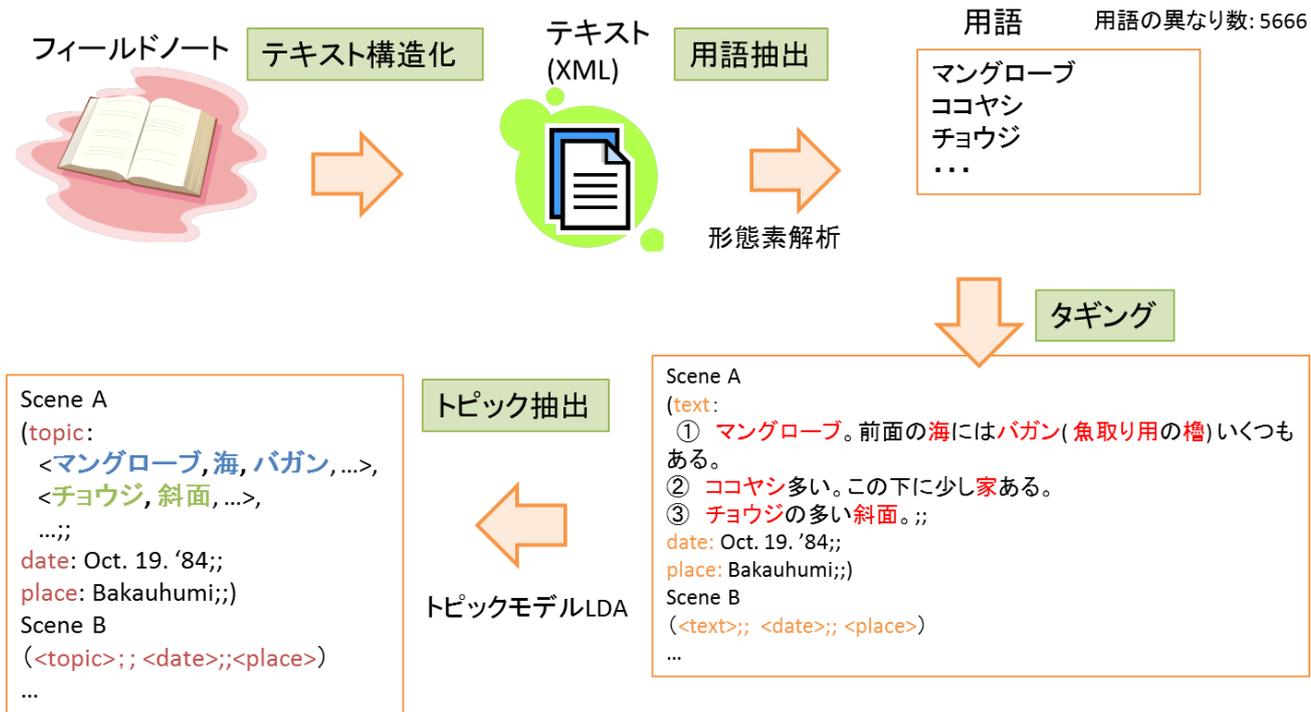


図 1 フィールドノートからのトピック検出
 Fig. 1 Topic detection from a field note

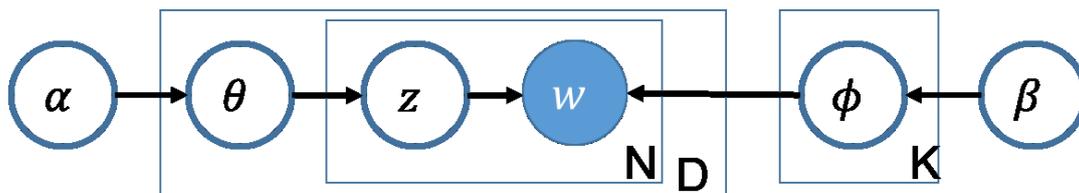


図 2 LDA のグラフィカルモデル
 Fig. 2 Graphical model for LDA

『フィールドノート集成』として8冊(合計4405ページ)が刊行された。『フィールドノート集成』のうち「フィールドノート集成2」にあるスマトラ(1984.10-19-1985.1.18)(以下、高谷ノート)は、分量としてはA4サイズ198ページ程度であるが、スマトラ島全域をカバーしている。

高谷ノートは、日ごとの記録であり、その日の記録の始めに日付と該当日の概要が記されており、場面ごとの記録、スケッチ画像、写真画像が掲載されている。記録として、場面ごとの観察記録以外にも、現地の人から聞いたことや調べたことなどもある。

著者らは[2],[3]にて、高谷ノートに対してトピックモデルの1つであるLDA(Latent Dirichlet Allocation)[4]を用いることで、高谷ノートの本文に潜在するトピック(潜在トピック)を検出していくことで、高谷ノートに記述されている場面ごとに特徴づけて行く手法を述べた。この概要を図1に示す。用語抽出では形態素解析器として

mecab*1, 形態素解析器用辞書としてIPADic*2を用いた。場面のトピックはLDAを用いて検出した。LDAは、統計的に共起しやすい用語の集合がいくつか存在しており、これを潜在トピックとして扱う。一つの文書に複数の潜在トピックが存在することを仮定しており、そのトピックの分布をモデル化していく。図2はここで用いたLDAのグラフィカルモデル表現を示す。ここで、灰色の円は観測変数、白色の円は未知変数を示し、矩形は繰り返しを、矩形の右下の数字はこの矩形で表す繰り返しの回数を示す。wは先に述べた用語抽出の結果、つまり用語を示す。ここでは唯一観測される変数である。zはトピック、thetaはトピック分布、phiは用語分布を示す。またalphaおよびbetaはthetaおよびphiのパラメータ、つまりハイパーパラメータを示す。文書数をD、文書dの用語数をN_dとしたとき、theta_dおよびphi_kは

$$\theta_d \sim Dir(\alpha) \quad (d = 1, \dots, M),$$

*1 <http://taku910.github.io/mecab/>

*2 <https://github.com/neologd/mecab-ipadic-neologd>

表 1 Solok 周辺のトピック
Table 1 Topics around Solok

トピック	用語数
V5	1
V6	1
V16	2
V17	1
V21	32
V24	35
V27	1
V28	1
総計	74

$$\phi_k \sim \text{Dir}(\beta) \quad (k = 1, \dots, K).$$

により生成されると仮定する。ここで $\text{Dir}(\cdot)$ はディリクレ分布を示す。トピック $z_{d,i}$ は下記のように生成されることにする。

$$z_{d,i} \sim \text{Multi}(\theta_d) \quad (i = 1, \dots, N_d)$$

ここで $\text{Multi}(\cdot)$ は多項分布を示す。さらに用語 $w_{d,i}$ は下記による生成を仮定する。

$$w_{d,i} \sim \text{Multi}(\phi_{z_{d,i}}) \quad (i = 1, \dots, N_d)$$

トピック数を 30 として高谷ノートに対して本モデルを適用した。

高谷ノートから、著者である高谷氏は Solok という町へ 3 度訪れている (1984 年 10 月 22-23 日, 1984 年 12 月 19 日, 1985 年 1 月 12-13 日)。1984 年 10 月 22-23 日では、スマトラ島の東にある Jambi から向かい、北側へ抜けている。1984 年 12 月 19 日は、Lubuk Jambi というスマトラ島の中央部にある町から向かい、西の海沿いにある Padang という町へ抜けている。1985 年 1 月 12-13 日は、Baringin という Solok の北に位置する町から向かい、南へ抜けている。Solok の町およびその周辺に関する記述から検出したトピック (表 1 の多くはトピック 21 とトピック 24 だった。トピック 21 の上位語は (水田, 広い, 稲, 多い, 幅, 棚田, 川, 谷地田, 周り, 柵) であり, 水田の情景を表すトピックであろうと思われる。トピック 24 の上位語は (多い, 家, コーヒー, ココヤシ, 村, ゴム, 周り, ランプータン, マンゴー, チョウジ) であり, 主に水田以外の畑の情景を表すトピックであろう。これにより Solok 周辺の様子を, あくまでも高谷ノートでの範囲ではあるが, 把握することができた。

3. 古記録

3.1 古記録の例 1 : 『上井覚兼日記』

『上井覚兼日記』は薩摩島津家の家臣であった上井覚兼による日記である。原本は東京大学史料編纂所 (以下, 史料編纂所) 所蔵の重要文化財である。原本は 28 冊で構成

されている。天正 2 年 8 月 1 日から天正 14 年 10 月 15 日までの約 12 年間の日記であるが, このうち, 天正 4 年 9 月 7 日から天正 10 年 11 月 3 日までの日条は存在しない。天正 2 年 8 月 1 日から天正 4 年 9 月 6 日までは鹿児島における奏者時代の, 天正 10 年 11 月 4 日から天正 14 年 10 月 14 日までは日向国宮崎城主時代の日記として見るができる。沖田暉の戦いや岩屋城の戦いに代表される島津家と龍造寺家・大友家との関係や, 羽柴秀吉による九州征伐寸前までの島津家の動向など, 天正期における島津家や九州の情勢を知る上で重要な史料として位置づけられている。『上井覚兼日記』は『大日本古記録 上井覚兼日記』(全 3 冊) として編纂史料集が発刊されている。これをベースとして, SHIPSDB「古記録フルテキストデータベース」(以下, 古記録 FT) に, この全テキストが格納されており, テキスト検索サービスを提供している。

例として古記録 FT に対して“五島”で検索してみると, 3 件の検索結果を得た。これらはいずれも『上井覚兼日記』での結果だった。

- 天正 12 年 6 月 27 日「一、廿七日、出仕如常、五嶋 (肥前、宇久純定) より、今度御高來 (肥前) 表御勝利之御祝言使書也、并太刀・巻物進上也、...」
 - 天正 13 年 4 月 24 日「... 出仕帰に鎌刑へ参し候、種々会尺共也、山新・鎌雲同心申候也、拙宿へ各参上申之由候て御礼共也、御酒参会閑談共也、五嶋 (肥前) 宇久和州 (純玄) より、使者貞方右衛門佐進上也、吾々へも書状并太刀・馬・織筋預候、書面、遙御無音所存之候、...」
 - 天正 14 年 2 月 20 日「... 此日、五島 (肥前) 宇久大和守 (純玄) 殿、去年御勝利之御祝言被申上候、到吾吾も使書預候、并太刀一腰・百疋・魚預候也、...」
- いずれも五島の地名ではなく、人名でヒットしている。

1 つ目は天正 12 年 3 月 24 日に勃発した沖田暉の戦いの後の出来事を記したものである。沖田暉の戦いは肥前の龍造寺家と薩摩の島津家との合戦であり, 九州の覇権をかけた戦いという側面も持つ。この戦いで勝利した島津家は, この後島原周辺の安定化に向けた動きを見せていた。ある程度の安定を見出していたところに宇久純定より勝利の祝言の使者が到来したことがわかる。宇久純定は当時肥前江川城主だった。宇久家は肥前松浦郡宇久島が出身地であり, この頃には五島列島全域を掌握していたという。沖田暉の戦いを契機に肥前の国衆は龍造寺家から島津家への寝返りが相次いで起こっていた。島津家では宇久家の島津家の接近により五島列島においても島津の影響下に置かれることを意味する。

2 つ目, 3 つ目も宇久家と島津家との関係を示すものであり, その信頼関係を強化していることを意味する。宇久純玄 (すみはる) は宇久純定の子である宇久純堯の子とも, 宇久純定の長男の遺児とも言われる。宇久純定はキリスト

<図>家久上京記主要宿泊地 (除、山城・近江・伊勢・大和)

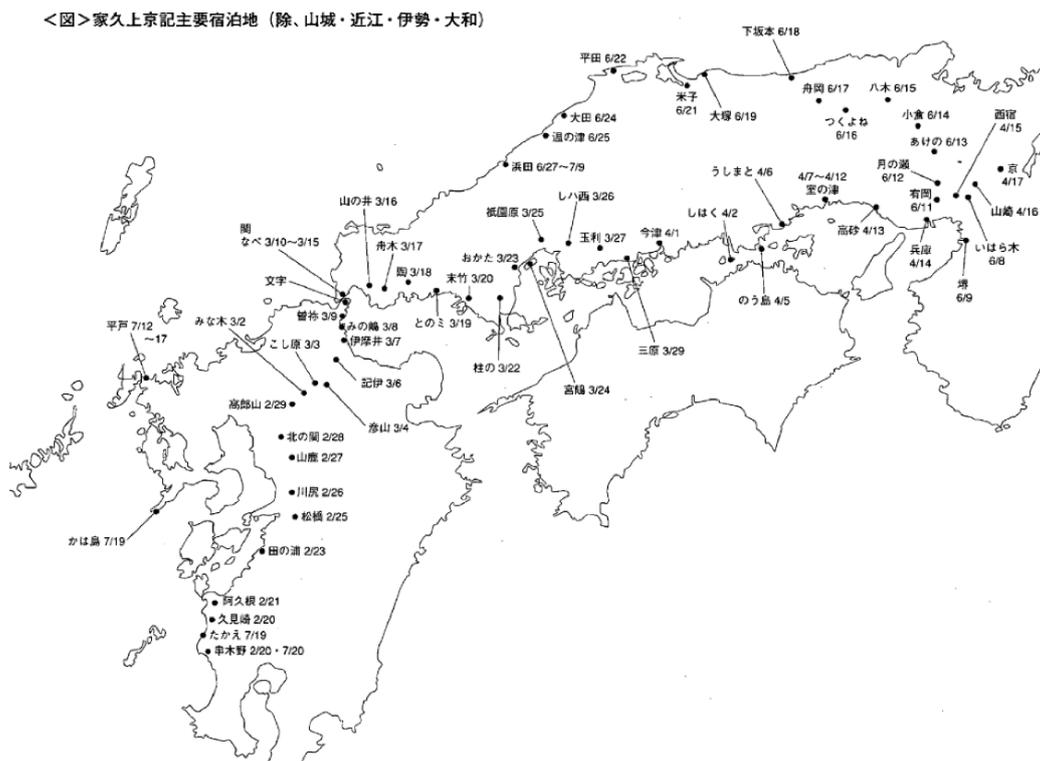


図 3 『中務大輔家久公御上京日記』における島津家久の宿泊地 ([6] より抜粋)

Fig. 3 Accommodation location by Shimazu Iehisa in “the old diary of when Nakat-sukasataifu Iehisa went to Kyoto”

教に入信しており、純玄もキリスト教の洗礼を受けている。後の天正 15 年に家督を継ぐ。その後は五島列島を平定し、五島氏に改名し、宇久純玄の継嗣である五島玄雅が初代福江藩主となる。

3.2 古記録の例 2 : 『中務大輔家久公御上京日記』

『中務大輔家久公御上京日記』は戦国期島津氏当主義久の弟である又七郎家久の伊勢参詣道中記である。この史料は『旧記雑録』所収および玉里文庫本『家久公上京日記』あるいは島津家本『中書家久公御上京日記』の書名である。この史料については [6] にその詳細が紹介されており、かつ翻刻も掲載されている。

島津家久は上京および伊勢参詣のため、天正 3 年 2 月 20 日に串木野 (現在の鹿児島県いちき串木野市) を立ち、瀬戸内ルートで同年 4 月 17 日に上京、往路は日本海側を通り、同年 7 月 20 日に再び串木野に到着した。家久の滞在先は図 3 (この図は [6] にある) のとおりである。この上京については『大日本史料』10 編 28 冊に記載されており、綱文 (歴史上の事件の概要を表す文章) として「島津義久の弟家久、上洛せんとし、是日、薩摩串木野を立出す、」とある。これを裏付ける史料の 1 つがこの『中務大輔家久公御上京日記』であり、もう 1 つは先に上げた『上井覚兼日記』である。

この記録にも五島に関する記述がある。帰路の途中、平

戸に 7 月 12 - 17 日の間滞在した後、7 月 19 日にはかは島 (長崎県樺島) に滞在している。この直前の日の記録は次のとおりである。

一、十八日午刻二出舟候へハ、肥州舟にておくり候に、拙者あなたの舟二乗移、酒ゑんさまくにて、さていとまこひ仕、本舟にのり候へハ、舟へ樽・食籠種々の肴おくられ候、せとのと (瀬戸渡) へいへる所迄、肥州おくられ候、其より九十九嶋 (松浦郡) を左の方に見て打過、右方にこたう (五島)、福田・礼崎夜中ニ打過候、

これにより大雑把ではあるが通過したルートを知ることができる。『上井覚兼日記』でも共通するが、記述内で地名ともにそこに関連する人名が登場する。人物を把握する上でも地名などの地理空間情報はとても重要である。例えば、先の『中務大輔家久公御上京日記』7 月 17 日条に“肥州”が登場する。これは地理空間を表すというよりも人名を示しており、松浦隆信を示す。この月の 13 日条・14 日条は、一、十三日、唐舟に乗見物仕候、なんはん (南蛮) より豊後殿へ進物とて虎の子四疋、それをめつらしく見帰り候へは、加治木衆彦太郎といへる者、樽・食籠持来候、亦肥州 (松浦隆信) より樽ニツ、肴取合、平松七郎左衛門といへる者使者、

一十四日、あきめの勘介樽、又とまりの又十郎樽持来候、亦平戸の薩戸とひすゝ持来り、さて町寺家くなど一見候へは、普門寺といへる寺にて肥州出合、頻にと候間、堅く斟

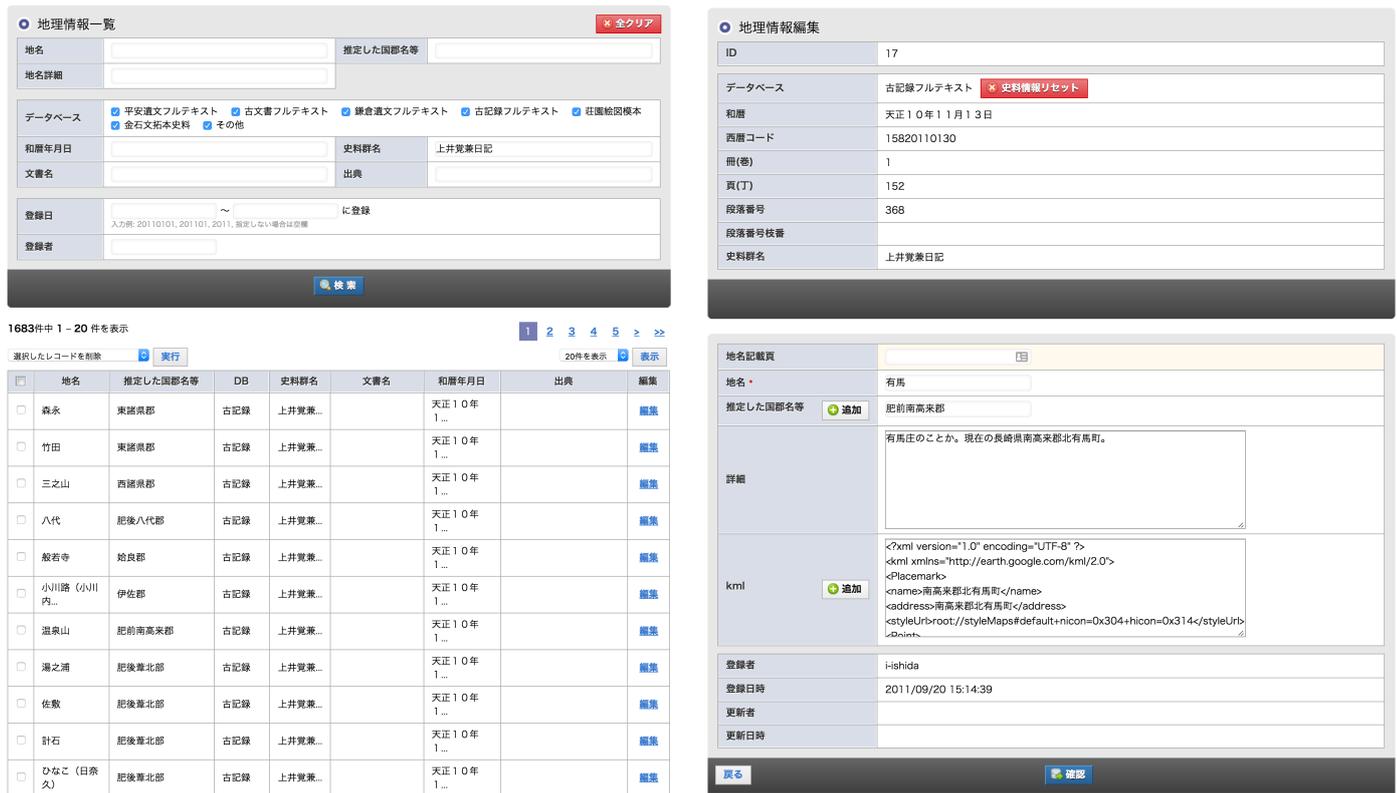


図 4 地理情報蓄積システム
Fig. 4 Geographic information accumulation system

酌仕候へ共、猶来るへく由申され候間、無余義まかり候て、肥州・同捨弟兩人へも見参仕、帰り候へは、臆而肥州礼に來られ候、其よりとひの宿にて酒寄合、肥州より太刀あつかり候、である。13日に豊後殿への進物である虎を見ていたことがわかる。ここで“豊後殿”は大友義鎮（宗麟）を示す。同じ日に松浦隆信から酒と肴を頂いているが、“肥州”と表現しており、14日に松浦隆信にあっては、相変わらず“肥州”と表現している。

4. 地理情報蓄積システム

史料編纂所では史料に登場する地名を収集し、その史料に関連付けて管理する地理情報蓄積システムを構築した(図 4)。このシステムで蓄積していく地名データは、一意に決められた地名を蓄積していくのではなく、史料に登場した地名を、その史料の出典、日付、経度・緯度などとともに蓄積していく。経度・緯度は KML 形式で記述したものをそのまま用いる。蓄積された地名データはこのシステムで検索し、検索結果として地名データを閲覧するだけでなく、データ検索およびデータ取得のための API を作成しており、この API を介して他のシステムからこのシステムに蓄積されている地名データに対して、検索・取得などの利用を可能としている。

史料編纂所データベース SHIPS DB において、このデー

タを利用しているデータベースシステムとしては、「金石文拓本史料データベース」および「史料編纂所所蔵荘園絵図摸本データベース」(以下、荘園 DB)がある。荘園 DB (図 5) の場合、検索結果として得られた荘園データを、付与された経度・緯度をもとに地図システム(実際には地理院地図^{*3})上にマッピングすることができる。また検索結果詳細からその荘園に関するデータについて地図システム上に表示することも可能である。

この地図システム上に検索結果を表示するための仕組みは、昨今目新しいものではなくなった。ここで重要な事は、歴史的地名のようにまだ地名辞書には記載されていないようなデータについても蓄積していき、すでに記載されている地名についても付加すべき情報があれば追記し、それを利活用可能な状態にしておくことが重要である。この点において、地名情報の蓄積および地理情報蓄積システムでの管理は重要であると考えている。

5. おわりに

本稿では、地名という観点から、記録類を用いてその地域に関する情報を収集し、また分析していった事例について述べた。地理空間情報は、たとえば日本史のみ、といった特定の分野のみではなく、人文科学、さらにそれ以外の多くの分野で求められている情報である。その観点から、

^{*3} <http://maps.gsi.go.jp/>



図 5 史料編纂所所蔵荘園絵図摸本データベース
Fig. 5 Database of the Shoen ezu reproductions

分野横断となりうるアイテムであり、異分野連携による研究促進や、新たな研究領域につながるかもしれない。また、ある地域に関するあらゆる分野の情報を統合的に扱うことができれば、地域貢献にもつながる活動とも言える。そのような観点からも地理空間情報の蓄積と提供の方式について改めて考えていく予定である。

謝辞 本研究の成果の一部は、日本学術振興会科学研究費若手研究 (B) (26730167)、基盤研究 (A) (26240049)、基盤研究 (A) (15H01722)、京都大学地域研究統合情報センター共同利用・共同研究拠点個別共同研究ユニット「フィールドノートにおける場面特徴の表現手法の深化と利活用に関する研究」の助成を受けたものによる。『中務大輔家久公御上京日記』の翻刻は村井祐樹氏から提供いただいたものによる。

参考文献

- [1] 柳澤雅之: フィールドノート・プロジェクト, Seeder11 号, pp.14-22(2014).
- [2] 山田太造: 地域研究資料と対象とした時空間情報に着目したデータの構造化, 研究報告人文科学とコンピュータ (CH), 2015-CH-105(4), pp.1-6(2015).
- [3] 山田太造: 地域研究史資料を対象とした時空間的特徴の抽出と場面の構造化, 情報科学技術フォーラム講演論文集, 14(4), pp.409-410(2015).
- [4] D.M.Blei, A.Y.Ng, and M.I.Jordan: Latent Dirichlet Allocation, Journal of Machine Learning Research, vol.3, pp.993-1022(2003).
- [5] T.L.Griffiths and M.Steyvers: Finding scientific topics, Proc. of the National Academy of Sciences of the United States of America, vol.101, pp.5228-5235(2004).
- [6] 村井祐樹: 東京大学史料編纂所所蔵『中務大輔家久公御上京日記』, 東京大学史料編纂所研究紀要, Vol.16, pp.99-118(2006).