2015年12月12日発行　第25巻4号　　学術刊行物 ISSN 0917-1436

Journal of Japan Society of Information and Knowledge

情報知識学会誌

Vol.25　No.4　(Dec. 2015)

～～～～～～～～～～～～～～～～～～～～目　次～～～～～～～～～～～～～～～～～～～～

～～～

JSIK

情報知識学会

http://www.jsik.jp/

開催に当たって
「地域情報学における知識情報基盤の構築と活用」

原正一郎
京都大学地域研究統合情報センター

地域研究は、地域を包括的に理解する知的体系であり、歴史学・民族学・人類学・社会学・経済学・農学・工学・医学などの多様な研究領域を包含する学際的な研究分野です。多様化とグローバル化の急速な進展に伴う今日的な諸課題の解決に貢献する研究分野として期待されています。地域研究には、大学だけではなく、NPO・NGO・企業・行政などの多様な組織が参加していることも大きな特徴であり、さまざまな地域でさまざまな活動を展開し、膨大なデータや情報を収集しています。しかし、これらのデータや情報は、収集や利用の目的・視点・概念・手法・機器・理論などを異にしているため、知識としての共有は進んでいません。地域情報学は、このような地域研究データや情報を知識として蓄積・共有・活用するための情報学的課題であり、学際研究としての地域研究を支援する情報基盤の構築を目指しています。

本情報知識学フォーラム「地域情報学における知識情報基盤の構築と活用」では、多様な地域研究データや情報を「貯める」「繋げる」「活用する」という視点から、地域情報学を俯瞰します。地域研究データには、HTML/XML の構造化文書、シソーラスの樹形図、データベースのテーブルなど構造の異なるデータが混在しています。これらを同じ枠組みで「貯める」ために RDF による Linked Open Data に注目しています。知識を「繋ぐ」主要な情報は語彙ですが、地域研究では語彙情報の乏しい資料も対象とします。このような場合、位置と時間の情報処理が重要となります。多様なデータ構造を語彙・位置・時間に注目して「繋げる」ことにより、センサデータや IoT データの対極に位置するビッグデータ、すなわち単体としてはスモールデータであっても、繋ぐことにより複雑かつ多様な学術ビッグデータを構成します。このようなビッグデータのナビゲーションや可視化も地域情報学の重要なテーマとなっています。

地域情報学の成果は人文社会科学全般への展開も可能と考えています。大量・多様な学術ビッグデータの活用により、大量データや質的に異なるデータを人文社会科学に導入できる、地域間・時代間で比較しながらさまざまなスケールでの分析できる、文理研究分野を繋ぐインタフェースを提供できるなどの可能性が開かれると期待しています。

情報知識学会
第20回　情報知識学フォーラム
「地域情報学における知識情報基盤の構築と活用」

2015年12月12日（土）
京都市・同志社大学今出川キャンパス

13:00-13:10　開会

13:10-13:45　Linked Dataによる分野連携型データベースの枠組み－大規模知能データベース－　武田英明（国立情報学研究所、情報学プリンシプル研究系）

13:45-14:20　人文社会系大規模データベースへのLinked Dataの適用－推論による知識処理－　後藤真（国立歴史民俗博物館、研究部）

14:20-14:55　さまざまな社会公共サービスを支える共通データ基盤としての空間情報－場所による知識処理－　柴崎亮介（東京大学、空間情報科学研究センター）

－－－　休憩　(20分)　－－－

15:15-15:50　時間情報基盤の構築と活用－時間に基づく知識処理－　関野樹（総合地球環境学研究所、研究高度化支援センター）

15:50-16:25　フィールドノートに記述された場面を特徴付ける－語彙による知識処理－　山田太造（東京大学、史料編纂所）

16:25-17:00　地域情報学のこれまでとこれから－地域研究統合情報センターの実践事例を通して－　亀田尭宙（京都大学、地域研究統合情報センター）

17:00-17:25　総合討論

第20回情報知識学フォーラム予稿

Linked Dataによる分野連携型データベースの枠組み

The framework for cross-disciplinary databases with Linked Data

武田英明[1*], 加藤文彦[2], 大向一輝[3]
Hideaki TAKEDA[1*], Fumihiro KATO[2], Ikki OHMUKAI[3]

1 国立情報学研究所, 総合研究大学院大学
National Institute of Informatics
〒101-8430 東京都千代田区一ツ橋2-1-2
E-mail: takeda@nii.ac.jp
2 情報・システム研究機構, 国立情報学研究所
Research Organization of Information and Systems
〒105-0001 東京都港区虎ノ門4-3-13 ヒューリック神谷町ビル2階
E-mail: fumi@nii.ac.jp
3 国立情報学研究所, 総合研究大学院大学
National Institute of Informatics
〒101-8430 東京都千代田区一ツ橋2-1-2
E-mail: i2k@nii.ac.jp

本稿ではLinked Dataの仕組みがいかに分野を横断して連携するデータベースを構築することを可能としているかを述べる。Linked DataはセマンティックWebの技術を使い、Web of Data、すなわちDataが相互につながったような世界をつくることを可能とする。実際、世界においてはDBpediaを中心として500以上のデータセットが相互につながっている。国内においては著者らが2012年よりDBpedia Japaneseを公開・運営することで、同様の活動を広めている。実際、2012年以降、50以上の研究報告があり、20件のデータセットがリンクされており、15件のアプリケーションが報告されている。またその内容も生物学から文化的コンテンツまで幅広く、様々なデータベースがリンクできることがわかる。

In this article, we discuss how Linked Data is suitable to create cross-disciplinary database with respect to the mechanism and the existing datasets. Linked Data mechanism is to provide "web of data" by describing data with RDF while the traditional Web pages consist of "web of documents" with HTML. In particular, DBpedia, LOD of Wikipedia, has accelerated creation of other LODs since

DBpedia works as a hub to link cross-disciplinary datasets. We have launched DBpedia Japanese with expecting the same effect in Japan. Indeed, we observed many new activities to expend Linked Data cloud in Japan, more than 50 technical reports, 20 directly linked datasets and 15 applications.
キーワード: Linked Data, Linked Open Data (LOD), DBpedia, DBpedia Japanese, 分野横断型データベース
Keyword: Linked Data, Linked Open Data (LOD), DBpedia, DBpedia Japanese, cross-disciplinary database

1 はじめに

Linked Open Data(LOD)とはWeb技術、ことにセマンティックWebの技術を用いたデータの公開・共有の技術を持いたデータベースのことである。World Wide Web (以下Web)がその技術によって、世界中の文書がインターネットを介して繋がった世界(Web of Documents)を作ったのに対して、LODはデータの繋がった世界(Web of Data)を作ることを可能としている。Webの画期的な点は、これまで本や個別の文書に閉じていた情報同士の関係をオープンかつフラットに結びつける仕組みを提供したことであり、このことによりこれまでの社会にあった様々な部門や組織、分野にあった様々な情報が一様に結びつくようになった。LODも同様にこれまでの部門や組織、分野などを超えてデータが結びつくことを可能としている。

本稿では、まずLODの概要を述べた上で、LODが実際に使われている状況についての報告を行う。ことに日本においてはこの2年ぐらいでLODが急速に利用されるようになっている。その点について調査結果を報告する。

2 Linked Open Data (LOD)とは

セマンティック Web とは Tim Berners-Lee 氏が提唱した現行の Web より高度に知識を記述できる Web をつくるというビジョンである。そのポイントは Web のグローバルな情報共有空間はそのままに、その上に標準的なメタデータの記法である RDF やそのスキーマを記述する言語 (RDF Schema や OWL) を用意することで、グローバルに知識を共有する仕組みである。Linked Open Data (LOD)は基本的にこのセマンティック Web の技術を使っている。ただし、記述する対象が文章からなる文書的情報(HTML 文書)ではなく、データであることだけが違いである。

LOD を構成する仕組みが簡単である。まず、RDF(Resource Description Framework)という言語で全てを書く。RDF はとても簡単な言語で、「主語」「述語」「目的語」に相当する 3 つ組で全ての情報を書く言語である。例えば「A さんは B さんを知っている」という関係は「A knows B」という 3 つ組で表現される。このとき表現したい事物(例えば、前述の例では A、B) には個別の URI を与える。URI は URL を一般化したもので URL も含んでいるものであるが、URL は Web ページの場所を示す (その URL を見にいけば Web ページがある)のに対して、URI は必ずしも Web ページがなくてもよい。ここでは個々の事物に URI を振ることで、URI が世界中でユニークな ID として使えることが重要である。また事物だけでなく関係も URI で表現することができる。そうすることで、ある 3 つ組の中で使っている関係 (例えば knows) が別の 3 つ組の中で使っている関係と同一であるというこ

とを示すことができる。
原理は基本的にこれだけである。このRDFで表現された3つ組の情報を組み合わせていくことで、複雑なデータも記述することができるし、さらには異なるデータセットの中にあるデータ間の関係も書くことができる。

3 LODが作る世界

URIとRDFを使うことでデータはデータがどこに含まれているといったことを気にせずに相互につなげることができる。これは新しいデータの世界である。Tim Berners-LeeはこのLODを普及させるために4つの原則を提唱している。

1．事物をURIを使って名前付けしよう
2．名前の参照がHTTP URIでできるようにしよう
3．URIを参照したときに関連情報が手に入るようにしよう
4．外部へのリンクも含めよう

この原則に基づくとデータセットは相互につながり、データセットのネットワークができる。これをLOD Cloudと呼んでいる（図1参照）[1]。図1において、丸が個別のデータセットを示し、データセットとデータセットを結んでいる線は、データセット内のデータ同士にリンクがあることを示している。
この図ではデータセットを9種類に分けている。政府関係(Government)、出版関係(Publications)、ライフサイエンス(Life sciences)、ユーザ生成コンテンツ(User-generated content)、分野横断(Cross-domain)、メディア(Media)、地理関係(Geographic)、ソーシャルWeb(Social web)である。
中心にあるのはDBpediaで、Wikipediaの情報をLOD化したデータセットである。Wikipediaは百科事典なので、様々な分野の項目が含まれている。このため、他の多くのデータセットと結びつきやすい。このため、DBpediaはこの図の中心にある。
なおこの図にあるはData Hubと呼ばれるデータカタログに登録されたデータセットのうち、以下の基準を満たしたものである[Cyganiak 11]。

1. 解決可能なhttp://(またはhttps://) URIsでなければならない。
2. content-negotiation等でよく使われるRDF形式（RDFa, RDF/XML, Turtle, N-Triples）のいずれかでRDFデータを解決できなければならない。
3. 1000トリプル以上含んでいる。
4. 他の既存データセットとのRDFリンクが50以上ある。
5. RDFクローリングまたはRDFダンプ，あるいはSPARQLエンドポイントによってデータセット全体にアクセスできる。

加えて
6. 認証なしかつ無料でアクセスできる。
も事実上の基準である。この図において500個以上のデータセットが相互にリンクし合っている。

4 DBpedia Japaneseの運用

先に述べたようにDBpediaは様々なデータセットからリンクされるLODのハブとして機能している。このようなハブがあることが

[1] "Linking Open Data cloud diagram 2014, by Max Schmachtenberg, Christian Bizer, Anja Jentzsch and Richard Cyganiak. http://lod-cloud.net/"

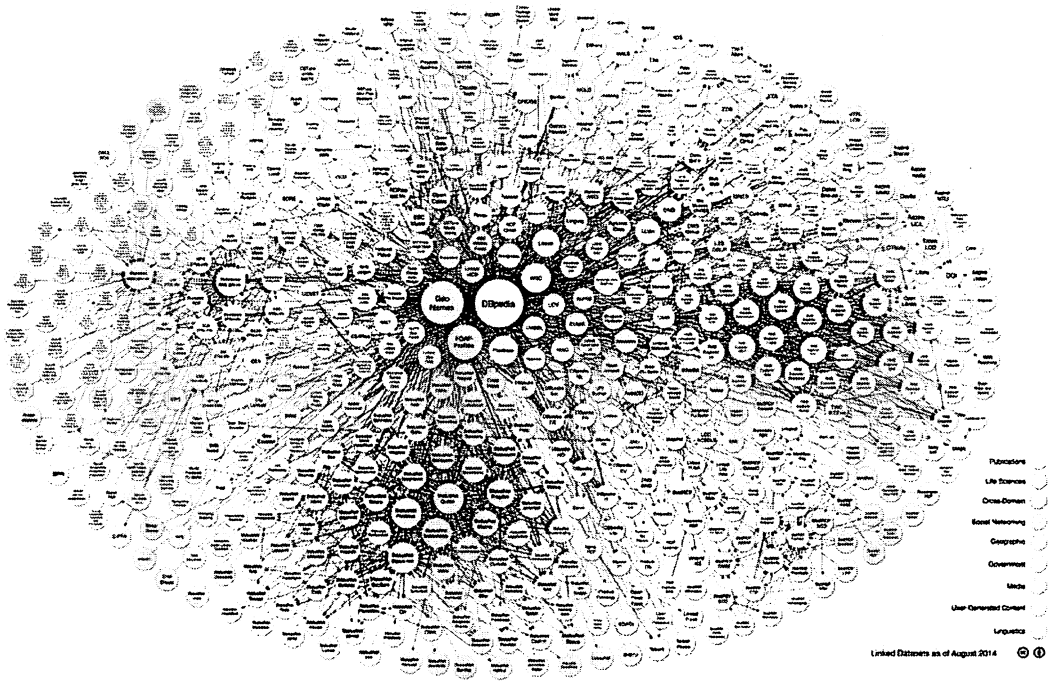


図 1：LOD Cloud

LODの相互参照を促す仕組みとして有効である。このため、我々はDBpedia Japaneseの運用を行なっている。

DBpediaは英語版Wikipediaから作られている。このため英語版Wikipediaにある項目は存在し、その情報もLODとして提供されている。さらにWikipediaの言語間リンクの情報を利用して、もし項目に各国語版がある場合、DBpediaはその各国語版での表記をその言語のラベルとしてRDFとして記述している。このためある程度日本語で表記された項目を対応づけることはできる。しかし、日本版Wikipedia独自の項目は当然存在しないし、また各項目の属性や属性値も英語版と日本語版では異なるにもかかわらず、英語版から取得した属性・属性値しか用意されていない。このため、我々は日本語Wikipediaから構築するDBpedia Japaneseを運用することにした。

DBpedia構築に当たってはDBpedia Information Extraction Framework[2]と呼ばれる一連のソフトウエアを使うことで、Wikipediaからほぼ自動的に構築することができる。これは言語の異なるWikipediaに適用可能である。

ただし、いくつかのことはDBpediaを有用にするために言語ごとに行う必要があった。

- i18n(国際化)に関わるコード改善
- Wikpedia内の表記に依存した処理の対象言語への置き換え（例：日付の処理）
- オントロジーの対応（新しい概念の追加と既存の概念とinfoboxやその属性への対応付け）

2012年から運用を始めている。Wikipediaのほうは日々更新されているものの、DBpedia

2 https://github.com/dbpedia/extraction-framework

Japaneseは現在は自動更新ではなく、適宜更新されている。

5 DBpedia Japaneseの利用

先に述べたようにDBpediaは様々なデータを結びつけるハブとして期待されている。2012年から論文等でDBpedia Japaneseが利用されている件数を図2に示す[3]。現在のところ、全体で58件であった。

データセットに関しては20件であった（図3参照）。出版・文化関係、生物学・ライフサイエンス、政府関係が数的に多く、これは本家のLOD Cloud掲載のデータセットの傾向と似ている。

利用しているアプリケーションとしては15件を発見している（表2）。アプリケーションは汎用ツール（検索・可視化など）と特定分野の検索に大別される。とくに特定分野の検索においては音楽、スポーツから歴史まで様々な分野に適用されていることがわかる。これはDBpediaの汎用性が活かされているといえる。

6 日本におけるLOD

図1に示したLOD Cloudには国内で作られ、公開されたLODはほとんど含まれていない。国立国会図書館が公開する典拠と件名標目（Web NDL Authorities）とDBpedia Japaneseのみが掲載されている。これは本家LOD Cloudは事実上オープンデータが求められていたり、一定数の外部リンクが必要などがあるため、国内のデータセットがなかなか基準を満たせないことがある。

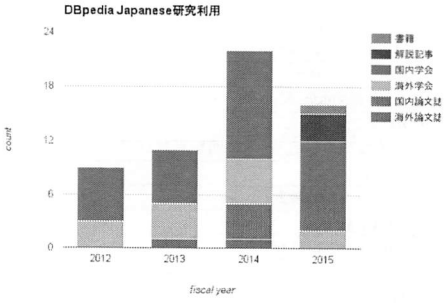


図2：DBpedia Japaneseの研究利用推移

そこで、我々はやや基準を緩めてリンクする国内のデータセットを調査した[2]。

- データ公開者が日本にいる人・組織等である
- 日本語ラベルを含んでいる
- 1000トリプル以上含んでいる
- 本家図かJLDC図の既存のデータセットとのRDFリンクが10以上ある
- 参照解決可能な状態，データダンプ，あるいはSPARQLエンドポイントのいずれかによってデータセットを公開している

この基準に従ったデータセットのネットワークを図4に示す。

7 LOD化に関する考察

7.1 繋がるデータのメリット

[3] DBpedia Japaneseが利用されているかどうかを知るの容易でなく、関連する学会、国際会議や検索などを用いて調査した。このため、必ずしも網羅しているといえない。

表 1：DBpedia Japanese にリンクしているデータセット

データセット名称	分野	URL
J-GLOBAL knowledge	出版、文化(Publication)	https://stirdf.jglobal.jst.go.jp/
LODAC Museum	出版、文化(Publication)	http://lod.ac
日本語Wikipediaオントロジー	汎用(Cross-domain)	http://www.wikipediaontology.org
LODAC Species	生物学、ライフサイエンス (Life Science)	http://lod.ac/species/
京都国際マンガミュージアム書誌情報LOD	出版、文化(Publication)	http://mdlab.slis.tsukuba.ac.jp/lodc2012/kmm/
GeoLOD	地理(Geographic)	http://geolod.ex.nii.ac.jp/
ライフサイエンス辞書	生物学、ライフサイエンス (Life Science)	http://lsd.dbcls.jp
WordNet-ja	汎用(Cross-domain)	http://wordnet.jp
Open DATA METI LOD	政府(Government)	http://datameti.go.jp
東日本大震災アーカイブFukushima	政府(Government)	http://fukushima.archive-disasters.jp
議員LOD	政府(Government)	http://mdlab.slis.tsukuba.ac.jp/lodc2013/senkyo/
ヨコハマ・アート・LOD	出版、文化(Publication)	http://fp.yafjp.org/yokohama_art_lod
青空文庫 Linked Open Data	出版、文化(Publication)	http://mdlab.slis.tsukuba.ac.jp/lodc2012/aozoralod/
NHK映像マップみちしるLOD	メディア(Media)	http://mdlab.slis.tsukuba.ac.jp/lodc2013/michishiru/
ねじLOD	産業(Indutry)	http://monodzukurilod.org/neji/
LSJ: Location Site of Japanimation	地理(Geographic)	http://cheese-factory.info/
環境リポジトリプロトタイプシステム	生物学、ライフサイエンス (Life Science)	http://rihnexers.chikyu.ac.jp
Geonames.jp	地理(Geographic)	http://geonames.jp
lod4all自治体名からdbpediaのエントリーにリンク	地理(Geographic)	http://lod4all.net/lodc/city-comp
Evacva	政府(Government)	http://evacva.net

表 2：DBpedia Japanese を利用したアプリケーション

名前	分類	URL
ここペディア	地図ナビゲーション	https://play.google.com/store/apps/details?id=com.midoriit.kokopedia&hl=ja
プロ野球囲碁サーチ	特定分野検索(スポーツ)	http://inmen-search.com
MusicSPARQL	特定分野検索(音楽)	http://musicsparql.appspot.com
日本の大学	特定分野検索(大学)	http://uedayou.net/SPARQLTimeliner/view/?url=1385715471-802.json
SPARQL Timeliner	検索・可視化ツール	https://github.com/uedayou/SPARQLTimeliner, http://uedayou.net/SPARQLTimeliner/
n次の隣たり	検索・可視化ツール	http://nzinohedatari.azurewebsites.net/
DBpedia Japanese簡単サーチ	検索・可視化ツール	http://lodosaka.hozo.jp/EasyLOD/index_dbpedia_ja.html
Wikipedia Category Consistency Checker	検索・可視化ツール	http://wnews.ist.hokudai.ac.jp/wc3jp
LOD4ALL	検索・可視化ツール	http://lod4all.net
電子書籍リマインダー	検索・可視化ツール	http://pinpoint-reminder.appspot.com/epub-intro
DashSearch LD	検索・可視化ツール	http://www.ahlrulab.com/DashSearchSparql/
疾患コンパス	特定分野検索(疾患)	http://lodc.med-ontology.jp/
バイオミメティクス・オントロジーによるキーワード探索システム	特定分野検索(生物)	http://biomimetics.hozo.jp/
東日本大震災アーカイブFukushima	特定分野データベース	http://fukushima.archive-disasters.jp
ウィキ町史ビュアー	特定分野検索(歴史)	http://oc-study.appspot.com/my-town-timeline

1 章で述べたように LOD を使うことで繋がるデータを公開できる。データベース公開側にとって二つの意味がある。一つは、外部へのリンクを含むことで、自分のデータベース利用者の便益を上げるということである。例えばある美術作品のメタデータを公開している時、その作者を文字列だけで示すのではなく、DBpedia の該当項目へリンクをつけることで、データベース利用者はより便利に使うことができる。に番目の意味は、他のデータベースから自分のデータへリンクを貼ってもらうことによる、データベースの利用の促進である。自分のデータベースにある美術作品が他で言及されている時に、この項目にリンクをつけてもらえれば、これまで自分のデータベースの利用者でない人も結果的に利用することになる。

7.2 LOD に向くデータベース

どんなデータベースでも原理的に LOD 化することは可能であるが、ときに LOD 化のメリットが大きいデータベースは以下のようなものである。これは前章のメリットに対応するものである。

1. 自身がハブ的データになっているもの：すなわち、他から参照される利用の仕方が多いものは、LOD 化のメリットは大きい。DBpedia もそうであるし、辞書なもの、あるいは分野ごとの基礎的なデータベース（例えば医薬品のデ

ータベース、生物種のデータベース、人名データベース）などが含まれる。

2. 外部参照を含むことで自分のデータの価値が上がるもの：美術品データベース、書誌などは典拠や人名などのデータベースにリンクすると自身の利用価値が上がる。
3. そもそも構造がネットワーク的なデータベース：外部参照、被参照がなくて、例えばSNSのデータなど、自分自身がネットワーク構造のデータあるものはLOD化したほうが利用しやすいデータベースになる。

7.3 LODを念頭においたときのデータベース設計

このような外部参照、被参照を前提としたとき、データベースの設計も自ずと変わってくる。

1. データのエンティティの単位を明確にする。一つの事象、事物に対応するようなエンティティを用意して、それに明確な方法で（できれば識別子に基づいて）URIを与える。そのエンティティにその事象、事物に直接関係しないようなことは含めないようにする。
2. データ構造を複雑にしない。データ構造が複雑になる場合、往々にして本来分離すべき別の事象・事物をまとめて構造にしていることがある。そういう事物・事象は別のエンティティとして管理し、リンクで関係付ける。
3. データ構造は汎用のスキーマを参照して設計する[4]。そのまま使えなくても対応関係を書くなど、汎用スキーマとの関係を記述することが望ましい。
4. 個々のデータベースはその主題に関するデータ提供に徹する。もし、データベースの主題ではないが、ユーザにとって有用な情報があり、それが他のデータベースによって提供される情報であるならば、他のデータベースへのリンクによって補うようにする。

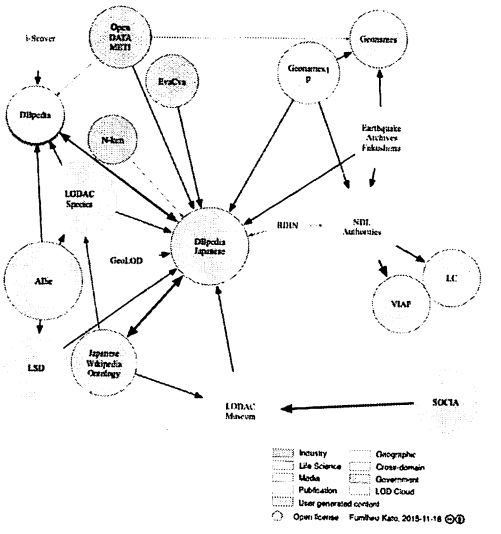


図3：Japanese Linked Data Cloud (JLDC)

これはLODとして利用価値の高いデータベース設計の方法論であるが、またこれは管理しやすく、長期にわたって維持更新のしやすいデータベースの設計方法論でもある。

8. まとめ

本稿ではLinked Dataがいかに分野横断的データベースを作りうるかついて、実例を交えながら説明を行った。データがリンクするこ

[4] 例えば、Linked Open Vocabulary (LOV, http://lov.okfn.org/dataset/lov/)には世界のLODで使われているスキーマが登録されている。また経済産業省/IPAが開発している標準語彙基盤 (http://goikiban.ipa.go.jp/)もある。

とで、様々な分野が繋がるというのは、単に便利になる以上の価値がある。思いがけない分野の組み合わせが生まれ、新しい研究が生まれる可能性がある。このためもまずはよく整備されたデータベースを持っている研究者・機関はそれをLinked Data化することで、このweb of dataの世界を広げる先陣となることが望まれる。

参考文献

[Cyganiak 11] Cyganiak, R, Jentzsch, A: The Linking Open Data cloud diagram, http://lod-cloud.net (2011)

[2] 加藤文彦, 武田英明, 小出誠二, 大向一輝： 日本語Linked Data Cloud の現状, 人工知能学会全国大会(第28回), No. 1G5-OS-19b-7, 松山市 (2014).

2015年情報知識フォーラム予稿

人文社会系大規模データベースへのLinked Dataの適用

― 推論による知識処理 ―

Apply Linked Data to Large-Scale Humanities Database

後藤真*
Makoto GOTO*

1 国立歴史民俗博物館
National Museum of Japanese History
〒285-8502 千葉県佐倉市城内町 117 国立歴史民俗博物館
E-mail: m-goto@rekihaku.ac.jp
*連絡先著者 Corresponding Author

人間文化研究機構では、2008年よりnihuINTという、機構を構成する6機関のデータベースの横断検索システムを提供してきた。その後、7年ほど経過し、データが大量になる中で、より効果的な情報発見が求められることとなった。そこで、プロトタイプとして、Linked Dataによるプロトタイプを作成し、実験に供している。特に荘園のデータベースという日本史の文献と、国文学の古典籍・緯度経度情報をマッチさせ、それぞれにリンクをはることで、より高度な情報発見を可能とできるように試みた。そのうえで、いくつかの実際の研究シナリオを作成し、人文系研究において効果があるかどうかの検証を開始しつつある。実際に情報発見の新たなモデルが提唱可能である見通しを述べると同時に、リンクの付し方など、大規模なデータベースとして実用可能なものとなる水準にするための、課題についても浮き上がってきた。

The National Institutes for the Humanities (NIHU) have been developing an integrated retrieval system (nihuINT) using existing research databases and the sharing of resources for application purposes since 2005. However, 10 years have elapsed since nihuINT was constructed, and some problems have been identified.

To solve these problems, we began to construct a database as a prototype for a new nihuINT. We focused on two databases for our experiment. One deals with "rare Japanese classical books," and the other deals with "Shoen" (a field or manor in Medieval Japan). We took these databases and rebuilt them as RDF (Resource Description Framework)Store/RDF databases.

キーワード: 人文系大規模データベース, Linked Data, RDF, Semantic Web, 歴史系データベース
Large-Scale Humanities Database, Linked Data, RDF, Semantic Web, Historical Database

1 nihuINTが可能にしたことと課題

人間文化研究機構では、機構を構成する6機関(国立歴史民俗博物館・国文学研究資料館・国立国語研究所・国際日本文化研究センター・総合地球環境学研究所・国立民族学博物館)を横断するデータベースを2008年より公開している(nihuINT)[1]。これは6機関が持つ、およそ150のデータベースを同時に検索可能とするものであり、2015年11月現在、およそ160のデータベースの横断検索が可能となっている。レコード数は550万を超え、人文系のデータベースとしては、その多様性からも、数からも有数のものとなっている。

この大量のデータは人間文化の研究を支えるための重要な基盤であり、多くの情報へとアクセスするための重要な手段として位置づけることができるまでになりつつある。

しかし、nihuINTも、システムの構想からおよそ10年、実際に公開しておよそ8年が経過し、さまざまな問題が指摘されている。特に以下のような問題点については、現状のシステムをそのまま維持しているだけでは困難な部分が生まれてきつつある。

例えば、以下のような問題が指摘されている。

①柔軟な統合検索の問題

現在のシステムにおいて、共有メタデータはダブリンコアのみであり、各データベースと共有メタデータとマッピング規則も固定されている。つまり、共有化のパターンが固定されてしまっている状況がある。書誌的データベース同士の共有検索であれば、それなりに有効な部分はあるがたとえば地名辞書の階層情報を共有することは困難である。このような構造を記述することで、より柔軟な検索が可能となると考えられる。

②データベース連携の問題

データベース連携においてはデータベース間でのテーブル結合や副クエリなどの検索操作が必要であるが、現在の共有化システムでは実現されていない。外部データベースとの連携はより困難である状況がある。

③データマイニングの問題

現在のシステムでは、各データベースをプログラムにより検索するAPI(Application Program Interface)が実運用されていないため、データベースごとに検索画面をいちいち開いてキーワードを入力しなければならない。そのため、あるデータから関連するデータを検索するための手間が多く、検索速度の問題と相まって、およそそのような行動がとれないという問題がある。

これらの問題を解決すべく、人間文化研究機構では、資源共有化事業の中にワーキンググループを作り、後述するプロトタイプを作成し、実験を行うこととした。このプロトタイプ構築は、これらの課題を解決するための情報学的な手法を具体的に検討することにあり、次期の資源共有化システムへとつなげることを目指している。もし、情報学的な課題を解決できることとなれば、次期システムへの適用も可能となるため、現在、詳細にその優位性を検討している。

2.プロトタイプの構築

実験対象とするデータベース(人間文化研究機構の内部にあるもの)は以下の通りである。いずれもnihuIntに収録されているものである。

1. 日本荘園データベース(国立歴史民俗博物館)[2]
2. 古典籍総合目録(国文学研究資料館)[3]
3. デジタル地名歴史辞書(総合地球環境学

研究所）

これらのデータベースをRDFストアとして再構築し、新しい共有化システムのプロトタイプとして構築を開始した。また、SPARQLエンドポイントを構築し、検索エンジンとして作成を行った。これにより、上記の問題点を解決することができるか検証を行っている。

具体的には、以下の通りである。

1. 上記データベースの中から、荘園データベースの情報のうちから備中に関係する100件程度の情報を抽出し、それぞれの項目ごとにメタデータを付与し、RDF化した。（図1）
2. 上記抽出情報から、関係する古典籍総合目録データベースの資料情報約200件を抽出。同様にRDFデータを作成した。
3. さらに、デジタル歴史地名辞書の情報を抽出し、同様にRDFデータを作成した。

そのうえで、以下のようなデータベースのリンクによる連結とインターフェース作成を試みている。なお、日本荘園データベースは、以下のような項目を持っている。

荘園コード／国名／郡名／荘園名／重複コード／参考市町村／市町村コード／明治村字名／史料村郷名／領家・本家／初見年和暦／初見年西暦／出典／遺文番号／記録類／地名辞典／備考／関係文献

これらの各項目から以下のような連携を行うこととした。

1. 郡名、明治村字名、史料村郷名から地名の情報を抽出し、デジタル地名歴史辞書とリ

```
<?xml version="1.0" encoding="utf-8" ?>
<rdf:RDF（中略）
        xmlns:rdfp-data-shoen="http://pkg.rdfp.infocom.co.jp/rdfp/load/map/shoen#" >
  <rdf:Description rdf:about="http://rihnexers.chikyu.ac.jp/rdfp/api/data/resource_share_proto/shoen/5210002">
    <rdfp-data-shoen:フリガナ>ニイミ</rdfp-data-shoen:フリガナ>
    <rdfp-data-shoen:出典>東寺百合文書</rdfp-data-shoen:出典>
    <rdfp-data-shoen:出典>教王護国寺文書</rdfp-data-shoen:出典>
    <rdfp-data-shoen:出典>東寺領</rdfp-data-shoen:出典>
    <rdfp-data-shoen:初見年和暦>承久三年</rdfp-data-shoen:初見年和暦>
    <rdfp-data-shoen:初見年西暦>1221</rdfp-data-shoen:初見年西暦>
    <rdfp-data-shoen:参考市町村>新見市</rdfp-data-shoen:参考市町村>
    <rdfp-data-shoen:史料村郷名>併守忠</rdfp-data-shoen:史料村郷名>
    <rdfp-data-shoen:国名>備中</rdfp-data-shoen:国名>
    <rdfp-data-shoen:明治村字>新見</rdfp-data-shoen:明治村字>
    <rdfp-data-shoen:荘園コード>5210002</rdfp-data-shoen:荘園コード>
    <rdfp-data-shoen:荘園名>新見荘</rdfp-data-shoen:荘園名>
    <rdfp-data-shoen:遺文番号>カ３２３３</rdfp-data-shoen:遺文番号>
    <rdfp-data-shoen:郡名>哲多</rdfp-data-shoen:郡名>
    <rdfp-data-shoen:領家本家>相国寺領（東方）</rdfp-data-shoen:領家本家>
    <rdfp-data-shoen:領家本家>皇室領（本家職、大覚寺統伝領）</rdfp-data-shoen:領家本家>
```

図１　荘園データベースから作成したRDFデータのサンプル

ンクする。

2. 出典、記録類から、荘園が出てくる史料名の情報を抽出し、日本古典籍総合目録データベースとリンクする。

3. 荘園データベース同士でも、何度も同じ検索をすることを避けるために、領家・本家の部分で、同一の領家本家の別荘園の様子を一見できるように両者にリンクをはった。これはループすることになるが、一方で同一データベースでも再帰的にリンクを張ることで新たな発見が起こることが期待できるため、あえて実験的に実施した。

また、機構外のデータベースについても、プロトタイプからのリンクを作成し、連携が可能かどうかの実験を試みている。具体的には、以下のようなデータベースにリンクを行っている。

1. 出典から、京都府立総合資料館が所蔵する東寺百合文書Webの個別画像へのリンクを実験的に行っている。

2. 関係文献の論文名から、国立情報学研究所のCiNiiおよび国立国会図書館のNDLサーチへのリンク。

3. 荘園名からDBPediaへのリンク。

これらのリンクを張ることで、全体像は図2のようになった。内部のデータベースから外部の資源へのリンクについては試験的に作成しているものであり、現状では外からの検証は行っていない。ただし、プロトタイプ段階から、1レコードにつき、1URIを付すことも実験しているので、このプロトタイプが公開に至った段階では、外部のデータベースから、これらのデータ

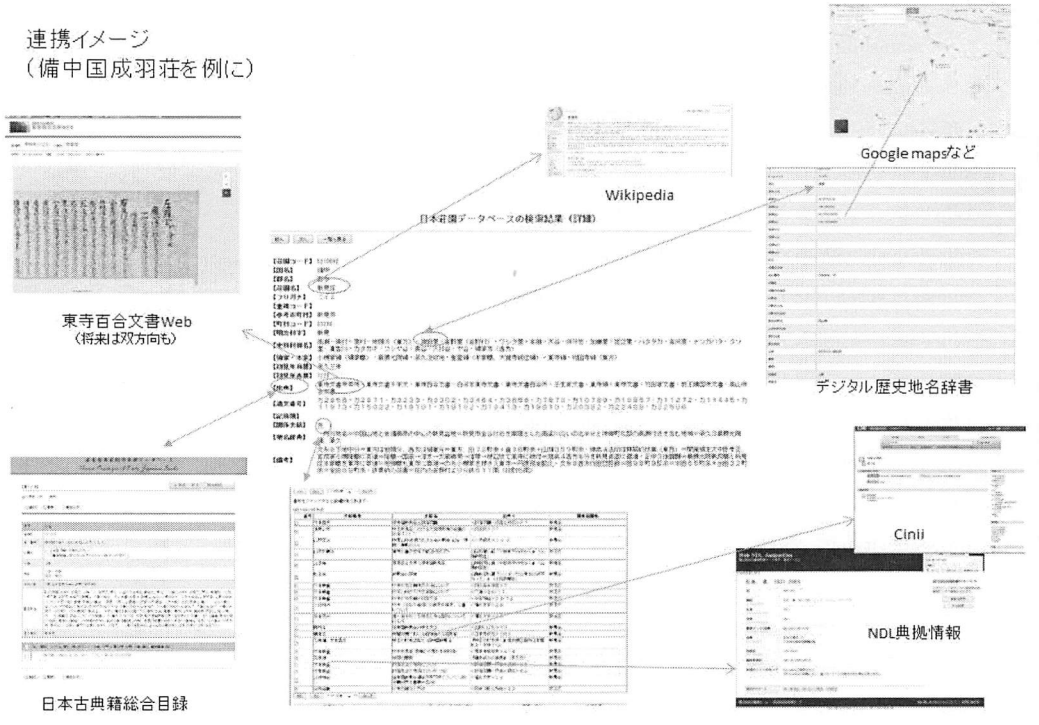


図2 データベースの連携イメージ

ベースへのリンクを行うことも可能となる。

3. 研究シナリオ

これらのプロトタイプをもとに、以下のような研究上の要求シナリオを想定した。それらの要求を満たすリンクの方法と例を示す。

a.中世荘園を調べた際にデータベースで提示された史料の特徴・性質を確かめたい場合。

日本古典籍総合目録データベースと荘園データベースの「出典」「記録類」に関係を作ることを試みることで、荘園情報から容易に資料情報のデータベースへとリンクすることが可能になると思われる。

たとえば、備中国の「井原荘」の情報が出てくる「九条家文書」には、ほかにどのような荘園が出てくるのか、この史料にアクセスするには、どうすればよいかなどが、データベースを行ったり来たりすることなくシンプルに成羽荘の情報にアクセスすればよく、あとは、リンクをたどることで情報を得ることができる。(図3)

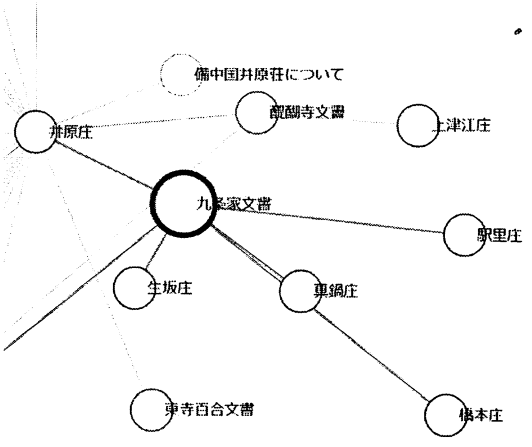
b. 中世荘園に関係すると推定される地名がどのような広がりを持っているか、視覚的に地図で確認したい場合

荘園データベースの「郡名」「明治村字名」「史料村郷名」と地名辞書に関係を作ればよい。たとえば、成羽荘にある地名情報(宮地)から、地名辞書を通じて、緯度経度を確認し地図上で時期的な限界はあるというものの、当該データを見ることができる。

例えば、宮地については、岡山県高梁市のある場所にある。同様に日名という地名であっても、高梁市の一定の場所にあることが確認できる。もちろん、このこと自体は歴史地理学的な研究成果によって明らかにすることができるのだが、これらの情報を、いちいち検索することなく、クリックでリンクをたどり、複数の候補を同時に閲覧できることによって、より容易な情報取得が期待できる。

c. 現在の地名で、ある地域の荘園を探したい。さらに関係する史料も見たい場合

この場合には、aの作業に加えて「郡名」「明治村字名」「史料村郷名」と地名辞書に関係を作ることを試みることにより、可能となる。これ



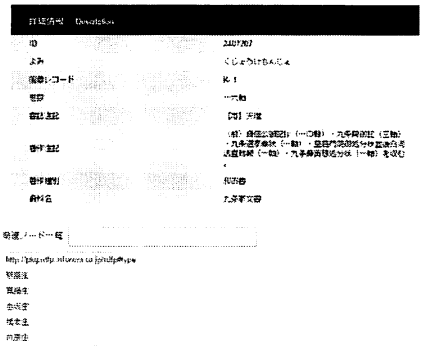


図3 井原荘のリンク先である「九条家文書」のリンク情報 九条家文書に記述されている荘園名が表示されている。右側には所蔵など古典籍総合目録の情報が記される。

はbの逆引きとなる。したがって、具体的には、以下のような成果が期待できることになる。

「現在の岡山県高梁市にある荘園および、その荘園の史料・典籍一覧」という検索条件を、設定し、検索することができるようになる。現在の高梁市→宮地→成羽荘→成羽荘関係の史料という導線での検索が可能となる。研究者の探索行動の一部を代替することが期待できる。

d.ある荘園を治めている領主が、ほかにどのような荘園を治めているかを確認したい場合

この場合は、循環するリンクになるという課題があるが、荘園データベースの領家・本家を再帰的につなぐシナリオが有益である。(図4)

たとえば、ある荘園が皇室領であった場合、ほかに皇室領はどのようなものがあるのか。さらにその皇室領の関係で摂関家領の存在も指摘されていた場合には、摂関家領にどのようなものがあるのか、これらをリンクで一望できることで、当時の荘園の所持関係がある程度確認できるものとなる。

e.ある史料について(荘園関係で)言及している論文や著者の情報を詳しく知りたい場合

古典籍総合目録と、荘園データベースを結びつけたうえで、荘園データベースの「関係文献」の論文名から、CiNiiの論文へと関係を作ることとなる。たとえば、「『東鑑』に出てくる荘園のことを書いている論文の詳細な書誌や(ある場合には)そのPDFデータ」を探すことが可能になる。

f. 荘園データベースの情報にかかわる外部の情報資源を見たい。また、外部の情報資源には、荘園のデータベースの情報資源とかかわるものはないのか確かめたい場合

以下のようなリンクとなる。東寺百合文書Webと荘園データベースの史料名間で対応を取ればよい。たとえば、新見荘に関係する東寺百合文書の画像がリンク等で閲覧できるようになるであろう。

また、1レコード1URIを実装することによって、一般の検索エンジンからも発見が可能となる土台を作ることができ、個別データへの容易なアクセスを実現することになると考えられる。

図 4　ある荘園から皇室領関係の情報のテーブルを抽出し、さらにその中から摂関家領の荘園データを抽出している

4. プロトタイプの課題

本システムで実現できた研究シナリオは上述のとおりであるが、以下のような課題がある。

4. 1. リンクの設定の問題（内部）

Linked Dataのメリットはデータとデータのつなぎによる新たな情報発見である。しかし、どの情報とどの情報を結びつけることが有益であるのかは、項目の内容を検証する必要がある。本検証結果は、荘園名と文献・地理情報など比較的明瞭な目的があったため、その有益性を確認することができた。

今後、nihuINTのような巨大データベースに対応させるためには。これら150を超えるデータベースの項目内容を精査する必要が生じてくるであろう。

また、大量のリンクか自動生成された場合の問題も検討する必要がある。例えば、今回の荘園には、「大谷」という地名がある。しかし、この長谷という地名は、地名辞書では全国におよそ数百あり、これらが正確な情報をどれほど反映しているのか、困難な部分があるであろう。（図5）

とりわけ、最後の「大量のリンク」の問題は解決を要する内容であり、例えば、地名である場合には、ある程度地域の範囲を絞ることで対応するなどの、セマンティクスを考慮した設計を行う必要がある。

4. 2. リンクの設定の問題（外部）

外部リンクでも同様の問題を抱えている。外部リンクの場合、そこへの適切なリンクを設定する場合、機械的に情報を取得する方法が難しい。たとえば、ある荘園からその荘園のことが書かれている東寺百合文書へのリンクを張る場合に、該当する資料を発見することが困難である。一度リンクを作ると、きわめて効果的で有用性の高いシステムになるのだが、そのリンクを作る労力を誰が負担するかという問題は依然として残されている。

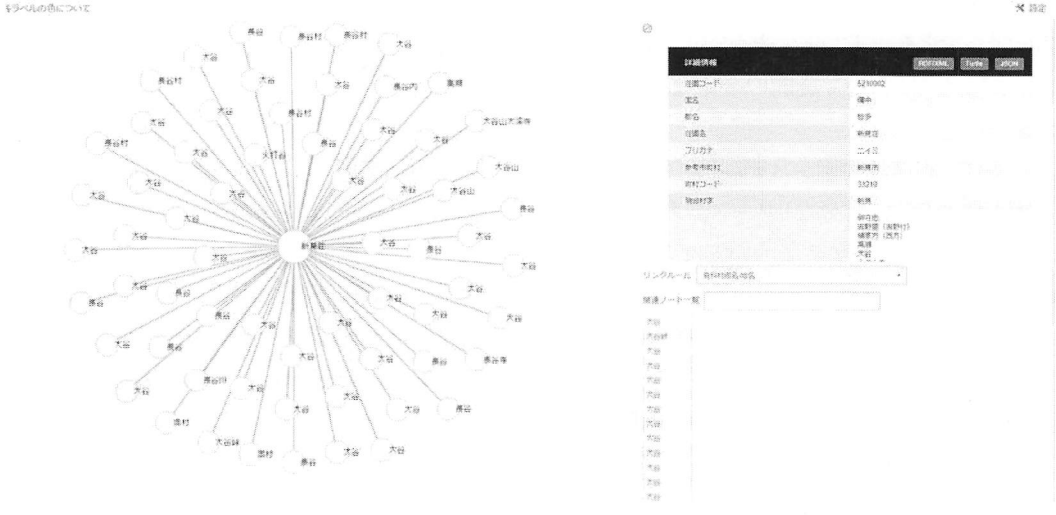


図 5：例えば、大谷という地名で必要以上のリンクが張られてしまうと、ノイズが多くなりすぎ、かえって見にくくなる

5. 大規模データベースとLinked Data

上記のような課題はあるが、大規模データの発見という点においてはLinked Dataのしくみはきわめて有益である。あまりに大規模となり、データの発見が困難になりつつある巨大DBの運用において、Semantic Webの思考方法に代わるモデルは、少なくとも2015年段階では存在しない。無論、Googleのように、ある種のランク付けを検討することは、きわめて重要ではあるが、「たくさんアクセスがあったり、ランク付けが高いものが有益であるとは限らない」人文系のデータベースにおいては、そのようなモデルも困難となる。

また、資料一つ一つが学術資源であると考えた場合、URLをふって、Web上でのエビデンスを確保できるようであれば、それに越したことはなく、さらにそこから論文にアクセスできるモデルは、人文系研究の、より新たなオープン化の形であるといえるだろう。

また、Webリソースに対する検索を一つの機関に依存しないモデルである点も重要である。リソースを公開する機関が、高度な検索システムを開発できるようになるかどうかは、今後、さまざまな状況において疑問の余地なしとは言えない。

一方で、高度な検索システムを開発できる技術を持つ側に、有効なリソースが多数ある状況は少なくとも人文系については考えにくいであろう。そのため、これらシステムとリソースは分割して開発するようなモデルを今後構築していく必要があると考えられる。その際に、より浅いところにあるWebの仕組みを適用することは、今後必ず求められることとなる。機構の資源を、大学の情報工学を専門とする研究者に開放できれば、今までとは異なる発見の方法も期待できるであろう。

これらのRDFによるデータの提供は、より複雑なものやより広範な検索システム・サービスへのデータ提供を可能にするものあるともいえる。大学共同利用機関法人として、各大学での人文系データの発見やそれを支援するための技術開発の重要な資源提供になるとも考えられる。

このプロトタイプの成功が、あらたな情報資源の共有と資源基盤の確立になると考えられる。

謝辞

本論文執筆のためのプロトタイプ構築は、人間文化研究機構本部資源共有化事業のプロトタイプ構築ワーキンググループによるものである。ワーキンググループメンバーである京都大学・原正一郎先生、総合地球環境学研究所・関野樹先生、東京大学山田太造先生には記して御礼申し上げる。本研究はJSPS科研費60191467, 25730199の助成を受けたものである。

参考文献

[1] nihuINT http://int.nihu.jp (2015年11月19日参照。以下同様)
[2] 国立歴史民俗博物館・荘園データベース
https://www.rekihaku.ac.jp/up-cgi/login.pl?p=param/soue/db_param
[3] 国文学研究資料館・古典籍総合目録データベース
http://base1.nijl.ac.jp/~tkoten/about.html

情報知識フォーラム

場所による知識の集積と統合;空間情報科学研究センターの取組

Collection and Integration of Knowledge by Location: Activities of CSIS/UT for Research Promotion

柴崎亮介[1]
Ryosuke SHIBASAKI[1*]

1 東京大学・空間情報科学研究センター
Center for Spatial Information Science, the University of Tokyo)
〒277-8568 千葉県柏市柏の葉5-1-5
E-mail: shiba@csis.u-tokyo.ac.jp

空間情報科学は一般的な情報科学とは異なりデータ駆動型のサイエンス，あるいはテクノロジーという色彩が強い．つまりデータ無しにはなかなか研究が進まない．そのため，東京大学・空間情報科学研究センターは，より地図、地域統計、様々な位置情報など研究利用できるデータを集め、提供するという活動を行ってきた．それらから得られた研究成果がさらに研究を進展させる基盤となるように、研究成果に限らず地図上に展開された様々な情報や知識を、地図を媒介に整理・統合し、あるいは新たな知識獲得を支援する参照データとして提供するなどの活動を行っている．本稿はこれらについて紹介する．

Spatial information science is a broad discipline originating geography, driven by data on real world phenomena. To promote the science and technology, it is very necessary to provide data platform where researchers can easily find base maps, statistics and fragmentary data material describing the changes or dynamics of the real world. Center for Spatial Information Science (CSIS), the University of Tokyo, provides such a data platform. To further promote or encourage the advances of the science, CSIS collect, organize and integrate research achievements and data/knowledge product based on "space" or location. This article introduces such activities of CSIS, UT.

キーワード: 空間、知識の俯瞰化、データ統合、参照データ、研究促進
Keywords: Space, Overview of knowledge, data integration, reference data, research promotion

1 はじめに

本空間情報科学研究センターは 1998 年に設置された研究センターである．センターメンバーが研究を推進するというミッションに留まらず、当初より全国の研究者に空間情報科学を「使ってもらう」あるいは本来の専門分野と併せて研究してもらうことを大きな目標に掲げて活動をしている．空間情報科学は一般的な情報科学とは異なりデータ駆動型のサイエンスあるいはエンジニアリングという色彩が強い．つまりデータ無しにはなかなか研究が進まない．そのためより制約少なく研究利用できるデータを集め、提供するという活動を行ってきた．データはいわゆる学術的な調査・研究等で得られるものに限らず、商用目的で提供されているものも数多く含まれている．商業プロダクトの研究利用については、空間情報科学の底辺拡大、深化の重要性をご理解いただいた上で、空間情報科学研究センターのコントロール下で外部提供可能な契約を、データ提供各社を結ぶことができたことが非常に大きい．

データ提供を通じて蓄積された研究成果は毎年のシンポジウム（CSIS Days）で発表される．これはもちろんデータを提供いただいた企業等に成果をオープンな形でフィードバックするという意味もあるが、同時に、さまざまな分野の研究者がデータを媒介としてつながる機会となっている．また空間情報科学研究センターにとっても新たなデータ整備への要望をいただいたり、データ利用研究の方向性を俯瞰したりと，貴重な機会になっている．

こうした活動は、新しい社会公共的な利用、さらに産業的な利用、今風に言えばニューエコノミーの創造へもつながるが、これは産官学で運営される G 空間情報センターが展開することになっている．なお G 空間情報センターは来年度の立上げが予定されている．

本稿では空間情報科学研究センターの研究あるいは研究支援活動の中で、場所に紐付けた情報や知識の整理、提示方法に関する成果、あるいは知識等を生成するために必要なレファレンスデータの整備・公開事業等について報告する．

2 GIS関連論文の引用･被引用関係に関する空間解析(小野2015)

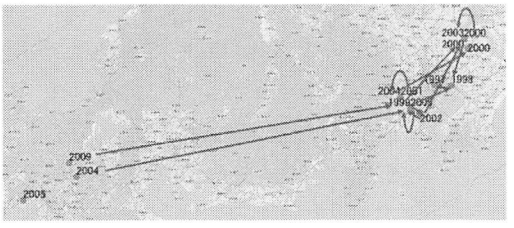
近年のテキストマイニングやビックデータ解析技術の進歩により、多量の学術論文を集合的・総合的に扱い、その結果得られる統計情報を解釈することで、従来とは異なる視点による知見を得ることが可能となった（那須川 2014)．

こうした背景のもと、これまでに地理情報システム学会が発行する論文誌「GIS-理論と応用」の研究論文を対象にして、地理情報分野における研究トピックに関する解析を行ってきたが(小野 2014)，特に論文内に含まれる地理空間情報に注目し、その活用可能性について述べる．具体的には、参考文献の引用により関連づけられた論文間のネットワークを論文の全数処理を通して構築し、論文中の地理空間情報を組み合わせることで、時空間上の知識の可視化を試みた．

地理情報システム学会(GISA)が 1993 年に発行された「GIS理論と応用」Vol.1から最新の 2015 年 Vol.23 No.1までの全研究論文を対象とする．

GISA において最大の深さをもつキーワード「地震」に関する地理空間情報を地図上で表現した結果を図 1 に示す．ラベルは

年代を示しており、矢印は引用元を指している．これを見ると「GIS-理論と応用」における「地震」研究では、関西は関東で行われた研究をフォローしているものの、関東は関東内の研究者しかフォローしていない様子などが伺える．



3「人間」中心の地図表現

地図は空間の上に様々な情報や知識を載せることで、それらを俯瞰化し同時に利用者の周辺にどのような情報が提供されているのかを「見える化」するメディアとして重要な役割を果たしている．しかしスマートフォンなどで提供されているウェブ地図は座標系で表現される幾何学的・位置的な正確性を第一に表現した地図を背景としており、利用者にとって直感的にわかりやすいものとは必ずしもなっていない．

図２「手書き」案内マップの例
出典（株）昭文社：『ことりっぷ　ベルギー・オランダ ルクセンブルク』2015年2月13日

また図２に示すように、「手書き地図」は単なる知識の空間的な俯瞰だけではなく、そこを移動する際に期待される経験をストーリーとしても提示している．こうした空間的ストーリーを、実空間を移動する際に直接参照できるように、実空間座標を示すGPSデータを「手書きマップ」にマッピングする研究を行っている(Min 2014)．

具体的には「手書きマップ」と実空間座標を参照点との相対位置（角度、相対距離）を利用してマッピングしている（図３参照）．さらに自らの体験（移動履歴等）情報を地図と結びつけて整理しておき、そこから様々なサービスにデータを提供する情報の自己管理型のプラットフォーム（pTalk）に関する研究(Kaji, 2013)や、さらに自らのパーソナルデータを自ら管理しつつ、利活用を図るための「情報銀行」コンセプトに関する研究(インフォメーションバンクコンソーシアム、2015)等を行っている．

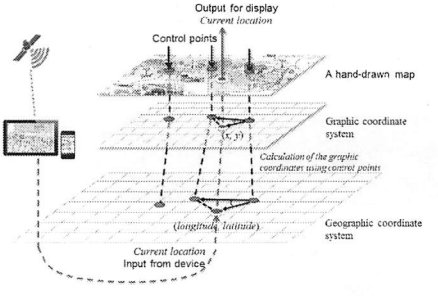


図３　参照点との相対位置による「手書きマップ」と実空間座標との関連づけ

3 人の流れプロジェクト(関本2015)

近年登場した空間データに人や車両等の移動データがある．これらの多くは携帯電話や車載システムに搭載されたGPS（Global Positioning System）などから

得られるデータであり，多数の人々等の移動を詳細に示しているものの，利用交通機関や移動目的などの属性データは添付されていないことがほとんどである．またどのような属性を有する人が携帯する端末からデータがあげられているのかは不明なことが多く，そのため母集団やサンプルバイアスなども明らかではない．そのため，サンプリングスキームが明らかで偏りを定量的に評価できるデータと比較することで，その偏りを明らかにすることが必要である．

また移動軌跡から利用交通手段等を推定するための検証データや学習データも必要となる．移動ビッグデータをその限界も理解しながら適切に利用するためには，上記のような条件を満たす参照データが必要となる．そこでパーソントリップ調査と呼ばれる交通需要調査データを収集し，アンケートにより得られた移動記録データを地図に対応付け，必要に応じて時空間内挿をし，幅広く利用できる参照データとして公開している．

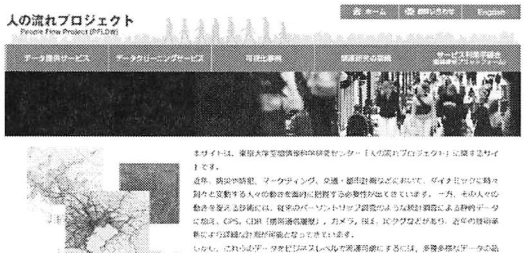


図４　人の流れプロジェクトホームページ（http://pflow.csis.u-tokyo.ac.jp/）

2014年10月現在、日本国内20箇所(22調査)、海外4箇所で合計430万人分を提供しており、公益に資する範囲であれば、産官学のどの立場でも、CSISとの共同研究申請（JoRAS）を通じて取得することができる．データ作成方法の詳細や提供の流れ等は（人の流れプロジェクト，2015）を参照されたい（図４参照）．

参考文献

[1] 那須川哲哉,西山莉紗,吉田一星:学術文献のテキストマイニング，言語処理学会第 20 回年次大会 発表論文集、2014

[2] 小野雅史、柴崎亮介:地理情報科学論文データベースによる研究トピックと地名情報の分析，地理情報システム学会第 23 回学術研究発表大会2014

[3] 小野雅史・柴崎亮介:学術論文マイニングから得られる地理空間情報の可能性，，地理情報システム学会第 24 回学術研究発表大会2015

[4] Min Lu: Human-Centered Cartography in Mobile Environment, Department of Socio-Cultural Environmental Studies, Graduate School of Frontier Science, The University of Tokyo、東京大学博士論文, 2014

[5]Hideki Kaji and Masatoshi Arikawa: Blog based Personal LBS (Location-Based Services), Proceedings of HCI 2013 International, 2013

[6]インフォメーションバンクコンソーシアムhttp://www.information-bank.net/ 2015

[7]関本義秀：人の流れプロジェクトのあゆみ. DICOMO2015招待論文　2015

[8]人の流れプロジェクト http://pflow.csis.u-tokyo.ac.jp/ 2015

第20回情報知識学フォーラム予稿

時間情報基盤の構築と活用－時間による知識処理

Construction of Temporal Information Platform and its Utilization - Knowledge Processing by Time

関野樹[1*]
Tatsuki SEKINO[1*]

1 総合地球環境学研究所
Research Institute for Humanity and Nature
〒603-8047 京都市北区上賀茂本山457-4
*連絡先著者 Corresponding Author

時間情報については，地理情報システム(GIS)のような可視化や解析を統合的に扱う環境が整備されていない．また，地理情報のベースマップや地名辞書に相当する基盤情報も時間情報については未整備のままである．その一方，さまざまな研究分野で時間情報が用いられており，地域研究や環境研究などの学際的な研究分野では，異なる情報同士の接点としても時間情報は重要である．本稿では，時間情報を統合的に扱うソフトウェアツールや基盤データの構築を進めているHuTimeプロジェクトの活動を中心に，時間情報基盤の実情を紹介する．

Integrated environment for visualization and analysis of temporal information is immature, while there is geographic information system (GIS) which is integrated environment for spatial information analysis. Additionally, basic data, corresponding to base map and gazetteer for spatial information, is not available for temporal information. Needless to say, temporal information is an essential element in various scientific fields, and is important to link between different kinds of data in interdisciplinary studies such as area study and environmental study. In this paper, activities of HuTime project which is trying to develop software tools and to construct basic data for visualization and analysis of temporal information are introduced.

キーワード: 時系列，時空間情報，HuTime，可視化，Linked Data
time series, spatiotemporal information, HuTime, visualization, Linked Data

1 はじめに

時間情報は，さまざまな学問分野において必須の要素であり，その重要性は誰もが認識するところであろう．さらに，地域研究や環境研究などの学際的な取り組みが求められる分野では，時間情報が異なる分野の情報を繋ぐ接点として重要な役割を果たす．つまり，データ形式（文字，数値，画像，音声など）や媒体（紙，電子データ，フィルムなど）が異なっていても，同じ時間属性を持っていれば，その関係を類推する手がかりとなり得る．

ところが，時間情報を扱うにあたって，可視化や解析を行う統合的な基盤が十分整備されていない．空間情報であれば，地理情報システム（GIS）が広く普及しており，地域研究のみならず幅広い分野で不可欠のものとなっており，データの可視化，解析，さらに，測地系やデータフォーマットの変換など，空間情報を扱うためのさまざまな機能が統合的に，かつ，GUIにより容易に利用できる．

一方，時間情報では，時系列解析などを行うソフトウェアや可視化を行うソフトウェアなどが存在するものの，それぞれ特定の機能に特化したものがほとんどで，GISのような統合的な環境がない．

H-GIS 研究会[1]の下で開発が始められたHuTimeは[2,3]，このような時間情報の扱いにかかる問題を解決するためのソフトウェアであり，時間情報に関する可視化や解析をGISのように統合的に行う環境を目指した，いわば，「時間情報システム」とも呼ぶべきものである[4]．現在HuTimeは，利用者の端末にインストールして利用するスタンドアロン版（Desktop HuTime）とWebページに埋め込んで使用するWeb版（Web HuTime）があり[5]，それぞれ，地域研究などの学際的な分野での活用されている（地域研究[6,7]，環境[8,9]，保健[10]，歴史[11]）．本稿では，このHuTimeの開発や暦などの基盤データの構築を進め，時間情報の生成から利活用までのすべての過程を総合的に扱うHuTimeプロジェクトの取り組みついて紹介する．

2 時間情報の可視化と解析

2.1 可視化

時間情報の可視化はさまざまな試みがなされ，多様なソフトウェアが開発されてきた[12]．さらに，Webページに年表を埋め込めるSIMILE Timeline[13]やTimeglider[14]，折れ線グラフなどの時系列のグラフを埋め込めるSimile Timeplot[15]，dygraphs[16]，Highstock JS[17]など，多くのWebアプリケーションも提供されている．近年は，ブログやSNSの普及に従って，時系列で写真や動画などの多彩なコンテンツを年表上に表示するdipity[18]やTilo Toki[19]といった各種Webアプリケーションが用いられるようになっている．

しかしながら，これらの多くは，単一の年表もしくはグラフを表示するのみで，複数の年表やグラフを並べて表示できるものはごく少数である．さらに，年表とグラフを同時に扱えるソフトウェアがないため，数値データである経済指標の変化と文字データである政治に関するできごとを同じ時間軸上で並べて比較するといった，研究上しばしば試みられる検討

を容易に行うことができない．このような既存のソフトウェアを研究（特に地域研究のような学際研究）に用いる場合の問題点を解決するために開発されたのがHuTimeである．

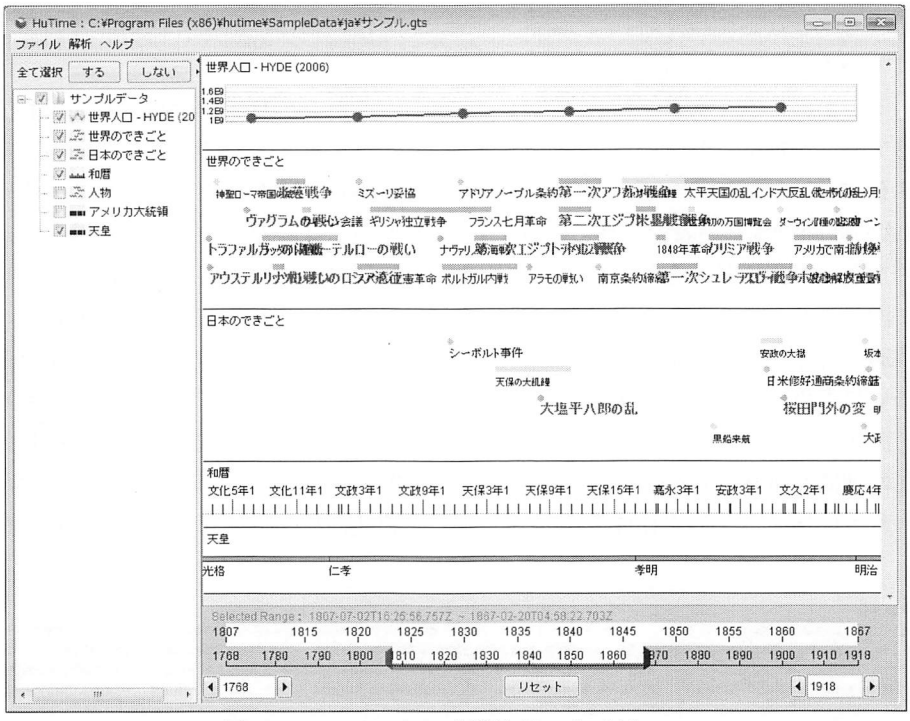


図1 HuTime による時間情報の表示例．

HuTimeでは，文字データは年表として表示され，各レコードは，その期間を示す帯とタイトルで年表内に表示される．期間が定まらないレコード（○○以降など）は期間を示す帯がグラデーションで示される．これらの年表上のレコードをクリックすると，そのレコードの詳細情報を示すウィンドウが開く．データ作成者はここに任意のデータを書き込むことが可能であり，レコードの説明，関連する場所，Web リンクなどを表示することができる．Web HuTime では，Google Maps などの地図や動画などを埋め込むことも可能である．年表上のレコードの書式（色やフォントなど）は，データ作成者が任意に設定可能であり，レコードを複数のグループに分け，それぞれ書式を設定することもできる．

一方，数値データは折れ線グラフ，棒グラフ，プロットグラフのいずれかの形式で表示される．年表と同様に，レコードをいくつかのグループに分け，それぞれ異なる書式で表示することができる．年表内のレコードをクリックした場合と同様に，グラフ内のプロット等をクリックすることにより，詳細情報が表示される．

HuTimeに特徴的な表示の１つに，デー

タ作成者による任意の時間軸目盛がある．一般的なソフトウェアやWebアプリケーションの多くは，日付は西暦（ユリウス/グレゴリオ暦）で扱われる．しかしながら，調査対象となる地域や時代で実際に使われている暦に基づく時間目盛が示されている方が，年表やグラフに表示されているデータを理解しやすいことも多い．図1の例では，江戸期の和暦（太陰太陽暦）による時間軸目盛が表示されている．和暦のように朔（新月）を基準にし，しかも，閏月の存在するような暦は，一般的な西暦による表現では難しいものの，当時の史料を扱う場合には必須のものである．図1のもう1つの例では，歴代天皇の在位期間が帯状の目盛で示されている．これも同様に，データ作成者が任意に作れる時間軸目盛で，時代区分や政権の移り変わりを示すのに用いることができる．また，歴史だけでなく，雨季と乾季，潮の満ち引きなどを時間軸目盛として用いれば，農業や漁業などのデータを扱い場合に有効である．

これらの年表，グラフ，時間軸目盛は，それぞれ，GISのレイヤに相当するものであるが，HuTimeでは，GISのように重ねる（オーバーレイ）のではなく，同じ時間軸上で画面の上下方向に並べられる形で表示される．Desktop HuTimeでは，これらをGISのようにプロジェクトとしてグループ化して管理することも可能である．表示される時間範囲の変更は，HuTime下部に表示されるGUIで操作し，範囲の移動や拡大・縮小が可能である．

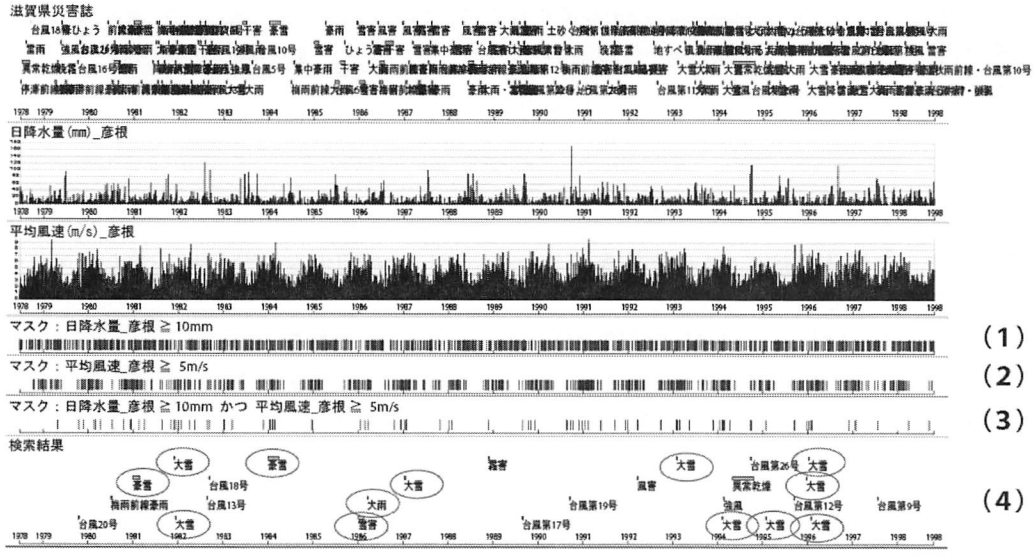


図2 HuTimeを使った時間情報の解析例．「降水量が多く，風速が大きいときに起きた災害はなにか？」という問いに対し，大雪が多いという回答が得られる（一番下の年表(4)の丸で囲まれたレコード）．

2.2 解析

HuTime には，GIS のクリップやユニオンといった機能に相当する解析機能がある．これらは，マスクと呼ばれる時間範囲だけを示す独自のレコードにより進められ，解析はまずこのマスクを検索結果として作成するところから始まる．年表であれば，指定された検索語を含むレコード，数値データであれば，指定した数値以上，以下などの条件を満たすレコードが抽出され，それらの時間範囲がマスクとなる．

図 2 はこれらの機能を使った時間情報の解析例である．ここでは，滋賀県の災害年表[20,21]と彦根気象台の観測データ（日降水量と平均風速）[22]を用い，「降水量が多く，風速が大きいときに起きた災害はなにか？」という問いに対する回答を HuTime の解析機能を使って得ようとしている．まず，「降水量が多く，風速が大きいとき」という条件を満たす時間範囲を得る．日降水量 10 mm 以上の時間範囲と平均風速 5 m/s 以上の時間範囲が検索され，マスクとして示される(図 2－(1)(2))．次に，これらの時間範囲の論理積をとることにより，「降水量が多く，風速が大きいとき」に相当する時間範囲が得られる（図 2－(3)）．最後に，この時間範囲を使って，「・・・ときに起きた災害はなにか？」の回答を得る．このため，得られた時間範囲に含まれる災害を災害年表から抽出し，年表として示す(図 2－(4))．この結果，「降水量が多く，風速が大きいときに起きた災害はなにか？」という問いに対し，滋賀県の場合は台風よりもむしろ大雪による災害が比較的多いという結果が得られる．

3 時間基盤情報

空間情報の場合，データ構築や可視化のためにベースマップや地名辞書といった基盤情報が用いられる．たとえば，地図上にデータを表示する場合には，背景として用いられるベースマップにより，行政界や海岸線，水部や山地などの自然地形が示さる．これにより，地図上に表示されたデータの空間的な属性や相対的な位置を知ることができる．このような基盤情報の必要性は，時間情報でも同じである．つまり，年表上にデータを表示する場合に，時代区分，戦争や災害といった主要な出来事などの情報が背景や隣接する年表として示されていれば，データの時間的な属性や相対的な位置を知ることができる．

代表的な空間基盤情報である地名辞書についても，同様の情報が時間情報で必要である．空間情報では，地名が場所を示す識別子として用いられる．これらの地名から緯度経度などの空間座標を得る操作（ジオコーディング）により，空間情報を地図上に正確に表示できる．この，地名と緯度経度を結びつけているのが地名辞書である．同じように，時間情報でも，できごとの名称が時間を表すために用いられることがしばしばある．たとえば「戦後」といった場合は，1945 年 8 月 15 日以降を指すし，江戸期といえば一般的には慶長 8 年（1603 年）から明治元年（1868 年)の期間を指す．しかしながら，これらのできごとの名称と時間軸上の座標値，つまり年月日とを結びつける地名辞書に相当する一般的な仕組みが無い．

さらに，時間情報で不可欠の基盤情報として暦がある．地域や時代によってさまざまな暦が用いられており，それぞれの暦で記述された時間情報を地域間，時代間で比較するには，それらを相互に変換する仕組みが必要である．これは，空間情報において座標系の変換が必要になることと同様である．

これら３つの時間基盤情報のうち，暦に関しては，Web 上で利用可能な資料や変換サービスがあるものの，基盤年表や時間名辞書を含めた時間情報のための体系的な整備はほとんどなされていない．このため，HuTime プロジェクトでは，HuTime の開発に加え，時間基盤情報の構築をあわせて進めている．３つの時間基盤情報のうち，基盤年表や時間名辞書については，試験的なシステムが構築され，時間名辞書で得られた情報を直接HuTimeに表示するなどの機能が試験的に構築された[23]．だが，暦に関する情報が他の基盤情報を構築するために必須となることから，現在の HuTime プロジェクトでの基盤情報構築は，暦に関する情報が優先的に進められている．

暦に関する情報として，まず，和暦と西暦を対応させることが行われた．一般的にも，西暦が標準的な暦として用いられているものの，問題点も多い．西暦と呼ばれているものの実体は，ユリウス暦とグレゴリオ暦を組み合わせたものである．ユリウス暦が想定する１年（365.25日）と実際の太陽年の長さ（365.2422）との違いを修正するために，１年の長さを 365.2425 日としたグレゴリオ暦が 16 世紀末に採用されたのだが，これが日付を表記するにあたっての問題の原因となっている[24]．最も大きな問題は，ユリウス暦からグレゴリオ暦の改暦の時期の違いであり，最初に改暦が行われたイタリアやスペインはユリウス暦 1582 年 10 月 4 日の翌日をグレゴリ暦 10 月 15 日とした．また，イギリス連邦では，ユリウス暦 1752 年 9 月 2 日の翌日をグレゴリオ暦 9 月 14 日とした．ちなみに，コンピュータはデフォルトでは後者の改暦時期を採用しており，Linux ではコマンド“cal 9 1752”を実行すると，確認できる．この改暦時期の違いにより，同じ日付であっても異なる日を指していることがしばしば生じる．このため，日付表記の標準的な規格となっている ISO 8601 では，グレゴリオ暦への改暦以前に遡ってグレゴリオ暦を適用する先発グレゴリオ暦（proleptic Gregorian Calendar）を採用することとなっている[25]．しかしながら，この規定も徹底されているわけでなく，ユリウス暦の日付を ISO 8601 形式で表記するといったものも散見される．

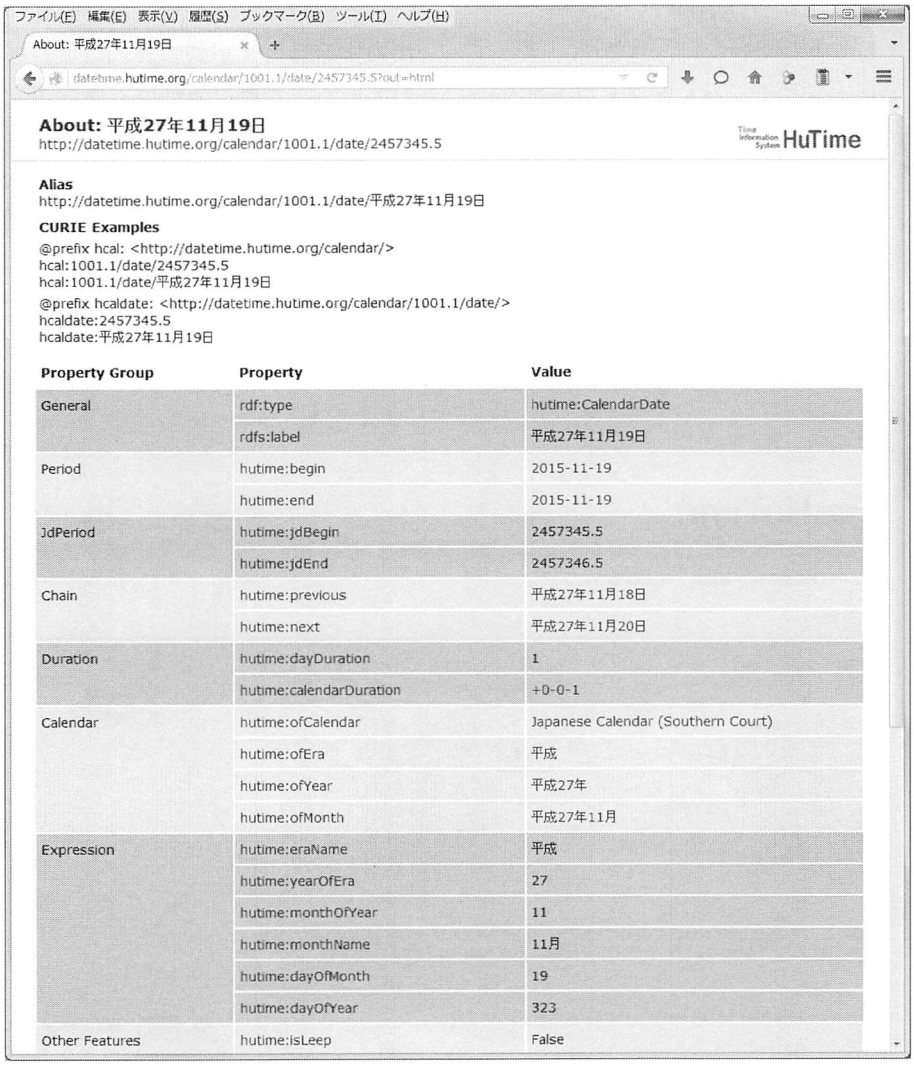
このような混乱を避けるため，HuTime プロジェクトでは，標準の暦としてユリウス通日を用いている．ユリウス通日は，紀元前 4713 年 1 月 1 日正午からの通算日数で，小数点以下で時刻を表現することもできる[24]．また，改暦などによる不連続がなく，実数で表現されるため，前後関係の比較が容易であるといった特徴もある．HuTime プロジェクトでは，上記のユリウス/グレゴリオ暦を含むさまざまな暦をこのユリウス通日と対応付けることで暦の変換を実現している．

図３は HuTime プロジェクトの Web サイトで公開されている暦変換サービスである．このサービスには２つの大きな特徴

がある．１つは，日付を表す文字列を解釈する機能である[26]．暦を変換するための様々なサービスがWeb上に存在するが，その多くは，年，月，日を別々のテキストボックスに入力することを要求する．しかしながら，この入力作業が手間であり，多くのデータを変換する際に障害となる．また，変換しようとする日付の年月日の区切りが分からないとデータを入力することができない．つまり，「平成二十七年十二月十二日」をグレゴリオ暦の日付に変換しようとすると，どこまで年で，どこまでが月かが分かること，そして，漢数字をアラビア数字に直せることが要件となる．日本人であれば問題はないが，漢字が読めない外国人にとっては全く使えないサービスとなってしまう．HuTimeプロジェクトの暦変換サービスでは，この「平成二十七年十二月十二日」を直接読みとって解釈し，年号，年，月，日を抽出して変換データとして用いることができる．したがって，漢字が読めなくても，日付と考えられる文字列をコピー&ペーストすれば利用することができる．さらに，この機能は拡張されて，出力する日付文字列の書式指定も可能にしている．和暦であれば，漢数字や干支を用いるなどの指定を利用者が自由に行うことができる．このため，暦変換だけではなく，同一の暦で書式を揃えるためにもこのサービスが利用されている．

図3　HuTimeプロジェクトの暦変換サービス
(http://www.hutime.jp/basicdata/calendar/form.html)．

この日付文字列を解釈する機能はもう一方の特徴を生み出すきっかけとなっている．つまり，年，月，日と分かれたテキストボックスが不要となることで，複数のデータを一度に変換することができるようになる．入力データのテキストボックスに，改行で区切った複数の日付文字列を入力すれば一度に変換処理が行われる．現在は，ユリウス/グレゴリオ暦，和暦，明治以降の太陰太陽暦（いわゆる「旧暦」），ユリウス通日に正式対応しており，そのほかに，ヒジュラ暦（イスラム暦），ユダヤ暦，タイ仏暦が試験版として利用可能である．



Property Group	Property	Value
General	rdf:type	hutime:CalendarDate
	rdfs:label	平成27年11月19日
Period	hutime:begin	2015-11-19
	hutime:end	2015-11-19
JdPeriod	hutime:jdBegin	2457345.5
	hutime:jdEnd	2457346.5
Chain	hutime:previous	平成27年11月18日
	hutime:next	平成27年11月20日
Duration	hutime:dayDuration	1
	hutime:calendarDuration	+0-0-1
Calendar	hutime:ofCalendar	Japanese Calendar (Southern Court)
	hutime:ofEra	平成
	hutime:ofYear	平成27年
	hutime:ofMonth	平成27年11月
Expression	hutime:eraName	平成
	hutime:yearOfEra	27
	hutime:monthOfYear	11
	hutime:monthName	11月
	hutime:dayOfMonth	19
	hutime:dayOfYear	323
Other Features	hutime:isLeep	False

図4 HuTime プロジェクトの暦に関する Linked データ（http://datetime.hutime.org）.

これらの暦に関するデータを Linked Data として公開する取り組みも進められている[27,28]. Linked Data において基礎となる RDF では，日付はリテラル値として扱うものとして規定されている[29]. 一方で，リテラル値は RDF の主語になることができない．したがって，ある出来事が起きた日付を RDF で表現することは容易であるが，反対に，ある日に起きたさまざまな出来事を RDF で表現することは難しい. また，日付のリテラル値は ISO 8601 形式で表現することになっているため[30]，和暦を含む地域や時代に固有の暦に基づく日付を表現することもできない. このため，HuTime プロジェクトでは，日をリテラル値ではあく，時間範囲を持つリソースとして扱い，そのリソースにさまざまな情報を付加する方法をとっている．すでに暦変換サービスでサポートしている各暦法について，Linked Data の公開が始まっており，日だけでなく，年号，年，月に関するデータも提供されている（図 4）．この HuTime プロジェクトによる暦の Linked Data はクリエイティブ・コモンズライセンス，“CC By”の下での公開である.

4 今後の課題

HuTime は，年表とグラフを同じ時間軸上に表示するという点で，時間情報の独自の可視化を実現した．一方で，いくつかの課題も残っている．１つは，レコード間の関係をうまく表現できないということである．時間情報は，一次元でかつ一方向であることから，２つのできごとの時間的な前後関係が因果関係を知るための手がかりとなる．このようなレコード間の関係をうまく表現した例として，過去には Time Wheel[31]などの試みがある．現在，レコード間の関係を表現するためのしくみとして，HuTime のデータに RDF を埋め込み（XML+RDFa[32]），それを HuTime 上で表現するための改良が進んでいる.

時間情報は，循環性を持つことも大きな特徴であり，周期の抽出や周期間での比較といったことが行われる．この点についても，Spiral Graph[33]や Time Wave[34]などの取り組みがある．一方，HuTime ではこの時間の循環性を活かした表現は，周期を記した時間軸目盛を用いるなどの方法は考えられるが，根本的な対策が必要であり，今後の検討課題である.

時間基盤情報では，HuTime プロジェクトの取り組みにより，特に暦に関する情報が充実してきた．これらの情報を HuTime や他のアプリケーションと連携させるため，API の開発が進められている. この機能により，HuTime との連携にして，任意の基盤年表を呼び出して表示するといった操作が可能になる.

最後に，空間情報との連携が時間情報の活用という点でも大きな課題である. GPS などの位置情報を取得する端末の普及や，各種測器などが生み出す多量のデータ（いわゆるビッグデータ）により，時間情報と空間情報を組み合わせて，時空間情報として扱われる機会が増えている．これらの多くは空間情報を主体にしており，地図上で表現されることが多い. ところが，時間変化や物事の経緯の詳細を異なる場所で比較するといったニーズ

もあるものの，複数の年表や時系列グラフを手軽に比較する環境は十分には整備されていない．HuTime と GIS を組み合わせた時空間情報の可視化や解析については検討されているものの，データ形式や使い勝手の点でまだまだ不完全である[35,36]．今後，時空間情報の解析にHuTimeなどの時間情報に特化したツールが必須となることは確実であり，進展が期待される．

HuTimeプロジェクトでは，今回紹介した活動を通じて，時間情報を扱うための総合的な基盤構築を進めている．これらが地理情報基盤と遜色ないレベルまで充実し，地理情報学と同様に「時間情報学」として体系化されることにより，真の意味で時間と空間が連携した時空間情報の可視化や解析が実現されるものと考えられる．なお，HuTiemプロジェクトの最新の情報は，プロジェクトのホームページで確認できる(http://www.hutime.jp/)．

謝辞

本研究は，JSPS 科研費 基盤研究 (A)「セマンティック・クロノロジー：時間軸に沿った知識の可視化と利用に向けた基盤構築」(15H01723)，および，京都大学地域研究統合情報センター・共同研究プロジェクト(地域情報学プロジェクト)「地域研究データにおける時空間情報の実践的活用」の助成を受けたものである．

参考文献

[1] H-GIS研究会：「H-GIS」，http://www.h-gis.org/ (2015年11月20日参照)

[2] 関野樹：「時間情報システム」，(総合地球環境学研究所編)『地球環境学マニュアル２』 朝倉書店，pp. 116-117，2014.

[3] 原正一郎；関野樹：「時空間情報処理ツールHuTime・HuMapの開発と利用」(HGIS研究協議会編)『歴史GISの地平．勉誠出版』，pp. 13-24，2012.

[4] Sekino, Tatsuki: "Tools and basic data for temporal information analysis", Proceedings of ANGIS and CRMA Bangkok meeting 2015, pp.55-58, 2015.

[5] Sekino, Tatsuki: "Time Information System on the Web", Proceedings of ANGIS Taipei meeting, in press, 2015.

[6] 柴山守：『地域情報マッピングからよむ東南アジア』，勉誠出版，317p.，2012.

[7] 久保正敏：「雲南県誌の分析から：HuTimeで歴史文書の利用を考える」，(関野樹編)『HuTimeを使った時間情報解析の現状，2010年11月12日 H-GIS研究会 報告書』，pp. 5-8，2010.

[8] 関野樹：「琵琶湖の水環境の時間に基づく情報解析」，東南アジア研究，Vol. 46, No. 4，pp. 593-607，2012.

[9] 関野樹：「時間情報に基づく情報の収集と解析」，(秋道智彌；小松和彦；中村康夫編)『水と環境．人と水』，勉誠出版，pp. 74-104，2010.

[10] 福士由紀；東城文柄；顧雅文；西田涼子；駒野恭子；門司和彦；飯島 渉：中国における日本住血吸虫症史研究へのHuTimeの利用．(関野樹編)『HuTime/Mapを使った研究事例と将来展望，2012年3月20日 H-GIS研究会 報告書』，pp. 13-14，2012.

[11] 後藤真：「HuTime/Mapの日本史研究

への応用の試み―続日本紀を題材に―」，（関野樹編）『HuTime/Mapを使った研究事例と将来展望，2012年3月20日 H-GIS研究会 報告書』，pp. 15-20，2012.
[12] Aigner, Wolfgang; Miksch, Silvia; Müller,Wolfgang; Schumann, Heidrun; Tominski, Christian: "Visualizing time-oriented data-A systematic view", Computers & Graphics Vol.31 (2007), pp.401-409, 2007.
[13] SIMILE Project, Massachusetts Institute of Technology: "SIMILE Widgets | Timeline", http://www.simile-widgets.org/timeline/（2015年11月20 日参照）
[14] Timeglider: "Timeglider: web-based timeline software", http://timeglider.com/（2015年11月20 日参照）
[15] SIMILE Project, Massachusetts Institute of Technology: "SIMILE Widgets | Timeplot", http://www.simile-widgets.org/timeplot/（2015年11月20 日参照）
[16] Vanderka, Dan: "dygraphs", http://dygraphs.com/（2015年11月20 日参照）
[17] Highsoft: "Highstock product | Highchart", http://www.highcharts.com/products/highstock（2015年11月20 日参照）
[18] Underlying, Inc.: "Dipity - Find, Create, and Embed Interactive Timelines:", http://www.dipity.com/（2015年11月20 日参照）
[19] Webalon Ltd: "Beautiful web-based timeline software", http://www.tiki-toki.com/（2015年11月20 日参照）
[20] 滋賀県：『滋賀県災害誌第三部』，156p.，1990.
[21] 滋賀県：『滋賀県災害誌第四部』，123p.，2000.
[22] 気象庁：「気象統計情報 過去の気象データ検索」，http://www.data.jma.go.jp/obd/stats/etrn/index.php（2015年11月20 日参照）.
[23] Sekino, Tatsuki: "Basic Knowledge for Temporal Analysis", PNC 2012 Annual Conference and Joint Meetings, UC Berkeley, 2012.
[24] Dershowitz, Nachum; Reingold, Edward M.: "Calendrical Calculations 3rd ed.", Cambridge University Press, 479p., 2007.
[25] ISO: "ISO 8601:2004 Data elements and interchange formats – Information interchange – Representation of dates and times, Third edition", 33p., 2004.
[26] 関野樹；山田太造：「日付を表す文字列の解釈と暦の変換－暦に関する統合基盤の構築に向けて」，情報処理学会シンポジウムシリーズ Vol. 2013, No. 4, pp. 161-166, 2013.
[27] 関野樹：「Linked Dataにおける日の取り扱い－時間に基づくデータ連携」，情報処理学会シンポジウムシリーズ, Vol. 2014, No. 3, pp. 125-130, 2014.
[28] 関野樹：「暦に関するLinked Dataとその活用」情報処理学会シンポジウムシリーズ，印刷中，2015.
[29] W3C: "RDF 1.1 Concepts and Abstract Syntax", http://www.w3.org/TR/rdf11-concepts/（2015年11月20 日参照）.
[30] W3C: XML Schema Part 2: Datatypes Second Edition, http://www.w3.org/TR/xmlschema-2/（2015年11月20 日参照）.
[31] Tominski, Christian; Abello, James;

Schumann, Heidrum: “Axes-based visualizations with radial layouts”, 2004 ACM Symposium on Applied Computing, pp.1242-1247, 2004.
[32] W3C: RDFa Core 1.1 - Third Edition, http://www.w3.org/TR/rdfa-syntax/ (2015年11月20 日参照).
[33] Weber, Marc; Alexa, Marc; Müller, Wolfgang: “Visualizing Time-Series on Spirals”, Information Visualization, 2001. INFOVIS 2001. IEEE Symposium on, pp.7-13, 2001.
[34] Li, Xia; Kraak, Menno-Jan: ”Explore multivariable spatio-temporal data with the time wave case study on meteorological data”, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. 38, Part II, pp.575-580, 2008.
[35] 久保正敏；原正一郎；関野樹：「三次元時空間モデルとその展開－歴史知識を構築するために」，人工知能学会誌 Vol. 25, No. 1, pp. 50-55, 2010.
[36] Sekino, Tatsuki: “Tools to Realize Spatiotemporal Analysis in the Humanities”, Proceedings of GIS in the Humanities and Social Sciences International Conference, pp.151-159, 2009.

第20回情報知識学会フォーラム予稿

フィールドノートに記述された場面を特徴付ける

ー語彙による知識処理ー

Characterizing Scenes in Field Note:

Knowledge Processing Using Vocabulary

山田太造[1*]
Taizo YAMADA[1*]

1 東京大学 史料編纂所
The University of Tokyo, Historiographical Institute
〒133-0033 東京都文京区本郷7-3-1
E-mail: t_yamada@hi.u-tokyo.ac.jp
*連絡先著者 Corresponding Author

フィールドノートは調査したフィールドの観察記録，観察したフィールドの場所・日時，その風景に関するスケッチ・写真などで構成されたものである．調査対象であるフィールドについて特定の日時での様子を詳細に理解することができるため，地域研究において非常に重要な研究資源の１つといえる．われわれはこれまでに，地域研究進展のためにフィールドノートを効率的かつ効果的に利用していく手法を模索しており，本研究ではテキストマイニングを用いてフィールドノートから記述されている場面を特徴づけ，かつセマンティックウェブを利用して表現する手法を提案する．

Field note consists of observation notes, locations, dates, drawings and images of fields which were observed by researchers of Area Studies. The note is one of important research resources for Area Studies, because it enables to understand details of the states of the field at the time. Due to progress of Area Studies we have explored some methods and/or techniques to utilize the note efficiently and effectively. In the paper we introduce a method which a scene described in the note can be characterized by text mining technique, and a method which the characterized scene can be represented by semantic web technique.

キーワード: 地域研究，トピックモデル, LDA, RDF，セマンティックウェブ
area studies, topic model, LDA, RDF, semantic web

1 はじめに

フィールドノートは調査したフィールドの観察記録，観察したフィールドの場所・日時，その風景に関するスケッチ・写真などで構成されたものである．記録されたフィールドについて特定の日時での様子を詳細に理解することができるため，地域研究において非常に重要な研究資源の１つといえる．記述されている内容は地域研究史資料に関する目録ベースの情報よりも詳細かつ膨大であり，地域研究推進において欠かせないエッセンスになりうる．

地域研究推進のために，ここ十年間で，さまざまなデータベースが構築され，公開されている．その多くは，研究資源の目録であり，さらには風景を撮影した写真・映像に関するデータベースの公開もされている．しかしながら，フィールドノートに関するデータベースとしては，公開されたとしても目録までであり，そのテキストに関するデータベースはあまり公開されていない．その理由としては，テキストの作成のコストの問題もあると考えられるが，テキストの地域研究等での利用方法が議論されていないためではないかと考えている．

著者らのプロジェクトでは，地域研究を展開していくために，フィールドノートを効率的かつ効果的に利用していく手法を模索している．そのため，フィールドノートから得られる情報がどのようなものがあるかを，情報学における手法を用いて提示し，提示された結果の地域研究における有益性を検証し，フィールドノート活用を地域研究における研究手法の１つとして確立することを目的としている．

本研究では，フィールドノートから記述されている場面の特徴を，テキストマイニングにより解析し，表現する手法を提案する．特に次の2点に着目する．

1) フィールドノートからの時空間的特徴の抽出と場面の構造化：テキストから地名・日付とともに各場面を特徴付ける用語を抽出する．

2) フィールドノート情報の利用：1) で表現した場面を検索できるフィールドノート検索システムのプロトタイピングを行う．また1)で表現した場面データの2次利用を図る仕組みを導入する．

一般的に，テキストからの情報抽出・話題検出などのテキストマイニングは，永らくデータ工学や自然言語処理などの分野で，提案され，改良されてきた．しかしながら，地域研究史資料に対する，地域研究への寄与を目的としたテキストマイニングは皆無に等しい．本研究が地域研究に与えるインパクト，特にフィールドノートやテキストデータの利活用の点で，地域研究の新たな研究パラダイムに繋がる可能性を秘めているというシーズとしての役割に展開していくと期待している．

本論文は以降，次のように構成している．2節で本研究にて用いたフィールドノートを示す．3章でフィールドノートのテキストから生成した場面データの構造を示す．4章で場面データを特徴づけていく手法を示す．5章で場面データの提示手法を示す．

2 対象のフィールドノート

京都大学地域研究統合情報センターは，東南アジア地域研究において著名な高谷好一氏によるフィールドノートを書籍と

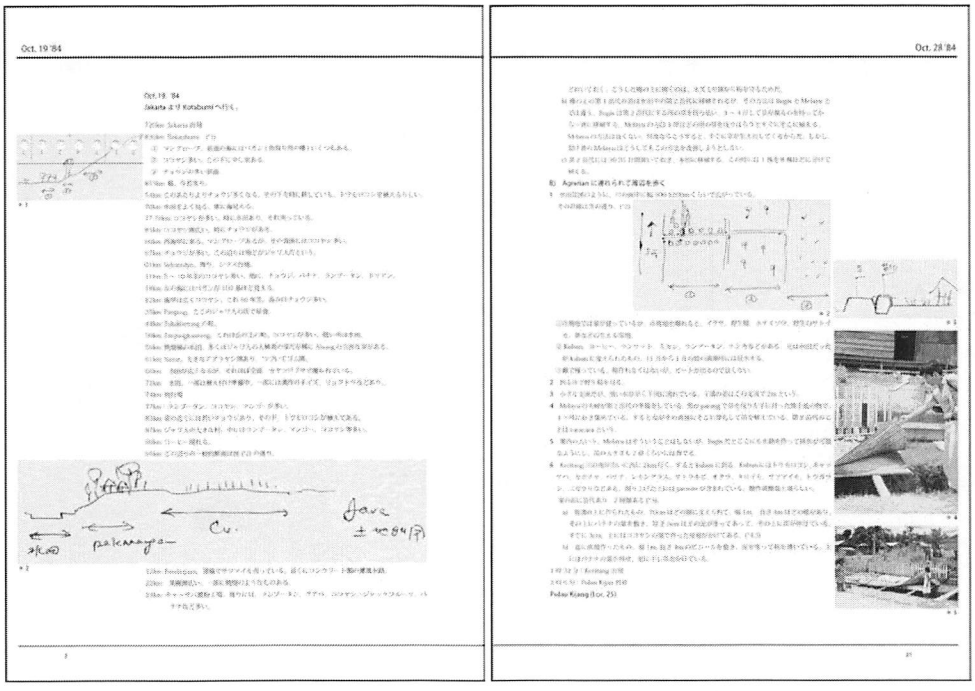


図1 フィールドノートの例．左図は Oct. 19 '84 の，右図は Oct. 28 '84 の記述．

して2012年3月に5冊，2013年3月に3冊合計で4,405ページを出版した．このシリーズは『地域研究アーカイブズ フィールドノート集成』と冠しており，調査対象地域は，東南アジア，インド，ヨーロッパ，アフリカなど様々である．[2]によれば，このフィールドノート集成は，記録者である高谷氏の全面的な協力の下，現場で観察した記録（調査した当日の夜にその日の観察記録をカード化し整理したもの）に対する文字起こしを行い，さらに写真やイラストの整理も行うことで，出版に至ったようだ．

本研究ではこのうち，「スマトラ 1984.10.19 - 1985.1.18 スマトほぼ全域」[1]を利用した．該当のフィールドノートは，分量としてはA4サイズ198ページ，165,757文字程度であるが，スマトラ島全域をカバーしている．図1は実際のフィールドノート集成のページを示す．前述のとおり，この観察記録は日を単位に整理されており，その日の観察した風景，現地の人へのインタビュー，土地利用などに関する記述，現場でスケッチしたイラスト，撮影した写真などで構成されている．図1左はこの調査全体の初日の冒頭を示す．書き出しは“Oct. 19 '84”であり，これは日付を示す．日付に続く“Jakarta より Kotabumiへ行く。”はこの後に続く観察した記録の主題に相当する．主題は１日に複数回登場することもあれば，全く登場しない場合もある．移動しながら風景を記録している場合，端的に１行から数行程度の短い記述が続く．ここで，行が変われば記録している場面が変わることを意味する．図

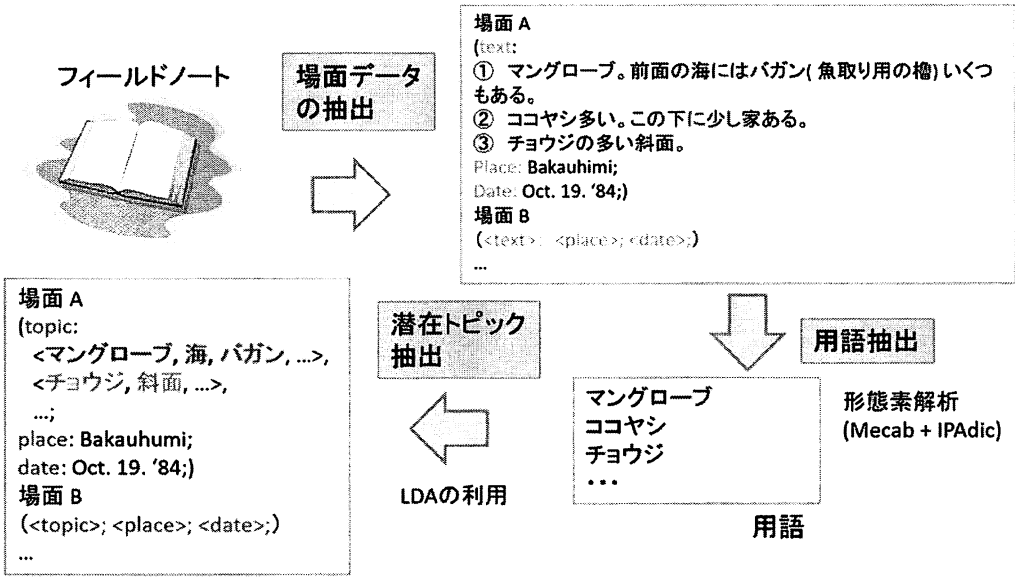


図 2 場面データの作成フロー

1右に“ B) Agrarian に連れられて周辺を歩く”という主題があり，その後に記述が続く．このような記述はその場所の風景，土地利用または統計的情報を集中的に記録している場合に見られる．現地の人へのインタビューの際も同様の記述形態である．本研究では，記録の最小単位であるこの記録を“場面”と呼ぶことにする．

3 場面を構造化する

本節では場面の構造化について述べる．前節で述べたとおり，本フィールドノートは日単位にまとめられており，記述内容は場面の景観，土地利用，現地の人へのインタビューであるため，データとしての場面は時間および場所に関する情報を持つ．そこで，本研究では，これらの要素を用いて各場面を構造化することにした．

図2は場面データの作成フローを示す．本フィールドノートは書き起こしたテキストがあるため[2]，テキスト化は不必要だった．まずフィールドノートのテキストを場面ごとに区切った．その後，場面データを作成した．前述のとおり，場面データは場面の記述内容，時間と場所を表すデータを用いて表現した．

次に，時間軸・空間軸で識別される場面を特徴付けた．マングローブ，チョウジ，ココヤシ，赤土，丘陵のようにその場面で観察された“もの”・“こと”が記述されており，場面の特徴を端的に表現している．本研究では，これらを抽出し，分類することでその場面を特徴付けていくことにした．まず，場面を特徴付けるこれらを用語としてテキストから抽出した．このとき形態素解析などを行うことで抽出した．場面の分類を行うため，抽出した用語とその出現頻度をもとに，トピックモデルを用いて場面を特徴付けた．このように特徴付けら

れた場面データを多目的に利活用可能な形式で表現する．その方式としてRDFを用いて表現した．これらは次節以降に述べる．

4 場面を特徴付ける

4.1 用語を抽出する

場面データを特徴付けるため，まずテキストから用語を抽出する．用語抽出ではMeCab[3]を用いて形態素解析を行い，そのうち名詞および形容詞を用語として抽出する．ただし名詞のうち，代名詞，数，接尾，副詞可能，形容動詞語幹，ナイ形容詞語幹，接続詞的，非自立は対象外とした．形態素解析辞書としてIPADICを用いた．また，頻出するがIPADICには登録されていない用語（ランブータン，ジャックフルーツ，キャッサバ，サゴヤシ）をユーザ辞書に登録した．

形態素解析の結果からチャンキングを行った．固有名詞や専門用語の出現をカバーするためである．日本語係り受け解析器であるCaboCha[4]は固有表現解析の機能を有しており，これを用いてチャンキングを行った．CaboChaはIREX（Information Retrieval and Extraction Exercis）の定義にもとづき固有表現タグをIOB2形式で出力することができる．本フィールドノートでは“Banda Aceh”や“Sungai Tumbesi”のようにローマライズされた表記が出現するが，残念ながらこれらに対して有効に機能しない．そこで，CaboChaでの固有表現解析の結果に対して，さらに下記のルールを適用した．これは汎用的ではなく，本フィールドノートにおいてのみ有効なヒューリスティックな手法である．

・抽出対象が連続する場合
・名詞の直後に名詞-接尾が出現する場合
・[a-zA-Z]+にマッチする形態素が連続する場合

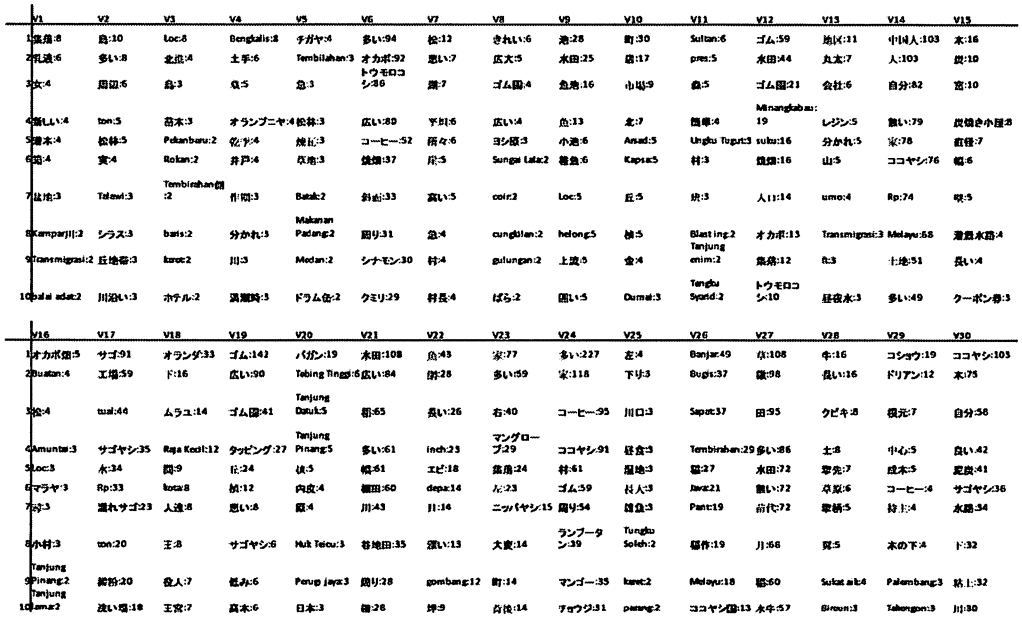
4.2 場面からトピックを検出する

次に，場面を特徴付ける．情報検索等ではしばしばベクトル空間モデルを用いて特徴付けることがある．文書を対象に特徴づける場合，ある文書での各用語の出現頻度等で算出される重みで文書ベクトルを作成する．本研究では文書は場面に相当する．場面における記述は1行程度である場合や数ページに渡る場合もあり，長さが不均一である．そこで本研究では，各場面のトピックにもとづいてその場面を分類していくことで特徴づけていくことにした．

予めどのようなトピックがあるかが判断することができなかった．そこで，教師なし学習(unsupervised Learning)の手法であるLDA（Latent Dirichlet Allocation）[5]を用いて潜在トピックの検出を行った．LDAを用いると，場面と潜在トピックの関係，および潜在トピックと用語の関係を検出することができる．LDAは（1）式により表現することができる．

$$p(d|\alpha,\beta) = \int Dir(\theta|\alpha) \cdot \left(\prod_{n=1}^{|d|}\sum_{k=1}^{C} p(w_n|z_{k,\beta})p(z_k|\theta)\right) d\theta \quad (1)$$

ここでα,βはパラメータ，$z = z_1, z_2, \ldots, z_C$は潜在トピック，$\theta = \theta_1, \theta_2, \ldots, \theta_C$は潜在トピックの生成確率，$Dir(\theta|\alpha)$はディリクレ分布，$\mathrm{d} = (\mathrm{w}_1, \mathrm{w}_2, \ldots, \mathrm{w}_{|d|})$は場面，$w_n$は用語，$|\mathrm{d}|$は場面 d の総用語数を示す．LDA は潜在トピックの生成確率がディリクレ分布に従うと仮定した文書生成モデルといえる．



	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15
1	住居:8	島:10	Loc:8	Bengkalis:8	チガヤ:4	多い:94	校:12	きれい:6	池:28	町:30	Sultan:6	ゴム:59	地区:11	中国人:103	木:16
2	礼遇:6	多い:8	北部:4	土手:6	Tembilahan:3	オカボ:92	悪い:7	広大:5	水田:25	島:17	pres:5	水田:44	丸太:7	人:103	畑:10
3	女:4	周辺:6	島:3	草:5	急:3	トウモロコシ:36	歳:7	ゴム園:4	急地:16	市場:9	桑:5	ゴム園:21	会社:6	自分:82	宮:10
4	寂しい:4	ton:5	苗木:3	オランプニヤ:4	松林:3	広い:89	下肘:6	広い:4	魚:13	北:7	簡単:4	Minangkabau:19	レジン:5	難しい:79	炭焼き小屋:8
5	種木:4	松林:5	Pekanbaru:2	亿万:4	焼畑:3	コーヒー:52	荒々:6	ヨシ原:3	小池:6	Arsad:5	Ungku Tugut:3	suku:16	分かれ:5	家:78	直径:7
6	国:4	実:4	Rokan:2	井戸:4	草地:3	焼畑:37	岸:5	Sungai Lata:2	着急:6	Kapus:5	村:3	焼畑:16	山:5	ココヤシ:76	幅:6
7	住地:3	Talawi:3	Tembirahan個:2	作期:3	Batak:2	斜面:33	高い:5	coir:2	Loc:5	丘:5	坊:3	人口:14	umo:4	Rp:74	根:5
8	Kamparjiji:2	シラス:3	baris:2	分かれ:3	Makanan Padang:2	周り:31	急:4	cungkilan:2	helong:5	袖:5	Blasting:2	オカボ:13	Transmigrasi:3	Melayu:68	湧昇水路:4
9	Transmigrasi:2	丘地帯:3	karet:2	川:3	Medan:2	シナモン:30	村:4	gulungan:2	上流:5	金:4	Tanjung enim:2	集落:12	8:3	土地:51	長い:4
10	balai adat:2	川沿い:3	ホテル:2	満潮時:3	ドラム缶:2	クミリ:29	村長:4	ばら:2	固い:5	Dumai:3	Tengku Syarif:2	トウモロコシ:10	昼夜水:3	多い:49	クーポン券:3

	V16	V17	V18	V19	V20	V21	V22	V23	V24	V25	V26	V27	V28	V29	V30
1	オカボ畑:5	サゴ:91	オランダ:33	ゴム:142	バガン:19	水田:108	魚:43	家:77	多い:227	左:4	Banjar:49	草:108	牛:16	コショウ:19	ココヤシ:103
2	Buton:4	工場:59	下:16	広い:90	Tebing Tinggi:6	広い:84	団:28	多い:59	家:118	下り:3	Bugis:37	種:98	長い:16	ドリアン:12	木:75
3	役:4	tual:44	ムラユ:14	ゴム園:41	Tanjung Datuk:5	都:65	長い:26	右:40	コーヒー:95	川口:3	Sapat:37	田:95	クピキ:8	根元:7	自分:58
4	Amuntai:3	サゴヤシ:35	Raja Kecil:12	タッピング:27	Tanjung Pinang:5	多い:61	inch:23	マングローブ:29	ココヤシ:91	昼食:3	Tembilahan:29	多い:86	土:8	中心:5	良い:42
5	Loc:3	木:34	間:9	住:24	状:5	幅:61	エビ:18	集落:24	林:61	湿地:3	猫:27	水田:72	家先:7	成木:5	肥良:41
6	マラヤ:3	Rp:33	kota:8	枝:12	内皮:4	植田:60	depa:14	左:23	ゴム:59	長大:3	Java:21	難い:72	草原:6	コーヒー:4	サゴヤシ:36
7	苏:3	湿れサゴ:23	人造:8	悪い:8	殿:4	川:43	川:16	ニッパヤシ:15	周り:54	雑食:3	Pant:19	前代:72	家柄:5	幹上:4	水路:34
8	小村:3	ton:20	王:8	サゴヤシ:6	Nuk Teiou:3	谷地田:35	強い:13	大変:14	ランブータン:39	Tungku Soleh:2	稼作:19	川:68	質:5	木の下:4	下:32
9	Tanjung Pinang:2	統治:20	役人:7	低み:6	Perup jaya:3	周り:28	gombang:12	町:14	マンゴー:35	karet:2	Melayu:18	語:60	Sukat ak:4	Palembang:3	結上:32
10	Tanjung Lamu:2	浅い環:18	王宮:7	高木:6	日本:3	種:26	坤:9	首長:14	チョウジ:31	perang:2	ココヤシ園:13	水牛:57	Biroon:3	Tabongon:3	川:30

図 3 LDA による潜在トピック検出の結果

本研究に適用した場合，LDA は，1 場面におけるトピックは複数あり，トピックはそれぞれに複数の用語を生成することをモデル化している．(1) 式をそのまま計算することはかなり困難であるが，崩壊形ギブスサンプラを用いた解法が知られており[6]，本研究ではこれを用いて潜在トピックを算出する．

トピックモデルとして LDA 以外にも LSI (Latent Semantic Indexing) [7]や pLSI (probabilistic LSI) [8]がある．LSI は 1 場面につき 1 トピックを仮定するため，多角的な関連性を考慮できない．pLSI は LDA と同様に LSI を拡張し 1 場面につき複数トピックを仮定する．しかしながら，潜在トピックの生成確率，つまり式 (1) における $p(z|\theta)$ を最尤推定するなどして事前に算出する必要がある．そのため，学習データにはない場面への対応は高コストになってしまう．また $p(z|\theta)$ は学習データの量に応じて計算コストが増大してしまうためアドホックな手法で求めることが多い．これに対し LDA は $p(z_k|\theta)$ を確率的に算出する生成モデルである．

4.3 フィールドノートへ適用する

本フィールドノートに対し，LDA を適用することで潜在トピックの検出を行った．LDA のパラメータとしてはトピック数を 30，ギブスサンプリングの回数を 2,000 とした．

この結果を図 3 に示す．この図では 30 のトピックと各トピックを構成する用語のうち出現頻度順に上位 10 件を示している．各トピックの概要を端的に示すことは難しい．そこで，水田とココヤシが出現するトピックに着目してみる．水田

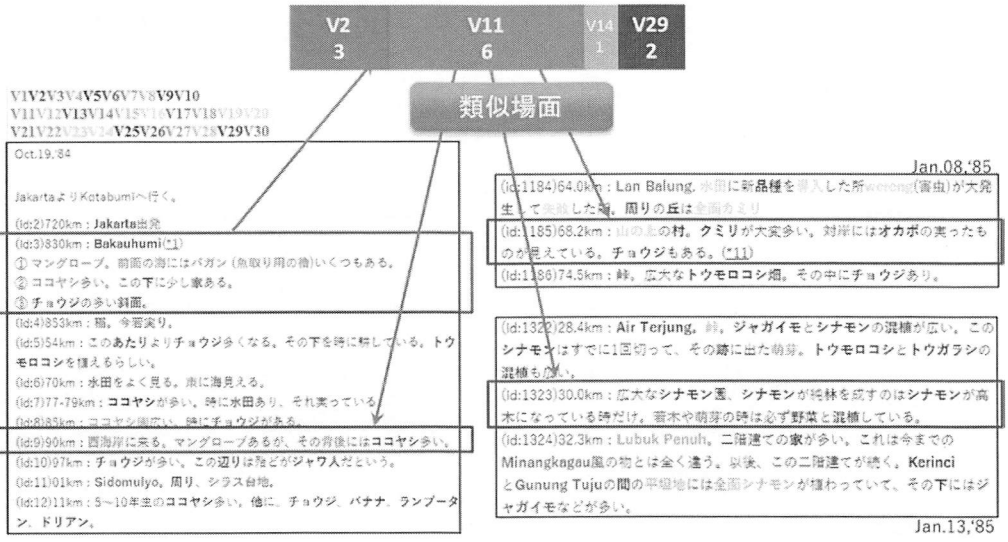


図 4　類似する場面の検索例

が出現するトピックは 9，12，21，27 である．トピック 9 では，水田以外に，池，魚池，小池，稚魚などがある．同様に，トピック 12 では，ゴム・焼畑・オカボ・集落・トウモロコシなど，トピック 21 では，稲・幅・棚田・川・谷地田など，トピック 27 では，草・鍬・苗代・水牛などがあった．これにより，トピック 9 は池や水田と魚関係，トピック 12 はゴムや水田などの作目がある風景，トピック 21 は水田のある風景，トピック 27 は田畑・耕作等関係を示すと考えられる．ココヤシが出現するトピックは，14，24，30 だった．トピック 14 では中国人・家・Melayu・土地など，トピック 24 ではコーヒー・村・ゴム・ランブータン・マンゴー・チョウジなど，トピック 30 では木・泥炭・サゴヤシ・水路・粘土・川などがあった．これらの検出された用語より，トピック 14 は土地利用関係，トピック 24 はコーヒー・ココヤシなどがある風景，トピック 30 は水辺・川辺での風景を示すと考えられる．また，文字列もしくは指し示す“もの”・“こと”は同じであっても，トピックが異なればその用語の意味合いが異なることがわかった．

5 場面を提示する

5.1 類似する場面を提示する

前節で示した LDA の結果を用いて，類似する場面を提示する方法について考えてみる．各場面から抽出した用語に対し，LDA を適用することにより，それぞれがどのトピックとして検出されたかが結果として得られる．検出された用語の頻度を場面ごとに集計し，これをベースに場面ごとに各トピックの重み付けを行えば場面ベクトルを作成することが可能である．ある場面xにおけるトピックiの重み$\mathrm{weight}(x_i)$を次式により算出する．

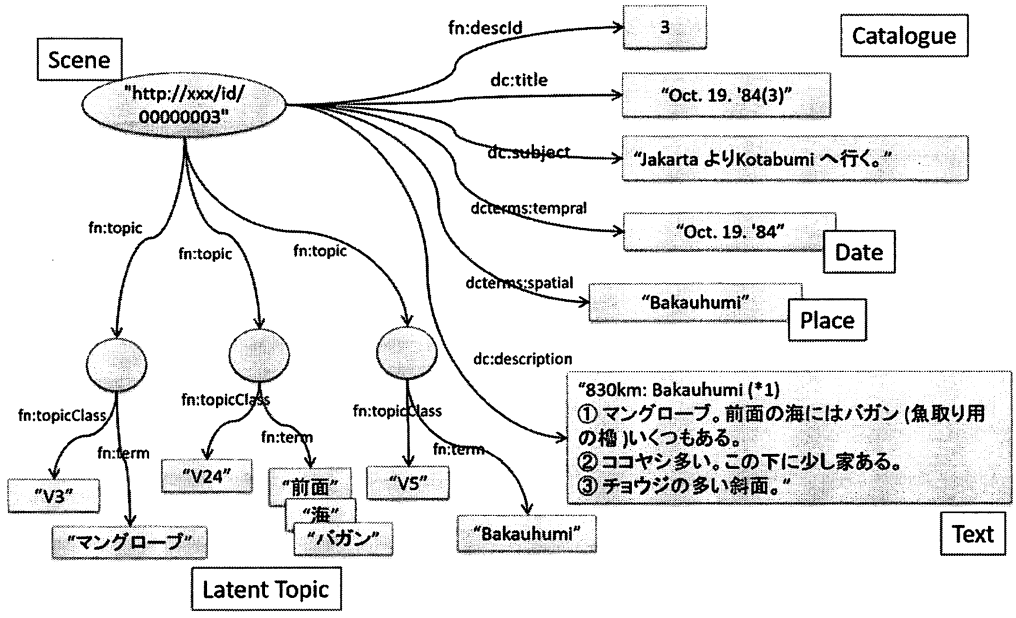


図 5 RDF を用いた場面データ

$$\mathrm{weight}(x_i) = tf(x_i) \cdot \left(log \frac{N}{df(i)+1}\right) \quad (2)$$

これはtf-idf重み付けに相当する．ここで，$tf(x_i)$は場面xにおけるトピックiの出現頻度，$\mathrm{df}(i)$はトピックiを含む場面数，Nは場面の総数を示す．次に場面の類似度次式で算出する．

$$sim(x,y) = \frac{\sum_i weight(x_i) \cdot weight(y_i)}{\sqrt{\sum_i weight(x_i^2) \cdot \sum_i weight(y_i^2)}} \quad (3)$$

（3）式は場面ベクトルxとyのコサイン類似度を示している．図4は0ct. 19 ’84の2つ目の場面（場面ID：3，青の矩形で囲んだ部分）と類似する場面を検索したときの検索結果（赤の矩形で囲んだ部分）の一部を示す．この場面における各トピックの出現頻度を図4の上部に示している．これは場面ベクトルの一例であり，これを用いて（2）式および（3）式により，場面ベクトル間の類似度が算出され，類似性の高い場面を検索することができる．

5.2 場面をデータとして提示する

図2に示したフローにより場面データを作成していく．このフローにおける各処理については前節までに示したとおりである．作成された場面データを単一の目的，例えば5. 1に示した類似する場面データを提示するための利用目的，もしくは単一のシステムでの利用だけではせっかく作成した場面データを広く共有・利活用するのは限界があると考えている．場面データをシステムとは切り離し，汎用的な形式で記述することができれば，既存のシステム・ツールで利用するなど多様な目的で利活用することが可能だと考えられる．

本研究では，特にweb上でのデータ取得・共有，あるいはwebアプリケーション・システムでの利用を考慮し，場面データを

RDF（Resource Description Framework）を用いて表現することにした．図5はRDFモデル化した場面データのグラフを示す．RDFはWeb上で扱う情報（リソース）を広く共有・利用するための共通化した標準的な枠組みである．RDFではリソースを主語・述語・目的語の3つ組（トリプル）で表現する．主語は表現対象である．場面データを対象としているため，場面（図5はScene）自体が主語である．目的語も主語と同じくリソースである．述語は主語と目的語の関係を明示した情報と見なすことができる．述語はプロパティと呼ばれることもある．図5では楕円は主語，矢印は述語，矩形は目的語を示す．RDFではリソースの識別のためURI（Uniform Resource Identifier）が標準化されている．本研究でも同様に，場面の識別子としてURIを用いる．目的語はURIもしくはIRI（Internatinalized Resource Identifier）で表現する．述語はDublin Coreなど既知・既存のメタデータ語彙を用いた．しかしながら潜在トピックの表現では，同様の語彙が見当たらなかったため独自語彙として定義した．この独自語彙は，図5において，接頭辞がfnである語彙がこれに相当する．図5は，場面Sceneは，その場面の識別子（fn:descId），タイトル（dc:title），主題（dc:subject），日時（dcterms:temporal），場所（dcterms:spatial），記述内容（dc:description）および潜在トピック（fn:topic）で構成されていることを示している．ここで，fn:termは用語，fn:topicClassは潜在トピックのクラスを示す．

6 おわりに

本研究では，高谷好一氏著者『地域研究アーカイブズ フィールドノート集成』の一部を用いて，フィールドノートに記録された場面の記述をもとに，フィールドの各場面を自動的に特徴付けていく手法について述べた．また，その成果を特定のシステム・アプリケーションだけではなく，広く利活用するため，RDFによるデータ表現についても述べた．

地域研究におけるテキストデータの利活用，特に地域研究における研究的側面に耐えうる利用方法を目指していくことを想定し本プロジェクトを開始した．しかしながら現在では，地域研究としてのニーズに耐えうるだけのアプローチを提案することは大変困難であるかもしれないと考えるに至った．テキストの利用方法についてはシーズ的に研究者に提供するほうがもしかしたらニーズに寄り添ったものになるかもしれないと考えるに至り，現在の研究過程を経た．これをより，研究者サイドで利用しやすいものへと転換する必要もある．これはテキストおよびテキスト分析結果等の提示もしくは新たな加工・分析を可能とするユーザインターフェースも必要かもしれない．もしくは，さらに掘り下げたデータ分析結果かもしれない．それについて，地域研究者との対話を重ねながら本研究を進めていく予定である．

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（B）（26280122），および京都

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「フィールドノートを対象としたテキストマイニングに関する研究」の助成を受けたものによる.

参考文献

[1] 高谷好一：「スマトラ」, 地域研究アーカイブズ フィールドノート集成2, 京都大学地域研究統合情報センター, Vol. 2, No. 22, pp. 1-198, 2012.
[2] 柳澤雅之：「フィールドノート・プロジェクト」, Seeder, 昭和堂, No. 11, pp. 14-22, 2014.
[3] MeCab: Yet Another Part-of-Speech and Morphological Analyzer, http://taku910.github.io/mecab/.
[4] CaboCha/南瓜: Yet Another Japanese Dependency Structure Analyzer, https://taku910.github.io/cabocha/.
[5] D. M. Blei; A. Y. Ng,; M. I. Jordan: Latent Dirichlet Allocation, Journal of Machine Learning Research, vol. 3, pp. 993-1022, 2003.
[6] T. L. Griffiths; M. Steyvers: Finding scientific topics, Proc. of the National Academy of Sciences of the United States of America, vol. 101, pp. 5228-5235, 2004.
[7] S. Deerwester; S. T. Dumais; G. W. Furnas; T. K. Landauer; R. Harshman: Indexing by Latent Semantic Analysis, Journal of the American Society of Information Science, Vol. 41, No. 6, pp. 391-407, 1990.
[8] T. Hofmann: Probabilistic Latent Semantic Indexing, Proc. of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval, pp. 50-57, 1999.

第20回情報知識学フォーラム予稿

地域情報学のこれまでとこれから
ー地域研究統合情報センターの実践事例を通してー

Past and Future of Area Informatics
ーfrom practices in the CIAS, Kyoto Universityー

亀田尭宙[1*]
Akihiro KAMEDA[1*]

1 京都大学
Kyoto University
〒606-8501 京都市左京区吉田下阿達町46
E-mail: kameda@cias.kyoto-u.ac.jp
*連絡先著者 Corresponding Author

本論文では，地域情報学のこれまでの研究を，地域情報学を構成する４つのパーツ（情報基盤，ツール，知識ベース・オントロジ，地域に関するデータ）と３つの軸（主題，空間，時間）の下に位置づけて俯瞰した．主に，京都大学地域研究統合情報センターに関わる実践事例を中心にしつつ周辺の技術動向も含めて紹介することで，そこから課題とこれからの展望を描いた．

In this paper, I described researches in Area Informatics domain, mainly related with Center for Integrated Area Studies, Kyoto University. For structurizing those researches, 4 components (information architecture / processing software / ontology and knowledgebase / area data) and 3 axis (topic / spatial / temporal) are introduced. Related technical trend is also surveyed, and, based on that, future of Area Informatics is also described.

キーワード: 地域情報学，地域研究，情報学, GIS, LOD，自然言語処理
Keyword: Area Informatics, Area Studies, Informatics, Geographic Information System, Linked Open Data, Natural Language Processing

1 はじめに

本論文では，京都大学地域研究統合情報センター（以下，地域研）に関わる研究を中心にして，地域情報学のこれまでの研究をまとめるとともに，そこから課題とこれからの展望を描く．科学自体が本質的にそうであるという指摘もあるが[1]，地域情報学のような学際領域は，長い歴史を持った学問分野と比べてより一層その枠組みが捉えづらく，固定され明文化可能な規範を持っているわけではない．それでも，ある角度からその分野の研究を俯瞰して描写すること，それが蓄積されていくことは，分野が体系化されることや取り組まれるべき問題をあぶりだすのに有用であると考えられる．

2004年4月に京都大学東南アジア研究所（以下，東南研）が附置研究所となるのに伴って地域情報学研究部門が作られたのが，はじめて公的な名称として地域情報学が使われた例である．同年，地域情報学の創出を目的とした議論の場であるH-GIS（Humanities GIS）研究会が発足した．地域情報学元年とも言える年である．その後，様々な活動を展開し，2008年度にはそれまでの活動を俯瞰するような大きな取り組みがいくつか行われた[2, 3, 4]．地域研は2006年4月に創設された後，2010年度に地域情報学プロジェクトを発足させセンターの大きな活動として位置づけるなど地域情報学に積極的に取り組んできた組織の１つであり，センターの活動として直接，また共同利用・共同研究拠点としての研究支援を通して間接的にも地域情報学に関わる多くの活動を行ってきた．

まず，2章で地域情報学の定義やモデルについて紹介し，3章では地域研の事例を中心にして，これまでの研究をまとめると共に近年の動向と展望を描く．

2 地域情報学とは

2.1 地域情報学の目的と方法

まず，地域情報学については[2]では次のように定義されている．

地域情報学とは，地域に存在する情報・知識・知恵を明確に定義された形式や手順に従って，比較可能な方法で構造や意味を理解し，体系化する過程を言う．この体系化された知識体系を『地域の知』と呼ぶ

また，手法としては以下のものが挙げられている．

1. 情報学的手法を導入した実証的な地域研究の展開
2. 地域情報の高次情報処理やツール開発
3. 地域を「語る」ための地域の「知」の構築の基礎となる情報学的な枠組みの体系化
4. 資源共有化と情報基盤の構築と地域情報のメタ情報の整備

これらを図1のようにまとめなおして，以後，各研究を位置づける．

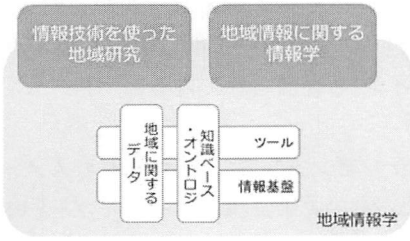


図1 地域情報学の構造

地域情報学が拠って立つ情報学と地域研究[1]は，それぞれが学際的な性格をもちつつも，それぞれに多くの研究者と研究の蓄積があり査読付きの学会誌を有するなど，分野として確立されている．そこで，地域情報学の成果は主に，情報技術を使った地域研究と，地域情報に関する情報学という相互乗り入れ的な形で論文化されている．1の「情報学的手法を導入した実証的な地域研究の展開」はまさに前者に対応する．また，その相互乗り入れの実態を図示したのが地域情報学の枠で囲った井桁状の4つのパーツである．情報学側がツールや基盤を構築して提供している(それぞれ，2と4の「資源共有化と情報基盤の構築」に対応)．そこで処理・蓄積される情報として各地域研究のデータや，それらのデータをつなぎ合わせるための知識ベースやオントロジ(4の「地域情報のメタ情報の整備」と 3の一部に相当)が共同で作られる．3における枠組みの体系化は，本論文を含め地域情報学を俯瞰してモデル化しようとするメタな取り組みによって実践されているといえる．

[1]日本語の地域研究はいくつかの分野を指す多義語であるが，ここでは世界の諸地域ごとに政治，経済，社会，文化などを横断的に研究し，またそれらを地域間比較する学際領域（英語の Area Studies に相当）を扱っている．

図1の構造は各研究そのものを MECE に分類するものではない．一つ一つの研究を詳細にみていくと，ある情報基盤は多くのデータを横断的に抱えており，一つのデータが複数のツールを用いて分析されていることもある．研究をそれらのパーツに分解して位置づけなおすことで，動向と課題を考察する．

2.2 地域情報学の時空間モデル

ツールや情報基盤によって新しい地域研究を切り開くためのデータモデルとして，4D-GIS空間というモデルが提唱されている[5]．これは，主題軸・2次元の空間・時間軸で地域のできごとを記述する時空間概念モデル[6]を拡張したもので，2次元の空間をGIS(地理情報システム)における空間一般に置き換えたものである．

そこで，前述のパーツの他に，この主題・空間・時間という軸をどう扱った研究であるかということも見ていきたい．また，このデータモデルとツールや情報基盤がどのようなかかわりにあるかということも，次章で考察する．

3 実践事例と周辺動向の分析

3.1 情報基盤の構築

情報基盤の構築として最も大きい成果はMyデータベースシステムである．これは，表形式のデータをCSVフォーマットで読み込み，各カラムなどについての設定を追加することで，容易に検索機能やマッピング機能およびAPIの提供を実現できるようにしたシステムであり，地域研のウェブサイト[2]からアクセスできるデータベースの多

[2] データベース一覧
http://www.cias.kyoto-u.ac.jp/database/

くが，このシステムの上に載っている．シンプルな例として，20世紀年表[3]は，20世紀前半の情勢を政治，経済，社会，文化に区分した年表のデータベースで，Myデータベースシステム上で主題，時期，地域を指定した検索ができるようになっている．また，地域研究資源共有化データベース[4]は同様のAPIを備えた他組織との横断検索を可能にしている．これらは，4D-GISの情報を扱え，検索の際に地域を矩形で指定したり，年代を期間で指定したりといった検索ができるようになっている．

また，APIの提供は，他の情報基盤やツールとの連携のために有用であり，例えば1950～60年代のマレー世界におけるムスリム社会の動向を理解するのに重要な史料であるマレー・イスラム雑誌『カラム』のデータベース[5]は別のサーバからMyデータベースのAPIにアクセスすることで，書籍の対訳を閲覧するのに適したインタフェースを提供している．

3.2 ツールのオープン化とウェブ化

ツールの提供を主とするプロジェクトとしては，時空間情報処理ツールの提供がある．4D-GIS 空間での解析を実現するために，時間情報解析ツールのHuTime[7]とそれと連携を図るGISツールHuMap[8]が開発され，公開中である．一方，GIS に関しては近年，QGIS[6] のように世界規模で使われているオープンな GIS ツールが新興してきており，そこで時空間分析を支援するプラグイン[7]も提供されている．

寺院マッピング[8]，センデロ・ルミノソ・マッピング[9]，アチェ津波アーカイブ[10]といったマッピングプロジェクトでは，ブラウザに Google Earth[11]のプラグインを入れるだけで地図上に3Dでマップされ，それぞれの共同研究者によって色分けや関連性の表示などの工夫がなされたインタフェースが閲覧できるようになっている．関連プロジェクトのヒロシマ・アーカイブ[12]では，Chromeブラウザ等のNPAPI打ち切りに伴って，ブラウザプラグイン および Google Earthに頼らずJavaScriptのライブラリで表現するように変更された[13]．

地図についてはオープンな地図を共同で作成するプロジェクト OpenStreetMap[14] のデータが充実してきており，アチェ津波モバイル博物館[15]や災害と社会情報マッピ

---

3 http://app.cias.kyoto-u.ac.jp/infolib/meta_pub/G0000020NPY

4 http://app.cias.kyoto-u.ac.jp/GlobalFinder-lg/cgi/Start.exe

5 http://majalahqalam.kyoto.jp/

6 http://qgis.org/

7 Time Manager http://anitagraser.com/projects/time-manager/

8 http://temple.mapping.jp/

9 http://peru.mapping.jp/

10 http://aceh.mapping.jp/

11 https://www.google.co.jp/intl/ja/earth/

12 http://hiroshima.mapping.jp/index_jp.html

13 技術は平和活動もアップデートする～原爆証言アーカイブの OSS 活用に見る，戦争の記憶を「みんなで遺す」意味 - エンジニア type http://engineer.typemag.jp/article/hiroshimaarchive2015（2015 年 11 月 15 日参照）

14 https://www.openstreetmap.org

15 http://disaster.net.cias.kyoto-u.ac.jp/Aceh/

ング・システム[16]でも使われている．
このようにツールのオープン化やウェブ化が進展している中で，それらのツールと連携するようにツールやデータベースのインタフェースを作ることで，重要な機能のみに集中して開発できるとともに，他のツールのユーザを呼び込むことが今後の課題となる．

3.3 ツールの寿命とデータの寿命

対象として歴史などを扱う地域情報学では，その成果物も十分な寿命を見込めることが望ましい．記憶媒体の寿命の問題を脇によけたとしても，ツールやデータにも寿命がある．ツールはそのツールがメンテナンスされなくなって数年たつと，多くの場合使えなくなってしまったり，より利便性の高いツールに代替されてしまったりする．

データの寿命は一般的にツールよりも長い．また，データの構造をシンプルに保つことで，データ規格に寿命が来た時に，次世代の規格への変換がしやすくなる．データの寿命はそのフォーマットに対応するツールやそのフォーマットで書かれたデータの数といったもので決まる．つまり，そのデータフォーマットの周辺環境が衰退すると，そのデータを可視化したり，検索したりといったことが難しくなり，データが「死んで」しまう．例えば，XMLのパーサは多くのプログラミング言語で既にあるが，その元になった SGMLのパーサライブラリを見つけるのはより困難なので，SGMLで書かれたデータを処理するツールを記述するには多大なコストがかかる．

データとツールや情報基盤を適切に分離することで，データの長期安定的な提供が可能になるとともに，その時々で最良のツールを利用できるようになる．

これは，4D-GIS空間のようなデータのモデルやそれを表現するためのフォーマットや語彙（CSVの場合はそのカラム名や日付フォーマット，RDF/XMLの場合はその述語・クラスなど）を，できるだけオープンで普及した仕様に基づいて策定し，関連ツールとの互換性を図っていく必要性があることを意味している．また，そのドキュメンテーションの共有が急務である．Myデータベースは CSVを基本にしており，ヘッダの拡張などに関するドキュメンテーションは現在準備中であり，他機関からのデータの受け入れを含めより外へ開かれたシステムとしていく．

3.3 フィールドでの支援

ここまで紹介してきたツールはデータの処理に関するものだったが，作成の段階から情報技術のツールによって支援することができる．GPSやコンパスは一般的なカメラやスマートフォンにも搭載されて多くの地域研究者が使っており，殊更地域情報学として言及するものではなくなった．しかし，加速度センサーなど他のデバイスとデータを接合する試みは地域情報学の中で行われている[9]．また，阪神大震災やアチェの津波の被災時の写真に重ねて現在を撮るアクティブ・ファインダーの仕

[16] http://disaster.net.cias.kyoto-u.ac.jp/Indonesia/

組みを用いたメモリーハンティング[17]アプリは，防災教育や災害ツーリズムの分野で実際に活用されている．

3.4 知識ベースとオントロジ

分野の知識を体系化した知識ベースやオントロジは，データを連携させるために重要な役割を果たす．地域研ではトピックマップによるオントロジの作成に取り組んできた．具体的には日本図書館協会および国立国会図書館の件名標目表，農林水産関連分野の語彙集（AGROVOC），世界各地の民族・社会・文化に関する文献語彙集（HRAF）などがある．

近年はLinked Open Data（以下，LOD）による知識ベースの共有が盛んになってきており，博物館資料や場所，生物種といった情報をLOD化した LODAC Project[18]を含め多くの知識ベースがウェブ上で連携するようになってきており[10]，トピックマップの LOD 化や利活用に取り組み始めている．

暦間の日付変換ツールである HuTime は，LOD 化がなされており変換ツールが時間という実質無限の情報に関する知識ベースを生成する仕組みは画期的である．歴史地名辞書データベース[19]も他のデータ間を結びつけるための知識ベースの役割を果たす．つまり，4D-GISのそれぞれの軸で知の体系化が進んでおり，特に主題軸では件名標目のように汎用的なものから地域研究の各分野の語彙にわたるまで地域情報学にとって必要不可欠なパーツとなっている．

3.5 多言語対応の周辺環境

各地域の言語を身につけてフィールドワークを行う地域研究にとって，資料の多言語性は必然であり，またそれを適切に保存・変換・検索できるようにすることは地域情報学に強く求められている．

例えば三印法典データベースは，アユタヤ時代からの社会・経済活動を網羅したタイ最古の成文法典である三印法典を，全文テキスト及びこれとリンクされた原本画像データベースとして整備したものである．この三印法典データベースの作成を遡ると，1983年当時に子音・母音・声調を適切に印字できるワープロさえない状況から始まっているが[5]，近年では UTF8 のような多言語に対応した文字コードの普及や，IME のような文字入力環境の向上により，漢字やタイ文字のような文字の入力とデータの保存は飛躍的に簡単になるとともにデータ間の齟齬が少なくなった．よって文字のレベルでの多言語の問題は解決しつつあると考えてよいが，例えば，翻字の際に声調記号を残した場合には，東南アジア逐次刊行物総合目録データベースと共に公開されている音標変換表[20]などを参考に検索の実装に工夫が必要である．

また，ここ最近の動向としては，語のレベ

---

17 http://dsr.nii.ac.jp/memory-hunting/
18 http://lod.ac/
19人権問題に抵触するため内部利用に限っている

20 音標処理と音標変換表の公開
http://www.cseas.kyoto-u.ac.jp/info/db/sealib/lang/ja/ja-spec.htm?lang=ja&footer=footer.htm

ルについて知識ベースによる多言語対応環境の急速な整備が挙げられる．BabelNet[21] は 1400万語 の シンセット (synset，表記ではなく意味によって同定される一語) を 272 の言語について横断的にデータ化している．また，この中にも含まれているが，GeoNames[22] も一つの地名に関しての複数の外名 (exonym，第三者による特定の土地・民族の呼称) や緯度経度といった情報を保持している．時間に関しては多言語性に近いものとして，暦とその表記の多様性がある．HuTime システムはそういった情報を統合する機能を有しており，時間に関する多様な表記をユリウス通日を介在させてまとめられる．

これらの知識ベースを生かして，データの言語間横断検索や関連情報の推薦を実現するような情報基盤やツールの整備が展望される．現在既に[地域研究資源共有化データベース多言語対応版[23]では言語グリッド[24]を用いて言語横断検索ができるようになっている．また，文のレベルでは，翻訳サービスや翻訳アルゴリズムの利活用も見込めるが，実用に十分な精度を提供するのは困難なのが現状である．

3.6 文書の解析

一人の地域研究者の蔵書を中心とする研究資料を整理した石井米雄コレクション[25]は，キーワード分類として，LC分類，大分類，地域分類，分野分類，言語分類で検索できるようになっている．地域分類などの分類体系は東南研の図書室の分類を活用しそこに書誌情報を対応付けている．また，ロシア帝国の中央アジア征服の後，単行本，新聞・雑誌記事，統計，地図，図版などを集め，ロシア人のための中央アジア百科として全594巻に及ぶ規模で作られたトルキスタン集成をデータベース化した「トルキスタン集成」データベース[26]では，その規模の大きさから，全体像を把握するための構造化手法が望まれている．そこで，帝政ロシア期の行政区分や現代の産業分類といった外部の知識体系を持ち込んで書誌情報を体系化する試みを行っている[11]．

データベースの考え方では，知識ベースやオントロジをデータ間の連携に用いる場合に，事前にデータとそれらの対応付けを行っておくのが一般的であるが，これらの試みはツールによって柔軟に対応を発見している．

また，三印法典と同様，世界的に貴重な資料をデジタル化したカラムデータベースでは，語の頻度分析によってワードクラウドから興味のある記事を探すことができるようになっている．高谷が収集したフィールドノートをデータベース化したフィールドノート・データベースでは LDA と呼ばれるトピックモデルを用いて背後のトピックを推定することで，フィールドノート間の関係や単語同士の関係を測り可視化している[12]．

---

21 http://babelnet.org/
22 http://www.geonames.org/
23 http://app.cias.kyoto-u.ac.jp/GlobalFinder-lg/cgi/Start.exe
24 http://langrid.org/jp/
25 http://www.cias.kyoto-u.ac.jp/collection/
26 http://app.cias.kyoto-u.ac.jp/turkestan/

こういった自然言語処理技術との連携は，情報技術を使った地域研究において従来データの読み解きは地域研究者に任されていたところを，主題の発見や割り当てという形で新たな読み解きを支援するところへ一歩踏み出していると言える．この試みは文書を含む他のデータベースにも広げていく予定である．

4 おわりに

地域情報学のこれまでの取り組みを俯瞰し，課題と展望を描いた．もちろん，地域研に関わる研究であっても紙面の都合上ここで紹介することができなかった研究も多くあるし，地域研の外でも多くの魅力的な地域情報学研究がなされている．さらには，引用したそれぞれの俯瞰の取り組みと同様，このまとめ方も一つの可能性でしかないことは付言しておきたい．一方で，それぞれのテーマごとの課題と展望やテーマ間の関連，また地域情報学を語る上で欠かせない地域研究と情報学の間のかかわりに関して，近隣分野からの参入を招き，地域情報学を活性化させるのに有用な視座を提供できたのではないかと考えている．

参考文献

[1] P.K.ファイヤアーベント(村上陽一郎・村上公子＝共訳)：『自由人のための知 ー科学論の解体へ』，新曜社，1982.
[2] 柴山守；原正一郎：「地域情報学の目指すところ 地域研究におけるＧＩＳの応用」，『アジア遊学』113号，勉誠出版，pp. 28-35，2008.
[3] 石井米雄；田中耕司；柴山守；貴志俊彦：「地域研究における情報学を考える」，『アジア遊学』113号，勉誠出版，pp. 4-25，2008.
[4] 柴山守：「〈特集〉地域情報学—地域研究と情報学の新たな地平— 序論」，『東南アジア研究』46巻4号，京都大学東南アジア研究所，2009.
[5] 柴山守：『地域情報マッピングからよむ東南アジア』，勉誠出版，2012.
[6] 久保正敏：「時空間統合アーカイブズの構築を目指して」，『アジア遊学』113号，勉誠出版，pp. 152-161，2008.
[7] 関野 樹：「明治以降の「旧暦」のデータベース化」研究報告人文科学とコンピュータ（CH），2015-CH-107，pp. 1-4，2015
[8] 原正一郎：「HuMapの使い方」，『アジア遊学』113号，勉誠出版，pp. 136-139，2008.
[9] Ishikawa, Masatoshi; Umezaki, Masahiro; Hoshikawa, Keisuke: “Integration/Visualization of Data for Location, Physical Activity, and Landscape: An Application in a Field Study in Bangladesh”, PNC 2010 Annual Conference, 2010.
[10] Schmachtenberg, Max; Bizer, Christian; Jentzsch, Anja; Cyganiak, Richard:” Linking Open Data cloud diagram”, http://lod-cloud.net/, 2014.
[11] 帯谷知可編：『書誌情報データベースの地域情報学的新展開を探る』，CIAS Discussion Paper Series No. 51，2015.
[12] 「地域研データベース探訪 第1回 フィールドノート・データベース」，京都大学地域研究統合情報センター ニューズレターNo. 16，2015

事務局からのお知らせ

［1］事務局住所変更のお知らせ（再掲載）

2015年1月より，事務局の住所を変更いたしました。これは2014年4月から(株)アドスリーに委託している個人会員管理・年会費徴収事務に加えて，同年11月には会計管理事務，郵便物管理事務，賛助会員・学会誌定期購読者管理事務も同社に委託することが決定し，その後，実際に委託が進んできたためです。

変更前住所　〒110-8560　東京都台東区台東 1-5-1 凸版印刷㈱内

変更後住所　〒164-0003　東京都中野区東中野 4-27-37 ㈱アドスリー内

事務局からの郵便物は変更後住所からお送りしています。

会員の皆様が事務局宛てに郵便を送られる場合は変更後住所にお送りください。

凸版印刷㈱様のご厚意により，変更前住所宛ての郵便物も当分の間，受け取っていただけますが，郵便物を転送するため受け取りが遅くなります。そのため，郵送先は上述の変更後住所にしてください。宜しくお願いいたします。

なお，事務局の住所は変わりましたが，年会費振込み先の郵便振替口座番号，銀行口座番号は従来どおりで変更ありません。

［2］個人会員の皆様へ，年会費納入のお願い

1年分の年会費は正会員8千円，学生会員・ユース会員・シニア会員は4千円です。お手元に届いた学会誌の封筒の宛名ラベルには，ご自分の年会費の納入日が年度毎に西暦下2桁，月（2桁），日（2桁）の6桁の数字で印字されています。その数字が印字されていない年度は未納ですので，次に示す郵便局または銀行口座へお振込願います。郵便局で払込取扱票にご記入のうえ，ATMから読み込ませてお振込みくだされば，手数料も80円と安価で，窓口が閉まった時間帯でも可能のため便利です。

振込の後，事務局に通知が届くまで10日掛りますので，ご了承ください。

1. 郵便振替口座 00150-8-706543 情報知識学会
2. ゆうちょ銀行 〇一九店(ｾﾞﾛｲﾁｷｭｳ店) 当座 0706543 情報知識学会

請求書が必要な方はその旨，情報知識学会事務局にメールなどでお知らせください。

［3］新規入会申込方法

入会ご希望の方は情報知識学会ホームページ http://www.jsik.jp/から「本会について」→「入会案内」→「入会申込フォーム」に必要事項を入力・送信してください。

あるいは申込用紙をpdf形式，doc形式でダウンロードし，ご記入のうえ下記の事務局へ電子メール・FAX送信または郵送などでお願いいたします。

情報知識学会事務局
〒164-0003　東京都中野区東中野 4-27-37 ㈱アドスリー内
FAX:050-3730-8956　E-Mail:office@jsik.jp　URL:http://www.jsik.jp

情報知識学会誌 Vol. 25, No.4　2015 年 12 月 12 日発行 編集・発行 情報知識学会

頒布価格 3000 円

情報知識学会(JSIK: Japan Society of Information and Knowledge)

会長　石塚　英弘
事務局　〒 164-0003　東京都中野区東中野 4-27-37 (株)アドスリー内
FAX: 050-3730-8956
E-mail: office@jsik.jp　URL: http://www.jsik.jp/

データシェアリングシンポジウム

科学の発展への起爆剤

～データ駆動型科学の推進に向けて～

Open Symposium
Data-driven Science - The trigger of Scientific development

2016年2月29日(月)/Feb 29, 2016　10:00～17:00

一橋講堂 / Hitotsubashi Hall

主催:国立研究開発法人科学技術振興機構 / Organized by Japan Science and Technology Agency

アジア地域初開催！オープンサイエンス推進への議論を深める４日間
「第７回リサーチデータ・アライアンス（RDA）総会」および
「データシェアリングシンポジウム」の開催について

「科学の発展への起爆剤～データ駆動型科学の推進に向けて～」がシンポジウムのテーマ。第７回 RDA 総会に先駆けて開催される、JST 主催の主に日本人に向けた１日限りの貴重なイベントです。※同時通訳あり。

データ駆動型科学の推進はこれからの日本におけるイノベーションのための起爆剤となっていくはずです。

午前は政府、アカデミアなどの各界の有識者をお招きし、日本のデータシェアリングの「今」を語りつくす講演会を開催。午後は分野ごとのデータシェアリングの可能性と取り組みの議論を深めるセッションです。

①データシェアリングシンポジウム

【 テーマ 】科学の発展への起爆剤ーデータ駆動型科学の推進に向けてー

【 日　程 】平成 28 年 2 月 29 日（月）

【 参加費 】無料

【 主　催 】国立研究開発法人科学技術振興機構
※共催：国立研究開発法人産業技術総合研究所／国立研究開発法人情報通信研究機構／国立研究開発法人物質・材料研究機構／大学共同利用機関法人情報・システム研究機構／国立研究開発法人理化学研究所（予定）
※後援：内閣府（予定）/文部科学省（予定）/日本学術会議 /駐日欧州連合代表部 /米国国立科学財団　他

【参加方法】詳細確認・参加登録はこちらから：https://jipsti.jst.go.jp/rda

【お問合せ】運営事務局（株式会社アイ・エス・エス内）山本、安部
E-mail：datasharing@issjp.com　Tel：03-6369-9984

②第７回ＲＤＡ総会

【 テーマ 】Making data sharing work in the era of Open Science

【 日　程 】平成 28 年 3 月 1 日（火）ー平成 28 年 3 月 3 日（木）

【 参加費 】有料（料金は RDA のウェブサイトに掲載）

【 主　催 】RDA、科学技術振興機構による共同開催

【参加方法】RDA サイトの登録ページよりお申し込み下さい。 https://rd-alliance.org
※総会への参加にはメンバー登録（無料）が必要となります。

★本件に関するお問い合わせ先
科学技術振興機構 知識基盤情報部 計画管理グループ
Tel：03-5214-7980　Fax：03-5214-8460　E-mail：rda@jst.go.jp

Journal of Japan Society of Information and Knowledge

Contents

情報知識学会誌 第25巻4号　2015年12月12日発行
編集兼発行人 情報知識学会 〒164-0003 東京都中野区東中野4-27-37 ㈱アドスリー内
E-mail : office@jsik.jp
URL : http: //www. jsik. jp/　　(振替 : 00150-8-706543)
学術刊行物 ISSN 0917-1436