

研究論文

複合的な情報資源のLinked Open Data化における構造分析と スキーマ定義手法

Structure analysis and a schema definition method for creating

Linked Open Data from complex information resources

林賢紀^{1*}, 瀬尾崇一郎, 阪口哲男²

Takanori HAYASHI^{1*}, Soichiro SEO, Tetsuo SAKAGUCHI²

1 国際農林水産業研究センター

Japan International Research Center for Agricultural Sciences

〒305-8686 茨城県つくば市大わし 1-1

E-mail: tzhaya@affrc.go.jp

筑波大学大学院図書館情報メディア研究科

Graduate School of Library, Information and Media Studies, University of Tsukuba

〒305-8550 茨城県つくば市春日 1-2

2 筑波大学 図書館情報メディア系 / 知的コミュニティ基盤研究センター

Faculty of Library, Information and Media Science / Research Center for Knowledge Communities, University of Tsukuba

〒305-8550 茨城県つくば市春日 1-2

E-mail: saka@slis.tsukuba.ac.jp

*連絡先著者 Corresponding Author

Web技術によるデータ公開の方法としてLinked Open Data (LOD)が注目されている。しかし、既存のWeb上の情報資源の多くは人が読む利用形態に適したデータ構造のままであるなど、構造化が不十分であることが指摘されている。本研究においては、異なる性質の要素を持つ複合的な情報資源に対し、相互運用性を持ちかつ情報損失を起さずにLODを適用する方法について検討を行った。この結果、対象となる情報資源に記載されている情報を元にして、文書の構造や使用されている語彙などを分析することにより、LODへの再構成を効率的に行うことが可能であることを明らかにした。また、関連付けが可能な他の情報資源を用いて不足している情報を補うことを前提とした構造化により、人が読む利用形態に適したデータ構造に基づいていても適切なLODの適用を可能とし利活用しやすくするための一手法を示した。

Linked Open Data (LOD) have attracted much attention as a method for publishing data using Web technology. However, since many information resources on the Web are organized for human reading, they are not well structured for LOD. In this study, we investigate how to create LOD from complex information resources with different types of elements, achieving interoperability without any information loss. We found that by analyzing the structure and vocabulary based on the description of the information resource, it is possible to structure information efficiently to create LOD. Furthermore, we propose an approach for applying the appropriate LOD structure that compensates lack of information with linking other resources when using information resources designed for human reading.

キーワード: Linked Open Data, 情報組織化, 構造化, オープンデータ, 研究情報

Linked Open Data, Information organization, Structuring method, Open data, Research information

1 はじめに

Web の技術を利用したデータ公開の方法として、Linked Open Data (以下、「LOD」という.) が注目されている。Web ページがリンクにより相互に関係づけられるのと同様に、LOD は従来の Web 及びセマンティック Web の技術を用いて個々のデータを機械的に解釈可能とし、かつ関連付けることでデータの利活用度を高めることを目的としている[1]。

しかし、日本国内、とりわけ政府における Web での情報発信の現状を見ると、公開されているデータセット数を形式別に見た場合、その 41%が PDF, 37%が HTML と LOD には適さないデータ形式が多いことが示されている[2]。その内容も、「電子行政オープンデータ推進のためのロードマップ」(平成 25 年 6 月 14 日高度情報通信ネットワーク社会推進戦略本部 (IT 総合戦略本部) 決定) によれば、現在の Web 上のデータの多くは「人が読む (画面上又は印刷して) という利用形態に適したデータの構造 (タグの付け方, 表

の形式等) やデータ形式」で公開され、「検索も難しく、大量・多様なデータをコンピュータで高速に、横断的に又は組み合わせて処理・利用しようとした場合、データの構造やデータ形式を変換するためにコストや手間がかかっている」など構造化が不十分な状態にあることが指摘されている。

これを解消し、すでに存在するデータについて利便性を向上させるべく LOD を適用する取り組みがなされている。たとえば、後藤[3]は花園大学歴史博物館の資料情報を材料として、より簡易な資料情報のデジタル化と Linked Data 化を行っている。また、高橋ら[4]は、写真間を歴史的関連で関連付けられるよう、函館の歴史に関する文献を利用し写真資料のメタデータを補完し LOD として作成している。このほか、中田らは明石工業高等専門学校においてそのシラバスを LOD として公開している[5]。他の高専や大学が同様のオープンデータを公開すれば、学生が興味のある他校の科目を聴講するきっかけともなるとしている。さらに、汎用的に LOD を作成できる環境作りへ向けた取り組みとして、

「オープンデータ活用支援プラットフォーム」である LinkData.org では、既存のテーブルデータをオープンデータとして活用するためのサービスを提供している。

しかし、これまでの取り組みの多くは、すでに資料目録のような情報資源の組織化に適した形態で構造化がなされているデータを対象としており、先の指摘のように「人が読む利用形態に適した構造」を処理・利用するためには多くの手間が必要となっていることが課題である。

2 本研究で取り組む課題

LOD の適用によりデータを機械的に解釈可能にするためには、個々のデータに含まれる意味や文脈を把握し、適切に構造化を行うことが必要である。また、LOD を適用しその効果を最大限に発揮するためには、識別可能な実体に URI を与えて参照可能とし、インターネット上に存在する他の情報資源との関連付けを容易にすることで、相互運用性を高めることが必要である。このため、構造化にあたっては、他の情報資源などでも使用されているなど、一般的かつ汎用的な語彙の活用が考えられる。

一方、これまでデジタル化されたデータについては、先の指摘のように「人が読む利用形態に適した構造」のままの場合もあれば、その意味について検討がなされ、既存の語彙を適用する、あるいは対応する語彙を新たに定義するなどして構造化が行われていることも考えられる。従って、LOD の適用にあたっては、データが持っている元々の意味を分析、把握して適切な語彙を適用する一方、これまでに定義されている語彙とその情報を失うことがないように留意する必要がある。

そこで、本稿においては、すでに XML により構造化が行われた実データを使用し、効率的な LOD の適用手法について検討を行う。具体的には、すでに適用されている XML で使用されている語彙の分析と再定義、また汎用的な語彙の適用の可否の検討を行う。これにより、相互運用性の確保という点から語彙を見直すほか、紙の様式上で表現されていた、独自に定義された語彙によって表されていた情報について、その意味を失うことがないよう可能な範囲で LOD を適用する手法を明らかにする。

また、LOD の適用にあたっての検討過程で得られたこのような知見は、これまで Web で蓄積されてきた各種の研究情報について、データとしての利便性を高めることに寄与できると考える。

3 対象

大学や研究機関においては、その研究成果を様々な方法で発信している。研究の成果は論文として学術雑誌に掲載されるほか、研究機関独自の刊行物などにより提供されている。また、特許や実用新案、新品種などの知的財産などとしても公開されている。近年では、これら研究成果の多くは Web により発信されている。

農林水産省関係の試験研究機関においては、研究成果を他の研究機関や行政機関で応用する、あるいは一般で事業化するため、「成果が特に顕著で、広く利用を図ることが望ましいと考えられる」研究成果を要約し紹介する研究成果情報が当初は冊子で、現在は Web で刊行されている。国立研究開発法人日本原子力研究開発機構などでも同様の刊行物がある。以下、本稿では農林水産省関係の試験

研究機関で刊行されているものを「研究成果情報」という。

公的機関が保有する情報を広く一般に公開するというオープンデータの理念を考えれば、長年公開と利用がなされているこの研究成果情報は意義のある情報といえる。特に、研究成果情報は、行政機関の担当者や農業改良普及所の普及員、また農家や製造・流通業など、研究者ではない者へも新たな知見を紹介し普及を図る目途で刊行されており、その受け手は論文とは異なっている。このことから、従来の学術情報流通の手段の一つである論文誌への掲載と公開に加え、様々な情報資源との関連付けと活用が望まれる。

しかし、現状においては、冊子体の記載様式をほぼそのまま Web に置き換えた状態で

公開されている。たとえば、発表論文や特許などが研究成果情報には記載されているが、特にオンライン上で参照するためのリンクなどは行われておらず、より詳細な情報である論文等への容易なアクセスはできないなど、利便性は高くない。

そこで、単なるハイパーリンクではなく LOD を適用することにより、他の情報資源と意味関係を用いて関連付けることが容易となり、論文や特許などの研究成果と行政、普及、研究を結びつけるという本来の目的がより達成しやすくなると考えられる。

3.1 研究成果情報の構成

図 1 に紙媒体で刊行された研究成果情報の例を示す[6]。研究成果情報は、冊子見開

11. タイ在来野菜で汁により抗酸化性を付与した加工米飯

【要約】パジル等の野菜・葉草等の熱水抽出物を用いて炊飯することにより、米飯に抗酸化性や風味・色などを付与することができる。抗酸化性を付与した加工米飯をレトルト包装した製品や、様々な風味の米飯フライ加工品を製造することができる。

【キーワード】抗酸化性、加工米飯、タイ野菜、レトルト包装米飯、米飯フライ加工品

【所属】国際農林水産業研究センター 利用加工領域

【分類】技術 A'

【背景・ねらい】産業利用があまり進んでいない東南アジア在来野菜の生理機能特性や加工特性を明らかにして、機能性食品等の原料として利用することにより、在来農産物の高付加価値化を目指したものである。タイ・カセサート大学と共同で、抗酸化性（食品中や生体内で起こる有害な酸化反応を防ぐ働き）に富む在来野菜を用いて、流通・保存に適した食べやすい機能性食品を開発する。

【成果の内容・特徴】

- タイ・ラオスでは、ゆでた野菜をナムブリック（調味料ペースト類）とともに食べるのが好まれるが、この際、一般的にゆで汁は捨てられている。一方、野菜をゆでることにより抗酸化成分の一部が煮汁中に漏出する。煮汁の抗酸化性は再度加熱しても失われない。従って、野菜のゆで汁を用いて炊飯することにより、抗酸化性や野菜の風味を米飯に付与できる。
- ゆで汁の調製には、抗酸化性を有し、かつ、好ましくない風味や強い苦み・渋みを持たない野菜を材料として選ぶ。例としては、バックブーン (*Ipomoea aquatica*、在来型のアサガオナ)、カブラオ (*Ocimum sanctum*、ホーリーバジル)、マクアプロ (*Solanum xanthocarpum*、在来型のナス) などである。野菜は、下準備をした後に適当な大きさに切り、数倍容の水で柔らかくなるまでゆでる。冷却後、ろ過により熱水抽出液を得る。ゆでた野菜は副食として利用できる。
- うるち米、もち米、または混合米を上記抽出液とともに、鍋または炊飯器により炊飯することにより抗酸化性付加工米飯を製造できる (図 1、2)。また、この抗酸化性付加工米飯を用いてライスナゲットなどの加工品を製造できる (図 3、4)。
- バンコクの一般消費者 200 名以上を対象に試食およびアンケート調査を行った結果、嗜好的に概ね好評であり、ほとんどの回答者が購入を希望した。ベンチスケールでの製造コスト (人件費含む) は、レトルト包装米飯が 1 パック (135 g) 当たり約 8 パーツ (約 22 円)、ライスナゲットが 1 個 (30g) 当たり約 4 パーツ (約 11 円) であり、現地の経済水準から見ても、十分に市販可能と判断される。

【成果の活用面・留意点】

- 米に機能性等の付加価値を付与するタイの先行特許としては、様々な葉草の抽出物または粉末により表面をコートした米がある。
- レトルト米飯は室温で、ライスナゲットは-20℃で6ヶ月以上、抗酸化性を保った状態で保存可能である。

【具体的データ】

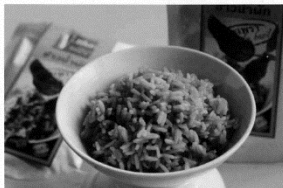


図 1 抗酸化性レトルト米飯




図 3 抗酸化性ライスナゲット

3 種類の在来野菜 (アサガオナ、ホーリーバジル、在来のナス) を使った製品を開発した。

3 種類の製品 (グリーンカレー風味、豚角煮風味、モチ米を使ったタロイモとココナツ風味) を開発した。

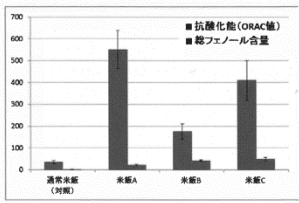


図 2 野菜熱水抽出物による米飯への抗酸化性の付与

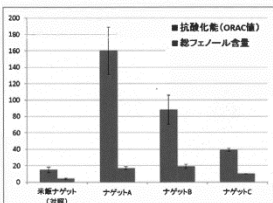


図 4 ライスナゲット製品の抗酸化性

【その他】研究課題：アジア農産物の高付加価値化
中課題番号：A-1)-(5)
予算区分：運営費交付金 (高付加価値化)
研究期間：2009~2010 年度
研究担当者：中原和彦・G. Trakoontivakorn・P. Tangkanukul (カセサート大学食品研究所)
発表論文等：1) Trakoontivakorn et al. (2010) Thai petty patent application No.1003001338.
2) Trakoontivakorn et al. (2011) JARQ 45(2), 211 - 218.

図 1 研究成果情報の例

きの左ページに概要が、右ページに図表という様式で構成されており、また図表を多く用いるなど学術論文より平易に記述されている。なお、発行機関や発行時期により、記載項目の名称や内容には若干の差異があるが、研究成果情報の様式としてはおおむね図 1 の通りである。

図 1 を元に研究成果情報の内容を概観すると、以下のような情報が記載されている。

- A) 研究成果の名称とその概要
- B) 研究担当者とその所属機関
- C) 実施した研究課題とその予算
- D) 研究を実施した期間

たとえば、研究成果について発表した論文等と A)~D)の情報について、相互に関連付けを可能にすることで、「誰がどのような研究成果を発表したか」また「ある研究課題からどのような研究成果が得られたか、またその概要」などを把握できる。また、これらの情報のうち、B)のような研究者や機関に関する情報についてはすでにデータセットが公開されているほか、D)には論文、特許、Web サイトなど様々な情報が記述されており、こ

れらについてもインターネット上の情報資源と関連付けが可能であると考えられる。

しかし、研究成果情報自体を一つのオブジェクトと見なし、図 1 に記載された名称や研究担当者、所属機関などの記述を研究成果情報のタイトルや著者を表すリテラルとして LOD を適用した場合、個々の研究担当者などについて実体という概念がないため外部の情報資源とは関連付けが難しく、LOD を適用する効果は得にくい。このため、研究成果情報を、研究担当者や所属機関などいくつかの実体とそれらについての関連の情報が複合した情報資源としてとらえて構造化を行うこととした。

一方、研究成果情報については、論文の書誌情報などとの横断検索を意図して XML による構造化も過去に行われている。この構造や使用されている語彙、また用法について他の情報資源との関連付けという観点から分析することで、LOD の適用方法について効率的な検討を行うことができると考えられる。

研究成果情報の XML データについては、

表 1 研究成果情報の項目名と使用されている語彙、記載例（抜粋）

| 大項目 | 項目 | 出力項目 | | | 記載例 |
|-------|-------------|-----------------|-------------------------|-----------|---|
| | | 親要素名 | 子要素名 | 属性 | |
| タイトル | タイトル(日) | dc:title | seika:title | lang="ja" | <seika:title lang="ja">タイ在来野菜ゆで汁により抗酸化性を付与した加工米飯</seika:title> |
| | タイトル(英) | | seika:title | lang="en" | |
| 担当機関 | 担当機関コード | dc:publisher | seika:institute_code | | <seika:institute_code>11454</seika:institute_code> |
| | 担当機関名 | | seika:institute | | <seika:institute>国際農林水産業研究センター</seika:institute> |
| | 担当部名(グループ名) | | seika:institute_group | | <seika:institute_group>利用加工領域</seika:institute_group> |
| | 担当室名(チーム名) | | seika:institute_team | | |
| | 備考 | | seika:institute_reserve | | |
| 研究担当者 | 氏名 | dc:creator | seika:person | | <dc:creator> <seika:person>中原和彦</seika:person> </dc:creator> <dc:creator> <seika:person>G. Trakoontivakorn</seika:person> </dc:creator> <dc:creator> <seika:person>P. Tangkanakul</seika:person> <seika:position>カセサート大学食品研究所</seika:position> </dc:creator> |
| | 所属 | | seika:position | | |
| 発表論文 | 発表論文 | seika:reference | seika:thesis | | <seika:reference> <seika:theses>1) Trakoontivakorn et al. (2010) Thai petty patent application No. 1003001338. </seika:theses> <seika:theses>2) Trakoontivakorn et al. (2011) JARQ 45 (2) :211 - 218. </seika:theses> </seika:reference> |
| 品種 | 品種出願(登録) | | seika:hinsyu | | |
| 特許 | 特許出願(公開) | | seika_patent | | |

農林水産研究情報総合センターが提供するデータベース AgriKnowledge の WebAPI により取得することができる[7]。また、出力される XML で使用されている語彙や定義内容については、農林水産研究情報総合センターより個別に提供を受けたものである。なお、研究成果情報について試行的に RDF を適用した語彙定義は、メタデータ基盤協議会が提供するメタデータ基盤システム Merabridge (<https://www.metabridge.jp/>) より参照できる。

3.2 XML 化された研究成果情報の構造

3.2.1 構造の分析

XML 化された研究成果情報は 51 の項目で構成される。本稿では、人物、組織名、書誌情報など、この研究成果情報以外にもインターネット上の情報資源が存在し関連付けの可能性があると考えられる「研究担当者」、「所属」、「発表論文等」に関連する項目について詳細を分析した。これらの項目を抜粋し

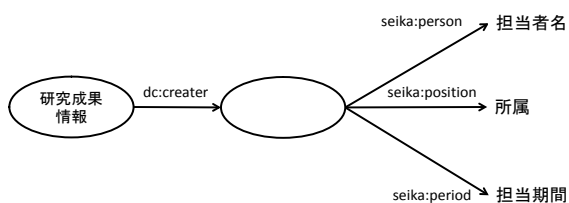


図 2 「研究担当者」の表現例

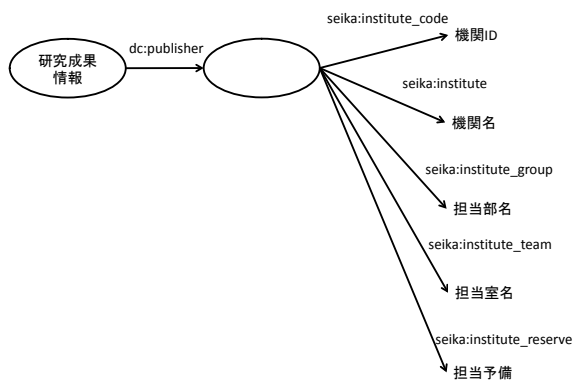


図 3 「所属」の表現例

表 1 に示す。名前空間については DCMI Metadata Terms[8]と seika[7]が使用されている。

「研究担当者」は、研究成果情報の著作者として dc:creator で表されており、この子要素は次の 3 つの属性から構成されている

- ・ 「担当者名」 (seika:person)
- ・ 「所属」 (seika:position)
- ・ 「担当期間」 (seika:period)

これらにより、研究成果情報について「ある所属の担当者が、担当期間の間に作成した」ことが表現されている (図 2)。

「所属」については、研究成果情報を出版した実体として dc:publisher で表されている。機関名にあたる記述は、子要素として以下の部署等で分割され格納されている。

- ・ 「機関名」 (seika:institute)
- ・ 「担当部名」 (seika:institute_group)
- ・ 「担当室名」 (seika:institute_team)
- ・ 「担当予備」 (seika:institute_reserve)

このほか、「機関名」に対してユニークな「機関 ID」 (seika:institute_code) が与えられている (図 3)。

「発表論文」は研究成果情報に対して参照する文献として seika:reference として定義され、その種類により以下の 3 つの子要素のいずれかに細分することで表現している (図 4)。

- ・ 「発表論文」 (seika:thesis)
- ・ 「品種出願」 (seika:hinsyu)
- ・ 「特許出願」 (seika:patent)

これらの 3 要素については、いずれも元の紙媒体の研究成果情報で一つの項目であっ

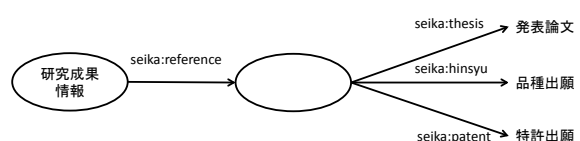


図 4 「発表論文」、「品種出願」、「特許出願」の表現例

たもの親要素とし、その詳細な項目を子要素に置き換えて木構造で表現しているが、子要素間の意味はそれぞれで異なる。

「研究担当者」の子要素のうち、「担当者名」は責任を表す実体であり、`dc:creator` で表現できるが、「担当期間」や「所属」は責任を表す実体というよりは「担当者名」に対する属性である。また、「所属」で記載されている「機関名」、「担当部名」、「担当室名」は一般的な組織を考慮すれば階層構造を取り得るが、木構造ではこの階層構造を表現し切れていない。「機関 ID」については、システム上は「機関名」に対して付与されているが、この構造ではその他の要素と同じ子要素の一つであり、どの子要素に対するものかは表されていない。一方、「発表論文」の子要素は全く別個のものであり、研究成果情報に関連して発表された情報ではあるが階層構造をとるものではない。

このように、各項目の記述内容について詳細には分析がなされず、人が読むことに最適化された記述を研究成果情報に特化して定義した語彙 `seika` により記述することで構造化が行われた箇所がある。また、XML の構造上、木構造による記述であるため、枝葉に相当する子要素同士の関連付けについては考慮されていない。このため、他の情報資源と項目名などを統一しての横断検索は可能となったが、情報資源相互の関連付けは困難で、十分な効果を得ることはできていない。

3.2.2 問題点

XML 化における問題点には、前項での分析の結果明らかになった元々の木構造に依存するという点のほか、紙の様式に記載された情報をそのまま XML に置き換えたことによる問題点も含まれる。たとえば、表 1 の構

造を持つデータに対して図 1 では「研究担当者」として「中原和彦・G. Trakoontivakorn・P. Tangkanakul (カセサート大学食品研究所)」と記載されている。この記載を人間が読解したとき、「・」を区切り記号として 3 名の研究担当者が存在すること、末尾に付加されている (カセサート大学食品研究所) が所属機関に含まれると推定することができる。

一方、研究成果情報のうち、「所属」に「国際農林水産業研究センター」と記載されていることから、この 3 名の所属機関は「国際農林水産業研究センター」とも考えられる。ここで、「P. Tangkanakul (カセサート大学食品研究所)」とあることから、最後の者はカセサート大学食品研究所が所属機関である可能性が高い。しかし、記述規則が明らかではないため、残りの 2 名の研究担当者の所属機関がこの 2 つのいずれかであることを機械的に区別することは困難である。

このため、XML 化された研究成果情報においては、親要素である `dc:creator` は 3 回繰り返され、それぞれの子要素の `seika:person` でその氏名が表されている。しかし、所属を表す `seika:position` は最後の 1 名のみが付与されており、残りの 2 名については、所属機関を表すことができていない。この例では「P. Tangkanakul は研究担当者で所属機関はカセサート大学食品研究所である」ことを表しており、厳密には異なる可能性もあるが「所属機関はカセサート大学食品研究所である」という情報を失うことなく構造化するにはやむを得ないと考えられる。

このように、紙の様式上では具体的な記載がなく、かつ人間の目では推定や解釈が可能な情報については、XML 化を行った結果としてこれらの情報が失われてしまう可能性がある。

4 LODによる再構成

本項では、XML化された研究成果情報のうち3.2.項で分析した「研究担当者」、「所属」、「発表論文等」に関連する項目について、LODの適用について検討を行った。

検討と結果の整理にあたっては、RDFグラフに準じた図を用いた。「研究成果情報」をリソースとして円で表し、その記述項目名とを矢印で結んでいる。四角はリテラルを表す。このとき、XMLで構造化した際に親要素または子要素で使用されていた語彙または一般的に利用されている語彙をプロパティとした。

この記述例を図5に示す。この例では、リソース「研究成果情報」(<http://example.com/642259>、以下の図においても同じ)の研究成果情報名(`dc:title`)が「タイ在来野菜ゆで汁により抗酸化性を付与した加工米飯」、ID(`dc:identifier`)が「642259」であることを表す。

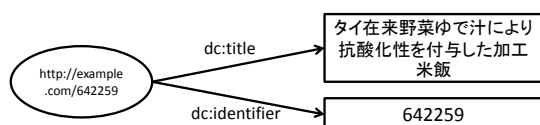


図5 研究成果情報のグラフによる図示

4.1 適用の方針

一般的に使用されている語彙を使用することで、構造化されたデータ同士の相互の関連付けがより容易になる。そこで、相互運用性の観点から、研究成果情報で使用されている語彙について一般的な語彙かどうか、異なる場合には置き換えが可能か否かを検討した。

3.2.項で分析したように、XMLによる研究

成果情報の構造化では、その語彙として主として親要素に DCMI Metadata Terms が、子要素に seika が使用されている。DCMI Metadata Terms については様々な分野で使われている語彙であり、問題なく使用することができると考えられるが、seika は一般的な語彙とは言えない。しかし、単純に使用する語彙を seika から DCMI Metadata Terms に置き換えた場合、元の研究成果情報で表現されていた意味や情報が失われる可能性がある。このため、3.2.項での分析結果を踏まえ、FOAF など一般的な語彙がある場合はその語彙を、適切な語彙がない場合は現に使用されている seika をそのまま使用することとした。

また、図2~4で示したとおり、親要素に DCMI Metadata Terms に使用してその項目の属性を、また子要素において seika により詳細な属性をそれぞれ表現している例が多い。このような場合は、リソースに対するクラスまたは一般的な語彙のサブクラスとして用いることで、seika にもともと有していた意味を表現することとした。

これらの方針の適用により、研究成果情報への LOD の適用にあたって相互運用性の確保と元の XML で表現されている意味や情報の保持ができると考える。

4.2 LODの適用

4.2.1 研究担当者

元の木構造を重視して LOD を適用した場合、図2の空白ノードに対して URI を割り当てる方法も考えられる。しかし、この方法では、子要素であった担当者名と所属の関係、また記述されている内容が人を表すものか組織を表しているのかなど、その属性を定義することができない。

そこで「担当者名」と「所属」について、

それぞれ人物情報の表現に多く用いられている FOAF (Friend of a Friend)[9]の語彙を使用して属性（人 (foaf:Person) または組織 (foaf:Organization)）と名称 (foaf:name) を表現した。

この 2 者の相互の関係については、XML では所属する組織を表す語彙として seika:position で表されていたが、より一般的な語彙に置き換えても元の意味は失われなと考え、schema:memberOf を使用した。担当期間については XML で使用されていた語彙 seika:period をそのまま使用した。この関係を図 6 に示す。なお、本来は研究成果情報を提供する者のドメイン名を URI に使用するべきであるが、本稿においてはドメイン名として http://example.com/ を用いる。

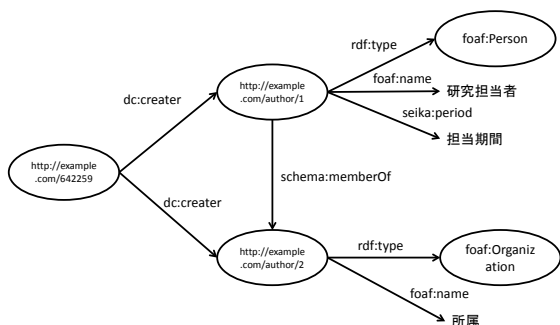


図 6 「研究担当者」の LOD

4.2.2 所属

「所属」については、研究担当者と同様 FOAF を語彙として用いたほか、組織とその階層を表すことができる schema:Organization を用いて表現した。この例を図 7 に示す。このとき、「担当部」(seika:Institute_group)「担当室」(seika:Institute_team) など、XML で使用されていた語彙が持っていた組織の属性に関する情報を失うことがないように、また意味を明確にするため、これらの語彙をそのまま組織を表す schema:Organization のサブクラスとして定義した。紙の様式や XML で推定できる組織の階層構造については、schema.org の語彙である schema:memberOf によりオブジェクト同士の関連付けを行った。これにより、「機関」と「担当部」の機械的な区別を可能とした。

例えば「機関」が「国際農林水産業研究センター」, 「担当部」が「利用加工領域」である場合は次のように表現できる。

- http://example.com/pub/1 で表される実体の名称 (foaf:name) は「国際農林水産業研究センター」である
- http://example.com/pub/1 はクラス「機関」

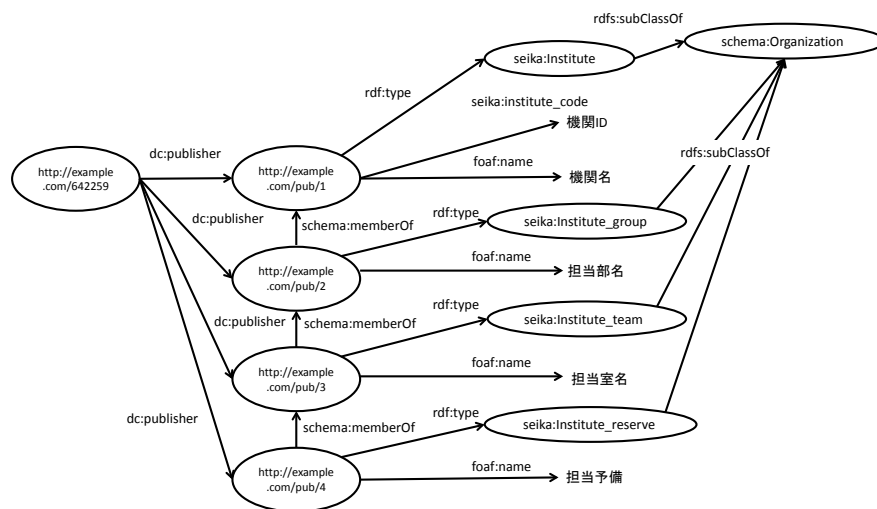


図 7 「所属」の LOD

(seika:Institute) に属する

- <http://example.com/pub/2> は、「利用加工領域」という名称をもち、かつクラス「担当部」(seika:Institute_group) に属する
- クラス「機関」及び「担当部」は組織を表すクラス schema:Organization のサブクラスである
- クラス「担当部」はクラス「機関」の一部 (schema:memberOf) である。従って、「利用加工領域」は「国際農林水産業研究センター」の一部である。

4.2.3 発表論文等

「発表論文等」には、図 1 及び表 1 に示すほかに以下のような記載が見られる。

例 1: 1. Yoshimichi Fukuta et al.(2012). Breeding Science 62:27-37.

例 2: 小宮山博 (2005): モンゴル国酪農家実態調査及び牛乳・乳製品消費状況調査の分析. 2005 年度統計関連学会連合大会講演報告集, 396-397.

これらの例から、「発表論文等」では著者名、発行年、論題、掲載誌名、巻号、掲載ページなどの書誌情報が区切り記号によって区切られて記載されていると判断できる。また、例 1 には論題と見なせる情報がないが、例 2 には記載がある。このように、記述様式はそれぞれ異なり一様でない。

このため、著者名、掲載誌などと関連付けが行えるよう LOD を適用するためには、1 件 1 件の記述内容自体を分析し詳細に構造化する方法が考えられる。しかし、ここでは過去の全てのデータに対して正確な記述と構造化を行うのではなく、まずは現在記述さ

れている書誌情報を実体として識別可能とした。また、書誌情報そのものは、一般的な内容記述に用いられる dc:description により表現した。これにより、他の LOD との関連付けから記述の精緻化を行うことが期待できる。

論文、特許など、記述されている書誌情報の属性については、次のように定義した。まず、個々の書誌情報に実体を与え、これらを研究成果情報に対する参考文献 (dcterms:references) として定義した。次に、XML で使用されていた seika:Ronbun などの語彙をクラスとして定義し、個々の属性を識別可能とした。さらに、これらのクラスについて、より一般的な語彙で定義されている属性のサブクラスであることを rdfs:subClassOf により定義し、この意味を明確にした。

図 8 に示す通り、「特許出願」<http://example.com/ref/13> の内容は dc:description により表されるが、その属性 (rdf:type) は seika:Patent である。このとき、seika:Patent は jst:Patent[10] のサブクラスであるため、特許に関する情報であることを表す。

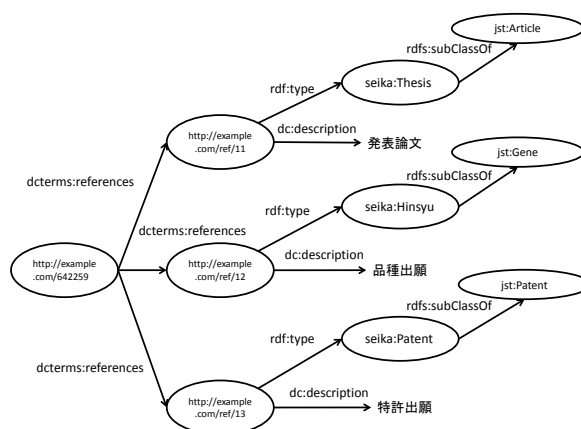


図 8 「発表演文」、「品種出願」、「特許出願」の LOD

5 考察

4章では、XMLで構造化された研究成果情報の分析結果を踏まえて DCMI Metadata Terms や FOAF などの一般的な語彙により LOD の適用を行った。これにより、他の情報資源との関連付けによる相互運用性の向上や、XMLでは明確に表すことができなかった人名、組織名といった属性の区別とその関係性を表現した。また、もともと表現されていた属性などの情報を保持したまま LOD の適用を行うことができた。

また、記述が省略されているなど情報がいまいで関連付けが困難な場合でも、時間をかけて精緻化を図るのではなく可能な範囲で構造化を行うことで、効率的に LOD を適用することができた。

本項では、LOD の適用により外部の情報資源との関連付けがどの程度可能になり、また相互運用性を向上させることができたかについて検討する。

5.1 対象とする情報資源

研究成果情報については、3章で分析したとおり、この研究成果情報以外にも研究者や機関に関する既存のデータセットや、論文、特許、Web サイトなどとの関連付けが可能と考えられる。これらの情報については、1章で触れたようにすでに構造化がなされ LOD が適用されているデータセットもある。

研究課題については、科学研究費助成事業データベース (KAKEN) について LOD の適用がすでに行われている[11]。以下、本稿では、科学研究費助成事業データベースについて LOD を適用した語彙及びそのデータセットを KAKEN と称する。KAKEN は、研究分野、研究機関、研究課題、研究者などのクラ

スとその概要や研究分野名、研究課題名、研究課題番号などを表すプロパティから構成される語彙を有している。また、オープンデータのカタログサイトである datahub 上で KAKEN のデータセットの情報が公開されている[12]。

組織を表す語彙としては、W3C が組織構造のためのコア・オントロジーとして Organization Ontology を勧告している[13]。このオントロジーでは、ある階層構造のもとに他の複数の組織で編成されている組織の概念をサポートし、組織の階層構造やその分類、沿革などを表現することができる。

図書館においては人物や組織、論文などの情報資源について、メタデータの作成と構造化がすでに行われている。W3C 図書館 Linked Data インキュベータグループが取りまとめた Library Linked Data Incubator Group Final Report[14]においては、図書館がこれまで収集、作成したこれら書誌データ・典拠データ・件名標目表・分類表といった情報資源を可視性と再利用性の高いものにするために、どのようにセマンティックウェブ標準や Linked Data の原則が適用可能であるかを検証している。

書誌データなどについては、すでに国立国会図書館ダブリンコアメタデータ記述 (DC-NDL) [15]や英国、ドイツ、フランスなど各国の図書館において LOD の適用が行われている。これらそれぞれの図書館で作成された書誌 URI から得られるグラフについて、神崎[16]が比較検討を行っている。神崎は、書誌、典拠の情報項目は目録規則に準じて概ね共通しているが、述語として用いる語彙 (プロパティ) が異なることなどを指摘している。たとえば、タイトルの語彙については、dcterms:title, dc:title, schema:name の三

通りがあり、大半は Dublin Core を使用しているが、図書館以外の領域との共有から WorldCat のように schema.org を採用する有用性もあるとしている。

以下、本稿で行った LOD の適用について、相互の運用可能性の検証のため、前述の書誌データなどのスキーマを対象に比較と検討を行う。これまで明らかにしたとおり、使用する語彙を単純に特殊なものから一般的なものに置き換えた場合、元の研究成果情報で表現されていた意味や情報が失われる可能性がある。そこで、情報の損失を起ささないか否かを主眼に置いて比較検討を行う。

5.2 比較検討

5.2.1 人物

研究成果情報においては、人物に相当する「研究担当者」に対して FOAF を語彙として用いたが、KAKEN においても人物を表現する語彙を有している。また、より詳細なプロパティが KAKEN では定義されている。

KAKEN ではクラス `Researcher` が `foaf:Person` のサブクラスとして定義されており、研究成果情報と同様に研究課題を実施した研究者を指すものとされている。いずれも、研究者という実体が人物であることを `foaf:Person` により表現している点では共通している。

同様に、図 6 で示した「研究担当者」に関連付けられる「所属」に相当するクラスとして、KAKEN では `Institution` が定義されている。このクラス `Institution` は、研究課題を実施する研究機関を指し、クラス `Researcher` とは研究課題実施時に研究者が所属した機関を表すプロパティ `affiliatedInstitution` で関連付けられている。加えて、KAKEN のクラス `Researcher` ではプロパティ `role` により研究

代表者、研究分担者、連携研究者などの研究課題における役割を表現することができる。

このように、研究成果情報と KAKEN では類似するスキーマを用いて研究者という実体と所属組織との関係性を表現しており、相互の関連付けも容易となる可能性が考えられる。

しかし、それぞれの「研究担当者」の範囲は異なっている。

研究成果情報については、図 1 によれば予算区分として「運営費交付金〔高付加価値化〕」と記載されているように、科学研究費補助金以外の資金による研究課題が含まれる。したがって、「研究担当者」には科学研究費補助金における研究者以外の者が含まれていることがわかる。

一方、KAKEN の語彙では、「研究課題」はクラス `Project` として定義されているが、`Comment` には何も記載されていない。このため、科学研究費補助金における研究課題であると解釈するのが自然である。ここで、クラス `Researcher` が指す「研究者」は `Comment` によれば「研究課題を実施した研究者を指す。研究代表者、研究分担者、連携研究者、協力研究者、受入研究者、特別研究員、外国人特別研究員などはこれに属す。」とあることから、科学研究費補助金の研究課題を実施した研究者のみを示すものと考えられる。

このように、同じ「ある研究課題を実施した研究担当者」を指す実体に対し、研究成果情報と KAKEN ではその範囲が異なるため、より広い範囲を含むことができる語彙を用いる必要がある。

また、3 章で示したとおり、研究成果情報においては複数の研究担当者とその所属を機械的に対応づけることは困難である。

これらの点から、研究成果情報においては

より foaf:name などの一般的な語彙を使用したほか、元の研究成果情報で有していた意味を表す独自の語彙を定義した。

5.2.2 組織

研究成果情報のうち「所属」では、研究担当者の組織に関する情報について、`schema:Organization` のサブクラスとして「機関」「担当部」「担当室」の3つのクラスを独自に定義した。また、推定できる組織の階層構造を `schema:memberOf` により定義した。

一方、`Organization Ontology` では、組織を表すクラス `org:Organization` と階層構造を表すプロパティ `org:subOrganizationOf` と `org:hasSubOrganization` を用いて一般的な組織とその階層を表すことができる。

また、前項で検討した KAKEN も組織を表す語彙を有している。研究課題実施時に研究者が所属した機関とその部局はプロパティ `affiliatedInstitution` 及び `affiliatedDepartment` で、プロパティ `jobTitle` が職名をそれぞれ表す。この3者は「所属機関と職名」として定義されたクラス `AffiliationTitle` で列記される。

研究成果情報と `Organization Ontology` を比較すると、いずれも組織体であること、またその階層構造を表現することができる。同様に、KAKEN の語彙によっても組織体であることは表現できるが、階層構造を表すプロパティは含まれていない。このため、階層構造を表現するためには、KAKEN ではなく、`schema:Organization` や `Organization Ontology` を使用することが考えられる。

ここで、研究成果情報と他の情報資源との関連付けを容易にするという観点から、`schema:Organization` を用いて組織とその階層を定義した。`schema:Organization` は各国の図書館で作成された典拠データを相互に関連

付けている VIAF[17]で使用されていることから、より多くの情報資源との関連付けやこれらを用いた情報の精緻化が期待できる。

5.2.3 文献

研究成果情報のうち「発表論文等」には著者名、掲載誌名など書誌事項が含まれることから、書誌データのための語彙である DC-NDL などの利用が期待できる。たとえば、著者名については DC-NDL のうち「作成者に関する語彙」である当該情報資源の作成者を URI で表現する `dcterms:creator` やリテラルで記述する `dc:creator` の適用が考えられる。

しかし、4章で検討したとおり、「発表論文等」は引用文献の形式で書誌事項を表しているが、その記述様式は一様ではない。たとえば、区切り記号を元にして機械的に記述内容を分割する手法も考えられるが、記述様式が異なることから誤った分割が行われる可能性がある。

このように、現状において書誌事項の個々の要素に対し適切な語彙を適用することは困難であるため、引用文献形式の記述全体を一般的な内容記述に用いられる `dc:description` により表現した。

情報損失を防ぐために記述されている情報を分割せずそのまま表現する手法は、KAKEN においても見られる。KAKEN 上の研究実績報告書では、発表文献を表すクラス `Publication` において、研究課題の発表文献リストを表すプロパティ `publications` が定義されている。この中で引用文献形式の記述全体がプロパティ `dcterms:bibliographicCitation` により表されており、この内容を人が見ることで発表文献を把握できる。また、文献によっては、`rdfs:seeAlso` により DOI や CiNii Article 上の情報資源が示されている。

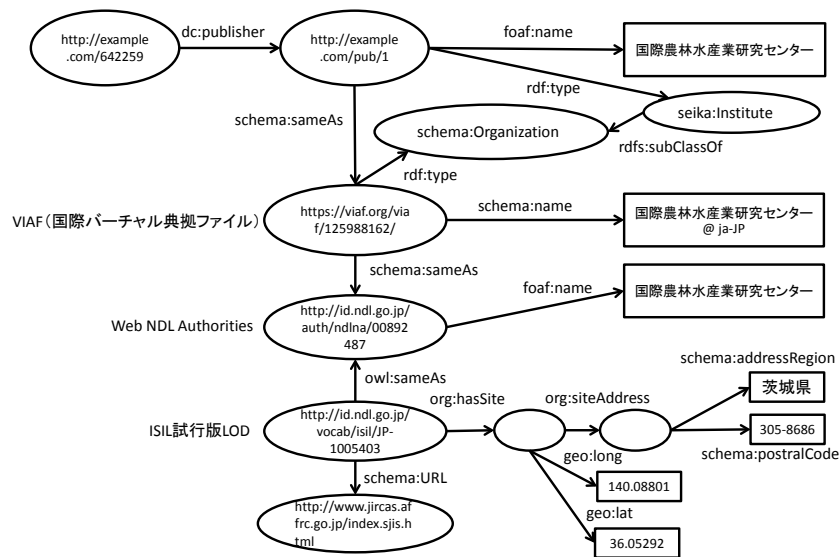


図9 他の情報資源との関連付けの例

これらのことから、将来的に適切な要素の分割が行われるとしても、記述された情報の全体を残すことにより、構造化された語彙の適用と情報の損失を防ぐことを両立できる。

5.3 他の情報資源との相互接続

本項では、研究成果情報について適用した LOD について、他の情報資源との相互接続を検討する。

研究成果情報の「所属」のうち、「組織名」については一般にその名称が重複していることが少なく、名称と属性が同一であれば実体も同一であると見なすことができる。このため、図書館で作成している典拠データなどとの関連付けも可能であると考えられる。

前項で触れた VIAF を介して、Web NDL Authorities (国立国会図書館典拠データ) [18] 及び「図書館及び関連組織のための国際標準識別子 (ISIL)」 試行版 LOD [19] との関連付けを行った例を図9に示す。

まず、研究成果情報の作成者 (dc:publisher) で表される実体 (<http://example.com/pub/1>) は、名称 (foaf:name) が「国際農林水産業研

究センター」であり、組織を表すクラス schema:Organization のサブクラス seika:Institute に属する。

一方、VIAF には日本語 (ja-JP) の名称 (schema:name) が「国際農林水産業研究センター」である実体 (<https://viaf.org/viaf/125988162/>) があり、同じく組織を表すクラス schema:Organization に属する。

VIAF 上の実体は、Web NDL Authorities の実体 (<http://id.ndl.go.jp/auth/entity/00892487>) と同一 (schema:sameAs) であると関連付けがなされている。また、ISIL 試行版 LOD の実体 (<http://id.ndl.go.jp/vocab/isil/JP-1005403>) は、Web NDL Authorities の実体と同一 (schema:sameAs) と関連付けがされている。

以上のことから、これらの実体は、名称も属性も同じであり、またすでに相互に関連付けが行われている実体があることから、すべて同一であるとして関連付けることができる。

また、この関係を利用して、研究成果情報の作成者と ISIL 試行版 LOD が持つ住所 (org:siteAddress) や緯度 (geo:lat) 経度

(geo:long), 組織情報の URL (schema:URL) といった情報を関連付けることができる。

このように, LOD を適用したことで外部の情報資源との相互接続が可能となる。また, 外部の情報資源を利用して, 研究成果情報が持つ記述の精緻化などに利用できる。

5.4 相互運用性の向上

これまで示したとおり, 本稿の手法では紙の様式に記載された情報をそのまま XML に置き換えたことによる問題点の全てを解消することはできていない。

では, 元の情報がどのような記述であれば適切に LOD を適用でき, また相互運用性を向上させることができるのか検討する。

3.2.2 項で示した「研究担当者」と「所属」のように, 紙の様式上で相互の関係について具体的な記載がない場合には, 人と組織の関連付けを行うことはできていない。あるいは 4.2.3 項の「発表論文」のように, 他の情報資源との関連付けが行える可能性を有しながら個々の要素に対し適切な語彙を適用することが困難な例もある。

3.2.2 項の例では, 仮に「中原和彦 (国際農林水産業研究センター)・G. Trakoontivakorn (カセサート大学食品研究所)・P. Tangkanakul (カセサート大学食品研究所)」のような記述であれば, 研究担当者名と所属の関連付けを明確にすることができる。4.2.3 項の「発表論文」についても, 記述されている引用文献記述の項目やこれらを区切る記号について一定のルールが用いられていれば, 機械的に分割することが可能である。

このように, 記述を省略し人間の目で情報が補われることを前提とするのではなく, 必要な情報を具体的に記述することで, 各項目

の内容についての分析と情報の意味の把握をより詳細に行うことができると考える。さらに, 各項目の情報について相互の関係や意味が明らかである場合には, LOD の適用にあたって適切な語彙を与えることができ, また相互運用性の向上が期待できる。このことは, 所属について分析した 4.2.2 項にて示したとおりである。

これらの点から, LOD の適用対象となる情報資源においては, 相互関係やその意味を適切に把握しうる記述が必要であるといえる。

一方で, 記述されている要素を分析した結果, 元の情報資源にはない ID などの識別子を加える必要が生じることも考えられる。

たとえば, 人名については, 同姓同名などの問題が考えられるため, 自動的な名寄せ作業のみでは同一性の判断や関連付けは困難である。CiNii Article では, 著者名に対して, ユニークな ID を付与し自動的な名寄せを行い, これに本人など人間が判断した結果を加えて著者の同定を行っている[20]。これらの事例から, 研究成果情報においても氏名に関連する所属や ID に相当する項目を持つことができれば, 機械的な同定作業が可能であると考えられる。

他の情報資源との関連付けを行い相互運用性を向上させるためには, 他の LOD と共通的に使われる一般的な語彙を使用するなど, 語彙の選定も重要である。先に触れた書誌データの LOD における神崎の分析[16]の通り, タイトルで使用する語彙が Dublin Core と schema.org と分かれているような場合には, それぞれ関連付けが行える情報資源が異なるなど, 相互運用性に影響すると考えられる。

本稿における LOD の適用においては, 語

彙 *seika* について一般的な語彙に置き換えるかサブクラスとして再定義した。一方、語彙 *seika* と一般的な語彙を両方使用し、1つのリソースに対して複数のクラスに属するよう構造化する方法も考えられる。DBpedia でも同様の方法で構造化が行われており、例えば <http://ja.dbpedia.org/resource/牧野富太郎> では `owl:Thing` , `schema:Person` , `foaf:Person` , `dbpedia-owl:Agent` などが `rdf:type` として表されている。同様の手法を用いた場合、語彙 *seika* についてその意味を失うことはない。しかし、他の LOD との関連付けに用いる一般的な語彙との関係を特定することが難しい、また全体のデータ量が増加するなどの問題が考えられる。研究成果情報への LOD の適用においては、相互運用性の向上を意図したことから、リソースの持つ性質を可能な限り特定することを重視した。

6 まとめ

本稿では、構造化にあたって多くのコストや手間がかかるとされてきた、「人が読む利用形態に適した構造」を持つ情報資源について、相互運用性を有する LOD による構造化を情報の損失なく効率的に行う手法の一つを示した。

具体的には、対象となる情報資源に記載されている情報を元に文書の構造や使用されている語彙などを分析することにより、内在している情報の意味を把握し、人が読む利用形態に適した構造から機械処理が可能な構造の一つである LOD への再構成を効率的に行うことができた。本稿で対象とした研究成果情報は、XML により構造化が行われた際にある程度の構造の分析と適切な語彙の定義がなされていたため、内容の分析と情報の

意味の把握が容易に行えた。再構成の過程においては、一つの文書を複数のオブジェクトに分割しそれぞれに実体を与えることで、外部の情報資源との容易な関連付けを可能とした。

また、研究成果情報のうち「所属」への LOD の適用で示したとおり、関連付けが可能な他の情報資源を用いて不足している情報を補うことを前提とした構造化を行った。これにより、VIAF, Web NDL Authorities, ISIL 試行版 LOD など、インターネット上のその他の LOD と関連付けを行うことで元のデータを修正することなく精緻な情報を得ることが可能となった。

一方、研究成果情報における「研究担当者」と「所属」の関係のように、紙の様式上で具体的な記載がない場合には現状では関連付けを行うことできていない。しかし、実体を与えた上で「人に関する情報である」という属性を付加することで、後から人手や他の情報資源上の情報を元にした同定作業を行う際に役立てることができると考える。「発表論文等」でリテラルとした書誌情報についても同様である。

実際の同定作業については、クラウドソーシングの利用も考えられる。例えば、森嶋らはクラウドソーシングプラットフォーム Crowd4U を利用して、国立国会図書館の書誌データの書誌誤同定の判定をマイクロタスク（短時間で作業ができる小さなタスク）の集合として分割し不特定多数の人々に委託することで、週平均平日タスク数 600～1000 程度の作業を行っている[21]。このような手法を用いて同定を行うことで、機械的な同定処理だけでは正確な判定が困難な作業についても正しい判定を行うことが可能であり、本研究で対象とした研究成果情報につ

いても応用性は高いと考えられる。

日本国内の Web 上の情報資源について LOD に適さない形式が多いことは冒頭で指摘した通りである。今後は LOD に適した形式による情報提供が増えることも想定されるが、これまで 20 年以上にわたり提供されてきた大量の情報資源をどう取り扱うかについては重要な課題である。たとえ一部であっても本稿で示した手法により LOD を適用することで、相互運用性や利便性の向上を図るなど、長年にわたって提供されてきた情報資源の価値をさらに高めることが必要である。

謝辞

本項の執筆にあたりご討論頂いた、筑波大学図書館情報メディア系の杉本重雄教授に深く感謝いたします。

本稿にて分析対象とした研究成果情報の XML データで使用されている語彙や定義内容については、農林水産省農林水産技術会議事務局筑波産学連携支援センター研究情報活用促進課の村田竜一課長補佐、土井亮文献情報係長、相川美由紀技術移転係長よりご提供いただきました。感謝申し上げます。

参考文献

- [1] Tim Berners-Lee: "Linked Data", <http://www.w3.org/DesignIssues/LinkedData.html>, 2006. (2016 年 1 月 28 日参照)
- [2] 内閣官房情報通信技術 (IT) 総合戦略室: 「データカタログサイト試行版「DATA.GO.JP」の最新の状況」, <http://www.kantei.go.jp/jp/singi/it2/densi/da/i6/sankou1.pdf>. (2016 年 1 月 28 日参照)
- [3] 後藤真: 「博物館資料情報の LinkedOpenData 化へ向けたモデル試作-

- 花園大学歴史博物館資料を題材に-」, 情報処理学会研究報告.人文科学とコンピュータ研究会報告, 2013(5), pp. 1-6, 2013.
- [4] 高橋正輝; 奥野拓; 川嶋稔夫: 「函館の歴史資料を用いた地域写真アーカイブの編纂」, 情報処理学会研究報告. DD, [デジタル・ドキュメント], Vol. 2013-DD-88, No.9, pp. 1-6, 2013.
 - [5] 中田季利; 新井イスマイル: 「明石高専のシラバス」, LOD Challenge2013, http://lod.sfc.keio.ac.jp/challenge2013/show_status.php?id=d049. (2016 年 1 月 28 日参照)
 - [6] 国際農林水産業研究センター: 「タイ在来野菜ゆで汁により抗酸化性を付与した加工米飯」, 国際農林水産業研究成果情報, 第 18 号, pp.21-22, 2010. Web 上のものは <http://agriknowledge.affrc.go.jp/RN/3010025669> を参照のこと. (2016 年 1 月 28 日参照)
 - [7] 農林水産研究情報総合センター: AgriKnowledge ウェブサービス (WebAPI), <http://agriknowledge.affrc.go.jp/api/>. (2016 年 1 月 28 日参照)
 - [8] Dublin Core Metadata Initiative (DCMI) が定めるインターネットの資源のメタデータを記述するために定義される語彙. 詳細は <http://dublincore.org/documents/dcmi-terms/> を参照のこと. (2016 年 1 月 28 日参照)
 - [9] Dan Brickley, Libby Miller: "FOAF Vocabulary Specification 0.99", <http://xmlns.com/foaf/spec/>. (2016 年 1 月 28 日参照)
 - [10] プレフィクス jst は J-GLOBAL knowledge で使用されているメタデータ語彙 <http://vocab.jst.go.jp/terms/sti> を表す. J-GLOBAL knowledge の詳細については, https://stirfd.jglobal.jst.go.jp/#learn_more を参照のこと. (2016 年 1 月 28 日参照)
 - [11] National Institute of Informatics, Research and Development Center for Scientific Information Resources: "KAKEN Vocabulary Specification", <https://kaken.nii.ac.jp/ns>. (2016 年 1 月 28

- 日参照)
- [12] National Institute of Informatics : "KAKEN", <http://datahub.io/ja/dataset/kaken>. (2016年1月28日参照)
- [13] W3C: "The Organization Ontology", <http://www.w3.org/TR/vocab-org/>. (2016年1月28日参照)
- [14] Thomas Baker, et al. : "Library Linked Data Incubator Group Final Report", <http://www.w3.org/2005/Incubator/ld/XGR-ld-20111025/>. (2016年1月28日参照)
- [15] 国立国会図書館 : 「国立国会図書館ダブリンコアメタデータ記述 (DC-NDL)」, <http://www.ndl.go.jp/jp/aboutus/standards/meta.html>. (2016年1月28日参照)
- [16] 神崎正英 : 「MLA とリンクするデータ 各国図書館の書誌・典拠グラフを中心に」, <http://www.kanzaki.com/works/2014/pub/1107lbf.html>. (2016年1月28日参照)
- [17] Virtual International Authority File (国際バーチャル典拠ファイル) の略. 各国の国立図書館等から典拠データの提供を受け、個人、団体といった同一の実体に対する典拠レコードを同定し、相互にリンクさせるシステム. OCLC が維持管理を行っている. <http://viaf.org/>. (2016年1月28日参照)
- [18] 国立国会図書館 : 「国立国会図書館典拠データ検索・提供サービス (Web NDL Authorities)」, <http://id.ndl.go.jp/auth/ndla>
- [19] 福山樹里 : 「NDL における LOD の作成と試行提供 : ISIL を対象として」, カレントアウェアネス-E, No.281, 2015, <http://current.ndl.go.jp/e1675>. (2016年1月28日参照)
- [20] 武田英明 : 「「名寄せ」の最前線 国内と海外の最新動向 - NII 研究者リゾルバーと ORCID プロジェクト -」, 第10回図書館総合展フォーラム, http://www.nii.ac.jp/userimg/libraryfair2010/nayose_takeda.pdf. (2016年1月28日参照)
- [21] 森嶋厚行 ; 川島隆徳 ; 原田隆史 ; 宇陀則彦 : 「クラウドソーシングってどうですか? Crowd4U × NDL データの事例」, 情報処理学会研究報告人文科学とコンピュータ (CH), Vol. 2015-CH-106, No. 13, pp. 1-4, 2015.