

「地球環境学リポジトリ」－セマンティック技術による 研究資源の異分野連携

関野 樹^{†1} 原 正一郎^{†2} 近藤康久^{†1} 窪田順平^{†1} 秋道智彌^{†1}

総合地球環境学研究所を中心に進められている「大学間連携を通じた広域アジアにおける地球環境学リポジトリの構築－環境保全と地域振興を目指す新たな知の拠点形成事業－」事業は、30余りの国公立大学の研究センターなどが参加し、地球環境学に関連するそれぞれの研究資源へのアクセスを可能にする、いわば学問分野のためのリポジトリの構築を目指すものである。研究資源の対象としては、地域研究や歴史などの人文科学分野の資料からリモートセンシングや環境モニタリングなどの自然科学系のデータまで多様なものが含まれる。本報告では、これらの研究資源をセマンティック技術による「知識」という共通の枠組みで捉える試みについて紹介する。

Repository for humanity and nature – Interdisciplinary resource sharing using semantic technology

TATSUKI SEKINO^{†1} SHOICHIRO HARA^{†2} YASUHISA KONDO^{†1}
JUMPEI KUBOTA^{†1} TOMOYA AKIMICHI^{†1}

“Repository for Humanity and Nature in Asia Region- Inter-University Knowledge Base for Global Environmental Issues” is a project of Research Institute for Humanity and Nature collaborating with more than 30 research institutes of universities in Japan. Main objective of the project is interdisciplinary resource sharing among the institutes, and the project will develop a repository for environmental study, using semantic technology.

1. はじめに

総合地球環境学研究所（地球研）では、地球環境問題の解決に資するために、自然科学だけでなく人文科学、社会科学を含めた学際的な学問分野である「地球環境学」の構築を目指している。地球研自身による地球環境学に関連する研究成果については、論文やデータなどが地球研の機関リポジトリである「地球研アーカイブス」[1]に蓄積されているものの、この他にも地球環境学に関連する研究資源が国内外各地の研究機関に所蔵されている。そこで、地球研を含む30余りの国公立大学の研究センターなどが参加し、それらが所蔵する地球環境学に関連した研究資源の相互利用を可能にするために始められたのが「大学間連携を通じた広域アジアにおける地球環境学リポジトリの構築」事業である[2]。

この事業では、各機関の研究資源に関するメタデータを集約した「地球環境学リポジトリ」が研究資源の相互利用を実現するための中心的な役割を果たすものとして構築される予定である。研究資源に関するリポジトリとしては、300以上の機関で機関リポジトリが運用されている[3]。これらの機関リポジトリが機関ごとの枠組みで研究資源を提供しているのに対し、この「地球環境学リポジトリ」は、各機関の機関リポジトリやデータベースに含まれる研究資源を地球環境学という視点で結びつけ、横断的な検索機能

を提供しようとしている。言うなれば、地球環境学という一学問分野のための仮想的なリポジトリを構築しようという構想である。

2. 異分野連携に向けて

2.1 モノとコト

地球環境学リポジトリを構築する上で問題となるのが、多様な研究資源をどのようにして活用するのかである。たとえ横断的な検索により研究資源を見つけられたとしても、専門性が増すほどにその研究資源が何を示しているか、何に使えるのかが門外漢には分かりにくい。その手がかりとして、一般的には、十進分類などの規則に従って研究資源を分類したり、キーワードを付与したりすることが行われる。これにより、その資源が「何」について書かれているのかを知ることは可能である。しかしながら、研究資源を取捨選択する際に本当に知りたいのは、その「何」が「どうしたのか」である。

例えば、人口に関する論文について考えてみる。この場合、論文の主題はもちろん人口であり、「人口」というキーワードが付与されていると想像される。だが、この論文を引用などで利用する際には、人口が増えたのか、減ったのかといった「人口」が「どうしたのか」が関心事である。つまり、分類やキーワードから得られる「何」だけでなく、「何」が「どうしたのか」に関する情報が存在することで、より適切な研究資源を迅速に見つけることが可能になる。特に、利用者にとって専門外の研究資源を見つける際には、

^{†1} 総合地球環境学研究所
Research Institute for Humanity and Nature
^{†2} 京都大学地域研究統合情報センター
Center for Integrated Area Studies, Kyoto University

その数居を下げることに役立つと予想される。

そこで、地球環境学リポジトリでは、研究資源に「何」が「どうしたのか」といった情報を付与し、研究資源の探索やそこから得られる知識を活用するための仕組みとして盛り込むこととした。なお、地球環境学リポジトリでは、「何」の部分は「モノ」と称している。これは従来の keyword に相当する。一方「何」の部分も含めて「何がどうしたのか」は「コト」と称しており、keyword に対して、key phenomenon もしくは key event ともいべきものと位置付けられる。

2.2 コトの詳細



図 1 「コト」とそれに関するモノ、コト、研究資源の関係。橙色の楕円はコト、緑色の四角はモノ、ピンク色の四角は研究資源を示す。以降の図も同様。

Figure 1 “Key phenomenon” and links to related keyword, key phenomenon and resource. Orange ellipses, green and pink squares show key phenomena, keywords and resources, respectively. These are the same in the following figures.

「二酸化炭素の増加」というコトを中心にモノ、コト、研究資源の関係を考えてみる(図 1)。「二酸化炭素の増加」というコトに関連するモノとして、二酸化炭素や大気が考えられる。また、「二酸化炭素の増加」に関連する研究資源として、その事実を裏付ける観測データやそれについて言及した論文、報告書などがあげられる。研究資源側から見れば、これらの研究資源には、キーワードとしての「二酸化炭素」に加えて、「二酸化炭素の増加」というコトの情報が付与されることになる。さらに、「二酸化炭素の増加」というコトに関連するコトも考えられる。一つは「二酸化炭素の増加」を引き起こした原因となるコトであり、「化石燃料の使用」や「火山の噴火」などがある。一方、「二酸化炭素の増加」が引き起こす結果となるコトとして「温暖化」などがあげられる。

こうしたコト同士のつながりは、因果関係による連鎖として捉えることができる(図 2)。さらに、コト同士をつなぐ因果関係についても、それを裏付けるデータや論文など

があるはずであるから、因果関係を示すパスに対して研究資源を紐付することもできる。

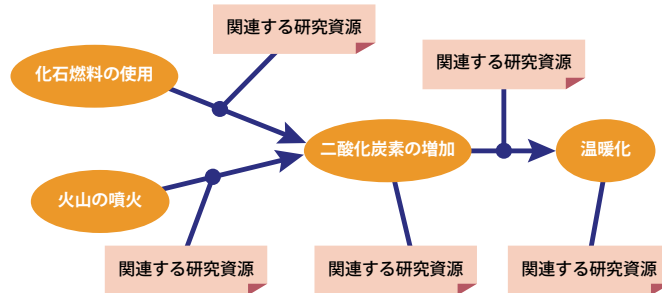


図 2 コトの連鎖による因果関係の表現。
 Figure 2 Expression of causal relationship by “key phenomenon” chain.

結果として、研究資源には、keyword に相当するモノだけでなく、研究資源で扱われているコトに関するメタデータが付与され、さらに、研究資源において扱われている場合には、コトとコトの因果関係に関するメタデータも付与されることになる。

2.3 コト連鎖の応用

コトの連鎖により表現される因果関係を使って、着目するコトが及ぼす結果、もしくは原因を追いながら、それぞれに紐付されている研究資源を探索することが可能になる。

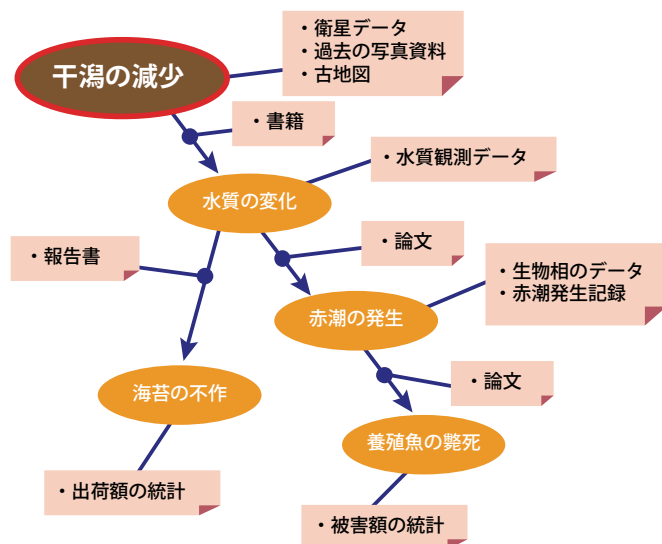


図 3 コト連鎖により特定の原因からその結果を辿る例。
 Figure 3 Retrieval of effects from a certain cause through “key phenomenon” chain.

図 3 は、「干潟の減少」というコトを起点にして、これ

によって起こりうるコト辿りながら研究資源を探索する例を想定したものである。まず、「干潟の減少」に直接紐付いている衛星データなどの研究資源が見つかる。次に、コトの連鎖から「干潟の減少」が「水質の変化」を引き起こすことが分かり、この因果関係について書かれた書籍が見つかる。同様に、いくつかの過程を経ながら、「干潟の減少」が「養殖魚の斃死」「海苔の不作」などを引き起こすことが分かり、それぞれのコトやコト同士の因果関係を裏付ける統計資料、データ、論文などの研究資源を見つけることができる。

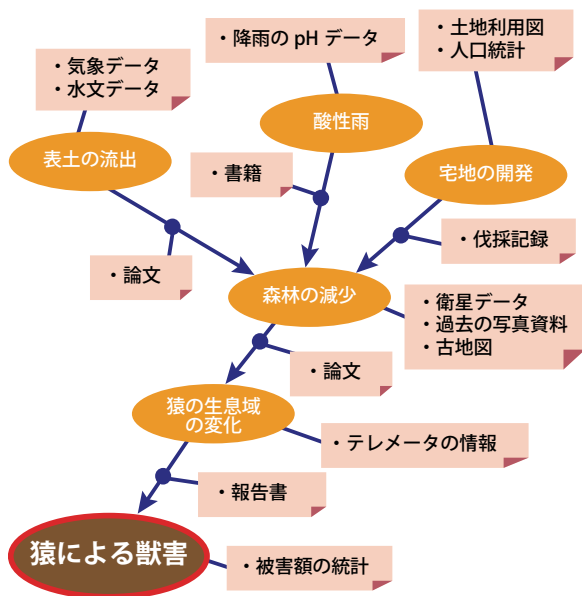


図 4 コト連鎖により特定の結果からその原因を辿る例。
 Figure 4 Retrieval of causes from a certain result through “key phenomenon” chain.

図 4 は図 3 とは逆方向にコトの連鎖を辿る例である。最初に「猿による獣害」という最終的な結果が与えられ、コトの連鎖により原因となるコトを次々と遡って辿りながらそれらに紐付いている研究資源を探索する。この結果、「猿による獣害」の遠因として、「表土の流出」「酸性雨」「宅地の開発」などのコト、それを裏付けるデータや統計資料などを見つけることができる。

このように、研究資源を探索するための仕組みとして検討されている研究資源へのコトの付与とその利用であるが、これを発展させて、因果関係そのものを探索するための仕組みとしての役割も期待できる。因果関係の連鎖として有名な慣用句に「風が吹けば桶屋が儲かる」がある。これに対して、「風が吹いて儲かったのは桶屋だけか?」、また、「桶屋が儲かる原因は風が吹くことだけか?」といった問いが可能になる。また、「風が吹く」コトと「桶屋が儲かる」コトの間に今まで気付かなかった複数の経路を見出すこともできるかもしれないし、因果関係がループしていれば、

フィードバックが起きていることが示唆される。さらに、矛盾する経路や十分な研究資源が関連付けられていない経路などは、新たな研究材料のヒントになり得る。コトに注目することで、研究資源の探索にとどまらず、動的なストーリーとして研究対象を理解したり、その中で欠落しているデータを発見したりするといった、新たな「知」を発見するためのプラットフォームとしても利用できる可能性がある。

3. セマンティック Web 技術の応用

3.1 データ構造

地球環境学リポジトリでは、モノ・コト及び研究資源のメタデータを表現する方法として RDF を採用した。その理由として、Linked Data の仕組みにより容易に他の情報資源と結びつけられること、広く利用されている形式であり相互運用性が確保できることなどがあげられる[4]。さらに、コト連鎖のような連鎖の数があらかじめ決まっておらず、かつ無制限につながってゆくような関係を RDF では表現しやすく、それを検索するための SPARQL などの仕組みも整っていることがあげられる[5]。

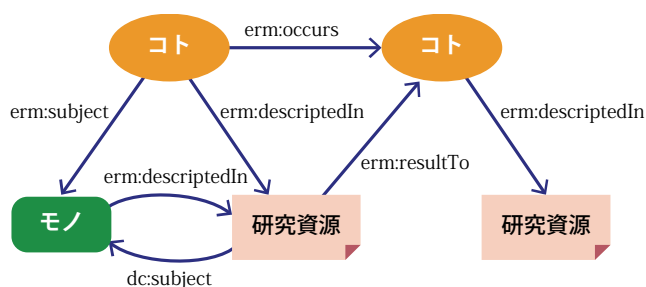


図 5 コトの連鎖による因果関係の表現。プレフィックス erm 及び dc は、それぞれ地球環境学リポジトリメタデータおよびダブリンコア。

Figure 5 Expression of causal relationship by “key phenomenon” chain. Prefixes erm and dc are metadata for repository for humanity and nature and dublin core, respectively.

モノ・コトを表現するために、地球環境学リポジトリメタデータとしていくつかの語彙が定義された(図 5)。コトからコトへの直接の因果関係は occurs で表し、主語となるコトが原因、目的語となるコトが結果となる。コトそのものに関する研究資源の存在は describedIn で表す。これに対し、因果関係に関する研究資源の存在は、因果関係(コトの間のリンク)そのものに直接研究資源を紐付することができないため、2組のトリプルにより間接的に表現される。1つは前述の原因となるコトから研究資源への describedIn もう一つは研究資源から結果となるコトへの

resultTo である。また、コトに関連するモノが存在する場合は、コトからモノへの subject で表す。以下は、RDF によるコトの記述例である (レコードサンプル 1)。

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>.
@prefix dcq: <http://purl.org/dc/terms/>.
@prefix dct: <http://purl.org/dc/dcmitype/>.
@prefix geo: <http://www.w3.org/2003/01/geo/wgs84_pos#>.
@prefix schema: <http://schema.org/>.
@prefix erm: <http://rihnexers.chikyu.ac.jp/erm/elements/1.0/>.
```

```
<http://rihnexers.chikyu.ac.jp/id/phenomenon/0000001234>
  rdf:type dct:Event ;
  rdfs:label "温暖化";
  dcq:spatial "シベリア";
  erm:describedIn <http://rihnexers.chikyu.ac.jp/id/6>;
  schema:temporal "1980-04-01, 2004-04-01";
  geo:polygon "19. 418, 100. 795 19. 418, 100. 795/15. 938, 104. 299
  15. 938 104. 299";
  erm:occurs <http://rihnexers.chikyu.ac.jp/id/phenomenon/0000
  001235>.
```

レコードサンプル 1 コトの記述例

(RDF/Turtle による記述) .

Record Sample 1 Example for description of "key phenomenon" (description by RDF/Turtle).

コト「温暖化」が URI <http://rihnexers.chikyu.ac.jp/id/6> で表される研究資源で扱われていることが erm:describedIn により記述されている。さらに、コト「温暖化」が erm:occurs により URI <http://rihnexers.chikyu.ac.jp/id/phenomenon/0000001235> で表される別のコトを引き起こすことが記述されている。この他に、ここで取り上げている温暖化の時間空間範囲が dterms (dcq), Geo Cocabulary (geo), Schema.org (schema) などの語彙により記述されている。

一方、研究資源はレコードサンプル 2 のような記述になる。基本的には、ダブリンコア等による資料に関する一般的なメタデータであるが、URI <http://rihnexers.chikyu.ac.jp/id/phenomenon/0000001235> で表されるコトへの因果関係についてこの研究資源の中で扱っていることが erm:resultTo により記述されている。

このレコードサンプル 2 の研究資源は、レコードサンプル 1 のコトに関する記述の中で erm:describedIn により紐付けられている研究資源と同じであり、かつ、レコードサンプル 2 の研究資源において erm:resultTo により因果関係の結果として示しているコト <http://rihnexers.chikyu.ac.jp/id/phenomenon/0000001235> とレコードサンプル 1 のコトにおいて erm:occurs により因果関係の結果として示しているコトが一致している(図 6 上)。このことから、レコードサンプル 2 の研究資源は、両者のコトの間の因果関係に関する内容を扱っていることが示されている(図 6 下)。

```
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>.
@prefix dcq: <http://purl.org/dc/terms/>.
@prefix dc: <http://purl.org/dc/elements/1.1/>.
@prefix owl: <http://www.w3.org/2002/07/owl#>.
@prefix geo: <http://www.w3.org/2003/01/geo/wgs84_pos#>.
@prefix schema: <http://schema.org/>.
@prefix erm: <http://rihnexers.chikyu.ac.jp/erm/elements/1.0/>.
```

```
<http://rihnexers.chikyu.ac.jp/id/6>
  dc:title "シベリアの温暖化";
  dc:source "地球研ニュース 00";
  dc:creator "地球太郎/Sample 著者1/サンプル著者2";
  dcq:hasFormat <http://archives-contents.chikyu.ac.jp/1472/
  NL40web.pdf>;
  dc:type "論文";
  schema:temporal "2004-04-01, 2004-04-02";
  dcq:spatial "シベリア";
  geo:polygon "19. 418, 100. 795 19. 418, 100. 795/15. 938, 104. 299
  15. 938 104. 299";
  dc:subject [
    rdfs:label "湖沼";
    dcq:subject "<http://id.ndl.go.jp/auth/ndlsh/湖沼>";
    dcq:subject "<http://jp.dbpedia.org/resource/湖沼>";
    owl:sameAs <http://rihnexers.chikyu.ac.jp/id/thing/0000
    001234>
  ];
  erm:resultTo <http://rihnexers.chikyu.ac.jp/id/phenomenon/
  0000001235>.
```

レコードサンプル 2 研究資源の記述例

(RDF/Turtle による記述) .

Record Sample 2 Example for description of resource (description by RDF/Turtle).

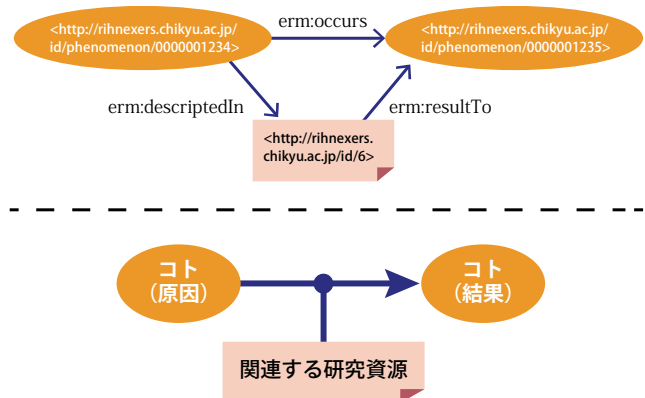


図 6 レコードサンプル 1, 2 により記述される因果関係と研究資源。RDF による表現 (上) と模式図 (下) .

Figure 6 Causal relationship and related resource described in record samples 1 and 2. Expression by RDF (upper panel) and Schematic diagram (lower panel).

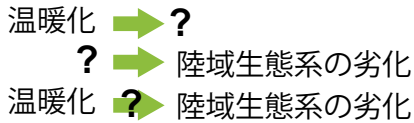
3.2 プロトタイプ

地球環境学リポジトリの想定している機能の実現可能性について、プロトタイプを構築して検証した。プロトタイプには、著者やタイトルなどにより研究資源を探索するため一般的な機能が実装されるとともに、研究資源に紐付けられたモノやコトを検索する機能、そして、コト連鎖により因果関係の経路を検索する機能が実装された。

コト連鎖による経路の検索では、経路の始点、終点のいずれか一方もしくは両方を指定することにより行う(図 7)。SPARQL を用いた検索により、当該経路の中での各ステップの情報が経路番号、その経路の始点および終点となるコトなどととも出力される。これを表示用に整理し、コト連鎖による経路の検索結果として出力する。

検索結果として表示された経路上のコトまたはコトとコトの間の矢印をクリックすると、紐付けられている研究資源がそれぞれ表示される(図 8)。この研究資源をクリックすることで、研究資源に関する詳細情報をグラフ表示 または文字による詳細表示として得ることができる。これらの詳細情報に付されたリンクをたどって、実際の研究資源を取得することも可能である。

検索条件の入力



経路の検索結果

start	end	path id	steps	route	label	relation
http://rihnexters.chikyuaac.jp/id/phenomenon/0000000004	http://rihnexters.chikyuaac.jp/id/phenomenon/0000000011	85	0	yuac.jp/id/phenomenon/0000000004	温暖化	
http://rihnexters.chikyuaac.jp/id/phenomenon/0000000004	http://rihnexters.chikyuaac.jp/id/phenomenon/0000000011	85	1	yuac.jp/erm/element/10/occurs	湿潤化	http://rihnexters.chikyuaac.jp/erm/element/10/occurs
http://rihnexters.chikyuaac.jp/id/phenomenon/0000000004	http://rihnexters.chikyuaac.jp/id/phenomenon/0000000011	85	2	yuac.jp/erm/element/10/occurs	永久凍土の劣化	http://rihnexters.chikyuaac.jp/erm/element/10/occurs
http://rihnexters.chikyuaac.jp/id/phenomenon/0000000004	http://rihnexters.chikyuaac.jp/id/phenomenon/0000000011	85	3	yuac.jp/erm/element/10/occurs	陸域生態系の劣化	http://rihnexters.chikyuaac.jp/erm/element/10/occurs

経路の可視化

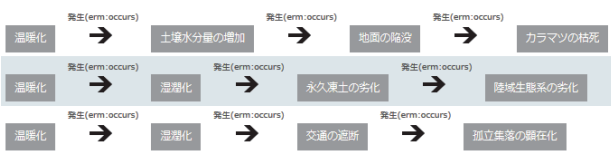


図 7 コトの連鎖を使った因果関係の経路検索。

Figure 7 Route search of causal relationship using “key phenomenon” chain.

コトや因果関係に関連する研究資源の表示



研究資源の詳細

研究資源の取得

図 8 経路検索結果からのコトや因果関係に関連する研究資源の取得。

Figure 8 Resources related with a “key phenomenon” and causal relationship between them in a route search result.

プロトタイプでは、これらの検索機能の他に、データの入力・編集機能、複数の機関による運用を前提とした個別の RDF ストアを運用する機能、さらに地球環境学リポジトリの機能を外部から利用するための API などが試験的に実装された。

4. 課題と将来展望

地球環境学リポジトリのプロトタイプでは、基礎的な技術検証ではあるものの、当初想定していた機能が実現可能であることが明らかになった。特に、コト連鎖経路検索においては、考え方としては従来の因果ネットワーク[6]などに近いものではあるものの、セマンティック Web の技術を応用したことや、研究資源と結びつけることなどの点で、従来とは異なる実装を実現できた。しかしながら、今後実際に地球環境学リポジトリを構築・運用する上では、多くの課題もある。中でも、コトを含む研究資源のメタデータをどのように構築するかが大きな問題である。この点で、地球環境学リポジトリの構築を推進するため、本事業に参加している大学等の研究機関を中心とした予備研究が実施されており、これらの実施過程や成果がメタデータの構築

のヒントになり得る。検討の結果、(1) データが示す変化を直接読み解きながらそれらが示すコトを抽出するボトムアップの方法、(2) はじめにその分野や研究対象で一般的に知られている現象(=コト)を描き、それぞれ過程に研究資源を張り付けてゆくトップダウンの方法など、いくつかアプローチが見出された。いずれにせよ、現段階では試行錯誤を繰り返しながら個別に対応するしかないが、将来的には自然言語処理などによるコト抽出の補助などが期待される。

このほかにも、分野間で異なる語彙の統制や同じコトであってもデータなどにより立証されている現象に対して仮説や推測などをどのように系統的に扱うかなどといった技術的な課題が存在する。これらの技術的な課題は、予備研究の成果や地球研の研究プロジェクトの成果を地球環境学リポジトリに登録しながら検討を進めてゆくことになる。

地球環境学リポジトリは、環境研究への利用を目指して構築されるものであるが、その仕組みは「知識」という共通の基盤で自然系と人文系のデータをつなぐハブともなり得る。強いて言えば、何が起きたのかというデータの解釈の結果としての自然科学や社会科学の知識と、何をどう捉えるのかという論や説などに基づく人文科学の知識とをつなぐ学際的な共通基盤として、その仕組みを応用することも考えられる。

謝辞 本事業に参画いただいている各大学の関係者、地球研の事務担当者、ならびにシステム開発にご協力いただいたインフォコム社の皆様に謹んで感謝の意を表す。

参考文献

- 1) 関野 樹, 安富 奈津子: 異分野混在の研究資源をいかに残すか? - 「地球研アーカイブス」の試み - . 研究報告人文科学とコンピュータ (CH) 2014-CH-101(6):1-6 (2014).
- 2) 地球環境学リポジトリ事業, <http://www.chikyu.ac.jp/repository/>
- 3) 学術機関リポジトリ構築連携支援事業: 国内の機関リポジトリ一覧, <http://www.nii.ac.jp/irp/list/>
- 4) Dean Allemang, James Hendler (萩野達也 訳): 実践セマンティック Web, ジャストシステム (2010).
- 5) Eric Prud'hommeaux and Andy Seaborne: SPARQL Query Language for RDF. W3C Recommendation (2008). <http://www.w3.org/TR/2008/REC-rdf-sparql-query-20080115/>
- 6) 西田豊明: 定性推論の諸相, 朝倉 AI らいぶらり, 朝倉書店 (1993).