

LinkData.org を使った RDF 教育とデータ公開化運動の推進

Promotion and education of open data by creation and hosting RDF on LinkData.org repository

下山 紗代子*¹
Sayoko Shimoyama西方 公郎*¹
Koro Nishikata吉田 有子*¹
Yuko Yoshida豊田 哲郎*¹
Tetsuro Toyoda*¹ 独立行政法人 理化学研究所 生命情報基盤研究部門
Bioinformatics And Systems Engineering division, RIKEN

LinkData aims to promote open data by providing a simple way for anyone to create Linked Open Data and publish it. LinkData's unique approach is easy to use for both novice and expert. It provides a simple field based template creation system combined with hosting that differs from any other Converter to RDF. We aim to generate new knowledge with this service by creating a linked repository for any type of web data ranging from human related information to scientific biological data, as well as by bringing diverse data creators and application developers together in collaboration. LinkData provides data as a free service using Linked Open Data (LOD) open standards for RDF document works. Here we show some evidence that LinkData is efficient in promoting opening data through educating those who without knowledge of RDF.

1. はじめに

Resource Description Framework (RDF) はグラフ形式でデータを表現するデータモデルであり、ウェブ上でデータを共有するための標準形式として広がりつつある。しかし、RDF の認知度は HTML に比べて低く、ウェブの専門家を除いてあまり知られていない。幅広い分野で活躍する研究者や多くの人々が RDF でデータを公開していけるようにするためには、RDF に変換するツールを単に提供するだけでなく、彼らが RDF とは何かを分かりやすく学ぶことができ、他者の作成した RDF のデータ構造を参考にしつつ、自らのデータに最適な RDF モデルを自らデザインできるようにするための教育の場をウェブ上で提供する必要がある。

そこで我々は、誰でもデータを RDF で公開できるプラットフォームとして、LinkData.org (<http://linkdata.org>) を構築し、テーブル形式と対比させながら RDF を理解させることで、RDF によるデータ公開化運動の推進を試み、その有用性を検証した。

2. RDF 教育とデータ公開化運動の推進

2.1 RDF によるデータ公開の課題

データを公開するにあたり、データとデータの間を表現する“スキーマ”を定義することは不可欠である。しかし、一般の人はスキーマを定義することに慣れていないことが、データ公開における障壁となっていた。

RDF の利点は、データ間の関係を明示的に記述することが可能であり、機械的な解釈が容易であることが挙げられる。しかし、その一方で RDF は人による可読性が低く、扱うには専門的な知識が必要であるという欠点がある(表 1)。

また、データ公開に際しては RDF を提供するウェブサーバの構築が必要であるが、技術面・コスト面の問題があり個人に求めるのは困難である。

2.2 LinkData の概要

上記の課題を解決してデータ公開化運動を推進するために

データモデル	RDF モデル	テーブルモデル
メリット	<ul style="list-style-type: none"> データ間の関係をリンクで明確に表現できる ウェブとの相性が良い 	<ul style="list-style-type: none"> 人が直感的に理解しやすい 行単位でテキスト処理できて簡単に扱える
デメリット	<ul style="list-style-type: none"> 人がデータ構造の全体像を把握しにくい 扱うには専門知識が必要 	<ul style="list-style-type: none"> データ間の関係を表すための意味論が不明確になりやすい 機械的に解釈させにくい
モデルの模式図		

表 1. RDF モデルとテーブルモデルの比較

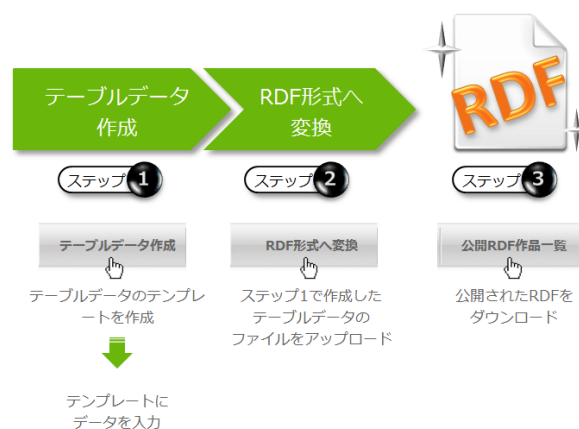


図 1. LinkData におけるデータ公開の手順

連絡先: 豊田 哲郎, 理化学研究所 生命情報基盤研究部門,
〒230-0045 神奈川県横浜市鶴見区末広町 1-7-22,
toyoda@base.riken.jp

LinkData では一般の人が慣れているテーブル形式でスキーマを定義できる支援機能を提供した。さらに、他者が既に定義したスキーマを参考にして再利用できる機能を提供することで、誰でも簡単に自分のデータを RDF として公開できるようにした(図 1)。

2.3 LinkData が提供する支援機能

(1) スキーマ定義の支援機能

LinkData ではユーザが簡単にスキーマを定義できるようにするための支援機能として、Excel 形式のテンプレートを作成してダウンロードできる GUI を提供している。ユーザはこの GUI から自分のデータに関する必要最小限のメタデータを入力すると、RDF プロパティをカラム名とするテーブル形式の Excel ファイルが生成され、ダウンロードすることができる。このファイルをテンプレートにして自分のデータを流し込み、テーブルデータとして完成させることで、ユーザは RDF 作成の基礎となる3つ組データ(サブジェクト・プロパティ・オブジェクト)の関係をテーブル形式で分かりやすく定義することができる(図 2)。

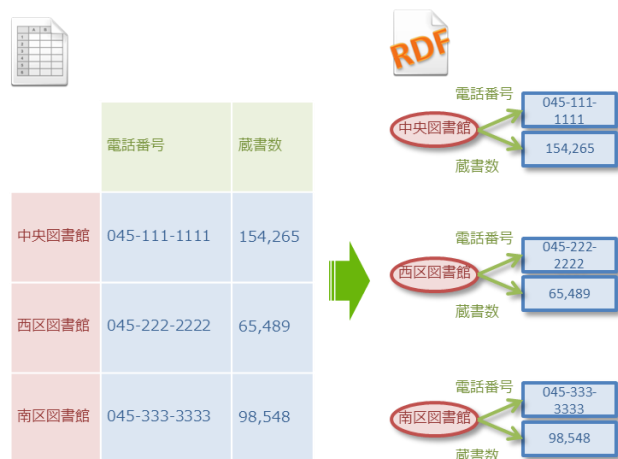


図2. テーブル形式から RDF への変換例

(2) 標準プロパティのサジェスト機能

LinkData では、ユーザが行うテンプレート作成において、プロパティを正しく選択できるように支援するために、世界中で汎用的に利用されている Dublin Core, RDFS, FOAF 等のプロパティを候補として表示する機能を提供している。ユーザはこのリストからプロパティを選択することができるだけでなく、任意の文字列でプロパティを指定することも可能である。ユーザが入力したプロパティが URL として不完全な場合は、後述の URI 自動生成機能で補完する。

(3) URI の自動生成機能

データをウェブ上のリソースとして表現するためには、リソースを URI で表す必要があるが、すべてのリソースにあらかじめ URI が割り当てられているとは限らない。そこで、LinkData のテーブルデータでは、サブジェクトやプロパティをリテラルで指定することを許容し、リテラルで指定されたサブジェクトやプロパティを RDF 変換時に LinkData ドメインを用いた独自 URI に自動変換することで、URI に関する知識のないユーザでも容易にデータ公開できるようにした。リテラルから URI への変換方法を以下に示す。

```
Subject : http://linkdata.org/resource/[作品 ID]#[リテラル]
Property : http://linkdata.org/property/[作品 ID]#[リテラル]
```

(4) テーブル形式と RDF 形式の対比機能

LinkData では、ユーザが作成したテーブルデータがどのように RDF に変換されるかをチュートリアルで分かりやすく解説するとともに、実際に変換された RDF をグラフ構造で可視化できるようにすることでテーブルデータとの対比を容易にした。また、他者が既に LinkData から公開したデータのスキーマを再利用して自らのデータ構造を定義できるようにした。これにより、一般のユーザが RDF とは何かを容易に理解できるようにした。

(5) RDF 形式への変換・公開機能

LinkData では、アップロードされたテーブル形式のデータを RDF 形式に変換してそのまま公開できる機能を提供した。LinkData ではユーザがデータ公開を積極的に行うためのインセンティブを与える目的で「RDF 作品」という概念を用いて、あるひとまとまりのデータセットを1つの創作物品として公開する場を提供する。ユーザは上記で作成したテーブルデータをアップロードし、登録先となる RDF 作品を作成する。アップロードされたテーブルデータは RDF に変換され、LinkData のウェブサイト上に RDF 作品として公開される。RDF 作品のページでは、誰でもダウンロード可能な形で元のテーブルデータ、テーブルデータのテンプレート、及び RDF が提供される。

(6) データのつながりの可視化機能

LinkData では、ユーザが他者の URI を積極的に活用して繋がり合う linked data を作成するためのインセンティブとして、RDF 作品間のつながりを可視化する機能を提供している。RDF 作品をノードとして、公開された RDF 作品で使われている URI のドメインを解析し、関連の強いドメインとつなげることで LinkData と外部ドメインとの LOD クラウドを自動構築する。

(7) アプリケーション開発者支援機能

LinkData のデータを利用するアプリケーション開発者のために外部からのデータ取得を容易にする API を提供している。アプリケーション開発者は LinkData に公開されている RDF 作品の各ファイルについて、TSV, RDF, RDF/JSON, RSS でデータを取得し、外部のアプリケーションにおいて利用することができる。

(8) 国際化にむけた多言語化対応のための支援機能

LinkData では、様々な国の人から協力を得て、ウェブサイトの多言語化対応を進めている。このために、ローカルオーガナイザ制度を導入し、各言語を担当する外部の有志を募ってウェブサイトの言語翻訳を担当してもらうとともに、不適切なコンテンツのチェックなどを言語ごとに監督してもらうようにした。このための支援機能として、サイトの翻訳を行うための GUI をローカルオーガナイザだけに提供している。

2.4 データ公開化運動による本手法の有効性の検証

データ公開化運動の流れは世界的にも広がりを見せており、Open Data のコンテストが世界各地で開催されている。日本でも本格的なデータ公開化運動として Linked Open Data Challenge Japan 2011 のコンテストが初開催されたため、我々の LinkData を出展することで、本手法の有効性の検証を試みた。

LinkData を体験するハンズオンセミナーを同コンテストのイベントとして開催したところ、RDF 作成の経験がない約 40 名の参加者にその場で RDF 作成の体験をさせることに成功した。また、セミナー終了後も多くの RDF 作品が LinkData に登録された。こうした活動を通して同コンテストの審査員からは、「LOD 作成におけるハードルを下げる効果がある」、「LOD の普及に貢献するツールである」、「データの公開と利用の促進に有用なプラ

ットフォームである」といった高い評価を受け、LinkData は同コンテストのアプリケーション部門最優秀賞を受賞した。

また、LinkData は実際の教育現場においても用いられた。大学の講義において、テーブル形式から変換されたデータが、RDF としてどう表現されているのかを学生に理解させるために、テーブル形式と RDF 形式の対比機能が利用された。

また、LinkData は理化学研究所が開発・運営を行っている統合データベースシステム SciNetS (Scientists' Networking System; <http://scinets.org>) [Masuya 2011] へのデータ登録の窓口としても機能している。例えば、理化学研究所免疫・アレルギー科学総合研究センターのデータベース RAPID (Resource of Asian Primary Immunodeficiency Diseases; <http://rapid.rcai.riken.jp/RAPID>) [Keerthikumar 2009] のデータが RDF 作成の経験のない研究者から RDF 作品として LinkData に投稿された。このデータを SciNetS のキュレータが受け取って、アノテーションの付加、オントロジーによる体系化を SciNetS 上で行うことで OWL 形式に変換し、BioLOD (Biological Linked Open Data database; <http://biolod.org/>)からダウンロードすることや、さらに BioSPARQL (<http://biosparql.org/>)による SPARQL 検索が可能となった。

LinkData は 2011 年 11 月に公開され、ユニークユーザ数は 1 か月あたりおよそ 1,340 人であった。2012 年 3 月の国別訪問数の合計は 1,964 であり、うち海外からのアクセスは 21% を占めていた。2012 年 4 月 13 日時点における LinkData に登録された RDF 作品数は 121 作品であり、そのうち公開されているものが 52 作品、限定公開が 13 作品、非公開が 56 作品であった。

2.5 LinkData の有用性に関する考察

Linked Open Data において、Tim Berners Lee は 5 つの星を用いた評価体系を提唱した (<http://www.w3.org/DesignIssues/LinkedData.html>)。この 5 つ星を用いて評価すると、LinkData に最初にユーザがアップロードするテーブル形式のデータは 2-Star に分類される。このデータが LinkData により変換され RDF で提供されることで 4-Star もしくは 5-Star にレベルアップされる。したがって LinkData は W3C の基準において有用性がある。

また、LinkData は web アプリケーションとして使いやすいインターフェースを持つ点が評価され、大学教育の場で活用されるとともに、データの公開と利用を促進するプラットフォームとしての有用性が認められた。

さらに、理研内においては、LinkData を用いることで、RDF の知識のない科学者からの複雑な科学データの受け渡しを容易に行うプラットフォームとしての有用性が認められた。

2012 年 4 月 13 日現在、LinkData に登録されている RDF 作品の総数は 121 作品であり、地域、地理、観光、工業、グルメ、アニメ、ゲーム、ライフサイエンス等非常に多様な分野のデータが含まれている。このことから、LinkData は分野を越えてデータ公開の場を提供する効果がある。また、LinkData では公開されたデータを取得する API を提供していることから、データ公開者とアプリケーション開発者をつなぐコミュニケーションサイトとして有用である。さらに、多言語化をユーザに行わせる仕組みによって、多数の言語による利用を可能としたため、データ公開化運動を様々な国に波及させていける可能性がある。

3. おわりに

LinkData はデータの公開において最も障害となるスキーマ定義から公開までの作業工程を専門知識のない人でも分かりや

すくできるように工夫することで、RDF 教育やデータ公開化運動への効果が認められた。

今後、さらにデータ公開を促進するためには、データを公開する上での強いインセンティブを与える仕組みが必要である。その1つとして、データ価値の評価体系の整備が挙げられる。データの価値を定量的に評価することができれば、ユーザが信頼できるデータを選ぶ際の基準を与えることができる。また、データ提供者に対し、より品質・信頼性の高いデータを提供させるためのインセンティブを与えることができるだろう。

参考文献

- [Masuya 2011] Masuya H, Makita Y, Kobayashi N, et al.: The RIKEN integrated database of mammals, *Nucleic Acids Res.*, Vol. 39, D861-870, (2011)
- [Keerthikumar 2009] Keerthikumar S, Raju R, Kandasamy K, et al.: RAPID: Resource of Asian Primary Immunodeficiency Diseases, *Nucleic Acids Res.*, Vol. 37, D863-867, (2009)