

Linked Data における日の取り扱い—時間に基づくデータ連携

関野 樹
総合地球環境学研究所

Linked Data において、日はリテラル値として記述される。しかしながら、これが実体としての日を示しているのか、その実体を指し示すラベル（日付）であるのか混乱した状態にある。本研究ではこの関係を整理して、実体としての日を「日の URI」として示す方法を検討し、これを通じて、日を基軸とした情報連携や各暦法に基づく日付などの情報を適切に扱う仕組みを構築した。

Entities of date in the Linked Data – Data linkage based on temporal information

Tatsuki Sekino
Research Institute for Humanity and Nature

Date is described as a literal value in the Linked Data. However, it is unclear whether the value indicates an entity of the date or a label of the entity. Relationship between the entity of date and its label are reconsidered in this study, and are described using a date URI. These results apply to develop new mechanism which realizes resource retrieval based on date and appropriate linkage between date and related information.

1. はじめに

Linked Data が広がりを見せている。時間は、空間と同様、様々な情報に関係する要素であり、Linked Data においても情報同士のリンクの接点となり得る。しかしながら、Linked Data や RDF が使われる中では、時間を使った情報検索や解析のため環境が十分に整っていない。そこで本研究では、時間が持つ性質を踏まえながら Linked Data で時間を適切に取り扱うための課題の抽出と整理、および解決に向けた実装実験を行った。なお、本研究では時間のなかでも日に関する課題を中心とし、時刻や時差については将来の課題とする。このため、地域や時代にかかわらず、協定世界時 (UTC) (UTC 制定前にあってはグリニッジ標準時 (GMT)) での正子 (午前 0 時) を日のはじまりとし、これを代表値として用いる。

2. Linked Data と時間

(1) 時間の性質

まず、Linked Data で情報を扱う上で考慮すべき時間の性質を整理する。

1) さまざまな情報に時間が付随する

上述のとおり、時間はさまざまな情報に付随する要素であり、時間を基軸としてこれらの情報を連携させ、検索や解析に用いることが可能である。たとえば、時系列に沿って情報を整理し、関係する文書、写真、データなどを連携させることができる。さまざまなデータを連携させることを目指

す Linked Data において、時間はその目標を達成するための重要な要素である。

2) さまざまな形で時間が表現される

時間軸上の点または範囲を指し示すためにさまざまな表現が用いられる。日という時間範囲については、日付がそれを指し示す代表的な表現であり、“2014 年 12 月 13 日”という日付が特定の時間軸上の時間範囲 (日) を指し示している。日付は何らかの暦法に従っており、上述の例はグレゴリオ暦の日付である。他にも和暦による“平成 26 年 12 月 13 日”やタイ仏暦による“13 ธันวาคม 2557”も指し示す時間範囲は同じである。また、昨日、今日、明日といった相対的な表現や“じんもんこん 2014 の 1 日目”といった表現も同じ時間範囲を指し示している。

3) 時間に関連する情報がある

時間軸上の特定の点や範囲が何らかの属性を持つことがある。典型的なものとして曜日があり、日という時間範囲に対して曜日を一意に決めることができる。他にも、日の干支、新月や春分などの天体現象などは、暦法にかかわらず特定の日に関連付く属性と考えることができる。

(2) Linked Data における時間の取り扱い

Linked Data では、情報同士の関係を記述するのに RDF が用いられる。この中で、日は ISO 8601 に則って表記されたりリテラル値として扱われており [1, 2], 2014 年 12 月 13 日であれば、“2014-12-13”¹と表記される。しかし

ながら、上述の時間の性質を踏まえながらその活用を考えると、いくつかの課題が浮上してくる。

1) 日を RDF の主語にできない

RDF の記述規則では、RDF の主語は URI (厳密には IRI) もしくは空白ノードである[1]。したがって、リテラル値として表現される日を主語とすることができない。一方で、上述のとおり、日には暦法の違いなどにより同じ時間範囲を指し示す複数の表記(日付)が存在する。これらの表記はリテラルとせざるを得ないので、こちらも RDF の主語にはできない。したがって、"2014-12-13"^^xsd:date が"平成 26 年 12 月 13 日"^^xsd:string と同じ時間範囲を指し示すことを RDF で直接関連付けて記述することができない。また、時間が曜日などの属性を持ちうることも上述のとおりであるが、"土曜日"^^xsd:string を RDF で直接関連付けることも同様に不可能である。現行の方法で日にかかる様々な情報を関連付けようとすれば、空白ノードを介して記述するしかない(図 1)。

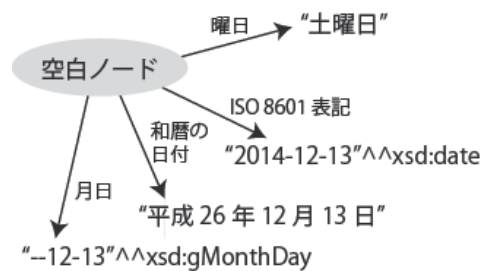


図 1 空白ノードによって記述された 2014 年 12 月 13 日にかかるグラフ。

Figure 1 Graph about the date of December 13, 2014 described using a blank node.

2) 日に関する情報を集約できない

人物を例にすれば、特定の人物をエンティティとして捉え、それを示す URI を起点にして、その人物に関する RDF を得る仕組みが一般的である。これによって、さまざまな情報にリンク可能な Linked Data が実現されている。しかしながら、日はリテラル値として記述されているので、関連する情報の起点となることができない。たしかに、SPARQL を用いれば、目的語として同じ日付を持つトリプルを抽出するという操作は可能ではある。だが、この方法でも異なるグラフ、異なるトリプルストアにまたがる操作は容易でないし、Linked Data ブラウザを使ってリンクを次々と辿るような操作もできない。結局現状では、ある出来事がいつ起きたのかを知ることは容易だが、特定の日に何が起きたのかを知ることは難しい。時間を基軸にして様々な情報をリンクでき

るにもかかわらず、現状の Linked Data では、時間はリンクの袋小路になっている。

3) ISO 8601 の運用にかかる課題

RDF での日の記述方法として用いられる ISO 8601 であるが、これに則って記述された日付が必ずしも同じ日を指すとは限らない。これは、日付の表記に用いられている暦法(いわゆる西暦)が過去にユリウス暦からグレゴリオ暦に改暦されたことに起因する。つまり、この改暦の際に、ユリウス暦が想定する 1 年(365.25 日)と実際の太陽の動き(1 平均太陽年: 365.24218951 日[3])の違いにより蓄積されたズレを修正するため、10 日以上の日付を読み飛ばす操作が行われること、そして、この改暦が国によって異なる時期に行われていることが問題である[4]。たとえば、イタリアは 1582 年に改暦し、イギリスは 1752 年に改暦している。このため、両者がそれぞれ異なる暦法を用いている期間は、史資料に記載されている日付がたとえ同じであっても、10~11 日も異なる日を指している可能性がある。

ISO 8601 の中でもこの問題について触れられている[5]。ISO 8601 では、暦日を特定するためにグレゴリオ暦を用いること、また、グレゴリオ暦の基準点はメートル条約がパリで署名された 1875 年 5 月 20 日であることが示されている。その上で、データ交換の当事者同士の合意があれば、グレゴリオ暦導入前にも遡ってグレゴリオ暦を適用することを要求している(先発グレゴリオ暦: proleptic Gregorian calendar)。しかしながら、先発グレゴリオ暦を実際のデータや史資料に適用する例はあまりなく、計算機上で扱われる情報においても、それぞれの国の暦に応じた暦日を用いられている。まして、Linked Data のような仕組みでは ISO 8601 が求めるような「当事者同士の合意」が成立している保証もない。

(3) 日と日付の関係

日の取り扱いに関するこれらの問題は、実体としての「日」とそれを指し示す「日付」とが混同されて用いられていること、さらには、これらを分けて記述するための仕組みが Linked Data や RDF が使われる環境で十分整っていないことが原因である。つまるところ、ISO 8601 は本来、あくまでも日付の書式を規定するものであって、日そのもの、つまり日の実体を示しているわけではないのである。

したがって、上述の問題を解決するには、実体としての日と日付の関係を整理し、その上で実体としての日に対して、日付、曜日、各暦法での表記といった情報を関連付ける仕組みが必要である。

この実体としての日と日付の関係を整理したものが図2である。まず時間軸上の一定の範囲を実体としての日として捉える。そうすると、日付はその範囲を識別するために一定に規則に従って付けられた名称もしくはラベルのようなものとして位置づけられる。この規則に相当するのが暦法であり、ユリウス暦、グレゴリオ暦をはじめ、和暦やヒジュラ暦（イスラム暦）などがある。さらに、実体としての日に曜日などのさまざまな情報が関連付けられる。

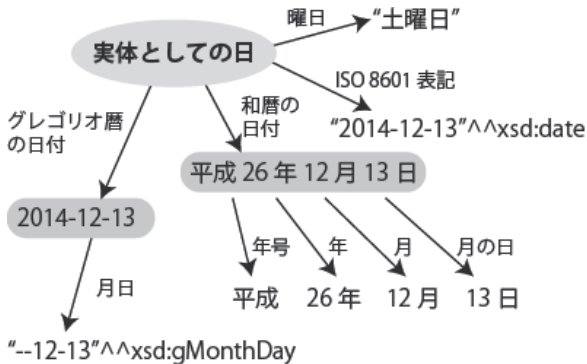


図2 実体としての日と関連情報のグラフ。
Figure 2 A graph about an entity of date with related information.

このような捉え方は RDF を使って物事を記述する一般的な方法と同じである。たとえば、人物に関する RDF データを構築する場合、人物の実体を表す URI を設定し、それに名前 (foaf:name) やニックネームなどの別名 (foaf:nick)、性別 (foaf:gender) などの属性が関連付けられる。同様に考えれば、実体としての日についても URI (以下、「日の URI」) をまず設定し、ISO 8601 による表記や他の暦による表記はそれに関連付けられた情報ということになる。結局、図1の空白ノードが日の URI に置き換わった形になる。

一方で、各暦に基づく日付は、日の URI を指し示す情報として暦法ごとに作られる。表記だけであれば、「2014-12-13」や「平成 26 年 12 月 13 日」のようなりテラル値で問題ないが、日付を構成する要素の年号、年、月、月の日がそれぞれ別のリソースにリンクされる可能性がある。また、解析や検索の中で特定の暦法に基づく特定の日付を直接参照することも考えられる。したがって、各暦法にも日 (暦日) の実体があるものとし、それ指し示す日付 ("平成 26 年 12 月 13 日" など)、や年号、年、月、月の日などの情報を関連付けられる URI (以下、「暦日の URI」) を用いた方が扱いやすい。なお、時刻や時差を考慮しなければ特定の日についての日の URI と暦日の URI は同じ時間範囲を示しているの、owl:sameAs の関

係にある。

3. 日を取り扱うための環境

(1) 日の URI

日の URI では実体としての日を一意に特定する必要があるが、ユリウス/グレゴリオ暦に基づく日付に問題があることは上述のとおりである。これに代わって、日を一意に特定する識別子としてユリウス通日が広く用いられている[6]。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、不連続が無いことから、日食の予測といった天文学での計算にも使用されている[7]。小数部分は時刻を表現し、0.5 が午前 0 時、0.75 で午前 6 時、0.0 (整数部分は+1) で午後 0 時となり、一般的には協定世界時 (UTC) もしくはグリニッジ標準時 (GTM) を指しているものとして解釈される。ただ、本稿では日に議論を絞っていることから、地域や時代にかかわらず、正子 (午前 0 時) のユリウス通日をその日の代表値として用いる。図3はこのユリウス通日を用いた日の URI の記述例で、実体としての日を識別する。

<http://datetime.hutime.org/date/2457004.5>
ユリウス通日

図3 日の URI の例。"2014-12-13"^^xsd:date が指し示す実体としての日を表す。

Figure 3 An example of the date URI indicating an entity of date
"2014-12-13"^^xsd:date.

(2) 暦日の URI

暦日の URI では、まず、暦法自体の URI を設定する (図4上)。calendar/の後の"0001"は暦法の ID で、実験段階である本研究では、暫定的に通し番号としている。この暦法の URI に、その暦法が適用可能な時間範囲、利用されている地域などの情報を関連付けることが可能になる。次に、この暦法での日 (暦日) の URI を、ユリウス通日を用いて表現する (図4下)。

<http://datetime.hutime.org/calendar/0001>
暦法 ID
<http://datetime.hutime.org/calendar/0001/date/2457004.5>
暦法 ID ユリウス通日

図4 暦法の URI (上) と暦日の URI (下) の例。"2014-12-13"^^xsd:date が指し示す実体としての日を表す。

Figure 4 An example of the calendar URI (upper) and the calendar date URI (lower) indicating an entity of date
"2014-12-13"^^xsd:date.

この暦日の URI は上述の日の URI と同様に実体としての日をしめしており、図 4 の例は図 3 の例と同じ時間範囲を示している。

(3) 日の URI にかかる情報

図 5 は、日の URI にかかるトリプルの例である。日の URI の type については、今後、実体としての日のクラスを定義していく必要はあるが、ここではいったん owl:Thing とした。次に、この日を ISO 8601 の形式で表記した場合の値を hutime:iso8601 により関連付けた。先述のとおり、Linked Data の既存のリソースでは日は ISO 8601 によるリテラル値として表記されている。したがって、日の URI と ISO 8601 に則った表

記との関係を記述しておくことで、既存のリソースが日の URI、さらには各暦日の URI と連携するための糸口を与えることになる。暦法については、ISO 8601 の要求通り、グレゴリオ暦（先発グレゴリオ暦を含む）で記述する。

暦法に関わらず日が決まれば機械的に決まる属性は、日の URI に直接関連付けることができる。例として、曜日を hutime:dayOfWeek により DBpedia の該当する曜日を示す URI と直接関連付けた。

各暦法の日付との関係は、図 2 のグラフをそのまま表現する形で、日の URI と暦日の URI とを owl:sameAs で結びつけることで記述した。

```
@prefix hutime: <http://datetime.hutime.org/ontology#> .
<http://datetime.hutime.org/date/2457004.5> a owl:Thing ;
  hutime:iso8601 "2014-12013"^^xsd:date ;
  hutime:dayOfWeek <http://dbpedia.org/resource/Saturday>;
  owl:sameAs <http://datetime.hutime.org/calendar/0001/date/2457004.5> ;
  owl:sameAs <http://datetime.hutime.org/calendar/1001/date/2457004.5> .
```

図 5 日の URI にかかるトリプル (RDF/turtle による表記).
Figure 5 Triples about a date URI (described by RDF/Turtle).

(4) 暦日の URI にかかる情報

図 6 はユリウス／グレゴリオ暦で暦日の URI にかかるトリプルの例である。type は日の URI と同様に owl:Thing とした。

label は各暦での標準的な表記を示す。ユリウス／グレゴリオ暦では ISO 8601 形式の表記を標準的な表記として記述した。なお、label は表記の例であるので、型は xsd:date ではなく xsd:string である。次に、日の URI との関係をおowl:sameAs で記述した。

その次は、日付の構成要素を別々に示したもので、hutime:gYear および hutime:gMonthDay はそれぞれ年および月日 (月と月の日) である。両者とも ISO 8601:2000 [6] の truncated representation に則った表記である (最新の ISO 8601:2004 [5]には規定されていない)。このような、年のみ、月日などで整理された情報が DBpedia などで作成・公開されており、他の情報との連携という点で有用である。他の情報との連携の例として DBpedia へのリンクを次に記述した。

```
@prefix hutime: <http://datetime.hutime.org/ontology#> .
<http://datetime.hutime.org/calendar/10001/date/2457004.5> a owl:Thing ;
  rdfs:label "2014-12-13" ;
  owl:sameAs <http://datetime.hutime.org/date/2457004.5> ;
  hutime:gYear "2014"^^xsd:gYear ;
  hutime:gMonthDay "--12-13"^^xsd:gMonthDay ;
  dc:relation http://dbpedia.org/resource/2014 ;
  dc:relation http://dbpedia.org/resource/December_13 .
```

図 7 は和暦での暦日の URI にかかるトリプルの例である。type および日の URI との関係の記述は、ユリウス／グレゴリオ暦の場合と同じであるが、label は標準的な表記として、日本語による日付の表記を記述している。

日付の構成要素である年号、年、月、月の日は、それぞれ、hutime:era、hutime:year、hutime:month、hutime:dayOfMonth により記述した。明治 5 年以前の改暦前の和暦では閏月が出現すること、また、グレゴリオ暦の年や月日と期間が一致しないことから xsd:gYear や xsd:gMonthDay といった型の値は意味をなさない。このため、日付の構成要素については、図 6 とは異なる記述となっている。一方で、外部の情報にリンク可能なものがあれば、それを関連付けることは可能である。暦日"平成 26 年 12 月 13 日"については、グレゴリオ暦の年および月日とその期間が一致するため、DBpedia へのリンクを記述している。無論、明治の改暦前の和暦の暦日には、これらのリンクを記述することはできない。

図 6 暦日の URI (ユリウス／グレゴリオ暦) にかかるトリプル (RDF/turtle による表記).
Figure 6 Triples about a calendar date URI (Julian / Gregorian Calendar) (described by RDF/Turtle).

```
@prefix hutime: <http://datetime.hutime.org/ontology#>.
<http://datetime.hutime.org/calendar/10001/date/2457004.5> a owl:Thing ;
  rdfs:label "平成 26 年 12 月 13 日";
  owl:sameAs <http://datetime.hutime.org/date/2457004.5>;
  hutime:era "平成";
  hutime:year "26 年";
  hutime:month "12 月";
  hutime:dayOfMonth "13 日";
  dc:relation http://dbpedia.org/resource/2014;
  dc:relation http://dbpedia.org/resource/December_13.
```

図7 暦日の URI (和暦) にかかるトリプル (RDF turtle による表記).
Figure 7 Triples about a calendar date URI (Japanese Calendar) (described by RDF/Turtle).

4. 実装と検証

(1) 実装

日の URI および暦日の URI と関連するトリプルを実際に生成する環境を構築した. なお各トリプルは, トリプルストアにあらかじめ蓄積されたものではなく, クライアントの要求に応じてその都度生成される. また, データは, HTTP ステータスコード 303 (リダイレクト) を使った仕組みにより, http リクエストヘッダ (Accept ヘッダ) の記述内容に従って html または RDF/XML で返される.

(2) 検証

XHTML+RDFa で記述されている時間情報システム HuTime の日本語 Web サイト (<http://www.hutime.jp>) に, 日の URL および暦日の URL の導入を試みた (図8). さらに, そこから Linked Data 用のブラウザ等[8][9][10]を使って RDF を抽出し, 構築した環境を用いて日の URI および暦日の URI から DBpedia などの外部のリソースにリンクを辿れるかを検証した (図9). この結果, HuTime の Web サイトがソフトウェアやサービスを公開した年や月日に起きた出来事の一覧などを, リンクを辿って得ることが可能であることを確認した.

```
<p>
  <a href="/basicdata/calendar/form.html">
    暦変換機能 (Web フォーム) を供用開始
    <span class="dateOfUpdate" rel="dc:date" resource
      ="http://datetime.hutime.org/date/2456397.5">
      2013-04-15 更新
    </span>
  </a>
</p>
```

図8 XHTML+RDFa により導入された日の URI の例.

Figure 8 An example of date URI introduced according to XHTML+RDFa.

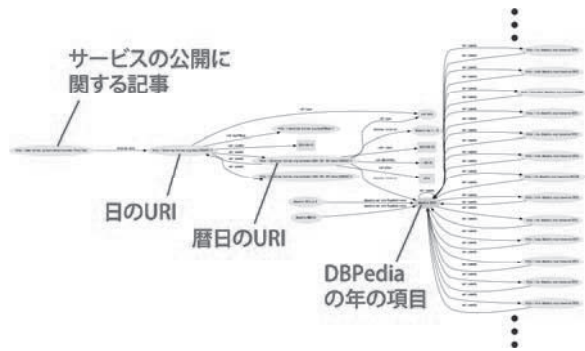


図9 Web サイト <http://www.hutime.jp> に導入された日の URI と暦日の URI による日を起点とした外部のリソースへのリンク. (Linked Data Browser [10] により出力されたグラフの一部)
Figure 9 Links to external resources starting from a date by the date and the calendar date URIs introduced into a web site <http://www.hutime.jp/>. (A part of a graph drawn by Linked Data Browser [10])

5. まとめと課題

本研究では, (1) Linked Data において実体としての日と日付とが混同されている現状が整理され, (2) 日の URI と暦日の URI を通じて, 実体としての日と日付の関係を RDF により記述する仕組みが構築された, さらに (3) その仕組みを介して, 日を基軸としたデータ連携や日に関するデータを適切に扱うことが可能であることが示された.

これにより, 日などの時間に基づいた検索や解析がより推進されることが期待される. さらに, 日の URI や暦日の URI について, SPARQL endpoint が構築されれば, 外部のリソースとも連携しながら「金曜日起きた出来事」, 「和暦での閏月に生まれた人物」といった複雑な検索が可能になる.

また, 本研究で構築された環境は, 時間に基づいたデータ連携という当初の目的だけでなく, 暦の変換機能やデータを Linked Data として提供するという別の側面もある. これにより, 異なる暦法で記述された史資料を適切に関連付けなが

ら利用することが実現される。

一方で残された課題や新たに浮上した課題もある。1つは URI にユリウス通日を利用したものの、ユリウス通日が分からないと日や暦日の URI を記述できないという問題である。この点は、日付を表す文字列を解釈する機能[11]を応用し、任意の表記の日付文字列からユリウス通日を使った URI にリダイレクトされる仕組みを構築すれば、解決できる。

また、本研究では実体としての日や暦日の type をいったん owl:Thing とした。他にも、曜日や暦日の構成要素（年号、年、月、月の日）に対して、それぞれ外部のリソースを関連付けたりリテラル値で記述したりするなどしている。しかし、和暦の月であれば、12 月に対して、師走、極月、臘月などの別名があること[12]、その年の 12 月は大小どちらであるか、前後の月が何月であるか（閏月があるので一定ではない）といった、情報を関連付けることも考えられ、こういった情報をどのように集約し提供するかといった課題もある。これらを勘案しながら、実体としての日や暦日をどのように扱うかを改めて検討する必要がある。

最後に、時刻や時差に関する問題は残されたままである。本研究では単純に日の始まりを正子としたが、日の区切りは時代や地域によって異なるし、当然、史資料もそれに基づいて記述されている。時刻まで扱おうとすれば、こういった問題まで解決しないと、実際の研究で活用できる実用的なものにはならないであろう。

今後、利用事例を積み重ねながら、本格的な運用に向けたこれらの課題の検討・解決が必要である。

参考文献

- 1) W3C: RDF 1.1 Concepts and Abstract Syntax, <<http://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/>> (参照 2014-10-06) .
- 2) W3C: XML Schema Part 2: Datatypes Second Edition, <<http://www.w3.org/TR/xmlschema-2/>> (参照 2014-09-11) .
- 3) 天文年鑑編集委員会（編）. 天文年鑑 2014 年版, 誠文堂新光社 (2013).
- 4) Dershowitz, N. and Reingold, E.M.: C alendrical Calculations, Cambridge University Press (2007).
- 5) ISO: ISO 8601:2004 Data elements and interchange formats – Information interchange – Representation of dates and times, Third edition, (2004).
- 6) ISO: ISO 8601:2000 Data elements and interchange formats – Information interchange – Representation of dates and times, Second edition, (2000).

7) NASA: NASA Eclipse Web Site, <<http://eclipse.gsfc.nasa.gov/eclipse.html>> (参照 2014-10-07)

8) W3C: RDFa 1.1 Distiller and Parser, <<http://www.w3.org/2012/pyRdfa/>> (参照 2014-09-11)

9) OpenLink Software: The OpenLink Data Explorer Extension, <<http://ode.openlinksw.com/>> (参照 2014-09-11)

10) Kanzaki, M.: Linked Data Browser, <<http://www.kanzaki.com/works/2014/pub/ld-browser>> (参照 2014-11-06)

11) 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.145-152 (2013).

12) 吉川弘文館編集部（編）: 日本史必携, 吉川弘文館 (2006).