

RDF版「人名一覧」の構築

— LinkedDataによる人文学のための基盤データ整備 —

大内英範（人間文化研究機構本部・総合情報発信センター）

後藤真（国立歴史民俗博物館）

鈴木卓治（国立歴史民俗博物館／総合研究大学院大学）

関野樹（国際日本文化研究センター）

「統合検索システム nihuINT」を運用する人間文化研究機構（NIHU）は、既存データベースのRDF化を進めている。本論文では、NIHUが2014年に構築した、主として芳賀矢一編『日本人名辞典』から抽出した人物等を対象とする「人名一覧表示システム」のRDF版について述べる。前システムと比べてより柔軟に、多くのリソースへ連携でき、人文科学の研究情報のハブとなることが期待される。また、日本の歴史人名をRDFとして記述するための標準的な語彙がまだないため、これを整備することも目指す。

Listing System of pre-modern Personal Names by RDF

- Construction of data infrastructure for humanities by LinkedData -

OUCHI Hidenori (National Institutes for the Humanities)

GOTO Makoto (National Museum of Japanese History)

SUZUKI Takuzi (National Museum of Japanese History / SOKENDAI)

SEKINO Tatsuki (International Research Center for Japanese Studies)

National Institutes for the Humanities (NIHU), which operates the "integrated search system nihuINT", is promoting RDF of existing databases. This paper describes the RDF version of "Listing System of pre-modern Personal Names". It is expected to be more flexible than previous system, to link with many resources, and to be a hub of research information on humanities. Also, there is no standard vocabulary for describing Japanese historical names as RDF, so we also aim to prepare that.

1. はじめに

人間文化研究機構（以下「機構」）は、人文系の6つの研究機関（それぞれが大学共同利用機関法人）および本部からなる大学共同利用機関法人であり、人文系としては唯一の研究機構である。

機構では、2005年度より「研究資源共有化事業」を推進し、機構内の各機関が作成した人間文化に関する多くのデータベースを一元的に検索・活用するための研究および環境整備を行ってきた。機構内機関および地域研究拠点のみならず、国立国会図書館と京都大学東南アジア地域研究研究所のデータベースも検索対象とした横断検索システムである「統合検索システム nihuINT (nihu INTegrated retrieval system) 」 (<https://int.nihu.jp>)（以下「nihuINT」）は、現在174データベース、500万レコード以上の検索が可能である。同事業は2016年度より「高度連携事業」に衣替えし、事業内容の高次化を進めている。その主要なところは既存データベースの

RDF化によって、人文学研究のための基盤データを整備することである。本論文ではその1つである「人名一覧」のRDF化について述べる。

人文科学研究において求められる基盤データとしては、まず人名・書名・地名であると考えている。これらはさまざまなリソースへの入り口になるものであり、これらを起点として情報を辿ることができるような仕組みが求められる。

また、日本の歴史人名をRDFとして記述するための標準的な語彙がまだないため、これを整備することも目標として重要な点である。

2. 旧「人名一覧」について

機構では2014年、nihuINT上に「人名一覧表示システム」（以下「旧「人名一覧」」）を公開した。旧「人名一覧」は、『古事類苑』および芳賀矢一編『日本人名辞典』から抽出した近代以前の人物等の一覧中に、『日本人名辞典』に掲載されている説明文を「解説」として提示したものであ

った。また、旧「人名一覧」では、表示された姓名等をクリックすることで、それをキーワードとして nihuINT で直接検索する機能を備えていた。

つまり旧「人名一覧」は、単なる解説付きの人名の一覧なのではなく、人名をキーワードとしてさまざまな人文系研究資源にアクセスできる入り口の提供というところに意義があったわけである。

3. RDF 版「人名一覧」

3.1 『日本人名辞典』の RDF 化

現在、前項に述べた旧「人名一覧」のデータの RDF 化を進めている。旧「人名一覧」は、見出し語としての「人名」、その「よみ」、クリックして nihuINT 検索するための「姓名、姓、名」、「解説」の4つのフィールドから構成される表形式の一覧であった。つまり大雑把に言えば、データは人名と解説の2つに分かれており、そのうち人名で nihuINT を検索できるというものであったということになる。本研究ではその前出『日本人名辞典』による「解説」中の要素も取り出し、nihuINT をはじめとする他のリソースと関連付けていくことが大きな目標である。そこで、「解説」の部分要素を分解し、各要素が見出しの「人名」に対してどのような関係にあるのかを整理し、RDF 化することとした。

3.2 「空海」の例

例えば「クウカイ 空海」について、「解説」すなわち『日本人名辞典』の説明は、以下の通りである。

高僧。佐伯氏。讃岐多度津郡の人。父は田公、母は阿刀氏。入唐して慧果阿闍梨に見え、其の教を受く。弘仁七年高野山に金剛峯寺を建て、承和二年三月二十一日寂す。年六十二。延喜二十一年弘法大師の号を贈らる。性霊集、文鏡秘府論の外伝教に関する、著作甚だ多し。新勅撰、続千載、風雅諸集の作家。

ここに記述されている情報の主だったものを整理するために、まず句点で分解して内容を検討してみる。

高僧。…A

佐伯氏。…B

讃岐多度津郡の人。…C

父は田公、母は阿刀氏。…D

入唐して慧果阿闍梨に見え、其の教を受く。

…E

弘仁七年高野山に金剛峯寺を建て、承和二年三月二十一日寂す。…F

年六十二。…G

延喜二十一年弘法大師の号を贈らる。…H

性霊集、文鏡秘府論の外伝教に関する、著作甚だ多し。…I

新勅撰、続千載、風雅諸集の作家。…J

「解説」には記述のパターンがあり、まず「高僧」「歌人」「国学者」など、その人物をひと言で表す言葉 (A) からはじまり、出自や家族を示す言葉が続く (BCD)。そして主だった事蹟等が述べられるが (EFGH)、没年月日は「寂す」「卒す」「歿す」などの言葉で記述され、死没時齢が示される。また、著作がある場合、勅撰和歌集に和歌が採られている場合には、その作品名・歌集名が記される (IJ)。

なお、『日本人名辞典』では、その人物の和歌が勅撰和歌集に採られている場合に、その歌集名を記載している。『平家物語』『忠度都落』のエピソードを想起するまでもなく、歌人としての評価としては、勅撰歌人かどうかということが大変重いものであったためである。

上記を見出し「クウカイ 空海」とあわせて、ひとまず整理すると、下記のようになる。

人名：空海

よみ：クウカイ

父：田公

母：阿刀氏

別名：弘法大師

出身地：讃岐多度津郡

没年月日：承和二年三月二十一日

死没時齢：六十二

著作：性霊集

著作：文鏡秘府論

勅撰和歌集：新勅撰

勅撰和歌集：続千載

勅撰和歌集：風雅

「母：阿刀氏」のように個人までは特定できない不完全な情報もあるし、他のキーワードとして「慧果阿闍梨」「高野山」「金剛峯寺」などを採ることも可能である。また、死没時齢「年六十二」を没年月日から生年を導くことも可能であろう。

3.3 「北条政子」の例

もう1例、「マサコ 政子」の「解説」を検

討してみる。

源頼朝の室。北条時政の長女。頼家、実朝及び二女を生む。頼朝の薨後、自ら政を視る。従二位に進み、嘉禄元年薨ず。年六十九。尼將軍と称せらる。

やはり句点で分解してみると下記のような。

源頼朝の室. …K

北条時政の長女. …L

頼家、実朝及び二女を生む. …M

頼朝の薨後、自ら政を視る. …N

従二位に進み、嘉禄元年薨ず. …O

年六十九. …P

尼將軍と称せらる. …Q

すなわち、家族関係を表す言葉 (KLM) , 事蹟・没年・死没時齡 (NOP) , 別名 (Q) となっている。見出しと合わせて、以下のように整理できよう。

人名：政子

よみ：マサコ

夫：源頼朝

父：北条時政

子：頼家

子：実朝

子：二女

没年：嘉禄元年

死没時齡：六十九

別名：尼將軍

位階である「従二位」もキーワードとして採ることが可能である。なお、著作や勅撰和歌集への入集などについての記載はない。

なお、これらの要素への分解であるが、解説の記述にある程度の規則性があるため、簡単なスクリプトによって、大まかなところまでは自動処理が可能である。もちろん最終的には手作業が必要である。

3.4 項目の整理

上記「空海」「政子」について、必要に応じてデータをきれいにしつつ、項目を揃えてみる (表 1)。省略された姓を補ったり、出身地は旧国名や旧郡名などで分割した。これは後述する地名での再検索のためである。また父・母は

「親」とし、同様に夫も「夫／妻」とした。また、「弘法大師」は空海の諡号、「尼將軍」は世の人が政子を称して言った呼び方であり、かなり位相が異なると考えられるが、そのあたりの違いは判断しないこととし、同じ「別名」とした。また、やはり後述する他の DB での再検索に関わる項目をひとまず立てることとし、「死没時齡」はここでは項目として立てていない。

表 1：項目で整理

人名	空海	北条政子
よみ	くうかい	ほうじょう まさこ
親	佐伯田公	北条時政
親	阿刀氏	
夫／妻		源頼朝
子		源頼家
子		源実朝
子		二女
別名	弘法大師	尼將軍
出身地	讃岐	
出身地	多度津	
没年月日	承和二年三月二十一日	嘉禄元年
著作	性靈集	
著作	文鏡秘府論	
勅撰和歌集	新勅撰和歌集	
勅撰和歌集	続千載和歌集	
勅撰和歌集	風雅和歌集	

「人名」としての「空海」に対する上記各要素の関係をひとまずグラフにして例示したものが図 1、「北条政子」について示したのが図 2 である。なお、実際に RDF で記述する際には、Subject は人物 ID を用いた URI とし、「空海」「北条政子」という人名そのものは、その Object となる。それについては後述する。

3.5 検索可能な RDF データベースへ

前節までに整理した項目・データとモデルを実際に RDF として定義し、nihuINT-LD に格納、検索可能な状態にする。

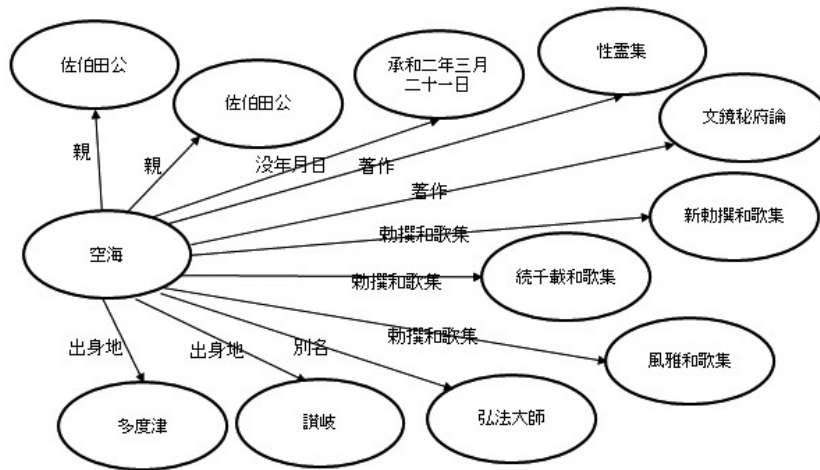


図 1:「空海」のグラフ

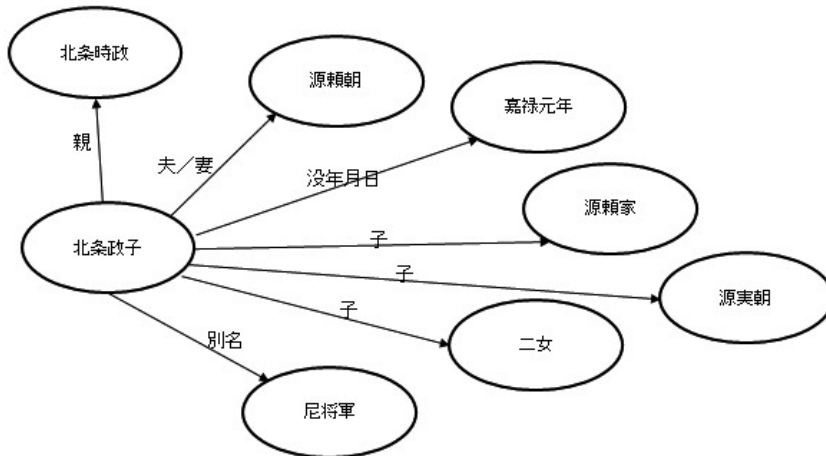


図 2:「北条政子」のグラフ

まずこれまで見た項目に対応した語彙「jinmeilist」を定義する. jinmeilist はクラスとして Jinmei を持つものとし, その下に以下にあげるプロパティを持ち, 各項目に対応するものとする. なお, 「jinmeiID」は, その人名を表現する URI に使用するために任意で付すユニークな数字, 「commentary」は, 解説の文章に対応するものである. なお, 語彙 jinmeilist については次節でも述べるが, 今後も検討を重ね, 修正および拡張する予定である.

表 2: Jinmei のプロパティ

プロパティ	対応する項目
jinmeiID	人物 ID
personalName	人名
personalNameYomi	人名よみ
parent	親
spouse	夫/妻
children	子
alias	別名
birthPlace	出身地
deathDate	没年月日
work	著作
chokusenshu	勅撰和歌集
commentary	説明

これを「空海」に適用してみると、表3のようになる。

表3: 「空海」に適用した例

プロパティ	対応するデータ
jinmeiID	515
personalName	空海
personalNameYomi	くうかい
parent	佐伯田公
parent	阿刀氏
alias	弘法大師
birthplace	讃岐
birthPlace	阿度津
deathDate	承和二年三月二十一日
work	性霊集
work	文鏡秘府論
chokusenshu	新勅撰和歌集
chokusenshu	続千載和歌集
chokusenshu	風雅和歌集
commentary	高僧. 佐伯氏. 讃岐多度津郡の人. 父は田公, … (略)

これを改めてグラフ表示したものが図3である。

以上のような定義に基づき、RDF ファイルを作成、nihuINT-LD に格納した。検索結果を表示したものが図4である。※開発途中のため仮の

項目ラベル。

3.6 語彙「jinmeilist」

前節までに RDF 化した人名データを nihuINT-LD で検索できるようにしたのであるが、そこで用いた語彙についてももう少し説明しておく。

RDF で人名・人物を示す場合の語彙については foaf, schema.org, dbpedia 等、既にさまざまあり、本研究の対象である古典人名についてもそれらの組み合わせによってある程度記述可能である。ただし先に見た勅撰和歌集への入集についての情報など、特有の情報もあり、既存の語彙だけでは記述することができない。

现阶段の jinmeilist はひとまず shema.org (の Person) になるべく対応させながら、chokusenshu などのプロパティを追加する形としたが、なお検討を要する点もある。たとえば work で示される「著作」も、chokusenshu で示される入集した「勅撰和歌集」も、作品のタイトルという意味では同じであるので、そこはきちんと示す必要がある。また、本論文では対象外とした所属組織や官位・官職なども現状ある程度関連付けを行なっているが、人名一覧には貴族・僧・刀工などさまざまな職種・階層の人物がおり、それらの整理はこれからである。

そうしたいくつかの課題をクリアして、より

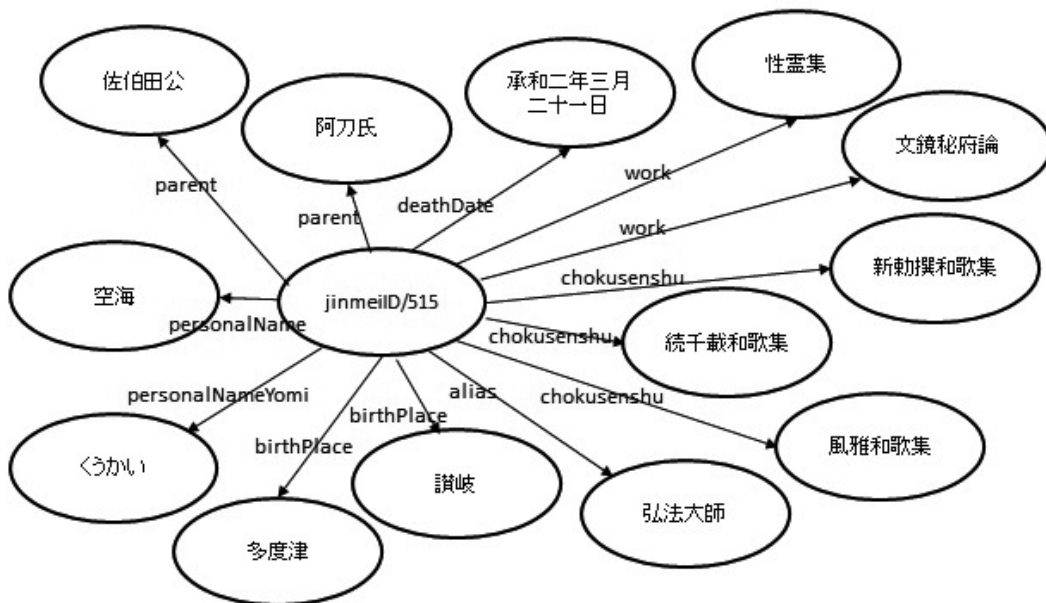


図3: 語彙 jinmeilist を適用した「空海」のグラフ

使いやすい日本の歴史人名に関する語彙定義を作成したいと考えており、今後も検討を重ねて修正・拡張する予定である。

4. RDF 版「人名一覧」の展開

前項に述べたような、RDF 版「人名一覧」はどのように利活用できるだろうか。

旧「人名一覧」では、一覧表から人物の姓名等をクリックすることで nihuINT を検索することができた。RDF 版では人名だけではなく、項目に分解した解説中の語句を用いてさまざまな DB を検索することが可能である。「空海」の例であれば、著作である「性霊集」で nihuINT を検索し、検索結果をデータベースごとにジャンプできる機能を使用して、「国文学論文目録データベース」で研究成果を調べたり、「日本古典籍総合目録データベース」で書誌情報や所蔵情報を調べることが容易である。(なお「日本古典籍総合目録データベース」については一部を試験的に RDF 化しており、そちらを検索することも可能)。出身地の地名で「地名辞書」、親や子の人名で再び人名一覧を検索することも可能である。また、dbpedia 等の外部リソースへの接続も含め、RDF 版「人名一覧」を入口として、よりさまざまな情報へのアクセスが容易になるだろう。旧「人名一覧」よりも、さらに接続先が広がり、人名を起点として、よりさまざまなリソースを辿ってゆくことが可能になる。

(図 5)

5. おわりに

機構は、人文系唯一の研究機構として、これまでも nihuINT 等により人文科学研究のための基盤データ整備を推進してきた。冒頭に述べたように、人文系の研究に特に重要な基盤データとして、人名・地名・書名についての整備が必要であると考えている。それらは人文科学研究において、調査の起点となる基礎データだからである。これまでの事業を継続しながらも、より便利で使い勝手の良い情報基盤を目指してゆく。

人名については本論文で述べた RDF 版「人名一覧」を、データ整備とメタデータ語彙整備の両面で急ぐ。書名については、すでに機構には「日本古典籍総合目録データベース」(国文学研究資料館)があり、地名についても機構では 2017 年度末に「地名辞書」を公開した。今後これらの RDF 化も見据えたより柔軟な連携を進め、さらには dbpedia 等外部リソースとの接続など、ユー

ザがより多くの情報に接することができるよう、整備を進めていきたいと考えている。

参考文献

- [1] 清野陽一, 山田太造, 高田智和, 古瀬蔵: 人文科学データベースからの人名一覧表示システムの構築, 研究報告人文科学とコンピュータ (CH) 2014-CH-103(4), pp.1-6, (2014)
- [2] 芳賀八一: 日本人名辞典, 思文閣出版, (1914, 1972 復刻)
- [3] <http://www.foaf-project.org/>, (参照 2018-11-01)
- [4] <https://schema.org/>, (参照 2018-11-01)
- [5] <https://wiki.dbpedia.org/>, (参照 2018-11-01)



図 4: 検索結果例



図 5: dbpedia に連携した例