

国立歴史民俗博物館日本荘園データベースにおける 時間情報のデータクレンジング (第1報)

—HuTime Web API を用いた西暦-和暦整合性チェック—

鈴木 卓治

国立歴史民俗博物館研究部 〒285-8502 千葉県佐倉市城内町 117

E-mail: suzuki@rekihaku.ac.jp

あらまし 国立歴史民俗博物館が公開する日本荘園データベースにおいて、同じ時間情報を表す和暦情報と西暦情報の間の不整合が見つかった。速やかに時間情報のデータクレンジングを行う必要がある。本稿では、和暦から西暦への“正しい”変換を実現するために検討すべき課題とその解決について、検討を試みる。

キーワード 国立歴史民俗博物館, 日本荘園データベース, 時間情報, データクレンジング, 和暦-西暦変換, HuTimeWeb API, 元号, 南北朝

Data Cleansing of Temporal Information in the National Museum of Japanese History's Japanese Manor Database (First Report)

—Checking Gregorian and Japanese Calendar Consistency
using HuTime Web API—

Takuzi SUZUKI

Research Department, National Museum of Japanese History 177 Jonai-cho, Sakura-shi, Chiba, 285-8502 Japan

E-mail: suzuki@rekihaku.ac.jp

Abstract In the Japanese Manor Database published by the National Museum of Japanese History, inconsistencies were found between Japanese calendar information and Gregorian calendar information, which represent the same time information. Data cleansing of the time information is required immediately. In this paper, we attempt to consider the issues that need to be considered in order to achieve "correct" conversion from the Japanese calendar to the Gregorian calendar and how to solve them.

Keyword National Museum of Japanese History, Japanese manor database, time information, data cleansing, Japanese-Western calendar conversion, HuTime Web API, era names, Northern and Southern Courts.

1. はじめに

1981年に設置、1983年に開館した国立歴史民俗博物館は、その計画段階より、コンピュータを用いた日本歴史学情報のデータベースの構築と公開を主要な任務のひとつとして与えられており[1]、1990年より「データベースれきはく」[2]の公開を開始した。その中に日本荘園データベースならびに荘園関係文献目録データベース（いずれも1993年12月より公開）がある。

日本荘園データベース（レコード数9,116件）は日本中世の社会的・経済的な基盤であった「荘園」について、それぞれの荘園ごとに所在や荘園領主・史料出典などの主要な情報を検索・表示するものである。また、荘園関係文献目録データベース（レコード数7,284

件）は、荘園に関する主要な論文・文献類を収録したものである[3]。

2024年に、大井らにより、日本荘園データベース・荘園関係文献目録データベース収録データのRDF (Resource Description Framework)によるLOD (Linked Open Data)が作成され、高等学校の「日本史探究」の授業への活用が試みられた[4]が、その過程で、日本荘園データベースの項目で、同じ時間情報を表す「初見年和暦」（和暦情報）と「初見年西暦」（西暦情報）とで、情報が整合しないものがある、との報告を受けた。

和暦から西暦への変換は、データの入力時に手作業で行われており、誤りが混入したと思われる。「データベースれきはく」に収録されている他のデータベースでも、オリジナルデータとしての和暦に、検索の便を

図るための西暦情報が付与されているものが少なからずあり、潜在的に同様の問題を含んでいる。

問題解決のため、時間情報に関するデータクレンジング (data cleansing) を速やかに実施する必要がある。本稿では、その予備作業として、日本荘園データベースを例題に、和暦から西暦への“正しい”変換を実現するための課題とその解決方法について、限られた事例とはなるが、検討を試みる。

2. 日本荘園データベースの時間情報

日本荘園データベースの項目一覧 ([3]より抜粋) を表 1 に示す。本稿ではこのうち、「初見年和暦」と「初見年西暦」の整合の問題についてとりあげる。

「初見年和暦」は、その荘園が歴史資料に最初に現れた年を表す(荘園が成立した年ではないことに注意)。一方「初見年西暦」は、時期範囲を指定した検索のために、「初見年和暦」を西暦で表したものである。

全 9,116 件のレコードのうち、「初見年和暦」のデータが空でない(何らかの文字列が入っている)ものは 8,065 件、空のものは 1,051 件であった。

3. HuTime Web API による和暦-西暦変換

3.1. HuTime Web API

HuTime Web API (Calendar Calculation)[5,6]は、2017 年に公開された、暦法の変換や暦法に基づく期間の計算を行う機能を提供する Web API である。暦日だけでなく、年号、暦年、暦月など、暦法に基づいて定められるさまざまな期間(以下、「暦日等」)を対象とした操作が可能である。HuTime Web API の使い方は Web ページ[6]に詳しい。また、改元日等の和暦の細かい扱いについては、Web ページ[7]を参照されたい。

3.2. 元号で指定した年間情報の問い合わせ

HuTime Web API の URL は <http://ap.hutime.org/cal/> である。GET メソッドも POST メソッドも受け付けるが、ここでは POST メソッドを用いる。

元号「慶長(けいちょう)」を与えて、始まりの日と終わりの日を求める HTTP リクエストの JSON データは次のようになる。ここで、「method": "info"」は「情報の取得」を、「ical": "1001.1"」は、「南朝暦(一般的に用いられる和暦)」を、「itype": "era"」は、「～年間(～時代)に関する問い合わせ」を、「oprop": "begin;end"」は、「(その年間)の始まりと終わり」を、それぞれ指示している。また「id": "2025022602"」は、リクエストとレスポンスとの間の対応を明確にするための識別子として機能する。

```
{
  "jsonrpc": "2.0",
  "method": "info",
  "params": {
    "ical": "1001.1",
    "itype": "era",
    "ival": "慶長",
    "oprop": "begin;end",
  },
  "id": "2025022601"
}
```

そのレスポンスは次のようになる。リクエストに対応するレスポンスに同じ「id": 2025022601」が含まれていることに注意されたい。

始まりの日と終わりの日としてそれぞれ、改元の行われた日、次の元号の改元日の前日が返答される。

```
{
  "jsonrpc": "2.0",
  "result": [
    {
      "begin": "慶長 1 年 10 月 27 日",
      "end": "慶長 20 年 7 月 12 日"
    }
  ],
  "id": 2025022601
}
```

3.3. 特定の年を指定しての問い合わせ

つぎに「慶長元年」という年について、同じように始まりと終わりの日を求めてみる。問い合わせのための JSON データは次のようになる。「itype": "year"」で「～年に関する問い合わせ」を指示している。

```
{
  "jsonrpc": "2.0",
  "method": "info",
  "params": {
    "ical": "1001.1",
    "itype": "year",
    "ival": "慶長元年",
    "oprop": "begin;end",
  },
  "id": "2025022602"
}
```

そのレスポンスは次のようになる。年間の場合と異なり、始まりの日と終わりの日としてそれぞれ、改元が行われた年の 1 月 1 日、大晦日が返答される。この仕様は、日本の改元における慣行と合致している。すなわち日本では明治の改元まで、改元が布告された時点で、布告された年の元日に遡って新元号の元年と見なしていた(立年改元) [8]。

```
{
  "jsonrpc": "2.0",
  "result": [
    {
      "begin": "文禄 5 年 1 月 1 日",
      "end": "慶長 1 年 12 月 30 日"
    }
  ],
  "id": 2025022602
}
```

3.4. 和暦から西暦を求める問い合わせ

いま求めた「慶長元年」の始まりの日と終わりの日を、和暦から西暦に変換するには、以下のように問い合わせる。「method: conv」は「情報の変換」を、「ocal: 101.1」は（出力の暦が）「ユリウス／グレゴリオ暦(1582年改暦)」であることを、「oprop: text」「oform: yyyy-MM-dd」によって、出力を「西暦-月-日」の形で出力することを、それぞれ指示している。

```
{
  "jsonrpc": "2.0",
  "method": "conv",
  "params": {
    "ical": "1001.1",
    "itype": "date",
    "ival": "文禄 5 年 1 月 1 日¥r¥n 慶長 1 年 12 月 30 日",
    "ocal": "101.1",
    "otype": "date",
    "oprop": "text",
    "oform": "yyyy-MM-dd"
  },
  "id": "2025022603"
}
```

そのレスポンスは次のようになる。これで、与えられた和暦の年／年間情報から、対応する西暦表記を得る準備ができた。

```
{
  "jsonrpc": "2.0",
  "result": [
    {
      "text": "1596-01-30"
    },
    {
      "text": "1597-02-16"
    }
  ],
  "id": 2025022603
}
```

3.5. 南北朝期の和暦の取り扱いに関する注意

和暦を取り扱う場合は、それが南朝暦によるものか北朝暦によるものかを正しく指示しなければならない。

たとえば南朝の元号「正平（しょうへい）」を北朝暦（ical:1001.2）で問い合わせると以下のように誤った結果を返す。

```
{
  "jsonrpc": "2.0",
  "method": "info",
  "params": {
    "ical": "1001.2",
    "itype": "era",
    "ival": "正平",
    "oprop": "begin;end",
  },
  "id": "2025022604"
}
```

```
{
  "jsonrpc": "2.0",
  "result": [
    {
      "begin": "貞和 1 年 10 月 21 日",
      "end": "貞和 6 年 2 月 26 日"
    }
  ],
  "id": 2025022604
}
```

北朝暦で問い合わせると、以下のように正しい結果が得られる。

```
{
  "jsonrpc": "2.0",
  "method": "info",
  "params": {
    "ical": "1001.1",
    "itype": "era",
    "ival": "正平",
    "oprop": "begin;end",
  },
  "id": "2025022605",
}
```

```
{
  "jsonrpc": "2.0",
  "result": [
    {
      "begin": "正平 1 年 12 月 8 日",
      "end": "正平 25 年 2 月 4 日"
    }
  ],
  "id": 2025022605
}
```

表 2 に、南北朝期の元号で指定される時間範囲の HuTime Web API への問い合わせについて、(a) 南朝／北朝の別を正しく指定した場合、(b) 誤って指定した場合、(c) 南朝・北朝に共通する元号について、南朝／北朝を指定した場合の出力結果を示す。

なお、2013年に公開された Web インターフェイスによる暦変換サービス[9]では、北朝暦を指定できない問題がある(2025年2月現在)ので注意を要する¹。

4. 「初見年和暦」における時間情報の表記とその解釈

日本荘園データベースの「初見年和暦」に現れる時間情報の表記は、その大半が「保元二年」のように年を指定したものか、「永久年間」のように元号で指定する年を指定したものである。これはそのまま HuTime Web API に渡して和暦-西暦変換を実行できる。

表記の中には、以下のように何らかの処理を施す必要があるものが含まれている(8,065件中93件)：

- 「乾元□年」のように年が不詳のもの(これは年間と解釈できるか)。
- 「文永十一年頃」のように曖昧さを含むもの。
- 「(明応八年)」のように括弧書きされているもの(間接的な証拠等から類推したものか)。
- 「嘉永四年か」、「文明十四年か」、「応保元年か」、「承久二年？」のように、推定を含むもの。

これらは以下のように解釈して問い合わせを実施することとする：

- 前処理として、先頭の「(」、末尾の「)」「か」「か」「カ」「カ」「?」を除去する。
- 元号+“○○年”の形のもの「年」とする。
- 元号+“○○年頃”の形のもの「年」とする。
- 元号+“□年”の形のもの「年間」とする。

また、以下の表記(8,065件中96件)については、歴史研究者による判定が必要として、本稿における和暦-西暦変換の対象から除外し、残り7,969件を和暦-西暦変換の対象とする：

- 応永[元号のみの場合年間と解釈してよいか]、
- 嘉応年中[年間と読み替えてよいか]、
- 仁治元年以前[～よりは前、の表記]、
- 寿永中[元号で指定した年間の中でさらに期間を指定]、
- 平安末期、室町時代[年を越える範囲で指定]、
- 年未詳、年欠[不明]、
- 年次「年欠の誤記か」。

5. 「初見年和暦」と「初見年西暦」の整合性チェック

前節の方針に従い変換の対象とした7,969件のレコードについて西暦-和暦変換を行ったところ、「初見年和暦」からの変換値(以下、新値)が「初見年西暦」にすでに入力されている値(以下、旧値)と一致した

レコードは7,685件(96.4%)であった。

新値と旧値が一致しない284件について、その差を調べたところ、プラスマイナス1年の差に収まるものは86件、収まらないものは198件であった。前者には改元日の解釈の違いに起因するものが含まれるであろう。一方後者は入力ミスに起因するものが多いと考えられる。その一部を表3に示した。

最終的には歴史研究者による判定を仰いでデータ校正を進めていくことになるが、Hutime Web API を用いて、時間情報のデータクレンジングが効率的に実施できることが確かめられた。

6. おわりに

Hutime Web API により、確かな根拠に基づく暦変換サービスを簡単に利用できる道が開かれた。「データベースれきはく」は、現在のようにパーソナルコンピュータが便利に利用できる以前に企画・作成されたこともあり、近年の各種情報サービスを活用したデータチェックはすすんでいない。近年のDH研究の進展にともない、研究の基礎となるデータの供給源として「データベースれきはく」がその役割を果たすために、そのデータクレンジングは地味ではあるが確実に実施していくべき業務であり、精進を重ねたい。

日本荘園データベースには、地名や市町村コード等の空間情報も含まれており、これらのデータクレンジングも今後の重要な課題である。

文 献

- [1] 国立歴史民俗博物館(編):国立歴史民俗博物館十年史,国立歴史民俗博物館,1991年10月。
- [2] https://www.rekihaku.ac.jp/research/document/database_rekihaku/,データベースれきはく,国立歴史民俗博物館 Web サイト,2025年2月掲載確認。
- [3] <https://www.rekihaku.ac.jp/doc/gaiyou/syouen.html>,データベース概要-日本荘園データベース,荘園関係文献目録データベース,国立歴史民俗博物館 Web サイト,2025年2月掲載確認。
- [4] 大井将生,中村覚,大野健人,高橋傑:“荘園関係データベースを活用した「nihuBridge LOD」の構築と学校教育への応用,”じんもんこん 2024 論文集, Vol.2024, pp. 45-52(2024-11-30)。
- [5] 関野樹:暦に関する Web API - 暦法の変換と期間の計算,じんもんこん 2017 論文集, Vol.2024, pp. 45-52(2024-11-30)。
- [6] <https://ap.HuTime.org/cal/>, HuTime Web API - Calendar Calculation, 2025年2月掲載確認。
- [7] <https://www.HuTime.jp/basicdata/calendar/calendars/Japanese.html> 時間基盤情報-暦変換(和暦), HuTime Web サイト, 2025年2月掲載確認。
- [8] <https://ja.wikipedia.org/wiki/改元>, 改元, WikiPedia, 2025年2月掲載確認。
- [9] <https://www.HuTime.jp/basicdata/calendar/form.html>, HuTime 暦変換サービス, HuTime Web サイト, 2025年2月掲載確認。

¹ 内部で HuTime Web API を呼び出す実装になっており、修正は容易であろう。

表 1 日本荘園データベースの項目一覧

フィールド名	データ型	解説
荘園コード	半角英数字	各荘園に付けた固有のコード番号
国名	全角漢字	荘園の所在国
郡名	全角漢字	荘園の所在郡
荘園名	全角漢字	荘園の名称
フリガナ	全角カタカナ	荘園名のうち「庄」等を抜いた部分のふりがな
重複コード	半角英数字	異称・分割等の関係にある荘園データの荘園コード
参考市町村	全角漢字	荘園の比定地（現在の自治体名）
市町村コード	半角英数字	参考市町村の市町村コード
明治村字名	全角漢字	荘園の比定地（明治廿年代の市町村字名）
史料村郷名	全角漢字	中世におけるその荘園の内部の地名
領家・本家	全角漢字	荘園の領家・本家
初見年和暦	全角漢字	その荘園の史料的な初見年
初見年西暦	半角英数字	史料初見年（西暦）
出典	全角漢字	主要な史料の出典
遺文番号	全角カタカナ	「平安遺文」「鎌倉遺文」等の文書番号
記録類	全角漢字	関係史料がみえる記録類の名称と日付
地名辞典	全角漢字	荘園の立地・比定地などを記した文献名
備考	全角漢字	荘園についての特徴的な事項
関係文献	全角漢字	荘園に関する文献が「荘園文献DB」に収録されているか

表 2 HuTime Web API による南北朝期の和暦出力

(a) 南朝・北朝を正しく指定した場合

北朝				
正慶	正慶 1 年 4 月 28 日	正慶 2 年 5 月 16 日	1332-05-23	1333-06-28
暦応	暦応 1 年 8 月 28 日	暦応 5 年 4 月 26 日	1338-10-11	1342-05-31
康永	康永 1 年 4 月 27 日	康永 4 年 10 月 20 日	1342-06-01	1345-11-14
貞和	貞和 1 年 10 月 21 日	貞和 6 年 2 月 26 日	1345-11-15	1350-04-03
観応	観応 1 年 2 月 27 日	観応 3 年 9 月 26 日	1350-04-04	1352-11-03
文和	文和 1 年 9 月 27 日	文和 5 年 3 月 27 日	1352-11-04	1356-04-28
延文	延文 1 年 3 月 28 日	延文 6 年 3 月 28 日	1356-04-29	1361-05-03
康安	康安 1 年 3 月 29 日	康安 2 年 9 月 22 日	1361-05-04	1362-10-10
貞治	貞治 1 年 9 月 23 日	貞治 7 年 2 月 17 日	1362-10-11	1368-03-06
応安	応安 1 年 2 月 18 日	応安 8 年 2 月 26 日	1368-03-07	1375-03-28
永和	永和 1 年 2 月 27 日	永和 5 年 3 月 21 日	1375-03-29	1379-04-08
康暦	康暦 1 年 3 月 22 日	康暦 3 年 2 月 23 日	1379-04-09	1381-03-19
永徳	永徳 1 年 2 月 24 日	永徳 4 年 2 月 26 日	1381-03-20	1384-03-18
至徳	至徳 1 年 2 月 27 日	至徳 4 年 8 月 22 日	1384-03-19	1387-10-04
嘉慶	嘉慶 1 年 8 月 23 日	嘉慶 3 年 2 月 8 日	1387-10-05	1389-03-06
康応	康応 1 年 2 月 9 日	康応 2 年 3 月 25 日	1389-03-07	1390-04-11
南朝				
元弘	元弘 1 年 8 月 9 日	元弘 4 年 1 月 28 日	1331-09-11	1334-03-04
延元	延元 1 年 2 月 29 日	延元 5 年 4 月 27 日	1336-04-11	1340-05-24
興国	興国 1 年 4 月 28 日	興国 7 年 12 月 7 日	1340-05-25	1347-01-19
正平	正平 1 年 12 月 8 日	正平 25 年 2 月 4 日	1347-01-20	1370-03-02
建徳	建徳 1 年 2 月 5 日	建徳 3 年 4 月 27 日	1370-03-03	1372-05-30
文中	文中 1 年 4 月 28 日	文中 4 年 5 月 26 日	1372-05-31	1375-06-25
天授	天授 1 年 5 月 27 日	天授 7 年 2 月 9 日	1375-06-26	1381-03-05
弘和	弘和 1 年 2 月 10 日	弘和 4 年 11 月 4 日	1381-03-06	1384-12-16
元中	元中 1 年 11 月 5 日	元中 9 年閏 10 月 4 日	1384-12-17	1392-11-18

(b) 誤って指定した場合

南朝の年号を北朝として変換				
元弘	元弘 3 年 5 月 17 日	元弘 4 年 1 月 28 日	1333-06-29	1334-03-04
延元	建武 1 年 1 月 29 日	建武 5 年 8 月 27 日	1334-03-05	1338-10-10
興国	暦応 1 年 8 月 28 日	暦応 5 年 4 月 26 日	1338-10-11	1342-05-31
正平	貞和 1 年 10 月 21 日	貞和 6 年 2 月 26 日	1345-11-15	1350-04-03
建徳	応安 1 年 2 月 18 日	応安 8 年 2 月 26 日	1368-03-07	1375-03-28
文中	応安 1 年 2 月 18 日	応安 8 年 2 月 26 日	1368-03-07	1375-03-28
天授	永和 1 年 2 月 27 日	永和 5 年 3 月 21 日	1375-03-29	1379-04-08
弘和	康暦 1 年 3 月 22 日	康暦 3 年 2 月 23 日	1379-04-09	1381-03-19
元中	至徳 1 年 2 月 27 日	至徳 4 年 8 月 22 日	1384-03-19	1387-10-04
北朝の年号を南朝として変換				
正慶	元弘 1 年 8 月 9 日	元弘 4 年 1 月 28 日	1331-09-11	1334-03-04
暦応	延元 1 年 2 月 29 日	延元 5 年 4 月 27 日	1336-04-11	1340-05-24
康永	興国 1 年 4 月 28 日	興国 7 年 12 月 7 日	1340-05-25	1347-01-19
貞和	興国 1 年 4 月 28 日	興国 7 年 12 月 7 日	1340-05-25	1347-01-19
観応	正平 1 年 12 月 8 日	正平 25 年 2 月 4 日	1347-01-20	1370-03-02
文和	正平 1 年 12 月 8 日	正平 25 年 2 月 4 日	1347-01-20	1370-03-02
延文	正平 1 年 12 月 8 日	正平 25 年 2 月 4 日	1347-01-20	1370-03-02
康安	正平 1 年 12 月 8 日	正平 25 年 2 月 4 日	1347-01-20	1370-03-02
貞治	正平 1 年 12 月 8 日	正平 25 年 2 月 4 日	1347-01-20	1370-03-02
応安	正平 1 年 12 月 8 日	正平 25 年 2 月 4 日	1347-01-20	1370-03-02
永和	文中 1 年 4 月 28 日	文中 4 年 5 月 26 日	1372-05-31	1375-06-25
康暦	天授 1 年 5 月 27 日	天授 7 年 2 月 9 日	1375-06-26	1381-03-05
永徳	弘和 1 年 2 月 10 日	弘和 4 年 11 月 4 日	1381-03-06	1384-12-16
至徳	弘和 1 年 2 月 10 日	弘和 4 年 11 月 4 日	1381-03-06	1384-12-16
嘉慶	元中 1 年 11 月 5 日	元中 9 年 閏 10 月 4 日	1384-12-17	1392-11-18
康応	元中 1 年 11 月 5 日	元中 9 年 閏 10 月 4 日	1384-12-17	1392-11-18

(c) 南朝・北朝両方で使われた元号に関する時間範囲の違い

北朝として変換				
元徳	元徳 1 年 8 月 29 日	元徳 4 年 4 月 27 日	1329-09-22	1332-05-22
建武	建武 1 年 1 月 29 日	建武 5 年 8 月 27 日	1334-03-05	1338-10-10
明德	明德 1 年 3 月 26 日	明德 5 年 7 月 4 日	1390-04-12	1394-08-01
南朝として変換				
元徳	元徳 1 年 8 月 29 日	元徳 3 年 8 月 8 日	1329-09-22	1331-09-10
建武	建武 1 年 1 月 29 日	建武 3 年 2 月 28 日	1334-03-05	1336-04-10
明德	明德 3 年 閏 10 月 5 日	明德 5 年 7 月 4 日	1392-11-19	1394-08-01

表 3 「初見年西暦」の値が「初見年和暦」から求めた計算値と異なるレコードの例

荘園コード	初見年和暦	初見年西暦	計算値	計算値 (和暦日)	計算値 (西暦日)
0709080	文治三年	1987	1187	文治 3 年 1 月 1 日	1187-02-10
1910005	建久八年	1208	1197	建久 8 年 1 月 1 日	1197-01-20
2216001	正安二年	1167	1300	正安 2 年 1 月 1 日	1300-01-23
3702030	承久三年	1721	1221	承久 3 年 1 月 1 日	1221-01-25
5402018	文永十年	1263	1273	文永 10 年 1 月 1 日	1273-01-21
6707005	応長元年	1211	1311	延慶 4 年 1 月 1 日	1311-01-21