



JAPAN SOCIETY FOR THE PROMOTION OF SCIENCE

日本学術振興会

2020年度統計関連学会連合大会
2020年9月10日13:00-15:00 C会場

人文学・社会科学分野におけるデータインフラストラクチャー
—データカタログの整備—

[2CPM1-03]

人文学分野におけるメタデータの整備状況

独立行政法人日本学術振興会

人文学・社会科学データインフラストラクチャー構築推進センター研究員 (PO)

池内有為 (文教大学文学部・専任講師)

「JDCat*メタデータスキーマ」

1. 人文学分野の検討プロセス
2. 検討方針
3. 追加項目
 - ① 統制語彙 (CV) : トピック
 - ② 統制語彙 (CV) : データタイプ
 - ③ エレメントの拡張 : データの言語
 - ④ 新規エレメント : 所蔵者・寄託者

*JDCat= Japan Data Catalog for the Humanities and Social Sciences
人文学・社会科学総合データカタログ

1. 人文学分野の検討プロセス

JDCatメタデータスキーマ



- Unit of Analysis
- Universe / Population
- Sampling Procedure
- Collection method
- Date of collection
- Sampling Rate
- Title
- Author
- Distributor
- URI
- Time Period(s)
- Geographic Coverage

社会科学分野



J P C O A R

オープンアクセスリポジトリ推進協会

1. 人文学分野の検討プロセス

JDCatメタデータスキーマ



?????

- Unit of Analysis
- Universe / Population
- Sampling Procedure
- Collection method
- Date of collection
- Sampling Rate
- Title
- Author
- Distributor
- URI
- Time Period(s)
- Geographic Coverage

- ???????

社会科学分野

人文学分野



J P C O A R

オープンアクセスリポジトリ推進協会

1. 人文学分野の検討プロセス

JDCatメタデータスキーマ



????

- Unit of Analysis
- Universe / Population
- Sampling Procedure
- Collection method
- Date of collection
- Sampling Rate

社会科学分野

- Title
- Author
- Distributor
- URI
- Time Period(s)
- Geographic Coverage

(共通)

- ??????

人文学分野



J P C O A R

オープンアクセスリポジトリ推進協会

国際的な相互運用性

2. メタデータスキーマの検討方針

■基本方針

- ① 社会科学分野およびJPCOARメタデータスキーマをベースとする
- ② 人文学分野の拠点機関である東京大学史料編纂所のデータを過不足なく記述する
- ③ 国際的な相互運用性を高める
- ④ 人文学分野全般に適用できるように設計する
- ⑤ 長期にわたり使用できるように設計する

2. メタデータスキーマの検討方針

■ユーザビリティ

- ① 国内外の人文・社会科学を中心とした**研究者**にとって必要な情報を記述する
- ② さまざまな分野の**データ登録者**（データ提供者やデータキュレータ）にとって理解や登録が容易なスキーマとする

3. 人文学分野の追加項目

JDCatメタデータスキーマ



????

- Unit of Analysis
- Universe / Population
- Sampling Procedure
- Collection method
- Date of collection
- Sampling Rate

社会科学分野

- Title
- Author
- Distributor
- URI
- Time Period(s)
- Geographic Coverage

(共通)

- **Topic CV**
- **Data Type CV**
- **Data Language**
- **Provider**

人文学分野



J P C O A R

オープンアクセスリポジトリ推進協会

国際的な相互運用性

3.-① トピック統制語彙 (CV) の検討

■方針

- ✓ 史料編纂所のデータ、及び今後データ登録が予想される人文学分野全般をカバーするCVとする
- ✓ 国際的に広く利用されているCVが望ましい

■検討対象

- ✓ UKDA - HASSET、 UNESCO、 Getty
- ✓ RDA¹⁾: CARARE、 DDI、 EAD、 MIDAS
- ✓ SSHOC²⁾、 CO-OPERAS: Dublin Core、 DataCite、 E-RIHS (Heritage Sciences)、 CLARIN (Language Sciences)、 DARIAH (Arts and Humanities)
- ✓ Europeana、 DPLA、 JAPAN SEARCH

1. Research Data Alliance “Metadata Standards Catalog” <https://rdamsc.bath.ac.uk/subject-index>
2. SSHOC “Mapping (meta)data interoperability problems, building the SSHOC Interoperability Hub” <https://www.sshopencloud.eu/news/mapping-metadata-interoperability-problems-building-sshoc-interoperability-hub>

3.-① トピック統制語彙 (CV) の検討

■既存CVの主な問題点

- ✓ 人文学全般を扱い、かつ国際的に広く利用されているCVはみあたらない
- ✓ 日本やアジア系の語彙が不十分
- ✓ 粒度にムラがある

■そこで . . .

- ✓ 幅広い分野の資料を扱う **図書館の分類** を使用する
- ✓ 本来、資料を一意に分類するために用いるが、キーワード的に用いる (複数付与を可とする)

■候補

- ✓ Dewey Decimal Classification (DDC)
- ✓ **日本十進分類法 (NDC)** → 日本の史資料を含む

3.-① トピックCV：日本十進分類法第10版

第2次区分表



第3次区分表 (芸術・美術／文学)

00 総記	General works	50 技術、工学	Technology, Engineering
01 図書館、図書情報学	Libraries. Library and information sciences	51 建設工学、土木工学	Construction, Civil engineering
02 図書、書誌学	Books. Bibliography	52 建築学	Architecture. Building
03 百科事典、用語索引	General encyclopedias. General concordances	53 機械工学、原子力工学	Mechanical engineering. Nuclear engineering
04 一般論文集、一般講演集、雑著	General collected essays. General miscellanies	54 電気工学	Electrical engineering
05 逐次刊行物、一般年鑑	General serial publications. General yearbooks	55 海洋工学、船舶工学、兵器、軍事工学	Maritime engineering. Weapons. Military engineering
06 団体、博物館	General societies. Museums	56 金属工学、鉱山工学	Metal and mining engineering
07 ショーナリズム、新聞	Journalism. Newspapers	57 化学工業	Chemical technology
08 叢書、全集、選集	General collections	58 製造工業	Manufactures
09 貴重品、郷土資料、その他の特別コレクション	Rare books. Local collections. Special collections	59 家政学、生活科学	Domestic arts and sciences
10 哲学	Philosophy	60 産業	Industry and commerce
11 哲学各論	Special treatises on philosophy	61 農業	Agriculture
12 東洋思想	Oriental thought	62 園芸、造園	Horticulture. Landscape gardening
13 西洋思想	Western philosophy	63 養糸業	Sericulture. Silk industry
14 心理学	Psychology	64 畜産業、獣医学	Animal husbandry. Veterinary medicine
15 倫理学、道徳	Ethics. Morals	65 林業、狩猟	Forestry. Hunting and shooting
16 宗教	Religion	66 水産業	Fishing industry. Fisheries
17 神道	Shinto	67 商業	Commerce
18 仏教	Buddhism	68 運輸、交通、観光事業	Transportation services. Tourist industry
19 キリスト教、ユダヤ教	Christianity. Judaism	69 通信事業	Communication services
20 歴史、世界史、文化史	General history, World history	70 芸術、美術	The arts, Fine arts
21 日本史	General history of Japan	71 彫刻、オブジェ	Sculpture. Plastic arts. Objet
22 アジア史、東洋史	General history of Asia	72 絵画、書、書道	Painting. Pictorial arts. Shodo. Calligraphy
23 ヨロッパ史、西洋史	General history of Europe	73 版画、印章、篆刻、印譜	Engraving. Seal engraving
24 アフリカ史	General history of Africa	74 写真、印刷	Photography and photographs. Printing. Graphic arts
25 北アメリカ史	General history of North America	75 工業	Industrial arts
26 南アメリカ史	General history of South America	76 音楽、舞踊、バレエ	Music. Theatrical dancing. Ballet
27 オセアニア史、南極地方史	General history of Oceania. General history of Polar regions	77 演劇、映画、大衆芸能	Theater. Motion pictures. Public entertainments
28 伝記	General biography	78 スポーツ、体育	Sports and physical training
29 地理、地誌、記行	General geography, Description and travel	79 雑芸、娯楽	Accomplishments and amusements
30 社会科学	Social sciences	80 言語	Language
31 政治	Political science	81 日本語	Japanese
32 法律	Law	82 中国語、その他の東洋の諸言語	Chinese. Other Oriental languages
33 経済	Economics	83 英語	English
34 財政	Public finance	84 ドイツ語、その他のゲルマン諸語	German. Other Germanic languages
35 統計	Statistics	85 フランス語、プロバンス語	French. Provençal
36 社会	Society	86 スペイン語、ポルトガル語	Spanish. Portuguese
37 教育	Education	87 イタリア語、その他のロマンス諸語	Italian. Other Romanic languages
38 風俗習慣、民俗学、民族学	Customs, folklore and ethnology	88 ロシア語、その他のスラブ諸語	Russian. Other Slavic languages
39 国防、軍事	National defense. Military science	89 その他の諸言語	Other languages
40 自然科学	Natural sciences	90 文学	Literature
41 数学	Mathematics	91 日本文学	Japanese literature
42 物理学	Physics	92 中国文学、その他の東洋文学	Chinese literature. Other Oriental literatures
43 化学	Chemistry	93 英米文学	English and American literature
44 天文学、宇宙科学	Astronomy. Space sciences	94 ドイツ文学、その他のゲルマン文学	German literature. Other Germanic Literatures
45 地球科学、地学	Earth sciences	95 フランス文学、プロバンス文学	French literature. Provençal literature
46 生物学、一般生物学	Biology	96 スペイン文学、ポルトガル文学	Spanish literature. Portuguese literature
47 植物学	Botany	97 イタリア文学、その他のロマンス文学	Italian literature. Other Romanic literatures
48 動物学	Zoology	98 ロシア・ソビエト文学、その他のスラブ文学	Russian literature. Other Slavic literatures
49 医学、薬学	Medical sciences, Pharmaceutics	99 その他の諸言語文学	Literatures of other languages

700 芸術、美術	The arts, Fine arts	750 工業	Industrial arts
701 芸術理論、美学	Theory of arts. Aesthetics	751 陶磁工業	Ceramic arts
702 美術史、美術史	History of arts	752 漆工業	Lacquer ware
703 参考図書 [レファレンスブック]	Reference books	753 染織工業	Textile arts
704 論文集、評論集、講演集	Essays and lectures	754 木竹工業	Art wood work and bamboo-work
705 逐次刊行物	Serial publications	755 宝石・牙角・皮革工業	Jewelry arts. Ivory and horn carving
706 団体	Organizations	756 金工業	Art metalwork
707 研究法、指導法、芸術教育	Study and teaching. Art education	757 デザイン、装飾美術	Design. Decorative arts
708 叢書、全集、選集	Collected works. Collections	758 美術家具	Artistic furniture
709 芸術政策、文化財	Art and state. Cultural assets	759 人形、玩具	Dolls. Toys
710 彫刻	Sculpture, Plastic arts	760 音楽	Music
711 彫塑材料・技法	Materials and techniques	761 音楽の一般理論、音楽学	Musicology
712 彫刻史、各国の彫刻	History of sculpture	762 音楽史、各国の音楽	History of music
713 木彫	Wood carving	763 楽器、器楽	Musical instruments. Instrumental music
714 石彫	Stone carving	764 器楽合奏	Instrumental ensembles
715 金属彫刻、鋳造	Metal sculpture and casting	765 宗教音楽、聖楽	Religions music. Sacred music
716		766 劇音楽	Dramatic music
717 粘土彫刻、塑造	Ceramic sculpture	767 声楽	Vocal music
718 仏像	Buddhist image	768 邦楽	Japanese music
719 オブジェ	Objet	769 舞踊、バレエ	Theatrical dancing, Ballet
720 絵画	Painting, Pictorial arts	770 演劇	Theater
721 日本画	Japanese painting	771 劇場、演出、演技	Stage. Direction. Acting
722 東洋画	Oriental painting	772 演劇史、各国の演劇	History of theater
723 洋画	Western painting	773 能楽、狂言	Noh play and Noh comedy
724 絵画材料・技法	Materials and techniques	774 歌伎	Kabuki play
725 素描、描画	Dessin. Drawing	775 各種の演劇	Other theaters. Stage
726 漫画、挿絵、児童画	Caricatures. Illustration	776	
727 グラフィックデザイン、図案	Graphic designs	777 人形劇	Puppetry
728 書、書道	Shodo, Calligraphy	778 映画	Motion pictures
729		779 大衆演芸	Public entertainments
730 版画	Engraving	780 スポーツ、体育	Sports and physical training
731 版画材料・技法	Materials and techniques	781 体操、遊戯	Gymnastics. Plays
732 版画史、各国の版画	History of engraving	782 陸上競技	Track and field athletics
733 木版画	Wood engraving	783 球技	Ball games
734 石版画 [リトグラフ]	Lithography	784 冬季競技	Winter sports
735 銅版画、鋼版画	Copperplate and steel engraving	785 水上競技	Aquatic sports
736 ノリウム版画、ゴム版画	Soft-ground engraving	786 戶外レクリエーション	Outdoor recreations
737 写真版画、孔版画	Photo-engraving, Mimeograph	787 釣魚、遊猟	Angling and hunting sports
738		788 相撲、拳闘、競馬	Combat sports. Racing
739 印章、篆刻、印譜	Seal engraving	789 武術	Military arts
740 写真	Photography and photographs	790 雑芸、娯楽	Accomplishments and amusements
741		791 茶道	Tea ceremony
742 写真器械・材料	Camera and photographic materials	792 香道	Burning incense
743 撮影技術	Photography	793 花道 [華道]	Art of flower arrangement
744 複製、印刷	Developing and printing	794 ビリヤード	Billiards
745 複製技術	Photoduplication, Photocopying	795 囲碁	Game of go
746 特殊写真	Specific field of photography	796 将棋	Game of shogi
747 写真の応用	Applications of photography	797 射撃ゲーム	Games of chance
748 写真集	Collections of photographs	798 その他の室内娯楽	Other indoor games
749 印刷	Printing, Graphic arts	799 ダンス	Dancing

3.-① トピックCV：日本十進分類法第10版

■階層構造

- ✓芸術・美術、文学はより詳細な語彙を使用可能に
- ✓データタイプや対象時期等のエレメントを組み合わせて詳細な検索を実現する

第2次区分表

70 芸術. 美術
71 彫刻. オブジェ
72 絵画. 書. 書道
73 版画. 印章. 篆刻. 印譜
74 写真. 印刷
75 工芸
76 音楽. 舞踊. バレエ
77 演劇. 映画. 大衆芸能
78 スポーツ. 体育
79 諸芸. 娯楽

第3次区分表

720 絵画
721 日本画
722 東洋画
723 洋画
724 絵画材料・技法
725 素描. 描画
726 漫画. 挿絵. 児童画
727 グラフィックデザイン. 図案
728 書. 書道

3.-② データタイプ統制語彙 (CV) の検討

■社会科学分野

✓内容＋データ形式

Number	Subject	日本語
1	quantitative research	量的調査
1.1	quantitative research: micro data	量的調査: ミクロデータ
1.2	quantitative research: aggregate data	量的調査: 集計データ、統計表
1.3	quantitative research: other	量的調査: その他
2	official statistics	公的統計
2.1	official statistics: micro data	公的統計: ミクロデータ
2.2	official statistics: aggregate data	公的統計: 集計データ、統計表
2.3	official statistics: other	公的統計: その他
3	qualitative research	質的調査
3.1	qualitative research: digital audio data	質的調査: 音声データ
3.2	qualitative research: digital video data	質的調査: 映像データ
3.3	qualitative research: text data	質的調査: 文書データ

3.-② データタイプ統制語彙 (CV) の検討

■方針

- ✓社会科学分野と同様の体系とすること
- ✓人文学分野のデータを広くカバーできること
- ✓国内外の研究者、及びデータ登録者が理解しやすいこと

■検討対象

- ✓史料編纂所Hi-Catのカテゴリ
- ✓HASSET Information Materials
- ✓JPCOARスキーマ：資源タイプ

■問題点

- ✓粒度
- ✓社会科学分野とのバランス

3.-② データタイプCV：オリジナル

■ 史料編纂所からの提案

4	materials	史資料
4.1	material: text	テキスト
4.2	material: image	画像
4.3	material: audio	音声
4.4	material: video	映像
4.5	material: other	その他
5	publication	出版物
5.1	publication: text	テキスト
5.2	publication: image	画像
5.3	publication: audio	音声
5.4	publication: video	映像
5.5	publication: other	その他
6	fact	事項
7	database	データベース
8	other	その他

3.-③④ エレメントの拡張・追加

■必要な要素

- ✓データの文字
- ✓所蔵者・寄託者

■検討

1. 既存エレメントへのあてはめ

2. 既存エレメントの拡張：Data Language

- ✓「データの文字」をデータの言語とあわせてISO15924で記載 [例：jpn, jpn-Hira, jpn-Kana]

3. 新規エレメントの追加：Provider

- ✓「所蔵者・寄託者」は、Author（作成者）、Distributor（配布者）、Publisher（編集者）のいずれにもあてはまらない

まとめ：人文学分野の検討結果

- JDCatメタデータスキーマ
✓ 合計31のエレメント

- 人文学分野の統制語彙 (CV)

- ① Topic: トピック
- ② Data Type: データタイプ

- エレメントの拡張

- ③ Data Language: データの言語

- 新規エレメント

- ④ Provider: 所蔵者・寄託者

No.	JDCat element	エレメント名	記入レベル (社会科学)	記入レベル (人文学)
1	Title	タイトル	○	○
2	Study ID	整理番号	○	○
3	Author	作成者	○	○
4	Distributor	配布者	○	○
5	URI	URI	○	○
6	Topic	トピック	○	○
7	Summary	概要	○	○
8	Time Period(s)	対象時期	○	○
9	Geographic Coverage	対象地域	○	○
10	Data Type	データタイプ	○	○
11	Access	アクセス権	○	○
12	License	権利情報	○	○
13	DOI	DOI	○	○
14	Data Language	データの言語	○	○
15	Version	バージョン情報	○	
16	Unit of Analysis	観察単位	○	
17	Universe / Population	母集団	△	
18	Sampling Procedure	サンプリング方法	△	
19	Collection method	調査方法	△	
20	Funding Agency	研究助成機関		
21	Grant ID	研究費番号		
22	Alternative title	その他のタイトル		
23	Series	シリーズ		
24	Date of collection	調査日		
25	Sampling Rate	回収率		
26	Bibliographic Citation	引用上の注意		
27	Datafile URI	データファイルURI		
28	Related Studies	関連情報		
29	Related Publications	関連文献		
30	Publisher	編集者		
31	Provider	所蔵者・寄託者		