

## 《資料》

# RUDA の DDI 対応に向けた取り組み

前田 豊  
朝岡 誠

【要旨】 社会調査データは今後の社会の発展に寄与する公共財であり、広くデータが利活用される環境を整備する必要がある。しかし、複数の社会調査データアーカイブが複数並列している現在の日本のアーカイブ状況は、広範なデータの利活用を妨げている状況であると理解でき、そのためには社会調査データを記述するメタデータの統一化、そして各データアーカイブを横断するデータ検索環境の整備が急務の課題となっている。本稿では、事実上の国際基準である DDI を取り上げ、RUDA が現在行っている DDI 導入に向けた取り組みの詳細について報告を行う。また、データの相互検索環境の実現に向けて、現在 RUDA と国立情報学研究所オープンサイエンス基盤研究センター (NII RCOS) が行っている共同プロジェクトに触れ、その基盤となるインフラについて報告する。

キーワード : RUDA, DDI, 相互運用性, オープンサイエンス

## I はじめに

社会調査は、複雑な現在社会の状況を見通すための重要なツールであり、また、そこから得られた社会調査データは、その実施目的が学術目的なのか、それとも政策提言に向けた基礎事実の確認のためなのかという目的の違いがあるものの、等しく今後の社会の発展に寄与する公共財であると言えよう。この意味において、幅広い社会調査データの利活用環境を整えるデータアーカイブの存在は極めて大きい。

しかし、現在の日本のデータアーカイブ環境は、広範なデータ利活用が見込める環境に至ってはいない。日本では、それぞれ趣意の異なるデータアーカイブが並列的に複数存在し、それぞれに特徴をもった(メタ)データのアーカイビングを進めている。こうした複数のデータアーカイブが並列している状況は、大局的に見れば、必ずしも小さくないデータアーカイブの運営・管理コストを分散化している状況として肯定的に評価できるが、データ利用者の観点からみれば、データを探し出すためまでに複数のデータアーカイブにアクセスする必要がある、データの利活用を潜在的に妨げている状況としても理解することができる。

それゆえ、より広範なデータ利活用を支える環境整備に向けては、各データアーカイブを横断したデータ検索環境の構築が一つの方策として考えられよう<sup>2)</sup>。では、こうした環境を整備するために、データアーカイブはどのような取り組みを実施するべきであろうか。我々は以前、この問題関心のもと CESSDA (Consortium of European Social Science Data Archives) が実施している、各国のナショナルアーカイブを横断したデータ検索環境である”Data Catalogue”を紹介し、その根幹となっているメタデータ基準 DDI (Data

Documentation Initiative) を国内アーカイブの共通メタデータとして導入することを喚起した(朝岡・前田 2015)。

本稿は、この前稿の続編として位置づけられるもので、前稿の内容を詳しく補完するとともに、RUDA が現在行っている DDI 導入に向けた取り組みの詳細について報告を行う。加えて、DDI 導入後の横断的検索システムの構築に向けて、現在 RUDA と国立情報学研究所オープンサイエンス基盤研究センター(以下、NII RCOS) が協同して行っている、データ公開環境基盤の整備に向けた研究プロジェクトについても簡単に触れる。

## II DDI の概要と RUDA の取り組み

社会調査データの統計的処理から十分に信頼できる結果を導くためには、例えば調査の実施方法、使用された質問文のワーディングや提示された回答項目、計画サンプルの規模や回収率などといった、実際に統計処理の対象となる社会調査データ(に含まれる変数)そのものだけではなく、その社会調査データが収集されたプロセスや環境についての情報も極めて重要になる。こうした多岐に亘る社会調査データのメタ的な情報、すなわち「データのデータ」であるメタデータを包括的かつ体系的に記述する基準として、今日国際的に広く使われているのが DDI と呼ばれるメタデータ基準である。本章では、DDI の形式である xml に簡単に触れ、導入を試みている DDI Codebook、およびマッピング案を紹介する。

### 1. xml の概括

xml は情報資源を効率的に保持するためのデータ形式の一つであり、要素 `element` と属性 `attribute` の 2 つを用いて情報を記述する。属性は、基本的に何らかの値に意味付けを行うときに使用される。例えば、「訪問面接調査」という文字列そのものは、量的な社会調査法についての知識を持つ人が見れば「社会調査の実施方法としての訪問面接調査」であることは即座に理解できるものの、そうではない人が見れば「訪問面接調査」が一体何についての情報なのかは理解できず、ともすれば文字列そのもの以上の意味は持たない。この文字列が「社会調査の実施方法」に関する情報であることを宣言するのが要素であり、「社会調査の実施方法」を含意する要素に紐づけることで、「調査の実施方法としての訪問面接調査」という情報を十分に表現することができる。一方、属性は要素に付随する情報を記述する目的で使用されることが多く、例えば「調査実施者」を含意する要素と「立教太郎」という値を紐づけることで、「調査実施者として立教太郎」という情報を表現することができるが、それに加えて、例えば所属などの付随的な情報を付け加えたいときに属性は使用される。

基本的には、記述する人が自由に要素・属性を決定することが許されているが、データの異なる主体間での相互運用を想定したときには、共通の意味付けが成された要素・属性を用いる必要があり、そうではない場合には、たとえ同じ意味付けが行われていたとしても、結果として読み取られる情報には齟齬が発生してしまう。こうした一義的な意味付けが行われた要素・属性の集合を提供するのが、名前空間(namespace)と呼ばれるパッケージであり、ある名前空間で用意された要素・属性を使用する限りにおいて、値を理解する枠組みを共有することが可能になる。いうなれば、情報を記述するための「語彙」に相当するものと

してイメージすることができるだろう。

また、データの相互運用を考える際に有用な xml の構造として、xml スキーマも挙げることができる。例えば、社会調査データに関わる重要な情報として、調査実施者や調査の実施方法、計画サンプルサイズといった情報が考えられるが、これらの情報は独立して並列的に評価されるのではなく、すべからく「社会調査データの収集」という大きなカテゴリに含まれる個別的な情報として理解できる。上述した通り、要素を用いることによって各値の意味付けを行うことが可能になるが、この要素と要素との関係性（など）を明確に定義するのが xml スキーマの役割である<sup>3)</sup>。いふなれば名前空間が「語彙」に相当するものだったのに対して、xml スキーマは「文法」に相当するものとして理解することができるだろう。

## 2. DDI Codebook について

このように特定の xml スキーム、そして名前空間を共有する限りにおいて、xml 形式で記述される情報は統一的な枠組みのもとで体系的に理解することが可能である。つまるところ、DDI とは特定の xml スキームが付与された、社会調査データを記述することに特化した名前空間そのものに他ならず、DDI をメタデータ基準とする限りにおいては、異なるデータアーカイブ間でのメタデータ記述の統一化だけではなく、記述されたメタデータの相互運用も（付与された意味を損ねることなく）同時に可能となる。加えて、xml そのものがウェブ上での情報資源の共有をイメージして作成されたマークアップ言語なので、翻っては DDI で記述されたメタデータ情報の検索環境も容易に設定することができる。まさしく DDI とは、社会調査データのメタデータを記述する (Document) ためだけではなく、記述されたメタデータ情報を容易に検索すること (Discovery)、そしてメタデータ情報の相互運用 (Interoperate) も可能にするメタデータ基準なのである<sup>4)</sup>。

ただし、いま現在 (2017 年 12 月)、DDI には異なる設計思想による 3 つのバージョンが存在している。これら 3 つのバージョンのなかでも、ここでは現在 RUDA で導入を計画している、通称 DDI Codebook (以下 DDI-C) について説明する<sup>5)</sup>。

2003 年にはじめて公開された DDI-C は、その名前のおりコードブックに記載する情報の記述を設計思想にもつバージョンで、初版の公開後、DTD (Documentation Type Definition) から xml スキームへの変更や要素の変更などのマイナーチェンジを経て、2014 年に現段階での最新のバージョン (DDI 2.5.1) が公開されている。

この DDI-C は全部で 6 つの名前空間から構成されているが、本質的に社会調査データを記述する目的に特化した名前空間は 1 つだけ ("ddi:codebook:2\_5") で、この名前空間に用意されている 252 の要素を基本的には用いて、当該の社会調査データのメタデータ情報が記述される<sup>6)</sup>。ただし、これらの要素は並列的に存在するのではなく、対応する xml スキーム (codebook.xsd) によって整序されており、大きくは 5 つの親要素 (ルート要素) を起点とするフィールドの中に、子要素・孫要素との階層的な関係づけから包括的かつ体系的に記述を行うことになる。以下が 5 つのルート要素と、それに紐づけられた各フィールドの簡単なイメージである。

1. Document Description : “docDscr”をルート要素とする一連のフィールドで、「DDI 対応のドキュメントそのもの」についての情報、いふなれば、メタデータの「ヘッダー」

に相当する情報を記入する箇所である。

2. **Study Description** : “studyDscr”をルート要素とする一連のフィールドで、例えば調査の時期や対象、方法などといった、広い意味での当該の社会調査データの「調査」に関わる情報を記述対象としている。
3. **Files Description** : “fileDscr”をルート要素とする一連のフィールドで、供されるデータファイルそのものに関係した情報を記入する箇所。
4. **Data Description** : “dataDscr”をルート要素とする一連のフィールドで、当該の社会調査データに含まれる変数レベルの情報、例えば質問文や回答カテゴリ、簡単な記述統計などがここに記入される。
5. **Other Related Material** : “otherMat”をルート要素とするフィールドで、上の4つのいずれにも該当しない関連資料についての情報を記載する箇所。

各フィールドに許容される具体的な要素、およびその意味づけについては、DDIの公式機関であるDDI Allianceが公開しているTag Library<sup>7)</sup>に詳しい説明が記載されている。また、ICPSRやUKDAといった海外の大規模データアーカイブではDDI-Cで記述されたメタデータが標準的に公開されているので、実際の応用例についても確認することが可能になっている。

### 3. Dublin Core と DDI-C とのマッピングについて

現在RUDAでは、DSpaceと呼ばれるリポジトリシステムをカスタマイズしてアーカイブ環境を構築し、Dublin Core (DC) と呼ばれるメタデータ基準を用いてメタデータを作成している。DCについては、公式機関であるDCMI (Dublin Core Metadata Initiative) のサイト<sup>8)</sup>や杉本 (2009) などに詳細が記載されているので、ここではごく簡単な概略を述べておけば、ウェブ上での情報資源を基本的には15個の要素から捉えるメタデータ基準であり、より汎用的なメタデータの記述、および情報の効率的な発見を目的に設計されている。その利用用途も多岐に亘り、例えば、国会図書館では書誌情報を記述する目的で使用され、またICPSRなどの一部海外データアーカイブでは、DDIによるメタデータだけではなく、DCによるメタデータも公開している。

RUDAでは、社会調査データの詳細なメタデータ情報を表現するため、DCの15個の基本要素そのものを使用するのではなく、基本要素に紐づけられた装飾子 (qualifier) を利用した独自のメタデータ基準を整備し、社会調査データ情報の管理、および検索環境の整備を行っている。しかし、こうしたRUDA独自のDCの仕様を前提とした場合、アーカイブ「内」での整合的な情報資源の管理、および検索環境の整備という点では特に問題を孕まないものの、アーカイブ「間」での情報資源の相互運用という点から見れば、DCの基本要素レベルでは共有可能だが、装飾子レベルのスキームを共有しない限りにおいては、装飾子で識別される情報の意味を損ねることなく共有することができない。そのため、メタデータの相互運用、および横断的な検索環境の実現に際しては、RUDA独自の (DCベースの) 基準で記述されたメタデータをそのままの形で転用することはできず、より構造がリジットに共有されているメタデータ基準、すなわちDDIに移し替えるマッピング作業が必要となる。

すでに、DDI Allianceが「推奨」という形でDCとDDI-Cのマッピング案を提唱してい

るものの<sup>9)</sup>、基本的には DC の基本要素と DDI-C の要素との対応付けになっているため、装飾子レベルでのスキームを応用している RUDA のメタデータ情報にそのまま適用した場合には、装飾子で表現した情報が適切に反映はされない。そこで、我々は、DC から DDI へのマッピングにあたり、DDI Alliance の推奨マッピングを利用するのではなく、RUDA で使用している装飾子に込められた意味内容 (セマンティックス) を踏まえつつ、新たなマッピング案を作成することにした。

マッピング案の作成にあたっては、まず正確な DDI-C の語彙と文法を把握するため、DDI-C の名前空間で提供される要素について、その定義づけ、および xml スキームに従う階層構造を Tag Library をもとに網羅的に精査を行うことから始め、また、並行して具体的な要素の使用イメージを掴むために、ICPSR や UKDA で実際に公開されている DDI-C ベースのメタデータを適宜参照し、マッピング先となる DDI-C の全体的かつ個別的な枠組みを用意した。この理解をもとに、今度はマッピング元となる DC の装飾子について、その意味内容と具体的な利用事例を (再度) 確認したうえで、DC の装飾子と DDI-C の要素との暫定的なマッピング案を作成した。

次に、暫定的なマッピング案に従って xml 形式に移されたメタデータが、妥当な文章 (xml valid document) であるかを確認する目的で、(1) ICPSR が公開している DDI-C 対応のスタイルシート<sup>10)</sup>を用いて HTML に変換し、正しく狙い通りの対応ができているのか、そして (2) オープンの妥当性検証用のチェッカー<sup>11)</sup>を利用し、xml スキーマに即して正しく記述されているのか、の 2 点について検討を行い、マッピング案の修正を行った。以上の手順を踏まえて最終的に完成したのが、図表 1 にて示したマッピング案である<sup>12)</sup>。(紙幅の都合上、ここでは Document Description、および Study Description の 2 つのフィールドに限定している)。

現在、RUDA では前章で作成した暫定的なマッピング案に基づき、DDI-C をベースとするメタデータの作成を順次進めている。DDI 文書の作成・編集ソフトとして知られているものに、Collectica や Nesstar Publisher、SSJDA が開発した Easy DDI Organizer があるが、RUDA では汎用的な xml の編集ソフトを用いてメタデータを記述している<sup>13)</sup>。DDI 形式のメタデータは、来年度より順次 RUDA のホームページにて公開する予定になっている。

### III 横断的なデータ検索環境の整備にむけて

前章でも触れたとおり、DDI をメタデータの統一基準とすることで、異なるデータアーカイブ間でのメタデータの相互運用が、その意味内容を損ねることなく可能となる。しかし、各データアーカイブがそれぞれに DDI ベースのメタデータを公開したとしても、それがすなわち、データ利用者にとっての効率的なデータ検索環境が整備されたことを意味するわけではなく、それらのメタデータ情報を一元的に集約することで、はじめて効率的なデータ検索環境の実現につながる。したがって、どのようにメタデータ情報を一元的に集約するのか、という問題は、改めて別途考える必要があるだろう。

## (30) 社会情報教育研究センター研究紀要『社会と統計』第4号

DDI-C	DC 基本要素	DC 装飾子	意味
codebook			
docDscr			
citation			
titlStmt			
IDNo	identifier	StudyNo	調査番号
titl	-	-	(メタデータタイトル)
parTitl	-	-	(メタデータタイトル・多言語)
stdyDscr			
citation			
titlStmt			
IDNo	identifier	StudyNo	調査番号
titl	Title	display	調査名
altTitl	Title	alternative	調査略称
parTitl	-	-	(調査名・多言語)
prodStmt			
fundAg	description	sponsorship	調査資金
grantNo	-	-	(基金の番号)
rspStmt			
AuthEnty	Creator	-	調査主体 / 調査代表者
serStmt			
serName	relation	ispartofseries	シリーズ
distStmt			
depositor	publisher	depositor	寄託者
distDate	Date	issued	公開日
distrbtr	publisher	distributor	配布者
verStmt			
version	identifier	version	ファイルのバージョン
holdings	identifier	uri	URL
dataAccs			
useStmt			
conditions	Rights	AccessRights	利用条件
citReq	identifier	citation	引用時の表記
method			
anlyInfo			
respRate	description	ResponseRate	有効回収率
dataColl			
collMode	description	ModeOfCollection	調査方法
dataCollector	description	investigator	調査実施者
sampProc	description	SamplingProcedure	標本抽出法
targetSampleSize			
sampleSize	description	SampleSize	標本サイズ
timeMeth	-	-	(クロスセクショナル・パネルなどの種別)
notes	description	NoOfValidResponses	有効回収数
stdyInfo			
abstract	description	abstract	調査概要
subject			
keyword	subject	-	キーワード
topcClas	description	discipline	研究分野
sumDscr			
anlyUnit	description	UnitOfObservation	観察単位
collDate	coverage	temporal	調査時期
geogCover	coverage	spatial	調査地域
geogUnit	-	-	(最小の調査範囲水準)
nation	-	-	(国)
universe	coverage	-	母集団
dataKind	Type	-	資料種別
othrStdyMat			
relPubl	relation	-	関連資料 (うち、データを使った研究成果)
relMat	relation	-	関連資料 (うち、データを使った研究成果以外)
notes			備考

図表1 DDI-C と DC とのマッピング ※カッコ内で示された項目は、現在 RUDA では使用していない項目。

例えば、寄託・公開データそのものを一つのサーバー上に集約し、検索システムを構築するという試みが考えられるが、前稿でも論じたように、セキュリティ上のリスクが高く、ま

た、検索システムを運営する人的・金銭的コストの観点からも難しいと言える。加えて、こうした試みは、これまでに各データアーカイブに寄託されたデータを、改めて(他の)管理基盤に寄託する必要があるので、寄託データが複数のアーカイブで保持されることになり、各アーカイブが寄託者に対して果たすべき責任の範囲という点で、慎重な議論が必要になると思われる。そこで、RUDAはNII RCOSと共同し、NII RCOSが進めている研究データ基盤を利用した横断的なデータ検索環境の可能性を模索するプロジェクトを開始した。本章では、そのベースラインのインフラである研究データ基盤を簡単に紹介し、RUDAが現在、そして将来行うべき取り組みについて概説する。

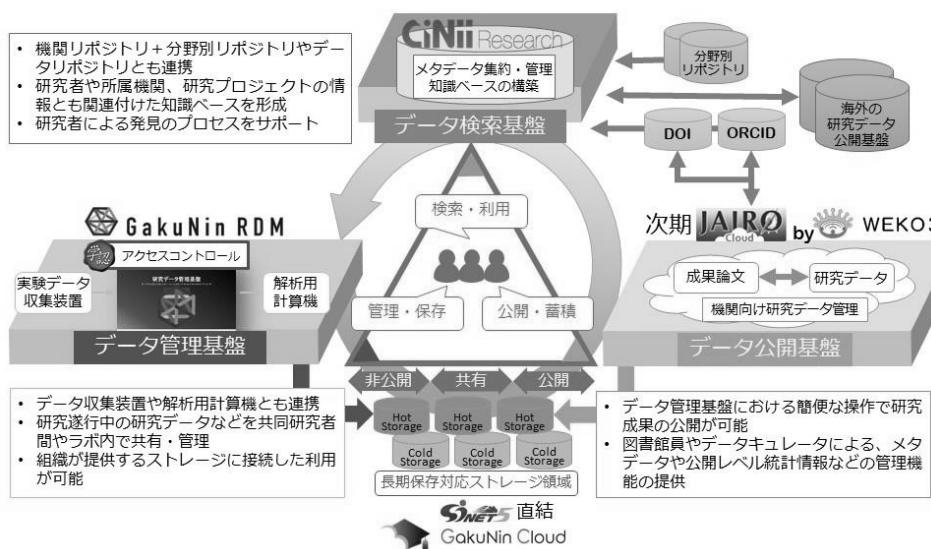
2013年のG8首脳会合、科学技術大臣会合において研究データのオープン化が合意されて以降、オープンサイエンスに関する関心が国際的に高まっている(村山・林 2016)。2016年1月閣議で決定した第5期科学技術基本計画では「オープンサイエンスの推進」の項目が盛り込まれ、国内でもオープンサイエンスのための環境整備が進められている。

オープンサイエンスの流れの一つに、研究成果と同時に根拠となる研究データの一般公開を目指す、オープンリサーチデータがある。データを公開することにより、研究上の不正の抑止や研究データの再利用による新たなイノベーションの進展が期待されるが、各専門分野の研究者が自分たちの所持する研究データを公開するためには、メタデータの標準化や組織化をはじめとした相互運用性を重視したインフラが必要となる。

NII RCOSが開発した研究データ基盤は、こうした相互運用性を重視して設計されたシステムであり、研究データや関連の資料を管理・公開・検索するための情報インフラを提供する。



## RCOSによるサービスの概要



図表 2 NII RCOS 研究データ基盤の概要 (<https://rcos.nii.ac.jp/service/>より引用)

研究データ基盤は、それぞれ独立したデータ管理基盤 (GakuNin RDM)、データ公開基盤 (WEKO3)、データ検索基盤 (CiNii Research) の3つのシステムで構成され、研究活動の研究ライフサイクルの段階に応じてそれぞれの基盤が利用される<sup>14)</sup>。

データ管理基盤は、研究プロジェクト実施中に収集されたデータを管理し、研究グループ内でファイルを管理・共有するシステムである。利用者が所属する機関のストレージとの接続が可能であり、またデータ解析ソフトやデータ収集装置との連携も視野に収めて設計されている。

研究プロジェクト終了後、研究者が公開すると判断したデータや関連の資料は公開基盤を用いて公開する。現在日本の学術機関リポジトリサービス JAIRO Cloud の研究データ版ととらえることができ、各研究分野に合わせて柔軟にメタデータ情報の入力と DOI の付与が可能である。

そして、公開基盤で公開された研究データや他のレポジトリで公開されたメタデータ情報は検索基盤にて集約・管理され、利用者は CiNii Research にて横断的に研究データを検索することができ、検索したリンク先より公開基盤にアクセスできる。これらの研究データ基盤は 2020 年度より本格的に運用され、JAIRO Cloud と同様に利用を希望する学術機関にサービスが提供される予定である。

このように現在、国内のオープンサイエンス環境が急ピッチで整備され、各研究者が自分たちの研究データを公開し、自分たちの研究成果を広く発信する土壌が整いつつある。こうした基盤のなかで、とくにデータ公開・管理基盤に連携することで、各データアーカイブがそれぞれに保有しているメタデータの相互検索環境を実現すると理解できる。現在 RUDA では、とくに公開基盤のインフラとなっている WEKO3 の実験的環境において、DDI によるメタデータの作成・編集を試み、今後のデータ寄託が見込まれる各研究者、そしてデータアーカイブ運営者にとって、どのようなインターフェイスであれば負担の少ない環境が構築できるのか、について模索しているところである。加えて、検索基盤にてクエスチョン・バンクなどの追加的な情報検索を実装できるかどうかを検討し、さらに管理基盤からデータが寄託されることを想定した新たな業務フローの作成を考える予定である。

しかし、データ公開におけるハードの側面であるインフラが整備されたとしても、ただちにオープンリサーチデータの実現が見込めるわけではない。研究者の中でもオープンサイエンスの考えが浸透しつつあるが、学術機関や研究分野内で研究データの管理、保管、公開の方針は定まっていないのが実情である(倉田・松林・武田 2017)。社会調査データにおいても、公開の目安やアナログ/デジタルデータの保存体制、データの所有権、個人情報の取り扱いについての一貫した見解は共有されておらず、これらの錯綜が、翻ってはオープンリサーチデータの実現を阻害する可能性も否定できない。そのため、オープンサイエンスのためのハードの側面であるインフラ整備とは同時に、ソフトの側面にあたるデータ公開・管理に関するポリシーの整備も同時に必要となると考えられる<sup>15)</sup>。

#### IV さいごに

冒頭にも述べたが、社会調査は、今日の社会状況を理解するためのツールであり、社会調



査データは社会の発展に寄与する公共財である。オープンサイエンスの理解は、こうした社会調査（データ）のありかたを的確にとらえたものであり、それに向けたインフラは整いつつあると言えるだろう。その中でデータアーカイブの役割は、従前の社会調査データを独自に管理し、また公開するといったものに限定されず、より広く調査データの利活用を促進するための能動的な役割を果たす必要がある。

本稿では、とくにデータアーカイブ間を横断する検索システムの構築を取り上げ、現在、そして将来 RUDA が行う取り組みについて議論してきた。しかし、国内数あるデータアーカイブの一つだけが取り組んだとしても、データの効率的な検索環境を構築できるわけではなく、まず国内データアーカイブの網羅的、そして有機的な連携が、データの効率的な検索環境の実現には必須である。

## 注

- 1) ここではデータ利用者の視点を想定して議論を展開しているが、データ寄託者の観点からも、多様な目的でのデータアーカイブ利用が予測される（前田・朝岡 2017）。
- 2) 日本学術会議社会学委員会社会統計アーカイブ分科会が、2014 年に発表した提言の中でも（<http://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-22-t193-1.pdf> アクセス日：2017 年 12 月 7 日）、データアーカイブおよび関連機関との間に有機的なネットワークを形成し、データ公開基盤を整備する必要性を訴えている。
- 3) 厳密に言えば、許容される値なども xml スキーマで定義することができるが、ここでは割愛する。
- 4) さらにクエスチョン・バンクや変数比較といった、DDI ドキュメントを利用、もしくは基盤とする広範なデータ検索環境も整備されつつある。
- 5) 最新のバージョンである DDI.3 については、佐藤・米倉（2011）に詳しい。また、異なるバージョンであっても互換性は担保されている。
- 6) 要素の再利用もされるので、捉えられる意味的にはより多くの情報が記述できる。また、すべての要素を使用する必要はなく、多くの要素の使用は任意である。
- 7) [http://www.ddialliance.org/Specification/DDI-Codebook/2.5/XMLSchema/field\\_level\\_documentation.html](http://www.ddialliance.org/Specification/DDI-Codebook/2.5/XMLSchema/field_level_documentation.html)（アクセス日：2017 年 12 月 7 日）
- 8) <http://dublincore.org/>（アクセス日：2017 年 12 月 7 日）
- 9) <http://www.ddialliance.org/resources/ddi-profiles/dc>（アクセス日：2017 年 12 月 7 日）
- 10) <http://www.ddialliance.org/sites/default/files/codebook2-0.xml>（アクセス日：2017 年 12 月 7 日）
- 11) <https://www.corefilng.com/opensource/schemaValidate/>（アクセス日：2017 年 12 月 7 日）
- 12) 朝岡・前田（2015）で示したマッピング案をベースとしつつも、上述の手順に基づく検討から一部修正・追加を加え、要素の階層構造を明記したものをここで示した。
- 13) 現在作成しているメタデータは公開を前提としており、ICPSR や UKDA をならい Document Description と Study Description のみを記述している。そのため、1 データにつき 100 行ほどの文量なので、xml の編集ソフトを利用したほうが簡便に作業を進められる。ただし Data Description も作成するとなると 1000 行を超える膨大な文量となるので DDI 文書作成・編集ソフトを利用しないと難しいと思われる。
- 14) 詳細については NII RCOS のサイトを参照（<https://rcos.nii.ac.jp/> アクセス日 2018 年 1 月 18 日）。
- 15) こうしたデータ管理に関わる一つの基準として、データアーカイブの認証機関である

Core Trust sealed のチェックリストが参考になるだろう

(<https://www.coretrustseal.org/> アクセス日 2017年12月7日).

#### 参考文献

- 朝岡誠・前田豊, 2015, 「データアーカイブ事業の展望に関する一考察」『社会と統計』1:49-62.
- 倉田敬子・松林麻実子・武田将季, 2017, 「日本の大学・研究機関における研究データの管理, 保管, 公開—質問紙調査に基づく現状報告」『情報管理』60(2):119-127.
- 前田豊・朝岡誠, 2017, 「海外データアーカイブの動向—IASSIST 年次大会の報告から—」『社会と統計』3:27-35.
- 村山泰啓・林和弘, 2016, 「欧州オープンサイエンスクラウドに見るオープンサイエンス及び研究データ基盤政策の展望」『STI Horizon』2(3):49-54.
- 杉本重雄, 2009, 「ダブリンコアの現在」『「デジタル図書館」ワークショップ第36回発表論文』32-45.

## Summary

# On the Effort to Implement DDI in RUDA

Yutaka Maeda and Makoto Asaoka

Social Research data is valuable public resources that contribute to the development of society, and favorable circumstance that facilitate border data usage should be prepared. However, the current environment of Japanese data archives in which each archive independently manage is not such favorable circumstance. Standardizing the format for describing metadata and introducing the cross-searchable system covering whole archives in Japan are considered to be imperative tasks. Based on this understanding, this paper reports the details of the projects aiming to introduce DDI, as a de facto global standard for describing the metadata of social research data, into RUDA. In addition, we introduce the collaborative projects between RUDA and NII RCOS (National Institute of Informatics Research Center for Open Science and Data Platform) and report the baseline infrastructure that would be expected to realize the cross-searchable environment in Japan.