《論文》

データアーカイブ事業の展望に関する一考察

A Thought on The Development of Data Archive Project

朝岡　誠 Makoto Asaoka
前田　豊 Yutaka Maeda

RUDA is a data archive maintained by CSI. In its 2010 launch, there are three motives: data archiving, provisions of opportunity for secondary, and educational usage. The first aim of this paper is to confirm the extent of RUDA's contribution to these motives in the past five years. In addition, the paper discusses the tasks that need to be executed for the further development of data archives. As for the first aim, RUDA has certainly made progress on its data repository and publication on a yearly basis. Further, RUDA has been especially utilized for educational usage. As for the second aim, the contemporary situation in Japan is such that there are several data archives; thus, it is required to implement a comprehensive searching system that covers all data archives for efficient data accessibility, similar to the manner in which CESSDA functions. For this requirement, DDI is a suitable metadata standard because of its structure, concept, and prevalence. For further development, we propose that national data archives, including RUDA, agree to create a DDI-based record metadata and to collect metadata so that users can easily find the data of interest.

Keywords：Data Archive, RUDA, DDI
キーワード: データアーカイブ，RUDA，DDI

Ⅰ　はじめに

RUDA（Rikkyo University Data Archive）は，「調査技法，情報技法及び統計技法の活用による本学（立教大学）における研究活動の高度化への寄与及び学生に対する研究基礎能力の涵養を目的」とする立教大学社会情報教育研究センター（CSI）が運営する社会調査データアーカイブである．SSJDA（Social Science Japan Data Archive：東京大学社会科学研究所附属社会調査・データアーカイブ研究センター）や SORD（Social and Opinion Research Database：札幌学院大学），SRDQ（Social Research Database on Questionnaire：大阪大学大学院人間科学研究科 SRDQ 事務局）といったデータアーカイブの後発として 2010 年に学内公開，2011 年に一般公開をはじめ，2015 年には一般公開から 5 年目の節目の時期を迎える．

今日，データアーカイブに期待される機能は多岐に亘る．調査データの収集・整理・保存という「アーカイブ」そのものとしての機能はもちろんのこと，二次分析の機会を供することで斯界における研究発展に寄与し，すでに存在する実証的研究の「再現性」を担保する環境を整えるという意味での学術的貢献や，既存の社会調査への参照機会を提供することで，効率的に質の高い社会調査の実施につながるという社会調査に対する貢献も指摘されている（佐藤・佐藤 2006，佐藤 2012）．加えて，社会調査教育に対する貢献にも期待が集められており，実査までの時間的ウェイトが大きかった社会調査教育（収集過程重視型教育）に代わり，公開データの利用を前提する社会調査教育（公開データ利用型教育）の提案がなされている（稲葉 2000）．

CSI の研究支援事業の一環として立ち上がったデータアーカイブ事業も，「アーカイブ」そのものの機能に加えて，そこから派生的に実現する機能も企図として設立された．RUDA の起案者である松本（2012）によれば，データアーカイブ設立の背景には，日本では本来公共財としての性格を持つはずの社会調査データを保存し，公開する環境が不十分であるため，欧米では浸透している二次分析や既存の実証研究に対する批判的検討の機会が失われている状況があった．また自身の経験に基づき，社会調査教育の観点から既存の社会調査データ公開環境の重要性も，データアーカイブ事業の立ち上げに向けた動機に含まれていた．本稿では，こうしたモチーフで設立された RUDA の，設立から今日までの社会調査データの公開実績と公開データの二次利用・教育利用実績の観点からの変遷を振り返り，RUDA がデータアーカイブとして期待されている機能をどの程度実現してきたのか，という点について検討する．

加えて，本稿では，今後の持続的なデータアーカイブ事業に向けて必要となる課題を検討したい．というのも，永続的な機関であることが明にも暗にも含意されているデータアーカイブだが，その持続的な運営には，現在，そして今後迎えうる状況に対応できる仕組づくりが不可欠だからである．

近年，寄託者・利用者の水準でデータアーカイブの認知は広がっている．例えば，社会調査を題材にしたセミナー論文では，データアーカイブへの寄託を社会調査プロセスの一環として位置づけるものも存在し（中野・小松 2003），調査企画／設計・データ収集・データ分析までの内容が基本であった社会調査法の教科書でも，二次分析の観点からデータアーカイブの内容に触れる機会が多くなっている（大谷ら 2005，轟・杉野 2013）．上述したデータアーカイブの機能が十分に実現するためには，少なくない数の公開データとその利用が必要になるが，少なくともデータの寄託者，そしてデータアーカイブの利用者の水準では着実にその土壌が形成されつつあると理解できよう [1)]．

しかし，今日の日本におけるデータアーカイブの状況は，少なくとも利用者の水準で涵養された期待に必ずしも十全に応えるものではない．日本で活動を行っているデータアーカイブは，それぞれの趣意の異なるアーカイブ運営を行っている．確かに，SSJDA のように分野に特化しない大規模データアーカイブも存在するが，例えば，SORD は「地域」データのアーカイブを推進し，日本女子大学の RIWAC-DA では女性の労働に関連したアーカイビングを標榜している．また，労働政策研究・研修機構の JILPT データアーカイブも同様に労働に特化したデータをアーカイビングしている．こうした特定分野に特化したアーカイビング・公開を行うデータアーカイブの並列している状況が，今日の日本のデータアーカイブ状況を表す一つの特徴であると理解できる．

こうした趣意の異なるデータアーカイブが並列的に存在している状況は，一方で利用可能なデータの多様性を示すと肯定的に理解できるだろう．しかし，利用者のヒューリスティックの観点から見れば，その多様さゆえに，利用者の利用実現性を低減させる一つの要因であるとも理解できる．後者の点を敷衍すれば，たとえ一般にデータアーカイブの存在・利用価値が膾炙し，その利用に向けた動きがあるといっても，必ずしも並列的に存在しているデータアーカイブ状況がそれを十全な形でくみ取れるような状況にはないとも言えよう．この点を踏まえ，本稿では，現在の国内データアーカイブの状況を踏まえつつ，今後のデータアーカイブ活動に向けた課題を述べ，その中で RUDA として果たすべき課題を議

論したい.

以下，第 2 節では，RUDA における寄託からデータ利用までのプロセスについて，システム管理の観点から紹介を行う．第 3 節では，一つ目の目的について，寄託データ数，利用者実績，利用目的に関する時間的な推移を述べ，続く 4 節では，それまでの議論を踏まえつつ，今後のデータアーカイブ事業に向けた課題を，特にメタデータの観点から議論する．

Ⅱ　RUDA について

本節では，データの寄託から公開・データ利用までのプロセスについて，その管理・運営体制を中心に概説する．

1. 寄託に際して

RUDA では社会調査データと使用された質問紙，（あれば）コードブックとともに，メタデータの提出を寄託者にお願いしている．メタデータとは「データに関するデータ」を意味し，社会調査データに限れば，調査実施者や調査目的，母集団・回収率といった当該社会調査そのもののレベルでのメタデータと，当該社会調査で用いられた質問文や回答カテゴリといった変数に付随するメタデータの 2 つが想定されうる．これらのうち，RUDA への寄託に際しては，前者の社会調査レベルでのメタデータの記入を寄託者にお願いしている．

表 1 記入項目

記入項目	Dublin Core element
寄託者	publisher
調査名	title
調査略称	title
キーワード	subject
研究分野	description
主要変数	description
調査目的	description
関連資料	relation
調査主体	creator
調査代表者	creator
調査資金	description
母集団	coverage
調査時期	coverage
標本抽出法	description
標本サイズ	description
有効回収数	description
有効回収率	description
観察単位	description
調査方法	description

表 1 は具体的な記入項目を示したものである．これらの記入項目は Dublin Core という

メタデータ基準の項目に対応づけされている．Dublin Core については，DCMI（Dublin Core Metadata Initiative）のサイト[2)]や杉本（2009）などに詳しいが，大枠では（ウェブ上での）情報資源を 15 個の基本項目（element）から捉えるメタデータ基準であり，国立国会図書館をはじめとした全国の図書館・リポジトリなどでも採用されている汎用性の高い基準である．RUDA は社会調査データのアーカイビングを目的としているので，Dublin Core の基本項目のそれぞれに付随する限定子（qualifier）をカスタマイズして，社会調査データに即したフォーマットを作成している．検索に関しては後述するが，RUDA ではこのように標準化されたメタデータ基準を用いることで，一元的な水準での寄託・公開データ管理と検索体系の整備を行っている．

2. 変数レベルのメタデータ作成とクリーニングについて

寄託されたデータはクリーニング作業に入り，次に変数レベルのメタデータ作成と（必要があれば）データのクリーニング作業が行われる．これらの作業には，RA として雇用された立教大学の院生数名，および学術調査員と称するスタッフと社会調査部会の助教があたる．

大まかな作業の内容と流れは図 1 に示す通りである．まずは質問紙・コードブックと照らし合わせつつ，寄託されたデータには含まれるが，質問紙・コードブックに記載されていない冗長な変数（寄託者が作成した分析用の新規変数など）を削除する．加えて，個人が特定されうる可能性がある変数についても，秘匿性の観点から同様に削除する．

次に SPSS を用いた変数レベルのメタデータ作成を行う．変数レベルのメタデータを質問文と回答カテゴリの 2 つから理解し，質問紙・コードブックに記載されている情報に忠実に従いつつ，各変数の変数ラベルに使用された質問文を，値ラベルに使用された回答カテゴリをそれぞれ入力する．

図 1　作業の流れ

続いて，変数のクリーニング作業を行う．大きくは 2 つの観点からクリーニング作業が行われる．一つは非該当に関するもので，先行する質問の回答結果に応じて，以降の質問に対する回答が制限されている場合に相当する．通常，回答が制限されているケースについては欠損処理が行われるが，欠損値が一つの値で表されている場合，その欠損値が回答しなかった／拒否したことによるのか，それとも回答の制限に起因したものなのかを区別することができないため，欠測メカニズムを所与とする統計的な欠損処理に際してノイズが発生する．この点に鑑み，RUDA では質問紙で明示的に回答が制限されている場合に限り，欠損値を無回答と非該当に区別する作業を行う．

もう一つの処理は複数回答に関わる処理で，用意されているいくつかの回答カテゴリか

ら複数の選択をお願いする質問，とくに，例えば「その他」や「どれにも該当しない」といった補集合をカバーする回答カテゴリが含まれる場合に該当する．通常，各回答カテゴリに対して選択・非選択を指示する 2 値変数で回答結果が表されるが，当該の質問に対して回答を行わなかった場合は，いずれの回答カテゴリにも非選択を表す値が入ってしまい，非選択と無回答を識別することができない．こういったケースに対して，RUDA では非該当と同じ理由から無回答と非選択を区別する作業を行っている．

これらクリーニング作業の画一性・信頼性を担保するため，具体的な作業手順を記したクリーニングマニュアルに基づいて作業は行われ，作業従事者による作業内容のダブルチェックを義務づけている．もちろん，寄託段階でのデータに十全な処理が行われている場合は，これらのクリーニング作業は省かれる．

3. 公開について

変数レベルのメタデータ作成（とクリーニング作業）が終われば，そのデータは当該調査で用いられた質問紙とともに RUDA 上に公開される. RUDA では Dublin Core をデフォルトのメタデータ基準とする Dspace と呼ばれるオープンソースをカスタマイズしてアーカイブシステムを構築している．公開データごとに「アイテム」と呼ばれる一つの層を作成し，寄託者に記入してもらったメタデータ情報，および運営側で付与するメタデータ情報（調査番号や公開日，調査地域など）がそのアイテムに記入される．こうしたメタデータ情報は，ユーザーインターフェイスにも援用されており（図 2），利用者はメタデータ情報レベルで公開データの検索を行うことができる．

4. データ利用

データの利用に際して，利用者にはアカウントの作成とともに，利用申請書の提出をお願いしている．利用申請書の記入内容は，利用するデータの調査番号・調査名，申請者の情報（氏名，住所，連絡先，身分，共同利用者の有無，（身分が学生の場合）承認者の情報）といった項目に加え，そのデータがどういった目的で使用されるのか（利用目的），どのような分析計画に基づいて使用するのか（利用計画）についても記入するように求めている．

利用申請書の記載事項に不備がないことを運営側で確認したのちに，利用申請者にはデータへのアクセス権が付与され，RUDA 上からの当該データのダウンロードが可能になる．データの提供は，主に SPSS ユーザーを想定し，txt ファイルでのローデータとセットアップ用の sps ファイルによる提供を行っている．ただし，今日の多様な分析環境に鑑み，一部の公開データでは，SPSS ユーザーを想定した por（ないしは sav）ファイル，R ユーザーを念頭においた rda ファイル，より広範な分析環境を念頭においた csv ファイルによる同時提供を試みている．

データの利用期限は，利用承認が行われた翌年の 3 月末を設定している．期限までに，利用者は利用報告書の提出を義務づけられており，当該データの利用延長／破棄希望，研究目的の利用であれば，その成果物の有無と詳細などについての報告が求められる．公開データの利用は，利用申請が承認された時点でデータベースに記録されており，利用申請書の記載事項レベルでの情報管理が行われている．また，利用報告書が届いた時点で，各レコードはその内容に従い更新される．

図 2　RUDA の画面

Ⅲ　RUDA の実績について

このように，RUDA では寄託・公開データを Dublin Core ベースのメタデータを基準として，Dspace で管理しており，利用実績に関しては利用申請・報告書の内容に基づき管理している．これらの情報を用いて，この節では，本稿の一つ目の目的である公開・利用実績について時系列的に確認し，その特徴を概説する．

1. 公開データ数の推移

図 3 は，2010 年から 2014 年（11 月末現在）までの公開データ数の推移を表したものである．学内公開を開始した 2010 年時点では公開データ数が 15 件であったのが，その後一般公開を始めた 2011 年度には 21 件へと公開データ数を増やし，その後 30 件，36 件，44 件と，おおよそ年 10 件弱のペースで堅調にデータの公開を進めている．

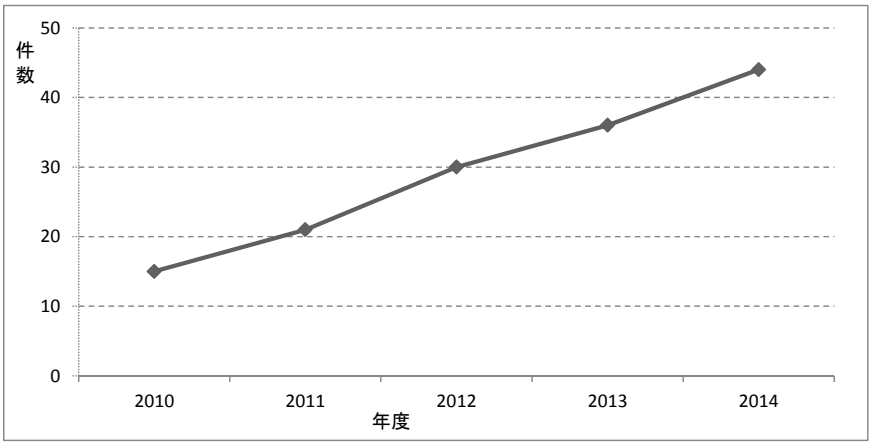

図 3 公開データ数の推移

2. アイテムアクセス数の推移

こうした公開データ数の漸次的な増加は，RUDA が社会調査データの「アーカイブ」としての役割，そして，二次分析・教育利用に向けた公開環境の整備を着実に進めていることを意味している．では，どれだけ一般に RUDA はそのデータアーカイブとしての存在を認知されるようになったのか．この点を確認するために，図 4 では，公開データの情報が表記されているアイテムへのアクセス数について，年次別にその推移を示した．

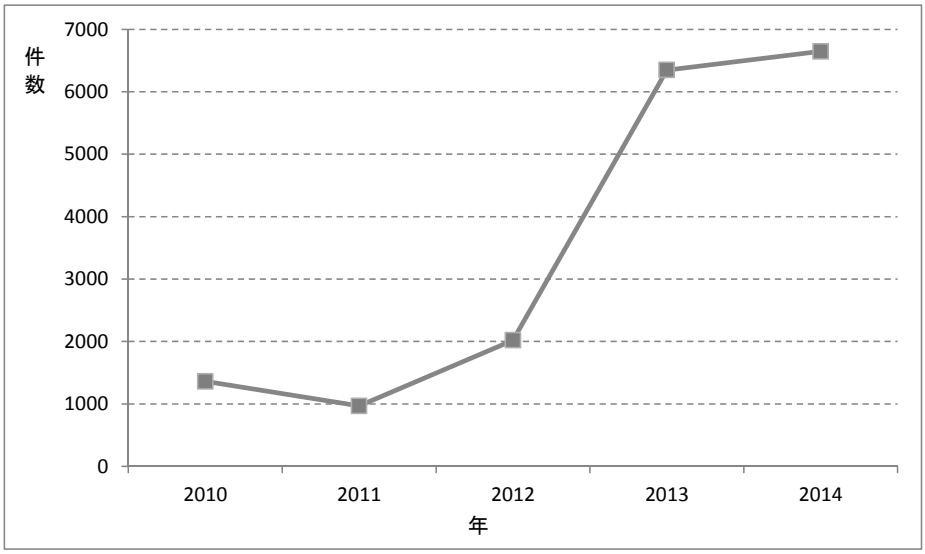

図 4　RUDA アイテムアクセス数の推移

一般公開を始めた 2011 年以降，着実にアクセス数は増加している [3]．2012 年には，2011 年度のアクセス数の倍にあたる約 2,000 件のアクセス数が，また，2013 年には 6,000 件強という著しいアクセス数の増加が確認でき，2014 年（11 月末現在）も 2013 年度のアクセス数に比肩する値（約 6,500 件）で推移している．

こうしたアイテムアクセス数の推移を踏まえれば，一般公開以降，徐々に RUDA の存在が広く膾炙しつつあると理解でき，とくに 2013 年度以降でその傾向は顕著であると判断できる．もちろん，先に確認した公開データ数の増加に伴い，RUDA 上にあるアイテム数が増加したこともアクセス数の増加に影響している可能性もある．しかし，公開データ数の増加が年に約 10 件ペースで安定していることに鑑みれば，アクセス数の増加が単純にアイテム数の増加のみに起因するものではなく，一般的な RUDA への認知拡大を反映したものだと理解できよう [4]．

3. 利用件数の推移

では，どれだけ RUDA は実際に二次分析・教育利用の利用環境として利用されてきたのだろうか．この点を確認するため，図 5 では，利用申請書の記載事項に基づき，利用目的別の利用件数の推移を示した [5]．

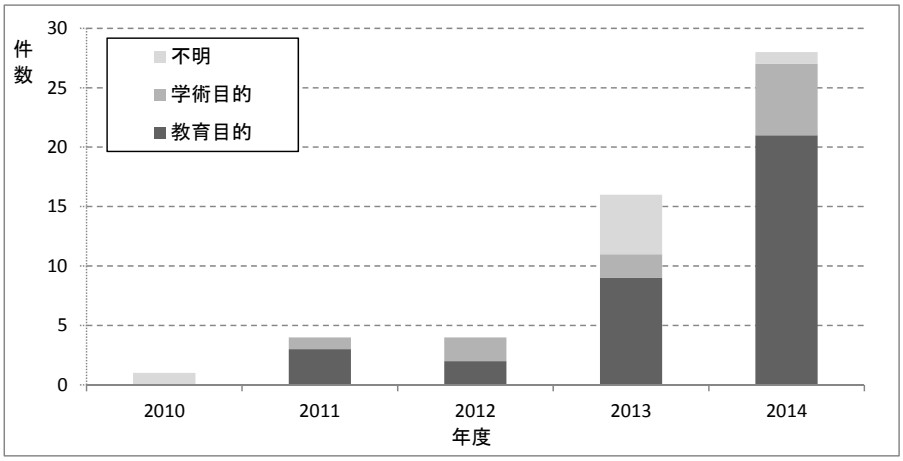

図 5　目的別利用件数の推移

図 5 に示されている通り，アクセス数と同様に利用件数についても，2013 年度以降の著しい増加が確認できる．2011･2012 年では等しく 4 件の利用件数にとどまっていたものが，2013 年度にはその 4 倍にあたる 16 件にまで利用件数が増加している．また，アクセス数が 2013 年から 2014 年にかけてあまり変化がなかったのに対して，利用件数は 2013 年から 2014 年（11 月末現在）にかけて増加しており，28 件の利用が確認できる．また，利用目的別に確認すれば，学術目的よりも教育目的のウェイトが高いこともあわせて確認できる．これらの点から，アクセス数に比して必ずしも数が多いわけではないが，RUDA は確かに社会調査データの提供環境としての役割を漸次的に実現しており，特に教育利用の機会提供において顕著であると判断できるだろう [6]．

Ⅳ　今後の展望に関する考察

前節では，RUDA の公開データ数・アイテムアクセス数・利用件数の時間的推移を確認し，それらの量的拡大から，確かに RUDA が社会調査データのアーカイブを進め，特に教育利用の機会を供する機関として活用されつつあることを確認した．これらの点から，RUDA は設立当初のモチーフに沿う形で着実にデータアーカイブとしての実績を積んでいると判断できるだろう．しかし，こうした過去の実績は，必ずしも将来的な発展を約束するものではない．本節では，国内のデータアーカイブ状況に鑑み，今後のデータアーカイブ事業の発展に向けた課題，およびその中での RUDA の課題について検討する．

1. 国内データアーカイブの現状

冒頭で述べたとおり，日本にはデータアーカイブが複数存在しており，それぞれ特色のもったデータの保管・公開を行っている．こうした分散的にデータアーカイブが存在する日本の状況は，利用機会の観点からは否定的に，国内全体でのデータアーカイブ運営から見た場合は肯定的に受け止めることができる．まず前者の点については，端的には，「どこのデータアーカイブにどのようなデータが存在しているのか」が必ずしも利用者に自明で

はないという点から，せっかくの二次利用・教育利用の需要に対して効率的に機会を供することが叶わないという問題として理解できる．換言すれば，潜在的なデータ利用を十全にくみ取ることができない非効率的な状況とも言えるだろう．

この問題に対しては，分散する国内のデータアーカイブをより少数のデータアーカイブに集約するという方法が考えられる．しかし，少数のデータアーカイブに集約させることは，その集約先のデータアーカイブの運営コストを高めることを意味する．前々節に紹介した RUDA の取組みで紹介したように，クリーニング作業などの寄託から公開に至るまでの作業，利用申請・利用報告の確認・記録，また，寄託・公開データの管理や，簡便なデータ提供を支えるシステム保全がデータアーカイブ事業の運営には必要な業務となる．これらの業務はデータ寄託・利用の増大とともに，負担量が増加することは想像に難くない．加えて，アーカイブシステムそのものに対する専門的知識，さらにその背後にあるメタデータ基準に関する理解といった専門知識も必要となる．だが，アーカイブ専門の常勤スタッフ・安定的な財源を確保して行われている海外の大規模データアーカイブと比較し，専門性を伴う業務の増加に対して国内データアーカイブの人的資源・財源は盤石ではないのが実情である（佐藤・佐藤 2006，佐藤 2012，中澤ら 2009）．この現実が変わらないのであれば，少数のデータアーカイブに集約させる試みは，ひるがえって円滑なデータアーカイブ運営の妨げになる危険性を持つ．この意味において，分散してデータアーカイブが存在する状況は，個々のデータアーカイブの運営コストを引き下げていることを意味するので，俯瞰的にみれば国内全体での円滑なデータアーカイブ事業を支えている状況であると肯定的に理解できる．

2. CESSDA の Data Catalogue

分散的に並列するデータアーカイブ状況が，円滑なデータアーカイブ事業に必要であるとの理解が正しいのであれば，「分散的にデータアーカイブが存在する状況を所与としつつも，いかに利用者の効率的なデータへのアクセスが可能になるのか」という問いを現実的に対応すべき課題として位置づけることができる．この課題に対して，我々は CESSDA（Consortium of European Social Science Data Archives）の取組みが参考になると考えている．

CESSDA は，（2014 年 11 月現在）ノルウェーの NSD やイギリス UKDA，ドイツ GESIS といった 13 か国のナショナル・データアーカイブ機関の連合体であり，国や言語，リポジトリの垣根を越えたシームレスなデータアクセスを一つの目的として標榜する機関である．CESSDA は Data Catalogue という検索システムを用いて加盟データアーカイブにそれぞれ所蔵されているデータ情報を共有し，利用者は用語，ないしはブラウザからトピック，キーワード，または配布アーカイブごとに調査データとその調査で用いられている変数を検索することができる（前田 2011）．利用者の希望に合致するデータがあった場合，その検索結果からそのデータが所蔵されている各国のデータアーカイブに移行することができる．

このように CESSDA の Data Catalogue は並列的に存在するヨーロッパのデータアーカイブと利用者との間を一つのプラットフォームで連結するものであり，並列的なデータアーカイブの存在を所与としつつも，利用者は効率的に目的に応じたデータの検索，そしてデータ利用が可能となる．

3. DDI について

CESSDA の Data Catalogue に範を求めるのであれば，まず日本に存在するデータアーカイブが取り組むべき課題として，一元的な検索体系の構築に向けた現在の公開データに関するメタデータ情報の標準化が挙げられる．というのも，メタデータは「記述対象となる情報資源に関して，決められた属性についてその属性値を書き表したもの」（杉本 2009：33）であるが，その「決められた属性」が用いるメタデータ基準で異なりうるため，メタデータ基準の標準化を行わない限り，メタデータの相互運用性が損なわれる危険性があるためである．

では，社会調査データのアーカイブにどのメタデータ基準が最適なのだろうか．メタデータ基準には，現在 RUDA でも利用している Dublin Core のように特定の分野に特化していない汎用性の高いメタデータ基準もあれば，例えば生物学データアーカイブに用いられる Darwin Core のような専門分野に特化したメタデータ基準もある．こうした様々なメタデータ基準のなかでも，現在，CESSDA をはじめとした多くの海外の社会科学系データアーカイブで使用されているのが，DDI（Data Documentation Initiative）と呼ばれるメタデータ基準である．

DDI は社会調査データに特化したメタデータ基準で，社会調査データに関わるメタデータ情報を詳細なレベルで包括的に記述する枠組みをもつ[7]．xml を用いることで，特にインターネット上でのデータの検索や情報の相互参照，さまざまな目的に合わせた出力を効率的に行うことを念頭に設計されており（前田 2011），CESSDA の Data Catalog は DDI の特性を生かした例といえる．Data Catalog と同じく CESSDA で開発された Nesster システムは DDI ベースの CMS であり，データアーカイブの所有するデータの表示や検索だけでなく，リモート集計機能を有している．

今日，DDI は CESSDA に加盟しているデータアーカイブのみならず，アメリカの ICPSR や，韓国の KOSSDA，台湾の SRDA でも導入されており，事実上，社会調査データにおけるメタデータ基準の国際規格となりつつある．国内でも SSJDA が先鞭をきって導入しており，SSJDA では一部データを Nesstar にて公開し，リモート集計を導入している．また，DDI 規格のメタデータ編集ソフトの EDO（Easy DDI Organizer）を先日公開し，国内での DDI の普及に努めている[8]．

4. RUDA の今後の取組み

こうした社会調査データアーカイブに特化した DDI の理念，および国内での動き，さらには国際的な動向に鑑みれば，「どのメタデータ基準を用いるのか」については DDI がその最有力の候補に挙げられる．しかし，ここで問題になるのが，「どの水準で DDI を導入するのか」についてである．

この課題に対する一つの案として，CESSDA や SSJDA のようにシステムレベルで Nesstar を導入する方法が挙げられる．だが，確かに Nesstar は Publisher・Server・Viewer を有機的に連結することで一元的な管理・公開，さらにはリモート集計を可能にできる魅力的なシステムであるが，調査データそのものをサーバ上に置いてしまうことになるのでセキュリティ上のリスクが存在する．また，各アーカイブで既存の安定的に運営しているシステムが存在するのに関わらず，それをシステムレベルで DDI ベースのそれに移行する

のは現実的に難しく，コストもかかる [9]．あくまで，データ検索環境の整備を目的として DDI を用いるのであれば，システムそのものを変える必要はなく，現在それぞれのアーカイブで保管しているデータの DDI 基準による xml でのメタデータを作成し，それと検索エンジンを用意すればブラウザベースでの検索が容易に実現する．

この DDI 基準のメタデータ作成方針に従うのであれば，RUDA としての今後の課題は，現在用いている Qualified Dublin Core で管理しているメタデータを，DDI 基準のそれに移し替えることが挙げられる．すでに Dublin Core の基準項目と DDI での要素へのマッピングに関しては，公式に「推奨」という形で提示されているが [10]，第 1 章で述べたとおり，RUDA では限定子を用いているため，既存の推奨マッピングでは，現実的に齟齬が発生する．こうした問題に鑑み，現在，RUDA では設立当初に企図とされていた記入要項と Dublin Core の限定子との関係を踏まえ，付表に示した DDI（codebook 2.5）ベースに対応するマッピング案を策定した．このマッピング案を踏まえて RUDA が管理するメタデータの DDI への移し替え，そして DDI に根ざした検索システムの構築が今後の課題になる．

5. 国内アーカイブの連携に向けて

前小節までの議論で，「分散的にデータアーカイブが存在する状況を所与としつつも，いかに利用者の効率的なデータへのアクセスが可能になるのか」という問いに対して，DDI 基準のメタデータ作成，そしてそれに基づくデータ検索環境の整備を提案し，この方針に従う今後の RUDA の具体的な取組みを提示した．しかし，RUDA 単体の取組みだけでは，効率的なデータ検索が行える環境を整備することは不可能で，他のデータアーカイブとも連携して行う必要がある．

これまでにもその必要性が繰り返し喚起されてきたデータアーカイブ間の連携だが（真鍋 2012，今田 2015），アーカイブ間でのメタデータ共有はその第一歩として理解できるだろう．しかし，それぞれの特色を生かした運営を行っている日本のデータアーカイブ状況においては，具体的にそれぞれのアーカイブで必要となる情報が異なるために，たとえメタデータ共有の重要性は理解していたとしても，それを実現するための具体的なスキームが共有されづらいという環境にあった．この問題に対して，メタデータ共有を企図とする DDI の導入は，解決に向けた契機になりえる．というのも，社会調査データを包括的，かつ詳細に記述することができる DDI を用いることで，個々のアーカイブで利用している独自のメタデータを共通のフレームで記述することができるからである．

しかし，この豊富な情報を表現できる DDI の導入は，翻って「何が社会調査データの記述に重要なのか」という困惑を与える可能性がある．事実，DDI を導入している海外アーカイブ間でも記述の様式は様々であり，DDI が用意している項目を全て使用しているアーカイブは存在しない．それぞれのアーカイブが培ってきた経験をもとに独自のメタデータを用いているのが現状である．だが，DDI という共通の枠組みを導入し，上記の困惑が表出したとしても，それは（DDI を基準としたときの）データアーカイブ間での相違が表出したと理解でき，その相違こそがデータアーカイブ間での連携に向けて調整すべき一つの問題であると理解できる．そして，その相違は DDI の枠組みのなかにおいて明確に表すことができる．この意味で，DDI の導入は，連携という視点からアーカイブ間で討議するための共通の土壌に他ならない．それゆえ，少なくとも共通の参照点のないことによる散逸

した議論，そして具体的な取り組みまで落とし込めない大言壮語的な結論に終わることなく，より具体的で建設的なアーカイブ間連携に向けた議論を行うことができるだろう．

このように，DDI の導入は，データアーカイブ間での連携にも寄与する．加えて，アーカイブ間の事業提携や業務の規格化の機会をもたらし，アーカイブ運営のコストを下げる働き [11]や，データの寄託者にとってデータを寄託がしやすいプラットフォームの基盤，さらには将来的に社会調査の実施環境を整備することも期待できる．このように今後のデータアーカイブの持続的な発展において，ひいては社会調査全体にまで波及しうる DDI のメリットは大きく，RUDA としてもその発展に主体的に関与していく予定である．

謝辞

本稿の内容は，立教大学データアーカイブ RUDA の設立から今日に至る約 6 年間の歩みについて綴ったものです．この期間，RUDA の設立・運営に尽力し，今日のアーカイブ業務の基盤を整備していただいた社会調査部会メンバーのみなさまに深く感謝します．

注

1)　持続的なデータアーカイブ事業を行なうためには，「データを寄託してもらう」という受動的な姿勢ではなく，データアーカイブが現在進行中の社会調査プロジェクトに関わる必要がある．例えば，アメリカの ICPSR が標榜する「データライフサイクル」は，調査プロジェクトの立案からデータ寄託までの一貫した社会調査プロセスを提唱したモデルだが，データアーカイブ運営者はこのプロセスの鼻緒からアドバイスを行う役割で介入することが明記されている（ICPSR 2012 : 8）．また，寄託者との関係だけではなく，例えば GESIS の GESIS Spring seminar や，ICPSR の summer seminar，SSJDA の計量分析セミナーに代表されるような統計教育の機会，さらにはコンサルティングの実施などを通じて，データ利用者の統計的知識・スキルを涵養することも，広くはデータアーカイブ（リサーチ・データ・センター）の機能であることが議論されている（ヤゴチンスキー 2012）．

2)　http://dublincore.org/（最終アクセス日：2015 年 1 月 30 日）

3）学内公開を始めた 2010 年時点では，約 1,300 件のアクセス数が確認できるが，一般公開を始めた 2011 年には，約 1,000 件へとアクセス数が減少している．この背景には，学外公開に向けたシステムの動作確認といった目的によるアクセス数が，2010 年時点では混在していたことが存在している．

4)　RUDA の認知拡大の背景には，現在，RUDA が行っている広報活動の存在が挙げられる．RUDA では日本語・英語版のリーフレットを作成し，量的調査・統計分析に関連のある学会での普及活動を行っている．これまでにリーフレットを配布した学会として，日本社会学会，数理社会学会，International Sociological Association Annual Conference などがある．

5)　ただし，ここでの数値は新規利用申請のみをカウントしており，利用延長の分は省いている．また，利用申請書に利用目的が正しく記入されていないケースについては，利用計画の内容に鑑みて利用目的を判断したが，一部には判断ができないケースがあったため，それらについては不明として処理をしている．

6)　利用申請書の身分を，学生（学部生・大学院生）と一般（研究機関に属する，ないしは関連する研究者）に分類し，利用件数の分布を確認したところ，2013 年度以降では，一定数の学生身分による利用申請が確認された．こうした学生身分の利用者からの活用も，RUDA の普及を裏付ける一つの事実として理解できよう．

7)　文中では説明を省いたが，DDI はヴァージョンによって規格が異なるために導入の際

には注意が必要である．ヴァージョン 1 と 2 ではインターネット上でのデータの検索や情報の相互参照を効率的に行うことを念頭に設計されているが，ヴァージョン 3 からは ICPSR が標榜する「データライフサイクル」（注 1 参照）を念頭に設計されている．そのため，社会調査を行う研究者自身が社会調査立ち上げ期からの計画から調査実施，さらにはデータ分析から寄託・公開といった一連の流れを網羅的に記述することを前提として設計されており，各段階での様々な要項を xml の要素として特定的に記述することができる．ICPSR やドイツの GESIS などの大きなアーカイブではヴァージョン 2 と 3 の両方を記載している．

8)　このエディタはヴァージョン 3 に対応したソフトであり，将来のデータ寄託者となる社会調査を学ぶ学生や若い研究者を対象にしている．調査項目を入力すると，word や pdf 形式で調査票を作成する機能を有している．DDI のメリットを具現化するためには，データアーカイブが DDI を導入するだけでなく，寄託者・利用者レベルで DDI が普及し，調査設計に活用される環境が必要になると思われるが，このソフトはその環境作りに貢献すると思われる．なお，利用法については佐藤・米倉（2011）を参照．

9)　加えて，「どの水準で Nesstar を導入するのか」という決定の必要性もコストの一部になる．事実，Nesstar の利用法については CESSDA に加盟するアーカイブ内でも違いがあり，例えばフィンランドの FSD ではメタデータ情報のみを公開し，データそのものは Nesstar 上に置いていない(朝岡 2011)．また Nesstar はヴァージョン 1 に対応したシステムであり，この点も導入コストの一部として理解できる．

10）http://www.ddialliance.org/resources/tools/dc を参照（最終アクセス日：2015 年 1 月 30 日）

11)　DDI をベースにした Nesstar もデータアーカイブのオンライン化のコストを下げる目的に作成されたものである．現在，アフリカの公的機関を中心にリモート集計がオミットされた Micro data Tool Kit という無償の CMS が利用されている(朝岡 2011)．

参考文献

朝岡誠，2011,「Nesstar の特徴と海外データアーカイブにおける利用状況」前田幸雄・佐藤慶一・安藤理・米倉祐貴・朝岡誠・入山浩一・角井祐『Data Documentation Initiative の利用可能性』SSJDA リサーチペーパーシリーズ 46：88-98.

今田高俊，2015,「調査科学リサーチ・コモンズの構築」統計数理研究所・調査科学研究センター・シンポジウム「調査科学リサーチ・コモンズの構築に向けて」基調講演資料.

Inter-university Consortium for Political and Social Research，2012，*Guide to Social Science Data Preparation and Archiving – Best Practice Through the Data Life Cycle [5th edition].* Michigan: Institute for Social Research University of Michigan.

稲葉昭英, 2000,「公開データ利用型の調査教育の勧め」佐藤博樹・石田浩・池田謙一編　『社会調査の公開データ　2 次分析への招待』東京大学出版会，35-50.

前田幸男，2011,「情報技術と統計メタデータ：DDI についての概観」前田幸雄・佐藤慶一・安藤理・米倉祐貴・朝岡誠・入山浩一・角井祐『Data Documentation Initiative の利用可能性』SSJDA リサーチペーパーシリーズ 46：40-51.

真鍋一史, 2012,「社会科学はデータアーカイブに何を求めているか」『社会と調査』8:18-23.

松本康，2012,「立教大学データ・アーカイブ RUDA の始動」『社会と調査』8:24-30.

中野康人・小松洋, 2003,「データの作成・公開と実査時の注意点」『理論と方法』18(2):237-51.

中澤秀雄・西城戸誠・大國充彦・新國三千代・祐成保志・新藤慶・小内純子・高橋徹，2009,「「社会調査のアーカイブズ学」の必要性―札幌学院大学 SORD が取り組んだ「夕張調査資料集成」作成経験からの提言―」『理論と方法』24(1):121-8.

大谷信介・木下栄二・後藤範章・小松洋・永野武編，2005,『社会調査へのアプローチ――論理と方法　第 2 版』ミネルヴァ書房.

佐藤博樹，2012,「実証研究におけるデータアーカイブの役割と課題 ：SSJ データアーカイブの活動実績を踏まえて」『フォーラム現代社会学』11: 103-12.
佐藤朋彦・佐藤博樹，2006,「データアーカイブの役割と SSJ データアーカイブの現状――実証研究における再現性を担保するために」『日本労働研究雑誌』48(6):42-54.
杉本重雄，2009,「Dublin Core の現在」『ディジタル図書館』36:32-45.
轟亮・杉野勇，2013,『入門・社会調査法――2 ステップで基礎から学ぶ　第 2 版』法律文化社.
ヤゴンチンスキー・ヴォルフガング，2012,「社会科学のためのインフラストラクチャ―の基盤としてのリサーチ・データ・センター」『社会と調査』8:5-15.
米倉佑貴・佐藤慶一，2011,「社会調査メタデータ管理ソフト Easy DDI Organizer（EDO）の設計」前田幸雄・佐藤慶一・安藤理・米倉祐貴・朝岡誠・入山浩一・角井祐『Data Documentation Initiative の利用可能性』SSJDA リサーチペーパーシリーズ 46：69-87.

付表 マッピング案

Dublin Core		意図	DDI 2.5 element
element	qualifier		
identifier	StudyNo	調査番号	IDNo
title	display	調査名	titl
title	alternative	調査略称	altTitl
creator		調査主体 / 調査代表者	AuthEnty
description	investigator	調査実施者	producer
description	sponsorship	調査資金	fundAg
publisher	depositor	寄託者	depositr
relation	ispartofseries	シリーズ	serName
publisher		出版社または出版者	※1
type		資料種別	dataKind
description	abstract	調査概要	abstract
subject		キーワード	keyword
description	discipline	研究分野	topcClas
coverage		母集団	universe
coverage	temporal	調査時期	timePrd
coverage	spatial	調査地域	geogCover
description	SamplingProcedure	標本抽出法	sampProc
description	SampleSize	標本サイズ	SampleSize
description	NoOfValidResponses	有効回収数	※2
description	ResponseRate	有効回収率	respRate
description	UnitOfObservation	観察単位	anlyUnit
description	ModeOfCollection	調査方法	collMode
description	variables	主要変数	※1
relation		関連資料	othrStdyMat のいずれかの要素
rights	AccessRights	利用条件	conditions
identifier	uri	URI	holdings の url
date	issued	公開日	prodDate
identifier	version	ファイルのバージョン	version
date	modified	更新履歴	※1
identifier	citation	引用時の表記	citReq
description	note	備考	notes
publisher	distributor	配布者	distrbtr

※1　対応する要素が確認できなかったため割愛.
※2　標本サイズと有効回収率から計算できるため割愛.