

Data Documentation Initiative の利用可能性

前田幸男 佐藤慶一 安藤理 米倉佑貴
朝岡誠 入山浩一 角井佑

SSJDA-46

May 2011

Data Documentation Initiative の利用可能性

—目次—

第1章	SSJDA の情報システムの状況と DDI 利用可能性検討の経緯	3
		佐藤慶一
第2章	情報技術と統計メタデータ：DDI についての概観	40
		前田幸男
第3章	DDI の技術的側面，利用イメージ，導入ステップ	52
		佐藤慶一 米倉佑貴
第4章	社会調査メタデータ管理ソフト Easy DDI Organizer (EDO) の設計	69
		米倉佑貴 佐藤慶一
第5章	Nesstar の特徴と海外データアーカイブにおける利用状況	88
		朝岡誠
第6章	SSJDA における Nesstar 試験導入	99
		入山浩一 角井佑 米倉佑貴 佐藤慶一
終章	Data Documentation Initiative の利用可能性	135
		佐藤慶一
資料編	東アジアのアーカイブにおける Nesstar の利用	141
		安藤理

はじめに

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は、共同利用・共同研究拠点として文部科学省から認定を受け、2010年4月より活動を開始した。

拠点としての活動の一環として、「データのより開かれた利用を促進するため、Web上でのダウンロードによるデータ提供や、DDI (Data Documentation Initiative) 対応による国際標準化などデータファイル形式の多様化を進めていく。また、近年、設置が進められている国内の他のデータ提供機関との連携のありかたについても検討していく」ことになっている。

本レポートは、上記の計画に対する2009年度後半から2010年度における検討状況を取りまとめたものである。DDIは現在進行形のプロジェクトで情報量も多く、しかしながら日本国内における取組みはほとんど見られない。SSJDAにおける検討は手探りの状態から始まったが、DDIに関する情報整理から、メタデータ管理ソフトEDOの基本設計と、最も普及しているDDI対応情報システムであるNesstarの導入実験まで、基礎的な検討を進めることができた。

下記検討会を組織し、2010年度後半を中心として情報交換や議論を行い、本レポートを取りまとめた。なお、各章において表明されている見解はあくまで各執筆者のものであり、社会調査・データアーカイブ研究センターの公式見解ではないことをお断りしておく。

DDI利用可能性検討会（肩書きは2011年5月9日現在）

前田幸男 東京大学社会科学研究所准教授
佐藤慶一 東京大学社会科学研究所准教授
安藤 理 東京大学社会科学研究所助教
米倉佑貴 東京大学社会科学研究所特任研究員
朝岡 誠 東京大学社会科学研究所特任研究員
入山浩一 東京大学社会科学研究所学術支援専門職員
角井 佑 前 東京大学社会科学研究所特任研究員（株式会社セキュアソフト技術本部
技術サポートグループ課長）

佐藤慶一・前田幸男

第1章 SSJDA の情報システムの状況と DDI 利用可能性検討の経緯

佐藤慶一

1. SSJDA の情報システムの状況

1.1 SSJDA の情報システムの概要

東京大学社会科学研究所では、1998年よりSSJデータアーカイブ（略称SSJDA）を運用し、社会調査の個票データ等を収集・保管、学術目的での利用への提供を行っている。活動を支える情報システムについては、適宜、構築や改変が積み重ねられ、2011年5月現在、図1に示すようなシステム構成となっている。

SSJDAでは、調査者（機関・研究者）による社会調査の個票データを、担当者を通じて寄託いただくことから活動がはじまる。通常、調査者の一次分析が終了した後に、寄託いただくこととなる。アーカイブに寄託された個票データは、一次分析結果との照合、関連メタデータの整備、必要な秘匿処理等を行なった上で、二次利用可能な形式に整理する。それらのデータアーカイブ室での作業進捗はエクセル表で管理されている。作成した公開データは、寄託者に確認いただいた上で、公開の手続きがされる。

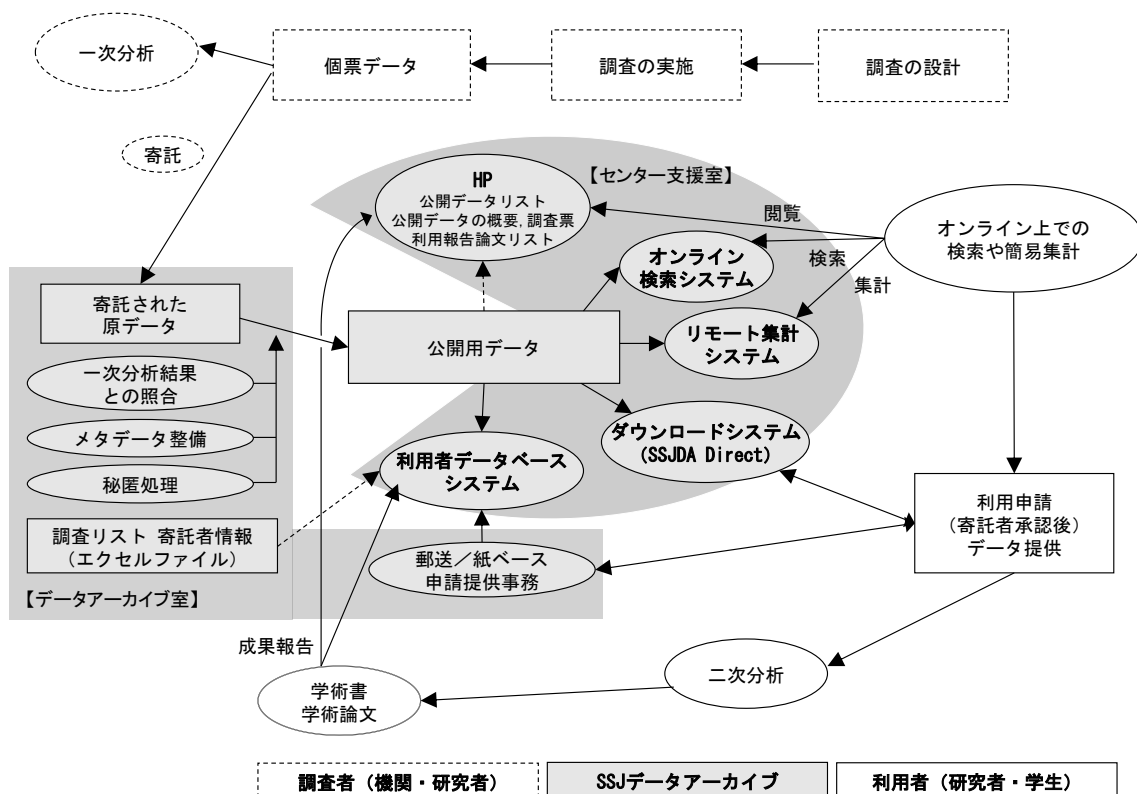


図 1 SSJDA の情報システムの概要

公開されたデータは、ホームページトップページでアナウンスされると同時に、公開データリストに掲載される。調査の概要や調査票様式も掲載され、オンライン検索システムで検索を可能とする。寄託者から許諾いただいた一部データは、リモート集計システムに掲載され、オンライン上でのデータ分析を可能とする。また、利用回数が多い一部データは、オンライン上で利用申請、データ入手、利用報告が可能な SSJDA Direct に掲載される。

データの提供は、提出された利用申請の内容をデータアーカイブが審査し、寄託者の承認を得た上で行われる。利用終了後は、データアーカイブに利用報告をしていただき、論文等を発表した場合は2部郵送いただき、1部はアーカイブで保管し、1部は寄託者へ郵送する。以上の利用に関する情報は、寄託者情報と併せて利用者データベースシステムで管理されている。

1.2 調査リストと寄託者情報

アーカイブに寄託された個票データについて、二次利用可能な形式に整理し、寄託者に確認いただいた上で、公開の手続きがされるまでの作業は、データアーカイブ室で行われている。調査リストやそれぞれの作業進捗についての情報は、エクセル表で管理されている。このエクセル表については、データが寄託され調査リストが更新されるごとにバックアップが取られている。過去のものを見ると、2004年5月のものが最も古く、微細な変更はあるものの、その頃より同様の形式で、これまでデータ公開準備作業の情報が蓄積されている。内容は、大きく、基礎情報、寄託情報、作業情報、公開情報の4つに分かれる。基礎情報には、調査番号と寄託日、寄託者名、調査名がある。寄託情報には、寄託書の情報（提供方法、利用者リストの年間提出回数、教育目的や学部生利用の可否、報告書の部数や複写の取り扱いなど）、調査票数や個票数がある。作業情報には、調査概要や集計表などメタデータ整備作業の担当者と作業依頼日、納品日、同じく SPSS データの集計チェックや秘匿処理の担当者と作業依頼日、納品日がある。一通り公開用のデータ整備が済んだら、校正担当がチェックをした後、データアーカイブ室担当者で集まり、最終確認作業が行なわれている。その後、寄託者への公開データの確認をして寄託者からの許諾をいただいた後、公開に至る。その間の公開に関する情報として、寄託者確認依頼の送付日や寄託者によるチェックが完了した日付、公開した日付が管理されている。

寄託者情報についても、担当助教によりエクセルで管理されているが、同様の情報を後述する利用者データベースシステムのためのアクセスデータベースでも管理しており、その集約が課題となっている。

1.3 ウェブサイト

(1) 経緯

SSJDA では、2002年8月にウェブサイトをリニューアルし（槇田 2006）、このサイトが2007年10月までは運用されていた。入れ替え直前の状態のトップページを図2、図3に示

す。トップページ上部に「個票データのプライバシー保護」「誓約事項」「担当スタッフ」へのリンクがあり、その下に、「個票データの寄託」「個票データの利用」「データ検索」「データアーカイブについて」「お知らせ」「新着情報」「リンク集」があり、さらに下部に「謝辞」が掲載されており、現在のサイトを構成する主要な情報が整っていたことが確認できる。情報の作成や編集は、当時 IBM が販売していたホームページ・ビルダーという WYSIWYG 形式の Web オーサリングツールが用いられていた。


社会科学研究所 附属日本社会研究情報センター
TO ENGLISH SITE


SSJデータアーカイブ
ICPSR
ICPSR国内利用協議会

日本の社会科学研究のためのデータ資源

SSJデータアーカイブでは、我が国の社会科学分野の研究活動を支援するために、統計調査、社会調査の個票データと調査に関する情報を収集、保存し、それらを学術目的での二次分析のために提供します。

調査データ寄託へのご協力をお願いするとともに、皆様の積極的なご利用をお待ちしております。

[個票データのプライバシー保護](#)
[誓約事項](#)
… 利用者が寄託者に報いるために


 当データアーカイブへの問い合わせは[こちら](#)をお読みください。

データアーカイブ担当スタッフ

<p>個票データの寄託 調査実施機関の方へ</p> <p>○寄託のメリット</p> <p>○寄託の方法</p> <p>FAQ</p>	<p style="text-align: center;">○データの検索 ここから当アーカイブ所蔵データを検索できます。</p> <p><お知らせ></p> <p>2007/10/02</p> <ul style="list-style-type: none"> SSJDAのウェブサイト更新作業を行う関係で、本日より数日間、接続しにくくなる場合があります。ご迷惑をおかけしますが、よろしくお願いたします。 <p>2007/07/25</p> <ul style="list-style-type: none"> <大切なお願い> 現在、申請時のCDの破損が目立っております。また粗悪な質のCDをお送りいただくと、データの焼き付けができません。CDは傷のつかないよう、ケースに入れて、クッションつきの封筒でお送りください。 <p><新着情報> (→更新履歴)</p> <p>2007/09/27</p> <ul style="list-style-type: none"> 0552「連合生活アンケート調査, 2002」(日本労働組合総連合会)を公開しました。 <p>2007/09/26</p> <ul style="list-style-type: none"> 0536「管理者を対象とした両立支援策に関する意識調査, 2005」(内閣府男女共同参画局)を公開しました。 0537「男女の働き方と仕事と生活の調和(ワーク・ライフ・バラ 	<p>個票データの利用 ↑ 初めにお読み下さい 大学又は研究機関の研究者、大学院生(学部学生)の方へ</p> <p>○利用申請書</p> <p>○利用報告書</p> <p>FAQ</p> <p>ISSPの利用について</p> <p>=>リモート集計システム ※一度にアクセスが集中するとシステムが動かなくなりま す。ご注意ください。</p>
<p>データアーカイブについて</p> <p>意義 構築の方針 諸外国のDA 統計 手続きの流れイメージ</p> <p></p> <p>ニュースとイベント</p> <p>*SSJデータアーカイブのデー</p>	<p><リンク集></p> <p>アーカイブ、調査、機関などの関連リソース</p> <p>国内</p> <p>日本版General Social Surveys JGSS</p> <p></p>	

図 2 2002～07 年まで運用されていたウェブサイト 1

文 を利用して作成された論文 *SSJデータアーカイブ優秀論文表彰 *二次分析研究会 *研究報告会・公開セミナー リサーチペーパー・シリーズ テキストなど 社会調査の公開データ From "SSJ Newsletter" 社会科学データの準備と保存の手引き NEW!!	ンス)に関する調査、2006」(内閣府男女共同参画局)を公開しました。 2007/09/20 <ul style="list-style-type: none"> 0540「小企業の金融機関借入に関する調査、2006」(国民生活金融公庫総合研究所)を公開しました。 2007/09/18 <ul style="list-style-type: none"> a179「モノグラフ小学生ナウ 小学生にとっての「勉強」, 2003」(ベネッセコーポレーション)を公開しました。 	... 国内リンクをもっと見る 社研パネル調査プロジェクト 海外 ICPSR / ICPSR国内利用協議会 ... 海外リンクをもっと見る
---	---	--

謝辞

当アーカイブの趣旨をご理解いただき、個票データを寄託いただいた下記機関等の皆様に厚くお礼申し上げます。

21世紀職業財団、CNEP日本チーム、アジア経済研究所、丸川知雄、ドイツ日本研究所・U・マーワルト、第一生命経済研究所、リクルートワークス研究所、関西生産性本部、高齢・障害者雇用支援機構、国民生活金融公庫総合研究所、司法制度改革審議会、社会経済生産性本部、生命保険文化センター、選挙とデモクラシー研究会、全国大学生生活協同組合連合会、中小企業総合研究機構、大阪商業大学比較地域研究所、日本家族社会学会、日本経済新聞社、日本能率協会経営革新研究所、日本能率協会総合研究所、労働政策研究・研修機構、連合総合生活開発研究所、中小企業経営者の実態に関する調査研究会、日本リサーチセンター、野村総合研究所、ニッセイ基礎研究所、旧東京都立労働研究所、三菱総合研究所、法政大学大原社会問題研究所、人材ニーズ調査専門委員会、電機連合総合研究企画室、雇用・能力開発機構、エンジニアリング振興協会、1995年SSM調査研究会、石黒格、日本フランチャイズチェーン協会、エンタテインメントビジネス総合研究所、ベネッセコーポレーション、日本人材派遣協会、日本労働組合総連合会、情報サービス産業協会、日本リサーチ総合研究所、アデコ株式会社、(株)アイデム 人と仕事研究所、家計経済研究所、経済同友会、猪口孝、全国中小企業団体中央会、天野郁夫、三宅一郎、大木栄一、中村圭介、佐藤博樹、藤田英典、(株)日本総合研究所、中小企業基盤整備機構、東京都老人総合研究所、ミシガン大学、三菱UFJリサーチ&コンサルティング、産業雇用安定センター、社会安全研究財団、韓オーエムエムジー(社)新情報センター、経済産業省、内閣府男女共同参画局、JESⅢ研究会、商工総合研究所、介護労働安定センター、2003年階層調査研究会、早稲田大学21世紀COE「開かれた政治経済制度の構築」、内閣府男女共同参画局

(敬称略)

当アーカイブの構築に当たっては、平成9(1997)年度から18(2006)年度にかけて文部科学省・独立行政法人日本学術振興会科学研究費補助金(研究成果公開促進費)の交付を受けました(平成11年度502002、平成12年度128051、平成13年度138045、平成14年度148052、平成15年度158053、平成16年度168052、平成17年度178050、平成18年度188056)。

当アーカイブは、2001年に平成13年度統計活動奨励賞(財団法人 日本統計協会)、2006年に第2回日本統計学会統計

ファイル形式

本サイトで提供する各種申請様式には、Microsoft Word 及び Adobe Portable Document Format を使用しています。Microsoftは、米国 Microsoft Corporationの米国及びその他の国における登録商標です。Acrobat Readerは、米国Adobe Systems Corporationの商標です。



SSJデータアーカイブが提供する個票データは、基本的には 日本語版SPSS(エス・ピー・エス・エス株式会社)によるsav形式を使用しています(もっと詳しく)

[社会科学研究所に戻る](#)

All Rights Reserved, Copyright (c)2002-2007 Information Center for Social Science Research on Japan, Institute of Social Science, The University of Tokyo

図 3 2002～07年まで運用されていたウェブサイト2

ホームページ・ビルダーで編集する形式で、概要や調査票の html ファイルも整えられている。図 4 に概要の html ファイルを示す。「調査番号」「調査名」「寄託者」「利用方法」「調査の概要」「調査対象」「データ数」「調査時点」「調査地域」「標本抽出」「調査方法」「調査実施者」「委託者」「報告書」「関連論文等」「主要調査事項」「調査データ区分」「利用上の注意」といったメタ情報が整理されており、2011年5月現在まで一貫して同じ形式でメタ

情報を蓄積している。概要を英訳した abstract の html ファイルも同様にひな形が用意されているが、2011年4月現在、236調査について英訳が完了した状態で、公開されている全ての調査概要の英訳は完了していない。調査票の html ファイルは、一部調査を除いて、集計結果も含めて独自に作成をしていたが、その編集作業が膨大であった。2010年度より、データアーカイブ室での作業量軽減のため、寄託された調査票をそのまま PDF 化してウェブに掲載する方式へと変更された。集計結果が掲載されていない調査票が寄託された場合は、エクセルで簡易な集計表を作成して PDF 化し掲載することとしている。

[概要][調査票][Abstract][検索]	
概要	
調査番号	0001
調査名	新規卒業者(中卒)労働市場調査, 1963
寄託者	東京大学社会科学研究所
利用方法	利用申込先: SSJデータアーカイブ SSJデータアーカイブが利用申請を承認したとき利用できる。 ■教育目的(授業など)の利用: 不可 ■学術目的(卒業論文など)の利用: 不可
調査の概要	<p>新制中学校卒業生と日本の産業の現在および将来がいかにかわり合っていくかという問題を、「一因の資本の蓄積量」と「方向」、産業と学校・技能教育の関係、社会階層の特殊な生活水準と生活慣習、とくに農民層の分科、「労働市場の職業労働化」、「未就業失業問題と低賃金産業問題」=「産業予備軍の存在問題」の観点から分析し、産業別・規模別・地域別労働市場の類型化をおこなう。</p> <p>調査の結論は次のとおりである。神奈川県中卒(男子)労働力の市場は、3階層・5地域—A、最上層=「全県をおおむね京浜地区労働市場」、B、第2階層=「横浜地区」、「西湘地区」を中心とする中部地区、小田原市を中心とする「西部地区」、C、最下層=高層・変質・途次非即にまたがる「農村地区」—からなる構造をもち、労働移動は、各地域市場間では主として3階層間で—G→A、B→Aという「連鎖型」の2型をとりこなしおこなわれる。こうした「労働移動の連鎖」になるのは、賃金に集中的に表現される労働条件の相対的格差であって、その背景に、地域別の産業構造、企業規模別構造の相違がある。女子中卒者の労働市場は、「地域的好適性」が大きく、京浜地区労働市場も全県をおおむねおこなう点で男子のそれと異なる。また、いずれの地域でも女子の初任給は男子より低く、地域別格差は男子より大きく、分科もまた大きい。</p>
調査対象	神奈川県立中学校1963年3月卒業生全員(188校男約20,000名、女約19,000名)
データ数	「調査対象」参照
調査時点	1963年7月
調査地域	神奈川県
原簿抽出	
調査方法	神奈川県および神奈川県教育委員会の名で各学校に調査票〔甲・乙・丙・丁〕を配布し、学校ごとに就職指導教員に記入を依頼(1963年10月31日)し、11月30日に回収した。さらに「個人別調査票」をもって、若干の地域で、生徒の自宅を訪問し、面接、記入した。
調査実施者	労働問題調査研究会(大内一男、氏原正治郎、江口英一、関谷耕一、高梨昌)
委託者(経費)	文部省科学研究費、神奈川県企画審議会専門委員調査費
報告書(論文)	<p>神奈川県編、氏原正治郎、江口英一、高梨昌、関谷耕一編『労働力需給調査報告書』(神奈川県, 1965年3月, 418頁, 非売品)</p> <p>大内一男、氏原正治郎編、江口英一、氏原正治郎、高梨昌、関谷耕一編『労働市場の研究——中学校卒業生の就職問題』(1965年9月, 東京大学出版会, 411頁)</p> <p>神奈川県企画審議会編『昭和28年新制中学校卒業生の就職状況調査(集計結果表紙巻)』(1964年2月, 非売品)</p>
関連論文	<p>氏原正治郎、中西洋編『神奈川県における新規卒業者労働市場の変遷』(『社会科学研究』14巻3・4合併号, 1962年12月所収)</p> <p>中西洋編『労働市場の研究』について(『社会科学研究』第22巻5・8合併号, 1971年)</p>
主要調査事項	出身地域、生業の職業、進学就職状況、就職先(産業・規模・地域・労働条件)など
調査データ区分	戦後日本の労働調査、登録: 1998年4月1日
利用上の注意	<p>保存されていた調査票から再入力した調査データを公開している。</p> <p>次の熟字・選択枝による誤問については、提供する個票データにはありません。ご了承ください: 乙票 (4)2、(1)12、丙票 (4)2、丁票 3</p>

図 4 調査概要.html

2007年10月より、情報構成を整理してメニューを設け、デザインを一新しCSS(Cascading Style Sheets)を用いたサイトが用いられた。CSSは、文書の構造と体裁を分離させたスタイルシートの仕様の一つで、複数のページのデザインをそれぞれに編集するのではなく、体裁

を管理する部分を変更すれば統一的にデザインやレイアウトがコントロールできる。このサイトが2010年3月まで運用されており、入れ替え直前の状態のトップページを図5に示す。管理は、アドビシステムズが販売しているDreamweaverというWebオーサリングツールが用いられていた。ホームページ・ビルダーに比べ、高機能なWebオーサリングツールで、CSSの編集機能も強化されているものであった。しかしながら、ニュースの更新が手動であったり、Dreamweaverが専門業界向けで使い勝手が必ずしも簡便でなく、ページの維持管理更新に必要な作業が煩雑であるという問題があった。

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

English »

■ メニュー

- SSJDAについて
- データの寄託
- データの利用
- セミナー
- 調査・研究
- 謝辞・リンク集
- スタッフ・問い合わせ先
- 公開データリスト

■ 検索・分析・ダウンロード

- データ詳細検索(Namazu)
- リモート集計システム
- ダウンロードシステム

■ サイト内検索

Search

- [東京大学](#)

- [社会科学研究所](#)

- [ICPSR国内利用協議会](#)

- [パネル調査](#)

- [生涯成長型雇用システム](#)

- [JGSS](#)

Search in Archive

in [概要全文](#)

表示件数 [10](#) 件

並び替え順 [調査番号\(昇順\)](#)

[検索](#)

[>>詳細検索画面](#)

SSJDA web-site Renewal / Information

2010-3-30:

[0655](#)「食育に関する意識調査, 2009」(内閣府政策統括官(共生社会政策担当))を公開しました。

2010-3-17:

ダウンロードシステムは、サーバー入れ替え作業のため、**2010年3月23日(火)17時00分～3月24日(水)18時00分まで、停止いたします。**

SSJDA Direct(データダウンロードシステム)の運用開始

2009年4月より[SSJDA Direct](#)(データダウンロードシステム)の運用を開始しました。

SSJDAで公開されている[JGSS](#)(日本版総合社会調査)および[NFRJ](#)(全国家族調査)について、オンライン上での利用申請・ダウンロード・利用報告等が可能となっております。

[更新履歴](#) | [個人情報保護方針](#)

Copyright © SSJDA committee All Rights Reserved.

図5 2007～10年まで運用されていたウェブサイト

(2) 公開データリストの作成

図3に示した調査概要のhtmlファイルや調査票のhtmlファイルが蓄積されるに従い、後述するデータ検索システムだけでは一覧性が低いという問題があった。また、学生利用や教育目的利用が許可されたデータを探す際にも、それらの調査番号のみが利用条件下部に示されているだけで、一つ一つ調査概要の利用方法を確認していかねばならず、利便性が低いという問題もあった。アーカイブされていても利用されないデータも多く見られることも、公開されている調査情報へのアクセス環境に起因している可能性も考えられた。そこで、調査2008年6月に、図6に示す公開データリストのページ、さらには、学生利用、教育目的利用が許可されたデータのみについてもリストしたページを作成し公開した。これにより、過去に利用されていなかったデータの利用申請が出てくるなど、一定程度の調査情報へのアクセス改善が見られた。公開データリストは、次に書くCMS化の際にも引き継がれ、図7のようにデザインが変更された形で継続して公開されている。

公開データリスト(調査番号順)

(一部調整のためSSJDA公開データの全てではありません)

[\[1-100\]](#) [\[101-200\]](#) [\[201-300\]](#) [\[301-400\]](#) [\[401-500\]](#) [\[501-599\]](#)
[\[601-\]](#)[\[a001-\]](#)[\[b001-\]](#)[\[c001-\]](#)[\[M001-\]](#)[\[K001-\]](#)[\[P-\]](#)

学部学生や
教員の方へ

学生利用、教育目的利用が許可されたデータリストをご覧ください。→[こちら](#)

調査番号 1-100

調査番号	調査名	寄託者(寄託時名称)
0001	新規学卒者(中卒)労働市場調査, 1953	東京大学社会科学研究所
0002	中小企業における賃金決定の実態に関する調査, 1995	連合総合生活開発研究所
0003	若年労働者のキャリアと学習歴に関する調査, 1995	連合総合生活開発研究所
0004	小学生・中学生の生活に関するアンケート調査, 1995	連合総合生活開発研究所
0005	ボランティア活動についてのアンケート調査, 1995	連合総合生活開発研究所
0006	勤続5年以上の女性労働者のキャリア形成と人事処遇の運用実態に関する調査, 1995	連合総合生活開発研究所

図6 2008年より公開されたデータリスト一覧のウェブページの一部

公開データリスト(調査番号順)

(一部調整中のためSSJDA公開データの全てではありません)

[[1-100](#)] [[101-200](#)] [[201-300](#)] [[301-400](#)] [[401-500](#)] [[501-599](#)]
[[601-700](#)] [[701-](#)] [[a001-](#)] [[b001-](#)] [[c001-](#)] [[H001-](#)] [[M001-](#)] [[K001-](#)] [[P-](#)]

学部学生や教員の方へ

学生利用, 教育目的利用が許可されたデータリストをご覧ください。→[こちら](#)

調査番号 1-100

調査番号	調査名	寄託者(寄託時名称)
0001	新規学卒者(中卒)労働市場調査, 1953	東京大学社会科学研究所
0002	中小企業における賃金決定の実態に関する調査, 1995	連合総合生活開発研究所
0003	若年労働者のキャリアと学習歴に関する調査, 1995	連合総合生活開発研究所
0004	小学生・中学生の生活に関するアンケート調査, 1995	連合総合生活開発研究所
0005	ボランティア活動についてのアンケート調査, 1995	連合総合生活開発研究所
0006	勤続5年以上の女性労働者のキャリア形成と人事処遇の運用実態に関する調査, 1995	連合総合生活開発研究所

図 7 2010年5月時点の公開データリスト一覧のウェブページの一部

(3) ウェブサイトのCMS化

2005年あたりから一般的に普及したと言われているウェブサイトを総合・体系的に管理するCMS(Content Management System)は、HTMLやCSS等の技術的な知識や、ウェブサイトが複雑に階層化した場合のディレクトリ構造の詳細な把握がなくても、簡便に情報作成や更新がおこなえるように工夫されており、ホームページの品質を維持しつつ多人数での共同作業を効率よく行うことが可能である。SSJDAではそれまで担当助教が一人でHPの管理を行っていたが、内容がデータアーカイブ全般だけでなく、計量セミナー分析セミナーの案内や二次分析研究会の情報、さらには拠点化に伴う共同利用・共同研究の情報や社会調査研究分野の情報など、扱う情報がより広範にわたり複数の担当者が関らざるを得ない状況となってきたため、2009年度より、CMSの導入に向けた作業を開始した。

しかしながら、CMS化しつつ、扱う情報を増やして管理人数も増やすことを同時に進めることは、作業を複雑にするため、まずは、それまで運用していたウェブサイトに必要な最

小限の変更を加えて CMS 化して、その後に拠点化に伴うコンテンツの再構成を行う手順とした。表 1 に、CMS 化する際に整理したサイト構成を示す。第 1 階層として、従来より整備していた「SSJDA について」「データの寄託」「データの利用」「セミナー」「調査・研究」「謝辞・リンク集」「公開データリスト」に、「共同利用・共同研究の募集」を加えた。第 2 階層、第 3 階層も、基本的に従来整備された情報を整理することで整えた。

この階層構造に従って CMS によるウェブサイト構築を進めた。いくつか他の汎用 CMS の利用を検討したが、最終的に、一般によく利用されている Movable Type を、利用利便性や安定性も高いと判断し、また、予算制約もクリアすることから、利用することとした。

Movable Type は、シックス・アパート社が開発・提供するブログソフトウェアで、主に Perl と PHP で書かれており、動作に必要なのは Perl のみである。データストアにはリレーショナルデータベースを採用しており、利用する Movable Type 5 からは MySQL のみ対応となっている。同ソフトのユーザー管理機能等を用いて、ウェブサイトの構築を行った。

図 8 に、作成したトップページを示す。最上部に従来からの SSJDA のロゴと新しいセンター名等を配置し、その下に第 1 階層に相当するメニューバーを設けた。メニューバーの下に新しいロゴを掲載した。新しいロゴの下は、3 分割して、左側に「検索・分析・ダウンロード」と「リンク」および「SSJDA Labs」のウィンドウを配置した。中央に「Web Site 更新情報」を配置し、ブログ形式で自動更新される仕組みとした。右側に「SSJDA からのお知らせ」を配置し、データ公開以外の SSJDA からのアナウンスを掲載する欄を設けた。ページ下部には、第 1 階層に相当するメニューをテキスト形式でリンクしておく形とした。

従来、Dreamweaver での CSS 管理では、ソフトに習熟していないものがページ更新するとデザインが崩れてしまうことが多かったが、新しいホームページではそういったことが無くなった。「共同利用・共同研究の募集」についても、複数スタッフが関与しながら、ホームページの品質が維持されつつ新しいコンテンツの追加ができることが確認された。また、データ公開情報の更新の際には、毎度手作業で過去の更新情報を移動する必要があったが、その作業が省力化された。一定期間の運用を経て、多人数での共同作業が可能であることを確認し、2010 年度後半から拠点化に伴うコンテンツの再構成に着手することとした。

表1 CMS化移行第1段階でのウェブコンテンツの階層構造

ページno.	第一階層	第二階層	第三階層
1	TOP		
2	SSJDAについて	SSJDAとは？	
3		アーカイブの歴史	
4		アーカイブの機能	
5		SSJDA関連資料	沿革
6			業務統計資料
7			関連テキスト
8		これまでのイベント	
9	データの寄託	寄託の手引	
10		寄託に必要なもの	
11		データの処理方法	
12		寄託FAQ	
13	データの利用	利用条件	
14		利用手引き	
15		利用申請書	
16		利用報告書	
17		利用論文リスト	書籍
18			学術雑誌論文
19			その他
20		個人情報保護	
21		ISSPの利用	
22		利用FAQ	申請編
23			分析編
24			報告編
25	セミナー	経緯	
26		計量分析セミナー	2009年
27			2008年(特別講演)
28			2008年
29			2007年
30		公開セミナー	第6回(2002年)
31			第5回(2001年)
32			第4回(2000年)
33			第3回(1998年)
34			第2回(1997年)
35			第1回(1997年)
36	調査・研究	二次分析研究会	
37		二次分析研報告会	2010年
38			2009年
39			2008年
40			2007年
41			2006年
42			2005年
43			2004年
44			2003年
45			2002年
46		社研パネル調査	
47		JGSS	
48		委託・共同研究	
49		リサーチペーパー	
50	謝辞・リンク集	謝辞	
51		国内のリンク	
52		海外のリンク	データアーカイブの国際組織
53			アメリカのデータアーカイブ
54			ヨーロッパのデータアーカイブ
55			その他の地域のデータアーカイブ
56			国際比較調査のプロジェクト
57	共同利用・共同研究の募集	共同利用・共同研究公募	2011年度公募
58			2010年度公募
59		参加者公募型共同研究	2011年度公募
60			2010年度公募
61		課題公募型共同研究	2011年度公募
62			2010年度公募
63	公開データリスト	公開データリスト	
64		学生・教育利用許可	
65	お問い合わせ・スタッフ		
66	サイトマップ		
67	プライバシーポリシー		

Social Science Japan Data Archive



検索・分析・ダウンロード

- データ検索システム
- リモート集計システム
- ダウンロードシステム

リンク

- 東京大学
- 社会科学研究所
- 社会調査・データアーカイブ研究センター
- ICPSR国内利用協議会
- 社研 パネル調査
- 社研 近未来事業
- 日本版総合社会調査

SSJDA Labs

構築中

Web Site 更新情報

2011-04-07 0705「労働関係法制度の知識の理解状況に関する調査, 2008」(厚生労働省 政策統括官付労働政策担当参事官室)を公開しました。
 0706「業務委託契約従事者に関する調査, 2010」(厚生労働省 政策統括官付労働政策担当参事官室)を公開しました。

2011-03-03 0702「今後の仕事と家庭の両立支援に関する調査, 2007」(厚生労働省 雇用均等・児童家庭局 職業家庭両立課)を公開しました。

2011-02-24 0692「介護労働実態調査2008」(介護労働安定センター)を公開しました。
 0693「介護労働実態調査特別調査2008」(介護労働安定センター)を公開しました。

2011-02-17 0800「日本版General Social Surveys」(大阪商業大学 JGSS研究センター)を公開しました。

2011-01-13 0707「第1回子ども生活実態基本調査, 2004」(ベネッセコーポレーション)を公開しました。

SSJDAからのお知らせ

2011年度共同利用・共同研究公募の開始のお知らせ
 2011年度共同利用・共同研究の公募を開始いたしました。詳細は[こちら](#)をご参照ください。

ダウンロードシステムについて
 ダウンロードシステムの督促メールについて、システムが自動的にメールを送信しておりますので、被災地域の利用者様にも送信されております。大変申し訳ありませんが、何卒ご容赦願います。

リモート集計システム再開のお知らせ
 リモート集計システムを再開いたしました。

共同利用・共同研究拠点としての活動開始
 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は、社会調査・データアーカイブ共同利用・共同研究拠点として文部科学省から認定を受け、2010年4月より拠点としての活動を開始しました。

SSJDA Direct(データダウンロードシステム)について
 2009年4月よりSSJDA Direct(データダウンロードシステム)の運用を開始しています。現在、JGSS、NFRJおよび政治関連のデータを掲載しています。

社研メールニュースのご案内
 月1度、データ公開を含む社研関連情報メールを受信できます。ご希望される方は、[こちら](#)をクリックください。

図 8 2010年4月にリニューアルしたウェブサイト

表2に、再度整理したサイト構成を示す。これまで、主にデータアーカイブのホームページに二次分析研究会のページが付随するような形で、社研パネル調査のページは別に作られているなど、社会調査・データアーカイブ研究センターのウェブサイトとしては十分な構成ではなかった。そこで、まず、「センター」「調査基盤」「社会調査」「計量社会」「国際調査」の5つの分類を明確にして、その上で第1階層として、従来のものに加えて、「センターについて」「社研パネル調査」「共同調査と共同研究」「国際連携」を加えた。「センターについて」では、拠点化に伴い作成したセンターを紹介する文章やパンフレットで作成した情報を整理して掲載する。「社研パネル調査」は従来別にあったパネル調査のホームページを移行するものである。「共同調査と共同研究」は、社会調査研究分野で進めているJGSS（日本版総合社会調査）との協力や職業・産業の自動コーディングに関する共同研究について掲載するために設けた。「国際連携」は、これまでのSSJDAにおける国際活動を整理しつつ、本リサーチペーパー後半で進めているDDIの利用可能性検討に関する情報を掲載するものとして設けた。これにより、社会調査・データアーカイブ研究センターで進めている作業がほぼ網羅される形でのホームページが整うこととなる。表2は、サイト移行検討時に整理されたもので、今後、第2階層以下については各担当者による調整があった後に公開される予定である。その際に若干変更の可能性のあることを付記しておくが、これも、CMSにより多人数での共同作業が可能となっているが故のこととなる。

図9に、作成中の新しいホームページのトップ画面を示す。従来のデータアーカイブのページを維持しつつ、社会調査・データアーカイブ研究センターとしての活動がより広範にフォローされるようになっている。最上部に、新たにCSRDAというセンターのロゴを配置しつつ、センター名と従来からのSSJDAのロゴ等を配置し、その下に第1階層に相当するメニューバーを設けた。メニューバーの上には、センターの部門構成を明確にしてある。メニューバーの下には2010年度より利用しているイメージを掲載した。イメージの下には、2分割して、左側に「センターからのお知らせ」を、右側に「SSJDA データ公開情報」を配置し、ブログ形式で自動更新される仕組みとした。ページ下部には、第1階層に相当するメニューをテキスト形式でリンクしておく形としてある。

表2 CMS化移行第2段階でのウェブコンテンツの階層構造

ページno	分類	第1階層(上部メニュー)	第2階層	第3階層		
1	センター	センターについて	センター長あいさつ			
2			社会調査・データアーカイブ研究センターについて			
3			SSJDAリサーチペーパーシリーズ			
4			委託・共同研究			
5			スタッフ			
6			個人情報保護			
7			これまでのイベント			
8	調査基盤	SSJDAについて	SSJDAとは？			
9			データアーカイブをめぐる状況			
10			謝辞・リンク集	謝辞		
11				国内のリンク		
12				海外のリンク		
13				SSJDA関連資料		
14				業務統計資料		
15				関連テキスト		
16				利用論文リスト		
17			データの寄託	寄託の手引き		
18					寄託に必要なもの	
19					データの処理方法	データクリーニング
20					個票データについて	
21					無回答・非該当の具体例	
22				寄託FAQ		
23	データの利用(共同利用)	利用条件	利用条件			
24			利用の流れ			
25			データ検索			
26			オンライン申請			
27			郵送利用申請			
28			二次分析と利用報告			
29			公開データ	公開データリスト		
30				学生・教育利用許可データリスト		
31				申請編		
32				分析編		
33				報告編		
34			社会調査	社研パネル調査	働き方とライフスタイルの変化に関する全国調査	
35	高校卒業後の生活と意識に関する調査					
36	プロジェクトの概要	プロジェクトの目的				
37		3つのパネル調査				
38		研究資金				
39		プロジェクトの組織				
40		研究委員会の開催				
41		調査票の集計				
42		個人情報保護				
43		ディスカッションペーパー				
44		研究成果報告会			研究成果報告会	
45					研究成果報告会2011	
46					研究成果報告会2010	
47					研究成果報告会2009	
48					研究成果報告会2008	
49		プレスリリース				
50		データの2次利用				
51	共同調査と共同研究	JGSS				
52			自動コーディング(職業・産業)			
53	計量社会	二次分析研究会(共同研究)	参加者公募型			
54			課題公募型			
55			一般公募型			
56			過去の研究会			
57			二次分析研究報告会	2010年		
58				2009年		
59				2008年		
60				2007年		
61				2006年		
62				2005年		
63				2004年		
64		2002年				
65	計量分析セミナー	経緯	経緯			
66			公開セミナー			
67			計量分析セミナー	計量分析セミナー2011		
68				計量分析セミナー2010(特別講演)		
69				計量分析セミナー2010		
70				計量分析セミナー2009		
71				計量分析セミナー2008(特別講演)		
72				計量分析セミナー2008		
73				計量分析セミナー2007		
74	国際調査	国際連携	国際活動のこれまで			
75			東アジアDAシンポジウム			
76			DDI利用可能性検討			
77	サイトマップ					
78	プライバシーポリシー					

社会調査・データアーカイブ共同利用・共同研究拠点
 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

SSJDA

HOME ENGLISH お問い合わせ

センター SSJDAデータアーカイブ (調査基礎) 社会調査 計量社会 国際調査

センターについて SSJDAについて データの寄託 データの利用 (共同利用) 社研パネル調査 共同調査と共同研究 二次分析研究会 (共同研究) 計量分析セミナー 国際連携

Social Science Japan Data Archive

センターからのお知らせ

2011年計量分析セミナーのご案内
[2011年計量分析セミナーのご案内](#)を掲載しました。なお、申込は終了いたしました。

共同利用・共同研究拠点としての活動開始
 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は、社会調査・データアーカイブ共同利用・共同研究拠点として文部科学省から認定を受け、2010年4月より拠点としての活動を開始しました。

SSJDA Direct(データダウンロードシステム)について
 2009年4月よりSSJDA Direct(データダウンロードシステム)の運用を開始しています。現在、JGSS、NFRJおよび財政関連のデータを掲載しています。

社研メールニュースのご案内
 月1度、データ公開を含む社研関連情報メールを受信できます。ご希望される方は、[こちら](#)をクリックください。

SSJDA データ公開情報

2011-03-03 0702「今後の仕事と家庭の両立支援に関する調査, 2007」(厚生労働省・雇用均等・児童家庭局職業家庭両立課)を公開しました。

2011-02-24 0692「介護労働実態調査, 2008」(介護労働安定センター)を公開しました。
 0693「介護労働実態調査特別調査, 2008」(介護労働安定センター)を公開しました。

2011-02-17 0800「日本版General Social Surveys」(大阪商業大学 JGSS 研究センター)を公開しました。

2011-01-13 0707「第1回子ども生活実態基本調査, 2004」(ベネッセコーポレーション)を公開しました。

2011-01-06 0701「仕事と生活の両立支援策と企業業績に関する調査, 2005」(厚生労働省・雇用均等・児童家庭局職業家庭両立課)を公開しました。
 0742「勤労者の仕事と暮らしについてのアンケート, 2010.4」(連合総合生活開発研究所)を公開しました。

Page Top

センターについて | SSJDAについて | データの寄託 | データの利用(共同利用) | 社研パネル調査プロジェクト | 共同調査と共同研究 | 二次分析研究会(共同研究) | 計量分析セミナー | 国際連携 | お問い合わせ:スタッフ | サイトマップ | プライバシーポリシー

Copyright 2010 © SSJDA committee All Rights Reserved.

図 9 2011 年度前半に公開予定の内容を拡大したウェブサイト

1.4 Namazu によるデータ検索システム

(1) システム概要

SSJDA では、2002 年 11 月より、オープンソースの全文検索システムである Namazu を用いて、ウェブ上での公開データの検索システムを実現している。Namazu は、1997 年、小中規模のメールアーカイブや WWW サーバ用の全文検索システムとして高林氏によって開発され、その後日本語圏における検索エンジンの代表格として普及したものである。現在は、Namazu Project によって開発されている。大量の文書を高速に検索するためにあらかじめインデックスの作成を行ない、インデックスを作成するプログラムは Perl で記述されている。検索コマンドは C 言語で記述されており、検索にかかる時間はインデックスの大きさにあ

まり影響されず検索ができる。セキュリティ面の強化も繰り返されており¹、最近では 2009 年 9 月に公開された version2.0.20 で、バッファオーバーランと呼ばれるプログラムの不具合によりデータ領域外のメモリを書き換えることで誤動作を起こすセキュリティホールへの対応がなされている。

NAMAZU の基本要件については、<http://www.namazu.org/index.html.ja> もしくは、馬場肇『Namazu システムの構築と活用』（ソフトバンク社 2001）に詳しいので参照されたい。

SSJDA では、2010 年からの CMS によるウェブサイトのリニューアルに併せて、データ検索システムの調整を行った。具体的には、①従来、作成していた調査票と集計結果が掲載されている html ファイルは、データアーカイブ室での作業軽減のため PDF 化されたため、調査概要の html ファイルのみを対象とすること、②調査年やトピック、代表的なキーワードによる検索機能を追加すること、③レイアウトを新しい HP に併せたものに変更すること、の 3 点に取り組んだ。

システム要件は、Windows server 2003 R2 standard SP2 において、NAMAZU-2.20、形態素解析の Mecab-0.98、ActivePerl-5.8.9 を搭載する形となっている。

全ての調査概要の html ファイルに、①トピック検索を実装するために topic タグを追加、②調査年検索を実装するために year タグを追加、③検索結果画面で調査名を表示させるために datatitle タグを追加、④キーワード検索を実装するために keyword タグを追加、という作業を加えた。

(2) 利用方法

図 10 に、データ検索システムの検索画面を示す。検索は、収録調査ごとにまとめている「概要」（調査名、寄託者名、調査対象、主要調査事項など）の情報を検索する。①概要全文、②調査名、③寄託者名、④調査年、⑤トピック、⑥キーワード、の 6 つの方法で検索でき、一度検索語、他の方法を組み合わせて絞り込んでいくこともできるようにしてある。

表示件数は、10～50 件を用意し、表示順は、調査番号、調査名、寄託者名、トピック、調査年をそれぞれ昇順と降順を用意した。検索ボタンは、「検索」「リセット」「全データ表示」を用意した。「検索」ボタンでは、検索条件のアンド検索を実施する形としており、未入力検索項目は無視される。「リセット」ボタンは、入力中の条件を、未入力状態へと初期化する。「全データ表示」ボタンは、検索範囲すべての調査データが表示する。

「調査年」では、1990 年以前、1991～95 年、1996～2000 年、2001～05 年、2005～10 年と区分してある。今後、2011 年～のカテゴリが追加される。「トピック」は、すでに公開されているデータを見ながら、担当スタッフで図 11 に示すように整理した。図中にも明示し

¹ 詳細は、<http://www.namazu.org/security.html> 参照

詳細検索

概要全文:
調査年:

調査名:
トピック: [\(トピック一覧\)](#)

寄託者名:
代表的なキーワード:

表示件数:
表示順:

*ブラウザは、Internet Explorer8.xとFirefox3.6.xに対応しています。
 *JavaScriptとCookieを利用しておりますので、ご利用のブラウザの設定を有効にしてください。

検索方法

収録調査ごとにまとめている「概要」(調査名、寄託者名、調査対象、主要調査事項など)の情報を検索します。
 次の6つの方法で検索できます。一度検索後、他の方法を組み合わせて、絞り込んでいくこともできます。

概要全文	概要全文: <input type="text" value="教育"/>	教育 が含まれる概要を検索
調査名	調査名: <input type="text" value="雇用"/>	調査名に 雇用 が含まれる概要を検索
寄託者	寄託者名: <input type="text" value="センター"/>	寄託者に センター が含まれる概要を検索
調査年	調査年: <input type="text" value="2006~2010"/>	調査年が 2006年から2010年 の概要を検索
トピック	トピック: <input type="text" value="雇用・労働"/>	トピックが 雇用・労働 の概要を検索。研究分野別に、概要を検索します。
キーワード	代表的なキーワード: <input type="text" value="若年"/>	キーワードが 若年 の概要を検索。SSJDAでよく検索に使われているキーワードで検索します。

図 10 データ検索システムの検索画面

であるように、SSJDA では雇用・労働関係の調査が多く、これについては別に、社会科学研究所で実施されている近未来の課題解決を目指した実証的社会科学研究推進事業「生涯成長型雇用システムプロジェクト」と連携をして、詳細なキーワード検索を実現している²。「キーワード」については、試験的に「高齢者」「ジェンダー」「若年」の3つを調査概要のhtmlファイルに埋め込み検索可能としている。図11に示すように、他にも代表的なキーワードが想定されるが、作業量の制約から、2011年5月時点では3つのキーワードにとどめている。

² http://ssjda.iss.u-tokyo.ac.jp/future/search_top.html 参照のこと

トピック一覧		
トピック	定義	代表的なキーワード
情報・通信・マスメディア	情報通信技術の利用状況、マスメディアの利用状況に関するもの。	インターネット、電話、携帯電話、テレビ、ラジオ、新聞等
経済・産業・経営	個人・企業の生産・消費活動を扱ったもの。企業等の活動のうち、雇用・労働以外のもの。	金融、投資、消費、資産、景気、家計
教育・学習	学校教育、人材育成に関するもの。自己啓発等も含む。	学校、育成、人材開発、成長
法律	司法制度に関するもの。(裁判、訴訟)法律そのものに対する意識等を調査しているもの。	裁判、訴訟
保健・医療	健康に関する意識や医療制度、医療に対する考え方等。	健康、疾患、障害
社会保障・福祉	年金、医療保険等社会保障に関するもの、児童、障害者、経済的困窮者等の福祉に関するもの。	生活保障、福祉サービス、年金、介護
国際比較・外交	国際比較調査、国家間の関係性等。	国際比較、国際関係
政治・行政・選挙	政治・行政に関する意識、投票行動に関するもの。	投票、政党、官僚、選挙、自治
雇用・労働*	雇用・労働に関する制度、意識等に関するもの。 * 雇用・労働については、詳細な分類が 近未来事業-雇用・労働調査データベース に記載されています。	人事、労働時間、派遣、労働組合、キャリア、ワークライフバランス
社会・文化	社会一般に対する意識、価値観や生活に関するもの。	社会意識、価値観、 高齢者 、 ジェンダー 、家族、結婚、生活、 若年
環境	地球環境、自然環境に関するもの。	地球環境、環境保護
安全・犯罪・災害	社会安全、犯罪、自然災害に関するもの。	犯罪、治安、地震、噴火
その他	上記のいずれにも当てはまらないもの。	

図 11 データ検索システムのトピック一覧

検索に成功すると、調査番号、調査名、寄託者名、トピックが表示される。図 12 に、トピックを「雇用・労働」として、キーワードに「若年」をつけて検索した結果例を示す。

検索結果画面の調査名のリンクをクリックすると、該当した調査の概要が表示される。検索結果画面の寄託者名のリンクをクリックすると、該当した寄託者による調査一覧が表示される。このとき、寄託者名以外の検索条件は初期化される。検索結果画面のトピックのリンクをクリックすると、該当したトピックの調査一覧が表示される。このとき、トピック以外の検索条件は初期化される。

検索結果
 検索ヒット数: [+id:: 882] [+topic:雇用・労働: 267] [+id:: 882] [+datatitle:: 882] [+depo:: 882] [+year:: 882]
 [+keyword:若年: 32]

検索の結果、9 個の調査が該当しました。

*寄託者／トピックのリンクをクリックすると、該当する寄託者／トピックの一覧が表示されます。なお、他の検索条件は初期化されます。

調査番号	調査名	寄託者(寄託時名称)	トピック
0286	若年アルバイトの仕事と意識に関するアンケート調査	旧東京都立労働研究所	雇用・労働
0321	新入社員「働くことの意味」調査,2003	社会経済生産性本部	雇用・労働
0389	若年者の採用方法の多様化と人材育成に関するアンケート,2003	連合総合生活開発研究所	雇用・労働 教育・学習
0405	若年者のキャリア形成に関する実態調査,2003	三菱UFJリサーチ&コンサルティング(寄託時 UFJ総合研究所)	雇用・労働 教育・学習
0421	若年者の採用・雇用管理の現状に関する調査,2004	労働政策研究・研修機構	雇用・労働
0423	第二新卒者の採用実態調査,2004	労働政策研究・研修機構	雇用・労働
0498	中小企業における若年労働者に関する調査,2004	連合総合生活開発研究所	雇用・労働
0525	インターンシップの実態に関するアンケート,2004	佐藤博樹	雇用・労働 教育・学習
0623	若者の生活と仕事に関する調査,2003	家計経済研究所	経済・産業・経営 雇用・労働 社会・文化

現在のリスト: 1 - 9
 ページ: [1]

図 12 検索結果画面の例

(3) 運用状況

データ検索システムの運用状況について、表 3 と図 13 に示す。表 3 には、年単位でデータ検索システムでの検索回数と、参考までに 2008 年から用意した公開データリストのアクセス件数を示す。2002 年は 11 月からの運用であったため、2003 年度からの数値としてある。2003 年度には約 3 万 1 千回の検索があったが、その後増加して、2007 年度には約 5 万 4 千回の検索となった。2008 年度には、公開データリストを設けたため、そちらの方でデータを探す人も増え、データ検索システムでの検索回数が約 4 万 3 千、公開データリストへのアクセス件数は約 5 千であった。公開データリストへのアクセス 1 回は、データ検索システムでの検索回数数回分に相当すると考えられ、2007 年以前との単純な比較は難しいが、2008 年度以降も SSJDA に所蔵されたデータの検索量は増加しているものと見受けられる。

図 13 に、月ごとのデータ検索システムでの検索回数の推移を示す。年度によって若干ばらつきが見られるものの、年度がはじまり少し経った 5 月から 6 月にかけて検索が最も増え、その後夏休みで検索が低下し、秋学期が開始する 10 月頃に検索が増えて、その後検索が低下していくという大きな傾向が見て取れる。データ検索システムのメンテナンスや入れ替え等は、利用が低下する 8 月頃か翌 2～3 月頃に行うことが適当であると見受けられる。

表 3 データ検索システムでの検索回数と公開データ一覧へのアクセス件数の推移

年度	検索回数	公開リストのアクセス数
2003年	31,014	-
2004年	40,583	-
2005年	38,770	-
2006年	43,011	-
2007年	53,986	-
2008年	43,613	5,406
2009年	40,943	10,255
2010年	41,425	12,642

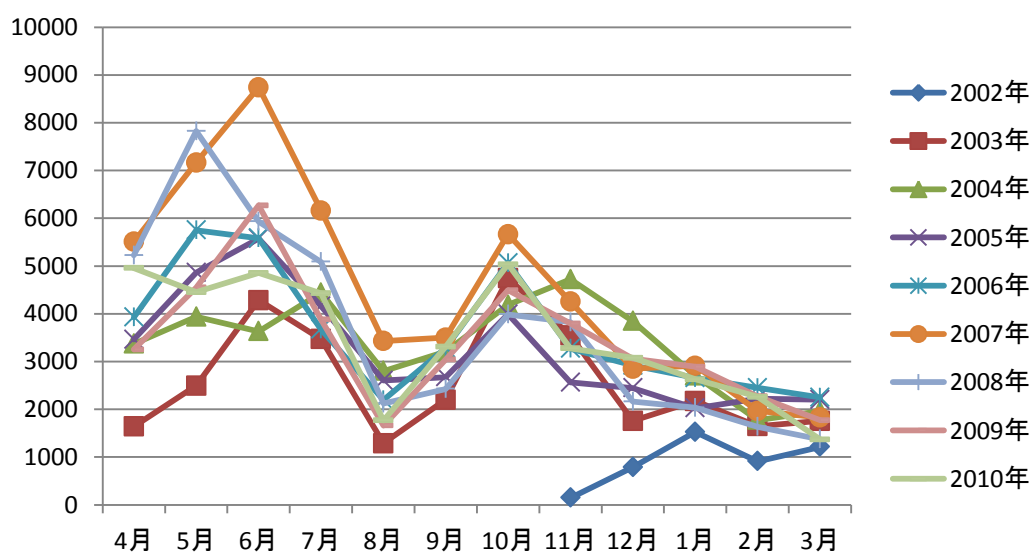


図 13 月ごとのデータ検索システムでの検索回数の推移

1.5 データダウンロードシステム Direct³

(1) 開発目的

SSJDA におけるデータダウンロードシステム（略称 SSJDA Direct）の開発目的は、①利

³ この部分は（佐藤慶一 2009）に、最新の運用状況等の情報を加えて加筆したものである。

用利便性の向上，②データアーカイブ側の事務作業量の軽減，の2点である。

1点目の，①利用利便性の向上の背景として，ICPSR や UKDA をはじめとした海外のデータアーカイブの大半で，オンライン上で個票データのダウンロードができるという状況があった。利用者から見ると，SSJDA の利用には，申請書や空 CD の郵送などといった手続きの煩雑さやデータ入手までに時間を要するなど不便さがあり，オンライン上でのデータ提供は利用利便性を向上させるものと考えられた。

もう一点の開発目的として，②データアーカイブ側の事務作業量の軽減が挙げられる。SSJDA では，個票データを CD-R で提供するという方法をとってきたが，運用開始した 1998 年には 11 だった利用申請件数が，2008 年には 481 まで増加しており，事務作業量の軽減が課題となっていた（佐藤博樹 2008）。1 つのデータ提供には，利用申請書類の受付・確認，申請情報のデータベースへの入力，寄託者への承認依頼，空 CD へのデータの焼付けと郵送，利用報告書の催促，利用報告情報のデータベースへの入力，成果物情報のデータベースへの入力，などの多くの事務作業が伴い，限られたスタッフの対応力の限界に近づきつつあった。オンライン上での利用申請やデータ提供，利用報告が実現できれば，相応の事務量の軽減が図れると考えられた。

（2）システム概要

前章に示した開発目的のもと，SSJ データアーカイブでは，2008 年度より SSJDA Direct の開発に着手した。従来の CD-R 提供で確立した利用手続きのプロセスは踏襲した上で，セキュアな環境でのオンラインデータ提供を実現することが求められた。基礎設計についての検討の結果，オンライン上で，利用諸手続き，データベース管理，管理業務を一元的に実現するべく，以下の基本設計要件を整えた。

- HTTP : Apache + OpenSSL
- データベース : MySQL
- 開発言語 : PHP 5
- セキュリティ : マイページ，管理者の全ページを SSL（暗号化）通信とする。

システムの基本設計の概要を図 14 に示す。概要は以下の 3 点である。①利用申請，申請承認，教育指導員承認（学生の場合），データダウンロード，利用報告，運用状況管理等のプロセスを，WEB 上の暗号化されたセキュアな環境で実現する。②ユーザー登録を行い，利用申請やダウンロード等は認証された人しかアクセスできないマイページで行う。③管理ページは，東京大学社会科学研究所内からのアクセスに限定する。

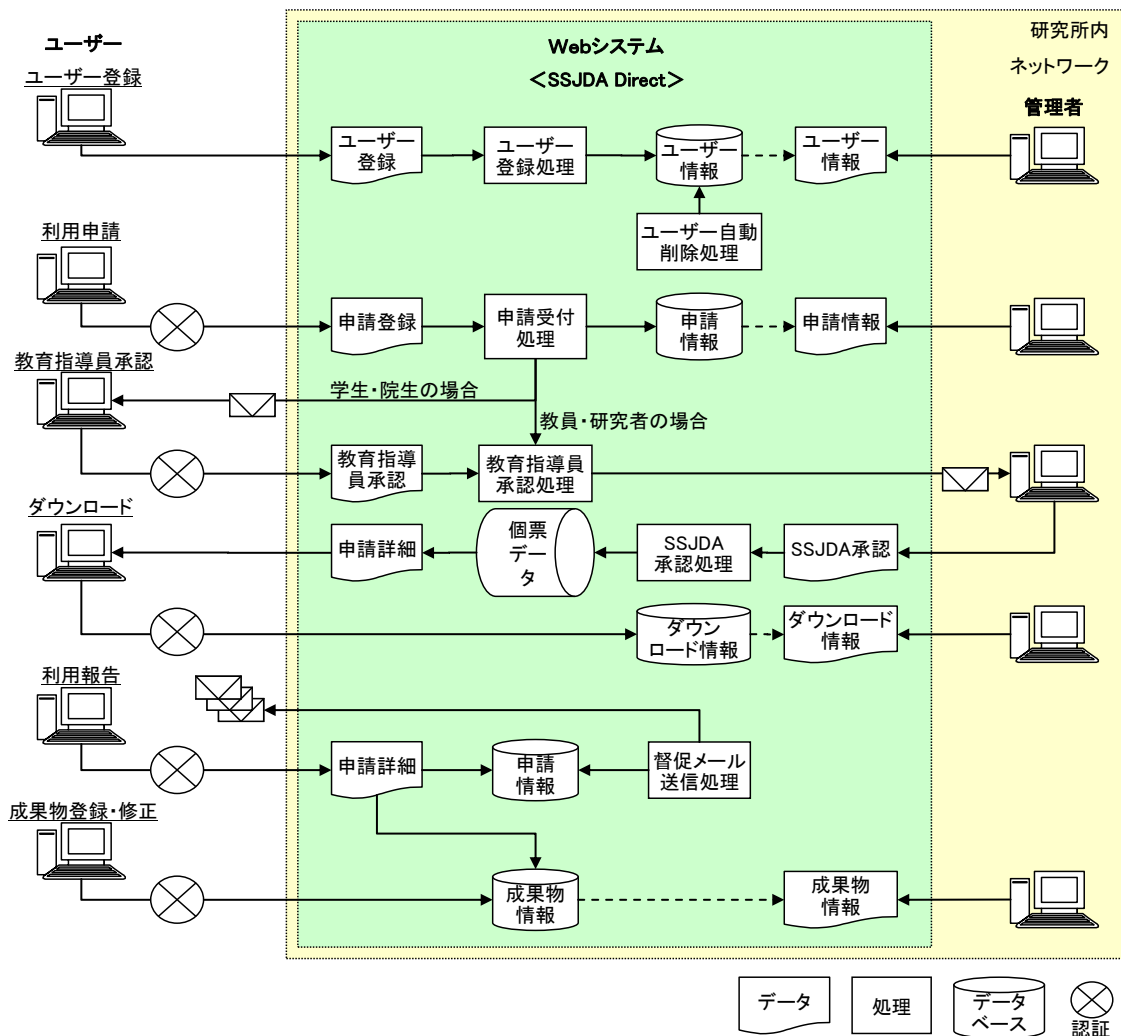


図 14 SSJDA Direct のシステム概要

(3) 利用方法

図 15 に利用者用の WEB 画面遷移を示す。計 35 ページが用意されており、新規登録，利用申請，ダウンロード，利用報告，成果物登録等が可能となっている。

「データ一覧」のページには、一覧表内に概要ファイルへのリンクを施してある。登載データは、2011 年 5 月現在で、SSJDA で公開されている JGSS（日本版総合社会調査）10 調査および NFRJ（全国家族調査）3 調査，政治関連データ 4 調査の計 17 調査である。

ログインページにアクセスすると、画面右上に「ご利用マニュアル」ボタンがあり、PDF 形式で利用手順手引きがある。

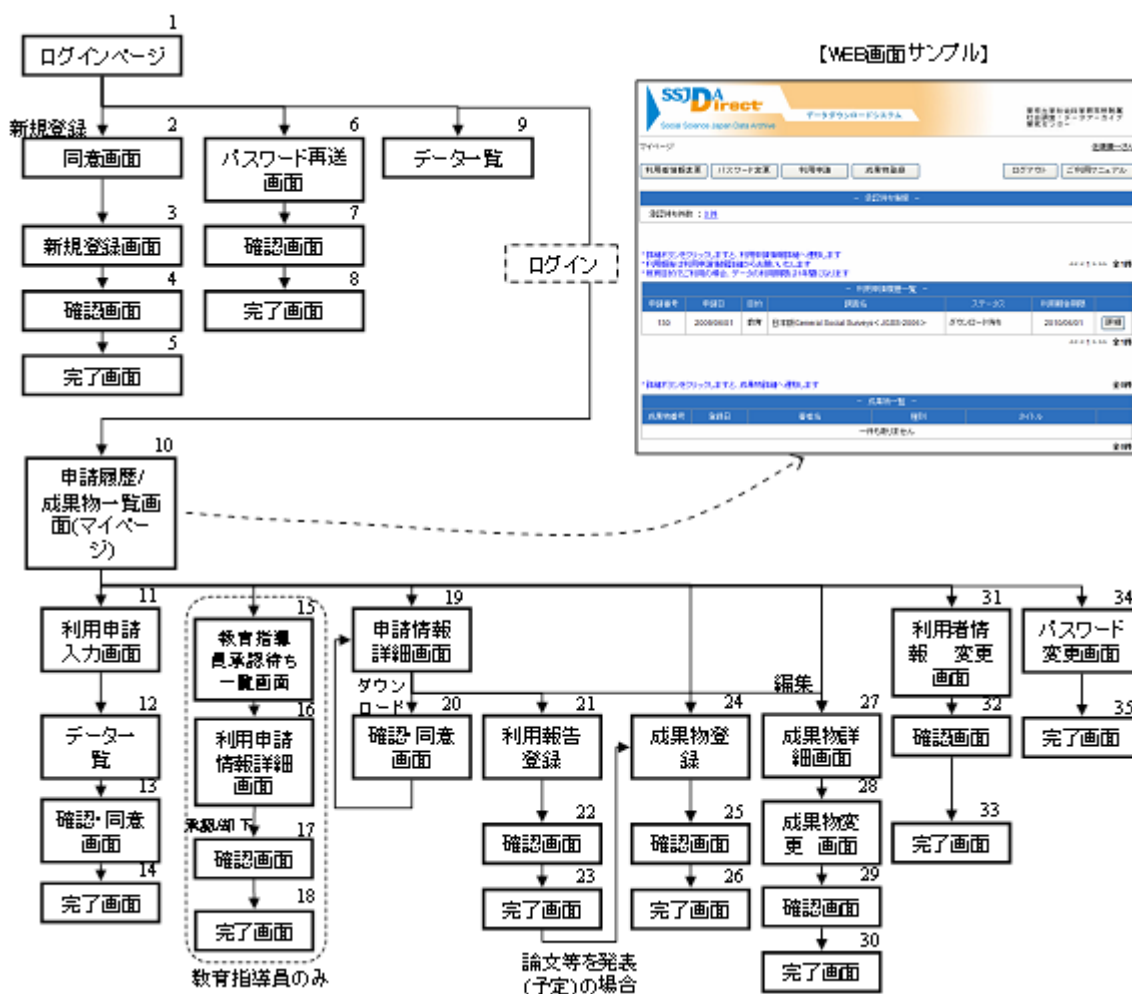


図 15 SSJDA Direct 利用者用の WEB 画面遷移

(4) 運用と課題

管理者用の WEB 画面遷移を図 16 に示す。WEB システム内で行う運用業務は、利用承認、利用状況の管理であり、システム内のデータベースは全て CSV ファイルで出力可能としている。

2009 年に運用開始して約 2 年が経過した。2009 年度には、Direct の HP およびメール画面の英語化を行い、2010 年 4 月より運用開始した。利用状況は表 4 に示す通りで、いずれの数値も 2009 年に比べ 2010 年の方が上回った。利用申請数（データセット数）は、2009 年が 448、2010 年が 520 で、合計 968 にのぼる。うち研究目的が 823 と大半であるが、145 の教育利用もなされた。ダウンロード数は、2009 年が 472、2010 年が 630 で、合計 1102 にのぼる。利用者数は、2009 年が 360、2010 年が 434 で、合計 794 人にのぼる。成果物登録数は、2009 年が 6 と少なかったが、2010 年は 37 と増加し、合計 43 の学術論文等が、SSJDA Direct からダウンロードしたデータを用いて、執筆され刊行された。

SSJDA Direct については、登載データを増やし検索機能等を拡充した形で拡張開発してい

く方向性と、登録データは利用申請件数が多いものに限って現行のまま継続運用しつつ、リモート集計システム（佐藤慶一 2008）や検索機能の刷新と併せて nesstar に代表される本格的なデータアーカイブシステムを導入していくという 2 つの方向性がある。SSJDA のスタッフ・予算の制約を睨みつつ、適切に情報システムをハンドリングし、利用利便性が高いデータアーカイブへと導いていくことが求められよう。

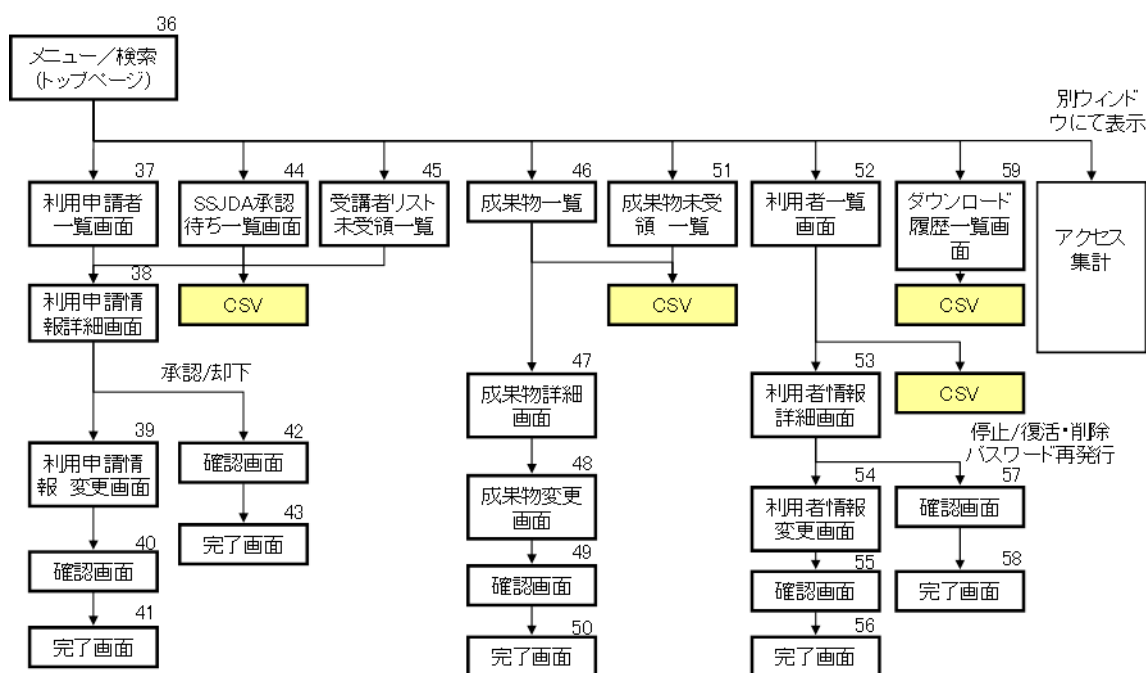


図 16 SSJDA Direct 管理者用の WEB 画面遷移

表 4 SSJDA Direct の利用状況

年度	利用申請数			ダウンロード数	利用者数			成果物登録数
	総数	研究目的	教育利用		総数	教員	教員以外	
2009年	448	386	62	472	360	178	182	6
2010年	520	437	83	630	434	191	243	37
合計	968	823	145	1102	794	369	425	43

1.6 リモート集計システム⁴

(1) 開発目的

SSJ データアーカイブにおけるリモート集計システム開発の目的は、①利用者への提供情報の拡大、②寄託データの教育活用、の 2 点に集約される。

⁴ この部分は（佐藤慶一 2008）に、最新の運用状況等の情報を加えて加筆したものである。

1点目の、①利用者への提供情報の拡大の背景として、増加するデータ利用申請数に対して、発表論文・著書数が少ないという状況が挙げられた。対策として、提供側の情報提供のあり方が再検討された。具体的には、WEB上で調査概要と調査票を眺めるだけでは、変数間の関係性を事前にチェックすることができず、データが手元に届いてから分析してみたら、利用目的を満たさない場合への対応が検討された。そこで、データ提供前の時点で、多次元の結果表を手元で集計したり、変数間の関係性をチェックできるようになれば、利用者の二次分析の研究計画の質的向上が図れ、それにより適切な分析の促進や成果報告数の増加も期待できると考えられた。

以上の観点から、SSJデータアーカイブでは、社会科学研究所図書室で調査報告書の閲覧対応を開始すると同時に、クロス集計や相関分析等の一歩進んだデータ分析をWebブラウザから行えるようにするリモート集計システムの構築を目指した。

もう一点の開発目的として、②寄託データの教育活用が挙げられる。寄託された多くの社会調査の個票データを、研究面だけでなく、教育面においても活用することは、アーカイブの意義として重要である。

実際の調査個票データを用いたデータ分析の演習を行うには、その前段階として、利用するデータの特徴の理解、分析手法の習得に加え、利用するデータの統計ソフトへの読み込み、利用する統計ソフトの操作方法の習熟等が必要で、多くの時間を要する。当然、データ分析の演習そのものにも多くの時間を要し、結果、初学者には、多大な学習時間が必要になる。さらに、個票データを自宅へ持ち帰ることや自宅PCへの統計ソフトの導入は難しい場合も想定され、実際のデータを用いた演習が不十分になることが危惧される。

さらには利用する統計ソフトの問題もある。多機能な統計パッケージは初学者向けでない場合が多く、ユーザーインターフェースの充実している有償の統計ソフトを自前のPCに導入するには相応の資金的コストを要する。

SSJデータアーカイブにリモート集計システムが導入されれば、ネットに接続したPCがあれば実際の個票データを用いた分析演習を従来に比べ簡便に短時間で実施可能となり、自宅等でのE-learningとしての利用が見込まれる。同時に、寄託データの教育活用の促進も図れる、ということが、リモート集計システム開発のもう一つの目的である。

(2) システムの概要

前章に示した開発目的のもと、SSJデータアーカイブでは、2004年度よりリモート集計システムの構築を開始し、2005年4月から試験運用を開始した。そして、2005年10月より正式な運用を開始した。図17に構築された情報システムの概要を示す。集計分析をSPSS社のWebAPPによって、ユーザ管理機能をOpenLDAPとよばれるディレクトリ・サービスのフリーウェアによって実現していた。具体的には、①SPSS WebAppによる集計分析や管理と、②OpenLDAPによって、ユーザー追加・削除やパスワード変更・再発行と関連するメール（ユーザーIDやパスワード、利用期限等）の自動送付、ならびに有効期限切れユー

ザーの一括削除の処理などの機能を実現している。

システムの本格運用に伴い、講習会等での利用事例も出てきて、15～6台以上で同時使用するとレスポンスが低下する、等の課題が出てきた。多数同時利用によるレスポンス低下は、開発時に、個人研究者の調査個票データのチェックのための利用や、社会統計を学習する学生の自宅での E-learning 利用を想定しており、授業での多人数同時利用は想定していなかったことにも起因した。不具合再現テストを重ねた結果、ユーザー管理のシステムと SPSS WebApp の連携部分の不具合が最大の要因であることを確認したが、SPSS WebApp の開発が終了していたこともあり、その調整には相応の労力がかかることと、将来的に nesstar など他のリモート集計システムの導入も視野に入ることから、2009年6月にユーザー認証機能を取り外し、登録データも複数のデータ⁵から JGSS のみと変更することで対応した。したがって、2009年6月以降は、図 17 中の「ユーザー管理 WEB システム」は取り外した形で運用されている。

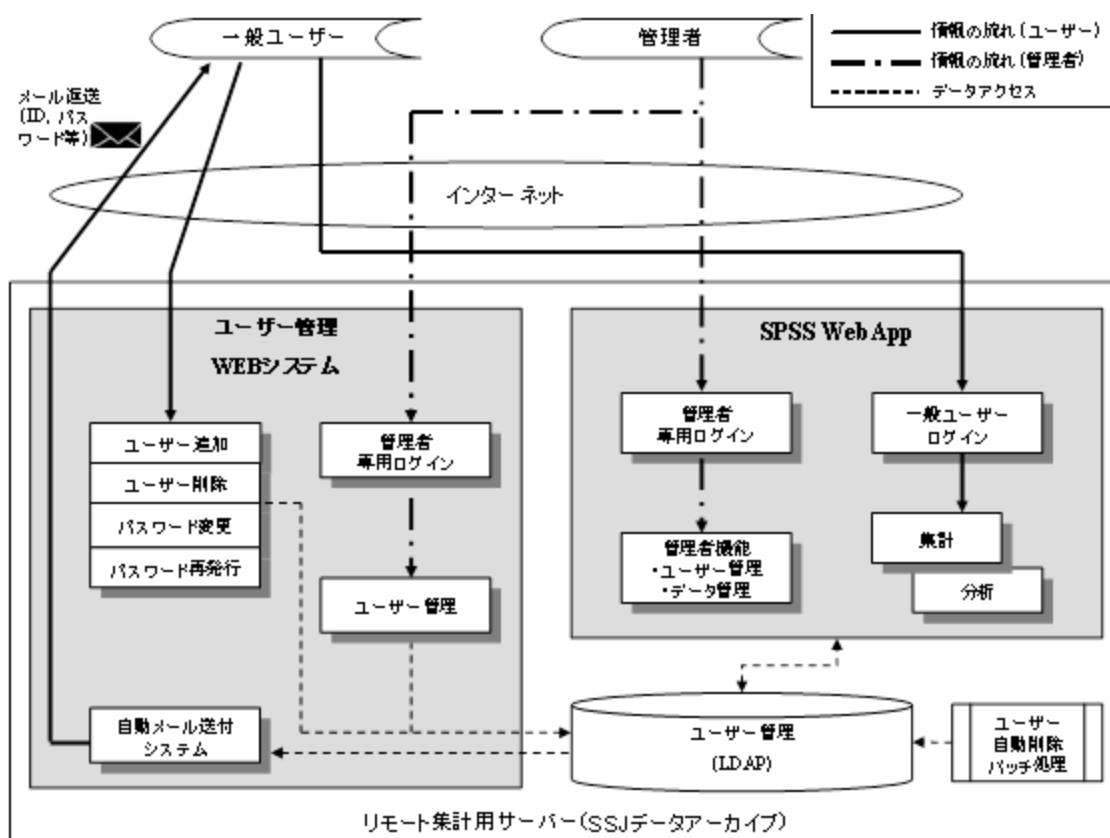


図 17 リモート集計システムのシステム概要

⁵ 2008年8月時点では、JGSS-2000/2001/2002/2003（大阪商業大学比較地域研究所・東京大学社会科学研究所寄託），AsiaBarometer2003（猪口孝氏寄託），社会生活における不安感に関するアンケート2004（社会安全研究財団寄託），「モノグラフ小学生ナウ」「モノグラフ中学生の世界」「モノグラフ高校生」（ベネッセコーポレーション寄託）が搭載されていた。

(3) 利用について

以前はユーザー登録や認証が必要であったが、2009年6月のシステム改変に伴い、現在はトップページで利用規定に同意後、自動でログインされ図18に示す利用画面トップが表示される仕組みとなっている。

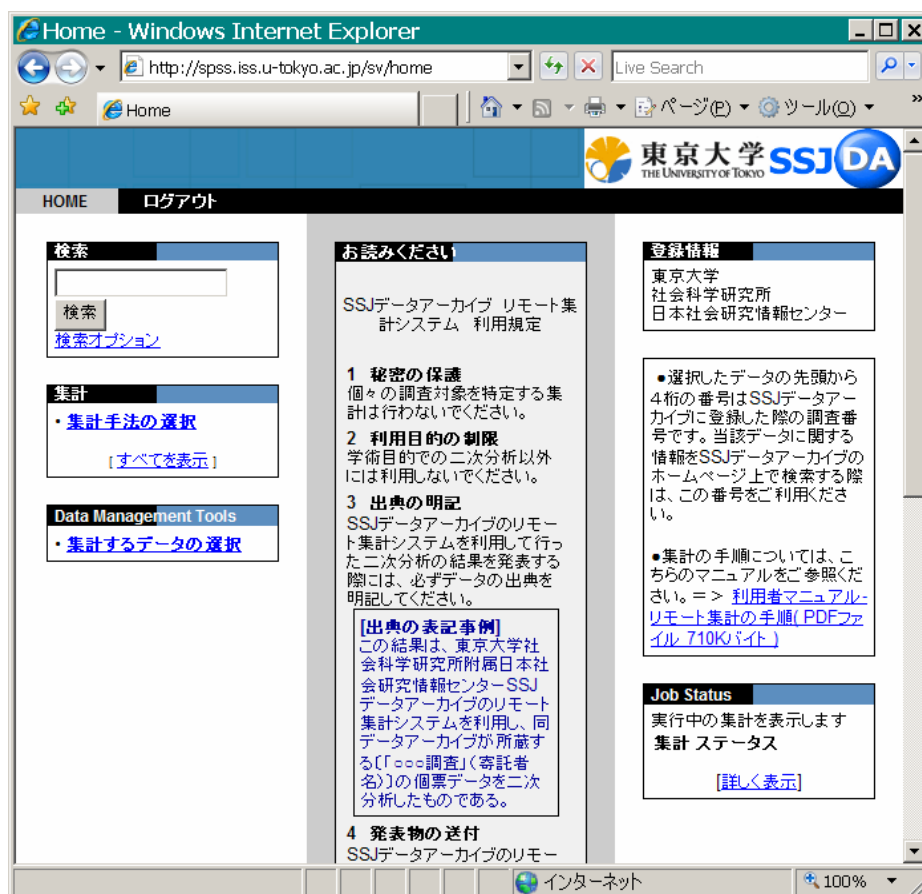


図 18 リモート集計システム利用画面トップ

分析は、「集計するデータの選択」の利用データを決めて、「集計手法の選択」で分析方法や利用変数を設定することで、データ分析結果をブラウザ上で実行することができる。

集計機能は、2011年5月現在、下記の6つを提供している；

1. 度数分布と関連統計量の集計
選択した集計項目の度数分布と関連する各種統計量を集計
2. 分類項目別度数分布の集計
選択した分類項目ごとの度数及び集計項目の合算値と平均値を集計（分類項目の選択は最大3変数まで）
3. 分類項目別度数分布の集計－2
選択した分類項目ごとの度数及び構成比のほか、 χ^2 検定の情報を集計（分類項目の選

択は最大3変数まで)

4. 分類項目別度数分布と関連統計量の集計

選択した分類項目ごとに集計項目の度数分布と関連する各種統計量を集計。(分類項目の選択は最大3変数, 行・列の最大2変数まで)

5. T検定

T検定の情報を集計

6. 相関係数

選択した項目間の相関係数を算出.

データを選び, 「集計手法の選択」で手法を選択すると, 手法に応じて変数等の設定画面が出てくる. 図19に, 「3. 分類項目別度数分布の集計-2」の画面を例示する. 行と列に対応する変数やオプション設定をして, 「集計」ボタンをクリックすることで, 分析結果を得ることができる.

集計対象データ: 0300-JGSS2002.sav

1.ウエイトの設定

2.分類項目

地域ブロック
都道府県名
市郡規模
年齢
年齢10歳刻み
アタック状況番号
生年
問01-1 先週の就労経験
問01-2a/b 就労日数/週
問01-2a/b 就労時間数/週
問01-2a/b 残業時間数/週
問02-1 就労地位
問02-2 業種
問02-3 職種
問02-4 従業員規模(就労事業別)

行
>> 性別
<<

列
>> 配偶者(DOMARRYの転記)
<<

層
>>
<<

3.統計量 (オプション)
 カイ2乗

4.表示順
 昇順 降順

5.
集計

図 19 リモート集計システムの分析設定画面の例

図20に分析結果画面を例示する. 「性別」と「配偶者」のクロス集計結果および Pearson のカイ2乗検定結果が表示される. 画面横に表示されるアイコンをクリックすることで, 当該 csv ファイルのダウンロードや, 表の単独ウィンドウでの表示等ができる.

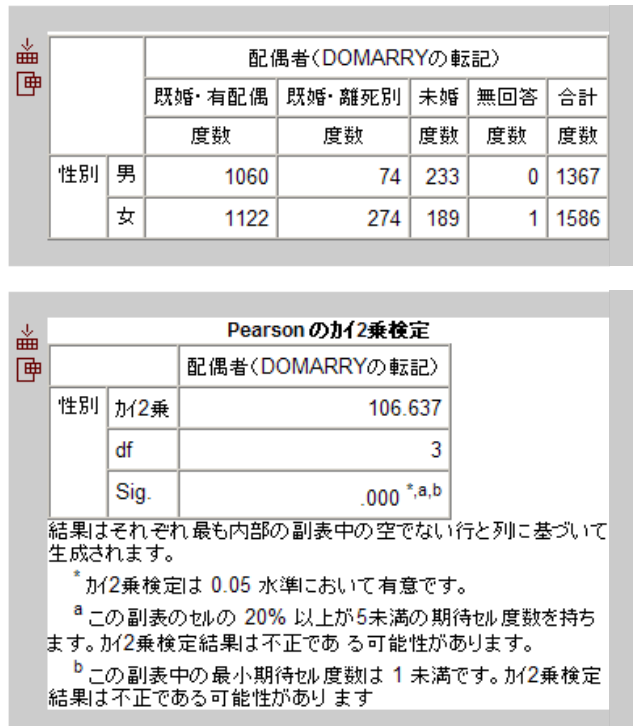


図 20 リモート集計システムの分析結果画面の例

その他、詳細な情報は、「利用者マニュアル-リモート集計の手順」として PDF ファイルを提供しているので、利用される際には参照されたい (図 18 右側にリンク有)。

また、リモート集計システムをデータ分析の演習に活用するための教科書も刊行されており、利用手順の紹介がある。(岩井・保田 2007)。

(4) 運用と課題

最後に、運用面からリモート集計システムの現状と課題について記す。

通常時のリモート集計システムの運用業務としては、システム用サーバーのメンテナンス、ホームページの管理、利用状況の管理、登載データの管理、が挙げられ、関連してシステム改善方策の検討がある。以下にそれらの概要を示す。

東京大学社会科学研究所では、システム管理室を設置し、この部署が研究所のネットワークおよびシステムのインフラ周りの運用・管理を行っている。リモート集計システムのサーバーも同室の管理下にある。

リモート集計システムのホームページに記載されている、利用規約、画面中の文言、利用マニュアル等の改定は、リモート集計システムを運用しているサーバーに保管されている当該ファイルを修正することで対応する。

利用状況は、月に一度のペースで、情報を管理している。表 5 に、2005 年 10 月から 2011 年 3 月までのリモート集計システムへのログイン件数の推移を整理した。2006 年に 909、

2007年に2572と利用が拡大し、2008年度に落ちこんだものの、2009年、2010年と2300～2400件で推移している。これまでのところ、大学の講義等と関係して、4～8月と10～12月あたりは利用が多いという傾向がある。

表5 リモート集計システムへのログイン件数の推移

月	2005年	2006年	2007年	2008年	2009年	2010年	2011年
1月	-	35	101	68	26	107	213
2月	-	49	49	9	5	42	123
3月	-	32	14	15	14	36	19
4月	-	66	102	295	493	264	-
5月	-	166	265	237	557	422	-
6月	-	113	256	144	163	362	-
7月	-	61	284	168	159	263	-
8月	-	106	604	43	35	33	-
9月	-	13	41	84	87	116	-
10月	75	174	209	178	290	331	-
11月	57	78	309	95	296	264	-
12月	68	16	338	138	133	232	-
合計	200	909	2572	1474	2258	2472	355

登録データの管理は、WEB上のSPSS WebAPPの管理者画面にログインして行う。新規データの追加は、SPSSファイルを用意して、図21に示すような管理画面において変数の種類（Analysis Variables, Groupby Variablesなど）を設定するなどの処理をすることで、可能となる。なお、管理ページにアクセスできるIPアドレスは、研究所内に限定している。

リモート集計システムについては、分析機能を拡大し、個票データの利用申請を軽減していく方向性と、本格的な分析は手元に個票データを入手して行うことを前提として、利用前のデータチェックに用いることを主眼におくという方向性の2つがある。

分析機能を拡大することには、相応の導入コスト等を要するが、利用者の利便性向上というメリットは大きいだろう。特に、非専門家にとって、商用ソフトウェアを購入せずに、実際の調査データを用いて複雑な分析が可能となることの意義は大きい。

一方で、統計処理に習熟した専門家は、適切なデータを入手できれば、本格的な作業は自身の分析環境で行うであろうし、社会調査等を学ぶ学生にとっても、導入としてリモート集計システムを用いることは有効としても、個票データを手元において、汎用的な分析ツールを用いて作業を行った方が、最終的には、本格的な学習に繋がる、というようにも考えられる。

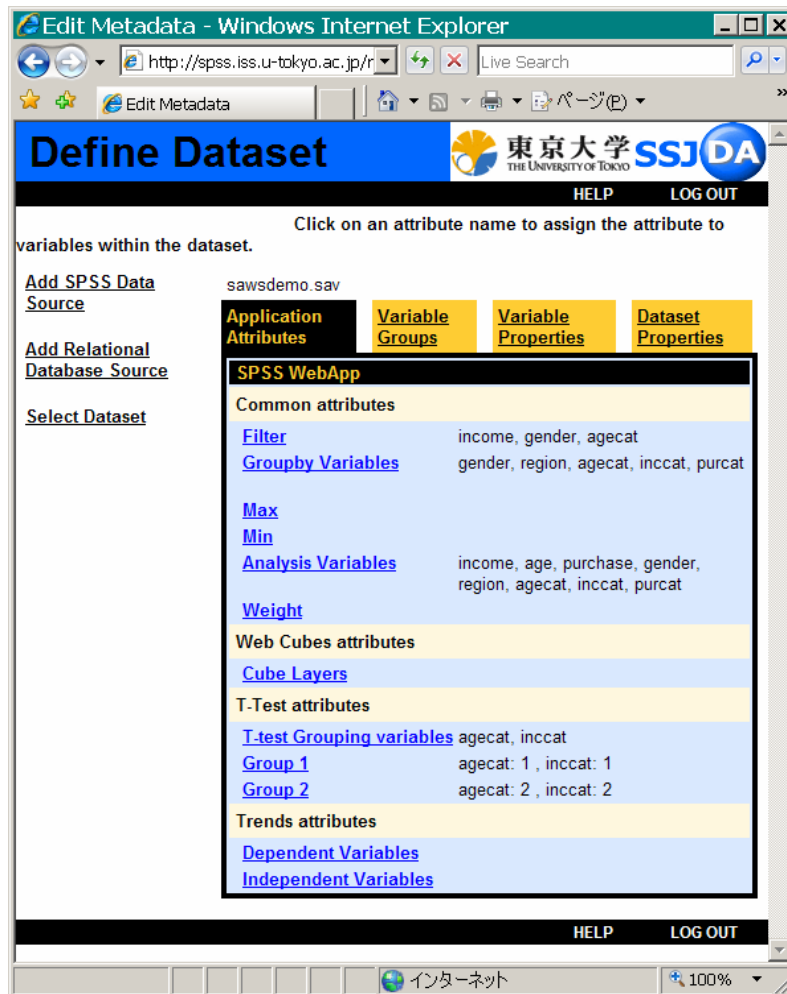


図 21 リモート集計システムの登録データ管理画面（変数種類の設定）

1.7 利用者データベースシステム

（1）開発経緯

SSJDA では、寄託データの増加に伴い、利用者が急激に増えてきたが、2005 年時点ではデータを提供する際の個々の管理ファイルがばらばらに存在しており、提供手続きが効率よく行われていなかった（佐藤朋彦 2006）。そこで、これらの管理ファイルのデータベースを構築し、提供手続きを正確に効率よく行うことを目指して、利用者データベースが開発された。開発は 2005 年中頃より着手され、2006 年より運用開始された。

具体的には、データの寄託者、提供している個票データ、利用者、成果物、それぞれの管理ファイルを統合する包括的なデータベースを開発し、データ提供作業の効率化や利用統計情報の集計作業を簡便化することなどの実現が目指された。

（2）システムの概要

データベースは、Microsoft Access の mdb 形式で作成され、データベースを管理するアプ

リケーションは Microsoft のスタンドアロン開発用プログラミング言語である Visual Basic を用いて開発されている。

システムとしての要求事項を整理すると、まず、データベースへとアクセス可能な範囲に制限を加えることがあり、具体的には、担当教員、特任専門職員、事務補佐員の一部までとすることがあった。それらの担当者により、画面ごとに次の情報入力・修正、閲覧、出力を可能とすることが図られた。

扱う情報としては、まず、寄託者情報（寄託者 ID、名称、部署、担当者名、郵便番号、住所、メールアドレス、電話番号、ファックス、承認手続きのチェックボックス、その他特記事項、最新更新年月日、変更前情報の保存）と提供データ情報（データ番号、データ名、寄託者 ID、公開年月日、利用目的のチェックボックス、受講者リストのチェックボックス、調査票数、個票数、報告書の複写許諾、リモート集計の許諾、その他特記事項、最新更新年月日、変更前情報の保存）があり、予め入力しておく必要がある。

利用申請があった際に、利用申請書にある氏名、所属大学・機関、学年・職名、連絡用メールアドレス、利用目的、受講者リスト提出予定日、共同利用者名、同所属、同職名、申請年月日、データ番号、申請番号等の情報を入力する。

その上で、担当教員のみアクセス可能とする利用審査画面で、申請ごとに内部での審査結果のチェックボックスや「決定」ボタンを押すと、審査結果や決済年月日が保存される。さらに、寄託者への検討依頼の文書を自動生成する「印刷」ボタンが用意され、寄託者から許諾をもらった後、寄託者決済年月日の入力と、「完了」ボタンによるステータスの変更がなされる。利用ステータスの遷移の詳細については、図 22 に示す。

データ提供画面では、申請ごとに「作成」ボタンや「発送」ボタンが用意され、提供年月日や使用期限年月日が自動入力される。データ利用報告画面では、申請番号ごとに「消去済」または「延長」のボタン、成果物の有無をあらわすチェックボックス、さらに成果物がある場合は、論文番号、申請番号、著者名、所属、職位・学年、タイトル、使用データ番号、雑誌・書籍・学会名、発表年月日、種別（雑誌/書籍/学会/その他）をあらわすチェックボックスに情報を入力する。

それらの一連の作業について、提供データ数、申請数、利用者数のカウントや、成果物統計の作成（論文数、論文に利用されたデータのカウントなど）などの集計結果をエクセル形式で出力できるような機能も用意された。

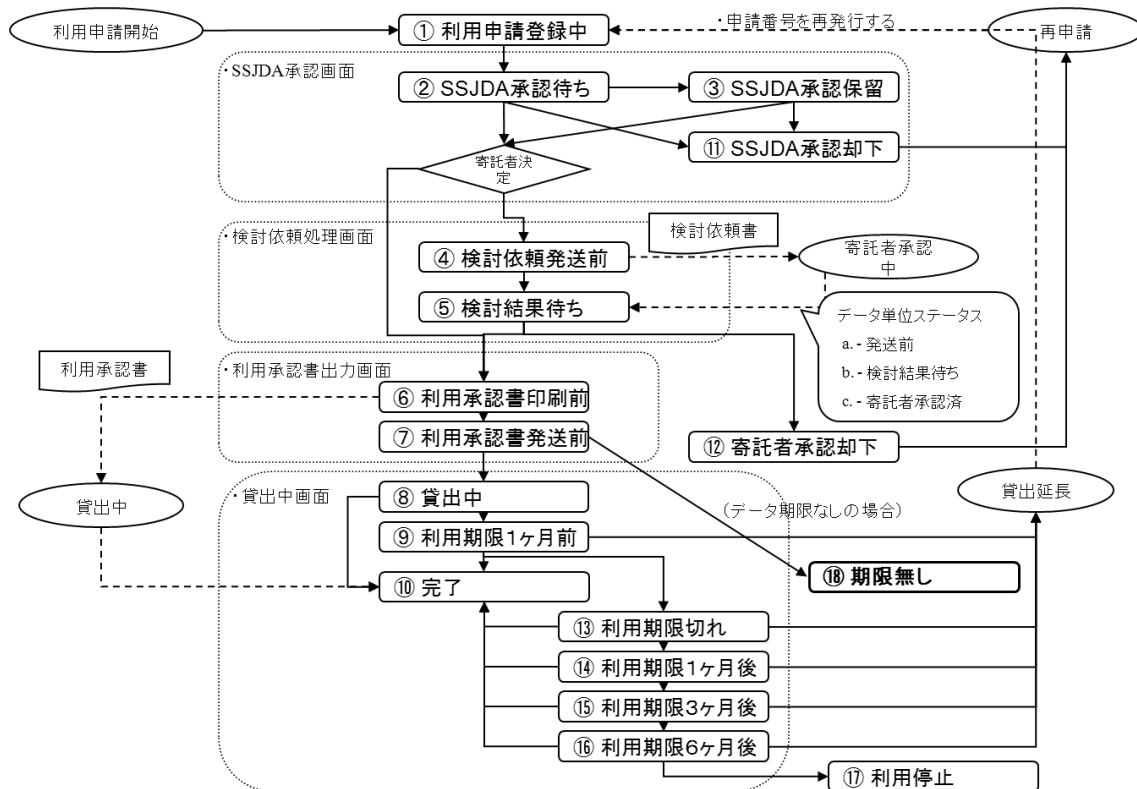


図 22 利用者データベースシステムにおける利用ステータスの遷移

(3) 利用画面

図 23 に利用者データベースシステムの画面構成を、図 24 にトップ画面を、図 25 に入力画面例として申請情報入力の画面を示す。

システムは、新規申請画面、登録中画面、担当教員による承認画面や検討依頼処理画面、利用承認書出力画面、貸し出し中画面、申請検索画面、成果物一覧画面、担当教員による利用統計作成画面から構成され、新規申請以外は、詳細画面や情報修正画面等を持つ。

アプリケーションを起動しトップ画面から、それらの画面へのアクセスが可能となっている。申請検索はトップから行う。実際のデータ入力や修正等を行う画面は、図 25 のような入力画面で、適宜情報を入力していくことで、アクセスデータベースが更新されていく。データベースは複数のテーブルから構成されており、id 等で関連づけられている。

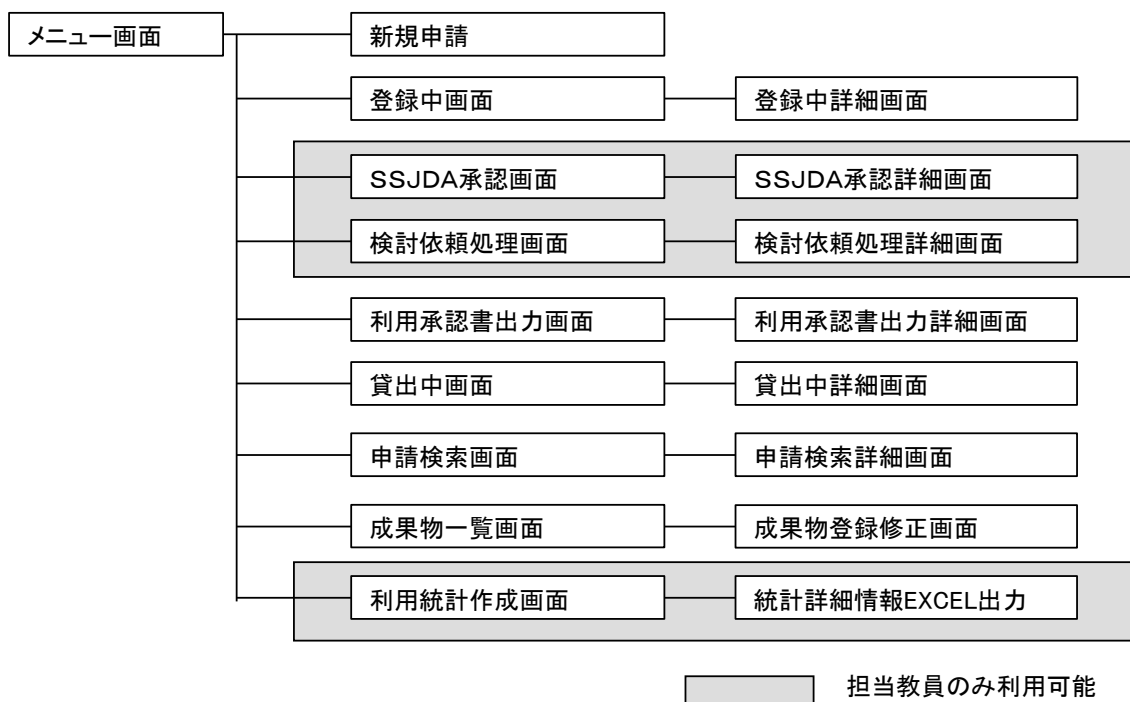


図 23 利用者データベースシステムの画面構成

東京大学 THE UNIVERSITY OF TOKYO

SSJDA利用者DB

A: 申請受付 E: 利用承認書

B: 登録中 F: 貸出中

C: SSJDA承認 G: 成果物


D: 検討依頼 H: 利用統計作成

申請番号開始 申請番号終了 名前 共同利用者を含む データ番号 ステータス

日付の種類 日付開始 日付終了 (YYYY/MM/DD) 検索 内容クリア

終了

図 24 利用者データベースシステムのトップ画面


東京大学
 THE UNIVERSITY OF TOKYO

★星印のついた項目は必須入力です。(登録中は除く)
 ※日付はYYYY/MM/DDの形式で入力して下さい。(例2006/03/24)

申請番号	旧申請番号	ステータス	登録日
★氏名	★大学・機関 / 所属	★職位	★メールアドレス
指導教員氏名	指導教員所属	指導教員メールアドレス	
申請年月日	★利用目的 ● 研究 ● 教育	受講者リスト提出日	データ提供日
報告予定日			
データ追加			データ番号
追加番号	寄託データ番号	寄託データ名	追加
			削除
共同利用者追加			
氏名	所属	職位	
追加番号	名前	所属	職位
			追加
			削除

キャンセル 内容クリア 印刷 登録中 登録

図 25 利用者データベースシステムの申請情報入力画面

(4) 課題と 2011 年度の取組み

利用者データベースシステムは、利用申請から成果報告までの一連のプロセスにおける情報管理の一元化を行ったが、その前段階で用意される寄託者情報や調査リストは、別にエクセルで管理されており、データ寄託時や変更ごとにエクセル情報を更新した後に、アクセスでデータベースの情報も連動して更新させる必要がある。

担当助教やデータアーカイブ室での情報管理が煩雑であるため、担当者のニーズを踏まえて 2011 年度中に若干の調整が加えられる予定である。具体的には、別に管理されていた寄託者情報のエクセルファイルをなくして、利用者データベースシステムの方で寄託者情報の追加・編集を行い、寄託者情報管理の一元化を行うことと、調査データ別に寄託担当者情報を一覧したり、寄託者への郵送ラベルやドキュメント作成機能を追加することが検討されている。

2. SSJDA の情報システムの構造的問題と DDI 利用可能性検討の経緯

2.1 SSJDA の情報システムの構造的問題

前章で見てきたように、SSJDA には、データアーカイブ室での担当スタッフによるデータクリーニングや公開用データの作成などの進捗管理を含めた調査リストや寄託者情報があり、公開されたデータは、ウェブサイト、Namazu による検索システム、データダウンロードシステム Direct、リモート集計システムと複数のシステムに搭載される。さらに、利用申請から承諾、成果報告までを管理する利用者データベースシステムが動いている。これらは、全体的なデザインのもとに順次相互の連携を意図しながら構築されてきたものではなく、任期付の教員がそれぞれ個別に構築を進めてきたものであった。それが故に、各システム間での連携が十分でなかったり、これまでこのリサーチペーパーで扱う DDI への検討が未着手であった等、全体的な見通しに欠ける部分もあった。

先に述べたように、現在、利用者データベースシステムの寄託者情報管理機能を拡充することで、これまでエクセルファイルで別に管理されていた寄託者情報を一元化することが取り組まれている。部分的にでも、個別のシステムや情報の連携不足を解消していくことは、作業効率性を上げることに繋がっていくが、一方で、マイクロソフト Access のデータベース形式でそういった情報を、これまでの延長線上で蓄積していくことは、将来的に他のシステム等との連携も考えたときに、フォーマットの面でも、扱う情報の項目の面でも、適切であるのかやや心もとない面もある。

後で触れるが、DDI3 では「データライフサイクルモデル」という、調査設計からデータアーカイビングまでに扱うメタ情報を一括して XML 形式で整備して、さらにデータアーカイブ等で活用していくことが提唱されている。膨大な情報を扱うデータアーカイブでは、機能別に個別にシステムを構築し、多様なフォーマットで情報を管理していると、その連携をさせることにも膨大な労力が発生する。DDI 対応には、一見コストが多いように見えても、何らかの統一された情報記述方法で、データアーカイブが利用するあらゆる情報を一元的に管理しておき、そこからデータクリーニングや寄託者・利用者対応業務に必要な情報を取り出したり、さらにはウェブサイトやデータ検索、オンライン集計やデータダウンロードシステムの情報が自動更新されていくような仕組みを考えることが、最終的には作業効率を高め、必要なメタデータが適切に蓄積されていくことにつながるものと考えられる。このような作業には、データアーカイブで扱っている情報を網羅的に把握しつつ、情報技術の展開にも適切な見通しを持った上で、総合的なシステムデザインを図り、事を進めていく必要があるものと考えられる。

2.2 DDI 利用可能性検討の経緯

前小節での認識を持った上で、まずは DDI について、その利用可能性を検討することが適切であると判断した。2010 年夏ごろより、佐藤を中心にアーカイブ内で非公式に「DDI

利用可能性検討会」を組織した。検討会は、情報技術に明るい角井，米倉，入山，安藤，朝岡各氏とともに、国際的なデータアーカイブやその技術動向について詳しい前田を加えて組織し、下表に示すように月1度程度集まり、DDI関連文書の読み込みや情報収集、Nesstarのトライアルバージョンのインストールと利用実験、海外のDDI関連ソフトウェアの試験利用、日本語での調査メタ情報管理ツールについての意見交換と基礎設計やシステムエンジニアへの発注作業等を、約半年間で進めた。最終段階で、東北地方太平洋沖地震が発生したことに起因して、リサーチペーパーの取りまとめが若干遅れたものの、DDIについての基礎的な情報収集は完了し、いくつかの試験的な作業も行え、今後のSSJDAでのDDI対応に向けて十分な検討が行えたものと考えている。

表 6 DDI 利用可能性の検討経緯

回数	月日	内容	報告者等	内容詳細
第1回	2010/8/10	キックオフ	佐藤慶一	検討背景，作業目的，作業分担等
			角井佑	Nesstar Webviewの操作画面の変更方法等
第2回	2010/9/17	分担作業報告	角井佑	Nesstar Publisherの使用方法等
			米倉佑貴	DDI関連文書とソフトウェアの紹介
			朝岡誠	基礎的な情報収集
第3回	2010/10/21	分担作業報告	角井佑	Nesstar・SDAのライセンス料について
			米倉佑貴	DDIで記録可能なメタデータ等
			朝岡誠	Microdata Management Toolkitについて等
第4回	2010/11/18	分担作業報告	角井佑	Nesstarバージョン4β版について
			米倉佑貴	DDI文書作成・編集用の既存ソフトウェアのレビュー
			安藤理	SDAとNesstarの利用方法，韓国・台湾での状況
			佐藤慶一	ドイツDDI Workshopの報告
第5回	2011/11/22	EDO開発	佐藤，米倉，入山，SE会社	基礎設計の調整
第6回	2010/12/27	分担作業報告	角井佑	Nesstar Publisher 4.0でのデータインポート作業等
			米倉佑貴	EDOの開発経過等
			朝岡誠	Microdata Management Toolkitについて等
			佐藤慶一	リサーチペーパーの執筆分担等
第7回	2011/12/28	EDO開発	佐藤，米倉，角井，入山，SE会社	基礎設計の調整
第8回	2011/1/27	リサーチペーパー	全員	原稿進捗報告，分担調整
第9回	2011/2/16	EDO開発	佐藤，米倉，角井，入山，SE会社	開発状況の報告
第10回	2011/3/4	リサーチペーパー	全員	原稿進捗報告
第11回	2011/3/23	EDO開発	佐藤，米倉，角井，入山，SE会社	2011年度分の開発作業の報告等

[参考文献]

榎田直木，2006，「楽屋話～品質は個票データに宿る」『ESTRELA』no.147：pp.32.

- 佐藤朋彦, 2006, 「私が担当した3つの仕事」『ESTRELA』 no.147 : pp.33.
- 佐藤博樹, 2008, 「SSJ データアーカイブの現状と課題」『統計』 2008年12月号, pp.10-15.
- 佐藤慶一, 2009, 「SSJ データアーカイブにおけるデータダウンロードシステムの開発」
『ESTRELA』 no.185, pp.20-23.
- 佐藤慶一, 2008, 「SSJ データアーカイブにおけるリモート集計システムについて」
『ESTRELA』 no.175, pp.2-9.
- 岩井紀子・保田時男編, 2007 「調査データ分析の基礎」有斐閣.

第2章 情報技術と統計メタデータ：DDIについての概観

前田 幸男

1. はじめに

情報技術の急速な発展により、社会科学データの提供方法は大きな変容を遂げつつある。インターネットを利用したデータの提供が半ば常識となりつつある一方、データ入手の前提となる調査概要などの情報を提供するための規格の統一が進みつつある。データ・アーカイブを通じたデータの利用件数は爆発的に増加しているが

¹、公文書館においては、統計分析のためではなく、歴史資料として、様々な電子データが利用されるようになってきている (Adams 2007)。従来は必ずしも緊密な連携関係になかった、公文書館、図書館、社会科学データ・アーカイブといった組織間においては、情報共有のための規格の平準化（共通化ではない）が進捗しつつある (Doorn and Tjalsma 2007)。

欧米諸国の主要なデータ・アーカイブが設立されたのは1960年代のことであるが、設立した直後から、各国のデータ・アーカイブの間で如何に利用可能なデータについての情報を共有し、また、相互にデータの提供するかについては、政治学者・社会学者を中心に議論が行われていた (Scheuch 2003)。また、国際比較分析を促すためには、調査単位ではなく、分析対象となる概念あるいは変数（質問文）を単位とした検索およびデータ入手のシステムが必要ということも指摘されていた (Scheuch and Stone 1964; Mochmann 1974-75)。

しかしながら、1990年代半ばまでは、簡単な調査概要と印刷されたコードブックのみが、データについての情報提供の手段であった。社会調査データを用いた国際比較、あるいは、一国における時系列の比較を行うためには、何冊ものコードブックを丁寧に見比べ、文言を確認する必要がある必要があったのである。そのような状況を考えるならば、変数あるいは概念・キーワードを利用して大量の社会調査データを横断検索するシステムは、「社会科学における夢の機械」であったと言えるだろう (Ryssevik and Musgrave 2001)。そして、インターネットの普及並びに情報技術の発達、そのような「夢の機械」を半ば現実にしつつあるように思える。

ただし、大量の情報が無秩序に存在しても、その効果的な利用は難しい。利用者が膨大なデータを検索して有益な情報を抽出するためには、事前に情報を整理し提供するための仕組みの構築が不可欠である。最近の情報技術の発展を取り込んだデータの検索・提供システムを構築するため考案され、統計メタデータの国際標準規格となりつつあるのが、DDI (Data Documentation Initiative) である。本稿は、データ提供方法の変化を理解する上で重要なDDIについて、その発展の背景を素描することを目的とする。²

¹ <http://www.icpsr.umich.edu/icpsrweb/ICPSR/curation/usage.jsp>

² 本稿では基本的に社会調査の個票データについての統計メタデータに集中する。実は、統計

2. メタデータとは—ダブリンコアと DDI

メタデータ (metadata) とは、日本語においては聞き慣れない言葉かもしれないが、データ・アーカイブあるいは情報検索システムの研究に携わる人々の間では日常用語になりつつある。メタデータという言葉が一般的に流通し始めるようになったのは、2000 年代に入ってからだと思われるが、実質的には 30 年以上も議論の対象であった (Carol A. Hert et al. 2004)。一般的には、メタデータとは、「データについてのデータ」を意味しており、字義通り、「対象となるデータに対して、そのメタレベル (高次レベル) の記録と捉えられることに由来」している (谷口 2010)。本稿の文脈においては、主に統計数値データを理解するために必須となる情報を、メタデータという。具体的には標本抽出の手続き、調査方法、質問文、コーディングの手続き等のデータについての記述内容を指す。かつては、それらのメタデータは冊子体のコードブックとして提供されるのが通例であった。データが機械可読形式で保存されていたとしても、コードブック、あるいはメタデータが無ければ、それは無用の長物である。メタデータがあつて初めて、人間は、データの内容を適切に理解できる。

数値データだけでなく、コードブックも機械可読形式で提供されていれば、検索は容易になる。しかし、文字情報のみが保存されたファイルの提供であれば、コンピュータによる検索は情報の内容如何を問わず検索対象の単語を全て拾い出すことになり、非効率的である。同様に、インターネットのサーチエンジンは膨大な情報を探し出すが、HTML ファイルについては、メタデータを構造化する基準がないため、不要な情報ノイズが大量に検索結果に入り込んでしまう。それに対して、メタデータを設計し、必要とされるデータ項目を事前に定義することで、精度の高い検索システムを実現できる (谷口 2010)。とりわけ、統計メタデータについては、データ記述の項目数が質量ともに多くならざるを得ないために、事前の整理・分類の有無はその後の利用しやすさに大きな影響を与える (金子 2002)。すなわち、セマンティックウェブの長所を生かし、利用者が必要とするデータの検索・提供を可能にするためには、メタデータの統一規格が必要なのである (Carol A. Hert et al. 2004)。

汎用性の高いメタデータの国際基準であるダブリンコア (Dublin Core) は、領域横断的に利用されることを念頭に置き、対象物に対する情報の発見やその取得を支援するために設計されている。ダブリンコアでは、対象を識別するために 15 のメタデータ項目が定義されているが、個別領域におけるメタデータはダブリンコアとの対応関係を念頭に設計され、共通の枠組みの中で検索のための仕組みが構築できるように工夫される (杉本 2009; 谷口 2010)。

ただし、ダブリンコアは対象となる情報資源とは独立して考案された枠組みであるため、

メタデータを考察する上では、政府統計の提供に関連して集計データのメタデータとして SDMX (Statistical Data and Metadata Exchange) が重要であるが、ここでは、筆者の能力の問題もあり、統計メタデータを考察する上で重要なもう一方の流れについては割愛する。

統計メタデータの仕様としては不十分である。DDIは、XMLに基づいた、社会科学・行動科学データの内容記述、保存、移植を容易にするために統計メタデータの国際基準を作ろうとするプロジェクトであり、DDIのデータ項目もダブリンコアとの対応付けがなされている (Vardigan et al. 2008)、両者のメタデータ要素を一覧にしたのが、表1である。

表 1 ダブリンコアと DDI (Ver 2) の項目対応表

ダブリンコアのデータ要素(項目)	日本語名	DDI Ver.2 のデータ要素(項目)	DDIにおける注記
Title	タイトル	<titl> 2.1.1.1	Title of Data Collection
Creator	作成者	<AuthEnty> 2.1.2.1	Authoring Entity of Data Collection
Subject	主題(キーワード)	<keyword> 2.2.1.1	Keyword(s)
		<topcClas> 2.2.1.2	Topic Classification
Description	内容記述	<abstract> 2.2.2	Abstract
Publisher	公開者	<producer> 2.1.3.1	Producer of Data Collection
Contributor	寄与者	<othId> 2.1.2.2	Other Identification/Acknowledgements - Data Collection
Date	日付	<prodDate> 2.1.3.3	Production Date - Data Collection
Type	資源タイプ	<dataKind> 2.2.3.10	Kind of Data
Format	記録形式(フォーマット)	<fileType> 3.1.5	Type of File
Identifier	資源識別子	<IDNo> 2.1.1.5	ID Number - Data Collection
		<holdings location="" callno="" URI=""> 2.1.8	Holdings Information - Data Collection
Source	出処	<sources> 2.3.1.8	Sources - Used for Data Collection
Language	言語		
Relation	関係	<othrStdyMat> 2.5	Other Study Description Materials
Coverage	時空間範囲	<timePrd> 2.2.3.1	Time Period Covered
		<collDate> 2.2.3.2	Date(s) of Data Collection
		<nation> 2.2.3.3	Country
		<geogCover> 2.2.3.4	Geographic Coverage
Rights	権利情報	<copyright> 2.1.3.2	Copyright - Data Collection

出典: DDIウェブサイト (<http://www.ddialliance.org/>)に掲載されている表を谷口(2010)を参考に修正

統計データを理解するために必要なメタデータの具体的な内容であるが、グレゴリーらによれば基本的には二つの内容に大別できる (Gregory et al. 2009)。第一は、主に数値を統計ソフトウェアに読み込ませるために必要となる情報であり、データ・ファイルが固定長形式か自由長形式か、1観測対象に1行が割り当てられるフラット形式かそれとも階層構造をなす形式か、等がそれにあたる。前者がデータのソフトウェアによる読み込みに必須であり、メインフレーム・コンピュータの時代からデータ・ディクショナリ (Data Dictionary)、あるいはデータ定義文 (Data Definition Statement) として提供されていた。

それに対して、もう一種のメタデータは、人間が読んで理解するために準備される。データの内容に関連する情報であり、標本抽出手続きや調査方法等が主で、その他、質問文や選択肢の内容、データ収集過程や公開準備の過程でデータに加えられた変更に関する記述、さらには調査データを利用した報告書や出版物の内容まで多岐にわたる。ただし、これらのデータに関する記述と統計ソフトウェアとの関連づけは、1980年代まではほとんど問題とならなかったようである。以下では、統計メタデータ特有の問題について、歴史的経緯を踏まえながら、考察する。

3. 統計ソフトウェアとメタデータ

統計ソフトウェアとメタデータの関係であるが、1970年代にメインフレーム・コンピュ

ータ上で稼働していたデータ管理及び統計分析のソフトウェアである OSIRIS³は、限定的ながら調査関連情報を含むコードブックを機械可読ファイルとして扱うことができた (Miller and Vardigan 2005). 1970~80 年代においては、OSIRIS の形式を前提に統計メタデータの基準が作成されていたようであるが、各アーカイブにおいて独自の修正がなされることが多く、統一性は低い状態であった。特に 1980 年代はパーソナル・コンピュータが普及したことで、各アーカイブや大学で様々なシステムが開発されたが、統一されたメタデータの規格が欠落したことで、データの検索や提供において様々な問題が生じていた (Gregory et al. 2009; Rasmussen and Blank 2007).

SPSS や SAS といった通常の統計パッケージによるメタデータへの対応は、さらに不十分であった。1990 年代には変数の補助情報 (Variable Notes) や、ファイル全体の情報 (File Notes) について、数値データと別にメタデータを保存する必要性は議論されていた (Blank 1993). 今日統計ソフトウェアのメタデータへの対応は、若干の改善は見られるものの、計算能力の向上と比べて、不十分と言わざるを得ない (Rasmussen and Blank 2007). 汎用統計ソフトウェアは、変数名と数値についてそれぞれ限定されたバイト数のラベルを許容するのみで、調査概要など、データセットの全体像を把握するために重要な情報を保存することはできない。あくまで数値の集計のみを前提にソフトウェアが設計されており、比較的最近までコードブックを自動作成することもできなかった⁴。例えば、選択肢が存在しても頻度がゼロの場合は出力に出てこない。また、欠損指定にしても、異なるタイプの欠損値を区別できないという難点がある。端的に言えば、統計ソフトウェアは、利用者側の視点からは、統計データとメタデータとの関連づけが不十分なのである。また、統計データとメタデータとの関連づけが不十分であると、刊行されたデータ分析の結果と、そのもととなったデータとの関連づけが困難になり、研究結果の再現性の観点からも大きな問題となる (Micah Altman et al. 2001).

従って、データの集計に関連した仕組みと、メタデータの検索や必要な情報の参照を有機的に組み合わせることで研究を効率化することが重要な課題となる。そこで、DDI では様々な組織が所蔵・提供する統計データが統一的・効率的に検索され、かつ、利用に供されると同時に、分析結果の公表・引用を円滑化することを念頭にそのメタデータ要素が設計されている (Blank and Rasmussen 2004; Miller and Vardigan 2005; Rasmussen and Blank 2007; Vardigan et al. 2008).

4. DDI の実践例

インターネットの利用が一般化した 1990 年代の後半はインターネット上での利用促進のために様々なシステムが構築された。大学で開発されたシステムに限っても、カリフォル

³ ミシガン大学の社会調査研究所で開発・維持されていたメインフレーム・コンピュータ上の統計データの管理・分析ソフトウェア。

⁴ SPSS や Stata では現在ではコードブックを機械的に作成するコマンドが準備されている。

ニア大学バークレー校で SDA (Survey Documentation and Analysis⁵) が開発され、ハーバード大学と MIT が共同で運営するデータ・センターで virtual data center (現在の Dataverse Network⁶) の運営が始まったのも同時期である。その中でも DDI に準拠して構築されたシステムが、Nesstar である (Ryssevick and Musgrave 2001)。ここでは CESSDA (Council of European Social Science Data Archive⁷) に加盟しているアーカイブのカタログを横断して検索するシステムである CESSDA catalogue⁸ に応用された例を紹介する。

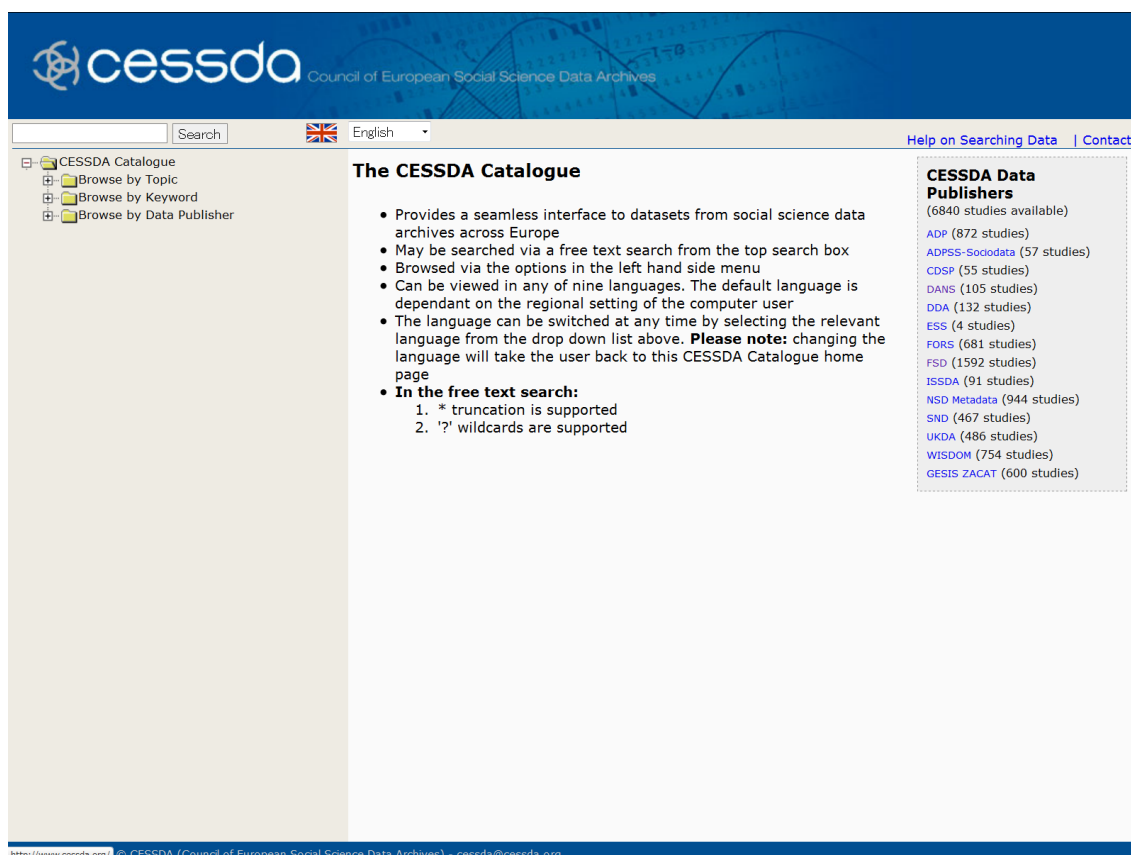


図 1 CESSDA Catalogue の入口

図 1 に示したのが、CESSDA Catalogue の入口の画面であるが、右脇の表示から 14 のデータ・アーカイブから合計で 6840 件のデータが利用可能であることが示されている。なお、これは 14 のアーカイブが持つデータの全てではなく、その一部であることは注意を要する。全てのデータを Nesstar 上に搭載しているアーカイブもあれば、その一部だけを搭載しているアーカイブもあり、対応はまちまちである。比較的新しい FSD が最多のデータセット数

⁵ <http://sda.berkeley.edu/index.html>

⁶ <http://thedata.org/home>

⁷ <http://www.cessda.org/>

⁸ <http://www.cessda.org/accessing/catalogue/>

を示しているのに対して、1960年代から活動をしている UKDA や GESIS のデータセット数が少ないことから、古いデータセットを全て Nesstar に搭載していないことがわかる。

図 2 に、CESSDA Catalogue の最初の画面で「vote」と入力した場合の検索結果を掲載している。調査やデータ・ファイルではなく、変数を単位とした検索結果が示されているが、最右列を見ると上から FSD (フィンランド)、SND (スウェーデン)、GESIS (ドイツ) が所蔵するデータに含まれる変数が拾い上げられていることが分かる。また、下段の表示から、欧米語のシソーラスから「vote」に対応する各国語(独語、仏語、オランダ語、スペイン語、ノルウェー語、スウェーデン語、フィンランド語)でも検索が行われていることが分かる。

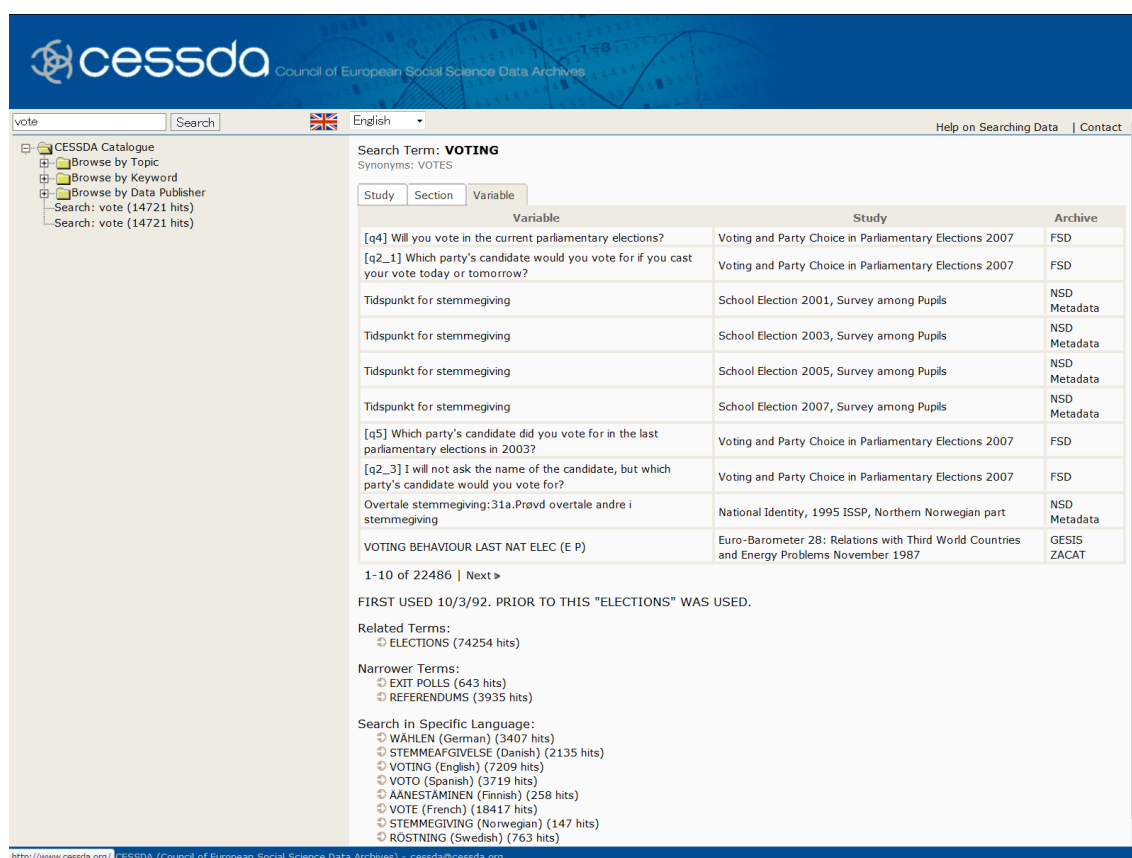


図 2 CESSDA Catalogue における検索結果

次に、一番上に示されている Finnish National Election Study 2007 を選択し、クリックすると、FSD (Finish Social Science Data Archive) のウェブサイトへと移行する。リンク先の FSD のウェブ上では直接 FNES2007 のメタデータが表示され、さらなる閲覧・検索へと円滑に移行できる。若干の閲覧後の画面を図 3 として掲載している。画面の上のロゴや色彩が違うのでわかりにくいかもしれないが、FSD のシステムも Nesstar を利用している⁹。

CESSDA Catalogue は各アーカイブで整理されているメタデータを自動的に読み込むので、

⁹ <http://www.fsd.uta.fi/english/index.html>

欧州のアーカイブを横断した検索が可能になると同時に、さらにデータの詳細を調べたり、データを手入したりするために利用者認証が必要となる段階においては、各国のアーカイブのシステム上からの利用へと誘導するように設計されている。

The screenshot shows the 'YHTEISKUNTATIETEELLINEN TIETOARKISTO' (Finnish Social Science Data Archive) interface. The main content area displays the following information for the dataset 'Voting and Party Choice in Parliamentary Elections 2007':

- Study Description**
- Bibliographic Citation**
- Title Statement**
- Full Title**: Voting and Party Choice in Parliamentary Elections 2007
- Parallel Title**: Äänestämisen ja puolueiden valintaperusteet eduskuntavaaleissa 2007
- Identification Number**: fi.fsd.ddi.2275e
- Responsibility Statement**
- Authoring Entity**:
 - Yleisradio
- Production Statement**
- Producer**:
 - Yleisradio. YLE news
- Copyright**: According to the agreement between FSD and the depositor.
- Distributor Statement**
- Data Distributor**:
 - Finnish Social Science Data Archive Abbreviation: FSD
- Depositor**:
 - Ainola, Olli
 - Affiliation: Yleisradio, YLE news

At the bottom of the page, it says 'Copyright © 2001-2006 NSD - All Rights Reserved - Contact Nesstar Support' and 'POWERED BY nesstar'.

図 3 Finish Social Science Data Archive の Nesstar 画面

なお、Nesstar にはリモート集計システムとしての側面もあるが（前田 2008）、本稿の文脈で言えば、変数－値ラベルと数値データとの関係がとれることで、リモート集計が容易となっていることを指摘しておく。この点、通常の統計パッケージは、手元にコードブック、最低でも調査票を持っていなければ分析が極めて面倒になるが、Nesstar 上でいわばコードブック（メタデータ）と数値データとの間の断絶を気にすることなく、基本的な集計を行うことができる。Nesstar が、開発者が銘打ったように「夢の機械」（Ryssevik and Musgrave 2001）と言うに値するかは評価が分かれると思われるが、従前よりも格段に利便性が向上したことは事実であろう。Nesstar の他の応用事例に関しては、本リサーチペーパーの安藤・朝岡、SSJDA における試験導入については、角井・入山を参照頂きたい。

以上は、データ・アーカイブの利用者側から見た例であるが、データのライフサイクルの他の局面における応用についても簡単に紹介しておきたい。データの収集を行う組織の側からみると、保存・対外提供用のデータを作成する過程で如何に DDI の要素に対応した

内容を効率的に保存し、移植・伝達するかが問題となる。(Gregory et al. 2009). 実地調査でコンピュータが利用される CATI (Computer Assisted Telephone Interview) で実施された調査については、質問文や選択肢は一旦機械可読形式となるので、それらのメタデータを、数値データとともに、DDI 形式に移植する仕組みの開発が行われている。また、定型的に使われる質問や選択肢については、メタデータの再利用も可能である (Rasmussen and Blank 2007).

ドイツの GESIS (German Social Science Infrastructure Services / Gesellschaft Sozialwissenschaftlicher Infrastruktureinrichtungen) では、マイクロセンサス・データのメタデータを DDI に沿って作成している。また、ドイツでは GESIS が積極的に利用促進の活動をしていることもあり、政府の統計を学術目的に利用するために整備されつつある RDC (Research Data Center)¹⁰においても、DDI が利用されている他、IAB (Institute for Employment Research)¹¹, IZA (Institute for the Study of Labor)¹²においても部分的には DDI に準拠してデータの整理が行われているようである (Gregory et al. 2009).

一方、データの検索段階では、DDI はダブリンコアだけではなく、図書館目録の基準である MARC との対応も念頭に置いて設計されているので (Vardigan et al. 2008), DDI に従ってメタデータを蓄積すると、データ目録を図書情報検索の枠組みに流し込むことも可能になる。メタデータ規格がダブリンコアを結節点として相互に共通であれば、例えば同一著者の書籍とデータを同時に検索することもできる。実際、ミシガン大学の図書館目録検索システム上では、ICPSR のデータも検索結果として拾い上げられるようになっている¹³。図 4 に示すのは、ミシガン大学の OPAC 上で、「JGSS」と「Michio Nitta」の2つをキーワードとして利用した際の検索結果である。

画面からはわかりにくいですが、仁田道夫が著者あるいはデータ作成者として登録されているものとして、JGSS (日本版総合的社会調査) に関連した書籍 (谷岡一郎, 仁田道夫, 岩井紀子編『日本人の意識と行動: 日本版総合的社会調査 JGSS による分析』東京大学出版会 2008) およびデータ・ファイルの両者が出力されていることがわかる。このように、異なる資源領域において共通の枠組みでメタデータを準備できれば、1つのキーワードにより複数の資源領域を横断した検索が可能となるのである。現段階では、まだ、書誌情報の検索と、データ・ファイルの検索システムの相互連携が十分進んでいるとは言いがたいが、研究結果の再現性を担保するという観点からは、両者の連携強化は重要な課題だと思われる。

データが二次利用される段階においても、再現性の担保という観点からは、二次分析者

¹⁰ <http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/EN/Navigation/Projekte/FDZ/ResearchDataCenter.psml>

¹¹ <http://www.iab.de/en/iab-aktuell.aspx>

¹² http://www.iza.org/en/webcontent/index_html

¹³ <http://mirlyn.lib.umich.edu/>

自身もメタデータを作成する必要がある。二次利用者がメタデータを作成し、それをデータ・アーカイブあるいはデータ収集者にフィードバックする仕組みについても検討が行われているようであるが、筆者の見聞の範囲内では、具体的に紹介できる例を見つけることができなかった。

The screenshot shows the Mirlyn Catalog interface. The search criteria are 'author:nitta michio AND; title:jgss'. The results are sorted by date (newest first). The left sidebar contains filters for Subject, Academic Discipline, Format, Language, Place of Publication, and Date of Publication. The main content area displays three search results, each with a 'Select' button and a table of location, status, and call number/description.

図 4 ミシガン大学図書館上の検索結果

5. おわりに

統計メタデータは、データ作成者・収集者と二次利用者とを媒介するものとして重要である。従前は基本的に一次データ収集者からメタデータを受け取り、必要な編集を施した上で、二次利用者に提供するのはデータ・アーカイブの任務と考えられてきた。今日でも基本的にその構図に変化はないが、データ・アーカイブが調査終了後に研究者からメタデータを入手して整理する場合、情報の遺失があると同時に、作業にも重複が多くなる。その観点から、近年ではデータのライフサイクルという考えを強調し、メタデータの作成・記録・保存をデータに関連した全ての過程と関連づける必要があると主張されている。また、DDI は、データのライフサイクルの各段階でメタデータの蓄積・移転を行うことで、データの質を向上させることも重要な目的としている。

しかしながら、著者が過去数年見聞している限りでは、研究者が DDI を効率的に利用し、

メタデータの蓄積をするソフトウェアの開発が追いついていない印象を受ける。商用及び無償のソフトウェア開発が進んでいるが、十分に普及していないのが実情である。この点、日本における社会調査データの提供を考える上で、DDI および関連するソフトウェアの動向を随時把握する必要があるだろう。また、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でも、DDI に準拠したメタデータ蓄積を支援するためのソフトウェアである EDO (Easy DDI Organizer) の開発を試験的に行っている。本リサーチペーパーの米倉・佐藤を参照されたい

その一方、DDI に基づいたメタデータの蓄積は、それ自体が目的ではなく、データが効率的に検索・利用され、かつ、研究を質量ともに発展させるための手段に過ぎない。その意味では、社会調査データの学術目的の共有が制度化されていなければ、DDI の導入自体意味をなさない。データのライフサイクルという発想に基づき、データの一次作成者がメタデータを蓄積し、その上で、データ・アーカイブがそのデータを二次利用に供するためには、データの保存と二次分析のための提供とが研究者の間で常識となる必要がある。日本において過去 10 年ほどの間に事態は相当改善したと思われるが、まだ、調査データの保存と学術目的提供が常識となったとは言えないように思われる。この点、筆者は、学会の査読誌の掲載条件に、再現性を担保するためにデータを公開すること、あるいは、研究助成金の条件として、データの保存と公開とを義務づけることが必要ではないかと考えている。

研究者にとって時間は極めて重要な資源であり、できるだけ時間の節約を図りたいという誘因を持つことはやむを得ないことである。ただし、何らの強制も課されない状態であれば、仮に最初は保存・公開するつもりがあったとしても、時間の経過とともに、その気力さらにはデータそのものも失うことも場合によっては生じるのではないか。その意味では、強制力を伴ったルールに基づく運用が長期的には学問全体の利益になる。筆者自身、他者からの強制により作業することの嫌悪感は深く感ずる。しかし、一定の強制と締切が無ければ、作業が進みにくいことも確かである。狭い意味では DDI の可能性から離れた論点であるが、DDI のような取り組みが波及効果を持つためには、学術目的で収集された社会調査データの保存と二次利用についての政策的対応が必要であろう。

実際、DDI と関連する形で行われている様々なデータ提供方法の開発は、いくつかのデータ・アーカイブが、単純に業務の効率化のために行っている訳ではない。その背景には、国境を越えてデータの相互利用を促進しようとする政策的な後押しがある。それは個人研究者では対応できない研究資源上の問題に対応するために研究拠点 (Research Infrastructure) を整備しようとする潮流のなかで起きていると理解するべきものである (Tanenbaum and Mochmann 1994; Scheuch 1990; ESFRI 2006)。学術目的で収集された社会調査データのさらなる蓄積と公開、および二次利用による研究の発展のためには、研究者あるいはアーカイブの熱意と努力だけではなく、政策的な対応が必要であることを強調して、本稿の結びとしたい。

[謝辞] 本稿の内容は、筆者が2009年度に日本学術振興会から頂いた特定国派遣・短期の費用でドイツのGESISに滞在すると同時に、フィンランド、ノルウェー、フランス、オランダのデータ・アーカイブを訪問した際に得た見聞に大きく依拠しています。日本学術振興会並びにドイツ学術交流会(DAAD)、聞き取りに快く応じてくださった各アーカイブの職員の皆さん、そしてGESISの受入責任者であったDr. Ekkehard Mochmannと実際の面談の調整にご尽力頂いたDr. Meinhard Moschnerに深く感謝します。

[参考文献]

- Adams, Margaret O'Neill. 2007. "Analyzing Archives and Finding Facts: Use and Users of Digital Data Records." *Archival Science* 7 (1):21-36.
- Blank, Grant. 1993. "Codebooks in the 1990s: or, Aren't You Embarrassed to Be Running a Multimedia-Capable, Graphical Environment Like Windows and Still Be Limited to 40-Byte Variable Labels?" *Social Science Computer Review* 11 (1):63-83.
- Blank, Grant, and Karsten Boye Rasmussen. 2004. "The Data Documentation Initiative: The Value and Significance of a Worldwide Standard." *Social Science Computer Review* 22 (3):307-18.
- Carol A. Hert, Sheila Denn, and Stephanie W. Haas. 2004. "The Role of Metadata in the Statistical Knowledge Network: An Emerging Research Agenda." *Social Science Computer Review* 22 (1):92-9.
- Doorn, Peter, and Heiko Tjalsma. 2007. "Introduction: Archiving Research Data." *Archival Science* 7 (1):1-20.
- ESFRI. 2006. "European Roadmap for Research Infrastructures – Report 2006." Luxembourg: European Strategy Forum on Research Infrastructures (Retrieved on May 7, 2011 at http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=home).
- Gregory, Arofan, Pascal Heus, and Jostein Ryssevik. 2009. "Metadata," (Retrieved on May 7, 2011 at http://www.ratswd.de/publ/workingpapers_09.php).
- Micah Altman, Leonid Andreev, Mark Diggory, Gary King, Akio Sone, Sidney Verba, and Daniel L. Kiskis. 2001. "A Digital Library for the Dissemination and Replication of Quantitative Social Science Research: The Virtual Data Center." *Social Science Computer Review* 19 (4):458-70.
- Miller, Kenneth, and Mary Vardigan. 2005. "How Initiative Benefits the Research Community - the Data Documentation Initiative." Delivered at the First International Conference on e-Social Science. Manchester, UK. (Retrieved on May 7 at <http://www.ddalliance.org/node/129>).
- Mochmann, Ekkehard. 1974-75. "Information Access at the Data Item Level: Approaches to Indicator Retrieval from Survey Archive Data Bases." *SIGSOC Bulletin* 6 (2-3):47-55.
- Rasmussen, Karsten Boye, and Grant Blank. 2007. "The Data Documentation Initiative: A Preservation Standard for Research." *Archival Science* 7 (1):55-71.

- Ryssevik, Jostein, and Simon Musgrave. 2001. "The Social Science Dream Machine: Resource Discovery, Analysis, and Delivery on the Web." *Social Science Computer Review* 19 (2):163-74.
- Scheuch, Erwin K, and Philip J Stone. 1964. "The General Inquirer Approach to an International Retrieval System for Survey Archives." *The American Behavioral Scientist* 7 (10):23-8.
- Scheuch, Erwin K. 1990. "From a Data Archive to an Infrastructure for the Social Sciences." *International Social Science Journal* 42 (1):93-111.
- . 2003. "History and Visions in the Development of Data Services for the Social Sciences." *International Social Science Journal* 55 (177):385-99.
- Tanenbaum, Eric, and Ekkehard Mochmann. 1994. "Integrating the European Database: Infrastructure Services and the Need for Integration." *International Social Science Journal* 46 (4):499-513.
- Vardigan, Mary, Pascal Heus, and Wendy Thomas. 2008. "Data Documentation Initiative: Toward a Standard for the Social Sciences." *The International Journal of Digital Curation* 3 (1):107-13.
- 金子康樹, 2002, 「統計数値メタデータのデータ要素と表現の標準化」『レコード・マネジメント』45: 55-68.
- 杉本重雄, 2009, 「Dublin Core の現在」『デジタル図書館』36: 32-45.
- 谷口祥一, 2010, 『メタデータの「現在」: 情報組織化の新たな展開』勉誠出版.
- 前田幸男, 2008, 「海外におけるリモート集計システムをめぐる動向」『エストレーラ』2008年10月: 18-27.

第3章 DDIの技術的側面，利用イメージ，導入ステップ

佐藤慶一 米倉佑貴

本稿の情報源は，DDI アライアンスのホームページ (<http://www.ddialliance.org/>) より取得した情報を翻訳したものが中心で，必要に応じて加筆したり，情報整理を加えたものである．個別の URL は脚注として本文内に示してあるので参照されたい．

1. DDIの技術的側面

1.1. DDIの概要¹

DDI は，社会科学，行動科学および経済学の調査データのメタ情報の記述に関する国際基準を作る取組みである．DDI は，XML 形式で表現され，研究におけるデータライフサイクル全体を支援することが目的とされている．具体的には，調査の概念，データ収集，データの処理や加工，配布，分析，再利用などにわたる一連のメタ情報を扱う．1995 年頃より初期の取組みが開始され，2000 年に DDI Version 1 が公開され，その後も，2003 年に DDI 2，2008 年に DDI 3 の公開と活動が継続している．DDI は，図 1 に占めすように，アメリカ，ヨーロッパをはじめ，世界中で利用がなされている．



図 1 DDI を利用している組織の分布 (2009 年 8 月時点)

¹ <http://www.ddialliance.org/what>

DDI を推進している DDI Alliance は任意の組織団体で、2009 年 8 月時点で、アメリカやヨーロッパの 32 の国や大学およびデータアーカイブが参加している²。

DDI Alliance は、会員機関からの年会費によって運営されており、財政的に自立している。2010 年度の年会費は、2500 ドルである。規約は ICPSR によって作成されており、Alliance の目的として、DDI 基準の継続的な開発や修正等の監視、DDI の関係機関での利用を促進すること、潜在的な利用者の開拓、他の分野でのメタデータの記述方法との互換性の確保などが挙げられている。ICPSR、Roper センターが財政および法的な責任を取る組織として位置付けられている。技術的な開発やその利用促進は、Council of European Social Science Data Archives、IASSIST (International Association for Social Science Information Service & Technology) と、IFDO (International Federation of Data Organizations in the Social Sciences) が担うとされている。

実際の運営や技術開発等は、Alliance に組織されている運営委員会とエキスパート委員会が担当している。運営委員会は、8 名程度の委員により構成され、予算や全体の運営方針などを担当している。エキスパート委員会は、40 名ほどの情報技術や社会調査に詳しい XML の専門家や研究者およびライブラリアンから構成されており、詳細な技術的作業を担当している。

1.2. DDI の技術的側面³

(1) DDI とは何か

DDI は社会調査のメタデータを XML 形式で記述する仕様である。特に、社会科学のデータファイルを文書化する形式の一つで、MS ワードやテキストファイルよりも有用である。タグ構造によりコンピュータでの情報管理や加工に適している。

現在のところ DDI はまだ ISO (国際標準化機構、International Organization for Standardization) の規格になっていない。ISO の規格になることは、DDI の取り組みの重要な目的とされている。

(2) XML

XML は、Extensible Markup Language の略で、データ形式のひとつである。XML は、タグ (tag) と呼ばれる情報をデータの中に埋め込む。たとえば社会調査であれば、データの中に <調査名><調査時点>・・・などのタグを埋め込む。

タグを埋め込んだデータを記述する言語は、マークアップ言語 (markup language) と呼ばれる。XML は、マークアップ言語のひとつである。

マークアップ言語にはさまざまな種類があり、現在最も普及している仕様は HTML (HyperText Markup Language) である。HTML は、Web ページの記述を目的とした言語で

² 加盟組織の詳細は、<http://www.ddialliance.org/alliance/structure#members> を参照

³ <http://www.ddialliance.org/resources/faq.html>

あり扱うタグが、W3C という非営利の標準化団体により定められているが、XML では、データを作成する人がタグを自由に決められる点に違いがある。図 2 に XML 文書の例として、DDI 文書の一部を示した。HTML には<r:Citation><s:Purpose>などのタグはないが、XML ではこのようなタグを取り決めて使うことができる。DDI では大量の要素（開始タグと終了タグとタグ間のコンテンツの総称）が用意されており、「r:」（この場合 reusable モジュール）や「s:」（この場合 study unit モジュール）などの名前空間(namespace)が用いられる。これにより、同名の要素が混在しても識別ができるようになり、また、利用者はただちにどのモジュールで定められている要素なのか把握できる。なお、DDI では、図 2 に見るように、要素によって、ID やバージョンを与えて管理することができる。

XML は、簡潔で厳密な文法をもつことから、コンピュータ上で扱いやすいという特徴を持つ。図 2 の文書例で検索機能を考えれば、<ModeOfCollection>のタグを利用して、郵送調査で行われた調査を検索して、検索結果にこの調査の情報を表示するという形となる。

タグの構成や決まりを標準化しておき、多くの調査関係者が利用すれば、多様で膨大な社会調査の情報を効率的に管理することが可能となる。社会調査の情報管理を行うための XML の独自仕様が DDI となる。XML を利用して、社会調査の情報管理を行おうとするのが、DDI ということになる。タグの構成や決まりを標準化しておき、多くの調査関係者が利用すれば、多様で膨大な社会調査の情報を効率的に管理することが可能となろう。

```

<ddi:DDIInstance id="0001 " version="1.0.0"
  versionDate="2011-03-11T13:56:11" agency="SSJDA">
  <r:Citation>
    <r:Title xml:lang="ja-JP">東大社研・高卒パネル調査・・・</r:Title>
    <r:Contributor xml:lang="ja-JP">東京大学社会科学研究所・・・</r:Contributor>
  </r:Citation>
  <s:Purpose id="0001">
    <r:Content xml:lang="ja-JP">本調査は、・・・ </r:Content>
  </s:Purpose>
  <d: DataCollection>
    <d: CollectionEvent>
      <d: ModeofCollection>
        <r: Content>郵送調査</r:Content>
      </d: ModeofCollection>
    </d: CollectionEvent>
  </d: DataCollection>
</ddi:DDIInstance>

```

図 2 DDI 文書の例

(3) DTD, XML スキーマ

タグ付けやマークアップは、適切なルールがなければ有効に活用することができない。例えばある調査者が<調査名>というタグを、調査名を表すのに使う一方で、他の調査者が<title>や<studyTitle>というタグを使っていれば、XML の有用性は下がる。こうした状況を避けるために、DTD(Data Type Definition)や XML スキーマを使う。これらの文書で、どのタグが使えて、タグが文書の中でどのような順番で出現し、そのタグが必須なものかオプションなのか、繰り返し使えるのか否かを定義する。DTD とスキーマの違いは、DTD が XML とは違う特別な構文で記述されるのに対して、スキーマは XML で記述されることである。DDI Alliance では、ユーザーが調査情報を XML で記述するルールを作成し、それを DTD, XML スキーマの形で無償公開している。この DTD, XML スキーマが DDI の仕様の本体となる。定義ファイルや関連ドキュメント一式は、DDI Alliance のウェブサイトよりダウンロード可能である。DDI ではバージョン 3 より前は DTD を用いて仕様が定義されていたが、バージョン 3.0 以降では XML スキーマが採用されている。

(4) DDI3 とデータライフサイクルモデル⁴

DDI は当初、コードブックを記述するモデルを採用しており、データアーカイブでの活用を想定していたが、最新の DDI3 ではこれをさらに拡張し、データの保存にとどまらず研究データのライフサイクル全体にわたった記述を支援することを目的とした「データライフサイクルモデル」が採用されている。

データライフサイクルとは図 3 に示すとおり、調査の概念、データ収集、データの処理や加工、配布、分析、再利用などにわたる研究の一連の流れを指す。DDI はデータアーカイブにおける調査データの管理だけではなく、調査の設計や調査の実施にも活用されることを目指すようになっている。

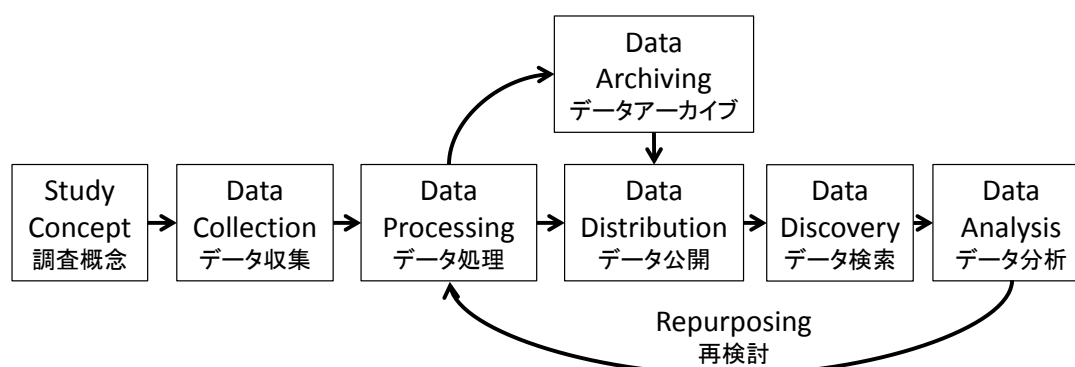


図 3 データライフサイクルモデル

⁴ (Data Documentation Initiative 2009) より

(5) モジュールとスキーム

DDI3 はデータライフサイクルの個々の段階に対応した機能のまとまりであるモジュールで構成されている。各モジュールで記述する内容は先に述べた XML スキーマで規定されている。これらのモジュールはその機能によって以下のように分類されている。

- **Packaging Modules** : 複数のモジュールを束ねるもので、Instance, Group, StudyUnit というモジュールがある。Instance は、すべての DDI 文書を含むルート要素を提供する。Group は、継続調査やパネル調査などを統合的に扱う際に利用する。StudyUnit は、一つの研究プロジェクトに関する情報を扱う。
- **Scheme-Based Modules** : 多くの研究単位で共有される可能性が高い情報のまとまりである「スキーム」を含むもので、Archive, ConceptualComponent, DataCollection, LogicalProduct, PhysicalDataProduct というモジュールがある。Archive は、データのライフサイクルイベントの情報やデータの作成に関わった人や組織の情報などを扱う。ConceptualComponent は、調査のコンセプトや研究対象、母集団などの情報を扱う。DataCollection は、データ収集に関する情報を扱う。LogicalProduct は、変数／カテゴリー／コード／集計値などに関する情報を扱う。PhysicalDataProduct は、データファイルの物理構造（レコードレイアウトなど）について扱う。
- **Non-Scheme Based Modules** : 「スキーム」を持たないもので、comparative, ddiprofile などのモジュールがある。comparative は、研究単位のグループ内での情報（母集団や変数など）の比較を扱う。ddiprofile では、扱う要素の指定やその属性について変更を加えることができる。
- **Shared Content** : 他のモジュールでよく使う要素を提供するためのもので、reusable, dcelements などのモジュールがある。reusable は、異なるモジュール間で再利用できる要素について記述する。dcelements では、書誌情報の国際標準である Dublin Core の要素を参照したり記述できる。

表 1 に、DDI3 におけるモジュールとスキームの全体およびその役割の概要を示すので、詳細はこちらを参照されたい。

表 1 DDI3 におけるモジュールとスキームの全体およびその役割の概要

分類	モジュール	スキーム	モジュール/スキームの役割
Packaging Module	Instance	-	全てのモジュールを含むルート要素
	Group		複数のStudyUnitをグループ化するために使用する
	StudyUnit		ひとつの研究プロジェクトに特化したメタ情報
Scheme Based Module	Archive	-	アーカイブ固有の情報、データのライフサイクルイベントの情報、データの作成に関わった組織や個人の情報に関するメタデータを扱うモジュール
		OrganizationScheme	データのライフサイクルを通して研究に関わったすべての組織または個人に関する情報を保持する
	ConceptualComponent	-	調査で使用する概念や研究対象・母集団などの情報を扱うモジュール
		ConceptScheme	使用された概念とその定義、概念の分類方法やそれらのスキームへの統合の仕方に関する情報
		UniverseScheme	研究の対象、母集団の構成、地理的 情報
		GeographicStructureScheme	研究対象の地理的構造
		GeographicLocationScheme	所在地情報
	DataCollection	-	サンプリングや質問紙、インタビューガイド等のデータ収集の過程に関する情報を扱うモジュール
		QuestionScheme	データ収集体具の中で使用されている一連の質問項目を記録する
		ControlConstructScheme	データ収集の道具における論理的流れを作成するスキーム
		InterviewerInstructionScheme	インタビュアーへの指示を ResourcePackageとして公開できる形式で記録するスキームである
	LogicaldataProduct	-	変数やコーディング、データの論理的な構造などを扱うモジュール
		CategoryScheme	使用される全てのカテゴリーから概念を表す特定のカテゴリーまでをカバー
		CodeScheme	変数や質問項目で使用されているカテゴリーにコードを当てはめ、それを階層的に統合することができる
		VariableScheme	変数のリストを保持する
		NCubeScheme	多元表のリストを保持する
	PhysicaldataProduct	-	データファイルで使用されている物理的なレイアウトを保持するモジュール
PhysicalStructureScheme		論理レコードの物理的側面が保持されるモジュール	
RecordLayoutScheme		論理レコードとレコードのレイアウトを関連付けたリスト	
Non-Scheme Based Module	Comparative	-	研究単位のグループやサブグループ内での比較や、外部基準との比較、resource package内での複数のスキームの比較に関する情報を扱う
	DDIprofile		DDI文書の中で使用するDDI要素やその属性を記述することができる
	PhysicalInstance		データセットの所在やデータセットの物理インスタンスに関する情報を提供する
Shared Content	Reusable	-	異なるモジュール間で再利用できる要素を記述しているモジュール
	DCElements		ダブリンコアの要素を参照またはメタデータのセットとして記述できる
	xml		XMLのクラスを使用するために含まれている
	ddi-xhtml11		DDI文書のテキスト記述を整形するためにXHTMLを使用している

2. DDI の利用イメージ

2.1. DDI の利点⁵

- ・相互運用性(interoperability)

DDI の仕様に沿ってマークアップしたコードブックは均一に交換、移動でき、これらの均質な文書を連携させるアプリケーションを作成できる。

- ・豊富な内容

DDI は社会科学のデータセットを可能な限り完全かつ全面的に記述するための包括的な要素を使えるようにデザインされている。それによって潜在的なデータ分析者にデータに関する幅広い情報を提供することができる。

- ・ひとつの文書に複数の目的を

DDI のコードブックは様々な異なるタイプの出力（例えば、伝統的な社会科学のコードブック、文献リスト、SAS/SPSS/STAT のデータ定義コード）を作成するために必要なすべての情報を保持する。

- ・オンラインでのサブグループ化や分析

DDI のマークアップは変数レベルまで拡張されており、変数群を一様な構造で管理できるとともに変数情報も提供できる。DDI 文書は簡単にオンライン分析システムに取り込むことができ、より多くのユーザーがより使用しやすくなるような準備を整えることができる。

- ・検索の正確性

DDI 準拠のコードブックに含まれる要素は特異的にタグ付けされているため、文書間や研究間で条件を特定して検索することができる。例えば、DDI コードブックのライブラリから 1960 年代に特定の州や国で行われた抗議活動を扱ったデータセットを検索することができる。

2.2. 利用イメージ

図 4 に 2010 年 10 月 24 日から 29 日にドイツで開催された Workshop 「The Data Documentation Initiative (DDI) XML Standard: Using DDI 3 to Support Production, Management, Dissemination, and Preservation Systems for Data in the Social Sciences and Economics」での配布資料⁶「6 : DDI 3 Use Cases」から主要なものをピックアップして、翻訳や調整を加えた一覧を示す。挙げられている Use Cases は実際の利用事例でなく、DDI を作成している側からの、DDI の利用イメージを示したものである。

⁵ Benefits of DDI (<http://ddi.icpsr.umich.edu/ddi-at-work/benefits>) より

⁶ 資料は、いくつかの DDI Workshop, IASSIST conference, GESIS の講習を通じて開発・改良がおこなわれたもので、主な作成者として、Wendy Thomas (Minnesota Population Center), Arofan Gregory (Open Data Foundation), Joachim Wackerow (GESIS), Pascal Heus (Open Data Foundation) があげられている。

左上から順に見ていくと、まず、調査設計段階では、Conceptual モジュールの Concepts や Universes でカバーされる概念設計部分を詰めていき、その後 datacollection モジュールの Questions や Flow Logic でカバーされる質問項目や設問順番を詰めていく。

質問生成やデータ収集・作成段階では、Concepts や Universes, Questions や Flow Logic の情報を用いて、質問紙やオンライン調査、CAI 調査を生成することが可能で、LogicalProduct モジュールの Variables や Categories/Codes などカバーされる変数情報を与えることで、データ作成に関する情報も整理できる。

コーディングや集計段階では、新しい変数やコードの情報を与えたり、集計情報を NCubes という要素へ与えることができる。Ncubes を用いた SDMX (Statistical Data and Metadata Exchange) という集計データのメタデータとの連携も検討されている。

整備された十分なメタデータを処理する情報システムが用意できれば、ウェブサイトやコードブックなどの関連資料や多様なアプリケーションでの利用のためのコマンドなどの自動生成が可能であり、データ公開における大半の作業が自動化されることが見込まれる。

データアーカイブ側で見れば、寄託されたメタデータが不十分であっても加えていくことができるし、Archival event という要素へ作業記録を残すこともできる。

整備された十分なメタデータから変数情報や調査情報を抽出して検索できるシステムが用意できれば、質問文や調査概念の検索など、従来よりも充実した検索機能をもたらすことが見込まれ、データ利用時だけでなく、新たな調査の設計時に既存調査を参考にする際にも有用となろう。DDI を用いていないが、Swedish National Data Services では、Questionbank⁷ という登録データの質問文の検索機能がある。DDI で整備されたメタデータがあれば、この様な検索機能を実現するのに新たにデータベースを作成する必要がない。

さらに、調査研究プロジェクトでの、企画段階から研究助成や関連成果、データの公開に至る多くの情報をまとめて管理することも可能である。従来、別に管理される研究体制や研究成果の情報も統合することで、より一元的な情報管理とその活用が可能となる。

継続調査やパネル調査のような複数にわたる調査についても、一つのグループとして扱ったり、調査による変数の変更などの情報も Comparative 要素で記述が可能である。別途作業が多く発生する累積データや教育用パッケージの作成も、DDI 形式でのメタデータ記述でサポートが可能とされている。

⁷ <http://www.snd.gu.se/en/qdb/>

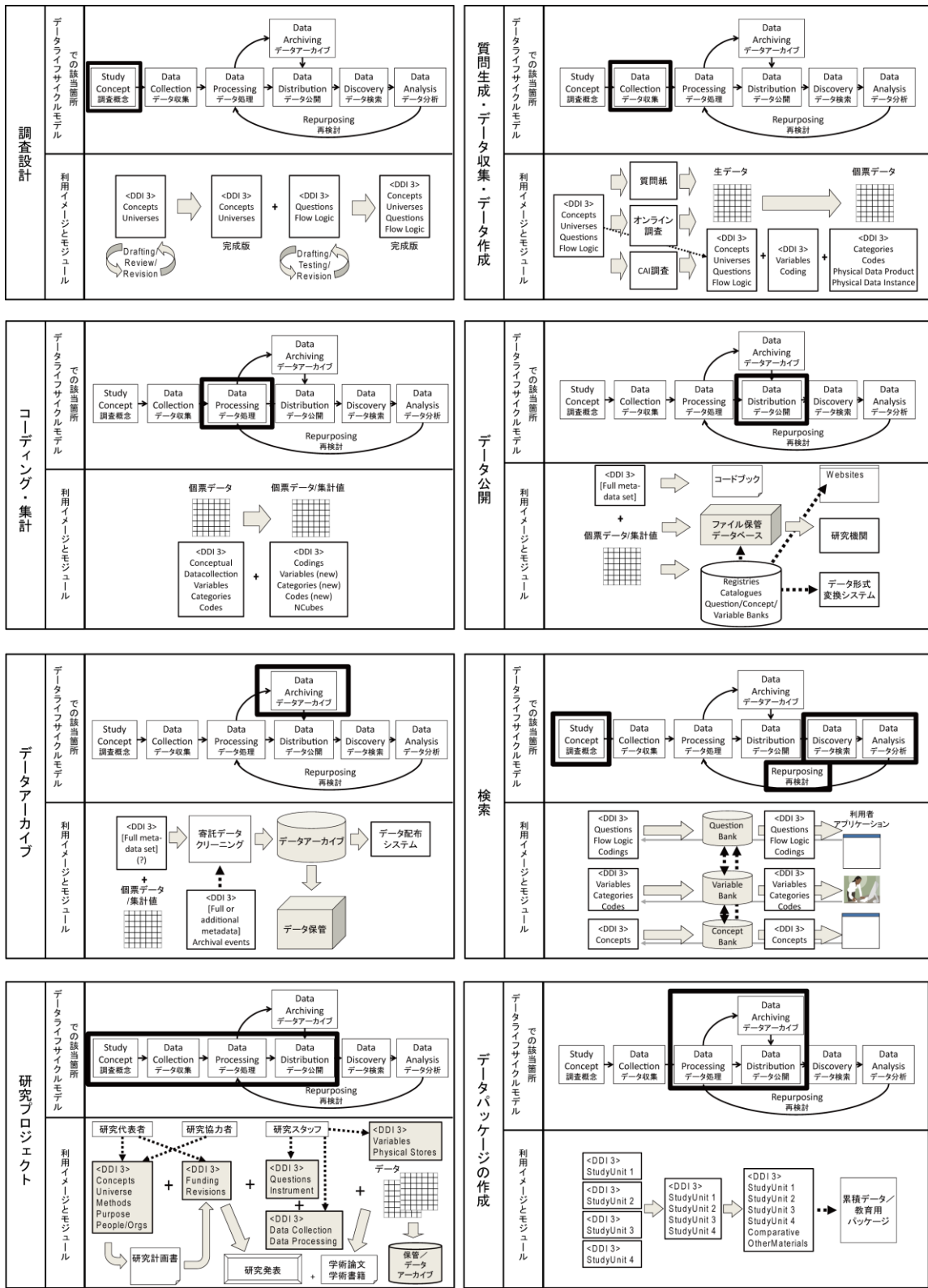


図 4 DDI の利用イメージ

3. DDI の導入ステップ⁸

3.1. 参照情報と留意点

質問がある場合や、特定の状況に対するアドバイス 欲しい場合は、DDI メーリングリストも用意されており、気軽にコンタクトを取ることが勧められている。他の組織で DDI がどのように活用されているかに興味がある場合は、DDI at Work⁹が参照できる。

DDI Best Practices¹⁰を読むと DDI 3 に慣れることができる。Best Practices の文書は DDI3 を最大限に活用する手順を定める最初の試みで、最終的なものではない。Best Practices の著者、編集者はこれらの文書の使いやすさに対するフィードバックや新しい Best Practices に対する要望を歓迎するとされており、ddi-bp-editors@icpsr.umich.edu までコメントや提案を送ることが推奨されている。

最新バージョンの DDI3 は、基本的にデータのライフサイクルの全ての段階においての情報保存をサポートできるようにデザインされている。手持ちのデータを DDI 形式で入力する前にいくつかの注意点は以下の通りである。

- DDI3 に従う XML ファイルの個々の主要なモジュール(Major Module)は他のモジュールと独立に作成することができる。一方で、Reusable schema 中の要素(タグ)は全ての主要なモジュール(Major Module)で再利用可能であり、他のモジュールとリンク可能である(他のモジュールでも呼び出せる)。
- 全ての要素に名前空間の接頭辞をつけることが推奨されている(名前空間の接頭辞の例:「l: variable」は logical product という名前空間の中の variable というタグを表す)。これによって全てのモジュールをひとつの XML ファイルに統合する際に作業が簡単になる。名前空間の接頭辞のリストは、Alliance のウェブサイト¹¹で公開されている(Data Documentation Initiative 2009) に収録されている。
- いくつかのモジュールは作成する時期に暗黙の順序があるため、他のモジュールの要素へのリンクを必要とする場合がある。たとえば、次の 2 例がある。
 - Physical Data Product モジュールは Logical Product へのリンクが必要となる。これは、データファイルのレコードレイアウトに相当する Physical Data Product を作成するためにはコーディングルール、変数の順番等に相当する Logical Product が必要となるためである。
 - Physical Instance モジュールは Physical Data Product モジュールへのリンクが必要となる。これは、データファイルそのものの場所やデータの集計結果を含む Physical Instance モジュールに先行してレコードレイアウトに相当する Physical Data Product が作成されていることが想定されるためである。
- モジュールでは他のモジュールへのリンクを作成することができる。例としては、

⁸ <http://www.ddialliance.org/resources/getting-started> より

⁹ <http://www.ddialliance.org/ddi-at-work>

¹⁰ <http://www.ddialliance.org/resources/publications/working/bestpractices>

Logical Product モジュールが Data Collection モジュールで使用されている Question 要素 (タグ) のデータを参照したり, Study Unit モジュール内の Concept 要素 (タグ) のデータを参照したりすることである。このリンクは特定の順番で作成する必要はないが, 先に作ったモジュールに後のモジュールへのリンクを作成することになる。

- Variable 要素は Concept 要素へのリンクを必要とする。もし Concept が作成されていなかった場合は, リンクを作成するためにおおまかな Concept を作成することもできる。この方法は Concept の情報がない過去のデータを DDI 形式に変換する際の, 理想的ではないが現実的なアプローチである。言い換えれば, 変数を作成するときにはその概念を記録しておく必要があるということである。
- データを公開する際に, “versionable”な要素 (タグ) やモジュールは, バージョン情報を提供する必要がある。公開前にバージョン管理が必要か否かは研究環境による。変更履歴を参照するために内部でのドラフトにバージョンをつけることができる。共同研究で DDI 文書を編集する際は, より大規模にバージョン管理を行う必要がある。

3.2. 新しい研究プロジェクトを始める際の DDI の使い方

(1) 研究組織を立ち上げる

ここで必要な要素は以下の項目のみである。

- ・ DDI 3 の Study Unit スキームに従って XML 文書を作成する。
- ・ 引用文献を用意する。

しかし, 少なくとも研究の目的をこの段階で記録しておいたほうが良い。また, この段階は研究で使用する概念を定義しておくのに良い時期である。研究で使用する概念はあとから見直すことができるが, 質問項目や変数名がこの概念を参照して作られるため, できるだけ早い時期に定義しておいたほうが賢明である。

(2) データ収集

ここで必要とされるのは XML 文書の同定のみである。質問項目の情報は変数情報から参照できるようにしておくことと, 調査用具の説明のところで使用可能にしておくことが強く推奨される。

質問項目については Question スキームの中で記述し, 「回答ドメイン」を参照することができる。「回答ドメイン」には回答の選択肢や値の範囲, 文章, その他質問に対する妥当な回答についての詳細情報が含まれる。調査用具の部分では質問項目の順番や項目を飛ばす際のパターン, 質問の前や後につく文章および, その質問項目を回答する (対象者の) 集団の単位, 分析する際の (対象者の) 集団の単位, 母集団に関する情報が示される必要がある。Question スキームは維持可能 (Maintainable) な要素であり, そのため, 調査用具の中の質問項目の使用に関する詳細情報は Question スキームの中よりも, 調査用具の中に含まれている必要がある。

新しい研究プロジェクトの場合、研究に関する情報はそれが作成されたときに DDI 文書に記録されることが前提となる。そのため、研究プロジェクトで、研究目的をまず決定し、次にデータをどこから入手するか、データ収集の方法を決定し、その後概念を定義して質問項目を作成しデータを収集して、そのデータを加工したとしたら、この順番がこれらの情報を DDI 文書に記載する順番となる。Study Unit モジュールと Data Collection モジュールをそれぞれ独立した XML ファイルにしてもいいし、1つの XML ファイルにまとめても構わない (Data Collection モジュールは Study Unit モジュールの要素である)。

(3) データの論理構造(Logical Product)を決める

1 度のデータ収集でいくつかの論理的データが生成される。しかし、一般的には最初に収集されたデータの成文化が行われる。これらには対象者の入力や回答の記録、計算の結果が反映されたり、データ収集の際の状況 (地理的な情報や、データ収集の際の様子) が含まれたりする。それとは無関係に論理的データの構造には使用された変数の構成概念そのものや構成概念を多次元表に集約したものが必要となる。

DDI3.0 の枠組みで変数を作成する際には、変数作成の方法が通常とは少し異なる。変数を作るために以下の手順を踏む必要がある。

- ・ステップ 1 選択肢を定義する。
- ・ステップ 2 選択肢のコードの構造にその定義を集約する。
- ・ステップ 3 この構造を変数作成の際に使用する。

選択肢の定義と選択肢のコードの構造は ID を付与しておけば 1 度作成するだけでよい。必要なときに参照することでこれらを再利用することができる。

ひとつの方法としては、質問項目の選択肢の部分を通して、全ての選択肢パターンを特定し、選択肢に関する情報 (ラベルと定義) を入力してそれらをコードの構造に集約することが挙げられる。このコード構造は全体やその一部分 (水準、選択肢、値の範囲、特定の選択肢) を複数の変数で共有することができる。コード構造はひとつの完全な構造として一度だけ作られる。すべてのコード構造を作成したら、コード構造を使ったり、妥当な回答の範囲を定義することで変数を組み立てることができる。

もう一つの方法としては、それぞれの変数に対して必要なときに選択肢とコード構造を組み立てる方法である。この方法では、3つのセクション (変数、選択肢、コード構造) を同時に開き続ける必要がある。すなわち、ある変数で新しい項目が必要になったときにその項目の選択肢のリストやコード構造を追加する必要がある。

2つの方法どちらを採用するかは個人の好みであるが、重要なのはどちらの方法を採用するにしても、選択肢を用いて回答するタイプの変数はコード構造を先に作成しておく必要があるということである。選択肢リストとコード構造は欠損した項目を追加するときには修正することができる。

変数情報を作成する際に、Question モジュールや Concept モジュールが作成されていれば

それを参照する必要がある。Concepts, Questions, Categories, Category Coding Structures, Variables, NCubes 要素は維持可能な要素で、それを個別の要素として維持したり、多くの研究組織で参照したり再利用したりすることができる。個人や組織でこれらの要素のデータベースを作成・維持することで、研究プロジェクト内外での一貫性を保つことができる。

NCubes 要素は常に Variables 要素が作成された後に作成され、ひとつまたはいくつかの変数を次元とすることで組み立てられる。

(4) Physical Data Product モジュールの作成

物理データはデータファイルの実体構造を記述するものである。自由にデータ構造を採用可能である（固定長、区切り、表形式、リレーショナルデータベース）。物理データモジュールの目標は、データを構成する要素（変数、複合表のセル）が含む情報とそのデータファイルにおける位置を関連付けることである。すなわち、物理データモジュールによって物理アドレス（レコードの型、レコード内での位置）と、変数または複合表でのセル番地を結びつけることである。

(5) Physical Instance モジュールを作成する

Physical Instance モジュールではデータファイルと Physical Instance との1対1の関係を記述する。例えば、5カ国で調査を行った場合、5カ国分の調査データを1つのファイルに保存することもできるし（1つの Physical Instance に対して1つの Physical Data Product）、1つの国のデータをそれぞれ1ファイルに保存することもできる（5つの Physical Instance に対して1つの Physical Data Product）。それぞれの Physical Instance はレコードのサブセット（この場合は国）を定義することができ、そのファイルに特異的な要約統計量（度数分布、平均、標準偏差等）を記録することもできる。

(6) すべてを統合する

ここまでで、ひとつの研究組織の基礎的なモジュールが完成する。これらの基礎的なモジュール群はそのまま独立したファイルとして残すこともできるし、Instance Module を使って、ひとつの DDI インスタンスに集約することもできる。これは、各セクション（モジュール）を（参照やひとつの XML ファイルに集約することで）統合し、完全な（研究全体での）引用情報（引用・参考文献の情報）、保管・所蔵の情報、そして研究全体の対象範囲（トピック、時代、地理）の情報を付加するということである。

Study Unit モジュールの引用をインスタンスの引用の基礎として用いることもできるが、ここでの引用は基本的には公表されたものを集約したダブリンコア (Dublin Core) に従った記録である。もし、新たに対象範囲に関する情報を作成したのであれば、それを全体の対象範囲に関する情報の基礎とすることもできる。また、全体の対象範囲情報にはそれよりも下位のモジュールで記述されたものを含んでいる。

(7) グループを作成する

Group モジュールは既に作成してある Study Unit モジュールから構成される。(Study Unit モジュール間で) 共通の情報は Group モジュールに記述されるとともに, Study Unit モジュール間での比較情報が提供される。研究デザインにおいて比較可能な研究は個々の研究で使われている共通の情報を示すためにこのグループモジュールを使うことができる。事後比較のために作成したグループは Comparative モジュールを使用する必要がある。

3.3. 既存のデータから DDI 文書の作成方法

終了した既存の研究情報からも, DDI 文書を作成することができる。以下の手順のいずれからでも作成を開始することもできる。

- Study Unit モジュールの作成
- Data Collection モジュールの作成
- Logical Data Product モジュールの作成

既存のデータから DDI 文書を作成するには, ①セットアップファイルからはじめる, ②既存の構造化された電子ファイルを用いて DDI 文書を作成する, ③既存の DDI 文書を DDI3 に変換する, の3通りが考えられる。次に詳細を見ていく。

(1) セットアップファイルから始める

もし, SPSS/SAS/STATA のセットアップファイルとファイル変換ソフトを使っている場合, それだけでは不完全な Logical Data Product, Physical Data Product, および基礎的な Physical Instance モジュールに相当する部分しかできない。選択肢やコード構造の反復を楽しみ, 変数情報の記述の中で参照できるようにするためには大掛かりなクリーニング作業が必要となる。

おそらく DDI 文書から SPSS/SAS/STATA のセットアップファイルを作成する方がより現実的である。これは主にこれらの統計ソフトウェアのセットアップファイルの構造に関する情報が現在のところ限定されているためである。これは目新しい問題点ではなく, 現在の社会科学のデータアーカイブに収録されている SPSS/SAS のセットアップファイルの大部分を見れば, これらの多くは変数に関する情報をほとんど含んでいないどころか var1, var2, var3 といった方法でしか情報を含んでいない場合すらあることがわかる。

しかし, Physical Data Product と変数の識別情報を含む Physical Instance の情報を得るための手段としては, セットアップファイルは依然として有用である。この有用性は変換ソフトウェアとソフトウェアがどの程度 Logical Data Product の情報をつかみ, それを並び替えたり編集したりできるかによる。現在のところ SPSS/SAS/STATA は複合表の依存関係をサポートしていないため, 変換ツールもこのような形式での集計データの記述にはあまり有用ではないものと考えられる。

統計ソフトのセットアップファイルから Logical Data Product, Physical Data Product, および Physical Instance に関する情報を抽出する作業を手軽に行うソフトウェアは現在開発中で新たな機能も今後追加されていくものと考えられる。このような変数情報は DDI の骨子ではあるが、さらに概念リストの拡張や基本文献, 研究組織の情報のような追加の情報も含めるべきである。同じ論理・物理構造をもつデータのサブセットはそれぞれの物理データファイルに対応した Physical Instance を作成するだけで同じ文書が使い回すことができることには留意しておいた方がよい。

(2) 既存の構造化された電子ファイルを用いて DDI 文書を作成する

多くの表現形式があるため、コードブックやデータディクショナリといった他の形式のファイルを DDI に変換する標準的なツールは、今のところない。同じ構造を持った多数のファイルを扱っているプロジェクトの場合は、その構造化された情報の構文を解析して DDI の構造に変換し、それを整形して最終的な DDI を作成する方法が有用と考えられる。プログラミングやマクロ, 正規表現などを使えばこの方法が実行できる。この場合、特異的な ID の作成とそのリンクを作成すること、構造と要素の妥当性を検討する XML パーサーを使って整形できるよう個々の文書に必要な要素を適切に供給することができるプログラムを組むことが重要となる。小さな単位で作業し、それぞれを独立に整形したあとで統合することでエラーがあった場合にその箇所を特定することが容易になる。各セクション間の正確なリンクを作成できるツールは価値のある資産となるものと考えられる。

(3) 既存の DDI 文書を DDI3.1 に変換する

XML browser tool を使うと、過去の DDI 文書を DDI3.0 のタグで閲覧することができる。このツールを使うと、DDI3.0 形式に変換した文書をクリップボードに保存することができる。現在のところ、このツールは個票データの構造と過去の DDI の仕様と DDI3.0 で共通のモジュールしか扱うことができない。しかし、変数の再構造化と内部での参照関係の妥当性の確保においては、このツールは DDI2.x から DDI3.0 に変換する際の大きな助けとなる。最新の規格に対応するためにはさらに DDI3.1 に変換する必要がある。

長い目でみると、変換対象の DDI 文書を見なおしておいたほうが時間と労力を節約できるものと考えられる。見直す際のポイントとして以下がある。

- DDI の全てのモジュールを全てのユーザーが使うわけではないため、変換しなくてもいいと思われるモジュールを知っておく必要がある。
- 事前に編集が必要だったり、特別注意を払う必要があるモジュールを特定しておく。
- 特定の情報を記録するためにその組織だけで使われている要素や属性を特定しておく。

過去の DDI のバージョンでバリエーションがあることや DDI3.0 への移行に伴う大きな構造の変更があるため、さらに注意する点がある。参考まで、特に注意する点として挙げられている点を以下に示す。

- Citations
 - Location
 - Structure
 - Future home
- Notes and Text
 - Special use
- Large text sections
 - Should these be pre-parsed
- Missing information
 - ID
 - Absent content
- Repetitive variable content
 - Repeated use of category set
 - Use of complex category groups

これらの全てが DDI2.x から DDI3.x に変換する際に処理する必要がある。これらの処理を可能な限り自動化することで、時間を節約しエラーを減らすことができる。バージョン変換のいくつかの選択肢は次の通りである。

- ・古いフォーマットで中間的な文書を作成しておく。
- ・途中で変換を承認したり編集できるソフトウェアを使用する。
- ・デフォルトの選択肢を決めておく。変換後の編集の過程を決めておく。
- ・手動で変換するモジュールを特定しておく

変換した際には、作成した DDI の妥当性を XML editor でチェックすることが必要である。

3.4. DDI2.x と DDI3.x の違い¹¹

DDI バージョン 3.x では特定のユーザーの要望に答えるため、以下のようないくつかの新しい機能が追加されている。

- ・いくつかの次元で（時代、地理的条件など）共通点のある研究同士を研究グループとして記述することが簡便に行えるようになっている。それにより同じ研究グループに含まれる研究間で比較可能な要素（概念、質問文、変数）を記述することもできるようになっている。この機能は時系列で行われる研究や多国間研究の記述において特に有用であると同時に、統合されたデータの記述にも有用と考えられる。
- ・複雑な構造のデータファイルの記述をより完全に行えるようになっている。たとえば、レコード間で関連する階層的なデータなどが相当する。
- ・質問紙に関する記述に新しいセクションが追加されている。調査用具の特徴、質問文、その設問に回答する条件、質問の順番（設問を飛ばすパターン）などである。質問文は

¹¹ <http://www.ddialliance.org/resources/faq#project> より

変数と関連付けられるが、Version1 や 2 の時ように変数の子要素ではない。

- (コンマ) 区切りの形式や、行/列の形式でデータの場所を表現したスプレッドシートの形式で集計データを記述することができるようになっている。
- XML 形式で実データ (集計データや個票データ) を記述したり、移送したりできるようになっている。
- 地理的範囲の記述が大幅に拡張されている。例えば、指定が必須な最上位と最下位の水準を含む全ての水準の地理情報や、地理的な変数との関連付けなどである。
- 多言語で実施された研究も対象にするなど、翻訳関連の情報もより扱いやすくなっている。

このような機能をプロジェクトで使用したいのであれば、Version3.x を使うことがよい。一方、プロジェクトが単純な個票データや調査形式の研究であったり、研究の記述をしたいだけであれば Version1.x か 2.x を使えばよい。Version1.x または 2.x が現在および将来的なニーズに合うようであれば、それらを使うことで問題ない。

DDI3 に移行する場合、Open Data Foundation が開発した DDI 2 to 3 Converter Tool(<http://www.opendatafoundation.org/tools/dextris/>)を使うことができる。このツールは準備バージョンの DDI3 を元に開発されているので注意が必要だが、それでも移行作業をする上で有用なものである。

DDI2.1 と DDI3.0 の対応表(<http://www.ddialliance.org/specification/ddi3.0#3.0convert>)もある。

[参考文献]

Data Documentation Initiative, 2009, 「DDI Technical Specification Part1:Overview version3.1」,
<http://www.ddialliance.org/Specification/DDI-Lifecycle/3.1/>内よりダウンロード可能

第4章 社会調査メタデータ管理ソフト Easy DDI Organizer の設計

米倉佑貴 佐藤慶一

1. DDI 文書の編集環境の現状

DDI は社会科学分野における研究のメタ情報を XML (eXtensible Markup Language) 形式で記述するための規格である。DDI の利点として、仕様が統一されているためデータの交換やアプリケーションでの取り扱いが容易であること、豊富な内容を記録できること、ひとつのドキュメントで様々な用途で使用できること、オンラインでのサブセット化と解析が容易であること、正確な検索が行い易いことが挙げられている(DDI Alliance 2009)。

これまではこうした利点の多くはデータアーカイブ間のデータのやり取りやデータアーカイブでの二次データの提供などに向けたものであった。しかしこれらの利点の多くは個人の研究者や研究グループによる共同研究にもあてはまる。例えば、DDI 形式で記述された調査のメタデータから SAS, SPSS, STATA など様々な統計パッケージのセットアップスクリプト、コードブック、質問紙等様々な文書に加工することができることが挙げられる。さらに、データライフサイクルモデル(Thomas et al. 2008)[2]を採用した DDI3.0 以降は研究計画の段階からデータ分析の段階までの情報を記録することができ、調査の設計や調査終了後の研究グループ内でのデータ共有など様々な用途で活用できるようになった。

こうした利点を活かすためには、まず DDI の仕様に沿った形式でメタデータを用意する必要がある。しかしながら、我が国において一般の研究者が DDI 文書を作成する環境は十分に整っていない。その大きな要因として、DDI の仕様の複雑さが挙げられる。DDI 文書は XML で記述されるため、直接編集することも可能であるが、基本的な XML の知識と DDI の仕様を理解している必要があることである。これは一般の利用者にとっては困難であるため、DDI 文書の作成には編集用のソフトウェアを使用することが一般的である。こうしたソフトウェアは欧米を中心に開発され商用のものと無償のものが提供されている。次章で詳しく述べるが商用のソフトは、機能が充実しているものの個人の研究者が使用するにはやや高価であり、無償のものでは機能が制限されている。また両者に共通して欧米で開発されたソフトウェアであるため、日本語への対応が不十分である。

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では DDI の利用可能性検討の一環として、日本語で利用可能なメタデータ作成・編集用アプリケーションの基礎設計を行い、その実装を試みている。本稿では、基礎設計の際に整理した DDI 文書の編集環境の情報と、開発中の DDI 文書編集ソフト (Easy DDI Organizer; EDO) を紹介する。

2. 既存の DDI 文書作成・編集ソフト

DDI 編集ソフトの開発のため、まず既存のソフトウェア機能および特徴を調査した。調査対象は DDI の利用者・利用機関のコミュニティである DDI Alliance のホームページに掲載されているソフトウェアのうち、DDI 文書の作成・編集機能を有するものであった。調査対象のソフトウェアの機能・特徴を表 1 に示した。

これらのソフトウェアは大きく分けて、(1)DDI で記録できる情報の入力フォームをもつもの、(2)SPSS/SAS/STATA といった統計パッケージで作成したデータファイルから変数名やラベルといった変数情報を抽出し DDI 形式に変換できるもの、(3)上記の両方の機能を持つものの 3 種類がある。次節では(1)および(3)に分類される DDI 編集ソフトに注目して機能の詳細を見ていく。

表 1 既存の DDI 文書作成・編集ソフトウェアの概要

名称	用途	機能	価格	入手先
DDI 3.0 Editor-Lite	DDI文書の作成・編集	・DDI文書の作成	無料	http://www.ddialliance.org/sites/default/files/ddi3/editor-lite-3.zip
Colectica Express/ Colectica Designer	DDI文書の作成・編集	・DDI文書の作成 ・コードブック、調査票、セットアップスクリプト、旧バージョンのDDI文書のエクスポート ・統計パッケージのデータファイル、DDI文書のインポート	有料 (\$1000-\$2000)	http://www.colectica.com/products/designer/
Nesstar Publisher	DDI文書の作成・編集	・DDI文書の作成 ・コードブック、調査票、セットアップスクリプト、旧バージョンのDDI文書のエクスポート ・統計パッケージのデータファイル、DDI文書のインポート	無料	http://www.nesstar.com/software/publisher.html
Microdata Management Toolkit	DDI文書の作成・編集	・DDI文書の作成 ・コードブック、調査票、セットアップスクリプト、旧バージョンのDDI文書のエクスポート ・統計パッケージのデータファイル、DDI文書のインポート	無料	http://www.surveynetwork.org/home/index.php?q=tools/toolkit
StatProgs2DDI/ SPSSOMS12DDI30	変数情報作成	・統計パッケージのデータファイルからDDI3.0対応のXMLを出力(現在はSPSSのみ対応) ・SPSS出力ファイルから度数分布(physical product module)もDDI3.0形式に変換	無料	http://db.zuma-mannheim.de/DDI/StatsProgs2DDI/StatsProgs2DDI.html
XCONVERT	変数情報作成	SAS, SPSS, Stata のデータ定義ファイルをDDIに変換	無料	http://sda.berkeley.edu/ddi/tools/xconvert.html
spssDDI	変数情報作成	SPSSから変数情報を抽出してDDI形式にする。 Rのパッケージ。	無料	http://cran.r-project.org/web/packages/spssDDI/
DDI-DExT	変数情報作成	・SPSSファイルから変数情報を抽出してDDI2.0, 3.0で出力 ・データファイルをASCII, CSV, SAS, Stata, で出力	無料	http://opendatafoundation.org/?lvl=forge

2.1. 機能の比較

DDI 文書作成・編集ソフトウェアのうち DDI 3.0 Editor Lite, Colectica Express, Nesstar

Publisherの主要な機能を表2に示した。なおMicrodata Management ToolkitはNesstar Publisherと同一であるため省略している。また、Colectica 関連製品のうち DDI 文書を作成・編集可能なものは Colectica Express および Colectica Designer の2種類があるが、この2つの違いは Designer がサーバーアプリケーションである Colectica Repository と連携可能であることで、DDI 文書の編集機能には実質的な差異はない。

これらのソフトウェアのうち、データライフサイクルモデルを採用した DDI 3.0 以降を扱えるのは DDI 3.0 Editor Lite, Colectica Express の2つである。Nesstar Publisher は1世代前の DDI 1.2.2 までしか対応していないが、入力フィールドをカスタマイズすることで DDI 1.2.2 では扱えないメタ報を記録することができる。

表 2 DDI 編集ソフトウェアの機能の詳細

	Colectica Express	DDI Editor	Nesstar Publisher
DDI			
バージョン	3.1	3.0	1.2.2
入力フィールドのカスタマイズ	×	×	○
インポート			
データファイル			
SPSS	○	×	○
SAS	×	×	○
STATA	×	×	○
Statistica	×	×	○
NSDStat	×	×	○
dBase	×	×	○
DIF	×	×	○
テキスト	○	×	○
DDI文書	○	×	○
エクスポート			
DDI文書			
DDI2.0	○	×	○
DDI3.0	○	○	×
コードブック			
PDF	○	×	×
RTF	○	×	×
統計パッケージ用スクリプト/データファイル			
SPSS	○	×	○
SAS	○	×	○
STATA	○	×	×
R	○	×	×
Statistica	×	×	○
NSDStat	×	×	○
質問紙	○	×	×
その他			
インターフェース	英語	英語	英語
多言語対応	○	○	○

データファイルや既に作成した DDI 文書のインポート機能は Colectica Express および Nesstar Publisher で利用出来る。両者とも様々な統計パッケージのデータファイルに対応し

ており、統計パッケージを使用して編集した変数情報をインポートし、DDI 形式で出力することができる。一方で DDI 3.0 Editor Lite はインポート機能を持たず、またこのソフトを使用して作成した文書を読み込むことも不可能である。さらに、規定されている必須項目をすべて入力するまで文書を出力することができないため、研究の進度にあわせて DDI 文書を作成するという用途には不向きで、完了した研究のメタ情報を入力する用途に限られる。

入力したメタ情報を加工しコードブックや質問紙、統計パッケージ用のデータファイルやセットアップファイル(変数名やラベルをつけるスクリプト)を出力するエクスポート機能も Colectica Express, Nesstar Publisher で提供されている。特に Colectica Express はエクスポート機能が充実しており、入力した質問項目の情報やコーディングの情報から質問紙や CAI(Computer Assisted Interview)ソフト用の設定ファイル、コードブック等を出力でき、DDI3.0 以降のデータライフサイクルモデルに沿った機能を提供している。

2.2. 入力フィールドの比較

次に DDI3.0 以降を扱える DDIEditor Lite, Colectica Express に注目して記録できる情報を整理する。これらのソフトウェアで記録できる情報は表 3 の通りである。どちらも研究全体の目的や概要、研究資金等の情報を扱う StudyUnit モジュール、研究計画の段階で考慮する概念や母集団に関する情報を扱う ConceptualComponent モジュール、サンプリングや調査方法、質問項目等の実際の調査に関する情報を扱う DataCollection モジュール、変数のコーディングや変数の定義など収集したデータを整理する LogicalProduct モジュール、データファイルのフォーマット、レイアウトや保管場所に関する情報を扱う PhysicalProduct モジュールに対応する入力フィールドが用意されている。

一方で、DDI3.0 以降に採用された Group モジュール(複数の研究・調査のメタ情報の比較等に関する情報を扱う機能)には対応しておらず、これらのソフトウェアで扱うのは現状では単一の調査のメタ情報にとどまっている。

表 3 Colectica Express と DDI3.0 Editor-Lite の入力可能フィールドの比較

	Colectica Express	DDI3.0 Editor-Lite	対応するDDIモジュール
研究全体の情報			
要約	○	○	
目的	○	○	
研究タイトル	○	○	
研究がカバーする範囲			
トピック	○	×	StudyUnit
キーワード	○	×	
時代	○	×	
地理	○	○	
研究資金に関する情報	○	○	
研究に関わった個人・組織に関する情報	○	○	
研究で使用する概念に関する情報	○	○	ConceptualComponent
研究対象となる母集団に関する情報	○	○	
サンプリングに関する情報	×	○	
質問項目に関する情報			
質問文	○	○	DataCollection
調査員への指示	○	×	
回答形式	○	○	
質問の意図	○	×	
所要時間	○	×	
質問文の翻訳	×	○	
質問の順番	○	×	
論理データに関する情報			
選択肢に関する情報	○	○	LogicalProduct
コーディングに関する情報	○	○	
変数に関する情報	○	○	
含まれる変数	○	○	
物理データに関する情報			
文字コード	○	○	PhysicalProduct
データのフォーマット	○	○	
データの保管場所	○	○	
データの集計情報			
ケース数	○	○	
全レコード数	○	○	
要約統計量の算出	○	×	

3. Easy DDI Organizer の設計

以上に示した既存の DDI 文書作成・編集ソフトウェアの機能・特徴を参考に、DDI 文書作成ソフトウェア Easy DDI Organizer(EDO)を設計した。EDO は DDI3.0 以降に採用されたデータライフサイクルモデルの特長を生かし、研究計画からデータの保存までの一連の流れで記録しておくべき情報を入力するフィールドを用意し、調査の設計、メタ情報の保存を支援するというコンセプトで設計している。具体的な機能は以下のとおりである。

- メタ情報入力・保存機能

研究のメタ情報を入力するインターフェースを提供する。入力フィールドの概要は表 4 のとおりである。入力したデータは将来の DDI の仕様変更への対応が容易になるよう、独自形式の XML で保存しそれを DDI 形式の XML にエクスポートできるようにした。

- 入力データの検証機能
入力したデータが規定のデータ型に合致しているか、必須のフィールドの入力漏れがないか検証する
- インポート機能
DDI形式のXML文書を取り込み、表示する。対応するDDIのバージョンは1.2.2から3.1までとした。また、統計解析パッケージのデータファイルを読み込み、変数、カテゴリ、コーディングの情報を取り込む機能も搭載することとした。
- エクスポート機能
入力したメタ情報をもとに、調査票、DDI仕様に準拠したXMLファイルの出力、コードブック、統計パッケージ用のセットアップシンタックス、統計パッケージ用の集計データ作成シンタックスを出力する機能を搭載することとした。

表 4 EDO で記録可能なメタ情報

大分類	中分類	内容
イベント管理		研究開始からデータ作成までのイベントを記録する
調査の企画	調査メンバー 調査タイトル, 調査目的, 調査の概要 調査でカバーする範囲 研究資金	研究を実施する組織体制を記述する 研究対象となるトピック, 時代, 地理について記述する 研究を実施するための資金をどのように確保したか記述する
データの収集方法	母集団 サンプリング データ収集方法	母集団を明確に定義する サンプリング方法を記録する どのようにデータを収集するか記録する
設問設計	扱いたい変数のイメージ 質問項目の設計 質問の順番	収集したいデータを定義し大まかに分類する 質問文, 回答方法, コーディングを定義し互いに関連付ける 質問項目の順番を定義
変数の管理	変数情報	質問項目とデータとして扱う変数を関連付ける
データの管理	データセットの定義 データファイルの定義 データファイル情報	データファイルに格納する変数等を決める データファイルにどのようにデータを格納するか決める(文字コード, 変数の位置等) データファイルがどこに保存されているかやデータファイルの概要(ケース数, 要約統計量等)を記録する
調査のグループ化と管理	研究の関連性 変数イメージの大分類の対応 変数イメージの小分類の対応 変数の対応	グループに属する研究の関連性を記述する グループに属する研究に含まれる変数イメージの大分類の対応を記述する グループに属する研究に含まれる変数イメージの小分類の対応を記述する グループに属する研究に含まれる変数の対応を記述する

4. EDO の開発状況と機能の紹介

上述の設計に基づいて、2010年12月にEDOの開発を開始した。2011年3月現在、メタ情報入力・保存機能、保存したメタ情報をDDI3.1準拠のXML形式でエクスポートする機

能を実装したプログラムがテスト段階に入っている。2011年度には各種インポート/エクスポート機能を実装するとともに、英語インターフェースの実装も目指す。現段階でのEDOの機能と既存のメタ情報編集ソフトの機能および記録可能なメタ情報の比較は表5、表6のとおりである。

表5 EDOと既存のメタ情報編集ソフトの機能の比較

	Easy DDI Organizer	Colectica Express	DDI Editor	Nesstar Publisher
DDI				
バージョン	3.1	3.1	3.0	1.2.2
入力フィールドのカスタマイズ	×	×	×	○
インポート				
データファイル				
SPSS	△(実装予定)	○	×	○
SAS	×	×	×	○
STATA	×	×	×	○
Statistica	×	×	×	○
NSDStat	×	×	×	○
dBase	×	×	×	○
DIF	×	×	×	○
テキスト	×	○	×	○
DDI文書	△(実装予定)	○	×	○
エクスポート				
DDI文書				
DDI2.0	△(実装予定)	○	×	○
DDI3.0	△(実装予定)	○	○	×
コードブック				
PDF	△(実装予定)	○	×	×
RTF	△(実装予定)	○	×	×
統計パッケージ用スクリプト/データファイル				
SPSS	△(実装予定)	○	×	○
SAS	△(実装予定)	○	×	○
STATA	△(実装予定)	○	×	×
R	△(実装予定)	○	×	×
Statistica	×	×	×	○
NSDStat	×	×	×	○
質問紙	△(実装予定)	○	×	×
その他				
インターフェース	日本語	英語	英語	英語
多言語対応	○	○	○	○

表 6 EDO と既存のメタ情報編集ソフトの記録可能なメタ情報の比較

	Colectica Express	DDI3.0 Editor-Lite	Easy DDI Organizer	対応するDDIモジュール
研究全体の情報				
要約	○	○	○	
目的	○	○	○	
研究タイトル	○	○	○	
研究がカバーする範囲				StudyUnit
トピック	○	×	○	
キーワード	○	×	○	
時代	○	×	○	
地理	○	○	○	
研究資金に関する情報	○	○	○	
研究に関わった個人・組織に関する情報	○	○	○	
研究で使用する概念に関する情報	○	○	○	ConceptualComponent
研究対象となる母集団に関する情報	○	○	○	
サンプリングに関する情報	×	○	○	
質問項目に関する情報				DataCollection
質問文	○	○	○	
調査員への指示	○	×	×	
回答形式	○	○	○	
質問の意図	○	×	×	
所要時間	○	×	×	
質問文の翻訳	×	○	×	
質問の順番	○	×	△(実装予定)	
論理データに関する情報				LogicalProduct
選択肢に関する情報	○	○	○	
コーディングに関する情報	○	○	○	
変数に関する情報	○	○	○	
含まれる変数	○	○	○	
物理データに関する情報				PhysicalProduct
文字コード	○	○	○	
データのフォーマット	○	○	○	
データの保管場所	○	○	○	
データの集計情報				
ケース数	○	○	○	
全レコード数	○	○	×	
要約統計量の算出	○	×	×	
複数の調査のグループ化・比較				Group, Comparative
概念の比較	×	×	○	
変数の比較	×	×	○	

以下、既に EDO に実装されているメタ情報入力・保存機能、DDI3.1 準拠の XML の出力機能について簡単に紹介する。

4.1. メタ情報入力・保存機能

EDO には「イベント管理」、「調査メンバー」、「調査タイトル、調査目的、調査の概要」、「調査でカバーする範囲」研究資金、「母集団」、「サンプリング」、「データ収集方法」、「扱いたい変数のイメージ」、「質問項目の設計」、「変数情報」、「データセットの定義」、「データファイルの定義」、「研究の関連性」、「大分類で比較」、「小分類で比較」、「変数で比較」の画面がある。これらの画面は 3 章で紹介した DDI のモジュールやスキームと対応している。それぞれの画面の入力項目と DDI のタグとの対応については巻末の資料に添付した。

このうち、「イベント管理」から「データファイルの定義」までは個々の調査のメタ情報を編集するためのものである。「研究の関連性」から「変数で比較」までは DDI3.1 の Group

モジュールおよび Comparative モジュールに対応する機能であり、複数の調査を一つのグループにまとめ、そこに含まれるメタ情報の比較を行うためのものである。以下、それぞれの機能の概要を紹介する。

(1) 「イベント管理」画面 (図 1)

この画面は DDI の Archive モジュールのデータライフサイクル情報に対応している (DDI のモジュールについては 3 章参照)。ここでは、調査の開始から終了までに行ったことを記録することができる。初期状態では「研究プロジェクトの開始」、「研究助成を受けた日」、「サンプリング」、「研究で使用する変数のイメージの決定」、「調査票の作成」、「調査の実施」、「調査データファイルの作成」、「調査データの寄託」、「調査データの公開」の 9 種類が表示されている。これ以外にも自由に内容を編集して調査研究を行っている際の出来事を記録することができる。それぞれの出来事について、日付とメモを記入する欄も設けている。

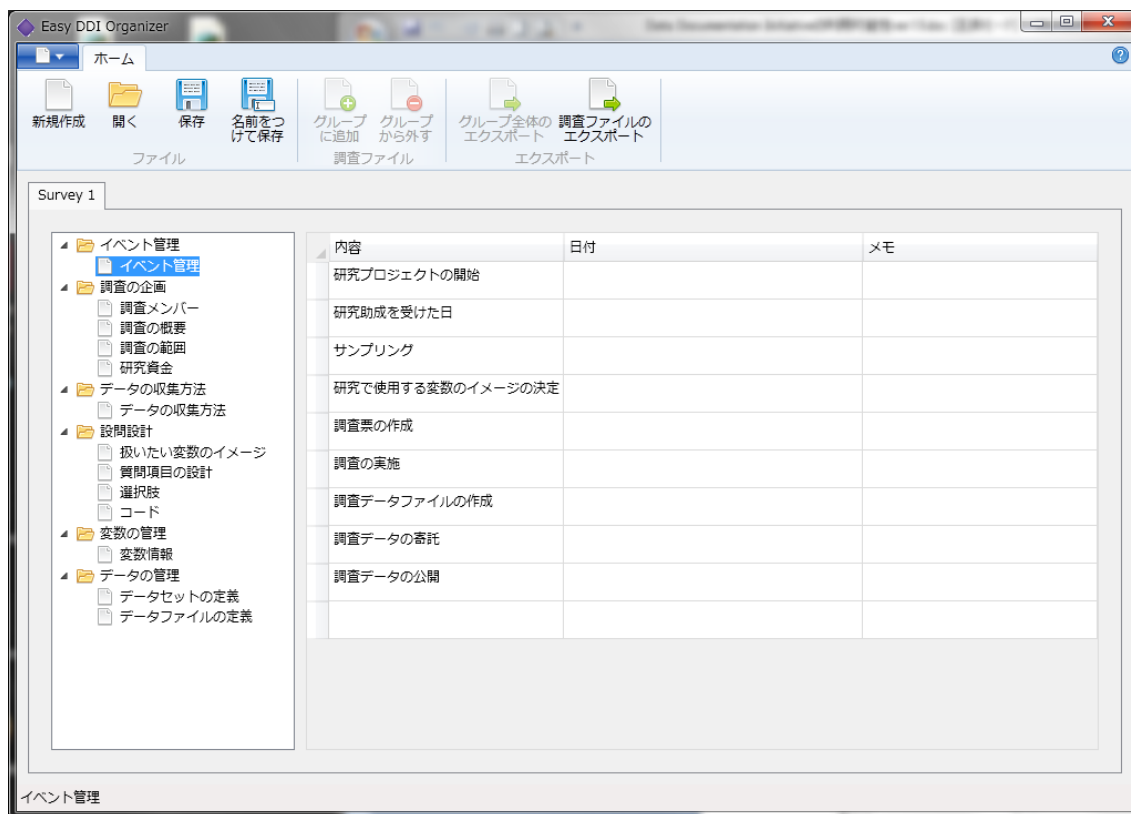


図 1 「イベント管理」画面

(2) 「調査メンバー」画面 (図 2)

この画面は DDI の Archive モジュールの OrganizationScheme に対応している。ここでは調査に関わった人、組織を記録することができる。それぞれのメンバーの名前や所属、研究組織の中での役割を入力できる。

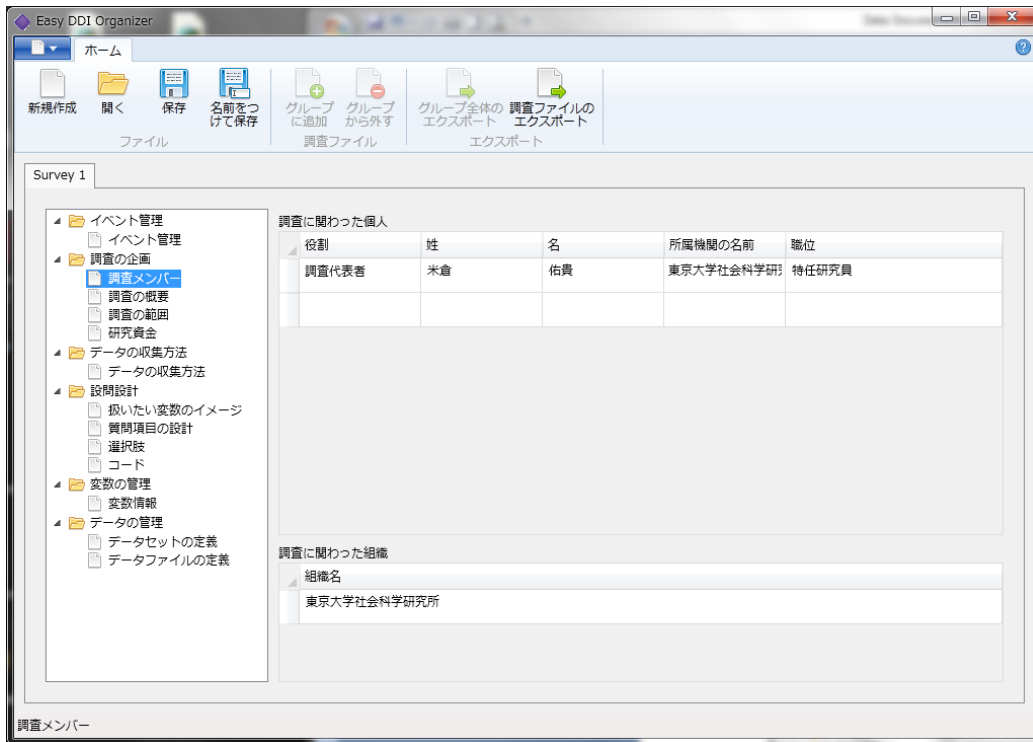


図 2 「調査メンバー」画面

(3) 「調査の概要」画面 (図 3)

ここでは調査タイトル，調査の目的，調査の概要を記録することができる。

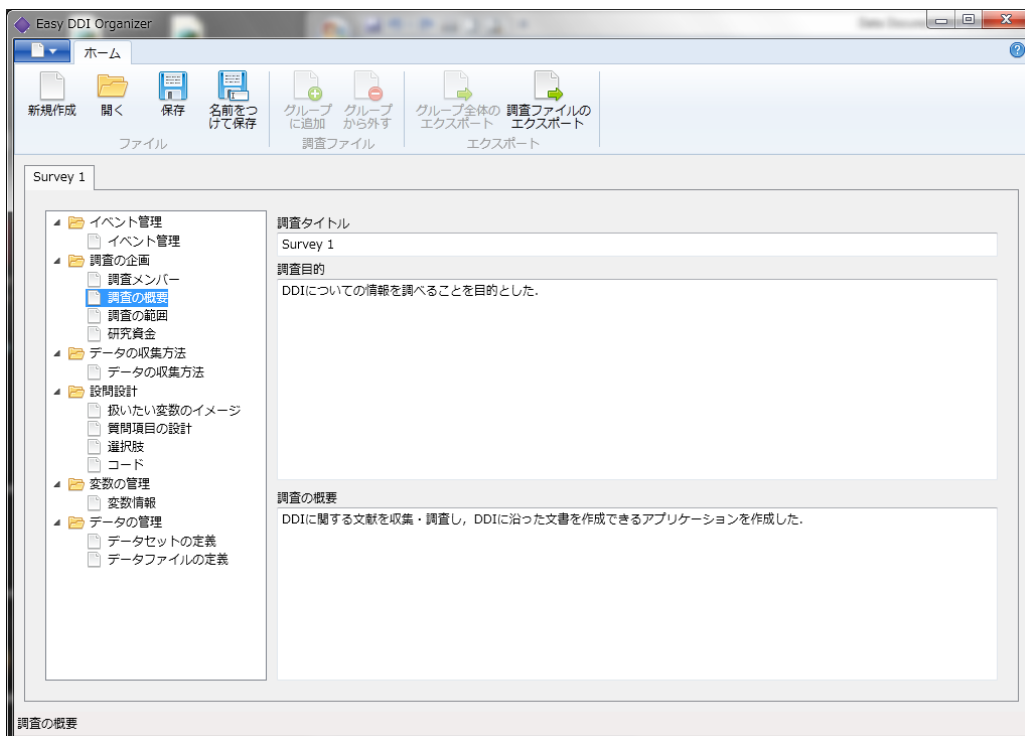


図 3 「調査の概要」画面

(4) 「調査の範囲」画面 (図4)

ここでは研究分野の大まかな分類, キーワード, 調査の対象となっている時期, 地域についての記録を残すことができる. 研究分野の大まかな分類は SSJDA で採用している研究分野の分類を使用している.

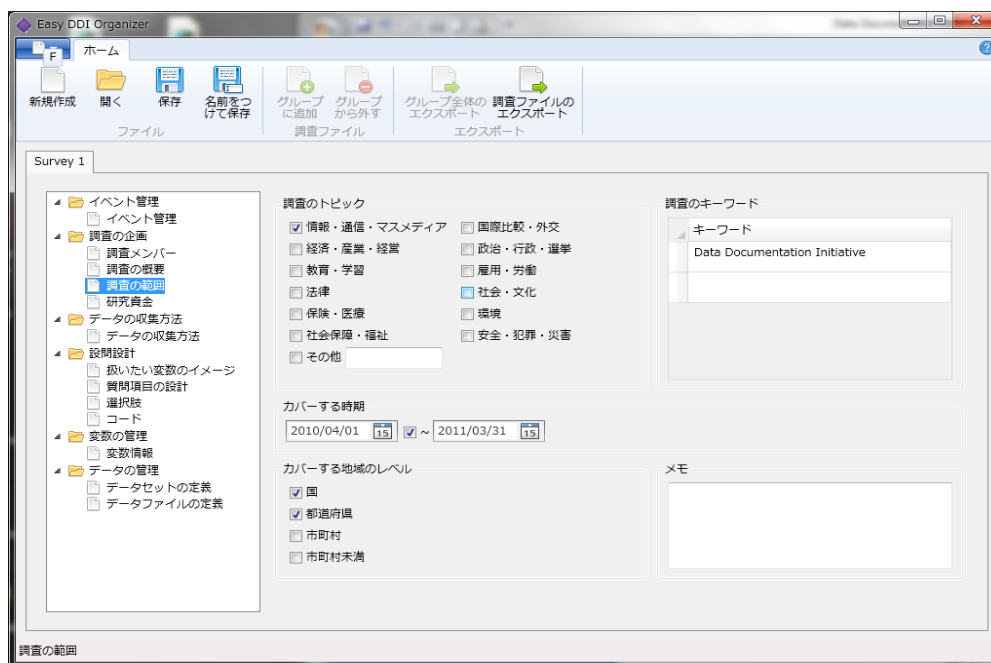


図4 「調査の範囲」画面

(5) 「研究資金」画面 (図5)

ここでは研究を行うための資金に関する情報を記録することができる. 資金の提供機関や助成期間, 助成金の名称や金額などを記録することができる.

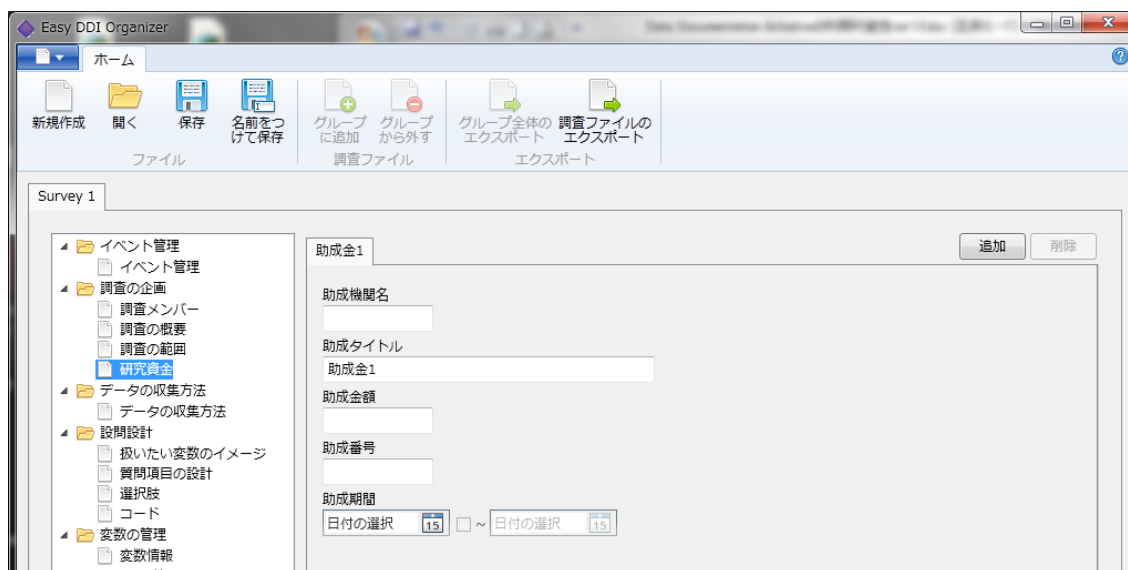


図5 「研究資金」画面

(6) 「データの収集方法」画面 (図6)

この画面は DDI の ConceptualComponent の UniverseScheme および DataCollection モジュールに対応している。ここでは母集団とそのサブグループ、サンプリング方法、調査日程、調査方法を記録することができる。

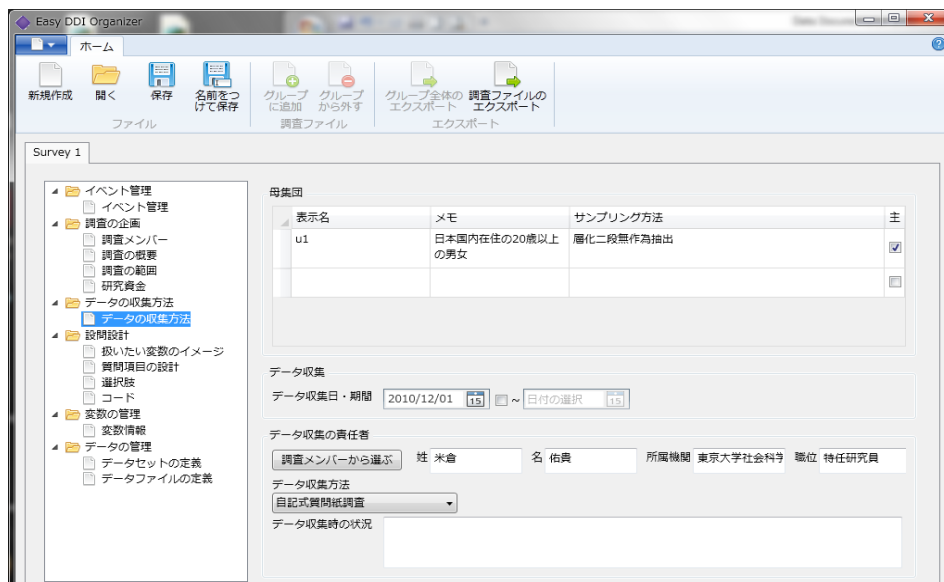


図 6 「データの収集方法」画面

(7) 「扱いたい変数のイメージ」画面 (図7)

この画面は DDI の ConceptualComponent モジュールの ConceptScheme に対応している。ここでは研究で使用する概念について記録することができる。以降の設問設計では基本的にここで設定した概念に基づいて設計できるようになっている。

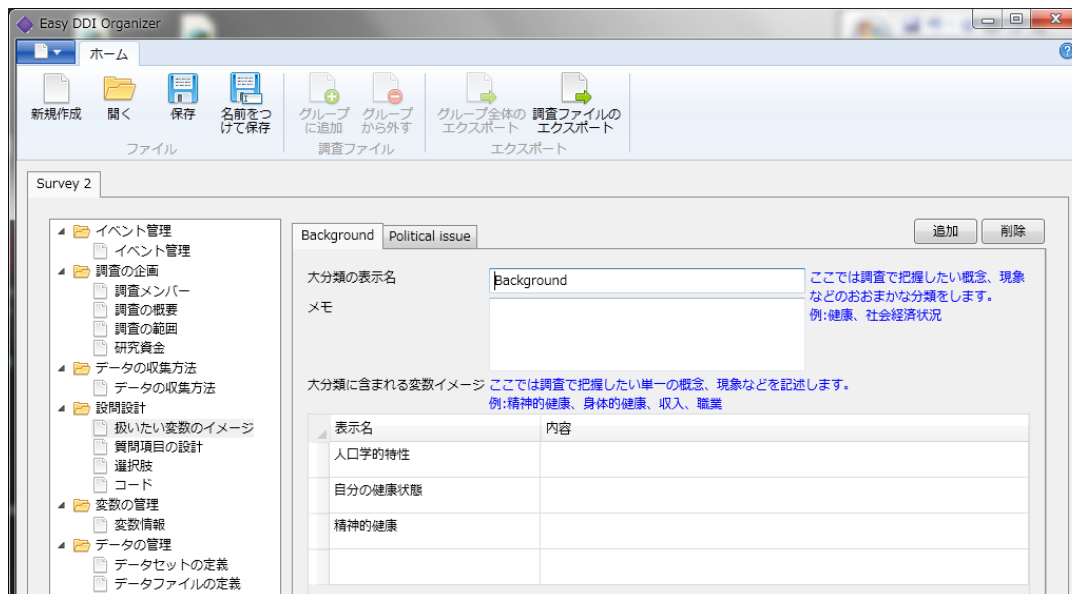


図 7 「扱いたい変数のイメージ」画面

(8) 「質問項目の設計」画面 (図 8)

この画面は DDI の DataCollection モジュールの QuestionScheme に対応している。ここでは調査に使用する質問項目の設計を行うことができる。図 7 で設定した概念ごとに質問項目を作ることができ、それぞれの質問の質問文や回答方法を記録することができる。また、回答方法については「選択肢」、「数値」、「自由記述」、「日付・時刻」の4種類が選択でき、それぞれについてさらに詳細な設定をすることができる。たとえば選択肢で回答する場合は選択肢の内容とコーディング、数値で回答する場合は数値の上限下限などを設定することができる。選択肢の内容とコーディングについては図 9 の「選択肢」画面、図 10 の「コード」画面で個別に設定することもできる。

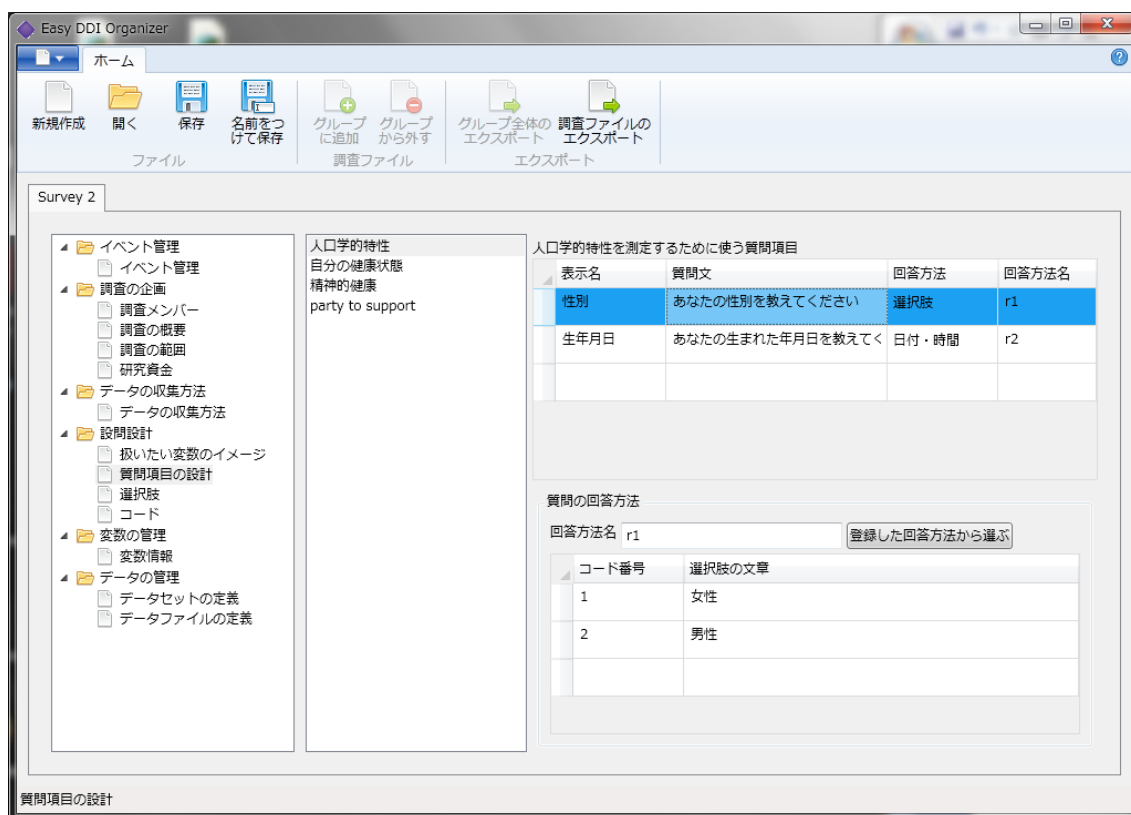


図 8 「質問項目の設計」画面

(9) 「選択肢」画面 (図 9)

この画面は LogicalProduct モジュールの CategoryScheme に対応している。ここでは調査で使用する選択肢を一元管理することができる。ここで予め選択肢を作成し図 10 でコーディングを決めておくことで複数の質問項目で使用することもできる。

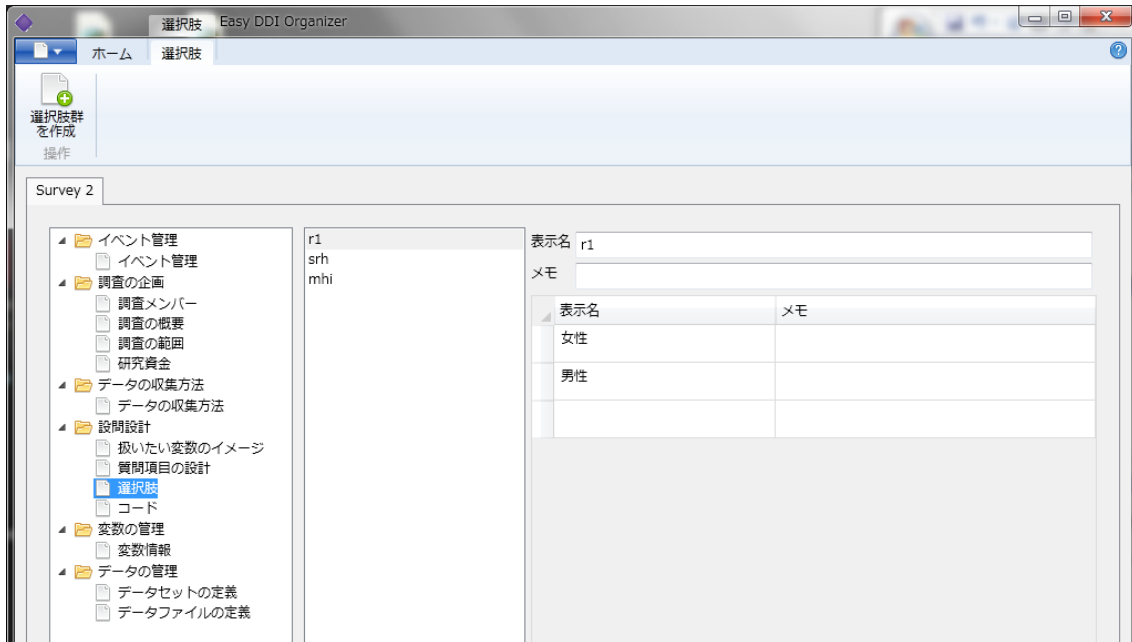


図 9 「選択肢」画面

(10) 「コード」画面 (図 10)

この画面は LogicalProduct モジュールの CodeScheme に対応している。ここでは調査で使用する選択肢とそのコーディングを管理することができる。

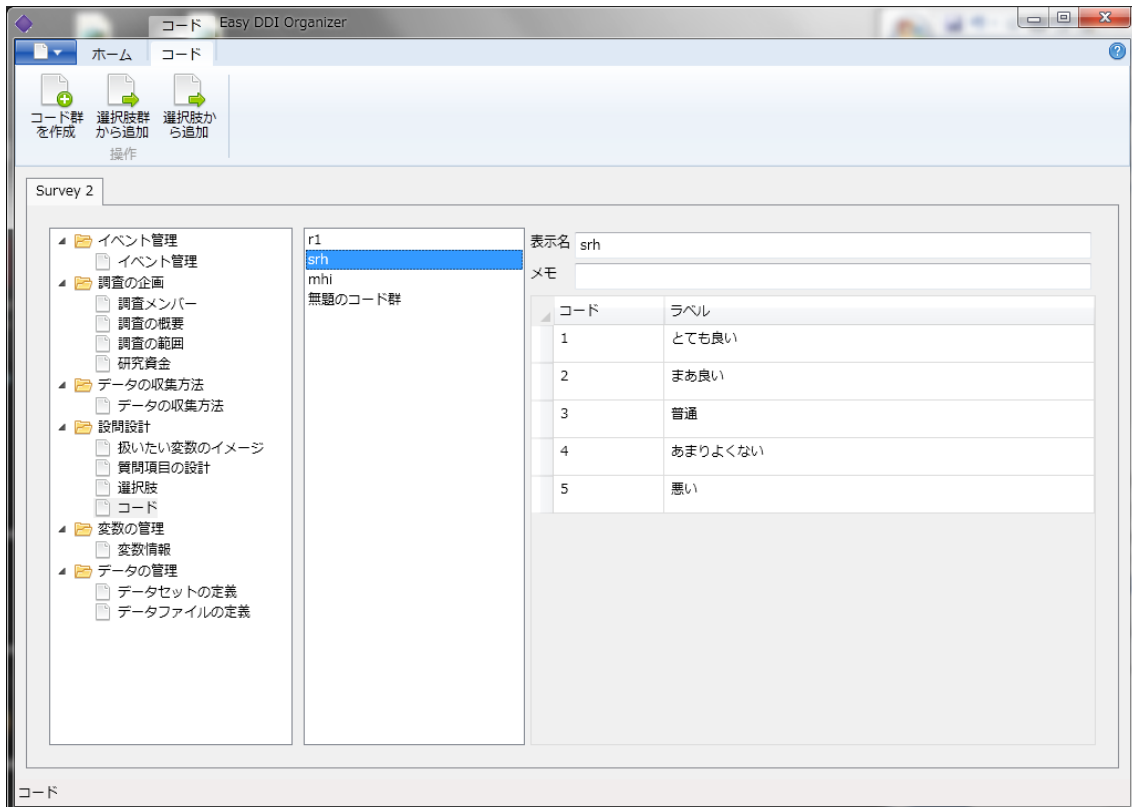


図 10 「コード」画面

(11) 「変数の管理」画面 (図 11)

この画面は LogicalProduct モジュールの VariableScheme に対応している。ここでは分析に使用する変数に関する情報を管理することができる。図 8 で作成した場合、その内容がここにも反映される。質問項目と対応していない変数や質問項目で得られた回答を加工して変数を作成する場合はここで作成することができる。

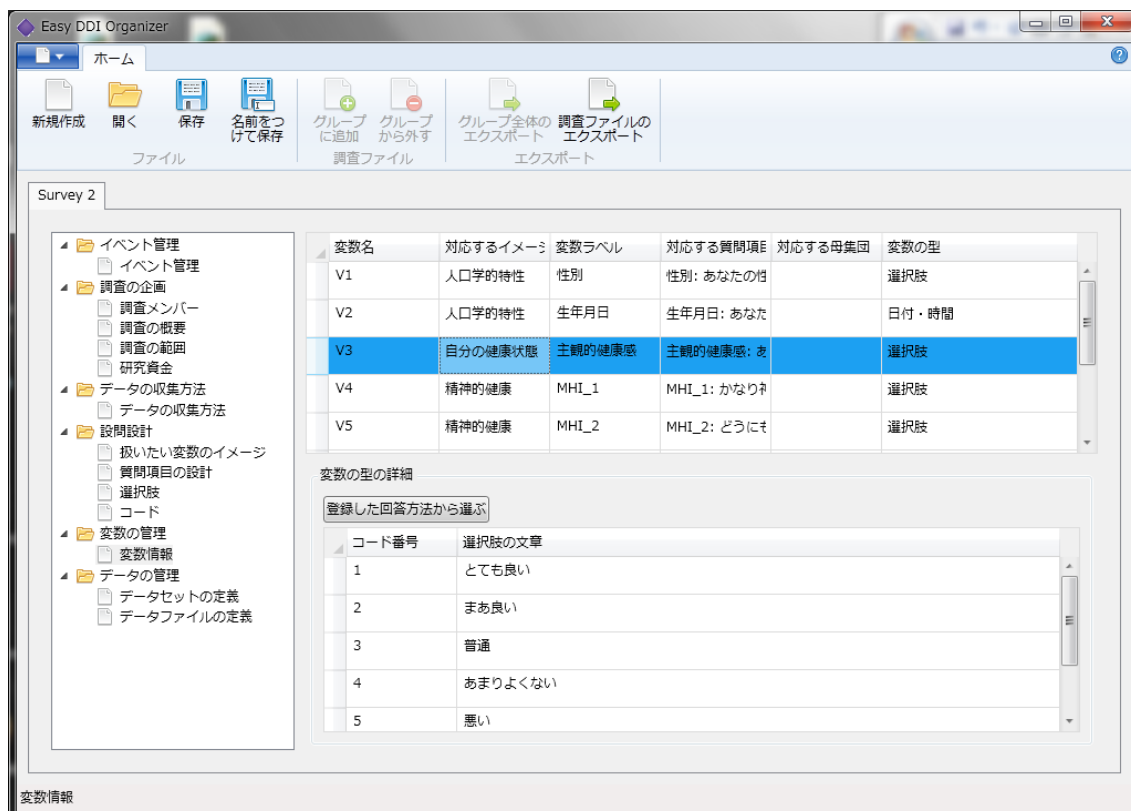


図 11 「変数の管理」画面

(12) 「データセットの定義」画面 (図 12)

この画面は LogicalProduct モジュールの論理レコードに関する部分に対応している。ここでは調査で得られたデータを、記録するデータファイルにどの変数を入れるか、変数の並び順を設定できる。

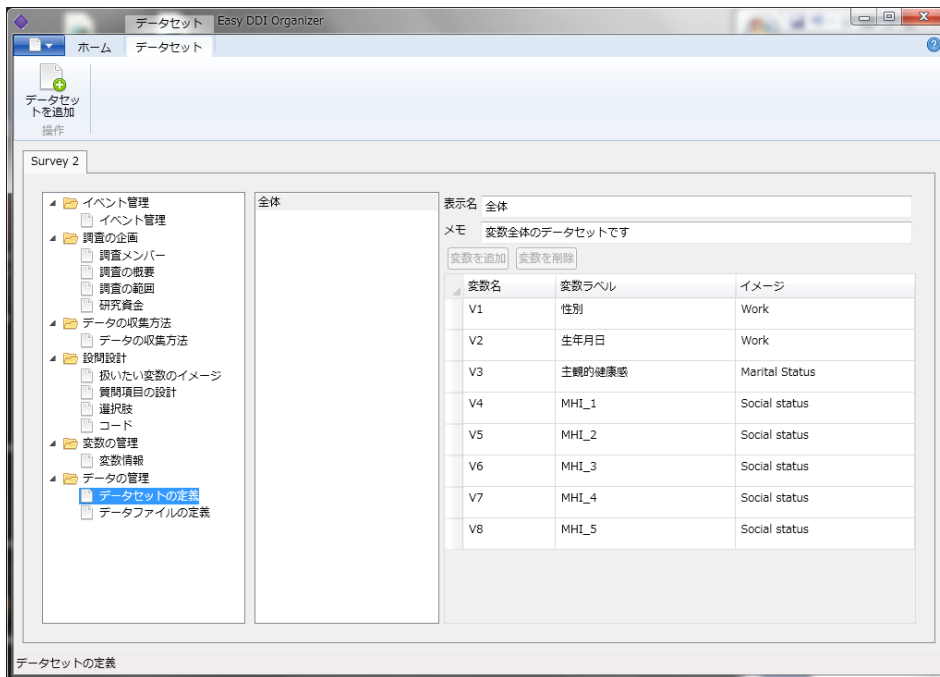


図 12 「データセットの定義」画面

(13) 「データファイルの定義」画面 (図 13)

この画面は PhysicalDataProduct モジュールの PhysicalStructure スキーム、RecordLayout スキームおよび PhysicalInstance モジュールに対応している。ここでは調査データを実際のファイルに保存する際の文字コードや区切り文字、ファイルの保存場所等の設定を記録できる。

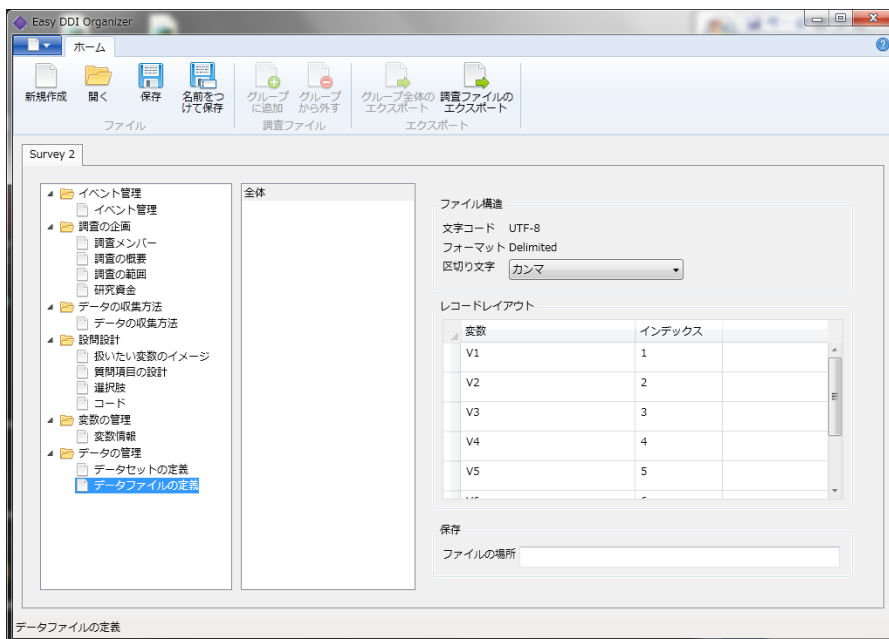


図 13 「データファイルの定義」画面

(14) グループ管理「詳細項目」画面 (図 14)

DDI の Group モジュールを利用することで複数の調査をひとつのグループにまとめ、グループ内で共通する要素をまとめたり、比較したりできる。この画面ではグループに含まれる調査やその調査同士の関係、グループの目的などを記録することができる。

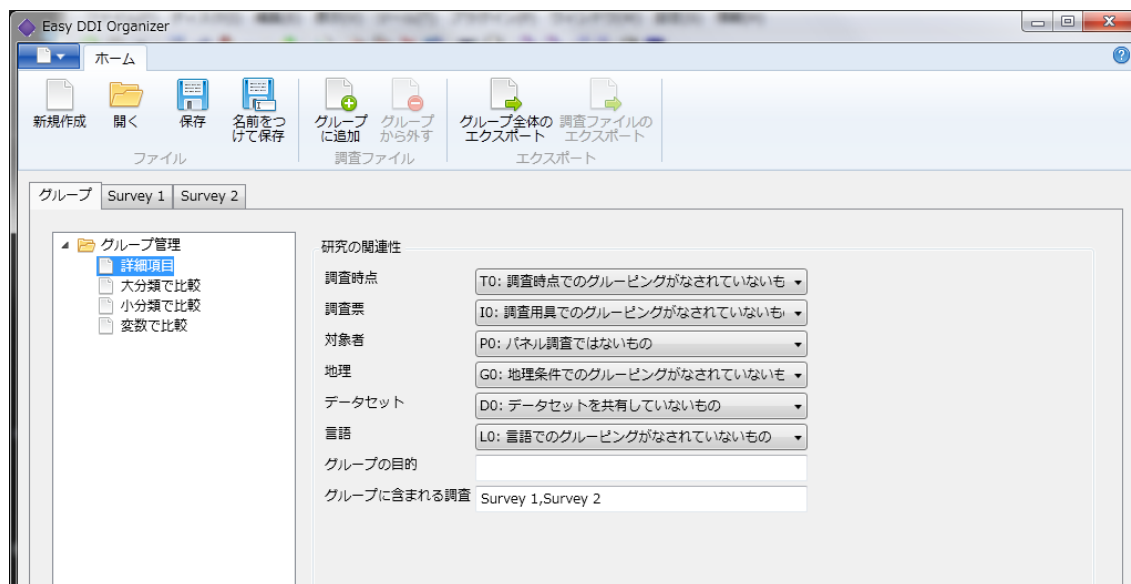


図 14 グループ管理「詳細項目」画面

(15) グループ管理「大分類で比較」 / 「小分類で比較」 / 「変数で比較」画面 (図 15)

これらの画面は Comparative モジュールと対応しており、グループに含まれる調査で使用されている「概念の大分類」、「概念」、「変数」を一覧でき、それぞれの調査に概念や変数が含まれているかや、似たような概念、変数の違いや共通性などを記述することができる。

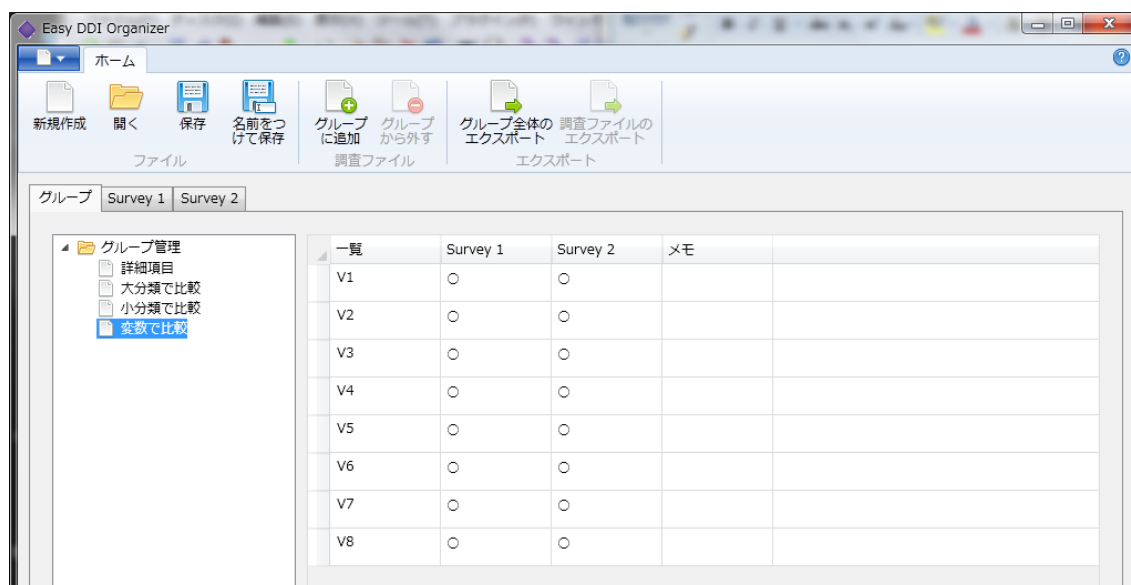
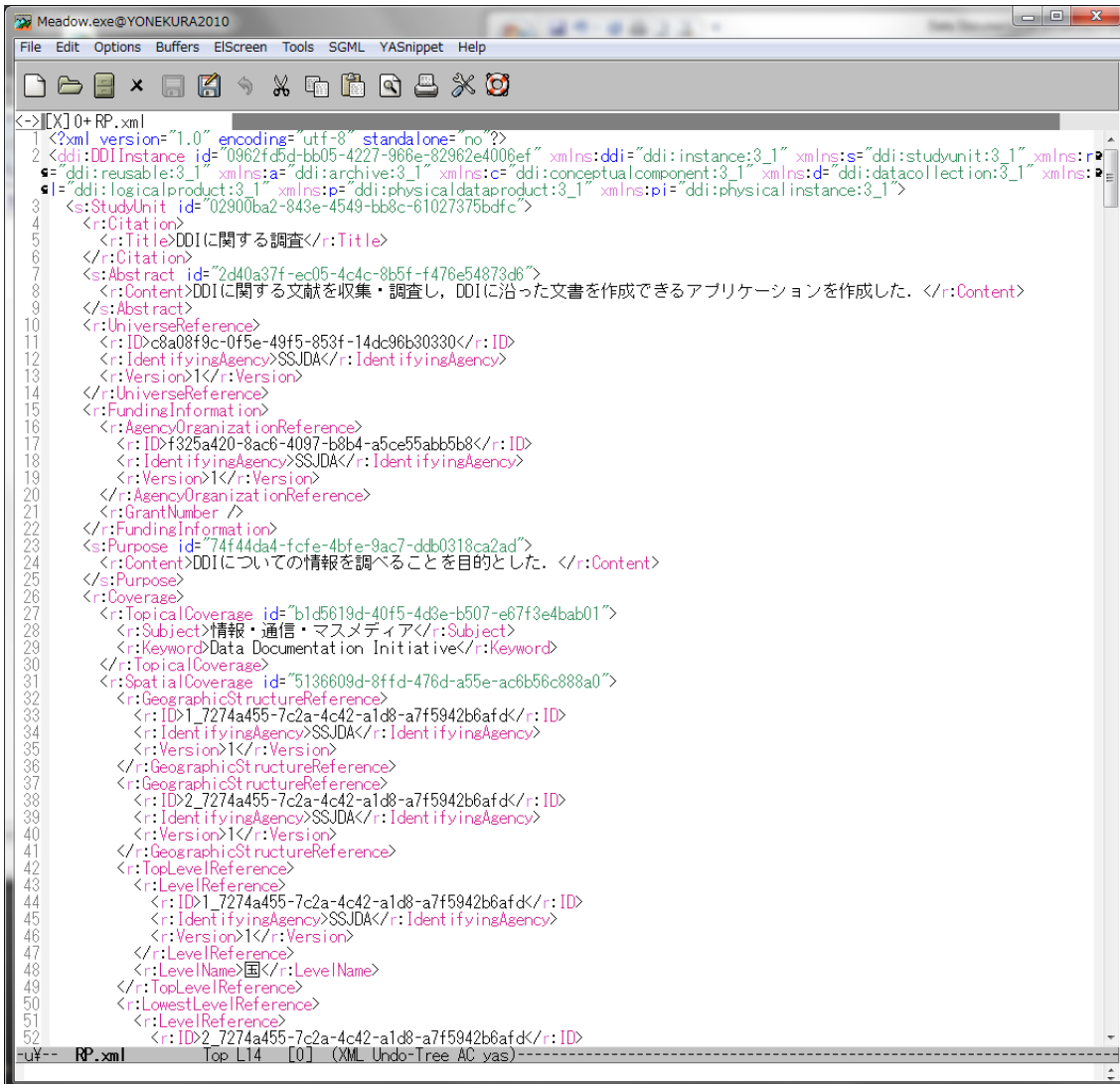


図 15 グループ管理「変数で比較」画面

4.2.DDI 準拠の XML 出力機能

メタ情報編集機能で記録した情報は DDI3.1 準拠の XML で出力することができる。図 16 は EDO で編集したメタデータを DDI3.1 準拠の XML で出力したものである。



```
<?xml version="1.0" encoding="utf-8" standalone="no"?>
<ddi:DDIInstance id="0962fd5d-bb05-4227-966e-82962e4006ef" xmlns:ddi="ddi:instance:3_1" xmlns:s="ddi:studyunit:3_1" xmlns:r="ddi:reusable:3_1" xmlns:a="ddi:archive:3_1" xmlns:c="ddi:conceptualcomponent:3_1" xmlns:d="ddi:datacollection:3_1" xmlns:l="ddi:logicalproduct:3_1" xmlns:p="ddi:physicaldataprod:3_1" xmlns:pi="ddi:physicalinstance:3_1">
  <s:StudyUnit id="02900ba2-843e-4549-bb8c-61027375bdfc">
    <r:Citation>
      <r:Title>DDIに関する調査</r:Title>
    </r:Citation>
    <s:Abstract id="2d40a37f-ec05-4c4c-8b5f-f476e54873d6">
      <r:Content>DDIに関する文献を収集・調査し、DDIに沿った文書を作成できるアプリケーションを作成した。</r:Content>
    </s:Abstract>
    <r:UniverseReference>
      <r:ID>c8a08f9c-0f5e-49f5-853f-14dc96b30330</r:ID>
      <r:IdentifyingAgency>SSJDA</r:IdentifyingAgency>
      <r:Version>1</r:Version>
    </r:UniverseReference>
    <r:FundingInformation>
      <r:AgencyOrganizationReference>
        <r:ID>f325a420-8ac6-4097-b8b4-a5ce55abb5b8</r:ID>
        <r:IdentifyingAgency>SSJDA</r:IdentifyingAgency>
        <r:Version>1</r:Version>
      </r:AgencyOrganizationReference>
      <r:GrantNumber />
    </r:FundingInformation>
    <s:Purpose id="74f44da4-fcfe-4bfe-9ac7-ddb0318ca2ad">
      <r:Content>DDIについての情報を調べることを目的とした。</r:Content>
    </s:Purpose>
    <r:Coverage>
      <r:TopicalCoverage id="b1d5619d-40f5-4d3e-b507-e67f3e4bab01">
        <r:Subject>情報・通信・マスメディア</r:Subject>
        <r:Keyword>Data Documentation Initiative</r:Keyword>
      </r:TopicalCoverage>
      <r:SpatialCoverage id="5136609d-8ffd-476d-a55e-ac6b56c888a0">
        <r:GeographicStructureReference>
          <r:ID>1_7274a455-7c2a-4c42-a1d8-a7f5942b6afd</r:ID>
          <r:IdentifyingAgency>SSJDA</r:IdentifyingAgency>
          <r:Version>1</r:Version>
        </r:GeographicStructureReference>
        <r:GeographicStructureReference>
          <r:ID>2_7274a455-7c2a-4c42-a1d8-a7f5942b6afd</r:ID>
          <r:IdentifyingAgency>SSJDA</r:IdentifyingAgency>
          <r:Version>1</r:Version>
        </r:GeographicStructureReference>
        <r:TopLevelReference>
          <r:LevelReference>
            <r:ID>1_7274a455-7c2a-4c42-a1d8-a7f5942b6afd</r:ID>
            <r:IdentifyingAgency>SSJDA</r:IdentifyingAgency>
            <r:Version>1</r:Version>
          </r:LevelReference>
          <r:LevelName>国</r:LevelName>
        </r:TopLevelReference>
        <r:LowestLevelReference>
          <r:LevelReference>
            <r:ID>2_7274a455-7c2a-4c42-a1d8-a7f5942b6afd</r:ID>
          </r:LevelReference>
        </r:LowestLevelReference>
      </r:Coverage>
    </r:StudyUnit>
  </ddi:DDIInstance>
```

図 16 EDO で出力した DDI3.1 準拠の XML ファイル

5. おわりに

本稿では現在開発中の、日本語で利用可能なメタデータ作成・編集用アプリケーション Easy DDI Organizer (EDO) を紹介した。EDO を用いて調査設計段階から情報管理をすることで、研究グループでのデータ、メタデータの共有が容易になることや、質問紙やコードブック作成の労力を軽減することが期待できる。また、調査が終了した後に、一から調査に関するメタデータを整理する作業が軽減され、調査に関するメタデータの散逸を防ぐとともに、データ共有や二次利用のためのデータ寄託につながる可能性も広がる。国際規格での社会調査のメタデータの蓄積は、我が国の社会調査ストックの質の向上にも寄与しう

ると考えられる。

[謝辞] 開発に協力いただいた有限会社エポック代表取締役の岩波毅氏，プログラマーの津田優氏に深く感謝申し上げます。

[参考文献・URL]

DDI Alliance, 2009, 「 Benefits of DDI | DDI - Data Documentation Initiative」, (2011 年 3 月 18 日取得, <http://ddi.icpsr.umich.edu/ddi-at-work/benefits>).

Thomas W, Gregory A, Gager J, Kuo I-L, Wackerow A, and Nelson C, 2008, Data Documentation Initiative (DDI) Technical Specification Part I : Overview, (2011 年 3 月 18 日取得, . http://ddi.icpsr.umich.edu/sites/default/files/DDI_3_1_2009-10-18_Documentation_XMLSchema.zip)

第5章 Nesstar の特徴と海外データアーカイブにおける利用状況

朝岡誠

1. Nesstar の特徴

1.1 概要

Nesstar は DDI によるメタデータを利用してウェブ上でデータの公開やオンライン集計を実現する統合システムとして、ヨーロッパの社会科学データアーカイブ統括機関である CESSDA (The Council of European Social Science Data Archives), イギリスの UKDA(UK Data Archive)とノルウェーの NDS(Norwegian Social Science Data Services)によって開発された。今日 Nesstar は世界中のデータアーカイブや学術機関で運用されている。

Nesstar は Nesstar Publisher , Nesstar Server , Nesstar Web View の3つのソフトウェアにより構成されている(図1参照)。SPSS や Stata といった一般的なデータセットに調査概要、サンプリング方法などのメタデータを追加作成し、DDI 様式に編集するのが Nesstar Publisher である。そして Nesstar Publisher によって作成された DDI 様式のデータを管理し、インターネット上に公開できるようにするのが Nesstar Server である。さらに Nesstar Server によって公開された DDI 様式のデータは Nesstar Webview によって検索やオンライン分析が可能となる。Nesstar は 1999 年よりテスト運用され、ヨーロッパをはじめ様々なデータアーカイブで用いられている。最近では 2010 年 12 月にバージョン 4.0 がリリースされ、Nesstar Publisher がフリーで配布されるようになった。しかし Nesstar4.0 では、DDI の最新のバージョンである 3.1 には対応しておらず、対応しているバージョンは DDI1.2.2.である。

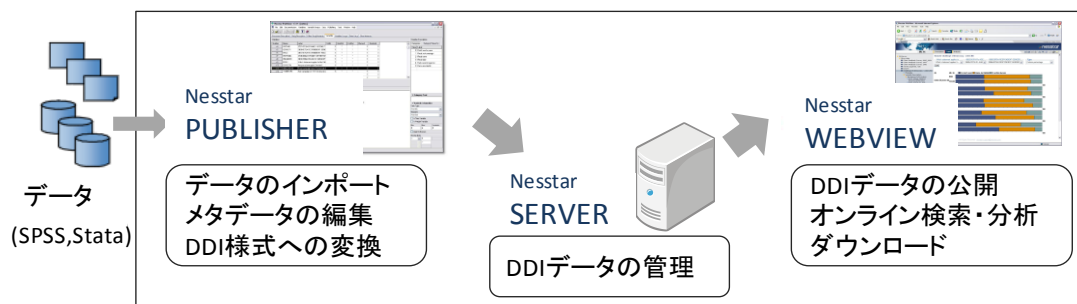


図 1 Nesstar の構成

1.2 利用方法

Nesstar は海外のデータアーカイブではオンライン集計ツールとして用いられているが、日本では馴染みが薄いシステムである。そこで、ドイツの中央アーカイブ (GESIS-ZA) のシステム ZACAT を例に、Nesstar の利用法を見ていく。

(1) データカタログ



図2はZACATで社会科学における国際比較調査(ISSP)の1985年データを選択し、研究概要を表示したものである。各データアーカイブによって多少の違いはあるが、Nesstarシステムによって構築されたウェブページのレイアウトは図2のような形式になる。画面は上のシステムバー、左下のサイドバー、右下のメインウィンドウの3画面で構成されている。サイドバーの  ボタンをクリックすることで下の階層へ移動し、 ボタンをクリックすると調査概要が表示され、さらにこの調査に関するメタデータや変数に関する情報にアクセスできるようになる。図2はISSP→調査年→1985年と選択し、研究概要を表示したところを示したものである。このISSP85の下に[Metadata]と[Variable Discription]が表示されていることが分かる。メタデータのうち[Study Discription]を選択することで調査責任者、サンプリング方法、関連する調査など情報を知ることができ、[Data Fife Description]を選択すると調査データに関する情報を知ることができる¹。また[Variable Description]を選択することでその調査で使われている変数一覧が表示され、その変数の記述統計を知ることができる。次頁図3はCOUNTRYの記述統計を表示したものである²。



図2 ZACAT ウェブサイト(<http://zacat.gesis.org/webview/index.jsp>)

¹ データによってはメタデータ情報のみで変数情報がないものもある。

² Nesstar3.5以降ではこれに加え平均値も表示される。

The screenshot displays the SPSS 'Variable View' for the variable 'v3: COUNTRY'. The left pane shows a tree structure of variables, with 'v3: COUNTRY' selected. The right pane provides details for this variable.

Variable v3 : COUNTRY

Literal Question
Country

1 Australia (AUS)
2 Germany (D)
3 Great Britain (GB)
4 USA (USA)
5 Austria (A)
8 Italy (I)

Values Categories

Value	Category
1	AUS
2	D
3	GB
4	USA
5	A
8	I

N = Frequencies based on unweighted data; NW = Frequencies based on weighted data

Summary statistics

Valid cases	7350
Missing cases	0
Minimum	1.0
Maximum	8.0

This variable is numeric

図 3 記述統計の表示

(2) 検索システム

Nesstar では、複数のデータセットのなかから、必要な変数を検索することができる。詳細検索では、調査の概要や変数の説明など、検索範囲を指定したうえで必要な変数を探ることができる。また、検索対象も、調査の概要、変数、図表というかたちで絞っていくことができる。調査の概要をとっても、対象地域やサンプリング過程など、より詳細な検索の指定が可能である。

たとえば、図 4 は **»** ボタンを押し、詳細検索画面を呼び出し、変数の説明のなかに「education (教育)」という言葉を含む変数を検索するところである。この画面は図 1 にて黒丸で囲んだアイコンをクリックすることで表示することができる。このように分析に使いたい変数がある調査票を逆引きで検索することができる。

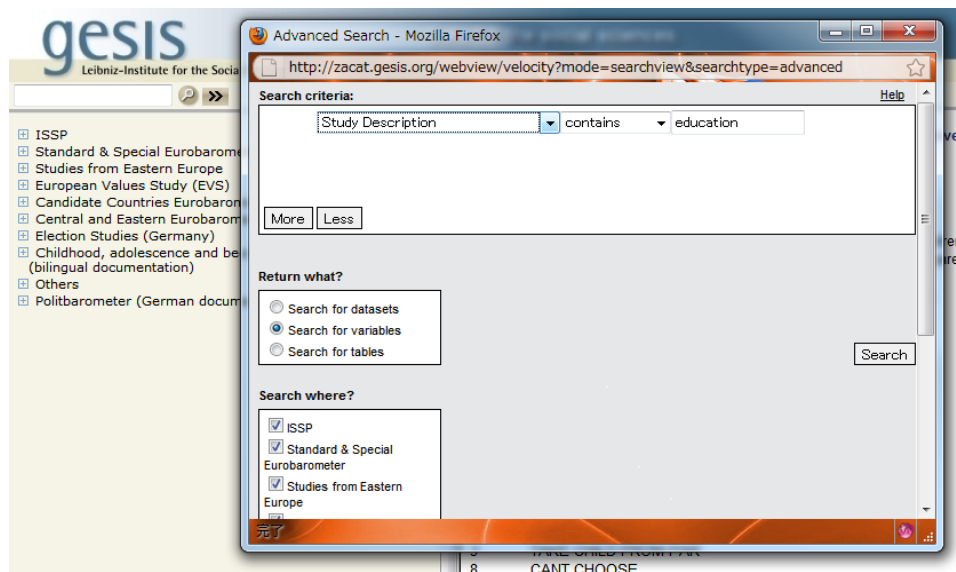



図 4 詳細検索

(3) ダウンロードシステム

Nesstar はダウンロードシステムを備えており、調査ページから調査データを直接ダウンロードすることができる。システムバーにあるフロッピーのアイコン() をクリックすることで図 5 のようなダウンロード画面が表示され、ダウンロードをするデータの形式 (SPSS,SAS など) を選択することができる。データの中には特定の変数のみを選んでダウンロードすることができるものもある。ユーザー管理は可能であるが、利用登録をしたユーザーは、掲載されているデータを即時にダウンロードができるため、寄託者承認などが必要なデータを搭載するには、相応のシステム調整が求められるという問題がある。

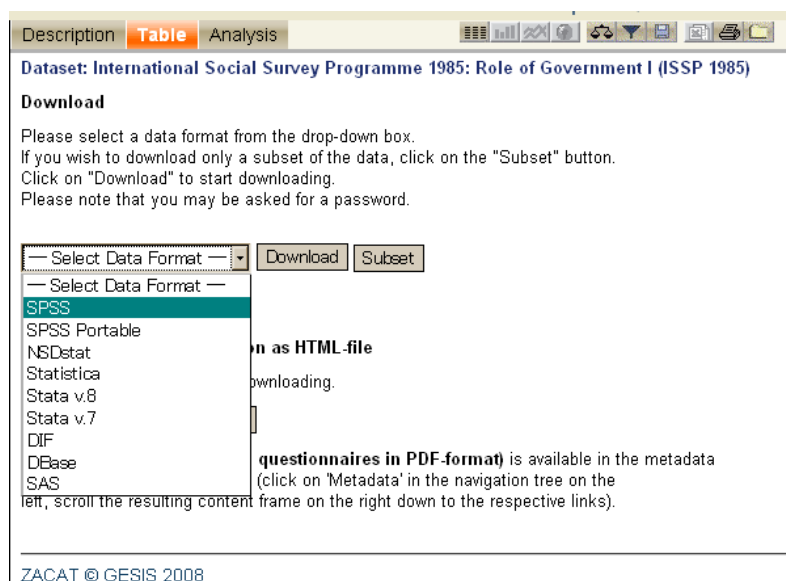


図 5 ダウンロード画面

(4) オンライン集計

- ・ クロス表, グラフの作成

Nesstar はサイドバーに表示された変数を用いて簡単なオンライン分析を行うことができる。メイン画面の「Table」を選択した後、変数をクリックすると、その変数をクロス表に投入することができる。図 6 は ISSP1985, 1990, 1996, 2006 の合同データを用いてクロス表を作成したものである。ここでは列に調査年, 行に政治に対する関心度(How much interested in politics)」を指定している。さらに、クロス表を表示した状態でグラフアイコンを選択することで棒グラフが表示される³。なお、行や列に複数の変数を指定することで多重クロス表も作成することができる。しかも層の指定も可能であり、層を指定することで条件別のクロス表を表示することもできる。そして、これらの分析の結果は xml 形式でダウンロード、または印刷を行うことで保存することができる。

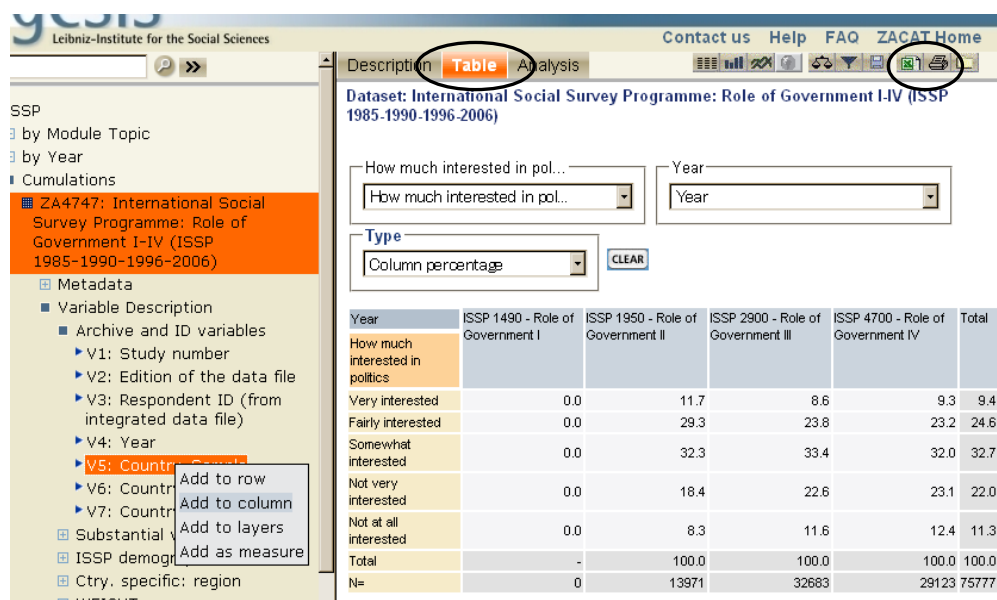


図 6 クロス表の表示

³時系列に関連する変数を指定している場合、折れ線グラフを表示することもできる

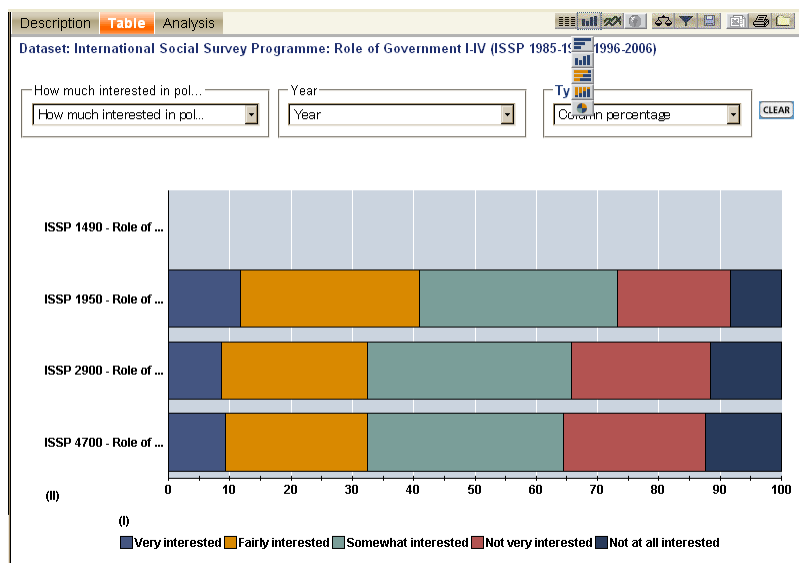


図 7 棒グラフの表示

・ 変数の加工

そして、バージョン 3.5 以上の Nesstar では、相関分析や回帰分析を行うことができる。ZACAT で用いられている Nesstar は 3.0 バージョンであるためこれらの分析を行うことは出来ない⁴。そこで、NORC で運用されている Nesstar を利用し、変数加工について見る。なお変数の加工を行うためには HP 上で利用者登録を行う必要がある。

図 8 は、○内をクリックし、変数の加工画面が現れたところである。四則演算に平均、増加率の計算の他に値の再割当て(Recode)も可能である。図 9 は Recode を選択し、値の再割当てを選択した画面である。この画面から変数の割り当てができる。また、変数の割り当ての際「as missing」をクリックしておけば、欠損値として指定することができる。

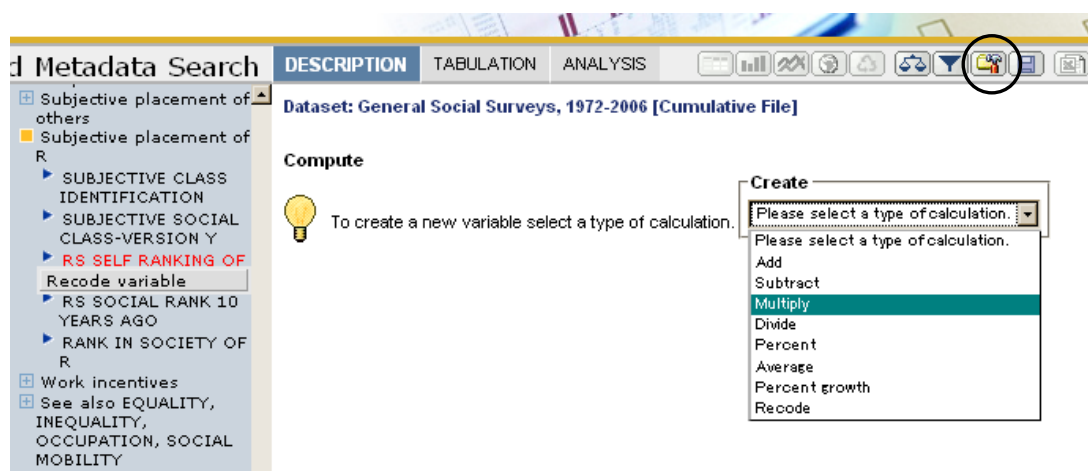


図 8 変数の加工

⁴ Nesstar 3.0 でも重回帰分析が可能だが、変数の加工ができない。したがって欠損値処理ができないため、分析の際には大きな制約がある。

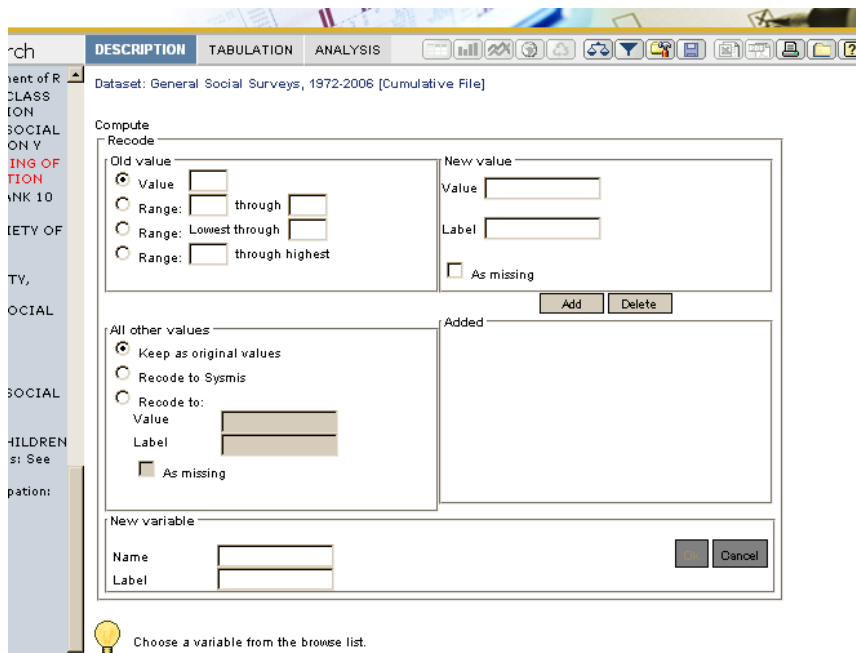




図 9 値の再割当て

・ 統計分析

Nesstar では、相関分析と回帰分析を行うことができる。図 10 は世帯収入と教育年数と年齢の相関分析の結果であり、図 11 は世帯収入を従属変数とし、教育年数と年齢を独立変数とした重回帰分析の結果である。また、 アイコンをクリックすることでウェイト付け、 アイコンをクリックすることでケースの選択が可能である。

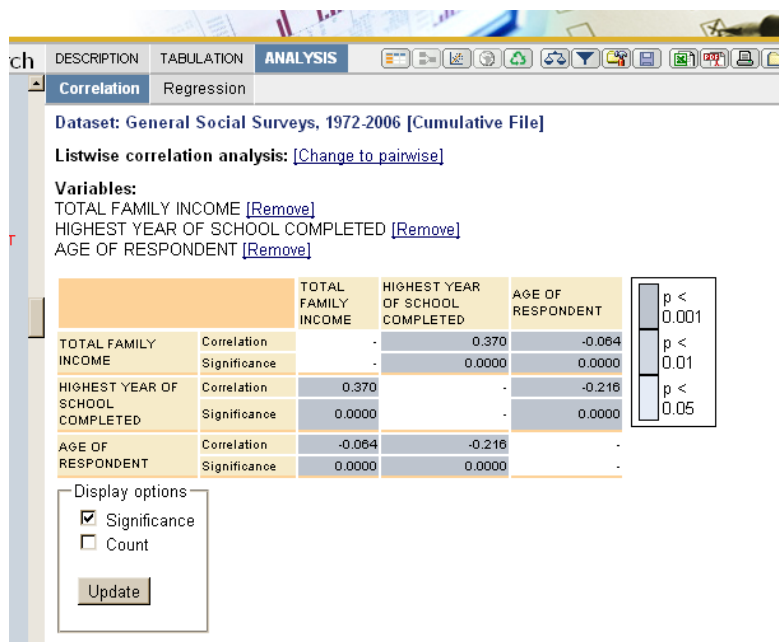


図 10 相関分析

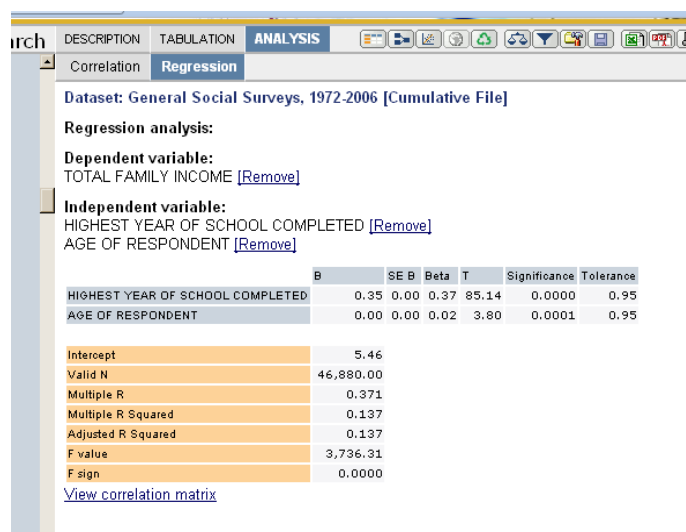


図 11 回帰分析

(5) 多言語対応

Nesstar は積極的に多言語対応しており, 英語以外の言語にも対応している. これは Nesstar がヨーロッパの各データアーカイブの環境にマッチすることを前提に開発されたためであり, 漢字やハングル語といったマルチバイト文字にも対応している⁵. 図 12 は台湾のデータアーカイブ SRDA(Survey Research Data Archive)における Nesstar 利用をみたものである. このように変数名やラベルを漢字で表記することができる. また, 様々な表記に関してカスタマイズが可能とされている.



図 12 RSDA(台湾)の Nesstar 利用

⁵ Nesstar Publisher を用いて SPSS データを DDI 様式にインポートする際, 変数ラベルが文字化けしてしまうため完全に対応しているとはいえない.

2. 海外アーカイブにおける Nesstar の利用状況

2.1 CESSDA における Nesstar の利用

CESSDA は各組織が所有するデータの情報を共有化し、アーカイブ同士で協力してアーカイブを運営するために 1976 年に結成された組織である。図 13 は CESSDA のメンバーを示したものである。メンバー全体で 25000 ほどのデータを所有し、毎年 30000 人ほどの研究者に対して、延べ 70000 のデータを提供し、毎年 1000 ほどのデータを追加している。Nesstar はこれらの 25000 のデータを統合するウェブ上の仮想アーカイブを設立するためのシステムとして開発されたという経緯がある。第二章でみたように CESSDA のウェブサイト上の CESSDA Catalogue から CESSDA に加盟するアーカイブのデータを検索し、各アーカイブのウェブサイトにアクセスすることができる。最近では GESIS を介して EDAN(The East European Data Archive Network)に加盟する東欧・ロシアのアーカイブのデータも検索できるようになっている。



図 13 CESSDA に加入しているデータアーカイブ

(出所：<http://www.cessda.org/about/members/>)

次に CESSDA に加盟するアーカイブ内で Nesstar がどのように利用されているのかを見てみよう。Nesstar は先に見たようにデータカタログ、検索、ダウンロード、オンライン集計

の 4 つの機能を合わせ持つシステムであるが、その機能を制限して運用しているアーカイブもある。例えばフィンランドのデータアーカイブ FSD(Finnish Social Science Data Services)では Nesstar のデータカタログ機能に特化し、オンライン集計をオミットしている。ダウンロードはメタデータのみを制限し、データのダウンロードは CD-ROM で行っている。しかし多くのアーカイブでは Nesstar をオンライン集計ツール、ダウンロードによるデータ提供ツールとして用いており、一部のデータのみ Nesstar で提供するというスタイルを採用している。このようなデータアーカイブでは Nesstar 以外の検索システムやデータカタログシステムを持っており、これらのシステムから Nesstar のページに誘導している。

2.2 その他の組織における Nesstar の利用

ヨーロッパに比べ、アメリカでは Nesstar を利用しているアーカイブは少ない。シカゴ大学の NORC (National Opinion Research center)では、オンライン集計システムとして SDA とともに Nesstar も利用されており、またウィスコンシン大学の BADGIR (Better Access to Data for Global Interdisciplinary Research) においても運用されている。その他にはオーストラリアのデータアーカイブ ASSDA (Australian Social Science Data Archive)、ニュージーランドのデータアーカイブ NZSSDS (New Zealand Social Science Data Service) でも統合環境として運用されている。また、イスラエルのデータアーカイブ ISDC (Israel Social Science Data Center) でも Nesstar を用いている。ISDC では、Nesstar をデータカタログとして用いるために、オンライン集計機能やダウンロード機能を除外している。なお、最近では台湾のデータアーカイブ SRDA(Survey Research Data Archive)、韓国のデータアーカイブ KOSSDA(Korea Social Science Data Archive)でも Nesstar が運用されている。詳細は、巻末資料編を参照頂きたい。

データアーカイブ以外の公的組織では、イースト・サセックス州議会や東イングランド公衆衛生協会、ノルウェーの国立公衆衛生センター、カナダ統計局などで用いられており、特に公衆衛生の分野で用いられている。また、最近では ESS (European Social Survey) などの国際比較調査プロジェクト単位で Nesstar を導入し、自分たちの行った調査を積極的に公開している例もある。このように Nesstar はデータアーカイブ以外でも利用されている事例が見られる。

2.3. Micro data Management Toolkit について

近年、アフリカなどの発展途上国のアーカイブを中心に世界銀行等の支援によって開発された Micro data Management Toolkit が用いられている。Micro data Management Toolkit は Nesstar の簡易版として開発されている。Nesstar がデータカタログ、検索、オンライン集計、ダウンロード機能を持っているのに対し、Micro data Tool Kit では DDI 形式でメタデータの編集は行いが、オンライン集計やダウンロード機能はもたず、データ配布は CD-ROM による。CD-ROM によってデータを配布された利用者は、同封されている Nesstar Explorer を使

って DDI 様式のファイルのメタデータ情報等を読み込む。この Micro data Management Toolkit は現在では無料で配布しているため、誰でもフリーで入手することができる。

[参考にしたウェブサイト一覧 (閲覧日 2011 年 3 月 25 日)]

- ASSDA (Australian Social Science Data Archive) <http://assda.anu.edu.au>.
- BADGIR (Better Access to Data for Global Interdisciplinary Research) <http://nesstar.ssc.wisc.edu/>.
- CESSDA (Council of European Social Science Data Archive) <http://www.CESSDA.org/>.
- ESSDA (Estonian Social Science Data Archives) <http://psych.ut.ee/esta/>.
- GESIS-ZA (German Social Science Infrastructure Services / Gesellschaft Sozialwissenschaftlicher Infrastruktureinrichtungen ZACAT) <http://zacat.gesis.org/webview/index.jsp>.
- ICPSR (*Interuniversity Consortium for Political and Social Research*)
<http://www.icpsr.umich.edu/icpsrweb/ICPSR/>.
- IHSN (International Household Survey Network) Microdata Management Toolkit
<http://www.surveynetwork.org/home/index.php?q=tools/toolkit>.
- ISDC (Israel Social Science Data Center) <http://isdc.huji.ac.il/>.
- NORC (National Opinion Research Center) <http://www.norc.org/GSS+Website/>.
- Nesstar <http://www.nesstar.com/>.
- NZSSDS (New Zealand Social Science Data Service) <http://www.nzssds.org.nz/>.
- NSD (Norwegian Social Science Data Services) <http://www.nsd.uib.no/index.html>.
- SRDA (Survey Research Data Archive) <http://srda.sinica.edu.tw/>.
- UKDA (UK Data Archive) <http://www.data-archive.ac.uk/>.

第6章 SSJDA における Nesstar 試験導入

入山浩一 角井佑 米倉佑貴 佐藤慶一

1. はじめに

本節では、Nesstar Server 4.0 (以後 Server), Nesstar Webview 4.0 (以後 Webview), Nesstar Publisher (以後 Publisher) で構成されている Nesstar を実験サーバにインストールし、サンプルデータ¹を用いて Nesstar の導入実験を行う。インストールから調査データの導入と実操作までの手順を以下で記述する。

- 1.1 Nesstar のインストールの手順
- 1.2 Publisher と Server による調査データの公開手順
- 1.3 Webview での分析手順

1.1 Nesstar のインストールの手順

Getting Started Guide Version4.0² を参考に、Server と Webview, Publisher のインストールを行う。

(1) インストールのシステム要件

インストールするコンピュータのシステム要件は下記の通りである。

- OS : Windows XP, 2000 or NT4, Windows Server 2003
- CPU : 1 GHz CPU
- メモリ : 1GB
- その他 : graphics card

今回のインストールは下記の環境で実施している。

- OS : Microsoft Windows Server 2003 R2 Standard Edition Service Pack 2
- CPU : 2core 2.8GHz CPU
- メモリ : 2GB

(2) Server と Webview のインストールの手順

- ① 入手したNesstar-Server-4.0.exeをダブルクリックする。

インストール実行ファイルの入手方法は下記 URL を参照されたい。

<http://www.nesstar.com/>

¹東大社研・若年パネル調査 (JLPS-Y) wave1, 2007

² Nesstar の WEB サイト (<http://www.nesstar.com/>) では、version4.0 のマニュアルをダウンロードできないため、カスタマーサポート (nesstar-support@nesstar.com) に問い合わせて入手している。

- ② 「はい」 をクリックする。

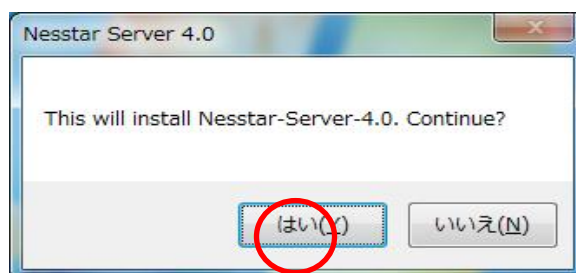


図 17 インストール継続確認画面

- ③ ライセンスに同意する。

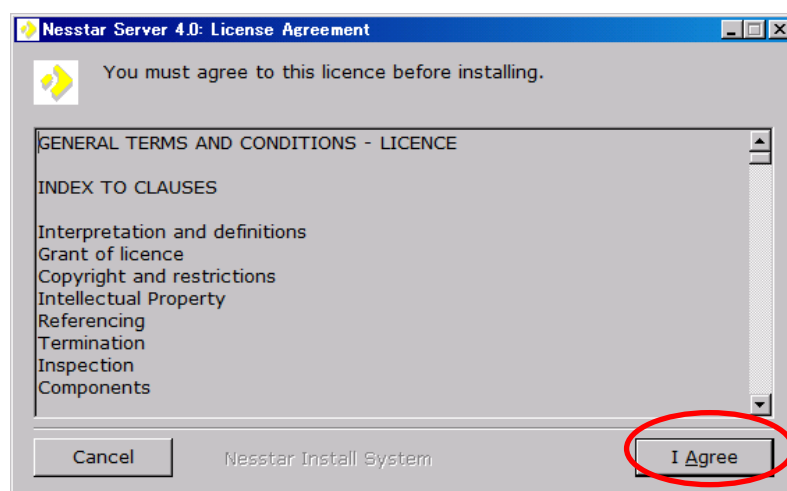


図 18 Server ライセンス同意画面

- ④ インストール先の設定，デスクトップへのショートカット作成，スタートメニューのフォルダ名を設定後，Configuration Tool が開始される。
表 7 の項目を入力し，Apply Settings をクリックする。

表 7 Configuration Tool の入力項目

入力項目	入力内容
License Code	メールで届いたキーを入力する。
Server Name	サーバ名を入力する。
IP Address	サーバの IP アドレスを入力する。 他の PC に接続させないローカルな環境でよい場合は、127.0.0.1 または localhost を選択する。
HTTP Port	通常は 80 だが，別ポートの指定も可能である。
Administrator name	管理者の名前を入力する。
Administrator password	管理者のパスワードを入力する。

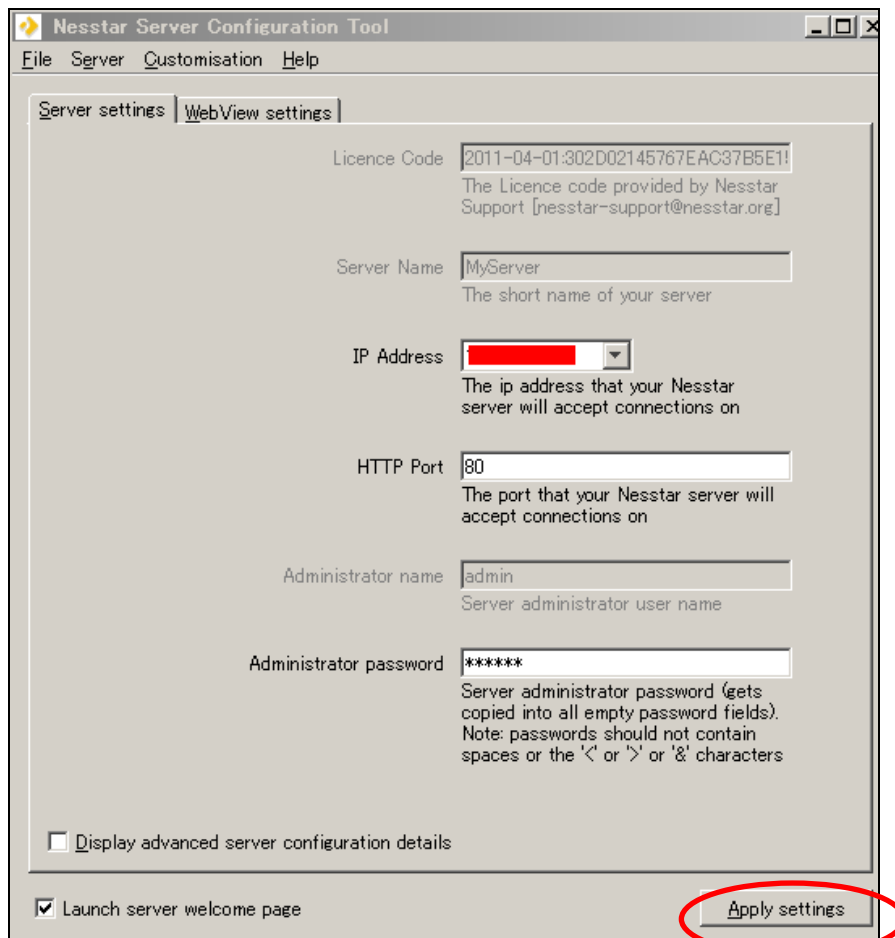


図 19 Configuration Tool 設定画面

- ⑤ インストールが完了すると、WebView の初期ページが表示される。

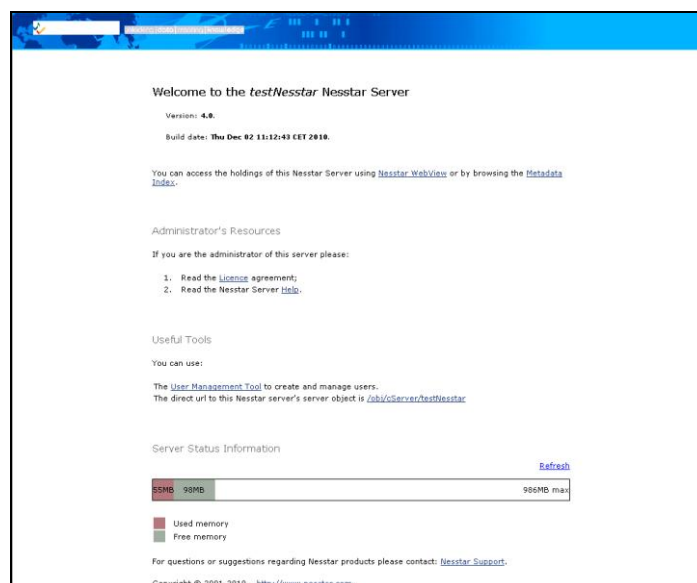


図 20 Webview 初期ページ画面

(3) Publisher のインストールの手順

① 入手した NesstarPublisherInstaller_v4.0.2.exe をダブルクリックする。

インストール実行ファイルの入手方法は下記 URL を参照されたい。

<http://www.nesstar.com/>

② 「Next」をクリックする。

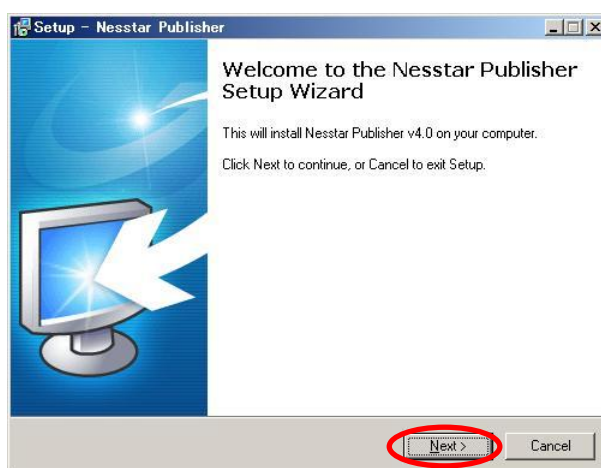


図 21 Publisher セットアップスタート画面

③ ライセンスに同意する。

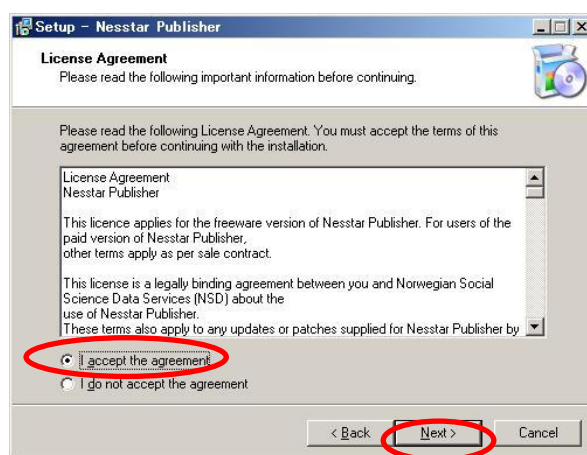


図 22 Publisher ライセンス同意画面

④ インストール先の設定，スタートメニューのフォルダ名を設定，アイコン追加の設定後， Publisher が起動する。

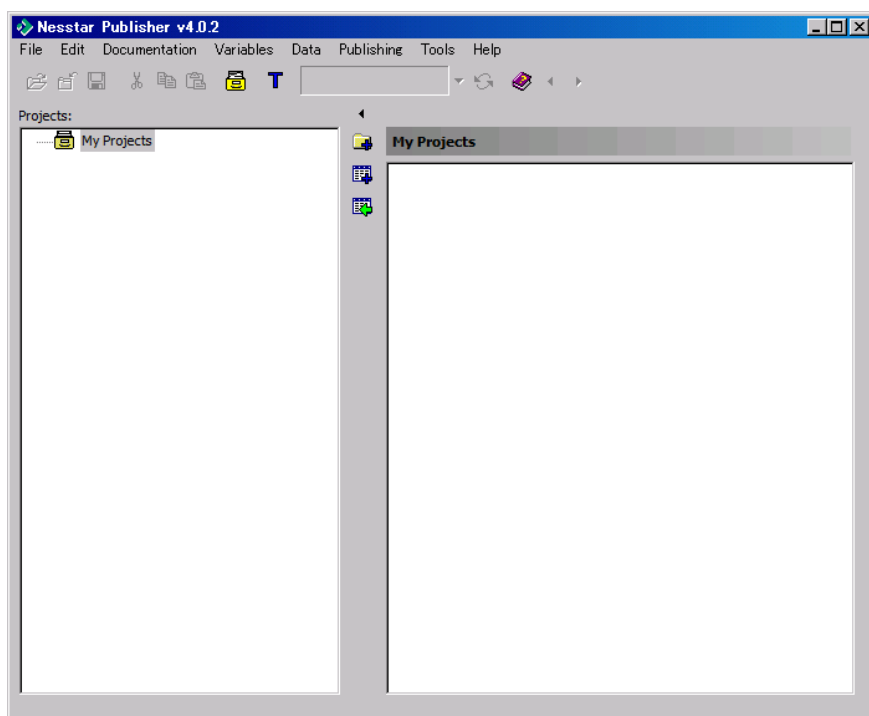


図 23 Publisher 起動初期画面

Nesstar のインストール作業は以上で完了である。

1.2 Publisher と Server による調査データの公開手順

調査データを公開し Webview で分析できるようにするには、(1) Publisher を用いてデータをインポートして DDI 形式のファイルを作成し、(2) 次にメタデータの追加を行い、(3) 最後に Server の公開の設定を行う必要がある。以下では、この手順を説明する。

(1) Publisher を用いたデータのインポートと DDI 形式ファイルの作成

Publisher を起動して、図 8 の通りに画面上のメニューバーの「File」→「Import Study」をクリックすると、インポートするファイルを選択するダイアログが表示される。なお、インポートできるデータの形式は以下の通りである。

- Nesstar(.Nestar)
- NSDstat(.NSDstat)
- DDI Document(.xml)
- SPSS(.sav,.por)
- SPSS Syntax(.sps)
- STATA(*.dta)
- Statistica(*.sta)
- NSDstat(*nsf)

- dBase(*.dbf)
- DIF(*.dif)
- Delimited Text(*.txt, *.csv, *.sdv, *.cdv, *.prn,)
- PC-Axis(*.px)
- Excel(*.xls)

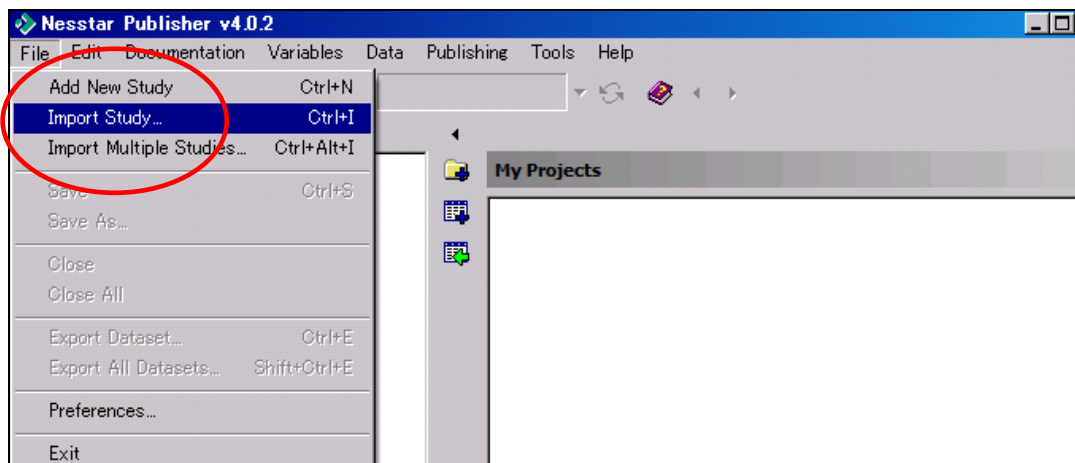


図 24 データのインポート選択

インポートされたデータはプロジェクトという単位で管理される。プロジェクトには複数のデータの他に、多次元表（Cube data）も格納することができる。図 9 は spss ファイル形式のサンプルデータをインポートした結果の画面であり、画面左のリストメニュー「My Projects」配下にデータが格納され、データにはメタデータ情報と実データを含んだ項目（Document Description, Study Description, Other Study Materials, Datasets, Variable Groups, Other Materials, External Resources）が追加される。項目の詳細については、(2) で後述する。

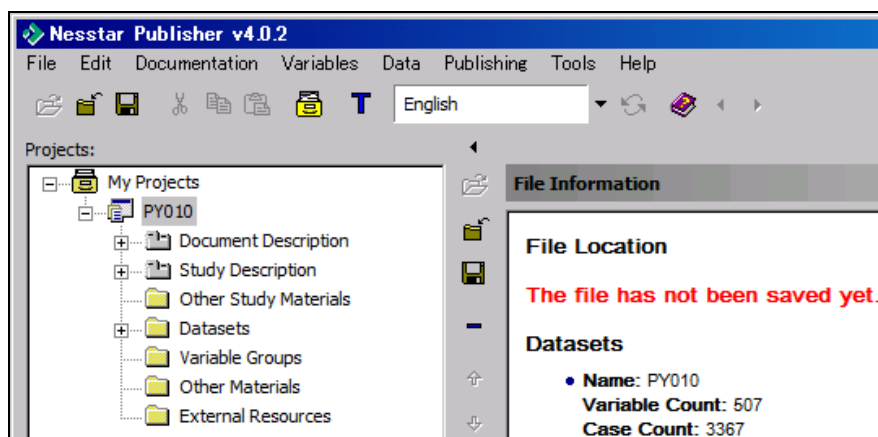


図 25 データインポート後の画面

インポートされたデータの変数情報が知りたいときには、図 10 の通りに、画面左のリス

トメニューの「Dataset」の配下にある「Variables」をクリックする。インポートされた変数情報の一覧表が表示され、変数名をクリックすると記述統計が表示される。図 11 は、画面中央の変数一覧表の「問 51_結婚時の満年齢」をクリックし、記述統計を出した画面である。記述統計では、度数分布、有効なケースの数、最大値、最小値が出力される。

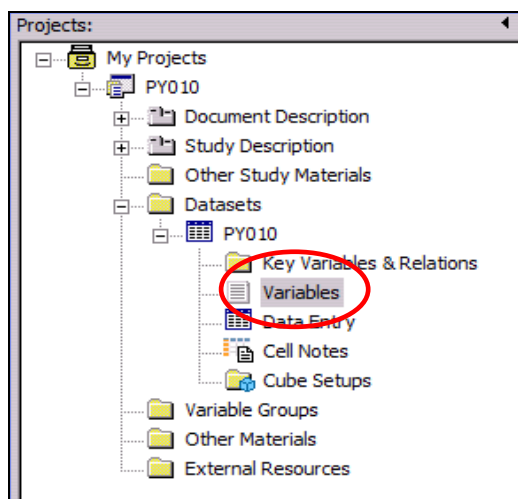


図 26 Variables の選択

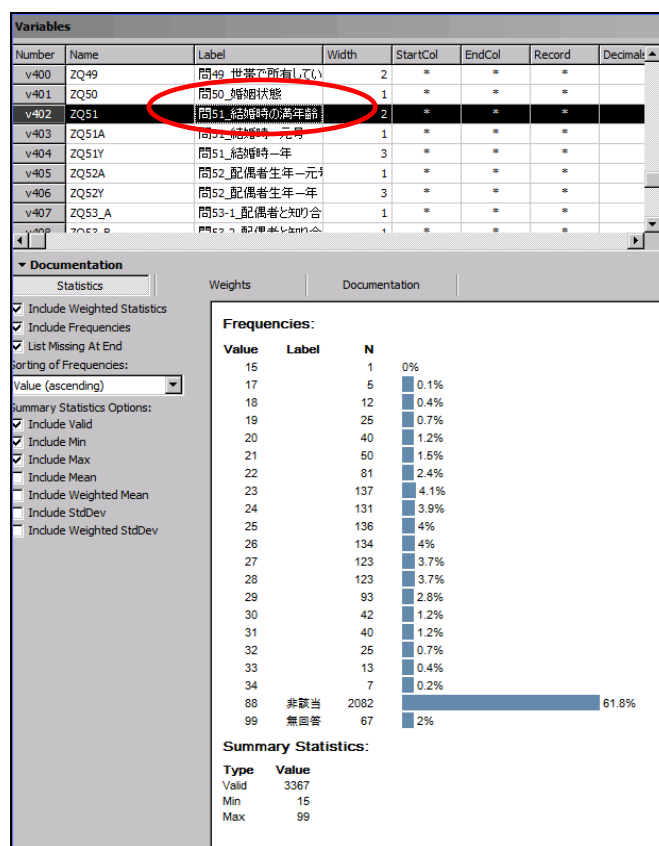


図 27 変数情報の表示

最後に、インポートされたデータを保存する。図 12 の通りに「File」→「Save」をクリ

ックして、Nesstar 形式 (.Nesstar) で保存する。

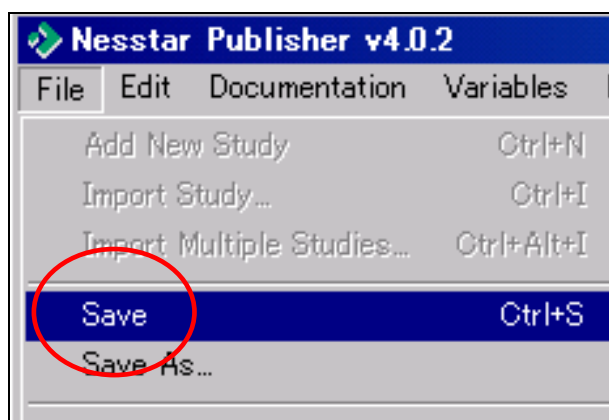


図 28 保存の選択

調査データのインポートは以上で完了である。

(2) メタ情報の追加

本節では、前節で **Publisher** にインポートしたサンプルデータにメタ情報を追加していく。なお、付加情報については、現在、同データアーカイブで公開している概要情報と調査票の情報をもととする。

Publisher には付加情報を追加するためのテンプレートが準備されており、それによってメタ情報を編集することができる。このテンプレートの入力項目は DDI 1.2.2 で定義されている項目の一部が使用されている。カスタムテンプレート機能を使うことで DDI 1.2.2 で定義されている他の入力項目をテンプレートに追加したり、DDI 1.2.2 にはない入力項目を追加することもできる。**Publisher** の標準のテンプレートの入力項目は大きくは表 8 に示す 7 項目からなっている。

表 8 Publisher の入力項目

分類	内容
Document Discription	マークアップしたドキュメントおよびそれを作成するために使用した文書 (コードブック等)に関する情報を記録するセクションである。また、文書の使用方法和文書の状態についても記録することもできる。
Study Description	DDIに準拠したドキュメントファイルに記載されているデータについての名前および関連するすべての情報が含まれている。このセクションでは、調査データに関する引用・主な研究員、提供者、キーワード、要約、データの収集方法、地理的・時間的情報、処理方法についての情報を入力する。また、本項目では、それ以外にも豊富な情報を入力でき、調査の背景の情報を網羅できる。
Datasets	変数などの具体的内容についてDDIおよびPublisher独自の項目を使用して情報を入力する項目である。この項目では変数情報の他に多元表データや各変数の階層構造なども設定できる。また、データは複数設定でき、また既存のデータを編集した(変数などを加工した)統計データを新規作成することもできる。
Variable Groups	本項目では、各変数をグループ化する設定が行うことができる。
Other Study Description Materials	Study Descriptionに関連した資料に関する情報を記述する。たとえば、サンプリングに関する情報、weightの詳細、方法論・技術の詳細などである。
Other Materials	other study-related materialsに関する情報を記述する。これに含まれるのは、質問紙(調査票)、コーディング情報、ソフトウェアの詳細情報、ユーザマニュアル、その他のガイド、プログラムのサンプル、用語集、インタビュー/プロジェクト手順、マップ、データベーススキーマ、データ辞書、showカード、誤値の情報、インタビューのスケジュール、度数ファイルなどである。
External Resources	Dublin Coreやe-GMSのメタ情報で記述されたWordファイルPDFファイル、地図のようなものを扱う項目である。

以下それぞれについて詳細を述べていく。

① Document Description

Document Description は、編集するメタ情報の文書の引用情報に関するセクションである。標準のテンプレートで入力できる情報は表 9 のとおりである。

表 9 Document Description の入力項目

項目グループ	入力項目	内容
Citation-Title Statement	Title	DDIドキュメントの正式なタイトルを入力する。(大抵は調査名を入力する。)
	ID Number	アーカイブなどでつけるような一意となるIDを入力する。
Citation-Responsibility Statement	Authoring Entity/Primary Investigator	編者、主な研究者(研究代表)を入力する。
	Other Identifications/ Acknowledgments	その他の貢献者を入力する。
Citation-Production Statement	Copyright	文書の著作権に関する情報を入力する。
	Producers	文書の作成のために出資した人/組織または文書の作成を管理していた人/組織に関する情報を入力する。

Document Description の項目を入力し、Document Description をクリックすると図 29 のように表示される。

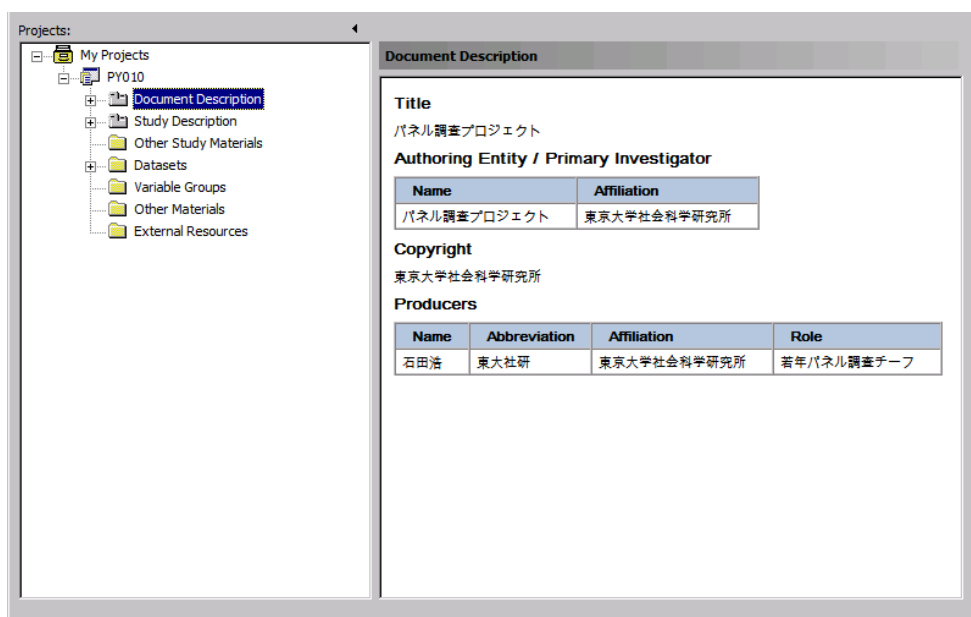


図 29 Document Description 画面

② Study Description

Study Description は、研究の背景情報を記述するセクションである。ここで入力できる情報は表 10 の通りである。

表 10 Study Description の入力項目

項目グループ	入力項目	内容
Citation	Title	DDIドキュメントの正式なタイトルを入力する。(大抵は調査名を入力する。)
	ID Number	アーカイブなどでつけるような一意となるIDを入力する。
	Authoring Entity/Primary Investigator	編者、主な研究者(研究代表)を入力する。
	Distributors	データを提供する個人・組織を入力する。
	Version	調査データのバージョンを入力する。
Citation-Production Statement	Producers	データ作成の財政面の責任者/責任組織またはデータの作成を管理していた人/組織に関する情報を入力する。
	Fundings	データ作成のための資金に関する情報を入力する。
Scope-Subject Information	Keywords	調査を特徴づけるキーワードを入力する。
	Topic Classifications	調査を分類するトピックを入力する。
Abstract	Abstract	調査の概要を記述する。入力形式は自由記述であり、主には調査の要約や目的、特質、研究分野、主な変数などについての情報を記録する。
Scope-Summary Data Description	Countries	調査が行われた国を入力する。
	Geographic Coverage	調査がカバーする地域の範囲を入力する。地域を識別する変数がある場合はその情報も加えてもよい。
	Unit of Analysis	分析や観察の単位を記述する。個人、家族、集団、組織等。
	Universe	調査の対象とした母集団に関する情報を記述する。
Methodology-Data Collection	Time Method	データ収集の時期を入力する。
	Sampling Procedure	サンプリング方法を入力する。
	Mode of Data Collection	データ収集の方法を入力する(郵送調査、電話インタビュー等)。
	Weighting	データ解析の時に使用するweightに関する情報を記録する。weightの計算式や係数などを作成した場合はそれも記述する。

③ Other Study Materials

Other Study Materials は、研究の背景情報に関する資料のタイトル、所在、内容を記述するセクションである。資料は表 11 のとおり、Related Materials, Related Studies, Related Publications, References の 4 つに分類されている。

表 11 資料の分類項目

分類	内容
Related Materials	研究の背景情報を補足する情報について記述する。附録やサンプリングに関する情報など。
Related Studies	本調査データと他のデータ(先行, その後のデータ, 他のウェブ, 他の段階)もしくは同一ファイルの他のエディションとの関係を入力する項目である。
Related Publications	このデータを使用した論文やレポートについての書誌情報(図書目録)やアクセス情報を入力するものである。
References	その他の参考文献をここでは提示する。

それぞれの資料に対して以下の 4 つの入力フィールドが用意されている。

- Title : 関連資料のタイトルを入力する。
- Location(URI) : 引用元の場所 (URI) を入力する。
- Description : 関連資料の説明を入力する。
- Note : 備考

図 30 に赤丸で示してある「+」をクリックすると資料を追加することができる。また「+」を再度クリックすればクリックした回数だけ選択した項目の入力が可能である。(後述の Related Publications を参照)

編集したい資料を選択すると図 31 のように Title 等の入力が可能となる。

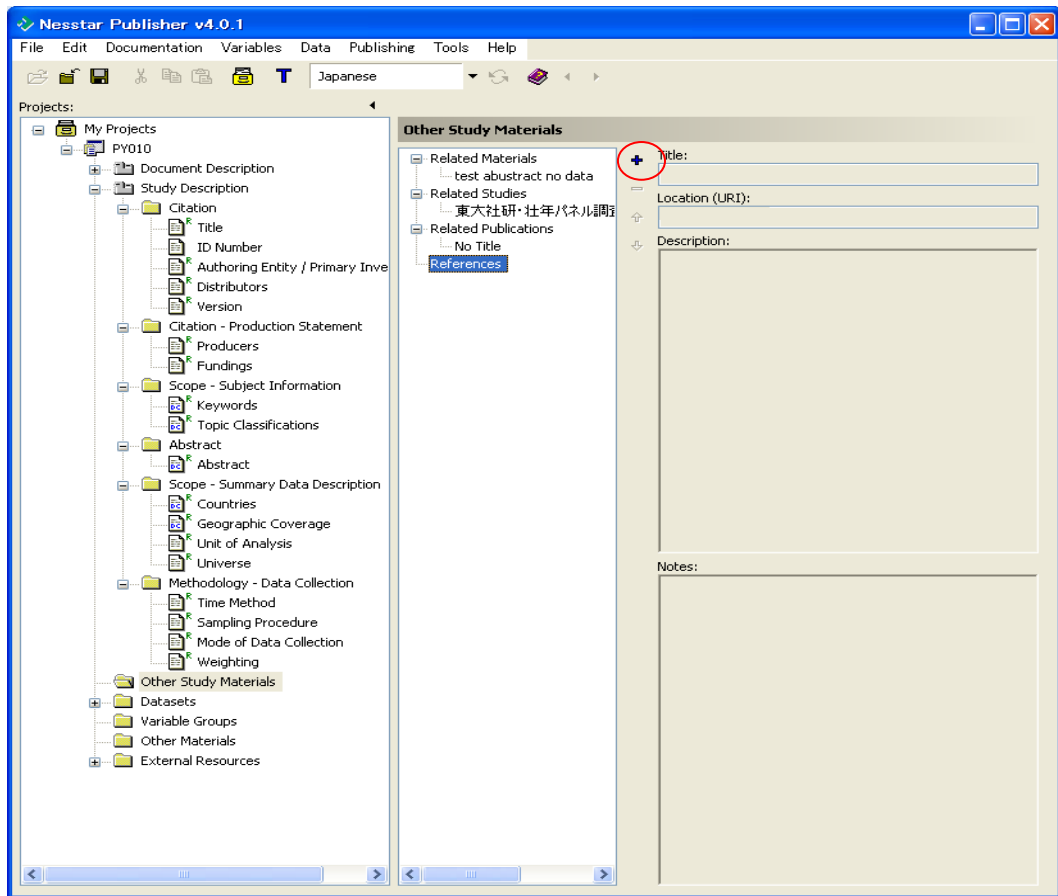


図 30 資料の追加

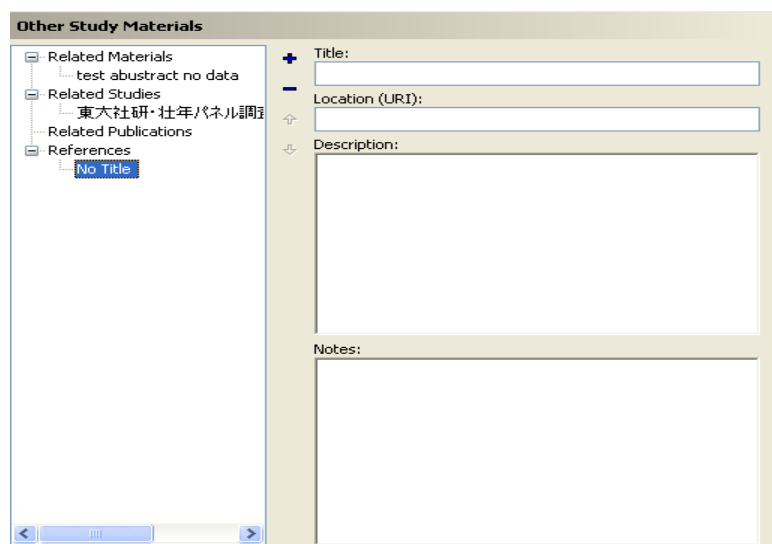


図 31 Title 等の入力

④ DataSets

DataSets は、データセットに含まれる変数やその元となる質問項目、集計情報等データセットに関する情報を記述するセクションである。このセクションは、Key variables & Relations, Variables, Data Entry, Cell Notes, Cube setups の5つの部分から構成される。以下では各部分の設定画面について解説する。

・ Key variables & Relations

Datasets セクションでは複数のデータセットを扱うことができる、この画面では複数のデータセットの関係性やマージ（名寄せ）する際のキーとなる変数の情報を残すことができる。

複数のデータセットがある場合、Relations テキストボックスの右側にある+をクリックすることで図 32 のように他のデータセットを選択する画面が表示され、関連のあるデータセットを追加することができる。

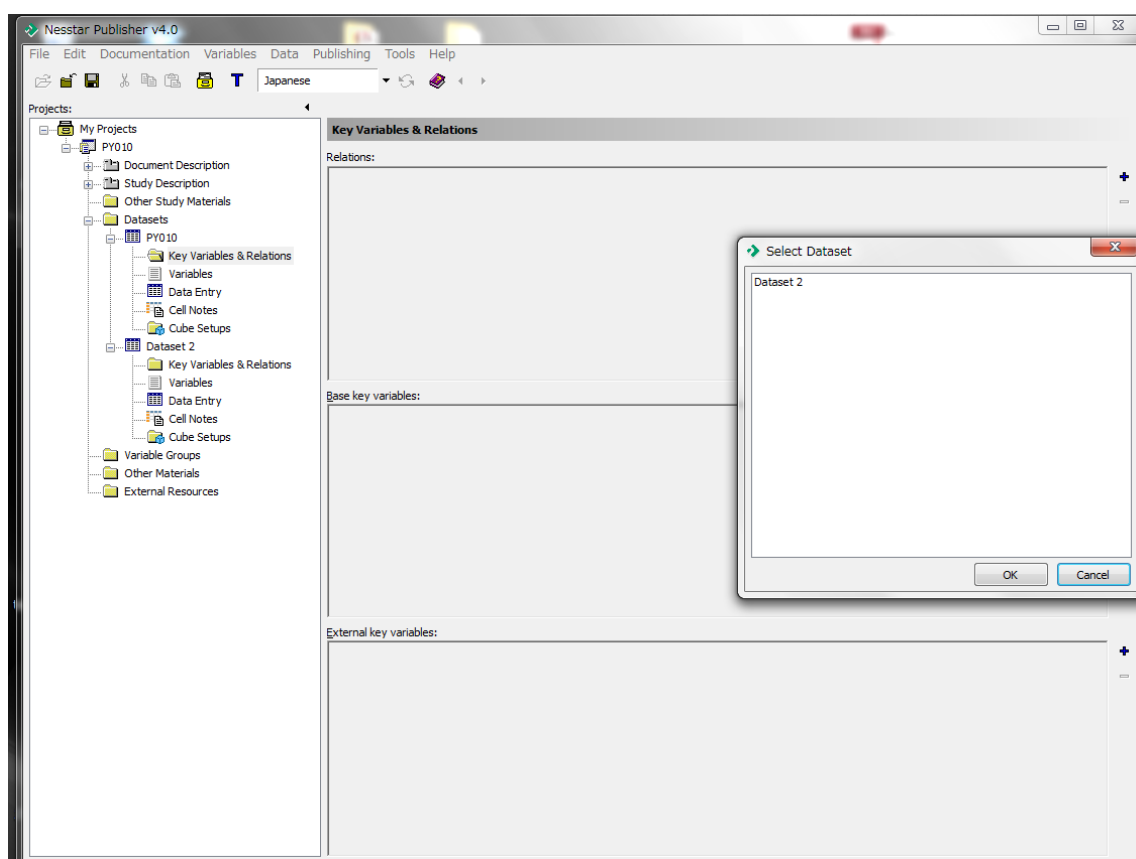


図 32 関連のあるデータセットの追加

Base key variables にはデータセットでレコードを一意に識別するための変数（個人 ID など）を指定する。Relations と同様、右側の+をクリックすると図 337 のようにデータセッ

トに含まれるすべての変数が表示され、その中からキー変数を選択すれば指定できる。

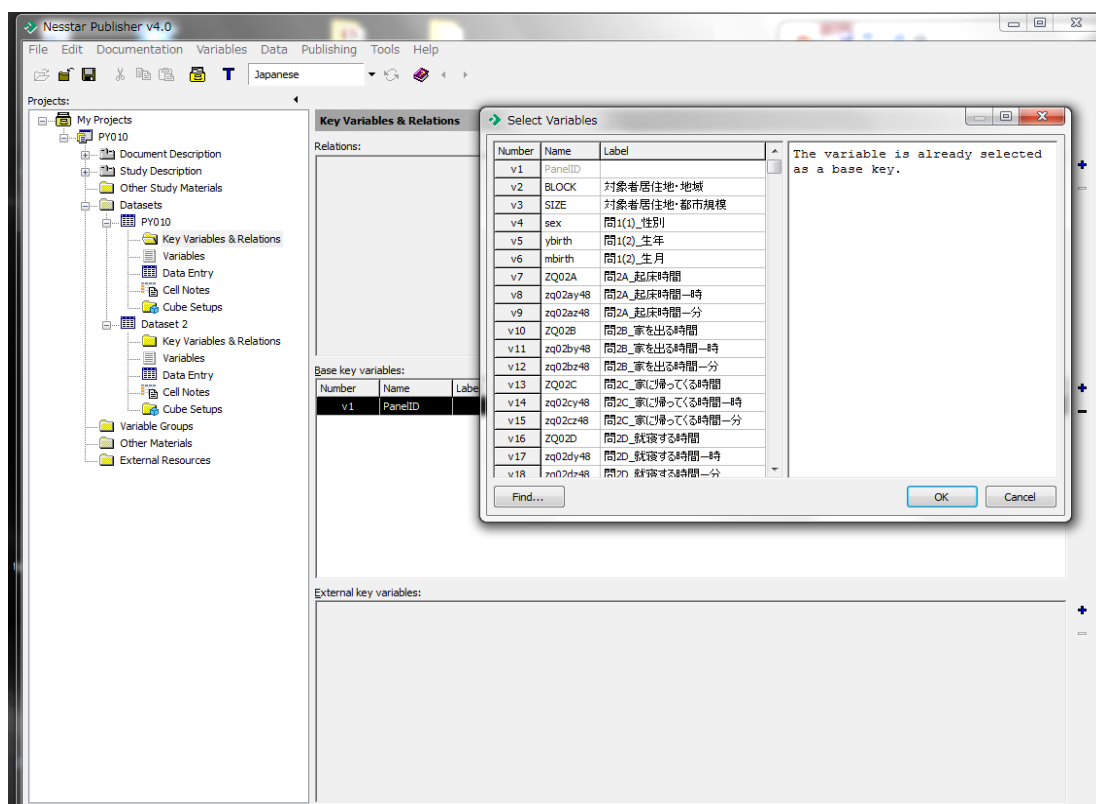


図 33 キー変数の選択画面

External key variables では他のデータセットとマージする際のキーとなる変数を指定する。Base key variables と同様、右側の+をクリックすれば追加することができる。

• Variables

ここではデータセットに含まれる変数に関する情報を記録することができる。

本稿のようにSPSS等の統計パッケージのデータファイルからデータセットをインポートした場合、元のデータファイルに記録されていた変数名や変数ラベル、変数の型などは図34のように既に入力された状態になっている。

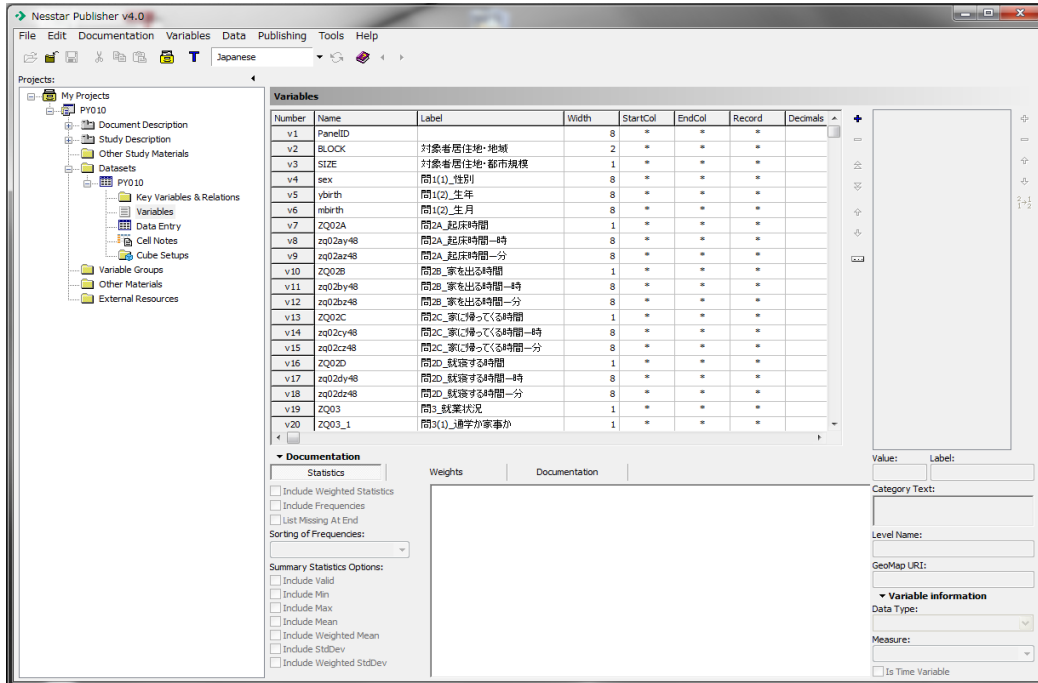


図 34 データファイルをインポートした場合の変数ビュー

図 35 のように Publisher の Variables タブには、変数の一覧とそれぞれ変数の情報が表示される。変数情報には、度数、データタイプ、ウェイト、説明、欠損値、質問、テキストなどが含まれており、Variables 部 (図 35 の A) と Category Hierarchy 部 (図 35 の B) と Documentation 部 (図 35 の C) の三つの部分で構成されている。

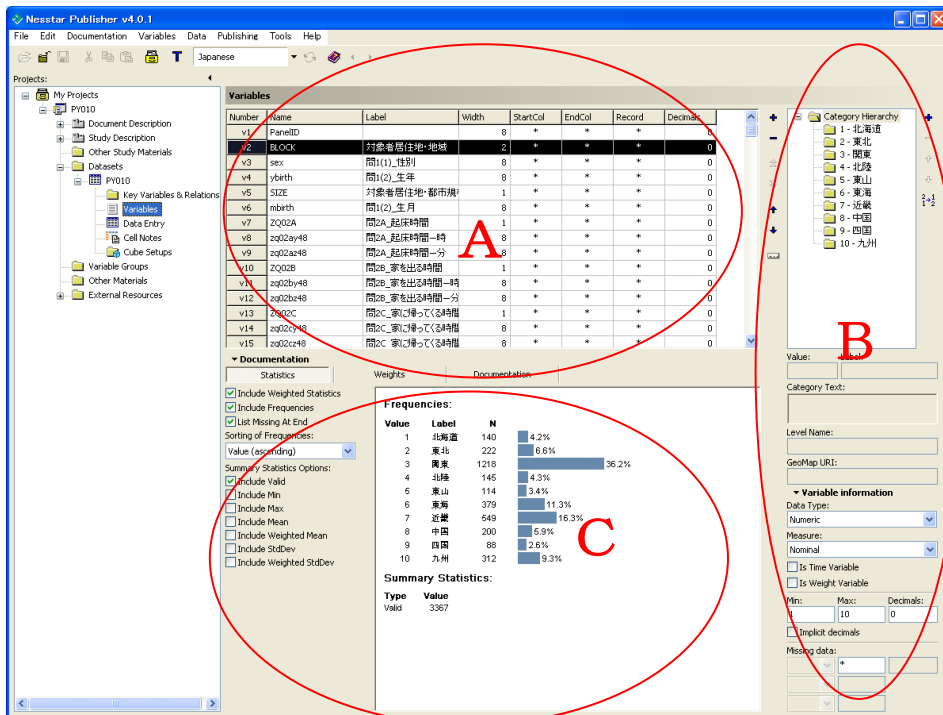



図 35 Variables タブ画面

Variables 部では、変数名(Name)、変数ラベル(Label)、変数の桁数(Width)等の情報が表示される。表示する項目は右側の  ボタンを押して表示されるメニューで指定することができる。

Category Hierarchy 部では、各変数のうちカテゴリカル変数については図 36 のように階層的にカテゴリが表示される。「+」を選択するとカテゴリを追加することができ、追加したカテゴリの詳細情報を下段の Value, Label など各項目に必要なに応じて入力することができる。

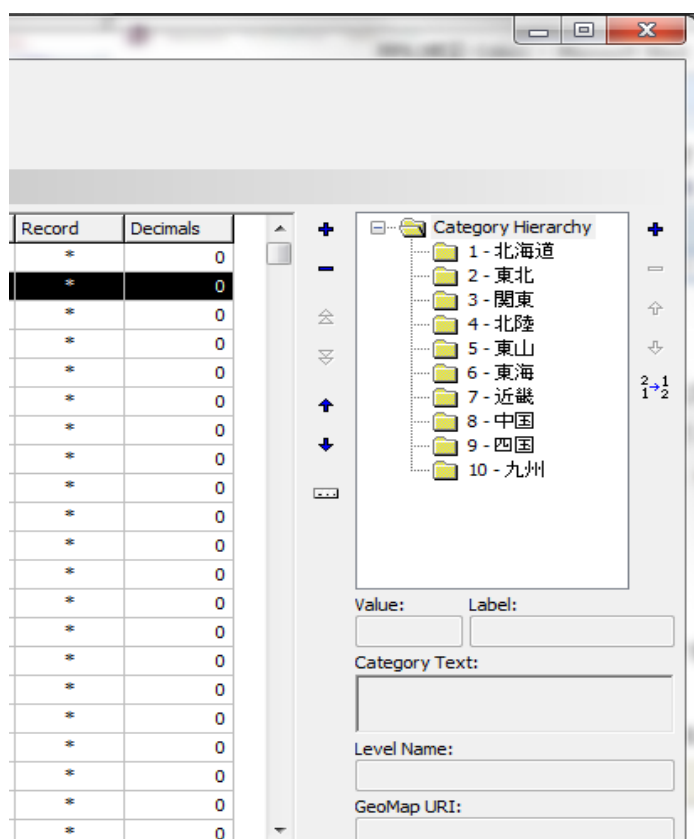


図 36 カテゴリ情報の編集

Documentation 部では Statistics, Weights, Documentation の三つのタブで編集を行う。

Statistics タブでは、図 37 のように各変数の集計項目を設定し表示することができる。集計項目に全てにチェックを入れると図 38 のように表示される。

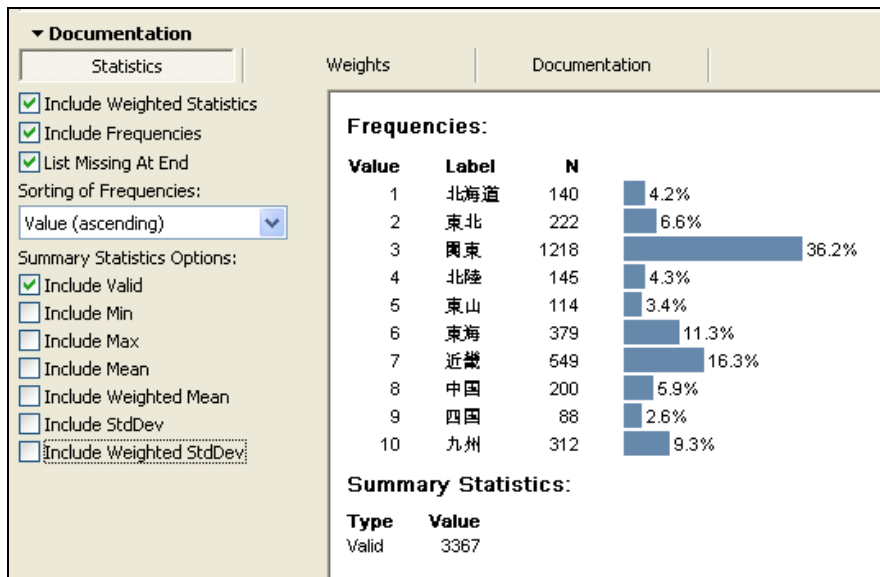


図 37 集計項目の選択

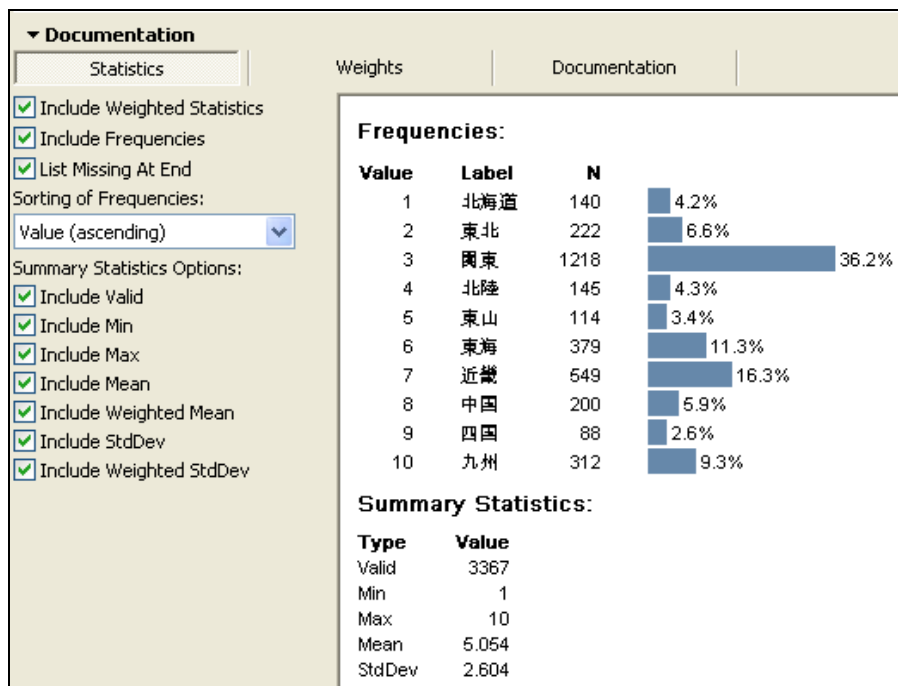


図 38 集計項目の表示

Weights タブでは、重み付けをする変数を設定できる。その結果は Statistics タブに反映される。

図 39 の赤丸で示されている「+」をクリックすると図 40 のように weight 変数の選択画面が表示され、weight 変数を設定することができる。

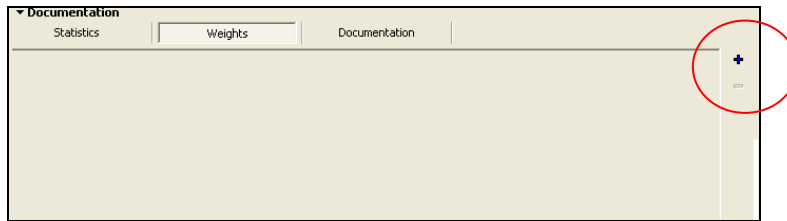


図 39 Weights の「+」の位置

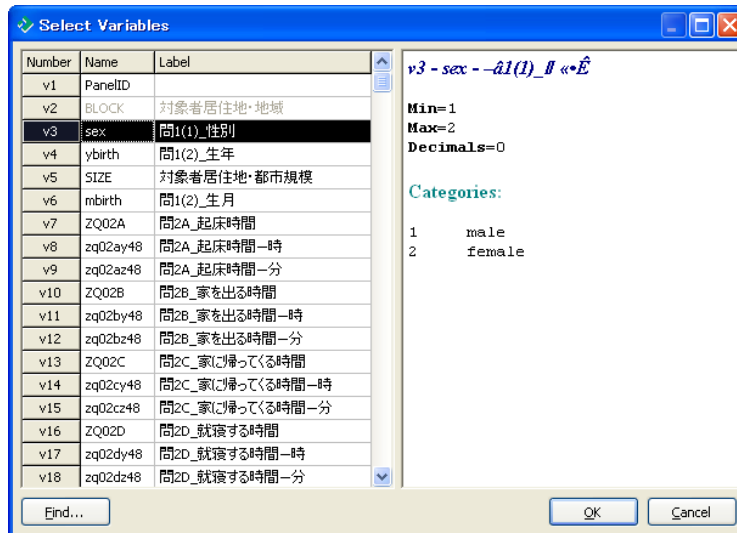


図 40 weight 変数の設定

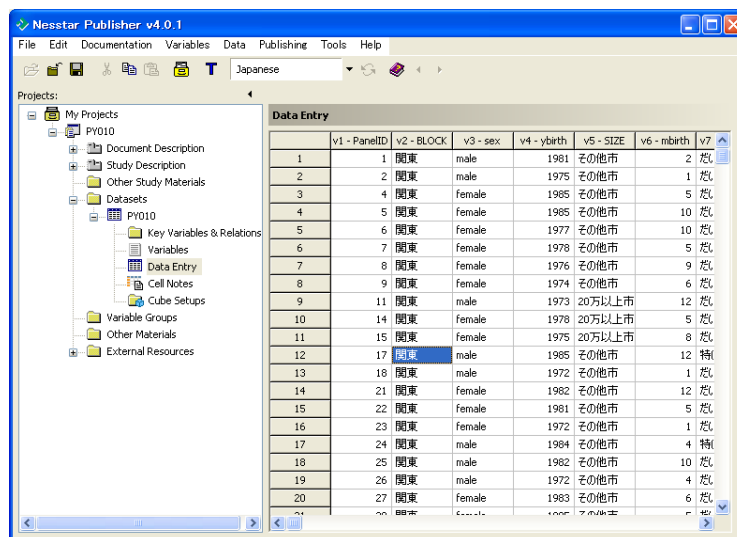
Documentation タブは DDI の Variable Description に基づいた項目であり、変数レベルの質問および付加情報を入力する。入力項目の内容は、表 12 の通りである。

表 12 Datasets の Variables の Documentation タブの入力項目

項目グループ	入力項目	内容
Question	Pre-Question Text	質問をする前に置く文章等を入力する項目。
	Literal Question	実際に行った質問について入力する。
	Post-Question Text	質問のあとに置く文章等について記述する。
	Interviewer instructions	インタビューを行う際にインタビュー実施者に行った教育内容などを入力する。例えば、なかなか質問に答えてくれない方への対処方法などである。
Description	Variable Text	変数名や変数ラベル以外の付加的情報を入力項目である。
	Concepts	変数が属する主なテーマについて入力する項目である。これは、keywords とトピックと同様だが、あくまで変数レベルについて入力する。

• Data Entry

本項目は実際のデータを表示する。ここでは各セルを選択してポップアップを表示すると読み取り専用の設定・解除、コピーなどができる。読み取り専用を解除すれば、データの編集も可能である。



	v1 - PanelID	v2 - BLOCK	v3 - sex	v4 - ybirth	v5 - SIZE	v6 - mbirth	v7
1	1	関東	male	1981	その他市	2	枚
2	2	関東	male	1975	その他市	1	枚
3	4	関東	female	1985	その他市	5	枚
4	5	関東	female	1985	その他市	10	枚
5	6	関東	female	1977	その他市	10	枚
6	7	関東	female	1978	その他市	5	枚
7	8	関東	female	1976	その他市	9	枚
8	9	関東	female	1974	その他市	6	枚
9	11	関東	male	1973	20万以上市	12	枚
10	14	関東	female	1978	20万以上市	5	枚
11	15	関東	female	1975	20万以上市	8	枚
12	17	関東	male	1985	その他市	12	特
13	18	関東	male	1972	その他市	1	枚
14	21	関東	female	1982	その他市	12	枚
15	22	関東	female	1981	その他市	5	枚
16	23	関東	female	1972	その他市	1	枚
17	24	関東	male	1984	その他市	4	特
18	25	関東	male	1982	その他市	10	枚
19	26	関東	male	1972	その他市	4	枚
20	27	関東	female	1983	その他市	6	枚
21	28	関東	female	1983	その他市	7	枚

図 41 実際のデータの表示

• Cell Notes

本項目は、多元表データ（Cube）で設定されたデータについてのみ適用される。設定された多元表の各セルやグループセルについてメタ情報を追加することができる。なお、データをインポートしても Cell Note は自動的に作成されないため、入力する必要がある。

• Cube Setups

Nesstar Publisher version 4 になって、個票データを集計した多元表データ（Cube）を追加できるようになった。本項目では、多元表データの設定ができる。なお、設定を行うには、Cell notes を先に設定する必要がある。

⑤ Variable Groups

Variable Groups は、各変数をグループ化する設定を行うセクションである。また、本項目で各変数を特定のグループに設定しない場合、Nesstar Webview と連携する際に、各変数が「…」として表示される。

図 42 の「+」ボタンをクリックすると新規グループのフォルダが追加されるので、グループの情報を Type 等に入力する。

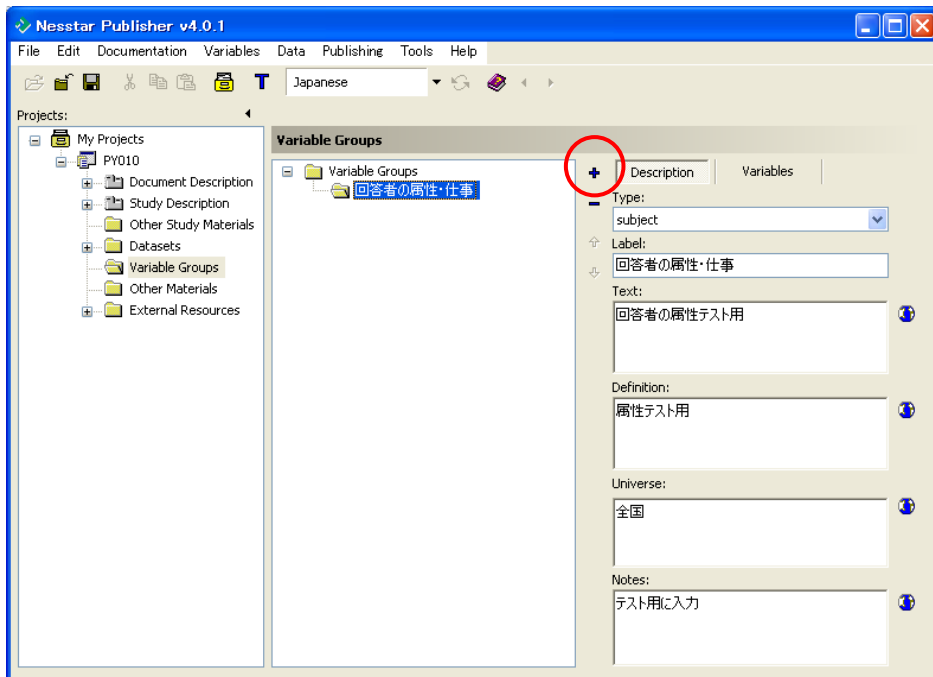


図 42 新規グループの追加

次に Variables タブを選択して、グループに所属させる変数を設定する。

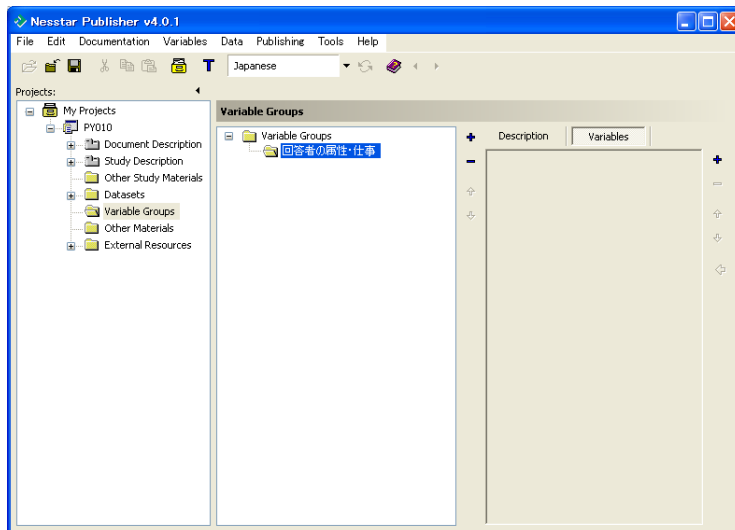


図 43 グループに所属させる変数の設定

「+」をクリックすると図 44 のように変数選択画面が表示されるので、グループに所属させる変数を選択する。ただし、Find 機能は英数のみの対応である。

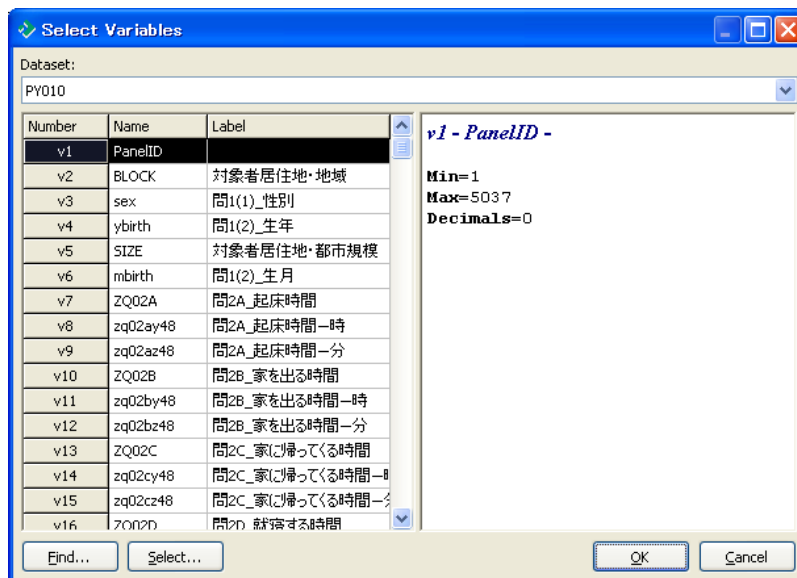


図 44 変数選択画面

「OK」をクリックすると、図 45 のように選択した変数が追加される。

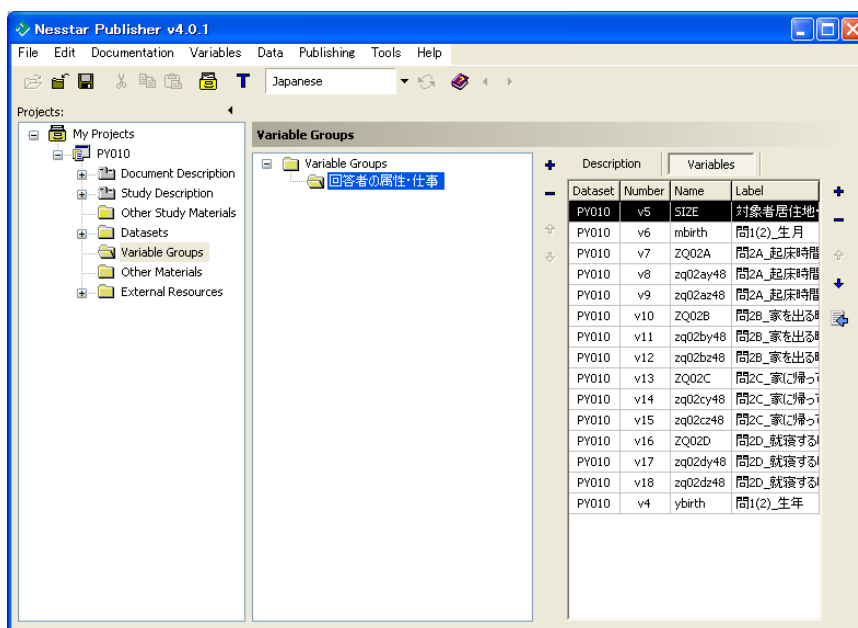


図 45 追加された変数

⑥ Other Materials

Other Materials は、DDI に準拠したファイルに関連情報を保存しておくために利用する。次のようなものが含まれる。

質問紙（調査票）、コーディング情報、ソフトウェアの詳細情報、ユーザマニュアル、そ

の他のガイド、プログラムのサンプル、インタビュー/プロジェクト手順、マップ、データベーススキーマ、データ辞書、誤値の情報、インタビューのスケジュール、度数ファイルなど。

⑦ External Resource

External Resource は、本調査に関連する外部資源についての情報を入力するセクションである。Dublin Core や e-GMS のメタ情報で記述された Word ファイルや PDF ファイル、地図などを扱う。

以上でメタ情報の追加は完了である。

(3) Server による Webview の設定手順と Publisher による公開手順

Server には Webview の分析や表示の設定機能や、ユーザー管理機能が実装されている。図 46 は、Webview で変数の加工を行えるようにするため、「Web View settings」タブの「Allow calculations and compute.」にチェックしている画面である。Server の設定の詳細については、Nesstar Server Administrator Guide Version3.5 を参照されたい。

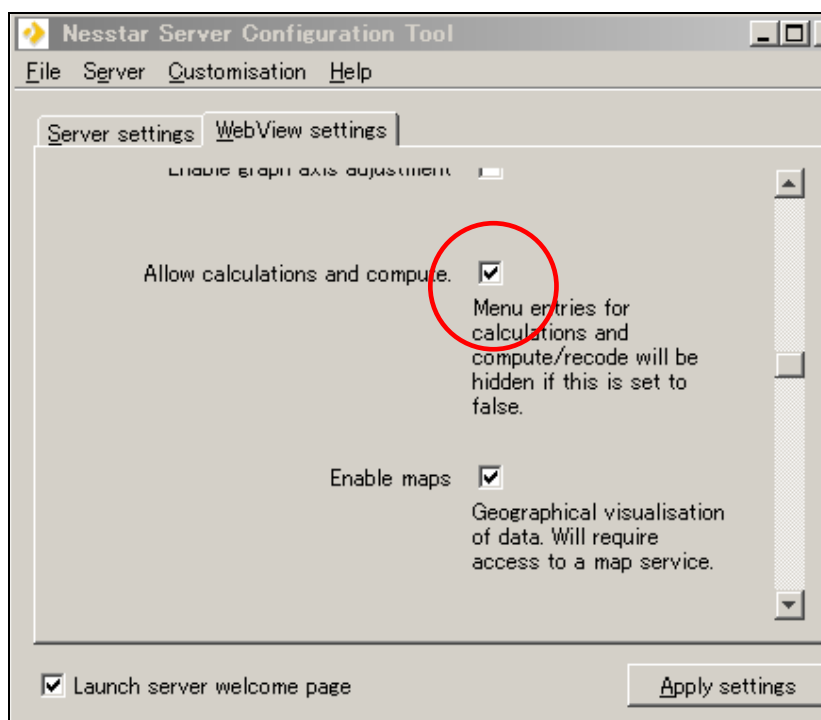


図 46 Webview の設定画面

Publisher による公開手順は以下の通りである。

① Publisher での公開用サーバの登録

Publisher に公開用サーバを登録する。図 47 の通りに、「Publishing」→「Add server ...」を選択する。図 19 で設定したサーバのアドレスとユーザー名、パスワードを入力したのが、図 48 である。「OK」をクリックすると、Publisher への公開用サーバ登録は完了である。

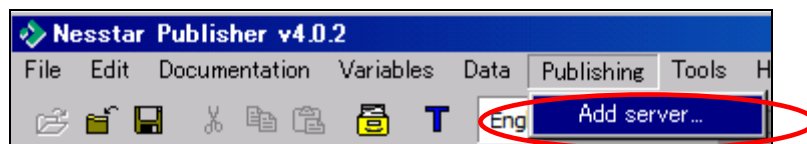


図 47 「Add server」の選択

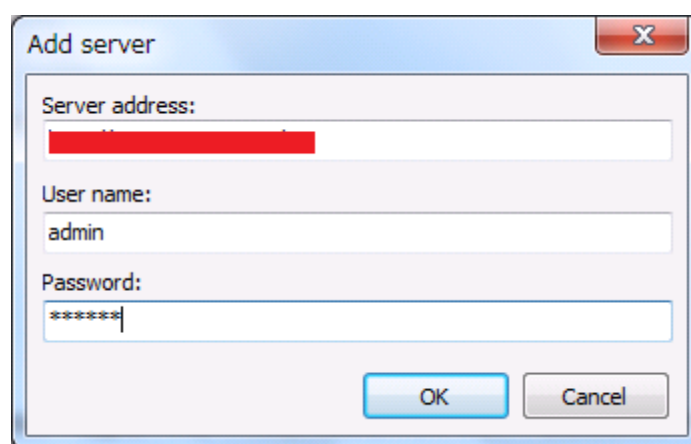


図 48 Publisher に登録する公開用サーバ設定画面

② Web でのデータ公開

インポートしたデータを、Web 上で公開する。図 49 の通りに、「Publishing」→「サーバ名」→「Study」→「Publish data and metadata...」を選択する。その結果表示された公開設定画面が図 50 である。サーバ名にチェックをつけて「Publish」をクリックすると、データが公開される。Server を再起動すると、公開したデータを Webview で閲覧することができる。

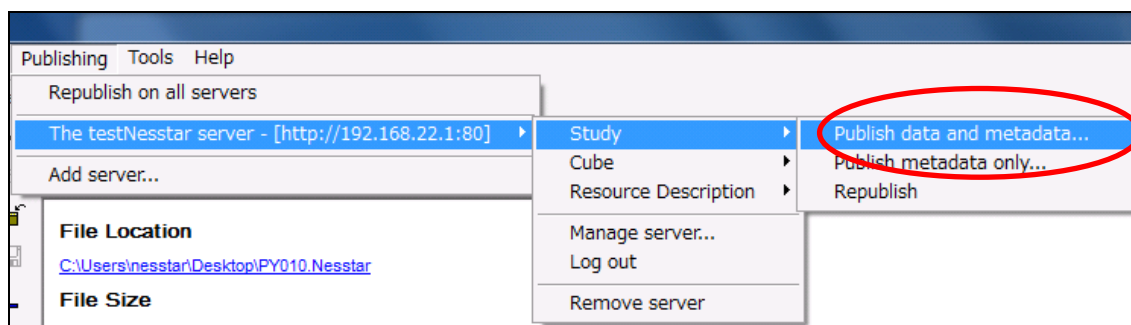


図 49 「Publish data and metadata...」の選択

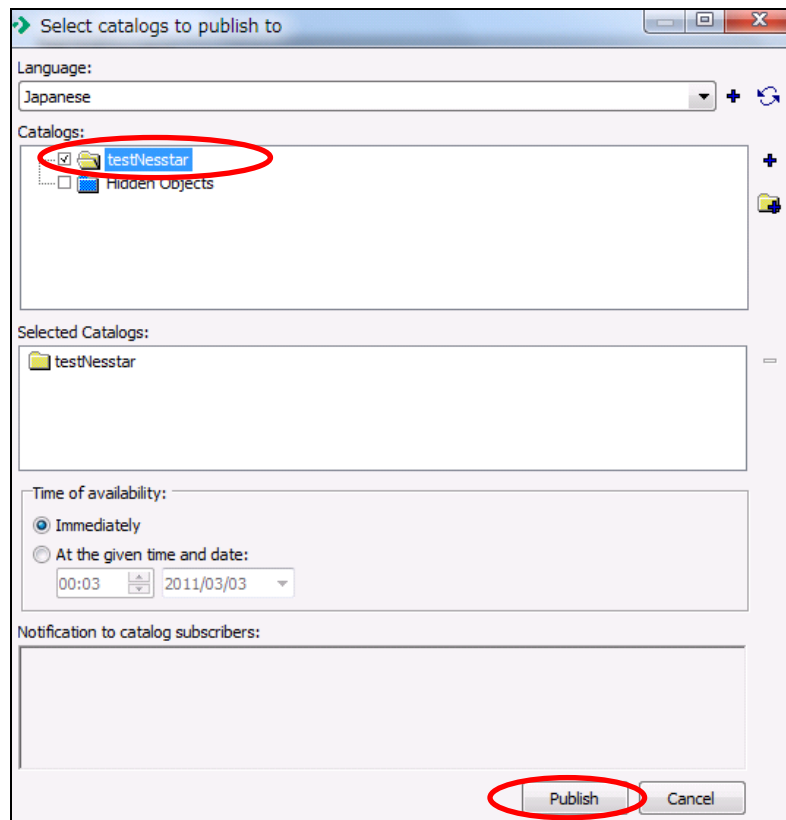


図 50 データの公開設定画面

データの公開作業は以上で完了である。

1.3 Webview の分析手順

Webview では、変数の検索、記述統計、クロス分析、各種グラフの作成、変数の加工、相関係数、回帰分析などのアウトプットを出すことができる。以下では、サンプルデータを用い、各種機能の概略を説明する。

図 51 は Webview の標準的な画面構成である。画面左側にデータ一覧が表示され、そのなかにメタデータ（調査の概要や調査票など）と変数の一覧が表示される。見たい項目をクリックすることで画面右側にアウトプットが表示されるかたちとなっている。Webview では変数名をクリックすると記述統計と棒グラフが自動的に表示される設定となっているため、画面右側にその結果が表示されている。

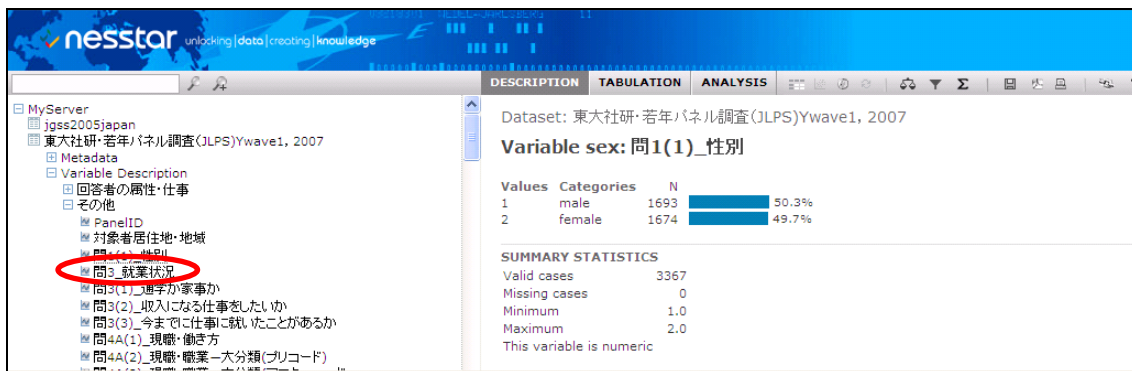


図 51 Webview での標準的な画面構成

(1) 変数の検索

Webview では、複数のデータセットのなかから、必要な変数を検索することができる。詳細検索では、調査の概要や変数の説明など、検索範囲を指定したうえで必要な変数を探ることができる。また、検索対象も、調査の概要、変数、図表というかたちで絞っていくことができる。調査の概要をとっても、対象地域やサンプリング過程など、より詳細な検索の指定が可能である。

たとえば、図 52 は、「Advanced Search」アイコンをクリックして詳細検索画面を別ブラウザで起動し、変数の説明のなかに「年齢」という言葉を含む変数を検索している画面である。その結果の画面が図 53 である。サンプルデータからは、「問 51_結婚時の満年齢」など 3 つの変数が検索されている。

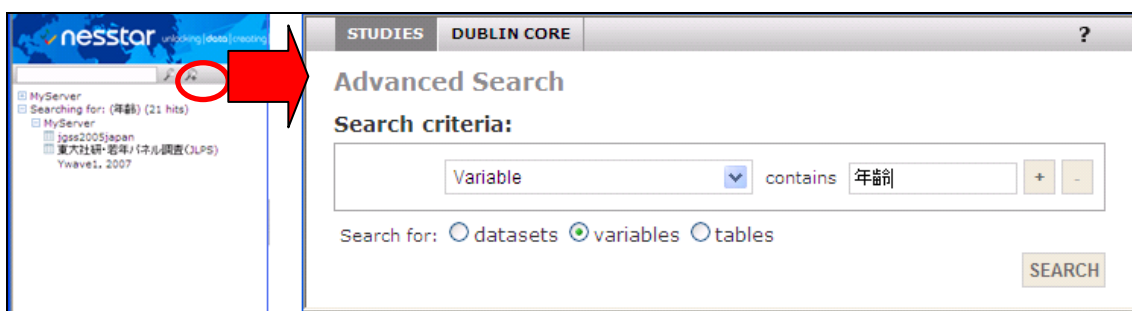


図 52 変数の詳細検索画面

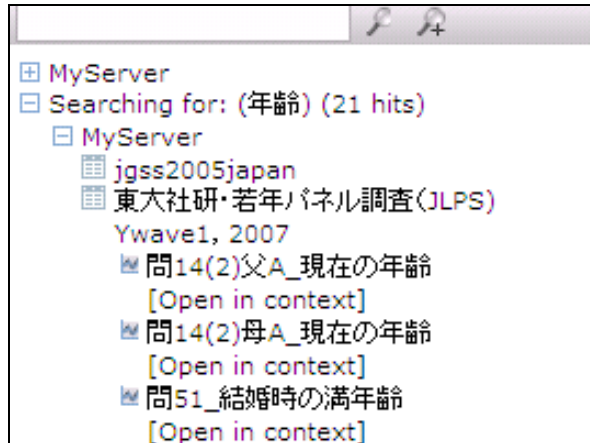


図 53 検索結果表示画面

(2) 記述統計

Webview では、変数をクリックすると自動的に記述統計が表示されるようになっている。図 54 は、画面左の「問 51_結婚時の満年齢」をクリックし、記述統計を出した画面である。連続変数である結婚時の満年齢については、有効なケースの数、欠損値の数、最小値、最大値、標準偏差、平均値がアウトプットされている。

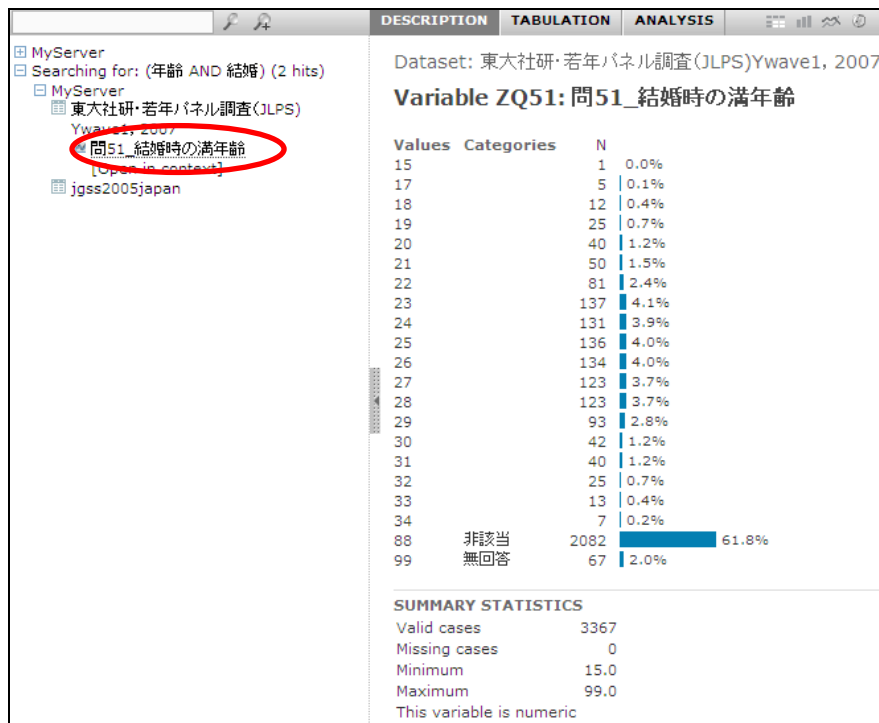


図 54 記述統計の出力画面

(3) クロス表と棒グラフ

Webview では、画面左の変数一覧から独立変数と従属変数を指定することでクロス表をアウトプットできる。図 55 は、「問 1__性別」を列に入れたのちに、「問 30D__現在の満足度 - 生活全般」をクリックして「Add to row」を選択している画面である。その結果が図 56 である。なお、作成されたクロス表上部にある変数コンボボックスをクリックすると、変数の削除や行と列の入れ替えなどが行える。また、変数コンボボックスの下にある percentage コンボボックスをクリックすると、percentage を行・列・行列のいずれで表示するかを選択できる。

Dataset: 東大社研・若年パネル調査(JLPS)Ywave1, 2007

問1(1)_性別: Categories

問1(1)_性別	male	female	Total
Code	1	2	
Frequency	1,693	1,674	3,367
% of all	50.3	49.7	100.0
% of valid	50.3	49.7	100.0

図 55 クロス表作成画面

Dataset: 東大社研・若年パネル調査(JLPS)Ywave1, 2007

問30D_現在の満足度-生活全般: Categories 問1(1)_性別: Categories

Type: Column percentage

問1(1)_性別	male	female	Total
問30D_現在の満足度-生活全般			
満足している	15.1	18.6	16.9
どちらかといえば満足している	37.6	45.2	41.4
どちらともいえない	28.6	23.5	26.1
どちらかといえば不満である	11.8	8.7	10.3
不満である	5.4	3.5	4.5
無回答	1.4	0.5	1.0
Total	100.0	100.0	100.0
N=	1,693	1,674	3,367

図 56 出力されたクロス表

さらに、クロス表をアウトプットした状態でグラフを選択すると棒グラフや円グラフがアウトプットできる。図 57 では表示させるグラフの種類を指定しており、その結果が図 58 である。

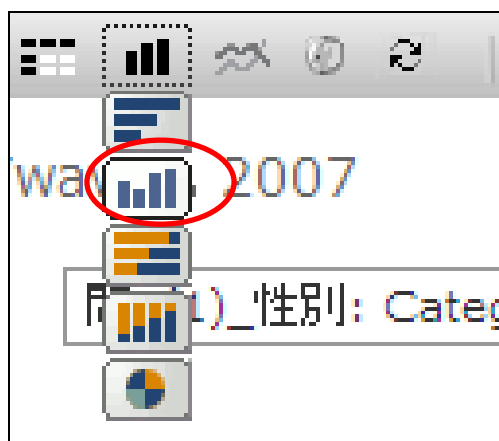


図 57 グラフ選択バー

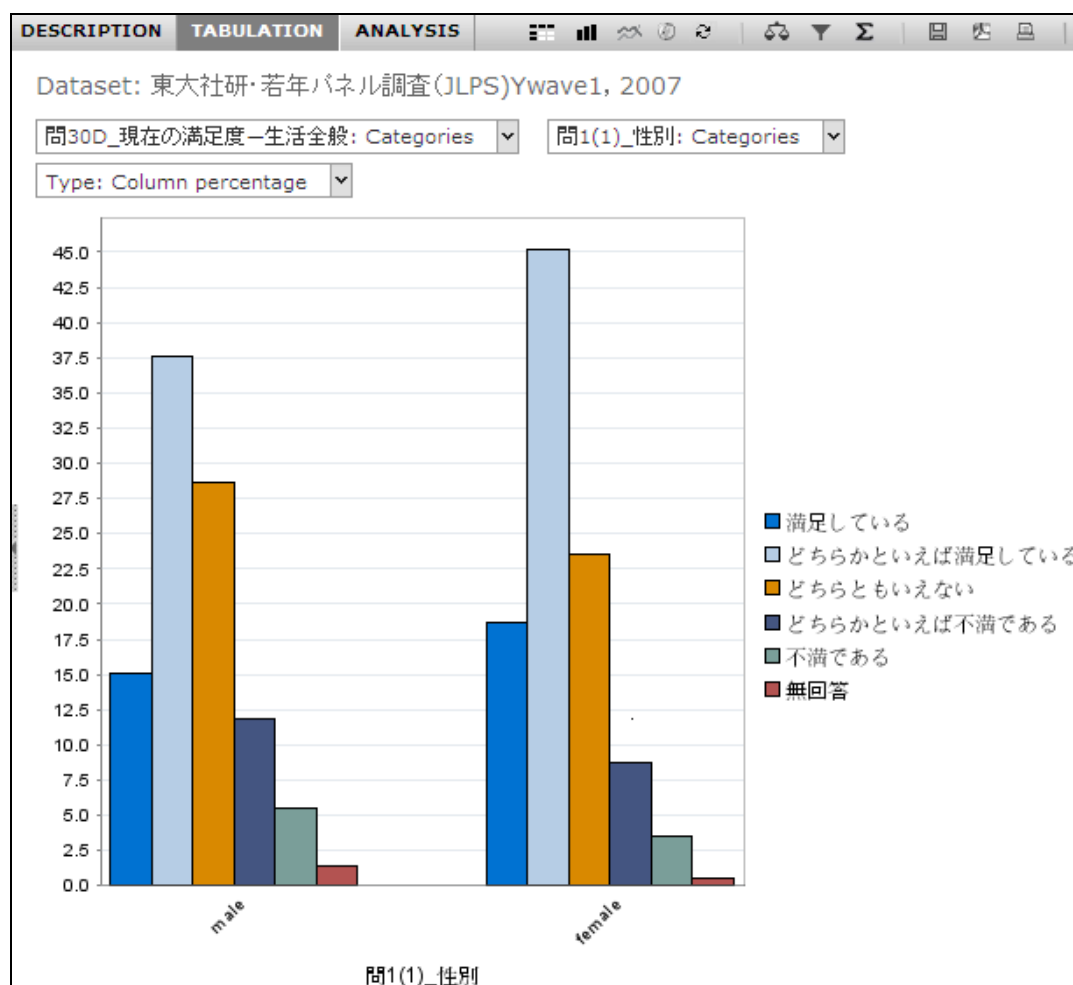


図 58 出力された棒グラフ

また、層に三つ目の変数を投入することで、三重クロス表も作成することができる。
 図 59 では「問 12_現在の暮らしむき」を三つ目の変数で選択している。 図 60 がその結果である。

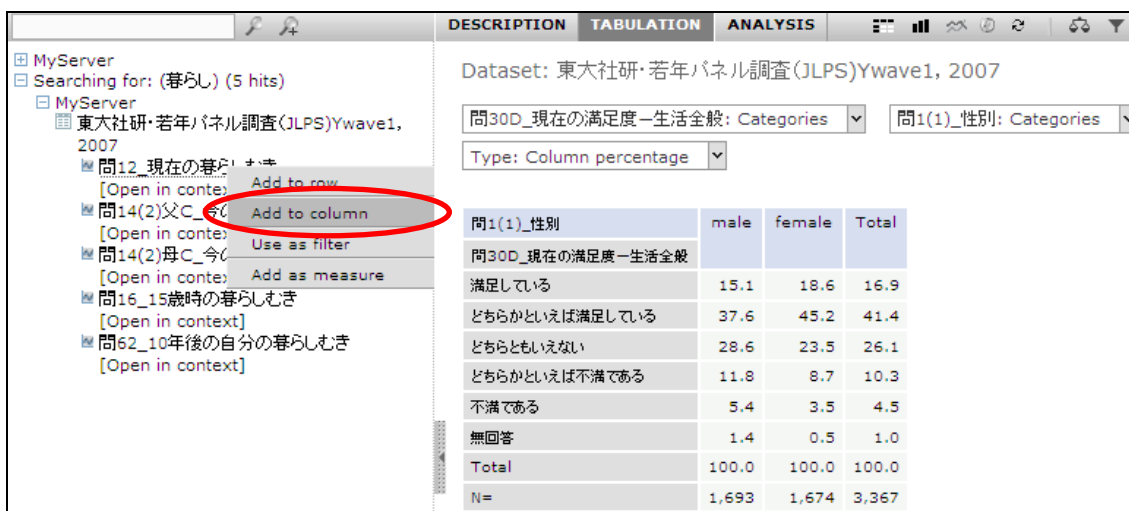


図 59 三つ目の変数の追加画面

問1(1)_性別	male						female						Total
	豊か	やや豊か	ふつう	やや貧しい	貧しい	無回答	豊か	やや豊か	ふつう	やや貧しい	貧しい	無回答	
問12_現在の暮らしむき													
問30D_現在の満足度-生活全般													
満足している	40.0	21.5	14.5	5.7	4.4	0.0	41.5	21.5	18.4	4.3	0.0	57.1	16.9
どちらかといえば満足している	40.0	48.6	40.9	22.1	12.1	20.0	41.5	47.3	49.8	32.1	13.0	14.3	41.4
どちらともいえない	12.0	19.0	30.4	36.9	29.7	50.0	13.0	22.2	22.9	33.0	28.3	21.4	26.1
どちらかといえば不満である	4.0	8.1	9.8	23.8	22.0	0.0	2.4	6.2	6.5	22.5	30.4	0.0	10.3
不満である	2.0	2.4	3.2	9.4	30.8	10.0	1.6	1.8	2.2	7.2	28.3	7.1	4.5
無回答	2.0	0.4	1.3	2.0	1.1	20.0	0.0	1.1	0.3	1.0	0.0	0.0	1.0
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
N=	100	247	1,001	244	91	10	123	275	1,007	209	46	14	3,367

図 60 出力された三重クロス表

(4) 変数の加工

SSJDA で運用している SPSS によるリモート集計システムでは、欠損値の指定ができないなど変数の加工に課題が残されていた。その点、Nesstar では、SPSS と同レベルの変数の加工が可能である。足し算・引き算・掛け算・割り算・割合・平均・増加率などの計算の他に値の再割り当てもできる。図 61 の通りに「compute」をクリックして、「値の再割り当て (recode)」を選択すると変数を加工する画面となる。図 62 では、「問 4A (1) 現

職・働き方」を選択し加工を行っている。画面上部には「Compute」パネルがあり、SPSSの値の再割り当てと同様の方法で変数の加工ができる。「Old value」パネルと「New value」パネルに変更前と変更後の値を入力している。さらに「Add」をクリックすると、「Added」パネルには再割り当てした値が追加される。

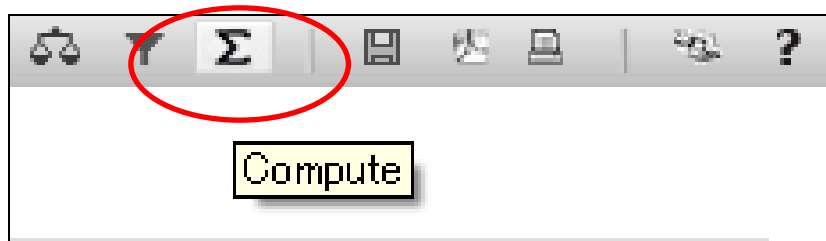


図 61 「Compute」の選択

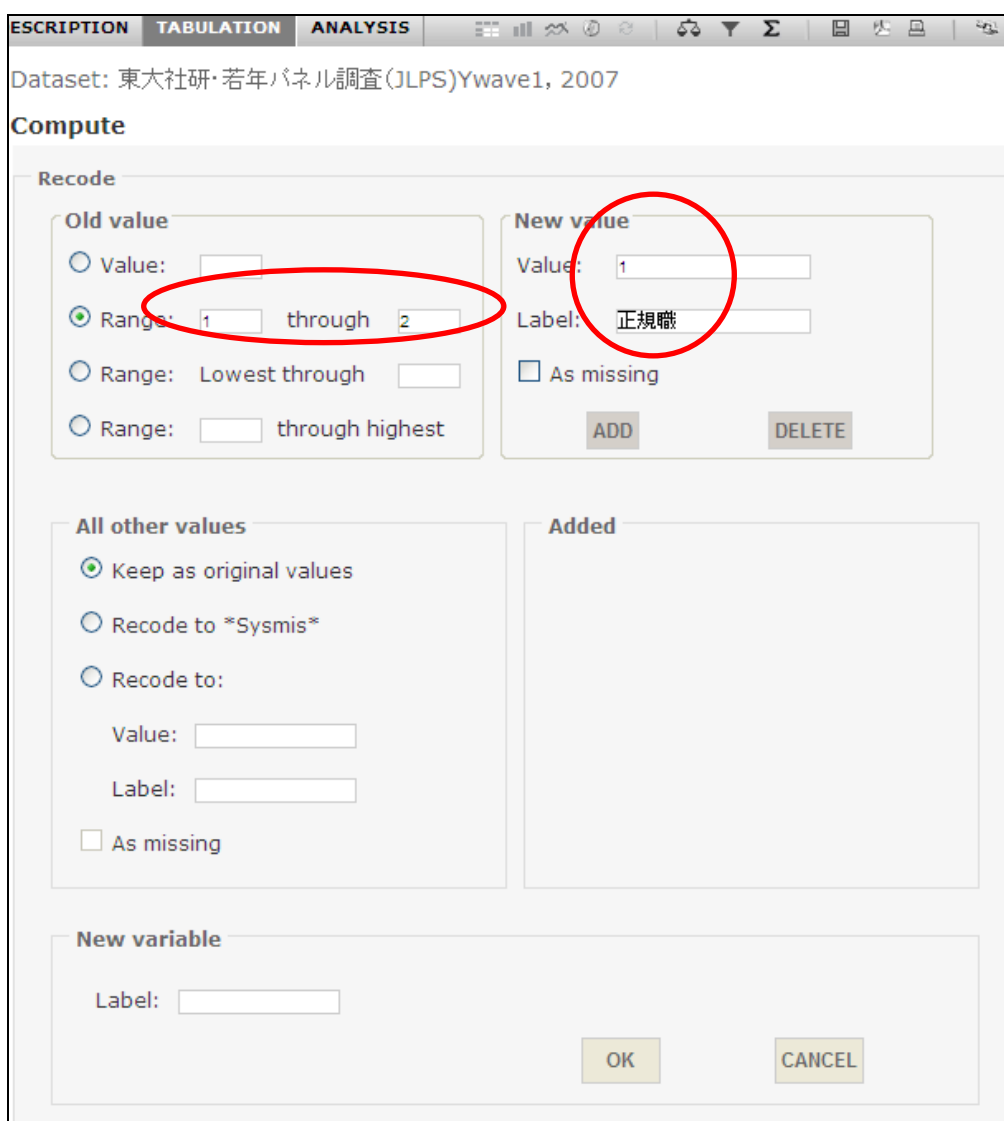


図 62 再割り当ての条件の入力

また、「as missing」をクリックしておけば、欠損値として処理することができる。図 63 は、これまでの値が「99」だったものをそのまま「99」として指定し、かつ、欠損値として指定した画面である。

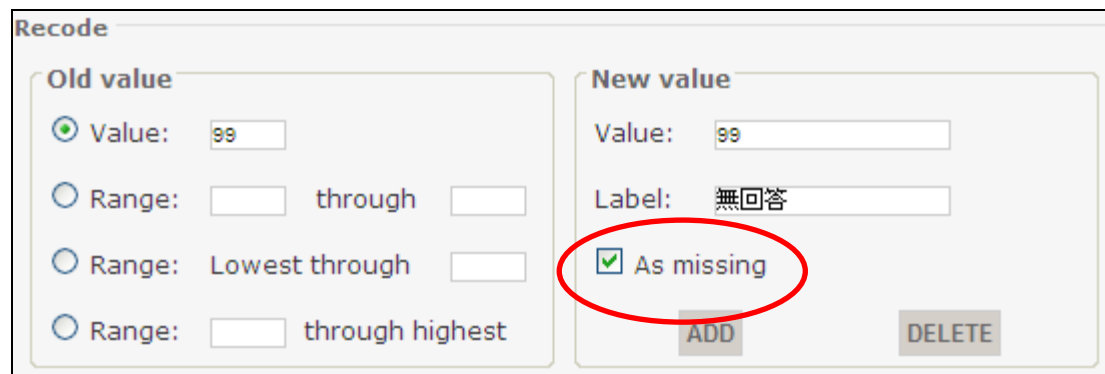


図 63 欠損値の指定

図 64 は、指定以外の値を欠損値として処理するように設定し、新しい変数のラベルに「正規職ダミー」を入力して、変数を作成している画面である。「OK」をクリックすると変数が作成される。その結果の画面が図 65 であり、作成された変数は「User defined variables」に格納される。画面右側の「As missing」の項目では、欠損値指定を確認することができる。

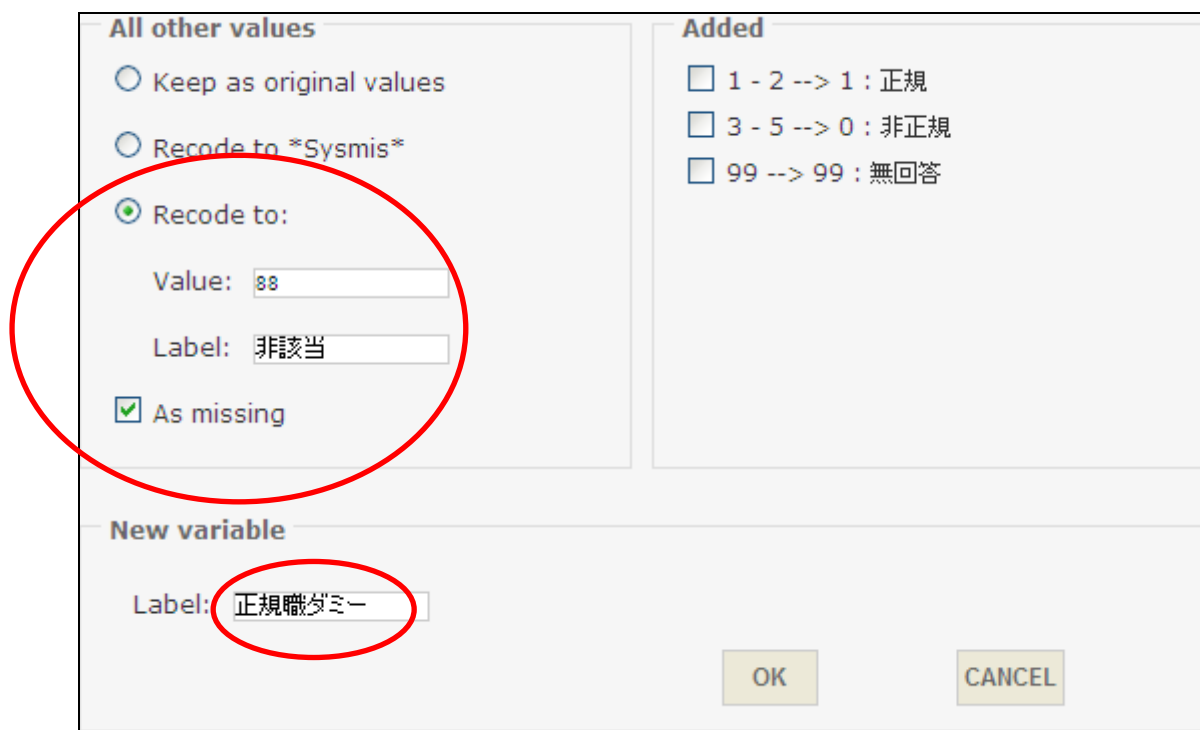


図 64 指定外の値の一括設定と新しい変数のラベル作成画面



図 65 新しく作成した変数の記述統計

(5) 相関係数と回帰分析

Webview では、変数を指定することで相関係数と回帰分析の結果をアウトプットできる。図 66 の通りに、「ANALYSIS」タブをクリックしてから「CORRELATION」タブを選択すると相関係数の計算画面になる。相関係数の計算に変数を追加するため、欠損値指定した「本人年収」を選択して「Add to correlation」をクリックしているのが図 67 である。さらに「通勤時間」を追加して相関係数が出力されている画面が、図 68 である。出力された相関係数はピアソンの積率相関係数であり、スピアマンの順位相関係数やケンドールの順位相関係数を選択することはできない。また、変数除去方法をリストワイズで結果を出力しているが、画面上部の「Change to pairwise」をクリックすると、ペアワイズに変更して相関係数を出力できる。

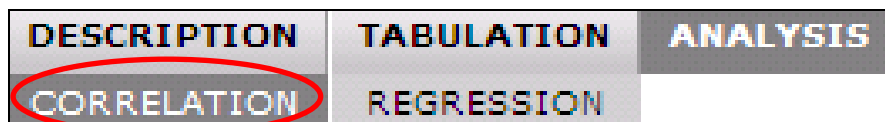


図 66 「CORRELARION」の選択

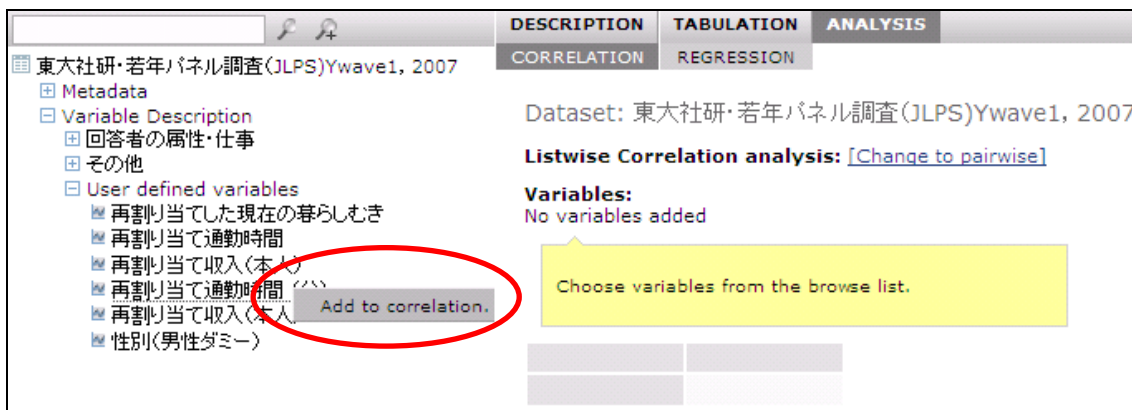


図 67 相関係数への変数追加画面

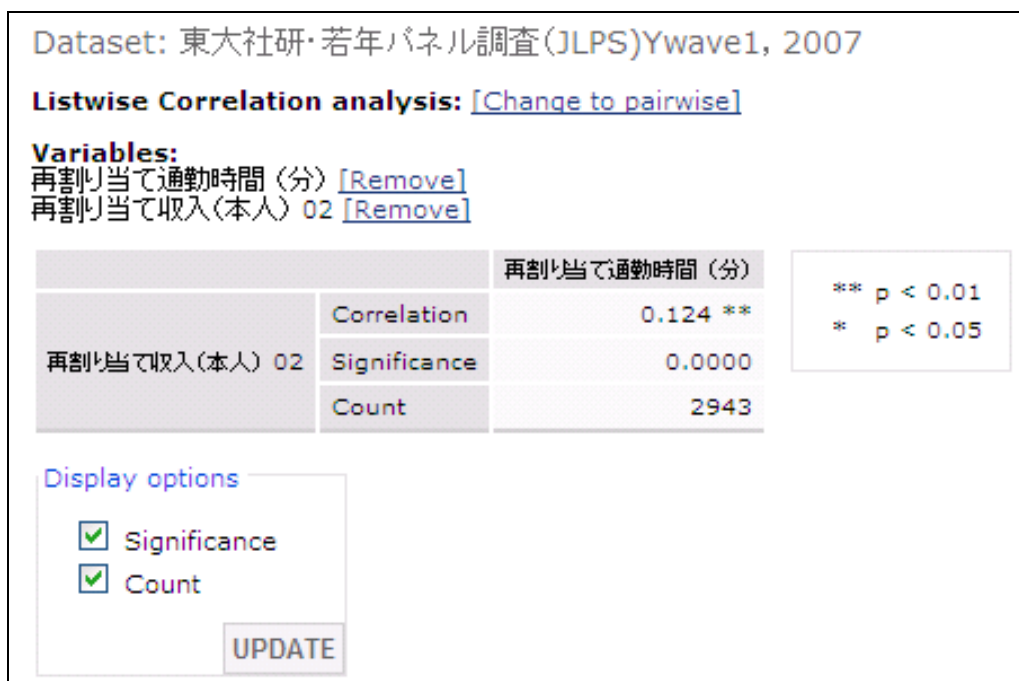


図 68 出力された相関係数

図 69 の通りに、「ANALYSIS」タブをクリックしてから「REGRESSION」を選択すると回帰分析の画面となる。図 70 は、「労働従事時間（1 週間あたり）」を従属変数として追加している画面である。同様に、「男性ダミー」と「正規雇用ダミー」を独立変数として追加し、重回帰分析を行った結果の画面が図 71 である。

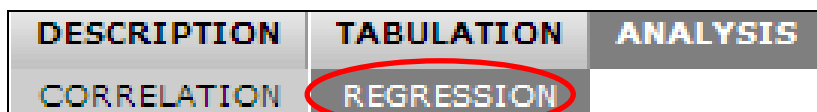


図 69 「REGRESSION」の選択

Dataset: 東大社研・若年パネル調査(JLPS)Ywave1, 2007

Regression analysis:

Dependent variable:
No dependent variable added

Independent variable:
No independent variable added

Add variables to the regression by selecting them from the browse list. Choose a dependent variable, and add one or more independent variables.

	B	SE B	Beta	T	Significance	Tolerance
Intercept	-	-	-	-	-	-
Valid N	-	-	-	-	-	-
Multiple R	-	-	-	-	-	-
Multiple R Squared	-	-	-	-	-	-
Adjusted R Squared	-	-	-	-	-	-
F value	-	-	-	-	-	-
F sign	-	-	-	-	-	-

図 70 従属変数の投入画面

Dataset: 東大社研・若年パネル調査(JLPS)Ywave1, 2007

Regression analysis:

Dependent variable:
従事時間(週あたり) [Remove]

Independent variable:
男性ダミー [Remove]
正規職ダミー [Remove]

	B	SE B	Beta	T	Significance	Tolerance
正規職ダミー	11.23	0.75	0.30	15.02	0.0000	0.92
男性ダミー	7.67	0.67	0.23	11.44	-0.0000	0.92
Intercept	37.54					
Valid N	2,289.00					
Multiple R	0.419					
Multiple R Squared	0.176					
Adjusted R Squared	0.175					
F value	244.00					
F sign	0.0000					

[View correlation matrix](#)

図 71 出力された相関係数

(6) ケースの選択

Webview では、ケースの選択も容易にできる。図 72 の通りに、「Subset」をクリックすると、ケースの選択画面となる。図 73 は、「性別=1」つまり、男性のみを選択する画面である。



図 72 「Subset」の選択

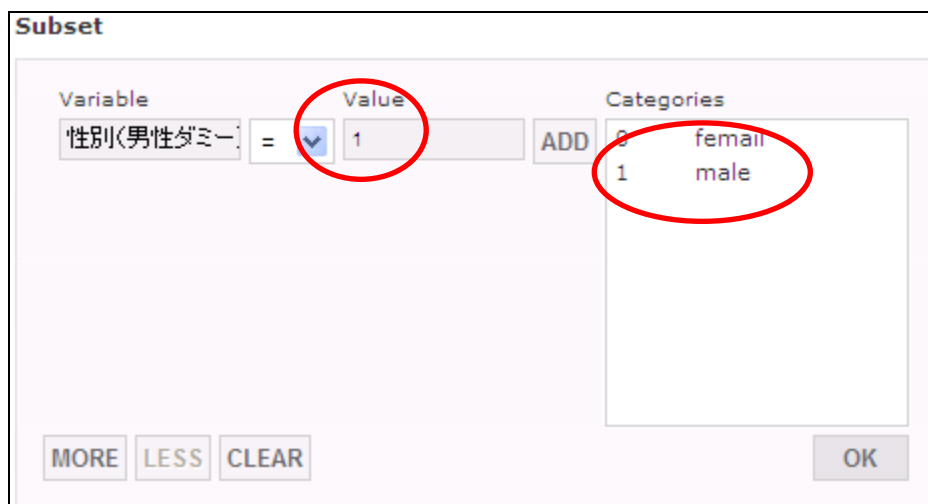


図 73 男性のみを選択

ケースを選択して図 68 と同様の相関係数を出したのが図 74 である。

Dataset: 東大社研・若年パネル調査(JLPS)Ywave1, 2007

Listwise Correlation analysis: [\[Change to pairwise\]](#)

Variables:
 再割り当て通勤時間(分) [\[Remove\]](#)
 再割り当て収入(本人) 02 [\[Remove\]](#)

		再割り当て通勤時間(分)	
再割り当て収入(本人) 02	Correlation	0.087 **	** p < 0.01
	Significance	0.0008	* p < 0.05
	Count	1461	

Display options

Significance
 Count

▼ Filter is on

図 74 ケースを選択して出力された相関係数

(7) その他の機能

以上の基本的な機能に加えて、Webview ではウェイト付け、時系列グラフのアウトプット、アウトプットの XML へのエクスポートなども可能である。詳細については、Nesstar Webview User Guide Version 3.5 を参照されたい。

[参考文献]

Norwegian Social Science Data Services,2010,「Getting Started Guide Version4.0」(2010年10月29日取得, nesstar-support@nesstar.com に問い合わせ)

Norwegian Social Science Data Services,2006,「Nesstar Server Administrator Guide Version 3.50」(2010年11月20日取得, <http://www.nesstar.com/support/documentation3-5.html>)

Norwegian Social Science Data Services,2010,「Nesstar Publisher v4.0 User Guide」(2010年11月11日取得,http://www.nesstar.com/help/4.0/publisher/download_resources/Publisher_UserGuide_v4.0.pdf)

Norwegian Social Science Data Services,2006,「Nesstar Webview User Guide Version 3.5」(2010年11月20日,<http://www.nesstar.com/support/documentation3-5.html>)

終章 Data Documentation Initiative の利用可能性

佐藤慶一

ここでは、各章の分析から明らかになった事項をまとめるとともに、SSJDA における Data Documentation Initiative の利用可能性について論点を整理することにした。

1. DDI 利用可能性検討から明らかにされた事項

第1章では、DDI 利用可能性検討の背景として、SSJDA における情報システムの状況が整理されている。SSJDA では、寄託された原データの整理からデータ公開や利用管理までの多様な情報が、複数のシステム等を用いて管理されている。原データが寄託された際には、エクセル形式の調査リストで公開データ準備作業進捗が管理され、別途寄託者情報がエクセル形式で管理されている。データが公開されると、ウェブページでアナウンスされ、公開データリストに掲載されたり、調査概要や調査票に関する個別のウェブページが作成される。公開データを用いた論文や書籍が刊行されると、利用報告論文リストのウェブページを更新する。利用者がオンライン上での検索や簡易集計を行うために、Namazu を用いたオンライン検索システム、SPSS WebAPP を用いたリモート集計システムが用意されている。さらに、一部利用頻度が高いデータについて、CD-ROM の郵送配布業務を軽減するため、SSJDA Direct というオンラインデータ提供システムを用意している。利用申請、利用承認、利用報告や成果物登録などのデータ利用管理は、MS Access データベースを管理するアプリケーションを独自開発して行っている。これらは、全体的なデザインのもとに順次相互の連携を意図しながら構築されてきたものではなく、それぞれ個別に構築を進めてきたものであった。それが故に、各システム間での連携が十分でなかったり、これまでこのリサーチペーパーで扱う DDI への検討が未着手であった等、全体的な見通しに欠ける部分もあったことが指摘されている。

第2章では、DDI 利用可能性検討の前提として、DDI 発展の背景をまとめてある。まず、DDI の記述対象であるメタデータについて、「データについてのデータ」を意味しており、具体的には、標本抽出の手続き、調査方法、質問文、コーディングの手続き等のデータについての記述内容を指すと整理されている。そして、メタデータに相当するコードブックが機械可読形式で提供されても、文字情報のみが保存されたファイルである場合、コンピュータによる検索は情報の内容如何を問わず検索対象の単語を全て拾い出すことになり非効率的となる点、利用者が必要とするデータの検索・提供を可能にするためにはメタデータの統一規格が必要とされるという DDI の基礎的背景が解説されている。さらに、SPSS や SAS といった通常の統計パッケージではメタデータへの対応が不十分であり、刊行されたデータ分析の結果と、そのもととなったデータとの関連づけが困難になるなどの問題が指

摘されていること、データの集計に関連した仕組みとメタデータの検索や必要な情報の参照を有機的に組み合わせることで研究を効率化することの重要性が指摘されていることが整理されている。その上で、CESSDA catalogue の利用例や、ミシガン大学の図書館目録検索システムで DDI のメタデータを図書情報検索の枠組みに流し込んだ事例が紹介されている。最後に、学術目的で収集された社会調査データのさらなる蓄積と公開、および二次利用による研究の発展のためには、研究者あるいはアーカイブの熱意と努力だけではなく、制度的・政策的な対応が必要であることが指摘されている。

第3章では、DDI の技術的側面、利用イメージ、導入ステップについて、DDI Alliance のウェブページ (<http://www.ddialliance.org/>) の情報を中心にまとめられている。DDI は社会調査のメタデータを XML 形式で記述する仕様であり、DI Alliance では、ユーザーが調査情報を XML で記述するルールを作成し、それを無償公開していること、定義ファイルや関連ドキュメント一式は、DDI Alliance のウェブサイトよりダウンロード可能であることなどが整理されている。また、2008年に DDIversion3 が公開され、データの保存にとどまらず研究データのライフサイクル全体にわたった記述を支援することを目的とした「データライフサイクルモデル」が採用されたこと、DDI はデータアーカイブにおける調査データの管理だけではなく、調査の設計や調査の実施にも活用されることを目指すようになっていくことなどが書かれている。その上で、DDI の利用イメージについて、2010年10月24日から29日にドイツで開催された Workshop 「The Data Documentation Initiative (DDI) XML Standard」での配布資料から主要なものをピックアップして、翻訳や調整を加えた一覧が作成され、また、導入ステップとして、新しい研究プロジェクトを始める際の DDI の使い方と、既存のデータから DDI 文書の作成方法について情報が整理されている。

第4章では、SSJDA で検討した社会調査メタデータ管理ソフト Easy DDI Organizer (EDO) の設計についてまとめられている。はじめに、DDI Alliance のホームページに掲載されているソフトウェアのうち、DDI 文書の作成・編集機能を有するものを対象として、既存の DDI 文書作成・編集ソフトウェアの機能および特徴を整理した。それらのうち、データライフサイクルモデルを採用した DDI 3.0 以降を扱えるのは DDI 3.0 Editor Lite (無料)、Colectica Express (有料) の2つであった。DDI 3.0 Editor Lite はインポート機能を持たず、またこのソフトを使用して作成した文書を読み込むことも不可能で、さらに、規定されている必須項目をすべて入力するまで文書を出力することができないため、研究の進度にあわせて DDI 文書を作成するという用途には不向きで、完了した研究のメタ情報を入力する用途に限られることが分かった。Colectica では、データファイルや既に作成した DDI 文書のインポート機能、入力したメタ情報を加工しコードブックや質問紙、統計パッケージ用のデータファイルやセットアップファイル(変数名やラベルをつけるスクリプト)を出力するエクスポート機能も提供されていることが分かった。ただし、DDI3.0 以降に採用された Group モジュール(複数の研究・調査のメタ情報の比較等に関する情報を扱う機能)には対応していないことも分かった。以上を参考に、DDI 文書作成ソフトウェア Easy DDI Organizer(EDO)を設

計した。EDO は DDI3.0 以降に採用されたデータライフサイクルモデルの特長を生かし、研究計画からデータの保存までの一連の流れで記録しておくべき情報を入力するフィールドを用意し、調査の設計、メタ情報の保存を支援するというコンセプトで設計された。2010年12月にEDOの開発を開始し、2011年5月現在、メタ情報入力・保存機能、保存したメタ情報をDDI3.1準拠のXML形式でエクスポートする機能を実装したプログラムのテストが最終段階に入っている。ペーパーでは既に実装された機能について逐次解説がなされた。

第5章では、ヨーロッパのCESSDA等で開発されたNesstarの特徴と海外データアーカイブにおける利用状況が整理されている。はじめ、システムの基礎的な特徴について要約されている。Nesstarは、Publisher、Server、Web Viewの3つのソフトウェアにより構成されている。SPSSやStata等のデータセットを読み込み、メタデータを追加作成し、DDI形式とするのがPublisherである。そしてPublisherによって作成されたDDI形式のデータを管理し、インターネット上に公開できるようにするのがServerである。さらにServerによって公開されたデータはWebviewによって検索やオンライン分析が可能となる。1999年よりテスト運用され、2010年12月にバージョン4.0がリリースされ、Publisherはフリーで配布されるようになった。ただし、最新のversion4.0に対応しているのはDDI1.2.2である。次に、ドイツの中央アーカイブ(GESIS-ZA)のシステムZACATの利用方法、アメリカのNORC(National Opinion Research Center)で運用されているNesstarの利用方法などが概観されている。

第6章では、Nesstarの導入方法について、実際の手順がまとめられている。具体的には、インストールの手順、PublisherとServerによる調査データの公開手順、Webviewでの分析手順である。主要なポイントのみを書き出すと、まず、インストールするコンピュータのシステム要件は、OS: Windows XP, 2000 or NT4, Windows Server 2003, CPU: 1 GHz, メモリ: 1GBと汎用的である。Publisherでインポートできるデータ形式は、Nesstar(.Nestar), NSDstat(.NSDstat), DDI Document(.xml), SPSS(.sav, .por), SPSS Syntax(.sps), STATA(*.dta), Statistica(*.sta), NSDstat(*.nsf), dBase(*.dbf), DIF(*.dif), Delimited Text(*.txt, *.csv, *.sdv, *.cdv, *.prn), PC-Axis(*.px), Excel(*.xls)と多い。インストール手順、データのインポート作業、追加するメタ情報についてはペーパー内で逐次紹介されている。さらに、ServerによるWebviewの設定手順とPublisherによる公開手順について記述されている。2章、5章でも見られているが、最後にあらためて、変数の検索、記述統計、クロス分析、各種グラフの作成、変数の加工、相関係数、回帰分析といったWebviewの分析機能について、一通り確認がなされている。以上の作業を通じて、Nesstarのライセンスと必要なサーバーおよび適切なメタデータを含めた調査公開データがあれば、SSJDAにおいてもNesstarの導入が可能であることが確認された。

資料編では、東アジアのアーカイブにおけるNesstarの利用事例として、韓国KOSSDA(Korea Social Science Data Archive)と台湾SRDA(Survey Research Data Archive)についてレポートされている。KOSSDAでは、KGSS(韓国版総合社会調査)、韓国青少年政策研究院の調査、長期経済統計が搭載されており、nesstar利用に課金をしている点が特徴的である。

SRDA は、中央研究院の大規模な調査（華人家庭の動向調査、社会変遷に関する基礎調査、教育に関する長期追跡調査等）が搭載されており、2011年5月現在で最新バージョンである Nesstar4.0 を導入している。当初は中国語（繁体字）データのみを扱っていたが、現在では英語版データも扱われている。

2. Data Documentation Initiative の利用可能性

SSJDA における Data Documentation Initiative の利用可能性に関して、主要なポイントを整理して締めくくりとしたい。

- ① DDI には大きく 2 つの使い方がある。その 1 つは、新たな社会調査活動の支援である。調査研究に関する情報を、既に整えられた XML 形式で一元的に管理することで、必要な情報を適切に保存することができる。個人研究者や小規模な研究グループが必要と判断したメタデータが、報告書や論文等の成果物に記載されるだけでは、メタ情報として不十分な場合も想定される。一連の調査研究活動終了後にあらためてメタ情報を整理する際には、必要な情報が散逸してしまうことも想定される。DDI 形式で社会調査のメタ情報を保存できる簡易なソフトウェアがあれば、様々な社会調査のメタ情報が適切に保存されていくことに寄与するだろう。また、整備したメタ情報を用いて、調査票やコードブックを自動生成することが技術的には可能であり、そのような機能が実装されれば、調査実施者の作業効率性に寄与することとなるだろう。このようなメタ情報の管理は、SPSS や Stata 等既存の統計ソフトウェアでは対応していない。第 4 章で紹介した Collectica というソフトウェアでは一部機能が実装されているが、日本語対応が十分でなく、また有償であり、日本で一般の研究者や研究機関で普及するとは考えにくい。
- ② 4 章に見たように SSJDA では、社会調査メタデータ管理ソフト Easy DDI Organizer (EDO) の開発を行っている。2011 年度は、整備したメタデータから、SPSS ファイルから変数ラベル等を読み込む機能、設問項目のシーケンスを管理する機能を追加した上で Word 形式の調査票自動生成が行われる機能、コードブックが自動生成される機能の実現に向けて検討を継続している。大きな調査会社では独自の調査支援システムを持っていることも想定されるが、調査ノウハウの蓄積が十分でない中小の研究機関や研究者等にとっては、既に整えられた DDI 形式で調査のメタ情報を適切に保管できること、それにより調査票やコードブックの作成にかかる手間が軽減されるということは、相応のメリットを有するものと考えられる。また、良質な社会調査のメタ情報の蓄積に寄与するという社会的なメリットもある。今後の作業イメージとしては、開発に応じて、利用テストを行いながら利用メリットを確保していくこと、計量分析セミナーなどで β 版を紹介する場を作っていくこと、β 版をフリーで公開していくことなどを検討している。
- ③ DDI の使い方のもう 1 つとして、既にある調査情報を変換して、データアーカイブの情報システムに活用する方法がある。SSJDA から見ると、現時点でデータ寄託時に DDI 形式のメタ情報を受け入れても、不足している情報を再度問い合わせることが少なくない。

ることや、既に印刷された文書の中から必要な情報を探すなどの労力軽減に寄与するものの、それを利用するシステムを持たなければ、本質的なメリットは少ない。SSJDAにおいてDDIを活用するとなると、いずれ、蓄積したメタ情報を活用するための情報システムが必要となってくる。5～6章で紹介したNesstar Publisherでは、SPSSやStata等の既存フォーマットから必要な変数情報等を抽出し、必要なメタデータを加えてDDI形式とすることが可能である。Nesstar Webviewでは、掲載したデータの検索やオンライン集計機能も実現されており、導入することでDDIを活用することができる。Nesstarは現在DDI1.2.2という古いバージョンにのみ対応しており、今後の開発動向について留意が必要と考えられる。2章で少し触れられているが、カリフォルニア大学バークレー校で開発されたSDA (Survey Documentation and Analysis) では、Nesstar Webviewよりも高度な分析機能が提供されており、DDIとの互換性にも配慮がある。オランダのUnivesity of Tilburgにある研究機関であるCentERdataでは、DDI3を用いたQuestasyというオンラインデータ公開システムを開発中であり、SPSSデータをインポートして必要なメタデータを与えると、ウェブサイトが自動生成される仕組みの開発が目指されている¹。

Questasyについては、現時点ではCentERdataでの活用が目的であり、他のデータアーカイブでの活用がどの程度意図されているのかは不明瞭である。今後も、IASSIST等で、他の国際機関や研究機関におけるDDIを活用するシステムの開発動向に留意していく必要があるものと見受けられる。

- ④ SSJDAでは、韓国や台湾でも既に利用されているなど、システムの普及がある程度進んでいるNesstarについて、試験的な利用を検討している。2011年度には、Nesstar用のサーバー購入やNesstarの利用ライセンスを、共同利用・研究拠点の予算に計上した。Nesstarについては、利用者管理が不十分で、SSJDA Directで実現している利用申請時の承認プロセスを再現することが難しく、用意されているダウンロード機能については削除して、現在SPSS WebAPPで実現しているリモート集計システムの代替としての利用を想定している。今後、掲載するデータについて調整を行い、2011年度中に試験運用サイトを公開する予定である。
- ⑤ 1章で触れたデータライフサイクルモデルとデータアーカイブの関係について再掲する。DDI3では「データライフサイクルモデル」という、調査設計からデータアーカイビングまでに扱うメタ情報を一括してXML形式で整備して、さらにデータアーカイブ等で活用していくことが提唱されている。膨大な情報を扱うデータアーカイブでは、機能別に個別にシステムを構築し、多様なフォーマットで情報を管理していると、その連携をさせることにも膨大な労務が発生する。DDI対応には、一見コストが多いように見えても、何らかの統一された情報記述方法で、データアーカイブが利用するあらゆる情報を一元的に管理しておき、そこからデータクリーニングや寄託者・利用者対応業務に必要な情報を取り出したり、さらにはウェブサイトやデータ検索、オンライン集計やデー

¹ <http://www.iassistdata.org/downloads/iqvol3312amin.pdf>

ダウンロードシステムの情報が自動更新されていくような仕組みを考えることが、最終的には作業効率を高め、必要なメタデータが適切に蓄積されていくことにつながるものと考えられる。このような作業には、データアーカイブで扱っている情報を網羅的に把握しつつ、情報技術の展開にも適切な見通しを持った上で、総合的なシステムデザインを図り、事を進めていく必要があるものと考えられる。

資料編 東アジアのアーカイブにおける Nesstar の利用

安藤理

1. Korea Social Science Data Archive(KOSSDA)

KOSSDA (Korea Social Science Data Archive) は、韓国社会科学図書館 (Korea Social Science Library) と韓国社会科学データアーカイブ (Korean Social Survey Data Archive) を統合して設立された非営利社会科学データアーカイブである。KOSSDA の主要な目的は、韓国における量的・質的なデータおよび文献資料をデジタルなカタチで収集・提供することであり、大学、政府、企業からデータの寄託を受けている。サーベイデータは 2011 年 2 月現在で 700 を超え、インタビューなどの質的データは 20000 を超えている。その分野は、政治学、経済学、社会学、文化論など多岐にわたる。量的データは質問紙調査や集計データである。質的データとしては公文書、フィールドノート、インタビュー、ナラティブ・ヒストリーのデータが挙げられる。文献資料としては、学術雑誌、調査報告書が含まれる。KOSSDA のデータは、標準的なメタデータ化のプロセスに沿ってデジタル化され、KOSSDA データベースシステムを通して利用者に提供可能となっている。2011 年 2 月現在、1000 近く（そのうち英語ページから利用可能なのは 600 ほど）の量的データ、80 ほど（そのうち英語で読めるのは 30 ほど）の質的データが利用可能となっている。この一部のデータは Nesstar にてオンライン分析を行うことができる。Nesstar で分析可能なデータは表 1 の通りである。インターフェイスは英語のままであるので、海外の利用者も利用しやすい環境となっている。

検索システムは、Nesstar と別に設けられており、タイトル、調査主体、寄託者、変数、アブストラクトからの通常のキーワード検索が可能である。また、量的データか質的データかというデータ形式、学術分野、年度、言語による検索も行えるようになっている。さらには、シソーラス検索（類語検索）も可能になっている。たとえば、図 1 は、「Welfare」に関連するデータを検索した画面である。その結果、「Child Welfare」「Social Welfare」などの関連語が画面中段に表示される。そして、これらのうち「Educational Welfare」をクリックすると、それに関連する語が「Broader terms」「Narrower terms」「Related terms」の三つに分類されて表示される。このうち興味のある用語にチェックを入れて検索をすると、その用語に関連したデータが絞り込まれることになる。

表 1 Nesstar で利用可能な主要データ

	KGSS 韓国総合社会調査、2003	青少年の生活・意識に関する実態調査	戦前のデータ:「政府部門の主要統計:1910年から1943年まで」 韓国における経済成長:1910年から1945年まで
研究課題の名称	韓国人の国家アイデンティティ、不平等への知覚、労働志向、家族価値:「国際社会調査 (ISSP)」および「韓国総合社会調査 (KGSS)」の資料を利用した国際比較の研究	国際基準対比の青少年人権の実態に関する調査研究:その2	
資料シリーズの名称	韓国総合社会調査 (KGSS)	青少年の生活・意識に関する実態調査	韓国における長期的経済統計
研究実施機関・資料の寄託者	韓国成均館大学サイバーリサーチセンター	韓国青少年政策研究院	落星堡経済研究所
シンーラス	KGSS;韓国総合社会調査;政治満足度;経済満足度;政治的志向;支持政党;投票;国政評価;対北朝鮮関係;国家親密度;南北統一;資本主義;企業;財閥;労働組合;新聞;階層意識;経済展望;家族構造;職業;収入;国家アイデンティティ;集団アイデンティティ;国家アイデンティティ;国際関係;外国人;人種;開放;家族価値;結婚;結婚満足度;性役割;女性就職;業務志向;労働志向;職務満足度;組織満足度;転職意志;不平等;葛藤;所得不平等;主観的階層意識;一般的信頼	青少年生活;青少年の価値観;青少年の逸脱;青少年の非行;アルバイト;進路選択;体罰;ストレス;親との関係;教師との関係;セクハラ;友人関係;同僚集団;校則;差別;生活満足度;課外活動;規律;悩みの相談;私教育;青少年団体	財政;公共部門;貯蓄;地方財政;移転支出;朝鮮総督府;公務員;植民地経済;日本統治期;消費支出
アブストラクト	「韓国総合社会調査」は、毎年、全国を対象にしたサンプル調査を行い、韓国社会の構造および変化に関する研究、そして国際比較の研究に利用されるよう、社会基礎データを生成させ、研究目的であればだれでも自由に使用できるようにする、公益目的の社会科学資料の創出のための事業である。この調査は、アメリカのシカゴ大学におけるNational Opinion Research Center (NORC) の、General Social Survey (GSS)を基本的モデルにして、ヨーロッパおよび東アジアの各国で実施されるGSSの形をした調査の内容の一部受容するとともに、韓国社会の特殊性も考慮するという方式で設計された。 この質問紙は、毎年継続される質問群と国際比較のために追加されるモジュールで構成されている。モジュールは、International Social Survey Programme (ISSP)の会員国が共同で調査する年次テーマモジュール、韓国、中国、日本、台湾などの東アジア4カ国が隔年の周期で共同的に開発して調査するEast Asian Social Survey (EASS)のテーマモジュール、そして特別研究のためのテーマモジュールなどが有機的に結合され構成されている。	韓国青少年政策研究院は、青少年の人権実態に関する国家間の比較研究を通して、韓国の青少年の人権水準を把握し、国際的レベルの青少年人権環境を作るために、2006年から、「国際基準対比韓国青少年の人権水準の研究」を実施している。「青少年生活・意識に関する実態調査2007」は、「国際基準と比べた韓国青少年の人権水準の研究」の2次年度資料を収集するために実施された。この調査資料は、生存権、保護権、発達権、参与権などの4つの大きな領域から構成されており、領域ごとの主な調査項目は以下のようである。 1) 生存権: 栄養および健康実態、飲酒/喫煙の頻度、犯罪からの安全の確保の程度 2) 保護権: 体罰された経験、差別された経験、逸脱/非行の経験、逸脱/非行による被害の経験など。 3) 発達権: 進路計画、親との関係、教諭との関係、勉強への態度、余暇時間、生活満足度、団体への加入/活動など。 4) 参与権: 青少年に対する親の態度、校則への認識、頭髪/服装への規制に関して、校則に生徒の意見が反映されるか、生徒会の活性化はされているか、など。	表1: 政府部門の歳出構成、1911年から1943年まで 表2: 政府サービス生産者の付加価値、1911年から1943年まで 表3: 政府サービス生産者の算出および最終的消費支出、1911年から1943年まで 表4: 政府部門の固定的資本の形成、1911年から1942年まで 表5: 総督府の歳入および歳出、1911年から1942年まで 表6: 政府サービス生産者の貯蓄、1911年から1943年まで 表7: 総督府歳出の経済的分類、1911年から1943年まで 表8: 地方財政の歳出の経済的分類、1911年から1943年まで 表9: 官吏・公吏の給料および人数、1913年から1942年まで 表10: 朝鮮人の官吏・公吏の人数と職級ごとの割合、1913年から1942年まで 表11: 総督府の目的別支出、1913年から1943年まで 表12: 総督府の目的別固定資本の形成、1913年から1943年まで 表13: 総督府の目的別移転支出、1913年から1943年まで 表14: 総督府の固定資本形成の資本形態ごとの分類、1913年から1943年まで
調査対象者	18歳以上の男女	中学生	
資料収集の周期	2003年から毎年	韓国青少年政策研究院	
資料収集方法	直接訪問による面接調査として、調査員による記入方式が原則	集団面接による自己記入式調査	
サンプル抽出方法	多段階地域確立サンプル抽出 (multi-stage area probability sampling)	多段階層化集落サンプル抽出	
サンプルの規模	1,315 (有効回答)	3,413 (有効回答)	
備考	「韓国総合社会調査」は、予備調査 (A1-2003-0005)、1次 (A1-2003-2006)、2次 (A1-2004-0001)、3次 (A1-2005-0001)、4次 (A1-2006-0003)、5次 (A1-2007-0026)、6次 (A1-2008-0009) までの、7回の調査が実施され、全て KOSSDA で利用可能である。	「韓国青少年開発院」は、2007年5月から、「韓国青少年政策研究院」として名称が変更された。	

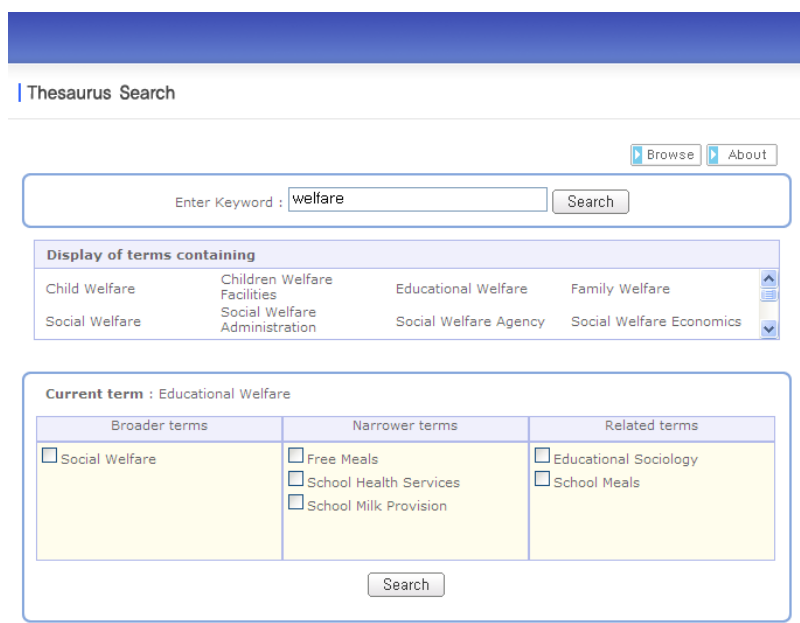


図 1 シソーラス検索 (KOSSDA ウェブサイト http://www.kosdda.or.kr/eng/w02_01s.asp)

データサービスを受けるためには、有料利用者か無料利用者として登録をする必要がある。①Web上で利用規約に同意し、②所属や連絡先などの個人情報をWeb登録し、③会費を支払い、④会員として認定されると登録される。会員のタイプと年会費および必要書類は表2にまとめた。機関会員の年会費は、Group A (博士課程のある大学) は1000ドル、Group B (修士課程のある大学) は500ドル、Group C (学士課程のみの大学) は300ドル、Group D (研究所) は200ドルである。

表 2 KOSSDA の利用者タイプ

利用者タイプ		年会費	備考
機関会員利用者		1000ドル	博士課程まである大学
		500ドル	修士課程まである大学
		300ドル	学部のみある大学
		200ドル	研究所
個人有料会員	教授、研究員など	100ドル	
	学生	50ドル	学生であることを示す証拠書類が必要
無料会員			

出所：KOSSDA ウェブサイト http://www.kosdda.or.kr/eng/w01_04c.asp

会員資格には、機関会員と個人会員がある。個人会員は学生、研究者がふくまれ、機関会員には大学図書館や研究所やその他の組織が含まれる。個人会員と機関会員の構成員はともに、KOSSDAのすべてのサービスに直接アクセスできる。会員資格ごとに利用可能な

サービスは、表 3 にまとめた。「カスタム資料請求」とは、利用者が目的とする分野の情報と雑誌のリストを提供するサービスである。

表 3 会員資格ごとに利用可能なサービス

サービス		会員資格		
		無料会員	有料会員	
検索と分析	閲覧と検索		○	○
	オンライン分析 (NESSTAR)	検索 記述統計	○	○
		クロス表 相関分析回帰分析		○
	プレビュー(音声、ビデオなどの質的データ)			○
MY KOSSDA	E-mail によるKOSSDAニュースレター		○	○
	カスタム資料請求(SDI)		○	○
	DDS(文献コピー申請)			○
	データのダウンロード			○

出所：KOSSDA ウェブサイト http://www.kosdda.or.kr/eng/w01_04c.asp

データ申請の目的は、研究、教育、探索に分けられる。KOSSDA 協会に属する利用者は 3 つのすべての目的で申請ができるのに対し、非構成員の大学の教員は教育目的の申請に限られる。

量的・質的データの利用申請をする際には、ホームページからダウンロードできる申請書に必要事項を記入し、目的別に分けて申請を行う。研究利用の場合は、半角 200 字以内で研究テーマを記入する必要がある。その後、KOSSDA の利用許可を受けてダウンロードが可能となる。

2. 台湾 SRDA (Survey Research Data Archive)

SRDA は、台湾のデータアーカイブとして 1994 年 11 月に設立された。保有データは、学術質問紙調査、世論調査、政府質問紙調査である。社会学、経済学、教育学、政治学、公衆衛生といった分野における学術調査を収集している。データの入手経路は、a)研究者からの寄託、b)協働して実施した調査プロジェクト、政府機関などである。政府調査に関するデータセットは、行政院主計処(Directorate-General of Budget Accounting and Statistics)によって行われた人的資源調査 (Survey on Human Resources)、社会趨勢発展調査 (Social Development Trends Survey)、家庭収支調査 (Survey on Family Income and Expenditure) や内政部が行った重要な調査のデータなどである。Nesstar で分析可能なデータは表 4 に示す通りで、非常に大規模な経年調査等が搭載されている。

表 4 Nesstar で分析可能なデータ

調査名	華人家庭の動向調査データベース	台湾の社会変遷に関する基本調査	台湾の教育に関する長期追跡調査データベース	台湾青少年の成長過程に関する調査
年度	1999年～	第一期:1984～1985年 第二期:1990～1994年 第三期:1995～1999年 第四期:2000～2004年 第五期:2005～2009年 第六期:2010～2014年	2001年～2007年	2000年3月～2007年6月
サンプル数	台湾:4000人強 中国(上海・浙江・福建):4370人	第一期: 質問紙Ⅰ:4233人 質問紙Ⅱ:4199人 第二期以後: 質問紙Ⅰ、Ⅱ 各2000人前後	第一回:39,337人 第二回:37,227人 第三回:20,173人	5586人
調査主体	中央研究院 人文社会科学センター	中央研究院 社会学研究所	中央研究院 社会学研究所、 欧米研究所	中央研究院 社会学研究所
調査方式	質問紙	質問紙	質問紙 (保護者に対しては 電話調査も実施した)	質問紙
概要	華人の家庭における成人を無作為に抽出し、対象としたパネル調査。また、その調査対象から派生する子女などの親族にも追跡調査を行った。このデータベースの整備を通して、経済・社会・心理・慣習などの面から華人家庭について全面的な研究を行う。	1984年から台湾全域でサンプル調査を今までに21回行った。その内容としては、家庭・教育・社会階層と移動・政治文化・選挙意識・マスメディア・文化価値・宗教などがある。この調査の特徴は、同じアンケートを二つ以上の違う期間で行うことである。台湾社会の変遷を多方面から研究するデータベースを作成することが目的である。	中央研究院、教育部と行政院国家科学委員会が共同企画した長期的研究計画である。長期的な調査により作成された台湾の代表的な教育領域に関するデータベースである。この計画は、自記式質問紙で調査が行われ、調査範囲は、台湾地域における中学校・高等学校(三年制の職業高校を含む)及び五年制職業学校の生徒を対象としている。これらの生徒を研究の核としながらも、生徒の学習経験に最も影響を与えるいくつかの要素:保護者、教師や学校にも研究範囲を拡張、調査を行った。	台湾の青少年に対して長期的・系統的に調査を行い、データを収集し、それにより台湾青少年の生活経験を分析する。また、家庭、学校及び地域が及ぼす影響が極めて大きいと考えられるので、それらを考慮した分析が必要不可欠である。調査には、教育期待・教育達成・就業の機会及び各成長過程においての社会資本が如何に活用されるかなどの重要なテーマが含まれる。

出所: SRDA ウェブサイト <http://140.109.171.222/webview/>

ユーザー登録は、①Web 上で利用規約に同意し、②所属や連絡先などの個人情報 Web 登録し、③登録画面のコピーと教育研究機関所属証明書のコピーを郵送し、④会員として認定されてはじめて利用できる。登録は無料で、期間は2年間である。登録が許可されると、データダウンロードおよびオンライン分析(Nesstar)が利用できるようになる。データダウンロードには、研究テーマの登録などの特別な手続きは必要なく、検索したデータを直接ダウンロードできるようになっている。

SRDA では、2011年5月現在で最新バージョンである Nesstar4.0 を導入している。当初は中国語(繁体字)データのみを扱っていたが、現在では英語版データも取り扱っている。

そのほか、政府統計資料についても、中央研究院調査研究センターのリンクを通じて、行政院主計処による「労働力調査」「女性の結婚・育児及び就業に関する調査」「労働力シミュレーションー追跡調査データベース」が Nesstar を用いてオンライン分析可能である。