

第4回京都大学研究データマネジメントワークショップ

～科学研究費補助金基盤研究（A）キックオフイベント～

セッション2 「科研費プロジェクト『葛桜』」

京都大学東南アジア地域研究研究所

原 正一郎

地域研究における研究データの特徴・現状

- 人文社会学研究データが多い
 - デジタル化は遅れ気味か（例：デジタルテキストの作成には予算も時間もかかる）
 - 公開が困難な研究データも多い（著作権，肖像権，政治的配慮等）
- デジタル化されている研究データは，データとしてよりもデータベースとして公開されている事例が多い
 - 書誌的データベースが多い ⇒ 現物へのアプローチが困難な事例も多い
 - 個人研究者や研究プロジェクトに関わる研究データのデジタル化やデータベース公開は遅れている，仮に公開されても停止してしまうことが多い（予算，人員等）
 - ⇒ 長期保存や長期利活用が困難
 - 参考：JCASニューズレター，No.15 (2013)，特集 JCAS加盟機関所蔵の研究資源一覧
- 研究データ管理に対する理解や関心は分野や世代（？）によりさまざま



対象を少し広げて

～ 人文科学研究における研究データの特徴・現状～

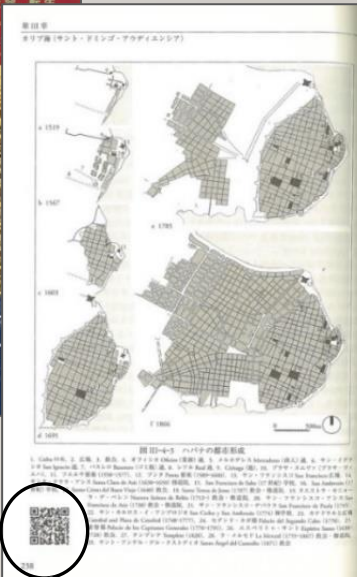
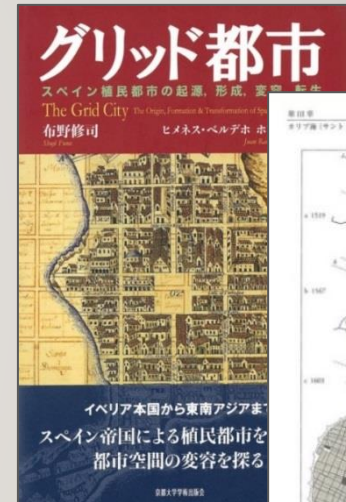
- 多様性
 - 対象（文書，絵画，モノ，辞書，報告書，数値，統計，文字等）
 - 記述（メタデータの標準化は進んでいない，無理？）⇒ 広範囲な共有や利活用が困難
 - 公開者（大学，研究所，学協会，企業，協会，自治体，個人等）⇒ 標準化が進まない，長期保存が困難
- 書誌的データが多い ⇒ 現物へのアプローチが困難な事例も多い
 - 参考：人文系データベース構築事例のポータルサイト・データベース (<http://www.jinbun-db.com/database>)
人間文化研究機構 nihulNT (<https://int.nihu.jp/>)
⇒ 機構外研究者のデータベースは入っていない？

- 研究データ管理への関心は概して低いと考えられる



東南アジア地域研究研究所の取り組み 1

- まずはデータベースサービスから
 - データベース構築支援システムの提供 (My データベース)
 - 利活用支援 (My データベース APIの提供)
 - 共同利用・共同研究拠点としての機能 (対象を研究所や学内に限定しない)
- 次に高度化
 - 時空間インタフェースの導入
 - RDF+SPARQLによる再構築 (Open Dataへの対応)
- サービスニーズの多様性
 - マルチメディア的な利用：研究書籍とデータベース (地図や動画) の連携：データベースはデータ置場
 - データ置場としてのニーズ：今後は数値データ等が増える見通し
⇒ レポジトリ機能の開発 (中)

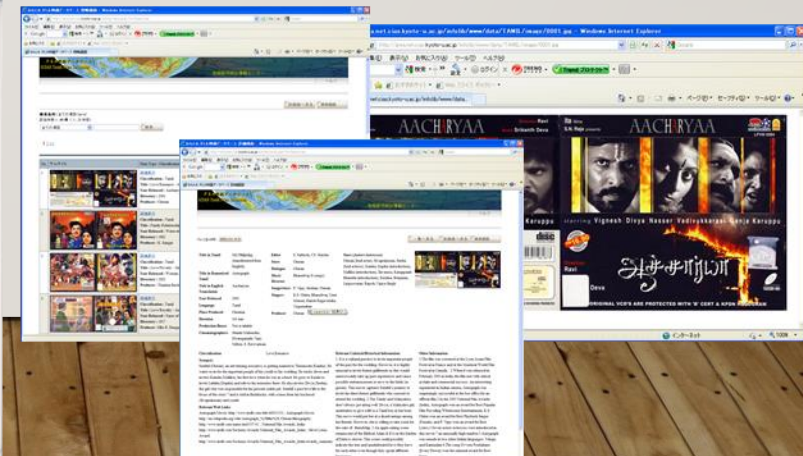


東南アジア地域研究研究所の取り組み 1

～東南アジア地域研究研究所のデータベースの例～

No.	Title cover	Classification/ Key Words:	Title in Tamil	Title in Romanised Tamil	Title in English Translation	Language	Colour	Year Released	Place Produced	Duration	Stars	Director	Cinematographers	Editor	Story
1		Love/Romance	Ml;Nlhfpuhg; (transliterated from English)	Autograph	Aacharyaa	Tamil	Colour	2004	Chennai	161 min	Cheran (lead actor), Sivaprakasam, Sneha (lead actress), Kanika, Gopika (introduction), Mallika (introduction),	Cheran	Shanki Mahendra, Dwaraganath, Vijay Milton, S. Ravivarman	S. Sathesh, J.N. Harsha	Cheran

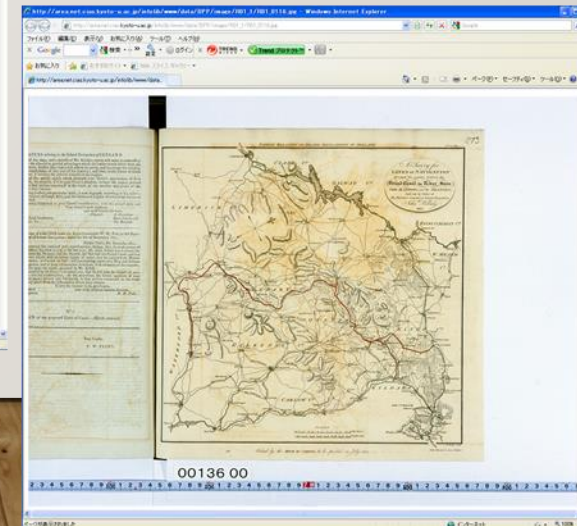
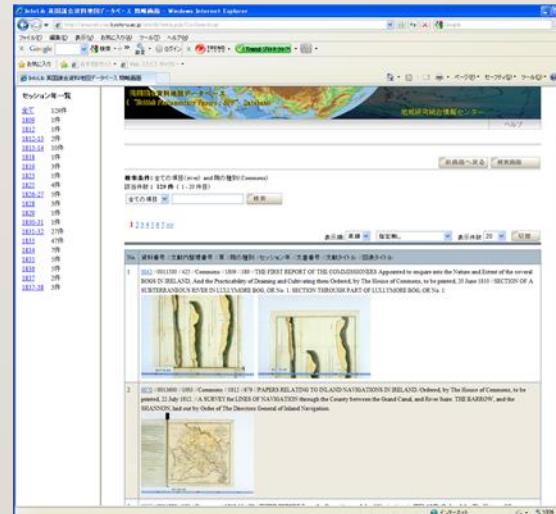
Dialogue	Music Director	Songwriters	Singers	Producer	Production House	Synopsis	Relevant Cultural/Historical Information:	Other Relevant Details	Relevant Web Links
Cheran	Bharadwaj (6 songs)	P. Vijay, Snehan, Cheran	K.S. Chitra, Bharadwaj, Unni Menon, Harish Ragavendra, Yugaendran	Cheran	Not available	Senthil (Cheran), an advertising executive, is getting married to Thenmozhi (Kanika). He wants to invite the important people of his youth to his wedding. He tracks down and invites Kamala.....	1. It is a cultural practice to invite important people of the past for the wedding. However, it is highly unusual to invite former girlfriends as this would unnecessarily rake up past experiences and cause possible embarrassment or envy to the bride (or groom).	1.The film was screened at the Lyon Asian Film Festival in France and at the Montreal World Film Festival in Canada. 2. When it was released in February 2004 in India, the film met with critical acclaim and commercial success.	Autograph Movie: http://www.imdb.com/title/tt0411131/ , Autograph Movie: http://en.wikipedia.org/wiki/Autograph_%28film%29 , Cheran filmography.....



東南アジア地域研究研究所の取り組み 1

～東南アジア地域研究研究所のデータベースの例～

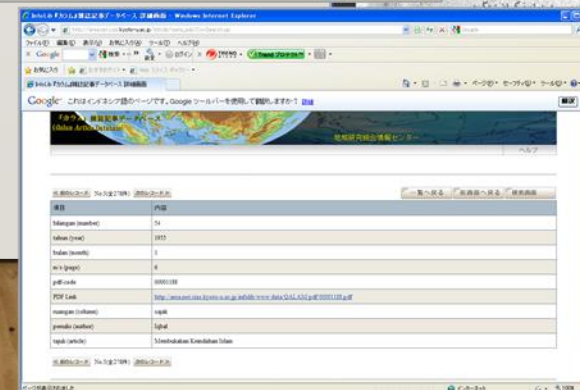
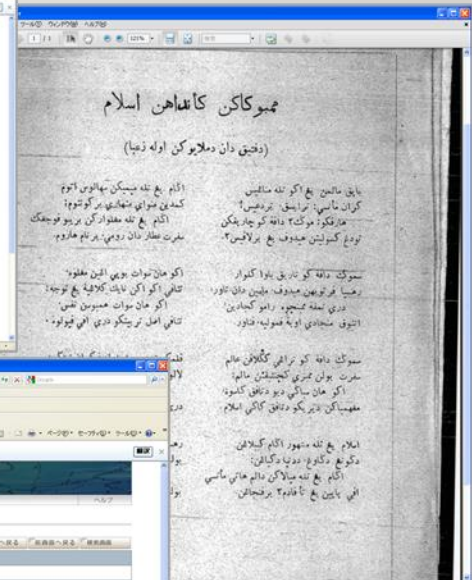
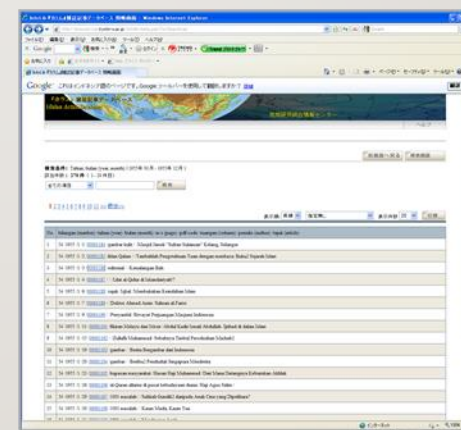
ID	Survey No.	page	Parts	Cat.	Year	Book	Text	Caption	
21	6100	408	1	Commons	1806	365	ACCOUNTS, &c. PRESENTED TO THE HOUSE OF COMMONS, RELATING TO THE INLAND NAVIGATIONS OF IRELAND, AND TO THE PORT OF DUBLIN./Ordered to be printed 10th July 1806.	PLAN of Proposed Improvements in DUBLIN HARBOUR.	D:\DISK1\YR01_1\YR01_0022.jpg



東南アジア地域研究研究所の取り組み 1

～東南アジア地域研究研究所のデータベースの例～

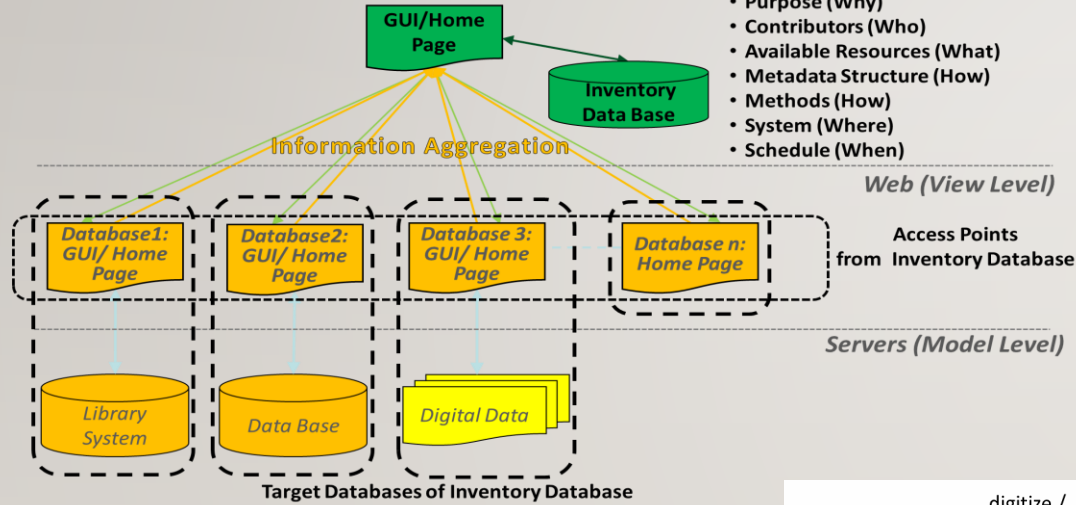
bil.	tahun	bln.	m/s	ruangan	tajuk	penulis	mikrofilm
1	1950	7-8	1	gambar kulit	Bu Karno (Isteri President Sukarno)		
1	1950	7-8	2		(keterangan gambar kulit & gambar hadiah)		



東南アジア地域研究研究所の取り組み 2

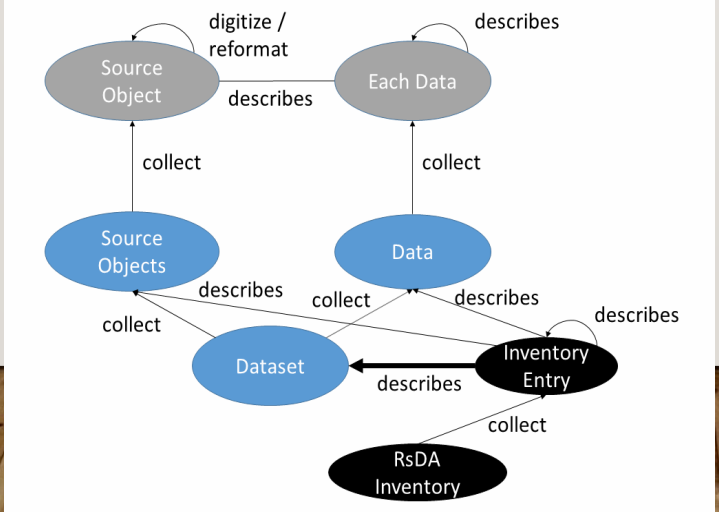
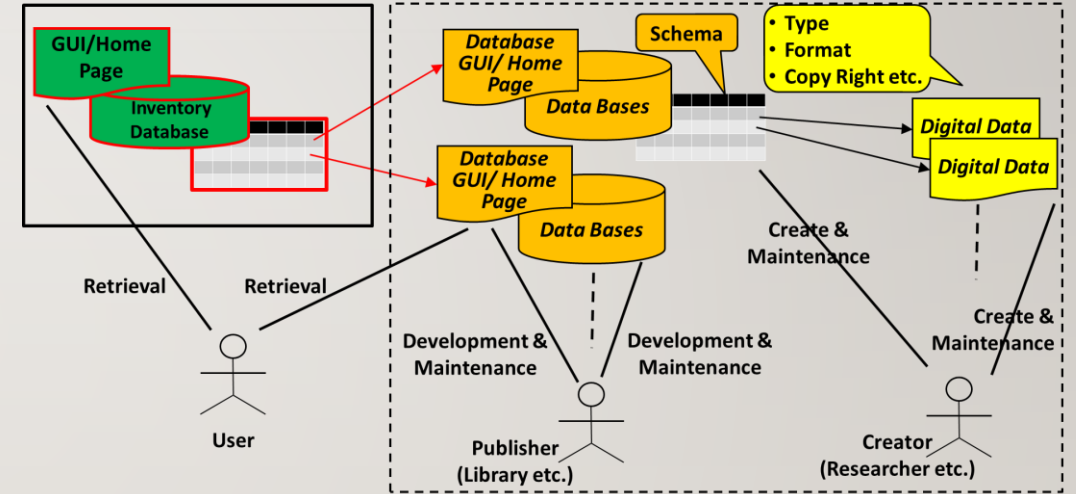
～ 研究デジタルデータに関するINVENTORY データベースの構築～

Phase1 of the Project



Discussion

- Purpose (Why)
- Contributors (Who)
- Available Resources (What)
- Metadata Structure (How)
- Methods (How)
- System (Where)
- Schedule (When)



Dataset
URL
alternative URLs
title (original)
title (English)
title (other language)
subject
description (English)
description (other language)
contributor
rights
accessibility
issued
modified
contact

Data
data language
data schema
Source Object
source (source language)
source (English)
source (other language)
source type
source format
source spatial coverage
source temporal coverage
source language
RsDA Entry
comment

東南アジア地域研究研究所の取り組み 2

～ 研究データのINVENTORY データベース構築～

The screenshot displays the RSDA Search website interface. At the top, there is a search bar labeled "Search title:". Below it, a grid of search results is presented, each in a white box with a blue header and purple tags. The results include:

- Thai Language "The Three Seals Law (Kotmai Tra Sam Duang)" (Royal Institute version)**: Includes tags for Thailand, Laws and Decrees, Ayutthaya, タイ, 法令・布告, and アユタヤ. Description: The Thai language "The Three Seals Law" (Kotmai Tra Sam Duang) is an aggregation of laws, ordinances and edicts compiled from the mid-14th century until the beginning of the 19th century and based on the manuscripts of the code of all existing things. These documents were saved af.
- Map Digital Collection Search**: Includes tags for Map, Taiwan, 地図, 台湾. Description: Map Digital Collection Search of Taiwan.
- Asia Depicted On Postcards**: Includes tags for Asia, postcards, world war 2, Imperial Japan. Description: "Asia Depicted on Postcards" is a collection of postcards related to Asia that a group of researchers led by Prof. Toshihiko Kishi of the Department of Political and Economic Coexistence, the Center for Southeast Asian Studies, Kyoto University (CEAS), have gathered since 2004. A
- Thai Language "The Three Seals Law (Kotmai Tra Sam Duang)" (Thammasat University Version)**: Includes tags for Thailand, Laws and Decrees, Ayutthaya, タイ, 法令・布告, and アユタヤ. Description: This is the Thammasat University's own compilation into a full text database, using photographic consultation, of laws, ordinances and edicts from the mid-14th century up until the beginning of the 19th century. The compilations is based on the manuscripts of the code of all exis
- Southern Part of the Northeast Thai Palm-Leaf Manuscripts Database**: Includes tags for Palm-Leaf Manuscripts, Khmer character, 貝葉, クメール文字. Description: This is a database of digital images of manuscripts on palm-leaves in old-local languages created by the local researcher Mr. Cheymongkol Chalernsukjitsri. He took high-resolution pictures of the palm-leaf documents at Buddhist temples in the Southern part of the Northeast Thaila
- Malaysian Movie Database**: Includes tags for movie, CD, DVD, Malaysia, Communist, 映画, マレーシア, 共産主義者. Description: A database of posters and catalogues of movies produced in Malaysia (in the video CD and DVD formats). In addition to about 900 feature films, the collection also includes films that are banned from screening in Malaysia, for example, The Last Communist (dir. Amir Muhammad, 2006)
- Qalam Article Database**: Includes tags for Muslim Brotherhood, Muslim, Jaui, Magazine, ムスリム同徳団, ムスリム, ジャワイ, 雑誌. Description: A database of all articles of Qalam, a monthly magazine on Islam and Muslim peoples and cultures, written in the Malay language and the Jawi script. Qalam was fi rst published in 1950 in Singapore and was widely read among Muslims in the Malay world until it ceased publication fo
- Topic Map Database of NDL Subject Headings**: Includes tags for NDL, subject index, 国立国会図書館, 件名標目表. Description: Subjects are keywords that describe primary them of materials. However, as there are many equivalent and resemble words (synonyms and quasi-synonyms) that express the same theme, information retrieval sometimes becomes diffi cult. Therefore, the subjects must be controlled under
- AGROVOC Topic Map**: Includes tags for AGROVIC, thesaurus, agriculture, 農畵, シソーラス. Description: AGROVOC is a multilingual thesaurus that covers vocabularies that are relevant to agriculture, forestry and fi sheries. The thesaurus contains approximately 40,000 entries in 20 languages. AGROVOC arranges the vocabularies in relation to synonyms, broader and narrower terms, rela
- "BOYS over FLOWERS" topic map**: Includes tags for manga, Boys over Flowers, マンガ, 漫画, 花より男子. Description: Japanese comic books (manga) have been widely read not only in Japan but also in abroad, and applied into various media from weekly magazines to monographs, animations, movies, internet books, and mobile-phone apps, that is, manga are considered to have already established a cult
- Japanese Historical Gazetteer**: Includes tags for Japan, Historical Gazetteer, 日本, 地名辞書. Description: To visualize and analyze data on a map, latitude and longitude information is often necessary. In recent years, geocoding data services to obtain latitude and longitude information from a place name have been prepared, and various data can be easily handled on a map. However, the

RDMにかかる現状（地域研究） 1

- RDMに関して地域研究分野でコンセンサスとなっている事項は？
 - 全ての研究者は少なくとも独自の方法でデータの整理・保存を行っている
 - しかし、地域研究は境界領域あるいは複合領域であるためコンセンサスを得ることは容易ではない
 - データを管理・保存することの重要性やデータ公開の有用性については誰しも合意するが・・・
 - データ公開に消極的な分野もある
 - 論文の根拠となる基礎データを公開しにくい（したくない？）雰囲気（？）がある
 - 査読システムがない（少ない）分野もある
 - データから結論に至る過程の再現あるいは追試が必ずしも必須・可能ではない分野もある
- 若手研究者（特に、大学院生）はどのようにRDMスキルを学んでいるか？
 - 研究所としては指導等を行っていない（必要に応じて指導教員が個別に指導することはある）

RDMにかかる現状（地域研究） 2

- 査読付論文発表時にデータ添付を求められているか？
 - 医学・自然科学系では求められた事例がある（一般的になりつつあるらしい）
 - 投稿時は必要に応じてデータを添付する
 - 論文発表時は全データを載せることはできないのでappendixとして掲載することもある。全データを参照するのであればサーバ機能が必要となるであろう
- データは自身のパソコン，外付けディスク，Googleドライブ等に保存している
 - 長期保存や利活用を考えると，今後はサーバー等置く必要がある
 - 長期保存の仕掛けがない（メタデータ，管理者，予算等）
 - 東南アジア地域研究研究所の場合はMyデータベースの利用も可能

RDMにかかる現状（地域研究） 3

～ CDL RDM RUBRIC に照らして当該分野の状況はどうなっているか～

	その都度	1度のみ	有効に活用	再利用のために最適化
データマネジメント計画	データを入手した段階で、自分なりに考える。標準化や文書化はしていない	最初に計画を作成するが、通常は、途中で見直すことはしない	詳細な計画を作成し、プロジェクト期間を通じ参照と見直しを行う	自分あるいは他人が将来スムーズにデータを利用できるように計画している
データの体系的整理	一貫した手法に従っていない。そのため、しばしば探し出すのに時間を要する	手法はあるが、それを適用するのは、プロジェクト終了時のみ	先を見越した手法を導入しているが、必ずしも標準化されていない	他人が自分でデータを検索、内容を理解し、利用できる
データの保存とバックアップ	作業中にどのデータが重要であるかを決め、通常はこれを特定の1か所に保存する	どのデータを保存すべきかを理解している。損失のリスクを緩和するため、作業後にバックアップをとる	作業中でも定期的にデータを保存するシステムを用いている。複数のバックアップを保持している	再利用する機会を最大限生かせるよう考慮した方法と場所にデータを保存している。
データを解析可能な状態にする	標準化もしくは十分に文書化された手続きはない	解析に適した形式を検討し整備するが、事例ごとに異なる方法をとっている。	データ処理方法は標準化され、かつ十分に文書化されている	将来も自分自身及び第3者が利用しやすい方法で、データ処理を行っている
データ解析と成果の取扱い	解析や実験がどのような手順やパラメータにより実施されたか確認するために、これらをやり直すことが多々ある	解析作業終了後、パラメータの詳細や解析手順、プロトコルを文書化している	データ解析の際、決められた方法で解析のワークフローや判断のプロセスの詳細をいつも記録している	自身が行った解析のワークフローや判断プロセスの詳細が他人によっても実行できることが確認できている
データの共有と公開	研究成果を共有しているが、その根拠となるデータについては共有できていない	リクエストがあった場合にのみ研究データの共有を行っている	自身の研究成果や根拠となるデータは、第三者が利用できる形式で共有できる	優れたデータマネジメントを実践し、自身のデータをいつでもだれとでも効率的に共有できる

•この表は医療系・自然科学系の教員に尋ねた事例

•人文社会学研究にこの尺度を適用できるのか検討が必要

RDMにかかる現状（地域研究） 4

～ RDM に関して当該分野の現状・課題は？～

- 現状は研究データの散逸・損失を防ぐ手立てで精一杯
 - デジタル化支援, メタデータ構築支援等
- 「地域研究」で括ることは不可能ではないか？
 - 分野によりRDMに対する理解や認識の差が大きい
 - CDL RDM Rubric は人文社会分野に適用できるか？
- 多様性が高い
 - 研究データにはモノが含まれることもある
 - 個別データではなくアーカイブになるケースもある（高名な研究者の全史資料等）
 - データをアーカイブする際の基準等が定まっていない（何を残し何を棄てるのか等）

RDMにかかる現状（地域研究） 5

～ RDM に関して当該分野の現状・課題は？～

- 解析の過程を残せるか（自然科学系より）？
 - 試行錯誤の過程は？
 - データ処理のプログラムを保存していれば、あるていど解析過程を辿ることもできるか？
- 入れ物（ディスク）だけでは不十分か？
 - 研究データの長期保存・長期利用（将来にわたり他人によるデータ利用）を可能とするには、どのような付帯情報が必要か
 - 母集団, 計測装置, 手技, データ記述ルール（単位系）等
 - 装置のマニュアル等
 - 底本, 翻刻の手順, 記述スキーマ, 異体字の同定ルール等
 - 手順書等
 - レポジトリ（あるいはRDMシステム）に研究データをアップしてくれても・・・
 - この付帯データを作成・入力してくれるか

RDMにかかる現状（地域研究） 6

～ RDM に関して当該分野の現状・課題は？～

- 研究所としてのRDMポリシーは未策定
 - 理解の促進, 合意形成, 具体化作業をどのように進めるのか
 - 支援が必要 (URA?)
- RDMを実施する研究所内の体制が未定
 - 研究所内における担当 (者) : 情報システム, ポリシーの実装・実施支援, 予算等
 - 広報委員会?, 図書委員会?, 情報整備委員会?, 情報処理室?, グローカル情報ネットワーク?
 - 研究者, 研究所, 大学における役割・責任等の分界点
- 共同利用・共同研究拠点としての機能をどう果たすか