

RDA東京大会参加報告

「語解」と「誤解」

—理解の透明な壁—

東京大学 地球観測データ統融合連携研究機構
/空間情報科学研究センター 柴崎研究室

小野 雅史

自己紹介: 専門

- 専門は「空間情報科学」
- 様々な分野と接する、学際性が特徴
- 従って、他分野の支援を目標の一つとする

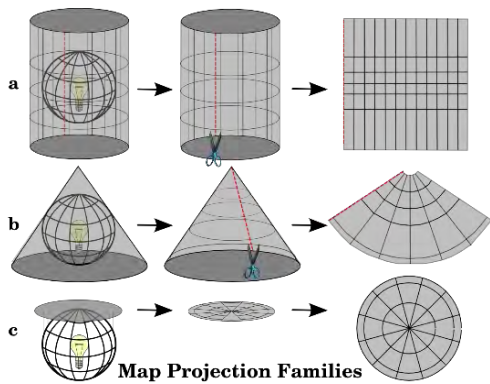
空間情報科学

座標系

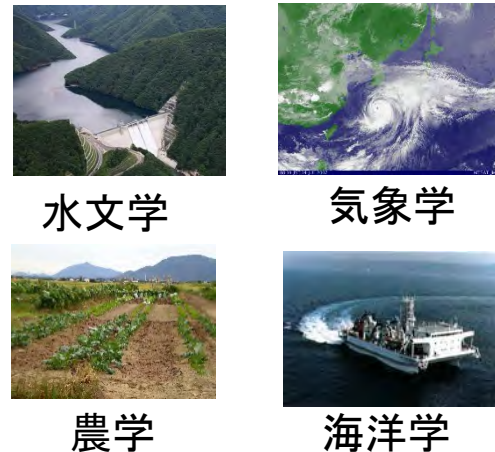
地理情報

地域分析の手法

デジタル地図、
ナビゲーション技術



基礎科学



自然科学

社会動態



人文・社会科学



情報学・工学

自己紹介: 過去の活動からRDAへ

過去の活動

- **ISO/TC211**
 - 地理情報の国際標準
 - **メタデータ**, **場所識別子** (Place Identifier) などの検討
- **DIAS/GRENE-ei**
 - 地球環境分野を対象にした、**DIASメタデータ** (ISO準拠) の設計
 - **語彙管理システム** の開発を担当
- **GEO (Group on Earth Observation)**
 - 各国の**地球観測機関**を中心とした国際グループ
 - オントロジー、ベストプラクティス・レジストリ等のタスクチームに参加
- **Belmont Forum e-Infrastructure and Data Management**
 - 資本と人的リソースを国際連携するための枠組み
 - WP4「**Data Sharing**」グループに所属

参加機関の一つ

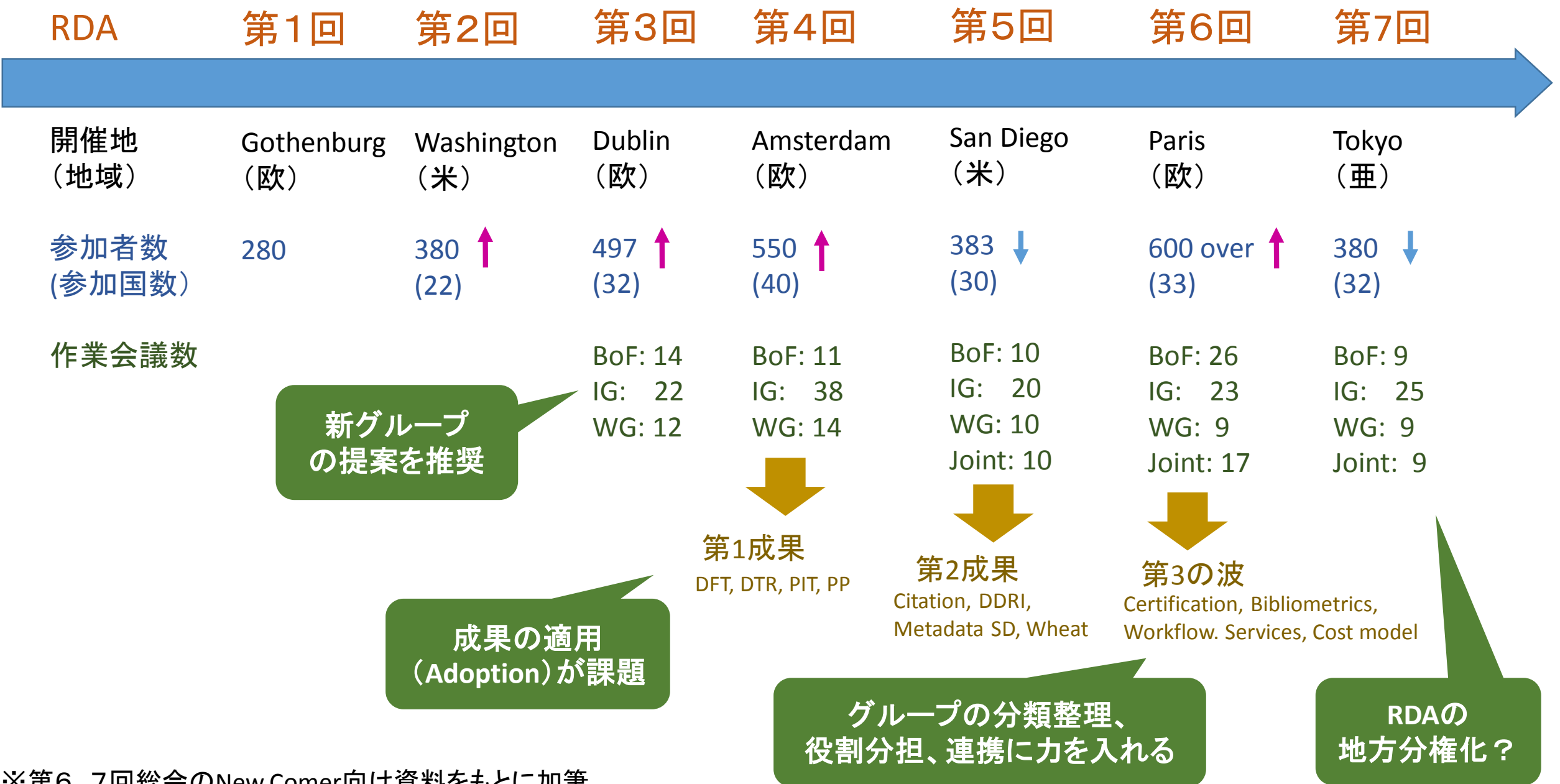
コア・メンバー
が重複

- 2014年: **第3回**、**第4回** 総会
- 2015年: ~~第5回~~、**第6回** 総会
- 2016年: **第7回** 総会

に参加



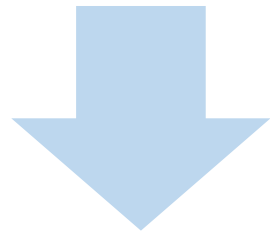
RDAの変遷



※第6、7回総会のNew Comer向け資料をもとに加筆

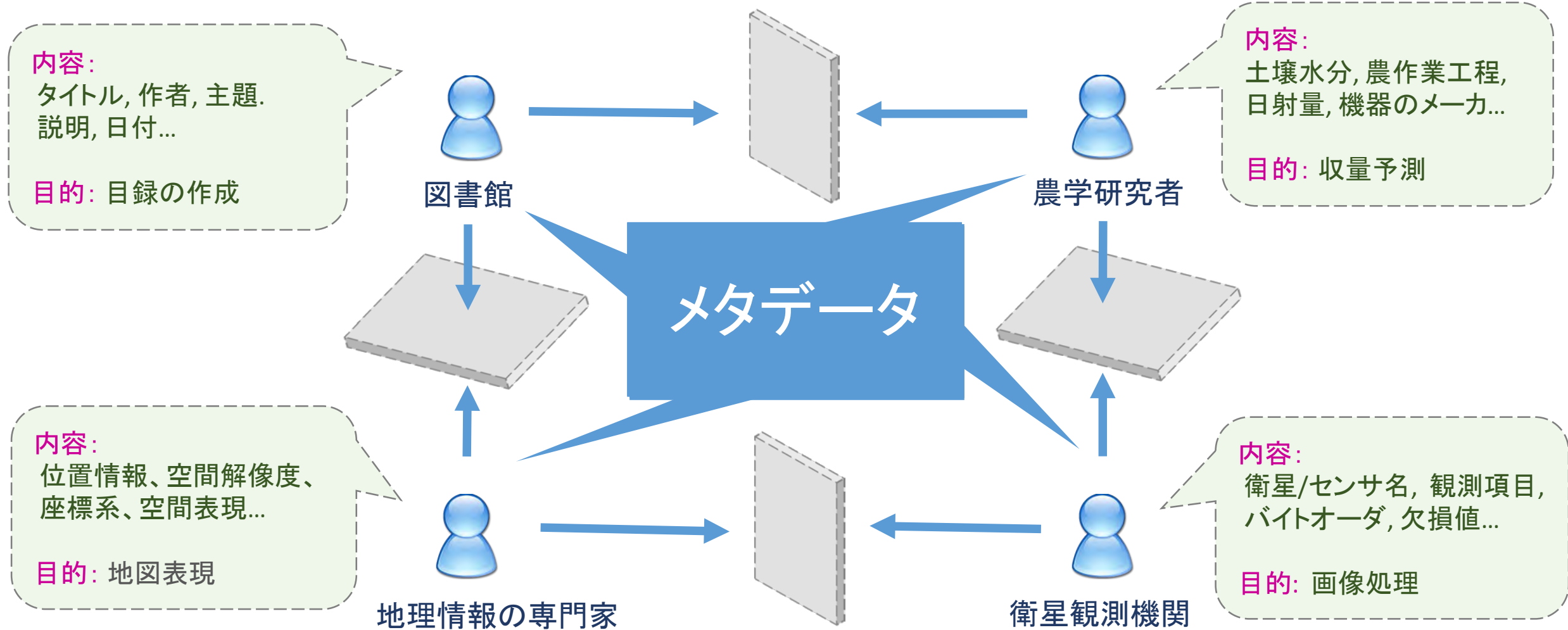
これまでの経験を振り返って

- (DIAS, GEO, RDAなど) 多様な人々が集まるプロジェクトに参加
- 背景が違くと、その人の言葉の意味が変わる



「語解」が「誤解」を生む状況が起こる

「語解」と「誤解」: 例えば、「メタデータ」の意味は？

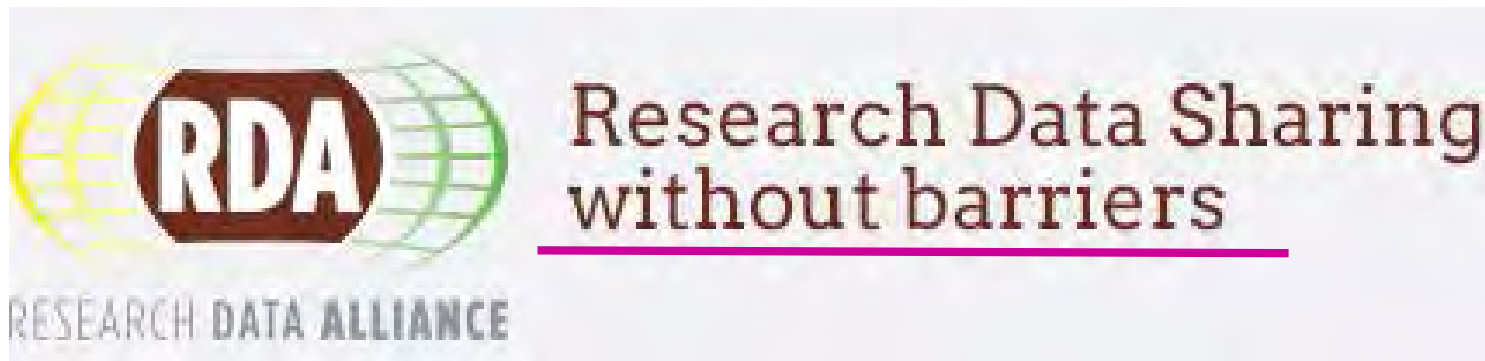


その人の背景によってイメージする内容が異なる。

よって、相手に意味が正確に伝わらない。

相互理解の間には、目に見えない壁がある

「理解の透明な壁」を超えるため、
Terminology/Vocabularyの整備が不可欠



RDA Working Meeting Session参加リスト(全7スロット)

- 3月1日:

- #1: IG Vocabulary Services: Community Use Cases for Vocabulary Services
- #2: WG QoS (Quality-of-service)-DataLC(Life-cycle)Definitions
- #3: IG Geospatial

- 3月2日:

- #4: IG Data Foundations and Terminology: expanding vocabulary coverage and services
- #5: IG Vocabulary Services: Access Methods Review of Existing Vocabulary Services

- 3月3日:

- #6: Joint meeting of IG Vocabulary Services, IG Data Foundations and Terminology:
- #7: BoF e-Infrastructure for Global Change Research
 - (↑Belmont Forum e-Infraメンバーが立ち上げたグループ)



”Terminology”と”Vocabulary”を扱う
2つのIGによる会合に集中参加

IG Data Foundations and Terminology (DFT)

- 概要:

- WGとして成果公開後、活動を継続しているグループ
- 参加者は、10数人程度

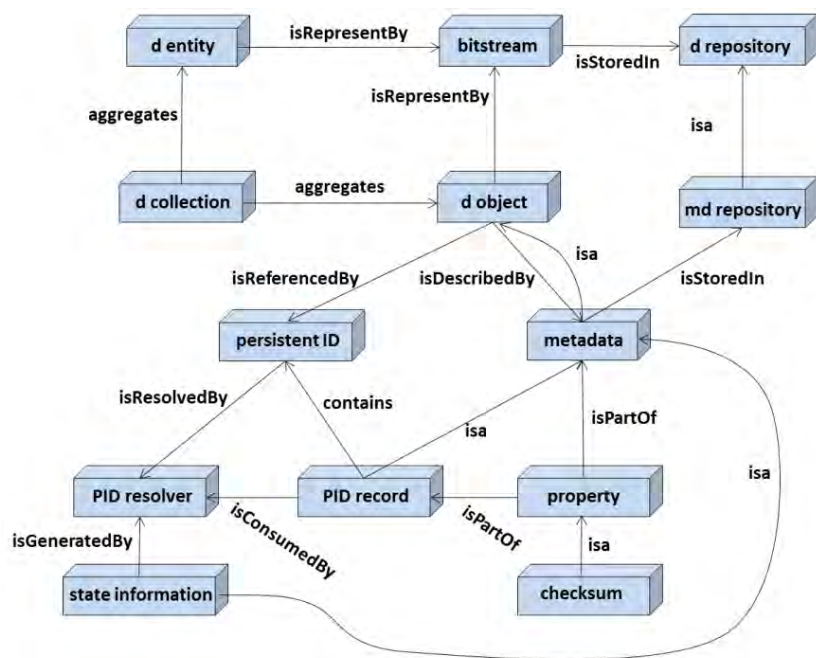
- 旧WGの成果:

1. 分野横断的にデジタルオブジェクト(DO)を扱うための、概念モデル
2. 中核となる用語集
3. Semantic MediaWikiによる語彙管理ツール

- 新IGの目的:

- RDAの他グループと連携し、
 - 概念モデルの拡張
 - 新規用語の追加
 - 各コミュニティに対して語彙の理解/利用を支援

RDA WG DFTの成果



1. 概念モデル

1. DFT Core Term Definitions

1.1 Digital Object (DO)

Definition

A digital object (DO) is represented by a bitstream, is referenced and identified² by a persistent identifier and has properties that are described by metadata.

Note: As indicated we only talk about registered DOs in the context of this document.

Note: Properties included in metadata include discovery, contextual, schema, rights, curation and provenance information.

Note: A DO is said to be a dynamic DO when the information content represented in a DO is changing for some period of time or even for an indefinite duration.

1.2 Persistent Identifier (PID)

Definition

A persistent identifier is a long-lasting ID represented by a string that uniquely identifies a DO and that is intended to be persistently resolved to meaningful state information about the identified DO³.

Note: We use the term Persistent Resolvable Identifier as a synonym.

2. 用語集

The screenshot shows the RDA Metadata page. It includes a navigation bar with links like 'Page', 'Discussion', 'Read', 'View form', 'View source', and 'View history'. The main content area is titled 'Metadata' and contains a table with definitions, explanations, examples, references, scope, and status for the term 'Metadata'. The table is structured as follows:

Term	Definition	Explanation	Examples	References	Scope	Status
Metadata	Metadata is data that plays the role (is used for) data/resource discovery, description/documentation, contextualisation. metadata provides contextual information on for data collections to re-use and re-purpose data.	Data/Resource discovery allows resources to be found by relevant criteria. Identifying resources; Bringing similar resources together; Distinguishing dissimilar resources; Giving location information.	Examples include data about related datasets (including provenance metadata), software, publications, organisations, persons (such as organizer of the data). Typically descriptive metadata includes such things as source & time of creation. For data and report publication it may include administrative metadata such as authors & date of submission. A PID is an example of metadata used to reference data. An example is retention period metadata which defines the date when retention of the data object should be evaluated.	RDA MD and PP WG discussions.	RDA Term Collection Core	New
Metadata	Metadata is a type of data object that contains attributes describing properties of an associated data or digital object. "Metadata" represent the set of instructions or documentation that describe the content, context, quality, structure, and accessibility of a data set." Mitchner (2006)	MD can be used for Discovery, Access, Selection, Licensing, authorization, Quality, usability and Provenance; reproducibility.	It may contain as key the persistent identifier of that associated object. The association between a data object and metadata is that the content of the metadata describes the data object. Metadata may serve different purposes, such as help people to find data of relevance - discovery or to bring data together - federation.	Data properties, both internal and external, are types of metadata as is transactional information about data. Mitchner, William K. "Meta-information concepts for ecological data management." Ecological Informatics 1.1 (2006): 2-7. http://www.esajournals.org/doi/pdf/10.1890/1551-0751(1987)0017620[0000MIFTE]2.0.CO;2	RDA Term Collection Core	In discussion

3. 語彙管理ツール

IG Vocabulary Services

- 概要:

- 比較的新しいIG
- 参加者は10数人程度で、DFTメンバーとの重複が多い

- 本IGの目的:

- 各研究コミュニティが統制語彙を開発するためのレコメンデーションの作成
- ソフトウェア工学的な視点が中心

- 主な議題:

- 既存の語彙リソース/サービスの情報収集と調査
- 何を調査する?
 - ユースケース
 - システム要件を分析
 - 共通要素を整理
 - 既存サービスの仕様/アクセス方法など

調査用資料

Use Case: #####
Point of Contact: ##### <#####@#####>
Version:
Date:

Use Case Name <i>A name that succinctly states the user task in the form of "verb + object" such as "Place and Order".</i>
Goal <i>The goal briefly describes what the user intends to achieve with this use case.</i>
Summary <i>Give a summary of the use case to capture the essence of the use case (no longer than a page). It provides a quick overview and includes the goal and principal actor.</i>
Actors <i>List actors, people or things outside the system that either acts on the system (primary actors) or is acted on by the system (secondary actors). Primary actors are ones that invoke the use case and benefit from the result. Identify sensors, models, portals and relevant data resources. Identify the primary actor and briefly describe role.</i>
Preconditions <i>Here we state any assumptions about the state of the system that must be met for the trigger (below) to initiate the use case. Any assumptions about other systems can also be stated here, for example, weather conditions. List all preconditions.</i>
Post Conditions <i>Here we give any conditions that will be true of the state of the system after the use case has been completed.</i>
Triggers <i>Here we describe in detail the event or events that brings about the execution of this use case. Triggers can be external, temporal, or internal. They can be single events or when a set of conditions are met. List all triggers and relationships.</i>

	A
1	Name of Software
2	URL
3	Documentation URL
4	Open Source
5	Based on SKOS
6	Uses ISO 25964
7	Supports OWL ontologies
8	Multiple languages
9	Multiple vocabularies
10	mapping/matching vocabulari
11	SPARQL endpoint
12	API for browse
13	API for searching vocab
14	API for tagging (auto-indexing
15	API for searching content
16	Server installation
17	Cloud version
18	Editing tool
19	Workflow
20	Export formats

1.ユースケース調査用のテンプレート

アクター、条件、イベントトリガーなどの項目

内容をFixしてから調査開始

2.既存サービス/ソフトウェアに対する調査項目

オープン/商用、対応フォーマット、
オントロジー/SPARQLサポート、など

調査・情報収集中

RDA参加報告：まとめ

- 研究データ共有には、多様性がある
 - 様々なステークホルダー、分野の専門家が関わる
- つまり、自分のメッセージが正しく相手に届くとは限らない
 - 「語解」と「誤解」問題に注意
- 従って、Terminology/Vocabularyの整備は重要。今回は、この意識を持ってRDAに参加。
 - 「理解の透明な壁」を超えるため
- 日本の場合には特に、RDAの議論を日本語に翻訳する際にも注意が必要
 - 単純な直訳はミスリードを誘発する可能性あり
- 現場経験を積むことも大事
 - 実際の議論を通して、どういう文脈でその表現が使われるか、効率よく学習できる

RDA東京大会後...これからの話

- 自分の役割
- 図書館への期待

OECDレポート意識(tentative) : オープンサイエンス (OSci) 推進の役割

1. 研究コミュニティ

- Osciの主体
- 「OSciに関する知識・理解の促進」に貢献
- ただし、課題はインセンティブ
 - 従来の研究活動から離れた「無駄な仕事」になる可能性があるため

2. 省・政府

- OSci推進イニシアチブを組織し、それを評価
- 国家レベルのアクションプランの作成
- 規制およびインセンティブの枠組みを定義

3. ファンディング・エージェンシー

- (データ管理計画などの) 研究成果に対するメカニズム・要件を定義し、OSciを促進
- オープンアクセスリポジトリへの投資

4. 大学・公的研究機関

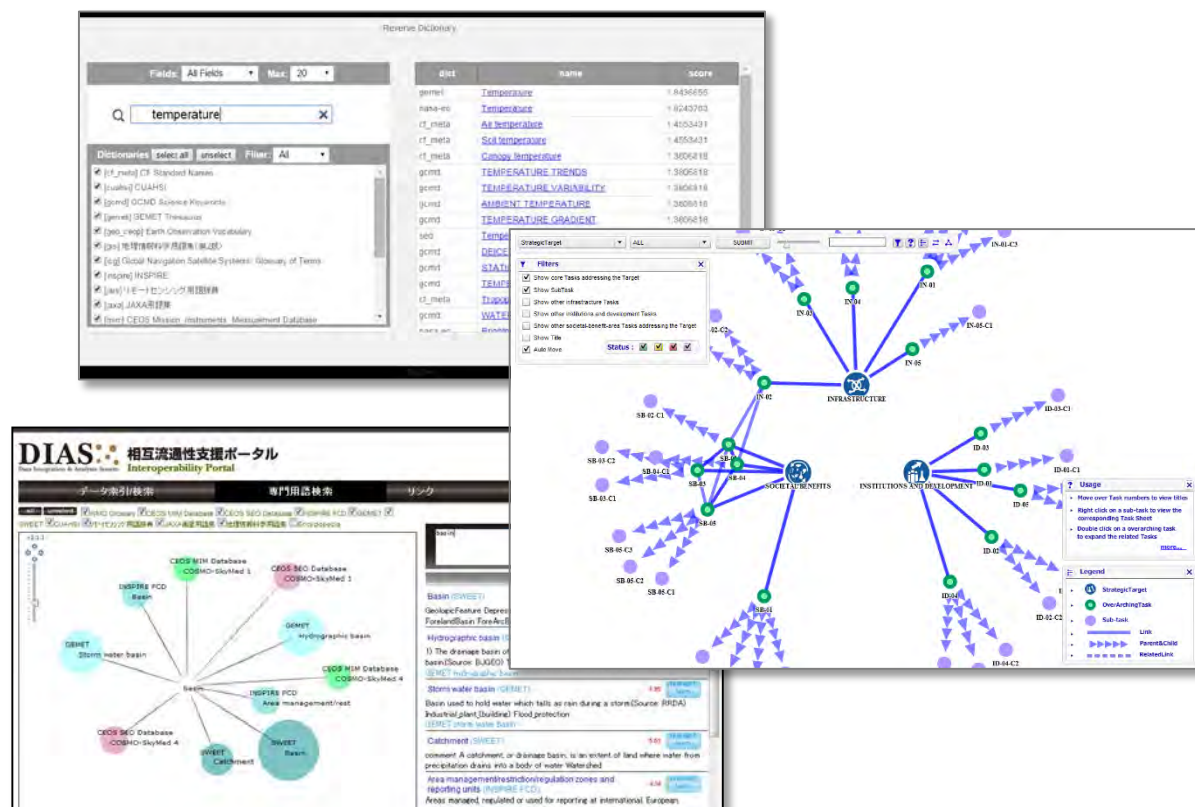
- 大学・研究機関が独自に有するポリシーや開発基盤(リポジトリ、プラットフォーム)をOSci対応に移行
- OSciに関する知識・技術を有する学生・研究者の育成
 - 特に、データ管理システム、キュレーション、品質管理の実装スキル

5. 図書館・リポジトリ

- 科学資料(論文、データなど)の保存・キュレーション・出版・普及に関して、新たな役割を持つ
- グリーンオープンアクセス(セルフアーカイブ)に関する重要な役割
- 基盤およびシステム間相互流通の開発・維持のための、持続的な投資が課題

自分の役割1 : Vocabulary Service支援

- DIASで、地球環境分野向け語彙管理システムを開発
- これを、オープンサイエンス対応に拡張



開発したソフトウェア

The image shows a screenshot of the DIAS website. The main heading is 'Temperature'. Below it, there is a definition: 'A property that determines the direction of heat flow when an object is brought into thermal contact with other objects: heat flows from regions of higher to those of lower temperatures. (Source: PITT)'. There is also a section for 'relations' and 'Physical property'. The page includes navigation links like 'Main page', 'Recent changes', 'Random page', and 'Help'. At the bottom, there are logos for 'Powered by MediaWiki' and 'Content by Commons MediaWiki'.

No.	名前	提供機関など
1-	CF Standard Standardized Region Name	CF Community
2-	CF Standard Area Type Table	CF Community
3-	Glossary of Oceanography and the Related Geosciences with References	Texas A&M University
4-	GCMD (Global Change Master Directory) Science Keywords	NASA
5-	地理情報科学用語集 (第2版)	地理情報システム学会
6-	CF Standard Names	CF Community
7-	NASA Earth Observatory	NASA
8-	Global Navigation Satellite Systems: Glossary of Terms	NASA
9-	Marine Metadata Interoperability	Harte Research Institute, Texas A&M University
10-	WMO Glossary	WMO
11-	CEOS Missions, Instruments and Measurements (MIM) Database	CEOS
12-	CEOS Systems Engineering Office (SEO)	CEOS
13-	GEMET	ETC/CDS (European Topic Centre on Catalogue of Data Sources)
14-	INSPIRE Feature Concept Dictionary	EU Joint Research Center
15-	SWEET Ontology	NASA
16-	CUASHI Ontology	CUASHI
17-	International Glossary of Hydrology/UNESCO	UNESCO
18-	JAXA 用語・解説集	JAXA
19-	多国語防災用語集	ADRC
20-	Earth Observation Vocabulary (GEO)	GEO, 東京大学
21-	Glossary of biotechnology for food and agriculture	FAO
22-	Water Science Glossary of Terms	USGS (アメリカ地質調査所)

収集したVocabulary一覧

各用語の定義の閲覧環境

自分の役割2: 研究コミュニティの現場の実態調査

- RDAのポスターセッションで、DIASで実施した調査の一部を発表。これをフォローアップ
⇒ ミスリードを避けるため、エビデンスづくり
⇒ オープンサイエンスの実効性(本当に良い方向に向かっている?)を評価

Comparison between Japanese and International results

- Here shows the comparison between two survey results: **this survey** (Fig.4) and **DataOne's survey** (Fig.5) published as an article "Data Sharing by Scientists: Practices and Perceptions" (Tenopir, 2011 ※).
- We refer to a series of questions about the **requirements on data sharing consensus** from both viewpoints of **data provider and user**.
- We summarize and classify the original question in the DataOne's survey into **4 group**: Contact, Publication, Agreement and Others.
- By comparing both, we can find several **mind gaps** between Japanese and International Community.

(※) doi:10.1371/journal.pone.0021101

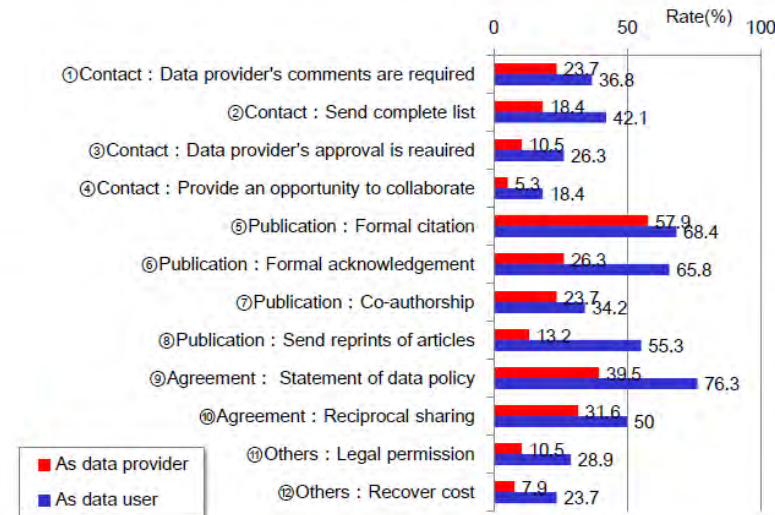


Fig.4: Data sharing requirements (this survey, Japan)

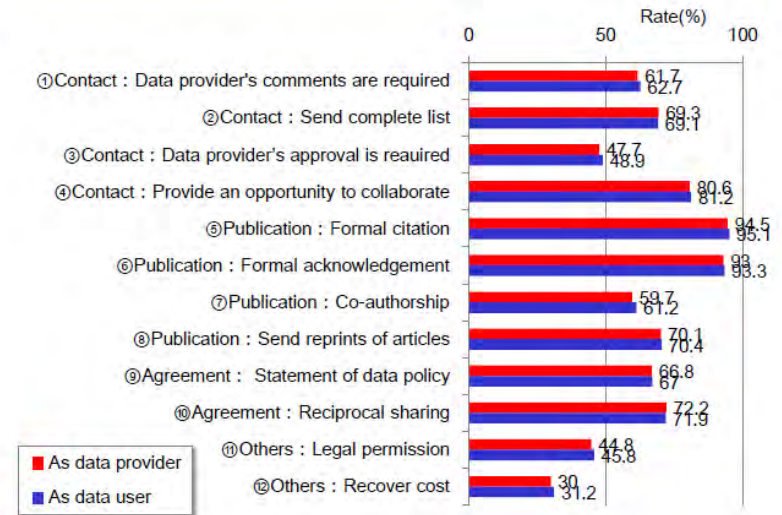


Fig.5: Data sharing requirements(DataOne, international)

(個人的な)図書館への期待

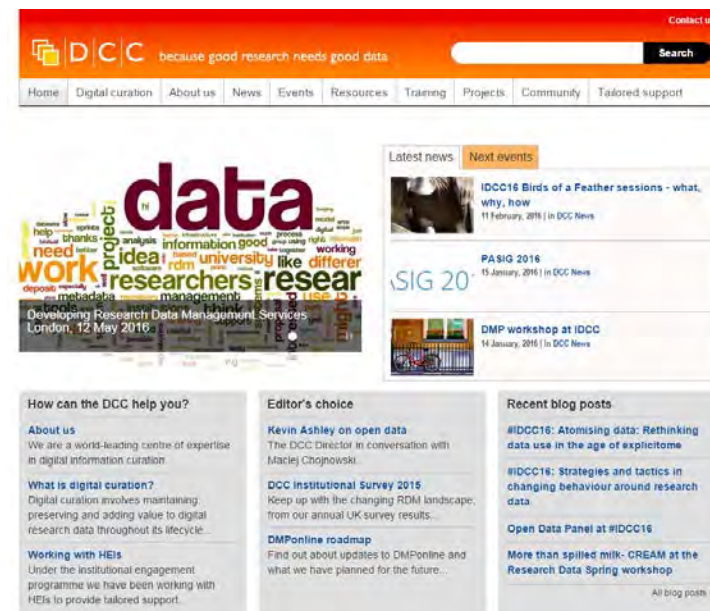
新しい役割に期待



- データ保存より、**キュレーション**
- オープンサイエンスに関する**メタ情報管理**
 - イベント情報
 - データジャーナルの情報
 - 分野毎のリポジトリの情報など



要するに、
日本版DCC(Digital Curation Center)を希望



ご清聴ありがとうございました。