

研究データへのDOI登録実験プロジェクト - プロジェクトの概要・進捗状況 -

国立情報学研究所 武田英明

(ジャパンリンクセンター運営委員長)
(「研究データへのDOI登録実験プロジェクト」リーダー)

科学技術振興機構 中島律子 (ジャパンリンクセンター事務局)

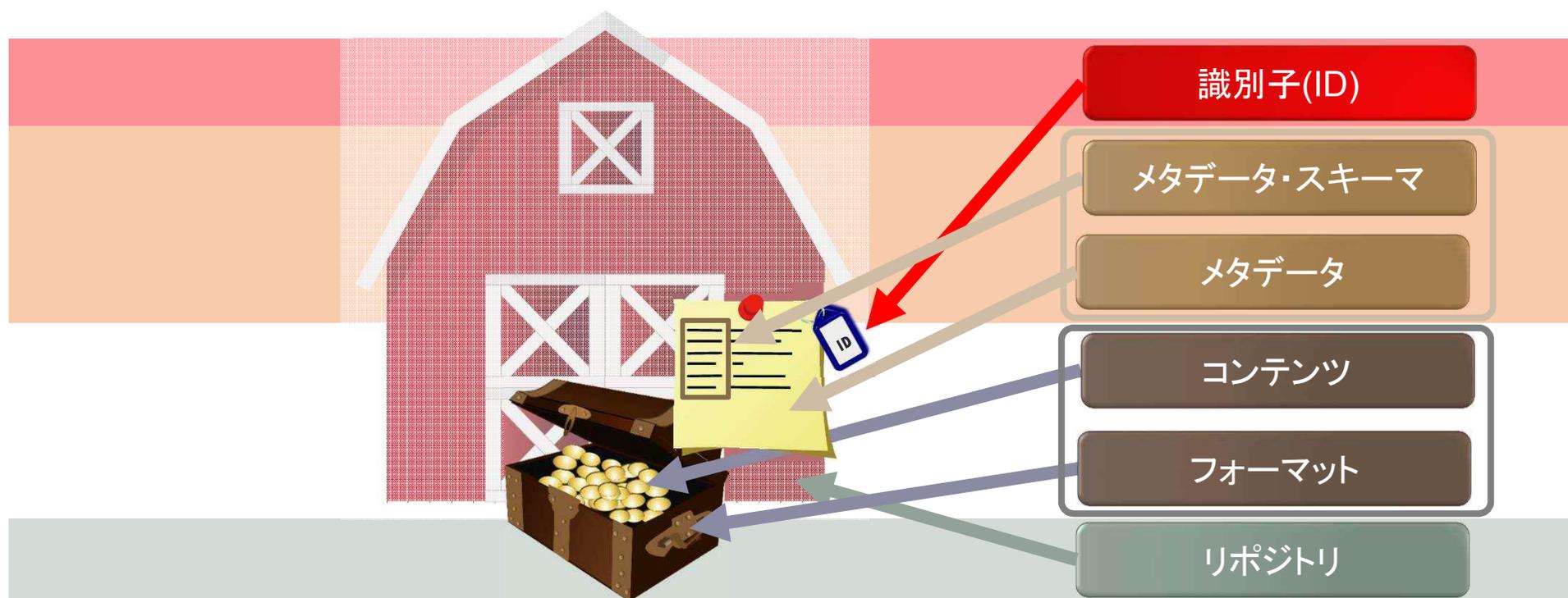


<http://japanlinkcenter.org/>

プロジェクトの背景

- JaLCにおけるDOI登録対象コンテンツの拡大
 - ・ 従来は主にジャーナルアークティクルにDOIを登録
 - ・ 平成26年12月に新システムリリース
 - 研究データ、書籍、報告書、e-learning 等へ対象を拡大
- 研究データへのDOI登録
 - ・ 研究データの公表や再利用への関心の高まり
 - ・ DOI登録における研究データ特有の課題検討の必要性
 - 研究データへのDOI登録実験プロジェクト実施を決定
(平成26年度第1回運営委員会)

研究データ流通におけるDOIの位置付け



研究データ流通におけるDOIの位置付け



プロジェクトの目的

●目的

- 日本における研究データへのDOI登録の仕組みを本プロジェクトへの参加機関とともに新たに構築する
- 研究データへのDOIの持続性、対象の粒度、DOIの活用などを検討する
- 運用方法の検討やDOI登録のテストを実施し、今後の安定運用につなげる

●ゴール

- 研究データへのDOI登録のポリシーを策定する
- 新JaLCシステムにおいて適切な運用フローを確立する。フロー確立に当たっては、DOI登録のテストを行って確認をとる

実施にあたっての方針等(1)

●参加機関

- ・参加機関は公募によって選定
- ・ただし、公募期間後も参加要望があれば、随時参加機関を追加（運用上問題がない範囲で）

●参加機関の公募要件

- ① 研究データに対するDOI登録を希望し、本プロジェクトの検討に積極的に参加していただける機関
- ② JaLC会員、もしくは本プロジェクト終了後にJaLC会員となる意思を有し、ジャパンリンクセンター参加規約におけるJaLC会員の資格を満たす機関

・・・ DOIの持続性を担保するための要件

実施にあたっての方針等(2)

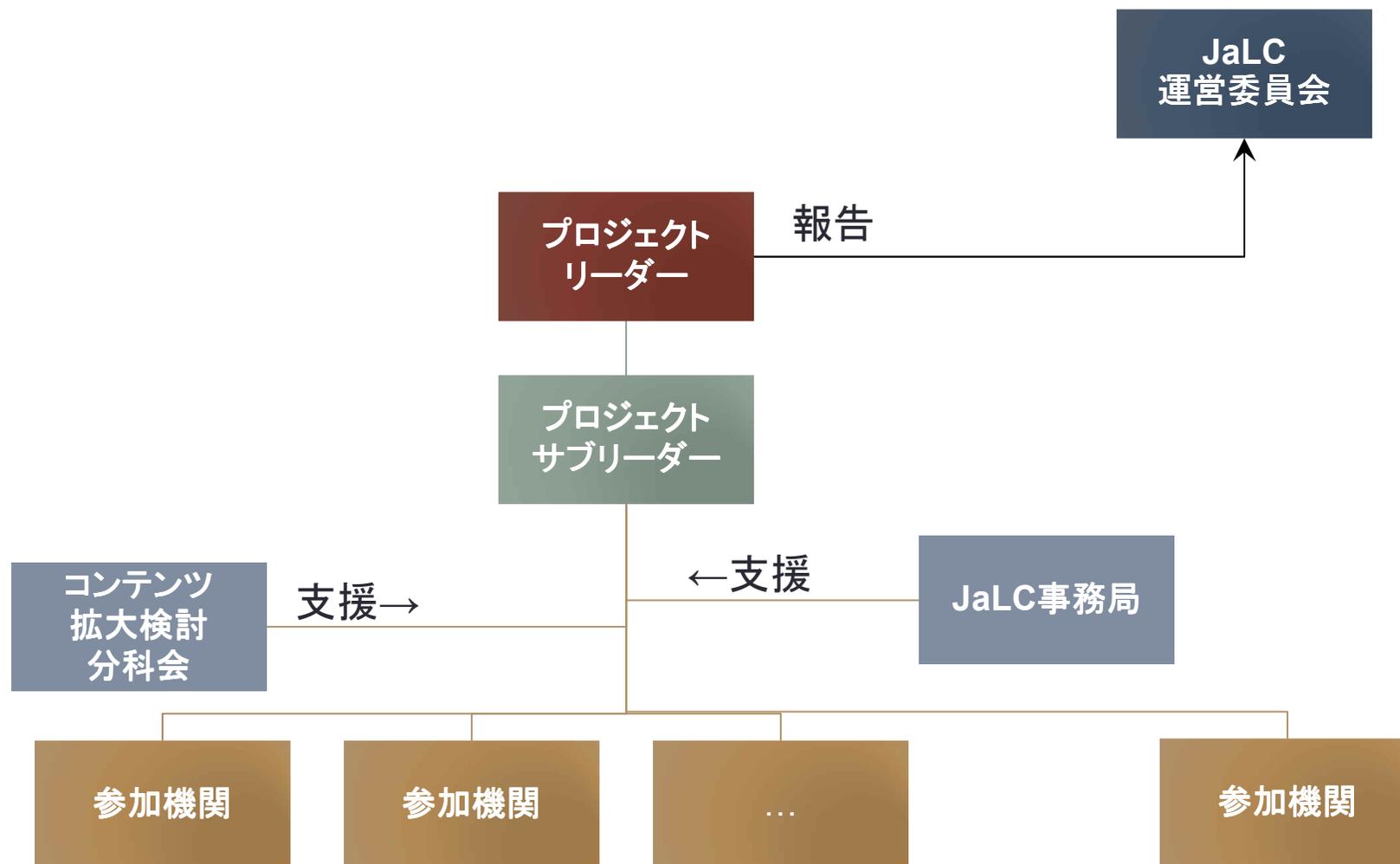
●DOI登録対象

- 研究データ(実験データ、観測データ、調査データ)
- JaLC DOI / DataCite DOI(JaLCが仲介)

●成果物

- 研究データへのDOI登録実験プロジェクト計画書
- 研究データへのDOI登録実験プロジェクト報告書
- JaLCにおける研究データの登録ポリシー
- JaLCにおける研究データの登録手順書

プロジェクト体制



※ 参加機関は、公募により決定した研究機関、大学、図書館、等。

実施内容

- 方式検討

- ・ 研究データに対するDOI登録の方式検討

- DOI登録テスト

- ・ JaLCへのデータ登録、更新、削除等のテスト
- ・ ※テスト用のDOIを使用、テスト終了後は削除

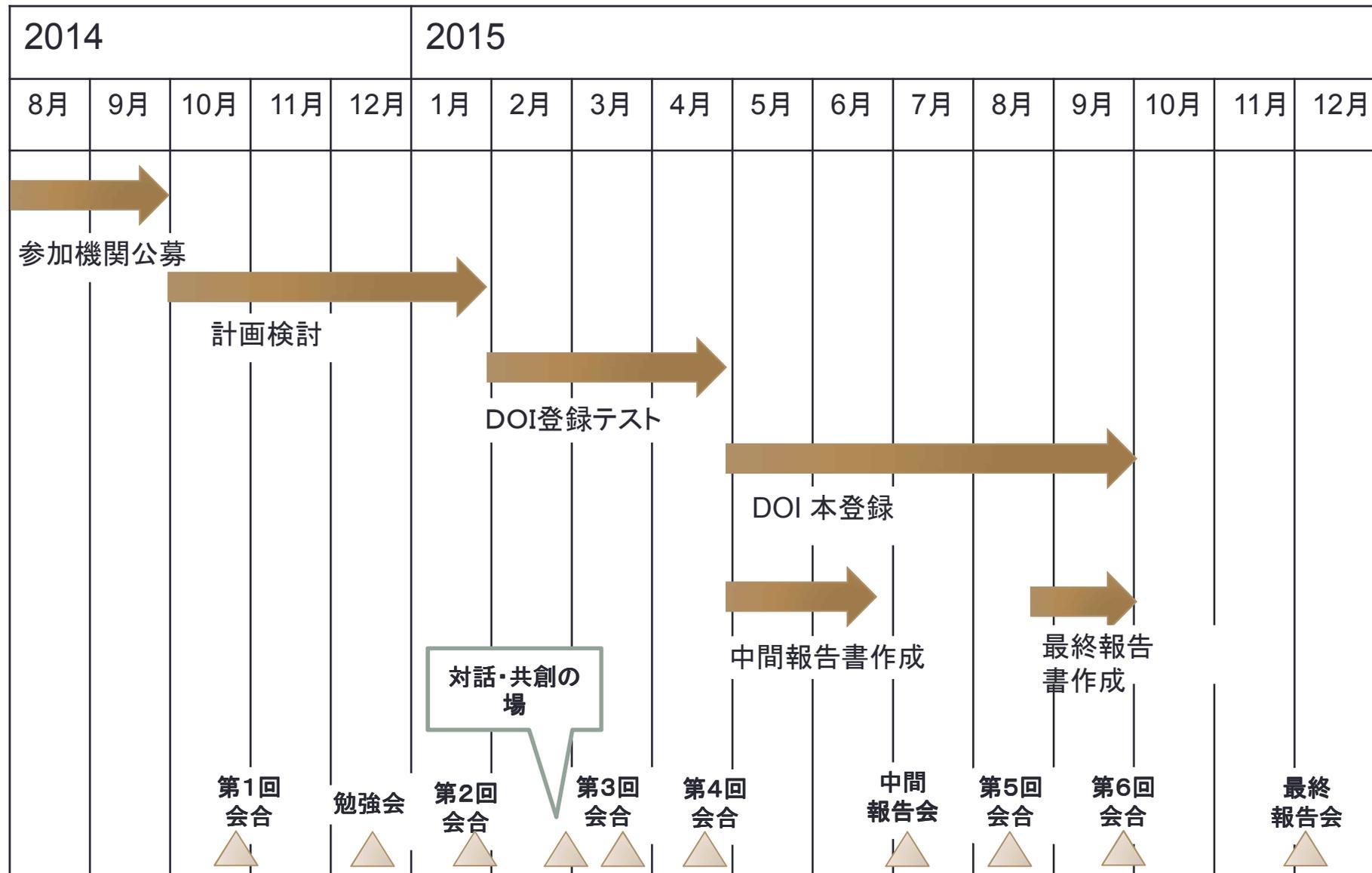
- DOI本番登録

- ・ 報告

- ・ プロジェクトの中間報告、最終報告書作成

参加機関は原則として研究データに対するDOI登録を実施
(必須ではない)

スケジュール

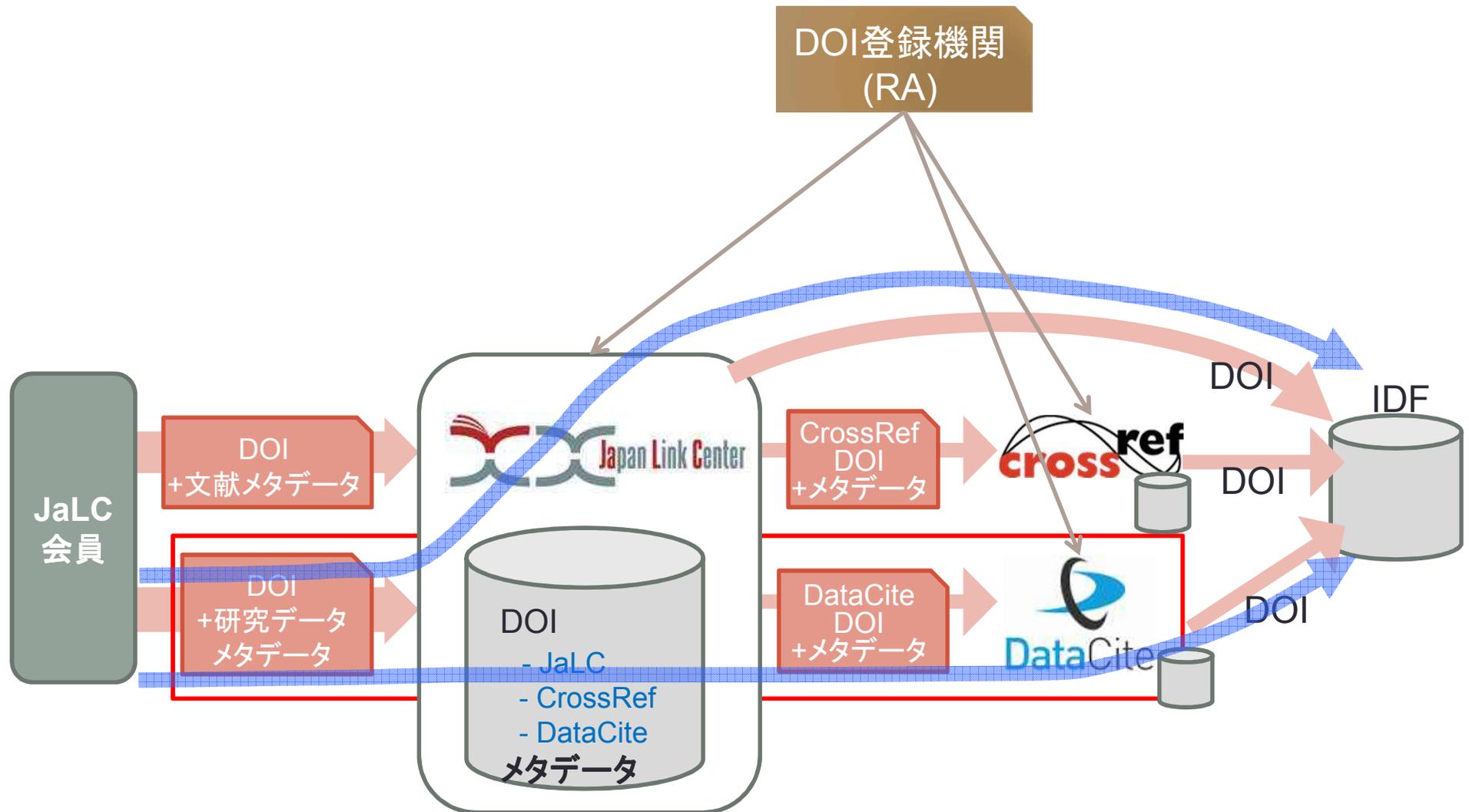


参加機関

リーダー 武田英明 (国立情報学研究所教授)
 サブリーダー 村山泰啓 (情報通信研究機構 室長)

機関	DOI登録対象とするデータの概要
科学技術振興機構	生命科学分野
国立極地研究所	地球科学・環境分野 (北極域、南極域における実験及び観測データ)
国立情報学研究所 (地球環境情報統融合プログラム(DIAS-P))	地球科学データ(地上観測データ,衛星観測データ、 気象予測モデル,気候変動予測モデル、その他社会データ)
国立情報学研究所	機関リポジトリの保有する研究データ等 (テキスト、画像、音声)
産業技術総合研究所	材料系
情報通信研究機構	超高層物理学分野(地磁気データ、電離層データ、オーロラ画像データ、衛星データ)
千葉大学附属図書館	一般的な研究データ、外部資金の助成による研究成果、 観測データ等のアーカイブ
物質・材料研究機構	<計画変更により再検討中>
理化学研究所 脳科学総合研究センター 神経情報基盤センター	脳・神経科学分野

データDOIの登録フロー



メタデータ

● 定義にあたっての考え方

- 原則として、DataCiteの定義にあわせる
(Metadata Schema 3.0。現在3.1への対応作業を行っているところ)
- ただし、可能な範囲で、JaLCで対象としている他のデータ種別
(ジャーナルアーティクル等)のメタデータ定義とそろえる
 - … タグ名の付け方、複数言語の持ち方、
ファンド情報やマルチプルレゾリューションの項目追加
- 一部JaLC独自のタグを導入(「signature」)

主なメタデータ項目

項目	参考:項目(英語)	説明
DOI	DOI	DOI
URL	URL	URL
タイトル	Title	データのタイトル
主題	Subject	主題、キーワード、分類、等
作成者	Creator	データ作成者の名前
所属機関	Affiliation	データ作成者の所属機関
研究者ID	Researcher ID	ORCID、e-Rad、科研費、等の番号
出版年月日	Publication date	研究データの出版・発行年
出版者	Publisher	研究データの出版者(研究機関等)
寄与者	Contributor	データ管理者、プロジェクトマネージャ、研究資金配分機関、等
版	Edition	Variation(出版版、著者版、等) Version(1.0、2.1等) Format(ファイルフォーマット)

項目	参考:項目(英語)	説明
関連コンテンツ	Related content	DOI、URL、relation type(引用、補足、別バージョンの存在、等)
代替識別子	Alternate ID	ローカルDBのIDなど
言語	Content language	コンテンツの言語
日付	Date	受理日、収集日、更新日、等
リソース種別	Resource type	データセット、ソフトウェア、等
権利	Rights	Creative Commons 等
追加情報	Description	抄録、研究方法、シリーズ情報、等
署名	Signature	データの品質保証等(JaLC独自)
位置情報	Geolocation	地域、場所の名前、緯度経度
研究資金情報	Fund	研究資金提供機関、課題番号、等
マルチプルレゾリューション優先度	Multiple resolution priority	ミラーサイトやアーカイブサイトがある場合のランディングページ優先度

検討課題

- ポリシー策定のために、検討課題ごとに論点出しを実施中
- 今後、DOIテスト登録・本登録作業と並行して、検討を進めていく予定

「対話・共創の場」等を通して、プロジェクト外からもご意見を伺っていく

検討中の課題

- a) 運用フロー
 - ・ 取りまとめ機関の位置付け
 - ・ データのライフサイクルと担当者・担当機関
 - ・ DOI登録の対象データ
 - ・ Prefixの割り当て方針
- b) アクセスの持続性の保証
- c) DOI登録対象の粒度
- d) DOIのランディングページの要件の検討
- e) 研究データの特성에応じた取り扱い
 - a) DOI登録後データ追加がある場合
 - b) データ量が多い場合
- f) 研究データに登録したDOIの活用

検討中の課題

a) 運用フロー

- 取りまとめ機関

- 自身の保有する研究データに対してDOI登録
- 傘下の機関の保有する研究データを取りまとめてDOI登録
 - ・ データ作成者との合意

- データのライフサイクルと担当者・担当機関

- ・ 人・機関が異なる役割で関わる

- DOI登録の対象データ

- 品質管理等： 査読有無

- Prefixの割り当て方針

データのライフサイクルと担当者・担当機関

データライフサイクル	実施者／責任者
データ作成	研究者
データ保存	研究者／図書館等
メタデータ作成	研究者／図書館等
データ修正	研究者
メタデータ修正	研究者／図書館等
データ公開化	研究者／図書館等／研究機関
データ破棄(非公開化)	研究者／図書館等／研究機関
DOI登録	研究者／図書館等／研究機関／プロジェクト等

検討中の課題

b) アクセスの持続性の保証

●有期のプロジェクトで保有するデータの扱い

DOI登録で求められる持続的なアクセスの保証をどのように得るか

- プロジェクト終了後にデータリポジトリで保存
 - …データ寄託先の候補は？
- プロジェクトの代表機関が引き続き管理

●DOI登録後のデータ取り下げの取り扱い

DOI登録後に研究データの取り下げが行われた場合も
持続的なアクセス保証を維持できるようポリシーを決める必要

- メタデータは、データ取り下げを明記のうえ引き続き公開
 - ➡ 管理者はだれか？
- 取り下げる予定のないものに限って公開
- データ登録ポリシーに削除の条件を明記。ポリシーは「研究活動における不正行為への対応等に関するガイドライン」とも整合性を持たせる

検討中の課題

c) DOI登録対象の粒度

- 適切な粒度の検討

- データセット単位、測定レコード単位、図表単位、プロジェクト単位・・・

- 粒度を決める観点

- 引用： データ登録者が引用を期待する粒度、
- データ種別等： 単独で意味のある粒度、
- アクセスの容易さ： データ参照、再利用、
- 管理の容易さ： 提供機関、管理者、DBのテーブル構造、

d) DOIのランディングページの要件の検討

- データの公開、非公開のケース

- 記載内容

- 内容の維持管理

検討中の課題

e) 研究データの特성에応じた取り扱い

●DOI登録後データ追加がある場合

➤別DOIの登録(バージョン情報の明記)

- データを追加したら原則別DOIとする
- 長期的に取得している観測データセットの場合、新たなデータを追加したデータセットはシリーズとして別名をつけ、年単位ごとにDOIを登録する
- 時間経過に即したデータ追加は、原則バージョンニングしない
- データ取得・処理や補正、誤差の修正はバージョンニング管理する

➤更新履歴の記述方法

- 同一DOIでランディングページに履歴を記載

➤更新前のデータの扱い

- 厳選したデータのみDOIを登録するとすれば運用上問題にならない

●データ量が多い場合

- データセットとしてまとめて一つとして扱い、必要に応じて枝番をつける
- 全体データと個別のデータなどを関連づける

検討中の課題

f) 研究データに登録したDOIの活用

- 研究データDOIの引用

- コンテンツホルダーにとってのDOI登録のベネフィットを示す
- ジャーナルがデータDOIの表示を論文採択の条件とするよう働きかける

- アプリケーション

- データ利用状況の把握ができるサービス

Webサイトに随時プロジェクトの情報を掲載していきます

<http://japanlinkcenter.org/>

