

JaLC実験プロジェクト 研究データDOI登録ガイドラインの検討

国立研究開発法人科学技術振興機構

中島律子

(ジャパンリンクセンター事務局)

2015/7/3
JaLC研究データDOI登録実験プロジェクト
中間報告会



<http://japanlinkcenter.org/>

背景等

DOI名の割り当て原則 (DOIハンドブックより)

- どんなオブジェクトにもDOI名を割り当てることができる
- DOI名は、登録者が適切と考える任意の精度と粒度で割り当てることができる
- すでにDOI名を持つアイテムに変更があった場合、そのアイテムに新たなDOIを割り当てるか否か、IDFはルールを定めていない

研究データに対するDOI登録の先行状況

DataCite

JaLCと同様DOI登録機関のひとつ

2004: ドイツ科学技術図書館で科学技術データに対するDOI登録開始

2009: DataCite設立

2015: 正会員23機関、準会員8機関



出典: <https://www.datacite.org/>

研究データに対してどのようにDOIを登録するか？

- 日本ではまだ知見、経験が少ない
- DataCiteのドキュメントには、考え方の説明はあるが
実例等はない
- DOI登録に当たっては海外機関も各事情に合わせて検討

**➡ 実験プロジェクトにおいて、DOI登録を行い、
知見・経験を得ることにより登録ガイドライン
を作成する**

研究データに対する DOI 登録ガイドライン

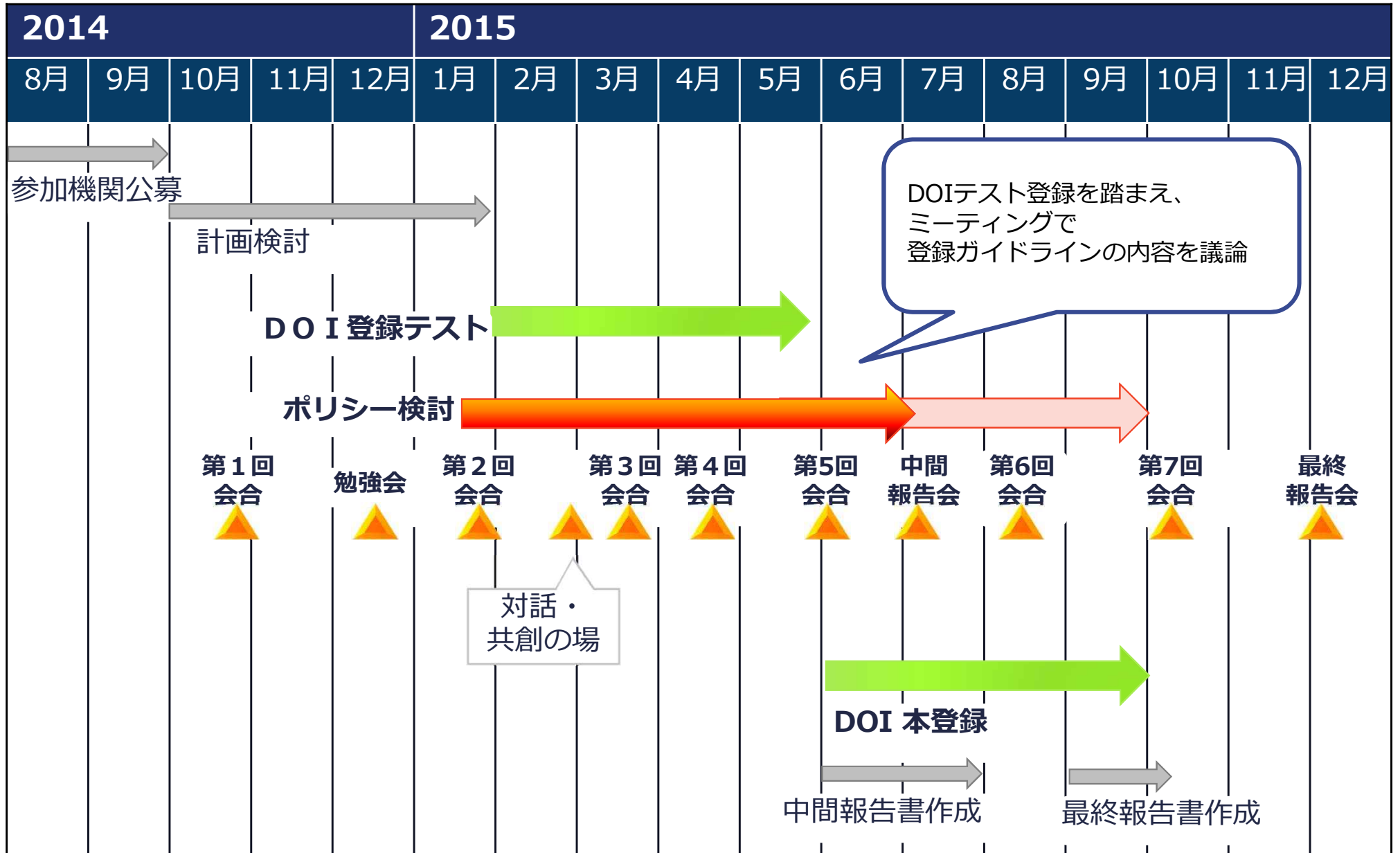
プロジェクトにて作成中

- ◆ 研究機関等がデータに DOI登録を開始する際の指針に
- ◆ DOI登録実作業への参考に

2015年10月

ジャパンリンクセンター
研究データへの DOI 登録実験プロジェクト

検討の経緯



ガイドライン目次

◆ ワークフロー

- データライフサイクルとDOI
- Prefixの割り当て方針

◆ DOI登録の対象データ

- DOI登録の対象とするデータ、品質に関する考え方

◆ アクセスの持続性の保証

- 研究プロジェクト終了後
- DOI登録後にデータの公開をやめる場合

◆ DOI登録対象の粒度

- 粒度を決める観点
- DOI登録後のデータ追加
- suffixのつけ方

◆ DOIのランディングページ

- 必要な情報

◆ 研究データに登録したDOIの活用

◆ 事例集、参考文献

ワークフロー

データのライフサイクルを考える

だが、いつ、どのように

- **研究データの生成**
- **リポジトリへの登録**
- **メタデータの作成**
- **DOI登録**
- **メタデータの更新**
- **データの取り下げ**

等を行うか



**研究者
図書館・データセンター
プロジェクト**

データライフサイクルと各工程の実施者

- 文献の場合 -

出版社等のケース

出版社

識別子(ID)

登録

メタデータ

作成

登録

修正

研究者

コンテンツ

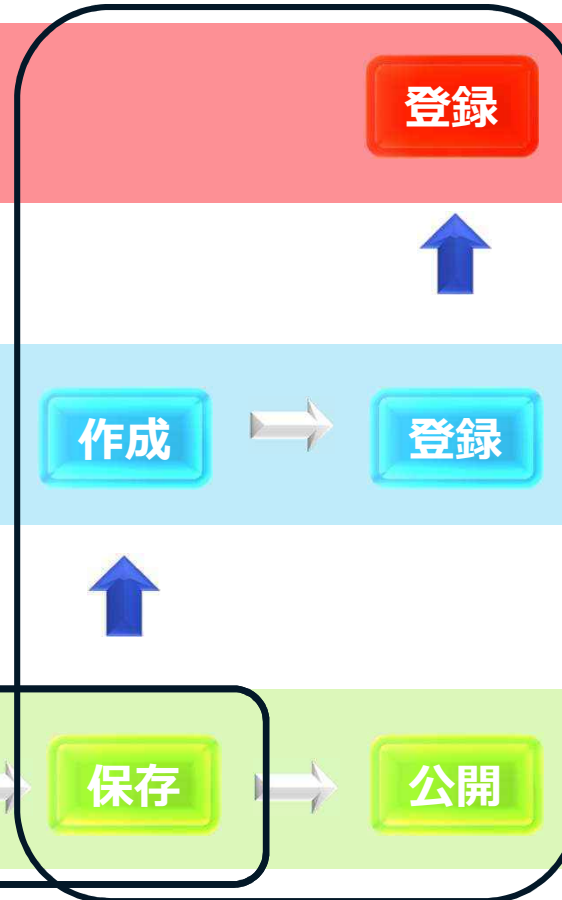
作成

保存

公開

修正

破棄



データライフサイクルと各工程の実施者

- 文献の場合 -

機関リポジトリのケース

図書館

識別子(ID)

登録

メタデータ

作成

登録

修正

研究者

コンテンツ

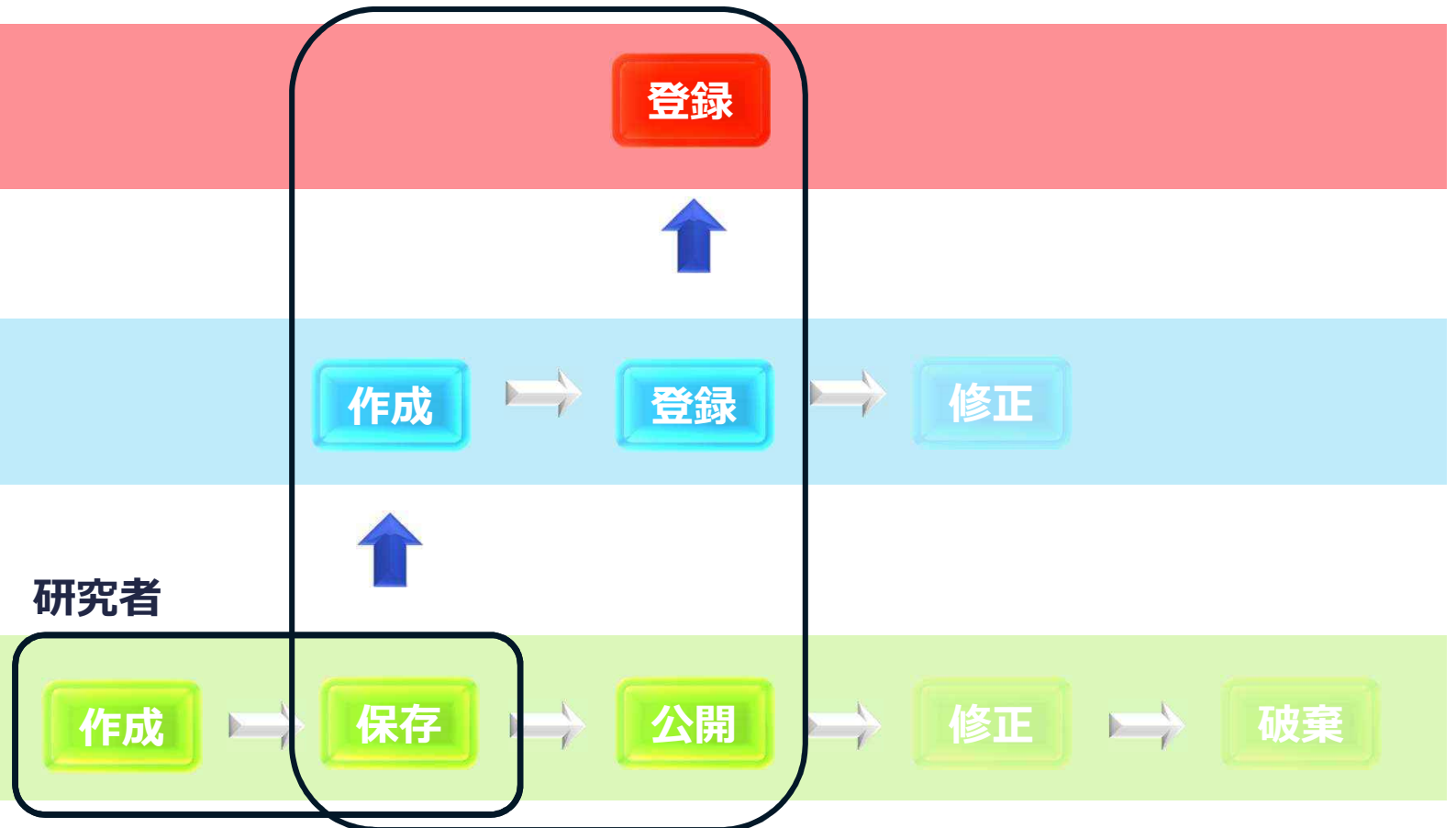
作成

保存

公開

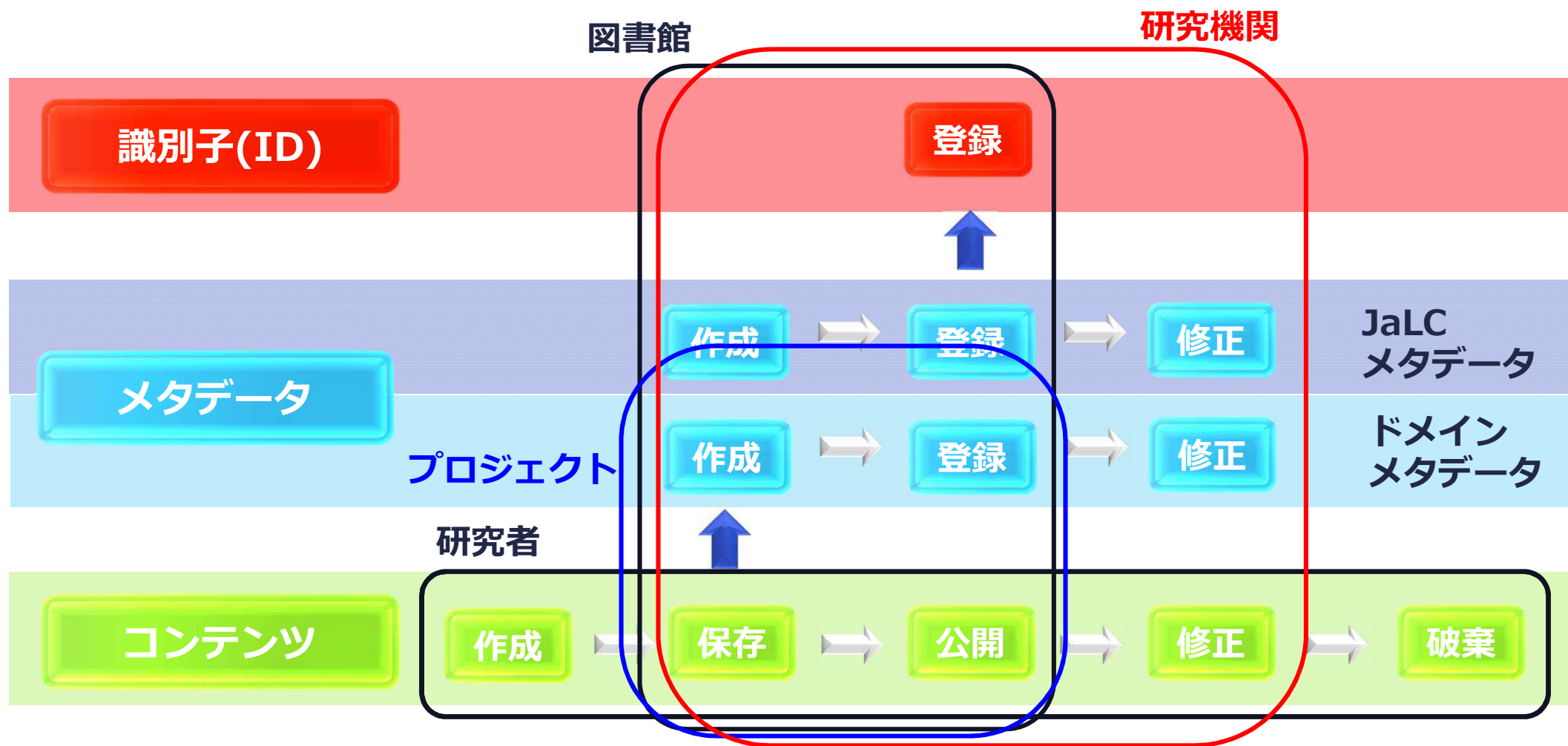
修正

破棄



データライフサイクルと各工程の実施者

- データの場合 -



DOI登録の対象データ

<アクセス保証>

- ◆ DOIが登録されたデータへのアクセス保証をしなければならない
 - 長期間にわたる管理体制の確保が可能なデータ
 - 組織として保全が必要であると判断した、取り下げの可能性が少ないデータ

<品質保証>

- ◆ DOIが保証するのはアクセシビリティ
- ◆ 品質については、メタデータに記述することは可能
第三者による査読が行われたものをDOI登録対象とし、メタデータに品質に関する記述をしている例もある

JaLCのメンバーシップ 研究プロジェクトの場合

◆JaLC会員の資格は法人または団体

- DOIを登録したコンテンツの継続的な管理が必要
- 論文にDOIを登録する際は特に問題なし

しかし、研究データは、プロジェクト（有期）により生成されるケースも多数

- プロジェクト終了後は？
- 複数機関から成るプロジェクトの場合の責任範囲は？

➔ 原則は、所属機関が管理

（複数機関の場合はプロジェクト代表機関または参加各機関）
あらかじめプレフィックスを分割しておくなどの措置



アクセスの持続性の保証

◆プロジェクト終了後のデータ保管

＜対応例＞

① 代表機関がプロジェクトによりDOI登録したデータをすべて引受。
管理Prefixの移管を伴うこともある

② DOI登録を行った機関がデータを保管

③ 引き受け先がないときは、外部のデータリポジトリ等へ寄託

- プロジェクト実施中から、終了後のアクセス保証のため
の体制等を、規約により定めておくことも有効
- 将来の管理体制を予測して、あらかじめその単位でPrefixを
割り当てておく

◆DOI登録後にデータの公開をやめる場合

DOIのランディングページは維持してメタデータの公開を継続し、
アクセスを保証。ランディングページには、データ公開を取り下げ
た事実、理由を明記

DOI登録対象の粒度

<粒度を決める観点>

◆データ提供者が引用を期待する粒度

◆単独で意味のある粒度

データ種別等によって判断が変わりうる

▶観測データ、実験データ、計算データ等

◆アクセスの容易さ

コンテンツが効果的に参照され、利用されるために

◆管理の容易さ

持続的なアクセス保証を容易に

◆DOI登録・解決のシステムが適切に運用できる数量

膨大な測定レコードの一つ一つに対してDOIを登録するようなことは行わない

DOI登録後のデータ追加

データを追加した後も、もとのDOIをそのまま使用する場合

- ランディングページに、データを追加した事実と内容を履歴として記載
- データ取得・処理や補正、誤差の修正は、バージョン管理
- 観測中で適宜データが追加される場合は、もとのDOIをそのまま保持し、バージョン管理もしない

あらためて別のDOIを登録する場合

- バージョン管理を行う。もとのデータのDOIと、データ追加後のDOIのメタデータそれぞれにバージョン情報を記載
- 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のDOI名を登録

参考【DataCite資料】

継続的、急速な更新が行われるデータセットの引用と保存には特別な課題あり

引用に関しては3つのアプローチ

- a) ある特定のスライスを引用（特定の期間に作成されたデータセットや、データセットの特定のエリアに対する更新のセット）
- b) ある特定のスナップショットを引用
（特定の時刻に作成されたデータセット全体のコピー）
- c) 継続的に更新されているデータセットを引用、ただしアクセス日時を追記

DataCite Metadata Schema for the Publication and Citation of Research Data
doi:10.5438/0010

DOIのランディングページ

<データ量やファイル数との関連>

- DOIを登録する単位は、対象データの量やファイル数には関係しない
- 1つのDOIに対してメタデータおよびランディングページは1 : 1 対応。登録対象データの実体は複数ファイルも可

<suffixのつけ方>

- suffixの文字列は、元来、意味も構造も持たない。
- ただし、管理者が扱いやすい付け方をしてよい。
 - 文字列がデータの意味や構造を示す
 - バージョンを示す枝番号を付加
 - データが格納されているDBにアクセスできるような構造を持つ例

研究データに登録したDOIの活用

議論はこれから。。。。

- コンテンツホルダにとってのDOI登録のインセンティブ
- アプリケーション開発
 - データの所在マップ
 - データ利用状況の把握ができるサービスなど

課題

<ガイドライン完成までの目標>

- 実例を豊富に
 - プロジェクトにて登録されたDOI、メタデータ、ランディングページを活用
- 義務、推奨、任意を明確に

<プロジェクト実施後>

- 実際に研究データへのDOIが登録され、活用されるようになるための活動が必要
 - たとえば
 - 関連データ可視化アプリケーションの開発
 - データリポジトリとの連携 など

「研究データへのDOI登録実験プロジェクト」の情報コーナー

http://japanlinkcenter.org/top/index.html#top_project

