

研究データへのDOI付与 が意味すること

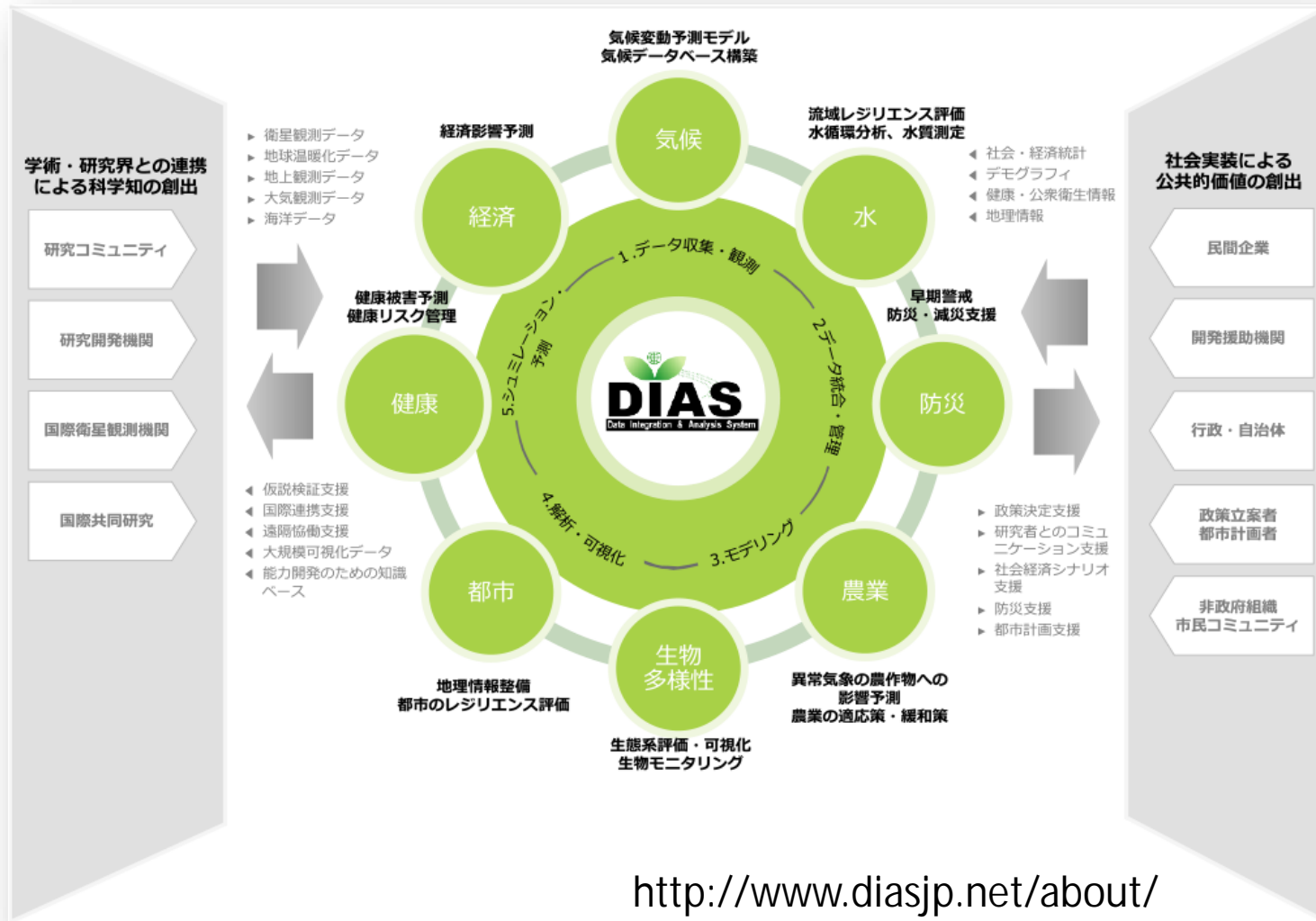
～ DIASにおける経験を踏まえて～

北本 朝展 (KITAMOTO Asanobu)

国立情報学研究所

<https://dias.ex.nii.ac.jp/>

DIAS第三期（2016～）



オープンサイエンス対応

- **オープンサイエンスの潮流**にDIASも対応していく必要がある。
- DOI付与を通じて、**永続性を確保したデータアクセス**を提供する。
- **データジャーナル**などの動きにも対応し、データの再利用を促進する。
- **データ引用**を通じて、データ公開者に報酬を与えられる文化を生み出す。

DOIとは？

Doi:10.20783/DIAS.1

この文字列をデータセットに紐付けることで、データへの永続的なアクセスを確保する。

- 単純化すれば、たったこれだけ！
- こんなに単純な話が、なぜ大きな話になってしまうのか？

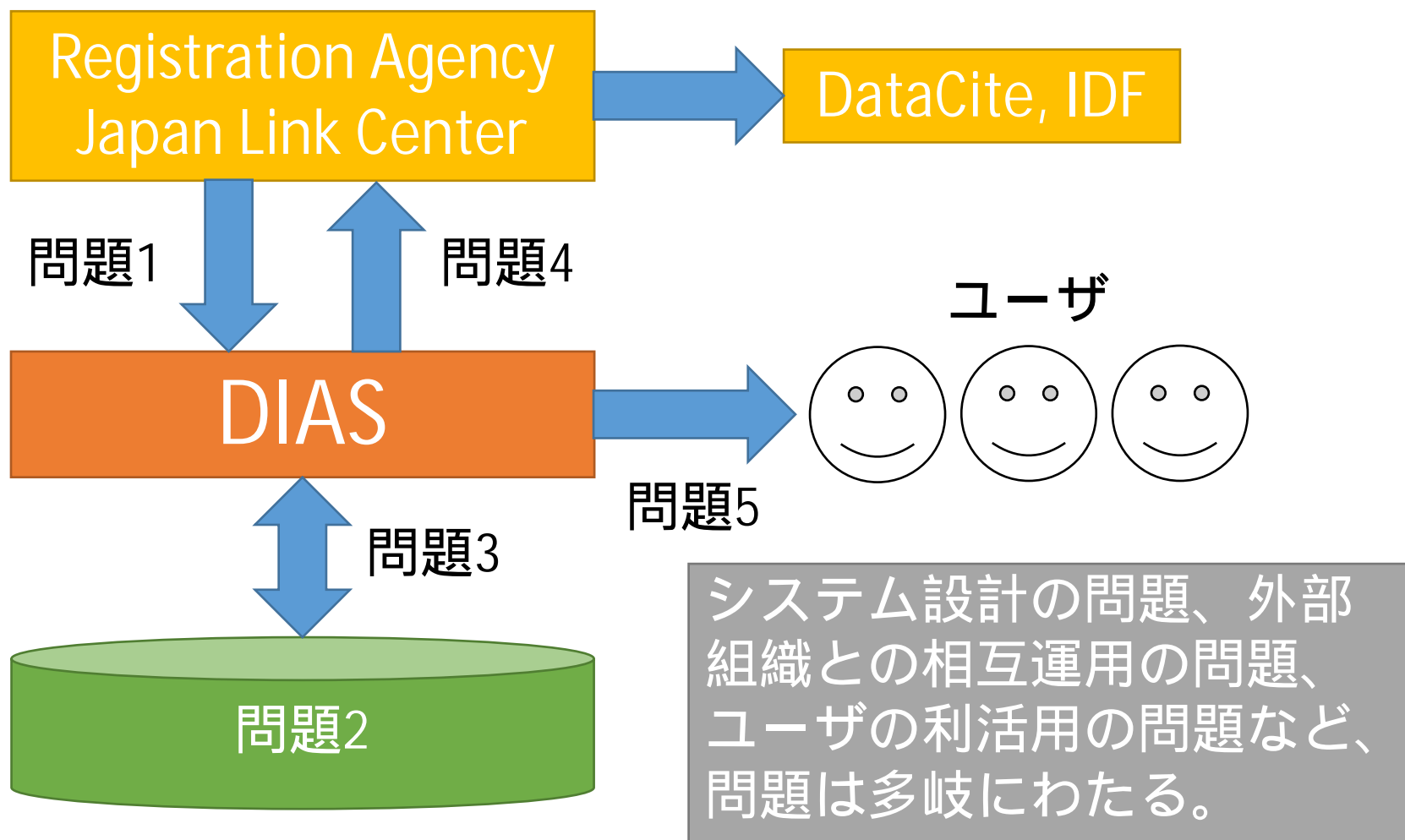
結論から言えば

DOIの付与は、最終的には「**信頼できるデータリポジトリ**」への成長を促す。

- 信頼できるリポジトリ（trustworthy repository）は、目的やワークフロー、ガバナンス等を、透明性ある形で公表している。
- 持続的な運営体制を備え、データへのアクセスを促進する仕組みを備えている。

リポジトリ全体を見直し、より高いレベルのリポジトリに変容させる必要がある。

全体像の図解



1. Prefix問題

- **10.20783の部分**。Registration Agency (Japan Link Center / JaLC) を通して申し込み、番号の割り当てを得る。
- **永続的にデータを提供できる組織のみが対象**。プロジェクトや個人は、永続性に問題があるため除外される。
- DIASの場合は、政府の資金提供に永続性が期待できるが、主体をどこにするかなど、**プロジェクト特有の問題**がある。

2. Suffix問題

- DIAS.1の部分。データリポジトリが独自にユニークな文字列を付与する。そこに意味を持たせる場合は慎重に決める。
- データライフサイクルの中で、公開データのユニークIDを一貫性ある方法で管理する機能が必要になる。
- IDはシステムの根幹を成すため、設計変更は影響範囲が大きく、システム全体の見直しに波及する可能性が高い。

データライフサイクル

- データは、受け入れ、システム登録、メタデータ整備、公開など、**いくつかの段階を経る**。
- DOIは永続的であるため、**変更や取り下げ要求**に対しても、一貫性ある方法で対応する必要がある。
- 既存システムをルール変更すると、**レガシーデータへの対応**も必要となり、新規システムに比べるとコストが大きい。

3.付与対象問題

- **粒度**：データ提供者が定義する「まとまりのあるデータ単位」が対象だが、イメージとしてはデータ引用の単位。
- **バージョン管理**：データを更新してもDOIは変更せず、メタデータの記述欄に変更内容を自然言語で記載。
- **データ選定**：DOIの二重登録は避ける。付与対象とするデータセットは、簡単な審査を経たものだけとする。

4. DOI登録問題

- **メタデータ** : DIASはもともと国際標準を参照していたため、JaLCメタデータ形式への変換はXSLTで自動化可能。
- **非同期性** : JaLC登録からDataCiteやIDFまでは非同期にデータが更新されるため、既存のシステムとは分離して構築。
- **共有化** : NICT/WDS国内委員会が構築したJaLCアップローダ（仮称）を利用。長期的にはコミュニティの共有リソースに？

5. 利活用促進問題

- **ランディングページ**：既存のデータセットページにDOI表示を追加。
- **データ引用支援**：データセットの標準的な引用方法をテキスト表示。
- **ライセンス整備**：研究データ特有の利用条件の整理は今後の課題。
- **何でもID**：複数のIDを統合的に活用したサービスの構築が長期的な課題。

信頼できるリポジトリ

1. 利用しやすく、ワークフローも透明性が高く、運営が持続的なサービス。
2. DOIは波及範囲が広いコア機能なので、DOI導入を梃子に設計を再構築する。
3. **認証**：現実的なレベルの仕組み = WDS や Data Seal of Approval (DSA) など。
4. **国際水準のリポジトリ**として、認証は一つの目標（日本は意識が低い？）。

オープンサイエンスの展望

1. DIASは地球環境データへのアクセスを、適切な条件の下で**オープン化**する。
2. DIASはデータ駆動型科学を支えるための**情報基盤**の参照モデルを構築する。
3. DIASは研究データ基盤を支える**人材**の育成にも取り組む。
4. DIASはデータ提供者が高く**評価**されるような文化を作り出す。

今後の計画

1. 現在、**3月末の運用開始**（DOI第一号付与）に向けて準備作業中。
2. 来年度以降は、DOI付与基準を明確化し、**DOIの件数を増加**させる計画である。
3. DIASにデータをデポジット→**データジャーナルに投稿**、も可能としたい。
4. **DOIの利活用**は模索中。データ引用、アプリ連携、利用調査、論文リンクなど。

謝辞

- DIASオープンサイエンス分科会メンバー（絹谷 弘子、清水 敏之、中原 陽子、吉川 正俊、小野 雅史）の協力を得て、オープンサイエンス対応を進めている。
- JaLCアップローダには、NICT/WDS国内委員会が開発したコードを利用している。
- より詳しい情報：<http://www.diasjp.net/>
または <https://dias.ex.nii.ac.jp/>