

# 研究データディスカバリーにおける 引用情報の利活用

大向 一輝

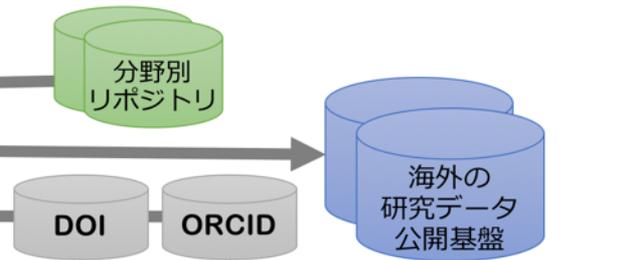
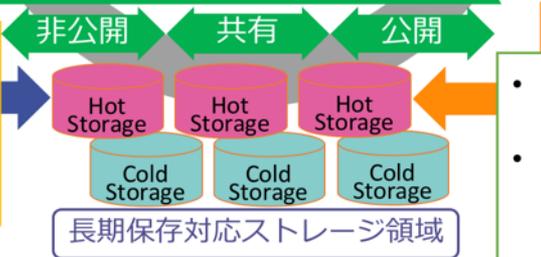
国立情報学研究所

# 背景

- 機関リポジトリ + 分野別リポジトリやデータリポジトリとも連携
- 研究者や所属機関、研究プロジェクトの情報とも関連付けた知識ベースを形成
- 研究者による発見のプロセスをサポート



- データ収集装置や解析用計算機とも連携
- 研究遂行中の研究データなどを共同研究者間やラボ内で共有・管理
- 組織が提供するストレージに接続した利用が可能



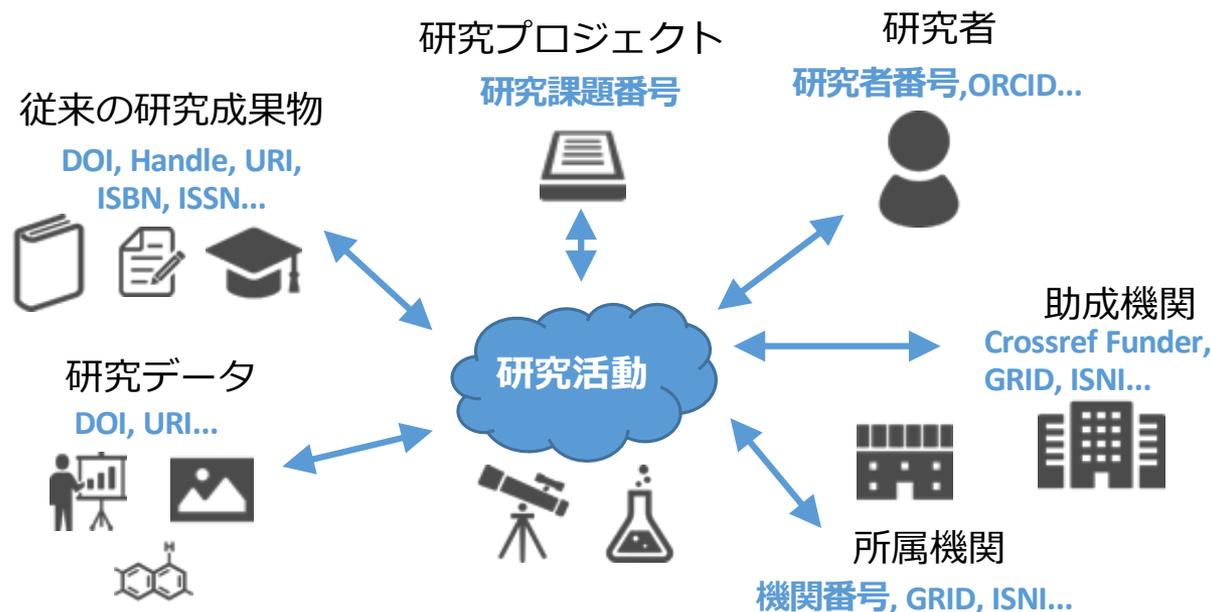
- データ管理基盤における簡便な操作で研究成果の公開が可能
- 図書館員やデータキュレータによる、メタデータや公開レベル統計情報などの管理機能の提供

# 研究データディスカバリー

- 「研究データ」を検索する
  - Google Dataset Search ・ Elsevier DataSearch ・ OpenAIRE…
  - 分野ごとに不揃いなメタデータ
  - 横断的・統合的検索が困難
  - 具体的なニーズは？
- 「研究データを取り巻く情報」を検索する
  - 論文・書籍・研究者・組織・ファンド…
  - 日常のワークフローの中から研究データを発見する
    - 「この論文で用いられたデータは？」
    - 「このデータを用いた研究は？」
- ナレッジグラフの構築
  - 本質的にネットワーク状の情報
  - ナレッジグラフを通じた研究データの検索

# ナレッジグラフ

- **情報**とその**関係性**を明示的に定義したもの
  - 情報： 論文・データ・研究者・組織・ファンド…
  - 関係性：引用・被引用・同一・所属・助成…
  - 関係性の記述に特化した組織は存在しない
- 学術系：SciGraph (Springer Nature) ・ Microsoft Academic Graph…
- 汎用：Google Knowledge Graph



# ナレッジグラフ

- 日本のナレッジグラフを作る
  - 論文：CiNii Articles 2000万件
  - 書籍：CiNii Books 1200万件
  - 研究者・ファンド：KAKEN 20万人・60万プロジェクト
  - 研究データ：機関リポジトリ・JaLC・分野ごとのデータベース
- 情報の関係性
  - 文献と著者：CiNii Articles名寄せ・KAKEN報告書
  - 著者と組織：機関リポジトリ
  - データと文献：データ引用
    - JPCOARスキーマ
    - Event Data (Crossref・DataCite)
    - Scholix
    - 国内ではこれから…

# 研究データ「被」引用

- データリポジトリによる研究成果の把握
  - データ公開の効果測定
  - コミュニティの拡大→イベント開催など
  - 事例
    - 情報学研究データリポジトリ（国立情報学研究所）
    - Social Science Japan Data Archive（東京大学社会科学研究所）
    - 国立国語研究所コーパス
- CiNiiナレッジグラフへの適用
  - データ書誌の作成・データIDの付与
  - 成果リストからの論文IDの引き当て



# まとめ

- データ引用情報の利活用
  - データを取り巻く情報からの検索・発見
  - データ提供者のビジビリティ向上
- 課題
  - データ・文献双方へのID付与
  - 引用・被引用のワークフロー
    - 「CiNii Researchかんたん登録キット」の提供？
  - 国際分業・分野間分業