

Title	欧州における科学イノベーション政策研究のための研究インフラ（RISIS）と日本の現状についての考察
Author(s)	林, 信濃; 中川, 尚志; 原田, 裕明; 松尾, 敬子
Citation	年次学術大会講演要旨集, 32: 735-739
Issue Date	2017-10-28
Type	Conference Paper
Text version	publisher
URL	http://hdl.handle.net/10119/14987
Rights	本著作物は研究・イノベーション学会の許可のもとに掲載するものです。This material is posted here with permission of the Japan Society for Research Policy and Innovation Management.
Description	一般講演要旨

2 H O 5

欧州における科学イノベーション政策研究のための研究インフラ（RISIS）と日本の現状についての考察¹

○林 信濃（国立研究開発法人 科学技術振興機構 研究開発センター）、
中川 尚志（国立研究開発法人 科学技術振興機構 研究開発センター）、
原田 裕明（国立研究開発法人 科学技術振興機構 研究開発センター）、
松尾 敬子（国立研究開発法人 科学技術振興機構 研究開発センター）

I. はじめに

欧州における科学イノベーション政策研究のための研究インフラ：RISIS（Research infrastructure for science and innovation policy studies）は、欧州連合（EU）の Framework Programme 7 によって 2014 年から 2017 年までの 4 年間資金提供を受けているプロジェクトである。RISIS は既存データ基盤のネットワーク化や共同研究の推進により新しい視点を与えられたデータ基盤の構築を目的にしている。参加機関は 9 カ国 13 大学および研究所であり、主幹事はパリ大学東校である。2004 年から 2009 年の間、欧州における科学技術イノベーションのコミュニティ統合に貢献した Prime Network of Excellence の関係者も多く RISIS に参加しており、総予算は 5 百万ユーロ²が配分されている。本稿では、事業の概観と日本のデータ基盤との比較を行い、今後のデータ基盤政策への示唆について述べる。

II. RISIS の目的

政策に寄与する科学技術研究を支援するデータ基盤等の研究インフラは欧州には点在しているが、連結や共有化がいままで考慮されることはないという現状を改善するための事業が RISIS の大きな目的である。この事業では、欧州に分散して存在しているデータベースの連携と統合を行い、European Research Area (ERA) を含む様々な研究課題への貢献を目指している。

近年、インターネットの活用により欧州各地に点在する【分散型】のデータ基盤として連結が可能になった。さらに、欧州で求められている研究課題に対応した形でのデータの再構築が予定されている。このような背景から、RISIS により欧州の科学を強化し、①イノベーション政策、②研究評価、そして③政策に関する指標、に必要なエビデンスを強化に寄与することを事業目標としている。

III. RISIS 事業内容

今まで、欧州におけるデータセットは各地の様々な団体によって運営されており使用も制限されていたため欧州の研究者全体に共有されていなかった。RISIS は既存データ基盤のネットワーク化や共同研究の推進により新しい視点を与えられたデータ基盤の構築を目的にしている。また、学生、様々な研究者、行政担当者に分析のためのデータの扱い方、ソフトウェアの使い方の研修も行っている。同時に多様なデータを分析するためのツール、ソフトウェア開発も行っている。

1. データセットの拡充

1-1 欧州研究領域に関連するもの

- i. EUPRO データベース（欧州のプロジェクトベースの協力関係についてのデータ）オーストリア工科大学
- ii. JOREP データベース（国境を越えた研究資金プログラムに関するデータ）イタリア学術会議持続的経済成長研究所
- iii. ナノテクノロジーデータベース パリ大学東校

1-2 大学および人材に関するもの

- i. ライデン大学ランキング ライデン大学
- ii. EUMIDA (European MicroData collection) と ETER (European Tertiary Education Register) : 欧州の研究者人材に関するデータベース ルガーノ大学
- iii. MORE データベース：研究者人材の移動に関するもの 北欧イノベーション・リサーチ・教育研

¹ 本稿は（調査報告書）「欧州における科学技術イノベーションネットワーク形成とデータ連結・拡張の動向」CRDS-FY2016-RR-04 2017 年 3 月 (<http://www.ist.go.jp/crds/report/report04/CRDS-FY2016-RR-04.html>) を基にまとめられた。

² 約 5 億 7500 万円（2016 年 7 月平均レートによる換算）

- 研究所
- iv. 若手研究者に関するデータベース ドイツ国立情報学研究所
 - 1-3 企業イノベーション
 - i. Corporate Board Invention:大企業の発明に関するデータベース パリ大学東校
 - ii. VICO データベース:操業開始企業とベンチャーキャピタルに関するデータベース ミラノ工科大学

2. データを運用するための2つのプラットフォームの創設

既存のデータだけでなく、ソーシャルネットワークサービス（SNS）に代表されるインターネット上の様々なテキスト、データを取り込み、研究者がコミュニケーションと協働をし易くするためのソフトウェア・プラットフォームを創設する。

2-1 SMS platform:インターネットを活用したデータセットの運用 アムステルダム自由大学ネットワーク・インスティテュート

2-2 CorTexT Manager platform: 膨大で様々な言語の科学技術関連のテキストから、関連する単語、表現をヒントに科学技術イノベーションにおける「知」の伝播や発展を分析する、としている。様々なデータセットに対応するためのオンラインプラットフォーム パリ大学東校

3. データの再構築のための共同研究

RISIS ではデータのオープンアクセス化を図るだけでなく、独自の視点からデータの再構築も計画している。RISIS の共同研究として、前述した既存の9つのデータを、企業、評価、欧州研究領域、公的研究、人材資源、そしてデータの活用支援の6つの観点から再編集する取組みが2016年から開始されている。

3-1 企業に関するデータセットの深化と連結 サセックス大学科学技術政策研究所

3-2 評価のための IPER レポジトリの構築 マンチェスター大学

3-3 欧州研究領域関連の EUPRO データベースと JOREP データベースの拡充 イタリア学術会議持続的経済成長研究所

3-4 EU における公的研究機関についてのデータセット構築 スペイン国立研究協議会

3-5 研究者人材のキャリア分析のためのデータセット構築 ドイツ国立情報学研究所

3-6 データ活用のための支援: 洗浄、統合、分析、品質管理 アムステルダム自由大学ネットワーク・インスティテュート

IV. RISIS データセットの特徴と日本との対比

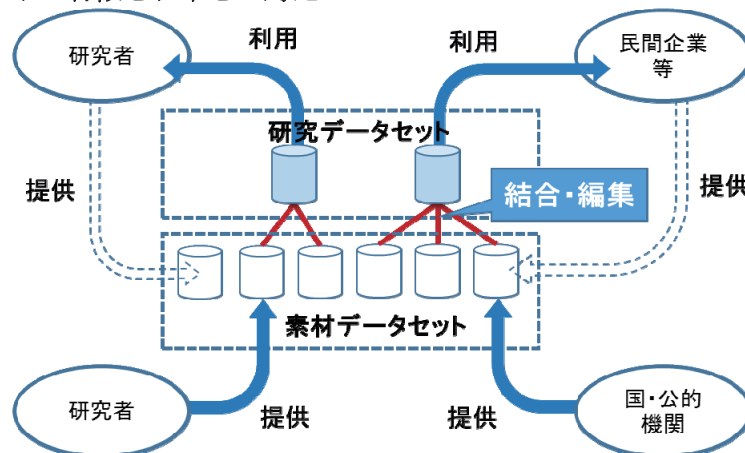


図1: 素材データセットと研究データセット

素材データセット

政府など公的機関から公開されているデータセット、また研究者が自ら作成または収集した一次情報のデータセットを指す。

研究データセット

素材データセットを加工編集して、研究のために利用しやすとしたデータセットを指す。自らの素材データセットだけでなく、他

者から公開されたデータセットを組み合わせて二次利用したものを含む。

図1は公的機関や研究者(研究組織)から提供された素材データセットを各研究者が独自に組み合わせ、編集して研究データセットを作成し、それを自分の研究に用いる経路を示している。各研究者(研究組織)はオープンサイエンスのポリシーに沿って素材データセットを提供するが、独自に作成した研究データセットをそのまま提供する場合もあり得る。

RISIS プロジェクトは各国の科学技術政策の研究プロジェクトが集合して、相互に研究データセットを利用し合うことによって、新たな研究課題を探るものである。すなわち、すでに各研究組織が自らの研究のために、素材データセットを編集加工して、独自の研究データセットを蓄積しているという背景がある。

たとえば CIB (大企業情報) や Nano (ナノテク知財情報) のデータセットがそれにあたる。このうち CIB の場合は、PATSTAT (欧州特許庁が頒布している学術用特許情報) と ORBIS (Bureau Van Dijk 社が頒布しているグローバル企業情報) の二つの素材データセットを組み合わせて作られている。

RISIS プロジェクトではそれらをさらに横断的に組み合わせるフェーズに入っている。たとえば CIB と VICO (スタートアップ企業情報) を結合した上で、あらたに FGFM (中規模企業のデータセット) を作ろうとしているのがそれにあたる。FGMF は欧州 28 カ国 (イスラエルを含む) の中で売上高または総資産が中規模の 67,439 社の情報を集約している (2016 年初時点)³。そのようなデータセット間の結合を可能としようとしているのが RISIS プラットフォーム (Cortex, SMS) である。

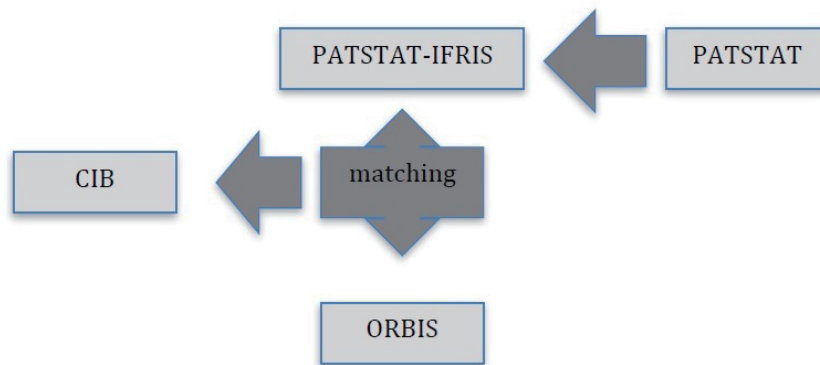


図2: CIB データセットの成り立ち⁴

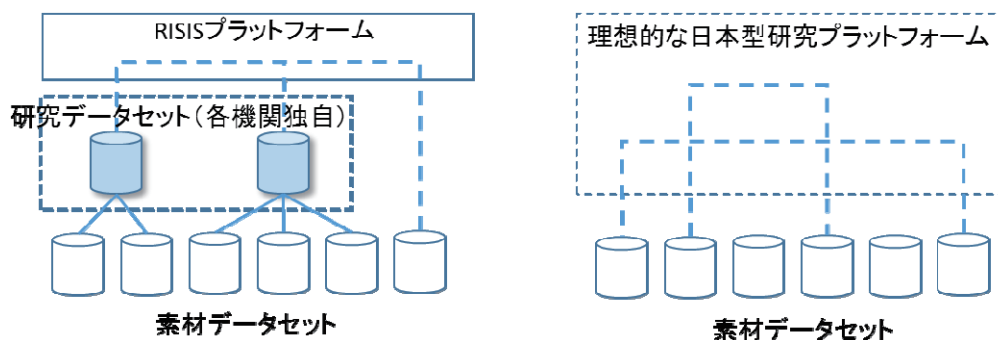


図3: RISIS と日本のプラットフォーム

図3はRISISの状況と日本の状況を模式的に対比させて描いたものである。図8のRISISプラットフォームはCIBやVICOのような既存の独自研究データセット同士をさらに上位で整合させることに注力している。RISISプロジェクトの各研究機関には、それぞれの研究データセットをこれまで構築してきたさまざまなノウハウ、経験も蓄積されていると考えられる。

一方、日本では政府が率先してオープンデータ化を推進してきており、政府統計データはじめ、相当量

³ <http://risis.eu/wp-content/uploads/2016/03/20160126-c-firms-3-FGMF.pdf>

⁴ <http://risis.eu/wp-content/uploads/2016/09/Del-6.4-Report-CIB.pdf>

の公的情報がオープンになってきた⁵。その他に、論文、特許、企業情報など「素材データセット」はかなり揃っており、それらを組み合わせた独自の「研究データセット」を作成すること自体はある程度可能な状況にある。

このような日本の現状において、もし素材データセットが自由に組み合わせられ、さらに素材データセット間の不整合が吸収されるような「日本型研究プラットフォーム」がいったん構築できれば、多面的に素材データセットを組合せて研究に用いる例が一気に増加することが期待できる。

しかし日本には研究データセットを作成し、公開してきたような蓄積が少ないため、プラットフォームの姿を具体的に設計するためには、なお検討が必要である。日本の研究機関が実際に素材データセットを組み合わせ、有用な研究データセットを作成し、それを公開する事例を増やすことによって、技術的な課題や知財上の課題を一つずつ解決していく必要がある。

V. 日本のデータ基盤への示唆

欧州の RISIS プロジェクトと日本の状況を比較すると、少なくとも政策研究におけるオープンサイエンシ化に相当な開きがある。その理由はいくつか存在する。

第一に、大学等の研究機関が持つであろう個々の研究データセットはいまだ十分に公開、共有できているとはいえない。研究データセットの公開のためにはデータの整理や手引書の作成など余分な作業も発生するが、そのようなコストを負担することも公開に後ろ向きになる原因であろう。

第二に、公開されている素材データセットであっても容易に相互連携できないという問題が残るとの声がある。公開されている素材データセットを使って、あらたな研究課題に挑戦しようとしても、データの欠損など不備点の確認と、加工編集の手間など間接的な作業に研究者が相当の時間を費やしてしまう結果になり、チャレンジがしにくい。

第三に、RISIS プロジェクトのように政策を取り巻くさまざまな研究者がコミュニティを形成し、それぞれの研究データセットを持ち寄ることにより、新しい観点からの研究課題を見いだすという文化が十分成熟していない。専門分野の学会は研究成果を発表、討議する場として機能しているが、研究データセットを中心として、技術的に発見を導くような議論の場がない。

RISIS に関わる研究者たちのモチベーションの根底にはモビリティ（国境を越えた移動可能性）の高さがあると考えられる。また欧州全体を研究対象にできることも、研究者にとっての魅力となるであろう。ひるがえって日本の科学技術イノベーション研究を俯瞰したとき、「国境を越えたデータ基盤構築の必要性の認識」は希薄である。しかしながら日本が直面している大きな問題に視点を移すことによって、日本ならではのデータ基盤の必要性が強く認識されるのではないか。

たとえば、日本にとっての大きな問題として、地方における少子高齢化・過疎化がある。各自治体がその対策として産学公連携に熱心に取り組んでいる事例は多い。それらの事例情報を共有し、さまざまな素材データセットと合わせて分析し、それぞれに最適なイノベーション施策を導出できるようなデータ基盤が生まれれば、地域活性化に大きく役立つことが期待できよう。その際には、自治体施策の状況をデータベース化した「地域科学技術政策支援システム RESIDENS」や、人口や産業に関わるデータセットを組み合わせた「地域経済分析システム RESAS」のようなツールがデータ基盤として有用になるかもしれない。この他に、日本においてスタートアップ企業がまだ少ないため、産業の新陳代謝が進まないという問題もある。それを制度面、経済面、文化面などから多角的に分析して、支援策を検討できるようなデータ基盤も有益ではないだろうか。

以上の考察をもとに、日本のデータ基盤事業にとって重要と考える項目を以下にまとめた。

- (1) 国の公開データを「二次利用」を前提として再整備する
オープンデータの5段階モデルにおいて、第5レベル（外部連携可能）に近づけていく必要がある。政府のオープンデータ戦略に沿って進める中で、特に民間企業や研究者から公開の要望が高いデータセットから優先的に第5レベル化をめざすことが望ましい。
- (2) 研究コミュニティを増大、活性化する

⁵ 各府省 CIO 連絡会議：「日本のオープンデータ憲章アクションプラン」（2013.10.29）の「別添 データセット別の公開の現状と今後の取組予定」参照

Open Data METI や Residens がおこなっているように、利用者の輪を広げ、その協力によって研究の質を向上したり、新しい課題提案を聞いたりする仕組みが必要である。もちろん、これらのコミュニティは研究分野ごとの特性に応じて個別に活動することが前提であるが、異分野を横断したところからイノベーションが生まれることを期待して、日本全体のオープンサイエンスの流れの中で緩やかに連携して、相互にデータセットを交換できる場づくりが望まれる。

（３） プラットフォームを開発する

データ不整合が見つかった際の自動調整、データアクセス構文の統一と単純化等、データセットを連携させる際に生じる技術的な共通課題をなるべく低コストで解決できることが望まれる。それにはデータマイニングや機械学習のようなコンピュータ技術の支援も有用である。

またプラットフォームの運用においてもコスト低減につながる工夫の余地がある。たとえば、調査票に回答を記入する際にメニュー選択方式を採れば、自由表記のゆれを後処理で検査・補正する作業コストは大幅に減ると推測される。