

## IRが目指すべきもう一つの仕事

### ベンチマークプラットフォームとしてのIRデータ形式の標準化

森 雅生<sup>1</sup>, 小野寺 理香, 大石 哲也 (東京工業大学)

#### 1.はじめに

昨今の大学 IR で注目されているのは、最新のデータサイエンス技術を用いた教育・学習効果の測定である。そうした教育・学習成果の測定には、しっかりとした教育情報の収集や管理が必要であり、その努力は各大学やその連携団体、関係機関で行われている[1][3]。

研究情報については、政府主導の一元的なデータベース[2][4]が運用されており、情報公開の観点から、詳細な情報が得られるようになった。

これらのデータベースの取り組みは、説明責任を果たす情報公開を推進する意味で一定の成果を収めていると言える。しかし今後は、教育情報および研究情報をはじめとした大学組織の情報を包括的に収集管理し、国内外のベンチマークを可能にすることが要求されてくるだろう。

国立大学の法人化および認証評価の義務化以降、高等教育機関や研究機関の組織評価における根拠情報を中心とした、ベンチマークなど活用志向の強いナショナルデータベースの提案が言われ続けているものの、技術的および財政的な困難が伴い、実装と運営について現実的なものとなっているとは言い難い。

本稿では、むしろデータベースそのものではなく、大学情報の標準データフォーマット

### Who needs Research Information?



図 1 研究情報を取り巻くアクター (出典：参考文献[6])

<sup>1</sup> Contact author: mori@irds.titech.ac.jp

の可能性を議論し、欧米の先進的な取り組みの状況について紹介する。

高等教育機関や研究機関が、教育・研究・社会貢献・国際交流といった大きな役割からブレークダウンし、実際の業務システムのデータスキームにつながるオントロジーを持ちつつ、データフォーマットへの変換をスムーズに行い、ベンチマークなどに対応できる情報流通のあり方の議論の端緒となることを本稿の目的とする。

## 2. 大学 IR 情報とデータベース、データフォーマット

教育や研究に関する機関の情報はどのように捉えれば良いのであろうか。大学など高等教育や科学研究を担う機関の情報は、必要とするアクターに対応して多面性があることが認識されている。例えば、euroCRIS (後述) では、研究情報に関しては情報を必要とするアクターによって多面性があることを認識されている (図 1)。

研究情報に関する、これらの多面性を集約すると、次のようにまとめることができる。

- 研究者情報データベース (研究者を軸、例: researchmap)
- 学術情報リポジトリ (論文を軸、例: CiNii)
- 研究プロジェクトデータベース (研究プロジェクトを軸、例: KAKEN)

しかし、これらのデータはもともと単一の研究活動から得られる情報であり、情報の単位 (軸) を何に設定しているかの違いで、データの形式が大きく異なっている。情報源が一つで、そこからデータ変換でこれらのデータベースへ移入されているのであれば問題は無いが、日本の場合、こうした個別データベースがそれぞれの目的で作成され運用を続けられているため、統合的な取扱いが困難となっているのである。

解決策は、全て一から統合データベースを作成してやり直す方法 (物理レベルでの統一) が考えられるが、日本の大学が置かれた状況から、それは現実的ではない。故にそれぞれのデータベースから情報が出される際に、その形式を統一することがより現実的であると考えられるのである。

## 3. 米国の事例

米国では高等教育機関における教育情報の統一的データベースとして IPEDS (the Integrated Postsecondary Education Data System) が知られている。このデータは、日本における学校基本調査に類似するが、春秋とデータ提出が義務化されており、情報項目は教育のみならず財務、教職員の給与情報にも及ぶ (表 1)。これらの情報項目に従い、個別大学の基本統計量が示されている。例えば、給与情報であれば、雇用期間別に職位、性別で各大学における平均給与額が示される。

IPEDS は国立のデータベースとして運営されているため、本稿の研究目的であるデータ交換などを目的としたデータフォーマットは存在しないが、米国の個別大学の教育情報が得られるので、ベンチマークの観点でこれらの情報項目に注目する価値はあると思われる。

他に高等教育における教育技術の高度化を推進する協会 EDUCAUSE では、高等教育機関における IT 設備環境に関する定期的調査を行い、ベンチマークする取り組みを行っている (Core Data Service)。[5]

#### 4. 欧州の事例

欧州では、研究情報の統一的取扱いについて早くから取り組まれている。EU 圏内の複数の国から複数の研究者による共同研究などが進められており、その成果報告書の作成については、言語の違いや様式の違いが困難の原因となっていたため、標準フォーマットの開発が 1991 年から行われている[6]。このフォーマットは CERIF (Current European Research Information Format) と呼ばれ[7]、欧州における研究情報システムに関する推進団体 euroCRIS (European Current Research Information System Association) により開発が行われている。

CERIF は、データベースを構築する際に、取り扱う情報の意味論を抽象的なレベルから定義するオントロジー工学に基づいて整備されている。その特徴は次の 3 つである。

- **概念レベルの明確化**…通念として通用されている概念の構成や関係を明示する。
- **論理レベルの明確化**…データベース設計を助ける。
- **用語の明確化**…研究情報で使われる専門用語の正確な定義を助ける。

一方で、本稿の観点からの教育情報の標準データフォーマットは未だ整備されていない。CERIF 開発関係者は、欧州の各方面から CERIF の教育情報版 CEEIF (Current European Education information format) の期待があることを認識しているが、CERIF 自体ボランティアで行なっているため、教育情報までの余裕がないことが実態のようである。

ただし、EU 圏内で転学や院留学進学する学生に向けて、成績情報のデータ交換を可能にするための取り組み EMREX が試行されており[8]、これが教育情報版の CERIF の基礎となることも期待される。

また、米国同様、IT 施設設備のベンチマークがフィンランドを中心に進められており、

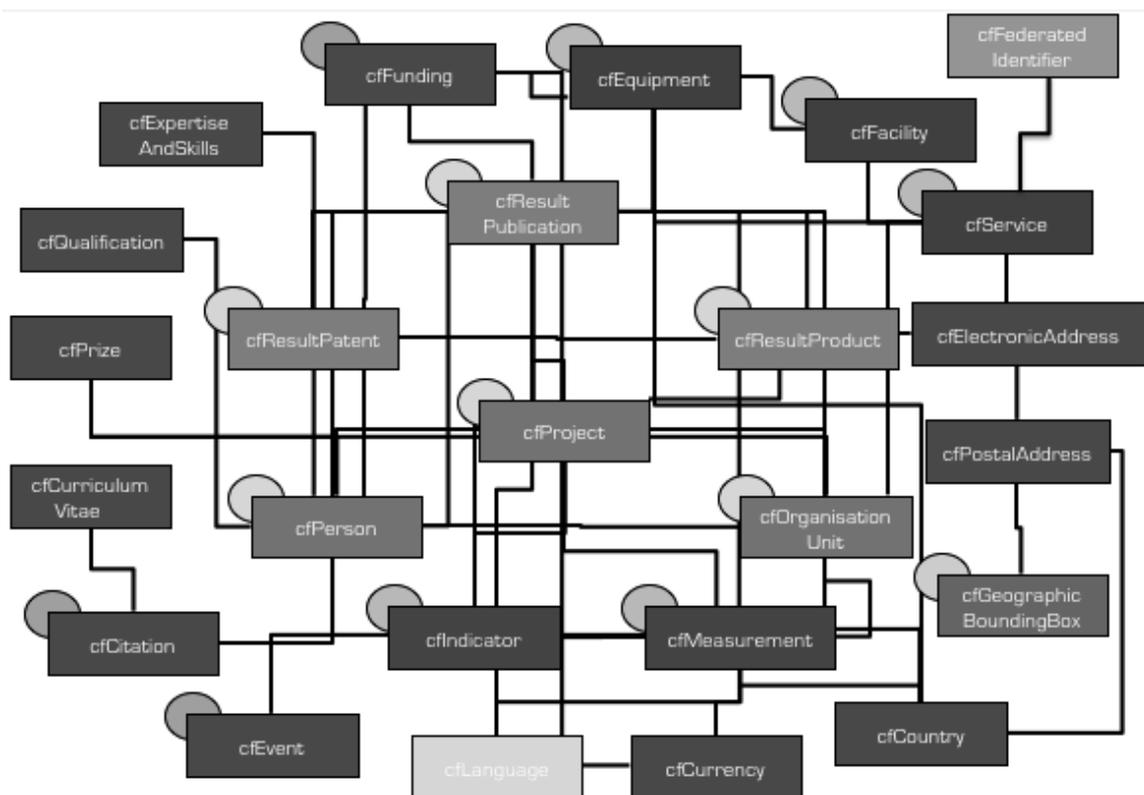


図 2 CERIF1.6 概要図 (出典：参考文献[6])

IT 設備に対する投資についての意思決定に活用されている (BenchHEIT) [9]。

## 5. 議論

欧米の状況をまとめると、

- 欧州は、研究情報について標準フォーマット (CERIF) が発達しているが、教育情報についてはデータ項目の統一化に着手されたばかり (EMREX)。
- 米国は、教育情報はとういつてきなデータベースが整備されており、ベンチマークも可能 (IPEDS)。研究情報については不明。
- IT 施設設備のベンチマークは、調査形式で欧米共に行われている。

である。

日本において、このような標準フォーマットの提案をする際に注意すべきポイントは、次の 3 つであろう。

1. 多くの高等教育機関は、教育研究を両輪で進めることを念頭に置いており、この観点では欧米にはない。よって、教育と研究を併合したデータフォーマットにすべきか、別に考えるかを判断する必要がある。
2. すでに運用されているデータベースや情報収集の取り組みが欧米に比べかなり多い。標準フォーマットを提案しても、これらの取り組みからの賛同がなければ実質化しない恐れがある。
3. CERIF は、関係する企業や機関のボランティアで開発されているが、日本ではそのようなコミュニティを構成するのは困難である。特に、情報技術が必要であるので、これを持つ人材をどのように見出すかが課題である。

最初の課題については、まずは日本の大学や研究機関の研究情報を CERIF に沿ってレビューし、教育情報のフォーマットについては、それとは別に構築するほうが良いだろう。すなわち、現存する各種データベースやデータセットの構造を包括するよう、教育研究それぞれのオントロジーの定義を試みる。その後で、教育活動と研究活動について関連性のあるものを定義すれば良いと思われる。

例えば、大学院生が共著者として関わる論文は、教育の成果と同時に研究の成果とも言える。論文自体は、学術リポジトリの中で論理的に情報が定義されており、教育成果と研究成果が学術リポジトリを通して関連し合う、と言った具合である。

二つ目の課題については、Researchmap や KAKEN を運営する JST や NII、リポジトリを推進する JPCORE、機関別認証評価の根拠資料を持つ認証機関など、コンテンツを保持し管理する主体や連携組織による積極的な関与が必要であろう。こうした組織が関与すれば、大学や研究機関も標準フォーマットの導入を検討しやすくなる。

また、最近では研究者に世界唯一の ID を付与する ORCID[10]という取り組みが普及しつつある。個人情報と繋がらないが、研究者個人を特定し異動しても追跡できるような ID である。さらに、これを活用すれば、論文書誌情報を整理する際に必ず立ちはだかる共著分析の作業が劇的に減る。こうした ID を利用したフォーマットの開発は関係組織の導入をさらに促進すると考えられる。

三つ目の課題について、最も困難な問題は、オントロジー工学などを踏まえて教育・研究情報の標準フォーマットを開発できる人材が乏しいことである。IT ベンダーの開発者や

情報学の研究者からの貢献が期待されるが、現在の日本の高等教育機関や研究機関の環境で、欧州のようなボランティアでの貢献を期待することは無意味であろう。よって、この標準フォーマットの開発は、大学 IR を推進する一つの研究事業として位置付けて研究費を取得し、専門人材を雇用して開発することが現実的かと思われる。

## 6. 最後に

大学 IR に取り組み、その成果を教育学系の学会などでの発表を多く目にするようになって久しい。ただし、それらの多くは歴史研究や事例発表に過ぎず、日本の大学 IR を後押しするような活力になっているとは言い難い。また、IR の成果を論文などにすることで、IR 自体に学術的価値を見出すことは、かなり難しいと思われる。抽象論・一般論やその大学だけで有効な事例研究も必要であるとは感じるが、標準フォーマットの開発のように IR を推進するための基盤づくりを進めることも、大学 IR 従事者の使命だろう。

※ 本研究は、科研費課題「課題番号 17K04677 日本の大学における諸活動を可視化するオントロジーマップの構築」の一部として行われています。

### 【参考文献】

1. IRiS, 大学 IR コンソーシアム, <http://www.irnw.jp/irsystem.html>
2. Researchmap, JST, <http://researchmap.jp/>
3. 大学基本情報, 大学改革支援・学位授与機構, <http://portal.niad.ac.jp/ptrt/table.html>
4. KAKEN データベース, JST, <http://kaken.org/>
5. Core Data Service, EDUCAUSE, <https://www.educause.edu/research-and-publications/research/core-data-service>
6. Jan Dvořák, “CERIF Tutorial”, euroCRIS membership meeting 2017.
7. euroCRIS, “CERIF1.6”, <http://www.eurocris.org/cerif/main-features-cerif>
8. Mats Lindstedt et al., “Supporting Student Mobility — Expected and unexpected results from the EMREX Field Trial”, EUNIS 2017 conference, 2017.
9. Ilkka Siissalo et al., “Bencheit – Benchmarking Higher Education It - An Update On The Work Of The Eunis Bencheit Task Force”, EUNIS 2017 conference, 2017.
10. ORCID-ID, <https://orcid.org/>, ORCID, Inc.

表 1: IPEDS の情報項目

調査内容	テーブルタイトル
機関の特性	教育の提供、組織、入試、サービス、運動競技団体
	年次プログラムの学生費用
	プログラム別の学生費用(職業プログラム)
	基本方針
入学	入学検討者(Admission considerations)、志願者、入学者、および入試点数
在籍者数	重複しない年間在籍者数
	年間教育活動
秋学期入学者	性別、出席状況、および学生のレベル
	学生の人種/民族、性別、出席状況とレベル:
	学生の年齢区分、性別、出席状況、レベル
	新入生の住居と移住
	全体の入学クラス、学生在籍率、学生数と教員数の比率
	学生の遠隔教育状況とレベル
受賞歴	賞のレベル、人種/民族、およびジェンダー関係で授与された賞や学位
	人種/民族、およびジェンダーごとの賞/学位を受ける学生数
	賞のレベルとジェンダー、人種/民族、年齢区分による賞や学位を受ける学生数
	賞のレベルにより発生したプログラム数と遠隔教育を経て発生したプログラム数
学生財政援助	学生財政援助
	学生財政援助と実際の学費
	従軍者サービスおよび退役軍人給付
卒業率	卒業率データ、標準終了年限 150%
	卒業率データ、標準終了年限 100%
	卒業率データ、標準終了年限 200%
財務情報	公共機関-GASB 34/35:1 会計年度
	FASB を使っている私立の非営利組織または公共機関: 1 会計年度
	私立の営利組織: 1 会計年度
人的資源	職種、所属、および雇用契約別の教職員数
	ジェンダーおよび学術ランク別の常勤教員数と給与(医学以外)
	職種別(医学以外)の常勤職員数と給与
	職種別、人種/民族およびジェンダー別の常勤・パートタイムスタッフ
	学術区分、所属、および契約状況ごとの常勤教員数(学位授与機関)
	契約状況、学術区分、人種/民族およびジェンダー別常勤教員数(学位授与機関)
	職業区分、人種/民族およびジェンダーごとの新規採用者(学位授与機関)
	頻用される派生変数・人的資源(HR)
学術図書館	学術図書館蔵書数