

国公立大学の大学ポータル関連データを利用した情報 活用支援

— 積極的支援への転換とその深化のために—

周東夏希、筒井優子、佐々木伸、金原英徳（大学改革支援・学位授与機構）

1. 本発表の目的と課題

大学改革支援・学位授与機構（以下、「機構」）では、大学ポータルによる教育情報の公表にあたり、学校基本調査に相当するデータ（大学基本情報）を国公立大学、公立短期大学、株式会社立大学から収集している。大学基本情報は、大学ポータルで公表する情報の調査票に転記することで大学の入力負担の軽減を図るとともに、各個票を集約し、国公立大学等の情報が一覧できるよう Web サイトにて Excel ファイルを公表してきた[1]。これによって、以前は文部科学省が公表する集計値しか利用できなかったデータについて、個別の大学・学部等の値が分かるデータを提供できた。

機構では、大学基本情報の活用を促進するため、BI ツールによるレポートを作成して国公立大学等の教職員を対象として 2019 年 10 月から閲覧に供するとともに、読み込み元となる加工済み元データも併せて提供することとした（「国公立大学情報活用サイト」）。本稿は、事業の背景と作業の概要、期待される効果について事例紹介することを第一の目的とする。また、レポートと元データの提供は事業の終着点では決してない。大学関係者の利用の裾野を広げ、真に役立ててもらうために機構がクリアすべき課題はなお山積しているため、本稿では担当者としての所感を交えながら今後の課題も併せて展望し、事業の充実に向けた検討の一助としたい。

2. 取組の背景 —情報活用支援という目標を達成するために—

大学ポータルは、Web により国公立大学等の教育情報を一元的に公表しており、情報の収集とデータベースへの登録については、国公立大学等は当機構が、私立大学等は日本私立学校振興・共済事業団が、それぞれ行っている。事業の開始にあたって取りまとめた大学ポータル構築のための「論点整理」では、集められた情報を各大学の教育改善に活用してもらうことも目標とした[2]。機構では、収集した国公立大学等のデータを IR 等に活用してもらう方策を検討し、その最初の取組として、上述のとおり大学基本情報を集約し、Excel データを公表してきた。

ただし、この大学基本情報のデータは、各大学から収集したデータを未加工で公表したに過ぎず、組織のコードや名称の誤りが多く含まれている。また、閲覧者の可読性に配慮していることなどにより、そのまま BI ツール等で分析することはできないものとなっている。そのため、IR 担当者等が大学基本情報を利用しようとすると、データのクリーニングや整形に手間がかかることになる[3]。

情報活用支援という目標からすれば、このように不完全なデータを提供することは、大

学担当者にとって「気の利かない取組」となっていた感が否めない。大学ポートレートに参加する国公立大学等は約180校ある。クリーニング・整形・レポート作成に掛かる作業量が10人日だとすれば、仮に180大学が大学基本情報を利用しようとするなら1800人日の作業負担を発生させてしまう（もちろん全大学が利用するという想定はいささか現実離れしているが、情報活用支援を企画する前提として、可能な限り多くの人々の利用を期待するのは当然である）。今回、機構がBIレポートと元データを提供しようとしたのは、担当者の習熟度に依らず簡易に情報が活用できれば、IR担当者の作業負担が軽減されるとともに、情報の利用者が増加するのではないかと考えたためである。

3. 提供に向けた作業 —自動化できる手順は何か、どのように共有するか—

データ加工の具体的な作業内容は[3]に詳しい。本稿では今回機構が行った作業を簡単に記すにとどめる。なお、使用したツールは、(1)及び(7)はExcel、(2)～(6)はExcelに付属のPower Queryというクエリエディタ、(9)～(11)のBIツールはPower BIである。

- (1) 組織名とコードが一意に対応するよう変換
誤記や表記揺れ（例：「1/I」、「--システム工学科/--システム工学科」）を修正する。
- (2) テーブルの細分化、不要な列の削除
元ファイルを複製し、列を削除して複数テーブルに分割する。また、合計列等は不要なので削除する。
- (3) null 値や不要情報の置換
BIツールでの読込エラーとなるため、値は「0」、テキストは「-」などに置換する。
- (4) ヘッダの調整
元データでは非表示行として1行にヘッダが集約されている。今後の変換が適切に行われるよう、必要に応じてヘッダ名の修正や列の分割をする。
- (5) ワイド型からロング型への変換
BIツールで扱いやすい形式にするため、複数列に渡っている値情報（ワイド型）を単一系列に集約して縦長（ロング型）のテーブルに変換する。ただし、この際に互いに足し合わせない数値（例えば内数、「留学生数と社会人学生数」）は複数列のままとする。
- (6) 組織マスタの作成
テーブルを複製して値列等を削除し、組織名と組織コードの重複行を削除して、名称・コードが一对一に対応する組織マスタを作成する。
- (7) その他マスタテーブルの作成
カテゴリに並び順を指定する場合など、デフォルトである文字コード順では見づらいことが多いため、マスタを作成する（例：都道府県名が文字コード順では「その他」→「愛知」→「愛媛」…となる。これを北から順に表示するために、「北海道」＝「01」、「青森」＝「02」…と定義したマスタを準備する）。

(8) Excel ファイルから CSV ファイルへの変換

BI ツールでの読込速度が Excel よりも CSV の方が速いため、保存しなおす。

(9) BI ツールで読み込み

上記で作成した CSV ファイルを、BI ツールにてデータソースとして読み込む。この際、データソースを機構内でなく SharePoint 上（後述）に配置しておくことで、大学担当者とも共有可能となる。また、カスタム関数を設定しておくことで、データソースの変更が容易となる。

(10) DAX 関数による指標の計算

Power BI 特有の関数である DAX 関数により指標を計算する。例えば、「学生数」のうち「性別」カテゴリ＝「女」であるデータを集計し、全「学生数」で割ることで「女性学生の割合」が計算できる。

(11) グラフ、表、スライサーの作成

計算した指標等をグラフ、表等のビジュアルにして配置する。レポート閲覧時に組織やカテゴリを絞り込むためのスライサーを配置する。

以上の作業のうち、(1)～(6)については、新年度のデータが追加される際に繰り返し同じ作業をする必要がある。ただし、(2)～(6)はクエリエディタを用いて作業手順を記録することで、新年度データの追加等にあたっての作業を自動化することができる。また、(8)はマクロで一括変換が可能である。(7)及び(9)～(11)の手順は、新年度データの追加にあたって作業すべきことは多くない。よって、新年度データ追加にあたっては、(1)以外の手順ではそれほど作業の労力はかからないものと期待される。

次にレポート等の共有方法について紹介する。機構が Power BI Premium を契約することで、クラウドサービスを通じて、アカウント数の制約なく大学教職員にレポートを共有できる。また、IR 担当者がレポート等を独自に作成したい場合には、Power BI Desktop のファイル (.pbix) やデータソース (CSV ファイル)、または Excel ファイルを利用することになるため、これらを Office365 の SharePoint により共有する。

なお、これらの環境を構築した上で、大学への提供開始（2019年10月）に先立って、7月に先行的に一部の大学教職員にレポートを閲覧いただき、ユーザビリティを調査した。スライサー機能や、指標追加に関する意見をいただき、必要な改善を施した。

4. 期待される効果 —内部質保証にどのような貢献ができるか—

BI ツールによるレポートを提供する利点として、「誰でも簡単に分析できる」ことが挙げられる。IR 担当者でない教職員であっても、基本的な画面構成と操作方法を理解すれば、スライサー選択するなどの操作だけで動的に可視化された分析を取得できる。

ある部局長に伺った事例であるが、「学長ヒアリングにあたって、自部局の立ち位置や現状についてエビデンスを添えて説明するようにとの指示があった」という。部局長が自部局について客観的なデータに基づいた分析を行おうとしても、元となるデータの入手や分析のための人員の確保には、それなりの困難を伴うだろう。その点では、大学基本情報という素朴なデータではあるが、BI ツールによるレポートがあれば部局教職員でもすぐ

に他大学とのベンチマーキングを行うことが可能となる。

機構の研究会が取りまとめた「教育の内部質保証に関するガイドライン」では、内部質保証における情報収集や分析の体制として、全学的なIR機能の有効性を示しつつも、「質保証のために意思決定を行う主体が現状を把握するために必要な情報を入手でき、その状況が分析できるのであれば、分散的な体制をとることも考えられる」（[4] pp.10-11）とし、全学レベルよりもミクロな主体による分析が有効となる可能性も示唆している。上述のように、BIレポートは、場面に応じて個々の組織や構成員が、エビデンスに即して自大学や自部局を把握することを容易にする。内部質保証システムを有効に機能させるには様々な要素が介在するが、その一場面としてのミクロレベルの主体による分析・自己点検を助けることが期待できる。

他方、大学本部のIR担当者にとっては、データ加工に掛かる手間が減る効果がある。今回の取組ではレポートのみならず元データも提供するため、IR担当者が独自にレポートを作成しなおすことができるので、より発展的な分析に進むことも可能である。以前はデータ加工に手間がかかるという理由で大学基本情報を利用していなかったIR担当者も少なくないだろうが、今後は利用のための障壁が小さくなるため、利用者が増加する（そして望むらくは、各担当者がそこから新たな洞察を得る）ことが期待できる。

最後に、機構内部のことではあるが、3で述べた一連の作業手順が確立されたことにより、同様の手法で他のデータを用いてレポートを提供することも視野に入れられた。大学基本情報では届かない部分（例えば研究費に関するデータ等）をカバーしたデータがあれば、新たに有用なレポートを提供できるかもしれない。機構は公立大学協会と覚書を既に締結しており、同協会が実施する「公立大学実態調査」のデータをレポート化して大学ポータルサイトに参加する公立大学に提供することを目指している。また、機構が収集している、国立大学法人評価に使用するデータをレポート化することも技術的には可能である。

5. 今後の課題 一周知と改良により利用を促進する

レポートや元データを提供することで期待される効果を述べてきたが、当然のことながら、これらは「大学教職員がレポートや元データを利用する」というステップを経て初めて実現する効果である。

機構として、大学教職員の反応をどう想定すればよいか。「提供したのだから使ってくれるだろう」とか、「使う気が無い人がいるのは仕方ない」といった考えでは、到底利用は広がらないだろう。逆に、「提供したけれども届いていないはず」→「いかに周知するか」という方向と、「使う気が無い人がいるなら、使う気が起きるものにするべき」→「いかに改良するか」という方向の2点から利用を促進することが今後の課題となる。

(1) いかに周知するか

提供にあたっては、大学ポータルサイト担当部署に事務文書を送付してアカウントを配付しているが、これは回答責任の生じるような「依頼」ではなく、単なる「通知」の文書である。受け取った部署が、通知を見て瞬時に「これは有益なものだ」と考えてくれればよいが、そうでなければ、学内の各部局や各事務担当課にアカウントを配付する手間をかけることはないのであって、通知はそこでストップしてしまう。

そこで、機構としては積極的に機会を設けて、国公立大学等の教職員に向けてレポート等の存在を周知していくことが必要となる。具体的な方策については未定だが、例えば、各地域ブロックでのセミナーの開催、国立大学協会等の大学団体が行うイベントや部局長等会議での紹介といった取組が有効となりうる。この際、「こういうレポートがある」という周知だけではなく、実際に利用した大学教職員が「レポートを利用して、こういう洞察が得られた」という事例も併せて伝えることが効果的だろう。

(2) いかに改良するか

(1)で述べた周知戦略が幸いにも奏功して、大学教職員がレポートにアクセスしてくれたとしよう。しかし、閲覧したレポートが直ちには興味を惹かないものであったり、使い方や意味がよく分からないものであったりしたら、そこから何か洞察を引き出すには至らず、軽く眺めただけで終わってしまう。

現在は提供開始から間もないため着手できていないが、利用状況のログの解析（誰が、どのレポートを、どのくらい見ているか）や、利用者へのヒアリングにより、反応を見極めることがまず必要となる。また、それに基づき、操作性やビジュアルの改善、レポートの充実等を絶えず試みて、利用者にとって洞察を引き出しやすいもの（使いたくなるもの、欲しい情報が簡単に出てくるもの）にすることが必要である。

さらに、これ以上に解決が難しい問題として、元データの誤りがある。レポートは、各大学から提供を受けたデータから作成しているが、データ誤り（データの収集漏れや異常値）と思われる値が散見される。これは利用者にとって興味を削がれる事態であるため、修正することが望ましい。

一方で、データ誤りについて、機構が提供元の大学に対して誤りを指摘し、差替提出を求めることも難しいという現実がある。大量の業務を抱える大学担当者の仕事をいたずらに増やしてしまうことは、機構としては避けたい。そこで少なくとも当面は、各大学の担当者が自大学のデータ誤りに気づいた場合には差替提出いただき、機構がレポートに修正を反映するという対応にとどめるしかないと考えている。やや楽観的な展望を言えば、(1)周知と(2)改良により、多くの大学教職員がレポートを日々閲覧し、データ誤りが多く発見されることで修正が行き渡る、というサイクルが確立されることが理想である。

6. 結び —データ収集機関がどこまでのリスクを取れるか—

データを収集しフィードバックしている機構のような機関が、未加工のデータを提供するという従前の取組から一歩踏み込み、データを加工してレポート等を提供する（データへの関与の仕方を変化させる）にはそれなりのリスクを取る必要があった。未加工データを提供するのであれば、データ誤りがあっても「大学から提供されたものをそのまま提供しているだけ」という弁明が可能である。しかし、データを加工して提供する場合、機構による加工ミスリスクが生じるし、ミスが実際に明らかになった場合には、機構の責任において必要な対応をする必要がある。

また、レポート等を提供するにあたって、データから指標を作成し、様々な可視化を施すということ自体には、何らかのメッセージが含まれてしまうし、それが誤解を招くリスクもある。例えば、「留学生割合」を指標としてグラフ化したレポートがある。ここには、

「大学が留学生割合を参考にすることがあるだろう」というメッセージが含まれてしまう。実際のところ、機構としては留学生割合を指標としていることにさほど深い意図はないのだが、誤解されて、「留学生割合が重要指標だと機構が考えている」とか「ランキングを作ろうとしている」と受け取られる恐れも無いとは言えない。

今回の取組では、こうしたリスクがあることは認識しつつも、レポート等の提供によるリターンが大きいのではないかと判断した。リスクとリターンの評価にあたっては、大学関係者へのヒアリングによって感触を確かめた。また、付随コスト（データ加工やレポート作成・提供とその継続に掛かる費用や作業量）を見極める必要があった。

一方で、いささか言い訳めいたことではあるが、機構として負にくいリスクも確かに存在し、現時点では実施できない支援もある。例えば、レポート等は大学ポर्टレートに参加する国公立大学等の教職員向けに提供しているが、これを一般公表することは今のところ難しいと考えている。一般公表により私立大学関係者や進学希望者、研究者等がレポートにアクセスできることは一定の利益となるだろうが、意図せぬ受容の仕方（例えば安易なランキング化）やデータ自体の誤りがあることにより、大学に不利益が及ぶリスクが大きいためである。同様に、例えば情報活用のために収集データ項目を新たに追加調査することは、レポートの充実にはつながるだろうが、回答のための負担を大学に掛けることで、大学にとって有益でない取組になるリスクが大きすぎるため、現実的ではない。

このように、様々な要望や、それによって期待されるリターンがあることを機構として認識していても、トレードオフとしてのリスクが存在するために取組を実施できないということが往々にしてある。機構のようなデータ収集機関が情報活用を支援しようとするとき、各大学を取り巻く様々な環境に照らしてリスクとリターンの大きさを評価し、施策を決定することになる。今後も、機構の情報活用支援が大学関係者にとって「スピード感に欠ける」とか「不十分だ」と映るところは多々あるだろうが、こうした難しさを抱えつつ「大学改革支援」に取り組んでいることを理解いただければ幸いである。

【参考文献】

- [1] 大学改革支援・学位授与機構、「大学基本情報」。
<https://portal.niad.ac.jp/ptrt/table.html>
- [2] 大学ポर्टレート（仮称）準備委員会、「大学ポर्टレート（仮称）構築のための論点整理（その2）」、2014年。
https://portraits.niad.ac.jp/files/20160311_ss1.pdf
- [3] 船守美穂・中鉢直宏、「「大学基本情報」のBIツール利用の可能性と課題—複合的統計表のデータベース化と解析に向けて—」、『第6回大学情報・機関調査研究会論文集』、pp.26-31、2017年。
https://mjir.info/download/articles_2017/2017-4.pdf
- [4] 大学改革支援・学位授与機構質保証システムの現状と将来像に関する研究会、「教育の内部質保証に関するガイドライン」、2017年。
https://www.niad.ac.jp/n_shuppan/project/_icsFiles/afieldfile/2017/06/08/guideline.pdf