教学IR室におけるデータ基盤構築と活用

― IRシステムの構築と効率的かつ効果的なIRに向けた取り組み ―

西　出　　　崇

1．はじめに

近年，大学を取り巻く環境がますます厳しくなるなかで，教育の成果や質について，以前にも増して根拠やデータに基づいて検証することや社会的な説明責任を果たすことが求められるようになっている。これに対応するために，大学ではアドミッション・ポリシー，カリキュラム・ポリシー，ディプロマ・ポリシーといった教学の指針を設定したり，教学マネジメントや内部質保証の体制を整備するなど，様々な取り組みが行われている。このような中で，大学内外に存在する教学の過程や成果などに関連するデータを体系的に収集，整理，蓄積し，教学マネジメントや内部質保証をはじめ大学運営において必要な情報提供や分析などの情報活動を一元的に担う機能としてIR（Institutional Research）が注目され，多くの大学で専門の部署や専任の担当者が置かれるようになっている。小樽商科大学（以下，本学）においても2020年２月にグローカル戦略推進センターに教学IR室を設置し，専任の教員を配置して本格的なIR活動を開始した。教学IR室ではこれまでに様々な活動に取り組んできているが，その中で本稿ではIRを効果的かつ効率的に推進するために構築を進めているデータおよび分析の基盤について，その構成や活用状況および課題を整理する。

日本の大学におけるIRは概念として定着しつつあるが，その具体的な機能や目的，対象，活動内容はそれぞれの機関におけるIRの位置づけや体制，学内状況や課題，担当者の専門性や能力などによって様々である。本学の教学

IR室の立ち上げにあたっても，当初から一般的なIRの概念以上に具体的な構想が存在していたわけではなく，データの収集や整理，部署間の調整，分析ニーズへの対応や情報提供などを手探りで行っている。この状況は本稿の執筆時点でも同様であり，現在進行形で進めているものであるが，ここではこれまでに行ってきたIRにおけるデータ基盤整備の主な取り組みについて記述する。

教学IR室の現在までの取り組みは，大きく①IR活動を遂行するために必要なデータの収集や整理，蓄積および分析やモニタリングと情報提供の基盤となる情報システム構築，②執行部や各部署からの指示や依頼，相談，および学内の課題に対応して行うデータ分析や情報提供，③教学マネジメントや内部質保証などを中心とした大学運営におけるIRの役割や位置づけの整理およびそれに関連する規定等の整備の三つに分けられる。本来であれば③のIRの役割や立ち位置を十分に議論したうえで，必要な①データ基盤整備や②分析および情報提供などが行われるべきである。しかし，組織内の情報活動やデータマネジメントの形を大きく変えるIRという概念を導入し，組織において教学IR室の機能や役割を定位させるためには現状を把握したうえで試行錯誤を行うことが不可欠であり，IRの枠組みの整理とデータの整理や分析，情報提供などの実践的活動の両面を相互に行き来しながら並行して進めている。データ収集や分析等の作業を進めるにあたっては，データの取り扱いや各部署の役割分担，規定の整備や権限等で問題があることも認識しているが，現時点では大学執行部の了承を得たうえで関係部署と綿密な連絡・調整を行いながら慎重に進めている。このような状況を踏まえて，以下では教学IR室のデータ基盤整備とその活用について概観していく。

2．学内データの把握・収集・整理とIRシステムの構築

IRを進めるうえで，情報システムとしてのデータベースは必ずしも必要であるわけではないが，例えば浅野(2016)，高田・大石・森(2015)，藤原・大野(2015)，松田(2014)など，IRの先行事例や研究においても必要性が指摘され

ているように，効果的かつ効率的にIRを推進するためには，学内データを統合し一元的に整理，蓄積できる統合型データベースの構築は欠かせない。一般にIR部門の作業において，分析に至るまでのデータ取得や整理・加工の作業は，分析そのものに対して大きな比重を占める場合が多い。例えば，学部ごとに成績が一定水準を上回る学生の割合を集計する場合を想定すると，データが学部ごとに管理されていれば学部の数だけデータ取得の手続きが必要となり，各学部から提供されたデータの形式を統一してデータセットを作成しなければならない。また，成績データは１人の学生についてそれぞれ数が異なる複数科目のデータが存在しており，「成績が一定水準を上回る」といったとき，学生ごとにGPAや平均点を計算したり，計算対象の科目区分などがあれば科目を絞り込む必要があるなど，データ抽出と事前処理に多大な労力を要する。他方でデータセットが手に入れば，場合によっては集計や分析そのものは統計解析ソフトウェア等によって瞬時に行うことができる。つまり，IR部門の作業負荷の大部分を占めるのはデータマネジメントにあるといえる。そこで，本学においても教学IR室が迅速かつ適時的にデータ分析および情報提供を行うための基盤として，学内データを統合して一元的に利用できるデータベースシステムの構築を進めている。

IRを目的とした統合型データベースの構築には，学内に散在するデータの所在や形式，管理の体制や方法の把握，データの取得や管理方法，利用などについて各部局との交渉と調整，データの整理や統合，蓄積，管理方法の検討，情報システムとしてのデータベースやデータ連携システムの構築など，各ステップの作業を複合的に進める必要がある。これらの作業は現在進行形で進めているが，現時点では学籍の基本情報，成績や履修などの教務情報，入試関連情報，就職・進路情報など，主にIRを推進するための中核的なデータの状況を把握し，主要な部分を教学IR室で構築したデータベースシステムに試行的にデータ連携するところまで進捗している。以下ではこのデータ基盤構築の各ステップの進め方や課題について，データ取得の手続きの側面とテクニカルな側面から整理する。

⑴　データマネジメント上の課題と学内データの把握

大学におけるデータ管理のあり方は機関によって様々であるが，表面的には各部署がそれぞれの業務に応じてデータを個別に管理し，必要に応じて部署間でやり取りされる場合が多い[1)]。教学IR室が各種データの状況について各部署に聞き取りを行った結果，本学でも学生の学籍データ，履修や成績データ，入試データ, 就職・進路データ, 留学や海外プログラム等への参加データなどが，それぞれの担当部署で縦割りに管理されていることが明らかになった。これらのデータは担当部署に依頼すれば随時利用することが可能であるが，IR業務においてはこれらのデータに対して横断的なアクセスが日常的に生じるため，IRを迅速かつ効果的に進めるためにも，各部署の業務負担の面でも，データの管理方法や利用手続きを見直す必要がある。

IRによるデータ取得手続きにおける課題として挙げられるのは，データへのアクセスや利用に対する権限や責任の範囲が，現時点では規定等が整備されておらず不明瞭である点である。現状では，データを所管する部署は定められているが，部署間でのデータのやりとりの手続きや利用範囲などについては明確に定められておらず，他部署がデータを利用する場合は担当部署にデータの提供をその都度依頼し，提供の可否が逐一判断されている。また，データの形式についても集計前のいわゆる「生データ」を直接やり取りすることはあまりなく，データを管理する部署がデータの加工や集計まで行い，利用する側が必要とする形式で提供される場合が多い。例えば，データを利用する部署は「語学検定で一定以上のスコアを取得した学生の割合」を要求し，担当部署はデータの集計までを行い提供する。このように，データを利用する部署と管理や加工をする部署が分離し, 業務ごとに断片的にデータが処理されているため, デー

1）現場の業務レベルではそれぞれの部署がデータを個別に管理している形になっているが，データ自体は統一的な学務システム等に一元化されている場合もある。各部署は業務に対応したシステムの機能を利用するだけで，背後のデータ構造を意識しないため，表面的にはデータが担当部署ごとに個別管理さているような状態となっていることがしばしばみられる。

タの構造や関係, 利用目的, 処理や集計の内容などが部署横断的に把握できず, 大学全体としてデータが統合的に活用できていない状態となっている。

このような組織のデータマネジメントの現状に対して, 教学IR室は大学内のデータを統合的に把握し, 定常的にモニタリング, 分析し, 大学運営における各場面で必要に応じて情報提供を行うことが期待される組織である。その機能を十分に果たすためには, 個々の分析ニーズに応じてその都度データを取得するようなデータ管理の枠組みを見直し, IRが学内データを横断的に把握し統合的に利用できるデータ基盤が必要である。そのためには, 各部署が管理するデータへのアクセスの方法や権限, これまで各部署が行ってきた集計や加工などの役割についても整理し, IR機能を念頭に教学IR室および各部署のデータマネジメントにおける役割分担や責任の所在を明確にする組織構造的な改革が求められる。すなわちIRを導入するということは, 単にデータの収集と分析を行う部署を設置するということではなく, 大学全体での統合的なデータの有効活用に向けて, データマネジメント体制や関連する制度, 組織構造を変化させることだといえる。

データマネジメント体制の改革を進めるにあたって, 一般的な観点からあるべき姿を設計することはある程度可能であるが, データをどのような組織的目的に対応してどのように活用するのかは, それぞれの組織の状況や課題に依存する。しかし本学の現状として, データ活用の組織的目的や課題がまだ明確になっておらず, 改革を進めるためにはまず, いつ, どの部署が, どのような目的でデータを利用し, その処理が実質的にどう行われているのかなど, データ利用の現状把握が必要な段階である。そのため, 先述のように現段階では制度設計や規定の整備に先行して, 現場での実質的なデータの把握や整理などの作業と, データ活用の方向性を探るための分析や情報提供が先行している状態である。

このように, 制度や枠組みの整備に先行して実質的なデータ収集や整理の作業が先行せざるを得ない状況にあるが, データの把握や利用にあたっては大学内の個人情報や機密性の高いデータにもアクセスする必要があるため, データの取り扱いにおけるセキュリティや利用の権限等には十分に留意する必要があ

る。この点については，大学執行部の承認を得たうえで，教学IR室長を兼ねる副学長のリーダーシップの下に教学担当副学長などの協力も得ながら，報告・連絡・相談を密に慎重に進めることで対応している。また，これらの作業過程での問題点なども考慮しながら，データの取り扱いやIRの位置づけなどに関連した規定の議論を進めている。

ところで，このように枠組みや制度の整備に先行して，現場での実践的なデータマネジメントの作業やデータ利用を柔軟に進めることができる要因として，本学の規模や組織文化についても指摘しておきたい。小樽商科大学は小規模な単科大学であることから，教員組織および事務組織の構成が比較的シンプルで指示や依頼，情報共有等を比較的スムーズに進めることができる。この点は，一方で業務が属人的になることや責任の所在があいまいになってしまうという問題点をはらむが，他方で各部署の業務に対応して発生，管理されるデータを把握する際には，それぞれの担当者がデータの全体像を把握しており，IRの目的や趣旨を共有しコミュニケーションを密に行うことができるため，学内データの把握という作業においては利点となる。教学IR室が2020年2月に設置されてから約半年程度で学内の主要なデータを把握できたのは，大学の組織的な条件によるところが大きいといえる。インフォーマルな側面に過度に依存するのは好ましくないが，問題点を認識しつつ，それぞれの組織の特性に応じたIRの進め方を模索することも重要だろう。

⑵ IRシステムの構築

IRを推進するための大学内のデータマネジメントの体制や制度の構築については，先述の通り現在進行形で進めているが，主要なデータについて所在や形式，管理方法等を概ね把握したところで，これらのデータを統合的に利用・管理する基盤として，「IRシステム」の構築を並行して進めている。ここでの「IRシステム」とは，データの連携・取得，蓄積・管理，分析，モニタリング，情報発信など，図1に示すようにIRに関連する一連の機能を全体として捉えた概念であり，いくつかのサブシステムから構成されている。これらのうち，こ

こではシステムの中核として構築を進めている統合型データベースおよびそこにデータを投入するデータ連携の仕組みと，蓄積されたデータから自動的に集計レポートを生成するシステムの構成や機能を示す。

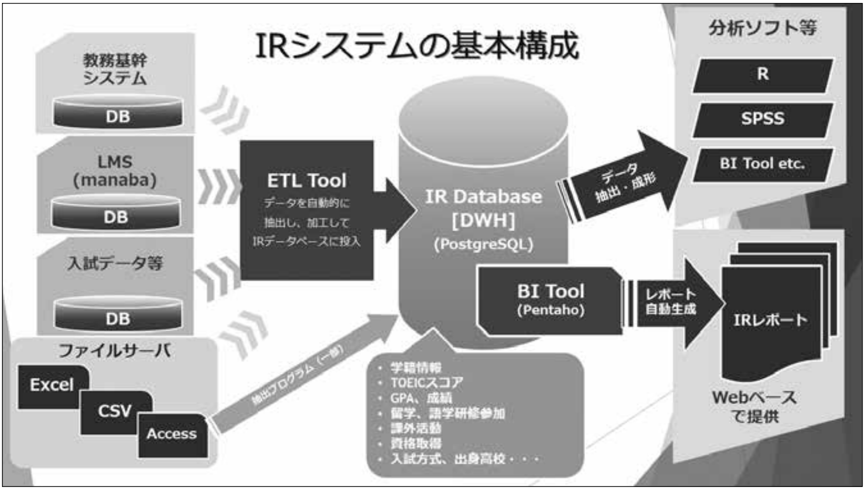

図１　IRを目的とした統合的なデータベース・分析システムの構成

a)　統合型データベース

組織におけるデータの効率的な管理や有効利用のために，統合型データベースを中心とした情報システムを構築することは，IRのみならず，昨今急速に注目を集めている「DX（デジタル・トランスフォーメーション）」の文脈でも欠かせないものといえる。他方で，このようなシステムを開発，導入するためには，コストや技術などの面で課題が多く，財政状況や人員の厳しい小規模大学では容易にシステムを構築することができない場合が多い。本学においてもコストの問題が大きく，システムを専門業者に委託して開発することや，既成の製品を導入することが難しい状況にある。そこで，システム全体を一括して開発するのではなく，機能ごとにいくつかのサブシステムに分割し，それぞれを学内資源やオープンソースソフトウェア等を組み合わせることで教学IR室

として必要なデータ基盤の構築を進めている[2)]。

このような自前でのシステム構築は予算の面ではかなりコストを抑えることが可能になるが，他方で構築を進めるために高度な技術的知識が必要になることや,維持管理の手間の増大,システムの運用が属人化するなどの課題もある。したがって，初期的にIRの体制や基盤の構築を進める段階では有効な方法であるが，システムの有用性や必要性が検証された段階で，学務システム等の大学運営における中核的なシステムと統合して整備するなどの見通しも必要だろう。ただし，大学全体のデータを統合的に把握しデータマネジメントの体制を見直すことと，IRに必要な統合型データベースシステムの設計や開発とは表裏をなすため，既成製品の導入や外部委託で開発を行うとしても，学内データの把握やその活用のあり方，システムの位置づけなどについては大学が主体的に定義する必要がある。したがって，既製品の利用や外部委託によって情報システムの構築や運用におけるテクニカルな部分以外のコストが大幅に削減されるわけではない[3)]。

ここでのIRシステムの中核をなすのは，データを整理して蓄積するデータベースシステムである。データベースにもいろいろな形式が存在するが，学内に存在する多くのデータが一般的な表形式の構造化データであるため，ここではオーソドックスなリレーショナルデータベースを導入した。データベース製品にも様々なものがあるが，オープンソースソフトウェアとして提供されているものでは，PostgreSQLやMySQLといった製品が利用実績も多く有名である。そこでこの2製品に絞って機能等を比較し，IRにおける分析や集計，デー

2) 2020年度の教学IR室の予算は，人件費を除外して全体で30万円である。この予算規模では実質的に商用製品の利用や外部委託等は難しいため，予算の範囲内に収まるサーバ用コンピュータを調達し，ソフトウェアは後述するようにオープンソース製品等を組み合わせて必要な機能やサービスを整備している。また，2021年度は予算が割り当てられていない。

3) 学内データの把握や利用の目的，方法の見通しがなければ，いくら先進的で高機能なシステムを構築，導入したとしても効果的なIRやデータマネジメントの改善にはつながらないだろう。

タの加工などの用途に鑑みて利用できる機能が整っていると判断したPostgreSQLを採用した[4)]。

リレーショナルデータベースの構築にあたっては，格納するデータの構造を分析し，適切にテーブルを設計しなければならない。データの一貫性や整合性を保つためには，データの正規化が欠かせないが，ここでは厳密に正規化を行わずデータ抽出や分析など利用上の利便性を優先して設計した。これは，後述するように多くのデータが学務システムなどのデータベースから直接取得されるものであるため，整合性や一貫性についてはあらかじめ担保されたものとして扱っているためである。データベース設計の観点からは問題があるが，管理コストとの兼ね合いで判断している。また，IR側のデータベースはあくまでも各部署で適切に管理されているデータをそのまま取得することを基本とし，IRデータベースにオリジナルなデータは可能な限り保持しない形で運用している[5)]。つまり，学内に散在しているデータを一つのデータベースに集約し統合的に利用可能にすることを主眼としており，仮にIRデータベースに不具合が生じても失われるデータはほとんどない。

本稿の執筆時点で，IRデータベース上には学籍情報，履修情報，成績，進路，入試やこれらに関連したもので約40のテーブルおよび仮想テーブルが存在している。また，分析や集計において利用頻度が高いデータベース関数もいくつか定義し，IRに特有の処理に対応できるようになっている。例えば，学籍管理の業務においてはほとんど必要とされないがIRでは利用頻度が高い処理とし

4）一般にPostgreSQLは分析系の機能が充実し複雑で大規模なデータ分析に適しており，MySQLはシンプルで高速であることからWebサイトのバックエンドでオンライントランザクションの処理によく用いられる。現在では両製品とも機能が充実し，ここで想定している用途においてはいずれかの製品に決定的な優位性があったわけではないが，window関数など分析に利用できる関数群においてPostgreSQLに優位性を認めたため，こちらを選定した。PostgreSQL: https://www.postgresql.org/，MySQL: https://www.mysql.com/

5）したがって，各部署におけるデータ生成や管理の方法を整備することなどと合わせて，組織のデータマネジメントの全体像を考える必要がある。データの一次的な管理者はあくまでもデータが生成される各部署で，IRはそれらを統合的に管理するという原則は崩すべきではないと考えている。

て，過去のデータとの比較が挙げられる。具体的には，ある年度の２年生と別の年度の２年生の成績を比較するような場合である。この比較を行うためには，ある時点での学籍状態や成績，履修状態を復元する必要がある。このような目的のために，日付を指定してその時点の学籍状態を再現する関数を定義している。その他にも，語学検定のスコアなど一人の学生に対して複数のデータが存在するようなデータについて，学生ごとの最高点や最新のスコアを抽出する関数なども定義し，データ抽出や分析の要望に迅速かつ柔軟に対応できるように機能の整備を進めている。

b）データの取得

次に，各部署で管理されているデータを取得しIRデータベースに展開するデータ連携の仕組みについて述べる。IR部局が各部署からデータを取得する方法はいくつか考えられるが，しばしばみられるのがCSVやExcelなどの形式でのデータのやり取りである。しかし，このような方法ではデータの受け渡しや手続きに手間や時間がかかり，データのバージョン管理や形式の統一，変換などが煩雑であるなど管理上の問題も多い。そこで，ここでは効率的なデータマネジメントの仕組みを構築するための基盤としてETLツールを導入した。

ETLとは「Extract（抽出）」「Transform（変換）」「Load（格納）」の頭文字で，データソースからデータを抽出し，必要に応じてデータを変換・加工し，目的のデータベース等にデータを格納する一連の作業を自動的，定型的に処理するものである。ETLを実現するソフトウェアにも多様な製品が存在するが，ここではコストの制約からオープンソースソフトウェアとして提供され無償で利用できる「Pentaho Data Integration（Community Edition）[6]」を利用してい

6）Pentaho Data Integrationとは，Hitachi Vantaraが開発しオープンソースソフトウェアとして公開している統合的なBI（Business Intelligence）ツール「Pentaho」の一部で，ETLを担うソフトウェアである。PentahoはETLの他に，OLAP（OnLine Analytical Processing），レポート生成，ダッシュボード構築，データマイニング等の包括的な機能を有している。Pentaho: https://sourceforge.net/projects/pentaho/

る。Pentaho Data Integrationは，様々なデータベース製品に接続することができ，ExcelやCSV，Accessなどからもデータを抽出することができる。ETL工程の定義は，Javaベースのデスクトップアプリケーションを用い，データの取得や加工，展開などの機能に対応したアイコンを並べ，データの流れや加工処理に応じてそれらを接続していくことで，直感的に行うことができる。図２は，学務システムのデータベースに接続して学籍データを抽出し，所属情報の結合や文字種の統一，データ形式の変換，テスト用データの除外などを行い，IRデータベースに投入する一連の処理を定義した例を示している。このように定義したETL処理は，IRデータベースと同一のサーバ上に実行環境を構築し，定時実行や何らかのトリガで実行するように設定し，定期的・自動的にそれぞれのETL処理が実行されるようになっている。

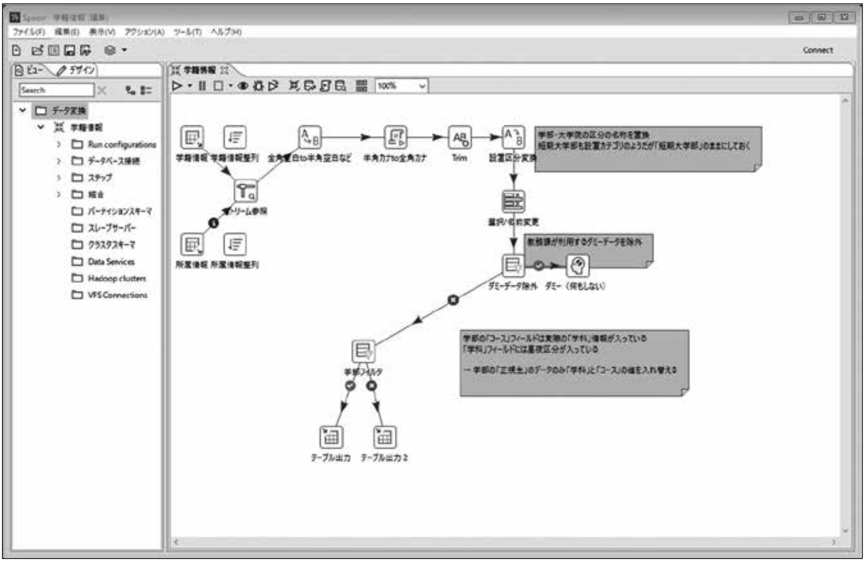

図２　ETLの設定例

データ取得においてETLツールを介在させることで，IRと各部署の間でデータの受け渡しの手続きを定型化・自動化することができる。各部署との間

で，データの形式や保管場所等についてルールを定め，部署の業務の一環としてデータを整理してもらうことで，部署ごとに管理され散在しているデータを定型的な手順でIRデータベースに統合することが可能となる。この仕組みを機能させるためには，部署内のデータ管理方法や関連する業務手順の変更なども伴うため，部署ごとに交渉や手順の整理が必要となる。IRの基盤となる統合的なデータマネジメント体制の確立においては，ETL処理の定義などテクニカルな問題よりも，このような部署との交渉や業務フローの見直しなどが本質的な作業となる。

現在は，各部署で管理しているデータの把握を進めるとともに，データの管理方法の改善や変更の交渉を進めている。取得するデータのうち，これまで各部署で依頼に応じて学務システム等から抽出し，加工や集計を行っていたものについては，大元のデータソースが一貫性の担保された学務システムのデータベースであるため，直接データベースに接続してIRデータベースに連携している[7]。学内の中核的なデータの大部分は何らかの業務システムで管理されているため，データベースから直接データ連携する仕組みを整えることで，IRに必要な主要なデータはほぼ統合することができる。また，これまで各部署で行われていた抽出や加工のステップを無くすことができるため，データ取得や加工における部署側のミスや負荷が低下する。

他方で，各部署がExcelファイルなどで管理しているデータについては，ETLによるデータ連携の仕組みに落とし込むまでに大きなハードルがある。まず，Excel等のデータは年度や担当者によって微妙に形式や項目が異なる，最低限の正規化が行われていない，セルが結合されるなど二次元の表になっていない，項目が横方向に追加されたいわゆる「横持ち」の形式になっているなど，データベースに取り込むことができない状態である場合も多い。また，場合によっては「データ」として提供されるものが，いわゆる「生のデータ」で

7）学務管理システムのデータベースに接続するにあたってはセキュリティに十分留意し，システムのベンダーの協力を得て行っている。

はなく「集計表」であるなど，「データ」の認識自体がIR側と齟齬が生じている場合もある。このように，部署内で管理されているデータについては根本的に管理方法や保管形式を見直す必要があることがほとんどである。

各部署で管理されるデータが，データベースに格納するのに適さない形で管理，運用される原因はいくつも考えられるが，一つは直感的にデータを集計したり加工する場合など人間が取り扱いやすいデータと，リレーショナルデータベースに保管されるデータの形式が異なることと，データ処理の知識やスキルが不足していることが挙げられる。各部署の業務に求められるのは，例えば会議資料に掲載する集計表などデータを加工した具体的な成果物であるため，日常的な業務の遂行において扱いやすい形式でデータが生成，管理される。他方で，データベースでは一貫性や整合性が担保され再利用性が高い汎用的な形式でデータが保管されるため，本質的には同じデータであるにも関わらず，両者にとって取り扱いやすい形式に齟齬が生じる。そして，データベースに格納される汎用的なデータ形式から，効率的に業務上必要な形式に加工したり集計する方法を現場の職員が持たないため，日常的に利用する成果物に近い形式でデータが管理され非効率な状態となっている。これを解消するためには，現場におけるデータの取り扱いや運用に関する認識，データ処理のスキルの向上，業務プロセスそのものの改善などが求められる。この点は主に組織改革やSD（Staff Development）の課題であり，教学IR室の業務の範疇を超える部分でもあるが，大学におけるデータ基盤やデータマネジメント体制の確立とも密接に関連することから，教学IR室としても対応しなければならない課題だと考えている[8)]。

ここでは組織内のデータ管理やフローを，ETLを媒介に定型化・自動化する形で整備したが，データの取得のみに関心をもつならば，IRが指定した形式のデータを各部署で追加的に作成してもらうという形で進めることも可能で

8）同様の問題意識での取り組みとして，東京工業大学のIR部門の事例が挙げられる（今井・森・富樫, 2020）。

ある。しかし，IRの導入がそれぞれの部署の業務負荷の増加になるならば，本末転倒である。そのため本学の教学IR室では，データの取得だけではなくデータの発生源である部署の業務改善を含めたデータ管理方法の見直しまでを視野に，各部署と協働して学内データの統合作業を進めている。

3．IRシステムを中核としたIRの取り組み

以上のように，本学の教学IR室では部署の設置からこれまでに，IRを目的とした統合型データベースを構築し，学内データの把握と整理を進めてきた。この作業は現在も引き続き行っているが，学籍情報や科目履修，成績，入試，進路等の主要なデータについては，試験的な段階ではあるが概ねIRのデータベースに集約できており，教学IR室の情報提供や分析において中核的な役割を果たしている。ここでは，このデータ基盤に基づいた教学IR室の情報提供や分析の取り組みを示す。

⑴　IRシステムによるデータの自動集計

これまでに述べたように，IRを目的とした統合型データベースやETLの仕組みを構築することで，学内に散在している各種データが一定の間隔で自動的にIRデータベースに同期され，常に最新のデータが一元的に利用できるようになる。学内データの集約が自動化されたことで，各種分析ニーズに迅速に対応できるとともに，データの効率的な活用を進めることが可能になる。IRにおける分析や情報提供は，大きく分けると個々の課題に対応して行われるものと，例えば成績分布や退学者数，入試の志願者数など，アセスメントポリシーや年間スケジュールなどに対応して定期的かつ定型的に集計しモニタリングするものがある。このうち，データを常に一定の手順で処理する後者は，これまで担当部署の職員が手作業で集計を行っていた。定型的なデータ集計を手作業で行うことにはいくつか問題がある。まず，毎回同じ手順の処理を手作業で繰り返す点が効率性の面で問題である。また，集計目的に照らして処理方法があ

らかじめ定まっているが，手作業では毎回処理が微妙に異なることや，担当者の交代で処理内容が変わってしまうこともある。その他にも，処理手順に再現性がないため，ミスや間違い，集計方法を検証することが難しいなどの問題がある。そこで，このような定型的な集計について，IRシステムによって自動生成して提供する仕組みの構築を試行的に進めている。

必要なデータをデータベースから抽出し，一定の手順で集計を行い所定のフォーマットで集計レポートをオンデマンドで自動生成するために，ここでは先述のオープンソースソフトウェアとして提供されている統合的BIツール「Pentaho」のレポーティング機能を用いてシステム構築を進めている。集計レポートの自動生成においては，まず所定の集計におけるレポートの定義として，データベース接続情報，データ抽出や集計手順，レイアウト，データ抽出や集計に対してユーザが指定できるパラメータ等を設定したレポート定義を作成する。これを「BIサーバ」上に展開すると設定したレポートがWebベースで提供され，ユーザがそれぞれのレポートにアクセスすると，オンデマンドでデータベースに蓄積されたデータからレポートが自動生成される。またBIサーバはユーザ管理やアクセス制御機能を備えており，利用者ごとに適切に閲覧権限などを設定することも可能である。

現時点では，BIサーバを構築していくつかの部署の職員に対して自動生成レポートを試験的に提供している。BIサーバは，情報システム部門の協力を得て大学の認証基盤に接続しており，各ユーザは他のシステムと共通のIDとパスワードで利用することができ，権限に応じたレポートを閲覧することができる。レポートへのアクセス画面および生成されたレポートの例を図３に示す。図３左のように教学IR室のWebサイト上に自動生成レポートへのリンク一覧があり，そこから各種レポートにアクセスするとユーザ認証を経てレポート画面が表示され，必要なパラメータを指定すると図３右のような集計レポートが生成される。例として示しているのは，商学部の昼間コースにおける学年別のGPA分布を集計する自動生成レポートである。

これらのレポートは基本的にPDF形式で生成され，ダウンロードすればそ

のまま会議資料等に利用することができる。またPDF形式以外にも，CSVやExcel等の形式でも出力が可能で，集計レポートではなくデータベースから業務に必要な形式に抽出・加工したデータの提供も行うことができる。現段階では，集計方法や値の正確性などが十分に検証されていないため参考情報として提供しているが，将来的には大学の基本的な情報を網羅し，正式な情報として利用できるようにするとともに，各部署の効率化，省力化など業務の改善にも結び付けたいと考えている9)。さらに中長期的には，このような仕組みを通して各部署の業務を「データ中心」に変えるDX（デジタル・トランスフォーメーション）にもつなげていくことを目指す。

このように学内データの定型的な集計・分析やそのモニタリングは，教学マネジメントや内部質保証，アセスメントポリシー等との関係において，そもそもIRの重要な役割の一つである。現状ではこれらの集計や分析を各部署や各種委員会などが個別に行っているが，データの統合と合わせて集計や分析，情報提供における教学IR室の役割や各部署との関係も整理する必要がある。各部署等の要請に応じて個別的にデータ集計や分析の支援を行うとともに，教学IR室は大学全体としてデータが有効に活用されるための中心的な役割を果たさなければならない。その意味では，ここでのIRシステムによるデータの統合と自動集計は，単に業務の効率化や省力化の取り組みではなく，学内データの共有や流通の「ハブ」として教学IR室を機能させるための試みとして位置づけられる。

9）ここでの改善のポイントは，「データ」と「処理」を分離し，「処理」を明示的にレポート設定ファイルとして記述した点にある。手作業では「処理」が常に一過性の再現性に乏しいものであるのに対して，レポート設定としてデータ抽出や集計の手順が記述されることで，処理の内容を事後に検証することができ，同じ手順を正確に何度でも適用することが可能になる。また，処理手順を記述するためには，従来の手作業で行っていた処理内容を整理する必要があり，そのプロセスで集計の問題点や曖昧だった点を発見し，改善することにもつながるなど，副次的な効果も見込める。このような業務の改善については稿を改めて議論したい。

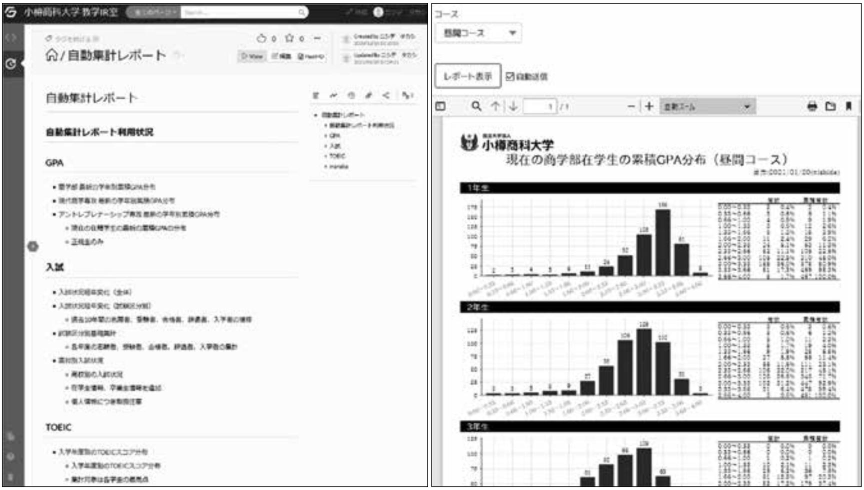

図 3　自動集計レポートのダッシュボードとレポート例

(2)　統合型データベースを中核とした迅速かつ効果的なIR情報の提供

教学IR室が設置されて日が浅く，加えて新型コロナウイルス感染症の世界的流行という特殊な事情のため，教学IR室の学内的な認知がまだ低く通常のIR活動の中心となる調査や分析は十分に行えていないが，執行部の指示や部署からの分析依頼にも徐々に対応している。これらの分析を進めるにあたっても，IRシステムとして構築した統合型データベースや分析環境が効率的かつ適時的な情報提供につながっている。特に，新型コロナウイルス感染症対策における危機対策に関連するものについては，判断や意思決定のために迅速な対応が求められるが，指示を受けてからデータの抽出，分析，結果の報告までを短時間で行い，効果的な情報提供を行うことができた。ここではコロナ禍におけるIRの取り組みとして，授業のオンライン化に関する教学への影響検証の事例を示し，教学IR室の役割やデータ基盤の重要性について述べる。

2019年末に中国で最初の感染者が報告されて以降，新型コロナウイルス感染症が世界規模で急速に流行拡大し，パンデミックといわれる事態に陥ったのは

周知のとおりである。感染症の流行拡大の影響を受けて，感染症拡大防止対策として多くの大学では2020年度の授業を何らかの形でオンライン化しており，本学においても前期はすべての科目を，後期は一部の科目を除き授業をオンラインで実施した（田島他，2020；田島他，2021)。詳細は本稿の主題ではないため踏み込まないが，十分な準備期間が取れないままに授業方法をオンラインに変更したため,教育の成果に対して影響があることは当初から懸念があった。これを受けて，2020年度前期が終了した段階で危機対策本部から教学IR室に対して教育成果への影響の検証の指示があり分析と情報提供を行った（西出，2021a；西出2021b)。以下ではこの事例を通してIRにおけるデータ基盤の有用性の一端を示す。

授業のオンライン化による教育成果への影響は様々に考えられるが，この時点で利用可能であったのは成績や履修のデータのみであったため，検証の対象を成績に絞り成績分布の変化を通常の対面授業が実施されていた過年度と比較することで検討を行った。分析の方法自体は成績分布の年度間比較というシンプルなものであるが，この作業のためのデータの抽出や加工には煩雑な作業を要する。例えば検証のひとつとして，2020年度の前期における学生ごとの成績評点の平均値の分布を過去４年間のデータと学年ごとに比較したが，学年別に比較するためには各年度の前期の時点での名簿が必要となる。しかし学務システムに「学年」の項目は存在しているが，常に最新の学年が表示されるため2020年度前期時点の学籍情報は抽出できても，過去の任意時点の名簿は基本的な機能として抽出できない。そのため，入学年度や進級情報を加工してデータを作成する必要がある。また，休学などの学籍状態についても考慮する必要がある。さらに，各年度の前期の成績情報を抽出し，学生ごとに評点の平均値を算出したうえで，各年度の学生名簿と結合しなければならないなど，データの加工が非常に煩雑になる。

これに対して，これまでに述べたように教学IR室では統合型データベースを構築し，IRに必要な機能を整備しているため，迅速にデータを抽出・加工することが可能であった。各年度の学籍情報については，任意の日付を指定すれば

その時点の学籍状態を復元する関数をデータベースに用意しており，各年度の５月１日時点の名簿に基づいて検証を行った。また，成績の抽出と平均値の集計についても，データベース上でSQLを用いることで容易に処理することが可能である。従来ならば，担当部署がこれらの作業を手作業で実施する必要があり多大な手間と時間を要したが，データベースを整備することによって，実質的には作業開始から１時間程度で必要なデータセットの作成が可能になった。

IRの役割は，データや分析結果に基づいて判断や意思決定を支援したり，そのための情報共有やコミュニケーションを促進することにあるため，情報提供の迅速性,適時性とともに,情報を効果的に提示することも重要である。データの提示においては，例えば統計の知識など受け手のスキルに依存する部分も大きいため，できるかぎりわかりやすく直感的に把握・理解できるように，高度な統計分析手法を用いることや精緻な数値を示すことなどよりも，場面に応じて「可視化」や「表現方法」にも留意し工夫を重ねている。

例えば，各年度の学年ごとに学生の成績評点の平均値の分布を比較するために，平均や標準偏差といった集計値のみを示すのではなく，以下に示すようにデータの分布を直感的に把握できるようなグラフを用いた。また冗長であるように思えるが，同じデータに対しても複数の表現方法を用いて示すことで，短時間でも効果的にデータの特徴を読み取れるように報告を行った。

図４は各学生の評点の平均値の分布を箱ひげ図で示し，その平均値を「+」記号で示したものである。これと併せて，図５では同じデータをヒストグラムで示し，中央値や最頻値の位置も線分で表示した。これによって，オンライン授業が行われた2020年度の成績分布は，通常授業が行われていた過年度とは明らかに異なり,他の年度よりも高くなっていることが一目で把握できる。また,箱ひげ図の箱の中の線が示す中央値と「+」マークが示す平均値が乖離していることから，分布が左右対称ではないことが示唆される。これは，図５のヒストグラムからも確認できる。箱ひげ図では，分布の範囲を捉えやすいが，ヒストグラムでは度数の多い得点域や分布の形が捉えやすい。また，ヒストグラムに示した平均値が度数の多い得点域と乖離していることから，平均値や標準偏

差だけでは成績の状態が十分に把握できないことも理解できる。

繰り返しになるが，IRの役割は，意思決定や状況判断のためにデータに基づいて認識共有やコミュニケーションを促進することであるため，このような直感的でわかりやすいデータの提示は欠かせない。これにより，オンライン授業によって成績分布に何らかの影響があることが危機対策本部会議のメンバーにおいてスムーズに共有され，その要因を探るために教員に対して成績評価に関するアンケート調査を実施するなど次のステップの検証にもつながっている。

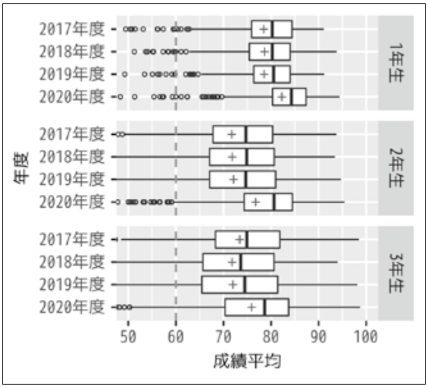

図４　データの可視化事例１「年度・学年別の平均成績の分布」（箱ひげ図）

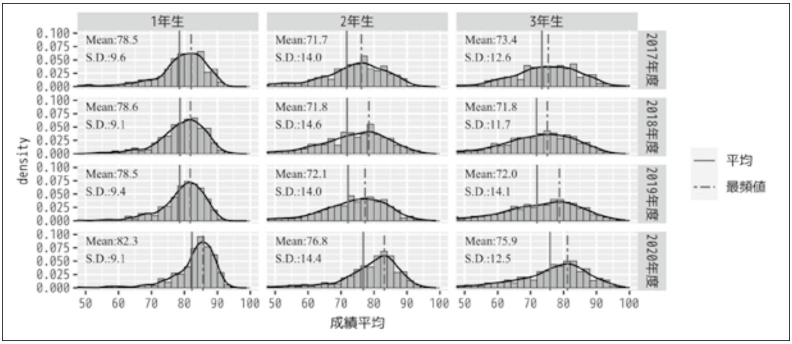

図５　データの可視化事例２「年度・学年別の平均成績の分布」（ヒストグラム）

全体的な成績分布の変化を踏まえて，その要因を検討することなどを目的に個別科目の成績分布の変化についても検証を行った。ただし，各レベルの意思決定や状況把握に必要な情報の量や範囲，集計単位を考慮した適切な情報提供を行うことが重要であることから，個別科目の状況については全体の状況把握や判断を行う危機対策本部会議では詳細を提示せず，教務課などの関係する部署における現場レベルへの情報提供とした。

個別科目の成績の検証については，過年度から同じ教員が同一科目を連続して担当する科目を対象に，図６に示すように年度ごとの成績分布を箱ひげ図と平均値，各学生のデータポイントを重ねてプロットして確認した。この科目では，2020年度の成績が過去３年間の平均的な成績水準よりも高い傾向がうかがえる。このような図を全ての対象科目について作成し，2020年度の成績分布の変化を確認した。本稿の主旨とはずれるため詳細に言及しないが，全体の傾向としては2020年度の成績が他の年度よりも上昇している科目が比較的多い一方で，ほとんど変化がない科目や毎年度傾向が異なる科目など担当者や科目によってバリエーションが見られた。これを踏まえて，何名かの教員に成績評価方法などについてヒアリングを行い，成績評価に関する教員アンケート等の参考にした。

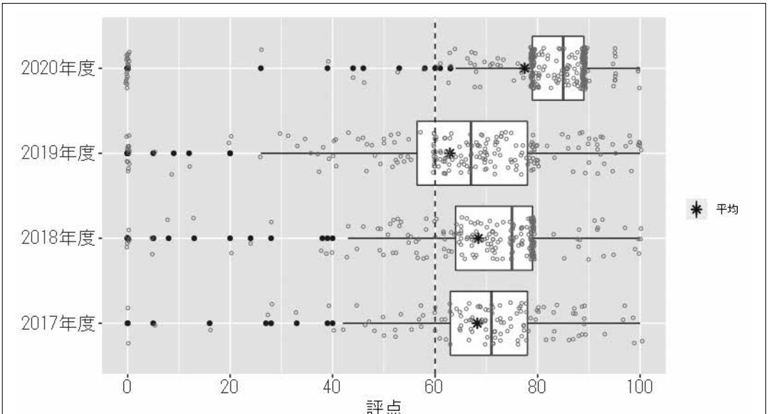

図６　個別科目における成績分布の年度間比較のために作成したグラフの例

以上のような検証において，IRが構築を進めているデータ基盤および分析システムが迅速かつ適時的な情報提供や分析に重要な役割を果たしている。前述のように，ここでの検証において分析自体はさほど高度なものではないが，データセットの準備やデータの可視化，資料の作成等においては，従来の方法では非常に手間の多い作業となる。データセットの準備については，先述のとおりデータベースを整備することで非常に効率的に行うことが可能になっている。分析と可視化については，統計解析用プログラミング言語であるRとその統合開発環境であるRStudioを利用し，RからIRデータベースに直接接続してデータを取得し，分析やグラフ作成等を行ったうえで，会議等に用いる資料をPDF形式で直接出力できる環境を用意している。この環境では，これらの一連の作業をすべてスクリプト形式でプログラムとして記述するため，分析の再現性や手順の検証が容易であり，効率化につながっている。

この方式のメリットは，データの抽出や加工，分析，グラフ作成，レポート生成をすべてスクリプトとして記述し，「データ」と「手順」を分離して手順を再現可能な形で定式化するため，異なるデータに同じ処理を正確に適用するなど処理の再利用性が高い。また，手順が記述されることで，分析方法の間違いや手順を事後に検証したり，データの処理方法を確認できる。さらにプログラミング言語であるため，繰り返しや条件分岐などの制御構造を利用することで，同一の処理を多数の対象に連続的に適用したり，集計表を定期的に自動生成することなども可能である。例えば，先述の個別科目の成績分布の検証においては多くの対象科目について図6に示したグラフを生成したが，ここでは対象科目を切り替えて繰り返し処理を行うことで効率的に資料を作成し，素早く確認を行うことができた。

ここでのコロナ禍におけるオンライン授業の成績への影響検証の事例では，危機対策本部会議からの指示をうけて以上のような作業をほぼ1両日程度で実施し，翌週の会議で報告を行うことができた。具体的には，まず危機対策本部会議の事務局と検証内容や方向性について打ち合わせを行い，その日のうちにデータ抽出と基本的な集計と可視化を行ったうえで事務局にフィードバック

し，事務局と協議して危機対策本部会議で報告する内容を固め，追加的に必要な集計や分析を行った。限られた時間の中で検証を進めなければならない状況において，これらをすべて手作業で行うことは難しいだろう。このように，迅速性が求められる分析ニーズに対して，効率的かつ適時的に効果的な情報提供を行うことができたのは，学内データを統合型データベースに一元化し，それを利用して効率的にデータを抽出して分析を行い，グラフや会議資料等の生成を行う環境を整えていたためである。このようなデータベースや分析システムがあれば直ちに効果的なIRを行うことができるわけではないが，少なくとも効率的なIRを行うための必要条件であるといえるだろう。効果的なIRを行うためには，まず必要な時に必要なデータを迅速に利用できることが大前提である。

ここで紹介したのは一つの事例であるが，教学IR室では学内データを横断的に把握し，IRのためのデータ基盤やデータマネジメント体制，統合的な分析環境を整備することで，分析課題に対して迅速かつ効率的に対応するとともに，効果的な情報提供を行うことができるような体制を構築している。効率的かつ効果的なIRの推進のためには，地味で地道なデータマネジメント体制やデータ基盤の構築と整備が欠かせない。また，データ基盤整備とともにIRを有効に機能させるためには，このような具体的でわかりやすい実践事例や成果を発信し，大学執行部や各学科，部署等にIRについて理解してもらうことが重要である。学内データの整理や体制整備は現在進行形で進めている作業であるが，このように現時点でも一定の成果を上げており，こうした成果を示しながら各方面の協力や理解を得て教学IR室の機能や役割を確立していくことが，現在の教学IR室の取り組みの大きな柱である。

4．むすびにかえて

小樽商科大学は2020年２月に教学IR室を設置し，本格的なIRを開始した。本稿では教学IR室の設置からこれまでの活動のうち，IRの基盤となるデータマネジメントの体制やデータ基盤の確立に向けた作業とその到達点を中心に整

理を行った。また，新型コロナウイルス感染症対策に係る危機対策において，IRのデータ基盤を用いて迅速かつ効果的な情報提供を行った事例から，その有用性や必要性を示した。

IRの役割や機能は組織の状況や課題などによって異なるが，本学の教学IR室がどのように活動を進めていくかは，現在も様々な取り組みを行いながら模索しているところである。ここではコロナ禍における危機対策の事例を取り上げたが，こうした個別的な課題への対応もIRの重要な役割であるが，本来の役割である教学の内部質保証や教学マネジメントにおいてどのような役割を果たすのかが今後の課題である。これは，教学IR室だけの問題ではなく，大学組織の中のIRの位置づけや役割をどう定義していくかという問題であり，本学のIRのあり方そのものを大学全体として議論しながら確立していく必要がある。学内データを統合するデータ基盤は，そのIRの役割を議論する上での基礎であり出発点となるものである。

参考文献

［1］浅野茂（2016）「データベースの構築とIRの課題」『高等教育研究』第19集，pp.49-66

［2］今井匠太朗・森雅生・富樫勝彦（2020）「学内データインフラ整備にむけた業務改善の取り組み ―Ranabaseを使った業務フローの可視化―」『第9回 大学情報・機関調査研究集会 論文集』，pp.148-153

［3］高田英一・大石哲也・森雅生（2015）「大学におけるIRを目的とするデータベースの運用の現状と課題 ―国立A大学の『大学評価情報システム』の事例を中心に―」『教育情報研究』31巻3号，pp.51-60

［4］田島貴裕・大津晶・西出崇・藤原健祐・佐野博之（2020）「オンライン授業の導入へ向けたFDの取り組みと課題」『PCカンファレンス北海道2020論文集』，pp.25-28

［5］田島貴裕・大津晶・西出崇・藤原健祐・佐野博之（2021）「全学的なオンライン授業の導入へ向けたFDの実施と課題―小樽商科大学の取り組み事例―」『コンピュータ&エデュケーション』Vol.50，pp.20-23

［6］西出崇（2021a）「コロナ禍における授業のオンライン化と成績」『大学教育学会第43回大会 発表要旨集録』，pp.185-186

［7］西出崇（2021b）「コロナ禍における授業のオンライン化の教育に対する影響検証の試み ―成績分布の変化とその要因の検討―」『第10回 大学情報・機関調査研究集会 論文集』，pp.54-59

［8］藤原宏司・大野賢一（2015）「全学統合型データベースの必要性を考える」『大学評価とIR』第1号，pp39-47

［9］松田岳士（2014）「教学IRの役割と実践事例 ―エビデンスベースの教育質保証をめざして―」『教育システム情報学会誌』Vol.31 No.1，pp.19-27