



## (11) 組織の意思決定のためのデータ共有の仕組みの構築

学習教育センター 教授 大石 哲也

### 1 はじめに

九州工業大学では2003年度に紙ベースでの自己評価を開始し、2005年度に情報工学部でJABEE認定以降、自己評価を全学展開し、2007年度には学修自己評価システムの開発を始めた。2009年度には正課外活動も記録できるeポートフォリオを全学的に稼働させ、現在に至るまでシステムの改良を続けながら全学に定着させてきた<sup>(1)</sup>。このように教学データが集められ個人にフィードバックする仕組みが存在している一方で、本学では組織として意思決定をするためのサポートの仕組みは存在しておらず、2022年度より教学IR (Institutional Research) の仕組みを強化すべく本学の学習教育センターに教学IR支援グループを置くこととなった。

学習自己評価システムだけでなく、学内にはさまざまなデータが存在しているが、これらのデータを活用するための統合的なシステムは存在せず、各担当部署によって管理している。そこで、本学ではまずIRに関する最初のシステムとしてデータを結合して活用するためにETL (Extract Transform Load) ツールを導入した。続いて、集中レポジトリとしてデータレイクシステムを導入した。

本稿ではIRの概要、ETLツールの導入やデータレイクの構築など、本学で推進しているIRシステムについて紹介する。最後に、データウェアハウスやビジネスインテリジェンスツールの導入も視野に入れた今後の展開について述べる。

### 2 IR (Institutional Research) とは

近年の18歳人口の減少の影響もあり、多くの高等教育機関において各々の機関を存続させるためにIR (Institutional Research) が注目され、各機関においてIRの組織や機能を備えるようになった。IRは決まった遂行の仕方が存在しないが、よくモデルとして採用される情報支援サイクル<sup>(2)</sup>という考え方がある。情報支援サイクルは以下の5つのフェーズを繰り返して組織の意思決定をサポートする一連の流れである。

- ・ 課題・ニーズの特定
- ・ データ収集・蓄積
- ・ データ再構築・分析
- ・ データ報告
- ・ 意思決定

このサイクルは何かしらの意思決定をした後に再び課題・ニーズの特定のフェーズに戻り新たな意思決定に繋げることを意味する。これらのフェーズの内、「データ収集・蓄積」、「データ再構築・分析」、「データ報告」がIR実務者の業務であり、「課題・ニーズの特定」、「意思決定」は組織を動かす大学執行部等の業務である。IRはステークホルダーの意思決

定をサポートするための仕組みであるが、特に教育データを扱った意思決定をする場合を教学IRと呼ぶことがある。

上記により、IR実務者はデータを扱うことが主たる業務であることがわかるが、ここにデータ管理を含まないことが多い。データ管理はデータが発生する業務に依存しており、それらの業務はそれぞれの業務に特有のデータの扱いがあることから、IR実務者は管理には関わらずにデータを提供してもらうだけに留める。

以下、IR実務者の関わる業務について説明する。

## 2.1 データ収集・蓄積

高等教育機関に存在するデータを分類すると、テキストなどの非構造化データと数値等で表現される構造化データが存在している<sup>(3)</sup>。非構造化データには原著論文や学則などがあり、主にテキストデータで構成されている。また構造化データには競争的外部資金や学生アンケート調査などの数値データが含まれる。IR担当者はこれらのうち各担当部署が管理している構造化データをCSV等のファイルやデータベースを直接参照するなどしてデータレイクと呼ばれるシステムにデータを収集して蓄積する。

## 2.2 データ再構築・分析

データレイクに集められたデータをETL (Extract Transform Load) ツールを用いて結合し、データ報告のために分析しやすいように再構築する。この時、分析しやすくなったデータをデータウェアハウスと呼ばれるシステムに格納する。データウェアハウスにはテーマごとに整理され、統合され、時系列で、削除されず、更新されないデータが保管される。つまり、組織の意思決定に資するデータを保管している。IRの経験が浅い場合、データ再構築の部分に関心がない傾向があるが、特にデータ再構築に利用するETLツールについてはIRの経験が長くなるほど、その重要性が増す傾向があることがわかっている<sup>(4)</sup>。

## 2.3 データ報告

データウェアハウスに蓄えられたデータを用いて必要に応じてグラフ等で可視化して意思決定者に対して報告する。データを報告する際には、理解を深めるために様々な視点からデータを可視化する必要がある。多角的に可視化するためにデータウェアハウスに格納するデータは様々な要素を含む必要がある。報告のリアルタイム性を上げ、データの可視化を効率化するためにBI (Business Intelligence) ツールを導入することもある。BIツールにより意思決定者がより迅速に意思決定を下せるようになり、最終的に大学の機能向上につながる。

## 3 九州工業大学における教学IR

本学の学習教育センターでは学修自己評価システムやコースポートフォリオにより教育に関するデータを収集し、教員や学生の個人的な意思決定をサポートしてきた。しかし、執行部等が組織的な意思決定をサポートする仕組みは存在していなかった。そこで、学習教育センターではIRに力を入れるために教学IR支援グループが設置された。

### 3.1 初期のIRシステム（2022年度）

本学では、まず、2.2節で述べたETLツールを導入し、学内に存在する教学データを結合して分析しやすい形にすることから開始した。2.1節で述べたデータレイクを用いたデータ収集・蓄積や、2.2節で述べたデータウェアハウスを用いて再構築したデータを格納する仕組みがあるのが理想であるが、本学ではETLツールが稼働しているサーバにこれらの機能を包含することで当面のIRシステムの機能を実現した。

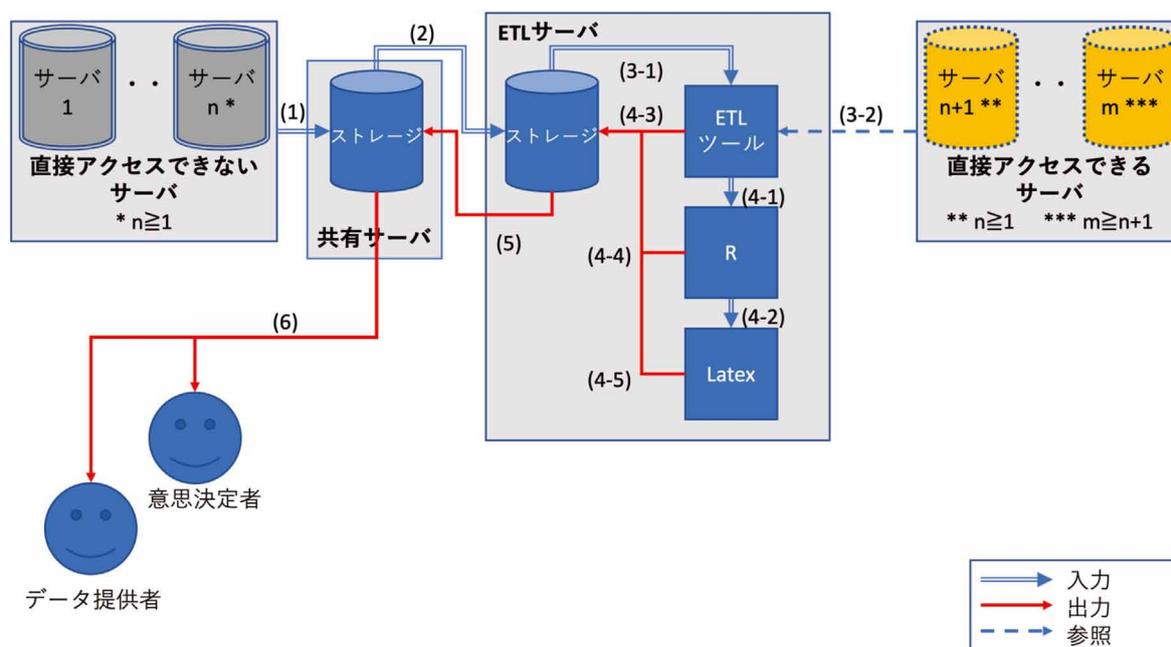


図1 初期のIRシステム

ETLツールを利用して結合されたデータを、統計解析ソフトRを用いて半自動的に可視化する仕組みを導入した。執筆時には教学IRの観点から成績分布や退学状況の半自動的に可視化を実現し、組織としての意思決定に資する情報を提供している。

半自動でデータを可視化する仕組みを実現するために、図1のような初期のIRシステムを導入した。まず(1)データ提供者が直接アクセスできないサーバのデータを共有サーバに転送し、(2)IR実務者がこれらのデータをETLサーバ内のストレージに手動で格納する。続いて(3-1)ETLサーバ内のストレージのデータだけでなく、(3-2)直接アクセスできるサーバのデータもETLツールを使って結合する。これらのデータは、(4-1)統計解析ソフトRで解析・可視化し、(4-2)LaTeXでPDF化する。その後、(4-3)ETLツールで結合されたデータ、(4-4)Rで可視化されたデータ、(4-5)LaTeXでPDF化されたデータは、ETLサーバ内のストレージに格納される。そして、(5)IR実務者が共有サーバに保管する。(6)データ提供者や意思決定者は、報告されたデータを共有サーバから取り出すことができる。

### 3.2 現状のIRシステム（2023年度）

本稿執筆時点では、データレイクサーバは導入済みだが、データウェアハウスサーバとビジネスインテリジェンスサーバは導入していない。したがって、データレイクサーバはデータウェアハウスサーバとビジネスインテリジェンスサーバの両方の役割を担う必要がある。本学のデータレイクサーバは、以下の機能を果たすように設計した。

- ・ 様々なデータの蓄積：データレイクの主要な機能
- ・ クレンジングされたデータの提供：データウェアハウスの機能
- ・ 可視化されたデータの提供：ビジネスインテリジェンスの機能

これらの機能を考慮すると、データレイクサーバへのアクセスを制御する必要があるため、以下のような領域が提供できることが重要であった。

- ・ データ提供者がアクセスできる領域
- ・ クレンジングされたデータを提供できる領域
- ・ 可視化されたデータを提供できる領域

IR担当者は、意思決定に資するデータ報告の際にデータレイクに蓄えられたデータを使用するため、最も重要な条件は、IR担当者がすべての領域にアクセスできることである。

これらの条件を満たすNextcloudをデータレイクサーバとして導入した。Nextcloudはフリーソフトでオープンソースのソフトでもあり、学内のプライベートサーバにインストールできる。Nextcloudは転送時にファイルを暗号化することもでき、LDAPでユーザを管理し、ファイル権限も定義できるのも特徴である。

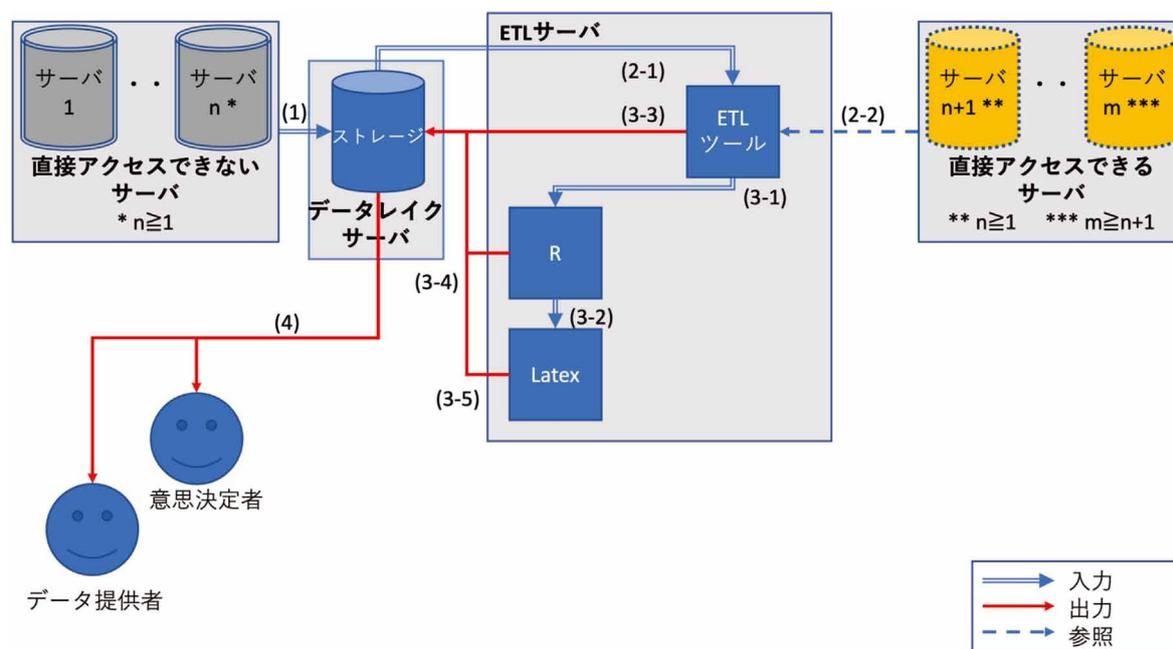


図2 現状のIRシステム

現状のIRシステムは図2に示すとおりである。(1) データ提供者は、直接アクセスできないサーバからデータレイクサーバのストレージにデータを転送する。次に (2-1) データレイクサーバ内のストレージにあるデータと (2-2) 直接アクセスできるサーバからの

データを、ETLツールを使って結合する。(3-1) ETLツールで結合されたデータは、統計解析ソフトRで分析・可視化され、(3-2) LaTeXでPDF化される。(3-3) ETLツールで結合されたデータ、(3-4) Rで可視化されたデータ、(3-5) LaTeXでPDF化されたデータは再度データレイクサーバ内のストレージに格納される。(4) データ提供者や意思決定者はデータレイクサーバから報告されたデータを取り出すことができる。

また、現状のIRシステムにおいて一部の可視化したデータの提供を自動化した。本学ではこのIRシステムを利用して年に数回、図3と図4に示すような成績分布を各学部提供している。

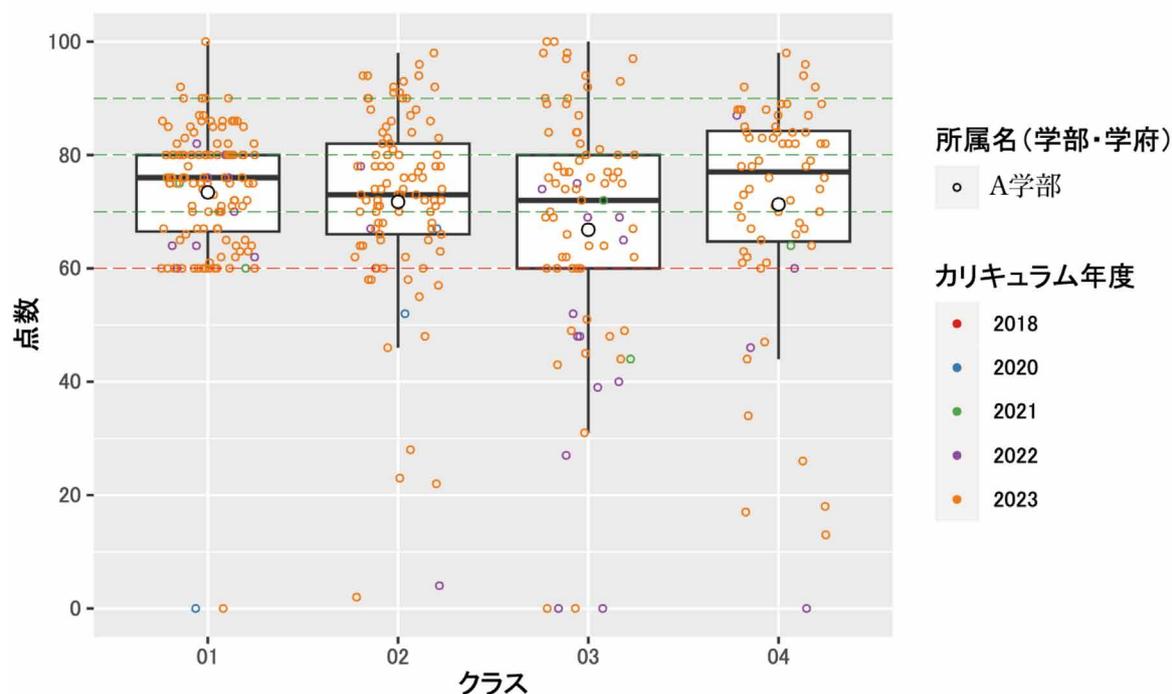


図3 成績分布 (点数)

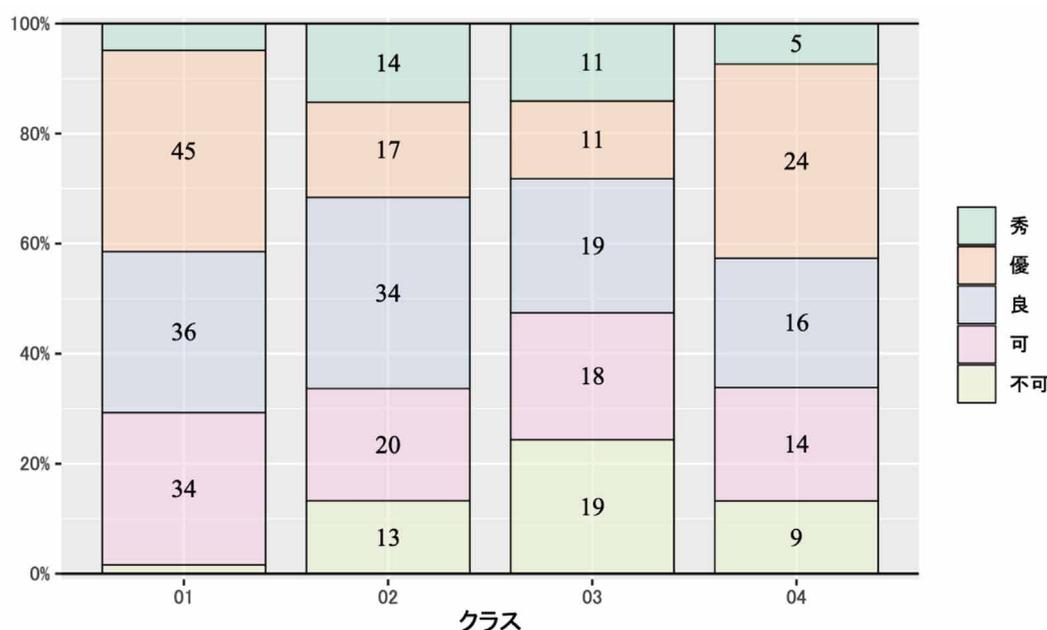


図4 成績分布 (評定)

### 3.3 今後のIRシステム

図5は、自動可視化の仕組みを実現するための将来のサーバ構成を示したものである。

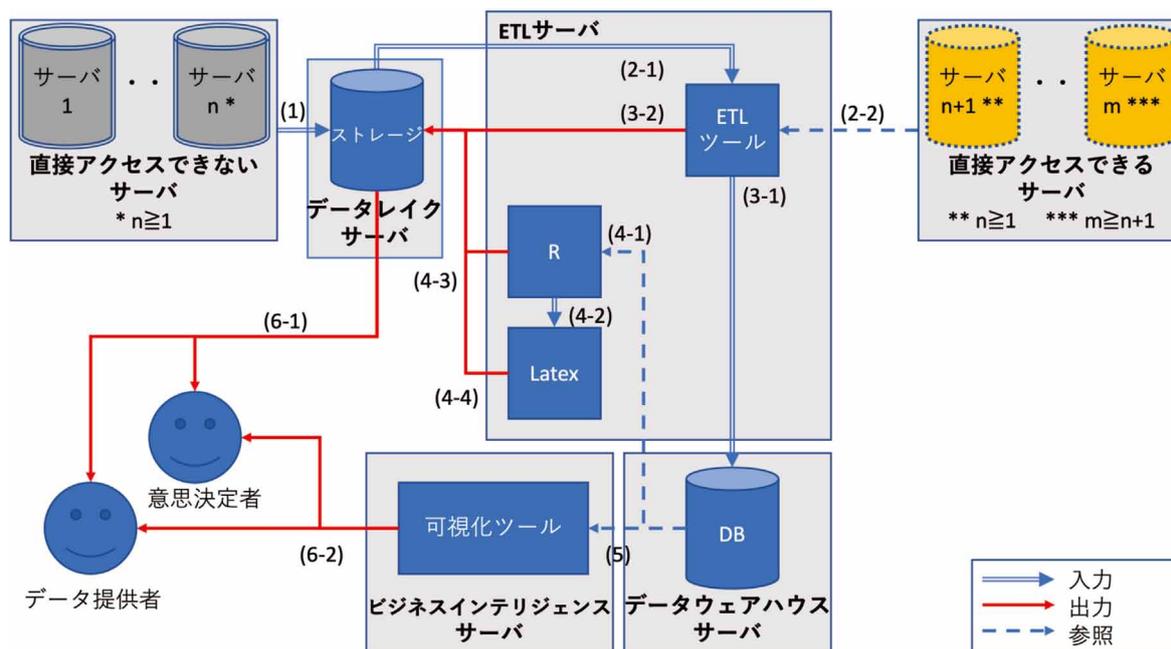


図5 今後のIRシステム

(1) データ提供者は、直接アクセスできないサーバからデータレイクサーバのストレージにデータを転送する。次に (2-1) データレイクサーバ内のストレージのデータだけでなく、(2-2) 直接アクセスできるサーバのデータもETLツールで結合する。このとき、データレイクサーバ内のストレージはETLサーバにマウントされるため、あたかもETLサーバ内のストレージにデータがあるかのように利用できる。また、ETLツールで結合されたデータは、(3-1) データウェアハウスサーバ内のデータベースに格納されるだけでなく、(3-2) 再びデータレイクサーバ内のストレージに格納される。続いて (4-1) データウェアハウスサーバ内のデータベースのデータは統計解析ソフトRを用いて分析・可視化され、(4-2) LaTeXを用いてPDF化される。(4-3) Rで可視化されたデータおよび (4-4) LaTeXでPDF化されたデータは再びデータレイクサーバ内のストレージに格納される。ETLサーバにRとLaTeXがあることで、後述するビジネスインテリジェンス機能では間に合わないような急な要望にも柔軟に対応できる可視化が可能となる。一方で、(5) データウェアハウスサーバ内のデータベースのデータはビジネスインテリジェンスサーバ上でインタラクティブに可視化できる。データ提供者や意思決定者は (6-1) データレイクサーバから報告されたデータを取得できるだけでなく、(6-2) ビジネスインテリジェンスサーバで動作する可視化システムを使用して必要に応じて可視化されたデータを取得することができる。

## 4 まとめ

本稿ではIRの概要と九州工業大学における教学IRの状況について説明した。本学における教学IRは発展途上にあり、試行錯誤で教学IRシステムの構築を進めている。今後は

データレイクを活用しつつ、他大学によるIRの事例を参考にし、本学の状況を考慮して、データウェアハウスおよびBIツールの導入の検討を進める。

## 謝辞

本稿で説明したIRシステムの構築にあたり、初期段階でETLサーバが必要であることを2022年1月の段階でお認めいただきました梶原誠司教授（前教育担当理事・副学長）に感謝の意を表明いたします。また、本項の執筆に至るまでサポートいただきました坂本寛教授（現教育担当副学長）、データレイクの構築にご尽力いただいた中村豊教授、大西淑雅准教授、林豊洋准教授（情報統括本部 情報基盤センター）、学習教育センターの教学IR支援グループにおけるサポートをしていただいた林朗弘准教授（大学院情報工学研究院）、福丸浩史上席高度専門職員（学習教育センター）に併せて感謝の意を表明いたします。また山出崇課長、小林朱美係長、高田将英主任、藤井佳奈子主任（教育支援課）にはIRシステムの検証に協力いただき感謝申し上げます。

## 参考文献

- (1) 坂本 寛, 福丸 浩史, 宮浦 崇, 西野 和典, 林 朗弘: “eポートフォリオによる学修成果の可視化の事例報告－教育の内部質保証の確立と多様な教育機関によるコンソーシアム形成－”, 大学 e ラーニング協議会 / 日本リメディアル教育学会 合同フォーラム 2019 (2020)
- (2) R.D.Howard 編, 大学評価・学位授与機構IR研究会 訳: “IR実践ハンドブック” (2012)
- (3) 相生 芳晴, 井芹 俊太郎, 今井 匠太郎, 大石 哲也, 岡田 佐織, 近藤 伸彦, 杉原 亨, 田尻 慎太郎, 椿本 弥生, 西山 慶太, 松田 岳士, 森 雅生: “データの収集”, 大学IR標準ガイドブック, 日本IR協会, pp.122-126 (2022)
- (4) Tetsuya Oishi: “What is the Essential Curriculum for IR in Japan?”, 12th International Conference on Data Science and Institutional Research (DSIR 2022), Kanazawa (2022)