

全学統合型データベースの必要性を考える

藤原 宏司¹・大野 賢一²

概要：日米のIR（インスティテューショナル・リサーチ）室には、大学内に存在するデータの管理方法や全学規模の統合型データベースの有無など、大きく異なる点が幾つか存在する。本稿では、効率的なIR業務の遂行ならびに、学内データの整合性の保証といった観点から、全学統合型データベース導入の必要性を考えてみたい。また、そのようなデータベースを導入できず、データが分散管理されている場合におけるIR機能の発展方法についても提案する。

キーワード：IR（インスティテューショナル・リサーチ）、データベース、日米比較、米国IR

1. はじめに

米国のIR業務は、統合型データベースの存在抜きでは語ることはできない。なぜなら、米国でIR業務に従事している者にとって、学内データは、各部署から集めてくるものではなく、予めデータベース上に集約されているものだからである。

ミネソタ州には、ミネソタ大学システム（University of Minnesota System）とミネソタ州立大学機構（Minnesota State Colleges and Universities System、以下「MnSCU」という。）の2つの高等教育機関の集合体があり、ベミジ州立大学（以下「BSU」という。）及びノースウェスト技術短期大学（以下「NTC」という。）は後者のMnSCUに属している。MnSCUでは、Integrated Statewide（Student）Record System（以下「ISRS」という。）と呼ばれる大規模な統合型データベースを運用しており、MnSCUを構成している7つの大学と24の短期大学では、ほとんど全ての学内データ（学生、教務、財務、人事、総務、その他）をISRSで管理している。

本稿では、ISRSのような統合型データベースを、「統一された情報システム（サーバー）上に複数のデータベース（教務系、財務系、人事系等）が構築されており、利用者が1つのアカウント（ユーザーIDとパスワード）を使用し、各データベース内に存在する全データに対して適切な権限によりアクセスできるもの」と定義する。この定義に従うと、米国では大学の規模に関係なく上記のように統合されたデータベースが整備されているのに対して、日本では、大規模大学を除けば、多くの大学において学内分散型データベースを運用しているのが現状ではなかろうか。また、そのような学内分散型データベースも整備されていない場合は、Excel等のファイルによりデータが分散管理されているのではないかと推察する。

ISRSは、MnSCUのIT部門によって独自に開発され、現在も同部門によって管理、運用及び保守がされている。また、データベース上にあるデータについては適切なアクセス権

¹ ミネソタ州立大学機構 ベミジ州立大学・ノースウェスト技術短期大学 IR/IE 室 リサーチアナリスト
責任著者 電話：+1-218-755-4606 メール：kfujiwara@bemidjistate.edu

² 鳥取大学 大学評価室 准教授 電話：0857-31-5706 メール：ohno@tottori-u.ac.jp

限があれば、誰でも利用することが可能で、日本の大学のように特定の部署の人間しかアクセスできない、というようなデータは存在しないと言ってよい。なお、扱うデータの範囲が広いことから、IR 担当者には学内に存在するほぼ全てのデータにアクセスできる権限が与えられている。

本稿では、全学規模の統合型データベース導入の必要性について、効率的な IR 業務の遂行ならびに、学内データの整合性の保証といった観点から、第一筆者が経験したことや、実際に受けたデータリクエストを例に挙げつつ、考えてみたい。

IR やその業務についての定義に関しては、日米問わず様々な提案がなされているが、今回は、IR 業務を「必要とする方に、必要な情報を、必要な時に提供する業務およびそのためのデータの情報への変換業務」（嶋田ほか，2014）とし、IR 室（オフィス）は「IR 業務を（全学規模で）より効果的、効率的に行う部署」と定義する。また、データとは「ある状態や結果等を示す数値や文字列」、情報とは「何らかの目的のためにデータを加工して意味を付加したもの」と定義する。さらに、データの管理方法としては、表 1 に示す「データベース」及び「データファイル」の 2 種類を想定している。つまり、データベースでは複数の利用者が同時に使用することを想定しているのに対し、データファイルでは利用者が個別に使用することを想定している点が大きく異なっている。

表 1 データ管理方法

区分	説明
データベース	生データを対象とし、各データの定義及び項目は管理者により事前に決定されており、データの正規化（データの重複を無くし、データの管理を容易にすること）を行うことで、複数のデータを組み合わせて使うことが可能なもの。 主なデータベースとして、MS Access、MS SQL Server、Oracle Database、MySQL、PostgreSQL 等がある。
データファイル	生データ及び加工されたデータを対象とし、各データの定義及び項目は利用者により随時変更可能であり、データの重複が可能なもの。 主なデータファイルとして、テキストファイル、CSV ファイル、MS Excel 等がある。

2. 全学統合型データベースの必要性

BSU では、新入生に対し、少なくとも初年度は学生寮で生活することを強く推奨している。これは学生寮に住んでいる学生の方が Retention 率³や卒業率が高い（Astin, 1984 など）という有名な研究結果から来ている。数年前に当時の学生部（Office of Student Development and Enrollment）担当副学長から、BSU の学生でもそのような傾向が実際に見られるかどうかを調べて欲しい、という依頼があった。統計解析の手法であるロジスティック回帰分析

³ 学生在籍率のことで、新入生のうち翌年秋学期に在籍している学生の割合を意味する。BSU の Retention 率は 68%前後。

と、データマイニングで良く使われる決定木を用いた分析を行った結果、Astin の研究等と同様の傾向が見られたが、詳しい分析内容については紙面の関係上、ここでは割愛する。

分析に用いたデータは、初年度の入寮状況に加え、First-Year Experience (FYE) と呼ばれる新入生向けの特別クラス (大学生活や学業の成功のために必要な知識を学び、それを実際に体験してもらい導入プログラム) の受講状況、学生のデモグラフィックデータ (性別、年齢、人種、社会・経済的地位⁴、奨学金の受給状況)、学期ごとの在籍状況と成績データ (在籍している学部、GPA や単位取得率)、高校の時の成績データ (GPA、クラスランク、ACT⁵の点数) 等であった。これらデータは、米国では一般的な Retention 率や卒業率に関する分析で良く使われているものであり、何ら特殊なものではない、という点に留意していただきたい。

本章では、統合型データベースの必要性について、1) IR 業務の効率化、2) 学内データの整合性の観点から検討し、統合型データベースの運用における IT 部門との連携についても考察する。

2. 1 IR 業務の効率化による観点

上記例の場合、分析に用いたデータは、既に ISRS 上に集約されている。個々のデータは、専用の Web 入力画面を通して自動的に、もしくは担当部署によって手動で ISRS に入力され、必要に応じて修正される (図 1)。

例えば、学生の性別や生年月日等の情報は、学生が入学願書をオンライン (Web 入力) で提出した場合、それら項目がデータベースに直接記録される。また、学生が入学願書を書面で提出した場合、アドミッション・オフィスの職員により、願書の各記載項目が ISRS の専用画面から入力される。このようにして収集されたほとんど全てのデータが、ISRS によって統合管理されていることと、IR 担当者に与えられたアクセス権限によって、第一筆者は分析に必要なデータセットを誰の許可も必要とせずに素早く構築することができた。

一方、日本の IR 室の場合はどうだろうか。日本の大学では、学生の人種や両親のうち少なくともどちらかが学士号を持っているか、といったデータを持っているとは考え難いが、仮に「似たような」リクエストがあった場合、どのようにして分析に必要なデータセットを作るのかを考えてみていただきたい。

大学評価コンソーシアムがまとめた「データ収集作業のガイドライン」によると、IR 室に来るデータリクエストの大半は新たにデータを集めなくてはならないもので、その都度、担当部署からデータを貰い、IR 室でそれらデータを組み合わせ、最終的な分析用のデータセットを作っている、ということであった。しかも、場合によっては分析に必要なデータが貰えないこともあるという (大学評価コンソーシアム, 2013)。つまり、各担当部署でデータが分散管理されている状況が容易にイメージできる (図 2)。そして、IR 室から出されたデータリクエストは、データの所有権を持つ担当部署で検討され、データを提供するかどうか判断されているのだろう。

⁴ PELL 奨学金の受給資格や両親のうち、少なくともどちらかが学士号を持っているか (First Generation かどうか) などを考慮しステータスを決定した。

⁵ American College Testing の略、大学進学適性試験の一つ。

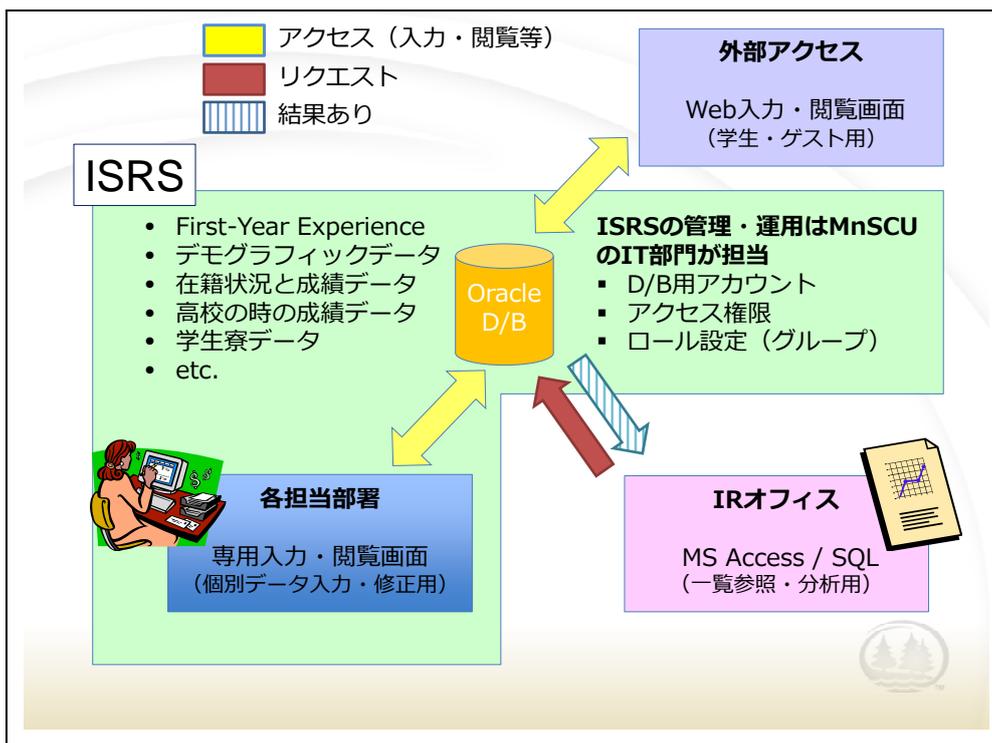


図 1 ISRS と IR 室が出すデータリクエスト (データが統合管理されている場合)

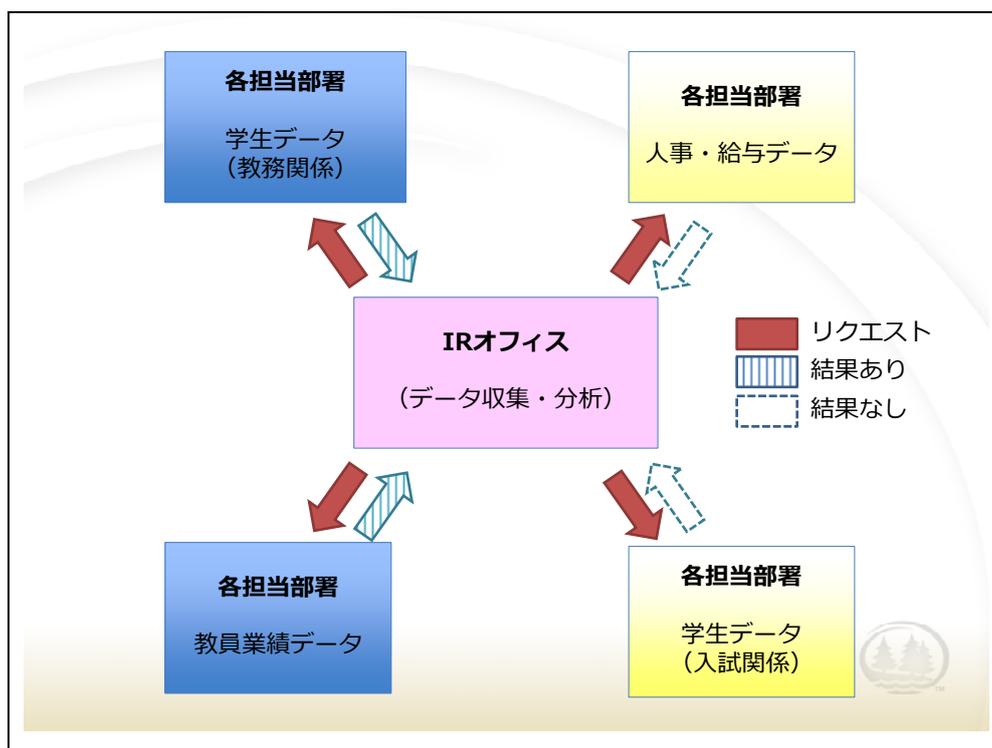


図 2 IR 室が出すデータリクエスト (データが分散管理されている場合)

図 1 と図 2 で大きく異なる点は、IR 室が出すデータリクエスト (図では IR オフィスから出ている矢印) の数である。もし統合型データベースが整備され、IR 室が必要なデータへのアクセス権限を持っている環境であれば、分析に必要なデータセットは、データベ

スへの 1 回のリクエストで入手できる。その反面、データが分散管理されている場合、必要なデータリクエストの数はデータを持っている担当部署の数に等しい。しかも、リクエスト先は各担当部署であり、そこで働く担当者である。どこの部署にも業務の優先順位というものがあり、経営陣からの直接的な指示でもない限り、IR 室からのアドホックなリクエストにすぐに対応してくれるという保証はないだろう。このような状況は、業務の遂行上、効率的なデータの集め方ではないと考えられる。

BSU と NTC で共有されている IR 室では、通常業務に加えて、年間に約 300 件のアドホックなリクエストが来る。大変規模の小さい IR 室であるが、ISRS と広範囲に渡るデータへのアクセス権限があるお蔭で、何とかそれらのリクエストをこなしつつ、IR 室独自のリサーチができる時間を確保し、米国 IR 協会（以下「AIR」という。）の年次大会等で 2012 年から口頭発表⁶を行っている。

日本の大学の IR 室も、学内でのプレゼンスが高まってくるにつれて、突発的なデータリクエストが増えると考えられる。その際、上記のような非効率的なデータ収集では、リクエストに対応できなくなる恐れがある。全学規模の統合型データベースを導入し、IR 室がダイレクトに必要なデータを収集できるようになれば、IR 業務の利便性を向上させるだけでなく、他の担当部署への負担も減るだろう。つまり、学内全体の業務の効率化の実現に貢献できる。そして、IR 室がその名の通り、「リサーチ」も行える部署になっていくのではなかろうか。

2. 2 学内データの整合性による観点

個人情報保護の観点からは、全学統合型データベースを敢えて導入せずに、データを分散管理することは、決して悪いことではない。その場合、担当部署ごとが限定的な IR の機能を担い、学内外に管轄する分野のレポートングをしていると考えられるが、学内データの一貫性や整合性を確保することが難しくなる。

例えば、BSU や NTC では 8 桁の数値からなる学籍番号を文字列として管理している。稀に、他の部署から分析のため ISRS に入力されていないデータのファイル（主にエクセルファイル）を受け取ることがあるが、学籍番号が数値型であったり、同じ列に文字列型と数値型が混在していたり、というケースも珍しくはなかった。他の部署の担当者が見やすさを考えて作ったエクセルファイルのほとんどは、ISRS 上のデータと統合する前に手作業での修正が必要になる。もちろん、この程度のデータ型やフォーマット形式の修正は、時間と手間はかかるものの、それほど問題にはならない。より深刻なのは、同じデータでも担当部署ごとにその定義が異なる場合である。

⁶ AIR の年次大会では、口頭発表（45 分）、パネルセッション（60 分）、テーブルセッション（45 分）、ポスターセッションの 4 通りの発表形式がある。プロポーザルは AIR の会員によって査読され、口頭発表での採択率は、4 割を切る。

IPEDS⁷では、一般的な新入生（First-time student）を、その年の秋学期からその高等教育機関に入学した者、または、夏学期から入学し、翌秋学期にも在籍している者、と定義している。BSU と NTC の IR 室でもこの定義に従って新入生数の報告をしていたが、ある時、執行部から IR 室が発表している数字と他の部署が出している数字がかなり違う、という指摘があった。詳しく調べてみると、その部署では秋学期から入学した学生「のみ」を新入生としてカウントしていたことが分かった。

統合型データベースを使っている、学内で統一された厳密なデータの定義が共有されていなければ、このような問題は起こる。しかし、上記の例では、修正は簡単であった。BSU と NTC では、IR 室が発表する数字が大学の公式な数字である、と定められていたので、その部署に新入生を定義するための SQL コード（データベースからデータを抽出するための条件文や定義を明記したプログラム言語）を提供し、次のレポートからはその定義を使ってもらうよう依頼しただけで済んだのである。

一方、担当部署ごとにデータが分散管理されている大学の場合、データ・オーソリティとしての IR の機能も各所に分散されていると考えられる。したがって、上記のような問題が起きた場合、どの部署の定義が大学としての「公式」な定義なのかを最初に決めなければならない。

IR 室の主たる役割は、執行部への効果的な意思決定支援である。部署ごとに異なるデータ定義に基づいて作成されたレポートは、徒に執行部を混乱させるだけでなく、誤った意思決定がなされるリスクも増大させる。その結果、レポートを作成した IR 室や各担当部署への執行部や学内からの信頼は著しく低下することになる。統合型データベースの導入またはその検討作業は、明治大学の例（山本, 2013）のように、大学として「公式なデータ定義書」の作成を促し、より信頼度の高いデータが学内で共有されていく良いきっかけになるのではなかろうか。

2. 3 米国の大学における統合型データベースの運用及び体制

全学統合型データベースを導入して IR 業務を行うためには、IT 部門との連携が必要不可欠と思われる。米国では、IR 室はあくまでも MS Access や SQL を用いてデータベースからデータを持ってくることが仕事であり、データベースの構築や管理は IT 部門が担う、という形が一般的だ。ISRS の場合、テーブルやフィールド等の追加といった、比較的簡単と思われる作業でさえも IT 部門が行っている。

日本の IR 担当者からデータベース構築に関する質問をされたことがあるが、IR に必要な知識としてのデータベースの使い方と、全学規模のデータベース構築の技術は、全く以て別物である。業務の発展向上のため、IR 室の中に IT の専門家を配置している大学もある。しかしながら、これも根幹となるデータベースシステムが IT 部門によって運用されて

⁷ Integrates Postsecondary Education Data Systems (<http://nces.ed.gov/ipeds/>) の略。連邦政府のデータコレクションプログラムのことで、Higher Education Act of 1965 によって規定された連邦政府の奨学金プログラムに参加している大学は、毎年 IPEDS にデータを提出しなければならない。奨学金を受けている学生が一人もない大学がある、とは考えられないので、事実上、全米にあるア krediteーションを受けている全ての大学が対象である。

いるからこそ実現できているのであって、IR 室はデータベースを「作る」のではなく、あくまでも「使いこなす」側である、と考えていただきたい。

米国の一部の短期大学では、ルミナ財団 (Lumina Foundation) が中心となり、低収入層またはマイノリティの短期大学生を主に対象とした、学生の学業成就や学業を通しての資格取得を支援するためのプロジェクト (Achieving the Dream⁸) が行われている。そのプロジェクトがまとめたレポートの一つに、プロジェクトに参加している短期大学が、どのようにして IT 部門と協力しながら IR 機能を拡充していったのかを紹介している事例がある (Glover, 2009)。本事例は、中小規模 (フルタイム換算で 3 人前後) の IR 室を既に運用、もしくは志向されている日本の大学や短期大学には参考になるかもしれない。

3. データ分散型管理における IR 機能発展のための提案

本稿では、全学統合型データベースの必要性について、業務の効率化及びデータの整合性の観点から論じてきた。しかしながら、一口で導入と言っても、その実現には多大な労力を必要とし、またそれを維持するために、莫大なコストがかかることもまた事実である (Kroc and Hanson, 2003)。

MnSCU では、ISRS の開発から運用開始まで 5 年以上を要し、また運用開始直後はトラブルもかなりあったと聞く。現在は、ISRS を管理、運用するために、人件費込で毎年約 25 億円を費やしている。2014 年度の学生総数が約 42 万人 (重複⁹有り、ノンクレジットクラス¹⁰の受講者を含む) かつ、7 つの大学と 24 の短期大学が使用している大規模な統合型データベースシステムである、ということ を考慮すると、運用コストである約 25 億円という数字はそれほど高いとは感じないかもしれない。しかし、ISRS を維持し続けるよりも、商用のデータベース (データウェアハウス) へ移行した方がコスト面で有利だという試算結果もあり、MnSCU 内で議論が行われている最中である。

では、費用や組織的な事情から全学統合型データベースを導入せずに、各部署がそれぞれのデータを管理する「分散型」を維持したままで、IR 機能を発展させたい場合はどうすればよいのだろうか。最低限必要なのは、学内での公式なデータ定義書を策定することではなかろうか。前節 (2. 2) で述べたように、部署ごとに異なるデータの定義が使われている状況が「もしも」学内に存在するのであれば、まずはその改善から始めることを奨めたい。

IR 業務の根幹は、データを有益な情報へ「正しく」変換することであり、そのためにも統一された厳密なデータ定義が求められる。その後、IR 室に正式なデータ・オーソリティとしての機能を与え、IR 室が、学内の他部署やそこから提出されるレポートにより深く関われる環境を作り、その先の発展に期待するのはどうだろうか。

⁸ 学生支援のために、継続的なデータを用いた意思決定モデルを提唱している。

(<http://achievingthedream.org/>)

⁹ MnSCU 内の大学・短期大学に入学を許可された学生は、他の MnSCU を構成する大学・短期大学も提供するクラスを自由に受講することができる。例えば、BSU の学生がある学期に BSU と NTC が提供しているクラスをそれぞれ受講したとする。この場合、その学生は BSU の学生としてカウントされると同時に NTC の学生としてもカウントされる。

¹⁰ 聴講用の授業のこと。基本的に大学を卒業するための単位としては認定されない。

担当部署ごとにデータが分散管理されていても、データ定義さえ正確に共有されていれば、レポートの一貫性や整合性は保たれる。しかしながら、効率性の問題は依然として残る。筆者らは、データリクエストに対し、迅速に正確な答えを返すことが、IR 室が学内から信頼されるための第一歩だと考えている。だが、各担当部署から手動でデータを集めなければならない状況であれば、その実現は難しいだろう。

また、経営陣は、IR 室に対して意思決定に必要な情報をタイムリーに提供することを求めている。その要求に答えるためにも、IR 室に他部署が管理するデータへのアクセス権限を与えることが重要ではなかろうか。これにより、IR 室と担当部署間の交渉にかかる時間が省略でき、データリクエストに対し、より素早い対応が可能となる。

上記の2点を達成するためには、大学執行部からの強力な支援が必要である。なぜなら、大学のIR 室を特徴付けるのはその大学の執行部であり、IR 室で働いている教員や職員ではないからだ。大学に必要とされていないIR 室が機能することはなく、必要とされて初めてIR 室が機能する。

4. まとめ

米国の大学では、学内にあるデータは大学全体のもの、という考え方が一般的である (Luna and Pearson, 2003)。それは、全ての州にある情報公開に関する法律の存在や、連邦法である情報自由法 (Freedom of Information Act) の趣旨から、公立・私立を問わず、学生や教職員のプライバシーを守りつつ、情報を公開しなければならない、という考えから来ていると思われる。大学総体の価値を高めるIR (山本, 2013) の実現のためにも、もし学内に、データのアクセス権限に関する部局や部署間の壁があるのならば、それを取り除く努力が必要ではなかろうか。

一部の大規模大学では、大学全体をカバーする従来のIR 室に加えて、各学部にもIR 担当者を置いているところもある。MnSCUでも同じような仕組みをとっており、機構本部にあるIR 部門がMnSCUを構成する各大学・短期大学の基本的なデータ (学生数、男女比、Retention 率、卒業率や就職率等) を取りまとめ、MnSCUの執行部や連邦・州政府への基本的なレポート業務を担い、各大学や短期大学に置かれているIR 室が、その機関の事情に則したデータ集計や分析を行っている。

組織の規模が大きくなると、セントラルのオフィスではきめ細やかな対応が難しくなる。このようにIRの機能を中央に置きながら各所にも分散するというやり方は、統合型データベースの存在が前提となっているが、この方式であれば日本でも有効かもしれない。

AIRが2012年に刊行したIRに関するハンドブック (Howard et al., 2012) には、米国で行われているIR 室の活動が多岐にわたることが示されており、第一筆者は、IR 室のことをデパートの総合案内所になぞらえて「データの総合案内所」と呼んでいる。大学評価コンソーシアム主催の勉強会「米国におけるIR 実践を通して考える日本型IR」 (大学評価コンソーシアム, 2014) の報告書57頁でも述べているが、データの総合案内所であるIR 室は、特定の分野のデータを扱うのではなく、データが存在する全ての分野をカバーする、という捉え方もほとんど全てのデータにアクセスできる環境で働いていることからきてい

る。将来的に、日本の大学でもこのような環境が整えば、さらに効果的かつ効率的な IR 室の運用が可能になるだろう。

謝辞

本稿作成にあたり、茨城大学大学戦略・IR 室の寫田助教にご助言をいただきました。また、査読者の方々には、貴重なご示唆をいただきました。ありがとうございました。

引用文献

- 寫田敏行, 藤原宏司, 浅野茂, 大野賢一, 関隆宏, 小湊卓夫, 土橋慶章, 本田寛輔 (2014) 「米国の中規模州立大学の IR オフィスおよび国立大学の評価・IR 部署における業務の現状と今後の展開に関する一考察」, 『日本高等教育学会第 17 回大会発表要旨集録』, pp.46-47.
- 大学評価コンソーシアム (2013) 「データ収集作業のガイドライン – 効率的・効果的な評価作業のためのデータ収集の課題と対応 – (平成 25 年 2 月 12 日版)」, 6p.
http://iir.ibaraki.ac.jp/jcache/documents/guideline/h25-0212_Hyouka_guideline_data_management.pdf (最終閲覧日 : 2015 年 1 月 26 日)
- 大学評価コンソーシアム (2014) 『勉強会 : 米国における IR 実践を通して考える日本型 IR 報告書』 (平成 25 年 11 月 12 日実施), 158p., 2014 年 3 月.
<http://iir.ibaraki.ac.jp/jcache/index.php?page=ir20131112> (最終閲覧日 : 2015 年 1 月 7 日)
- 山本幸一 (2013) 「PDCA サイクルの基盤となるマネジメント志向 IR の開発と定着 ~ 明治大学における内部質保証システムの整備を事例に ~」, 大学マネジメント, 9(3), pp. 22-28.
- Astin, A. (1984). Student involvement: A developmental theory for higher education. *Journal of College Student Personnel*, 25, 297–308.
- Glover, R. (2009). *Strengthening Institutional Research and Information Technology Capacity through Achieving the Dream: Principles and Practices of Student Success*. Retrieved from http://www.achievingthedream.org/sites/default/files/resources/PrinciplesAndPracticesofStudentSuccess_StrengtheningInstitutionalResearchAndInformationTechnologyCapacityThroughATD.pdf. (最終閲覧日 : 2015 年 1 月 7 日)
- Howard, R. D., McLaughlin, G. W., & Knight, W. E. (2012). *The Handbook of Institutional Research* (p. 768). San Francisco: Jossey-Bass.
- Kroc, R., & Hanson, G. (2003). Enrollment Management. In W. Knight (Ed.), *The Primer for Institutional Research* (p. 161). Tallahassee: Association for Institutional Research.
- Luna, A. L., & Pearson, T. P. (2003). Records Management. In W. Knight (Ed.), *The Primer for Institutional Research* (p. 161). Tallahassee: Association for Institutional Research.

[受付 : 平成 27 年 1 月 10 日 受理 : 平成 27 年 2 月 5 日]