

データカタログによるデータの所在把握と定義の統一 ～データ収集に対する評価担当者と IR 担当者の違いとは～

鳥取大学 大学評価室
大野 賢一

平成29年8月24日
大学評価・IR担当者集会2017 全体会
立命館大学 大阪いばらきキャンパス

データカタログによるデータの 所在把握と定義の統一

～データ収集に対する評価担当者とIR担当者の違いとは～

鳥取大学 大学評価室
大野 賢一

話題提供 2 を担当させていただきます鳥取大学の 大野です。よろしくお願いたします。

今回、私が説明するのは『データカタログによるデータの所在把握と定義の統一』です。皆さまのお手元にお配りしている資料（A3の横置き）は、鳥取大学で今作りつつある（運営はまだ始めてないんですけど）「データカタログ」の例になります。要は、そういうものを作ったらどうかという提案が、今回の

プレゼンの趣旨になります。そもそも、このデータカタログ、先ほど浅野さんの説明にあったファクトブックもそうなんですけど、これといった定義はないです。定義がない中で、データを学内で収集したときに困った点、今後こういうふうに取り組んだらいいのかなという点を、例示しながら説明していきたいと思います。

本日の内容

データ収集に係る業務の効率化を実現する方法として「データカタログ」について説明するとともに、収集するデータに対する評価担当者とIR担当者の考え方の違いを示し、各担当者の役割分担（連携）について提案する。

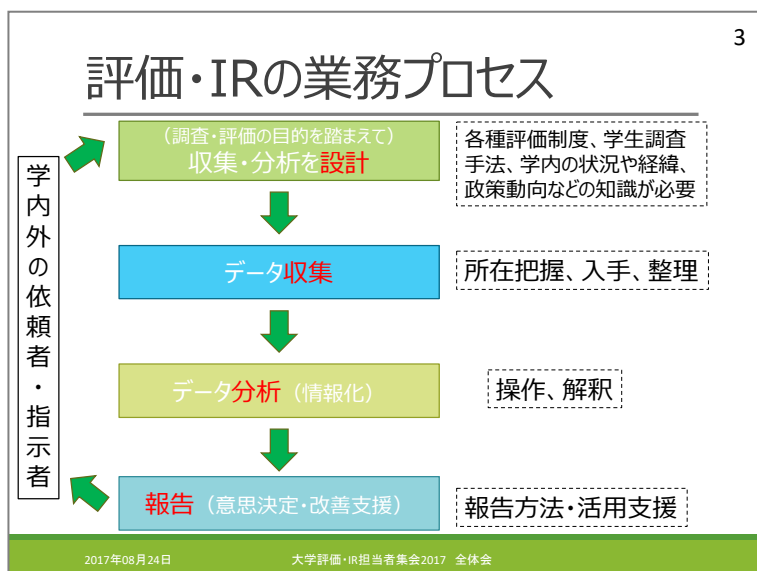
- データに係わる問題点
- データを収集するとは？
- データの利活用
- 評価担当者とIR担当者の役割分担

2017年08月24日 大学評価・IR担当者集会2017 全体会

『本日の内容』として、「データに係わる問題点」、「データを収集するとは？」、「データの利活用」、「評価の立場と IR 担当の役割分担」という 4 点を説明させていただきます。

私は、国立大学法人鳥取大学の大学評価室に所属していますが、業務実績報告書、第二期中期目標に係る評価書等、法人評価全般の報告書を書きましたし、あと、大学機関別認証評価でも自己評価書を

書いています。それとは別に、学長特別補佐（IR 担当）という立場がありますので、評価を業務としてやりつつ、IR としてこういったデータを活用するのかといった点を、経験を踏まえて少し説明させていただきます。



こちらのスライド『評価・IRの業務プロセス』は全体会の趣旨説明でも出ましたので、飛ばします。

4

データに係わる問題点

- データに関して、本学では以下のような問題がある。
 - データの定義が曖昧な状態で図・表が作成されており、毎回数値が異なる場合がある。
 - 様々な調査票の作成（データ→情報への変換）が行われているが、全学ではほとんど共有されていない。
 - 担当部署が作成したデータ（情報）がノーチェックで学外に提供・公表されている場合がある。
- 評価関係の報告書作成において、データを根拠資料（エビデンス）とする際に、結構頻繁に発生。

2017年08月24日 大学評価・IR担当者集会2017 全体会

『データに係わる問題点』についてです。評価の報告書を作成する際に、一番下のほうに書いてありますけれども、データを根拠資料として収集します。収集するときは、主に評価部署が各部署に対して、このデータについて過去何年分のデータをください、というふうにお願いします。そうすると、大体、集計結果が上がってきます。それを元に報告書を書いていくんですが、明らかにデータが間違っ

ていたり、経年変化で見たら少しデータがおかしかったりと、頻繁に間違いがあります。そのことを担当者に問い合わせると、データの定義が違っていたとか、集計をちょっとミスしていたというような回答が結構ありました。こういうことがありましたので、やはりデータの定義っていうのをハッキリしておく必要があると思います。上に書いてありますように、曖昧な状態で図とか表が作成されており、毎回数値が異なる場合があります。本学の場合でも、大学概要に載せた数値データが、実際は違っていったことも過去にありました。評価でも当然同様のことが起きます。

また、様々な調査票を作成する際に、このようなデータを集計して情報に変換するんですけど、そういったものを作ってもなかなか学内で共有されなかったり、担当部署で作ったものがそのまま学外に出ていたりする場合もありました。そのため、全学として、どこでどのような調査票の作成が行われ、こういったデータが学外に出ているかということが、なかなか把握しづらいという状況にありました。こういった状況を改善するために本学では色々取り組んでいますが、後で詳しく説明します。

5

過去に発生した問題点（評価）①

基準 3 教員及び教育支援者

3-1-② 学士課程において、教育活動を展開するために**必要な教員が確保**されているか。また、教育上主要と認める授業科目には、専任の教授又は准教授を配置しているか。

3-1-③ 大学院課程において、教育活動を展開するために**必要な教員が確保**されているか。

【根拠となる資料・データ等例】

- ・認証評価共通基礎データ様式【大学用】様式 1（提出必須）
- 専任教員数、研究指導教員及び研究指導補助教員**

大学改革支援・学位授与機構「大学機関別認証評価 自己評価実施要項（平成30年度実施分）」より抜粋

2017年08月24日 大学評価・IR担当者集会2017 全体会

まずは、実際どういったところが問題だったのかという点を、評価初心者向けに、認証評価の例で少し説明したいと思います。大学改革支援・学位授与機構の大学機関別認証評価では、こういった観点でデータを出しなさいというものがありません。根拠となる資料、データ等の例も示されており、ここでは専任教員数、研究指導教員数等を根拠資料として出しなさいということです。

6

過去に発生した問題点（評価）①

学部・学科等の名称	専任教員等										非常勤教員	専任教員一人あたり5000円相当の学生数	備考
	教授	准教授	講師	助教	計	基準数	うち前経数	助手	教員	非専任			
〇〇学部〇〇学科	人	人	人	人	0人	人	人	人	人	人	人	人	
△△課程 (大学全体の収容定員に応じた教員数)	—	—	—	—	—	—	—	—	—	—	—	—	
計	0人	0人	0人	0人	0人	0人	0人	0人	0人	0人	0人	0人	

大学院課程	研究指導教員及び研究指導補助教員										非常勤教員	備考
	研究指導教員	うち教授数	研究指導補助教員	計	基準数	うち教授数	研究指導補助教員基準数	基準数計	助手	非専任		
〇〇研究科〇〇専攻(M)	人	人	人	人	人	人	人	人	人	人	人	
〇〇専攻(D)	0	0	0	0	0	0	0	0	0	0	0	
計	0	0	0	0	0	0	0	0	0	0	0	

専門職学位課程	専任教員										非常勤教員	備考
	専任教員	うち教授数	3%未満の専任教員数	基準数	うち教授数	3%未満の専任教員数	うちみなし教員数	助手	非専任			
〇〇研究科〇〇専攻	人	人	人	人	人	人	人	人	人	人	人	
法務研究科法務専攻	0	0	0	0	0	0	0	0	0	0	0	
計	0	0	0	0	0	0	0	0	0	0	0	

大学改革支援・学位授与機構「大学機関別認証評価 自己評価実施要項（平成30年度実施分）」より抜粋

2017年08月24日 大学評価・IR担当者集会2017 全体会

実際この帳票の数値を各学部で埋めてもらったんですが、当初は、この専任教員という定義があまり理解できていなかったせいか、専属の教員数をカウントする部署があったり、研究指導教員と研究指導補助教員の違いがいまいちよく分かっていなかった部署があったりしました。事前に、担当部署とは打ち合わせを行っていたんですが、結果的にこういったミスがありました。このような単純な集計でも

ミスは発生しており、評価に関係する部署ではちよくちよく起きている現象だと思えます。

次に、過去に発生した問題点の2つ目です。今度は「基準6 学習成果」になります。ここでは、卒業の状況について、少し注目していただきたいと思えます。「標準修業年限内の卒業率」と

7

過去に発生した問題点（評価）②

基準 6 学習成果

6-1-1-① 各学年や卒業（修了）時等において学生が身に付けるべき知識・技能・態度等について、単位修得、進級、卒業（修了）の状況、資格取得の状況等から、あるいは卒業（学位）論文等の内容・水準から判断して、学習成果が上っているか。

【根拠となる資料・データ等例】

・学部、研究科等ごとの標準修業年限内の卒業（修了）率（※1）及び「標準修業年限×1.5」年内卒業（修了）率（※2）（過去5年分）〔提出必須〕

大学改革支援・学位授与機構「大学機関別認証評価 自己評価実施要項（平成30年度実施分）」より抜粋

2017年08月24日 大学評価・IR担当者集会2017 全体会

「標準修業年限×1.5」年内卒業率について、データの提出を求められました。

「標準修業年限×1.5」年内卒業率については、普通の学部（4年制）であれば、1.5倍の6年かかって卒業した率というものを求めます。機構の定義では、上の「標準修業年限内の卒業率」は卒業生数ですが、下の「標準修業年限×1.5」年内卒業率は学位を取得した者の数ということで、若干定義が変わって

ます。

8

過去に発生した問題点（評価）②

※1 標準修業年限内卒業（修了）率 = $\frac{\text{標準修業年限で卒業（修了）した者の数（注1）}}{\text{標準修業年限（例：4年制学部であれば4年）前の入学者数}}$

（注1）長期履修制度を利用している学生については、長期履修年限を標準修業年限とする。

（例）4年制学部についての平成28年度における標準修業年限内卒業率
平成25年度入学者数 200人
平成25年度入学者のうち、平成28年度卒業生 175人
標準修業年限内卒業率 = $\frac{175}{200} = 87.5\%$

※2 (注2) $\frac{\text{Aのうち、(標準修業年限×1.5)年間に学位を取得した者の数(注3)}}{\text{(標準修業年限×1.5)年前の入学者数(A)}}$

（注2）「標準修業年限×1.5」の算出において、端数がある場合は、1年として切り上げる。
（注3）博士課程においては、便宜上、単位取得満期退学後に学位を取得した者を含める。

（例）4年制学部についての平成28年度における「標準修業年限×1.5」年内卒業率
平成23年度入学者数 100人
平成23年度入学者のうち、卒業生（平成26年度80人、平成27年度7人、平成28年度3人）
「標準修業年限×1.5」年内卒業率 = $\frac{80+7+3}{100} = 90\%$

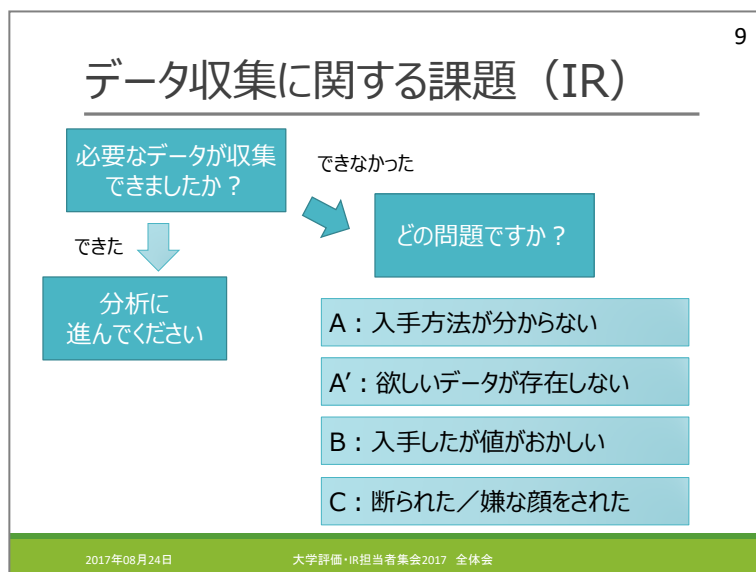
大学改革支援・学位授与機構「大学機関別認証評価 自己評価実施要項（平成30年度実施分）」より抜粋

2017年08月24日 大学評価・IR担当者集会2017 全体会

また、上については、長期履修者はちゃんと年限を考慮して標準修業年限に加えなさい、といった注意書きがあります。なお、この点について担当部署とやり取りしたときに、この定義について少し理解していただけなかったため、結果として数値がばらついたといったことがありました。

このように、データを集計するときには、何らかの条件や設定というものがされています。各部署

ではそれらを考慮して集計するんですが、評価部署には集計した結果が上がってきますので、実際、それを検算することが難しい状態になっています。なので、各部署から上がってきたデータは多分合っているだろうと信じてやっていくんですよね。けど、例えば、評価の報告書を出す直前に、最終版のチェックを各部署にお願いしたら、そのときになって初めて「これデータ違っています」ということもたまにあります。評価部署では、なるべくそういったことが発生しないように各段階で気をつけてはいるのですが、やはりこういったことが発生するのが現状だろうと思います。



ここからは、IR の観点になります。これは、昨年、一昨年の担当者集会の全体会で、今回と同じように付箋ワークをしていただいたときに出てきた課題や問題点を集約した結果になります。例えば、『データ収集に関する課題』として、必要なデータが収集できなかった場合です。そのときに、「どの問題ですか」という問いに対して、こういった4つの解を考えています。「入手方法が分からない」、「欲しいデータが存在しない」、「入手したが値がおかしい」、あとは「断られた/嫌な顔をされた」となっています。

「入手したが値がおかしい」、あとは「断られた/嫌な顔をされた」となっています。

10

A：入手方法が分からない

- 詳しい人に訊いてください。
 - 職員の場合、教務系のデータならば教務系経験者、財務系なら財務系の経験者など、その畑の人に聞く。
 - 教員がデータを入手したい場合、事務同士のネットワークで探してもらうのも手です。
 - 兼務型IRオフィスなら、その辺は楽か？
- **そもそも何を入手すれば良いのだろうか？**
 - クライアントとよく話してください。
 - 話をしながら、相手の求めている背景も読みとれるとなおよいです。プラスアルファを出そうと思うなら、ここが重要。
 - 信頼関係の醸成もここが重要かも。

2017年08月24日 大学評価・IR担当者集会2017 全体会

まず『A：入手方法が分からない』についてですが、どんな状況だったらどういうふうに対応すればいいのか、という点についてまとめてあります。

「詳しい人に訊いてください」というのがあります。教員がデータを収集するときは、どこが事務担当なのかが分かりにくいと思いますので、職員の方にお聞きするっていうのは手だと思います。ただ、それ以外に「そもそも何を入手すれば良いのだろうか」といった場合もあります。こちらが収集したいデータが分からなければ、クライアントである執行部だったり学部の先生だったり、どういう情報を欲しがっているのかを探る必要があります。その際、クライアントと必ずやり取りをしたいと思いますので、そのやりとりの中で相手の求めている背景もうまく読み取ってください。つまり、話をしながら、実際にどういったデータが欲しい、っていうことを絞り込んでいくことになります。これらは、直接的なコミュニケーションになりますけれど、これが結構重要で、一番肝心なことになります。この出だし（ステップ）を間違えると、全く違うデータを収集することになりますので、こういったところに少し時間を割いてみてはどうでしょうかということです。

こちらが収集したいデータが分からなければ、クライアントである執行部だったり学部の先生だったり、どういう情報を欲しがっているのかを探る必要があります。その際、クライアントと必ずやり取りをしたいと思いますので、そのやりとりの中で相手の求めている背景もうまく読み取ってください。つまり、話をしながら、実際にどういったデータが欲しい、っていうことを絞り込んでいくことになります。これらは、直接的なコミュニケーションになりますけれど、これが結構重要で、一番肝心なことになります。この出だし（ステップ）を間違えると、全く違うデータを収集することになりますので、こういったところに少し時間を割いてみてはどうでしょうかということです。

B：入手したが値がおかしい

11

■ 定義、基準日をはっきりさせよう。

- 利用目的がはっきりすれば、定義も自ずと決まることが多い。
- データベースの設計、各種学内調査の調査表作成に、IRも加えてもらえるよう根回しも重要？

■ 他人を信用しないのも選択肢。

- 発生源入力データベースの精度が信用できないなら、自分たちで入れる。→ 他部署から楽になった実感があればパート職員くらい増員できる？ → データリクエストへの対応業務を本務にするかどうか。

■ 生データももらっておけば、用途に応じてデータを加工すればよい。

2017年08月24日

大学評価・IR担当者集会2017 全体会

ってというのは、そもそもその利用目的をちゃんと相手に伝えているかどうか、そこに大きな問題があります。なぜなら、利用目的というものがはっきりしていたら、定義というものは自ずと決まってくることが多いからです。こういった目的の部分の部分をちゃんと設計してデータ収集ができる、生データの場合だったら、データベースの設計にもつながりますし、単純集計したデータの場合でも、定義とか基準日が揃ったものが集まることになります。

ただ、IRの立場からすると、集計データももらっても、その次の分析段階ではほとんど使えませんので、できる限り生データももらっておくことをお勧めします。そうすれば、後々こちらで集計もできますし、集計した結果を、担当部署に提供することもできます。なので、やはりIRとしては、生データももらっておきたいということになり、それがIR活動の目的にもなるかと思えます。

次に『B：入手したが値がおかしい』についてです。この点も先ほど少し説明しましたが、データの定義や基準日が変わると、当然データの値も変わってきます。集計されたものだと検算はできませんし、後で説明しますが、生データと呼ばれるリスト形式のデータについても、入っている値がどんな定義なのかという点について、はっきりしておく必要があります。

あと、入手したが値がおかしい

なぜこのような状況が発生する？

12

■ 評価担当者の場合

- 大学機関別認証評価にしろ、国立大学法人評価にしろ、法令で定められているため、自己評価書を作成するために根拠資料を収集すること（目的）は明確。
- 根拠資料が指定されている場合もあるが、データや情報を収集する必要あり。

■ IR担当者の場合

- 執行部や学部等の依頼によるため、求められるもの（目的）が明確でない場合もある。
- 依頼内容により収集するデータ形式が異なったり、複数の部署から収集する場合もある。また、収集したデータを単に集計するだけでなく、加工・統合する必要あり。

どちらの担当者（業務）でも「データ収集」は必要だが、対象とするデータの「種類」や「その後の活用」が異なる？

2017年08月24日

大学評価・IR担当者集会2017 全体会

このような問題点は他にもたくさんあると思うんですが、じゃあなぜこのような状況が発生するのかというのを、私なりに「評価担当者の場合」と「IR担当者の場合」に分けて示します。

評価業務にはいろんな評価がありますが、評価でやることは決まっています。自己評価書を作成するためには、根拠資料を収集するというのは当然ですし、エビデンスに基づいた分析というのは、当

然やらなければいけないし、やらないとなかなか評価結果も良くならないということです。

根拠資料については、先ほどの認証評価の例で示したように指定されている場合もありますが、国立大学の法人評価だと、自分の大学が立てた中期計画に対して、根拠資料を明示する必要がある

ります。そういったときには、自ら、こういう定義で収集し、こういったデータを元に集計しています、ってことも明確にする必要があります。

IR の場合だと、そういった評価における「外圧」と呼ばれるようなものがなく、大本の目的が明確になっていない場合があります。なので、そういった目的をより明確にしておく必要があるかと思います。例えば、依頼される内容にもよりますが、IR に求められるものが、単純に集計して現状を示すものだけでいいのか、それとも、あるデータと他のデータを組み合わせて、何か新しい現象を発見するようなものを求められているのか、そういった目的に応じて、収集するデータだったり、そのデータの形式だったり、そういったものが変わってくると思います。例えば、学生の入試データと就職時データを収集しようとする、入試課と就職支援課のような2つの部署に依頼する必要がありますし、それぞれのデータを組織ごとに集計するだけではなくて、学生個人のデータとして結合して分析する必要もあろうかと思います。

先ほど、ワークをしていただいた結果（付箋）を、評価系と IR 系に分類したところですが、評価系だと、大体、単一項目の集計結果（学生数とか教員数とか）、そういったものが多かったと思います。けど、IR 系だと、それ以外の、他の項目と組み合わせたクロス集計（複数の項目による集計）というようなものが主だったかと思います。評価担当者と IR 担当でデータの見方の違いはありますが、その辺りをうまく組み合わせていけば、学内のデータっていうものを、もう少しうまく活用できるようになるんじゃないかと考えています。

13

データを収集するとは？

- データは、何を対象に、どこから集めるのか？
 - 収集の対象（目的）、所在把握
- データは、どのように集めて、どのように蓄積するのか？
 - 収集方法（形式）、管理方法
- 疑問点：
 - ✓データの定義は同じもの？
 - ✓収集したデータは正しいの？（データの正確性）
 - ✓担当者が代わってもデータは同じもの？（データの信頼性）

2017年08月24日
大学評価・IR担当者集会2017 全体会

ただ、どちらもやはりデータ収集に関しては大きな課題があります。じゃあ実際にデータを収集する際に、どんな種類のデータを集めて、その後データをどういうふうに活用するのかという点について、少し説明させていただきます。

『データを収集するとは？』として、主に2つあります。「データは、何を対象に、どこから集めるのか」という観点。こちらは、収集の対象とか目的、所在の把握って

うのがメインになっております。その後、「データは、どのように集めて、どのように蓄積するのか」という観点については、収集の方法とか管理の方法というのが対象になってきます。疑問点として、先ほどから少し説明していますが、「データの定義が同じものなのかどうか」、「収集したデータは実際に合っているのかどうか」、担当者が代わっても本当に同じデータが出てくるのかどうかです。こういった点は実際にデータ収集を行う時点で課題になってきます。

14

データは、何を対象に、どこから集めるのか？

■収集の対象

- 加工されたデータ（集計値）：大学概要、学校基本調査、大学ポートレート等
- 生データ：学生の成績リスト、教員の業績リスト、個別のアンケート結果等

■所在把握

- データの保有部署、上記調査の作成先、業務システムの運用先等で作成されるデータ及びフォーマット

2017年08月24日

大学評価・IR担当者集会2017 全体会

個別に見ていきますと、「収集の対象」では、加工されたデータとして、大学概要だったり、学校基本調査や大学ポートレートだったりします。そういったところに提出したデータを、調査票から集めるといのが手っ取り早い方法かと思えます。それ以外では、生データとして、学生の成績リスト（例えば、学務支援システムなどと呼ばれる業務系システムのデータベースで管理しているデータ）、教員の業績

リスト、そういったものがあるかと思えます。どちらを集めるかは、目的によって異なると思えます。あと、それらのデータがどこにあるのかを示す「所在把握」です。データ保有部署、調査票の作成先、業務システムのデータフォーマット、そういったものも把握しておく、情報を整理してうまく使えるようになるかと思えます。例えば、自分の部署で人事異動があったとしても、他部署の担当者が代わったとしても、どこにどんなデータがあるというのが分かります。

15

(参考) 学内にある公開済の数量データ

	自大学の経年データ	他大学のデータ	短評
1) 大学概要	○：通常、5、6年分なら容易に入手可能。(印刷物としても存在する)	△：国立大学に限って言えば入手しやすい。ただし一般にPDFである。	国立大学に限って言えば、掲載内容は似通っている。公立、私立との比較はやや難しい。
2) 学校基本調査	○：通常、5、6年分なら容易に入手可能。(一般に数年間は保管しているため)	△：簡単には入手できないが、グループごとの平均値等はwebで入手可能である。	国公立は大学基本情報として公表することである。
3) 大学ポートレート	△：1～3年分程度のデータしかないため、経年比較は難しい。	△：多くのデータが比較可能であると考えられるが、詳細は不明である。	大学ポートレートセンターのwebサイト等も見あたらないため仕様が不明である。
4) 教育情報の公開	△：散逸していなければ、入手可能。(過去のデータを保管しておく必要性がないため)	△：単年度ならば入手可能だが、フォーマットが異なることも多い。	このデータは当該年度のものしか公表されていない可能性が高そうである。
5) 学外からの調査依頼	△：調査を回答した部署には、数年分は残っているはず。	△：公的機関の調査の場合、集計済のデータは公表される場合が多い。	学校基本調査と比べると回収率は低くなることが多い。
6) 自己点検評価書	△：叙述資料が大半である。用いる数量データも異なる。	△：同じ認証評価機関で受審しているも用いるデータは異なる。	認証評価等では、例外的に共通の教表があるので、公表されていれば比較は容易である。

情報誌「大学評価とIR」第1号 高田(2015)より

2017年08月24日

大学評価・IR担当者集会2017 全体会

ここに、学内にある公開済みの数量データの一覧表を、サンプルとして挙げてますので、また時間があるときでも見ていただいたらと思います。

16

データの種類と質

■データは大きく分けて 2 種類

- **加工されたデータ（集計値）**
 - 例) 大学概要、学校基本調査、大学ポートレート等
 - ✓ データの集計方法や結果の真偽について、こちらで十分な説明ができない
 - ✓ 既存のものが利用できるため、収集は早い（部局間の壁がある場合も）
- **生データ（数値、テキスト）**
 - 例) 学生の成績情報、教員の業績リスト、アンケート結果等
 - ✓ データの集計方法や結果の真偽について、こちらで説明ができる
 - ✓ 業務システムへのアクセス権限の付与、保有部署への依頼文書等が必要となる場合が多く、収集するまで時間がかかる

2017年08月24日
大学評価・IR担当者集会2017 全体会

では、『データの種類と質』についてです。先ほどから少し説明していますが、データは大きく分けて2種類あり、「加工されたデータ」と「生データ」があります。

例えば、加工されたデータとは、大学概要とか学校基本調査等の、要は調査票に入っているデータになります。しかし、これらのデータについては、相手先の部署が集計した結果になり、こちらで集計していないため、実際の集計方法は

よく分からない状況ですし、どうしてもこの数値なのかということの説明も難しい状況になります。しかし、これらのデータについてはそのまま使うことができるため、収集するまでにかかる時間は少し早くなります。生データとは、学生の成績情報とか教員の業績リストになります。これらのデータは、収集後にこちらで集計できるんですが、なかなかこのようなデータを提供していただけないということがあります。それは、例えば、データベースシステムからデータの抽出に少し手間がかかるとか、担当部署に依頼文書を出さないとこれらのデータがもらえないとか（特に入試データはお願いしてもすぐにもらえるようなデータではない）、収集するまでに、結構時間がかかる場合があります。

こういったことから、データを集めるといっても2種類ありますので、どういう目的で、どういうふうを集めて、どのように活用するのかといった点を明らかにしておくことが、やはり重要になろうかと思えます。

17

データの種類と質

■データの質を左右するものとして 4 項目

- **データの定義**
 - 例) データカタログ、データベース設計書等
- **データの正確性**
 - 例) 入力値のチェック、プルダウン等による入力のルール化、複数人によるダブルチェック体制等
- **データの信頼性**
 - 例) データ追加・変更・削除処理のマニュアル化、データのクレンジング・クレンジング方針の決定等
- **データの鮮度**
 - 例) 更新の頻度、業務の定型化、定期的な報告等

2017年08月24日
大学評価・IR担当者集会2017 全体会

次の『データの種類と質』については、ここにいろんな例を示してありますが、大体こういうものが該当するのではないかなと思っていきます。今回は、このスライドを飛ばします。

18

データは、どのように集めて、どのように蓄積するのか？

収集方法や管理方法を、以下の2種類で区分。

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。

情報誌「大学評価とIR」第1号 藤原・大野（2015）より

2017年08月24日 大学評価・IR担当者集会2017 全体会

『データは、どのように集めて、どのように蓄積するのか？』では、データを収集して管理する方法を説明します。データ収集を議論するときによく出てくるキーワードに、データベースというものがあります。データベースというのは、ここに書いてある定義のようなもので、使用するソフトであれば、個人向けには「MS Access」というものがあります。他にも、ここに挙げているような製品がありますが、

要は、データの定義が決まったものが、規則正しく正確なデータとして入っているというイメージをお持ちください。データファイル、これはよく皆さまが使われている「Excel ファイル」とか「CSV ファイル」、そういったものが該当します。これらに入っているデータに関しては、利用者がデータの定義を変更でき、比較的加工しやすいものになります。

19

データベースがない／ある場合

IRオフィス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)

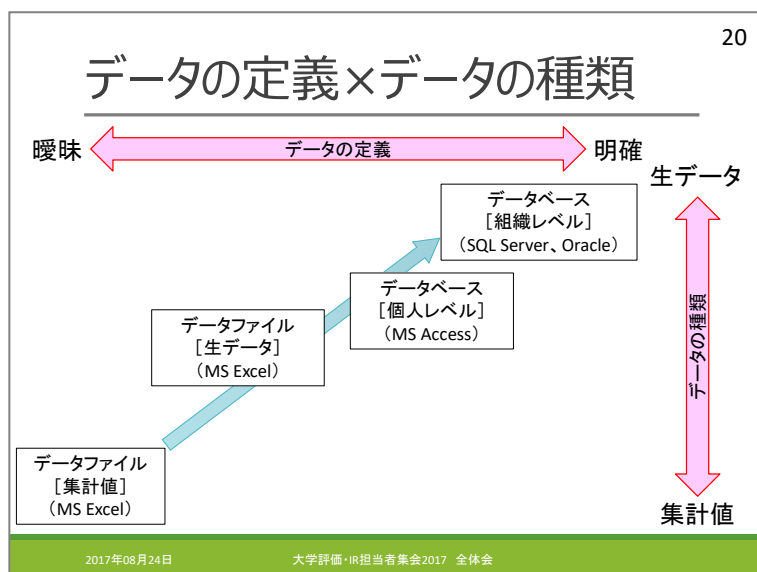
IRオフィスが出すデータリクエスト
(データが統合管理されている場合)

情報誌「大学評価とIR」第1号 藤原・大野（2015）より

2017年08月24日 大学評価・IR担当者集会2017 全体会

『データベースがない／ある場合』の IR 業務について、少しまとめたものになります。この図は『大学評価と IR』の情報誌に載せたものですけれども、データベースがあると、右側の図になります。その場合、担当者はデータベースにデータを取りに行つて、引き出して分析だけするということになります。左図のようにデータベースがないと、IR 部署は、各部署に依頼してデータを収集するっていうこ

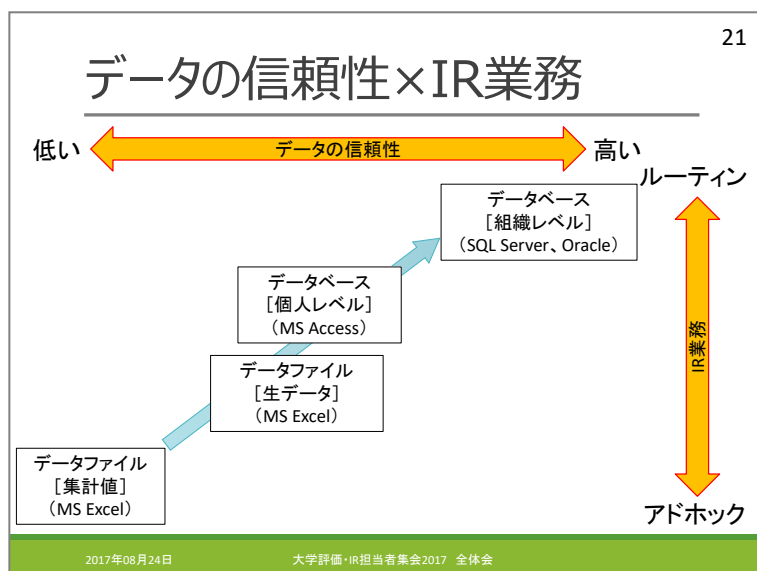
とを、依頼ごとに行う必要があります。そうすると、複数の依頼先が発生しますので、業務量が増えていくこととなります。それが、データベースがあれば1回のデータリクエストだけになりますので、業務というものは効率化でき、さらに各部署へ依頼していた業務の負担が抑えられるといったメリットがあります。



その辺りを少し概念図としてまとめたものが、『データの定義×データの種類』になります。横軸をデータの定義として、「曖昧」になるのは左側、「明確」になるのは右側としています。使用するソフトがデータベースなら右側に行きますし、Excel 等のソフトだと自分で定義を変えることもできますので左側に行きます。また、縦軸をデータの種類として、下側が集計値で、上側に行くほど生データに近づく

というものです。データベース、Excel や Access とかの使用するソフトを並べると、大体こんな感じになろうかと思えます。

各部署から Excel 等でデータを集めたときに、データ定義が異なっているってことが多々ありますので、Excel を使って収集する部分、こういったところのデータ定義をそろえるために、「データカタログ」というものを作って、この部分をなるべく無駄がないように改善していきましょうということを目指しています。



次にまとめたものが『データの信頼性×IR 業務』になります。今度は、横軸をデータの信頼性として、左側が「低い」、右側側が「高い」としています。縦軸を IR 業務とし、上側をルーティン（定型業務）、下側をアドホック（一時的な業務）としています。上と同じく、データベース、Excel や Access とかの使用するソフトを並べると、少し位置を変えていますが、大体こんな感じになろうかと思えます。

実際、データベースというものはこの辺に相当するだろうと思えます。要は、IR 業務が定型化した場合、データベースというものが導入されていると、業務を効率良く捌くことができると思えます。

また、データベースに入っているデータというのは、通常、データの信頼性が高くなっていますので、個別のデータ収集もより効率的になっていくと思えます。ただ、日本の大学だと、なかなかデータベースっていう話にはいかずに、もうちょっと手前の部分に課題があるのかなと思

ます。なので、データカタログのようなものを作って、データベースが導入できるレベルまでデータの定義を明確にし、データの信頼性を高められればいいのですが、そこに持っていけるかどうかは、IR業務が定型化されているかどうかが一番重要だと思います。

22

データベースは必要か？

- 最終的には、あった方がIR業務が効率的・効果的に行える。
- ただし、以下の条件が整ってからでも遅くない。
 - ・IR業務の大半が、定型業務（ルーティン）である。
 - ・定期的にデータ（分析結果）の提供先がある。
 - ・複数の利用者（職員や役員等も含む）が、最新データをリアルタイムで共有する必要がある。
- データベースには、多大なコストがかかることを忘れずに。
 - ・【お金】導入費、保守費、カスタマイズ費等
 - ・【人員】サーバ管理者、データベース管理者、データ運用者等
 - ・【技術】システム設計、データベース設計、情報セキュリティ等

2017年08月24日
大学評価・IR担当者集会2017 全体会

『データベースは必要か？』についてです。データベースを最初に入れたほうがいいですかという質問をよく聞きます。私は、評価業務に就く前は、学内の情報処理センターに所属してしまして、学内のネットワーク管理とかサーバー構築、データベースの設計とか、そういった業務をやっていました。結論から言うと、データベースは最初からは必要ないです。

やはり、データベースを導入す

るとなると、先ほど説明した定型業務ができてからだと思いますし、導入に伴って多大なコストがかかります。【お金】なら、やはり導入費用、入れるだけで何千万円ってかかるものもありますし、他にも年間保守費やカスタマイズ費（1回カスタマイズするごとに100万とか飛んでいく場合も）があります。【人員】や【技術】について、学内の情報センター系の部署との協力が得られれば、サーバーの運用とか任せられるんですけど、なかなか日本の国立大学等だと、業務を持っている所が導入した情報システムは、業務を持っている部署（事務）が管理するというような状況になって居ると思います。じゃあ、その部署でサーバーのセキュリティ管理とかアクセス権限、データベースの設定やテーブルの変更、そういったものができるのかということ、それは難しいと思います。実際、データベースの運用を始めると、そういった情報に関するスキルっていうのが結構必要になってきます。また、技術的な設計も必要になりますし、担当部署の職員さんができるのかということ、なかなか難しいと思います。なので、やはりこういった環境づくりは学内の体制として連携できる部署、そういったものが出来上がってから、データベースを導入したほうがいいのではないかと思います。

データの利活用

23

- このような問題点を改善するために、本学では評価業務及びIR業務において、以下に取り組む。
 - データ（情報）の所在把握と定義の統一
 - ✓鳥取大学データカタログ
 - 評価業務で収集するデータや外部に公表されているデータ（情報）の利活用
 - ✓鳥取大学ファクトブック
- 「継続的なデータ収集・蓄積体制の構築」と「意思決定や改善に資するデータ分析の提供」を目指す。

2017年08月24日 大学評価・IR担当者集会2017 全体会

ちょっと駆け足になりますけど、『データの利活用』についてです。実際データを収集した後の活用、これを考えないと、データを収集しても、単に集めているだけになりますので、その辺りをどういふふうに検討したかということです。先ほどから少し説明しているように、データは、どこにどんなものがあるかを把握し、データの定義の統一を図ることを目的に、本学では「データカタログ」というもの

を作成しています。他の大学だと、「データカタログ」という名称を使っている場合もありますが、全く違うものもあります。本学では、このデータカタログに収録されたデータを対象に、収集したものを活用するために、「ファクトブック」というものを考えております。このようなデータの収集、蓄積を行う体制の構築と、あとはデータをいかに活用し、意思決定や改善に資する情報（データ分析等）の提供、これらを目指すために、日々取り組んでいるところです。

データ利活用の一例 (ファクトブック作成のフロー例)

24

対象となるデータや根拠資料

- データ元資料名：学校基本調査（文部科学省）、大学情報データベース（大学改革支援・学位授与機構）、大学ポータル（大学ポータルセンター）
- 活用先資料名：国立大学法人評価（文部科学省）、大学機関別認証評価（大学改革支援・学位授与機構）、鳥取大学ファクトブック

データカタログ

- 定義：「いつどの部署で、どのようなフォーマットでデータを作成しているのかをまとめたもの」（馬田ほか、2015）
- 目的：学内におけるデータの所在を把握し、使用するデータ・指標の定義を統一する

ファクトブック

- 定義：「ファクトブックは、その大学の運営上、重要なデータ（経年、ときには他大学との比較）を見やすい表やグラフの形に整理し、大学執行部、学部執行部や各現場などへ意思決定や判断の支援のために供されるものである。」（馬田、2015）
- 目的：経年データを可視化することで、大学全体を数量的に俯瞰する。

2017年08月24日 大学評価・IR担当者集会2017 全体会

『データ利活用の一例』についてです。これはファクトブック作成の流れですが、やはり評価業務を中心にやっていますので、評価で集めるデータの定義を統一したいというのと、評価で収集するデータっていうのを、評価にだけ使うのではなくて、データの二次利用ということも考えたいということです。

実際、評価等で集めているデータの保有部署、データ元資料と活

用先資料名等をまとめたデータカタログを作成します。そのデータカタログには、例えば、法人評価で使う指標はこれとか、認証評価で使うのはこれというように、項目のリストを作ります。その中から、学内外に公表する項目を何個かピックアップして作ったのが、本学で言う「ファクトブック」になります。ただ、ファクトブックもデータカタログと同様、各大学によってそれらの定義自体がまちまちでしょうから、各大学で、こういう形で作成すると決めて運用したらいいかと思います。

25

鳥取大学データカタログ

- 完成形のイメージ（当日配付します）
- データカタログには、以下の2種類があると考えている。
 - 集計表対応型：既存調書の集計データが対象
 - データ対応型：生データが対象
- 本学では、評価業務で収集したデータを活用（2次利用）するため、「集計表対応型」から着手。

2017年08月24日
大学評価・IR担当者集会2017 全体会

『鳥取大学データカタログ』は、配布している資料（A3横）のようなイメージになります。このデータカタログというのは、あくまで調査票に記入された集計値をかき集めてファクトブックを作ろうという「集計値対応型」と呼んでいるデータカタログになります。集計値の代わりに、生データが集まるようなものになってくると、「データ対応型」になります。例えば、このデータは数値型で何桁というよ

うな細かな定義がされたものです。要は、データ対応型のデータカタログができれば、それはすなわちデータベースの設計書にそのまま転用できますので、まずは集計値で集めて、項目リストと所在把握を行い、さらに詳細なものを作ったら、データベースの設計まで持っていけるという流れで本学はデータカタログを作っています。

26

鳥取大学データカタログ

- データカタログは、以下の2つで構成される。
 - データ一覧表：データ・指標の項目、所在、基準日等をまとめたもの（データの所在把握）
 - データ定義書：各データの定義と収集フォーマットを定めたもの（データ定義の統一）
- データ一覧表で割り振る項目番号は、以下のルールで作成。

項目番号	説明	備考
1～2桁目	種別番号	数字2桁
3～5桁目	項目番号	数字3桁
6桁目	区分記号	英字1桁

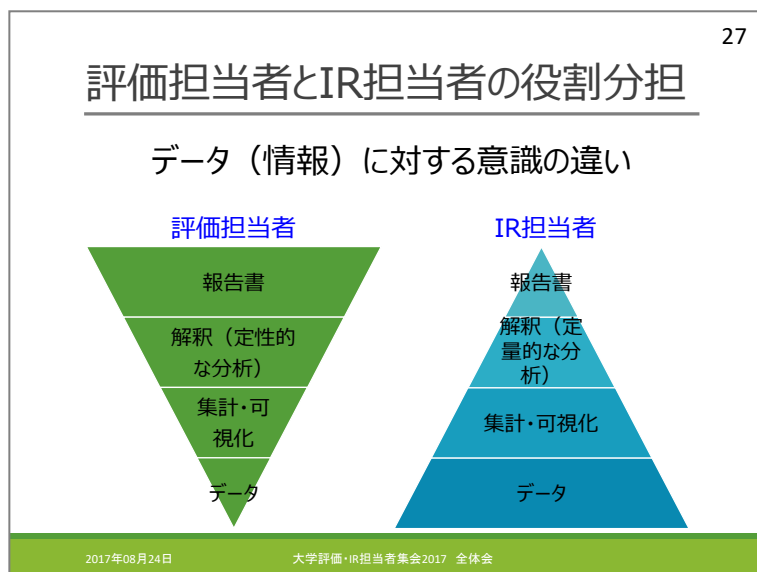
※種別名については、「新潟大学ファクトブック2015」項目一覧を参考に作成

種別番号	種別名
01	教職員
02	入試
03	学生
04	学習成果
05	進路
06	学生支援
07	研究
08	社会貢献
09	国際交流
10	財務

区分記号	区分名
A-Z	データ
a-z	指標

2017年08月24日
大学評価・IR担当者集会2017 全体会

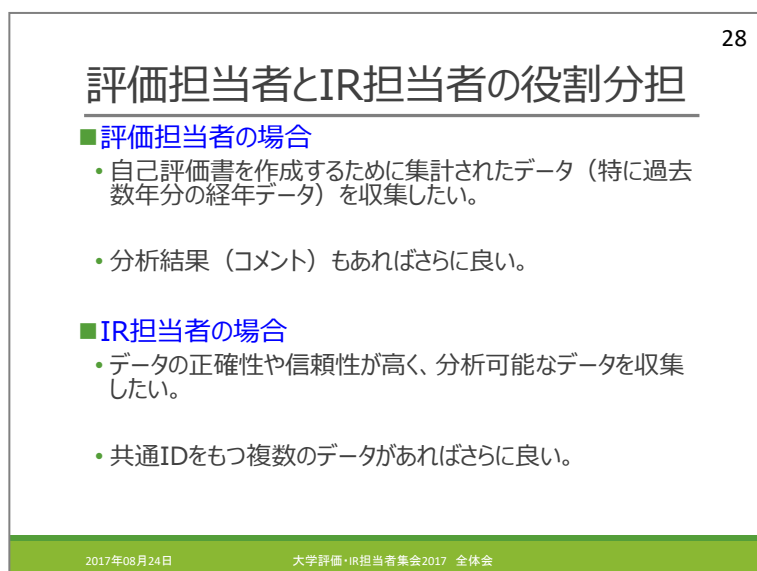
データ一覧表やデータ定義の細かな内容とかは、配布した資料を見てください。このスライドでは、こういった形でコードを振っていますという説明になります。



かなり時間が押していますが、『評価担当者と IR 担当者の役割分担』についてです。これは私の個人的な見解になりますが、評価担当者が一番重きを置いているものは、やはり報告書の作成ではないかと思えます。その際にデータは集めるんですけども、あくまで報告書を書くための根拠として集めているので、ざっくりと言えば、集計された経年変化のデータがあれば、実績を示す文章（テキストのデー

タ）と上手く組み合わせながら、それらをうまく解釈（分析）しながら報告書を作っている、ということでないかと思えます。

けど、IR 担当者だと、評価担当者のようにそもそも報告書を書かなければいけないわけではなくて、データから現象を数量的に明らかにするっていうことが、主な目的になります。そのためには、当然、データの集計、可視化、分析というものを含めて、いかにデータを分かりやすいような形で提示できるか、といった点に重きを置くことになろうかと思えます。右図だと、中盤から下のところが該当する部分であり、データの質がかなり良くないと、解釈（定量的な分析）まではたどり着けない。こういったところも、両者の意識の違いがあるのかなと思ひ、ここでは上下が異なる図で示しています。



評価担当者の場合、自己評価書を作成するために、集計されたデータを収集します。担当者は、担当部署である相手先がデータの集計を行うものと考えており、できれば集計したデータを踏まえて傾向とか分析もして欲しい、っていうのを考えていると思ひます。これが、IR 担当者の場合だと、生データさえ収集できれば、相手先がやるよりは自分たちで集計したほうが、当然、色々な分析ができます

ので、集計っていう作業は自分たちの組織がやるもんだと考えていると思ひます。

例えば、収集した複数のデータがあっても、それらを結合しようとなると、キーとなるような共通 ID が必要になります。それぞれのデータに、共通 ID が含まれるデータの形式に統一されて

いるとありがたいと IR 担当者は考えるでしょう。

29

評価担当者とIR担当者の役割分担

- **評価担当者**には、以下の点を意識して欲しい。
 - 収集するデータの絞り込み、関係部署とのデータ定義の共有、提供されたデータ（情報）が示す意味、データ（情報）から考える改善支援策
- **IR担当者**には、以下の点を意識して欲しい。
 - データ収集や分析に係る目的の明確化、関係者へのデータ提供、分かりやすい分析結果の提示、関係部署とのコミュニケーション

評価担当者（部署）とIR担当者（部署）の連携例として、自己評価書の作成において、**データ収集・蓄積は評価部署**が担当し、**データ分析はIR部署**が担当し、分析結果を評価部署へフィードバックする、など。

2017年08月24日
大学評価・IR担当者集会2017 全体会

最後に、ここでは評価担当者及びIR担当者に意識してほしい点を挙げています。

評価担当者には、単にデータが集計されたものを集めるだけでなく、やはり、そのデータはどのような方法で集計されたのかについて、ちゃんと認識していただいて、データ（情報）が示す意味についてものを見てほしいということです。

IR 担当者には、データ収集や分析に係る目的がなかなか明確になっていないと思われるのですが、やはり、読んでいただく方に分かりやすい分析結果の提示を意識してほしいということです。例えば、詳細な分析結果を提出しても執行部になかなか理解していただけないのなら、まずはクロス集計等の簡単な結果を出しておいて、その場での議論を確認しながらどんどんニーズを掘り下げていき、最終的に詳しい分析結果を出していくというような流れにするなど、データや分析結果の出し方を工夫するとか、事前に関係部署とのコミュニケーションを密に行っておくとか、こういったことを重点的にやっていたらどうかと思います。

今回の全体会テーマは「評価と IR の結節点」なので、例えば、認証評価への対応として、評価と IR の部署が連携する場合、評価部署では目的に応じてデータ収集・蓄積を行い、IR 部署ではそれらデータを使って分析を行い、その分析結果を評価部署にフィードバックし、最終的に評価部署が分析結果を踏まえて報告書を完成させる。ちょっと個人的な提案になりますが、こういった連携の仕方もあるんじゃないかなと考えています。

他にも各大学によって、色々な連携の仕方はあると思いますので、評価部署と IR 部署、その2つの部署が学内にあるようでしたら、そこで共通するような業務や取組、そういったものを話し合われて、何らかの連携を模索されてはいかがでしょうか、というのを最後のまとめとして、発表を終わらせていただきます。