

# 大学基本情報の分析環境の紹介と BIレポート作成・管理のtips

独立行政法人 大学改革支援・学位授与機構  
大学ポータルセンター事務局



周東 夏希 (大学情報推進係長)  
筒井 優子 (大学情報推進係特任スタッフ)  
佐々木 伸 (大学情報システム係員)  
金原 英徳 (室長補佐)  
2019年8月22日@大学評価・IR担当者集会  
IR実務担当者セッション

## 本報告の構成

0. 大学ポータルセンターと情報活用
1. 分析環境の提供の概要
2. BIレポート化のためのデータ加工作業
  - (1) 準備
  - (2) データ整形・クリーニングとその自動化
  - (3) データ形式の選択と管理方法

## 0. 大学ポートレートと情報活用

ざっくりいうと…

- ① これまで大学ポートレートシステムでデータを集めてきた。
- ② 集めた関連データを有効活用できるようにすべきという目標がある。
- ③ Excel等でデータを共有してきたが、使いやすさに限界がある。

2

## 0. 大学ポートレートと情報活用 (1) 計画等

### 機構の第4期（2019～2023年）中期目標

大学や評価機関等と連携し、大学等の教育研究の質の維持向上、運営基盤の強化の促進、大学等が社会に対する説明責任を果たすための教育情報の公開・活用等を支援する。

### 機構の第4期中期計画

大学等における教育研究の質の維持向上に資する情報等を収集、蓄積し、大学等が評価活動やIR（インスティテューショナル・リサーチ）活動等に活用できるよう提供する。また、大学等の教職員向けの研修の開催等により、質保証に関わる人材の能力向上を支援する。

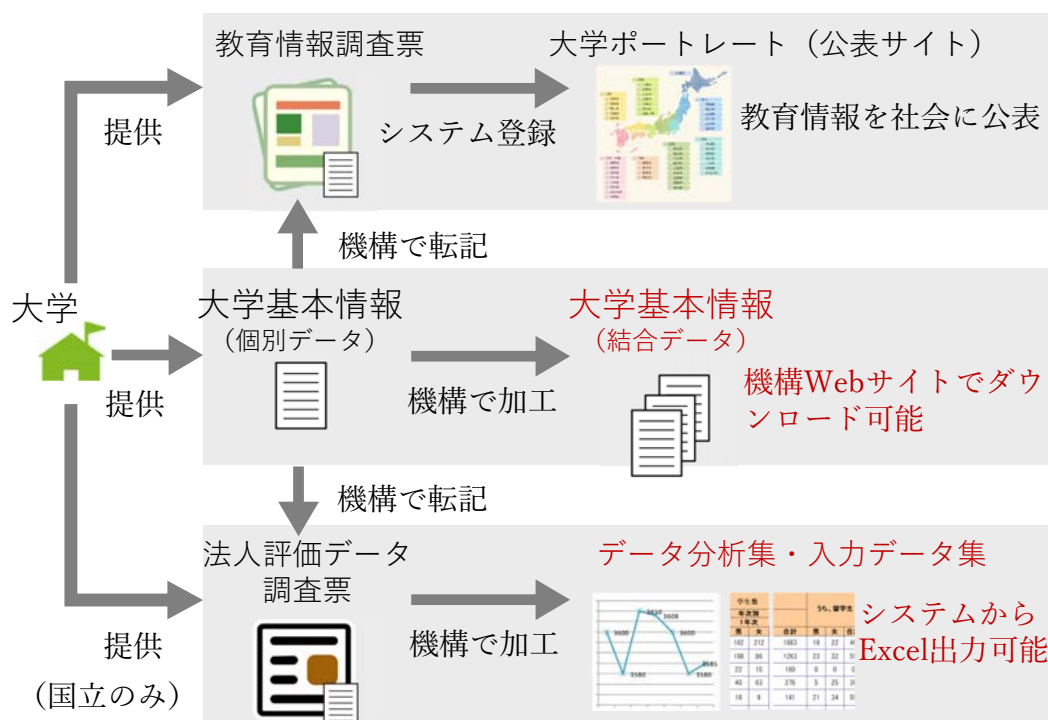
### 大学ポートレート（仮称）構築のための論点整理（2014年2月） （大学ポートレート（仮称）準備委員会）

各大学が教育改善のサイクルを機能させていく上で、そのための検討、判断の材料として、自大学の状況の把握・分析を客観的な情報に基づいて行うことは極めて重要である。大学ポートレート（仮称）に集められる情報は、関係大学の理解の下に、そのような各大学の取組にも活用できるようにし、我が国の大学教育全体の質の保証や向上を図っていくことが望ましい。…各大学における教育改善のための情報の有効活用を促進するため、大学ポートレートセンター（仮称）において、日本私立学校振興・共済事業団や他の大学関係団体の取組とも連携しつつ、情報分析を効果的に行うことのできる人材（IR人材）の育成に取り組むことが期待される。

3

## 0. 大学ポートレートと情報活用 (2) 過去の取組

国公立大学の「大学基本情報」と、国立大学の法人評価用データ（「データ分析集」・「入力データ集」）を大学等に提供。



4

## 1. 分析環境の提供の概要 (1) 環境の構成

### ●これまでの課題

Excelで分析作業するために手間がかかる。特に部局等では活用しにくい。

### ●解決策

大学基本情報等について、大学ポートレート参加大学を対象に、**BIツール**を利用した分析環境を提供したい（今秋提供開始予定）。

- Office 365の環境により、アカウントごとに閲覧権限を設定して、大学間だけでデータ共有が可能。
- 利用者は学校名の選択などの簡易な操作だけで分析を出力できる。

**分析環境について  
当日デモをします**

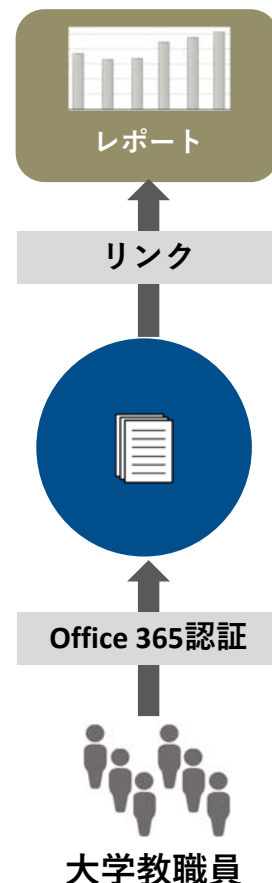
#### Power BI Service (Office 365)

機構が作成したレポートを大学が閲覧。  
大学基本情報については、公立だけでなく国立大学等のデータも閲覧可能。

#### Share Pointサイト (Office 365)

お知らせ・関連ドキュメント（マニュアル、元データ等）

Office365の招待アカウントを機構が払い出し（数の制限なし＝部局教職員なども利用可）



5

## 1. 分析環境の提供の概要 (2) 提供データの種類と範囲

データの種類	提供範囲
大学基本情報	国・公・公短・株立間
法人評価データ	国立間
公立大学実態調査データ	公立間

\* 「公立大学実態調査」：公立大学協会が実施する調査。科研費など、大学基本情報よりも詳細なデータが含まれている。新たに機構に提供いただき、分析環境を構築予定。

6

## 1. 分析環境の提供の概要 (3) BI化にあたっての配慮

- レポートだけではなく、元ファイルを提供する  
→ **大学担当者が再利用しやすいような加工が必要**
- データ修正対応、毎年のデータ追加を滞りなく行う
- 新しい指標・ビジュアルの追加要望に対応する  
(頓挫することが許されない対外的な責任を伴う)  
→ **可能な限り、データ加工・読込手順の半自動化が必要**  
→ **業者任せでなく、担当者自らの技術習得と継承が必要**

7

## 2. BIレポート化のためのデータ加工作業

- (1) 準備
- (2) データ整形・クリーニングとその自動化
- (3) データ形式の選択と管理方法

8

### 2. データ加工 (1) 準備

(一般的に) データの管理とは…

★組織で保持しているデータをユーザーが使いやすいようにしておく。

- ・ データの目録および保管場所を明示
- ・ 同系統のデータについてフォーマット統一
- ・ 常に最新の情報に更新

9

## 2. データ加工 (1) 準備

立ちはだかる“ハードル”

- ・ データ分析に使うデータって何だ...
- ・ データはあるけどどうやって分析するんだ...
- ・ 分析するためには何が必要なんだ...
- ・ そもそも分析って何だ... \ (^o^ ) /



→ データ分析を行うための準備について、一例をご紹介します。

10

## 2. データ加工 (1) 準備

BIツールを使ってみましょう

★いろいろなビジュアルで、データを鮮やかに視覚化します。

- ・ デモでお見せしました。種々のデータを取り込めます
- ・ データ同士を共通の項目で関連付けて動的にグラフを表現
- ・ ビジュアルの種類も豊富

→ BIツールを使うためにはいくつか準備をしなければいけません。

11

## 2. データ加工 (1) 準備

プロセスをつくる (準備)

★データを入手したら、どのような処理を行うのが整理する。

- ・分析しやすい形に加工する
- ・ファイル形式、フォーマットを揃える
- ・データ項目を整備する           etc...

これらを行うことで、BIツールを使うハードルを下げるすることができます。

それぞれ具体的にどのような作業をしたのかご紹介します。

12

## 2. データ加工 (2) データ整形・クリーニングとその自動化

- 正確で効果的な分析ができるためには、データソースがクリーンで整っていることが必要  
→データ整形、データクリーニング
- **データ整形**  
→数値の列名を適切に変換し、統一  
→横並びのデータを縦並びの行データに変換
- **データクリーニング**  
→正確なコードと値、余分な行の削除、データ型の設定とデータ型に合わないデータの変換、データが未入力のセルの対応。
- データは毎年増える  
→毎年、データ整形とクリーニングが必要  
→ **データ整形・クリーニングの自動化**

13

## 2. データ加工 (2) ①データ整形・クリーニング (Power Query)

### ● 2012~2018年までの大学基本情報「(7-A) 学生数」のデータ整形・クリーニング

年度	学校調査番号	学校名	昼夜	学生数														計				
				大学院						学部・本科		専攻科		別科		科目等履修生・聴講生・研究生						
				博士課程		修士課程		専門職学位課程								学部卒以上		左記以外				
男	女	男	女	男	女	男	女	男	女	男	女	男	女	男	女	男	女	計				
2018	0100	北海道大学	昼間	1,661	687	2,680	1,104	154	53	8,055	3,291					190	189	305	236	13,045	5,560	18,605
2018	0100	北海道大学	夜間																	0	0	0
2018	0104	北海道教育大学	昼間			111	108	64	15	2,397	2,642			0	21	10	9	16	46	2,598	2,841	5,439
2018	0104	北海道教育大学	夜間																	0	0	0
2018	0108	室蘭工業大学	昼間	50	20	427	47			2,287	355	0	0	0	0	11	9	2	0	2,777	431	3,208
2018	0108	室蘭工業大学	夜間							182	13	0	0	0	0	0	0	0	0	182	13	195



年度	学校調査番号	学校名	昼夜別	課程区分	性別	学生数
2018/5/1	0100	北海道大学	昼間	大学院_博士	男	1661
2018/5/1	0100	北海道大学	昼間	大学院_博士	女	687
2018/5/1	0100	北海道大学	昼間	大学院_修士	男	2680
2018/5/1	0100	北海道大学	昼間	大学院_修士	女	1104
2018/5/1	0100	北海道大学	昼間	大学院_専門職学位	男	154
2018/5/1	0100	北海道大学	昼間	大学院_専門職学位	女	53
2018/5/1	0100	北海道大学	昼間	学部_本科	男	8055
2018/5/1	0100	北海道大学	昼間	学部_本科	女	3291

14

## 2. データ加工 (2) ①データ整形・クリーニング (Power Query)

### ExcelのPower Queryの概要

- 分析用のデータに変換する豊富な機能を備えている
- Power Query エディタにデータを読み込み、次のような操作をすることで、データ変換コード（クエリ）を作成する
  - ・ 行や列の削除、列の追加
  - ・ データ一括変換、エラーデータ変換
  - ・ 横列の項目を縦列データへ変換
  - ・ 列データの分割（区切り記号、文字数等）
  - ・ …等
- 「Excelに読み込む」機能を選択すると、クエリを実行し変換されたデータをExcelの新しいシートに読み込む

15



## 2. データ加工 (2) ①データ整形・クリーニング (Power Query)

### データ整形・クリーニング手順

- 必要な操作をあらかじめ、定義する  
→**再利用可能**なPower Queryコードを記録できる

操作手順	追加/削除/修正	列名	整形内容	データ値	処理モード (Excel/Query エディタ)
1	修正	学校名	[学校調査番号]を検索キーとし、[学校マスタ]から再検索	学校マスタの[学校]	Excel
2	削除	全て	黄色背景色のヘッダーを削除する。		Queryエディタ
3	設定	全て	列のデータ型を設定する。 学生数の列は整数、その他は文字列にする。		Queryエディタ
4	追加、削除	年度	①[年度2]カラムを追加し、計算式「年度&"/05/01"」でデータを生成する。 ②[年度]列を削除する。 ③[年度2]列の名前を「年度」に変更する。		Queryエディタ
5	削除	計_男、計_女、計_計	列削除		Queryエディタ
6	変換	[学生数]列グループ	①[学生数]列グループのすべての列に対し、以下の変換を行う。 数値セルの「null」を0へ変換。エラーを0へ変換。		Queryエディタ
7	修正	ヘッダー	①[学生数]列グループのヘッダーを次のように修正する。 博士_男、博士_女 ⇒ 大学院_博士_男、大学院_博士_女 修士_男、修士_女 ⇒ 大学院_修士_男、大学院_修士_女 専門_男、専門_女 ⇒ 大学院_専門職学位_男、大学院_専門職学位_女 本科_男、本科_女 ⇒ 学部_本科_男、学部_本科_女 専攻_男、専攻_女 ⇒ 専攻科_男、専攻科_女 学卒上_男、学卒上_女 ⇒ 科目等履修生・聴講生・研究生_学部卒以上_男、科目等履修生・聴講生・研究生_学部卒以上_女 以外_男、以外_女 ⇒ 科目等履修生・聴講生・研究生_左記以外_男、科目等履修生・聴講生・研究生_左記以外_女	画像参照	Queryエディタ
8	追加	学生数	①[学生数]列グループをピボット解除する。 ②[値]列名を「学生数」とする。		Queryエディタ
9	追加	課程区分	①ピボット解除後に作成された[属性]列を次のように列分割する。 区切り文字：「」 分割方法：一番右の区切り記号 ②分割して作成された[属性.1]の列名を「課程区分」に変更する。	画像参照	Queryエディタ
10	追加	性別	①分割して作成された[属性.2]の列名を「性別」に変更する。		Queryエディタ

16

## 2. データ加工 (2) ①データ整形・クリーニング (Power Query)

### Power Query利用時の注意点

- 列名の変更と追加をすると、クエリを再利用できない
- 列名（ヘッダー）の変更が発生しない工夫  
→データファイル名・シート名ごとに列名を定義  
→データ整形・クリーニング手順を作成する上で利用

ファイル名	シート名	column1	column2	column3
00_共通マスタ	00_G_学科系統マスタ (短大)	学科系統番号	並べ替え用番号	学科系統
00_共通マスタ	00_H_主たる学部系統マスタ (大学・短)	主たる学部系統番号	並べ替え用番号	主たる学部系統
00_共通マスタ	00_I_学科系統マスタ (大学・短大)	学科系統番号	並べ替え用番号	学科系統
00_共通マスタ	00_J_入学者前歴マスタ	入学者前歴番号	入学者前歴	
07_1_本務教員数再掲	07_1_1_本務教員数再掲	年度	学校調査番号	学校名
07_A_学生総数	07_A_1_学生総数	年度	学校調査番号	学校名
07_A_学生総数	07_A_2_課程区分マスタ	課程区分番号	課程区分	課程中区分番号
07_B_本務教員数	07_B_1_本務教員数	年度	学校調査番号	学校名
07_B_本務教員数	07_B_2_所属マスタ	所属番号	所属	

17

## 2. データ加工 (2) ②データ整形・クリーニングの自動化

- 2019年度のデータ整形・クリーニングを  
2018年度までのデータと同様に、  
漏れなく正確に行い、迅速に更新する。

年度	学校調査番号	学校名	昼夜	学生数																計		
				大学院						学部・本科		専攻科		別科		科目等履修生・聴講生・研究生				男	女	計
				博士課程		修士課程		専門職学位課程		男	女	男	女	男	女	学部卒以上		左記以外				
				男	女	男	女	男	女							男	女	男	女	男	女	男
2019	0100	北海道大学	昼間	1,675	689	2,722	1,184	127	48	8,044	3,267					177	178	312	290	13,057	5,656	18,713
2019	0100	北海道大学	夜間																	0	0	0
2019	0104	北海道教育大学	昼間			94	95	56	12	2,384	2,671	0	0	1	36	15	9	21	46	2,571	2,869	5,440
2019	0104	北海道教育大学	夜間																	0	0	0
2019	0108	室蘭工業大学	昼間	55	23	458	62			2,227	372					20	5	3		2,763	463	3,226
2019	0108	室蘭工業大学	夜間							185	15					0	0	0		185	15	200



年度	学校調査番号	学校名	昼夜	課程区分	性別	学生数
2019	0332	岡山大学	夜間	科目等履修生・聴講生・研究生・左記以外	女	1
2019	0336	広島大学	夜間	大学院_博士	男	21
2019	0336	広島大学	夜間	大学院_博士	女	24
2019	0336	広島大学	夜間	大学院_修士	男	28
2019	0336	広島大学	夜間	大学院_修士	女	28
2019	0336	広島大学	夜間	学部_本科	男	267
2019	0336	広島大学	夜間	学部_本科	女	155
2019	0336	広島大学	夜間	科目等履修生・聴講生・研究生_学部卒以上	男	1
2019	0336	広島大学	夜間	科目等履修生・聴講生・研究生_学部卒以上	女	3
2019	0336	広島大学	夜間	科目等履修生・聴講生・研究生_左記以外	男	3

18

## 2. データ加工 (2) ②データ整形・クリーニングの自動化

- 手順1. Excelのマクロ記録を開始する。
- 手順2. Excelで①の手順どおりに操作を行う。
- 手順3. Power Queryのコード（再利用）を貼り付ける。
- 手順4. Excelのマクロ記録を終了する。
- 手順5. Excelシートにボタンを配置し、マクロを関連付ける。

2019\_07\_A\_学生総数.xlsm - 保存済み

ホーム 挿入 描画 ページレイアウト 数式 データ 校閲 表示 開発 アドイン

データ整形・クリーニング手順

- 1 整形対象の「07\_A\_学生総数」データが格納されているシートを本ファイルにコピーします。
- 2 マスターファイルを開きます。
- 3 学校名を[学校マスタ]で正規のものに置き換えます。
- 4 整形操作をします。〈データ整形・クリーニング スタート〉ボタンをクリックします。

データ整形・クリーニング スタート

- 5 [ヘッダー等名称仕様]シートのとおりシート名を設定します。

19

## 2. データ加工 (3) データ形式の選択と管理方法

★いろいろな“かたち”のデータがあります。

- ・ ファイル形式の違い (.xlsx, .accdb, .csv, .xmlなど)
- ・ フォーマットの違い

1	2	3	4	5	6	
1	年度	学部等コード	学部等名	役職	男	女
2	2010/5/1	01-1200	北海道大学 水産学部	教授	1	0
3	2010/5/1	01-1200	北海道大学 水産学部	准教授	4	0
4	2010/5/1	01-1200	北海道大学 水産学部	講師	0	0
5	2010/5/1	01-1200	北海道大学 水産学部	助教	4	0
6	2010/5/1	01-1200	北海道大学 水産学部	助手	0	0
7	2010/5/1	01-1200	北海道大学 水産学部	教授	51	3
8	2010/5/1	01-1200	北海道大学 水産学部	准教授	33	7
9	2010/5/1	01-1200	北海道大学 水産学部	講師	0	0
10	2010/5/1	01-1200	北海道大学 水産学部	助教	10	5
11	2010/5/1	01-1200	北海道大学 水産学部	助手	0	0

1	2	3	4	5	6	
1	年度	学部等コード	学部等名	職名	性別	教員数
2	2010/5/1	01-1200	北海道大学 水産学部	教授	男	1
3	2010/5/1	01-1200	北海道大学 水産学部	教授	女	0
4	2010/5/1	01-1200	北海道大学 水産学部	准教授	男	4
5	2010/5/1	01-1200	北海道大学 水産学部	准教授	女	0
6	2010/5/1	01-1200	北海道大学 水産学部	講師	男	0
7	2010/5/1	01-1200	北海道大学 水産学部	講師	女	0
8	2010/5/1	01-1200	北海道大学 水産学部	助教	男	4
9	2010/5/1	01-1200	北海道大学 水産学部	助教	女	0
10	2010/5/1	01-1200	北海道大学 水産学部	助手	男	0
11	2010/5/1	01-1200	北海道大学 水産学部	助手	女	0
12	2010/5/1	01-1300	北海道大学 文学研究科	教授	男	51
13	2010/5/1	01-1300	北海道大学 文学研究科	教授	女	3
14	2010/5/1	01-1300	北海道大学 文学研究科	准教授	男	33
15	2010/5/1	01-1300	北海道大学 文学研究科	准教授	女	7
16	2010/5/1	01-1300	北海道大学 文学研究科	講師	男	0
17	2010/5/1	01-1300	北海道大学 文学研究科	講師	女	0
18	2010/5/1	01-1300	北海道大学 文学研究科	助教	男	10
19	2010/5/1	01-1300	北海道大学 文学研究科	助教	女	5
20	2010/5/1	01-1300	北海道大学 文学研究科	助手	男	0
21	2010/5/1	01-1300	北海道大学 文学研究科	助手	女	0

20

## 2. データ加工 (3) データ形式の選択と管理方法

目指すところ

★データに馴染みがない人にも気軽にデータ分析に触れてほしい。

- ・ 可能な限り簡単な仕組みで
- ・ データの正確性を失わない
- ・ 何度でもやり直し操作ができる

ような管理方法を作りたい。

21

## 2. データ加工 (3) データ形式の選択と管理方法

### Excel-CSV方式

★私たちはExcelファイルと.csvファイルの併用を選びました。

- ・ Excelは最も馴染みのあるソフトウェアの1つ
- ・ .csvファイルは非常に軽い
- ・ .xlsxから.csvに簡単に変換できる



.csvファイルはBIツールへの読み込みも非常に速い。  
→データ更新の機会が多くなっても対応しやすい。

22

## 2. データ加工 (3) データ形式の選択と管理方法

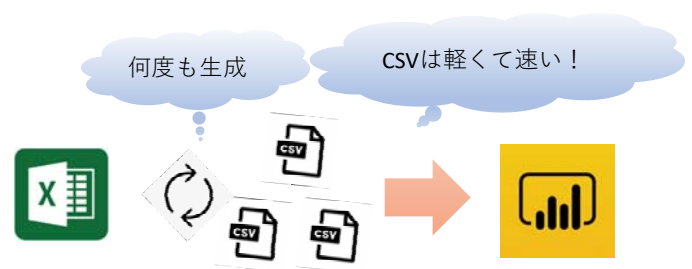
### 作業のながれ

1. 大元のデータはExcelで管理する。

1Book - 1Sheetでも複数Sheetでも良い。

2. Excelファイルを.csvに変換する。

自動化すると速い。



3. BIツールに取り込み分析する。

csvファイルは読み込みが速く、便利。

※データに修正が生じた場合は、Excelファイルを修正する。

Excel修正時はcsvを一旦破棄し、再度Excelファイルからcsvを生成する。

23

## 2. データ加工 (3) データ形式の選択と管理方法

Excel-CSV方式をとりいれて

1. データを修正するときはExcelファイルを修正する。

→ 慣れているため修正が簡単。

1つのセルの値だけ直したいときなどはExcelが最も簡単。

2. Excelを更新するたびにcsvを更新する。

→ データ更新時、csvは必ず破棄。

毎度作り直すことで、常に最新の状態を保つことができる。

3. Excelファイルのデータ項目を一覧化しておく。

→ 一覧表を見ることでデータの種類、在り処が瞬時にわかる。

24

## 2. データ加工 (3) データ形式の選択と管理方法

“しくみ”の確立

★データを加工し形式を揃える、というプロセスを確立しました。

- ・作業に迷いがなくなった。
- ・どんなデータ分析をするかという課題に専心できるようになった。
- ・新しく業務に取り組む人たちに説明しやすくなった。

→ データ分析の裾野を広げるための第一歩であると感じています。

25

## 2. データ加工 (3) データ形式の選択と管理方法

慣れてきたら…

★世の中には便利なサービスがたくさんあります。

- ・クラウドサービス (AWS、Azure...等)
- ・商用データベース (Oracle、MySQL、PostgreSQL...等)

★いきなり本格的なデータベースから入るのは難しいと感じました。

- ・そのための“Excel-CSV方式”
- ・データの扱い方についてはweb上に多くの情報が転がっています
- ・色々な人の色々な取り組みを知る

→ 自分たちの組織に合った自分たちのメソッドをつくりあげる

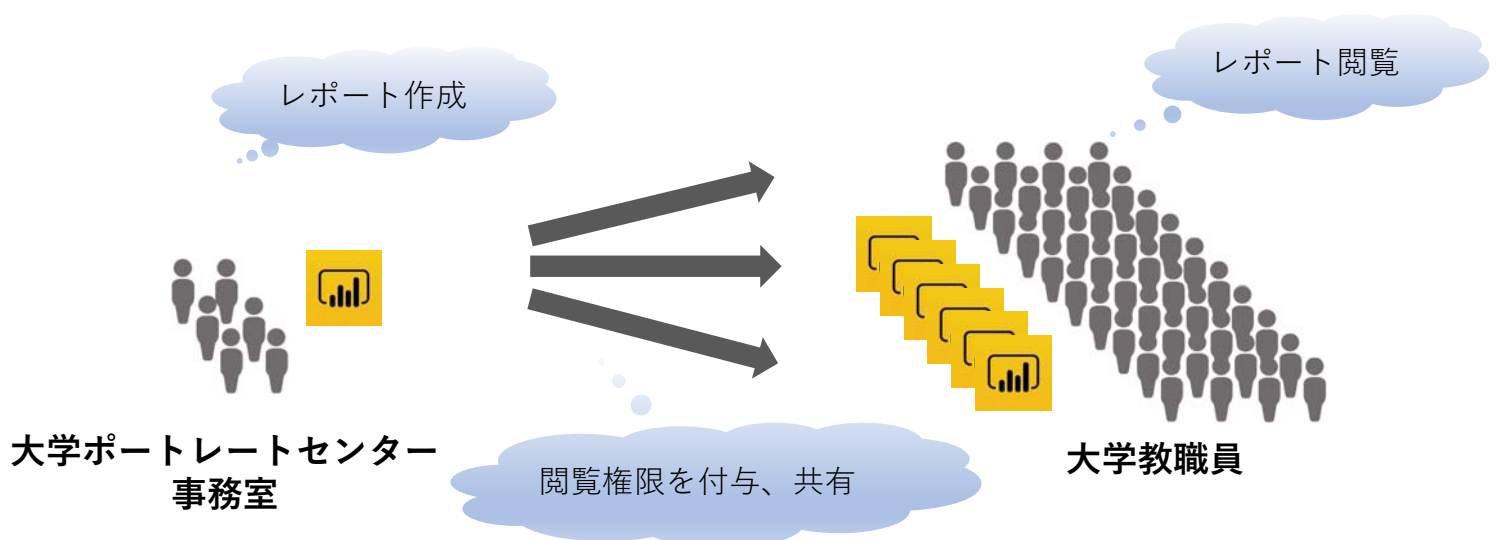
※挫折する可能性を最小限に留めることが最重要課題！

26

## 参考

★私たちのデータ分析環境をご紹介します。

- ・ PowerBIというBIツールを使用しています。
- ・ PowerBI Proというサービスに加入することでメンバーとレポートを共有できます。
- ・ PowerBI Premiumというサービスを契約し、閲覧環境を多くの人に開放します。



27