



山形大学
Yamagata University

IR担当者に データマネジメントのスキルが 必要な理由

山形大学 学術研究院 教授

藤原 宏司

2020.9.24

大学評価・IR担当者集会2020
IR初級セッション

今日の内容

- データマネジメントとは？
IR (Institutional Research) とは？
- IRとデータマネジメント
- データマネジメントの例
- まとめ

自己紹介

専門はIRと応用統計学 (Ph.D.)。米国の大学・短期大学等で統計解析、IRおよび大学評価対応業務に従事。2016年8月から現職。フロリダ州立大学大学院IR履修証明プログラム修了。

データマネジメントとは？
IR (Institutional Research) とは？



データマネジメントとは？

■ データマネジメントとは、分析に必要なデータを

- **分析可能な形で準備**すること
(含：データ収集、変形、修正等)
- **将来に備えて整理**すること
(含：データの更新、追加、修正等)

■ なぜデータマネジメントが必要なのか？

- **意思決定には、高品質なデータが必要だから**
(Data Informed Decision Making)

→ Quality in, quality out (Garbage in, garbage out)

IR (Institutional Research) とは？

- **IR**とは、**米国の大学**で発展してきた、**大学経営**や**教育改善**を**Support**する**機能**

ポイント

- 日本の大学で生まれた機能・考え方ではない
 - **輸入された考え方**（ゆえに、人によって解釈が違う）
 - i.e., GPA、**Support**、Research
- **IRにおける「Research」**をどのように解釈するか
 - 米国の大学におけるIR業務や、**米国、英国での「Research」の使われ方から考えると、「ちょっとした調査やデータ（情報）収集」**では？
- **Support機能**
 - 裏方・黒子であり、花形ではない

IRとデータマネジメント



IRとデータマネジメント（1）

■ IRの定義（一つの考え方）

IRとは、**計画立案のサポート**、**政策形成のサポート**、**意思決定のサポート**に資する**情報提供**を目的とした、**学内外データの収集・分析（活動／機能）**のこと

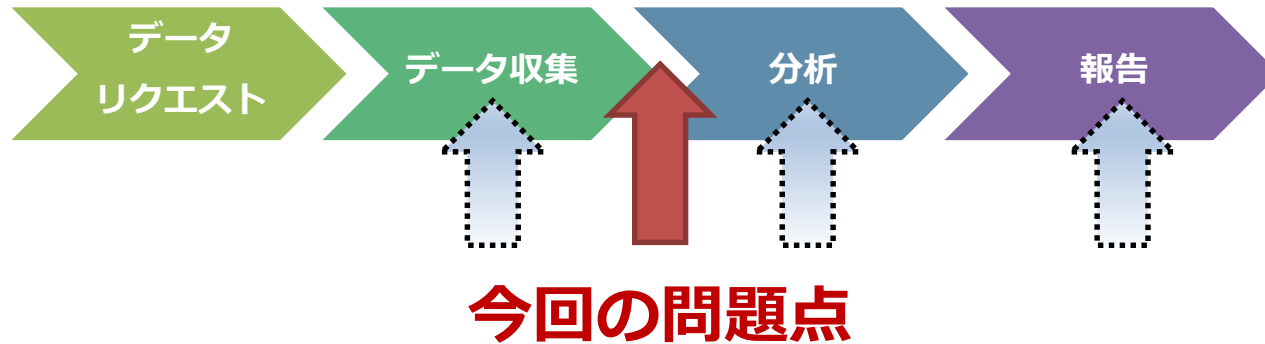
→ **意思決定者**（i.e., 執行部）の**サポート**

ポイント

- 研究が主目的ではない
(Educational Research ≠ Institutional Research)
- **米国FERPA**：IR担当者であっても、個人の研究目的が理由の場合は、学生データにアクセスすることはできない

IRとデータマネジメント（2）

IRの代表的な役割（とされているもの）



データ収集と分析の間にある問題点

1. **データの内容に問題**があつて、分析ができない
2. **データの形式に問題**があつて、分析ができない

データマネジメントが必要になる

※ 「データが収集できない」などの問題は、ここでは取り扱いません

データマネジメントが必要な例

【データの内容に問題がある】表記ゆれ

氏名	大学名	住所
山本幸一	学校法人明治大学	東京都千代田区
鈴木 達哉	山形大学	山形市小白川町1丁目4番12号
浅野 茂	国立大学法人 山形大学	山形県山形市小白川町1-4-12
藤原 宏司	山形大 法人本部	山形市

主な原因

- DBではなく、エクセルで管理している
- データ入力担当者が定期的に変わる

【データの形式に問題がある】複数選択式問題の回答データ

ID	併願候補大学	ID2
1	岩手大学;東北大学	00000001
2	岩手大学;東北大学;宮城教育大学	00000002
3	秋田大学	00000003
4	弘前大学;秋田大学	00000004
5	弘前大学;岩手大学;東北大学;秋田大学;福島大学	00000005

余談：米国のIR担当者とデータマネジメント

- 米国の一般的な大学では、
データが統合管理されているので、

「**データの内容に問題**があつて、分析ができない」

というケースに直面することは、ほとんどありません

■ 理由

- データ入力担当者に、データの取り扱い（i.e, データ入力）に関する**裁量の自由が無い**
- 担当者が定期的に異動する、という文化が無い（大学を辞めることは頻繁にあります）

IRとデータマネジメント (3)

IRの役割

- 意思決定者のサポート



意思決定者の要求

- I need the data yesterday!



意思決定者を効果的にサポートするためには、**高品質なデータ** (= エラーが無いデータ) を手際よく準備して、**効率的にデータ分析と情報提供**を行う必要がある

DMA: Data Management & Analysis

そのデータマネジメント、手作業でやりますか？

- IR担当者に必要なスキル：
エラーが無いデータを手際よく準備
- 僕は、「R」と「Access」（たまにVBA）を使ってデータマネジメントを行っています
 - 山形大学のIRオフィスにおける他のメンバーは、データマネジメントの際、それぞれの得意な手法を使っています（多分）
 - 鈴木達哉さんの場合
 - VBA, MS Power Query, Access, Python(?)

比較：プログラム（R） vs. 手作業

比較項目	R	手作業
単純な手作業	自動化できる	ひたすらコピペ
エラー	規則的	ランダム
エラーの特定	比較的容易	難しい (作業をやり直す?)
同じ作業の再現 (追加リクエスト等)	簡単	根性と時間があれば (人間関係が心配)
スキルの共有	コードを共有する	手作業のコツは 伝えられるけど...
ダブルチェック	同僚にコードを 見てもらう	同僚に同じ作業を してもらう?
習熟スピード	時間がかかる	それほどかからない
根性	Rを覚えるのに必要	作業をするのに必要
充実を感じる時	プログラムが動いた!	徒労感の方が 大きいのでは?

データマネジメントの例



分析しやすいデータ: Tidy Data

■ Tidy Data (Tidyなデータ)

- データ分析・可視化ツールが「扱いやすい・分析しやすい」データのこと
- 「Power BI」等のデータ分析・可視化ツールは、「Tidyなデータ」が準備されていることが前提
- ∴ 「Tidyなデータ」が作れないと、ツールを使ったデータ分析・可視化ができない
- データマネジメント ≅ Tidyなデータを作成する (管理も)

データのTidy化 (1)

IR大学入学生：県内出身者 vs. 県外出身者

	A_県内	A_県外	B_県内	B_県外
2017	97	268	76	120
2018	88	273	71	122
2019	95	265	79	118
2020	91	272	74	120

Why not tidy?

- 列に **2つの情報**が含まれている。



それでも、集計データですが...

年度	学部	県内県外	入学者数
2017	A	県内	97
2018	A	県内	88
2019	A	県内	95
2020	A	県内	91
2017	A	県外	268
2018	A	県外	273
2019	A	県外	265
2020	A	県外	272
2017	B	県内	76
2018	B	県内	71
2019	B	県内	79
2020	B	県内	74
2017	B	県外	120
2018	B	県外	122
2019	B	県外	118
2020	B	県外	120

データのTidy化 (2)

IR大学入学生：県内出身者 vs. 県外出身者

	A学部入学者 (県内入学者)	B学部入学者 (県内入学者)
2017	365 (97)	196 (76)
2018	361 (88)	193 (71)
2019	360 (95)	197 (79)
2020	363 (91)	194 (74)

それでも、集計データですが...

年度	学部	県内県外	入学者数
2017	A	県内	97
2018	A	県内	88
2019	A	県内	95
2020	A	県内	91
2017	A	県外	268
2018	A	県外	273
2019	A	県外	265
2020	A	県外	272
2017	B	県内	76
2018	B	県内	71
2019	B	県内	79
2020	B	県内	74
2017	B	県外	120
2018	B	県外	122
2019	B	県外	118
2020	B	県外	120

Why not tidy?

- 列に**3つの情報**が含まれている。
- 1つのセルに**2つの観測値**がある。



名寄せ

- イベント1の参加者が、イベント2にも参加しているか確認したい

イベント1			イベント2			同一人物?
名前		生年月日	名前		生年月日	
藤原	宏司	YYMMDD	藤原	宏司	YYMMDD	同じ
高橋	〇〇	S490906	高橋	〇〇	S490906	同じ (漢字のミス)
高崎	××	S511213	高崎	××	S511213	同じ (漢字のミス)
山崎	△△	S530426	山崎	△△	S530426	同じ (漢字のミス)
齊藤	□□	S600919	齋藤	□□	S600919	同じ (漢字のミス)
山形	三郎	H021114	山形	三郎	S540628	別人
鈴木	達哉	S1M1D1	浅野	茂	S2M2D2	該当者なし


Rで合致率を計算

Name. x	Name. y	Prob_Matching
藤原宏司_YYMMDD	藤原宏司_YYMMDD	100.0%
高橋〇〇_S490906	高橋〇〇_S490906	94.4%
高崎××_S511213	高崎××_S511213	94.4%
山崎△△_S530426	山崎△△_S530426	94.4%
齊藤□□_S600919	齋藤□□_S600919	94.4%
山形三郎_H021114	山形三郎_S540628	71.5%
鈴木達哉_S1M1D1	高崎××_S511213	62.4%

複数選択式解答のDM (1)

- 架空の設問：（山形大学の学生が対象と想定）
東北地方にある国立大学のうち、併願を考えたことのある大学を「全て」選択してください。
- 出力されたデータ（適当に作成）

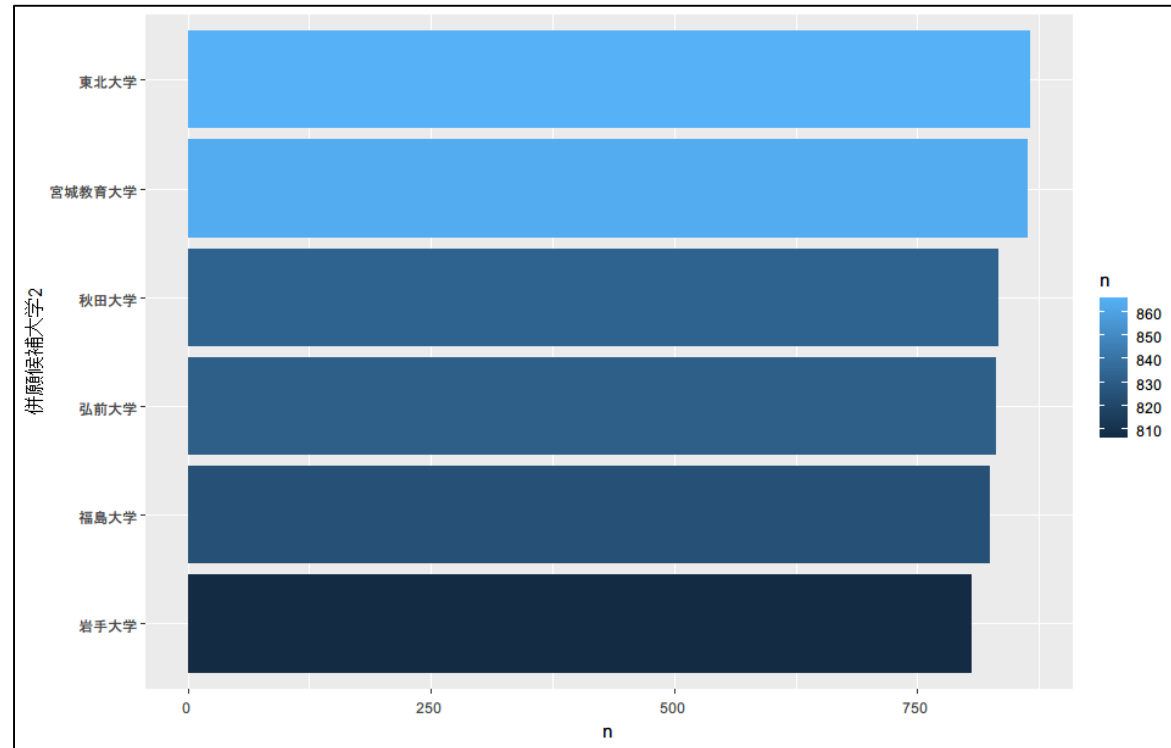
ID	併願候補大学	ID2
1	岩手大学;東北大学	00000001
2	岩手大学;東北大学;宮城教育大学	00000002
3	秋田大学	00000003
4	弘前大学;秋田大学	00000004
5	弘前大学;岩手大学;東北大学;秋田大学;福島大学	00000005
7	弘前大学;岩手大学;東北大学;秋田大学;福島大学	00000007
8	弘前大学;東北大学;宮城教育大学	00000008
9	弘前大学;東北大学;宮城教育大学;福島大学	00000009
10	弘前大学;宮城教育大学;秋田大学	00000010

複数選択式解答のDM (2)

データマネジメント後のデータ
(Rで処理)

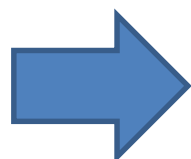
ID2	併願候補大学2
00000001	岩手大学
00000001	東北大学
00000002	岩手大学
00000002	東北大学
00000002	宮城教育大学
00000003	秋田大学
00000004	弘前大学
00000004	秋田大学
00000005	弘前大学
00000005	岩手大学
00000005	東北大学
00000005	秋田大学
00000005	福島大学

可視化例



住所から都道府県・市町村等を抽出する

大学名	住所
山形大学	当日投影
山形県立保健医療大学	
山形県立米沢栄養大学	
東北芸術工科大学	
東北公益文科大学	
東北文教大学	
山形県立米沢女子短期大学	
東北文教大学短期大学部	
羽陽学園短期大学	



Rで処理

大学名	住所	都道府県	市町村	番地以降
山形大学	当日投影	山形県	山形市	当日投影
山形県立保健医療大学		山形県	山形市	
山形県立米沢栄養大学		山形県	米沢市	
東北芸術工科大学		山形県	山形市	
東北公益文科大学		山形県	酒田市	
東北文教大学		山形県	山形市	
山形県立米沢女子短期大学		山形県	米沢市	
東北文教大学短期大学部		山形県	山形市	
羽陽学園短期大学		山形県	天童市	

使える場面：学生がどこから通っているか集計したい

余談: Rで楽ができた例 (with Excel)

1. データをグループごとに分けて、
個別のエクセルファイルに保存したい

→ それほど時間はかかりません

2. 各教員から帰ってきた
個別のエクセルファイルを一つにまとめたい

→ すぐに終わります
シートの結合も簡単です

まとめ



まとめ

- **IR担当者の役割：意思決定者のサポート**
 - エラーが無いデータを手際よく準備して、
効率的にデータ分析と情報提供を行う必要がある

- 日本の大学におけるIR担当者あるある
 - データが「すぐ利用」できる形で管理されていない

- データマネジメントとは、分析に必要なデータを
 - a. 分析可能な形で準備すること
 - b. 将来に備えて整理すること

- IRを効率的に機能させるためには、
データマネジメントのスキルが必要
(全ての大学職員に必要なスキルだと思います)

THANK YOU!

ANY QUESTIONS, COMMENTS OR SUGGESTIONS?

藤原 宏司 | Koji Fujiwara, Ph.D.

kfujiwara@cc.yamagata-u.ac.jp



参考資料

藤原宏司（2013）「IR って何？ - こんなことやっていたり、思ったりしています -」,
大学評価コンソーシアム 勉強会『米国におけるIR実践を通して考える日本型IR 実施報告書』, 7-156.

藤原宏司（2017）「データマネジメントの必要性と実践」, 大学評価・IR担当者集会2017 講演資料.

藤原宏司（2018）「FERPAおよび米国大学におけるデータの収集・管理・活用」,
大学評価コンソーシアム 大学評価・IR担当者集会2018 講演資料.

藤原宏司ほか（2019）「IR履修証明プログラムの開発について」, 『大学評価とIR』, 第10号, 3-15.

藤原宏司（2020）「IR (Institutional Research) におけるデータマネジメント」, IR初級人材育成研修会 講演資料.

Saube, J. L. (1981). The functions of institutional research. Association for Institutional Research.

Wickham, H. (2014). Tidy Data. Journal of Statistical Software, 59(10), 1 - 23.

Swing, R. L., Jones, D., and Ross, L. E. (2016). The AIR National Survey of Institutional Research Offices.
Association for Institutional Research, Tallahassee, Florida.

※ 本講演資料には、山形大学IR担当者向け実践プログラム用の講義資料（藤原宏司作成）の一部が使われています。