



# データマネジメントの 必要性と実践

山形大学 学術研究院 (IR担当)

藤原 宏司

kfujiwara@cc.yamagata-u.ac.jp

*August 24, 2017*

大学評価・IR 担当者集会 2017 全体会  
(於：立命館大学 大阪いばらきキャンパス)

# データマネジメントの必要性

- データマネジメント (DM) ≒ 分析に必要なデータを用意する (or 準備しておく) こと
- 効果的かつ効率的な**データ分析**には、**高品質なデータを手際よく整理**することが必要
  - データ分析の例 : FACTBOOK・データ集作成、退学者の特徴分析
  - 高品質なデータの例 : エラーが少ない  
(i.e., 藤原 宏司 vs. 藤原 宏司)
  - データを手際よく準備 : データを分析可能な状態にする
  - Data scientists spend around **80% of their time on preparing and managing data for analysis**<sup>1</sup>.
  - 米国の大学には、データマネジメントの専門科目がある

<sup>1</sup> Press, G (2016, March). data scientists spend most of their time massaging rather than mining or modeling data. Forbes.

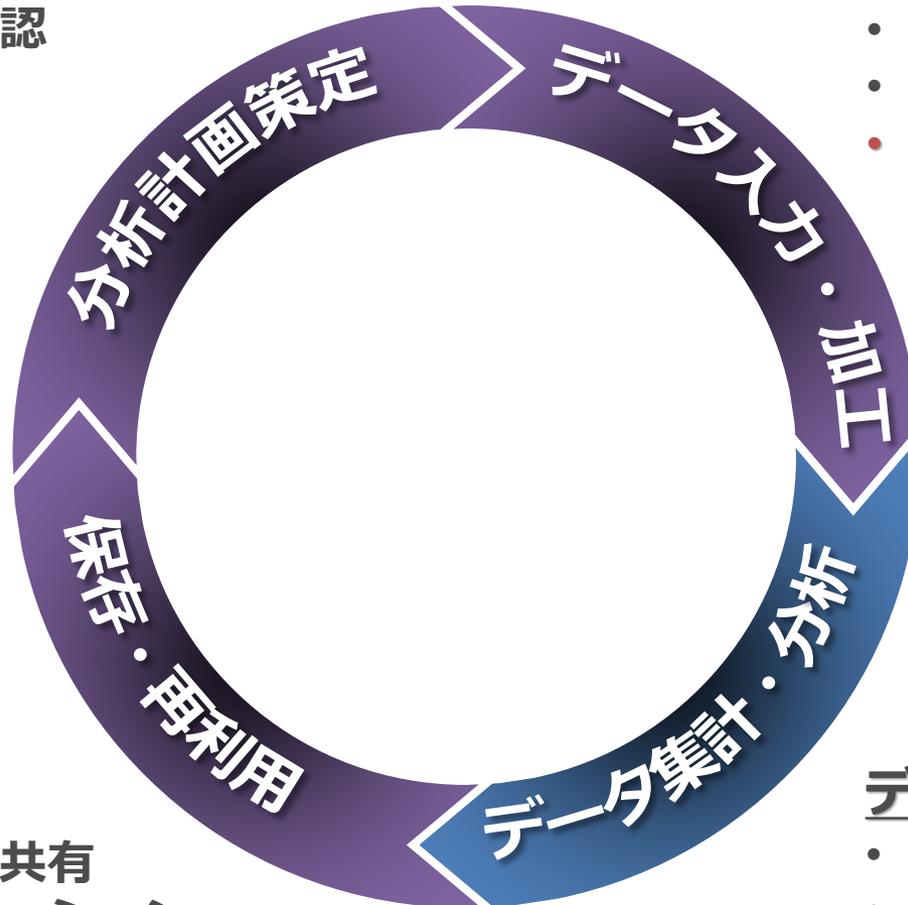
# Data Life Cycle & Data Management (DM)

## 分析計画策定

- ・ データ所在確認
- ・ データ収集

## データ入力・加工

- ・ クリーニング
- ・ エラーチェック
- ・ **データ変形**



## 保存・再利用

- ・ データ移行／共有
- ・ ドキュメンテーション

## データ集計・分析

- ・ 解釈
- ・ レポートディング

# 今日の活動に照らし合わせると

データマネジメント	FB・データ集
<ul style="list-style-type: none"><li>■ <u>分析計画策定</u><ul style="list-style-type: none"><li>・データ所在確認</li><li>・データ収集</li></ul></li></ul>	<ul style="list-style-type: none"><li>■ FB・データ集の目次作成<ul style="list-style-type: none"><li>・データカタログを参照</li><li>・同上</li></ul></li></ul>
<ul style="list-style-type: none"><li>■ <u>データ入力・加工</u><ul style="list-style-type: none"><li>・データクリーニング</li><li>・エラーチェック</li><li>・<b>データ変形</b></li></ul></li></ul>	
<ul style="list-style-type: none"><li>■ <u>データ集計・分析（≠ DM）</u><ul style="list-style-type: none"><li>・結果の解釈</li><li>・レポートニング</li></ul></li></ul>	<ul style="list-style-type: none"><li>・<b>成果物の公開</b></li></ul>
<ul style="list-style-type: none"><li>■ <u>保存・再利用</u><ul style="list-style-type: none"><li>・データ移行／共有</li><li>・ドキュメンテーション</li></ul></li></ul>	<ul style="list-style-type: none"><li>・<b>データカタログのUpdate</b></li></ul>

# DMの一例：データ変形（Reshaping Data）

- データ変形 = BIツール等の分析ソフトが**利用**できるように、**データの形式を変換**すること
- 主なデータ形式
  - **ワイド（Wide）型**：横に長く、直感的に理解しやすい
    - ✓ 公開データはワイド型が多い
  - **ロング（Long）型**：縦に長く、分析ソフト等で扱いやすい
    - ✓ **BIツールを使うにはロング型のデータが必要**

ワイド型の例			
番号	性別	英語	数学
001	女性	95	80
002	男性	78	86
...	...	...	...



ロング型の例			
番号	性別	科目	点数
001	女性	英語	95
001	女性	数学	80
002	男性	英語	78
002	男性	数学	86
...	...	...	...

# 公開データ（ワイド型）の例

16 出身高校の所在地県別 入学者数（8-1）

1 計		16 出身高校の所在地県別 入学者数（8-1）																				出身高校の所在地				
出身高校の所在地		計	北海道	青森	岩手	宮城	秋田	山形	福島	茨城	栃木	群馬	埼玉	千葉	東京	神奈川	新潟	富山	石川	福井	山梨	長野	岐阜	静岡	愛知	出身高校の所在地
大学の所在地		大学の所在地																								
平成28年度		618,423	20,555	4,949	4,612	10,145	3,600	4,242	7,668	14,919	9,211	9,192	33,864	29,188	76,039	43,077	9,459	4,527	5,435	3,817	4,930	9,194	9,292	16,890	38,244	平成28年度
北海道		18,716	13,794	357	302	150	138	78	76	146	96	89	209	240	542	250	88	49	68	22	52	82	58	193	238	北海道
青森		3,309	474	1,845	280	82	203	53	34	23	21	11	29	15	33	12	22	6	1	1	3	7	5	16	11	青森
岩手		2,458	103	245	1,187	290	196	72	60	21	31	9	20	14	37	25	21	3	1	2	5	10	-	18	18	岩手
宮城		11,602	172	540	817	5,827	527	831	923	195	170	100	137	106	237	99	200	46	40	5	32	81	12	120	68	宮城
秋田		2,074	63	112	122	99	844	63	49	55	71	38	39	25	69	34	65	11	2	8	13	25	17	55	77	秋田
山形		2,824	93	89	131	557	112	856	195	72	101	30	46	39	65	26	107	10	5	2	10	53	7	72	55	山形
福島		3,253	30	60	117	215	70	172	1,451	247	205	61	49	34	71	38	114	5	7	3	15	33	10	67	28	福島
茨城		7,331	108	72	71	80	50	60	294	2,852	203	134	420	729	582	143	108	46	30	24	49	109	22	126	100	茨城
栃木		4,459	35	82	72	134	60	89	331	447	2,028	240	222	55	90	30	65	14	6	8	19	74	9	47	16	栃木
群馬		6,353	90	41	86	87	53	79	152	202	453	2,685	548	133	299	106	208	44	34	13	44	306	24	127	60	群馬
埼玉		30,803	362	182	171	309	154	212	660	1,364	1,193	1,392	10,495	1,918	6,468	952	706	122	71	45	245	649	51	510	204	埼玉
千葉		26,116	337	183	138	220	134	208	437	2,012	562	371	1,880	9,482	4,543	1,034	426	100	71	42	214	470	79	631	215	千葉
東京		149,860	2,155	659	615	1,165	522	661	1,773	5,153	2,689	2,424	17,008	13,433	49,926	21,526	2,121	653	480	270	1,606	2,212	574	3,461	2,104	東京
神奈川		48,328	631	251	238	422	186	324	656	1,155	762	785	1,926	2,124	10,569	17,430	862	230	191	93	612	991	192	2,190	673	神奈川
新潟		5,948	70	40	54	95	198	252	264	73	136	216	48	29	73	43	3,305	197	94	42	30	279	31	61	44	新潟
富山		2,387	19	8	1	9	13	13	6	20	19	26	18	12	41	17	93	800	360	105	9	144	117	44	244	富山
石川		5,895	87	14	18	14	11	35	14	33	26	92	28	18	58	32	269	823	2,353	342	27	267	199	183	246	石川
福井		2,285	16	1	-	4	1	5	1	6	4	8	6	3	13	7	38	110	144	1,100	5	31	94	37	238	福井
山梨		4,269	78	21	48	45	23	32	51	91	48	59	111	101	472	261	64	49	29	30	1,312	336	45	323	119	山梨
長野		3,496	62	17	7	14	5	9	22	56	39	71	85	64	175	88	100	46	37	19	133	1,568	91	110	231	長野
岐阜		4,608	35	4	6	4	-	7	1	9	7	6	10	8	21	14	20	84	60	55	9	105	1,752	148	1,411	岐阜
静岡		8,000	98	28	21	37	15	25	49	105	77	56	88	127	230	209	50	25	29	48	118	117	148	4,716	636	静岡
愛知		42,047	195	19	14	44	12	16	15	98	24	39	45	47	127	77	56	390	253	335	74	638	4,519	1,951	27,288	愛知
三重		3,149	10	1	2	4	1	2	3	6	7	4	8	2	15	10	2	10	12	9	7	25	98	85	669	三重
滋賀		7,295	59	2	1	20	5	3	4	28	20	13	13	11	46	31	27	61	77	114	9	47	251	130	453	滋賀

出典：学校基本調査 - 高等教育機関《報告書掲載集計》 - 大学・大学院

注意：このデータを「そのまま」BIツールで読み込んでも分析できません。



データ変形（DM）が必要

# データ変形：手動 vs. 自動化

---

## ■ データ変形を手動（例：コピー&ペースト）で行なうのは非現実的

- 前述した学校基本調査データの場合、1年分のデータ変形でさえも膨大なコピペが必要
- 単純作業なため、ミスが起こる可能性大
- エラーチェックも困難

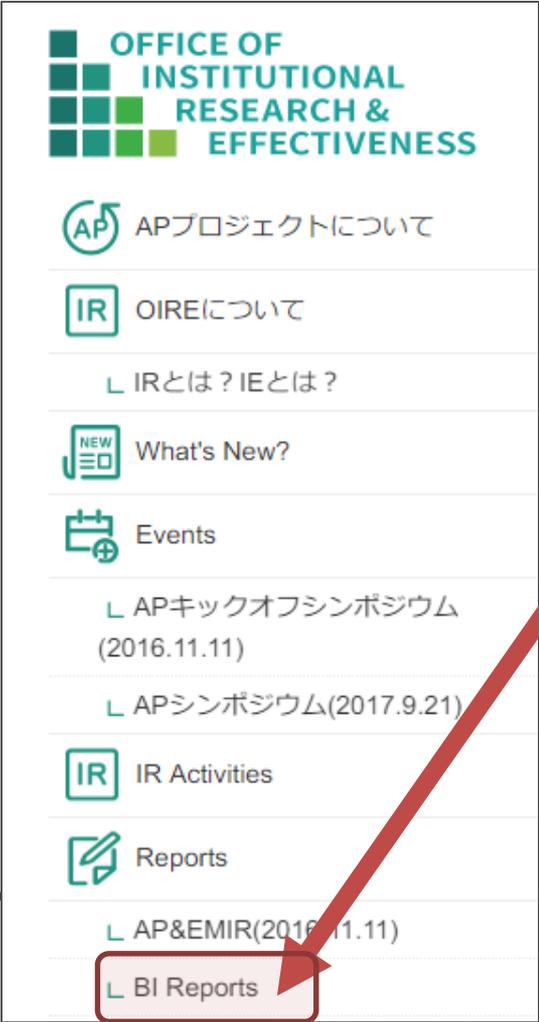
## ■ 自動化による変形作業の効率化

- Reshaping data made easy | Tableau Software : <https://www.tableau.com/about/blog/2012/03/reshaping-data-made-easy-16353>
- STATA (reshape)
- R (reshape2, etc)
- Excel VBA

# OIRE Power BI Report

- ① Google等の検索エンジンで「OIRE」と検索
  - a. [山形大学 Office of Institutional Research & Effectiveness](#)

②



OFFICE OF  
INSTITUTIONAL  
RESEARCH &  
EFFECTIVENESS

- APプロジェクトについて
- OIREについて
  - IRとは? IEとは?
- What's New?
- Events
  - APキックオフシンポジウム (2016.11.11)
  - APシンポジウム(2017.9.21)
- IR Activities
- Reports
  - AP&EMIR(2016.11.11)
  - BI Reports**

③

## 保護中: BI Reports

このコンテンツはパスワードで保護されています。閲覧するには以下に

パスワード

④

## パスワードを入力

**oireyu2016**

# まとめ

---

- 評価・IRにおけるルーチン業務を効率的に実施するには、**質の高いデータマネジメントが必要**
- 質の高いデータマネジメントは、
  - a. データの質の向上 → 分析の質の向上
  - b. 他部署への負担軽減（余計な問い合わせが減る）
  - c. 評価・IR部署における業務の効率化
    - ✓ BIツールの有効活用も必要（？）に繋がる可能性大

ありがとうございました