

「大学基本情報」のBIツール利用の可能性と課題

ー複合的統計表のデータベース化とデータ解析に向けてー

第6回 大学情報・機関調査研究会

国立情報学研究所 船守美穂

帝京大学 中鉢直宏

2017年8月19日

Today's Talk

- 問題設定
- データの形式変換
- 課題と今後の展開

問題設定

問題設定

□ 伝統的な統計表形式のデータセットを、BIツールで解析可能なように、データ形式を変換したい。

◆ 対象のデータセット

- 大学改革支援・学位授与機構(NIAD)のウェブサイトにて公開される「大学基本情報」(2012-15年度分)

◆ BIツールとは

- Business Intelligenceツールの略。
- 大量のデータの可視化や分析が容易なデータ解析ツール

BIツール利用のメリット

- データベースやデータ解析の専門的な知識がなくとも、容易に使える。
- データの可視化が容易。
- 複数のデータセットを容易に連結可能。
- ドリルダウン機能により、自身の興味関心に応じて、分析を深めていくことができる。
- データ解析機能の公開により、解析体験自体を共有できる。

BIツールで必要とされるデータ形式

データの属性情報は全て列情報としてまとめる

西暦年度	学校調査番号	大学名	番号別大学名	昼夜別	集計対象	学生所属区分	課程区分	性別	学生数
2012年度	100	北海道大学	0100北海道大学	昼間	学生数	大学院	博士課程	男	1675
2012年度	100	北海道大学	0100北海道大学	昼間	学生数	大学院	博士課程	女	741
2012年度	100	北海道大学	0100北海道大学	昼間	学生数	大学院	修士課程	男	2678
2012年度	100	北海道大学	0100北海道大学	昼間	学生数	大学院	修士課程	女	895
2012年度	100	北海道大学	0100北海道大学	昼間	学生数	大学院	専門職学位	男	245
2012年度	100	北海道大学	0100北海道大学	昼間	学生数	大学院	専門職学位	女	54
2012年度	100	北海道大学	0100北海道大学	昼間	学生数	学部・本科	学部・本科	男	8173
2012年度	100	北海道大学	0100北海道大学	昼間	学生数	学部・本科	学部・本科	女	3257
2012年度	100	北海道大学	0100北海道大学	昼間	学生数	専攻科	専攻科	男	0
2012年度	100	北海道大学	0100北海道大学	昼間	学生数	専攻科	専攻科	女	0
2012年度	100	北海道大学	0100北海道大学	昼間	学生数	別科	別科	男	0
2012年度	100	北海道大学	0100北海道大学	昼間	学生数	別科	別科	女	0
2012年度	100	北海道大学	0100北海道大学	昼間	学生数	聴講生・選科生・科目等履修生		男	191

データ見出しは、原則一行

集計列は、原則一列

※ データ見出しと属性値を、複数の表で統一することにより、複数表の連結が可能となる。

元の統計表の形式

見出しと、
データの不一致



年度	学校調査番号	大学名	学部・学科名	教員数（本務者）														計 (a+b+c+d+e+f+g)		
				学長 (a)		副学長 (b)		教授 (c)		准教授 (d)		講師 (e)		助教 (f)		助手 (g)				
				男	女	男	女	男	女	男	女	男	女	男	女	男	女	計男	計女	計計
年度(西暦)	学校調査番号	大学	D列	学長_男	学長_女	副学長_男	副学長_女	教授_男	教授_女	准教授_男	准教授_女	講師_男	講師_女	助教_男	助教_女	助手_男	助手_女	計男	計女	計計
2012	0100	北海道大学	大学院担当者					704	25	564	69	121	17	530	110	6	10	1,925	231	2,156
2012	0100	北海道大学	休職者																1	1
2012	0100	北海道大学	外国人					11	1	19	10	1	2	22	17			53	30	83
2012	0104	北海道教育大学	大学院担当者					164	25	117	29	12	6					293	60	353
2012	0104	北海道教育大学	休職者					1										1		1
2012	0104	北海道教育大学	外国人					1		1	4		2					2	6	8
2012	0108	室蘭工業大学	大学院担当者					76	1	57	3	13		29	1			175	5	180
2012	0108	室蘭工業大学	休職者																	
2012	0108	室蘭工業大学	外国人					1		3	1	1						5	1	6
2012	0112	小樽商科大学	大学院担当者					35	1	16	1							51	2	53
2012	0112	小樽商科大学	休職者							1								1		1
2012	0112	小樽商科大学	外国人					5	2	3				1				9	2	11
2012	0116	帯広畜産大学	大学院担当者					53	1	31	4	7	1	15	3			106	9	115
2012	0116	帯広畜産大学	休職者																	
2012	0116	帯広畜産大学	外国人					1		1		2		1				5		5
2012	0120	北見工業大学	大学院担当者					46	1	59	3	2		11	1			118	5	123
2012	0120	北見工業大学	休職者					1										1		1
2012	0120	北見工業大学	外国人							1	1	1	1		1			2	3	5
2012	0124	弘前大学	大学院担当者					202	22	160	25	81	20	99	15	10	15	552	97	649

データ見出し
が複数行

簡略表現によ
る一行見出し

相互に排他的で
はない(内数の)
属性データ

集計列が複数

総計列も複数

元の統計表の全体像(1)

大学基本情報 2012 (H24)

学生教職員等	(7-A) 学生数 [大学院、学部、専攻科、別科、科目等履修生・聴講生・研究生]	DOWNLOAD
	(7-B) 教員数(本務者) [学長、副学長、教授、准教授、講師、助教、助手]	DOWNLOAD
	(7-1) 教員数(本務者)(再掲) [大学院担当者、退職者、外国人]	DOWNLOAD
	(7-Z) 教員数(兼務者) [学長、副学長、学長・副学長以外の教員 外国人]	DOWNLOAD
	(7-C) 職員数 [事務系、記述技能系、医療系、教務系、その他 医療系のうち看護師]	DOWNLOAD
学部学生内訳	(8-D) 学科別学生数 [1年次～6年次 入学者数]	DOWNLOAD
	(8-2) 学科別学生数のうち休学者数	DOWNLOAD
	(8-3) 学科別学生数のうち最低在学年限超過学生数(編入学者は除く。)	DOWNLOAD
	(8-E) 入学状況 [高等学校卒業年度]	DOWNLOAD
	(8-G) 出身高校の所在地県別入学者数	DOWNLOAD
	(8-7) 専攻科・別科及び聴講生等の学生数	DOWNLOAD
	(8-R) 短期大学・高等専門学校・専修学校(専門課程)からの編入学者数 [2年次編入学、3年次編入学]	DOWNLOAD

← 年度ごとの集計

ファイル名は、
 ・07go_A.xls
 ・07go_B.xls
 ・07go_1.xls 等



不自然で飛びのあるナンバリング

元の統計表の全体像(2)

大学院学生内訳	(9-H) 専攻別学生数 [1年次～5年次 社会人]	DOWNLOAD
	(9-4) 専攻別学生数のうち休学者数 [1年次～5年次]	DOWNLOAD
	(9-5) 専攻別学生数のうち最低在学年限超過学生数(編入学者は除く)	DOWNLOAD
	(9-I) 入学状況 [入学志願者数(当該大学、他大学、外国) 入学者数(当該大学、他大学、外国)]	DOWNLOAD
	(9-5) 年齢別入学者数 [21歳以下 22歳 23歳 24歳 25歳 26歳 27歳 28歳 29歳 30-34歳 35-39歳 45-49歳 50-54歳 55-60歳 61歳以上]	DOWNLOAD
	(9-8) 科目等履修生等の学生数	DOWNLOAD
本科学生内訳	(10-J) 学科別学生数、入学状況(本科) [1年次、2年次、3年次 入学志願者数、入学者数]	DOWNLOAD
	(10-6) 学科別学生数のうち休学者数 [1年次、2年次、3年次]	DOWNLOAD
	(10-K) 出身高校の所在地県別入学者数 [北海道 ～ 沖縄]	DOWNLOAD
	(10-9) 専攻科、別科及び科目等履修生等の学生数	DOWNLOAD
外国人学生	(11) 国費留学生、私費留学生、留学生以外の外国人学生 [人文科学 社会科学 理学 工学 農学 保健 商船 家政 教育 芸術 その他]	DOWNLOAD
	(11 別掲) 国費留学生、私費留学生、留学生以外の外国人学生(専攻科・別科の学生、科目等履修生・聴講生・研究生 [人文科学 社会科学 理学 工学 農学 保健 商船 家政 教育 芸術 その他]	DOWNLOAD

元の統計表の全体像(3)

学校施設	<p>(20) 学校施設</p> <p>[学校土地の用途別面積 学校建物の用途別面積、構造別面積、新築等増加の面積、被害等減少の面積 職員宿舍の用途別土地面積、建物面積]</p>	DOWNLOAD
卒業後の状況	<p>(30-1) 状況別卒業者数、入学年度別卒業者数</p> <p>[大学院研究科、大学学部、短期大学本科、専攻科、別科、就職者、臨床研修医、専修学校・外国の学校等入学者、一時的な仕事に就いた者、左記以外、不詳・死亡の者 大学院博士課程のうち再掲 各入学年度、その他(編入)]</p>	DOWNLOAD
	<p>(30-2) 職業別就職者数、産業別就職者数</p> <p>[専門的・技術的職業従事者、管理的職業従事者、事務従事者、販売従事者、サービス職業従事者、保安職業従事者、農林産業従事者、生産工程従事者、輸送・機械運転従事者、建設・採掘従事者、運搬・清掃・包装等従事者、左記以外]</p> <p>[農業・林業・漁業、鉱業・採石業・砂利採取業、建設業、製造業、電気・ガス・熱供給・水道業、情報通信業、運輸業・郵便業、卸売業、小売業、金融業、保険業、不動産業・賃貸・管理業、物品賃貸業、学術研究、専門・技術サービス業、宿泊業、飲食サービス業、生活関連サービス業、娯楽業、教育、学習支援業、医療、福祉、複合サービス事業、サービス事業、公務、左記以外]</p>	DOWNLOAD

元の統計表の全体像...年度横断(1)

(ファイル名)	2012	2013	2014	2015	2016	リシェイプデータ
(7-A) 学生数	○	○	○	○	○	07go1_学生数
(7-B) 教員数 (本務者)	○	○	○	○	○	07go2_教員数 (本務者) 組織区分別
(7-1) 教員数 (本務者) (再掲)	○	○	○	○	○	07go3_教員数 (本務者) 教員状態別
(7-Z) 教員数 (兼務者)	○	○	○	○	○	07go4_1_教員数 (兼務者) 職位別
						07go4_2_外国人教員数 (兼務者)
(7-C) 職員数	○	○	○	○	○	07go5_1_職系別職員数
						07go5_2_医療系職員数
(8-D) 学科別学生数	○	○	○	○	○	08go1_1_学科別学部生数
	○	○	○	○	○	08go1_2_学科別学部入学者数
	(8-E) 入学状況から抽出 (学科情報なし)			未公開	○	08go1_3_学科別学部入学志願者数
(8-2) 学科別学生数のうち休学者数	○	○	○	○	○	08go2_学部別休学者数
(8-3) 学科別学生数のうち最低在学年限超過学生数 (編入学者は除く。)	○	○	○	○	○	08go3_学部別最低在学年限超過学生数
(8-E) 入学状況	○	○	○	未公開	×	08go4_学部入学状況 (志願者数、入学者数)
(8-O) 年齢別入学者数	×	×	×	×	○	08go4_1_年齢別学部入学者数
	(8-E) 入学状況から抽出			未公開	○	08go4_2_出身高校種類別学部入学者数
	×	×	×	×	○	08go4_3_学部留学生入学者数
(8-G) 出身高校の所在地県別入学者数	○	○	○	○	○	08go5_出身高校の所在地県別学部入学者数
(8-7) 専攻科・別科及び聴講生等の学生数	○	○	○	○	○	08go6_専攻科、別科、学部科目等履修生等の学生数
(8-R) 短期大学・高等専門学校・専修学校 (専門課程) からの編入学者数	○	○	○	○	○	08go7_学部編入学者数

年度により異なる集計

分かりやすくした、
データ加工後のファイル名

元の統計表の全体像...年度横断(2)

(ファイル名)	2012	2013	2014	2015	2016	リシェイプデータ
(9-H) 専攻別学生数	○	○	○	○	○	09go1_1_専攻別院生数
						09go1_2_専攻別社会人院生数
(9-4) 専攻別学生数のうち休学者数	○	○	○	○	○	09go2_研究科別休学院生数
(9-5) 専攻別学生数のうち最低在学年限超過学生数 (編入学者は除く)	○	○	○	○	○	09go3_研究科別最低在学年限超過院生数
(9-I) 入学状況	○	○	○	○	○	09go4_1_専攻別大学院入学者数
						09go4_2_専攻別大学院入学志願者数
(9-S) 年齢別入学者数	○	○	○	○	○	09go5_1_年齢別大学院入学者数
						09go5_2_大学院社会人入学者数
						09go5_3_大学院留学生入学者数
(9-8) 科目等履修生等の学生数	○	○	○	○	○	09go6_大学院科目等履修生等の学生数
(10-J) 学科別学生数、入学状況 (本科)	○	○	○	未公開	○	10go1_1_学科別本科学生数
						10go1_2_学科別本科入学者数
						10go1_3_学科別本科入学志願者数
(10-6) 学科別学生数のうち休学者数	○	○	○	未公開	○	10go2_本科休学者数
(10-K) 出身高校の所在地県別入学者数	○	○	○	未公開	○	10go3_出身高校の所在地県別本科入学者数
(10-9) 専攻科、別科及び科目等履修生等の学生数	○	○	○	未公開	○	10go4_専攻科、別科、本科科目等履修生等の学生数
(10-Q) 年齢別入学者数(再掲)	×	×	×	×	○	10go5_1_年齢別本科入学者数
	(10-J) 学科別入学状況 から抽出			未公開	○	10go5_2_出身高校種類別本科入学者数
	×	×	×	×	○	10go5_3_本科留学生入学者数
(11) 国費留学生、私費留学生、留学生以外の外国人学生	○	○	○	○	○	11go1_外国人学生数
(11 別掲) 国費留学生、私費留学生、留学生以外の外国人学生 (専攻科・別科の学生、科目等履修生・聴講生・研究生	○	○	○	○	○	11go2_外国人学生数 (専攻科、別科、科目等履修生等)

元の統計表の全体像...年度横断(3)

(ファイル名)	2012	2013	2014	2015	2016	リシェイプデータ
(20) 学校施設	○	○	○	○	○	20go_1_学校土地の用途別面積 (㎡)
						20go_2_学校建物の用途別延面積 (㎡)
						20go_3_厚生補導施設の延面積 (㎡)
						20go_4_学校建物の構造別延面積 (㎡)
						20go_5_学校建物増減面積 (㎡)
						20go_6_職員宿舍土地の用途別面積 (㎡)
						20go_7_職員宿舍建物の用途別面積 (㎡)
(30-1) 状況別卒業者数、入学年度別卒業者数	○	○	○	○	○	30go1_1_状況別卒業者数
						30go1_2_就職状況別卒業者数
						30go1_3_博士課程満期退学者数
						30go1_4_ポストドクター数
						30go1_5_入学年度別卒業者数
(30-2) 職業別就職者数、産業別就職者数	○	○	○	○	○	30go2_1_職業別就職者数
						30go2_2_産業別就職者数

➤ 年度あたり概ね27の表を、54の表に変換

元の統計表の課題の整理

1. データセットの年度別欠落と集計方法の変化
2. データセット間の列見出しとデータの不統一
3. 1つの表における複数の集計値の横並び
4. 複数行にわたるデータ見出し
5. 不規則かつナンバリングのみのファイル名
6. 複合的または細かすぎるコード

データの形式変換

データ形式変換...対応内容(全体像)

A. BIツールで利用可能とするためのデータ加工

- A-1. データセットの年度別欠落や集計方法の変化への対処
- A-2. データセット間の用語(列見出し)の不統一への対応
- A-3. 複数の異なる集計値を含む単表の分割
- A-4. 複合見出しの単行化

B. BIツールの利用しやすさのためのデータ加工

- B-1. 分かりやすく、規則性あるファイル名の付与
- B-2. コードの分割と内容情報の付加
- B-3 小分類コードへの、大中分類の付加
- B-4. 合計行および無回答行の削除

A-1. データセットの年度別欠落や集計方法の変化への対処

(対応内容)

- 年度別欠落や集計方法の変化については、最新のデータセット(2016年度データ)に合わせて整形。
 - 2016年度データに合わせて、計算し直しや該当のセルのレイアウト変更。
- 対応不能の場合は、「データ(ファイル)欠如」。

(今後の課題)

- ◆ 今後追加される年度別データの集計方法や公開されるデータセットが異なる場合、作成した全データセットの作り替え、あるいは2016年度の基準に合わせたデータ加工が必要となる。
- ⇒ 手間が煩雑な上、フォーマット統一化不能なデータセットが出現する可能性あり。

A-2. データセット間の用語(列見出し)の不統一への対応

(対応内容)

- 複数表が連結可能なように、列見出しの統一を図った。
 - 元表では、個々の表で自己完結的に見出しが、特に簡潔性に留意して、ふられている。
- 「学科・専攻」の表記揺れには対応せずに、「学科・専攻コード」で利用されることを想定。

(今後の課題)

- ◆ 「学科・専攻」別の分析については、(学科・専攻名が大学ごとに自由に名称設定されていることもあり)、「学科・専攻コード」の利用で問題ないが、「学科・専攻コード」が細かく振られすぎていて、使いづらい可能性がある。

⇒ 自身の分析目的に応じた「学科・専攻」の分類が必要になる可能性あり。

A-3. 複数の異なる集計値を含む単表の分割

(対応内容)

➤ 異なる複数の集計値を含む表を分割。

- たとえば「入学者数」「入学志願者数」は別の表にし、それぞれのファイル名も適切に付与。

年度	学校調査番号	研究科番号	所在地	昼夜別	課程別	大学名	研究科名	専攻	符号	入学状況（入学志願者数、入学者数）																											
										入学志願者数														入学者数													
										当該大学 出身者	他大学出身者						外国の 学校卒		その他		計		当該大学 出身者	他大学出身者						外国の 学校卒		その他		計			
											国立		公立		私立									国立		公立		私立									
男	女	男	女	男	女	男	女	男	女	男	女	男	女	男	女	男	女	男	女	男	女	男	女	男	女	男	女	男	女	男	女						
西暦年学校調										志願	志願	志願	志願	志願	志願	志願	志願	志願	志願	志願	志願	志願	志願	志願	志願	志願	志願	志願	志願	志願	志願	志願	志願	志願			
2012	0100	1M01	57	1	1	北海道大学	医学研究科	医科学	M137	5	6	1	3		2	3	3	1	1		12	13	4	2	1	1			3	3	1	1	9	7			
2012	0100	2A01	57	1	2	北海道大学	文学研究科	思想文化学	A346	6	8	1	4		7	1	4	4			18	17	5	4		3		1		2	3		8	10			
2012	0100	2A01	57	1	2	北海道大学	文学研究科	歴史地域学	B991	11	6	3	3	1	13	7	5	7			33	23	7	4	2	2		5	2	2	3		16	11			
2012	0100	2A01	57	1	2	北海道大学	文学研究科	言語文学	A1B3	6	6	4	3		7	10	7	11			24	30	5	5	2			3	6	2	5		12	16			
2012	0100	2A01	57	1	2	北海道大学	文学研究科	人間シスラ	X345	6	6	4	1	1	10	5	7	21			28	33	4	5	3	1		1	2	3	15		11	23			
2012	0100	2C01	57	1	2	北海道大学	法学研究科	法学（・）	C140	11	3	2			14	5	9	3			36	11	3	1	1			3	1	6	2		13	4			
2012	0100	2C09	57	1	2	北海道大学	経済学研究科	現代経済学	C288	22	1	3			10	5	12	13			47	19	10	1	1			5	1	7	8		23	10			
2012	0100	2E28	57	1	2	北海道大学	理学院	数学	E101	27	1	6		4	18	3	1				56	4	22	1	4		1	10	1	1			38	2			
2012	0100	2E28	57	1	2	北海道大学	理学院	化学	E301																												
2012	0100	2E28	57	1	2	北海道大学	理学院	物性物理学	E219	23	3	1	1		2						26	4	17	1				2					19	1			
2012	0100	2E28	57	1	2	北海道大学	理学院	宇宙物理学	E210	18	5	3	1		2						23	6	12	5	1			1					14	5			
2012	0100	2E28	57	1	2	北海道大学	理学院	自然史科学	F9B7	44	11	13	2		2	1			1	2	60	16	32	9	3	1					1	1	36	11			
2012	0100	2E31	57	1	2	北海道大学	総合化学院	総合化学	F9D1	122	37	12	1	1	1	4		1		3	143	39	112	30	3		1	1	1		1	3	121	31			

A-4. 複合見出しの単行化

(対応内容)

- 複数行にまたがる見出しは、①属性情報を列に写し、②集計列が一行となるように変換。
- 同時に、「集計対象」の列を設け、集計内容が明確になるように配慮。

西暦年度	学校調査	学部学	所在地	所在地	八地方区	昼夜別	課程別	大学名	番号別	大学学部・研究科	学科・専攻	学科・専攻	学校種別	集計対象	出身大学種別	性別	専攻別	大学
2015年度	1136	2 D07	10群馬県	群馬県	3関東地方	昼間	博士課程(前期)	高崎経済大学	1136高崎	経済・経営研究科	(現代) 経済社会シ	C 2 9 3	大学院	入学者数	他大学出身者 (公立)	女		0
2015年度	1136	2 D07	10群馬県	群馬県	3関東地方	昼間	博士課程(前期)	高崎経済大学	1136高崎	経済・経営研究科	(現代) 経済社会シ	C 2 9 3	大学院	入学者数	他大学出身者 (私立)	男		1
2015年度	1136	2 D07	10群馬県	群馬県	3関東地方	昼間	博士課程(前期)	高崎経済大学	1136高崎	経済・経営研究科	(現代) 経済社会シ	C 2 9 3	大学院	入学者数	他大学出身者 (私立)	女		0
2015年度	1136	2 D07	10群馬県	群馬県	3関東地方	昼間	博士課程(前期)	高崎経済大学	1136高崎	経済・経営研究科	(現代) 経済社会シ	C 2 9 3	大学院	入学者数	外国の学校卒	男		0
2015年度	1136	2 D07	10群馬県	群馬県	3関東地方	昼間	博士課程(前期)	高崎経済大学	1136高崎	経済・経営研究科	(現代) 経済社会シ	C 2 9 3	大学院	入学者数	外国の学校卒	女		0
2015年度	1136	2 D07	10群馬県	群馬県	3関東地方	昼間	博士課程(前期)	高崎経済大学	1136高崎	経済・経営研究科	(現代) 経済社会シ	C 2 9 3	大学院	入学者数	その他	男		0
2015年度	1136	2 D07	10群馬県	群馬県	3関東地方	昼間	博士課程(前期)	高崎経済大学	1136高崎	経済・経営研究科	(現代) 経済社会シ	C 2 9 3	大学院	入学者数	その他	女		0
2016年度	1136	2 D07	10群馬県	群馬県	3関東地方	昼間	博士課程(前期)	高崎経済大学	1136高崎	経済・経営研究科	(現代) 経済社会シ	C 2 9 3	大学院	入学者数	当該大学出身者	男		3
2012年度	1288	1 C09	35山口県	山口県	6中国地方	昼間	修士課程	下関市立大学	1288下関	経済学研究科	(現代) 経済社会シ	C 2 9 3	大学院	入学者数	当該大学出身者	男		0
2012年度	1288	1 C09	35山口県	山口県	6中国地方	昼間	修士課程	下関市立大学	1288下関	経済学研究科	(現代) 経済社会シ	C 2 9 3	大学院	入学者数	当該大学出身者	女		0
2012年度	1288	1 C09	35山口県	山口県	6中国地方	昼間	修士課程	下関市立大学	1288下関	経済学研究科	(現代) 経済社会シ	C 2 9 3	大学院	入学者数	他大学出身者 (国立)	男		0
2012年度	1288	1 C09	35山口県	山口県	6中国地方	昼間	修士課程	下関市立大学	1288下関	経済学研究科	(現代) 経済社会シ	C 2 9 3	大学院	入学者数	他大学出身者 (国立)	女		0
2012年度	1288	1 C09	35山口県	山口県	6中国地方	昼間	修士課程	下関市立大学	1288下関	経済学研究科	(現代) 経済社会シ	C 2 9 3	大学院	入学者数	他大学出身者 (公立)	男		0
2012年度	1288	1 C09	35山口県	山口県	6中国地方	昼間	修士課程	下関市立大学	1288下関	経済学研究科	(現代) 経済社会シ	C 2 9 3	大学院	入学者数	他大学出身者 (公立)	女		0
2012年度	1288	1 C09	35山口県	山口県	6中国地方	昼間	修士課程	下関市立大学	1288下関	経済学研究科	(現代) 経済社会シ	C 2 9 3	大学院	入学者数	他大学出身者 (私立)	男		0
2012年度	1288	1 C09	35山口県	山口県	6中国地方	昼間	修士課程	下関市立大学	1288下関	経済学研究科	(現代) 経済社会シ	C 2 9 3	大学院	入学者数	他大学出身者 (私立)	女		0
2012年度	1288	1 C09	35山口県	山口県	6中国地方	昼間	修士課程	下関市立大学	1288下関	経済学研究科	(現代) 経済社会シ	C 2 9 3	大学院	入学者数	外国の学校卒	男		0
2012年度	1288	1 C09	35山口県	山口県	6中国地方	昼間	修士課程	下関市立大学	1288下関	経済学研究科	(現代) 経済社会シ	C 2 9 3	大学院	入学者数	外国の学校卒	女		0
2012年度	1288	1 C09	35山口県	山口県	6中国地方	昼間	修士課程	下関市立大学	1288下関	経済学研究科	(現代) 経済社会シ	C 2 9 3	大学院	入学者数	その他	男		0

見出しの変換作業(1)...対応すべき内容の確認

(7-A) 学生数

original

年度	学校調査番号	学校名	昼夜	学生数																		
				大学院						学部・本科		専攻科		別科		聴講生・選科生・研究生等				計		
				博士課程 (博士後期+博士一貫)		修士課程 (修士+博士前期)		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
年度(西暦)	学校調査番号	大学	昼夜別	博士_男	博士_女	修士_男	修士_女	専門_男	専門_女	本科_男	本科_女	専攻_男	専攻_女	別科_男	別科_女	学卒上_男	学卒上_女	以外_男	以外_女	計_男	計_女	計_計

formatted

07go1_学生数

年度	学校調査番号	学校名	学校調査番号&大学名	昼夜	学生数															
					大学院						学部・本科		専攻科		別科		聴講生・選科生・研究生等			
					博士課程（博士後期+博士一貫）		修士課程（修士+博士前期）		専門職学位課程								科目等履修生等（学部卒以上）		科目等履修生等（学部卒未満）	
					男	女	男	女	男	女							男	女	男	女
西暦年度	学校調査番号	大学名	番号別大学	昼夜別	博士_男	博士_女	修士_男	修士_女	専門_男	専門_女	本科_男	本科_女	専攻_男	専攻_女	別科_男	別科_女	学卒上_男	学卒上_女	以外_男	以外_女

(要変換)

(7-B) 教員数 (本務者)

original

年度	学校調査番号	大学名	学部・学科名	符号	教員数（本務者）																	
					学長 (a)		副学長 (b)		教授 (c)		准教授 (d)		講師 (e)		助教 (f)		助手 (g)		計 (a+b+c+d+e+f+g)			
					男	女	男	女	男	女	男	女	男	女	男	女	男	女	男	女	男	女
年度(西暦)	学校調査番号	大学	D列	学部名	符号	学長_男	学長_女	副学長_男	副学長_女	教授_男	教授_女	准教授_男	准教授_女	講師_男	講師_女	助教_男	助教_女	助手_男	助手_女	計_男	計_女	計_計

formatted

07go2_教員数 (本務者)

年度	学校調査番号	大学名	学校調査番号&大学名	学部・学科名	符号1	符号2	本務・兼務別	教員数													
								学長		副学長		教授		准教授		講師		助教		助手	
								男	女	男	女	男	女	男	女	男	女	男	女	男	女
西暦年度	学校調査番号	大学名	番号別大学:組織区分	学部・研究科学部・研究科学部・研究科本務・兼務別				学長_男	学長_女	副学長_男	副学長_女	教授_男	教授_女	准教授_男	准教授_女	講師_男	講師_女	助教_男	助教_女	助手_男	助手_女

(要変換)

(算出) L (算出) R 本務者

2016年データ

(算出) "": (算出) 学部・研究科または""(学部・研究科="学長・副学長","教養部(一般教養)","大学院","付属研究所","付属病院","その他","計"の場合)

(7-1) 教員数 (本務者) (再掲)

original

年度	学校調査番号	大学名	学部・学科名	教員数（本務者）																
				学長 (a)		副学長 (b)		教授 (c)		准教授 (d)		講師 (e)		助教 (f)		助手 (g)		計 (a+b+c+d+e+f+g)		
				男	女	男	女	男	女	男	女	男	女	男	女	男	女	男	女	計
年度(西暦)	学校調査番号	大学	D列	学長_男	学長_女	副学長_男	副学長_女	教授_男	教授_女	准教授_男	准教授_女	講師_男	講師_女	助教_男	助教_女	助手_男	助手_女	計_男	計_女	計_計

formatted

07go3_教員数 (本務者)

年度	学校調査番号	大学名	学校調査番号&大学名	学部・学科名	本務・兼務別	教員数													
						学長		副学長		教授		准教授		講師		助教		助手	
						男	女	男	女	男	女	男	女	男	女	男	女	男	女
西暦年度	学校調査番号	大学名	番号別大学:教員状態	本務・兼務別		学長_男	学長_女	副学長_男	副学長_女	教授_男	教授_女	准教授_男	准教授_女	講師_男	講師_女	助教_男	助教_女	助手_男	助手_女

(要変換)

本務者

見出しの変換作業(2)...変換内容

ファイル名	07go1_学生数											
修正前	西暦年度	学校調査番号	大学名	番号別大学名	昼夜別	Col6	Col7	Col8	Col9	Col10	Col11	
修正後	西暦年度	学校調査番号	大学名	番号別大学名	昼夜別	集計対象	学生所属区分	課程区分	性別	非表示ラベル	学生数	
ファイル名	07go2_教員数（本務者）	組織区分別										
修正前	西暦年度	学校調査番号	大学名	番号別大学名	組織区分	学部・研究科	学部・研究科番号	学部・研究科番号（下本務・兼務別	Col10	Col11	Col12	
修正後	西暦年度	学校調査番号	大学名	番号別大学名	組織区分	学部・研究科	学部・研究科番号	学部・研究科番号（下本務・兼務別	集計対象	教員職位	性別	
ファイル名	07go3_教員数（本務者）	教員状態別										
修正前	西暦年度	学校調査番号	大学名	番号別大学名	教員状態	本務・兼務別	Col7	Col8	Col9	Col10	Col11	
修正後	西暦年度	学校調査番号	大学名	番号別大学名	教員状態	本務・兼務別	集計対象	教員職位	性別	非表示ラベル	教員数（本務者）	教員状態別
ファイル名	07go4_1_教員数（兼務者）	職位別										
修正前	西暦年度	学校調査番号	大学名	番号別大学名	本務・兼務別	Col6	Col7	Col8	Col9	Col10	Col11	
修正後	西暦年度	学校調査番号	大学名	番号別大学名	本務・兼務別	集計対象	教員職位	兼務分類	性別	非表示ラベル	教員数（兼務者）	職位別
ファイル名	07go4_2_外国人教員数（兼務者）											
修正前	西暦年度	学校調査番号	大学名	番号別大学名	本務・兼務別	Col6	Col7	Col8	Col9	Col10		
修正後	西暦年度	学校調査番号	大学名	番号別大学名	本務・兼務別	集計対象	教員状態	性別	非表示ラベル	外国人教員数（兼務者）		
ファイル名	07go5_1_職系別職員数											
修正前	西暦年度	学校調査番号	大学名	番号別大学名	本務・兼務別	Col6	Col7	Col8	Col9	Col10		
修正後	西暦年度	学校調査番号	大学名	番号別大学名	本務・兼務別	集計対象	職員職系	性別	非表示ラベル	職系別職員数		
ファイル名	07go5_2_医療系職員数											
修正前	西暦年度	学校調査番号	大学名	番号別大学名	本務・兼務別	職員職系	Col7	Col8	Col9	Col10	Col11	
修正後	西暦年度	学校調査番号	大学名	番号別大学名	本務・兼務別	職員職系	集計対象	看護師該非	看護師業務	非表示ラベル	医療系職員数	

見出しの変換作業(3)...見出しの統一状況

見出し（統一後）	見出し（original）	凡例セット	07go1_学	07go2_教	07go3_教	07go4_1_学	07go4_2_学	07go5_1_学	07go5_2_学	08go1_1_学	08go1_2_学	08go1_3_学
大学名	大学/大学名/短期大学名/学校名	(各大学・大学院・短期大学の名称)	○	○	○	○	○	○	○	○	○	○
学校種別	学校種別	大学/大学院/短期大学（/高等専門学校） *「凡例」シート参照									○	○
学校・課程種別	学校(課程)種別	大学/短期大学/修士課程/博士課程/専門職学位課程（/高等専門学校） *「凡例」シート参照										
学部・研究科	学部・学科名/学部名/研究科名/学部・研究科名	(各学部・研究科の名称)		○						○	○	○
学部・研究科番号	学部番号/研究科番号/学部(研究科)番号	4桁数値（大学・大学院は2桁目がアルファベット）		○						○	○	○
学科・専攻	学科/専攻/学科・専攻名	(各学科・専攻の名称)								○	○	○
学科・専攻コード	符号									○	○	○
学生所属区分	なし/国籍名	大学院/学部・本科/専攻科/別科/聴講生・選科生・研究生等	○									
課程区分	なし	博士課程(博士後期+博士一貫)/修士課程(修士+博士前期)/学部・本科/専攻科/別科/科目等履修生等（学部卒以上）/科目等履修生等（学部卒未満）	○									
課程別	課程別	修士課程/博士課程(前期)/博士課程(後期)/博士課程(一貫)/専門職学位課程 *「凡例」シート参照										
本務・兼務別	なし/区分	本務者/兼務者		○	○	○	○	○	○			

B-1. 分かりやすく、規則性あるファイル名の付与

(対応内容)

- 人間に理解しやすく、かつ規則性のあるファイル名を付与。
- ① 元のファイル名のナンバリングは、初めの一桁については踏襲。
 - ② その後のアルファベットについては、順番に並ぶように、ナンバリング。
 - ③ 分割した表については、枝番を付与。
 - ④ 集計値の意味を示すファイル名を付与。

ファイル内容	元ファイル名	新ファイル名	備考
(7-A) 学生数	07go_A	07go1_学生数	
(7-B) 教員数（本務者）	07go_B	07go2_教員数（本務者）組織区分別	* 組織区分が「計」の行を削除 * 学部名が学部・学科名の場合は組織区分に「学部（科）」を、学部名が学部・学科名以外の場合はその内容を組織区分に移動（2016年） * 集計値が「／」の場合は空欄に置換（2016年）
(7-1) 教員数（本務者） （再掲）	07go_1	07go3_教員数（本務者）教員状態別	教員状態：大学院担当者/休職者/外国人
(7-Z) 教員数（兼務者）	07go_Z	07go4_1_教員数（兼務者）職位別	
		07go4_2_外国人教員数（兼務者）	教員状態：外国人/外国人以外
(7-C) 職員数 要分割	07go_C	07go5_1_職系別職員数	
		07go5_2_医療系職員数	看護師該非：「看護師」/「看護師以外」 看護師業務：「学生の健康管理」/「付属病院」

B-2. コードの分割と内容情報の付加

(対応内容)

- 「学部・研究科番号」など、一桁目が課程、下3桁が分野を擬似的に意味すると把握されたものについては、コードを2つに分けた。
- コード番号とともに、その意味を示す列も挿入した。

年度	学校調査番号	研究科番号	所在地	昼夜別	課程別	大学名	研究科名	専攻	符号
2012	0100	6M01	57	1	4	北海道大学	医学研究科	脳科学	M176
2012	0100	6M01	57	1	4	北海道大学	医学研究科	病態制御学	M194
2012	0100	2Y51	57	1	2	北海道大学	環境科学院	環境起学	X182
2012	0100	4Y51	57	1	3	北海道大学	環境科学院	環境起学	X182
2012	0100	2Y51	57	1	2	北海道大学	環境科学院	環境物質科学	F9A8
2012	0100	4Y51	57	1	3	北海道大学	環境科学院	環境物質科学	F9A8
2012	0100	2Y51	57	1	2	北海道大学	環境科学院	生物圏科学	X183
2012	0100	4Y51	57	1	3	北海道大学	環境科学院	生物圏科学	X183



上一桁と下三桁に分割

(参考)学部・研究科番号の一桁目の意味

学部番号1桁目

- | | |
|---|------------|
| 1 | 昼間（修業年限4年） |
| 2 | 夜間（修業年限4年） |
| 3 | 夜間（修業年限5年） |
| 4 | 昼間（修業年限6年） |

学部(研究科)番号 短期大学の場合（30号のみ）

- | | |
|------|-----------|
| 1002 | 昼間・修業年限2年 |
| 1003 | 昼間・修業年限3年 |
| 2002 | 夜間・修業年限2年 |
| 2003 | 夜間・修業年限3年 |

研究科番号1桁目

- | | |
|---|------------------------------|
| 1 | 修士課程 |
| 2 | 博士課程（前期） |
| 3 | 医歯薬学，獣医学関係以外の博士一貫制の課程（1，2年次） |
| 4 | 博士課程（後期） |
| 5 | 医歯薬学，獣医学関係以外の博士一貫制の課程（3～5年次） |
| 6 | 医歯薬学，獣医学関係の博士一貫制の課程 |
| 7 | 夜間の博士課程（前期） |
| 8 | 夜間の博士課程（後期） |
| 9 | 夜間の修士課程 |
| A | 専門職学位課程 |
| B | 夜間の専門職学位課程 |
| C | 専門職学位課程法科大学院 |
| D | 夜間の専門職学位課程法科大学院 |
| G | 専門職学位課程教職大学院 |
| H | 夜間の専門職学位課程教職大学院 |

(参考) 文部科学省「平成28年度 学校基本調査の手引(大学, 短期大学, 高等専門学校)」
(http://www.mext.go.jp/b_menu/toukei/chousa01/kihon/sonota/1355787.htm)

(参考)学科系統分類表

2 学科系統分類表

1 大学(学部)

大分類	中分類	小 分 類 (学 科)					
人文科学	文 学 関 係	文学	日本文学	国語国文学	国語学国文学	文芸学	言語学
		語学文学	英語英文学教育	国文中国文学	国文学国語学	国文学	語学
		特設日本語学	日本語(・)日本文化学	外国語学(2参照)	外国語・外国文学(2参照)	外国文学(2参照)	比較文化
		文化史学	児童文学	日本語・日本文化学類	日本語学	文	日本学
		日本語教育学	日本語(・)日本文学	英語(・英米)文化学	外国語文化学	日本アジア言語文化学	米英言語文化学
		国際文化(学)	英語英米文化(学)	中国語中国文化学	ヨーロッパ・アメリカ学	表現文化	実践英語
		言語文化学	英語英米文学	東アジア言語文化学	日本文化学	社会文化	国際コミュニケーション教育
		ドイツ語圏文化学	フランス語圏文化学	グローバル英語学	英語教育学	日本文学(・)文化学(類)	英語・英文学類
		英語文化コミュニケーション学	日本・中国文学科	国際言語学	国際コミュニケーション学	ヨーロッパ文化学	日本・東アジア文化学
		英語キャリア学	日本語・日本語教育学	アジア言語学	イベロアメリカ言語学	中国言語文化学	日本文化コミュニケーション学
		グローバル地域文化学	メディア表現学	ヨーロッパ言語学	英語国際学	外国学	
	史 学 関 係	史学	国史学	東洋史学	西洋史学	史学地理学	地理歴史学
		地理学	美術史学	歴史学	文化財学	歴史社会学	日本文化史学
		歴史文化(文化歴史)学	歴史遺産学	総合歴史	歴史地理学	文化財歴史学	日本史学
		史学・文化財学					
	哲 学 関 係	哲学	中国哲学	中国哲学文学	心理学(類)	美学美術史学	宗教学
		宗学	仏教学	基督(キリスト)教学	神学	神道学	禅学
		密教学	真宗学	美学	東洋哲学	西洋哲学	印度哲学
		臨床心理学	人間心理学	芸術文化学	国際禅学	文化心理学	心理臨床学
		社会臨床心理学	総合心理学	実践心理学	臨床福祉心理学	心理カウンセリング学	子どもコミュニケーション学
		国際キリスト教福祉学	宗教文化学	心理コミュニケーション学	応用健康心理学	対人・社会心理学	多元心理学
		仏教文化学	現代応用心理学	コミュニケーション心理学	人間心理応用学		
		人文学	人文科学	文化学	図書館学	人文・社会・自然	文
	そ の 他	社会心理教育学	キリスト教文化学	文明学	図書館・情報学	生活文化学	人間関係学
		欧米文化学	日本文化(学)	心理社会学	国際関係学	東洋文化	人文学類
		比較文化学類	人間科学	国際学	ヨーロッパ文化学	マスコミュニケーション学	総合文化学
		国際文化学	行動科学	地域科学	行動科学課程	文化課程	英米文化
		国際言語文化学	コミュニケーション学	言語文化(文化言語)学	図書館情報学	行動学	工芸文化学
		地域文化学	人間文化学(課程)	文化行動学	心理・教育学	英米地域研究	人文社会学

(参考)文部科学省「学科系統分類表(高等教育機関)」
http://www.mext.go.jp/b_menu/toukei/chousa01/kihon/shiryo/sh_detail/1375044.htm

B-3 小分類コードへの、大中分類の付加

(対応内容)

- 「所在地(都道府県＋政令市)」や「国籍」などは区分数が多く、BIツールの利用にあたって人による選択が困難となるため、大中分類を付加した。
- コード番号とともに、その意味を示す列も挿入した。

original	formatted		original	formatted				original	formatted	
学校調査番号	番号別大学名	大学名	所在地	所在地コード	所在地県	八地方区分		種類・国籍	国籍	地域
0100	0100北海道大学	北海道大学	1	01北海道（指定都市を	北海道	1北海道地方		A17	カンボジア	東南アジア
0104	0104北海道教育大学	北海道教育大学	2	02青森県	青森県	2東北地方		A19	ラオス	東南アジア
0108	0108室蘭工業大学	室蘭工業大学	3	03岩手県	岩手県	2東北地方		A20	ブルネイ	東南アジア
0112	0112小樽商科大学	小樽商科大学	4	04宮城県（指定都市を	宮城県	2東北地方		A22	台湾	東南アジア
0116	0116帯広畜産大学	帯広畜産大学	5	05秋田県	秋田県	2東北地方		A24	東ティモール	東南アジア
0120	0120北見工業大学	北見工業大学	6	06山形県	山形県	2東北地方		B01	パキスタン	西アジア
0124	0124弘前大学	弘前大学	7	07福島県	福島県	2東北地方		B02	インド	西アジア
0128	0128岩手大学	岩手大学	8	08茨城県	茨城県	3関東地方		B03	ネパール	西アジア
0132	0132東北大学	東北大学	9	09栃木県	栃木県	3関東地方		B18	ブータン	西アジア
0136	0136宮城教育大学	宮城教育大学	10	10群馬県	群馬県	3関東地方		B21	モルディブ	西アジア
0140	0140秋田大学	秋田大学	11	11埼玉県（指定都市を	埼玉県	3関東地方		B19	バーレーン	中近東
0144	0144山形大学	山形大学	12	12千葉県（指定都市を	千葉県	3関東地方		B20	オマーン	中近東
0148	0148福島大学	福島大学	13	13東京都（23区を除く	東京都	3関東地方		C01	イラン	中近東
0152	0152茨城大学	茨城大学	14	14神奈川県（指定都市	神奈川県	3関東地方		C02	トルコ	中近東

B-4. 合計行および無回答行の削除

(対応内容)

- 元の統計表に挿入されている、「小計」や「合計」などの行は削除した。
 - BIツールで集計する際、誤りが起こるため。
- 無回答の行や、集計列を一行に変換する過程で生成される無回答行は、削除した。
 - 行数が膨大となり、動作が遅くなるため。
 - もともと「0(ゼロ)」と入力されてあったものと、分別不能となる。

(今後の課題)

- ◆ 「0(ゼロ)」と入力されてある欄について、大学が「ゼロ」と入力したのか、もともと無回答なのに学校基本調査システムにより自動的に「ゼロ」がセットされてしまったのか、判別不能なケースがある。

⇒ 判別は不能であるが、可能な限り、もともと無回答であったと判断されるものについては削除した。

課題と今後の展開

「大学基本情報」のデータの特徴

- 大学別のデータ
- 研究科・学部、専攻・学科別のデータ
- 細かい属性別
 - 男女別、外国人、社会人、休学者
 - 設置形態、年次等

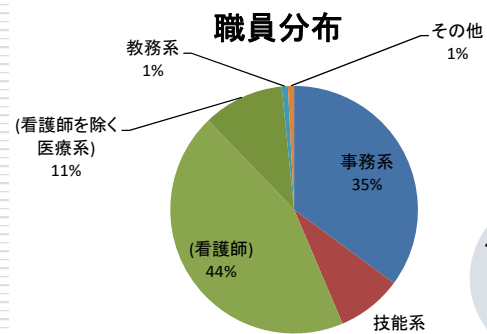


大学別のデータが
公開されたことで、
ようやく大学の現状
把握が行えるよう
になった！

「大学基本情報」のデータの可能性

1. 大学の多様性の広がり が可視化可能
2. 類似する大学を抽出・比較可能
3. クロス分析により、副次的データを分析可能
 - 異なるデータセットを組み合わせた分析
4. クロス分析等により、データの背景を分析可能
 - 飛び値などが発見しやすく、更なる詳細分析が可能
5. 課程別のデータが比較可能
6. (時系列分析が可能となる)
7. 設置形態別の大学の特色が明確となる

多様なデータ、多様な切り口

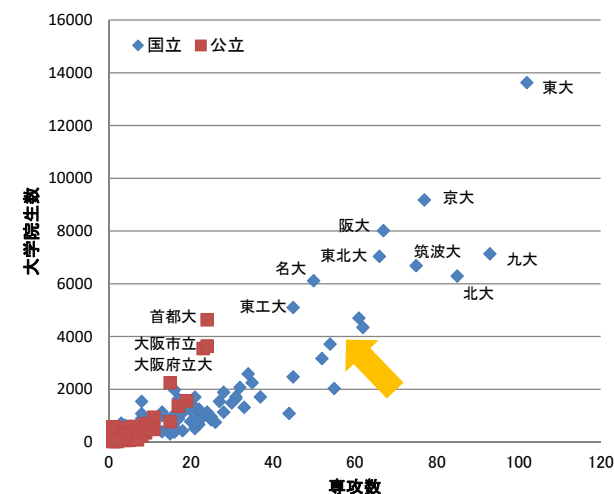


色々なデータに
色々な切り口がある。
それぞれで自大学の
ポジションは
分かるけど……。

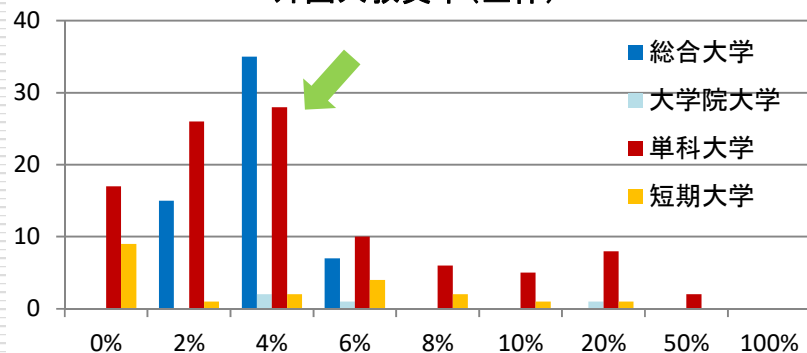
なんか
統一のアプローチが
ないと、バラバラの
分析があるだけ！



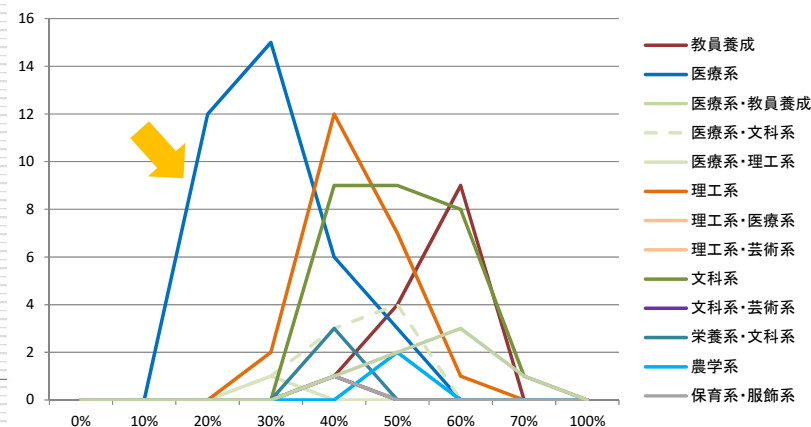
専攻数 vs. 大学院生数



外国人教員率(全体)



教授率(単科大学 分野別)



大学のグルーピングと、分野別(研究科・専攻)比較

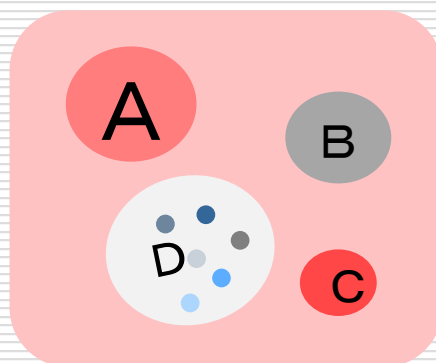
□ 大学のグルーピングと、グループごとの特徴理解

STEP1: 類似大学へのグルーピング

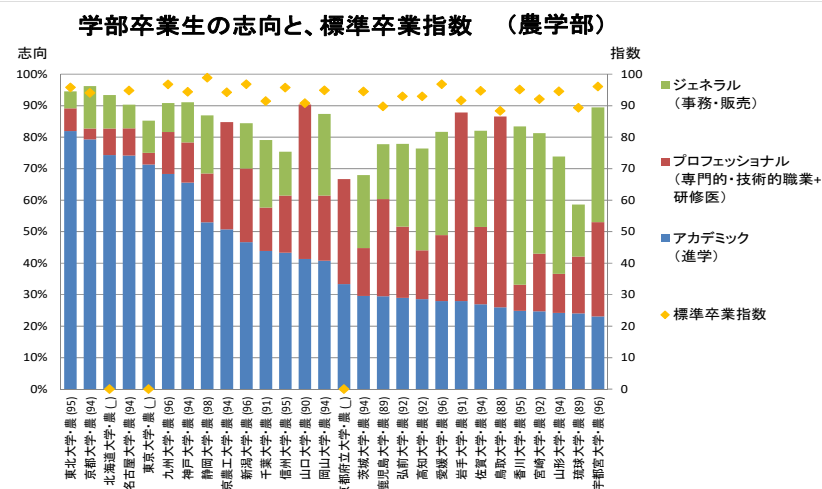
- ・ クラスタ分析等を利用

STEP2: グループ内で大学の特徴を抽出

- ① ユーザによる探索的アプローチ
- ② システムによる特徴の自動検出

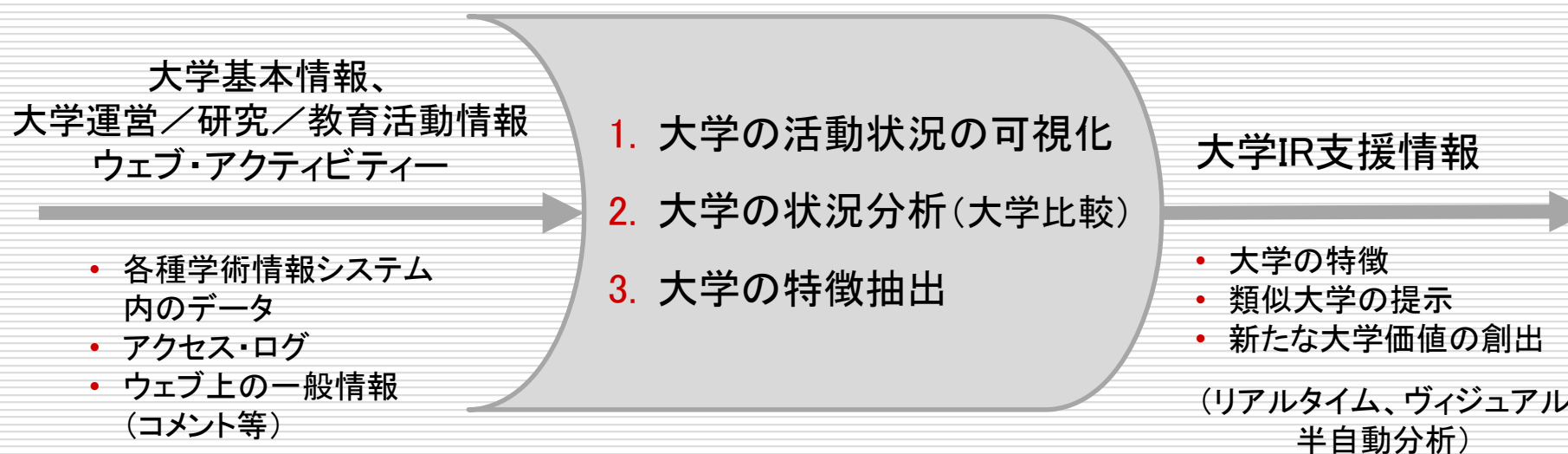


■ 全体像のなかで、大学の位置づけを把握



全国版大学IR支援システム(IRSSJU)(概念図)

全国版大学IR支援システム(IRSSJU)(概念図)



データの形式変換後の、 「大学基本情報」データセット利用の留意点

1. データセットの年度別欠落や集計方法の変化

- 「大学基本情報」は、永続的な分析に耐えるのか？

2. 大学別の集計—「番号別大学名」の利用の勧め

- 大学の並び順が分かりやすくなり、かつ大学の表記揺れにも対応。

3. 「学部・研究科」「学科・専攻」の分野別分析における留意点

- コードは便利であるが、細かすぎるので気をつける必要
 - 「商学(C201)」「経済学(C203)」「経営学(C205)」「経営経済学(C208)」「国際経済学(C209)」「応用経済学(C215)」など

4. 空欄データへの注意

- 単表で粒度の異なる集計がなされている場合がある。
- BIツールだと見逃す危険性があるため、注意要。

分析をしていく上での課題

1. 「学部・研究科」や「学科・専攻」の分野別分析の際のグルーピング

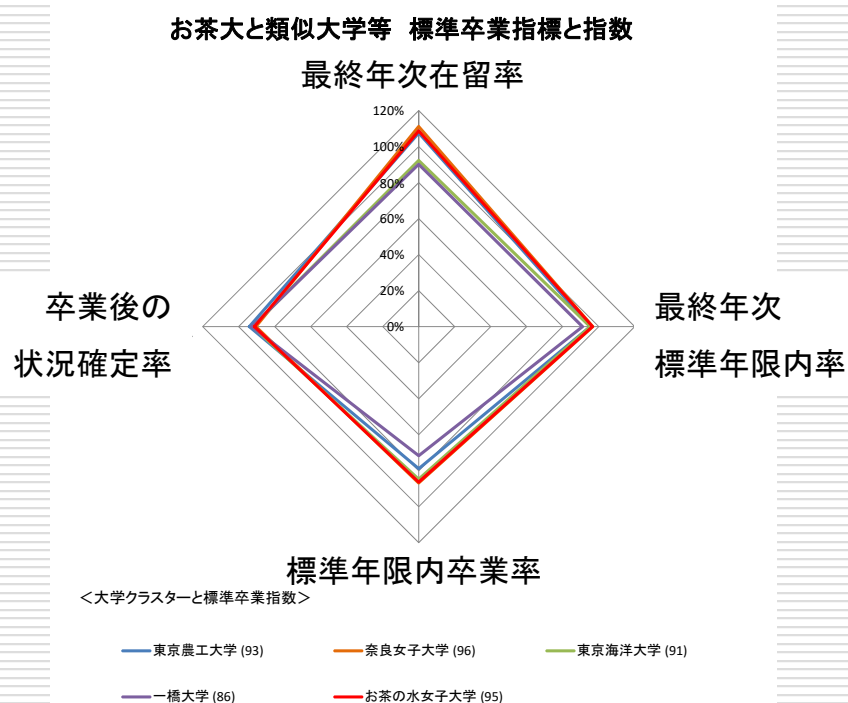
- 原則としては、IRerが自身の分析目的に応じて、コードとKWを組み合わせて、比較対象をグルーピングする必要がある。
- ただし、よく使われ可能性のあるグルーピングは、事前に提示されてよい。

2. 規格化された成果指標の開発

- 「大学基本情報」は、学生数や教員数などの基本的データのみので、これをそのまま可視化しても、つまらないだけ。
- しかし、「女子学生比率」「留学生比率」「休学者数」等、細かく分析すれば意味が生まれる。
- 一方、IRerの力量に依存しないために、規格化された成果指標の開発が望まれる。

成果指標開発の重要性

□ いくつかの重要な成果指標等を類似大学と比較。



(注1) 括弧内は標準卒業指数(=4つの指標の和)
(注2) 100%を越えている最終年次在留率は、1年次の学生数と4年次の学生数を比較しているため。

大学改革支援・学位授与機構

データ分析集 指標一覧 44指標

参考 3

区分	指標番号	データ・指標	指標の計算式
1. 学生入学・在籍状況データ	1	女性学生の割合	女性学生数/全学生数
	2	社会人学生の割合	社会人学生数/全学生数
	3	留学生の割合	留学生数/全学生数
	4	正規課程学生に対する科目等履修生等の比率	科目等履修生数/全学生数
	5	海外派遣率	海外派遣学生数/全学生数
	6	受験者倍率	受験者数/募集人員
	7	入学定員充足率	入学者数/定員
	8	学部生に対する大学院生の比率	大学院生総数/学部学生総数
2. 教職員データ	9	専任教員あたり学生数	学生数/専任教員数
	10	専任教員に占める女性教員の割合	女性専任教員数/専任教員数
	11	本務教員あたり研究員数	研究員総数/本務教員数
	12	本務教員あたり職員数	職員数/本務教員数
3. 進級・卒業データ	13	本務教員あたり職員数(常勤・常勤以外別)	職員数(常勤)/本務教員数 職員数(常勤以外)/本務教員数
	14	留年率	留年者数/学生数
	15	退学率	退学者・除籍者数/学生数
	16	休学率	休学者数/学生数
	17	標準修了年限内卒業・修了率	標準修了年限での卒業・修了者数/標準修業年限前の入学者数
	18	標準修了年限×1.5年以内での卒業・修了率	標準修了年限×1.5年以内での卒業・修了者数/標準修業年限×1.5年前の入学者数
	19	受験者数に対する資格取得率	合格者数/受験者数
	20	卒業・修了者数に対する資格取得率	合格者数/卒業・修了者数
	21	進学率	進学者数/卒業・修了者数
	22	就職率	就職者数/卒業・修了者数
4. 卒業後の進路データ	23	職業別就職率	職業区分別就職者数/就職者数合計
	24	産業別就職率	産業区分別就職者数/就職者数合計
5. 競争的外部資金データ	25	本務教員あたりの科研費申請件数(新規)	申請件数(新規)/全本務教員数
	26	本務教員あたりの科研費採択内定件数	内定件数(新規)/全本務教員数 内定件数(新規・継続)/全本務教員数
	27	科研費採択内定率(新規)	内定件数(新規)/申請件数(新規)
	28	本務教員あたりの科研費内定金額	内定金額/全本務教員数 内定金額(間接経費含む)/全本務教員数
	29	本務教員あたりの競争的資金採択件数	採択件数/全本務教員数
	30	本務教員あたりの競争的資金受入金額	受入金額/全本務教員数

今後の展開

- 加工したデータセットを全国のIRerの利用に供する。
 - 同時に、オープンサイエンスの動向に合わせて、「データ・パブリッシング」の経験を積む機会とする。
-
1. データセットをダウンロードした人が利用できるように、説明ファイルや適切なメタデータを設定
 2. DOI設定と、データ公開
 3. データ・ジャーナルへの論文投稿

「大学基本情報」のBIツール利用の可能性と課題

— 複合的統計表のデータベース化とデータ解析に向けて —

船守美穂（国立情報学研究所）

中鉢直宏（帝京大学）

1. 本発表の目的と課題

大学改革支援・学位授与機構（NIAD）のウェブサイトには、国内国公立大学の学生数や教職員数などに関わるデータが「大学基本情報」として公開されている。これは、国の指定統計である「学校基本調査」の個票データを各大学から得ることで、これまでの集計値レベルの公表から一步踏み込み、大学別／学部・研究科別／学科・専攻別にデータを公開したもので、各大学が自身の大学経営戦略を検討することを可能とした、極めて意義深いデータセットである。

一方、このデータセットは、「学校基本調査」という伝統的な統計表をベースとして作成され、利用の方法としても所謂統計表として、人間の読みやすさや理解のしやすさに配慮して作成されたものである。このため、統計表の行や列の途中に「小計」の欄があり、また見出し行も複合的かつ複数行にわたっているなど極めて複雑で、これらのデータをデータベースのようにBIツールなどで機械的に読み込むことが困難なレイアウトとなっている。また単一の年度において20以上の表が提供されているが、それらの表の間のデータ項目の用語統一も十分になされていないため、複数の表を機械的に相互連携させることも現状では出来ない。さらに表が年度ごとに公表されているため、単一の表を複数年にわたって解析することも出来ない。しかし一方でこれらの表は、学科・専攻別のデータが列挙され、行数が膨大で、表形式のままで人間が表の内容を読み取るには限界がある。

このため、この「大学基本情報」のデータセットを可視化とデータ解析が可能となるように、BIツールで読み込むためのデータ加工を行った。この加工したデータセットは公開し、全国の大学IR担当者等の利用に供する予定である。本小論は第一義的には、このデータセットを利用する担当者が、どのようなデータ加工が行われたかの把握可能なように、データ加工のプロセスやそのような加工の判断理由について、ドキュメント化したものである。また今年度以降、新たなデータセットが加わった際も、このドキュメントをもとに同様の加工ができることを期待している。さらに、本「大学基本情報」のBIツール利用に向けての加工の手法が、一般的な複合的統計表のためのデータベース化とデータ解析においても汎用的に利用可能であることを、期待している。

2. 「大学基本情報」データセットの概要

2016年度末において、表1に挙げるデータセットが2012-16年度の5年度分、公開されている。一方、年度によっては欠落しているデータセットや、集計方法が途中から変更となったデータセットがある。集計方法が変更となったデータセットについては、他のデータセットから類似のデータセットを生成可能な場合、そのようなデータを挿入した。

発表要旨

表 1 : 「大学基本情報」の年度別データセットと、リシェイプ後のデータセット

(ファイル名)	2012	2013	2014	2015	2016	リシェイプデータ
(7-A) 学生数	○	○	○	○	○	07go1_学生数
(7-B) 教員数（本務者）	○	○	○	○	○	07go2_教員数（本務者）組織区分別
(7-1) 教員数（本務者）（再掲）	○	○	○	○	○	07go3_教員数（本務者）教員状態別
(7-Z) 教員数（兼務者）	○	○	○	○	○	07go4_1_教員数（兼務者）職位別 07go4_2_外国人教員数（兼務者）
(7-C) 職員数	○	○	○	○	○	07go5_1_職系別職員数 07go5_2_医療系職員数
(8-D) 学科別学生数	○	○	○	○	○	08go1_1_学科別学部生数 08go1_2_学科別学部入学者数 (8-E) 入学状況から抽出 (学科情報なし) 未公開 08go1_3_学科別学部入学志願者数
(8-2) 学科別学生数のうち休学者数	○	○	○	○	○	08go2_学部別休学者数
(8-3) 学科別学生数のうち最低在学年限超過学生数 (編入学者は除く。)	○	○	○	○	○	08go3_学部別最低在学年限超過学生数
(8-E) 入学状況	○	○	○	未公開	×	08go4_学部入学状況（志願者数、入学者数）
(8-O) 年齢別入学者数	×	×	×	×	○	08go4_1_年齢別学部入学者数 (8-E) 入学状況から抽出 未公開 08go4_2_出身高校種類別学部入学者数 08go4_3_学部留学生入学者数
(8-G) 出身高校の所在地県別入学者数	○	○	○	○	○	08go5_出身高校の所在地県別学部入学者数
(8-7) 専攻科・別科及び聴講生等の学生数	○	○	○	○	○	08go6_専攻科、別科、学部科目等履修生等の学生数
(8-R) 短期大学・高等専門学校・専修学校（専門課程）からの編入学者数	○	○	○	○	○	08go7_学部編入学者数
(9-H) 専攻別学生数	○	○	○	○	○	09go1_1_専攻別院生数 09go1_2_専攻別社会人院生数
(9-4) 専攻別学生数のうち休学者数	○	○	○	○	○	09go2_研究科別休学院生数
(9-5) 専攻別学生数のうち最低在学年限超過学生数 (編入学者は除く)	○	○	○	○	○	09go3_研究科別最低在学年限超過院生数
(9-I) 入学状況	○	○	○	○	○	09go4_1_専攻別大学院入学者数 09go4_2_専攻別大学院入学志願者数
(9-S) 年齢別入学者数	○	○	○	○	○	09go5_1_年齢別大学院入学者数 09go5_2_大学院社会人入学者数 09go5_3_大学院留学生入学者数
(9-8) 科目等履修生等の学生数	○	○	○	○	○	09go6_大学院科目等履修生等の学生数
(10-J) 学科別学生数、入学状況（本科）	○	○	○	未公開	○	10go1_1_学科別本科学生数 10go1_2_学科別本科入学者数 10go1_3_学科別本科入学志願者数
(10-6) 学科別学生数のうち休学者数	○	○	○	未公開	○	10go2_本科休学者数
(10-K) 出身高校の所在地県別入学者数	○	○	○	未公開	○	10go3_出身高校の所在地県別本科入学者数
(10-9) 専攻科、別科及び科目等履修生等の学生数	○	○	○	未公開	○	10go4_専攻科、別科、本科科目等履修生等の学生数
(10-Q) 年齢別入学者数(再掲)	×	×	×	×	○	10go5_1_年齢別本科入学者数 (10-J) 学科別入学状況から抽出 未公開 10go5_2_出身高校種類別本科入学者数 10go5_3_本科留学生入学者数
(11) 国費留学生、私費留学生、留学生以外の外国人学生	○	○	○	○	○	11go1_外国人学生数
(11 別掲) 国費留学生、私費留学生、留学生以外の外国人学生（専攻科・別科の学生、科目等履修生・聴講生・研究生	○	○	○	○	○	11go2_外国人学生数（専攻科、別科、科目等履修生等）
(20) 学校施設	○	○	○	○	○	20go_1_学校土地の用途別面積（㎡） 20go_2_学校建物の用途別延面積（㎡） 20go_3_厚生補導施設の延面積（㎡） 20go_4_学校建物の構造別延面積（㎡） 20go_5_学校建物増減面積（㎡） 20go_6_職員宿舍土地の用途別面積（㎡） 20go_7_職員宿舍建物の用途別面積（㎡）
(30-1) 状況別卒業者数、入学年度別卒業者数	○	○	○	○	○	30go1_1_状況別卒業者数 30go1_2_就職状況別卒業者数 30go1_3_博士課程満期退学者数 30go1_4_ポストドクター数 30go1_5_入学年度別卒業者数
(30-2) 職業別就職者数、産業別就職者数	○	○	○	○	○	30go2_1_職業別就職者数 30go2_2_産業別就職者数

3. 「大学基本情報」データセットのBIツール利用にあたっての課題と対応策

NIADのウェブページで公開されている「大学基本情報」は冒頭に説明したように、人が見ることを想定とした統計表として基本的には作成されており、BIツールの利用に適していない。また、複数表を連携して分析する上でも不整合があり、データベースとしての体をなしていない。

以下に、「大学基本情報」のデータセットをBIツールで利用可能とするために必要な加工の観点を挙げる。

A. BIツールで利用可能とするためのデータ加工

A-1. データセットの年度別欠落や集計方法の変化への対処

年度によっては欠落しているデータセットや、集計方法が途中から変更となったデータセットがある。これらについては、過去のデータセットを可能な限り最新のデータセット(2016年度)のフォーマットに合わせて、整形した。しかし、データが存在せず、そのような加工が不可能であった場合もある(表1「×」)。

今後、大学改革支援・学位授与機構によりデータセットが毎年度追加されていくにあたり、集計方法や公開されるデータセットが変更となってくると、今回作成したデータセットを全て作り替える、あるいは追加されたデータセットを現在のフォーマットに変換する必要があることには、留意が必要である。この場合、データセットのBIツール利用への加工の手間が煩雑な上、そのようなフォーマットの統一化が不可能なデータセットが複数出現することは、避けられない。

A-2. データセット間の用語（列見出し）の不統一への対応

BIツールを利用する魅力の一つに、複数の表を連結させてデータ解析できることが挙げられる。たとえば、学科別ST比などは、学生数と教員数の表を連結させることにより、算出、表示可能となる。他方、これを実現するためには、異なる表の見出し、および、その値の表記が同一でないと、これらの表を機械的に連結することができない。学科別ST比の例では、「学科・専攻」という見出しと各データが、教員数と学生数の表の双方で、統一されている必要がある。

表の見出しは、もともとは個々の表において理解しやすいように振られており（特に、見出しが長くなりすぎないように、簡略に表現されている）、複数表横断的に統一性をもって振られているわけではない。今回のデータセットの加工においては、「大学基本情報」の20以上の表横断的に統一感をもった列見出しを振った。なお、上述の例の「学科・専攻」の各データ（＝学科名）の表記ゆれには対処せず、学科名を用いた複数表の連結には、「学科・専攻コード」が利用されることを想定している。

A-3. 複数の異なる集計値を含む単表の分割

「大学基本情報」の表には、異なる性格の集計値が単一の表に含まれているものがある。たとえば、「入学状況」の表には、「入学者数」と「入学志願者数」のデータが含まれている。このように、異なる性格の集計値が単一の表に含まれる場合は、表を複数に分割することとした。表1の「リシェイプデータ」に、複数に分割した表を一覧した。

A-4. 複合見出しの単行化

「大学基本情報」の表の一部は、人間の読みやすさに配慮して、列見出しが複数の行に

わたっている。たとえば学生数の表において「年次」や「性別」は、学生数の列の見出しに組み込まれており、複数年次の男女の学生数が横比較可能となっている。

一方 BI ツールを利用するためには、集計値を示す列は1列にまとまり、残りの表は属性を示す列として構成されている必要がある。つまり「年次」や「性別」などの属性は、列見出しとして組み込まれているのではなく、列情報として各行に、提示されている必要がある。

なお NIAD で公表する「大学基本情報」の表には、隠れた見出しの行が存在し、複数行にわたる列見出しを単一の見出しで示すラベル（たとえば「1年_男」「3年_女」等）が提供されているが、これは利用せず、見出しにおける全ての属性情報を列情報として組み込む作業を行った。また、「集計対象」という列を新たに設け、具体的に何を集計しているのか（たとえば「学生数」「休学者数」等）を記載した。

見出しにおける属性情報を列情報として組み込むにあたっては、“Tableau Excel Add-In”を利用した[1]。このツールは整合性のとれたクロス集計を BI ツールで扱うデータ形式に、ワンタッチで並び替えることができる。今回の作業は、このツールを利用することにより、大幅に軽減された。

B. BI ツールの利用しやすさのためのデータ加工

B-1. 分かりやすく、規則性あるファイル名の付与

NIAD で公表されている「大学基本情報」の各表のファイル名は、年度とコードとが組み合わさったものとなっている。またそのコードは、データの元となった「学校基本調査」に準拠しており、(7-A), (7-B), (7-1), (7-Z), (7-C)などという並びとなっている。これは「学校基本調査」を知らない者にとっては意味が理解できないだけでなく、パソコン等でファイルが表示される場合、ファイルの並び順がおかしくなる。また、A-4 節に示したように、今回は一つの表を複数に分割したこともあり、ファイルが順番に表示されるようにコードを振り直した。またファイル名からファイルの内容がくみ取れるよう、ファイルの保持するデータの内容についても、ファイル名に含めた。

B-2. コードの分割と内容情報の付加

たとえば「研究科番号」のコードなど、1桁目が課程（修士課程、博士課程等）、下3桁が分野を擬似的に意味すると把握されたものについては、コードを2つに分けた。また、人の可読性に配慮して、そのコードの内容を表現する列を付加した。ものによっては、コード情報と文字情報を組み合わせ、文字情報の並びが分かりやすくなるよう工夫をした。

「所在地コード」や「課程別」、「学校種別」などのコードの内容については、「学校基本調査の手引き」を参照した[2]。他方、情報として最も有用であるはずの学部・研究科や学科・専攻の分野を示すコードについては、コードと内容の対応表が公開されていない。文部科学省生涯政策局政策課調査統計企画室学校基本調査係によると、コード表は公開はしていないものの、「学科系統分類表」（コードなし）は公開しており、これを参考にされたいとのことであった[3]。今回のデータ加工では、「大学基本情報」に元からあった、学部・研究科や学科・専攻の分野を示すコードと、各大学において手入力された学部・研究科や学科・専攻の名称（表記ゆれ含む）との双方を残し、これらの対応関係が確認できるようになっている。

B-3 小分類コードへの、大中分類の付加

たとえば「所在地（都道府県＋政令市）」や「国籍」などは区分数が多く、BIツールの利用にあたって人による選択が困難となるため、大中分類を付加した。「所在地（都道府県＋政令市）」については「所在地県」や「八地方区分」、「国籍」については「地域区分」などが付加され、絞り込みが容易になっている。

B-4. 合計行および無回答行の削除

「大学基本情報」の表には、人の可読性に配慮してときどき、小計や合計などの行が挿入されている。これらは「小計」等のフラグを立てておけば、その他の行と分離可能であるが、BIツールでデータを分析する際に意図せず、このような行も含めて合算等をしてしまう危険性もあるため、それら行については削除をした。

また無回答行は、NIADで公表する「大学基本情報」の表上「0（ゼロ）」が入力され、「0（ゼロ）」と回答されたものと混在しているケースが見られる。このことによりBIツールを利用した分析において不整合が起きる可能性があるため、もともと「無回答」であったと把握可能な行については、削除した。

4. BIツール向けに変換後の「大学基本情報」データセット利用にあたっての留意点

（1）データセットの年度別欠落や集計方法の変化

A-1節に詳説したように、データセットにより年度別の欠落や集計方法の変化があり、これらについては可能な限り対処はしたものの、データを何も考えずに利用や可視化すると、データの飛び等で問題が生じる可能性がある。

（2）大学別の集計―「番号別大学名」の利用の勧め

「大学名」のフィールドは、各大学が入力した値がはいっており、表記ゆれがあることがある。このため、元データから存在する「学校調査番号」と「大学名」とを機械的に組み合わせで作成した「番号別大学名」を、大学別の集計等では利用することを推奨する。

（3）「学部・研究科」「学科・専攻」の分野別分析における留意点

「大学基本情報」のデータセットの魅力の一つは、「学部・研究科」や「学科・専攻」別の個票データが存在し、これらの分野別にデータ分析が可能なことである。同じランクの大学とみなされていても、学部・研究科構成等が異なる大学は、学生数や教職員数などの基本的に数値さえも、相互比較することができない。

一方「学部・研究科」や「学科・専攻」の名称は、たとえば同じ経済学系の学科が、大学によって「経済学」や「経済学類」とされるなど、同じ内容であっても名称が異なり、学科名称でグルーピングすることができない。しかし学科・専攻コードではこれらは

“C203”で統一されているため、グルーピング可能である。他方、コード側も極めて細かく振られており、たとえば経済・経営系の学科では、「商学（C201）」「経済学（C203）」「経営学（C205）」「経営経済学（C208）」「国際経済学（C209）」「応用経済学（C215）」「経営情報（C217）」「経済工学（C219）」「国際商学（C224）」「国際経営学（C225）」「流通学（C226）」「経済情報学（C227）」などとなっており、単一のコードに依存するのは危険である。分析したい対象に応じて、たとえばコードの1桁目のみを参照する、「経営」というKWでグルーピングする等の配慮が必要である。

（4）空欄データへの注意

BI ツールを利用すると、データの入っている表を確認せずに、グラフ表示のためのデータラベル（見出し）のみを操作する危険性が高い。しかしたとえば「教員数」のデータの大部分は学科・専攻別に存在するものの、学部・研究科レベルで教員の所属が管理されていたり、大学執行部など大学本部に所属があり、「学科・専攻」欄は空欄である大学も存在する。これらに気づかずに、教員数を学科・専攻別に集計すると、大学別の総教員数と齟齬が生じるなど、問題が発生する危険性がある。これを回避するには、頻繁にデータに立ち戻りながらデータの可視化を行うなどの注意が必要である。

5. 「大学基本情報」データセット利用上の課題と今後の展開

「大学基本情報」のデータセットを BI ツールで利用可能なように整形したが、これを大学 IR に利用するという観点からみた場合、課題も多く残されている。

まず「大学基本情報」に含まれるデータは、学生数や教員数などの基本的なデータのみであり、たとえば大学の競争力などを示す指標は明示的には存在しない。このため、これらデータセットを単に可視化しても、つまらないだけである。しかしたとえば女子学生比率や留学生比率、休学者数、学年別の留年者数など、細かく分析すると大学の現状把握につながる指標は抽出可能であり、つまり、どのような目的でデータを分析するかというリサーチクエスションの立て方において、IRer の力量が大きく問われる状況となっている。

こうした個人の力量への依存を多少でも回避するためには、たとえば、ある程度規格化された「大学のヘルスチェックレポート」の作成・提示、大学の多様性や学生の標準年以内修了状況などの「大学の課題別テーマ」に基づいた一連の分析流れの提示などを明確にし、示していくことが必要と考えられる。

同時に、4.(3)節で示した、「学部・研究科」や「学科・専攻」の分野別分析についても、個々の大学や担当者の判断でグルーピングするだけでなく、ある程度大ぐくりであっても、よく使われる可能性のあるグルーピングは事前に提示することが重要である。

なお今回は「大学基本情報」という、国の指定統計である「学校基本調査」をベースとした一連の統計表を BI ツール利用可能なようにデータ加工したが、ここで見いだされた3節に挙げたデータ加工の手続きは、他の伝統的な統計表のデータ加工にも適用できる方法である。近年、「公的統計調査の調査票情報等の学術研究等への活用」に見られるように、伝統的な統計調査の個票が、一定程度の匿名化措置はなされていても、利用可能となっていく上で、有用な知見となっていくと考えられる。

【参考文献】

- [1] Tableau Community, “Tableau Add-In for Reshaping Data in Excel”
(<http://kb.tableau.com/articles/knowledgebase/addin-reshaping-data-excel>)
- [2] 文部科学省「平成28年度 学校基本調査の手引（大学，短期大学，高等専門学校）」
(http://www.mext.go.jp/b_menu/toukei/chousa01/kihon/sonota/1355787.htm)
- [3] 文部科学省「学科系統分類表（高等教育機関）」
(http://www.mext.go.jp/b_menu/toukei/chousa01/kihon/shiryo/sh_detail/1375044.htm)