



公開データの効率的な データ変形 (RESHAPING) について

山形大学 学術研究院 (ICT・IR担当)

鈴木 達哉

suzukit@jm.kj.yamagata-u.ac.jp

March 2, 2018

継続的改善のためのIR/IEセミナー (於：九州工業大学 戸畑キャンパス)

本日の内容

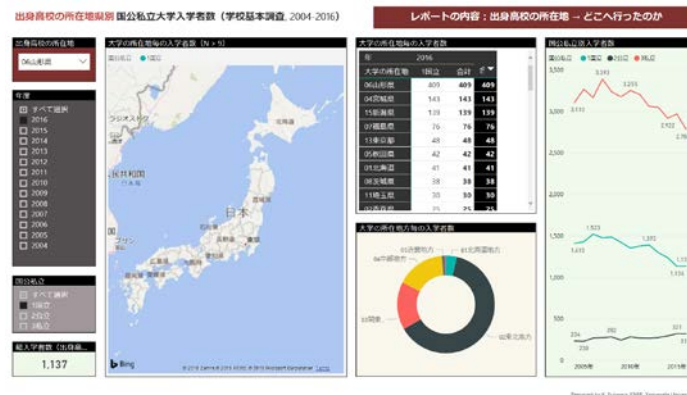
1. 概要の確認
2. データマネジメントについて
3. データマネジメントツールとして「クエリ エディター」を使用した公開データの加工例

1. 概要の確認

公開データをBIツールなどのデータ分析ツールを活用して可視化するためには、データマネジメントが不可欠である。

中でも、Reshapingと呼ばれるデータ形式の変形技術は、特に重要だと考えられる。

本発表では、Microsoft Excel 2016付属の「クエリエディター」を用いて、具体的かつ効率的なデータ変形の方法を紹介する。



2. データマネジメントについて

■ データマネジメントの必要性

- 学内で使用しているデータは、レポート用に見やすい形になって場合が多い。
- 文部科学省学校基本調査等の公開データもエクセルファイルではあるが、印刷を想定している形式がある。

The image shows a screenshot of an Excel spreadsheet with a large grid of data. Several red boxes are drawn around the spreadsheet to highlight specific features: one box is around the top-left corner (title area), another is around the top-right corner (total area), and several others are around the bottom edge (tab area). A blue callout box points to the top-right corner, and a green arrow points downwards from the bottom of the spreadsheet.

タイトル、合計、罫線、タブ、等の見やすくするための工夫がされている場合が多い。

しかし

(例：学校基本調査 平成20年度 16 出身高校の所在地別_入学者数_大学・大学院のエクセルファイル)

これらは、そもそもデータ分析を目的として作成されている資料ではないものがあるため、BIツールなどで使用するには、これらの「印刷時に見やすくするための要素」を取り除き、データを使いやすいように変形し管理しつづける必要がある。

2. データマネジメントについて

■ データマネジメントのプロセス

データを使いやすいように変形し管理し続ける流れ

- データの入手・更新
データがある場所、形式・内容、いつ更新されるか等把握する
- データの変形
分析しやすい形、経年で見やすい形に変形する
- データの管理
データベース、Excel、サーバー、共有・閲覧権限等設計する

「集計・分析」のために必要なプロセス

2. データマネジメントについて

■ 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						ロング型			
計	北海道	青森	岩手	宮城	...	年度	大学の所在地	出身高校の所在地	入学者数
北海道	14941	410	219	161	...	2016	北海道	北海道	14941
青森	394	1783	272	94	...	2016	北海道	青森	410
岩手	99	257	1337	278	...	2016	北海道	岩手	219
宮城	157	599	842	6289	...	2016	北海道	宮城	161
秋田	52	119	171	87	...	2016	青森	北海道	394
山形	68	73	162	474	...	2016	青森	青森	1783
...

変形

3. クエリ エディターについて

- Power BI、Excelに付属の「クエリ」を作成するための「エディター」。
- 操作内容を「ステップ」として記録して、それをまとめて「クエリ」を作成する。これによって、変更の履歴が記録される。また、その履歴をステップごとに変更したり、削除したりすることができる。元となるデータをクエリ エディターで読み込んで変形し、エクセルに読み込ませて使用することで、「データ変形」が可能である。
- 多くの業務で使用しているExcelのツールなので、共有しやすい。

方法	費用	難易度	手間
手作業	無料	低い	苦痛（件数が少なければ楽だが、件数多い場合）
クエリ エディター	PowerBI（無料）・ Excel（有料） 付属のツール	Excelがわかっていると低い	楽（操作を覚えれば）
プログラミング	無料	未経験者には高い	楽（書けるようになれば、自由に操作可能）



3. クエリ エディターについて

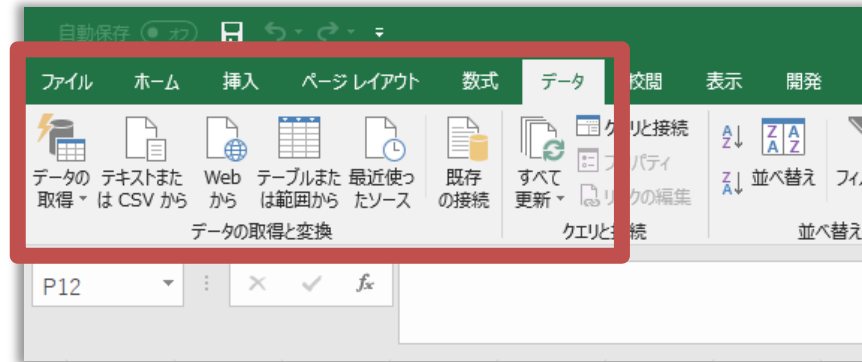
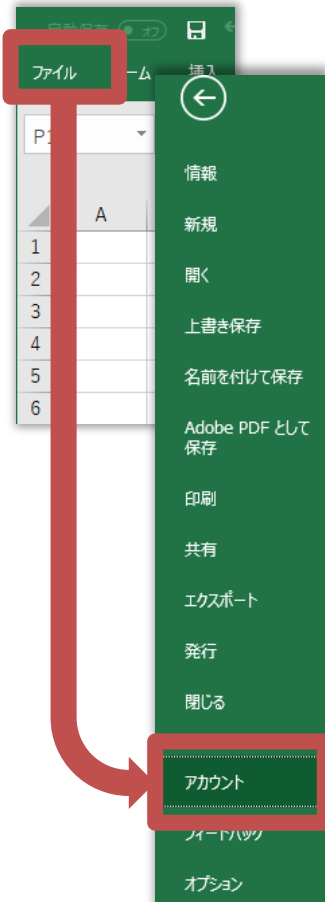
■ 主な機能

- データの型の変更：10進数、整数、パーセンテージ、日付、テキスト等
- 行・列の変更（追加・削除）
- **列の分割：区切り記号、文字数による分割**
- 値の置換
- **列のピボット解除（ワイド型をロング型に変換）**
- **ステップ・クエリの記録：手順を記録するため、戻ることができる**
- Power Query 式言語による式の作成

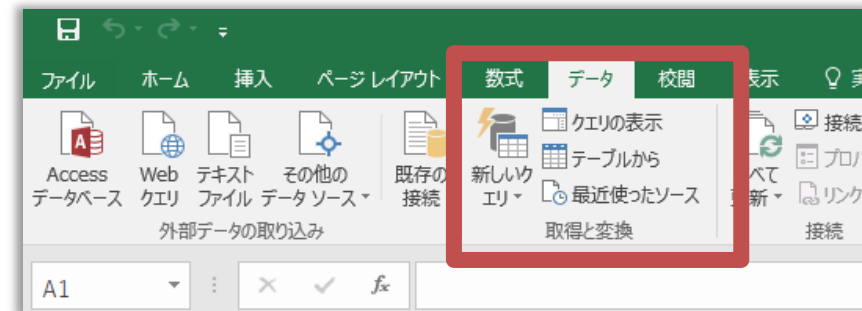
3. クエリ エディターについて

- Excelのバージョンによるメニューの違いがある。（表示は違うが、機能は同一）

- Office 365 の場合



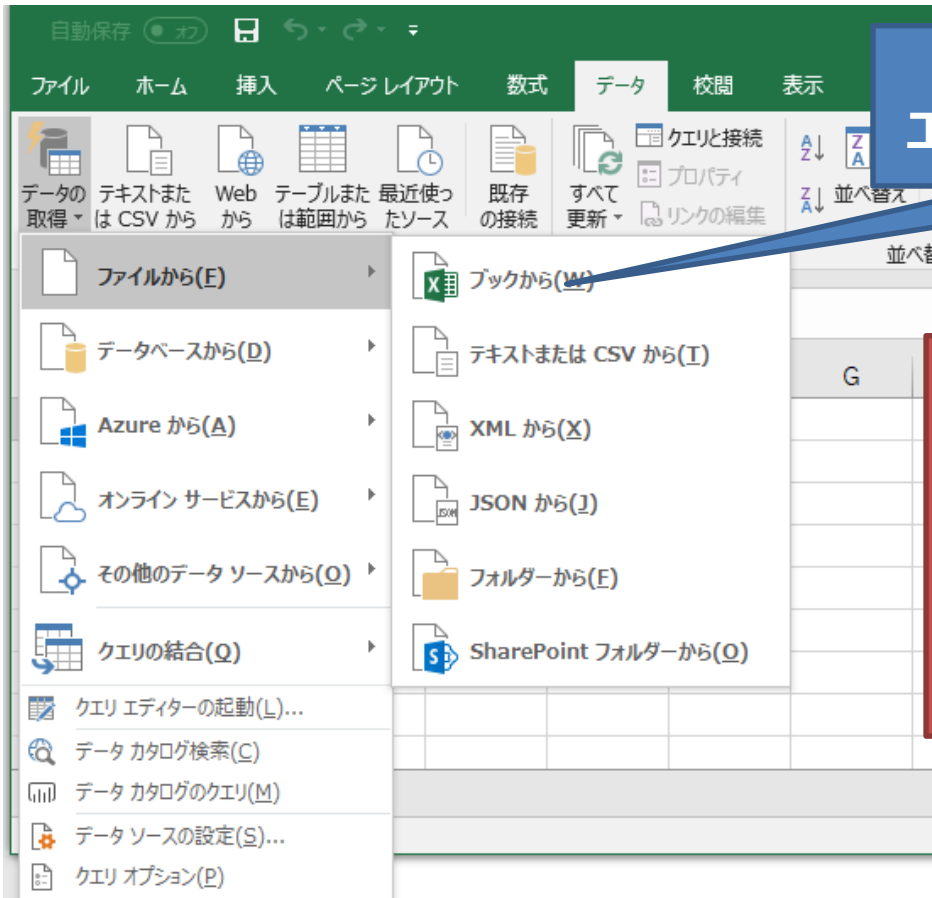
- Excel 2016 の場合



3. クエリ エディターについて

■ データの読み込み

「データの取得」 - 「ファイルから」 - 「ブックから」 - エクセルファイルを指定する。



「ブックから」は、
エクセルファイル指定を意味する

- 取得できるデータは多種多様
エクセルファイルのほかにも、
CSV,XML,JSON,
SQL Server,Access,WEB,
ODBC,Azure,Facebook
などが選択可能。

3. クエリ エディターについて

■ 画面の見方

The screenshot shows the Query Editor interface with several callouts:

- タブ・リボン部分**: Points to the top ribbon menu containing options like 'クエリのマージ' and 'クエリの追加'.
- クエリウィンドウ (ファイルに記録されたクエリの選択ができる)**: Points to the left sidebar where saved queries are listed.
- クエリの設定ウィンドウ (ステップが記録される)**: Points to the right sidebar where query settings and steps are configured.
- プレビューウィンドウ (ステップの結果をプレビューできる)**: Points to the central data table showing the results of the query.

出身高校の所在地	Column2	計	北海道	青森	岩手	宮城	秋田	山形	福島
大学の所在地	29	45148	14965	3354	3106	7306	2086	2154	2953
北海道		45148	14965	3354	3106	7306	2086	2154	2953
青森		3421	1948	1948	277	104	189	44	31
岩手		2625	2625	2625	2625	2625	2625	2625	2625
宮城		11845	171	171	171	171	171	171	171
秋田		2059	62	62	62	62	62	62	62
山形		2794	60	60	60	60	60	60	60
福島		3351	31	31	31	31	31	31	31

3. クエリ エディターについて

■ 使用データについて

学校基本調査、大学基本情報等のデータはワイド型のため、BIツールで使用するために、データ変形が必須となる。

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AIAI	AD	AI															
1	19 出身高校の所在地																県別 入学者数 (8-1)																												
2																																													
3	1 計																																												
4	出身高校の所在地																大学所在地											出身高校の所在地																	
5	大学の所在地																大学の所在地											大学の所在地																	
6	平 成 20 年 度	607,159	21,033	5,338	4,804	10,569	4,104	5,008	8,357	15,454	9,953	9,385	31,651	27,105	70,274	38,358	10,660	4,754	5,327	4,029	4,989	9,764	9,891	17,486	35,063						平 成 20 年 度														
7	北 海 道	19,370	14,941	410	219	161	166	108	83	154	120	97	176	146	340	196	126	80	62	24	27	93	46	157	209						北 海 道														
8	青 森 県	3,158	394	1,783	272	94	194	55	43	26	27	9	13	14	11	5	13	4	2	-	1	7	2	10	9						青 森 県														
9	岩 手 県	2,632	99	257	1,337	278	211	83	67	12	35	4	18	11	30	9	22	2	3	2	2	7	8	12	18						岩 手 県														
10	宮 城 県	12,189	157	599	842	6,289	538	867	994	175	181	161	92	68	146	77	223	42	40	6	46	91	18	116	34						宮 城 県														
11	秋 田 県	1,852	52	119	171	87	908	69	41	30	67	23	8	10	18	10	59	12	5	14	5	27	8	44	70						秋 田 県														
12	山 形 県	2,586	68	73	162	474	95	879	196	61	103	42	36	21	41	13	82	7	5	1	9	25	8	42	44						山 形 県														
13	福 島 県	3,578	32	85	116	307	115	177	1,644	195	235	86	35	39	40	25	125	12	3	3	15	54	9	50	19						福 島 県														
14	茨 城 県	7,285	90	82	79	99	76	98	289	2,942	257	141	327	618	440	119	105	41	35	33	48	97	44	161	115						茨 城 県														
15	栃 木 県	4,718	47	93	60	102	69	98	393	496	2,091	225	192	62	85	32	105	9	12	6	24	98	6	94	22						栃 木 県														
16	群 馬 県	6,250	97	61	94	85	71	118	151	221	508	2,400	592	144	226	84	269	43	31	19	37	264	19	158	57						群 馬 県														
17	埼 玉 県	33,677	370	263	186	315	224	336	701	1,585	1,326	1,566	10,555	2,108	7,246	983	853	104	101	55	251	811	93	671	218						埼 玉 県														
18	千 葉 県	27,363	275	181	163	223	155	263	532	2,170	610	460	2,023	9,307	4,789	990	461	88	68	37	250	575	121	705	185						千 葉 県														
19	東 京 都	137,859	2,021	786	663	1,135	744	874	1,893	4,991	2,850	2,436	14,703	11,295	43,731	17,993	2,368	722	519	319	1,657	2,351	672	3,812	2,037						東 京 都														
20	神 奈 川 県	52,629	671	283	240	435	287	404	763	1,469	940	961	2,259	2,590	11,261	16,811	1,093	338	212	128	683	1,198	274	2,484	797						神 奈 川 県														
21	新 潟 県	5,366	27	30	17	63	123	268	202	38	101	153	35	19	62	14	3,504	144	80	31	26	157	9	38	26						新 潟 県														
22	富 山 県	2,281	14	5	1	6	10	15	8	8	12	13	9	5	20	5	105	863	413	113	13	77	119	47	149						富 山 県														
23	石 川 県	5,674	64	22	9	31	37	35	26	25	37	72	26	24	44	16	340	727	1,943	448	22	185	180	156	220						石 川 県														
24	福 井 県	1,882	9	1	1	1	5	4	4	4	1	3	-	2	10	2	25	67	103	971	4	30	102	27	214						福 井 県														
25	山 梨 県	4,348	63	24	34	53	32	41	64	97	68	58	90	70	390	222	104	37	28	33	1,343	331	56	390	143						山 梨 県														
26	長 野 県	3,336	51	12	11	13	11	3	13	40	37	65	50	36	80	48	97	51	51	39	101	1,602	99	114	224						長 野 県														
27	岐 阜 県	4,322	27	8	10	8	11	8	9	9	4	7	6	3	24	13	28	75	47	75	12	113	1,650	136	1,188							岐 阜 県													
28	静 岡 県	7,549	68	23	15	32	17	45	53	100	92	59	61	100	217	175	68	39	46	53	108	163	138	4,282	570						静 岡 県														
29	愛 知 県	40,894	139	23	17	43	16	23	30	95	31	47	35	28	117	67	83	417	342	390	48	727	4,755	2,156	24,995							愛 知 県													
30	三 重 県	3,355	11	5	-	3	-	3	1	6	3	5	5	6	11	4	3	9	10	11	6	45	113	96	650						三 重 県														
31	滋 賀 県	7,922	80	9	8	16	6	11	12	27	17	33	26	23	41	24	25	62	77	117	13	56	247	137	437						滋 賀 県														
32	京 都 府	32,825	387	36	33	102	31	49	65	174	73	132	130	132	325	174	160	311	472	382	66	255	468	569	1,021						京 都 府														
33	大 阪 府	49,141	223	24	16	42	16	32	22	88	38	36	39	59	136	65	75	240	322	367	52	114	247	295	485						大 阪 府														
34	兵 庫 県	26,731	171	11	5	18	7	10	12	48	20	17	19	34	70	38	50	113	157	184	31	35	130	134	290						兵 庫 県														
35	奈 良 県	5,307	34	6	3	8	4	5	7	18	9	9	9	10	17	14	26	37	38	42	11	40	43	64	95						奈 良 県														
36	和 歌 山 県	1,726	6	2	1	-	1	2	2	2	3	5	3	1	6	3	2	6	8	8	1	10	15	16	25						和 歌 山 県														
37	鳥 取 県	1,324	5	-	2	-	-	-	-	1	7	-	7	4	1	7	3	4	5	9	8	8	1	12	10	8	26					鳥 取 県													
38	島 根 県	1,448	7	-	1	-	3	-	1	2	-	3	1	4	6	5	2	5	4	9	3	10	7	23	33						島 根 県														
39	岡 山 県	8,337	30	1	6	2	-	2	11	1	3	4	4	17	10	9	7	18	18	6	15	20	56	44	100						岡 山 県														
40	広 島 県	12,787	54	1	-	4	-	4	2	15	3	4	2	5	24	8	5	9	19	34	4	21	42	56	72						広 島 県														
41	山 口 県	3,964	20	2	-	-	-	1	4	3	6	3	2	1	5	8	6	3	4	7	1	7	12	20	25						山 口 県														
42	徳 島 県	2,608	9	1	1	-	-	-	2	4	1	2	3	6	9	4	1	3	3	2	1	2	7	9	23						徳 島 県														
43	香 川 県	2,054	3	-	-	-	1	1	-	2	-	2	2	5	4	2	-	-	1	7	-	3	5	6	16						香 川 県														
44	愛 媛 県	3,689	8	-	1	-	1	4	-	7	4	4	3	7	3	2	1	1	4	-	-	7	8	19	27						愛 媛 県														
45	高 知 県	1,770	19	-	2	2	-	2	1	6	5	1	3	3	12	5	4	3	1	3	5	8	16	16	27						高 知 県														

(例：学校基本調査 平成20年度 16 出身高校の所在地県別_入学者数_大学・大学院のエクセルファイル)

3. クエリ エディターについて

■ データの変形

「ワイド型」から「ロング型」へ変換

列を指定して、「変換」タブの「任意の列」の中にある「列のピボット解除」

16(8-1) - クエリエディター

ホーム 変換 列の追加 表示

データ型: 整数 値の置換 列のピボット解除
データ型の検出 フィル 移動
名前の変更 列のピボット リストに変換

任意の列

1 ² 北海道	1 ² 青森	1 ² 岩手	1 ² 宮城
14053	350	270	147
483	1948	277	104
105	256	1359	260
171	570	848	588
62	88	127	83
60	84	120	594
31	58	23	180

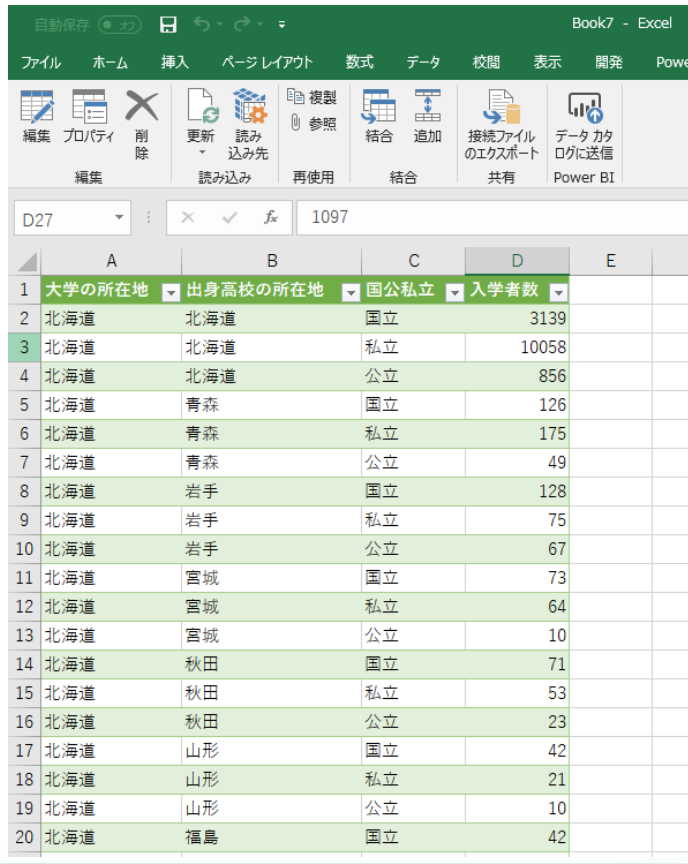
ABC 123 大学の所在地	A ^B 属性	1.2 値
1 北海道	北海道	14053
2 北海道	青森	350
3 北海道	岩手	270
4 北海道	宮城	147
5 北海道	秋田	147
6 北海道	山形	73
7 北海道	福島	76
8 青森	北海道	483
9 青森	青森	1948
10 青森	岩手	277
11 青森	宮城	104
12 青森	秋田	189
13 青森	山形	44
14 青森	福島	31
15 岩手	北海道	105
16 岩手	青森	256
17 岩手	岩手	1359
18 岩手	宮城	260
19 岩手	秋田	180

変換したい列を指定することで、その列のみ変換される。

3. クエリ エディターについて

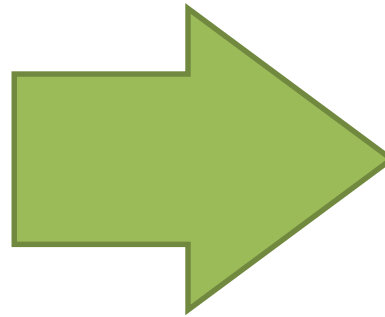
■ データの読み込み

ワイド型からロング型へ変換し、修正完了後のデータをエクセルファイルに読み込む。このデータをPower BIやその他データベースに取り込み活用する。



The screenshot shows the Microsoft Excel interface with a table containing university data. The table has four columns: '大学の所在地' (University Location), '出身高校の所在地' (Origin High School Location), '国公立私立' (Public/Private/Independent), and '入学者数' (Number of Enrollments). The data is as follows:

大学の所在地	出身高校の所在地	国公立私立	入学者数
北海道	北海道	国立	3139
北海道	北海道	私立	10058
北海道	北海道	公立	856
北海道	青森	国立	126
北海道	青森	私立	175
北海道	青森	公立	49
北海道	岩手	国立	128
北海道	岩手	私立	75
北海道	岩手	公立	67
北海道	宮城	国立	73
北海道	宮城	私立	64
北海道	宮城	公立	10
北海道	秋田	国立	71
北海道	秋田	私立	53
北海道	秋田	公立	23
北海道	山形	国立	42
北海道	山形	私立	21
北海道	山形	公立	10
北海道	福島	国立	42



データベース等で
使いやすい形

3. クエリ エディターについて

■ 効率的な入手方法

例) 文科省の学校基本調査データを入手する
e-Statから、

1. 「ファイルから探す」でエクセルファイルをダウンロード
2. 「データベースから探す」でDB機能でダウンロード
3. 「API機能」でAPI登録してダウンロード

The screenshot shows the e-Stat website interface. At the top, the logo 'e-Stat' is on the left, and '統計で見る日本' and 'お問い合わせ | ヘルプ | English' are on the right. Below the logo, there are navigation tabs for '調査項目' and 'その他'. A blue callout box on the left says 'ウェブ上で、データ変形をある程度行う'. A blue callout box on the right says 'クエリ エディター「WEBから」情報を取得する機能を使う'. In the main content area, there are several search options: 'データベースから探す' (highlighted with a red box), 'ファイルから探す' (highlighted with a red box), '分野から探す', '組織から探す', and '地域から探す'. At the bottom, there is a search bar with the text 'キーワードで探す 例: 国勢調査' and a '検索' button. On the right side, there is a vertical menu with options: '統計GIS', 'API機能' (highlighted with a red box), '統計LOD', and '統計ダッシュボード'.

まとめ

- データマネジメントの中でもデータ変形は重要な要素

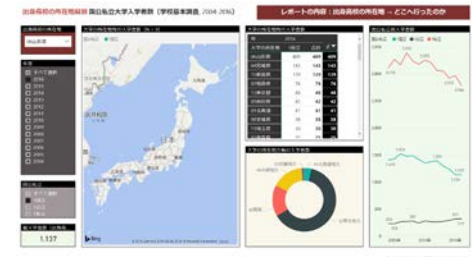
データ分析を目的として作成されているわけではない資料も多いため、**データを使いやすいように変形し管理しつづける**ことが必要。

- 公開データを効率的にデータ変形するために

データ変形は手軽なツールから始めることが重要。

エクセルに付属したソフトウェア等の**手軽なツールから導入**することで、安価にデータ変形をはじめることができる。

API等のさまざまな提供方法を活用する。



クエリ エディターを活用した公開データの整形

Google等の検索エンジンで「OIRE」と検索

山形大学 Office of Institutional Research & Effectiveness

The image shows a screenshot of the OIRE website on the left and a flyer for a workshop on the right. A red arrow points from the 'IR Activities' link on the website to the flyer.

Website Screenshot:

- OFFICE OF INSTITUTIONAL RESEARCH & EFFECTIVENESS
- APプロジェクトについて
- IR OIREについて
 - IRとは？IEとは？
- What's New?
- Events
 - APシンポジウム
 - 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勉強会
 - BI勉強会 (2017.11.20)
 - データ変形 (2018.2.26)
- IR Activities

Flyer Content:

- 山形大学 Yamagata University
- クエリ エディターを活用した公開データのデータ整形
- 鈴木 達哉
- suzukit@jm.kj.yamagata-u.ac.jp
- February 26, 2018
- 「データマネジメント」に関する勉強会 (於：東京工業大学キャンパスイノベーションセンター (CIC) 509号室)