

オープンデータ推進講座

統計データをオープンデータに変える

2019年1月11日

長野大学・前川道博研究室 & ゼミ
GEEKLAB.NAGANO



本資料は、クリエイティブ・コモンズ 表示4.0 国際 (CC BY 4.0) に従って利用できます。
(<http://creativecommons.org/licenses/by/4.0/legalcode.ja>)

本日のプログラム

統計データをオープンデータに変える

- 14:00 統計データのオープンデータ化
前川道博(長野大学企業情報学部教授)
- 14:30 「神エクセル」自動整形ツール試作版
シュク・メイスイ(前川ゼミ)
- 15:00 Excel→CSV変換実習
サポート: 中澤祐一(日本システム技研) + 前川ゼミ学生
- 15:20 休憩
- 15:40 情報交換
 - 統計業務の課題
 - ツール改善の要望・課題
 - 今後に向けて
- 16:20 アンケート
- 16:30 終了

オープンデータ推進講座

統計データのオープンデータ化

2019年1月11日

前川道博
長野大学・企業情報学部教授

昨今の状況

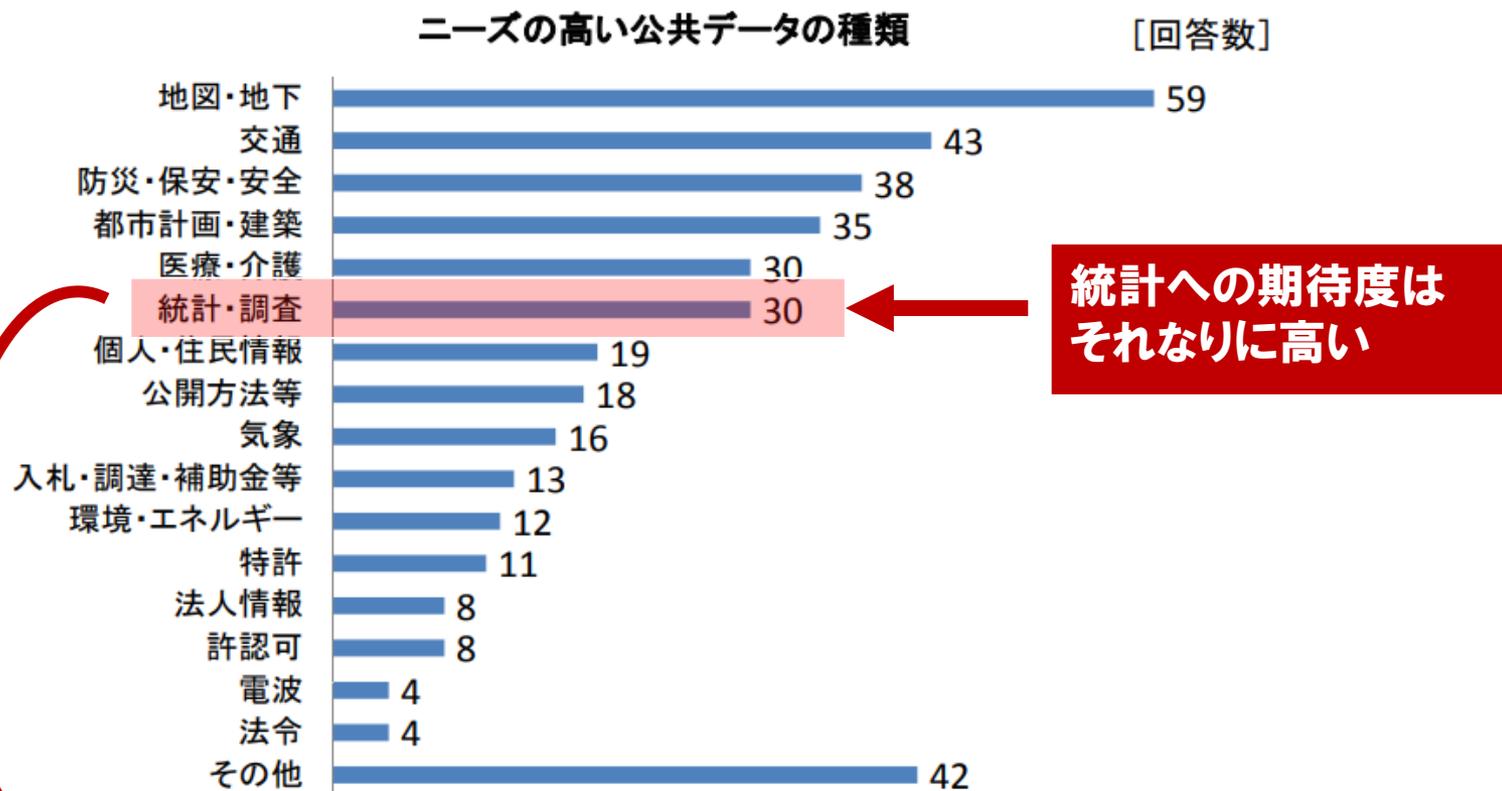
オープンデータ化の要求と対極の無関心

- ・ **オープンデータ化の政策的推進**
 - 官民データ活用推進基本法
 - ・ データ流通の拡大
- ・ **社会の知識基盤化の進展**
 - 旧来の業務スタイル・慣習
 - ICTの普及による情報授受文化の進展
 - まだ圧倒的に多い「無関心」「わからない」層

統計の実態の疑問

- 統計調査はどの程度、どう役立っているか
- 統計資料はどれぐらい参照されているか
- 統計データは何に役立つのか
- 統計データは誰が使うのか
- どうすると統計データはもっと活用されるか
- 統計をオープンデータにするメリットがわからないという大方の受け止め

統計のオープンデータ化への期待度



統計への期待度は
それなりに高い

統計・調査	5	国勢調査	基本統計区あるいは町丁字別に集計した指定統計や防災関連データ、固定資産の概要調査、自治体の保有する1/2500地図
		複合データ(地図・交通・気象等)	空間線量測定結果を各自治体が計測したモニタリングデータ、火葬場情報
		課税対象所得額データ	課税対象所得額データ(町丁目単位の所得階級別人数)

出典「公共データの産業利用に関する調査結果」(2013年、日本経済団体連合会)

http://www.keidanren.or.jp/policy/2013/020_honbun.pdf

前川研究室 & ゼミの取り組み メディア環境「信州デジタルコモンズ」試行

地域デジタル知識基盤プラットフォームのコモンズ (知識・データの共有地)

多様なデジタルアーカイブ群の包摂プラットフォーム



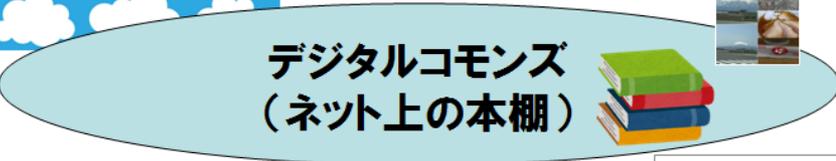
信州デジタルコモンズ みんなで共有できるネット上の本棚

信州学の学（知識）と学び（学習プロセスと成果）
を載せ合える「ネット上の本棚」



クラウド(雲上の世界?)

統計も



知識・データを棚上げて

棚卸し(閲覧・活用)

家・図書館・どこでも
PC・スマホどれでも

各地域ごとに

オープンデータにして



図書も

デジタルデータも



地域・全国・世界の人へ

地域の皆さんが

当事者になって統計データを使ってみると 上田城跡・観光者数の推移を知りたい

- ・ 上田市統計にはなし
- ・ 「長野県の統計情報」掲載の「観光地利用者統計調査」各年データ

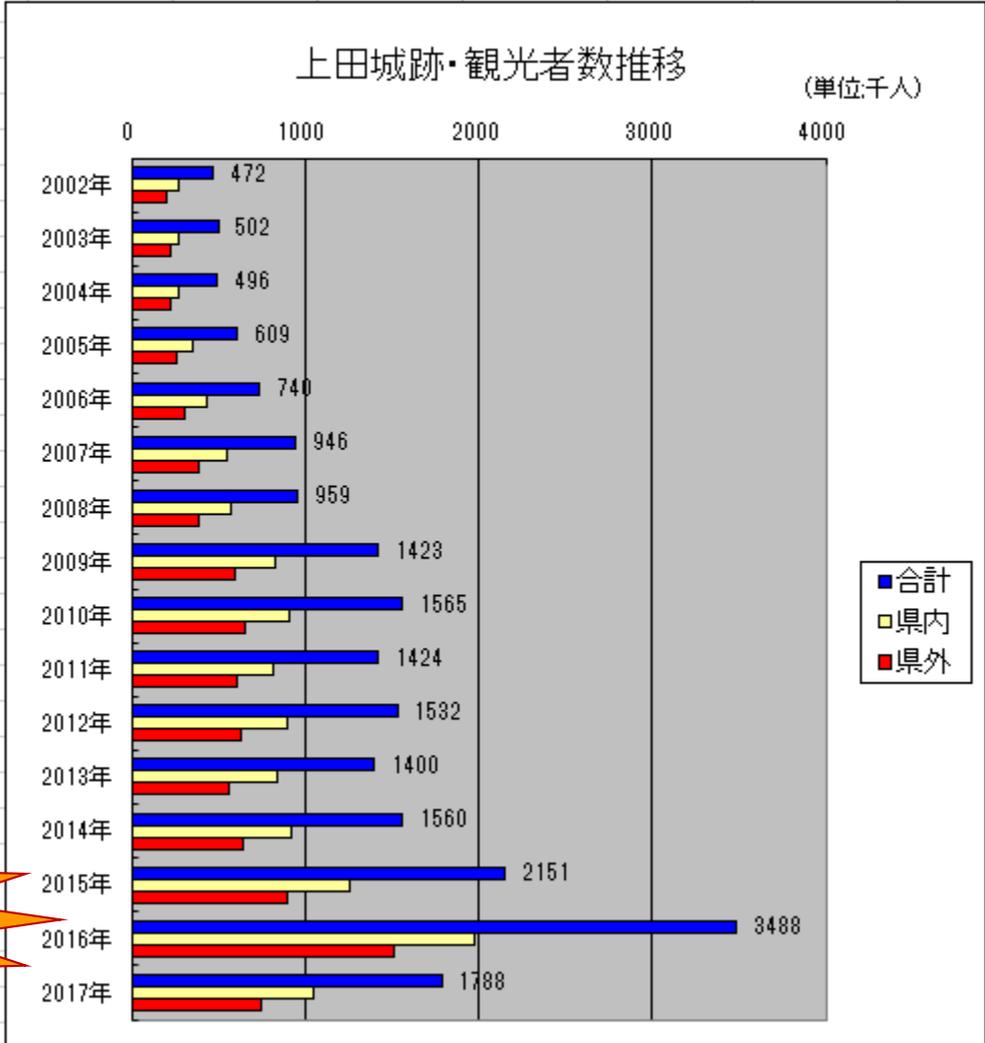
	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	平成29年観光地利用者統計調査結果													
2	観光地利用者延数（百人）													
3	市町村名	観光地名	合計	県内・県外別		日帰り・宿泊別		月						
県内				県外	日帰り	宿泊	1月	2月	3月	4月	5月	6月	7月	
40	上田市	上田城跡	17,876	10,442	7,434	17,876	0	333	587	659	4,600	2,282	1,306	1,212
41	上田市	信州の鎌倉・塩田平	2,981	1,259	1,722	2,981	0	155	151	265	240	269	171	226
42	上田市	別所温泉	7,324	3,194	4,130	4,447	2,877	826	662	698	638	559	448	757
43	上田市	丸子温泉郷	3,818	1,977	1,841	418	3,400	359	228	278	271	306	235	289
44	上田市	信州国際音楽村	782	608	174	700	82	21	18	32	71	71	219	93
45	上田市	菅平高原	10,892	3,042	7,850	2,549	8,343	535	850	435	98	1,200	564	2,150
46	上田市	海間温泉	305	114	191	114	191	1	1	1	12	21	23	32
47	上田市	美ヶ原高原	4,749	3,840	909	3,840	909	586	583	304	153	430	319	475
48	上田市	真田氏ゆかりの地	562	186	376	562	0	34	22	39	49	87	42	48

- ・ **問題：たった1スポットの観光者数を調べるだけなのに、年別ファイルから逐一データを取り出す必要があり大変手間がかかる。あえて調べようとは思わない。**

上田城跡公園・観光者数推移

Q: 戦国無双 & BASARA、『真田丸』の影響を知りたい

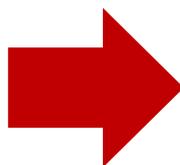
	A	B	C	D	E	F	G	H	I	J	K	L	M
1	◆上田城跡・観光者数推移 2002～2017 (単位:千人)												
2	和暦	年	合計	県内	県外								
3	平成14年	2002年	472.3	267.6	204.7								
4	平成15年	2003年	501.5	277.3	224.2								
5	平成16年	2004年	495.8	274.9	220.9								
6	平成17年	2005年	609.1	350.7	258.4								
7	平成18年	2006年	739.5	432.8	306.7								
8	平成19年	2007年	945.9	554.7	391.2								
9	平成20年	2008年	958.6	571.0	387.6								
10	平成21年	2009年	1423.2	824.7	598.5								
11	平成22年	2010年	1564.7	911.4	653.3								
12	平成23年	2011年	1424.4	812.2	612.2								
13	平成24年	2012年	1531.5	894.0	637.5								
14	平成25年	2013年	1399.8	835.0	564.8								
15	平成26年	2014年	1560.0	920.1	639.9								
16	平成27年	2015年	2151.2	1256.6	894.6								
17	平成28年	2016年	3487.7	1975.4	1512.3								
18	平成29年	2017年	1787.6	1044.2	743.4								
19													
20	参考「長野県の統計情報」掲載の「観光地利用者統計調査」各年データ												
21	https://tokei.pref.nagano.lg.jp/statistics-info/search-result												
22	【オープンデータ宣言】												
23	このデータはCCライセンス4.0 CC-BY-SA(表示・継承)で提供します。												
24													
25													
26													
27													
28													
29													
30													



簡単なことでも調べるのに手間がかかる

オープンデータ化の卑近な課題

- ・ オープンデータデイ2017
 - 上田市統計データを手作業でCSVにし、LinkDataに登録
 - 上田市職員と一緒に課題を検討



- ・ 階層的な見出し行列と結合セルの処理が作業効率を阻む
- ・ 力作業でやれることには限界がある
- ・ これでは本当の問題解決にはつながらない

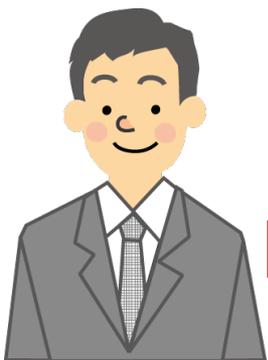
上田市統計担当とゼミで課題検討

- ・ 上田市の統計業務の実情をヒアリングし、解決策を一緒に探る



大きな障壁「紙文化」と「神エクセル」

・ 統計資料作成がアウトプットの業務フロー



印刷物版下作成

地区名	平成29(17)年度現在			平成28(16)年度現在			平成27(15)年度現在		
	世帯数	計	人口	世帯数	計	人口	世帯数	計	人口
計	16,217	119,426	279,011	16,217	119,426	279,011	16,217	119,426	279,011
市	11,240	82,769	193,624	11,240	82,769	193,624	11,240	82,769	193,624
町	5,077	36,657	85,387	5,077	36,657	85,387	5,077	36,657	85,387
村	999	7,000	16,000	999	7,000	16,000	999	7,000	16,000
市	1,000	7,000	16,000	1,000	7,000	16,000	1,000	7,000	16,000
町	1,000	7,000	16,000	1,000	7,000	16,000	1,000	7,000	16,000
村	1,000	7,000	16,000	1,000	7,000	16,000	1,000	7,000	16,000



ついでにネットにも



手作業変換

主にExcel, PDF
CSVはちょっとだけ

問題点

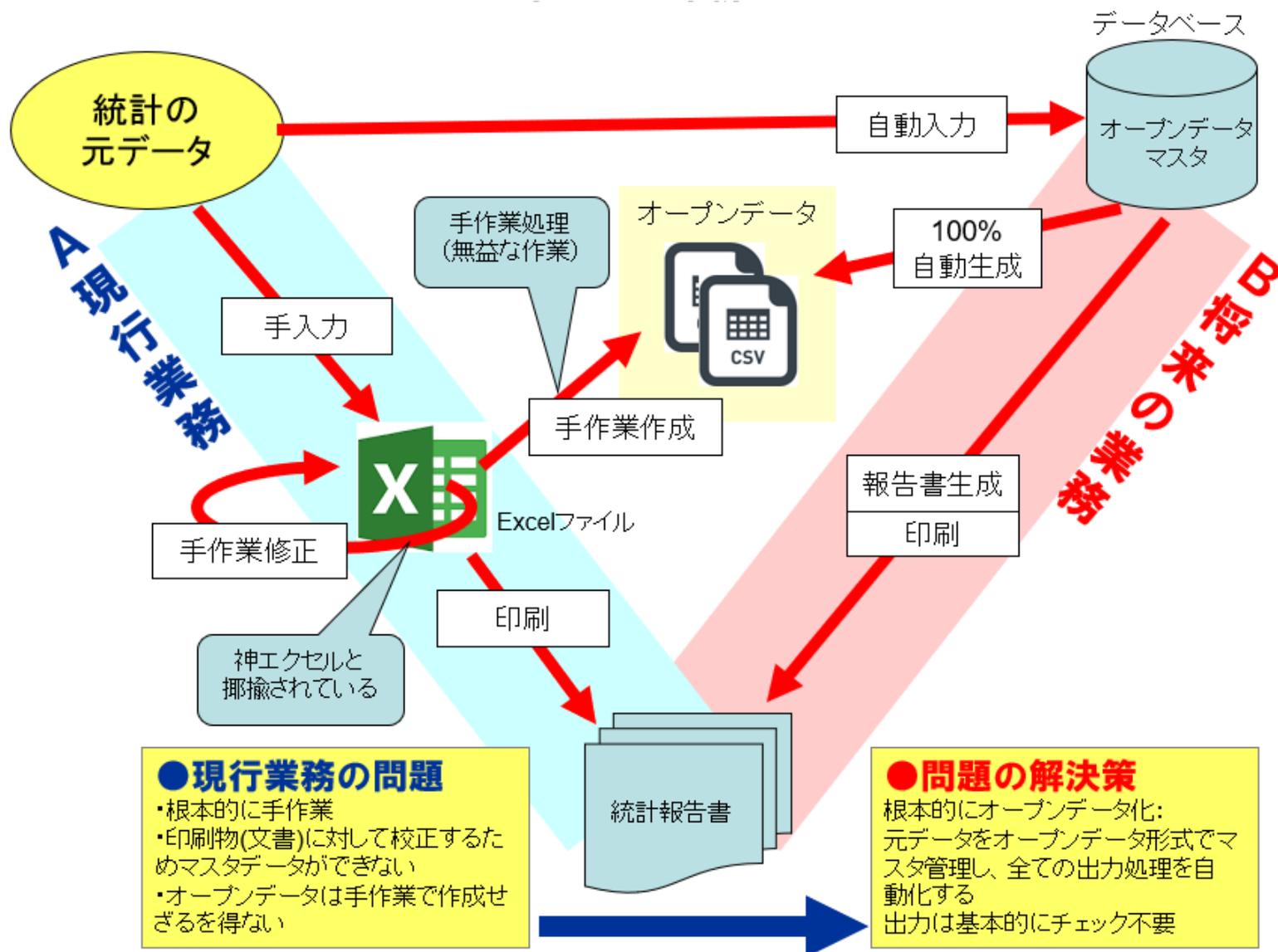
- ① 統計データは印刷のために整形されている
- ② 業務フローにマスタがなくその都度の対応

統計データの実情

- **データの源泉(〇〇市のケース)**
 - 市が独自に作成
 - 総務省統計局からのフィードバック(国政調査等)
 - 県からのデータ
- **統計データのオープンデータ化はデータの源泉(データフロー)ごとに自動化処理されるのが望ましい**

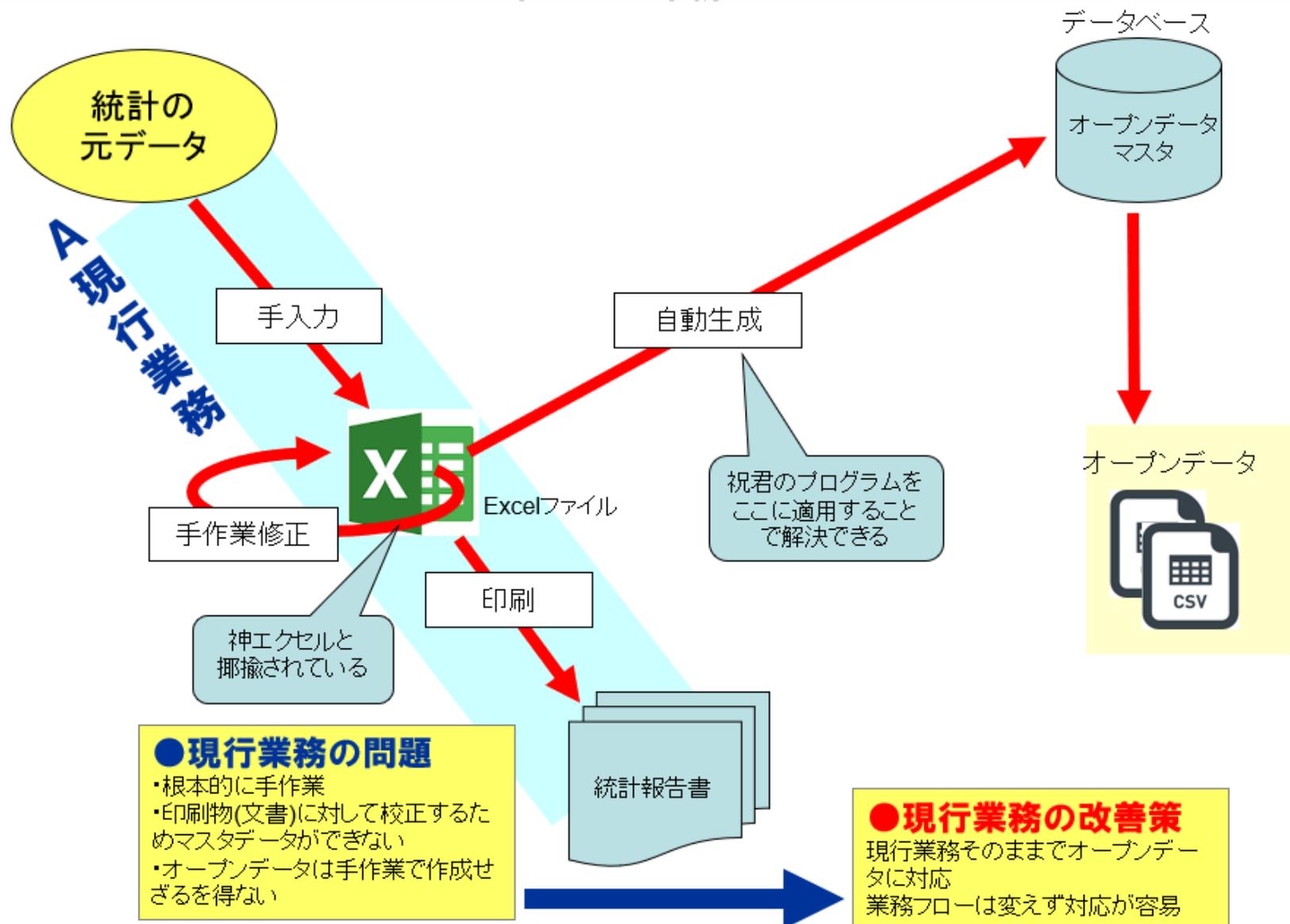
統計業務の改善：A to B

オープンデータ化による業務イノベーション



統計業務の改善：現行業務のままの改善

オープンデータ化による業務イノベーション



人が読みやすいデータ 統計データ

入れ子の表(結合セル)



国勢調査年次別の世帯数及び人口

地区名	大正9年10月1日現在				大正14年10月1日現在				
	世帯数	計	人口 男	人口 女	世帯数	計	人口 男	人口 女	
総数	22,358	112,001	52,896	59,105	23,116	114,756	54,697	60,059	
計	15,093	71,928	34,784	37,144	15,521	73,535	35,802	37,733	
上田	計	9,885	46,533	22,427	24,106	10,446	48,788	23,727	25,061
上田	計	5,656	26,271	12,588	13,683	7,120	32,589	15,675	16,914
上田	城下	817	3,681	1,806	1,875	0	0	0	0
上田	塩尻	619	2,959	1,402	1,557	627	3,006	1,466	1,540
上田	川辺	590	2,851	1,410	1,441	549	2,722	1,370	1,352
上田	泉田	360	1,615	803	812	330	1,439	715	724

機械が読みやすいデータ

#property	西暦	面積(平方キロメートル)	世帯数	人口_総数	人口		
#object_type_xsd	int	float	int	int	int	float	
#property_context	Assertion	Assertion	Assertion	Assertion	Assertion	Assertion	Asse
大正9年	1920	551.09	22358	112001	52896	59105	5.00943733786
大正14年	1925	551.09	23116	114756	54697	60059	4.96435369441
昭和5年	1930	551.09	23598	119271	57097	62174	5.05428426137
昭和1							92024616930
昭和1							95700433784
昭和2							98275684047
昭和2							95498536249
昭和3							82010108998
昭和3							52615717233
昭和40年	1965	551.09	33421	138001	65132	72869	4.12917028215
昭和45年	1970	551.09	36404	139770	66455	73315	3.83941325129
昭和4							5347943908
昭和4							8387755102
昭和5							3901889625
昭和5							1615414941
昭和5							1400599839
昭和5							9169608337
昭和54年	1979	551.09	42125	150701	72952	77749	3.57747181008
昭和55年	1980	551.09	44288	151804	73332	78472	3.42765534682
昭和56年	1981	551.09	44879	152942	73937	79005	3.40787450701

入れ子でない表(セマンティックウェブ)
単純なテキストデータ

テキストなので扱いがとっても簡単
実はExcelで読んでデータ処理しやすい

統計データの大まかな構成パターン

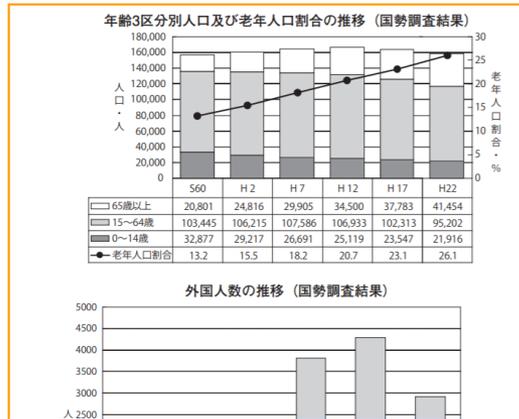
- 大きくは次の3パターンである

表データ

平成18年(2006) 長野県推計人口(毎月人口異動調査)による上田市

年月日	世帯数	人口総数	男性	女性	対前月人口増減
H18.1.1	60,712	163,809	79,815	83,994	36
H18.2.1	60,897	163,713	79,781	83,932	-96
H18.3.1	61,170	163,670	79,754	83,916	-43
H18.4.1	61,809	163,262	79,520	83,742	-408
H18.5.1	61,992	163,341	79,577	83,764	79
H18.6.1	61,973	163,298	79,532	83,766	-43
H18.7.1	61,984	163,285	79,525	83,760	-13
H18.8.1	61,980	163,321	79,581	83,740	36
H18.9.1	62,000	163,337	79,607	83,730	16
H18.10.1	61,958	163,211	79,539	83,672	-126
H18.11.1	62,003	163,283	79,579	83,704	72
H18.12.1	61,961	163,190	79,549	83,641	-93

グラフ混在



説明(テキスト)

- 7 現金給与総額 (統計表冊子 第3表)**
- 現金給与総額は、862億524万円で、前年に比べ26億848万円(前年比3.1%)増加しました。
- また、製造品出荷額等に占める現金給与総額の割合は、15.7%で前年(15.2%)となっています。
- (1) 産業別現金給与総額
- 産業別では、構成比の高い順に、電機 20.2% (174億2,119万円)、輸送機械 15.7% (134億9,250万円)で、これら3産業で全体の45.9%を占めています。
- (2) 従業者規模別現金給与総額及び常用労働者1人当たり現金給与総額
- 従業者規模別では、4~29人が148億6,710万円(構成比17.2%、前年比1.2%増加)が713億3,814万円(構成比82.8%、前年比△3.9%)となっています。
- 常用労働者1人当たりの現金給与総額は443万円で、前年に比べ1.2%の増加となっています。

望ましくはグラフ・表を単体データに分離する

表データの構成パターン

- 大きくは次の4パターンである

A. 単一型 (数値)

平成18年(2006) 長野県推計人口(毎月人口異動調査)による上田市

年月日	世帯数	人口総数	男性	女性	対前月人口増減
H18.1.1	60,712	163,809	79,815	83,994	36
H18.2.1	60,897	163,713	79,781	83,932	-96
H18.3.1	61,170	163,670	79,754	83,916	-43
H18.4.1	61,809	163,262	79,520	83,742	-408
H18.5.1	61,992	163,341	79,577	83,764	79
H18.6.1	61,973	163,298	79,532	83,766	-43
H18.7.1	61,984	163,285	79,525	83,760	-13
H18.8.1	61,980	163,321	79,581	83,740	36
H18.9.1	62,000	163,337	79,607	83,730	16
H18.10.1	61,958	163,211	79,539	83,672	-126
H18.11.1	62,003	163,283	79,579	83,704	72
H18.12.1	61,961	163,190	79,549	83,641	-93

C. イテレート (横反復) 型

年月日	表1	表2	表3	...
H18.1.1				
H18.2.1				
H18.3.1				
H18.4.1				
H18.5.1				
H18.6.1				
H18.7.1				

B. 単一型 (テキスト)

平成18年(2006) 長野県推計人口(毎月人口異動調査)による上田市

年月日	世帯数	人口総数	男性	女性	対前月人口増減
H18.1.1	60,712	163,809	79,815	83,994	36
H18.2.1	60,897	163,713	79,781	83,932	-96
H18.3.1	61,170	163,670	79,754	83,916	-43
H18.4.1	61,809	163,262	79,520	83,742	-408
H18.5.1	61,992	163,341	79,577	83,764	79
H18.6.1	61,973	163,298	79,532	83,766	-43
H18.7.1	61,984	163,285	79,525	83,760	-13

A. B. はともに構造的には同じではあるが、データ行がテキストの場合、見出し列かどうかの識別がしにくい。

D. イテレート (縦反復) 型

平成18年(2006) 長野県推計人口(毎月人口異動調査)による上田市

年月日	表1
H18.1.1	
H18.2.1	
H18.3.1	
H18.4.1	
H18.5.1	
H18.6.1	
H18.7.1	

イテレート型の本来の構造は？

C. イテレート (横反復) 型

平成18年(2006年)	表1	表2	...
年月日			
H18.1			
H18.2			
H18.3			
H18.4			
H18.5			
H18.6			
H18.7			

a. 縦反復型



b. 連続型



印刷物の制約により、横反復するケース

a.縦反復は見出し列に横反復の項目を見出し列に加える形で変換できる

b.連続型はデータ行を単純に単一型の表に変換する

表データの構成パターン

ヘッダ部(タイトル部)

地区名	大正9年10月1日現在				大正14年10月1日現在				
	世帯数	人口			世帯数	人口			
		計	男	女		計	男	女	
		人	人	人		人	人	人	
総数	22,358	112,001	52,896	59,105	23,116	114,756	54,697	60,059	
計	15,093	71,928	34,734	37,194	15,521	73,535	35,862	37,673	
上田	計	9,885	46,533	22,427	24,106	10,446	48,789	23,727	25,061
上田	計	5,656	26,271	12,588	13,683	7,120	32,589	15,675	16,914
上田	東部								
上田	南部								
上田	中央								
上田	北部								
上田	西部								
上田	城下	817	3,817	1,766	2,051			1,466	1,540
上田	塩尻	619	2,619	1,234	1,385			1,370	1,352
上田	川辺	590	2,590	1,234	1,356			1,370	1,352
上田	泉田	360	1,615	803	812	330	1,439	715	724
上田	神科	1,103	5,336	2,584	2,752	1,086	5,312	2,629	2,683
上田	神川	740	3,820	1,834	1,986	734	3,720	1,872	1,848
上田	計	778	3,964	1,939	2,025	768	3,929	1,936	1,993
上田	豊里	391	2,050	1,001	1,049	391	2,026	991	1,035
上田	殿城	387	1,914	938	976	377	1,903	945	958
上田	計	3,100	14,763	7,180	7,583	3,008	14,404	7,043	7,361
上田	東塩田	1,052	5,039	2,435	2,604	993	4,775	2,362	2,413
上田	西塩田	970	4,758	2,304	2,454	945	4,648	2,220	2,428
上田	塩田	733	3,462	1,711	1,751	722	3,496	1,722	1,774

見出し行

データ行

見出し列

フッタ部

Excel→CSV一括変換

統計Excelファイル



✓ jinko_jukijinko_2015.xls

大字別	世帯数	人口	世帯統計	大字
総計	1,198	5,017	1,261	2,000世帯
中央	2,225	4,946	2,211	5,000世帯
北	667	272	801	1,213世帯
北東	400	171	476	1,000世帯
北西	200	87	266	500世帯
東	24,840	104,641	4,001	1,000世帯
南	898	3,156	1,211	2,213世帯
南東	490	200	730	1,213世帯
南西	560	230	700	1,213世帯
西	790	341	1,213世帯	1,213世帯
東	280	285	321	1,213世帯
南	1,198	5,017	1,261	1,213世帯
北	266	104	316	1,213世帯
南	701	285	341	1,213世帯

複数ワークシート

Excel→CSV
一括変換



✓ jinko_jukijinko_2015

CSVファイル (ワークシート別)

_2015.1.csv



✓ 大字別、人口、世帯統計_2015.5.csv



✓ 大字別、人口、世帯統計_2015.9.csv



✓ 大字別、人口、世帯統計_2015.2.csv



✓ 大字別、人口、世帯統計_2015.6.csv



✓ 大字別、人口、世帯統計_2015.10.csv



✓ 大字別、人口、世帯統計_2015.3.csv



✓ 大字別、人口、世帯統計_2015.7.csv



✓ 大字別、人口、世帯統計_2015.11.csv



✓ 大字別、人口、世帯統計_2015.4.csv



✓ 大字別、人口、世帯統計_2015.8.csv



✓ 大字別、人口、世帯統計_2015.12.csv

オープンデータ化の2ケース

統計データ



手作業変換

大変



ほんのちょっと
a little

徒労？
学習と思えば意義がある

▼大幅な作業効率化を図り、担当者は本来の業務に専念する

統計データ



自動変換

どっさり
many many



常時公開

ネットで
閲覧・活用



地域・全国・世界の人へ

統計のオープンデータ化の進む方向

CSVへの手作業変換
ODとは何ぞやを学習

CSVの一括自動変換
当面のニーズに応える

データベース化によるデ
ータ活用への踏み出し

統合化されたデジタルコ
モンズでの横断的活用

2018年度

2019年度

2020年度

統計オープンデータの意義

・ 民間・企業とのコラボ

経済誘導

・ 統計業務の手間を減らす

事務合理化

・ 見て面白い・発見
interestの誘発

教育

・ データを収集分析・
シミュレーション

政策立案

オープンデータ化：当面の進め方

市町村ホームページ
Excelファイル



CC-BY

自動変換



一括登録

CC-BY

DB
マスターデータ

常時公開

ネットで
活用



地域・全国・世界の人へ

いくつかのデータを
試験的に変換

ツールの改善

- ・一括選択
- ・対応パターンの拡充

「信州デジタルコモンズ」
サイトに寄せ換え

多数のデータを一覧化

民間・大学等の力を活かして実のある取り組みを！