

「ネ申 Excel」問題

奥村 晴彦*

2013年7月12日

概要

Excel に代表される表計算ソフトは、簡便なデータ入力・解析・可視化ツールであるが、これを柔軟な罫線の引ける DTP ソフトとして用い、データとしての再利用の困難な複雑な帳票を作成してしまうことがよくある。この類の「ネ申 Excel」(神 Excel) が引き起こす問題点をまとめ、解決策を提案する。

Hilarious uses of Excel

Haruhiko Okumura*

Spreadsheet software, notably Excel, is widely used in Japan to create paperwork with complex ruled tables, sacrificing reusability of the data. This practice is analyzed with a number of examples.

1 はじめに

著者ら [1] は、2011 年の東日本大震災後に見られた情報リテラシーにかかわる問題点を論じた際に、公開されたデータの機械可読性の低さを指摘した。それらのデータは、Excel に入力されているが、公開情報が PDF だけであったり、時にはデータ抽出禁止の設定がなされていたり画像化されていたりして、集約や再解析にたいへん手間がかかった。

しかし、震災データに限らず、政府統計の総合窓口 e-Stat [3] や自治体などで公開されているデータを調べるにつれて、そもそも Excel への入力の段階でデータの再利用を考えず、罫線を多用した紙の帳票作成を最終目的とするものが多いことに気づいた。こうしたデータは、PDF でなく Excel ファイル (またはそれから生成した CSV ファイル) で公開されたとしても、データを再解析するには余計な手間がかかる。

こうした巧妙だがデータとしての再利用性の低い

Excel ファイルが自治体等で多用されていることに言及した井ノ口 [4] は、このような Excel ファイルを「神的 Excel」と呼んだ。本稿ではすでにネットで一部に浸透している「^{かみ}ネ申 Excel」[5, 6] という呼び方を使うことにする。ここで「ネ申」は「神」の隠語であるが、「紙 (への出力しか考えていない) Excel」の意味も含んでいる。

なお、セルを小さい正方形 (後述の Excel 方眼紙) にして描く「ドット絵」や、「図形」(Excel 2003 までは「オートシェイプ」) 機能での描画 (例: [7]) など、明らかにデータ処理から外れた Excel の利用法については、本稿では扱わない。

また、本稿はあくまで Excel に代表される表計算ソフトの使い方の話である。Excel の問題や、Excel で描いた統計グラフの問題については、別稿 [2] を参照されたい。

2 罫線ツールとしての Excel

日本独特の帳票には縦横の罫線を多用するという特徴がある。特に自由に罫線を引くためにあらかじめセルを正方形に設定しておくことがあり (例:

* 三重大学 (Mie University)

[8, 9]), Excel 方眼紙と挿入される。マイクロソフトのサイトからも方眼紙テンプレートがダウンロードできる [10]。Excel 方眼紙の実例を図 1, 2 に挙げる。

図 1 Excel 方眼紙の例: 「平成 24 年度科学研究費助成事業 実績報告書 (研究実績報告書)」 https://www.jsps.go.jp/j-grantsinaid/16_rule/index.html 様式 Z-7

図 2 Excel 方眼紙の例: 法務省「登記事項証明書」, 文字はすべてテキストボックスになっている。 <http://www.moj.go.jp/ONLINE/COMMERCE/11-2.html>

Excel 方眼紙の方眼に 1 文字ずつ入力したものを Excel 原稿用紙とも呼ぶ。次の例 (図 3) でも一部でこの技法が使われている。

また, うまく 1 枚の紙に収まるように, 複数の表を 1 シートにまとめたり, 逆に一つの表を複数に分割したりすることもよく行われている (例: 図 4)。

区 分	給 与 所 得 者 数				
	3月末	6月末	9月末	12月末	年間月平均
	千人	千人	千人	千人	千人
平成18年度	54,752	55,013	54,065	53,396	54,506
19	54,957	55,311	54,417	53,768	54,817
20	54,672	55,269	55,000	54,739	55,124
21	56,232	54,909	54,035	53,884	54,967
22	55,817	54,887	54,367	54,153	54,792
23	55,169	54,688	54,459	54,273	54,647
人	人	人	人	人	人
10人未満	10,010,414	10,022,094	9,900,419	9,627,198	9,890,052
10人以上	8,333,724	7,886,247	7,863,073	7,735,850	7,954,723
30人 "	8,725,936	8,668,139	8,630,664	8,653,029	8,669,395
100人 "	11,820,473	11,718,332	11,658,236	11,727,615	11,731,189
500人 "	4,039,942	4,061,021	4,072,913	4,081,849	4,063,946
1,000人 "	7,178,060	7,244,851	7,248,209	7,290,875	7,240,466
5,000人 "	5,060,628	5,086,941	5,085,593	5,156,124	5,097,320
計	36,825,039	36,779,284	36,695,615	36,909,492	36,802,316
合 計	55,169,177	54,687,625	54,459,107	54,272,540	54,647,091

図 3 「事業所別規模」が 1 文字 1 セルに入っている。e-Stat「民間給与と実態統計>結果表>年次>2011 年」の最初の Excel ファイル。

平成25年4月1日現在 住民基本台帳による年齢別・男女別人口																
年齢別・男女別人口	前年比				平成25年4月1日				平成24年4月1日				各年齢別・男女別人口			
	年齢	計	男	女	年齢	計	男	女	年齢	計	男	女	年齢	計	男	女
0	1,647	792	855	0	1,647	792	855	0	1,647	792	855	0	1,647	792	855	
1	1,702	870	832	1	1,702	870	832	1	1,702	870	832	1	1,702	870	832	
2	1,727	883	844	2	1,727	883	844	2	1,727	883	844	2	1,727	883	844	
3	1,791	897	894	3	1,791	897	894	3	1,791	897	894	3	1,791	897	894	
4	1,640	831	809	4	1,640	831	809	4	1,640	831	809	4	1,640	831	809	
5	1,613	853	760	5	1,613	853	760	5	1,613	853	760	5	1,613	853	760	
6	1,573	830	743	6	1,573	830	743	6	1,573	830	743	6	1,573	830	743	
7	1,514	779	735	7	1,514	779	735	7	1,514	779	735	7	1,514	779	735	
8	1,522	798	724	8	1,522	798	724	8	1,522	798	724	8	1,522	798	724	
9	1,549	797	752	9	1,549	797	752	9	1,549	797	752	9	1,549	797	752	
10	1,481	748	733	10	1,481	748	733	10	1,481	748	733	10	1,481	748	733	
11	1,443	731	712	11	1,443	731	712	11	1,443	731	712	11	1,443	731	712	
12	1,487	754	733	12	1,487	754	733	12	1,487	754	733	12	1,487	754	733	
13	1,362	702	660	13	1,362	702	660	13	1,362	702	660	13	1,362	702	660	
14	1,484	717	767	14	1,484	717	767	14	1,484	717	767	14	1,484	717	767	
15	1,422	725	697	15	1,422	725	697	15	1,422	725	697	15	1,422	725	697	
16	1,372	706	666	16	1,372	706	666	16	1,372	706	666	16	1,372	706	666	
17	1,424	709	715	17	1,424	709	715	17	1,424	709	715	17	1,424	709	715	
18	1,362	702	660	18	1,362	702	660	18	1,362	702	660	18	1,362	702	660	
19	1,391	744	647	19	1,391	744	647	19	1,391	744	647	19	1,391	744	647	
20	1,382	693	689	20	1,382	693	689	20	1,382	693	689	20	1,382	693	689	
21	1,540	820	720	21	1,540	820	720	21	1,540	820	720	21	1,540	820	720	
22	1,474	779	695	22	1,474	779	695	22	1,474	779	695	22	1,474	779	695	
23	1,869	794	775	23	1,869	794	775	23	1,869	794	775	23	1,869	794	775	
24	1,619	847	772	24	1,619	847	772	24	1,619	847	772	24	1,619	847	772	
25	1,607	808	799	25	1,607	808	799	25	1,607	808	799	25	1,607	808	799	
計	168,024	83,311	84,713	総数	168,195	82,163	86,032									

図 4 二つの表を一つのシートに入力し, しかも 2 番目の表が長いので二つに分割した例。流山市 <http://www.city.nagareyama.chiba.jp/10763/011144.html>

3 情報集約に使われる Excel

Excel ファイルをメール添付で提出するアンケートの類が今でも少なからず存在する。入力欄以外をロックし, 選択肢もリストから選ぶように設定してあれば, 自動集計も可能であるが, 多くは回収した時点で紙に印刷して人力で集計しているようである。

人力で集計している証拠として, 「いずれかを〇で囲む」という指示がよくある (例: 図 5)。回答に

は、「挿入」→「図形」で「円/楕円」を選んで該当部分の対角線をマウスで指定し、「塗りつぶしなし」にして、「図形の枠線」を適当な色にする。「配置」→「グリッドに合わせる」を設定しておけば、レイアウトが崩れて図形が移動することを防げる。

A	B	C	D	E	F	G	H	I	J	K	L	M	N
※VDT作業(パソコン等を使用し、データ入力、検索、文書等作成、プログラミング等を行う作業)についてお答えください。													
5	フリガナ			性別		生年月日		年	月	日	部署名		
6	氏名			男・女							職名		
8	職務内容(該当箇所○をつける)	1. 医療事務	2. 医療技術	3. 図書業務	4. 情報処理・技術	5. 一般事務	6. 教員	7. その他()					
10	VDT作業の有無		有	無	【有】の方の○以下の質問にお答えください。								
11	肩	責任が高いと診断されたことがありますか		有	無	VDT作業歴		年 月 ~ 年 月					
12	腕	肘内縁と診断されたことがありますか		有	無								
13	手	指の付け根に痛みを感じますか		有	無								
15	VDT作業の状況	(注1) VDT作業区分(該当箇所○をつける)	A		(注2) 作業の種類(該当箇所○をつける)	1. 単純入力型							
16		B		2. 拘束型									
17		C		3. 対話型									
18		1日の平均VDT従事時間数	時間	分	4. 技術型								
19	1日の平均VDT従事時間数	時間	分	5. 監視型									
20	1週間の平均VDT従事日数	日		6. その他の型									
21	項目	有無・割合	項目	有無・割合									
22	目が疲れる	有 無	肩が重たい(痛い)	有 無									
23	目が痛い	有 無	腕や指がこる	有 無									
24	目がかすむ	有 無	手指の力が人引にくい	有 無									
25	目が乾く	有 無	手指がはげれる(痛い)	有 無									
26	作業疲れ	有 無	腕がだるい(痛い)	有 無									

図5 「図形」で○を付けることを想定したExcelアンケート

会議などのスケジュールを調整するためにも頻りにExcelが使われる。複数の委員会等に属していれば、同じ情報を微妙に様式の異なる複数のExcelシートに入力することを強いられる。

報告書や論文リストもExcelで提出することが多い。論文リストの本人の名前にアンダーラインを付けるなど、単純なテキストの流し込みではできないルールが課されることもよくある。同じ内容をあちこちに入力しなければならないことも多く、現場の生産性を低下させている。

4 数値を文字列として入力したもの

2011年3月の東京電力福島第一原子力発電所の事故で、東京電力は大量の放射線データを公表したが、その値は数値としてではなく単位付きの文字列としてExcelに入力されたものであった。そのため、数値として解釈できないものを入力してもフィードバックが得られない。実際、ある個所では

「0」が「お」に変換されていた(図6)。他にも、全角・半角の混在、午前・午後の間違い、単位の不揃い(「μ」の抜け)などが見られた。

【別紙】福島第一原子力発電所モニタリングカーによる計測状況													
計測日: 3月19日													
計測時間	計測場所	γ線	中性子線	風向	風速(m/s)								
午前7時30分	西門	288.9 μSv/h	0.01 μSv/h未満	西	0.8								
午前7時20分	西門	289.1 μSv/h	0.01 μSv/h未満	西	0.7								
午前7時10分	西門	289.8 μSv/h	0.01 μSv/h未満	西	0.3								
午前7時00分	西門	290.6 μSv/h	0.01 μSv/h未満	北西	0.7								
午前6時50分	西門	290.9 μSv/h	0.01 μSv/h未満	東	0.4								
午前6時40分	西門	291.5 μSv/h	0.01 μSv/h未満	南南東	0.3								
午前6時30分	西門	292.3 μSv/h	0.01 μSv/h未満	東南東	0.4								
午前6時20分	西門	292.6 μSv/h	0.01 μSv/h未満	東北東	0.5								
午前6時10分	西門	293.6 μSv/h	0.01 μSv/h未満	南東	0.7								
午前6時00分	西門	293.8 μSv/h	0.01 μSv/h未満	西	0.9								
午前5時50分	西門	294.3 μSv/h	0.01 μSv/h未満	南東	0.6								
午前5時40分	西門	295.4 μSv/h	0.01 μSv/h未満	東	0.9								
午前5時30分	西門	295.1 μSv/h	0.01 μSv/h未満	北西	0.9								
午前5時20分	西門	295.8 μSv/h	0.01 μSv/h未満	東	0.9								
午前5時10分	西門	296.4 μSv/h	0.01 μSv/h未満	南	0.5								
午前5時00分	西門	297.5 μSv/h	0.01 μSv/h未満	南	0.5								
午前4時50分	西門	298.5 μSv/h	0.01 μSv/h未満	南	0.4								
午前4時40分	西門	299.2 μSv/h	0.01 μSv/h未満	西	0.6								
午前4時30分	西門	299.2 μSv/h	0.01 μSv/h未満	南東	0.6								
午前4時20分	西門	300.5 μSv/h	0.01 μSv/h未満	北北西	0.4								
午前4時10分	西門	301.3 μSv/h	0.01 μSv/h未満	西	0.5								
午前4時00分	西門	301.7 μSv/h	0.01 μSv/h未満	東	0.6								
午前3時50分	西門	303.1 μSv/h	0.01 μSv/h未満	北北西	0.6								
午前3時40分	西門	303.6 μSv/h	0.01 μSv/h未満	南南東	0.9								
午前3時30分	西門	304.3 μSv/h	0.01 μSv/h未満	東	0.7								
午前3時20分	西門	305.1 μSv/h	0.01 μSv/h未満	南南東	0.7								
午前3時10分	西門	306.0 μSv/h	0.01 μSv/h未満	南西	0.7								
午前3時00分	西門	306.9 μSv/h	0.01 μSv/h未満	北北西	0.6								
午前2時50分	西門	308.6 μSv/h	0.01 μSv/h未満	北北西	0.4								
午前2時40分	西門	309.1 μSv/h	0.01 μSv/h未満	西南西	0.3								
午前2時30分	西門	310.0 μSv/h	0.01 μSv/h未満	西南西	0.6								
午前2時20分	西門	311.1 μSv/h	0.01 μSv/h未満	北	0.3								
午前2時10分	西門	312.2 μSv/h	0.01 μSv/h未満	南	0.3								
午前2時00分	西門	313.7 μSv/h	0.01 μSv/h未満	北	3.0								
午前1時50分	事務本館北	3181.0 μSv/h	0.01 μSv/h未満	西南西	1.3								
午前1時40分	事務本館北	3188.0 μSv/h	0.01 μSv/h未満	南西	1.6								
午前1時30分	事務本館北	3195.0 μSv/h	0.01 μSv/h未満	西南西	1.4								
午前1時20分	事務本館北	3247.0 μSv/h	0.01 μSv/h未満	西南西	1.3								
午前1時10分	事務本館北	3279.0 μSv/h	0.01 μSv/h未満	西南西	1.3								
午前1時00分	事務本館北	3248.0 μSv/h	0.01 μSv/h未満	西南西	1.3								
午前0時50分	事務本館北	3284.0 μSv/h	0.01 μSv/h未満	西	1.4								
午前0時40分	事務本館北	3342.0 μSv/h	0.01 μSv/h未満	西南西	0.9								
午前0時30分	事務本館北	3231.0 μSv/h	0.01 μSv/h未満	南西	1.1								
午前0時20分	事務本館北	3219.0 μSv/h	0.01 μSv/h未満	南西	1.2								
午前0時10分	事務本館北	3224.0 μSv/h	0.01 μSv/h未満	西南西	1.4								
午前0時00分	事務本館北	3229.0 μSv/h	0.01 μSv/h未満	西南西	1.4								

図6 東京電力が2011年3月19日に測定し同日公表した放射線量。午前3時20分のγ線の値が「3お5.1 μSv/h」になっている。現在入手できる修正版 http://www.tepco.co.jp/cc/press/betu11_j/images/110528d.pdf では「305.1 μSv/h」に訂正されている。

この表をPDFからExcelにコピーし、「μSv/h」を一括削除すると、図7のようになり、「3お5.1」以外にも4カ所、数値でないものが入っていたことがわかる(小数点がコンマになっている)。最初からこのように入力していれば、わざわざ「μSv/h」を入力する手間が不要になるだけでなく、入力と同時にこのページだけで5カ所の入力間違いに気がついたらはずである。

「午前7時30分」等々も文字列として入力したと思われる。日付は「3月19日」としか書かれていないが、何年も残す文書であるから、「2011年」を省

23	2011-3-19 3:10	西門		306
24	2011-3-19 3:20	西門	3お5.1	
25	2011-3-19 3:30	西門		304.3
26	2011-3-19 3:40	西門		303.6
27	2011-3-19 3:50	西門		303.1
28	2011-3-19 4:00	西門		301.7
29	2011-3-19 4:10	西門		301.3
30	2011-3-19 4:20	西門		300.5
31	2011-3-19 4:30	西門		299.2
32	2011-3-19 4:40	西門		299.2
33	2011-3-19 4:50	西門	298.5	
34	2011-3-19 5:00	西門		297.5
35	2011-3-19 5:10	西門		296.4
36	2011-3-19 5:20	西門	295.8	
37	2011-3-19 5:30	西門	295.1	
38	2011-3-19 5:40	西門	295.4	
39	2011-3-19 5:50	西門		294.3

図7 図6を時刻順に並べ替え、日時を省略せず入れ、「 $\mu\text{Sv/h}$ 」を一括削除したもの。左揃えになったセルは数値でないものが入っている。

略してはならない。さらに、すべての行に「年」から始まる日時を入れれば、機械可読性が良くなるだけでなく、入力も簡単になる。また、時系列データの一般的な入力ルールに従って、日時の昇順に並べるべきである。上の図7では、逆順に並べ替え、先頭に「2011-3-11 0:00」、次に「2011-3-11 0:10」と入れて両者を選択し、フィルハンドルをドラッグすることにより、10分刻みの日時を入力した。こうすれば午前・午後の間違いもない。ただし、Excelの内部形式の関係で、ある程度以上長くフィルすると、誤差が出る。

5 セル結合

Excelのセル結合は、項目の入れ子関係を表すためや、不定個の値を表すため、あるいは単に見栄えのために、多用されている。東日本大震災後のデータでも、2種類の放射性物質 (^{137}Cs , ^{134}Cs) の量を個別に表示したものと、合わせて表示したものが混在したデータを表すため、合わせて表示した場合にはセルを結合することがよくあった(例:図8)。

日時は、日と時が入れ子関係にあるので、日をセル結合することがよくあった(例:図9)。これも機械処理には向かない。

また、特にセルを結合するわけではないが、セル

検査法 (Ge/NaI)	採取日 (購入日)	結果 判明日	結果(Bq/kg)		
			ヨウ素-131	セシウム-134	セシウム-137
Ge	H24.3.22	H24.3.22	< 5	< 5	< 5
Ge	H24.3.22	H24.3.22	< 5	< 5	< 5
Ge	H24.3.21	H24.3.21	< 5	< 5	< 5
Ge	H24.3.21	H24.3.21	< 5	< 5	< 5
Ge	H24.3.14	H24.3.22	< 12	105	161
Ge	H24.3.14	H24.3.22	< 10	67	137
Ge	H24.3.14	H24.3.22	< 9	31	172
Ge	H24.3.21	H24.3.22	< 4.7	< 4.9	< 4.9
Ge	H24.3.21	H24.3.22	< 4.1	< 4.3	< 5.0
Ge	H24.3.21	H24.3.22	< 3.9	< 4.6	< 4.5
Ge	H24.3.21	H24.3.22	< 1.9	< 1.8	< 1.4
Ge	H24.3.22	H24.3.22	< 5	< 5	
Ge	H24.3.21	H24.3.21	< 2	< 3	< 2
NaI	H24.3.19	H24.3.20	-	< 25	
NaI	H24.3.19	H24.3.20	-	< 25	
NaI	H24.3.19	H24.3.20	-	< 25	

図8 セルを結合した表の例:厚生労働省「食品中の放射性物質の検査結果について(第350報)」(2012年3月22日)。なお、この表では「<5」(検出限界5Bq/kgで不検出)のように数値に不等号を付けた文字列が多用されている。

結合と同様な見栄えに仕上げた表もよくある(例:図10)。

これに類した表は、*Bad Data Handbook* [12]の、特にPaul Murrellによる第3章“Data Intended for Human Consumption, Not Machine Consumption”にも紹介されている。

6 どうすればよいか

表計算ソフトで機械可読なデータを作成するには、セル結合・罫線・文字飾りなどの視覚的効果にデータとしての意味を持たせず、一定の方式(通常は1行目に項目名、2行目以降にデータ)で入力する。その際に、数値データは(単位を付けず)数値として、日時データは日時として解釈できる文字列(例えば2013/07/01 12:34:56)として入力する。保存は、CSVなどの軽いファイル形式にするのが

○平成23年

月日	測定時刻	測定値	月日	測定時刻	測定値	月日	測定時刻	測定値
3月13日 (日)	0:00	0.12	3月15日 (火)	0:00	0.06	3月17日 (木)	0:00	3.71
	1:00	0.12		1:00	0.06		1:00	5.76
	2:00	0.11		2:00	0.06		2:00	5.71
	3:00	0.11		3:00	0.06		3:00	5.63
	4:00	0.11		4:00	0.24		4:00	5.59
	5:00	0.06		5:00	0.06		5:00	5.56
	6:00	0.06		6:00	0.06		6:00	5.54
	7:00	0.06		7:00	0.06		7:00	5.36
	8:00	0.06		8:00	0.06		8:00	5.66
	9:00	0.06		9:00	0.24		9:00	5.69
	10:00	0.06		10:00	14.71		10:00	5.68
	11:00	0.06		11:00	31.82		11:00	5.66
	12:00	0.06		12:00	13.11		12:00	5.61
13:00	0.06	13:00	10.19	13:00	5.51			
14:00	0.06	14:00	9.80	14:00	5.51			
15:00	0.06	15:00	9.38	15:00	5.46			
16:00	0.06	16:00	8.09	16:00	5.52			
17:00	0.06	17:00	7.47	17:00	5.35			
18:00	0.06	18:00	8.53	18:00	5.22			
19:00	0.06	19:00	8.86	19:00	5.08			
20:00	0.06	20:00	8.40	20:00	5.06			
21:00	0.06	21:00	8.27	21:00	4.99			
22:00	0.06	22:00	7.99	22:00	4.90			
23:00	0.06	23:00	6.71	23:00	4.86			
3月14日 (月)	0:00	0.06	3月16日	0:00	6.40	3月18日 (金)	0:00	4.88
	1:00	0.06		1:00	6.37		1:00	4.88
	2:00	0.06		2:00	10.29		2:00	4.85
	3:00	0.06		3:00	9.09		3:00	4.77
	4:00	0.06		4:00	8.15		4:00	4.76
	5:00	0.06		5:00	7.73		5:00	4.77
	6:00	0.06		6:00	7.36		6:00	4.69
	7:00	0.06		7:00	7.11		7:00	4.56
	8:00	0.06		8:00	6.94		8:00	4.61
	9:00	0.06		9:00	6.71		9:00	4.60
	10:00	0.06		10:00	6.46		10:00	4.65
	11:00	0.06		11:00	6.22		11:00	4.62
	12:00	0.06		12:00	6.20		12:00	4.61
13:00	0.06	13:00	5.85	13:00	4.71			
14:00	0.06	14:00	5.62	14:00	4.66			
15:00	0.06	15:00	6.01	15:00	4.65			
16:00	0.06	16:00	6.09	16:00	4.65			
17:00	0.06	17:00	6.03	17:00	4.67			
18:00	0.06	18:00	6.03	18:00	4.66			
19:00	0.06	19:00	5.98	19:00	4.55			
20:00	0.06	20:00	5.88	20:00	4.47			
21:00	0.06	21:00	5.86	21:00	4.42			
22:00	0.06	22:00	5.80	22:00	4.35			
23:00	0.06	23:00	5.78	23:00	4.29			

図9 セルを結合した表の例：福島県「可搬型モニタリングポスト未公表データ一覧表について」<http://www.cms.pref.fukushima.jp/download/1/jikoshokinokahanMPdate.pdf> ちなみに未公表部分は背景色を変えてある。背景色やフォントに意味を持たせた表は機械可読性が劣る。

	A	B	C	D	E	F	G
1	[基本集計]				長期時系列表1 (1) 労働力人口 - 全国、月別結果		
2	[Basic Tabulation]				Historical data 1 (1) Labour force - Whole Japan, Monthly Data		
3					(万人)		
4					(in thousand persons)		
5					原数値 (2010年国勢調査基準切り替え以前の既公表値)		
6					Original series (initially released data before 2010-Census base revision)		
7					男女計	男	女
8		年 月			Both sexes	Male	Female
9		Year and month					
10							
12	平成17年	10月	Oct.		6713	3930	2783
13	2005	11月	Nov.		6636	3901	2736
14		12月	Dec.		6580	3881	2699
15	平成18年	1月	Jan.		6561	3864	2697
16	2006	2月	Feb.		6549	3855	2694
17		3月	Mar.		6597	3887	2710
18		4月	Apr.		6652	3901	2751
19		5月	May		6725	3925	2799
20		6月	June		6717	3924	2793
21		7月	July		6688	3908	2780
22		8月	Aug.		6699	3912	2788
23		9月	Sept.		6711	3917	2794
24		10月	Oct.		6718	3921	2797
25		11月	Nov.		6669	3892	2777
26		12月	Dec.		6598	3873	2725
27	平成19年	1月	Jan.		6542	3861	2681
28	2007	2月	Feb.		6572	3878	2694
29		3月	Mar.		6632	3898	2733
30		4月	Apr.		6712	3925	2788
31		5月	May		6757	3942	2814

図10 暗にセルを結合した表

望ましい*1 が、Excel 形式 (特に *.xlsx) で保存しても、オープンソースのライブラリ (例えば R の XLConnect [13]) を利用して、CSV と同様に読めることが多い。以上の作業に、技術的に難しい箇所はまったくなく、むしろ作業時間は減るはずである。

いったん機械可読な形式で保存できれば、同じデータを人間が何度も再入力しなくても、さまざまな形での解析が可能になる。このことは、情報教育の中で何らかの自動処理 [14] を体験していれば、想像に難くない。

高校「情報」や大学の一般情報教育では、たとえ Excel を扱うにしても、単に「紙」をコンピュータで再現する方法を学ぶのでなく、「データ・リテラシー」を学ぶことが大切であろう。

一方、「紙」文化圏の大人にとって、「データ」文化への切り替えは容易ではない。オープンデータの追い風をうまく利用して、文化を変えるように運動するしかない。折しも、本稿執筆の最終段階で、内閣官房からオープンデータに関するパブリックコメント [15] が出た。特に資料中の「数値 (表)、文章、地理空間情報のデータ作成に当たっての留意事項 (案)*2 の「1. 数値 (表形式) データの作成に当たっての留意事項」は、その大部分が本稿の趣旨と合致するものである。

ソフトウェアの側にも工夫の余地がある。文書の内容と見せ方を分離する HTML + CSS や LaTeX と同様に、データとその見せ方を分離するという発想のソフトが期待される。例えば Apple の Numbers は複数の表を 1 枚のシートにまとめ、個々の表のスタイルを簡単に指定できる。CSV 形式で保存する際にも一つのフォルダに入った複数の CSV ファイルの形で保存できる。ただ、これを使えばクリーンなデータが自動的にできるわけではない。

また、少し異なる問題として、東日本大震災後の自治体などには、せっかくデータをオープンにして

*1 ただし Excel で CSV を読み込む際に問題が生じることがある。例えば 1-2-3 は日付 2001 年 2 月 3 日に、OCT1 という遺伝子名は 10 月 1 日に変換される。

*2 この「留意事項 (案)」は、オープンデータコンソーシアムの提案 [16] に拠っているようである。

も、古くなったデータは不要なので削除するという行動様式も見られた。これはデータ文化にも関連するが「アーカイブズ」の文化が欠けているのが原因と思われる。東日本大震災アーカイブズの類がいろいろ構築されているが、どんどん失われている数値データのアーカイブズを作るという観点は未だに欠けたままである（伊藤 [17] 参照）。

謝辞

たくさんの有用なご指摘をいただいた査読者の方々に感謝します。

参考文献

- [1] 奥村晴彦, 辰己丈夫, 藤間 真「大震災で見えてきた情報教育の課題」情報処理学会情報教育シンポジウム SSS2011 論文集, pp. 25–32 (2011年 8月)
- [2] 奥村晴彦「情報教育と統計」情報処理学会研究報告「コンピュータと教育」2008-CE-97 (情処技報 Vol. 2008, No. 128), pp. 81–88 (2008年)
- [3] 政府統計の総合窓口 <http://www.e-stat.go.jp/>
- [4] 井ノ口宗成「ウェブサービス型「岩手県被災者台帳システム」を用いた被災者生活再建支援」情報処理学会第 75 回全国大会, 2013 年 3 月 8 日
- [5] 「公務員が公開するネ申 Excel が日本の生産性を落としている話」<http://togetter.com/li/469017>
- [6] 「Excel をワープロ代わりに使用することが日本の生産性を落としている？ 考察まとめ」<http://togetter.com/li/469143>
- [7] 「エクセルで本格的な絵画を描く 油絵や日本画の画家が「仲間」」『日経パソコン』2013 年 3 月 11 日号 pp. 8–9, <http://pc.nikkeibp.co.jp/article/interview/20130319/1083744/>
- [8] 日経ウーマンオンライン「美文書作りには「方眼紙」シートを使う」<http://wol.nikkeibp.co.jp/article/column/20130523/153361/>
- [9] NAVER まとめ「日経ウーマンが美文書「Excel 方眼紙」特集、古傷を刺激された皆さんの叫び声が響きわたる」<http://matome.naver.jp/odai/2136948486965928001>
- [10] Office.com 「方眼紙 - テンプレート」<http://office.microsoft.com/ja-jp/templates/TC001018375.aspx>
- [11] 「Excel 方眼紙 Advent Calendar 2013」<http://atnd.org/events/38387>
- [12] Q. Ethan McCallum, *Bad Data Handbook*, O'Reilly, 2012.
- [13] CRAN - Package XLConnect, <http://cran.r-project.org/web/packages/XLConnect/>
- [14] 情報処理学会情報処理教育委員会「日本の情報教育・情報処理教育に関する提言 2005 (2006.11 改訂/追補版)」<http://www.ipsj.or.jp/12kyoiku/teigen/v81teigen-rev1a.html>
- [15] 内閣官房 IT 担当室「電子行政オープンデータ推進のためのロードマップ (案)」及び「二次利用の促進のための府省のデータ公開に関する基本的考え方 (ガイドライン) (案)」に関するパブリックコメントの募集について (2013 年 5 月 24 日) <http://www.kantei.go.jp/jp/singi/it2/info/h250524.html>
- [16] オープンデータ流通推進コンソーシアム「技術委員会 資料・議事録」<http://www.opendata.gr.jp/committee/technical/documents.php>
- [17] 伊藤好孝「福島放射線測定データの現状とメタデータベース作り」日本アーカイブズ学会 2013 年度大会 <http://www.slideshare.net/yoshitakaitow/130421kokai-19477468>