

การสร้างชุดข้อมูลเปิดสำหรับแพลตฟอร์ม Open-D (Open Data Publishing Guideline for Open-D Platform)

ดร.มารุต บุรณรัช

ห้องปฏิบัติการวิจัยเทคโนโลยีภาษาธรรมชาติและความหมาย
ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ

marut.bur@nectec.or.th

หัวข้อนำเสนอ

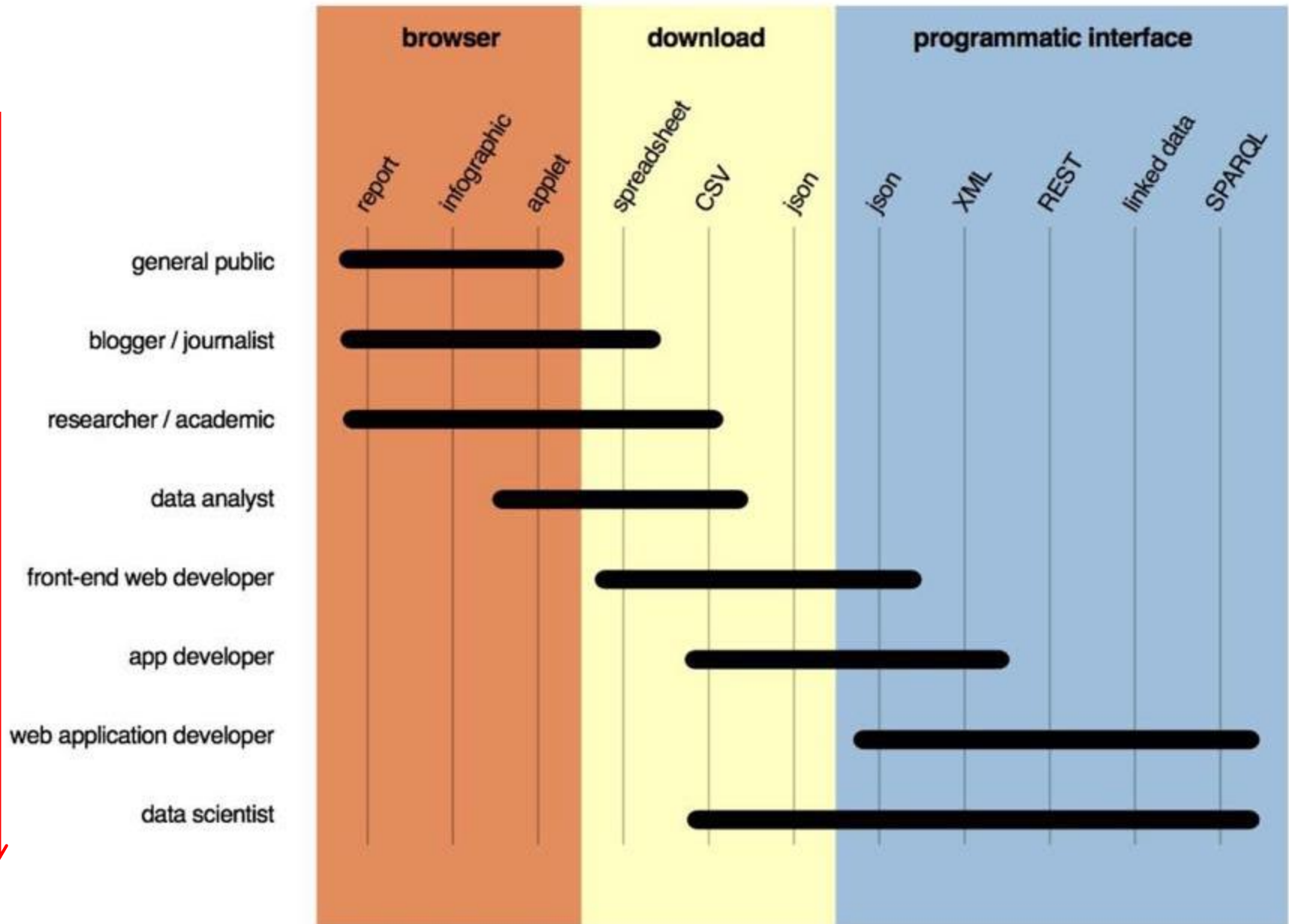
- วิธีการเปิดเผยข้อมูล (Publishing Open Data)
 - Metadata ของชุดข้อมูล
 - หมวดหมู่ของข้อมูล
 - Open Government License (ประเทศไทย)
- การสร้างข้อมูลแบบตารางข้อมูล (table data)
 - ตารางข้อมูล vs. ตารางรายงาน
 - รูปแบบตารางข้อมูลที่ดี
 - แนวปฏิบัติสำหรับการสร้างชุดข้อมูลเปิดสำหรับ Open-D

วิธีการเปิดเผยข้อมูล (Publishing Open Data)

อ้างอิง: แนวปฏิบัติและมาตรฐานเชิงเทคนิคสำหรับศูนย์กลางข้อมูลเปิดภาครัฐ
(Data.go.th) <https://data.go.th/document/14>

รูปแบบ (format) ของการเปิดเผยข้อมูล

ความซับซ้อนของการใช้งานข้อมูล



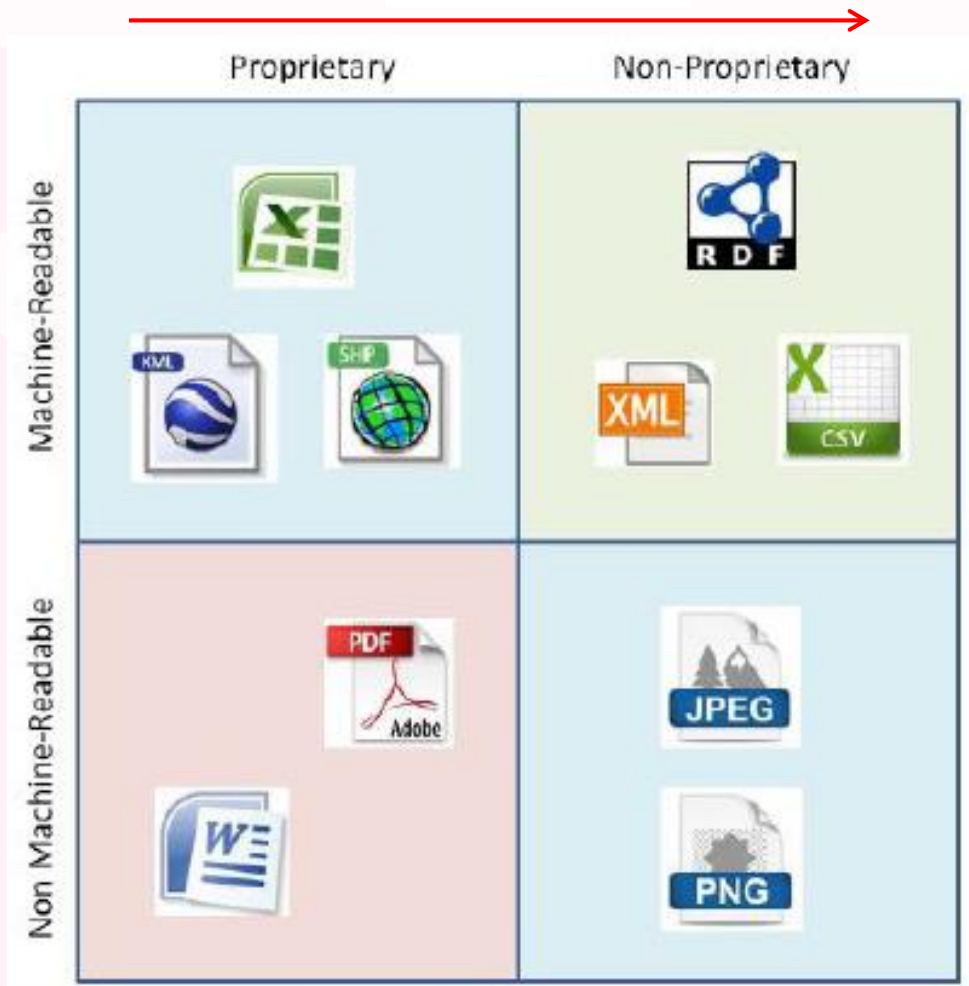
รูปแบบของไฟล์ (file format)

- รูปแบบของไฟล์ที่เหมาะสมสำหรับการเปิดเผยข้อมูล
พิจารณาจาก
 - เป็นมาตรฐานเปิด (open standard)
 - โปรแกรมคอมพิวเตอร์นำไปใช้งานได้ง่าย (machine readable)

รูปแบบการเปิดเผยข้อมูลตามความง่ายต่อการนำไปใช้งาน

มาตรฐานเปิด

การนำไปใช้ซ้ำได้ง่าย



รูปแบบไฟล์ชุดข้อมูลบน Data.go.th

Data.go.th

ค้นหาชุดข้อมูล...

หน้าแรก ชุดข้อมูล แอปพลิเคชัน เอกสารเผยแพร่ ข่าวสารและประกาศ แนะนำ

ค้นหาชุดข้อมูล...

1,063 รายการ

เรียงลำดับตาม: การเรียกดูชุดข้อมูล

ข้อมูลพิกัด LAT/LONG ที่ตั้งตำบล

ข้อมูลพิกัด LAT/LONG ที่ตั้งตำบล อำเภอ จังหวัด (ในรูปแบบ Microsoft Excel) ซึ่งได้จัดทำจากข้อมูลขอบเขตการปกครองของกรมการปกครอง

รายได้เฉลี่ยต่อเดือนต่อครัวเรือน จำแนกตามภาค และจังหวัด พ.ศ. 2541 - 2558

รายได้เฉลี่ยต่อเดือนต่อครัวเรือน พ.ศ. 2541 - 2558

จำนวนรถจดทะเบียน (สะสม) ณ วันที่ 29 กุมภาพันธ์ 2559

สถิติจำนวนรถจดทะเบียนสะสม ณ วันที่ หมายถึง จำนวนรถที่จดทะเบียนตามกฎหมายว่าด้วยรถยนต์ และกฎหมายว่าด้วยการขนส่งทางบก ที่มีอยู่ทั้งสิ้น ณ วันที่ โดยไม่...

ข้อมูลการจดทะเบียนนิติบุคคล พ.ศ. 2558 รายชื่อนิติบุคคลจัดตั้งใหม่ เดือนมกราคม

รายชื่อนิติบุคคลจดทะเบียนจัดตั้งใหม่ประจำเดือน เป็นข้อมูลจากการรับจดทะเบียนนิติบุคคลที่สำนักงานของ กรมพัฒนาธุรกิจการค้าทั่วประเทศ ประกอบด้วยข้อมูลดังนี้ - เลขทะเบียน 13 หลัก - ช...

ดูเพิ่มเติม

ชนิดของไฟล์

csv (528)

xls (436)

xlsx (398)

pdf (128)

xml (77)

html (20)

ods (8)

ข้อมูล ณ วันที่ 13 ก.พ. 61

เมทาเดตาของชุดข้อมูล

- ช่วยให้ผู้ใช้สามารถสืบค้นชุดข้อมูลได้จากข้อมูลที่เกี่ยวข้อง เช่น ชื่อและรายละเอียดชุดข้อมูล หน่วยงาน หมวดหมู่ เป็นต้น
- ช่วยให้ผู้ใช้เข้าใจขอบเขตเนื้อหาของชุดข้อมูลโดยไม่ต้องเปิดไฟล์ชุดข้อมูล

เมตาเดตาของชุดข้อมูล data.go.th

ลำดับที่	รายการ	ฟิลด์บังคับ	คำอธิบาย
1	Unique Identifier	Yes	รหัส (ID) สำหรับข้อมูลในแต่ละชุด (Auto-generated)
2	Title	Yes	ชื่อหัวข้อของชุดข้อมูล
3	Description	Yes	บ่งบอกถึงเนื้อหาของข้อมูลอย่างสั้น
4	Tags	Yes	Keywords ที่เกี่ยวข้องกับชุดข้อมูล สำหรับการค้นหา
5	Last Updated	Yes	วันที่ทำการเปลี่ยนแปลงหรืออัปเดตข้อมูลครั้งล่าสุด (Auto-generated)
6	Publisher	Yes	ชื่อหน่วยงานเจ้าของชุดข้อมูลที่เผยแพร่
7	Contact Name	Yes	ชื่อผู้ติดต่อหลักสำหรับข้อมูลที่นำมา เผยแพร่
8	Contact Email	Yes	E-mail ของผู้ติดต่อหลัก
9	Contact Number	Yes	เบอร์โทรของผู้ติดต่อหลัก

เมตาเดตาของชุดข้อมูล data.go.th (2)

ลำดับที่	รายการ	ฟิลด์บังคับ	คำอธิบาย
10	Format	Yes	รูปแบบของไฟล์ข้อมูล
11	Frequency	Yes	ความถี่ในการเผยแพร่ข้อมูล
12	Category	Yes	หมวดหมู่ของชุดข้อมูล
13	Language	Yes	ภาษาของชุดข้อมูล
14	Public Access Level	Yes	ระดับในการเข้าถึงข้อมูล (มี/ ไม่มีข้อจำกัดในการใช้งานชุดข้อมูล)
15	URL		URL ของชุดข้อมูล
16	License		Open Government License
17	Rights		อธิบายเพิ่มเติมจากข้อ 14

หมวดหมู่ (category) ของชุดข้อมูล

หมวดหมู่	จำนวนชุดข้อมูล
สังคมและสวัสดิการ	200
ศาสนา ศิลปะ และวัฒนธรรม	166
คมนาคม และโลจิสติกส์	118
เศรษฐกิจ การเงิน และอุตสาหกรรม	108
สาธารณสุข	81
พลังงาน ทรัพยากรธรรมชาติ และสิ่งแวดล้อม	76
งบประมาณ และการใช้จ่ายของภาครัฐ	72
กฎหมาย ศาล และอาชญากรรม	67
เกษตรกรรม และชลประทาน	36
สถานที่ การท่องเที่ยว และกีฬา	33
วิจัย วิทยาศาสตร์ เทคโนโลยีและนวัตกรรม	31
เทคโนโลยีสารสนเทศ และการสื่อสาร	26
การศึกษา	20
การเมือง และการปกครอง	18
แผนที่	4
ภูมิอากาศ และภัยพิบัติ	4
ตัวระบุวัตถุดิจิทัล และรหัสมาตรฐาน	3

Data.go.th

กำหนดหมวดหมู่
ของชุดข้อมูล
17 หมวดหมู่

(ดัดแปลงจาก
หมวดหมู่ที่กำหนด
โดยสำนักงานสถิติ
แห่งชาติ)

Open Government License – Thailand

สิทธิและเสรีภาพของผู้ใช้ข้อมูล

1. คัดลอก เผยแพร่ แจกจ่าย และส่งข้อมูล เพื่อวัตถุประสงค์อันชอบด้วยกฎหมาย
2. ปรับใช้ข้อมูล โดยวิธีการหรือรูปแบบใดๆ ซึ่งไม่เป็นการปรับเปลี่ยนเนื้อหาสาระของข้อมูล เพื่อวัตถุประสงค์อันชอบด้วยกฎหมาย
3. ใช้ประโยชน์ของข้อมูล ซึ่งรวมถึง การรวมข้อมูลเข้ากับข้อมูลอื่นๆ หรือการนำข้อมูลมาเป็น ส่วนหนึ่งของผลิตภัณฑ์หรือแอปพลิเคชันของผู้ใช้ข้อมูล ในเชิงสาธารณะประโยชน์ หรือในเชิงพาณิชย์ เพื่อวัตถุประสงค์อันชอบด้วยกฎหมาย

Open Government License – Thailand (2)

- ผู้ใช้ข้อมูล**ต้องอ้างอิงถึงผู้อนุญาต** โดยใส่ข้อความเฉพาะของผู้ให้บริการข้อมูลที่ระบุถึงผู้อนุญาต
 - หรือ อาจระบุเป็น “สิทธิการใช้ข้อมูล ภายใต้สัญญาอนุญาตให้ใช้ข้อมูลภาครัฐแบบเปิดของประเทศไทย (Open Government License - Thailand) จาก “Data.go.th””
- การอนุญาตให้ใช้ข้อมูลนี้ไม่ครอบคลุมถึง “**ข้อมูลส่วนบุคคลที่อยู่ในเนื้อหาข้อมูล**”

License อื่นๆ – Creative Commons

สัญลักษณ์	ชื่อเรียก	คำอธิบาย
	CC BY	อนุญาตให้ใช้ข้อมูลอย่างอิสระ โดยต้องระบุที่มาข้อมูล (Credit original creator)
	CC Share Alike	อนุญาตให้ใช้ข้อมูลอย่างอิสระ โดยต้องระบุที่มาข้อมูล และต้องอนุญาตให้ใช้ข้อมูลที่สร้างขึ้นภายใต้สิทธิ์เดียวกัน (Credit original creator + credit the same license)
	CC NoDerivs	ห้ามดัดแปลงแก้ไข
	CC NonCommercial	ห้ามนำไปใช้ในเชิงพาณิชย์

การสร้างข้อมูลแบบตารางข้อมูล (table data) สำหรับแพลตฟอร์ม Open-D

ตารางข้อมูล (data) vs. ตารางรายงาน (report)

- ตารางข้อมูล (data table)
 - มีจำนวนแถวข้อมูลเป็นจำนวนมาก เป็นข้อมูลดิบที่ยังไม่มีการสรุป
 - สามารถนำข้อมูลไปวิเคราะห์ข้อมูลต่อได้อย่างยืดหยุ่น
 - สามารถสืบค้นกรองข้อมูล (filter) ได้ตามฟิลด์ของข้อมูล
 - เป็นรูปแบบที่เหมาะสมสำหรับการประมวลผลด้วยเครื่อง (machine consumption)

ตารางข้อมูล (data) vs. ตารางรายงาน (report)

- ตารางรายงาน (report table)
 - มีจำนวนแถวข้อมูลไม่มาก เป็นข้อมูลที่สรุปแล้ว
 - เป็นข้อมูลที่สรุปแล้ว ไม่นิยมนำไปวิเคราะห์ต่อ
 - เป็นข้อมูลที่มีปริมาณไม่มาก ไม่จำเป็นในการสืบค้น
 - เป็นรูปแบบที่เหมาะสมสำหรับมนุษย์ (human consumption)

ตัวอย่างตารางรายงาน (report table)

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	รายได้เฉลี่ยต่อเดือนต่อครัวเรือน จำแนกตามภาค และจังหวัด พ.ศ. 2539 - 2556												
2	AVERAGE MONTHLY INCOME PER HOUSEHOLD: 1996 - 2013												
3	บาท Baht												
4	ภาค และจังหวัด	2539	2541	2543	2545	2547	2549	2550	2552	2554	2556	Region and province	
5		(1996)	(1998)	(2000)	(2002)	(2004)	(2006)	(2007)	(2009)	(2011)	(2013)		
6													
7	ทั่วราชอาณาจักร	10,779	12,492	12,150	13,736	14,963	17,787	18,660	20,904	23,236	25,194	Whole Kingdom	
8	ตอนพิเศษ	21,947	24,929	25,242	28,239	28,135	33,088	35,007	37,732	41,631	43,058	Greater Bangkok	
9	กรุงเทพมหานคร	21,550	26,054	26,909	29,589	29,843	36,658	39,020	42,380	48,951	49,191	Bangkok	
10	สมุทรปราการ	17,146	18,100	15,745	19,680	19,946	20,382	21,302	23,359	23,798	29,575	Samut Prakan	
11	นนทบุรี	36,888	24,211	24,566	29,119	26,658	31,152	32,743	34,626	35,120	30,664	Nonthaburi	
12	ปทุมธานี	15,539	21,793	19,282	22,838	21,530	25,143	26,107	26,686	21,616	33,461	Pathum Thani	
13	ภาคกลาง	10,907	12,643	13,012	14,128	16,355	19,279	18,932	20,960	20,822	26,114	Central Region	
14	พระนครศรีอยุธยา	11,376	12,918	14,904	13,319	14,980	19,676	21,676	25,820	22,302	26,482	Phra Nakhon Si Ayutthaya	
15	อ่างทอง	11,414	10,878	12,544	11,653	12,855	18,300	17,704	25,506	21,140	28,641	Ang Thong	
16	ลพบุรี	9,331	10,587	10,649	11,010	15,003	19,935	16,852	22,405	17,178	23,426	Lop Buri	
17	สิงห์บุรี	9,525	10,786	11,894	14,434	14,788	18,510	20,558	25,419	26,068	28,115	Sing Buri	
18	ชัยนาท	9,575	11,159	12,693	11,119	13,383	13,058	13,995	18,719	20,181	32,754	Chai Nat	
19	สระบุรี	10,582	13,281	12,319	14,677	18,742	20,468	22,363	21,947	27,114	32,834	Saraburi	
20	ชลบุรี	12,223	14,075	13,807	16,797	22,286	21,869	22,260	24,052	23,007	28,367	Chon Buri	
21	ระยอง	13,254	12,491	14,739	13,255	18,962	20,018	25,090	22,983	21,929	30,401	Rayong	
22	จันทบุรี	12,411	15,548	11,527	15,959	15,897	20,606	18,866	19,442	24,278	27,284	Chanthaburi	
23	ตราด	11,696	11,703	7,426	13,822	14,080	20,286	16,664	16,949	28,118	21,653	Trat	
24	ฉะเชิงเทรา	11,820	14,010	13,781	14,829	16,938	16,770	20,665	21,252	23,031	34,548	Chachoengsao	
25	ปราจีนบุรี	8,117	9,852	10,477	12,720	15,032	16,031	18,263	22,548	25,338	21,039	Prachin Buri	
26	นครนายก	9,526	11,809	12,595	10,722	13,096	14,104	15,983	17,503	17,042	23,391	Nakhon Nayok	
27	สระแก้ว	7,189	6,958	8,355	9,951	10,777	11,577	13,593	15,525	17,781	24,805	Sa Kaeo	
28	ราชบุรี	12,621	12,868	15,174	14,692	19,867	23,202	17,576	20,994	18,713	25,742	Ratchaburi	
29	กาญจนบุรี	10,230	13,082	11,680	15,380	12,122	15,606	15,326	17,571	15,210	18,746	Kanchanaburi	
30	สุพรรณบุรี	7,427	12,409	13,334	12,984	15,781	14,783	15,112	15,797	15,928	17,260	Suphan Buri	

ตัวอย่างตารางข้อมูล (data table)

	A	B	C	D
1	จังหวัด	ปี	รายได้เฉลี่ยต่อครัวเรือน	
2	กรุงเทพมหานคร	2539	21,550	
3	สมุทรปราการ	2539	17,146	
4	นนทบุรี	2539	36,888	
5	ปทุมธานี	2539	15,539	
6	พระนครศรีอยุธยา	2539	11,376	
7	อ่างทอง	2539	11,414	
8	ลพบุรี	2539	9,331	
9	สิงห์บุรี	2539	9,525	
10	ชัยนาท	2539	9,575	
11	สระบุรี	2539	10,582	
12	ชลบุรี	2539	12,223	
13	ระยอง	2539	13,254	
14	จันทบุรี	2539	12,411	
15	ตราด	2539	11,696	
16	ฉะเชิงเทรา	2539	11,820	
17	ปราจีนบุรี	2539	8,117	
18	นครนายก	2539	9,526	
19	สระแก้ว	2539	7,189	
20	ราชบุรี	2539	12,621	
21	กาญจนบุรี	2539	10,230	
22	สุพรรณบุรี	2539	7,427	
23	นครปฐม	2539	14,737	

ประโยชน์ของการสร้างข้อมูลในแบบตารางข้อมูล

- เป็นข้อมูลชนิดเครื่องคอมพิวเตอร์ประมวลผลได้ (machine-readable data) ที่สามารถนำไปต่อยอดใช้ประโยชน์ได้ง่าย
 - สามารถนำเข้าสู่ฐานข้อมูลได้ทันที
 - สามารถสืบค้นได้ตามฟิลด์ของข้อมูลได้อย่างยืดหยุ่น (search)
 - สามารถนำไปวิเคราะห์ข้อมูลได้อย่างยืดหยุ่น (analysis)
 - สามารถแปลงไปเป็นข้อมูลเปิดในระดับ 4-5 ดาวได้ต่อไป (4-5 star open data)

รูปแบบตารางข้อมูลที่ถูกรูปแบบ (well-formed table data)

- ไฟล์ชุดข้อมูลที่อยู่ในรูปแบบของ ตารางฐานข้อมูล
 - แถวที่หนึ่งต้องเป็น header ที่ระบุชื่อคอลัมน์ (ชื่อฟิลด์)
 - แถวที่ตามมาต้องเป็นข้อมูลที่มีรายละเอียดแยกตามคอลัมน์
 - 1 คอลัมน์ 1 ฟิลด์ (ไม่มีการ span คอลัมน์)
 - ไม่มีคำอธิบายข้อมูลอยู่เหนือ หรือ ท้ายตาราง

	A	B	C	D	E
1	ชื่อ-สกุล	สาขาสำนักงาน	ที่อยู่	จังหวัด	เบอร์โทรศัพท์
2	นายสมชาย ใจดี	รัชดาภิเษก	1 ถ.รัชดาภิเษก เขตดินแดง	กรุงเทพมหานคร	02-123-4567
3	นายสมศักดิ์ สุภาพ	คลองหลวง	111 ถ.พหลโยธิน ต.คลองหนึ่ง อ.คลองหลวง	ปทุมธานี	02-564-9999
4	นายสมพงษ์ สวัสดิ์	รัชดาภิเษก	1 ถ.รัชดาภิเษก เขตดินแดง	กรุงเทพมหานคร	02-123-4567
5					
6					



ข้อผิดพลาดที่พบบ่อยในการสร้างข้อมูลตาราง (1)

- มีคำอธิบายอยู่เหนือ หรือ ใต้ตาราง

	A	B	C	D	E
1	รายชื่อพนักงานบริษัท ABC ในสาขาต่างๆ			ต้องลบแถวออก	
2	ชื่อ-สกุล	สาขาสานักงาน	ที่อยู่	จังหวัด	เบอร์โทรศัพท์
3	นายสมชาย ใจดี	รัชดาภิเษก	1 ถ.รัชดาภิเษก เขตดินแดง	กรุงเทพมหานคร	02-123-4567
4	นายสมพงษ์ สวัสดิ์	รัชดาภิเษก	1 ถ.รัชดาภิเษก เขตดินแดง	กรุงเทพมหานคร	02-123-4567
5	นายสมศักดิ์ สุภาพ	คลองหลวง	111 ถ.พหลโยธิน ต.คลองหนึ่ง อ.คลองหลวง	ปทุมธานี	02-564-9999
6					
7	ที่มา: เว็บไซต์บริษัท http://www.abc.in.th/			ต้องลบแถวออก	
8	ปรับปรุงล่าสุด: 28 ก.พ. 2559				
9					

วิธีแก้ไข: ควรลบคำอธิบายอยู่เหนือ หรือ ใต้ตารางออก และนำไปใส่ใน Description ในส่วนของ metadata แทน

ข้อผิดพลาดที่พบบ่อยในการสร้างข้อมูลตาราง (2)

- มีมากกว่า 1 ตารางใน 1 worksheet

	A	B	C	D	E	
1	ชื่อ-สกุล	สาขาสำนักงาน	ที่อยู่	จังหวัด	เบอร์โทรศัพท์	ตารางที่ 1
2	นายสมชาย ใจดี	รัชดาภิเษก	1 ถ.รัชดาภิเษก เขตดินแดง	กรุงเทพมหานคร	02-123-4567	
3	นายสมพงษ์ สวัสดิ์	รัชดาภิเษก	1 ถ.รัชดาภิเษก เขตดินแดง	กรุงเทพมหานคร	02-123-4567	
4	นายสมศักดิ์ สุภาพ	คลองหลวง	111 ถ.พหลโยธิน ต.คลองหนึ่ง อ.คลองหลวง	ปทุมธานี	02-564-9999	
5						
6	ชื่อสาขา	จังหวัด	ที่อยู่	เบอร์โทรศัพท์		ตารางที่ 2
7	รัชดาภิเษก	กรุงเทพมหานคร	1 ถ.รัชดาภิเษก เขตดินแดง	02-123-4567		
8	พญาไท	กรุงเทพมหานคร	1 ถ.พญาไท เขตราชเทวี	02-234-5678		
9	คลองหลวง	ปทุมธานี	111 ถ.พหลโยธิน ต.คลองหนึ่ง อ.คลองหลวง	02-564-9999		
10						

วิธีแก้ไข: ควรแยกเป็น 1 ตาราง ต่อ 1 ไฟล์

ข้อผิดพลาดที่พบบ่อยในการสร้างข้อมูลตาราง (3)

- มีการละค่าที่มีค่าเดียวกันกับเซลล์ของแถวก่อนหน้า

	A	B	C	D	E	F
1	จังหวัด	ชื่อ-สกุล	สาขาสำนักงาน	ที่อยู่	เบอร์โทรศัพท์	
2	กรุงเทพมหานคร	นายสมชาย ใจดี	รัชดาภิเษก	1 ถ.รัชดาภิเษก เขตดินแดง	02-123-4567	
3		นายสมพงษ์ สวัสดิ์	รัชดาภิเษก	1 ถ.รัชดาภิเษก เขตดินแดง	02-123-4567	
4		นายสวัสดิ์ สมพงษ์	พญาไท	1 ถ.พญาไท เขตราชเทวี	02-234-5678	
5	ปทุมธานี	นายสมศักดิ์ สุภาพ	คลองหลวง	111 ถ.พหลโยธิน ต.คลองหนึ่ง อ.คลองหลวง	02-564-9999	
6		นายสุภาพ ดีมาก	คลองหลวง	112 ถ.พหลโยธิน ต.คลองหนึ่ง อ.คลองหลวง	02-564-10000	
7						
8						



	A	B	C	D	E	
1	จังหวัด	ชื่อ-สกุล	สาขาสำนักงาน	ที่อยู่	เบอร์โทรศัพท์	
2	กรุงเทพมหานคร	นายสมชาย ใจดี	รัชดาภิเษก	1 ถ.รัชดาภิเษก เขตดินแดง	02-123-4567	
3	กรุงเทพมหานคร	นายสมพงษ์ สวัสดิ์	รัชดาภิเษก	1 ถ.รัชดาภิเษก เขตดินแดง	02-123-4567	
4	กรุงเทพมหานคร	นายสวัสดิ์ สมพงษ์	พญาไท	1 ถ.พญาไท เขตราชเทวี	02-234-5678	
5	ปทุมธานี	นายสมศักดิ์ สุภาพ	คลองหลวง	111 ถ.พหลโยธิน ต.คลองหนึ่ง อ.คลองหลวง	02-564-9999	
6	ปทุมธานี	นายสุภาพ ดีมาก	คลองหลวง	112 ถ.พหลโยธิน ต.คลองหนึ่ง อ.คลองหลวง	02-564-10000	



วิธีแก้ไข: ควรใส่ค่าซ้ำให้ครบทุกช่อง โดยไม่ละไว้ในฐานที่เข้าใจ

ข้อผิดพลาดที่พบบ่อยในการสร้างข้อมูลตาราง (4)

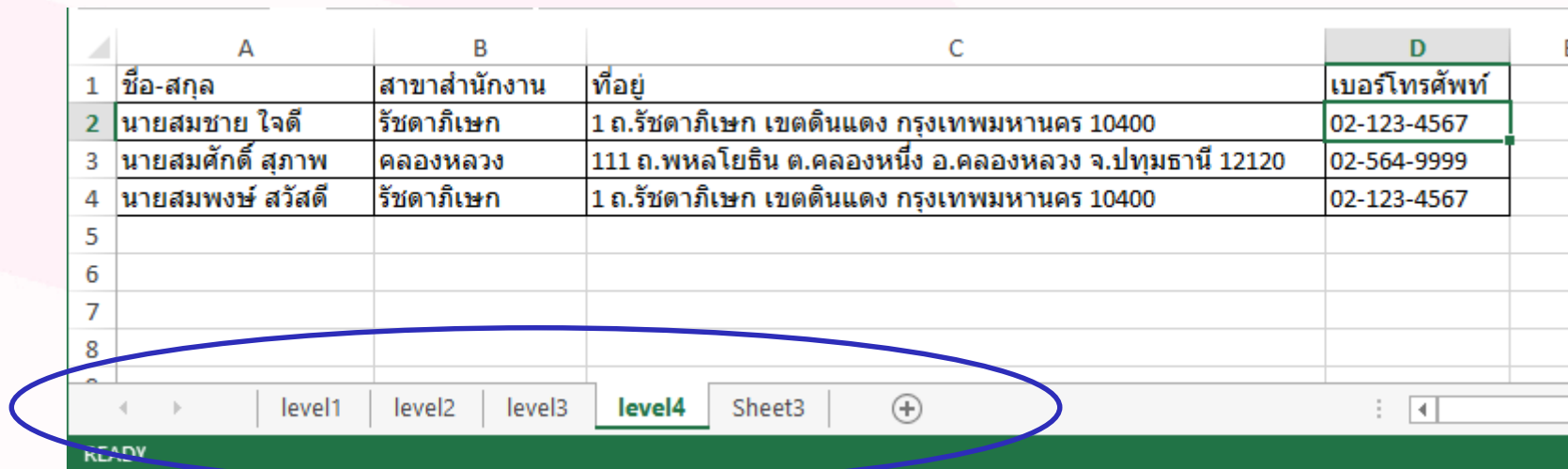
- มีการ merge cell ในส่วน header

	A	B	C	D	E
1			ข้อมูลติดต่อ		
2	ชื่อ-สกุล	สาขาสำนักงาน	ที่อยู่	จังหวัด	เบอร์โทรศัพท์
3	นายสมชาย ใจดี	รัชดาภิเษก	1 ถ.รัชดาภิเษก เขตดินแดง	กรุงเทพมหานคร	02-123-4567
4	นายสมศักดิ์ สุภาพ	คลองหลวง	111 ถ.พหลโยธิน ต.คลองหนึ่ง อ.คลองหลวง	ปทุมธานี	02-564-9999
5	นายสมพงษ์ สวัสดิ์	รัชดาภิเษก	1 ถ.รัชดาภิเษก เขตดินแดง	กรุงเทพมหานคร	02-123-4567

วิธีแก้ไข: สร้าง header เพียง 1 แถว ที่ไม่มีการ merge cell ทั้งใน แนวตั้ง และแนวนอน

ข้อผิดพลาดที่พบบ่อยในการสร้างข้อมูลตาราง (5)

- 1 ไฟล์มีมากกว่า 1 worksheet



	A	B	C	D
1	ชื่อ-สกุล	สาขาสำนักงาน	ที่อยู่	เบอร์โทรศัพท์
2	นายสมชาย ใจดี	รัชดาภิเษก	1 ถ.รัชดาภิเษก เขตดินแดง กรุงเทพมหานคร 10400	02-123-4567
3	นายสมศักดิ์ สุภาพ	คลองหลวง	111 ถ.พหลโยธิน ต.คลองหนึ่ง อ.คลองหลวง จ.ปทุมธานี 12120	02-564-9999
4	นายสมพงษ์ สวัสดิ์	รัชดาภิเษก	1 ถ.รัชดาภิเษก เขตดินแดง กรุงเทพมหานคร 10400	02-123-4567
5				
6				
7				
8				

วิธีแก้ไข: 1 ไฟล์ควรมีเพียง 1 worksheet หากเป็นข้อมูลต่าง sheet ที่มีฟิลด์เหมือนกัน ควรนำมาต่อกันใน sheet เดียวกัน หากมีรายการฟิลด์ต่างกันควรแยกเป็นคนละไฟล์

ข้อผิดพลาดอื่นที่พบในการสร้างข้อมูลตาราง

- ไม่มีแถว header
- มีแถวว่างทั้งแถว หรือ คอลัมน์ว่างทั้งคอลัมน์ ล้อมรอบตาราง
- มีข้อมูลหัวตาราง (header) มากกว่า 1 แถว
- แถวที่เป็นหัวตารางมีจำนวนคอลัมน์ไม่เท่ากับแถวที่เป็นข้อมูลในตาราง
- ข้อมูลหัวตาราง (header) มีบางคอลัมน์ที่ไม่มีข้อมูล

Open-D: Dataset Error codes

Error No.	Error Code	ชนิด	คำอธิบาย
100	F_multiple_sheets	ไฟล์	มีหลาย Worksheet อยู่ใน 1 ไฟล์
101	F_invalid_encoding	ไฟล์	มีการระบุ encoding ของภาษาไม่ถูกต้อง
102	F_extra_size	ไฟล์	ไฟล์มีขนาดใหญ่เกินกำหนด
201	T_metadata	ข้อมูล	มีคำอธิบายอยู่เหนือ หรือ ใต้ตาราง
202	T_multiple	ข้อมูล	ข้อมูลหลายตารางที่มีโครงสร้างต่างกัน ถูกรวมเป็นตารางเดียวกัน
203	T_whitespace	ข้อมูล	ตารางมีแถวว่างเปล่าที่ไม่มีข้อมูลอยู่เหนือหรือใต้ตาราง หรือ มีคอลัมน์ที่ไม่มีข้อมูลอยู่คอลัมน์ก่อนหน้าหรือหลังข้อมูลตาราง
302	H_incomplete	หัวตาราง	ข้อมูลหัวตาราง (header) มีบางคอลัมน์ที่ไม่มีข้อมูล
303	H_duplicate	หัวตาราง	มีข้อมูลหัวตาราง (header) มากกว่า 1 แถว
304	H_cardinality	หัวตาราง	แถวที่เป็นหัวตารางมีจำนวนคอลัมน์ไม่เท่ากับแถวที่เป็นข้อมูลในตาราง
305	H_col_span	หัวตาราง	บางคอลัมน์ของหัวตารางมีการควบคอลัมน์หรือควบแถว (column span หรือ row span) ที่เกิดจากการ merge cell
900	ok	ปกติ	เป็นไฟล์ชุดข้อมูลแบบที่เครื่องสามารถประมวลผลได้ (machine-readable dataset)

แนวปฏิบัติสำหรับการสร้างชุดข้อมูลเปิดที่ดี (Best Practices)

1. ข้อมูลควรสื่อความหมาย (understandable)
 - ผู้ใช้สามารถเข้าใจข้อมูลได้จากชื่อฟิลด์ เพื่อประโยชน์ต่อการค้นคืน/วิเคราะห์ข้อมูลในภายหลัง
2. ข้อมูลที่มีรายละเอียด (ข้อมูลดิบ) มีประโยชน์มากกว่าข้อมูลสรุป
 - ผู้ใช้สามารถนำข้อมูลดิบไปสรุปได้เองในภายหลัง
3. หากเป็นไปได้ควรใช้รหัสมาตรฐานในอ้างอิงข้อมูล เช่น รหัสหน่วยงาน รหัสจังหวัด รหัสเลขเรียกหนังสือ (ISBN) เป็นต้น
 - สามารถใช้ในการเชื่อมโยงข้อมูลข้ามชุดข้อมูลได้ในภายหลัง

แนวปฏิบัติสำหรับการสร้างชุดข้อมูลเปิดที่ดี (Best Practices) (2)

4. ไฟล์ชนิด CSV ให้ Save ไฟล์โดยใช้ encoding เป็นแบบ UTF-8

- เพื่อให้แสดงผลข้อมูลภาษาไทยได้ถูกต้องเสมอ

5. ไม่ควรแยกชุดข้อมูลเดียวเป็นหลายไฟล์ย่อย (เช่น 1 ไฟล์ ต่อข้อมูล 1 ปี หรือ ต่อ 1 จังหวัด) ยกเว้นในกรณีที่เป็น เช่น ขนาดไฟล์ใหญ่เกินไป

- 1 ชุดข้อมูล = 1 API

6. การลบชุดข้อมูลในระบบ Open-D จะมีผลให้ API ของชุดข้อมูลนั้นไม่สามารถใช้งานได้อีกต่อไป

- หากต้องการ update ไฟล์ชุดข้อมูล ควรใช้วิธีการ update แทนการลบแล้วสร้างใหม่ (ซึ่งจะมีผลให้รหัสบ่งชี้ (id) ของชุดข้อมูลเปลี่ยน)

สรุป

- การสร้างไฟล์ชุดข้อมูลเพื่อนำเข้าสู่ระบบ Open-D
 - เมทาเดตาของชุดข้อมูล
 - Open Government License
 - รูปแบบตารางข้อมูลที่ถูกรูปแบบ (well-formed table data)
 - แนวปฏิบัติสำหรับการสร้างชุดข้อมูลเปิดที่ดี (best practices)