NAC2018
14th NSTDA Annual Conference
การประชุมวิชาการประจำปี สวทช. ครั้งที่ ๑๔

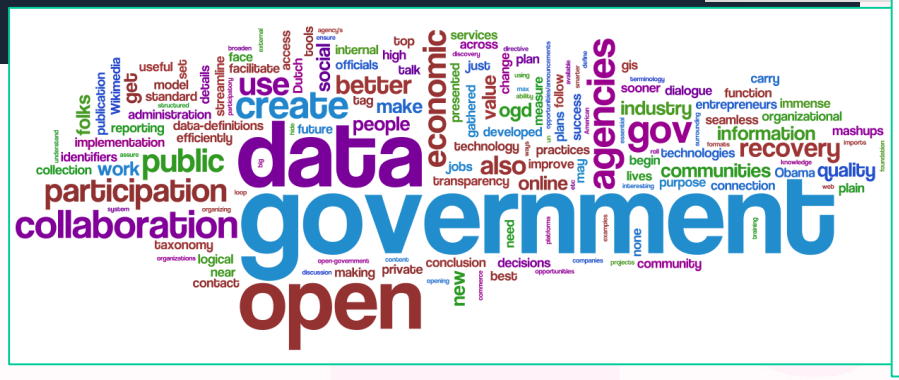
NECTEC
a member of NSTDA

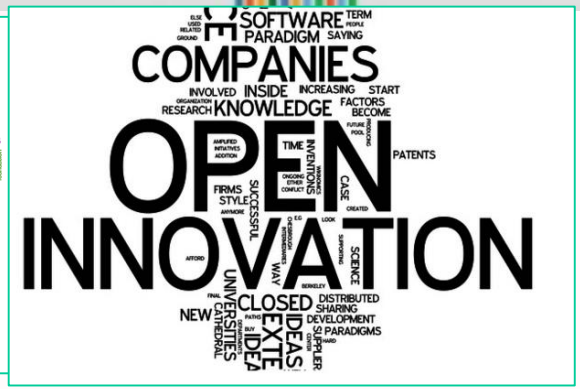
แนะนำระบบแพลตฟอร์มสร้าง API จากไฟล์ชุดข้อมูลเปิด Open Data Service Platform (Open-D)

ดร.มารุต บูรณรัช
ห้องปฏิบัติการวิจัยเทคโนโลยีภาษาธรรมชาติและความหมาย
ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ
marut.bur@nectec.or.th

การเปิดเผยข้อมูลภาครัฐ
(Open Government Data)

Thailand 4.0 – เน้นการสร้างนวัตกรรมด้วยเทคโนโลยีดิจิทัล





Open and Connected Government เพื่อวันพรุ่งนี้ที่ดีกว่า

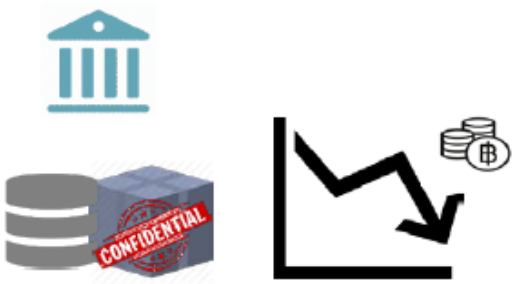
เมื่อก่อน

เข้าถึงยาก

การเข้าถึงข้อมูลภาครัฐ

Open Government

เข้าถึงง่าย ตรวจสอบการดำเนินงานรัฐได้



นำไปใช้ประโยชน์ได้น้อย

โอกาสที่เอกชนจะมีส่วนร่วมในการช่วยงานภาครัฐ

นักพัฒนาคิดค้นนวัตกรรมบริการใหม่
ตลาดเสรี ไม่เปลืองงบประมาณแผ่นดิน

ระบบเปิดเผยข้อมูลภาครัฐเพื่อความโปร่งใส และนวัตกรรมใหม่ๆในสังคมไทย – กวีศักดิ์ กออนันตกูล ๙ พย. ๒๕๖๐

กรณีศึกษา: การใช้ข้อมูลเปิดภาครัฐเพื่อเพิ่มความปลอดภัยในชีวิต

เปิดเผยข้อมูล: สถิติอาชญากรรมของ UK

Mobile App แสดงค่าพฤติกรรมความรุนแรงของชุมชน

https://data.gov.uk/dataset/crime_statistics

https://data.gov.uk/apps/asborometer

ตัวอย่างข้อมูลภาครัฐที่มีคุณค่าสูงหากนำมาเปิดเผยได้ (high-valued dataset)

ภูมิอากาศ (weather)

จราจรและขนส่ง

(transportation & traffic)

สิ่งแวดล้อม

(environment)

การใช้จ่ายภาครัฐ

(spending)

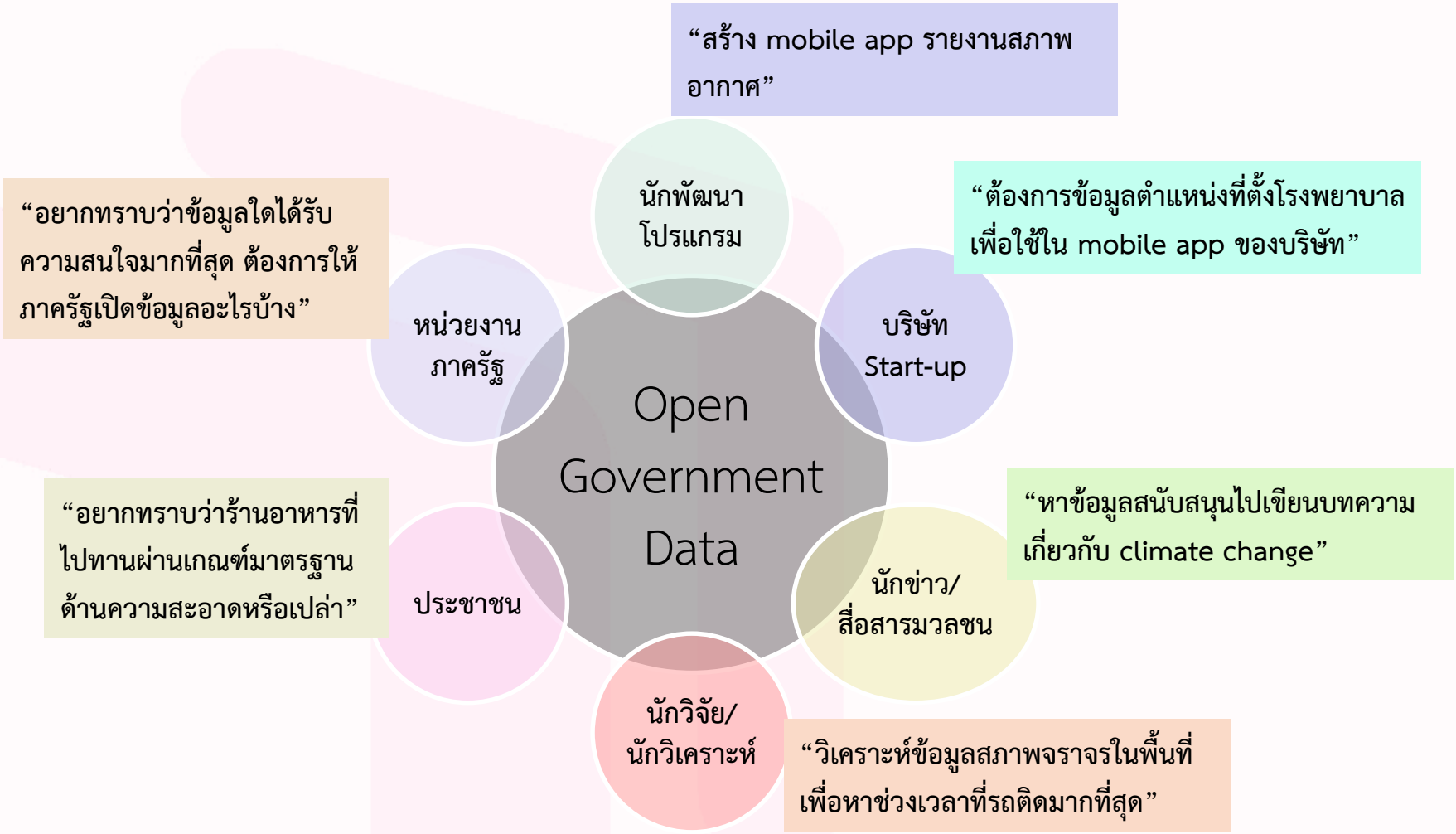
สถานพยาบาล

(hospitals)

อาหารปลอดภัย

(food safety)

ตัวอย่างการใช้ประโยชน์ข้อมูลเปิดภาครัฐ



ที่มา: http://www.rockinst.org/observations/martine/2017-05-10_martin.aspx

A Driving Force for National Science and Technology Capability

คุณสมบัติที่สำคัญของข้อมูลแบบเปิด (Open Data)

- Open License
  - สามารถนำข้อมูลไปใช้ประโยชน์ได้โดยไม่ต้องขออนุญาตเจ้าของข้อมูล
- Machine-readable Data
  - ข้อมูลอยู่ในรูปแบบที่เครื่องคอมพิวเตอร์นำไปประมวลผลได้ง่าย

การเปิดเผยข้อมูลในแบบชุดข้อมูล (dataset)

- การเปิดเผยข้อมูลดิบ (raw data) ในแบบไฟล์ตารางคำนวณ (spreadsheet)

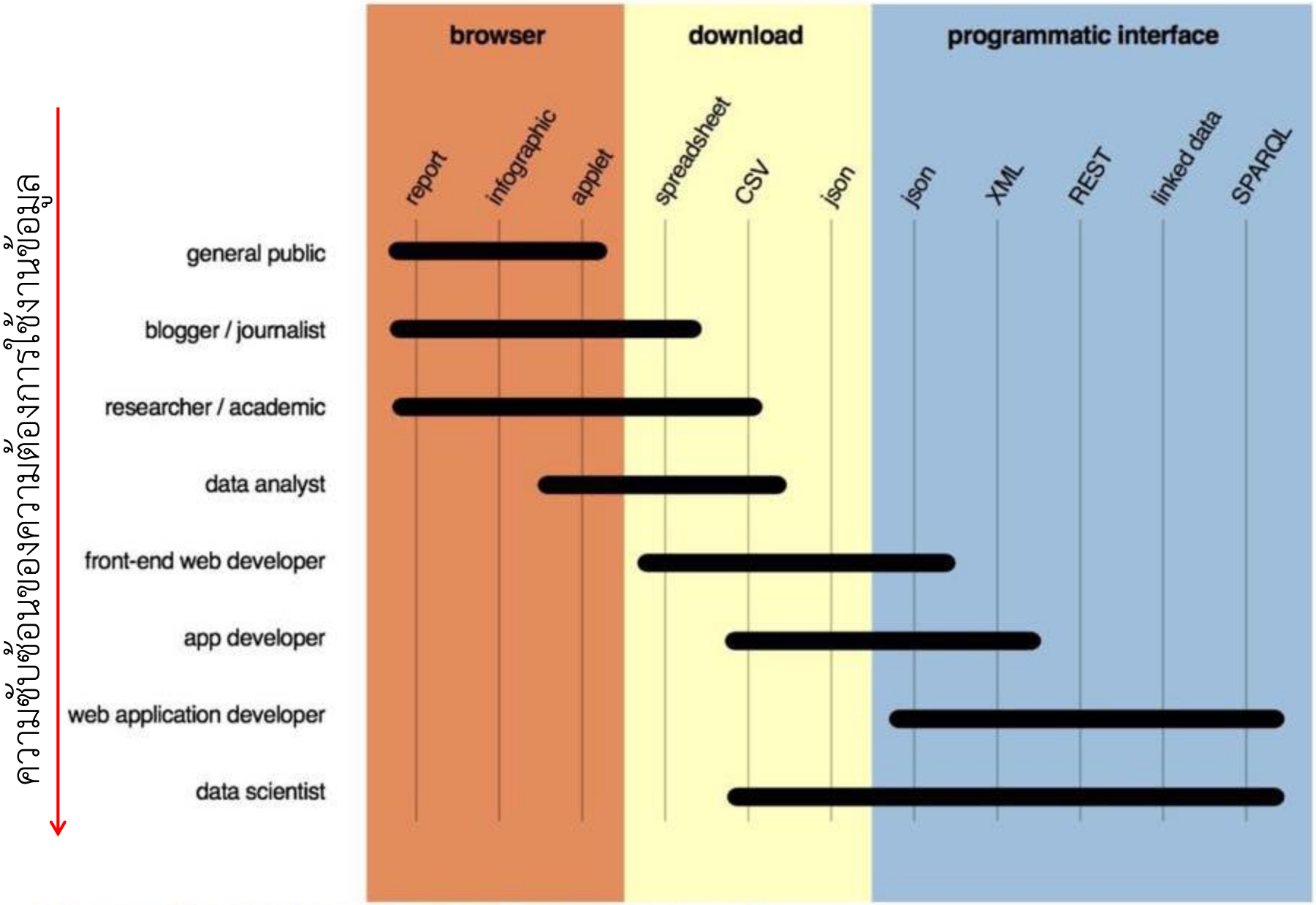
A1 | Data Source

	A	B	C	D	E	F	G	H	I
1	Data Source	World Development Indicators							
2	Last Updated Date	17-02-16							
3									
4	Country Name	Country Code	Indicator Name	Indicator Code	1960	1961	1962	1963	1964
5	Aruba	ABW	Population, tota	SP.POP.TOTL	54208	55435	56226	56697	5702
6	Andorra	AND	Population, tota	SP.POP.TOTL	13414	14376	15376	16410	1747
7	Afghanistan	AFG	Population, tota	SP.POP.TOTL	8994793	9164945	9343772	9531555	972864
8	Angola	AGO	Population, tota	SP.POP.TOTL	5270844	5367287	5465905	5565808	566570
9	Albania	ALB	Population, tota	SP.POP.TOTL	1608800	1659800	1711319	1762621	181413
10	Arab World	ARB	Population, tota	SP.POP.TOTL	92495902	95041593	97691498	100438281	10327392
11	United Arab Emirates	ARE	Population, tota	SP.POP.TOTL	92612	100985	112240	125216	13822
12	Argentina	ARG	Population, tota	SP.POP.TOTL	20619075	20953079	21287682	21621845	2195392
13	Armenia	ARM	Population, tota	SP.POP.TOTL	1867396	1934239	2002170	2070427	213813
14	American Samoa	ASM	Population, tota	SP.POP.TOTL	20012	20478	21118	21883	2270
15	Antigua and Barbuda	ATG	Population, tota	SP.POP.TOTL	54681	55403	56311	57368	5850
16	Australia	AUS	Population, tota	SP.POP.TOTL	10276477	10483000	10742000	10950000	1116700
17	Austria	AUT	Population, tota	SP.POP.TOTL	7047539	7086299	7129864	7175811	722380
18	Azerbaijan	AZE	Population, tota	SP.POP.TOTL	3897889	4030130	4167558	4307315	444565
19	Burundi	BDI	Population, tota	SP.POP.TOTL	2786740	2840375	2894510	2950903	301195
20	Belgium	BEL	Population, tota	SP.POP.TOTL	9153489	9183948	9220578	9289770	937811
21	Benin	BEN	Population, tota	SP.POP.TOTL	2431620	2466002	2503232	2543335	258636
22	Burkina Faso	BFA	Population, tota	SP.POP.TOTL	4829291	4894578	4960325	5027818	509889
23	Bangladesh	BGD	Population, tota	SP.POP.TOTL	48200702	49593610	51030604	52532595	5412939
24	Bulgaria	BGR	Population, tota	SP.POP.TOTL	7867374	7943118	8012946	8078145	814434
25	Bahrain	BHR	Population, tota	SP.POP.TOTL	162501	167924	173107	178048	18277
26	Bahamas, The	BHS	Population, tota	SP.POP.TOTL	109526	115108	121083	127331	13369
27	Bosnia and Herzegovina	BIH	Population, tota	SP.POP.TOTL	3214520	3277096	3341809	3406466	346808

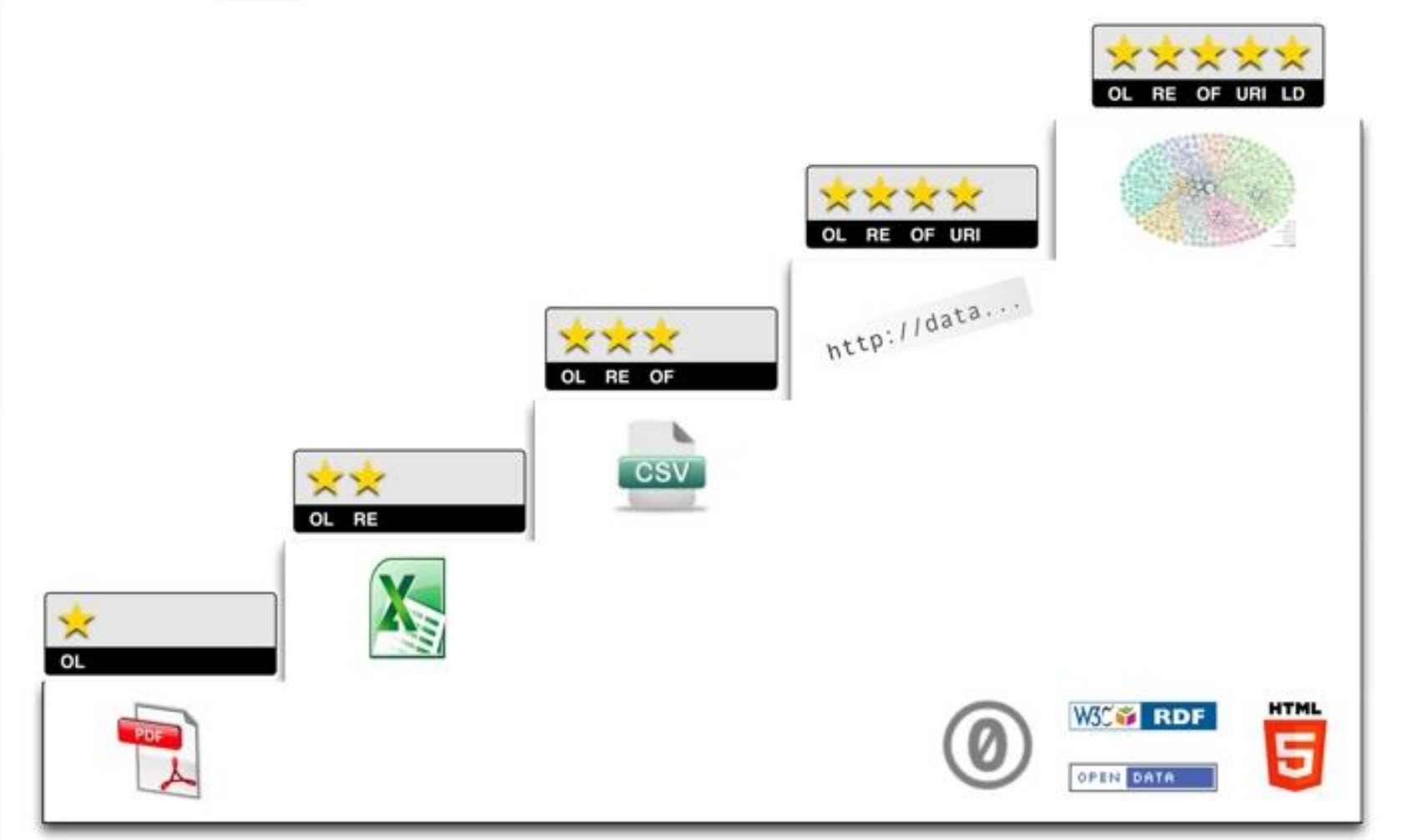
ข้อมูลอยู่ในแบบที่นำไปใช้ซ้ำได้ง่าย ในโปรแกรมคอมพิวเตอร์ที่ต้องการนำข้อมูลนี้ไปใช้ (machine-readable)

ระดับของข้อมูลแบบเปิด
(Levels of open data)

รูปแบบ (format) ของการเปิดเผยข้อมูล



การเปิดเผยข้อมูล 5 ระดับ



ความหมายของการเปิดเผยข้อมูล 5 ระดับ

	★	★★	★★★	★★★★	★★★★★
"OL" On-Line	✓	✓	✓	✓	✓
"RE" (machine) Readable	-	✓	✓	✓	✓
"OF" Open Format	-	-	✓	✓	✓
"URI" As identification for things	-	-	-	✓	✓
"LD" Linked Data	-	-	-	-	✓

ตัวอย่างรูปแบบข้อมูลตามการเปิดเผยข้อมูล 5 ระดับ

ประเภทข้อมูล	ระดับการเปิดเผยข้อมูล
PDF, DOC, TXT, TIFF, JPEG	★☆☆☆☆ (1 ดาว)
XLS	★★☆☆☆ (2 ดาว)
CSV, ODS, XML, JSON, KML, SHP, KMZ	★★★☆☆ (3 ดาว)
RDF (URIs)	★★★★☆ (4 ดาว)
RDF (Linked Data)	★★★★★ (5 ดาว)

ชุดข้อมูลตัวอย่าง

- หน่วยงาน ก. ต้องการเปิดเผยข้อมูลรายชื่อเจ้าหน้าที่ในหน่วยงาน สาขาที่สังกัด ที่อยู่ จังหวัดที่ตั้ง และเบอร์ติดต่อ

การเปิดเผยข้อมูล ระดับที่ 1 -> .pdf

การเปิดเผยข้อมูล ระดับที่ 2 -> .xls (Excel format)

การเปิดเผยข้อมูล ระดับที่ 3 -> .csv (Comma Separated Format)

การเปิดเผยข้อมูล ระดับที่ 4 และ 5 -> .rdf (Resource Description Framework)

ตัวอย่างการเปิดเผยข้อมูล ระดับที่ 1

ข้อมูลรายชื่อเจ้าหน้าที่และสังกัดของหน่วยงาน ก.

ชื่อ-สกุล	สาขาสำนักงาน	ที่อยู่	จังหวัด	เบอร์โทรศัพท์
นายสมชาย ใจดี	รัชดาภิเษก	1 ถ.รัชดาภิเษก เขตดินแดง	กรุงเทพมหานคร	02-123-4567
นายสมศักดิ์ สุภาพ	คลองหลวง	111 ถ.พหลโยธิน ต.คลองหนึ่ง อ.คลองหลวง	ปทุมธานี	02-564-9999
นายสมพงษ์ สวัสดี	รัชดาภิเษก	1 ถ.รัชดาภิเษก เขตดินแดง	กรุงเทพมหานคร	02-123-4567

ข้อมูลปรับปรุงล่าสุด: 23 พ.ค. 2558

employee.pdf

OL = Open License

การเปิดเผยข้อมูล ระดับที่ 1

ข้อดี

- สร้างง่าย
- ง่ายต่อการอ่าน ดูสวยงาม

ข้อเสีย

- นำข้อมูลไปต่อยอดใช้ประโยชน์ยาก (not reusable)

A Driving Force for National Science and Technology Capability

ตัวอย่างการเปิดเผยข้อมูล ระดับที่ 2

	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 สวัสดี	รัชดาภิเษก	1 ถ.รัชดาภิเษก เขตดินแดง	กรุงเทพมหานคร	02-123-4567
5					

** ข้อมูลดิบ (raw data)

employee.xls

OL = Open License
RE = Machine-readable

การเปิดเผยข้อมูล ระดับที่ 2

ข้อดี

- สร้างง่าย
- นำข้อมูลไปต่อยอดใช้ประโยชน์ได้ง่าย (reusable)

ข้อเสีย

- ไม่ใช่มาตรฐานแบบเปิด (not open format)

ตัวอย่างการเปิดเผยข้อมูล ระดับที่ 3

example_xls2.csv - Notepad

File Edit Format View Help

ชื่อ-สกุล,สาขาสำนักงาน,ที่อยู่,จังหวัด,เบอร์โทรศัพท์
นายสมชาย ใจดี,รัชดาภิเษก,1 ถ.รัชดาภิเษก เขตดินแดง,กรุงเทพมหานคร,02-123-4567
นายสมศักดิ์ สุภาพ,คลองหลวง,111 ถ.พหลโยธิน ต.คลองหนึ่ง อ.คลองหลวง,ปทุมธานี,02-564-9999
นายสมพงษ์ สวัสดี,รัชดาภิเษก,1 ถ.รัชดาภิเษก เขตดินแดง,กรุงเทพมหานคร,02-123-4567

employee.csv

OL = Open License
RE = Machine-readable
OF = Open Format

การเปิดเผยข้อมูล ระดับที่ 3

ข้อดี

- สร้างง่าย
- นำข้อมูลไปต่อยอดใช้ประโยชน์ได้ง่าย (reusable)
- มาตรฐานแบบเปิด (open format)

ข้อเสีย

- ยากต่อการบูรณาการเชื่อมโยงข้อมูลข้ามชุดข้อมูล

ตัวอย่างการเปิดเผยข้อมูล ระดับที่ 4

NECTEC a member of NSTDA

```
@prefix a: <http://org1.or.th/data#> .
                                        URI
a:person_1    a:name              "นายสมชาย ใจดี".
a:person_1    a:branch_province   "กรุงเทพมหานคร ".
a:person_1    a:branch_address    "1 ถ.รัชดาภิเษก เขตดินแดง กรุงเทพมหานคร 10400".
a:person_1    a:branch_phone      "02-123-4567".

a:person_2    a:name              "นายสมศักดิ์ สุภาพ".
a:person_2    a:branch_province   "ปทุมธานี".
a:person_2    a:branch_address    "111 ถ.พหลโยธิน ต.คลองหนึ่ง อ.คลองหลวง จ.ปทุมธานี
12120".
a:person_2    a:branch_phone      "02-564-9999".

a:person_3    a:name              "นายสมพงษ์ สวัสดี".
a:person_3    a:branch_province   "กรุงเทพมหานคร".
a:person_3    a:branch_address    "1 ถ.รัชดาภิเษก เขตดินแดง กรุงเทพมหานคร 10400".
a:person_3    a:branch_phone      "02-123-4567".
```

W3C RDF

employee.rdf

★★★★

OL = Open License
RE = Machine-readable
OF = Open Format
URI = Uniform Resource Identifier

การเปิดเผยข้อมูล ระดับที่ 4

ข้อดี

- นำข้อมูลไปต่อยอดใช้ประโยชน์ได้ง่าย (reusable)
- มาตรฐานแบบเปิด (open format)
- รองรับการบูรณาการเชื่อมโยงข้อมูลข้ามชุดข้อมูล

ข้อเสีย

- สร้างได้ยากกว่าระดับที่ 1-3

ตัวอย่างการเปิดเผยข้อมูล ระดับที่ 5

NECTEC a member of NSTDA

```
@prefix a: <http://org1.or.th/data#> .
@prefix b: <http://data.go.th/thai_location#> .

a:person_1      a:name              "นายสมชาย ใจดี".
a:person_1      a:branch_province   b:province_10.
a:person_1      a:branch_address    "1 ถ.รัชดาภิเษก เขตดินแดง กรุงเทพมหานคร 10400".
a:person_1      a:branch_phone      "02-123-4567".

a:person_2      a:name              "นายสมศักดิ์ สุภาพ".
a:person_2      a:branch_province   b:province_13.
a:person_2      a:branch_address    "111 ถ.พหลโยธิน ต.คลองหนึ่ง อ.คลองหลวง จ.ปทุมธานี 12120".
a:person_2      a:branch_phone      "02-564-9999".

a:person_3      a:name              "นายสมพงษ์ สวัสดี".
a:person_3      a:branch_province   b:province_10.
a:person_3      a:branch_address    "1 ถ.รัชดาภิเษก เขตดินแดง กรุงเทพมหานคร 10400".
a:person_3      a:branch_phone      "02-123-4567".
```

URI

อ้างอิงข้อมูลในชุดข้อมูลอื่น

W3C RDF

employee.rdf

★★★★★

OL = Open License
RE = Machine-readable
OF = Open Format
URI = Uniform Resource Identifier
LD = Linked Data

ตัวอย่างการเปิดเผยข้อมูล ระดับที่ 5 (2)

```
@prefix b: <http://data.go.th/thai_location#> .

b:province_10 b:name        "กรุงเทพมหานคร".
b:province_10 b:name_en     "Bangkok".
b:province_10 b:code        "10".

b:province_13 b:name        "ปทุมธานี".
b:province_13 b:name_en     "Pathum Thani".
b:province_13 b:code        "13".

...
```

thai_location.rdf

A Driving Force for National Science and Technology Capability

การเปิดเผยข้อมูล ระดับที่ 5

ข้อดี

- นำข้อมูลไปต่อยอดใช้ประโยชน์ได้ง่าย (reusable)
- มาตรฐานแบบเปิด (open format)
- มีการบูรณาการเชื่อมโยงข้อมูลข้ามชุดข้อมูล (linked data)

ข้อเสีย

- สร้างได้ยากกว่าระดับที่ 1-3
- ต้องมีมาตรฐานรหัสบ่งชี้ (standard identifier)

NECTEC a member of NSTDA
A Driving Force for National Science and Technology Capability

ประโยชน์ของ open data ระดับที่ 4-5

- เพิ่มคุณค่าให้กับชุดข้อมูลที่มีอยู่
  - สามารถนำไปสู่ความเชื่อมโยงของข้อมูล (Linked Data)
    - ง่ายต่อการอ้างอิงไปยังชุดข้อมูลอื่นโดยใช้ URI
    - ชุดข้อมูลอื่นสามารถอ้างอิงมายังชุดข้อมูลของเราโดยใช้ URI
  - ผู้พัฒนาโปรแกรมสามารถ**บูรณาการข้อมูล**จากหลายๆ ชุดข้อมูล มาสร้างเป็นแอพพลิเคชันใหม่ได้ง่ายยิ่งขึ้น

A Driving Force for National Science and Technology Capability

โครงสร้างพื้นฐานของการเชื่อมโยงข้อมูล

- ข้อมูลชนิดมีโครงสร้างที่อยู่ในรูปแบบตารางที่ถูกต้อง (well-formed table data)
- รหัสกลางในการอ้างอิงถึงสิ่งต่างๆ (standard codes)
- เว็บเอพีไอ (Web API)

ข้อมูลชนิดมีโครงสร้างที่อยู่ในรูปแบบตารางที่ถูกต้อง

- ไฟล์ชุดข้อมูลที่จะสามารถแปลงเป็นข้อมูลชนิด RDF ต้องมีคุณภาพดี และมีโครงสร้างข้อมูลในแบบของ**ตารางฐานข้อมูล**
  - แถวแรกเป็นชื่อคอลัมน์/ แถวที่เหลือเป็นค่าข้อมูล
  - 1 คอลัมน์ 1 ชื่อ (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 สวัสดี	รัชดาภิเษก	1 ถ.รัชดาภิเษก เขตดินแดง	กรุงเทพมหานคร	02-123-4567
5					

Ivan Ermilov, Sören Auer, and Claus Stadler. 2013. User-driven semantic mapping of tabular data. In Proceedings of the 9th International Conference on Semantic Systems (I-SEMANTICS '13), Marta Sabou, Eva Blomqvist, Tommaso Di Noia, Harald Sack, and Tassilo Pellegrini (Eds.). ACM, New York, NY, USA, 105-112.

A Driving Force for National Science and Technology Capability

ตัวอย่างข้อมูลมีโครงสร้างตารางไม่ถูกต้อง

NECTEC a member of NSTDA

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

มีการ span คอลัมน์ ในชื่อคอลัมน์ หรือในค่าข้อมูล

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

เก็บข้อมูลในแต่ละคอลัมน์

	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		นายสมพงษ์ สวัสดี	รัชดาภิเษก	1 ถ.รัชดาภิเษก เขตดินแดง	02-123-4567
4	ปทุมธานี	นายสมศักดิ์ สุภาพ	คลองหลวง	111 ถ.พหลโยธิน ต.คลองหนึ่ง อ.คลองหล	02-564-9999

ปล่อยให้ข้อมูลที่มีค่าเหมือนแถวก่อนหน้าเป็นค่าว่าง

ตัวอย่างข้อมูลมีโครงสร้างไม่ถูกต้อง (2)

	A	B	C	D	E	F	G
1		รายชื่อพนักงานบริษัท ABC ในสาขาต่างๆ					
2							
3							
4		ชื่อ-สกุล	สาขาสำนักงาน	ที่อยู่	จังหวัด	เบอร์โทรศัพท์	
5		นายสมชาย ใจดี	รัชดาภิเษก	1 ถ.รัชดาภิเษก เขตดินแดง	กรุงเทพมหานคร	02-123-4567	
6		นายสมพงษ์ สวัสดี	รัชดาภิเษก	1 ถ.รัชดาภิเษก เขตดินแดง	กรุงเทพมหานคร	02-123-4567	
7							
8		ชื่อ-สกุล	สาขาสำนักงาน	ที่อยู่	จังหวัด	เบอร์โทรศัพท์	
9		นายสมศักดิ์ สุภาพ	คลองหลวง	111 ถ.พหลโยธิน ต.คลองหนึ่ง อ.คลองหล	ปทุมธานี	02-564-9999	
10							
11		ที่มา: เว็บไซต์บริษัท http://www.abc.in.th/					
12		ปรับปรุงล่าสุด: 28 ก.พ. 2559					
13							

มีคำอธิบายข้อมูลอยู่เหนือ หรือท้ายตาราง

มีแถว หรือคอลัมน์ว่าง (ไม่มีข้อมูล)

มีแถว ที่เป็นชื่อคอลัมน์หลายครั้ง

A Driving Force for National Science and Technology Capability

รหัสกลางในการอ้างอิงถึงสิ่งต่างๆ (standard codes)

- ต้องมีข้อตกลงเรื่องรหัสบ่งชี้มาตรฐาน (standard identifier) ที่ชุดข้อมูลจากต่างหน่วยงานต้องใช้งานร่วมกัน
  - รหัสบุคคล
  - รหัสหน่วยงาน
  - รหัสสถานที่
  - รหัสถนน
  - รหัสตำบล/อำเภอ
  - รหัสจังหวัด
  - ฯลฯ

	A	B
1	PROVINCE_ID	PROVINCE_NAME
2	10	กรุงเทพมหานคร
3	11	สมุทรปราการ
4	12	นนทบุรี
5	13	ปทุมธานี
6	14	พระนครศรีอยุธยา
7	15	อ่างทอง
8	16	ลพบุรี
9	17	สิงห์บุรี
10	18	ชัยนาท
11	19	สระบุรี
12	20	ชลบุรี
13	21	ระยอง
14	22	จันทบุรี
15	23	ตราด
16	24	ฉะเชิงเทรา
17	25	ปราจีนบุรี
18	26	นครนายก
19	27	สระแก้ว
20	30	นครราชสีมา
21	31	บุรีรัมย์
22	32	สุรินทร์
23	33	ศรีสะเกษ
24	34	อุบลราชธานี
25	35	ยโสธร
26	36	ชัยภูมิ
27	37	อำนาจเจริญ

รหัสจังหวัด

4	2667	กรมการขนส่งทางน้ำและพาณิชยนาวี (กรมเจ้าท่า)
5	2668	กรมการขนส่งทางบก
6	2669	กรมการค้าต่างประเทศ
7	2670	กรมการค้าภายใน
8	2671	กรมการจัดหางาน
9	2672	กรมการเจรจาการค้าระหว่างประเทศ
10	2673	กรมการบินพลเรือน
11	2674	กรมการปกครอง (สำนักบริหารการทะเบียน)
12	2675	กรมการปกครอง คลอง 9
13	2676	กรมการพัฒนาชุมชน
14	2677	กรมการแพทย์
15	2678	กรมการศาสนา
16	2679	กรมการอุตสาหกรรมทหาร
17	2680	กรมควบคุมมลพิษ
18	2681	กรมควบคุมโรค
19	2682	กรมคุ้มครองสิทธิและเสรีภาพ
20	2683	กรมคุมประพฤติ
21	2684	กรมชลประทาน
22	2685	กรมเชื้อเพลิงธรรมชาติ
23	2686	กรมตรวจบัญชีสหกรณ์
24	2687	กรมทรัพย์สินทางปัญญา
25	2688	กรมทรัพยากรทางทะเลและชายฝั่ง
26	2689	กรมทรัพยากรธรณี
27	2690	กรมทรัพยากรน้ำ
28	2691	กรมทรัพยากรน้ำบาดาล
29	2692	กรมทางหลวง
30	2693	กรมทางหลวงชนบท
31	2694	กรมที่ดิน
32	2695	กรมธนารักษ์
33	2696	กรมธุรกิจพลังงาน

รหัสหน่วยงาน

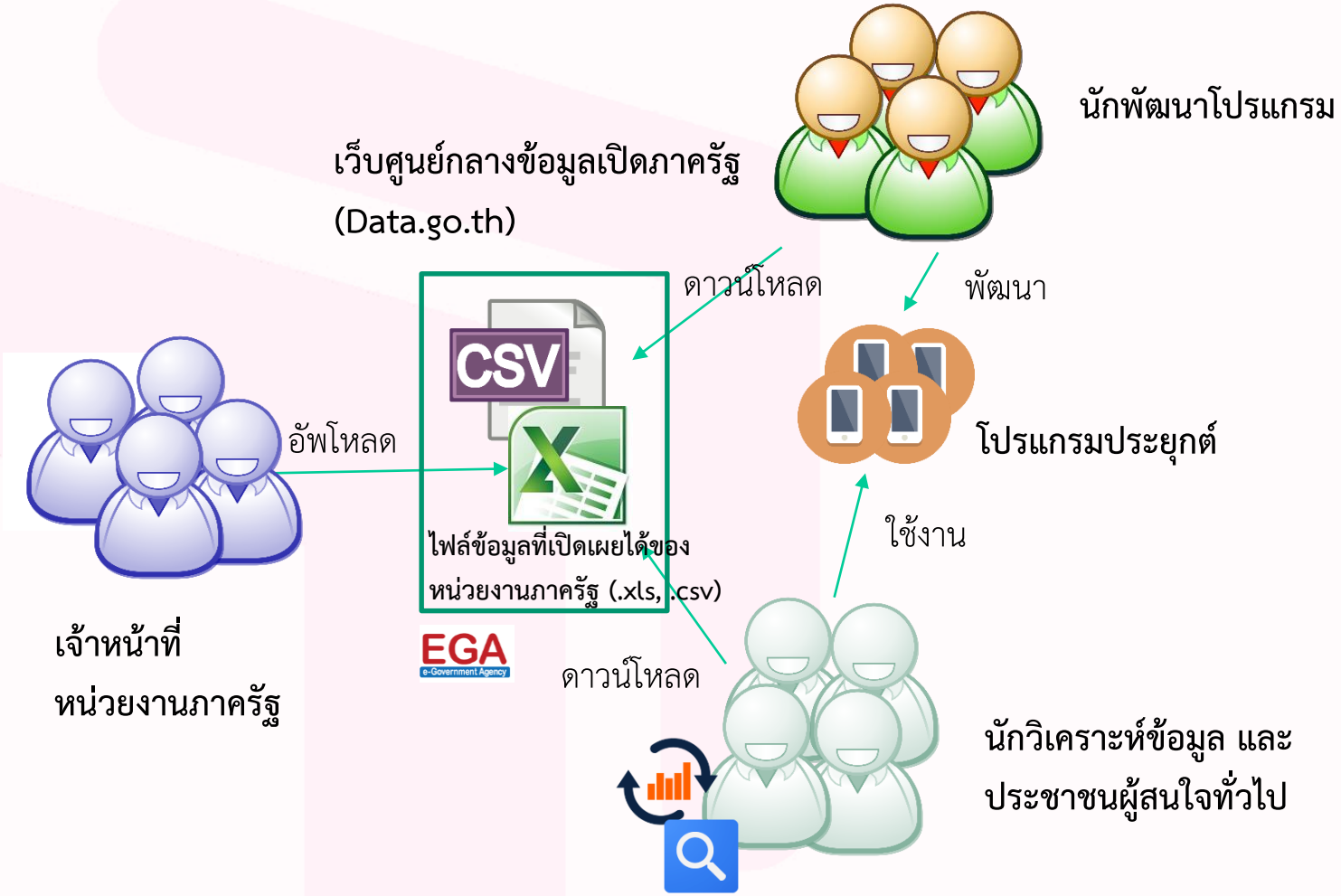
A Driving Force for National Science and Technology Capability

Web API vs. ไฟล์ชุดข้อมูล (Full Dataset)

- **ไฟล์ชุดข้อมูล (Full Dataset)**
  - การใช้งานต้องอ่านข้อมูลทั้งไฟล์ เช่น ข้อมูลทั้ง 10 ปี
  - วิธีการประมวลผลข้อมูลเช่น การสืบค้นข้อมูล ทำโดยนักพัฒนา
  - ต้อง download ชุดข้อมูลใหม่เมื่อมีการปรับปรุงข้อมูล
- **Web API**
  - ใช้งานชุดข้อมูลเฉพาะส่วนที่ต้องการ เช่น เลือกเข้าถึงเฉพาะข้อมูลของปีปัจจุบันเท่านั้น
  - นักพัฒนาไม่ต้องเขียนวิธีประมวลผลข้อมูลเอง เพียงส่งผ่านตัวแปรเท่านั้น
  - เมื่อมีการปรับปรุงข้อมูล สามารถเรียกใช้ API เดิมได้

งานวิจัยแพลตฟอร์มให้บริการข้อมูลแบบเปิด
Open Data Service Platform (Open-D)

การเปิดเผยข้อมูลภาครัฐในแบบไฟล์ชุดข้อมูล (ปัจจุบัน)



การเปิดเผยข้อมูลในแบบไฟล์ชุดข้อมูล (dataset)

- **ข้อดี**
  - จัดเตรียมข้อมูลได้ง่าย
  - ผู้เผยแพร่ข้อมูลไม่จำเป็นต้องมีทักษะด้าน IT ในระดับสูง
- **ข้อเสีย**
  - การใช้งานต้อง download ข้อมูลทั้งไฟล์ เช่น ข้อมูลทั้ง 10 ปี
  - ผู้ใช้ต้อง download ชุดข้อมูลใหม่ทุกครั้ง เมื่อมีการปรับปรุงข้อมูล
  - ผู้คนยุคปัจจุบัน นิยมเข้าถึงข้อมูลผ่าน Social Media และ Smart Phone ซึ่งไฟล์ dataset ในแบบ Excel ไม่ใช่รูปแบบที่ประมวลผลได้โดยสื่อเหล่านี้

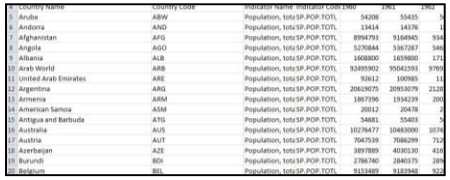
A Driving Force for National Science and Technology Capability

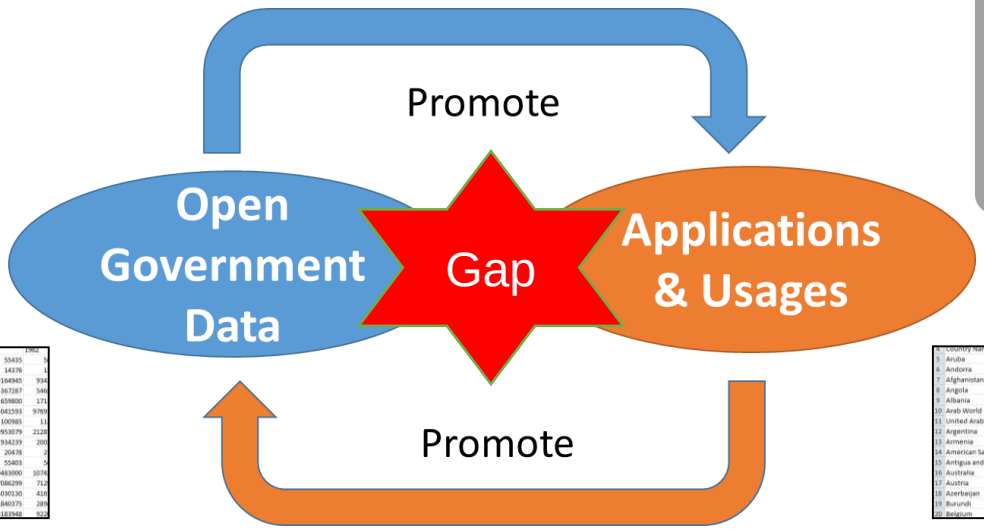
ที่มาของงานวิจัย Open Data Service Platform

- ปัญหา: Open data ในแบบของไฟล์ตารางคำนวณไม่สะดวกต่อการนำไปใช้งาน

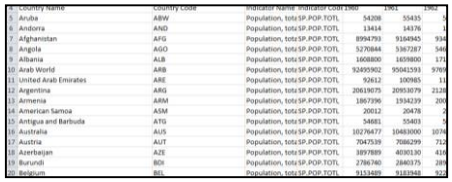
upload

Spreadsheets





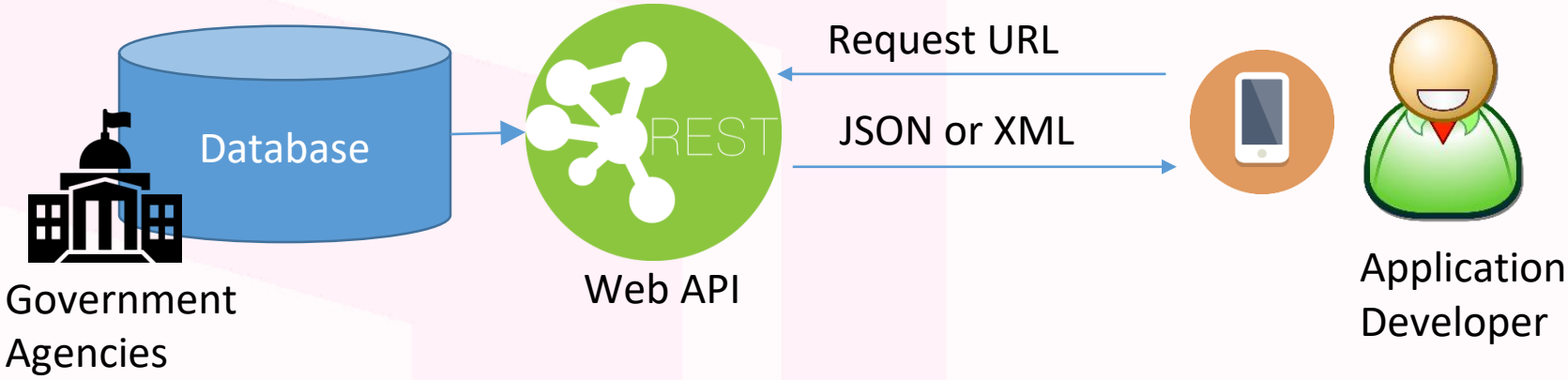
Not easy to consume by



Spreadsheets

- **แนวทางแก้ปัญหา:**
  - แปลงไฟล์ชุดข้อมูล (dataset) ให้เป็น บริการข้อมูล (data service)

การเปิดเผยข้อมูลในแบบ API (Application Programming Interface)

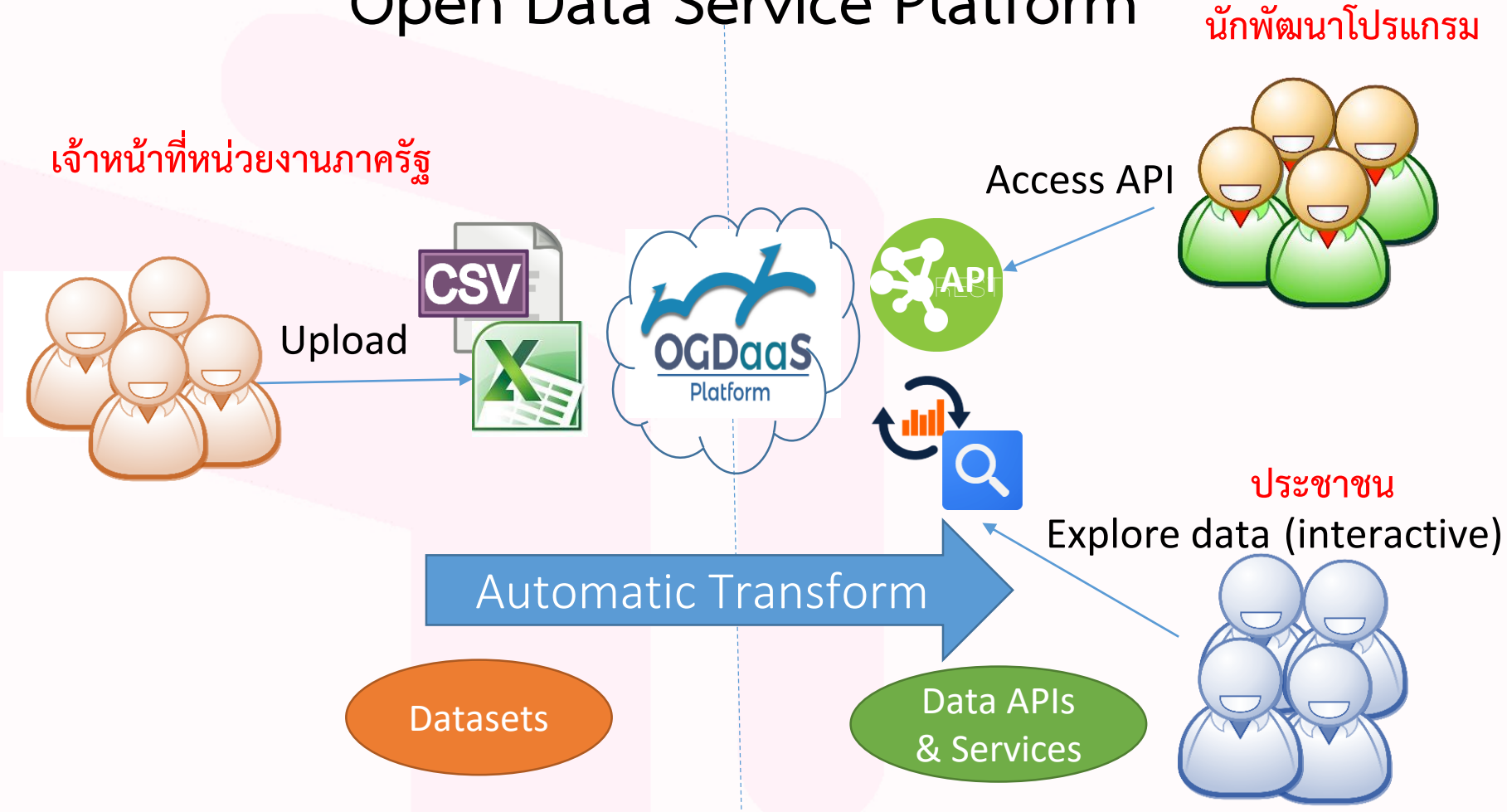


* การแบ่งปันข้อมูลในแบบ API เป็นวิธีที่ช่วยให้นักพัฒนาโปรแกรมสามารถนำข้อมูลไปใช้ได้ง่ายที่สุด

การเปิดเผยข้อมูลในแบบ API (2)

- **ข้อดี**
  - ใช้งานชุดข้อมูลเฉพาะส่วนที่ต้องการ เช่น เลือกเข้าถึงเฉพาะข้อมูลของปีปัจจุบันเท่านั้น
  - เมื่อมีการปรับปรุงข้อมูล สามารถเรียกใช้ API เดิมได้
  - นักพัฒนาไม่ต้องเขียนวิธีประมวลผลข้อมูลเอง เช่น การสืบค้นข้อมูล สามารถเรียกคำสั่งประมวลผลผ่าน API ได้
- **ข้อเสีย**
  - หน่วยงานต้องลงทุนค่าใช้จ่าย (cost) ในการพัฒนา API

ระบบสร้าง API จากชุดข้อมูลเปิด
Open Data Service Platform



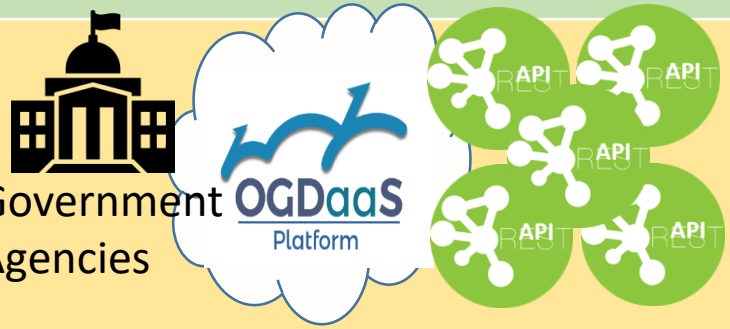
A Driving Force for National Science and Technology Capability

ระบบสร้าง API จากชุดข้อมูลเปิดภาครัฐ
Open Government Data Service Platform (2)

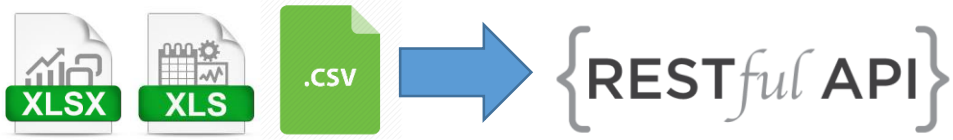
- **<u>หน่วยงานที่เปิดเผยข้อมูล</u>** สามารถสร้าง API ได้เองจากไฟล์ชุดข้อมูลเปิดที่มีอยู่ โดยไม่ต้องเขียนโปรแกรม
- **<u>นักพัฒนา</u>** สามารถนำข้อมูลเปิดภาครัฐไปพัฒนาโปรแกรมประยุกต์ได้สะดวกยิ่งขึ้นผ่าน API
- **<u>ผู้ใช้</u>** สามารถสืบค้นข้อมูลและสร้างรายงานจากไฟล์ชุดข้อมูลโดยไม่ต้องใช้โปรแกรมภายนอก

A Driving Force for National Science and Technology Capability

ประโยชน์ต่อภาครัฐด้านการประหยัดงบประมาณ

การผลิตทีละชิ้น (Tailor-made Production)	การผลิตแบบโรงงาน (Factory Production)
• ผลิตได้จำนวนน้อย (Low volume) • ค่าใช้จ่ายในการผลิตต่อชิ้นสูง (High-cost per unit)	• ผลิตได้จำนวนมาก (High volume) • ค่าใช้จ่ายในการผลิตต่อชิ้นต่ำ (Low-cost per unit)
	
• พัฒนาโดยโปรแกรมเมอร์ (Manual construction) • ผลิตได้น้อย (Low volume) • ค่าใช้จ่ายการพัฒนาต่อหน่วยสูง (High-cost per unit)	• สร้างโดยอัตโนมัติ (Automatic construction) • ผลิตได้มาก (High volume) • ค่าใช้จ่ายการพัฒนาต่อหน่วยต่ำ (Low-cost per unit)

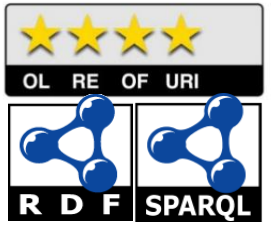
ฟังก์ชันการทำงาน



- สร้าง API อย่างอัตโนมัติจากไฟล์ตารางคำนวณ (xls, xlsx, csv)
- ตรวจสอบว่าไฟล์ชุดข้อมูลอยู่ในรูปแบบตารางที่ถูกต้องหรือไม่ (Dataset Validation)

- เป็น API ชนิด RESTful Web APIs
  - พัฒนาโปรแกรมด้วยภาษาใดก็ได้
  - สะดวกต่อการนำไปพัฒนาเป็น Mobile App

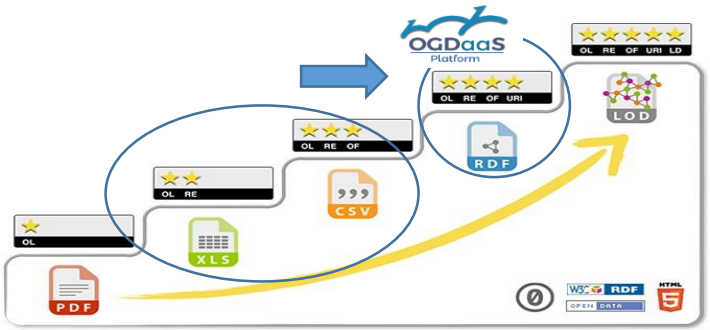
ฟังก์ชันการทำงาน (2)



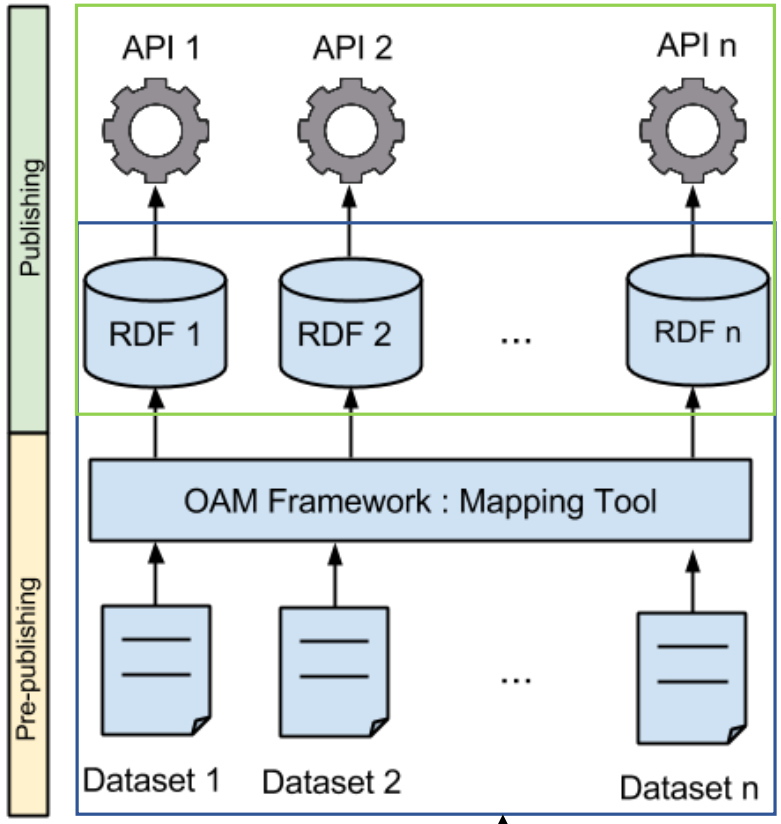
- แปลงชุดข้อมูลให้อยู่ในแบบ 4 ดาว (4-Star Open Data Compliant)
- ผู้ใช้สืบค้นและสร้างรายงานจากข้อมูลได้

- แบ่งปันผลการสืบค้นหรือรายงานข้อมูลผ่าน Social Media -- “Crowdsourcing”
  - เพิ่มช่องทางการมีส่วนร่วมของประชาชนต่อข้อมูลภาครัฐ

5-Star Open Data (by Sir Tim Berners-Lee)



สถาปัตยกรรมระบบ



Sources:

Data go.th

Dataset catalog & dataset files

Dataset Validator

Valid datasets (well-formed tabular data)

Output data formats

RESTful API

RDF SPARQL OWL

Tools

semantic web framework Jena

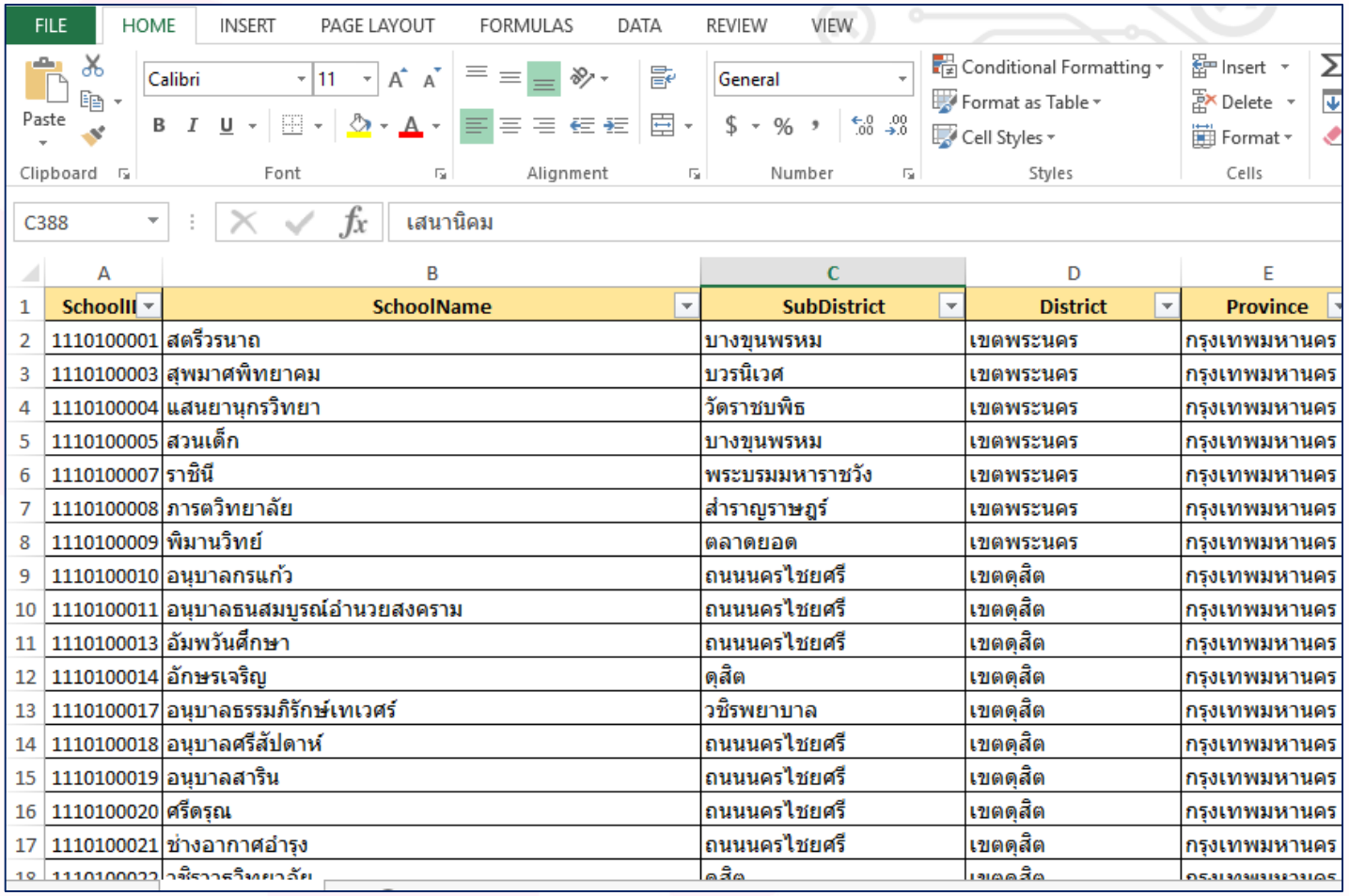
OPENLINK VIRTUOSO UNIVERSAL SERVER

D2RQ Accessing Relational Databases as Virtual RDF Graphs

Input data formats

XLSX XLS .CSV

ตัวอย่างไฟล์ชุดข้อมูล “ข้อมูลพื้นฐานสถานศึกษา ปี 2558” (กระทรวงศึกษาธิการ)



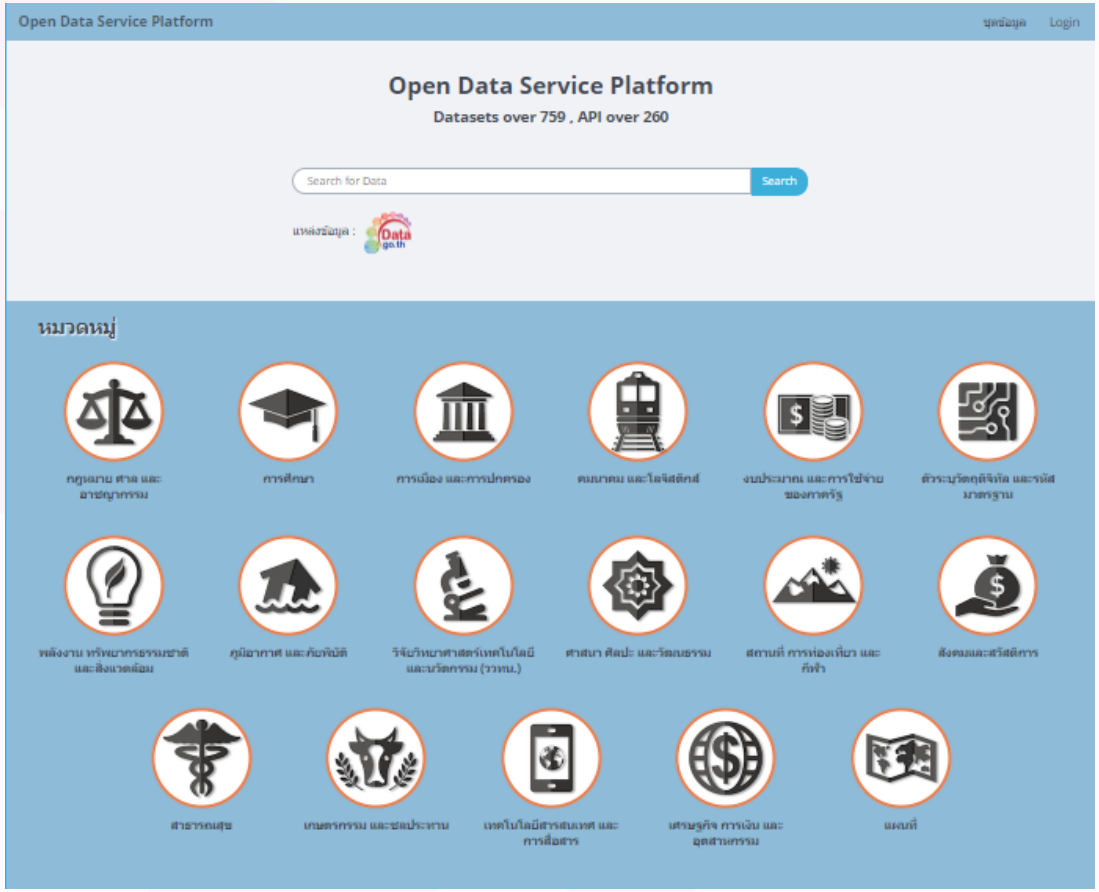
	A	B	C	D	E
1	SchoolI	SchoolName	SubDistrict	District	Province
2	1110100001	สตรีวรนาถ	บางขุนพรหม	เขตพระนคร	กรุงเทพมหานคร
3	1110100003	สุพมาศพิทยาคม	บวรนิเวศ	เขตพระนคร	กรุงเทพมหานคร
4	1110100004	แสนยานุกรวิทยา	วัดราชบพิธ	เขตพระนคร	กรุงเทพมหานคร
5	1110100005	สวนเด็ก	บางขุนพรหม	เขตพระนคร	กรุงเทพมหานคร
6	1110100007	ราชินี	พระบรมมหาราชวัง	เขตพระนคร	กรุงเทพมหานคร
7	1110100008	ภารตวิทยาลัย	สำราญราษฎร์	เขตพระนคร	กรุงเทพมหานคร
8	1110100009	พิมานวิทย์	ตลาดยอด	เขตพระนคร	กรุงเทพมหานคร
9	1110100010	อนุบาลกรแก้ว	ถนนนครไชยศรี	เขตดุสิต	กรุงเทพมหานคร
10	1110100011	อนุบาลธนสมบูรณ์อำนวยสงคราม	ถนนนครไชยศรี	เขตดุสิต	กรุงเทพมหานคร
11	1110100013	อัมพวันศึกษา	ถนนนครไชยศรี	เขตดุสิต	กรุงเทพมหานคร
12	1110100014	อักษรเจริญ	ดุสิต	เขตดุสิต	กรุงเทพมหานคร
13	1110100017	อนุบาลธรรมภิรักษ์เทเวศร์	วชิรพยาบาล	เขตดุสิต	กรุงเทพมหานคร
14	1110100018	อนุบาลศรีสัปดาห์	ถนนนครไชยศรี	เขตดุสิต	กรุงเทพมหานคร
15	1110100019	อนุบาลสาริน	ถนนนครไชยศรี	เขตดุสิต	กรุงเทพมหานคร
16	1110100020	ศรีดรุณ	ถนนนครไชยศรี	เขตดุสิต	กรุงเทพมหานคร
17	1110100021	ช่างอากาศอำรุง	ถนนนครไชยศรี	เขตดุสิต	กรุงเทพมหานคร

A Driving Force for National Science and Technology Capability

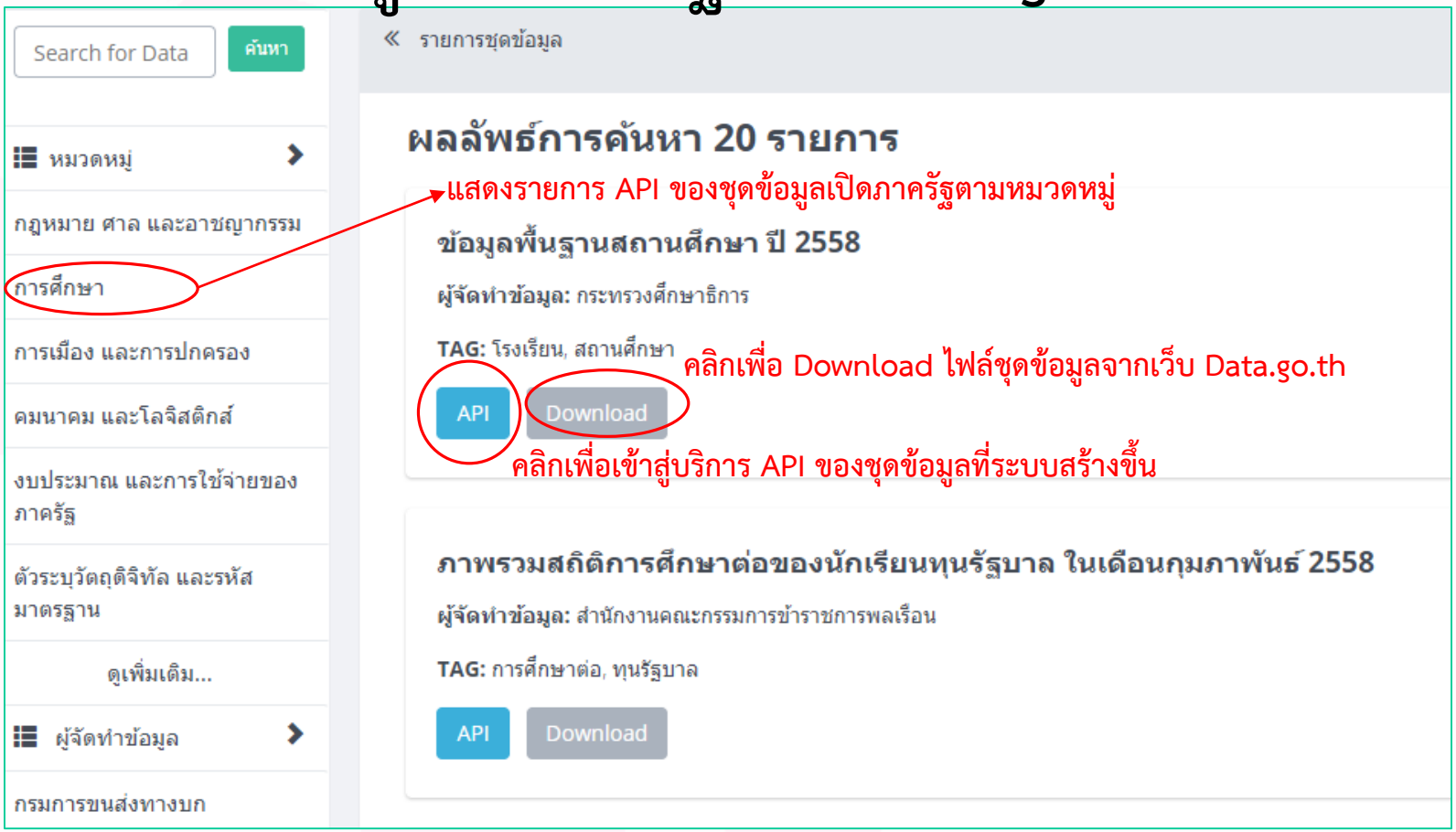
ระบบสร้าง API จากชุดข้อมูลเปิดภาครัฐ

Open Government Data Service Platform (3)

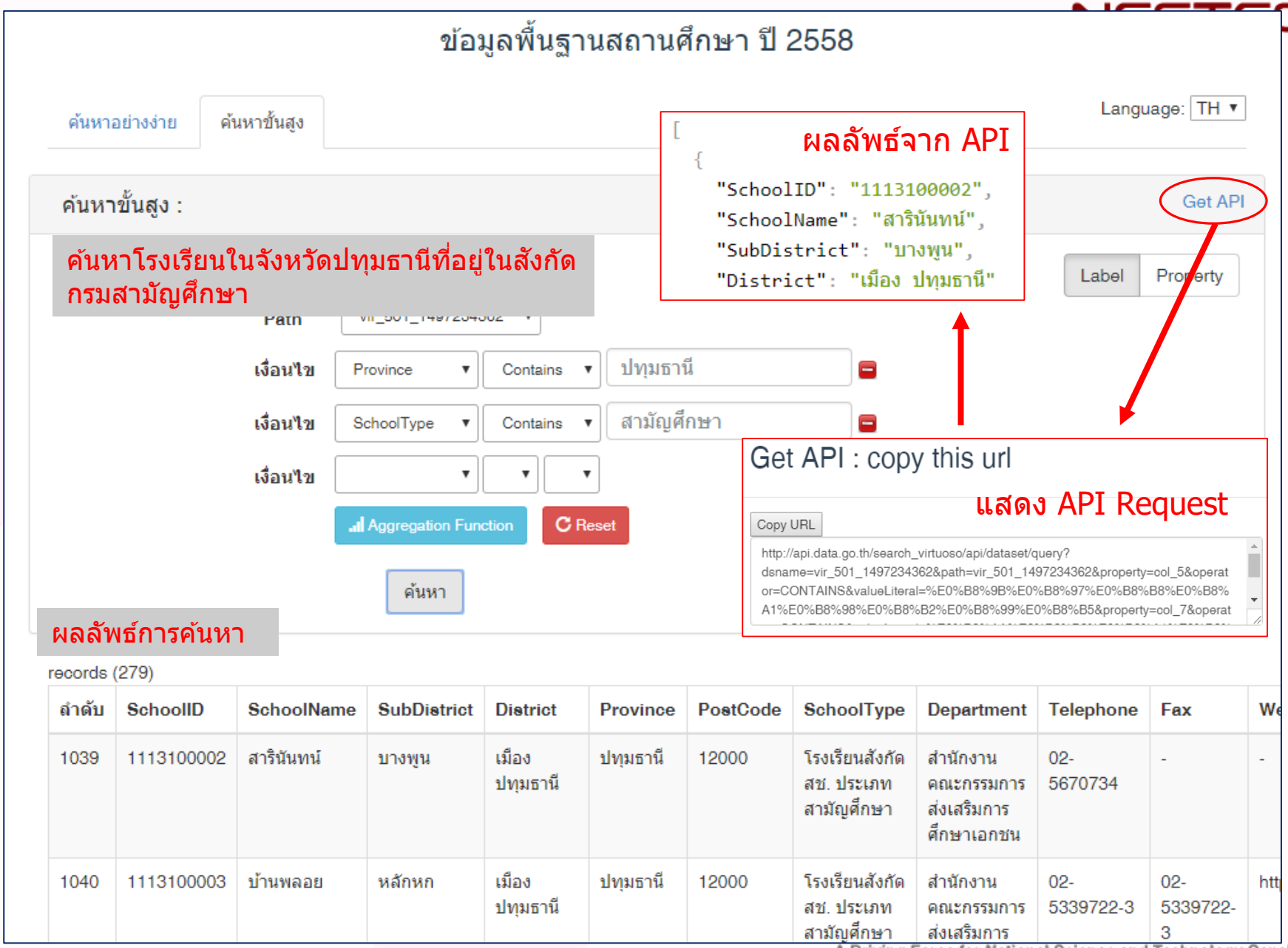
http://api.data.go.th/demo/



ตัวอย่างการใช้งาน - รายการ API ที่สร้างจากไฟล์ชุดข้อมูลเปิดภาครัฐจาก Data.go.th



ตัวอย่างการเข้าถึง API ของชุดข้อมูลเปิดภาครัฐเพื่อสืบค้นข้อมูล



ลำดับ	SchoolID	SchoolName	SubDistrict	District	Province	PostCode	SchoolType	Department	Telephone	Fax	We
1039	1113100002	สารินันทน์	บางพูน	เมือง ปทุมธานี	ปทุมธานี	12000	โรงเรียนสังกัด สช. ประเภท สามัญศึกษา	สำนักงานคณะกรรมการส่งเสริมการศึกษาเอกชน	02-5670734	-	-
1040	1113100003	บ้านพลอย	หลักหก	เมือง ปทุมธานี	ปทุมธานี	12000	โรงเรียนสังกัด สช. ประเภท สามัญศึกษา	สำนักงานคณะกรรมการส่งเสริมการ	02-5339722-3	02-5339722-3	htt

ตัวอย่างการเข้าถึง API ของชุดข้อมูลเปิดภาครัฐเพื่อแสดงรายงานสรุป

แสดงจำนวนโรงเรียนที่อยู่ในสังกัดกรมสามัญศึกษา ในแต่ละจังหวัด (เรียงลำดับจากมากไปน้อย)

เงื่อนไข SchoolType Contains สามัญศึกษา

เงื่อนไข

Aggregation Function | Reset

แสดง จำนวน SchoolName — จัดเรียงตาม มากไปหาน้อย

แสดง --select--

แบ่งตาม Province

แบ่งตาม --select--

ค้นหา

ผลลัพธ์การค้นหา

records (78)

Province	count(SchoolName)
นครราชสีมา	1499
อุบลราชธานี	1260
ขอนแก่น	1244
ศรีสะเกษ	953
บุรีรัมย์	949

ผลลัพธ์จาก API


```
[
  {
    "col_5": "นครราชสีมา",
    "col_2:count": "1499"
  },
  {
    "col_5": "อุบลราชธานี",
    "col_2:count": "1260"
  },
  {
    "col_5": "ขอนแก่น",
    "col_2:count": "1244"
  },
  {
    "col_5": "ศรีสะเกษ",
    "col_2:count": "953"
  },
```


A Driving Force for National Science and Technology Capability

ผลลัพธ์การสร้าง API ไฟล์ชุดข้อมูลจาก Data.go.th

- ประมวลผลไฟล์ชุดข้อมูลทั้งหมดจากเว็บ Data.go.th
  - ได้รับความอนุเคราะห์ข้อมูลจากสำนักงานรัฐบาลอิเล็กทรอนิกส์ (องค์การมหาชน) – EGA
  - จำนวนทั้งสิ้น 1,366 ไฟล์ชุดข้อมูลจาก 799 ชุดข้อมูล (ข้อมูล ณ. วันที่ 31 ตุลาคม 2559)
  - จำนวน API ของชุดข้อมูลที่มีอยู่ในระบบ มีจำนวน 267 API คิดเป็นสัดส่วนประมาณ 20% (267/1,366) ของไฟล์ทั้งหมด
- ไฟล์ชุดข้อมูลเปิดที่อยู่ในรูปแบบตารางที่ถูกต้อง จะถูกนำมาสร้างเป็น Web API ของชุดข้อมูลนั้นให้อย่างอัตโนมัติ

ตัวอย่างข้อมูลแบบตารางที่นำเข้าได้

	A	B	C	D	E	F	G	H	I
1	TAMBON_E	AM_ID	AMPHOE_T	AMPHOE_E	CH_ID	CHANGWAT_T	CHANGWAT_E	LAT	LONG
2	Ko Sarai	9101	อ. เมืองสตูล	Mueang Satun	91	จ. สตูล	Satun	6.546	99.706
3	Maptaphut	2101	อ. เมืองระยอง	Mueang Rayong	21	จ. ระยอง	Rayong	12.646	101.171
4	Maptaphut	2101	อ. เมืองระยอง	Mueang Rayong	21	จ. ระยอง	Rayong	12.645	101.170
5	Hat Sai Ri	8601	อ. เมืองชุมพร	Mueang Chumphon	86	จ. ชุมพร	Chumphon	10.459	99.403
6	Dan Sawi	8607	อ. สวี	Sawi	86	จ. ชุมพร	Chumphon	10.291	99.241
7	Tha Tum	4401	อ. เมืองมหาสารคาม	Mueang Maha Sarakham	44	จ. มหาสารคาม	Maha Sarakham	16.162	103.484

ตัวอย่างข้อมูลแบบตารางที่ไม่สามารถนำเข้าได้

	A	B	C
1			
2	ข้อมูลสถิติการจับกุมคดียาเสพติด ปีงบประมาณ 2558		
3	ปี	เดือน	จำนวน(ราย)
4	พ.ศ. 2557	ตุลาคม	14,724
5		พฤศจิกายน	14,877
6		ธันวาคม	14,527
7	พ.ศ. 2558	มกราคม	13,248
8		กุมภาพันธ์	15,399
9		มีนาคม	16,210
10		เมษายน	13,087
11		พฤษภาคม	14,340
12		มิถุนายน	15,449
13		กรกฎาคม	14,615
14		สิงหาคม	13,985
15		กันยายน	11463
16	ยอดรวม		171,924
17			
18	หมายเหตุ ข้อมูล ณ วันที่ 2 ตุลาคม 2558		
19	ที่มาของข้อมูล : ระบบ บช.ปส.ผ่านระบบ Polis		

T_metadata
มีคำอธิบายข้อมูล
อยู่เหนือหรือท้ายตาราง

พระภิกษุ และสามเณร เข้าสอบนักธรรม ปีการศึกษา 2556			
ประเภท	ผู้เรียน	เข้าสอบ	สอบได้
[illegible]ชั้นตรี	106,888	84,117	62,469
[illegible]ชั้นโท	46,325	32,431	17,554
[illegible]ชั้นเอก	17,379	13,229	8,937
รวม	170,592	129,777	88,960

ผู้เข้าสอบธรรมศึกษา ปีการศึกษา 2556			
ประเภท	ผู้เรียน	เข้าสอบ	สอบได้
ธรรมศึกษาชั้นตรี	956,943	755,642	502,987
ธรรมศึกษาชั้นโท	461,145	374,464	212,351
ธรรมศึกษาชั้นเอก	191,017	150,446	97,380
รวม	1,609,105	1,280,552	812,718

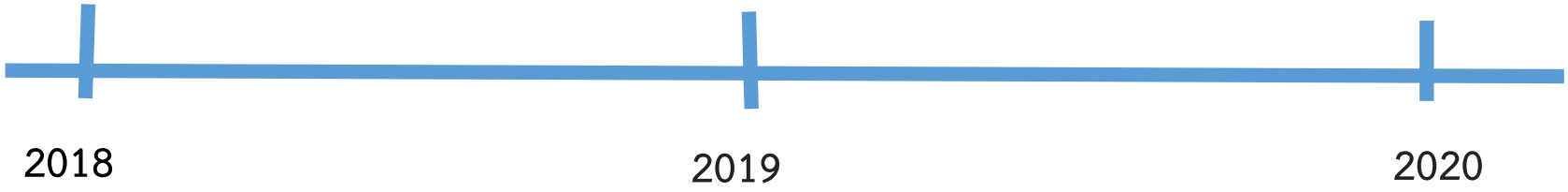
T_multiple
ข้อมูลหลายตารางที่มี
โครงสร้างต่างกัน
ถูกรวมเป็นตารางเดียวกัน

แผนการพัฒนาเทคโนโลยี (Technology Roadmap)

- เพิ่มจำนวน API ของชุดข้อมูล (ปรับปรุง data quality ของชุดข้อมูล)
- เพิ่มการแสดงผลรายงานของข้อมูลในแบบกราฟและแผนที่ (data visualization)

- พัฒนาเวอร์ชันสำหรับการใช้งานภายในองค์กร (Intranet version)
- เพิ่มเสถียรภาพการให้บริการ API สำหรับการใช้งานทางธุรกิจ (Premium-grade API for Startups)

- พัฒนาแพลตฟอร์มการเชื่อมโยงข้อมูลแบบเปิด (Linked Open Government Data Platform – 5-Star Open Data)



สรุปจุดเด่นของ Open Data Service Platform

- ช่วยสร้างคุณค่าเพิ่มให้กับข้อมูลเปิดภาครัฐ:
  - ภาครัฐสามารถสร้างบริการเข้าถึงข้อมูล เช่น API จำนวนมาก โดยไม่ต้องลงทุนสูง (high volume of data services with lower cost)
  - ช่วยเพิ่มโอกาสและ ช่องทางการเข้าถึงข้อมูลเปิดภาครัฐสำหรับประชาชน ผ่าน Social Media และ App ต่างๆ ที่เรียกใช้ข้อมูลผ่าน API