กระบวนการจัดการข้อมูลในรูปแบบ Open Government Data โดยใช้มาตรฐาน RDF

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

หัวข้อนำเสนอ

- การเปิดเผยข้อมูล 5 ระดับ (5-star Open Data)
  - Open Data ระดับที่ 1-3
  - Open Data ระดับที่ 4 และ 5
- กระบวนการจัดการ Open Government Data ด้วยมาตรฐาน RDF
  - ขั้นตอนการจัดการชุดข้อมูล
- ต้นแบบและตัวอย่างการประยุกต์ใช้งาน Open Government Data ด้วยมาตรฐาน RDF
  - demo-api.data.go.th

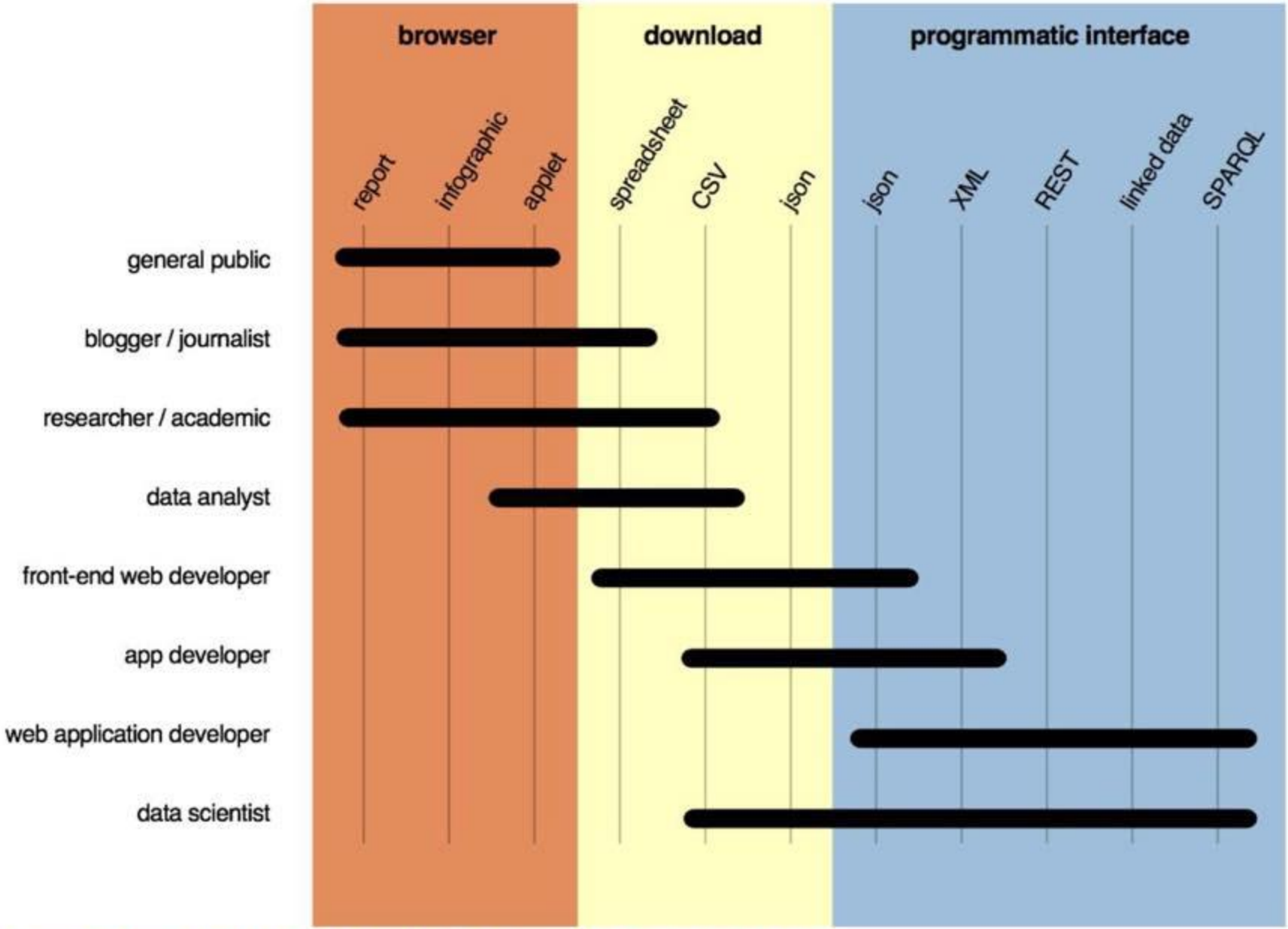
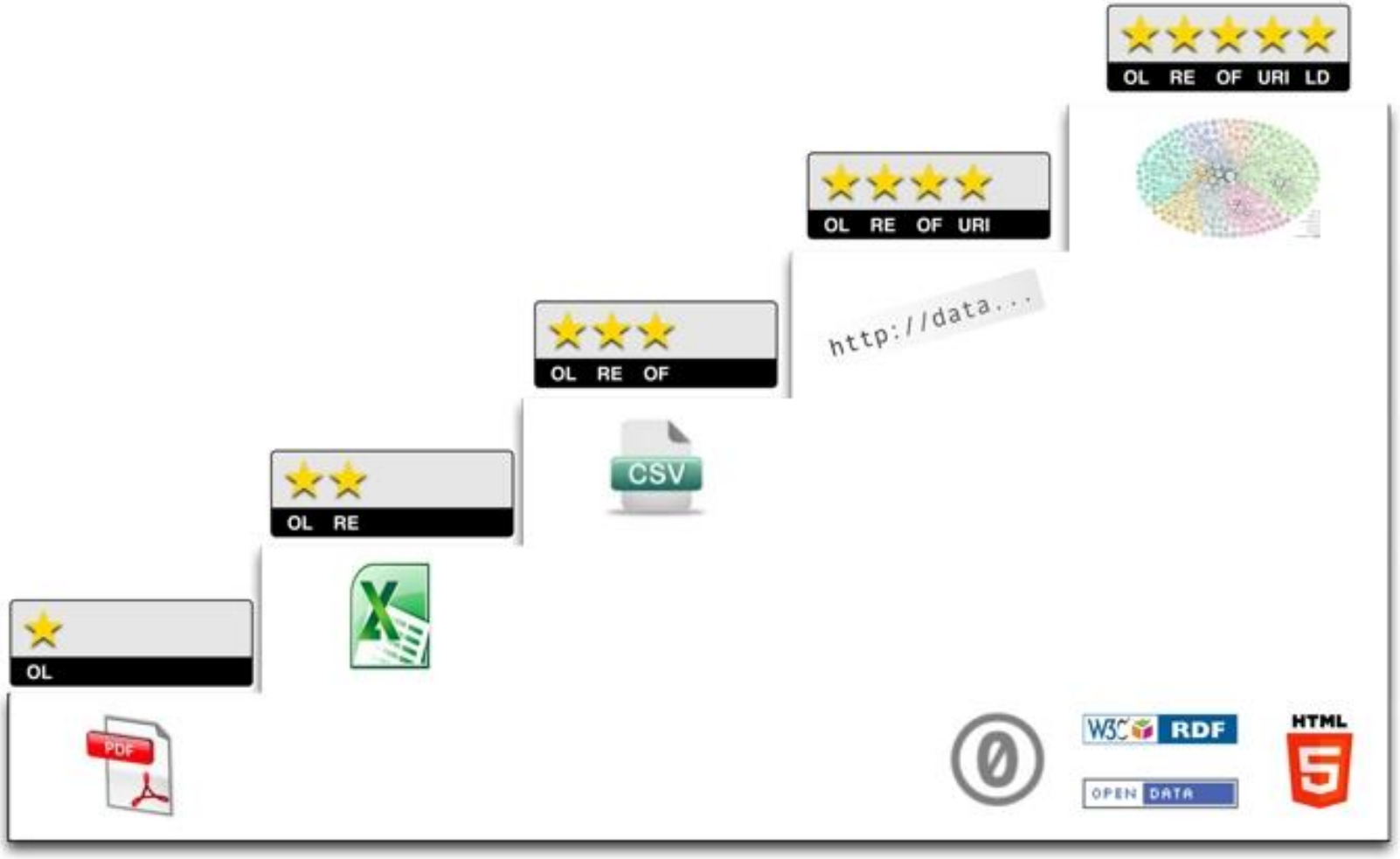


Figure 2 Publishing open data for multiple audiences

การเปิดเผยข้อมูล 5 ระดับ
(5-star Open Data)

การเปิดเผยข้อมูล 5 ระดับ



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)

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 = Reusable

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 = Reusable
OF = Open Format

การเปิดเผยข้อมูล ระดับที่ 3

ข้อดี

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

ข้อเสีย

- ยากต่อการบูรณาการเชื่อมโยงข้อมูลข้ามชุดข้อมูล

ตัวอย่างการเปิดเผยข้อมูล ระดับที่ 4

```
@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 = Reusable
OF = Open Format
URI = Uniform Resource Identifier

การเปิดเผยข้อมูล ระดับที่ 4

ข้อดี

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

ข้อเสีย

- สร้างได้ยากกว่าระดับที่ 1-3

ตัวอย่างการเปิดเผยข้อมูล ระดับที่ 5

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

		URI
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".

อ้างอิงข้อมูลในชุดข้อมูลอื่น

W3C RDF

employee.rdf

OL = Open License
RE = Reusable
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

การเปิดเผยข้อมูล ระดับที่ 5

ข้อดี

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

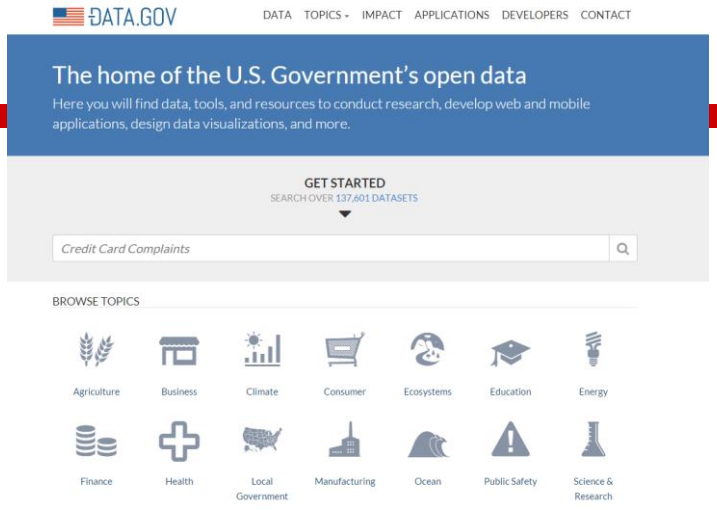
ข้อเสีย

- สร้างได้ยากกว่าระดับที่ 1-3
- ต้องมีมาตรฐานรหัสบ่งชี้ (standard identifier)

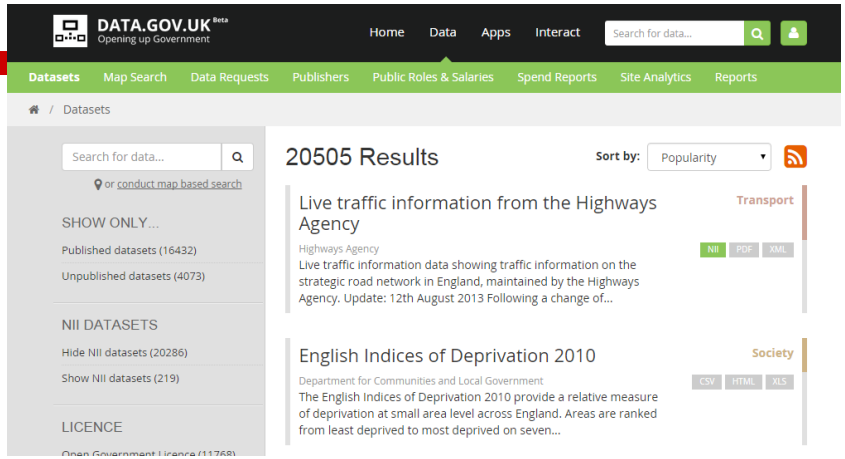
ประโยชน์ของ open data ระดับที่ 4-5

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

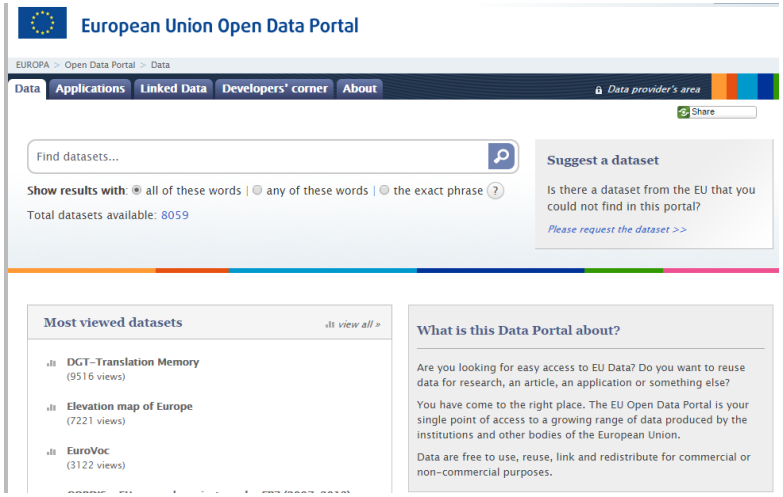
Open Government Data ในปัจจุบัน



Data.gov



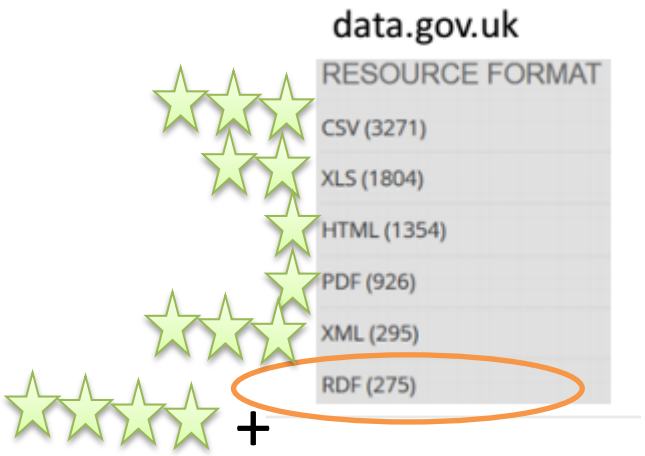
Data.gov.uk

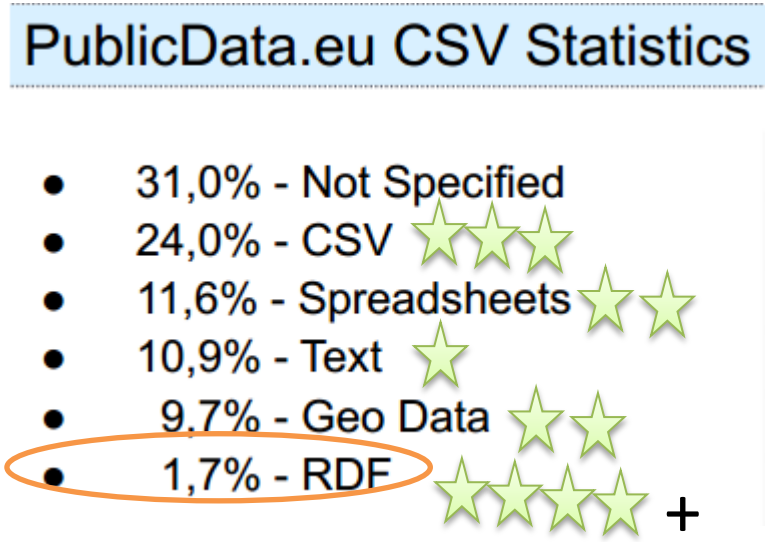


Open-data.europa.eu

Data.go.jp

สถิติของรูปแบบข้อมูลที่เปิดเผยข้อมูลบน Government Open Data Portal





ชุดข้อมูลแบบ RDF มีปริมาณ < 5%

References:
- *Sören Auer, http://www.ega.or.th/TrainingContent.aspx?CourseID=178*
- *Dumitru Roman,* https://sites.google.com/site/2014ldop/

สถิติของรูปแบบข้อมูลที่เปิดเผยข้อมูลบน Government Open Data Portal (2)

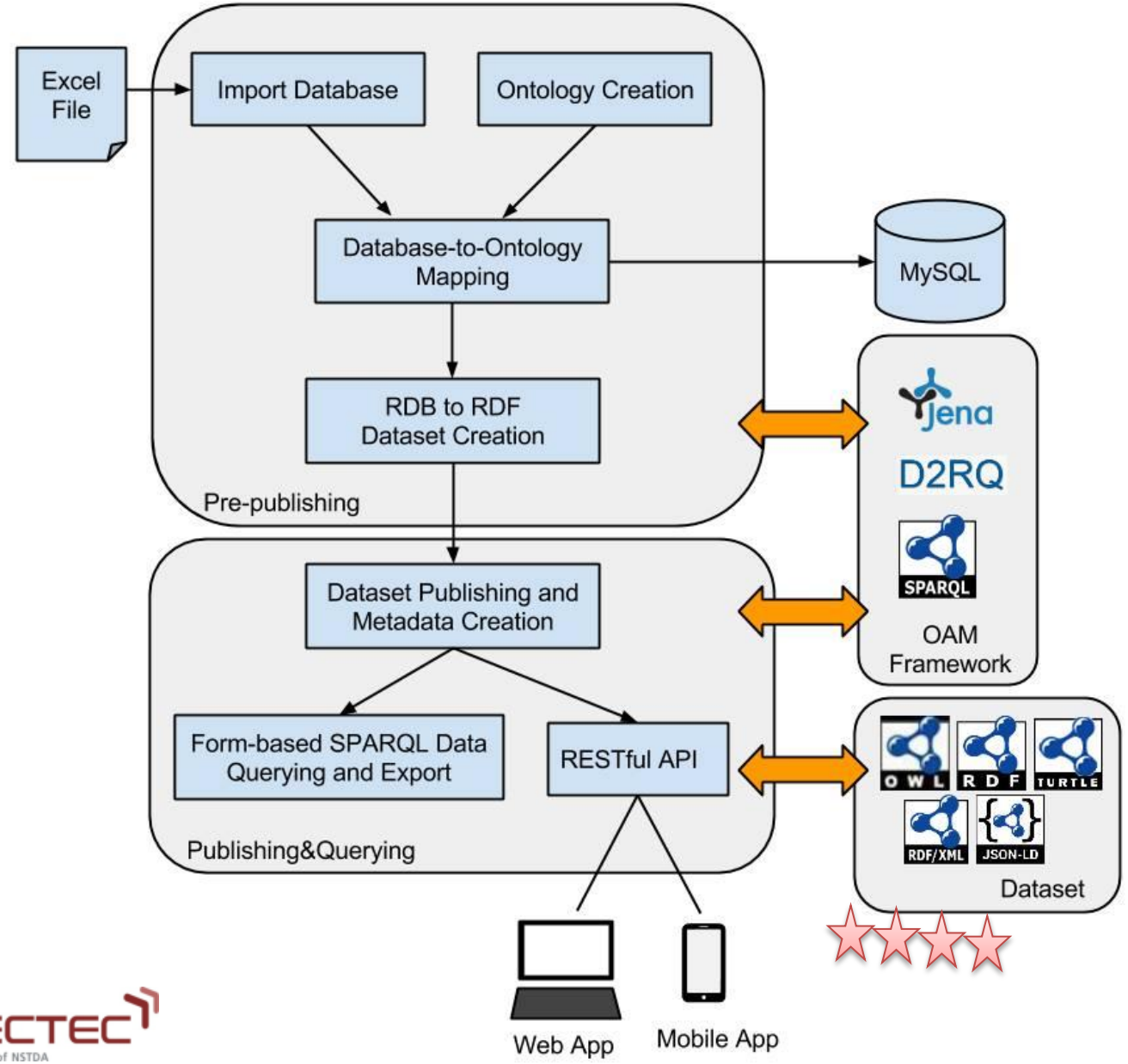
Open Data Portal	Datasets	Applications
data.gov	~ 110 000	~ 350
publicdata.eu	~ 50 000	~ 80
data.gov.uk	~ 20 000	~ 350
data.norge.no	~ 300	~ 40

- **Application** ยังมีน้อยเมื่อเทียบกับจำนวนชุดข้อมูล
- การใช้ประโยชน์จาก **Government Open Data Portal** ยังทำได้ยาก

References:
- *Sören Auer, http://www.ega.or.th/TrainingContent.aspx?CourseID=178*
- *Dumitru Roman,* https://sites.google.com/site/2014ldop/

กระบวนการจัดการ Open Government Data ด้วยมาตรฐาน RDF

กระบวนการจัดการชุดข้อมูลในแบบ RDF สำหรับชุดข้อมูลจาก data.go.th



โปรแกรมเครื่องมือสำหรับจัดการชุดข้อมูล Open Data ชนิด RDF

สร้างชุดข้อมูลชนิด RDF (Dataset Publishing)

- โปรแกรมสร้างชุดข้อมูล (RDF Dataset Creation Tool)
- เว็บท่าศูนย์รวมชุดข้อมูล (RDF Dataset Portal)

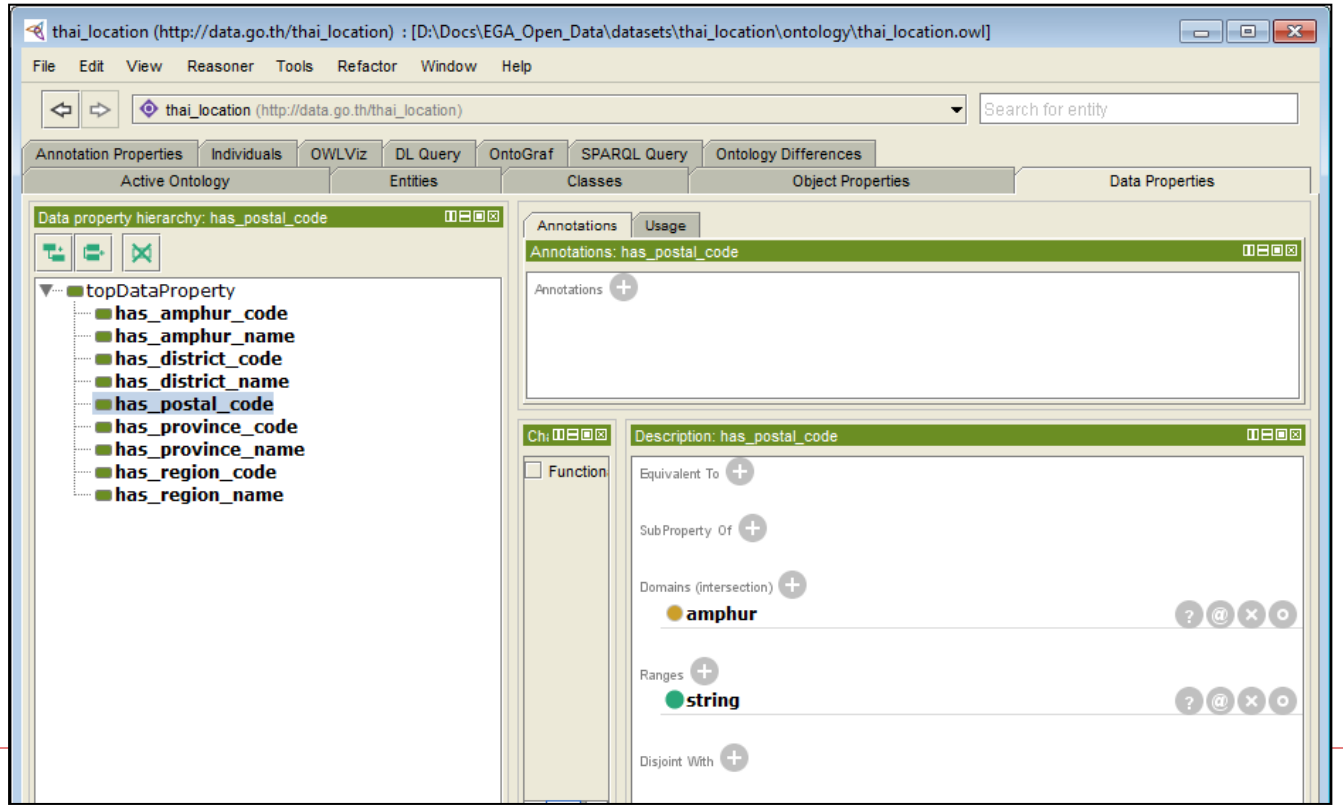
ใช้ประโยชน์จากชุดข้อมูลชนิด RDF (Dataset Application)

- โปรแกรมสืบค้นชุดข้อมูล (RDF Data Browsing/ Search)
- API สำหรับชุดข้อมูล (RDF Dataset API)

การสร้างชุดข้อมูลชนิด RDF

โปรแกรมสร้างชุดข้อมูล (RDF Dataset Creation Tool)

ขั้นตอนที่ 1 - กำหนดโครงสร้างของข้อมูลในแบบมาตรฐาน OWL (Web Ontology Language) โดยใช้โปรแกรม Ontology Editor



Protégé - http://protege.stanford.edu/

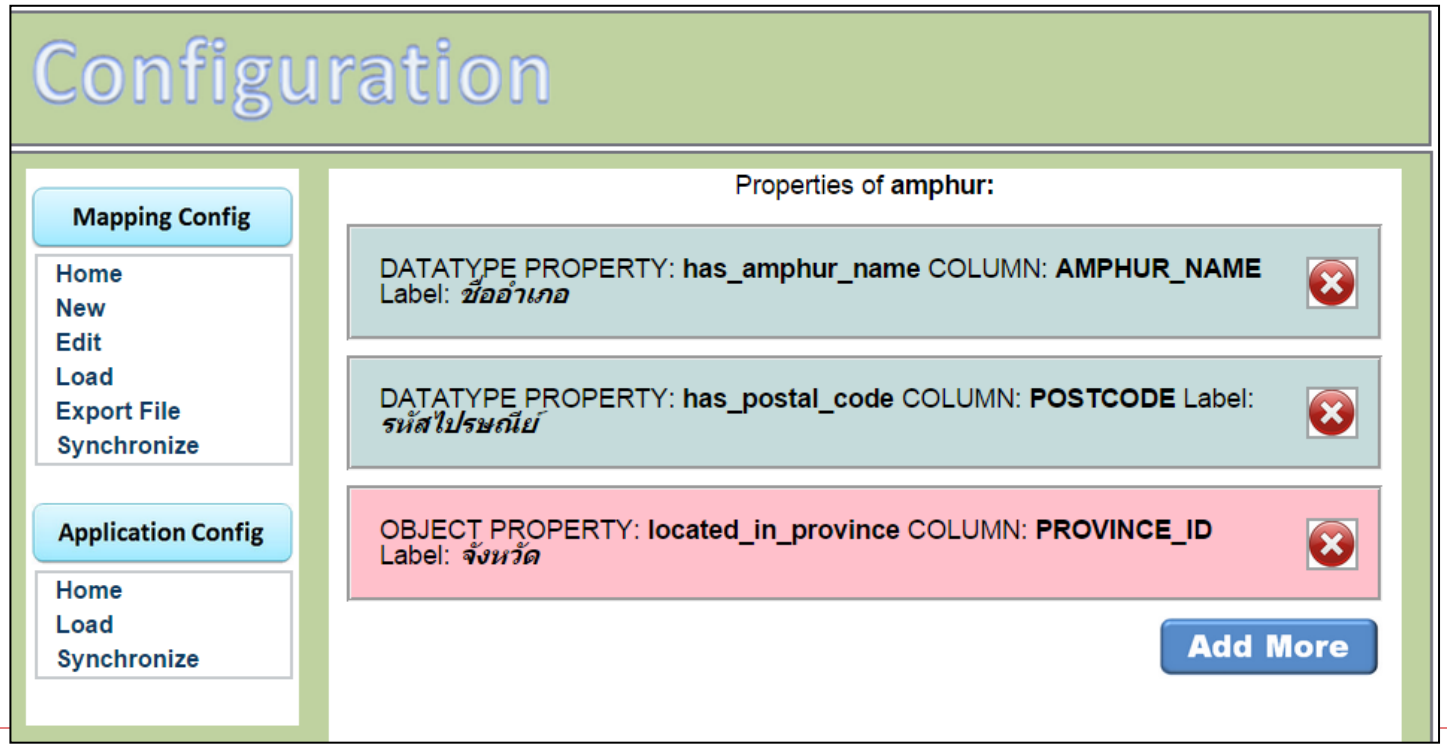
โปรแกรมสร้างชุดข้อมูล (RDF Dataset Creation Tool)

ขั้นตอนที่ 2 - จัดเตรียมชุดข้อมูลในแบบตารางฐานข้อมูล

ชื่อจังหวัด	ชื่อจังหวัดถอดเป็นอักษรโรมัน	รหัสจังหวัด
กระบี่	Krabi	81
กรุงเทพมหานคร	Krung Thep Maha Nakhon	10
กาญจนบุรี	Kanchanaburi	71
กาฬสินธุ์	Kalasin	46
กำแพงเพชร	Kamphaeng Phet	62
ขอนแก่น	Khon Kaen	40
จันทบุรี	Chanthaburi	22
ฉะเชิงเทรา	Chachoengsao	24

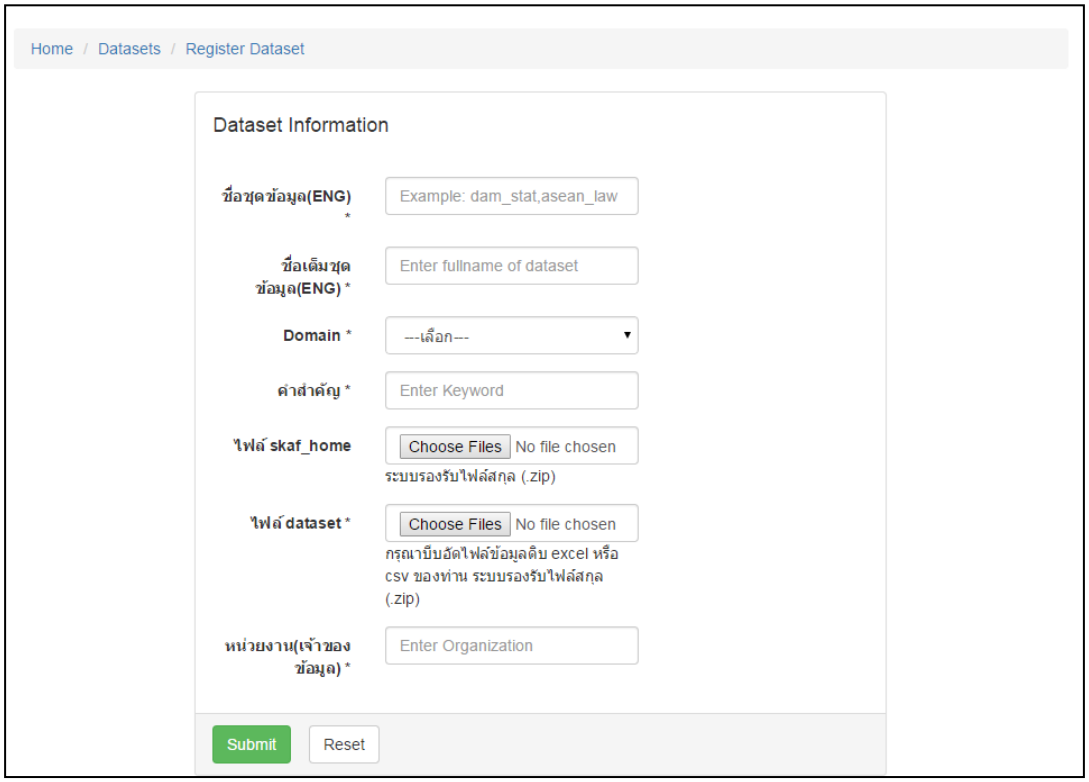
โปรแกรมสร้างชุดข้อมูล (RDF Dataset Creation Tool)

ขั้นตอนที่ 3 – เชื่อมโยงโครงสร้างข้อมูลแบบ OWL กับโครงสร้างฐานข้อมูล และสร้างชุดข้อมูลแบบ RDF



เว็บท่าศูนย์รวมชุดข้อมูล (RDF Dataset Portal)

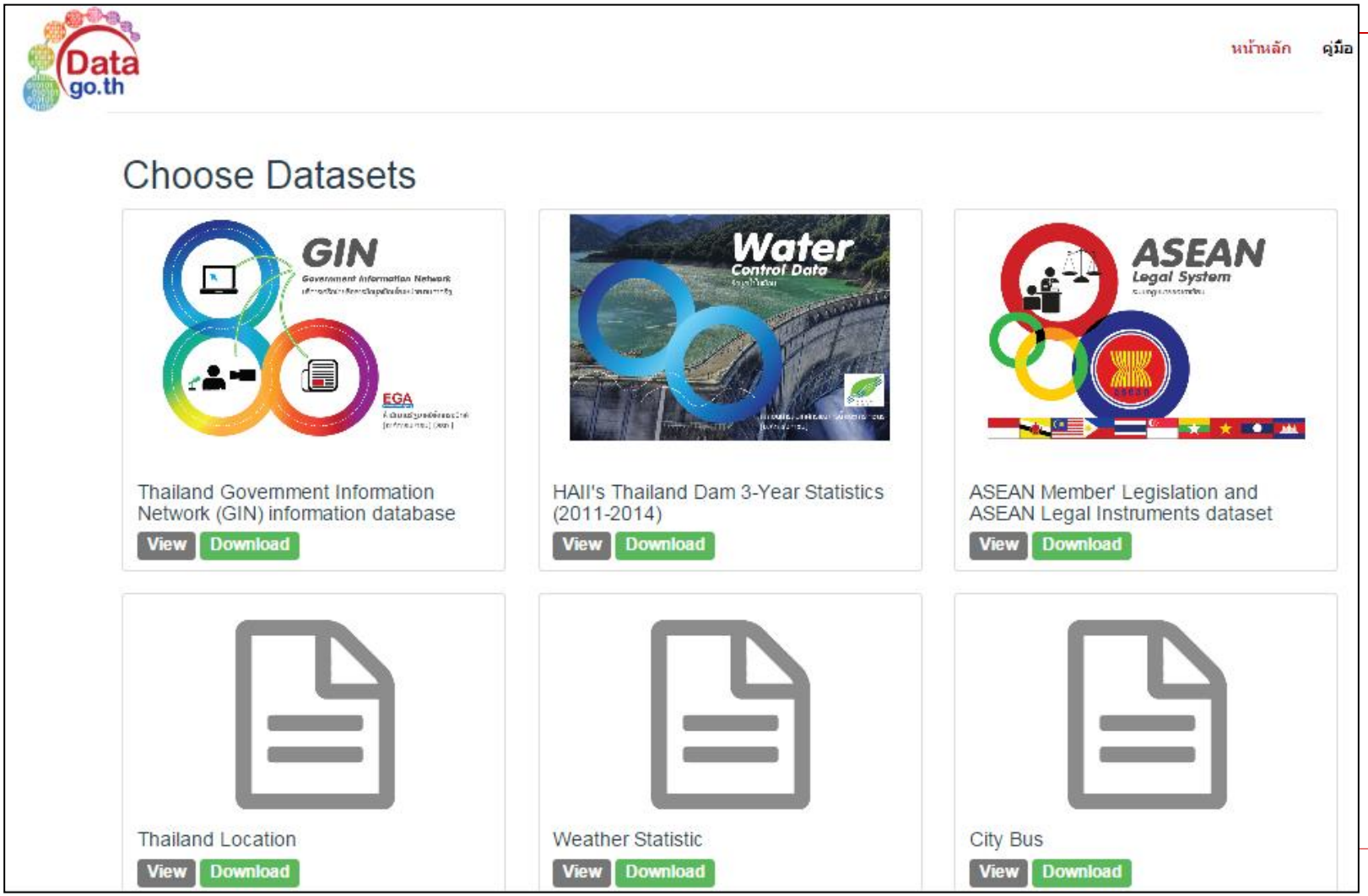
ขั้นตอนที่ 4 – Upload ไฟล์ชุดข้อมูลและกำหนดเมทาเดทาของชุดข้อมูล



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

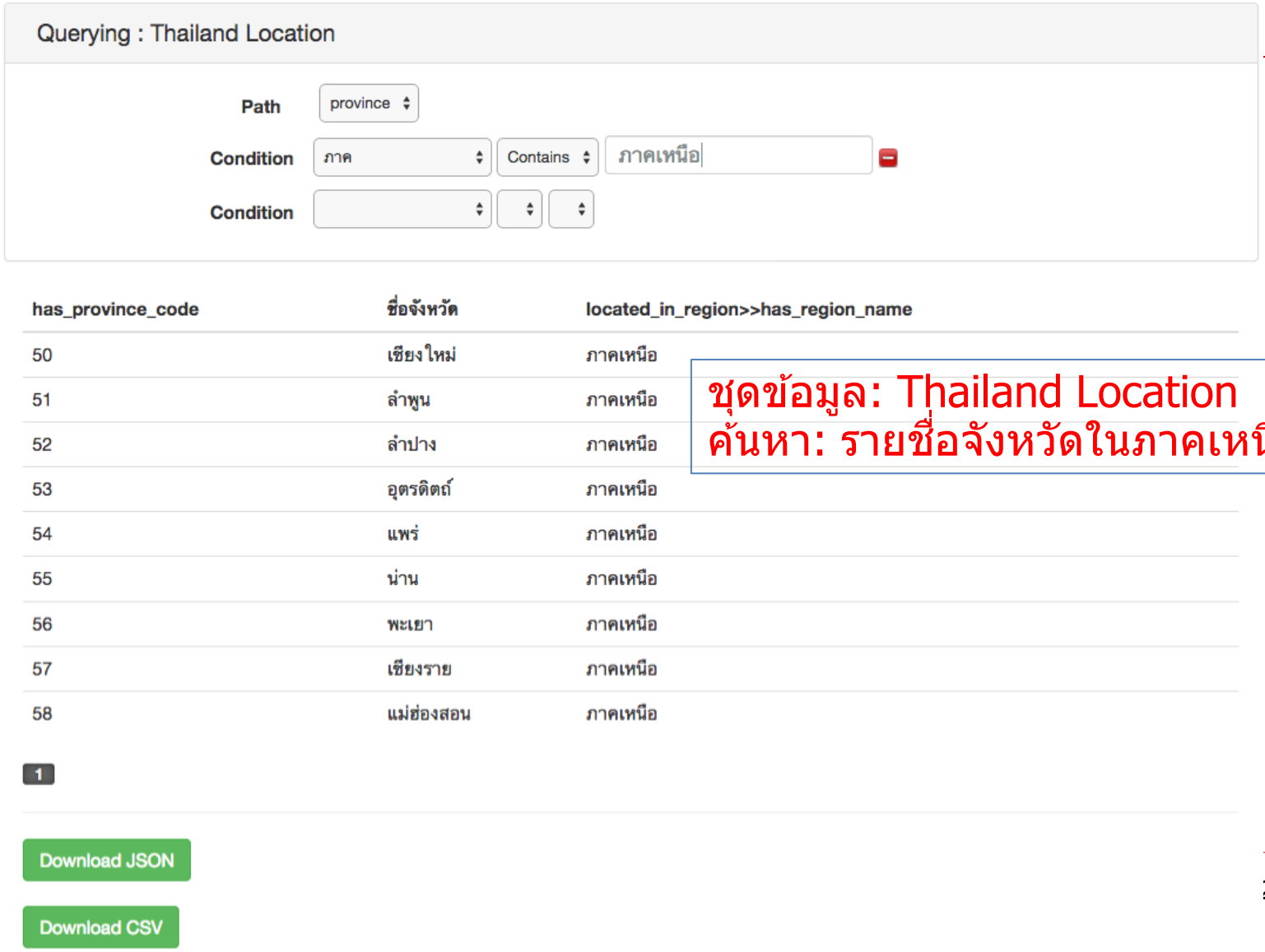
Open Data Portal สำหรับชุดข้อมูลชนิด RDF (Beta)

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



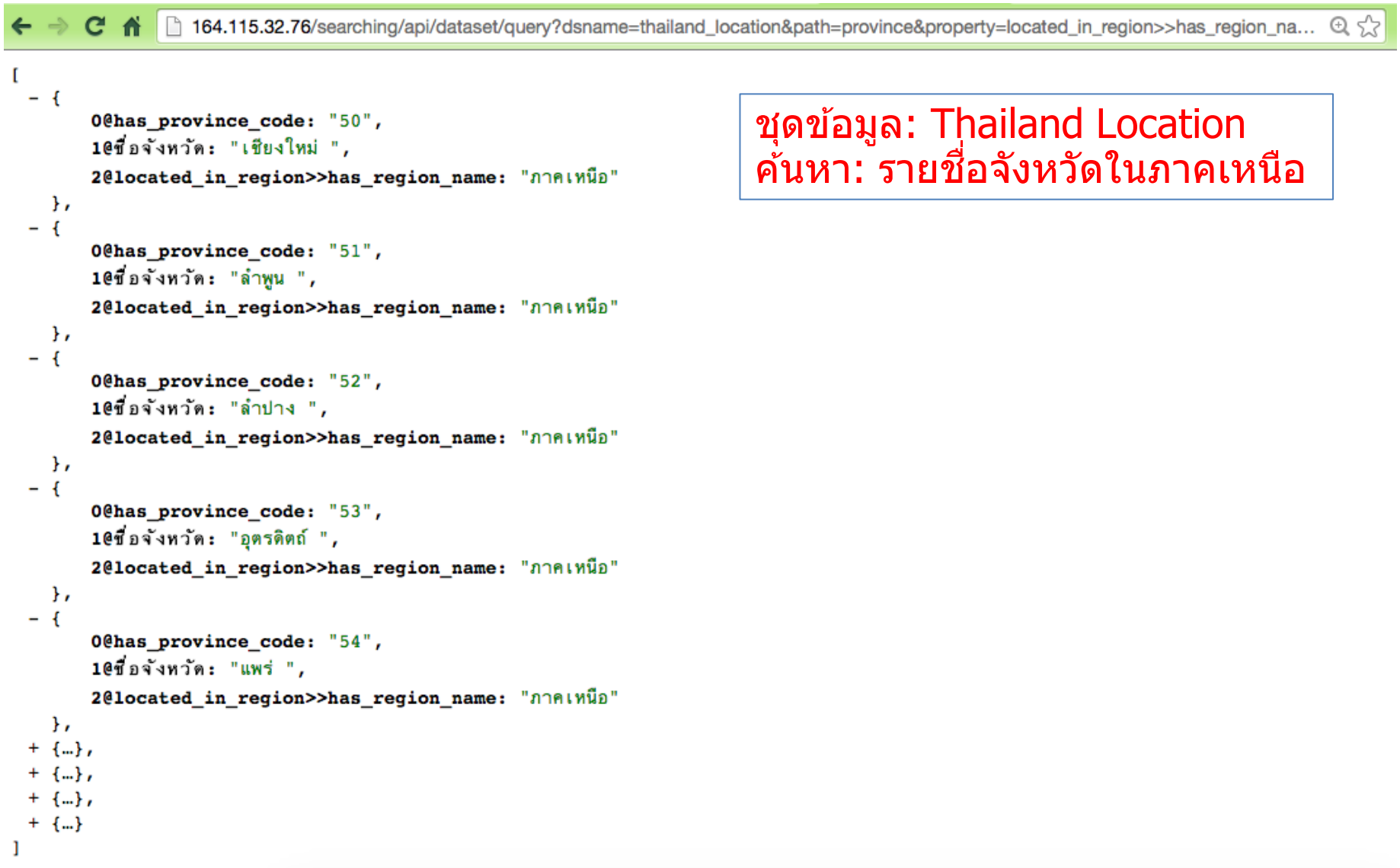
การใช้ประโยชน์ชุดข้อมูลชนิด RDF

โปรแกรมสืบค้นชุดข้อมูล (RDF Data Browsing/ Search)



ชุดข้อมูล: Thailand Location
ค้นหา: รายชื่อจังหวัดในภาคเหนือ

API สำหรับชุดข้อมูล (RDF Dataset API)



ชุดข้อมูล: Thailand Location
ค้นหา: รายชื่อจังหวัดในภาคเหนือ

ตัวอย่างการค้นหาข้อมูลจากชุดข้อมูลผ่าน Web API

ข้อมูลสายรถเมล์

1. ค้นหาข้อมูลป้ายรถเมล์บนถนนเยาวราช

Dataset : city_bus

API: /searching/api/dataset/query?dsname=city_bus&path=city_bus&
property=road&operator=CONTAINS&value=เยาวราช

2. ค้นหาข้อมูลป้ายรถเมล์ที่รถเมล์สาย1 วิ่งผ่านบนถนนเจริญกรุง

Dataset : city_bus

API: /searching/api/dataset/query?dsname=city_bus&path=city_bus&
property=Number_bus&operator=EQUALS&value=1&property=road&
operator=CONTAINS&value=เจริญกรุง

ข้อมูลระดับน้ำในเขื่อน

3. ค้นหาข้อมูลระดับน้ำในเขื่อนลำตะคองวันที่มีระดับน้ำที่ใช้ได้น้อยกว่าร้อยละ 30

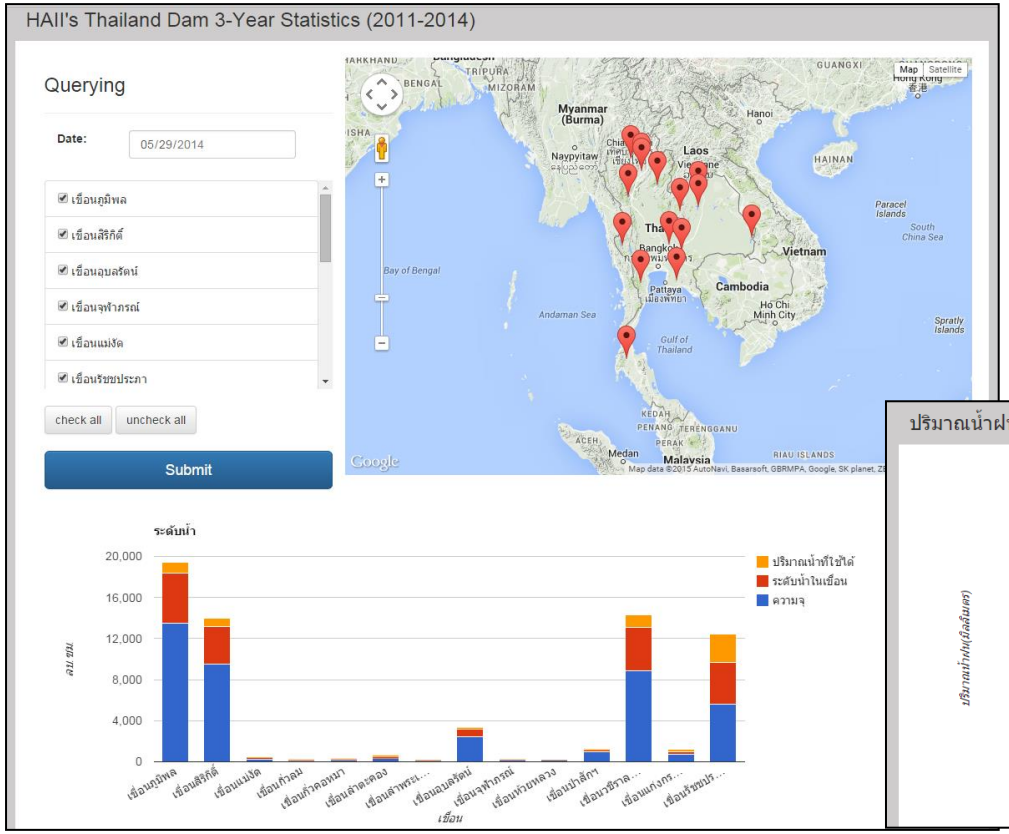
Dataset : dam_stat

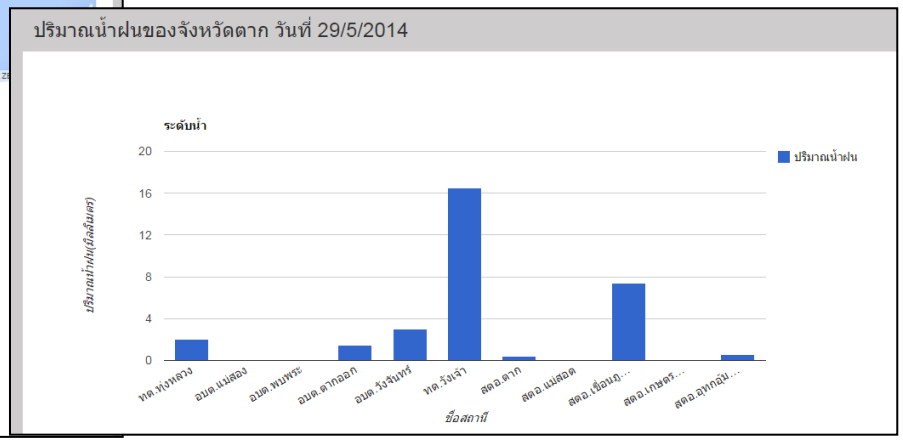
API: /searching/api/dataset/query?dsname=dam_stat&path=DamStat&
property=damName&operator=CONTAINS&value=ลำตะคอง&
property=usablePercentage&opeator=LT&value=30

OpenData Hackathon#2

31 ม.ค. – 1 ก.พ. 2558

ตัวอย่าง Demo Application ของ demo-api.data.go.th





รายงานระดับน้ำในอ่างเก็บน้ำขนาดใหญ่ทั่วประเทศ

ชุดข้อมูล: dam_stat

รายงานปริมาณน้ำฝนในแต่ละพื้นที่

ชุดข้อมูล: rainfall_stat

ประโยชน์ของการสร้างข้อมูลบน demo-api.data.go.th

- เพิ่มคุณค่าให้กับชุดข้อมูลที่มีอยู่
  - สร้างเป็นชุดข้อมูล Open Data ระดับ 4 ดาว (ข้อมูลชนิด RDF)
  - ผู้ใช้สามารถใช้ประโยชน์จากชุดข้อมูลในการสืบค้นชุดข้อมูลได้
  - นักพัฒนาแอพพลิเคชันสามารถใช้ประโยชน์และบูรณาการข้อมูลในชุดข้อมูลต่างๆ ผ่าน API (Restful Web API)

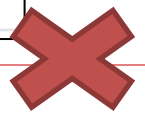
Web API vs. ไฟล์ชุดข้อมูล (Full Dataset)

- 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

ข้อจำกัดในปัจจุบัน

- การสร้างไฟล์ชุดข้อมูลชนิด RDF มีความซับซ้อนกว่าการสร้างชุดข้อมูลชนิด .XLS หรือ .CSV
- ไฟล์ชุดข้อมูลที่จะสามารถแปลงเป็นข้อมูลชนิด RDF ต้องมีคุณภาพดี และมีโครงสร้างข้อมูลในแบบของฐานข้อมูล
  - ต้องมีขั้นตอนการปรับรูปแบบของชุดข้อมูล (data reformatting) ก่อน

	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				



ก้าวสู่ข้อมูล Open Data ระดับ 5 ดาว (Linked Data)

- ต้องมีข้อตกลงเรื่องมาตรฐานข้อมูลในแต่ละสาขา
  - มาตรฐานเมตาเดตา (metadata) สำหรับข้อมูลสาขาต่างๆ เช่น ข้อมูลบุคคล ข้อมูลหน่วยงาน ข้อมูลสภาพอากาศ ข้อมูลสถานที่ ฯลฯ
  - รหัสบ่งชี้มาตรฐาน (identifier) เช่น รหัสสถานที่ ถนน ตำบล อำเภอ จังหวัด รหัสหน่วยงาน รหัสที่ใช้บ่งชี้บุคคล ฯลฯ
- เพื่อนำไปสู่การเชื่อมโยงข้อมูลแบบเปิดของภาครัฐ (Linked Open Government Data) ได้ต่อไป

สรุป

- การเปิดเผยข้อมูลสามารถทำได้ในหลายรูปแบบไฟล์ (การเปิดเผยข้อมูล 5 ระดับ)
  - แต่ละระดับต่างกันที่ความง่ายต่อการนำไปพัฒนาต่อยอด (reusability)
- Demo-api.data.go.th เป็นการนำร่องการสร้างชุดข้อมูลในแบบที่สามารถนำไปใช้ประโยชน์ได้ง่ายยิ่งขึ้นสำหรับนักพัฒนา
  - Application ของชุดข้อมูล RDF และ Restful Web API