

แนวปฏิบัติการเปิดเผยข้อมูลภาครัฐ (Publishing Open Government Data)

ดร.มารุต บุณรัช

กลุ่มวิจัยวิทยาการข้อมูลและการวิเคราะห์

ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ (NECTEC)

สำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ (สวทช.)

หัวข้อนำเสนอ

- แนวทางการจัดทำข้อมูลเปิด (Publishing Open Data)
 - รูปแบบข้อมูลที่เครื่องอ่านได้ (machine-readable data)
- แนวทางการจัดทำชุดข้อมูลจากตารางฐานข้อมูล
 - รูปแบบข้อมูลแบบตารางฐานข้อมูล
 - การประมวลผลและจัดเตรียมชุดข้อมูลจากตารางฐานข้อมูล

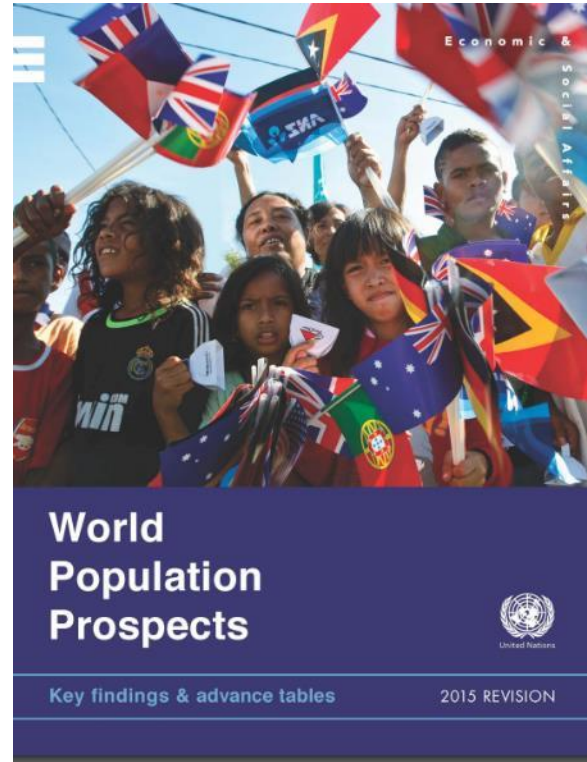
การเปิดเผยข้อมูลต่างจากการเผยแพร่ข้อมูลผ่านเว็บไซต์อย่างไร ?

การเปิดเผยข้อมูลในแบบรายงาน

การเปิดเผยข้อมูล
ในรูปแบบรายงาน
(pdf)



ข้อมูลอยู่ในแบบที่
นำไปใช้ซ้ำได้ยาก
(non machine-
readable)



WORLD POPULATION PROSPECTS: THE 2015 REVISION

SUMMARY AND KEY FINDINGS

Understanding the demographic changes that are likely to unfold over the coming years, as well as the challenges and opportunities that they present for achieving sustainable development, is important for designing and implementing the post-2015 development agenda. The *2015 Revision of World Population Prospects* is the twenty-fourth round of official United Nations population estimates and projections that have been prepared by the Population Division of the Department of Economic and Social Affairs of the United Nations Secretariat. The *2015 Revision* builds on the previous revision by incorporating additional results from the 2010 round of national population censuses as well as findings from recent specialized demographic and health surveys that have been carried out around the world. The *2015 Revision* provides the demographic data and indicators to assess population trends at the global, regional and national levels and to calculate many other key indicators commonly used by the United Nations system.

Snapshot of global population in 2015

According to the results of the *2015 Revision*, the world population reached 7.3 billion as of mid-2015 (table 1), implying that the world has added approximately one billion people in the span of the last twelve years. Sixty per cent of the global population lives in Asia (4.4 billion), 16 per cent in Africa (1.2 billion), 10 per cent in Europe (738 million), 9 per cent in Latin America and the Caribbean (634 million), and the remaining 5 per cent in Northern America (358 million) and Oceania (39 million). China (1.4 billion) and India (1.3 billion) remain the two largest countries of the world, both with more than 1 billion people, representing 19 and 18 per cent of the world's population, respectively.

TABLE 1. POPULATION OF THE WORLD AND MAJOR AREAS, 2015, 2030, 2050 AND 2100, ACCORDING TO THE MEDIUM-VARIANT PROJECTION

Major area	Population (millions)			
	2015	2030	2050	2100
World	7 349	8 501	9 725	11 213
Africa	1 186	1 679	2 478	4 387
Asia	4 393	4 923	5 267	4 889
Europe	738	734	707	646
Latin America and the Caribbean	634	721	784	721
Northern America	358	396	433	500
Oceania	39	47	57	71

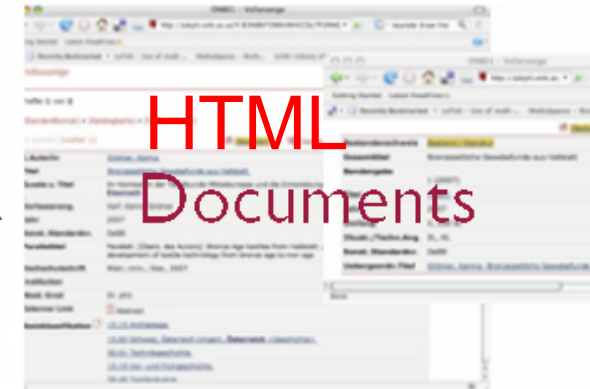
Source: United Nations, Department of Economic and Social Affairs, Population Division (2015). *World Population Prospects: The 2015 Revision*. New York: United Nations.

In 2015, 50.4 per cent of the world's population is male and 49.6 per cent is female (figure 1). The median age of the global population, that is, the age at which half the population is older and half is younger, is 29.6 years. About one-quarter (26 per cent) of the world's people are under 15 years of age, 62 per cent are aged 15-59 years, and 12 per cent are 60 or over.

การเปิดเผยข้อมูลในแบบหน้าเว็บ (web page)

TABLE S.1. TOTAL POPULATION BY SEX IN 2015 AND SEX RATIO BY COUNTRY IN 2015

Country or area	Population (thousands)			Sex ratio (males per 100 females)
	Total	Male	Female	
World.....	7 349 472	3 797 206	3 642 266	102
Afghanistan.....	32 527	16 774	15 753	106
Albania.....	2 897	1 437	1 460	98
Algeria.....	39 667	19 958	19 709	101
American Samoa.....	56
Andorra.....	70
Angola.....	23 022	12 416	12 606	98
Anguilla.....	15
Antigua and Barbuda.....	92	44	48	92
Argentina.....	43 417	21 245	22 172	96
Armenia.....	3 018	1 400	1 618	86
Aruba.....	104	49	55	91
Australia.....	23 969	11 976	11 993	100
Austria.....	8 545	4 196	4 349	97
Azerbaijan.....	9 754	4 856	4 898	99
Bahamas.....	388	190	198	96
Bahrain.....	1 377	854	523	163
Bangladesh.....	160 996	81 277	79 719	102
Barbados.....	284	136	148	92
Belarus.....	9 496	4 414	5 082	87
Belgium.....	11 299	5 559	5 740	97
Belize.....	359	179	180	99
Benin.....	10 880	5 426	5 454	99
Bermuda.....	62
Bhutan.....	775	416	359	116
Bolivia (Plurinational State of).....	10 725	5 371	5 354	100
Bosnia and Herzegovina.....	3 810	1 896	1 914	99
Botswana.....	2 262	1 130	1 132	100
Brazil.....	207 848	102 201	105 647	97
British Virgin Islands.....	30
Brunei Darussalam.....	423	218	205	106
Bulgaria.....	7 150	3 473	3 676	94
Burkina Faso.....	18 106	8 984	9 121	98
Burundi.....	11 179	5 524	5 655	98
Cabo Verde.....	521	257	264	97
Cameroon.....	15 578	7 598	7 980	95
Canada.....	35 940	17 672	18 268	100



- นำข้อมูลเข้าสู่ฐานข้อมูล เพื่อให้ผู้ใช้สืบค้นและเข้าถึง Web browser ได้

ข้อมูลอยู่ในแบบที่นำไปใช้ซ้ำได้ยาก ในโปรแกรมคอมพิวเตอร์ที่ต้องการนำข้อมูลนี้ไปใช้ (non machine-readable)

การเปิดเผยข้อมูลในแบบชุดข้อมูล (dataset)

- การเปิดเผยข้อมูลดิบ (raw data) ในแบบไฟล์ตารางคำนวณ (spreadsheet)

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, total	SP.POP.TOTL	54208	55435	56226	56697	5702
6 Andorra	AND	Population, total	SP.POP.TOTL	13414	14376	15376	16410	1747
7 Afghanistan	AFG	Population, total	SP.POP.TOTL	8994793	9164945	9343772	9531555	972864
8 Angola	AGO	Population, total	SP.POP.TOTL	5270844	5367287	5465905	5565808	566570
9 Albania	ALB	Population, total	SP.POP.TOTL	1608800	1659800	1711319	1762621	181413
10 Arab World	ARB	Population, total	SP.POP.TOTL	92495902	95041593	97691498	100438281	10327392
11 United Arab Emirates	ARE	Population, total	SP.POP.TOTL	92612	100985	112240	125216	13822
12 Argentina	ARG	Population, total	SP.POP.TOTL	20619075	20953079	21287682	21621845	2195392
13 Armenia	ARM	Population, total	SP.POP.TOTL	1867396	1934239	2002170	2070427	213813
14 American Samoa	ASM	Population, total	SP.POP.TOTL	20012	20478	21118	21883	2270
15 Antigua and Barbuda	ATG	Population, total	SP.POP.TOTL	54681	55403	56311	57368	5850
16 Australia	AUS	Population, total	SP.POP.TOTL	10276477	10483000	10742000	10950000	1116700
17 Austria	AUT	Population, total	SP.POP.TOTL	7047539	7086299	7129864	7175811	722380
18 Azerbaijan	AZE	Population, total	SP.POP.TOTL	3897889	4030130	4167558	4307315	444565
19 Burundi	BDI	Population, total	SP.POP.TOTL	2786740	2840375	2894510	2950903	3011195
20 Belgium	BEL	Population, total	SP.POP.TOTL	9153489	9183948	9220578	9289770	937811
21 Benin	BEN	Population, total	SP.POP.TOTL	2431620	2466002	2503232	2543335	258636
22 Burkina Faso	BFA	Population, total	SP.POP.TOTL	4829291	4894578	4960325	5027818	509889
23 Bangladesh	BGD	Population, total	SP.POP.TOTL	48200702	49593610	51030604	52532595	5412939
24 Bulgaria	BGR	Population, total	SP.POP.TOTL	7867374	7943118	8012946	8078145	814434
25 Bahrain	BHR	Population, total	SP.POP.TOTL	162501	167924	173107	178048	18277
26 Bahamas, The	BHS	Population, total	SP.POP.TOTL	109526	115108	121083	127331	13369
27 Bosnia and Herzegovina	BIH	Population, total	SP.POP.TOTL	3214520	3277096	3341809	3406466	346808

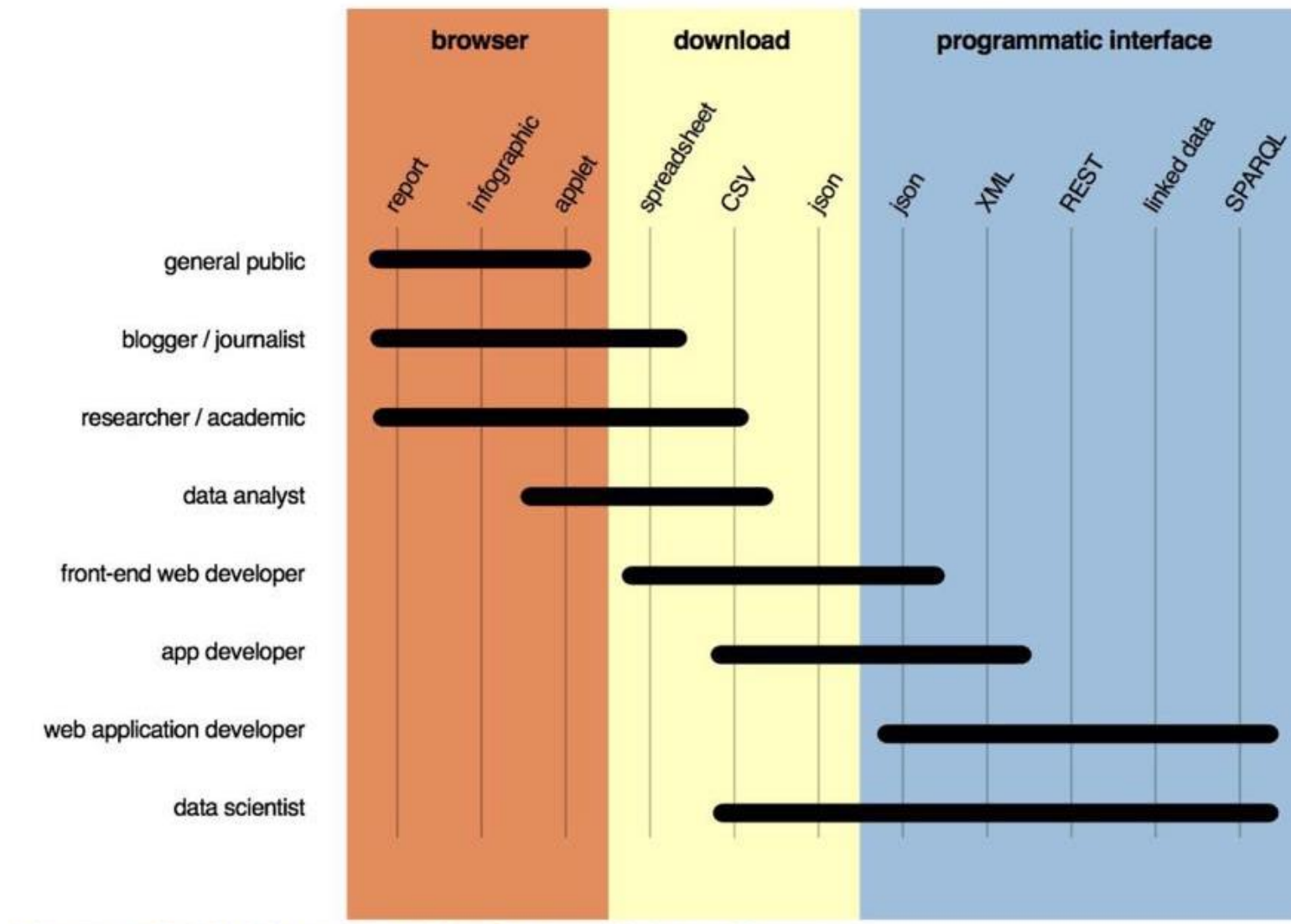
ข้อมูลอยู่ในแบบที่นำไปใช้ซ้ำได้ง่าย ในโปรแกรมคอมพิวเตอร์ที่ต้องการนำข้อมูลนี้ไปใช้ (machine-readable)

คุณสมบัติที่สำคัญของข้อมูลแบบเปิด (Open Data)

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

แนวทางการจัดทำข้อมูลเปิด (Publishing Open Data)

รูปแบบ (format) ของการเปิดเผยข้อมูล



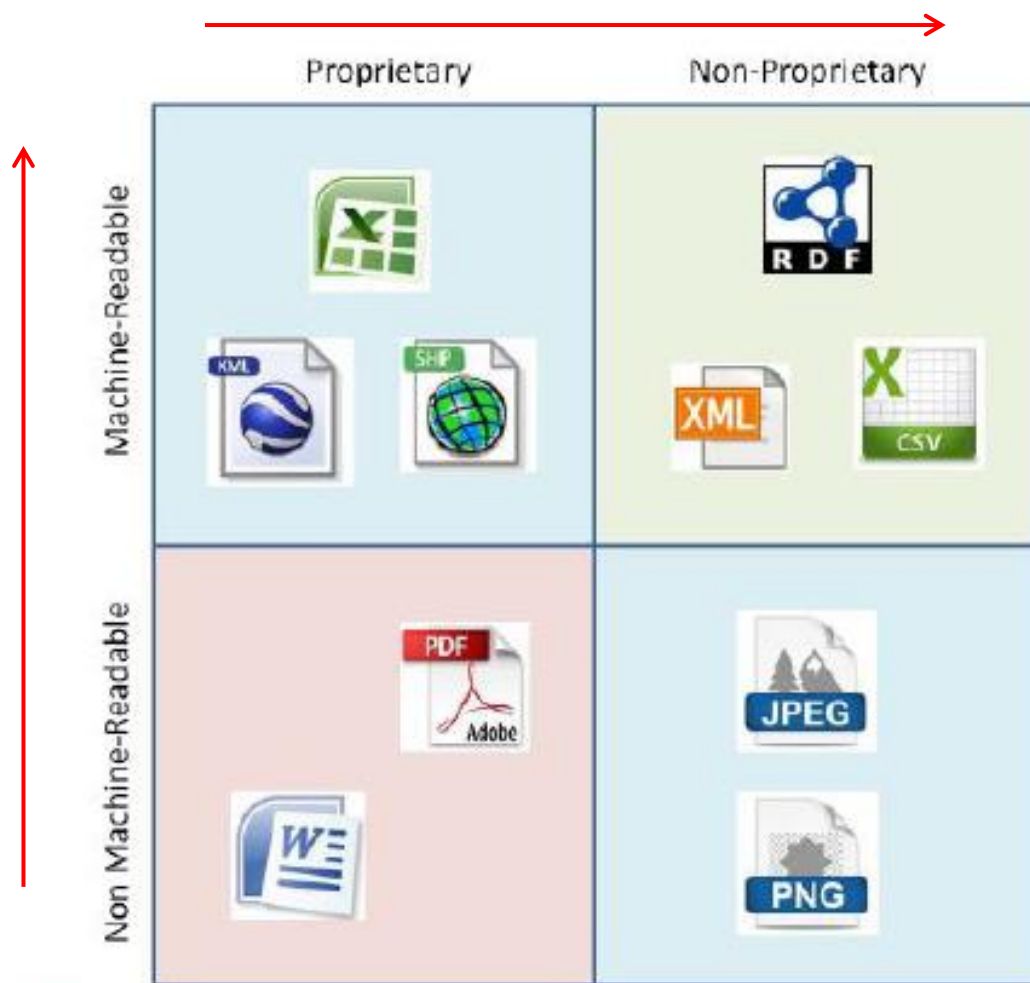
รูปแบบของไฟล์ (file format)

- รูปแบบของไฟล์ที่เหมาะสมสำหรับการเปิดเผยข้อมูลพิจารณาจาก
 - เป็นมาตรฐานเปิด (open standard)
 - โปรแกรมคอมพิวเตอร์นำไปใช้งานได้ง่าย (machine readable/reusability)

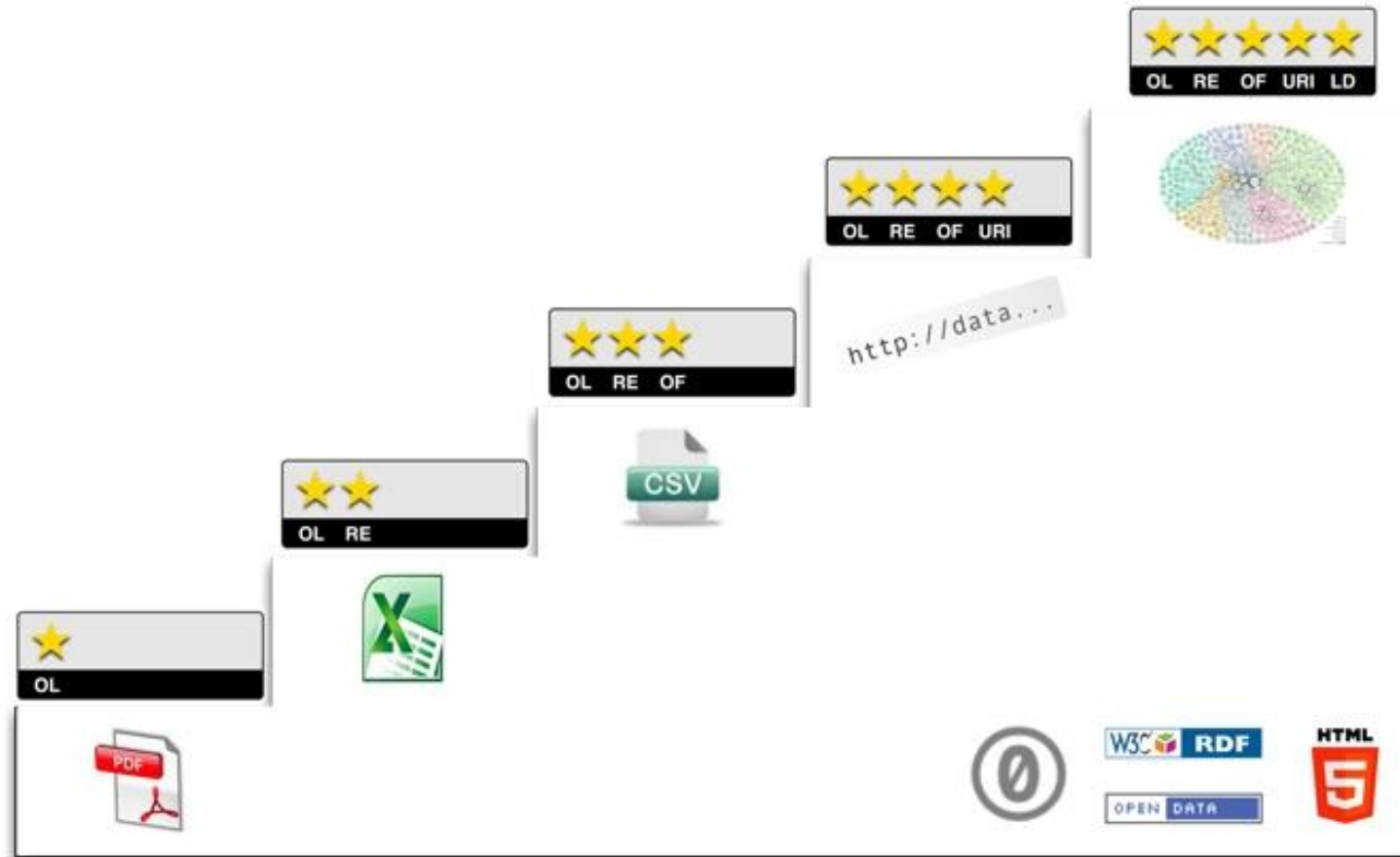
รูปแบบการเปิดเผยข้อมูลตามความง่ายต่อการนำไปใช้งาน

มาตรฐานเปิด

การนำไปใช้ซ้ำได้ง่าย



การเปิดเผยข้อมูล 5 ระดับ



ความหมายของการเปิดเผยข้อมูล 5 ระดับ

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

GDCatalog.go.th

รูปแบบ		
XLSX	510	★ ★
CSV	501	★ ★ ★
PDF	195	★
HTML	128	★
JSON	110	★ ★ ★
XLS	106	★ ★
database	60	
URL	19	★
WORD	15	★
ArcGIS Rest Services	12	★ ★ ★
แสดง รูปแบบ เพิ่มเติม		

รูปแบบ		
●	CSV (2,143)	★ ★ ★
●	XLSX (1,954)	★ ★
●	PDF (1,204)	★
●	XLS (729)	★ ★
●	URL (425)	★
●	JSON (371)	★ ★ ★
●	API (256)	★ ★ ★
●	XML (170)	★ ★ ★
●	JPG (124)	★
●	SHP (115)	★ ★ ★

Data.go.th

รูปแบบข้อมูลที่เครื่องอ่านได้ (machine-readable data)

ตัวอย่างของข้อมูลแบบที่เครื่องอ่านได้ (machine-readable data)

- ข้อมูลในแบบของ ตารางฐานข้อมูล
 - แถวแรกเป็นชื่อคอลัมน์/ แถวที่เหลือเป็นค่าข้อมูล
 - 1 คอลัมน์ 1 ชื่อ (ไม่มีการ span คอลัมน์)
 - ไม่มีคำอธิบายข้อมูลอยู่เหนือ หรือ ท้ายตาราง

	A	B	C	D	E
1	ชื่อ-สกุล	สาขาส่งงาน	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					



ตัวอย่างข้อมูลแบบตารางที่เครื่องอ่านไม่ได้ (non machine-readable data)

	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.คลองหลวง	ปทุมธานี	02-564-9999
4	นายสมพงษ์ สวัสดิ์	รัชดาภิเษก	1 ถ.รัชดาภิเษก เขตดินแดง กรุงเทพมหานคร		
5					



มีการ span
คอลัมน์ ในชื่อ
คอลัมน์ หรือ
ในค่าข้อมูล

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



เก็บข้อมูลในแต่ละ
คอลัมน์

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



ปล่อยให้ข้อมูลที่มี
ค่าเหมือนแถวก่อน
หน้าเป็นค่าว่าง

ตัวอย่างข้อมูลแบบตารางที่เครื่องอ่านไม่ได้ (non machine-readable data) (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 สวัสดิ์	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.คลองหลวง	ปทุมธานี	02-564-9999	
10							
11		ที่มา: เว็บไซต์บริษัท http://www.abc.in.th/					
12		ปรับปรุงล่าสุด: 28 ก.พ. 2559					
13							



มีคำอธิบายข้อมูล
อยู่เหนือ หรือ
ท้ายตาราง

มีแถว หรือ
คอลัมน์ว่าง (ไม่มี
ข้อมูล)

มีหลายตารางอยู่ใน
sheet เดียวกัน

ตัวอย่างของข้อมูลแบบที่เครื่องอ่านได้ (machine-readable data) ในระบบบัญชีข้อมูล

สถิติการผลิตเมล็ดพันธุ์ข้าว 291264

URL: http://catalog.ricethailand.go.th/dataset/2bb3b8be-f048-4961-8340-e4666f799f01/resource/5a4cef83-1fd0-429e-9181-4469182f1c44/download/09-curpro_291264.csv

Dataset description:

ปริมาณรวมการผลิตเมล็ดพันธุ์ข้าวปีปัจจุบัน

แหล่ง: สถิติการผลิตเมล็ดพันธุ์ข้าว

Data Explorer

Machine-readable data

Grid กราฟ แผนที่ 62 records

_id	วันเดือนปี	ศูนย์เมล็ดพันธุ์ข้าว	พันธุ์	ผลการผลิตปัจจุบัน (กก.)
1	25641229	พิษณุโลก	กข79	544,000
2	25641229	พิษณุโลก	พิษณุโลก 2	1,068,000
3	25641229	นครราชสีมา	ขาวดอกมะลิ 105	146,000
4	25641229	นครราชสีมา	ชัยนาท 1	490,300
5	25641229	ลำปาง	ลำปางแดง 1	902,700
6	25641229	ชัยนาท	กข41	509,750
7	25641229	ชัยนาท	กข57	132,525
8	25641229	ฉะเชิงเทรา	กข42	5,000

Data Dictionary

หลัก	ประเภท	ป้ายกำกับ	รายละเอียด
วันเดือนปี	numeric		
ศูนย์เมล็ดพันธุ์ข้าว	text		
พันธุ์	text		
ผลการผลิตปัจจุบัน (กก.)	text		

หลักเกณฑ์การตรวจสอบคุณภาพของชุดข้อมูลบน Data.go.th

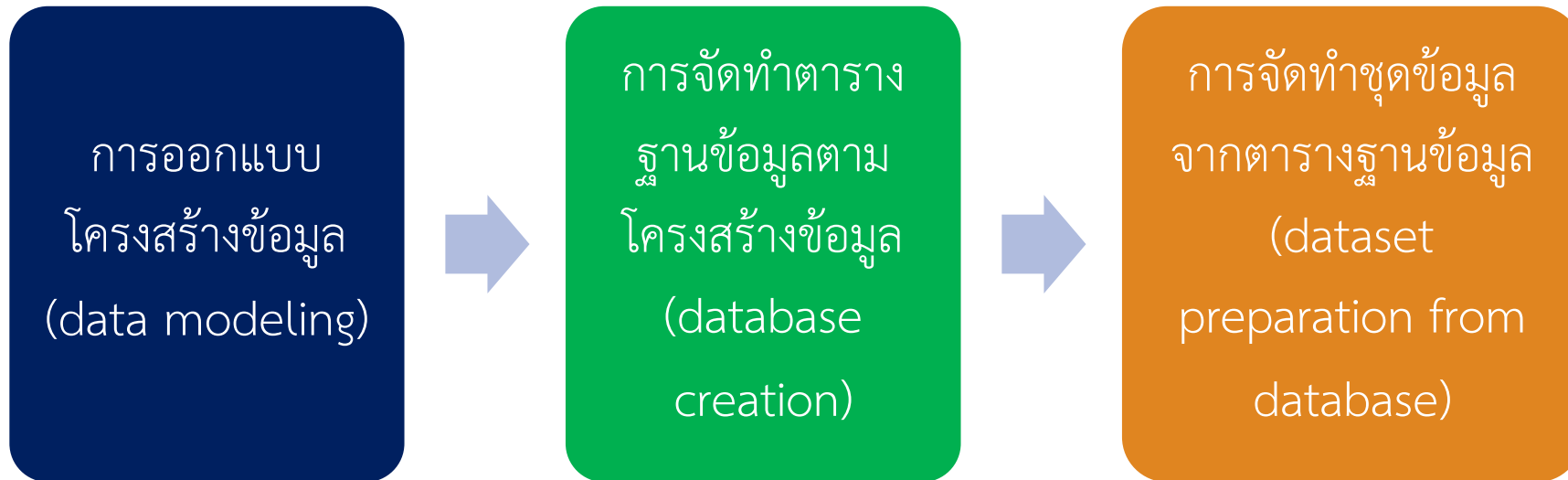
ตารางเกณฑ์การประเมินคุณภาพข้อมูล

ลำดับ	รายการเกณฑ์การประเมินทรัพยากร(ไฟล์)	คะแนนที่ตัดออก
1.	ไม่ระบุหัวตาราง (Header) ของไฟล์	20
2.	หัวตารางไฟล์มีมากกว่าหนึ่ง และมีการประสานเซลล์ (Merge Cells)เกิดขึ้น	5
3.	ข้อมูลตามในคอลัมน์ (Columns) ไม่มีความสม่ำเสมอ** (consistency) มีการผสมระหว่างตัวเลขและตัวอักษรภายในคอลัมน์ (Columns)	15
4.	ทรัพยากรหรือไฟล์ไม่ได้เป็นไปตามมาตรฐานการเข้ารหัสตัวอักษรแบบ Unicode (UTF-8)	10

อ้างอิง: <https://data.go.th/pages/open-government-data-of-thailand>

แนวทางการจัดทำชุดข้อมูลจากตารางฐานข้อมูล (Dataset Preparation from Database)

ภาพรวมการจัดทำชุดข้อมูลเพื่อเผยแพร่บนระบบบัญชีข้อมูล



แนวคิดการออกแบบโครงสร้างข้อมูล (Data Modelling)

การออกแบบโครงสร้างของข้อมูล (Data Modelling)

- ในยุคของข้อมูล (data) การออกแบบโครงสร้างของข้อมูล (Data Modelling) มีความจำเป็นอย่างยิ่งต่อการออกแบบและพัฒนาฐานข้อมูลของหน่วยงาน
- การออกแบบโครงสร้างข้อมูลที่ดี จะช่วยให้การจัดเก็บข้อมูลมีความสอดคล้องกับความต้องการของหน่วยงานอย่างแท้จริง (organization requirements)
 - สามารถเก็บข้อมูล (data) ของโลกความจริง (real-world objects) ได้ตามที่ต้องการ

ข้อมูล (Data)

ชื่อคอลัมน์ (ฟิลด์)

ชื่อตารางข้อมูล

Account		
No.	Name	Balance
1	John Smith	200.00
2	Joe Brown	50.00
3	Peter Clark	120.00
4	Paul Bright	230.00
5	Steve Thomas	1.00

โครงสร้างข้อมูล vs. ข้อมูล

- โครงสร้างข้อมูล (data schema)
 - โครงสร้างข้อมูลสำหรับการจัดเก็บข้อมูลในฐานข้อมูล
 - ปกติจะกำหนดขึ้นครั้งเดียวตอนสร้างฐานข้อมูล
- ข้อมูล (data record)
 - ข้อมูลที่จัดเก็บจริงในฐานข้อมูลที่สอดคล้องกับโครงสร้างข้อมูลที่ออกแบบไว้
 - เปลี่ยนแปลง/ปรับปรุงได้ตลอดเวลา (update)

โครงสร้างข้อมูล vs. ข้อมูล (2)

Account			โครงสร้างข้อมูล (data schema)
No.	Name	Balance	
1	John Smith	200.00	ข้อมูล (data record)
2	Joe Brown	50.00	
3	Peter Clark	120.00	
4	Paul Bright	230.00	
5	Steve Thomas	1.00	

ทดสอบ: โครงสร้างข้อมูล หรือ ข้อมูล

- ปีที่เกิด
- เบิร์ด ธงไชย
- ชื่อศิลปิน
- นครปฐม
- สถานที่เกิด
- 025640000
- 2563

โครงสร้างข้อมูลแบบตารางฐานข้อมูล

โครงสร้างข้อมูลแบบสัมพันธ์ (Relational data model)

- โครงสร้างข้อมูลแบบสัมพันธ์ (relational data model) กำหนดโครงสร้างข้อมูลในรูปแบบตาราง (tables)
- โครงสร้างข้อมูลแบบสัมพันธ์ประกอบด้วยข้อมูลต่อไปนี้:
 - ชื่อของตาราง (names of the tables)
 - คอลัมน์ของตาราง (columns of each table)
 - ชื่อคอลัมน์ (names)
 - ชนิดข้อมูล (data types)
 - ข้อกำหนดของตารางฐานข้อมูล (integrity constraints)

ข้อมูลแบบตาราง (Table)

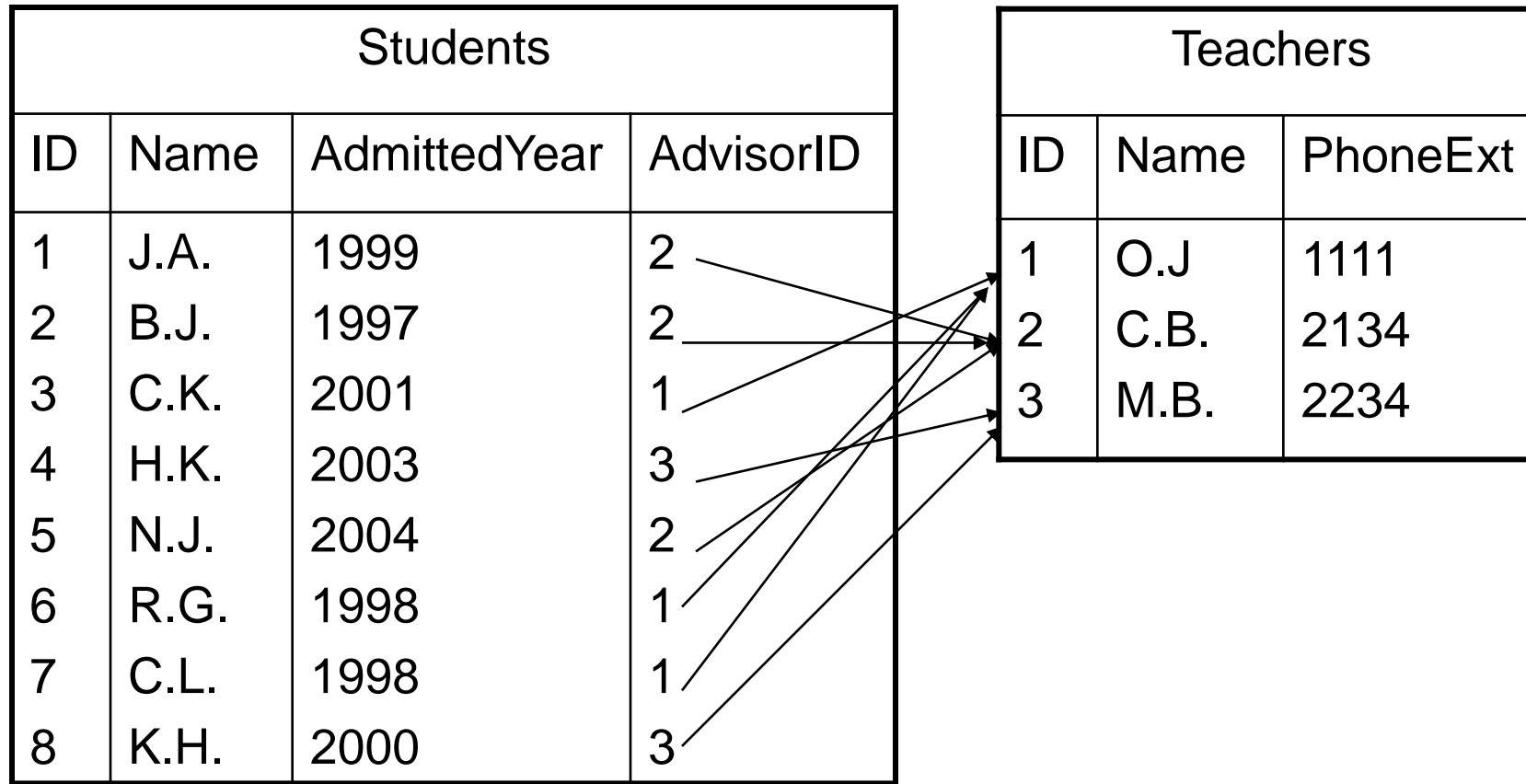
ตาราง		
คอลัมน์ 1 (ฟิลด์ 1)	...	คอลัมน์ n (ฟิลด์ n)
ค่าแอททริบิวต์ 1,1	...	ค่าแอททริบิวต์ 1,N
ค่าแอททริบิวต์ 2,1	...	ค่าแอททริบิวต์ 2,N
...
ค่าแอททริบิวต์ M,1	...	ค่าแอททริบิวต์ M,N

แถว (เรคคอร์ด) →

โมเดลข้อมูลแบบสัมพันธ์ (Relational data model) (2)

- คำศัพท์ที่ใช้เรียก
 - ตาราง (Table)
 - แถว (Row), เรคคอร์ด (Record)
 - คอลัมน์ (Column), แอททริบิว (Attribute), ฟิลด์ (Field)
 - ค่าแอททริบิว (Attribute Value)

ตัวอย่างฐานข้อมูลแบบสัมพันธ์



วิธีเขียนโครงสร้างของข้อมูลแบบตาราง

- แสดงหัวตาราง

Students			
ID	Name	AdmittedYear	AdvisorID

Teachers		
ID	Name	PhoneExt

- ชื่อตาราง(ชื่อคอลัมน์1, ชื่อคอลัมน์ 2, ...)
 - Students(ID, Name, AdmittedYear, AdvisorID),
Teachers(ID, Name, PhoneExt)

วิธีเขียนโครงสร้างของข้อมูลแบบตาราง (2)

- ชื่อตาราง(ชื่อคอลัมน์1: ชนิดข้อมูล1, ชื่อคอลัมน์2: ชนิดข้อมูล2)
 - Students(ID: decimal(2), Name: char(20), AdmittedYear: decimal(4), AdvisorID: decimal(2)), Teachers(ID: decimal(2), Name: char(20), PhoneExt: decimal(4))
- วิธีเขียนโครงสร้างของข้อมูลแบบตารางแบบนี้ เป็นแบบเดียวกับที่ใช้ในภาษา SQL
 - CREATE TABLE students(ID decimal(2), Name char(20), ...)

วิธีเขียนโครงสร้างของข้อมูลแบบตาราง (3)

- ตารางแสดงข้อมูลของแต่ละคอลัมน์ในตาราง (เช่น ชื่อคอลัมน์ ชนิดข้อมูล เป็นต้น)
- วิธีเขียนโครงสร้างของข้อมูลแบบตารางแบบนี้ เป็นแบบเดียวกับที่ใช้ใน data dictionary

Students	
Column	Type
ID	decimal(2)
Name	char(20)
AdmittedYear	decimal(4)
AdvisorID	decimal(2)

Teachers	
Column	Type
ID	decimal(2)
Name	char(20)
PhoneExt	decimal(4)

คุณสมบัติของข้อมูลแบบตารางฐานข้อมูล

- ห้ามมีแถวที่ข้อมูลซ้ำกัน (no duplicate rows)
- ค่าแอททริบิวต์ไม่สามารถแยกย่อยได้ (atomic attribute value)
 - ห้ามมีค่าแอททริบิวต์ที่มีโครงสร้าง (no structured attribute value)
 - ห้ามมีค่าแอททริบิวต์ที่มีหลายค่าซ้อนอยู่ภายใน (no multi-value attribute value)

name	birthyear
J.J.	1974
J.J.	1974

name	address	
J.J.	Muang	Pathumthani
P.P.	Muang, Nonthaburi	

name	phone
J.J.	02-111-1111
	02-222-2222
P.P.	02-111-1111, 02-222-2222

ข้อกำหนดของตารางฐานข้อมูล (integrity constraint)

- ข้อกำหนด 4 ชนิดของตารางฐานข้อมูล
 - คีย์ หรือ คีย์หลัก (Key หรือ Primary Key)
 - คีย์อ้างอิง (Foreign Key)
 - ห้ามมีค่าว่าง (NOT NULL)
 - ตรวจสอบค่าตามเงื่อนไข (CHECK)

ข้อกำหนดตารางฐานข้อมูล: ค่าว่าง (null value)

- ตารางฐานข้อมูลอนุญาตให้มีค่าว่าง (null value) ได้
- ค่าว่าง นิยมใช้ในกรณีต่อไปนี้:
 - มีค่าแต่ไม่ทราบ
 - เช่น ฟิลด์ เบอร์โทรศัพท์ของหน่วยงาน (มีแต่เราไม่ทราบ)
 - ไม่มีค่า
 - เช่น ฟิลด์ เบอร์แฟกซ์ของหน่วยงาน (ไม่มีเบอร์แฟกซ์)
 - ค่านั้นไม่เกี่ยวข้องกับข้อมูลในแถวดังกล่าว
 - เช่น ฟิลด์ จำนวนที่นั่ง ใช้กับแถวที่เป็นข้อมูลรถยนต์หรือรถตู้เท่านั้น ถ้าเป็นแถวที่เป็นรถจักรยานยนต์จะไม่ต้องระบุค่าในฟิลด์ดังกล่าว

ข้อกำหนดตารางฐานข้อมูล: ค่าว่าง (null value) (2)

- ค่าว่างช่วยให้เกิดความยืดหยุ่นในการป้อนข้อมูล แต่ค่าว่างอาจนำไปสู่ปัญหาสำคัญในการประยุกต์ใช้งาน เช่น ค่าในฟิลด์สำคัญขาดหายไป (missing value) เช่น ชื่อบุคคล ชื่อหน่วยงาน เบอร์ติดต่อ เป็นต้น
- ดังนั้นตารางในระบบฐานข้อมูลจึงอนุญาตให้ระบุข้อกำหนด NOT NULL สำหรับฟิลด์ที่ไม่ต้องการให้มีค่าว่างได้
- ‘NOT NULL’
 - ช่วยป้องกันปัญหาที่เกิดจากค่าว่าง
 - หากใช้กับฟิลด์ที่ไม่เหมาะสม อาจนำไปสู่ปัญหาภายหลังได้ เช่น ผู้บันทึกข้อมูลถูกบังคับให้ต้องระบุค่า แม้จะไม่ทราบค่า หรือไม่มีค่าก็ตาม เช่น เบอร์แฟกซ์ หรือ อีเมล เป็นต้น

ข้อกำหนดตารางฐานข้อมูล: คีย์ (Keys)

- คอลัมน์ (ฟิลด์) ในตารางฐานข้อมูล ที่ค่าในแต่ละแถวไม่ซ้ำกัน เช่น หมายเลขบัตรประชาชน
 - คีย์สามารถประกอบด้วยการผสมกันของหลายคอลัมน์ได้ (composite key) เช่น คอลัมน์ Firstname + คอลัมน์ Lastname เป็นคีย์ของตาราง Student เป็นต้น
- หากตารางมีมากกว่า 1 คีย์:
 - จะต้องเลือก 1 คอลัมน์ มากำหนดเป็นคีย์หลัก (Primary Key)
 - คีย์อื่นๆ จะกำหนดเป็น คีย์ทางเลือก (Alternate Keys) หรือ คีย์รอง (Secondary Keys)

ข้อกำหนดตารางฐานข้อมูล: คีย์ (Keys) (2)

- คีย์ และ ค่าว่าง
 - คีย์หลัก ห้ามเป็น ค่าว่าง เช่น ฟิลด์ หมายเลขบัตรประชาชน ชื่อ-นามสกุล
 - คีย์ทางเลือก/คีย์รอง สามารถเป็นค่าว่างได้ เช่น ฟิลด์ อีเมล เป็นต้น
- คีย์หลักควรเป็นฟิลด์ที่ไม่มีการปรับปรุงแก้ไข เช่น ฟิลด์ หมายเลขบัตรประชาชน เป็นต้น
- หากไม่สามารถหาฟิลด์ที่เป็นคีย์หลักของตารางได้ สามารถใช้ทุกฟิลด์ในตารางรวมกัน เพื่อกำหนดให้เป็นคีย์หลักของตารางได้

ตัวอย่าง คีย์ (Keys)

Primary Key

Alternate Key*

Students			
ID	Name	AdmittedYear	AdvisorID
1	J.A.	1999	2
2	B.J.	1997	2
3	C.K.	2001	1
4	H.K.	2003	3
5	N.J.	2004	2
6	R.G.	1998	1
7	C.L.	1998	1
8	K.H.	2000	3

Primary Key

Alternate Key*

Teachers		
ID	Name	PhoneExt
1	O.J	1111
2	C.B.	2134
3	M.B.	2234

* กรณีที่ชื่อนักเรียน ชื่อ

อาจารย์ เบอร์โทรศัพท์อาจารย์

มีค่าไม่ซ้ำกัน

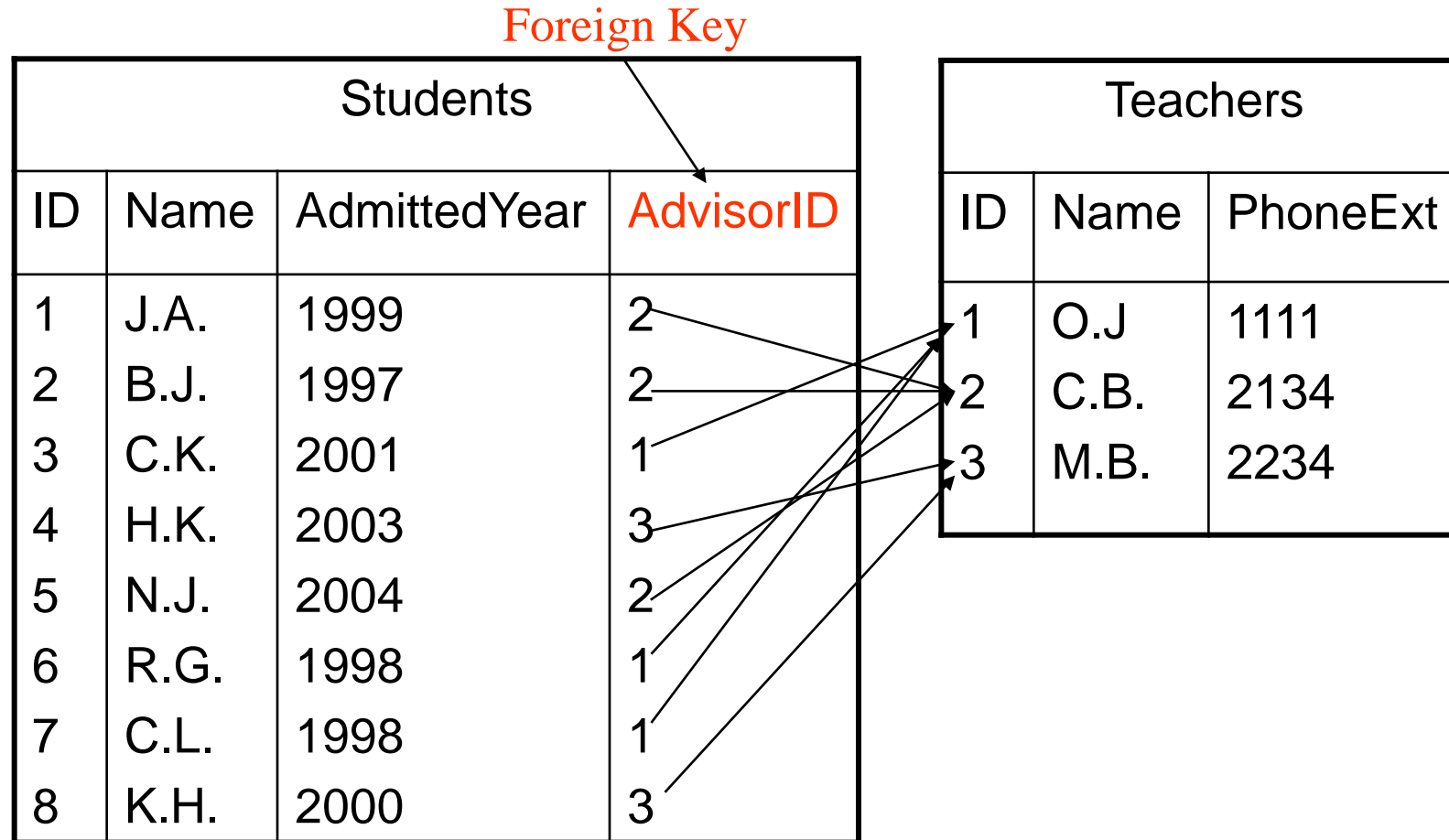
ทดสอบ: ระบุคีย์ของตารางฐานข้อมูล ‘สนามบิน’

- รหัสสนามบิน
- ชื่อสนามบิน (ไทย)
- ชื่อสนามบิน (อังกฤษ)
- จังหวัดที่ตั้ง
- ประเภทสนามบิน

ข้อกำหนดตารางฐานข้อมูล: คีย์อ้างอิง (Foreign Keys)

- คีย์อ้างอิง เป็นคอลัมน์ที่ค่าในคอลัมน์นั้นเป็นค่าที่อ้างอิงไปยังแถวต่างๆของตารางอื่น เช่น รหัสหน่วยงานที่บุคคลสังกัด
 - การเชื่อมโยงข้อมูลของตารางที่อ้างอิงไปยังตารางอื่น กับตารางที่ถูกอ้างอิงจะใช้วิธีการรวมตารางฐานข้อมูล (“Join”)
- ความถูกต้องสมบูรณ์ของการอ้างอิงข้อมูล (Referential Integrity)
 - ข้อกำหนดของคีย์อ้างอิง (Foreign key constraint)
 - ค่าข้อมูลในคอลัมน์ที่เป็นคีย์อ้างอิงต้องอ้างอิงค่าที่มีอยู่จริงในตารางอื่นเท่านั้น

ข้อกำหนดตารางฐานข้อมูล: คีย์อ้างอิง (Foreign Keys) (2)



ข้อกำหนดตารางฐานข้อมูล: คีย์อ้างอิง (Foreign Keys) (3)

- คีย์อ้างอิง กับค่าว่าง
 - คีย์อ้างอิงสามารถเป็นค่าว่างได้ เช่น ฟิลด์ หน่วยงานที่สังกัด เป็นต้น
- คีย์อ้างอิง มีความหมายแตกต่างจากคีย์ เนื่องจากค่าของฟิลด์ที่เป็นคีย์อ้างอิง สามารถมีค่าซ้ำกันได้
- คีย์อ้างอิง ต้องอ้างอิงไปยังฟิลด์ที่เป็นคีย์ ของตารางอื่นเท่านั้น (ไม่สามารถอ้างอิงฟิลด์ที่ไม่ใช่คีย์)

ข้อกำหนดตารางฐานข้อมูล: การตรวจสอบเงื่อนไขข้อมูล (CHECK)

- ตัวอย่างข้อกำหนดตารางฐานข้อมูล เพื่อการตรวจสอบเงื่อนไขข้อมูล (CHECK)
 - CHECK(AGE > 0 and AGE < 200)
 - CHECK(BIRTH_YEAR < GRADUATION_YEAR)
 - CHECK(GENDER='ชาย' OR GENDER='หญิง')
- ค่าว่าง (null value) จะไม่ถูกตรวจสอบเงื่อนไข (ถือว่าผ่านสำหรับทุกเงื่อนไข)

ทดสอบ: คีย์หลัก (primary key) หรือ คีย์อ้างอิง (foreign key) หรือไม่ใช้ทั้งสองอย่าง

ตารางฐานข้อมูล:

- เพลง (หมายเลขเพลง, ชื่อเพลง, แนวเพลง, หมายเลขศิลปินเจ้าของเพลง)
- ศิลปิน (หมายเลขศิลปิน, ชื่อศิลปิน, ค่ายเพลง)

- หมายเลขเพลง
- ชื่อเพลง
- แนวเพลง
- หมายเลขศิลปินเจ้าของเพลง

ตัวอย่างข้อมูล

เพลง				
<u>s_id</u>	ชื่อเพลง	แนวเพลง	ความยาวเพลง	artist_ID
<u>1</u>	เมตอินไทยแลนด์	เพื่อชีวิต	3.25	1
<u>2</u>	สบาย สบาย	ป๊อป	3.12	2
<u>3</u>	เดือนเพ็ญ	เพื่อชีวิต	2.58	1

ศิลปิน			
<u>a_id</u>	ชื่อศิลปิน	ประเภท	ค่ายเพลง
<u>1</u>	คาราบาว	วงดนตรี	XX
<u>2</u>	เบิร์ด ธงไชย	ศิลปินเดี่ยว	YY

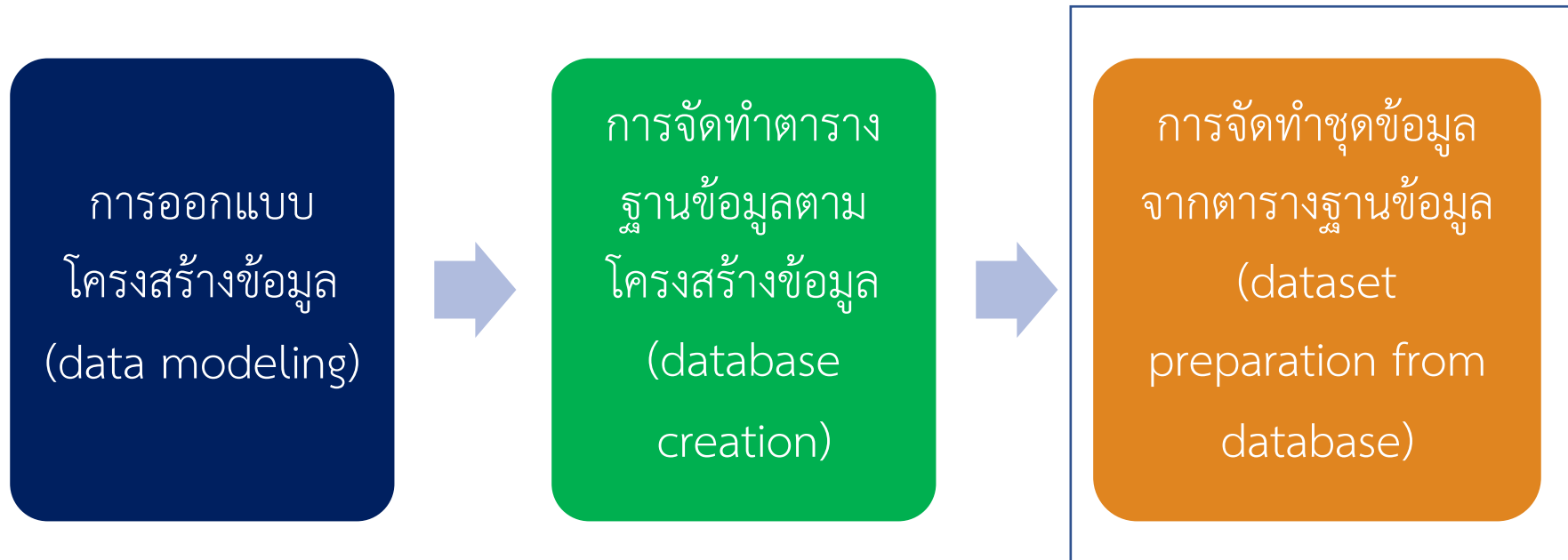
ตัวอย่างข้อมูล

ภาพยนตร์		
<u>m_id</u>	ชื่อภาพยนตร์	ปีที่สร้าง
1	คนเลี้ยงช้าง	xxxx
2	องค์บาก 2	yyyy

ดารานักแสดง		
<u>a_id</u>	ชื่อดารานักแสดง	ปีเกิด
1	สรพงษ์ ชาตรี	aaaa
2	จา พนม	bbbb

ภาพยนตร์_นักแสดง		
<u>m_id</u>	<u>a_id</u>	บทบาท
1	1	แสดงนำ
2	1	สมทบ
2	2	แสดงนำ

ภาพรวมการจัดทำชุดข้อมูลเพื่อเผยแพร่บนระบบบัญชีข้อมูล



การประมวลผลข้อมูลตารางด้วยภาษา SQL

รูปแบบคำสั่ง SQL (Syntax)

- คำสั่งพื้นฐานในการค้นคืนข้อมูล (SQL query syntax)

SELECT <ชื่อคอลัมน์1>, <ชื่อคอลัมน์2>, ...

FROM <ชื่อตาราง1>, <ชื่อตาราง 2>

WHERE <เงื่อนไขการกรองข้อมูล>

ตัวอย่างข้อมูล

students			
<u>id</u>	firstname	lastname	phone
<u>1</u>	สมชาย	ใจดี	1111
<u>2</u>	สมศักดิ์	สุขมาก	2222



SELECT firstname, lastname
FROM students
WHERE id='1'

<u>id</u>	firstname	lastname
<u>1</u>	สมชาย	ใจดี



SELECT *
FROM students
WHERE id='1'

<u>id</u>	firstname	lastname	phone
<u>1</u>	สมชาย	ใจดี	1111

การเชื่อมโยงข้อมูลตาราง (Join)

- เมื่อต้องการเชื่อมโยงข้อมูลมากกว่า 1 ตาราง (join) สามารถระบุชื่อตารางข้อมูลที่ต้องการเชื่อมโยงได้ในส่วน FROM
- ในกรณีดังกล่าวจะต้องระบุชื่อคอลัมน์ที่ต้องการเชื่อมโยงกัน ให้เป็นเงื่อนไขในการเชื่อมโยงตาราง (join condition) ในส่วน WHERE ด้วย

ตัวอย่างข้อมูล

students			
<u>id</u>	firstname	lastname	advisor_id
<u>1</u>	สมชาย	ใจดี	2
<u>2</u>	สมศักดิ์	สุขมาก	2

teachers	
<u>id</u>	name
<u>1</u>	สมหวัง สุขใจ
<u>2</u>	สมคิด ดีมาก



```
SELECT s.id, s.firstname, s.lastname, t.name  
FROM students s, teachers t  
WHERE s.advisor_id = t.id AND s.id = '1'
```

<u>s.id</u>	s.firstname	s.lastname	t.name
<u>1</u>	สมชาย	ใจดี	สมคิด ดีมาก

ตัวอย่างข้อมูล

students		
<u>id</u>	firstname	lastname
<u>1</u>	สมชาย	ใจดี
<u>2</u>	สมศักดิ์	สุขมาก

courses		
<u>id</u>	title	tname
<u>1</u>	คณิตศาสตร์	สมคิด ดีมาก
<u>2</u>	ภาษาไทย	สมหวัง สุขใจ

students_courses	
<u>s_id</u>	<u>course_id</u>
<u>1</u>	<u>2</u>
<u>2</u>	<u>1</u>



SELECT s.id, s.firstname, s.lastname, c.title, c.tname
FROM students_courses sc, courses c, student s
WHERE sc.course_id = c.id AND sc.s_id = s.id AND c.tname = 'สมคิด ดีมาก';

<u>s.id</u>	s.firstname	s.lastname	c.title	c.tname
<u>2</u>	สมศักดิ์	สุขมาก	คณิตศาสตร์	สมคิด ดีมาก

คำสั่ง SQL ในการค้นคืนข้อมูลแบบคำนวณสรุป

- ฟังก์ชันการคำนวณสรุป (Aggregation functions)
- เงื่อนไข GROUP BY, HAVING

การคำนวณสรุป (Aggregations)

- การประมวลผลกลุ่มของข้อมูลหลายค่า (set) ให้เหลือเพียงค่าเดียว เช่น $\min\{20,55,16,73\} = 16$
- ฟังก์ชันการคำนวณสรุปของภาษา SQL (SQL Aggregation functions)
 - COUNT: นับจำนวนทั้งหมด (Counting total number)
 - SUM: ค่าผลรวมทั้งหมด (Sum total value)
 - AVG: ค่าเฉลี่ย (Average value)
 - MIN: ค่าต่ำสุด (Minimum value)
 - MAX: ค่าสูงสุด (Maximum value)

ตัวอย่างการใช้งาน Aggregations

students			
id	firstname	lastname	province
1	สมชาย	ใจดี	เชียงใหม่
2	สมศักดิ์	สุขมาก	เชียงราย
3	สมนึก	สุขสบาย	เชียงใหม่



SELECT COUNT(*)
FROM students

COUNT(*)
3



SELECT province, COUNT(*)
FROM students
GROUP BY province

province	COUNT(*)
เชียงใหม่	2
เชียงราย	1



SELECT province, COUNT(*)
FROM students
GROUP BY province
HAVING COUNT(*) >= 2

province	COUNT(*)
เชียงใหม่	2

ชนิดของการเชื่อมโยงตาราง (join type)

- [INNER] JOIN: Join แบบปกติ
- LEFT [OUTER] JOIN: ข้อมูลทุกแถวของตารางทางด้านซ้ายจะถูกแสดงผลทั้งหมด
- RIGHT [OUTER] JOIN: ข้อมูลทุกแถวของตารางทางด้านขวาจะถูกแสดงผลทั้งหมด
- FULL [OUTER] JOIN: ตารางที่ใช้ในการประมวลผล join ทั้งหมดจะถูกแสดงผล

ตัวอย่าง OUTER JOIN

teachers	
<u>tid</u>	name
1	สมหวัง สุขใจ
2	สมคิด ดีมาก
3	สมนึก สุขสบาย



name	title
สมหวัง สุขใจ	ภาษาไทย
สมคิด ดีมาก	คณิตศาสตร์
สมคิด ดีมาก	ภาษาอังกฤษ
สมนึก สุขสบาย	

SELECT name, title

FROM teachers LEFT OUTER JOIN courses ON teachers.tid=courses.teacher_id



SELECT name, COUNT(title)

FROM teachers LEFT OUTER JOIN courses ON teachers.tid=courses.teacher_id

GROUP BY name

courses		
<u>cid</u>	title	teacher_id
1	คณิตศาสตร์	2
2	ภาษาไทย	1
3	ภาษาอังกฤษ	2

name	COUNT(title)
สมหวัง สุขใจ	1
สมคิด ดีมาก	2
สมนึก สุขสบาย	0

ตารางข้อมูลเสมือน (View)

- ตารางข้อมูลเสมือน (View) เป็นการนำผลลัพธ์ของการประมวลผลคำสั่งค้นคืนภาษา SQL มาใช้ประโยชน์ในการสร้างตารางข้อมูลที่มีความยืดหยุ่น (dynamic)

```
CREATE VIEW my_course (cid, title) AS  
SELECT cid, title  
FROM courses  
WHERE teacher_id = 2
```



my_course	
cid	title
1	คณิตศาสตร์
3	ภาษาอังกฤษ

ตารางข้อมูลเสมือน (View) (2)

- ในคำสั่งภาษา SQL นั้น View สามารถค้นคืนได้เสมือนเป็นตารางฐานข้อมูลปกติ, e.g.

SELECT *

FROM my_course

WHERE title LIKE 'ภาษา%'



cid	title
3	ภาษาอังกฤษ

ประโยชน์ของตารางข้อมูลเสมือน (Advantages of Views)

- สะดวกต่อการใช้งานและใช้คำสั่งค้นคืน SQL ซ้ำได้โดยไม่ต้องเขียนใหม่ (Convenience/ Reusability)
 - ประหยัดเวลาในการเขียนคำสั่งค้นคืน SQL
- สามารถใช้ในการจัดเตรียมชุดข้อมูลนำออกและเพื่อการแบ่งปันข้อมูลได้อย่างยืดหยุ่น
 - export ตารางเสมือนเป็นไฟล์ชุดข้อมูล (เช่น CSV) จากฐานข้อมูล

สรุป

- ภาพรวมการจัดทำชุดข้อมูลเพื่อเผยแพร่บนระบบบัญชีข้อมูล
 - การออกแบบโครงสร้างข้อมูล (data modeling)
 - การจัดทำตารางฐานข้อมูลตามโครงสร้างข้อมูล (database creation)
 - การจัดทำชุดข้อมูลจากตารางฐานข้อมูล (dataset preparation from database)

สรุป (2)

- การจัดทำชุดข้อมูลจากตารางฐานข้อมูล
 - การเชื่อมโยงข้อมูลจากหลายตาราง (join)
 - การสร้างตารางเสมือน (view) โดยระบุคำสั่งค้นคืนข้อมูล (query) ที่ต้องการ
 - การ export view ให้เป็นไฟล์ชุดข้อมูล (เช่น ไฟล์ CSV)