

MODULE 1.2

การค้นหาและรวบรวมข้อมูล

Data Search and Collection for Policy Analysis



2 ชั่วโมง



22 สไลด์

1 SMART Questions

สามารถแปลงปัญหานโยบายให้เป็นคำถามวิจัยที่มีคุณสมบัติ SMART ได้อย่างเป็นระบบ

- ระบุองค์ประกอบ 5 ประการ
- แยกคำถามหลักเป็นย่อย
- Map ไปสู่ตัวแปรที่วัดได้

2 Search Strategy

วางแผนการค้นหาและเข้าถึงข้อมูลจากแหล่งต่างๆ อย่างมีกลยุทธ์

- ใช้ Data Catalog อย่างมีประสิทธิภาพ
- ขอข้อมูลตาม พ.ร.บ.
- ประเมินความเป็นไปได้

3 Data Integration

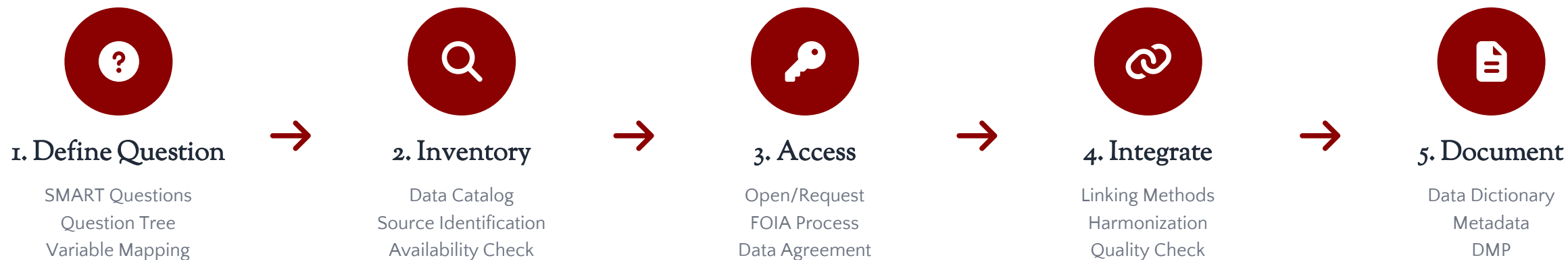
ออกแบบการรวมข้อมูลจากหลายแหล่งและจัดทำ data dictionary ที่สมบูรณ์

- Linking & Merging
- Standardization
- Privacy & Ethics

หมายเหตุ: ผู้เรียนจะได้ฝึกปฏิบัติจริงผ่านกิจกรรมเขียน SMART questions และออกแบบแผนการค้นหาข้อมูลตามโจทย์นโยบายจริง

Workflow ภาพรวม: 5 ขั้นตอนหลัก

End-to-End Data Collection Process



🔑 หลักการสำคัญ

- เริ่มจากคำถามที่ชัดเจน ไม่ใช่จากข้อมูลที่มีอยู่
- วางแผนย้อนกลับจากคำถามสู่แหล่งข้อมูล
- บันทึกทุกขั้นตอนเพื่อการตรวจสอบซ้ำ

⚠️ ข้อควรระวัง

- อย่าปรับคำถามให้เข้ากับข้อมูลที่มี
- ตรวจสอบความสอดคล้องของนิยามก่อน merge
- คำนึงถึงจริยธรรมและความเป็นส่วนตัวเสมอ

SMART Questions คืออะไร

Framework for Well-Defined Research Questions

S**Specific**

ชัดเจน
เฉพาะเจาะจง

- ใครเกี่ยวข้อง?
- อะไรต้องการ?
- ที่ไหน?
- เมื่อไร?
- ทำไม?

M**Measurable**

วัดได้
มีตัวชี้วัด

- ตัวชี้วัดอะไร?
- วัดอย่างไร?
- เกณฑ์ความสำเร็จ?
- หน่วยวัด?

A**Achievable**

บรรลุได้
มีทรัพยากร

- ข้อมูลมีจริง?
- เข้าถึงได้?
- มีเวลา?
- มีงบประมาณ?

R**Realistic**

สมจริง
สอดคล้อง

- สอดคล้องเป้าหมาย?
- เป็นไปได้จริง?
- ข้อยกเว้นอะไร?
- ความเสี่ยง?

T**Time-bound**

มีกรอบเวลา
ชัดเจน

- เริ่มเมื่อไร?
- สิ้นสุดเมื่อไร?
- ระยะเวลา?
- Milestones?

ที่มา: CDC Healthy Schools Training Framework และ SAMHSA Guidelines

SMART framework ถูกพัฒนาขึ้นเพื่อช่วยให้การตั้งคำถามวิจัยและวัตถุประสงค์มีความชัดเจน สามารถติดตามความก้าวหน้า และประเมินผลได้อย่างเป็นระบบ

ตัวอย่าง SMART Question

จากปัญหานโยบายสู่คำถามวิจัย

✕ คำถามที่ไม่ดี (Bad)

"อัตราการมารดาในประเทศไทยลดลงหรือไม่หลังจากมีนโยบายรักษาฟรี?"

- ⚠ **ไม่ชัดเจน:** ไม่ระบุช่วงเวลาและกลุ่มเป้าหมาย
- ⚠ **วัดยาก:** ไม่ระบุตัวชี้วัดและนิยาม
- ⚠ **ไม่มี baseline:** ไม่ระบุช่วงเวลาก่อน intervention
- ⚠ **ไม่สมจริง:** อาจมีปัจจัยอื่นมีผล

✓ คำถาม SMART (Good)

"อัตราการมารดาต่อ 100,000 การคลอดมีความแตกต่างกันระหว่างช่วงก่อน (2555-2559) และหลัง (2560-2564) การใช้สิทธิ UC สำหรับการฝากครรภ์ และคลอดในหญิงตั้งครรภ์อายุ 15-49 ปี ในจังหวัดที่มีอัตราการมารดาสูงกว่าค่าเฉลี่ยประเทศหรือไม่?"

- ✓ **Specific:** ระบุกลุ่มเป้าหมาย อายุ จังหวัด
- ✓ **Measurable:** ตัวชี้วัดชัดเจน (ต่อ 100,000 การคลอด)
- ✓ **Achievable:** ข้อมูลจาก MOPH และ NHSO มีอยู่
- ✓ **Realistic:** สอดคล้องกับนโยบาย UC
- ✓ **Time-bound:** ระบุช่วงเวลา baseline และ intervention

? คำถามหลัก (Main Question)

"อัตราการตายมารดาเปลี่ยนแปลงอย่างไรหลังนโยบาย UC?"

→ คำถามย่อย 1: อัตราตายมารดาคืออะไร?

ตัวแปร:

- จำนวนการตายมารดา
- จำนวนการคลอดมีชีพ
- อัตราต่อ 100,000 การคลอด

แหล่งข้อมูล:

- MOPH - การตายมารดา
- NHSO - การคลอดมีชีพ

→ คำถามย่อย 2: การใช้สิทธิ์ UC เปลี่ยนไปหรือไม่?

ตัวแปร:

- จำนวนผู้ใช้สิทธิ์ฝากครรภ์
- จำนวนผู้ใช้สิทธิ์คลอด
- อัตราการใช้สิทธิ์

แหล่งข้อมูล:

- NHSO - ข้อมูลสิทธิ์ UC
- สปสช. - ข้อมูลการใช้สิทธิ์

→ คำถามย่อย 3: ปัจจัยอื่นมีผลหรือไม่?

ตัวแปร:

- อายุมารดา
- การศึกษา
- รายได้
- พื้นที่ (urban/rural)

แหล่งข้อมูล:

- NSO - สำมะโนรายปี
- MOPH - ข้อมูลประชากร

🔑 หลักการสำคัญ

- ทุกคำถามย่อยต้อง map ไปที่ตัวแปรได้
- ตัวแปรต้องมีแหล่งข้อมูลที่เข้าถึงได้
- ควรมี 3-7 คำถามย่อยต่อคำถามหลัก
- ตรวจสอบความสอดคล้องระหว่างคำถาม

! ข้อควรระวัง

หลีกเลี่ยงคำถามย่อยที่ไม่สามารถตอบได้ด้วยข้อมูลที่มี เช่น "ทำไมอัตราการตายมารดาลด?" (ต้องใช้ qualitative data)

Mapping ตัวแปรตามกรอบ IPO/OI

Input-Process-Output-Outcome-Impact Framework

หมวดหมู่	คำอธิบาย	ตัวอย่างตัวแปร
IInput	ทรัพยากรและปัจจัยนำเข้าที่ใช้ในการดำเนินงาน	• งบประมาณ• บุคลากร• อุปกรณ์• สิทธิ UC
PProcess	กิจกรรมและกระบวนการที่ดำเนินการ	• จำนวนการให้บริการ• จำนวนผู้รับบริการ• ความถี่การติดตาม
OOutput	ผลผลิตที่เกิดขึ้นจากกระบวนการ	• จำนวนการคลอดมีชีพ• จำนวนการฝากครรภ์• รายงานที่ผลิต
OOutcome	ผลลัพธ์ระยะสั้น-กลางที่เกิดกับผู้รับบริการ	• อัตราตายมารดา• อัตราภาวะแทรกซ้อน• ความพึงพอใจ
IImpact	ผลกระทบระยะยาวต่อสังคม/ระบบ	• อายุขัยเฉลี่ย• คุณภาพชีวิต• ความยั่งยืน

✓

ประโยชน์ของกรอบ IPO/OI

- ช่วยให้ไม่หลุดตัวแปรสำคัญ
- แสดงความสัมพันธ์เชิงเหตุผล
- สนับสนุนการออกแบบ intervention
- ใช้สื่อสารกับผู้มีส่วนได้ส่วนเสีย

⚠

ข้อควรระวัง

- อย่าสับสนระหว่าง Output กับ Outcome
- Impact อาจต้องใช้เวลานานในการวัด
- ควรมีตัวแปรทุกหมวดเพื่อความสมบูรณ์

ระดับการวิเคราะห์ (Granularity)

Choosing the Right Level of Analysis



Individual

บุคคล

รายบุคคล
รายผู้ป่วย

ใช้เมื่อ:

- วิเคราะห์ผลกระทบรายบุคคล
- ติดตาม longitudinal
- ทำ matching



Facility

หน่วยบริการ

โรงพยาบาล
รพ.สต.

ใช้เมื่อ:

- เปรียบเทียบหน่วยงาน
- วิเคราะห์ capacity
- ประเมินคุณภาพ



District

อำเภอ

อำเภอ
เขต

ใช้เมื่อ:

- วิเคราะห์พื้นที่
- เปรียบเทียบระดับท้องถิ่น
- กระจายทรัพยากร



Province

จังหวัด

77 จังหวัด
กรุงเทพฯ

ใช้เมื่อ:

- เปรียบเทียบจังหวัด
- นโยบายระดับจังหวัด
- รายงานผู้ว่าฯ



National

ประเทศ

ทั้งประเทศ
ภูมิภาค

ใช้เมื่อ:

- นโยบายระดับชาติ
- เปรียบเทียบนานาชาติ
- SDGs indicators

📌 หลักการเลือกระดับการวิเคราะห์

1. สอดคล้องกับคำถาม

ถ้าคำถามเกี่ยวกับผลกระทบรายบุคคล ต้องใช้ข้อมูลระดับบุคคล ไม่ใช่ข้อมูลรวม

2. ข้อมูลมีจริง

บางครั้งข้อมูลระดับละเอียดไม่มี ต้องปรับคำถามหรือหา proxy

3. ความเป็นส่วนตัว

ข้อมูลระดับบุคคลต้องมีการปกป้อง privacy ตามกฎหมาย

Search Strategy: เริ่มจาก Data Catalog

Where to Start Looking for Data



Data.go.th

Open Government Data Portal

- ✓ บทบาท: ศูนย์กลางข้อมูลเปิดภาครัฐ
- ✓ จัดการโดย: DGA (สพร.)
- ✓ ข้อมูล: 11,000+ datasets
- ✓ การใช้งาน: ค้นหา ดาวน์โหลด API

เหมาะสำหรับ: ข้อมูลที่เปิดเผยสาธารณะ มี metadata เบื้องต้น



GD Catalog

Government Data Catalog

- ✓ บทบาท: บัญชีข้อมูลภาครัฐ
- ✓ จัดการโดย: NSO (สสช.)
- ✓ ข้อมูล: 8,390+ datasets (276 หน่วยงาน)
- ✓ การใช้งาน: ค้นหา metadata เชิงลึก

เหมาะสำหรับ: ระบุเจ้าของข้อมูล ข้อมูลที่ไม่เปิด ติดต่อขอใช้

๑.๕ แนวทางการค้นหาข้อมูล

1

เริ่มจาก
GD Catalog

2

ตรวจสอบ
สถานะ

3

ติดต่อ
เจ้าของ

4

ดาวน์โหลด
หรือขอใช้





Data.go.th: Open Data Portal ของไทย

Thailand's Central Open Data Platform

ข้อมูลทั่วไป

ชื่อเต็ม	URL
Open Government Data Portal	data.go.th
หน่วยงาน	จำนวนข้อมูล
DGA (สำนักงานพัฒนารัฐบาลดิจิทัล)	11,000+ datasets
เริ่มดำเนินการ	มาตรฐาน
พ.ศ. 2556	CKAN Platform

ฟังก์ชันการใช้งาน

 ค้นหาข้อมูล	 ดาวน์โฮลด์
ด้วย keywords, tags, categories	CSV, Excel, JSON, XML
 API Access	 Data Visualization
CKAN Data API, Open-D API	แสดงผลข้อมูลเบื้องต้น

หมวดหมู่ข้อมูล

การศึกษา	1,200+
สาธารณสุข	980+
เศรษฐกิจ	850+
สิ่งแวดล้อม	720+
คมนาคม	650+
อื่นๆ	6,600+

ที่มา:

www.dga.or.th – Digital Government Development Agency

เริ่มตามมติคณะรัฐมนตรี พ.ศ. 2556 เพื่อเตรียมความพร้อมสู่ประชาคมอาเซียน

GD Catalog: Government Data Catalog

Comprehensive Government Data Inventory

🏛️ โครงสร้างระบบ

- 1 **หน่วยงานเจ้าของข้อมูล**
276 หน่วยงานภาครัฐ
- 2 **ระบบบัญชีข้อมูล**
แต่ละหน่วยงานจัดการเอง
- 3 **GD Catalog (ศูนย์กลาง)**
รวบรวม metadata จากทุกหน่วย
- 4 **ผู้ใช้ข้อมูล**
ค้นหาและติดต่อเจ้าของ

📁 ข้อมูลที่ลงทะเบียน

Open Data

ข้อมูลที่เปิดเผยแพร่สาธารณะ ดาวน์โหลดได้ทันที

Restricted Data

ข้อมูลที่มีจำกัดการเข้าถึง ต้องขออนุญาต

Metadata Only

ข้อมูลที่มีเฉพาะรายละเอียด ไม่มีไฟล์

★ ความสำคัญของ GD Catalog

1. รู้ว่ามีข้อมูลอะไรบ้าง

แม้ข้อมูลจะไม่เปิด แต่รู้ว่าหน่วยงานไหนมีข้อมูลอะไร

2. ติดต่อเจ้าของข้อมูล

มีข้อมูลติดต่อ ช่องทางการขอใช้ข้อมูล

3. ลดการซ้ำซ้อน

หน่วยงานรู้ว่าหน่วยงานอื่นมีข้อมูลอะไรแล้ว

ช่องทางเข้าถึงข้อมูล (Access Pathways)

Three Types of Data Access



Open Data

ข้อมูลเปิดเผยสาธารณะ

- ✓ เข้าถึงได้ทันที
- ✓ ไม่ต้องขออนุญาต
- ✓ ดาวน์โหลดผ่าน data.go.th

ตัวอย่าง: สถิติประชากร ข้อมูลภูมิอากาศ แผนที่



Request

ขอตาม พ.ร.บ.ข้อมูลข่าวสาร

- ✓ ยื่นคำร้องเป็นหนังสือ
- ✓ หน่วยงานพิจารณา 15-30 วัน
- ✓ อาจมีค่าใช้จ่าย

ตัวอย่าง: รายงานการประชุม สถิติการให้บริการ



Restricted

ข้อมูลจำกัดการเข้าถึง

- ✓ ต้องมี MOU/NDA
- ✓ ระบุวัตถุประสงค์ชัดเจน
- ✓ มีเงื่อนไขการใช้งาน

ตัวอย่าง: ข้อมูลส่วนบุคคล ข้อมูลความมั่นคง

🔑 หลักการสำคัญ

1. เปิดเผยเป็นหลัก

ตาม พ.ร.บ.ข้อมูลข่าวสาร ข้อมูลภาครัฐต้องเปิดเผย เว้นแต่มีข้อยกเว้น

2. ระบุวัตถุประสงค์

ยิ่งระบุวัตถุประสงค์ชัดเจน ยิ่งมีโอกาสได้รับข้อมูลมากขึ้น

3. เริ่มจาก Open Data

ค้นหาข้อมูลเปิดก่อนเสมอ อาจมีข้อมูลที่ต้องการแล้ว

การขอข้อมูลตาม พ.ร.บ.ข้อมูลข่าวสาร

Official Information Act B.E. 2540

🔧 หลักการสำคัญ

”เปิดเผยเป็นหลัก ปกปิดเป็นข้อยกเว้น”

- 1 สิทธิของประชาชน**
มีสิทธิรู้ข้อมูลข่าวสารของราชการ (มาตรา 5)
- 2 หน้าที่ของหน่วยงาน**
ต้องจัดทำข้อมูลข่าวสารให้ประชาชนเข้าถึง (มาตรา 7)
- 3 ข้อยกเว้น**
ข้อมูลที่ไม่ต้องเปิดเผย 14 ประเภท (มาตรา 14)

⚠️ ข้อมูลที่ไม่ต้องเปิดเผย (ตัวอย่าง)

- | | |
|--------------------|-----------------------------|
| 🔒 ความมั่นคงของรัฐ | 🔒 ความสัมพันธ์ระหว่างประเทศ |
| 🔒 ความลับทางการค้า | 🔒 ข้อมูลส่วนบุคคล |
| 🔒 การดำเนินคดี | 🔒 ความลับทางวิชาชีพ |

📋 กระบวนการขอข้อมูล

- 1 ยื่นคำร้อง**
เป็นหนังสือหรืออิเล็กทรอนิกส์
- 2 พิจารณา**
ภายใน 15 วัน (ขยายได้อีก 15 วัน)
- 3 แจ้งผล**
อนุมัติ/ปฏิเสธ พร้อมเหตุผล
- 4 อุทธรณ์**
ถ้าไม่เห็นด้วย อุทธรณ์ได้

ที่มา:

oic.go.th – สำนักงานคณะกรรมการข้อมูลข่าวสารของราชการ
พ.ร.บ.ข้อมูลข่าวสารของราชการ พ.ศ. 2540 และแก้ไขเพิ่มเติม

I.2

Data Access Checklist

Systematic Approach to Data Access

หัวข้อ	รายการตรวจสอบ	หมายเหตุ	สถานะ
ผู้ถือครองข้อมูล	• ระบุหน่วยงานเจ้าของข้อมูล	GD Catalog	
	• ข้อมูลติดต่อผู้รับผิดชอบ	โทร/อีเมล	
	• กระบวนการขอข้อมูล	FOIA/ขอโดยตรง	
เงื่อนไขการใช้	• วัตถุประสงค์การใช้ข้อมูล	ระบุให้ชัดเจน	
	• ข้อจำกัดการใช้งาน	NDA/MOU	
	• การอ้างอิงแหล่งที่มา	Citation required	
Timeline	• ระยะเวลาพิจารณา	15-30 วัน	
	• กำหนดส่งมอบข้อมูล	ตกลงร่วมกัน	
ค่าใช้จ่าย	• ค่าธรรมเนียมข้อมูล	ถ้ามี	
	• ค่าใช้จ่ายอื่นๆ	จัดส่ง, ประมวลผล	
เอกสาร	• บันทึกการติดต่อ, อีเมล, MOU	เก็บไว้ตรวจสอบ	

Data Management Plan (DMP) แนวทางสากล

Planning for Data Lifecycle

1 Data Description

- ประเภทข้อมูล (quantitative/qualitative)
- รูปแบบไฟล์ (CSV, Excel, SPSS)
- ขนาดข้อมูลประมาณการ
- แหล่งที่มาของข้อมูล

2 Documentation

- Metadata มาตรฐาน
- Data dictionary
- Codebook สำหรับตัวแปร
- คำอธิบายวิธีการเก็บข้อมูล

3 Storage & Backup

- สถานที่จัดเก็บ (local/cloud)
- การสำรองข้อมูล
- ระบบรักษาความปลอดภัย
- การควบคุมการเข้าถึง

4 Data Sharing

- แผนการเผยแพร่ข้อมูล
- เงื่อนไขการใช้งาน
- การระบุสิทธิ์ (license)
- แพลตฟอร์มเผยแพร่

5 Quality Control

- ขั้นตอนตรวจสอบคุณภาพ
- การจัดการ missing data
- การตรวจสอบ outliers
- มาตรฐานความถูกต้อง

6 Ethics & Privacy

- การขออนุมัติจริยธรรม
- การปกป้องข้อมูลส่วนบุคคล
- การ anonymization
- การเก็บรักษาตามกฎหมาย

ที่มา: NIEHS – National Institute of Environmental Health Sciences

DMP ควรจัดทำตั้งแต่เริ่มโครงการวิจัย และทบทวนเป็นระยะตลอดวงจรชีวิตของข้อมูล

Metadata และ Data Dictionary

Making Data Interpretable and Reusable

Metadata (ข้อมูลเกี่ยวกับข้อมูล)

Descriptive Metadata

ชื่อชุดข้อมูล, คำอธิบาย, ผู้สร้าง, วันที่สร้าง, keywords

Structural Metadata

รูปแบบไฟล์, โครงสร้างตาราง, ความสัมพันธ์ระหว่างตาราง

Administrative Metadata

สิทธิ์การเข้าถึง, เงื่อนไขการใช้, ประวัติการแก้ไข

มาตรฐาน: Dublin Core, DDI (Data Documentation Initiative), DCAT

Data Dictionary (พจนานุกรมข้อมูล)

Field	Type	Description
patient_id	string	รหัสผู้ป่วย (ไม่ระบุตัวตน)
age	integer	อายุ (ปี)
gender	string	เพศ (M/F)
province	string	รหัสจังหวัด (2 หลัก)
diagnosis	string	รหัสโรค ICD-10

ควรมี: Field name, Data type, Description, Values/Range, Source, Calculation method

ทำไมต้องมี Metadata และ Data Dictionary?

1. ตีความเข้าใจได้

ผู้อื่นหรือตัวเราเองในอนาคตสามารถเข้าใจข้อมูลได้

2. รวมข้อมูลได้

ช่วยให้ harmonization และ integration ง่ายขึ้น

3. ตรวจสอบได้

รองรับการตรวจสอบคุณภาพและการทวนสอบ

✕ Before Harmonization

ตัวอย่าง 1: นิยามเบาหวาน

MOPH:

 $FPG \geq 126 \text{ mg/dL}$

NHSO:

 $HbA1c \geq 6.5\%$

⚠ ผลลัพธ์ต่างกัน!

ตัวอย่าง 2: รหัสจังหวัด

แหล่งที่มา A:

10 (2 หลัก)

แหล่งที่มา B:

1001 (4 หลัก)

⚠ Merge ไม่ได้!

ตัวอย่าง 3: หน่วยวัด

แหล่งที่มา A:

kg

แหล่งที่มา B:

pound

⚠ เปรียบเทียบผิด!

✓ After Harmonization

ตัวอย่าง 1: นิยามเบาหวาน

มาตรฐาน: ใช้นิยามเดียวกัน

 $FPG \geq 126 \text{ mg/dL}$ หรือ $HbA1c \geq 6.5\%$ หรือ กำลังรับประทานยาเบาหวาน

✓ ผลลัพธ์สอดคล้องกัน

ตัวอย่าง 2: รหัสจังหวัด

มาตรฐาน: ใช้รหัส 2 หลัก

แปลง $1001 \rightarrow 10$, $1002 \rightarrow 10$ (กรุงเทพฯ)

✓ Merge ได้

ตัวอย่าง 3: หน่วยวัด

มาตรฐาน: แปลงเป็นหน่วยเดียวกัน

 $\text{pound} \times 0.453592 = \text{kg}$

✓ เปรียบเทียบถูกต้อง

☰ Checklist ก่อน Harmonization

☐ ตรวจสอบนิยามตัวแปร☐ ตรวจสอบหน่วยวัด☐ ตรวจสอบรหัสมาตรฐาน☐ ตรวจสอบช่วงเวลา

Data Integration คืออะไร

Combining Data from Multiple Sources



Linking

เชื่อมโยงระเบียนจากหลายแหล่งโดยใช้ตัวระบุ (identifiers)

ตัวอย่าง: เชื่อมข้อมูลผู้ป่วยจากโรงพยาบาล A และ B ด้วย CID



Merging

รวมตารางข้อมูลเข้าด้วยกันตามตัวระบุที่ตรงกัน

ตัวอย่าง: LEFT JOIN, INNER JOIN, FULL OUTER JOIN



Governance

กำกับดูแลคุณภาพ ความปลอดภัย และการใช้งานข้อมูล

ตัวอย่าง: Privacy protection, Access control, Audit trail

ทำไมต้องทำ Data Integration?

1. มุมมองครบถ้วน

ได้ภาพรวมจากหลายแหล่ง ไม่ใช่มุมมองเดียว

2. วิเคราะห์ลึกขึ้น

สามารถทำ causal analysis, matching ได้

3. ลด bias

ตรวจสอบความสอดคล้องจากหลายแหล่ง

Linking Methods: Deterministic vs Probabilistic

Two Approaches to Record Linkage



Deterministic

การเชื่อมโยงแบบกำหนดได้

หลักการ

ใช้ unique identifier ที่ตรงกันเป๊ะ 100%

ตัวอย่าง Identifiers

- CID (เลขบัตรประชาชน)
- HN (Hospital Number)
- รหัสผู้ป่วยในระบบ

ข้อดี

- แม่นยำ 100%
- เร็ว ไม่ซับซ้อน
- ง่ายต่อการตรวจสอบ

ข้อจำกัด

- ต้องมี unique ID
- ถ้า ID ผิด = ไม่ match
- ไม่รองรับข้อผิดพลาด

เหมาะสำหรับ: ข้อมูลที่มี unique identifier ที่น่าเชื่อถือ



Probabilistic

การเชื่อมโยงแบบน่าจะเป็น

หลักการ

ใช้หลายตัวแปรร่วมกัน คำนวณความน่าจะเป็นว่าเป็นคนเดียวกัน

ตัวอย่าง Variables

- ชื่อ + นามสกุล
- วันเดือนปีเกิด
- เพศ + ที่อยู่

ข้อดี

- ไม่ต้องมี unique ID
- รองรับข้อผิดพลาด
- ยืดหยุ่นกว่า

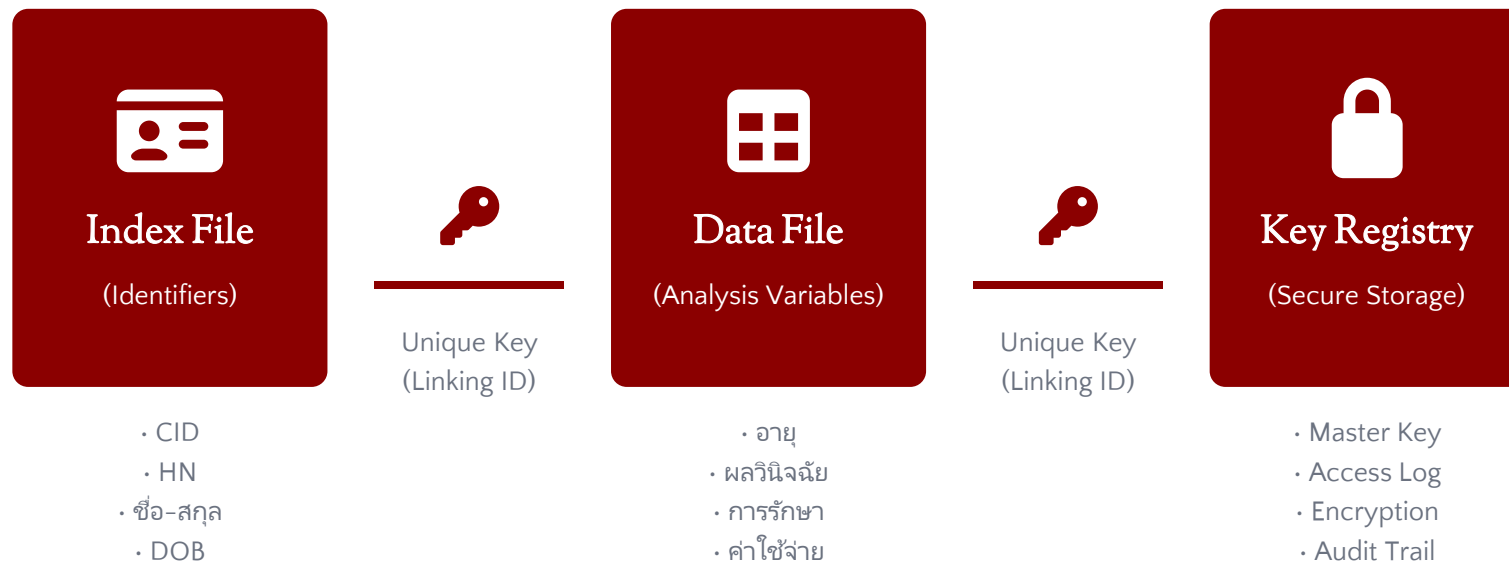
ข้อจำกัด

- ซับซ้อน ใช้เวลา
- ต้องกำหนด threshold
- อาจมี false positive/negative

เหมาะสำหรับ: ข้อมูลที่ไม่มี unique ID หรือ ID ไม่สมบูรณ์

Separation of Identifiers

Privacy-Preserving Data Architecture



ความปลอดภัย (Security)

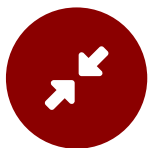
- Identifiers แยกเก็บ (Separate storage)
- ข้อมูลวิเคราะห์ไม่มี PID (Analysis data without PID)
- เข้าถึง Key Registry จำกัด (Limited access to Key Registry)
- บันทึกทุกการเข้าถึง (Log all access)

การเชื่อมโยง (Linkage)

- ใช้ Linking ID แทน PID (Use Linking ID instead of PID)
- สามารถ link กลับได้ (Can link back)
- หลาย Data File ใช้ Key เดียวกัน (Multiple Data Files use the same Key)
- ง่ายต่อการตรวจสอบ (Easy to check)

การควบคุม (Control)

- กำหนดสิทธิ์การเข้าถึง (Define access rights)
- แยกหน้าที่ Data Analyst (Separate Data Analyst roles)
- ตรวจสอบย้อนกลับได้ (Traceable)
- ลดความเสี่ยงรั่วไหล (Reduce risk of leakage)



Collection Limitation

เก็บเท่าที่จำเป็น

หลักการ

เก็บข้อมูลส่วนบุคคลเท่าที่จำเป็นต่อวัตถุประสงค์ที่ระบุไว้

การปฏิบัติ

- ระบุว่าทำไมต้องเก็บ
- ไม่เก็บข้อมูลส่วนเกิน
- ตรวจสอบความจำเป็น



Purpose Specification

ระบุวัตถุประสงค์ชัดเจน

หลักการ

วัตถุประสงค์ต้องระบุตั้งแต่ตอนเก็บข้อมูล

การปฏิบัติ

- บันทึกวัตถุประสงค์
- แจ้งให้ผู้ให้ข้อมูลทราบ
- ไม่ใช่นอกวัตถุประสงค์



Use Limitation

จำกัดการใช้งาน

หลักการ

ไม่ใช่ข้อมูลนอกวัตถุประสงค์ที่ระบุไว้

การปฏิบัติ

- ตรวจสอบก่อนใช้
- ขอความยินยอมถ้าจะใช้อย่างอื่น
- บันทึกการใช้งาน

📌 หลักการอื่นๆ จาก OECD Privacy Guidelines

Data Quality

ข้อมูลต้องถูกต้อง ครบถ้วน ทันสมัย

Security Safeguards

ป้องกันการเข้าถึงโดยไม่ได้รับอนุญาต

Openness

เปิดเผยแนวทางการจัดการข้อมูล

Individual Participation

สิทธิเข้าถึงและแก้ไขข้อมูลตนเอง

ที่มา: OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data

✓ สรุปบทเรียน I.2

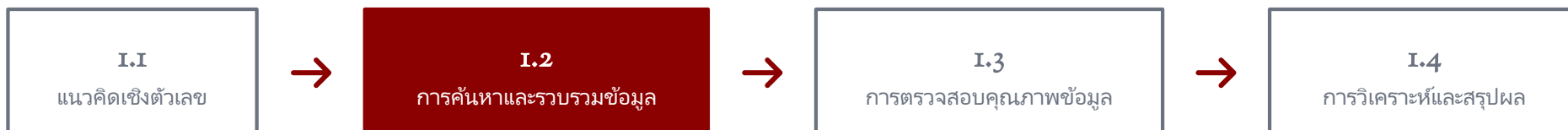
- 1 SMART Questions
แปลงปัญหานโยบายเป็นคำถามวิจัยที่ชัดเจน วัดได้
- 2 Search Strategy
ค้นหาข้อมูลจาก Data.go.th และ GD Catalog
- 3 Data Access
ขอข้อมูลตาม พ.ร.บ. ข้อมูลข่าวสาร
- 4 Data Integration
รวมข้อมูลหลายแหล่งด้วย linking methods
- 5 Documentation
จัดทำ metadata และ data dictionary

→ หัวข้อต่อไป: I.3

การตรวจสอบคุณภาพข้อมูล

Data Quality Assessment

- ✓ เมื่อมีข้อมูลแล้ว ต้องตรวจสอบความน่าเชื่อถือ
- ✓ ประเมินความสมบูรณ์ ความถูกต้อง ความสอดคล้อง
- ✓ จัดการ missing data และ outliers
- ✓ เตรียมข้อมูลสำหรับการวิเคราะห์



ขอบคุณสำหรับการเข้าร่วม

พร้อมสำหรับ Workshop การเขียน SMART Questions และ Data Source Mapping