

Organizing Data

จัดระเบียบข้อมูล

The secret weapon to solid data analysis
อาวุธลับเพื่อการวิเคราะห์ข้อมูล

เข้าใจ...มาตรฐานข้อมูล

Data standard หรือ **มาตรฐานข้อมูล** คือ การจัดทำข้อมูลให้อยู่ในลักษณะรูปแบบกลาง (Neutral format) โดยเนื้อหาของมาตรฐานข้อมูลหลัก ๆ ได้แก่

- ระบบอ้างอิง (Reference systems)
 - โครงสร้างข้อมูล (Data models) = เนื้อหา + คุณลักษณะของข้อมูล
- รวมทั้งวิธีการผลิต การจัดเก็บข้อมูล และการใช้ข้อมูล เป็นต้น

ระบบนี้ช่วยให้คอมพิวเตอร์ สามารถรับรู้และวิเคราะห์ข้อมูล ทำให้สามารถเข้าถึงข้อสรุปทั่วไปเกี่ยวกับข้อมูล



ต้องมีการกำหนด **คำศัพท์ (Data dictionaries)** และข้อมูลจะต้องมี
คุณภาพของข้อมูล (data quality) และในด้านคำอธิบายข้อมูล (Metadata)
ซึ่งเป็นรายละเอียดบอกถึงเนื้อหาคุณลักษณะของข้อมูล



Species	Habitats
AH	Shrubland, Grassland
AS	Grassland
SH	Grassland, Woodland, Shrubland

Species	Habitat 1	Habitat 2	Habitat 3
AH	Shrubland	Grassland	
AS	Grassland		
SH	Grassland	Woodland	Shrubland

Species	Habitat
AH	Shrubland
AH	Grassland
AS	Grassland
SH	Grassland
SH	Woodland
SH	Shrubland

ข้อมูลไม่มีคุณภาพ เกิดจาก...

1. การขาดการระบุถึงแหล่งข้อมูลหลัก และ นโยบายของการป้อนข้อมูลที่ชัดเจน ส่งผลให้ผู้ใช้งานในแต่ละฝ่ายงานมีความเข้าใจข้อมูลที่แตกต่างกัน โดยเฉพาะอย่างยิ่ง องค์กรขนาดใหญ่ที่ประกอบด้วยหน่วยงานหลายส่วน และพนักงานในแต่ละฝ่ายก็มักให้ความสนใจเฉพาะกับหน่วยงานของตนเองเท่านั้น
2. การขาดการกำหนดผู้ที่มีหน้าที่รับผิดชอบเกี่ยวกับการบริหารจัดการข้อมูล หรือ มีการกำหนดผู้รับผิดชอบที่ไม่เหมาะสม
3. ความไม่สม่ำเสมอของข้อมูลในระบบงานต่างๆ
4. การขาดการควบคุมภายในที่มีประสิทธิภาพ



1. เมื่อทำงานกับฐานข้อมูล ข้อมูลอาจมาจากแหล่งต่าง ๆ ซึ่ง
 - โครงสร้างข้อมูลอาจไม่สมบูรณ์
 - ข้อผิดพลาดบางอย่าง เช่น มีข้อมูลที่ซ้ำกัน หรือ การสะกดผิดสิ่งนี้จะทำให้กระบวนการวิเคราะห์ซับซ้อน และแม้ว่าเราจะรับรู้ถึงข้อผิดพลาด แต่คอมพิวเตอร์ไม่รู้
2. การกำหนดมาตรฐาน หรือ การทำความสะอาดข้อมูล เป็นขั้นตอนสำคัญในการสื่อสารข้อมูล
3. หนึ่งในขั้นตอนแรกในการทำความสะอาดข้อมูล คือ การทำให้แน่ใจว่าส่วนหัวตารางทั้งหมดถูกต้องและครบถ้วน และเนื้อหาข้อมูลในแต่ละแถวตรงกัน

The Header | ส่วนหัวตาราง

1. แถวแรกควรมีชื่อตัวแปร
2. อย่าใช้ชื่อตัวแปรมากกว่าหนึ่งแถว
3. เต็มส่วนหัวที่หายไป



Species	Habitats
AH	Shrubland, Grassland
AS	Grassland
SH	Grassland, Woodland, Shrubland

Species	Habitat1	Habitat2	Habitat3
AH	Shrubland	Grassland	
AS	Grassland		
SH	Grassland	Woodland	Shrubland

The Cell | เนื้อหาในเซลล์

1. ช่องตาราง = เซลล์ ในสเปรดชีต ควรมีข้อมูลหนึ่งอย่าง
ไม่ควรใส่เนื้อหามากกว่าหนึ่งอย่าง
2. หลีกเลี่ยงการผสานเซลล์
3. จัดรูปแบบข้อมูลให้เหมือนกัน



Species	Habitats
AH	Shrubland, Grassland
AS	Grassland
SH	Grassland, Woodland, Shrubland

Empty cells | เนื้อหาในเซลล์ที่ว่างเปล่า (ถ้ามี)

1. ลบแถวที่ว่างเปล่า แต่ไม่ใช่เซลล์แต่ละเซลล์
ยกเว้น คุณแน่ใจเกี่ยวกับสาเหตุที่เซลล์ว่างเปล่า
2. ใช้รหัสที่เป็นที่รู้จัก (เช่น n/a) สำหรับข้อมูลที่หายไป
(missing data)
3. เต็มช่องว่างในการแยกข้อมูล ด้วยค่าเฉลี่ยข้อมูล



การสร้างข้อมูลให้มีมาตรฐาน

1. หลักการพื้นฐาน คือ ข้อมูลทั้งหมดจะถูกป้อนในรูปแบบเดียวกัน บ่อยครั้งในตัวพิมพ์ใหญ่/พิมพ์เล็กเหมือนกันทั้งหมด โดยไม่มีช่องว่างเพิ่มเติม
2. ข้อมูลต่างประเภท ควรมีคอลัมน์แยกออกจากกัน

หากฐานข้อมูลมี

- ที่อยู่
- วันที่
- อายุ

ต้องตัดสินใจเลือกใช้วิธีมาตรฐาน
ในการป้อนข้อมูลเหล่านี้ลงในฐานข้อมูล

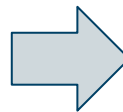
Date
07 February 2020
07/2/2020
2/07/2020
07/2/20
07/feb/2020

ไม่ได้มีรูปแบบเดียวที่ถูกต้อง
แต่วันที่ทั้งหมดต้องถูกจัดรูปแบบที่
เหมือนกันและคอมพิวเตอร์จะ
เข้าใจรูปแบบวันนั้นได้

การเลือกรูปแบบที่เหมาะสมที่สุด
สำหรับฐานข้อมูล จึงเป็นสิ่งสำคัญ

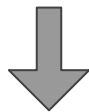
ในกรณีนี้เลือก DD/MM/YYYY
ดังนั้น เมื่อจัดระเบียบข้อมูลแล้ว
จะได้เป็นรูปแบบ ดังนี้:

Date
07 February 2020
07/2/2020
2/07/2020
07/2/20
07/feb/2020



DATE
07/02/2020
07/02/2020
07/02/2020
07/02/2020
07/02/2020

Name	Date of Birth	Address	Salary
Dy Saveth	16 April 1975	Corner Street 13 & 102, SangKat Wat Phnom, Khan Daun Penh, Phnom Penh	\$1250
Virak Dara	31/5/1980	# 217, St. Monica Ray, Sangkat Boeng Salang, Tuol Kork, Phnom Penh	US\$1000



NAME	DATE OF BIRTH	ADDRESS	COMMUNE	DISTRICT	CITY	SALARY (USD)
Dy Saveth	16/04/1975	41 12TH STREET	Daun Penh	Khan Daun Penh	PHNOM PENH	1250
Virak Dara	31/05/1980	6 SANCHAUN G ROAD	Boeng Salang	Tuol Kork	PHNOM PENH	1000

เคล็ดลับ สำหรับการตรวจสอบข้อมูลเบื้องต้น

- ตรวจสอบให้แน่ใจว่า ข้อมูลมีจำนวนทั้งหมดเท่าไร และคุณมีข้อมูลทั้งหมดแล้ว
- ตรวจสอบ ความสอดคล้อง - ข้อมูลทุกอัน
- ตรวจสอบให้แน่ใจว่าส่วนหัวคอลัมน์มีความชัดเจน และ หน่วยต่าง ๆ สอดคล้องกัน
- พิจารณาว่า ช่องว่าง หรือ N/As หมายถึงอะไร
- หากมีการคัดลอกตารางข้อมูล / ดึงข้อมูลจากแหล่งต่าง ๆ ให้คิดวิธีที่ดีที่สุดในการจัดโครงสร้างข้อมูลของคุณสำหรับการวิเคราะห์

Thank You



This lesson was adapted from the World Bank's Introduction to Data Literacy training manual by Eva Constantaras, and adapted by Yan Naung Oak, Open Development Cambodia and Open Development Initiative, and is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License](#). For full terms of use, see [here](#).