



# อภิปรายคำนิยามของคำว่า Data, Big Data, Records และ Archives

เลิศชัย วาสนานิกรกุลชัย

โครงการประชุมวิชาการประจำปี ๒๕๖๒

เรื่อง บทบาทงานจดหมายเหตุในกระแส Big Data และสังคมแห่งการเรียนรู้

จัดโดย สมาคมจดหมายเหตุไทย ร่วมกับ มูลนิธิหอจดหมายเหตุพุทธทาส อินทปัญโญ

ณ ศูนย์การเรียนรู้ธนาคารแห่งประเทศไทย

๙ สิงหาคม ๒๕๖๒



## ลักษณะของเอกสาร (Records)

- ความจริงแท้ **Authenticity** → Genuine
- ความสมบูรณ์ **Integrity** → Trustworthy
- สื่อความหมายได้ **Usability** → Accessible over time + sufficient evidence of context
- ความน่าเชื่อถือ **Reliability** → Accurate content

ตามมาตรฐาน ISO 15489-1 2016



## นิยามเอกสารและข้อมูล

- Document
  - Work in progress, not yet captured as a record
  - The smallest unit of filing ( เช่น จดหมาย แบบฟอร์ม รายงาน ฯลฯ )
  - Not all documents are records
    - ใบสมัครงาน จดหมาย ใบเสร็จ สมุดทะเบียน สมุดบัญชี เป็น document และ record
    - แบบฟอร์มเปล่า ไปรษณีย์บัตรภาพ แผ่นโฆษณา หนังสืออ้างอิง เป็น document แต่ไม่ใช่ record
    - ไปรษณีย์บัตรที่เขียนข้อความ และส่งไปรษณีย์ถึงผู้รับ จึงจะเป็น record ได้
    - Record ต้องเกี่ยวข้องกับกิจกรรมทางธุรกิจ
- Data
  - ข้อมูลดิบ ตัวเลข ชื่อ ตำแหน่ง วันที่ ปริมาณ ต้นทุน ฯลฯ
  - ประกอบด้วย element (คอลัมน์ในฐานข้อมูล) และ value (ข้อมูลภายในคอลัมน์นั้น)

Shepherd, E., & Yeo, G. (2003). *Managing Records: A Handbook of Principles and Practice*. Facet Publishing.



# หลักการสร้างเอกสาร(Records)

- เอกสารควรสร้างเพื่อกำกับหรือประกอบการทำกิจกรรมในองค์กร
- ในทางปฏิบัติ เอกสารจะถูกสร้างขึ้นจากความต้องการใช้งานขององค์กร
  - บางองค์กรอาจมีการบันทึกบทสนทนาทางโทรศัพท์ หรือการสื่อสารด้วยวาจาในลักษณะอื่น
- การประเมินความต้องการใช้งาน
  - ความต้องการขององค์กร เพื่อสร้างหลักฐานและข้อมูลสำหรับปฏิบัติงาน
  - ความต้องการของผู้มีส่วนได้ส่วนเสียภายนอก เพื่อเป็นหลักฐานเพื่อแสดงความรับผิดชอบ
  - คำนึงถึงต้นทุนการสร้าง ต้นทุน **capturing** ต้นทุนการเก็บรักษาเอกสาร และความเสี่ยงที่อาจเกิดขึ้นกับองค์กรหากไม่มีเอกสาร

Shepherd, E., & Yeo, G. (2003). *Managing Records: A Handbook of Principles and Practice*. Facet Publishing.



# Capturing

- มาจากคำว่า **Capture** ใช้ในสาขาวิชาการจัดการเอกสาร หมายถึง กลไกที่ใช้เพื่อนำเอกสารเข้าสู่ระบบจัดการเอกสารเพื่อเก็บรักษาอย่างปลอดภัยและยาวนาน สามารถเข้าถึงได้เมื่อต้องการ
- เอกสารบางชนิด จะต้อง **Capture** ตั้งแต่ตอนสร้างเอกสาร บางชนิด ถูก **Capture** หลังจากเวลาผ่านไปช่วงหนึ่ง
- เอกสารทั่วไป สามารถทำลายทิ้งได้
- คำว่า **records** มักใช้เรียก เอกสารที่จะต้องได้รับการสงวนรักษาเพื่อใช้เป็นหลักฐานอ้างอิง ในขณะที่เอกสารทั่วไป จะใช้คำว่า **document**

Shepherd, E., & Yeo, G. (2003). *Managing Records: A Handbook of Principles and Practice*. Facet Publishing.



# ลักษณะของ Big Data-ข้อมูลเกินนับ

- ลักษณะตั้งต้น
  - มีปริมาณมหาศาล (**Volume**-ข้อมูลมีปริมาณมาก)
  - มีการเปลี่ยนแปลงรวดเร็ว (**Velocity**-ข้อมูลเปลี่ยนแปลงอย่างรวดเร็วตลอดเวลา)
  - มีรูปแบบหลากหลาย (**Variety** -ข้อมูลมีรูปแบบหลากหลาย)
- ลักษณะเพิ่มเติม
  - มีความถูกต้องมากหรือน้อย (**Veracity**-ความถูกต้อง)
  - มีประโยชน์ที่จะวิเคราะห์หรือไม่ (**Value**-คุณค่า)

Zgurovsky, M. Z., Zhurovs'kyi, M. Z., & Zaychenko, Y. P. (2019). *Big Data: Conceptual Analysis and Applications*. Springer.



# สื่อ

- สื่อที่ใช้เพื่อสร้างและบันทึกเนื้อความ โดย**ทั่วไป** คือ กระดาษ
- ในปัจจุบัน นิยมใช้ **digital media**
- เอกสารที่บันทึกใน **digital media** เรียกว่า **digital records** หรือ **electronic records**
- เนื้อความที่บันทึกในสื่อ มีสองประเภท
  - ตัวอักษร
  - ไม่ใช่ตัวอักษร



## กิจกรรมและการสร้างเอกสาร

- เอกสาร คือผลผลิตจากกิจกรรมขององค์กร โดยอาจถูกสร้างหรือได้รับ ระหว่าง หรือ หลังเสร็จสิ้นกิจกรรม
- เอกสาร บรรจุข้อเท็จจริงไว้ในเนื้อหาเพื่อใช้อ้างอิงการกระทำกิจกรรม
- โดยทั่วไป เอกสารสร้างโดยมนุษย์ ใช้ปากกา และหมึก หรือเครื่องพิมพ์ดีด แต่ในปัจจุบัน เอกสารสร้างโดยโปรแกรมคอมพิวเตอร์
- การรับเอกสาร อาจได้รับผ่านทางแฟกซ์ อีเมล จดหมาย หรือระบบอื่น ๆ





# ประเด็นศึกษาเกี่ยวกับ Big Data

- Capturing data
- Data storage
- Data analysis
- Search
- Sharing
- Transfer
- Visualization
- Querying
- Updating
- Information privacy
- Data source



# ภารกิจหอจดหมายเหตุเพื่อ Big Data

- Capturing data
- Data storage
- Search
- Transfer
- Information privacy
- Data source



## ความสัมพันธ์ของข้อมูล

- แม้ว่า **Big Data** จะใช้ข้อมูลจากแหล่งที่หลากหลาย ข้อมูลเหล่านั้นต้องมี**ความสัมพันธ์**บางส่วน (Attribute, Element, Column, Field) เช่น ข้อมูลการซื้อขาย กับข้อมูลการพยากรณ์อากาศ มีข้อมูลวันที่ประกอบอยู่



# การค้นคืน (Search)

- สร้างดัชนี (Concordance and indexing)
  - บัญชีคำสำคัญและตำแหน่งที่พบคำนั้น
- แยกคำสำคัญ (Term extraction)
  - The **diagnosis** is **chronic viral hepatitis**
- สร้างรหัสอ้างอิงเอง (Autocoding)
  - กำหนดรหัสและชุดคำที่อยู่ในกลุ่มเดียวกัน

```
C0206708|Cervical Intraepithelial Neoplasms
C0206708|Cervical Intraepithelial Neoplasm
C0206708|Intraepithelial Neoplasm, Cervical
C0206708|Intraepithelial Neoplasms, Cervical
C0206708|Neoplasm, Cervical Intraepithelial
C0206708|Neoplasms, Cervical Intraepithelial
C0206708|Intraepithelial Neoplasia, Cervical
C0206708|Neoplasia, Cervical Intraepithelial
C0206708|Cervical Intraepithelial Neoplasia
```



# Data และ Record ?

- แหล่งกำเนิด ผู้สร้าง?
- การแก้ไขข้อมูล?
- ความน่าเชื่อถือ?
- การรับมอบ/การส่งมอบ?



# Big Data และ Archives ?

- ปริมาณข้อมูล?
- ข้อมูลเปลี่ยนแปลงเร็ว?
- ข้อมูลมีความหลากหลาย?
- ระดับความถูกต้อง?
- คุณค่า?



## บทสรุป

- ข้อมูล **Data** ใน **Big Data** และเอกสาร **Record** ใน **Archival Science** อาจถือเป็นสิ่งเดียวกัน
- เอกสารจดหมายเหตุ **Archives** เป็นส่วนหนึ่งของข้อมูลเกินนับ **Big Data**
- เอกสารจดหมายเหตุดิจิทัล **Digital Archives** มีลักษณะเสมือน **Big Data**
- หอจดหมายเหตุเป็นแหล่งข้อมูลคุณภาพดีสำหรับการประมวล **Big Data**
- เอกสารจดหมายเหตุเป็นเอกสารที่ค้นพบได้ยาก หากไม่มีนักจดหมายเหตุคอยช่วยเหลือ
- กระบวนการจัดการจดหมายเหตุในปัจจุบัน จะต้องแทรกแซงวงจรชีวิตเอกสารตั้งแต่ตอนเริ่มสร้าง