



GBDi

Government Big Data Institute

สถาบันส่งเสริมการวิเคราะห์และบริหารข้อมูลขนาดใหญ่ภาครัฐ (สวช.)



งาน “จดหมายเหตุ” ความท้าทายเมื่อเข้าสู่ยุค Big Data

กฤตพัฒน์ รัตนภูผา

GBDi Team Meeting

29 พฤษภาคม 2566

การบริหารจัดการข้อมูลภาครัฐตามกรอบธรรมาภิบาลข้อมูลภาครัฐ

Data Governance : การกำกับดูแลข้อมูล

Accuracy Completeness Timeliness Security Privacy Connectedness Worthiness

Organization & Stewardship

Function of

- DG Committee
- Steward team
- Data Controller/Processor/User

Policy & Standard

- Data Management
- Data Quality Assurance
- Data Exchange

Data Risk Management Compliance & Data Audit

- Data Risk Management
- Data Law & Regulation
- Data Quality Audit

Building Knowledge & Awareness

- Program Coverage
- Measurement
- Ongoing

Data Management : การบริหารจัดการข้อมูล

Data Lifecycle

Foundation

Create

Store

Use

Public

Archive

Destroy

Data Catalog

- What data ?
- Who's the owner ?
- Search Tags
- Data sources

Data Security & Privacy

- Confidentiality
- Integrity
- Availability

Data Quality

- Accuracy
- Validity
- Timeliness
- Completeness
- Uniqueness
- Consistency

นิยาม

ประโยชน์

ความท้าทาย

จดหมายเหตุ (Archives) คืออะไร

“เอกสารจดหมายเหตุ” หมายความว่า เอกสารซึ่งเกิดจากการปฏิบัติงานของบุคคลหรือนิติบุคคล โดยเอกสารเหล่านั้นได้สิ้นกระแสการใช้งานแล้ว และได้รับการประเมินว่ามีคุณค่าอย่างต่อเนื่องต่อผู้ครอบครอง

— สมบัติ พิกุลทอง (2563), “การประเมินความต้องการจำเป็นในการสงวนรักษาเอกสารจดหมายเหตุดิจิทัลของหน่วยงานบริการจดหมายเหตุในประเทศไทย”



ทำลายทิ้ง
(Destroy)



จัดเก็บถาวร
(Archive)



การประเมิน



องค์กร

ผลิต



เอกสาร

สิ้นกระแสใช้งานแล้ว เช่น

1. ระเบียบนั้นยกเลิกบังคับใช้แล้ว เพราะมีระเบียบใหม่
2. โครงการที่เกี่ยวข้องกับเอกสารนี้ได้จบลง

จดหมายเหตุ (Archives) คืออะไร

“เอกสารจดหมายเหตุ” หมายความว่า เอกสารซึ่งเกิดจากการปฏิบัติงานของบุคคลหรือนิติบุคคล โดยเอกสารเหล่านั้นได้สิ้นกระแสการใช้งานแล้ว และได้รับการประเมินว่ามีคุณค่าอย่างต่อเนื่องต่อผู้ครอบครอง (Archival material)

ทั้งนี้ Archives ยังสามารถใช้ในความหมายได้อีก 2 นัยยะ คือ

- Archival agency เป็นองค์กรที่เก็บ archival material ซึ่งเรียกว่า Archives เช่นกัน
- Archival repository เป็นบริเวณ/สถานที่ที่เก็บ archival material ซึ่งเรียกว่า Archives เช่นกัน



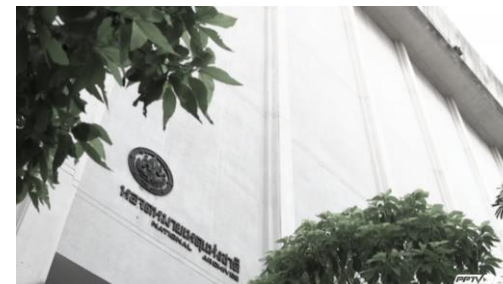
สำนักหอจดหมายเหตุแห่งชาติ
NATIONAL ARCHIVES OF THAILAND



หอภาพยนตร์ (องค์การมหาชน)
Film Archive (Public Organization)



Das
Bundesarchiv



อะไรเป็นจดหมายเหตุได้บ้าง



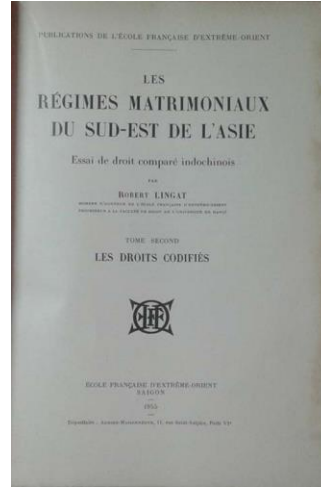
ศิลาจารึกหลักที่ 1
เป็นจดหมายเหตุได้
เพราะไม่อยู่ในกระแสการ
ใช้งานและก่อกำเนิดขึ้นจากการ
ปฏิบัติงานของนิติบุคคล
(รับคำสั่งจากราชวงศ์)



ประมวลกฎหมายตราสามดวง

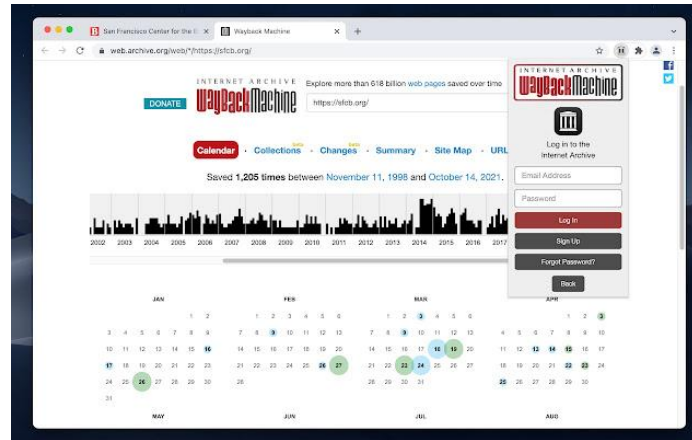
เป็นจดหมายเหตุได้

เพราะไม่อยู่ในกระแสการใช้งานและก่อกำ
เนิดขึ้นจากการชำระตามพระราชวินิจฉัย



หนังสือกฎหมายสมรส
เปรียบเทียบในเอเชียอาคเนย์
ของ โรแบร์ แลงกาต์

ไม่สามารถเป็นจดหมายเหตุได้
เพราะมีการตีพิมพ์ซ้ำ เป็นการ
เผยแพร่ ไม่ใช่การปฏิบัติงาน



เว็บไซต์ที่เก็บไว้ใน Wayback Machine

เป็นจดหมายเหตุได้

เพราะไม่อยู่ในกระแสการใช้งานและก่อกำ
เนิดขึ้นจากการพัฒนาเพื่อใช้ host ในที่ ๆ เดียว

ซอฟต์แวร์กับไลบรารี

ไม่สามารถเป็นจดหมายเหตุได้

เพราะมีการจำหน่ายซ้ำ เป็นการ
เผยแพร่ ไม่ใช่การปฏิบัติงาน

นายโจชัว ไม่ชื่อนิล
เลขที่ 52/33 หมู่ที่ 2 ตำบลบางเขน อำเภอเมืองนนทบุรี จังหวัดนนทบุรี
เลขที่ประจำตัวผู้เสียภาษีอากร 1101500075390
Tel : 081-1711963 Fax : 02-5741577

ใบเสร็จรับเงิน		
ต้นฉบับ		
ชื่อลูกค้า:	บริษัท ตัวอย่าง จำกัด	เลขที่: RV6105001
ที่อยู่:	999/99 ซอยแจ้งวัฒนะ 9 ถนนแจ้งวัฒนะ แขวงตลาดบางเขน เขตหลักสี่ กรุงเทพมหานคร 10210	วันที่: 1 พ.ค. 2561
TAX ID :	0105552133333	
ลำดับ Item	รายการ Description	จำนวนเงิน (บาท) Amount (Bath)
1	ค่าบริการประจำปีเดือนพฤษภาคม 2561	10,000.00
(รวมเงินบาทถ้วน)		จำนวนรวม 10,000.00
ชำระโดย : เงินสด จำนวนเงิน _____		
เช็ค	ธนาคาร _____ เลขที่ _____ ลงวันที่ _____	
จำนวนเงิน _____		
ลงชื่อ _____ วันที่ _____	ผู้รับเงิน _____	ลงชื่อ _____ ผู้มีอำนาจ _____

ใบเสร็จรับเงิน

เป็นจดหมายเหตุได้

เพราะไม่อยู่ในกระแสการใช้งาน (การชำระ
หนี้เสร็จสิ้นตามกฎหมายแพ่งและพาณิชย์)
และเกิดขึ้นจากการปฏิบัติของทั้งฝ่ายผู้ซื้อ
และผู้ขาย ซึ่งเป็นบุคคล/นิติบุคคล

งาน “จดหมายเหตุ”

นิยาม

ประโยชน์

ความท้าทาย



บันทึกความทรงจำของบุคคล/นิติบุคคล/
สังคมเป็นคุณค่าทางความทรงจำ เพื่ออนุรักษ์
ให้ดำรงสืบไป



เป็นส่วนหนึ่งในการยกระดับคุณภาพ
กระบวนการยุติธรรมและความโปร่งใส
ทำให้สามารถเห็นกระบวนการทำงานของบุคคล/
นิติบุคคล/สังคมทั้งหมดในอดีตและเป็นบทเรียน
ของบุคคล/นิติบุคคล/สังคม



งาน “จดหมายเหตุ”

นิยาม

ประโยชน์

ความท้าทาย

ความท้าทายที่จะทำให้บรรลุผลในยุคสมัยนี้



1. ข้อมูลมีความหลากหลายมากขึ้น ในยุค big data



3. ข้อมูลมีมากทำให้ เทคนิคในการหา insights และจัดการเอกสารที่ซับซ้อนยิ่งขึ้นจึงเป็นเรื่องสำคัญ



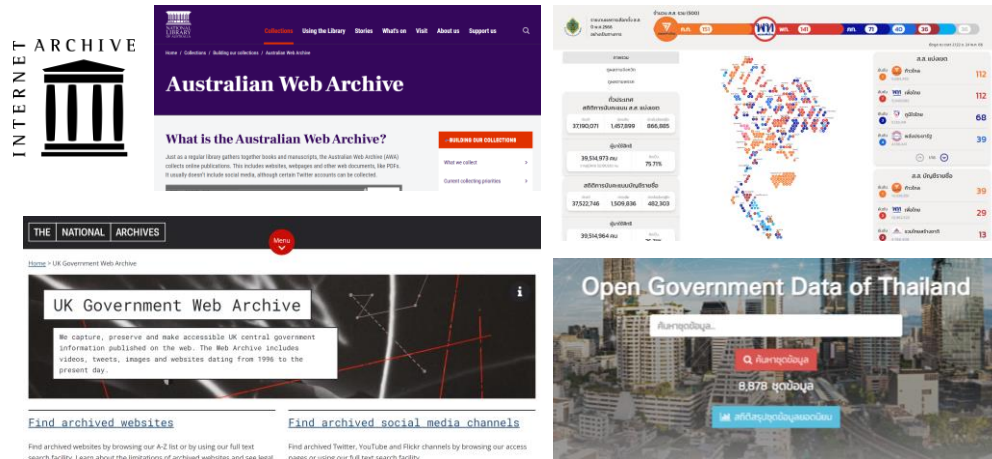
2. หากต้องการใช้ข้อมูลเก่าในการวิเคราะห์ digitization จึงมีบทบาทสำคัญ



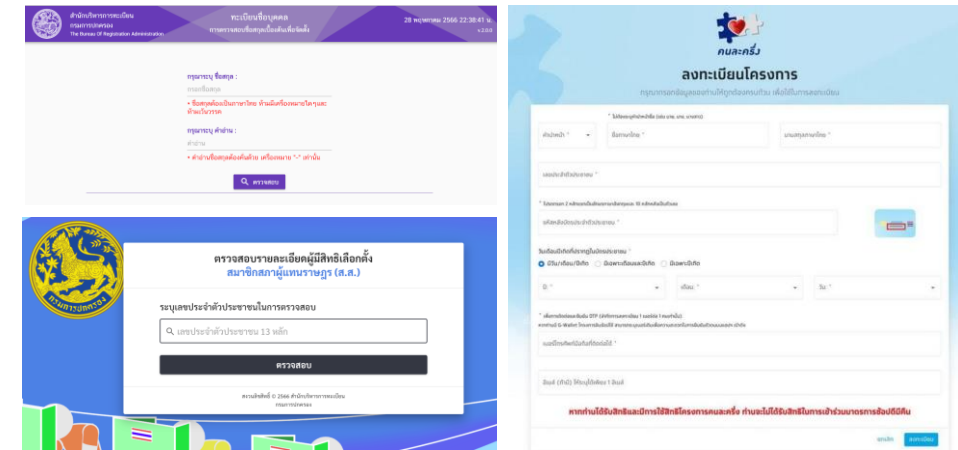
4. ข้อมูลมีบริบทที่หลากหลายทำให้ คนที่ดีความต้องมีความรู้ที่รอบด้าน

ความท้าทายที่จะทำให้บรรลุผลในยุคสมัยนี้

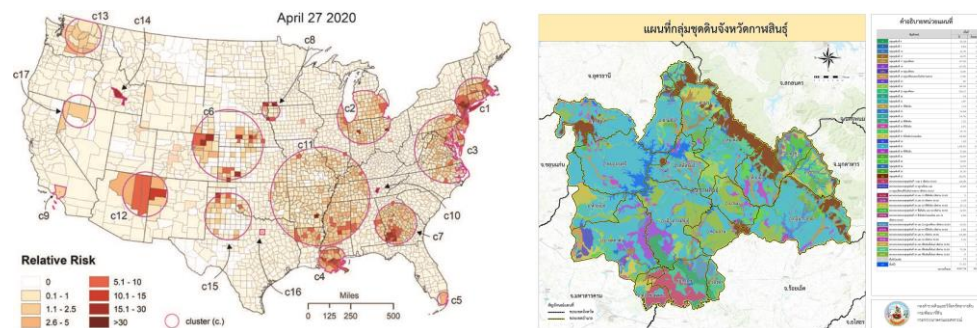
1. ข้อมูลมีความหลากหลายมากขึ้นในยุค Big data



ข้อมูลเว็บไซต์



ข้อมูล e-government services, e-billing, e-payment



ข้อมูลสถิติ ทั้งในเชิงภูมิสารสนเทศและสารสนเทศอื่น ๆ



ข้อมูลประเภทอื่น ๆ ที่อาจเป็น streaming data, graph data ฯลฯ

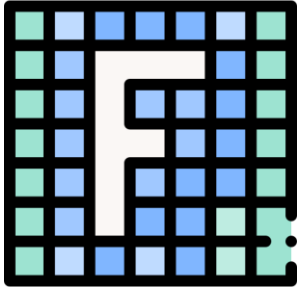
ความท้าทายที่จะทำให้บรรลุผลในยุคสมัยนี้



1. ข้อมูลมีความหลากหลายมากขึ้น ในยุค big data



3. ข้อมูลมีมากทำให้ เทคนิคในการหา insights และจัดการเอกสารที่ซับซ้อนยิ่งขึ้นจึงเป็นเรื่องสำคัญ



2. หากต้องการใช้ข้อมูลเก่าในการวิเคราะห์ digitization จึงมีบทบาทสำคัญ



4. ข้อมูลมีบริบทที่หลากหลายทำให้ คนที่ตีความต้องมีความรู้ที่รอบด้าน

ความท้าทายที่จะทำให้บรรลุผลในยุคสมัยนี้



2. หากต้องการใช้ข้อมูลเก่าในการวิเคราะห์ digitization จึงมีบทบาทสำคัญ

๑ คำนำ กฎหมาย ลักษณะ ทาส ๖๖

๑ ท่าน ทั้งปวง อัน ต้อง การ ใช้ เรื่อง กฎหมาย ยอยาก รู้ ความ เดียด หมด จดใน เรื่อง ความ ที่ จะ เรียบ ที่ จะ ให้ ชำนาญ เพราะ เหตุ เช่น นี้ ข้าพเจ้า ครุ สมิต ได้ คิด ความ แต่ กฎหมาย เก็บ เอา กฎหมายเก่า พระราชกำหนด แล หมายถึง ประกาศเก่า แต่ ใหม่ มา ประมวล เข้า เป็น เรื่อง หนึ่งใน หนังสือ เล่ม นี้ ว่า แต่ เพียง ลักษณะ ทาส เด็ดจ กรมหมื่น พิศิต ได้ ช่วย จัด เรียง ข้อ ก่ง ๆ ประสงค์ ความ จะ ให้ ท่าน ทั้งหลาย รู้ กฎหมาย ลักษณะ ทาส ให้ เดียด กว่า แต่ ก่อน เป็น อัน มาก

๑ พระบาทสมเด็จพระจุลจอมเกล้าเจ้ากรุงสยาม ใน บัดจุบัน นี้ ได้ มี พระไทย เหน็ด ปรานี แก่ อาณาประชาราษฎร์ เป็น อัน ยิ่ง แต่ ได้ ดำรง พระราช อหิยาไศรย แก่ แผ่นดิน บ้าน เมือง ประเทศ สยาม นี้ หลาย ครั้ง หนัก หนา ควร ที่ คน ทั้งปวง ทั้ง ตลอด พระราช อาณา เขต จะ ดำรง ความ มีดี ยินดี ตั้ง ใจ กระตัญญู ประพฤติ การ โดย ชื่อ สุจริต ให้ สม ความ ดี ที่ ใน หลวง แผ่นดิน นี้ หมายถึง พระไทย ทำ คุณ ให้ แก่ คน ทั้งปวง

๑ ไม่ช้า ไม่นาน ทาส ก็ จะ หมด สูญ จาก ประเทศ สยาม เมื่อ การ นั้น จะ ดำรง แล้ว นาย เงิน เจ้า ทาส แล ลูก นี้ เป็น ทาส เรือน เบี้ย จะ เป็น อัน เล็ก ก้น เจ้า ทาส จะ ชุม ชี เบียด เบี้ยว ขาย หญิง ว่า เป็น ทาส เกณฑ์ ใช้ ให้ ทำ การ ของ เจ้า เงิน เปลา ๆ ก็ ไม่ได้ ด้วย ว่า หมด ทาส แล้ว บันดา คน ที่ เกิด เป็น ลูก ทาส ตั้งแต่ จุลศักราช ๑๒๓๐ ปี มะโรง สัมฤทธิศก ถัด มา ช้าง หน้า ตาม หมายถึง ประกาศ แล กฎหมาย ซึ่ง พระบาทสมเด็จพระจุลจอมเกล้าเจ้ากรุงสยาม ตั้ง มา แล้ว นั้น มี เวลา กำหนด ที่ จะ หลุด พัน จาก ซึ่ง เป็น ทาส ลูก ทาส ทั้ง หญิง ทั้ง ชาย เหล่า นี้ คง มี คำ คัว ตั้งแต่ วันเกิด จน กระทั่ง อายุ ๒๑ ปี ครั้น ถึง กำหนด อายุ ๒๑ ปี หมด คำ คัว ทุก คน คำ คัว ลูก ทาส เหล่า นั้น มี แต่ จะ มาก ช้น ตั้งแต่ วันเกิด จน กระทั่ง ถึง อายุ แปด ขวบ ตั้งแต่ อายุ แปด ขวบ คำ คัว นั้น มี แต่ จะ ตก น้อย ลง ไป จน ถึง อายุ ๒๑ ปี คำ คัว นั้น ก็ หมด ไม่ มี อีก แล้ว เมื่อ หมด คำ คัว แล้ว กฎหมาย ห้าม แล้ว ว่า ลูก ทาส ที่ หมด คำ คัว แล้ว จะ เป็น ทาส คือ ไป อีก ไม่ได้ ผู้ ที่ จะ ขาย จะ ซื้อ คน ที่ หลุด คำ คัว แล้ว คง มี โทษ ตาม พระราชกำหนด บันดา คน อัน มี ได้ เป็น ลูก ทาส แต่ เกิด ใน ปี มะโรง สัมฤทธิศก แล ถัด ปี นั้น มา ช้าง หน้า ท่าน ตั้ง พระราชกำหนด ห้าม ไม่ ให้ เป็น อัน ขาย ได้ ผู้ ที่ จะ ช้น ขาย ช้น ซื้อ คน เช่น นั้น แล้ว คง เป็น โทษ ตาม พระราชกำหนด

กฎหมายตราสามดวงลักษณะทาส ฉบับครุสมิต (2423)

หน้า ๕

เล่ม ๑๔๐ ตอนที่ ๒๓ ก ราชกิจจานุเบกษา ๒๒ มีนาคม ๒๕๖๖

ประกาศคณะกรรมการการเลือกตั้ง

เรื่อง จำนวนสมาชิกสภาผู้แทนราษฎรแบบแบ่งเขตเลือกตั้งแต่ละจังหวัดจะมี และจำนวนเขตเลือกตั้งแบบแบ่งเขตเลือกตั้งของแต่ละจังหวัด

ด้วยได้มีพระราชกฤษฎีกาอุปสภาผู้แทนราษฎร พ.ศ. ๒๕๖๖ เพื่อให้มีการเลือกตั้งสมาชิกสภาผู้แทนราษฎรใหม่เป็นการเลือกตั้งทั่วไป

อาศัยอำนาจตามความในมาตรา ๑๒ แห่งพระราชบัญญัติประกอบรัฐธรรมนูญว่าด้วยการเลือกตั้งสมาชิกสภาผู้แทนราษฎร พ.ศ. ๒๕๖๑ และที่แก้ไขเพิ่มเติม ประกอบข้อ ๔๖ ของระเบียบคณะกรรมการการเลือกตั้งว่าด้วยการเลือกตั้งสมาชิกสภาผู้แทนราษฎร พ.ศ. ๒๕๖๖ และประกาศสำนักทะเบียนกลาง เรื่อง จำนวนราษฎรทั่วราชอาณาจักร ตามหลักฐานการทะเบียนราษฎร ณ วันที่ ๓๑ ธันวาคม พ.ศ. ๒๕๖๕ ลงวันที่ ๕ มกราคม พ.ศ. ๒๕๖๖ คณะกรรมการการเลือกตั้ง จึงประกาศให้ทราบโดยทั่วกัน ดังนี้

๑. ราษฎรทั่วราชอาณาจักร ณ วันที่ ๓๑ ธันวาคม พ.ศ. ๒๕๖๕ มีจำนวน ๖๕,๑๐๖,๔๘๑ คน

๒. จำนวนราษฎรโดยเฉลี่ย ๑๖๒,๗๖๖ คน ต่อสมาชิกสภาผู้แทนราษฎรหนึ่งคน

๓. จำนวนสมาชิกสภาผู้แทนราษฎรและเขตเลือกตั้งแบบแบ่งเขตเลือกตั้งแต่ละจังหวัด จะมีและจำนวนเขตเลือกตั้งแบบแบ่งเขตเลือกตั้งของแต่ละจังหวัด

ลำดับที่	จังหวัด	จำนวนสมาชิกสภาผู้แทนราษฎร	จำนวนเขตเลือกตั้ง	เขตเลือกตั้งที่	ท้องที่ที่ประกอบเป็นเขตเลือกตั้ง
๑	ตราด	๑	๑	๑	ทั้งจังหวัด
๒	ระนอง	๑	๑	๑	ทั้งจังหวัด
๓	สมุทรสงคราม	๑	๑	๑	ทั้งจังหวัด
๔	สิงห์บุรี	๑	๑	๑	ทั้งจังหวัด
๕	ชัยนาท	๒	๒	๑	๑. อำเภอเมืองชัยนาท ๒. อำเภอสรรพยา ๓. อำเภอโนนชัย ๔. อำเภอวัดสิงห์ (เฉพาะตำบลชะครามเก่า ตำบลหนองน้อย ตำบลหนองบัว เทศบาลตำบลวัดสิงห์ และตำบลหนองจันท)

ประกาศ กกต. ว่าด้วยจำนวน ส.ส. เขต (2566)

ตัวอักษรที่ถูกทำ Romanization แล้ว

1. เกิดขึ้นในสมัยรัชกาลที่ 3 (ฉบับแรกสุด คือ ประกาศห้ามสูบฝิ่นในราชอาณาจักร)
2. ตัวพิมพ์ที่นิยมมีไม่กี่รูปแบบ
 1. ตัวบรัดเลย์
 2. ตัว ฝ.ศ. (คล้าย Angsana)
 3. ตัวทอมโลห์ (คล้าย Cordia)
 4. ตัวโป่งแซ
 5. ตัวไม่มีหัว (เช่น Sukhumvit หรือ Kanit)
3. มีลักษณะการวางบรรทัดชัดเจน

ทำให้การทำ OCR เป็นเรื่องง่าย

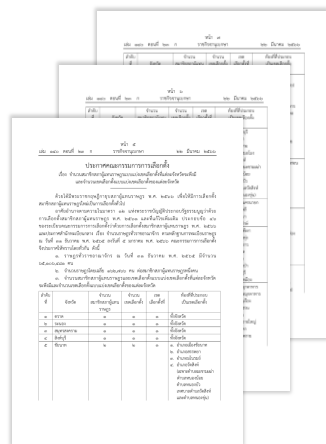
ความท้าทายที่จะทำให้บรรลุผลในยุคสมัยนี้

2. หากต้องการใช้ข้อมูลเก่าในการวิเคราะห์ digitization จึงมีบทบาทสำคัญ

ทำให้การทำ OCR เป็นเรื่องง่าย



ไฟล์ PDF



ไฟล์ภาพ lossless (PNG)

ตรวจจับเส้นตาราง
(Straight Line Detection)



ตรวจจับกลุ่มตัวอักษร
(Region of interest detection)



GBDi
แปลงตารางเป็น CSV



สร้างโครงสร้างเป็น
Markdown

๓. อำเภอกันทรวิชัย

รู้จำตัวอักษร
(Text recognition)



Tesseract OCR

Customized
Tesseract
Engine

๓. อำเภอกันทรวิชัย

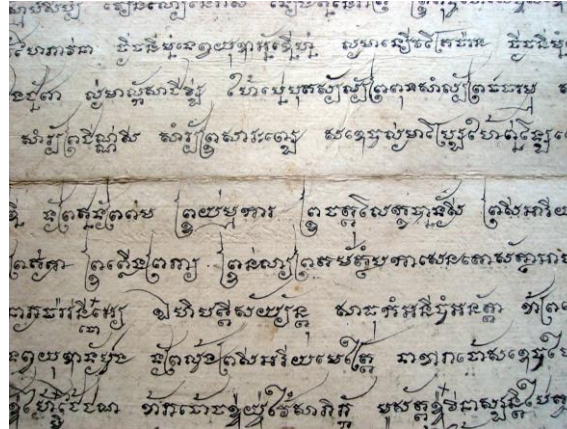
Text region

ความท้าทายที่จะทำให้บรรลุผลในยุคสมัยนี้

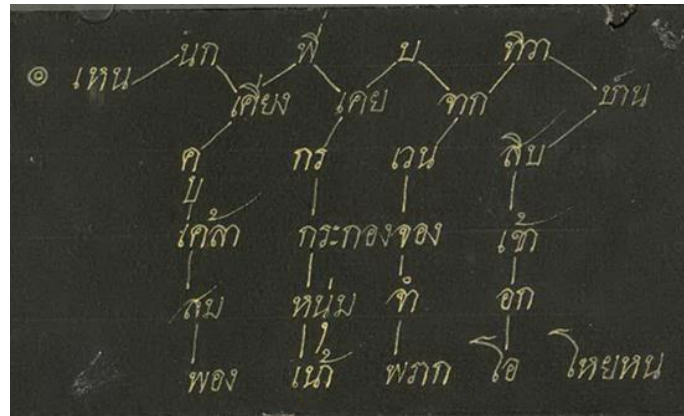
2. หากต้องการใช้ข้อมูลเก่าในการวิเคราะห์ digitization จึงมีบทบาทสำคัญ



พงศาวดารฉบับหลวงประเสริฐ



สมุดไทยแบบอักษรขอมไทย



ตัวอย่างกลบท



คัมภีร์แผนผีดาษ

ตัวอักษรอื่น ๆ ที่ยังไม่ได้ทำ romanization
ตัวอักษรที่เป็นลายมือ
หรือตัวอักษรอื่นที่ไม่ใช้อักษรไทย

1. ยากต่อการทำ transcribe ออกมาเป็น text
2. อาจให้ผู้เชี่ยวชาญเป็นคนแกะ
3. ทำ digitization โดยไม่มีการแกะใด ๆ ทั้งสิ้น

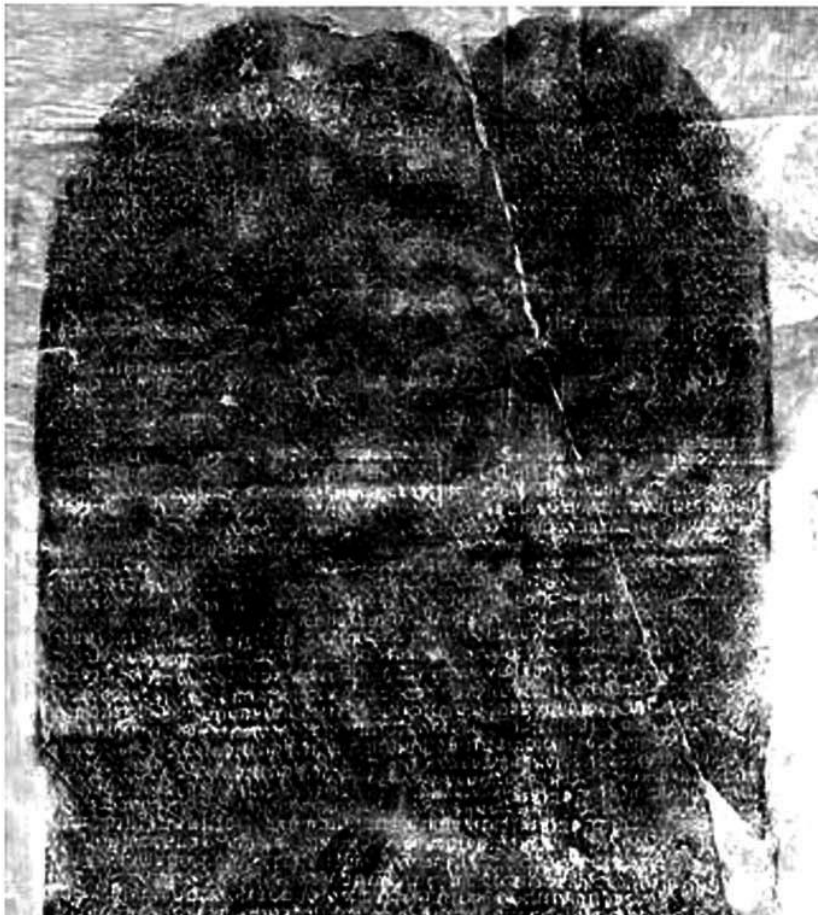
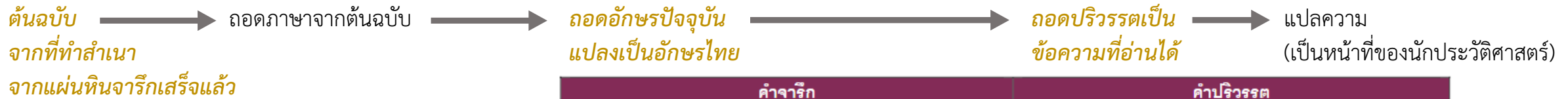
กระบวนการในการแกะโดยปกติในทางโบราณคดี คือ

- **ต้นฉบับ**
- ถอดภาษาจากต้นฉบับ
- **ถอดอักษรปัจจุบันแปลงเป็นอักษรไทย**
- **ถอดปริวรรตเป็นข้อความที่อ่านได้**
- แปลความ

ความท้าทายที่จะทำให้บรรลุผลในยุคสมัยนี้

2. หากต้องการใช้ข้อมูลเก่าในการวิเคราะห์ digitization จึงมีบทบาทสำคัญ

กระบวนการในการแกะโดยปกติในทางโบราณคดี คือ



คำจารึก	คำปริวรรต
๑. ^(๑)	๑.
๒.	๒.
๓.สเด็จพระตากตเป็น.....	๓.สมเด็จพระตากตเป็น.....
๔.มท(ยง)คน เมื่อสเด็จพระมหาราษฎร์มีสรีมหาราตุ	๔.มทียังคนะ เมื่อสมเด็จพระมหาราษฎร์มีสรีมหาราตุ ^(๒)
๕. อยู่.....สเด็จพระมหาราษฎร์มีสรีมหาราษฎร์ สรีรัตนลงกาที่บมหลามมีเป	๕. อยู่.....สมเด็จพระมหาราษฎร์มีสรีรัตนลงกาที่บ ^(๓) มหลามมีเป-
๖. (นเจ้า).....ไกลฝั่งน้ำมาวลิกคงคาในลงกาทวิป นอกเมืองกำพลไถเลิง...ียสเด็จพระ	๖. นเจ้า.....ไกลฝั่งน้ำมาวลิกคงคา ^(๔) ในลงกาทวิป นอกเมืองกำพลไถ ^(๕) เลิง ^(๖) ..ย สมเด็จพระ
๗.ที่นบกพระทนต์ราตุคนเจดีย์มีสาม(อ)นน มีญอคอนน	๗.ที่นบกพระทนต์ราตุคนเจดีย์มีสามอัน มีญอคอนน -
๘. (งพระมหาราษฎร์มีสรีมหาราษฎร์) เกิดในนครหลวงสองแควปู่ชื้อพระยาศรีนาวนาถ (เป็นขุน)เป็นพ	๘. ง พระมหาราษฎร์มีสรีรัตนลงกาที่บมหลามมีเป เกิดในนครหลวงสองแคว ^(๗) ปู่ชื้อพระยาศรีนาวนาถ เป็นขุน เป็นพ
๙.เสวยราชในนครสองอนนอนนิงซี	๙.เสวยราชในนครสองอัน อันหนึ่งชื้อ

ฐานข้อมูลจารึกในประเทศไทย | จารึกวัดศรีชุม ด้านที่ 1 (sac.or.th)



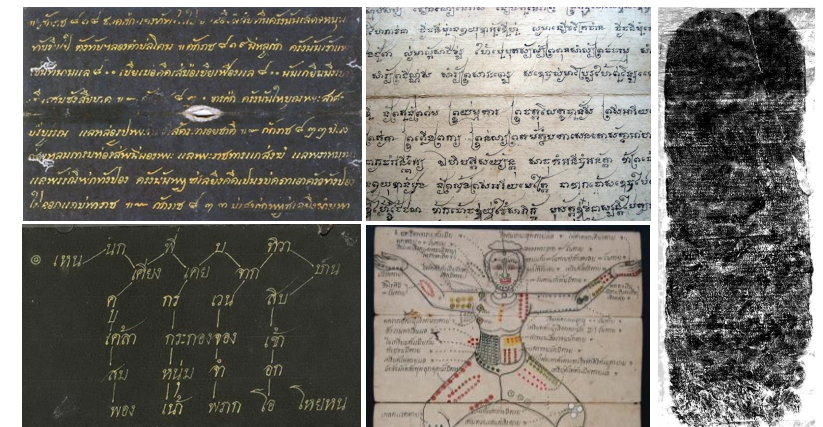
Non-digital-born archives

Unstructured documents

[illegible][illegible]

ปัญหาที่เกิดขึ้น

1. เก็บไว้ที่ไหน
2. จัดการบัญชีอย่างไร
3. หาอย่างไรให้พบได้ง่ายที่สุด
4. จะเชื่อมโยงความสัมพันธ์ระหว่างเอกสารอย่างไร



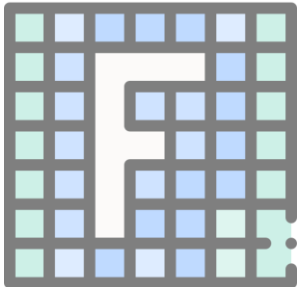
ความท้าทายที่จะทำให้บรรลุผลในยุคสมัยนี้



1. ข้อมูลมีความหลากหลายมากขึ้น ในยุค big data



3. ข้อมูลมีมากทำให้ เทคนิคในการหา insights และจัดการเอกสารที่ซับซ้อนยิ่งขึ้นจึงเป็นเรื่องสำคัญ



2. หากต้องการใช้ข้อมูลเก่าในการวิเคราะห์ digitization จึงมีบทบาทสำคัญ



4. ข้อมูลมีบริบทที่หลากหลายทำให้ คนที่ตีความต้องมีความรู้ที่รอบด้าน

ความท้าทายที่จะทำให้บรรลุผลในยุคสมัยนี้

3. ข้อมูลมีมากทำให้เทคนิคในการหา insights ที่ซับซ้อนยิ่งขึ้นจึงเป็นเรื่องสำคัญ



การใช้งานฐานข้อมูลภาครัฐเพื่องานด้านสังคมศาสตร์และมนุษยศาสตร์

แบบสอบถามนี้จัดทำเพื่อสอบถามข้อมูลเบื้องต้นเกี่ยวกับการค้นหาข้อมูลในฐานข้อมูลภาครัฐเพื่อทำงานทางสังคมศาสตร์และมนุษยศาสตร์ ซึ่งอาจนำไปต่อยอดวิธีการใช้ประโยชน์ข้อมูลสาธารณะภาครัฐสำหรับงานด้านสังคมศาสตร์และมนุษยศาสตร์ในอนาคต

จากการสอบถามทั้งสิ้น 9 คน

โดยเป็นการสอบถามแบบตอบรายละเอียด

Facts ที่ได้จากแบบสอบถามนี้

1. คำตอบที่ได้มาจากนิสิต/นักศึกษา อาจารย์ และเจ้าหน้าที่รัฐ
2. เกือบทุกคนใช้บริการอย่างน้อยเดือนละครั้ง
3. ส่วนใหญ่จะหยิบอ้างอิงราชกิจจานุเบกษาเป็นหลัก
ชุดข้อมูลอื่นที่ใช้มีเช่น
 - data.go.th
 - หอสมุดแห่งชาติ
 - หอจดหมายเหตุแห่งชาติ
 - สำนักงานคณะกรรมการกฤษฎีกา
 - หอสมุดรัฐสภา
 - หอภาพยนตร์
 - มติคณะรัฐมนตรี
 - สำนักงานสถิติแห่งชาติ
 - ฐานข้อมูลวัดในประเทศไทย
 - GIS กรมศิลปากร
 - หอสมุดมหาวิทยาลัย
 - สำนักผังเมือง กทม.
 - TDC ฯลฯ

ประสบการณ์ของนักค้นคว้า



ห้องบรรณนิเอกสารลายลักษณ์
หอจดหมายเหตุแห่งชาติ



ห้องตู้พระธรรมและห้องเอกสารโบราณ
หอสมุดแห่งชาติ



ตู้ส่งไมโครฟิล์ม

จากแบบสอบถาม

1. ฐานข้อมูลไม่เป็นระบบ กระจายกระจาย
นำมาใช้วิเคราะห์ยาก
2. ค้นหายาก ไม่พบข้อมูลที่ต้องการ
3. ข้อมูลเก่าตกหล่น
4. ข้อมูลใหม่ไม่อัปเดต

ความท้าทายที่จะทำให้บรรลุผลในยุคสมัยนี้

3. ข้อมูลมีมากทำให้เทคนิคในการหา insights ที่ซับซ้อนยิ่งขึ้นจึงเป็นเรื่องสำคัญ

National Archives by the Number

(as of September 30, 2022)

13.5 billion pieces of paper

>700K artifacts

448 million feet of film

40 million photographs

40 million aerial images

10 million maps, charts, and architectural/engineering drawings

835 terabytes electronic records



หอจดหมายเหตุแห่งชาติ

สรุปยอดปริมาณเอกสารในคลังเก็บเอกสาร

- เอกสารลายลักษณ์ จำนวน 10,285 ชุด ประมาณ 15 ล้านแผ่น
- เอกสารโสตทัศนจดหมายเหตุ
 - ฟิล์มกระจก จำนวน 31,289 แผ่น
 - ภาพ จำนวน 451,504 ภาพ
 - ฟิล์ม จำนวน 816,118 ภาพ
 - แผนที่ แผนที่ แบบแปลน จำนวน 20,187 แผ่น
 - โปสเตอร์ จำนวน 2,496 แผ่น
 - ปฏิทิน จำนวน 4,472 แผ่น
 - แถบบันทึกเสียง จำนวน 4,467 ม้วน
 - วีดิทัศน์ จำนวน 3,941 ม้วน
 - ไมโครฟิล์ม จำนวน 9,731 ม้วน
 - ซีดี จำนวน 734 แผ่น
 - แถบบันทึกภาพดิจิทัล จำนวน 34 ม้วน
- เอกสารเขียนเล่ม สิ่งพิมพ์รัฐบาล หนังสือหายาก จำนวน 44,023 เล่ม ประมาณ 1.5 ล้านแผ่น
- เอกสารการประชุม บันทึกความทรงจำ บันทึกเหตุการณ์ จำนวน 1,867 เรื่อง
- เอกสารข่าว จำนวน 677,443 ข่าว

ความท้าทายที่จะทำให้บรรลุผลในยุคสมัยนี้

3. ข้อมูลมีมากทำให้เทคนิคในการหา insights ที่ซับซ้อนยิ่งขึ้นจึงเป็นเรื่องสำคัญ

มีเทคนิคมากมายที่ใช้ในการจัดการข้อมูลให้ ง่ายยิ่งขึ้นและเพิ่มศักยภาพในการค้นหา

ตาม NARA Datasets for Artificial
Intelligence/Machine Learning

1. Face recognition, Image captioning
2. Recommender system ตามความสนใจ
ของผู้ค้นคว้า
3. Catalog API

ตามหลักค้นคว้าวางจดหมายเหตุ/บันทึกสากล

1. ISAD(G)
2. Function analysis และ File plan
3. Continuum model (ธรรมชาติของข้อมูล)

ตาม UK National Web Archives ร่วมกับ Alan
Turing Institute

1. Knowledge graph construction
2. Semantic Searching
3. Algorithm transparency and trust

Semantic annotation

1. Logic-based model: [Petr Křemen และ Martin Nečaský \(2019\)](#)
2. Pattern-based and Machine Learning-based model: [Reeve และ Han \(2005\)](#)

Tools อื่น ๆ

1. ElasticSearch
2. Linked data

ความท้าทายที่จะทำให้บรรลุผลในยุคสมัยนี้



1. ข้อมูลมีความหลากหลายมากขึ้น ในยุค big data



3. ข้อมูลมีมากทำให้ เทคนิคในการหา insights และจัดการเอกสารที่ซับซ้อนยิ่งขึ้นจึงเป็นเรื่องสำคัญ



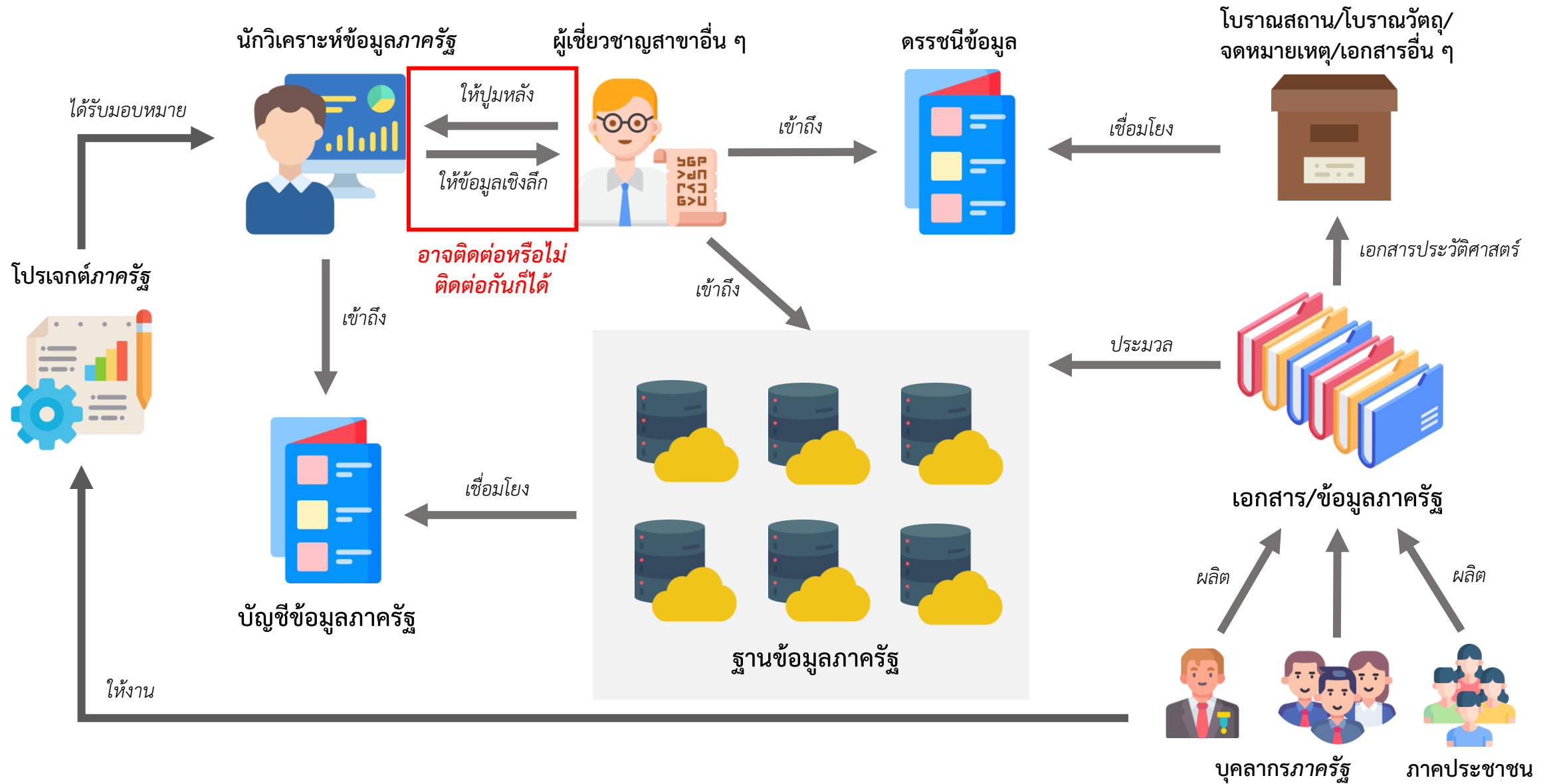
2. หากต้องการใช้ข้อมูลเก่าในการวิเคราะห์ digitization จึงมีบทบาทสำคัญ



4. ข้อมูลมีบริบทที่หลากหลายทำให้ คนที่ดีความต้องมีความรู้ที่รอบด้าน

ความท้าทายที่จะทำให้บรรลุผลในยุคสมัยนี้

4. ข้อมูลมีบริบทที่หลากหลายทำให้คนที่มีความต้องการมีความรู้ที่รอบด้าน



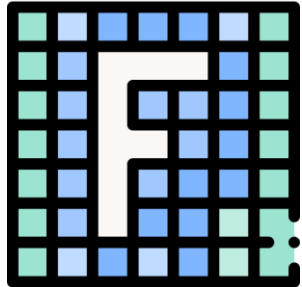
ความท้าทายที่จะทำให้บรรลุผลในยุคสมัยนี้



1. ข้อมูลมีความหลากหลายมากขึ้น ในยุค big data



3. ข้อมูลมีมากทำให้ เทคนิคในการหา insights และจัดการเอกสารที่ซับซ้อนยิ่งขึ้นจึงเป็นเรื่องสำคัญ



2. หากต้องการใช้ข้อมูลเก่าในการวิเคราะห์ digitization จึงมีบทบาทสำคัญ



4. ข้อมูลมีบริบทที่หลากหลายทำให้ คนที่ดีความต้องมีความรู้ที่รอบด้าน

นอกจากนี้ ความท้าทายก็มีใน Open Government Data Project เช่นกัน

ODI, DDI, OGD, Interoperable Europe, data.gouv.fr,
Linked Open Government Data, National Data Strategy, etc.

การบริหารจัดการข้อมูลภาครัฐตามกรอบธรรมาภิบาลข้อมูลภาครัฐ

Data Governance : การกำกับดูแลข้อมูล

Accuracy Completeness Timeliness Security Privacy Connectedness Worthiness

Organization & Stewardship

Function of

- DG Committee
- Steward team
- Data Controller/Processor/User

Policy & Standard

- Data Management
- Data Quality Assurance
- Data Exchange

Data Risk Management Compliance & Data Audit

- Data Risk Management
- Data Law & Regulation
- Data Quality Audit

Building Knowledge & Awareness

- Program Coverage
- Measurement
- Ongoing

Data Management : การบริหารจัดการข้อมูล

Data Lifecycle

Foundation

Create

Store

Use

Public

Archive

Destroy

Data Catalog

- What data ?
- Who's the owner ?
- Search Tags
- Data sources

Data Security & Privacy

- Confidentiality
- Integrity
- Availability

Data Quality

- Accuracy
- Validity
- Timeliness
- Completeness
- Uniqueness
- Consistency

ผู้ที่มีส่วนเกี่ยวข้องต่อการค้นคว้าในบทความนี้

ผู้เรียบเรียง
และที่ปรึกษา



ผู้อุเคราะห์ข้อมูล

สังคมศาสตร์และ
มนุษยศาสตร์



คุณศุภวิชญ์ บุตรดีสุวรรณ
นักประวัติศาสตร์สมัครเล่น



ศ. (พิเศษ) ธงทอง จันทรางศุ
อดีตปลัดสำนักนายกรัฐมนตรี



คุณหฤษฎ์ ทะวะบุตร
นิสิตคณะอักษรศาสตร์
จุฬาลงกรณ์มหาวิทยาลัย



คุณสิทธิกันต์ ธีระวัฒนชัย
Data Storyteller (PunchUp)



ผู้ตอบแบบสอบถามเกี่ยวกับการใช้
ข้อมูลภาครัฐในการทำวิจัยด้าน
สังคมศาสตร์และมนุษยศาสตร์

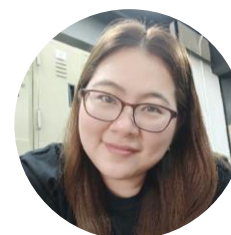
ดิจิทัล



ดร.พนชิต กิตติปัญญางาม
ผู้ก่อตั้งและกรรมการ
ผู้จัดการใหญ่ ZTRUS



อาจารย์เทพชัย ทรัพย์นิธิ
อุปนายกสมาคม AIAT



งานห้องสมุด
จดหมายเหตุ และ
เอกสารโบราณ



อาจารย์ ดร.พิมพ์พจน์ สีลาเขต
ภาควิชาบรรณารักษศาสตร์
จุฬาลงกรณ์มหาวิทยาลัย



เจ้าหน้าที่สำนักหอสมุดแห่งชาติ
(งานกลุ่มหนังสือตัวเขียนและจารึก)
และหอจดหมายเหตุแห่งชาติ