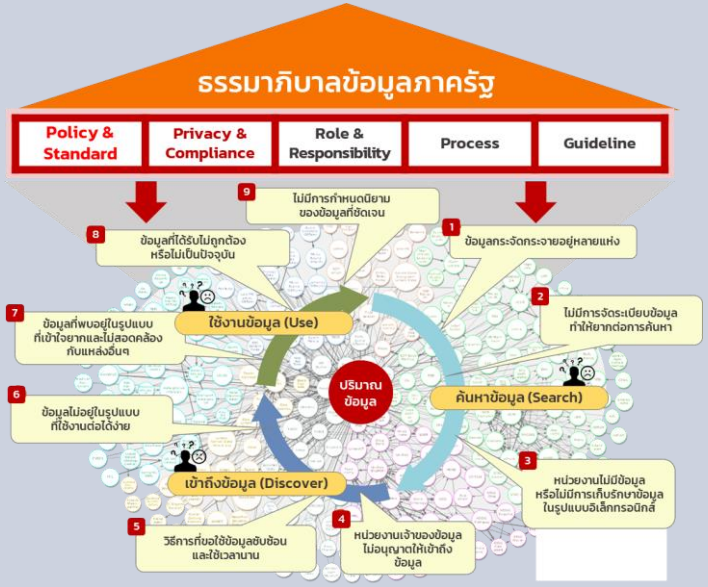
DIGITAL GOVERNMENT DEVELOPMENT AGENCY

มสพร. 14-2567

มาตรฐานสำนักงานพัฒนารัฐบาลดิจิทัล (องค์การมหาชน)

ว่าด้วยแนวทางการจัดทำข้อมูลนิรนาม (DATA ANONYMIZATION)

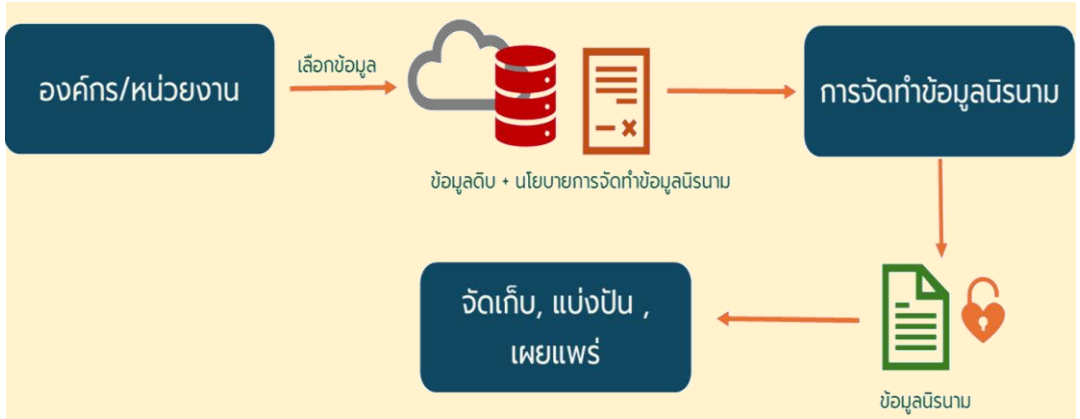
ความเป็นมา



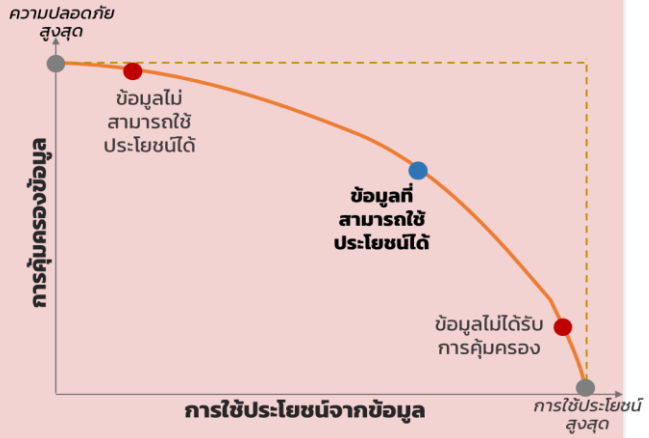
ธรรมาภิบาลข้อมูลจะช่วยให้ข้อมูลได้รับการป้องกันการละเมิดข้อมูล ***โดยการลดความสามารถในการระบุตัวตนลง*** *เพื่อให้สามารถนำข้อมูลไปใช้ประโยชน์ได้*

การจัดทำข้อมูลนิรนาม

ไม่เพียงแต่เกี่ยวข้องกับการลบข้อมูลที่สามารถระบุตัวตนได้โดยตรงเท่านั้น แต่ยังรวมถึงหลักคิดในการวิเคราะห์ว่าข้อมูลที่เหลืออยู่สามารถถูกใช้ในการระบุตัวตนได้หรือไม่ หากมีความเป็นไปได้ จำเป็นต้องมีการปรับเปลี่ยนหรือเพิ่มเทคนิคการป้องกันเพื่อลดความเสี่ยงนี้



การสร้างสมดุลระหว่างการคุ้มครองข้อมูลและการใช้ประโยชน์



*ข้อมูลที่มี**การคุ้มครองที่เคร่งครัด**ย่อมนำไปใช้ประโยชน์ได้น้อย ในขณะที่ข้อมูลที่ใช้ประโยชน์ได้เต็มที่ย่อมเสี่ยงต่อการถูกละเมิด ดังนั้นจึงควรมีการหาสมดุลระหว่างการคุ้มครองข้อมูลและการใช้ประโยชน์*

บทที่1 แนวคิด และ เทคนิคพื้นฐานในการไม่เปิดเผยข้อมูล

DIGITAL GOVERNMENT DEVELOPMENT AGENCY

❑ วัตถุประสงค์

1. การปกป้องข้อมูลส่วนบุคคล
2. การรักษาความเป็นส่วนตัว
3. การใช้ข้อมูลอย่างมีจริยธรรม
4. การปฏิบัติตามกฎหมายและข้อกำหนด
5. การเพิ่มมูลค่าข้อมูล

❑ ขอบข่าย

1. ประกาศคณะกรรมการพัฒนารัฐบาลดิจิทัล เรื่อง ธรรมาภิบาลข้อมูลภาครัฐ เวอร์ชัน 1.0
2. มรด. 6 : 2566 ว่าด้วยกรอบธรรมาภิบาลข้อมูลภาครัฐ ฉบับปรับปรุง: แนวปฏิบัติ
3. มสพร. 8-2565 ว่าด้วยหลักเกณฑ์การจัดระดับชั้นและการแบ่งปันข้อมูลภาครัฐ เวอร์ชัน 1.0
4. แนวปฏิบัติเกี่ยวกับการคุ้มครองข้อมูลส่วนบุคคล 3.0 (TDPG 3.0)
5. มาตรฐาน NISTIR 8053: De-identification of Personal Information
6. Guide to Basic Data Anonymization

❑ บทนิยาม

- **ข้อมูลนิรนาม (Anonymous Data) หมายความว่า** ข้อมูลที่ผ่านกระบวนการซึ่งทำให้ไม่สามารถระบุตัวตนหรือแสดงตัวตนได้ทั้งในปัจจุบันและอนาคต และไม่เป็นข้อมูลส่วนบุคคลตามกฎหมายว่าด้วยการคุ้มครองข้อมูลส่วนบุคคล
- **การจัดทำข้อมูลนิรนาม (Data Anonymization)** หมายความว่า กระบวนการทำให้ข้อมูลส่วนบุคคลไม่สามารถระบุหรือเชื่อมโยงข้อมูลไปถึงตัวบุคคลได้ทั้งในปัจจุบันและอนาคต โดยใช้เทคนิคหลายเทคนิคร่วมกันจนมั่นใจว่าไม่สามารถระบุตัวตนได้ตามกฎหมายว่าด้วยการคุ้มครองข้อมูลส่วนบุคคล
- **การจัดทำข้อมูลแฝง (Data Pseudonymization)** หมายความว่า กระบวนการเปลี่ยนแปลงข้อมูลส่วนบุคคลด้วยการใช้อักขระแฝงหรือวิธีการอื่นใด เช่น การเข้ารหัสข้อมูล (Encryption) การเข้าฟังก์ชันแฮช (Hashing) การเก็บข้อมูลแยกส่วนโดยเชื่อมผ่านโทเค็น (Tokenization) โดยยังสามารถเชื่อมโยงข้อมูลเพื่อระบุตัวตนได้เมื่อมีข้อมูลเพิ่มเติมประกอบและไม่ถือเป็นข้อมูลนิรนาม

คำนิยามอื่นๆ อ้างอิงจาก : พ.ร.บ. คุ้มครองข้อมูลส่วนบุคคล พ.ศ. 2562

- **ผู้ประมวลผลข้อมูลส่วนบุคคล (Data Processor/Personal Data Processor)**
- **ผู้ควบคุมข้อมูลส่วนบุคคล (Data Controller)**
- **ข้อมูลส่วนบุคคล (Personal Data)**
- **ข้อมูลส่วนบุคคลรั่วไหล (Personal Data Breach)**

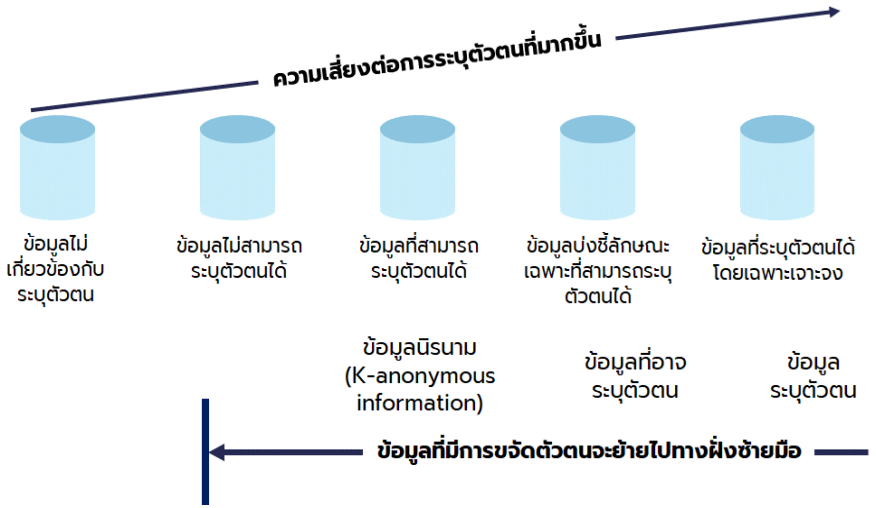
บทที่ 2 แนวคิดการจัดทำข้อมูลนิรนาม

DIGITAL GOVERNMENT DEVELOPMENT AGENCY

ข้อมูลส่วนบุคคล ถือเป็น 1 ใน 5 หมวดหมู่ข้อมูลการจัดทำธรรมาภิบาลข้อมูลภาครัฐ

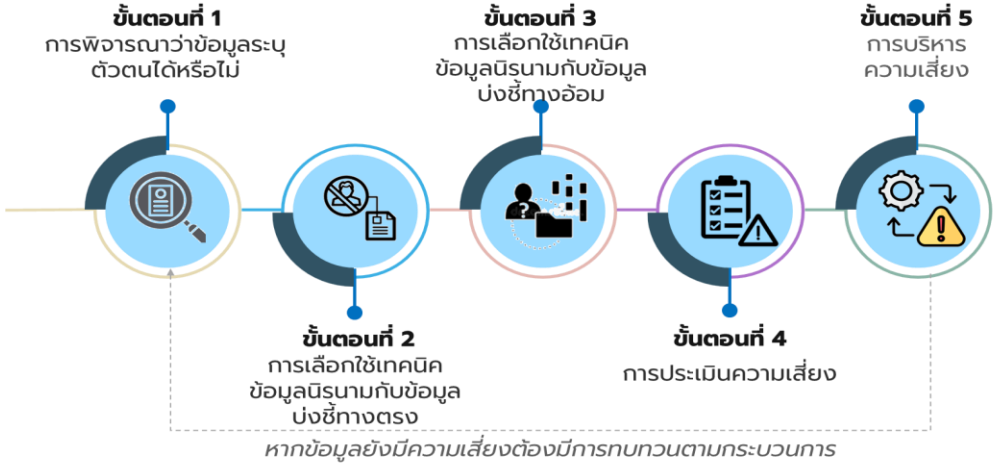
หน่วยงานจำเป็นต้องมี → การจัดระดับชั้นข้อมูล ซึ่งเป็นการบริหารจัดการข้อมูลก่อนการเปิดเผยหรือแบ่งปัน โดยพิจารณาจากความอ่อนไหวของข้อมูล ทั้งนี้ยังพบว่ายังคงมีความเสี่ยงในการรั่วไหลของข้อมูลดังนั้น การจัดทำข้อมูลนิรนามจึงเป็นเครื่องมือสำคัญที่ช่วยในการปกป้องคุ้มครองข้อมูลส่วนบุคคลและลดความเสี่ยงที่อาจก่อให้เกิดความเสียหายที่อาจเกิดขึ้นกับตัวบุคคลได้ ในขณะเดียวกันเป็นการสร้างความมั่นใจแก่หน่วยงานภาครัฐ ให้สามารถนำข้อมูลมาใช้สำหรับวิเคราะห์ วิจัย และพัฒนานวัตกรรม หรือเทคโนโลยีใหม่ ๆ โดยไม่เปิดเผยตัวตนของบุคคลผู้ที่เกี่ยวข้องกับข้อมูล

การลดความเสี่ยงด้วยการทำให้เป็นข้อมูลนิรนาม



การนำข้อมูลไปใช้ประโยชน์

ขั้นตอนการจัดทำข้อมูลนิรนาม

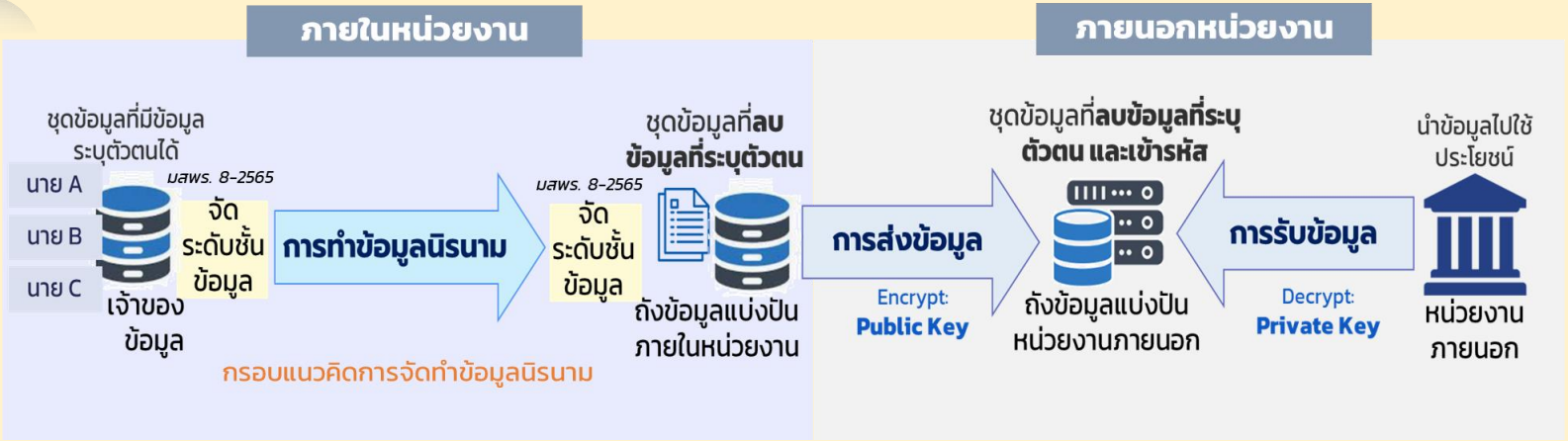


บทที่ 2 แนวคิดการจัดทำข้อมูลนิรนาม

วิธีการจัดทำข้อมูลนิรนามที่เป็นนิยมมี จำนวน 9 วิธี

01 + 02	**Attribute Suppression** การลบคุณลักษณะเฉพาะ (การลบข้อมูลรายคอลัมน์) + **Record Suppression** การลบข้อมูลรายบันทึก (การลบข้อมูลรายแถว)
03	**Character Masking** การปิดทับลักษณะข้อมูล
04	**Pseudonymization** การแฝงข้อมูล
05	**Generalization** การทำให้ข้อมูลเป็นสามัญ
05	**Swapping/Shuffling/ Permutation** การสลับข้อมูล
07	**Data Perturbation** การรบกวนข้อมูล
08	**Data Aggregation การรวมข้อมูล**
09	**Synthetic Data** การสังเคราะห์ข้อมูล

ตัวอย่างการจัดทำข้อมูลนิรนามเพื่อการแบ่งปันข้อมูล



ข้อมูลนิรนามควรพิจารณาเป็น รายคอลัมน์
โดยลบข้อมูลระบุตัวตนได้ เพื่อนำไปแบ่งปัน/ใช้ประโยชน์ต่อไป

ข้อมูลที่ผ่านการทำข้อมูลนิรนามควรพิจารณาจัดระดับชั้นข้อมูลอีกครั้ง (Declassification) โดยพิจารณาจากบริบท องค์กร ตาม มสพร. 8-2565 ว่าด้วย หลักเกณฑ์การจัดระดับชั้นและการแบ่งปันข้อมูลภาครัฐ

- การทำข้อมูลนิรนามควรพิจารณาตามวัตถุประสงค์การนำข้อมูลไปใช้ประโยชน์
- ชุดข้อมูลควรมีการเข้ารหัสทั้งฝ่ายส่งข้อมูลและรับข้อมูลเพื่อคุ้มครองข้อมูล

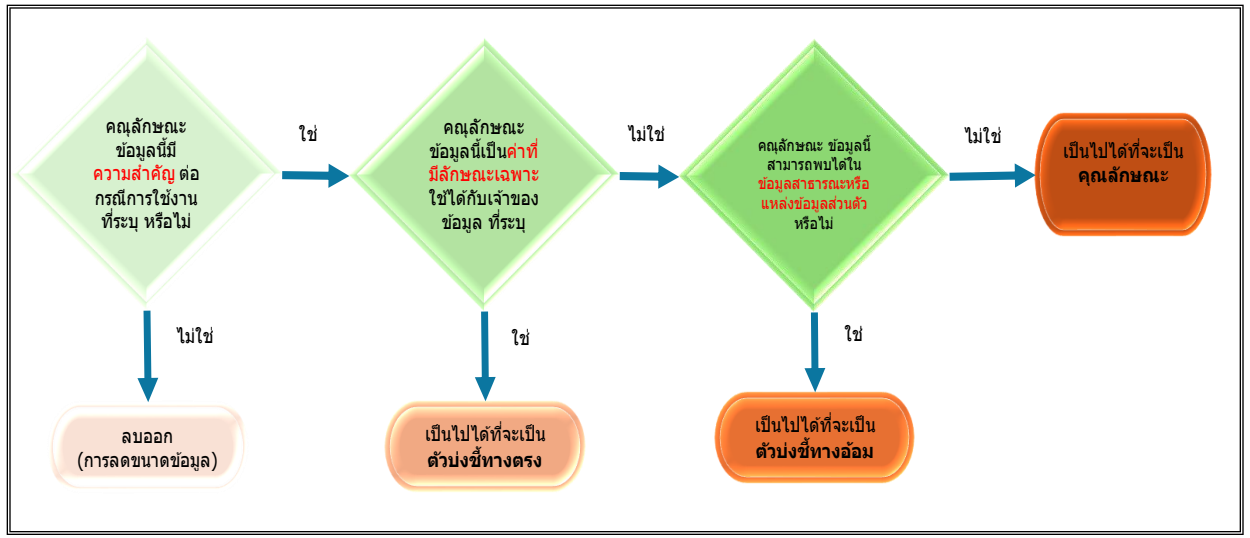
บทที่ 3 กระบวนการจัดทำข้อมูลนิรนาม

การพิจารณาตามคุณลักษณะข้อมูล และ การพิจารณาสถานการณ์ของข้อมูล

การพิจารณาความรับผิดทางกฎหมาย	การพิจารณาตัวข้อมูล	การพิจารณาการใช้งานของข้อมูล	การพิจารณาการขอใช้ข้อมูลโดยชอบ
ผู้ควบคุมข้อมูลส่วนบุคคล	ผู้ควบคุมข้อมูลส่วนบุคคล	ผู้ครอบครองข้อมูล/ผู้จัดทำข้อมูล	ผู้ควบคุมข้อมูลส่วนบุคคล / ผู้ประมวลผลข้อมูล
เป็นข้อมูลส่วนบุคคลหรือไม่	ใครเป็นเจ้าของข้อมูล?	ทำไม? (ทำไมถึงอยากที่จะเปิดเผยข้อมูล หรือเปิดเผยข้อมูลให้กับผู้อื่น หรือสาธารณะ)	ความโปร่งใสในการใช้ข้อมูล
มีหน้าที่เป็นผู้ควบคุม หรือผู้ประมวลผลข้อมูลหรือไม่?	ข้อมูลเป็นข้อมูลประเภทใด?	ใคร? ใครบ้างที่จะมีสิทธิเข้าถึงข้อมูล	การมีระบบธรรมาภิบาลในด้านข้อมูลที่ดี
	ตัวแปรในข้อมูล? (ทางตรง,ทางอ้อม)	อย่างไร? ผู้ที่จะเข้าถึงข้อมูลจะนำข้อมูลไปใช้อย่างไร (โดยต้องใช้การสอบถามโดยละเอียด)	
	คุณสมบัติของชุดข้อมูล? (คุณภาพของการวัด, อายุของข้อมูล, โครงสร้างของข้อมูล เป็นข้อมูลประชากร หรือ กลุ่มตัวอย่าง)		

แนวทางการลดขนาดของข้อมูล

เริ่มพิจารณาคุณลักษณะของข้อมูลที่ไม่จำเป็นในชุดข้อมูลผลลัพธ์**ควรถูกลบออก**



ตัวอย่าง

การปิดทับลักษณะข้อมูล (Character Masking)

ตัวอย่างข้อมูล**ก่อน**การทำ Character Masking

ชื่อ – นามสกุล	รหัสพนักงาน	เพศ	อายุงาน/ปี
แดง เป็นคนไทย	11000	ชาย	10
ไข่ ขยันทำงาน	110011	ชาย	2
สวย ใจดี	110012	หญิง	3
น้ำใจ รักงาน	110013	หญิง	2

ตัวอย่างข้อมูล**หลัง**การทำ Character Masking

ชื่อ – นามสกุล	รหัสพนักงาน	เพศ	อายุงาน/ปี
XX XXX	1100XX	ชาย	10
XX XXX	1100XX	ชาย	2
XX XXX	1100XX	หญิง	3
XX XXX	1100XX	หญิง	2

บทที่ 3 กระบวนการจัดทำข้อมูลนิรนาม

การเลือกวิธีการจัดทำข้อมูลนิรนามให้เหมาะสม

หลักเกณฑ์พิจารณา

ประเภทข้อมูล + การใช้ประโยชน์จากข้อมูล = วิธีการทำข้อมูลนิรนาม

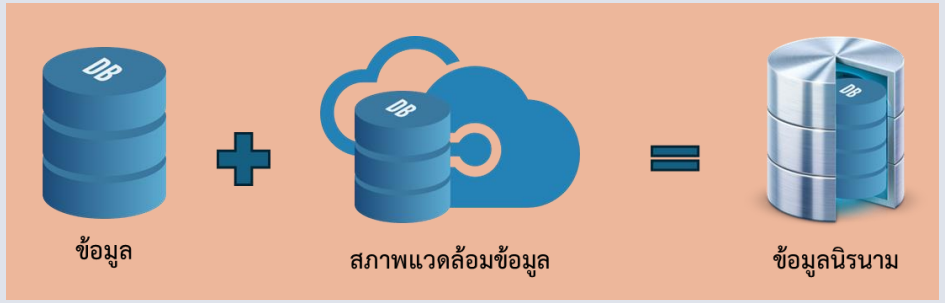
วิธีการจัดทำข้อมูลนิรนาม / ข้อเสนอแนะวิธีการจัดทำข้อมูลนิรนาม	ข้อเสนอแนะสำหรับการเลือกวิธีการจัดทำข้อมูลนิรนามให้เหมาะสมกับประเภทข้อมูล			ข้อเสนอแนะเพื่อการเลือกวิธีการจัดทำข้อมูลนิรนามให้เหมาะสมกับการใช้ประโยชน์		
	ข้อมูลบ่งชี้ทางตรง (Direct identifiers)	ข้อมูลบ่งชี้ทางอ้อม (Indirect identifiers)	ข้อมูลเชื่อมโยงไปข้อมูลบ่งชี้ได้ (Target attributes)	ข้อมูลนำมาวิเคราะห์ได้	ข้อมูลแปลงย้อนกลับได้	ข้อมูลสามารถเชื่อมโยงกับชุดข้อมูลอื่นได้
การลบคุณลักษณะข้อมูล (Suppression)	✓	✓	✓	!	!	!
การปิดทับลักษณะข้อมูล (Character Masking)	✓	✓	✓	!	!	!
การแฝงข้อมูล (Pseudonymization)	✓	✓	✓	!	✓	✓
การทำให้ข้อมูลเป็นสามัญ (Generalization)	!	✓	✓	✓	✓	!
การสลับข้อมูล (Swapping/Shuffling/Permutation)	!	✓	✓	✓	!	!
การรบกวนข้อมูล (Data Perturbation)	!	✓	✓	✓	!	!
การรวมข้อมูล (Data Aggregation)	✓	✓	✓	✓	!	!

*** การเลือกใช้วิธีการทำข้อมูลนิรนามรายคอลัมน์ ✓ เหมาะสม ! ไม่เหมาะสม

การพิจารณาสถานการณ์ของข้อมูล

คุณสมบัติ	ความเสี่ยงต่ำ	ความเสี่ยงสูง
คุณภาพของข้อมูล	ต่ำ	สูง
อายุของข้อมูล	เก่า	ใหม่
ระดับของข้อมูล	ข้อมูลรวมกลุ่ม	ข้อมูลรายบุคคลหรือรายหน่วยย่อย
โครงสร้างของข้อมูล	มีมิติเดียว	มีหลายมิติ
ความครบถ้วนข้อมูล	ข้อมูลตัวอย่าง	ข้อมูลประชากร
ข้อมูลที่มีความอ่อนไหว	น้อย	มาก
จำนวนตัวแปรหลัก	น้อย	มาก

การวิเคราะห์ความเสี่ยง และมาตรการจัดการความเสี่ยง



บทที่ 4 ภาคผนวก

ตัวอย่าง การจัดทำข้อมูลนิรนามเบื้องต้น โดยการใช้โปรแกรม Microsoft Excel

ตัวอย่าง รายการเครื่องมือการจัดทำข้อมูลนิรนามเชิงพาณิชย์หรือโอเพ่นซอร์ส (Open Source)

ตัวอย่าง กรณีศึกษาการจัดทำข้อมูลนิรนาม

BIG DATA INSTITUTE

TRAVEL LINK

คณะทำงานโครงการ Travel Link ภายใต้หน่วยงานสถาบันข้อมูลขนาดใหญ่ (องค์การมหาชน) ร่วมกับ หน่วยงาน พันธมิตร อาทิ สำนักงานปลัดกระทรวงการท่องเที่ยวและกีฬา การท่องเที่ยวแห่งประเทศไทย สำนักงานตรวจคนเข้าเมือง กรมการปกครอง ฯลฯ

ปัจจัยที่ส่งผลต่อการนิรนามข้อมูล

ความปลอดภัยของข้อมูล
ความเร็วในการประมวลผล
การนำข้อมูลไปใช้ต่อ

การนิรนามฟิลด์ข้อมูล

ประเภทฟิลด์ข้อมูล	รูปแบบการนำไปใช้งาน	วิธีการ
ข้อมูลส่วนบุคคล	ต้องการแยกแยะตัวตนแต่ละบุคคล	เข้ารหัสข้อมูลทางเดียวขั้นสูง
ข้อมูลส่วนบุคคล	เผยแพร่สู่สาธารณะ	ลบฟิลด์ข้อมูล
ข้อมูลที่เชื่อมโยงกับข้อมูลส่วนบุคคล	เผยแพร่สู่สาธารณะ	ลบฟิลด์ข้อมูล หรือ ลดความละเอียดข้อมูล
ข้อมูลที่ไม่ใช่ข้อมูลสาธารณะ	เผยแพร่สู่สาธารณะ	ผสมข้อมูล

***การนิรนามข้อมูลจึงถูกนำมาใช้ก่อนการเชื่อมโยงข้อมูลเพื่อสร้างความปลอดภัยสูงสุดในการเชื่อมโยงข้อมูล

อัลกอริทึม SHA-512

การเข้ารหัสข้อมูลทางเดียวขั้นสูง

สามารถเพิ่มองค์ประกอบต่าง ๆ ประกอบกับข้อมูลตั้งต้นได้ดังนี้
- ค่าประกอบการเข้ารหัส (salt) เป็นค่าสุ่มเพื่อทำให้การเดานั้นยากขึ้น โดยเป็นค่าที่ควรประกอบด้วยตัวอักษรพิมพ์เล็ก ตัวอักษรพิมพ์ใหญ่ ตัวเลข และเครื่องหมาย เช่น Q65e;=Cyx$hr8+?H
- วิธีการสลับหรือเรียงข้อมูล

>> การเข้ารหัสไฟล์เพื่อใช้ในการรับส่งข้อมูล → (Encryption) 2 รูปแบบ ได้แก่แบบสมมาตร (Symmetric Encryption) และไม่สมมาตร (Asymmetric Encryption)

1 คู่ แบ่งเป็น Public key และ Private key

ธนาคารแห่งประเทศไทย
BANK OF THAILAND

ข้อมูลของธนาคารแห่งประเทศไทยที่มีการจัดทำข้อมูลนิรนามเพื่อการใช้งาน

- ข้อมูลที่ ธปท. ได้รับจากหน่วยงานภายใต้การกำกับดูแล โดยอาศัยอำนาจตามกฎหมายหรืออาศัยข้อตกลงความร่วมมือระหว่างกันให้ผู้ประกอบธุรกิจจัดส่งข้อมูลดังกล่าวให้ ธปท.
- ข้อมูลที่ได้จากการสำรวจ
- ข้อมูลที่ได้จากหน่วยงานอื่นตามข้อตกลงความร่วมมือการแลกเปลี่ยน

รูปแบบการทำข้อมูลนิรนาม

- **การเข้าฟังก์ชันแฮช (Hashing)** ตามมาตรฐานสากล SHA256 โดยแปลงข้อมูลให้อยู่ในอักขระรูปแบบอื่นซึ่งจะไม่สามารถแปลงกลับเป็นข้อมูลเดิมได้ (One-way function) โดย ธปท. จะใช้ค่า salt ร่วมกับ key ในการเข้าฟังก์ชันแฮช เพื่อลดความเสี่ยงในการคาดเดา
- **การรวมข้อมูล (Data Aggregation)** โดยแสดงข้อมูลด้วยค่าผลรวมตัวเลขแยกตามมิติต่าง ๆ วิธีนี้มักจะใช้ในการแบ่งปันข้อมูลให้ฝ่ายงานที่ไม่จำเป็นต้องเห็นรายละเอียดข้อมูลในระดับรายบุคคล และใช้ในการแบ่งปันข้อมูลกลับให้แก่ผู้ที่จัดส่งข้อมูลให้ ธปท. หรือทำข้อมูลสถิติเผยแพร่ต่อสาธารณชนบนเว็บไซต์

*****การรวมข้อมูลจะมีการพิจารณาปัจจัย K-anonymity และ L-diversity1/ ด้วย เพื่อลดความเสี่ยงที่ผู้รับข้อมูลจะคาดเดาและระบุตัวตนบุคคลจากข้อมูลรวม หากไม่ผ่านค่าที่กำหนด ธปท. จะเพิ่มข้อมูลรบกวน (noise) หรือยุบรายละเอียดของมิติ (attribute) นั้นให้มีความละเอียดลดลง**

ธปท. มีข้อกำหนดให้ผู้ที่ได้รับอนุญาตให้ใช้ข้อมูล ผ่านการจัดทำเป็นข้อมูลนิรนามจะต้องไม่ดำเนินการหรือพยายามดำเนินการเพื่อคาดเดาหรือระบุตัวบุคคลในข้อมูลนั้น รวมถึงไม่ดำเนินการหรือพยายามดำเนินการเพื่อให้เกิดการจับคู่หรือเชื่อมโยงกับข้อมูลแวดล้อมอื่น

แนะนำเว็บไซต์
แหล่งข้อมูล เสริมพลัง สร้างความรู้

standard.dga.or.th

DIGITAL GOVERNMENT DEVELOPMENT AGENCY

DGA Thailand

DGA Thailand

DGA Thailand

contact@dga.or.th