งานประชาพิจารณ์

(ร่าง) มาตรฐานสำนักงานพัฒนารัฐบาลดิจิทัลว่าด้วยแนวทางการจัดทำข้อมูลนิรนาม เวอร์ชัน 1.0

โดย คุณสุภัทรา เรืองวานิช

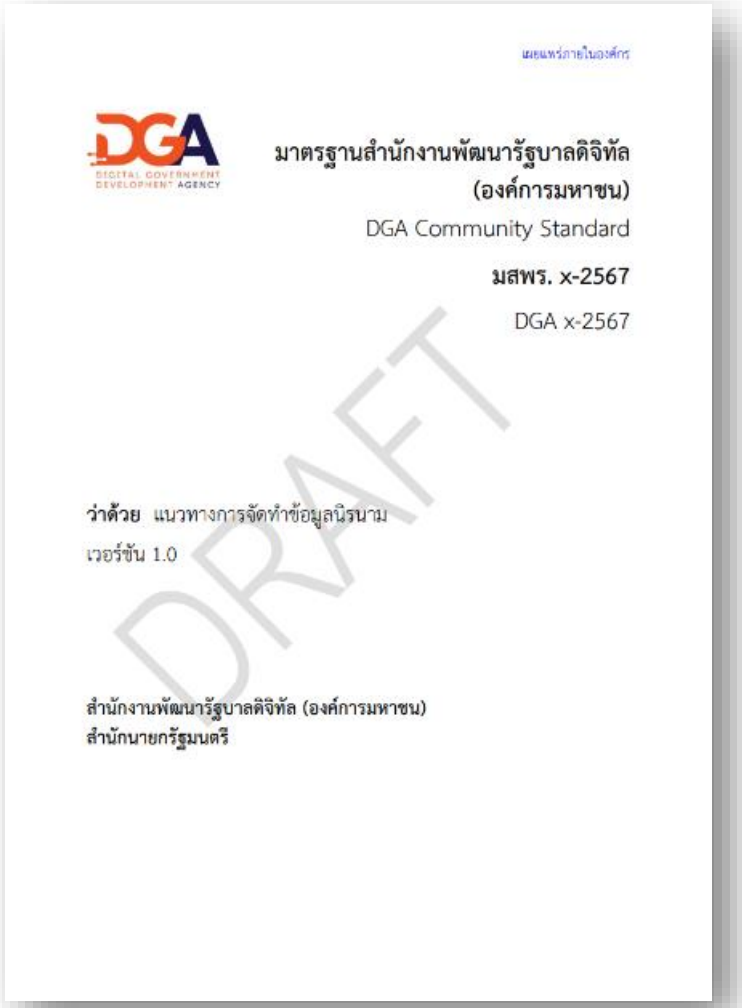
ผู้เชี่ยวชาญ ฝ่ายมาตรฐานดิจิทัลภาครัฐ

สำนักงานพัฒนารัฐบาลดิจิทัล (องค์การมหาชน)

วันที่ 28 มิถุนายน 2567

DIGITAL GOVERNMENT DEVELOPMENT AGENCY

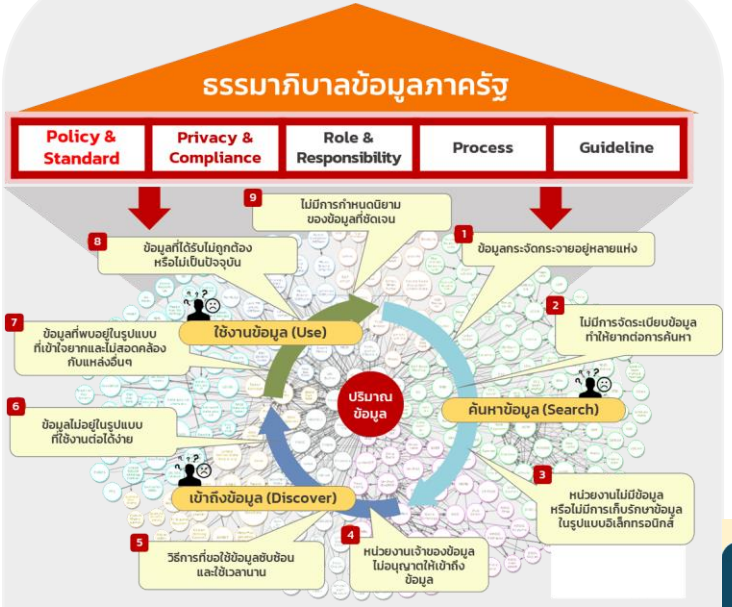
(ร่าง) มสพร. ว่าด้วยแนวทางการจัดทำข้อมูลนิรนาม



มสพร. คือ มาตรฐานสำนักงานพัฒนารัฐบาลดิจิทัล เพื่อยึดถือเป็นแนวทางปฏิบัติภายในของสำนักงานพัฒนารัฐบาลดิจิทัล (องค์การมหาชน)

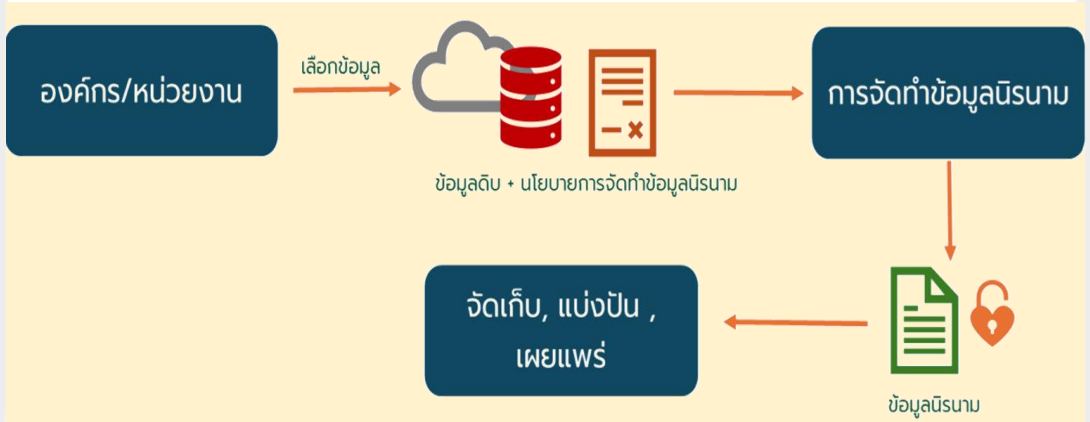
(ร่าง) มสพร. ว่าด้วยแนวทางการจัดทำข้อมูลนิรนาม

บทที่1 แนวคิด และ เทคนิคพื้นฐานในการไม่เปิดเผยข้อมูล >> 1.1 ความเป็นมา

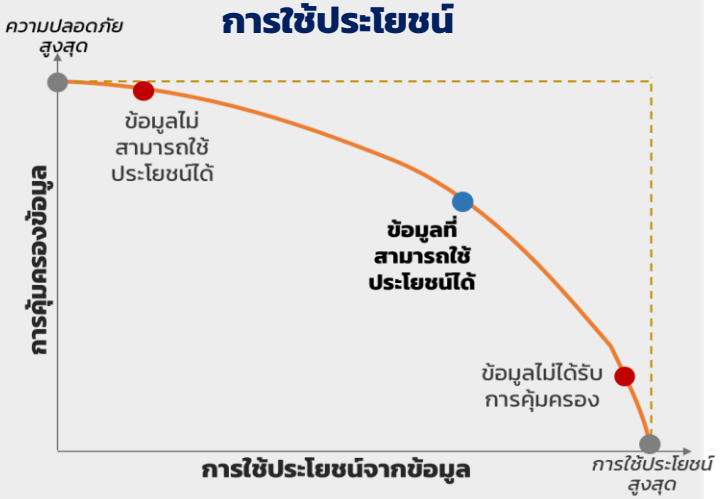


ธรรมาภิบาลข้อมูลจะช่วยให้ข้อมูลได้รับการป้องกันการละเมิดข้อมูล ***โดยการลดความสามารถในการระบุตัวตนลง*** *เพื่อให้สามารถนำข้อมูลไปใช้ประโยชน์ได้*

การจัดทำข้อมูลนิรนาม ไม่เพียงแต่เกี่ยวข้องกับการลบข้อมูลที่สามารถระบุตัวตนได้โดยตรงเท่านั้น แต่ยังรวมถึงหลักคิดในการวิเคราะห์ว่าข้อมูลที่เหลืออยู่สามารถถูกใช้ในการระบุตัวตนได้หรือไม่ หากมีความเป็นไปได้จำเป็นต้องมีการปรับเปลี่ยนหรือเพิ่มเทคนิคการป้องกันเพื่อลดความเสี่ยงนี้



การสร้างสมดุลระหว่างการคุ้มครองข้อมูลและการใช้ประโยชน์



ข้อมูลที่มีการคุ้มครองที่เคร่งครัดย่อมนำไปใช้ประโยชน์ได้น้อย ในขณะที่ข้อมูลที่ใช้ประโยชน์ได้เต็มที่ย่อมเสี่ยงต่อการถูกละเมิด ดังนั้นจึงควรมีการหาสมดุลระหว่างการคุ้มครองข้อมูลและการใช้ประโยชน์

DGA DIGITAL GOVERNMENT DEVELOPMENT AGENCY

(ร่าง) มสพร. ว่าด้วยแนวทางการจัดทำข้อมูลนิรนาม

บทที่1 แนวคิด และ เทคนิคพื้นฐานในการไม่เปิดเผยข้อมูล

1.2. วัตถุประสงค์

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

1.3. 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

1.4. 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)**

(ร่าง) มสพร. ว่าด้วยแนวทางการจัดทำข้อมูลนิรนาม

บทที่ 2 แนวคิดการจัดทำข้อมูลนิรนาม >> 2.1. ความสำคัญของการจัดทำข้อมูลนิรนาม

ข้อมูลส่วนบุคคลถือเป็น 1 ใน 5 หมวดหมู่ของข้อมูล การจัดทำธรรมาภิบาลข้อมูลภาครัฐ

ตาม มรด. 6 : 2566 ว่าด้วยกรอบธรรมาภิบาลข้อมูลภาครัฐ ฉบับปรับปรุง: แนวปฏิบัติ

Data Class. Level / Data Category	เปิดเผย (Open)	เผยแพร่ภายในองค์กร (Private)	ลับ (Confidential / sensitive)	ลับมาก (Secret / Medium Sensitive)	ลับที่สุด (Top secret / Highly Sensitive)
ข้อมูลสาธารณะ	▪ พ.ร.บ. ข้อมูลข่าวสารของทางราชการ 2540 (มาตรา 7 และมาตรา 9) ▪ มาตรฐานและหลักเกณฑ์การเปิดเผยข้อมูลภาครัฐ				
ข้อมูลใช้ภายใน		ISO 27001: 2013			
ข้อมูลส่วนบุคคล			พ.ร.บ. คุ้มครองข้อมูลส่วนบุคคล 2562 (มาตรา 24 – มาตรา 27)		
ข้อมูลข่าวสารลับ			พ.ร.บ. ข้อมูลข่าวสารของทางราชการ 2540 (มาตรา 9 และมาตรา 15 ที่เปิดเผยได้) ระเบียบว่าด้วยการรักษาความลับของทางราชการ 2544		▪ พ.ร.บ. ข้อมูลข่าวสารของทางราชการ 2540 (มาตรา 14 – มาตรา 15 อาจมีคำสั่งมิให้เปิดเผย)
ข้อมูลความมั่นคง			ระเบียบสำนักนายกรัฐมนตรีว่าด้วยการรักษาความปลอดภัยแห่งชาติ พ.ศ. 2552 นโยบายและแผนระดับชาติ ว่าด้วยความมั่นคงแห่งชาติ (พ.ศ. 2562-2565)		

หน่วยงานจำเป็นต้องมี → การจัดระดับชั้นข้อมูล
ซึ่งเป็นการบริหารจัดการข้อมูลก่อนการเปิดเผยหรือแบ่งปัน
โดยพิจารณาจากความอ่อนไหวของข้อมูล
ตาม มสพร. 8-2565 ว่าด้วยหลักเกณฑ์การจัดระดับชั้นและการแบ่งปันข้อมูลภาครัฐ

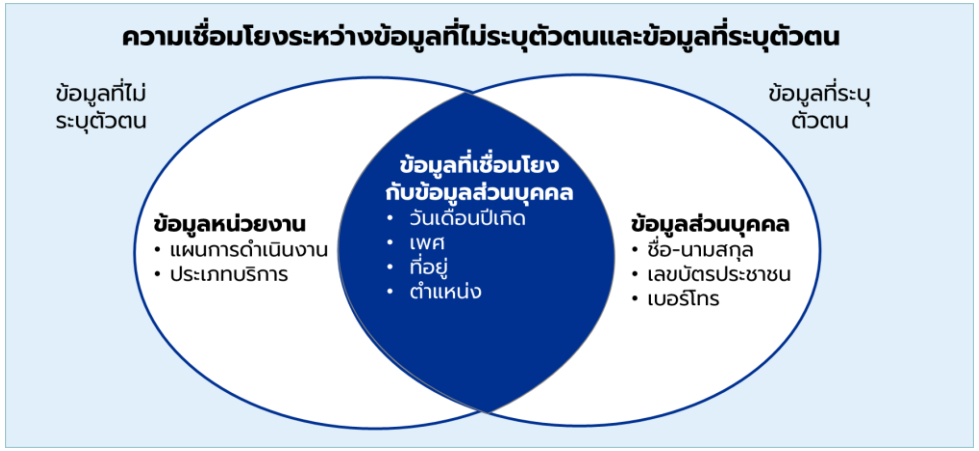
สิ่งที่เกิดขึ้น !!! หน่วยงานภาครัฐหรือหน่วยงานเอกชนยังคงมีความเสี่ยงในการถูกโจมตี ก่อให้เกิดการรั่วไหลของข้อมูลได้ **ดังนั้น การจัดทำข้อมูลนิรนามจึงเป็นเครื่องมือสำคัญที่ช่วยในการปกป้องคุ้มครองข้อมูลส่วนบุคคล และลดความเสี่ยงที่อาจก่อให้เกิดความเสียหายที่อาจเกิดขึ้นกับตัวบุคคลได้** ในขณะเดียวกันเป็นการสร้างความมั่นใจแก่หน่วยงานภาครัฐ ให้สามารถนำข้อมูลมาใช้สำหรับวิเคราะห์ วิจัย และพัฒนานวัตกรรม หรือเทคโนโลยีใหม่ ๆ โดยไม่เปิดเผยตัวตนของบุคคลผู้ที่เกี่ยวข้องกับข้อมูล

DIGITAL GOVERNMENT DEVELOPMENT AGENCY

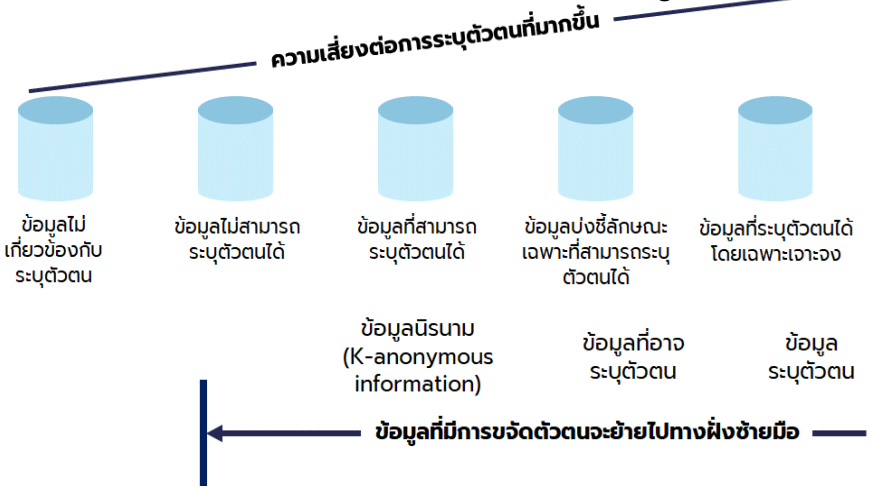
(ร่าง) มสพร. ว่าด้วยแนวทางการจัดทำข้อมูลนิรนาม

บทที่ 2 แนวคิดการจัดทำข้อมูลนิรนาม >> 2.2 แนวคิดการจัดทำข้อมูลนิรนาม

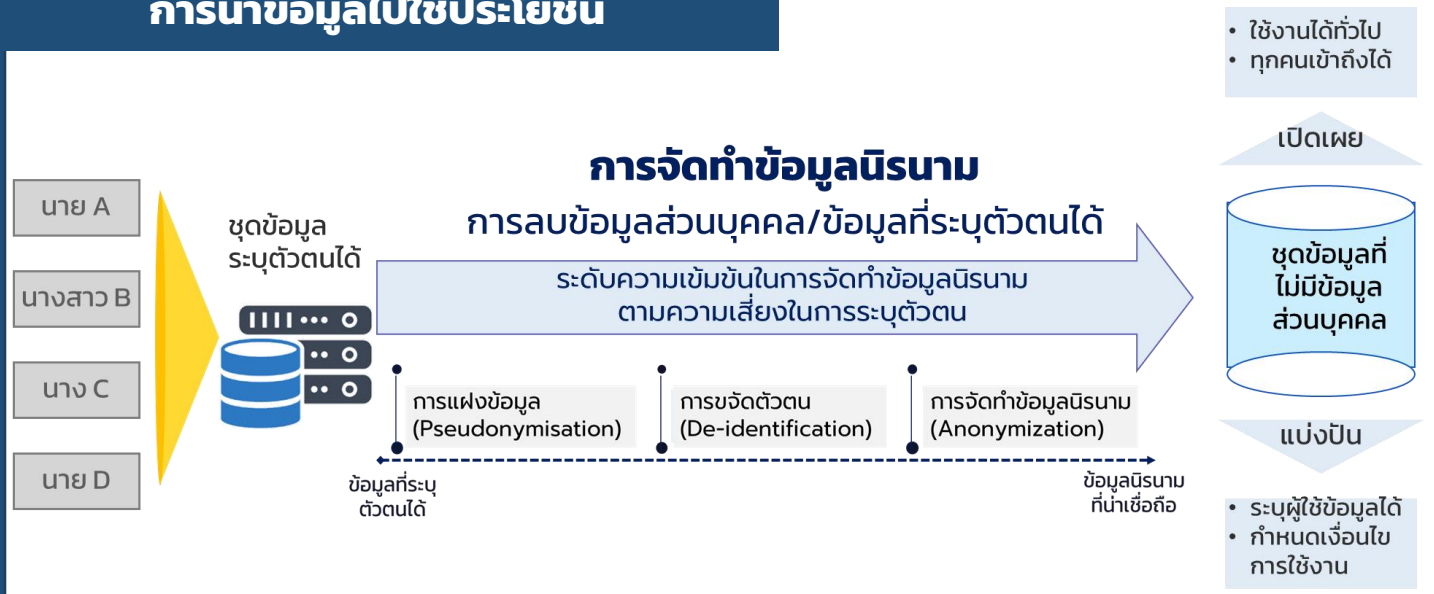
ความเสี่ยงที่จะนำไปสู่การระบุตัวตนได้ : 1. การแปลกแยกออกจากกลุ่ม 2.ความสามารถเชื่อมโยง และ 3.การอนุมาน



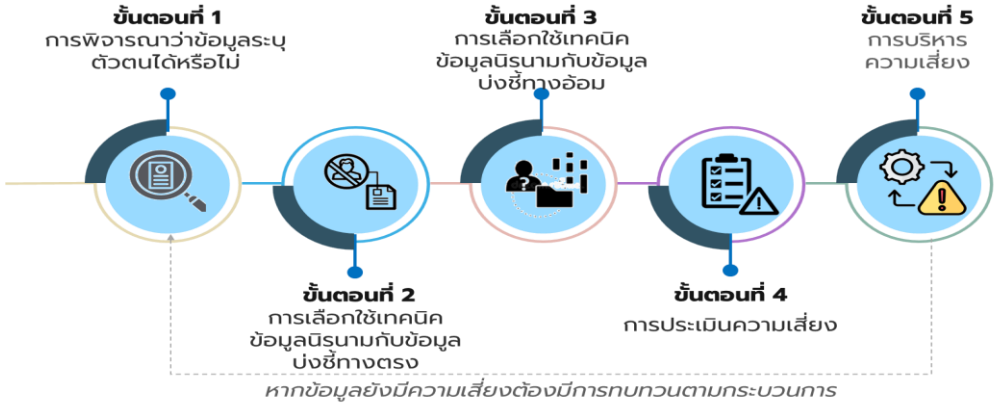
การลดความเสี่ยงด้วยการทำให้เป็นข้อมูลนิรนาม



การนำข้อมูลไปใช้ประโยชน์



ขั้นตอนการจัดทำข้อมูลนิรนาม



Guide to Basic Data Anonymization

(ร่าง) มสพร. ว่าด้วยแนวทางการจัดทำข้อมูลนิรนาม

บทที่ 2 แนวคิดการจัดทำข้อมูลนิรนาม >> 2.3. วิธีการจัดทำข้อมูลนิรนาม

จากการศึกษา : วิธีการจัดทำข้อมูลนิรนามที่เป็นนิยมมี **จำนวน 9 วิธี**

>> โดยในบทที่ 3.2 หลักเกณฑ์การจัดทำข้อมูลนิรนาม ได้คัดเลือกมา 7 วิธี (วิธีที่ 1+2 มีการนำมารวมกัน)

01 **Attribute Suppression**
การลบคุณลักษณะเฉพาะ
(การลบข้อมูลรายคอลัมน์)

+

02 **Record Suppression**
การลบข้อมูลรายบันทึก
(การลบข้อมูลรายแถว)

03 **Character Masking**
การปิดทับลักษณะข้อมูล

04 **Pseudonymization**
การแฝงข้อมูล

05 **Generalization**
การทำให้ข้อมูลเป็นสามัญ

06 **Swapping/Shuffling/ Permutation**
การสลับข้อมูล

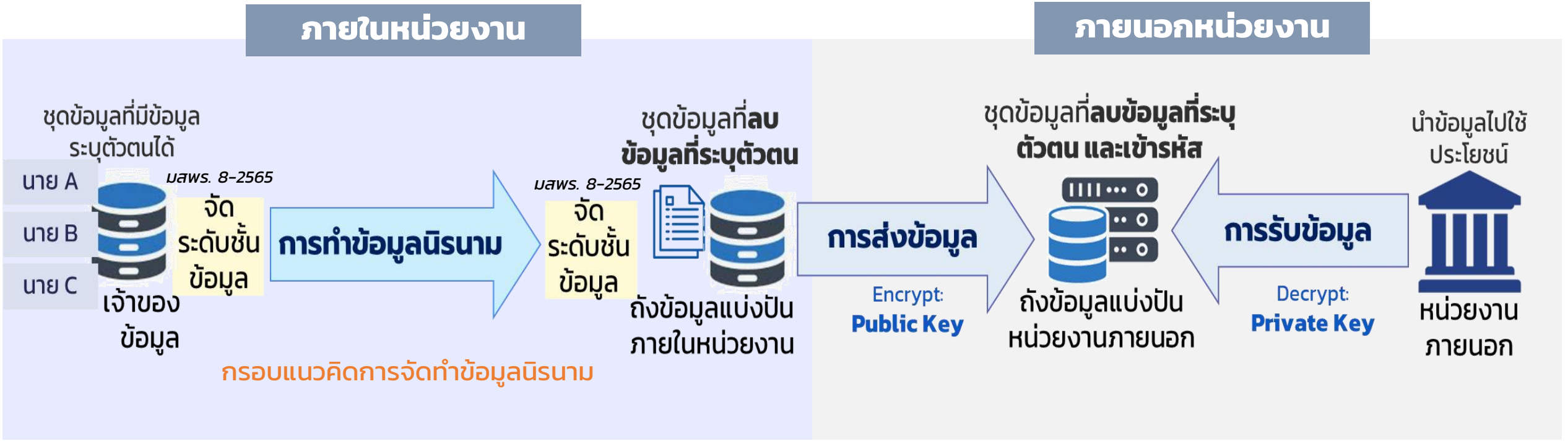
07 **Data Perturbation**
การรบกวนข้อมูล

08 **Data Aggregation**
การรวมข้อมูล

09 **Synthetic Data**
การสังเคราะห์ข้อมูล

DGA DIGITAL GOVERNMENT DEVELOPMENT AGENCY

ตัวอย่างการจัดทำข้อมูลนิรนามเพื่อการแบ่งปันข้อมูล



ข้อมูลนิรนามควรพิจารณาเป็น **รายคอลัมน์**
โดยลบข้อมูลระบุตัวตนได้ เพื่อนำไปแบ่งปัน/ใช้ประโยชน์ต่อไป

- ข้อมูลที่ผ่านการทำข้อมูลนิรนามควรพิจารณาจัดระดับชั้นข้อมูลอีกครั้ง (Declassification) โดยพิจารณาจากบริบท องค์กร ตาม มสพร. 8-2565 ว่าด้วยหลักเกณฑ์การจัดระดับชั้นและการแบ่งปันข้อมูลภาครัฐ
- การทำข้อมูลนิรนามควรพิจารณาตามวัตถุประสงค์การนำข้อมูลไปใช้ประโยชน์
- ชุดข้อมูลควรมีการเข้ารหัสทั้งฝ่ายส่งข้อมูลและรับข้อมูลเพื่อคุ้มครองข้อมูล

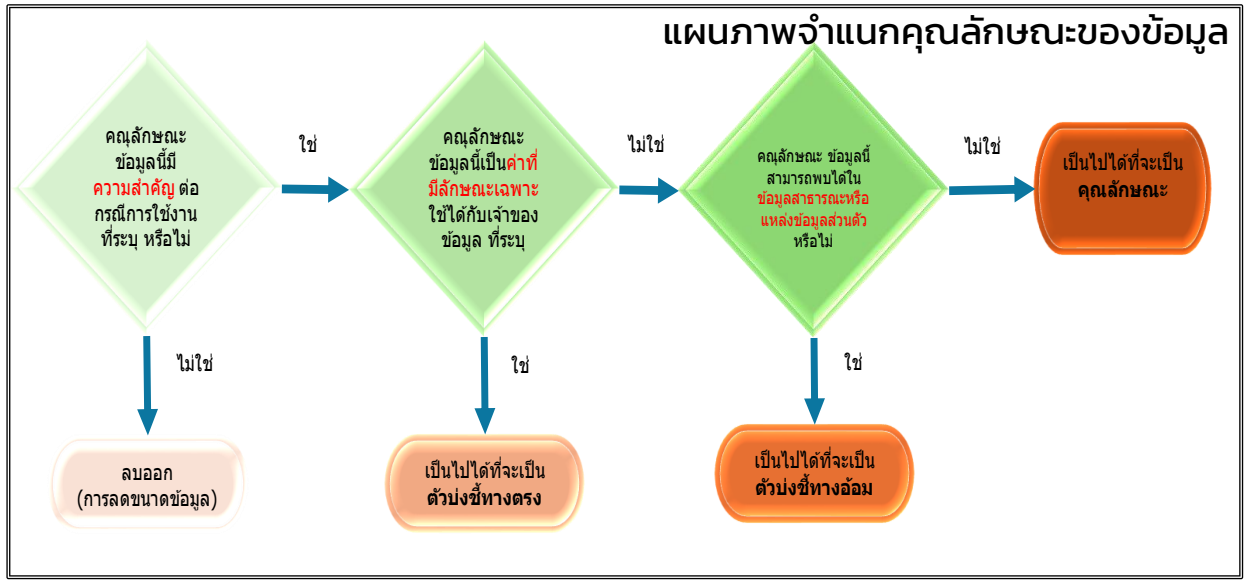
บทที่ 3 กระบวนการจัดทำข้อมูลนิรนาม >> 3.1. การพิจารณาข้อมูล และ การขจัดข้อมูล

1. การพิจารณาตามคุณลักษณะข้อมูล

ระดับการระบุตัวตนของชุดข้อมูล	คุณลักษณะของข้อมูล	การเข้าถึงข้อมูล	ตัวอย่างในชุดข้อมูล
1.ตัวบ่งชี้ทางตรง (Direct identifiers)	เป็นคุณลักษณะข้อมูลที่เฉพาะเจาะจงแต่ละบุคคลและสามารถใช้เป็นคุณลักษณะข้อมูลหลักในการระบุตัวตนของบุคคลนั้นได้อีกครั้ง	คุณลักษณะข้อมูลเหล่านี้มักจะเป็นข้อมูลสาธารณะหรือข้อมูลที่เข้าถึงได้ง่าย	ชื่อ นามสกุล , ที่อยู่อีเมล ,หมายเลขโทรศัพท์มือถือ , หมายเลขหนังสือเดินทาง หมายเลขบัญชี , หมายเลขสูติบัตร , หมายเลขใบอนุญาตทำงาน ,ชื่อผู้ใช้งานโซเชียลมีเดีย
2. ตัวบ่งชี้ทางอ้อม (Indirect identifiers)	เป็นคุณลักษณะข้อมูลที่ไม่เฉพาะเจาะจงแต่ละบุคคล แต่อาจทำให้สามารถระบุตัวตนของบุคคลนั้นได้เมื่อรวมกับข้อมูลอื่น (เช่น การรวมกันของอายุ, เพศ และรหัสไปรษณีย์)	คุณลักษณะข้อมูลเหล่านี้มักจะเป็นข้อมูลสาธารณะหรือข้อมูลที่เข้าถึงได้ง่าย	อายุ , เพศ , เชื้อชาติ , วันเดือนปีเกิด ,ที่อยู่ รหัสไปรษณีย์ , ตำแหน่งงาน ,ชื่อบริษัท , สถานภาพการสมรส , ส่วนสูง , น้ำหนัก ที่อยู่อินเทอร์เน็ตโปรโตคอล (IP Address) ,หมายเลขทะเบียนรถ , ตำแหน่งพิกัดบนพื้นโลก (GPS)
3. คุณลักษณะเป้าหมาย (Target attributes)	คุณลักษณะของข้อมูลนี้อาจจะลักษณะละเอียดอ่อน และอาจส่งผลให้เกิดผลเสียต่อบุคคลได้สูงเมื่อถูกนำไปเปิดเผย	คุณลักษณะข้อมูลเหล่านี้มักจะไม่เป็นข้อมูลสาธารณะ หรือไม่สามารถเข้าถึงได้ ซึ่งข้อมูล เหล่านี้ไม่สามารถใช้เพื่อระบุตัวตนของบุคคลนั้นอีกครั้งได้ เนื่องจากโดยทั่วไปแล้วจะเป็นข้อมูลที่มีกรรมสิทธิ์	ธุรกรรม (เช่น การซื้อของ) , เงินเดือน , อัตราเครดิต , กรมธรรม์ประกัน , การวินิจฉัยทางการแพทย์ , สถานะการฉีดวัคซีน

แนวทางการลดขนาดของข้อมูล โดยเริ่มจากการพิจารณาคุณลักษณะของข้อมูลที่ไม่จำเป็นในชุดข้อมูลผลลัพธ์ควรถูกลบออก

แผนภาพจำแนกคุณลักษณะของข้อมูล



2. การพิจารณาสถานการณ์ของข้อมูล

ผู้จัดทำข้อมูลนิรนาม จะต้องจัดทำผังการเคลื่อนที่ข้อมูล (Data Flowchart) โดยระบุถึงสิ่งแวดล้อมทั้งหมดที่ข้อมูลอาจมีการเคลื่อนย้ายโดยอาจระบุถึง

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)		

***การพิจารณาตามคุณลักษณะข้อมูล ร่วมกับ การพิจารณาสถานการณ์ของข้อมูล เป็นเกณฑ์สำคัญ ในการจัดทำข้อมูลให้เหมาะสม ต่อการใช้งานภายใน หรือ การใช้ภายนอกหน่วยงาน

(ร่าง) มสพร. ว่าด้วยแนวทางการจัดทำข้อมูลนิรนาม

บทที่ 3 กระบวนการจัดทำข้อมูลนิรนาม >> 3.2. หลักเกณฑ์การจัดทำข้อมูลนิรนาม

ตัวอย่าง: การปิดทับลักษณะข้อมูล (Character Masking)

หลักเกณฑ์การพิจารณา	• เป็นข้อมูลบ่งชี้ทางตรง หรือข้อมูลที่มีความอ่อนไหว ใช่หรือไม่ • เป็นข้อมูลที่สามารถแทนที่ด้วยตัวอักษรหรือตัวเลขหรือไม่ • เป็นข้อมูลที่ต้องการเชื่อมโยงกับข้อมูลอื่น ๆ ในชุดข้อมูลเดียวกัน ใช่หรือไม่ • เป็นข้อมูลที่ปิดทับไปแล้วไม่ส่งผลต่อการใช้ประโยชน์ข้อมูลอื่น ๆ ในชุดข้อมูล ใช่หรือไม่
ข้อแนะนำ	• ควรเลือกใช้กับข้อมูลบ่งชี้ทางตรง • สามารถปิดทับได้มากกว่า 1 คอลัมน์ในชุดข้อมูล และสามารถใช้การจัดทำข้อมูลนิรนามวิธีอื่น ๆ ประกอบกัน เพื่อป้องกันการเชื่อมโยงหรืออนุมานกลับการระบุตัวตน • ข้อมูลในชุดข้อมูลควรมีปริมาณมากพอเพื่อป้องกันไม่ให้เกิดการเชื่อมโยงกลับไปยังเจ้าของข้อมูลส่วนบุคคลได้ • สามารถประยุกต์ใช้วิธีนี้กับข้อมูลประเภท 1) ข้อมูลใช้ภายใน 2) ข้อมูลที่มีการแบ่งปันให้กับบุคคลภายนอกองค์กร ซึ่งเจ้าของข้อมูล/ผู้ควบคุมข้อมูลสามารถพิจารณาได้ตามความเหมาะสม

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

(ร่าง) มสพร. ว่าด้วยแนวทางการจัดทำข้อมูลนิรนาม

บทที่ 3 กระบวนการจัดทำข้อมูลนิรนาม >> 3.2 หลักเกณฑ์การจัดทำข้อมูลนิรนาม

การเลือกวิธีการจัดทำข้อมูลนิรนามให้เหมาะสม

หลักเกณฑ์พิจารณา

ประเภทข้อมูล + **การใช้ประโยชน์จากข้อมูล** = **วิธีการทำข้อมูลนิรนาม**

*** การเลือกใช้วิธีการทำข้อมูลนิรนามรายคอลัมน์

✓ เหมาะสม ⚠ ไม่เหมาะสม

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)	✓	✓	✓	✓	⚠	⚠

ในการประยุกต์ใช้วิธีการจัดทำข้อมูลนิรนามทั้ง 7 วิธี

>> ในชุดข้อมูลแต่ละชุดข้อมูล ควรใช้วิธีการจัดทำข้อมูลนิรนามมากกว่า 1 วิธี ซึ่งจะพิจารณาได้จากประเภทข้อมูลและการนำข้อมูลไปใช้ประโยชน์

DGA DIGITAL GOVERNMENT DEVELOPMENT AGENCY

(ร่าง) มสพร. ว่าด้วยแนวทางการจัดทำข้อมูลนิรนาม

บทที่ 3 กระบวนการจัดทำข้อมูลนิรนาม >> 3.3. การวิเคราะห์ความเสี่ยงในการเปิดเผยข้อมูล

1. การพิจารณาสถานการณ์ของข้อมูล

แนวทางการพิจารณาถึงคุณสมบัติหลัก ๆ ที่เกี่ยวข้องกับข้อมูล ดังนี้

- ใครเป็นเจ้าของข้อมูล
- ข้อมูลเป็นข้อมูลประเภทใด
- ประเภทของตัวแปรของข้อมูล
- คุณสมบัติของชุดข้อมูล

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

2. การวิเคราะห์ความเสี่ยง และมาตรการจัดการความเสี่ยง

- พิจารณาภาพรวมของข้อมูล
- การกำหนดมาตรการในการควบคุมความเสี่ยงที่สอดคล้องกับสถานการณ์ของข้อมูล ซึ่งทำได้ 2 วิธี

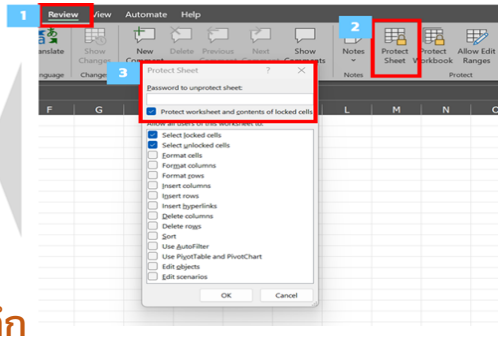
- การเปลี่ยนข้อมูล ต้องคำนึงถึง 2 ปัจจัย คือ
  - ความง่ายต่อการเปิดเผยข้อมูลส่วนบุคคล
  - ความอ่อนไหวของข้อมูลส่วนบุคคล

- การปรับสิ่งแวดล้อม คือการควบคุมการเข้าถึงข้อมูล มี 4 วิธี ซึ่งเรียงลำดับในการควบคุมการเข้าถึงและใช้ข้อมูลตามความจำเป็นจากน้อยไปมาก
  - การเปิดให้ใช้ข้อมูลโดยทั่วไป (open access)
  - การจัดส่งข้อมูลให้เป็นรายกรณี (delivered access)
  - การใช้ข้อมูล ณ สถานที่ที่จัดเตรียมไว้ (on-site safe settings)
  - การใช้ใบอนุญาต (Licenses)

ตัวอย่าง การจัดทำข้อมูลนิรนามเบื้องต้น โดยการใช้โปรแกรม Microsoft Excel

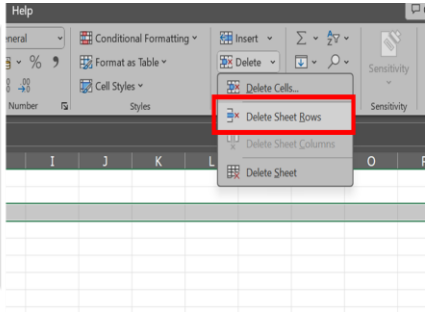
- กรณีการซ่อนคอลัมน์

	A	B	C	D
1	ชื่อ - นามกุล	ตำแหน่ง	เพศ	อายุงาน/ปี
2	แดง เป็นคนไทย	ผู้บริหาร	ชาย	10
3	ไข่ ขยันทำงาน	พนักงาน 1	ชาย	2
4	สวย ใจดี	พนักงาน 2	หญิง	3
5	น้ำใจ รักงาน	พนักงาน 3	หญิง	2



- กรณีต้องการลบข้อมูลรายบันทึก (Record Suppression)

	A	B	C	D
1	ชื่อ - นามกุล	ตำแหน่ง	เพศ	อายุงาน/ปี
2	แดง เป็นคนไทย	ผู้บริหาร	ชาย	10
3	ไข่ ขยันทำงาน	พนักงาน 1	ชาย	2
4	สวย ใจดี	พนักงาน 2	หญิง	3
5	น้ำใจ รักงาน	พนักงาน 3	หญิง	2



- กรณีต้องการการปิดทับลักษณะข้อมูลเบื้องต้น

การใช้ฟังชั่น	ข้อมูลก่อนการ Masking	ข้อมูลหลังการ Masking
=("**** ") & RIGHT (C2, 2))	11001	**** 01

คอลัมน์ที่ต้องการแสดง
จำนวนอักขระที่ต้องการปิดทับ ซึ่งควรมีจำนวนเท่ากันกับข้อมูลเดิม
จำนวนอักขระตัวท้ายที่ต้องการแสดง

ตัวอย่างการใช้ฟังก์ชัน RIGHT

	A	B	C	D	
1	เพศ	อายุงาน /ปี	รหัสพนักงาน ก่อนการ Masking	รหัสพนักงาน หลังการ Masking	ตัวอย่างการใช้ฟังชั่น
2	ชาย	10	515-43-51101	****-**-51101	=("****-**-"&RIGHT(C2,5))
3	ชาย	2	515-43-51102	@@@-@@-51102	=("@@@-@@-"&RIGHT(C3,5))
4	หญิง	3	515-43-51103	XXX-XX-51103	=("XXX-XX-"&RIGHT(C4,5))

การใช้ฟังชั่น	ข้อมูลก่อนการ Masking	ข้อมูลหลังการ Masking
=(REPT("*", LEN (C2)))	11001	*****

ฟังชั่น
อักขระที่ต้องการแสดงเมื่อมีการปิดทับ
คอลัมน์ที่ต้องการปิดทับเนื้อหา

ตัวอย่างการใช้ฟังก์ชัน REPT

	B	C	D	
	อายุงาน /ปี	รหัสพนักงาน ก่อนการ Masking	รหัสพนักงาน หลังการ Masking	ตัวอย่างการใช้ฟังชั่น
	10	515-43-51101	**********	=(REPT("*",LEN(C2)))
	2	51102	WWWW	=(REPT("W",LEN(C3)))
4 หญิง	3	001	@@@	=(REPT(@",LEN(C2)))

ตัวอย่าง รายการเครื่องมือการจัดทำข้อมูลนิรนามเชิงพาณิชย์หรือโอเพ่นซอร์ส (Open Source)

รายการ	คำอธิบาย	ที่มา
PDPC	เป็นเครื่องมือเพื่อตามกระบวนการจัดทำข้อมูลนิรนามในการแปลงให้เป็นข้อมูลนิรนาม ตั้งแต่การเริ่ม การพิจารณาว่าข้อมูล การเลือกใช้เทคนิคข้อมูลนิรนามกับข้อมูลบ่งชี้ทางตรง ไปจนถึงการประเมินความเสี่ยง	https://www.pdpc.gov.sg/help-and-resources/2018/01/basic-anonymisation
Amnesia	เป็นเครื่องมือในการแปลงให้เป็นข้อมูลนิรนาม โดยลบข้อมูลบ่งชี้ทางตรงและข้อมูลอ่อนไหว โดยมีการประเมินโดยใช้ k-anonymity และ km-anonymity	https://amnesia.openaire.eu/
ARGUS	เครื่องมือนี้ใช้วิธีการลบข้อมูลระบุตัวบุคคลทางสถิติ โดยการจัดทำข้อมูลนิรนามด้วยวิธี 1) การทำข้อมูลให้เป็นสามัญ 2) การรบกวนข้อมูล และ 3) การรวมข้อมูล	https://research.cbs.nl/casc/mu.htm
ARX	เป็นแบบจำลอง (Model) เพื่อแสดงการจัดทำข้อมูลนิรนามในหลากหลายมิติ เช่น การเลือกใช้วิธีการจัดทำข้อมูลนิรนาม การวิเคราะห์การใช้ประโยชน์ของข้อมูล และ การวิเคราะห์ความเสี่ยงในการระบุตัวตัวตน	https://arx.deidentifier.org/

- อ้างอิงจาก The Personal Data Protection Commission , **Guide to Basic Data Anonymization (31 March 2022)** และแนวทางสำหรับการจัดทำข้อมูลนิรนามขั้นพื้นฐาน โดยสำนักงานคณะกรรมการคุ้มครองข้อมูลส่วนบุคคล (สคส.) หรือ PDPC ร่วมกับ สวทช.

บทที่ 4. ภาคผนวก (เครื่องมือ Open Source) >> 4.2. - 4.3 กรณีศึกษาการจัดทำข้อมูลนิรนามของ BDI และ BOT

อัลกอริทึม SHA-512

BIG DATA INSTITUTE

TRAVEL LINK

คณะทำงานโครงการ Travel Link ภายใต้หน่วยงานสถาบันข้อมูลขนาดใหญ่ (องค์การมหาชน) ร่วมกับ หน่วยงานพันธมิตร 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

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 ke

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

DGA
DIGITAL GOVERNMENT DEVELOPMENT AGENCY

ขอเชิญร่วมแสดงความเห็นต่อ (ร่าง) มาตรฐานสำนักงานพัฒนารัฐบาล ดิจิทัลว่าด้วยแนวทางการจัดทำข้อมูลนิรนาม เวอร์ชัน 1.0

Data Anonymization

ขอเชิญเข้าร่วมงานรับฟังความคิดเห็นต่อ (ร่าง) มสพร. X-256X มาตรฐานสำนักงานพัฒนารัฐบาลดิจิทัล ว่าด้วยแนวทางการจัดทำข้อมูลนิรนาม เวอร์ชัน 1.0

ร่วมประชุมการรับฟังความคิดเห็น
28 มิ.ย. 67 | 13.30 - 16.00 น.
ผ่านทาง Microsoft Teams

พร้อมเสวนาในหัวข้อ:

ขับเคลื่อนประเทศด้วย "การจัดทำข้อมูลที่มีประสิทธิภาพ"

โดยเปิดรับฟังความคิดเห็น 12 มิ.ย. 67 – 12 ก.ค. 67

QR CODE สำหรับดาวน์โหลดเอกสาร ร่างมาตรฐานฯ แบบแสดงความคิดเห็น แบบฟอร์มลงทะเบียน Link งานประชุม และเอกสารอื่นที่เกี่ยวข้อง

DGA DIGITAL GOVERNMENT DEVELOPMENT AGENCY

แนะนำเว็บไซต์
แหล่งข้อมูล เสริมพลัง สร้างความรู้

standard.dga.or.th

DGA Thailand
DGA Thailand
DGA Thailand
contact@dga.or.th