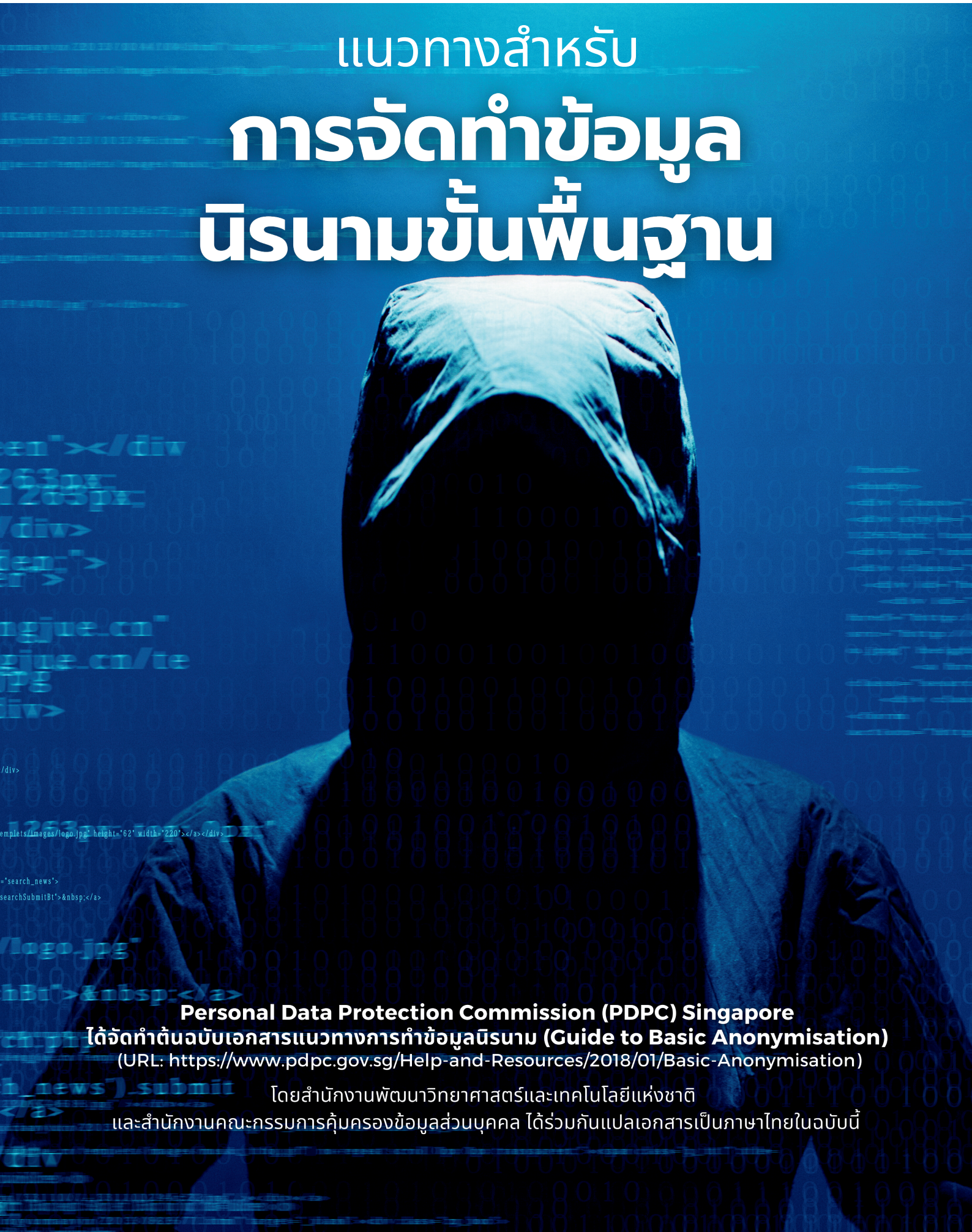


แนวทางสำหรับ
การจัดทำข้อมูลนิรนามขั้นพื้นฐาน

Personal Data Protection Commission (PDPC) Singapore
ได้จัดทำต้นฉบับเอกสารแนวทางการทำข้อมูลนิรนาม (Guide to Basic Anonymisation)
(URL: https://www.pdpc.gov.sg/Help-and-Resources/2018/01/Basic-Anonymisation)

โดยสำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ
และสำนักงานคณะกรรมการคุ้มครองข้อมูลส่วนบุคคล ได้ร่วมกันแปลเอกสารเป็นภาษาไทยในฉบับนี้


แนวทางสำหรับ
การจัดทำข้อมูล
นิรนามขั้นพื้นฐาน
www.pdpc.or.th

สารบัญ

บทนำ

บทนำ

คู่มือฉบับนี้เป็นความรู้เบื้องต้นและแนวทางปฏิบัติสำหรับองค์กรที่ยังไม่คุ้นเคยกับการจัดทำข้อมูลนิรนาม (Anonymization) วิธีดำเนินการจัดทำข้อมูลนิรนามขั้นพื้นฐานและการจัดทำข้อมูลไม่ระบุตัวตน (De-identification) ของชุดข้อมูลที่เป็นโครงสร้าง[1] เป็นข้อความ [2] และไม่ซับซ้อน[3] อย่างเหมาะสม โดยจะนำเสนอขั้นตอนการทำ Anonymization ในบริบทของสี่ยูสเคสทั่วไป

คู่มือฉบับนี้ไม่ได้ครอบคลุมทุกวิธีในการจัดการกับปัญหาทั้งหมดที่เกี่ยวข้องกับการทำ Anonymization การทำ De-identification และการย้อนรอยเพื่อระบุตัวตน (Re-identification) ของชุดข้อมูล ดังนั้นองค์กรต่างๆ จึงควรว่าจ้างผู้เชี่ยวชาญด้านการทำ Anonymization นักสถิติ หรือผู้ประเมินความเสี่ยงอิสระ เพื่อรับรองว่าจะเลือกใช้เทคนิคการทำ Anonymization ที่เหมาะสม หรือการประเมินความเสี่ยงในการทำ Re-idenficiation ซึ่งปัญหาในการทำ Anonymization เป็นปัญหาที่ซับซ้อน (เช่น ชุดข้อมูลขนาดใหญ่ที่เป็นข้อมูลส่วนบุคคลซึ่งถูกเก็บรักษาในระยะยาวหรือเป็นข้อมูลละเอียดอ่อนต่างๆ)

การปฏิบัติตามคำแนะนำในคู่มือฉบับนี้ไม่ได้หมายความถึงการปฏิบัติตามพระราชบัญญัติคุ้มครองข้อมูลส่วนบุคคล (PDPA)

เขตอำนาจศาลต่างๆ มีมุมมองในการทำ Anonymization แตกต่างกัน ดังนั้นคำแนะนำในคู่มือฉบับนี้อาจใช้ไม่ได้กับกฎหมายคุ้มครองข้อมูลในประเทศอื่นๆ

ควรอ่านคู่มือฉบับนี้ร่วมกับ*แนวทางคำแนะนำ*ของคณะกรรมการคุ้มครองข้อมูลส่วนบุคคล (PDPC) *เกี่ยวกับพระราชบัญญัติคุ้มครองข้อมูลส่วนบุคคลสำหรับหัวข้อที่เลือก*

---

[1] “เป็นโครงสร้าง” หมายถึงข้อมูลในรูปแบบที่กำหนดไว้และเป็นตาราง เช่น สเปรดชีตหรือฐานข้อมูลเชิงสัมพันธ์ (เช่น XLSX และ CSV)

[2] “เป็นข้อความ” หมายถึงข้อความ ตัวเลข วันที่ ฯลฯ ซึ่งก็คือข้อมูลตัวอักษรและตัวเลขที่อยู่ในรูปแบบดิจิทัลอยู่แล้ว เทคนิคการทำ Anoymization สำหรับข้อมูลที่ไม่ใช่ข้อความ เช่น เสียง วิดีโอ รูปภาพ ข้อมูลไบโอเมตริกซ์ ฯลฯ มความซับซ้อนมากขึ้นและต้องใช้เทคนิคการทำ Anonymization ต่างๆ ซึ่งไม่ได้กล่าวถึงในคู่มือฉบับนี้

[3] แนวทางที่แนะนำในคู่มือฉบับนี้ใช้กับการจัดการข้อมูลที่เป็นโครงสร้างเท่านั้น (เช่น ข้อมูลที่เป็นข้อความในรูปแบบตารางในคอลัมน์และแถว เช่น สเปรดชีต Microsoft Excel) ตัวอย่างเช่น ภาพถ่ายดิจิทัล ไม่จัดอยู่ในหมวดหมู่นี้

การจัดทำข้อมูลนิรนาม กับการจัดทำข้อมูลไม่ระบุตัวตน

การจัดทำข้อมูลนิรนามกับการจัดทำข้อมูลไม่ระบุตัวตน

การทำ Anonymization หมายถึงการแปลงข้อมูลส่วนบุคคลเป็นข้อมูลที่ไม่สามารถใช้ในการระบุตัวบุคคลได้ PDPC มองว่าการทำ Anonymization เป็นกระบวนการที่อิงตามความเสี่ยง ซึ่งรวมถึงการใช้ทั้งเทคนิคการทำ Anynomization และการป้องกันการทำ Re-identification

การทำ **De-identification**[4] หมายถึงการลบข้อมูลระบุตัวตน (เช่น ชื่อ ที่อยู่ หมายเลขบัตรประจำตัวประชาชน (NRIC)) ที่ระบุตัวบุคคลโดยตรง บางครั้งอาจมีการเข้าใจผิดว่าการทำ De-identification เทียบเท่ากับการทำ Anonymization แต่ความจริงคือกระบวนการนี้เป็นเพียงขั้นตอนแรกของการทำ Anynomization เท่านั้น ชุดข้อมูลที่ผ่านการทำ De-identification อาจถูกทำ Re-identification ได้อย่างง่ายดายเมื่อนำไปรวมกับข้อมูลสาธารณะหรือข้อมูลที่เข้าถึงได้ง่าย

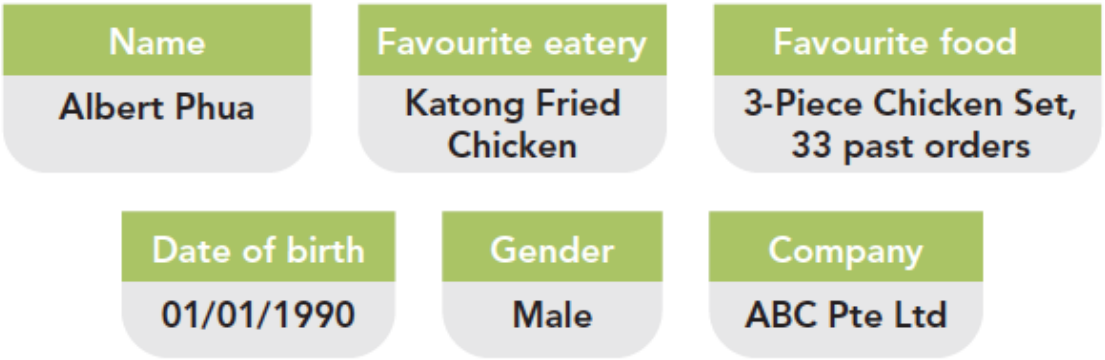
การทำ Re-identification หมายถึงการระบุตัวตนของบุคคลจากชุดข้อมูลที่ผ่านการทำ De-identification หรือ Anonymization ก่อนหน้านี้

ข้อมูลที่ผ่านการทำ Anonymization ไม่ถือเป็นข้อมูลส่วนบุคคล ดังนั้นจึงไม่อยู่ภายใต้การควบคุมของ PDPA หากต้องการข้อมูลเพิ่มเติม โปรดดูหัวข้อเกี่ยวกับการทำ Anonymization ใน *แนวทางคำแนะนำ*ของ PDPC *เกี่ยวกับการคุ้มครองข้อมูลส่วนบุคคลสำหรับหัวข้อที่เลือก*

ตัวอย่างการทำ DE-IDENTIFICATION

อัลเบิร์ตใช้แอปสั่งอาหารบ่อยๆ ที่มีชื่อว่า SuperHungry ซึ่งเป็นแอปสั่งอาหารที่เขาใช้ประจำ แอปดังกล่าวได้เผยแพร่ข้อมูลบางอย่างเกี่ยวกับผู้ใช้ในงานแฮกกาธอน

บันทึกข้อมูลของอัลเบิร์ตในแอป SuperHungry:

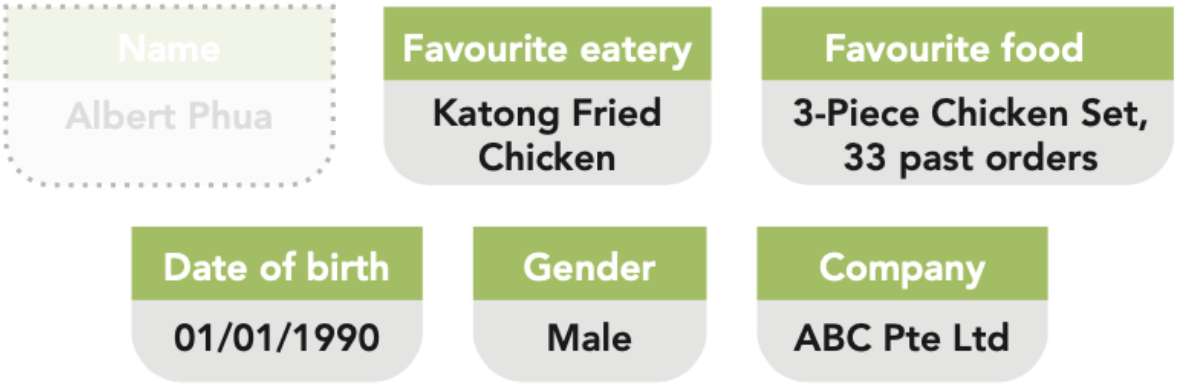


[4] กระบวนการทำ De-identification อาจรวมถึงการใช้นามแฝง

ชื่อ	ร้านอาหารโปรด	อาหารโปรด
อัลเบิร์ต พัว	Katong Fried Chicken	ชุดไก่ 3 ชิ้น การสั่งอาหาร 33 ครั้งล่าสุด
วันเกิด	เพศ	บริษัท
01/01/1990	ชาย	ABC Pte Ltd

SuperHungry ได้ทำ De-identification กับชุดข้อมูลโดยการลบชื่อออกก่อนที่จะเผยแพร่ โดยคิดว่าวิธีนี้เทียบเท่ากับการทำ Anonymization กับชุดข้อมูล

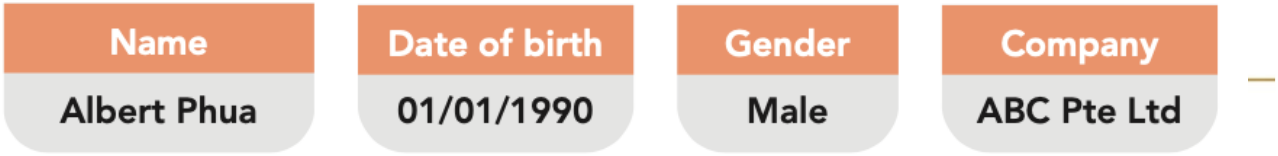
บันทึกที่ผ่านการทำ De-identification ของอัลเบิร์ตที่ SuperHungry เผยแพร่:



ชื่อ	ร้านอาหารโปรด	อาหารโปรด
อัลเบิร์ต พัว	Katong Fried Chicken	ชุดไก่ 3 ชิ้น การสั่งอาหาร 33 ครั้งล่าสุด
วันเกิด	เพศ	บริษัท
01/01/1990	ชาย	ABC Pte Ltd

อย่างไรก็ตาม สามารถทำ Re-idenfication กับข้อมูลของอัลเบิร์ตได้โดยการรวมบันทึกของเขาที่ผ่านการทำ De-identification เข้ากับบันทึกอื่นๆ (เช่น ข้อมูลส่วนบุคคลจากโปรไฟล์ในโซเชียลมีเดียของเขา)

โปรไฟล์ในโซเชียลมีเดียของอัลเบิร์ต:



ชื่อ	วันเกิด	เพศ	บริษัท
อัลเบิร์ต พัว	01/01/1990	ชาย	ABC Pte Ltd

บุคคลใดก็ตามที่มีแรงจูงใจเพียงพอสามารถระบุ[5]ได้ว่าบุคคลนั้นเป็นอัลเบิร์ตได้อย่างง่ายดายจากข้อมูลที่ผ่านการ De-identification หากมีข้อมูลอื่นที่ถูกเปิดเผยต่อสาธารณะหรือที่เข้าถึงได้ง่ายที่สามารถนำไปใช้เพื่อทำ Re-identification ได้ หากชุดข้อมูลหรือชุดข้อมูลรวมมีความละเอียดอ่อน จะต้องทำ Anonymization เพิ่มเติม

[5] ตัวอย่างนี้จะทำการ Re-identification เฉพาะในกรณีที่ข้อมูลนี้เป็นข้อมูลเฉพาะของอัลเบิร์ตในหน่วยข้อมูลทุกหน่วย

บทนำเกี่ยวกับแนวคิดการจัดทำข้อมูลนิรนามขั้นพื้นฐาน

บทนำเกี่ยวกับแนวคิดการจัดทำข้อมูลนิรนามขั้นพื้นฐาน

การทำ Anonymization จำเป็นต้องมีความเข้าใจที่ดีในองค์ประกอบต่อไปนี้เพื่อเลือกเทคนิคและระดับการทำ Anonymization ที่เหมาะสม

ก วัตถุประสงค์ของการทำ Anonymization และอรรถประโยชน์

วัตถุประสงค์ของการทำ Anonymization จะต้องชัดเจน เพราะการทำ Anonymization ควรทำเพื่อวัตถุประสงค์เฉพาะ กระบวนการทำ Anonymization ไม่ว่าจะใช้เทคนิคใดก็ตาม จะลดข้อมูลดั้งเดิมในชุดข้อมูลลงในระดับหนึ่ง ดังนั้นเมื่อระดับการทำ Anonymization เพิ่มขึ้น โดยทั่วไปอรรถประโยชน์ (เช่น ความชัดเจนและ/หรือความแม่นยำ) ของชุดข้อมูลก็จะลดลง ดังนั้น องค์กรจำเป็นต้องเลือกว่าจะให้ความสำคัญในด้านใดระหว่างอรรถประโยชน์ที่ยอมรับได้ (หรือที่คาดหวัง) กับความเสี่ยงในการทำ Re-identification

โปรดทราบว่าไม่ควรประเมินอรรถประโยชน์ในระดับของชุดข้อมูลทั้งหมด เนื่องจากโดยทั่วไปแล้วอรรถประโยชน์จะแตกต่างกันสำหรับคุณลักษณะต่างๆ ในขณะที่สิ่งสำคัญประการหนึ่งคือความถูกต้องของคุณลักษณะของข้อมูลเฉพาะและไม่ควรนำเทคนิคการลดรายละเอียดข้อมูล (Generalization) หรือการทำ Anonymization ไปใช้ (เช่น อาการป่วยและยาที่จ่ายให้กับคนไข้อาจเป็นข้อมูลสำคัญในการวิเคราะห์แนวโน้มการรับเข้ารักษาในโรงพยาบาลได้) คุณลักษณะของข้อมูลไม่มีประโยชน์ตามวัตถุประสงค์ที่ต้องการและอาจถูกลบทิ้งได้โดยไม่ส่งผลกระทบต่อการใช้ข้อมูลของผู้รับ (เช่น วันเกิดของลูกค้าอาจไม่มีความสำคัญในการวิเคราะห์แนวโน้มการซื้อสินค้า)

ข้อควรพิจารณาที่สำคัญอีกประการหนึ่งในการกำหนดว่าจะให้ความสำคัญในด้านใดระหว่างอรรถประโยชน์และการทำ Anonymization คือ จะมีความเสี่ยงเพิ่มเติมหรือไม่หากผู้รับข้อมูลรู้ว่าใช้เทคนิคและระดับความละเอียดอ่อนใดในการทำ Anonymization ในอีกด้านหนึ่ง การรู้ข้อมูลนี้อาจช่วยให้นักวิเคราะห์เข้าใจผลลัพธ์และตีความได้ดีขึ้น แต่ในทางกลับกัน อาจมีเบาะแสซึ่งอาจทำให้มีความเสี่ยงสูงขึ้นที่จะมีการทำ Re-identification

ข การย้อนกลับได้

โดยทั่วไปแล้ว กระบวนการทำ Anonymization จะ "ไม่สามารถย้อนกลับได้" และผู้รับชุดข้อมูลที่ผ่านการทำ Anonymization จะไม่สามารถสร้างข้อมูลดั้งเดิมขึ้นมาใหม่ได้ อย่างไรก็ตาม อาจมีกรณีที่องค์กรที่ทำ Anonymization ยังคงสามารถสร้างชุดข้อมูลดั้งเดิมขึ้นมาใหม่จากข้อมูลที่ผ่านการ**ทำ** Anonymization **ได้ ในกรณีดังกล่าว กระบวนการทำ** Anonymization **จะ“สามารถย้อนกลับได้”**

ค ลักษณะของเทคนิคการทำ Anonymization

ลักษณะต่างๆ ของเทคนิคการทำ Anonymization ต่างๆ หมายความว่าเทคนิคบางอย่างอาจเหมาะสมกับบางสถานการณ์หรือบางประเภทข้อมูลมากกว่าเทคนิคอื่นๆ ตัวอย่างเช่น บางเทคนิค (เช่น การปิดบังอักขระ (Masking)) อาจเหมาะสำหรับการใช้กับตัวระบุโดยตรงมากกว่า และเทคนิคอื่นๆ (เช่น การรวมข้อมูล (Aggregation)) เหมาะสำหรับตัวระบุโดยอ้อม อีกลักษณะหนึ่งที่ต้องพิจารณาคือค่าคุณลักษณะเป็นค่าต่อเนื่อง (เช่น ส่วนสูง = 1.61 ม.) หรือไม่ต่อเนื่อง (เช่น "ใช่" หรือ "ไม่ใช่") เนื่องจากเทคนิค เช่น การเพิ่มสัญญาณรบกวนในข้อมูล (Perturbation) มีประสิทธิภาพมากกว่าสำหรับค่าต่อเนื่อง

เทคนิคการทำ Anonymization ต่างๆ ยังปรับเปลี่ยนข้อมูลด้วยวิธีที่แตกต่างกันมาก บางเทคนิคแก้ไขคุณลักษณะเพียงบางส่วนเท่านั้น (เช่น การทำ Masking) บางเทคนิคเปลี่ยนค่าของคุณลักษณะในบันทึกต่างๆ (เช่น การทำ Aggregation) บางเทคนิคเปลี่ยนค่าของคุณลักษณะด้วยค่าที่ไม่เกี่ยวข้องแต่มีลักษณะเฉพาะ (เช่น การใช้นามแฝง (Pseudonymization)) และบางเทคนิคลบคุณลักษณะออกทั้งหมด (เช่น การกำจัดคุณลักษณะ (Suppression))

บางเทคนิคการทำ Anonymization สามารถใช้ร่วมกันได้ (เช่น การทำ Suppression หรือการลบบันทึก (ค่าผิดปกติ) หลังทำ Generaliztion)

ง ข้อมูลอนุมาน

ข้อมูลบางอย่างอาจถูกอนุมานจากข้อมูลที่ผ่านการทำ Anonymization ได้ ตัวอย่างเช่น การทำ Masking อาจซ่อนข้อมูลส่วนบุคคล แต่ไม่ได้ซ่อนความยาวของค่าเดิมในแง่ของจำนวนอักขระ

องค์กรควรพิจารณาลำดับการนำเสนอข้อมูลที่ผ่านการทำ Anonymization ด้วย ตัวอย่างเช่น หากผู้รับรู้ว่าบันทึกข้อมูลถูกรวบรวมตามลำดับ (เช่น การลงทะเบียนของผู้ที่มาใช้บริการตามลำดับ) ควรใช้ความระมัดระวัง (โดยไม่ส่งผลกระทบต่อออรรถประโยชน์) ในการสับเปลี่ยนชุดข้อมูลทั้งหมดใหม่เพื่อไม่ให้สามารถอนุมานตามลำดับการบันทึกข้อมูลได้

การอนุมานไม่ได้จำกัดเพียงแค่คุณลักษณะเดียว แต่ยังสามารถนำไปใช้กับคุณลักษณะต่างๆ ได้ด้วย แม้ว่าเทคนิคการทำ Anonymization จะถูกนำไปใช้กับคุณลักษณะทั้งหมดก็ตาม ดังนั้น กระบวนการทำ Anonymization จะต้องคำนึงถึงความเป็นไปได้ทั้งหมดที่อาจเกิดขึ้นจากการอนุมาน ทั้งก่อนและหลังการเลือกเทคนิคที่จะใช้

จ ความเชี่ยวชาญในเนื้อเรื่อง

เทคนิคการทำ Anonymization โดยทั่วไปจะลดความสามารถในการระบุตัวตนของบุคคลตั้งแต่หนึ่งรายขึ้นไปจากชุดข้อมูลดั้งเดิมให้อยู่ในระดับที่องค์กรยอมรับความเสี่ยงได้

ควรประเมินความสามารถในการระบุตัวตนและความสามารถในการทำ Re-idenfitication[6] ก่อนและหลังการใช้เทคนิคการทำ Anonymization ขั้นตอนนี้จำเป็นต้องมีความเข้าใจในเนื้อเรื่องที่เกี่ยวข้องกับข้อมูลเป็นอย่างดี ตัวอย่างเช่น หากชุดข้อมูลเป็นข้อมูลด้านการดูแลสุขภาพ องค์กรควรมอบหมายให้บุคคลที่มีความรู้ด้านการดูแลสุขภาพที่เพียงพอเป็นผู้ประเมินเอกลักษณ์ของบันทึก (เช่น ข้อมูลสามารถระบุตัวตนหรือทำ Re-identification ได้ในระดับใด)

การประเมินก่อนกระบวนการทำ Anonymization ช่วยให้แน่ใจว่าได้ระบุและทำความเข้าใจโครงสร้างและข้อมูลภายในคุณลักษณะอย่างชัดเจนแล้ว และได้ประเมินความเสี่ยงของการอนุมานทั้งโดยชัดแจ้งและโดยนัยจากข้อมูลดังกล่าวแล้ว ตัวอย่างเช่น คุณลักษณะที่เป็นปีเกิดจะบ่งบอกถึงอายุได้โดยนัย เช่นเดียวกับหมายเลข NRIC ในระดับหนึ่ง การประเมินหลังจากกระบวนการทำ Anonymization จะบอกถึงความเสี่ยงที่เหลืออยู่ของการทำ Re-identification จากข้อมูลที่ผ่านการทำ Anonymization แล้ว

อีกกรณีหนึ่งคือเมื่อคุณลักษณะของข้อมูลถูกสลับระหว่างบันทึกต่างๆ เฉพาะผู้เชี่ยวชาญในเนื้อหาเท่านั้นที่จะเข้าใจบันทึกที่ผ่านการทำ Anonymization แล้ว

ดังนั้นการเลือกเทคนิคการทำ Anonymization ที่ถูกต้องจึงขึ้นอยู่กับการรับรู้ข้อมูลโดยชัดเจนและโดยนัยที่มีอยู่ในชุดข้อมูลและจำนวนหรือประเภทของข้อมูลที่ต้องการทำ Anonymization

ญ ความสามารถในกระบวนการและเทคนิคการทำ Anonymization

องค์กรที่ต้องการแชร์ชุดข้อมูลที่ผ่านการทำ Anonymization แล้วควรตรวจสอบให้แน่ใจว่ากระบวนการทำ Anonymization นั้นดำเนินการโดยพนักงานที่ได้รับการฝึกอบรมและคุ้นเคยกับเทคนิคดังกล่าวและหลักการในการทำ Anonymization หากไม่มีผู้เชี่ยวชาญภายในองค์กร ควรขอความช่วยเหลือจากผู้เชี่ยวชาญภายนอก

[6] “ความสามารถในการระบุตัวตน” หมายถึงระดับที่สามารถระบุตัวบุคคลได้จากชุดข้อมูลตั้งแต่หนึ่งชุดขึ้นไปที่มีทั้งตัวระบุโดยตรงและโดยอ้อม ในขณะที่ “ความสามารถในการทำ Re-identification” หมายถึงระดับที่สามารถทำ Re-identification กับตัวบุคคลจากชุดข้อมูลที่ผ่านการทำ Anonymization แล้ว

ช ผู้รับข้อมูล

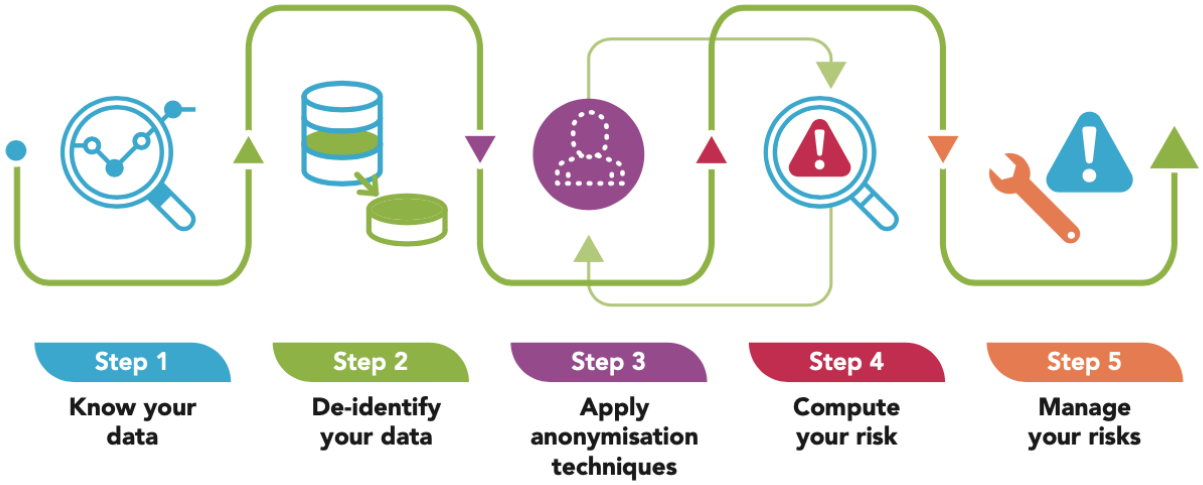
ปัจจัยต่างๆ เช่น ความเชี่ยวชาญของผู้รับข้อมูลในเนื้อเรื่องและการควบคุมที่นำมาใช้เพื่อจำกัดจำนวนผู้รับและเพื่อป้องกันไม่ให้มีการแชร์ข้อมูลกับบุคคลที่ไม่ได้รับอนุญาต มีบทบาทสำคัญในการเลือกเทคนิคการทำ Anonymization โดยเฉพาะอย่างยิ่ง การใช้ข้อมูลที่ผ่านการทำ Anonymization แล้ว โดยผู้รับอาจทำให้มีข้อจำกัดในเทคนิคที่จะนำมาใช้ เนื่องจากอรรถประโยชน์ของข้อมูลอาจสูญหายเกินระดับที่ยอมรับได้ ต้องใช้ความระมัดระวังเป็นพิเศษเมื่อเผยแพร่ข้อมูลต่อสาธารณะ และองค์กรต่างๆ จะต้องมีรูปแบบการทำ Anonymization ที่เข้มงวดมากขึ้น เมื่อเทียบกับข้อมูลที่แชร์ภายใต้ข้อตกลงตามสัญญา

ซ เครื่องมือต่างๆ

เครื่องมือซอฟต์แวร์มีประโยชน์มากในการใช้เทคนิคการทำ Anonymization โปรดดูภาคผนวก จ สำหรับตัวอย่างเครื่องมือในการทำ Anonymization ที่มีอยู่ในตลาด

กระบวนการการจัดทำข้อมูลนิรนาม

กระบวนการทำ Anonymization



ขั้นตอนที่ 1	ขั้นตอนที่ 2	ขั้นตอนที่ 3	ขั้นตอนที่ 4	ขั้นตอนที่ 5
รู้จักข้อมูลของคุณ	ทำ De-identification กับข้อมูลของคุณ	ใช้เทคนิคการทำ Anonymization	คำนวณความเสี่ยงของคุณ	บริหารความเสี่ยงของคุณ

คุณสามารถใช้ห้าขั้นตอนเหล่านี้เพื่อทำ Anonymization กับชุดข้อมูลของคุณได้ตามความเหมาะสม ขึ้นอยู่กับยูสเคสของคุณ ในคู่มือฉบับนี้ เราจะอธิบายขั้นตอนเหล่านี้โดยใช้กรณีการใช้ข้อมูลทั่วไปห้ากรณีโดยองค์กรต่างๆ

ในการใช้ข้อมูลทุกกรณี คุณควรตรวจสอบให้แน่ใจว่า:

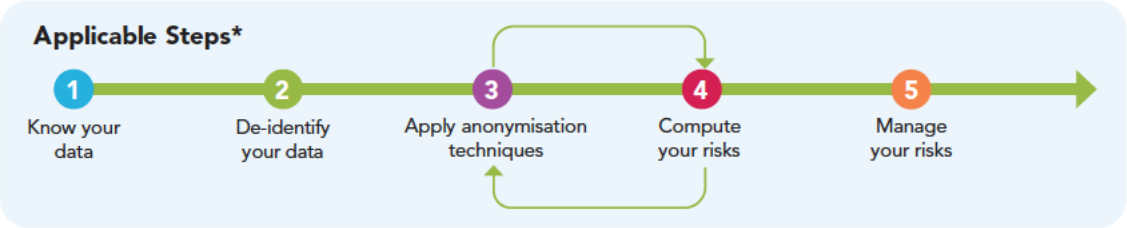
มีการประมวลผลข้อมูลเท่าที่จำเป็น (Minimization) โดยจะแชร์เฉพาะคุณลักษณะของข้อมูลและข้อมูลที่ดึงออกมา (หากเป็นไปได้) กับบุคคลที่สามเท่าที่จำเป็นเท่านั้น

ข้อมูลที่ใช้ระบุตัวตนของชุดข้อมูลที่คุณทำ Anonymization ไม่ควรถูกเปิดเผยต่อสาธารณะ (เช่น หากคุณทำ Anonymization กับข้อมูลบนฐานข้อมูลสมาชิก ก็ไม่ควรเปิดเผยโปรไฟล์ของฐานสมาชิกต่อสาธารณะ) และ

ชุดข้อมูลที่ผ่านการทำ Anonymization แล้วและตารางจับคู่ข้อมูลระบุตัวตนสำหรับนามแฝงมีระดับการป้องกันที่เหมาะสมเพื่อป้องกันการทำ Re-identification โดยทั่วไป ยิ่งคุณแก้ไขชุดข้อมูลด้วยการทำ Anonymization น้อยเท่าใด คุณก็ยิ่งจำเป็นต้องปกป้องชุดข้อมูลมากขึ้นเท่านั้น เนื่องจากมีความเสี่ยงในการทำ Re-identification สูงขึ้น

ยูสเคส: คุณสามารถใช้ข้อมูลที่ผ่านการทำ Anonymiztion แล้วหรือข้อมูลที่ผ่านการทำ De-identification แล้วได้อย่างไร

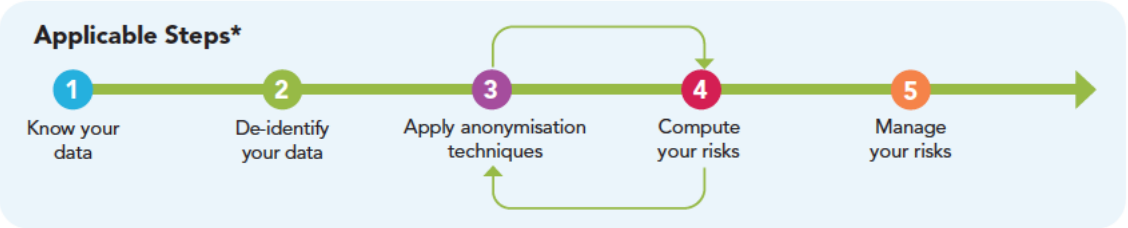
ต่อไปนี้เป็นตัวอย่างวิธีการใช้ข้อมูลที่ผ่านการทำ Anonymization แล้วหรือข้อมูลที่ผ่านการทำ De-identification แล้วในองค์กรของคุณ



ขั้นตอนที่เหมาะสม*				
1	2	3	4	5
รู้จักข้อมูลของคุณ	ทำ De-identification กับข้อมูลของคุณ	ใช้เทคนิคการทำ Anonymization	คำนวณความเสี่ยงของคุณ	บริหารความเสี่ยงของคุณ

ยูสเคส	**การแชร์ข้อมูลภายใน (ข้อมูลที่ผ่านการทำ De-identification แล้ว) (เช่น ข้อมูลลูกค้าที่ผ่านการทำ De-identification แล้วที่แชร์ระหว่างฝ่ายขายกับฝ่ายการตลาดเพื่อการวิเคราะห์และการพัฒนาแคมเปญการตลาดแบบกำหนดเป้าหมายภายในองค์กร)**
คำอธิบาย	ข้อมูลจะถูกทำ De-identification เพื่อรองรับการแชร์ข้อมูลระดับบันทึกและการใช้งานภายในองค์กรเท่านั้น ซึ่งอาจจำเป็นต้องห้ามแก้ไขรายละเอียดส่วนใหญ่ในข้อมูล ข้อมูลที่ผ่านการทำ De-identification แล้วยังคงเป็นข้อมูลส่วนบุคคลเนื่องจากมีแนวโน้มที่จะสามารถทำ Re-identification ได้อย่างง่ายดาย อย่างไรก็ตาม การทำ De-identification กับข้อมูลเป็นแนวทางปฏิบัติที่ดีเนื่องจากเป็นการป้องกันเพิ่มเติมอีกชั้นหนึ่ง

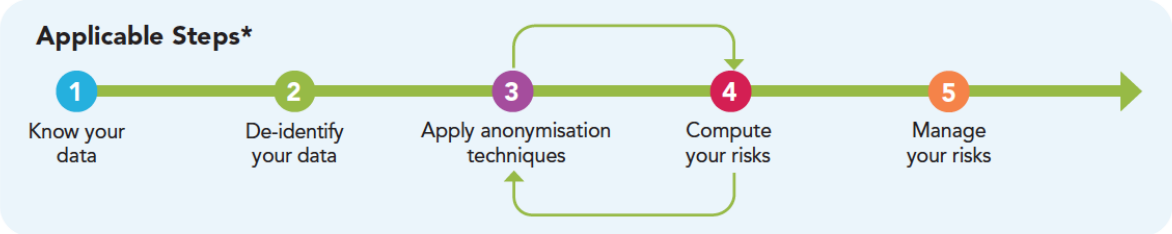
จำเป็นต้องมีการควบคุมเพิ่มเติมเพื่อป้องกันการทำ Re-identification หรือไม่	ใช่
ผลลัพธ์สุดท้ายถือเป็นข้อมูลที่ผ่านการทำ Anonymization แล้วหรือไม่	ไม่
***ขั้นตอนที่เหมาะสม**	1 2 5
ยูสเคส	**การแชร์ข้อมูลภายใน (ข้อมูลที่ผ่านการทำ Anonymization แล้ว)** (เช่น ข้อมูลที่ผ่านการทำ Anonymization แล้วเกี่ยวกับข้อมูลประชากรของผู้บริโภครายใหญ่และรูปแบบการใช้จ่ายตามลำดับที่แชร์กับทีมดูแลความภักดีเพื่อพัฒนาคุณค่าที่ส่งมอบให้ลูกค้าที่แตกต่าง)
คำอธิบาย	องค์กรสามารถใช้ข้อมูลที่ผ่านการทำ Anonymization แล้วแทนข้อมูลที่ผ่านการทำ De-identification แล้วสำหรับการแชร์ภายในในกรณีต่อไปนี้ได้: **• การแชร์ข้อมูลภายในไม่จำเป็นต้องมีข้อมูลส่วนบุคคลที่ผ่านการทำ De-identification แล้วโดยละเอียด (เช่น สำหรับการวิเคราะห์แนวโน้ม)** **• ข้อมูลที่เกี่ยวข้องมีลักษณะที่ละเอียดอ่อนมากกว่า (เช่น ข้อมูลทางการเงิน) หรือ** **• ชุดข้อมูลขนาดใหญ่ที่ใช้ร่วมกันกับแผนกอื่นๆ** **ในกรณีดังกล่าว องค์กรสามารถใช้กระบวนการทำ Anonymization ที่แนะนำสำหรับการแชร์ข้อมูลภายนอกกับยูสเคสที่มีการแชร์ข้อมูลภายในเพื่อลดความเสี่ยงในการทำ Re-identification และการเปิดเผยข้อมูล**
จำเป็นต้องมีการควบคุมเพิ่มเติมเพื่อป้องกันการทำ Re-identification หรือไม่	ใช่
ผลลัพธ์สุดท้ายถือเป็นข้อมูลที่ผ่านการทำ Anonymization แล้วหรือไม่	ใช่
***ขั้นตอนที่เหมาะสม**	1 2 3 4 5



ขั้นตอนที่เหมาะสม*				
1	2	3	4	5
รู้จักข้อมูลของคุณ	ทำ De-identification กับข้อมูลของคุณ	ใช้เทคนิคการทำ Anonymization	คำนวณความเสี่ยงของคุณ	บริหารความเสี่ยงของคุณ

ยูสเคส	**การแชร์ข้อมูลภายนอก (เช่น ข้อมูลลูกค้าที่ผ่านการทำแล้ว Anonymization แล้วที่แชร์ระหว่างฝ่ายขายกับพันธมิตรทางธุรกิจภายนอกสำหรับการวิเคราะห์โปรไฟล์ลูกค้าและการพัฒนาผลิตภัณฑ์ที่มีตราสินค้าร่วม)**
คำอธิบาย	ข้อมูลระดับบันทึกที่แชร์กับบุคคลภายนอกที่ได้รับอนุญาตเพื่อวัตถุประสงค์ในการทำงานร่วมกันทางธุรกิจ ใช้เทคนิคการทำ Anonymization เพื่อแปลงข้อมูลส่วนบุคคลเป็นข้อมูลที่ไม่ระบุตัวตน
จำเป็นต้องมีการควบคุมเพิ่มเติมเพื่อป้องกันการทำ Re-identification หรือไม่	ใช่
ผลลัพธ์สุดท้ายถือเป็นข้อมูลที่ผ่านการทำ Anonymization แล้วหรือไม่	ใช่
***ขั้นตอนที่เหมาะสม**	1 2 3 4 5

ยูสเคส	**การเก็บรักษาข้อมูลระยะยาวเพื่อการวิเคราะห์ข้อมูล (เช่น การวิเคราะห์ประวัติแนวโน้มของลูกค้า)**
คำอธิบาย	ใช้เทคนิคการทำ Anonymization เพื่อแปลงข้อมูลส่วนบุคคลเป็นข้อมูลที่ไม่ระบุตัวตน และทำให้สามารถเก็บข้อมูลไว้ที่ระดับบันทึกเกินระยะเวลาการเก็บรักษาเพื่อการวิเคราะห์ข้อมูลในระยะยาวได้
จำเป็นต้องมีการควบคุมเพิ่มเติมเพื่อป้องกันการทำ Re-identification หรือไม่	ใช่
ผลลัพธ์สุดท้ายถือเป็นข้อมูลที่ผ่านการทำ Anonymization แล้วหรือไม่	ใช่
***ขั้นตอนที่เหมาะสม**	1 2 3 4 5



ขั้นตอนที่เหมาะสม*				
1	2	3	4	5
รู้จักข้อมูลของคุณ	ทำ De-identification กับข้อมูลของคุณ	ใช้เทคนิคการทำ Anonymization	คำนวณความเสี่ยงของคุณ	บริหารความเสี่ยงของคุณ

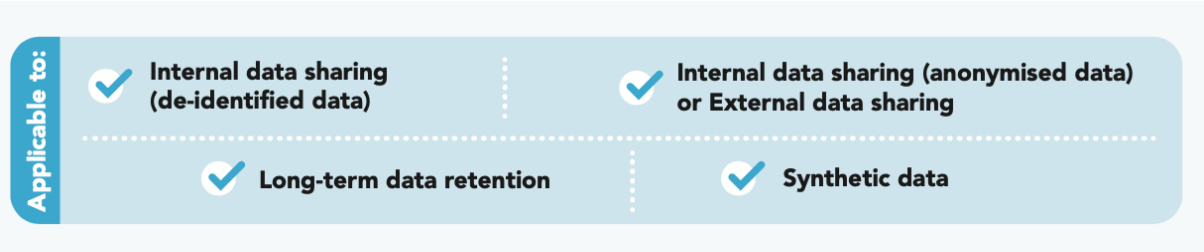
ยูสเคส	**ข้อมูลสังเคราะห์[7] เพื่อวัตถุประสงค์ในการพัฒนาและทดสอบแอปพลิเคชัน โดยไม่จำเป็นต้องจำลองลักษณะทางสถิติของข้อมูลดั้งเดิม (เช่น ใช้สำหรับการทดสอบโดยผู้จำหน่ายภายนอกที่ถูกจ้างให้พัฒนาและทดสอบแอปพลิเคชันสำหรับบัญชีเงินเดือน)**
คำอธิบาย	สามารถสร้างข้อมูลสังเคราะห์ระดับบันทึกได้จากข้อมูลดั้งเดิมโดยการทำ Anonymization กับคุณลักษณะของข้อมูลทั้งหมดอย่างละเอียดโดยใช้เทคนิคการทำ Anonymization ในคู่มือฉบับนี้ เพื่อให้คุณลักษณะของข้อมูลทั้งหมดได้รับการแก้ไขอย่างมีนัยสำคัญ และเพื่อให้บันทึกทั้งหมดที่สร้างขึ้นไม่ตรงกับบันทึกของบุคคลใด ๆ ในข้อมูลดั้งเดิม ในกรณีนี้ การใช้เทคนิคการทำ Anonymization จะไม่คงลักษณะทางสถิติของข้อมูลดั้งเดิมไว้ จึงไม่เหมาะสำหรับวัตถุประสงค์ที่ซับซ้อน เช่น การฝึกโมเดลปัญญาประดิษฐ์หรือการวิเคราะห์ข้อมูล
จำเป็นต้องมีการควบคุมเพิ่มเติมเพื่อป้องกันการทำ Re-identification หรือไม่	ไม่[8]
ผลลัพธ์สุดท้ายถือเป็นข้อมูลที่ผ่านการทำ Anonymization แล้วหรือไม่	ใช่
***ขั้นตอนที่เหมาะสม**	1 2 3

หมายเหตุ: ในข้อมูลสังเคราะห์ ตัวระบุโดยตรง "ปลอม" ที่ใช้ไม่ควรเกี่ยวข้องกับบุคคลจริง กล่าวคือ NRIC ที่สร้างขึ้นแบบสุ่มที่มีชื่อที่สร้างแบบสุ่มไม่ควรเหมือนกับ NRIC และชื่อของบุคคลจริง

[7] อีกวิธีหนึ่งที่ไม่ได้กล่าวถึงในคู่มือฉบับนี้คือการสร้างข้อมูลสังเคราะห์ตั้งแต่ต้น วิธีนี้สามารถทำได้โดยการสุ่มสร้างชุดข้อมูลที่ตอบสนองความต้องการด้านรูปแบบข้อมูลเท่านั้น หรือโดยการสร้างชุดข้อมูลที่ยังคงรักษาลักษณะทางสถิติของชุดข้อมูลดั้งเดิมโดยใช้ปัญญาประดิษฐ์หรือวิธีการอื่นๆ

[8] ดูสมมติฐานในขั้นตอนที่ 5: การจัดการความเสี่ยงในการทำ Re-identification และการเปิดเผยข้อมูลของคุณ

ขั้นตอนที่ 1 รู้จักข้อมูลของคุณ



เหมาะสมกับ:	การแชร์ข้อมูลภายใน (การทำ De-identification กับข้อมูล)	การแชร์ข้อมูลภายใน (การทำ Anonymizationกับข้อมูล) หรือการแชร์ข้อมูลภายนอก
	การเก็บรักษาข้อมูลระยะยาว	ข้อมูลสังเคราะห์

บันทึกข้อมูลส่วนบุคคลประกอบด้วยคุณลักษณะของข้อมูลที่มีระดับการระบุตัวตนและความอ่อนไหวของแต่ละบุคคลที่แตกต่างกัน

โดยทั่วไปแล้ว การทำ Anonymization จะมีการลบตัวระบุโดยตรงออกและการแก้ไขตัวระบุโดยอ้อม โดยปกติแล้วคุณลักษณะเป้าหมายจะไม่เปลี่ยนแปลง ยกเว้นในกรณีที่มีวัตถุประสงค์เพื่อสร้างข้อมูลสังเคราะห์ ตารางและตัวอย่างด้านล่างแสดงให้เห็นวิธีการจัดประเภทคุณลักษณะของข้อมูลในบันทึกข้อมูลโดยทั่วไป

	ตัวระบุโดยตรง	ตัวระบุโดยอ้อม	คุณลักษณะเป้าหมาย
การจำแนกประเภทของคุณลักษณะของข้อมูลในชุดข้อมูล	คุณลักษณะของข้อมูลเหล่านี้เป็นข้อมูลเฉพาะของบุคคล และสามารถใช้เป็นคุณลักษณะของข้อมูลหลักเพื่อทำ Re-identification ได้	คุณลักษณะของข้อมูลเหล่านี้ไม่ใช่ข้อมูลเฉพาะของบุคคล แต่อาจทำ Re-identification ได้เมื่อนำไปรวมกับข้อมูลอื่นๆ (เช่น การรวมอายุ เพศ และรหัสไปรษณีย์เข้าด้วยกัน)	เหล่านี้เป็นคุณลักษณะของข้อมูลที่มีอรรถประโยชน์หลักของชุดข้อมูล ในบริบทของการประเมินความเพียงพอของการทำ Anonymization คุณลักษณะของข้อมูลนี้อาจมีความละเอียดอ่อนและอาจเป็นไปได้สูงที่จะส่งผลเสียต่อบุคคลได้เมื่อถูกนำไปเปิดเผย

	ตัวระบุโดยตรง	ตัวระบุโดยอ้อม	คุณลักษณะเป้าหมาย
ความสามารถในการเข้าถึงข้อมูล	คุณลักษณะของข้อมูลเหล่านี้มักจะเป็นข้อมูลสาธารณะหรือสามารถเข้าถึงได้ง่าย	คุณลักษณะของข้อมูลเหล่านี้อาจเป็นข้อมูลสาธารณะหรือสามารถเข้าถึงได้ง่าย	คุณลักษณะของข้อมูลเหล่านี้มักจะไม่ใช่ข้อมูลสาธารณะหรือไม่สามารถเข้าถึงได้ง่าย ข้อมูลเหล่านี้ไม่สามารถใช้เพื่อทำ Re-identification ได้เนื่องจากโดยทั่วไปแล้วเป็นข้อมูลที่มีกรรมสิทธิ์
ตัวอย่างทั่วไปของชุดข้อมูล	• ชื่อ • ที่อยู่อีเมล • **หมายเลขโทรศัพท์มือถือ** • **หมายเลข** NRIC • หมายเลขหนังสือเดินทาง • **หมายเลขบัญชี** • หมายเลขสูติบัตร • หมายเลขประจำตัวต่างประเทศ (FIN) • **หมายเลขใบอนุญาตทำงาน** • **ชื่อผู้ใช้โซเชียลมีเดีย**	• อายุ • เพศ • **เชื้อชาติ** • วันเกิด • ที่อยู่ • รหัสไปรษณีย์ • **ตำแหน่งงาน** • **ชื่อบริษัท** • **สถานภาพการสมรส** • ส่วนสูง • น้ำหนัก • ที่อยู่อินเทอร์เน็ตโปรโตคอล (IP) • **หมายเลขทะเบียนรถ** • หมายเลขหน่วยในรถยนต์ (IU) • ตำแหน่งของระบบกำหนดตำแหน่งบนพื้นโลก (GPS)	• **ธุรกรรม (เช่น การซื้อ)** • เงินเดือน • **อันดับเครดิต** • **กรมธรรม์ประกันภัย** • **การวินิจฉัยทางการแพทย์** • **สถานะการฉีดวัคซีน**

ตัวอย่างที่ 1: การจำแนกประเภทคุณลักษณะของข้อมูลในบันทึกข้อมูลพนักงาน

Staff ID	Name	Department	Gender	Date of birth	Start date of service	Employment type
39192	Sandy Thomas	Research & Development	F	08/01/1971	02/03/1997	Part-time
37030	Paula Swenson	Engineering	F	15/05/1976	08/03/2015	Full-time
22722	Bosco Wood	Engineering	M	31/12/1973	30/07/1991	Full-time
28760	Stef Stone	Engineering	F	24/12/1970	18/03/2010	Part-time
13902	Jake Norma	Human Resource	M	15/07/1973	28/05/2012	Part-time

Direct identifiers | Indirect identifiers | Target variables

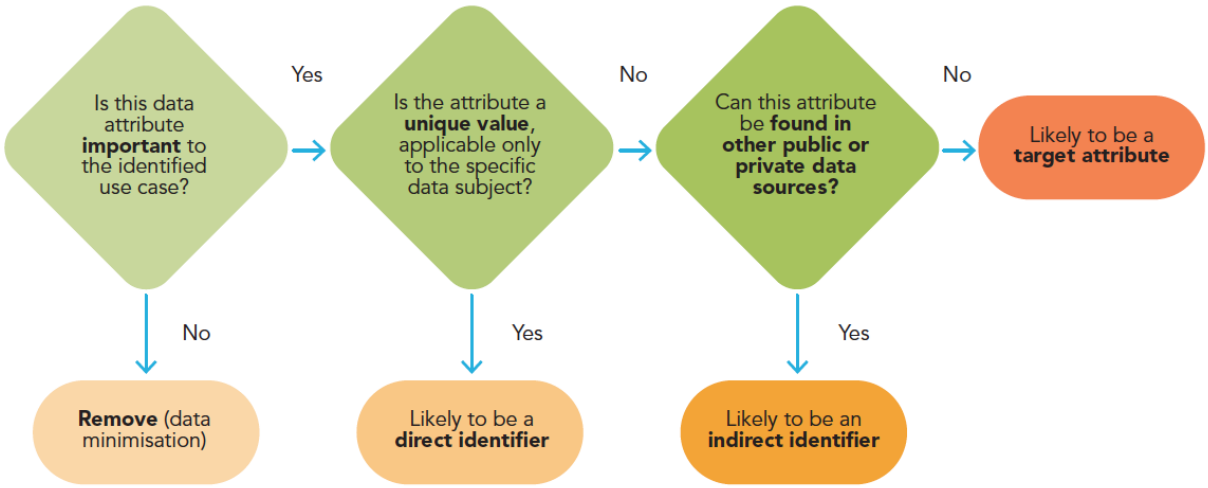
เลขประจำตัวพนักงาน	ชื่อ	แผนก	เพศ	วันเกิด	วันเริ่มงาน	ประเภทการจ้างงาน
	แซนดี้ โธมัส	วิจัยและพัฒนา	ญ			ชั่วคราว
	พอล่า ซเวนสัน	วิศวกรรม	ญ			ประจำ
	บอสโค วู้ด	วิศวกรรม	ช			ประจำ
	สเตฟ สโตน	วิศวกรรม	ญ			ชั่วคราว
	เจค นอร์มา	ทรัพยากรมนุษย์	ช			ชั่วคราว
ตัวระบุโดยตรง		ตัวระบุโดยอ้อม			ตัวแปรเป้าหมาย	

ตัวอย่างที่ 2: การจำแนกประเภทคุณลักษณะของข้อมูลในบันทึกข้อมูลลูกค้า

Customer ID	Name	Gender	Date of birth	Postal code	Occupation	Income	Education	Marital status
56833	Jenny Jefferson	F	05/08/1975	570150	Data scientist	$13,000	Masters	Widowed
50271	Peter G	M	14/12/1973	787589	University lecturer	$12,000	Doctorate	Married
53041	Tim Lake	F	02/03/1985	408600	Researcher	$7,000	Doctorate	Divorced
17290	Remy Bay	M	27/03/1968	570150	Database administrator	$8,000	Bachelor	Married
52388	Walter Paul	M	25/06/1967	199588	Architect	$10,000	Masters	Single
Direct identifiers		Indirect identifiers				Target variables	Indirect identifiers	

เลขประจำตัวลูกค้า	ชื่อ	เพศ	วันเกิด	รหัสไปรษณีย์	อาชีพ	รายได้	การศึกษา	สถานภาพสมรส
	เจนนี เจฟเฟอสัน	ญ			นักวิทยาศาสตร์ข้อมูล		ปริญญาโท	ม่าย
	ปีเตอร์ จี	ช			อาจารย์มหาวิทยาลัย		ปริญญาเอก	แต่งงาน
	ทีม เลค	ญ			นักวิจัย		ปริญญาเอก	หย่า
	เรมี เบย์	ช			ผู้ดูแลระบบฐานข้อมูล		ปริญญาตรี	แต่งงาน
	วอลเตอร์ พอล	ช			สถาปนิก		ปริญญาโท	โสด
ตัวระบุโดยตรง		ตัวระบุโดยอ้อม				ตัวแปรเป้าหมาย	ตัวระบุโดยอ้อม	

ควรลบคุณลักษณะของข้อมูลใดๆ ที่ไม่จำเป็นในชุดข้อมูลผลลัพธ์ออก ซึ่งเป็นส่วนหนึ่งของการทำ Minimization กับข้อมูล ด้านล่างคือผังงานอย่างง่ายเพื่อช่วยคุณในการจำแนกคุณลักษณะของข้อมูลของคุณได้อย่างเหมาะสม

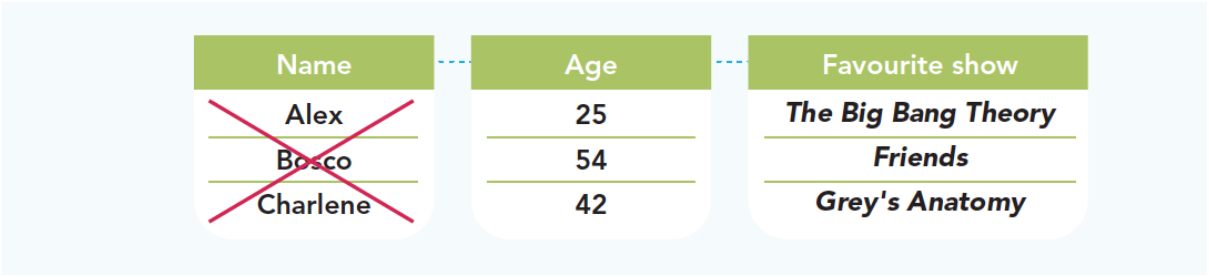


คุณลักษณะของข้อมูลนี้**สำคัญ**กับยูสเคสที่มีการระบุตัวตนหรือไม่	ใช่	คุณลักษณะเป็น**ค่าที่มีลักษณะเฉพาะ**ที่เหมาะสมกับเจ้าของข้อมูลเฉพาะรายเท่านั้นหรือไม่	ไม่ใช่	คุณลักษณะนี้**มีอยู่ในแหล่งข้อมูลสาธารณะหรือแหล่งข้อมูล**ส่วนตัวหรือไม่	ไม่ใช่	เป็นไปได้ที่จะเป็น**คุณลักษณะเป้าหมาย**
ไม่ใช่		ใช่		ใช่		
ลบออก (ทำ Minimization กับข้อมูล)		เป็นไปได้ที่จะเป็น**ตัวระบุโดยตรง**		เป็นไปได้ที่จะเป็น**ตัวระบุโดยอ้อม**		

ขั้นตอนที่ 2 การทำ De-identification กับข้อมูลของคุณ

เหมาะสมกับ:	การแชร์ข้อมูลภายใน (การทำ De-identification กับข้อมูล)	การแชร์ข้อมูลภายใน (การทำ Anonymizationกับข้อมูล) หรือการแชร์ข้อมูลภายนอก
	การเก็บรักษาข้อมูลระยะยาว	ข้อมูลสังเคราะห์

ขั้นตอนนี้เป็นส่วนหนึ่งของกระบวนการทำ Anonymization เสมอ
ขั้นตอนแรก ลบตัวระบุโดยตรงออกทั้งหมด ในตัวอย่างต่อไปนี้ ชื่อทั้งหมดจะถูกลบออก ในกรณีที่ชุดข้อมูลมีตัวระบุโดยตรงอื่นๆ เช่น หมายเลข NRIC และที่อยู่อีเมล ควรลบข้อมูลเหล่านี้ออกด้วย



ชื่อ	อายุ	รายการโปรด
อเล็กซ์		The Big Bang Theory
บอสโค		Friends
ชาร์ลีน		Grey’s Anatomy

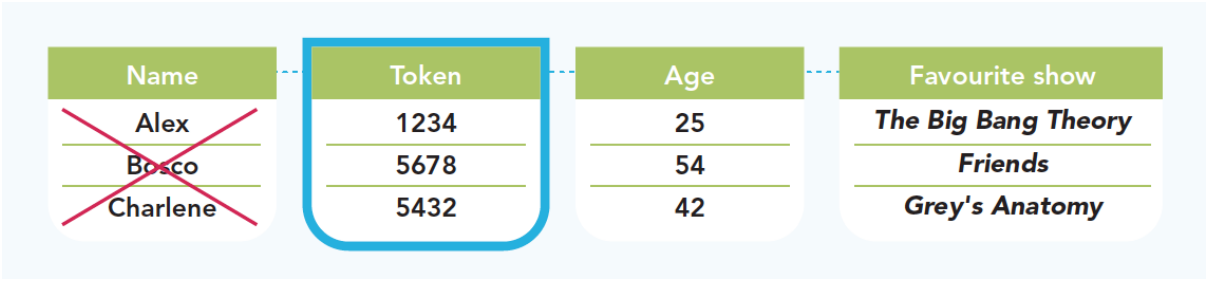
อีกทางเลือกหนึ่ง คือการกำหนดนามแฝงให้กับบันทึกแต่ละรายการ หากจำเป็นต้องเชื่อมโยงบันทึกกลับไปยังบุคคลที่มีลักษณะเฉพาะหรือกับบันทึกต้นฉบับสำหรับยูสเคสต่างๆ เช่น:

ก. การรวมข้อมูล

ข. การวิเคราะห์บันทึกหลายรายการที่เกี่ยวข้องกับบุคคลที่มีลักษณะเฉพาะ หรือ

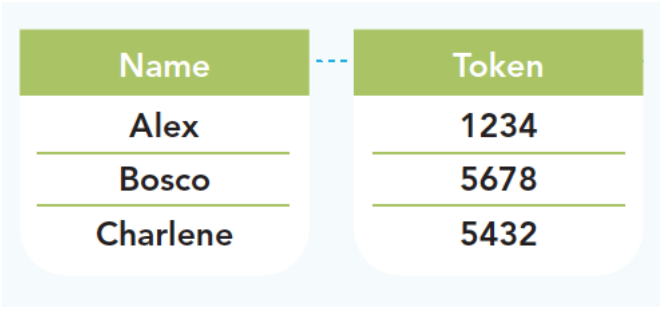
ค. การสร้างชุดข้อมูลสังเคราะห์ที่ต้องใช้ค่าตัวระบุโดยตรงสำหรับการพัฒนาและการทดสอบแอปพลิเคชัน สำหรับยูสเคสนี้ ให้เปลี่ยนตัวระบุโดยตรงที่จำเป็นทั้งหมดเป็นนามแฝง

นามแฝงควรจะมีลักษณะเฉพาะสำหรับตัวระบุโดยตรงที่มีลักษณะเฉพาะแต่ละตัว (ดังภาพด้านล่าง) การกำหนดนามแฝงควรมีความเข้มงวดด้วยเช่นกัน (เช่น บุคคลที่ไม่ได้รับอนุญาตไม่สามารถย้อนกลับได้โดยการเดาหรือคำนวณค่าตัวระบุโดยตรงดั้งเดิมจากนามแฝง)



ชื่อ	โทเคน	อายุ	รายการโปรด
อเล็กซ์			The Big Bang Theory
บอสโค			Friends
ชาร์ลีน			Grey’s Anatomy

หากคุณต้องการเชื่อมโยงบันทึกข้อมูลที่ทำ De-identification กลับไปยังบันทึกต้นฉบับในภายหลัง คุณจำเป็นต้องเก็บการจับคู่ระหว่างตัวระบุโดยตรงกับนามแฝงไว้ ควรเก็บรักษาตารางจับคู่ข้อมูลระบุตัวตน (ตามภาพประกอบด้านล่าง) ไว้อย่างปลอดภัยเนื่องจากสามารถใช้ทำ Re-identification ได้



ชื่อ	โทเคน
อเล็กซ์	
บอสโค	
ชาร์ลีน	

ขั้นตอนที่ 3 ใช้เทคนิคการทำ Anonymization

เหมาะสมกับ:	การแชร์ข้อมูลภายใน (การทำ De-identification กับข้อมูล)	การแชร์ข้อมูลภายใน (การทำ Anonymization กับข้อมูล) หรือการแชร์ข้อมูลภายนอก
	การเก็บรักษาข้อมูลระยะยาว	**ข้อมูลสังเคราะห์**

ในขั้นตอนนี้ คุณจะใช้เทคนิคการทำ Anonymization กับตัวระบุโดยอ้อม เพื่อไม่ให้สามารถนำไปรวมเข้ากับชุดข้อมูลอื่น ๆ ที่อาจมีข้อมูลเพิ่มเติมเพื่อทำ Re-identification ได้ง่าย สำหรับกรณีการใช้ข้อมูลสังเคราะห์ ควรใช้เทคนิคการทำ Anonymization กับคุณลักษณะเป้าหมายด้วย

โปรดทราบว่าการใช้เทคนิคเหล่านี้จะปรับเปลี่ยนค่าข้อมูลและอาจส่งผลต่ออรรถประโยชน์ของข้อมูลที่ผ่านการทำ Anonymization แล้วสำหรับบางยูสเคส (เช่น การวิเคราะห์ข้อมูล) เทคนิคการทำ Anonymization ที่แนะนำด้านล่างนี้คำนึงถึงอรรถประโยชน์ที่เป็นไปได้ที่จำเป็นสำหรับข้อมูลระดับบันทึกในการใช้งานแต่ละกรณี องค์กรสามารถใช้เทคนิคการทำ Anonymization อื่นๆ นอกเหนือจากที่แนะนำได้ หากเกี่ยวข้องกับยูสเคสของตน

ยูสเคส	เทคนิคการทำ Anonymization ที่แนะนำสำหรับข้อมูลระดับบันทึก
การแชร์ข้อมูลภายใน (ข้อมูลที่ผ่านการทำ Anonymization แล้ว) หรือการแชร์ข้อมูลภายนอก	• **การทำ Suppression กับบันทึก:** การลบบันทึก (เช่น แถวข้อมูลโดยเฉพาะอย่างยิ่งเมื่อข้อมูลดังกล่าวอาจมีค่าข้อมูลที่มีลักษณะเฉพาะที่ไม่สามารถทำ Anonymization ได้อีกต่อไป) • **การทำ Suppression กับคุณลักษณะ:** การลบคุณลักษณะของข้อมูล (เช่น คอลัมน์ข้อมูล โดยเฉพาะอย่างยิ่งในกรณีที่ไม่จำเป็นต้องใช้ข้อมูลดังกล่าวในชุดข้อมูล และอาจมีค่าข้อมูลที่มีลักษณะเฉพาะที่ไม่สามารถทำ Anonymization ได้อีกต่อไป) • **การทำ Masking กับอักขระ:** การเปลี่ยนอักขระบางตัวของค่าข้อมูลเป็นสัญลักษณ์ที่สอดคล้องกัน (เช่น * หรือ x) ตัวอย่างเช่น การทำ Masking กับรหัสไปรษณีย์จะต้องเปลี่ยนจาก "235546" เป็น "23xxxx” • **การทำ Generalization:** การลดรายละเอียดของข้อมูล (เช่น การแปลงอายุของบุคคลเป็นช่วงอายุ) เช่น การทำ Generalization กับอายุของบุคคลจาก “26 ปี” เป็น “25-29 ปี”

ยูสเคส	เทคนิคการทำ Anonymization ที่แนะนำสำหรับข้อมูลระดับบันทึก
การเก็บรักษาข้อมูลในระยะยาว	• **การทำ Suppression กับบันทึกหรือคุณลักษณะ** • **การทำ Masking กับอักขระ** • **การทำ Generalization** • **การทำ Perturbation กับข้อมูล:** การปรับเปลี่ยนค่าในข้อมูลโดยการเพิ่ม “สัญญาณรบกวน” ให้กับข้อมูลดั้งเดิม (เช่น +/- ค่าสุ่มให้กับข้อมูล) ระดับของการทำ Perturbation ควรมีสัดส่วนตามช่วงค่าของคุณลักษณะ ตัวอย่างเช่น การทำ Perturbation กับข้อมูลจะเกี่ยวข้องกับการปรับเปลี่ยนข้อมูลเงินเดือนของแต่ละบุคคลจาก “256,654 ดอลลาร์” เป็น “260,000 ดอลลาร์” โดยการปัดเศษข้อมูลขึ้นหรือลงให้เศษไม่ต่ำกว่า 10,000 ดอลลาร์ หรืออีกวิธีหนึ่ง สามารถปรับเปลี่ยนเงินเดือนของแต่ละบุคคลเป็น “250,554 ดอลลาร์” โดยการลบตัวเลขสุ่มจำนวนไม่เกิน 10,000 ดอลลาร์จากค่าเดิม *หมายเหตุ: สามารถทำ Aggregation กับข้อมูลได้สำหรับการใช้งานในกรณีนี้เมื่อไม่จำเป็นต้องใช้ข้อมูลระดับบันทึก (ดูตัวอย่างในภาคผนวก ก)*
ข้อมูลสังเคราะห์	ใช้การทำ Anonymization อย่างละเอียดกับข้อมูลดั้งเดิมเพื่อสร้างข้อมูลสังเคราะห์เพื่อให้คุณลักษณะของข้อมูลทั้งหมด (รวมถึงคุณลักษณะเป้าหมาย) ได้รับการแก้ไขอย่างมีนัยสำคัญ ชุดข้อมูลผลลัพธ์และบันทึกแต่ละรายการที่สร้างขึ้นโดยใช้วิธีการนี้จะไม่มีความคล้ายคลึงกับบันทึกของบุคคลใดๆ และจะไม่มีลักษณะพิเศษของชุดข้อมูลดั้งเดิม เนื่องจากชุดข้อมูลผลลัพธ์ไม่เหมือนกับต้นฉบับ จึงเหมาะสำหรับการพัฒนา/ทดสอบแอปพลิเคชัน แต่ไม่เหมาะสำหรับการฝึกโมเดลปัญญาประดิษฐ์ • **การทำ Perturbation กับข้อมูล** • **การแลกเปลี่ยน:** การจัดเรียงข้อมูลในชุดข้อมูลใหม่แบบสุ่ม โดยที่ค่าคุณลักษณะแต่ละรายการยังคงอยู่ในชุดข้อมูล แต่โดยทั่วไปแล้วจะไม่สอดคล้องกับบันทึกเดิม

โปรดดูภาคผนวก ก สำหรับข้อมูลเพิ่มเติมเกี่ยวกับเทคนิคการทำ Anonymization ต่างๆ และวิธีการนำเทคนิคเหล่านั้นไปใช้ โปรดดูภาคผนวก ข สำหรับเทคนิคการทำ Anonymization ที่แนะนำเพื่อนำไปใช้กับรายการคุณลักษณะของข้อมูลทั่วไป

ขั้นตอนต่อไป: หลังจากใช้เทคนิคการทำ Anonymization ที่เหมาะสมแล้ว ให้ดำเนินการตามขั้นตอนที่ 4 เพื่อประเมินระดับความเสี่ยง ทำขั้นตอนที่ 3 และ 4 ซ้ำจนกว่าคุณจะได้ค่า *k*-anonymity เป็น 3 5 หรือมากกว่า

หมายเหตุ: คุณยังสามารถลบบันทึกหรือคุณลักษณะที่ผิดปกติ (โดยใช้การทำ Suppression กับบันทึกหรือคุณลักษณะ) ที่ "ต้านทาน" ต่อเทคนิคการทำ Anonymization อื่นๆ ที่นำมาใช้ได้ โดยเฉพาะอย่างยิ่งหากมีจำนวนค่าผิดปกติดังกล่าวค่อนข้างต่ำ และการลบออกจะไม่ส่งผลกระทบอย่างมีนัยสำคัญต่อคุณภาพของข้อมูลสำหรับยูสเคสของคุณ

ขั้นตอนที่ 4 คำนวณความเสี่ยงของคุณ

เหมาะสมกับ:	การแชร์ข้อมูลภายใน (การทำ De-identification กับข้อมูล)	การแชร์ข้อมูลภายใน (การทำ Anonymization กับข้อมูล) หรือการแชร์ข้อมูลภายนอก
	การเก็บรักษาข้อมูลระยะยาว	ข้อมูลสังเคราะห์

k-anonymity[9] เป็นวิธีการง่ายๆ [10,11] ในการคำนวณระดับความเสี่ยงในการทำ Re-identification ของชุดข้อมูล โดยพื้นฐานแล้วหมายถึงจำนวนบันทึกที่เหมือนกันจำนวนน้อยที่สุดที่สามารถจัดกลุ่มเข้าด้วยกันในชุดข้อมูลได้ โดยปกติแล้วกลุ่มที่เล็กที่สุดจะถูกนำไปใช้เพื่อแสดงสถานการณ์กรณีที่เลวร้ายที่สุดในการประเมินความเสี่ยงโดยรวมในการทำ Re-identification

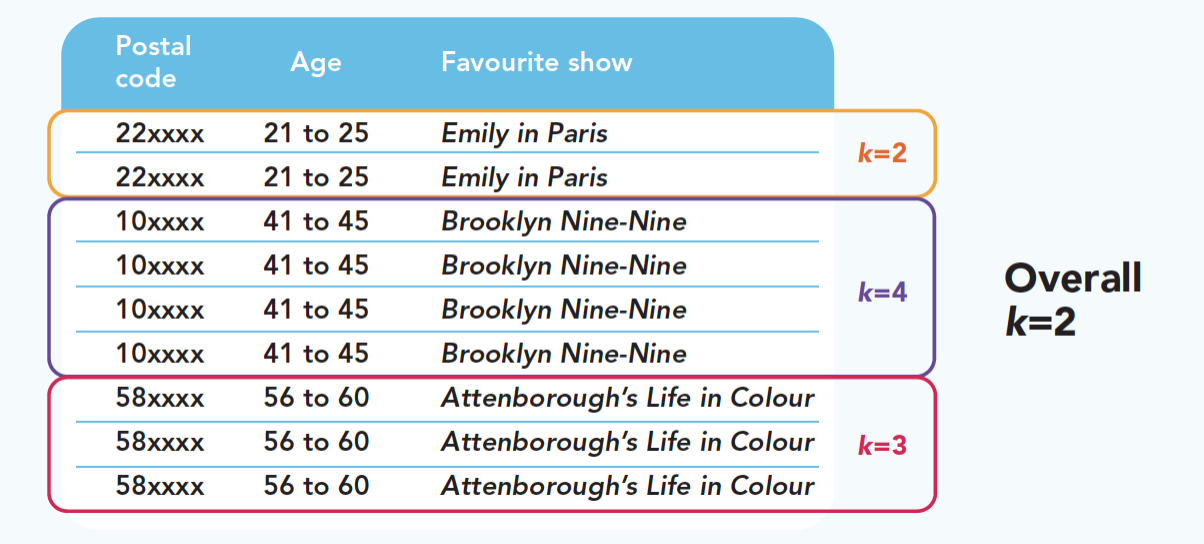
[9] สามารถดูข้อมูลเพิ่มเติมเกี่ยวกับ *k*-anonymity และวิธีใช้ *k*-anonymity เพื่อประเมินความเสี่ยงในการทำ Re-identification ได้ในภาคผนวก ค และภาคผนวก ง

[10] *k*-anonymity อาจไม่เหมาะกับชุดข้อมูลทุกประเภทหรือยูสเคสที่ซับซ้อนอื่นๆ (เช่น ข้อมูลระยะยาวหรือข้อมูลธุรกรรมที่อาจมีตัวระบุโดยอ้อมเดียวกันในบันทึกหลายรายการ) อัลกอริธึมการตรวจจับเฉพาะพิเศษ (SUDA) และ μ-Argus เป็นวิธี/เครื่องมืออื่นๆ สำหรับประเมินความเสี่ยงของชุดข้อมูลที่ใช้ร่วมกัน

[11] ข้อจำกัดของการใช้ *k*-anonymity คือการเปิดเผยคุณลักษณะโดยการโจมตีความเป็นแบบเดียวกัน (Homogeneity Attack) ซึ่งสามารถแก้ไขได้โดยใช้ส่วนขยาย *k*-anonymity *l*-diversity และ *t*-closeness หัวข้อเหล่านี้ไม่ได้กล่าวถึงในคู่มือฉบับนี้

ของชุดข้อมูล ค่า *k*-anonymity เท่ากับ 1 หมายความว่าบันทึกมีลักษณะเฉพาะ โดยทั่วไป เฉพาะตัวระบุโดยอ้อมเท่านั้นที่จะถูกนำมาคำนวณค่า *k*-anonymity[12]

ค่า *k*-anonymity ที่สูงกว่าหมายความว่ามีความเสี่ยงในการทำ Re-identification ต่ำกว่า ในขณะที่ค่า *k*-anonymity ที่ต่ำกว่าหมายถึงความเสี่ยงที่สูงกว่า โดยทั่วไปเกณฑ์ทางอุตสาหกรรมสำหรับค่า *k*-anonymity จะอยู่ที่ 3 หรือ 5[13] หากเป็นไปได้ควรตั้งค่าเกณฑ์ *k*-anonymity ให้สูงขึ้นเพื่อลดความเสี่ยงในการทำ Re-identification



ดูบทที่ 3 (การทำ Anonymization) ของ*แนวทางคำแนะนำ*ของ PDPC *เกี่ยวกับพระราชบัญญัติคุ้มครองข้อมูลส่วนบุคคลสำหรับหัวข้อที่เลือก*เกี่ยวกับเกณฑ์การพิจารณาว่าข้อมูลนั้นเป็นข้อมูลที่ผ่านการทำ Anonymization อย่างเพียงพอหรือไม่

รหัสไปรษณีย์	อายุ	รายการโปรด		
	21 ถึง 25	Emily in Paris	k=2	โดยรวม k=2
	41 ถึง 45	Brooklyn Nine-Nine	k=4	
	56 ถึง 60	Attenborough's Life in Colour	k=3	

[12] ควรลบตัวระบุโดยตรงออกในขั้นตอนที่ 2 และไม่ควรรวมนามแฝงในการคำนวณ มิฉะนั้นบันทึกทุกรายการจะมีลักษณะเฉพาะ

[13] อ้างอิงจาก*วิธีการตัดสินใจเกี่ยวกับการทำ De-identification* โดยสำนักงานกรรมาธิการข้อมูลแห่งออสเตรเลีย, CSIRO และข้อมูล 61

แผนภาพด้านบนแสดงชุดข้อมูลที่มีบันทึกที่เหมือนกันสามกลุ่ม ค่า *k* ของแต่ละกลุ่มมีตั้งแต่ 2 ถึง 4 โดยรวมแล้วค่า *k*-anonymity ของชุดข้อมูลคือ 2 ซึ่งสะท้อนถึงค่าต่ำสุด (ความเสี่ยงสูงสุด) ภายในชุดข้อมูลทั้งหมด[14]

ขั้นตอนต่อไป: หากถึงเกณฑ์ค่า *k*-anonymity ให้ทำตามขั้นตอนที่ 5 หากค่า *k*-anonymity ต่ำกว่าเกณฑ์ที่ตั้งไว้ให้กลับไปที่ขั้นตอนที่ 3 และทำซ้ำ

หมายเหตุ: หากเป็นไปได้คุณควรตั้งค่า k-anonymity ให้สูงขึ้น (เช่น 5 หรือมากกว่า) สำหรับการแชร์ข้อมูลภายนอก ในขณะที่สามารถใช้ค่าที่ต่ำกว่า (เช่น 3) สำหรับการแชร์ข้อมูลภายในหรือการเก็บรักษาข้อมูลในระยะยาวได้ อย่างไรก็ตามหากคุณไม่สามารถทำ Anonymization กับข้อมูลของคุณเพิ่มเติมเพื่อให้บรรลุเป้าหมายดังกล่าวได้ คุณควรใช้มาตรการป้องกันที่เข้มงวดมากขึ้นเพื่อให้แน่ใจว่าข้อมูลที่ผ่านการทำ Anonymization แล้วจะไม่ถูกเปิดเผยต่อบุคคลที่ไม่ได้รับอนุญาต และความเสี่ยงในการทำ Re-identification จะลดลง หรือคุณควรว่าจ้างผู้เชี่ยวชาญเพื่อเสนอวิธีการประเมินทางเลือกเพื่อให้ได้ความเสี่ยงในการทำ Re-identification ที่เทียบเท่ากัน

ขั้นตอนที่ 5 จัดการความเสี่ยงในการทำ Re-identification และการเปิดเผยข้อมูลของคุณ

เหมาะสมกับ:	การแชร์ข้อมูลภายใน (การทำ De-identification กับข้อมูล)	การแชร์ข้อมูลภายใน (การทำ Anonymization กับข้อมูล) หรือการแชร์ข้อมูลภายนอก
	การเก็บรักษาข้อมูลระยะยาว	ข้อมูลสังเคราะห์

โดยทั่วไปแล้วควรใช้มาตรการที่เหมาะสมเพื่อปกป้องข้อมูลของคุณจากความเสี่ยงในการทำ Re-identification และการเปิดเผยข้อมูล นี่เป็นมุมมองของความก้าวหน้าทางเทคโนโลยีในอนาคต และชุดข้อมูลที่ไม่รู้จักที่สามารถนำไปใช้เพื่อจับคู่กับชุดข้อมูลที่ผ่านการทำ

[14] คู่มือฉบับนี้ใช้วิธีที่รอบคอบในการพิจารณาความเสี่ยงสูงสุด นอกจากนี้ยังมีวิธีอื่นๆ (เช่น ความเสี่ยงโดยเฉลี่ย และความเสี่ยงโดยเฉลี่ยที่เข้มงวด)

Anonymization ของคุณ และช่วยให้การทำ Re-identification ง่ายกว่าที่คาดไว้ในขณะที่ทำ Anonymization
ตามแนวทางปฏิบัติที่ดี ควรบันทึกรายละเอียดของกระบวนการทำ Anonymization ตัวแปรที่ใช้ และการควบคุมไว้อย่างชัดเจนเพื่อใช้อ้างอิงในอนาคต เอกสารดังกล่าวอำนวยความสะดวกในการทบทวน บำรุงรักษา ปรับแต่ง และการตรวจสอบ โปรดทราบว่าควรเก็บเอกสารดังกล่าวไว้อย่างปลอดภัย เนื่องจากการเผยแพร่ตัวแปรอาจเอื้อต่อการทำ Re-identification และการเปิดเผยข้อมูลที่ผ่านการทำ Anonymization ได้

การทำ Re-identification และความเสี่ยงในการเปิดเผยข้อมูลมีหลายประเภท ข้อมูลต่อไปนี้จะอธิบายปัจจัยพื้นฐานบางประการที่คุณควรประเมินเมื่อทบทวนความเพียงพอของมาตรการป้องกันที่นำมาใช้

1	**การทำ Re-identification (การเปิดเผยตัวตน)**
	การระบุตัวตนของบุคคลที่อธิบายโดยบันทึกเฉพาะด้วยความมั่นใจในระดับสูง การดำเนินการนี้อาจเกิดขึ้นจากสถานการณ์ต่างๆ เช่น การทำ Anonymization ไม่เพียงพอ การทำ Re-identification โดยการเชื่อมโยงหรือการย้อนกลับนามแฝง (Pseudonym Reversal) ตัวอย่างเช่น กระบวนการทำ Anonymization ซึ่งสร้างนามแฝงโดยใช้อัลกอริธึมที่คาดเดาได้ง่ายและย้อนกลับได้ เช่น การเปลี่ยน “1” เป็น “ก” “2” เป็น “ข” เป็นต้น
2	**การเปิดเผยคุณลักษณะ**
	การกำหนดด้วยความมั่นใจในระดับสูงว่าคุณลักษณะที่อธิบายไว้ในชุดข้อมูลเป็นของบุคคลใดบุคคลหนึ่ง แม้ว่าจะไม่สามารถแยกแยะบันทึกของบุคคลนั้นได้ก็ตาม ตัวอย่างเช่น ชุดข้อมูลที่มีบันทึกของลูกค้าที่ผ่านการทำ Anonymization ของศัลยแพทย์ด้านความงามรายหนึ่งซึ่งเผยให้เห็นว่าลูกค้าของเขาทุกคนที่มีอายุต่ำกว่า 30 ปีได้ผ่านขั้นตอนหนึ่งมาแล้ว หากทราบว่าบุคคลใดบุคคลหนึ่งมีอายุ 28 ปีและเป็นลูกค้าของศัลยแพทย์รายนี้ เราจะทราบว่าบุคคลดังกล่าวได้ผ่านขั้นตอนดังกล่าวมาแล้ว แม้ว่าบันทึกของบุคคลนั้นจะไม่สามารถแยกแยะจากบุคคลอื่นในชุดข้อมูลที่ผ่านการทำ Anonymization ได้ก็ตาม
3	**การเปิดเผยข้อมูลอนุมาน**
	การทำการอนุมานเกี่ยวกับแต่ละบุคคลด้วยความมั่นใจในระดับสูง แม้ว่าเขาหรือเธอจะไม่ได้อยู่ในชุดข้อมูลตามคุณสมบัติทางสถิติของชุดข้อมูลก็ตาม ตัวอย่างเช่น หากชุดข้อมูลที่เผยแพร่โดยนักวิจัยทางการแพทย์พบว่า 70% ของบุคคลที่อายุเกิน 75 ปีมีอาการป่วยบางอย่าง ข้อมูลนี้อาจถูกอนุมานได้เกี่ยวกับบุคคลที่ไม่อยู่ในชุดข้อมูล

โดยทั่วไป เทคนิคการทำ Anonymization แบบดั้งเดิมส่วนใหญ่มีจุดมุ่งหมายเพื่อป้องกันการทำ Re-identification แต่อาจไม่ป้องกันความเสี่ยงในการเปิดเผยข้อมูลประเภทอื่น

ตารางต่อไปนี้จะอธิบายว่าเมื่อใดควรใช้มาตรการป้องกันการทำ Re-identification และความเสี่ยงในการเปิดเผยข้อมูล ชุดมาตรการป้องกันขั้นพื้นฐาน (การควบคุมด้านเทคนิค กระบวนการ และกฎหมาย) สำหรับยูสเคสมีระบุไว้ในย่อหน้าต่อไปนี้

ยูสเคส	คุณจำเป็นต้องจัดการความเสี่ยงในการทำ Re-identification และการเปิดเผยข้อมูลสำหรับชุดข้อมูลที่ผ่านการทำ De-identification หรือการทำ Anonymization หรือไม่?
การแชร์ข้อมูลภายใน (ข้อมูลที่ผ่านการทำ De-identification)	เนื่องจากมีการใช้การทำ De-identification เพียงอย่างเดียวเพื่อรักษาอรรถประโยชน์ของข้อมูลในระดับสูง จึงมีความเสี่ยงที่สูงขึ้นที่จะมีการทำ Re-identification และการเปิดเผยข้อมูลที่ผ่านการทำ De-identification ดังนั้น จึงจำเป็นต้องมีการป้องกันสำหรับชุดข้อมูลที่ผ่านการทำ De-identification ควรรักษาความปลอดภัยของตารางจับคู่ข้อมูลระบุตัวตน (ถ้ามี) ในกรณีที่มีการละเมิดข้อมูล การใช้เทคนิคการทำ De-identification วิธีป้องกันชุดข้อมูลที่ผ่านการทำ De-identification และวิธีรักษาความปลอดภัยของตารางจับคู่จะถือว่าเป็นส่วนหนึ่งของกลไกการป้องกันที่นำไปใช้
การแชร์ข้อมูลภายใน (ข้อมูลที่ผ่านการทำ Anonymization)	เพื่อลดความเสี่ยงในการทำ Re-identification และการเปิดเผยข้อมูล หากจำเป็น ควรใช้การทำ Anonymization กับข้อมูลเพื่อการแชร์ภายใน ในกรณีต่อไปนี้ ได้แก่ (ก) ในกรณีที่ไม่จำเป็นต้องใช้ข้อมูลส่วนบุคคลโดยละเอียด (ข) ในกรณีที่ข้อมูลที่ละเอียดอ่อนอาจถูกแชร์ หรือ (ค) ในกรณีที่มีการแชร์ชุดข้อมูลขนาดใหญ่กับหลายแผนก จำเป็นต้องมีการป้องกันขั้นพื้นฐานสำหรับชุดข้อมูลที่ผ่านการทำ Anonymization ควรรักษาความปลอดภัยของตารางจับคู่ข้อมูลระบุตัวตน (ถ้ามี) และไม่ควรแชร์กับแผนกภายในอื่นๆ
การแชร์ข้อมูลภายนอก	จำเป็นต้องมีการป้องกันขั้นพื้นฐานสำหรับชุดข้อมูลที่ผ่านการทำ Anonymization ควรรักษาความปลอดภัยของตารางจับคู่ข้อมูลระบุตัวตน (ถ้ามี) และไม่ควรแชร์ภายนอก
การเก็บรักษาข้อมูลในระยะยาว	จำเป็นต้องมีการป้องกันขั้นพื้นฐานสำหรับชุดข้อมูลที่ผ่านการทำ Anonymization ควรทำลายตารางจับคู่ข้อมูลระบุตัวตนทั้งหมดอย่างปลอดภัย

สำหรับกรณีการใช้ข้อมูลสังเคราะห์ ความเสี่ยงในการทำ Re-identification จะถือว่ามีน้อยที่สุดเมื่อมีการใช้การทำ Anonymization อย่างละเอียดกับตัวระบุโดยอ้อมและคุณลักษณะเป้าหมายทั้งหมดจนบันทึกไม่เหมือนกับชุดข้อมูลดั้งเดิม ด้วยเหตุนี้จึงไม่จำเป็นต้องมีการปกป้องชุดข้อมูลนี้เพิ่มเติม

การควบคุมทางเทคนิคและกระบวนการ: คุณควรใช้มาตรการป้องกันทางเทคนิคเพื่อจัดการการทำ Re-identification และความเสี่ยงในการเปิดเผยข้อมูลที่ผ่านการทำ De-identification หรือการทำ Anonymization ตารางต่อไปนี้แสดงตัวอย่างแนวทางปฏิบัติที่ดีที่แนะนำ

คุณควรทบทวนแนวทางปฏิบัติที่ดีเหล่านี้ว่าเพียงพอที่จะปกป้องข้อมูลที่ผ่านการทำ De-identification/Anonymization ของคุณหรือไม่ โดยพิจารณาจากระดับของการทำ Anonymization ที่ใช้ ความละเอียดอ่อนของข้อมูลที่ผ่านการทำ De-identification/ Anonymization และยูสเคส คุณสามารถดู*แนวทางปฏิบัติในการปกป้องข้อมูลสำหรับระบบ ICT* ของ PDPCเพื่อดูมาตรการป้องกันเพิ่มเติมตามที่เกี่ยวข้อง

ในตาราง "*Y*" หมายความว่าแนะนำให้คุณใช้การควบคุมทางเทคนิคที่เกี่ยวข้อง และ "*NA*" หมายความว่าการควบคุมทางเทคนิคนั้นใช้ไม่ได้กับยูสเคสนั้น

การควบคุมทางเทคนิค		**การแชร์ข้อมูลภายใน (ข้อมูลที่ผ่านการทำ De-identification)**	**การแชร์ข้อมูลภายใน (ข้อมูลที่ผ่านการทำ Anonymization)**	**การแชร์ข้อมูลภายนอก**	**การเก็บรักษาข้อมูลในระยะยาว**
การควบคุมการเข้าถึงและรหัสผ่าน	ใช้การควบคุมการเข้าถึงในระดับแอปพลิเคชันเพื่อจำกัดการเข้าถึงข้อมูลในระดับผู้ใช้ ระดับความซับซ้อนต่ำสุดของรหัสผ่าน (เช่น อักขระตัวอักษรและตัวเลขอย่างน้อย 12 ตัว โดยผสมระหว่างตัวพิมพ์ใหญ่ ตัวพิมพ์เล็ก ตัวเลข และอักขระพิเศษ)	Y	Y	Y	Y
	ทบทวนบัญชีผู้ใช้เป็นประจำเพื่อให้แน่ใจว่าบัญชีทั้งหมดใช้งานได้และสิทธิ์ที่ได้รับมอบหมายมีความจำเป็น (เช่น ลบบัญชีผู้ใช้เมื่อผู้ใช้ออกจากองค์กรไปแล้ว หรืออัปเดตสิทธิ์ของผู้ใช้เมื่อเขาหรือเธอเปลี่ยนบทบาทของตนภายในองค์กรแล้ว)	Y	Y	Y	Y
ความปลอดภัยสำหรับอุปกรณ์	ป้องกันคอมพิวเตอร์โดยใช้ฟังก์ชันรหัสผ่าน ตัวอย่างของวิธีเหล่านี้รวมถึงการป้อนรหัสผ่านระหว่างการบูตเครื่องการขอให้ล็อกอินเพื่อเข้าสู่ระบบ	Y	Y	Y	Y

การควบคุมทางเทคนิค		การแชร์ข้อมูลภายใน (ข้อมูลที่ผ่านการทำ De-identification)	การแชร์ข้อมูลภายใน (ข้อมูลที่ผ่านการทำ Anonymization)	การแชร์ข้อมูลภายนอก	การเก็บรักษาข้อมูลในระยะยาว
จัดเก็บข้อมูล/ฐานข้อมูล	ปฏิบัติการ การล็อคหน้าจอหลังจากไม่มีการใช้งานช่วงระยะเวลาหนึ่ง เป็นต้น				

การควบคุมทางเทคนิค		การแชร์ข้อมูลภายใน (ข้อมูลที่ผ่านการทำ De-identification)	การแชร์ข้อมูลภายใน (ข้อมูลที่ผ่านการทำ Anonymization)	การแชร์ข้อมูลภายนอก	การเก็บรักษาข้อมูลในระยะยาว
ความปลอดภัยสำหรับอุปกรณ์จัดเก็บข้อมูล/ฐานข้อมูล	เข้ารหัสชุดข้อมูล ทบทวนวิธีการเข้ารหัส (เช่น อัลกอริธึมและความยาวของคีย์) เป็นระยะๆ เพื่อให้มั่นใจว่าวิธีดังกล่าวมีความเกี่ยวข้องและปลอดภัยที่เป็นที่ยอมรับตามมาตรฐานของอุตสาหกรรม	Y *(ในกรณีที่ข้อมูลที่เกี่ยวข้องมีความละเอียดอ่อนหรือมีการแชร์ชุดข้อมูลขนาดใหญ่กับหลายแผนกแต่ไม่มีการทำ Anonymization กับชุดข้อมูล)*	NA	NA *(ในกรณีที่มีการประเมินความเสี่ยงในการทำ Re-identification ว่าต่ำ (เช่น k-anonymity มีค่าตั้งแต่ 5 ขึ้นไป) ไม่จำเป็นต้องใช้การเข้ารหัสกับชุดข้อมูลที่ผ่านการทำ Anonymization)*	NA
	เข้ารหัสตารางจับคู่ข้อมูลระบุตัวตน ควรเก็บรักษาตารางจับคู่ข้อมูลระบุตัวตนให้ปลอดภัยและไม่ควรแชร์ในการใช้งานทุกกรณี	Y	Y	Y	NA *(ควรลบตารางจับคู่ข้อมูลระบุตัวตนออก)*

การควบคุมทางเทคนิค		การแชร์ข้อมูลภายใน (ข้อมูลที่ผ่านการทำ De-identification)	การแชร์ข้อมูลภายใน (ข้อมูลที่ผ่านการทำ Anonymization)	การแชร์ข้อมูลภายนอก	การเก็บรักษาข้อมูลในระยะยาว
	แจ้งคีย์ถอดรหัสของชุดข้อมูลแยกต่างหากให้ผู้รับเป้าหมายของข้อมูลที่แชร์/ส่งออกทราบ	Y	NA	NA	NA

การควบคุมกระบวนการ		การแชร์ข้อมูลภายใน (ข้อมูลที่ผ่านการทำ De-identification)	การแชร์ข้อมูลภายใน (ข้อมูลที่ผ่านการทำ Anonymization)	การแชร์ข้อมูลภายนอก	การเก็บรักษาข้อมูลในระยะยาว
การจัดการเหตุการณ์	พัฒนาแผนการจัดการการละเมิดข้อมูลเพื่อตอบสนองต่อการละเมิดข้อมูลและจัดการการสูญเสียชุดข้อมูลได้อย่างมีประสิทธิภาพมากขึ้น แผนควรรวมถึงวิธีการจัดการการสูญเสียตารางจับคู่ข้อมูลระบุตัวตนหรือข้อมูลที่อาจทำให้การย้อนการทำ De-identification/การทำ Anonymization กับข้อมูลกลับสู่รูปแบบเดิม ส่งผลให้สามารถทำ Re-identification เพื่อระบุข้อมูลที่สูญหายได้ โปรดดูข้อมูลเพิ่มเติมด้านล่างเกี่ยวกับการจัดการเหตุการณ์	Y	Y	Y	Y
	เก็บทะเบียนข้อมูลส่วนกลางของข้อมูลที่ผ่านการทำ De-identification/Anonymization ที่แชร์ไว้ทั้งหมด เพื่อให้แน่ใจว่าข้อมูลที่แชร์ที่รวมกันจะไม่ส่งผลให้มีการทำ Re-identification กับข้อมูลที่ผ่านการทำ De-identification/ Anonymization	Y	Y	y	NA

การควบคุมกระบวนการ		การแชร์ข้อมูลภายใน (ข้อมูลที่ผ่านการทำ De-identification)	การแชร์ข้อมูลภายใน (ข้อมูลที่ผ่านการทำ Anonymization)	การแชร์ข้อมูลภายนอก	การเก็บรักษาข้อมูลในระยะยาว
การควบคุมการกำกับดูแลภายใน	ดำเนินการทบทวนการทำ Re-identification ของข้อมูลที่ผ่านการทำ De - identification /Anonymization	Y	Y	Y	y
	ตรวจสอบให้แน่ใจว่าผู้รับ (บุคคลหรือแผนก) และวัตถุประสงค์ของการใช้ข้อมูลที่ผ่านการทำ De-identification/Anonymization ได้รับการอนุมัติจากหน่วยงานที่เกี่ยวข้องภายในองค์กร	Y	y	NA	NA
	ห้ามผู้รับที่ได้รับอนุญาต (บุคคลหรือแผนก) แชร์ข้อมูล Anonymization ให้กับบุคคลที่ไม่ได้รับอนุญาต หรือพยายามที่จะทำ Re-identification โดยไม่ได้รับการอนุมัติจากหน่วยงานที่เกี่ยวข้องภายในองค์กร	Y	Y	NA	NA
	กำจัดข้อมูลที่ผ่านการทำ De-identification/Anonymization ภายในองค์กรเป็นประจำ เมื่อบรรลุวัตถุประสงค์และไม่จำเป็นต้องใช้ข้อมูลนั้นอีกต่อไป	Y	Y	NA	NA
	ดำเนินการตรวจสอบภายในเป็นระยะๆ เพื่อให้มั่นใจว่าสอดคล้องกับกระบวนการต่างๆ	Y	Y	Y	Y

การจัดการเหตุการณ์: องค์กรควรระบุความเสี่ยงของการละเมิดข้อมูลที่เกี่ยวข้องกับตารางจับคู่ข้อมูลระบุตัวตน ข้อมูลที่ผ่านการทำ De-identification และข้อมูลที่ผ่านการทำ Anonymization และรวมสถานการณ์ที่เกี่ยวข้องไว้ในแผนการจัดการเหตุการณ์ ข้อควรพิจารณาต่อไปนี้อาจเกี่ยวข้องกับการรายงานการละเมิดข้อมูลและการสืบสวนภายใน:

การสูญเสียข้อมูลที่ผ่านการทำ De-identification และตารางจับคู่ข้อมูลระบุตัวตน

การละเมิดทั้งข้อมูลที่ผ่านการทำ De-identification และตารางการจัดการข้อมูลระบุตัวตนจะคล้ายกับการละเมิดข้อมูลส่วนบุคคล ในกรณีดังกล่าว องค์กรต้องประเมินว่าต้องแจ้งการละเมิดข้อมูลและแจ้งให้บุคคลที่ได้รับผลกระทบและ/หรือคณะกรรมการทราบหรือไม่ ซึ่งประเมินว่าต้องแจ้งได้ภายใต้หน้าที่การแจ้งเตือนการละเมิดข้อมูล

การสูญเสียข้อมูลที่ผ่านการทำ De-identification เท่านั้น

จำเป็นต้องมีการประเมินหากข้อมูลที่ผ่านการทำ De-identification ถูกละเมิดจากภายนอก องค์กรต้องประเมินว่าต้องแจ้งการละเมิดข้อมูลหรือไม่ เนื่องจากข้อมูลที่ผ่านการทำ De-identification มีความเสี่ยงสูงที่จะถูกทำ Re-identification อย่างไรก็ตาม การใช้การทำ De-identification และการป้องกันอื่นๆ เพื่อปกป้องข้อมูลและตารางจับคู่ข้อมูลระบุตัวตนอาจถือได้ว่าเป็นส่วนหนึ่งของกลไกการป้องกันที่องค์กรนำไปใช้

การสูญเสียข้อมูลที่ผ่านการทำ Anonymization และตารางจับคู่ข้อมูลระบุตัวตน

องค์กรต้องประเมินความเสี่ยงในการทำ Re-identification ในกรณีที่พบว่ามีความเสี่ยงสูง องค์กรต้องประเมินว่าต้องแจ้งการละเมิดข้อมูลและแจ้งให้บุคคลที่ได้รับผลกระทบและ/หรือคณะกรรมการทราบหรือไม่ภายใต้หน้าที่การแจ้งเตือนการละเมิดข้อมูลจำเป็นต้องมีการประเมินหากข้อมูลที่ผ่านการทำ De-identification แล้วถูกละเมิดจากภายนอก องค์กรต้องประเมินว่าต้องแจ้งการละเมิดข้อมูลหรือไม่ เนื่องจากข้อมูลที่ผ่านการทำ De-identification แล้วมีความเสี่ยงสูงในการทำ Re-identification อย่างไรก็ตาม การใช้การทำ De-identification และการป้องกันอื่นๆ เพื่อปกป้องข้อมูลและตารางจับคู่ข้อมูลระบุตัวตนอาจถือได้ว่าเป็นส่วนหนึ่งของกลไกการป้องกันที่องค์กรนำไปใช้

เฉพาะการสูญเสียข้อมูลที่ผ่านการทำ Anonymization แล้วเท่านั้น

ในกรณีที่องค์กรใช้เทคนิคการทำ Anonymization อย่างเหมาะสม องค์กรก็ไม่จำเป็นต้องรายงานการละเมิดว่าเป็นการละเมิดที่ต้องแจ้งให้ทราบ อย่างไรก็ตามควรดำเนินการตรวจสอบเหตุการณ์ดังกล่าวต่อไปเพื่อหาสาเหตุเพื่อปรับปรุงการป้องกันภายในจากเหตุการณ์การละเมิดข้อมูลในอนาคต

เฉพาะการสูญเสียการจับคู่ข้อมูลระบุตัวตนเท่านั้น

หากชุดข้อมูลที่มีการใช้ตารางจับคู่ข้อมูลระบุตัวตนยังคงได้รับการปกป้อง องค์กรต่างๆ ไม่จำเป็นต้องรายงานการละเมิด เนื่องจากตารางจับคู่ข้อมูลระบุตัวตนเพียงลำพังไม่ใช่ข้อมูลส่วนบุคคล อย่างไรก็ตามองค์กรควรสร้างนามแฝงใหม่สำหรับชุดข้อมูลและตารางจับคู่ข้อมูลระบุตัวตนใหม่ทันที ควรดำเนินการตรวจสอบเหตุการณ์ดังกล่าวต่อไปเพื่อหาสาเหตุเพื่อปรับปรุงการป้องกันภายในจากเหตุการณ์การละเมิดข้อมูลในอนาคต

การควบคุมทางกฎหมาย: องค์กรควรปกป้องตนเองโดยตรวจสอบให้แน่ใจว่าผู้รับข้อมูลที่ผ่านการทำ Anonymization แล้วซึ่งเป็นบุคคลที่สามมีการป้องกันที่เกี่ยวข้องกับข้อมูลที่ผ่านการทำ Anonymization แล้วที่แชร์ไว้เพื่อลดความเสี่ยงในการทำ Re-identification แนวทางปฏิบัติที่ดีในตารางต่อไปนี้นำมาจาก*วิธีการแชร์ข้อมูลที่เชื่อถือได้*ของ PDPC

การควบคุมทางกฎหมาย		การแชร์ข้อมูลภายใน (ข้อมูลที่ผ่านการทำ De-identification แล้ว)	การแชร์ข้อมูลภายใน (ข้อมูลที่ผ่านการทำ Anonymization แล้ว)	การแชร์ข้อมูลภายนอก	การเก็บรักษาข้อมูลในระยะยาว
ข้อตกลงการแชร์ข้อมูล	ตรวจสอบให้แน่ใจว่าข้อมูลถูกนำไปใช้เพื่อวัตถุประสงค์ที่ได้รับอนุญาตเท่านั้น (เช่น ไม่มีการเปิดเผยต่อบุคคลที่ไม่ได้รับอนุญาต) และกำหนดความรับผิดในกรณีที่มีการละเมิดสัญญา	NA	NA	Y	NA
	ห้ามมิให้ผู้รับบุคคลที่สามพยายามทำ Re-identification กับชุดข้อมูลที่ผ่านการทำ Anonymization แล้วที่ถูกแชร์	NA	NA	y	NA
	ตรวจสอบให้แน่ใจว่าผู้รับบุคคลที่สามปฏิบัติตามการป้องกันที่เกี่ยวข้องกับข้อมูลที่ผ่านการทำ Anonymization แล้วที่แชร์ตามการควบคุมภายในขององค์กร	NA	NA	Y	NA

ภาคผนวก

ภาคผนวก ก: เทคนิคการจัดทำข้อมูลนิรนาม (Anonymization) ขั้นพื้นฐาน

การทำ Suppression กับบันทึก	
คำอธิบาย	การทำ Suppression กับบันทึกหมายถึงการลบบันทึกทั้งหมดในชุดข้อมูล ซึ่งตรงกันข้ามกับเทคนิคอื่นๆ ส่วนใหญ่ เทคนิคนี้ส่งผลต่อคุณลักษณะหลายรายการในเวลาเดียวกัน
ควรใช้วิธีนี้เมื่อใด	ใช้การทำ Suppression กับบันทึกเพื่อลบบันทึกที่ผิดปกติซึ่งมีลักษณะเฉพาะหรือไม่ตรงตามเกณฑ์อื่นๆ เช่น *k*-anonymity ออกจากชุดข้อมูลที่ผ่านการทำ Anonymization แล้ว เนื่องจากค่าผิดปกติอาจทำให้มีการทำ Re-identification ได้ง่าย สามารถนำเทคนิคนี้ไปใช้ก่อนหรือหลังการใช้เทคนิคอื่นๆ (เช่น การทำ Generalization) ได้
ควรใช้วิธีนี้อย่างไร	ลบบันทึกทั้งหมด โปรดทราบว่าการทำ Suppression ควรทำแบบถาวรและไม่ใช่แค่ฟังก์ชัน "ซ่อนแถว"[15] อย่างไรก็ตาม "การแก้ไข" อาจไม่เพียงพอหากยังคงสามารถเข้าถึงข้อมูลพื้นฐานได้
คำแนะนำอื่นๆ	- โปรดดูตัวอย่างการทำ Generalization เพื่อดูการทำ Suppression กับบันทึก - โปรดทราบว่าการลบบันทึกอาจส่งผลต่อชุดข้อมูลได้ (เช่น ในแง่ของสถิติ เช่น ค่าเฉลี่ยและค่ามัธยฐาน)

การทำ Masking กับอักขระ	
คำอธิบาย	การทำ Masking กับอักขระหมายถึงการเปลี่ยนอักขระของค่าข้อมูล ซึ่งสามารถทำได้โดยใช้สัญลักษณ์ที่สอดคล้องกัน (เช่น “*” หรือ “x”) โดยทั่วไปการทำ Masking จะใช้กับอักขระบางตัวในคุณลักษณะเท่านั้น
ควรใช้วิธีนี้เมื่อใด	การทำ Masking กับอักขระควรใช้เมื่อค่าข้อมูลเป็นอักขระเรียงกันและการซ่อนบางส่วนของข้อมูลนั้นเพียงพอสำหรับการทำ Anonymization ที่ต้องการ

[15] นี่หมายถึงการใช้ฟังก์ชัน "ซ่อนแถว" ในซอฟต์แวร์สเปรดชีตของคุณ

การทำ Masking กับอักขระ	
ควรใช้วิธีนี้อย่างไร	เปลี่ยนอักขระที่เหมาะสมด้วยสัญลักษณ์ที่เลือกตามลักษณะของคุณลักษณะ คุณสามารถเปลี่ยนจำนวนอักขระคงที่ (เช่น หมายเลขบัตรเครดิต) หรือเปลี่ยนจำนวนอักขระ (เช่น ที่อยู่อีเมล) ทั้งนี้ขึ้นอยู่กับประเภทของคุณลักษณะ
คำแนะนำอื่นๆ	• โปรดทราบว่าการทำ Masking อาจต้องคำนึงว่าความยาวของข้อมูลดั้งเดิมบอกอะไรเกี่ยวกับข้อมูลดั้งเดิมหรือไม่ ความรู้ในเนื้อหาเป็นสิ่งสำคัญ โดยเฉพาะอย่างยิ่งในการทำ Masking บางส่วนเพื่อให้แน่ใจว่าได้ทำ Masking กับอักขระที่ถูกต้องแล้ว สามารถใช้ผลรวมตรวจสอบ (Checksum) ภายในข้อมูลได้ บางครั้งสามารถใช้ Checksum เพื่อกู้คืน (ส่วนอื่น ๆ) ของข้อมูลที่ผ่านการทำ Masking แล้วได้ หากต้องการทำ Masking อย่างสมบูรณ์ สามารถทำ Suppression กับคุณลักษณะได้ เว้นแต่ความยาวของข้อมูลบอกอะไรบางอย่าง • สถานการณ์ของข้อมูลที่ผ่านการทำ Masking แล้วที่เจ้าของข้อมูลต้องการให้รับรู้ข้อมูลของตนเองนั้นเป็นสถานการณ์พิเศษ และไม่ได้อยู่ในวัตถุประสงค์ปกติของการทำ Anonymization กับข้อมูล เช่น การเผยแพร่ผลการจับรางวัล ซึ่งโดยปกติแล้วจะมีการเผยแพร่ชื่อและหมายเลข NRIC ของผู้โชคดีจากการจับรางวัลที่ได้ทำ Masking แล้วบางส่วนเพื่อให้ผู้อื่นรู้ว่าตนเป็นผู้ชนะ ตัวอย่างที่สองคือ ข้อมูล เช่น หมายเลขบัตรเครดิตของบุคคลที่ผ่านการทำ Masking แล้วในแอปหรือใบแจ้งยอดที่ส่งถึงบุคคลนั้น โปรดทราบว่าโดยทั่วไปไม่ควรจำข้อมูลที่ผ่านการทำ Anonymization แล้วได้แม้แต่เจ้าของข้อมูลเองก็ตาม

ตัวอย่าง

ตัวอย่างนี้เป็นร้านขายของชำออนไลน์ที่กำลังศึกษาความต้องการในการจัดส่งจากข้อมูลในอดีตเพื่อปรับปรุงประสิทธิภาพการดำเนินงาน บริษัทได้ทำ Masking กับรหัสไปรษณีย์ 4 หลักสุดท้าย โดยเหลือ 2 หลักแรกไว้ซึ่งตรงกับ “รหัสภาค” ภายในสิงคโปร์

ก่อนทำ Anonymization:

รหัสไปรษณีย์	ช่วงเวลาจัดส่งที่นิยม	จำนวนคำสั่งซื้อเฉลี่ยต่อเดือน
100111	20.00 น. ถึง 21.00 น.	2
200222	11.00 น. ถึง 12.00 น.	8
300333	14.00 น. ถึง 15.00 น.	1

หลังทำ Masking กับบางส่วนของรหัสไปรษณีย์:

รหัสไปรษณีย์	ช่วงเวลาจัดส่งที่นิยม	จำนวนคำสั่งซื้อเฉลี่ยต่อเดือน
10xxxx	20.00 น. ถึง 21.00 น.	2
20xxxx	11.00 น. ถึง 12.00 น.	8
30xxxx	14.00 น. ถึง 15.00 น.	1

การทำ Pseudonymization	
คำอธิบาย	การทำ Pseudonymization หมายถึงการเปลี่ยนข้อมูลระบุตัวตนเป็นค่าที่สร้างขึ้น หรืออีกชื่อหนึ่งเรียกว่าการโค้ดดิ้ง นามแฝงอาจย้อนกลับไม่ได้เมื่อลบค่าดั้งเดิมออกอย่างเหมาะสม และการทำ Pseudonymization จะทำในลักษณะที่ไม่สามารถทำซ้ำได้ นามแฝงสามารถย้อนกลับได้ (โดยเจ้าของข้อมูลดั้งเดิม) เมื่อค่าดั้งเดิมถูกเก็บไว้อย่างปลอดภัยสามารถดึงข้อมูลและเชื่อมโยงกลับไปยังนามแฝงได้หากจำเป็น[16] นามแฝงแบบถาวรช่วยให้สามารถเชื่อมโยงได้โดยใช้ค่านามแฝงเดียวกันเพื่อเป็นตัวแทนของบุคคลเดียวกันในชุดข้อมูลต่างๆ อย่างไรก็ตาม สามารถใช้นามแฝงต่างๆ เพื่อเป็นตัวแทนของบุคคลเดียวกันในชุดข้อมูลที่แตกต่างกันเพื่อป้องกันการเชื่อมโยงชุดข้อมูลที่แตกต่างกันได้ สามารถสร้างนามแฝงแบบสุ่มหรือจะกำหนดอย่างชัดเจนก็ได้
ควรใช้วิธีนี้เมื่อใด	ควรทำ Pseudonymization เมื่อจำเป็นต้องแยกแยะค่าข้อมูลตามลักษณะเฉพาะ และไม่มีการเก็บอักขระหรือข้อมูลโดยนัยอื่นใดเกี่ยวกับตัวระบุโดยตรงของคุณลักษณะดั้งเดิมไว้
ควรใช้วิธีนี้อย่างไร	เปลี่ยนค่าคุณลักษณะที่เกี่ยวข้องเป็นค่าที่สร้างขึ้น วิธีหนึ่งในการใช้วิธีนี้คือสร้างรายการค่าที่สร้างขึ้นล่วงหน้าและสุ่มเลือกจากรายการนี้เพื่อเปลี่ยนค่าเดิมแต่ละค่า ค่าที่สร้างขึ้นควรมีลักษณะเฉพาะและไม่ควรมีความสัมพันธ์กับค่าดั้งเดิม (ในลักษณะที่สามารถหาค่าดั้งเดิมจากนามแฝงได้)
คำแนะนำอื่นๆ	• เมื่อใช้นามแฝง ตรวจสอบให้แน่ใจว่าไม่ได้ใช้นามแฝงที่เคยใช้ในชุดข้อมูลเดียวกันซ้ำ โดยเฉพาะอย่างยิ่งเมื่อมีการสร้างนามแฝงแบบสุ่ม นอกจากนี้ หลีกเลี่ยงการใช้โปรแกรมสร้างนามแฝงเดียวกันสำหรับคุณลักษณะต่างๆ โดยไม่มีการเปลี่ยนแปลง (เช่น อย่างน้อยควรใช้เลขตั้งต้นแบบสุ่มที่แตกต่างกัน)

[16] ตัวอย่างเช่น ในกรณีที่การศึกษาวิจัยให้ผลลัพธ์ที่เป็นประโยชน์ที่สามารถแจ้งเตือนให้เจ้าของข้อมูลทราบได้

การทำ Pseudonymization	
	○ นามแฝงแบบถาวรมักจะให้อรรถประโยชน์มากกว่าโดยการรักษาความสมบูรณ์ของการอ้างอิงระหว่างชุดข้อมูล ○ สำหรับนามแฝงแบบย้อนกลับได้ ต้องไม่แชร์ตารางจับคู่ข้อมูลระบุตัวตนกับผู้รับ ควรเก็บตารางไว้อย่างปลอดภัย องค์กรสามารถใช้ตารางได้เมื่อจำเป็นต้องทำ Re-identification เท่านั้น • ในทำนองเดียวกัน หากใช้การเข้ารหัสหรือฟังก์ชันแฮชเพื่อทำ Pseudonymization กับข้อมูล ต้องปกป้องคีย์การเข้ารหัสหรืออัลกอริทึมแฮชและค่าสุ่ม (Salt Value) สำหรับแฮชให้ปลอดภัยจากการเข้าถึงที่ไม่ได้รับอนุญาต เนื่องจากการรั่วไหลของข้อมูลดังกล่าวอาจส่งผลให้เกิดการละเมิดข้อมูลโดยการทำให้สามารถถอดรหัสหรือใช้ตารางที่คำนวณล่วงหน้าเพื่ออนุมานข้อมูลที่ถูกแฮชได้ (โดยเฉพาะอย่างยิ่งสำหรับข้อมูลที่เป็นไปตามรูปแบบที่กำหนดไว้ล่วงหน้า เช่น NRIC) เช่นเดียวกับเครื่องสร้างตัวเลขสุ่มเทียม ซึ่งต้องใช้เลขตั้งต้น ต้องมั่นใจในความปลอดภัยของคีย์ที่ใช้เช่นเดียวกับการเข้ารหัสหรือกระบวนการที่สามารถย้อนกลับได้ประเภทอื่น[17] องค์กรควรทบทวนวิธีการเข้ารหัส (เช่น อัลกอริธึมและความยาวของคีย์) และฟังก์ชันแฮชเป็นระยะๆ เพื่อให้มั่นใจว่าวิธีดังกล่าวเหมาะสมปลอดภัย และเป็นที่ยอมรับตามมาตรฐานของอุตสาหกรรม • ในบางกรณี นามแฝงอาจต้องเป็นไปตามโครงสร้างหรือประเภทข้อมูลของค่าดั้งเดิม (เช่น เพื่อนำนามแฝงไปใช้ในแอปพลิเคชันซอฟต์แวร์) ในกรณีดังกล่าวอาจจำเป็นต้องใช้เครื่องสร้างนามแฝงพิเศษเพื่อสร้างชุดข้อมูลสังเคราะห์ หรือในบางกรณีสามารถใช้วิธีที่เรียกว่า “การเข้ารหัสแบบรักษารูปแบบ” ได้ ซึ่งจะสร้างนามแฝงที่มีรูปแบบเดียวกันกับข้อมูลดั้งเดิม

[17]โปรดทราบว่าการใช้กระบวนที่เป็นกรรมสิทธิ์หรือกระบวนการย้อนกลับ "ลับ" (ไม่ว่าจะมีหรือไม่มีคีย์) มีความเสี่ยงสูงกว่าที่จะถูกถอดรหัสเมื่อเปรียบเทียบกับการเข้ารหัสหรือการแฮชแบบใช้คีย์มาตรฐาน

ตัวอย่าง

ต่อไปนี้คือตัวอย่างการทำ Pseudonymization ที่ใช้กับชื่อของบุคคลที่ได้รับใบอนุญาตขับขี่และข้อมูลที่เกี่ยวข้อง ในตัวอย่างนี้ ชื่อจะถูกเปลี่ยนเป็นนามแฝงแทนที่จะทำ Suppresssion กับคุณลักษณะ เนื่องจากองค์กรต้องการให้สามารถย้อนกลับการทำ Pseudonymization ได้หากจำเป็น

ก่อนทำ Anonymization:

บุคคล	ผลก่อนการประเมิน	ชั่วโมงเรียนก่อนที่จะสอบผ่าน
โจ แพง	ก	20
แซค ลิม	ข	26
ยู เซ็ง ซาน	ค	30
ลินนี มอค	ง	29
เจสลิน แทน	ข	32
แชน สิว ลี	ก	25

หลังทำ Pseudonymization กับคุณลักษณะของ “บุคคล”:

บุคคล	ผลก่อนการประเมิน	ชั่วโมงเรียนก่อนที่จะสอบผ่าน
416765	ก	20
562396	ข	26
964825	ค	30
873892	ง	29
239976	ข	32
943145	ก	25

สำหรับการทำ Pseudonymization แบบย้อนกลับได้ ตารางจับคู่ข้อมูลระบุตัวตนจะถูกเก็บไว้อย่างปลอดภัยสำหรับกรณีที่มีความจำเป็นตามกฎหมายในอนาคตที่จะต้องทำ Re-identification กับบุคคล ควรใช้การควบคุมความปลอดภัย (รวมถึงด้านการบริหารและด้านเทคนิค) เพื่อปกป้องตารางจับคู่ข้อมูลระบุตัวตน

ตารางจับคู่ข้อมูลระบุตัวตน (โค้ดดิ้งเดี่ยว):

นามแฝง	บุคคล
416765	โจ แพง
562396	แซค ลิม
964825	ยู เช็ง ซาน
873892	ลินนี มอค
239976	เจสลิน แทน
943145	แชน สิว ลี

เพื่อเพิ่มความปลอดภัยเกี่ยวกับตารางจับคู่ข้อมูลระบุตัวตน สามารถใช้โค้ดดิ้งคู่ได้ ต่อจากตัวอย่างก่อนหน้านี้ ตัวอย่างนี้จะแสดงตารางเชื่อมโยงเพิ่มเติม ซึ่งเก็บรักษาโดยบุคคลที่สามที่เชื่อถือได้ เมื่อใช้โค้ดดิ้งคู่ จะทราบข้อมูลระบุตัวตนของบุคคลได้ก็ต่อเมื่อบุคคลที่สามที่เชื่อถือได้ (ซึ่งมีตารางเชื่อมโยง) และองค์กร (ซึ่งมีตารางจับคู่ข้อมูลระบุตัวตน) นำข้อมูลของตนมารวมกันเท่านั้น

หลังทำ Anonymization:

บุคคล	ผลก่อนการประเมิน	ชั่วโมงเรียนก่อนที่จะสอบผ่าน
373666	ก	20
594824	ข	26
839933	ค	30
280074	ง	29
746791	ข	32
785282	ก	25

ตารางเชื่อมโยง (เก็บรักษาโดยบุคคลที่สามที่เชื่อถือได้เท่านั้น และองค์กรก็จะลบตารางเชื่อมโยงออกในที่สุด นอกจากตารางเชื่อมโยงแล้วบุคคลที่สามจะไม่ได้รับข้อมูลอื่นใด):

นามแฝง	นามแฝงชั่วคราว
373666	OQCPBL
594824	ALGKTY
839933	CGFFNF
280074	BZMHCP
746791	RTJYGR
785282	RCNVJD

ตารางจับคู่ข้อมูลระบุตัวตน (องค์กรจะเก็บไว้อย่างปลอดภัย)

นามแฝงชั่วคราว	บุคคล
OQCPBL	โจ แพง
ALGKTY	แซค ลิม
CGFFNF	ยู เช็ง ซาน
BZMHCP	ลินนี มอค
RTJYGR	เจสลิน แทน
RCNVJD	แชน สิว ลี

หมายเหตุ: ในทั้งตารางเชื่อมโยงและตารางจับคู่ข้อมูลระบุตัวตน แนวทางปฏิบัติที่ดีคือการสลับลำดับของบันทึก แทนที่จะปล่อยให้อยู่ในลำดับเดียวกันกับชุดข้อมูล ในตัวอย่างนี้ บันทึกในทั้งสองตารางจะยังคงอยู่ในลำดับเดิมเพื่อให้เห็นภาพได้ง่ายขึ้น

การทำ Generalization	
คำอธิบาย	การทำ Generalization คือการลดความแม่นยำของข้อมูลโดยเจตนา เช่น การแปลงอายุของบุคคลเป็นช่วงอายุหรือการแปลงตำแหน่งที่แม่นยำให้เป็นตำแหน่งที่แม่นยำน้อยลง เทคนิคนี้มีอีกชื่อหนึ่งเรียกว่าการบันทึก
ควรใช้วิธีนี้เมื่อใด	การทำ Generalization ใช้สำหรับค่าที่สามารถทำ Generalization ได้และยังคงมีประโยชน์ตามวัตถุประสงค์ที่ต้องการ
ควรใช้วิธีนี้อย่างไร	ออกแบบหมวดหมู่ข้อมูลและกฎที่เหมาะสมสำหรับการแปลข้อมูล ทำ Suppression กับบันทึกใดๆ ที่ยังคงโดดเด่นหลังจากการแปล (กล่าวคือ การทำ Generalization)

การทำ Generalization	
คำแนะนำอื่นๆ	• เลือกช่วงข้อมูลที่เหมาะสม ช่วงข้อมูลที่มีขนาดใหญ่เกินไปอาจทำให้สูญเสียอรรถประโยชน์ของข้อมูลอย่างมีนัยสำคัญได้ ในขณะที่ช่วงข้อมูลที่เล็กเกินไปอาจทำให้ข้อมูลแทบไม่มีการแก้ไขเลยจึงยังคงสามารถทำ Re-identification ได้ง่าย หากใช้ *k*-anonymity ค่า *k* ที่เลือกจะส่งผลต่อช่วงข้อมูลด้วย โปรดทราบว่าควรเลือกช่วงแรกและช่วงสุดท้ายที่ใหญ่กว่าเพื่อรองรับจำนวนบันทึกที่ต่ำกว่าในส่วนนี้ ซึ่งมักเรียกว่าการโค้ดดิ้งด้านบน/ด้านล่าง

ตัวอย่าง

ในตัวอย่างนี้ ชุดข้อมูลประกอบด้วยชื่อของบุคคล (ซึ่งผ่านการทำ Pseudonymization แล้ว) อายุ (หน่วยเป็นปี) และที่อยู่อาศัย

ก่อนทำ Anonymization:

หมายเลขเครื่อง	บุคคล	อายุ	ที่อยู่
1	357703	24	700 Toa Payoh Lorong 5
2	233121	31	800 Ang Mo Kio Avenue 12
3	938637	44	900 Jurong East Street 70
4	591493	29	750 Toa Payoh Lorong 5
5	202626	23	5 Tampines Street 90
6	888948	75	1 Stonehenge Road
7	175878	28	10 Tampines Street 90
8	312304	50	50 Jurong East Street 70
9	214025	30	720 Toa Payoh Lorong 5
10	271714	37	830 Ang Mo Kio Avenue 12
11	341338	22	15 Tampines Street 90
12	529057	25	18 Tampines Street 90
13	390438	39	840 Ang Mo Kio Avenue 12

สำหรับคุณลักษณะ "อายุ" วิธีที่ใช้คือการทำ Generalization ให้เป็นช่วงอายุต่อไปนี้

< 20	21-30	31-40	41-50	51-60	> 60

สำหรับ “ที่อยู่” วิธีหนึ่งที่สามารถทำได้คือการลบบล็อก/เลขที่บ้านออก และคงไว้แต่ชื่อถนนเท่านั้น

หลังทำ Generalization กับคุณลักษณะ "อายุ" และ "ที่อยู่":

หมายเลขเครื่อง	บุคคล	อายุ	ที่อยู่
1	357703	21-30	Toa Payoh Lorong 5
2	233121	31-40	Ang Mo Kio Avenue 12
3	938637	41-50	Jurong East Street 70
4	591493	21-30	Toa Payoh Lorong 5
5	202626	21-30	Tampines Street 90
6	888948	> 60	Stonehenge Road
7	175878	21-30	Tampines Street 90
8	312304	41-50	Jurong East Street 70
9	214025	21-30	Toa Payoh Lorong 5
10	271714	31-40	Ang Mo Kio Avenue 12
11	341338	21-30	Tampines Street 90
12	529057	21-30	Tampines Street 90
13	390438	31-40	Ang Mo Kio Avenue 12

ตัวอย่างเช่น สมมติว่าในความเป็นจริงมียูนิตพักอาศัยเพียงยูนิตเดียวบนถนนสโตนเฮนจ์ จึงทำให้สามารถหาที่อยู่ที่แน่นอนได้แม้ว่าข้อมูลจะผ่านการทำ Generalization แล้วก็ตาม นี่อาจถือได้ว่า "มีลักษณะเฉพาะเกินไป”

ดังนั้น ในขั้นตอนถัดไปของการทำ Generalization คือ ลบบันทึกที่ 6 ออก (กล่าวคือ การใช้เทคนิคการทำ Suppression กับบันทึก) เนื่องจากที่อยู่ยังคง "มีลักษณะเฉพาะเกินไป" หลังจากลบหมายเลขยูนิตแล้ว อีกวิธีหนึ่ง คือ ทำ Generalization กับที่อยู่ทั้งหมดให้มีขอบเขตกว้างมาก

ขึ้น (เช่น เมืองหรือเขต) เพื่อให้ไม่จำเป็นต้องทำ Suppression อย่างไรก็ตามวิธีนี้อาจส่งผลกระทบต่ออรรถประโยชน์ของข้อมูลมากกว่าการทำ Suppression กับบันทึกบางรายการจากชุดข้อมูล

การแลกเปลี่ยน (Swapping)	
คำอธิบาย	จุดประสงค์ของการทำ Swapping คือเพื่อจัดเรียงข้อมูลในชุดข้อมูลใหม่โดยที่ค่าคุณลักษณะแต่ละรายการยังคงแสดงอยู่ในชุดข้อมูล แต่โดยทั่วไปแล้วจะไม่สอดคล้องกับบันทึกเดิม เทคนิคนี้มีอีกชื่อหนึ่งเรียกว่าการสับเปลี่ยน (Shuffling and Permutation)
ควรใช้วิธีนี้เมื่อใด	การทำ Swapping จะใช้เมื่อการวิเคราะห์ในภายหลังต้องการดูข้อมูลรวมเท่านั้น หรือการวิเคราะห์อยู่ที่ระดับคุณลักษณะภายใน อีกนัยหนึ่งคือ ไม่จำเป็นต้องวิเคราะห์ความสัมพันธ์ระหว่างคุณลักษณะในระดับบันทึก
ควรใช้วิธีนี้อย่างไร	ขั้นแรก ระบุคุณลักษณะที่จะทำ Swapping จากนั้น สำหรับแต่ละค่าในคุณลักษณะ ให้ทำ Swapping หรือกำหนดค่าใหม่ให้กับบันทึกอื่นในชุดข้อมูล
คำแนะนำอื่นๆ	• ประเมินและตัดสินใจว่าจะต้องทำ Swapping กับคุณลักษณะ (คอลัมน์) ใด ตัวอย่างเช่น องค์กรอาจตัดสินใจว่าเฉพาะคุณลักษณะ (คอลัมน์) ที่มีค่าที่สามารถระบุตัวตนได้ค่อนข้างมากเท่านั้นที่จำเป็นต้องทำ Swapping ขึ้นอยู่กับสถานการณ์

ตัวอย่าง

ในตัวอย่างนี้ ชุดข้อมูลประกอบด้วยข้อมูลเกี่ยวกับบันทึกของลูกค้าสำหรับองค์กรธุรกิจ

ก่อนทำ Anonymization:

บุคคล	ตำแหน่งงาน	วันเกิด	ประเภทสมาชิก	การใช้บริการเฉลี่ยต่อเดือน
ก	อาจารย์มหาวิทยาลัย	3 มกราคม 1970	เงิน	0
ข	พนักงานขาย	5 กุมภาพันธ์ 1972	แพลตตินัม	5
ค	ทนายความ	7 มีนาคม 1985	ทอง	2
ง	ผู้เชี่ยวชาญด้านไอที	10 เมษายน 1990	เงิน	1
จ	พยาบาล	13 พฤษภาคม 1995	เงิน	2

หลังทำ Anonymization:

ในตัวอย่างนี้ ค่าของคุณลักษณะทั้งหมดผ่านการทำ Swapping แล้ว

บุคคล	ตำแหน่งงาน	วันเกิด	ประเภทสมาชิก	การใช้บริการเฉลี่ยต่อเดือน
ก	ทนายความ	10 เมษายน 1990	เงิน	1
ข	พยาบาล	7 มีนาคม 1985	เงิน	2
ค	พนักงานขาย	13 พฤษภาคม 1995	แพลตตินัม	5
ง	ผู้เชี่ยวชาญด้านไอที	3 มกราคม 1970	เงิน	2
จ	อาจารย์มหาวิทยาลัย	5 กุมภาพันธ์ 1972	ทอง	0

หมายเหตุ: ในทางกลับกัน หากวัตถุประสงค์ของชุดข้อมูลที่ผ่านการทำ Anonymization คือเพื่อศึกษาความสัมพันธ์ระหว่างโปรไฟล์งานและรูปแบบการบริโภค วิธีการทำ Anonymization อื่นอาจเหมาะสมมากกว่า (เช่น การทำ Generalization กับตำแหน่งงานซึ่งอาจส่งผลให้ “อาจารย์มหาวิทยาลัย” ถูกแก้ไขเป็น "ผู้ให้การศึกษา")

การทำ Perturbation กับข้อมูล	
คำอธิบาย	ค่าจากชุดข้อมูลดั้งเดิมถูกแก้ไขให้แตกต่างออกไปเล็กน้อย
ควรใช้วิธีนี้เมื่อใด	การทำ Perturbation กับข้อมูลใช้สำหรับตัวระบุโดยอ้อม (โดยทั่วไปคือตัวเลขและวันที่) ซึ่งอาจระบุได้เมื่อนำไปรวมกับแหล่งข้อมูลอื่น แต่การเปลี่ยนแปลงค่าเล็กน้อยเป็นสิ่งที่ยอมรับได้สำหรับคุณลักษณะ ไม่ควรใช้เทคนิคนี้ในกรณีที่ความถูกต้องของข้อมูลเป็นสิ่งสำคัญ
ควรใช้วิธีนี้อย่างไร	ขึ้นอยู่กับเทคนิคการทำ Perturbation กับข้อมูลที่ใช้ ซึ่งรวมถึงการปัดเศษและการเพิ่มสัญญาณรบกวนแบบสุ่ม ตัวอย่างในส่วนนี้แสดงการปัดเศษฐาน x

การทำ Perturbation กับข้อมูล	
คำแนะนำอื่นๆ	• ระดับของการทำ Perturbation ควรเป็นสัดส่วนกับช่วงค่าของคุณลักษณะ หากฐานมีจำนวนน้อยเกินไป ผลของการทำ Anonymization จะลดลง ในทางกลับกันหากฐานมีจำนวนมากเกินไป ค่าสุดท้ายจะแตกต่างจากค่าเดิมมากเกินไป และอรรถประโยชน์ของชุดข้อมูลจะมีแนวโน้มที่จะลดลง • โปรดทราบว่าเมื่อมีการคำนวณค่าคุณลักษณะที่ผ่านการทำ Perturbation มาก่อน ค่าผลลัพธ์อาจได้รับผลจากการทำ Perturbation มากกว่าเดิม

ตัวอย่าง

ในตัวอย่างนี้ชุดข้อมูลประกอบด้วยข้อมูลที่จะใช้สำหรับการวิจัยเกี่ยวกับความเชื่อมโยงที่เป็นไปได้ระหว่างส่วนสูง น้ำหนัก อายุ บุคคลนั้นสูบบุหรี่หรือไม่ และบุคคลนั้นมี “โรค ก” และ/หรือ “โรค ข” หรือไม่ ชื่อของบุคคลนั้นผ่านการทำ Pseudonymization แล้ว

จากนั้นจึงใช้การปัดเศษดังต่อไปนี้:

คุณลักษณะ	เทคนิคการทำ Anonymization
ส่วนสูง (ซม.)	การปัดเศษฐาน 5 (เลือก 5 เนื่องจากค่อนข้างได้สัดส่วนกับค่าส่วนสูงโดยทั่วไปที่ 120 ถึง 190 ซม.)
น้ำหนัก (กก.)	การปัดเศษฐาน 3 (เลือก 3 เนื่องจากค่อนข้างได้สัดส่วนกับค่าน้ำหนักโดยทั่วไปที่ 40 ถึง 100 กก.)
อายุ (ปี)	การปัดเศษฐาน 3 (เลือก 3 เนื่องจากค่อนข้างได้สัดส่วนกับค่าอายุโดยทั่วไปที่ 10 ถึง 100 ปี)
คุณลักษณะที่เหลือ	ไม่มี เนื่องจากไม่ใช่ตัวเลขและยากที่จะแก้ไขโดยไม่ให้เปลี่ยนแปลงค่าอย่างมีนัยสำคัญ

ชุดข้อมูลก่อนทำ Anonymization:

บุคคล	ส่วนสูง (ซม.)	น้ำหนัก (กก.)	อายุ (ปี)	สูบบุหรี่หรือไม่	เป็นโรค ก หรือไม่	เป็นโรค ข หรือไม่
198740	160	50	30	ไม่	ไม่	ไม่
287402	177	70	36	ไม่	ไม่	ใช่
398747	158	46	20	ใช่	ใช่	ไม่
498732	173	75	22	ไม่	ไม่	ไม่
598772	169	82	44	ใช่	ใช่	ใช่

ชุดข้อมูลหลังการทำ Anonymization:

บุคคล	ส่วนสูง (ซม.)	น้ำหนัก (กก.)	อายุ (ปี)	สูบบุหรี่หรือไม่	เป็นโรค ก หรือไม่	เป็นโรค ข หรือไม่
198740	160	51	30	ไม่	ไม่	ไม่
287402	175	69	36	ไม่	ไม่	ใช่
398747	160	45	18	ใช่	ใช่	ไม่
498732	175	75	21	ไม่	ไม่	ไม่
598772	170	81	42	ใช่	ใช่	ใช่

หมายเหตุ: สำหรับการปัดเศษฐาน x ค่าคุณลักษณะที่จะปัดเศษจะถูกปัดเศษให้เป็นพหุคูณที่ใกล้เคียงที่สุดของ x

การทำ Aggregation กับข้อมูล	
คำอธิบาย	การทำ Aggregation กับข้อมูลหมายถึงการแปลงชุดข้อมูลจากรายการบันทึกเป็นค่าสรุป
ควรใช้วิธีนี้เมื่อใด	ใช้วิธีนี้เมื่อไม่ต้องการบันทึกแต่ละรายการและข้อมูลที่ผ่านการทำ Aggregation แล้วเพียงพอสำหรับวัตถุประสงค์ดังกล่าว
ควรใช้วิธีนี้อย่างไร	รายละเอียดเกี่ยวกับการวัดผลทางสถิติอยู่นอกเหนือขอบเขตของคู่มือฉบับนี้ อย่างไรก็ตามวิธีทั่วไปรวมถึงการใช้ผลรวมหรือค่าเฉลี่ย เป็นต้น นอกจากนี้การหารือกับผู้รับข้อมูลเกี่ยวกับอรรถประโยชน์ที่คาดหวังและการประนีประนอมที่เหมาะสม
คำแนะนำอื่นๆ	• ควรระวังในกรณีของกลุ่มที่มีบันทึกน้อยเกินไปหลังทำ Aggregation ในตัวอย่างด้านล่างหากข้อมูลที่ผ่านการทำ Aggregation แล้วมีบันทึกเดียวในหมวดหมู่ใดๆ ผู้ที่มีความรู้เพิ่มเติมบางอย่างสามารถระบุตัวตนของผู้บริจาคได้อย่างง่ายดาย • ดังนั้นการทำ Aggregation อาจจำเป็นต้องใช้ร่วมกับการทำ Suppression และอาจจำเป็นต้องลบคุณลักษณะบางอย่างออก เนื่องจากมีรายละเอียดที่ไม่สามารถทำ Aggregation ได้ นอกจากนี้ยังอาจจำเป็นต้องเพิ่มคุณลักษณะใหม่ (เช่น เพื่อให้มีค่ารวมที่คำนวณใหม่) ด้วย

ตัวอย่าง

ในตัวอย่างนี้ องค์กรการกุศลมีบันทึกการบริจาค รวมถึงข้อมูลบางอย่างเกี่ยวกับผู้บริจาค

องค์กรการกุศลประเมินว่าข้อมูลที่ผ่านการทำ Aggregation แล้วเพียงพอสำหรับที่ปรึกษาภายนอกจะทำการวิเคราะห์ข้อมูล จึงดำเนินการทำ Aggregation กับข้อมูลในชุดข้อมูลดั้งเดิม

ชุดข้อมูลดั้งเดิม:

ผู้บริจาค	รายได้ต่อเดือน ($)	จำนวนเงินบริจาค ในปี 2016 ($)
ผู้บริจาค ก	4000	210
ผู้บริจาค ข	4900	420
ผู้บริจาค ค	2200	150
ผู้บริจาค ง	4200	110
ผู้บริจาค จ	5500	260
ผู้บริจาค ฉ	2600	40
ผู้บริจาค ช	3300	130
ผู้บริจาค ซ	5500	210
ผู้บริจาค ฌ	1600	380
ผู้บริจาค ญ	3200	80
ผู้บริจาค ฎ	2000	440
ผู้บริจาค ฏ	5800	400
ผู้บริจาค ฐ	4600	390
ผู้บริจาค ฑ	1900	480
ผู้บริจาค ฒ	1700	320
ผู้บริจาค ณ	2400	330
ผู้บริจาค ด	4300	390
ผู้บริจาค ต	2300	260
ผู้บริจาค ถ	3500	80
ผู้บริจาค ท	1700	290

ชุดข้อมูลที่ผ่านการทำ Anonymization แล้ว:

รายได้ต่อเดือน ($)	จำนวนเงินบริจาคที่ได้รับ (2016)	ยอดจำนวนเงินบริจาคในปี 2016 ($)
1000-1999	4	1470
2000-2999	5	1220
3000-3999	3	290
4000-4999	5	1520
5000-6000	3	870
ยอดรวมทั้งหมด	**20**	**5370**

ภาคผนวก ข: คุณลักษณะของข้อมูลทั่วไปและเทคนิคการจัดทำข้อมูลนิรนามที่แนะนำ

ตัวระบุโดยตรง

ต่อไปนี้คือตารางคำแนะนำเกี่ยวกับเทคนิคการทำ Anonymization ที่สามารถนำไปใช้กับตัวระบุโดยตรงทั่วไปบางประเภทได้ โดยทั่วไปควรทำ Suppression (ลบ) หรือทำ Pseudonymization กับตัวระบุโดยตรง หากจำเป็นต้องกำหนดนามแฝงโดยปกติแล้วนามแฝงหนึ่งชุด (เช่น หนึ่งคอลัมน์) ต่อชุดข้อมูลถือว่าเพียงพอแล้ว

สำหรับกรณีการใช้ข้อมูลสังเคราะห์ สามารถเก็บคอลัมน์ตัวระบุโดยตรงทั้งหมดไว้ได้ แต่ต้องเปลี่ยนเป็นค่าที่ผ่านการทำ Pseudonymization แล้ว

การทำ Suppression กับบันทึก	**เทคนิคที่ใช้กันทั่วไป**	**ตัวอย่าง**	
		ก่อน	**หลัง**
• ชื่อ • ที่อยู่อีเมล • หมายเลขโทรศัพท์มือถือ • หมายเลข NRIC • หมายเลขหนังสือเดินทาง • หมายเลขบัญชี • หมายเลขสูติบัตร • หมายเลขประจำตัวต่างประเทศ (FIN) • หมายเลขใบอนุญาตทำงาน	**การทำ Suppression กับคุณลักษณะ**	จอห์น แทน	(ถูกลบออก)
	การกำหนดนามแฝง เช่น:		
	• เปลี่ยนค่าตัวระบุโดยตรงเป็นค่าสุ่มที่มีลักษณะเฉพาะ หรือ	จอห์น แทน	123456
	• เปลี่ยนค่าตัวระบุโดยตรงเป็นค่าที่สร้างขึ้นแบบสุ่มซึ่งเป็นไปตามรูปแบบของข้อมูล	John.tan@gmail.com S8822311H	123456@abc.com S8512345A

ตัวระบุโดยอ้อม

ต่อไปนี้คือตารางคำแนะนำเกี่ยวกับเทคนิคการทำ Anonymization ที่สามารถนำไปใช้กับตัวระบุโดยอ้อมทั่วไปบางประเภทได้ คุณควรเลือกใช้อย่างน้อย 1 เทคนิคกับตัวระบุโดยอ้อมแต่ละตัว (เช่น ใช้การทำ Generalization และการทำ Swapping กับอายุ ขึ้นอยู่กับยูสเคสของคุณ)

สำหรับกรณีการใช้ข้อมูลสังเคราะห์ สองเทคนิคที่มีประโยชน์คือการทำ Swapping และทำการ Perturbation กับข้อมูล เทคนิคเหล่านี้ใช้กับตัวระบุโดยอ้อมทั้งหมด

ตัวระบุโดยอ้อม	เทคนิคที่ใช้กันทั่วไป	ตัวอย่าง	
		ก่อน	หลัง
• **อายุ** • **ส่วนสูง** • **น้ำหนัก**	**การทำ Generalization:** ทำ Generalization กับอายุ/ส่วนสูง/น้ำหนักให้อยู่ในช่วง 5 หรือ 10 ปี/ซม./กก.	บันทึกที่ 1: 24 บันทึกที่ 2: 39 บันทึกที่ 3: 18	การทำ Generalization (ช่วงอายุ 5 ปี): บันทึกที่ 1: 21 ถึง 25 บันทึกที่ 2: 36 ถึง 40 บันทึกที่ 3: 16 ถึง 20
	การทำ Perturbation กับข้อมูล: เพิ่มค่าสุ่ม (+/- 5) ให้กับค่าเดิม		**การทำ Perturbation กับข้อมูล:** บันทึกที่ 1: 25 บันทึกที่ 2: 36 บันทึกที่ 3: 17
	การทำ Swapping: สุ่มเปลี่ยนอายุ/ส่วนสูง/น้ำหนักที่เกี่ยวข้องกับแต่ละบันทึก		**การทำ Swapping:** บันทึกที่ 1: 39 บันทึกที่ 2: 18 บันทึกที่ 3: 24

ตัวระบุโดยอ้อม	เทคนิคที่ใช้กันทั่วไป	ตัวอย่าง	
		ก่อน	หลัง
• **เพศ**	โดยทั่วไปคุณลักษณะของข้อมูลทางอ้อมนี้จะมีค่าที่ไม่ระบุตัวตนทั่วไปเพียงสองค่าเท่านั้น คือ ช หรือ ญ ดังนั้นโดยทั่วไปแล้วจึงปลอดภัยที่จะคงไว้เหมือนเดิม สำหรับกรณีการใช้ข้อมูลสังเคราะห์สามารถใช้เทคนิคต่อไปนี้กับคุณลักษณะนี้เพิ่มเติมได้ **การทำ Swapping:** สุ่มเปลี่ยนเพศภายในชุดข้อมูล	บันทึกที่ 1: ช บันทึกที่ 2: ช บันทึกที่ 3: ญ บันทึกที่ 4: ช	**การทำ Swapping:** บันทึกที่ 1: ช บันทึกที่ 2: ญ บันทึกที่ 3: ช บันทึกที่ 4: ช
• **เชื้อชาติ** • **สถานภาพการสมรส**	**การทำ Generalization:** คุณสามารถรวมและทำ Generalization กับกลุ่มชาติพันธุ์หรือสถานภาพการสมรสที่เลือกให้เป็นหมวดหมู่ที่เรียกว่า “อื่นๆ” ได้ ทั้งนี้ขึ้นอยู่กับชุดข้อมูลของคุณ จะต้องใช้วิธีนี้หากมีกลุ่มชาติ	บันทึกที่ 1: อินเดีย บันทึกที่ 2: จีน บันทึกที่ 3: จีน บันทึกที่ 4: มาเลย์ บันทึกที่ 5: ยูเรเชียน	**การทำ Generalization:** บันทึกที่ 1: อื่นๆ บันทึกที่ 2: จีน บันทึกที่ 3: จีน บันทึกที่ 4: อื่นๆ บันทึกที่ 5: อื่นๆ

ตัวระบุโดยอ้อม	เทคนิคที่ใช้กันทั่วไป	ตัวอย่าง	
		ก่อน	หลัง
	พันธุ์/สถานภาพการสมรสที่มีลักษณะเฉพาะ หรือมีกลุ่มชาติพันธุ์/สถานภาพสมรสเดียวกันน้อยเกินไปภายในชุดข้อมูลของคุณ		
	การทำ Swapping: สุ่มเปลี่ยนเชื้อชาติหรือสถานภาพสมรสภายในชุดข้อมูล		**การแลกเปลี่ยน:** บันทึกที่ 1: มาเลย์ บันทึกที่ 2: จีน บันทึกที่ 3: อินเดีย บันทึกที่ 4: ยูเรเชียน บันทึกที่ 5: จีน
• **วันเกิด**	**การทำ Generalization:** ทำ Generalization กับวันเกิดให้เป็นปีหรือเดือนกับปี	บันทึกที่ 1: 1 ก.พ. 2003 บันทึกที่ 2: 15 ส.ค. 1990 บันทึกที่ 3: 30 ธ.ค. 1998	การทำ Generalization (เดือนและปี): บันทึกที่ 1: ก.พ. 2003 บันทึกที่ 2: ส.ค. 1990 บันทึกที่ 3: ธ.ค. 1998
	การทำ Perturbation กับข้อมูล: สุ่มแก้ไขวันที่ (เช่น +/- 30 วันนับจากวันที่เดิม)		**การทำ Perturbation กับข้อมูล:** บันทึกที่ 1: 20 ม.ค. 2003 บันทึกที่ 2: 18 ส.ค. 1990 บันทึกที่ 3: 6 ม.ค. 1999

ตัวระบุโดยอ้อม	เทคนิคที่ใช้กันทั่วไป	ตัวอย่าง	
		ก่อน	หลัง
	การทำ Swapping: สุ่มเปลี่ยนวันที่ภายในชุดข้อมูล		**การทำ Swapping:** บันทึกที่ 1: 30 ธ.ค. 1998 บันทึกที่ 2: 1 ก.พ. 2003 บันทึกที่ 3: 15 ส.ค. 1990
• **ที่อยู่**	**การทำ Generalization:** ทำ Generalization กับที่อยู่ให้เป็นโซนที่กำหนดไว้ล่วงหน้า (เช่น โดยอ้างอิงถึงแผนแม่บทของหน่วยงานด้านการพัฒนาเมือง (URA)[18])	71 Punggol Central, Singapore 828755	**การทำ Swapping:** บันทึกที่ 1: 35 Mandalay Road, #13-37 Singapore 208215 บันทึกที่ 2: 71 Punggol Central, #10-1122, Singapore 828755
	การทำ Swapping: สุ่มเปลี่ยนที่อยู่ภายในชุดข้อมูล *หมายเหตุ: สำหรับที่อยู่ หมายเลขยู*	บันทึกที่ 1: 71 Punggol Central, #10-1122, Singapore 828755 บันทึกที่ 2: 35 Mandalay Road,	

[18] https://www.ura.gov.sg/maps/?service=MP.

ตัวระบุโดยอ้อม	เทคนิคที่ใช้กันทั่วไป	ตัวอย่าง	
		ก่อน	หลัง
	นิติอาจระบุตัวตนได้ ในกรณีที่ไม่จำเป็นควรลบหมายเลขยูนิตออกจากชุดข้อมูล	#13-37 Singapore 208215	
• **รหัสไปรษณีย์**	**การทำ Masking กับอักขระ:** ทำ Masking กับรหัสไปรษณีย์สี่หลักสุดท้าย (สิงคโปร์มีเขตไปรษณีย์ 80 แห่ง)	117438	
	การทำ Swapping: สุ่มเปลี่ยนรหัสไปรษณีย์ภายในชุดข้อมูล	บันทึกที่ 1: 117438 บันทึกที่ 2: 828755	**การทำ Swapping:** บันทึกที่ 1: 828755 บันทึกที่ 2: 117438
• **ตำแหน่งงาน**	**การทำ Generalization:** ไม่มีวิธีง่ายๆ ในการทำ Anonymization กับตำแหน่งงานด้วยวิธีอัตโนมัติ เนื่องจากตำแหน่งงานไม่มีมาตรฐานและองค์กรต่างๆ ก็สามารถคิดตำแหน่งงานของตนเองขึ้นมาได้ วิธีหนึ่งคือการทำ	ประธานเจ้าหน้าที่บริหาร หัวหน้าทีมพัฒนาซอฟต์แวร์	**การทำ Generalization:** ผู้บริหารระดับสูง ผู้จัดการไอที

ตัวระบุโดยอ้อม	เทคนิคที่ใช้กันทั่วไป	ตัวอย่าง	
		ก่อน	หลัง
	Generalization กับตำแหน่งงานให้เป็นอนุกรมวิธานที่กำหนดไว้ล่วงหน้าของลักษณะงานและ/หรือระดับงาน อย่างไรก็ตาม เป็นไปได้ที่จะต้องทำการจับคู่ด้วยตนเอง **การทำ Swapping:** สุ่มเปลี่ยนตำแหน่งงานภายในชุดข้อมูล	บันทึกที่ 1: ซีอีโอ บันทึกที่ 2: กรรมการ บันทึกที่ 3: ผู้จัดการ	**การทำ Swapping:** บันทึกที่ 1: ผู้จัดการ บันทึกที่ 2: ซีอีโอ บันทึกที่ 3: กรรมการ
• **ชื่อบริษัท**	**การทำ Generalization:** การทำ Generalization กับชื่อบริษัทตามภาคอุตสาหกรรม (เช่น โดยอ้างอิงถึงระบบจำแนกประเภทอุตสาหกรรมของประเทศสิงคโปร์ (SSIC))[19]	บริษัท สปีดี้แท็กซี่ จำกัด	**การทำ Generalization:** การขนส่งและการจัดเก็บ

[19] https://www.singstat.gov.sg/standards/standards-and-classifications/ssic.

ตัวระบุโดยอ้อม	เทคนิคที่ใช้กันทั่วไป	ตัวอย่าง	
		ก่อน	หลัง
	การทำ Swapping: สุ่มเปลี่ยนชื่อบริษัทภายในชุดข้อมูล	บันทึกที่ 1: บริษัท สปีดี้ แท็กซี่ จำกัด บันทึกที่ 2: เบสท์ ฟู้ด จำกัด บันทึกที่ 3: บริษัท นัมเบอร์ 1 โคลด์ แวร์ เอกชนจำกัด	**การทำ Swapping:** บันทึกที่ 1: เบสท์ ฟู้ด จำกัด บันทึกที่ 2: บริษัท นัมเบอร์ 1 โคลด์ แวร์ เอกชนจำกัด บันทึกที่ 3: บริษัท สปีดี้ แท็กซี่ จำกัด
• **ที่อยู่ไอพี**	**การทำ Masking กับอักขระ:** ทำ Masking กับสองออคเต็ตสุดท้าย[20]ของที่อยู่ไอพี IPv4 และ 80 บิตสุดท้ายของที่อยู่ไอพี IPv6 *หมายเหตุ: สามารถใช้การ***ทำ Swapping** *นอกเหนือจากการทำ Masking กับอักขระได้*	IPv4: 12.120.210.88 IPv6: 2001:0db8:85a3:0000:0 000:8a2e:0370:7334	**การทำ Masking กับอักขระ:** IPv4: 12.120.xxx.xxx IPv6: 2001:0db8:85a3:xxxx-:xxxx:xxxx:xxxx:xxxx
• **หมายเลขทะเบียนรถ**	**การทำ Masking กับอักขระ:** ทำ Masking กับอักขระสี่ตัวสุดท้ายของหมายเลขทะเบียนรถ *หมายเหตุ: สามารถใช้การ***ทำ Swapping**	SMF1234A	**การทำ Masking กับอักขระ:** SMF1xxxx

[20] เราแนะนำให้ทำ Masking กับสองออคเต็ตสุดท้ายโดยไม่คำนึงถึงระดับหมายเลขที่อยู่ของเครือข่าย (ก/ข/ค) เพื่อป้องกันไม่ให้มีการระบุว่าที่อยู่ที่ผ่านการทำ Masking แล้วอยู่ในเครือข่ายย่อยระดับ ข หรือ ค วิธีนี้ยังทำให้ยากต่อการจัดกลุ่มบุคคลที่อยู่บนเครือข่ายย่อยเดียวกันอีกด้วย

ตัวระบุโดยอ้อม	เทคนิคที่ใช้กันทั่วไป	ตัวอย่าง	
		ก่อน	หลัง
	นอกเหนือจากการทำ Masking กับอักขระได้		
• **หมายเลขหน่วยในรถยนต์ (IU)**	**การทำ Masking กับอักขระ:** ทำ Masking กับตัวเลขสามหลักสุดท้ายของหมายเลข IU *หมายเหตุ: สามารถใช้การ***ทำ Swapping** *นอกเหนือจากการทำ Masking กับอักขระได้*	1234567890	**การทำ Masking กับอักขระ:** 1234567xxx
• **ตำแหน่งของระบบกำหนดตำแหน่งบนพื้นโลก (GPS)**	**การทำ Generalization:** *ปัดเศษพิกัด GPS (เป็นองศาทศนิยม) เป็นทศนิยมสองตำแหน่งที่ใกล้เคียงที่สุด (เทียบเท่าความแม่นยำ 1.11 กม.) หรือทศนิยมสามตำแหน่ง (เทียบเท่าความแม่นยำ 111 ม.)*	1.27434, 103.79967	**การทำ Generalization:** 1.274, 103.800 (ปัดเศษองศาทศนิยมเป็นทศนิยมสามตำแหน่ง)

ตัวระบุโดยอ้อม	เทคนิคที่ใช้กันทั่วไป	ตัวอย่าง	
		ก่อน	หลัง
	การทำ Perturbation กับข้อมูล: เพิ่มค่าสุ่มระหว่าง 0.005 กับ -0.005 หรือระหว่าง 0.0005 กับ -0.0005		**การทำ Perturbation กับข้อมูล:** 1.27834, 103.79767
	การทำ Swapping: สุ่มเปลี่ยนตำแหน่ง GPS ภายในชุดข้อมูล	*บันทึกที่ 1: 1.27434, 103.79967* *บันทึกที่ 2: 1.26421, 103.80405* *บันทึกที่ 3: 1.26463, 103.82226*	**การทำ Swapping:** บันทึกที่ 1: 1.26463, 103.82226 บันทึกที่ 2: 1.27434, 103.79967 บันทึกที่ 3: 1.26421, 103.80405

คุณลักษณะเป้าหมาย

คุณลักษณะเป้าหมายเป็นข้อมูลที่เป็นกรรมสิทธิ์ซึ่งมีความสำคัญต่อการเก็บรักษาไว้เพื่ออรรถประโยชน์ของข้อมูล ดังนั้นในยูสเคสส่วนใหญ่เทคนิคการทำ Anonymization จะไม่นำไปใช้กับคุณลักษณะเป้าหมาย อย่างไรก็ตามสำหรับกรณีการใช้ข้อมูลสังเคราะห์ เนื่องจากข้อมูลระดับบันทึกมักใช้ในการพัฒนาและการทดสอบซึ่งอาจไม่ได้รับการรักษาความปลอดภัยอย่างเหมาะสม จึงขอแนะนำให้ใช้เทคนิคการทำ Anonymization อย่างน้อยหนึ่งเทคนิคกับคุณลักษณะเป้าหมายเพื่อให้แน่ใจว่าจะไม่มีการทำ Re-identification ในกรณีที่มีการละเมิดข้อมูล

สิ่งสำคัญคือต้องตรวจสอบและทำให้แน่ใจว่าหลังจากใช้เทคนิคการทำ Anonymization แล้ว ไม่มีบันทึกในชุดข้อมูลสังเคราะห์ที่คล้ายกับบันทึกใดๆ ในชุดข้อมูลดั้งเดิม

คุณลักษณะเป้าหมาย	**เทคนิคที่ใช้กันทั่วไป**	**ตัวอย่าง**	
		ก่อน	**หลัง**
• **ธุรกรรม** • **เงินเดือน** • **อันดับเครดิต** • **กรมธรรม์ประกันภัย** • **การวินิจฉัยทางการแพทย์** • **สถานะการฉีดวัคซีน**	**การทำ Perturbation กับข้อมูล:** สุ่มแก้ไขข้อมูลตัวเลข (เช่น บวกหรือลบค่าสุ่มจากข้อมูลดั้งเดิม) ไม่สามารถทำ Perturbation กับข้อมูลได้สำหรับข้อมูลที่เป็นตัวอักษรและตัวเลขหรือข้อมูลที่เป็นข้อความแบบไม่มีโครงสร้าง	มูลค่าการซื้อ: $38.05 เงินเดือน: $6,200	**การทำ Perturbation กับข้อมูล:** มูลค่าการซื้อ: $42 เงินเดือน: $7,500
	การทำ Swapping: สุ่มเปลี่ยนข้อมูลภายในชุดข้อมูล *หมายเหตุ: สามารถใช้การทำ* **Swapping** *นอกเหนือจากการทำ Perturbation กับข้อมูลได้*	สถานะการฉีดวัคซีน: บันทึกที่ 1: ฉีดวัคซีนแล้ว บันทึกที่ 2: เข็มแรก บันทึกที่ 3: ยังไม่ได้ฉีดวัคซีน	**การทำ Swapping:** สถานะการฉีดวัคซีน: บันทึกที่ 1: เข็มแรก บันทึกที่ 2: ยังไม่ได้ฉีดวัคซีน บันทึกที่ 3: ฉีดวัคซีนแล้ว

ภาคผนวก ค: การปิดบังชื่อด้วยเทคนิค *k*-ANONYMITY

เทคนิค *k*-anonymity (และส่วนขยายที่คล้ายกัน เช่น *l*-diversity และ *t*-closeness) เป็นมาตรการที่ใช้เพื่อให้แน่ใจว่าความเสี่ยงจะไม่เกินเกณฑ์ซึ่งเป็นส่วนหนึ่งของวิธีการทำ Anonymization

เทคนิค *k*-anonymity ไม่ใช่มาตรการเดียวที่ใช้ได้ และไม่ได้ไม่มีข้อจำกัด แต่ก็ค่อนข้างเป็นที่เข้าใจอย่างกว้างขวางและนำไปใช้ได้ง่าย *k*-anonymity อาจไม่เหมาะกับชุดข้อมูลทุกประเภทหรือยูสเคสที่ซับซ้อนอื่นๆ วิธีการและ/หรือเครื่องมืออื่นๆ เช่น อัลกอริทึมตรวจจับลักษณะเฉพาะพิเศษ (SUDA) และ μ-Argus อาจเหมาะสมกว่าสำหรับการประเมินความเสี่ยงของชุดข้อมูลขนาดใหญ่ นอกจากนี้ยังมีวิธีการทางเลือก เช่น ความเป็นส่วนตัวที่แตกต่างกัน (Differential Privacy)[21] เริ่มถูกนำมาใช้ในช่วงไม่กี่ปีที่ผ่านมาอีกด้วย

***k*-anonymity**	
คำอธิบาย	แบบจำลอง *k*-anonymity ถูกใช้เป็นแนวทางก่อนที่จะนำเทคนิคการทำ Anonymization (เช่น การทำ Generalization) มาใช้ และสำหรับการตรวจสอบหลังจากนั้นด้วยเช่นกันเพื่อให้แน่ใจว่าตัวระบุโดยอ้อมของบันทึกใด ๆ จะถูกแชร์โดยบันทึกอื่น ๆ *k*-1 รายการเป็นอย่างน้อย นี่คือการป้องกันที่สำคัญโดย *k*-anonymity จากการโจมตีด้วยการเชื่อมโยง เนื่องจากบันทึก *k* รายการ (หรือตัวระบุโดยอ้อมที่แตกต่างกันน้อยที่สุด) จะเหมือนกันในคุณลักษณะที่ระบุตัวตน ดังนั้นจึงสร้างระดับ[22]ที่เทียบเท่ากับสมาชิก *k* ราย ดังนั้นจึงไม่สามารถเชื่อมโยงหรือแยกบันทึกของแต่ละบุคคลได้ เนื่องจากมีคุณสมบัติที่เหมือนกัน *k* รายการเสมอ ชุดข้อมูลที่ผ่านการทำ Anonymization แล้วอาจมีระดับ *k*-anonymity ที่แตกต่างกันสำหรับชุดตัวระบุโดยอ้อมที่แตกต่างกัน แต่สำหรับการป้องกัน "ความเสี่ยงสูงสุด" จากการเชื่อมโยง ค่า *k* ต่ำสุดจะถูกใช้เป็นค่าตัวแทนเพื่อเปรียบเทียบกับเกณฑ์

[21] Differential Privacy เกี่ยวข้องกับหลายแนวคิด รวมถึงการตอบคำถามแทนที่จะจัดเตรียมชุดข้อมูลที่ผ่านการทำ Anonymization แล้ว การเพิ่มสัญญาณรบกวนแบบสุ่มเพื่อปกป้องบันทึกแต่ละรายการ การรับประกันทางคณิตศาสตร์ว่าจะไม่เกิน "งบประมาณความเป็นส่วนตัว" ที่กำหนดไว้ล่วงหน้า ฯลฯ

[22] “ระดับที่เทียบเท่า” หมายถึงบันทึกในชุดข้อมูลที่ใช้ค่าเดียวกันภายในคุณลักษณะบางอย่าง ซึ่งโดยทั่วไปแล้วจะเป็นตัวระบุโดยอ้อม

***k*-anonymity**	
ควรใช้วิธีนี้เมื่อใด	*k*-anonymity ใช้เพื่อยืนยันว่ามาตรการในการทำ Anonymization ที่นำมาใช้เป็นไปตามเกณฑ์ที่ต้องการในการป้องกันการโจมตีด้วยการเชื่อมโยง
ควรใช้วิธีนี้ อย่างไร	ขั้นแรก เลือกค่าสำหรับ *k* (ซึ่งเท่ากับหรือสูงกว่าค่าผกผันของขนาดระดับที่เทียบเท่า) ซึ่งให้ค่า *k* ต่ำสุดที่จะได้มาจากระดับที่เทียบเท่าทั้งหมด โดยทั่วไป ยิ่งค่า *k* สูงเท่าใด การระบุตัวตนของเจ้าของข้อมูลก็จะยิ่งยากขึ้นเท่านั้น อย่างไรก็ตามอรรถประโยชน์อาจลดลงเมื่อ *k* เพิ่มขึ้น และอาจจำเป็นต้องทำ Suppression กับบันทึกมากขึ้น หลังจากทำ Anonymization แล้ว ให้ตรวจสอบว่าแต่ละบันทึกมีบันทึกอื่นๆ อย่างน้อย *k*-1 รายการที่มีคุณลักษณะเดียวกันซึ่งจัดการโดยการทำ *k*-anonymization ควรทำ Suppression กับบันทึกในระดับเทียบเท่าที่มีบันทึกน้อยกว่า *k* รายการ หรือจะทำ Anonymization กับชุดข้อมูลเพิ่มเติมก็ได้
คำแนะนำอื่นๆ	• นอกจากการทำ Generalization แล้ว ยังสามารถสร้างข้อมูลสังเคราะห์เพื่อให้ได้ *k*-anonymity ได้ด้วยเช่นกัน บางครั้งสามารถใช้เทคนิคเหล่านี้ (และอื่นๆ) ร่วมกันได้ แต่โปรดทราบว่าวิธีการเฉพาะที่เลือกอาจส่งผลต่ออรรถประโยชน์ของข้อมูล ควรพิจารณาข้อดีข้อเสียระหว่างการลบค่าผิดปกติหรือการแทรกข้อมูลสังเคราะห์ • *k*-anonymity ถือว่าแต่ละบันทึกเกี่ยวข้องกับบุคคลอื่น หากบุคคลคนเดียวกันมีบันทึกหลายรายการ (เช่น ไปโรงพยาบาลหลายครั้ง) *k*-anonymity จะต้องสูงกว่าบันทึกที่ซ้ำกัน มิฉะนั้นบันทึกอาจไม่เพียงแต่สามารถเชื่อมโยงกันได้เท่านั้น แต่ยังสามารถทำ Re-identification ได้อีกด้วย แม้ว่าจะดูเหมือนจะมี "ระดับเทียบเท่า *k*"

ตัวอย่าง

ในตัวอย่างนี้ ชุดข้อมูลประกอบด้วยข้อมูลเกี่ยวกับคนที่ใช้บริการแท็กซี่

ใช้ *k* = 5 (กล่าวคือ แต่ละบันทึกควรใช้คุณลักษณะเดียวกันกับบันทึกอื่นอีกสี่รายการในที่สุดหลังทำ Anonymization)

ใช้เทคนิคการทำ Anonymization ต่อไปนี้ร่วมกัน ระดับรายละเอียดเป็นแนวทางหนึ่งเพื่อให้ได้ระดับ *k* ที่ต้องการ

คุณลักษณะ	เทคนิคการทำ Anonymization
อายุ	การทำ Generalization (ช่วง 10 ปี)
อาชีพ	การทำ Generalization (เช่น ทั้ง “ผู้ดูแลระบบฐานข้อมูล” และ “โปรแกรมเมอร์” ต่างถูกทำ Generalization ให้เป็น “ไอที”)
การทำ Suppression กับบันทึก	บันทึกที่ไม่ตรงตามเกณฑ์การทำ Anonymization 5 ข้อหลังจากใช้เทคนิคการทำ Anonymization แล้ว (ในกรณีนี้คือการทำ Generalization) จะถูกลบออก ตัวอย่างเช่น บันทึกของนายธนาคารจะถูกลบออกเนื่องจากเป็นเพียงค่าเดียวภายใต้ "อาชีพ”

ชุดข้อมูลก่อนทำ Anonymization:

หมายเลขเครื่อง	อายุ	เพศ	อาชีพ	จำนวนการเดินทางเฉลี่ยต่อเดือน
1	21	หญิง	ผู้ช่วยเจ้าหน้าที่คุ้มครองข้อมูลส่วนบุคคล	15
2	38	ชาย	หัวหน้าที่ปรึกษาด้านไอที	2
3	25	หญิง	นายธนาคาร	8
4	34	ชาย	ผู้ดูแลระบบฐานข้อมูล	3
5	30	หญิง	หัวหน้าเจ้าหน้าที่ดูแลข้อมูลส่วนบุคคล	1
6	29	หญิง	เจ้าหน้าที่คุ้มครองข้อมูลส่วนบุคคลประจำภูมิภาค	5
7	38	ชาย	โปรแกรมเมอร์	3
8	32	ชาย	นักวิเคราะห์ไอที	4
9	25	หญิง	ตัวแทนเจ้าหน้าที่คุ้มครองข้อมูลส่วนบุคคล	2
10	23	หญิง	ผู้จัดการสำนักงานเจ้าหน้าที่คุ้มครองข้อมูลส่วนบุคคล	11
11	31	ชาย	นักออกแบบ UX	0

ชุดข้อมูลกลายเป็นชุดข้อมูลที่ไม่มีลักษณะเฉพาะ 5 รายการ หลังจากทำ Anonymization กับอายุและอาชีพ และทำ Suppression กับค่าผิดปกติ (ระดับความเท่าเทียมกันตามลำดับจะถูกไฮไลท์ด้วยสีที่ต่างกัน):

หมายเลขเครื่อง	อายุ	เพศ	อาชีพ	จำนวนการเดินทางเฉลี่ยต่อเดือน
1	21 ถึง 30	หญิง	เจ้าหน้าที่คุ้มครองข้อมูลส่วนบุคคล	15
2	31 ถึง 40	ชาย	ไอที	2
3	21 ถึง 30	หญิง	นายธนาคาร	8
4	31 ถึง 40	ชาย	ไอที	3
5	21 ถึง 30	หญิง	เจ้าหน้าที่คุ้มครองข้อมูลส่วนบุคคล	1
6	21 ถึง 30	หญิง	เจ้าหน้าที่คุ้มครองข้อมูลส่วนบุคคล	5
7	31 ถึง 40	ชาย	ไอที	3
8	31 ถึง 40	ชาย	ไอที	4
9	21 ถึง 30	หญิง	เจ้าหน้าที่คุ้มครองข้อมูลส่วนบุคคล	2
10	21 ถึง 30	หญิง	เจ้าหน้าที่คุ้มครองข้อมูลส่วนบุคคล	11
11	31 ถึง 40	ชาย	ไอที	0

หมายเหตุ: นี่เป็นตัวอย่างจำนวนการเดินทางโดยเฉลี่ยต่อสัปดาห์สำหรับคุณลักษณะเป้าหมาย โดยไม่จำเป็นต้องทำ Anonymization กับคุณลักษณะนี้เพิ่มเติม

ภาคผนวก ง: การประเมินความเสี่ยงในการย้อนรอยเพื่อระบุตัวตน (Re-identification)

มีหลายวิธีในการประเมินความเสี่ยงในการทำ Re-identification และอาจต้องใช้การคำนวณที่ค่อนข้างซับซ้อนซึ่งเกี่ยวข้องกับการคำนวณความน่าจะเป็น

ในส่วนนี้จะอธิบายแบบจำลองอย่างง่ายโดยใช้ *k*-anonymity[23] และตั้งสมมติฐานต่อไปนี้:

1. แบบจำลองการเผยแพร่ไม่ใช่ข้อมูลสาธารณะ
2. ผู้โจมตีมีแรงจูงใจที่จะเชื่อมโยงบุคคลกับชุดข้อมูลที่ผ่านการทำ Anonymization แล้วและ
3. เนื้อหาของข้อมูลที่ผ่านการทำ Anonymization แล้วจะไม่ถูกนำมาพิจารณา และความเสี่ยงที่คำนวณได้นั้นไม่ขึ้นอยู่กับประเภทของข้อมูลที่ผู้โจมตีมีอยู่จริง

ขั้นแรก ควรกำหนดเกณฑ์ความเสี่ยง ค่าที่สะท้อนถึงความน่าจะเป็นนี้จะมีค่าอยู่ระหว่าง 0 ถึง 1 ค่านี้จะสะท้อนถึงระดับความเสี่ยงที่องค์กรยอมรับได้ ปัจจัยหลักที่ส่งผลต่อเกณฑ์ความเสี่ยงควรรวมถึงอันตรายที่อาจเกิดขึ้นได้กับเจ้าของข้อมูล เช่นเดียวกับอันตรายต่อองค์กรหากมีการทำ Re-identification นอกจากนั้นยังคำนึงถึงการควบคุมอื่นๆ ที่ถูกนำมาใช้เพื่อลดความเสี่ยงที่เหลือด้วย ยิ่งอันตรายที่อาจเกิดขึ้นสูงเท่าใด เกณฑ์ความเสี่ยงก็จะยิ่งสูงขึ้นเท่านั้น ไม่มีกฎตายตัวว่าควรใช้ค่าเกณฑ์ความเสี่ยงใด ต่อไปนี้เป็นเพียงตัวอย่างเท่านั้น

อันตรายที่อาจเกิดขึ้นได้	เกณฑ์ความเสี่ยง
ต่ำ	0.2
กลาง	0.1
สูง	0.01

ในการคำนวณความเสี่ยงในการทำ Re-identification คู่มือฉบับนี้ใช้ “ความเสี่ยงของผู้กล่าวหา (Prosecutor Risk)” ซึ่งสมมุติว่าผู้โจมตีรู้จักบุคคลใดบุคคลหนึ่งในชุดข้อมูล และกำลังพยายามหาว่าบันทึกใดในชุดข้อมูลอ้างถึงบุคคลนั้น

กฎง่ายๆ ในการคำนวณความน่าจะเป็นของการทำ Re-identification สำหรับบันทึกเดียวในชุดข้อมูล คือการใช้ตัวผกผันของขนาดระดับเทียบเท่าของบันทึก:

[23] การคำนวณจะแตกต่างออกไปหากใช้ Differential Privacy หรือการควบคุมการเปิดเผยข้อมูลทางสถิติแบบดั้งเดิม เป็นต้น

P (เชื่อมโยงบุคคลกับบันทึกเดียว) = 1 / ขนาดระดับเทียบเท่าของบันทึก

ในการคำนวณความน่าจะเป็นของการทำ Re-identification ของบันทึกใดๆ ในชุดข้อมูลทั้งหมด สมมุติว่ามีความพยายามที่จะทำ Re-identification วิธีที่ปลอดภัยคือการเทียบเคียงกับความน่าจะเป็นสูงสุดของการทำ Re-identification ระหว่างบันทึกทั้งหมดในชุดข้อมูล

P (การทำ Re-identification กับบันทึกใดๆ ในชุดข้อมูล) = 1 / ขนาดระดับที่เทียบเท่าในชุดข้อมูลที่น้อยที่สุด

หมายเหตุ: หากชุดข้อมูลผ่านการทำ k-anonymization แล้ว (การทำ Re-identification กับบันทึกใด ๆ ในชุดข้อมูล) <= 1 / k

เราสามารถพิจารณาสถานการณ์การโจมตีของผู้บุกรุกที่มีแรงจูงใจได้สามสถานการณ์:

1. การโจมตีโดยเจตนาโดยใช้ข้อมูลภายใน
2. การรับรู้โดยไม่ได้ตั้งใจจากคนรู้จัก และ
3. การละเมิดข้อมูล

P (การทำ Re-identification) = P (การทำ Re-identification | การพยายามทำ Re-identification) x P (การพยายามทำ Re-identification)

โดยที่ P (การทำ Re-identification | การพยายามทำ Re-identification) หมายถึงความน่าจะเป็นของการทำ Re-identification ได้สำเร็จ สมมุติว่ามีการพยายามในการทำ Re-identification ตามที่กล่าวไว้ก่อนหน้านี้เราสามารถกำหนด P (การทำ Re-identification | การพยายามทำ Re-identification) เป็น (1 /ขนาดระดับเทียบเท่าน้อยที่สุดในชุดข้อมูล)

ดังนั้น P (การทำ Re-identification) = (1 / ขนาดระดับเทียบเท่าน้อยที่สุดในชุดข้อมูล) x P (การพยายามทำ Re-identification)

สำหรับสถานการณ์ที่ 1 ซึ่งเป็นการโจมตีโดยเจตนาโดยใช้ข้อมูลภายใน สมมุติว่าฝ่ายที่ได้รับชุดข้อมูลพยายามทำ Re-identification ในการประมาณค่า P (การพยายามทำ Re-identification): ความน่าจะเป็นของการพยายามทำ Re-identification ปัจจัยที่ต้องพิจารณา รวมถึงขอบเขตของการควบคุมการบรรเทาผลกระทบที่มี ตลอดจนแรงจูงใจและทรัพยากรของผู้โจมตี ตารางต่อไปนี้แสดงค่าตัวอย่าง ขอย้ำอีกครั้งว่าฝ่ายที่ทำ Anonymization กับชุดข้อมูลจะต้องเลือกค่าที่เหมาะสมที่จะใช้

สถานการณ์ที่ 1—การโจมตีโดยเจตนาโดยใช้ข้อมูลภายใน P (การพยายามทำ Re-identification) = P(การโจมตีโดยเจตนาโดยใช้ข้อมูลภายใน)

		แรงจูงใจและทรัพยากรของผู้โจมตี		
		ต่ำ	**กลาง**	**สูง**
ขอบเขตของการควบคุมบรรเทาผลกระทบ	**สูง**	0.03	0.05	0.1
	กลาง	0.2	0.25	0.3
	ต่ำ	0.4	0.5	0.6
	ไม่มี	1.0	1.0	1.0

ปัจจัยที่ส่งผลต่อแรงจูงใจและทรัพยากรของผู้โจมตีอาจรวมถึง:

1. การจงใจที่จะละเมิดสัญญา (สมมติว่ามีสัญญาป้องกันการทำ Re-identification)
2. ข้อจำกัดทางการเงินและเวลา
3. การรวมบุคคลที่เป็นเป้าสายตาของสาธารณะชน (เช่น คนดัง) หรือข้อมูลที่ละเอียดอ่อน (เช่น ข้อมูลเครดิต) ในชุดข้อมูล
4. ความง่ายในการเข้าถึงข้อมูล "ที่เชื่อมโยงได้" ไม่ว่าจะเป็นข้อมูลสาธารณะหรือเป็นข้อมูลส่วนตัว ซึ่งอาจช่วยให้สามารถทำ Re-identification กับชุดข้อมูลที่ผ่านการทำ Anonymization แล้วได้

ปัจจัยที่ส่งผลต่อขอบเขตของการควบคุมบรรเทาผลกระทบ ได้แก่:

1. โครงสร้างองค์กร
2. การควบคุมด้านการบริหาร/กฎหมาย (เช่น สัญญา)
3. การควบคุมทางเทคนิคและกระบวนการ

สำหรับสถานการณ์ที่ 2 ซึ่งเป็นการรับรู้โดยไม่ได้ตั้งใจโดยคนรู้จัก เราสมมุติว่าฝ่ายที่ได้รับชุดข้อมูลได้ทำ Re-identification กับข้อมูลและรู้ตัวตนของเจ้าของข้อมูลโดยไม่ได้ตั้งใจในขณะที่ตรวจสอบชุดข้อมูล ซึ่งอาจเกิดขึ้นได้เนื่องจากฝ่ายดังกล่าวมีความรู้เพิ่มเติมเกี่ยวกับเจ้าของข้อมูลเนื่องจากความสัมพันธ์ของพวกเขา (เช่น เป็นเพื่อน เพื่อนบ้าน ญาติ เพื่อนร่วมงาน ฯลฯ) ในการประมาณค่า P (การพยายามทำ Re-identification): ความน่าจะเป็นของการพยายามทำ Re-identification ปัจจัยหลักที่ต้องพิจารณาคือความเป็นไปได้ที่ผู้รับข้อมูลจะรู้จักใครบางคนในชุดข้อมูล

สถานการณ์ที่ 2—การรับรู้โดยไม่ได้ตั้งใจโดยคนรู้จัก P (การพยายามทำ Re-identification) = P (ผู้รับข้อมูลที่รู้จักบุคคลคนหนึ่งในชุดข้อมูล)

+สำหรับสถานการณ์ที่ 3 สามารถประมาณความน่าจะเป็นของการละเมิดข้อมูลที่เกิดขึ้นในระบบ ICT ของผู้รับข้อมูลได้โดยอิงตามสถิติที่มีอยู่เกี่ยวกับความชุกของการละเมิดข้อมูลในอุตสาหกรรมของผู้รับข้อมูล ซึ่งขึ้นอยู่กับสมมติฐานที่ว่าผู้โจมตีที่ได้รับชุดข้อมูลจะพยายามทำ Re-identification

สถานการณ์ที่ 3—การละเมิดข้อมูล

P (การพยายามทำ Re-identification) = P (การละเมิดข้อมูลในอุตสาหกรรมของผู้รับข้อมูล)

ความน่าจะเป็นสูงสุดในสามสถานการณ์ควรใช้เป็น P (การพยายามทำ Re-identification)

P (การพยายามทำ Re-identification) = (P (การโจมตีโดยใช้ข้อมูลภายใน) สูงสุด, P (ผู้รับข้อมูลที่รู้จักบุคคลภายในชุดข้อมูล), P (การละเมิดข้อมูลในอุตสาหกรรมของผู้รับข้อมูล))

รวมทั้งหมดเข้าด้วยกัน

P (การทำ Re-identification) = (1 / ขนาดระดับเทียบเท่าต่ำสุดในชุดข้อมูล) x P (การพยายามทำ Re-identification) = (1 / k) x P (การพยายามทำ Re-identification) สำหรับชุดข้อมูลที่ผ่านการทำ *k*-anonymization แล้ว

ในกรณีที่ P (การพยายามทำ Re-identification) = (P (การโจมตีโดยใช้ข้อมูลภายใน) สูงสุด , P (ผู้รับข้อมูลที่รู้จักบุคคลภายในชุดข้อมูล), P (การละเมิดข้อมูลในอุตสาหกรรมของผู้รับข้อมูล))

ภาคผนวก จ: เครื่องมือจัดทำข้อมูลนิรนาม

ต่อไปนี้คือตัวอย่างรายการเครื่องมือสำหรับทำ Anonymization เชิงพาณิชย์หรือโอเพ่นซอร์ส

เครื่องมือ	คำอธิบาย	URL
Amnesia	Amnesia เป็นเครื่องมือสำหรับทำ Anonymization ซึ่งเป็นซอฟต์แวร์ที่ใช้ในท้องถิ่นเพื่อทำ Anonymization กับข้อมูลส่วนบุคคลและข้อมูลละเอียดอ่อน ปัจจุบันเครื่องมือนี้รองรับการรับประกันด้วย k-anonymity และ km-anonymity	https://amnesia.openaire.eu/
Arcad DOT-Anonymizer	DOT-Anonymizer เป็นเครื่องมือที่รักษาความลับของข้อมูลทดสอบโดยการปกปิดข้อมูลส่วนบุคคล เครื่องมือนี้จะทำ Anonymization กับข้อมูลส่วนบุคคลในขณะที่ยังคงรักษารูปแบบและประเภทของข้อมูลไว้	https://www.arcadsoftware.com/dot/data-masking/dot-anonymizer/
ARGUS	ARGUS ย่อมาจาก “Anti Re- Identification General Utility System” เครื่องมือนี้ใช้วิธีการทำ Anonymization ทางสถิติที่หลากหลาย เช่น การเปลี่ยนรหัสสากล (การจัดกลุ่มหมวดหมู่) การทำ Suppression เฉพาะที่ การสุ่ม การเพิ่มสัญญาณรบกวน การรวมกลุ่มย่อย การเข้ารหัสบนและล่าง เครื่องมือนี้ยังสามารถใช้เพื่อสร้างข้อมูลสังเคราะห์ได้อีกด้วย	https://research.cbs.nl/casc/mu.htm
ARX	ARX เป็นซอฟต์แวร์โอเพ่นซอร์สสำหรับการทำ Anonymization กับข้อมูลส่วนบุคคลที่ละเอียดอ่อน	https://arx.deidentifier.org/
Eclipse	Eclipse เป็นชุดเครื่องมือจาก Privacy Analytics ที่อำนวยความสะดวกในการทำ Anonymization กับข้อมูลด้านสุขภาพ	https://privacy-analytics.com/health-data-privacy/health-data-software/
sdcMicro	sdcMicro ใช้เพื่อสร้างข้อมูลระดับย่อยที่ผ่านการทำ Anonymization แล้ว เช่น ไฟล์สำหรับการใช้งานในสาธารณะและทางวิทยาศาสตร์	https://cran.r-project.org/web/packag

เครื่องมือ	คำอธิบาย	URL
	เครื่องมือนี้รองรับวิธีการประมาณความเสี่ยงต่างๆ	es/sdcMicro/index.html UTD
UTD Anonymisation Toolbox	UT Dallas Data Security และ Privacy Lab ได้รวบรวมเทคนิคการทำ Anonymization ต่างๆ ไว้ในกล่องเครื่องมือสำหรับการใช้งานในสาธารณะ	http://cs.utdallas.edu/dspl/cgi-bin/toolbox/index.php?go=home

กิตติกรรมประกาศ

กิตติกรรมประกาศ

PDPC สิงคโปร์ และสำนักงานพัฒนาสารสนเทศและการสื่อสารของรัฐบาลสิงคโปร์ (Infocomm Media Development Authority หรือ IMDA) ขอบคุณอย่างจริงใจต่อองค์กรต่อไปนี้สำหรับความเห็นที่เป็นประโยชน์ในการพัฒนาสิ่งพิมพ์ฉบับนี้

- AsiaDPO
- BetterData Pte Ltd
- ISACA (Singapore Chapter)—Data Protection SIG
- Law Society of Singapore—Cybersecurity and Data Protection Committee (CSDPC)
- กระทรวงสาธารณะสุข (MOH)
- Replica Analytics
- Privitar Ltd
- SGTech
- Singapore Business Federation (SBF)—Digitalisation Committee
- Singapore Corporate Counsel Association (SCCA)—Data Protection, Privacy and Cybersecurity (DPPC) Chapter
- Singapore Department of Statistics (DOS)
- Smart Nation and Digital Government Group (SNDGO)

คู่มือฉบับนี้มีการอ้างอิงจากคู่มือต่อไปนี้

- UKAN. *The Anonymisation Decision Making Framework 2nd Edition: European Practitioners' Guide, by* Mark Elliot, Elaine Mackey and Kieron O'Hara, 2020.
- CSIRO and OAIC. *The De-Identification Decision-Making Framework*, by Christine M O'Keefe, Stephanie Otorepec, Mark Elliot, Elaine Mackey and Kieron O'Hara, 18 September 2017.
- IPC. *De-identification Guidelines for Structured Data*, June 2016, *https://www.ipc.on.ca/wp-content/uploads/2016/08/Deidentification-Guidelines-for-Structured-Data.pdf.*
- El Emam, K. *Guide to the De-Identification of Personal Health Information*, CRC Press, 2013.
- Article 29 Data Protection Working Party (European Commission). "Opinion 05/2014 on Anonymisation Techniques". 10 April 2014, *http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf*.
- NIST. *NISTIR 8053: De-Identification of Personal Information*, by S L Garfinkel, October 2015, *http://nvlpubs.nist.gov/nistpubs/ir/2015/NIST.IR.8053.pdf.*

แนวทางสำหรับ
การจัดทำข้อมูลนิรนามขั้นพื้นฐาน

สำนักงานคณะกรรมการคุ้มครองข้อมูลส่วนบุคคล (สคส.)

120 หมู่ 3 ชั้น 7 อาคารรัฐประศาสนภักดี ศูนย์ราชการเฉลิมพระเกียรติ 80 พรรษา
5 ธันวาคม 2550 ถนนแจ้งวัฒนะ แขวงทุ่งสองห้อง เขตหลักสี่ กรุงเทพฯ 10210

เว็บไซต์: www.pdpc.or.th FB: "PDPC Thailand"

ไปรษณีย์อิเล็กทรอนิกส์: saraban@pdpc.or.th

www.pdpc.or.th