

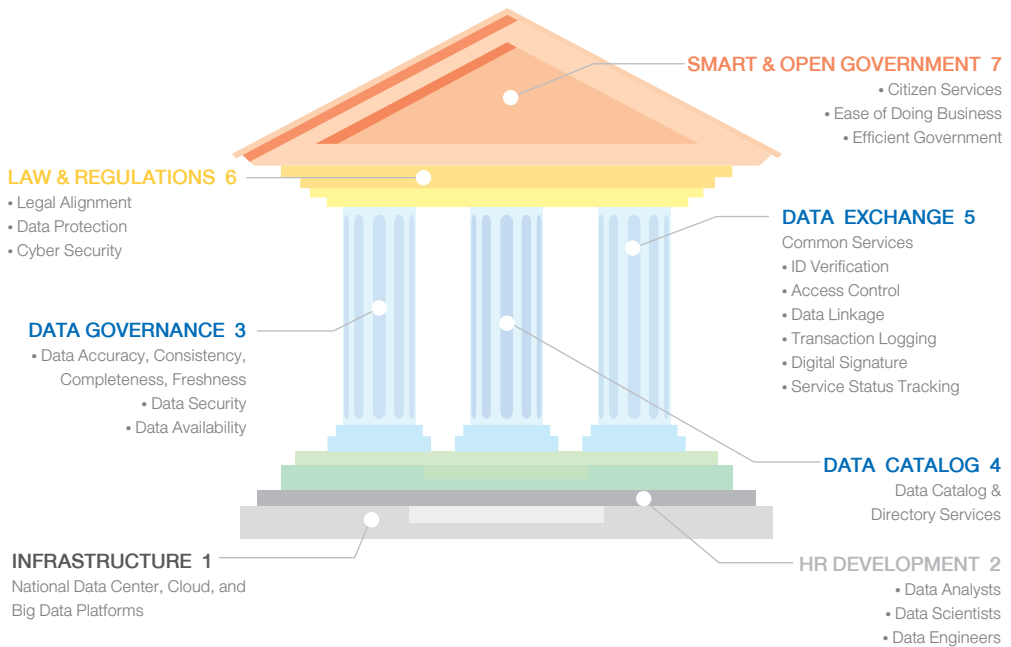


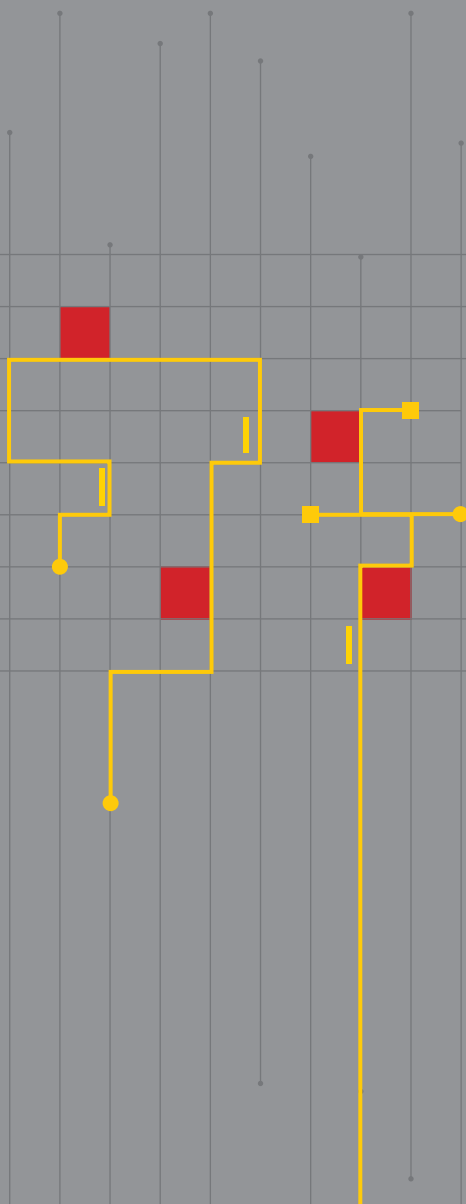
กรอบการวิเคราะห์ข้อมูล ขนาดใหญ่ภาครัฐ

Government Big Data Analytics Framework

GBDi เป็นหน่วยงานภายใต้สำนักงานส่งเสริมเศรษฐกิจดิจิทัล กระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม ได้รับมติเห็นชอบจากคณะรัฐมนตรีให้จัดตั้งขึ้นเมื่อวันที่ ๗ พฤษภาคม ๒๕๖๒ ให้มีการปฏิบัติงานส่งเสริมและสนับสนุนการใช้ประโยชน์จากข้อมูลขนาดใหญ่ (Big Data) เพื่อให้หน่วยงานภาครัฐและภาคเอกชน มีศักยภาพในการจัดเก็บ ประมวลผล และวิเคราะห์ข้อมูลได้อย่างเป็นระบบ บุคลากรของ GBDi ส่วนมากเป็น นักวิทยาศาสตร์และวิศวกรข้อมูล ที่ให้บริการแก่องค์กรต่าง ๆ ในด้านการใช้ประโยชน์ข้อมูลเชิงวิเคราะห์ เพื่อประกอบการตัดสินใจและวางแผนนโยบาย (Data Intelligence) นอกจากนี้ GBDi ยังมีเป้าหมายในการสนับสนุน บ่มเพาะ และพัฒนานวัตกรรมด้านข้อมูล รวมถึงให้คำปรึกษาในการวิเคราะห์ ข้อมูลด้านเทคนิคผ่านกรณีศึกษา เพื่อสนับสนุนการสร้างนวัตกรรมด้าน การใช้ประโยชน์ข้อมูลขนาดใหญ่ของ ประเทศไทย

กรอบการวิเคราะห์ข้อมูลขนาดใหญ่ภาครัฐ (Government Big Data Analytics Framework)





Government Big Data Analytics Framework

ที่มาและความสำคัญ	5
บทนำ	6
บทที่ 1 โครงสร้างพื้นฐานด้านสารสนเทศภาครัฐ	11
1.1 กรอบมาตรฐานดาต้าเซ็นเตอร์/คลาวด์ภาครัฐ	12
1.2 โครงสร้างพื้นฐานเพื่อการประมวลผลของหน่วยงานภาครัฐ	16
1.3 ระบบคลาวด์กลางภาครัฐ	19
บทที่ 2 การพัฒนาบุคลากรภาครัฐด้านการวิเคราะห์ข้อมูล	23
2.1 กรอบการพัฒนาบุคลากร	23
2.2 เครือข่ายมหาวิทยาลัยด้านวิทยาศาสตร์ข้อมูล	28
บทที่ 3 การกำกับดูแลข้อมูลภาครัฐ	33
3.1 กรอบธรรมาภิบาลข้อมูล	34
3.2 ขั้นตอนการจัดทำกรอบเพื่อกำกับดูแลข้อมูลดิจิทัลในเบื้องต้น	40
บทที่ 4 รายการข้อมูลภาครัฐ	43
4.1 บริการรายการข้อมูลภาครัฐ	43
4.2 การทำงานร่วมกับหน่วยงานภาครัฐเรื่องการกำกับดูแลข้อมูล	51
บทที่ 5 การให้บริการด้านการแลกเปลี่ยนข้อมูล	55
5.1 มาตรฐานข้อมูล	55
5.2 การปรับปรุงข้อมูลเพื่อเตรียมความพร้อม	58
5.3 ระบบบริการและแลกเปลี่ยนข้อมูลกลาง	59
บทที่ 6 พระราชบัญญัติคุ้มครองข้อมูลส่วนบุคคล พ.ศ. 2562	65
6.1 ภาพรวมของพระราชบัญญัติคุ้มครองข้อมูลส่วนบุคคล พ.ศ. 2562	65
6.2 พระราชกฤษฎีกากำหนดหน่วยงานและกิจการที่ผู้ควบคุมข้อมูลส่วนบุคคลไม่อยู่ภายใต้บังคับแห่งพระราชบัญญัติคุ้มครองข้อมูลส่วนบุคคล พ.ศ. 2562 และ พ.ศ. 2563	67
6.3 หลักการสำคัญตามพระราชบัญญัติคุ้มครองข้อมูลส่วนบุคคล พ.ศ. 2562	68

บทที่ 7 การเป็นรัฐบาลเปิดและอัจฉริยะ	73
7.1 การเพิ่มศักยภาพด้านการท่องเที่ยวเมืองรอง	73
7.2 การวิเคราะห์สื่อสังคมออนไลน์และพัฒนาเครื่องมือเพื่อรวบรวมข้อมูล สำรวจความคิดเห็นของภาครัฐ : กรณีศึกษาเน็ตประชารัฐ	78
7.3 การวิเคราะห์กระบวนการงานภาครัฐเพื่อนำไปสู่การพัฒนาศูนย์บริการ เบ็ดเสร็จ	81
7.4 การพยากรณ์ปริมาณผลผลิตพืชเศรษฐกิจหลัก 5 ประเภท ของประเทศไทย	83
7.5 การเตรียมความพร้อมสำหรับการบริหารจัดการทรัพยากร ทางด้านสาธารณสุขในเขตพัฒนาพิเศษภาคตะวันออก (EEC)	85
7.6 การวิเคราะห์ข้อมูล จำนวน และประเภทครอบครัวในประเทศ	88
7.7 การสกัดหัวข้องานเขียนเชิงวิชาการเพื่อศึกษาแนวโน้มหัวข้อ งานวิจัยไทย	91
7.8 ระบบสารสนเทศเพื่อการแลกเปลี่ยนข้อมูลสุขภาพ	93
7.9 การวิเคราะห์ความสัมพันธ์ระหว่างลักษณะทางประชากรของผู้เข้าพัก กับประเภทโรงแรมที่เข้าพัก	95
7.10 การหากลุ่มสมาชิกที่มีแนวโน้มในการออมเพิ่มและเลือกแผนการลงทุน ของกองทุนบำเหน็จบำนาญข้าราชการ	98
ภาคผนวก ก โครงสร้างกรอบหลักสูตร	103
ภาคผนวก ข.1 โครงสร้างและการใช้งานของรายการข้อมูล (Data Catalog)	127
ภาคผนวก ข.2 แนวทางปฏิบัติทางเทคโนโลยีสำหรับการแปลงข้อมูลให้อยู่ ในรูปแบบพร้อมใช้งาน	141

ที่มาและ ความสำคัญ

ในคราวประชุมคณะรัฐมนตรี เมื่อวันที่ 23 มกราคม 2561 นายกรัฐมนตรีได้มอบหมายให้รองนายกรัฐมนตรี (พลอากาศเอก ประจิน จั่นตอง) กำกับการขับเคลื่อนการดำเนินนโยบายการจัดการข้อมูลขนาดใหญ่ (Big Data) รวมทั้งติดตามและเร่งรัดการดำเนินการของคณะกรรมการขับเคลื่อนการบูรณาการฐานข้อมูลกลางภาครัฐ เพื่อให้การดำเนินการเป็นไปอย่างมีประสิทธิภาพและบรรลุผลสำเร็จอย่างเป็นรูปธรรม

นายกรัฐมนตรีได้มีคำสั่งสำนักนายกรัฐมนตรีที่ 55/2561 แต่งตั้งคณะกรรมการขับเคลื่อนการดำเนินนโยบายเพื่อใช้ประโยชน์ข้อมูลขนาดใหญ่ (Big Data) ศูนย์ข้อมูล (Data Center) และคลาวด์คอมพิวติ้ง (Cloud Computing) ณ วันที่ 26 กุมภาพันธ์ 2561

ในการประชุมคณะกรรมการฯ เมื่อวันที่ 11 พฤษภาคม 2561 ได้มีมติเห็นชอบให้จัดตั้งคณะอนุกรรมการ 3 คณะ ได้แก่ คณะอนุกรรมการออกแบบสถาปัตยกรรม (Enterprise Architecture) ระบบบูรณาการข้อมูลของภาครัฐ คณะอนุกรรมการด้านกฎหมาย กฎระเบียบ เพื่อการใช้ประโยชน์ข้อมูลขนาดใหญ่ และคณะอนุกรรมการด้านการพัฒนาบุคลากร เพื่อการใช้ประโยชน์ข้อมูลขนาดใหญ่ คณะทำงานได้ดำเนินการสำรวจข้อมูลเกี่ยวกับโครงสร้างพื้นฐานจากหน่วยงานภาครัฐ จำนวน 159 แห่ง เพื่อนำไปวิเคราะห์หาช่องว่าง (Gap Analysis) และวางแผนเพื่อให้หน่วยงานภาครัฐสามารถใช้ประโยชน์จากข้อมูลขนาดใหญ่

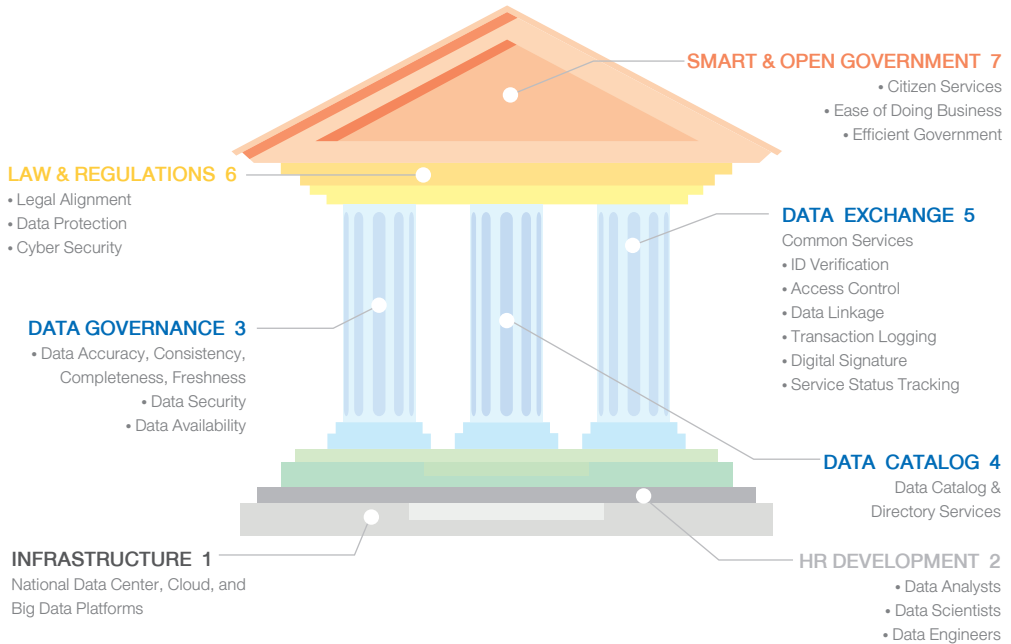
คณะทำงานของคณะอนุกรรมการออกแบบสถาปัตยกรรม (Enterprise Architecture) ระบบบูรณาการข้อมูลของภาครัฐ ได้จัดทำ (ร่าง) กรอบการวิเคราะห์ข้อมูลขนาดใหญ่ ภาครัฐ (Government Big Data Analytics Framework) โดยมีวัตถุประสงค์หลักคือ การพัฒนาการใช้ประโยชน์ข้อมูลภาครัฐให้เป็นไปในทิศทางเดียวกัน ให้สามารถบูรณาการและให้บริการข้อมูลข้ามหน่วยงานได้อย่างเป็นระบบ รวดเร็ว ปลอดภัย ควบคุมได้ ซึ่งจะนำไปสู่การวิเคราะห์และใช้ประโยชน์ข้อมูลอย่างเป็นรูปธรรม ซึ่งร่างนี้ได้รับการเห็นชอบในการประชุมคณะกรรมการฯ เมื่อวันที่ 29 มกราคม 2562 ต่อมา **สถาบันส่งเสริมการวิเคราะห์และบริหารข้อมูลขนาดใหญ่ภาครัฐ (GBDi)** ได้ปรับปรุงแก้ไขให้สมบูรณ์ชัดเจน และไม่มีข้อมูลความลับของราชการ อันจะนำไปสู่การเผยแพร่สู่สาธารณะต่อไป

บทนำ

ในปัจจุบันประเทศไทยมีความตื่นตัวเรื่องบิ๊กดาต้า (Big Data) หรือข้อมูลขนาดใหญ่อย่างกว้างขวางตามกระแสโลก องค์กรทั้งภาครัฐและภาคเอกชนเริ่มปรับตัวและเตรียมพร้อมเข้าสู่ยุคของการวิเคราะห์และใช้ประโยชน์ข้อมูลขนาดใหญ่เพื่อประกอบการตัดสินใจ (Data Driven Decision) และการดำเนินงาน (Insight to Operation) ดังนั้น เพื่อให้เกิดประโยชน์อย่างเป็นรูปธรรม องค์กรภาครัฐซึ่งเป็นเจ้าของข้อมูลสำคัญที่มีผลต่อการพัฒนาประเทศ จึงควรต้องคำนึงถึงการบูรณาการและการให้บริการข้อมูลข้ามหน่วยงานภาครัฐ ตลอดจนจนถึงการให้บริการข้อมูลที่ไม่มีชั้นความลับแก่ภาคเอกชน

คณะกรรมการขับเคลื่อนการดำเนินนโยบายเพื่อใช้ประโยชน์ข้อมูลขนาดใหญ่ (Big Data) ศูนย์ข้อมูล (Data Center) และคลาวด์คอมพิวติ้ง (Cloud Computing) ได้มีมติเห็นชอบต่อกรอบการวิเคราะห์ข้อมูลขนาดใหญ่ภาครัฐ (Government Big Data Analytic Framework) และให้แต่ละกระทรวงนำกรอบฉบับนี้ไปศึกษารายละเอียดและทดลองใช้ ทั้งยังได้มีมติให้จัดตั้งหน่วยงานใหม่ที่ใช้ชื่อว่า **สถาบันส่งเสริมการวิเคราะห์และบริหารข้อมูลขนาดใหญ่ภาครัฐ (Government Big Data Institute: GBDi)** เป็นหน่วยงานภายใต้สำนักงานส่งเสริมเศรษฐกิจดิจิทัล (Digital Economy Promotion Agency: depa) (องค์การมหาชน) GBDi ได้รับการจัดตั้งเพื่อส่งเสริมและสนับสนุนการดำเนินงานด้านบิ๊กดาต้าสำหรับหน่วยงานภาครัฐ นอกจากนี้ GBDi จะดำเนินงานร่วมกับสำนักงานพัฒนารัฐบาลดิจิทัลในการวางกรอบด้านการใช้ประโยชน์ข้อมูลของประเทศโดยมีวัตถุประสงค์หลัก คือ การพัฒนาการใช้ประโยชน์ข้อมูลภาครัฐให้เป็นไปในทิศทางเดียวกัน สามารถบูรณาการและให้บริการข้อมูลข้ามหน่วยงานได้อย่างเป็นระบบ รวดเร็ว ปลอดภัย และควบคุมได้ ซึ่งจะนำไปสู่การวิเคราะห์และใช้ประโยชน์ข้อมูลอย่างเป็นรูปธรรมในอนาคต ดังอธิบายในเอกสารฉบับนี้

การออกแบบแนวคิดของการให้บริการข้อมูลภาครัฐ (Government Data Service Framework) แสดงในภาพที่ 1



ภาพที่ 1 แนวคิดของการให้บริการข้อมูลภาครัฐ (Government Data Service Framework)



สามารถติดตามข้อมูลด้าน Big Data ได้ทาง youtube



กรอบการให้บริการข้อมูลภาครัฐประกอบด้วย 7 ส่วนงานหลักที่จะสร้างให้เกิดระบบนิเวศข้อมูลที่สามารถถูกใช้ประโยชน์ร่วมกันได้อย่างเป็นระบบ โดยทุกส่วนงานควรเริ่มดำเนินการไปพร้อม ๆ กัน ดังนี้

1. การพัฒนาโครงสร้างพื้นฐานด้านสารสนเทศภาครัฐ (Government Infrastructure) ที่เหมาะสมมีประสิทธิภาพในการใช้งาน ทั้งในมิติด้านความต่อเนื่องของบริการ การรักษาความปลอดภัย และการแลกเปลี่ยนเชื่อมโยงข้อมูลอย่างเป็นระบบ สถาปัตยกรรมโครงสร้างพื้นฐานควรเป็นแบบแบ่งปัน ที่ผู้รับบริการมองเสมือนว่าเป็นโครงสร้างเดียวที่ใช้ร่วมกัน ในขณะที่ทางกายภาพโครงสร้างพื้นฐานเป็นแบบกระจายที่มีการเข้าใช้คลาวด์ ผสมกับการสร้างดาต้าเซเตอร์ หน่วยงานภาครัฐสามารถวางแผนการปรับโครงสร้างพื้นฐานการประมวลผลที่เหมาะสมให้กับหน่วยงานได้ โดยทำการวิเคราะห์ช่องว่าง (Gap Analysis) ตามแนวทางที่กระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคมกำหนดขึ้น ผลสรุปการออกแบบโครงสร้างพื้นฐานภาครัฐแสดงในบทที่ 1

2. การพัฒนาบุคลากรภาครัฐด้านการวิเคราะห์ข้อมูล (Human Resource Development) โดยการวางแผนและดำเนินงานด้านการพัฒนาบุคลากรเพื่อเตรียมความพร้อมสำหรับการประยุกต์ใช้ข้อมูลในการวิเคราะห์เพื่อวางแผนยุทธศาสตร์ การปฏิบัติงานของเจ้าหน้าที่ การบริหารงานของหน่วยงานภาครัฐ และการให้บริการประชาชน โดยเน้นการอบรมเพื่อพัฒนานักวิเคราะห์ข้อมูล นักวิทยาศาสตร์ข้อมูล และวิศวกรข้อมูล โดยอาศัยความร่วมมือกับเครือข่ายมหาวิทยาลัย และสำนักงานคณะกรรมการข้าราชการพลเรือน ราชการและกระบวนการสร้างบุคลากรแสดงในบทที่ 2

3. การกำกับดูแลข้อมูลภาครัฐ (Government Data Governance) เพื่อส่งเสริมให้หน่วยงานภาครัฐสามารถกำกับดูแลข้อมูลของตนให้มีความถูกต้อง ตรงกัน สมบูรณ์ และทันสมัย รวมทั้งอยู่ในสภาพที่พร้อมใช้งานโดยมีการกำกับดูแลความปลอดภัยของข้อมูลให้ได้มาตรฐานตามชั้นความลับ เพื่อให้การใช้ประโยชน์และการบริการข้อมูลเกิดขึ้นได้อย่างเป็นรูปธรรม ทั้งนี้กรอบการกำกับดูแลข้อมูลภาครัฐแสดงในบทที่ 3

4. การพัฒนาระบบรายการข้อมูลภาครัฐ (Government Data Catalog)

โดยรวบรวมรายการชุดข้อมูลสำคัญพร้อมคำอธิบายชุดข้อมูล (Metadata) จากหน่วยงานภาครัฐทั้งหมด เพื่อจัดสร้างเป็นรายการชุดข้อมูลกลางภาครัฐ (Government Data Catalog) จากนั้นพัฒนาระบบสารสนเทศเพื่อบริหารรายการชุดข้อมูลกลางพร้อมทั้งกลไกการสืบค้นในมิติต่าง ๆ เพื่อให้บริการสืบค้นข้อมูล (Directory Service) ที่ชาญฉลาด ทั้งนี้ระบบรายการข้อมูลภาครัฐนี้จะเหมือนสมุดหน้าเหลือง (Yellow Pages) สำหรับให้บุคลากรภาครัฐค้นหาแหล่งข้อมูลเพื่อใช้ประกอบการวิเคราะห์ หรือ ให้บริการได้อย่างสะดวก รวมไปถึงการออกแบบเครื่องมือต่าง ๆ ที่หน่วยงานภาครัฐต้องใช้ในการกำกับดูแลและให้บริการข้อมูลของหน่วยงานได้อย่างเป็นระบบ โดยรายละเอียดแสดงในบทที่ 4

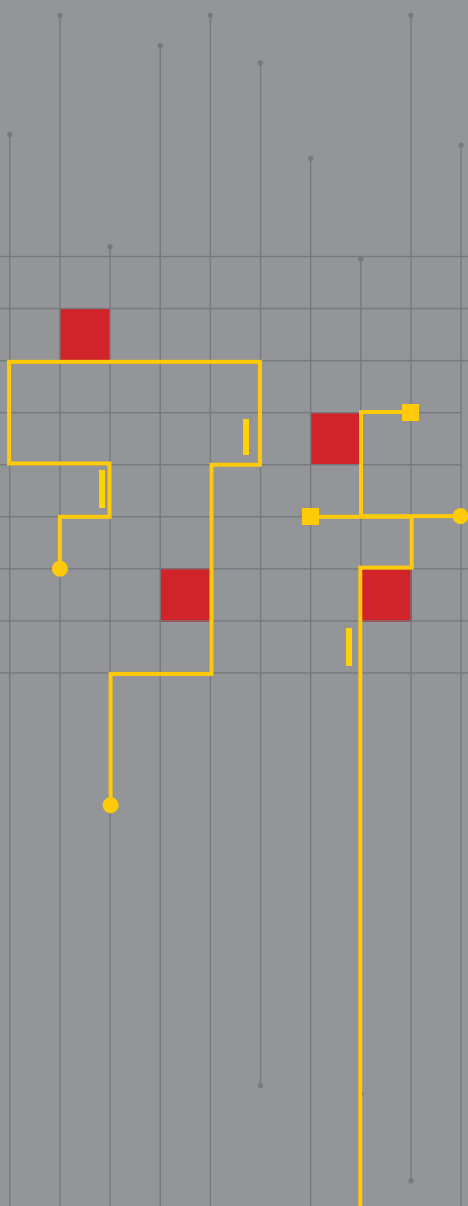
5. การให้บริการด้านการแลกเปลี่ยนข้อมูล (Data Exchange) สำหรับ

การวิเคราะห์ข้อมูลขนาดใหญ่ (Big Data Analytics) ของหน่วยงานภาครัฐ โดยมีการพัฒนาบริการพื้นฐานกลาง เช่น การยืนยันตัวตน การจำกัดสิทธิในการเข้าถึงข้อมูล บริการส่งข้อมูลข้ามหน่วยงานผ่านระบบสารสนเทศ บันทึกธุรกรรม (Transaction) การเข้าถึงข้อมูลเพื่อเป็นหลักฐาน และการบริการข้อมูลประเภทอื่น โดยการให้บริการด้านการแลกเปลี่ยนข้อมูลแสดงในบทที่ 5

6. การศึกษากระบวนการทางกฎหมายและกฎกระทรวงต่าง ๆ (Law and Regulations) ของหน่วยงานผู้ให้บริการข้อมูล รวมถึงศึกษาพระราชบัญญัติคุ้มครองข้อมูลส่วนบุคคล และพระราชบัญญัติการรักษาความมั่นคงปลอดภัยไซเบอร์ พ.ศ. 2562 เพื่อออกแบบแนวทางปฏิบัติที่เหมาะสมเพื่อเอื้อให้เกิดการใช้ประโยชน์ข้อมูลและการบูรณาการข้อมูลข้ามหน่วยงานได้อย่างเป็นรูปธรรม โดยแนวคิดดังกล่าวนำเสนอในบทที่ 6

7. การเป็นรัฐบาลเปิดและอัจฉริยะ (Smart and Open Government)

ให้สอดคล้องกับเป้าหมายระยะยาวของกรอบการวิเคราะห์ข้อมูลขนาดใหญ่ภาครัฐ คือ การที่กระทรวง รัฐวิสาหกิจ และส่วนราชการทุกแห่งมีศักยภาพในการจัดเก็บและประมวลผลข้อมูลได้อย่างเป็นระบบ รวมทั้งสามารถใช้ประโยชน์ข้อมูลขนาดใหญ่และเชื่อมโยง/แลกเปลี่ยนข้อมูลได้อย่างเป็นรูปธรรมเพื่อให้เกิดการให้บริการประชาชนที่มีประสิทธิภาพ พัฒนาศักยภาพการแข่งขันของอุตสาหกรรมโดยรวมและสร้างให้เกิดการบริการราชการที่มีประสิทธิภาพ ต้นทุนลดลง และเพิ่มความโปร่งใสเป็นธรรม โดยตัวอย่างการใช้ประโยชน์ข้อมูลเชิงวิเคราะห์เพื่อก้าวเข้าสู่การเป็นรัฐบาลอัจฉริยะในแสดงในบทที่ 7



โครงสร้างพื้นฐานด้านสารสนเทศภาครัฐ
Government Infrastructure



โครงสร้างพื้นฐานด้านสารสนเทศภาครัฐ Government Infrastructure

การวางแผนภาพรวมของโครงสร้างพื้นฐานดาต้าเซ็นเตอร์ (Data Center) และ คลาวด์ (Cloud) มีความสำคัญในการเพิ่มประสิทธิภาพของระบบงานภาครัฐโดยรวม ทั้งในด้านการให้บริการระบบงานที่ต่อเนื่องและการรักษาความปลอดภัยทางไซเบอร์ให้ได้ตามมาตรฐานสากล การจัดการทรัพยากรประมวลผลที่คุ้มค่า การวางแผนงบประมาณที่มีประสิทธิภาพ และการแลกเปลี่ยนข้อมูลข้ามหน่วยงานอย่างเป็นระบบ รูปแบบแนวคิด (Conceptual Design) สำหรับโครงสร้างพื้นฐานการประมวลผลภาครัฐจะเป็นแบบกระจาย โดยมีการเข้าใช้คลาวด์ ผสมกับการสร้างดาต้าเซ็นเตอร์เพื่อเพิ่มความยืดหยุ่นในการพัฒนาโครงสร้างพื้นฐาน หน่วยงานที่เข้าใช้ระบบคลาวด์จะไม่จำเป็นต้องจัดซื้อครุภัณฑ์ หรือสร้างอาคารดาต้าเซ็นเตอร์ เพื่อรองรับการขยายตัวของระบบงานทั้งหมด และบุคลากรสารสนเทศสามารถมุ่งเน้น การกำกับดูแลระบบงานและข้อมูลแทนการใช้เวลาในการดูแลเครื่องประมวลผลและอุปกรณ์พื้นฐานได้ จากข้อมูลในต่างประเทศการใช้คลาวด์นั้นจะประหยัดงบประมาณได้ถึง 30-70 เปอร์เซ็นต์ ขึ้นอยู่กับลักษณะการใช้งานโดยการใช้แบบ Software-as-a-Service หรือ SaaS จะประหยัดได้สูงที่สุด¹

หน่วยงานภาครัฐแต่ละหน่วยงานมีลักษณะการใช้ทรัพยากรประมวลผลที่แตกต่างกันไปตามพันธกิจและชั้นความลับของข้อมูล อีกทั้งยังมีความพร้อมด้านสารสนเทศที่ไม่เท่ากัน โครงสร้างพื้นฐานด้านการประมวลผลจึงจะมีความแตกต่างกันออกไป ดังนั้น เพื่อให้การออกแบบโครงสร้างพื้นฐานมีความเหมาะสมกับการใช้งานของแต่ละหน่วยงาน คณะกรรมการขับเคลื่อนการดำเนินนโยบายเพื่อใช้ประโยชน์ข้อมูลขนาดใหญ่ (Big Data) ศูนย์ข้อมูล (Data Center) และคลาวด์คอมพิวติ้ง (Cloud Computing) จึงได้ขอให้หน่วยงานภาครัฐต่าง ๆ ทำการสำรวจโครงสร้างพื้นฐานที่มีอยู่เทียบกับมาตรฐานสากล และ นำข้อมูลดังกล่าวมาประเมินวิเคราะห์หาช่องว่าง (Gap Analysis) ร่วมกับผู้เชี่ยวชาญ ในกระบวนการออกแบบสถาปัตยกรรม (Architecture Design) ระบบบูรณาการข้อมูล

¹ “Cloud vs. On-Premises cost” by www.fusionalliance.com, “Comparing Costs of Managed Hosting vs. On-Prem DIY” by Rackspace, “SaaS VS On-Premise HR Systems” by www.peoplehr.com

ภาครัฐ ซึ่งผลจากการวิเคราะห์ได้ถูกนำมาใช้วางแผนเพื่อปรับปรุงโครงสร้างพื้นฐานสำหรับแต่ละหน่วยงาน รวมไปถึงการวางแผนการโอนย้ายระบบงานเก่า (Legacy System) เข้าสู่เทคโนโลยีใหม่ และการวางแผนการพัฒนาบุคลากรสารสนเทศอีกด้วย

การดำเนินงานออกแบบโครงสร้างการประมวลผลพื้นฐานในครั้งนี้ มีหน่วยงานภาครัฐเข้าร่วมให้ข้อมูลเพื่อประเมินวิเคราะห์หาช่องว่างจำนวน 159 หน่วยงาน² ข้อมูลจากการสำรวจและสัมภาษณ์ได้ถูกนำมาออกแบบกรอบโครงสร้างสารสนเทศและการประมวลผลข้อมูลพื้นฐาน กรอบดังกล่าวถูกออกแบบโดยมีวัตถุประสงค์หลักดังนี้

- เพื่อให้หน่วยงานภาครัฐสามารถเข้าถึงทรัพยากรคอมพิวเตอร์ได้อย่างรวดเร็ว และทั่วถึง ทันต่อความต้องการในการเปลี่ยนแปลงเข้าสู่รัฐบาลดิจิทัล
- เพื่อเพิ่มความต่อเนื่องในการให้บริการระบบงานภาครัฐ (Availability) ให้ถึงมาตรฐานสากลคือ ร้อยละ 99.9
- เพื่อให้การดูแลรักษาความปลอดภัยระบบงาน และข้อมูลที่มีชั้นความลับ เป็นไปตามมาตรฐานสากล
- เพื่อให้เกิดการแลกเปลี่ยนข้อมูลข้ามหน่วยงานได้อย่างมีมาตรฐานและเป็นระบบ (Repeatable Process) รองรับการใช้งานการวิเคราะห์และใช้ประโยชน์ข้อมูลขนาดใหญ่
- เพื่อประหยัดงบประมาณด้านโครงสร้างพื้นฐานการประมวลผล

1.1 กรอบมาตรฐานดาต้าเซนต์อร์/คลาวด์ภาครัฐ

กรณีที่ระบบงานของหน่วยงานภาครัฐไม่มีชั้นความลับข้อมูล หน่วยงานสามารถจัดหาหรือเช่าซื้อทรัพยากรประมวลผลได้อย่างอิสระ ผู้ให้บริการจะเป็นองค์กรใดก็ได้ อย่างไรก็ตาม ในกรณีที่ระบบงานมีระดับชั้นความลับข้อมูลประเภท “ข้อมูลความลับราชการ” หรือ “ข้อมูลส่วนบุคคล” ผู้ให้บริการการประมวลผลจะต้องเป็นองค์กรที่รัฐบาลสามารถกำกับดูแลได้ ส่วนในกรณีที่ระบบงานมีชั้นความลับระดับที่เป็นข้อมูลความมั่นคง ผู้ให้บริการการประมวลผลจะต้องเป็นองค์กรที่รัฐบาลสามารถกำกับดูแลได้ที่มีมาตรฐานเรื่องความปลอดภัยสูงมาก ทั้งด้านความปลอดภัยพื้นที่และความปลอดภัยเครือข่าย

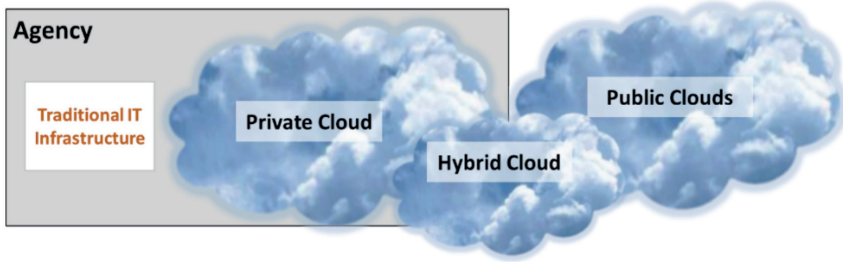
² ผลการสำรวจเมื่อเดือนกันยายนถึงตุลาคม 2561 โดยคณะทำงานภายใต้คณะอนุกรรมการออกแบบสถาปัตยกรรม (Architecture Design) ระบบบูรณาการข้อมูลภาครัฐ

ระบบงานและฐานข้อมูลที่มีชั้นความลับของภาครัฐจะต้องถูกติดตั้งอยู่ในดาต้าเซ็นเตอร์ที่ได้มาตรฐานสากลทั้งในด้านพื้นที่ แหล่งจ่ายไฟ การรักษาความปลอดภัยข้อมูล (Data Protection) และความปลอดภัยไซเบอร์ (Cyber Security) รวมทั้งมีการรับประกันบริการ (Service Level Agreement - SLA) ที่เหมาะสม สามารถบริหารจัดการเครื่องประมวลผลและที่จัดเก็บข้อมูลให้ได้ตามชั้นความลับของระบบงานและฐานข้อมูล โดยหน่วยงานภาครัฐในระดับต่าง ๆ สามารถเลือกใช้บริการดาต้าเซ็นเตอร์/คลาวด์ได้ ดังนี้

- ดาต้าเซ็นเตอร์ที่ได้มาตรฐาน Tier 4 (หรือใกล้เคียงโดยยกเว้นเรื่องการจ่ายไฟจาก 2 แหล่ง)³ ที่หน่วยงานจะสร้างขึ้นหรือมีอยู่แล้ว ซึ่งหน่วยงานจะกำกับดูแลพื้นที่ ระบบประมวล ระบบเครือข่าย และระบบรักษาความปลอดภัยด้วยตัวเอง
- ผู้ให้บริการภายนอก (Outsource) บริการดาต้าเซ็นเตอร์/คลาวด์ ไปยังผู้ให้บริการคลาวด์สาธารณะ (Public Cloud) โดยระบบงานและฐานข้อมูลที่มีชั้นความลับต้องใช้บริการจากองค์กรที่รัฐบาลสามารถกำกับดูแลได้ ในขณะที่ระบบงานและฐานข้อมูลที่ไม่ใช่ชั้นความลับ (ข้อมูลสาธารณะ) สามารถถูกนำไปติดตั้งในพื้นที่ของผู้ให้บริการจากภาคเอกชนหรือจากต่างประเทศได้อย่างอิสระ ทั้งนี้หน่วยงานภาครัฐสามารถที่จะเลือกผู้ให้บริการดาต้าเซ็นเตอร์/คลาวด์สำหรับระบบงานต่าง ๆ ได้เอง อาจจะใช้บริการจากผู้ให้บริการหลายเจ้าได้ แต่ให้เป็นไปตามเกณฑ์เรื่องชั้นความลับข้อมูลที่กำลังกล่าวในข้างต้น
- สร้างระบบคลาวด์ผสมเอง (Hybrid Cloud) มีทั้งดาต้าเซ็นเตอร์และระบบคลาวด์ส่วนตัว (Private Cloud) และมีบางส่วนเชื่อมต่อไปยังคลาวด์สาธารณะเป็นแบบคลาวด์ผสม การบริหารจัดการเรื่องชั้นความลับของระบบงานและฐานข้อมูลให้เป็นไปตามเกณฑ์เรื่องชั้นความลับข้อมูลที่กำลังกล่าวในข้างต้น รูปแบบสถาปัตยกรรมแสดงในภาพที่ 1.1

³ หนึ่งในคุณสมบัติของคุณภาพคลัสเตอร์ประมวลผล ระดับที่ 4 (สูงมาก) คือ มีระบบสำรองการจ่ายไฟและทำความเย็นสามารถรองรับช่วงไฟตกได้นาน 96 ชั่วโมง อ้างอิงจาก มาตรฐานและแนวทางปฏิบัติการออกแบบโครงสร้างพื้นฐานทางด้านสารสนเทศเพื่อการประมวลผลข้อมูลภาครัฐ ของคณะกรรมการออกแบบสถาปัตยกรรม (Architecture Design) ระบบบูรณาการข้อมูลภาครัฐ

Traditional IT Infastucture



ภาพที่ 1.1 คลาวด์แบบผสม (Hybrid Cloud)

ผู้ให้บริการภายนอก ให้บริการดาต้าเซนเตอร์/คลาวด์ ไปยังระบบคลาวด์กลาง ภาครัฐ (Government Cloud) ซึ่งในกรณีนี้กระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคมจะกำกับดูแลผู้ให้บริการคลาวด์ภาครัฐอีกต่อหนึ่ง โดยผู้ให้บริการคลาวด์ภาครัฐสามารถบริหารจัดการเช่าทรัพยากรต่อจากภาคเอกชนเพื่อสร้างเป็นมัลติคลาวด์ (Multi-Cloud) เพื่อให้บริการระบบงานและฐานข้อมูลที่ไม่มีชั้นความลับได้ โดยประเด็นเรื่องชั้นความลับข้อมูลจะถูกจัดการโดยอัตโนมัติ หน่วยงานภาครัฐไม่จำเป็นต้องบริหารผู้ให้บริการคลาวด์ (Cloud Provider) เอง ในกรณีที่หน่วยงานเลือกที่จะบริหารจัดการการเช่าซื้อระบบคลาวด์เอง ขอให้อ้างอิงราคากลางที่กำหนดโดยกระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม ในการจัดซื้อครั้งนี้

บริการเครื่องเสมือน (Virtual Machine - VM) โดยใช้ไฮเปอร์ไวเซอร์ (Hypervisor) แบบลิขสิทธิ์จากต่างประเทศ ค่าเช่าเฉพาะเครื่องเสมือนขนาดมาตรฐานราคาไม่เกิน 12,250 บาท/เดือน/เครื่อง* สำหรับเครื่องเสมือนที่มีคุณสมบัติแตกต่างออกไปให้คำนวณโดยปรับจากค่ากลาง รายละเอียดการให้บริการมาตรฐานอธิบายได้ดังต่อไปนี้

- เครื่องเสมือนขนาด 4 vCPU**, RAM 8 GB, HDD 1 TB
- Core Network ของผู้ให้บริการขั้นต่ำ 100 Gbps Domestic Bandwidth (NIX) และ 30 Gbps International Bandwidth (IIG)
- มาตรฐานศูนย์ข้อมูล ISO 27001:2013 Data Center Edition เป็นอย่างน้อย
- ระบบปฏิบัติการ (Operating System) และฐานข้อมูล (Database) แบบโอเพ่นซอร์ส (Opensource)

- บริการเสริมด้านลิขสิทธิ์ซอฟต์แวร์จากต่างประเทศ
 - o Windows Server 2016 ราคาประมาณ 800 บาทต่อเดือน*
 - o Redhat Server ราคาประมาณ 800 บาทต่อเดือน*
 - o Microsoft SQL STD edition ราคาประมาณ 12,000 บาทต่อเดือน*
 - o Microsoft SQL ENT edition ราคาประมาณ 50,000 บาทต่อเดือน*
- บริการเครื่องเสมือน (VM) โดยใช้ไฮเปอร์ไวเซอร์ (Hypervisor) แบบลิขสิทธิ์ของบริษัทไทยที่คนไทยพัฒนาต่อยอดจากไฮเปอร์ไวเซอร์โอเพนซอร์ส ตัวอย่างเช่น Openstack ค่าเช่าเฉพาะเครื่องเสมือนขนาดมาตรฐาน ไม่เกิน 8,000 บาท/เดือน/เครื่อง* สำหรับเครื่องเสมือนที่มีรายละเอียดคุณสมบัติแตกต่างออกไปให้คำนวณโดยปรับจากค่ากลาง รายละเอียดการให้บริการมาตรฐานอธิบายได้ดังต่อไปนี้

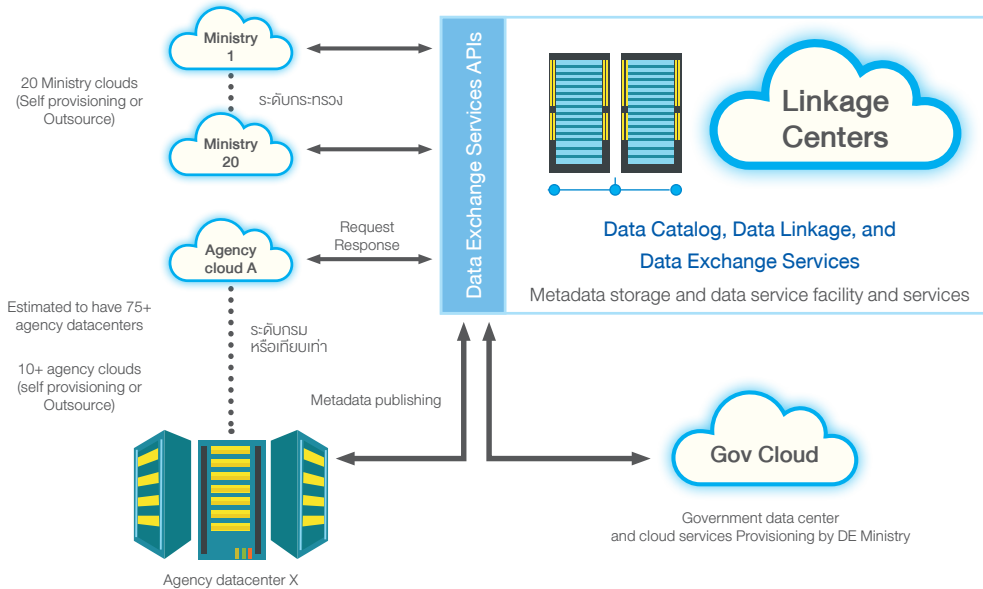
- เครื่องเสมือนขนาด 4 vCPU**, RAM 8 GB, HDD 1 TB
- Core Network ของผู้ให้บริการขั้นต่ำ 100 Gbps Domestic Bandwidth (NIX) และ 30 Gbps International Bandwidth (IIG)
- มาตรฐานศูนย์ข้อมูล ISO 27001:2013 Data Center Edition เป็นอย่างน้อย
- ระบบปฏิบัติการ (Operating System) และฐานข้อมูล (Database) แบบโอเพนซอร์ส (Opensource)
- เป็นระบบคลาวด์แบบ Multi-region แบบเดียวกับ Microsoft Azure, Google Cloud, Amazon AWS, และ GDCC ของกระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม
- บริการเสริมด้านลิขสิทธิ์ซอฟต์แวร์จากต่างประเทศ
 - o Windows Server 2016 ราคาประมาณ 800 บาทต่อเดือน*
 - o Redhat Server ราคาประมาณ 800 บาทต่อเดือน*
 - o Microsoft SQL STD edition ราคาประมาณ 12,000 บาทต่อเดือน*
 - o Microsoft SQL ENT edition ราคาประมาณ 50,000 บาทต่อเดือน*

หมายเหตุ * ราคาสำรวจจากท้องตลาด ต้นปี 2563

** Share Ratio Physical: vCPU = 1: 2-4

1.2 โครงสร้างพื้นฐานเพื่อการประมวลผลของหน่วยงานภาครัฐ

ข้อมูลที่ได้จากการสำรวจ ศึกษา และสัมภาษณ์หน่วยงานภาครัฐ จำนวน 159 แห่ง ดำเนินงานในคณะกรรมการออกแบบสถาปัตยกรรมระบบบูรณาการข้อมูลภาครัฐ ได้ออกแบบกรอบโครงสร้างสารสนเทศและการประมวลผลข้อมูลพื้นฐานภาครัฐดังแสดงใน ภาพที่ 1.2



ภาพที่ 1.2 สถาปัตยกรรมโครงสร้างพื้นฐานเพื่อการประมวลผลของหน่วยงานภาครัฐ

จากภาพดาต้าเซ็นเตอร์ และระบบคลาวด์จะถูกแบ่งออกเป็น 4 แบบ คือ 1. ดาต้าเซ็นเตอร์/คลาวด์ระดับกรมหรือหน่วยงานเทียบเท่า (Agency Cloud) 2. ดาต้าเซ็นเตอร์/คลาวด์ระดับกระทรวง (Ministry Cloud) 3. ดาต้าเซ็นเตอร์กลางเพื่อให้บริการข้อมูลภาครัฐ (Data Linkage and Exchange) และ 4. ระบบคลาวด์กลางภาครัฐ (Government Data Center and Cloud services: GDCC) โดยมีรายละเอียดดังต่อไปนี้



1. ดาต้าเซนเตอร์/คลาวด์ระดับกรมหรือหน่วยงานเทียบเท่า (Agency Cloud)

ข้อมูลจากการสำรวจประเมินทรัพยากรและบุคลากรของหน่วยงานจากผู้เชี่ยวชาญ ในคณะกรรมการออกแบบสถาปัตยกรรมระบบบูรณาการข้อมูลภาครัฐ พบว่าภาครัฐ มีหน่วยงานระดับกรมหรือเทียบเท่าจำนวน 90 (± 5) แห่ง ที่มีดาต้าเซนเตอร์หรือใช้บริการระบบคลาวด์อยู่แล้ว ซึ่งหน่วยงานเหล่านี้มีความประสงค์จะดูแลดาต้าเซนเตอร์/คลาวด์ เพื่อให้บริการระบบงานของหน่วยงานเอง โดยแบ่งได้เป็น 2 รูปแบบ ดังนี้

- มีดาต้าเซนเตอร์อยู่แล้วและวางแผนที่จะพัฒนาต่อเพื่อให้ดาต้าเซนเตอร์ได้มาตรฐานที่ใกล้เคียง Tier 4 และขยายกำลังการประมวลผลด้วยการขยายดาต้าเซนเตอร์ โดยหน่วยงานกลุ่มนี้มีประมาณ 75 (± 5) แห่ง
- มีดาต้าเซนเตอร์อยู่แล้วและวางแผนที่จะขยายกำลังการประมวลผลไปยังระบบคลาวด์ ดังแสดงในภาพที่ 1.1 หรือวางแผนย้ายไประบบคลาวด์ทั้งหมดผ่านการเช่าใช้คลาวด์จากผู้ให้บริการคลาวด์สาธารณะที่รัฐบาลกำกับดูแลได้หรือระบบคลาวด์กลางภาครัฐ (GDCC) ซึ่งหน่วยงานในกลุ่มนี้มีอีกประมาณ 10 (± 5) แห่ง

2. ดาต้าเซนเตอร์/คลาวด์ระดับกระทรวง (Ministry Cloud)

รัฐบาลมีแผนจะสร้างดาต้าเซนเตอร์คลาวด์ระดับกระทรวงทั้งสิ้น 20 แห่ง เพื่อให้บริการระดับกรมและหน่วยงานภายใต้กระทรวงที่ไม่ประสงค์จะสร้างหรือดูแลดาต้าเซนเตอร์/คลาวด์เอง โดยคลาวด์ระดับกระทรวงส่วนมากจะถูกสร้างขึ้นโดยการเช่าใช้บริการพื้นที่และทรัพยากรการประมวลผลและจัดเก็บข้อมูลจากผู้ให้บริการคลาวด์สาธารณะที่รัฐบาลสามารถกำกับดูแลได้ หรือ "ระบบคลาวด์กลางภาครัฐ" (Government Data Center and Cloud Services : GDCC) และนำทรัพยากรที่เช่าใช้เหล่านั้นมาให้บริการต่อหน่วยงานภายใต้กระทรวง สำหรับบางกระทรวงที่มีดาต้าเซนเตอร์อยู่แล้ว จะสามารถให้บริการหน่วยงานภายใต้กระทรวงได้ในลักษณะคลาวด์ส่วนตัว และยังสามารถขยายโดยการต่อเชื่อมไปยังคลาวด์สาธารณะได้อีกด้วย ดังแสดงในภาพที่ 1.1 รูปแบบบริการ มีดังนี้

- **Co-location** : หน่วยงานภายใต้กระทรวงนำทรัพยากรประมวลผลที่มีอยู่แล้วไปติดตั้งในสถานที่ของผู้ให้บริการคลาวด์ระดับกระทรวง เพื่อเพิ่มมาตรฐานด้านความปลอดภัยและความต่อเนื่องบริการของระบบงาน และลดภาระเรื่องการดูแลทรัพยากรพื้นฐาน

- **Infrastructure-as-a-Service (IaaS)** : ผู้ดูแลระบบคลาวด์กระทรวงจัดสรรเครื่องประมวลผลเสมือน (Virtual Machine) ที่เก็บข้อมูล และเครือข่ายเชื่อมต่อให้กับระบบงาน ตามที่หน่วยงานภายใต้กระทรวงร้องขอ
- **Software-as-a-Service (SaaS)** : ผู้ดูแลคลาวด์กระทรวงจัดหาหรือพัฒนาระบบซอฟต์แวร์กลางเพื่อติดตั้งในระบบคลาวด์ และจัดสรรให้ผู้ใช้สำหรับให้หน่วยงานภายใต้กระทรวงเข้าใช้งานระบบซอฟต์แวร์กลางร่วมกัน

3. ดาต้าเซนเตอร์กลางภาครัฐเพื่อให้บริการข้อมูล (Data Linkage and Exchange Services)

การใช้ประโยชน์จากข้อมูลภาครัฐในอนาคตจะเกิดขึ้นได้อย่างเป็นระบบ สะดวก รวดเร็ว และเป็นไปตามข้อบังคับ/ข้อกำหนดของหน่วยงานเจ้าของข้อมูลนั้น ภาครัฐควรสร้างดาต้าเซนเตอร์กลางสำหรับให้บริการเป็นจุดเชื่อมโยงและแลกเปลี่ยนข้อมูล ผลจากการสำรวจยังพบว่าในปัจจุบันหน่วยงานที่มีดาต้าเซนเตอร์ที่ให้บริการเชื่อมโยงข้อมูลนั้นมี 2 หน่วยงาน คือ 1) ระบบ Data Linkage ของกรมการปกครอง (ข้อมูลบุคคล) และ 2) ระบบ National Single Window ของกรมศุลกากร (ข้อมูลนำเข้า/ส่งออก) ดังนั้นการลงทุนขยายบริการที่มีอยู่เพื่อให้สามารถรองรับการแลกเปลี่ยนข้อมูลสำหรับการวิเคราะห์ข้อมูลขนาดใหญ่ และการให้บริการประชาชนแบบ One Stop Service จึงเป็นการลงทุนที่คุ้มค่า คณะอนุกรรมการออกแบบสถาปัตยกรรมได้เสนอให้ทั้งสองหน่วยงานพิจารณาขยายบริการเพื่อให้ครอบคลุมข้อมูลส่วนอื่น เช่น ข้อมูลนิติบุคคลในด้านอื่นนอกเหนือจากการนำเข้า/ส่งออก ข้อมูลด้านพลังงาน เป็นต้น และขอให้หน่วยงานที่เกี่ยวข้องหรือได้รับมอบหมาย ให้สร้างดาต้าเซนเตอร์เพื่อให้บริการเชื่อมโยงข้อมูลเพิ่มเติมให้ครบสมบูรณ์ในอนาคต นอกจากนี้ภาครัฐควรสร้างบริการบัญชีรายชื่อข้อมูลกลางภาครัฐหรือแค็ตตาล็อกข้อมูลกลาง (Government Data Catalog) และบริการสืบค้นข้อมูล (Directory Service) เพิ่มเติมเพื่อให้การใช้ประโยชน์ข้อมูลเกิดขึ้นได้อย่างเป็นระบบ

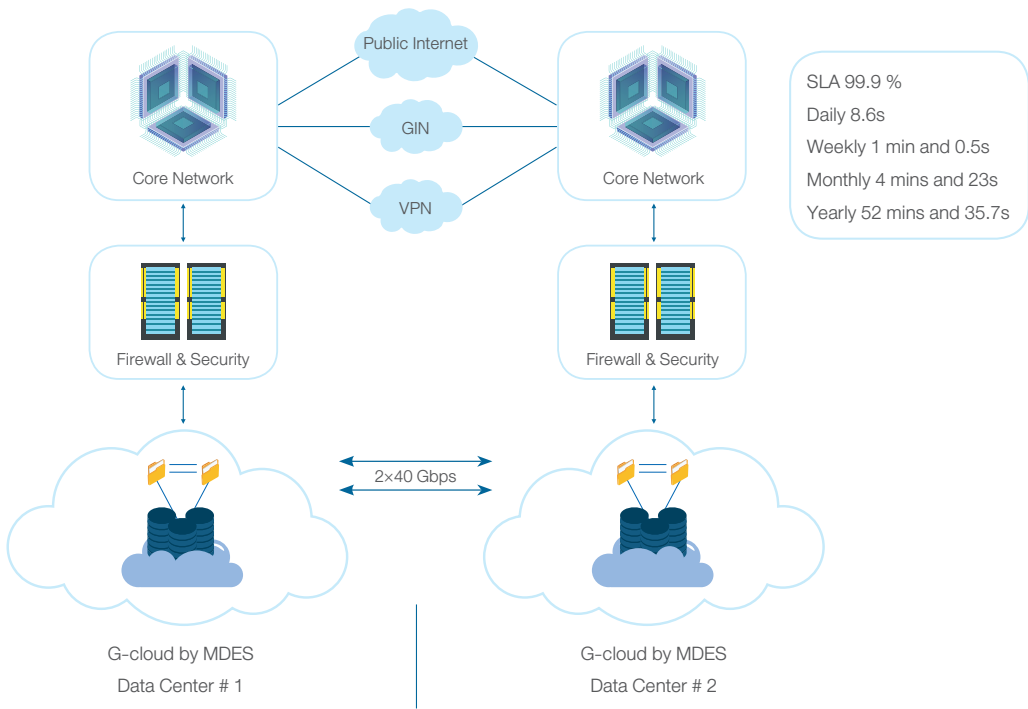
4. คลาวด์กลางภาครัฐ (Government Cloud)

บริษัท กสท โทรคมนาคม จำกัด (มหาชน) สร้างคลาวด์กลางภาครัฐขึ้น ภายใต้การกำกับดูแลของกระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม เพื่อให้ประเทศมีระบบกลางในการให้บริการคลาวด์แก่หน่วยงานภาครัฐที่มีมาตรฐานและปลอดภัย หน่วยงานภาครัฐจะสามารถเข้าถึงทรัพยากรคอมพิวเตอร์ได้อย่างรวดเร็วทันต่อความต้องการในการเปลี่ยนแปลง

เข้าสู่รัฐบาลดิจิทัล บริการของคลาวด์กลางภาครัฐจะมี 2 ระดับ คือ การให้บริการเต็มรูปแบบ (Managed Service) และการให้บริการผ่าน Self Service Portal ผู้ใช้จะสามารถเข้าถึงและปรับบริการได้เอง โดยรายละเอียดจะนำเสนอในหัวข้อต่อไป

1.3 ระบบคลาวด์กลางภาครัฐ (Government Data Center and Cloud services : GDCC)

ระบบคลาวด์กลางภาครัฐจะถูกออกแบบและพัฒนาขึ้นโดย บริษัท กสท โทรคมนาคม จำกัด (มหาชน) ภายใต้การกำกับดูแลของกระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม ซึ่งจะ เป็นระบบกลางในการให้บริการคลาวด์สำหรับหน่วยงานภาครัฐที่มีมาตรฐานและปลอดภัย หน่วยงานภาครัฐสามารถเข้าถึงทรัพยากรคอมพิวเตอร์ได้อย่างรวดเร็ว ทัวถึง และทันต่อความต้องการในการเปลี่ยนแปลงเข้าสู่รัฐบาลดิจิทัล รวมทั้งเป็นการรองรับการใช้งานการวิเคราะห์และใช้ประโยชน์ข้อมูลขนาดใหญ่อีกด้วย โดยระบบคลาวด์กลางภาครัฐ ที่สร้างขึ้นจะสามารถให้บริการผ่านเครือข่ายได้หลายประเภท รวมถึงเครือข่ายกลางภาครัฐ (Government Information Network : GIN) เครือข่ายสาธารณะและเครือข่ายส่วนตัว (Private หรือ Virtual Private Network : VPN) และมีการออกแบบให้การบริการ มีความต่อเนื่องสูงถึงร้อยละ 99.9 โดยดาต้าเซนเตอร์อย่างน้อย 2 แห่งจะถูกใช้ในการ ให้บริการระบบคลาวด์กลางร่วมกัน เมื่อดาต้าเซนเตอร์หนึ่งเกิดปัญหาขึ้นระบบงานและ ฐานข้อมูลจะถูกปรับให้ทำงานบนอีกดาต้าเซนเตอร์หนึ่งได้อย่างรวดเร็ว นอกจากนี้ ระบบ คลาวด์กลางภาครัฐจะมีข้อตกลงเรื่องการบริการ (SLA) ที่เหมาะสม มีผู้ดูแลระบบแบบ ตลอดเวลา (24 ชั่วโมงต่อวัน และ 7 วันต่อหนึ่งสัปดาห์) มีการบันทึกข้อมูล (Log) และ ให้บริการหน้าจอแสดงผลข้อมูลการใช้งาน (Dashboard) ของทุกระบบงานและฐานข้อมูล เพื่อให้หน่วยงานต่าง ๆ สามารถทำรายงานสรุปการใช้ประโยชน์ทรัพยากรประกอบการ พิจารณาปรับปริมาณทรัพยากรหรือเพิ่มประสิทธิภาพการใช้งานและงบประมาณโดยรวม อีกด้วย สถาปัตยกรรมระบบคลาวด์กลางภาครัฐ ดังแสดงในภาพที่ 1.3



ภาพที่ 1.3 สถาปัตยกรรมระบบคลาวด์กลางภาครัฐ

จากการอ้างอิงข้อมูลการสำรวจประเมินทรัพยากรของหน่วยงานโดยผู้เชี่ยวชาญ ในขณะอนุกรรมการออกแบบสถาปัตยกรรมระบบบูรณาการข้อมูลภาครัฐนั้น ระบบคลาวด์กลาง ภาครัฐ (GDCC) จะถูกวางแผนเพื่อให้บริการแก่หน่วยงานภาครัฐ ดังนี้ 1.) ให้บริการ คลาวด์ระดับกระทรวง 2.) ให้บริการหน่วยงานภาครัฐบางส่วนที่ประสบปัญหาความล่าช้าใน การจัดซื้อเพื่อขยายดาต้าเซนเตอร์/เช่าคลาวด์เพื่อขยายระบบงาน 3.) จัดสรรเพื่อเตรียม รองรับหน่วยงานภาครัฐที่ระบุว่าการใช้บริการคลาวด์กลางภาครัฐ โดยบริษัท กสท โทรคมนาคม จำกัด (มหาชน) จะให้บริการตามราคาและคุณสมบัติของผู้ให้บริการกลาง ดังที่ระบุในหัวข้อ 1.1

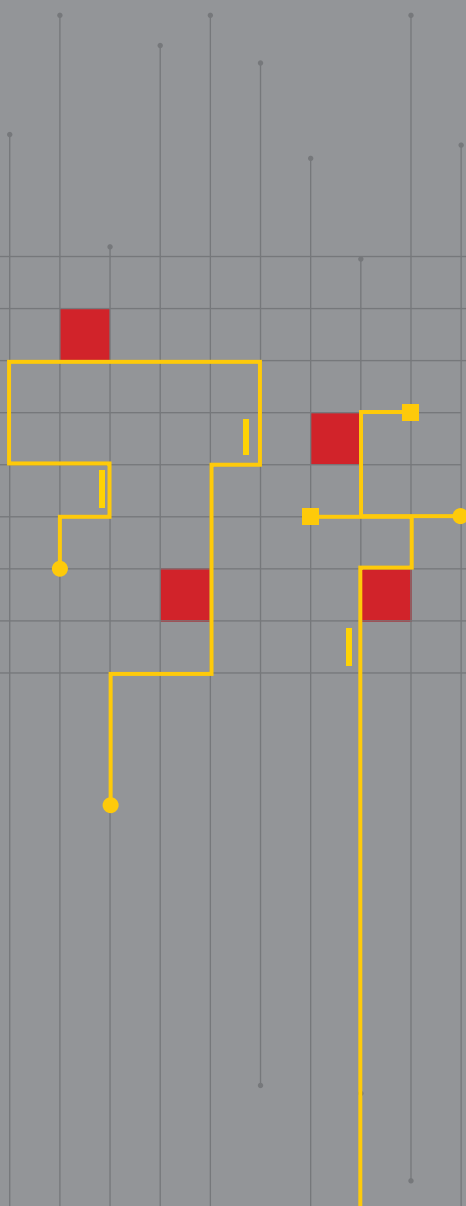
ทั้งนี้ เพื่อให้การใช้ประโยชน์ระบบคลาวด์กลางภาครัฐเกิดขึ้นได้อย่างเป็นรูปธรรม กระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคมจะให้บริการการฝึกอบรมเจ้าหน้าที่สารสนเทศภาครัฐในเรื่องการใช้งานระบบคลาวด์ แบ่งออกเป็น 3 ระดับ และมีการออกเกียรติบัตรให้กับเจ้าหน้าที่ ทั้ง 3 ระดับ คือ 1) ระดับพื้นฐานที่สามารถใช้งานระบบคลาวด์ได้ 2) ระดับกลางที่สามารถใช้งานระบบและดูแลระบบคลาวด์ขั้นพื้นฐานได้ และ 3) ระดับสูงที่สามารถใช้งานระบบดูแลระบบคลาวด์ขั้นสูงและสามารถนำไปถ่ายทอดและขยายผลต่อได้

นอกจากนี้กระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคมจะร่วมกับบริษัท กสท โทรคมนาคม จำกัด (มหาชน) ให้บริการโอนย้ายระบบงานขึ้นสู่ระบบคลาวด์กลางภาครัฐ โดยหน่วยงานภาครัฐที่จะนำระบบงานขึ้นสู่ระบบคลาวด์ ควรจะต้องจัดเตรียมงบประมาณในการโอนย้ายระบบ ส่วนค่าใช้จ่ายจะขึ้นอยู่กับความซับซ้อนของระบบงาน การโอนย้ายระบบแบ่งออกได้เป็น 3 ประเภท ดังนี้

1. Re-Deploy เป็นการย้ายระบบซอฟต์แวร์ที่ใช้อยู่ในปัจจุบันไปประมวลผลบนระบบคลาวด์โดยไม่มีการออกแบบใหม่หรือพัฒนาเพิ่มเติม ซึ่งการโอนย้ายระบบมี 2 รูปแบบ คือ 1) การโอนย้ายจาก VM ในดาต้าเซ็นเตอร์ ไปยัง VM บนคลาวด์ โดยใช้ VM เวอร์ชันเดิมของบริษัทเดียวกัน 2) การโอนย้ายจากเครื่องแม่ข่าย (Physical Server) ไปสู่ VM หรือ VM ต่างเวอร์ชันและต่างบริษัท โดยราคากลางขึ้นอยู่กับความยากง่ายและจำนวน VM

2. Re-Architect เป็นการโอนย้ายระบบงานที่อยู่บนเทคโนโลยีเก่า (Legacy System) และจะต้องได้รับการออกแบบสถาปัตยกรรมระบบซอฟต์แวร์ใหม่ ซึ่งส่งผลให้บางโมดูลต้องถูกพัฒนาใหม่บางส่วนหรือทั้งหมดก่อนที่จะสามารถวางระบบซอฟต์แวร์ลงบนคลาวด์ได้ การโอนย้ายจึงมีความยุ่งยากสูง และค่าใช้จ่ายสูงขึ้นอยู่กับเทคโนโลยี หน่วยงานจึงควรประเมินราคาและวางแผนงบประมาณล่วงหน้า

3. Replace เป็นการใช้บริการซอฟต์แวร์ที่จัดให้โดยผู้ให้บริการบนระบบคลาวด์ทดแทนการใช้ซอฟต์แวร์เดิมขององค์กร การทดแทนระบบซอฟต์แวร์ในลักษณะนี้ ผู้ดูแลระบบจะมีภาระในการโอนย้ายฐานข้อมูลเข้าสู่ระบบใหม่ หน่วยงานควรประเมินราคาและวางแผนงบประมาณล่วงหน้าในกรณีที่ข้อมูลมีความซับซ้อนสูง



การพัฒนาบุคลากรภาครัฐ
ด้านการวิเคราะห์ข้อมูล
Human Resource Development

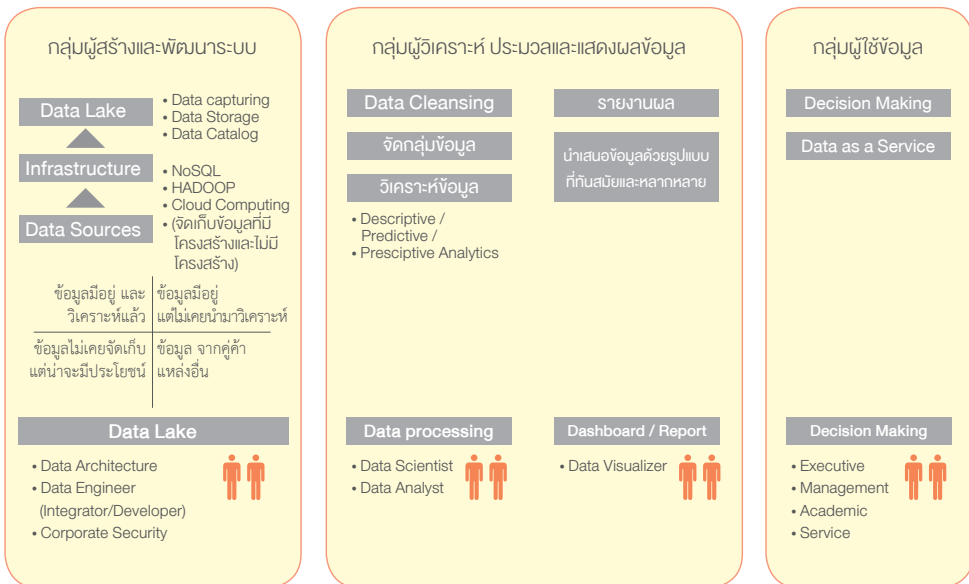


การพัฒนาบุคลากรภาครัฐด้านการวิเคราะห์ข้อมูล Human Resource Development

ในการปรับตัวและเตรียมความพร้อมเข้าสู่ยุคของการวิเคราะห์และใช้ประโยชน์ข้อมูลขนาดใหญ่เพื่อวางแผนยุทธศาสตร์ ปฏิบัติงาน และบริหารงานของภาครัฐนั้น หน่วยงานภาครัฐจำเป็นต้องวางแผนและดำเนินงานด้านการพัฒนาบุคลากรล่วงหน้า โดยเน้นการอบรมเพื่อสร้างนักวิทยาศาสตร์ข้อมูล วิศวกรข้อมูล และนักวิเคราะห์ข้อมูล เนื่องจากบุคลากรในสาขาดังกล่าวเป็นที่ต้องการสูงมาก การจัดจ้างอัตราใหม่ในจำนวนที่เพียงพอจึงเป็นเรื่องที่ท้าทาย สำนักงานคณะกรรมการข้าราชการพลเรือนร่วมกับกระทรวงดิจิทัล เพื่อเศรษฐกิจและสังคมจึงได้เสนอกลไกการพัฒนาบุคลากรเพื่อการใช้ประโยชน์ข้อมูลขนาดใหญ่ ดังนี้

2.1 กรอบการพัฒนาบุคลากร

ข้อเสนอการพัฒนาบุคลากรเพื่อการใช้ประโยชน์ข้อมูลขนาดใหญ่ ได้กำหนดเป้าหมายว่า ภายในปี 2565 บุคลากรภาครัฐของไทยจะตระหนักถึงคุณค่าของข้อมูลและสามารถนำข้อมูลขนาดใหญ่มาสร้างมูลค่าเพิ่ม และมีการพัฒนาเครือข่ายการทำงานอย่างเป็นพันธมิตรธุรกิจระหว่างภาครัฐและเอกชน (Public-Private Partnership: PPP) รวมทั้งมีการกำหนดนโยบาย ยุทธศาสตร์ และการบริหารจัดการที่ตั้งอยู่บนข้อมูลและหลักฐานเชิงประจักษ์ กลุ่มเป้าหมายของการพัฒนาประกอบด้วย 3 กลุ่ม ดังแสดงในภาพที่ 2.1



ภาพที่ 2.1 กลุ่มเป้าหมายการพัฒนาบุคลากรด้านการใช้ประโยชน์ข้อมูลขนาดใหญ่

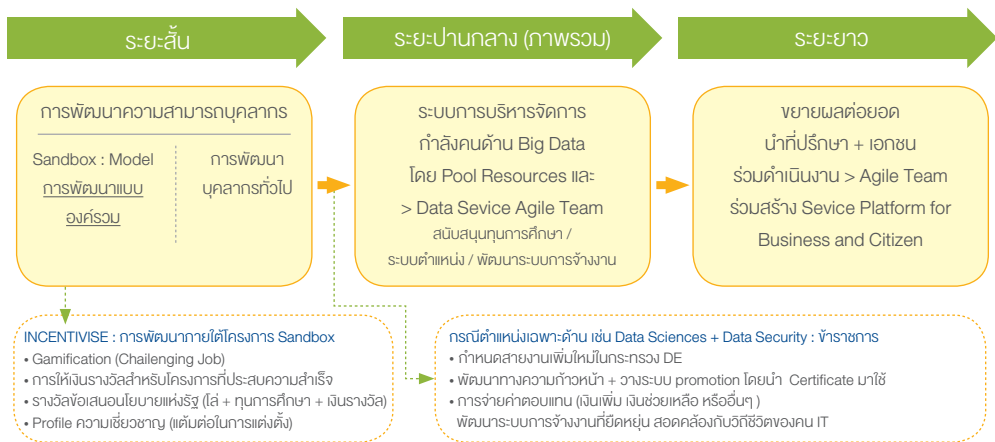
ภาพที่ 2.1 แสดงกลุ่มงานและทักษะหลักที่ต้องการเพื่อทำโครงการข้อมูลขนาดใหญ่ ภาครัฐให้ได้ประสิทธิผลโดยแบ่งเป็น 3 เรื่อง คือ งานด้านการสร้างและดูแลระบบสารสนเทศ ข้อมูลขนาดใหญ่ งานด้านการวิเคราะห์และแสดงผลข้อมูล และงานด้านการนำเอาข้อมูลไป ใช้ประโยชน์ประกอบการตัดสินใจ ดังนั้น กลุ่มเป้าหมายของการพัฒนาบุคลากรจึงแบ่งออกเป็น 3 กลุ่ม เพื่อให้สอดคล้องกับงาน ดังนี้

1) กลุ่มผู้ใช้ข้อมูล (Business Domain) ได้แก่ ผู้บริหารระดับสูง ผู้อำนวยการกอง ผู้ทำงานด้านนโยบายและวิชาการ ผู้ทำงานด้านบริการ มีหน้าที่กำหนดโจทย์หรือประเด็น ที่ต้องการวิเคราะห์ข้อมูล และนำสารสนเทศที่เป็นผลจากการวิเคราะห์และนำเสนอข้อมูล มาใช้ประกอบการดำเนินงาน

2) กลุ่มผู้วิเคราะห์ ประมวลผลและแสดงผลข้อมูล ได้แก่ นักวิทยาศาสตร์ข้อมูล (Data Scientist) นักวิเคราะห์ข้อมูล (Data Analyst) และนักนิเทศน์ข้อมูล (Data Visualizer) มีหน้าที่ในการนำข้อมูลมาจัดกลุ่ม วิเคราะห์ และประมวลผล พร้อมทั้งพัฒนาภาพแบบ แสดงผลข้อมูลหรือ Dashboard สำหรับการนำเสนอข้อมูล

3) กลุ่มผู้สร้างและพัฒนาระบบ ได้แก่ วิศวกรข้อมูล (Data Engineer) สถาปนิกข้อมูล (Data Architect) นักวิเคราะห์ธุรกิจ (Business Analyst) ผู้จัดการโครงการ (Project Manager) และเจ้าหน้าที่เทคโนโลยีสารสนเทศด้านความปลอดภัยไซเบอร์ (Corporate Security IT Operator) มีหน้าที่ในการออกแบบและพัฒนาโครงสร้างพื้นฐานด้านข้อมูล รวมทั้งดูแลและบริหารจัดการข้อมูลให้อยู่ในสภาพที่พร้อมใช้งานอย่างต่อเนื่องและปลอดภัย

นอกจากนี้ กรอบแนวทางการดำเนินงานการพัฒนาบุคลากรในการใช้ประโยชน์ข้อมูลขนาดใหญ่ จะถูกแบ่งออกเป็นระยะสั้น ระยะกลาง และระยะยาว เพื่อให้การดำเนินการบรรลุเป้าหมายอย่างเป็นขั้นตอน ดังแสดงในภาพที่ 2.2



ภาพที่ 2.2 กรอบการดำเนินงาน 3 ระยะ

ระยะสั้น : การพัฒนาแพลตฟอร์ม (Platform) เสริมสร้างทักษะที่เรียกว่า Sandbox สำหรับการพัฒนาบุคลากรในระยะสั้น เริ่มตั้งแต่การกำหนดโจทย์ การพัฒนาโครงสร้างและระบบข้อมูล การวิเคราะห์และประมวลผลข้อมูล การแสดงผลข้อมูลเชิงประจักษ์ตลอดไปจนถึงการนำเสนอสารสนเทศและองค์ความรู้ที่ได้จากการวิเคราะห์และประมวลผลมาใช้ประกอบการตัดสินใจและการดำเนินงานในรูปแบบต่าง ๆ แพลตฟอร์มดังกล่าวเน้นการพัฒนาบุคลากรผ่านหลักสูตรที่นำรูปแบบการพัฒนาแบบการเรียนรู้แบบใช้โครงงานเป็นฐาน (Project Based Learning) มาปรับใช้ เพื่อให้ผู้เข้ารับการฝึกอบรมสามารถ **“ทำได้ ทำเป็น”** มากกว่าเรียนรู้จากทฤษฎี โดยการดำเนินงานร่วมกับหลักสูตรนักบริหารภาครัฐเพื่อการบูรณาการการพัฒนาประเทศไทย ตามกรอบการปฏิรูปประเทศ ยุทธศาสตร์ชาติ และการสร้างความสามัคคีปรองดอง (ป.ย.ป.) โดยมีหน่วยงานหลักที่เกี่ยวข้อง ได้แก่ กระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม (ดศ.) สำนักงานคณะกรรมการดิจิทัลเพื่อเศรษฐกิจและสังคมแห่งชาติ (สดช.) สำนักงานคณะกรรมการข้าราชการพลเรือน (สกพ.) สำนักงานพัฒนารัฐบาลดิจิทัล (องค์กรมหาชน) (สพร.) และสถาบันสถาบันส่งเสริมการวิเคราะห์และบริหารข้อมูลขนาดใหญ่ภาครัฐ (GBDi)

ระยะกลาง : การวางแผนการบริหารจัดการกำลังคนเพื่อการใช้ประโยชน์ข้อมูลขนาดใหญ่ที่มีความคล่องตัว สามารถสนับสนุนการพัฒนานโยบาย การตัดสินใจ การบริหารจัดการ การให้บริการ และการพัฒนานวัตกรรมภาครัฐ มีกรอบการดำเนินงานดังนี้

- วางระบบการบริหารจัดการกำลังคนแบ่งเป็น 3 กลุ่มหลัก ดังนี้ 1) กลุ่มหน่วยงานที่มีความพร้อมด้านบุคลากรที่สามารถดูแลบริหารจัดการระบบข้อมูล และมีบุคลากรที่มีความสามารถด้านการวิเคราะห์ ประมวลและแสดงผลข้อมูล แต่ยังคงมีความต้องการการสนับสนุนเชิงเทคนิคหรือการดำเนินการในบางประการ 2) กลุ่มหน่วยงานที่อยู่ระหว่างการพัฒนา มีบุคลากรที่มีความเข้าใจและสามารถให้ข้อมูลความต้องการได้ชัดเจนระดับหนึ่ง ต้องการการสนับสนุนเชิงเทคนิคในการสร้างและพัฒนาระบบและการวิเคราะห์ ประมวลและแสดงผลข้อมูล 3) กลุ่มหน่วยงานที่ขาดแคลนบุคลากรด้านข้อมูล แต่มีความจำเป็นต้องนำข้อมูลมาใช้ประโยชน์เพื่อการตัดสินใจ การกำหนดนโยบาย หรือการบริหารจัดการต่าง ๆ

- นำบุคลากรที่มีทักษะและความเชี่ยวชาญเฉพาะทางจากหน่วยงานต่าง ๆ มารวมไว้ภายในหน่วยงานเดียว เช่น พัฒนารูปแบบการทำงานในลักษณะเป็น “ทีมที่ปรึกษา” (Agile Team) เพื่อให้บริการหน่วยงานภาครัฐ โดยอาจนำที่ปรึกษาภายนอกหรือหน่วยงานเอกชนมาร่วมดำเนินงานในรูปแบบต่าง ๆ หรือให้มีการพัฒนารูปแบบการจ้างงานใหม่นอกเหนือจาก “ข้าราชการ/พนักงานราชการ” เพื่อดึงดูด/จูงใจ และอาจมีการเพิ่ม “สายงานเฉพาะทาง” สำหรับผู้ปฏิบัติงานที่ต้องใช้องค์ความรู้เกี่ยวกับการวิเคราะห์ข้อมูล มีการดูแลความก้าวหน้าในอาชีพ มีการสร้างโอกาสการพัฒนาที่ต่อเนื่อง มีการให้ค่าตอบแทนที่เหมาะสม มีการปรับกรอบอัตราค่าจ้างให้สอดคล้องกับลักษณะงานที่มีความซับซ้อนและความต้องการบุคลากรที่มีสมรรถนะสูงและมีความเป็นมืออาชีพ ทั้งนี้อาจมีการนำ วุฒิบัตรหรือประกาศนียบัตรรับรองมาตรฐานหรือสมรรถนะด้านการวิเคราะห์ข้อมูลมาใช้ในการบริหารจัดการกำลังคนด้วย
- พัฒนาขีดความสามารถของผู้ใช้ข้อมูลกลุ่มต่าง ๆ ได้แก่ ผู้บริหารระดับสูง ผู้อำนวยการกอง ผู้ทำงานด้านนโยบายและวิชาการ และผู้ทำงานด้านบริการให้สามารถกำหนดโจทย์หรือประเด็นที่ต้องการวิเคราะห์ข้อมูล และนำเสนอสารสนเทศที่เป็นผลจากการวิเคราะห์และนำเสนอข้อมูลมาใช้ในการประกอบการดำเนินงานได้อย่างเหมาะสม รวมทั้งยกระดับศักยภาพ ผู้ปฏิบัติงานด้านเทคโนโลยีสารสนเทศของหน่วยงาน โดยอย่างน้อย ให้มีความสามารถในการกำกับควบคุมงานจ้างที่ปรึกษา (Project Management) และในการบริหารจัดการข้อมูลอย่างปลอดภัยและมีธรรมาภิบาล (Data Governance)
- นำกลไกการให้ทุนรัฐบาลมาใช้สนับสนุนการสร้างและพัฒนากำลังคนด้านการวิเคราะห์และใช้ประโยชน์ข้อมูลขนาดใหญ่ทั้งในระยะสั้นและระยะยาว
- วางระบบการบริหารองค์ความรู้และเก็บรวบรวมข้อมูลการดำเนินงาน เพื่อสร้างความต่อเนื่องในการดำเนินงาน

ระยะยาว : การต่อยอดขยายผลโดยมีเป้าหมายเพื่อให้เกิดการแบ่งปันและสร้างประโยชน์ในทรัพยากรข้อมูลของทุกภาคส่วน เพื่อการพัฒนาต่อยอดการสร้างมูลค่าเพิ่มให้กับระบบเศรษฐกิจ สร้างความมั่นคงให้กับสังคม และสร้างการพัฒนาที่ยั่งยืน โดยเน้นส่งเสริมสนับสนุนการพัฒนาพันธมิตรธุรกิจระหว่างภาครัฐและเอกชน ทั้งภายในและระหว่างประเทศภายใต้เจตจำนงค์ร่วมกัน โดยร่วมสร้าง Open Government Data Platform for Business and Citizen ให้ความรู้และสร้างภูมิคุ้มกันเพื่อให้ประชาชนมีความเข้าใจที่ถูกต้องเกี่ยวกับการดูแลความปลอดภัยของข้อมูลส่วนบุคคล และการนำข้อมูลเปิดภาครัฐมาใช้เพื่อพัฒนาต่อยอดและสร้างมูลค่าเพิ่ม

2.2 เครื่องข่ายมหาวิทยาลัยด้านวิทยาศาสตร์ข้อมูล

การสนับสนุนการดำเนินการตามกรอบการพัฒนาบุคลากรเพื่อการใช้ประโยชน์ข้อมูลขนาดใหญ่ สถาบันส่งเสริมการวิเคราะห์และบริหารข้อมูลขนาดใหญ่ภาครัฐ (GBDI) ได้สร้างความร่วมมือกับเครือข่ายมหาวิทยาลัยชั้นนำในประเทศไทย เพื่อวางแผนในการสร้างให้บุคลากรภาครัฐมีทักษะในด้านการวิเคราะห์ข้อมูลและสามารถปฏิบัติงานได้อย่างเป็นรูปธรรม โดยการสร้างเครือข่ายมหาวิทยาลัยด้านวิทยาศาสตร์ข้อมูล (Data Science) เพื่อการพัฒนาบุคลากรภาครัฐในการวิเคราะห์และจัดการข้อมูลขนาดใหญ่

เครือข่ายมหาวิทยาลัยดังกล่าว มีวัตถุประสงค์เพื่อระดมผู้เชี่ยวชาญและอาจารย์ผู้สอนในมหาวิทยาลัยมาร่วมออกแบบโครงสร้างกรอบหลักสูตรกลาง⁴ และจัดการเรียนการสอนแบบนอกห้องเรียนให้กับเจ้าหน้าที่ของหน่วยงานภาครัฐ กรอบหลักสูตรกลางรวมถึงเนื้อหาที่จำเป็น (Materials) วิธีการเรียนการสอนที่เหมาะสม (Pedagogy) ทักษะ/องค์ความรู้ที่จะได้ (Learning Outcome) และการวัดผล/ประเมินทักษะ (Assessment) จะเน้นการสร้างทักษะด้านวิทยาศาสตร์ข้อมูล ทั้งการวิเคราะห์และจัดการข้อมูลขนาดใหญ่ และเน้นพัฒนาให้คนทำเป็นมากกว่ารู้เพียงแค่เรียนเท่านั้น ผู้สอนจากแต่ละมหาวิทยาลัยจะปรับเนื้อหากรอบหลักสูตรกลางให้เข้ากับโจทย์ปัญหาของหน่วยงานภาครัฐแต่ละแห่งตามความเหมาะสม

การเรียนการสอนจะมีรูปแบบการฝึกเชิงปฏิบัติ (Bootcamp) โดยมีแนวทางการดำเนินการ ดังนี้

⁴ ในการประชุมหารือเรื่องการสร้างเครือข่ายมหาวิทยาลัยด้านวิทยาศาสตร์ข้อมูลเพื่อการพัฒนาบุคลากรภาครัฐในการวิเคราะห์และจัดการข้อมูลขนาดใหญ่ ครั้งที่ 2 เมื่อวันที่ 29 มีนาคม 2563 ณ สำนักงานปลัดกระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม

1) มหาวิทยาลัยเสนอกลไกการเรียนการสอนในรูปแบบ Bootcamp ที่เหมาะสม เช่น

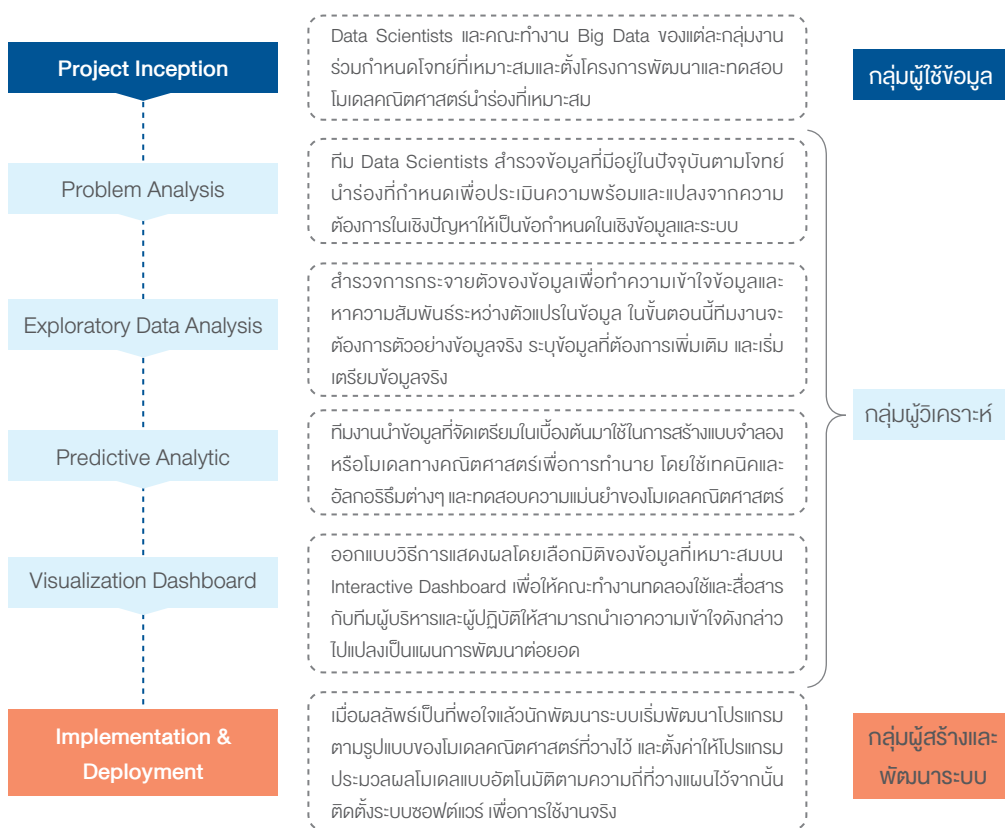
- การบรรยาย (Lecture) โดยใช้การเรียนรู้แบบใช้ปัญหาเป็นฐาน (Problem-Based Learning) เนื้อหาอยู่ในบริบทของงานที่เกี่ยวข้องโดยตรงกับผู้เรียน
- การประชุมเชิงปฏิบัติการที่เน้นการร่วมกันลงมือปฏิบัติ (Hands-On Workshop)
- การเรียนรู้แบบใช้โครงงานเป็นฐาน (Project-Based Learning) โดยใช้โครงงานที่อยู่ในบริบทของงานที่เกี่ยวข้องโดยตรงกับผู้เรียน
- การอภิปรายหรือการสัมมนาแก้ปัญหาในห้องเรียน
- การนำเสนอผลงาน (Showcases) ตอนจบหลักสูตรอย่างเป็นทางการ

2) แต่ละ Bootcamp ดำเนินงานเป็นอิสระต่อกัน แต่อยู่ภายในกรอบของโครงสร้างหลักสูตรกลางที่เครือข่ายมหาวิทยาลัยเห็นชอบ แต่ละ Bootcamp มีการออกแบบชุดองค์ความรู้และทักษะ ที่รวมเรียกว่าผลลัพธ์การเรียนรู้ (Learning Outcome) ที่เหมาะสม

3) มีการให้ประกาศนียบัตร (Certificate) สำหรับผู้ที่ผ่านการทดสอบความรู้ความเข้าใจและการลงมือปฏิบัติในแต่ละหลักสูตร

ทั้งนี้ การออกแบบกรอบหลักสูตรจะต้องสอดคล้องกับกระบวนการการดำเนินโครงการบิกดาต้าที่เหมาะสม ขั้นตอนแรกของกระบวนการดำเนินงานดังกล่าว นักวิทยาศาสตร์ข้อมูล ผู้บริหาร และคณะทำงานด้านบิกดาต้าของแต่ละกลุ่มงานจะร่วมกันกำหนดโจทย์ที่เหมาะสมสำหรับการนำร่อง (Project Inception) และสำรวจข้อมูลที่มีอยู่ในปัจจุบันตามโจทย์นำร่องที่กำหนดเพื่อประเมินความเป็นไปได้และแปลงจากความต้องการในเชิงปัญหาให้เป็นข้อกำหนดในเชิงข้อมูลและระบบ (Problem Analysis) จากนั้นนักวิทยาศาสตร์ข้อมูลทำการศึกษาการกระจายตัวของข้อมูลและหาความสัมพันธ์ระหว่างตัวแปรโดยใช้ตัวอย่างต้องการตัวอย่างข้อมูลจริงเพื่อระบุข้อมูลที่ต้องการเพิ่มเติม (Exploratory Data Analysis) และเตรียมข้อมูลมาใช้ในการสร้างแบบจำลองหรือโมเดลทางคณิตศาสตร์เพื่อการทำนายโดยใช้เทคนิคและอัลกอริทึม (Algorithm) ต่าง ๆ รวมทั้งทดสอบความแม่นยำของโมเดลคณิตศาสตร์ที่สร้างขึ้น (Predictive Analytics) และออกแบบวิธีการแสดงผลโดยเลือกมิติของข้อมูลที่เหมาะสม เพื่อให้คณะทำงานทดลองใช้และสื่อสารกับทีมผู้บริหารและผู้ปฏิบัติ

ให้สามารถนำเอาความเข้าใจดังกล่าวไปแปลงเป็นแผนการพัฒนาต่อยอด (Visualization Dashboard) จากนั้นนักพัฒนาระบบสารสนเทศจะสามารถพัฒนาโปรแกรมตามรูปแบบของโมเดลคณิตศาสตร์ที่วางไว้ ตั้งค่าโปรแกรมให้ประมวลผลโมเดลแบบอัตโนมัติตามความถี่ที่วางแผนไว้ และติดตั้งระบบซอฟต์แวร์ เพื่อการใช้งานจริง (Implementation and Deployment) ซึ่งขั้นตอนต่าง ๆ ดังกล่าวจะต้องการบุคลากร 3 กลุ่มอย่างชัดเจน ดังแสดงในภาพที่ 2.3



ภาพที่ 2.3 ขั้นตอนการทำโครงการปิกดาต้า

กรอบหลักสูตรได้รับการออกแบบสำหรับกลุ่มเป้าหมายทั้ง 3 กลุ่ม ดังนี้ [รายละเอียดของกรอบหลักสูตรแสดงในภาคผนวก ก.]

1) กลุ่มผู้ใช้ข้อมูล ได้แก่ ผู้บริหารระดับสูง ผู้อำนวยการกอง ผู้ทำงานด้านนโยบายและวิชาการ และผู้ทำงานด้านบริการ ควรเข้าอบรมสูตร ดังนี้

- Basic Data Analysis in Public Management
- Data Design for Executives

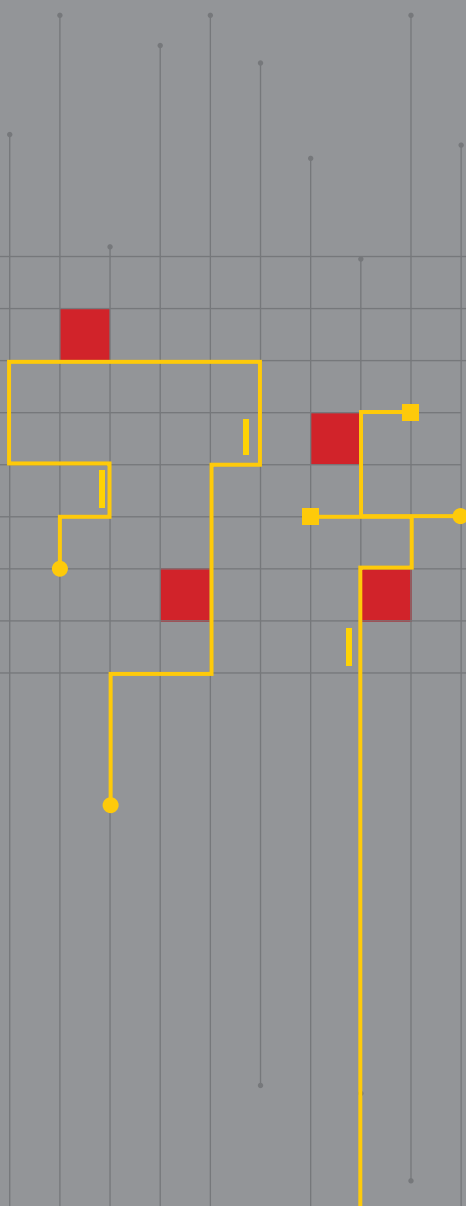
2) กลุ่มผู้วิเคราะห์ ประมวลและแสดงผลข้อมูล ได้แก่ นักวิทยาศาสตร์ข้อมูล (Data Scientist), นักวิเคราะห์ข้อมูล (Data Analyst), และ Data Visualizer ควรเข้าอบรมสูตร ดังนี้

- Basic Data Science
- Intermediate Data Science
- Advanced Data Science

3) กลุ่มผู้สร้างและพัฒนาระบบ ได้แก่ วิศวกรข้อมูล (Data Engineer), สถาปนิกข้อมูล (Data Architect), นักวิเคราะห์ธุรกิจ (Business Analyst), หัวหน้าโครงการ (Project Manager), ผู้รักษาความปลอดภัยสารสนเทศ (Corporate Security), เจ้าหน้าที่สารสนเทศ (IT Operator) ควรเข้าอบรมสูตร ดังนี้

- Basic Data Engineering
- Intermediate Data Engineering
- Advance Data Engineering

นอกจากนี้ GBDi ได้สร้างความร่วมมือกับสถาบันพัฒนาบุคลากรดิจิทัลภาครัฐ สำนักงานพัฒนารัฐบาลดิจิทัล เพื่อเตรียมความพร้อมสำหรับระบบ E-Learning ด้านการวิเคราะห์ข้อมูลขนาดใหญ่ในอนาคต และได้ทำบันทึกข้อตกลงร่วมกับเครือข่ายมหาวิทยาลัยชั้นนำของประเทศ จำนวน 16 มหาวิทยาลัย และคาดหวังว่าจะมีการขยายเครือข่ายในปีต่อ ๆ ไปอย่างต่อเนื่อง โดยมหาวิทยาลัยที่ร่วมโครงการจะเสนอรายชื่อผู้สอนที่มีความเชี่ยวชาญและมีประสบการณ์ในการทำงานด้านการวิเคราะห์ข้อมูลขนาดใหญ่ เพื่อเข้าร่วมในการออกแบบโครงสร้างหลักสูตรกลางพร้อมจัดทำสื่อการสอนพื้นฐาน รวมทั้งวิธีการประเมิน ทักษะ/องค์ความรู้ย่อยต่าง ๆ โดย GBDi จะเป็นเจ้าภาพในการบริหารจัดการและจัดสัมมนาเรื่องวิธีการสอนบุคลากรภาครัฐให้กับผู้สอนจากมหาวิทยาลัยในเครือข่าย รวมทั้งออกแบบวิธีการประเมินผู้สอนที่เหมาะสม



การกำกับดูแลข้อมูลภาครัฐ Government Data Governance

บทที่ 3



การกำกับดูแลข้อมูลภาครัฐ Government Data Governance

พระราชบัญญัติ การบริหารงานและการให้บริการภาครัฐผ่านระบบดิจิทัล พ.ศ. 2562 ได้กำหนดให้หน่วยงานภาครัฐทุกแห่งมีหน้าที่ในการกำกับดูแลข้อมูลขององค์กรให้ถูกต้อง ตรงกัน สมบูรณ์ เป็นปัจจุบัน ปลอดภัย และพร้อมใช้ตามกรอบธรรมาภิบาลข้อมูล (Data Governance Framework) ของสำนักงานพัฒนารัฐบาลดิจิทัล (สพร.) โดยมีวัตถุประสงค์เพื่อนำไปสู่การให้บริการตามภารกิจขององค์กรอย่างมีประสิทธิภาพ และสนับสนุนให้เกิดการใช้ประโยชน์ข้อมูลเชิงวิเคราะห์ประกอบการตัดสินใจวางแผน บริหารงาน การจัดทำยุทธศาสตร์และกำหนดนโยบาย รวมทั้งการพัฒนากระบวนการทำงาน ทำให้มีประสิทธิภาพและเพิ่มความสามารถในการให้บริการประชาชนได้อย่างเป็นรูปธรรม

นอกจากนี้ การกำกับดูแลข้อมูลที่เหมาะสมจะเอื้อให้หน่วยงานภาครัฐสามารถ แลกเปลี่ยนข้อมูลระหว่างกันได้อย่างมีมาตรฐาน และพร้อมสำหรับการให้บริการข้อมูลเปิด (Open Data) อันจะเป็นประโยชน์แก่สาธารณะและนักวิจัยในประเทศเป็นอย่างมาก เนื้อหาในบทนี้นำเสนอบทสรุปเรื่องการกำกับดูแลข้อมูลภาครัฐ

รายละเอียดของกรอบธรรมาภิบาลข้อมูล



Data Governance [ตอนที่ 3] 
การให้บริการข้อมูลและการกำกับดูแลข้อมูล

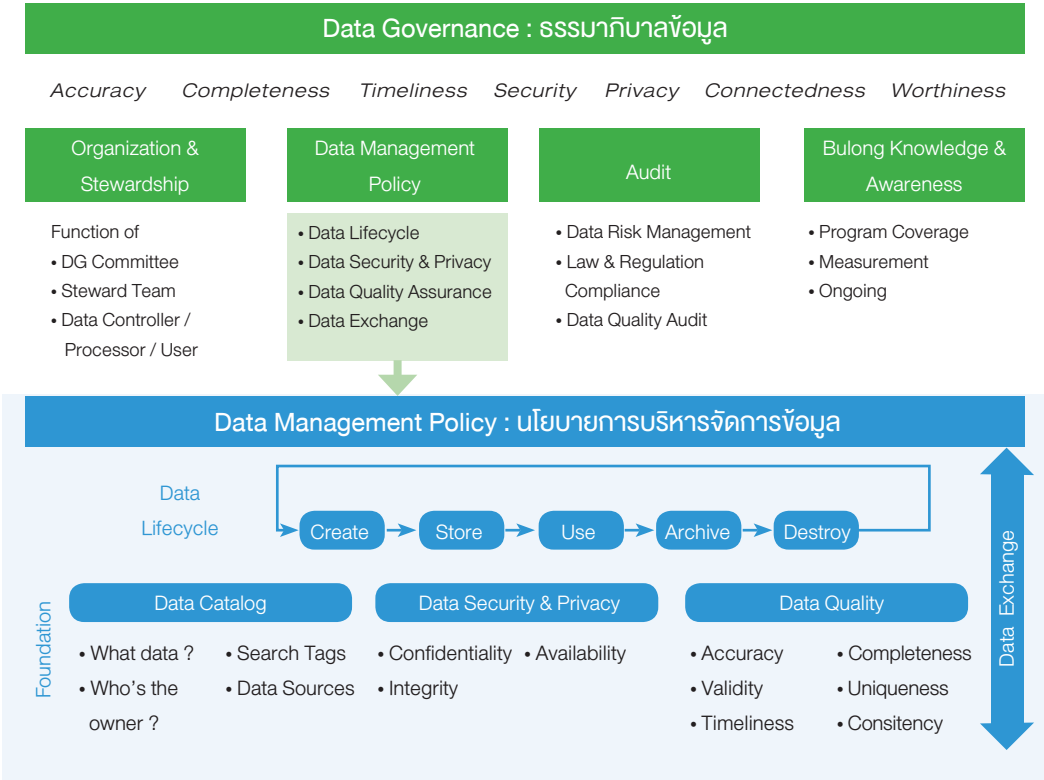
 ธรรมาภิบาลข้อมูลภาครัฐ
(Data Governance for Government)



3.1 กรอบธรรมาภิบาลข้อมูล

กรอบธรรมาภิบาลข้อมูลมีจุดประสงค์หลักเพื่อกำกับให้ข้อมูลมีความน่าเชื่อถือ ได้รับการกำกับดูแลอย่างเป็นระบบ และได้รับการปรับปรุงสม่ำเสมอ โดยเน้นให้ข้อมูลมีคุณภาพ ถูกต้อง สมเหตุสมผล เป็นปัจจุบัน ครอบคลุมสมบูรณ์ และสอดคล้องกัน

แนวทางการกำกับดูแลข้อมูลมีทั้งสิ้น 4 ขั้นตอน ดังต่อไปนี้



3.1.1 การวางโครงสร้างบุคลากรและระบุมความรับผิดชอบ (Organization and Stewardship)

การบริหารจัดการเรื่องธรรมาภิบาลข้อมูลนั้นมีความซับซ้อนและมีความต่อเนื่องทั้งในมิติของการกำหนดกฎเกณฑ์ นโยบาย และการวางระบบสารสนเทศเพื่อการกำกับดูแลข้อมูลให้เป็นไปตามนโยบายองค์กร ดังนั้นการวางโครงสร้างบุคลากรเพื่อเป็นคณะกรรมการข้อมูล (Data Committee) ที่เหมาะสมจึงมีความจำเป็น โครงสร้างคณะกรรมการข้อมูลควรประกอบด้วยบุคลากร 3 กลุ่ม คือ

- ผู้บริหารระดับกลางและระดับสูงขององค์กร มีหน้าที่ในการประเมิน (Assess) และอนุมัติ (Approve) นโยบายด้านการกำกับดูแลข้อมูลขององค์กร ซึ่งควรมีการทบทวนเป็นระยะ ๆ เพื่อปรับให้ทันสมัย รวมทั้งมีหน้าที่ในการตัดสินใจเรื่องที่เกี่ยวข้องกับข้อมูลในมิติต่าง ๆ โดยเฉพาะการตัดสินใจเลือกหน่วยงานเจ้าของข้อมูล (Data Owner) สำหรับชุดข้อมูลที่มีการเก็บเข้าข้อมูลภายในองค์กร ทั้งนี้เพื่อให้เกิดความสอดคล้องของข้อมูลสำหรับการวิเคราะห์
- ตัวแทนกลุ่มผู้ใช้ข้อมูลหรือผู้มีส่วนเกี่ยวข้องหลัก (User/Stakeholder) จากส่วนงานต่าง ๆ มีหน้าที่ในการกำหนดสิทธิการเข้าถึงชุดข้อมูลของบุคลากรภายในและภายนอก รวมทั้งกำหนดกฎเกณฑ์ในการกำกับดูแลชุดข้อมูลที่ส่วนงานตนเป็นเจ้าของ และร่างนโยบาย เงื่อนไข การอนุมัติใช้ข้อมูลของบุคลากรภายนอกองค์กร
- เจ้าหน้าที่สารสนเทศข้อมูล (Data Custodian) มีหน้าที่ในการพัฒนา/ จัดหาระบบสารสนเทศเพื่อกำกับดูแลข้อมูลและจำกัดสิทธิในการเข้าถึงข้อมูลอย่างเป็นรูปธรรมตามนโยบายที่กำหนด

นอกจากนี้ คณะกรรมการข้อมูลยังมีส่วนในการกำหนดโทษจากปัญหาหรือความเสียหายของแต่ละฝ่าย/แผนก/ส่วนงาน เพื่อใช้เป็นจุดตั้งต้นในการระบุชุดข้อมูลสำคัญขององค์กร และกำหนดผู้ควบคุมข้อมูล (Data Controller) ผู้ใช้ข้อมูล (Data User) และผู้ประมวลผลข้อมูล (Data Processor) สำหรับแต่ละชุดข้อมูล รวมทั้งมีหน้าที่ในการทำงานร่วมกับบริการข้อมูล (Data Steward) เพื่อตรวจสอบและแก้ไขปัญหาด้านคุณภาพข้อมูลและความมั่นคงปลอดภัยอีกด้วย

3.1.2 การจัดทำนโยบายการบริหารจัดการข้อมูล (Data Management Policy)

สิ่งที่คณะกรรมการข้อมูลต้องดำเนินการในลำดับแรกคือ การกำหนดนโยบายเพื่อการบริหารจัดการข้อมูลอย่างเป็นระบบแบ่งได้เป็น 4 มิติหลัก ดังนี้

1) การจัดการวงจรชีวิตข้อมูล (Data Lifecycle) : ดำเนินการวางกลไกเพื่อจัดการข้อมูลตามวงจรชีวิตที่เริ่มจากการสร้าง (Create) จัดเก็บ (Store) ใช้ประโยชน์ (Use) สำรองข้อมูล (Archive) และทำลายข้อมูล (Destroy) โดยระบุกระบวนการและกฎเกณฑ์ที่จำเป็นในแต่ละช่วงชีวิตของข้อมูลสำหรับชุดข้อมูลดิจิทัลขององค์กร ทั้งนี้เพื่อประสิทธิภาพในการบริหารจัดการ คณะกรรมการข้อมูลสามารถเริ่มจากการนำร่องชุดข้อมูลสำคัญบางส่วนก่อนเพื่อลดความยุ่งยากและประหยัดงบประมาณ ในช่วงแรกควรเลือกส่วนงานนำร่องและรายชุดข้อมูลเพื่อจัดทำ “รายการชุดข้อมูล” หรือ Data Catalog โดยมีการใส่ “รายละเอียดชุดข้อมูล” (Metadata) ตามที่ สพร. กำหนด [รายละเอียดเรื่องการจัดทำ รายการชุดข้อมูล นำเสนอในบทที่ 4]

2) การวางนโยบายด้านความมั่นคงปลอดภัยและความเป็นส่วนตัวของข้อมูล (Data Security and Privacy) : ร่างนโยบายด้านการป้องกันข้อมูลในบริบทของการรักษาความลับ ความพร้อมใช้งาน ความถูกต้องของข้อมูลตามมาตรฐาน ISO/IEC27001 มีรายละเอียดดังนี้

- การรักษาความลับ (Confidentiality) หมายถึง การรักษาข้อมูลตามสภาพของการจัดชั้นความลับ และมีการกำหนดสิทธิการเข้าถึงชุดข้อมูลอย่างชัดเจนตามระดับชั้นความลับโดยส่วนงานที่เป็นเจ้าของข้อมูล ทั้งนี้กฎการใช้ข้อมูลรวมถึงการเพิ่ม ลบ แก้ไข ประมวลผล นำไปใช้ แลกเปลี่ยน และเปิดเผยข้อมูล ซึ่งต้องสอดคล้องกับกฎหมาย เงื่อนไข และข้อกำหนดต่าง ๆ รวมถึงการคำนึงถึงมูลค่า ความสำคัญ และความอ่อนไหวของข้อมูล การกำหนดกฎเกณฑ์ที่เป็นรูปธรรมที่ชัดเจนจะทำให้เจ้าหน้าที่สารสนเทศข้อมูลสามารถพัฒนาเงื่อนไขหรือข้อกำหนดในโปรแกรมทำให้ควบคุมการเข้าถึงเพื่อรักษาความลับ และลดความเสี่ยงของการถูกคุกคาม เพื่อเป็นการป้องกันการดัดแปลง แก้ไข แต่งเติมโดยไม่ได้รับอนุญาต และการรั่วไหลของข้อมูลอีกด้วย

- ความพร้อมใช้งาน (Availability) หมายถึง การพร้อมในการใช้งาน อยู่เสมอ ผู้ใช้งานสามารถใช้ข้อมูลตามสิทธิได้ทันทีและได้อย่างต่อเนื่อง รวมถึงมีการสำรองข้อมูลไว้เมื่อเกิดภัยพิบัติหรือเหตุการณ์ที่ไม่คาดคิด
- ความถูกต้องของข้อมูล (Integrity) หมายถึง การคงสภาพของข้อมูล หรือการรักษาความถูกต้องสมบูรณ์ของข้อมูลให้มีความถูกต้องและ น่าเชื่อถือ รวมถึงมีการปกป้องข้อมูลให้ปราศจากการถูกเปลี่ยนแปลง โดยผู้ไม่มีสิทธิ เช่น ข้อมูลที่ใช้จะต้องเป็นข้อมูลที่ถูกต้องอย่างแท้จริง ไม่มีการดัดแปลงหรือแก้ไข

3) การประกันคุณภาพของข้อมูล (Data Quality Assurance): เพื่อประกัน ความน่าเชื่อถือของข้อมูลและประสิทธิภาพของการนำข้อมูลไปใช้ คณะกรรมการข้อมูลควรมีการวางแผนการด้านการประกันคุณภาพข้อมูลโดยมีการระบุเกณฑ์การวัดคุณภาพสำหรับ ข้อมูลแต่ละชุด และมีการกำหนด KPIs ที่ใช้วัดคุณภาพ รวมถึงกลไกการตรวจสอบคุณภาพที่ ชัดเจนเป็นระบบ เพื่อเป็นการตรวจสอบผลลัพธ์หรือความสำเร็จจากการกำกับดูแลข้อมูล องค์ประกอบในการประเมินคุณภาพ ตามกรอบการกำกับดูแลข้อมูลของ สพร. ประกอบด้วย

- ข้อมูลมีความถูกต้อง (Accuracy) ขึ้นอยู่กับวิธีการที่ใช้ในการควบคุม ข้อมูลนำเข้า และการควบคุมการประมวลผล ซึ่งวิธีการควบคุม การประมวลผล ได้แก่ การตรวจสอบยอดรวมที่ได้จากการประมวลผล แต่ละครั้ง หรือการตรวจสอบผลลัพธ์ที่ได้จากการประมวลผลด้วย เครื่องคอมพิวเตอร์ ส่วนการควบคุมการนำเข้าสามารถใช้วิธีสุ่มตรวจ ข้อมูลได้
- ข้อมูลมีความครบถ้วน (Completeness) ข้อมูลบางประเภทหากไม่ ครบถ้วนก็จัดเป็นข้อมูลด้อยคุณภาพได้เช่นกัน เช่น ข้อมูลประวัติผู้รับ การรักษาพยาบาล หากไม่มีหมู่เลือดของผู้รับการรักษาพยาบาล จะไม่ สามารถใช้ได้ในกรณีที่ต้องการข้อมูลหมู่เลือดของผู้รับการรักษาพยาบาล
- ข้อมูลมีความต้องกัน (Consistency) หมายถึง ค่าของข้อมูลในชุดเดียวกัน ควรสอดคล้องกับค่าในชุดข้อมูลอื่น หากมีการดึงข้อมูลสองค่าจากชุด ข้อมูลต่างแหล่งกันข้อมูลต้องไม่ขัดแย้งกัน

- ข้อมูลมีความเป็นปัจจุบัน (Timeliness) หมายถึง ผู้ใช้สามารถนำเอาผลลัพธ์ที่ได้ไปใช้ได้ทันเวลา นั่นคือ จะต้องมีการจัดเก็บข้อมูลที่รวดเร็ว เพื่อให้ทันความต้องการของผู้ใช้
- ข้อมูลตรงตามความต้องการของผู้ใช้ (Relevancy) หมายถึง ระดับของข้อมูล que บริหารจัดการต้องนำเสนอได้ตรงและมีประสิทธิภาพสามารถใช้งานได้ตามวัตถุประสงค์
- ข้อมูลมีความพร้อมใช้ (Availability) หมายถึง ข้อมูลมีความพร้อมในการใช้งานอยู่เสมอ เข้าถึงได้ง่าย ใช้งานได้จริง และใช้งานได้ตลอดเวลา

4) การแลกเปลี่ยนข้อมูล (Data Exchange): การแลกเปลี่ยนข้อมูลกับหน่วยงานภายนอกองค์กรจะเป็นรูปธรรมได้นั้น คณะกรรมการข้อมูลต้องมีการวางนโยบายที่ชัดเจน โดยมีการจัดทำข้อกำหนดและเงื่อนไข (Terms and Conditions) ที่จะใช้สำหรับการแลกเปลี่ยนข้อมูลทั้งการส่งออกและนำเข้าข้อมูล มีมาตรการและกลไกการแลกเปลี่ยนข้อมูล รวมถึงประเด็นเรื่องความจำเป็นในการเข้ารหัส (Encrypt) และผู้ดูแลจัดเก็บกุญแจถอดรหัส (Key) กรณีข้อมูลมีชั้นความลับ นโยบายและแนวทางที่ชัดเจนจะลดความเสี่ยงให้ส่วนงานสารสนเทศข้อมูล และสนับสนุนให้การบริหารจัดการแลกเปลี่ยนข้อมูลมีประสิทธิภาพ ประเด็นเรื่องการเชื่อมโยงแลกเปลี่ยนข้อมูลตามกรอบการกำกับดูแลข้อมูลของ สพร. กำหนดหัวข้อไว้ดังนี้

- กำหนดแนวปฏิบัติในการจัดการเรื่องความมั่นคงปลอดภัย คุณภาพข้อมูล และผู้ประสานงานหรือศูนย์ติดต่อ
- กำหนดกระบวนการในการเชื่อมโยงและแลกเปลี่ยนข้อมูลให้ชัดเจน เริ่มต้นตั้งแต่ขั้นตอนการเตรียมการ ขั้นตอนเริ่มดำเนินการ ขั้นตอนระหว่างดำเนินการ และขั้นตอนสิ้นสุดการดำเนินการ
- กำหนดเมทาเดตาของชุดข้อมูลที่ต้องการเชื่อมโยงและแลกเปลี่ยน ที่จำเป็นให้ครบถ้วน
- ทำสัญญาอนุญาตหรือข้อตกลงในการเชื่อมโยงและแลกเปลี่ยนที่จำเป็นให้ครบถ้วน
- กำหนดเทคโนโลยีและมาตรฐานทางเทคนิคที่ใช้ในการเชื่อมโยงและแลกเปลี่ยนข้อมูล

- บันทึกรายละเอียดและจัดเก็บข้อมูลการดำเนินงานที่เกิดขึ้นในแต่ละครั้งที่มีการเชื่อมโยงและแลกเปลี่ยนข้อมูล (Log File) ระหว่างหน่วยงาน เพื่อให้สามารถตรวจสอบย้อนกลับได้
- สามารถตรวจสอบได้ว่าการเชื่อมโยงและแลกเปลี่ยนได้ดำเนินการอย่างเหมาะสมหรือเป็นไปตามแนวทางปฏิบัติ กระบวนการเชื่อมโยงและแลกเปลี่ยนและมาตรฐานตามที่กำหนด

[รายละเอียดเพิ่มเติมเรื่องการให้บริการการแลกเปลี่ยนข้อมูลและกลไกนำเสนอ

ในบทที่ 5]

3.1.3 การตรวจประเมินธรรมาภิบาลข้อมูล (Audit)

องค์กรควรมีหน่วยงานภายในที่ได้รับมอบหมายให้ตรวจสอบธรรมาภิบาลข้อมูล (Internal Auditor) เพื่อเป็นการประกันคุณภาพโดยการตรวจสอบผลลัพธ์และความสำเร็จจากการกำกับดูแลข้อมูล นอกจากนี้หน่วยงานกลางภาครัฐโดยสำนักงานสถิติแห่งชาติ (สสช.) และ สพร. (External Auditor) จะมีกระบวนการตรวจสอบด้านธรรมาภิบาลข้อมูลหน่วยงานภาครัฐเพื่อเป็นการประกันคุณภาพอีกต่อหนึ่ง โดยองค์ประกอบในการตรวจประเมินคุณภาพ ประกอบด้วย

- การปฏิบัติตามการจัดการความเสี่ยงด้านข้อมูล (Data Risk Management) ตามนโยบายบริหารความเสี่ยงที่กำหนดโดยคณะกรรมการข้อมูล
- การปฏิบัติตามระเบียบและกฎหมายที่เกี่ยวข้องกับข้อมูล (Law and Regulation Compliance)
- การตรวจสอบคุณภาพข้อมูล (Data Quality Audit) เป็นการตรวจสอบกระบวนการที่เกี่ยวข้องกับข้อมูล

3.1.4 การสร้างความตระหนักและองค์ความรู้ในองค์กร (Building Knowledge and Awareness)

การกำกับดูแลข้อมูลเป็นสิ่งที่ทุกคนในองค์กรต้องให้ความร่วมมือ หากบุคลากรในองค์กรไม่มีความเข้าใจในวัตถุประสงค์ของการกำกับดูแลข้อมูล หรือไม่ตระหนักถึงความสำคัญของการกำกับดูแลข้อมูล ก็จะส่งผลให้การดำเนินการด้านนี้ไม่มีประสิทธิภาพและไม่เกิดประโยชน์อย่างแท้จริง ดังนั้นองค์กรควรมีแผนและนโยบายในการสร้างความตระหนัก รวมทั้งให้ความรู้เกี่ยวกับธรรมาภิบาลข้อมูล และมีการกระตุ้นให้บุคลากรทุกส่วนให้ความสำคัญและมีโอกาสได้นำความรู้ที่ได้ไปปรับใช้ในการทำงาน โดยมีการจัดอบรมหลักสูตรที่รวมถึงการสร้างองค์ความรู้ (Program Coverage) การวัดผล (Measurement) และความต่อเนื่อง (Ongoing) ของการดำเนินงานด้านธรรมาภิบาลข้อมูล

3.2 ขั้นตอนการจัดทำกรอบเพื่อกำกับดูแลข้อมูลดิจิทัลในเบื้องต้น

ในหัวข้อ 3.1 ได้อธิบายถึงกรอบธรรมาภิบาลในภาพกว้าง หน่วยงานภาครัฐที่อยู่ในช่วงเริ่มต้นของการวางนโยบายด้านนี้ สามารถนำร่องโดยกำหนดกระบวนการเชิงปฏิบัติสำหรับข้อมูลที่ขึ้นความลับ 4 ระดับ คือ กลุ่มข้อมูลเปิด กลุ่มข้อมูลส่วนบุคคล กลุ่มข้อมูลความลับราชการและกลุ่มข้อมูลความมั่นคงของรัฐ โดยเริ่มจากการกำหนดกรอบ ดังนี้

- 1) กำหนดกระบวนการจัดทำและคุ้มครองข้อมูล (Procedure) ที่ทำให้มั่นใจว่าปริมาณและคุณภาพของข้อมูลตรงตามมาตรฐาน
- 2) กำหนดกระบวนการตรวจสอบประเมินที่ ทำให้มั่นใจว่าข้อมูลมีความทันสมัยตลอดเวลา
- 3) กำหนดกระบวนการในการใช้ข้อมูลและการใช้ข้อมูลร่วมกัน⁶ (Sharing of Data)
- 4) กำหนดข้อตกลงและเงื่อนไข⁷ (Terms and Conditions) ในการเผยแพร่ข้อมูล (Data Release)
- 5) กำหนดเกณฑ์วิธี (Protocol) สำหรับการถ่ายโอนข้อมูล และวิธีการจัดการกับข้อมูล (Data Handling Method) เช่น แอปพลิเคชัน (Application Programming Interface : API) โพรโทคอล (File Transfer Protocol : FTP) หน่วยงานจำแนกประเภท อีเมล ฯลฯ

⁶ <http://www.stou.ac.th/Schools/Sst/main/eLearning/Oa/html/charpter3.html>

⁷ https://www.etda.or.th/file_storage/uploaded/Etda_Website/file/article_law_4.pdf

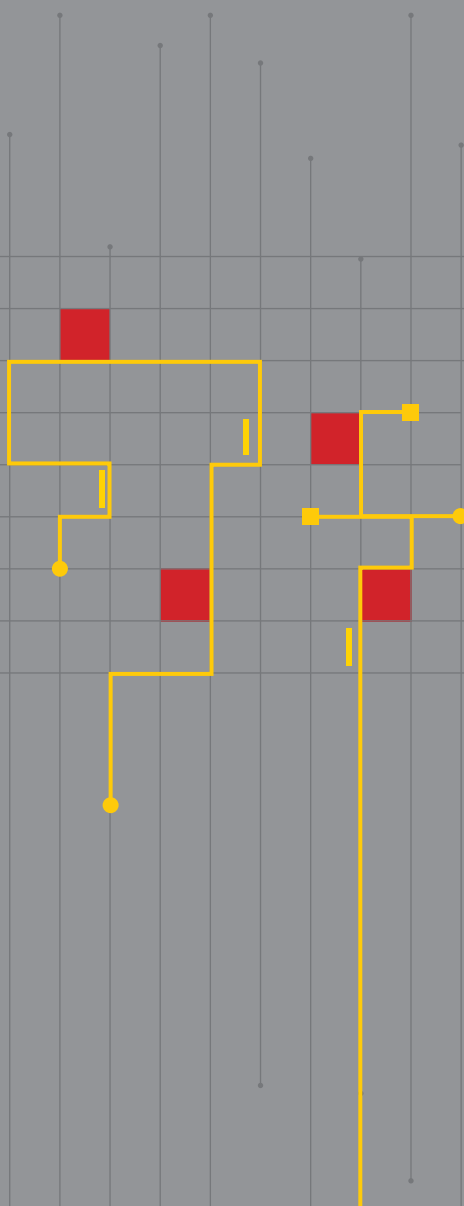
6) กำหนดนโยบายควบคุมการเข้าถึง (Access Control Policy) ว่าบุคคลใดสามารถเข้าถึงข้อมูลได้บ้าง โดยอาจกำหนดตามหน้าที่ความรับผิดชอบของบุคคล

7) กำหนดกลุ่มของฟิลเตอร์ (Filter) ที่ทำให้ข้อมูลไม่สามารถใช้ระบุตัวตนได้ (Anonymize) หรือ ลดระดับ (De-Classify) ขึ้นความลับของชุดข้อมูล (Dataset) โดยตัวอย่างการลดระดับขึ้นความลับของข้อมูล ได้แก่ กรณีที่ชุดข้อมูลมีข้อมูลที่สามารถระบุถึงตัวบุคคลได้ (Personally Identifiable Information : PII) การดำเนินการนำชื่อของบุคคล และเลขประจำตัวประชาชนออกจากชุดข้อมูลนั้น จะทำให้ชุดข้อมูลดังกล่าวถูกลดระดับขึ้นความลับลงได้

8) ออกแบบแบบฟอร์ม (Template) นโยบายการป้องกันข้อมูล (Data Protection Policy)

ตัวอย่างขั้นตอนการจัดทำธรรมาภิบาลข้อมูลและตัวอย่างเอกสารที่เกี่ยวข้องสามารถดูได้จาก





รายการข้อมูลภาครัฐ Government Data Catalog



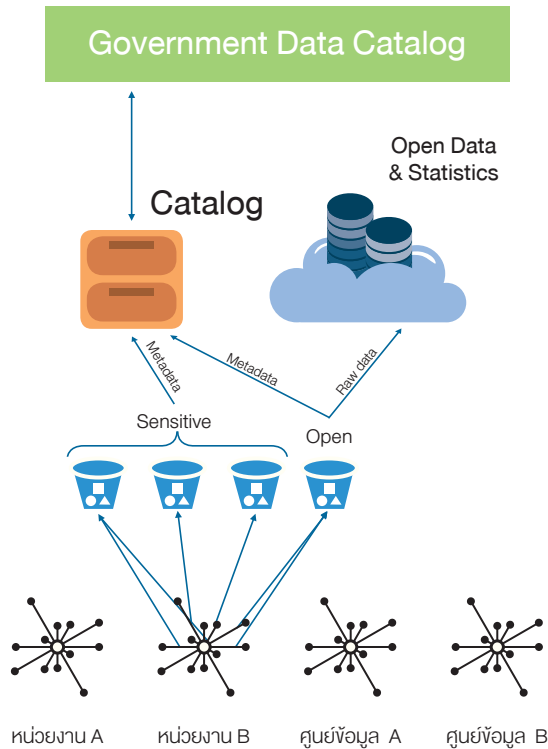
รายการข้อมูลภาครัฐ Government Data Catalog

จากความต้องการใช้ประโยชน์ข้อมูลภาครัฐเพิ่มมากขึ้นในปัจจุบัน และผู้ใช้ข้อมูลที่มีอยู่หลากหลาย ทั้งกลุ่มเจ้าหน้าที่ภาครัฐ นักวิจัยจากมหาวิทยาลัย นักวิชาการ/นักวิเคราะห์ ข้อมูลทั้งภาคเอกชนและภาคประชาสังคม รวมถึงประชาชนทั่วไป อย่างไรก็ตามปริมาณข้อมูลที่ถูกผลิตขึ้นอย่างต่อเนื่องทำให้การเลือกใช้ข้อมูลและการสืบค้นหาแหล่งข้อมูลที่เหมาะสมสำหรับการวิเคราะห์หรือให้บริการมีความซับซ้อนและยุ่งยากสูง ดังนั้นเพื่อให้การใช้ประโยชน์ข้อมูลภาครัฐเกิดขึ้นได้อย่างเป็นรูปธรรม สำนักงานสถิติแห่งชาติและสำนักงานพัฒนารัฐบาลดิจิทัลจึงวางแผนที่จะพัฒนารายการข้อมูลภาครัฐ (Government Data Catalog) และระบบสารสนเทศเพื่อการสืบค้น ดังจะนำเสนอต่อไป

4.1 บริการรายการข้อมูลภาครัฐ (Government Data Catalog Service)

ในกระบวนการการวิเคราะห์ข้อมูลขนาดใหญ่ นักวิทยาศาสตร์ข้อมูลจำเป็นต้องสำรวจหาแหล่งข้อมูลสำคัญตามหัวข้อการวิเคราะห์และทำความเข้าใจการใช้ข้อมูลในมิติต่าง ๆ รายการข้อมูลภาครัฐและเครื่องมือสารสนเทศข้อมูลจึงเป็นสิ่งจำเป็นเพื่อช่วยให้ผู้ใช้ประโยชน์ข้อมูลสามารถสืบค้น ร้องขอ เข้าถึง ทราบแหล่งที่มา ทราบถึงชั้นความลับ ประเภท รูปแบบ และสามารถใช้อนุพันธ์ของข้อมูลภาครัฐทั้งหมดได้ รายการข้อมูลดังกล่าวจะเป็นเสมือนสมุดหน้าเหลือง (Yellow Pages) ของข้อมูลภาครัฐที่สำคัญทั้งหมดของประเทศไทย ซึ่งเป็นจุดเริ่มต้นที่จำเป็นอย่างยิ่งในการพัฒนาการใช้ประโยชน์ข้อมูลภาครัฐให้มีประสิทธิภาพและเป็นไปในทิศทางเดียวกัน สามารถบูรณาการ ให้บริการ และใช้ประโยชน์ข้อมูลข้ามหน่วยงานได้อย่างเป็นระบบ

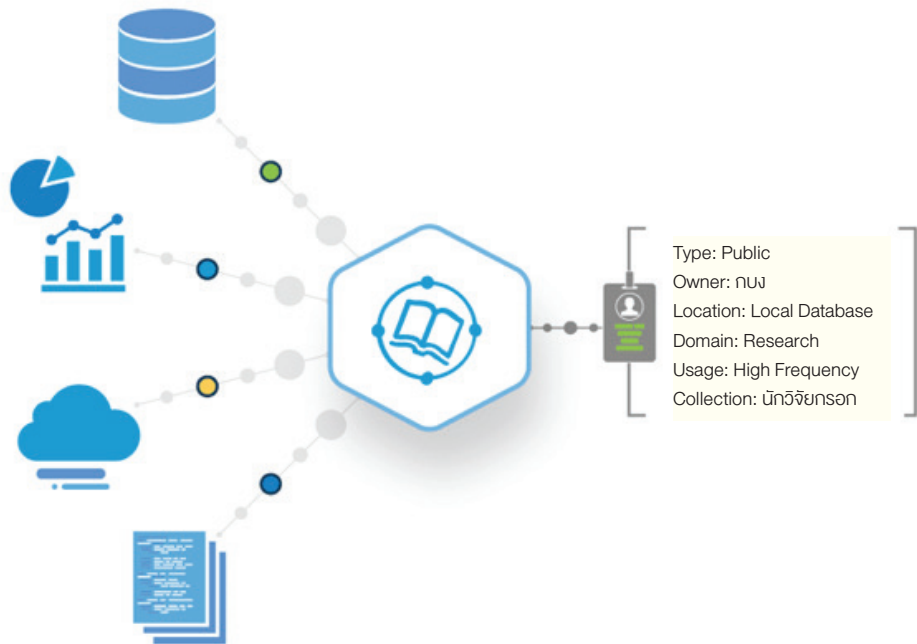
ระบบบริการรายการข้อมูลภาครัฐ (Government Data Catalog) แสดงในภาพที่ 4.1



ภาพที่ 4.1 การให้บริการรายการข้อมูลภาครัฐ

การจัดทำระบบรายการข้อมูลภาครัฐ เริ่มต้นจากการส่งเสริมและสนับสนุนให้หน่วยงานภาครัฐและศูนย์ข้อมูลทั้งหมดจัดทำรายการข้อมูลสำคัญภายในหน่วยงานเองในรูปแบบที่เป็นมาตรฐานเดียวกัน ชุดข้อมูลจะถูกจัดกลุ่มตามชั้นความลับและมีการบันทึกคำอธิบายรายการข้อมูลหรือที่เรียกว่า “เมทาดาท้า” (Metadata) สำหรับทุกชุดข้อมูล จากนั้นนำมาสร้างเป็นระบบ “แค็ตตาล็อกข้อมูล” (Data Catalog) หรือรายงานชุดข้อมูลของหน่วยงาน

หน่วยงานกลาง (สำนักงานพัฒนารัฐบาลดิจิทัล สพร. และสำนักงานสถิติแห่งชาติ) รวบรวมชุดข้อมูลต่าง ๆ จากระบบแค็ตตาล็อกระดับหน่วยงานมาสร้างเป็นรายการข้อมูลกลางภาครัฐเพื่อการใช้ประโยชน์ร่วมกันได้ ทั้งนี้การจัดทำระบบรายการข้อมูล ควรจัดทำผ่านระบบเทคโนโลยีสารสนเทศเพื่อความสะดวกในการสืบค้นและการใช้ประโยชน์อย่างเป็นรูปธรรม โครงสร้างระบบสารสนเทศอธิบายได้ดังแสดงในภาพที่ 4.2



ภาพที่ 4.2 ระบบสารสนเทศรายการข้อมูลภาครัฐ

โครงสร้างในภาพที่ 4.2 แบ่งออกได้เป็น 3 ส่วนคือ ระบบพอร์ทัล (Portal) บริการนามานุกรมข้อมูล (Directory Services) หรือสืบค้นโปรแกรมเพื่อเชื่อมโยงพอร์ทัลกับฐานข้อมูล และเมทาดาต้าของชุดข้อมูล [รายละเอียดของโครงสร้างสำคัญของระบบสารสนเทศรายการข้อมูลแสดงใน ภาคผนวก ข.1]

เมทาดาต้าเป็นส่วนสำคัญที่ช่วยให้การสืบค้นชุดข้อมูลที่ต้องการเกิดขึ้นได้อย่างรวดเร็ว ทั้งนี้เพื่อความสะดวกในการสืบค้นและแลกเปลี่ยนข้อมูลข้ามหน่วยงาน การจัดทำเมทาดาต้าควรอ้างอิงมาตรฐานเดียวกัน ซึ่ง สพร. ร่วมกับ GBDi ได้ออกแบบมาตรฐาน แบบฟอร์ม เอกสารกระบวนการ และเครื่องมือสำหรับให้ความสะดวกในการจัดทำรายการชุดข้อมูลพร้อมเมทาดาต้า หน่วยงานภาครัฐต่าง ๆ จะสามารถนำมาตรฐานดังกล่าวไปใช้งานและจัดทำชุดข้อมูลของหน่วยงานได้อย่างสอดคล้องกัน นอกจากนี้จะมีการออกแบบกลไกด้านข้อมูลที่หน่วยงานภาครัฐสามารถนำไปประยุกต์ใช้ในเรื่องการกำกับดูแลข้อมูล การให้บริการข้อมูล และการใช้ประโยชน์ข้อมูลอีกด้วย

ขั้นตอนการจัดทำรายการข้อมูล อธิบายได้ดังนี้

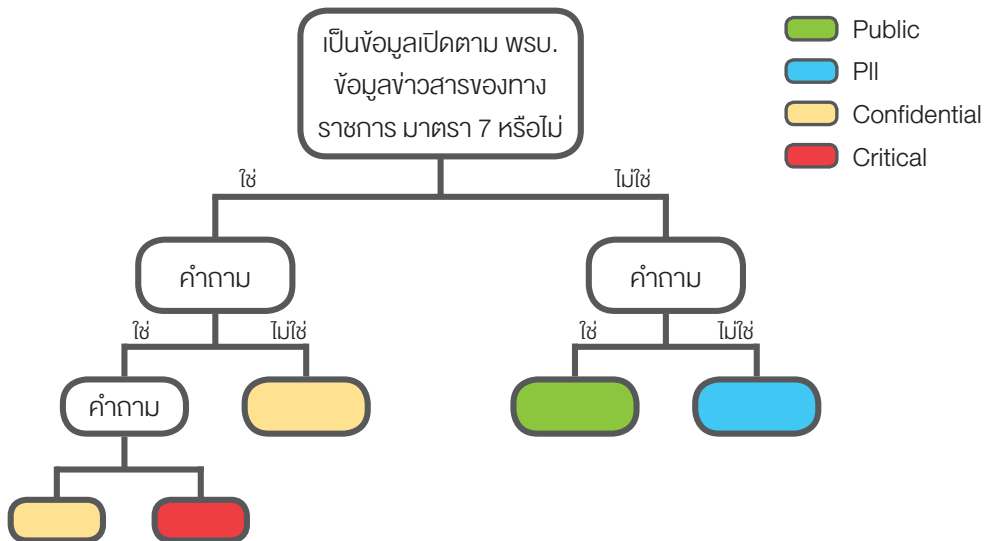
1) เลือกชุดข้อมูลสำคัญของหน่วยงาน สามารถทำได้โดยการระดมสมองจากผู้ใช้และผู้ดูแลข้อมูล เลือกรายการชุดข้อมูลเพื่อมาจัดทำเป็นแค็ตตาล็อกข้อมูล อาจพิจารณาจากการใช้ข้อมูลที่ใช้วางแผนและนโยบายที่ตอบ KPI ของหน่วยงาน หรือข้อมูลที่ใช้จัดทำรายงานประจำเดือน/ปี หรือ ข้อมูลที่ใช้วิเคราะห์หาคำตอบสำหรับคำถามที่ถูกลบปล่อยโดยผู้บริหารระดับสูง

2) ศึกษาและระบุเมทาดาต้าของชุดข้อมูลที่เลือกตามมาตรฐานกลาง โดยเมทาดาต้าที่จะจัดทำขึ้นควรมีการระบุข้อมูลต่าง ๆ อย่างน้อยดังนี้ ชื่อ-ประเภทชุดข้อมูล รูปแบบการจัดเก็บข้อมูล (Format) คำอธิบายรายการข้อมูล (Data Description) วัตถุประสงค์การจัดเก็บข้อมูล ชื่อหน่วยงานเจ้าของข้อมูลและผู้ติดต่อ คำสำคัญกำกับชุดข้อมูลเพื่อการสืบค้น (Keywords, Tags) กลุ่มผู้ใช้ข้อมูล (Business Domain) แหล่งที่มาของข้อมูล (Data Source) วิธีการ/กระบวนการเก็บรวบรวมข้อมูล (Collection Method) ความถี่ในการจัดเก็บ และขอบเขตเชิงภูมิศาสตร์หรือเชิงพื้นที่ของข้อมูล รวมทั้งสัญญาอนุญาตให้ใช้ข้อมูล รายการในเมทาดาต้าอาจจะมีการปรับเปลี่ยนได้ตามความเหมาะสม

หลังจากหน่วยงานมีการจัดทำเมทาดาต้าของชุดข้อมูลสำคัญแล้ว หน่วยงานกลางจะรวบรวมเมทาดาต้ามาจากหน่วยงานต่าง ๆ รวมถึงเมทาดาต้าของข้อมูลภาครัฐจากศูนย์ข้อมูล เช่น ศูนย์ข้อมูลน้ำ ศูนย์ข้อมูลประชาชน ฯลฯ มาเก็บไว้ เพื่อประโยชน์ในการสืบค้นข้อมูลภาครัฐ โดย

- ข้อมูลที่ไม่มีชั้นความลับ สามารถเผยแพร่ได้หรือข้อมูลเปิด (Open Data) การจัดเก็บจะรวบรวมทั้งเมทาดาต้าและข้อมูล ทั้งนี้เพื่อให้ความสะดวกแก่ผู้ให้บริการ กล่าวคือ หากข้อมูลที่ต้องการ ใช้ประโยชน์เป็นข้อมูลเปิด ผู้ร้องขอข้อมูลจะสามารถดาวน์โหลดข้อมูลไปใช้ได้โดยตรงผ่านทางบริการข้อมูล (Data Portal) เดียวกัน
- ข้อมูลที่มีชั้นความลับ การจัดเก็บจะรวบรวมเฉพาะเมทาดาต้าเท่านั้น ไม่จัดเก็บตัวข้อมูลจริง

3) จัดทำเกณฑ์สำหรับการระบุชั้นความลับของข้อมูล โดยแต่ละหน่วยงานควรออกแบบชุดคำถามเพื่อใช้เป็นเครื่องมือในการระบุชั้นความลับ โดยการสร้างชุดคำถามนั้นควรเริ่มจากการศึกษาพระราชบัญญัติและกฎกระทรวงต่าง ๆ ของหน่วยงานโดยละเอียด และสร้างชุดคำถามเรื่องความลับข้อมูลให้อยู่ในรูปแบบของต้นไม้ประกอบการตัดสินใจ (Decision Free) เพื่อให้ทุกส่วนงานในองค์กรสามารถใช้ต้นไม้ดังกล่าวเป็นเครื่องมือเพื่อแบ่งชั้นความลับได้อย่างสอดคล้องกันภายในองค์กร ตัวอย่างของต้นไม้ประกอบการตัดสินใจแสดงในภาพที่ 4.3 ชั้นความลับข้อมูลจะจำแนกได้เป็น 4 ระดับตามกรอบการกำกับดูแลข้อมูลของสพร. ได้แก่ ข้อมูลสาธารณะ ข้อมูลที่สามารถระบุถึงตัวบุคคลได้ ข้อมูลความลับราชการ และข้อมูลที่เกี่ยวข้องกับความมั่นคงของประเทศ



ภาพที่ 4.3 ต้นไม้ประกอบการตัดสินใจจำแนกชั้นความลับของข้อมูล

4) จัดทำดัชนีข้อมูล (Data Index) ซึ่งเป็นส่วนสำคัญที่จะเพิ่มความสะดวกรวดเร็วแก่บริการสืบค้นรายการข้อมูล โดยผู้ใช้บริการจะสามารถสืบค้นได้แบบหลายมิติ (Multi-Dimensional Search) ตัวอย่างเช่น

- การสืบค้นตามหัวข้อเรื่อง เช่น การวิเคราะห์ข้อมูลด้านสุขภาพและการสาธารณสุข การเงินและการธนาคาร ทรัพยากรธรรมชาติ การศึกษา การท่องเที่ยว กระบวนการยุติธรรม คมนาคม และบริการสังคม เป็นต้น
- การสืบค้นตามรายชื่อกระทรวงและหน่วยงานระดับกรมหรือเทียบเท่า เช่น การวิเคราะห์ข้อมูลจากกรมการขนส่งทางบก การบินไทย กระทรวงคมนาคม เป็นต้น
- การสืบค้นตามประเภทข้อมูล (Data Type) หรือ รูปแบบข้อมูล (Data Format)
- การสืบค้นตามชั้นความลับของข้อมูลและสิทธิการเข้าถึงข้อมูล เช่น
 - กลุ่มข้อมูลเปิด (Open Data) ที่ทุกคนมีสิทธิในการเข้าถึงและนำไปใช้ได้
อย่างเสรี กลุ่มข้อมูลนี้จะเป็นประโยชน์ต่อภาครัฐ ภาคเอกชน และประชาชน
 - กลุ่มข้อมูลส่วนบุคคล (Personally Identifiable Information) ที่อาจต้อง
มีการนำรายละเอียดส่วนที่ระบุตัวตนได้ออกก่อนที่จะนำไปวิเคราะห์ และ
หน่วยงานเจ้าของข้อมูลสามารถระบุข้อจำกัดการใช้ประโยชน์ข้อมูลได้
 - กลุ่มข้อมูลความลับราชการ (Government Data) ที่ให้สิทธิเฉพาะเจ้าหน้าที่
ที่มีหน้าที่รับผิดชอบเกี่ยวข้องเท่านั้น
 - กลุ่มข้อมูลความมั่นคงของรัฐ (Secured Government Data) ที่ให้สิทธิ
เฉพาะเจ้าหน้าที่ที่มีหน้าที่รับผิดชอบเกี่ยวข้องเท่านั้น

5) ออกแบบเกณฑ์กลางและกระบวนการตัวอย่างในการประเมินคุณภาพชุดข้อมูล ซึ่งตัวอย่างเกณฑ์ด้านคุณภาพข้อมูลตามมาตรฐานสากล ได้แก่ ความถูกต้องแม่นยำ (Accuracy) ความต้องกัน (Consistency) ความพร้อมใช้งาน (Availability) ความครบถ้วน (Completeness) ความสอดคล้องกับรูปแบบมาตรฐาน (Conformance) ความน่าเชื่อถือได้ (Credibility) ความสามารถนำไปใช้งาน (Processability) ความเข้าประเด็น (Relevance) ความเป็นปัจจุบัน (Timeliness) และบริบท (Context) ตารางที่ 4.1 อธิบายความหมายของเกณฑ์ตัวอย่าง เมื่อส่วนงานที่เป็นเจ้าของชุดข้อมูลเลือกเกณฑ์การประเมินผลข้อมูลที่เหมาะสมแล้วควรตั้ง KPI ด้านการวัดคุณภาพข้อมูล ตัวอย่างเช่น กรณีเจ้าหน้าที่บันทึกข้อมูลผ่านหน้าจอคอมพิวเตอร์ เกณฑ์ด้านความถูกต้องของชุดข้อมูลจะเป็นเกณฑ์สำคัญ ค่า KPI อาจกำหนดเป็นร้อยละ 75 ในปีแรก และปรับเป็นร้อยละ 90 ในปีต่อ ๆ ไป จากนั้นควรมีการระบุวิธีการในการสุ่มตรวจเพื่อให้ข้อมูลเป็นไปตามเกณฑ์ด้านคุณภาพอย่างชัดเจน

ตารางที่ 4.1 เกณฑ์การประเมินคุณภาพข้อมูล

Accuracy	ข้อมูลถูกต้อง เป็นไปตามความจริง
Consistency	ข้อมูลมีความสอดคล้อง ไม่ขัดแย้งกันเอง
Availability	สามารถเข้าถึง/ใช้งานข้อมูลได้ทั้งในปัจจุบันและในอนาคต
Completeness	ข้อมูลมีความสมบูรณ์ ครบถ้วน
Conformance	ข้อมูลเป็นไปตามมาตรฐานที่เป็นที่ยอมรับ
Credibility	ข้อมูลได้มาจากแหล่งข้อมูลที่เชื่อถือได้
Processability	ข้อมูลสามารถถูกใช้งานและประมวลผลได้โดยระบบสารสนเทศ
Relevance	ข้อมูลมีจำนวนมากพอที่จะใช้งานได้อย่างมีประสิทธิภาพ
Timeliness	ข้อมูลมีความทันสมัย ใหม่พอที่จะใช้ประโยชน์ได้อย่างมีประสิทธิภาพ
Context	ข้อมูลเหมาะสมที่จะใช้งานในบริบทที่ต้องการ

6) ระบุสิทธิการเข้าถึงชุดข้อมูล โดยส่วนงานเจ้าของข้อมูลพิจารณาชุดข้อมูลตามชั้นความลับ และออกแบบกลุ่มผู้ใช้ข้อมูล จากนั้นเลือกว่ากลุ่มผู้ใช้ข้อมูลกลุ่มใดควรมีสิทธิในการเข้าถึงชั้นความลับข้อมูลใดบ้างและใส่เกณฑ์การเข้าถึงข้อมูลในเมทาดาต้าขอชุดข้อมูล ซึ่งเจ้าหน้าที่สารสนเทศจะสามารถนำเกณฑ์ดังกล่าวไปพัฒนาเป็นโปรแกรมเพื่อการระบุสิทธิการเข้าถึงชุดข้อมูล ดังนั้นการเข้าถึงบริการข้อมูลจะเกิดขึ้นได้ตามสิทธิของผู้ใช้ผ่านระบบสารสนเทศเท่านั้น จัดเป็นการป้องกันการรั่วไหลของข้อมูลและจะส่งผลให้ผู้ใช้สามารถเข้าถึง/แลกเปลี่ยนข้อมูล ได้สะดวก รวดเร็ว และเป็นรูปธรรม

7) ออกแบบและจัดทำระบบบริการนามุกรมข้อมูล (Directory Services) และระบบบริหารจัดการเมทาดาต้า (Metadata Management System) [รายละเอียดเชิงเทคนิคและตัวอย่างหน้าจอแสดงในภาคผนวก ข.1] โดยระบบดังกล่าวจะมีกลไกในการค้นหาข้อมูลผ่านการสืบค้นในมิติต่าง ๆ กลไกการร้องขอข้อมูล (กรณีข้ามหน่วยงาน) และกลไกการอนุมัติใช้ข้อมูลสำหรับคำขอประเภทต่าง ๆ ทั้งนี้การขอข้อมูลที่ไม่มีชั้นความลับ (ข้อมูลเปิด) จะไม่จำเป็นต้องขออนุมัติ นอกจากนี้ระบบนามุกรมจะทำหน้าที่ยืนยันตัวตน (Authentication) และตรวจสอบสิทธิในการเข้าถึงข้อมูล (Authorization) ของผู้ใช้งานตามกฎเกณฑ์ที่หน่วยงานเจ้าของข้อมูลระบุ รวมทั้งบันทึกรายการการสืบค้นและคำขอข้อมูลทั้งหมด (Search and Request Log) เพื่อให้หน่วยงานเจ้าของข้อมูลสามารถตรวจสอบได้ ระบบนามุกรมนี้สามารถเชื่อมโยงไปยังศูนย์เชื่อมโยงและแลกเปลี่ยนข้อมูล (Data Linkage and Exchange Centers) เพื่ออำนวยความสะดวกให้การเข้าถึงข้อมูล (Data Access) และถ่ายโอนข้อมูล (Data Transfer) ข้ามหน่วยงานในอนาคต

4.2 การทำงานร่วมกับหน่วยงานภาครัฐเรื่องการกำกับดูแลข้อมูล

ความสำเร็จของการจัดทำรายการข้อมูลภาครัฐอย่างเป็นรูปธรรมนั้น จะต้องเกิดจากความเข้าใจและเห็นถึงประโยชน์ร่วมกัน การแลกเปลี่ยนข้อมูลข้ามหน่วยงานจะเกิดประโยชน์ได้ หน่วยงานภาครัฐต้องสามารถดูแลข้อมูลของตนได้อย่างเป็นระบบ กระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคมและสำนักงานพัฒนารัฐบาลดิจิทัลจึงสนับสนุนให้หน่วยงานภาครัฐจัดตั้งคณะทำงานด้านข้อมูล (Data Stewardship Team) เพื่อดำเนินงานด้านนโยบายข้อมูลของหน่วยงานเอง โดยสถาบันส่งเสริมการวิเคราะห์และบริหารข้อมูลขนาดใหญ่ภาครัฐ (GBDi) และสำนักงานพัฒนารัฐบาลดิจิทัล สามารถให้การสนับสนุนทางด้านวิชาการเพื่อให้ได้นโยบายข้อมูลที่เหมาะสมกับหน่วยงานและสอดคล้องกันในภาครัฐ ตัวอย่างเช่น การกำหนดนโยบายเพื่อกำกับดูแลให้ข้อมูลแต่ละประเภทมีวงจรชีวิต (Data Life Cycle) ที่เหมาะสม บริหารจัดการได้อย่างเป็นรูปธรรม และตอบโจทย์ด้านการรายงานและการวิเคราะห์เพื่อประกอบการตัดสินใจ

ในเบื้องต้น กระบวนการสร้างความเข้าใจและจัดเตรียมมาตรฐานด้านข้อมูลจำเป็นต้องได้รับความร่วมมือจากหน่วยงานภาครัฐอย่างเข้มแข็ง ดังต่อไปนี้

- 1.) จัดตั้งคณะทำงานด้านข้อมูล เพื่อประสานกับคณะทำงานจากสำนักงานสถิติแห่งชาติและสำนักงานพัฒนารัฐบาลดิจิทัล ซึ่งเป็นหน่วยงานที่รับผิดชอบเรื่องการจัดทำรายการข้อมูลภาครัฐ
- 2.) ระบุชุดข้อมูลสำคัญ (Critical Dataset) ของหน่วยงาน โดยมีเกณฑ์ที่ชัดเจนในการกำหนดระดับความสำคัญของชุดข้อมูล
- 3.) จำแนกชุดข้อมูลสำคัญตามชั้นความลับและจัดทำดัชนีข้อมูลในมิติต่าง ๆ เพื่อความสะดวกในการสืบค้น
- 4.) สร้างกฎเกณฑ์ว่าด้วยสิทธิการเข้าถึงข้อมูล (Access Insight) โดยระบุประเภทบุคคล ที่จะมีสิทธิเข้าดูข้อมูลในแต่ละชั้นความลับ และสิทธิในการขอรับข้อมูลไปใช้ประโยชน์ผ่านกลไกการถ่ายโอนข้อมูล

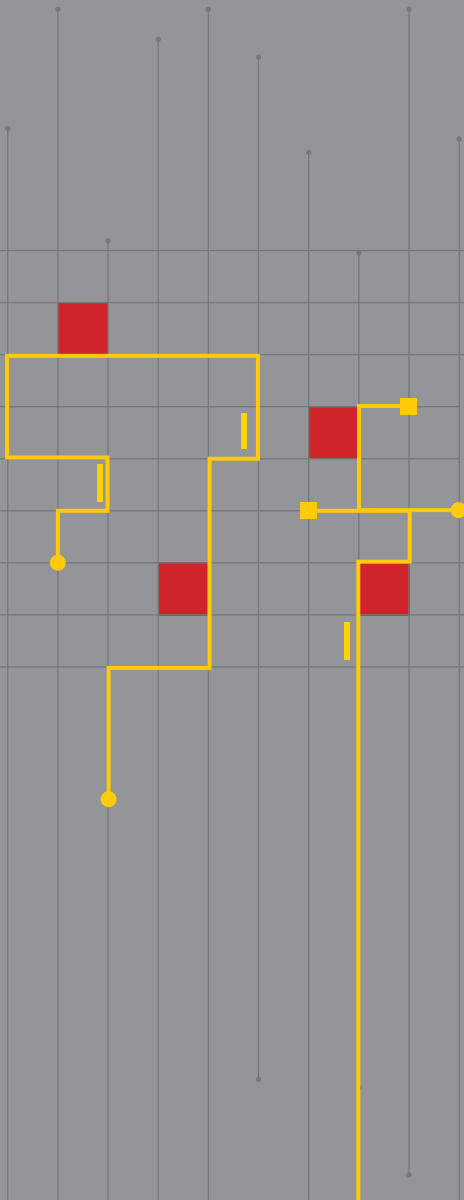
- 5.) จัดทำเมทาเดตาของชุดข้อมูลต่าง ๆ เพื่อให้หน่วยงานภาครัฐมีเมทาเดตาในรูปแบบเดียวกันสามารถรองรับการสืบค้น เชื่อมโยง และแลกเปลี่ยนข้อมูลภาครัฐได้ในอนาคต
 - 6.) พิจารณาสร้างพื้นที่สำหรับแลกเปลี่ยนข้อมูลแยกต่างหากจากฐานข้อมูลหลัก และแปลงรูปแบบการจัดเก็บข้อมูลสำคัญในพื้นที่ดังกล่าวให้อยู่ในสกุลไฟล์ที่ทั้งคนและเครื่องคอมพิวเตอร์สามารถอ่านได้อย่างสะดวกเพื่อให้เกิดการแลกเปลี่ยนข้อมูลเกิดขึ้นได้แบบอัตโนมัติ
 - 7.) ออกแบบกระบวนการประเมินคุณภาพและปริมาณของชุดข้อมูลสำคัญที่เหมาะสมกับหน่วยงาน โดยระบุเกณฑ์ที่ใช้ในการประเมินคุณภาพอย่างชัดเจนและเป็นไปตามมาตรฐานสากล จากนั้นทำการประเมินช่องว่างที่เกิดขึ้น ระหว่างสถานะข้อมูลปัจจุบันกับสถานะที่ควรจะเป็นตามเกณฑ์ที่กำหนด เพื่อวางแผนดำเนินการ (Roadmap) การพัฒนาชุดข้อมูลสำคัญให้มีคุณภาพ ภายในระยะเวลา 3-5 ปี
 - 8.) วางกรอบนโยบายด้านกำกับดูแลวงจรชีวิตและบริหารจัดการข้อมูลดิจิทัล และสามารถใช้ได้กับข้อมูลทุกประเภท (4 ประเภท) โดยเฉพาะประเด็นเรื่องกระบวนการจัดทำและคุ้มครองข้อมูล และกระบวนการใช้และแลกเปลี่ยนข้อมูล
- แผนการดำเนินงานเรื่องการให้บริการรายการข้อมูลภาครัฐจะแบ่งเป็น 3 ช่วง โดยเริ่มต้นช่วงที่ 1 ในปีงบประมาณ 2563 ช่วงที่ 2 ในปีงบประมาณ 2564 และช่วงที่ 3 ในปีงบประมาณ 2565 และ 2566

ช่วงที่ 1 กิจกรรมรวมถึงการศึกษาและจัดทำชุดเครื่องมือที่ประกอบด้วย กรอบมาตรฐาน กรอบนโยบาย กลไกการจัดการ และแบบฟอร์มด้านข้อมูลต่าง ๆ เพื่อให้หน่วยงานภาครัฐใช้เป็นจุดเริ่มต้นในการจัดทำมาตรฐานเมทาเดตาและกำกับดูแลข้อมูล เช่น มาตรฐานและตัวอย่างเมทาเดตา วิธีการประเมินคุณภาพและปริมาณข้อมูล เกณฑ์สำหรับการระบุชั้นความลับของข้อมูล กรอบนโยบายด้านการบริหารจัดการข้อมูลดิจิทัล ฯลฯ โดยศึกษาจากรูปแบบที่ใช้ในต่างประเทศ และรูปแบบของเมทาเดตาที่มีการดำเนินการแล้วของหน่วยงานภาครัฐ ในขั้นตอนนี้คณะทำงานอาจจะร้องขอเมทาเดตาจากหน่วยงานภาครัฐเท่าที่มีเพื่อมาถ่วงน้ำหนักและจัดทำรายการข้อมูลเบื้องต้นได้อีกด้วย หลังจากนั้น คณะทำงานจากสำนักงานสถิติแห่งชาติและสำนักงานพัฒนารัฐบาลดิจิทัลจะนำชุดเครื่องมือที่จัดทำไปใช้เพื่อจัดทำเมทาเดตาข้อมูลและวางแผนการจัดการนโยบายด้านข้อมูลกับหน่วยงานนำร่อง

รวมทั้งจัดฝึกอบรมบุคลากรเพื่อเตรียมความพร้อมในการให้คำแนะนำให้กับหน่วยงานภาครัฐอื่นในการดำเนินงานระยะถัดไป

ช่วงที่ 2 สำนักงานสถิติแห่งชาติและสำนักงานพัฒนารัฐบาลดิจิทัลจะนำชุดเครื่องมือที่ผ่านกระบวนการทดสอบกับหน่วยงานนำร่องแล้วไปใช้ร่วมกับหน่วยงานภาครัฐอื่น ๆ รวมทั้งพัฒนาระบบบริการนิตยสารข้อมูลภาครัฐ และระบบบริหารจัดการเมทาดาต้า

ช่วงที่ 3 ให้บริการระบบบริการนิตยสารข้อมูลภาครัฐอย่างเต็มรูปแบบ และทำการขยายผลให้ครอบคลุมหน่วยงานภาครัฐทั้งหมด



การให้บริการด้านการแลกเปลี่ยนข้อมูล
Data Exchange



การให้บริการด้านการแลกเปลี่ยนข้อมูล Data Exchange

การใช้ประโยชน์ข้อมูลข้ามหน่วยงานต้องอาศัยระบบบริการนามาตรฐานข้อมูลภาครัฐ เพื่อการค้นหาและระบุข้อมูลที่ต้องการสำหรับแต่ละโจทย์ปัญหา และต้องมีกลไกด้านระบบสารสนเทศที่เอื้อให้เกิดการถ่ายโอนข้อมูลข้ามหน่วยงานได้อย่างสะดวก ดังนั้นจึงต้องมีการพัฒนาระบบสารสนเทศเพื่อบริหารจัดการการเชื่อมโยงและแลกเปลี่ยนข้อมูลสำหรับการวิเคราะห์ข้อมูลขนาดใหญ่ของหน่วยงานภาครัฐ ตัวอย่างบริการพื้นฐานของระบบดังกล่าวรวมถึงการยืนยันตัวตน การจำกัดสิทธิในการเข้าถึงข้อมูลบริการส่งข้อมูลข้ามหน่วยงานผ่านระบบสารสนเทศ การบันทึกทรานแซกชัน (Transaction) การเข้าถึงข้อมูลเพื่อเป็นหลักฐาน และการบริการข้อมูลประเภทอื่น เนื้อหาในบทนี้จะนำเสนอแนวทางสำหรับการให้บริการด้านการแลกเปลี่ยนข้อมูล

5.1 มาตรฐานข้อมูล

เพื่อให้การแลกเปลี่ยนข้อมูลเกิดขึ้นได้อย่างมีประสิทธิภาพ หน่วยงานภาครัฐควรสร้างพื้นที่แลกเปลี่ยนข้อมูล (Data Exchange Zone) แยกต่างหากจากฐานข้อมูลหลัก และพิจารณาแปลงรูปแบบข้อมูลดิบที่จะนำมาใส่ในพื้นที่แลกเปลี่ยนนี้ให้เอื้อต่อการแลกเปลี่ยนข้อมูล มาตรฐานรูปแบบข้อมูลแบ่งออกได้เป็น 4 ระดับ ดังแสดงในตารางที่ 5.1

ตารางที่ 5.1 สกุลไฟล์สำหรับการแลกเปลี่ยนข้อมูล

สกุลไฟล์ที่เครื่องคอมพิวเตอร์สามารถอ่านได้สะดวก	สกุลไฟล์เชิงแผนที่ที่เครื่องคอมพิวเตอร์สามารถอ่านได้	สกุลไฟล์ที่สามารถถูกใช้งานได้อย่างมีข้อจำกัด	สกุลไฟล์ที่ไม่แนะนำ	สกุลไฟล์ที่ใช้ประโยชน์ได้ยาก ⊘
★★★★		★★★★	★★	★
JSON	Shapefile	CSV	HTML	Images (PNG, JPG)
XML	GeoJSON	Excel	Word	Charts
RDF	GML		TXT	PDF
ODF	KML			เอกสาร/กระดาษ
TSV	WKT			

ระดับ 4 ดาว : รูปแบบที่ดีที่สุดสำหรับการแลกเปลี่ยนข้อมูล ตัวอย่างเช่น เจซัน (JavaScript Object Notation : JSON) เอ็กซ์เอ็มแอล (Extensible Markup Language : XML) และ อาร์ดีเอฟ (Resource Description Framework : RDF) ฯลฯ ซึ่งเป็นไฟล์ข้อมูลที่ถูกเก็บในลักษณะข้อความ (Plain Text) ที่ทั้งคนและคอมพิวเตอร์สามารถอ่านเข้าใจได้ รวมทั้งระบบคอมพิวเตอร์สามารถประมวลผลข้อมูลได้อัตโนมัติ

ระดับ 3 ดาว : ไฟล์ข้อมูลถูกจัดเป็นแถวหรือคอลัมน์ (Column) ที่คนและคอมพิวเตอร์สามารถอ่านเข้าใจได้ถ้ามีคำอธิบายจากผู้ให้บริการข้อมูล ได้แก่ ซีเอสวี (Comma-Separated Value : CSV) เอ็กซ์แอลเอส (XLS) หรือ เอ็กซ์แอลเอสเอ็กซ์ (XLSX) ทั้งนี้สกุลไฟล์ CSV สามารถถูกแปลงเป็น JSON, XML และ RDF ได้อัตโนมัติโดยใช้โปรแกรมสำเร็จรูป [ตัวอย่างวิธีการแสดงในภาคผนวก ข.2] ขณะที่ไฟล์เอ็กซ์เซลจะมีข้อดีน้อยกว่าเมื่อเทียบกับ CSV เนื่องจากต้องใช้ซอฟต์แวร์ลิขสิทธิ์จึงจะสามารถเปิดและแปลงไฟล์ได้ อย่างไรก็ตามการแลกเปลี่ยนข้อมูลยังสามารถทำได้อย่างเป็นระบบแม้มีข้อจำกัดอยู่บ้าง

ระดับ 2 ดาว : การใช้ประโยชน์ข้อมูลรูปแบบด็อก (DOC) หรือ ด็อกเอ็กซ์ (DOCX) เฮชทีเอ็มแอล (HTML) และไฟล์เท็กซ์ (TXT) โดยคอมพิวเตอร์จะต้องแปลงไฟล์หรือกระบวนการประมวลผลภาษาธรรมชาติ (Natural Language Processing : NLP) เพื่อสกัดข้อมูลจากไฟล์เท็กซ์ก่อน จัดเป็นภาระเพิ่มในการประมวลผลและไม่สะดวกสำหรับการแลกเปลี่ยนข้อมูลมากนัก

ระดับ 1 ดาว : รูปแบบข้อมูลที่ไม่เหมาะสมกับการแลกเปลี่ยนข้อมูลเพื่อการนำไปใช้ประโยชน์คือ ไฟล์ PDF ภาพ กราฟ และ ข้อมูลที่ยังไม่อยู่ในรูปแบบดิจิทัล รูปแบบเหล่านี้ต้องอาศัยเทคโนโลยีขั้นสูง เช่น เทคโนโลยีปัญญาประดิษฐ์ ในการสกัดข้อมูลสำคัญสำหรับการวิเคราะห์และประมวลผล

นอกจากนี้ ข้อมูลดิบที่หน่วยงานภาครัฐจัดเก็บอาจจะไม่สอดคล้อง ตัวอย่างเช่น ข้อมูล ‘วันที่’ หน่วยงานที่ 1 เก็บในรูปแบบ dd/MM/YYYY หน่วยงานที่ 2 เก็บในรูปแบบ MM/dd/YYYY และปีที่เก็บจะมีโอกาสเป็นได้ทั้งปีคริสต์ศักราช (ค.ศ.) และปีพุทธศักราช (พ.ศ.) การใช้ประโยชน์ข้อมูลจึงต้องมีการแปลงรูปแบบเข้าหากัน ซึ่งจะเพิ่มค่าใช้จ่ายและไม่สะดวกต่อการทำงาน ดังนั้นกระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคมจะทำหน้าที่สร้างและรวบรวมมาตรฐานกลางสำหรับข้อมูลประเภทต่าง ๆ และจัดทำคลังมาตรฐานข้อมูล (Master Data) ที่สืบค้นอ้างอิงได้ โดยมีการตรวจสอบผู้เป็นเจ้าของข้อมูลหลักก่อน ตัวอย่างเช่น ชื่อคนและที่อยู่ จะมีกระทรวงมหาดไทยเป็นเจ้าของข้อมูลหลัก รูปแบบการจัดเก็บจึงควรยึดตามกระทรวงมหาดไทย การนำข้อมูลไปใช้ต่อจึงจะสามารถอ้างอิงได้อย่างถูกต้องโดยไม่ต้องแปลงรูปแบบ กรณีข้อมูลทั่วไป (ไม่มีเจ้าของตามพระราชบัญญัติ เช่น ‘วันที่’) กระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคมจะสร้างมาตรฐานรูปแบบข้อมูลเพื่อการจัดเก็บตามมาตรฐานสากล

5.2 การปรับปรุงข้อมูลเพื่อเตรียมความพร้อม

สำหรับหน่วยงานเจ้าของข้อมูลที่จะเตรียมจะเข้าร่วมในกลไกการแลกเปลี่ยนข้อมูลภาครัฐ หน่วยงานควรที่จะปรับปรุงข้อมูลสำคัญเพื่อเตรียมพร้อมสำหรับการให้บริการข้อมูล ก่อนที่นำข้อมูลมาวางในพื้นที่แลกเปลี่ยนข้อมูล (Data Exchange Zone) ของหน่วยงานเอง โดยเริ่มจากเกณฑ์ประเมินคุณภาพพื้นฐาน 2 ประการ ดังนี้

1.) ความสอดคล้องกับรูปแบบมาตรฐานของข้อมูล (Conformance)

- ตรวจสอบให้ชุดข้อมูลมีแถว “Header” สำหรับอธิบายรายละเอียดของคอลัมน์ (Column) เพื่อการนำไปใช้ประโยชน์ของหน่วยงานที่ขอข้อมูล
- ตรวจสอบให้ชุดข้อมูลมีรหัสที่แสดงเวอร์ชัน เพื่อให้ผู้ขอข้อมูลสามารถติดตามการเปลี่ยนแปลงของชุดข้อมูลได้ โดยเวอร์ชันสามารถใช้ตรวจสอบข้อมูลพื้นฐาน เช่น วันที่เก็บข้อมูล วันที่ปรับปรุงข้อมูล วิธีจัดเก็บข้อมูล และหน่วยงานเจ้าของข้อมูล ทั้งนี้เพื่อให้ผู้ใช้ประโยชน์ข้อมูลสามารถตรวจสอบความใหม่ของข้อมูลได้

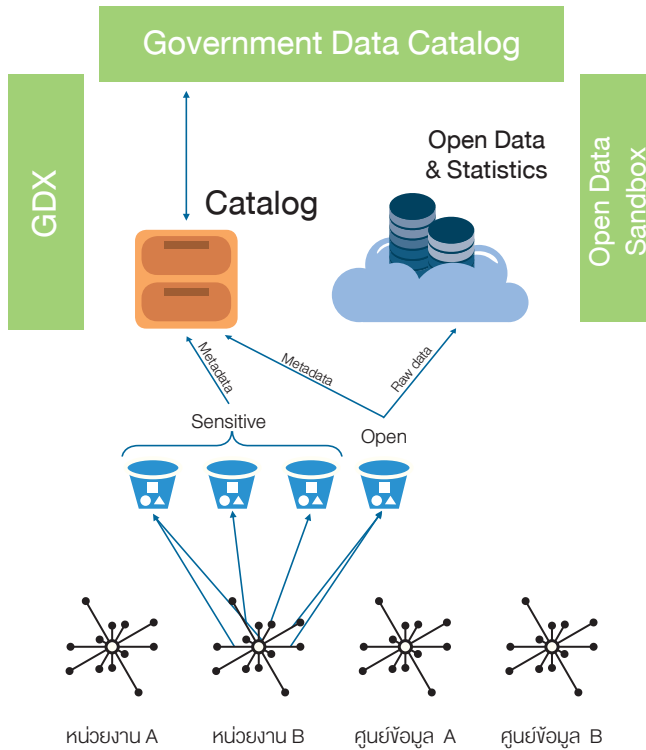
2.) ความถูกต้องของข้อมูล (Accuracy)

- ตรวจสอบหาค่าที่ผิดปกติจากกลุ่มทั่วไปในชุดข้อมูล หากพบข้อมูลที่มีช่วงของค่าที่เป็นไปไม่ได้ เช่น ข้อมูลอายุมีค่าหลักพัน ควรปรับแก้ เป็นต้น
- ตรวจสอบหาข้อมูลที่ซ้ำกัน ข้อมูลที่ขาดหายไปบางส่วน (Missing Data) และความผิดปกติทั่วไปจากกฎที่ตั้งไว้ เช่น การเก็บตัวอักษรในคอลัมน์ตัวเลข หากมีควรปรับค่ารวมทั้งระบุกระบวนการจัดการแก้ไข
- ตรวจสอบว่าชุดข้อมูลมีคอลัมน์ที่ถูกสร้างขึ้นมาเพื่อทดสอบระบบงานหรือไม่ หากมีควรลบทิ้ง

การตรวจสอบคุณภาพแบบพื้นฐานนี้ สามารถทำแบบออฟไลน์ได้ เพื่อไม่ให้เกิดการใช้งานระบบฐานข้อมูล รวมทั้งเจ้าหน้าที่สารสนเทศของหน่วยงานสามารถพัฒนาโปรแกรมย่อยเพื่อใช้ตรวจสอบข้อมูลแบบกึ่งอัตโนมัติได้อีกด้วย

5.3 ระบบบริการและแลกเปลี่ยนข้อมูลกลาง

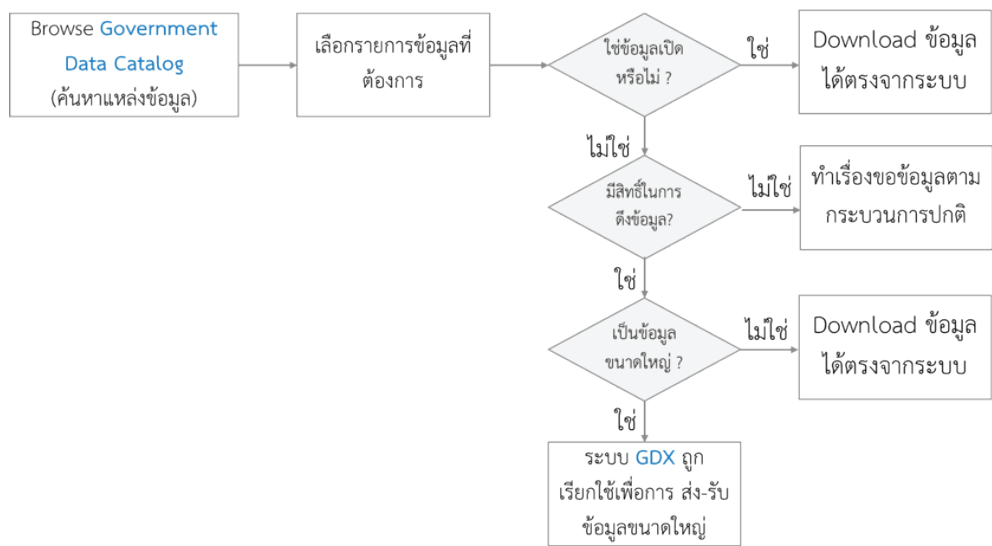
การกำกับดูแลชั้นความลับข้อมูล สิทธิการเข้าถึงข้อมูล กลไกและระบบสารสนเทศ เพื่อให้การให้บริการและแลกเปลี่ยนข้อมูลกลาง จำเป็นจะต้องเชื่อมต่อกับระบบนามาตรฐานข้อมูลภาครัฐ เพื่อยืนยันตัวตนผู้ใช้ และตรวจสอบสิทธิในการเข้าถึงข้อมูล ให้เป็นไปตามกฎหมายที่หน่วยงานเจ้าของข้อมูลระบุ ระบบบริการและแลกเปลี่ยนข้อมูลจะต้องบันทึกรายการการเข้าถึงข้อมูลและถ่ายโอนข้อมูลทั้งหมด (Access Log) เพื่อให้หน่วยงานเจ้าของข้อมูลสามารถตรวจสอบได้ และเพื่อให้สามารถสรุปผลเชิงสถิติเรื่องการร้องขอใช้ข้อมูลของทั้งหน่วยงานภายในและภายนอก ภาพที่ 5.1 แสดงระบบสารสนเทศที่จำเป็นในการให้บริการข้อมูลกลางต่อยอดจากภาพที่ 4.1 (การให้บริการรายการข้อมูลภาครัฐ)



ภาพที่ 5.1 ระบบสารสนเทศกลางเพื่อให้บริการเชื่อมโยงและแลกเปลี่ยนข้อมูล

จากภาพที่ 5.1 จะเห็นได้ว่ามีระบบสารสนเทศที่จำเป็นต้องเชื่อมต่อและทำงานร่วมกัน 3 ระบบ คือ 1.) ระบบรายการข้อมูลภาครัฐ (Government Data Catalog) [รายละเอียดนำเสนอในบทที่ 4] 2.) ระบบเชื่อมโยงแลกเปลี่ยนข้อมูลภาครัฐ (Government Data Exchange หรือ GDX) [รายละเอียดนำเสนอในหัวข้อที่ 5.3.1] และ 3.) ระบบบริการข้อมูลเปิด (open data sandbox) [รายละเอียดนำเสนอในหัวข้อที่ 5.3.2]

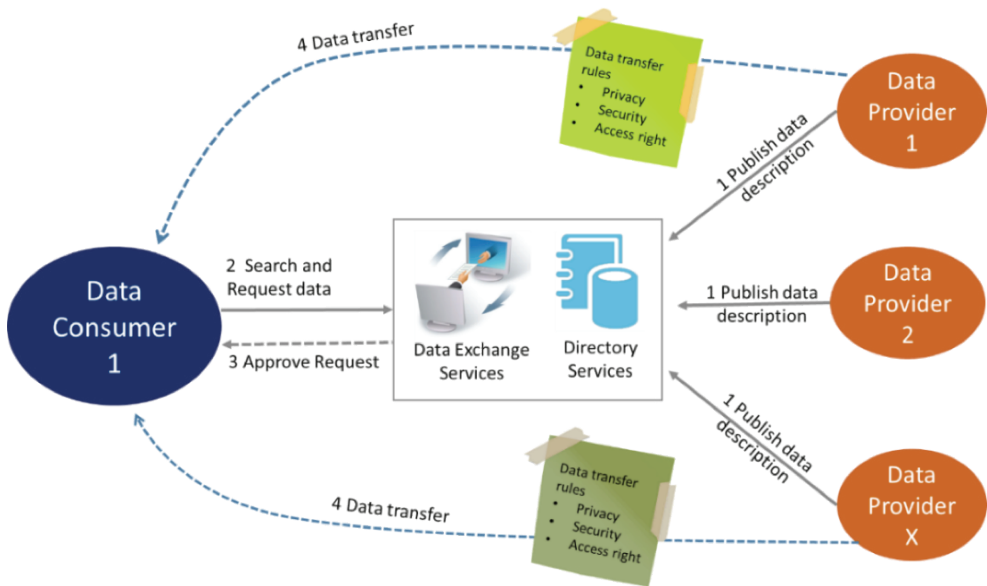
การทำงานร่วมกันของระบบสารสนเทศในภาพที่ 5.1 จะช่วยให้ผู้ใช้ข้อมูล (User หรือ Program) สามารถทำการสืบค้น ระบุ และ ร้องขอข้อมูลได้อย่างเป็นระบบ ควบคุมได้ ส่งผลให้การใช้ประโยชน์ข้อมูลในอนาคตเป็นรูปธรรม ภาพที่ 5.2 แสดงการไหล (Workflow) ของการใช้งานระบบสารสนเทศทั้ง 3 ระบบ



ภาพที่ 5.2 Workflow การใช้งานระบบสารสนเทศเพื่อการบริหารข้อมูล

5.3.1 ระบบเชื่อมโยงแลกเปลี่ยนข้อมูลภาครัฐ (Government Data Exchange)

กลไกการบริการและการแลกเปลี่ยนข้อมูลเพื่อการวิเคราะห์ข้อมูลขนาดใหญ่ นั้น ควรถูกออกแบบในลักษณะที่หน่วยงานเจ้าของข้อมูลไม่จำเป็นต้องส่งข้อมูลดิบ มาจัดเก็บที่ศูนย์บริการข้อมูลกลาง เพียงแต่เผยแพร่เมตาดาต้าขึ้นระบบนามุกรมข้อมูล ภาครัฐเพื่อให้ระบบสารสนเทศอื่น ๆ สามารถสืบค้นและขอเชื่อมโยงข้อมูลได้ โดยขั้นตอน การแลกเปลี่ยนข้อมูลระหว่างหน่วยงาน อธิบายได้ตามภาพที่ 5.3



ภาพที่ 5.3 กลไกการบริการและการแลกเปลี่ยนข้อมูล

จากภาพการบริการและแลกเปลี่ยนข้อมูลจะมี 4 ขั้นตอน ดังต่อไปนี้

- 1.) หน่วยงานภาครัฐเผยแพร่เมตาดาต้ามาที่ระบบนามานุกรมข้อมูลภาครัฐ
- 2.) ผู้ใช้งานที่ต้องการร้องขอข้อมูลจะทำการสืบค้น ระบุข้อมูลที่ต้องการ และยื่นคำร้องขอข้อมูลผ่านระบบนามานุกรมข้อมูลภาครัฐ ซึ่งระบบดังกล่าวจะยืนยันตัวตน ตรวจสอบสิทธิของผู้ร้องขอ และตรวจสอบว่าคำขอข้อมูลมีความเหมาะสมตามกฎหมายเกณฑ์ที่ระบุ โดยหน่วยงานเจ้าของข้อมูลหรือไม่ ตัวอย่างเช่น ข้อมูลที่มีชั้นความลับระดับใด ผู้ขอมีสิทธิหรือไม่ คำขอต้องการปริมาณข้อมูลมากเพียงใด หากคำขออยู่ในเกณฑ์ปกติการอนุมัติจะสามารถทำได้โดยระบบอัตโนมัติ มิเช่นนั้นจะต้องมีการทำหนังสือเพื่อขอข้อมูลอย่างเป็นทางการไปที่หน่วยงานเจ้าของข้อมูล นอกจากนี้คำร้องขอข้อมูลทั้งหมดจะถูกบันทึกเพื่อประโยชน์ในการตรวจสอบในภายหลัง ทั้งนี้การระบุและร้องขอข้อมูลสามารถทำได้ทั้งผ่านส่วนติดต่อผู้ใช้งาน (Graphical User Interface : GUI) ของระบบนามานุกรม หรือให้ระบบสารสนเทศเชื่อมต่อตรงผ่านส่วนติดต่อโปรแกรม API
- 3.) หากคำร้องขอข้อมูลได้รับการอนุมัติ ระบบจะมีการแจ้งไปยังผู้ร้องขอ และคำร้องพร้อมรายละเอียดจะถูกส่งผ่านไปยังระบบบริการและแลกเปลี่ยนข้อมูลกลาง โดยอัตโนมัติ
- 4.) ระบบบริการและแลกเปลี่ยนข้อมูลกลาง⁸ จะมีการเชื่อมต่อกับหน่วยงานเจ้าของข้อมูลและหน่วยงานผู้ให้บริการทั้งหมด รวมทั้งเก็บข้อมูลด้านเครือข่าย เช่น IP Address และเกณฑ์ในการส่ง-รับข้อมูลของหน่วยงานเจ้าของข้อมูล เพื่อจุดประสงค์ในการถ่ายโอนข้อมูล ซึ่งการถ่ายโอนข้อมูลจากต้นทางไปยังปลายทางสามารถทำได้ทั้งแบบผ่านและไม่ผ่านระบบกลาง การที่ข้อมูลไม่จำเป็นต้องวิ่งผ่านระบบบริการข้อมูลกลางจะสามารถลดปัญหาคอขวดที่อาจเกิดขึ้นในกรณีที่มียารายการคำขอจำนวนมาก ทั้งนี้ช่องทางการส่งข้อมูลจะขึ้นอยู่กับเกณฑ์ที่หน่วยงานเจ้าของข้อมูลระบุไว้ ตัวอย่างเช่น

⁸ หมายเหตุ: ระบบนามานุกรมข้อมูลภาครัฐ และระบบบริการข้อมูลกลาง ไม่จำเป็นต้องติดตั้งในดาต้าเซนเตอร์เดียวกัน ระบบทั้งสองสามารถทำงานร่วมกันได้ผ่านกลไก Web Service ได้

- Application Programming Interface หรือ API คือ การโอนย้ายข้อมูลอัตโนมัติผ่านการเรียกใช้งานส่วนโปรแกรมที่หน่วยงานเจ้าของข้อมูลจัดเตรียมไว้ให้ โดยระบบงานของผู้ร้องขอข้อมูลจะสามารถดึงข้อมูลตรงจากระบบฐานข้อมูล (Database) หรือระบบทะเลสาบข้อมูล (Data Lake) ของหน่วยงานเจ้าของข้อมูลได้โดยตรง
- Secured File Transfer Protocol หรือ SFTP ทำได้ 2 วิธี คือ 1) เจ้าของข้อมูลทำการอัปโหลด (Upload) ไฟล์ข้อมูลเข้าไปที่พื้นที่จัดเก็บชั่วคราวของระบบบริการข้อมูลกลาง เพื่อให้ผู้ร้องขอข้อมูลเข้ามาดึงข้อมูลไปใช้ในภายหลัง 2) หน่วยงานเจ้าของข้อมูลถ่ายโอน (Transfer) ข้อมูลไปยังผู้รับโดยตรงผ่านระบบอัตโนมัติด้วยโปรโตคอล SFTP การส่งข้อมูลลักษณะนี้ต้องมีรหัสผู้ใช้และรหัสผ่านสำหรับการเข้าสู่ระบบที่ชัดเจน
- Removable Storage หรือ Email คือการทำสำเนาไฟล์ข้อมูลโดยใส่รหัสป้องกันการเปิดไฟล์ และนำไฟล์ดังกล่าวใส่ลงใน Removable Storage เพื่อจัดส่ง หรือส่งทาง Email การส่งลักษณะนี้เหมาะกับหน่วยงานที่มีข้อจำกัดทางด้านระบบเครือข่ายคอมพิวเตอร์

5.3.2 ระบบบริการข้อมูลเปิด (Open Data Sandbox)

จากภาพที่ 5.1 หน่วยงานกลาง (สำนักงานพัฒนารัฐบาลดิจิทัล สำนักงานสถิติแห่งชาติ และสถาบันส่งเสริมการวิเคราะห์และบริหารข้อมูลขนาดใหญ่ภาครัฐ) จะร่วมกันส่งเสริมให้หน่วยงานภาครัฐต่าง ๆ สร้างรายการชุดข้อมูลเพื่อนำมาวางไว้ในระบบนามาณุกรมข้อมูลภาครัฐ และหากชุดข้อมูลใดไม่มีชั้นความลับ จัดเป็นข้อมูลเปิดที่ให้ประโยชน์กับสาธารณะได้ หน่วยงานกลางจะกระตุ้นและส่งเสริมให้หน่วยงานเจ้าของข้อมูลทำสำเนาข้อมูลเปิดมาไว้ในระบบกลางที่เรียกว่าระบบบริการข้อมูลเปิด (Open Data Sandbox) เพื่อประโยชน์ต่อภาครัฐ ภาคเอกชนและประชาชนในการศึกษา วิเคราะห์ วิจัย และวางแผนในมิติต่าง ๆ โดยรูปแบบการสร้างระบบบริการข้อมูลเปิดควรอ้างอิงจากระบบข้อมูลเปิดของประชาคมโลกที่ชื่อว่า Kaggle.com ซึ่งเป็นประชาคมด้านวิทยาศาสตร์ข้อมูลที่ใหญ่ที่สุดในโลก⁹ ซึ่งให้บริการข้อมูลเปิด สอนเรื่องวิธีวิเคราะห์ และให้บริการเครื่องมือสำหรับวิเคราะห์ วิจัย จำนวนมาก รวมทั้งมีเนื้อหาและตัวอย่างกรณีศึกษาการใช้ประโยชน์ข้อมูลในทุกโดเมน

⁹ <https://en.wikipedia.org/wiki/Kaggle>

ซึ่งการบริการครบวงจรในลักษณะนี้จะเป็นประโยชน์ในการศึกษาด้านข้อมูลของบุคลากรภาครัฐและประชาชนทั่วไปเป็นอย่างมาก

นอกจากระบบสารสนเทศที่กล่าวข้างต้น กลไกการบริการและแลกเปลี่ยนข้อมูลกลางยังต้องมีกระบวนการบนระบบสารสนเทศที่ให้ผู้ใช้งานสามารถจัดการข้อมูลที่มีความผิดปกติได้ตามขั้นตอนดังต่อไปนี้

- 1.) Confirm : ตรวจสอบและยืนยันว่าชุดข้อมูลมีปัญหา
- 2.) Notify : แจ้งเตือนให้ผู้ที่เกี่ยวข้องหรือหน่วยงานเจ้าของข้อมูลนี้เกี่ยวกับปัญหาในชุดข้อมูล
- 3.) Log : ทำการเก็บ log ปัญหาที่เกิดขึ้นไว้เพื่อติดตามและสรุปผล
- 4.) Diagnose : วิเคราะห์ปัญหาของข้อมูลพื้นฐานได้
- 5.) Evaluate Option : ประเมินต้นตอของปัญหา เช่น การเก็บข้อมูล อัลกอริทึมในการคำนวณข้อมูล
- 6.) Correct: แก้ไขข้อมูลที่ผิดปกติให้ถูกต้อง

ในช่วงแรกของการให้บริการแลกเปลี่ยนข้อมูล กระบวนการจัดการข้อมูล จะช่วยให้เกิดการแจ้งเตือนระหว่างหน่วยงานเพื่อให้เกิดการพัฒนาคุณภาพข้อมูลโดยรวม และทำให้หน่วยงานสามารถให้บริการข้อมูลที่มีระดับคุณภาพที่ยอมรับ (Data Quality Service Level Agreements : DQ SLAs) ได้ในอนาคต



พระราชบัญญัติคุ้มครองข้อมูลส่วนบุคคล พ.ศ. 2562 Law and Regulations

เพื่อให้หน่วยงานภาครัฐสามารถดำเนินการแลกเปลี่ยนข้อมูลหรือเปิดเผยข้อมูลให้หน่วยงานอื่นได้อย่างมีประสิทธิภาพและมั่นคงปลอดภัย หน่วยงานจำเป็นต้องศึกษากฎ ระเบียบกระทรวง กฎหมายที่เกี่ยวข้อง ทั้งของหน่วยงาน พระราชบัญญัติข้อมูลข่าวสารของราชการ พ.ศ. 2540 และพระราชบัญญัติคุ้มครองข้อมูลส่วนบุคคล พ.ศ. 2562 ในบทนี้จะนำเสนอบทสรุปจากพระราชบัญญัติคุ้มครองข้อมูลส่วนบุคคล โดยเฉพาะเรื่องการเก็บรวบรวม ใช้ หรือเปิดเผยข้อมูลส่วนบุคคลได้โดยชอบด้วยกฎหมาย เพื่อใช้เป็นแนวทางในการจัดการเรื่องธรรมาภิบาลข้อมูล

6.1 ภาพรวมของพระราชบัญญัติคุ้มครองข้อมูลส่วนบุคคล พ.ศ. 2562

พระราชบัญญัติคุ้มครองข้อมูลส่วนบุคคล พ.ศ. 2562 ลงประกาศในราชกิจจานุเบกษา เมื่อวันที่ 27 พฤษภาคม 2562 ว่าด้วยการกำหนดหลักการคุ้มครองข้อมูลส่วนบุคคลเป็นการทั่วไป ครอบคลุมทุกภาคส่วน โดยกำหนดหลักเกณฑ์ กลไก หรือมาตรการกำกับดูแลเกี่ยวกับการให้ความคุ้มครองข้อมูลส่วนบุคคล เพื่อให้การคุ้มครองข้อมูลส่วนบุคคลมีประสิทธิภาพและเพื่อให้มีมาตรการเยียวยาเจ้าของข้อมูลส่วนบุคคลจากการถูกละเมิดสิทธิที่มีประสิทธิภาพ

พ.ร.บ. ดังกล่าวประกอบด้วย 7 หมวด และบทเฉพาะกาล รวม 96 มาตรา โดยมีการแบ่งวันที่มีผลบังคับใช้เป็น 2 กลุ่ม คือ กลุ่มที่ 1 บังคับใช้ตั้งแต่วันที่ 28 พฤษภาคม 2562 รวม 50 มาตรา (หมวด 1 คณะกรรมการคุ้มครองข้อมูลส่วนบุคคล และหมวด 4 สำนักงานคณะกรรมการคุ้มครองข้อมูลส่วนบุคคล) และกลุ่มที่ 2 บังคับใช้ตั้งแต่วันที่ 27 พฤษภาคม 2563 รวม 46 มาตรา (หมวด 2 การคุ้มครองข้อมูลส่วนบุคคล หมวด 3 สิทธิของเจ้าของข้อมูลส่วนบุคคล หมวด 5 การร้องเรียน หมวด 6 ความรับผิดชอบทางแพ่ง และหมวด 7 บทกำหนดโทษ)

กลุ่มที่ 1 บังคับใช้ 28 พฤษภาคม 2562 รวม 50 มาตรา	
มาตรา 1 - 7	ชื่อพระราชบัญญัติ วันที่มีผลบังคับใช้ ขอบเขตการบังคับใช้ นิยาม รัฐมนตรี ผู้รักษาการ
หมวด 1 คณะกรรมการคุ้มครองข้อมูลส่วนบุคคล	
มาตรา 8 - 18	การสรรหา คุณสมบัติ การแต่งตั้ง และอำนาจหน้าที่ของคณะกรรมการคุ้มครองข้อมูลส่วนบุคคล
หมวด 4 สำนักงานคณะกรรมการคุ้มครองข้อมูลส่วนบุคคล	
มาตรา 43 - 47	การจัดตั้ง วัตถุประสงค์ และอำนาจหน้าที่ของสำนักงานฯ
มาตรา 48 - 56	การสรรหา คุณสมบัติ การแต่งตั้ง และอำนาจหน้าที่ของคณะกรรมการกำกับสำนักงานคณะกรรมการคุ้มครองข้อมูลส่วนบุคคล
มาตรา 57 - 65	การสรรหา คุณสมบัติ การแต่งตั้ง และอำนาจหน้าที่ของเลขาธิการคณะกรรมการคุ้มครองข้อมูลส่วนบุคคล
มาตรา 66 - 70	การบริหารงานของสำนักงาน
บทเฉพาะกาลที่เกี่ยวข้องกับคณะกรรมการฯ และสำนักงานฯ	
มาตรา 91 - 94	แนวทางในการดำเนินการในวาระเริ่มแรกก่อนจัดตั้งสำนักงานให้แล้วเสร็จ
กลุ่มที่ 2 บังคับใช้ 27 พฤษภาคม 2563 รวม 46 มาตรา	
หมวด 2 การคุ้มครองข้อมูลส่วนบุคคล	
มาตรา 19 - 27	การขอความยินยอม และข้อมูลส่วนบุคคลของผู้อย่าง การเก็บรวบรวม ใช้ และเปิดเผยข้อมูลส่วนบุคคล ข้อยกเว้นในการเก็บรวบรวม ใช้ และเปิดเผยข้อมูลส่วนบุคคล การเก็บรวบรวม ใช้ และเปิดเผยข้อมูลส่วนบุคคลที่มีความละเอียดอ่อน
มาตรา 28 - 29	การส่งหรือโอนข้อมูลส่วนบุคคลไปต่างประเทศ
หมวด 3 สิทธิของเจ้าของข้อมูลส่วนบุคคล	
มาตรา 30 - 36	การกำหนดสิทธิของเจ้าของข้อมูล ได้แก่ การขอเข้าถึงและขอรับสำเนาข้อมูล (Access) การขอโอนหรือรับข้อมูลที่เก็บแบบอัตโนมัติ (Data Portability) การขอคัดค้านการเก็บข้อมูล (Objection) การขอให้ลบหรือทำลาย (Erasure) การระงับการใช้ (Restriction) การขอทำให้ข้อมูลถูกต้อง (Rectification)
มาตรา 37 - 39	หน้าที่ของผู้ควบคุมข้อมูลส่วนบุคคล (Data Controller) รายการของกิจกรรม และต้องแจ้งเหตุการละเมิดข้อมูลส่วนบุคคลแก่สำนักงาน
มาตรา 40	หน้าที่ของผู้ประมวลผลข้อมูลส่วนบุคคล (Data Processor) ต้องจัดให้มีมาตรการรักษาความมั่นคงปลอดภัย และจัดทำบันทึกรายการของกิจกรรม
มาตรา 41 - 42	หน้าที่ของเจ้าหน้าที่คุ้มครองข้อมูลส่วนบุคคล (Data Protection Officer, DPO)

หมวด 5 การร้องเรียน	
มาตรา 71 - 76	การแต่งตั้งคณะกรรมการผู้เชี่ยวชาญ เพื่อพิจารณาตรวจสอบเรื่องร้องเรียนและไกล่เกลี่ยการแต่งตั้งและอำนาจหน้าที่ของพนักงานเจ้าหน้าที่
หมวด 6 ความรับผิดชอบทางแพ่ง	
มาตรา 77 - 78	การกำหนดค่าสินไหมทดแทนและอายุความ
หมวด 7 บทกำหนดโทษ	
มาตรา 79 - 81	การกำหนดโทษอาญาสูงสุดจำคุกไม่เกิน 1 ปี หรือปรับไม่เกิน 1,000,000 บาท หรือทั้งจำทั้งปรับ
มาตรา 82 - 90	การโทษทางปกครอง สูงสุด 5,000,000 บาท
บทเฉพาะกาล	
มาตรา 95	การเก็บรวบรวมและใช้ข้อมูลเดิม
มาตรา 96	การออกกฎหมายลำดับรอง

6.2 พระราชบัญญัติกำหนดหน่วยงานและกิจการที่ผู้ควบคุมข้อมูลส่วนบุคคลไม่อยู่ภายใต้บังคับแห่งพระราชบัญญัติคุ้มครองข้อมูลส่วนบุคคล พ.ศ.2562 และ พ.ศ. 2563

เนื่องจากพระราชบัญญัติคุ้มครองข้อมูลส่วนบุคคลกำหนดหลักเกณฑ์ วิธีการ และเงื่อนไขในการคุ้มครองข้อมูลส่วนบุคคลไว้โดยละเอียด เพื่อให้การคุ้มครองข้อมูลส่วนบุคคลเป็นไปอย่างมีประสิทธิภาพ ผู้ควบคุมข้อมูลส่วนบุคคลทุกรายทั่วประเทศทั้งภาครัฐและภาคเอกชนมีหน้าที่ต้องดำเนินการตามหลักเกณฑ์ วิธีการ และเงื่อนไขดังกล่าวโดยเคร่งครัด ซึ่งการปฏิบัติตามกฎหมายกำหนดนั้นมีรายละเอียดมากและซับซ้อน กับต้องใช้เทคโนโลยีช่วยให้การคุ้มครองข้อมูลส่วนบุคคลเป็นไปอย่างมีประสิทธิภาพสมดังเจตนารมณ์ของกฎหมาย แต่ปรากฏว่า ผู้ควบคุมข้อมูลส่วนบุคคลซึ่งเป็นหน่วยงานและกิจการต่าง ๆ ทั้งภาครัฐและเอกชนจำนวนมากทั่วประเทศยังไม่พร้อมที่จะปฏิบัติตามกฎหมายดังกล่าว จึงได้มีการตราพระราชบัญญัติกำหนดหน่วยงานและกิจการที่ผู้ควบคุมข้อมูลส่วนบุคคลไม่อยู่ภายใต้บังคับแห่งพระราชบัญญัติคุ้มครองข้อมูลส่วนบุคคล พ.ศ. 2562 และ พ.ศ. 2563 มาใช้บังคับตั้งแต่วันที่ 27 พฤษภาคม 2563 จนถึงวันที่ 31 พฤษภาคม 2564 เพื่อยกเว้นให้ผู้ควบคุมข้อมูลส่วนบุคคล 22 หน่วยงานหรือกิจการตามบัญชีท้าย ยังไม่ต้องปฏิบัติตามบางหมวดในกฎหมาย ได้แก่ การคุ้มครองข้อมูลส่วนบุคคล สิทธิของเจ้าของข้อมูล การร้องเรียน ความรับผิดชอบทางแพ่ง และบทกำหนดโทษ แต่ผู้ควบคุมข้อมูลส่วนบุคคลยังคงต้องจัดให้มีมาตรการรักษา

ความปลอดภัยของข้อมูลส่วนบุคคลให้เป็นไปตามประกาศกระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม เรื่อง มาตรฐานการรักษาความมั่นคงปลอดภัยของข้อมูลส่วนบุคคล พ.ศ. 2563 ซึ่งบังคับใช้ตั้งแต่ 18 กรกฎาคม 2563 จนถึง 31 พฤษภาคม 2564 เพื่อให้มีระยะเวลาในการเตรียมความพร้อมสำหรับปฏิบัติตามหลักเกณฑ์ในกฎหมายได้อย่างครบถ้วน เหมาะสม และมีประสิทธิภาพ

6.3 หลักการสำคัญตามพระราชบัญญัติคุ้มครองข้อมูลส่วนบุคคล พ.ศ. 2562

6.3.1 ข้อมูลส่วนบุคคล (Personal Data)

ข้อมูลเกี่ยวกับบุคคลซึ่งทำให้สามารถระบุตัวบุคคลนั้นได้ไม่ว่าทางตรงหรือทางอ้อม เช่น ชื่อ-สกุล ที่อยู่ เลขบัตรประชาชน ข้อมูลสุขภาพ หมายเลขโทรศัพท์ e-mail ประวัติอาชญากรรม เป็นต้น

6.3.2 บุคคลที่เกี่ยวข้องกับข้อมูลส่วนบุคคล

๐ เจ้าของข้อมูลส่วนบุคคล (Data Subject)

ตามกฎหมายไม่ได้ให้คำนิยามไว้ แต่โดยหลักการทั่วไปแล้วหมายถึง บุคคลที่ข้อมูลนั้นระบุไปถึง

๐ ผู้ควบคุมข้อมูลส่วนบุคคล (Data Controller)

๐ บุคคลหรือนิติบุคคลซึ่งมีอำนาจหน้าที่ตัดสินใจเกี่ยวกับการเก็บรวบรวม ใช้ หรือเปิดเผยข้อมูลส่วนบุคคล เช่น หน่วยงานของรัฐ หรือเอกชนโดยทั่วไป ที่เก็บรวบรวม ใช้หรือเปิดเผยข้อมูลส่วนบุคคลของประชาชนหรือลูกค้าที่มาใช้บริการ

๐ ผู้ควบคุมข้อมูลส่วนบุคคลมีหน้าที่สำคัญที่กฎหมายกำหนดไว้ เช่น จัดให้มีมาตรการรักษาความมั่นคงปลอดภัยข้อมูลส่วนบุคคล ดำเนินการเพื่อป้องกันมิให้ผู้อื่นใช้หรือเปิดเผยข้อมูลส่วนบุคคลโดยมิชอบ แจ้งเหตุการละเมิดข้อมูลส่วนบุคคลให้สำนักงานคุ้มครองข้อมูลส่วนบุคคลทราบภายใน 72 ชั่วโมงนับแต่ทราบเหตุ แต่งตั้งเจ้าหน้าที่คุ้มครองข้อมูลส่วนบุคคล (Data Protection Officer : DPO) เพื่อตรวจสอบการทำงานของตน เป็นต้น

๐ ผู้ประมวลผลข้อมูลส่วนบุคคล (Data Processor)

บุคคลหรือนิติบุคคลซึ่งดำเนินการเกี่ยวกับการเก็บรวบรวม ใช้ หรือเปิดเผยข้อมูลส่วนบุคคลตามคำสั่งหรือในนามของผู้ควบคุมข้อมูลส่วนบุคคล เช่น การให้บริการคลาวด์ (Cloud Service) เป็นต้น

ผู้ประมวลผลข้อมูลส่วนบุคคลมีหน้าที่หลัก คือ ดำเนินการตามคำสั่งที่ได้รับจากผู้ควบคุมข้อมูลส่วนบุคคลเท่านั้น เว้นแต่คำสั่งนั้นขัดต่อกฎหมายหรือบทบัญญัติในการคุ้มครองข้อมูลส่วนบุคคล

6.3.3 การเก็บรวบรวม ใช้ หรือเปิดเผยข้อมูลส่วนบุคคลได้โดยชอบด้วยกฎหมาย

การเก็บรวบรวม ใช้ หรือเปิดเผยข้อมูลส่วนบุคคลดังกล่าวจะชอบด้วยกฎหมาย หากดำเนินการตามหลักการใดหลักการหนึ่ง ดังต่อไปนี้

- **ความยินยอม (Consent) ของเจ้าของข้อมูล [มาตรา 19]**
 - เจ้าของข้อมูลส่วนบุคคลให้ความยินยอมในการเก็บรวบรวม ใช้ หรือเปิดเผยข้อมูลส่วนบุคคล
 - ต้องแจ้งวัตถุประสงค์ของการเก็บรวบรวม ใช้หรือเปิดเผยข้อมูลส่วนบุคคล
 - มีแบบหรือข้อความที่อ่านแล้วเข้าใจได้โดยง่าย และต้องไม่เป็นการหลอกลวง
 - เจ้าของข้อมูลส่วนบุคคลจะถอนความยินยอมเมื่อใดก็ได้ ถ้าไม่มีข้อจำกัดสิทธิ เช่น มีกฎหมายที่กำหนดให้เก็บรวบรวมข้อมูลส่วนบุคคลนั้นไว้ก่อน
- **การศึกษาวิจัยทางวิทยาศาสตร์หรือประวัติศาสตร์ (Scientific or Historical Research) [มาตรา 24 (1)]**
 - จัดทำเอกสารประวัติศาสตร์ จดหมายเหตุ การศึกษาวิจัย และสถิติ
- **ความสำคัญต่อชีวิต (Vital Interest) ของเจ้าของข้อมูล [มาตรา 24 (2)]**
 - เพื่อป้องกันหรือระงับอันตรายต่อชีวิต ร่างกาย หรือสุขภาพของบุคคล เช่น การเข้ารับบริการทางการแพทย์ ณ โรงพยาบาล
- **สัญญา (Contract) ระหว่างผู้ควบคุมข้อมูลกับเจ้าของข้อมูล [มาตรา 24 (3)]**
 - เป็นการจำเป็นเพื่อการปฏิบัติตามสัญญา เช่น เจ้าของข้อมูลส่วนบุคคลทำสัญญากู้ยืมเงินจากธนาคาร ธนาคารสามารถเก็บรวบรวม ใช้ หรือเปิดเผยข้อมูลส่วนบุคคลนั้นได้ตามวัตถุประสงค์ของสัญญา

○ **ประโยชน์สาธารณะ (Public Task) ของผู้ควบคุมข้อมูล [มาตรา 24 (4)]**

- เป็นการจำเป็นเพื่อการปฏิบัติหน้าที่ในการดำเนินการกิจเพื่อประโยชน์สาธารณะ หรือปฏิบัติหน้าที่ในการใช้อำนาจอธิปไตย เช่น หน่วยงานของรัฐ จัดทำข้อมูลขนาดใหญ่เพื่อแก้ปัญหาความยากจนของเกษตรกร

○ **ผลประโยชน์อันชอบด้วยกฎหมาย (Legitimate Interest) ของผู้ควบคุมข้อมูล [มาตรา 24 (5)]**

- เป็นการจำเป็นเพื่อประโยชน์โดยชอบด้วยกฎหมายของผู้ควบคุมข้อมูลส่วนบุคคล หรือของบุคคลหรือนิติบุคคลอื่น เช่น บริษัทเอกชนติดตั้งกล้องวงจรปิดภายในอาคารเพื่อรักษาความปลอดภัย ซึ่งบริษัทสามารถเก็บรวบรวมภาพถ่ายซึ่งเป็นข้อมูลส่วนบุคคลของบุคคลที่อยู่ในบริเวณดังกล่าวได้

○ **การปฏิบัติตามกฎหมาย (Legal Obligations) ของผู้ควบคุมข้อมูล [มาตรา 24 (6)]**

- เป็นการปฏิบัติตามกฎหมาย

นอกจากหลักการข้างต้นแล้ว มีข้อมูลส่วนบุคคลอีกประเภทซึ่งเรียกว่า ข้อมูลส่วนบุคคลที่มีความละเอียดอ่อน (Sensitive Personal Data) เช่น เชื้อชาติ, ประวัติอาชญากรรม, ข้อมูลพันธุกรรม, พฤติกรรมทางเพศ เป็นต้น การเก็บรวบรวม ใช้ หรือเปิดเผยข้อมูลส่วนบุคคลดังกล่าวจะมีหลักการที่เข้มงวดกว่าข้อมูลส่วนบุคคลทั่วไป โดยจะกระทำได้หากดำเนินการตามหลักการใดหลักการหนึ่ง เช่น ได้รับความยินยอมโดยชัดแจ้ง (Explicit Consent) จากเจ้าของข้อมูลส่วนบุคคล เพื่อป้องกันหรือระงับอันตรายต่อชีวิต ร่างกาย หรือสุขภาพของบุคคลซึ่งเจ้าของข้อมูลส่วนบุคคลไม่สามารถให้ความยินยอมได้ เป็นต้น

6.3.4 การส่งหรือโอนข้อมูลส่วนบุคคลไปยังต่างประเทศ

ประเทศปลายทางที่รับข้อมูลส่วนบุคคลต้องมีมาตรฐานการคุ้มครองข้อมูลส่วนบุคคลที่เพียงพอ ทั้งนี้ ต้องเป็นไปตามหลักเกณฑ์ที่คณะกรรมการประกาศ

6.3.5 สิทธิของเจ้าของข้อมูลส่วนบุคคล (Data Subject Right) เช่น

○ **สิทธิขอเข้าถึงข้อมูลส่วนบุคคล (Right of Access)**

เจ้าของข้อมูลส่วนบุคคลมีสิทธิขอเข้าถึงและขอรับสำเนาข้อมูลส่วนบุคคลที่เกี่ยวข้องตนซึ่งอยู่ในความรับผิดชอบของผู้ควบคุมข้อมูลส่วนบุคคล

- สิทธิขอให้ลบหรือทำลาย หรือทำให้ข้อมูลส่วนบุคคลเป็นข้อมูลที่ไม่สามารถระบุตัวบุคคล (Right to Erasure (Also Known as Right to be Forgotten))

เจ้าของข้อมูลส่วนบุคคลมีสิทธิขอให้ลบหรือทำลาย หรือทำให้ข้อมูลส่วนบุคคลเป็นข้อมูลที่ไม่สามารถระบุตัวบุคคลได้ หากข้อมูลส่วนบุคคลที่หมดความจำเป็น หรือข้อมูลส่วนบุคคลที่ขอถอนความยินยอมแล้วได้

6.3.6 การร้องเรียน

เจ้าของข้อมูลส่วนบุคคลที่ถูกละเมิดข้อมูลส่วนบุคคลสามารถร้องเรียนต่อคณะกรรมการผู้เชี่ยวชาญซึ่งมีหน้าที่พิจารณาเรื่องร้องเรียนตามพระราชบัญญัตินี้ได้

6.3.7 ความรับผิดชอบและบทลงโทษ

1) ความรับผิดชอบทางแพ่ง

- ผู้กระทำละเมิดข้อมูลส่วนบุคคลต้องชดเชยค่าสินไหมทดแทนให้กับเจ้าของข้อมูลส่วนบุคคล ไม่ว่าการดำเนินการนั้นจะเกิดจากการกระทำโดยจงใจหรือประมาทเลินเล่อหรือไม่ก็ตาม
- ศาลมีอำนาจสั่งให้ชดเชยค่าสินไหมทดแทนเพิ่มเติมได้สองเท่าของค่าสินไหมทดแทนที่แท้จริง

2) โทษอาญา

- กำหนดบทลงโทษทางอาญาไว้สำหรับความผิดร้ายแรง เช่น การใช้หรือเปิดเผยข้อมูลส่วนบุคคลที่มีความละเอียดอ่อนโดยมิชอบ ล่วงรู้ข้อมูลส่วนบุคคลของผู้อื่นแล้วนำไปเปิดเผยแก่ผู้อื่นโดยมิชอบ
- ระวางโทษสูงสุดจำคุกไม่เกินหนึ่งปี หรือปรับไม่เกินหนึ่งล้านบาท หรือทั้งจำทั้งปรับ
- ในกรณีที่ผู้กระทำความผิดเป็นนิติบุคคล กรรมการหรือผู้จัดการ หรือบุคคลใดซึ่งรับผิดชอบในการดำเนินงานของนิติบุคคลนั้นอาจต้องร่วมรับผิดชอบในความผิดอาญาที่เกิดขึ้น

3) โทษทางปกครอง

- กำหนดโทษปรับทางปกครองสำหรับการกระทำความผิดที่เป็นการฝ่าฝืนหรือไม่ปฏิบัติตามที่กฎหมายกำหนด เช่น ไม่แจ้งวัตถุประสงค์ในการเก็บรวบรวมข้อมูลส่วนบุคคลให้เจ้าของข้อมูลส่วนบุคคลทราบ ขอความยินยอมโดยหลอกลวงเจ้าของข้อมูลส่วนบุคคล ไม่แต่งตั้ง DPO เป็นต้น
- โทษปรับทางปกครองสูงสุดห้าล้านบาท



การเป็นรัฐบาลเปิดและอัจฉริยะ Smart and Open Government

เพื่อเป็นจุดเริ่มต้นของการพัฒนาสู่ความเป็นรัฐบาลอัจฉริยะ สถาบันส่งเสริมการวิเคราะห์และบริหารข้อมูลขนาดใหญ่ภาครัฐ หรือ GBDi ร่วมมือกับหน่วยงานต่าง ๆ ภายใต้กระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคมและสำนักงานพัฒนารัฐบาลดิจิทัล ในการให้บริการการวิเคราะห์และกำกับดูแลข้อมูลของภาครัฐ เพื่อเป็นกรณีศึกษาการใช้ประโยชน์ข้อมูลขนาดใหญ่อย่างเป็นรูปธรรม โดยทำงานร่วมกับหน่วยงานภาครัฐในการพัฒนาระบบนำร่องในหลากหลายสาขา เช่น การส่งเสริมการท่องเที่ยว การแพทย์และการสาธารณสุข การใช้ข้อมูลจากสื่อสังคมออนไลน์กับการสำรวจความคิดเห็น การวิเคราะห์บริการภาครัฐ การส่งเสริมการเกษตร ฯลฯ การนำร่องดังกล่าวมีจุดประสงค์เพื่อให้บุคลากรภาครัฐได้เห็นถึงประโยชน์ของการใช้ข้อมูลประกอบการสร้างนโยบายและแผนงาน รวมถึงการเพิ่มประสิทธิภาพและพัฒนาการดำเนินงานในปัจจุบันเพื่อนำไปสู่ความเป็นรัฐบาลอัจฉริยะในอนาคต โดยตัวอย่างของกรณีศึกษา สรุปลงได้ดังต่อไปนี้

7.1 การเพิ่มศักยภาพด้านการท่องเที่ยวเมืองรอง

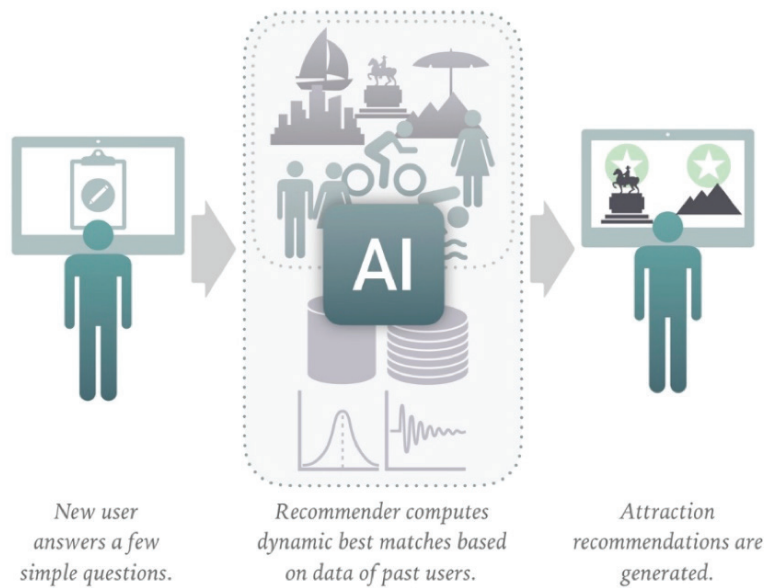
GBDi ดำเนินงานร่วมกับการท่องเที่ยวแห่งประเทศไทย และกระทรวงการท่องเที่ยวและกีฬาในการวิเคราะห์และนำเสนอสถานที่ท่องเที่ยวที่เหมาะสมกับนักท่องเที่ยวแต่ละกลุ่ม โดยมีจุดมุ่งหมายเพื่อเพิ่มความพึงพอใจและกระตุ้นนักท่องเที่ยว รวมถึงช่วยกระจายนักท่องเที่ยวสู่เมืองรอง (Emerging Destinations) ผ่านระบบแนะนำสถานที่ท่องเที่ยวอัตโนมัติ ใช้ข้อมูลของนักท่องเที่ยว (Demand Side) ได้แก่ ข้อมูลภาพรวมลักษณะและพฤติกรรมของนักท่องเที่ยว ร่วมกับข้อมูลของผู้ให้บริการ (Supply Side) ได้แก่ ข้อมูลสถานที่ท่องเที่ยว กิจกรรมการท่องเที่ยว ฯลฯ เพื่อนำมาวิเคราะห์และสร้างโมเดลทางคณิตศาสตร์ที่สามารถให้คำแนะนำสถานที่ท่องเที่ยวอัตโนมัติ รวมทั้งสามารถนำมาสนับสนุน

การท่องเที่ยวในมิติต่าง ๆ ได้ด้วย เช่น การจัดทำแคมเปญโฆษณาแหล่งท่องเที่ยวในเมืองรองที่มีประสิทธิภาพ การนำเสนอแหล่งท่องเที่ยวเมืองรองให้นักท่องเที่ยว เป็นต้น ทั้งนี้เป้าหมายหลักของโจทย์นี้ในเบื้องต้นเป็นกลุ่มของนักท่องเที่ยวชาวต่างชาติ

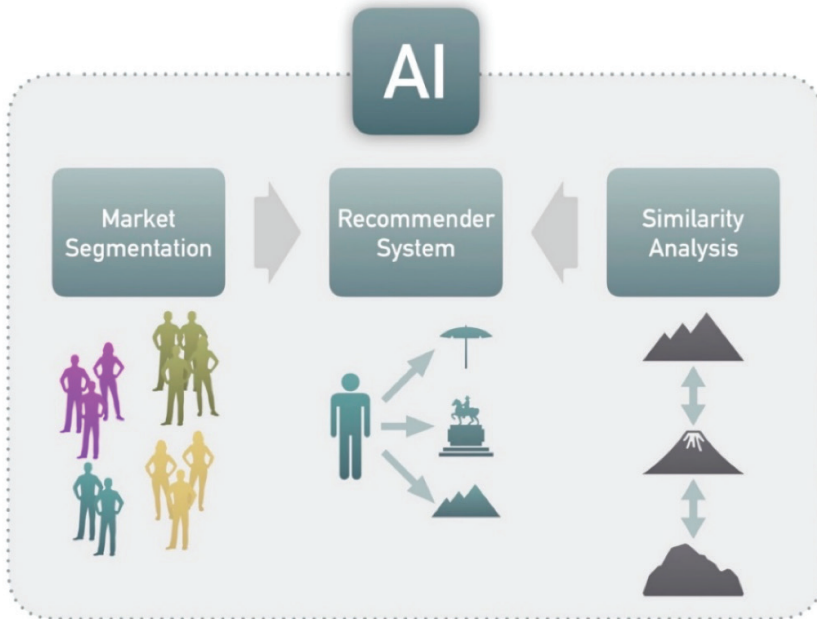
ระบบแนะนำสถานที่ท่องเที่ยวอัตโนมัตินี้ ประกอบด้วยขั้นตอนหลัก 3 ขั้นตอน ได้แก่ 1) ขั้นตอนการให้ข้อมูล ผ่านการตอบคำถามของนักท่องเที่ยวแต่ละราย 2) ขั้นตอนการประมวลผลโดยระบบปัญญาประดิษฐ์เพื่อคัดเลือกสถานที่ท่องเที่ยวที่เหมาะสมกับลักษณะเชิงพฤติกรรมของนักท่องเที่ยว และ 3) ขั้นตอนนำเสนอข้อมูลสถานที่ท่องเที่ยวที่ถูกแนะนำให้นักท่องเที่ยวได้เลือกพิจารณา ดังแสดงในภาพที่ 7.1

ระบบปัญญาประดิษฐ์ที่กล่าวขั้นตอนที่ 2 มีองค์ประกอบหลักพัฒนาจากโมเดลทางคณิตศาสตร์ มี 3 องค์ประกอบ ดังแสดงในภาพที่ 7.2

องค์ประกอบแรก คือ โมเดลการจัดกลุ่มนักท่องเที่ยว (Market Segmentation) เป็นการพิจารณาข้อมูลจากหลายมิติ เช่น กลุ่มหนึ่งอาจมีพฤติกรรมท่องเที่ยวในจังหวัดข้างเคียงกัน กลุ่มหนึ่งเน้นการใช้จ่ายในแหล่งท่องเที่ยวที่หรูหรา ในขณะที่อีกกลุ่มหนึ่งมีพฤติกรรมชอบท่องเที่ยวเชิงผจญภัยในราคาประหยัด เป็นต้น โดยใช้เทคนิควิธีการวิเคราะห์จัดกลุ่ม (Clustering Analysis) เพื่อจัดจำแนกประเภทของนักท่องเที่ยว จากข้อมูลผลสำรวจนักท่องเที่ยวต่างชาติ 120,000 คนในช่วง พ.ศ.2557-2560 สืบมาจากการท่องเที่ยวแห่งประเทศไทยใน 4 มิติ ได้แก่ เชิงลักษณะประชากร (Demographic) เชิงภูมิศาสตร์ (Geographic) เชิงลักษณะจิตวิทยา (Psychographic) และเชิงพฤติกรรม (Behavioral)



ภาพที่ 7.1 ภาพรวมของระบบการแนะนำสถานที่ท่องเที่ยวแบบอัตโนมัติในมุมมองของผู้ใช้งาน



ภาพที่ 7.2 องค์ประกอบในการสร้างระบบการแนะนำสถานที่ท่องเที่ยวอัตโนมัติ

องค์ประกอบที่สอง คือ โมเดลแนะนำสถานที่ท่องเที่ยว (Recommender System) พิจารณาจากข้อมูลของนักท่องเที่ยวรายใหม่ เพื่อเลือกกลุ่มที่มีลักษณะใกล้เคียงกับนักท่องเที่ยวรายนี้มากที่สุด และใช้กลุ่มนักท่องเที่ยวนี้สำหรับค้นหาสถานที่ท่องเที่ยวที่เหมาะสม โดยอ้างอิงกับข้อมูลจาก TripAdvisor ซึ่งเป็นแหล่งรวมข้อมูลการให้คะแนนสถานที่ท่องเที่ยวขนาดใหญ่และรวบรวมความเห็นจากนักท่องเที่ยวจำนวนหลายล้านคนทั่วโลก ข้อมูลคะแนนดังกล่าวจะช่วยให้ระบบปัญญาประดิษฐ์สามารถแนะนำสถานที่ท่องเที่ยวที่มีแนวโน้มจะสร้างความพึงพอใจให้กับนักท่องเที่ยวรายดังกล่าวมากที่สุด โดยอ้างอิงกับลักษณะกลุ่มนักท่องเที่ยวที่ได้รับการจำแนกจากองค์ประกอบแรกที่ตรงกับลักษณะของนักท่องเที่ยวรายนั้น ๆ

องค์ประกอบสุดท้าย คือ โมเดลแนะนำสถานที่ท่องเที่ยวที่คล้ายคลึงกัน (Similarity Analysis) เพื่อแนะนำสถานที่ท่องเที่ยวที่ยังไม่แพร่หลายหรือยังไม่เป็นที่รู้จักมากนักในหมู่นักท่องเที่ยว การนิยามความคล้ายคลึงในที่นี้อาจมีความเป็นไปได้ในหลายมิติ เช่น อยู่ในจังหวัดข้างเคียง มีขนาดใกล้เคียงกัน เป็นแหล่งท่องเที่ยวประเภทเดียวกัน ฯลฯ องค์ประกอบนี้จะช่วยเสริมข้อมูลองค์ประกอบที่สองให้สามารถแนะนำสถานที่ท่องเที่ยวในเมืองรอง (ซึ่งได้จากการสำรวจข้อมูลแหล่งท่องเที่ยวและกิจกรรมกว่า 8,000 รายการ โดยกระทรวงการท่องเที่ยวและกีฬา) เพิ่มเติมนอกเหนือจากแหล่งท่องเที่ยวหลักเดิม

ระบบปัญญาประดิษฐ์เพื่อแนะนำสถานที่ท่องเที่ยวนี้ อาจมีการปรับปรุงข้อมูลที่ใช้เพื่อสร้างระบบให้มีประสิทธิภาพตามความเหมาะสม โดยตัวอย่างข้อมูลที่ใช้ในการวิเคราะห์ร่วม ได้แก่ ข้อมูลโรงแรมที่จดทะเบียนจากกระทรวงมหาดไทย ข้อมูลโรงแรมจากการสำรวจจากสำนักงานสถิติแห่งชาติ ข้อมูลการเข้าออกประเทศของนักท่องเที่ยวจากสำนักงานตรวจคนเข้าเมือง ข้อมูลการเข้าพักโรงแรมของนักท่องเที่ยวจากกระทรวงการท่องเที่ยวและกีฬา เป็นต้น

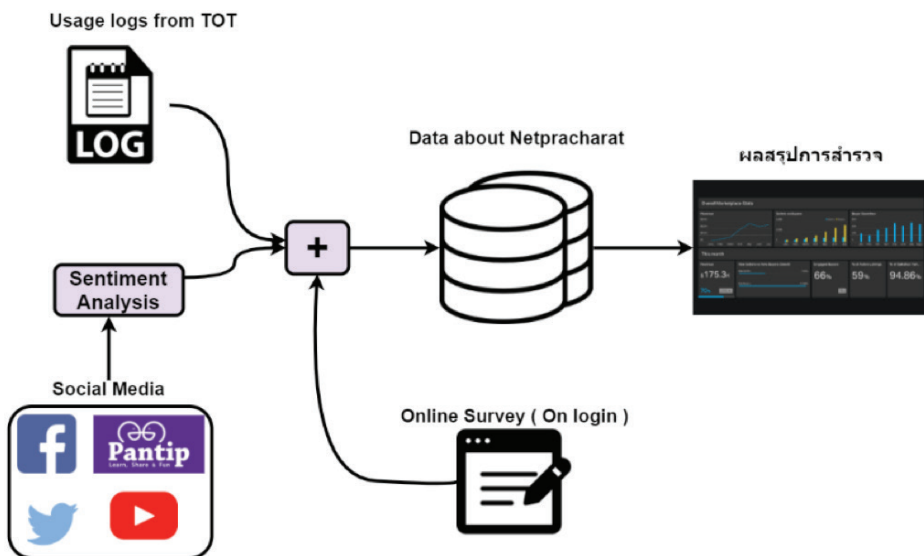
ระบบแนะนำสถานที่ท่องเที่ยวอัตโนมัตินี้ สามารถเปิดให้บริการข้อมูลสู่ภาคเอกชนได้ โดยใช้เสริมกับระบบหรือแอปพลิเคชันที่เอกชนมีอยู่ก่อนแล้ว เช่น แอปพลิเคชัน “ทักทาย” ที่เปิดโอกาสให้ผู้คนจากทั่วโลกสามารถเข้ามาค้นหาและวางแผนการเข้ามาเที่ยวในประเทศไทยได้สะดวกมากยิ่งขึ้น โดยระบบสามารถช่วยเพิ่มพูนความพึงพอใจและกระตุ้นนักท่องเที่ยวให้สนใจได้ ในขณะเดียวกันยังช่วยกระจายนักท่องเที่ยวไปยังเมืองรอง และแหล่งท่องเที่ยวใหม่ ๆ ได้อีกด้วย



7.2 การวิเคราะห์สื่อสังคมออนไลน์และพัฒนาเครื่องมือเพื่อรวบรวมข้อมูลสำรวจความคิดเห็นของภาครัฐ : กรณีศึกษาเน็ตประชารัฐ

สืบเนื่องจากที่สำนักงานสถิติแห่งชาติ ภายใต้กระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม ได้เริ่มศึกษาวิธีการ เครื่องมือ หรือแพลตฟอร์มสำหรับสำรวจความคิดเห็นและความพึงพอใจต่อนโยบายต่าง ๆ ของรัฐบาล แทนรูปแบบการสำรวจเดิมที่ต้องอาศัยการลงพื้นที่พบประชาชนกลุ่มตัวอย่างโดยตรง ซึ่งจำเป็นต้องใช้กำลังคน เวลา และงบประมาณค่อนข้างสูง ในการนี้ GBDi จึงได้มีส่วนร่วมศึกษาและสร้าง/เลือกใช้เครื่องมือสำหรับสำรวจข้อมูลความคิดเห็นผ่านการใช้เทคโนโลยีดิจิทัลและพัฒนากระบวนการสำรวจความพึงพอใจให้มีประสิทธิภาพมากยิ่งขึ้น

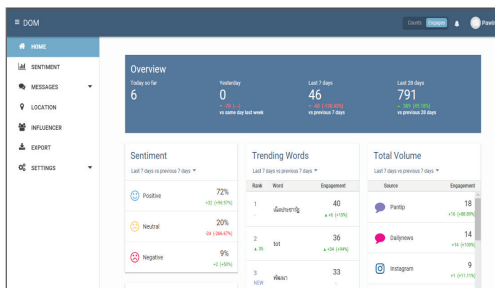
GBDi ได้พัฒนาแพลตฟอร์มเพื่อสำรวจความคิดเห็นและความพึงพอใจของประชาชนต่อโครงการเน็ตประชารัฐภายใต้การดูแลของกระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม โดยเลือกใช้ข้อมูลที่มีอยู่แล้วบนโลกดิจิทัล ได้แก่ ข้อมูลความพึงพอใจหรือความคิดเห็นในรูปแบบภาษาเขียนที่ปรากฏบนสื่อสังคมออนไลน์ และข้อมูลการใช้งานเน็ตประชารัฐจากเครื่องแม่ข่าย (Server Log) ผู้ให้บริการเน็ตประชารัฐ ร่วมกับการสำรวจข้อมูลผ่านแบบสอบถามออนไลน์โดยใช้เทคโนโลยีการตลาดแบบดิจิทัลในการคัดเลือกกลุ่มสำรวจเพื่อให้ได้ข้อมูลที่ครบถ้วนแม่นยำในประเด็นที่ไม่ได้ถูกพูดถึงในสื่อสังคมออนไลน์ โดยมีกรอบกระบวนการและขอบเขตของแหล่งข้อมูลดังแสดงในภาพที่ 7.3



ภาพที่ 7.3 โครงสร้างแพลตฟอร์มสำหรับการสำรวจความคิดเห็นของประชาชนแบบดิจิทัล

จากการศึกษาลักษณะของแบบสอบถามสำหรับการเดินสำรวจเดิมพบว่าแบบสอบถามดังกล่าวถูกออกแบบขึ้นเพื่อหาคำตอบใน 5 ประเด็น ได้แก่ 1) ข้อมูลลักษณะทางประชากรของผู้ใช้งานเน็ตประชารัฐ 2) ประสิทธิภาพการใช้งานและปัญหาที่พบในการใช้งาน 3) วัตถุประสงค์การใช้เน็ตประชารัฐของประชาชน 4) บริการอื่น ๆ ที่ประชาชนอยากให้ผ่านทางเน็ตประชารัฐ เช่น การฝึกอบรมเกี่ยวกับใช้เน็ตสำหรับการค้าขาย และ 5) ค่าบริการที่เหมาะสมสำหรับการต่อเน็ตประชารัฐเข้าบ้านโดยตรง

การใช้ข้อมูลจากแหล่งที่มีอยู่แล้วในโลกดิจิทัลมีศักยภาพเพียงพอที่จะสามารถนำมาใช้ตอบคำถามในบางประเด็นที่กล่าวมา ดังนี้ 1) เชื่อมโยงข้อมูลเลขประจำตัวประชาชนที่ลงทะเบียนใช้งานเน็ตประชารัฐเข้ากับระบบสารสนเทศของสำนักงานทะเบียนราษฎร กรมการปกครอง เพื่อให้ทราบข้อมูลลักษณะทางประชากรของผู้ใช้ 2) และ 3) ประมวลผลข้อมูลจากบันทึกการใช้งานจากเครื่องแม่ข่ายผู้ให้บริการและข้อมูลจากสื่อสังคมออนไลน์ในแง่ของความเห็นและความพึงพอใจอันประกอบด้วยความเห็นทางด้าน ความเร็ว ความเสถียร วัตถุประสงค์การใช้งาน และความเหมาะสมของจุดติดตั้ง Wi-Fi ฟรี ซึ่งในงานศึกษาชิ้นนี้ได้ใช้โปรแกรม Web Crawler เก็บข้อมูล 2 ปีย้อนหลังจากเว็บไซต์และสื่อสังคมออนไลน์ทั้งสิ้น 74 แห่ง แล้วนำมาแสดงผลในรูปแบบของ Dashboard ที่สามารถอัปเดตได้โดยอัตโนมัติดังตัวอย่างที่แสดงในภาพที่ 7.4



ภาพ Dashboard สรุปการวิเคราะห์อารมณ์/ความรู้สึก (Sentimental Analysis) ของข้อความจากภาษาที่เขียน

Last Time	Source IP	Destination	Host	URL	Category	User Agent
3/12/2018 18:05	10.150.7.2	157.240.11	443	edge-mqt	https://ec	Social Networking
3/12/2018 18:05	10.150.17.104	65.195	443	104.65.195	https://10	Content Server
3/12/2018 18:05	10.150.24.161	117.71	80	mon.byte	https://mo	Content S Dalvik/2.1.0 (Linux; U
3/12/2018 18:05	10.150.25.157	240.11	443	graph.facebook	https://gr	Social Networking
3/12/2018 18:05	10.150.25.179	60.194	443	179.60.194	https://17	Social Networking
3/12/2018 18:05	10.150.166.152	195.61	443	static.exo	https://st	Web Ads
3/12/2018 18:05	10.150.6.54	64.77.1	80	log16.byte	https://log	Content S Dalvik/2.1.0 (Linux; U
3/12/2018 18:05	10.150.25.103	21.25	443	lan.line.m	https://la	Messaging
3/12/2018 18:05	10.150.55.103	21.25	443	i.ytimg.co	https://i.y	Content Server
3/12/2018 18:05	10.150.2.1	23.56.157	443	obs.line-s	https://ot	Parked Domain
3/12/2018 18:05	10.150.4.1	173.194.2	443	r5---sn-np	https://r5	Streaming Media, Media Sharing
3/12/2018 18:05	10.150.166.172	217.31	443	googleads	https://gc	Web Ads
3/12/2018 18:05	10.150.8.5	157.240.11	443	edge-mqt	https://ec	Social Networking
3/12/2018 18:05	10.150.7.6	216.58.20	443	www.google	https://www	Internet Services

ภาพตัวอย่างบันทึกการใช้งานจากเครื่องแม่ข่ายผู้ให้บริการ

ภาพที่ 7.4 ตัวอย่าง Dashboard และ ไฟล์บันทึกการใช้งาน

ทั้งนี้ในส่วนของข้อมูลบันทึกการใช้งานจากเครื่องแม่ข่าย ภาพที่ 7.4 บ่งบอกถึงประเภทของกิจกรรมและวัตถุประสงค์การใช้งานได้อย่างชัดเจน โดยข้อมูลที่บันทึกนี้เป็นข้อมูลที่เกิดจากการใช้งานจริง และถูกจัดกลุ่มวิเคราะห์การใช้งานได้ในหลากหลายมิติ เช่น ช่วงเวลาที่เข้าใช้ หรือพื้นที่ที่เข้าใช้งาน เป็นต้น

อย่างไรก็ดีในประเด็นเรื่องบริการอื่น ๆ ที่ประชาชนอยากให้มี และค่าบริการที่เหมาะสม นั้นเป็นส่วนที่ไม่สามารถพบคำตอบได้จากทั้งในสื่อสังคมออนไลน์และบันทึกการใช้งานจากระบบแม่ข่าย การตอบคำถามในประเด็นเหล่านี้จึงยังจำเป็นต้องอาศัยการเก็บข้อมูลในรูปแบบของแบบสอบถาม ซึ่งอาจทำได้ในรูปแบบของแบบสอบถามออนไลน์ โดยสามารถออกแบบให้มีแบบสอบถามสั้น ๆ สำหรับผู้ใช้อีก่อนการเข้าใช้งานในทุกครั้งที่มีการเชื่อมต่ออินเทอร์เน็ตผ่านเน็ตประชารัฐ การศึกษานี้ได้ออกแบบให้มีการสุ่มข้อความ 4-5 คำถามจาก 20 คำถามที่ต้องการทราบข้อมูล และใช้เทคโนโลยีปัญญาประดิษฐ์เข้ามาช่วยคัดเลือกคำถามที่เหมาะสมกับผู้ตอบ

เพื่อตรวจสอบคุณภาพของข้อมูลที่ได้จากแพลตฟอร์มที่ออกแบบขึ้นนี้ คณะทำงานได้นำผลที่ได้จากการเก็บข้อมูลผ่านแพลตฟอร์มดังกล่าวมาเปรียบเทียบกับผลที่ได้จากผลของแบบสำรวจข้อมูลเดิม พบว่าข้อมูลที่ได้จากแพลตฟอร์มและจากแบบสอบถามเดิมนั้นมีแนวโน้ม (Correlation) ไปในทิศทางเดียวกัน

เพื่อขยายผลให้ต้นแบบเบื้องต้นสำหรับแพลตฟอร์มสามารถใช้เก็บข้อมูลในหัวข้ออื่นได้ คณะทำงานจึงออกแบบแพลตฟอร์มดังกล่าวให้สามารถเชื่อมโยงข้อมูลจาก 4 แหล่ง ได้แก่ ข้อมูลลักษณะของประชากรจากกรมการปกครอง สื่อสังคมออนไลน์ บันทึกการใช้งานจากเครื่องแม่ข่าย และแบบสอบถามออนไลน์ โดยแสดงผลรวมในรูปแบบ Dashboard นอกจากนี้ยังสามารถสร้างแบบสอบถามออนไลน์ใหม่บนแพลตฟอร์มนี้ได้โดยตรงอีกด้วย

7.3 การวิเคราะห์กระบวนการงานภาครัฐเพื่อนำไปสู่การพัฒนาศูนย์บริการเบ็ดเสร็จ (One-Stop Service Government Services Analysis)

เพื่อพัฒนาบริการภาครัฐให้มีประสิทธิภาพและสะดวกสำหรับประชาชนและภาคธุรกิจ หน่วยงานภาครัฐจำเป็นต้องการพัฒนากระบวนการและบริการสาธารณะให้อยู่ในรูปแบบอิเล็กทรอนิกส์รวมทั้งเชื่อมโยงข้อมูลระหว่างหน่วยงานเพื่อให้ประชาชนไม่จำเป็นต้องใช้สำเนาเอกสารจากหน่วยงานหนึ่งเพื่อจะเข้าไปรับบริการในอีกหน่วยงานหนึ่ง ทั้งนี้สำนักงานคณะกรรมการพัฒนาระบบราชการ (ก.พ.ร.) จึงได้มีนโยบายส่งเสริมให้เกิดการพัฒนาศูนย์บริการเบ็ดเสร็จ (One-Stop Service: OSS) สำหรับแต่ละกระบวนการ อย่างไรก็ตาม ด้วยข้อจำกัดทางทรัพยากรทั้งในด้านบุคลากรและด้านจำนวนบริการหรือกระบวนการภาครัฐที่มากกว่า 4,000 บริการ จึงเป็นเรื่องยากที่จะพัฒนาระบบกลางศูนย์บริการเบ็ดเสร็จสำหรับแต่ละกระบวนการทั้งหมด ด้วยเหตุนี้ GBDi จึงศึกษาความจำเป็นเร่งด่วนรวมถึงความพร้อมของแต่ละกระบวนการเพื่อจัดลำดับความสำคัญและความเป็นไปได้สำหรับตัดสินใจคัดเลือกกระบวนการที่สามารถพัฒนาเป็นศูนย์บริการเบ็ดเสร็จเป็นลำดับต้น ๆ

งานศึกษาชิ้นนี้ได้รับความร่วมมือด้านข้อมูลกระบวนการของภาครัฐในด้านต่าง ๆ เช่น สถิติการใช้งาน ระยะเวลาการดำเนินงานของกระบวนการ เอกสารที่เกี่ยวข้องกับกระบวนการ หน่วยงานที่ให้บริการกระบวนการ จากผลสำรวจของ สพร. ซึ่งคณะทำงานได้ดำเนินการรวบรวมข้อมูลจากหน่วยงานที่เกี่ยวข้องและจัดการข้อมูลเพื่อนำเสนอในรูปแบบของ Dashboard ดังแสดงในภาพที่ 7.5 โดยหน้าจอแสดงข้อมูลเพื่อประกอบการตัดสินใจดังกล่าว จะแบ่งออกเป็น 4 ส่วน ได้แก่

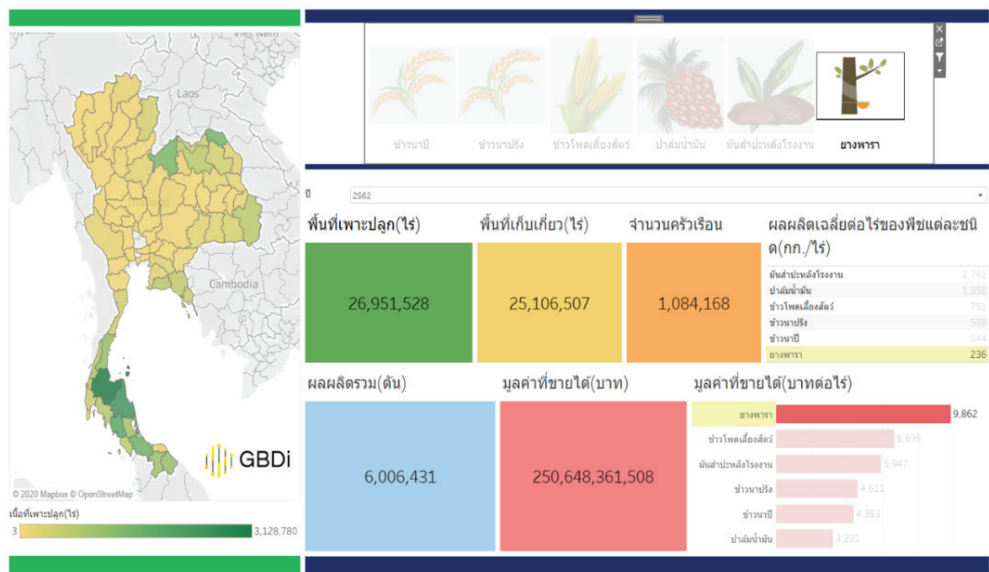
1. กระบวนการภายใต้ความรับผิดชอบของแต่ละหน่วยงาน สำหรับให้ผู้ใช้งานค้นหาข้อมูลกระบวนการของแต่ละหน่วยงาน เช่น จำนวนกระบวนการ รายการเอกสาร หน่วยงานเจ้าของเอกสารที่เกี่ยวข้องสำหรับบริการหรือกระบวนการนั้น ๆ เป็นต้น
2. สถิติการใช้งานของเอกสารต่าง ๆ เช่น บัตรประชาชน ทะเบียนบ้าน โฉนดที่ดิน เป็นต้น ซึ่งแสดงผลให้ผู้ใช้งานสามารถเห็นว่าเอกสารชิ้นที่สนใจนั้นเกี่ยวข้องกับกระบวนการและหน่วยงานใด รวมถึงสถานะความเป็นดิจิทัลและความพร้อมในการเชื่อมโยงของเอกสารชิ้นนั้น ๆ
3. จำนวนเอกสารที่แต่ละหน่วยงานเป็นเจ้าของ สำหรับใช้วางแผนเตรียมการติดต่อขอความร่วมมือเพื่อเชื่อมโยงข้อมูล

7.4 การพยากรณ์ปริมาณผลผลิตพืชเศรษฐกิจหลัก 5 ประเภทของประเทศไทย

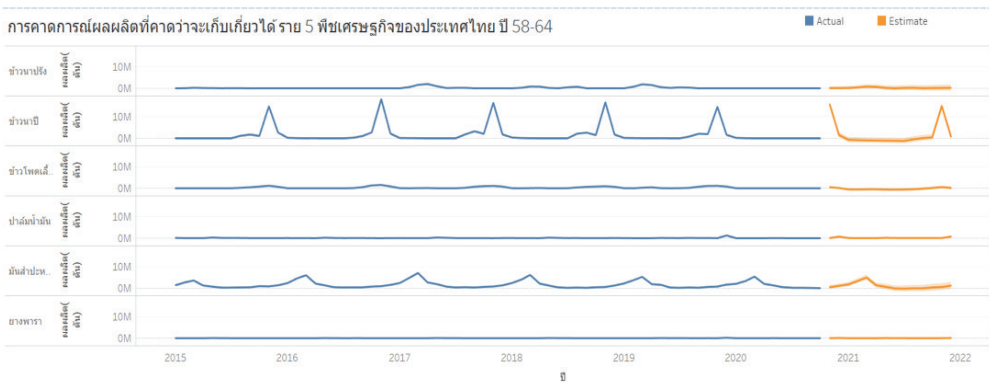
เพื่อส่งเสริมการวางแผนเศรษฐกิจด้านการเกษตรของประเทศ ความสามารถในการคาดการณ์ปริมาณผลผลิตของพืชผลทางการเกษตรที่จะถูกป้อนเข้าสู่ตลาดในแต่ละปีเป็นสิ่งสำคัญยิ่งต่อการวางแผนนโยบายเพื่อการบริหารจัดการคลังสินค้าการเกษตรให้มีประสิทธิภาพ รวมถึงการวางแผนนโยบายรณรงค์และให้ข้อมูลการปลูกพืชที่เหมาะสมแก่เกษตรกรด้วยเหตุนี้ GBDi ร่วมกับสำนักงานเศรษฐกิจการเกษตร กระทรวงเกษตรและสหกรณ์จึงได้นำข้อมูลปัจจัยทางการเกษตรต่าง ๆ ที่เก็บรวบรวมไว้โดยสำนักงานเศรษฐกิจการเกษตรมาใช้พยากรณ์ปริมาณผลผลิตของพืชเศรษฐกิจหลักของไทย 5 ประเภท ได้แก่ ข้าว ยางพารา ปาล์ม มันสำปะหลัง และอ้อย ผ่านเทคโนโลยีปัญญาประดิษฐ์ เพื่อพัฒนาปรับปรุงการพยากรณ์ปริมาณผลผลิตพืชเศรษฐกิจให้มีความแม่นยำมากยิ่งขึ้น

ในกรณีศึกษา GBDi ได้พัฒนาโมเดลคณิตศาสตร์สำหรับพยากรณ์ปริมาณพืชผลทางการเกษตรด้วยเทคนิค Random Forest โดยใช้ข้อมูลปัจจัยที่ส่งผลต่อปริมาณพืชผล 7 อย่าง ได้แก่ ช่วงเวลาเพาะปลูก ภูมิภาคที่ปลูก ปริมาณน้ำฝน ปริมาณน้ำเขื่อน อุณหภูมิ ความชื้นอากาศ และสายพันธุ์ที่เพาะปลูก (เฉพาะในโมเดลสำหรับข้าว) ของแต่ละเดือนในความละเอียดระดับภูมิภาค ซึ่งรวบรวมจากข้อมูลของกรมการข้าว สำนักงานเศรษฐกิจการเกษตร ร่วมกับข้อมูลอุตุนิยมวิทยาจากเครื่องตรวจวัดที่เผยแพร่ต่อสาธารณะ โดยนำมาเปรียบเทียบกับข้อมูลผลผลิตทางการเกษตรจากโครงการการพัฒนารฐานข้อมูลเกษตรกรกลาง (Farmer One) โดยกระทรวงเกษตรและสหกรณ์ ดังแสดงในภาพที่ 7.6 ซึ่งสามารถพยากรณ์ปริมาณผลผลิตของพืชเศรษฐกิจหลัก 5 ประเภทในแต่ละเดือนได้ดังแสดงในภาพที่ 7.7 เครื่องมือนี้น่าจะสามารถใช้เป็นส่วนหนึ่งในการวิเคราะห์เพื่อวางแผนส่งเสริมการเกษตรต่อไป





ภาพที่ 7.6 ผลผลิตพืชเศรษฐกิจหลัก 5 ประเภทของประเทศไทย จำแนกรายปี รายจังหวัด

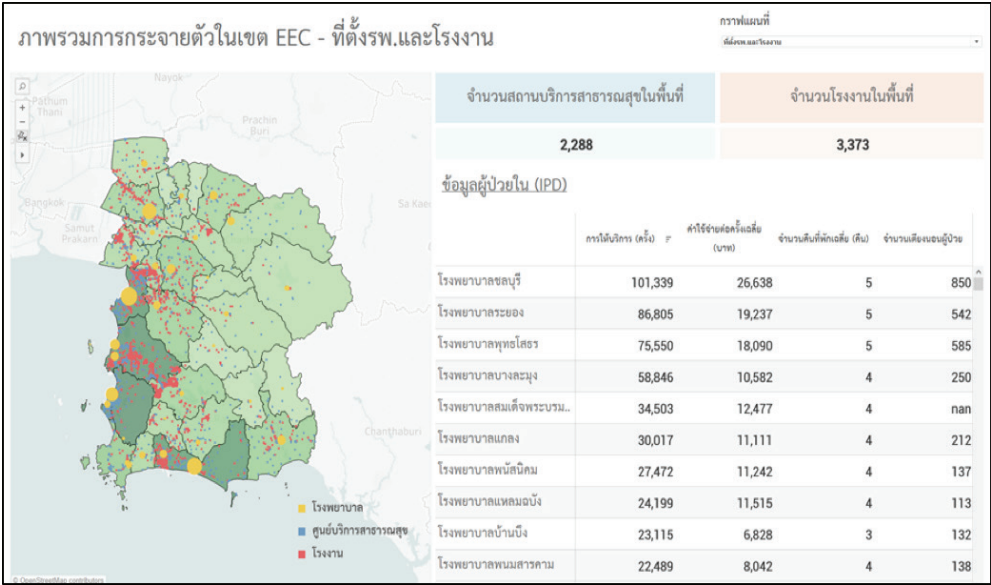


ภาพที่ 7.7 แนวโน้มผลผลิตพืชเศรษฐกิจหลัก 5 ประเภทของประเทศไทย

7.5 การเตรียมความพร้อมสำหรับการบริหารจัดการทรัพยากรทางด้านสาธารณสุข ในเขตพัฒนาพิเศษภาคตะวันออก (Eastern Economic Corridor : EEC)

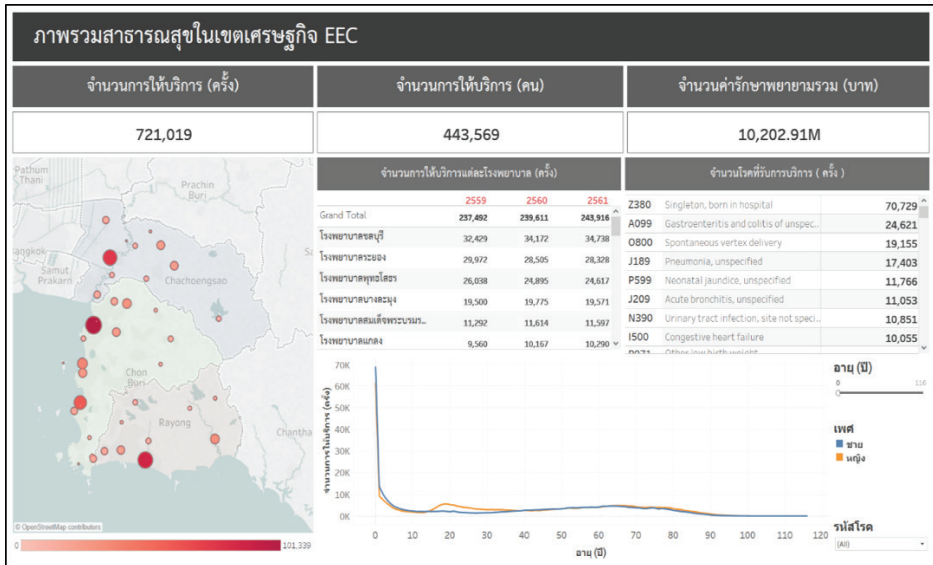
จากนโยบายส่งเสริมอุตสาหกรรมในเขตพัฒนาพิเศษภาคตะวันออก (EEC) เพื่อยกระดับพื้นที่เศรษฐกิจใน 3 จังหวัด ได้แก่ ระยอง ชลบุรี และฉะเชิงเทราให้เป็นเขตเศรษฐกิจชั้นนำของเอเชีย ซึ่งจะส่งผลต่อการเพิ่มขึ้นของจำนวนแรงงานในพื้นที่โดยรอบอย่างหลีกเลี่ยงไม่ได้ การวิเคราะห์ข้อมูลโรงงานอุตสาหกรรมร่วมกับข้อมูลด้านสาธารณสุขจึงเป็นสิ่งจำเป็นเพื่อเตรียมความพร้อมด้านสาธารณสุขให้เพียงพอต่อความต้องการของประชาชนในพื้นที่และพร้อมรับมือกับภาวะสุขอนามัยที่จะเปลี่ยนแปลงไปเนื่องจากการพัฒนาทางเศรษฐกิจด้วยเหตุนี้ GBDi จึงได้เข้าร่วมประชุมกับคณะกรรมการนโยบายเขตพัฒนาพิเศษภาคตะวันออกด้านสาธารณสุข โดยมี ศาสตราจารย์คลินิกเกียรติคุณ นพ.ปิยะสกล สกลสัตยาทร อดีตรัฐมนตรีว่าการกระทรวงสาธารณสุขเป็นประธานในที่ประชุม และร่วมงานกับสำนักงานหลักประกันสุขภาพแห่งชาติเพื่อรวบรวมข้อมูลจาก สำนักงานหลักประกันสุขภาพแห่งชาติ สำนักงานสถิติแห่งชาติ กรมอุตุนิยมวิทยา กรมควบคุมมลพิษ และสำนักงานพัฒนาเทคโนโลยีอวกาศและภูมิสารสนเทศ เพื่อศึกษาและพัฒนาเครื่องมือสำหรับการความพร้อมทางด้านสาธารณสุขในเขต EEC

การศึกษานี้ได้เก็บรวบรวมข้อมูลและแสดงผลในรูปแบบของ Dashboard สำหรับประกอบการวิเคราะห์เพื่อศึกษาการกระจายตัวของโรงพยาบาล ร่วมกับโรงงานและจำนวนครัวเรือนในเขตพัฒนาพิเศษตะวันออกทั้ง 3 จังหวัดดังกล่าว ดังภาพที่ 7.8 ซึ่งจะช่วยให้สามารถประเมินความสอดคล้องของจำนวนประชากรและความสามารถในการรองรับผู้รับการรักษาพยาบาลของโรงพยาบาลในแต่ละพื้นที่ อันจะนำไปสู่การยกระดับการบริการสาธารณสุขในพื้นที่ เช่น ลดความแออัดของจำนวนผู้รับการรักษาพยาบาลในแต่ละโรงพยาบาล จัดการเวลารักษาสำหรับผู้รับการรักษาพยาบาลที่ถูกส่งต่อ/นัดหมาย/ฉุกเฉินให้เหมาะสม และเพิ่มคุณภาพบริการสุขภาพปฐมภูมิ ผ่านการจัดสรรทรัพยากรด้านบริการสุขภาพให้เหมาะสมตามความต้องการ

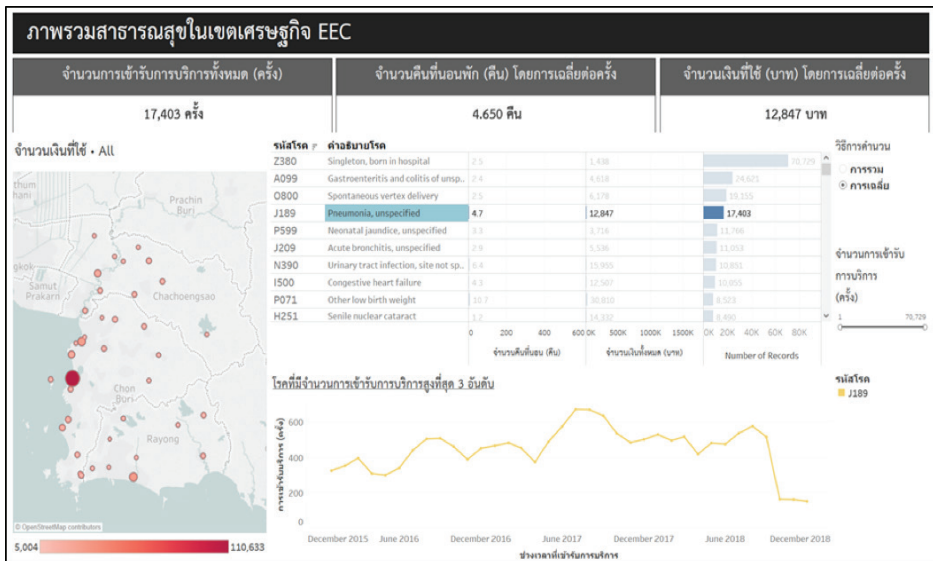


ภาพที่ 7.8 ภาพรวมการกระจายตัวของโรงพยาบาลและโรงงานในเขต EEC

จากข้อมูลการให้บริการของแต่ละโรงพยาบาลในเขต EEC ช่วงปี พ.ศ. 2559-2561 พบว่าการเข้ารับบริการหลักของประชาชนคือการคลอดบุตร ซึ่งสามารถสังเกตได้จากปริมาณผู้ใช้บริการในช่วงอายุ 0-1 ปีที่มีจำนวนมากเป็นพิเศษดังแสดงให้เห็นในภาพที่ 7.9 ลำดับถัดมา คือ โรคเกี่ยวกับทางเดินอาหาร และตามด้วยโรคทางเดินหายใจ การนำเสนอข้อมูลในรูปแบบของ Dashboard เช่นนี้จะช่วยให้สามารถตรวจสอบรายละเอียดอายุของผู้รับการรักษาพยาบาล โรงพยาบาลที่ผู้รับการรักษาพยาบาลนิยมเข้าทำการรักษา จำนวนคืนเฉลี่ยสำหรับการเข้าพัก จำนวนเงินที่ใช้ในการรักษา และจำนวนครั้งในการเข้ารับบริการสำหรับแต่ละกลุ่มโรคได้ ดังแสดงให้เห็นในภาพที่ 7.10



ภาพที่ 7.9 ภาพรวมด้านสาธารณสุข (โรงพยาบาล) ในเขต EEC



ภาพที่ 7.10 ภาพรวมด้านสาธารณสุข (โรค) ในเขต EEC

7.6 การวิเคราะห์ข้อมูล จำนวน และประเภทครอบครัวในประเทศ

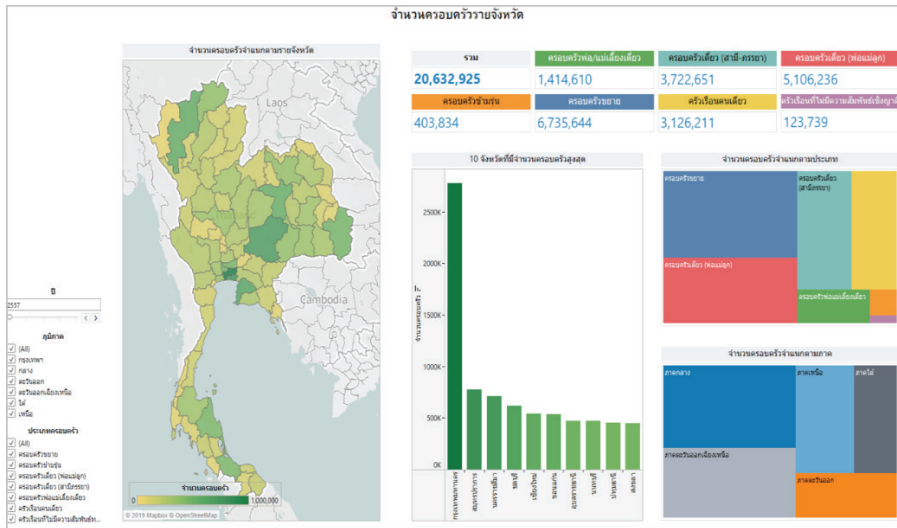
สืบเนื่องจากสภาพการเปลี่ยนแปลงของโครงสร้างครอบครัว และสัมพันธภาพของครอบครัวในปัจจุบัน กรมกิจการสตรีและสถาบันครอบครัว กระทรวงพัฒนาสังคมและความมั่นคงของมนุษย์ได้ดำเนินการทบทวนนิยามและประเภทครอบครัว¹⁰ เพื่อปรับปรุงความรู้ความเข้าใจในด้านรูปแบบและโครงสร้าง รวมถึงลักษณะพิเศษของครอบครัวในปัจจุบัน อันเป็นจุดเริ่มต้นที่สำคัญของการดำเนินงานโครงการคลังข้อมูลสารสนเทศด้านครอบครัว (Family Big Data) โดยมีเป้าหมายที่จะวิเคราะห์จำนวนครอบครัวแต่ละประเภทเป็นลำดับถัดไป

เพื่อต่อยอดการดำเนินงานให้สามารถเข้าถึงเป้าหมายการพัฒนาความมั่นคงของครอบครัว การมองเห็นภาพรวมของการกระจายตัวของครอบครัวในแต่ละประเภทตามนิยามใหม่จึงเป็นสิ่งจำเป็นเพื่อช่วยให้ผู้ที่มีอำนาจตัดสินใจสามารถตัดสินใจงานนโยบายที่เหมาะสมด้วยเหตุนี้ GBDi จึงได้จัดทำ Dashboard แสดงผลข้อมูลจำนวนครอบครัวตามประเภทของครอบครัวที่สอดคล้องกับความสนใจตามนิยามของกรมกิจการสตรีและสถาบันครอบครัว โดยใช้ข้อมูลผลสำรวจภาวะการทำงานของประชากรไตรมาสที่ 3 (กรกฎาคม-กันยายน) โดยสำนักงานสถิติแห่งชาติ ซึ่งแม้จะขาดข้อมูลสำหรับบางประเภทครอบครัวที่เป็นที่สนใจของกรมกิจการสตรีและสถาบันครอบครัว ได้แก่ ครอบครัวผู้สูงอายุอยู่ด้วยกันตามลำพัง และครอบครัวคู่รักเพศเดียวกัน แต่ก็มีอีกหลายประเภทครอบครัวที่สามารถใช้เทียบเคียงกับประเภทครอบครัวตามนิยามของกรมกิจการสตรีและสถาบันครอบครัวได้

งานศึกษานี้ได้นำเสนอข้อมูลจำนวนครอบครัวในรายจังหวัดของประเทศไทยย้อนหลัง 5 ปี (พ.ศ.2557-2561) สำหรับครอบครัวประเภทต่าง ๆ อ้างอิงตามนิยามของสำนักงานสถิติแห่งชาติ โดยจากข้อมูลของสำนักงานสถิติแห่งชาติสามารถแบ่งการนำเสนอออกเป็น 3 ส่วน ได้แก่

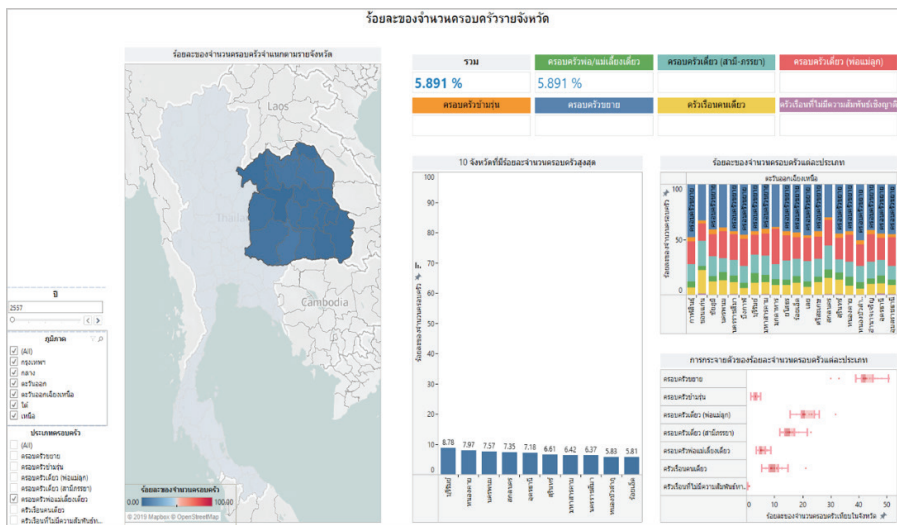
1) จำนวนครอบครัวสุทธิ ดังภาพที่ 7.11 สำหรับใช้เปรียบเทียบจำนวนครอบครัวแต่ละประเภทในแต่ละจังหวัด

¹⁰ กรมกิจการสตรีและสถาบันครอบครัว, การทบทวนนิยามและประเภทครอบครัว



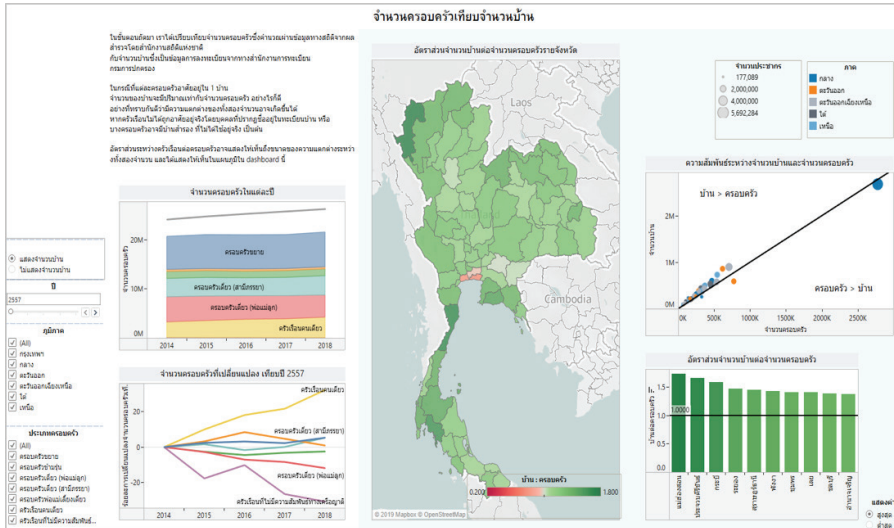
ภาพที่ 7.11 Dashboard แสดงข้อมูลจำนวนครอบครัวรายจังหวัด

2) ร้อยละของจำนวนครอบครัว ดังภาพที่ 7.12 สำหรับใช้เปรียบเทียบสัดส่วนของครอบครัวประเภทต่าง ๆ ในแต่ละจังหวัด



ภาพที่ 7.12 Dashboard แสดงร้อยละของจำนวนครอบครัวรายจังหวัด
ในภาพเลือกให้แสดงผลเฉพาะข้อมูลครอบครัวพ่อแม่เลี้ยงเดี่ยว
ในพื้นที่ภาคตะวันออกเฉียงเหนือ

3) การเปลี่ยนแปลงของจำนวนครอบครัวประเภทต่าง ๆ ในปี พ.ศ.2557-2561 ดังแสดงในภาพที่ 7.13 ซ้าย



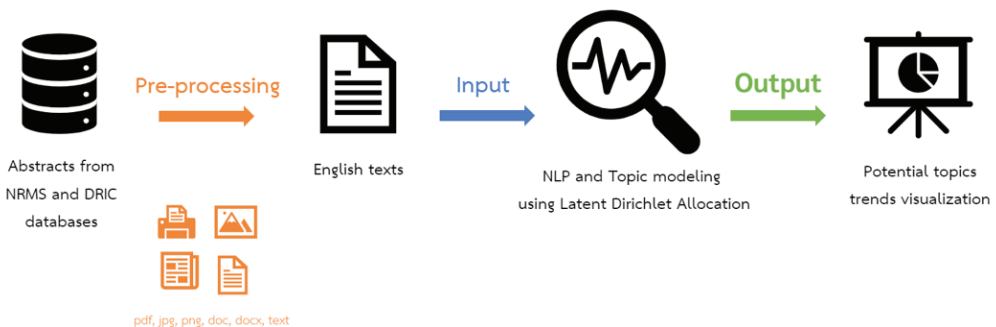
ภาพที่ 7.13 Dashboard แสดง (ซ้าย) ข้อมูลจำนวนครอบครัวรายปี และ (ขวา) ข้อมูลจำนวนครอบครัวเทียบกับข้อมูลจำนวนบ้านที่จดทะเบียน

นอกจากนี้เมื่อนำข้อมูลจากสำนักงานสถิติแห่งชาติ มาประกอบกับข้อมูลจำนวนบ้านที่จดทะเบียน ซึ่งเป็นข้อมูลสาธารณะภายใต้การดูแลของสำนักงานบริหารการทะเบียน กรมการปกครอง กระทรวงมหาดไทย ทำให้สามารถสร้างมิติการเปรียบเทียบลักษณะข้อมูลจากสองแหล่งซึ่งมีความแตกต่างในจุดประสงค์และวิธีการเก็บรวบรวม ดังแสดงในภาพที่ 7.13 ขวา แสดงการเปรียบเทียบข้อมูลจากสองส่วน คือ จำนวนครอบครัว และจำนวนบ้านที่จดทะเบียน ซึ่งความแตกต่างของสองจำนวนข้างต้นอาจสะท้อนถึงการมีบ้านสำรองนอกเหนือจากบ้านหลักของครอบครัวหนึ่ง ๆ หรือการอาศัยร่วมบ้านกันของครอบครัวมากกว่าหนึ่งครอบครัว เป็นต้น

7.7 การสกัดหัวข้องานเขียนเชิงวิชาการเพื่อศึกษาแนวโน้มหัวข้องานวิจัยไทย

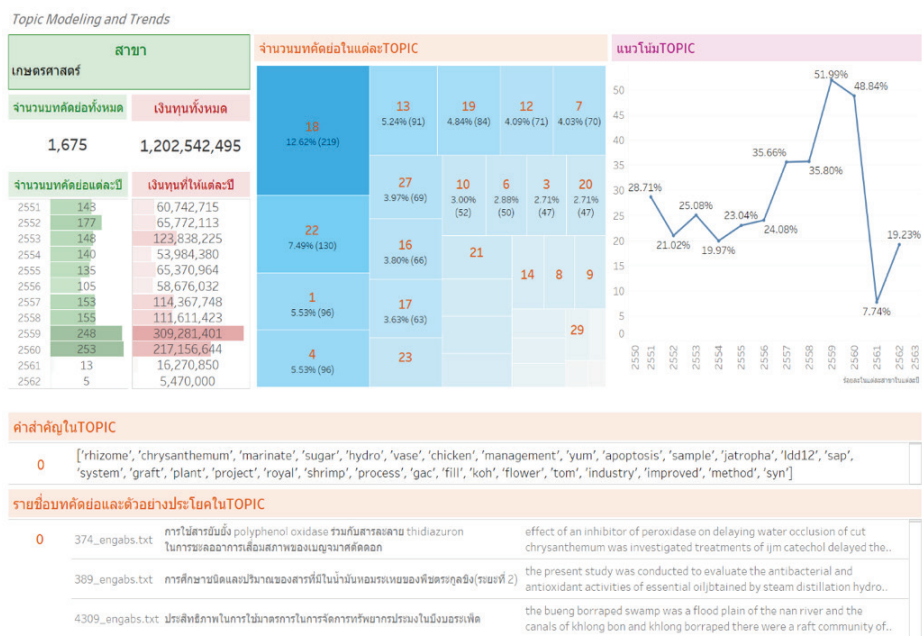
ด้วยภาระหน้าที่ในการให้ทุนวิจัยและนวัตกรรม การจัดทำฐานข้อมูลและดัชนีด้านวิทยาศาสตร์ วิจัยและนวัตกรรมของประเทศ รวมถึงทำการริเริ่มขับเคลื่อนและประสานการดำเนินงานโครงการวิจัยและนวัตกรรมที่สำคัญของประเทศ สำนักงานการวิจัยแห่งชาติ (วช.) จึงได้มีความพยายามในการจำแนกหมวดหมู่ประเภทงานวิจัยในฐานข้อมูล วิเคราะห์สัดส่วน และกลุ่มหัวข้อเรื่องงานวิจัยในประเทศในช่วงเวลาที่ผ่านมา ซึ่งเป็นสิ่งสำคัญที่จะช่วยให้ทราบถึงแนวโน้มของงานวิจัยที่เป็นมาในอดีตจนถึงปัจจุบัน สำหรับใช้วางแผนการให้ทุนการวิจัยในปีถัดไป โดยผ่านการกระตุ้นให้เกิดการวิจัยในด้านที่สำคัญหรือขาดแคลนของประเทศและเพิ่มความกวดขันในการพิจารณางานวิจัยในสาขาที่ได้รับความนิยม

ด้วยธรรมชาติของงานวิชาการซึ่งเกิดการศึกษาในศาสตร์ใหม่ที่หลากหลายอยู่ตลอดเวลา การจัดจำแนกประเภทงานวิจัยให้เหมาะสมนั้น นอกจากจำเป็นต้องอาศัยผู้มีความรู้ความเข้าใจในศาสตร์หลากหลายแขนงแล้วยังจำเป็นต้องมีการพัฒนาปรับปรุงความรู้ให้ทันสมัยอยู่เสมอ ข้อจำกัดนี้อาจถูกแก้ไขได้ด้วยการจัดจำแนกประเภท ผ่านการสกัดหัวข้อสำหรับแต่ละงานวิจัยโดยอัตโนมัติด้วยการใช้ประมวลผลภาษาธรรมชาติ (Natural Language Processing: NLP) กับข้อมูลบทคัดย่อของแต่ละงานวิจัย ซึ่งปัจจุบัน GBDI ร่วมมือกับ วช. ในการออกแบบเพื่อปรับวิธีการเก็บบทคัดย่อของโครงการวิจัยและงานวิจัยในประเทศไทยให้มีความครบถ้วนสมบูรณ์ สามารถใช้ระบบปัญญาประดิษฐ์ (AI) วิเคราะห์เนื้อหาพื้นฐานได้โดยมีการกำหนดรูปแบบของไฟล์และรูปแบบของเนื้อหาอย่างชัดเจนเพื่อความสะดวกในการประมวลผล



ภาพที่ 7.14 ขั้นตอนกระบวนการวิเคราะห์หัวข้องานวิจัยจากบทคัดย่อภาษาอังกฤษ

ทั้งนี้ ระบบนำร่องสำหรับการวิเคราะห์จัดกลุ่มหัวข้อวิจัยและสรุปแนวโน้มงานวิจัยในหัวข้อต่าง ๆ ได้ถูกออกแบบและพัฒนาขึ้นในช่วงปีที่ผ่านมา โดยเริ่มจากการวิเคราะห์บทความภาษาอังกฤษที่ได้รับการเก็บรวบรวมไว้ในระบบภายในของสำนักงานการวิจัยแห่งชาติ ภาพที่ 7.14 แสดงขั้นตอนการศึกษาและวิเคราะห์ข้อมูลบทความดังกล่าว ซึ่งต้องผ่านกระบวนการจัดเตรียมข้อมูลบทความเพื่อให้ได้ข้อมูลข้อความภาษาอังกฤษ ก่อนจะนำข้อความที่ได้ไปวิเคราะห์ด้วยโมเดลทางคณิตศาสตร์และการประมวลผลภาษาธรรมชาติ ซึ่งเป็นระบบ AI ประเภทหนึ่ง เพื่อสกัดหัวข้อที่เกี่ยวข้องกับบทความนั้น ๆ และนำผลลัพธ์ที่ได้จากการประมวลผลมานำเสนอบน Dashboard ดังภาพที่ 7.15 เพื่อแสดงให้เห็นภาพรวมแนวโน้มหัวข้องานวิจัยในอดีตที่ผ่านมา



ภาพที่ 7.15 Dashboard ผลลัพธ์การสกัดหัวข้องานวิจัยจากบทความภาษาอังกฤษ แสดงแนวโน้มและปริมาณงานวิจัยในหัวข้อต่าง ๆ

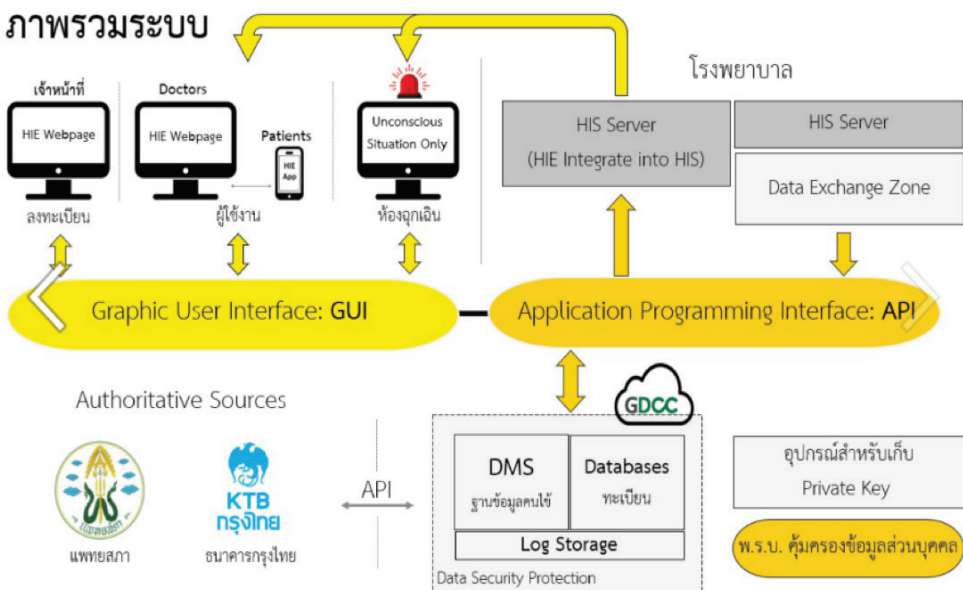
7.8 ระบบสารสนเทศเพื่อการแลกเปลี่ยนข้อมูลสุขภาพ

เพื่ออำนวยความสะดวกให้ประชาชนสามารถเข้าถึงประวัติการรักษาพยาบาลของตนเองได้อย่างรวดเร็ว กระทรวงสาธารณสุขร่วมกับกระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคมจึงได้ร่วมมือกันพัฒนาการเชื่อมโยงข้อมูลขนาดใหญ่ด้านสาธารณสุข และมอบหมายให้บริษัท กสท โทรคมนาคม จำกัด (มหาชน) ร่วมกับ GBDi ดำเนินการศึกษา ออกแบบ พัฒนา และให้บริการระบบสารสนเทศเพื่อการแลกเปลี่ยนข้อมูลสุขภาพส่วนบุคคลระหว่างโรงพยาบาล โดยปัจจุบันได้รับความสนับสนุนจากภาคีด้านสุขภาพทั้งจากโรงพยาบาลรัฐบาลภายใต้กระทรวงสาธารณสุข โรงพยาบาลโรงเรียนแพทย์ โรงพยาบาลตำรวจ และโรงพยาบาลเอกชน จำนวน 17 เครือข่ายโรงพยาบาลในเบื้องต้น

ข้อมูลผู้รับการรักษาพยาบาลและประวัติการรักษาเป็นข้อมูลสำคัญเพื่อใช้ประกอบการวินิจฉัยโรคและให้การรักษา การเชื่อมโยงข้อมูลสุขภาพส่วนบุคคลระหว่างโรงพยาบาลที่ผู้รับการรักษาพยาบาลเข้ารับการรักษามานับระบบดิจิทัลจะเพิ่มความสะดวกและย่นระยะเวลาการใช้บริการด้านสุขภาพของผู้รับการรักษาพยาบาล โดยเฉพาะในกรณีของผู้รับการรักษาพยาบาลเข้ารับการรักษฉุกเฉิน ซึ่งการเข้าถึงประวัติผู้รับการรักษาพยาบาลอย่างสะดวกรวดเร็ว และแม่นยำ มีผลสำคัญที่จะช่วยให้ผู้รับการรักษาพยาบาลได้รับการรักษาที่ถูกต้อง

โครงการนี้มีการวางกรอบการสร้างกลไกเพื่อเชื่อมโยงข้อมูลผู้รับการรักษาพยาบาลผ่านเลขประจำตัวประชาชน 13 หลัก โดยประชาชนจะสามารถใช้บัตรประชาชนเพื่อยืนยันคำขอข้อมูลประวัติการรักษาของตนเองจากโรงพยาบาลที่เข้าร่วมในภาคีสุขภาพภายใต้โครงการนี้ โดยจะสร้างระบบนำร่องด้านสุขภาพที่จะเป็นแพลตฟอร์มกลางที่สามารถรองรับระบบนิเวศน์ในการเชื่อมโยงและจัดเก็บข้อมูล พิสัยส่วนตัว มีความปลอดภัยในการรับ-ส่งข้อมูล รองรับความหลากหลายของชนิดและปริมาณของข้อมูล ที่คาดว่าจะเติบโตอย่างก้าวกระโดด ได้อย่างแม่นยำ ครบถ้วน ถูกต้อง และเชื่อถือได้ จึงจัดเป็นโครงการปีถัดมาที่มุ่งเน้นการเชื่อมโยงข้อมูลระหว่างองค์กรด้านสุขภาพ ส่งเสริมให้ประชาชนตระหนักถึงสิทธิในการเข้าถึง และสามารถประโยชน์จากข้อมูลของตนเองได้อย่างเต็มที่

ภาพรวมของแพลตฟอร์ม Health Information Exchange หรือ HIE แสดงในภาพที่ 7.16 ออกแบบให้ข้อมูลในระบบ Hospital Information System (HIS) ใช้จัดการข้อมูลผู้รับการรักษาพยาบาลในแต่ละโรงพยาบาล มีการเชื่อมต่อ (ตามความยินยอมของผู้รับการรักษาพยาบาล) เข้ากับฐานข้อมูลกลางที่เก็บรักษาอยู่ที่ Government Data Center and Cloud Services เฉพาะส่วนข้อมูลการรักษาพื้นฐาน เช่น การแพทย์ ข้อมูลยาที่ใช้ ข้อมูลประวัติการรักษาพื้นฐาน ฯลฯ โดยผ่านส่วนต่อประสานโปรแกรมประยุกต์ (API) เมื่อเกิดเหตุการณ์ที่แพทย์จากโรงพยาบาลอื่นต้องการใช้ข้อมูลประวัติการรักษาที่จำเป็น แพทย์สามารถขอความยินยอมจากผู้รับการรักษาพยาบาลเพื่อเรียกดูข้อมูลที่เก็บรักษาที่ส่วนกลางผ่านส่วนต่อประสานกราฟิกกับผู้ใช้ (Graphical User Interface: GUI) และอาจมีสิทธิเข้าถึงข้อมูลจากส่วนกลางได้โดยไม่ต้องรอการยินยอมจากผู้รับการรักษาพยาบาลเฉพาะในกรณีเข้ารับการรักษาฉุกเฉิน (ผู้รับการรักษาพยาบาลหมดสติ) ทั้งนี้ผู้รับการรักษาพยาบาลต้องลงทะเบียนเพื่อเข้าร่วมโครงการและมีการพิสูจน์และยืนยันตัวตนผ่านระบบ E-KYC และ KYC บนแพลตฟอร์ม “เปาตัง” ของธนาคารกรุงไทยก่อน ส่วนแพทย์ผู้ขอข้อมูล (โดยผู้รับการรักษาพยาบาลยินยอม) จะผ่านการยืนยันตัวตนกับระบบทะเบียนแพทย์ของแพทยสภา



ภาพที่ 7.16 แสดงภาพรวมผังการเชื่อมโยงระบบสารสนเทศเพื่อการแลกเปลี่ยนข้อมูล

7.9 การวิเคราะห์ความสัมพันธ์ระหว่างลักษณะทางประชากรของผู้เข้าพักกับประเภทโรงแรมที่เข้าพัก

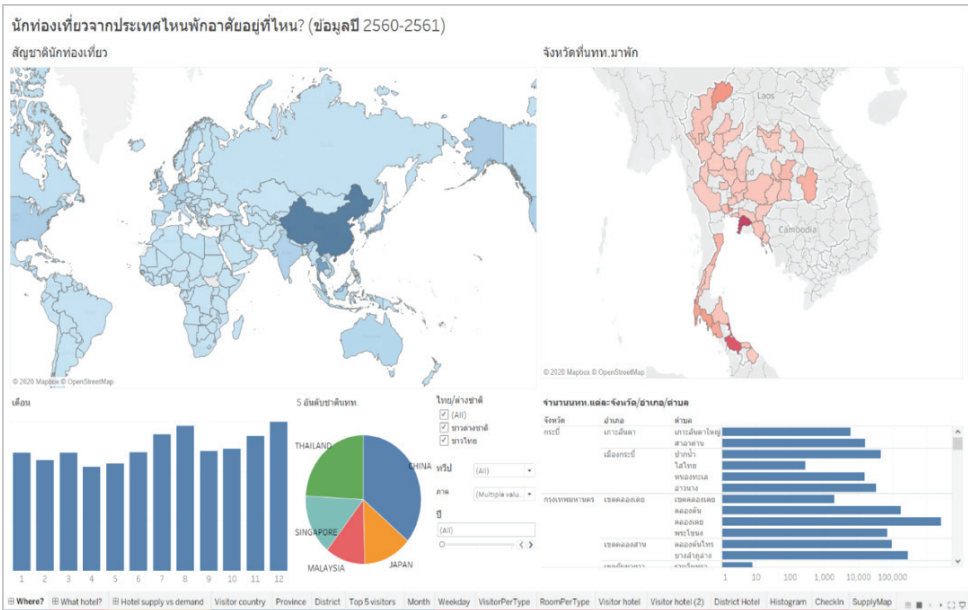
ข้อมูลด้านที่พักและการพักรวมเป็นส่วนหนึ่งที่มีความสำคัญต่อการวางแผนส่งเสริมอุตสาหกรรมท่องเที่ยว เพื่อตอบสนองนโยบายส่งเสริมการท่องเที่ยวโดยอ้างอิงจากข้อมูลรอบด้าน GBDi จึงได้ร่วมมือกับกรมการปกครองในการรวบรวมข้อมูลด้านที่พักและการพักรวม โดยข้อมูลผู้เข้าพักและแหล่งที่พักซึ่งเป็นที่พักที่ขึ้นทะเบียนกับกรมการปกครองมากกว่าหนึ่งหมื่นแห่งนั้น สามารถนำมาใช้วิเคราะห์เพื่อหาความสัมพันธ์ระหว่างลักษณะทางประชากรของผู้เข้าพักกับประเภทโรงแรมที่เข้าพัก สำหรับใช้ในการทำความเข้าใจลักษณะทางประชากรและวางแผนการประกอบการให้ตรงกับกลุ่มนักท่องเที่ยวได้

เพื่อตอบสนองความสนใจในด้านการส่งเสริมอุตสาหกรรมท่องเที่ยวของประเทศ GBDi ได้เลือกวิเคราะห์ทะเบียนพักรวมใน 3 มิติ ได้แก่

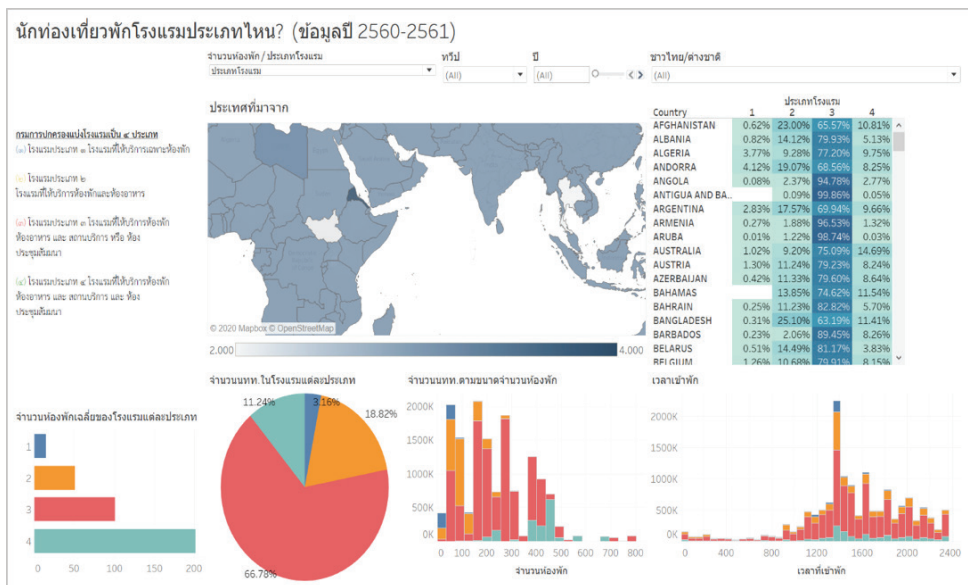
1) สถานที่พักที่นักท่องเที่ยวจากแต่ละประเทศนิยมเข้าพัก ดังภาพที่ 7.17 แสดงข้อมูลสถิติจำนวนนักท่องเที่ยวในแต่ละเดือนสำหรับแต่ละที่พัก จำนวนนักท่องเที่ยวในแต่ละจังหวัด อำเภอ ตำบล รวมถึงอันดับสัญชาติของนักท่องเที่ยวที่นิยมเข้าพักในแต่ละท้องถิ่น ทั้งนี้สามารถเลือกพิจารณาข้อมูลรายสัญชาติของนักท่องเที่ยวได้

2) ลักษณะโรงแรมที่นักท่องเที่ยวแต่ละสัญชาตินิยมพัก ดังภาพที่ 7.18 แสดงข้อมูลความนิยมของนักท่องเที่ยวรวมและจำแนกรายสัญชาติต่อแต่ละประเภทของโรงแรม และเวลาการเข้าพัก

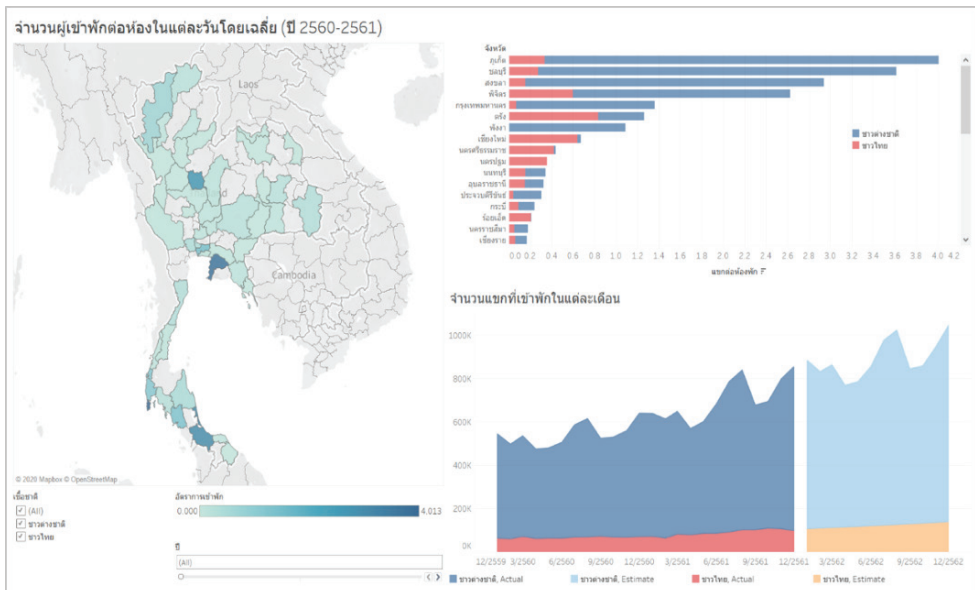
3) การเข้าพักของนักท่องเที่ยวในแต่ละเดือน แยกข้อมูลการเข้าพักของนักท่องเที่ยวไทยและนักท่องเที่ยวต่างชาติ ดังภาพที่ 7.19 แสดงรายละเอียดเกี่ยวกับอัตราส่วนระหว่างจำนวนผู้เข้าพักและห้องพักในแต่ละจังหวัด อัตราส่วนระหว่างจำนวนผู้เข้าพักและห้องพักในแต่ละจังหวัดโดยมีการแยกระหว่างนักท่องเที่ยวไทยและนักท่องเที่ยวต่างชาติ ประกอบกับสถิติจำนวนนักท่องเที่ยวไทยและนักท่องเที่ยวต่างชาติในแต่ละเดือน รวมถึงพยากรณ์นักท่องเที่ยวไทยและนักท่องเที่ยวต่างชาติในเดือนถัดไป



ภาพที่ 7.17 Dashboard แสดงการเข้าพักตามสัญชาติหรือประเทศต้นทาง
ของนักท่องเที่ยว



ภาพที่ 7.18 Dashboard แสดงลักษณะโรงแรมที่นักท่องเที่ยวแต่ละสัญชาตินิยมพัก



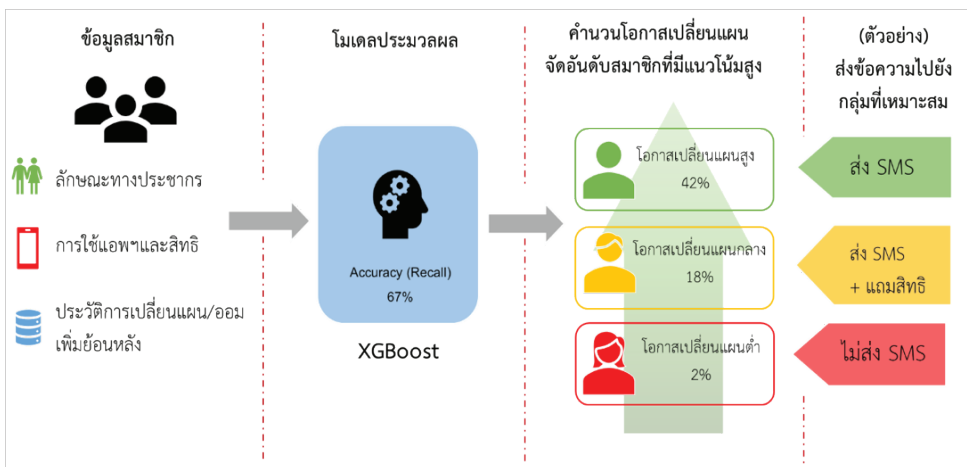
ภาพที่ 7.19 Dashboard แสดงการเข้าพักของนักท่องเที่ยวในแต่ละเดือน

7.10 การหากลุ่มสมาชิกที่มีแนวโน้มในการออมเพิ่มและเลือกแผนการลงทุนของกองทุนบำเหน็จบำนาญข้าราชการ

เพื่อส่งเสริมและสนับสนุนสมาชิกของกองทุนบำเหน็จบำนาญข้าราชการ (กบข.) ให้มีข้อมูลเพื่อเลือกการลงทุนที่เหมาะสม กองทุนบำเหน็จบำนาญข้าราชการจึงได้ร่วมมือกับ GBDi เพื่อพัฒนาต้นแบบโมเดลทางคณิตศาสตร์เชิงพยากรณ์ (Predictive Model) ด้วยเทคโนโลยีปัญญาประดิษฐ์ สำหรับใช้ประเมินแนวโน้มการเปลี่ยนเปอร์เซ็นต์และแผนการลงทุนของสมาชิกรายบุคคลและรายกลุ่ม เพื่อใช้ประกอบแผนการกระตุ้นการปรับเปลี่ยนเปอร์เซ็นต์การลงทุนและปรับเปลี่ยนแผนการลงทุนของสมาชิก โดยใช้ข้อมูลที่มีอยู่ เช่น ข้อมูลลักษณะทางประชากรของสมาชิก ข้อมูลประวัติการปรับเปลี่ยนเปอร์เซ็นต์และเปลี่ยนแผนการลงทุนของสมาชิก และ ข้อมูลประวัติการเข้าใช้บริการระบบ My GPF Application ในภาพที่ 7.20

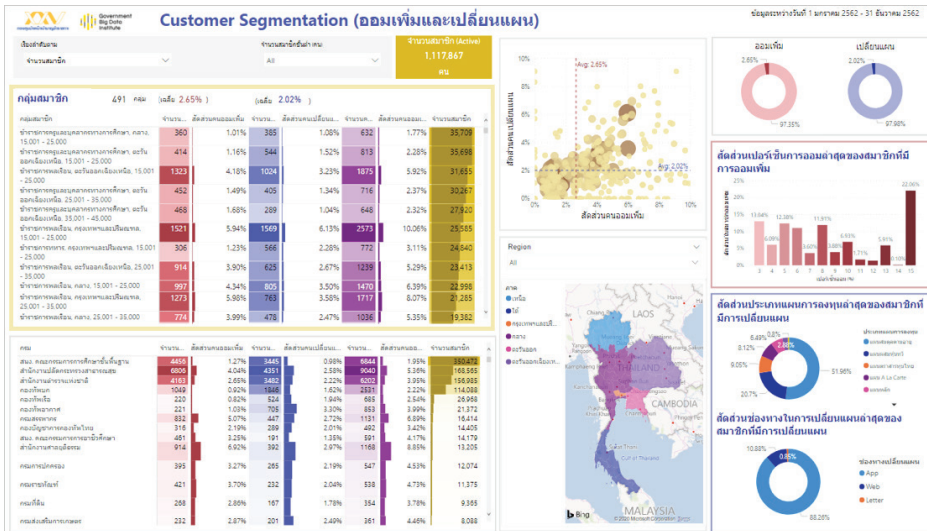
โมเดลทางคณิตศาสตร์ที่พัฒนาสามารถทำนายแนวโน้มที่สมาชิกคนหนึ่ง ๆ จะตัดสินใจเปลี่ยนเปอร์เซ็นต์หรือแผนการลงทุน ซึ่งจะช่วยให้กองทุนฯ สามารถเลือกกิจกรรมส่งเสริมการตลาดออนไลน์ เช่น ข้อความโทรศัพท์ อีเมล หรือไลน์ เฉพาะต่อกลุ่มสมาชิกที่มีแนวโน้มสูงที่จะเปลี่ยนแผนหรือสัดส่วนการลงทุนโดยเฉพาะ

จากการทดสอบผลการทำนายของโมเดลกับประวัติสมาชิกในปี พ.ศ. 2562 พบว่าร้อยละ 70 ของสมาชิกในกลุ่มที่มีผลทำนายว่ามีแนวโน้มสูงที่จะเปลี่ยนแผน ได้เปลี่ยนแผนจริง ซึ่งแม่นยำกว่าการประมาณจากค่าเฉลี่ยโดยทั่วไปถึง 18 เท่า ในขณะที่พบว่าร้อยละ 21 ของสมาชิกในกลุ่มที่มีผลทำนายว่ามีแนวโน้มปานกลางที่จะเปลี่ยนแผน ได้เปลี่ยนแผนจริง แม่นยำกว่าการประมาณจากค่าเฉลี่ยโดยทั่วไป 5 เท่า



ภาพที่ 7.20 ภาพรวมของโมเดลประเมินสมาชิกรายบุคคล

นอกจากนี้ ยังสามารถวิเคราะห์ข้อมูลแนวโน้มการเปลี่ยนแปลงรายกลุ่มได้จาก Dashboard แสดงข้อมูลการเปลี่ยนแปลงย้อนหลังของกลุ่มสมาชิกประเภทต่าง ๆ ดังภาพที่ 7.21 เพื่อใช้เลือกพิจารณากลุ่มที่มีแนวโน้มที่จะเปลี่ยนแปลงสูงได้ สำหรับใช้ส่งเสริมการตลาด ผ่านการทำกิจกรรม หรือการจัดสัมมนา เป็นต้น โมเดลต้นแบบเหล่านี้สามารถนำไปปรับใช้ และเชื่อมต่อกับข้อมูลในฐานข้อมูลจริงของกองทุนบำเหน็จบำนาญข้าราชการ และสามารถ นำผลวิเคราะห์ไปใช้ในกิจกรรมส่งเสริมการลงทุนตามเป้าหมายขององค์กร



ภาพที่ 7.21 ภาพรวม Dashboard ประเมินรายกลุ่มของสมาชิก
ที่มีแนวโน้มออมเพิ่มและเปลี่ยนแปลง



จะเห็นได้ว่าการใช้ประโยชน์ข้อมูลเชิงวิเคราะห์ของหน่วยงานภาครัฐใน 10 ตัวอย่างกรณีศึกษาที่กล่าวในบทนี้ เป็นก้าวสำคัญของการพัฒนาสู่ความเป็นรัฐบาลอัจฉริยะของประเทศไทยในอนาคต สำหรับผู้ที่สนใจกรณีศึกษาเพิ่มเติมจากหน่วยงานรัฐอื่น ๆ สามารถติดต่อขอข้อมูลมายัง GBDi หรือ สถาบันส่งเสริมการวิเคราะห์และบริหารข้อมูลขนาดใหญ่ภาครัฐได้ ตามช่องทางต่อไปนี้



Gbdi.depa.or.th หรือ
Bigdata.go.th



gbdi-info@depa.or.th



เลขที่ 80 ลาดพร้าวซอย 4
แขวงจอมพล เขตจตุจักร กทม.
10900



02026-2333 ต่อ 2508, 2525



Govbigdata



Govbigdata



Govbigdata



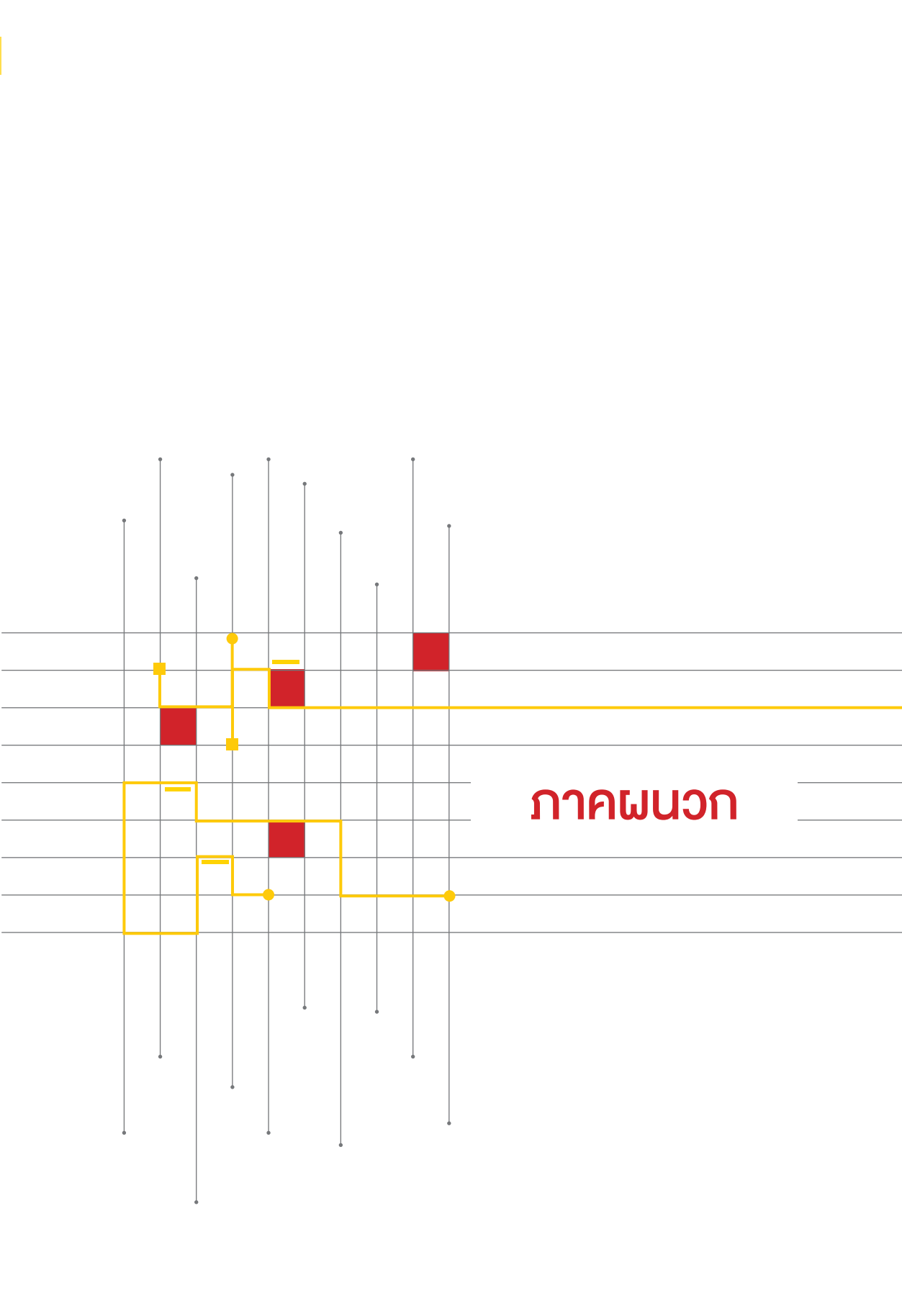
Govbigdata



@gbdi



LINE @gbdi



ภาคผนวก ก

โครงสร้างกรอบหลักสูตร

กรอบโครงสร้างหลักสูตรจำแนกตามกลุ่มเป้าหมายออกเป็น 3 กลุ่ม ได้แก่ นักวิเคราะห์ข้อมูล (Data Analyst) นักวิทยาศาสตร์ข้อมูล (Data Scientist) และวิศวกรข้อมูล (Data Engineer) โดยมีรายละเอียดของแต่ละหลักสูตร ดังนี้

1. หลักสูตรเพื่อพัฒนาเป็นนักวิเคราะห์ข้อมูล (Data Analyst)

1.1. หลักสูตรการวิเคราะห์ข้อมูลสำหรับผู้ปฏิบัติ (ระยะเวลา 7-13 วัน) Data Analytics for Analysts

เงื่อนไขที่ต้องมีมาก่อน (Prerequisite) : ไม่มี

รูปแบบการสอน : Lecture, Problem-based Learning, In-class Discussion, Workshop

ผลของการเรียน (Learning Outcome) :

ผลของการเรียน	วิธีวัดผล
ผู้เรียนสามารถออกแบบวิธีการและนำข้อมูลมาใช้ได้อย่างเหมาะสม	สนทนาในห้องเรียน
ผู้เรียนสามารถกำหนดประเด็นที่จะนำข้อมูลมาใช้และระบุประเภทข้อมูลที่จำเป็นได้อย่างถูกต้อง	สอบข้อเขียน
ผู้เรียนสามารถออกแบบวิธีการเก็บข้อมูลได้อย่างถูกต้องและสามารถสรุปผลเชิงสถิติจากข้อมูลได้อย่างถูกต้อง	สอบโดยการนำเสนอผลงาน (Presentation)
ผู้เรียนสามารถเล่าผลสรุปจากการวิเคราะห์และออกแบบการนำเสนอผลได้อย่างมีประสิทธิภาพและสามารถใช้เครื่องมือประเภท Data Visualization ขั้นพื้นฐานได้	สอบปฏิบัติ

องค์ประกอบของหลักสูตร :

เนื้อหา	รายละเอียด
1. การเก็บและการบูรณาการข้อมูล (Data Collection and Integration) ระยะเวลา 1-2 วัน	
- การสำรวจความเห็น สาธารณะและประชามติ (Poll and Public Opinion Survey)	การสำรวจด้วยตัวอย่าง การสำรวจความคิดเห็นสาธารณะ และประชามติ การกำหนดกรอบแนวคิดและวัตถุประสงค์ ในการสำรวจด้วยตัวอย่าง การออกแบบแบบสอบถาม และการร่างแบบสัมภาษณ์ การทดลองแบบสอบถาม และแบบสัมภาษณ์ก่อนการลงพื้นที่ วิธีการเก็บข้อมูลด้วย การสำรวจและการสัมภาษณ์ การตรวจสอบคุณภาพของ ข้อมูลภาคสนาม การจัดเตรียมข้อมูลและวิเคราะห์ข้อมูล สำรวจ การนำความคิดเห็นสาธารณะและประชามติไปใช้ ในการวางนโยบายและแผนการพัฒนา การประเมินผล กระทบของการบังคับใช้กฎหมาย (Regulatory Impact Assessment) ด้วยการสำรวจความคิดเห็นสาธารณะ
- การสำรวจออนไลน์ (Online Survey)	การออกแบบแบบสอบถามสำหรับการสำรวจออนไลน์ การจัดทำแบบสอบถามออนไลน์และ Platform สำหรับ การสำรวจออนไลน์ การบริหารจัดการการเก็บข้อมูล ออนไลน์ การตรวจสอบคุณภาพของข้อมูลจากการสำรวจ ออนไลน์
- เส้นทางไหลของข้อมูล (Data Pipeline)	การวิเคราะห์ระบบและเส้นทางไหลของข้อมูล การเชื่อมต่อข้อมูลเข้าสู่ฐานข้อมูล การออกแบบฐานข้อมูล เพื่อการเชื่อมต่อข้อมูล การบันทึกและการจัดเก็บข้อมูล การสำรองข้อมูล

เนื้อหา	รายละเอียด
2. บทนำสู่สถิติศาสตร์ (Introduction to Statistics) ระยะเวลา 2-4 วัน	
- สถิติเชิงบรรยาย (Descriptive Statistics)	ระดับการวัดตัวแปร ความรู้เบื้องต้นเกี่ยวกับสถิติเชิงพรรณนา ได้แก่ ค่าเฉลี่ยชนิดต่าง ๆ มัธยฐาน ฐานนิยม ความแปรปรวน ส่วนเบี่ยงเบนมาตรฐาน ความเบ้ ความโด่ง สหสัมพันธ์ระหว่างตัวแปรทั้งตัวแปรจัดประเภท และตัวแปรจัดกลุ่ม ดัชนีขนาดอิทธิพล การนำเสนอค่าสถิติให้เหมาะสมกับข้อมูลแต่ละประเภท
- สถิติเชิงอนุมานขั้นพื้นฐาน (Basic Statistical Inference)	สมมุติฐานการวิจัยและสมมุติฐานทางสถิติ ความรู้เบื้องต้นเกี่ยวกับสถิติเชิงอนุมาน ความผิดพลาดในการทดสอบสมมุติฐานทางสถิติ การประมาณค่าพารามิเตอร์แบบจุดและแบบช่วง การทดสอบค่าเฉลี่ย ค่าสัดส่วน และความแปรปรวนสำหรับประชากรเดียว สองประชากร และหลายประชากร การทดสอบความเป็นอิสระแก่กันของตัวแปรจัดประเภทด้วยไคกำลังสอง การทดสอบสมมุติฐานเกี่ยวกับค่าสหสัมพันธ์

เนื้อหา	รายละเอียด
3. การวิเคราะห์ข้อมูลเชิงบุกเบิกและภาพนิทัศน์จากข้อมูล (Exploratory Data Analysis and Data Visualization) ระยะเวลา 3-4 วัน	
การวิเคราะห์ข้อมูลเชิงบุกเบิก ได้แก่ การตรวจสอบความผิดปกติในการกรอกข้อมูล ค่าสูญหาย (Missing value) ค่าสุดโต่ง (Outlier) การตรวจสอบแจกแจงของข้อมูลและการแปลงไม่เชิงเส้นตรง การตรวจสอบความแปรปรวนของข้อมูล ภาพนิทัศน์สำหรับข้อมูลจัดประเภท ภาพนิทัศน์สำหรับการแจกแจงของข้อมูล ภาพนิทัศน์สำหรับสหสัมพันธ์ และการถดถอยรวมถึงความสัมพันธ์ที่ไม่เป็นเชิงเส้นตรงและการวิเคราะห์ส่วนเหลือ (Residual Analysis) ภาพนิทัศน์สำหรับข้อมูลอนุกรมเวลา ภาพนิทัศน์สำหรับข้อมูลเชิงพื้นที่ การออกแบบภาพนิทัศน์สำหรับข้อมูล การเล่าเรื่องจากภาพนิทัศน์ การใช้โปรแกรมสำเร็จรูปในการวิเคราะห์ข้อมูลเชิงบุกเบิกและการสร้างภาพนิทัศน์จากข้อมูล	
4. การวิเคราะห์เชิงทำนายและการพยากรณ์ขั้นพื้นฐาน (Basic Predictive Analytics and Forecasting) ระยะเวลา 2-3 วัน (ไม่บังคับ)	
- การวิเคราะห์ถดถอยและสหสัมพันธ์อย่างง่าย (Simple Correlation and Regression Analysis)	สหสัมพันธ์และแนวคิดของการวิเคราะห์การถดถอยเชิงเส้นอย่างง่าย
- การวิเคราะห์สหสัมพันธ์และการวิเคราะห์ถดถอยพหุคูณ (Multiple Correlation and Regression Analysis)	การวิเคราะห์สหสัมพันธ์และการวิเคราะห์ถดถอยพหุคูณ การสร้างแบบตัวแบบและการคัดเลือกตัวแปรในแบบจำลอง การวิเคราะห์ส่วนเหลือ (Residual Analysis) การวิเคราะห์ถดถอยพหุนาม (Polynomial Regression Analysis) ตัวอย่างการประยุกต์ใช้การวิเคราะห์สหสัมพันธ์และการวิเคราะห์ถดถอยพหุคูณในการบริหารรัฐกิจ อันได้แก่ การพยากรณ์เพื่อการบริหาร การศึกษาปัจจัยสัมพันธ์กับผลการปฏิบัติราชการ การศึกษาความสัมพันธ์ระหว่างต้นทุนและตัวขับเคลื่อน

เนื้อหา	รายละเอียด
- การวิเคราะห์อนุกรมเวลาแบบคลาสสิก (Classical Decomposition of Time Series)	ความรู้เบื้องต้นเกี่ยวกับข้อมูลอนุกรมเวลา ความคลาดเคลื่อนในการพยากรณ์อนุกรมเวลา การพยากรณ์อนุกรมเวลาโดยการวิเคราะห์การถดถอย การพยากรณ์โดยการปรับให้เรียบ ตัวแบบอนุกรมเวลาที่มีแนวโน้มและมีฤดูกาล ตัวแปรอนุกรมเวลาคลาสสิกแบบผลคูณ (Multiplicative Classical Decomposition Model) ตัวอย่าง การประยุกต์ใช้การวิเคราะห์อนุกรมเวลาแบบคลาสสิกในการวางแผนการจัดเก็บภาษี การประยุกต์ใช้การวิเคราะห์อนุกรมเวลาแบบคลาสสิกในการพยากรณ์ราคาสินค้าเกษตรและการวางแผนการผลิตทางการเกษตร
5. การวิเคราะห์ข้อมูลเชิงพื้นที่-เวลา (Spatial-Temporal Data Analysis) ระยะเวลา 1 วัน (ไม่บังคับ)	ลักษณะของข้อมูลเชิงเวลาและพื้นที่ ตัวอย่างการประยุกต์ใช้ข้อมูลเชิงพื้นที่และเวลาในการบริหารรัฐกิจ แหล่งข้อมูลเชิงพื้นที่และเวลาของภาครัฐ การจัดเตรียมข้อมูลเชิงพื้นที่-เวลา การสร้างภาพนิทัศน์สำหรับข้อมูลเชิงพื้นที่และเวลา สหสัมพันธ์เชิงพื้นที่ สหสัมพันธ์เชิงพื้นที่แบบท้องถิ่น (Local Indicator of Spatial Autocorrelation) การวิเคราะห์การถดถอยเชิงพื้นที่ การวิเคราะห์ถดถอยเชิงพื้นที่และเวลา

1.2. หลักสูตรการวิเคราะห์ข้อมูลสำหรับผู้บริหาร (ระยะเวลา 1-4 วัน) Data Design for Executives

เงื่อนไขที่ต้องมีมาก่อน (Prerequisite) : ไม่มี

รูปแบบการสอน : Lecture, In-class Discussion, Workshop

ผลของการเรียน (Learning Outcome) :

ผลของการเรียน	วิธีวัดผล
สามารถตัดสินใจจากการประยุกต์ใช้หลักฐานเชิงประจักษ์	นำเสนอผลงาน (Presentation)
ชี้ประเด็นในการนำข้อมูลมาใช้อย่างเหมาะสม	นำเสนอผลงาน (Presentation)
ให้แนวทางการดำเนินงานด้านข้อมูลต่อผู้ใต้บังคับบัญชาได้อย่างเหมาะสม	อภิปรายในชั้นเรียนด้วยกรณีศึกษา

องค์ประกอบของหลักสูตร :

หลักสูตรนี้ถูกออกแบบเป็นหลักสูตรระยะสั้นสำหรับผู้บริหารเพื่อสร้างความเข้าใจทางด้านการใช้ประโยชน์ข้อมูลขององค์กรผ่านการอธิบายกรณีศึกษา (Case Studies) ผู้สอนเลือกกรณีศึกษาเพียงบางส่วน (จากหัวข้อ 4-10) ที่เกี่ยวข้องกับงานของผู้เรียนเพื่อใช้ประกอบการอภิปราย นอกจากนี้เนื้อหา รวมถึงการปูพื้นฐานด้านการใช้ประโยชน์จากการวิเคราะห์ข้อมูลและการกำหนดโจทย์ปัญหาโดยอาศัยเครื่องมือ Design Thinking Canvas

เนื้อหา	รายละเอียด
1. Introduction to Data Analytics ระยะเวลา 1 ชั่วโมง	หลักการ ความหมาย และตัวอย่างการใช้งานของข้อมูลขนาดใหญ่และคลังข้อมูล ความรู้พื้นฐานด้านการวิเคราะห์ข้อมูลเพื่อประกอบการตัดสินใจ โดยกล่าวถึง การวิเคราะห์แบบ Descriptive, Predictive, และ Prescriptive ผ่านกรณีศึกษา
2. Data Project Management ระยะเวลา 2 ชั่วโมง	กระบวนการของวิทยาการข้อมูล กระบวนการในการพัฒนาผลิตภัณฑ์ข้อมูล การวางแผนโครงการที่เกี่ยวข้องกับข้อมูล หลักการระบบแบบน้ำตก (Water fall Model) และแบบอไจล์ (Agile Development) การบริหารทีม และการจัดการทรัพยากรมนุษย์ในการบริหารโครงการ การประเมินและการติดตามความก้าวหน้าของโครงการ การจัดสรรและการบริหารทรัพยากรในการวิเคราะห์ข้อมูล
3. Introduction to Data Analytics ระยะเวลา 3 ชั่วโมง	การทำ Workshop เพื่อการกำหนดโจทย์ปัญหาที่ชัดเจนและเป็นไปได้ ในกรอบเวลาและงบประมาณที่มีอยู่ โดยอาศัยเครื่องมือ Design thinking Canvas (ฉบับปรับปรุงสำหรับโครงการข้อมูล) ประกอบการเรียนรู้และระดมสมอง พร้อมทั้งนำเสนอผลงานและแลกเปลี่ยนความคิดเห็นระหว่างผู้สอนและผู้เข้าร่วมอบรม
4. การวิเคราะห์ทรัพยากรมนุษย์ (Human Resource Analytics) ระยะเวลา 3 ชั่วโมง	ตัวแปรเกณฑ์และตัวแปรพยากรณ์ในการจัดการทรัพยากรบุคคล การวิเคราะห์อัตราค่าจ้าง การวางแผนการสืบทอดตำแหน่ง การวัดความผูกพันใจมั่นของบุคลากร และเจตคติที่มีต่อองค์กร การศึกษาพฤติกรรมมององค์กร การศึกษาการเปลี่ยนพฤติกรรมองค์กร การศึกษาโค้งการเรียนรู้ของบุคลากร ตัวแบบทำนายการลาออกของบุคลากร ตัวอย่างและการประยุกต์ใช้การวิเคราะห์ทรัพยากรมนุษย์ในการจัดการภาครัฐ

เนื้อหา	รายละเอียด
5. การวิเคราะห์ห่วงโซ่อุปทานและโลจิสติกส์ (Supply Chain and Logistics Analytics) ระยะเวลา 3 ชั่วโมง	การวิเคราะห์ต้นทุน ด้วยการบัญชีต้นทุนตามกิจกรรม (Activity-Based Costing) การจัดสรรทรัพยากร ด้วยกำหนดการเชิงเส้น การจัดการโลจิสติกส์ด้วยตัวแบบการขนส่ง การบริหารสินค้าคงคลัง การจัดการการรอ (Waiting Time Management) การวางแผนและควบคุมโครงการ การวิเคราะห์การแข่งขันด้วยทฤษฎีเกมส์ ตัวอย่างและการประยุกต์ใช้การวิเคราะห์ห่วงโซ่อุปทานและโลจิสติกส์ในการจัดการภาครัฐ
6. การวิเคราะห์การเงิน (Financial Analytics) ระยะเวลา 3 ชั่วโมง	การประเมินมูลค่าของการลงทุน (Financial Valuation) และการวิเคราะห์ความคุ้มค่าทางการเงินด้วยคณิตศาสตร์การเงินขั้นพื้นฐาน การวิเคราะห์งบการเงิน ทฤษฎีการจัดพอร์ตการลงทุนและตัวแบบ Capital Asset Pricing การวิเคราะห์ทางเทคนิคด้วยการวิเคราะห์อนุกรมเวลา การพยากรณ์ทางการเงิน ตัวอย่างและการประยุกต์ใช้การวิเคราะห์การเงินในการจัดการภาครัฐ
7.การวิเคราะห์ความเสี่ยงเชิงปริมาณ (Quantitative Risk Analytics) ระยะเวลา 3 ชั่วโมง	การตรวจสอบและการตรวจจับทุจริต/การตรวจจับความผิดปกติ (Fraud and Anomaly Detection) การจัดอันดับความน่าเชื่อถือ (Credit Scoring) การวิเคราะห์พฤติกรรม ตัวแบบการล้มละลาย ตัวแบบจำแนกการจ่ายคืนหนี้ (Debt Aging) การวัดและจำลองความเสี่ยง ตัวอย่างและการประยุกต์ใช้การวิเคราะห์ความเสี่ยงเชิงปริมาณในการจัดการภาครัฐ

เนื้อหา	รายละเอียด
8. การวิเคราะห์เศรษฐมิติ (Econometric Analysis) ระยะเวลา 3 ชั่วโมง	การวิเคราะห์เศรษฐมิติเบื้องต้น การวิเคราะห์อัตโนมัติ (Autocorrelation) ปัญหาการร่วมเชิงเส้นพหุ (Multicollinearity) ตัวแบบ Autoregressive Integrated Moving Average (ARIMA) ตัวแบบ Generalized Autoregressive Conditional Heteroscedasticity (GARCH) การวิเคราะห์ความสัมพันธ์เชิงสาเหตุและ Cointegration และการวิเคราะห์ Vector Autoregressive ตัวอย่างและการประยุกต์ใช้การวิเคราะห์เศรษฐมิติในการจัดการภาครัฐ
9. การวิเคราะห์ประชากร (Demographic Analysis) ระยะเวลา 3 ชั่วโมง	การวิเคราะห์ทางประชากร ส่วนประกอบประชากร การสมรสและภาวะเจริญพันธุ์ การเจ็บป่วยและการตาย การวัดภาวะการตาย การกระจายของประชากรและการย้ายถิ่น การปรับเข้ามาตรฐาน การเปลี่ยนแปลงของประชากรและการคาดคะเน การสร้างตารางชีพ แบบจำลองทางประชากร การประมาณค่าสถิติประชากรจากข้อมูลที่ไม่สมบูรณ์ ตัวอย่างและการประยุกต์ใช้การวิเคราะห์ประชากรในการจัดการภาครัฐ
10. การวิเคราะห์การตลาดสำหรับการจัดการภาครัฐ (Marketing Analytics for Public Management) ระยะเวลา 3 ชั่วโมง	การศึกษาพฤติกรรมประชาชนและผู้ใช้บริการสาธารณะ ด้วยการวิจัยเชิงสำรวจและการออกแบบสอบถาม การวัดความพึงพอใจของประชาชนด้วยการวิเคราะห์องค์ประกอบเชิงยืนยัน การวิเคราะห์จัดกลุ่ม (Market segmentation) ผู้ใช้บริการสาธารณะ การกำหนดส่วนประสมทางการตลาด (Marketing Mix) สำหรับบริการสาธารณะด้วย Conjoint Analysis การวัดตำแหน่งทางการตลาด (Market Positioning) และแผนที่ความคิด (Perceptual Mapping) ของหน่วยงาน/บริการภาครัฐด้วยการวิเคราะห์ Multidimensional Scaling และ Correspondence Analysis

2. หลักสูตรเพื่อพัฒนาเป็นนักวิทยาศาสตร์ข้อมูล (Data Scientist)

2.1. หลักสูตรวิทยาศาสตร์ข้อมูลพื้นฐาน (ระยะเวลา 17-25 วัน) Basic Data Science

เงื่อนไขที่ต้องมีมาก่อน (Prerequisite) : ผู้เข้ารับการอบรมจำเป็นต้องมีทักษะภาษาโปรแกรมภาษาใดก็ได้ได้ผ่านการทดสอบตรรกะด้วยผังงาน (Flowchart)

รูปแบบการสอน : Lecture, Problem-based Learning, Project-based Learning, Workshop

ผลของการเรียน (Learning Outcome) :

ผลของการเรียน	วิธีวัดผล
ผู้เรียนสามารถทำความเข้าใจข้อมูลแต่ละประเภทและสรุปผลเชิงสถิติจากข้อมูลได้อย่างถูกต้อง	สอบข้อเขียน
ผู้เรียนสามารถเขียนโปรแกรมภาษา R หรือ Python เพื่อทำความสะอาดข้อมูลและขึ้นโมเดลวิเคราะห์พื้นฐานได้	สอบปฏิบัติ
ผู้เรียนสามารถสร้างโมเดลคณิตศาสตร์พื้นฐานตามบริบทของงานได้ด้วยตัวเอง	สอบโดยการอภิปรายกลุ่มและนำเสนอผลงาน
ผู้เรียนสามารถเล่าผลสรุปจากการวิเคราะห์และออกแบบการนำเสนอผลได้อย่างมีประสิทธิภาพ	สอบโดยการนำเสนอผลงาน (Presentation)

องค์ประกอบของหลักสูตร :

เนื้อหา	รายละเอียด
1. การเก็บและการบูรณาการข้อมูล (Data Collection and Integration) ระยะเวลา 1 วัน	ความรู้เบื้องต้นเกี่ยวกับการวิเคราะห์ข้อมูล รวมถึงการอธิบายเทคนิคทั่วไปที่สำคัญสำหรับการวิเคราะห์และแปลงให้ได้มาซึ่งสารสนเทศที่มีความหมายจากชุดข้อมูลต่าง ๆ อธิบายหลักการที่เกี่ยวข้องกับการดำเนินการเก็บรวบรวมข้อมูล การแปลง การทำความสะอาดข้อมูล (Data Cleaning and Integration) และการตัดสินใจข้อมูลโดยเน้นการคิดเชิงวิเคราะห์
2. Data Science with Basic Python Programing or R Programing ระยะเวลา 3 วัน	สอนการเขียนโปรแกรมพื้นฐานเฉพาะเพื่อการวิเคราะห์ข้อมูล โดยการพัฒนาความรู้ความสามารถในการเขียนโปรแกรมภาษา Python หรือภาษา R ซึ่งเป็นภาษาที่ใช้กันอย่างแพร่หลาย
3. Introduction to Statistics ระยะเวลา 2-4 วัน	
- Descriptive Statistics	ความรู้เบื้องต้นเกี่ยวกับสถิติเชิงพรรณนา ได้แก่ ค่าเฉลี่ย มัธยฐาน ฐานนิยม ความแปรปรวน และค่าเบี่ยงเบนมาตรฐาน
- Basic Statistical Inference	ความรู้เบื้องต้นเกี่ยวกับสถิติเชิงอนุมาน การประเมินค่าพารามิเตอร์ในประชากร การทดสอบสมมติฐาน ความคลาดเคลื่อนในการทดสอบสมมติฐาน การประมาณแบบช่วง การประเมินสมบัติและการควบคุมความผิดพลาดของการทดสอบสมมติฐานและการประมาณแบบช่วง ตัวอย่างเนื้อหาเช่น Frequency Distributions, Central Tendency, Correlation, Hypothesis Testing, ANOVA

เนื้อหา	รายละเอียด
4. Exploratory Data Analysis ระยะเวลา 2-3 วัน	การวิเคราะห์ข้อมูลเบื้องต้น ประกอบด้วย การตรวจสอบความผิดปกติในการกรอกข้อมูล ค่าสูญหาย (Missing value) ค่าผิดปกติ (Outlier) การแจกแจงของข้อมูล ความเท่ากันของความแปรปรวน (Homogeneity of Variance or Equality of Variance) ความสัมพันธ์เชิงเส้นตรง ภาวะร่วมเส้นตรงพหุ (Multicollinearity) และส่วนเหลือ (Residual) รวมทั้งการนำเสนอข้อมูล การแสดงความถี่ การแสดงการเปรียบเทียบและแนวโน้ม การแสดงการจัดลำดับ
5. Data Cleansing with Python or R ระยะเวลา 1-2 วัน	การทำความสะอาดข้อมูล เพื่อเตรียมการสำหรับการวิเคราะห์ โดยเรียนรู้หลักการพื้นฐานของการประเมินข้อมูลและทำความสะอาดข้อมูลด้วยการเขียนโปรแกรม ภาษา R หรือ Python

เนื้อหา	รายละเอียด
6. Basic Machine Learning ระยะเวลา 4-6 วัน	ความรู้เบื้องต้นเกี่ยวกับการเรียนรู้ของเครื่อง ขั้นตอน ที่จำเป็นสำหรับการสร้างแบบจำลองการเรียนรู้ของ เครื่อง แนวคิดการเรียนรู้ด้วยเครื่องและประเภทการเรียนรู้ แบบมีผู้สอน (Supervised Learning) และการเรียนรู้ แบบไม่มีผู้สอน (Unsupervised Learning) การประเมิน ขั้นตอนวิธีการเรียนรู้ของเครื่อง การวิเคราะห์ที่ไม่ขึ้นกับ ขั้นตอนวิธี การเรียนรู้แบบเสริมกำลัง (Reinforcement Learning) อัลกอริทึมสำหรับการจัดแบ่งประเภท เรียนรู้ คลังโปรแกรมที่ใช้อย่างแพร่หลายของอัลกอริทึมสำหรับ การเรียนรู้ด้วยเครื่อง การเตรียมประมุลผลข้อมูลเพื่อสร้าง Data Set ที่มีคุณภาพ การบีบอัดเพื่อลดมิติของข้อมูล การประเมินผลแบบจำลองและการปรับแต่งพารามิเตอร์ - Linear Regression (Generalized Linear Model, Linear Regression, Poisson Regression, Survival Analysis) - Classification (Decision Tree, Support Vector Machine, Random Forest) - Clustering (kNN, Hierarchical Clustering, K-Mean Clustering) - Co-Occurrence Analysis (Content-Based Filtering, Collaborative Filtering) - Performance Evaluation and Cost-Benefit (Accuracy, ROC, RMSE, Cost-Benefit)
7. Data Modeling Project ระยะเวลา 4-6 วัน	เลือกโจทย์ปัญหาพื้นฐานขององค์กรที่มีข้อมูลพร้อมอยู่ แล้วมาสร้างโมเดลคณิตศาสตร์ที่เหมาะสมเพื่อการวิเคราะห์เชิงพยากรณ์และนำเสนอผลงาน

2.2. หลักสูตรวิทยาศาสตร์ข้อมูลชั้นกลาง (ระยะเวลา 9-12 วัน) Intermediate Data Science

เงื่อนไขที่ต้องมีมาก่อน (Prerequisite) : ผู้เข้ารับการอบรมจำเป็นต้องผ่านหลักสูตรวิทยาศาสตร์ข้อมูลพื้นฐาน (Basic Data Science) หรือ มีความรู้พื้นฐานตามหลักสูตรวิทยาศาสตร์ข้อมูลพื้นฐานโดยผ่านการทดสอบแบบออนไลน์

รูปแบบการสอน : Lecture, Problem-Based Learning, Project-Based Learning, Workshop

ผลของการเรียน (Learning Outcome) :

ผลของการเรียน	วิธีวัดผล
ผู้เรียนสามารถทำความเข้าใจข้อมูลขนาดใหญ่ เทคโนโลยีที่เกี่ยวข้องและยกตัวอย่างการประยุกต์ใช้งานได้	การอภิปรายและการสอบข้อเขียน
ผู้เรียนสามารถพัฒนาโมเดลคณิตศาสตร์พื้นฐานด้วย Spark R หรือ PySpark	สอบปฏิบัติ
ผู้เรียนสามารถแสดงผลการวิเคราะห์ข้อมูลขนาดใหญ่และออกแบบการนำเสนอผลได้อย่างมีประสิทธิภาพ	สอบโดยการนำเสนอผลงาน (Presentation)

องค์ประกอบของหลักสูตร :

เนื้อหา	รายละเอียด
1. Basic Big Data Management ระยะเวลา 2 วัน	ความรู้เบื้องต้นเกี่ยวกับความหมายและคุณลักษณะของข้อมูลขนาดใหญ่ เรียนรู้การประยุกต์ใช้ข้อมูลขนาดใหญ่ผ่านการอภิปรายกรณีศึกษาในบริบทต่างๆ เทคโนโลยีและแพลตฟอร์มพื้นฐาน เช่น Hadoop, Spark วิธีการจัดเก็บข้อมูลขนาดใหญ่ หลักการใช้ข้อมูลเพื่อการวิเคราะห์ผ่านเครื่องมือด้าน Big Data การนำเสนอข้อมูลขนาดใหญ่ด้วยภาพ
2. Basic Big Data Analytics and Programming ระยะเวลา 7-10 วัน	ความรู้เบื้องต้นสำหรับการวิเคราะห์ข้อมูลขนาดใหญ่ และการเขียนโปรแกรมโดยใช้ Library และ Software Package สำหรับการวิเคราะห์ข้อมูลขนาดใหญ่ผ่านการลงทำจริง - Spark Ecosystem - PySpark หรือ Spark R - Basic Algorithm and Mathematics for Big Data Analysis - Basic Data Mining Over Massive Dataset Using Spark - Storytelling: Big Data Visualization

2.3. หลักสูตรวิทยาศาสตร์ข้อมูลขั้นสูง (ระยะเวลา 8-16 วัน) Advance Data Science

เงื่อนไขที่ต้องมีมาก่อน (Prerequisite) : ผู้เข้ารับการอบรมจำเป็นต้องผ่านหลักสูตรวิทยาศาสตร์ข้อมูลขั้นกลาง (Intermediate Data Science) หรือ มีความรู้พื้นฐานตามหลักสูตรวิทยาศาสตร์ข้อมูลพื้นฐานและขั้นกลางโดยผ่านการทดสอบแบบออนไลน์

รูปแบบการสอน : Lecture, Workshop

ผลของการเรียน (Learning Outcome) :

ผลของการเรียน	วิธีวัดผล
ผู้เรียนมีความเข้าใจและสามารถประมวลผล/วิเคราะห์ข้อมูลประเภทไฟล์อักษร (Text File) ได้	การสอบข้อเขียนและสอบปฏิบัติ
ผู้เรียนมีความเข้าใจและสามารถประมวลผล/วิเคราะห์ข้อมูลประเภทไฟล์ภาพได้	การสอบข้อเขียนและสอบปฏิบัติ
ผู้เรียนมีความเข้าใจและสามารถประมวลผล/วิเคราะห์ข้อมูลประเภทไฟล์วิดีโอได้	การสอบข้อเขียนและสอบปฏิบัติ
ผู้เรียนมีความเข้าใจและสามารถประมวลผล/วิเคราะห์ข้อมูลประเภทไฟล์เสียงได้	การสอบข้อเขียนและสอบปฏิบัติ

องค์ประกอบของหลักสูตร :

หลักสูตรนี้ถูกออกแบบเพื่อเสริมองค์ความรู้ในการประมวลผลข้อมูลประเภทไม่มีโครงสร้าง เช่น ไฟล์เสียง ไฟล์วิดีโอ และไฟล์อักษร รวมทั้งกล่าวถึงอัลกอริทึมทางสาขาปัญญาประดิษฐ์ (AI) และการประยุกต์ใช้อัลกอริทึมเพื่อการวิเคราะห์ข้อมูล โดยการสอนจะเน้นหลักการ ทฤษฎีพื้นฐาน และการพัฒนาโมเดลการวิเคราะห์โดยใช้ไลบรารีสำเร็จรูปและเครื่องมือ เช่น Tensorflow เป็นต้น เนื้อหาหลักสูตรแบ่งเป็นโมดูลย่อยโดยในแต่ละโมดูลผู้สอนจะบรรยายหลักการและทฤษฎีก่อนและ Workshop เพื่อให้ผู้เรียนได้ทดลองใช้เครื่องมือเพื่อทดสอบการประมวลผลข้อมูลประเภทต่าง ๆ โดยประมาณการเวลาการสอน 2 วันต่อโมดูล ผู้สอนจะเลือกโมดูลที่เหมาะสมกับบริบทในการทำงานของผู้เรียน

เนื้อหา	รายละเอียด
1. Neural Network	ระยะเวลา 2 วัน
2. Deep Learning	ระยะเวลา 2 วัน
3. GPU-Based Deep Learning Library and Tools	ระยะเวลา 2 วัน
4. Basic Natural Language Processing	ระยะเวลา 2 วัน
5. Text Mining	ระยะเวลา 2 วัน
6. Social Network Analysis	ระยะเวลา 2 วัน
7. Image Analytics	ระยะเวลา 2 วัน
8. Voice Analytics	ระยะเวลา 2 วัน
9. Voice Recognition	ระยะเวลา 2 วัน

3. หลักสูตรเพื่อพัฒนาเป็นนักวิศวกรข้อมูล (Data Engineering)

3.1. หลักสูตรนักวิศวกรข้อมูล (ระยะเวลา 17-25 วัน) Basic Data Engineering

เงื่อนไขที่ต้องมีมาก่อน (Prerequisite) : ผู้เข้ารับการอบรมจำเป็นต้องมีทักษะภาษาโปรแกรมภาษาใดก็ได้โดยผ่านการทดสอบตรรกะด้วยผังงาน (Flowchart)

รูปแบบการสอน : Lecture, Problem-based Learning, Project-based Learning, Workshop

ผลของการเรียน (Learning Outcome) :

ผลของการเรียน	วิธีวัดผล
ผู้เรียนมีความเข้าใจพื้นฐานเรื่องวิศวกรรมข้อมูล และการให้บริการข้อมูล	การสอบข้อเขียน
ผู้เรียนสามารถอธิบายถึงหลักการการกำกับดูแลข้อมูล	การอภิปรายหรือสอบปากเปล่า
ผู้เรียนสามารถติดตั้งและตั้งค่าพื้นฐานสำหรับระบบฐานข้อมูลแบบ SQL และ เข้าใจหลักการของคลังข้อมูล	การสอบปฏิบัติ
ผู้เรียนสามารถทำกระบวนการ ETL (SQL to Datawarehouse) ได้	การสอบปฏิบัติ

องค์ประกอบของหลักสูตร :

เนื้อหา	รายละเอียด
1. Introduction to Data Engineering ระยะเวลา 3 วัน	หลักการพื้นฐานและภาพรวมของวิศวกรรมข้อมูล (การแปลงข้อมูล การทำความสะอาดข้อมูลการบูรณาการข้อมูลพื้นฐาน) และวิทยาศาสตร์ข้อมูล (การตรวจวินิจฉัยข้อมูล การมองภาพข้อมูลการประมวลผลคุณลักษณะ) รวมไปถึงกระบวนการการกำกับดูแลคุณภาพข้อมูล และกฎเกณฑ์/กฎหมายต่าง ๆ ที่เกี่ยวข้องกับการรักษาความลับข้อมูล - Data Engineering Concept - Data Science Concept - Data Governance Methodology - Thai Data Privacy Law and Regulations
2. Relational Database / NoSQL / Data Warehouse Concept ระยะเวลา 2 วัน	หลักการเบื้องต้นและการเปรียบเทียบระหว่างฐานข้อมูลแบบสัมพันธ์ ฐานข้อมูลแบบ NoSQL และระบบบริหารจัดการคลังข้อมูล ผู้สอนอธิบายหลักการการประยุกต์ใช้เทคโนโลยีทั้ง 3 ร่วมกันเพื่อจัดการการประมวลผลข้อมูลขนาดใหญ่
3. Basic ETL and Data Cleans ระยะเวลา 3 วัน	พื้นฐานการ Extract, Transform, Load (ETL) ข้อมูลจากฐานข้อมูลเพื่อนำเข้าสู่ระบบคลังข้อมูลและการทำความสะอาดข้อมูลเพื่อเตรียมสำหรับการวิเคราะห์ โดยทดสอบทำจริงด้วยเครื่องมือสำเร็จรูป
4. Cloud Data Storage and Backup Process ระยะเวลา 2 วัน	หลักการและการทดลองเรื่องการใช้งานระบบคลาวด์และการดูแลกระบวนการด้านการสำรองข้อมูลพื้นฐาน
5. Open/Government Data and Data Services ระยะเวลา 1 วัน	หลักการและกรณีศึกษาด้านการให้บริการข้อมูลสาธารณะของหน่วยงานภาครัฐ รวมไปถึงการบรรยายสถาปัตยกรรมพื้นฐานที่เหมาะสมเรื่องการให้บริการข้อมูล

3.2. หลักสูตรวิศวกรรมข้อมูลชั้นกลาง (ระยะเวลา 12 วัน) Intermediate Data Engineering

เงื่อนไขที่ต้องมีมาก่อน (Prerequisite) : ผู้เข้ารับการอบรมจำเป็นต้องผ่านหลักสูตรวิศวกรรมข้อมูลพื้นฐาน (Basic Data Engineering) หรือมีความรู้พื้นฐานตามหลักสูตรวิศวกรรมข้อมูลพื้นฐาน โดยผ่านการทดสอบแบบออนไลน์

รูปแบบการสอน : Lecture, Project-based Learning, Workshop

ผลของการเรียน (Learning Outcome) :

ผลของการเรียน	วิธีวัดผล
ผู้เรียนสามารถค้นหา ดึง และ Preprocess ข้อมูลจากแหล่งข้อมูลทางเลือกที่เกี่ยวข้องกับบริบทงานได้	สอบโดยการนำเสนอผลงาน (Presentation)
ผู้เรียนเข้าใจหลักการและเทคนิคสำหรับการพัฒนาบริการข้อมูลภาครัฐ	สอบโดยการอภิปราย

องค์ประกอบของหลักสูตร :

หลักสูตรนี้ถูกออกแบบเพื่อเสริมองค์ความรู้ในการบริหารจัดการข้อมูลประเภทไม่มีโครงสร้าง โดยรวมถึงการจัดหา จัดเก็บ และจัดการข้อมูลจากแหล่งข้อมูลทั้งภายในและภายนอกองค์กร รวมทั้งเทคนิค เทคโนโลยีและวิธีการต่าง ๆ ที่เกี่ยวข้องกับการให้บริการข้อมูลอย่างเป็นระบบ โดยการสอนจะเน้นหลักการ ทฤษฎีพื้นฐาน และการทดลองใช้เครื่องมือสำเร็จรูปหรือระบบงานตัวอย่างเพื่อเสริมความเข้าใจ

เนื้อหา	รายละเอียด
1. Data Sources ระยะเวลา 1 วัน	เทคนิคและเทคโนโลยีสำหรับการจัดหา จัดการ และจัดเก็บข้อมูลที่ได้จากแหล่งข้อมูลที่ใช้งานกันอย่างแพร่หลาย เช่น Blog, IoTs, CCTV, Datawarehouse, Social Media, Mobile Apps, News and Websites, etc. รวมทั้งตัวอย่างการใช้ประโยชน์ข้อมูล
2. Web Scraping / Internet Crawler ระยะเวลา 2 วัน	เทคนิคพื้นฐานและเครื่องมือสำเร็จรูปสำหรับประยุกต์ใช้ในการดึงข้อมูลจากอินเทอร์เน็ต เพื่อการใช้ประโยชน์ในเชิงวิเคราะห์
3. Data Preprocessing ระยะเวลา 2 วัน	วิธีการ เทคนิคและเครื่องมือสำเร็จรูปต่างๆ ที่ใช้ในการประมวลผลก่อนการใช้งาน (Preprocess) ของข้อมูลที่ได้จากแหล่งข้อมูลหลากหลาย
4. Metadata Management System / Data Catalog ระยะเวลา 2 วัน	หลักการเรื่องเมทาดาทาของข้อมูลเพื่อจัดทำแค็ตตาล็อกข้อมูล และตัวอย่างบริการและ/หรือระบบสารสนเทศเพื่อการค้นหาข้อมูลในมิติต่าง ๆ
5. Data Exchange Services ระยะเวลา 1 วัน	หลักการและวิธีการที่ใช้กันอย่างแพร่หลายพร้อมตัวอย่างกรณีศึกษาเรื่องการแลกเปลี่ยนข้อมูลของหน่วยงานภาครัฐ
6. Micro-Service Concept, Programming, and API Design for Data Services ระยะเวลา 4 วัน	การสร้างบริการข้อมูลในรูปแบบพื้นฐาน

3.3. หลักสูตรวิศวกรรมข้อมูลขั้นสูง (ระยะเวลา 12 วัน) Advance Data Engineering
เงื่อนไขที่ต้องมีมาก่อน (Prerequisite) : ผู้เข้ารับการอบรมจำเป็นต้องผ่าน
หลักสูตรวิศวกรรมข้อมูลขั้นกลาง (Intermediate Data Science) หรือ มีความรู้พื้นฐานตาม
หลักสูตรวิศวกรรมข้อมูลพื้นฐานและขั้นกลางโดยผ่านการทดสอบแบบออนไลน์

รูปแบบการสอน : Lecture, Workshop

ผลของการเรียน (Learning Outcome) :

ผลของการเรียน	วิธีวัดผล
ผู้เรียนมีความเข้าใจและสามารถติดตั้งและตั้งค่า พื้นฐานระบบฐานข้อมูลแบบ NoSQL ได้	สอบปฏิบัติ
ผู้เรียนสามารถทำกระบวนการ ETL (NoSQL and File System to Data lake) ได้	สอบปฏิบัติ
ผู้เรียนสามารถออกแบบสถาปัตยกรรมระบบ บิกดาต้าตามบริบทการใช้งานได้	สอบโดยการนำเสนอผลงาน (Presentation)
ผู้เรียนเข้าใจหลักการเรื่องความปลอดภัยข้อมูล	สอบข้อเขียน
ผู้เรียนเข้าใจหลักการเรื่อง Data Virtualization	สอบข้อเขียน
ผู้เรียนสามารถจัดการและบูรณาการข้อมูลแบบ Streaming ได้	สอบปฏิบัติ

องค์ประกอบของหลักสูตร :

หลักสูตรนี้ถูกออกแบบเพื่อเสริมองค์ความรู้ในการบริหารจัดการระบบการประมวล บั๊กตาต้า โดยการสอนจะเน้นหลักการ ทฤษฎีพื้นฐาน และการทดลองใช้เครื่องมือสำเร็จรูป และ/หรือ การเขียน Script หลังจากผู้เรียนผ่านเนื้อหาในหัวข้อ 1-4 แล้ว ผู้สอนและผู้เรียน สามารถเลือกโมดูลย่อยเพื่อเสริมองค์ความรู้ให้เหมาะกับบริบทของงานได้

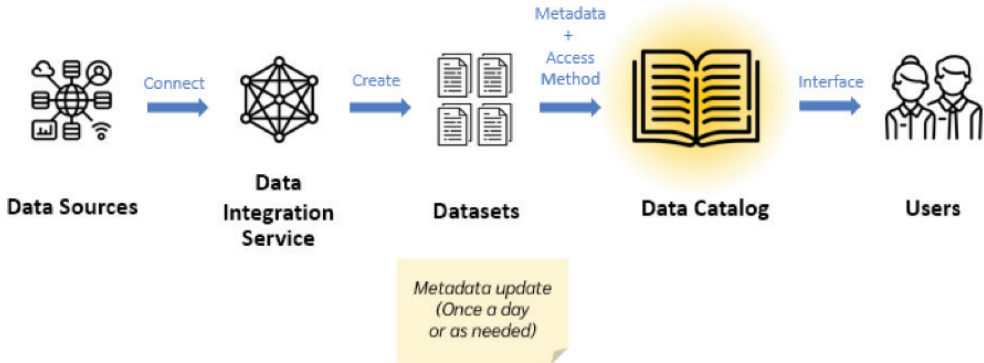
เนื้อหา	รายละเอียด
1. No SQL Database Management ระยะเวลา 3 วัน	
2. Big Data Ecosystem and Platform ระยะเวลา 2 วัน	
3. Data Integration and Service ระยะเวลา 2 วัน	
4. Big Data Architectural Design ระยะเวลา 2 วัน	การออกแบบสถาปัตยกรรมสำหรับ Big Data Platform โดยเน้น Hadoop Ecosystem และการเลือก Service Components ที่เหมาะสมสำหรับข้อมูล/โจทย์ปัญหา และระบบงานที่จะพัฒนา การคัดเลือกเทคโนโลยีด้านการการบูรณาการข้อมูลเพื่อนำเข้าข้อมูลและกลไกการเชื่อมต่อฐานข้อมูลทั้งแบบ Relational Database และแบบ NoSQL เข้ากับ Big Data Platform
5. Data Virtualization Concept and Tools ระยะเวลา 2 วัน	

เนื้อหา	รายละเอียด
6. Information Security and Tools ระยะเวลา 2 วัน	
7. Data Streaming Concept and Tools ระยะเวลา 2 วัน	

ภาคผนวก V.1

โครงสร้างและการใช้งานของรายการข้อมูล (Data Catalog)

ข.1.1 ขั้นตอนของการสร้างระบบสารสนเทศรายการข้อมูลโดยสรุป



ภาพที่ ข.1.1 ขั้นตอนของการสร้างระบบสารสนเทศรายการข้อมูล

ขั้นตอนในการสร้างระบบสารสนเทศรายการข้อมูลเริ่มจากหน่วยงานจะต้องสำรวจแหล่งข้อมูล (Data Source) ที่เป็นรูปแบบดิจิทัล ภายในหน่วยงานทั้งหมดไม่ว่าจะเป็นรูปแบบฐานข้อมูล ไฟล์ข้อความ ไฟล์ CSV ฯลฯ แหล่งข้อมูลเหล่านี้จะถูกเชื่อมต่อกันด้วยซอฟต์แวร์บูรณาการข้อมูล (Data Integration Tools) เพื่อทำการรวบรวมข้อมูลสำคัญในหน่วยงานพร้อมทั้งระบุแหล่งข้อมูลแต่ละชุด

จากนั้นข้อมูลเหล่านี้จะถูกจัดเป็นชุดข้อมูลตามความต้องการของผู้ใช้งาน โดยข้อมูลภายในชุดข้อมูลหนึ่ง ๆ สามารถมาจากแหล่งข้อมูลที่หลากหลายได้ และข้อมูลแต่ละชิ้นสามารถอยู่ในชุดข้อมูลได้มากกว่า 1 ชุดตามความเหมาะสม ชุดข้อมูลแต่ละชุดจะถูกกำหนดคำอธิบายข้อมูลหรือเมทาดาต้า (Metadata) เพื่อระบุรายละเอียดทั้งเชิงเทคนิคและเชิงการใช้งาน เช่น ชื่อเจ้าของข้อมูล สิทธิการใช้งาน ความถี่ในการปรับปรุงข้อมูล รวมไปถึงวิธีการร้องขอและเข้าถึงข้อมูลต่าง ๆ (อ้างอิงจากมาตรฐานเมทาดาต้าที่ สพร. กำหนด)

รายชื่อและเมทาดาต้าของชุดข้อมูลเหล่านี้จะถูกนำไปสร้างเป็นรายการข้อมูลและนำไปเผยแพร่ในรูปแบบที่ง่ายแก่การสืบค้นและใช้งาน เช่นในลักษณะระบบพอร์ทัลบริการ นามานุกรมข้อมูล (Directory Services) ที่มีกล่องให้ใส่คำสืบค้น (Search Box) ทั้งนี้ คำอธิบายข้อมูลของชุดข้อมูลแต่ละชุดต้องมีการปรับปรุงให้มีความเป็นปัจจุบันอย่างสม่ำเสมอตามความเหมาะสมของข้อมูล

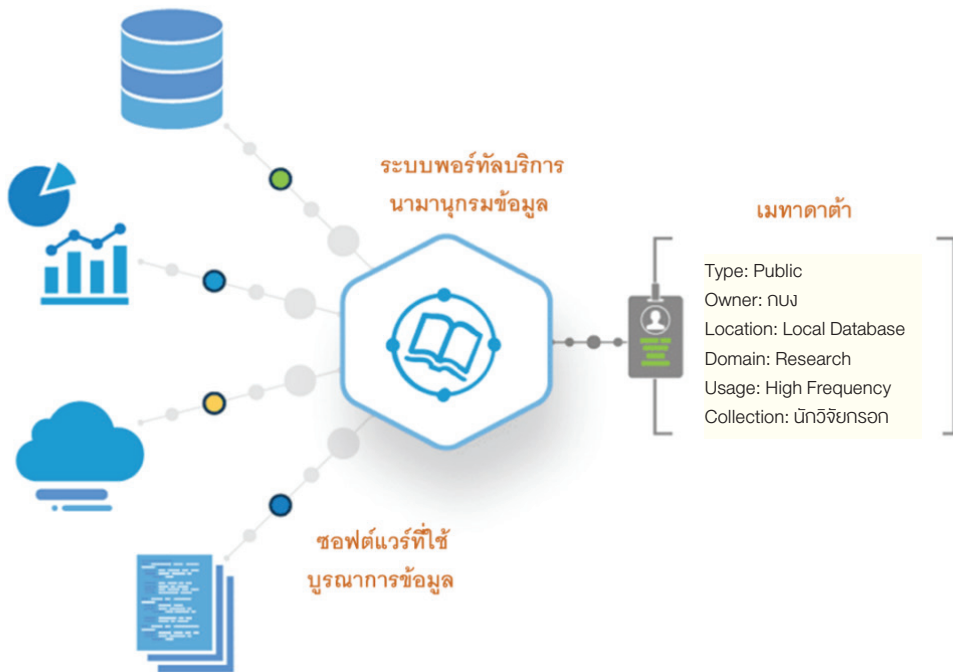
ว.1.2 โครงสร้างสำคัญของระบบสารสนเทศรายการข้อมูล

ส่วนประกอบหลักของระบบสารสนเทศรายการข้อมูลดังแสดงในภาพที่ ก.1.2 โดยมีองค์ประกอบสำคัญอยู่ทั้งหมด 3 ส่วน ได้แก่

1.) โปรแกรมเพื่อเชื่อมโยงพอร์ทัลบริการกับฐานข้อมูล ซึ่งจะเป็นซอฟต์แวร์ตระกูลที่ใช้ในการบูรณาการข้อมูล มีหน้าที่เชื่อมต่อแหล่งข้อมูลและรวบรวมข้อมูลทั้งหมดภายในหน่วยงาน

2.) ระบบพอร์ทัลบริการนามานุกรมข้อมูล สำหรับให้ผู้ใช้เข้าสืบค้นรายชื่อและวิธีร้องขอชุดข้อมูลที่มีสิทธิเข้าถึงได้

3.) เมทาดาต้าของชุดข้อมูลแต่ละชุด ซึ่งบอกถึงลักษณะของชุดข้อมูลทั้งในทางเทคนิคและการใช้งาน รวมไปถึงพจนานุกรมข้อมูล (Data Dictionary) ที่บอกรายละเอียดเชิงลึกของข้อมูลชิ้นต่าง ๆ ในแต่ละชุดข้อมูล



ภาพที่ ข.1.2 โครงสร้างสำคัญในระบบสารสนเทศรายการข้อมูลภาครัฐ

หน่วยงานสามารถเลือกที่จะสร้างระบบสารสนเทศรายการข้อมูลได้โดยการจัดซื้อจากเอกชนหรือสร้างต่อยอดจากซอฟต์แวร์ที่มีลักษณะที่เป็นโอเพ่นซอร์ส (Open Source) โดยปกติแล้วชุดซอฟต์แวร์สำหรับบริหารจัดการระบบข้อมูลขนาดใหญ่ที่จัดจำหน่ายในท้องตลาดปัจจุบันมักมีความสามารถในการสร้างระบบสารสนเทศรายการข้อมูลรวมอยู่ด้วย โดยจะมีทั้งระบบที่จัดการบูรณาการข้อมูลและหน้าจอสืบค้น รวมอยู่ในระบบนิเวศน์ (Ecosystems) เดียวกันเพื่อความสะดวกในการใช้งาน อย่างไรก็ตาม หากหน่วยงานภาครัฐเลือกใช้ซอฟต์แวร์ที่เป็นโอเพ่นซอร์สซึ่งโดยปกติมีศักยภาพที่จำกัดกว่า หน่วยงานอาจจะต้องบูรณาการซอฟต์แวร์ร่วมกันมากกว่า 1 ระบบ และอาจมีความจำเป็นต้องพัฒนาเพิ่มเติมบางส่วนจึงจะสามารถใช้งานได้สมบูรณ์

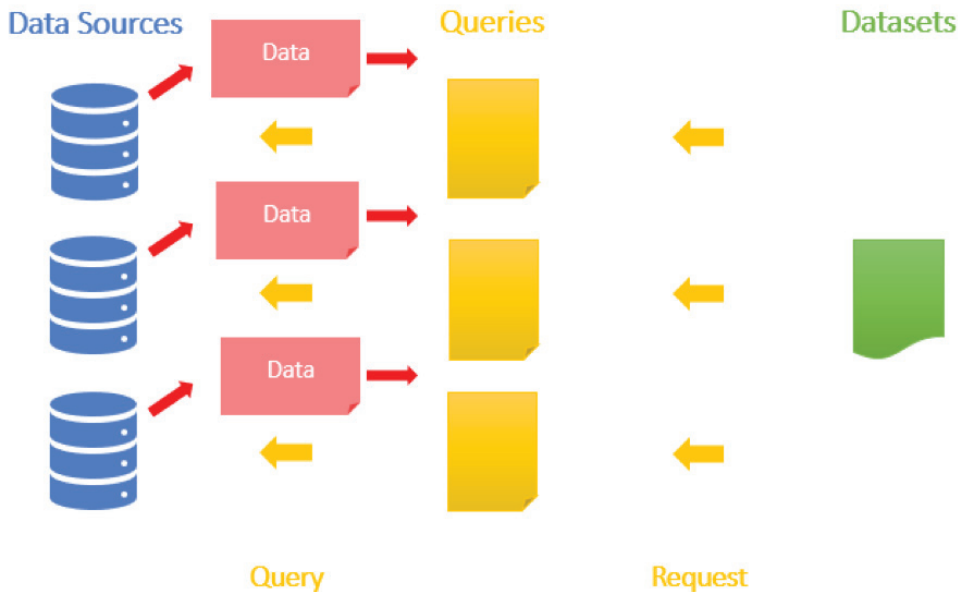
v.1.3 ซอฟต์แวร์สำหรับบูรณาการข้อมูล (Data Integration)

ซอฟต์แวร์สำหรับบูรณาการข้อมูลจะทำงานอยู่เบื้องหลัง (Backend) ของระบบสารสนเทศรายการข้อมูล โดยจะเชื่อมต่อและรวบรวมข้อมูลจากหลายแหล่งและหลายรูปแบบ หน้าหลักของซอฟต์แวร์นี้จะเป็นการเป็นตัวกลางในการเชื่อมโยงพอร์ทัลกับฐานข้อมูล มีคุณลักษณะดังต่อไปนี้

- 1.) ผู้ดูแลระบบสามารถนำข้อมูลจากแหล่งต่าง ๆ ที่โปรแกรมเชื่อมต่อมาจัดสร้างชุดข้อมูลใหม่ได้อย่างอิสระ ชุดข้อมูลแต่ละชุดสามารถประกอบด้วยข้อมูลจากแหล่งที่แตกต่างกันได้
- 2.) ซอฟต์แวร์ควรมีพื้นที่หลายส่วนสำหรับจัดเก็บชุดข้อมูลที่ถูกสร้างขึ้น โดยแบ่งเป็นสัดส่วนเพื่อช่วยให้ง่ายแก่การจัดระเบียบชุดข้อมูลตามระดับชั้นความลับและจุดประสงค์ในการใช้ข้อมูล
- 3.) เนื่องจากตัวซอฟต์แวร์มีหน้าที่เพียงดึงข้อมูลจากแหล่งข้อมูลมาจัดเตรียม (Preprocess) และส่งต่อให้ผู้ร้องขอข้อมูล ซอฟต์แวร์นี้จึงไม่มีความจำเป็นต้องทำหน้าที่เป็นแหล่งเก็บข้อมูลกลางของหน่วยงาน ชุดข้อมูลที่ถูกสร้างสามารถอยู่ในรูปแบบของคำร้องสำหรับฐานข้อมูล (Database Query) ซึ่งจะถูกระมวลผลเพื่อดึงข้อมูลจากแหล่งต่าง ๆ ตามความต้องการของผู้ใช้งาน

4.) มีการให้บริการ API (Application Programming Interface) สำหรับให้ผู้ใช้ดึงข้อมูลของชุดข้อมูลแต่ละชุด และสามารถทำหน้าที่เป็นตัวเชื่อมระหว่างข้อมูลจริงกับแอปพลิเคชันหรือผู้ร้องขอข้อมูล

เมื่อระบบได้รับคำร้องขอชุดข้อมูลจากผู้ใช้ ซอฟต์แวร์จะทำการส่งคำร้องขอข้อมูล (Query) ไปที่แหล่งข้อมูลต่าง ๆ เพื่อดึงข้อมูลส่วนที่ต้องการมาให้แก่ผู้ร้องขอแบบเรียลไทม์ โดยข้อมูลที่ได้จะมีความเป็นปัจจุบัน ดังแสดงในภาพที่ ก.1.3 ทั้งนี้ซอฟต์แวร์อาจจะสามารถทำการเก็บข้อมูลชั่วคราว (Caching) และปรับปรุงความทันสมัยเป็นระยะ ๆ เพื่อช่วยลดความคับคั่ง (Congestion) ในการดึงข้อมูลพร้อมทั้งเพิ่มประสิทธิภาพในการทำงานได้อีกด้วย



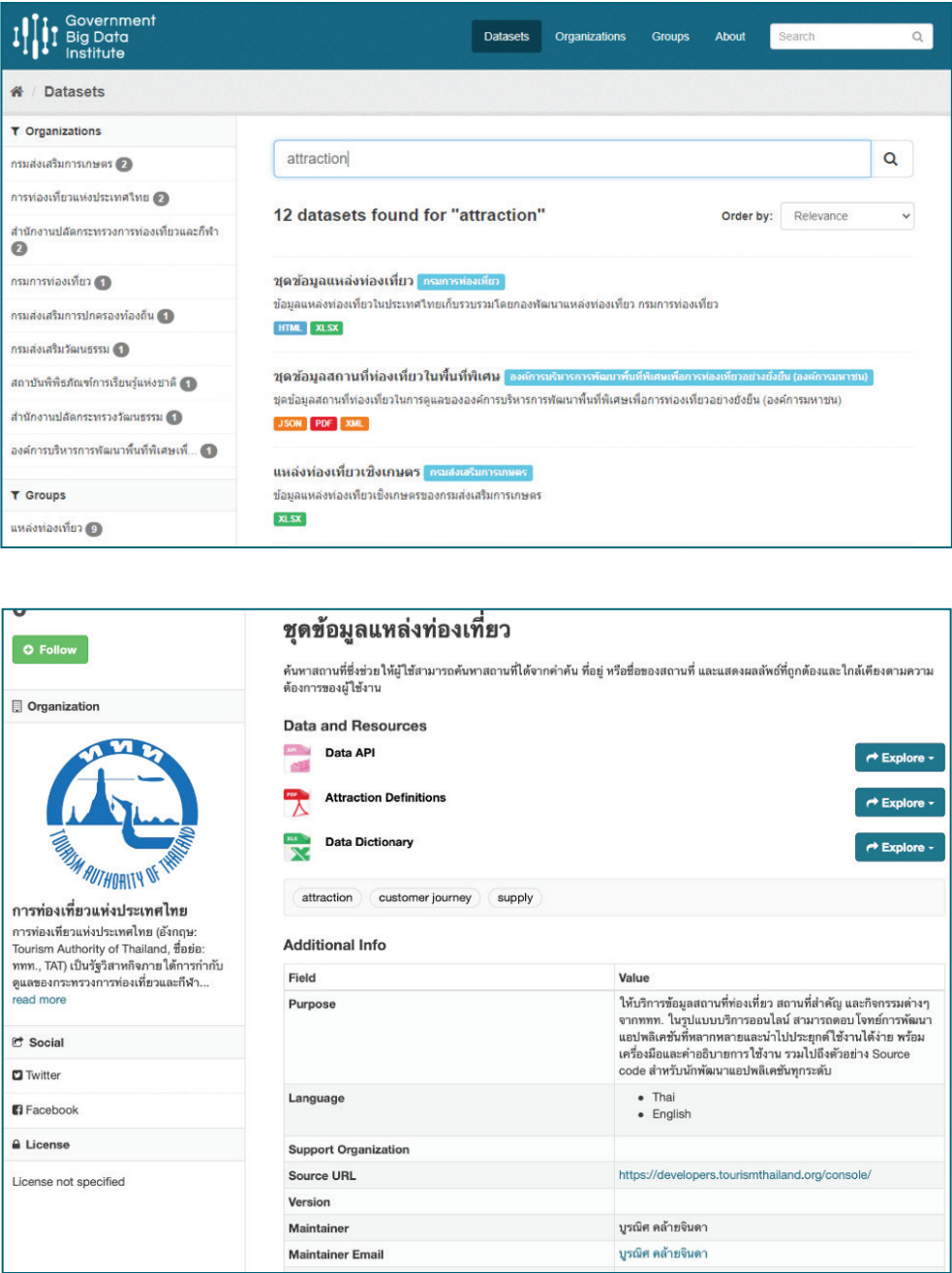
ภาพที่ ข.1.3 การส่งคำร้องขอข้อมูลจากชุดข้อมูล
โดยโปรแกรมที่เชื่อมโยงพอร์ตกับฐานข้อมูล

v.1.4 ระบบพอร์ทัลบริการนามานุกรมข้อมูล (Directory Services)

ระบบพอร์ทัลบริการนามานุกรมข้อมูลจะทำงานเป็นเบื้องหน้า (Frontend) ของระบบสารสนเทศรายการข้อมูลที่มีปฏิสัมพันธ์กับผู้ใช้โดยตรง ระบบพอร์ทัลมีหน้าที่ในการให้บริการสืบค้นข้อมูลภายในองค์กรและอำนวยความสะดวกในการดึงข้อมูลมาใช้งาน ระบบควรมีคุณสมบัติดังต่อไปนี้

- 1.) หน่วยงานสามารถจัดเก็บรายการชุดข้อมูลทั้งหมดภายในองค์กรได้อย่างสะดวกและครบถ้วน
- 2.) ระบบสามารถแสดงรายละเอียดเมทาดาทาของแต่ละชุดข้อมูล รวมไปถึงตัวอย่างข้อมูลให้ผู้ใช้งานทราบได้ โดยผู้สร้างชุดข้อมูลสามารถกำหนดเมทาดาทาของชุดข้อมูลให้สอดคล้องกับระเบียบแบบแผนได้ด้วยตนเอง
- 3.) หน่วยงานสามารถจัดหมวดหมู่ของรายชื่อชุดข้อมูลได้ตามความต้องการ
- 4.) ผู้ใช้งานสามารถสืบค้นชุดข้อมูลได้ด้วยตัวกรอง (Filter) ในมิติต่าง ๆ เช่น รูปแบบของข้อมูล ชื่อของแผนก/หน่วยงานที่เป็นเจ้าของ คำสำคัญ และอื่น ๆ แล้วแต่ลักษณะการใช้งาน
- 5.) ผู้สร้างชุดข้อมูลสามารถกำหนดคำค้น (Tag) ของชุดข้อมูลแต่ละชุดเพื่อช่วยในการสืบค้นชุดข้อมูลได้ โดยคำค้นอาจถูกสร้างโดยอัตโนมัติด้วยความสามารถของซอฟต์แวร์หรือผู้สร้างเป็นผู้กำหนดไว้เอง
- 6.) ผู้ดูแลระบบสามารถควบคุมสิทธิการใช้งานชุดข้อมูลต่าง ๆ ภายในระบบด้วยการกำหนดชั้นความลับของชุดข้อมูลได้ สามารถยืนยันตัวตนของผู้ใช้ (Authentication) ด้วยการล็อกอิน และสามารถจำกัดขอบเขตในการมองเห็นและเข้าถึงชุดข้อมูลต่าง ๆ ของผู้ใช้ได้
- 7.) ผู้ใช้สามารถดึงรายละเอียดชุดข้อมูลต่าง ๆ ได้อย่างสะดวก เช่น มี API สำหรับใช้งาน

หากหน่วยงานเลือกใช้ซอฟต์แวร์โอเพ่นซอร์สในการสร้างระบบ ซอฟต์แวร์ที่เลือกควรมีฐานผู้ใช้ที่มีขนาดใหญ่ มีการปรับปรุงโปรแกรมอย่างต่อเนื่อง และมีการพัฒนาปลั๊กอิน (Plugins) เสริมจากกลุ่มผู้ใช้งาน ทั้งนี้เนื่องจากโปรแกรมโอเพ่นซอร์สส่วนมากไม่มีการสนับสนุนโดยตรงจากทางผู้สร้าง การเลือกโปรแกรมที่มีจำนวนผู้ใช้สูงจะทำให้หน่วยงานมีทรัพยากรในการแก้ปัญหาที่อาจเกิดขึ้นจากการใช้งานมากขึ้น นอกจากนี้ การที่โปรแกรมมีปลั๊กอินเสริม ก็จะช่วยลดภาระของหน่วยงานในการพัฒนาความสามารถของระบบเพิ่มเติมในอนาคตอีกทางหนึ่ง



ภาพที่ ข.1.4 ตัวอย่างหน้าจอสืบค้น (บน) และแสดงรายละเอียดชุดข้อมูล (ล่าง) ของระบบพอร์ทัลบริการนามานุกรมข้อมูล

หมายเหตุ: หน้าจอสืบค้นและแสดงรายละเอียดชุดข้อมูลในรูปที่ ก.1.4 นั้นถูกพัฒนาจากซอฟต์แวร์โอเพ่นซอร์สที่มีชื่อว่า “ckan” ซึ่งเป็นเพียงตัวอย่างหนึ่งของซอฟต์แวร์ที่สามารถนำมาช่วยสร้างระบบพอร์ทัลบริการนามุกรมข้อมูลได้ โดยซอฟต์แวร์นี้ถูกใช้งานเป็นวงกว้างในการทำระบบเพื่อเผยแพร่ข้อมูลของภาครัฐในนานาประเทศเช่น ข้อมูลเปิดของรัฐบาลสหรัฐอเมริกา (Data.gov) และของสหภาพยุโรป (Europeandataportal.eu)

v.1.5 เมทาดาต้า (Metadata)

คำอธิบายข้อมูลหรือเมทาดาต้า มีหน้าที่ช่วยให้ข้อมูลเกี่ยวกับชุดข้อมูลต่าง ๆ เพิ่มเติม ทั้งในด้านการบริหารจัดการชุดข้อมูลและลักษณะของข้อมูลภายใน รายละเอียดเหล่านี้สามารถใช้เป็นข้อมูลเบื้องต้นเพื่อช่วยเหลือนักวิเคราะห์ข้อมูล นักวิทยาศาสตร์ข้อมูล และผู้ใช้ข้อมูล ในการวางแผนและใช้ประกอบการตัดสินใจเบื้องต้นในการใช้ประโยชน์ข้อมูล โดยเนื้อหาส่วนมากของเมทาดาต้าจะเป็นการอธิบายภาพรวมของทั้งชุดข้อมูล เช่น ชื่อเจ้าของชุดข้อมูล ความถี่ในการปรับปรุงข้อมูล ระดับชั้นความลับ เป็นต้น

ทั้งนี้สำนักงานพัฒนารัฐบาลดิจิทัลร่วมกับสำนักงานสถิติแห่งชาติและสถาบันส่งเสริมการวิเคราะห์และบริหารข้อมูลขนาดใหญ่ภาครัฐได้ออกเอกสารเพื่อกำหนดมาตรฐานขั้นต่ำให้หน่วยงานนำไปใช้จัดทำเมทาดาต้าภายในหน่วยงาน สำหรับการจัดทำรายการข้อมูล (Data Catalog) ของหน่วยงานและของประเทศเพื่อการเชื่อมโยงและแลกเปลี่ยนข้อมูลภาครัฐในอนาคต โดยแนะนำให้หน่วยงานภาครัฐใช้เมทาดาตาดังแสดงในตารางที่ ข.1.1

ตารางที่ ข.1.1 เมทาดาต้าที่สำคัญต่อการใช้ประโยชน์

No.	ชื่อรายการ ไทย	ชื่อทางเทคนิค	คำอธิบาย	ตัวเลือก / รูปแบบ
1	ประเภทข้อมูล	data_type	ชุดข้อมูลนี้เป็นข้อมูล ประเภทใด	Code (Character 1 digits (0-9)) หรือ CHAR(1)
2	ชื่อชุดข้อมูล	title	ชื่อของชุดข้อมูลที่กำหนด โดยองค์กรที่รับผิดชอบ ข้อมูล	Text (150 Characters)
3	องค์กร	data_owner	ชื่อองค์กรที่รับผิดชอบ ข้อมูล	Code (Character 6 digits (0-9)) หรือ CHAR(6) (ดูรายละเอียด ภาคผนวก ข.2)

No.	ชื่อรายการ ไทย	ชื่อทางเทคนิค	คำอธิบาย	ตัวเลือก / รูปแบบ
4	ชื่อผู้ติดต่อ	contact_person	ชื่อกอง สำนัก ฝ่าย หรือบุคคลที่ได้รับการ มอบหมายให้รับผิดชอบ ข้อมูล	Text (150 Characters) หรือ VARCHAR (150)
5	อีเมลผู้ติดต่อ	contact_email	อีเมลกอง สำนัก ฝ่าย หรือบุคคลที่ได้รับการ มอบหมายให้รับผิดชอบ ข้อมูล	Text (50 Charac- ters) หรือ VARCHAR (50)
6	คำสำคัญ	tag_string	หัวข้อ คำ วลี หรือแท็ก (tag) ที่ใช้ระบุคำสำคัญ	Text แยกแต่ละ keywords ด้วย “,”(comma) (200 Characters) หรือ VARCHAR (200)
7	รายละเอียด	notes	คำอธิบายรายละเอียด ที่สำคัญของชุดข้อมูล อย่างสั้น เช่น คำนิยาม ชุดข้อมูลเกี่ยวกับอะไร มีวิธีการจัดเก็บแบบใด กลุ่มเป้าหมายผู้ใช้งาน ข้อมูล เป็นใคร	Text (1,000 Characters) หรือ VARCHAR (1,000)

No.	ชื่อรายการ ไทย	ชื่อทางเทคนิค	คำอธิบาย	ตัวเลือก / รูปแบบ
8	วัตถุประสงค์	objective	อธิบายที่มาและวัตถุประสงค์ของการจัดทำชุดข้อมูล เช่น กฎหมาย ภารกิจ โครงการตามแผน ยุทธศาสตร์ และเพื่อใช้ในการวิเคราะห์หรือตอบโจทย์ในประเด็น ยุทธศาสตร์ในเรื่องใดที่ผู้ต้องการ	Code (Character 2 digits (0-9)) หรือ CHAR(2) (ดูรายละเอียดภาคผนวก ข.2)
9.1	หน่วยความถี่ของการปรับปรุงข้อมูล	update_ frequency_unit	สำหรับข้อมูลทะเบียน ข้อมูลระดับย่อย และ ข้อมูลภูมิสารสนเทศเชิงพื้นที่ : ความถี่ที่ข้อมูลในระบบคลังข้อมูลถูกปรับปรุง/เพิ่ม หรือเปลี่ยนแปลงสำหรับทั่วไป และสถิติทางการ : ความถี่ในการเผยแพร่ต่อผู้ใช้ข้อมูล	Code (Character 1 digits (A-Z)) หรือ CHAR (1) (ดูรายละเอียดภาคผนวก ข.2)
9.2	ค่าความถี่ของการปรับปรุงข้อมูล	update_ frequency_ interval	ใช้คุณสมบัตินี้ประกอบกับหน่วยความถี่ในการปรับปรุงข้อมูล ตัวอย่างเช่น ถ้าชุดข้อมูลมีการปรับปรุงทุก ๆ 2 ปี ท่านสามารถใส่ “2” สำหรับค่าความถี่ และ “รายปี” สำหรับหน่วยความถี่	Number หรือ เว้นว่างไว้ หรือ INTEGER (2)

No.	ชื่อรายการ ไทย	ชื่อทางเทคนิค	คำอธิบาย	ตัวเลือก / รูปแบบ
10	ขอบเขตเชิง ภูมิศาสตร์หรือ เชิงพื้นที่	geo_coverage	สำหรับข้อมูลทะเบียน ข้อมูลระดับย่อย และ ข้อมูลภูมิสารสนเทศเชิง พื้นที่: มิติการจัดจำแนก ข้อมูลพื้นที่ในระดับย่อย ที่สุดในการจัดเก็บข้อมูล สำหรับสถิติทั่วไปและ สถิติทางการ: มิติการจัด จำแนกข้อมูลพื้นที่ใน ระดับย่อยที่สุดในการ นำเสนอข้อมูล	Code (Character 2 digits (0-9)) หรือ CHAR(2) (ดูรายละเอียด ภาคผนวก ข.2)
11	แหล่งที่มา	data_source	แหล่งที่มาของข้อมูลที่น่า มาจัดทำชุดข้อมูล พร้อม หน่วยงานที่จัดทำ เช่น สำรวจภาวะการทำงาน ของประชากร (สำนักงาน สถิติแห่งชาติ) ฐานข้อมูล ทะเบียนราษฎร์ (กรม การปกครอง)	Text (200 Characters) หรือ VARCHAR (200)
12	รูปแบบการ เก็บข้อมูล	data_format	รูปแบบของการจัดเก็บ ข้อมูล	Code (Character 2 digits (0-9)) หรือ CHAR(2) (ดูรายละเอียด ภาคผนวก ข.2)

No.	ชื่อรายการ ไทย	ชื่อทางเทคนิค	คำอธิบาย	ตัวเลือก / รูปแบบ
13	หมวดหมู่ข้อมูลตาม ธรรมาภิบาล ข้อมูลภาครัฐ	data_category	หมวดหมู่ข้อมูลตาม ธรรมาภิบาลข้อมูลภาครัฐ	Code (Character 1 digits (1-4)) หรือ CHAR (1) (ดูรายละเอียด ภาคผนวก ข.2)
14	สัญญาอนุญาต ให้ใช้ข้อมูล	right_of_usage	สัญญาอนุญาตให้ใช้ข้อมูล ต้องสอดคล้องกับหมวด หมู่ข้อมูลตามธรรมาภิบาล ข้อมูลภาครัฐ	Code (Character 1 digits (0-9)) หรือ CHAR (1) (ดูรายละเอียด ภาคผนวก ข.2)

พจนานุกรมข้อมูล (Data Dictionary) เป็นส่วนหนึ่งของเมทาดาทาที่มีหน้าที่อธิบายข้อมูลภายในชุดข้อมูลอย่างละเอียดเป็นรายตัวแปร พร้อมทั้งอาจมีตัวอย่างของข้อมูลภายในชุดประกอบอยู่ด้วย รูปที่ ข.1.5 แสดงตัวอย่างของเมทาดาทาและพจนานุกรมข้อมูลของชุดข้อมูล

ข้อมูลเพิ่มเติม			
ฟิลด์	คำ		
วัตถุประสงค์เบื้องต้นในการจัดเก็บ	ให้บริการข้อมูลสถานที่ท่องเที่ยวในพื้นที่พิเศษของ อพท. ซึ่งสามารถสร้างการรับรู้ให้กับประชาชน และนักท่องเที่ยวได้ และเป็นการจัดเก็บข้อมูลเพื่อบูรณาการร่วมกับหน่วยงานอื่นๆ ได้ในอนาคต		
ภาษาที่ใช้	ไทย		
ผู้สนับสนุนหรือผู้ร่วมดำเนินการ			
Source URL	http://203.151.225.116:8080/services.asmx		
เวอร์ชัน			
ผู้ดูแลชุดข้อมูล	ศิริรณัท ทิพยมงคลกุล		

ชื่อฟิลด์ข้อมูล Attribute	คำอธิบายฟิลด์ Description	ประเภทข้อมูล Data Type	รูปแบบข้อมูล Data Format
gender	เพศของผู้ใช้	VARCHAR	"male", "female"
income	เงินได้ของผู้ใช้	FLOAT	numerical value
spending	ปริมาณการใช้จ่าย	FLOAT	numerical value
arrival_date	วันที่ผู้ใช้เดินทางถึง ปี พ.ศ.	DATE	yyyy-mm-dd

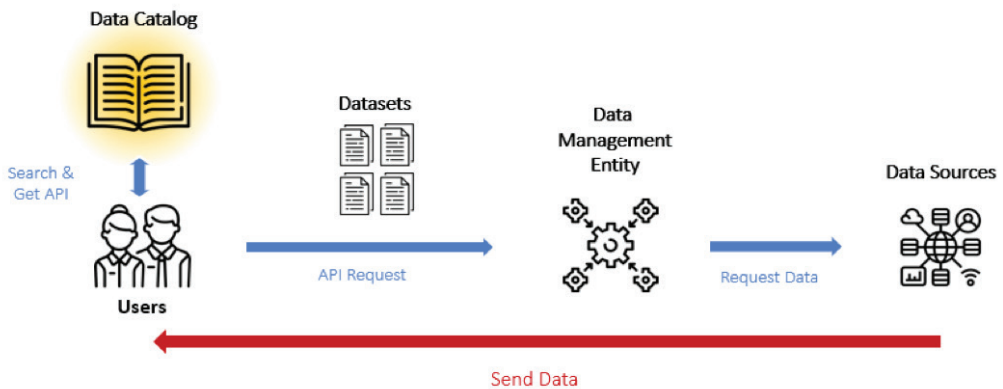
ภาพที่ ข.1.5 ตัวอย่างของเมทาดาทาของชุดข้อมูล (บน)
และตัวอย่างของพจนานุกรมข้อมูล (ล่าง)

v.1.6 การใช้ระบบสารสนเทศสายการข้อมูล

สำหรับการสร้างระบบสารสนเทศสายการข้อมูลนั้น นอกจากจะช่วยให้เจ้าหน้าที่ภายในหน่วยงานทราบถึงข้อมูลที่หน่วยงานครอบครองและแหล่งที่เก็บของข้อมูลเหล่านั้นแล้ว จุดประสงค์หลักอีกข้อหนึ่งคือการให้ความสะดวกผู้ใช้ในการค้นหาข้อมูลอย่างมีประสิทธิภาพ โดยระบบจะสามารถช่วยตอบผู้ได้ว่า ข้อมูลที่ต้องการนั้นมีอยู่ภายในหน่วยงานหรือไม่ และสามารถเข้าถึงได้อย่างไร การใช้ระบบสารสนเทศสายการข้อมูลของผู้ใช้ข้อมูลจะมีขั้นตอนดังต่อไปนี้

- 1) ผู้ใช้ที่ต้องการข้อมูลทำการเข้าสู่หน้าจอแสดงผลของระบบพอร์ทัลบริการ นามานุกรมข้อมูลของหน่วยงาน ซึ่งอาจอยู่ในรูปแบบของหน้าเว็บไซต์
- 2) ผู้ใช้ทำการค้นหาข้อมูลที่เกี่ยวข้องกับโจทย์ของตนผ่านทางกล่องใส่คำค้น โดยใช้กลุ่มคำที่มีความสำคัญกับข้อมูลที่ต้องการค้นหา ถ้าหากคำค้นที่ใช้ตรงกับชุดข้อมูลใดชุดข้อมูลเหล่านั้นจะถูกนำมาแสดงผลบนหน้าจอให้ผู้ได้พิจารณา ในลักษณะเดียวกับการสืบค้นบนหน้าจอของ Google
- 3) จากรายชื่อชุดข้อมูลที่ค้นเจอในขั้นตอนที่ 2) ผู้ใช้ระบบจะทำการเลือกชุดข้อมูล และตรวจสอบรายละเอียดของชุดข้อมูลเหล่านั้นเพื่อความแน่ใจว่าเป็นชุดข้อมูลที่ตนต้องการ โดยจะทำการสำรวจคุณสมบัติและพจนานุกรมข้อมูลของชุดข้อมูลนั้น ถ้าหากข้อมูลในชุดสามารถตอบโจทย์ได้ ผู้ใช้จะทำการร้องขอข้อมูลตามช่องทางที่ระบุไว้ในเมทาดาต้า

หลังจากผู้ใช้นั้นยื่นคำขอข้อมูลตามสิทธิระบบจะทำการประมวลผลคำร้องเพื่อดึงข้อมูลและส่งกลับไปหาผู้ร้องขอที่มีสิทธิในการเข้าถึงข้อมูลแบบเรียลไทม์ ผู้ใช้จึงได้รับข้อมูลที่มีความเป็นปัจจุบันมากที่สุด



ภาพที่ ข.1.6 การใช้ระบบสารสนเทศสายการข้อมูล

ภาคผนวก v.2

แนวทางปฏิบัติทางเทคโนโลยีสำหรับการแปลงข้อมูลให้อยู่ในรูปแบบพร้อมใช้งาน

เอกสารฉบับนี้อธิบายแนวทางปฏิบัติในการแปลงข้อมูลในรูปแบบ CSV (ย่อมาจาก Comma-Separated Values) ให้อยู่ในรูปแบบ XML (ย่อมาจาก Extensible Markup Language) และ JSON (ย่อมาจาก JavaScript Object Notation) ที่พร้อมสำหรับการนำไปพัฒนาต่อยอดเป็นรูปธรรม เช่น พัฒนาเป็นโปรแกรมพร้อมใช้งานผ่าน Web/Mobile Applications หรือการวิเคราะห์ข้อมูล Data Analytics โดยเอกสารฉบับนี้มีที่มาจากผลการสำรวจที่พบว่าข้อมูลในหน่วยงานภาครัฐส่วนใหญ่ยังถูกเก็บอยู่ในรูปแบบที่หลากหลาย แต่สามารถแปลงออกมา (Export) ให้ในรูปแบบ CSV ได้เป็นอย่างน้อย ดังนั้นแนวทางปฏิบัติจึงตั้งต้นด้วยการที่ข้อมูลในรูปแบบ CSV ถูกแปลงให้อยู่ในรูปแบบที่มีมาตรฐานพร้อมนำไปใช้งานมากขึ้น ได้แก่ XML หรือ JSON ซึ่งมีการใช้งานอย่างแพร่หลายทั้งในและต่างประเทศ สามารถใช้งานได้สะดวก และที่สำคัญคือ มีขั้นตอนการแปลงข้อมูลที่ชัดเจนพร้อมสำหรับการนำไปปฏิบัติใช้และเชื่อถือได้

วิธีการแปลงข้อมูลเป็น JSON และ XML ในเอกสารฉบับนี้ แบ่งออกเป็น 2 วิธี ได้แก่

1. การแปลงด้วยเครื่องมือออนไลน์ : เหมาะสำหรับข้อมูลที่ไม่ต้องการความปลอดภัยสูง และจำนวนไม่มากนัก สามารถแปลงข้อมูลได้อย่างรวดเร็วทั้งยังไม่ต้องติดตั้งโปรแกรมในเครื่อง

2. การแปลงด้วยเครื่องมือที่ติดตั้งบนเครื่อง : เหมาะสำหรับข้อมูลที่ต้องการความปลอดภัยสูงมีขั้นตอนที่เยาะกว่าการใช้เครื่องมือออนไลน์ เครื่องมือที่เอกสารฉบับนี้แนะนำคือ Openrefine ซึ่งเป็นโปรแกรม Open Source ที่ได้รับการยอมรับจาก European Open Data Portal <https://www.europeandataportal.eu> ผู้ใช้สามารถดาวน์โหลดซอฟต์แวร์ Openrefine สำหรับระบบปฏิบัติการต่าง ๆ ได้จาก <http://openrefine.org/download.html> จากนั้นทำการติดตั้งตามขั้นตอนที่ระบุในเว็บไซต์ดังกล่าว

v.2.1. วิธีการแปลงข้อมูลจาก CSV เป็น JSON

ข.2.1.1 การแปลงข้อมูลเครื่องมือออนไลน์ สามารถเลือกได้ 2 วิธีดังนี้

- กรณีข้อมูลถูกเก็บอยู่ในเครื่องอย่างเดียว สามารถแปลงข้อมูลผ่าน URL: <https://www.csvjson.com/csv2json> โดยผู้ใช้งานต้องอัปโหลดไฟล์ CSV โดยคลิก “Select a file..” จากนั้นคลิกปุ่ม “Convert” ระบบจะแปลงข้อมูลในรูปแบบ CSV ไปยัง JSON โดยอัตโนมัติ จากนั้นคลิก “Download” เพื่อเก็บข้อมูลในรูปแบบ JSON ไปใช้ในขั้นตอนต่อไป
- กรณีมีข้อมูลที่ถูกเก็บอยู่บนคลาวด์ สามารถแปลงข้อมูลผ่าน URL: <https://codebeautify.org/csv-to-json-converter> โดยผู้ใช้งานต้องอัปโหลดไฟล์ CSV โดยคลิก “Browse” หรือคลิก “Load Url” ในกรณีที่ไฟล์อยู่บนคลาวด์ จากนั้นคลิกปุ่ม “CSV to JSON” ระบบจะแปลงไฟล์ในรูปแบบ CSV ไปยัง JSON โดยอัตโนมัติ จากนั้นผู้ใช้งานคลิก “Download” เพื่อเก็บไฟล์ JSON ไปใช้ในขั้นตอนต่อไป

ข.2.1.2 การแปลงข้อมูลด้วยเครื่องมือที่ติดตั้งบนเครื่อง สามารถใช้ Openrefine 3.0 ตามขั้นตอนดังนี้

- ดับเบิลคลิกเพื่อเปิด Openrefine

OpenRefine A power tool for working with messy data. New version! Download OpenRefine V3.1 Final Release now.

Project name: diabetes.csv Tags Create Project »

« Start Over Configure Parsing Options

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
1.	1	85	66	29	0	26.6	0.351	31	0
2.	8	183	64	0	0	23.3	0.672	32	1
3.	1	89	66	23	94	28.1	0.167	21	0
4.	0	137	40	35	168	43.1	2.288	33	1
5.	5	116	74	0	0	25.6	0.201	30	0
6.	3	78	50	32	88	31	0.248	26	1
7.	10	115	0	0	0	35.3	0.134	29	0
8.	2	197	70	45	543	30.5	0.158	53	1
9.	8	125	96	0	0	0	0.232	54	1
10.	4	110	92	0	0	37.6	0.191	30	0
11.	10	168	74	0	0	38	0.537	34	1
12.	10	139	80	0	0	27.1	1.441	57	0
13.	1	189	60	23	846	30.1	0.398	59	1
14.	5	166	72	19	175	25.8	0.587	51	1
15.	7	100	0	0	0	30	0.484	32	1
16.	0	118	84	47	230	45.8	0.551	31	1

Parse data as Character encoding Update Preview

CSV / TSV / separator-based files

Columns are separated by
☒ commas (CSV)
☐ tabs (TSV)
☐ custom: ,
 Escape special characters with \

☐ Ignore first 0 line(s) at beginning of file
☒ Parse next 1 line(s) as column headers
☐ Discard initial 0 row(s) of data
☐ Load at most 0 row(s) of data
☒ Use character * to enclose cells containing column separators

☐ Parse cell text into numbers, dates, ...
☒ Store blank rows
☒ Store blank cells as nulls
☐ Store file source (file names, URLs) in each row

Version 3.0 (TRUNK)

Preferences
Help
About

• เลือกไฟล์ CSV ที่ต้องการแปลง

OpenRefine A power tool for working with messy data. New version! Download OpenRefine V3.1 Final Release now.

Create Project Open Project Import Project Language Settings

Create a project by importing data. What kinds of data files can I import?
TSV, CSV, *SV, Excel (.xls and .xlsx), JSON, XML, RDF as XML, and Google Data documents are all supported. Support for other formats can be added with OpenRefine extensions.

Get data from Locate one or more files on your computer to upload:

This Computer Choose Files No file chosen

Web Addresses (URLs) Next »

Clipboard

Data Package (JSON URL)

Database

Google Data

Version 3.0 [TRUNK]

Preferences Help About

• คลิก Create project

OpenRefine A power tool for working with messy data. New version! Download OpenRefine V3.1 Final Release now.

Create Project « Start Over Configure Parsing Options Project name diabetes csv Tags Create Project »

Open Project Import Project Language Settings

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
1.	1	85	66	29	0	26.6	0.351	31	0
2.	8	183	64	0	0	23.3	0.672	32	1
3.	1	89	66	23	94	28.1	0.167	21	0
4.	0	137	40	35	168	43.1	2.288	33	1
5.	5	116	74	0	0	25.6	0.201	30	0
6.	3	78	50	32	88	31	0.248	26	1
7.	10	115	0	0	0	35.3	0.134	29	0
8.	2	197	70	45	543	30.5	0.158	53	1
9.	8	125	96	0	0	0	0.232	54	1
10.	4	110	92	0	0	37.6	0.191	30	0
11.	10	168	74	0	0	38	0.537	34	1
12.	10	139	80	0	0	27.1	1.441	57	0
13.	1	189	60	23	846	30.1	0.398	59	1
14.	5	166	72	19	175	25.8	0.587	51	1
15.	7	100	0	0	0	30	0.484	32	1
16.	0	118	84	47	230	45.8	0.551	31	1

Parse data as

Character encoding

CSV / TSV / separator-based files

Columns are separated by
☒ commas (CSV)
☐ tabs (TSV)
☐ custom: ,
 Escape special characters with \

☐ Ignore first 0 line(s) at beginning of file
☒ Parse next 1 line(s) as column headers
☐ Discard initial 0 row(s) of data
☐ Load at most 0 row(s) of data
☒ Use character * to enclose cells containing column separators

☐ Parse cell text into numbers, dates, ...

☒ Store blank rows
☒ Store blank cells as nulls
☐ Store file source (file names, URLs) in each row

Update Preview

Version 3.0 [TRUNK]

Preferences Help About

● คลิ๊ก Templating

The screenshot shows the OpenRefine interface with a dataset named 'diabetes.csv'. The 'Export' menu is open, and 'Templating...' is highlighted. The main table displays 767 rows of data with columns: Pregnancies, Glucose, BloodPressure, SkinThickness, Insulin, BMI, DiabetesPedigreeFunction, Age, and Outcome. A sidebar on the left provides instructions on using facets and filters.

- การเปลี่ยน Template จาก CSV ไปเป็น JSON นั้น ผู้ใช้ต้องทำการแก้ไข Prefix และ Suffix โดย Prefix จะอยู่ในรูปแบบ “ [” และ Suffix จะอยู่ในรูปแบบ “] ” ทั้งนี้ผู้ใช้สามารถใช้ Row template เป็นค่า default ได้

The 'Templating Export' dialog box is shown. It contains fields for 'Prefix', 'Row Template', 'Row Separator', and 'Suffix'. The 'Row Template' field is populated with a JSON template using the `{{jsonize(cells["ColumnName"].value)}}` syntax. The 'Row Separator' field contains a comma. The 'Prefix' field is empty, and the 'Suffix' field contains a closing square bracket. A preview of the resulting JSON output is shown on the right. At the bottom, there are buttons for 'Reset Template', 'Export', and 'Cancel'.

- เมื่อแก้ไขเสร็จคลิ๊ก “Export” เพื่อดาวน์โหลดไฟล์

v.2.2 วิธีการแปลงข้อมูลจาก CSV เป็น XML

ข.2.2.1 การแปลงข้อมูลเครื่องมือออนไลน์ สามารถเลือกได้ 2 วิธี ดังนี้

- แปลงข้อมูลผ่าน URL: <http://convertcsv.com/csv-to-xml.htm> ผู้ใช้งานต้องอัปโหลดไฟล์ CSV โดยคลิก “Choose File” หรือคลิก “Enter URL” ในกรณีที่ไฟล์อยู่บนคลาวด์ จากนั้นคลิกปุ่ม “Convert CSV to XML” ระบบจะแปลงไฟล์ในรูปแบบ CSV ไปยัง XML โดยอัตโนมัติ จากนั้นผู้ใช้งานคลิก “Download” เพื่อเก็บข้อมูลในรูปแบบ XML ไปใช้ในขั้นตอนต่อไป
- แปลงข้อมูลผ่าน URL: <https://codebeautify.org/csv-to-xml-converter> ผู้ใช้งานต้องอัปโหลดไฟล์ CSV โดยคลิก “Browse” หรือคลิก “Load Url” ในกรณีที่ไฟล์อยู่บนคลาวด์ จากนั้นคลิกปุ่ม “CSV to XML” ระบบจะแปลงไฟล์ในรูปแบบ CSV ไปยัง XML โดยอัตโนมัติ จากนั้นผู้ใช้งานคลิก “Download” เพื่อเก็บข้อมูลในรูปแบบ XML ไปใช้ในขั้นตอนต่อไป

ข.2.2.2 การแปลงข้อมูลด้วยเครื่องมือที่ติดตั้งบนเครื่อง สามารถใช้ Openrefine

3.0 ตามขั้นตอนดังนี้

- ดับเบิลคลิกเพื่อเปิด Openrefine
 - เลือกไฟล์ CSV ที่ต้องการแปลง
 - คลิก Create project ตามด้วย Export
 - คลิก Templating
 - เพิ่ม Prefix, Row Template และ Suffix ดังตัวอย่างต่อไปนี้
- ข้อมูลที่ถูกเก็บเป็น XML โดยปกติจะมีรูปแบบดังนี้

```
<row>
  <ชื่อ Attribute1> ค่าของ Attribute1 </ชื่อ Attribute1>
  <ชื่อ Attribute2> ค่าของ Attribute2 </ชื่อ Attribute2>
  <ชื่อ Attribute3> ค่าของ Attribute3 </ชื่อ Attribute3>
</row>
```

ตัวอย่างเช่น

```
<food>
  <name> Belgian Waffles </name>
  <price> $5.95 </price>
  <description> Two of our famous Belgian Waffles with plenty of
real maple syrup
```

```
</description>
```

```
<calories> 650 </calories>
```

```
</food>
```

โดยความหมายของตัวอย่างนี้คือข้อมูลอาหาร (food) ถูกเก็บในรูปแบบของ XML และมี Attribute สี่ชนิดคือ

name มีค่า Belgian Waffles

price มีค่า \$5.95

description มีค่า Two of our famous Belgian Waffles with plenty of real maple syrup

calories มีค่า 650

ในกรณีที่ผู้ใช้งานต้องการแปลงข้อมูล CSV เป็น XML โดยใช้ Openrefine ซึ่งโปรแกรมจะอ่านค่า {{cells [“ชื่อ Attribute1”].value }} เป็นค่าของ Attribute นั้น ๆ ในขั้นตอนการเขียน row template ดังนั้นจากตัวอย่างผู้ใช้งานสามารถเขียน row template ได้ดังนี้

```
<food>
```

```
<name> {{cells[“name”].value}} </name>
```

```
<price> {{cells[“price”].value}} </price>
```

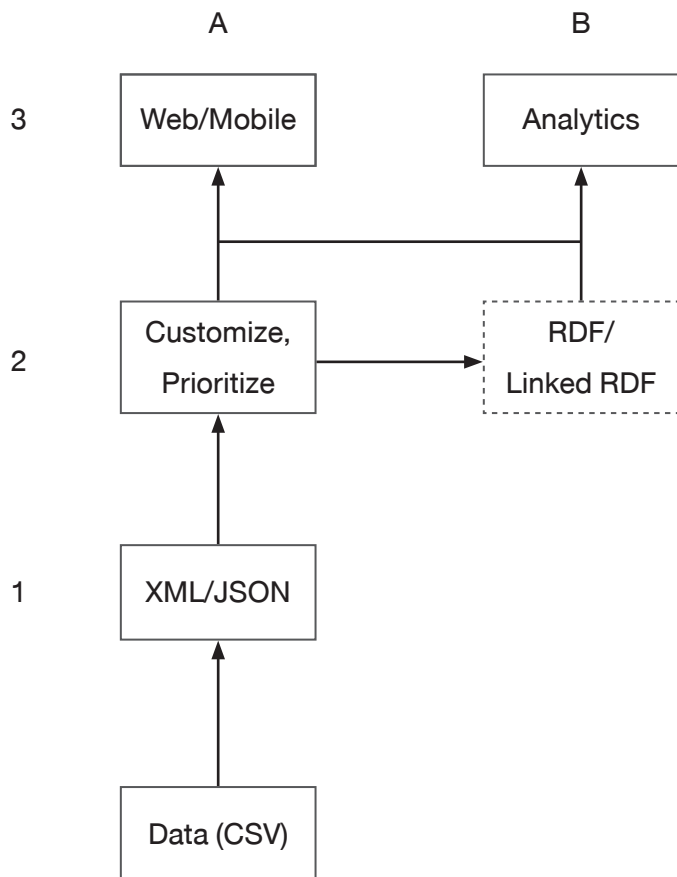
```
<description> {{cells[“description”].value}} </description>
```

```
<calories> {{cells[“calories”].value}} </calories>
```

```
</food>
```

- จากนั้นคลิก Export เพื่อดาวน์โหลดไฟล์

ว.2.3 ขั้นตอนหลังจากการแปลงไฟล์เป็น JSON และ XML เพื่อนำไปสู่การต่อยอดใช้งาน



แผนภาพแสดงขั้นตอนหลังจากแปลงไฟล์เป็น JSON และ XML เพื่อนำไปสู่การต่อยอดใช้งานใน Web/Mobile App หรือการวิเคราะห์ข้อมูล (Analytics) โดยสามารถเลือกใช้งานจากข้อมูลในรูปแบบ XML/JSON ได้โดยตรงหรือแปลงเป็นรูปแบบของ RDF / Linked RDF ก่อน



สถาบันส่งเสริมการวิเคราะห์และ
บริหารข้อมูลขนาดใหญ่ภาครัฐ (GBDi)
สำนักงานส่งเสริมเศรษฐกิจดิจิทัล

อาคารลาดพร้าว ฮิลส์
เลขที่ 80 ซ.ลาดพร้าว 4
ถนนลาดพร้าว แขวงจอมพล
เขตจตุจักร

กรุงเทพมหานคร 10900

โทรศัพท์ 0 20262333

www.bigdata.go.th

www.gbdi.depa.or.th

 Govbigdata

 Govbigdata

 Govbigdata

 Govbigdata

 @gbdi



กระทรวงศึกษาธิการ
เพื่อครูผู้ทรงคุณวุฒิ

