



(ร่าง)

กรอบการวิเคราะห์ข้อมูลขนาดใหญ่ภาครัฐ
(Government Big Data Analytics Framework)
Version 1

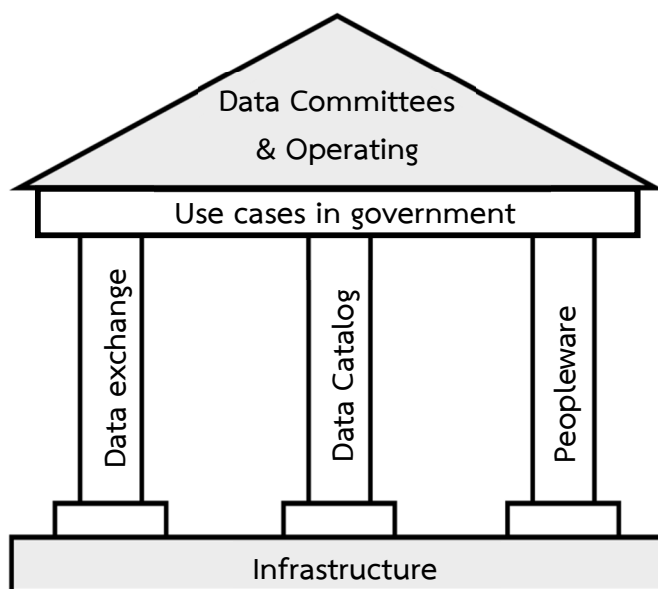
จัดทำโดย

คณะกรรมการออกแบบสถาปัตยกรรม (Architecture Design) ระบบบูรณาการข้อมูลภาครัฐ
ภายใต้คณะกรรมการขับเคลื่อนการดำเนินนโยบายเพื่อใช้ประโยชน์ข้อมูลขนาดใหญ่ (Big Data)
ศูนย์ข้อมูล (Data Center) และคลาวด์คอมพิวติ้ง (Cloud Computing)



(ร่าง)

กรอบการวิเคราะห์ข้อมูลขนาดใหญ่ภาครัฐ (Government Big Data Analytics Framework)



โดย

คณะอนุกรรมการออกแบบสถาปัตยกรรม (Architecture Design) ระบบบูรณาการข้อมูลภาครัฐ
ภายใต้คณะกรรมการขับเคลื่อนการดำเนินนโยบายเพื่อใช้ประโยชน์ข้อมูลขนาดใหญ่ (Big Data)
ศูนย์ข้อมูล (Data Center) และคลาวด์คอมพิวติ้ง (Cloud Computing)

Version 1.0



คำสั่งคณะกรรมการขับเคลื่อนการดำเนินนโยบายเพื่อใช้ประโยชน์ข้อมูลขนาดใหญ่ (Big Data)
ศูนย์ข้อมูล (Data Center) และคลาวด์คอมพิวติ้ง (Cloud Computing)

ที่ ๑ / ๒๕๖๑

เรื่อง แต่งตั้งคณะกรรมการออกแบบสถาปัตยกรรม (Architecture Design)
ระบบบูรณาการข้อมูลของภาครัฐ

ตามที่ได้มีการแต่งตั้งคณะกรรมการขับเคลื่อนการดำเนินนโยบายเพื่อใช้ประโยชน์ข้อมูลขนาดใหญ่ (Big Data) ศูนย์ข้อมูล (Data Center) และคลาวด์คอมพิวติ้ง (Cloud Computing) โดยมีอำนาจหน้าที่ในการกำหนดยุทธศาสตร์ กำหนดโจทย์ที่มีความสำคัญเร่งด่วน ติดตามการดำเนินการบูรณาการฐานข้อมูล กำกับให้ทุกส่วนราชการที่เป็นเจ้าของข้อมูลรับผิดชอบความถูกต้อง ตลอดจนเสนอแนะเกี่ยวกับระเบียบ ข้อบังคับ กฎหมายที่เกี่ยวข้อง อันเป็นปัญหาและอุปสรรคต่อการบูรณาการฐานข้อมูล นั้น

เพื่อให้การดำเนินการดังกล่าวเป็นไปอย่างมีประสิทธิภาพ และบรรลุผลสำเร็จอย่างเป็นรูปธรรม โดยอาศัยอำนาจหน้าที่ของคณะกรรมการขับเคลื่อนการดำเนินนโยบายเพื่อใช้ประโยชน์ข้อมูลขนาดใหญ่ (Big Data) ศูนย์ข้อมูล (Data Center) และคลาวด์คอมพิวติ้ง (Cloud Computing) ตามข้อ ๒.๗ จึงเห็นสมควรแต่งตั้งคณะกรรมการออกแบบสถาปัตยกรรม (Architecture Design) ระบบบูรณาการข้อมูลของภาครัฐ โดยมีองค์ประกอบ และอำนาจหน้าที่ ดังนี้

๑.องค์ประกอบ

๑.๑ ปลัดกระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม	ประธานคณะกรรมการ
๑.๒ นางอาทิตยา สุธารธรรม	อนุกรรมการ
รองปลัดกระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม	
๑.๓ ผู้แทนสำนักงานปลัดสำนักนายกรัฐมนตรี	อนุกรรมการ
๑.๔ ผู้แทนกระทรวงกลาโหม	อนุกรรมการ
๑.๕ ผู้แทนกระทรวงการคลัง	อนุกรรมการ
๑.๖ ผู้แทนกระทรวงการต่างประเทศ	อนุกรรมการ
๑.๗ ผู้แทนกระทรวงการท่องเที่ยวและกีฬา	อนุกรรมการ
๑.๘ ผู้แทนกระทรวงการพัฒนาสังคมและความมั่นคงของมนุษย์	อนุกรรมการ
๑.๙ ผู้แทนกระทรวงเกษตรและสหกรณ์	อนุกรรมการ
๑.๑๐ ผู้แทนกระทรวงคมนาคม	อนุกรรมการ
๑.๑๑ ผู้แทนกระทรวงพลังงาน	อนุกรรมการ
๑.๑๒ ผู้แทนกระทรวงพาณิชย์	อนุกรรมการ
๑.๑๓ ผู้แทนกระทรวงมหาดไทย	อนุกรรมการ

๑.๑๔ ผู้แทนกระทรวงยุติธรรม	อนุกรรมการ
๑.๑๕ ผู้แทนกระทรวงแรงงาน	อนุกรรมการ
๑.๑๖ ผู้แทนกระทรวงวัฒนธรรม	อนุกรรมการ
๑.๑๗ ผู้แทนกระทรวงวิทยาศาสตร์และเทคโนโลยี	อนุกรรมการ
๑.๑๘ ผู้แทนกระทรวงศึกษาธิการ	อนุกรรมการ
๑.๑๙ ผู้แทนกระทรวงสาธารณสุข	อนุกรรมการ
๑.๒๐ ผู้แทนกระทรวงอุตสาหกรรม	อนุกรรมการ
๑.๒๑ ผู้แทนบริษัท กสท โทรคมนาคม จำกัด (มหาชน)	อนุกรรมการ
๑.๒๒ รองศาสตราจารย์ ดร.ธีรณี อจลากุล	อนุกรรมการ
๑.๒๓ ผู้ช่วยศาสตราจารย์ ดร.ภูงศ อุตโยภาส	อนุกรรมการ
๑.๒๔ ผู้ช่วยศาสตราจารย์ ดร.สุภาภรณ์ เกียรติสิน	อนุกรรมการ
๑.๒๕ ผู้ช่วยศาสตราจารย์ ดร.ณัฐวุฒิ หนูไพโรจน์	อนุกรรมการ
๑.๒๖ นายไชยเจริญ อติแพทย์	อนุกรรมการ
๑.๒๗ นายศุภชัย จงศิริ	อนุกรรมการ
๑.๒๘ นายกำพล ศรชนะรัตน์	อนุกรรมการ
๑.๒๙ นายแพทย์สุธี ทวีรัตน์	อนุกรรมการ
๑.๓๐ ดร.กฤษชลิข ฐิติกมล	อนุกรรมการ
๑.๓๑ นายอดิศักดิ์ ศรีนครินทร์	อนุกรรมการและ
สำนักงานพัฒนารัฐบาลดิจิทัล (องค์การมหาชน)	เลขานุการร่วม
๑.๓๒ นายทรงพล ใหม่สาธิต	อนุกรรมการและ
สำนักงานสถิติแห่งชาติ	เลขานุการร่วม
๑.๓๓ นางสมศรี หอกันยา	อนุกรรมการและ
สำนักงานปลัดกระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม	เลขานุการร่วม
๑.๓๔ นางสาววาสนา หมั่นสระเกษ	ผู้ช่วยเลขานุการ
สำนักงานพัฒนารัฐบาลดิจิทัล (องค์การมหาชน)	
๑.๓๕ นางสาวธัญลักษณ์ มณีวัฒนา	ผู้ช่วยเลขานุการ
สำนักงานสถิติแห่งชาติ	
๑.๓๖ นายทรงวุฒิ โชติกาญจวิทย์	ผู้ช่วยเลขานุการ
สำนักงานปลัดกระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม	

๒. อำนาจหน้าที่

๒.๑ จัดทำและออกแบบสถาปัตยกรรมองค์กรของภาครัฐ (Government Enterprise Architecture) เพื่อให้เกิดบูรณาการในการปฏิรูปกระบวนการทางธุรกิจและสารสนเทศของหน่วยงานภาครัฐ ให้สามารถดำเนินการประมวลผลข้อมูลขนาดใหญ่ (Big Data) ศูนย์ข้อมูล (Data Center) และคลาวด์คอมพิวติ้ง (Cloud Computing) ได้ตรงตามวัตถุประสงค์และรายงานต่อคณะกรรมการขับเคลื่อนการดำเนินนโยบายเพื่อใช้ประโยชน์ข้อมูลขนาดใหญ่ (Big Data) ศูนย์ข้อมูล (Data Center) และคลาวด์คอมพิวติ้ง (Cloud Computing)

๒.๒ ออกแบบระบบเพื่อความมั่นคงปลอดภัยทางไซเบอร์ (Cyber Security) เพื่อความมั่นคง ปลอดภัยในการใช้ประโยชน์จากข้อมูลขนาดใหญ่ และเสนอแผนในการเตรียมการด้านโครงสร้างพื้นฐานระบบนิเวศน์ (Ecosystem) เพื่อรองรับการบูรณาการข้อมูลภาครัฐ

๒.๓ จัดทำข้อเสนอแนวทางการดำเนินงานที่เกี่ยวข้องเพื่อผลักดันการใช้ประโยชน์ข้อมูลขนาดใหญ่ (Big Data) ศูนย์ข้อมูล (Data Center) และคลาวด์คอมพิวติ้ง (Cloud Computing)

๒.๔ รายงานผลการดำเนินการที่เกี่ยวข้องให้คณะกรรมการขับเคลื่อนการดำเนินนโยบายเพื่อใช้ประโยชน์ข้อมูลขนาดใหญ่ (Big Data) ศูนย์ข้อมูล (Data Center) และคลาวด์คอมพิวติ้ง (Cloud Computing) ทราบเป็นระยะ

๒.๕ ทำหน้าที่ตัดสินใจกรณีมีทางเลือกเพื่อให้มีการเชื่อมโยง Integration กระบวนการราชการที่เหมาะสม

๒.๖ เชิญผู้แทนส่วนราชการ รัฐวิสาหกิจ หน่วยงานภาครัฐ หรือหน่วยงานภาคเอกชน มาชี้แจง หรือส่งข้อมูลใด ๆ มาให้คณะกรรมการได้ และประสานงานตรงกับหน่วยงานที่เกี่ยวข้อง เพื่อขอรับการสนับสนุนข้อมูล

๒.๗ แต่งตั้งคณะทำงาน เพื่อดำเนินการใด ๆ ได้ตามความจำเป็น

๒.๘ ดำเนินการอื่นใดตามที่คณะกรรมการขับเคลื่อนการดำเนินนโยบายเพื่อใช้ประโยชน์ข้อมูลขนาดใหญ่ (Big Data) ศูนย์ข้อมูล (Data Center) และคลาวด์คอมพิวติ้ง (Cloud Computing) มอบหมาย

สำหรับการเบิกจ่ายประชุมให้เป็นไปตามพระราชกฤษฎีกาเบี่ยงประชุมกรรมการ พ.ศ. ๒๕๔๗ และที่แก้ไขเพิ่มเติม ส่วนค่าใช้จ่ายในการเดินทางและค่าใช้จ่ายที่เกี่ยวข้องกับการบริหารจัดการอื่น ๆ ที่จำเป็นต่อการปฏิบัติงานให้เบิกจ่ายตามระเบียบของทางราชการ โดยให้เบิกจ่ายจากกระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม

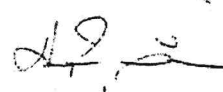
ทั้งนี้ ตั้งแต่บัดนี้เป็นต้นไป

สั่ง ณ วันที่ ๒๕ พฤษภาคม พ.ศ. ๒๕๖๑

สำเนาถูกต้อง


(นางสาวนงลักษณ์ โง้ววินชัย)
นักวิชาการสถิติชำนาญการพิเศษ
รักษาการในตำแหน่ง ผู้เชี่ยวชาญด้านวิชาการสถิติ

พลอากาศเอก


(ประจักษ์ จันทอง)

รองนายกรัฐมนตรี

และรัฐมนตรีว่าการกระทรวงยุติธรรม

ประธานคณะกรรมการขับเคลื่อนการดำเนินนโยบาย
เพื่อใช้ประโยชน์ข้อมูลขนาดใหญ่ (Big Data) ศูนย์ข้อมูล (Data Center)
และคลาวด์คอมพิวติ้ง (Cloud Computing)



คำสั่งคณะกรรมการขับเคลื่อนการดำเนินนโยบายเพื่อใช้ประโยชน์ข้อมูลขนาดใหญ่ (Big Data)
ศูนย์ข้อมูล (Data Center) และคลาวด์คอมพิวติ้ง (Cloud Computing)
ที่ ๕ / ๒๕๖๑

เรื่อง แก้ไขเพิ่มเติมรายชื่อคณะกรรมการออกแบบสถาปัตยกรรม (Architecture Design)
ระบบบูรณาการข้อมูลของภาครัฐ

อนึ่ง คำสั่งคณะกรรมการขับเคลื่อนการดำเนินนโยบายเพื่อใช้ประโยชน์ข้อมูลขนาดใหญ่ (Big Data) ศูนย์ข้อมูล (Data Center) และคลาวด์คอมพิวติ้ง (Cloud Computing) ที่ ๑/๒๕๖๑ ลงวันที่ ๒๕ พฤษภาคม ๒๕๖๑ แต่งตั้งคณะกรรมการออกแบบสถาปัตยกรรม (Architecture Design) ระบบบูรณาการข้อมูลของภาครัฐ จำนวน ๓๖ ราย ไปแล้ว นั้น

เพื่อให้การดำเนินการดังกล่าวเป็นไปอย่างมีประสิทธิภาพ และบรรลุผลสำเร็จอย่างเป็นรูปธรรม โดยอาศัยอำนาจหน้าที่ของคณะกรรมการขับเคลื่อนการดำเนินนโยบายเพื่อใช้ประโยชน์ข้อมูลขนาดใหญ่ (Big Data) ศูนย์ข้อมูล (Data Center) และคลาวด์คอมพิวติ้ง (Cloud Computing) ตามข้อ ๒.๗ จึงให้แก้ไขเพิ่มเติมรายชื่อคณะกรรมการออกแบบสถาปัตยกรรม (Architecture Design) ระบบบูรณาการข้อมูลของภาครัฐ จำนวน ๒ ราย ดังนี้

- | | |
|---|------------|
| ๑. ผู้แทนกระทรวงทรัพยากรธรรมชาติและสิ่งแวดล้อม | อนุกรรมการ |
| ๒. ผู้อำนวยการสำนักงานพัฒนารัฐบาลดิจิทัล (องค์การมหาชน) | อนุกรรมการ |

ทั้งนี้ ตั้งแต่บัดนี้เป็นต้นไป

สั่ง ณ วันที่ ๗ สิงหาคม พ.ศ. ๒๕๖๑

พลอากาศเอก

(ประจิน จั่นตอง)

รองนายกรัฐมนตรี

และรัฐมนตรีว่าการกระทรวงยุติธรรม

ประธานคณะกรรมการขับเคลื่อนการดำเนินนโยบาย
เพื่อใช้ประโยชน์ข้อมูลขนาดใหญ่ (Big Data) ศูนย์ข้อมูล (Data Center)
และคลาวด์คอมพิวติ้ง (Cloud Computing)

สำเนาถูกต้อง

(นางสาวนงลักษณ์ โง้ววัฒนชัย)

นักวิชาการสถิติชำนาญการพิเศษ

รักษาการในตำแหน่ง ผู้เชี่ยวชาญด้านวิชาการสถิติ



คำสั่งคณะกรรมการออกแบบสถาปัตยกรรม (Architecture Design) ระบบบูรณาการข้อมูลของภาครัฐ

ที่ ๑ /๒๕๖๑

เรื่อง แต่งตั้งคณะกรรมการด้านการจัดทำและออกแบบสถาปัตยกรรมองค์กรของภาครัฐ

(Government Enterprise Architecture)

ตามที่ได้มีการแต่งตั้งคณะกรรมการออกแบบสถาปัตยกรรม (Architecture Design) ระบบบูรณาการข้อมูลของภาครัฐ โดยมีอำนาจหน้าที่ในการจัดทำและออกแบบสถาปัตยกรรมองค์กรของภาครัฐ (Government Enterprise Architecture) เพื่อให้เกิดบูรณาการในการปฏิบัติงานกระบวนการทางธุรกิจและสารสนเทศของหน่วยงานภาครัฐให้สามารถดำเนินการประมวลผลข้อมูลขนาดใหญ่ (Big Data) ศูนย์ข้อมูล (Data Center) และคลาวด์คอมพิวติ้ง (Cloud Computing) ได้ตรงตามวัตถุประสงค์และรายงานต่อคณะกรรมการขับเคลื่อนการดำเนินนโยบายเพื่อใช้ประโยชน์ข้อมูลขนาดใหญ่ (Big Data) ศูนย์ข้อมูล (Data Center) และคลาวด์คอมพิวติ้ง (Cloud Computing) นั้น

เพื่อให้การดำเนินการดังกล่าวเป็นไปอย่างมีประสิทธิภาพ และบรรลุผลสำเร็จอย่างเป็นรูปธรรม โดยอาศัยอำนาจหน้าที่ของคณะกรรมการออกแบบสถาปัตยกรรม (Architecture Design) ระบบบูรณาการข้อมูลของภาครัฐ ตามข้อ ๒.๗ จึงเห็นสมควรแต่งตั้งคณะกรรมการด้านการจัดทำและออกแบบสถาปัตยกรรมองค์กรของภาครัฐ (Government Enterprise Architecture) โดยมีองค์ประกอบ และอำนาจหน้าที่ ดังนี้

๑. องค์ประกอบ

- | | |
|--|--------------------------------|
| ๑.๑ นายศุภชัย จงศิริ | ประธานคณะกรรมการ |
| ๑.๒ นางอาทิตย์ยา สุธารธรรม | รองประธานคณะกรรมการ |
| ๑.๓ นายแพทย์พลวรรธน์ วิฑูรกลชิต
สำนักงานปลัดกระทรวงสาธารณสุข | คณะกรรมการ |
| ๑.๔ นางสมศรี หอกันยา
สำนักงานปลัดกระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม | คณะกรรมการ |
| ๑.๕ นายอดิศักดิ์ ศรีนครินทร์
สำนักงานพัฒนารัฐบาลดิจิทัล (องค์การมหาชน) | คณะกรรมการ |
| ๑.๖ รองศาสตราจารย์ ดร.ธีรณี อจลากุล | คณะกรรมการ |
| ๑.๗ ผู้ช่วยศาสตราจารย์ ดร.ภูษงค์ อุทโยภาส | คณะกรรมการ |
| ๑.๘ ผู้ช่วยศาสตราจารย์ ดร.สุภาภรณ์ เกียรติสิน | คณะกรรมการ |
| ๑.๙ ผู้ช่วยศาสตราจารย์ ดร.ณัฐวุฒิ หนูไพโรจน์ | คณะกรรมการ |
| ๑.๑๐ นายไชยเจริญ อติแพทย์ | คณะกรรมการ |
| ๑.๑๑ นายกำพล ศรธนรัตน์ | คณะกรรมการ |
| ๑.๑๒ นายแพทย์สุธี ทุวิรัตน์ | คณะกรรมการ |
| ๑.๑๓ ดร.กฤษชัช ฐิติกมล | คณะกรรมการ |
| ๑.๑๔ นายทรงวุฒิ โชติกาญจน์วิทย์
สำนักงานปลัดกระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม | คณะกรรมการ
และเลขานุการร่วม |

- ๑.๑๕ ผู้แทนสำนักงานพัฒนารัฐบาลดิจิทัล (องค์การมหาชน) คณะทำงาน
และเลขานุการร่วม
- ๑.๑๖ นายทรงพล ใหม่สาลี
สำนักงานสถิติแห่งชาติ คณะทำงาน
และเลขานุการร่วม
- ๑.๑๗ นายจุฑาสีทธิ์ โรหิตรัตน์
สำนักงานปลัดกระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม คณะทำงาน
และผู้ช่วยเลขานุการ
- ๑.๑๘ ผู้แทนสำนักงานพัฒนารัฐบาลดิจิทัล (องค์การมหาชน) คณะทำงาน
และผู้ช่วยเลขานุการ
- ๑.๑๙ นางสาวธัญลักษณ์ มณีวัฒนา
สำนักงานสถิติแห่งชาติ คณะทำงาน
และผู้ช่วยเลขานุการ
- ๑.๒๐ นายณรงค์เดช วัชรภาสกร
สำนักงานปลัดกระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม ผู้ช่วยเลขานุการ
- ๑.๒๑ นายชิษณุพงศ์ ธนุทอง
สำนักงานปลัดกระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม ผู้ช่วยเลขานุการ
- ๑.๑๑ นายณัฐ พยงค์ศรี
สำนักงานปลัดกระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม ผู้ช่วยเลขานุการ

๒. อำนาจหน้าที่

๒.๑ จัดทำและออกแบบสถาปัตยกรรมองค์กรของภาครัฐ (Government Enterprise Architecture) เพื่อให้เกิดบูรณาการในการปฏิรูปกระบวนการทางธุรกิจและสารสนเทศของหน่วยงานภาครัฐให้สามารถดำเนินการประมวลผลข้อมูลขนาดใหญ่ (Big Data) ศูนย์ข้อมูล (Data Center) และคลาวด์คอมพิวติ้ง (Cloud Computing) ได้ตรงตามวัตถุประสงค์

๒.๒ จัดทำข้อเสนอแนวทางการดำเนินงานที่เกี่ยวข้องเพื่อผลักดันการใช้ประโยชน์ของสถาปัตยกรรมองค์กรของภาครัฐ (Government Enterprise Architecture)

๒.๓ รายงานผลการดำเนินงานตามข้อ ๒.๑ และ ๒.๒ เสนอคณะกรรมการออกแบบสถาปัตยกรรม (Architecture Design) ระบบบูรณาการข้อมูลของภาครัฐพิจารณา

๒.๔ เชิญผู้แทนส่วนราชการ รัฐวิสาหกิจ หน่วยงานภาครัฐ หรือหน่วยงานภาคเอกชนมาชี้แจง หรือส่งข้อมูลใด ๆ มาให้คณะกรรมการได้ และประสานงานตรงกับหน่วยงานที่เกี่ยวข้อง เพื่อขอรับการสนับสนุนข้อมูล

๒.๕ ดำเนินการอื่นใดตามที่คณะกรรมการออกแบบสถาปัตยกรรม (Architecture Design) ระบบบูรณาการข้อมูลของภาครัฐมอบหมาย

สำหรับการเบิกจ่ายประชุมให้เป็นไปตามพระราชกฤษฎีกาเบี่ยงประชุมกรรมการ พ.ศ. ๒๕๔๗ และที่แก้ไขเพิ่มเติม ส่วนค่าใช้จ่ายในการเดินทางและค่าใช้จ่ายที่เกี่ยวข้องกับการบริหารจัดการอื่นๆ ที่จำเป็นต่อการปฏิบัติงานให้เบิกจ่ายตามระเบียบของทางราชการ โดยให้เบิกจ่ายจากกระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม

ทั้งนี้ตั้งแต่บัดนี้เป็นต้นไป

สั่ง ณ วันที่ ๑๓ มิถุนายน ๒๕๖๑



(นางสาวอัจฉรินทร์ พัฒนพันธ์ชัย)

ปลัดกระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม

ประธานคณะกรรมการออกแบบสถาปัตยกรรม (Architecture Design)

ระบบบูรณาการข้อมูลของภาครัฐ

สารบัญ

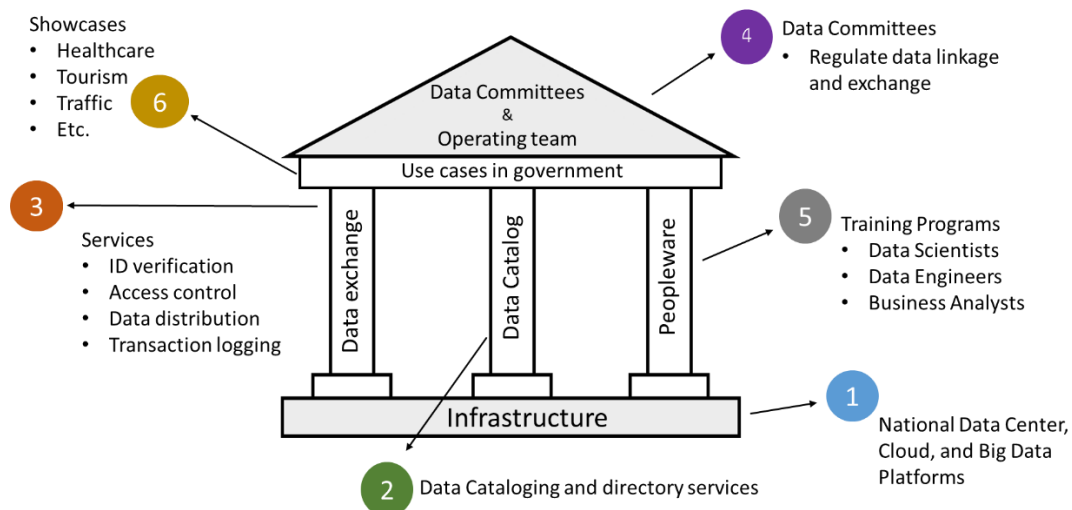
บทนำ.....	1
บทที่ 1 โครงสร้างพื้นฐานเพื่อการประมวลผลภาครัฐ.....	3
1.1 กรอบมาตรฐานดาต้าเซ็นเตอร์/คลาวด์ภาครัฐ.....	4
1.2 โครงสร้างพื้นฐานเพื่อการประมวลผลของหน่วยงานภาครัฐ.....	6
1.3 ระบบคลาวด์กลางภาครัฐ (Government Data Center and Cloud services : GDCC).....	9
บทที่ 2 รายการข้อมูลภาครัฐ (Government Data Catalog).....	13
2.1 บริการรายการข้อมูลภาครัฐ (Government Data Catalog Service).....	13
2.2 การทำงานร่วมกับหน่วยงานภาครัฐเรื่องการกำกับดูแลข้อมูล	18
บทที่ 3 การให้บริการด้านการแลกเปลี่ยนข้อมูล	21
3.1 มาตรฐานข้อมูล.....	21
3.2 การปรับปรุงข้อมูลเพื่อเตรียมความพร้อม.....	22
3.3 ระบบบริการและแลกเปลี่ยนข้อมูลกลาง	23
บทที่ 4 การกลั่นกรองคำร้องขอข้อมูลข้ามหน่วยงาน	26
บทที่ 5 การพัฒนาบุคลากรภาครัฐด้านการวิเคราะห์ข้อมูล	28
5.1 กรอบการพัฒนาบุคลากร	28
5.2 เครือข่ายมหาวิทยาลัยด้านวิทยาศาสตร์ข้อมูล.....	32
บทที่ 6 กรณีตัวอย่างการใช้ประโยชน์ข้อมูล.....	39
6.1 การเพิ่มศักยภาพด้านการท่องเที่ยวเมืองรอง.....	39
6.2 การวิเคราะห์สื่อสังคมออนไลน์และพัฒนาเครื่องมือเพื่อรวบรวมข้อมูลสำรวจความคิดเห็นของภาครัฐ : กรณีศึกษาเน็ตประชารัฐ.....	41
6.3 การจัดทำเครื่องมือประเมินช่วยคัดกรองผู้กระทำผิดที่มีโอกาสย้อนกลับไปกระทำผิดซ้ำ	43
6.4 โครงการวิเคราะห์ข้อมูลค่าเบิกจ่ายสิทธิการรักษาพยาบาลจากรัฐบาล	45
6.5 การศึกษาปัจจัยเพื่อทำนายระดับความสามารถของบิดามารดาในการดูแลบุตรวัย 0-3 ขวบ	46
6.6 การวิเคราะห์เสี่ยงสะท้อนจากนักท่องเที่ยว	46

6.7 การเพิ่มศักยภาพด้านการจัดเก็บภาษี.....	47
6.8 การวิเคราะห์รูปแบบภูมิอากาศเพื่อการวางแผนของท่าอากาศยาน	47
6.9 การจัดทำแพลตฟอร์มการรวบรวมข้อมูลเพื่อการวิเคราะห์ “Digital Transformation ตามมาตรฐาน” OECD (Digital Economy Outlook)	47
เอกสารแนบ ก1 สรุปรายชื่อหน่วยงานระดับกรมหรือเทียบเท่า ที่มี Agency Data Center / Cloud.....	50
เอกสารแนบ ก2 สรุปจำนวนหน่วยงานที่จะร่วมใช้คลาวด์ระดับกระทรวง (Ministry Cloud).....	53
เอกสารแนบ ข แนวทางปฏิบัติทางเทคโนโลยีสำหรับการแปลงข้อมูลให้อยู่ในรูปแบบพร้อมใช้งาน	55
ข.1. วิธีการแปลงข้อมูลจาก CSV เป็น JSON.....	55
ข.1.1 การแปลงข้อมูลเครื่องมือออนไลน์	55
ข.1.2 การแปลงข้อมูลด้วยเครื่องมือที่ติดตั้งบนเครื่อง	56
ข.2 วิธีการแปลงข้อมูลจาก CSV เป็น XML	58
ข.2.1 การแปลงข้อมูลเครื่องมือออนไลน์	58
ข.2.2 การแปลงข้อมูลด้วยเครื่องมือที่ติดตั้งบนเครื่อง	58
ข.3 ขั้นตอนหลังจากการแปลงไฟล์เป็น JSON และ XML เพื่อนำไปสู่การต่อยอดใช้งาน.....	60

บทนำ

ในปัจจุบันประเทศไทยมีความตื่นตัวเรื่องบิ๊กดาต้าอย่างกว้างขวางตามกระแสโลก องค์กรทั้งภาครัฐและภาคเอกชนเริ่มปรับตัวและเตรียมพร้อมเข้าสู่ยุคของการวิเคราะห์และใช้ประโยชน์ข้อมูลขนาดใหญ่ เพื่อประกอบการตัดสินใจ (Data Driven Decision) และการดำเนินงาน (Insight to Operation) ดังนั้น เพื่อให้เกิดประโยชน์อย่างเป็นรูปธรรม องค์กรภาครัฐซึ่งเป็นเจ้าของข้อมูลสำคัญที่มีผลต่อการพัฒนาประเทศ จึงควรต้องคำนึงถึงการบูรณาการและการให้บริการข้อมูลข้ามหน่วยงานภาครัฐ ตลอดไปจนถึงการให้บริการข้อมูลที่ไม่ขึ้นความลับแก่ภาคเอกชน คณะทำงานภายใต้คณะกรรมการขับเคลื่อนการดำเนินนโยบาย เพื่อใช้ประโยชน์ข้อมูลขนาดใหญ่ (Big Data) ศูนย์ข้อมูล (Data Center) และคลาวด์คอมพิวติ้ง (Cloud Computing) จึงได้มีมติให้จัดตั้งหน่วยงานใหม่ที่ใช้ชื่อว่า **สถาบันส่งเสริมการวิเคราะห์และบริหารข้อมูลภาครัฐ** ซึ่งเป็นหน่วยงานในรูปแบบ Service Delivery Unit หรือ SDU โดยหน่วยงานดังกล่าวมีวัตถุประสงค์หลักในการส่งเสริมให้เกิดโครงการและสนับสนุนการดำเนินงานด้านบิ๊กดาต้าสำหรับหน่วยงานภาครัฐ นอกจากนี้ SDU ดังกล่าวจะดำเนินงานวางกรอบการพัฒนาการใช้ประโยชน์ข้อมูลภาครัฐให้เป็นไปในทิศทางเดียวกัน สามารถบูรณาการและให้บริการข้อมูลข้ามหน่วยงานได้อย่างเป็นระบบ รวดเร็ว ปลอดภัย และควบคุมได้ ซึ่งจะนำไปสู่การวิเคราะห์และใช้ประโยชน์ข้อมูลอย่างเป็นรูปธรรมในอนาคต ดังอธิบายในเอกสารฉบับนี้

การออกแบบแนวคิดของการให้บริการข้อมูลภาครัฐ (Government Data Service Framework) แสดงในภาพที่ 1



ภาพที่ 1 แนวคิดของการให้บริการข้อมูลภาครัฐ (Government Data Service Framework)

กรอบการให้บริการข้อมูลภาครัฐประกอบด้วย 6 ส่วนงานหลักที่จะสร้างให้เกิดระบบนิเวศข้อมูลที่สามารถถูกใช้ประโยชน์ร่วมกันได้อย่างเป็นระบบ โดยทุกส่วนงานควรเริ่มดำเนินการไปพร้อมๆ กัน ดังนี้

1. **พัฒนาโครงสร้างพื้นฐานการประมวลผลภาครัฐที่เหมาะสม** มีประสิทธิภาพในการใช้งานทั้งในมิติ ด้านความต่อเนื่องของบริการ การรักษาความปลอดภัย และการแลกเปลี่ยนเชื่อมโยงข้อมูลอย่างเป็นระบบ ซึ่งสถาปัตยกรรมโครงสร้างพื้นฐานควรเป็นแบบแบ่งปัน ที่ผู้รับบริการมองเห็นว่าเป็นโครงสร้างเดียวที่ใช้ร่วมกัน ในขณะที่ทางกายภาพโครงสร้างพื้นฐานเป็นแบบกระจาย ที่มีการเข้าใช้คลาวด์ผสมกับการสร้างดาต้าเซนเตอร์ หน่วยงานภาครัฐสามารถวางแผนการปรับโครงสร้างพื้นฐานการประมวลผลที่เหมาะสมให้กับหน่วยงานได้ โดยทำการวิเคราะห์ช่องว่าง (Gap Analysis) ตามกรอบแนวทางที่กระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคมกำหนดขึ้น (ผลสรุปการออกแบบโครงสร้างพื้นฐานภาครัฐแสดงในบทที่ 1)

2. **พัฒนาระบบรายการข้อมูลภาครัฐ (Government Data Catalog)** โดยรวบรวมเมทาดาต้าของข้อมูลสำคัญจากหน่วยงานภาครัฐทั้งหมดเพื่อจัดสร้างเป็นแคตาล็อก (โดยแบ่งทำเป็นระยะ เริ่มจากกลุ่มหน่วยงานนำร่องก่อน) จากนั้นพัฒนากลไกการสืบค้นในมิติต่างๆ และหน้าจอสืบค้นที่ใช้งานง่ายสำหรับให้บุคลากรภาครัฐค้นหาแหล่งข้อมูลเพื่อใช้ประกอบการวิเคราะห์ หรือ ให้บริการได้อย่างสะดวก รวมไปถึงการออกแบบเครื่องมือต่างๆ ที่หน่วยงานภาครัฐต้องใช้ในการกำกับดูแลและให้บริการข้อมูลของหน่วยงานได้อย่างเป็นระบบ (รายละเอียดแสดงในบทที่ 2)

3. **พัฒนาระบบสารสนเทศเพื่อบริหารจัดการการเชื่อมโยงและแลกเปลี่ยนข้อมูล** สำหรับการวิเคราะห์ข้อมูลขนาดใหญ่ (Big Data Analytics) ของหน่วยงานภาครัฐโดยมีบริการพื้นฐาน เช่น การยืนยันตัวตน การจำกัดสิทธิในการเข้าถึงข้อมูล บริการส่งข้อมูลข้ามหน่วยงานผ่านระบบสารสนเทศ บันทึกทราบนโยบาย การเข้าถึงข้อมูลเพื่อเป็นหลักฐาน และการบริการข้อมูลประเภทอื่น (การออกแบบแสดงในบทที่ 3)

4. **ตั้งคณะกรรมการกลั่นกรองคำขอข้อมูลและคณะทำงาน** เพื่อให้มีนโยบายการใช้ข้อมูลของหน่วยงานภาครัฐและกำกับกรให้บริการข้อมูลในมิติต่างๆ ของหน่วยงานภาครัฐ (กรอบหน้าที่ของคณะกรรมการนำเสนอในบทที่ 4)

5. **วางแผนและดำเนินงานด้านการพัฒนาบุคลากรภาครัฐ** เพื่อเตรียมความพร้อมสำหรับการประยุกต์ใช้ข้อมูลในการวิเคราะห์เพื่อวางแผนยุทธศาสตร์ การปฏิบัติงานของเจ้าหน้าที่ และการบริหารงานของหน่วยงานภาครัฐ โดยเน้นการอบรมเพื่อพัฒนานักวิทยาศาสตร์ข้อมูล นักวิเคราะห์ข้อมูล และวิศวกรข้อมูล (รายละเอียดและกระบวนการสร้างบุคลากรในบทที่ 5)

6. **พัฒนาระบบนำร่องการวิเคราะห์ข้อมูลขนาดใหญ่** ในหลากหลายสาขา เช่น การแพทย์และการสาธารณสุข การบริหารงานการจราจร การพัฒนาการศึกษา การส่งเสริมการท่องเที่ยว ฯลฯ เพื่อให้บุคลากรภาครัฐเห็นถึงประโยชน์ของการใช้ข้อมูลประกอบการสร้างนโยบายและแผนงาน (ตัวอย่างกรณีศึกษาเสนอในบทที่ 6)

บทที่ 1 โครงสร้างพื้นฐานเพื่อการประมวลผลภาครัฐ

การวางแผนภาพรวมของโครงสร้างพื้นฐานดาต้าเซ็นเตอร์และคลาวด์มีความสำคัญในการเพิ่มประสิทธิภาพของระบบงานภาครัฐโดยรวม ทั้งในด้านการให้บริการระบบงานที่ต่อเนื่องและการรักษาความปลอดภัยทางไซเบอร์ให้ได้ตามมาตรฐานสากล การจัดการทรัพยากรประมวลผลที่คุ้มค่า การวางแผนงบประมาณที่มีประสิทธิภาพ และการแลกเปลี่ยนข้อมูลข้ามหน่วยงานอย่างเป็นระบบ ซึ่งรูปแบบแนวคิด (Conceptual Design) สำหรับโครงสร้างพื้นฐานการประมวลผลภาครัฐจะเป็นแบบกระจาย โดยมีการเช่าใช้คลาวด์สมกับการสร้างดาต้าเซ็นเตอร์เพื่อเพิ่มความยืดหยุ่นในการพัฒนาโครงสร้างพื้นฐาน หน่วยงานที่เช่าใช้ระบบคลาวด์จะไม่จำเป็นต้องจัดซื้อครุภัณฑ์ หรือสร้างอาคารดาต้าเซ็นเตอร์เพื่อรองรับการขยายตัวของระบบงานทั้งหมด และบุคลากรสารสนเทศสามารถมุ่งเน้นการกำกับดูแลระบบงานและข้อมูลแทนการใช้เวลาในการดูแลเครื่องประมวลผลและอุปกรณ์พื้นฐานได้ ซึ่งจากข้อมูลในต่างประเทศการใช้คลาวด์นั้นจะประหยัดงบประมาณได้ถึง 30-70% ขึ้นกับลักษณะการใช้งานโดยการใช้แบบ Software-as-a-service หรือ SaaS จะประหยัดได้สูงที่สุด [ref: “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]

ทั้งนี้ เนื่องจากหน่วยงานภาครัฐแต่ละหน่วยงานมีลักษณะการใช้ทรัพยากรประมวลผลที่แตกต่างกันไปตามพันธกิจและชั้นความลับของข้อมูล อีกทั้งยังมีความพร้อมด้านสารสนเทศที่ไม่เท่ากัน โครงสร้างพื้นฐานด้านการประมวลผลจึงจะมีความแตกต่างกันออกไป ดังนั้นเพื่อให้การออกแบบโครงสร้างพื้นฐานมีความเหมาะสมกับการใช้งานของแต่ละหน่วยงาน คณะกรรมการขับเคลื่อนการดำเนินนโยบายเพื่อใช้ประโยชน์ข้อมูลขนาดใหญ่ (Big Data) ศูนย์ข้อมูล (Data Center) และคลาวด์คอมพิวติ้ง (Cloud Computing) จึงได้ขอให้หน่วยงานภาครัฐต่างๆ ทำการสำรวจโครงสร้างพื้นฐานที่มีอยู่เทียบกับมาตรฐานสากล และนำข้อมูลดังกล่าวเข้าร่วมประเมินวิเคราะห์หาช่องว่าง (Gap Analysis) ร่วมกับผู้เชี่ยวชาญในคณะกรรมการออกแบบสถาปัตยกรรม (Architecture Design) ระบบบูรณาการข้อมูลภาครัฐ ซึ่งผลจากการวิเคราะห์ได้ถูกนำมาใช้วางแผนเพื่อปรับปรุงโครงสร้างพื้นฐานสำหรับแต่ละหน่วยงาน รวมไปถึงการวางแผนการโอนย้ายระบบงานเก่า (Legacy System) เข้าสู่เทคโนโลยีใหม่ และการวางแผนการพัฒนาบุคลากรสารสนเทศอีกด้วย

การดำเนินงานออกแบบโครงสร้างการประมวลผลพื้นฐานในครั้งนี้ มีหน่วยงานภาครัฐเข้าร่วมให้ข้อมูลเพื่อประเมินวิเคราะห์หาช่องว่างจำนวน 159 หน่วยงาน ข้อมูลจากการสำรวจและสัมภาษณ์ได้ถูกนำมาออกแบบกรอบโครงสร้างสารสนเทศและการประมวลผลข้อมูลพื้นฐาน กรอบดังกล่าวถูกออกแบบโดยมีวัตถุประสงค์หลักดังนี้

- เพื่อให้หน่วยงานภาครัฐสามารถเข้าถึงทรัพยากรคอมพิวเตอร์ได้อย่างรวดเร็วและทั่วถึง ทันต่อความต้องการในการเปลี่ยนแปลงเข้าสู่รัฐบาลดิจิทัล

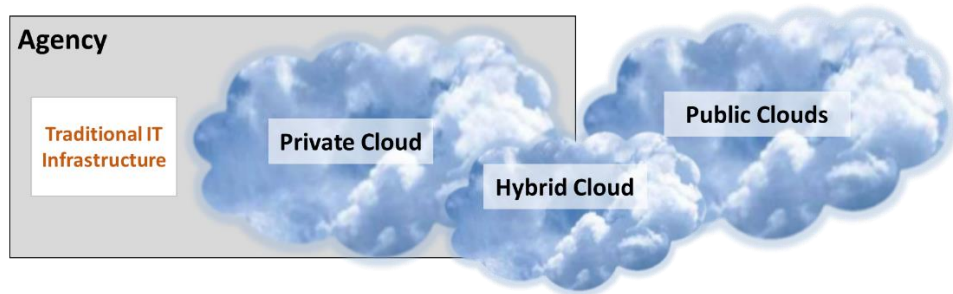
- เพื่อเพิ่มความต่อเนื่องในการให้บริการระบบงานภาครัฐ (Availability) ให้ถึงมาตรฐานสากลคือ 99.9%
- เพื่อให้การดูแลรักษาความปลอดภัยระบบงาน และข้อมูลที่มีชั้นความลับเป็นไปตามมาตรฐานสากล
- เพื่อให้เกิดการแลกเปลี่ยนข้อมูลข้ามหน่วยงานได้อย่างมีมาตรฐานและเป็นระบบ (Repeatable process) รองรับการใช้งานการวิเคราะห์และใช้ประโยชน์ข้อมูลขนาดใหญ่ (Big Data)
- เพื่อประหยัดงบประมาณด้านโครงสร้างพื้นฐานการประมวลผล

1.1 กรอบมาตรฐานดาต้าเซนเตอร์/คลาวด์ภาครัฐ

กรณีที่ระบบงานของหน่วยงานภาครัฐไม่มีชั้นความลับข้อมูล หน่วยงานสามารถจัดหาหรือเช่าซื้อทรัพยากรประมวลผลได้อย่างอิสระ ผู้ให้บริการจะเป็นองค์กรใดก็ได้ อย่างไรก็ตามในกรณีที่ระบบงานมีชั้นความลับข้อมูลระดับ ข้อมูลความลับราชการ หรือ ข้อมูลส่วนบุคคล ผู้ให้บริการการประมวลผลจะต้องเป็นองค์กรที่รัฐบาลสามารถกำกับดูแลได้ ส่วนในกรณีที่ระบบงานมีชั้นความลับระดับ ข้อมูลความมั่นคง ผู้ให้บริการการประมวลผลจะต้องเป็นองค์กรที่รัฐบาลสามารถกำกับดูแลได้ที่มีมาตรฐานเรื่องความปลอดภัยสูงมากทั้งความปลอดภัยพื้นที่และความปลอดภัยเครือข่าย

ระบบงานและฐานข้อมูลที่มีชั้นความลับของภาครัฐจะต้องถูกติดตั้งอยู่ในดาต้าเซนเตอร์ที่ได้มาตรฐานสากลทั้งในด้านพื้นที่ แหล่งจ่ายไฟ การรักษาความปลอดภัยข้อมูล (Data Protection) และความปลอดภัยไซเบอร์ (Cyber Security) รวมทั้งมีการรับประกันบริการ (Service Level Agreement - SLA) ที่เหมาะสม สามารถบริหารจัดการเครื่องประมวลผลและที่จัดเก็บข้อมูลให้ได้ตามชั้นความลับของระบบงานและฐานข้อมูล โดยหน่วยงานภาครัฐในระดับต่างๆสามารถเลือกใช้บริการดาต้าเซนเตอร์/คลาวด์ได้ ดังนี้

- ดาต้าเซนเตอร์ที่ได้มาตรฐาน tier 4 (หรือใกล้เคียงโดยยกเว้นเรื่องการจ่ายไปจาก 2 แหล่ง) ที่หน่วยงานจะสร้างขึ้นหรือมีอยู่แล้ว ซึ่งหน่วยงานจะกำกับดูแลพื้นที่ ระบบประมวล ระบบเครือข่าย และระบบรักษาความปลอดภัยด้วยตัวเอง
- Outsource บริการดาต้าเซนเตอร์/คลาวด์ ไปยังผู้ให้บริการคลาวด์สาธารณะ โดยระบบงานและฐานข้อมูลที่มีชั้นความลับต้องใช้บริการจากองค์กรที่รัฐบาลสามารถกำกับดูแลได้ ในขณะที่ระบบงานและฐานข้อมูลที่ไม่มีชั้นความลับ (ข้อมูลสาธารณะ) สามารถถูกนำไปติดตั้งในพื้นที่ของผู้ให้บริการจากภาคเอกชนหรือจากต่างประเทศได้อย่างอิสระ ทั้งนี้หน่วยงานภาครัฐสามารถที่จะเลือกผู้ให้บริการดาต้าเซนเตอร์/คลาวด์สำหรับระบบงานต่างๆได้เอง อาจจะใช้ผู้ให้บริการหลายเจ้าได้แต่ให้เป็นไปตามเกณฑ์เรื่องชั้นความลับข้อมูลที่กล่าวในข้างต้น
- สร้างระบบคลาวด์ผสมเอง โดยมีทั้งดาต้าเซนเตอร์และระบบคลาวด์ส่วนตัว จากนั้นมีบางส่วนเชื่อมต่อไปยังคลาวด์สาธารณะเป็นแบบคลาวด์ผสม โดยมีการบริหารจัดการเรื่องชั้นความลับของระบบงานและฐานข้อมูลให้เป็นไปตามเกณฑ์เรื่องชั้นความลับข้อมูลที่กล่าวในข้างต้น รูปแบบสถาปัตยกรรมแสดงในรูปที่ 1.1



รูปที่ 1.1 คลาวด์แบบผสม (Hybrid Cloud)

- Outsource บริการดาต้าเซ็นเตอร์/คลาวด์ ไปยังระบบคลาวด์กลางภาครัฐ (Government Cloud) ซึ่งในกรณีนี้กระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคมจะกำกับดูแลผู้ให้บริการคลาวด์ภาครัฐอีกต่อหนึ่ง โดยผู้ให้บริการคลาวด์ภาครัฐสามารถบริหารจัดการเช่าทรัพยากรต่อจากภาคเอกชนเพื่อสร้างเป็น Multi-cloud สำหรับให้บริการระบบงานและฐานข้อมูลที่ไม่มีชั้นความลับได้ โดยประเด็นเรื่องชั้นความลับข้อมูลจะถูกจัดการโดยอัตโนมัติ หน่วยงานภาครัฐไม่จำเป็นต้องบริหารผู้ให้บริการคลาวด์เอง ในกรณีที่หน่วยงานเลือกที่จะบริหารจัดการการเช่าซื้อระบบคลาวด์เอง ขอให้อ้างอิงราคากลางที่กำหนดโดยกระทรวงดิจิทัลฯ ในการจัดซื้อดังนี้

บริการเครื่องเสมือน (VM) โดยใช้ไฮเปอร์ไวเซอร์แบบลิขสิทธิ์จากต่างประเทศ ค่าเช่าเฉพาะเครื่องเสมือน ขนาดมาตรฐานราคาไม่เกิน **12,250 บาท/เดือน/เครื่อง** โดยสำหรับเครื่องเสมือนที่มีรายละเอียดสเปกแตกต่างออกไปให้คำนวณโดยปรับจากค่ากลาง 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) และ ฐานข้อมูล (Database) แบบ Opensource
- บริการเสริมด้านลิขสิทธิ์ซอฟต์แวร์จากต่างประเทศ
 - Windows Server 2016 ราคา 800 บาทต่อเดือน
 - Redhat Server ราคา 800 บาทต่อเดือน
 - Microsoft SQL STD edition ราคา 12,000 บาทต่อเดือน
 - Microsoft SQL ENT edition ราคา 50,000 บาทต่อเดือน

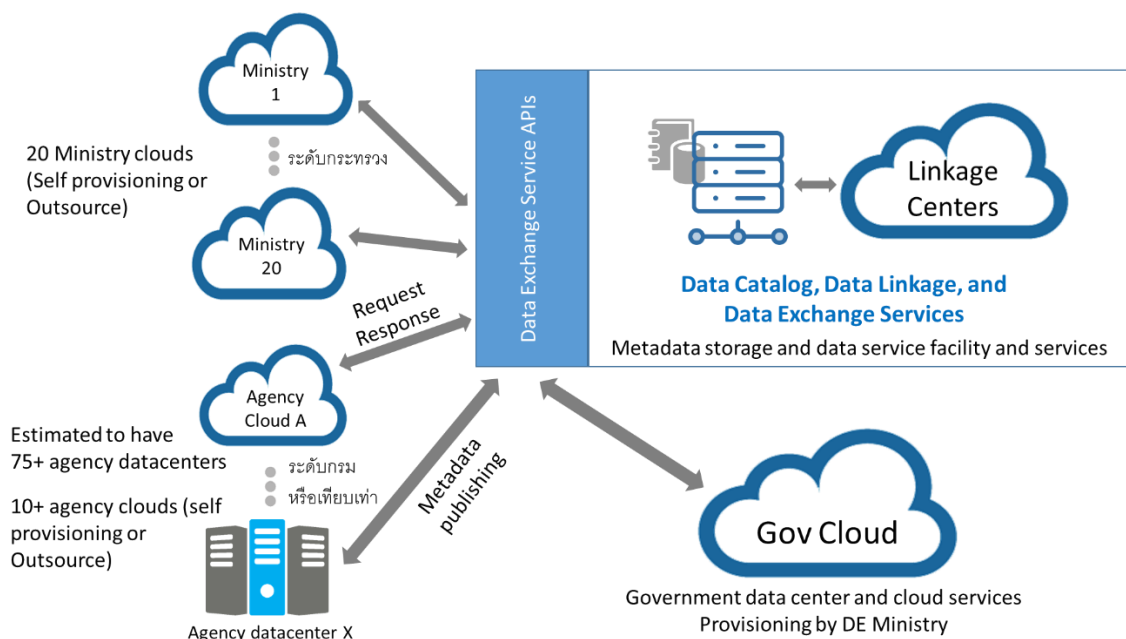
บริการเครื่องเสมือน (VM) โดยใช้ไฮเปอร์ไวเซอร์แบบลิขสิทธิ์ของบริษัทไทยที่คนไทยพัฒนา ต่อยอดจากไฮเปอร์ไวเซอร์โอเพ่นซอร์สตัวอย่างเช่น Openstack ค่าเช่าเฉพาะเครื่องเสมือนขนาดมาตรฐาน ไม่เกิน **8,000 บาท/เดือน/เครื่อง** โดยสำหรับเครื่องเสมือนที่มีรายละเอียดสเปกแตกต่างออกไปให้คำนวณ โดยปรับจากค่ากลาง รายละเอียดการให้บริการมาตรฐานอธิบายได้ดังต่อไปนี้

- 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) และ ฐานข้อมูล (Database) แบบ Opensource
- เป็นระบบคลาวด์แบบ Multi-region แบบเดียวกับ Microsoft Azure, Google Cloud, Amazon AWS, และ GDCC ของกระทรวงดิจิทัลฯ
- บริการเสริมด้านลิขสิทธิ์ซอฟต์แวร์จากต่างประเทศ
 - Windows Server 2016 ราคา 800 บาทต่อเดือน
 - Redhat Server ราคา 800 บาทต่อเดือน
 - Microsoft SQL STD edition ราคา 12,000 บาทต่อเดือน
 - Microsoft SQL ENT edition ราคา 50,000 บาทต่อเดือน

หมายเหตุ share ratio Physical: vCPU = 1: 2-4

1.2 โครงสร้างพื้นฐานเพื่อการประมวลผลของหน่วยงานภาครัฐ

จากข้อมูลที่ได้จากการสำรวจและสัมภาษณ์หน่วยงานเกือบ 160 แห่ง กรอบโครงสร้างสารสนเทศและการประมวลผลข้อมูลพื้นฐานภาครัฐได้ถูกออกแบบ ดังแสดงในรูปที่ 1.2



รูปที่ 1.2 สถาปัตยกรรมโครงสร้างพื้นฐานเพื่อการประมวลผลของหน่วยงานภาครัฐ

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

1. ดาต้าเซนเตอร์/คลาวด์ระดับกรมหรือหน่วยงานเทียบเท่า (Agency Cloud) ข้อมูลจากการสำรวจประเมินทรัพยากรและบุคลากรของหน่วยงานจากผู้เชี่ยวชาญในคณะกรรมการออกแบบสถาปัตยกรรมระบบบูรณาการข้อมูลภาครัฐ พบว่าภาครัฐมีหน่วยงานระดับกรมหรือเทียบเท่าจำนวน 90 (± 5) แห่ง (ร่างรายชื่อในเอกสารแนบ ก1) ที่มีดาต้าเซนเตอร์หรือใช้บริการระบบคลาวด์อยู่แล้วโดยมีการใช้งานหน่วยประมวลผลรวมอยู่ที่ประมาณ 35,000 คอร์ในปัจจุบัน ซึ่งหน่วยงานเหล่านี้มีความประสงค์จะดูแลดาต้าเซนเตอร์ และ/หรือ คลาวด์เพื่อให้บริการระบบงานของหน่วยงานเอง โดยแบ่งได้เป็น 2 รูปแบบดังนี้

1. มีดาต้าเซนเตอร์อยู่แล้วและวางแผนที่จะพัฒนาต่อยอดให้ดาต้าเซนเตอร์ได้มาตรฐานที่ใกล้เคียง tier 4 และขยายกำลังการประมวลผลด้วยการขยายดาต้าเซนเตอร์ โดยหน่วยงานกลุ่มนี้มีประมาณ 75 (± 5) แห่ง และมีความต้องการปริมาณหน่วยประมวลผลเพิ่มเติมรวมสำหรับปีงบประมาณ 2563 ประมาณ 7,000 ซีพียูคอร์ และเพิ่มอีกปีละ 3,500 ซีพียูคอร์ ในปีงบประมาณ 2564 และ 2565 (หน่วยประมวลผลที่มีอยู่เดิมจะต้องมีการปรับเปลี่ยนทุก 5 ปี)

หมายเหตุ ตัวเลขรวมเฉพาะหน่วยงานที่ให้ข้อมูลเรื่องทรัพยากรการประมวลผลแบบละเอียด และไม่รวม supercomputer ของกระทรวงวิทยาศาสตร์และเทคโนโลยี

2. มีดาต้าเซนเตอร์อยู่แล้วและวางแผนที่จะขยายกำลังการประมวลผลไปยังระบบคลาวด์ ดังแสดงในรูปที่ 1.1 หรือวางแผนย้ายไประบบคลาวด์ทั้งหมดผ่านการเช่าใช้คลาวด์จากผู้ให้บริการคลาวด์สาธารณะที่รัฐบาลกำกับดูแลได้หรือระบบคลาวด์กลางภาครัฐ ซึ่งหน่วยงานในกลุ่มนี้มีอีกประมาณ 10 (± 5) แห่ง และมีปริมาณความต้องการปริมาณหน่วยประมวลผลในปีที่ 1 (งบประมาณ 2563) คือ 6,000 VM และเพิ่มเป็น 9,000 VM และ 12,000 VM ในปีงบประมาณ 2564 และ 2565 ตามลำดับ

หมายเหตุ ตัวเลขรวมเฉพาะหน่วยงานที่ให้ข้อมูลเรื่องทรัพยากรการประมวลผลแบบละเอียด

2. ดาต้าเซนเตอร์/คลาวด์ระดับกระทรวง (Ministry Cloud) จะมีแผนสร้างทั้งสิ้น 20 แห่ง ซึ่งมีจุดประสงค์ในการให้บริการกรมและหน่วยงานภายใต้กระทรวงที่ไม่ประสงค์จะสร้างและดูแลดาต้าเซนเตอร์/คลาวด์เอง (จำนวนหน่วยงานเสนอในเอกสารแนบ ก2) โดยคลาวด์กระทรวงส่วนมากจะถูกสร้างขึ้นโดยการเช่าใช้บริการพื้นที่และทรัพยากรการประมวลผล และจัดเก็บข้อมูลจากผู้ให้บริการคลาวด์สาธารณะ (Public Cloud) ที่รัฐบาลสามารถกำกับดูแลได้ หรือระบบคลาวด์กลางภาครัฐ (GDCC) และนำทรัพยากรที่เช่าใช้เหล่านั้นมาให้บริการต่อหน่วยงานภายใต้กระทรวง สำหรับบางกระทรวงที่มีดาต้าเซนเตอร์อยู่แล้ว จะสามารถ

ให้บริการหน่วยงานภายใต้กระทรวงได้ในลักษณะคลาวด์ส่วนตัว (Private Cloud) และยังสามารถขยายโดยการต่อเชื่อมไปยังคลาวด์สาธารณะ แสดงในรูปที่ 1.1 ได้อีกด้วย รูปแบบบริการ คือ

a. Co-location : หน่วยงานภายใต้กระทรวงนำทรัพยากรประมวลผลที่มีอยู่แล้ว ไปติดตั้งในสถานที่ของผู้ให้บริการคลาวด์ระดับกระทรวง เพื่อเพิ่มมาตรฐานด้านความปลอดภัยและความต่อเนื่องบริการของระบบงาน และลดภาระเรื่องการดูแลทรัพยากรพื้นฐาน

b. Infrastructure-as-a-Service (IaaS) : ผู้ดูแลระบบคลาวด์กระทรวงจัดสรรเครื่องประมวลผลเสมือน (Virtual Machine) ที่เก็บข้อมูล และเครือข่ายเชื่อมต่อ ให้กับระบบงาน ตามที่หน่วยงานภายใต้กระทรวงร้องขอ

c. Software-as-a-service (SaaS) : ผู้ดูแลคลาวด์กระทรวงจัดหาหรือพัฒนาระบบซอฟต์แวร์กลางเพื่อติดตั้งในระบบคลาวด์ และจัดสรรรหัสผู้ใช้สำหรับให้หน่วยงานภายใต้กระทรวงเข้าใช้งานระบบซอฟต์แวร์กลางร่วมกัน

ปริมาณหน่วยประมวลผลที่ผู้ให้บริการภาครัฐต้องเตรียมพร้อมให้สำหรับกระทรวงต่างๆ ในปีที่ 1 (งบประมาณ 2563) คือ 11,000 VM และเพิ่มเป็น 16,500 VM และ 22,000 VM ในปีงบประมาณ 2564 และ 2565 ตามลำดับ

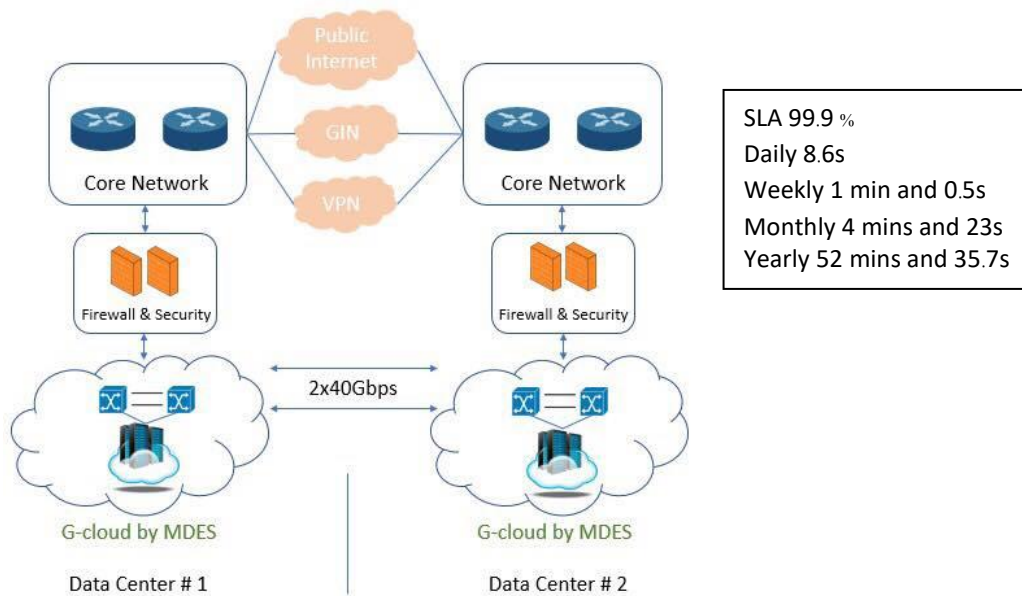
หมายเหตุ ตัวเลขรวมเฉพาะหน่วยงานที่ให้ข้อมูลเรื่องทรัพยากรการประมวลผลแบบละเอียด

3. ดาต้าเซนเตอร์กลางภาครัฐเพื่อให้บริการข้อมูล (Data Linkage and Exchange Services) : เพื่อให้การใช้ประโยชน์ข้อมูลภาครัฐในอนาคตเกิดได้อย่างเป็นระบบ สะดวกรวดเร็ว และเป็นไปตามข้อบังคับ/ข้อกำหนดของหน่วยงานเจ้าของข้อมูล ภาครัฐจึงควรสร้างดาต้าเซนเตอร์กลางสำหรับให้บริการเป็นจุดเชื่อมโยงและแลกเปลี่ยนข้อมูล ซึ่งจากการสำรวจศึกษาและสัมภาษณ์โดยคณะทำงานในคณะกรรมการออกแบบสถาปัตยกรรมระบบบูรณาการข้อมูลภาครัฐ พบว่าในปัจจุบันหน่วยงานที่มีดาต้าเซนเตอร์ที่ให้บริการเชื่อมโยงข้อมูลนั้นมีอยู่ทั้งสิ้น 2 หน่วยงานคือ 1) ระบบ Data linkage ของกรมการปกครอง (ข้อมูลบุคคล) และ 2) ระบบ National Single Window ของกรมศุลกากร (ข้อมูลนำเข้า/ส่งออก) ดังนั้นการลงทุนขยายบริการที่มีอยู่เพื่อให้สามารถรองรับการแลกเปลี่ยนข้อมูลสำหรับการวิเคราะห์ข้อมูลขนาดใหญ่ (Big Data Analytics) และการให้บริการประชาชนแบบ One Stop Services จึงเป็นการลงทุนที่คุ้มค่า โดยเสนอให้ทั้ง 2 หน่วยงานพิจารณาขยายบริการในปีงบประมาณ 2563 และ 2564 เพื่อให้ครอบคลุม สำหรับข้อมูลส่วนอื่น เช่น ข้อมูลนิติบุคคลในด้านอื่นนอกเหนือจากการนำเข้า/ส่งออก ข้อมูลด้านพลังงาน ฯลฯ ขอให้หน่วยงานที่เกี่ยวข้องหรือได้รับมอบหมาย ให้สร้างดาต้าเซนเตอร์เพื่อให้บริการเชื่อมโยงข้อมูลเพิ่มเติมให้ครบสมบูรณ์ในอนาคต นอกจากนี้ ในปีงบประมาณ 2563 ภาครัฐควรสร้างบริการแคตตาล็อกข้อมูลกลาง (National Data Catalog) และบริการสืบค้นข้อมูล (Directory Service) เพิ่มเติมเพื่อให้การใช้ประโยชน์ข้อมูลเกิดขึ้นได้อย่างเป็นระบบ โดยมีกระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคมเป็นผู้ให้บริการ

4. คลาวด์กลางภาครัฐ (Government Cloud) ซึ่งจะถูกสร้างขึ้นโดย บริษัท กสท โทรคมนาคม จำกัด (มหาชน) ภายใต้การกำกับดูแลของกระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม เพื่อให้ประเทศมีระบบกลางในการให้บริการคลาวด์แก่หน่วยงานภาครัฐที่มีมาตรฐานและปลอดภัย หน่วยงานภาครัฐจะสามารถเข้าถึงทรัพยากรคอมพิวเตอร์ได้อย่างรวดเร็วทันต่อความต้องการในการเปลี่ยนแปลงเข้าสู่รัฐบาลดิจิทัล บริการของคลาวด์กลางภาครัฐจะมี 2 ระดับ คือ การให้บริการเต็มรูปแบบ (Managed Service) และการให้บริการผ่าน Self Service Portal ผู้ใช้จะสามารถเข้าถึงและปรับบริการได้เอง โดยรายละเอียดจะนำเสนอในหัวข้อต่อไป

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

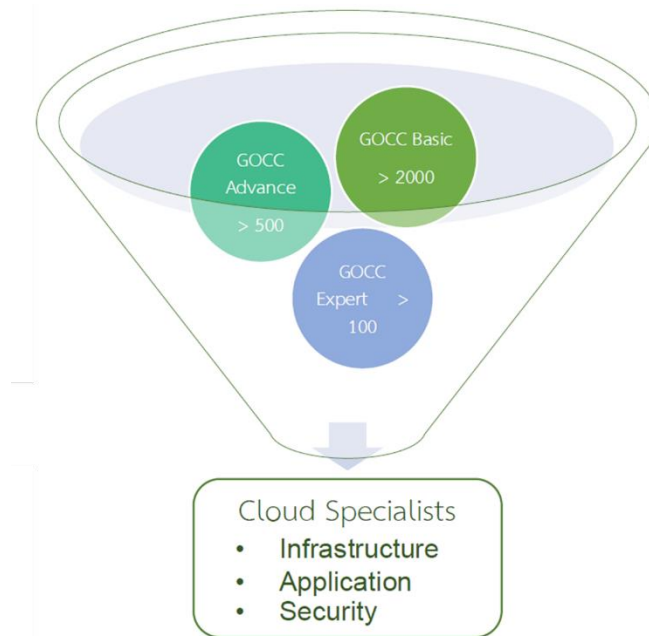
ระบบคลาวด์กลางภาครัฐจะถูกออกแบบและพัฒนาขึ้นโดย บริษัท กสท โทรคมนาคม จำกัด (มหาชน) ภายใต้การกำกับดูแลของกระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม ซึ่งจะเป็นระบบกลางในการให้บริการคลาวด์สำหรับหน่วยงานภาครัฐที่มีมาตรฐานและปลอดภัย หน่วยงานภาครัฐสามารถเข้าถึงทรัพยากรคอมพิวเตอร์ได้อย่างรวดเร็ว ทัวถึง และทันต่อความต้องการในการเปลี่ยนแปลงเข้าสู่รัฐบาลดิจิทัล รวมทั้งเป็นการรองรับการใช้งานการวิเคราะห์และใช้ประโยชน์ข้อมูลขนาดใหญ่อีกด้วย โดยระบบคลาวด์กลางภาครัฐที่สร้างขึ้นจะสามารถให้บริการผ่านเครือข่ายได้หลายประเภท รวมถึงเครือข่ายกลางภาครัฐ (GIN) เครือข่ายสาธารณะ (Public) และเครือข่ายส่วนตัว (Private/VPN) และมีการออกแบบให้บริการมีความต่อเนื่องสูงให้ถึง 99.9% โดยดาต้าเซ็นเตอร์อย่างน้อย 2 แห่งจะถูกใช้ในการให้บริการระบบคลาวด์กลางร่วมกัน เมื่อดาต้าเซ็นเตอร์หนึ่งเกิดปัญหาขึ้นระบบงานและฐานข้อมูลจะถูกปรับให้ทำงานบนอีกดาต้าเซ็นเตอร์หนึ่งได้อย่างรวดเร็ว นอกจากนี้ ระบบคลาวด์กลางภาครัฐจะมีข้อตกลงเรื่องการบริการ (Service Level Agreement - SLA) ที่เหมาะสม มีผู้ดูแลระบบแบบ 24x7 มีการบันทึกข้อมูล (Log) และให้บริการหน้าจอแสดงผลข้อมูลการใช้งาน (Dashboard) ของทุกระบบงานและฐานข้อมูลเพื่อให้หน่วยงานต่างๆสามารถทำรายงานสรุปการใช้ประโยชน์ทรัพยากรประกอบการพิจารณาปรับปริมาณทรัพยากรหรือเพิ่มประสิทธิภาพการใช้งานและงบประมาณโดยรวมอีกด้วย สถาปัตยกรรมระบบแสดงในรูปที่ 1.3



รูปที่ 1.3 สถาปัตยกรรมระบบคลาวด์กลางภาครัฐ

จากการอ้างอิงข้อมูลการสำรวจประเมินทรัพยากรของหน่วยงานโดยผู้เชี่ยวชาญในคณะกรรมการออกแบบสถาปัตยกรรมระบบบูรณาการข้อมูลภาครัฐนั้น คลาวด์กลางภาครัฐหรือ GDCC จะถูกวางแผนจะให้บริการแก่หน่วยงานภาครัฐดังนี้ 1. ให้บริการ คลาวด์ระดับกระทรวง อย่างน้อย 10 แห่งในเบื้องต้น 2. ให้บริการหน่วยงานภาครัฐบางส่วนที่ประสบปัญหาล่าช้าในการจัดซื้อเพื่อขยายดาต้าเซ็นเตอร์ หรือเช่าคลาวด์เพื่อขยายระบบงาน (จัดสรรเตรียมให้ 25 % ของปริมาณหน่วยประมวลผลรวม) 3. จัดสรรเพื่อเตรียมรองรับหน่วยงานภาครัฐที่ระบุว่าต้องการใช้บริการคลาวด์กลางภาครัฐ โดยปริมาณหน่วยประมวลผลที่คลาวด์กลางภาครัฐจะต้องเตรียมพร้อมเริ่มต้นจากขนาด 4,000 VM ในปีงบประมาณ 2562 และขยับขึ้นเป็น 8,000 12,000 และ 20,000 VM ในปีงบประมาณ 2563 2564 และ 2565 ตามลำดับ (การประมาณการโดย Extrapolation) ส่วนในปีงบประมาณถัดไปควรพิจารณาการใช้งานในช่วงแรกก่อนวางแผนขยาย โดยบริษัท กสท โทรคมนาคม จำกัด (มหาชน) จะให้บริการตามราคาและสเปคผู้ให้บริการกลางดังที่ระบุในหัวข้อที่ 1.1

ทั้งนี้เพื่อให้การใช้ประโยชน์ระบบคลาวด์กลางภาครัฐเกิดขึ้นได้อย่างเป็นรูปธรรมกระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคมจะให้บริการการฝึกอบรมเจ้าหน้าที่สารสนเทศภาครัฐในเรื่องการใช้งานระบบคลาวด์ โดยแบ่งออกเป็น 3 กลุ่มดังแสดงในรูปที่ 1.4 จากรูปการฝึกอบรมจะมีการออกไปเกียรติบัตร 3 ระดับคือ 1) ระดับพื้นฐานที่สามารถใช้งานระบบคลาวด์ได้ จำนวน 2,000 คน ในระยะเวลา 3 ปี 2) ระดับกลางที่สามารถใช้งานระบบและดูแลระบบคลาวด์ขั้นพื้นฐานได้ จำนวน 500 คนในระยะเวลา 3 ปี และ 3) ระดับสูงที่สามารถใช้งานระบบ ดูแลระบบคลาวด์ขั้นสูงและสามารถสอนคนอื่นต่อไปได้ เป็นจำนวน 100 คนในช่วงปีงบประมาณ 2563



รูปที่ 1.4 คลาสฝึกอบรมการใช้คลาวด์กลางภาครัฐใน 3 ระดับ

สำหรับหน่วยงานที่ไม่มีเจ้าหน้าที่สารสนเทศ ทางสำนักงานคณะกรรมการดิจิทัลเพื่อเศรษฐกิจและสังคมแห่งชาติ พร้อมจะทำเรื่องเพื่อสนับสนุนการขยายอัตรากำลังไปยังสำนักงานคณะกรรมการข้าราชการพลเรือนให้กับหน่วยงานภาครัฐ

นอกจากนี้กระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคมจะร่วมกับบริษัท กสท โทรคมนาคม จำกัด ในการให้บริการโอนย้ายระบบงานขึ้นสู่ระบบคลาวด์กลางภาครัฐ โดยหน่วยงานภาครัฐที่จะนำระบบงานขึ้นสู่คลาวด์ ควรจะต้องจัดเตรียมงบประมาณในการโอนย้ายระบบ ซึ่งค่าใช้จ่ายจะขึ้นอยู่กับความซับซ้อนของระบบงาน การโอนย้ายระบบแบ่งออกได้เป็น 3 ประเภท ดังนี้

1. Re-Deploy เป็นการย้ายระบบซอฟต์แวร์ที่ใช้อยู่ในปัจจุบันไปประมวลผลบนระบบคลาวด์โดยไม่มีการออกแบบใหม่หรือพัฒนาเพิ่มเติม ซึ่งการโอนย้ายลักษณะนี้ หากเป็นการโอนย้ายจาก VM ในดาต้าเซนเตอร์ ไปยัง VM บนคลาวด์ โดยใช้ VM เวอร์ชันเดิมของบริษัทเดียวกัน จะมีราคาค่าใช้จ่ายกลางอยู่ที่ **4,000 บาท/ครั้ง** หากในกรณีการโอนย้ายจากเครื่องแม่ข่าย (Physical Server) ไปสู่ VM หรือ VM ข้ามเวอร์ชันและบริษัท ซึ่งมีความยุ่งยากสูงกว่าจะมีราคาค่าใช้จ่ายกลางอยู่ที่ **4,000 – 10,000 บาท/ครั้ง** ขึ้นอยู่กับความยากง่ายและจำนวน VM
2. Re-Architect ในกรณีที่ระบบงานอยู่บนเทคโนโลยีเก่า (Legacy System) การโอนย้ายระบบงานมักจะต้องออกแบบสถาปัตยกรรมระบบซอฟต์แวร์ใหม่ ซึ่งส่งผลให้บางโมดูลต้องถูกพัฒนาใหม่ บางส่วนหรือทั้งหมดก่อนที่จะสามารถวางระบบซอฟต์แวร์ลงบนคลาวด์ได้ การโอนย้ายจึงมีความยุ่งยากสูง และค่าใช้จ่ายสูงขึ้นอยู่กับเทคโนโลยี หน่วยงานจึงควรประเมินราคาและตั้งงบประมาณล่วงหน้า

3. Replace คือการใช้บริการซอฟต์แวร์ที่จัดให้โดยผู้ให้บริการบนระบบคลาวด์ ทดแทนการใช้ซอฟต์แวร์เดิมขององค์กร การทดแทนระบบซอฟต์แวร์ในลักษณะนี้ ผู้ดูแลระบบจะมีภาระในการโอนย้าย ฐานข้อมูลเข้าสู่ระบบใหม่ หน่วยงานควรประเมินราคาและตั้งงบประมาณล่วงหน้าในกรณีที่ข้อมูลมีความซับซ้อนสูง

บทที่ 2 รายการข้อมูลภาครัฐ (Government Data Catalog)

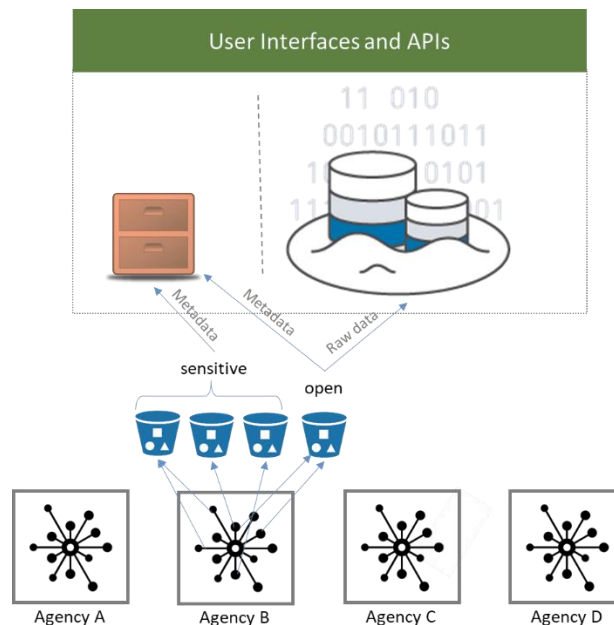
จากความต้องการใช้ประโยชน์ข้อมูลภาครัฐเพิ่มมากขึ้นในปัจจุบัน และผู้ใช้ข้อมูลที่มีอยู่หลากหลาย ทั้งกลุ่มเจ้าหน้าที่ภาครัฐ นักวิจัยจากมหาวิทยาลัย นักวิชาการ/นักวิเคราะห์ข้อมูลทั้งภาคเอกชนและภาคประชาสังคม รวมถึงประชาชนทั่วไป อย่างไรก็ตาม ปริมาณข้อมูลที่ถูกผลิตขึ้นอย่างต่อเนื่องทำให้การเลือกใช้ข้อมูลและการสืบค้นหาแหล่งข้อมูลที่เหมาะสมสำหรับการวิเคราะห์หรือให้บริการมีความซับซ้อนและยุ่งยากสูง ดังนั้นเพื่อให้การใช้ประโยชน์ข้อมูลภาครัฐเกิดขึ้นได้อย่างเป็นรูปธรรม กระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม จึงวางแผนที่จะพัฒนารายการข้อมูลภาครัฐ (Government Data Catalog) และระบบสารสนเทศเพื่อการสืบค้น ดังจะนำเสนอต่อไป

2.1 บริการรายการข้อมูลภาครัฐ (Government Data Catalog Service)

ในกระบวนการการวิเคราะห์ข้อมูลขนาดใหญ่ นักวิทยาศาสตร์ข้อมูลจำเป็นต้องสำรวจหาแหล่งข้อมูลสำคัญตามหัวข้อการวิเคราะห์และทำความเข้าใจการใช้ข้อมูลในมิติต่างๆ รายการข้อมูลภาครัฐ และเครื่องมือสารสนเทศข้อมูลจึงเป็นสิ่งจำเป็นเพื่อช่วยให้ผู้ใช้ประโยชน์ข้อมูลสามารถสืบค้น ร้องขอ เข้าถึง ทราบแหล่งที่มา ทราบถึงชั้นความลับ ประเภท รูปแบบ และสามารถใช้บริการของข้อมูลภาครัฐทั้งหมดได้ รายการข้อมูลดังกล่าวจะเป็นเหมือนสมุดหน้าเหลือง (Yellow pages) ของข้อมูลภาครัฐที่สำคัญทั้งหมดของประเทศไทย ซึ่งเป็นจุดเริ่มต้นที่จำเป็นอย่างยิ่งในการพัฒนาการใช้ประโยชน์ข้อมูลภาครัฐให้มีประสิทธิภาพและเป็นไปได้ในทิศทางเดียวกัน สามารถบูรณาการ ให้บริการ และใช้ประโยชน์ข้อมูลข้ามหน่วยงานได้อย่างเป็นระบบ

การจัดทำรายการข้อมูลภาครัฐ ดังแสดงในรูปที่ 2.1 เป็นการรวบรวมคำอธิบายรายการข้อมูลหรือเมทาดาทา (Metadata) ของข้อมูลสำคัญจากหน่วยงานภาครัฐทั้งหมดมาสร้างเป็นแคตตาล็อกเพื่อประโยชน์ในการสืบค้นผ่านระบบสารสนเทศบริการนามุกรมข้อมูลภาครัฐ (Directory Services) โดยเมทาดาทาที่รวบรวมจากทุกหน่วยงานควรมีรูปแบบเดียวกัน ซึ่งกระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคมจะทำหน้าที่ในการออกแบบกระบวนการและเครื่องมือเพื่อจัดทำมาตรฐานเมทาดาทาและรวบรวมเมทาดาทา รายการข้อมูลที่สำคัญจากหน่วยงานรัฐทั้งหมด โดยศึกษา ค้นคว้า เปรียบเทียบและวิเคราะห์ กรณีศึกษาด้านการจัดทำมาตรฐานเมทาดาทาของประเทศอื่น เพื่อปรับให้เหมาะสมและเอื้อต่อการแลกเปลี่ยนข้อมูลในประเทศไทย หน่วยงานภาครัฐต่างๆ จะสามารถนำมาตรฐานดังกล่าวไปใช้งานและจัดทำเมทาดาทาของหน่วยงานได้อย่างสอดคล้องกัน

นอกจากนี้กระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคมจะทำการออกแบบกลไกด้านข้อมูลที่หน่วยงานภาครัฐสามารถนำไปประยุกต์ใช้ในเรื่องการกำกับดูแลข้อมูล การให้บริการข้อมูล และการใช้ประโยชน์ข้อมูลอีกด้วย



รูปที่ 2.1 การให้บริการรายการข้อมูลภาครัฐ

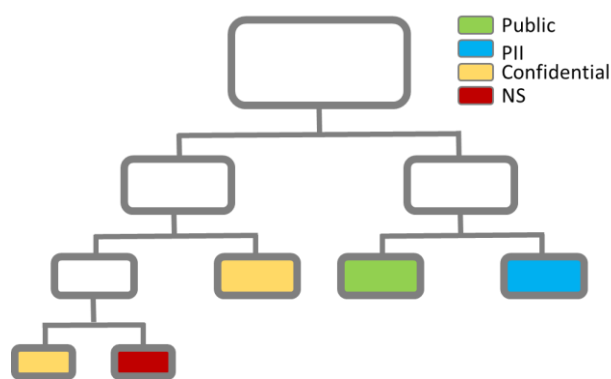
จากรูป 2.1 กระบวนการดิจิทัลเพื่อเศรษฐกิจและสังคมจะดำเนินการสร้างระบบการให้บริการรายการข้อมูลภาครัฐ ดังนี้

1) ออกแบบกระบวนการจัดทำมาตรฐานเมทาดาดาและรูปแบบที่เป็นมาตรฐาน (Template) รวมทั้งแนะนำให้หน่วยงานภาครัฐจัดทำเมทาดาดาของข้อมูลสำคัญตามแบบที่กำหนด ซึ่งควรมีรูปแบบเดียวกันเพื่อความสะดวกในการสืบค้นและแลกเปลี่ยนข้อมูล โดยเมทาดาดาที่จะจัดทำขึ้นควรมีการระบุข้อมูลต่างๆ ดังนี้ มาตรฐานของรูปแบบเมทาดาดา (Metadata model) คำอธิบายรายการข้อมูล (Data Description) แหล่งที่มาของข้อมูล (Data Source) หน่วยงานเจ้าของข้อมูล (Data Owner) วิธีการ/กระบวนการเก็บรวบรวมข้อมูล (Collection Method) จุดประสงค์การเก็บข้อมูลของหน่วยงานที่เป็นเจ้าของ (Purposes of Data Collection) ประเภทข้อมูล (Type) และรูปแบบการจัดเก็บข้อมูล (Format) คำกำกับ (Tags) โดเมนทางธุรกิจ (Business domain) และคำสำคัญ (Keywords) (รายการในเมทาดาดาอาจจะมีการปรับเปลี่ยน/ลดได้ตามความเหมาะสม)

หลังจากหน่วยงานมีการจัดทำเมทาดาดาแล้ว ส่วนกลางจะรวบรวมเมทาดาดาจากหน่วยงานต่างๆ รวมถึงเมทาดาดาของข้อมูลภาครัฐจากศูนย์ข้อมูล เช่น ศูนย์ข้อมูลน้ำ ศูนย์ข้อมูลประชาชน ฯลฯ มาเก็บไว้เพื่อประโยชน์ในการสืบค้นข้อมูลภาครัฐ โดย

- ข้อมูลที่ไม่มีชั้นความลับ สามารถเผยแพร่ได้ (Open data) การจัดเก็บจะรวบรวมทั้งเมทาดาดาและข้อมูล ทั้งนี้เพื่อความสะดวกแก่ผู้ให้บริการ กล่าวคือหากข้อมูลที่ต้องการใช้ประโยชน์เป็นข้อมูลเปิด ผู้ร้องขอข้อมูลจะสามารถดาวน์โหลดข้อมูลไปใช้ได้เลยผ่านช่องทางการให้บริการข้อมูล (Data portal) เดียวกัน
- ข้อมูลที่มีชั้นความลับ การจัดเก็บจะรวบรวมเฉพาะเมทาดาดาเท่านั้น ไม่จัดเก็บตัวข้อมูลจริง

2) จัดทำเกณฑ์สำหรับการระบุชั้นความลับของข้อมูล เพื่อให้หน่วยงานภาครัฐสามารถใช้เป็นเครื่องมือในการระบุจัดแบ่งประเภทข้อมูลตามชั้นความลับได้อย่างสอดคล้องกัน โดยเกณฑ์อาจจะอยู่ในรูปแบบของต้นไม้ประกอบการตัดสินใจดังแสดงในรูปที่ 2.2 จากรูปชั้นความลับข้อมูลจะจำแนกได้เป็น 4 ระดับตามกรอบการกำกับดูแลข้อมูลคือ ข้อมูลสาธารณะ (Public data) ข้อมูลที่สามารถระบุถึงตัวบุคคลได้ (Personally Identifiable Information: PII) ข้อมูลความลับราชการ (Confidential Government) และข้อมูลที่เกี่ยวข้องกับความมั่นคงของประเทศ (National Security)



รูปที่ 2.2 ต้นไม้ประกอบการตัดสินใจจำแนกชั้นความลับของข้อมูล

3) จัดทำดัชนีข้อมูล (data index) ซึ่งเป็นส่วนสำคัญที่จะเพิ่มความสะดวกรวดเร็วแก่บริการค้นหารายการข้อมูล โดยผู้ใช้บริการจะสามารถค้นหาแบบหลายมิติ (Multi-dimensional search) ได้ 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) ที่ทุกคนมีสิทธิในการเข้าถึงและนำไปใช้ได้อย่างเสรี กลุ่มข้อมูลนี้จะเป็นประโยชน์ต่อภาครัฐ ภาคเอกชน และประชาชน

- กลุ่มข้อมูลส่วนบุคคล (Personally Identifiable Information) ที่อาจต้องมีการนำรายละเอียดส่วนที่ระบุตัวตนได้ออกก่อนที่จะนำไปวิเคราะห์ และหน่วยงานเจ้าของข้อมูลสามารถระบุข้อจำกัดการใช้ประโยชน์ข้อมูลได้
- กลุ่มข้อมูลความลับราชการ (Government data) ที่ให้สิทธิเฉพาะเจ้าหน้าที่ที่มีหน้าที่รับผิดชอบเกี่ยวข้อง
- กลุ่มข้อมูลความมั่นคงของรัฐ (Secured Government data) ที่ให้สิทธิเฉพาะเจ้าหน้าที่ที่มีหน้าที่รับผิดชอบเกี่ยวข้อง

4) พิจารณาเลือกรูปแบบ (สกุลไฟล์) การจัดเก็บเมทาดาต้า โดยที่รูปแบบดังกล่าวจะต้องให้มนุษย์และเครื่องคอมพิวเตอร์สามารถอ่านเข้าใจได้อย่างสะดวก เพื่อให้การสืบค้นข้อมูลทำได้ง่ายและการแลกเปลี่ยนข้อมูลทำได้แบบอัตโนมัติ หรือกึ่งอัตโนมัติ โดยรูปแบบสากลของเมทาดาต้าจะมีอยู่หลายประเภทได้แก่ Common European Research Information Format (CERIF), Comprehensive Knowledge Archive Network (CKAN) ทั้งนี้รูปแบบที่ได้รับความนิยม คือ Data Catalog Vocabulary (DCAT) ซึ่งจะเป็นรูปแบบการเก็บเมทาดาต้าแบบ RDF จากนั้นทำการออกแบบกลไกและส่วนติดต่อผู้ใช้ (User Interface) ที่สะดวกสำหรับให้หน่วยงานใช้ในการจัดทำ (publish) เมทาดาต้าออนไลน์

5) ออกแบบเกณฑ์กลางและกระบวนการตัวอย่างในการประเมินคุณภาพข้อมูลหน่วยงาน ซึ่งตัวอย่างเกณฑ์ด้านคุณภาพข้อมูลตามมาตรฐานสากล ได้แก่ ความถูกต้องแม่นยำ (Accuracy), ความตรงกัน (Consistency), ความสามารถในการเข้าถึงข้อมูลที่เป็นปัจจุบัน (Availability), ความสมบูรณ์ของข้อมูล (Completeness), ความสอดคล้องกับรูปแบบมาตรฐานของข้อมูล (Conformance), ความน่าเชื่อถือได้ของข้อมูล (Credibility), ความเข้าประเด็น (Relevance), ความทันสมัยของข้อมูล (Timeliness) และบริบทข้อมูล (Context) ดังรูปที่ 2.3 อธิบายความหมายของเกณฑ์ตัวอย่าง

Accuracy	• Is the data correctly representing the real-world entity or event?
Consistency	• Is the data not containing contradictions?
Availability	• Can the data be accessed now and over time?
Completeness	• Does the data include all data items representing the entity or event?
Conformance	• Is the data following accepted standards?
Credibility	• Is the data based on trustworthy sources?
Processability	• Is the data machine-readable?
Relevance	• Does the data include an appropriate amount of data?
Timeliness	• Is the data representing the actual situation and is it published soon enough?
Context	• Why was the data collected in the first place?

รูปที่ 2.3 เกณฑ์การประเมินคุณภาพข้อมูล

6) จัดทำกรอบนโยบายต้นแบบด้านการกำกับดูแลวงจรชีวิตและบริหารจัดการข้อมูลดิจิทัลของหน่วยงานภาครัฐ เพื่อให้เอื้อต่อการใช้ประโยชน์ภายในหน่วยงานและแลกเปลี่ยนข้อมูลข้ามหน่วยงาน หน่วยงานภาครัฐสามารถนำไปดำเนินงานให้เป็นไปในทิศทางเดียวกันได้ โดยกรอบนโยบายจะต้องกำหนดกระบวนการเชิงปฏิบัติแยกกันสำหรับข้อมูลที่ขึ้นความลับต่างกัน : กลุ่มข้อมูลเปิด กลุ่มข้อมูลส่วนบุคคล กลุ่มข้อมูลความลับราชการ กลุ่มข้อมูลความมั่นคงของรัฐ

โดยกรอบดังกล่าวประกอบด้วยประเด็นดังนี้

- a. การกำหนดกระบวนการจัดทำและคุ้มครองข้อมูล (Procedure) ที่ทำให้มั่นใจว่าปริมาณและคุณภาพของข้อมูลตรงตามมาตรฐาน
- b. การกำหนดกระบวนการที่ทำให้มั่นใจว่าข้อมูลมีความทันสมัยตลอดเวลา
- c. กำหนดกระบวนการในการใช้ข้อมูลและการใช้ข้อมูลร่วมกัน¹ (Sharing of data)
 - i. ข้อตกลงและเงื่อนไข² (Terms and Conditions) ในการเผยแพร่ข้อมูล (Data Release)
 - ii. เกณฑ์วิธี (Protocol) สำหรับการถ่ายโอนข้อมูล และวิธีการจัดการกับข้อมูล (Data handling method) เช่น เอพียู (API: Application Programming Interface) เอฟทีพี (FTP: File Transfer Protocol) หน่วยความจำแบบพกพา อีเมล ฯลฯ
 - iii. นโยบายควบคุมการเข้าถึง (Access control policy) การกำหนดว่าบุคคลใดสามารถเข้าถึงข้อมูลใดได้บ้าง โดยอาจกำหนดตามหน้าที่ความรับผิดชอบของบุคคลนั้น
 - iv. การกำหนดกลุ่มของฟิลเตอร์ (Filter) ที่ทำให้ข้อมูลไม่สามารถระบุตัวตนได้ (anonymize) หรือ ลดระดับ (de-classify) ของชุดข้อมูล (Dataset) ที่เป็นความลับ (confidential) แต่เป็นชุดข้อมูลที่มีประโยชน์มาก โดยตัวอย่างการลดระดับชั้นความลับของข้อมูล ได้แก่ กรณีที่ชุดข้อมูลมีข้อมูลที่สามารถระบุถึงตัวบุคคลได้ (PII) การดำเนินการนำชื่อของบุคคลออกจากชุดข้อมูลนั้นจะทำให้ชุดข้อมูลดังกล่าวถูกลดระดับชั้นความลับจากระดับข้อมูลที่สามารถระบุถึงตัวบุคคลได้ (PII) เป็นระดับข้อมูลความลับ (Confidential)
- d. การกำหนดแบบฟอร์ม (Template) นโยบายการป้องกันข้อมูล (Data protection policy)

¹ <http://www.stou.ac.th/Schools/Sst/main/eLearning/Oa/html/charpter3.html>

² https://www.etda.or.th/file_storage/uploaded/Etda_Website/file/article_law_4.pdf

7) ออกแบบและจัดทำระบบบริการนามานุกรมข้อมูลภาครัฐ (Directory services) และระบบบริหารจัดการเมทาดาทา (Metadata Management System) โดยระบบจะมีกลไกในการค้นหาข้อมูลผ่านการสืบค้นในมิติต่างๆ กลไกการร้องขอข้อมูล (กรณีข้ามหน่วยงาน) และกลไกการอนุมัติใช้ข้อมูลสำหรับคำขอประเภทต่างๆ ทั้งนี้การขอข้อมูลที่ไม่มีชั้นความลับ (ข้อมูลเปิด) จะไม่ต้องขออนุมัติ นอกจากนี้ระบบนามานุกรมจะทำหน้าที่ยืนยันตัวตน (Authentication) และตรวจสอบสิทธิในการเข้าถึงข้อมูล (Authorization) ของผู้ใช้ตามกฎหมายที่หน่วยงานเจ้าของข้อมูลระบุ รวมทั้งบันทึกการสืบค้นและคำขอข้อมูลทั้งหมด (Search and Request Log) เพื่อให้หน่วยงานเจ้าของข้อมูลสามารถตรวจสอบได้ ระบบนามานุกรมนี้ จะเชื่อมต่อกับศูนย์เชื่อมโยงและแลกเปลี่ยนข้อมูล (Data Linkage and Exchange Centers) เพื่ออำนวยความสะดวกให้การเข้าถึง (Data Access) และถ่ายโอนข้อมูล (Data Transfer) สามารถทำได้ผ่านระบบทั้งหมดในอนาคต

8) จัดระบบเพื่อบันทึกธุรกรรมการล็อกอิน การร้องขอ การอนุมัติ ข้อมูลทั้งหมดที่เกิดขึ้นผ่านระบบนามานุกรมภาครัฐ เพื่อประโยชน์ในการติดตามสถานะและรายงานธุรกรรมการแลกเปลี่ยนข้อมูลทั้งหมด

2.2 การทำงานร่วมกับหน่วยงานภาครัฐเรื่องการกำกับดูแลข้อมูล

ความสำเร็จของการจัดทำรายการข้อมูลภาครัฐอย่างเป็นรูปธรรมนั้น จะต้องเกิดจากความเข้าใจและเห็นถึงประโยชน์ร่วมกันของหน่วยงานภาครัฐ นอกจากนี้การแลกเปลี่ยนข้อมูลข้ามหน่วยงานจะเกิดประโยชน์ได้นั้น หน่วยงานภาครัฐก็ต้องสามารถดูแลข้อมูลของตนได้อย่างเป็นระบบ กระทั่งดิจิทัลเพื่อเศรษฐกิจและสังคมจึงสนับสนุนให้หน่วยงานภาครัฐจัดตั้งคณะทำงานด้านข้อมูล (Data Stewardship team) เพื่อดำเนินงานด้านนโยบายข้อมูลของหน่วยงานเอง โดยกระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคมสามารถให้การสนับสนุนทางด้านวิชาการเพื่อให้ได้นโยบายข้อมูลที่เหมาะสมกับหน่วยงานและสอดคล้องกันในภาครัฐได้ ตัวอย่างเช่น การวางแผนนโยบายเพื่อกำกับดูแลให้ข้อมูลแต่ละประเภทมีวงจรชีวิต (Data life Cycle) ที่เหมาะสม บริหารจัดการได้อย่างเป็นรูปธรรม และตอบโจทย์ด้านการรายงานและการวิเคราะห์เพื่อประกอบการตัดสินใจ

ในเบื้องต้น กระบวนการสร้างความเข้าใจและจัดเตรียมมาตรฐานด้านข้อมูลของกระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม จำเป็นต้องได้รับความร่วมมือจากหน่วยงานภาครัฐอย่างเข้มแข็งในเรื่อง ดังต่อไปนี้

1. จัดตั้งคณะทำงานด้านข้อมูล (Data Stewardship team) เพื่อประสานกับทีมงานจากกระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคมที่รับผิดชอบเรื่องการจัดทำรายการข้อมูลภาครัฐ
2. ระบุชุดข้อมูลสำคัญ (Critical dataset) ของหน่วยงาน โดยมีเกณฑ์ที่ชัดเจนในการกำหนดระดับความสำคัญของชุดข้อมูล
3. จำแนกชุดข้อมูลสำคัญตามชั้นความลับและจัดทำดัชนีข้อมูลในมิติต่างๆ เพื่อความสะดวกในการสืบค้น

4. สร้างกฎเกณฑ์ว่าด้วยสิทธิการเข้าถึงข้อมูล (Access Right) โดยระบุประเภทบุคคล ที่จะมีสิทธิเข้าถึงข้อมูลในแต่ละชั้นความลับ และสิทธิในการขอรับข้อมูลไปใช้ประโยชน์ผ่านกลไกการถ่ายโอนข้อมูล
5. จัดทำเมทาเดตาร่วมกับกระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม เพื่อให้หน่วยงานภาครัฐมีเมทาเดตาในรูปแบบเดียวกันสามารถรองรับการสืบค้น เชื่อมโยง และแลกเปลี่ยนข้อมูลภาครัฐได้ในอนาคต
6. พิจารณาสารสร้างพื้นที่สำหรับแลกเปลี่ยนข้อมูลแยกต่างหากจากฐานข้อมูลหลัก และแปลงรูปแบบการจัดเก็บข้อมูลสำคัญในพื้นที่ดังกล่าวให้อยู่ในสกุลไฟล์ที่ให้ทั้งมนุษย์และเครื่องคอมพิวเตอร์สามารถอ่านได้อย่างสะดวกเพื่อให้การแลกเปลี่ยนข้อมูลเกิดขึ้นได้แบบอัตโนมัติ
7. ออกแบบกระบวนการประเมินคุณภาพและปริมาณของชุดข้อมูลสำคัญที่เหมาะสมกับหน่วยงาน โดยระบุเกณฑ์ที่ใช้ในการประเมินคุณภาพอย่างชัดเจนและเป็นไปตามมาตรฐานสากล (สามารถอ้างอิงจากเกณฑ์กลางและกระบวนการตัวอย่างของกระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคมได้) จากนั้นทำการประเมินช่องว่างที่เกิดขึ้น (Gap) ระหว่างสถานะข้อมูลปัจจุบันกับสถานะที่ควรจะเป็นตามเกณฑ์ที่กำหนด เพื่อวางแผนดำเนินการ (Roadmap) การพัฒนาชุดข้อมูลสำคัญให้มีคุณภาพภายในระยะเวลา 3-5 ปี
8. วางนโยบายเพื่อกำกับดูแลวงจรชีวิตและบริหารจัดการข้อมูลดิจิทัล และสามารถใช้ได้กับทุกประเภทข้อมูล (4 ประเภท) โดยเฉพาะประเด็นเรื่องกระบวนการจัดทำและคุ้มครองข้อมูล และกระบวนการใช้และแลกเปลี่ยนข้อมูล

แผนการดำเนินงานเรื่องการให้บริการรายการข้อมูลภาครัฐของกระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม จะแบ่งเป็น 3 ช่วง โดยเริ่มต้นช่วงที่ 1 ในปีงบประมาณ 2562 ช่วงที่ 2 ในปีงบประมาณ 2563 และช่วงที่ 3 ในปีงบประมาณ 2564 และ 2565

ในช่วงที่ 1 กิจกรรมรวมถึงการศึกษาและจัดทำชุดเครื่องมือที่ประกอบด้วย กรอบมาตรฐาน กรอบนโยบาย กลไกการจัดการ และแบบฟอร์มด้านข้อมูลต่างๆ สำหรับให้หน่วยงานภาครัฐใช้เป็นจุดเริ่มต้นในการจัดทำมาตรฐานเมทาเดตาและกำกับดูแลข้อมูล เช่น มาตรฐานและตัวอย่างเมทาเดตา วิธีการประเมินคุณภาพและปริมาณข้อมูล เกณฑ์สำหรับการระบุชั้นความลับของข้อมูล กรอบนโยบายด้านการบริหารจัดการข้อมูลดิจิทัล ฯลฯ โดยศึกษาจากรูปแบบที่ต่างประเทศใช้ และรูปแบบของเมทาเดตาที่มีการดำเนินการแล้วของหน่วยงานภาครัฐ

ในขั้นตอนนี้คณะทำงานอาจจะร้องขอเมทาเดตาจากหน่วยงานภาครัฐเท่าที่มีเพื่อมาถ่วงน้ำหนักและจัดทำรายการข้อมูลเบื้องต้นได้อีกด้วย หลังจากนั้น ทีมงานจากกระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม จะนำชุดเครื่องมือที่จัดทำขึ้นไปใช้เพื่อร่วมจัดทำเมทาเดตาข้อมูลและวางแผนการจัดการนโยบายด้านข้อมูลกับหน่วยงานนำร่อง 1-2 หน่วยงาน รวมทั้งจัดฝึกอบรมบุคลากรของกระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม เพื่อเตรียมความพร้อมในการให้คำแนะนำให้กับหน่วยงานภาครัฐอื่นในการดำเนินงานระยะถัดไป

ในช่วงที่ 2 กระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคมจะนำชุดเครื่องมือที่ผ่านกระบวนการทดสอบกับหน่วยงานนำร่องแล้วไปใช้ร่วมกับงานใน 1-2 รายสาขาของงานภาครัฐ (จากทั้งสิ้น 21 สาขา) รวมทั้งพัฒนาระบบบริการนิตินามุกรมข้อมูลภาครัฐ และระบบบริหารจัดการเมทาดาต้า เพื่อเริ่มต้นให้บริการ

ในช่วงที่ 3 กระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคมจะทำการขยายผลให้ครอบคลุมทั้ง 21 สาขาของงานภาครัฐ ให้ระบบบริการนิตินามุกรมข้อมูลภาครัฐ สามารถให้บริการได้อย่างสมบูรณ์

บทที่ 3 การให้บริการด้านการแลกเปลี่ยนข้อมูล

การใช้ประโยชน์ข้อมูลข้ามหน่วยงานต้องอาศัยระบบบริการนามานุกรมข้อมูลภาครัฐเพื่อการค้นหาและระบุข้อมูลที่ต้องการสำหรับแต่ละโจทย์ปัญหา และต้องมีกลไกสารสนเทศเพื่อเอื้อให้เกิดการถ่ายโอนข้อมูลข้ามหน่วยงานได้อย่างสะดวก ดังนั้นจึงต้องมีการพัฒนาระบบสารสนเทศเพื่อบริหารจัดการการเชื่อมโยงและแลกเปลี่ยนข้อมูลสำหรับการวิเคราะห์ข้อมูลขนาดใหญ่ (Big Data Analytics) ของหน่วยงานภาครัฐ ซึ่งตัวอย่างบริการพื้นฐานของระบบดังกล่าวรวมถึง การยืนยันตัวตน การจำกัดสิทธิในการเข้าถึงข้อมูลบริการส่งข้อมูลข้ามหน่วยงานผ่านระบบสารสนเทศ การบันทึกทราบนแซกชันการเข้าถึงข้อมูลเพื่อเป็นหลักฐาน และการบริการข้อมูลประเภทอื่น เนื้อหาในบทนี้จะนำเสนอรายละเอียดสำหรับการให้บริการด้านการแลกเปลี่ยนข้อมูล

3.1 มาตรฐานข้อมูล

เพื่อให้การแลกเปลี่ยนข้อมูลเกิดขึ้นได้อย่างมีประสิทธิภาพ หน่วยงานภาครัฐควรสร้างพื้นที่แลกเปลี่ยนข้อมูล (Data Exchange Zone) แยกต่างหากจากฐานข้อมูลหลัก และพิจารณาแปลงรูปแบบข้อมูลดิบที่จะนำมาใส่ในพื้นที่แลกเปลี่ยนนี้ให้เอื้อต่อการแลกเปลี่ยนข้อมูล โดยกระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคมได้กำหนดมาตรฐานรูปแบบข้อมูลเป็น 4 ระดับ ดังแสดงในตารางที่ 3.1

- **ระดับ 4 ดาว :** ตัวอย่างเช่น เจซัน (JSON: JavaScript Object Notation) เอ็กเอ็มแอล (XML: Extensible Markup Language) และ อาร์ดีเอฟ (RDF: Resource Description Framework) ฯลฯ ซึ่งเป็นไฟล์ที่ข้อมูลถูกเก็บในลักษณะข้อความ (plain text) ที่ทั้งมนุษย์และระบบคอมพิวเตอร์สามารถอ่านเข้าใจได้ รวมทั้งระบบคอมพิวเตอร์สามารถประมวลผลข้อมูลได้โดยอัตโนมัติ ซึ่งเป็นรูปแบบที่ดีที่สุดสำหรับการแลกเปลี่ยนข้อมูล
- **ระดับ 3 ดาว :** ซีเอสวี (CSV: Comma-Separated Value) เอ็กแอลเอส (XLS) หรือ เอ็กแอลเอสเอ็ก (XLSX) ซึ่งเป็นไฟล์ที่ข้อมูลถูกจัดเป็นคอลัม (column) ระบบคอมพิวเตอร์และมนุษย์สามารถอ่านเข้าใจได้ถ้ามีคำอธิบายคอลัม (column) จากผู้ให้บริการข้อมูล ทั้งนี้สกุลไฟล์ CSV สามารถถูกแปลงเป็น JSON, XML และ RDF ได้อัตโนมัติโดยใช้โปรแกรมสำเร็จรูป (ตัวอย่างวิธีการแสดงในเอกสารแนบ ข) ขณะที่ไฟล์เอ็กเซลจะมีข้อดีน้อยกว่าเมื่อเทียบกับ CSV เนื่องจากต้องใช้ออฟต์แวร์ลิขสิทธิ์จึงจะสามารถเปิดและแปลงไฟล์ได้ อย่างไรก็ตามการแลกเปลี่ยนข้อมูลยังสามารถทำได้อย่างเป็นระบบแม้มีข้อจำกัดอยู่บ้าง
- **ระดับ 2 ดาว :** ด็อก (DOC) หรือ ด็อกเอ็ก (DOCX) เฮชทีเอ็มแอล (HTML) และไฟล์เท็กซ์ (TXT) ซึ่งเป็นไฟล์ข้อมูลที่ใช้ประโยชน์โดยระบบคอมพิวเตอร์ต้องมีการแปลงไฟล์หรือกระบวนการประมวลผลภาษาธรรมชาติ (Natural language Processing) เพิ่มเติม ซึ่งไม่สะดวกสำหรับการแลกเปลี่ยนข้อมูลมากนัก
- **ระดับ 1 ดาว :** รูปแบบข้อมูลที่ไม่เหมาะสมกับการแลกเปลี่ยนข้อมูลเพื่อการใช้ประโยชน์คือ ไฟล์ PDF ภาพกราฟ และ ข้อมูลที่ยังไม่อยู่ในรูปแบบดิจิทัล

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

ตารางที่ 3.1 สกุลไฟล์สำหรับการแลกเปลี่ยนข้อมูล

นอกจากนี้ ข้อมูลดิบที่หน่วยงานภาครัฐจัดเก็บอาจจะไม่สอดคล้อง ตัวอย่างเช่นข้อมูล ‘วันที่’ หน่วยงานที่ 1 เก็บในรูปแบบ dd/MM/YYYY หน่วยงานที่ 2 เก็บในรูปแบบ MM/dd/YYYY และปีที่เก็บจะมีโอกาสเป็นได้ทั้งปีคริสต์ศักราช (ค.ศ.) และปีพุทธศักราช (พ.ศ.) การใช้ประโยชน์ข้อมูลจึงต้องมีการแปลงรูปแบบเข้าหากันซึ่งเพิ่มค่าใช้จ่ายและไม่สะดวกต่อการทำงาน ดังนั้นกระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคมจะทำหน้าที่สร้างและรวบรวมมาตรฐานกลางสำหรับข้อมูลประเภทต่างๆ และจัดทำคลังมาตรฐานข้อมูลที่สืบค้นอ้างอิงได้ โดยมีการตรวจสอบผู้เป็นเจ้าของข้อมูลหลัก (Master Data) ตัวอย่างเช่น ชื่อคนและที่อยู่จะมีกระทรวงมหาดไทยเป็นเจ้าของข้อมูลหลัก รูปแบบการจัดเก็บจึงควรยึดตามกระทรวงมหาดไทย การนำข้อมูลไปใช้ต่อจึงจะสามารถอ้างอิงได้อย่างถูกต้องโดยไม่ต้องแปลงรูปแบบ กรณีข้อมูลทั่วไปที่ไม่มีเจ้าของตามพระราชบัญญัติ เช่น ‘วันที่’ กระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคมจะสร้างมาตรฐานรูปแบบข้อมูลเพื่อการจัดเก็บขึ้นตามมาตรฐานสากล

3.2 การปรับปรุงข้อมูลเพื่อเตรียมความพร้อม

สำหรับหน่วยงานเจ้าของข้อมูลที่จะเตรียมจะเข้าร่วมในกลไกการแลกเปลี่ยนข้อมูลภาครัฐ หน่วยงานควรที่จะปรับปรุงข้อมูลสำคัญเพื่อเตรียมพร้อมสำหรับการให้บริการข้อมูล โดยเริ่มจาก 2 เกณฑ์ประเมินคุณภาพที่ทำได้ง่ายก่อน คือ ความสอดคล้องกับรูปแบบมาตรฐานของข้อมูล (Conformance) และความถูกต้องแม่นยำ (Accuracy) โดยเริ่มจากการตรวจสอบระดับพื้นฐานก่อน ดังนี้

ความสอดคล้องกับรูปแบบมาตรฐานของข้อมูลพื้นฐาน

- ตรวจสอบให้ชุดข้อมูลมีแถว Header สำหรับอธิบายรายละเอียดของคอลัม (column) เพื่อนำไปใช้ประโยชน์ของหน่วยงานที่ขอข้อมูล

- ตรวจสอบให้ชุดข้อมูลมีรหัสที่แสดงเวอร์ชัน เพื่อให้ผู้ขอข้อมูลสามารถติดตามการเปลี่ยนแปลงของชุดข้อมูลได้ โดยเวอร์ชันสามารถใช้ตรวจสอบข้อมูลพื้นฐาน เช่น วันที่เก็บข้อมูล วันที่ปรับปรุงข้อมูล วิธีจัดเก็บข้อมูล และหน่วยงานเจ้าของข้อมูล ทั้งนี้เพื่อให้ผู้ใช้ประโยชน์ข้อมูลสามารถตรวจสอบความใหม่ของข้อมูลได้

ความถูกต้องแม่นยำ

- ตรวจสอบหาค่าที่ผิดปกติจากกลุ่มทั่วไปในชุดข้อมูล หากพบข้อมูลที่มีช่วงของค่าที่เป็นไปไม่ได้ เช่น ข้อมูลอายุมีค่าหลักพัน ควรปรับแก้ เป็นต้น
- ตรวจสอบหาข้อมูลที่ซ้ำกัน ข้อมูลที่ขาดหายไปบางส่วน (Missing Data) และความผิดปกติทั่วไปจากกฎที่ตั้งไว้ เช่น การเก็บตัวอักษรในคอลัมตัวเลข หากมีควรปรับค่ารวมทั้งระบุกระบวนการการจัดการแก้ไข
- ตรวจสอบว่าชุดข้อมูลมีคอลัมข้อมูลที่ถูกสร้างขึ้นมาเพียงเพื่อการทดสอบระบบงานหรือไม่ หากมีควรลบทิ้ง

โดยการตรวจสอบคุณภาพแบบพื้นฐานนี้ สามารถทำแบบออฟไลน์ได้ เพื่อไม่ขัดจังหวะการใช้งานระบบฐานข้อมูล รวมทั้งเจ้าหน้าที่สารสนเทศหน่วยงานสามารถพัฒนาโปรแกรมย่อยเพื่อใช้ตรวจสอบข้อมูลแบบกึ่งอัตโนมัติอีกด้วย

3.3 ระบบบริการและแลกเปลี่ยนข้อมูลกลาง

เพื่อการกำกับดูแลชั้นความลับข้อมูล กลไกและระบบสารสนเทศเพื่อการให้บริการและแลกเปลี่ยนข้อมูลกลาง จำเป็นจะต้องเชื่อมต่อกับระบบนามาณุกรมข้อมูลภาครัฐ เพื่อทำการยืนยันตัวตนผู้ใช้ (Authentication) และตรวจสอบสิทธิในการเข้าถึงข้อมูล (Authorization) ให้เป็นไปตามกฎเกณฑ์ที่หน่วยงานเจ้าของข้อมูลระบุ ระบบบริการและแลกเปลี่ยนข้อมูลจะต้องบันทึกรายการการเข้าถึงข้อมูลและถ่ายโอนข้อมูลทั้งหมด (Access Log) เพื่อให้หน่วยงานเจ้าของข้อมูลสามารถตรวจสอบได้ และเพื่อให้สามารถทำผลสรุปเชิงสถิติเรื่องการร้องขอใช้ข้อมูลของทั้งภายในและภายนอกหน่วยงาน อีกด้วย

นอกจากนี้กลไกการบริการและแลกเปลี่ยนข้อมูลกลางยังต้องมีกระบวนการบนระบบสารสนเทศที่ให้ผู้ใช้งานสามารถจัดการข้อมูลที่มีความผิดปกติได้ตามขั้นตอนดังต่อไปนี้

1. Confirm: ตรวจสอบและทำการยืนยันว่าชุดข้อมูลมีปัญหา
2. Notify: แจ้งเตือนให้กับผู้ดูแลหรือหน่วยงานเจ้าของข้อมูลนี้เกี่ยวกับปัญหาในชุดข้อมูล
3. Log: ทำการเก็บ log ปัญหาที่เกิดขึ้นไว้เพื่อติดตามและสรุปผล
4. Diagnose: วิเคราะห์ปัญหาของข้อมูลพื้นฐานได้
5. Evaluate Option: ประเมินต้นตอของปัญหา เช่น การเก็บข้อมูล อัลกอริทึมในการคำนวณข้อมูล
6. Correct: แก้ไขข้อมูลที่ผิดปกติที่พบให้ถูกต้อง

ในช่วงแรกของการให้บริการแลกเปลี่ยนข้อมูล กระบวนการจัดการข้อมูลจะช่วยให้เกิดการแจ้งเตือนระหว่างหน่วยงานเพื่อให้เกิดการพัฒนาคุณภาพข้อมูลโดยรวม และทำให้หน่วยงานสามารถให้บริการข้อมูลที่มีระดับคุณภาพที่ยอมรับ (Data Quality Service Level Agreement) ได้ในอนาคต

กลไกการบริการและการแลกเปลี่ยนข้อมูลเพื่อการวิเคราะห์ข้อมูลขนาดใหญ่ นั้น ควรถูกออกแบบในลักษณะที่หน่วยงานเจ้าของข้อมูลไม่จำเป็นต้องส่งข้อมูลดิบเข้ามาจัดเก็บที่ศูนย์บริการข้อมูลกลาง เพียงแต่จัดทำ (publish) เมทาเดตาขึ้นระบบนามานุกรมข้อมูลภาครัฐเท่านั้น ขั้นตอนการแลกเปลี่ยนข้อมูลระหว่างหน่วยงาน อธิบายได้ตามรูปที่ 3.1



รูปที่ 3.1 กลไกการบริการและการแลกเปลี่ยนข้อมูล

จากรูปการบริการและแลกเปลี่ยนข้อมูลจะมี 4 ขั้นตอน ดังต่อไปนี้

1. หน่วยงานภาครัฐจัดทำ (publish) เมทาเดตาเข้ามาที่ระบบนามานุกรมข้อมูลภาครัฐ
2. ผู้ใช้งานที่ต้องการร้องขอข้อมูลจะทำการสืบค้น ระบุข้อมูลที่ต้องการ และยื่นคำร้องขอข้อมูลผ่านระบบนามานุกรมข้อมูลภาครัฐ (Directory Service) ซึ่งระบบดังกล่าวจะยืนยันตัวตน ตรวจสอบสิทธิ์ของผู้ร้องขอ และตรวจสอบว่าคำขอข้อมูลมีความเหมาะสมตามกฎเกณฑ์ที่ระบุโดยหน่วยงานเจ้าของข้อมูลหรือไม่ ตัวอย่างเช่น ข้อมูลที่ขอมีชั้นความลับระดับใด ผู้ขอมีสิทธิ์หรือไม่ คำขอต้องการปริมาณข้อมูลมากเพียงใด หากคำขออยู่ในเกณฑ์ปกติการอนุมัติจะสามารถทำได้โดยระบบอัตโนมัติ มิเช่นนั้นจะต้องมีการส่งผ่านคำขอไปยัง คณะอนุกรรมการกลั่นกรองคำขอข้อมูลอีกต่อหนึ่ง นอกจากนี้คำร้องขอข้อมูลทั้งหมดจะถูกบันทึกเพื่อประโยชน์ในการตรวจสอบในภายหลัง
3. หากคำร้องขอข้อมูลได้รับการอนุมัติ จะมีการแจ้งไปยังผู้ขอ และคำร้องพร้อมรายละเอียดจะถูกส่งผ่านไปยังระบบบริการและแลกเปลี่ยนข้อมูลกลาง (Data Exchange Service)
4. ระบบบริการและแลกเปลี่ยนข้อมูลกลาง จะมีการต่อเชื่อมกับหน่วยงานเจ้าของข้อมูลและหน่วยงานผู้ขอรับบริการทั้งหมด รวมทั้งเก็บข้อมูลด้านเครือข่าย เช่น IP Address และเกณฑ์ในการส่ง-รับข้อมูลของ

หน่วยงานเจ้าของข้อมูล เพื่อจุดประสงค์ในการถ่ายโอนข้อมูล ซึ่งการถ่ายโอนข้อมูลจากต้นทางไปยังปลายทางสามารถทำได้ทั้งแบบผ่านและไม่ผ่านระบบกลาง การที่ข้อมูลไม่จำเป็นต้องวิ่งผ่านระบบบริการข้อมูลกลางจะสามารถลดปัญหาข้อขัดข้องที่อาจเกิดขึ้นในกรณีที่มีรายการคำขอจำนวนมาก ทั้งนี้ช่องทางการส่งข้อมูลจะขึ้นอยู่กับเกณฑ์ที่หน่วยงานเจ้าของข้อมูลระบุไว้ ตัวอย่างเช่น

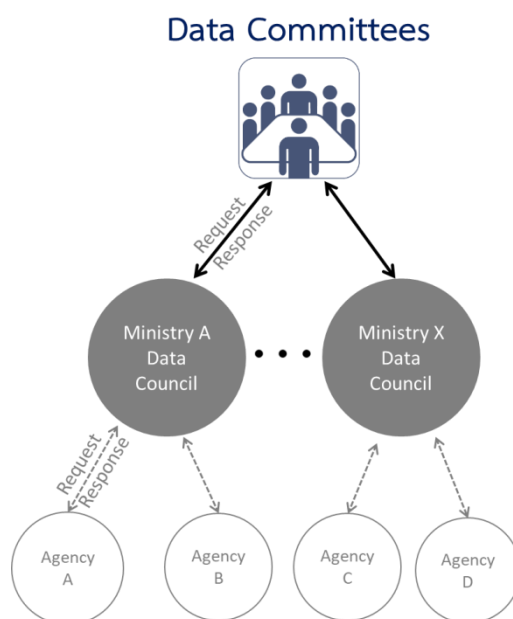
- Application Programming Interface หรือ API คือการโอนย้ายข้อมูลโดยอัตโนมัติ ผ่านการเรียกใช้งานส่วนโปรแกรมที่หน่วยงานเจ้าของข้อมูลจัดเตรียมไว้ให้ โดยระบบงานของผู้ขอข้อมูลจะสามารถดึงข้อมูลตรงจากระบบฐานข้อมูลของหน่วยงานเจ้าของข้อมูลได้โดยตรง การส่งลักษณะนี้เหมาะกับข้อมูลที่มีโครงสร้าง
- Secured File Transfer Protocol (SFTP) ทำได้ 2 วิธี คือ 1) เจ้าของข้อมูลทำการอัปโหลดไฟล์ข้อมูลเข้าไปที่พื้นที่จัดเก็บชั่วคราวของระบบบริการข้อมูลกลาง เพื่อให้ผู้ร้องขอข้อมูลเข้ามาดึงข้อมูลไปใช้ในภายหลัง 2) หน่วยงานเจ้าของข้อมูลถ่ายโอนข้อมูลไปยังผู้รับโดยตรงผ่านระบบอัตโนมัติด้วยโปรโตคอล SFTP การส่งข้อมูลลักษณะนี้ต้องมีรหัสผู้ใช้และรหัสผ่านสำหรับการเข้าสู่ระบบที่ชัดเจน การส่งลักษณะนี้เหมาะกับข้อมูลขนาดใหญ่และข้อมูลที่ไม่มีโครงสร้าง
- Removable storage หรือ Email คือการทำสำเนาไฟล์ข้อมูลโดยใส่รหัสป้องกันการเปิดไฟล์ และนำไฟล์ดังกล่าวใส่ลงใน Removable Storage เพื่อจัดส่ง หรือ ส่งทาง Email การส่งลักษณะนี้เหมาะกับหน่วยงานที่มีข้อจำกัดทางด้านระบบเครือข่ายคอมพิวเตอร์

หมายเหตุ ระบบนามานุกรมข้อมูลภาครัฐ และระบบบริการข้อมูลกลาง ไม่จำเป็นต้องติดตั้งในดาต้าเซนเตอร์เดียวกัน ระบบทั้ง 2 สามารถทำงานร่วมกันได้ผ่านกลไก Web Service

บทที่ 4 การกลั่นกรองคำร้องขอข้อมูลข้ามหน่วยงาน

เมื่อระบบบริการและแลกเปลี่ยนข้อมูลเริ่มถูกใช้งานอย่างเป็นรูปธรรม คำร้องขอข้อมูลจะมีหลากหลายประเภทและต่างวัตถุประสงค์ ภาครัฐจึงจำเป็นต้องวางแผนอย่างเป็นระบบเพื่อรองรับคำร้องขอพิเศษรูปแบบต่างๆ เช่น คำขอข้อมูลจำนวนมากกว่าปกติ คำขอข้อมูลจากหลายหน่วยงานในคราวเดียว คำขอข้อมูลที่มีชั้นความลับสูงจากกระทรวงอื่น เป็นต้น ซึ่งการสร้างมาตรฐานและกระบวนการในการอนุมัติคำร้องขอและการให้บริการข้อมูลที่เป็นระบบ ทำซ้ำได้ และสะดวกสำหรับผู้ใช้งานจึงเป็นสิ่งจำเป็น

คำขอข้อมูลจากหน่วยงานภายใต้กระทรวงเดียวกันสามารถพิจารณาตัดสินได้ในระดับกระทรวงโดยคณะทำงานด้านข้อมูลของกระทรวง อย่างไรก็ตามในกรณีการขอข้อมูลข้ามกระทรวงนั้น การพิจารณาอนุมัติจะเป็นเรื่องยาก รัฐบาลจึงควรแต่งตั้งหัวหน้าเจ้าหน้าที่ข้อมูลภาครัฐ (Government Chief Data Officer) และจัดตั้งคณะกรรมการเพื่อการกลั่นกรองคำร้องขอข้อมูล เพื่อร่วมพิจารณาคำร้องขอข้อมูลแบบพิเศษอย่างเป็นระบบดังแสดงในรูปที่ 4.1



รูปที่ 4.1 พิจารณาคำร้องขอข้อมูล

คณะกรรมการดังกล่าวควรมีตัวแทนจากคณะทำงานด้านข้อมูล (Data Stewardship team) ของทุกกระทรวง และมีคณะทำงานจากสำนักงานพัฒนารัฐบาลดิจิทัล และ/หรือ กระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคมเพื่อช่วยประสานให้เกิดการปฏิบัติตามมติจากคณะกรรมการฯ โดยคณะทำงานดังกล่าวจะทำหน้าที่ในการออกแบบและกำกับการทำงานบนระบบสารสนเทศเพื่ออำนวยความสะดวกในกระบวนการพิจารณาและอนุมัติคำร้องขอข้อมูล อีกด้วย

การมีบุคลากรในระดับต่างๆ ร่วมกำกับดูแลการให้บริการข้อมูลของหน่วยงานภาครัฐ จะส่งผลให้การแลกเปลี่ยนและใช้ประโยชน์ข้อมูลเกิดขึ้นได้อย่างมีประสิทธิภาพ หัวหน้าเจ้าหน้าที่ข้อมูลภาครัฐและคณะกรรมการฯ ดังกล่าวจะมีหน้าที่หลัก ดังนี้

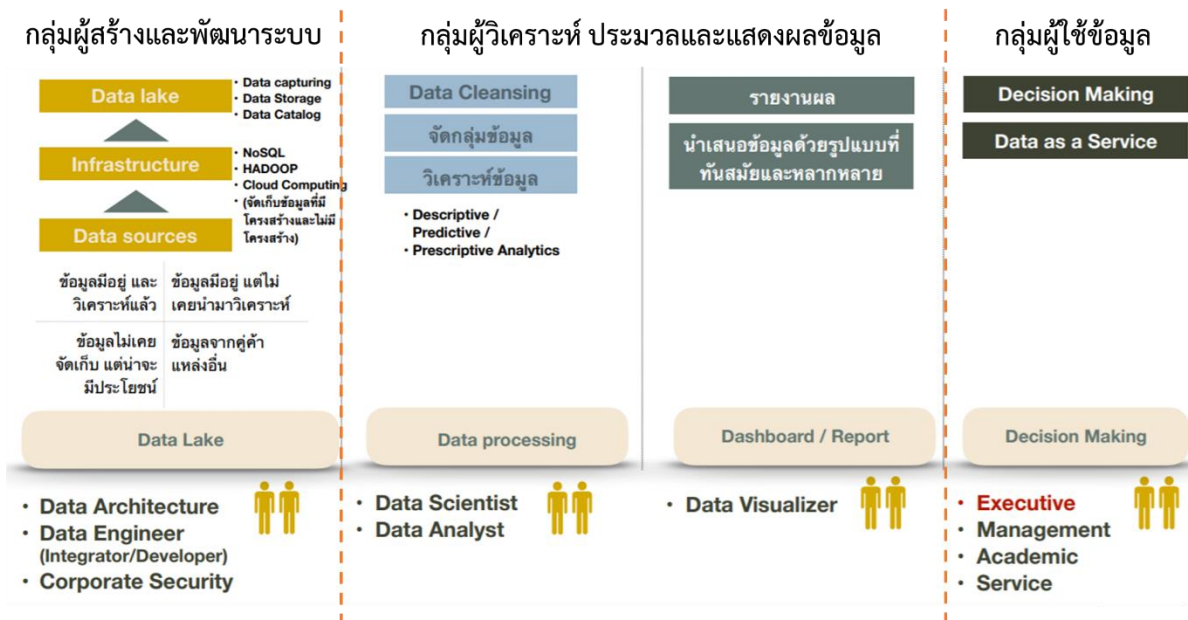
- พิจารณาประเด็นที่เกี่ยวข้องกับการสร้างศูนย์ข้อมูล
- พิจารณาคำร้องขอข้อมูลพิเศษ และให้นโยบายแก่หน่วยงานเจ้าของข้อมูลเรื่อง การอนุมัติ/ไม่อนุมัติ คำร้องขอ
- บริหารจัดการให้การใช้ข้อมูลภาครัฐเกิดประโยชน์สูงสุดและสนับสนุนให้การใช้ข้อมูลเพื่อการวิเคราะห์ เกิดขึ้นในวงกว้าง โดยไม่ขัดต่อพระราชบัญญัติด้านข้อมูล กฎกระทรวง และเกณฑ์ด้านข้อมูลอื่นๆ
- บริหารจัดการการแลกเปลี่ยนข้อมูลข้ามกระทรวง โดยออกนโยบาย วิธีการ และข้อตกลงการใช้ ประโยชน์ข้อมูลข้ามหน่วยงานที่เหมาะสม โดยมีตัวอย่างเอกสาร เช่น ข้อตกลงการแลกเปลี่ยนข้อมูล (Data Exchange Agreement) ข้อตกลงการให้บริการข้อมูล (Data Supply Agreement) และ กฎเกณฑ์/เงื่อนไข ในการปล่อยข้อมูลจากหน่วยงาน (Terms and Conditions of Data Release) ให้ หน่วยงานอื่นของภาครัฐนำไปใช้ประโยชน์
- ส่งเสริม สนับสนุน และประสานให้เกิดการปฏิบัติการอย่างเป็นรูปธรรมในเรื่องการบริการข้อมูล ภาครัฐในมิติต่างๆ คือ การทำรายการข้อมูลของหน่วยงาน (Data Inventory) กระบวนการ กำกับดูแลข้อมูล (Data Governance) การผลิตข้อมูล (Data Production) การสนับสนุนให้เกิด การใช้ประโยชน์ข้อมูล (Data Usage)
- นำเสนอกลยุทธ์ด้านข้อมูลให้แก่คณะรัฐมนตรี

บทที่ 5 การพัฒนาบุคลากรภาครัฐด้านการวิเคราะห์ข้อมูล

ในการปรับตัวและเตรียมความพร้อมเข้าสู่ยุคของการวิเคราะห์และใช้ประโยชน์ข้อมูลขนาดใหญ่ เพื่อวางแผนยุทธศาสตร์ ปฏิบัติงาน และบริหารงานของภาครัฐนั้น หน่วยงานภาครัฐจำเป็นต้องวางแผนและดำเนินงานด้านการพัฒนาบุคลากรล่วงหน้า โดยเน้นการอบรมเพื่อสร้างนักวิทยาศาสตร์ข้อมูล วิศวกรข้อมูล และนักวิเคราะห์ข้อมูล เนื่องจากบุคลากรในสาขาดังกล่าวเป็นสิ่งที่ต้องการสูงมาก การจัดจ้างอัตราใหม่ในจำนวนที่เพียงพอจึงเป็นเรื่องที่ท้าทาย สำนักงานคณะกรรมการข้าราชการพลเรือนร่วมกับกระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคมจึงได้เสนอกลไกการพัฒนาบุคลากรเพื่อการใช้ประโยชน์ข้อมูลขนาดใหญ่ดังนี้

5.1 กรอบการพัฒนาบุคลากร

ข้อเสนอการพัฒนาบุคลากรเพื่อการใช้ประโยชน์ข้อมูลขนาดใหญ่ ได้กำหนดเป้าหมายว่า ภายในปี 2565 บุคลากรภาครัฐของไทยจะตระหนักถึงคุณค่าของข้อมูลและสามารถนำข้อมูลขนาดใหญ่มาสร้างมูลค่าเพิ่ม และมีการพัฒนาเครือข่ายการทำงานอย่างเป็นพันธมิตรกับทุกภาคส่วน (Public-Private Partnership: PPP) รวมทั้งมีการกำหนดนโยบาย ยุทธศาสตร์ และการบริหารจัดการที่ตั้งอยู่บนข้อมูลและหลักฐานเชิงประจักษ์ กลุ่มเป้าหมายของการพัฒนาประกอบด้วย 3 กลุ่ม ดังแสดงในรูปที่ 5.1



รูปที่ 5.1 กลุ่มเป้าหมายการพัฒนาบุคลากรด้านการใช้ประโยชน์ข้อมูล

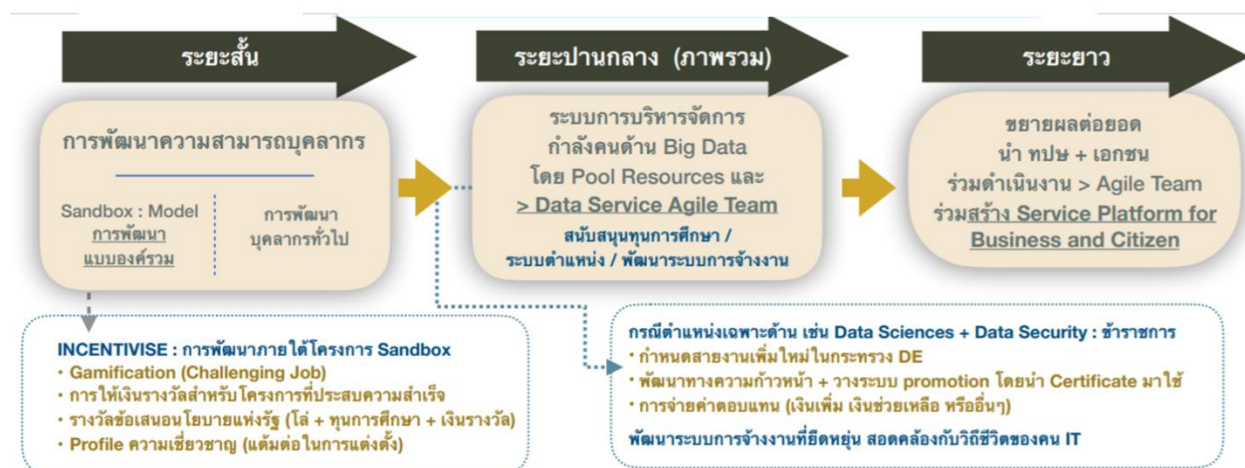
รูปที่ 5.1 แสดงกลุ่มงานและทักษะหลักที่ต้องการเพื่อทำโครงการบิ๊กดาต้าภาครัฐให้ได้ประสิทธิผล โดยแบ่งเป็น 3 เรื่องคือ งานด้านการสร้างและดูแลระบบสารสนเทศบิ๊กดาต้า งานด้านการวิเคราะห์และแสดงผลข้อมูล และงานด้านการนำเอาข้อมูลไปใช้ประโยชน์ประกอบการตัดสินใจ ดังนั้นกลุ่มเป้าหมายของการพัฒนาบุคลากรจึงมี 3 กลุ่มเพื่อให้สอดคล้องกับงาน ดังนี้

1) กลุ่มผู้ใช้ข้อมูล (Business Domain) ได้แก่ ผู้บริหารระดับสูง ผู้อำนวยการกอง ผู้ทำงานด้านนโยบายและวิชาการ ผู้ทำงานด้านบริการ มีหน้าที่กำหนดโจทย์หรือประเด็นที่ต้องการวิเคราะห์ข้อมูล และนำเสนอสารสนเทศที่เป็นผลจากการวิเคราะห์และนำเสนอข้อมูลมาใช้ในการประกอบการดำเนินงาน

2) กลุ่มผู้สร้างและพัฒนาระบบ ได้แก่ Data Engineer, Data Architecture, Business Analyst, Project Manager, Corporate Security, IT Operator มีหน้าที่ในการออกแบบและพัฒนาโครงสร้างพื้นฐานด้านข้อมูล รวมทั้งดูแลและบริหารจัดการข้อมูลให้อยู่ในสภาพที่พร้อมใช้งาน อย่างต่อเนื่องและปลอดภัย

3) กลุ่มผู้วิเคราะห์ ประมวลผลและแสดงผลข้อมูล ได้แก่ Data Scientist, Data Analyst, Data Visualizer มีหน้าที่ในการนำข้อมูลมาจัดกลุ่ม วิเคราะห์ และประมวลผล พร้อมทั้งพัฒนารูปแบบการแสดงผลรวมถึง Dashboard สำหรับการนำเสนอข้อมูล

นอกจากนี้ กรอบแนวทางการดำเนินงานการพัฒนาบุคลากรในการใช้ประโยชน์ข้อมูลขนาดใหญ่จะถูกแบ่งออกเป็นระยะต้น ระยะกลาง และระยะยาว เพื่อให้การดำเนินการบรรลุเป้าหมายอย่างเป็นขั้นตอน ดังแสดงในรูปที่ 5.2



รูปที่ 5.2 กรอบการดำเนินงาน 3 ระยะ

ระยะสั้น การพัฒนาแพลตฟอร์มเสริมสร้างทักษะที่เรียกว่า Sandbox สำหรับการพัฒนาบุคลากรในระยะสั้น โดยเริ่มตั้งแต่การกำหนดโจทย์ การพัฒนาโครงสร้างและระบบข้อมูล การวิเคราะห์และประมวลผลข้อมูล การแสดงผลข้อมูลเชิงประจักษ์ ตลอดไปจนถึงการนำเสนอสารสนเทศและองค์ความรู้ที่ได้จากการวิเคราะห์และประมวลผลมาใช้ในการประกอบการตัดสินใจและการดำเนินงานในรูปแบบต่างๆ แพลตฟอร์มดังกล่าวเน้นการพัฒนาบุคลากรผ่านหลักสูตรที่นำรูปแบบการพัฒนาแบบ Project based Learning มาปรับใช้ เพื่อให้ผู้เข้ารับการฝึกอบรมสามารถ “ทำได้ ทำเป็น” มากกว่าเรียนรู้จากทฤษฎี ดังนี้

ขั้นตอนที่ ๑ กำหนดโจทย์/ ประเด็นเชิงนโยบาย	ขั้นตอนที่ ๒ สรรหา/คัดเลือกบุคลากร เข้าร่วมโครงการ	ขั้นตอนที่ ๓ พัฒนาบุคลากร แบบ Project based Learning	ขั้นตอนที่ ๔ นำข้อมูลมาใช้ประโยชน์
กลไกการดำเนินงาน			
การดำเนินการร่วมกับ หลักสูตร ป.ย.ป.๑ (ปลัดกระทรวง) โดยจัด ให้มีการให้ความรู้ เกี่ยวกับการนำข้อมูล ขนาดใหญ่มาใช้ประโยชน์	การสรรหา/คัดเลือกจาก ๑) ผู้เข้ารับการอบรม/ เจ้าของโจทย์คัดเลือก บุคลากรที่มีคุณสมบัติที่ กำหนดเข้ารับการพัฒนา ทักษะบนแพลตฟอร์ม Sandbox ๒) การนำบุคลากรจาก โครงการ Policy Work and Study Team ของ ก.พ. มาสนับสนุน	การจัดการพัฒนาทักษะ บุคลากรผ่านการเรียนรู้ โดยการทำงานจริง	การให้ผู้เข้ารับการฝึกอบรม กลุ่มผู้ใช้ข้อมูล นำการ แสดงผลข้อมูลที่พัฒนาขึ้น ในขั้นตอนที่ ๓ มาจัดทำ ข้อเสนอนโยบายแห่งรัฐ
หน่วยงานที่เกี่ยวข้อง			
สกพ. นำกลไกการฝึกอบรม หลักสูตร ป.ย.ป. มา สนับสนุนการดำเนินงาน ดศ. ให้การสนับสนุนเชิง เทคนิคและวิธีการในการ กำหนดโจทย์/ประเด็น ให้กับกลุ่มผู้เข้ารับการ ฝึกอบรมหลักสูตร ป.ย.ป. ในการนำข้อมูลขนาด ใหญ่มาใช้ประโยชน์	สกพ. ประสานและคัดเลือก ผู้มีคุณสมบัติที่กำหนด เข้าร่วมโครงการ ดศ. กำหนดคุณสมบัติบุคคล ผู้เข้าร่วมโครงการ และ ร่วมคัดเลือกบุคลากร	ดศ. จัดดำเนินการพัฒนา บุคลากรผ่านเครือข่าย มหาวิทยาลัย สถาบันคุณวุฒิวิชาชีพ จัดการประเมินสมรรถนะ ผู้เรียน สศช. จัดประเมินผู้สอน สพร. พัฒนา ระบบ e-Learning สกพ. ออกแบบการสร้างแรงจูงใจ ให้บุคลากรมีความสนใจเข้า รับการฝึกอบรม เช่น การ Top-Up เงินเดือน การให้ รางวัลหรือทุนการศึกษา	สกพ. จัดให้การพัฒนา ความสามารถในการนำ ข้อมูลมาใช้ประโยชน์เป็น ส่วนหนึ่งของการฝึกอบรม หลักสูตรการพัฒนาผู้นำ ต่างๆ เช่น นบส. ป.ย.ป. เป็นต้น ดศ. จัดดำเนินการพัฒนา บุคลากรให้สามารถนำ ข้อมูลมาใช้ประโยชน์

ระยะกลาง วางแผนการบริหารจัดการกำลังคนเพื่อการใช้ประโยชน์ข้อมูลขนาดใหญ่ที่มีความคล่องตัว สามารถสนับสนุนการพัฒนานโยบาย การตัดสินใจ การบริหารจัดการ การให้บริการ และการพัฒนานวัตกรรมภาครัฐ โดยมีกรอบการดำเนินงานดังนี้

- การวางระบบการบริหารจัดการกำลังคนโดยแบ่งเป็น ๓ กลุ่มหลัก ดังนี้ (๑) กลุ่มหน่วยงานที่มีความพร้อมด้านบุคลากร สามารถดูแลบริหารจัดการระบบข้อมูล และมีบุคลากรที่มีความสามารถด้านการวิเคราะห์ ประมวลผลและแสดงผลข้อมูล แต่ยังคงมีความต้องการการสนับสนุนเทคนิคหรือการดำเนินการในบางประการ (๒) กลุ่มหน่วยงานที่อยู่ระหว่างการพัฒนา โดยมีบุคลากรที่มีความเข้าใจ และสามารถให้ข้อมูลความต้องการได้ชัดเจนระดับหนึ่ง ต้องการการสนับสนุนเชิงเทคนิคในการสร้างและพัฒนาระบบ และการวิเคราะห์ ประมวลผลและแสดงผลข้อมูล (๓) กลุ่มหน่วยงานที่ขาดแคลนบุคลากรด้านข้อมูล แต่มีความจำเป็นต้องนำข้อมูลมาใช้ประโยชน์เพื่อการตัดสินใจ การกำหนดนโยบาย หรือการบริหารจัดการต่างๆ
- นำบุคลากรที่มีทักษะและความเชี่ยวชาญเฉพาะทางจากหน่วยงานต่างๆ มารวมไว้ในหน่วยงานเดียว เช่น กระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม และพัฒนารูปแบบการทำงานในลักษณะเป็น “ทีมที่ปรึกษา” (Agile Team) สำหรับให้บริการหน่วยงานภาครัฐ โดยอาจนำที่ปรึกษาภายนอกหรือหน่วยงานเอกชนมาร่วมดำเนินงานในรูปแบบต่าง ๆ ด้วย ในการนี้ ให้มีพัฒนารูปแบบการจ้างงานใหม่นอกเหนือจาก “ข้าราชการ/พนักงานราชการ” เพื่อดึงดูดใจ และอาจมีการเพิ่ม “สายงานเฉพาะทาง” สำหรับผู้ปฏิบัติงานที่ต้องใช้องค์ความรู้เกี่ยวกับการวิเคราะห์ข้อมูล มีการดูแลทางก้าวหน้าในอาชีพ การสร้างโอกาสการพัฒนาที่ต่อเนื่องการให้ได้รับค่าตอบแทนที่เหมาะสม การพัฒนาระบบการแต่งตั้งให้สอดคล้องกับลักษณะงานที่มีความซับซ้อนและความต้องการบุคลากรที่มีสมรรถนะสูงและมีความเป็นมืออาชีพ ทั้งนี้ อาจมีการนำวุฒิปัตหรือประกาศนียบัตรรับรองมาตรฐานหรือสมรรถนะด้านการวิเคราะห์ข้อมูล มาใช้ในการบริหารจัดการกำลังคนด้วย
- พัฒนาขีดความสามารถของผู้ใช้ข้อมูลกลุ่มต่าง ๆ ได้แก่ ผู้บริหารระดับสูง ผู้อำนวยการกอง ผู้ทำงานด้านนโยบายและวิชาการ และผู้ทำงานด้านบริการให้สามารถกำหนดโจทย์หรือประเด็นที่ต้องการวิเคราะห์ข้อมูล และนำเสนอสารสนเทศที่เป็นผลจากการวิเคราะห์และนำเสนอข้อมูลมาใช้ประกอบการดำเนินงาน ได้อย่างเหมาะสม รวมทั้งยกระดับศักยภาพผู้ปฏิบัติงานด้านเทคโนโลยีสารสนเทศของหน่วยงาน โดยอย่างน้อยให้มีความสามารถในการกำกับควบคุมงานจ้างที่ปรึกษา (Project management) และในการบริหารจัดการข้อมูลอย่างปลอดภัยและมีธรรมาภิบาล (Data Management)
- นำกลไกการให้ทุนรัฐบาลมาใช้สนับสนุนการสร้างและพัฒนากำลังคนด้านการวิเคราะห์และใช้ประโยชน์ข้อมูลขนาดใหญ่ทั้งในระยะสั้นและระยะยาว
- วางระบบการบริหารองค์ความรู้และเก็บรวบรวมข้อมูลการดำเนินงานเพื่อสร้างความต่อเนื่องในการดำเนินงานด้วย

ระยะยาว การต่อยอดขยายผลโดยมีเป้าหมายเพื่อให้เกิดการแบ่งปันและสร้างประโยชน์ในทรัพยากรข้อมูลของทุกภาคส่วน เพื่อการพัฒนาต่อยอดการสร้างมูลค่าเพิ่มให้กับระบบเศรษฐกิจ สร้างความมั่นคงให้กับสังคม และสร้างการพัฒนาที่ยั่งยืน โดยเน้นส่งเสริมสนับสนุนการพัฒนาความร่วมมือกับหุ้นส่วนภาคี (PPP) ทั้งภายในและระหว่างประเทศภายใต้เจตจำนงค์ร่วมกัน โดยร่วมสร้าง Open Government Data Platform for Business and Citizen ให้ความรู้และสร้างภูมิคุ้มกันเพื่อให้ประชาชนมีความเข้าใจที่ถูกต้องเกี่ยวกับการดูแลความปลอดภัยของข้อมูลส่วนบุคคล และการนำข้อมูลเปิดภาครัฐมาใช้เพื่อพัฒนาต่อยอดและการสร้างมูลค่าเพิ่ม

5.2 เครือข่ายมหาวิทยาลัยด้านวิทยาศาสตร์ข้อมูล

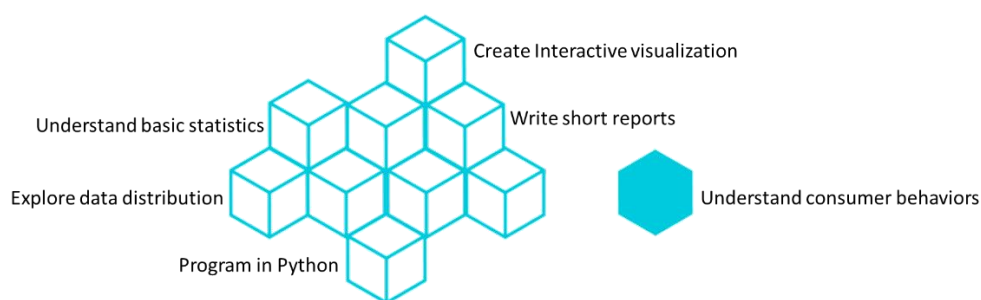
ทั้งนี้เพื่อเป็นการสนับสนุนการดำเนินการตามข้อเสนอการพัฒนาศูนย์กลางข้อมูลขนาดใหญ่ที่สำนักงานคณะกรรมการข้าราชการพลเรือนเสนอขึ้น กระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคมจึงได้สร้างความร่วมมือกับเครือข่ายมหาวิทยาลัยชั้นนำในประเทศไทยเพื่อวางแผนในการสร้างให้บุคลากรภาครัฐมีทักษะในด้านการวิเคราะห์ข้อมูลและสามารถปฏิบัติงานได้อย่างเป็นรูปธรรม โดยการสร้างเครือข่ายมหาวิทยาลัยด้านวิทยาศาสตร์ข้อมูล (Data Science) เพื่อการพัฒนาศูนย์กลางภาครัฐในการวิเคราะห์และจัดการข้อมูลขนาดใหญ่

เครือข่ายมหาวิทยาลัยดังกล่าว มีวัตถุประสงค์เพื่อระดมผู้เชี่ยวชาญและผู้สอนในมหาวิทยาลัยมาร่วมออกแบบโครงสร้างกรอบหลักสูตรกลางและจัดการเรียนการสอนแบบนอกห้องเรียนให้กับเจ้าหน้าที่ของหน่วยงานภาครัฐ ซึ่งกรอบหลักสูตรกลางรวมถึงเนื้อหาที่จำเป็น (Materials) วิธีการเรียนการสอนที่เหมาะสม (Pedagogy) ทักษะ/องค์ความรู้ที่จะได้ (Learning Outcome) และการวัดผล/ประเมินทักษะ (Assessment) โดยกรอบหลักสูตรจะสร้างทักษะด้านวิทยาศาสตร์ข้อมูลทั้งการวิเคราะห์และจัดการข้อมูลขนาดใหญ่ โดยเน้นพัฒนาให้คนทำเป็นมากกว่ารู้เพียงแค่เรียนเรื่องอะไรเท่านั้น ซึ่งคณะผู้สอนจากแต่ละมหาวิทยาลัยจะสามารถปรับเนื้อหาจากกรอบหลักสูตรกลางให้เข้ากับโจทย์ปัญหาของหน่วยงานภาครัฐแต่ละแห่งได้ตามความเหมาะสม

การเรียนการสอนจะเป็นเชิงปฏิบัติ (Bootcamp) โดยมีแนวทางการดำเนินการ ดังนี้

- 1) ให้มหาวิทยาลัยเสนอกลไกการเรียนการสอนในรูปแบบปฏิบัติ (Bootcamp) ที่เหมาะสม เช่น
 - a. Lecture โดยใช้การจัดการเรียนรู้แบบใช้ปัญหาเป็นฐาน (Problem-Based Learning) โดยจัดเนื้อหาให้อยู่ในบริบทของงานที่เกี่ยวข้องโดยตรงกับผู้เรียน
 - b. การประชุมเชิงปฏิบัติการโดยเน้นการร่วมกันลงมือปฏิบัติ (Hands-on Workshop)
 - c. การจัดการเรียนรู้แบบใช้โครงงานเป็นฐาน (Project-Based Learning) โดยมอบหมายโครงงานที่อยู่ในบริบทของงานที่เกี่ยวข้องโดยตรงกับผู้เรียน
 - d. การอภิปรายและการสนทนาอภิปรายในหัวข้อปัญหาในหัวข้อเรียน
 - e. การนำเสนอผลงาน (Showcases) ตอนจบหลักสูตรอย่างเป็นทางการ

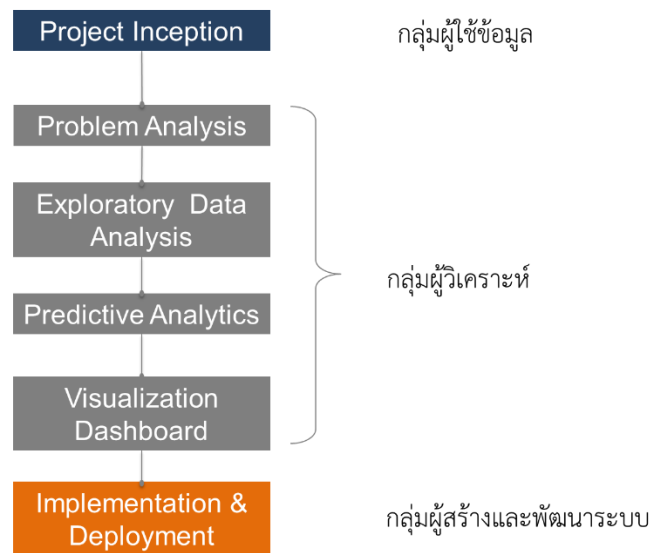
- 2) แต่ละ bootcamp ดำเนินงานเป็นอิสระต่อกัน โดยอยู่ภายในกรอบของโครงสร้างหลักสูตรกลางที่เครือข่ายมหาวิทยาลัยเห็นชอบ โดยแต่ละค่ายมีการออกแบบชุดองค์ความรู้และทักษะ ที่รวมเรียกว่า Learning Outcome ที่เหมาะสม
- 3) มีการให้หนังสือรับรอง (Certificate) สำหรับแต่ละ bootcamp สำหรับผู้ที่ผ่านการทดสอบความรู้ความเข้าใจและการลงมือปฏิบัติ
- 4) ผู้เรียนสามารถสะสมผลการเรียนรู้ (Satisfied learning outcome) ในรูปแบบของ Micro-credential สำหรับใช้เป็นส่วนหนึ่งของการได้หนังสือรับรองหรือประกาศนียบัตร และ/หรือเพื่อใช้เป็นส่วนหนึ่งของปริญญาระดับมหาบัณฑิตในภายหลัง ซึ่งการใช้กลไกในลักษณะนี้ เครือข่ายมหาวิทยาลัยฯ ต้องร่วมกันแตกเนื้อหาในแต่ละหลักออกเป็นทักษะ/องค์ความรู้ย่อย (Micro-credential) ดังตัวอย่างในรูปที่ 5.3 และออกแบบวิธีการประเมินทักษะ/องค์ความรู้ย่อยให้ได้ประสิทธิภาพ สำหรับผู้สอนใช้เป็นเครื่องมือในการประเมินผล



รูปที่ 5.3 ตัวอย่างทักษะ/องค์ความรู้ย่อย (Micro-credential)

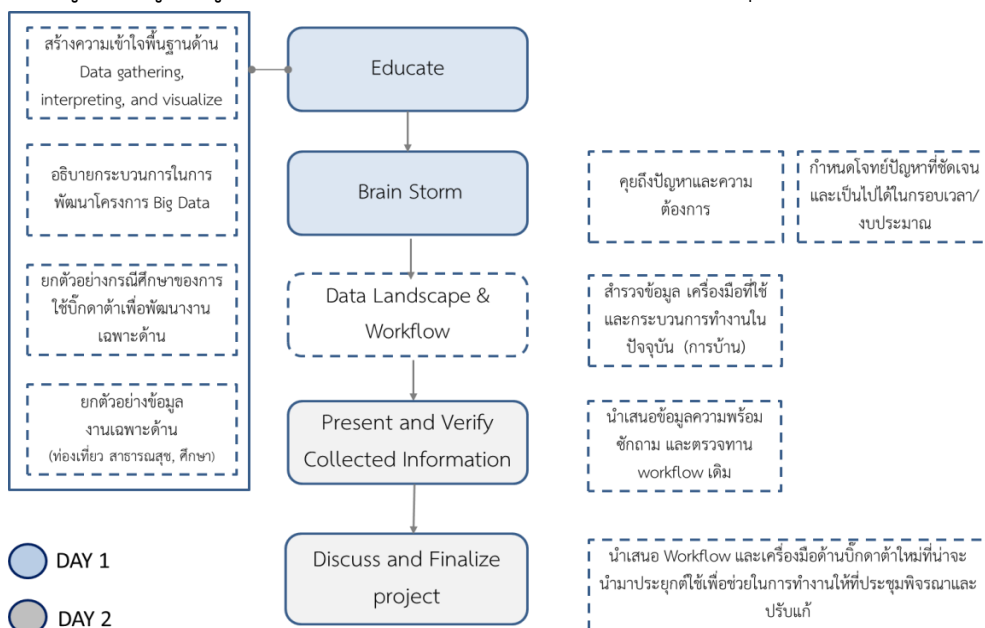
ทั้งนี้ การออกแบบกรอบหลักสูตรควรที่จะสอดคล้องกับกระบวนการการดำเนินโครงการปึกค้ำที่ เหมาะสม โดยขั้นตอนแรกของกระบวนการดำเนินงานดังกล่าว นักวิทยาศาสตร์ข้อมูล ผู้บริหาร และ คณะทำงานด้านปึกค้ำของแต่ละกลุ่มงานจะร่วมกันกำหนดโจทย์ที่เหมาะสมสำหรับการนำร่อง (Project Inception) และสำรวจข้อมูลที่มีอยู่ในปัจจุบันตามโจทย์นำร่องที่กำหนดเพื่อประเมินความเป็นไปได้และแปลง จากความต้องการในเชิงปัญหาให้เป็นข้อกำหนดในเชิงข้อมูลและระบบ (Problem Analysis) จากนั้น นักวิทยาศาสตร์ข้อมูลทำการศึกษาการกระจายตัวของข้อมูลและหาความสัมพันธ์ระหว่างตัวแปรโดยใช้ ตัวอย่างต้องการตัวอย่างข้อมูลจริงเพื่อระบุข้อมูลที่ต้องการเพิ่มเติม (Exploratory Data Analysis) และ เตรียมข้อมูลมาใช้ในการสร้างแบบจำลองหรือโมเดลทางคณิตศาสตร์เพื่อการทำนายโดยใช้เทคนิคและ อัลกอริทึมต่างๆ รวมทั้งทดสอบความแม่นยำของโมเดลคณิตศาสตร์ที่สร้างขึ้น (Predictive Analytics) และ ออกแบบวิธีการแสดงผลโดยเลือกมิติของข้อมูลที่เหมาะสมเพื่อให้คณะทำงานทดลองใช้และสื่อสารกับทีม ผู้บริหารและผู้ปฏิบัติ ให้สามารถนำเอาความเข้าใจดังกล่าวไปแปลงเป็นแผนการพัฒนาต่อยอด (Visualization Dashboard) จากนั้นนักพัฒนาระบบสารสนเทศจะสามารถพัฒนาโปรแกรมตามรูปแบบของ โมเดลคณิตศาสตร์ที่วางไว้ ตั้งค่าให้โปรแกรมให้ประมวลผลโมเดลแบบอัตโนมัติตามความถี่ที่วางแผนไว้ และ

ติดตั้งระบบซอฟต์แวร์เพื่อการใช้งานจริง (Implementation and Deployment) ซึ่งขั้นตอนต่างๆ ดังกล่าว จะต้องการบุคลากร 3 กลุ่มอย่างชัดเจนดังแสดงในรูปที่ 5.4



รูปที่ 5.4 ขั้นตอนการทำโครงการบิกดาต้า

ดังนั้น กรอบหลักสูตรจึงต้องถูกออกแบบสำหรับทั้ง 3 กลุ่มเป้าหมาย โดยตัวอย่างหลักสูตรของกลุ่มที่ 1 ผู้ใช้ข้อมูลแสดงในรูป 5.5 ตัวอย่างกรอบหลักสูตรดังกล่าวเป็นหลักสูตร 2 วัน กระบวนการการเรียนการสอนจะถูกแบ่งเป็นช่วงบรรยายและช่วงปฏิบัติ โดยช่วงบรรยายเน้นการสร้างความรู้เข้าใจเรื่องบิกดาต้าผ่านกรณีศึกษาในบริบทที่เกี่ยวข้องกับองค์กร และช่วงปฏิบัติจะเน้นให้ผู้เรียนเข้าใจและสามารถเลือกโจทย์ปัญหาที่เหมาะสมกับองค์กร เป็นไปตามวิสัยทัศน์ผู้บริหาร และเป็นโจทย์ที่สามารถดำเนินการได้อย่างเป็นรูปธรรม โดยคำนึงถึงข้อมูลที่มีอยู่ ข้อมูลที่หาเพิ่มเติมได้ และกระบวนการทำงานในปัจจุบัน

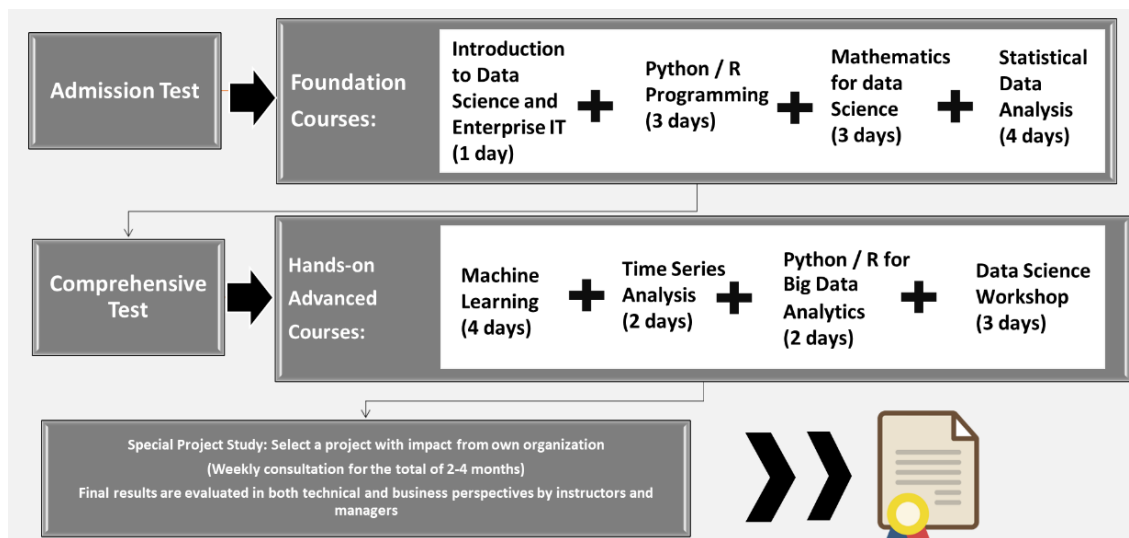


รูปที่ 5.5 ตัวอย่างกรอบหลักสูตรของกลุ่มผู้ใช้ข้อมูล (Problem Selection Workshop and Basic BI)

ส่วนกรอบหลักสูตรของกลุ่มที่ 2 ผู้วิเคราะห์ข้อมูล จะมี 2 ส่วนดังแสดงในรูปที่ 5.6 และ 5.7 ซึ่งรายละเอียดและลักษณะกลไกการฝึกอบรมของตัวอย่างหลักสูตร Data Science Basic ในรูปที่ 5.6 นั้น การฝึกอบรมเริ่มต้นด้วยการทดสอบก่อนการอบรม (Admission test) เพื่อคัดเลือกบุคลากรที่เหมาะสมที่จะเข้ารับการอบรม การทดสอบระหว่างการอบรม (Comprehensive Test) เพื่อทดสอบความรู้ความเข้าใจในเนื้อหาของผู้รับการอบรม โดยอาจจะเป็นช่วงกลางของระยะเวลาอบรมทั้งหมด รวมทั้งการทดสอบหลังการอบรมเสร็จสิ้นโดยผู้รับการอบรมจะได้รับการทดสอบโดยวิธีนำเสนอโครงการให้กับอาจารย์และผู้บริหารจากหน่วยงานต้นสังกัดเพื่อเป็นการประเมินด้านเทคนิคและมุมมองจากบริบทของงานที่เกี่ยวข้องกับผู้รับการอบรม ซึ่งจะคล้ายกับการสอบวิทยานิพนธ์ของนักศึกษาระดับมหาบัณฑิต หลักสูตรใช้เวลาเรียน 20-30 วันทำการ ในช่วงระยะ 5-6 เดือน

เนื้อหาหลักของหลักสูตรนี้จะเน้นการปูพื้นฐานด้านการวิเคราะห์ข้อมูลเชิงสถิติ การทดสอบสมมติฐาน การวิเคราะห์ปัจจัยและเทรนด์ ผู้เข้าอบรมจะได้ทำการวิเคราะห์ข้อมูลจริงโดยใช้หลักการทางสถิติเพื่อให้ได้ผลลัพธ์ที่เป็นรูปธรรม รวมทั้งมีการอบรมเชิงปฏิบัติการโดยการทำงานกับข้อมูลและเน้นการเข้าไปวิเคราะห์ข้อมูลจริง โดยผู้สอนจะให้คำแนะนำในการวิเคราะห์ข้อมูลในแต่ละขั้น ตั้งแต่การวิเคราะห์ปัญหา การสำรวจข้อมูล การเลือกใช้วิธีการในการวิเคราะห์ การสร้างแบบจำลองและการประเมินผล รวมถึงแนะนำในการนำเอาผลไปต่อยอด ตัวอย่างเนื้อหาคือ

- 1) Introduction to Data Science and data analytic mindset
- 2) Business and data analytical thinking
- 3) Problem analysis (Problem statement and decomposition, Best practices and solution)
- 4) Exploratory data analysis (Assumption checking, Feature selection, Data visualization and Preparation)
 - Statistical thinking and descriptive statistics (Introduction to statistics, data sampling, biases, terms, mean, median, mode, frequency, percentile and quartile, IQR, Histogram, Boxplot)
 - Probability theory (Probability distribution function, cumulative distribution function)
 - Hypothesis testing (Null and alternative hypotheses, z-test and t-test, paired t-test, chi-square test, ANOVA, correlation)



Learning Outcome Examples:

Being able to capture, clean, and manipulate data through programming interface

Understand data integration and gain basic hands-on experience (Tabular, file/documents, streaming)
Acquire skill to program in R/Python

Being able to analyze data using basic techniques

Understand statistics for data science
Explore data distribution using R or Python
Analyze data with data mining techniques
Being able to make decision using insights
Understand consumer behaviors
Understand business analytics and data-driven decisions

รูปที่ 5.1 ตัวอย่างกรอบหลักสูตรกลุ่มผู้วิเคราะห์ข้อมูล (Data Science Basic)

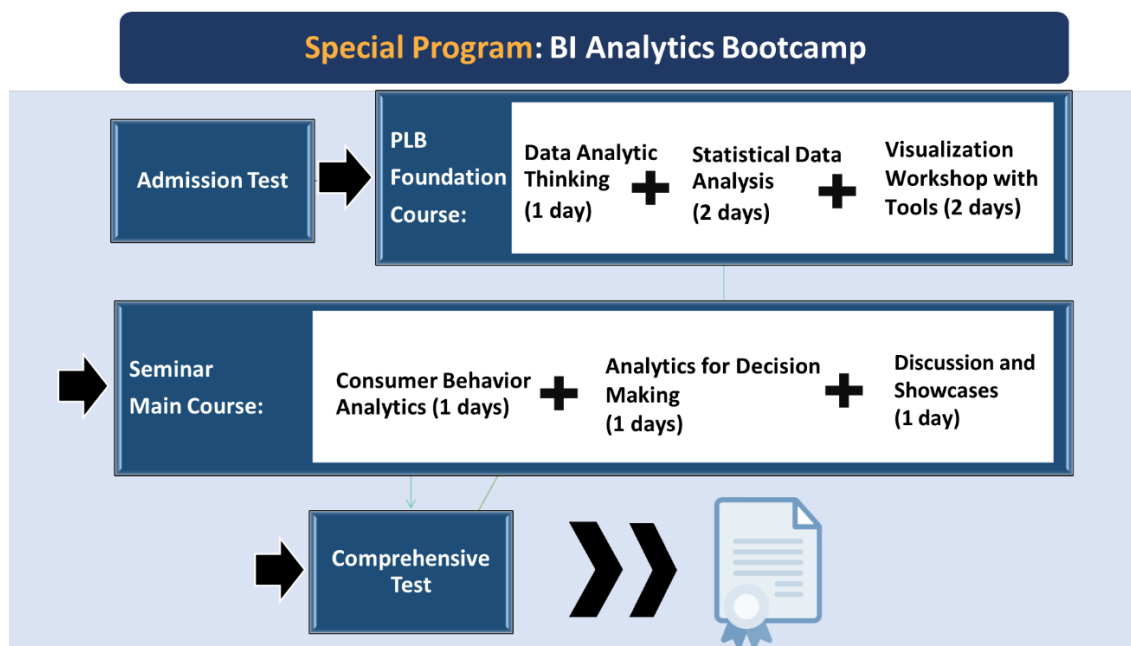
5) Model building, training, validation, and evaluation (Defining test data, Evaluation strategies, Result analysis)

- 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)

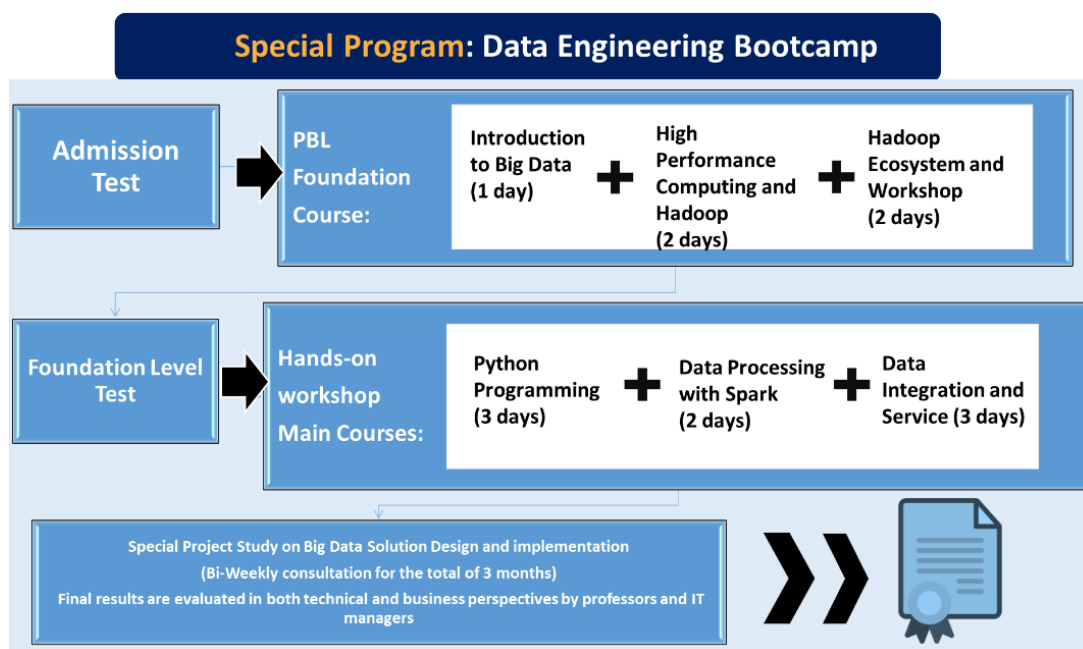
ผู้เรียนจะสามารถตั้งคำถามจากข้อมูล วิเคราะห์คำถาม เตรียมข้อมูล สร้างแบบจำลอง และวิเคราะห์ผลจากแบบจำลองได้

ตัวอย่างกรอบหลักสูตรด้าน Business Intelligence แสดงในรูปที่ 5.7 จะใช้เวลาเรียน 5-8 วันทำการ ในช่วงระยะเวลา 1-2 เดือน และตัวอย่างกรอบตัวอย่างหลักสูตรสำหรับกลุ่มผู้สร้างและพัฒนาระบบ แสดงในรูปที่ 5.8 ซึ่งต้องการเวลาเรียน 15-20 วันทำการในช่วงระยะเวลา 4-5 เดือน ทั้งนี้การออกแบบกรอบ

หลักสูตรทั้งหมดจะเป็นหน้าที่ของคณะทำงานในเครือข่ายมหาวิทยาลัยด้านวิทยาศาสตร์ข้อมูลต่อไป ซึ่งขณะนี้ยังอยู่ในกระบวนการทำงาน ดังนั้นรูปที่ 5.5-5.8 จึงเป็นการแสดงตัวอย่างกรอบหลักสูตรเท่านั้น



รูปที่ 5.7 ตัวอย่างกรอบหลักสูตรกลุ่มผู้วิเคราะห์ข้อมูล (Business Intelligence Analytics)



รูปที่ 5.8 ตัวอย่างกรอบหลักสูตรกลุ่มผู้สร้างและพัฒนาระบบ (Data Engineering)

นอกจากนี้กระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคมยังมีแผนที่จะสร้างความร่วมมือกับสถาบันพัฒนาบุคลากรดิจิทัลภาครัฐ สำนักงานพัฒนารัฐบาลดิจิทัล เพื่อเตรียมความพร้อมสำหรับระบบ E-Learning ด้านการวิเคราะห์ข้อมูลขนาดใหญ่ในอนาคต และขณะนี้ทางกระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคมได้เตรียมการ

ด้านการทำบันทึกข้อตกลงร่วมกับเครือข่ายมหาวิทยาลัยในปี 2562 จำนวน 14 มหาวิทยาลัย และคาดหวังว่า จะมีการขยายเครือข่ายในปีต่อไปอย่างต่อเนื่อง โดยมหาวิทยาลัยที่ร่วมโครงการจะเสนอรายชื่อผู้สอนที่มีความเชี่ยวชาญและมีประสบการณ์ในการทำงานด้านการวิเคราะห์ข้อมูลขนาดใหญ่ เพื่อเข้าร่วมในการ ออกแบบโครงสร้างหลักสูตรกลางพร้อมจัดทำสื่อการสอนพื้นฐาน รวมทั้งวิธีการประเมิน ทักษะ/องค์ความรู้ ย่อยต่างๆ โดยกระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคมจะเป็นเจ้าภาพในการบริหารจัดการ และจัดสัมมนาเรื่อง วิธีการสอนบุคลากรภาครัฐให้กับผู้สอนจากมหาวิทยาลัยในเครือข่าย รวมทั้งออกแบบวิธีการประเมินผู้สอนที่ เหมาะสม ทั้งนี้เป้าหมายของเครือข่ายความร่วมมือดังกล่าวคือ การพัฒนาบุคลากรภาครัฐทางด้าน วิทยาศาสตร์ข้อมูล (Data Sciences) อย่างน้อย 250 คน ด้านวิศวกรรมข้อมูล (Data Engineering) อย่างน้อย 250 คน และด้านวิเคราะห์ธุรกรรมข้อมูล (Business Analytics) อีกอย่างน้อย 800 คน โดยวางเป้าหมายให้ สามารถเริ่มดำเนินการได้ในเดือนมิถุนายน 2562 นี้

บทที่ 6 กรณีตัวอย่างการใช้ประโยชน์ข้อมูล

หลังจากที่กระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคมได้เริ่มโครงการดังที่กล่าวในข้างต้นของเอกสารฉบับนี้เพื่อสนับสนุนให้เกิดการใช้ประโยชน์ข้อมูลในภาครัฐแล้ว ทางกระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคมยังได้วางแผนที่จะร่วมมือกับหน่วยงานภาครัฐจำนวนหนึ่งในการพัฒนาระบบนำร่องการวิเคราะห์ข้อมูลขนาดใหญ่ในหลากหลายสาขา เช่น การส่งเสริมการท่องเที่ยว การแพทย์และการสาธารณสุข การใช้ข้อมูลจากสื่อสังคมออนไลน์กับการสำรวจความคิดเห็น การบริหารงานจราจร การพัฒนาการศึกษา ฯลฯ เพื่อให้บุคลากรภาครัฐได้เห็นถึงประโยชน์ของการใช้ข้อมูลประกอบการสร้างนโยบายและแผนงานในอนาคต รวมถึงการเพิ่มประสิทธิภาพและพัฒนากิจการดำเนินงานในปัจจุบัน โดยตัวอย่างของกรณีศึกษาต่างๆ สรุปได้ดังต่อไปนี้

6.1 การเพิ่มศักยภาพด้านการท่องเที่ยวเมืองรอง

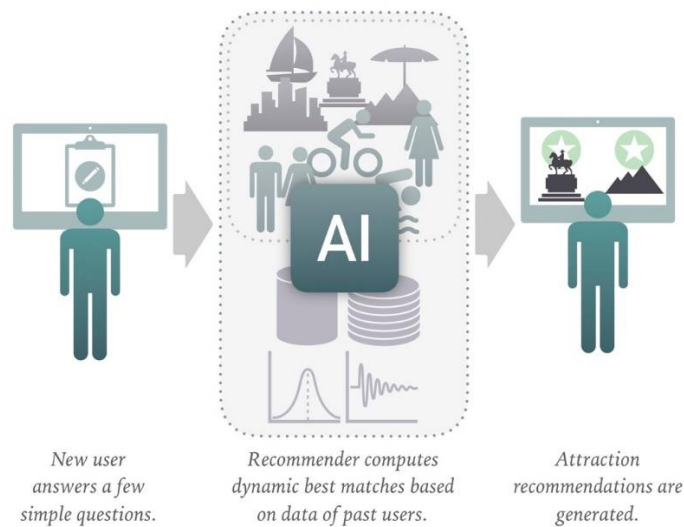
กระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคมดำเนินงานร่วมกับการท่องเที่ยวแห่งประเทศไทยและกระทรวงการท่องเที่ยวและกีฬาในการวิเคราะห์และนำเสนอสถานที่ท่องเที่ยวที่เหมาะสมกับนักท่องเที่ยวแต่ละกลุ่ม ด้วยระบบการแนะนำสถานที่ท่องเที่ยวอัตโนมัติ โดยมีจุดมุ่งหมายในการเพิ่มความพึงพอใจและกระตุ้นการใช้จ่ายของนักท่องเที่ยว รวมถึงช่วยกระจายนักท่องเที่ยวสู่เมืองรอง (Emerging Destinations) ระบบนี้จะใช้ข้อมูลลักษณะและพฤติกรรมของนักท่องเที่ยวในภาพรวม (Demand Side) ร่วมกับข้อมูลสถานที่ท่องเที่ยว กิจกรรมการท่องเที่ยว ฯลฯ (Supply Side) เพื่อนำมาทำการวิเคราะห์สร้างโมเดลทางคณิตศาสตร์ที่สามารถให้คำแนะนำสถานที่ท่องเที่ยวอัตโนมัติได้ และสามารถนำมาสนับสนุนการท่องเที่ยวในมิติต่างๆ ได้ด้วย เช่น การจัดทำแคมเปญโฆษณาแหล่งท่องเที่ยวในเมืองรองที่มีประสิทธิผล การนำเสนอแหล่งท่องเที่ยวเมืองรองให้แก่นักท่องเที่ยว เป็นต้น โดยเป้าหมายหลักของโจทย์นี้ คือ นักท่องเที่ยวชาวต่างชาติ

ในภาพรวม ระบบการแนะนำสถานที่ท่องเที่ยวอัตโนมัติ ประกอบด้วยขั้นตอนหลัก 3 ขั้นตอน คือ 1) นักท่องเที่ยวคนใหม่เข้ามาใช้งานระบบ (เพื่อให้การกล่าวถึงนักท่องเที่ยวรายใหม่นี้มีความกระชับและเข้าใจง่าย เราจะเรียกนักท่องเที่ยวคนนี้ว่า นาย ก.) ตอบคำถามสั้นๆ เพื่อให้ระบบทำความรู้จักกับนาย ก., 2) ระบบ AI คำนวณหาสถานที่ท่องเที่ยวที่เหมาะสมกับลักษณะเชิงพฤติกรรมของนาย ก. และ 3) ระบบนำเสนอสถานที่ท่องเที่ยวต่างๆ กลับมาให้นักท่องเที่ยว ดังแสดงในรูปที่ 6.1

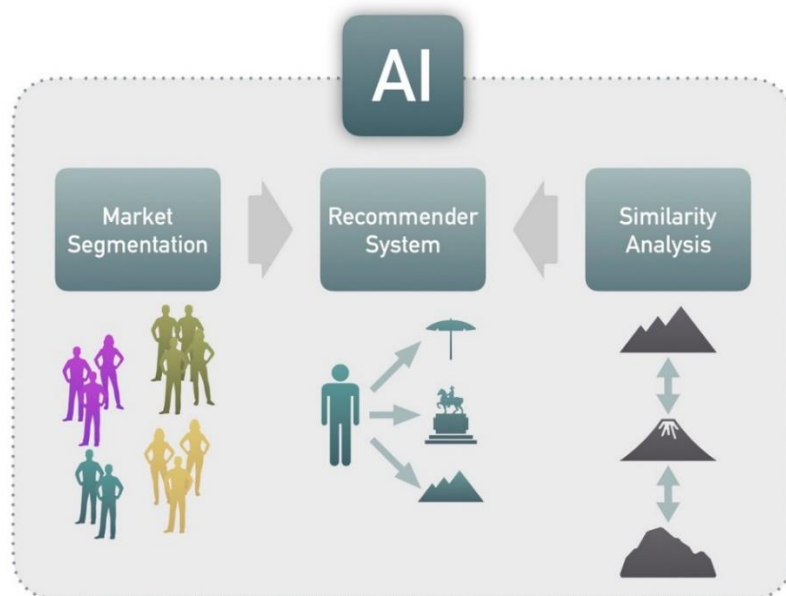
ระบบ AI ดังที่กล่าวไว้ข้างต้นนั้น ประกอบด้วยองค์ประกอบหลักซึ่งมีการสร้างจากโมเดลทางคณิตศาสตร์ อยู่ 3 องค์ประกอบด้วยกัน ดังแสดงในรูปที่ 6.2

องค์ประกอบแรก คือ โมเดลการจัดกลุ่มนักท่องเที่ยว (Market Segmentation) ซึ่งไม่จำเป็นจะต้องเป็นมิติเดียวกัน เช่น กลุ่มหนึ่งชอบเที่ยวจังหวัดใกล้ๆ กัน กลุ่มหนึ่งชอบเที่ยวเน้นความโอ่อ่าหรูหรา อีกกลุ่มหนึ่งชอบกิจกรรมผจญภัยแต่เน้นประหยัด เป็นต้น ซึ่งวิเคราะห์ข้อมูลการจัดกลุ่มจากข้อมูลการสำรวจนักท่องเที่ยว โดยการท่องเที่ยวแห่งประเทศไทย ซึ่งมีการสำรวจนักท่องเที่ยวต่างชาติมากกว่า 120,000 คนในเวลา 4 ปีที่ผ่านมา มีการเก็บข้อมูลทั้งหมด 4 มุมมองเพื่อใช้ในการแบ่งกลุ่มนักท่องเที่ยว คือ เชิงลักษณะประชากร (Demographic) เชิงภูมิศาสตร์ (Geographic) เชิงลักษณะจิตวิทยา (Psychographic)

และเชิงพฤติกรรม (Behavioral) ซึ่งสามารถจัดกลุ่มนักท่องเที่ยวจากข้อมูลเหล่านี้ด้วยเทคนิควิธีการวิเคราะห์จัดกลุ่ม (Clustering Analysis)



รูปที่ 6.1 ภาพรวมของระบบการแนะนำสถานที่ท่องเที่ยวอัตโนมัติในมุมมองของผู้ใช้งาน



รูปที่ 6.2 องค์ประกอบในการสร้างระบบการแนะนำสถานที่ท่องเที่ยวอัตโนมัติ

องค์ประกอบที่สอง คือ โมเดลคำนวณค้นหาสถานที่ท่องเที่ยว (Recommender System) จากกลุ่มที่เราคำนวณด้วยข้อมูลนักท่องเที่ยวในอดีตจากองค์ประกอบที่หนึ่ง ระบบสามารถหากลุ่มที่ใกล้เคียงกับนาย ก. มากที่สุด และใช้กลุ่มนักท่องเที่ยวนี้ เพื่อค้นหาสถานที่ท่องเที่ยวที่เหมาะสม โดยจะมีการอ้างอิงกับข้อมูลขนาดใหญ่อีกแหล่งหนึ่ง คือ TripAdvisor ซึ่งเป็นแหล่งร่วมข้อมูลการให้คะแนนสถานที่ท่องเที่ยวเขียนโดยนักท่องเที่ยวจำนวนหลายล้านคนจากทั่วโลก ข้อมูลคะแนนดังกล่าวจะช่วยให้ระบบ AI นำเสนอสถานที่ท่องเที่ยวที่มีแนวโน้มจะสร้างความพึงพอใจให้กับนักท่องเที่ยวรายใหม่ (นาย ก.) มากที่สุด โดยอ้างอิงกับลักษณะกลุ่มนักท่องเที่ยวที่นาย ก. ได้รับการจำแนกจากองค์ประกอบแรก

องค์ประกอบที่สาม คือ โมเดลคำนวณหาสถานที่ท่องเที่ยวที่คล้ายคลึงกัน (Similarity Analysis) เพื่อแนะนำสถานที่ท่องเที่ยวที่ยังไม่แพร่หลาย หรือไม่มีข้อมูลการไปเที่ยวมากนักในเมืองรอง ความคล้ายคลึงในที่นี้อาจมีความหมายในหลายมิติ เช่น อยู่จังหวัดใกล้เคียงกัน ขนาดสถานที่ท่องเที่ยวพอๆกัน หรือเป็นประเภทวัดประเภทภูเขา เหมือนกัน องค์ประกอบนี้จะช่วยเสริมองค์ประกอบที่สอง โดยนอกจากระบบ AI จะแนะนำสถานที่ท่องเที่ยวในเมืองรองเข้ามาแนะนำเพิ่มเติมด้วย โดยยังอิงกับกลุ่มเชิงพฤติกรรมและรสนิยมของนักท่องเที่ยวคนใหม่ (นาย ก.) ซึ่งในการวิเคราะห์เราจะใช้ข้อมูลจากกระทรวงการท่องเที่ยวและกีฬา ที่มีการสำรวจข้อมูลแหล่งท่องเที่ยวและกิจกรรมกว่า 8,000 รายการ

ระบบ AI เพื่อแนะนำนักท่องเที่ยวนี้ อาจมีการปรับเปลี่ยนข้อมูลที่ใช้เพื่อสร้างระบบให้มีประสิทธิภาพและประสิทธิผลตามความเหมาะสม ซึ่งข้อมูลจากแหล่งอื่นๆ ที่สามารถนำมาช่วยในการวิเคราะห์ได้แก่

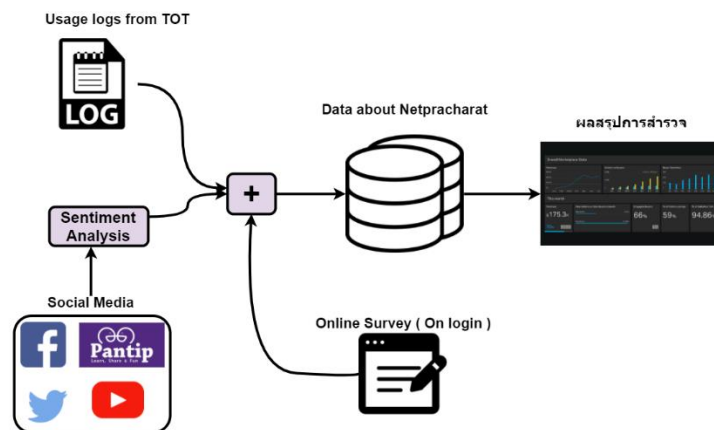
- ข้อมูลจากหน่วยงานภาครัฐ เช่น ข้อมูลโรงแรมที่จดทะเบียนจากกระทรวงมหาดไทยและข้อมูลสำรวจจากสำนักงานสถิติแห่งชาติ, ข้อมูลการเข้าออกประเทศของนักท่องเที่ยวจากสำนักงานตรวจคนเข้าเมือง, ข้อมูลการเข้าพักโรงแรมของนักท่องเที่ยวจากกระทรวงการท่องเที่ยวและกีฬา, รวมถึงข้อมูลอื่นๆ ที่ได้มีการเก็บรวบรวมไว้ในฐานข้อมูลของการท่องเที่ยวแห่งประเทศไทย และกระทรวงการท่องเที่ยวและกีฬา
- ข้อมูลจากแหล่งข้อมูลภายนอก เช่น ข้อมูลเที่ยวบินทั่วโลกจาก IATA (International Air Transport Association), ข้อมูลการเข้าพักใน Airbnb โดยบริษัท AirDNA, ข้อมูลบนสื่อสังคมออนไลน์ เช่น Instagram และ Wongnai, ข้อมูลการค้นหาเกี่ยวกับประเทศไทยจาก Google Trends, ข้อมูลการใช้จ่ายของนักท่องเที่ยวจากบริษัทบัตรเครดิต เช่น Mastercard, Visa

ระบบการแนะนำสถานที่ท่องเที่ยวอัตโนมัตินี้ สามารถนำไปสู่การพัฒนาแอปพลิเคชันหรือหน้าเว็บไซต์ที่การท่องเที่ยวแห่งประเทศไทย หรือกระทรวงการท่องเที่ยวและกีฬา ใช้เสริมกับระบบที่มีอยู่แล้วได้ เพื่อเปิดให้ผู้คนจากทั่วโลกสามารถเข้ามาใช้ระบบนี้เพื่อค้นหาและวางแผนการเข้ามาเที่ยวในประเทศไทย โดยระบบสามารถช่วยเพิ่มพูนความพึงพอใจและกระตุ้นการใช้จ่ายของนักท่องเที่ยว ในขณะที่เดียวกันสามารถกระจายนักท่องเที่ยวไปยังเมืองรองใหม่ๆ ได้ด้วย

6.2 การวิเคราะห์สื่อสังคมออนไลน์และพัฒนาเครื่องมือเพื่อรวบรวมข้อมูลสำรวจความคิดเห็นของภาครัฐ : กรณีศึกษาเน็ตประชารัฐ

กรณีศึกษาของสำนักงานสถิติแห่งชาติ ภายใต้กระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม ทำการศึกษาพัฒนาเครื่องมือหรือแพลตฟอร์มสำหรับการสำรวจความคิดเห็นและความพึงพอใจต่อนโยบายต่างๆ ของรัฐบาล โดยใช้โครงการเน็ตประชารัฐ ของกระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม เป็นกรณีศึกษานำร่องสำหรับการทดสอบแพลตฟอร์มนี้ ในปัจจุบันการสำรวจเกี่ยวกับความความเห็นและความพึงพอใจของประชาชนดำเนินการโดยการส่งเจ้าหน้าที่ลงพื้นที่เพื่อพบประชาชนกลุ่มตัวอย่างโดยตรง และทำการเก็บข้อมูลผ่านทางแบบสอบถามที่เป็นกระดาษ หรือผ่านแท็บเล็ตของเจ้าหน้าที่ ซึ่งวิธีการดังกล่าวใช้กำลังคน เวลา และ

งบประมาณค่อนข้างมาก โครงการนี้จึงมุ่งเน้นการสร้างเครื่องมือสำหรับการทำการสำรวจในลักษณะดังกล่าว โดยนำเทคโนโลยีดิจิทัลมาพัฒนากระบวนการสำรวจให้มีประสิทธิภาพมากยิ่งขึ้น โดยแพลตฟอร์มนี้มุ่งเน้นการใช้ข้อมูลที่มีอยู่ในโลกดิจิทัลอยู่แล้วนั่นคือ 1) ข้อมูลความพึงพอใจหรือความคิดเห็นในรูปแบบของภาษาเขียนจากสื่อสังคมออนไลน์ 2) ข้อมูลการใช้งานเน็ตประชารัฐจากเครื่องแม่ข่ายผู้ให้บริการเน็ตประชารัฐ 3) การสำรวจข้อมูลผ่านการตอบแบบสอบถามออนไลน์ โดยมีการเลือกกลุ่มที่จะสำรวจด้วยเทคโนโลยีการตลาดแบบดิจิทัล โดยแบบสอบถามออนไลน์จะใช้สำหรับประเด็นที่ไม่พบในสื่อสังคมออนไลน์ รูปที่ 6.3 แสดงแหล่งข้อมูลและกรอบวิธีการ



รูปที่ 6.3 โครงสร้างแพลตฟอร์มสำหรับการสำรวจความคิดเห็นของประชาชนแบบดิจิทัล

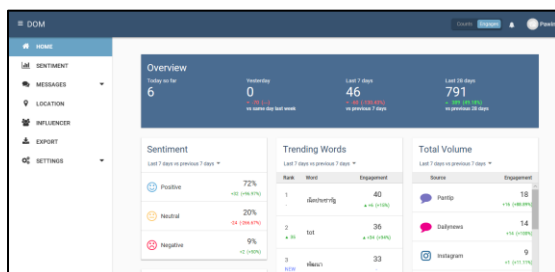
สำหรับการวิเคราะห์ข้อมูลและสำรวจในประเด็นต่างๆ จะอ้างอิงจากแบบสอบถามเดิมที่ดำเนินการโดยสำนักงานสถิติแห่งชาติ ในการสำรวจความคิดเห็นของประชาชนที่มีต่อโครงการเน็ตประชารัฐ โดยนำข้อมูลจากแพลตฟอร์มนี้เทียบกับข้อมูลจากวิธีเดิมเพื่อหาความเหมือนหรือความต่างกันในลักษณะแนวโน้มของข้อมูล (Correlation)

จากการศึกษาแบบสอบถามพบว่าข้อมูลทั้งหมด 5 ประเด็นด้วยกัน คือ 1) ข้อมูลลักษณะทางประชากรของผู้ใช้งานเน็ตประชารัฐ 2) ประสิทธิภาพการใช้งานและปัญหาที่พบในการใช้งาน 3) วัตถุประสงค์การใช้เน็ตประชารัฐของประชาชน 4) บริการอื่น ๆ ที่ประชาชนอยากให้มีผ่านทางเน็ตประชารัฐ เช่น การฝึกอบรมเกี่ยวกับใช้เน็ตสำหรับการค้าขาย 5) ค่าบริการที่เหมาะสมสำหรับการต่อเน็ตประชารัฐเข้าบ้านโดยตรง

การใช้ข้อมูลจากแหล่งที่มีอยู่แล้วในโลกดิจิทัลต่างๆ จะสามารถนำมาใช้ตอบในทั้ง 5 ประเด็นนี้ได้ ดังนี้ 1) ประเด็นข้อมูลลักษณะทางประชากรของผู้ใช้งานเน็ตประชารัฐ เนื่องจากในการใช้งานเน็ตประชารัฐจะต้องมีการเข้าสู่ระบบและลงทะเบียนด้วยบัตรประชาชน ในอนาคตจะสามารถเชื่อมโยงข้อมูลเหล่านี้กับระบบสารสนเทศของสำนักทะเบียนราษฎร กรมการปกครองเพื่อเก็บข้อมูลเกี่ยวกับผู้ใช้ได้ 2) ประสิทธิภาพการใช้งานและวัตถุประสงค์การใช้งาน สามารถนำข้อมูลจากสื่อสังคมออนไลน์และบันทึกการใช้งานจากเครื่องแม่ข่ายผู้ให้บริการได้ โดยข้อมูลจากสื่อสังคมออนไลน์จะอยู่ในรูปแบบของความคิดเห็นและความพึงพอใจ ประเด็นที่พบว่าการกล่าวถึงในสื่อสังคมออนไลน์ คือ ความเร็ว ความเสถียร วัตถุประสงค์การใช้งาน ความเหมาะสมของจุดติดตั้ง Wi-Fi ฟรี โดยข้อมูลที่เก็บนั้นจะเป็นข้อมูล 2 ปีย้อนหลังจากเว็บไซต์และสื่อสังคม

ออนไลน์ทั้งหมด 74 แห่ง ข้อมูลเหล่านี้เก็บโดยการใช้โปรแกรม Web Crawler และสามารถแสดงผลในรูปแบบของ Dashboard ที่อัปเดตโดยอัตโนมัติดังตัวอย่างในรูปที่ 6.4 สำหรับข้อมูลบันทึกการใช้งานจากเครื่องแม่ข่าย (Server Log) จะสามารถบ่งบอกได้อย่างชัดเจนว่าประชาชนใช้เน็ตประชารัฐเพื่อทำกิจกรรมใดหรือมีวัตถุประสงค์การใช้งานอย่างไร โดยข้อมูลบันทึกนี้เป็นข้อมูลที่เกิดจากการใช้งานจริง และสามารถถูกจัดกลุ่มวิเคราะห์การใช้งานได้หลากหลาย เช่น ช่วงเวลาที่เข้าใช้เน็ตประชารัฐ หรือพื้นที่ที่เข้าใช้งาน เป็นต้น ดังตัวอย่างในรูปที่ 6.4

สำหรับประเด็น 4) บริการอื่นๆ ที่ประชาชนอยากให้มีผ่านทางเน็ตประชารัฐ เช่น การฝึกอบรมเกี่ยวกับใช้เน็ตสำหรับการค้าขาย 5) ค่าบริการที่เหมาะสมสำหรับการต่อเน็ตประชารัฐเข้าบ้านโดยตรง เป็นข้อมูลที่ไม่พบในสื่อสังคมออนไลน์และบันทึกการใช้งานจากระบบแม่ข่าย จึงอาจยังจำเป็นต้องเก็บข้อมูลดังกล่าวในรูปแบบของแบบสอบถาม แต่อยู่ในรูปแบบของแบบสอบถามออนไลน์ โดยสามารถออกแบบให้ในทุกครั้งที่มีการเชื่อมต่ออินเทอร์เน็ตจากเน็ตประชารัฐ จะมีแบบสอบถามสั้นๆ ให้ผู้ใช้ตอบก่อนการใช้งาน โดยการตอบคำถามสามารถออกแบบการสุ่มคำถาม 4-5 คำถามจาก 20 คำถามที่ต้องการทราบข้อมูล และอาจใช้เทคโนโลยี AI เข้ามาช่วยในการเลือกคำถามที่เหมาะสมกับผู้ตอบ จากการดำเนินการดังกล่าวในเบื้องต้นของการศึกษาพบว่าในข้อมูลที่ได้จากแบบสอบถามด้วยวิธีนี้มีแนวโน้มในทิศทางเดียวกันกับข้อมูลจากแหล่งต่าง ๆ ที่นำมาใช้ในแพลตฟอร์มจากกรณีศึกษาในครั้งนี้



Last Time	Source IP	Destination 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	10.150.173	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 http://mo	Content S Dalvik/2.1.0 (Linux; U	
3/12/2018 18:05	10.150.25	157.240.11	443	graph.face 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	10.150.6.5	45.64.77.1	80	log16.bytc http://log	Content S Dalvik/2.1.0 (Linux; U	
3/12/2018 18:05	10.150.25	103.104.14	443	lan.line.m https://la	Messaging	
3/12/2018 18:05	10.150.55	103.21.25	443	lytimg.co https://l.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	173.194.22	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	10.150.2.6	216.58.20	443	www.go https://w	Internet Services	

ภาพตัวอย่าง บันทึกการใช้งานจากเครื่องแม่ข่ายผู้ให้บริการ

ภาพ Dashboard สรุปการวิเคราะห์อารมณ์/ความรู้สึก (Sentimental analysis) ของข้อความจากภาษาที่เขียน

รูปที่ 6.4 ตัวอย่าง Dashboard และ ไฟล์บันทึกการใช้งาน

6.3 การจัดทำเครื่องมือประเมินช่วยคัดกรองผู้กระทำผิดที่มีโอกาสย้อนกลับไปกระทำผิดซ้ำ

กระทรวงยุติธรรมเห็นความสำคัญของการใช้ประโยชน์ข้อมูลโดยเฉพาะการนำผลของการวิเคราะห์ข้อมูลมาพัฒนางานด้านกระบวนการยุติธรรม โดยในเบื้องต้น กระทรวงยุติธรรมมีความร่วมมือกับกระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคมในการกำหนดโจทย์การวิเคราะห์ข้อมูลขนาดใหญ่ รวมถึงการวิเคราะห์คัดเลือกข้อมูลในเบื้องต้นสำหรับการจัดทำโจทย์ และเตรียมข้อมูลสำหรับการวิเคราะห์ในเชิงลึก ซึ่งผลของการกำหนดโจทย์คือ การวิเคราะห์เชิงพยากรณ์ช่วยคัดกรองผู้กระทำผิดซ้ำ

ในเบื้องต้นได้มีการศึกษาข้อมูลจากฐานข้อมูลหลักของ 3 หน่วยงาน คือ กรมราชทัณฑ์ กรมคุมประพฤติและกรมพินิจและคุ้มครองเด็กและเยาวชน โดยทำ Exploratory Data Analysis และสร้างเป็น

Dashboard เพื่อการกระจายตัวของข้อมูลผู้กระทำผิดซ้ำ จากนั้นกระทรวงยุติธรรมได้ทำการรวบรวมข้อมูลจากเครื่องมือที่ใช้ในการปฏิบัติงานเพิ่มเติมเพื่อเสริมกับข้อมูลจากฐานข้อมูลหลัก ตัวอย่างเช่น

- แบบสอบถามเพื่อรวบรวมข้อเท็จจริง จัดทำประวัติ เพื่อจำแนกระดับความเสี่ยงต่อการกระทำผิดซ้ำ โดยแบบสอบถามนี้ถูกใช้เก็บข้อมูลในขั้นตอนการสืบเสาะ/พินิจ การจำแนก/วิเคราะห์เพื่อพัฒนาพินิจ และการกลับเข้าสู่สังคม
- แบบสำรวจสถานะจิตใจ ร่างกาย และพฤติกรรม ซึ่งถูกใช้ในขั้นตอนการสืบเสาะ/พินิจ
- แบบสัมภาษณ์เพื่อติดตามการพัฒนาสถานะ ซึ่งถูกใช้เก็บข้อมูลด้านการพัฒนาสถานะในขั้นตอนการจำแนก/วิเคราะห์เพื่อพัฒนาพินิจเป็นระยะหลังการพิจารณาคดี
- แบบสอบถามข้อมูลของผู้กระทำผิด เช่น ประวัติการกระทำผิด ประวัติการใช้สารเสพติด และการบำบัดรักษา ภูมิหลังทางครอบครัว ประวัติการศึกษาและอาชีพ ประวัติการใช้ความรุนแรง สถานะทางจิตและพฤติกรรม ชุมชนและสภาพแวดล้อม การคบเพื่อน ฯลฯ โดยข้อมูลในส่วนนี้สามารถใช้งานร่วมกันระหว่างหน่วยงานที่เกี่ยวข้องได้เพื่อระบุกรณีที่ผู้กระทำผิดย้อนกลับเข้ามาในระบบหลังการปล่อยตัว

หน่วยงานในกลุ่มพินิจของกระทรวงยุติธรรมได้วางแผนที่จะนำข้อมูลทั้งหมดที่ได้จัดเตรียมไว้เพื่อจัดทำโมเดลพยากรณ์ 2 ส่วนด้วยกัน คือ

ส่วนที่ 1 เครื่องมือประเมินความเสี่ยงในการช่วยคัดกรองกลุ่มผู้มีโอกาสย้อนกลับไปกระทำผิด

เป็นการสร้างโมเดลสำหรับการวิเคราะห์เชิงทำนาย เพื่อใช้ในการคัดกรองผู้กระทำผิดที่มีโอกาสกระทำผิดซ้ำ หนีการคุมประพฤติ หรือฝ่าฝืนกฎการภาคทัณฑ์ โดยในเบื้องต้นใช้ทฤษฎีทางด้านสถิติประยุกต์ใช้เพื่อสร้างโมเดลพื้นฐานที่สามารถคัดกรองแบ่งกลุ่มผู้กระทำผิด ออกเป็น 3 กลุ่มคือ กลุ่มความเสี่ยงต่ำ ความเสี่ยงปานกลาง และความเสี่ยงสูง (ในลักษณะเดียวกับการทำ Credit Scoring ของธนาคาร) ซึ่งการแบ่งกลุ่มในลักษณะนี้จะสามารถช่วยในการออกแบบวิธีการคุมประพฤติที่แตกต่างกันสำหรับนักโทษแต่ละกลุ่มเสี่ยง รวมถึงเป็นการเพิ่มประสิทธิภาพการดูแล คุมประพฤติ และฟื้นฟู ผู้กระทำผิดได้

โดยตั้งแต่ปี ค.ศ. 1970 เป็นต้นมานักสังคมศาสตร์ได้เริ่มศึกษาและสร้างเครื่องมือวิจัยเพื่อประเมินความเสี่ยงการกระทำผิดซ้ำ เครื่องมือเหล่านั้นได้ถูกนำมาประยุกต์ใช้อย่างเป็นรูปธรรมในประเทศแคนาดาดังนำเสนอในบทความเรื่อง “Risk-Need-Responsivity Model for Offender Assessment and Rehabilitation” โดย Andrew and Bonta ซึ่งจุดหลักที่เพิ่มประสิทธิผลในการทำนายคือการดึงเอา Dynamic Risk Factor เข้ามาใช้ประกอบการสร้างโมเดล

ข้อมูลที่ใช้ในการสร้างโมเดลเชิงคณิตศาสตร์จะแบ่งออกได้เป็น 3 กลุ่มคือ

- Static Factors เช่น ประเภทการกระทำผิด อายุปัจจุบัน อายุเมื่อถูกจับครั้งแรก ประวัติการมีงานทำ ประวัติการกระทำผิด ฯลฯ
- Dynamic Factors เช่น เจตคติ (Attitude) กลุ่มเพื่อนที่คบ สารเสพติดที่ใช้ บุคลิกในมิติด้านการต่อต้านสังคม ฯลฯ ซึ่งเป็นข้อมูลที่ต้องเก็บเพิ่มเติมและประเมินคะแนนโดยผู้เชี่ยวชาญ

- Real-time Data เช่น พฤติกรรม และสถานที่ที่มักสูม ในเวลาปัจจุบัน

ในปัจจุบันกลุ่มงานด้านพฤติกรรมศาสตร์มีข้อมูลบางส่วนใน 2 หมวดแรกแล้ว โดยมีความน่าจะเป็นที่สามารถทดสอบการสร้างโมเดลคณิตศาสตร์เพื่อนำร่องได้ จากนั้นค่อยเริ่มประสานงานกับหน่วยงานอื่นในกระบวนการยุติธรรมเพื่อรวบรวมข้อมูลในหมวดที่ 3

ส่วนที่ 2 การใช้ระบบ Recommender เพื่อแนะนำโครงการฟื้นฟูพฤติกรรมที่เหมาะสมสำหรับผู้กระทำผิดส่วนบุคคล

สร้างโมเดลสำหรับการวิเคราะห์เชิงทำนายเพื่อใช้ในการคัดเลือกวิธีการที่เหมาะสมเพื่อแทรกแซง (Intervention) แนวโน้มการกระทำผิดซ้ำ โดยพิจารณาลักษณะและความต้องการของแต่ละบุคคล ในเบื้องต้น ทฤษฎีทางด้านสถิติและความน่าจะเป็น จะถูกประยุกต์ใช้เพื่อสร้างระบบ Recommender ในลักษณะเดียวกับระบบที่ใช้กันอย่างแพร่หลายในภาคธุรกิจ ระบบ Recommender จะถูกสร้างและฝึก (Build and Train) จากพฤติกรรมหลังจากร่วมโครงการฟื้นฟูของผู้กระทำผิดในอดีต โดยจัดกลุ่มตามลักษณะนิสัยและเจตคติ จากนั้นเมื่อมีผู้กระทำผิดรายใหม่เข้ามาในระบบ โมเดล Recommender จะสามารถเลือกโครงการฟื้นฟูที่เหมาะสมจากข้อมูล รูปแบบ (Pattern) ที่มีอยู่โดยดูจากความคล้ายคลึงได้ ข้อมูลที่ใช้ในการสร้างระบบ recommender จะรวมถึง ประวัติการกระทำผิด สถานะแวดล้อมที่อยู่อาศัย การเข้าร่วมหรือความเกี่ยวข้อง กับกลุ่มมีงานอาชีพ กลุ่มเพื่อน และ ข้อมูลความเสี่ยงในการกระทำผิดซ้ำ (Credit Score) ฯลฯ

6.4 โครงการวิเคราะห์ข้อมูลค่าเบี่ยงเบนสิทธิการรักษาพยาบาลจากรัฐบาล

กระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคมได้ร่วมเจรจากับสำนักงานกลางสารสนเทศบริการสุขภาพ (สทส.) ในการวิเคราะห์ข้อมูลค่าเบี่ยงเบนสิทธิการรักษาพยาบาลจากรัฐบาล เพื่อให้กระทรวงสาธารณสุขสามารถนำผลการวิเคราะห์ที่ได้ไปช่วยในการวางแผนงานและดำเนินการในส่วนที่เกี่ยวข้องได้อย่างมีประสิทธิภาพ โดยการวิเคราะห์เป็นการนำศาสตร์ทางด้าน AI มาใช้เพื่อหาโมเดลที่เหมาะสมเพื่อให้หน่วยงานที่เกี่ยวข้องกับระบบประกันสุขภาพ เช่น สทส. สำนักงานหลักประกันสุขภาพแห่งชาติ (สปสช.) กรมบัญชีกลาง และสำนักงานประกันสังคม สามารถนำโมเดลดังกล่าวไปใช้พัฒนางานในมิติต่างๆ ได้ เช่น ใช้ในการประเมินงบประมาณกองทุนและการเบิกจ่ายตามพื้นที่ ใช้เพื่อทำนายหรือพยากรณ์โรคที่อาจจะเกิดขึ้นของผู้ป่วยเดิม ใช้เพื่อการจัดสรรทรัพยากรบุคคล สถานที่ และรถพยาบาลได้อย่างเหมาะสมมีประสิทธิภาพ เป็นต้น โดยปัจจุบันกระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคมอยู่ระหว่างการดำเนินการวิเคราะห์สร้างโมเดลข้อมูล เพื่อจัดกลุ่มข้อมูลค้นหาลักษณะความสัมพันธ์ของข้อมูลที่แอบแฝงอยู่และพัฒนาให้เป็นแบบจำลองข้อมูลเพื่อการพยากรณ์ โดยโมเดลที่ได้จะสามารถผลักดันให้ระบบประกันสุขภาพของประเทศไทยมีประสิทธิภาพมากขึ้น และสามารถไปถึงหรือเหนือกว่าระบบประกันสุขภาพในมาตรฐานสากล

ปัจจุบันคนไทยได้รับการคุ้มครองสิทธิการรักษาพยาบาลจากรัฐบาล โดยสิทธิการรักษาพยาบาลที่มีอยู่ 3 ระบบ คือ 1) สิทธิสวัสดิการการรักษาพยาบาลของข้าราชการ 2) สิทธิประกันสังคม และ 3) สิทธิหลักประกันสุขภาพ 30 บาท โดยองค์กรที่กำกับดูแลแต่ละระบบมีหน้าที่ดูแลจัดเก็บข้อมูลการเบิกจ่ายค่ารักษาพยาบาล ได้แก่ 1) กรมบัญชีกลางและกระทรวงการคลังดูแลค่าเบิกจ่ายสิทธิสวัสดิการการรักษาพยาบาล

ของข้าราชการ 2) สำนักงานประกันสังคมดูแลการเบิกจ่ายสิทธิประกันสังคม และ 3) สปสช. ดูแลสิทธิประกันสุขภาพ โดยการจัดเก็บข้อมูลผู้ใช้สิทธิทั้ง 3 สิทธิ ซึ่งสำนักงานกลางสารสนเทศบริการสุขภาพเป็นหน่วยงานกลางที่รวบรวมข้อมูลผู้ใช้สิทธิทั้ง 3 ประเภท แต่ในปัจจุบัน สกส. มีความร่วมมือกับกรมบัญชีกลาง และสำนักงานประกันสังคมในการพัฒนาระบบจัดเก็บข้อมูลและประมวลผลทางอิเล็กทรอนิกส์ ซึ่งในโครงการจะใช้ข้อมูลที่ได้จาก สกส. (กรมบัญชีกลาง และสำนักงานประกันสังคม) มาใช้ในการวิเคราะห์เป็นหลัก นอกจากนั้นจะขอความร่วมมือขอใช้ข้อมูลจาก สปสช. เพื่อเพิ่มเติมข้อมูลบางส่วนมาร่วมในการวิเคราะห์ในครั้งนี้ด้วย โดยข้อมูลที่น่ามาใช้คือ ข้อมูลการเบิกจ่ายค่ารักษาพยาบาลของข้าราชการ ข้อมูลแจ้งค่ารักษาพยาบาลของประกันสังคม ข้อมูลรายละเอียดธุรกรรมค่ารักษาพยาบาล ข้อมูลตรวจรักษา ข้อมูลการทำหัตถการ ข้อมูลการวินิจฉัย เป็นต้น

6.5 การศึกษาปัจจัยเพื่อทำนายระดับความสามารถของบิดามารดาในการดูแลบุตรวัย 0-3 ขวบ

กรณีศึกษาที่เกิดจากความร่วมมือของกระทรวงต่างๆ ทั้งหมด 5 กระทรวง คือ กระทรวงสาธารณสุข กระทรวงการพัฒนาสังคมและความมั่นคงของมนุษย์ กระทรวงศึกษาธิการ กระทรวงการคลัง และกระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม เพื่อตอบโจทย์การทำนายระดับความสามารถของบิดามารดาในการดูแลบุตรวัย 0-3 ปี โดยการบูรณาการข้อมูล

โดยนำข้อมูลของเด็กจากกระทรวงสาธารณสุข เช่น การได้รับวิตามิน การได้รับการตรวจไทรอยด์ การเข้ารับการตรวจครรภ์ ความถี่ในการพบแพทย์ ระดับพัฒนาการ ฯลฯ เป็นต้น มาทำการวิเคราะห์ด้วยเทคนิค Cluster Analysis เพื่อจัดกลุ่มเด็กตามคุณภาพการเลี้ยงดู โดยจัดกลุ่มได้เป็น 3 กลุ่ม คือ กลุ่มที่ได้รับการเลี้ยงดูในระดับดี กลุ่มที่ได้รับการเลี้ยงดูระดับปานกลาง และกลุ่มที่ได้รับการเลี้ยงดูระดับที่ควรได้รับคำแนะนำ จากนั้นนำข้อมูลของเด็กจับคู่กับข้อมูลของบิดามารดา โดยใช้ข้อมูลที่ได้จากการบูรณาการของ 5 กระทรวง เช่น ข้อมูลสุขภาพพ่อและแม่ ข้อมูลด้านสถานะทางเศรษฐกิจ การรับเงินกู้ ระดับการศึกษา เป็นต้น เพื่อทำการวิเคราะห์หาสาเหตุรากเหง้า (Root Cause Analysis) โดยใช้เทคโนโลยีปัญญาประดิษฐ์ (Artificial Intelligence: AI) สร้างสูตรสำหรับคัดกรองพ่อและแม่ที่มีโอกาสเสี่ยงที่จะไม่สามารถเลี้ยงดูบุตรให้มีคุณภาพที่ดีได้ ซึ่งประโยชน์ที่ได้รับคือสามารถทำนายปัญหาการเลี้ยงดูของพ่อและแม่รายใหม่ได้ตั้งแต่วันคลอด เพื่อให้หน่วยงานภาครัฐที่เกี่ยวข้องสามารถวางแผนเตรียมความพร้อมในการให้ความช่วยเหลือกับพ่อและแม่ เพื่อให้สามารถเลี้ยงดูเด็กได้อย่างมีคุณภาพ และพร้อมสำหรับก้าวเข้าสู่เยาวชนที่ดีในอนาคตได้

6.6 การวิเคราะห์เสียงสะท้อนจากนักท่องเที่ยว

กรณีศึกษาโดยการรวบรวมข้อมูลจากสื่อสังคมออนไลน์ เพื่อใช้นำมาใช้วิเคราะห์ความคิดเห็นของนักท่องเที่ยว ในกรณีศึกษานี้ใช้ข้อมูลที่รวบรวมจาก TripAdvisor.com ซึ่งเป็นเว็บไซต์ด้านการท่องเที่ยวอันดับ 1 ของโลก โดยเน้นความคิดเห็นของนักท่องเที่ยวที่มีต่อการท่องเที่ยวในจังหวัดเชียงราย ซึ่งเป็นจังหวัดที่มีผู้ให้ความคิดเห็นใน TripAdvisor.com มากที่สุด โดยมีการนำข้อมูล ความคิดเห็นนักท่องเที่ยวมาทำการวิเคราะห์ด้วยเทคนิค Descriptive Statistics, Spatial Analytics และ Sentiment Analytics

ประโยชน์ที่ได้รับจากการวิเคราะห์ คือ ทำให้ทราบแหล่งท่องเที่ยวที่นักท่องเที่ยวสนใจ สามารถนำไปใช้ในการวางแผนจัดการการท่องเที่ยว การบริหารจัดการสถานที่ท่องเที่ยวได้อย่างเหมาะสมทั้งสถานที่ท่องเที่ยวใหม่ที่กำลังได้รับความนิยม หรือสถานที่ท่องเที่ยวเดิมที่มีปริมาณนักท่องเที่ยวหนาแน่น

6.7 การเพิ่มศักยภาพด้านการจัดเก็บภาษี

เป็นการดำเนินการร่วมกับกรมสรรพากรในการจัดลำดับกรณีการตรวจสอบการจ่ายภาษี ซึ่งโจทย์ดังกล่าวเป็นการวิเคราะห์ข้อมูลโดยใช้เทคโนโลยีด้านการเรียนรู้ของเครื่อง (Machine learning) เพื่อสร้างแบบจำลองในการพยากรณ์รายรับของธุรกิจในอนาคตของนักท่องเที่ยว และพัฒนาระดับการจ่ายภาษีมาตรฐาน (benchmarks) ตามประเภทธุรกิจ จากนั้นแบบจำลองทางคณิตศาสตร์ที่ได้จากการเรียนรู้ของเครื่องจะถูกใช้ในการทำนายรายรับเพื่อนำมาเปรียบเทียบกับรายรับที่ธุรกิจได้รายงานแก่กรมสรรพากร หากธุรกิจใดมีโอกาสสูงที่จะมีรายจ่ายรายรับต่ำกว่าความเป็นจริงมาก โดยพิจารณาจากธุรกิจที่เป็นค่าผิดปกติ (Outlier) จากระดับมาตรฐานข้างต้น ระบบจะสามารถแจ้งเตือนเจ้าหน้าที่สรรพากรได้ ข้อมูลที่สามารถนำมาใช้ในการวิเคราะห์ เช่น ข้อมูลภาษีเงินได้ ภาษีการขาย ข้อมูลรายได้ รายจ่าย กำไรสุทธิของธุรกิจในอุตสาหกรรมท่องเที่ยว ข้อมูลโรงแรม ที่พัก ร้านอาหาร ข้อมูลทั่วไปเกี่ยวกับการดำเนินการของธุรกิจในอุตสาหกรรมท่องเที่ยว เป็นต้น

6.8 การวิเคราะห์รูปแบบภูมิอากาศเพื่อการวางแผนของท่าอากาศยาน

กรมอุตุนิยมวิทยาภายใต้กระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม จัดทำกรณีศึกษาการบูรณาการข้อมูลที่เกี่ยวข้องกับสภาพอากาศที่มีอยู่ ได้แก่ 1) ข้อมูลอุณหภูมิ/ความชื้นสัมพัทธ์/ทิศทางลม/ความกดอากาศ 2) ข้อมูลสภาพทัศนวิสัย 3) ข้อมูลปริมาณเมฆ 4) ข้อมูลลักษณะอากาศปัจจุบัน 5) มาทำการวิเคราะห์ด้วยเทคนิค Cluster Analysis อุตุนิยมวิทยา ในพื้นที่นาร่อง จ.สุโขทัย เพื่อหารูปแบบการจัดกลุ่มของภูมิอากาศ ซึ่งทำให้พบกับการเปลี่ยนแปลงของสภาพภูมิอากาศ ซึ่งประโยชน์ที่ได้รับคือ สามารถใช้ในการวางแผนด้านการบริหารความเสี่ยงเรื่องการบิน และใช้ในการจัดทำแผนส่งเสริมการท่องเที่ยวให้เหมาะสมกับสภาพอากาศที่เปลี่ยนแปลงได้

6.9 การจัดทำแพลตฟอร์มการรวบรวมข้อมูลเพื่อการวิเคราะห์ “Digital Transformation ตามมาตรฐาน” OECD (Digital Economy Outlook)

เนื่องจากประเทศไทยมีความร่วมมือกับองค์การเพื่อความร่วมมือทางเศรษฐกิจและการพัฒนา (Organization for Economic Co-operation and Development: OECD) ในประเด็นการขับเคลื่อนการพัฒนาและปฏิรูปประเทศอย่างมีประสิทธิภาพและยั่งยืน เพื่อสอดคล้องกับเป้าหมายการพัฒนาของประเทศ โดยเฉพาะอย่างยิ่งด้านการพัฒนาเศรษฐกิจดิจิทัล ในการดำเนินการตามกรอบความร่วมมือดังกล่าวประเทศไทยโดยสำนักงานคณะกรรมการดิจิทัลเพื่อเศรษฐกิจและสังคมแห่งชาติ จะดำเนินกิจกรรมความร่วมมือเพื่อศึกษาบริบทการพัฒนาเศรษฐกิจดิจิทัลของประเทศ (Thailand Digital Economy Outlook) ใน 6 มิติหลัก คือ การเข้าถึง การใช้นวัตกรรม การทำงาน สังคม ความเชื่อมั่น และการเปิดตลาด โดยผลการวิเคราะห์สามารถนำมาใช้เป็นข้อมูลสนับสนุนการวางแผนการพัฒนาดิจิทัลเพื่อเศรษฐกิจและสังคมในระดับมหภาค

ต่อไปได้ ซึ่งในการดำเนินการดังกล่าวจะต้องทำการรวบรวมข้อมูลจากแหล่งต่างๆ รวมทั้งการวิเคราะห์และนำเสนอข้อมูลปรากฏการณ์ Digital Transformation ให้ครอบคลุมทั้ง 6 มิติ

การรวบรวมข้อมูลจากหลายแหล่งเพื่อให้ครอบคลุมทุกมิตินั้น อาจมีความหลากหลายของแหล่งที่มาและความหลากหลายของรูปแบบการจัดเก็บข้อมูล ดังนั้นในการศึกษานี้จึงเป็นการสร้างแพลตฟอร์มในการรวบรวมข้อมูลจากแหล่งต่างๆ การจัดเก็บข้อมูล การวิเคราะห์ รวมไปถึงการนำเสนอข้อมูลในรูปแบบ Dashboard ซึ่งการทำแพลตฟอร์มดังกล่าว มีแนวโน้มที่จะพัฒนาต้นแบบให้กับจังหวัดราชบุรี และหากกรณีศึกษาประสบความสำเร็จกระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคมจะสามารถนำแพลตฟอร์มดังกล่าวนี้ไปประยุกต์ใช้ในพื้นที่อื่นต่อไปได้ เพื่อให้วิธีการรวบรวมข้อมูลปรากฏการณ์ Digital Transformation ที่เป็นไปในทิศทางเดียวกันทั่วประเทศ

เอกสารแนบ

เอกสารแนบ ก1 สรุปรายชื่อหน่วยงานระดับกรมหรือเทียบเท่า ที่มี Agency Data Center / Cloud

กระทรวงกลาโหม ● กองทัพบก¹ ● กองทัพเรือ¹ ● กองทัพอากาศ¹	กระทรวงการคลัง ● กรมบัญชีกลาง ● กรมศุลกากร ● กรมสรรพสามิต ● กรมสรรพากร ● สำนักงานเศรษฐกิจการคลัง
กระทรวงการท่องเที่ยวและกีฬา ● กรมท่องเที่ยว ● สถาบันการพลศึกษา	กระทรวงการพัฒนาสังคมและความมั่นคงของมนุษย์ ● กรมกิจการเด็กและเยาวชน ● กรมกิจการผู้สูงอายุ ● กรมส่งเสริมและพัฒนาคุณภาพชีวิตคนพิการ ● การเคหะแห่งชาติ ● สำนักงานธนาคุณเคราะห์
กระทรวงเกษตรและสหกรณ์ ● องค์การสะพานปลา	กระทรวงคมนาคม ● กรมการขนส่งทางบก ● กรมทางหลวง ● กรมทางหลวงชนบท ● กรมท่าอากาศยาน ● การทางพิเศษแห่งประเทศไทย ● การท่าเรือแห่งประเทศไทย ● การรถไฟแห่งประเทศไทย ● บริษัท ขนส่ง จำกัด ● บริษัท ท่าอากาศยานไทย จำกัด (มหาชน) ● บริษัท ไทยสมายล์แอร์เวย์ จำกัด ● บริษัท วิทยุการบินแห่งประเทศไทย จำกัด ● สถาบันการบินพลเรือน

¹ กองทัพบก กองทัพเรือ และกองทัพอากาศ เป็นหน่วยงานที่ขึ้นตรงต่อกองทัพบกไทย (ซึ่งเป็นหน่วยงานระดับกรม) ที่วางแผนจะจัดทำ data center ของหน่วยงานเอง

กระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม ● กรมอุตุนิยมวิทยา ● บริษัท ไปรษณีย์ไทย จำกัด ● สำนักงานพัฒนาธุรกรรมทางอิเล็กทรอนิกส์ (องค์การมหาชน) ● สำนักงานส่งเสริมเศรษฐกิจดิจิทัล ● สำนักงานสถิติแห่งชาติ	กระทรวงทรัพยากรธรรมชาติและสิ่งแวดล้อม ● กรมควบคุมมลพิษ ● กรมทรัพยากรน้ำ
---	---

กระทรวงพลังงาน ● กรมธุรกิจพลังงาน ● กรมพัฒนาพลังงานทดแทนและอนุรักษ์พลังงาน ● การไฟฟ้าฝ่ายผลิตแห่งประเทศไทย ● บริษัท ปตท. จำกัด (มหาชน) ● สำนักงานคณะกรรมการกำกับกิจการพลังงาน	กระทรวงพาณิชย์ ● กรมการค้าต่างประเทศ ● กรมการค้าภายใน ● กรมส่งเสริมการค้าระหว่างประเทศ
กระทรวงมหาดไทย ● กรมที่ดิน ● การประปานครหลวง ● การประปาส่วนภูมิภาค ● การไฟฟ้านครหลวง ● การไฟฟ้าส่วนภูมิภาค ● สำนักบริหารการทะเบียน กรมการปกครอง	กระทรวงยุติธรรม ● กรมสอบสวนคดีพิเศษ ● สำนักงานคณะกรรมการป้องกันและปราบปรามยาเสพติด
กระทรวงแรงงาน ● กรมการจัดหางาน	กระทรวงวิทยาศาสตร์และเทคโนโลยี ● ศูนย์ความเป็นเลิศด้านชีววิทยาศาสตร์ ● สถาบันวิจัยวิทยาศาสตร์และเทคโนโลยีแห่งประเทศไทย ● สถาบันวิจัยแสงซินโครตรอน (องค์การมหาชน) ● สถาบันสารสนเทศทรัพยากรน้ำและการเกษตร (องค์การมหาชน) ● สำนักงานพัฒนาเทคโนโลยีอวกาศและภูมิสารสนเทศ (องค์การมหาชน) ● สำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ

กระทรวงสาธารณสุข ● กรมควบคุมโรค ● กรมสนับสนุนบริการสุขภาพ ● กรมอนามัย ● สำนักงานคณะกรรมการอาหารและยา	กระทรวงอุตสาหกรรม ● กรมโรงงานอุตสาหกรรม ● กรมส่งเสริมอุตสาหกรรม
สำนักนายกรัฐมนตรี ● บริษัท อสมท. จำกัด (มหาชน) ● สำนักงบประมาณ ● สำนักงานคณะกรรมการพัฒนาการเศรษฐกิจและสังคมแห่งชาติ ● สำนักงานพัฒนารัฐบาลดิจิทัล (องค์การมหาชน)	หน่วยงาน/องค์กรอิสระ และอื่นๆ ● ธนาคารแห่งประเทศไทย ● สำนักงานการตรวจเงินแผ่นดิน ● สำนักงานคณะกรรมการป้องกันและปราบปรามการทุจริตแห่งชาติ ● สำนักงานคณะกรรมการป้องกันและปราบปรามการฟอกเงิน ● สำนักงานตำรวจแห่งชาติ ● สำนักงานศาลยุติธรรม

เอกสารแนบ ก2 สรุปจำนวนหน่วยงานที่จะร่วมใช้คลาวด์ระดับกระทรวง (Ministry Cloud)

สังกัดหน่วยงาน	จำนวน หน่วยงานที่จะ ร่วมใช้คลาวด์ ระดับกระทรวง	จำนวน หน่วยงานใน สังกัด	เปอร์เซ็นต์ของ หน่วยงานที่จะร่วมใช้ คลาวด์ระดับกระทรวง
กระทรวงกลาโหม ²	4	8	50%
กระทรวงการคลัง ³	14	19	74%
กระทรวงการต่างประเทศ ⁴	13	13	100%
กระทรวงการท่องเที่ยวและกีฬา	5	7	71%
กระทรวงการพัฒนาสังคมและความมั่นคงของมนุษย์	4	9	44%
กระทรวงเกษตรและสหกรณ์	21	22	95%
กระทรวงคมนาคม	10	22	45%
กระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม	4	9	44%
กระทรวงทรัพยากรธรรมชาติและสิ่งแวดล้อม	14	16	88%
กระทรวงพลังงาน	4	9	44%
กระทรวงพาณิชย์	10	13	77%
กระทรวงมหาดไทย	6	12	50%
กระทรวงยุติธรรม	14	16	88%
กระทรวงแรงงาน	5	6	83%
กระทรวงวัฒนธรรม ⁵	9	9	100%
กระทรวงวิทยาศาสตร์และเทคโนโลยี	9	15	60%
กระทรวงศึกษาธิการ ⁶	12	12	100%
กระทรวงสาธารณสุข	7	11	64%
กระทรวงอุตสาหกรรม	17	19	89%
สำนักนายกรัฐมนตรี ⁷	3	26	85%
รวม	185¹	273¹	75%

ข้อสังเกต:

1. สถานะปัจจุบัน ได้มีการสำรวจและวิเคราะห์ข้อมูลหน่วยงานแล้ว 159 หน่วยงานจากหน่วยงานภาครัฐต่างๆ (รวมหน่วยงานอิสระ) ซึ่งได้รวมถึงสำนักงานปลัดจากทุกกระทรวง ผู้เชี่ยวชาญในคณะอนุกรรมการออกแบบสถาปัตยกรรมระบบบูรณาการข้อมูลภาครัฐ ได้อนุมานหน่วยงานบางส่วนว่าจะร่วมใช้คลาวด์ระดับกระทรวง

ซึ่งหน่วยงานเหล่านี้สามารถให้ข้อมูลยืนยันในภายหลัง สำหรับจำนวนหน่วยงานทั้งหมดใน 20 สังกัด ได้มีการรวบรวมจากเว็บไซต์ของแต่ละสังกัดโดยตรงเพื่อความครบถ้วนและทันสมัย

2. จำนวนหน่วยงานในสังกัดกระทรวงกลาโหม นับกองทัพบก กองทัพเรือ และกองทัพอากาศ เสมือนเป็นกรมเพื่อจุดประสงค์ในการนับจำนวน Agency Data center / Cloud เนื่องจากเหล่าทัพมีข้อมูลชั้นความลับสูง จึงประสงค์จะดูแลดาต้าเซ็นเตอร์เอง

3. จำนวนหน่วยงานในสังกัดกระทรวงการคลัง ไม่รวมการพิจารณาธนาคารต่างๆ ในสังกัดกระทรวง

4. กระทรวงต่างประเทศ มีแผนจะใช้ระบบคลาวด์ผสม มีดาต้าเซ็นเตอร์และระบบคลาวด์ โดยสำนักงานปลัดจะให้บริการทุกหน่วยงานในสังกัด

5. กระทรวงวัฒนธรรม มีแผนจะใช้ระบบคลาวด์กลางภาครัฐ

6. กระทรวงศึกษาธิการ มีแผนจะสร้างระบบคลาวด์ผสม มีดาต้าเซ็นเตอร์และระบบคลาวด์ส่วนตัวและมีบางส่วนเชื่อมต่อไปยังคลาวด์สาธารณะ โดยส่วนกลางจะให้บริการทุกหน่วยงานในสังกัด

7. สำนักงานนายกรัฐมนตรี วางแผนจะจัดทำคลาวด์ระดับกระทรวงสำหรับส่วนราชการที่ขึ้นตรงต่อปลัดสำนักนายกรัฐมนตรีเท่านั้น ประกอบด้วย 3 หน่วยงาน คือ สำนักงานปลัดสำนักนายกรัฐมนตรี กรมประชาสัมพันธ์ และสำนักงานคณะกรรมการคุ้มครองผู้บริโภค

เอกสารแนบ ข แนวทางปฏิบัติทางเทคโนโลยีสำหรับการแปลงข้อมูลให้อยู่ในรูปแบบพร้อมใช้งาน

เอกสารแนบฉบับนี้อธิบายแนวทางปฏิบัติในการแปลงข้อมูลในรูปแบบ 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) ให้อยู่ในรูปแบบ CSV ได้เป็นอย่างดี 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 : เหมาะสำหรับข้อมูลที่ต้องการความปลอดภัยสูง มีขั้นตอนที่เยอะกว่าการใช้เครื่องมือออนไลน์ เครื่องมือที่เอกสารฉบับนี้แนะนำคือ Openrefine ซึ่งเป็นโปรแกรม Open Source ที่ได้รับการยอมรับจาก European Open Data Portal <https://www.europeandataportal.eu> ผู้ใช้สามารถดาวน์โหลดซอฟต์แวร์ Openrefine สำหรับระบบปฏิบัติการต่างๆได้จาก <http://openrefine.org/download.html> จากนั้นทำการติดตั้งตามขั้นตอนที่ระบุในเว็บไซต์ดังกล่าว

ข.1. วิธีการแปลงข้อมูลจาก CSV เป็น JSON

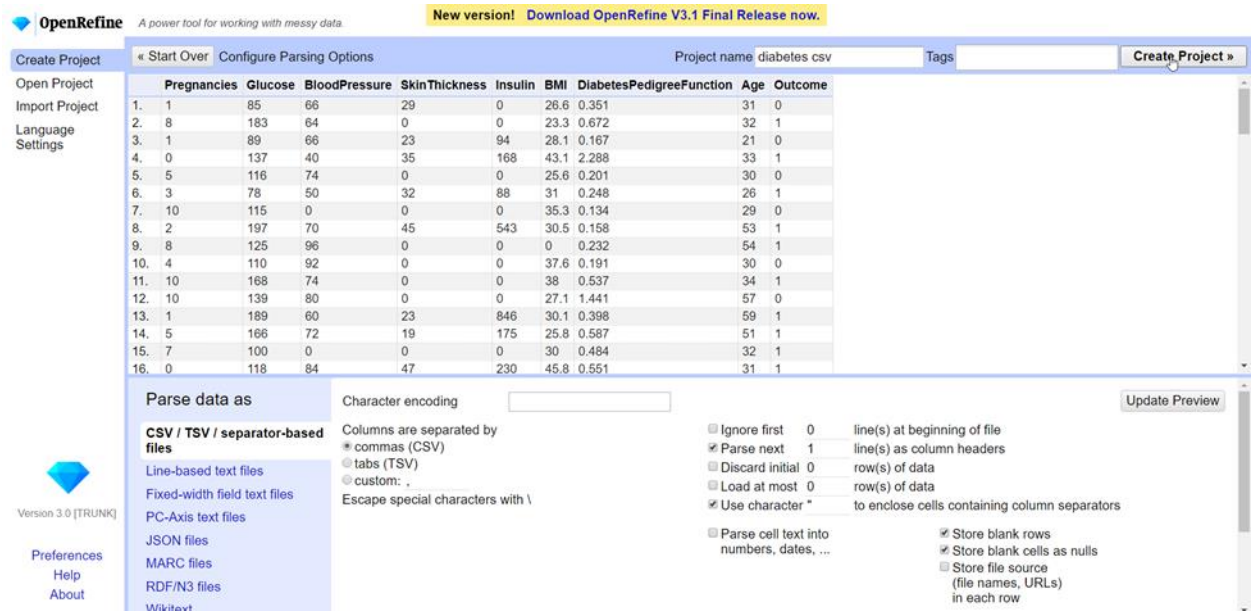
ข.1.1 การแปลงข้อมูลเครื่องมือออนไลน์ สามารถเลือกได้ 2 วิธีดังนี้

- 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”

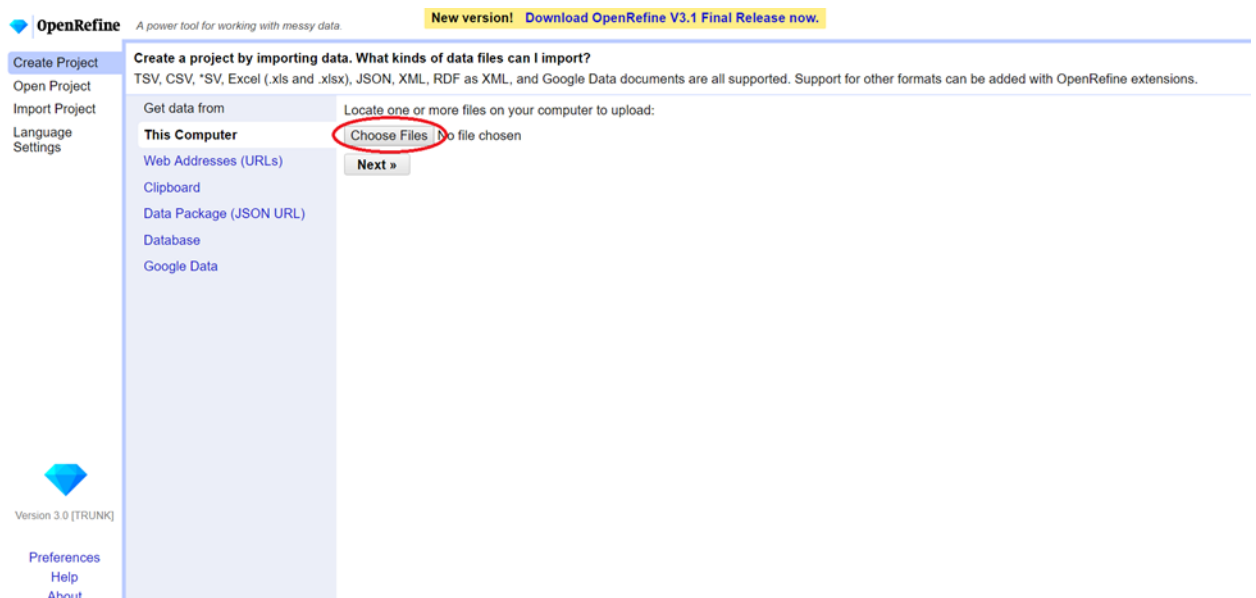
ในกรณีที่ไฟล์อยู่บนคลาวด์ จากนั้นคลิกปุ่ม “CSV to JSON” ระบบจะแปลงไฟล์ในรูปแบบ CSV ไปยัง JSON โดยอัตโนมัติ จากนั้นผู้ใช้งานคลิก “Download” เพื่อเก็บไฟล์ JSON ไปใช้ในขั้นตอนต่อไป

ข.1.2 การแปลงข้อมูลด้วยเครื่องมือที่ติดตั้งบนเครื่อง สามารถใช้ Openrefine 3.0 ตามขั้นตอนดังนี้

- ดับเบิลคลิกเพื่อเปิด Openrefine



- เลือกไฟล์ CSV ที่ต้องการแปลง



- 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

CSV / TSV / separator-based files

Line-based text files
Fixed-width field text files
PC-Axis text files
JSON files
MARC files
RDF/N3 files
Wikitext

Character encoding:

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

- คลิก Templating

OpenRefine diabetes.csv Permalink Open... Export Help

Facet / Filter Undo / Redo 0 / 0 **767 rows**

Show as: rows records Show: 5 10 25 50 rows

Using facets and filters
 Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.
 Not sure how to get started? [Watch these screencasts](#)

All	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedi	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

Export project
 Project data package
 Tab-separated value
 Comma-separated value
 HTML table
 Excel (.xls)
 Excel 2007+ (.xlsx)
 ODF spreadsheet
 Custom tabular exporter...
 SQL Exporter...
Templating...
 QuickStatements

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

Templating Export

Prefix
[

Row Template

```
{
  "Pregnancies" : {{jsonize(cells["Pregnancies"].value)
  "Glucose" : {{jsonize(cells["Glucose"].value)}}
  "BloodPressure" : {{jsonize(cells["BloodPressure"].v
  "SkinThickness" : {{jsonize(cells["SkinThickness"].v
  "Insulin" : {{jsonize(cells["Insulin"].value)}}
  "BMI" : {{jsonize(cells["BMI"].value)}}
  "DiabetesPedigreeFunction" : {{jsonize(cells["Diabet
  "Age" : {{jsonize(cells["Age"].value)}}
  "Outcome" : {{jsonize(cells["Outcome"].value)}}
}
```

Row Separator
,

Suffix
]

Reset Template Export Cancel

Preview:

```
{
  {
    "Pregnancies" : "1",
    "Glucose" : "85",
    "BloodPressure" : "66",
    "SkinThickness" : "29",
    "Insulin" : "0",
    "BMI" : "26.6",
    "DiabetesPedigreeFunction" : "0.351",
    "Age" : "31",
    "Outcome" : "0"
  },
  {
    "Pregnancies" : "8",
    "Glucose" : "183",
    "BloodPressure" : "64",
    "SkinThickness" : "0",
    "Insulin" : "0",
    "BMI" : "23.3",
    "DiabetesPedigreeFunction" : "0.672",
    "Age" : "32",
    "Outcome" : "1"
  },
  {
    "Pregnancies" : "1",
    "Glucose" : "89",
    "BloodPressure" : "66",
    "SkinThickness" : "23",
    "Insulin" : "94",
    "BMI" : "28.1",
    "DiabetesPedigreeFunction" : "0.167",
    "Age" : "21",
    "Outcome" : "0"
  }
}
```

- เมื่อแก้ไขเสร็จคลิก “Export” เพื่อดาวน์โหลดไฟล์

ข.2 วิธีการแปลงข้อมูลจาก CSV เป็น XML

ข.2.1 การแปลงข้อมูลเครื่องมือออนไลน์ สามารถเลือกได้ 2 วิธีดังนี้

- 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 การแปลงข้อมูลด้วยเครื่องมือที่ติดตั้งบนเครื่อง สามารถใช้ Openrefine ๓.๐ 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 Attribute 2>
  <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

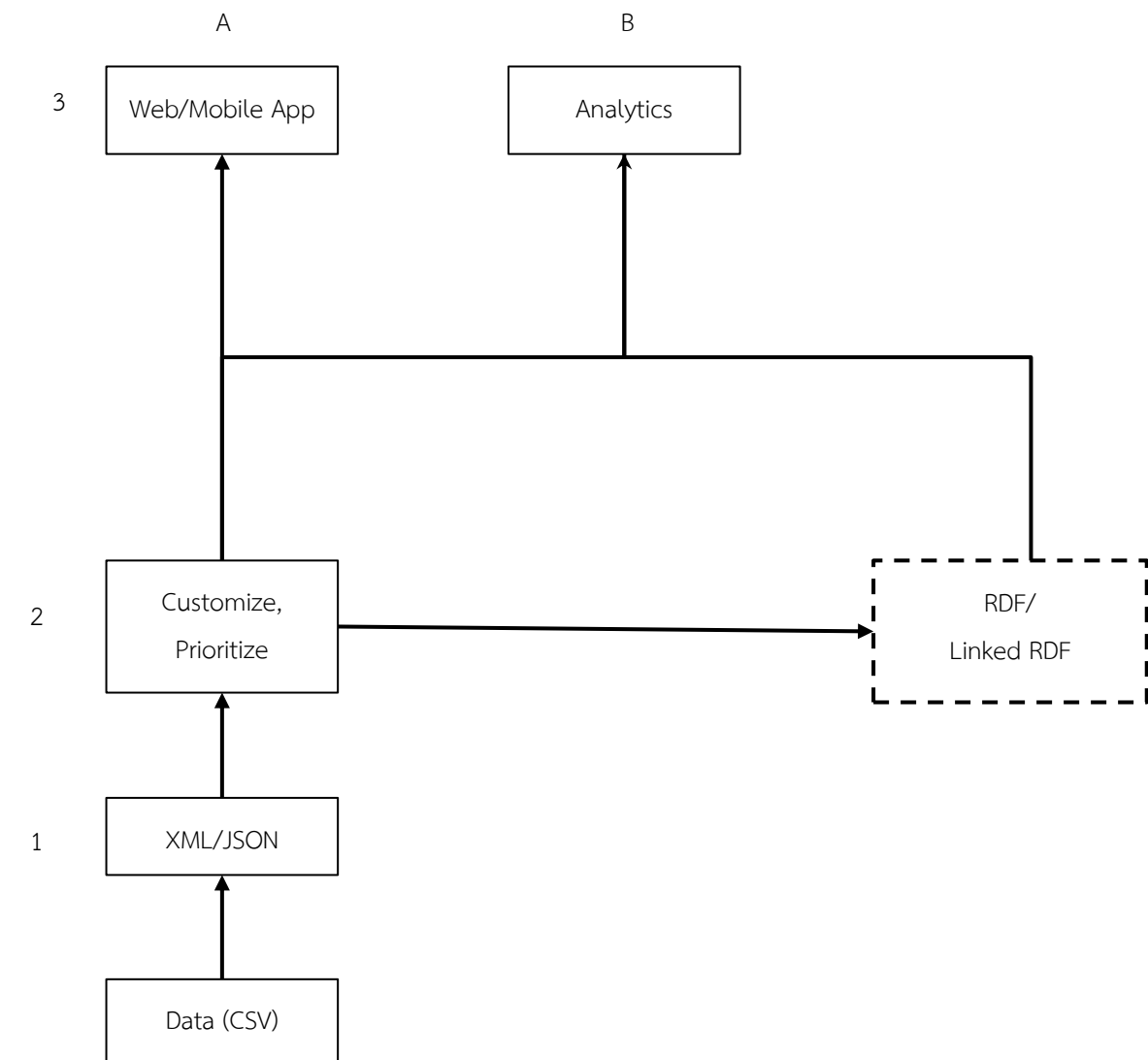
name มีค่า Belgian Waffles
price มีค่า \$5.95
description มีค่า Two of our famous Belgian Waffles with plenty of real maple
syrup
calories มีค่า 650

ในกรณีที่ผู้ใช้งานต้องการแปลงข้อมูล CSV เป็น XML โดยใช้ Openrefine ซึ่งโปรแกรมจะอ่านค่า {{cells ["ชื่อ Attribute1"].value }} เป็นค่าของ Attribute นั้นๆ 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

ข.3 ขั้นตอนหลังจากการแปลงไฟล์เป็น JSON และ XML เพื่อนำไปสู่การต่อยอดใช้งาน



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



(ร่าง) กรอบการวิเคราะห์ข้อมูลขนาดใหญ่ภาครัฐ (Government Big Data Analytics Framework)

ฉบับพิมพ์ครั้งที่ 1

ข้อเสนอแนะเพิ่มเติม

ควรปรับปรุงในเรื่อง

.....

.....

.....

.....

.....

ขอแก้ไขเนื้อหา

.....

.....

.....

.....

.....

ควรเพิ่มเติมเนื้อหา

.....

.....

.....

.....

.....

(กรณีที่กรอกข้อมูลไม่เพียงพอให้เขียนเพิ่มเติมในกระดาษเปล่าแนบมาพร้อมเอกสารนี้หรือส่งความเห็นมาที่ deinfra@onde.go.th)

จาก (นาย/นาง/นางสาว)

หน่วยงาน/บริษัท.....

โทรศัพท์.....โทรสาร.....อีเมล.....

(คณะกรรมการออกแบบสถาปัตยกรรม (Architecture Design) ระบบบูรณาการข้อมูลภาครัฐ ภายใต้คณะกรรมการการขับเคลื่อนการดำเนินนโยบายเพื่อใช้ประโยชน์ข้อมูลขนาดใหญ่ฯ ขอขอบพระคุณล่วงหน้ามา ณ ที่นี้)



ชำระค่าส่งเป็นรายเดือน
ใบอนุญาตที่ พิเศษ 65/2559
ปณท.ศูนย์ราชการเฉลิมพระเกียรติ

ส่งถึง

กองโครงสร้างพื้นฐานเทคโนโลยีดิจิทัล
สำนักงานคณะกรรมการดิจิทัลเพื่อเศรษฐกิจและสังคมแห่งชาติ
เลขที่ 120 หมู่ 3 ชั้น 6-9 อาคารรัฐประศาสนภักดี
ศูนย์ราชการเฉลิมพระเกียรติ 80 พรรษาฯ
ถนนแจ้งวัฒนะ กรุงเทพมหานคร 10210



กระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม
120 หมู่ที่ 3 อาคารรัฐประศาสนภักดี
ศูนย์ราชการเฉลิมพระเกียรติ 80 พรรษา 5 ธันวาคม 2550
ถนนแจ้งวัฒนะ แขวงทุ่งสองห้อง เขตหลักสี่ กรุงเทพมหานคร 10210
www.mdes.go.th

สงวนลิขสิทธิ์ © 2019 กระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม