doi: 10.14456/jiskku.2021.9

ข้อมูลที่เครื่องคอมพิวเตอร์สามารถประมวลผลได้สำหรับการนำไปใช้ในการเปิดเผยข้อมูลและการวิเคราะห์ข้อมูลของภาครัฐ

A Machine-Readable Framework for Open Government Data and Government Data Analytics

วิชญ์วิสิฐ เกษรสิทธิ์[1*]

Witwisit Kesornsit[1*]

*Corresponding author email: witwisit.kes@gmail.com

Received: January 4, 2021

Revised: March 18, 2021

Accepted: March 24, 2021

บทคัดย่อ

วัตถุประสงค์: การศึกษาครั้งนี้ใช้วิธีการศึกษาด้วยวิธีการสำรวจเอกสารและการวิเคราะห์เนื้อหา มีวัตถุประสงค์เพื่อศึกษาและพัฒนากรอบแนวทางการเตรียมข้อมูลให้อยู่ในรูปแบบข้อมูลที่เครื่องคอมพิวเตอร์สามารถประมวลผลได้สำหรับการเปิดเผยและวิเคราะห์ข้อมูลภาครัฐ และวัดผลชุดข้อมูลตามกรอบระดับข้อมูลที่เครื่องคอมพิวเตอร์สามารถประมวลผลได้ของประเทศที่โดดเด่นด้านการเปิดเผยข้อมูล

วิธีการศึกษา: การศึกษาเอกสารและวิเคราะห์เนื้อหาด้านคุณสมบัติของข้อมูลที่เครื่องคอมพิวเตอร์สามารถประมวลผลได้และพัฒนากรอบระดับข้อมูลที่เครื่องคอมพิวเตอร์สามารถประมวลผลได้ รวมทั้งทำการกำหนดคุณสมบัติของข้อมูลในแต่ละระดับ หลังจากนั้นดำเนินการวัดผลระดับการเปิดเผยข้อมูลตามกรอบที่พัฒนาขึ้นกับข้อมูลบนช่องทางการเข้าถึงข้อมูลเปิดของเอเชีย (AODP)

[1] กลุ่มงานเดตาโซลูชันส์ภาครัฐ ฝ่ายเดตาโซลูชันส์ สำนักงานพัฒนารัฐบาลดิจิทัล (องค์การมหาชน); Government Data Solution Division, Department of Data Solution, Digital Government Development Agency (Public Organization)

ข้อค้นพบ: คุณสมบัติของข้อมูลที่เครื่องคอมพิวเตอร์สามารถประมวลผลได้ประกอบด้วย 6 ประเด็น คือ (1) ข้อมูลบันทึกอยู่ในรูปแบบไฟล์ที่คอมพิวเตอร์สามารถนำไปประมวลผลได้แบบอัตโนมัติและต้องมีโครงสร้างแบบตารางในลักษณะคอลัมน์และแถว (2) ข้อมูลจะต้องไม่จำกัดสิทธิ์และการเข้ารหัสแบบเป็นสากลที่สามารถใช้งานได้กับทุกภาษา (3) ข้อมูลที่ได้จากการประมวลผลจะต้องไม่ผิดความหมายตามโครงสร้างข้อมูลหรือเนื้อหาของข้อมูล (4) ข้อมูลมีความคงเส้นคงวาของประเภทข้อมูลและรูปแบบข้อมูล (5) การตั้งชื่อตัวแปรควรให้สอดคล้องกับมาตรฐานการตั้งชื่อตัวแปร และ (6) ข้อมูลสามารถประมวลผลด้วยรหัสที่คอมพิวเตอร์อ่านได้โดยไม่ต้องผ่านกระบวนการแปลงให้อยู่ในรูปแบบรหัสคอมพิวเตอร์ ซึ่งจัดระดับของข้อมูลออกเป็น 4 ระดับ คือ ขั้นต้น ขั้นกลาง ขั้นสูง และขั้นเสริม ผลการสำรวจและวัดผลชุดข้อมูลเปิดพบว่าระดับการเปิดเผยมากที่สุดอยู่ในขั้นกลาง และข้อมูลเปิดที่มีคะแนนสูงสุดคือข้อมูลเปิดของไต้หวัน สำหรับข้อมูลเปิดของไทยมีระดับการเปิดเผยข้อมูลอยู่ในขั้นต้นและขั้นกลาง
การประยุกต์ใช้จากการศึกษานี้: ผลการศึกษาสามารถนำไปใช้ในการกำหนดแนวทางการเตรียมข้อมูลและการพัฒนามาตรฐานการเปิดเผยข้อมูลสู่สาธารณะในรูปแบบที่เครื่องคอมพิวเตอร์สามารถประมวลผลได้ นอกจากนี้สามารถนำไปใช้ในการบริหารจัดการข้อมูลตามแนวทางข้อมูลที่เครื่องคอมพิวเตอร์สามารถประมวลผลได้เพื่อให้รองรับการวิเคราะห์ข้อมูลขนาดใหญ่ของภาครัฐ

คำสำคัญ: ข้อมูลที่เครื่องคอมพิวเตอร์สามารถอ่านได้ ความสามารถในการอ่านได้ของคอมพิวเตอร์ ข้อมูลเปิดภาครัฐ การวิเคราะห์ข้อมูลของภาครัฐ

Abstract

Purpose of the study: This documentary research aimed at (1) examining research papers and other related materials on machine-readable and conducting their content analysis (2) developing machine-readable framework for transforming data into machine-readable data for computer processing and making it open to the public and (3) locating machine-readable format in the public open dataset of distinguished open data portal.

Methodology: Research papers and related works, including data guidelines were examined and analyzed. Then the machine-readable framework and characteristics of government open data were set. Moreover, the proposed framework was used to evaluate the machine-readable format in the public open dataset available on Asia Open Data Portal (AODP).

Main Findings: There were 6 characteristics of machine-readable dataset consisting of (1) data format which could be automatically processed by a computer, and being structured data (2) data format which no entity had exclusive control and must be encrypted internationally format (3) data format that could be processed while ensuring no semantic meaning was lost (4) consistency of data in format and type (5) variable naming that follows the rules and naming conventions and (6) being the data which could be interrogated and processed by computer code and classified in 4 levels: basic, medium, advanced and optional. In addition, the evaluation of the dataset revealed that most datasets were opened at a medium level. Taiwan dataset was found to be the highest open data while Thailand opened published data at the basic and medium levels.

Applications of this study: Findings from this investigation can be used to develop guidelines for data preparation and open data policy formulation. Also, the study is useful in data management which is a good preparation for future government big data analytics.

Keywords: Machine-Readable, Machine Readability, Open Government Data, Government Data Analytics

บทนำ

ในยุคที่โลกถูกขับเคลื่อนด้วยข้อมูลและสารสนเทศ หน่วยงานต่าง ๆ ไม่ว่าจะเป็นหน่วยงานภาครัฐหรือเอกชนต่างก็ให้ความสำคัญกับการใช้ข้อมูลเพื่อประโยชน์ในการตัดสินใจของผู้บริหาร การวิเคราะห์และออกแบบนโยบายการพัฒนาองค์กรที่สะท้อนความเป็นจริงจากผลดำเนินการขององค์กรและสภาพความเป็นไปของสังคมในยุคโลกาภิวัตน์ รวมทั้งการนำข้อมูลมาต่อยอดเพื่อสร้างสรรค์นวัตกรรมและการให้บริการใหม่ ๆ ที่สอดคล้องกับความต้องการพื้นฐานของผู้รับบริการเป็นสำคัญ โดยแต่ละหน่วยงานต่างมีการนำเทคโนโลยีที่ทันสมัยเข้ามาช่วยในการจัดเก็บและประมวลผลข้อมูล ซึ่งข้อมูลเหล่านี้สามารถนำมาใช้ประโยชน์ได้ทั้งภาครัฐ ภาคเอกชน รวมถึงภาคประชาชน ในปัจจุบันข้อมูลส่วนใหญ่ของภาครัฐถูกจัดเก็บอยู่ภายในหน่วยงานหรือระหว่างหน่วยงานเป็นหลัก ส่งผลให้การพัฒนาบริการหรือการใช้ประโยชน์จากข้อมูลยังมีข้อจำกัด การเปิดเผยหรือแบ่งปันข้อมูลเพื่อให้เกิดการใช้ประโยชน์จากข้อมูล จึงเป็นสิ่งจำเป็นต่อการดำเนินงานของหน่วยงานของรัฐในทุกระดับ ทั้งในระหว่าง

หน่วยงานของรัฐด้วยกันและระหว่างหน่วยงานของรัฐกับประชาชน เพื่อให้ทุกภาคส่วนเข้าใจถึงสถานการณ์และแนวทางการแก้ไขปัญหาต่าง ๆ ที่มีความซับซ้อน ปรับเปลี่ยนให้เป็นการทำงานเชิงรุก เน้นการยกระดับไปสู่ความร่วมมือกัน แสวงหาความคิดริเริ่มและสร้างนวัตกรรม แสดงถึงความโปร่งใสของการดำเนินงานของหน่วยงานของรัฐ และสร้างประโยชน์ในการอำนวยความสะดวกแก่ประชาชนในการเข้าถึงข้อมูล รวมถึงสนับสนุนให้ภาคเอกชนและภาคประชาชนนำข้อมูลที่เปิดเผยไปสร้างนวัตกรรมและบริการเพื่อยกระดับการพัฒนาประเทศต่อไป

การเปิดเผยและแบ่งปันข้อมูลในรูปแบบข้อมูลแบบเปิดหรือข้อมูลเปิด (Open data) เป็นแนวความคิดที่ว่าด้วยข้อมูลบางอย่างควรจะสามารถใช้ประโยชน์ได้โดยเสรีสำหรับทุกคน เพื่อใช้และเผยแพร่ตามความต้องการของผู้ใช้ โดยไม่มีข้อจำกัดจากลิขสิทธิ์ สิทธิบัตร หรือกลไกการควบคุมต่าง ๆ สามารถนำกลับมาใช้ใหม่ (Reuse) และแจกจ่ายต่อได้โดยต้องระบุแหล่งที่มาและต้องใช้เงื่อนไขเดียวกันกับที่มาหรือตามที่เจ้าของงานกำหนด เป้าหมายของขบวนการข้อมูลเปิดคล้ายคลึงกับแนวคิดการ "เปิด" ของแนวคิดอื่น ๆ เช่น โอเพนซอร์ซ (Open source) ฮาร์ดแวร์แบบเปิด เนื้อหาแบบเปิด สื่อการสอนและการเรียนรู้แบบเปิด หลักสูตรที่เรียนออนไลน์แบบเปิด และการเข้าถึงแบบเปิด (Digital Government Development Agency, 2020; United Nations, 2013) ซึ่งตามหลักการของข้อมูลเปิดถือได้ว่าเป็นกลไลสำคัญที่สนับสนุนให้เกิดการทำงานร่วมกัน (Interoperability) อย่างเป็นรูปธรรม การเปิดเผยข้อมูลในรูปแบบข้อมูลเปิดเป็นกระบวนการที่ช่วยสร้างความโปร่งใสให้กับองค์กร (Transparency) โดยเปิดโอกาสให้ประชาชนหรือประชาสังคมสามารถตรวจสอบการดำเนินงานตามนโยบายหรือสถิติผลประกอบการขององค์กรได้ นอกจากนี้ข้อมูลเปิดจัดเป็นแหล่งข้อมูลหนึ่งที่ช่วยผลักดันการสร้างนวัตกรรมและบริการใหม่ ๆ ให้สามารถเผยแพร่สู่สังคมและเชิงพาณิชย์ (Digital Government Development Agency, 2020; Tauberer, 2019)

การบรรลุเป้าประสงค์ของการเปิดเผยข้อมูลสำหรับภาครัฐและสนับสนุนการนำข้อมูลเปิดไปใช้ในการสร้างสรรค์นวัตกรรมและการวิเคราะห์ข้อมูลนั้น คุณลักษณะของข้อมูลที่อยู่ในรูปแบบข้อมูลที่เครื่องคอมพิวเตอร์สามารถประมวลผลได้จึงเป็นปัจจัยที่มีความสำคัญ เนื่องด้วยปัจจุบันเทคโนโลยีในการพัฒนานวัตกรรมและเทคโนโลยีการวิเคราะห์ข้อมูล เช่น เครื่องมือสร้างจิตทัศน์จากข้อมูล (Data visualization) การเรียนรู้ของเครื่องจักร (Machine learning) ดำเนินการโดยใช้คอมพิวเตอร์เป็นหลัก และในทุกกระบวนการของใช้งานข้อมูลจะต้องผ่านการประมวลผลโดยใช้เครื่องคอมพิวเตอร์เพื่อนำไปต่อยอดในการใช้ประโยชน์ข้อมูล ดังนั้นการเตรียมข้อมูลให้อยู่ในรูปแบบข้อมูลที่เครื่องคอมพิวเตอร์สามารถประมวลผลได้จะส่งเสริมให้ผู้ใช้สามารถนำข้อมูลไปใช้ได้ง่ายและเป็นอิสระ ไม่ต้องผ่านกระบวนการจัดการโดยมนุษย์ก่อนการประมวลผลซึ่งกระบวนการดังกล่าวเป็นกระบวนการที่ต้องใช้เวลาในการดำเนินการและต้องใช้กำลังคนในการทำงานค่อนข้างมาก

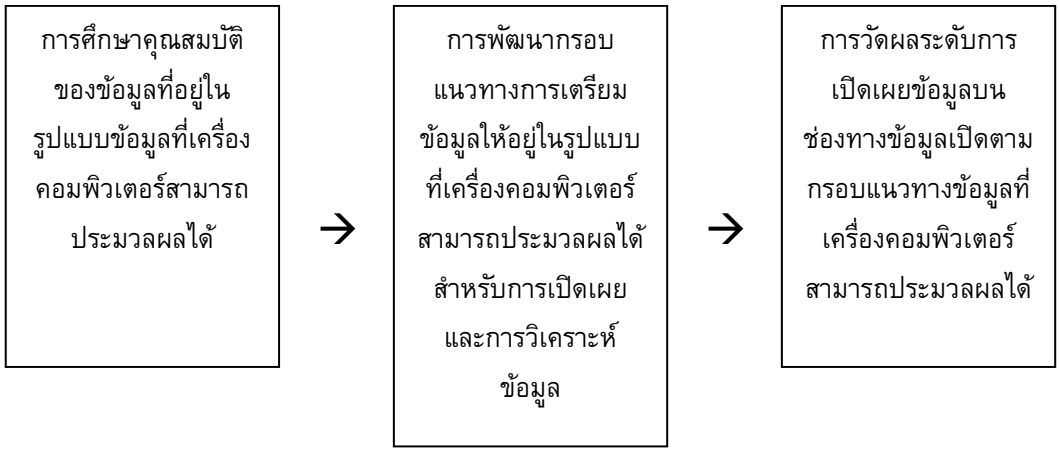
สำหรับปัจจุบันประเทศต่าง ๆ ได้มีการผลักดันให้มีการเตรียมข้อมูลให้อยู่ในรูปแบบที่เครื่องคอมพิวเตอร์สามารถประมวลผลได้เพื่อให้เกิดความคล่องตัวและความสามารถในการใช้งานได้ของข้อมูล ซึ่งแต่ละประเทศจะมีการกำหนดรูปแบบการจัดเตรียมที่เหมือนและแตกต่างกันขึ้นอยู่กับวัตถุประสงค์ของงานที่ต้องการนำข้อมูลไปใช้ ทางผู้วิจัยจึงได้ศึกษาคุณลักษณะของข้อมูลที่อยู่ในรูปแบบข้อมูลที่เครื่องคอมพิวเตอร์สามารถประมวลผลได้และพัฒนากรอบแนวทางการเตรียมข้อมูล พร้อมทั้งดำเนินการจัดระดับของการจัดเตรียมข้อมูล โดยนำเสนอระดับของข้อมูลที่เครื่องคอมพิวเตอร์สามารถประมวลผลได้กับความสามารถในการนำข้อมูลไปใช้ประโยชน์ ซึ่งจะก่อให้เกิดประโยชน์ต่อภาครัฐที่จะกำหนดนโยบายและวางแผนการพัฒนารูปแบบการเปิดเผยข้อมูลให้สอดคล้องกับวัตถุประสงค์ของการใช้ประโยชน์ เช่น ความสามารถในการประมวลผลได้แบบเบื้องต้น ความสามารถในการนำไปใช้ในเชิงฐานข้อมูล ความสามารถในการนำไปใช้ในการพัฒนาแอปพลิเคชัน ความสามารถในการนำไปใช้สำหรับการวิเคราะห์ข้อมูล และความสามารถในการเรียนรู้ของเครื่องจักร เป็นต้น

จากประเด็นที่ประเทศต่าง ๆ ได้มีการผลักดันให้มีการเปิดเผยข้อมูลในรูปแบบที่เครื่องคอมพิวเตอร์สามารถประมวลผลได้ โดยเฉพาะอย่างยิ่งการเปิดเผยข้อมูลบนช่องทางการเข้าถึงข้อมูลเปิดในประเทศต่าง ๆ ผู้วิจัยจึงได้นำกรอบแนวทางที่พัฒนาขึ้นไปใช้ในการวัดผลจำนวนข้อมูลที่เปิดเผยในรูปแบบที่เครื่องคอมพิวเตอร์สามารถประมวลผลได้ตามระดับที่กำหนดในประเทศที่มีศักยภาพในการส่งเสริมการเปิดเผยข้อมูล เพื่อนำเสนอประเทศที่สามารถเป็นต้นแบบในการดำเนินงานและส่งเสริมการเปิดเผยข้อมูลในประเทศไทยและองค์กรที่สนใจ ซึ่งสอดรับกับภารกิจการเปิดเผยข้อมูลเปิดภาครัฐที่หน่วยงานภาครัฐในประเทศไทยต้องดำเนินการตามกฎหมายที่กำหนดให้หน่วยงานของรัฐต้องเปิดเผยข้อมูลเพื่อประโยชน์แก่สาธารณะ ได้แก่ รัฐธรรมนูญแห่งราชอาณาจักรไทย พ.ศ. 2560 พระราชบัญญัติข้อมูลข่าวสารของทางราชการ พ.ศ. 2540 นอกจากนี้ยังมีพระราชบัญญัติการบริหารงานและการให้บริการภาครัฐผ่านระบบดิจิทัล พ.ศ. 2562 ซึ่งกำหนดให้หน่วยงานของรัฐจัดทำข้อมูลที่ต้องเปิดเผยตามกฎหมายว่าด้วยข้อมูลข่าวสารของราชการในรูปแบบข้อมูลดิจิทัลต่อสาธารณะ โดยต้องให้ประชาชนทั่วไปสามารถเข้าถึงได้อย่างเสรีโดยไม่เสียค่าใช้จ่าย และสามารถนำไปเผยแพร่ ใช้ประโยชน์ หรือพัฒนาบริการและนวัตกรรมในรูปแบบต่าง ๆ ได้ รวมถึงแสดงความโปร่งใสในการดำเนินงาน และสามารถตรวจสอบได้จากทุกภาคส่วน (Digital Government Development Agency, 2020)

วิธีดำเนินการศึกษา

การศึกษาครั้งนี้ใช้วิธีการศึกษาแบบเก็บรวบรวมข้อมูลโดยวิธีการสำรวจเอกสารและการวิเคราะห์เนื้อหา โดยมีขั้นตอนการศึกษาคือ (1) การศึกษาคุณลักษณะของข้อมูลที่อยู่ในรูปแบบข้อมูลที่เครื่องคอมพิวเตอร์สามารถประมวลผลได้ (2) พัฒนากรอบแนวทางการเตรียมข้อมูลให้อยู่ในรูปแบบที่เครื่องคอมพิวเตอร์สามารถประมวลผล

ได้ และ (3) ทำการวัดผลระดับการเปิดเผยข้อมูลบนช่องทางข้อมูลเปิดตามกรอบแนวทางการเตรียมข้อมูลให้อยู่ในรูปแบบข้อมูลที่เครื่องคอมพิวเตอร์สามารถประมวลผลได้ทั้งในและต่างประเทศ รวมถึงทำการเปรียบเทียบความสามารถในการเปิดเผยข้อมูลในรูปแบบที่เครื่องคอมพิวเตอร์สามารถประมวลผลได้ โดยกำหนดกรอบแนวคิดการศึกษาดังภาพที่ 1



ภาพที่ 1 แสดงกรอบแนวคิดการศึกษา

ผู้วิจัยได้ดำเนินการตามขั้นตอนต่าง ๆ ที่สอดคล้องกับกรอบแนวคิดการศึกษาจำนวน 3 ขั้นตอนดังนี้

1. ศึกษาคุณลักษณะของข้อมูลสำหรับการเปิดเผยข้อมูลภาครัฐ รูปแบบการเปิดเผยข้อมูลภาครัฐในรูปแบบดิจิทัล และคุณสมบัติของข้อมูลที่เครื่องคอมพิวเตอร์สามารถประมวลผลได้สำหรับการเปิดเผยข้อมูลและการวิเคราะห์ข้อมูล จากมาตรฐานและหลักการการเปิดเผยข้อมูลของศูนย์ข้อมูลเปิดในประเทศต่าง ๆ ที่มีการผลักดันให้มีศูนย์ข้อมูลเปิดและงานวิจัยที่เกี่ยวข้อง ดังนี้

1.1 คุณลักษณะของข้อมูลสำหรับการเปิดเผยข้อมูลภาครัฐ การเปิดเผยชุดข้อมูลเปิดมีการกำหนดรูปแบบของชุดข้อมูลเปิดออกเป็น 3 ประเภท คือ 1) ข้อมูลที่ไม่มีโครงสร้าง (Unstructured data) เป็นข้อมูลชุดข้อมูลที่มีโครงสร้างไม่ชัดเจน หรือไม่สามารถระบุความแน่นอนของโครงสร้างข้อมูลได้ หรือเป็นข้อมูลที่ ไม่สามารถจัดเก็บในลักษณะแถวและคอลัมน์ในฐานข้อมูลเชิงสัมพันธ์ได้ (Relational database) 2) ข้อมูลกึ่งโครงสร้าง (Semi-Structured data) เป็นข้อมูลที่ถูกจัดเก็บอย่างมีรูปแบบในระดับหนึ่งแต่ยังไม่สมบูรณ์ โดยทั่วไปจะมีการเก็บอยู่ในรูปแบบสายอักขระ (String) แต่สามารถระบุสคีมา (Schema) ของข้อมูลได้ โดยจะมีการเก็บชื่อตัวแปร ประเภทของตัวแปร และค่าของข้อมูลอยู่ในแต่ละบรรทัดของภาษา โดยคอมพิวเตอร์สามารถประมวลผลและแปลงเป็นโครงสร้างที่ชัดเจนได้ และ 3) ข้อมูลที่มีโครงสร้าง (Structured data) เป็นข้อมูลที่มี

คุณลักษณะมีโครงสร้างแบบตาราง (Tabular) สามารถระบุแถวและคอลัมน์ได้ชัดเจนและเป็นอัตโนมัติ (Sint, Schaffert, Stroka, & Ferstl, 2009)

1.2 รูปแบบการเปิดเผยข้อมูลภาครัฐในรูปแบบดิจิทัล (Open government data) มีการกำหนดให้สามารถเปิดเผยข้อมูลได้ทุกประเภทโดยข้อมูลที่มีโครงสร้างและกึ่งโครงสร้างจะมีระดับการเปิดเผยชุดข้อมูลที่สูงกว่าข้อมูลที่ไม่มีโครงสร้าง จากแนวทางการเปิดเผยข้อมูลเปิดภาครัฐในรูปแบบดิจิทัลต่อสาธารณะของสำนักงานพัฒนารัฐบาลดิจิทัลและแผนการปรับใช้ระดับ 5 ดาวสำหรับข้อมูลเปิดของ Tim Berners-Lee (2012) กำหนดแนวทางการเปิดเผยข้อมูลไว้เป็น 5 ระดับ โดยระดับ 1 ดาวสามารถเปิดเผยข้อมูลบนเว็บไซต์ได้ทุกรูปแบบและสามารถนำข้อมูลไปใช้ประโยชน์ได้โดยไม่ต้องขออนุญาตจากหน่วยงานของรัฐ (Open license) ซึ่งรูปแบบนี้สามารถสร้างได้ง่าย แต่นำข้อมูลไปต่อยอดใช้ประโยชน์ได้ยาก (Not reusable) ระดับ 2 ดาวสามารถเปิดเผยข้อมูลในรูปแบบข้อมูลที่มีโครงสร้าง สามารถใช้กับซอฟต์แวร์จำกัดสิทธิ์ (Proprietary software) และอ่านได้ด้วยเครื่อง (Machine readable) ระดับ 3 ดาวสามารถเปิดเผยข้อมูลในรูปแบบมาตรฐานแบบเปิด (Open format) และไม่จำกัดสิทธิโดยบุคคลใด (Non-proprietary) ระดับ 4 ดาวสามารถเปิดเผยข้อมูลในรูปแบบที่ใช้ยูอาร์ไอ (Uniform Resource Identifier: URI) ในการระบุตัวตนของข้อมูล และชี้ไปยังตำแหน่งของข้อมูลนั้น และระดับ 5 ดาวสามารถเปิดเผยข้อมูลในรูปแบบที่ข้อมูลสามารถเชื่อมโยงไปสู่แหล่งข้อมูลอื่น หรืออ้างอิงข้อมูลในชุดข้อมูลอื่นได้ (Linked data) (Berners-Lee, 2012; Digital Government Development Agency, 2020)

1.3 ข้อมูลที่เครื่องคอมพิวเตอร์สามารถประมวลผลได้ (Machine readable) จากการศึกษาคุณสมบัติของข้อมูลที่เครื่องคอมพิวเตอร์สามารถประมวลผลได้สำหรับการเปิดเผยข้อมูลและการวิเคราะห์ข้อมูล โดยได้ดำเนินการศึกษาและวิเคราะห์ข้อมูลจากมาตรฐานและหลักการการเปิดเผยข้อมูล ดังแสดงรายละเอียดในข้อ 1-5 และแนวทางการดำเนินงานของศูนย์ข้อมูลเปิดในประเทศต่าง ๆ ที่มีการผลักดันให้มีศูนย์ข้อมูลเปิด ดังแสดงรายละเอียดในข้อ 6-8 รวมทั้งงานวิจัยที่เกี่ยวข้อง ดังแสดงรายละเอียดในข้อ 9-10 ซึ่งสามารถอธิบายคำนิยามของข้อมูลที่เครื่องคอมพิวเตอร์สามารถประมวลผลได้ ดังนี้

(1) ประกาศในกฎหมายทางด้านข้อมูลเปิด (OPEN Government Data Act) ของประเทศสหรัฐอเมริกาให้ความหมายของข้อมูลที่เครื่องคอมพิวเตอร์สามารถประมวลผลได้ คือข้อมูลที่อยู่ในรูปแบบที่คอมพิวเตอร์สามารถประมวลผลได้ง่ายโดยไม่ต้องผ่านการจัดการข้อมูลโดยใช้มนุษย์ก่อนการประมวลผล ซึ่งข้อมูลที่ได้จากการประมวลผลจะต้องไม่ผิดความหมาย และสามารถเข้าถึงและนำไปใช้ประโยชน์ได้อย่างอิสระ (One Hundred Fifteenth Congress of the United States of America, 2018)

(2) แนวทางการส่งมอบและการดำเนินการของเอกสารทางงบประมาณ (Circular No A-11 Preparation, Submission and Execution of the Budget) ของหน่วยงานด้านการบริหารและงบประมาณ

สำนักงานบริหารของประธานาธิบดี ในประเทศสหรัฐอเมริกา กำหนดแนวทางการเตรียมข้อมูลในรูปแบบที่เครื่องคอมพิวเตอร์สามารถประมวลผลได้ คือรูปแบบมาตรฐานของภาษาคอมพิวเตอร์ซึ่งสามารถอ่านหรือประมวลได้อย่างอัตโนมัติ ได้แก่ ไฟล์จากภาษาที่กำหนดกฎหรือรูปแบบของการเข้ารหัสของข้อมูล (Extensible Markup Language: XML) รูปแบบการแลกเปลี่ยนข้อมูลคอมพิวเตอร์ (JavaScript Object Notation: JSON) ไฟล์ข้อความที่ใช้สำหรับเก็บข้อมูลในรูปแบบตารางและใช้เครื่องหมายจุลภาคในการแบ่งคอลัมน์ (Comma Separated Value: CSV) (Office of Management and Budget, 2020)

(3) คู่มือข้อมูลเปิด (Open data handbook) พัฒนาโดยองค์กรความรู้แบบเปิด (Open knowledge foundation) ให้ความหมายของข้อมูลที่เครื่องคอมพิวเตอร์สามารถประมวลผลได้ คือข้อมูลในรูปแบบที่สามารถประมวลผลได้อย่างอัตโนมัติโดยคอมพิวเตอร์ เป็นข้อมูลที่มีโครงสร้าง เช่น Extensible Markup Language (XML), CSV หรือ JSON นอกจากนี้ยังมีไฟล์ที่เครื่องสามารถประมวลผลได้ในรูปแบบอื่นที่แตกต่างไปจากที่ได้กล่าวไว้ เช่น ข้อมูลเชิงภูมิศาสตร์ (Geospatial data) ได้แก่ ไฟล์ที่สร้างจากไวยากรณ์และรูปแบบไฟล์ XML สำหรับการทำโมเดลและการจัดเก็บคุณลักษณะทางภูมิศาสตร์ ได้แก่ Keyhole Markup Language: KML และ Zipped KML Format: KMZ ซึ่งไฟล์ประเภทดังกล่าวจะมีความแตกต่างจากรูปแบบข้อมูลแบบตาราง (Open Knowledge Foundation, n.d.)

(4) แนวทางการดำเนินงานด้านข้อมูลเปิดของยุโรป (European commission) ได้รับรองกฎบัตร G8 Open Data กับสมาชิกและข้อมูลที่เครื่องคอมพิวเตอร์สามารถประมวลผลได้เป็นเงื่อนไขหนึ่งในการดำเนินงานเปิดเผยข้อมูลเพื่อให้สามารถเข้าถึงและนำข้อมูลไปใช้งานได้ (EU Open Data Portal, 2013) และหน่วยงานด้านหลักทรัพย์และตลาดของยุโรปกล่าวถึงข้อมูลที่เครื่องคอมพิวเตอร์สามารถประมวลผลได้ว่าเป็นข้อมูลที่สามารถแปลงเป็นข้อมูลในรูปแบบอื่นได้ง่าย เช่นภาษาที่ใช้ในการเขียนโปรแกรมเพื่อจัดการกับฐานข้อมูล (Structured Query Language: SQL) หรือ XLS โดยหลีกเลี่ยงการจัดการข้อมูลโดยการพิมพ์ข้อมูลใหม่ (rekeying) (European Securities and Markets Authority, 2017) และ

(5) แนวทางการเปิดเผยข้อมูลเปิดภาครัฐในรูปแบบดิจิทัลต่อสาธารณะของสำนักงานพัฒนารัฐบาลดิจิทัล (องค์การมหาชน) กำหนดรูปแบบของชุดข้อมูลเปิดในแต่ละประเภท เพื่อให้ผู้ใช้ข้อมูลสามารถนำไปใช้งานได้สะดวกมากขึ้น ซึ่งคุณลักษณะสำคัญของรูปแบบชุดข้อมูล ประกอบด้วย การเปิดเผยข้อมูลในรูปแบบที่อ่านได้ด้วยเครื่อง (Machine readable) เป็นข้อมูลที่มีโครงสร้าง ข้อมูลในรูปแบบมาตรฐานแบบเปิดและไม่จำกัดสิทธิ์โดยบุคคลใด ซึ่งได้กล่าวถึงข้อมูลที่อ่านได้ด้วยเครื่องว่าเป็นข้อมูลที่ต้องมีโครงสร้าง สามารถอ่านได้ด้วยเครื่อง และนำข้อมูลไปใช้งานต่อได้ (Digital Government Development Agency, 2020)

(6) ช่องทางการเข้าถึงข้อมูลเปิดของรัฐแคลิฟอร์เนีย ประเทศสหรัฐอเมริกา (California Open Data Portal) กล่าวถึงข้อมูลที่เครื่องคอมพิวเตอร์สามารถประมวลผลได้ว่าข้อมูลที่จัดอยู่ในรูปแบบที่คอมพิวเตอร์สามารถประมวลผลแบบอัตโนมัติได้และไม่ต้องผ่านการจัดการข้อมูลโดยมนุษย์ก่อนการประมวลผล ข้อมูลจะต้องจัดอยู่ในรูปแบบที่มีโครงสร้าง และบันทึกอยู่ในไฟล์นามสกุล CSV, XML หรือ JSON (California Open Data Portal, n.d.)

(7) ช่องทางการเข้าถึงข้อมูลเปิดภาครัฐของประเทศออสเตรเลีย (Open Data of Government Western Australia) กล่าวถึงข้อมูลที่เครื่องคอมพิวเตอร์สามารถประมวลผลได้ว่าเป็นข้อมูลที่สามารถอ่านและแปลผลโดยคอมพิวเตอร์ได้โดยไม่ต้องผ่านการจัดการข้อมูลโดยมนุษย์ก่อนการประมวลผล ข้อมูลจัดอยู่ในรูปแบบที่เป็นโครงสร้าง มีความคงเส้นคงวาของรูปแบบข้อมูล สามารถประมวลผลด้วยรหัสที่คอมพิวเตอร์อ่านง่าย และไม่ต้องมีการจัดซื้อซอฟต์แวร์หรือระบบปฏิบัติเฉพาะสำหรับการเปิดไฟล์ข้อมูล (Western Australian Land Information Authority, 2020) และ

(8) ช่องทางการเข้าถึงข้อมูลเปิดภาครัฐของประเทศเอสโตเนีย (Estonian Open Government Data Portal) กล่าวว่าข้อมูลเปิดควรดำเนินการในรูปแบบข้อมูลดิบที่เครื่องคอมพิวเตอร์สามารถประมวลผลได้ โดยเปิดให้ใช้งานได้อย่างอิสระ ไม่มีลิขสิทธิ์หรือสิทธิบัตรในการใช้งานหรือการแจกจ่าย เป็นข้อมูลที่จัดเก็บอยู่ในฐานข้อมูลหรือเว็บเซอร์วิซ (Web service) ซึ่งสามารถนำออกมาเป็นข้อมูลในไฟล์ CSV, XML หรือ JSON ได้ (Estonian Open Government Data Portal, n.d.)

(9) Stevens (2019) ในบทความเรื่อง “Open data, closed government: Unpacking data.gov.sg” ได้ศึกษาหลักการแบ่งปันข้อมูลสำหรับภาครัฐในประเทศสิงคโปร์ ประกอบด้วย ข้อมูลควรสามารถเข้าถึงได้โดยง่าย ข้อมูลจะต้องมีความพร้อมใช้สำหรับการนำไปสร้างสรรค์หรือต่อยอดต่อได้ ข้อมูลมีการเปิดเผยตามระยะเวลาที่เหมาะสม ข้อมูลจัดอยู่ในรูปแบบที่เครื่องคอมพิวเตอร์สามารถประมวลผลได้ เช่น CSV หรือ XLS ข้อมูลมีการจัดรูปแบบให้เป็นข้อมูลดิบหรือมีความละเอียดของรายการข้อมูลให้มากที่สุดโดยมีการจัดการข้อมูลส่วนบุคคล ซึ่งคุณลักษณะของความสามารถที่เครื่องคอมพิวเตอร์สามารถอ่านได้ (Machine Readable หรือ Machine Readability) สามารถช่วยส่งเสริมให้มีการนำข้อมูลไปใช้ซ้ำ เกิดการสร้างสรรค์ต่อยอดนวัตกรรมจากข้อมูลได้ง่าย (Stevens, 2019) และ

(10) Goëta และ Davies (2019) ในบทความเรื่อง “The Daily Shaping of State Transparency: Standards, Machine-Readability and the Configuration of Open Government Data Policies” ได้ศึกษามาตรฐานของข้อมูลเปิด (Open data standards) ข้อมูลที่เครื่องคอมพิวเตอร์สามารถ

ประมวลผลได้ (Machine-Readability) และการตั้งค่าหรือกำหนดนโยบายของข้อมูลเปิดภาครัฐ (Configuration of Open Government Data Policies) กล่าวว่าปัจจุบันศูนย์กลางข้อมูลเปิดภาครัฐมีการกำหนดนโยบายในการเปิดเผยข้อมูลในรูปแบบที่เครื่องคอมพิวเตอร์สามารถประมวลผลได้เพื่อให้สามารถนำไปใช้ประโยชน์ได้ง่าย ปราศจากข้อจำกัดในการเข้าถึงข้อมูล และรูปแบบของข้อมูลที่ไม่สนับสนุนการนำไปใช้งานต่อได้ (Fecher & Friesike, 2014) สำหรับไฟล์ที่สามารถประมวลผลได้ด้วยเครื่องคอมพิวเตอร์ ได้แก่ CSV, XML หรือ GTFS (General Transit Feed Specification) ซึ่งเป็นไฟล์ CSV ประเภทหนึ่งที่มีการเก็บข้อมูลตามมาตรฐาน GTFS (Goëta & Davies, 2019)

2. พัฒนากรอบแนวทางการเตรียมข้อมูลให้อยู่ในรูปแบบข้อมูลที่เครื่องคอมพิวเตอร์สามารถประมวลผลได้สำหรับการนำไปใช้ในการเปิดเผยข้อมูลตามกฎหมายที่เกี่ยวข้องที่กำหนดให้หน่วยงานของรัฐต้องเปิดเผยข้อมูลเพื่อประโยชน์แก่สาธารณะ และสามารถอำนวยความสะดวกต่อการนำไปวิเคราะห์ข้อมูลและการเรียนรู้ของเครื่องทั้งด้านความสามารถในการใช้งานได้ของข้อมูล (Usability) ด้านประสิทธิภาพการประมวลผลของเครื่อง (Performance) และด้านความถูกต้องของข้อมูลเมื่อผ่านการประมวลผลด้วยเครื่อง (Correctness) โดยทำการศึกษาจากแนวทางการเปิดเผยข้อมูลในรูปแบบที่เครื่องคอมพิวเตอร์สามารถประมวลผลได้ของศูนย์ข้อมูลเปิดในประเทศต่าง ๆ โดยทำการกำหนดคุณสมบัติของข้อมูล ระดับของรูปแบบของข้อมูล และความสามารถในการนำไปใช้ประโยชน์

3. ดำเนินการวัดผลระดับการเปิดเผยข้อมูลบนช่องทางการเข้าถึงข้อมูลเปิดตามกรอบการเตรียมข้อมูล ในการศึกษาครั้งนี้ผู้วิจัยดำเนินการวัดผลชุดข้อมูลบนช่องทางการเข้าถึงข้อมูลเปิดของเอเชีย (Asia Open Data Partnership, 2015) ของชุดข้อมูลที่เผยแพร่และปรับปรุงข้อมูลล่าสุดในปี 2563 จำนวน 3 องค์กร (Place) ที่มีคะแนนดัชนีการเปิดเผยข้อมูลระดับโลก (Global Open Data Index: GOI) สูงสุด โดยมีการวัดผลในประเด็นต่อไปนี้

(1) การเปิดเผยข้อมูลที่สอดคล้องกับนิยามของข้อมูลเปิดที่กล่าวว่าข้อมูลในรูปแบบข้อมูลดิจิทัลที่สามารถเปิดเผยให้ทุกคนสามารถเข้าถึงได้อย่างเสรี ไม่จำกัดแพลตฟอร์ม ไม่เสียค่าใช้จ่าย และนำไปเผยแพร่ ทำซ้ำหรือใช้ประโยชน์ได้โดยไม่จำกัดวัตถุประสงค์

(2) บทบาทของภาครัฐในการผลักดันให้เกิดการเปิดเผยข้อมูลในหมวดหมู่หลักที่สำคัญ โดยกำหนดให้มีชุดข้อมูลในแต่ละหมวดหมู่ครบตามที่กำหนดไว้

(3) การเปิดเผยข้อมูลที่สะท้อนข้อมูลระดับประเทศ และ

(4) การวัดผลดังกล่าวไม่ได้สะท้อนการดำเนินการทางด้านข้อมูลเปิดในระดับประเทศ แต่จะสะท้อนในระดับหน่วยงานหรือองค์กรที่ดำเนินการทางด้านข้อมูลเปิด โดยองค์กรที่เลือกศึกษาประกอบด้วยข้อมูล

เปิดของไต้หวัน (GOI = 90%) ข้อมูลเปิดของญี่ปุ่น (GOI = 61%) ข้อมูลเปิดของสิงคโปร์ (GOI = 60%) และเปรียบเทียบกับข้อมูลเปิดของไทย (GOI = 34%) (Global Open Data Index, 2017) และกำหนดขอบเขตของการวัดผลข้อมูลในกลุ่มข้อมูลด้านเศรษฐกิจและการเงิน (Economy and finance) เพื่อทดลองใช้กรอบที่ผู้วิจัยนำเสนอ นำข้อมูลที่ได้จากการทดลองในการวิเคราะห์และเปรียบเทียบผลการเปิดเผยข้อมูลจากตัวอย่างการวิจัย

ผลการศึกษา

1. คุณลักษณะของข้อมูลที่เครื่องคอมพิวเตอร์สามารถประมวลผลได้

จากการศึกษาและวิเคราะห์คำนิยามและแนวทางการเตรียมข้อมูลให้อยู่ในรูปแบบเครื่องคอมพิวเตอร์สามารถประมวลผลได้ สามารถพัฒนากรอบแนวทางการเตรียมข้อมูลซึ่งสามารถแสดงรายละเอียดคุณลักษณะที่สำคัญจำนวน 6 ประการดังนี้

1.1 ข้อมูลบันทึกอยู่ในรูปแบบไฟล์ที่คอมพิวเตอร์สามารถอ่านและนำไปประมวลผลได้แบบอัตโนมัติโดยไม่ต้องผ่านการจัดการข้อมูลโดยมนุษย์ก่อนการประมวลผล เป็นไฟล์ที่มีลักษณะเป็นโครงสร้างหรือกึ่งโครงสร้างโดยข้อมูลจะต้องมีการจัดเรียงแบบตารางที่มีรูปแบบชัดเจนและเป็นระเบียบ ซึ่งคอมพิวเตอร์สามารถประมวลผลคอลัมน์และแถวได้อัตโนมัติ สำหรับประเภทของไฟล์ที่จัดอยู่ในประเภทที่คอมพิวเตอร์สามารถอ่านได้ ได้แก่ CSV, XML, JSON, RDF, GeoJSON, KML และ KMZ เป็นต้น สำหรับข้อมูลที่มีโครงสร้างจะต้องกำหนดให้แถวแรกเป็นชื่อคอลัมน์หรือชื่อตัวแปร และแถวถัดไปเป็นค่าของข้อมูล ตัวแปรแต่ละตัวและค่าของข้อมูลจะต้องมีการจัดเก็บอยู่เพียงช่องเดียว (Single Cell) เท่านั้น

1.2 ข้อมูลสามารถเข้าถึงและนำไปใช้ประโยชน์ต่อได้โดยไม่มีข้อจำกัดด้านรูปแบบของข้อมูลหรือต้องมีการจัดซื้อซอฟต์แวร์หรือระบบปฏิบัติเฉพาะสำหรับการเปิดไฟล์ข้อมูล ซึ่ง (1) ข้อมูลจะต้องไม่จำกัดสิทธิ์ ซึ่งอยู่ในรูปแบบมาตรฐานเปิดที่สามารถใช้ได้หลายแพลตฟอร์ม และต้องไม่ถือครองกรรมสิทธิ์หลังจากนำข้อมูลเปิดไปใช้ประโยชน์ (Brickley, Burgess, & Noy, 2019) หากข้อมูลมีข้อจำกัดในการเข้าถึงและเปิดไฟล์จะทำให้เครื่องคอมพิวเตอร์จะไม่สามารถประมวลผลได้โดยง่ายและไม่เป็นอัตโนมัติ และจำเป็นจะต้องผ่านกระบวนการที่ใช้มนุษย์ในการจัดการก่อนการประมวลผล ซึ่งทำให้ไม่สอดรับกับนิยามข้อมูลที่เครื่องคอมพิวเตอร์สามารถประมวลผลได้ และ (2) ระบบการแสดงผลของระบบคอมพิวเตอร์ตัวอักษรในภาษาอังกฤษนั้นจะจัดเก็บอยู่ในลักษณะของรหัสมาตรฐานของสหรัฐอเมริกาเพื่อการแลกเปลี่ยนสารสนเทศ (ASCII) ในการแสดงผลภาษาอื่น ๆ บนการเข้ารหัส (encode) แบบ ASCII นั้นสามารถทำได้ แต่จะขึ้นอยู่กับระบบปฏิบัติการ (OS) ในภาษาไทยมีการใช้วิธีการเข้ารหัสที่หลากหลาย เช่น ในระบบปฏิบัติการแบบวินโดว์ (Windows) จะใช้การเข้ารหัสแบบ Windows874 ซึ่งใช้ได้เฉพาะระบบปฏิบัติการแบบวินโดว์เท่านั้น และเพื่อให้การใช้งานเป็นสากล ไม่ขึ้นอยู่กับ

ระบบปฏิบัติการ จึงได้มีการคิดค้นรหัสคอมพิวเตอร์ใช้แทนตัวอักขระ (Unicode) หรือ Universal Character Set (UCS) ซึ่งรหัสดังกล่าวจะมีการแทนตัวอักษรทุกตัวอักษร โดยไม่ขึ้นต่อระบบปฏิบัติการ โปรแกรม และภาษา สามารถใช้งานได้สะดวกและรองรับทุก ๆ ภาษา โดยปกติแล้วการส่งข้อมูลหรือการแสดงผ่านเว็บไซต์นั้น จะใช้การเข้ารหัสแบบ UTF-8 เพื่อความเป็นสากลและสามารถใช้งานได้กับทุกภาษา ซึ่ง UTF-8 จัดเป็นรหัสคอมพิวเตอร์ใช้แทนตัวอักขระโดยใช้จำนวน 8 บิตในการแทนตัวอักษร ส่วนบนระบบปฏิบัติการแบบวินโดว์จะใช้ UTF-16 และระบบปฏิบัติการลินุกซ์ (Linux) จะใช้ UTF-8 ดังนั้นเพื่อให้คอมพิวเตอร์สามารถประมวลผลได้จึงจำเป็นต้องกำหนดการเข้ารหัสข้อมูลให้สอดคล้องมาตรฐานข้อมูลเปิดและสอดคล้องกับภาษาที่ใช้ในข้อมูล (Yergeau, 2003)

dcode	sex	5	6	7	8	9
1033	1	451	479	506	468	506
1033	2	470	470	474	465	513
1018	1	255	271	267	304	336
1018	2	264	260	255	326	334

(ก) ตารางไขว้หรือตารางไพวอทระหว่าง 2 ตัวแปรคือเพศ (Sex) และอายุ (Age)

dcode	sex	age	amount
1033	1	5	451
1033	1	6	479
1033	1	7	506
1033	1	8	468
1033	1	9	506
1033	2	5	470
1033	2	6	470
...			

(ข) ตารางข้อมูลระเบียน (Transaction data) ซึ่งประกอบด้วยตัวแปร 3 ตัวแปรคือเพศ (Sex), อายุ (Age) และจำนวนประชากร (amount)

ภาพที่ 2 แสดง (ก) การจัดรูปแบบโครงสร้างข้อมูลในรูปแบบตารางไขว้ (Crosstab table) หรือตารางไพวอท (Pivot table) และ(ข) ตารางข้อมูลระเบียน (Transaction data)

1.3 ข้อมูลที่ได้จากการประมวลผลจะต้องไม่ผิดความหมายตามโครงสร้างข้อมูลหรือเนื้อหาของข้อมูล ตามที่ได้กล่าวมาแล้วตามข้อที่ 1.1 ว่าข้อมูลจะต้องมีโครงสร้างแบบตารางโดยมีการจัดรูปแบบโครงสร้างในแถวแรกเป็นชื่อคอลัมน์หรือชื่อตัวแปรและแถวถัดไปเป็นค่าของข้อมูล ทำให้ข้อมูลบางประเภท เช่น ข้อมูลสถิติ

ทางการ (United Nations, n.d.) ซึ่งเป็นข้อมูลที่ผ่านกระบวนการประมวลผลทางสถิติและมีการจัดรูปแบบโครงสร้างข้อมูลใหม่ ในรูปแบบตารางไขว้ (Crosstab table) หรือตารางไพวอท (Pivot table) ทำให้ข้อมูลมีการจัดเรียงโดยหมุนค่าข้อมูลเพื่อนำเสนอเป็น 2 มิติในแนวแกน x และ y ทำให้ข้อมูลไม่จัดอยู่ในรูปแบบที่มีโครงสร้างแบบตารางและไม่มีการจัดรูปแบบโครงสร้างให้แถวแรกเป็นชื่อคอลัมน์หรือชื่อตัวแปรและแถวถัดไปเป็นค่าของข้อมูล แต่ในแถวแรกจะประกอบไปด้วยชื่อตัวแปรและค่าข้อมูลปะปนกัน เมื่อคอมพิวเตอร์ประมวลผลจะทำให้ข้อมูลมีความผิดพลาดซึ่งจะเกิดการนำค่าข้อมูลมาเป็นชื่อตัวแปรตามภาพที่ 2 (ก) และส่งผลต่อการพิจารณาตัวแปรในการวิเคราะห์ข้อมูลและการเรียนรู้ของเครื่อง ดังนั้นการจัดรูปแบบข้อมูลจะต้องจัดให้อยู่ในรูปแบบข้อมูลระเบียนตามภาพที่ 2 (ข) โดยการทำการหมุนข้อมูลที่อยู่ในแกน x (คอลัมน์) ลงมาเป็นแกน y (แถว) หรือเรียกว่าอันไพวอท (Unpivot) เพื่อให้คอมพิวเตอร์สามารถประมวลผลได้โดยไม่ผิดความหมายตามโครงสร้างข้อมูลหรือเนื้อหาของข้อมูล

1.4 ข้อมูลมีความคงเส้นคงวาของประเภทข้อมูลและรูปแบบข้อมูล โดยข้อมูลที่มีการจัดเก็บอยู่ภายใต้ตัวแปรเดียวกันจะต้องจัดประเภทข้อมูล (Data type) และรูปแบบข้อมูล (Data format) ให้มีมาตรฐานเดียวกัน สำหรับประเภทข้อมูลต้องจัดเก็บค่าของข้อมูลให้สอดคล้องกับประเภทของข้อมูลในทุกระเบียน เช่น ตัวแปรจัดอยู่ในประเภทตัวเลขจะต้องจัดเก็บค่าของข้อมูลเป็นตัวเลขเท่านั้น ตัวแปรจัดอยู่ในประเภทวันที่จะต้องจัดเก็บค่าของข้อมูลเป็นวันที่เท่านั้น เนื่องด้วยค่าของข้อมูลที่มีการจัดเก็บแตกต่างไปจากประเภทที่กำหนดให้กับตัวแปร จะมีผลให้คอมพิวเตอร์ประมวลผลผิดพลาด รวมทั้งอาจส่งผลให้คอมพิวเตอร์ไม่สามารถประมวลผลข้อมูลได้ นำไปสู่การคัดทิ้งข้อมูลก่อนการประมวลผล สำหรับรูปแบบข้อมูลต้องจัดเก็บให้สอดคล้องกันในทุกรายการข้อมูล เช่น เพศ ควรมีการกำหนดมาตรฐานการจัดเก็บให้สอดคล้องกัน ได้แก่ ช/ญ, ชาย/หญิง, M/F, หรือ Male/Female (Cheong & Chang, 2007)

1.5 การตั้งชื่อตัวแปรควรให้สอดคล้องกับมาตรฐานการตั้งชื่อตัวแปร ควรตั้งชื่อโดยใช้ภาษาอังกฤษ เนื่องด้วยคอมพิวเตอร์สามารถประมวลผลได้โดยทันที ไม่ต้องผ่านกระบวนการแปลงข้อมูลให้อยู่ในรูปแบบภาษาคอมพิวเตอร์ก่อนประมวลผล ซึ่งโดยปกติแล้วชื่อตัวแปรในภาษาคอมพิวเตอร์จะเป็นภาษาอังกฤษและไม่ควรใช้อักขระพิเศษที่อาจซ้ำซ้อนกับอักขระที่เป็นไวยากรณ์ของภาษาคอมพิวเตอร์ เช่น $, #, @, หรือ & เป็นต้น นอกจากนี้ไม่ควรตั้งชื่อตัวแปรให้มีการเว้นวรรค เนื่องด้วยเมื่อมีการใช้ชื่อตัวแปรในการเขียนโปรแกรมหรือวิเคราะห์ข้อมูลด้วยโปรแกรมคอมพิวเตอร์อาจส่งผลต่อความสามารถหรือข้อจำกัดในการประมวลผลของโปรแกรมบางประเภท

1.6 ข้อมูลสามารถประมวลผลด้วยรหัสที่คอมพิวเตอร์สามารถประมวลผลได้โดยไม่ต้องผ่านกระบวนการแปลงให้อยู่ในรูปแบบรหัสคอมพิวเตอร์ เพื่อความสามารถในการประมวลผลอย่างมีประสิทธิภาพ ทั้ง

ด้านความเร็วในการประมวลผล การใช้ศักยภาพของเครื่องคอมพิวเตอร์ในการประมวลผล เช่น รหัสตัวเลขแทนตัวอักษรหรือข้อความ สำหรับความสามารถในการประมวลผลของคอมพิวเตอร์จะประมวลผลตัวเลขได้มีประสิทธิภาพกว่าตัวอักษร เนื่องด้วยคอมพิวเตอร์สามารถประมวลผลได้ทันทีโดยไม่ต้องผ่านกระบวนการแปลงการจัดเก็บข้อมูลใช้พื้นที่น้อยกว่าเมื่อเทียบกับตัวอักษร นอกจากนี้สนับสนุนการนำข้อมูลไปใช้ประโยชน์ในด้านการวิเคราะห์ได้และเป็นอัตโนมัติ โดยเฉพาะอย่างยิ่งการวิเคราะห์ข้อมูลโดยใช้หลักการเรียนรู้ของเครื่องจักร และรองรับความสามารถในการทำงานได้ของอัลกอริทึมที่ใช้ในการวิเคราะห์ข้อมูล

2. ระดับข้อมูลที่เครื่องคอมพิวเตอร์สามารถประมวลผลได้

จากการวิเคราะห์คุณสมบัติของข้อมูลที่เครื่องคอมพิวเตอร์สามารถประมวลผลได้ สามารถจัดระดับของข้อมูลได้เป็น 4 ระดับ โดยกำหนดระดับของข้อมูลตามรูปแบบความสามารถในการนำไปใช้ประโยชน์ต่อได้ แสดงรายละเอียดได้ดังนี้

1) ขั้นต้น (Basic level) มีรายละเอียดคือ (1) ข้อมูลบันทึกอยู่ในรูปแบบไฟล์ที่คอมพิวเตอร์สามารถอ่านและนำไปประมวลผลได้แบบอัตโนมัติ (2) ข้อมูลมีลักษณะเป็นโครงสร้างหรือกึ่งโครงสร้างโดยข้อมูลจะต้องมีการจัดเรียงแบบตารางที่มีรูปแบบชัดเจนและเป็นระเบียบ (3) ข้อมูลจะต้องไม่จำกัดสิทธิ์ และ (4) การเข้ารหัสข้อมูลอยู่ในรูปแบบที่เป็นสากลสามารถใช้งานได้กับทุกภาษา สำหรับข้อมูลที่จัดอยู่ในระดับขั้นต้นเป็นข้อมูลที่เครื่องคอมพิวเตอร์สามารถประมวลผลได้เบื้องต้น โดยข้อมูลจะมีคุณลักษณะที่สามารถเปิดและใช้งานได้อย่างอิสระ ไม่จำกัดการเข้าถึงและเปิดไฟล์ แต่อาจมีความผิดพลาดของข้อมูลในเชิงความหมายและโครงสร้างข้อมูล ซึ่งอาจต้องผ่านกระบวนการตรวจสอบโดยมนุษย์ก่อนการนำข้อมูลไปใช้ประโยชน์ต่อไป ข้อมูลในระดับขั้นต้นเป็นข้อมูลที่สอดคล้องกับคุณสมบัติของการเปิดเผยข้อมูล ซึ่งสอดคล้องกับมาตรฐานการเปิดเผยข้อมูลระดับ 3 ดาว (Berners-Lee, 2012; Digital Government Development Agency, 2020)

2) ขั้นกลาง (Medium level) มีรายละเอียดคือ ข้อมูลผ่านเกณฑ์ระดับขั้นต้น และ (1) ข้อมูลที่จัดเก็บอยู่ภายในไฟล์จะต้องไม่มีการแปลงหรือเปลี่ยนโครงสร้างของข้อมูลให้การประมวลผลผิดความหมายตามโครงสร้างข้อมูลหรือเนื้อหาของข้อมูล หรือไม่มีการจัดรูปแบบโครงสร้างข้อมูลใหม่ที่อยู่ในรูปแบบตารางไขว้หรือตารางไพวอท สำหรับข้อมูลที่จัดอยู่ในระดับขั้นกลางเป็นข้อมูลที่เครื่องคอมพิวเตอร์สามารถประมวลผลได้ในเชิงความหมายและโครงสร้างข้อมูล รวมทั้งมีความถูกต้องในด้านประเภทของข้อมูลและรูปแบบข้อมูลที่รองรับการนำข้อมูลไปวิเคราะห์และนำเข้าโปรแกรมคอมพิวเตอร์ สามารถนำไปใช้กับโปรแกรมวิเคราะห์เชิงภาพได้อย่างมีประสิทธิภาพ และสามารถนำเข้าจัดเก็บในฐานข้อมูลได้อย่างถูกต้อง

ตารางที่ 1 แสดงคุณสมบัติข้อมูลที่เครื่องคอมพิวเตอร์สามารถประมวลผลได้

รายละเอียด	ระดับของข้อมูลที่เครื่องคอมพิวเตอร์สามารถประมวลผลได้			
	ขั้นต้น	ขั้นกลาง	ขั้นสูง	ขั้นเสริม
ข้อมูลบันทึกอยู่ในรูปแบบไฟล์ที่คอมพิวเตอร์สามารถอ่านและนำไปประมวลผลได้แบบอัตโนมัติ (Automatically reading and processing)	✓	✓	✓	✓
ข้อมูลจะต้องมีคุณลักษณะที่มีโครงสร้างแบบตารางในลักษณะคอลัมน์และแถว ซึ่งค่าข้อมูลบรรจุอยู่เพียงช่องเดียว (Structured data)	✓	✓	✓	✓
ข้อมูลจะต้องไม่จำกัดสิทธิ์และการเข้ารหัสแบบเป็นสากลและสามารถใช้งานได้กับทุกภาษา (Non-proprietary and accessible)	✓	✓	✓	✓
ข้อมูลที่ได้จากการประมวลผลจะต้องไม่ผิดความหมายตามโครงสร้างข้อมูลหรือเนื้อหาของข้อมูล (No semantic meaning is lost)		✓	✓	✓
ข้อมูลมีความคงเส้นคงวาของประเภทข้อมูลและรูปแบบข้อมูล (Consistency)			✓	✓
การตั้งชื่อตัวแปรควรให้สอดคล้องกับมาตรฐานการตั้งชื่อตัวแปร (Naming convention)			✓	✓
ข้อมูลสามารถประมวลผลด้วยรหัสที่คอมพิวเตอร์อ่านง่ายเพื่อความสามารถในการประมวลผลอย่างมีประสิทธิภาพ (Easy interrogation by computer code)				✓

3) ขั้นสูง (Advanced level) มีรายละเอียดคือ ข้อมูลผ่านเกณฑ์ระดับขั้นกลาง และ (1) การจัดรูปแบบข้อมูลให้มีความคงเส้นคงวาของประเภทและรูปแบบข้อมูล ซึ่งจะมีผลให้คอมพิวเตอร์สามารถประมวลผลได้อย่างถูกต้อง ลดการการคัดทิ้งข้อมูลที่ไม่สอดคล้องก่อนการนำเข้าโปรแกรมหรือฐานข้อมูลและการนำไปวิเคราะห์ และ (2) มีการตั้งชื่อตัวแปรสอดคล้องกับมาตรฐานการตั้งชื่อ ซึ่งการตั้งชื่อตามมาตรฐานการตั้งชื่อจะเพิ่มขีดความสามารถในการประมวลผลของคอมพิวเตอร์ ซึ่งคอมพิวเตอร์สามารถประมวลผลได้ทันทีโดยไม่ต้องผ่านกระบวนการแปลงข้อมูลให้อยู่ในรูปแบบภาษาคอมพิวเตอร์ รวมทั้งทำให้สมรรถนะในการประมวลผล (Capacity) ที่มีประสิทธิภาพมากขึ้นทั้งด้านความเร็วและการใช้ทรัพยากรคอมพิวเตอร์ในการประมวลที่น้อยลง สำหรับข้อมูลที่จัดอยู่ในระดับขั้นสูงเป็นข้อมูลที่มีความพร้อมในการนำไปพัฒนาต่อยอดในการสร้างเอพีไอและ

การนำเข้าฐานข้อมูลและโปรแกรมคอมพิวเตอร์ และรองรับการเขียนโปรแกรมเพื่อเชื่อมโยงข้อมูลมายังระบบหรือแอปพลิชันได้อย่างอัตโนมัติ

4) ขั้นเสริม (Optional level) มีรายละเอียดคือ ข้อมูลผ่านเกณฑ์ขั้นสูง และ (1) ข้อมูลสามารถประมวลผลด้วยรหัสที่คอมพิวเตอร์สามารถประมวลผลได้โดยไม่ต้องผ่านกระบวนการแปลงให้อยู่ในรูปแบบรหัสคอมพิวเตอร์ เพื่อความสามารถในการประมวลผลอย่างมีประสิทธิภาพทั้งด้านความเร็วในการประมวลผล การใช้ศักยภาพของเครื่องคอมพิวเตอร์ในการประมวลผล สนับสนุนการนำข้อมูลไปใช้ประโยชน์ในด้านการวิเคราะห์โดยใช้หลักการเรียนรู้ของเครื่องจักร ข้อมูลในระดับขั้นเสริมเป็นการเตรียมข้อมูลให้อยู่ในรูปแบบที่พร้อมสำหรับการวิเคราะห์ข้อมูล โดยใช้สถิติหรือการเรียนรู้ของเครื่องจักร ซึ่งจำเป็นต้องแปลงข้อมูลให้อยู่ในรูปแบบตัวเลขหรือระดับการวัดของตัวแปรในระดับนามมาตรา (Nominal) หรืออันดับมาตรา (Ordinal)

3. ผลการสำรวจและวัดผลชุดข้อมูลเปิดตามกรอบแนวคิด

ผู้วิจัยสำรวจและวัดผลข้อมูลในกลุ่มข้อมูลด้านเศรษฐกิจและการเงิน (Economy and finance) จำนวน 1,728 ชุดข้อมูลจาก 4 ประเทศ แบ่งเป็น ไต้หวัน 985 ชุดข้อมูล (ร้อยละ 57.00) ญี่ปุ่น 162 ชุดข้อมูล (ร้อยละ 9.38) สิงคโปร์ 286 ชุดข้อมูล (ร้อยละ 16.55) และไทย 295 ชุดข้อมูล (ร้อยละ 17.07) โดยการนับจำนวนชุดข้อมูลทั้งหมดจากระบบข้อมูลเปิดของเอเชีย (คอลัมน์ที่ 2,3) ผู้วิจัยกำหนดขนาดตัวอย่างในการสำรวจอย่างน้อย 317 รายการอ้างอิงจากขนาดตัวอย่างในตารางของเครจซี่และมอร์แกน (Krejcie & Morgan) ซึ่งทำการสุ่มชุดข้อมูลแบบอย่างง่ายจากแต่ละประเทศตามจำนวนที่แสดงดังตารางที่ 2 โดยกำหนดการสุ่มเลือกจากชุดข้อมูลที่มีการเผยแพร่หรือแก้ไขล่าสุด ผู้วิจัยเก็บข้อมูลจำนวน 500 ชุดข้อมูล คิดเป็นร้อยละ 28.94 ของชุดข้อมูลทั้งหมด และคำนวณสัดส่วนจำนวนชุดข้อมูลตัวอย่างของแต่ละประเทศที่ใช้ในการวิจัย (คอลัมน์ที่ 4) และผู้วิจัยนำเสนอจำนวนชุดข้อมูลที่ทำการสำรวจที่ศึกษาจริงจากระบบข้อมูลเปิดเอเชียของแต่ละประเทศ (คอลัมน์ที่ 5) รวมทั้งคำนวณสัดส่วนของจำนวนชุดข้อมูลที่ศึกษากับจำนวนชุดข้อมูลทั้งหมดในแต่ละประเทศ (คอลัมน์ที่ 6) จากผลการสำรวจชุดข้อมูลและวัดผลชุดข้อมูลเปิดตามกรอบแนวคิดที่ผู้วิจัยนำเสนอ ดังแสดงตามตารางที่ 3 พบว่าชุดข้อมูลเปิดมีระดับการเปิดเผยข้อมูลในรูปแบบข้อมูลที่เครื่องคอมพิวเตอร์สามารถประมวลผลได้มากที่สุดอยู่ในขั้นกลาง คิดเป็นร้อยละ 33.00 ของชุดข้อมูลทั้งหมด โดยไต้หวันมีจุดเด่นในการเปิดเผยข้อมูลได้ในขั้นกลางและขั้นสูงตามลำดับ (ร้อยละ 68.00 และ 31.50 ตามลำดับ) สิงคโปร์มีจุดเด่นในการเปิดเผยข้อมูลได้มากที่สุดในขั้นกลาง (ร้อยละ 96.00) สำหรับประเทศไทยมีระดับการเปิดเผยข้อมูลมากที่สุดในขั้นกลางและขั้นต้นตามลำดับ (ร้อยละ 28.00 และ 27.00 ตามลำดับ) แต่สำหรับญี่ปุ่นมีการเปิดเผยชุดข้อมูลที่ยังไม่เป็นข้อมูลที่เครื่องคอมพิวเตอร์สามารถประมวลผลได้ เมื่อพิจารณาคะแนนการเปิดเผยข้อมูลในรูปแบบที่เครื่องคอมพิวเตอร์สามารถประมวลผลได้ โดยกำหนดให้ไม่เป็นข้อมูลที่อ่านได้ด้วยเครื่องเท่ากับ 0 คะแนน ขั้นต้นเท่ากับ 1 คะแนน ขั้นกลางเท่ากับ 2

คะแนน ขั้นสูงเท่ากับ 3 คะแนน และขั้นเสริมเท่ากับ 4 คะแนน พบว่าประเทศที่มีคะแนนสูงสุดคือไต้หวัน (230.50 คะแนน) ตามด้วยสิงคโปร์และไทย (215.00 และ 83.00 คะแนน ตามลำดับ)

ตารางที่ 2 แสดงสัดส่วนของขนาดตัวอย่างที่ใช้ในการศึกษา

ข้อมูลเปิด	จำนวนชุดข้อมูลทั้งหมด	สัดส่วนจำนวนชุดข้อมูลทั้งหมด (%)	สัดส่วนจำนวนชุดข้อมูลตัวอย่าง	จำนวนชุดข้อมูลที่ทำการสำรวจ	สัดส่วนของจำนวนตัวอย่างที่ทำการสำรวจกับจำนวนชุดข้อมูลทั้งหมด (%)
ไต้หวัน	985	57.00	181	200	20.30
ญี่ปุ่น	162	9.38	30	100	61.73
สิงคโปร์	286	16.55	52	100	34.97
ไทย	295	17.07	54	100	33.90
รวม	1728	100.00	317	500	28.94

ตารางที่ 3 แสดงผลการสำรวจและวัดผลชุดข้อมูลเปิดตามกรอบแนวคิด

ข้อมูลเปิด/ จำนวนชุดข้อมูล และร้อยละ		ไม่เป็นข้อมูลที่อ่านได้ด้วยเครื่อง (0)	ขั้นต้น (1)	ขั้นกลาง (2)	ขั้นสูง (3)	ขั้นเสริม (4)	รวม	คะแนน
ไต้หวัน	จำนวน	1	0	**136**	**63**	0	200	-
	ร้อยละ	0.50	0.00	**68.00**	**31.50**	0.00	100.00	**230.50**
ญี่ปุ่น	จำนวน	**100**	0	0	0	0	100	-
	ร้อยละ	**100.00**	0.00	0.00	0.00	0.00	100.00	0.00
สิงคโปร์	จำนวน	0	0	**95**	5	0	100	-
	ร้อยละ	0.00	0.00	**95.00**	5.00	0.00	100.00	215.00
ไทย	จำนวน	40	**27**	**28**	5	0	100	-
	ร้อยละ	40.00	**27.00**	**28.00**	5.00	0.00	100.00	83.00
รวม	จำนวน	**141**	27	**165**	67	0	500	-
	ร้อยละ	**28.20**	5.40	**33.00**	13.40	0	100.00	-

สรุปผลการศึกษา

สำหรับคุณสมบัติของข้อมูลที่เครื่องคอมพิวเตอร์สามารถประมวลผลได้ที่ผู้วิจัยได้ทำการศึกษาจากแนวทางการเปิดเผยและนโยบายการเปิดเผยข้อมูลในประเทศต่าง ๆ รวมทั้งงานวิจัยที่เกี่ยวข้องเพื่อประโยชน์ในการประยุกต์ใช้และส่งเสริมการเปิดเผยข้อมูลภาครัฐในประเทศไทย ผลการวิจัยพบว่าคุณสมบัติที่สำคัญประกอบด้วย 6 ประเด็น คือ 1) ข้อมูลบันทึกอยู่ในรูปแบบไฟล์ที่คอมพิวเตอร์สามารถอ่านและนำไปประมวลผลได้แบบอัตโนมัติ และจะต้องมีโครงสร้างแบบตารางในลักษณะคอลัมน์และแถว ซึ่งค่าข้อมูลบรรจุอยู่เพียงช่องเดียว 2) ข้อมูลจะต้องไม่จำกัดสิทธิ์และมีการเข้ารหัสแบบเป็นสากลที่สามารถใช้งานได้กับทุกภาษา 3) ข้อมูลที่ได้จากการประมวลผลจะต้องไม่ผิดความหมายตามโครงสร้างข้อมูลหรือเนื้อหาของข้อมูล 4) ข้อมูลมีความคงเส้นคงวาของประเภทข้อมูลและรูปแบบข้อมูล 5) การตั้งชื่อตัวแปรควรให้สอดคล้องกับมาตรฐานการตั้งชื่อตัวแปร และ 6) ข้อมูลสามารถประมวลผลด้วยรหัสที่คอมพิวเตอร์อ่านได้ง่ายเพื่อความสามารถในการประมวลผลอย่างมีประสิทธิภาพ นอกจากนั้นผู้วิจัยได้ทำการจัดระดับของข้อมูลตามคุณสมบัติของชุดข้อมูล โดยใช้หลักการจัดกลุ่มตามประสิทธิภาพของความสามารถในการนำข้อมูลไปใช้ประโยชน์ต่อได้ ซึ่งเป็นคุณลักษณะของข้อมูลที่เครื่องคอมพิวเตอร์สามารถประมวลผลได้ในข้อมูลที่มีคุณลักษณะเป็นโครงสร้าง ซึ่งเป็นการกำหนดคุณลักษณะของข้อมูลเพื่อให้สนับสนุนการเปิดเผยข้อมูลและการวิเคราะห์ข้อมูลได้อย่างมีประสิทธิภาพและประสิทธิผล ประกอบด้วย 4 ระดับ คือ ขั้นต้น ขั้นกลาง ขั้นสูง และขั้นเสริม โดยเมื่อมีการจัดทำข้อมูลเปิดเพื่อเผยแพร่สู่สาธารณะในรูปแบบข้อมูลที่เครื่องคอมพิวเตอร์สามารถประมวลผลได้จะทำให้ข้อมูลดังกล่าวมีคุณสมบัติสอดคล้องกับคุณลักษณะของข้อมูลเปิดใน 5 ประเด็นจากบทความเรื่อง 8 ประการหลักของคุณสมบัติข้อมูลเปิดของ Tauberer (2019) กล่าวคือ (1) ข้อมูลอยู่ในรูปแบบปฐมภูมิ (Primary) โดยข้อมูลมีการจัดรูปแบบในลักษณะข้อมูลรายการ ไม่มีการทำเป็นข้อมูลสรุป หากมีการประมวลผลของข้อมูลในรูปแบบข้อมูลจะสรุป (Aggregation) ก็จะต้องจัดอยู่ในรูปแบบข้อมูลรายการ (2) ข้อมูลสามารถเข้าถึงได้ง่าย (Accessible) สามารถค้นหา เข้าถึง และใช้งานชุดข้อมูลได้หลายช่องทาง (3) ข้อมูลอ่านได้ด้วยเครื่อง ซึ่งต้องมีโครงสร้าง สามารถอ่านได้ด้วยเครื่อง และนำข้อมูลไปใช้งานต่อได้ (4) ข้อมูลต้องไม่เลือกปฏิบัติ (Non-discriminatory) โดยต้องสามารถนำข้อมูลไปใช้ได้อย่างอิสระ โดยไม่ต้องระบุตัวตน หรือเหตุผลของการนำไปใช้งาน และ (5) ข้อมูลไม่จำกัดสิทธิ โดยต้องอยู่ในรูปแบบมาตรฐานเปิดที่สามารถใช้ได้หลายแพลตฟอร์ม และต้องไม่ถือครองกรรมสิทธิ์หลังจากนำข้อมูลเปิดไปใช้ประโยชน์ (Tauberer, 2019)

ผลการสำรวจและวัดผลชุดข้อมูลเปิดตามกรอบแนวคิดพบว่าชุดข้อมูลเปิดมีระดับการเปิดเผยข้อมูลในรูปแบบข้อมูลที่เครื่องคอมพิวเตอร์สามารถประมวลผลได้มากที่สุดอยู่ในขั้นกลาง ซึ่งเป็นข้อมูลที่อยู่ในรูปแบบมาตรฐานของการเปิดเผยข้อมูล นอกจากนี้เมื่อพิจารณาข้อมูลเปิดที่มีคะแนนสูงสุดในการเปิดเผยข้อมูลในรูปแบบ

ที่เครื่องคอมพิวเตอร์สามารถประมวลผลได้คือข้อมูลเปิดของไต้หวัน ซึ่งเป็นองค์กรที่มีการผลักดันในการเปิดเผยข้อมูล และมีการดำเนินการเปิดเผยข้อมูลได้อย่างมีประสิทธิภาพ โดยมีคะแนนตัวชี้วัดการเปิดเผยข้อมูลสากลเท่ากับร้อยละ 90 จัดเป็นอันดับ 1 ของเครือข่ายข้อมูลเปิดในทวีปเอเชียที่เข้าร่วมโครงการ (Global Open Data Index, 2017) สำหรับข้อมูลเปิดของไทยมีระดับการเปิดเผยข้อมูลในรูปแบบที่เครื่องคอมพิวเตอร์สามารถประมวลผลได้ประมาณร้อยละ 60 ซึ่งยังมีข้อมูลอีกร้อยละ 40 ที่จะต้องดำเนินการปรับปรุงให้มีคุณสมบัติตามเกณฑ์ นอกจากนี้เมื่อพิจารณาข้อมูลที่เครื่องคอมพิวเตอร์สามารถประมวลผลได้พบว่าส่วนใหญ่อยู่ในขั้นต้นและขั้นกลาง เนื่องด้วยตามเกณฑ์การเปิดเผยข้อมูลที่กำหนดโดยประกาศคณะกรรมการพัฒนารัฐบาลดิจิทัล เรื่อง มาตรฐานและหลักเกณฑ์การเปิดเผยข้อมูลเปิดภาครัฐในรูปแบบข้อมูลดิจิทัลต่อสาธารณะ กำหนดให้หน่วยงานรัฐเปิดเผยข้อมูลอย่างน้อยระดับ 3 ดาวขึ้นไปจากระดับ 5 ดาว (Digital Government Development Agency, 2020) ซึ่งเกณฑ์ดังกล่าวมีคุณลักษณะของความสามารถที่เครื่องคอมพิวเตอร์สามารถประมวลผลได้ในขั้นต้นเท่านั้น ดังนั้นแล้วจึงควรมีการผลักดันมีหน่วยงานรัฐเพิ่มระดับของการเปิดเผยข้อมูลให้อยู่ในระดับขั้นกลางขึ้นไป ซึ่งจะทัดเทียมกับประเทศมีจุดเด่นของการเปิดเผยข้อมูล ได้แก่ ข้อมูลเปิดของไต้หวันและสิงคโปร์ซึ่งมีระดับการเปิดเผยข้อมูลอยู่ในขั้นกลางขึ้นไป สำหรับการผลักดันให้ระดับการเปิดเผยข้อมูลให้ได้ในระดับขั้นกลางขึ้นไปจะต้องคำนึงถึงคุณลักษณะของข้อมูลเปิดภาครัฐ ที่ควรมีการจัดรูปแบบอยู่ในลักษณะข้อมูลรายการ ไม่มีการปรับแต่ง หรืออยู่ในรูปแบบข้อมูลสรุป

นอกจากนี้เมื่อพิจารณาจากผลการสำรวจและวัดผลข้อมูลพบว่าประเทศไทยยังคงมีระดับการเปิดเผยข้อมูลในขั้นสูงที่ค่อนข้างน้อยและขั้นเสริมยังไม่มีการดำเนินการ ซึ่งการเปิดเผยข้อมูลในระดับขั้นสูงและขั้นเสริมยังเป็นระดับที่ควรมีการส่งเสริมให้ชุดข้อมูลมีรูปแบบอยู่ในระดับดังกล่าวที่เพิ่มมากขึ้น เนื่องด้วยปัจจุบันเทคโนโลยีการวิเคราะห์ข้อมูลสำหรับข้อมูลขนาดใหญ่ (Big data) จำเป็นต้องใช้เครื่องคอมพิวเตอร์ในการประมวลผลและวิเคราะห์ข้อมูลเพื่อค้นหาและสกัดความรู้จากข้อมูลขนาดใหญ่ ดังนั้นการเสริมข้อมูลให้สนับสนุนการทำงานของเครื่องคอมพิวเตอร์จึงเป็นปัจจัยสำคัญที่จะช่วยส่งเสริมให้ข้อมูลมีการนำไปใช้ได้อย่างมีประสิทธิภาพมากยิ่งขึ้น สำหรับขั้นสูงจะเน้นให้มีการเตรียมข้อมูลให้สามารถใช้งานได้ผ่านโปรแกรมคอมพิวเตอร์และฐานข้อมูล และขั้นเสริมจะเน้นให้มีการเตรียมข้อมูลให้สามารถรองรับการวิเคราะห์ข้อมูลด้วยวิธีการเรียนรู้ของเครื่องจักร

กิตติกรรมประกาศ

ผู้วิจัยขอขอบคุณกลุ่มงานเดตาโซลูชันส์ภาครัฐ ฝ่ายเดตาโซลูชันส์ สำนักงานพัฒนารัฐบาลดิจิทัล (องค์การมหาชน) สำหรับคำแนะนำ สนับสนุน และชี้แนะแนวทางของการศึกษาชุดข้อมูลที่เครื่องคอมพิวเตอร์สามารถประมวลผลได้สำหรับการนำไปใช้ในการเปิดเผยข้อมูลและการวิเคราะห์ข้อมูลของภาครัฐ

เอกสารอ้างอิง

Asia Open Data Partnership. (2015). **Dataportal.asia Data Explorer in Asia**. Retrieved 1 December 2020, from https://dataportal.asia/home

Berners-Lee, T. (2012). **5-Star Deployment Scheme for Open Data**. Retrieved 16 March 2021, from https://5stardata.info/en/

Brickley, D., Burgess, M., & Noy, N. (2019). **Google Dataset Search: Building a search engine for datasets in an open Web ecosystem**. Paper presented at WWW '19: The World Wide Web Conference, May 2019 Pages 1365–1375 DOI:rg/10.1145/3308558.3313685

California Open Data Portal. (n.d.). **Open Data Glossary**. Retrieved 30 November 2020, from https://data.ca.gov/pages/open-data-glossary

Cheong, L. K., & Chang, V. J. A. P. (2007). **The need for data governance: a case study**. 8th Australasian Conference on Information System 5-7 Dec 2007, Toowoomba,999

Digital Government Development Agency. (2020).**Digital Government Standard in Open Government Data Guideline**. (In Thai). DGS-12001:2563 C.F.R.

Estonian Open Government Data Portal. (n.d.). **Why is it important to make data openly available?** Retrieved 1 December 2020, from https://opendata.riik.ee/en/about/

EU Open Data Portal. (2013). **EU implementation of the G8 Open Data Charter.** Retrieved 30 November 2020, from http://ec.europa.eu/information_society/newsroom/cf/dae/document.cfm?doc_id=3489

European Securities and Markets Authority. (2017). **New rules make EU issuers' annual financial reports machine-readabletest**. Retrieved 1 December 2020, from https://www.esma. europa.eu/press-news/esma-news/new-rules-make-eu-issuers%E2%80%99-annual-financial-reports-machine-readable

Fecher, B., & Friesike, S. (2014). **Open science: one term, five schools of thought**. In Opening science (17-47): Springer, Cham. DOI: 10.1007/978-3-319-00026-8_2

Global Open Data Index. (2017). **The Global Open Data Index provides the most comprehensive snapshot available of the state of open government data publication**. Retrieved 17 December 2020, from https://index.okfn.org/place/

Goëta, S., & Davies, T. (2019). **The Daily Shaping of State Transparency: Standards, Machine-Readability and the Configuration of Open Government Data Policies**. Science & Technology Studies, **29**(4), 9-30.

Tauberer, J. (2019). **The 8 Principles of Open Government Data.** Retrieved 15 December 2020, from https://opengovdata.org

Office of Management and Budget. (2019). **Circular No A-11 Preparation, Submission and Execution of the budget**. Retrieved 30 November 2020, from https://static.carahsoft.com /concrete/files/8415/7305/7362/a11_-_section_280_OMB_report.pdf

Open Knowledge Foundation. (n.d.). **Machine readable**. Retrieved 1 December 2020, from http://opendatahandbook.org/glossary/en/terms/machine-readable/

Stevens, H. (2019). **Open data, closed government: Unpacking data.gov.sg.** First Monday, 24(4). doi:https://doi.org/10.5210/fm.v24i4.9851

United Nations. (2013). **Guidelines on open government data for citizen engagement**. Retrieved 15 December 2020, from https://publicadministration.un.org/en/ogd

United Nations. (n.d.). **Economic and development statistics.** Retrieved 7 December 2020, from http://www.un.org/esa/progareas/stats.html

United State Government Publishing Office. (2018). **Foundations for Evidence-Based Policymaking Act of 2018**. Retrieved 30 November 2020, from https://stratml.us/ references/HR4174. htm#Agency_Responsibilities

Western Australian Land Information Authority. (2020). **The principles of open and machine-readable data.** Retrieved 30 November 2020, from https://data.wa.gov.au/junk/fact-sheets-and-toolkit /creating-machine-readable-data

Yergeau, F. (2003). **RFC 3629 - UTF-8, a transformation format of ISO 10646**. Retrieved 30 November 2020, from https://www.hjp.at/doc/rfc/rfc3629.html