

# แนวปฏิบัติในการจัดเก็บชุดข้อมูล (Dataset) ของ คลังปัญญา สถาบันบัณฑิตพัฒนบริหารศาสตร์

## แนวปฏิบัติการจัดเก็บและเผยแพร่ข้อมูลเปิด (Open Data)

### สำหรับคลังปัญญา สถาบันบัณฑิตพัฒนบริหารศาสตร์

#### วัตถุประสงค์

แนวปฏิบัติการจัดเก็บและเผยแพร่ข้อมูลเปิด (Open Data) สำหรับคลังปัญญา สถาบันบัณฑิตพัฒนบริหารศาสตร์ จัดทำขึ้นเพื่อใช้เป็นแนวปฏิบัติในการดำเนินการเกี่ยวกับการเปิดเผยข้อมูลประกอบการทำวิจัยของนักศึกษาศาสนาบัณฑิตพัฒนบริหารศาสตร์ให้เป็นไปในทิศทางเดียวกัน รวมถึงการกำหนดลักษณะของชุดข้อมูล (Datasets) ที่เหมาะสมจะนำมาจัดเก็บและเผยแพร่ ตลอดจนรูปแบบของชุดข้อมูลและวิธีการเผยแพร่ โดยยึดหลักหัวใจสำคัญของ Open Data ตามสำนักงานพัฒนารัฐบาลดิจิทัล (องค์การมหาชน) ได้แก่ 1) Availability and Access ข้อมูลที่เปิดเผยต้องสามารถใช้งานได้ทั้งหมดและการเผยแพร่อยู่ในรูปแบบที่ง่ายต่อการใช้งานและสามารถแก้ไขได้ สามารถดาวน์โหลดฟรีผ่านเว็บไซต์คลังปัญญา 2) Re-use and Redistribution ข้อมูลที่เผยแพร่ต้องถูกจัดเตรียมภายใต้เงื่อนไข การอนุญาตให้นำมาใช้ใหม่ และเผยแพร่ได้ รวมถึงการใช้ชุดข้อมูลร่วมกับชุดข้อมูลอื่น ๆ 3) Universal Participation คือ ทุกคนสามารถที่จะใช้ข้อมูลได้ ไม่ว่าจะเป็นการนำไปใช้ การใช้ซ้ำ การเผยแพร่ โดยไม่ติดเรื่องข้อจำกัดใด ๆ เพื่อให้สนับสนุนให้เกิดการทำงานร่วมกัน (interoperability) อย่างเป็นรูปธรรมสามารถนำข้อมูลไปพัฒนาต่อยอดงานวิจัยและนวัตกรรมในด้านต่าง ๆ ของประเทศต่อไป

#### คำนิยาม

**ข้อมูล** สิ่งที่สื่อความหมายให้รู้เรื่องราวข้อเท็จจริงหรือเรื่องอื่นใด ไม่ว่าจะการสื่อความหมายนั้นจะทำได้โดยสภาพของสิ่งนั้นเองหรือโดยผ่านวิธีการใด ๆ และไม่ว่าจะได้จัดทำไว้ในรูปของเอกสาร แฟ้ม รายงาน หนังสือ แผ่นผัง แผนที่ ภาพวาด ภาพถ่าย ภาพถ่ายดาวเทียม ภาพยนตร์ การบันทึกภาพหรือเสียง การบันทึกโดยเครื่องคอมพิวเตอร์ เครื่องมือตรวจวัด การสำรวจระยะไกล หรือวิธีอื่นใดที่ทำให้สิ่งที่บันทึกไว้ปรากฏได้

**ข้อมูลเปิด** ข้อมูลที่ทุกคนสามารถนำไปใช้ได้โดยอิสระ เช่น การนำไปใช้ การนำกลับมาใช้ใหม่ หรือนำไป

**(Open Data)** เผยแพร่ได้โดยใครก็ตาม ซึ่งข้อมูลทั้งหมดที่อยู่ในคลังปัญญานี้ เป็นข้อมูลเปิดจากการค้นคว้า วิจัยของประชาคมในสถาบันบัณฑิตพัฒนบริหารศาสตร์

**ชุดข้อมูล** ข้อมูลที่ถูกนำมาจัดเป็นชุดให้ถูกต้องตามลักษณะโครงสร้างข้อมูลที่กำหนดไว้

### **(Datasets)**

**ลักษณะของชุดข้อมูล (Datasets) ที่นำมาจัดเก็บและเผยแพร่ควรมีลักษณะดังนี้**

1) รูปแบบชุดข้อมูลควรอยู่ในรูปแบบที่สามารถนำไปใช้ประมวลผลด้วยคอมพิวเตอร์ต่อได้เลย โดยไม่ต้องการจัดเตรียมข้อมูลใด ๆ (Machine Readable) เช่น CSV, XLS, XLSX, XML, JSON เพื่อให้ข้อมูลที่เผยแพร่ได้มีการนำไปใช้ประโยชน์ได้อย่างเต็มที่

2) เนื้อหาข้อมูลควรมีลักษณะเป็นตัวเลขสถิติ ตัวอย่างเช่น ข้อมูลปริมาณขยะมูลฝอย ปริมาณน้ำฝนรายเดือน เป็นต้น ซึ่งสามารถสังเกตได้ว่าข้อมูลที่ยกตัวอย่างจะมีการเก็บข้อมูลในลักษณะตาราง หรือเป็นข้อมูลที่เป็น Transaction หรือเป็นรายการ ซึ่งเป็นชุดข้อมูลที่เหมาะสมสำหรับการนำไปใช้ประโยชน์ต่อได้อย่างเสรี

3) การเปิดเผยข้อมูลที่มีลักษณะเป็น URL ควรเป็น URL ที่ลิงค์ตรงไปยังข้อมูล ไม่ควรเป็นเว็บไซต์ของหน่วยงานหรือ Facebook Page เนื่องจากไม่ได้เป็นข้อมูลในรูปแบบที่เหมาะสมในการเผยแพร่บนระบบ

ทั้งนี้ชุดข้อมูลที่เผยแพร่ควรคำนึงถึงการนำไปใช้ประโยชน์เป็นสำคัญ และเนื้อหาข้อมูลต้องไม่มีขัดต่อกฎหมาย ระเบียบ ข้อบังคับต่าง ๆ เช่น มีข้อมูลส่วนบุคคลอันขัดต่อพระราชบัญญัติคุ้มครองข้อมูลส่วนบุคคล (PDPA) เป็นต้น

## รูปแบบมาตรฐานของชุดข้อมูล (Datasets) ที่รองรับได้แก่

1) CSV: Comma-separated Values คือ Text File สำหรับเก็บข้อมูลแบบตาราง โดยใช้จุลภาค (,) แบ่งข้อมูลในแต่ละหลัก (Column) และใช้การเว้นบรรทัดแทนการแบ่งแถว (Row)

2) XLS: (ไฟล์เอกสาร Microsoft Excel) คือ ไฟล์ประเภทสเปรดชีต (Spreadsheet) หรือตารางคำนวณอิเล็กทรอนิกส์ ที่ใช้ในการเก็บบันทึกข้อมูลในลักษณะต่าง ๆ ซึ่งส่วนใหญ่จะเก็บข้อมูลประเภทคำนวณ โดยเก็บข้อมูลลงในตารางสี่เหลี่ยมที่เรียกว่า เซล (Cell) ที่สามารถนำเอาเซลมาอ้างอิงใส่ในสูตรเพื่อให้โปรแกรมคำนวณหาผลลัพธ์จากข้อมูลที่บันทึกไว้ได้

3) PDF: Portable Document Format (PDF) คือไฟล์ที่ถูกสร้างจากโปรแกรม Adobe Acrobat หรือโปรแกรมประเภท PDF Creator เหมาะสำหรับการทำเอกสารตัวอย่าง, ใบเสนอราคา, Manual, Sample Picture หรือเอกสารอื่น ๆ ซึ่งปัจจุบันเป็นที่นิยมอย่างมากในการทำ e-Document หรือ e-Paper ต่าง ๆ เนื่องจากไฟล์ที่ได้นั้นจะมีคุณภาพสูง ไม่ผิดเพี้ยนจากต้นฉบับ และผู้ที่นำไฟล์ PDF ไปใช้งานไม่สามารถแก้ไขต้นฉบับได้ เนื่องจากเอกสารลักษณะนี้ มีรูปแบบ, ขนาดไฟล์ไม่ใหญ่มาก และสามารถทำงานข้ามระบบ (Cross Platform) ได้ทำให้เอกสารประเภท PDF เป็นอิสระจากซอฟต์แวร์ ฮาร์ดแวร์ และระบบปฏิบัติการ (OS) สามารถสร้างได้ทั้งจากเครื่องคอมพิวเตอร์ระบบ Macintosh และ PC Computer และยังสามารถอ่านข้อมูลผ่านทาง Web Page

4) DOC: (ไฟล์เอกสาร Microsoft Word) คือ ไฟล์เอกสารประเภทข้อความ ที่อยู่ในรูปแบบไฟล์ไบนารีและประกอบด้วยพื้นของกรอบและบันทึกที่มุ่งเน้นข้อความ, การจัดรูปแบบหน้า กราฟแผนภูมิ ตาราง ภาพ และเอกสารอื่น ๆ ที่มีเนื้อหาที่ถูกสร้างขึ้นและแก้ไขเนื้อหาของเอกสารได้ ทั้งยังสามารถพิมพ์ในขนาดต่าง ๆ และมีความสามารถในการแสดงบนอุปกรณ์ที่แตกต่างกัน ให้ระบบปฏิบัติการสามารถอ่านไฟล์ DOC รูปแบบไบนารีของไฟล์ Word

5) XML: Extensible Markup Language (XML) คือ ภาษาที่ใช้ในการแสดงผลข้อมูล หากเปรียบเทียบกับภาษา HTML จะแตกต่างกันที่ HTML ถูกออกแบบมาเพื่อการแสดงผลอย่างเดียวเท่านั้น เช่น ให้แสดงผลตัวเล็ก ตัวหนา ตัวเอียง แต่ภาษา XML นั้นถูกออกแบบมาเพื่อเก็บข้อมูล โดยทั้งข้อมูลและโครงสร้างของข้อมูลนั้น ๆ ไปด้วยกัน ส่วนการแสดงผลก็จะใช้ภาษาเฉพาะซึ่งก็คือ XSL (Extensible Stylesheet Language)

6) RDF: Resource Description Framework (RDF) ตามที่ W3C ได้บอกไว้ คือการอธิบายถึงทรัพยากรของเว็บ เช่น ชื่อโดเมน ผู้เขียน วันที่ปรับปรุง และข้อมูลลิขสิทธิ์ของเว็บเพจ ไฟล์ประเภท RDF ได้รับการออกแบบขึ้นมาเพื่อให้คอมพิวเตอร์ หรือ แอปพลิเคชันของคอมพิวเตอร์ อ่านและเข้าใจ แต่ไม่ได้

รับการออกแบบให้แสดงผลผ่านเว็บแก่ผู้ใช้ ต้องใช้ภาษา XML ซึ่งเมื่อมีการนำ XML มาใช้โดย RDF จะเรียกว่า RDF/XML

7) KML: Keyhole Markup Language คือไวยากรณ์และรูปแบบไฟล์ XML สำหรับการทำโมเดล และการจัดเก็บคุณลักษณะทางภูมิศาสตร์ เช่น จุด เส้น ภาพ รูปหลายเหลี่ยม เช่นเดียวกับโมเดลที่แสดงใน Google Earth, Google Maps และแอปพลิเคชันอื่น ๆ โดยสามารถใช้ KML เพื่อเผยแพร่สถานที่และข้อมูลกับผู้ใช้แอปพลิเคชันรายอื่น ๆ ซึ่งไฟล์ KML ได้รับการประมวลผลโดย Google Earth ด้วยวิธีเดียวกับที่ไฟล์ HTML และ XML ได้รับการประมวลผลโดยเว็บเบราว์เซอร์ มีโครงสร้างพื้นฐานเป็นแท็กที่มีชื่อและแอตทริบิวต์ที่ใช้เฉพาะสำหรับการแสดงผล

8) SHP: Esri Shape file คือ ไฟล์ที่ประกอบด้วยข้อมูลเวกเตอร์แต่ละประเภท ซึ่งแต่ละเวกเตอร์ประกอบ จะประกอบเป็น Shape File ที่อ้างอิงพิกัด UTM สำหรับ Shape File นั้นหมายถึง ข้อมูลสารสนเทศภูมิศาสตร์ประเภทหนึ่งที่เก็บข้อมูลอยู่ในรูปของเวกเตอร์ (Vector) ใน 3 ลักษณะ คือ จุด (Point) เส้น (Line) และรูปปิด (Polygon) มีการแยกเก็บออกเป็นแต่ละชั้นข้อมูล (Layer) โดยที่ Shape File หนึ่งจะประกอบด้วยไฟล์อย่างน้อย 3 ไฟล์ที่มีการอ้างอิงถึงกันและกันและไม่สามารถขาดไฟล์ใดไฟล์หนึ่งไปได้

9) ODS: The Open Document Format for Office Applications (ODF) คือ ไฟล์งานตารางคำนวณ ในรูปแบบ XML Based สำหรับ Spreadsheets ซึ่งมีลักษณะการทำงานเหมือนกับไฟล์ EXCEL เป็นโปรแกรมโอเพ่นซอร์สออฟฟิศ

10) KMZ: (KML-Zipped) คือ ไฟล์ KML ที่ถูกบีบอัด (ZIP) ไว้ ซึ่งภายใน KMZ จะประกอบไปด้วยไฟล์ KML และไฟล์รูปภาพ และอาจจะมีไฟล์อื่นๆที่เกี่ยวข้อง เช่น COLLADA Model File และ Texture File ซึ่งจุดเด่นของไฟล์ KMZ คือ เมื่อบีบอัดไฟล์ KML แล้วจะมีขนาดที่เล็กลงมาก ทำให้สามารถใส่รูปภาพได้มากขึ้น

11) JSON: JavaScript Object คือ ฟอर्मแมตสำหรับแลกเปลี่ยนข้อมูลคอมพิวเตอร์ ซึ่งฟอर्मแมต JSON อยู่ในรูปข้อความธรรมดา (Plain Text) ที่ทั้งมนุษย์และโปรแกรมคอมพิวเตอร์สามารถอ่านเข้าใจได้

12) TXT คือ ไฟล์เอกสารต่าง ๆ ที่มีแต่ตัวอักษร ใช้โปรแกรม Notepad หรือโปรแกรม Text Editor อื่น ๆ ในการเปิดใช้งาน

13) HTML: Hypertext Markup Language (HTML) คือ ไฟล์ข้อความที่เก็บแท็ก Markup (Markup Tag) ขนาดเล็ก สามารถใช้ในการรวมคุณลักษณะการจัดรูปแบบบางอย่างและข้อกำหนดรูปแบบในเนื้อหาของหน้าเว็บเพจได้ แต่ไฟล์ประเภท HTML ไม่สามารถใช้ทำเว็บอร์ด, Register Form หรืออะไรที่ต้องเก็บข้อมูลได้ โดยไฟล์ HTML ใช้สำหรับการดู Web Site ผ่าน Browser ต่าง ๆ เช่น Netscape, IE, Opera และ Browser อื่น ๆ รวมทั้งโปรแกรม Text Editor ทั่วไป

14) TIFF: Tagged-Image File Format คือ รูปแบบไฟล์ที่มีการจัดเก็บแบบ Bitmap เหมาะสำหรับการจัดเก็บภาพที่นิยมในอุตสาหกรรมกราฟิกและช่างภาพ ซึ่งมีขนาดไฟล์ที่ใหญ่กว่า JPG แต่สามารถเก็บข้อมูลภาพ ที่มีความลึกสี (Bit Depth) ได้หลายแบบ อาทิ 8-bit หรือ 16-bit หรือแม้แต่ 24-bit ทำให้สามารถนำไฟล์ไปใช้งานกับ สิ่งพิมพ์ หรืองานที่ต้องการคุณภาพมาก ๆ ได้

15) JPEG: Joint Photographic Experts Group คือ ไฟล์ JPG เป็นไฟล์ภาพที่ถูกนำมาใช้สำหรับการจัดเก็บภาพถ่ายดิจิทัล เนื่องจากมีความละเอียดภาพสูงและไฟล์มีขนาดเล็ก สามารถเก็บภาพสีได้หลากหลายระดับความแม่นยำของสี (Bit Depth) ซึ่งความสามารถในการย่อขนาดไฟล์ของแฟ้ม JPEG นั้น เกิดจากการใช้เทคนิคการย่อขนาดภาพแบบการบีบอัดคงข้อมูลหลัก (Lossy Compression) หรือการบีบอัดแบบมีความสูญเสียทำให้ไม่นิยมใช้กับภาพที่เป็นลายเส้นหรือไอคอนต่าง ๆ

### ขั้นตอนและกระบวนการเผยแพร่ข้อมูล

1) ชุดข้อมูล (Datasets) ที่นำมาเผยแพร่ ต้องไม่ขัดต่อกฎหมาย ระเบียบ ข้อบังคับ รวมถึงความเป็นส่วนตัว (Privacy) และการรักษาความมั่นคงปลอดภัย (Security)

2) จัดเตรียมเตรียมข้อมูลให้อยู่ในรูปแบบที่มาตรฐานกำหนด รวมถึงการจัดทำคำอธิบายชุดข้อมูล นั้น ๆ เพื่อให้ผู้ใช้ข้อมูลสามารถเข้าใจเกี่ยวกับบริบทของข้อมูล

3) นำส่งชุดข้อมูล (Datasets) ขึ้นเผยแพร่ โดยระบุชื่อเจ้าของข้อมูล คณะ ชื่อวิทยานิพนธ์หรืองานวิจัย ส่งมาที่ งานคลังปัญญา สถาบันบัณฑิตพัฒนบริหารศาสตร์ อีเมล [nidawr@nida.ac.th](mailto:nidawr@nida.ac.th)

\*\* การนำชุดข้อมูลเผยแพร่ในคลังปัญญา เจ้าหน้าที่จะอนุมัติเผยแพร่ชุดข้อมูลที่ไม่ขัดต่อกฎหมาย ระเบียบ ข้อบังคับ เช่น พระราชบัญญัติข้อมูลข่าวสารของราชการ พระราชบัญญัติคุ้มครองข้อมูลส่วนบุคคล (PDPA) ระเบียบว่าด้วยการรักษาความลับของทางราชการ (Classification) รวมถึงความเป็นส่วนตัว (Privacy) และการรักษาความมั่นคงปลอดภัย (Security) เท่านั้น