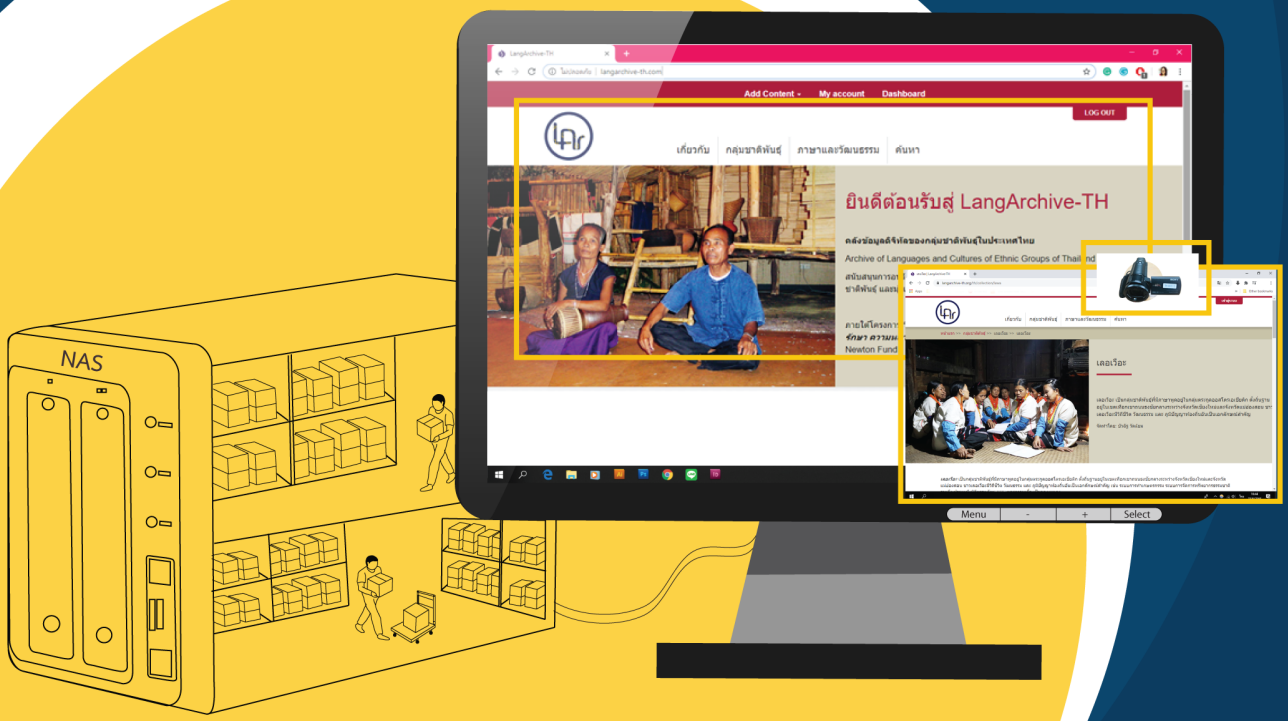
การจัดทำคลังข้อมูลดิจิทัล ด้านภาษาและวัฒนธรรม ของกลุ่มชาติพันธุ์ในประเทศไทย

สถาบันวิจัยภาษาและวัฒนธรรมเอเชีย มหาวิทยาลัยมหิดล



การจัดทำคลังข้อมูลดิจิทัล
ด้านภาษาและวัฒนธรรมของกลุ่มชาติพันธุ์
ในประเทศไทย

หนังสือการจัดทำคลังข้อมูลดิจิทัล
ด้านภาษาและวัฒนธรรมของกลุ่มชาติพันธุ์ในประเทศไทย

บรรณาธิการ ศิริเพ็ญ อึ้งสิทธิพูนพร, กุมารี ลาภอาภรณ์

กองบรรณาธิการ บัวฉัฐ วัดแย้ม, ณัฐมน โรจนกุล

พิมพ์ครั้งแรก จำนวน 200 เล่ม มิถุนายน 2563

ISBN 978-616-43-455-4

ผู้พิมพ์ โครงการคลังข้อมูลดิจิทัลของกลุ่มชาติพันธุ์: ส่งเสริมการปกป้องและดำรงรักษา ความหลากหลายทางภาษาและวัฒนธรรมในประเทศไทย
The ethnic group digital archive project: Promoting the protection and preservation of language and culture diversity in Thailand
สถาบันวิจัยภาษาและวัฒนธรรมเอเชีย มหาวิทยาลัยมหิดล
999 ถนนพุทธมณฑลสาย 4 ตำบลศาลายา อำเภอพุทธมณฑล
จังหวัดนครปฐม 73170
www.langarchive-th.org

ออกแบบปก / รูปเล่ม สุมาลี หาญยศ

พิมพ์ที่ บริษัท จรัลสนิทวงศ์การพิมพ์ จำกัด
219 ซอยเพชรเกษม 102/2 แขวงบางแคเหนือ
เขตบางแค กรุงเทพฯ 10160

ได้รับทุนสนับสนุนจาก สำนักงานคณะกรรมการส่งเสริมวิทยาศาสตร์ วิจัยและนวัตกรรม (สกสว.) สำนักงานกองทุนสนับสนุนการวิจัย (สกว.) เดิม

หน่วยงานร่วมทุนสนับสนุนการวิจัย Newton Institutional Links, British Council

คำนำ

โครงการคลังข้อมูลดิจิทัลของกลุ่มชาติพันธุ์: ส่งเสริมการปกป้อง และดำรงรักษาความหลากหลายทางภาษาและวัฒนธรรมในประเทศไทย ได้รับงบประมาณสนับสนุนการวิจัยจากสำนักงานกองทุนสนับสนุนการวิจัย (สกว.) เดิม ต่อมาเปลี่ยนชื่อเป็น สำนักงานคณะกรรมการส่งเสริมวิทยาศาสตร์ วิจัยและนวัตกรรม (สกสว.) และเป็นความร่วมมือวิจัยระหว่าง สถาบันวิจัยภาษาและวัฒนธรรมเอเชีย มหาวิทยาลัยมหิดล กับ Endangered Languages Archive (ELAR), SOAS University of London ภายใต้ทุนร่วมวิจัย Newton Institutional Links (โดยความร่วมมือด้านงบประมาณสนับสนุนการวิจัยระหว่าง British Council และสำนักงานกองทุนสนับสนุนการวิจัย (สกว. เดิม) มีระยะเวลาดำเนินงานตั้งแต่ 1 มกราคม 2562 - 30 มิถุนายน 2563 ซึ่งช่วงที่ขยายเวลาการวิจัย (1 มกราคม 2563 - 30 มิถุนายน 2563) เฉพาะงบที่ได้รับการสนับสนุนจากสำนักงานกองทุนสนับสนุนการวิจัย (สกว.) เดิม

โครงการฯ นี้มีวัตถุประสงค์หลักเพื่อจัดทำระบบคลังข้อมูลดิจิทัลให้เป็นมาตรฐาน และมีโครงสร้างที่เหมาะสมเพื่อรองรับการศึกษาวิจัยและรวบรวมองค์ความรู้จัดทำมาตรฐานการเก็บรักษาข้อมูลด้านภาษาและวัฒนธรรมของกลุ่มชาติพันธุ์ในประเทศไทยอันเป็นประโยชน์ต่อชุมชนกลุ่มชาติพันธุ์เอง และนักวิชาการทุกสาขาวิชาสามารถนำไปต่อยอดให้เกิดประโยชน์ได้ รวมทั้งสร้างบุคลากรให้สามารถจัดทำคลังข้อมูลดิจิทัล ฝึกอบรมและจัดการประชุมระดับชุมชน ระดับนานาชาติ เผยแพร่การจัดทำคลังข้อมูล เพื่อให้ตระหนักถึงองค์ความรู้ที่หลากหลายของกลุ่มชาติพันธุ์ในประเทศไทยและคุณค่าที่กำลังจะสูญหายไปพร้อมกับกลุ่มคน รวมทั้งคุณค่าของการเก็บรักษาข้อมูลเหล่านี้ที่จะเป็นประโยชน์ต่อมนุษยชาติ ในการส่งเสริมการเรียนรู้และการพัฒนาสังคมอย่างยั่งยืน

เนื้อหาในหนังสือเล่มนี้ประกอบด้วย ที่มาและความสำคัญของโครงการวิจัยฯ เนื้อหาเชิงวิชาการที่เกี่ยวกับการจัดทำคลังข้อมูล (Archiving) การบันทึกความรู้ทางภาษาและวัฒนธรรม (Language Documentation) การจัดการข้อมูล (Data Management) จนถึงการนำข้อมูลขึ้นเว็บไซต์คลังข้อมูล https://langarchive-th.org/ และคำถามที่พบบ่อย ๆ ระหว่างการจัดอบรมเชิงปฏิบัติการ “บันทึกองค์ความรู้และจัดทำคลังข้อมูลดิจิทัล” ให้แก่นักวิชาการของสถาบันการศึกษาท้องถิ่น เช่น วิทยาลัยชุมชนตราด มหาวิทยาลัยราชภัฏยะลา มหาวิทยาลัยสงขลานครินทร์ วิทยาเขตปัตตานี ฯลฯ หรือพื้นที่ที่มีกลุ่มชาติพันธุ์จำนวนมาก เช่น มหาวิทยาลัยราชภัฏเชียงราย โดยหวังว่านักวิชาการในพื้นที่จะมีโอกาสทำงานกับกลุ่มชาติพันธุ์นั้น ๆ ต่อไปในอนาคต

นอกจากนี้ยังได้รวบรวมเนื้อหาจากการถอดเทปงานสัมมนาความรู้การจัดทำคลังข้อมูลของหน่วยงานอื่น ๆ ในประเทศไทย อาทิ หอภาพยนตร์ (องค์การมหาชน) ศูนย์มานุษยวิทยาสิรินธร (องค์การมหาชน) ศูนย์เทคโนโลยีอีเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ (NECTEC) และคลังข้อมูลของโครงการฯ นี้ โดยอยู่ในส่วนของการประมวลผลที่เกิดขึ้นระหว่างดำเนินงานอีกด้วย ซึ่งเป็นการ

แลกเปลี่ยนการจัดทำคลังข้อมูลของหน่วยงานอื่น ๆ ในประเทศไทย ว่าเป็นอย่างไร มีความคล้ายคลึงหรือแตกต่างกันอย่างไรบ้าง

การจัดทำหนังสือเรื่อง **คลังข้อมูลดิจิทัลด้านภาษาและวัฒนธรรมของกลุ่มชาติพันธุ์ในประเทศไทย** มีวัตถุประสงค์เพื่อรวบรวมองค์ความที่รู้ได้จากโครงการวิจัยฯ และใช้เป็นแนวทางในการศึกษาวิจัย หรือฝึกอบรมให้แก่ผู้ที่สนใจการจัดทำคลังข้อมูลดิจิทัล ด้านภาษาและวัฒนธรรมของกลุ่มชาติพันธุ์ในประเทศไทย ซึ่งนับเป็นความต้องการที่สำคัญ เป็นมรดกทางวัฒนธรรมที่จับต้องไม่ได้ สำหรับจัดทำเป็นคลังข้อมูลขนาดใหญ่ (Big Data) ต่อไปในอนาคต

สมาชิกทีมวิจัยฯ หวังเป็นอย่างยิ่งว่า ข้อมูลในหนังสือเล่มนี้จะเป็นประโยชน์ต่อนักวิชาการและผู้สนใจทั่วไป ไม่มากก็น้อย หากมีความผิดพลาดใด ๆ ทีมวิจัยน้อมรับคำชี้แนะเพื่อการปรับปรุงให้ดีขึ้นต่อไป

ศิริเพ็ญ อึ้งสิทธิพูนพร หัวหน้าโครงการฯ และทีมวิจัย
มิถุนายน 2563

สารบัญ

หน้า

1. ที่มาของโครงการวิจัย 1

2. แนวคิดที่เกี่ยวข้องกับการบันทึกข้อมูลและการจัดทำคลังข้อมูลดิจิทัล 2

(Concept of Documentation and Digital Archiving)

2.1 Archive คืออะไร ทำไมจึงมีความสำคัญ

2.2 คำศัพท์ที่เกี่ยวข้อง

3. การบันทึกข้อมูลทางภาษาและวัฒนธรรม 6

3.1 วิธีการบันทึกข้อมูล

3.2 เครื่องมือที่ใช้ในการบันทึกข้อมูล (Equipment)

4. การจัดการข้อมูล (Data Management) 10

4.1 การจัดระบบข้อมูล (Filing System)

4.2 การจัดทำรายการข้อมูล (Metadata)

4.3 การแปลงไฟล์ข้อมูล

5. การนำข้อมูลขึ้นบน Mukurtu CMS 18

5.1 การ upload ข้อมูลไฟล์เสียงขึ้น SoundCloud

5.2 การ upload ข้อมูลไฟล์วิดีโอขึ้น Vimeo

5.3 การเพิ่มหัวข้อภาษาและวัฒนธรรม

(Digital heritage) บน LangArchive-TH

6. คำถามที่พบบ่อย 28

7. ประมวลผลที่เกิดขึ้นระหว่างทำโครงการและสิ่งที่จะทำในอนาคต 32

เอกสารอ้างอิง 65

ภาคผนวก 67

1. ที่มาของโครงการวิจัย

ประเทศไทยมีกลุ่มภาษาต่าง ๆ ทั้งสิ้นประมาณ 70 กลุ่ม ประกอบด้วย 5 ตระกูลภาษา ได้แก่ ตระกูลภาษาออสโตรเอเชียติก ตระกูลภาษาไต - กะได ตระกูลภาษาออสโตรนีเชียน ตระกูลภาษาจีน - ทิเบต และตระกูลภาษาม้ง - เมี่ยน (สุวิไล เปรมศรีรัตน์ และคณะ, 2547) ในจำนวนนี้มีภาษาที่อยู่ในภาวะวิกฤตถึง 15 กลุ่ม (Suwilai, 2007a) สาเหตุสำคัญเนื่องมาจากกระแสโลกาภิวัตน์ที่ทำให้ชุมชน สังคม และวัฒนธรรมเกิดความเปลี่ยนแปลง ประชากร คนรุ่นใหม่หรือวัยแรงงานต่างเคลื่อนย้ายเข้าสู่สังคมเมือง ส่งผลต่อการใช้ภาษาท้องถิ่นตลอดจนการสืบทอดความรู้ท้องถิ่นได้ถดถอยลง ขณะเดียวกันผู้คนในสังคมส่วนใหญ่ต่างมองข้ามความสำคัญของภาษาท้องถิ่น หากภาษาและภูมิปัญญาท้องถิ่นของชนกลุ่มต่าง ๆ สูญหายไป นั่นย่อมหมายถึงการสูญเสียตัวตน ภาษา วัฒนธรรม และองค์ความรู้ภูมิปัญญาท้องถิ่นของชนกลุ่มนั้น ๆ ไปด้วย ปัจจุบันชุมชนกลุ่มชาติพันธุ์จำนวนมากในโลกมีความพยายามหาวิธีที่จะธำรงรักษาภาษาและความรู้ทางวัฒนธรรมเหล่านั้นไว้ เพราะถือเป็นมรดกทางวัฒนธรรมอันล้ำค่ายิ่งของโลกใบนี้

โครงการวิจัย “คลังข้อมูลดิจิทัลของกลุ่มชาติพันธุ์ : ส่งเสริมการปกป้อง และดำรงรักษาความหลากหลายทางภาษาและวัฒนธรรมในประเทศไทย” ได้เล็งเห็นถึงความสำคัญของการสูญเสียความหลากหลายทางภาษาและภูมิปัญญาทางวัฒนธรรมที่นับวันจะเพิ่มมากขึ้นเรื่อย ๆ ส่งผลให้ชุมชนกลุ่มชาติพันธุ์สูญเสีย อัตลักษณ์ทางวัฒนธรรมและแหล่งเรียนรู้ที่สำคัญ การดำเนินงานในโครงการนี้จึงมุ่งเน้นการพัฒนาวิธีการอนุรักษ์ความหลากหลายทางภาษาและภูมิปัญญาทางวัฒนธรรมโดยชุมชนเจ้าของภาษา รวมทั้งพัฒนาศักยภาพชุมชนให้มีการพัฒนาคลังข้อมูลดิจิทัลที่ชุมชนท้องถิ่นสามารถเข้าถึงได้ ผ่านการแลกเปลี่ยนเรียนรู้งานวิจัยร่วมกับนักวิชาการ โดยมีกลุ่มเป้าหมายที่สำคัญ คือ กลุ่มชาติพันธุ์ต่าง ๆ ที่อยู่ในประเทศไทย โดยเฉพาะกลุ่มภาษาที่อยู่ในภาวะวิกฤต เป็นอันดับต้น ๆ แต่กระนั้นก็ตามทุกกลุ่มชาติพันธุ์ในประเทศไทยมีความตระหนักถึงสถานการณ์ที่ถดถอยของภาษาและองค์ความรู้ท้องถิ่นของตนเองเช่นกัน โครงการฯ นี้จึงไม่จำกัดการทำคลังข้อมูลเฉพาะกลุ่มภาษาในภาวะวิกฤตเท่านั้น แต่ได้สนับสนุนให้กลุ่มที่มีโอกาสและศักยภาพได้จัดทำคลังข้อมูลของตนเองในเบื้องต้น เพื่อเป็นแหล่งข้อมูลสำหรับผู้ที่สนใจและลูกหลานในกลุ่มของพวกเขาเองได้เรียนรู้ต่อไป

งานวิจัยด้านฟื้นฟูภาษาและวัฒนธรรมที่ผ่านมา ไม่ว่าจะเป็นงานของนักภาษาศาสตร์หรืองานวิจัยชุมชนที่ทำร่วมกับนักวิชาการ ดังเช่นที่บุคลากรของศูนย์ศึกษาและฟื้นฟูภาษาและวัฒนธรรมในภาวะวิกฤต สถาบันวิจัยภาษาและวัฒนธรรมเอเชีย มหาวิทยาลัยมหิดล ได้ทำงานร่วมกับชุมชนเจ้าของภาษามานานกว่า 20 ปี ได้จัดทำระบบเขียนภาษาท้องถิ่นอักษรไทย เพื่อเป็นเครื่องมือในการบันทึกองค์ความรู้ของกลุ่มนั้น ๆ และบางกลุ่มยังได้ผลิตสื่อการเรียนการสอน เพื่อใช้เป็นเครื่องมือในการถ่ายทอดภาษาสื่อรุ่นต่อไป ผลผลิตส่วนใหญ่จะเป็นการบันทึกองค์ความรู้ด้วยมือทำสื่อการเรียนการสอนด้วยมือ เช่น หนังสือแบบเรียนท้องถิ่น หนังสืออ่านภาษาท้องถิ่นเล่มเล็กเล่มยักษ์ เป็นต้น เมื่อเวลาผ่านไปไม่นานวัตถุ สิ่งของ หนังสือ หรือสื่อการเรียนรู้เหล่านั้นจะเสื่อม

สลาย ผุ เปื่อย หากไม่ใช้เทคโนโลยีสมัยใหม่มาช่วยเก็บรักษา ความรู้เหล่านั้นก็สูญหายในเวลาอันสั้น ปัจจุบันการเก็บรักษาข้อมูลในรูปแบบไฟล์ดิจิทัล เป็นวิธีการเก็บข้อมูลได้เป็นเวลานาน สามารถเข้าถึงข้อมูลได้ง่ายด้วยอินเทอร์เน็ต โครงการวิจัยฯ นี้ จึงมุ่งเน้นที่จะจัดการกับข้อมูลที่เคยเก็บไว้แล้วในรูปแบบอะนาล็อก ให้เป็นข้อมูลดิจิทัล และการเก็บข้อมูลใหม่ที่ทันสมัยและมีคุณภาพเพื่อเก็บรักษาได้ในระยะยาว

วัตถุประสงค์ของโครงการ ฯ

1. เพื่อจัดทำระบบคลังข้อมูลดิจิทัลให้เป็นมาตรฐาน และมีโครงสร้างที่เหมาะสมเพื่อรองรับการศึกษาวิจัยและรวบรวมองค์ความรู้ของกลุ่มชาติพันธุ์ในประเทศไทย โดยเฉพาะกลุ่มภาษาที่อยู่ในภาวะวิกฤต ผ่านแพลตฟอร์มเว็บไซต์ http://www.mukurtu.org

2. จัดทำมาตรฐานการเก็บรักษาข้อมูลด้านภาษาและวัฒนธรรมของกลุ่มชาติพันธุ์ในประเทศไทย (โดยรวบรวมจากผลงานที่ทำงานร่วมกับชุมชนกลุ่มชาติพันธุ์มากว่า 20 ปี) ให้พร้อมเผยแพร่ เพื่อประโยชน์ต่อชุมชนกลุ่มชาติพันธุ์ และนักวิชาการทุกสาขาวิชาสามารถนำไปต่อยอดให้เกิดประโยชน์ได้

3. พัฒนาศักยภาพบุคลากรให้มีความรู้ด้านการดูแลรักษาข้อมูล การรวบรวมข้อมูลอย่างเป็นระบบ และสามารถจัดทำคลังข้อมูลดิจิทัลได้

4. สร้างเครือข่ายการทำงานด้านกลุ่มชาติพันธุ์ ร่วมกับสถาบันการศึกษาอื่น ๆ

5. เผยแพร่การจัดทำคลังข้อมูล โดยผ่านการฝึกอบรมและการจัดประชุมทั้งระดับชุมชนและระดับนานาชาติ เพื่อทำให้คนในชุมชนและสังคมได้เกิดความตระหนัก เห็นคุณค่าและความสำคัญของภาษาและวัฒนธรรมที่กำลังจะสูญหาย รวมทั้งเห็นคุณค่าของการเก็บรักษาข้อมูล ซึ่งเป็นประโยชน์ต่อมนุษยชาติ ในการส่งเสริมการเรียนรู้และการพัฒนาสังคมอย่างยั่งยืน

ดังนั้นคู่มือเล่มนี้ จึงเป็นการถอดองค์ความรู้จากโครงการวิจัย ฯ เพื่อเป็นแนวทางในการจัดทำคลังข้อมูลดิจิทัลต่อไปในอนาคต เนื้อหาในคู่มือจะเป็นหัวข้อที่เกี่ยวข้องกับแนวคิดการบันทึกและการจัดทำคลังข้อมูลดิจิทัลในเชิงวิชาการ รวมทั้งกระบวนการบริหารจัดการข้อมูลเพื่อนำไปเป็นแนวทางจัดอบรมเชิงปฏิบัติการต่อไป ทั้งยังรวบรวมปัญหาที่พบบ่อยระหว่างดำเนินการอีกด้วย

2. แนวคิดที่เกี่ยวข้องกับการบันทึกข้อมูลและการจัดทำคลังข้อมูลดิจิทัล (Concept of Documentation and Digital Archiving)

2.1 Archive คืออะไร ทำไมจึงมีความสำคัญ

เนื่องจากข้อมูลภาษาและวัฒนธรรมของกลุ่มชาติพันธุ์เป็นข้อมูลที่มีความสำคัญ ส่วนใหญ่จะถูกบันทึกในรูปแบบเอกสาร หรือสื่อทำมือ ซึ่งไม่มีความคงทน หากปล่อยไว้นานวัน

ข้อมูลเหล่านี้ก็จะผุสลายไป การจัดทำคลังข้อมูล หรือ Archive เป็นวิธีการจัดการข้อมูลอย่างเป็นระบบ ซึ่งจะเป็นประโยชน์ต่อการนำมาใช้งาน ทั้งในกลุ่มนักวิชาการและชุมชนเจ้าของภาษา รวมทั้งยังสามารถนำข้อมูลเหล่านี้เผยแพร่ผ่านช่องทางออนไลน์ได้

แต่เดิมแหล่งข้อมูลส่วนใหญ่เป็นวัตถุที่จับต้องได้ เช่น หนังสือ เครื่องมือต่าง ๆ ภาพถ่าย งานศิลปะ แผ่นจารึก ซึ่งล้วนแล้วแต่มีคุณค่าเป็นอย่างมาก ข้อมูลเหล่านี้มักถูกจัดเก็บไว้ในพิพิธภัณฑ์ ห้องสมุด มหาวิทยาลัย หรือสถาบันการศึกษา ซึ่งมีข้อจำกัดในการเข้าถึงข้อมูลและการนำไปใช้งาน กล่าวคือ ข้อมูลเหล่านี้มักถูกนำมาใช้เมื่อมีคนต้องการ และต้องได้รับการอนุญาตเท่านั้น (Austin, 2011; Nordhoff & Hammarström, 2014; Trilsbeek & Wittenburg, 2006) ทำให้คนภายนอก หรือชุมชนเจ้าของภาษาไม่สามารถเข้าถึงข้อมูลเพื่อใช้งานได้ ทั้ง ๆ ที่ข้อมูลบางอย่างลูกหลานเจ้าของภาษาควรได้ใช้ประโยชน์เพื่อการเรียนรู้ภาษาและวัฒนธรรมของตนเอง

ต่อมาแนวคิดเรื่องการจัดทำคลังข้อมูลได้ขยายกว้างขึ้น และมีความเปลี่ยนแปลงอย่างรวดเร็ว โดยเฉพาะในช่วง 30 ปีที่ผ่านมานี้ มีการให้คำจำกัดความหรือนิยามการจัดทำคลังข้อมูลภาษาในภาวะวิกฤตไว้ว่า การจัดทำคลังข้อมูล หมายถึง การจัดทำคลังข้อมูลโดยองค์กร หรือสถาบันที่มีผู้รับผิดชอบที่ได้รับความไว้วางใจ เพื่อเก็บรักษาคลังข้อมูลเหล่านั้นให้อยู่ได้นาน ซึ่งในระยะแรกนั้นมุ่งเน้นเก็บรักษาวัตถุสิ่งของที่สำคัญของกลุ่มชาติพันธุ์เล็ก ๆ รวมถึงชนพื้นเมืองซึ่งมีอยู่ทุกประเทศ พัฒนาการของการพัฒนาคลังข้อมูลภาษาและวัฒนธรรมของภาษาในภาวะวิกฤต แบ่งออกเป็น 4 ช่วง ดังนี้

ช่วงที่ 1 ช่วงก่อนปี ค.ศ. 1990 มีชาวอเมริกันเป็นผู้บุกเบิก เช่น Franz Boas and Edward Sapir ได้บันทึกข้อมูลภาษาของชนพื้นเมือง ซึ่งอยู่ในรูปแบบสมุดบันทึกภาคสนาม การแปล การเก็บรวบรวมคำศัพท์ การถ่ายถอดข้อมูล และการพรรณนาไวยากรณ์ ข้อมูลต่าง ๆ ถูกรวบรวมอยู่ในรูปแบบเอกสาร รวมถึงเทปบันทึกเสียง ซึ่งถือเป็นการจัดเก็บข้อมูลในรูปแบบอะนาล็อก (Analogue) บุคคลอื่นสามารถเข้าถึงข้อมูลได้ยาก การเผยแพร่ในระยะนั้น ทำได้เพียงตีพิมพ์ในวารสารบางฉบับ เช่น International Journal of American Linguistics (Woodbury, 2011:163) คนที่เข้าถึงข้อมูลได้ส่วนใหญ่จึงเป็นนักวิชาการที่สนใจ

ช่วงที่ 2 ระหว่างปี ค.ศ. 1990-1999 ซึ่งนักภาษาศาสตร์มีความตระหนักในการเก็บบันทึกภาษาที่อยู่ในภาวะวิกฤตมากขึ้น จึงให้ความสำคัญกับการเก็บรักษาข้อมูลภาษาและการธำรงภาษา Himmelmann (1998:166) ให้คำนิยาม การบันทึกทางภาษา (Language Documentation) ว่าคือการบันทึกภาษาศาสตร์เชิงปฏิบัติและขนบธรรมเนียมการใช้ภาษาในชุมชน ซึ่งมีความหมายต่างจากการพรรณนาด้านภาษาศาสตร์ (Descriptive Linguistics) โดยที่การบันทึกทางภาษามีขอบเขตกว้างขวางกว่า และต้องใช้ความพยายามมากกว่า นอกจากนี้ Woodbury (2003:46–48) กล่าวว่า การบันทึกข้อมูลทางภาษาที่ดีควรประกอบไปด้วยข้อมูลหลายลักษณะ เช่น ข้อมูลมีขนาดใหญ่ เน้นการใช้ประโยชน์ของข้อมูลที่หลากหลาย เป็นต้น ซึ่งข้อมูลที่ดีนั้นไม่เพียงแต่เป็นประโยชน์

ต่อนักภาษาศาสตร์เท่านั้น แต่ยังเป็นประโยชน์ต่อนักวิชาการสาขาอื่น ๆ ด้วย ความพยายามในการบันทึกข้อมูลสมัยใหม่จะให้ความสำคัญในการเก็บรักษาข้อมูลปฐมภูมิให้มีอายุที่ยาวนาน

ช่วงที่ 3 เริ่มตั้งแต่ช่วงต้นศตวรรษที่ 21 ถือเป็นช่วงที่มีความก้าวหน้าด้านเทคโนโลยีมากขึ้น มีความพยายามที่จะระดมความคิดเพื่อพัฒนามาตรฐานการปฏิบัติงาน มีแหล่งทุนขนาดใหญ่ให้การสนับสนุนการบันทึกข้อมูลภาษาเพื่อทำเป็นคลังข้อมูล ซึ่งเป็นองค์ประกอบสำคัญในการเก็บบันทึกข้อมูลทางภาษา เป็นช่วงที่มีการพัฒนาแนวทางปฏิบัติที่ดี (best practices) สำหรับการทำคลังข้อมูลดิจิทัล และต้องคำนึงถึงจริยธรรมของการบันทึกข้อมูลทางภาษาด้วย

ในยุคนี้ได้ชี้ให้เห็นว่ามีปัจจัยสำคัญ 3 ประการที่ทำให้การบันทึกข้อมูลทางภาษาเพิ่มขึ้นอย่างมาก ได้แก่

1) การเห็นความสำคัญของภาษาที่อยู่ในภาวะวิกฤตมากขึ้น
2) มีจำนวนแหล่งทุนสนับสนุนจำนวนมากที่ให้ความสำคัญและส่งเสริมให้มีการบันทึกข้อมูลทางภาษา
3) ความทันสมัยด้านเทคโนโลยี ที่ทำให้การบันทึกข้อมูลมีประสิทธิภาพและมีคุณภาพสูง โดยเฉพาะอย่างยิ่ง การบันทึกเสียงและวิดีโอ และสามารถเผยแพร่ข้อมูลทางอินเทอร์เน็ตได้อย่างกว้างขวาง

ช่วงที่ 4 ยุคปัจจุบัน (ค.ศ. 2010 - ปัจจุบัน) ซึ่งเป็นการสนทนา หรือค้นหาข้อมูลระหว่างผู้ใช้คลังข้อมูล ผู้ให้ข้อมูล และผู้เก็บรักษาข้อมูลได้ขยายขอบเขตกว้างขวางอย่างไร้พรมแดน

สถานการณ์เมื่อไม่กี่ปีที่ผ่านมามีการเปลี่ยนแปลงไปอย่างรวดเร็ว กล่าวคือ เป็นการพัฒนาการทำคลังข้อมูลแบบมีส่วนร่วม (Participatory Archive) หมายถึง รูปแบบการจัดทำคลังข้อมูลโดยคณะ องค์กร หรือกลุ่มคน ไม่ใช่ทำโดยผู้เชี่ยวชาญคนใดคนหนึ่ง รูปแบบดังกล่าวสามารถเกิดขึ้นได้โดยอาศัยปัจจัยพื้นฐาน 4 ประการ ได้แก่

1) การพัฒนาองค์ความรู้ชุมชนเพื่องานวิจัยทางภาษาศาสตร์
2) เพิ่มอำนาจให้กับคนในชุมชนในการบริหารจัดการภาษาของตนเอง กล่าวคือ ให้ชุมชนเจ้าของภาษามีบทบาทในการทำคลังข้อมูลมากขึ้น เพื่อจุดประสงค์ในการธำรงรักษาและฟื้นฟูภาษาวัฒนธรรมของตนเอง
3) ผสานสื่อสังคมเข้ากับการทำคลังข้อมูล
4) พัฒนารูปแบบการทำงานร่วมกันในการทำคลังข้อมูล

จากข้อมูลข้างต้น จะเห็นได้ว่าพัฒนาการของการพัฒนาคลังข้อมูลภาษาและวัฒนธรรมของภาษาในภาวะวิกฤต มีความพยายามหาวิธีการบันทึกความรู้ของกลุ่มชาติพันธุ์ต่าง ๆ เพื่อเก็บรักษาให้มีอายุยาวนาน สำหรับเป็นแหล่งศึกษาค้นคว้า สืบทอดจากรุ่นสู่รุ่น ซึ่งเทคโนโลยีสมัยใหม่มีประโยชน์ต่อการทำคลังข้อมูลดิจิทัลเป็นอย่างมาก นอกจากนี้ยังมีพัฒนาการจากการบันทึกโดยนักวิชาการมาเป็นความร่วมมือระหว่างนักวิชาการกับเจ้าของภาษา และสุดท้ายมุ่งหวังให้เจ้าของภาษาเป็นผู้เก็บบันทึกองค์ความรู้ได้เองอย่างเป็นระบบ

2.2 คำศัพท์ที่เกี่ยวข้อง

ก่อนที่จะทำความเข้าใจเกี่ยวกับการจัดทำคลังข้อมูล (Archive) มีคำศัพท์สำคัญที่ควรรู้ ดังนี้ Metadata หมายถึง ข้อมูล หรือรายละเอียดของข้อมูล เป็นการบรรยายข้อมูลที่ได้เก็บบันทึกมา ซึ่งจะเป็นประโยชน์ต่อผู้ใช้ข้อมูล เช่น ใครเป็นผู้ให้ข้อมูล ใครเป็นผู้บันทึกข้อมูล มีใครเกี่ยวข้องบ้าง บันทึกข้อมูลเมื่อไหร่ มีเนื้อหาเกี่ยวกับอะไร เป็นต้น

Roles บทบาทของผู้ที่เกี่ยวข้อง ในที่นี้จะมีผู้เกี่ยวข้อง 3 ตำแหน่งที่สำคัญ ได้แก่

• Depositor เป็นผู้ที่จัดเตรียมข้อมูลเพื่อที่จะฝากในคลัง อาจจะเป็นผู้ให้ข้อมูลเอง หรือไม่ใช่ผู้ให้ข้อมูลก็ได้ แต่จะเป็นผู้เก็บรวบรวมและจัดการข้อมูลดังกล่าวให้พร้อมสำหรับนำเข้าคลังข้อมูล

• Digital Archivist เป็นคนที่รับผิดชอบการทำคลังข้อมูล รับฝากข้อมูลจากชุมชนหรือคนภายนอก พร้อมทั้งทำหน้าที่ตรวจสอบประเภทและรูปแบบของข้อมูลที่ถูกต้อง หากข้อมูลที่ได้รับยังมีรูปแบบที่ไม่ถูกต้องก็จะทำหน้าที่ประสานงานติดต่อกับผู้ฝากข้อมูลอีกครั้งหนึ่ง หรือแก้ไขให้ถูกต้องตามรูปแบบที่กำหนดของคลังข้อมูล

• Archive user ผู้ที่เข้ามาใช้คลังข้อมูล ซึ่งเป็นกลุ่มบุคคลทั่วไปในวงกว้าง หรือเป็นเจ้าของภาษาที่ต้องการเรียนรู้ภาษาของตนเอง หากผู้ใช้ต้องการใช้ข้อมูลที่ไม่สามารถเข้าถึงข้อมูลที่อยู่ในคลังได้ ก็จะต้องติดต่อประสานเพื่อขออนุญาตกับ Digital Archivist จากนั้น Digital Archivist จะทำหน้าที่ประสานงานเพื่อขออนุญาตจาก Depositor อีกครั้งหนึ่ง แล้วจึงประสานไปยัง Archive user Access to the materials คือการเข้าถึงเนื้อหาในคลังข้อมูล จะมีระดับการเข้าถึงข้อมูล อย่างน้อย 3 ระดับ ได้แก่

• Open to all users without registering ทุกคนสามารถเข้าถึงข้อมูลได้โดยไม่จำเป็นต้องลงทะเบียนหรือสมัครเป็นสมาชิก ส่วนใหญ่เป็นข้อมูลที่ต้องการเผยแพร่ให้คนทั่วไปได้รับรู้ จึงลดขั้นตอนการสมัครเป็นสมาชิกของเว็บไซต์

• Open only to registered users คือผู้ที่จะเข้าถึงข้อมูลได้จะต้องเป็นผู้ที่ผ่านการลงทะเบียนแล้วเท่านั้น ซึ่งจะทำให้ทราบว่าผู้เข้ามาใช้ข้อมูลเป็นใคร สามารถวัดสถิติการใช้ข้อมูลได้ รวมทั้งยังเป็นการป้องกันข้อมูลในระดับเบื้องต้นอีกด้วย

• Closed, available only upon request เป็นข้อมูลปิด ผู้ใช้จะต้องขออนุญาตในการเข้าถึงหรือเข้าใช้ข้อมูลจากผู้ดูแลคลังข้อมูล หรือ Digital Archivist ก่อน ซึ่งถือเป็นการเคารพและปกป้องสิทธิ์ของเจ้าของข้อมูล

Discoverability ความสามารถให้ผู้คนเข้าถึงหรือค้นพบ ซึ่งควรคำนึงถึงการใช้คำค้นหาง่าย เข้าถึงได้สะดวก

CMS & Archiving system

CMS คือ ระบบการจัดการข้อมูล ที่สามารถให้คนที่ไม่ค่อยมีความรู้ทางเทคโนโลยีสารสนเทศ (IT)

สามารถเข้าใช้ได้ แต่ CMS ไม่ใช่ระบบที่ใช้ทำ Archive ซึ่งในที่นี้จะใช้ CMS Mukurtu[1] โดยมีคุณลักษณะดังนี้

1) CMS Mukurtu (Archive) มีระบบ Version control หมายถึงการสำรองข้อมูล (back up) กล่าวคือ เมื่อมีการปรับปรุงข้อมูลใหม่ในไฟล์เดิม (update) ระบบนี้จะทำการสำรองข้อมูลไฟล์เดิมไว้ ซึ่งจะทำให้เห็นข้อมูลเดิม โดยจะระบุชื่อผู้ที่ปรับปรุงแก้ไขข้อมูล ส่วนไฟล์เดิมจะไม่ปรากฏ หากต้องการเข้าถึงไฟล์เดิมก็จะต้องขออนุญาตจากผู้ดูแลคลังข้อมูล

2) CMS Mukurtu (Archive) จะมีระบบ Preservation layer หมายถึงการเก็บรักษาระดับชั้นของข้อมูล กล่าวคือ หากมีการปรับปรุงระบบ IT ผู้ดูแลคลังข้อมูลไม่จำเป็นต้องแปลงไฟล์เพื่อให้เข้ากับระบบใหม่ เนื่องจากระบบ Preservation layer จะทำหน้าที่เก็บรักษาข้อมูลไว้

3) CMS Mukurtu (Archive) จะมีความสม่ำเสมอของการให้รายละเอียดข้อมูล (Metadata) ทำให้ข้อมูลมีมาตรฐานเดียวกัน

Archiving system มีข้อดีคือ สามารถดูออนไลน์ได้ตลอดเวลา เป็นมาตรฐาน Normalization and fixity/ integrity checks (version control) สามารถตรวจสอบได้ มีระบบสำรองข้อมูล (back up)

3. การบันทึกข้อมูลทางภาษาและวัฒนธรรม

นักภาษาศาสตร์ได้กล่าวไว้ว่า ***"ในทุกสองสัปดาห์จะมีภาษาที่หายไป ในยุคของคนรุ่นถัดไปจะมีภาษาที่หายไปครึ่งหนึ่งของจำนวนภาษาทั้งหมดในโลก"*** การบันทึกภาษาและวัฒนธรรมจึงมีความสำคัญที่สามารถเก็บรวบรวมองค์ความรู้เพื่อถ่ายทอดสู่รุ่นต่อไปได้

การบันทึกข้อมูล (Documentation) ไม่ใช่แค่การบันทึกภาษา หรือเสียงพูดเท่านั้น แต่ควรบันทึกลักษณะท่าทางการแสดงออก ของผู้บอกภาษาหรือผู้ให้ข้อมูลด้วย นอกจากนี้การบันทึกข้อมูลไม่สามารถบันทึกได้ทุกอย่าง แต่ผู้บันทึกจำเป็นต้องเลือกประเด็น หรือหัวข้อในการบันทึก โดยมีการวางแผนร่วมกันกับเจ้าของภาษา ซึ่งจะทำให้ได้ข้อมูลที่มีความสำคัญและจำเป็นต้องบันทึกอย่างเร่งด่วน ทั้งนี้จะต้องมีการทำงานร่วมกันเป็นทีมระหว่างเจ้าของภาษาและนักวิชาการ นักวิจัย หรือนักภาษาศาสตร์อีกด้วย

3.1 วิธีการบันทึกข้อมูล

1) การวางแผนและเตรียมการร่วมกับชุมชนเจ้าของภาษา โดยจะต้องวางแผนร่วมกันว่าจะบันทึกเรื่องใด มีใครบ้างที่เป็นผู้ให้ข้อมูล จำนวนกี่คน รวมทั้งเวลาและสถานที่ที่จะบันทึกข้อมูล

[1] Mukurtu พัฒนาขึ้นที่ประเทศสหรัฐอเมริกา (USA) เพื่อส่งเสริมให้คนในชุมชนสามารถจัดการ แชร์ แลกเปลี่ยนข้อมูลทางภาษาและวัฒนธรรมทางเว็บไซต์นี้ และมีการทดลองใช้ครั้งแรกที่ออสเตรเลีย Mukurtu สามารถติดตั้งผ่าน Cloud หรือ Locally ก็ได้

2) ก่อนทำการบันทึกข้อมูล ควรทำความเข้าใจกับชุมชนหรือผู้ให้ข้อมูลก่อนว่าการบันทึกข้อมูลนี้มีวัตถุประสงค์อะไร และอนุญาตให้นำข้อมูลที่จะบันทึกนี้ไปเผยแพร่ได้หรือไม่ หากผู้ให้ข้อมูลอนุญาตให้มีการบันทึกและเผยแพร่ จะต้องมีการจัดทำหลักฐานการยินยอม เช่น การลงนามในเอกสารยินยอมให้เผยแพร่ข้อมูล หรือใช้วิธีการบันทึกเสียง หรือบันทึกภาพเพื่อเป็นหลักฐาน แต่หากผู้ให้ข้อมูลไม่อนุญาตให้บันทึกหรือเผยแพร่ ก็ไม่สามารถบันทึกได้

ทั้งนี้ ***"สิทธิ์"*** เป็นของส่วนบุคคล เช่น ข้อมูลส่วนบุคคล ส่วน ***"ลิขสิทธิ์"*** จะหมายถึง สิ่งที่จับต้องได้ เช่น หนังสือ วิดีโอ เทปเสียง เป็นต้น

3) ก่อนที่จะบันทึกข้อมูลควรมีการจัดสถานที่ และติดตั้งเครื่องมือที่ใช้ในการบันทึกข้อมูล พร้อมกับตรวจสอบอุปกรณ์ให้สามารถใช้งานได้ ทั้งนี้การตั้งกล้องเพื่อถ่ายวิดีโอ ควรตั้งกล้องในระดับสายตาหรือระดับเดียวกับผู้ให้ข้อมูล รวมทั้งควรคำนึงถึงทิศทางของแสงที่อยู่ในระหว่างการบันทึก

4) วิธีการเก็บข้อมูลมี 2 ลักษณะด้วยกัน กล่าวคือ ลักษณะที่ 1 เป็นวิธีการเก็บข้อมูลในลักษณะการสังเกตอย่างไม่มีส่วนร่วม (Non – Participant Observation) โดยผู้เก็บข้อมูลจะไม่เข้าไปมีส่วนร่วมหรือทำกิจกรรมร่วมกับผู้ให้ข้อมูล และลักษณะที่ 2 คือ การสังเกตอย่างมีส่วนร่วม (Participant Observation) เป็นการเก็บข้อมูลอย่างมีส่วนร่วมกับผู้ให้ข้อมูล โดยส่วนใหญ่จะใช้วิธีการสัมภาษณ์หรือร่วมพูดคุยกับผู้ให้ข้อมูล หากใช้วิธีการสัมภาษณ์ ควรตั้งกล้องบันทึกภาพให้ผู้สัมภาษณ์และผู้ถูกสัมภาษณ์อยู่ในภาพเดียวกัน ไม่ควรตั้งกล้องให้ผู้บอกภาษาอยู่ในกล้องคนเดียว

ลักษณะการสัมภาษณ์มีหลายรูปแบบ ได้แก่

• Standard interview เป็นการสัมภาษณ์โดยมีการเตรียมคำถามไว้ล่วงหน้า รวมทั้งมีการเตรียมสภาพแวดล้อม และมีการดำเนินการอย่างเป็นระบบตามขั้นตอน

• Semi - directed interview เป็นการสัมภาษณ์ที่มีการเตรียมแนวคำถาม เปิดโอกาสให้ผู้สัมภาษณ์ได้พูดถึงเรื่องอื่น แต่มีความเกี่ยวข้องกับประเด็นที่ผู้สัมภาษณ์ต้องการ ผู้สัมภาษณ์จะต้องคอยกำกับการสนทนาด้วย เพื่อไม่ให้ออกนอกเรื่องหรือใช้เวลามากเกินไป

• Interview with stimuli เป็นการสัมภาษณ์ที่มีเครื่องมือช่วย เช่น รูปภาพ สถานที่จริง หรือสิ่งของที่สามารถนำมาได้ หากใช้ภาพควรเป็นภาพที่มีความใกล้เคียงกับบริบทของชุมชน ซึ่งจะสามารถช่วยให้ผู้บอกภาษาสามารถบอกภาษาได้

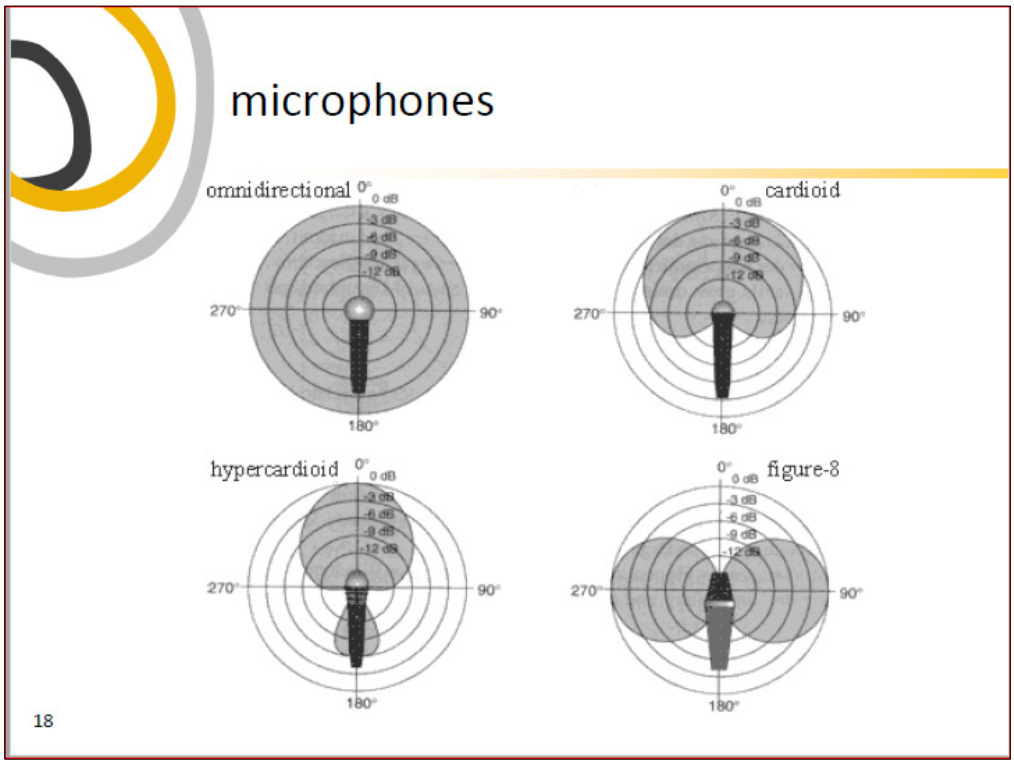
ทั้งนี้ควรมีคนในชุมชนที่ได้รับการฝึกฝนให้เป็นผู้บันทึกข้อมูล สามารถพูดภาษาในชุมชนได้ เรียกว่า ผู้อำนวยความสะดวก (Facilitator) คอยเป็นผู้กำกับหรือเป็นผู้สื่อสารกับผู้ให้ข้อมูล เพื่อให้การบันทึกข้อมูลนั้นเกิดผลสำเร็จ รวมทั้งจำเป็นต้องมีการทำงานร่วมกับนักวิชาการด้วย

5) เมื่อบันทึกข้อมูลเรียบร้อยแล้ว ผู้บันทึกจำเป็นต้องจดบันทึกรายละเอียดของข้อมูล (Metadata) โดยมีรายละเอียดต่าง ๆ เช่น ผู้ให้ข้อมูล รายละเอียดของข้อมูล ประเภทของข้อมูล บันทึกที่ไหน เมื่อไหร่ รวมทั้งต้องสอบถามเจ้าของข้อมูลว่าใครสามารถเข้าถึงข้อมูลได้บ้าง เป็นต้น

6) การจัดรวบรวมไฟล์ให้เป็นหมวดหมู่ หรือข้อมูลที่มีความเกี่ยวข้องกันไว้ด้วยกัน (Bundle) เช่น รูปภาพ ไฟล์เสียง ไฟล์วิดีโอ การถอดเสียงที่เป็นเรื่องเดียวกัน ควรรวมอยู่ในหมวดหมู่เดียวกัน หรือ Bundle เดียวกัน

3.2 เครื่องมือที่ใช้ในการบันทึกข้อมูล (Equipment)

เครื่องมือ หรืออุปกรณ์ที่ใช้ในการบันทึกข้อมูลที่สำคัญ ได้แก่ ไมโครโฟน เครื่องบันทึกเสียง และกล้องถ่ายวิดีโอ รวมทั้งขาตั้งกล้อง ซึ่งจำเป็นต้องคำนึงถึงคุณภาพของเครื่องมือเพื่อให้ได้ข้อมูลที่มีคุณภาพ ทั้งนี้การใช้ไมโครโฟน หรือเครื่องบันทึกเสียงมีหลายรูปแบบและหลายลักษณะ ควรเลือกใช้ให้เหมาะสมกับลักษณะของการรับเสียง เพื่อให้ได้คุณภาพเสียงที่ดี มีความคมชัด หากสามารถเลือกสถานที่ที่ไม่มีเสียงรบกวนจะดีมาก



ภาพที่ 1 : ลักษณะไมโครโฟนแบบต่าง ๆ (ภาพจาก PPT ของ Vera Ferreira ใช้ประกอบการอบรมเชิงปฏิบัติการ เรื่อง “Language Documentation and Archiving” ที่มหาวิทยาลัยมหิดล เมื่อวันที่ 15 พฤษภาคม 2562)

จากภาพ จะเห็นพื้นที่ที่รับเสียงของไมโครโฟนประเภทต่างๆ ซึ่งแต่ละประเภทมีวัตถุประสงค์การใช้งานต่างกัน เช่น ถ้ามีผู้ให้ข้อมูลคนเดียว ก็ควรใช้แบบ hypercardioid เพื่อลดเสียงรบกวนที่จะเข้ามาจากทิศทางอื่น ๆ หรือหากบันทึกการสนทนาระหว่างคนสองคนที่ยืนตรงข้ามกันก็ควรใช้แบบ figure-8 เป็นต้น

ภาพที่ 2 : ลักษณะของเครื่องบันทึกเสียงแบบต่าง ๆ (ภาพจาก PPT ของ Vera Ferreira ใช้ประกอบการอบรมเชิงปฏิบัติการเรื่อง “Language Documentation and Archiving” ที่มหาวิทยาลัยมหิดล เมื่อวันที่ 15 พฤษภาคม 2562)

ภาพที่ 3 : ลักษณะของเครื่องบันทึกภาพเคลื่อนไหวแบบต่าง ๆ (ภาพจาก PPT ของ Vera Ferreira ใช้ประกอบการอบรมเชิงปฏิบัติการเรื่อง “Language Documentation and Archiving” ที่มหาวิทยาลัยมหิดล เมื่อวันที่ 15 พฤษภาคม 2562)

4. การจัดการข้อมูล (Data management)

การจัดการข้อมูล (Data management) คือการจัดการไฟล์ข้อมูล ได้แก่ สื่อประเภทต่าง ๆ เช่น วิดีโอ เสียง รูปภาพ เอกสาร เป็นต้น ซึ่งได้รับการจัดกลุ่มตามหัวข้อภาษาและวัฒนธรรม (Digital Heritage Item) ให้เป็นระบบและอยู่ในมาตรฐานเดียวกัน ขั้นตอนการจัดการข้อมูลประกอบไปด้วย 3 ขั้นตอนหลักสำคัญ ได้แก่ การจัดระบบข้อมูล (Filing System) การจัดทำรายการข้อมูล (Metadata) และการแปลงไฟล์ข้อมูล

4.1 การจัดระบบข้อมูล (Filing System)

การจัดระบบข้อมูลภายในหนึ่งโฟลเดอร์ จะใช้บรรจุข้อมูลจากหนึ่งภาษาและหนึ่งผู้ฝากข้อมูล (Depositor) ซึ่งอาจมีหลายหัวข้อภาษาและวัฒนธรรม ทั้งนี้ภายในหนึ่งโฟลเดอร์จะต้องประกอบด้วยข้อมูล 3 ส่วน ได้แก่

- ไฟล์ข้อมูลทั้งหมด
- รายการข้อมูล หรือ Metadata ซึ่งบันทึกรายละเอียดของข้อมูลทุก ๆ ไฟล์ในโฟลเดอร์นั้น ควรตั้งชื่อไฟล์ว่า metadata.csv
- ไฟล์ที่อธิบายระบบการตั้งชื่อไฟล์ทั้งหมดในโฟลเดอร์นั้น ควรตั้งชื่อไฟล์ว่า readme.txt จะมีกี่หัวข้อภาษาและวัฒนธรรมก็ได้

ระบบการตั้งชื่อ

การตั้งชื่อไฟล์มีส่วนสำคัญมากในการกำหนดมาตรฐานการจัดการระบบของข้อมูลให้เป็นระเบียบและสืบค้นได้ง่าย การตั้งชื่อไฟล์นั้นจะใช้สัญลักษณ์ที่เป็น Unicode ซึ่งเป็นสัญลักษณ์มาตรฐานของเทคโนโลยีสารสนเทศ เช่น ตัวอักษรภาษาอังกฤษ a-z, ตัวเลข 0-9, เครื่องหมายยัติภังค์ (-), เครื่องหมาย underscore (_) เป็นต้น โดยจะต้องหลีกเลี่ยงการใช้ตัวอักษรภาษาไทย, อักขระพิเศษ และช่องว่าง คลังข้อมูลดิจิทัลของกลุ่มชาติพันธุ์ฯ LangArchive-TH ได้กำหนดระบบการตั้งชื่อไว้เป็นมาตรฐานเดียวกันดังต่อไปนี้

1) การตั้งชื่อโฟลเดอร์ (Folder Name)

ชื่อภาษา - ชื่อผู้ฝากข้อมูล / ผู้ให้ข้อมูล - หมายเลขผู้ฝากข้อมูล (Deposit Key)[1]

ตัวอย่าง

chung-waen-010

- chung = ภาษาชอุ้ง
- waen = ชื่อผู้ให้ข้อมูล นางแหวน กากี
- 010 = Deposit key

2) การตั้งชื่อไฟล์ (Filename)

รหัสภาษา_ชื่อผู้ฝากข้อมูล/ผู้ให้ข้อมูล_หมายเลขหัวข้อ-ลำดับไฟล์.นามสกุลไฟล์

ตัวอย่าง

scq_waen_001-002.WAV

- scq = รหัสภาษาจาก Ethnologue (www.ethnologue.com)
- waen = ชื่อผู้ฝากข้อมูล (depositor) หรือ ชื่อผู้ให้ข้อมูล (contributor)
- 001 = หมายเลขหัวข้อ (digital heritage) ได้แก่ ลำดับของกลุ่มข้อมูล
- 002 = ลำดับไฟล์ในกลุ่มข้อมูล
- WAV = นามสกุลของไฟล์เสียง

ข้อสังเกตในการจัดระบบข้อมูล

1) ไฟล์ที่มาจากหัวข้อเดียวกัน ให้ใช้ Base Name (ส่วนที่เป็นตัวหนา) เหมือนกัน **รหัสภาษา_ชื่อผู้ฝากข้อมูล/ผู้ให้ข้อมูล_หมายเลขหัวข้อ-ลำดับไฟล์.นามสกุลไฟล์**

2) การเรียงลำดับไฟล์นั้นให้ไล่ลำดับไปโดยไม่แยกประเภทของไฟล์ (media type) เช่น วิดีโอ (MP4), เสียง (MP3), รูปภาพ (JPG) และ เอกสาร (PDF)

ตัวอย่าง

scq_waen_001-001.MP4
scq_waen_001-002.JPG
scq_waen_001-003.JPG
scq_waen_001-004.PDF
scq_waen_002-001.MP4
scq_waen_002-002.JPG
scq_waen_002-003.JPG
scq_waen_002-004.PDF

ในกรณีที่ไฟล์มาจากแหล่งเดียวกัน เช่น ไฟล์วิดีโอและไฟล์เสียงที่ได้รับการบันทึกจากเหตุการณ์ทางภาษาและวัฒนธรรม (communicative event) เดียวกัน ให้ใช้ Base Name เดียวกันทั้งหมด ต่างกันแค่**นามสกุลไฟล์**เท่านั้น

ตัวอย่าง

scq_waen_001-001.MP4
scq_waen_001-001.WAV

ในตัวอย่างนี้ ไฟล์ลำดับ 001 มีข้อมูลวิดีโอ (MP4) และเสียง (WAV) ที่แปลงมาจากไฟล์วิดีโอ แล้วบันทึกในรูปแบบไฟล์เสียง อาจจะเพื่อประโยชน์ในการใช้เฉพาะไฟล์เสียง

3) ในกรณีที่ไฟล์ประเภทเดียวกันมีความต่อเนื่องกัน แต่แยกกันหลายไฟล์ ให้เพิ่มเครื่องหมาย underscore และลำดับ 01, 02, 03,... ต่อท้าย เพื่อให้ทราบว่าเป็นเนื้อหาที่ต่อเนื่องกัน

ตัวอย่าง

scq_waen_001-001_01.MP4
scq_waen_001-001_02.MP4
scq_waen_001-001_03.MP4

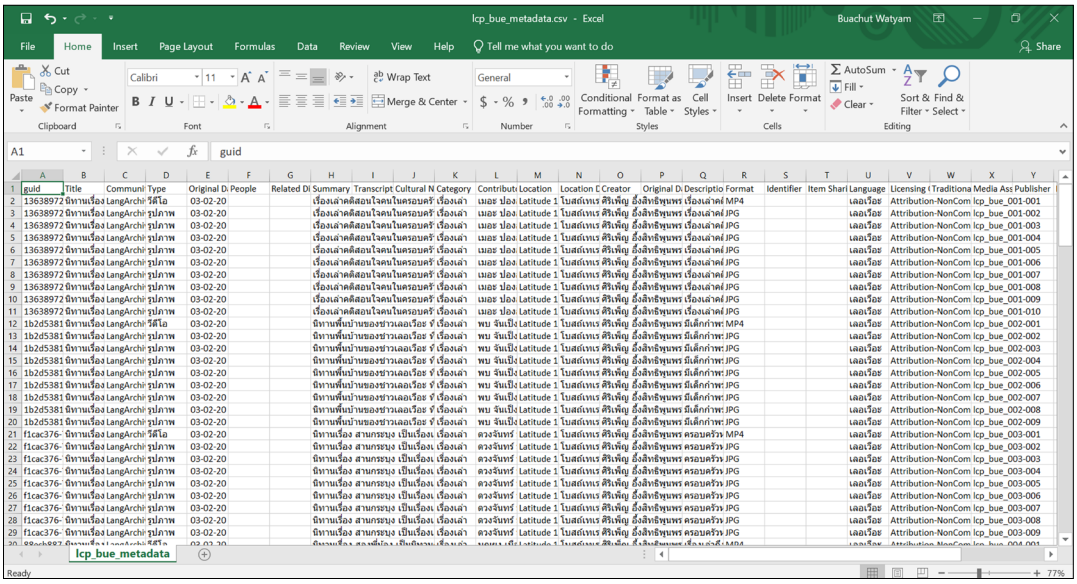
จากตัวอย่างข้างต้น หมายความว่า เนื้อหาของวิดีโอนี้มีความยาวเกินไป จึงตัดออกเป็น 3 ไฟล์ ได้แก่ 01, 02, 03 ตามลำดับ ดังนี้เมื่อผู้ใช้ข้อมูลเปิดดูเนื้อหาแต่ละไฟล์จะไม่ใช้เวลานานเกินไปจนรู้สึกเบื่อ และสามารถดูทีละไฟล์ หรือครั้งละไฟล์ก็ได้

4.2 การจัดทำรายการข้อมูล (Metadata)

การจัดทำรายการข้อมูล หรือ Metadata คือข้อมูลซึ่งบรรยายเนื้อหาและรายละเอียดต่าง ๆ ซึ่งจะเป็นประโยชน์ต่อการจัดการและสืบค้น คลังข้อมูลฯ LangArchive-TH ใช้โครงสร้าง Metadata ของ Mukurtu เป็นมาตรฐาน โดยใช้โปรแกรม Microsoft Excel ในการจัดการบันทึกรายละเอียดของข้อมูลลงใน Node - Digital Heritage.csv ซึ่งมีหัวข้อสำคัญดังต่อไปนี้

- TITLE: ชื่อหัวข้อสั้น ๆ สำหรับหัวข้อภาษาและวัฒนธรรม
- COMMUNITY: กลุ่มผู้มีสิทธิเข้าถึงข้อมูล
- TYPE: ประเภทของสื่อ เช่น วิดีโอ, รูปภาพ, ไฟล์เสียง
- ORIGINAL DATE: วันที่เก็บข้อมูล
- SUMMARY: สรุปข้อมูลอย่างสั้น
- CATEGORY: หัวข้ออย่างกว้างของข้อมูล เช่น เรื่องเล่า การละเล่น สื่อการสอน ภาษา และภูมิปัญญาท้องถิ่น เป็นต้น
- CONTRIBUTOR: ผู้ให้ข้อมูล ซึ่งอาจจะเป็นเจ้าของภาษาหรือผู้ฝากข้อมูล
- LOCATION: สถานที่ในการเก็บข้อมูล
- CREATOR: ผู้เกี่ยวข้องกับกระบวนการเก็บข้อมูล
- DESCRIPTION: รายละเอียดทั้งหมดของข้อมูล
- FORMAT: นามสกุลไฟล์
- LANGUAGE: ภาษาของข้อมูล และภาษาที่ใช้ในการทำงาน
- LICENSING: Attribution-Non-commercial 4.0 International (CC BY-NC 4.0)

- MEDIA ASSETS: ชื่อไฟล์ ยกเว้นนามสกุลไฟล์
- PUBLISHER: ผู้เผยแพร่ข้อมูลบนคลัง
- KEYWORDS: คำสำคัญ
- PROTOCOL: สิทธิการเข้าถึงข้อมูล เช่น เปิด ปิด
- USERNAME: ผู้ upload ข้อมูลขึ้นคลัง



ภาพที่ 4 : แสดงรายการข้อมูล Node - Digital Heritage.csv

4.3 การแปลงไฟล์ข้อมูล

เมื่อจัดระบบข้อมูลเรียบร้อยแล้ว ให้คัดลอกโฟลเดอร์ข้อมูลนั้นไว้ก่อนนำไปจัดเก็บไว้ในส่วนสำรองข้อมูล (backup) กล่าวคือ ผู้จัดการข้อมูลจะมีโฟลเดอร์ข้อมูลอยู่ด้วยกันสองชุด ได้แก่ โฟลเดอร์ข้อมูลจริง (สำหรับการสำรอง) และโฟลเดอร์ข้อมูลคัดลอก (สำหรับทำงาน) ข้อมูลชุดหลังนี้จะได้รับการนำไปแปลงไฟล์เพื่อบรรจุขึ้นคลังข้อมูลออนไลน์ต่อไป

การแปลงไฟล์ข้อมูลมีจุดประสงค์เพื่อแปลงนามสกุลไฟล์เดิมให้เป็นนามสกุลที่เหมาะสมและทนทานกับการบรรจุขึ้นคลังข้อมูลออนไลน์ โดยไฟล์ข้อมูลที่จะนำจัดเก็บขึ้นคลังข้อมูลดิจิทัลออนไลน์นั้นควรเป็นนามสกุลไฟล์ที่มีความเสถียร ทนทานและเข้าถึงได้ในอนาคต ดังต่อไปนี้

- เสียง (Audio): wav 48kHz, 16 bit (MP3 สำหรับการ streaming)
- วิดีโอ (Video): MP4 (H264)
- รูปภาพ: JPEG
- รูปแสกน: PDF (PDF-A)
- เอกสาร: PDF (PDF-A), TXT (Unicode UTF-8)
- Annotations: EAF

โฟลเดอร์ข้อมูลจริง (สำรอง)

- ได้รับการจัดระบบข้อมูลแล้ว
- **ไม่**ผ่านการแปลงไฟล์

ตัวอย่าง

scq_waen_001-001.AVI
scq_waen_001-002.WAV
scq_waen_001-003.TIF
scq_waen_001-004.TIF
scq_waen_001-005.DOCX

โฟลเดอร์ข้อมูลคัดลอก (ทำงาน)

- ได้รับการจัดระบบข้อมูลแล้ว
- อยู่ในกระบวนการแปลงไฟล์หรือผ่านการแปลงไฟล์แล้ว

ตัวอย่าง

scq_waen_001-001.MP4
scq_waen_001-002.MP3
scq_waen_001-003.JPG
scq_waen_001-004.JPG
scq_waen_001-005.PDF

ตัวอย่างในกล่องสี่เหลี่ยมข้างต้นแสดงให้เห็นกระบวนการสำรองไฟล์ ความแตกต่างของโฟลเดอร์ข้อมูลจริงที่ใช้เพื่อการสำรอง และโฟลเดอร์ข้อมูลคัดลอกที่ใช้เพื่อการทำงานซึ่งได้รับการแปลงไฟล์จากนามสกุลไฟล์เดิมให้เป็นนามสกุลที่เหมาะสมต่อการนำจัดเก็บขึ้นคลังข้อมูลดิจิทัลออนไลน์ นอกจากนี้ขั้นตอนการแปลงไฟล์ข้อมูลยังมีประโยชน์ในการลดขนาดของไฟล์เดิมที่ใหญ่เกินไปจนสร้างปัญหาในการบรรจุและเข้ารับชมออนไลน์

การแปลงไฟล์โดยใช้โปรแกรม FFMPEG

โปรแกรม FFMPEG เป็นโปรแกรมแปลงไฟล์ที่ใช้ระบบ command prompt ในการสั่งการ สามารถแปลงและลดขนาด เสียง และรูปภาพได้ทุกนามสกุล โดยไม่ลดความละเอียดของไฟล์นั้น ๆ ลง

การติดตั้ง FFMPEG (สำหรับ Windows)

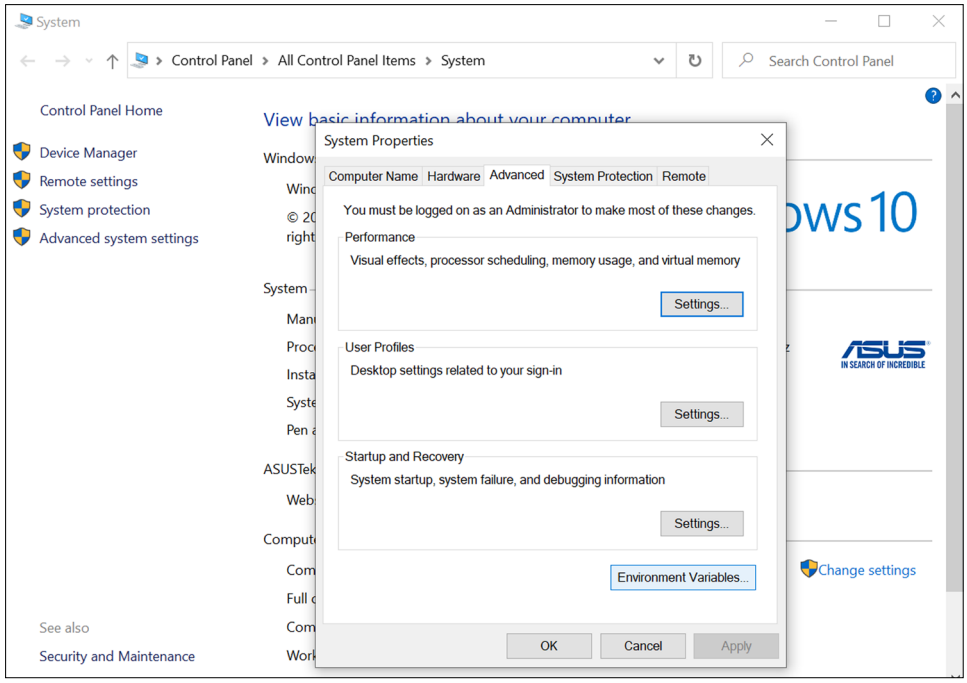
1) ดาวน์โหลดโปรแกรมที่ https://www.ffmpeg.org/download.html
2) คลิกขวาที่ไฟล์ ZIP แล้ว Extract to ffmpeg-4.2.2\
3) เมื่อได้โฟลเดอร์ ffmpeg-4.2.2 แล้ว ให้เปลี่ยนชื่อโฟลเดอร์เป็น FFmpeg
4) ย้ายโฟลเดอร์ FFmpeg ไปไว้ที่ Drive C:
5) คลิกขวาที่ This PC แล้วเลือก Properties
6) เลือก Advanced system settings ที่แถบด้านซ้าย
7) ไปที่ Environment Variables
8) คลิกที่ Path สองครั้ง
9) คลิก New แล้วพิมพ์คำสั่ง C:\FFmpeg\bin จากนั้นคลิก OK ทุกหน้าต่าง

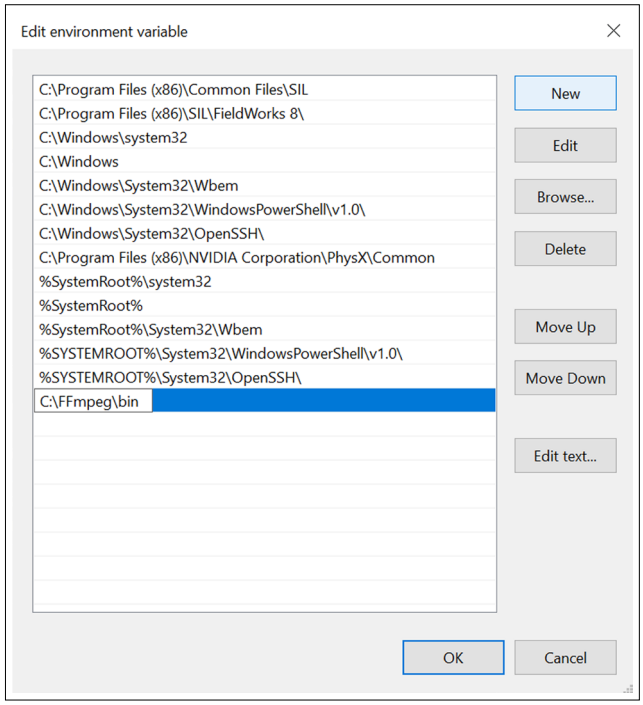
10) ทดสอบการติดตั้งด้วย command prompt

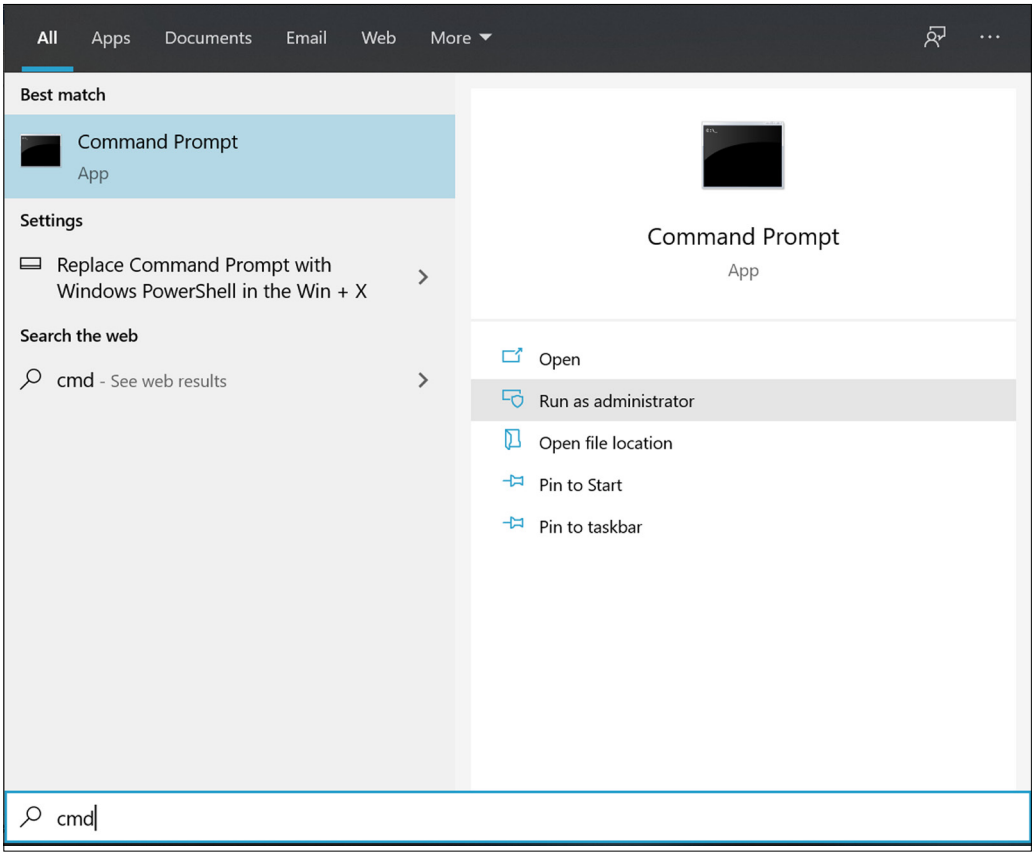
11) พิมพ์ Command Prompt หรือ cmd ที่ Start

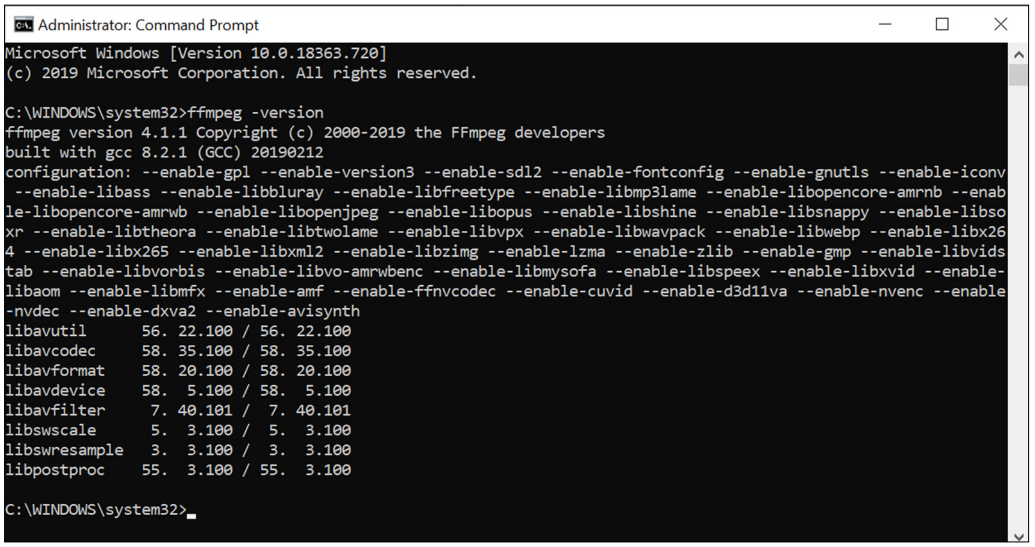
12) เลือก Run as administrator จะมีหน้าต่างขออนุญาตดำเนินการ ให้เลือก yes

13) พิมพ์คำสั่ง ffmpeg -version แล้วกด enter เพื่อทดสอบการติดตั้ง





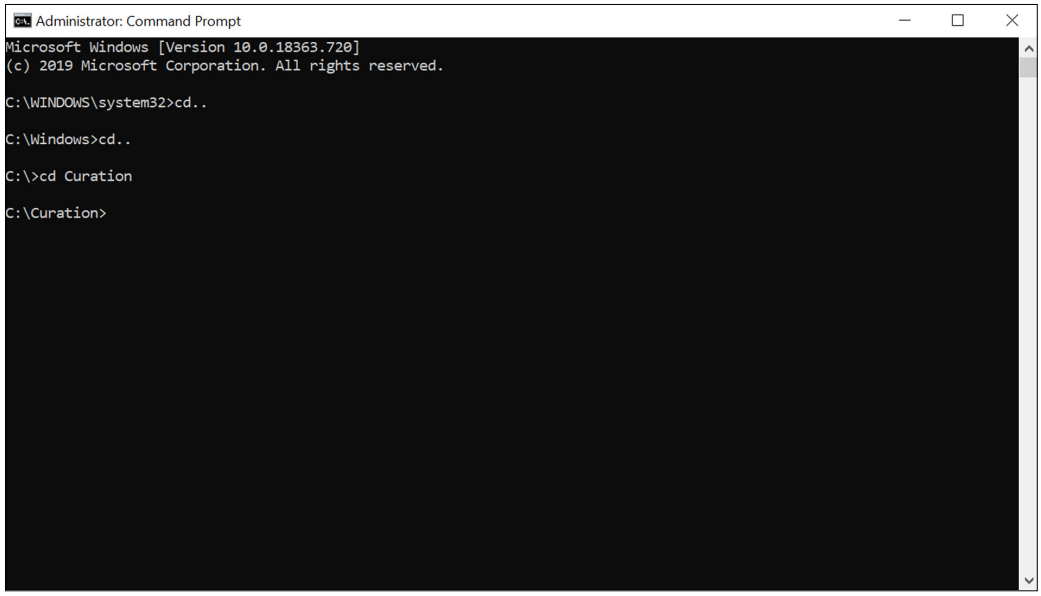




ภาพที่ 5 - 8 : แสดงขั้นตอนการติดตั้ง FFMPEG (สำหรับ Windows)

การใช้งาน FFmpeg

1) สร้างโฟลเดอร์ Curation ที่ Drive C:
2) สร้างอีกโฟลเดอร์ชื่อ Output ไว้ข้างในโฟลเดอร์ Curation
3) คัดลอกไฟล์ที่ต้องการแปลงไปไว้ในโฟลเดอร์ Curation
4) เริ่มการใช้งานโดยใช้ Command Prompt โดยเลือก Run as administrator เสมอ
5) พิมพ์คำสั่ง cd.. แล้ว enter สองครั้ง
6) พิมพ์คำสั่ง cd Curation แล้ว enter (การพิมพ์คำสั่งนี้ต้องพิมพ์ตามชื่อโฟลเดอร์ที่ได้ไว้สร้างใต้ C: รวมถึงตัวอักษรพิมพ์ใหญ่หรือพิมพ์เล็กด้วย)



7) เมื่อเห็นคำสั่ง C: \Curation> ดังภาพแล้ว จึงเริ่มใช้คำสั่งแปลงไฟล์ได้

การแปลงไฟล์ จำนวน 1 ไฟล์

ffmpeg -i **NameOfTheFile.extension** “Output**NameOfTheFile.mp4**”

- คำสั่งข้างต้นคือการสั่งแปลงไฟล์วีดีโอให้เป็น MP4
- เปลี่ยนนามสกุลไฟล์ด้านท้ายคำสั่ง เมื่อต้องการแปลงสื่อประเภทอื่น ตัวอย่างเช่น การเปลี่ยนไฟล์เสียงให้เป็นนามสกุล MP3

ffmpeg -i **NameOfTheFile.extension** “Output**NameOfTheFile.mp3**”

หมายเหตุ : การใช้คำสั่งนี้**ต้องเปลี่ยนชื่อและนามสกุลไฟล์**ที่ตัวหนาทั้งสองที่เป็นชื่อไฟล์ที่ต้องการแปลง

การแปลงไฟล์ จำนวนมากกว่าหนึ่งไฟล์

for %f IN **(*.*)** DO ffmpeg -i “%f” “Output\%~nf.**mp4**”

- คำสั่งข้างต้นคือการสั่งแปลงไฟล์วีดีโอทุกไฟล์ในโฟลเดอร์ให้เป็น MP4

for %f IN **(*.mov)** DO ffmpeg -i “%f” “Output\%~nf.**mp4**”

- คำสั่งข้างต้นคือการสั่งแปลงเฉพาะไฟล์นามสกุล MOV ทั้งหมดในโฟลเดอร์ให้เป็น MP4
- เปลี่ยนนามสกุลไฟล์ในวงเล็บและด้านท้ายคำสั่ง เมื่อต้องการแปลงสื่อประเภทอื่น ตัวอย่างเช่น การเปลี่ยนไฟล์ WAV เสียงให้เป็นนามสกุล MP3

for %f IN **(*.wav)** DO ffmpeg -i “%f” “Output\%~nf.**mp3**”

- ไฟล์ที่ได้รับการแปลงเรียบร้อยแล้วจะอยู่ในโฟลเดอร์ Output

หากเป็นไฟล์ที่มีขนาดใหญ่มาก สามารถทำซ้ำหลายรอบได้ เพื่อให้บีบอัดไฟล์ให้เล็กลง หากเป็นไฟล์ที่ใหญ่มาก เวลาที่ใช้ในการแปลงไฟล์ก็นานด้วยเช่นกัน

5. การนำข้อมูลขึ้นใน Mukurtu CMS

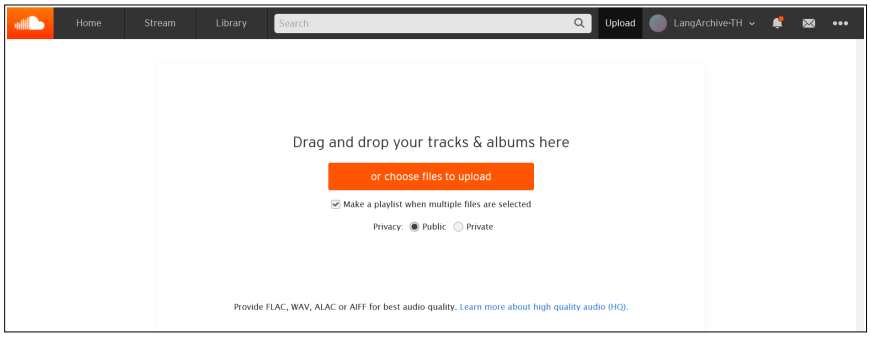
คลังข้อมูลดิจิทัลของกลุ่มชาติพันธุ์ในประเทศไทย LangArchive-TH (www.langarchive-th.org) คือเว็บไซต์คลังข้อมูลออนไลน์ซึ่งใช้ระบบการสร้างและบริหารเว็บไซต์แบบสำเร็จรูป Mukurtu CMS การนำข้อมูลขึ้นหรือ upload ข้อมูลขึ้นคลังทำได้โดยการเพิ่มเนื้อหาหัวข้อภาษาและวัฒนธรรม (Digital Heritage)

การนำข้อมูลประเภทรูปภาพ (JPG) และเอกสาร (PDF) ขึ้นคลังข้อมูลฯ นั้นสามารถ upload ขึ้นเว็บไซต์ได้โดยตรง แต่หากเป็นข้อมูลประเภทเสียง (MP3) และวีดีโอ (MP4) นั้น ผู้ฝากข้อมูลต้อง upload ข้อมูลผ่านเว็บไซต์ SoundCloud (https://soundcloud.com) สำหรับไฟล์เสียง และ Vimeo (https://vimeo.com) สำหรับไฟล์วีดีโอ เนื่องจากพื้นที่บนเว็บไซต์คลังข้อมูลฯ มีขนาดจำกัด

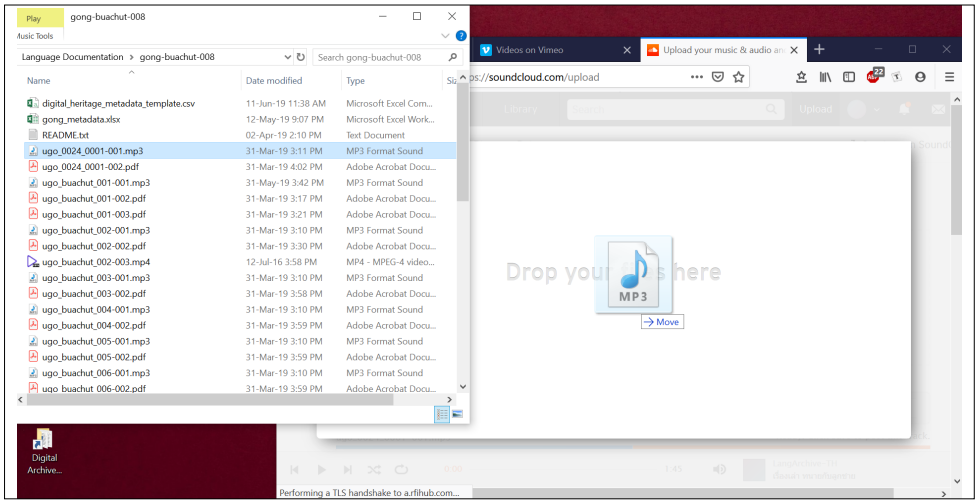
5.1 การ upload ข้อมูลไฟล์เสียงขึ้น SoundCloud

1) ไปที่ https://soundcloud.com และ log in เพื่อเข้าสู่ระบบ

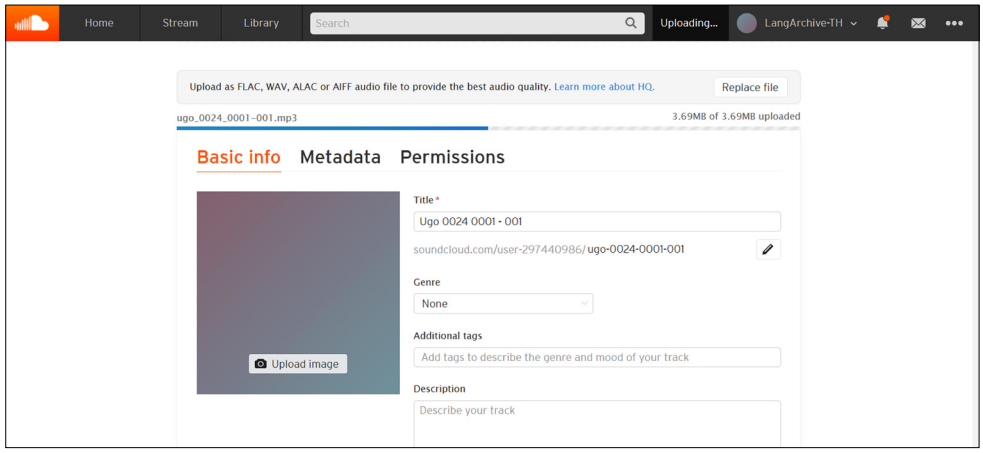
2) หลังจากเข้าสู่ระบบแล้ว ไปที่หน้า ‘Upload’



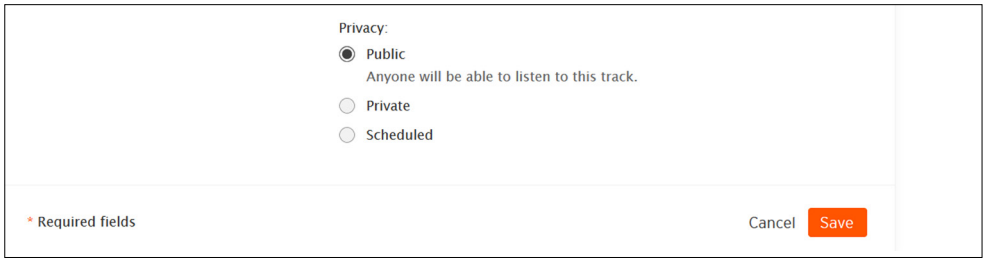
3) คลิกและลากไฟล์เสียงที่ต้องการมาวางที่หน้า ‘Upload’



4) กรอกข้อมูลชื่อ (Title) เลือกประเภท (Genre) ของไฟล์นั้น ๆ และ upload รูปภาพหน้าปก (ถ้ามี)



5) เลือก ‘Public’ สำหรับ ‘Privacy’

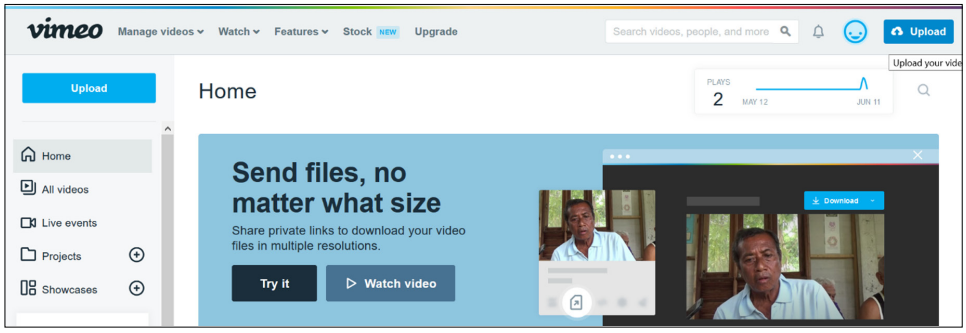


6) รอจนการ upload ไฟล์เสียงเสร็จสิ้น จึงคลิก ‘Save’

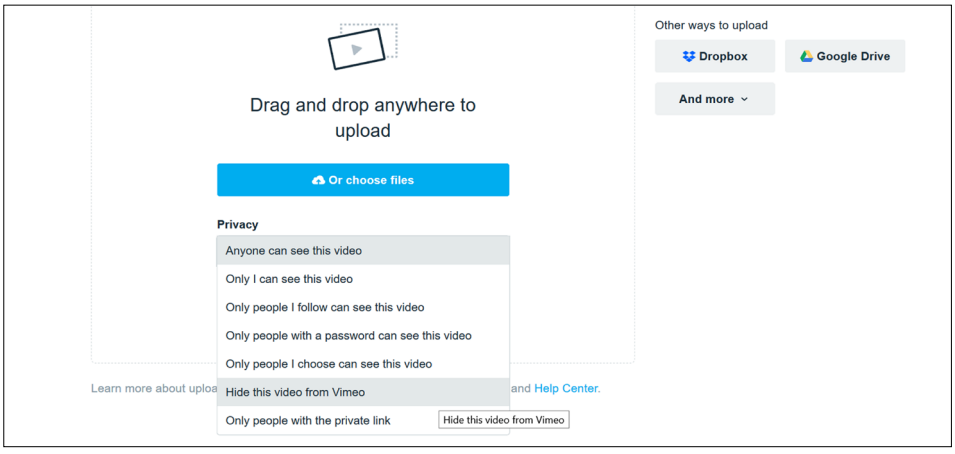
5.2 การ upload ข้อมูลไฟล์วิดีโอขึ้น Vimeo

1) ไปที่ https://vimeo.com และ log in เข้าสู่ระบบ

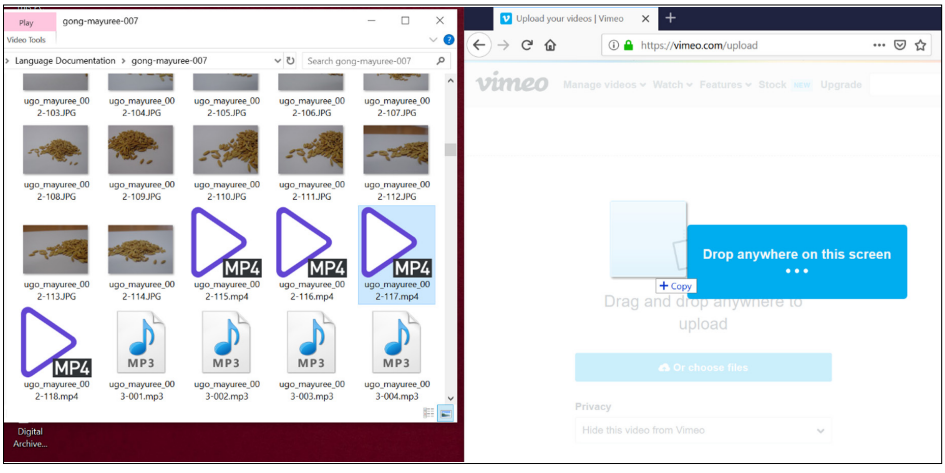
2) ไปที่หน้า ‘Upload’



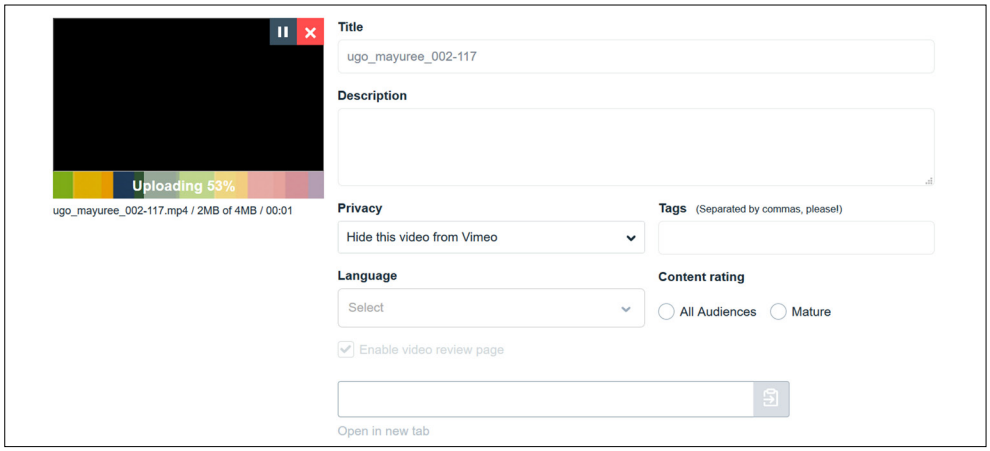
3) เลือก ‘Hide this video from Vimeo’ สำหรับ ‘Privacy’



4) คลิกและลากไฟล์วิดีโอที่ต้องการมาวางที่หน้า ‘Upload’



5) กรอกข้อมูลของไฟล์นั้น ๆ เช่น ชื่อหัวข้อ (Title) แล้วรอจนการ upload เสร็จสิ้น



5.3 การเพิ่มหัวข้อภาษาและวัฒนธรรม (Digital heritage) บน LangArchive-TH

1) เข้าไปที่ http://langarchive-th.org

2) เข้าสู่ระบบ ใส่ชื่อผู้ใช้ และรหัสผ่าน

3) ไปที่ ‘เพิ่มเนื้อหา’ แล้วคลิก ‘+ หัวข้อภาษาและวัฒนธรรม’

4) กรอกข้อมูลสำหรับหัวข้อภาษาและวัฒนธรรม ซึ่งมี 5 ส่วนด้วยกัน ได้แก่ Mukurtu Essentials, Mukurtu Core, Rights and Permission, Additional Metadata และ Relations

การกรอกข้อมูลซึ่งจะเป็น**คำค้นหา**บนคลังข้อมูลฯ เช่น ชื่อบุคคล ประเภทของสื่อ คำสำคัญ เป็นต้น **ให้ใช้เครื่องหมาย semi-colon (;) คั่นระหว่างคำนั้น ๆ** ตัวอย่าง ขั้นตอน; อาหารพื้นบ้าน; กลุ่มชาติพันธุ์ฮากกา

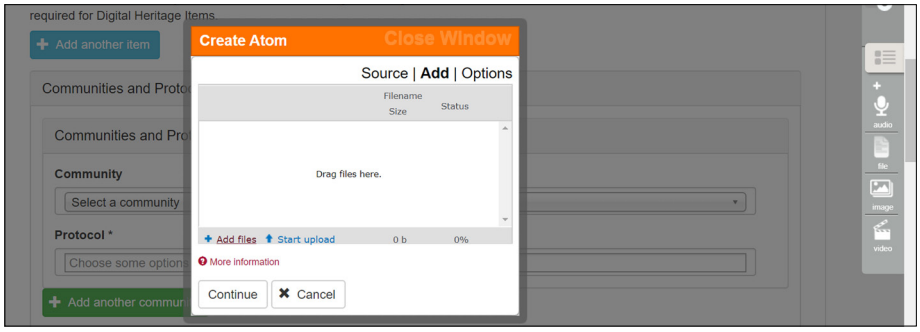
I. Mukurtu Essentials

- **Title**: ชื่อหัวข้อภาษาและวัฒนธรรม
- **Summary**: สรุป
- **Media Assets** (สื่อ) การเพิ่มสื่อในหัวข้อภาษาและวัฒนธรรม ไปที่แถบเครื่องมือด้านขวา แล้วเลือกประเภทของสื่อที่ต้องการ upload

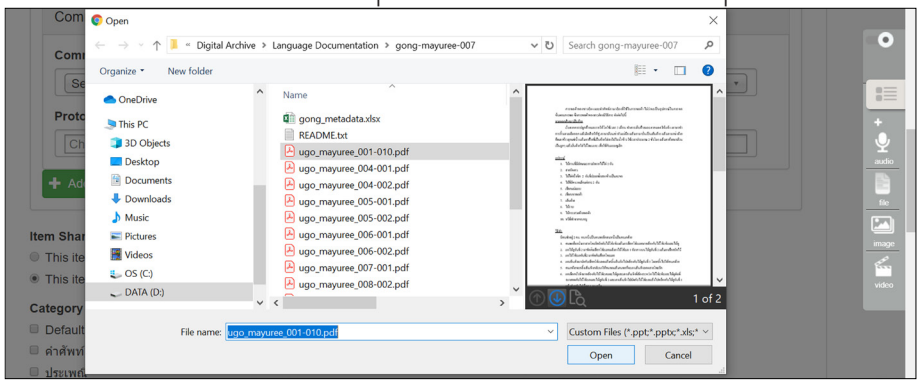
<u>การเพิ่มสื่อเอกสาร (PDF)</u> ไปที่แถบเครื่องมือด้านขวาแล้วคลิก ‘file’

<u>การเพิ่มสื่อข้อมูลรูปภาพ (JPG)</u> คลิก ‘image’

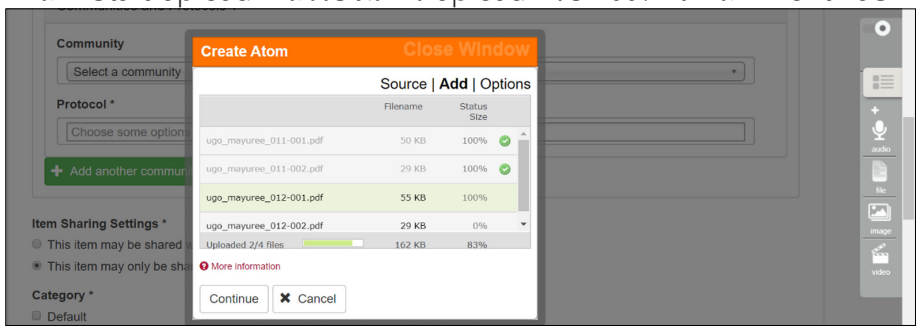
คลิก ‘Add files’



เลือกไฟล์ที่ต้องการ upload ทั้งหมด จากนั้นคลิก ‘Open’

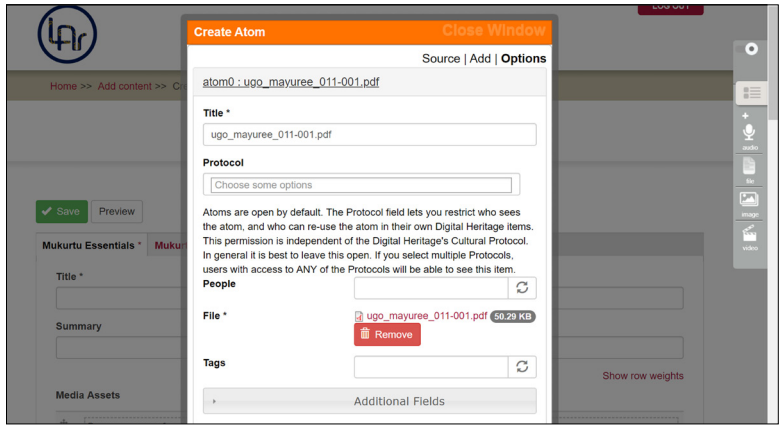


คลิก ‘Start upload’ แล้วรอจนการ upload ครบ 100% จึงคลิก ‘Continue’

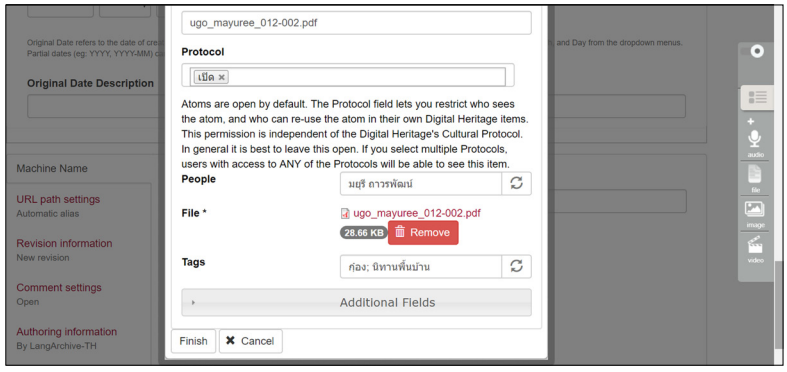


กรอกรายละเอียดสำคัญของไฟล์นั้น ๆ

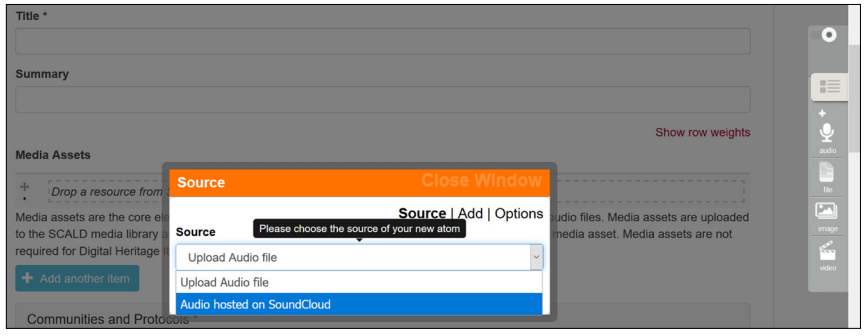
- Title (ชื่อ)
- Protocol (สิทธิการเข้าถึงข้อมูล เปิด หรือ ปิด)
- People (ชื่อผู้มีส่วนร่วมในข้อมูลทั้งหมด)
- Tags (คำค้นหา)



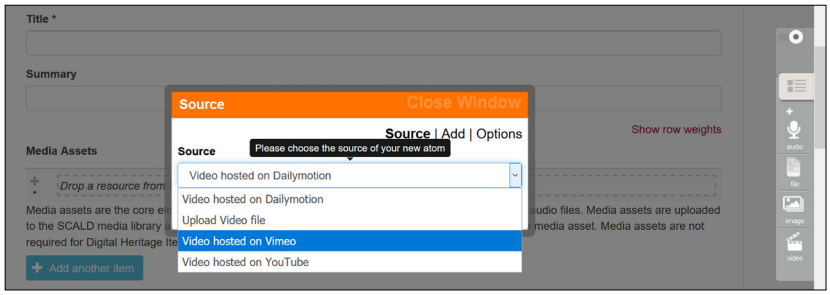
- คลิก ‘Finish’



<u>การเพิ่มสื่อไฟล์เสียงจาก SoundCloud</u> ให้เลือกที่ ‘audio’ จะมีหน้าต่าง ‘Source’ ขึ้นมา ให้เลือก ‘Audio hosted on SoundCloud’

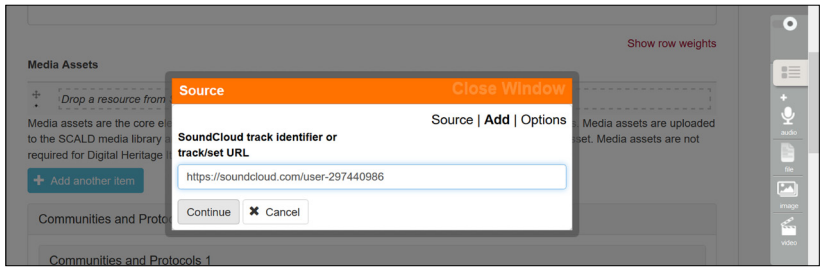


<u>การเพิ่มสื่อไฟล์วีดีโอจาก Vimeo</u> ให้ไปที่ ‘video’ แล้วเลือก ‘Video hosted on Vimeo’



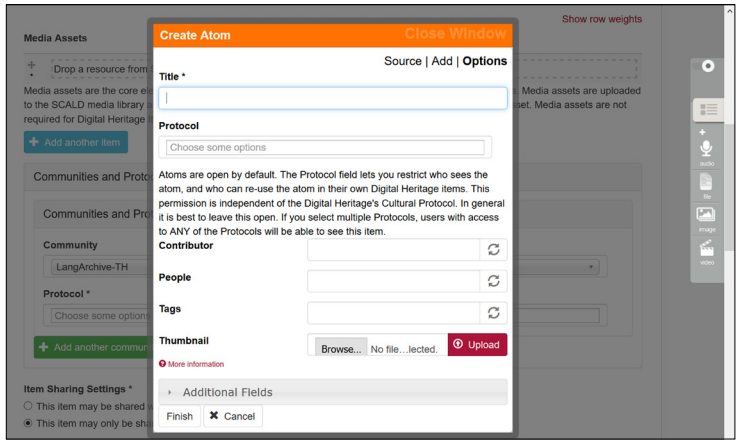
จากนั้นคลิก ‘Continue’

คัดลอก URL จากไฟล์เสียงหรือวีดีโอที่ upload ขึ้น SoundCloud หรือ Vimeo แล้วมาวาง จากนั้นคลิก ‘Continue’



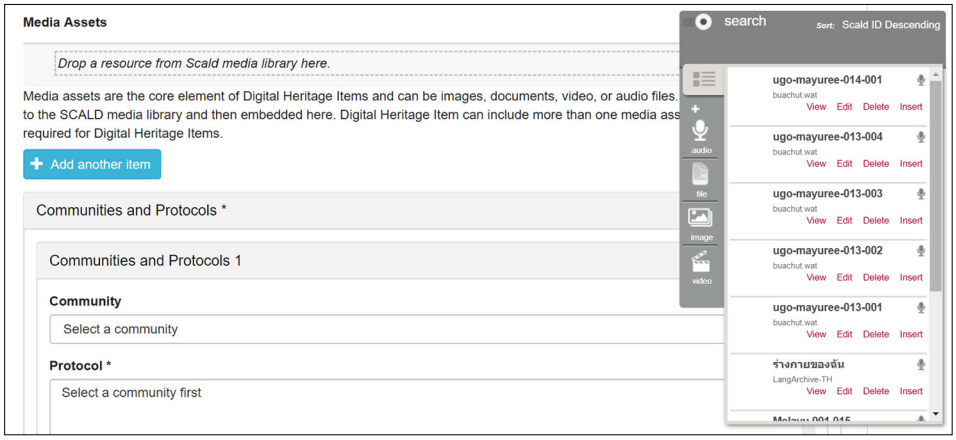
กรอกรายละเอียดข้อมูลของไฟล์นั้น ๆ ได้แก่

- Title (ชื่อหัวข้อ)
- Protocol (เลือกสิทธิการเข้าถึงข้อมูล เปิด หรือ ปิด)
- Contributor (ผู้ให้ข้อมูล เช่น ฉันทนา คำนาค; ยุพา หล่อรุ่งเรืองกุล)
- Tags (คำค้นหา เช่น อาหารพื้นเมือง; อาหารฮากกา; ขั้นตอน; ฮากกา)
- Thumbnail (upload รูปหน้าปกของสื่อนั้น ๆ ถ้ามี)

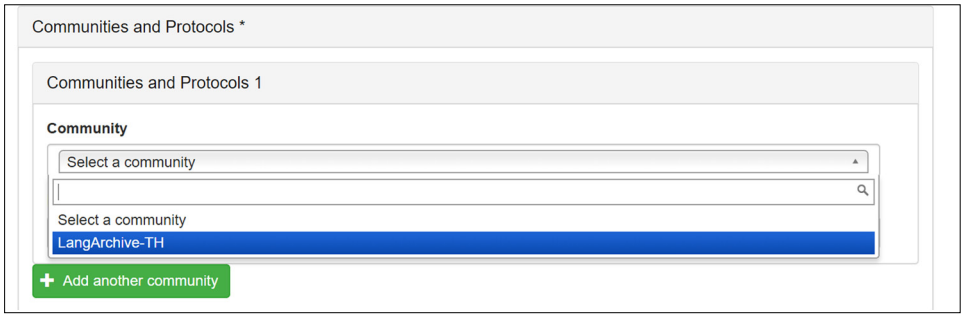


• คลิก ‘Finish

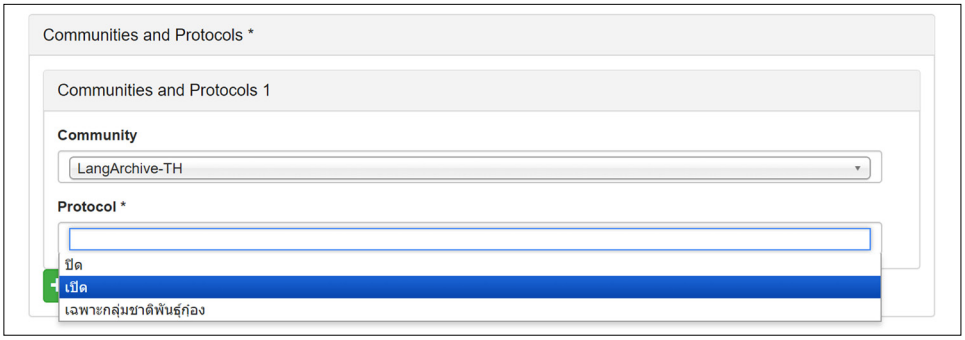
ที่แถบด้านขวา คลิกและลากสื่อที่ได้เพิ่มรายการไว้ แล้ววางลงที่ช่อง Media Assets



• **Communities and Protocols**: กลุ่มผู้มีสิทธิเข้าถึงข้อมูล และสิทธิการเข้าถึงข้อมูล เลือก Community (กลุ่มผู้มีสิทธิเข้าถึงข้อมูล) ได้แก่ LangArchive-TH



เลือก Protocol (สิทธิการเข้าถึงข้อมูล) ของ Community นั้น ได้แก่ เปิดหรือ ปิด หรือสิทธิเฉพาะกลุ่มอื่น ๆ ที่ได้ร้องขอให้ผู้ดูแลคลังข้อมูลฯ สร้างขึ้น



- **Category**: หมวดหมู่ เลือกหมวดหมู่ของข้อมูล เช่น พจนศิลป์ การละเล่น เรื่องเล่า เป็นต้น
- **Creator**: ผู้มีส่วนร่วมในการจัดเก็บข้อมูล เช่น ศิริเพ็ญ อึ้งสิทธิพูนพร
- **Contributor**: ผู้ให้ข้อมูล หรือ ผู้ฝากข้อมูล เช่น ฉันทนา คำนาค; ยุพา หล่อรุ่งเรืองกุล
- **Original Date**: วันที่ในการเก็บข้อมูล

II. Mukurtu Core

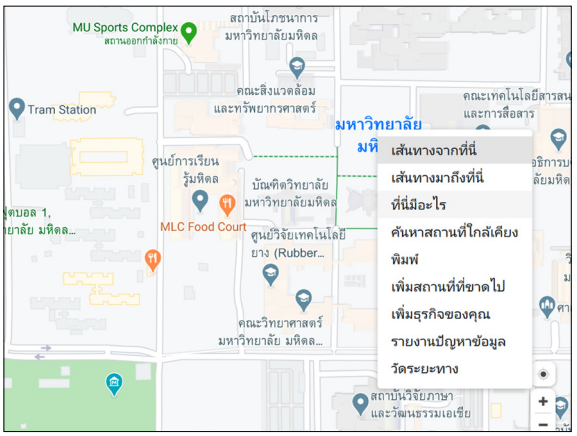
- Description: รายละเอียดทั้งหมดของข้อมูล
- Keywords: คำสำคัญสำหรับการค้นหา เช่น อาหารพื้นเมือง; อาหารฮากกา; เพลงพื้นบ้าน

III. Rights and Permissions

- **Publisher**: ชื่อผู้เผยแพร่หัวข้อภาษาและวัฒนธรรมบนคลังข้อมูล
- **Rights**: ผู้มีสิทธิในข้อมูล
- **Licensing Options**: เลือก Attribution-NonCommercial 4.0 International

IV. Additional Metadata

- Format: สกุลไฟล์ เช่น MP3; MP4; PDF
- Type: ประเภทของสื่อ เช่น ไฟล์เสียง; วิดีโอ; เอกสาร
- Language: ภาษาของข้อมูล หรือภาษาที่ใช้ในการทำงาน เช่น ฮากกา; ไทย
- People: บุคคลที่เกี่ยวข้องทั้งหมด
- Location: สถานที่ ใส่ Latitude และ Longitude การหา Latitude และ Longitude สามารถค้นหาได้จาก **Google Maps**[2] คลิกขวาที่สถานที่นั้น แล้วเลือก **ที่นี่มีอะไร** หรือ **What's here?** ตัวเลข Latitude, Longitude จะปรากฏอยู่ด้านล่าง



- Location Description: สถานที่ในการเก็บข้อมูล

[2] วิธีนี้ใช้กับ Google Maps บนคอมพิวเตอร์เท่านั้น

V. Relations

• **Collection**: กลุ่มชาติพันธุ์

• **Related Digital Heritage Items**: หัวข้อภาษาและวัฒนธรรมอื่นที่เกี่ยวข้อง ถ้ามี

5) เมื่อกรอกรายละเอียดของข้อมูลเรียบร้อยแล้ว คลิก 'Save'

สรุปการจัดการข้อมูล

การจัดการข้อมูลภาษาและวัฒนธรรมของกลุ่มชาติพันธุ์ มีขั้นตอนค่อนข้างมาก ดังนั้นจึงควรวางแผนตั้งแต่เตรียมเก็บบันทึกข้อมูล โดยมีผู้จดรายละเอียดข้อมูล (metadata) ตามแบบฟอร์มที่เตรียมไว้ (ดูแบบฟอร์มที่ภาคผนวก) เมื่อกลับจากภาคสนาม ควรกรอกรายละเอียดข้อมูลในไฟล์ excel ทันที

การเตรียมไฟล์ก่อน upload บนคลังข้อมูล ควรตรวจสอบขนาดและคัดเลือกไฟล์ที่จะนำบรรจุเข้าคลังข้อมูล โดยเฉพาะไฟล์รูปภาพที่มีจำนวนมาก เลือกเฉพาะภาพที่เกี่ยวข้องกับเนื้อหาหรือแสดงลำดับเหตุการณ์ได้ชัดเจน ส่วนไฟล์เสียงหรือวิดีโอ ควรแปลงไฟล์ให้มีขนาดเล็กลงให้เหมาะสมกับการ upload ขึ้นคลังฯ

เมื่อคัดเลือกไฟล์ที่จะบรรจุขึ้นคลังฯ ไว้เป็นแต่ละโฟลเดอร์พร้อมเผยแพร่แล้ว ตรวจสอบการตั้งชื่อไฟล์ให้เป็นระบบ ไฟล์วิดีโอ สามารถuploadที่ vimeo.com ส่วนไฟล์เสียง upload ที่ soundcloud.com ไว้ก่อนได้ เป็นการเตรียมพร้อม

ขั้นตอนสุดท้าย เมื่อกรอกข้อมูลขึ้น langarchive-th.org จะมีข้อมูลพร้อม โดยสามารถคัดลอกเนื้อหาจากไฟล์ excel (metadata) มาใส่ตามหัวข้อต่าง ๆ ได้ครบถ้วน พร้อมกับ upload ไฟล์ประเภทต่าง ๆ ที่ได้เตรียมไว้แล้ว

6. คำถามที่พบบ่อย

ถาม : อุปกรณ์ที่ใช้ในการบันทึกข้อมูล ควรเป็นแบบใดบ้าง ต้องใช้คุณภาพสูงเพียงใด

ตอบ : ขึ้นอยู่กับจุดประสงค์ในการเก็บข้อมูล กล่าวคือ นักวิชาการ และสมาชิกในชุมชน มีจุดประสงค์ในการเก็บข้อมูลที่แตกต่างกัน

ในกรณีของนักวิชาการซึ่งจัดเก็บข้อมูลเพื่อนำไปวิเคราะห์เพิ่มเติม เช่น นักภาษาศาสตร์ที่เน้นความคมชัดของเสียง จำเป็นต้องใช้อุปกรณ์ที่มีคุณภาพดีหรือคุณภาพสูงเท่าที่จะหาได้ เพื่อให้ได้ข้อมูลที่มีคุณภาพสูงสุด เหมาะสมสำหรับเก็บบันทึกเสียง จึงจะสามารถนำไปถ่ายถอดเสียงและวิเคราะห์เรื่องเสียงได้ต่อไป อีกทั้งยังเป็นการรักษาภาษาของกลุ่มชาติพันธุ์ที่อยู่ในภาวะวิกฤตอีกด้วย เพราะอาจจะหาผู้ให้ข้อมูลที่ดี ออกเสียงได้ชัดเจนค่อนข้างยาก ซึ่งข้อมูลเหล่านี้นับว่าเป็นสิ่งที่มีค่ามากต้องบันทึกด้วยเครื่องมือที่มีคุณภาพสูงและเก็บรักษาไว้ได้นาน อุปกรณ์สำคัญสำหรับการบันทึกเสียงได้แก่ ไมโครโฟน ควรใช้ประเภทของไมโครโฟนให้เหมาะกับการบันทึกเสียง เช่น บันทึกเสียงคนเดียว หรือบทสนทนาสองคน หรือหลายคน ส่วนการถ่ายภาพเคลื่อนไหว หรือสถานการณ์ที่ต้องจัดเตรียมคน สถานที่ ควรใช้เครื่องบันทึกวิดีโอที่สามารถต่อไมโครโฟนได้เช่นกัน และอีกอุปกรณ์ที่สำคัญคือขาตั้งกล้อง เพื่อไม่ให้ภาพสั่นไหว และควรจัดภาพให้เห็นองค์ประกอบครบถ้วน

แต่ในกรณีที่ชุมชนเป็นผู้บันทึกด้วยตนเอง ไม่สามารถซื้ออุปกรณ์เพื่อบันทึกคุณภาพระดับสูงได้ และมีจุดประสงค์การบันทึกเพื่อเก็บรักษาและถ่ายทอดองค์ความรู้ท้องถิ่น อนุโลมให้ใช้อุปกรณ์เท่าที่มี เช่น กล้องจากโทรศัพท์มือถือ ซึ่งปัจจุบันส่วนใหญ่มีคุณภาพที่ค่อนข้างดี ความคมชัดค่อนข้างสูง แต่สิ่งสำคัญที่ควรคำนึงถึงในการบันทึกข้อมูลทุกครั้งคือ ภาพไม่สั่นไหวเสียงมีความชัดเจน ไม่มีเสียงรบกวนจนฟังไม่เข้าใจ หากต้องการบันทึกสื่อประเภทวิดีโอด้วยกล้องมือถือ ควรมีตัวช่วยเป็นขาตั้งกล้องเพื่อป้องกันภาพสั่น ทั้งนี้ขอให้คำนึงถึงเทคนิคการบันทึกสื่อประเภทต่าง ๆ เป็นสำคัญ

ถาม : คลังข้อมูลรับฝากข้อมูลประเภทใดและหัวข้อใดบ้าง

ตอบ : ข้อมูลที่รับฝาก เป็นข้อมูลที่เกี่ยวข้องกับภาษาและวัฒนธรรมของกลุ่มชาติพันธุ์ในประเทศไทย รวมถึงภาษาไทยถิ่น โดยมีหัวข้อและประเด็นที่ชัดเจน ซึ่งชุมชนเห็นว่ามีความสำคัญหรือใกล้สูญหาย คุณภาพของสื่อคมชัด เช่น ไฟล์เสียง ต้องไม่มีเสียงรบกวนซึ่งกระทบกับเนื้อหาหลักของข้อมูล ไฟล์วิดีโอมีความคมชัดพอและไม่สั่นไหว มีเนื้อหาที่สามารถสื่อได้ตรงตามเนื้อหา มีความยาวประมาณ 10 นาทีต่อ 1 ไฟล์ มีขนาดไม่เกิน 2 GB และไฟล์ภาพมีความชัดเจน ไม่เบลอ สื่อความหมายที่สอดคล้องกับหัวข้อและประเด็นที่ต้องการ มีขนาดไม่เกิน 2 MB ต่อไฟล์ เพื่อความสะดวกต่อการupload ดาวน์โหลด และการรับชมสื่อออนไลน์ ยิ่งขนาดไฟล์ใหญ่มากจะใช้เวลาในการจัดการไฟล์มาก และต้องคำนึงถึงความเร็วของอินเทอร์เน็ตด้วย เมื่อuploadขึ้นเว็บคลังข้อมูลแล้ว ผู้รับชม

อาจมีปัญหาในขณะที่ชมวิดีโอที่เป็นไฟล์ขนาดใหญ่เช่นกัน นอกจากนี้ประเด็นสำคัญในการเผยแพร่ข้อมูลที่ต้องคำนึงถึงให้มากคือ ข้อมูลที่ต้องการฝากกับคลังข้อมูลออนไลน์ต้องได้รับการอนุญาตจากผู้เกี่ยวข้องทั้งหมด โดยเฉพาะเจ้าของข้อมูล

ถาม : การแปลงไฟล์ สามารถใช้โปรแกรมอื่นนอกจาก FFMPEG ได้หรือไม่

ตอบ : การแปลงไฟล์นั้นสามารถใช้โปรแกรมอื่นนอกจาก FFMPEG ได้ หากแต่โปรแกรมส่วนใหญ่มีค่าใช้จ่าย หรือจำกัดเวลาการทดลองใช้ อย่างไรก็ตาม ขอให้คำนึงถึงคุณภาพที่ได้จากผลิตภัณฑ์นั้น ๆ และให้เป็นไปตามข้อกำหนดของคลังข้อมูลฯ

ถาม : ทำไมต้องจัดระบบชื่อไฟล์ข้อมูล

ตอบ : การจัดระบบไฟล์ข้อมูลจะช่วยในการจัดการ จัดกลุ่ม จัดเรียงไฟล์ที่มีปริมาณมากอย่างเป็นระบบ เพื่อระบุข้อมูลเบื้องต้นของไฟล์นั้น ๆ ซึ่งเป็นประโยชน์ต่อการสืบค้น เช่น มีการระบุชื่อย่อที่เป็นของภาษาแต่ละภาษา (ดูจาก ethnologue) เช่น ภาษาบีซู ใช้รหัสภาษา bzi ภาษาไทยใช้รหัสภาษา tha ภาษาเขมรถิ่นไทยใช้รหัสภาษา kxm เป็นต้น และชื่อของผู้ที่ฝากข้อมูลของแต่ละไฟล์ก็จะสามารถหาได้ง่ายขึ้น อย่างน้อยก็จัดอยู่ในกลุ่มภาษาเดียวกัน

ถาม : ทำอย่างไรถ้ารายละเอียดของข้อมูล (Metadata) มีไม่ครบตามที่คลังข้อมูลฯ กำหนด

ตอบ : หัวข้อตามรายละเอียดข้อมูล เป็นเพียงแนวทางในการกรอกข้อมูล หากมีไม่ครบขอให้ใส่ข้อมูลสำคัญให้ได้มากที่สุดเท่าที่จะหาได้ เช่น ผู้ให้ข้อมูล ผู้ผลิต เนื้อหา ปีที่บันทึกข้อมูล เนื้อหาสรุปเป็นเรื่องเกี่ยวกับอะไร เป็นต้น ซึ่งบางหัวข้ออาจไม่จำเป็นต้องใส่ แต่ถ้าเป็นการเก็บบันทึกข้อมูลใหม่ ควรใส่รายละเอียดข้อมูลให้ครบตามหัวข้อที่มีในแบบฟอร์ม เพื่อประโยชน์ของผู้ใช้ข้อมูลและการอ้างอิงที่ถูกต้อง

ความหมายและรายละเอียดเพิ่มเติมสำหรับ Metadata

1. **GUID** หมายถึง เลขที่สร้างขึ้นเพื่อระบุรหัส bundle ใน 1 bundle จะมีกี่ไฟล์ก็ได้ แต่ละไฟล์มีเลข GUID เลขเดียว (bundle มีความหมายเดียวกับ digital heritage)

2. **Title** เปรียบเสมือนชื่อเรื่อง ควรตั้งให้สั้น กระชับ เช่น song หรือจะระบุชื่อเพลง rammana song ก็ได้

3. **Community = Archive user** หมายถึงผู้เข้ามาใช้คลังข้อมูล หมายถึงการอนุญาตให้ผู้ใช้งานเข้าถึงข้อมูล เช่น ถ้าไฟล์นั้นจำกัดเฉพาะคนที่ลงทะเบียนสมัครเป็นสมาชิก (registered and log in) เท่านั้นก็ต้องตั้งค่าเฉพาะคนที่ registered ถ้าให้ทุกคนเข้าถึงได้ ก็ตั้งเป็น open access แต่ถ้าเป็น administrator ก็สามารถเข้า Community นี้ได้หมด

4. **Type** ประเภทของไฟล์ เช่น Video; Image; Text.... แนะนำให้ใช้อย่างเป็นระบบสม่ำเสมอ เช่น อักษรตัวแรกเป็นตัวพิมพ์ใหญ่ ก็ใช้ให้เหมือนกันทุกคำเพื่อความสม่ำเสมอ

5. **Original date** เป็นวันที่ผลิตไฟล์นั้น ๆ ถ้ารู้วันที่แน่นอนก็ให้ใส่ทั้ง วัน เดือน ปี ถ้าไม่รู้วันที่แน่นอนก็ใส่เฉพาะเท่าที่จำได้ เช่น เดือน ปี หรือ ปี อย่างเดียว แต่ต้องใช้ระบบเดียวกันให้สม่ำเสมอ ในที่นี้จะใช้เป็น DDMMYYYY เป็นปี ค.ศ.

6. **People** ใส่ชื่อให้หมดทุกคนที่เกี่ยวข้อง ตั้งแต่ผู้ผลิตไฟล์ ผู้บันทึก ผู้ให้ข้อมูล เป็นต้น หรือใส่เป็นชื่อกลุ่มชาติพันธุ์ เพื่อการคั่นชื่อแต่ละคนให้ใช้ semi-colon (;) คั่นแต่ละชื่อ ทุกคอลัมภ์ที่ต้องการจะการแยกคำ แยกคน ... ใช้เฉพาะเครื่องหมาย semi-colon (;) ถ้าใส่เครื่องหมาย colon (:) ระบบจะไม่ทราบค่า

7. **Related Dictionary** ใส่ในกรณีที่มีพจนานุกรมเชื่อมโยงกัน

8. **Summary** ให้ใส่รายละเอียดมากกว่าหัวข้อ บอกให้คนรู้ว่าเรื่องนี้เกี่ยวกับอะไร แต่ไม่ควรยาวเกินไป เป็นเพียงสรุปเนื้อหาสั้น ๆ ให้ผู้ใช้ข้อมูลทราบคร่าว ๆ

9. **Transcription** อธิบายรายละเอียดได้มากเท่าที่อยากจะอธิบาย (แต่ไม่จำเป็นต้องใส่)

10. **Cultural narrative** จะเป็นข้อความที่นอกเหนือจาก Description โดยจะเป็นวัฒนธรรมเฉพาะของกลุ่มชาติพันธุ์นั้น ๆ ที่แตกต่างจากคนส่วนใหญ่หรือเป็นสิ่งที่ไม่พบในกลุ่มอื่น

สำหรับหัวข้อ 9 และ 10 ไม่จำเป็นต้องใส่ สามารถระบุรายละเอียดทุกอย่างได้ใน Description

11. **Category** ประเภทของข้อมูลอย่างกว้าง เช่น เพลง การละเล่น พิธีกรรม ขั้นตอน สื่อการสอน เรื่องเล่า เป็นต้น ถ้าใช้ภาษาอังกฤษให้ใช้ตัวพิมพ์ใหญ่ตัวแรก Songs เป็นต้น

12. **Contributor** เป็นเจ้าของข้อมูล, ผู้ให้ข้อมูล (ผู้บอกภาษา), เจ้าของภาษาผู้นำข้อมูล เข้าคลังข้อมูล หรือเป็นผู้ฝากข้อมูลแต่ไม่ได้ upload ด้วยตนเอง

13. **Creator** ทุกคนที่เกี่ยวข้องกับเนื้อหา อาจจะไม่ได้เป็นเจ้าของภาษา แต่ถ้ามีส่วนร่วมในขั้นตอนใดขั้นตอนหนึ่งจะเป็น Creator ได้ เช่น บันทึกเสียง ถ่ายถอดเสียง รวมทั้งการผลิตไฟล์นั้น ๆ และให้คั่นชื่อแต่ล่ะบุคคลด้วยเครื่องหมาย “;”

14. **Location** เป็นสถานที่ผลิตไฟล์นั้น ๆ โดยใส่เป็น Latitude / Longitude สามารถหาได้จาก google map

15. **Original date description** เป็นวันที่ มีการประมวลผล ไฟล์นั้น ๆ เช่น บันทึกวิดีโอมาแล้ว 2 เดือนแล้วค่อยมาถอดเสียง ให้ระบุวันที่ถอดเสียง

16. **Format** เป็นประเภทของไฟล์นั้น ๆ หรือ นามสกุลของไฟล์นั้นๆ เช่น MP4; MP3; JPG; PDF ควรใส่เป็นตัวพิมพ์ใหญ่ทั้งหมด

17. **Identifier** = GUID

18. **Language** ควรใส่เป็นภาษาของกลุ่มชาติพันธุ์ (ภาษาที่บันทึกวิดีโอ หรือเสียง) แต่สามารถระบุได้ว่า เนื้อหาเป็นภาษาอะไร ภาษาอะไรใช้ในการสื่อสาร

19. **Licensing option** ต้องระบุว่า non-commercial

- file ที่ upload ผ่าน Vimeo สามารถตั้งค่าป้องกันไม่ให้คนอื่น download ได้ แต่โดยปกติการ upload ขึ้น Mukurtu ก็ไม่สามารถ download ได้เช่นกัน
- ลิขสิทธิ์ ต้องเป็นของเจ้าของภาษา จึงต้องมีการลงนามในใบยินยอม consent form

20. **Media asset** ข้อนี้ไม่จำเป็นต้องระบุ

21. **Publisher** เป็นผู้ที่โหลดข้อมูลขึ้นเว็บ ซึ่งบางครั้ง Contributor เป็นคนละคนกับ Publisher

22. **Rights** อธิบายว่าใครเป็นคนที่สามารถเข้าถึงข้อมูล เพราะอะไร แต่ในโครงการนี้อาจจะไม่จำเป็น เพราะมีความประสงค์จะเผยแพร่ข้อมูลอยู่แล้ว

23. **Source** หัวข้อนี้ไม่จำเป็นต้องระบุ

24. **Keywords** เป็นคำสำคัญเพื่อให้การค้นหาเข้าถึงได้ง่าย สามารถระบุ Keywords ได้หลายคำ โดยคั่นด้วยเครื่องหมาย ";"

25. **Protocol** ให้ระบุว่า Open หรือ close

26. **Username** เป็นชื่อของคนที่ upload file เช่น Siripen; vf.4

27. **Depositor** คือคนที่ประสานระหว่างชุมชนกับ Archive (เป็นผู้ที่ใส่ข้อมูลใน Archive)

- Archivist เป็นผู้บริหารจัดการข้อมูลในคลังข้อมูล และดูแลคลังข้อมูล
- Legacy data เป็นข้อมูลภาคสนามที่บันทึกโดยบุคคลอื่น แต่ต้องการเก็บไว้ใน archive จะเป็นข้อมูลที่ถ่ายถอดเสียงแล้วหรือไม่ก็ได้

ถาม : การจำกัดสิทธิการเข้าถึงข้อมูล ทำได้อย่างไรบ้าง

ตอบ : คลังข้อมูลฯ จำกัดสิทธิการเข้าถึง หรือ protocol พื้นฐาน 2 ประเภทด้วยกัน ได้แก่ **เปิดและปิด** สมาชิกทั่วไปจะเข้ารับชมข้อมูลที่มีสิทธิการเข้าถึงแบบเปิดได้เท่านั้น ส่วนข้อมูลที่มีสิทธิการเข้าถึงข้อมูลแบบปิดนั้น สมาชิกต้องขออนุญาต**ผู้ให้ข้อมูล** ผ่านผู้ดูแลคลังข้อมูลฯ เพื่อเข้าถึงข้อมูลนั้น ๆ เป็นรายกรณีไป ซึ่งผู้ดูแลคลังข้อมูลจะถามถึงเหตุผลในการขอเข้าถึงข้อมูล

หากผู้ฝากข้อมูลต้องการจำกัดสิทธิการเข้าถึงข้อมูล สามารถติดต่อผู้ดูแลคลังข้อมูลให้เพิ่ม protocol หรือสิทธิการเข้าถึงเฉพาะสำหรับข้อมูลนั้น ๆ ได้ เช่น เฉพาะกลุ่มชาติพันธุ์ก่อง เป็นต้น

7. ประมวลผลที่เกิดขึ้นระหว่างทำโครงการและสิ่งที่จะทำในอนาคต

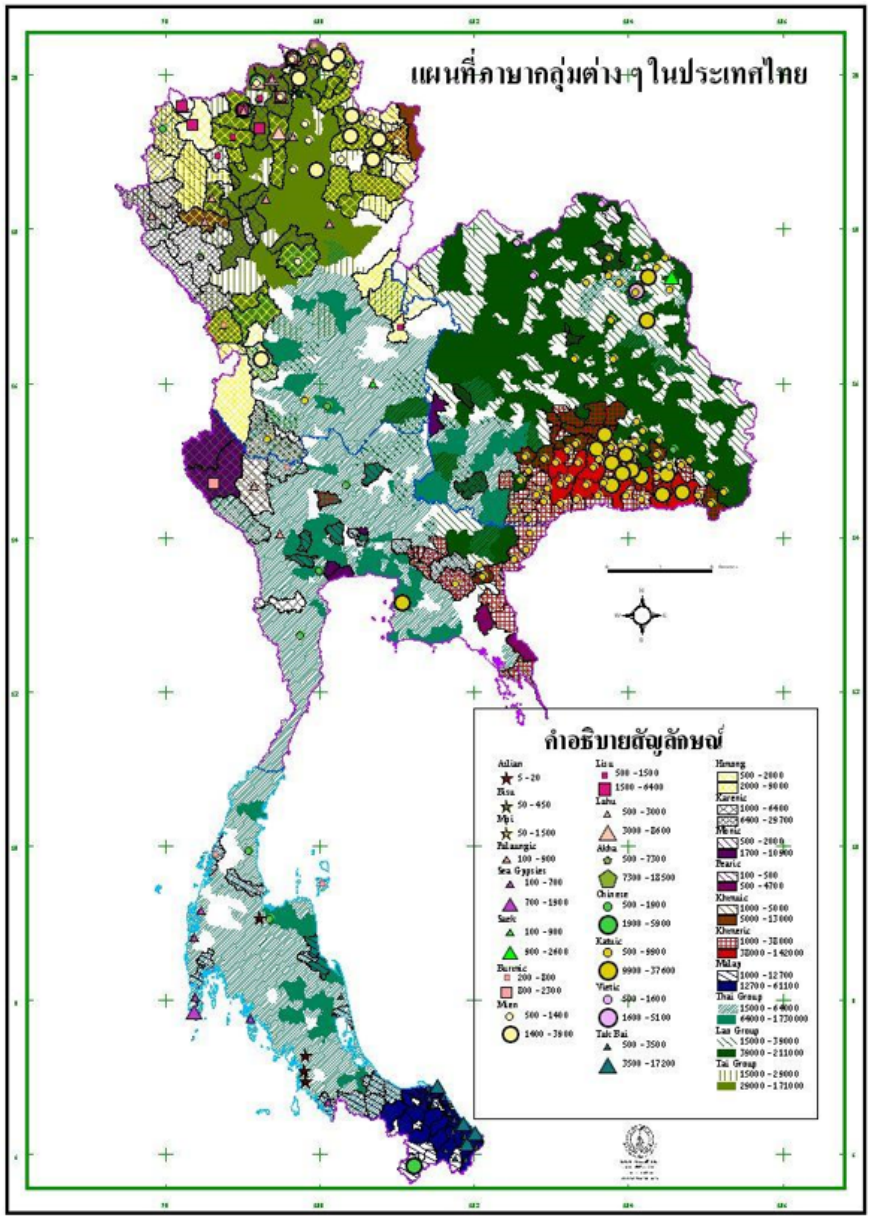
ประมวลความรู้จากการเสวนา “**การจัดทำและนำเสนอคลังข้อมูลดิจิทัลออนไลน์**” ซึ่งเป็นส่วนหนึ่งของกิจกรรมในโครงการฯ ได้จัดขึ้นเมื่อวันที่ 27 พฤศจิกายน 2562 ที่สถาบันวิจัยภาษาและวัฒนธรรมเอเชีย มหาวิทยาลัยมหิดล

โครงการคลังข้อมูลดิจิทัลของกลุ่มชาติพันธุ์ : ส่งเสริม การปกป้อง และดำรงรักษาความหลากหลายทางภาษาและวัฒนธรรมในประเทศไทย เป็นความร่วมมือระหว่างสถาบันวิจัยภาษาและวัฒนธรรมเอเชีย มหาวิทยาลัยมหิดล และ SOAS University of London ซึ่งมีหน่วยงานที่ทำงานด้านคลังข้อมูลชาติพันธุ์ที่อยู่ในภาวะวิกฤตทั่วโลก ชื่อว่า ELAR (Endangered Languages Archive) เนื้อหาในการเสวนามาจากภาคีเครือข่ายที่ทำงานด้านคลังข้อมูลในประเทศไทย ได้แก่ หอภาพยนตร์ (องค์การมหาชน) ศูนย์มานุษยวิทยาสิรินธร (องค์การมหาชน), ศูนย์เทคโนโลยีอีเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ (NECTEC) และการทำคลังข้อมูลดิจิทัลด้านภาษาและวัฒนธรรมของกลุ่มชาติพันธุ์ในประเทศไทย ของสถาบันวิจัยภาษาและวัฒนธรรมเอเชีย มหาวิทยาลัยมหิดล ซึ่งเป็นการสรุปภารกิจของแต่ละหน่วยงานว่ามีบทบาททำหน้าที่อย่างไรในการจัดทำคลังข้อมูลดิจิทัลและได้รับทุนสนับสนุนที่ชื่อว่า Newton Institutional Links ซึ่งเป็นการให้งบประมาณร่วมกันระหว่าง British Council (ฝ่ายอังกฤษ) และ สำนักงานกองทุนสนับสนุนการวิจัย (ฝ่ายไทย) ปัจจุบันเปลี่ยนเป็น สกสว. แล้ว มีระยะเวลาในการดำเนินโครงการ 1 ปี

1. สถาบันวิจัยภาษาและวัฒนธรรมเอเชีย มหาวิทยาลัยมหิดล

สถาบันวิจัยภาษาและวัฒนธรรมเอเชีย ได้มีการทำงานร่วมกับกลุ่มชาติพันธุ์มาอย่างต่อเนื่อง โดยปัจจุบันได้ทำงานด้านการอนุรักษ์ ฟื้นฟูภาษามากกว่า 20 กลุ่มชาติพันธุ์ในประเทศไทย เป็นระยะเวลาเกือบ 20 ปี นำโดยท่าน ศาสตราจารย์เกียรติคุณ ดร. สุวิไล เปรมศรีรัตน์

ผู้ช่วยศาสตราจารย์ ดร.ศิริเพ็ญ อึ้งสิทธิพูนพร (หัวหน้าโครงการฯ) ได้ทำงานร่วมกับท่านศาสตราจารย์เกียรติคุณ ดร.สุวิไล เปรมศรีรัตน์ มาเป็นระยะเวลานาน ได้รับประสบการณ์และคำแนะนำที่ดีมาโดยตลอด นำมาซึ่งการขอทุนวิจัยระหว่างประเทศ เนื่องจากมีเครือข่ายต่างประเทศและสถาบันการศึกษาที่ทำงานด้านการบันทึกข้อมูลภาษาและวัฒนธรรมของกลุ่มชาติพันธุ์ที่อยู่ในภาวะวิกฤต



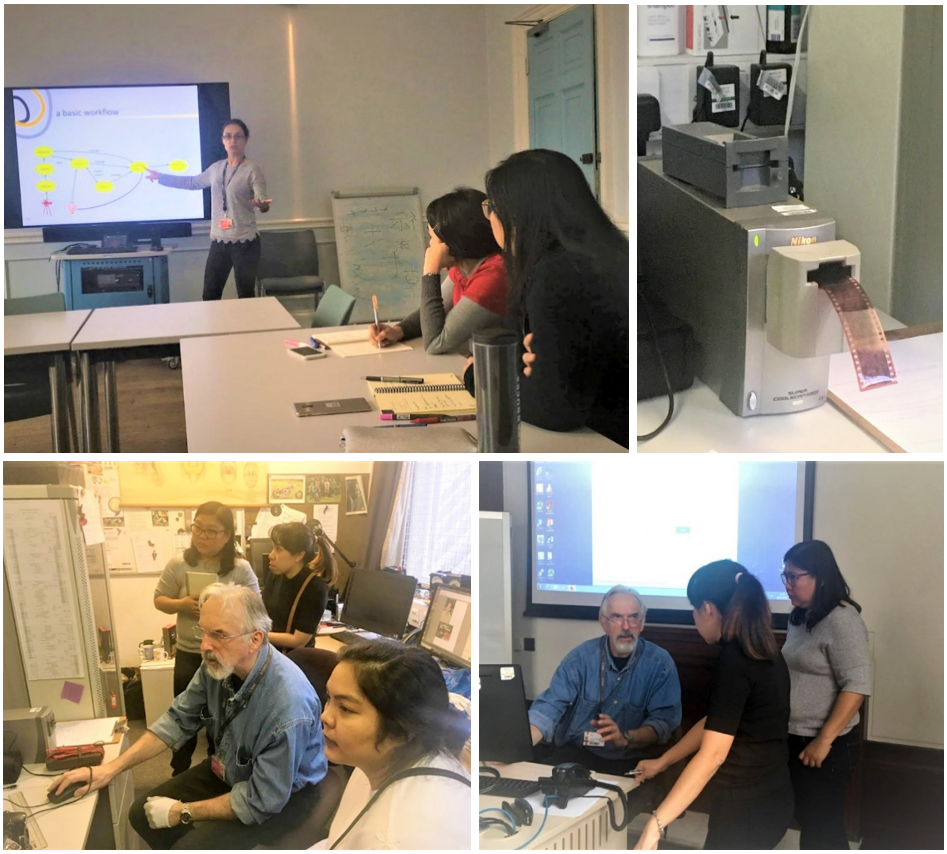
ภาพที่ 9 : แผนที่ภาษากลุ่มต่างๆในประเทศไทย (สุวิไล เปรมศรีรัตน์ และคณะ 2547:16)

จากแผนที่ภาษาในประเทศไทย พบว่า ประเทศไทยมีความหลากหลายทางภาษาและวัฒนธรรมมากมายกว่า 70 กลุ่ม ซึ่งมีภาษาที่อยู่ในภาวะวิกฤต จำนวน 15 กลุ่ม กระจายไปในภูมิภาคต่าง ๆ หากปล่อยทิ้งไว้ภาษา วัฒนธรรม รวมถึงความคิด และภูมิปัญญาก็จะสูญหายไป เพราะฉะนั้นเราจึงสนับสนุนให้กลุ่มชาติพันธุ์ได้เห็นความสำคัญ โดยใช้เครื่องมือในบันทึกภาษาและภูมิปัญญาท้องถิ่นของตนเอง ซึ่งมีหลายเครื่องมือ เช่น ระบบเขียนภาษาชาติพันธุ์ เนื่องจาก

บางภาษาไม่มีระบบเขียนมาก่อน เราจึงได้ทำงานร่วมกับเจ้าของภาษาพัฒนาระบบเขียนขึ้นมา โดยใช้มีวิธีการด้านภาษาศาสตร์ที่จะช่วยให้เจ้าของภาษาสามารถบันทึกภาษาของตนเองได้ใกล้เคียงมากที่สุด

วิธีการทำงานของโครงการ “คลังข้อมูลดิจิทัลของกลุ่มชาติพันธุ์: ส่งเสริมการปกป้องและดำรงรักษา ความหลากหลายทางภาษาและวัฒนธรรมในประเทศไทย”

กิจกรรมแรก คือ การเรียนรู้การจัดทำคลังข้อมูลกับเครือข่ายการทำงานจากต่างประเทศ เพราะมีประสบการณ์ด้านการจัดทำคลังข้อมูล แต่เดิมที่เรามีประสบการณ์การทำงานกับกลุ่มชาติพันธุ์จะใช้วิธีการเก็บข้อมูลแบบเครื่องบันทึกเสียง วิดีโอ หนังสือทำมือ สื่อต่าง ๆ ซึ่งยังไม่ได้อยู่ในรูปแบบดิจิทัล จนนานวันข้อมูลเหล่านี้เช่นสื่อ หรือหนังสืออาจจะเปื่อย เสื่อมสภาพ หรือเราเก็บข้อมูลในแผ่นดิสก์ เมื่อเทคโนโลยีพัฒนาไปก็ไม่สามารถอ่านข้อมูลจากแผ่นดิสก์ได้แล้ว เพราะฉะนั้นถ้าหากเราแปลงไฟล์เป็นดิจิทัล แล้วจัดเก็บเป็นระบบเราก็สามารถที่จะเผยแพร่ผ่านเว็บไซต์ได้ ก็จะสามารถรักษาข้อมูลได้นานกว่า ดังนั้นเราจึงได้ไปเรียนรู้วิธีการสร้างระบบคลังข้อมูลอย่างเป็นมาตรฐานที่ลอนดอน ประเทศอังกฤษ



ภาพที่ 10 -13 : กิจกรรมเรียนรู้กับเครือข่ายการทำงานจากต่างประเทศ

กิจกรรมที่สำคัญอีกกิจกรรมหนึ่งคือ การเก็บข้อมูลภาคสนาม อย่างเช่น กลุ่มชอุ้ง เป็นกลุ่มภาษาที่อยู่ในภาวะวิกฤตมาก แทบจะไม่มีผู้พูดได้แล้ว แต่มีนักวิชาการจากสถาบันวิจัยภาษา ฯ ได้ศึกษาไว้ เราจึงได้ไปค้นหาว่ามีผู้พูดได้เท่าไหร่ และได้พูดคุยว่าต้องการเก็บความรู้หรือไม่ ซึ่งก่อนการบันทึกจำเป็นต้องมีการเซ็นเอกสารอนุญาตยินยอมให้ข้อมูล และบันทึกข้อมูลด้วยเครื่องบันทึกเสียง วิดีโอ และนำไปจัดเก็บเป็นไฟล์ดิจิทัล ส่วนอีกกลุ่มหนึ่งคือ กลุ่มไทแสก ซึ่งเป็นภาษาที่อยู่ในภาวะวิกฤตเช่นกัน จัดอยู่ในตระกูลภาษาไต-กะได กลุ่มนี้ได้มีการทำวิจัยเพื่อท้องถิ่นมาก่อนและได้บันทึกภาษา เรื่องเล่า เนื้อเพลงด้วยมือ ทำเป็นหนังสือ แต่ยังไม่มีการบันทึกเสียงหรือบันทึกวิดีโอ ทีมวิจัยจึงเห็นโอกาสในการเติมเต็มข้อมูลส่วนนี้ และได้ติดต่อสอบถามความต้องการของชาวไทแสกว่าต้องการบันทึกวิดีโอเพิ่มเติมหรือไม่ ต้องการเผยแพร่ผ่านคลังข้อมูลออนไลน์หรือไม่ เมื่อทางชุมชนตอบตกลง ทีมวิจัยจึงได้ดำเนินการติดต่อไปยังชุมชนเพื่อบันทึกการแสดงทางวัฒนธรรมของชาวไทแสกดังกล่าว

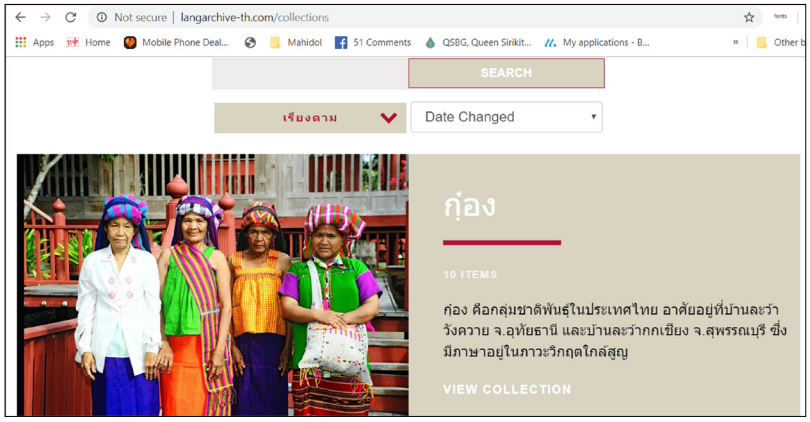
ภาพที่ 14 - 17 : กิจกรรม บันทึกข้อมูลภาคสนามกลุ่มชอุ้ง และไทแสก

นอกจากนี้ ยังมีการจัดอบรมเรื่องการบันทึกข้อมูลและทำคลังข้อมูล ซึ่งเป็นภารกิจหนึ่งของโครงการที่ต้องการเผยแพร่ความรู้ให้กับบุคลากรจากสถาบันวิชาการต่าง ๆ ให้มีความสามารถในการจัดเก็บคลังข้อมูลภาษาและวัฒนธรรมอย่างเป็นระบบ เช่น การจัดอบรมเชิงปฏิบัติการร่วมกับวิทยาลัยชุมชนตราด โดยมีการทดลองเก็บข้อมูลและนำข้อมูลเผยแพร่ผ่านเว็บไซต์ รวมทั้งการจัดอบรมเชิงปฏิบัติการร่วมกับ มหาวิทยาลัยราชภัฏยะลา มอ.ปัตตานี มหาวิทยาลัยหาดใหญ่ ในกลุ่ม

มลายูถิ่น ซึ่งเป็นเครือข่ายที่มีการทำงานร่วมกันมากว่า 10 ปี ได้ผลิตสื่อการเรียนการสอนมากมาย โดยเฉพาะสื่อทำมือที่มีการแปลงเป็นดิจิทัลแล้วบางส่วน

ภาพที่ 18 - 21 : กิจกรรมการจัดอบรมเรื่องการบันทึกข้อมูลและทำคลังข้อมูล

จากนั้นมีการนำเข้าคลังข้อมูลและเผยแพร่ผ่านเว็บไซต์ ซึ่งเรามีบุคลากรที่ทำหน้าที่ประจำ รวมทั้งการจัดกิจกรรมเสวนา “***การจัดทำและนำเสนอคลังข้อมูลดิจิทัลออนไลน์***” เพื่อการเผยแพร่ประชาสัมพันธ์ รวมทั้งแลกเปลี่ยนเรียนรู้ร่วมกันในด้านการจัดทำคลังข้อมูลด้านภาษาและวัฒนธรรมของกลุ่มชาติพันธุ์



ภาพที่ 22 - 23 : ตัวอย่างการนำเข้าคลังข้อมูลของกลุ่มฮากกา และกลุ่มก๋อง

หน้าเว็บไซต์คลังข้อมูลดิจิทัลกลุ่มชาติพันธุ์ในประเทศไทย ได้ใช้โดเมน http://langarchive-th.org ซึ่งมาจากคำว่า Language and Culture Archiving in Thailand ซึ่งการตั้งชื่อโดเมนจะต้องมีการตรวจสอบว่าซ้ำกับโดเมนอื่นหรือไม่ และควรตั้งชื่อให้ครอบคลุมกับสิ่งที่เราต้องการจะสื่อสาร ทั้งนี้เว็บไซต์ดังกล่าวยังสามารถใช้งานผ่านโทรศัพท์มือถือได้อีกด้วย

ภาพที่ 24 : หน้าเว็บคลังข้อมูล <https://langarchive-th.org/>

ลักษณะของเว็บไซต์คลังข้อมูล มีลักษณะเรียบง่าย เข้าถึงง่าย มีเมนูไม่มาก เน้นรูปภาพหน้าปกที่มีสีสัน เพื่อดึงดูดความสนใจ สามารถกำหนดผู้เข้าถึงข้อมูลได้ และมีการกำหนดข้อมูลชัดเจนว่าใครเป็นเจ้าของ ใครเป็นผู้จัดเก็บ

ประเภทข้อมูลที่จัดเก็บ มีทั้งรูปแบบไฟล์เสียง ไฟล์รูปภาพ ไฟล์วิดีโอ และไฟล์เอกสาร โดยใช้นามสกุล .MP3 .MP4 .JPG และ .PDF เพื่อจัดเก็บไฟล์ในขนาดที่ไม่ใหญ่เกินไป

ในเว็บไซต์มีเพียง 4 เมนูหลักเท่านั้น ได้แก่ ***เกี่ยวกับ*** แสดงรายละเอียดเกี่ยวกับเว็บไซต์ ***กลุ่มชาติพันธุ์*** แสดงคลังข้อมูลของกลุ่มชาติพันธุ์ที่มีการจัดเก็บแล้วมากกว่า 10 กลุ่มชาติพันธุ์ ***ภาษาและวัฒนธรรม*** แสดงข้อมูลด้านภาษาและวัฒนธรรม โดยไม่จำแนกกลุ่มชาติพันธุ์ และ ***ค้นหา*** แสดงการให้ค้นหาภาษาและวัฒนธรรมหรือกลุ่มชาติพันธุ์

ทั้งนี้จะมีระบบสมาชิก โดยผ่านการสมัครสมาชิก และผู้ดูแลคลังข้อมูลจะเป็นผู้ตรวจสอบ และ อนุญาตให้เข้าใช้งานในระบบสมาชิก ซึ่งไม่ใช่ใครก็ได้ที่จะเข้ามาปรับปรุงแก้ไข หรือทำอะไรก็ได้บนเว็บไซต์

การบันทึกข้อมูลทางภาษาและวัฒนธรรม ทางภาษาศาสตร์เรียกว่า Language Documentation คือการบันทึกข้อมูล ซึ่งเป็นศาสตร์ที่เริ่มมาประมาณ 20 กว่าปีมานี้เอง

การบันทึกข้อมูลทางภาษาและวัฒนธรรมมีวัตถุประสงค์สำคัญ คือการบันทึกข้อมูลทางด้านภาษาและวัฒนธรรมที่อยู่ในภาวะวิกฤต ซึ่งมีคุณค่า ความหลากหลาย แน่นอนว่าเราจำเป็นต้องรีบบันทึกก่อนที่ภาษานั้นจะสูญหายไป ข้อมูลที่ได้บันทึกเหล่านี้สามารถใช้เป็นสื่อในการเรียนรู้หรือถ่ายทอดไปยังลูกหลานเจ้าของวัฒนธรรมนั้น ๆ ถ้าสามารถเก็บรวบรวมข้อมูลด้านภาษาและวัฒนธรรมได้มาก โอกาสในการจะนำไปสู่การถ่ายทอดหรือใช้ประโยชน์ก็จะมากขึ้นตามไปด้วย

ก่อนที่จะบันทึกข้อมูล จะต้องมีการตกลงร่วมกัน โดยชี้แจงว่าข้อมูลที่จะบันทึกนี้จะบันทึกอะไร จะนำไปใช้อย่างไร ซึ่งหากเจ้าของภาษาอนุญาตหรือยินยอมผ่านการเซ็นใบยินยอม ก็สามารถบันทึกข้อมูลนั้นได้

การบันทึกภาษา ไม่ใช่เพียงแค่การบันทึกเสียงพูด แต่จำเป็นต้องบันทึกท่าทาง ลักษณะการพูดด้วย เพื่อจะทำให้ได้ข้อมูลที่มีความสมบูรณ์มากยิ่งขึ้น ดังนั้นจึงต้องใช้เครื่องบันทึกเสียงและกล้องวิดีโอร่วมด้วย

การบันทึกข้อมูลเป็นการทำงานร่วมกันระหว่างนักวิชาการหรือนักภาษาศาสตร์ ร่วมกับชุมชนกลุ่มชาติพันธุ์ รวมถึงนักเทคนิคที่มีความเชี่ยวชาญในด้านการใช้เครื่องมือ ก็จะทำให้ได้ข้อมูลที่มีความสมบูรณ์

การใช้ประโยชน์ของข้อมูลที่ได้บันทึกมานั้น สามารถนำไปใช้ได้หลายกลุ่ม เช่น ชุมชนเจ้าของภาษา กลุ่มนักวิชาการด้านอื่น ๆ นอกจากนี้ยังต้องคำนึงถึงชุมชนที่จะสามารถเข้าถึงข้อมูลของตนเองได้ง่าย

คำถามจากผู้เข้าร่วมเสวนา

จากมหาวิทยาลัยราชภัฏยะลา อยากให้เป็นข้อเสนอแนะหรือคำถาม เนื่องจากว่าในหนึ่งภาษามีหลากหลายพื้นที่ เช่น ภาษาเจ๊ะเห ในเขตพื้นที่นราธิวาส ถ้าหากปล่อยให้ชาวบ้านใส่ข้อมูลเข้าไปก็จะทำให้ควบคุมได้ค่อนข้างยาก แต่ขอเสนอว่าควรมีทีมงานของนักวิชาการในพื้นที่ต่าง ๆ เป็นผู้คัดกรองชั้นหนึ่ง หรือให้นักวิชาการในพื้นที่สมัครสมาชิกเพื่อ upload สิ่งเหล่านั้นมา เพื่อที่จะให้มีความรัดกุมมากยิ่งขึ้น นอกจากนี้ประเด็นเรื่องความเร็วของอินเตอร์เน็ตก็เป็นปัจจัยหนึ่งในการนำข้อมูลเข้าด้วย

ตอบโดย ผศ.ดร.ศิริเพ็ญ อึ้งสิทธิพูนพร

เรามีนักศึกษาปริญญาโทซึ่งเป็นสมาชิก กำลังเก็บรวบรวมข้อมูลกลุ่มเจ๊ะเห แล้วนำข้อมูลมาฝากกับเราส่วนหนึ่งแล้ว หากมหาวิทยาลัยราชภัฏยะลา ไปเก็บข้อมูลเพิ่มเติม สามารถที่จะทำข้อตกลงกับชุมชนว่าจะเป็นตัวแทนในการเผยแพร่ในรูปแบบคลังข้อมูลและจัดเก็บในเว็บไซต์ของเรา อาจจะทำในหน่วยงานของท่านได้ หรือจะนำมาฝากในเว็บไซต์ของเราก็สามารถที่จะพูดคุยได้ เพราะว่าตอนนี้เราก็เปิดรับ ขอบคุณค่ะ

คำถามจากผู้เข้าร่วมเสวนา

จากศูนย์มานุษยวิทยาสิรินธร เป็นประเด็นการสร้างระบบการจัดเก็บและจัดการ มี 2 ทางเลือก คือเราจะสร้างเอง หรือจะใช้ระบบสำเร็จที่เรียกว่า CMS ตรงนี้ศูนย์เราก็เคยผ่านระบบนี้มาเหมือนกัน ซึ่ง เราเคยใช้ Mukurtu มาแล้ว ซึ่งเลือกแบบไหนไม่ผิด แต่สิ่งที่จะต้องพิจารณาก่อนจะเลือกว่าจะสร้างเอง หรือใช้ระบบสำเร็จ มีอยู่ 3 ประเด็น

ประเด็นแรกคือเรื่องงบประมาณ การสร้างระบบใหม่จะต้องใช้งบประมาณในการจัดสร้างและ บำรุงรักษา ประการที่สอง คือ ถ้าใช้งานระยะยาว จะมีงบดูแลรักษาแค่ไหน ซึ่งบางกรณีที่จะต้องซื้อ account ที่ไม่ใช่ของฟรี ก็ต้องจ่ายเงินดูแลรักษาระยะยาวเช่นกัน ประการที่สามคือ เรื่องคน คนที่จะต้องดูแลระบบจะต้องเป็นคนที่มีความรู้ ซึ่งจริง ๆ จะต้องมีทั้ง 2 แบบ ไม่ใช่แค่ความรู้ทางด้าน IT อย่างเดียว แต่จะต้องมีความรู้ทางด้านเนื้อหาด้วย ขอให้ข้อมูลเบื้องต้นไว้ก่อนนะครับ

ตอบโดย ผศ.ดร.ศิริเพ็ญ อึ้งสิทธิพูนพร

ขอบคุณมากค่ะ ตอนแรกเราได้พูดคุยกับนักวิจัยฝั่งอังกฤษ แต่ในระยะแรกเราใช้ระบบนี้ไปก่อน และในอนาคตอาจจะต้องคุยกับมหาวิทยาลัยว่าจะสนับสนุนได้มากน้อย แค่ไหน อย่างไร ขอบคุณค่ะ

2. หอภาพยนตร์ (องค์การมหาชน)

โดยคุณ ชลิดา เอื้อบำรุงจิต ผู้อำนวยการหอภาพยนตร์ ซึ่งมีประสบการณ์การทำงานเกี่ยวกับด้านภาพยนตร์ ทั้งเป็นนักวิจารณ์ภาพยนตร์ อาจารย์พิเศษ รวมทั้งเป็นผู้คัดเลือกภาพยนตร์ของเทศกาลภาพยนตร์นานาชาติ และเป็นกรรมการตัดสินภาพยนตร์นานาชาติ ในงานเทศกาลภาพยนตร์จากหลายประเทศ ตลอดจนมีผลงานวิชาการซึ่งได้รับการตีพิมพ์ทั้งภาษาไทย และภาษาต่างประเทศ ได้กรุณามาแลกเปลี่ยนความรู้ของการทำคลังข้อมูลของหอภาพยนตร์

คุณชลิดา เอื้อบำรุงจิต ได้พูดถึงการจัดทำคลังข้อมูล และกระบวนการเปลี่ยนผ่านของหอภาพยนตร์

เมื่อก่อนหากพูดถึงหอภาพยนตร์ คนก็จะนึกถึงฟิล์มภาพยนตร์ ซึ่งก็เป็นจริงตามนั้น เพราะ collection ชุดแรก เป็น collection ที่ใหญ่มาก ๆ ของหอภาพยนตร์นั่นก็คือ ฟิล์มภาพยนตร์

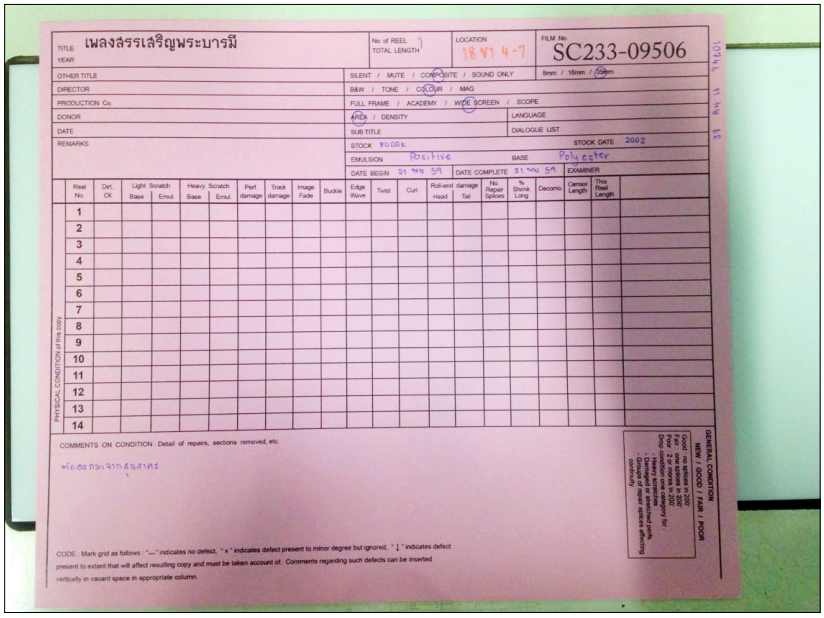
ตอนนี้ถ้าใครเจอฟิล์มภาพยนตร์จากที่อื่น แล้วไม่รู้จะจัดการหรือทำอย่างไร ก็สามารถส่งมาที่หอภาพยนตร์ได้ แต่ว่าคนที่ไม่เคยเจอฟิล์มภาพยนตร์อาจจะนึกไม่ออกว่า เวลาที่เราจะต้องจัดการกับฟิล์ม จะต้องใช้เวลามากเพียงใด

หนังเรื่องยาวเรื่องหนึ่งในสมัยก่อนจะมีปริมาณประมาณ 6 ม้วน เพราะฉะนั้นการจัดการพื้นที่ รวมทั้งแรงงาน ทุกอย่าง หลายคนอาจนึกไม่ออกว่าเป็นอย่างไร ตอนนี้เรามีห้องเก็บฟิล์มแล้วประมาณ 30,000 กว่าม้วน แต่นั่นไม่ใช่ทั้งหมด เรายังมีฟิล์มที่จะต้องจัดการและต้องเก็บเพิ่มอีกพอสมควร เพราะฉะนั้นคิดว่าเราคงมีไม่ต่ำกว่าแสนม้วน

นอกจากฟิล์มภาพยนตร์แล้ว เรายังมีสื่ออื่น ๆ อีกที่เรียกว่า สิ่งเกี่ยวเนื่องกับภาพยนตร์ อย่างเช่น โปสเตอร์หนัง 4,000 - 5,000 รายการ แต่มีจำนวนแผ่นรวมกันเป็นแสน อีกส่วนหนึ่งคือ เทป เกินแสนม้วนแน่นอน ปัจจุบันได้รับ Hard disk เป็นดิจิตอลไฟล์อีกหลายชิ้น เหล่านี้เป็นสิ่งที่เราจะต้องจัดการ เพราะฉะนั้นจึงต้องมีกระบวนการหรือหลักการในการบริหารจัดการ

ย้อนกลับไปในยุคแรก หอภาพยนตร์เป็นสมาชิกของ FIAF หรือสหพันธ์ภาพยนตร์นานาชาติ จะมีเรื่อง guide line หรือ กฎเกณฑ์ในการจัดทำ catalogue (The FIAF Cataloguing Rules For Film Archives) ในยุคแรกยังไม่มีดิจิทัล ก็จะมีคู่มือออกมาหรือเรียกว่าเป็น metadata ในปัจจุบัน

เมื่อเราได้ฟิล์มมาเราต้องลงทะเบียนฟิล์ม ซึ่งส่วนบนจะเหมือน metadata เลย เนื่องจากจะมีข้อมูลเกี่ยวกับหนัง ได้มาจากใคร ลักษณะกายภาพเป็นอย่างไร สมมติว่าฟิล์มแย่มากเลย เราจะไม่มีทางรู้เลยว่ามันคืออะไรนอกจากเราจะเปิดดู เพราะฉะนั้นสิ่งแรกเราต้องจัดการทางกายภาพก่อนก่อนที่เราจะดูได้ เมื่อก่อน preservation มาก่อน access แต่คนข้างนอกจะมองเห็นว่าเราทำอะไรก็ต่อเมื่อเขาได้เห็นสิ่งนั้น เพราะฉะนั้นบัตรทะเบียนฟิล์ม ก็จะระบุว่า มีลักษณะทางกายภาพเป็นอย่างไรบ้าง มีรอยต่อกี่รอย ตรงไหนบ้าง เพื่อที่เจ้าหน้าที่ที่จะต้องจัดการในลำดับต่อไปจะได้จัดการถูก เนื่องจากจะต้องซ่อม หรือต้องมี treatment ทางเคมี หรือเข้าห้องแล็บอย่างไร เบื้องต้นเราจะต้องบันทึกลงไปในบัตรทะเบียนฟิล์ม เพราะฉะนั้นเบื้องต้นในลักษณะ descriptive จึงยังมีไม่มาก ซึ่งเราอ้างอิงมาจากแคตาล็อกของ FIAF



ภาพที่ 25 : ตัวอย่างบัตรทะเบียนฟิล์ม[3]

ในขั้นต่อมาที่เริ่มมีการใช้คอมพิวเตอร์ เราได้ออกแบบ database ซึ่งเป็นการออบแบบตามบัตรทะเบียนฟิล์ม มี 3 สีด้วยกันคือ บัตรสีชมพู เป็นหนังสารคดี บัตรสีฟ้าเป็นหนังข่าว บัตรสีขาวเป็นหนัง เรื่อง field ต่าง ๆ ก็อ้างอิงมาจากบัตร ซึ่งเมื่อก่อนเรายังไม่มีความรู้ในเรื่องดิจิทัล หรือคอมพิวเตอร์มากนัก เราไม่เข้าใจเรื่องการออกแบบ เมื่อนำมาใช้จริง ๆ ก็พบว่ามีข้อจำกัดค่อนข้างมาก เพราะว่าโลกของ analogue และ digital มันคนละอย่างกัน แต่อย่างไรก็ตามสิ่งที่เราทำก็ไม่ได้เปล่าประโยชน์

ในหน้าสืบค้นเราก็นำคำค้น เหมือนสืบค้นสมัยนี้ได้ ตอนนั้นเรายังไม่ได้แปลงฟิล์มเป็นดิจิทัล แต่เราให้บริการดูเป็นวิดีโอ จะมีเลขกำกับว่าเลขอะไร แล้วเจ้าหน้าที่ก็จะไปหยิบตามเลข แล้วมาดูเหมือนในห้องสมุด ถือเป็นระยะแรกที่เราชำระข้อมูลจากกระดาษมาลงคอมพิวเตอร์ เมื่อใช้ไปสักระยะหนึ่งก็เริ่มรวน เพราะไม่ได้รับการพัฒนาต่อ เราพยายาม export ข้อมูลทั้งหลายที่เราใส่เข้าไปออกมาเป็นตาราง

เมื่อก่อนเป็นระบบทะเบียนซีดีและดีวีดี มีรุ่นสมุด ก็ให้เห็นว่ามันมาจากเราทำรายการขึ้นมาก่อน จากนั้นมาเราเริ่มมีระบบสืบค้นในแบบห้องสมุด และสืบค้นผ่านเว็บไซต์ออนไลน์ ซึ่งใช้ open vimeo รวมถึงข้อมูลที่เป็นสื่อโสตทัศน์ ที่สามารถเข้าไปสืบค้นข้อมูลได้ จะเห็นว่าในระบบแรกเราต้องบอกเจ้าหน้าที่เพื่อให้เจ้าหน้าที่สืบค้นให้ ต่อมาผู้ใช้บริการสามารถสืบค้นได้ด้วยตนเอง และนำไปบอกเจ้าหน้าที่

---

[3] ภาพสไลด์ของคุณชลิดา เอื้อบำรุงจิต จากการการประชุมเสวนาแลกเปลี่ยนความรู้ “การจัดทำและนำเสนอคลังข้อมูลดิจิทัลออนไลน์” วันที่ 27 พฤศจิกายน 2562 หน้าที่ 6

ตอนนี้มีข้อมูลทั้งหมด 23,364 รายการ และเป็นสื่อโสตทัศน์อีก 15,456 รายการ ซึ่งพร้อมให้บริการ เมื่อเราเข้าไปสืบค้นก็จะขึ้นหมายเลข แล้วไปบอกกับเจ้าหน้าที่ได้ บางรายการจะมี descriptive ที่ทำให้สามารถค้นข้อมูลได้ลึกมากขึ้น แต่บางรายการก็ยังไม่มี descriptive

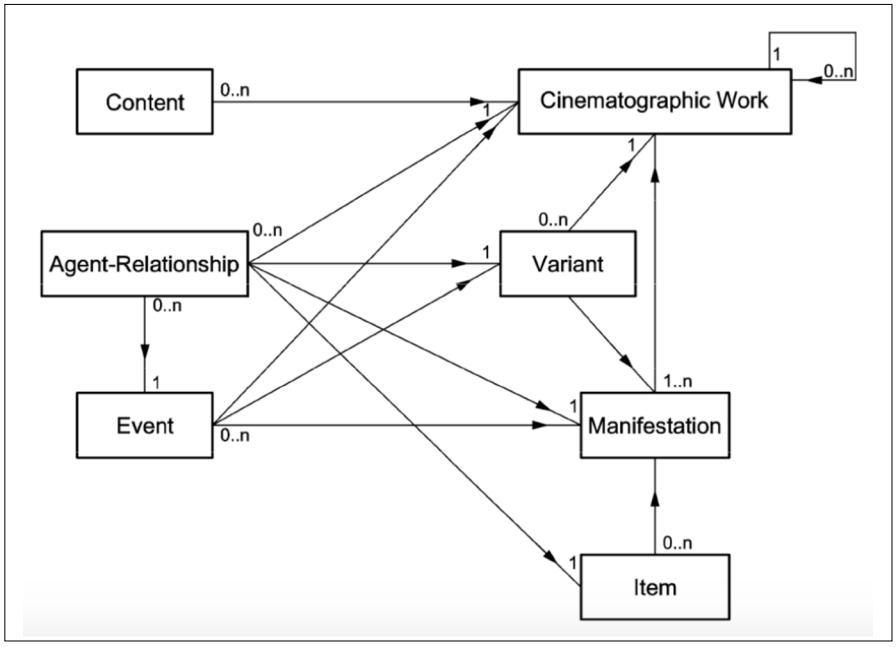
เนื่องจากการที่จะทำ descriptive ให้ทันมันยากมาก ซึ่งเราพยายามจ้างผู้เชี่ยวชาญก็ทำได้ไม่เกิน 1,000 รายการ เนื่องจากจะต้องมีความละเอียดและต้องค้นคว้าเพิ่มเติม

ต่อมาเข้าสู่ยุคที่เรียกว่าโลกเราเปลี่ยนเป็น youtube การให้บริการที่เป็นแผ่น เครื่องเล่นก็เริ่มเสียและหายากขึ้น จึงต้องเคลื่อนไปสู่ระยะต่อไป

หอภาพยนตร์เริ่มทำ digitized เรียกได้ว่าเป็น full option เมื่อประมาณ 3 ปีที่แล้ว และสิ่งที่เราเรียนรู้ เราค้นพบว่า กว่าจะแปลงสัญญาณฟิล์มได้ต้องผ่านการตรวจฟิล์ม ทำให้เป็นการทำงานที่ช้า คิดว่าการเตรียมฟิล์มน่าจะใช้เวลาประมาณ 70 - 80 % ในขณะที่การ digitized น่าจะประมาณ 20 % เพราะฉะนั้นใน 1 ปี ถ้าความยาวสั้น ๆ สามารถทำได้หลักพันเรื่อง แต่ถ้ามีความยาวมาก ก็ทำได้ไม่กี่ร้อยเรื่อง เพราะฉะนั้นจำนวนของการเพิ่มขึ้น กว่าจะได้เท่ากับ 15,000 เรื่อง น่าจะใช้เวลาเยอะมาก จึงได้ปรับมาทดลองใช้ระบบ VOD ซึ่งสามารถดูได้เฉพาะที่หอภาพยนตร์

เราเริ่มทำ data cleaning มาเช็คดูว่าอันไหนมันแย่มาก ๆ หรือซ้ำ ก็เอาออก นอกจากนี้สิ่งใหม่ที่เราพบคือ การที่เราจะนำเข้าระบบ ทั้ง media และ metadata จะต้องมาพร้อมกัน ทำให้เกิดแรงผลักดันใหม่ เราทำงานแบบเดิมไม่ได้แล้ว เพราะฉะนั้นเราต้องเปลี่ยนวิธีมอง โดยมองว่าอะไรที่มีความเป็นไปได้ที่จะทำให้ metadata เราพร้อม เราเริ่มจัดการกับหนังเรื่องก่อน ก็คือหนังที่เข้าโรง ซึ่งจะมีข้อมูล metadata จำนวนหนึ่ง ตอนนี้มีอยู่ประมาณเกือบ 2,200 เรื่อง ที่สามารถสืบค้นได้ในหอสมุด ซึ่งอาจจะใช้สำหรับดูในจอได้ แต่ไม่สามารถไปฉายหรือทำอย่างอื่นได้ เราจึงทำสิ่งที่ให้บริการประชาชนได้ก่อน และรวมทั้งหนังที่คนมาค้นบ่อย ๆ หนังข่าวมีประมาณ 5,000 รายการ ในขณะที่ด้านหลังก็จะทำการอนุรักษ์อย่างเต็มที่มันต้องใช้เวลามาก ซึ่งถือเป็นวิธีการแก้ปัญหาที่จะทำอย่างไรเราจึงสามารถทำให้ข้อมูลไล่ตามทันกับการให้บริการประชาชน แม้ว่าจะไม่สมบูรณ์แบบก็ตาม

ทีนี้กลับไปที่ FIAF เมื่อ Catalogue Rule จัดทำขึ้นสำหรับภาพยนตร์อย่างเดียว ไม่รวมสื่อใด ๆ แต่เมื่อมี digital เข้ามา หรือ metadata เข้ามา ก็มีการถกเถียงกันว่า data model แบบเดิมมันใช้ไม่ได้แล้ว เขาใช้เวลาเกือบ 10 ปี จึงออกคู่มือ The FIAF Moving Image Cataloguing Manual คือจะบอกอย่างละเอียด อย่างเช่น หนังเรื่องเดียวกันแต่มีหลายเวอร์ชั่น เช่น ฉายโรง 120 นาที ฉายผ่านทีวี 180 นาที ก็จะต้องมีข้อมูลเพื่อที่จะควบคุมทุกอย่างได้ ซึ่งเราก็ฝันอยากจะทำให้ได้ระบบแบบนี้



ภาพที่ 26 : ระบบ data model ของการจัดทำคลังด้านภาพยนตร์[4]

ในระหว่างกลางทางที่เราทำงาน เราก็ใช้ google drive อย่างที่บอกคือในตอนแรก เรามีเทป เรามีสิ่งเกี่ยวเนื่อง เครื่องแต่งกาย โปสเตอร์ การ์ด รูปถ่าย ในระหว่างที่เรารอระบบ เราก็ทำงานบน sheet ตอนนี้เราก็มีตาราง 20 - 30 ตาราง ซึ่งเมื่อมีระบบที่ดีแล้วก็สามารถนำข้อมูลมาใส่ได้เลย

นอกจากนี้มีข้อคิดอีกอย่างหนึ่ง คือ ผู้ที่จะสามารถออกแบบโปรแกรมจำเป็นจะต้องมีความเข้าใจโครงการในเรื่องของภาพยนตร์ ประจวบเหมาะกับปัจจุบัน BFI (The British Film Institute : สถาบันภาพยนตร์อังกฤษ) ได้นำโมเดลไปพัฒนาทำให้ได้โปรแกรมการจัดการข้อมูล ซึ่งปัจจุบันทางหอภาพยนตร์กำลังอยู่ในระหว่างทดสอบการใช้โปรแกรมนี้ ทำให้ลดระยะเวลาและต้นทุนในการทำงาน

ดังนั้นการทำ digital transformation มีอยู่ 2 ส่วน ส่วนที่หนึ่งคือ metadata ส่วนที่สองก็คือ media ซึ่งมี file format หลายรูปแบบ เช่น DCDM DPX TIFF JPEG2000 ซึ่งไฟล์แต่ละประเภทก็มีการใช้แตกต่างกัน เราต้องมองด้วยว่าไฟล์ที่เราจะจัดเก็บเป็นไฟล์ขนาดไหน ไฟล์อะไร และเราจะเก็บเอาไว้ที่ไหน อย่างน้อยควรเก็บ backup ตามกฎ 3 - 2 - 1 คือ 3 copies 2 formats และ 1 offsite อย่างน้อยก็ปลอดภัย

ในอนาคตถ้าเป็นไปได้หอภาพยนตร์ก็อยากจะเก็บภาพเคลื่อนไหว ตอนนี้เรากำลังจะขยาย tape library ให้ใหญ่ขึ้น เพื่อให้สามารถรองรับอะไรได้มากขึ้น

---

[4] ภาพสไลด์ของคุณชลิดา เอื้อบำรุงจิต จากการการประชุมเสวนาแลกเปลี่ยนความรู้ “การจัดทำและนำเสนอคลังข้อมูลดิจิทัลออนไลน์” วันที่ 27 พฤศจิกายน 2562 หน้าที่ 18

สิ่งที่อยากจะฝากไว้คือ digitization ไม่เท่ากับ preservation หลาย ๆ คนคิดว่า เมื่อเรา digitized อะไรไปแล้ว เราก็ทิ้ง master ไปก็ได้ ฟิล์มทำไมต้องเก็บ ห้องเก็บฟิล์มจะต้องเป็นห้องเย็น 4 องศา ความชื้นสัมพัทธ์ 40 เปิดแอร์ตลอด 24 ชม. บางทีก็จะรู้สึกว่ามันเปลืองไฟ แต่ว่าหากไฟล์ digitized เป็นอะไรขึ้นมา เราสามารถที่จะเอาฟิล์มม้วนเดิมกลับมาทำใหม่ได้ และตอนนี้ที่เป็นวิกฤตมาก คือเทป เนื่องจากต้องเล่นด้วยเครื่องเล่นเทป ซึ่งเครื่องเล่นเทป และ เครื่อง play back ก็กำลังจะหมดไป หน่วยงานที่มีเทปอยู่ ก็อยากให้ลองเช็คดูว่าอะไรสำคัญ ก็ลองแปลงออกมาเป็นไฟล์ดิจิทัล เพราะว่าในไม่ช้าอาจจะทำไม่ได้แล้ว

สิ่งสุดท้ายคือ มีช่อง youtube ของหอภาพยนตร์ เนื่องจากว่ามีเรื่องลิขสิทธิ์ เราจึงเผยแพร่ได้เพียงบางส่วน เราขึ้น youtube ยังไม่ถึง 1,000 รายการ เนื่องจากเราขึ้นเฉพาะที่ไม่เป็นปัญหาเรื่องลิขสิทธิ์

คำถามจากผู้ร่วมเสวนา

ไฟล์ที่จัดเก็บว่ามีการประเมินคุณค่าหรือการแบ่งประเภทอย่างไร รับมอบจากที่อื่นมาหรือไม่ หรือว่าเราผลิตขึ้นเอง

ตอบโดย คุณชลิดา เอื้อบำรุงจิต

คิดว่าเกือบ 100 % ที่เราอนุรักษ์คือเรารับมาจากที่อื่น เราผลิตเองน้อยมาก ส่วนการประเมินเรามักจะประเมินตอนรับ เรามีแผนกจัดหา หรือรับของ เรามี collection policy ว่าเราจะรับอะไรบ้าง ซึ่งก็คือภาพยนตร์หรือภาพเคลื่อนไหวที่ผลิตโดยคนไทยหรือเกี่ยวข้องกับคนไทย

เมื่อก่อนจะเป็นฟิล์มภาพยนตร์อย่างเดียว แต่สื่ออื่น ๆ ก็เป็นเรื่องที่เราคุยกัน ประเด็นหนึ่งของการอนุรักษ์ก็คือเราพยายามที่จะไม่ duplicate กับคนอื่น เมื่อ 20 ปีก่อนเราจะไม่เก็บโทรทัศน์ เพราะสถานีโทรทัศน์จะมีเก็บอยู่แล้ว แต่พอหลัง ๆ มองเห็นว่า สถานการณ์การเก็บโทรทัศน์มันไม่ได้ไปไกลมาก เขาเลือกเก็บ และมีอะไรเยอะแยะที่เขาไม่เก็บ และมีเทปที่เดินทางมาถึงเราเยอะเหมือนกัน หลาย ๆ ครั้งก็ตัดใจที่จะทิ้งไปยาก เพราะเรารู้ว่าจะไม่มีใครเก็บแล้ว คือเรามี collection policy ที่เราจะต้องทบทวนกันทุก 5 ปี และจัดลำดับความสำคัญ

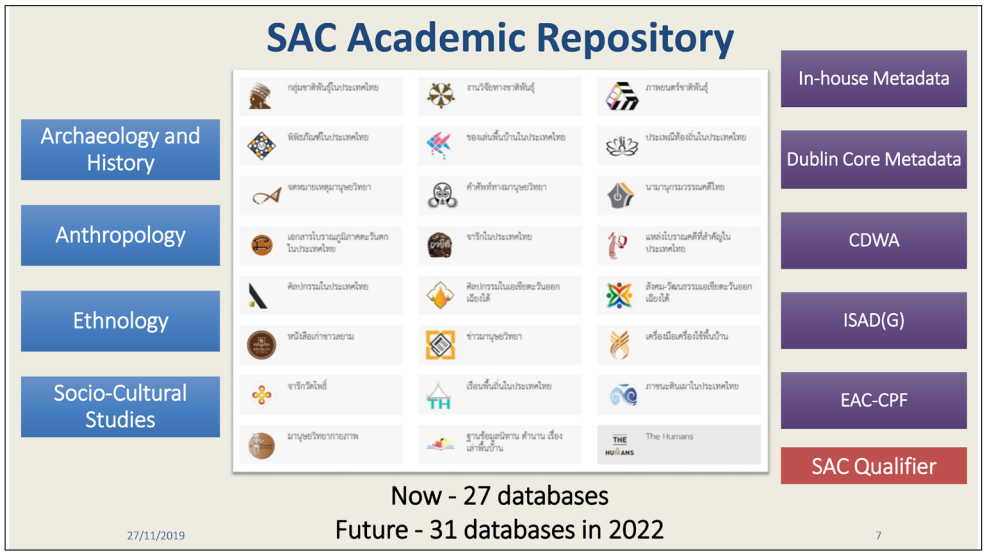
3. ศูนย์มานุษยวิทยาสิรินธร (องค์การมหาชน)

โดยคุณสิทธิศักดิ์ รุ่งเจริญสุขศรี นักวิชาการกลุ่มงานคลังข้อมูล สำนักวิชาการและสารสนเทศ จากศูนย์มานุษยวิทยาสิรินธร มีความเชี่ยวชาญด้านการวิเคราะห์และออกแบบโครงสร้างฐานข้อมูลวิชาการต่าง ๆ ขององค์กร และได้ดำเนินโครงการคลังข้อมูลชุมชนและเครือข่ายข้อมูล โครงการศึกษาและจัดทำระบบจัดเก็บรักษาเว็บถาวร และโครงการจัดทำแนวปฏิบัติการจัดการข้อมูลมานุษยวิทยา

คุณสิทธิศักดิ์ รุ่งเจริญสุขศรี ได้อธิบายถึง บทบาทและความท้าทายของศูนย์มานุษยวิทยากับการจัดการและดูแลสารสนเทศมรดกทางวัฒนธรรม อะไรคือความท้าทายที่เราเจอ และอะไรคือสิ่งที่เรียกว่าสารสนเทศทางวัฒนธรรม

ศูนย์มานุษยวิทยาสิรินธร (องค์การมหาชน) มีวิสัยทัศน์คือ เป็นองค์กรหลักในการจัดการข้อมูลความรู้ด้านมานุษยวิทยาเพื่อสร้างความเข้าใจ ความหลากหลายทางวัฒนธรรมของชาติ โดยมีภารกิจหลัก 3 ด้าน ก็คือ วิจัยและพัฒนา การทำคลังข้อมูลเพื่อให้บริการทางวิชาการ และนำข้อมูลความรู้ที่ได้มาสื่อสารสู่สาธารณะ

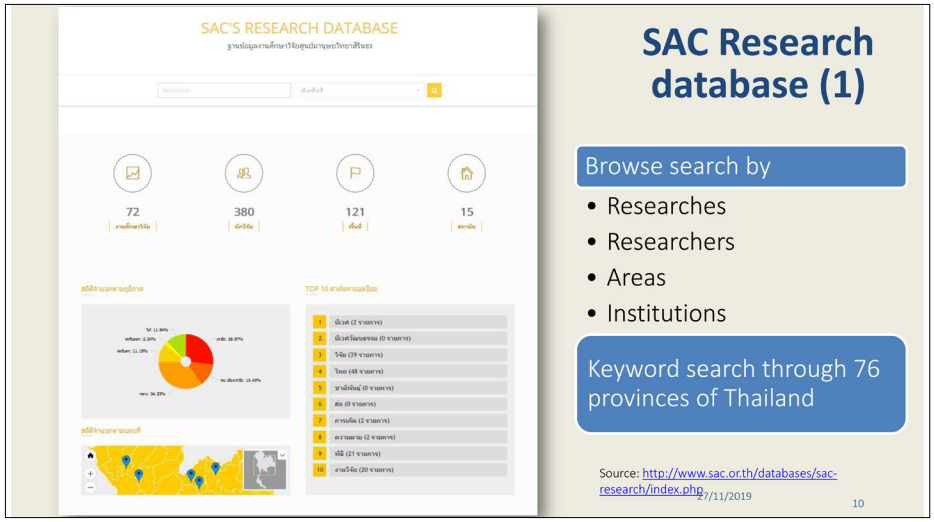
คลังข้อมูลของศูนย์ ฯ เมื่อปี 1999 มีแค่ 2 ฐานข้อมูล คือ ฐานข้อมูลพิพิธภัณฑ์และฐานข้อมูลกลุ่มชาติพันธุ์ ต่อมาหน้าตาปัจจุบันมีฐานข้อมูล 27 ฐาน และภายในปี 2022 จะมีเพิ่มเป็น 31 ฐาน



ภาพที่ 27 : ฐานคลังข้อมูลศูนย์มานุษยวิทยาสิรินธร [5]

หากแบ่งตามหมวดของเนื้อหา จะประกอบด้วยด้านประวัติศาสตร์และโบราณคดี มนุษยวิทยาและกลุ่มชาติพันธุ์ สังคมวัฒนธรรมร่วมสมัย หากแบ่งตามโครงสร้างในการทำฐานข้อมูลที่เรียกว่า Metadata Structure จะประกอบด้วย 5 หมวดหลัก ๆ คือ 1) Metadata แบบเดิมที่ได้ออกแบบขึ้นเอง ซึ่งไม่ได้อิงกับมาตรฐานอะไร 2) Dublin Core Metadata 3) CDWA ซึ่งจะใช้กับประเภทวัตถุในพิพิธภัณฑ์ 4) ISAD (G) จะใช้กับงานจดหมายเหตุ และ 5) EAC-CPF จะคล้ายกับ ISAD (G) แต่จะเน้นที่การอธิบายรายละเอียดข้อมูลส่วนบุคคลมากขึ้น เหล่านี้คือตัว Standard Metadata ที่ใช้อธิบายโครงสร้างคลังข้อมูลที่อยู่ในระบบคลังข้อมูลของศูนย์ฯ และมีสิ่งที่เพิ่มขึ้นที่พิเศษ เรียกว่า SAC qualifier ก็คือส่วนขยายสำหรับการอธิบายข้อมูลของศูนย์มานุษยวิทยาสิรินธรโดยเฉพาะ

[5] ภาพสไลด์ของคุณสิทธิศักดิ์ รุ่งเจริญสุขศรี จากการการประชุมเสวนาแลกเปลี่ยนความรู้ “การจัดทำและนำเสนอคลังข้อมูลดิจิทัลออนไลน์” วันที่ 27 พฤศจิกายน 2562 หน้าที่ 7



ภาพที่ 28 : แสดงฐานข้อมูลของศูนย์มานุษยวิทยาสิรินธร (องค์การมหาชน)[6]

ตัวอย่างฐานข้อมูล คือ ประเพณี พิธีกรรม ความเชื่อท้องถิ่น เมื่อเก็บข้อมูลนี้มาแล้วก็อธิบายภายใต้โครงการของ Dublin Core หรือเครื่องมือเครื่องใช้ในชีวิตประจำวัน ก็อธิบายภายใต้โครงสร้าง Dublin Core เช่นกัน

ภาพที่ 29 : ตัวอย่างฐานข้อมูลประเพณี และพิธีกรรม[7]

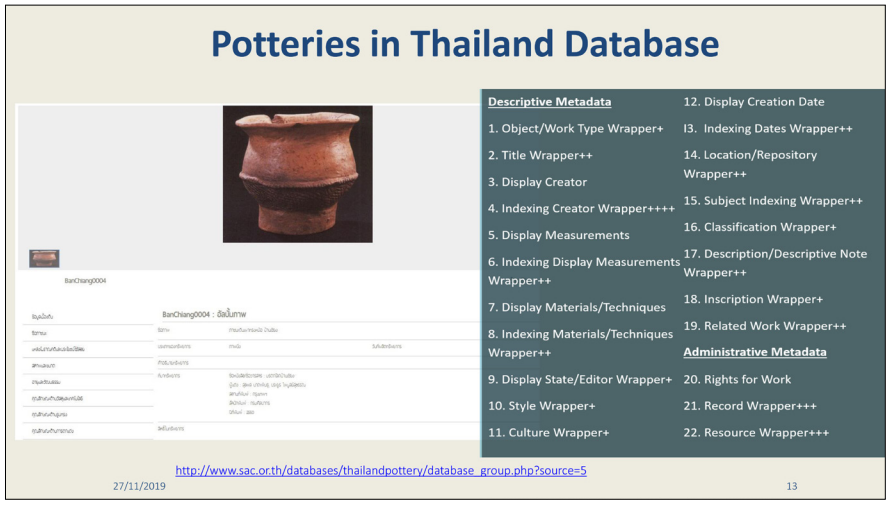
[6] ภาพสไลด์ของคุณสิทธิศักดิ์ รุ่งเจริญสุขศรี จากการการประชุมเสวนาแลกเปลี่ยนความรู้ “การจัดทำและนำเสนอคลังข้อมูลดิจิทัลออนไลน์” วันที่ 27 พฤศจิกายน 2562 หน้าที่ 10

[7] ภาพสไลด์ของคุณสิทธิศักดิ์ รุ่งเจริญสุขศรี จากการการประชุมเสวนาแลกเปลี่ยนความรู้ “การจัดทำและนำเสนอคลังข้อมูลดิจิทัลออนไลน์” วันที่ 27 พฤศจิกายน 2562 หน้าที่ 11

เหล่านี้เป็นฐานข้อมูลงานวิจัยของศูนย์มานุษยวิทยาสิรินธร ซึ่งก่อนหน้านี้ทางศูนย์ฯ ไม่ค่อยมีวิธีการจัดการและเผยแพร่ที่ดีเท่าไหร่ คือทำเป็นรายงานเป็นเล่ม แล้วเก็บไว้ใช้เองในองค์กร แต่เมื่อ 2 - 3 ปีก่อนมีแผนว่างานวิจัยนี้ศูนย์ฯ ควรจะตีพิมพ์เผยแพร่ แล้วให้คนนำไปใช้ในรูปแบบ full text ได้ เลยสร้างฐานข้อมูลนี้ขึ้นมา ซึ่งก็ใช้ตัว Dublin Core Metadata เป็นฐานข้อมูลในการอธิบาย

ศูนย์ฯ ได้เพิ่มส่วนขยายของศูนย์มานุษยวิทยาสิรินธร ที่เรียกว่า SAC qualifier พร้อมด้วยเปิดให้บริการ full text download และมีแยกเป็นบทได้ สิ่งพิเศษที่เราเพิ่มขึ้นมาเรียกว่า การสร้างระบบอ้างอิงอัตโนมัติ คือ ผู้ใช้บริการไม่จำเป็นต้องไปเขียนเอง สามารถคัดลอกนำไปใช้ได้เลย ซึ่งการที่จะทำตัว auto citation ได้ก็คือเกิดขึ้นจากการที่เราทำโครงสร้างหลังบ้านไว้ได้ดีในระดับหนึ่งแล้ว แล้วไปสื่อสารกับโปรแกรมเมอร์ให้เข้าใจว่า ภายใต้โครงสร้างแบบนี้เราต้องการให้ดึงข้อมูลนี้ออกมา เพื่อ generate ใหม่เป็นตัว bibliography ในรูปแบบ APA MLA หรือ HARVARD เพื่อให้ผู้ใช้สามารถนำไปใช้ได้เลย ซึ่งจะช่วยส่งเสริมการอ้างอิงทางวิชาการได้ด้วย เพราะปัญหาที่เราเจอก็คือ ถ้าเป็นตัวเว็บไซต์ ผู้ใช้ส่วนใหญ่จะละเลยการอ้างอิง จะเน้นไปที่หนังสือมากกว่า auto citation

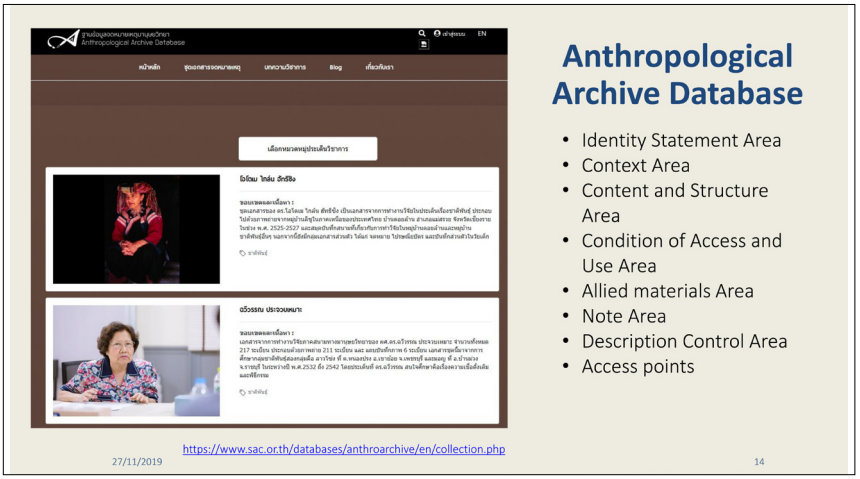
ตัวอย่างฐานข้อมูลภาชนะดินเผาในประเทศไทย ใช้ CDWA ในการอธิบาย ก็จะได้ 2 แกนหลักในการอธิบาย คือตัวที่เรียกว่า Descriptive Metadata ก็คือให้รายละเอียดของวัตถุ และ Admimistrative Metadata ก็คือให้รายละเอียดของการบริหารจัดการวัตถุทางพิพิธภัณฑ์นี้ว่าใครเป็นคนทำทะเบียน ทำเมื่อไหร่ รับมอบมาเมื่อไหร่ หากสนใจรายละเอียด Metadata เหล่านี้สามารถหาโหลดได้



ภาพที่ 30 : แสดงตัวอย่างฐานข้อมูลภาชนะดินเผา[8]

[8] ภาพสไลด์ของคุณสิทธิศักดิ์ รุ่งเจริญสุขศรี จากการการประชุมเสวนาแลกเปลี่ยนความรู้ “การจัดทำและนำเสนอคลังข้อมูลดิจิทัลออนไลน์” วันที่ 27 พฤศจิกายน 2562 หน้าที่ 13

ถัดมาเป็นฐานข้อมูลจดหมายเหตุมานุษยวิทยาในประเทศไทย ซึ่งศูนย์ฯ ใช้ ISAD (G) (General International Standard Archival Description) ถ้าใครอยู่ในจดหมายเหตุก็จะรู้ว่าอันนี้เป็นตัว Metadata ที่ใช้อธิบายจดหมายเหตุ ซึ่งมีความแตกต่างจากการอธิบายวัตถุพิพิธภัณฑ์ซึ่งต้องใช้ Standard ที่แตกต่างออกไป เช่น บริบท เนื้อหา เงื่อนไขในการเข้าถึง ซึ่งศูนย์เราก็ประยุกต์ ISAD (G) มาใช้ในการให้รายละเอียด collection งานวิจัยภาคสนามของนักวิจัยที่เราได้รับมอบหมายมา ถ้าสนใจอยากดูคู่มือภาษาไทยก็ได้เขียนไว้แล้วภายใต้งานจดหมายเหตุทางมานุษยวิทยา



ภาพที่ 31 : แสดงตัวอย่างฐานข้อมูลจดหมายเหตุมานุษยวิทยาในประเทศไทย [9]

ตัวอย่างข้อมูลจากฐานจดหมายเหตุ ภาพถ่ายจากกลุ่มชาติพันธุ์ เช่น ภาพไทดำ ซึ่งมาจากงานวิจัยของอาจารย์ฉวีวรรณ ประจวบเหมาะ ส่วนไทลื้อ มาจากงานวิจัยของ ไมเคิล มอร์แมน ซึ่งได้เข้ามาทำวิจัยในช่วง 1970 - 1980

ภาพที่ 32 : แสดงตัวอย่างภาพถ่ายกลุ่มชาติพันธุ์ไทดำ[10]

[9] ภาพสไลด์ของคุณสิทธิศักดิ์ รุ่งเจริญสุขศรี จากการการประชุมเสวนาแลกเปลี่ยนความรู้ “การจัดทำและนำเสนอคลังข้อมูลดิจิทัลออนไลน์” วันที่ 27 พฤศจิกายน 2562 หน้าที่ 14

[10] ภาพสไลด์ของคุณสิทธิศักดิ์ รุ่งเจริญสุขศรี จากการการประชุมเสวนาแลกเปลี่ยนความรู้ “การจัดทำและนำเสนอคลังข้อมูลดิจิทัลออนไลน์” วันที่ 27 พฤศจิกายน 2562 หน้าที่ 15

โดยที่ข้อมูลทั้งหมดที่ศูนย์ฯ มีอยู่ประมาณ 27 ฐาน ซึ่งถูกผูกเข้าด้วยกัน เพื่อการค้นหาหรือสืบค้น ผู้ใช้บริการสามารถเริ่มจากประเภท เนื้อหา และฐานข้อมูล อันนี้เราสร้างขึ้นมาเพื่อแก้ปัญหา ซึ่งก่อนหน้านั้นอย่างที่ทราบว่าฐานข้อมูลศูนย์เป็นคลังใหญ่ ๆ 1 คลัง แต่มีตัว database เยอะมาก และด้วยนโยบายของศูนย์ฯ ที่จะต้องผลิตข้อมูลทางวิชาการเพิ่มขึ้น ทำให้ประสบปัญหาว่ามีฐานข้อมูลเพิ่มขึ้น แล้วเมื่อผู้ใช้เข้ามาค้นข้อมูล จะค้นได้อย่างไร ที่ผ่านมาต้องใช้วิธีเข้าฐานหนึ่งออกฐานหนึ่ง ซึ่งมีความยุ่งยากมาก จึงทำให้ใช้ federated search มา แต่ ณ ตอนนี้ federated search ก็ประสบปัญหาในการใช้งานเหมือนกัน เนื่องจากถูก google search แย่งหน้าที่ในการใช้งาน มันไม่ตอบโจทย์พฤติกรรมการใช้งานของผู้ใช้ ณ ปัจจุบันแล้ว เนื่องจากหาได้ยากมากที่ผู้ใช้จะเข้ามาในเว็บไซต์ศูนย์ แล้วไปค้นข้อมูล กลายเป็นว่า ผู้ใช้เข้าจาก google สิ่งนี้ก็เป็นเรื่องที่ศูนย์ฯ พยายามแก้ไข

ความท้าทายในการทำงาน อย่างที่บอกว่าโดยพื้นฐานของศูนย์ฯ เป็นสถาบันทางวิชาการรับนักวิจัยทางด้านสังคมศาสตร์ มนุษยศาสตร์เป็นหลักเพื่อมาทำงานหาข้อมูล แต่สิ่งที่ต้องเผชิญจากการทำงาน กลายเป็นว่านักวิชาการที่จบทางด้านมานุษยวิทยาชาติพันธุ์ โบราณคดี ประวัติศาสตร์ ภาษาศาสตร์ หรือพิพิธภัณฑ์ คนพวกนี้จะต้องเป็นคนที่ต้องผลิตเนื้อหาและดูแลบริหารจัดการข้อมูลที่เกิดจากการทำงาน เพื่อนำไปสู่การสร้างฐานข้อมูล ซึ่งแน่นอนความรู้พวกนี้ไม่มีสอนในวิชาที่เรียน เป็นความรู้ที่เกิดขึ้นจากการทำงานที่นักวิชาการของเราต้องเรียนรู้ในระหว่างทำงาน

ตัวอย่างเช่น คุณดอกรัก เป็นนักวิจัยทางด้านเอกสารโบราณ แต่สุดท้ายภายใต้ทักษะที่เขาสร้างขึ้นมา เขาต้องกลายเป็นคนที่ออกแบบ metadata สำหรับเอกสารโบราณ

สิ่งที่เราต้องเรียนรู้เพิ่มเติมก็คือ เรารู้แล้วว่าถ้าจะจัดการข้อมูล เราจะต้องมี work flow ในการทำงาน ซึ่ง work flow ในการทำงานที่เรียกว่า Lifecycle เราก็พยายามหาโมเดลในการทำงานว่ามีโมเดลอะไรที่จะสอดคล้องกับการทำงานของเรา ซึ่งจะมี Digital Curation Lifecycle Model เป็นโมเดลในการจัดการข้อมูลงานวิจัยที่ใช้กันมาในฝั่งอังกฤษ ก็คือกิจกรรมที่เกิดขึ้นบ่อย ภาพ loop จากการทำงานจะเริ่มจากการวางแผน จะทำอะไร จะสร้างอะไร คัดเลือกข้อมูลที่เข้ามาอย่างไร นำเข้าอย่างไร สงวนรักษาอย่างไร จัดเก็บอย่างไร มันคือวงจรชีวิตที่ใช้ในระดับสากลที่เราต้องเรียนรู้

อย่างที่สอง ต้องรู้ว่ามี Standard Metadata สำหรับอธิบายโดเมนความรู้แต่ละแบบหรือวัตถุประสงค์การใช้งานแต่ละประเภท เช่น ภาพถ่ายจะใช้ metadata อะไร ฟิล์มใช้ metadata อะไร หรือถ้าอยากจะจัดการพิพิธภัณฑ์ใช้ metadata อะไร สิ่งเหล่านี้มี standard อยู่แล้ว สามารถหยิบมาใช้ได้ก่อน ถ้าไม่ตอบโจทย์ค่อยหาวิธีประยุกต์ ปรับใช้

สำหรับการจัดการที่เป็น collection ทางมรดกทางวัฒนธรรม อาจจะแบ่งไปได้ว่ามันมีตัวที่เป็นโครงสร้าง เนื้อหา format และการแลกเปลี่ยนที่จะต้องคำนึงถึง แต่สำหรับกลุ่มคนที่ทำงานด้านเนื้อหาอย่างเรา ก็คือเราใส่ใจแค่โครงสร้างและเนื้อหาก็พอ

ถ้าจะให้เปรียบเทียบให้เห็นภาพระหว่าง Data Structure และ Data Content ก็คือ

เหมือนเรามีของเหลว 1 ชนิด เราต้องการจัดของเหลวชนิดนั้น เราต้องการขวดแบบไหนที่จะใส่ของเหลว ขวดกลม ขวดเหลี่ยม ขวดเตี้ย ขวดสูง คือสิ่งที่เราทำควบคู่กันไป จากนั้นภาพใต้ขวดหลาย ๆ ใบ เราจะเก็บมันภายใต้ รูปแบบ อะไร เช่น ใส่ลัง ใส่กล่อง ก่อนที่จะนำไปสู่การแลกเปลี่ยนด้วยวิธีไหน เช่น ขนด้วยรถเข็น หรือขนด้วยคนยก นี่คือหลักการง่าย ๆ ก็คือ ถ้าคิดจะทำฐานข้อมูลสักหนึ่งฐาน น่าจะต้องคำนึงถึงองค์ประกอบ 4 ด้านนี้ก่อน คือ โครงสร้าง เนื้อหา รูปแบบ และการแลกเปลี่ยน

ในส่วนของโครงสร้างและเนื้อหา ก็จะมี Standard ในการใช้งาน เช่น ถ้าเป็น Material Culture ก็จะใช้ CDWA สำหรับการอธิบาย ซึ่งเป็นที่น่าดีใจว่าทางพิพิธภัณฑ์ธรรมศาสตร์ กำลังพัฒนาตัวนี้อยู่ ถ้าประสบความสำเร็จเราก็จะมี CDWA ฉบับไทยเป็นฉบับแรกที่นำไปใช้ได้ หรือห้องสมุดก็จะมี MARC หรือข้อมูลจดหมายเหตุ ก็จะใช้ ISAD

คำถามก็คือ แล้วการจัดการข้อมูลด้านสังคมศาสตร์ มนุษยศาสตร์ เราจะเลือกเสนอภายใต้หรือเนื้อหาอะไรที่เป็นตัวมาตรฐาน เพราะตอนนี้ถ้าเราแบ่งตามโดเมน หรือวัตถุประสงค์ในการใช้งานก็มีอยู่ แต่ในความเป็นจริง ถ้าเป็นคนที่เห็นข้อมูลจากสนามมาก ๆ เวลามาทำจะมีความงงว่าสามารถใช้ได้จริงหรือ เพราะมันอาจจะไม่สามารถอธิบาย หรือจัดเก็บข้อมูลจากภาคสนามได้ทั้งหมด นี่เป็นสิ่งที่ศูนย์ฯ เผชิญอยู่

ณ ตอนนี้ก็สามารถบอกได้ว่าในส่วนของประเทศไทย ยังไม่มีใครที่สามารถบอกได้ชัดเจนว่าข้อมูลทางด้านสังคมศาสตร์ มนุษยศาสตร์ควรจะใช้โครงสร้าง หรือเนื้อหาอะไรเป็นมาตรฐาน แต่โดยพื้นฐานรวม ๆ เราจะใช้ Dublin Core เป็นตัวทั่วไปในการอธิบายข้อมูล

ในอนาคตเราจะทำอะไร จากที่เราไปศึกษาแล้วว่ามันมีวงจรชีวิตในการจัดการข้อมูลแล้วจะเอามาประยุกต์ในการจัดการได้อย่างไร เราลองถอดบทเรียนมา คือหยิบตัวแกนของเขามา แต่ในตัวรายละเอียดเราก็ใส่ข้อมูลที่นักวิจัยของเราทำจริง ปฏิบัติจริง คือ ก่อนที่จะเริ่มโครงการสักหนึ่งเรื่อง แน่นอนว่าจะต้องมีการพูดคุยกันทั้งในระดับผู้บริหารและผู้ปฏิบัติงานว่าจะเก็บสะสมอะไร จะนำไปใช้ประโยชน์อะไร ใครเป็นกลุ่มเป้าหมาย จะเก็บไว้นานเท่าไหร่ จากนั้นจึงนำไปสู่การสร้างและรับมอบ ซึ่งอย่างที่ทราบว่าเนื่องด้วยภายใต้สภาวะแวดล้อมดิจิทัล ข้อมูลมันเกิดขึ้นง่าย จากเมื่อก่อนเคยใช้ฟิล์มม้วน 36 - 37 รูป เดี๋ยวนี้ไปภาคสนามกลับมามีรูปเป็นพัน ซึ่งจะต้องมีการประเมินและคัดเลือกเบื้องต้น ก่อนที่จะนำไปสู่การจัดการและจัดเก็บ เพราะฉะนั้นนโยบายในการจัดเก็บควรจะค่อนข้างชัดเจนตั้งแต่แรกเริ่มก่อนการทำงาน และนำไปสู่การให้รายละเอียดข้อมูล รวมไปถึงการอนุรักษ์ รักษา การจัดเก็บ การเผยแพร่ เข้าถึง และใช้ซ้ำ ซึ่งทั้งหมดนี้เป็นสิ่งที่จะต้องแจกแจงในรายละเอียด

ต่อมางานที่เราทำศูนย์ฯ มองว่าข้อมูลที่จะเก็บ ไม่ใช่เพียงข้อมูลทั่วไป หรือเป็นตัวเลขทางเศรษฐกิจ ข้อมูลที่เราเก็บคือสิ่งที่เกี่ยวข้องกับวิถีชีวิต เกี่ยวข้องกับคน ซึ่งมีทั้งรูปแบบที่จับต้องได้และจับต้องไม่ได้

ข้อมูลที่จับต้องได้ (Tangible) ก็คืออนุสรณ์สถาน โบราณคดี วัตถุทางวัฒนธรรม แต่วัฒนธรรมที่จับต้องไม่ได้ (Intangible) คือตัวความรู้ของผู้สร้าง หรือขั้นตอน พิธีกรรมในการประกอบของคนผลิต ทั้งหมดเหล่านี้จะจัดเก็บหรือจัดการได้อย่างไร ขั้นตอนเหล่านี้เราใช้การ Capturing และ Curating

Capturing เช่น เราลงภาคสนาม ถ่ายภาพ อัดเสียง จากนั้นกลับมาทำ Metadata หรือแปลงไฟล์ และ Curating ก็คือ ก่อนที่จะนำไปสู่การให้รายละเอียดสารสนเทศ ว่าข้อมูลที่เกิดขึ้นมีใครอยู่ในภาพ เขากำลังทำอะไร เกิดขึ้นที่ไหน เมื่อไหร่ ซึ่งเป็นข้อมูลพื้นฐาน ก่อนที่จะนำไปสู่การประเมินและคัดเลือกเพื่อจะวิเคราะห์ว่าจะนำไปสู่การสร้างข้อมูลอ้างอิงได้อย่างไร ซึ่งเราต้องตั้งคำถามกับข้อมูลชุดนี้ว่า ข้อมูลนี้ถูกทำขึ้นอย่างไร ทำขึ้นทำไม นี่คือหลักพื้นฐานในการให้รายละเอียดข้อมูล

แต่ทั้งนี้ทั้งนั้น ข้อมูลที่เกิดขึ้นมีสิ่งที่เรียกว่าบริบทของการเกิดขึ้น โดยเฉพาะบริบทของการเกิดขึ้นทางวัฒนธรรมที่ผูกติดอยู่กับชุมชน มีสิ่งที่เป็นทั้งภูมิปัญญาทางวัฒนธรรม การแสดงออกทางวัฒนธรรม วิถีชีวิต และความทรงจำของคน ซึ่งไม่ใช่ของใครคนใดคนหนึ่งในชุมชน แต่เป็นความทรงจำร่วมกันของคนในชุมชนที่ถ่ายทอดกันมาจากรุ่นสู่รุ่น ส่งต่อกันมา อันนี้เป็นคำถามว่าเราจะแชร์มันอย่างไรให้เกิดประโยชน์ทั้งต่อองค์กรและชุมชนที่เป็นเจ้าของทางวัฒนธรรม

ตัวอย่างที่จังหวัดภูเก็ต ที่ซอยรมณีย์ มีตึกสวย คนมักชอบไปถ่ายรูปกันมาก แต่น้อยคนที่จะรู้ว่าเมื่อก่อนซอยนี้คือย่านโคมเขียวโคมแดง แต่ทุกวันนี้กลายเป็นที่เช็คอิน ถ่ายรูป วัยรุ่นชอบไปกัน เทียบกับอีกรูปที่ซอยชบาดง ก็เป็นย่านโคมเขียว โคมแดงเหมือนกัน แต่อาจจะไม่ได้มีตึกให้ไปเช็คอินถ่ายรูป เป็นแค่ตึกแถวพาณิชย์อายุ 30 – 40 ปีที่เปิดให้บริการ เป็นที่ที่กลุ่มคนวัยรุ่น วัยหนุ่มในชุมชนได้มีประสบการณ์การใช้ชีวิตร่วมกัน เมื่อได้คุยกับกลุ่มผู้เฒ่าหรือวัยกลางคนทุกคนก็จะรู้จักว่านี่ คือพื้นที่ที่มีความทรงจำร่วมกันของคนในชุมชน แต่ถ้าไปคุยกับกลุ่มคนอาวุโสที่เป็นผู้หญิงที่อยากจะรักษาภาพลักษณ์ของคนในชุมชน ก็จะบอกว่าข้อมูลแบบนี้มันไม่ควรจะให้คนอื่นรู้ว่าที่นี่มีแบบนี้ ไม่ค่อยจะดีเท่าไหร่ เสียงชุมชนแตกเป็นสองเสียง ข้อมูลแบบนี้เราจะจัดการมันอย่างไร

อย่างที่สอง ถ้าข้อมูลสารสนเทศมรดกทางวัฒนธรรมถูกนำไปใช้ในทางที่ผิด จะเป็นอย่างไร ตัวอย่างเช่น “วิวาห์ เงาะป่าซาไก” แต่งตัวแฟนซี ถามว่าคนเงาะป่าซาไก หรือเรียกว่า มานิ ในชีวิตจริงแต่งตัวแบบนี้ใช่ไหม ไม่ใช่ อันนี้คือภาพมายาคติที่เราติดยึดจากตัวละครเวที ที่ใส่ชุดแดง ตัวดำ มีดอกไม้ทัดหู จริง ๆ เขาก็ใช้ชีวิตปกติ แต่เมื่อถูกนำเสนอโดยมีเป้าหมายเพื่อการท่องเที่ยว จึงถูกปรุงแต่ง ถือเป็นการลิดรอนวัฒนธรรมของเขา ซึ่งแม้แต่ชื่อเรียก “เงาะป่าซาไก” ก็มีความหมายในเชิงดูแคลนอยู่แล้ว ชื่อที่เขาใช้เรียกตัวเองจริง ๆ คือ “มานิ” ที่แปลว่าคน อันนี้คือสิ่งที่เราในฐานะที่เป็นหน่วยงานในการเผยแพร่ข้อมูล จะต้องให้ข้อมูลในสิ่งที่ถูกต้อง ซึ่งไม่ควรจะให้ข้อมูลไปในทางที่ทำให้เจ้าของวัฒนธรรมเกิดความเสียหาย

จากปัญหาเหล่านี้เองได้นำไปสู่โครงการที่ศูนย์ฯ กำลังดำเนินการอยู่ คือโครงการพัฒนาคลังข้อมูลชุมชนและเครือข่ายข้อมูล ความคิดหลักก็คือ ก่อนหน้านี้ผู้เก็บข้อมูลจะเป็นนักวิจัยถ่ายภาพ บันทึกเสียง เอาข้อมูลกลับมา เขียนรายงาน ตีพิมพ์ คำถามคือ ชุมชนได้อะไร นอกจากได้ความภาคภูมิใจว่าคนนอกเห็นคุณค่าของมรดกบ้านเรา โครงการนี้จึงเกิดขึ้นมาด้วยความคิดที่ว่า ให้ชุมชนเป็นคนเก็บข้อมูล จัดการข้อมูล และใช้ประโยชน์ข้อมูลด้วยตนเองดีกว่า สิ่งที่สถาบันวิชาการจะทำก็คือเข้าไปฝึกอบรมกระบวนการจัดเก็บและจัดการข้อมูล พร้อมกับสร้างสิ่งอำนวยความสะดวกอะไรก็ตามเพื่อให้ชุมชนสามารถทำงานบน platform ได้ ซึ่งโครงการนี้อยู่ในระยะปีที่ 2 แล้ว ก็คือข้อมูลที่ชุมชนเก็บมาจะใช้ประโยชน์ทำอะไร บางชุมชนก็สนใจจะทำพิพิธภัณฑ์ บางชุมชนก็อยากทำผลิตภัณฑ์ชุมชน บางชุมชนก็อยากทำหนังสือเรื่องเล่าของชุมชนตัวเอง

เริ่มจากกิจกรรมอบรมเชิงปฏิบัติการที่ศูนย์ฯก่อน แล้วก็ลงพื้นที่เพื่อจัดทำแผนที่เดินดิน คือการเดินสำรวจตามบ้าน ทำแผนผังว่าคนอยู่ที่ไหน อย่างไร จุดประสงค์ของการทำแผนที่เดินดินก็เพื่อให้ชุมชนที่คิดว่าเขารู้จักกันดีแล้ว เขารู้จักกันดีจริง หรือ ก็ให้เดินสำรวจอีกครั้งว่าจริง ๆ แล้วเกิดการเปลี่ยนแปลงชุมชนอย่างไร

ตัวอย่างพื้นที่ที่ อำเภอนาแห้ว จังหวัดเลย สิ่งที่เขาอยากจะเก็บก็คือ บ่อเกลือโบราณ เมื่อก่อนชุมชนนี้เป็นศูนย์กลางการทำเกลือ แล้วส่งไปโดยรอบ แต่ตอนหลังถูกเกลือจากอีสานใต้มาตีตลาด จึงทำให้ต้องหยุดไป ตระกูลที่ทำเกลือก็ไม่ค่อยทำกันแล้ว ก็ถูกละเลยไป เขาก็อยากเก็บข้อมูลชุดนี้ หรือชุมชนมัสยิดที่บางอ้อ สิ่งที่ทำคือประวัติศาสตร์ชุมชนว่า เขาเป็นใคร มาจากไหน เมื่อไหร่ แล้วก็ทำปฏิทินชุมชนว่าในรอบหนึ่งปีเขามีเหตุการณ์อะไรเกิดขึ้นบ้าง สิ่งที่โครงการลงไปทำก็คือ ให้คนในชุมชนเก็บบันทึกข้อมูลชุดนี้ขึ้นมา

ภาพที่ 33 : กิจกรรมฝึกอบรมการเก็บและจัดการข้อมูลชุมชน[11]

[11] ภาพสไลด์ของคุณสิทธิศักดิ์ รุ่งเจริญสุขศรี จากการการประชุมเสวนาแลกเปลี่ยนความรู้ “การจัดทำและนำเสนอคลังข้อมูลดิจิทัลออนไลน์” วันที่ 27 พฤศจิกายน 2562 หน้าที่ 29

Work flow ของงานแบ่งเป็น 4 ประเภทหลัก ๆ ก็คือ เก็บข้อมูลมาแล้วก็จัดกลุ่มจัดโฟลเดอร์ จากนั้นก็นำเข้าสู่ Community Drive ซึ่งมีฟังก์ชั่นคล้าย ๆ กับ onedrive แล้วถ้าอยากจะหยิบข้อมูลชิ้นไหนเป็นพิเศษเพื่อให้รายละเอียด หรือ description ก็นำไปสู่การสร้างหรือเผยแพร่บนเว็บไซต์ ก็เข้าไปสู่ Content Management System หรือ CMS แล้วก็เผยแพร่ในระบบ www อันนี้เป็นตัวระบบ ซึ่ง ณ ตอนนี้ระบบอยู่ในตัว onedrive

สิ่งที่เราสนใจอีกอย่างหนึ่งก็คือ นอกจากการให้รายละเอียดข้อมูลแล้ว เราควรจะให้ความรู้กับผู้นำไปใช้ด้วย ว่าเขาควรใช้ข้อมูลภายใต้บริบทข้อมูลดั้งเดิมได้อย่างไร สิ่งนี้เรียกว่า Traditional Knowledge Licenses ก็คือป้ายกำกับการใช้ข้อมูลทางวัฒนธรรม จริง ๆ ก็เป็นโครงการที่เชื่อมโยงกับ Mukurtu ที่ผู้ช่วยศาสตราจารย์ ดร.ศิริเพ็ญได้นำเสนอ ก็คือ Mukurtu เป็น CMS ในการจัดการแต่ตัว TK licenses ก็คือนักวิจัยในทีมเสนอ ที่เราคุ้นชิน เช่น CC licenses ที่ว่าด้วยการให้สละสิทธิ์ในการใช้ หรือจะใช้ด้วยวัตถุประสงค์อะไรก็ตาม แต่ TK licenses ก็คือให้ความรู้ทางวัฒนธรรมว่า ข้อมูลนี้เดิมทีถูกใช้ภายใต้ประโยชน์อะไร เช่น ใช้โดยผู้ประกอบพิธีกรรมในชุมชน ใช้เฉพาะผู้ชายเท่านั้น ใช้ในบางฤดูกาล อันนี้ศูนย์ฯ พยายามหยิบมาใช้ แต่ติดขัดปัญหาว่าบริบทของไทยกับบริบทของต่างประเทศซึ่งถูกคิดค้นจากกลุ่ม American Indian หรือ Aborigin มันค่อนข้างจะต่างกัน พอของเราพบว่าข้อมูลที่เขากังวลคือ ข้อมูลส่วนบุคคลที่ไม่ใช่ข้อมูลชุมชน

สรุปก็คือ ถ้าหน่วยงานของท่านจะต้องทำคลังข้อมูลหรือฐานข้อมูล ควรจะคุย ปรึกษาให้ดีก่อนว่าจะจัดเก็บอะไร จัดการอย่างไร ต้องมี work flow ต้องหาผู้ใช้งาน ใครเป็นกลุ่มเป้าหมายที่จะเข้ามาใช้ จะได้นำไปสู่การออกแบบ และที่สำคัญอีกอย่างคือ ถ้าเป็นข้อมูลที่มีเจ้าของข้อมูลทางวัฒนธรรม ชุมชนเป็นผู้ดูแลรับผิดชอบ เราไปเก็บไปดึงของเขามาใช้ เราจะมีวิธีการแชร์ข้อมูลกลับคืนสู่ชุมชนได้อย่างไร

นอกเหนือจากความรู้ในเชิงเนื้อหา นักวิจัยจะต้องมีความสามารถในการจัดการข้อมูล อันนี้เป็นสิ่งที่จะต้องเรียนรู้จากกระบวนการทำงาน และเป็นโจทย์ที่ท้าทายนักวิจัยทางด้านสังคมศาสตร์และมนุษยศาสตร์ว่า ในอนาคตภาระงานไม่ใช่แค่ลงไปเก็บข้อมูล เขียน และรายงานแล้ว แต่จะต้องจัดการข้อมูลที่เกิดขึ้นจากกระบวนการทำงานภาคสนามด้วย ซึ่งในต่างประเทศจะค่อนข้างชัดเจนว่า ถ้ารับทุนวิจัยจะต้องคืนข้อมูลที่เกิดจากการทำงานระหว่างการวิจัยส่งคืนเจ้าของทุนด้วย ซึ่งในอนาคตคนไทยก็จะใช้ระบบนี้เหมือนกัน จึงเป็นทักษะที่นักวิจัยในอนาคตจะต้องมี

ปิดท้าย อย่างที่บอกว่า การสื่อสาร การจัดการ และการจัดเก็บ เป็นสิ่งที่มนุษย์ทำกันอย่างต่อเนื่องและยาวนานแล้ว ไม่มีอะไรเปลี่ยนแปลง วัตถุประสงค์มีแค่นี้ว่าเราจะสื่อสารอย่างไร จัดการอย่างไร จัดเก็บอย่างไร แต่สิ่งที่เปลี่ยนก็คือรูปแบบและเครื่องมือที่ใช้ในการจัดเก็บ เปลี่ยนไปตามยุคสมัย อันนี้เป็นสิ่งที่เราจะต้องเรียนรู้ เพื่อที่จะได้ส่งผ่านความรู้ไปสู่นักวิจัยรุ่นใหม่หรือสังคมได้

4. ศูนย์เทคโนโลยีอีเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ (NECTEC)

โดยคุณวัชชิรา บูรณสิงห์ ผู้ช่วยวิจัยอาวุโส ศูนย์เทคโนโลยีอีเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ หรือ NECTEC ซึ่งเป็นผู้มีทักษะหรือมีประสบการณ์ในการจัดเก็บข้อมูลดิจิทัล รวมทั้งยังได้ทำโครงการที่เกี่ยวกับระบบการจัดการข้อมูลในหลายโครงการ อย่างเช่น โครงการ Information Technology System For Redcross Archive โครงการ Search Engine for Tourism Management System

คุณวัชชิรา บูรณสิงห์ได้อธิบายถึง ศูนย์เทคโนโลยีและคอมพิวเตอร์แห่งชาติได้ทำอะไรไปบ้าง และก้าวต่อไปจะทำอะไรเกี่ยวกับการจัดเก็บข้อมูลทางด้านวัฒนธรรม ซึ่งปัจจุบันได้ดำเนินงานโครงการเรื่องการจัดการท่องเที่ยวเชิงวัฒนธรรมด้วย concept ของโครงการชื่อว่า “นวนุรักษ์” มาจากคำว่า นวัต ที่แปลว่าใหม่ และ อนุรักษ์ ความหมายรวมคือการอนุรักษ์ในรูปแบบสมัยใหม่

โครงการนวนุรักษ์ เป็นโครงการที่เราพยายามจัดเก็บข้อมูลให้อยู่ในรูปแบบมาตรฐาน แล้วเราจะนำข้อมูลเหล่านั้นไปใช้ประโยชน์ได้อย่างไร มันคือการเก่าเก็บ เพื่อก่อเกิด

นวนุรักษ์แพลตฟอร์ม จะเป็น software platform ที่จะบริหารจัดการข้อมูลและให้บริการข้อมูลทางด้านวัฒนธรรม โดยจะเล่าว่าโครงการนี้ทำอะไรบ้าง และใครบ้างที่สามารถนำข้อมูลนี้ไปใช้ และใช้ในรูปแบบไหนได้บ้าง

นวนุรักษ์เป็นโครงการที่แปลงข้อมูลให้อยู่ในรูปแบบดิจิทัล หมายความว่านอกจากที่จะแปลงแล้ว ยังสามารถที่จะสกัดองค์ความรู้ที่อยู่ในคลังข้อมูลให้เราสามารถเห็นได้ในหลายมิติของข้อมูล และนำเสนอข้อมูลเหล่านั้นให้น่าสนใจ

ปัญหาและความสำคัญของโครงการที่ริเริ่มทำ เราเห็นว่ามีพิพิธภัณฑ์เกิดใหม่ขึ้นเป็นจำนวนมาก ซึ่งเป็นข้อมูลที่ได้จากศูนย์มานุษยวิทยาสิรินธร รวมถึงศูนย์การเรียนรู้ที่จัดเก็บข้อมูลด้านศิลปะและวัฒนธรรม ซึ่งเป็นเรื่องที่น่ายินดี แต่ว่ากระบวนการที่ใช้ในการจัดเก็บข้อมูล เครื่องมือที่จะนำมาจัดเก็บมีเพียงพอหรือไม่ สามารถที่จะนำมาใช้ได้อย่างมีประสิทธิภาพหรือไม่ และที่สำคัญเนื้อหาหรือข้อมูลที่ไม่ได้อยู่ในรูปแบบโครงสร้างจะนำมาทำประโยชน์ได้น้อย ฉะนั้นทำอย่างไรให้ข้อมูลอยู่ในรูปแบบที่เป็นโครงสร้าง สามารถไปต่อยอดทางนวัตกรรมได้เพิ่มมากขึ้น

นวนุรักษ์ มีประวัติความเป็นมาค่อนข้างจะยาวนาน ตั้งแต่ปี 2555 ตอนแรกเป็นแค่การพัฒนาให้แต่ละที่ ถ้าที่ไหนอยากทำเรื่องการบริหารจัดเก็บข้อมูลให้เป็นระบบในพิพิธภัณฑ์ เราก็ไปทำให้แต่ละที่ ต่อมาเราพัฒนาเป็น platform ให้ทุกคน หน่วยงาน หรือชุมชนสามารถจัดเก็บข้อมูลได้ด้วยตนเอง และนำข้อมูลเหล่านั้นไปเผยแพร่ได้ ปัจจุบันที่เป็น platform สิ่งที่เพิ่มขึ้นคือรูปแบบในการจัดเก็บเพิ่มมากขึ้น ลูกเล่นในการจัดเก็บที่มากขึ้น การแชร์ข้อมูลและการนำข้อมูลไปใช้ได้อย่างมีประสิทธิภาพมากขึ้น จะมีเทคโนโลยีต่าง ๆ ที่นำข้อมูลทางวัฒนธรรมไปใช้ได้มากยิ่งขึ้น ทั้งนำไปใช้ในการสร้างเกมส์ต่าง ๆ

มีหลายคนบอกว่าคนทำงาน IT ไม่เข้าใจโดเมน เวลาที่พัฒนาอะไรออกมา บางครั้งคนที่ทำงานด้านสังคมศาสตร์อาจจะบอกว่าใช้งานไม่ได้ อันนี้เป็นเรื่องที่ทีมเจอบ่อยมาก ดังนั้นทีมวิจัยนวนุรักษ์ จึงมีบทบาทในการทำระบบ ทำ platform ให้สอดคล้องกับคนที่ทำงานด้านจัดเก็บข้อมูลทำงานได้ง่ายยิ่งขึ้น และตอบโจทย์ที่เขาต้องการเพื่อให้ใช้งานได้มากขึ้น

Platform นี้เป็นภาพใหญ่ ประกอบไปด้วย 4 ส่วนด้วยกัน คือ ส่วนที่หนึ่ง **Local Community** เป็นตัวที่มีระบบบริหารจัดการให้เป็น platform ที่ทาง NECTEC พัฒนาขึ้น และเมื่อได้จัดเก็บข้อมูลในแต่ละที่แล้ว เรามีการบูรณาการร่วมกัน ฉะนั้นบทบาทของเราคือ ถ้าใครไม่มีที่จัดเก็บข้อมูล เรามีระบบจัดเก็บให้ และมีพื้นที่ให้ ส่วนใครที่มีข้อมูลอยู่แล้วก็ถือเป็นเรื่องที่น่ายินดี เพราะบางที่มีศักยภาพมากพอที่มีพื้นที่ในการจัดเก็บ มีบุคลากรมากพอที่สามารถจัดเก็บได้เอง จากนั้นเรามีศูนย์ข้อมูลตรงกลาง เพื่อกระจายข้อมูลออกไปให้คนอื่นมาใช้งานได้อย่างไร

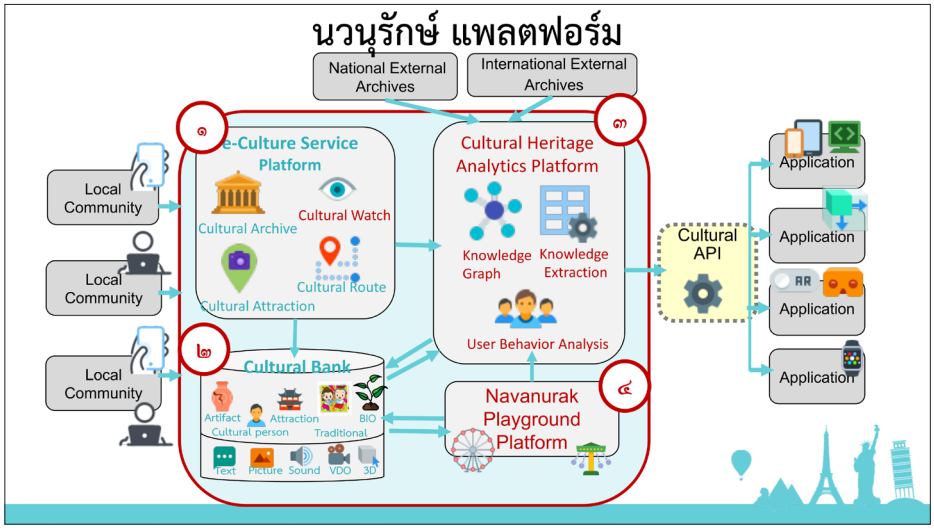
เราจะมีระบบการบริหารจัดการคลังข้อมูล มีระบบจัดเก็บข้อมูลการท่องเที่ยวเพื่อวัฒนธรรมด้วย และอีกเรื่องคือ Cultural Watch ซึ่งเกิดมาจากโจทย์ที่ว่า ทำอย่างไรเราถึงจะเห็นการเปลี่ยนแปลงทางด้านวัฒนธรรม ทำอย่างไรถึงจะรู้ว่าวัฒนธรรมที่เปลี่ยนแปลงดีขึ้นหรือแย่ลง ซึ่งเรามี application ที่ชื่อว่า Fondue ซึ่งจะถูกใช้งานในเรื่องการแจ้ง น้ำไม่ไหล ท่อประปาแตก แต่ ณ ปัจจุบัน Fondue จะมีฟังก์ชั่นให้คุณสามารถแจ้งสิ่งที่สุ่มเสี่ยงต่อการสูญเสียทางด้านวัฒนธรรมหรือไม่ และข้อมูลหนึ่งที่เราได้จาก Cultural Watch คือข้อมูลที่ได้จากทาง Social Network ว่าเกิดอะไรขึ้นบ้างที่เกี่ยวกับทางวัฒนธรรมของบ้านเรา

ส่วนที่สอง เรียกว่า **Cultural Bank** คือการสร้างคลังข้อมูลทางวัฒนธรรมขนาดใหญ่ ซึ่งมีหลากหลายโดเมน และหลากหลายรูปแบบ การมีหลากหลายรูปแบบหมายความว่า เราสามารถใช้ข้อมูลร่วมกันได้ และไปใช้เผยแพร่ได้น่าสนใจมากยิ่งขึ้น

ส่วนที่สาม เรียกว่า **Cultural Heritage Analytics Platform** คือเวลาที่เรามีข้อมูลขนาดใหญ่ และจะสกัดความรู้ออกมา ทำให้อยู่ในรูปแบบที่เข้าใจง่ายยิ่งขึ้น หรือเห็นความเชื่อมโยงของข้อมูลได้ง่ายยิ่งขึ้น

ส่วนที่สี่ เรียกว่า **Navanurak Playground Platform** มาจากที่ว่า ไม่ใช่ว่าทุกหน่วยงานหรือทุกชุมชนจะสามารถสร้าง application ด้วยตนเอง ตรงนี้จะตอบโจทย์ว่า ถ้าคุณมีข้อมูลดี ๆ ทำเนื้อหาดี ๆ platform นี้สามารถสร้าง application ให้คุณได้โดยที่ไม่ต้องลงทุน เพียงแต่ทำข้อมูลดี ๆ และเลือก application ที่เหมาะสม

นอกจากนี้จะมี API คือ นำข้อมูลออกไปใช้งานได้อย่างไร ซึ่งจะมี software developer หรือกลุ่มผู้พัฒนา software นำข้อมูลนี้ไปสร้าง application ต่าง ๆ นอกจากนี้ในระบบสามารถให้คุณเลือกปิด หรือ เปิดเผยข้อมูล หรือจะเผยแพร่หรือไม่เผยแพร่ก็ได้ ซึ่งเป็นสิ่งที่เราค่อนข้างกังวล เพราะเรารู้ว่าบางข้อมูลควรจะเก็บแต่ว่าไม่จำเป็นต้องเผยแพร่ก็ได้



ภาพที่ 34 : แสดง นวนุรักษ์ แพลตฟอร์ม[12]

กระบวนการทำเรื่องการจัดเก็บข้อมูลดิจิทัล ประกอบด้วย 5 กระบวนการ

ขั้นตอนแรก คือ Collection ซึ่งเป็นการจัดเก็บข้อมูลเบื้องต้น การแยกข้อมูลเบื้องต้น

ขั้นตอนที่สอง คือ Digitization คือการแปลงข้อมูลให้อยู่ในรูปแบบดิจิทัล ซึ่งปัจจุบันมีหลากหลายรูปแบบมาก เช่น ถ่ายรูป ทำคลิปวิดีโอ อัดเสียง หรือทำภาพสามมิติ ซึ่งนวนุรักษ์ platform สามารถที่จะรองรับได้

ขั้นตอนที่ 3 คือ Data Management ขั้นตอนนี้คือ การที่แปลงไฟล์เป็นดิจิทัลแล้วเราจะเก็บข้อมูลอย่างไร ก็จะมีระบบและเครื่องมือในการจัดเก็บข้อมูล ซึ่งคล้ายกับของศูนย์มานุษยวิทยา

ขั้นตอนที่ 4 คือ Data Integration and Analysis คือวิเคราะห์ข้อมูลให้เห็นมิติต่าง ๆ ของข้อมูล

ขั้นตอนที่ 5 Visualization เป็นขั้นตอนที่จะนำเสนอข้อมูลอย่างไรให้น่าสนใจ ซึ่งเป็นขั้นตอนที่ขายดีมาก เพราะหลายคนไม่รู้ว่าจะนำข้อมูลไปนำเสนออย่างไรให้น่าสนใจ บางคนมีพิพิธภัณฑ์หรือศูนย์การเรียนรู้ บางที่คือการเอาคอมพิวเตอร์ไปวาง ให้ผู้เข้าชมมาค้นหา ตรงจุดนี้อาจจะเพิ่มความน่าสนใจได้มากยิ่งขึ้น ข้อมูลจะได้ดูสนุกมากขึ้น

สุดท้ายสิ่งที่เราจะได้ร่วมกันก็คือคลังข้อมูลทางวัฒนธรรมขนาดใหญ่ ซึ่งมีการเชื่อมโยงข้อมูลซึ่งกันและกัน

สิ่งที่ทางทีมวิจัยทำคือ เรามีระบบการบริหารจัดการทางวัฒนธรรม มีระบบข้อมูลวัตถุที่อยู่ในพิพิธภัณฑ์ มีการจัดเก็บข้อมูลใบลาน จดหมายเหตุ และนิทรรศการ ซึ่งเป็นรูปแบบ web application สามารถจัดเก็บข้อมูลได้ทั้งข้อมูล ภาพ เสียง หรือวิดีโอ หรือถ้าเป็นพิพิธภัณฑ์

[12] ภาพสไลด์ของคุณวัชชิรา บูรณสิงห์ จากการการประชุมเสวนาแลกเปลี่ยนความรู้ “การจัดทำและนำเสนอคลังข้อมูลดิจิทัลออนไลน์” วันที่ 27 พฤศจิกายน 2562 หน้าที่ 8

อยากให้มีการนำชม ก็สามารถสร้าง QR Code ได้ด้วยตนเอง ซึ่งเป็นเทคโนโลยีที่ง่ายที่สุดและถูกที่สุด ไม่ต้องลงทุนมาก

ในนั้นจะจัดเก็บภาพ 360 องศา ภาพวัตถุที่สามารถหมุนได้โดยรอบ ซึ่งทางทีมวิจัยฝึกให้ชุมชนสามารถทำได้เอง สามารถใช้กล้องมือถือถ่ายวิดีโอได้เอง หากไม่มีกล้องราคาแพง มีไฟล์เอกสาร มีภาพพาโนรามา รวมทั้งภาพสามมิติ

นอกจากนี้ก็มีระบบการบริหารจัดการคลังข้อมูลทางวัฒนธรรมที่เก็บจดหมายเหตุและใบลาน ซึ่งมีการแปลพอสังเขป

โจทย์อีกข้อหนึ่งที่ทางทีมวิจัยเจอคือ เวลาที่ชุมชนจัดเก็บข้อมูลเอง เขาไม่สามารถจะนำรูปภาพที่อยู่ในกล้องมาใส่ลงคอมพิวเตอร์เองได้ หรือไม่มีอินเตอร์เน็ต เราก็มี Mobile Application ที่เรียกว่า “เดินทุ่ง Mobile” สามารถทำทุกอย่างอยู่ในมือถือได้

นอกจากนี้เรามีโครงการที่สร้างเส้นทางการท่องเที่ยวทางวัฒนธรรมด้วย ซึ่งสามารถสร้างเส้นทางการท่องเที่ยวทางวัฒนธรรมได้เอง เพราะเราเชื่อว่าชุมชนหรือหน่วยงานรู้ว่าเส้นทางหรือการท่องเที่ยวตรงไหนที่อยากจะแนะนำ

ถ้าอยากจะทำเส้นทางการท่องเที่ยวแต่ละจุด ก็สามารถสร้าง QR Code ได้เอง นอกจากนั้นถ้าสถานที่นั้นมีวัตถุหรือข้อมูลที่น่าสนใจ ก็สามารถแสดงข้อมูลเหล่านั้นได้

อีกอย่างหนึ่งคือ Navanurak Playground อย่างที่บอกไม่ใช่ทุกคน หรือทุกหน่วยงานจะสร้าง application เองได้ คุณแค่ทำเนื้อหาดี ๆ แล้วนำมาบริหารจัดการ สามารถเลือกข้อมูลไปทำเกมส์ หรือ playground แล้วสร้าง QR Code ซึ่งแค่ไปสถานที่นั้นแล้วอยากได้ข้อมูลก็สามารถใช้ QR Code ได้เลย ไม่ต้องโหลด application ก็ได้

เรามีกระบวนการทำงานร่วมกับมหาวิทยาลัย เพราะเราเชื่อว่ามหาวิทยาลัยหลายที่มีความใกล้ชิดกับชุมชน สิ่งเหล่านี้ก็จะเป็นเครื่องมือหนึ่งที่สามารถนำไปใช้ในการจัดเก็บข้อมูลของชุมชน โดยที่มีนักศึกษาลงไปเก็บข้อมูลร่วมกับชุมชน หรือบางที่ เช่น มหาวิทยาลัยราชภัฏเชียงใหม่ก็จะนำเครื่องมือนี้ไปอยู่ในรายวิชา หรือบางที่ชุมชนเข้มแข็ง สามารถดูแล บริหารจัดการได้ด้วยตนเองก็ใช้เป็นเครื่องมือในการจัดเก็บข้อมูล

นอกจากนี้ในแต่ละปี จะมีกิจกรรมส่งเสริมผู้พัฒนา software หรือ Hackathon คือการนำข้อมูลเหล่านี้มาพัฒนาหรือสร้าง application ได้อย่างไร จากโจทย์นี้จะเห็นว่าข้อมูลที่มีอยู่สามารถนำไปสร้าง application ดี ๆ ได้

ประโยชน์จากคลังข้อมูล ถ้าเราเก็บข้อมูลที่เป็นมาตรฐาน หรือดีพอ จะสามารถนำไปใช้ในงานอะไรได้บ้าง แน่นอนเรื่องการศึกษาที่เป็นเนื้อหาความรู้ สามารถใช้ในด้านการท่องเที่ยวก็ได้ สามารถที่จะไปสร้างนวัตกรรมใหม่ ๆ ได้ นอกจากนี้ยังเป็นส่วนหนึ่งที่จะสามารถเพิ่มมูลค่าสินค้าทางวัฒนธรรม อย่างเช่น สินค้าหนึ่งชิ้นจะมี QR Code ติดกับตัวสินค้า พอเข้าไปดูก็จะพบว่า ร่มแต่ละอันทำอย่างไร มีลวดลายแตกต่างกันอย่างไร

นอกจากนี้การนำข้อมูลไปต่อยอดทางเทคโนโลยี ซึ่งหลายคนอาจจะรู้สึกว่า การนำข้อมูลไปทำเว็บ หรือทำ application ต่าง ๆ ต้องเป็นสายคอมพิวเตอร์เท่านั้นหรือเปล่า จริง ๆ แล้วมีเครื่องมือหลายชิ้นมากที่จะสามารถนำข้อมูลไปต่อยอดได้ อย่างเช่น Application AR หรือ Hologram ก็สามารถทำได้

สามารถเข้าไปดูรายละเอียดได้ที่ www.navanurak.in.th ก็จะมีข้อมูลต่าง ๆ ว่าเก็บข้อมูลอะไรบ้าง มีใครที่เข้าไปใช้งาน เช่น ตัวอย่างศูนย์การเรียนรู้ไทลื้อ ซึ่งทำงานร่วมกันระหว่างชุมชนกับมหาวิทยาลัยราชภัฏเชียงใหม่

วันนี้ถือว่าเรามาเล่าให้ฟังกันว่าตอนนี้เรามีการทำงานอย่างไร มีรูปแบบการทำงานแบบไหนและมีการจัดเก็บข้อมูล และสามารถนำข้อมูลมาเสนอแบบไหนที่สามารถเผยแพร่และดูน่าสนใจมากยิ่งขึ้น

คำถามจากผู้เข้าร่วมประชุม

มีคำถามว่าเนื่องจากประเทศไทยมีชุมชนที่มีคุณค่าทางวัฒนธรรมมากมาย แล้วทาง NECTEC หรือศูนย์มานุษยวิทยาสิรินธร มีวิธีการที่จะทำให้เข้าถึงความรู้ที่มีเยอะแยะมากมายได้อย่างไร เพราะว่าจากที่เราทำงานเรื่องของภาษาและวัฒนธรรมในภาวะวิกฤตในประเทศไทยปีหนึ่งไปได้ไม่กี่ชุมชน อยากทราบว่าในอนาคตทาง NECTEC จะขยายความรู้ให้ชุมชนสามารถทำได้เองอย่างไร

ตอบโดย คุณวัชชิรา บูรณสิงห์

NECTEC ก็ไม่ได้มีแรงขนาดนั้นที่จะทำทุกที่ วิธีที่เราทำงานคือ เราสร้างแม่ไก่ ซึ่งก็คือกลุ่มที่มีมหาวิทยาลัยหรือนักศึกษาที่เกี่ยวข้อง หรือมหาวิทยาลัยที่มีการบริการชุมชน เราจะทำงานร่วมกับกลุ่มนี้เพื่อที่เขาจะสามารถถ่ายทอดเทคโนโลยีสู่ชุมชนได้ และอีกโมเดลหนึ่งคือ ถ้าชุมชนนั้นเข้มแข็งพอ เขาสามารถเป็นแม่ไก่เองในการที่จะกระจายไปสู่ชุมชนอื่น เราใช้โมเดลนี้ในการถ่ายทอดเทคโนโลยีให้ไวขึ้น เวลาที่มีการ update เราจะทำงานกับเฉพาะส่วนที่เป็นแม่ไก่ แล้วเขาก็จะไป update กันต่อ

คำถามจากผู้เข้าร่วมประชุม

จากมหาวิยาลัยราชภัฏเชียงราย ขออนุญาตถาม เนื่องจากเรามีเนื้อหาในเรื่องวัฒนธรรมภูมิปัญญา แต่ว่ามีส่วนหนึ่งที่ทางมหาวิทยาลัยกำลังให้ความสำคัญคือเรื่องการอ่าน การเขียน ซึ่งมีการเชื่อมโยงเรื่องภาษาและวัฒนธรรมมาก แต่ว่ายังขาดในส่วนของเทคโนโลยีที่จะต่อยอดไปสู่ IOT ต่าง ๆ ไม่ทราบว่าทางด้านนี้จะเข้าร่วมในส่วนของนวนุรักษ์ด้วยได้ไหม

ตอบโดย คุณวัชชิรา บูรณสิงห์

ทาง NECTEC ทำในส่วนที่เป็นวัฒนธรรมซึ่งมีทั้งที่จับต้องได้ และจับต้องไม่ได้ รวมทั้งมีเรื่องทรัพยากรธรรมชาติ แต่เราสามารถคุยกันได้ว่าอาจารย์จัดเก็บข้อมูลประเภทไหน มีวิธีการจัดเก็บเป็นอย่างไร เพื่อที่ว่าจะลองดูในตัวระบบว่าสามารถจัดเก็บได้เลยหรือไม่ อย่างที่เรียนว่า บทบาทของ NECTEC คือกระบวนการสร้าง platform เพื่อตอบโจทย์ผู้ใช้งาน ดังนั้นสิ่งที่เราต้องการคือ ต้องขอความรู้จากผู้เชี่ยวชาญเพื่อพัฒนา platform ถ้าอาจารย์สนใจก็สามารถคุยกันได้ค่ะ ว่าเราจะทำงานร่วมกันได้อย่างไรบ้าง

คำถามจากผู้เข้าร่วมประชุม

ขอถาม เรื่อง Cultural Watch ว่าวิธีการสังเกตหรือการเฝ้าระวังเรื่องวัฒนธรรม จะมีกระบวนการอย่างไรบ้าง เพราะเรากำลังทำเรื่องภาษาและวัฒนธรรมในภาวะวิกฤต คิดว่ากำลังจะทำ ทาง NECTEC มี platform หรือมีวิธีการการเฝ้าระวังหรือการมองเรื่องพวกนี้อย่างไรบ้าง

ตอบโดย คุณวัชชิรา บูรณสิงห์

อยากบอกว่าชวนมาทำด้วยกันดีกว่า เพราะว่า NECTEC จะทำเทคโนโลยีได้ สมมติว่ากำลังพูดถึงข้อมูลวันลอยกระทง เราจะรู้ว่าวันลอยกระทงมันเกิดขึ้นที่ไหน และเกิดอะไรบ้าง คนกำลังพูดถึงตรงนี้ในแง่ดีหรือแง่ลบ วัฒนธรรมบางอย่างก็จะถูกพูดไปในทางที่ไม่ดีด้วย

ที่จะตอบคำถามอาจารย์คือ ถ้าอาจารย์สนใจตรงนี้ เราสามารถมาคุยกันได้ เพราะตอนนี้ที่เราทำได้คือใช้วิธีจาก social network กวาดข้อมูลจาก social network ทางด้านวัฒนธรรม เพื่อที่จะมาสกัดว่าเกิดข้อมูลจากที่ไหนบ้าง มันดีหรือมันไม่ดี และ Application Fondue ที่จะทำให้คนสามารถเข้ามาแจ้งเตือนได้ จริง ๆ อยากจะขอข้อมูลจากผู้เชี่ยวชาญ เพราะเราไม่ได้เชี่ยวชาญทางด้านวัฒนธรรม

คำถามจากผู้เข้าร่วมประชุม

มีความคิดเห็นแลกเปลี่ยนและคำถามนะคะ ส่วนแรกก็คือ อาจจะต่อเนื่องจาก Cultural Watch ดีหรือไม่ดี มันก็มีหลายระดับ เพราะจริง ๆ ในแต่ละยุคมีการเปลี่ยนแปลงทางวัฒนธรรมอยู่ตลอดเวลา เป็นหัวข้อที่น่าจะคุยกันต่อไป ความหลากหลายดี หรือ ไม่ดี ก็ตอบได้หลายอย่าง ไม่ใช่ ขาว ดำ ขนาดนั้น

คำถามคือ เห็นจากหลายหน่วยงานสนใจประเด็นวัฒนธรรมชุมชน การจัดเก็บ การอนุรักษ์เพื่อนำไปสู่การผลิตสิ่งใหม่ จากผู้สนใจข้างนอกเห็นว่าหลายหน่วยงานทำ ไม่แน่ใจว่าแต่ละหน่วยงานแบ่งกันอย่างไร หรือว่าอย่างของศูนย์มานุษยวิทยาสิรินธร มีความคล้ายหรือเหมือนกับของ NECTEC อย่างไร จึงไม่แน่ใจว่าเวลาจะหาข้อมูลควรหาที่ไหนดี อาจจะขอความเห็นจากทาง NECTEC และ

ศูนย์มานุษยวิทยาสิรินธร ว่าในบริบทของประเทศไทยมีกี่หน่วยงานที่จะเก็บข้อมูลอย่างไร สำหรับคนที่สนใจจะเริ่มที่ไหนก่อน หรือมีหน้าเว็บไหนที่รวบรวมหลายหน่วยงานให้ได้เห็นภาพรวมไหม

ตอบโดย คุณวัชชิรา บูรณสิงห์

จริง ๆ เราไม่ได้แบ่งพื้นที่นะ แต่ว่าวิธีการทำงานก็อาจจะแตกต่าง ถ้าเป็นของ NECTEC เราจะเน้นที่ว่า ถ้าใครอยากจัดเก็บข้อมูล แล้วไม่มีที่จัดเก็บ เรามี platform ให้ แล้ว NECTEC ไม่ใช่เป็นคนจัดเก็บข้อมูลเองด้วย ให้เจ้าของชุมชนเป็นคนเก็บ เพราะว่า output สุดท้ายที่เราคาดหวังคือเราต้องการ Cultural Bank แต่ถ้าถามว่าไปรวมกันที่ไหน จริง ๆ ปัจจุบันมันไม่จำเป็นต้องรวมกันที่ไหน เพราะว่ามันมี link data ซึ่งสามารถจะเชื่อมโยงข้อมูลกันได้

ทาง NECTEC ก็คุยกับศูนย์มานุษยวิทยาสิรินธรเรื่องการเชื่อมโยงข้อมูลเหมือนกัน ด้วยความที่เราเก็บข้อมูลเป็นมาตรฐานเดียวกัน การเชื่อมโยงข้อมูลก็จะไปด้วยกันได้ง่าย ข้อมูลที่ซ้ำซ้อนกัน มันซ้ำในลักษณะที่เป็นหัวข้อเดียวกัน แต่ในรายละเอียดอาจจะแตกต่างกัน

คำถามจากผู้เข้าร่วมประชุม

ในอนาคตจะมีที่ไหนที่ user ไม่ต้องไปหลายที่ หรือจะรู้ว่าที่ไหนมีข้อมูลต่างกันอย่างไร

ตอบโดย คุณวัชชิรา บูรณสิงห์

เป็นเรื่องของนโยบาย ซึ่งยังตอบตรงนี้อย่างชัดเจนไม่ได้ เพราะว่ามันเป็นหลายหน่วยงาน

ตอบโดย คุณสิทธิศักดิ์ รุ่งเจริญสุขศรี

มันเป็นเรื่องนโยบายระดับประเทศจริง ๆ ซึ่งในประเทศไทยยังไม่ชัดเจนเท่าไหร่ แต่ที่ประเทศอังกฤษ จะมี UK Data Government ก็คือจะเป็นเว็บไซต์ที่รวมข้อมูลของหน่วยงานรัฐทุกอย่างเข้าไปอยู่ในนั้น ซึ่งผู้ใช้สามารถเข้าไปดูจากเว็บนั้น แต่สำหรับประเทศไทยจะมีเงื่อนไขของระบบ ความซับซ้อน รวมทั้งเงื่อนไขของแต่ละหน่วยงานที่จะบูรณาการร่วมกัน อันนี้เป็นเรื่องที่จะต้องใช้เวลาพอสมควร ยังไม่ค่อยเห็นภาพรวมของการคุยที่ชัดเจนมากนัก

ตอบโดย คุณชลิดา เอื้อบำรุงจิต

มันเป็นความพยายาม แต่ว่าปัญหาหนึ่งคือคิดว่าส่วนต่าง ๆ เมื่อเอามารวมกันแล้วเอาไปทำอะไร และแต่ละที่ไม่ได้มี website หรือ platform ที่มันเวิร์คแล้วจะไปต่อมากกว่า จริง ๆ มีหน่วยงานที่อยากจะได้ platform ของประเทศอยู่แล้ว เพียงแต่ข้างใต้มันค่อนข้างจะยุ่ง เพราะว่าเมื่อก่อนเราไม่มีสิ่งที่จะรองรับ และการพูดเรื่อง Big data มันเป็นสิ่งที่ไม่เคยทำ data อะไรเตรียมไว้มาก่อน มันเป็นสิ่งที่เป็นไปไม่ได้ แล้วมันทำให้ผู้ที่ทำงานอยู่ทุกคนมีปัญหา

ที่เราทำเรื่อง Big data ไม่ได้ เพราะเราต้องเริ่มจาก Collection Digitization Data Management ทั้งหลาย จะต้องมาจากตรงนั้นก่อน แล้วจึงไปสู่ Big data ได้ เพราะฉะนั้น เราเหมือนไม่ผ่าน 1 2 3 4 แต่คิดว่าในอนาคตเป็นเรื่องที่อยากจะคุย คือวันหนึ่งเราจะต้องคุยเรื่อง link open data ทุกคนทำของตัวเองอยู่ตรงไหนก็ได้ แต่มันเชื่อมโยงกันได้ มันไม่ใช่เรื่องของความเป็นเจ้าของ หรือส่วนตัว สุดท้ายเราแบ่งปันกันทุกคน ซึ่งมันต้องเริ่มจากต้นทางที่เราคุยกัน ถ้าเราไปดูที่ยุโรป link open data มีมากแล้ว ถ้าถนัดทำฐานของตัวเองก็ทำ หรือถ้าเราสบายใจที่จะทำ platform ของใครก็ทำไป

อีกอย่างหนึ่งที่จะตอบอาจารย์ บางครั้งเราให้ชุมชนทำเองทุกอย่างเขาอาจจะตกใจว่าเปิด platform ขึ้นมาแต่ไม่เคยทำอะไร อาจจะมีอีกวิธีหนึ่งคือ อาจจะมีอาจารย์หรือนักวิจัยที่เป็นคนกลาง แล้วเราอาจจะ simplify สิ่งที่จะให้ชุมชนทำ เช่น แค่ช่วยอัดเสียงมาแล้วกรอกข้อมูลเข้าไปในตาราง แล้วเราจะเข้าไป upload หรือ manage อีกทีเพื่อให้อยู่ในระบบ รวมทั้งจัดการ metadata บางอย่างที่เขาไม่รู้ว่าจะต้องกรอกอะไร เราอาจจะทำงานแบบแชร์กันก็ได้

ตอนที่เรายังไม่มีเครื่องมือพร้อม เราก็แบ่งกันทำงาน เราทำงานกันบน google drive ก็แบ่งกันว่าคนนี้ทำ data แค่นี้ อีกคนจะมาเติมต่อ เหมือนเราแบ่งกันเป็นท่อน ๆ ก็จะรู้สึกเบาลง เพื่อให้ฟังก์ชั่นอยู่บน platform ได้

ข้อคิดเห็นจากผู้เข้าร่วมประชุม

ตอนนี้ DGA หรือสำนักงานพัฒนารัฐบาลดิจิทัล (องค์การมหาชน) พยายามเต็มที่ที่จะเอาข้อมูลทั้งหมดมารวมเข้าด้วยกัน แต่นิสัยคนไทยคือ ของฉัน ฉันไม่ให้ นี่คือปัญหา ผมในฐานะที่เป็นมหาวิทยาลัยเล็ก ๆ ที่ขอข้อมูลคนละส่วน คนละหน่วยงานแทบไม่สามารถให้กันได้ ด้วยเหตุผลว่าข้อมูลมีแต่ให้กันไม่ได้ ด้วยเหตุผลหวงข้อมูลเป็นหลัก

ประเด็นที่สอง ข้อมูลเชิงวัฒนธรรม หรือภาษา ยังไม่เห็นทิศทางอะไร เพราะเขามองแค่ข้อมูลที่เป็น Data ที่เขาเคยเก็บกันแล้ว เป็นข้อมูลรายบุคคล ข้อมูลสารพัดที่มีอยู่ แต่ข้อมูลเชิงภาษาและวัฒนธรรมยังไม่เห็น ด้วยเหตุผลก็คือมันไม่เห็นเป็นเม็ดเงิน มันไม่เห็นเป็นอะไรที่ตอบแทนได้โดยทันที ถ้าข้อมูลเดิม ๆ ที่มีอยู่ ข้อมูลประชากร ข้อมูลเศรษฐกิจ เขาจะตีค่าได้ แต่ข้อมูลที่เป็นเรื่องของคุณค่าเชิงวัฒนธรรม การวิ่งนโยบายลักษณะนี้ค่อนข้างจะช้า

ข้อคิดเห็นจากผู้เข้าร่วมประชุม

สวัสดีครับ วีระพงศ์ มีสถาน จากสถาบันวิจัยภาษา ฯ ครับ ต่อไปนี้ไม่ใช่คำถามนะครับ แต่ว่ามีความเป็นห่วง ในฐานะที่เป็นคนทำงานด้านชาติพันธุ์ ก็เป็นห่วงว่าข้อมูลบางอย่าง ด้วยสังคมเปิดมากขึ้น มันจะเป็นช่องทางหนึ่งของคนบางประเภทที่นำเสนอเรื่องราวของตนเอง โดยไม่แยแสต่อความจริงทางวิชาการ

ก่อนหน้านี้กระทรวงศึกษาธิการ เขามีการให้คุณครูท้องถิ่นลุกขึ้นมาเขียนประวัติศาสตร์ท้องถิ่นแล้วก็ไปเจอว่าบางแห่ง ข้อมูลที่เราศึกษามาแบบนี้ แต่พอเข้าไปในชุมชนจริง ๆ ปรากฏว่าบางครอบครัวเขียนให้ตัวเองโดดเด่นในท้องถิ่นนั้น และให้ลูกศิษย์เรียนสิ่งที่คุณครูเขียนขึ้นมาแล้วข้อมูลเหล่านั้นก็จะถูก upload ไปใน youtube ในเว็บไซต์ แล้วมันก็เป็นการสร้างความชอบธรรมใน social ก็กลายเป็นว่าโอกาสแบบนี้ทำให้คนบางประเภทมีช่องทางเกิดขึ้นได้

ประเด็นต่อมาขอเพิ่มเติมว่า เราจะให้มีข้อมูลที่เป็นเอกภาพ หรือเป็น unity ยาก และคงจะไม่ได้ ตนเองทำพจนานุกรมหลายปีมาก มันมีคำศัพท์ที่คนท้องถิ่นหนึ่งเรียกแบบนี้ เราจะไปบีบคอ ให้คุณเรียกแบบเขาไม่ได้ เช่น ต้นข่อย บางภูมิภาคเรียก ต้นซมพ้อ เพราะฉะนั้นความเป็นท้องถิ่นเขาเป็นเช่นนั้น มันก็จะมีความหลากหลาย ผมคิดเช่นเดียวกันว่าในชุมชน ในชาติพันธุ์ท้องถิ่นต่าง ๆ มันก็มีความหลากหลาย มันก็ดีแล้วล่ะ เพียงแต่ว่าเราจะทำอย่างไรให้คนที่ไม่ใส่ใจทางวิชาการนึกสนุก แต่อยากจะเอาข้อมูล upload ขึ้นไป แล้วมันจะมีหน่วยงานอะไรที่จะมาสกรีนว่าอันนี้ไม่น่าจะถูก ขอบคุณมากครับ

ตอบโดย คุณสิทธิศักดิ์ รุ่งเจริญสุขศรี

สำหรับคำถามแรก เรื่องการผูกขาดของข้อมูล ที่ว่าข้อมูลมันเกิดขึ้นได้ง่ายแล้วมีใครคนใดคนหนึ่งสร้างข้อมูล แล้ว upload แล้วก็มีคนเชื่อ ความจริงกรณีแบบนี้ศูนย์มานุษยวิทยาสิรินธรก็ตั้งคำถามเหมือนกัน เช่น เวลาไปชุมชนก็จะมีป้าคนหนึ่งเป็น key maker สำคัญในการให้ข้อมูลกลายเป็นว่าป้าคนนี้เป็นคนรู้ข้อมูลทุกอย่าง แต่ในความเป็นจริงเรามารู้ทีหลังว่า มีความไม่พอใจเล็ก ๆ ของคนอื่น ๆ ในชุมชนว่า ทำไมต้องไปหาป้าคนนั้น ความจริงเรื่องนี้ฉันก็รู้ เรื่องนั้นฉันก็รู้ เราก็บอกตรง ๆ ว่าเราก็หาทางแก้ไม่ได้ แต่สิ่งที่เราพยายามทำก็คือ ทำให้เกิดกระบวนการมีส่วนร่วมมากที่สุด เวลาเราลงพื้นที่ทำกิจกรรมกับชุมชน พยายามจะกระจาย พยายามบอกข่าวให้คนในชุมชนรู้จะได้ไม่เกิดการผูกขาดข้อมูลคนใดคนหนึ่งในชุมชน นี่เป็นวิธีการแก้ปัญหาเบื้องต้นที่เราพยายามจะทำ

เรื่องที่สอง ที่อาจารย์บอกว่าในท้องถิ่นมีความหลากหลายในการเรียกชื่อต่าง ๆ เห็นด้วยอยู่แล้วครับว่า คำพวกนี้ควรจะเก็บไว้ เพราะมันเป็นคำเฉพาะถิ่น เฉพาะกลุ่ม ไม่ควรจะไปเปลี่ยนแปลง แต่ถ้ามองในแง่ของการจัดการข้อมูลเพื่อนำไปสู่การเข้าถึงหรือการไปใช้ในอนาคต คำพวกนี้จะต้องถูกนำมาชำระและมาเรียงความสำคัญกัน เช่น กลุ่มชาติพันธุ์ ชาวเขา ชนพื้นถิ่น ชนพื้นเมือง คำพวกนี้มีความหมายเดียวกัน แต่คนที่เข้ามาใช้เขาไม่สามารถรู้ได้ว่ามีคำกลุ่มชาติพันธุ์ เขาก็ใช้คำว่าชาวเขาในการค้น แต่สิ่งที่จะทำได้ก็คือ นักวิชาการจะต้องมาร่วมกันสร้างชุดคำควบคุม เช่น access สำหรับงานด้านสังคมศาสตร์อย่างไร ซึ่งตอนนี้งานด้านบรรณารักษ์ก็มีของเขามายาวนานแล้ว ซึ่งบางหน่วยงานก็หยิบมาใช้ แต่สำหรับงานวิจัยบางทีก็จะติดข้อปัญหาว่า subject access มันไม่ค่อยตอบโจทย์ของเราเท่าไหร่ ทางศูนย์ ฯ เราก็อยู่ในระหว่างการจะทำขึ้นมาเหมือนกัน ซึ่งจะเป็น keyword ในการเชื่อมโยงข้อมูลจากสถาบันหนึ่งไปสถาบันหนึ่งในการเรียกค้นคำ ๆ นี้ ที่พร้อมด้วยคำลูกที่มีความหมายใกล้เคียง

ตอบโดย คุณชลิดา เอื้อบำรุงจิต

คิดว่าเราคงต้องส่งเสริมเรื่อง Media Literacy ว่าทุกอย่างที่ค้นเจอมันไม่ใช่ว่าเชื่อได้เลย สมมติว่าเราค้นเจอ 10,000 รายการขึ้นมา เราจะเลือกเชื่ออันไหนมากกว่ากันล่ะ เราต้องฝึกตรงนั้นมากกว่าว่าอยู่ดี ๆ ใครสักคนจะมาบอกว่าตัวเองเป็นผู้รู้ คือส่วนตัวรู้สึกว่าตอนนี้เราให้ความเชื่อถือกับบุคคลมากเกินไป โดยที่เราไม่ใช้หลักฐานแวดล้อมอื่น ๆ คือถ้าเราเป็นนักศึกษา นักวิจัย มันต้องมีหลักฐานอย่างอื่น แล้วก็จะทำให้เรากรองข้อมูลได้ ตอนนี้ข้อมูลมีความหลากหลายมาก หรือแม้แต่เราให้ดูหนัง เพราะว่ามันเป็นหนัง หนังมันไม่ใช่ทุกอย่าง จะเถียงทำไมว่าเป็นความจริง หรือไม่เป็นความจริง สิ่งเหล่านี้จะต้องเป็นสิ่งที่ต้องรู้กันทั่วไป แล้วจะแก้ปัญหาเรื่องความเชื่อผิด ๆ ได้

ผศ.ดร.ศิริเพ็ญ อึ้งสิทธิพูนพร กล่าวสรุปปิดกิจกรรมเวทีเสวนา ฯ

กล่าวโดยสรุปจากงานเสวนานี้ จะเห็นได้ว่าเรื่อง concept ก็เป็นแบบเดียวกัน แต่ว่าแต่ละหน่วยงานก็มีภารกิจที่แตกต่างกันไปบ้าง และมีบางส่วนที่ทำงานคล้ายกัน ซึ่งแต่ละหน่วยงานก็จะมีข้อดี ข้อจำกัดต่างกัน อย่างวันนี้ทำให้เราได้รับรู้ว่าหน่วยงานทั้งหลาย เราพยายามให้ความสำคัญกับเจ้าของข้อมูล และให้เจ้าของข้อมูลสามารถที่จะจัดการข้อมูลของเขาได้ อย่าง NECTEC ก็มีทั้งพื้นที่ให้จัดเก็บข้อมูล และมีคนที่ไปให้ความรู้และเข้าถึงชุมชนจริง ๆ รวมทั้งศูนย์มานุษยวิทยาสิรินธรก็มีนักวิชาการเข้าไป หรือว่าสถาบันวิจัยภาษาและวัฒนธรรมเอเชีย เราก็ไม่ได้ทำแค่นักวิชาการเราให้ความสำคัญกับชุมชน แต่ก็มีข้อจำกัดว่าทางชุมชนก็ไม่สามารถจะทำเองได้ทุกอย่าง แต่ถ้าสอนถ่ายรูป หรือจัดเก็บข้อมูลด้านวัตถุในเชิงพิพิธภัณฑ์ ยังถือว่าง่าย แต่หากเป็นด้านภาษาและวัฒนธรรมที่ต้องมีการวิเคราะห์ ก็ยังต้องทำงานร่วมกัน รวมถึงกระบวนการทำงานของหอภาพยนตร์ ก็ยังมีความสำคัญ

คิดว่าวันนี้คงจะได้องค์ความรู้หลายอย่าง และท่านผู้ฟังทั้งหลายก็จะได้ทราบแหล่งข้อมูลต่าง ๆ อย่างไรก็ตามข้อมูลจะเชื่อถือได้หรือไม่ ก็ต้องไปสืบค้น และขึ้นอยู่กับผู้ที่ใช้ข้อมูล ไม่ใช่ว่าทุกอย่างจะเชื่อได้หมด ก็ขอให้เลือกใช้ข้อมูลอย่างพินิจพิจารณานะคะ

สรุปประเด็นจากการเสวนาแลกเปลี่ยนความรู้เรื่อง *"การจัดทำและนำเสนอคลังข้อมูลดิจิทัลออนไลน์"*

ประเด็นสำคัญของแต่ละหน่วยงานที่ทำงานด้านคลังข้อมูล ไม่ว่าจะเป็น หอภาพยนตร์ที่มุ่งเน้นเฉพาะเรื่องภาพยนตร์ ซึ่งมีความชัดเจนในตัวเอง ส่วนอีกสามหน่วยงานคือ ศูนย์มานุษยวิทยาสิรินธร ศูนย์เทคโนโลยีอีเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ (NECTEC) และสถาบันวิจัยภาษาและวัฒนธรรมเอเชีย มหาวิทยาลัยมหิดล ล้วนมีเป้าหมายที่คล้ายคลึงกันคือ เก็บรวบรวมข้อมูลด้านภาษาและวัฒนธรรมของชุมชนหรือความรู้ท้องถิ่น เพื่อการเผยแพร่ หรือเก็บไว้โดยไม่เผยแพร่ก็ตาม ต่างมุ่งหวังให้เกิดประโยชน์กับเจ้าของข้อมูลเป็นหลัก โดยอาจมีวิธีการที่แตกต่างกัน เช่น

นักวิชาการเก็บรวบรวมข้อมูลร่วมกับเจ้าของวัฒนธรรม หรือถ่ายทอดวิธีการเก็บข้อมูลแล้วให้คนในชุมชนเป็นคนทำเองทั้งหมด ก็ขึ้นอยู่กับบริบทที่สามารถทำได้อย่างเหมาะสม

ด้วยเทคโนโลยีที่ทันสมัยในปัจจุบัน การบริหารจัดการข้อมูลให้อยู่ในรูปแบบดิจิทัลเป็นสิ่งที่หลีกเลี่ยงไม่ได้ ทั้งยังมีประโยชน์ในแง่การเก็บรักษาให้ยาวนาน และสะดวกต่อการเผยแพร่ในวงกว้างถ้าหากพื้นที่นั้น ๆ มีอินเทอร์เน็ตเข้าถึง สิ่งที่คาดหวังในอนาคตคือ ต้องการให้มีหน่วยงานภาครัฐที่เห็นความสำคัญของการสร้างคลังข้อมูลขนาดใหญ่ (Big Data) โดยเฉพาะด้านสังคมศาสตร์ มนุษยศาสตร์ อย่างเชื่อมโยงเป็นระบบ ไม่ว่าหน่วยงานไหนจะเก็บบันทึกข้อมูลในหัวข้อที่คล้าย ๆ กัน แต่ทีมวิจัยเชื่อว่าเนื้อหารายละเอียดคงไม่เหมือนกัน ซึ่งจะเป็นการเติมเต็มให้ข้อมูลนั้นมีความสมบูรณ์ สามารถนำไปใช้ประโยชน์ต่อวงวิชาการและต่อคนในชุมชนเอง โดยอาจจะหยิบเนื้อหาเฉพาะส่วนที่ตนเองสนใจก็นับว่าเป็นประโยชน์มากมาย

สุดท้ายนี้ ขอฝากว่าผู้ค้นหาข้อมูลเพื่อนำไปใช้ประโยชน์ ไม่ว่าในแง่ใดก็ตาม ควรพิจารณาว่าเนื้อหาเชื่อถือได้มากน้อยเพียงใด การค้นหาข้อมูลจากหลายแหล่งสามารถช่วยตรวจสอบว่าข้อมูลที่ถูกต้องควรเป็นเช่นไร และสิ่งสำคัญที่ไม่ควรลืม คือการอ้างอิงแหล่งที่มาของข้อมูล เพราะหากเกิดความผิดพลาด ผู้อ่านก็ยังสามารถไปสืบค้นย้อนหลังได้ ทั้งยังแสดงถึงความเคารพสิทธิ์ของผู้ผลิตอีกด้วย

เอกสารอ้างอิง

Austin, Peter K. 2011. *Who uses digital language archives?* http://www.paradisec.org.au/blog/2011/04/who-uses-digital-language-archives/.

Ferreira, Vera. 2019. *PPT for Language Documentation and Archiving workshop* at Mahidol University.

Himmelmann, Nikolaus. 1998. Documentary and descriptive linguistics. *Linguistics* 36. 161–95.

Nordhoff, Sebastian & Harald Hammarström. 2014. *Archiving grammatical descriptions.* In David Nathan

Peter K. Austin (eds.), *Language Documentation and Description, Volume 12: Special Issue on Language Documentation and Archiving*, 164–186. London: SOAS.

Ryan E. Henke and Andrea L. Berez-Kroeker. 2016. A brief history of Archiving in Language Documentation, with an annotated bibliography. *Language Documentation & Conservation*. Vol.10 pp.411-457. https://scholarspace.manoa.hawaii.edu/bitstream/10125/24714/1/henke_berez.pdf

Trilsbeek, Paul & Peter Wittenburg. 2006. Archiving challenges. In Jost Gippert, Nikolaus P. Himmelmann & Ulrike Mosel (eds.), E*ssentials of language documentation* (Trends in Linguistics Studies and Monographs 178), 311–336. Berlin: Mouton de Gruyter.

Suwilai Premsrirat. 2007a. *Endangered Languages of Thailand. Interna¬tional Journal of Sociology of Language*. 186, 75–93.

Woodbury, Tony. 2003. Defining documentary linguistics. In Peter Austin (ed.), *Language Documentation and Description Volume* 1, 35–51. London: SOAS.

Woodbury, Anthony. 2011. Language documentation. In Peter K. Austin & Julia Sallabank (eds.), *The Cambridge handbook of endangered languages,* 159–211. Cambridge: Cambridge University Press.

สุวิไล เปรมศรีรัตน์ และคณะ. 2547. *แผนที่ภาษาของกลุ่มชาติพันธุ์ต่างๆในประเทศไทย (Ethnolinguistic Maps of Thailand)*. สำนักงานคณะกรรมการวัฒนธรรมแห่งชาติ. กรุงเทพฯ: โรงพิมพ์คุรุสภาลาดพร้าว

ภาคผนวก

• แบบฟอร์มสัญญาการฝากข้อมูล

หน้าที่ 1

คลังข้อมูลดิจิทัลของกลุ่มชาติพันธุ์ในประเทศไทย (LangArchive-TH)
สถาบันวิจัยภาษาและวัฒนธรรมเอเชีย มหาวิทยาลัยมหิดล

Deposit key:

วันที่:
สำหรับ LangArchive-TH เท่านั้น

สัญญาอนุญาตให้ใช้สิทธิ

ก. ข้อมูลของผู้ฝาก

รายละเอียดของผู้ฝากข้อมูล

1. ชื่อ

คำนำหน้า.....

ชื่อ.....

นามสกุล.....

2. สัญชาติ .....

3. วันเกิด .....

4. ที่อยู่ .....

.....

จังหวัด .....

รหัสไปรษณีย์.....

ประเทศ.....

5. อีเมล .....

6. เบอร์โทรศัพท์.....

หน้าที่ 2

ข้อมูลพื้นฐานของตัวแทนผู้ฝากข้อมูล

7. ชื่อ

คำนำหน้า.....

ชื่อ.....

นามสกุล.....

8. ที่อยู่ .....

.....

จังหวัด .....

รหัสไปรษณีย์.....

ประเทศ.....

9. อีเมล .....

10. เบอร์โทรศัพท์.....

11. ชุมชนหรือองค์กรที่เกี่ยวข้อง.....

รายละเอียดการฝากข้อมูล

12. Deposit Key:.....

13. Deposit Title: .....

หน้าที่ 3

ข. ข้อตกลง

สิ่งที่คลังข้อมูลดิจิทัลของกลุ่มชาติพันธุ์ได้รับอนุญาต

คลังข้อมูลดิจิทัลของกลุ่มชาติพันธุ์ในประเทศไทย (LangArchive-TH) อยู่ภายใต้โครงการคลังข้อมูลดิจิทัลของกลุ่มชาติพันธุ์: ส่งเสริมการปกป้อง และดำรงรักษา ความหลากหลายทางภาษาและวัฒนธรรมในประเทศไทย

คลังข้อมูลดิจิทัลของกลุ่มชาติพันธุ์ในประเทศไทย สนับสนุนการอนุรักษ์ความหลากหลายทางภาษาและภูมิปัญญาทางวัฒนธรรมของกลุ่มชาติพันธุ์ และมุ่งเน้นพัฒนาคลังข้อมูลที่ชุมชนในท้องถิ่นสามารถเข้าถึงได้ เราจะปกป้องผลประโยชน์ของผู้ฝากข้อมูลและกำหนดสิทธิการเข้าถึงข้อมูลตามที่ท่านกำหนด ทางคลังข้อมูลฯ จะรวบรวมและจัดเก็บข้อมูลส่วนตัวตามกฎหมายการคุ้มครองข้อมูลส่วนบุคคล (Data Privacy) ระบบข้อมูลอาจมีการเปลี่ยนแปลงถ้ามีการปรับเปลี่ยนหรือออกกฎหมายใหม่

การอนุญาต

คลังข้อมูลดิจิทัลของกลุ่มชาติพันธุ์ในประเทศไทย (LangArchive-TH) ไม่ได้เป็นเจ้าของลิขสิทธิ์ในข้อมูลซึ่งได้รับการนำฝากโดยเจ้าของข้อมูล สัญญาฉบับนี้กำหนดให้ LangArchive-TH ได้รับอนุญาตดังต่อไปนี้

- จัดการข้อมูลเพื่ออนุรักษ์
- เผยแพร่รายละเอียดของข้อมูลสู่สาธารณะ
- จัดการสิทธิการเข้าถึงหรือการคัดลอกข้อมูลตามที่ผู้ฝากกำหนด
- ดำเนินการบำรุงรักษา รายงาน และการเผยแพร่ข้อมูลสู่สาธารณะตามข้อกำหนดจากผู้ฝากข้อมูล

ผู้ใช้งานคลังทุกคนมีสิทธิเข้าถึงข้อมูลได้เท่าที่ผู้ฝากข้อมูลอนุญาตเท่านั้น

หน้าที่ 4

คำแถลงของผู้ฝากข้อมูล

ข้าพเจ้า ผู้ฝากข้อมูลกับคลังข้อมูลดิจิทัลของกลุ่มชาติพันธุ์ในประเทศไทย (LangArchive-TH) ยืนยันว่า

- ข้าพเจ้าเป็นเจ้าของข้อมูลหรือเป็นผู้มีสิทธิในการฝากข้อมูล ซึ่งรับหน้าที่ทำการตกลงอย่างเป็นทางการกับคลังข้อมูลฯ
- ข้าพเจ้าขอให้ LangArchive-TH อนุรักษ์และบำรุงรักษาข้อมูล โดยเข้าใจว่าการอนุรักษ์นี้อาจเกี่ยวข้องกับการคัดลอกและแปลงข้อมูลที่มีลิขสิทธิ์ไปสู่นามสกุลไฟล์ใหม่ในอนาคต และใช้พื้นที่การเก็บข้อมูลร่วมกับผู้ใช้งานคลังอื่น ๆ
- ข้าพเจ้าให้สิทธิ LangArchive-TH ในการจัดแจงสิทธิการเข้าถึงข้อมูลให้สอดคล้องกับเงื่อนไขและข้อตกลงที่ข้าพเจ้าได้ให้รายละเอียดไว้พร้อมกับข้อมูล
- นอกเหนือจากเงื่อนไขและข้อตกลงที่ข้าพเจ้าได้ให้รายละเอียดไว้ใน Metadata และกฎหมายที่เกี่ยวข้อง ข้าพเจ้าอนุญาตให้ผู้อื่นเข้าถึงข้อมูลเพื่อการวิจัย และจุดประสงค์อื่นซึ่งไม่เกี่ยวข้องกับเชิงพาณิชย์ ข้าพเจ้ารับทราบว่าการเผยแพร่ แสดง หรือแจกจ่ายข้อมูลบน World Wide Web หรือเทคโนโลยีอื่นในอนาคตนั้นได้รับการอนุญาตจากข้าพเจ้าอย่างเป็นลายลักษณ์อักษร ผู้ใช้ข้อมูลไม่ว่าส่วนใด และเผยแพร่หรือไม่ก็ตาม ต้องได้รับการยอมรับจากข้าพเจ้าและ LangArchive-TH
- ข้าพเจ้าจะปรับข้อมูลการติดต่อให้เป็นปัจจุบัน หากไม่สามารถติดต่อตัวแทนของข้าพเจ้าได้ ทาง LangArchive-TH สามารถดำเนินการจัดการและเข้าใช้ข้อมูล โดยคำนึงจากรายละเอียดที่ข้าพเจ้าได้ระบุไว้ใน Metadata และผลประโยชน์ต่อชุมชน
- ข้าพเจ้าจะรับผิดชอบในส่วนค่าใช้จ่าย หรือหนี้สิน อันเกิดจากการกระทำของข้าพเจ้าที่ทำให้เกิดการสูญหายหรือเสียหายของข้อมูล หรือนำข้อมูลไปใช้ในทางที่ผิด ไม่ว่าจะเป็นทางตรงหรือทางอ้อม
- LangArchive-TH ได้รับอนุญาตให้สร้างเวอร์ชั่นดัดแปลงของข้อมูลให้เหมาะสมในเพื่อการเผยแพร่โดยการเข้าถึงทาง website และ search software ในการจัดทำคลังข้อมูล มหาวิทยาลัยมหิดลได้รับการอนุญาตให้คัดลอกและแปลงข้อมูลทั้งหมดเพื่อการอนุรักษ์ โดยข้าพเจ้าอนุญาตให้มอบสำเนาข้อมูลแก่บุคคลที่สาม ได้แก่ มหาวิทยาลัยมหิดล เพื่อรับรองการอนุรักษ์ รวมทั้งแปลงข้อมูลจากนามสกุลไฟล์หนึ่งไปเป็นอีกสกุลหนึ่งเพื่อรับรองการเข้าถึงข้อมูลอย่างต่อเนื่องต่อไป
- LangArchive-TH ได้รับอนุญาตให้เผยแพร่ข้อมูลบนพื้นที่ออนไลน์เพื่อการวิจัยและการศึกษาจากทั่วโลก ภายใต้สัญญาอนุญาตครีเอทีฟคอมมอนส์ (Creative Commons Attribution/Non-

หน้าที่ 5

Commercial/NoDerivatives (CC BY-NC-ND 4.0) international license) และพระราชบัญญัติลิขสิทธิ์ พ.ศ. 2537

ลงนาม

ผู้ฝากข้อมูล		*LangArchive-TH*	
ชื่อตัวบรรจง	.....	ชื่อตัวบรรจง	.....
ลงนาม	.....	ลงนาม	.....
วันที่	.....	วันที่	.....

ติดต่อ *LangArchive-TH*

ที่อยู่

สถาบันวิจัยภาษาและวัฒนธรรมเอเชีย มหาวิทยาลัยมหิดล
999 พุทธมณฑล สาย 4 ตำบลศาลายา อำเภอพุทธมณฑล
จังหวัดนครปฐม 73170

อีเมล

langarchive.th@gmail.com

เว็บไซต์

www.langarchive-th.org

เบอร์โทรศัพท์

02 800 2319

• **แบบฟอร์มการบันทึกรายละเอียดข้อมูล (Metadata)**

หน้าที่ 1

MUKURTU METADATA

TITLE (ชื่อหัวข้อสั้น ๆ สำหรับ digital heritage item):.....

.....

COMMUNITY (กลุ่มผู้ใช้งานคลังข้อมูลที่ได้รับอนุญาตให้เข้าถึงข้อมูล):.....

.....

TYPE (ประเภทของข้อมูล เช่น วีดีโอ รูปภาพ เสียง):.....

ORIGINAL DATE (วันที่เก็บข้อมูล):.....

SUMMARY (สรุปข้อมูลอย่างสั้น):.....

.....

.....

CATEGORY (หัวข้ออย่างกว้างของข้อมูล เช่น เพลง กลอน การละเล่น):.....

.....

CONTRIBUTOR (ผู้ให้ข้อมูล):.....

.....

.....

LOCATION (สถานที่):.....

.....

CREATOR (ผู้เกี่ยวข้องกับการเก็บข้อมูล):.....

.....

.....

DESCRIPTION (รายละเอียด):.....

.....

.....

.....

.....

.....

หน้าที่ 2

FORMAT (นามสกุลไฟล์):

LANGUAGE (ภาษาของข้อมูล และภาษาที่ใช้ในการทำงาน):

LICENSING: Non-Commercial

MEDIA ASSETS (ชื่อไฟล์โดยไม่มีนามสกุลไฟล์):

PUBLISHER (ผู้เผยแพร่ข้อมูลบนคลัง):

KEYWORDS (คำสำคัญ):

PROTOCAL (สิทธิการเข้าถึงข้อมูล เช่น เปิด ปิด):

USERNAME (ผู้อัปโหลดข้อมูลขึ้นคลัง):

• ประมวลภาพการจัดอบรมบุคลากรทางการวิจัย / การศึกษา

การอบรมเชิงปฏิบัติการเรื่อง การบันทึกข้อมูลและการทำคลังข้อมูลดิจิทัล 3 ภาค ได้แก่ ภาคใต้ (มอ.ปัตตานี) ภาคตะวันออก (วิทยาลัยชุมชนตราด) และภาคเหนือ (มหาวิทยาลัยราชภัฏเชียงราย)

ภาพที่ 35 - 39 : การจัดอบรมเชิงปฏิบัติการที่ภาคใต้

ภาพที่ 40 - 45 : การจัดอบรมเชิงปฏิบัติการที่ภาคตะวันออก

ภาพที่ 46 - 50 : การจัดอบรมเชิงปฏิบัติการที่ภาคเหนือ

• **ประมวลภาพความรู้จากการลงภาคสนาม**

ภาพที่ 51 - 57 : การบันทึกข้อมูลภาคสนาม กลุ่มไทแสก

ภาพที่ 58 - 62 : การบันทึกข้อมูลภาคสนาม กลุ่มชอุ้ง

• **ประมวลภาพการจัดงานเสวนา**

ภาพที่ 63 - 67 : บรรยากาศงานวันเสวนาแลกเปลี่ยนเรียนรู้

• ประมวลภาพการจัดงานประชุมวิชาการนานาชาติ

ภาพที่ 68-77 : บรรยากาศวันประชุมวิชาการนานาชาติ

ดัชนีคำ

ก

กล้อง 7, 57
กล้องมือถือ 28, 57
กล้องวิดีโอ 38
กลุ่มชาติพันธุ์ 1, 27
ก๋อง 31, 37
การตั้งชื่อ 10
การตั้งชื่อไฟล์ 11
การละเล่น 12, 26
เก็บข้อมูล 7, 12, 35, 54, 57

ข

ข้อมูล
ข้อมูลคัดลอก 13
ข้อมูลจริง 13
ข้อมูลเดิม 6
ข้อมูลปฐมภูมิ 4
ข้อมูลปิด 5
ข้อมูลวัตถุ 56
ขาตั้งกล้อง 8, 28
เข้าถึงข้อมูล 3, 12, 23, 29, 31, 37

ค

คลังข้อมูล 1, 3, 12, 27, 32, 36, 44, 54
คลังข้อมูลแบบมีส่วนร่วม 4
คลังข้อมูลดิจิทัล 1, 2, 13, 18, 32
คลังข้อมูลชุมชน 44, 52
ความหลากหลาย 1, 33
คัดลอก 13
คำค้นหา 21, 23
คำถาม 28, 38
คำสำคัญ 13, 21, 26, 31
เครือข่าย 32
เครื่องมือ 8, 38
เครื่องหมายยัติภังค์ 10

จ

จดหมายเหตุ 48, 56
จริยธรรม 4
จัดการข้อมูล 5, 27, 54
จัดระบบ 9, 13, 29
จับต้องได้ 51, 58
จับต้องไม่ได้ 51, 58
จำกัดสิทธิ 31
เจ๊ะเห 38
เจ้าของข้อมูล 5, 29, 30, 53, 63
เจ้าของภาษา 3, 5, 6, 12, 34

ช

ชนพื้นเมือง 3
ชอุ้ง 10, 35
ชื่อไฟล์ 13, 29
ชื่อหัวข้อ 12, 22

ซ

ซีดี 40

ฐ

ฐานข้อมูล 45

ด

ดิจิทัล 34, 36, 40
ข้อมูลดิจิทัล 2, 54, 56
ไฟล์ดิจิทัล 2, 35, 44
ดีวีดี 40
ดาวน์โหลด 14, 28
เดินทุ่ง 57
โดเมน 37, 54

ต

ตระกูลภาษา 1
ตระกูลภาษาจีน-ทิเบต 1
ตระกูลภาษาไต-กะได 1, 35
ตระกูลภาษาม้ง-เมี่ยน 1
ตระกูลภาษาออสโตรนีเชียน 1
ตระกูลภาษาออสโตรเอเชียติก 1
ตัวพิมพ์ใหญ่ 30, 31

ถ

ถ่ายถอดข้อมูล 3
ถ่ายถอดเสียง 28

ท

ท่องเที่ยว 54, 57
ท้องถิ่น 1, 35, 46
ครูท้องถิ่น 62
ประวัติศาสตร์ท้องถิ่น 62
ภาษาท้องถิ่น 1
ทิศทางของแสง 7
เทคโนโลยีสารสนเทศ 5
เทปบันทึกเสียง 3
ไทดำ 48
ไทลื้อ 48, 58
ไทแสก 35
โทรศัพท์มือถือ 28, 37

น

นราธิวาส 38
นวนุรักษ์ 54, 56
นวัต 54
นักวิชาการ 3, 38
นามสกุล 11, 12, 30
นิทรรศการ 56
แนวทางปฏิบัติที่ดี 4

บ

บัตรทะเบียนฟิล์ม 41
บันทึกข้อมูล 3, 6, 8, 27, 38
บันทึกภาพ 7
เครื่องบันทึกภาพ 9
บันทึกเสียง 4, 28, 35, 38, 52
เครื่องบันทึกเสียง 8, 34, 38
บีซู 29
บูรณาการ 54, 60
แบบฟอร์ม 27, 29

ป

ประมวลผล 32
ประวัติศาสตร์ชุมชน 52
ประเทศไทย 32
ประเภทของสื่อ 12, 21

ประเภทของไฟล์ 11, 30
แปลงไฟล์ 13, 27, 29, 34
แปลงข้อมูล 56
โปรแกรม 12, 14, 29, 43

ผ

ผู้จัดการข้อมูล 13
ผู้ดูแลคลังข้อมูล 5, 6, 31, 38
ผู้บอกภาษา 7, 30
ผู้บันทึก 6, 28
ผู้เผยแพร่ 13, 26
ผู้ฝากข้อมูล 9, 10, 12, 26, 29, 30
ผู้มีส่วนร่วม 23, 26
ผู้สัมภาษณ์ 7
ผู้อำนวยความสะดวก 7
ผู้เกี่ยวข้อง 12
ผู้ใช้ 5
ผู้ให้ข้อมูล 7, 10, 12, 26, 30, 31
แผ่นดิสก์ 34
แผนที่ภาษา 33

ฝ

ฝากข้อมูล 5, 10, 26, 28
เฝ้าระวัง 59

พ

พจนศิลป์ 26
พจนานุกรม 30, 62
พิพิธภัณฑ์ 3, 45, 49, 52, 54, 56
แพลตฟอร์ม 1, 54

ฟ

ฟิล์มภาพยนตร์ 39
ฟื้นฟูภาษา 1, 32
ไฟล์
ไฟล์ข้อมูล 9, 13
ไฟล์เดิม 6, 13
ไฟล์ภาพ 28
ไฟล์วิดีโอ 28
ไฟล์เสียง 12, 26, 28
โฟลเดอร์ 13

ภ

ภาคผนวก 27
ภาคสนาม 3, 27, 31, 35, 48, 51, 53
ภาพเคลื่อนไหว 9, 28, 43
ภาพยนตร์ 39
ภาพวัตถุ 57
ภาพสามมิติ 56
ภาวะวิกฤต 1, 3, 28, 32
ภาษาท้องถิ่น 1
ภาษาศาสตร์ 34
นักภาษาศาสตร์ 1, 3, 6, 38
ภาษาศาสตร์เชิงปฏิบัติ 3
ภูมิปัญญา 1, 12, 33, 51

ม

มนุษยศาสตร์ 50, 53
มลายูถิ่น 36
มรดกทางวัฒนธรรม 44
มหาวิทยาลัยมหิดล 1, 32
มหาวิทยาลัยราชภัฏเชียงราย 58

มหาวิทยาลัยราชภัฏเชียงใหม่ 57
มหาวิทยาลัยราชภัฏยะลา 35, 38
มหาวิทยาลัยหาดใหญ่ 35
มอ.ปัตตานี 35
มานิ 51
แม่ไก่ 58
ไมเคิล มอร์แมน 48
ไมโครโฟน 8, 28

ย

ยินยอม 7, 31, 35, 38

ร

รหัสภาษา 11, 29
ระบบเขียน 1, 33
รายละเอียดของข้อมูล 5, 6, 27, 29, 51
รายการข้อมูล 10, 12
รูปภาพ 8, 12, 13
รูปแสกน 13

ล

ลงทะเบียน 5, 29
ลอนดอน 34
ลิขสิทธิ์ 7, 31, 44
ลำดับไฟล์ 11
โลกาภิวัตน์ 1

ว

วัตถุทางวัฒนธรรม 51
วิทยาลัยชุมชนตราด 35
วิเคราะห์ 28, 56
วิดีโอ 7, 8, 10, 12, 13, 26, 28, 29, 57
เว็บไซต์ 1, 18, 28, 35, 37, 49, 60, 62

ศ

ศูนย์การเรียนรู้ 54, 56, 58
ศูนย์เทคโนโลยีอีเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ 32, 54
ศูนย์มานุษยวิทยาสิรินธร 32, 38, 44, 58, 62
ศูนย์ศึกษาและฟื้นฟูภาษาและวัฒนธรรมในภาวะวิกฤต 1

ส

สกว. 32
สกสว. 32
สถานที่ 12, 26, 30
สถาบันภาพยนตร์อังกฤษ 43
สถาบันวิจัยภาษาและวัฒนธรรมเอเชีย 1, 32
สมาชิก 5, 29, 31, 38
สรุป 22
ส่วนร่วม 7
สหพันธ์ภาพยนตร์นานาชาติ 40
สังคมศาสตร์ 50, 53, 54, 62
สัมภาษณ์ 7
สารสนเทศ 44
สิทธิ์ 5, 7, 12, 13, 23, 26
สื่อ 22
สื่อการเรียนการสอน 1, 36
สื่อการสอน 12
สื่อทำมือ 2, 36
สำนักงานกองทุนสนับสนุนการวิจัย 32
สำนักงานพัฒนารัฐบาลดิจิทัล 61

สำรองข้อมูล 6, 13
เสวนา 32, 36, 63

ห

หมวดหมู่ 26
หมายเลขผู้ฝากข้อมูล 10
หอภาพยนตร์ 32, 39

อ

องค์การมหาชน 32, 39, 44, 61
องค์ความรู้ 1, 6, 28, 54
อนุญาต 29, 35, 38
อนุรักษ์ 32, 44, 50, 54, 59
อบรม 35
ออนไลน์ 3, 6, 13, 29, 35
ออสโตรนีเซียน ดูที่ ตระกูลภาษา
ออสโตรเอเชียติก ดูที่ ตระกูลภาษา
อะนาล็อก 1, 3
อักขระพิเศษ 10
อักษรไทย 1
อังกฤษ 32, 34, 60
อัตลักษณ์ 1
อินเตอร์เน็ต 38, 57
อาหารพื้นบ้าน 21, 26
อุปกรณ์ 8, 28
เอกภาพ 62
เอกสาร 2, 13, 26

ฮ

ฮากกา 21, 26, 37

A

Access 5, 62
Administrator 29
American linguistics 3
Analogue 3, 41
Analysis 56
Analytics 55
Annotation 13
Anthropological 48
API 55
Application 55, 56, 58
AR 58
Archive 2, 31, 54
Archive user 5, 29
Archiving 2, 37
Archiving system 5
Film archive 40
Participatory archive 4
Archivist 31
Digital archivist 5
Attribution 12, 26
Audio 13
Audio recorder 9
Austin 3

B

Back up 6, 13
Bank 55
Base name 11
Best practices 4
Big data 61, 64
BFI 43
British Council 32
Bundle 7, 29

C

Capturing 51
Catalogue 40
Category 12, 26, 30
CDWA 45, 50
Chung 9
Closed 5
Community drive 53
CMS 5, 6, 14, 39, 53
Collection 27, 56, 61
Collection policy 44
Colon 30
Command prompt 14
Communicative event 11
Community 12, 25, 29, 55
Consent form 31
Content 31, 53
Contributor 11, 12, 24, 26, 30, 31
Creator 12, 26, 30
Culture 37
Cultural bank 55, 58, 60
Cultural heritage 55
Cultural narrative 30
Cultural watch 55, 59
Curation 17, 49
Curating 51

D

Data content 50
Data cleaning 42
Data integration 56
Data management 9, 56, 61
Data model 43
Data structure 50
Database 41, 48
Deposit key 10
Depositor 5, 9, 11, 31
Description 12, 26, 30, 53
Descriptive linguistics 3
Developer 55
DGA 61
Dictionary 30
Digital 2, 41, 49
Digital heritage 11, 12, 15, 27, 29
Digital transformation 43
Digitized 42
Digitization 56, 61
Discoverability 5
Documentation 2, 6
Language documentation 3, 38
Download 31
Dublin core 45, 50

E

EAC-CPF 45
EAF 13
Edward Sapir 3
ELAR 32
Endangered 32
Equipment 8
Ethnologue 11, 29

F

Facilitator 7
Federated search 49
FIAF 40, 42
FFMPEG 17, 29
File name 11
Filing system 9
Fixity 6
Folder name 10
Fondue 55, 59
Format 12, 26, 31, 43
Franz Boas 3

G

Genre 19
Google drive 61
Google map 26, 30
Google search 49
GUID 29, 31

H

Hackathon 57
Hammarstörm 3
Heritage 9, 55
Himmelmann 3
Hologram 58
Hypercardioid 8

I

Identifier 31
Image 30
Intangible 51
Integrity check 6
Interview 7
Standard interview 7
Semi-directed interview 7
Interview with stimuli 7
ISAD (G) 45, 48, 50
IT 5, 6, 54
Items 10, 27

J

JPG 11
JPEG 13
Journal 3

K

Key maker 62
Keyword 13, 26, 31, 62

L

Langarchive-TH 10, 12, 14, 27, 37
Language 12, 26, 31, 37
Latitude 26, 30
Legacy 31
Licensing 12, 26, 31
Link data 60
Life cycle 49
Local community 55
Location 12, 26, 30
Longitude 26, 30

M

Management system 54
Master 44
Materials 5
Material culture 50
Media assets 13, 22, 25, 31
Media literacy 63
Media type 11
Metadata 5, 6, 7, 9, 12, 27, 29, 42, 45, 61
Additional metadata 21, 26
Administrative metadata 47
Descriptive metadata 47
Standard metadata 49
Microsoft excel 12
Mobile 57
MOV 18
MP3 11, 13, 26
MP4 11, 13, 26
Mukurtu 2, 6, 12, 14, 31, 39, 53
Mukurtu essentials 21, 22
Mukurtu core 21, 26

N

Navanurak 55, 57
NECTEC 32, 54, 58, 60
Network 55

Newton Institutional Links 32
Node 12
Non-commercial 12, 26, 31
Nordhoff 3
Normalization 6

O

Onedrive 53
Open data 61
Original date 12, 30

P

Participant observation 7
Non-participant observation 7
Participatory archive *ดูที่* Archive
PDF 11, 13, 26
PDF-A 13
People 23, 26, 30
Permission 21, 26
Platform 52, 54, 55, 60
Playground 55
Preservation 44
Preservation layer 6
Privacy 19
Protocol 13, 23, 25, 31
Publisher 13, 26, 31

Q

QR Code 57

R

Rammana song 29
Redcross 54
Register
Registered 29
Registering 5
Relations 21, 27
Request 5
Rights 21, 26, 31
Role 5

S

SAC qualifier 45
Search engine 54
Semi-colon 21, 30
SOAS 32
Social network 55, 59
Software 55, 57
SoundCloud 18, 23
Source 31
Streaming 13
Summary 12, 22, 30

T

Tags 23
Tangible 51
Technology system 54
Text 30, 47
Thailand 37
The British Film Institute 43
Thumbnail 24
Title 12, 14, 23, 29

Tourism 54
Traditional knowledge licenses 53
Transcription 30
Trilsbeek 3
TXT 13
Type 12, 26, 30

U

Underscore 10
Unicode 10, 13
University of London 32
Update 6
Upload 13, 27, 28, 38, 62
User 5
Username 13, 31

V

Vera Ferreira 8
Version control 6
Video 13, 30
Video camera 9
Vimeo 18, 24, 31
Visualization 56

W

Watch 55, 59
WAV 11, 13
Web application 56
Website 60
Windows 14
Wittenburg 3
Woodbury 3
Work flow 49, 53

Y

Youtube 42, 44, 62

ทีมวิจัยจากสถาบันวิจัยภาษาและวัฒนธรรมเอเชีย มหาวิทยาลัยมหิดล

1. ผู้ช่วยศาสจราจารย์ ดร.ศิริเพ็ญ อึ้งสิทธิพูนพร หัวหน้าโครงการ
2. นางสาวกุมารี ลาภอาภรณ์ นักวิจัย
3. นางสาวณัฐมน โรจนกุล นักวิจัย
4. นางสาวบัวฉัฐ วัดแย้ม นักวิจัย

ทีมวิจัยจาก Endangered Languages Archive (ELAR), SOAS University of London

1. Dr Mandana Seyfeddinipur
2. Vera Ferreira

องค์กร ภาคี เครือข่าย

ELAR, SOAS, University of London
University of Hawai‘i at Mānoa
หอภาพยนตร์ (องค์การมหาชน)
ศูนย์มานุษยวิทยาสิรินธร (องค์การมหาชน)
ศูนย์เทคโนโลยีอีเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ (NECTEC)
วิทยาลัยชุมชนตราด
มหาวิทยาลัยราชภัฏยะลา
มหาวิทยาลัยสงขลานครินทร์ วิทยาเขตปัตตานี
มหาวิทยาลัยราชภัฏเชียงราย
มหาวิทยาลัยหาดใหญ่
โรงเรียนคณะราษฎรบำรุง จังหวัดยะลา