การทำคลังความรู้ภาษาและวัฒนธรรมของกลุ่มชาติพันธุ์ในประเทศไทย

3 ธันวาคม 2562

ผศ.ดร.ศิริเพ็ญ อึ้งสิทธิพูนพร

สถาบันวิจัยภาษาและวัฒนธรรมเอเชีย

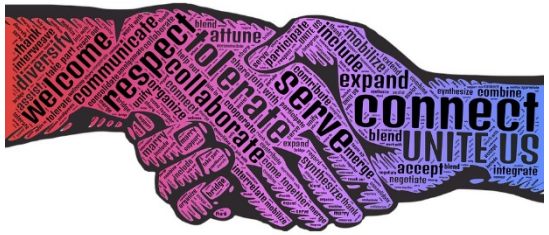
มหาวิทยาลัยมหิดล

เป็นส่วนหนึ่งของโครงการวิจัยเรื่อง

โครงการคลังข้อมูลดิจิทัลของกลุ่มชาติพันธุ์: ส่งเสริมการปกป้อง และดำรงรักษา ความหลากหลายทางภาษาและวัฒนธรรมในประเทศไทย

สนับสนุนโดย Newton Institutional Links
ซึ่งเป็นความร่วมมือระหว่าง

RILCA, MU

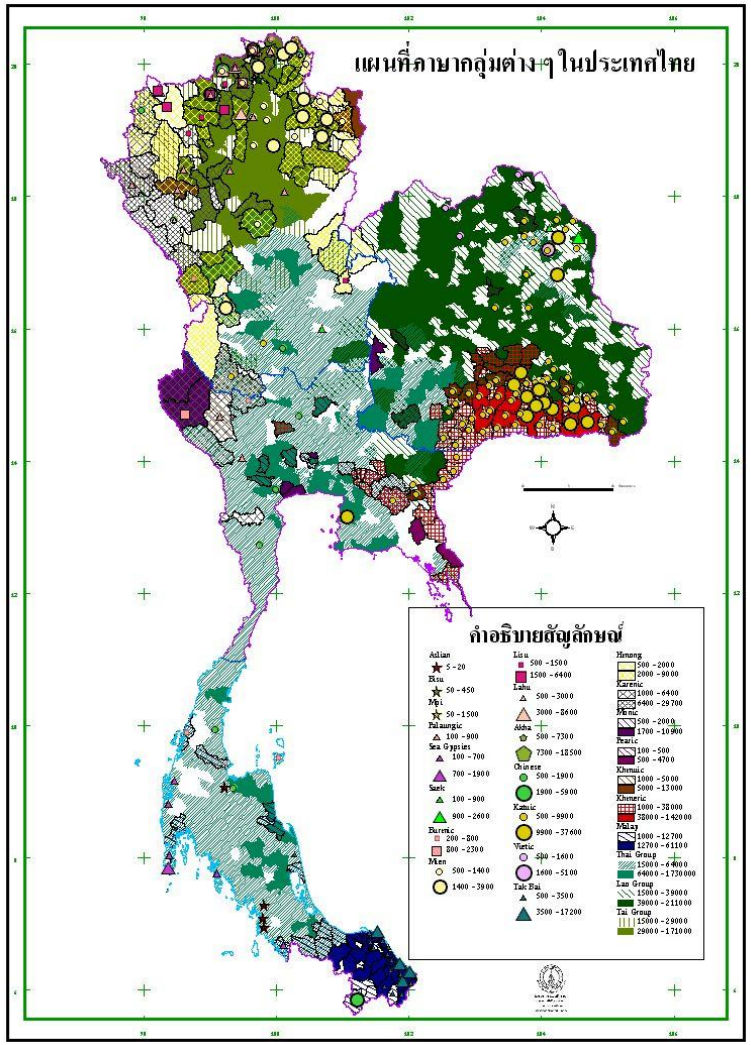


ELAR, SOAS, UK

ระยะเวลา 1 ปี

ที่มาและความสำคัญของโครงการ

ประเทศไทยมีกลุ่มภาษาทั้งสิ้นประมาณ 70 กลุ่ม ในจำนวนนี้มีภาษาที่อยู่ในภาวะวิกฤต ถึง 15 กลุ่ม



โครงการวิจัยนี้ให้ความสำคัญในเรื่อง

- การสูญเสียความหลากหลายทางภาษาและภูมิปัญญาทางวัฒนธรรม
- มุ่งเน้นที่จะพัฒนาวิธีการอนุรักษ์ความหลากหลายทางภาษาและภูมิปัญญาทางวัฒนธรรมโดยปราชญ์และคนในชุมชนเองรวมทั้งพัฒนาศักยภาพคนในระดับท้องถิ่น
- พัฒนาคลังข้อมูลดิจิทัลที่ชุมชนในท้องถิ่นสามารถเข้าถึงได้
- แลกเปลี่ยนเรียนรู้งานวิจัยกับนักวิชาการ ทำให้เกิดผลกระทบระดับสังคม

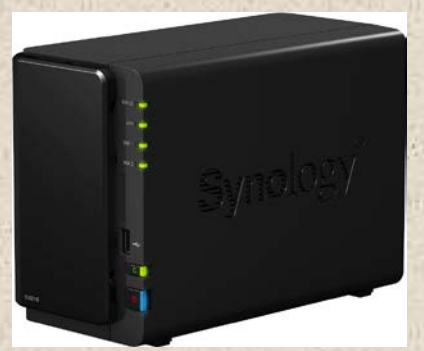
จุดประสงค์ของโครงการ

- จัดทำระบบคลังข้อมูลดิจิทัลให้เป็นมาตรฐาน เพื่อรวบรวมองค์ความรู้ของกลุ่มชาติพันธุ์ในประเทศไทย โดยเฉพาะกลุ่มภาษาที่อยู่ในภาวะวิกฤต โดยผ่านเว็บไซต์
- จัดทำมาตรฐานการเก็บรักษาข้อมูลด้านภาษาและวัฒนธรรมของกลุ่มชาติพันธุ์ในประเทศไทยให้พร้อมเผยแพร่
- สร้างบุคลากรให้สามารถจัดทำคลังข้อมูลดิจิทัล มีความรู้ด้านการดูแลรักษาข้อมูลให้สามารถใช้งานได้ในระยะยาว
- สร้างเครือข่ายการทำงานด้านกลุ่มชาติพันธุ์ ร่วมกับสถาบันการศึกษาอื่น ๆ
- ฝึกอบรมและจัดการประชุมระดับชุมชนและระดับนานาชาติเผยแพร่การจัดทำคลังข้อมูล

http://langarchive-th.com/

การจัดทำคลังข้อมูลภาษาและวัฒนธรรมของกลุ่มชาติพันธุ์

- แบบออฟไลน์
  - จัดการโดยใช้ระบบ **Network-Attached Storage (NAS)** คือ การจัดเก็บข้อมูลของเราบนเครือข่าย
  - ใช้สำหรับ เก็บข้อมูลต้นฉบับ (back up ข้อมูล)
  - ใช้สำหรับจัดการไฟล์ข้อมูลอย่างเป็นระบบ



- **แบบออนไลน์**
  - **ใช้สำหรับนำเสนอข้อมูล ผ่านเว็บไซต์ (Mukurtu)**
  - **ข้อมูลประเภทไฟล์เสียง ให้ upload ผ่าน soundcloud**
  - **ข้อมูลประเภทไฟล์วิดีโอ ให้ upload ผ่าน vimeo**
  - **ข้อมูลประเภทไฟล์ภาพ หรือไฟล์เอกสาร สามารถ upload ผ่าน mukurtu ได้เลย**

การจัดเก็บคลังข้อมูลภาษาและวัฒนธรรมของกลุ่มชาติพันธุ์

ตรวจสอบ-คัดแยกข้อมูล

การจัดการข้อมูล

จัดระบบฐานข้อมูล
Metadata (excel)

ข้อมูล analogue
- คาสเซ็ทเทป
- ม้วนฟิล์ม
- หนังสือ
- cd-rom

ข้อมูล digital
- ไฟล์ทุกประเภทที่จัดเก็บในคอมพิวเตอร์

แปลงข้อมูลด้วยอุปกรณ์ดิจิทัล
- แปลงเทปด้วยเครื่องบันทึกเสียงดิจิทัล
- สแกนฟิล์ม
- สแกน/ถ่ายภาพหนังสือ
- cd แปลงไฟล์

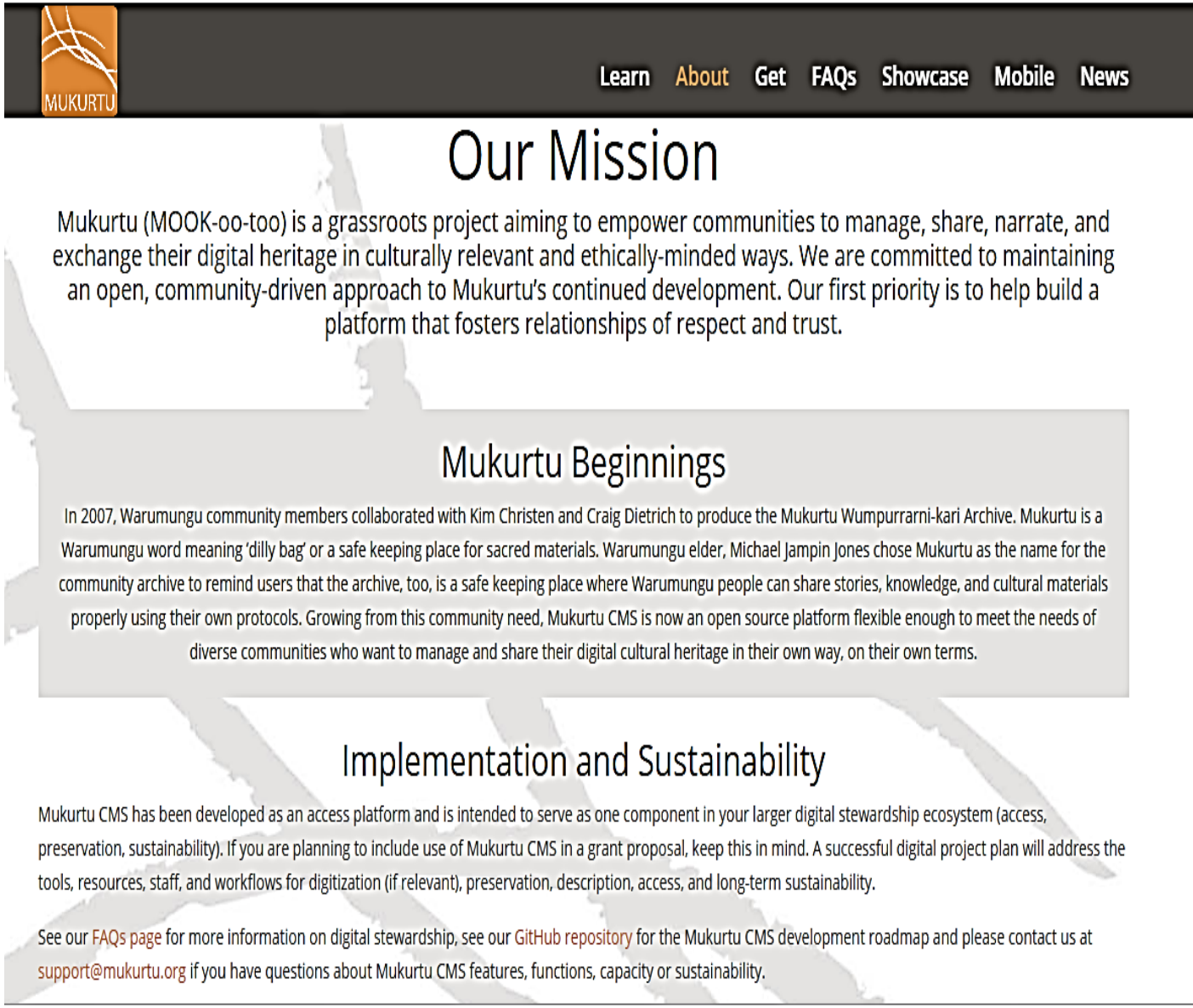
ไฟล์ดิจิทัล

เก็บใน hdd ของระบบ NAS

- การจัดระบบไฟล์ข้อมูล..เปลี่ยนชื่อไฟล์ทั้งหมดให้เป็นระบบเดียวกัน
- ใส่ข้อมูลทั้งหมดเกี่ยวกับไฟล์นั้นลง metadata sheet

Mukurtu

- ระบบที่สร้างเว็บไซต์มาเพื่อรองรับการนำเสนอข้อมูลด้านภาษาและวัฒนธรรมของกลุ่มชาติพันธุ์
- ออกแบบแพลตฟอร์มที่ง่ายต่อการนำเข้าข้อมูลด้วยตนเอง
- ผู้ใช้สามารถออกแบบหน้าตาของเว็บไซต์ได้เองตามรูปแบบที่มีหลากหลาย เลือกชมได้ที่ http://mukurtu.org/showcase/
- เว็บไซต์จะรองรับไฟล์ที่มีขนาดไม่ใหญ่
- เว็บไซต์รองรับภาษาไทย
- รองรับแพลตฟอร์มของมือถือ แท็บเล็ต

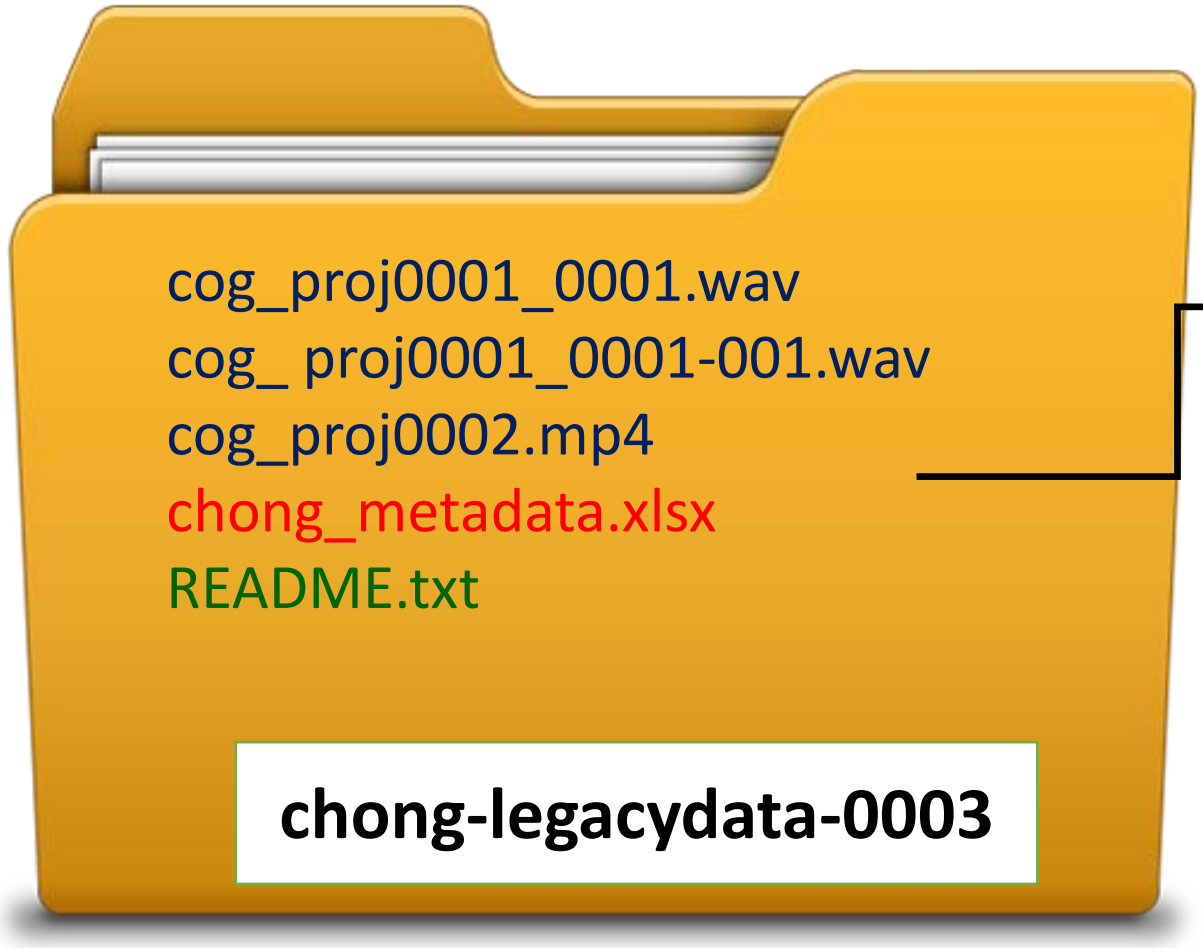


ขั้นตอนในการจัดการข้อมูล

- ข้อมูลเดิม (ข้อมูลเก่า)
  - แปลงไฟล์จาก analog ให้เป็น digital
    - Cassette tape ---- .mp3
    - VDO Tape ---- .mp4
    - Scan ภาพถ่าย / หนังสือเล่มเล็ก ----- .jpg
    - reversal films ---- .jpg
    - negative films ----- .jpg
    - หนังสือเล่มยักษ์ ใช้วิธีการถ่ายภาพ ------ .jpg
  - Back up ข้อมูล ลงใน NAS
  - จัดระบบไฟล์ข้อมูล

- ข้อมูลปัจจุบัน (ที่เป็นไฟล์ digital มาแล้ว)
  - Back up ข้อมูล ลงใน NAS
  - แปลงไฟล์ข้อมูล ให้มีขนาดเล็กลง
    - ไฟล์ภาพ ---- .jpg
    - ไฟล์เสียง ----- .mp3
    - ไฟล์วิดีโอ ------.mp4
    - ไฟล์เอกสาร ---- .pdf
    - eaf
    - Flex
  - จัดระบบไฟล์ข้อมูล

การจัดระบบไฟล์ข้อมูล



ภายใน folder ประกอบไปด้วยข้อมูล

1) เสียง, ภาพ, วิดีโอ, ข้อมูลเอกสาร ฯลฯ โดยข้อมูลแต่ละประเภทมีการตั้งชื่อไฟล์อย่างเป็นระบบ
2) ข้อมูล metadata ในรูปแบบ spreadsheet
3) ความหมายของ code

การตั้งชื่อไฟล์

- ชื่อไฟล์ประกอบด้วย

ชื่อภาษา	ลำดับเลข ระบุชุดไฟล์ข้อมูล	ลำดับเลข ระบุจำนวนไฟล์	นามสกุลไฟล์
• Melayu • Chong	• _0001 • _0002	• _00001 • _00002	• .jpg .pdf • .wav .doc • .mp4

melayu_0001_00001.jpg
melayu_0001_00002.jpg
melayu_0001_00001.pdf
melayu_0001_00001.mp4

melayu_0020_00001.jpg
melayu_0020_00002.jpg
melayu_0020_00001.pdf
melayu_0020_00001.mp4

melayu_0310_00001.jpg
melayu_0310_00002.jpg
melayu_0310_00001.pdf
melayu_0310_00001.mp4

- ข้อกำหนด
  - ✓ ตัวอักษรควรเป็น ภาษาอังกฤษ ตัวเล็ก
  - ✓ ตัวเลขควรนำด้วย 0 หรือ 00
  - ✓ สัญลักษณ์คั่น – (hyphen) หรือ _ (underscore)

- ข้อห้าม
  - ✗ ตัวอักษรเกิน 10
  - ✗ ตัวอักษรภาษาไทย
  - ✗ ช่องว่าง
  - ✗ อักขระพิเศษอื่น ๆ

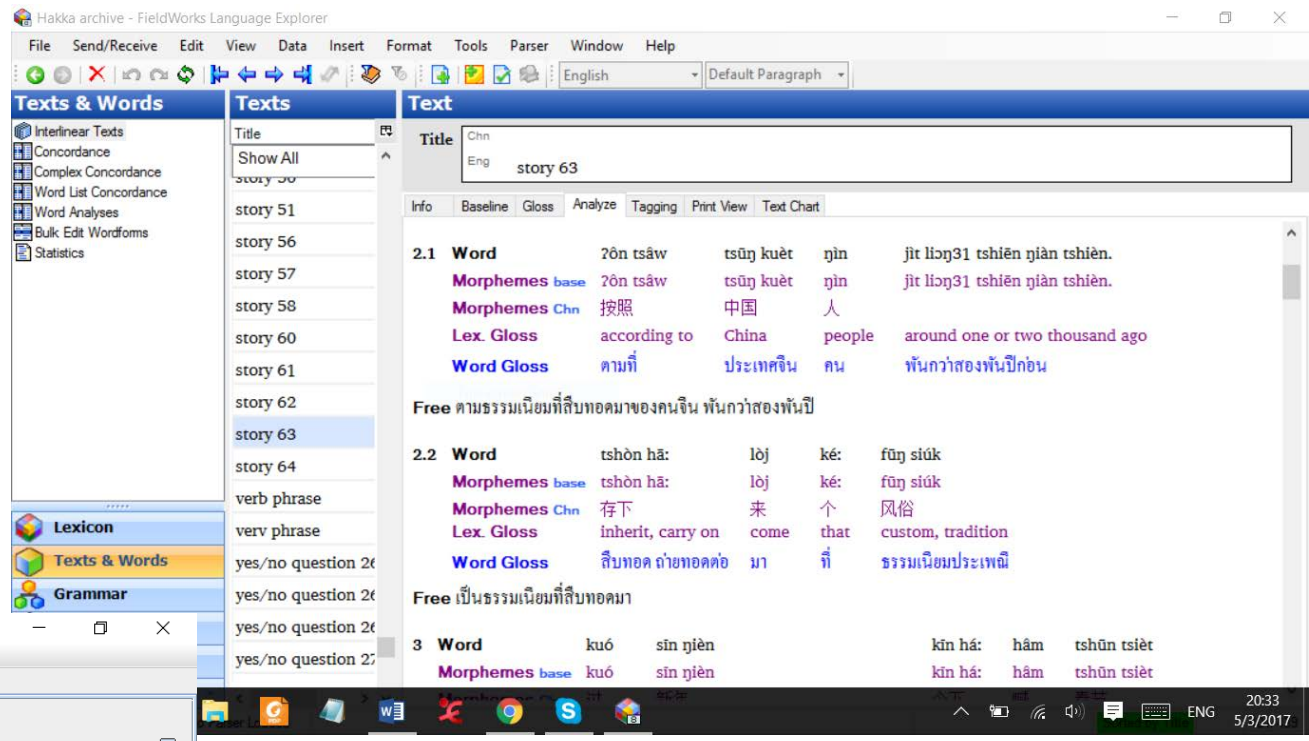
ข้อมูล Metadata

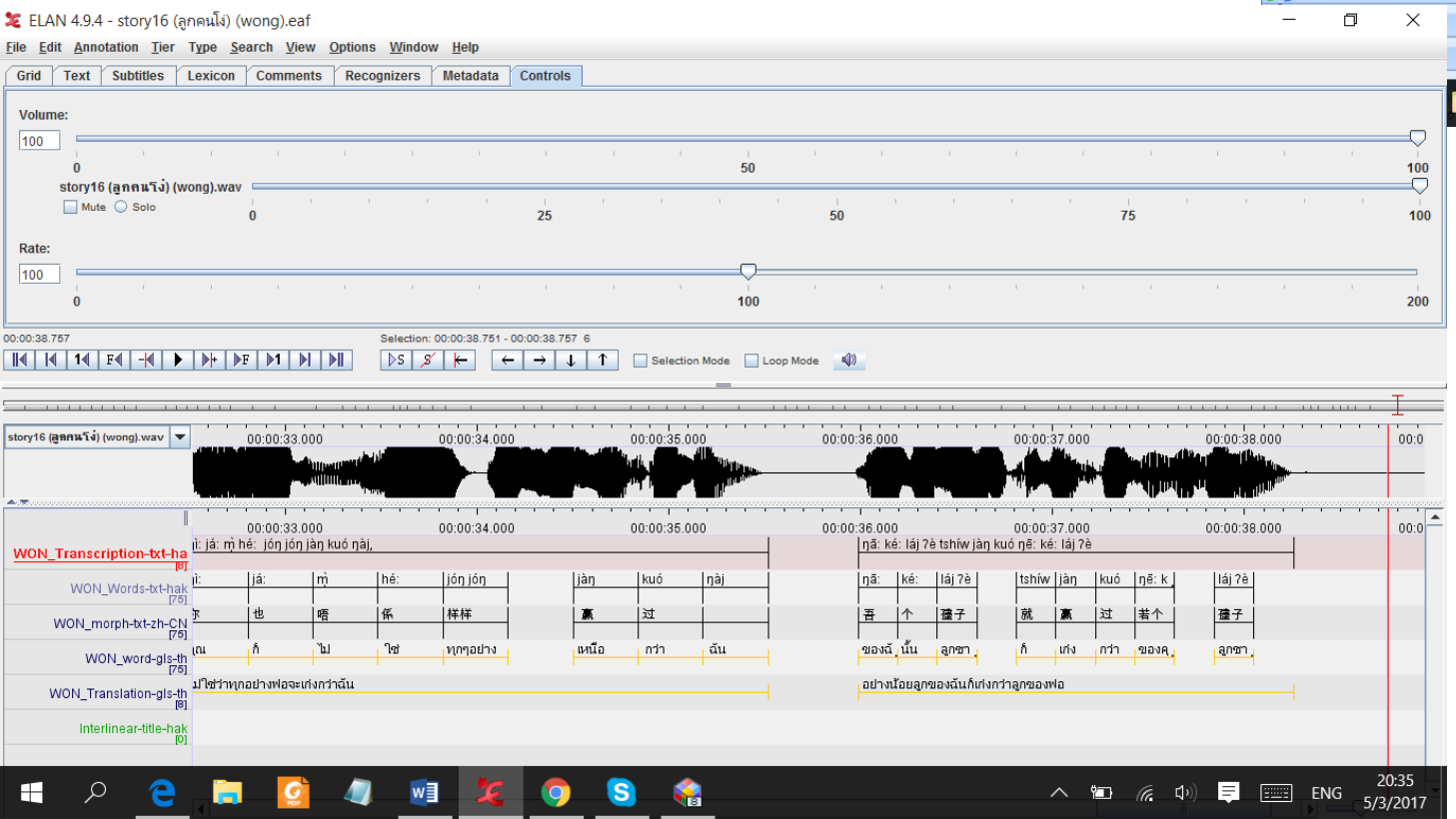
- **Title หัวเรื่อง / หัวข้อ**
- **Summary ข้อมูลที่สรุปอย่างสั้น**
- **Category ประเภทข้อมูล**
- **Creator ผู้ผลิตข้อมูล ทุกคนที่เกี่ยวข้องกับการผลิตข้อมูล**
- **Contributor ผู้ให้ข้อมูล เจ้าของข้อมูล**
- **Original Date วันที่บันทึกข้อมูล / วันที่ผลิตข้อมูล**
- **Description คำอธิบายรายละเอียดของข้อมูล**
- **Traditional Knowledge อธิบายรายละเอียดของข้อมูลเฉพาะที่ต่างจากกลุ่มอื่นๆ**
- **Keywords คำสำคัญ**
- **Publisher ผู้เผยแพร่ข้อมูล**
- **Rights ผู้มีสิทธิ์เข้าถึงข้อมูล**
- **Format รูปแบบของไฟล์ เช่น MP4 ; JPG ; PDF**
- **Type ประเภทของข้อมูล เช่น Video ; Image ; Text**
- **Identifier ได้จาก website www.gguid.com**
- **Language เนื้อหาของภาษา**
- **People ผู้เกี่ยวข้องทั้งหมด**
- **อื่นๆ เช่น ที่มาของไฟล์** (หนังสือ หรือวัตถุข้อมูล) **วัตถุ/ไฟล์อื่นที่เกี่ยวข้องกัน**

ประโยชน์ของคลังข้อมูล

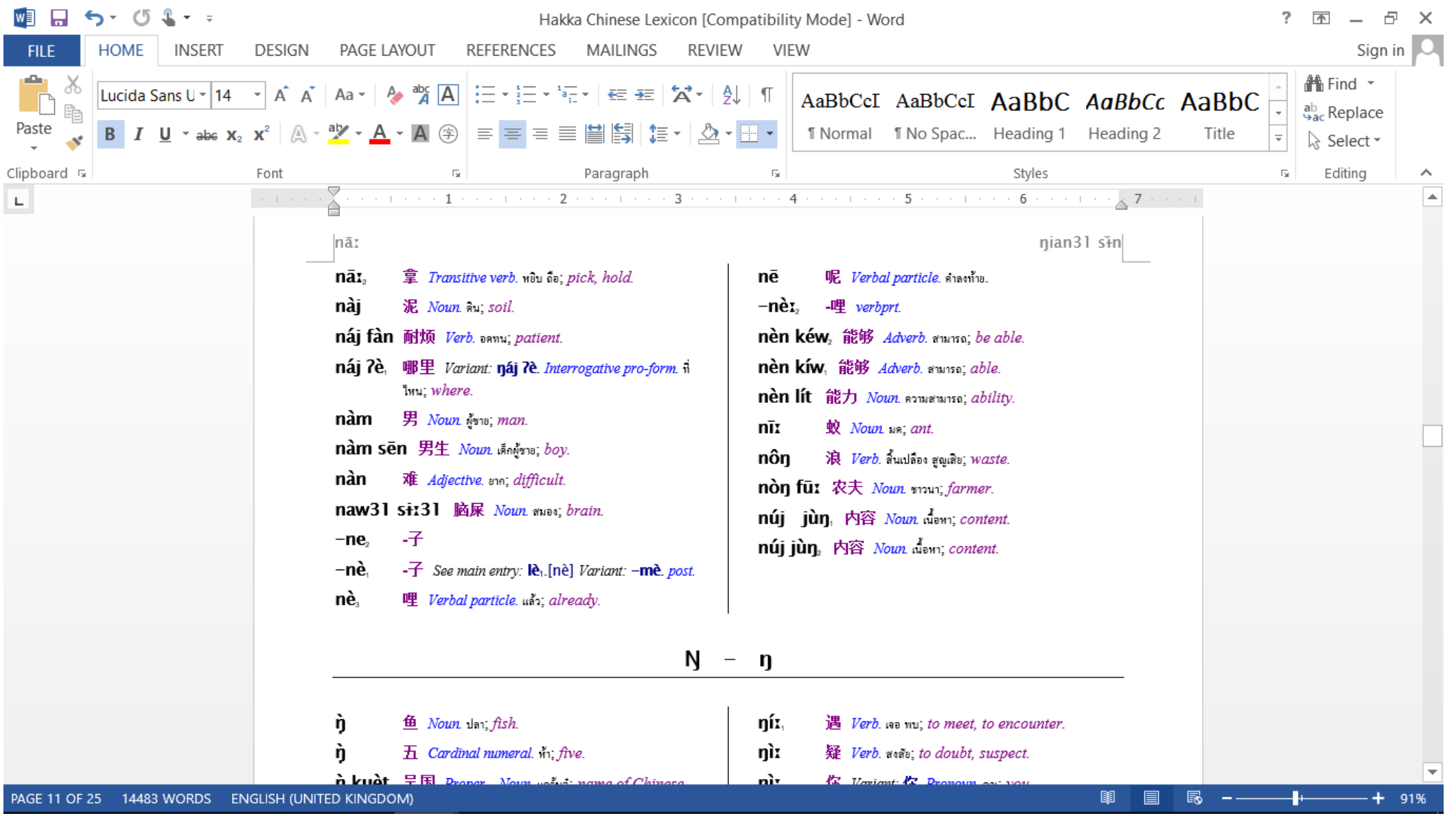
- เป็นการจัดการฐานข้อมูลขนาดใหญ่
- เป็นที่รวบรวมฐานข้อมูล และเป็นแหล่งเรียนรู้ทางภาษาและวัฒนธรรมของกลุ่มชาติพันธุ์ในประเทศไทย
- นักวิชาการสามารถใช้ข้อมูลจากคลังนี้ไปศึกษาเปรียบเทียบ หรือศึกษาต่อ
- เจ้าของภาษาสามารถศึกษาค้นคว้า และเป็นแหล่งเรียนรู้ของลูกหลานต่อไป
- ธำรงไว้ซึ่งความหลากหลายของสังคมพหุวัฒนธรรม
- เจ้าขององค์ความรู้เกิดความภูมิใจ ที่มีการบันทึกสิ่งที่มีค่าไม่ให้สูญหาย
- เป็นพันธกิจหนึ่งของสถาบันฯ และมหาวิทยาลัยมหิดล

การวิเคราะห์ข้อมูลทางภาษาศาสตร์

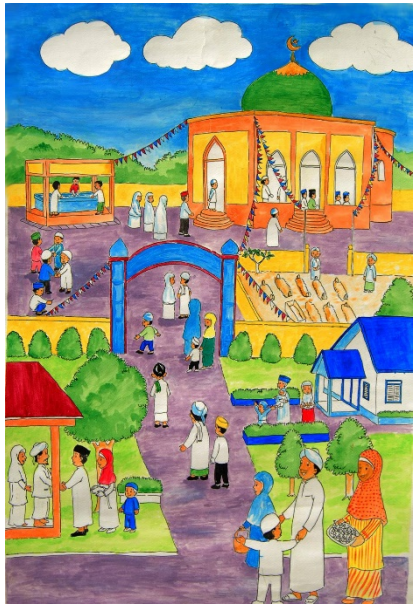




ตัวอย่างพจนานุกรม



ตัวอย่างการแปลงไฟล์ *หนังสือเล่มยักษ์หรือฉากภาพวัฒนธรรม*



ตัวอย่างการเก็บข้อมูลภาคสนาม

ตัวอย่างการเก็บข้อมูลภาคสนาม

ตัวอย่างการเก็บข้อมูลภาคสนาม

ตัวอย่างการจัดอบรมการบันทึกข้อมูลและทำคลังข้อมูล

ตัวอย่างการจัดอบรมการบันทึกข้อมูลและทำคลังข้อมูล

ข้อดีของการทำงานภาคสนาม: ทำบุญ ทัศนศึกษา

ขอบคุณค่ะ Thank you 谢谢！