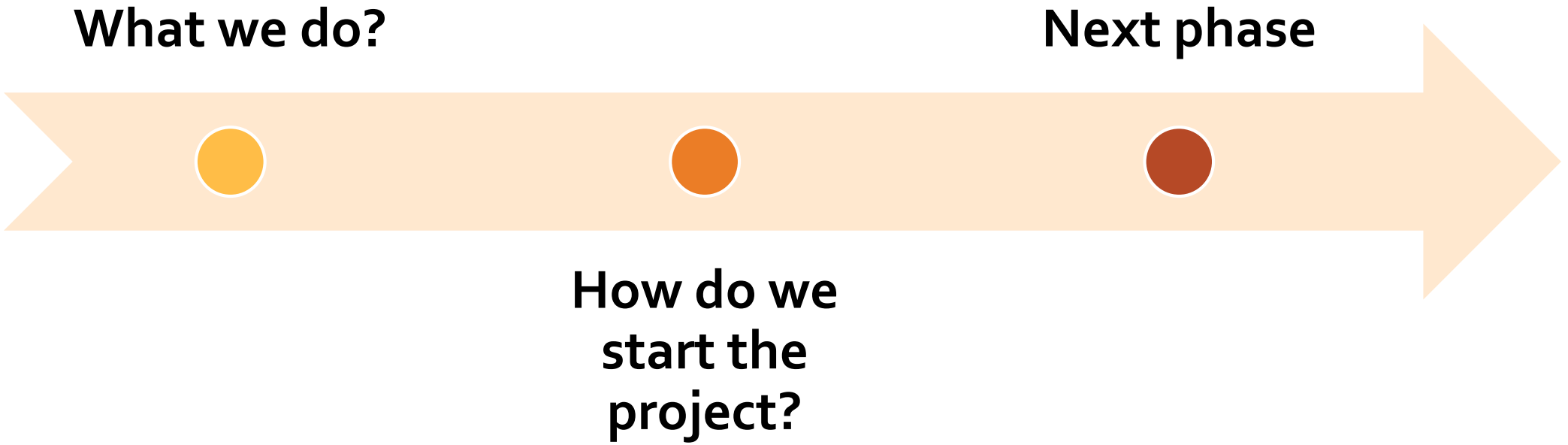
Archiving Web Content in Anthropology: Lessons Learned for a Step in the Right Direction

IIPC 2018, The National Library of New Zealand Te Puna Mātauranga o Aotearoa, Wellington, New Zealand

Chindarat Berpan, Asst. Prof.
Wachiraporn Klungthanaboon, PhD
Sittisak Rungcharoensuksri

Outline



"Through anthropology, humans learn about humanity, and learn to understand themselves and their fellow humans, forming the basis for mutual understanding and peaceful coexistence."

Her Royal Highness Princess Maha Chakri Sirindhorn

9th March, 1999

Databases & Resources

Research

Knowledge Sharing & Outreach Activities

Online databases

- Academic databases (21)
- Library

Tools

- SAC Portal
- SAC Channel
- SMART SAC
- The Humans (Web magazine)

Exhibition

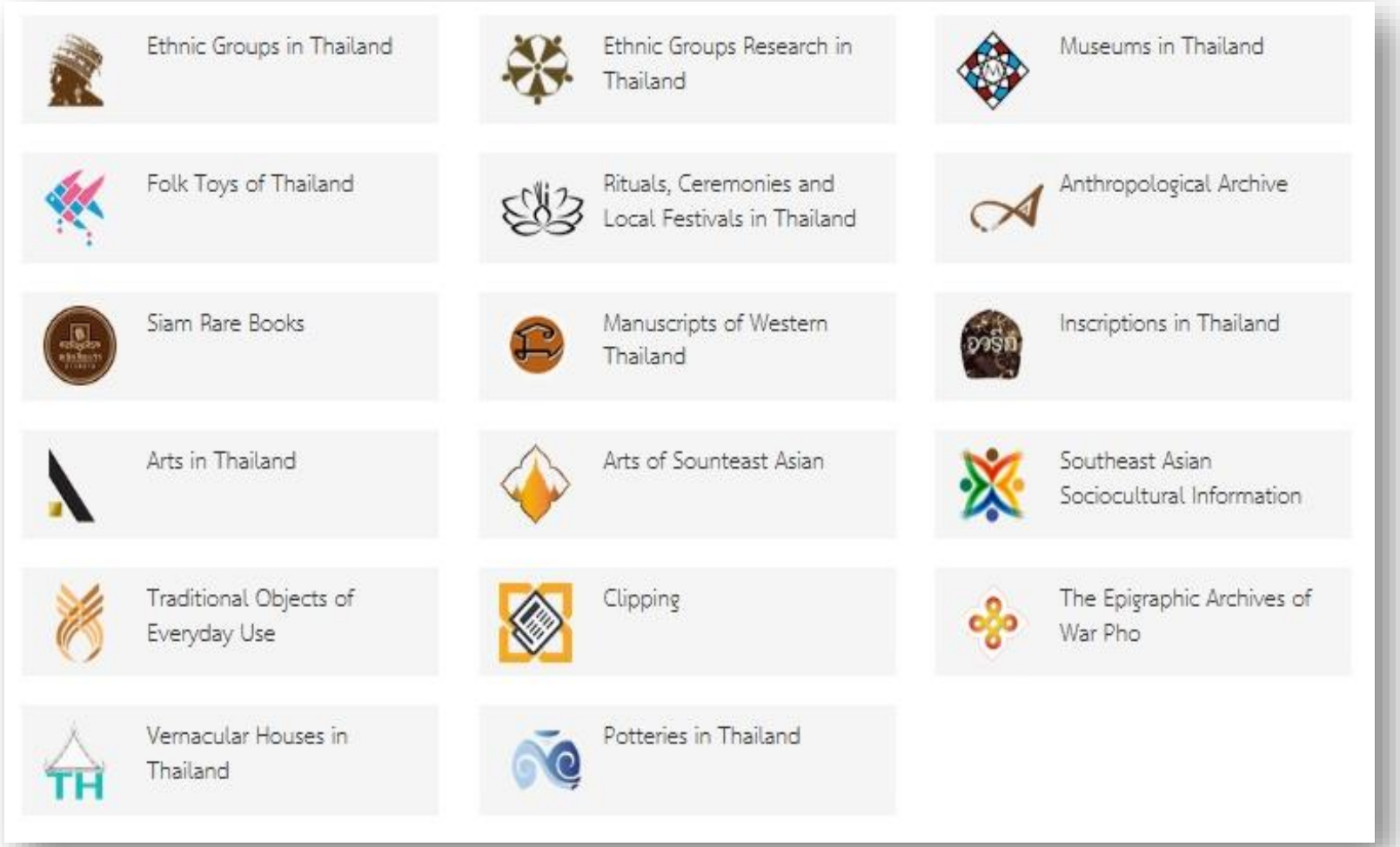
MIS Intranet (17 systems)

- SAC Archives
- SAC IR
- Room booking system
- Car booking system
- KPI evaluation system
- Project evaluation system
- Competency evaluation system
- Account system
- Etc.

Sac's digital repository (1999-2018)

WWW.SAC.OR.TH

SAC's academic databases



Archaeology and History

Anthropology

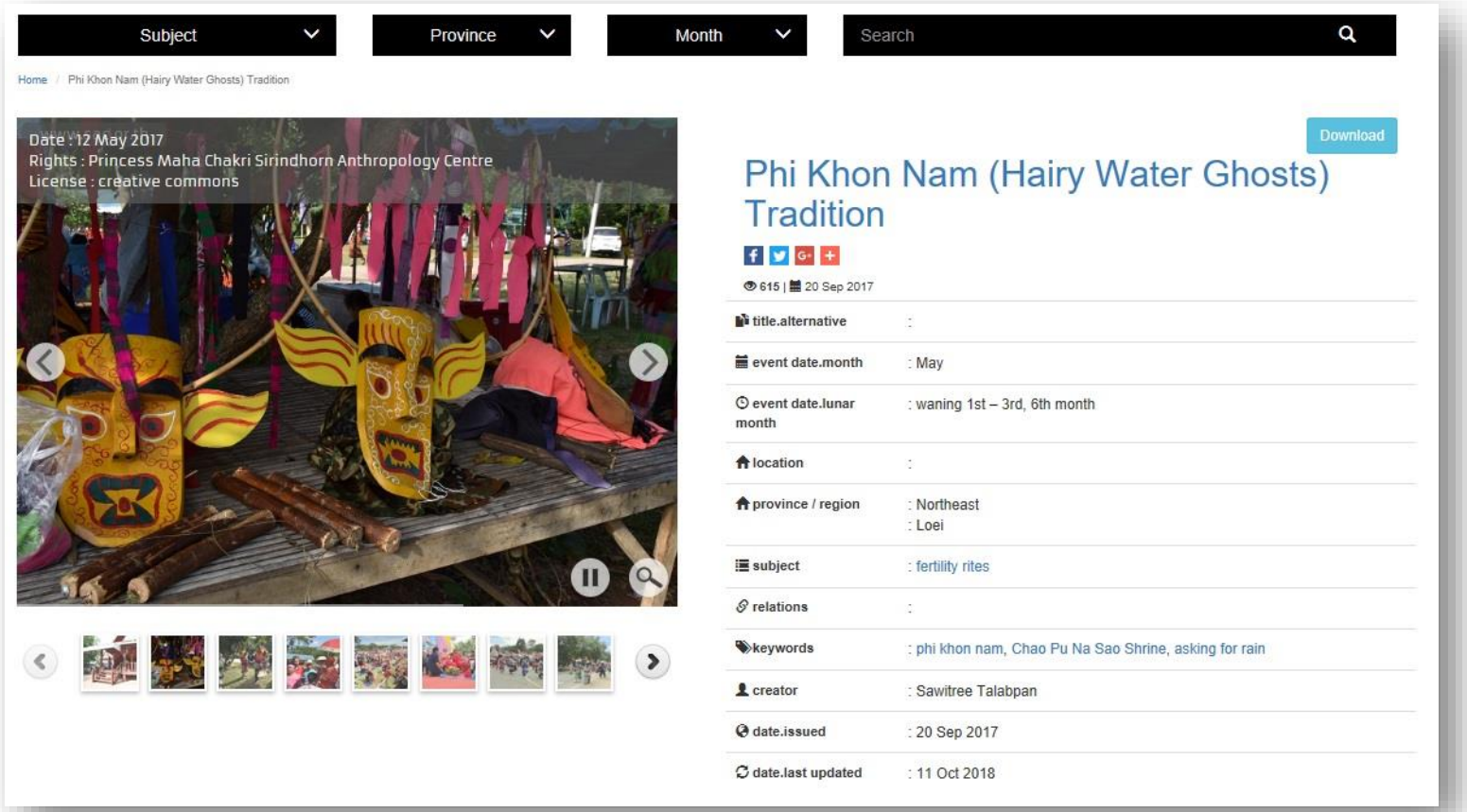
Ethnology

Socio-Cultural

Now - 21 databases

Future - 31 databases in 2022

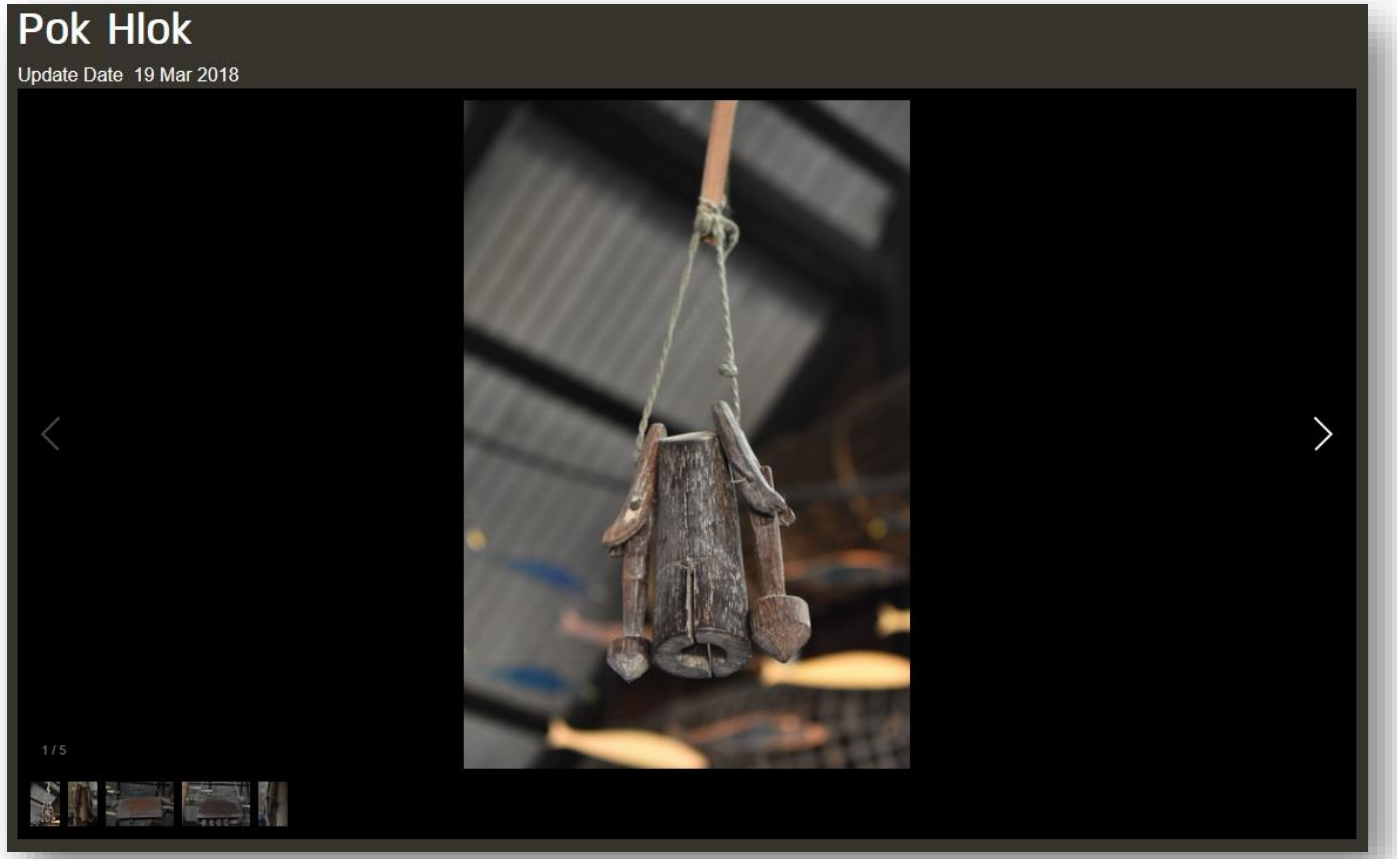
Examples of the SAC's digital resources (1)



The Rituals, Ceremonies and Local Festivals in Thailand Database

Source: http://www.sac.or.th/databases/rituals/en/detail.php?id=109

Examples of the SAC's digital resources (2)



Traditional Objects of Everyday Use Database

Source: http://www.sac.or.th/databases/traditional-objects/en/equipment-detail.php?ob_id=139

Alternative name
Mai Pong or Pok Ka Hlok or Look Kro

Creator
Bhijja Thongklib

Material
wood

Date.write
2017

Place
Sergeant Thawee Folk Museum

Address
26/138 Wisutkasat Rd., Muang District, Phitsanulok Province 65000

Region
Central region

Classification
Agricultural implements

Format
MsWord

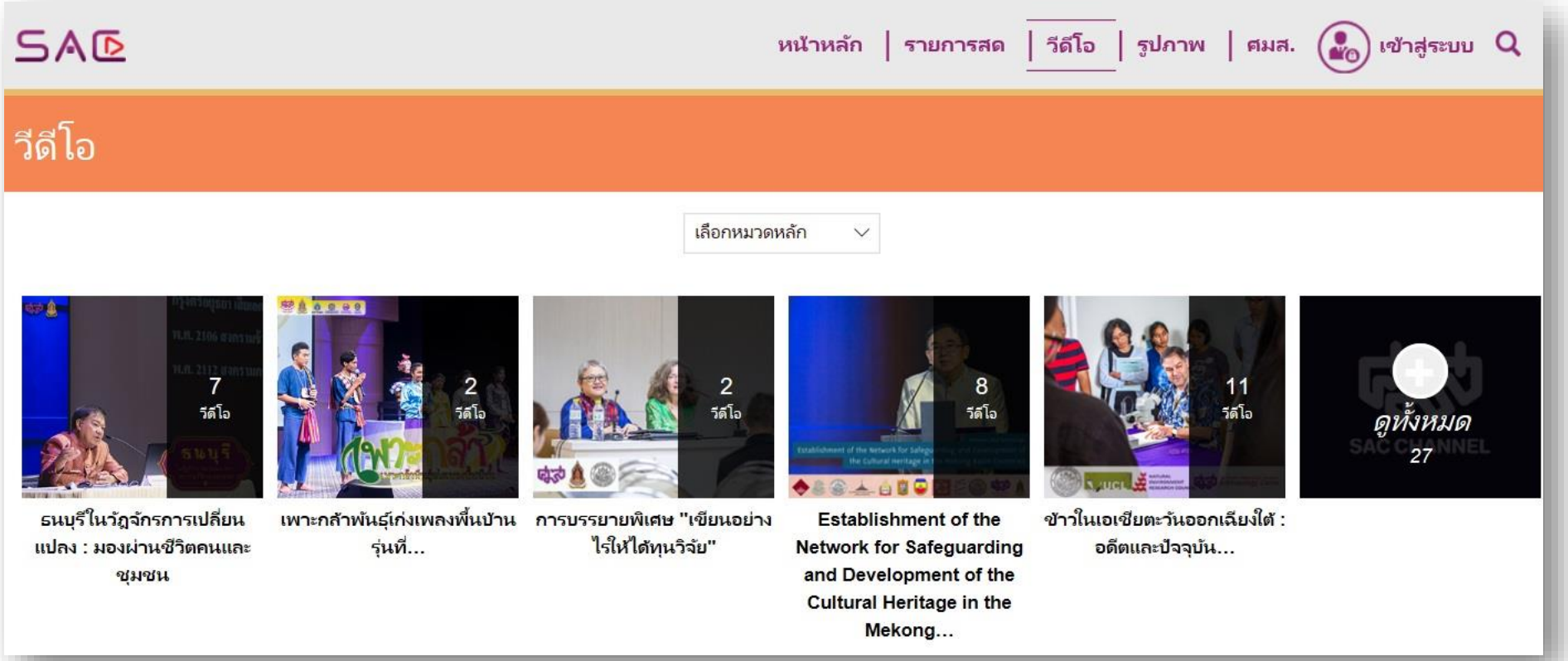
Type
text

Publisher
Princess Maha Chakri Sirindhorn Anthropology Centre (Public Organization)

Rights
Princess Maha Chakri Sirindhorn Anthropology Centre (Public Organization)

Contributors
Sergeant Thawee Folk Museum

Examples of the SAC's digital resources (3)



SAC CHANNEL

LIVE video streaming, VDOs, and Images from SAC's conferences and public activities

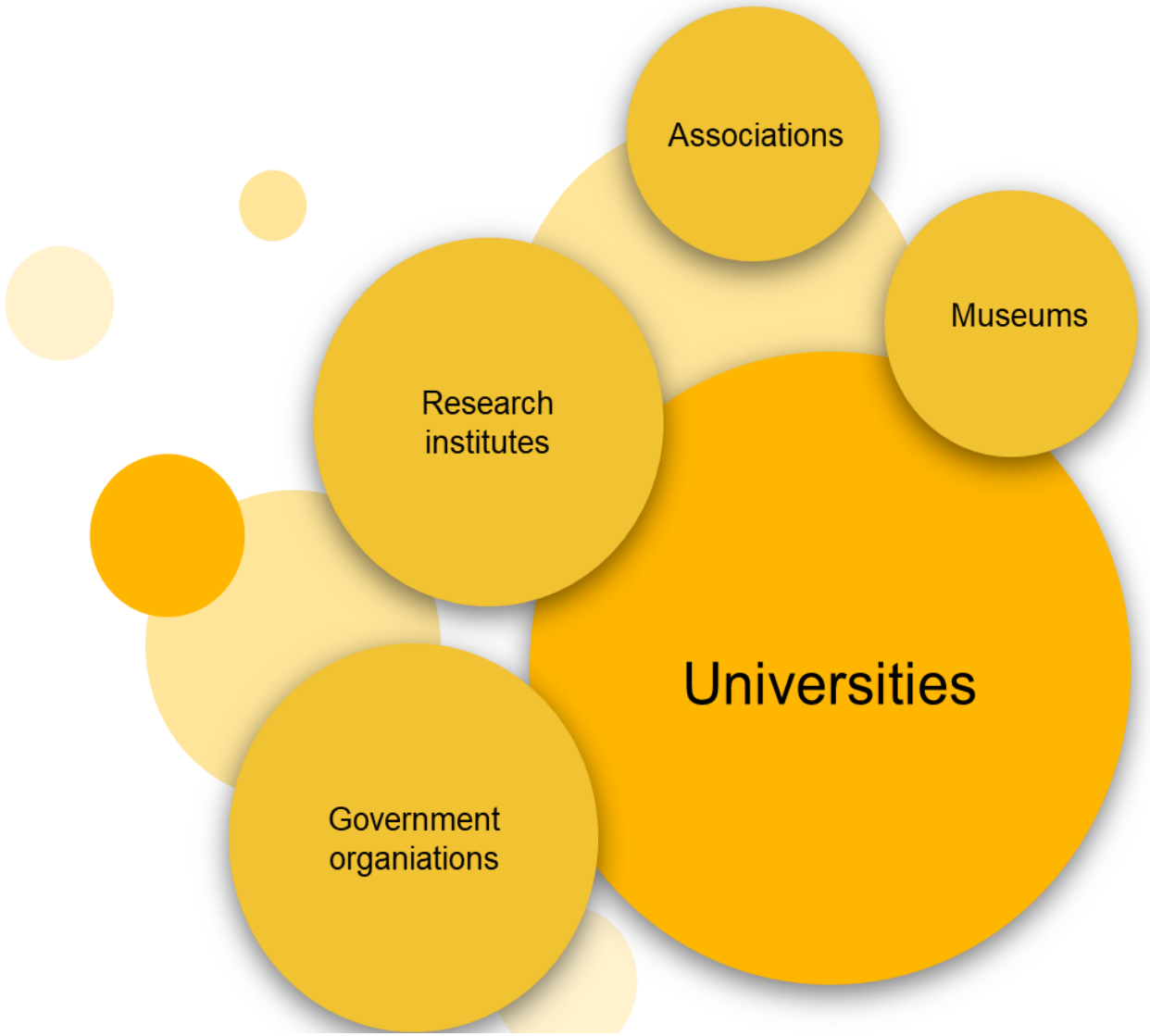
Old version of SAC's digital repository

- Recorded by the Internet Archive on 27 Apr. 1999
- 2 databases: Museum and Ethnic Groups Database

Source: https://web.archive.org/web/19990427133516/http://www.sac.or.th:80/

Process

Web content analysis
Literature review
Insights on web archiving lifecycle, components, and challenges



Anthropological content on websites

30 selected websites of organisations in Thailand and Southeast Asia

Genres

- Collection
- Dataset
- Event
- Image
- Interactive Resource
- Moving Image
- Sound
- Text

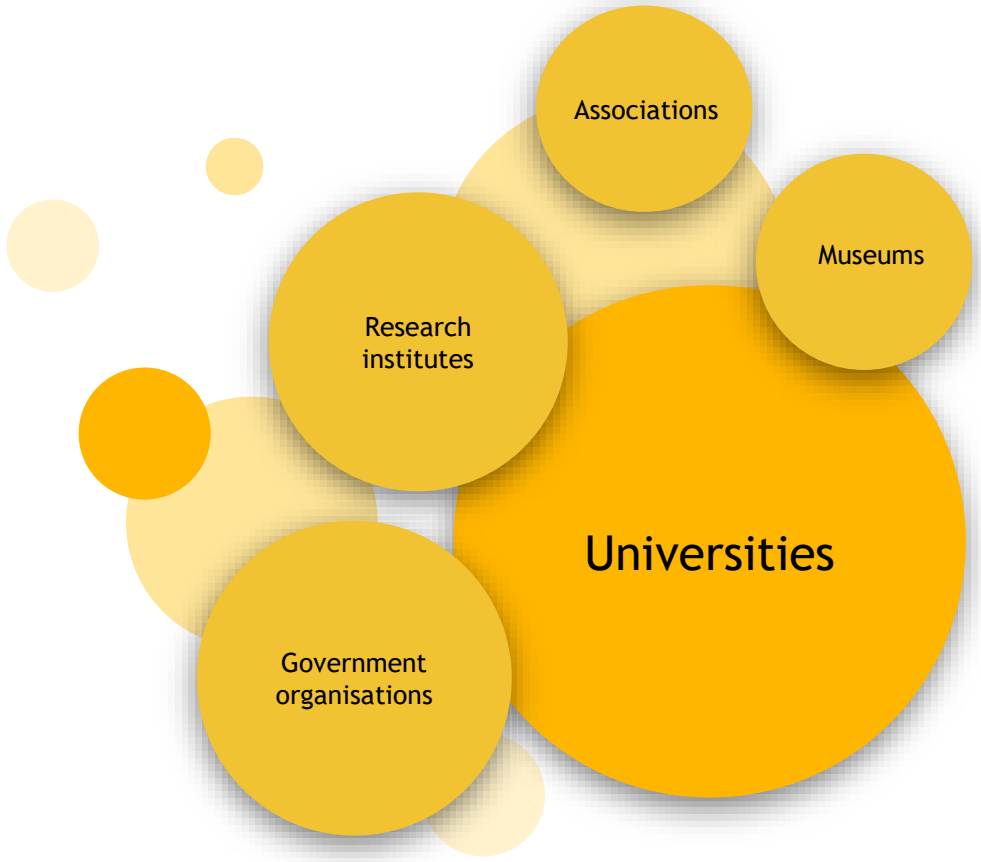
File formats

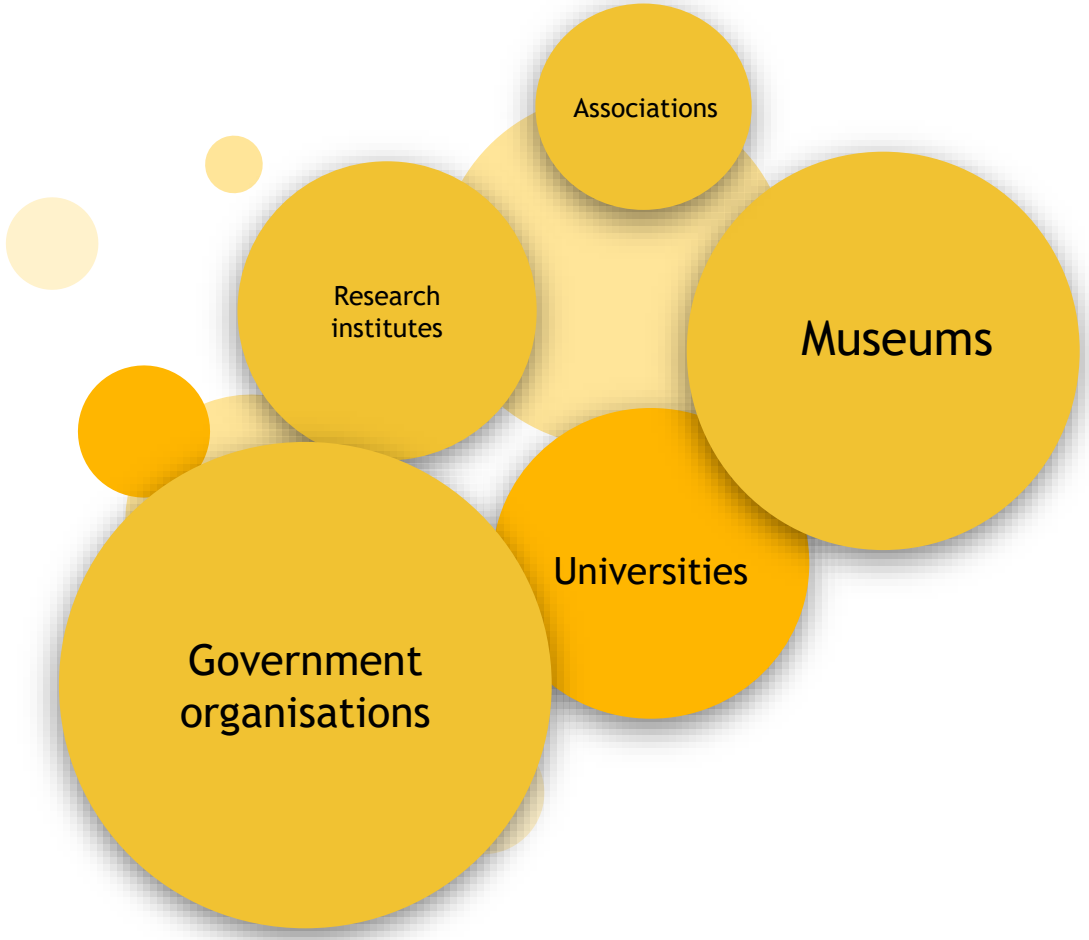
- Application
- Audio
- Video
- Text
- Image

Topics

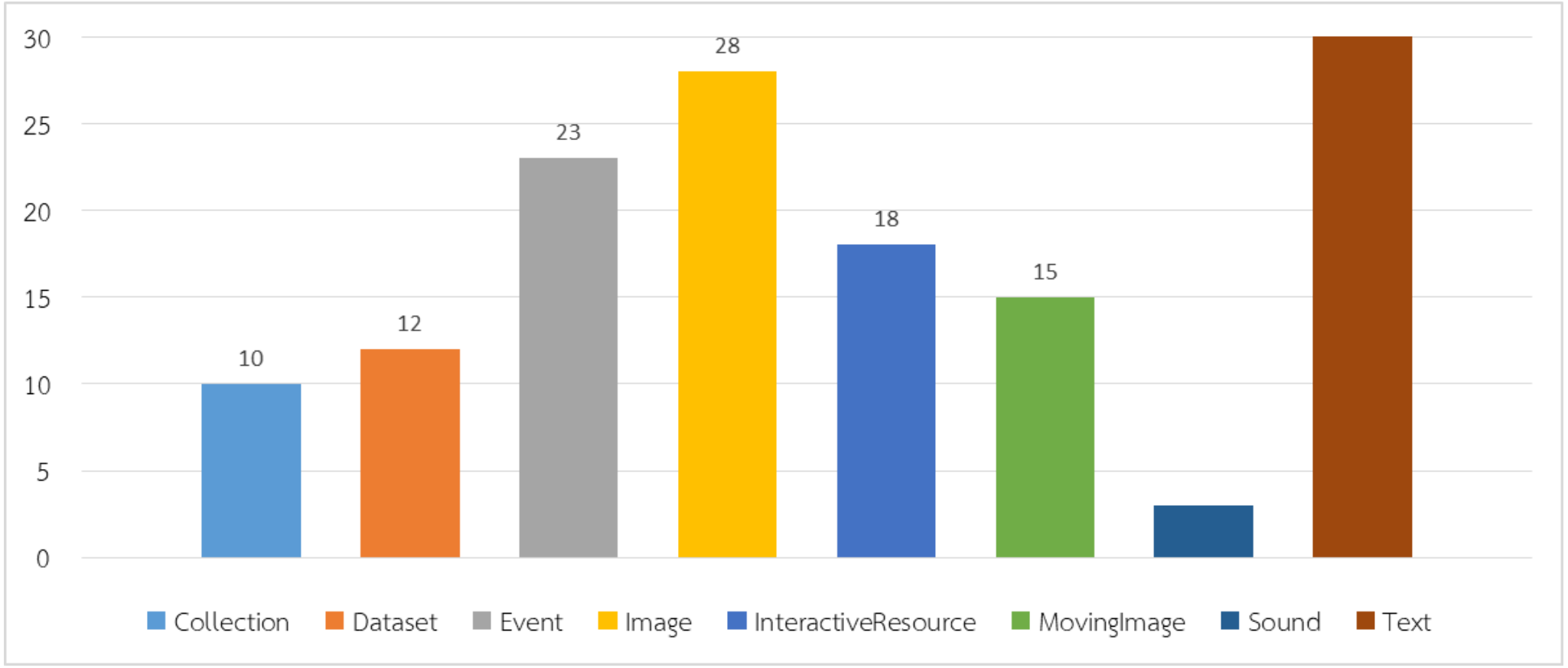
- Civilization and Archaeology
- Biographies
- Thai history
- Asian history
- Environments
- Languages and literatures
- Museums
- Education
- Technologies
- Agricultures
- Philosophy
- Society and social issues
- Ethnicity
- Folklore
- Etiquette and customs
- Music and recreations
- Tourism and travel
- Laws
- Arts

Types of organization providing content on their websites

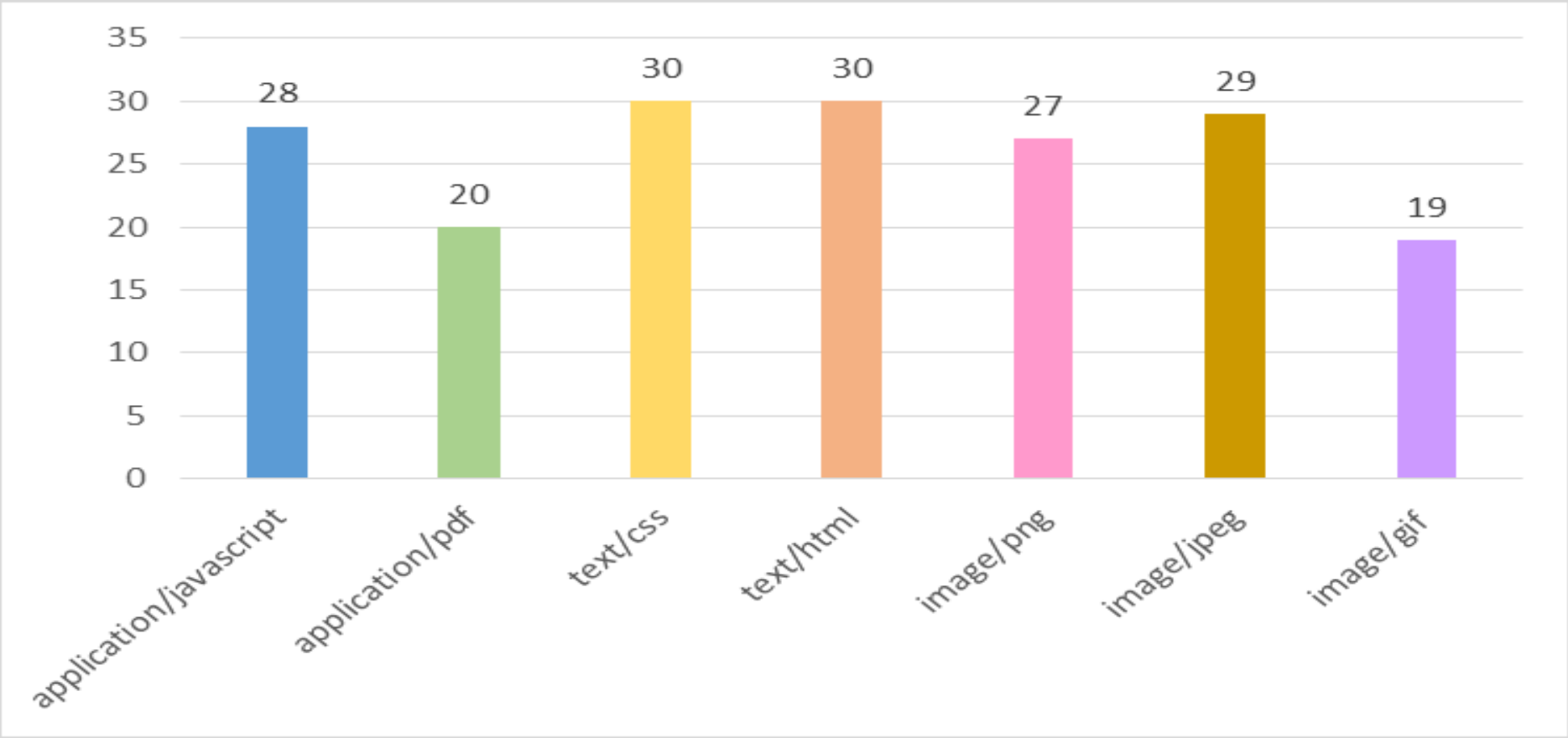




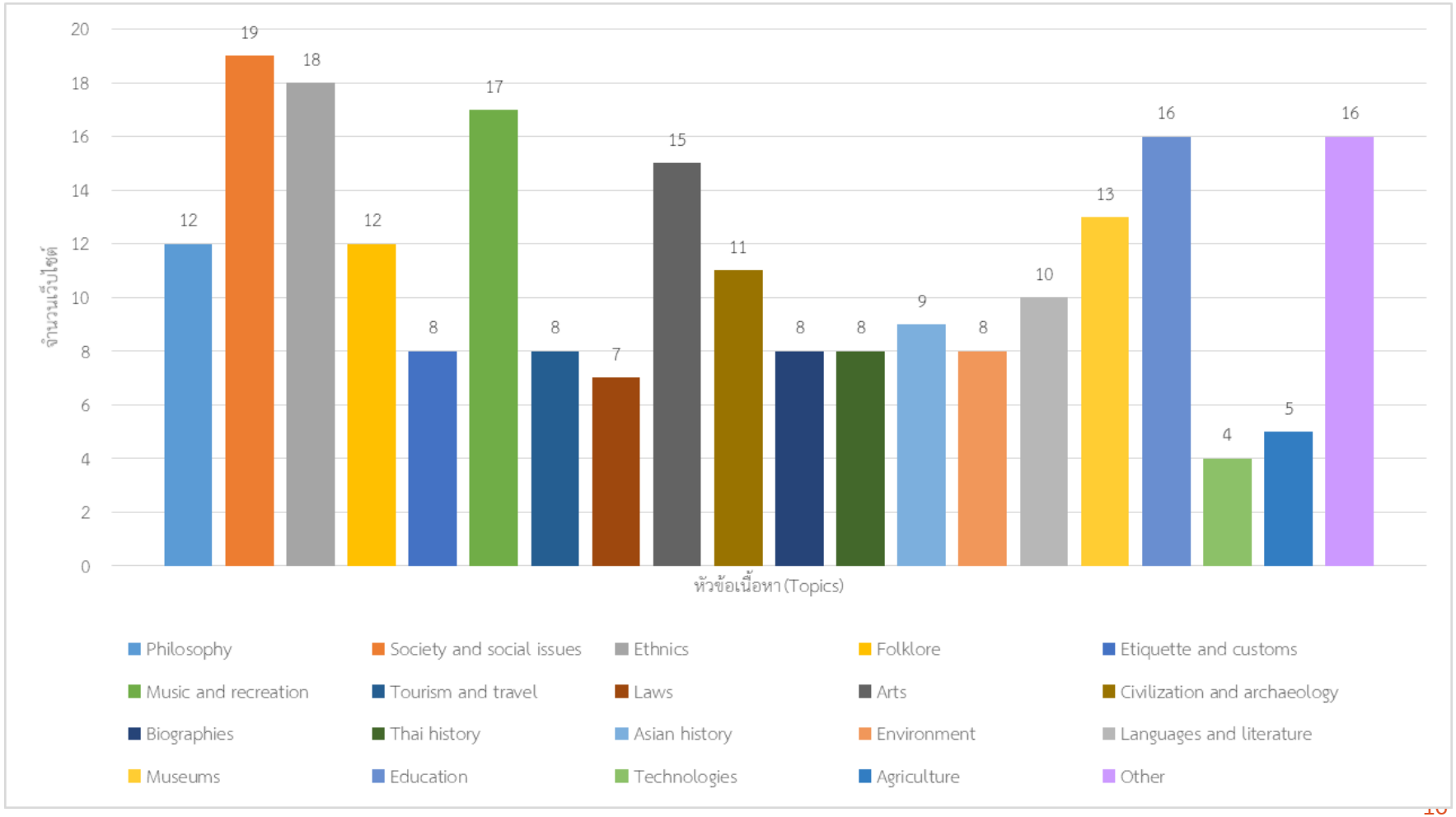
Genres of web content on the selected websites

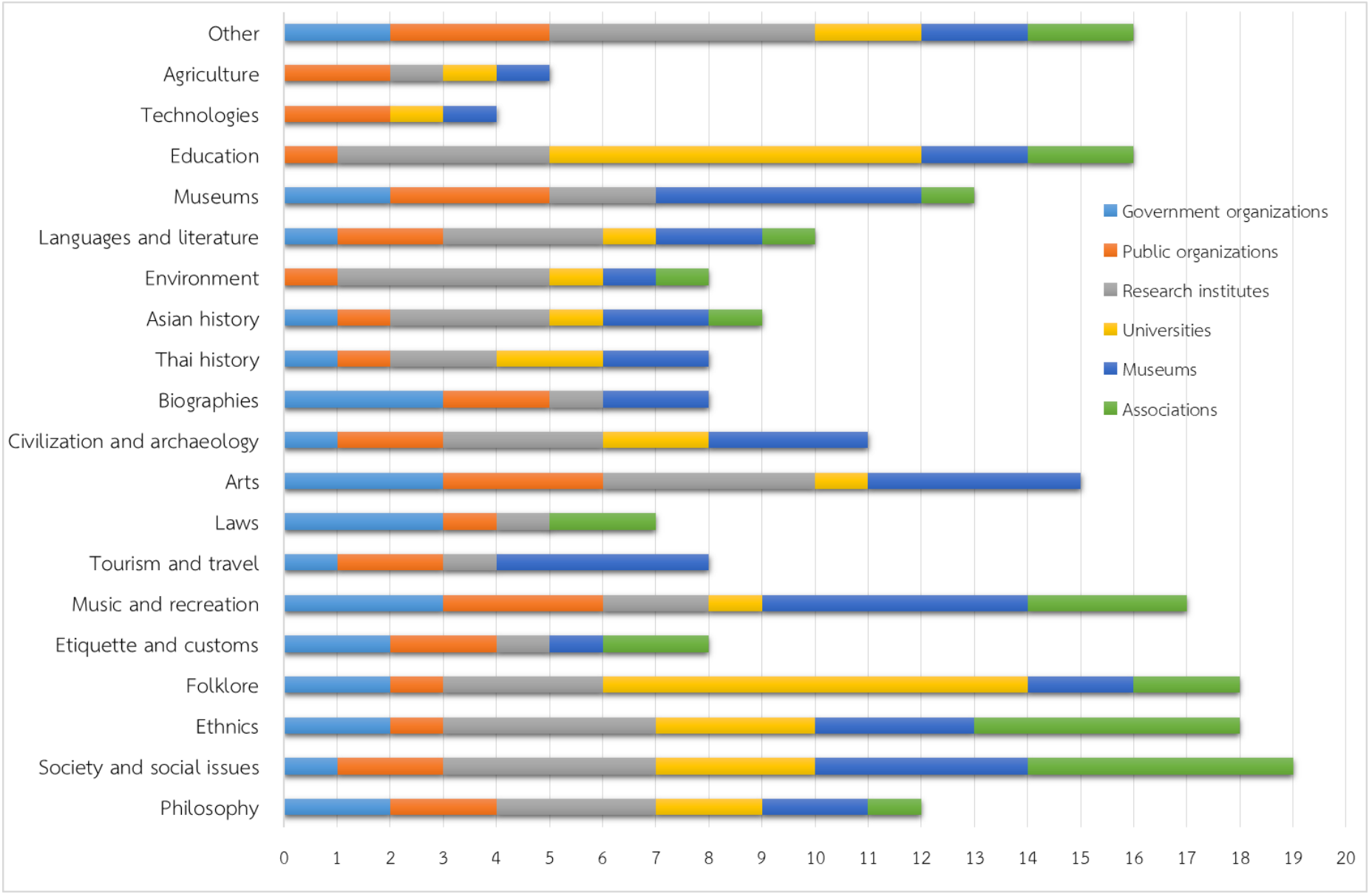


File formats of anthropological web content

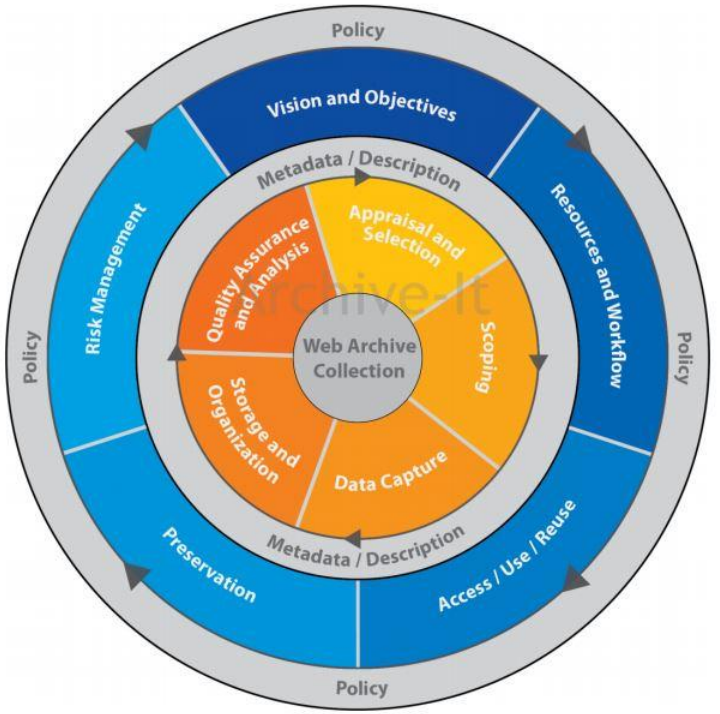


Anthropological web content available on the selected websites

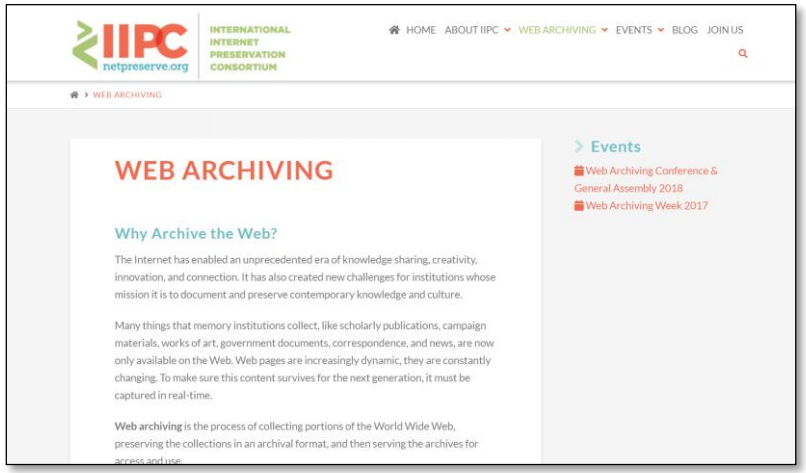



Other
Agriculture
Technologies
Education
Museums
Languages and literature
Environment
Asian history
Thai history
Biographies
Civilization and archaeology
Arts
Laws
Tourism and travel
Music and recreation
Etiquette and customs
Folklore
Ethnics
Society and social issues
Philosophy
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Government organizations
Public organizations
Research institutes
Universities
Museums
Associations

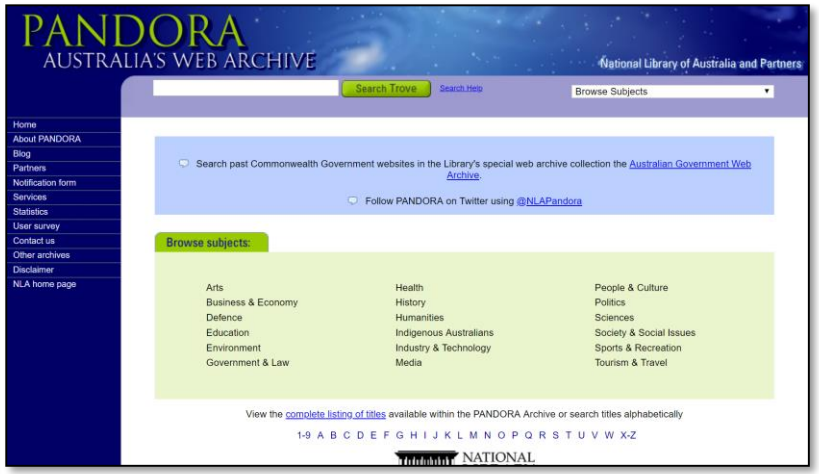
Web content analysis
Review literature
Insights on web archiving lifecycle, components, and challenges

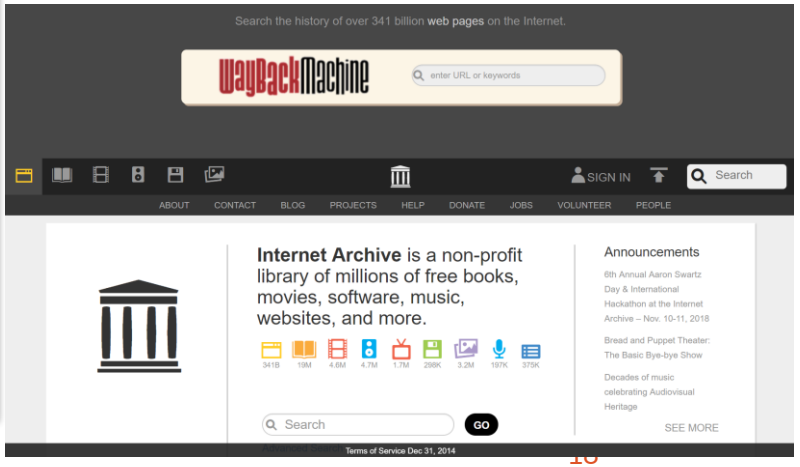

Policy
Vision and Objectives
Metadata / Description
Appraisal and Selection
Scoping
Data Capture
Metadata / Description
Storage and Organization
Quality Assurance and Analysis
Web Archive Collection
Resources and Workflow
Access / Use / Reuse
Preservation
Risk Management
Policy
Policy
Policy

Julien Masanès (Ed.)
Web Archiving
Springer


IIPC
INTERNATIONAL INTERNET PRESERVATION CONSORTIUM
netpreserve.org
HOME ABOUT IIPC WEB ARCHIVING EVENTS BLOG JOIN US
WEB ARCHIVING
Why Archive the Web?
The Internet has enabled an unprecedented era of knowledge sharing, creativity, innovation, and connection. It has also created new challenges for institutions whose mission it is to document and preserve contemporary knowledge and culture.
Many things that memory institutions collect, like scholarly publications, campaign materials, works of art, government documents, correspondence, and news, are now only available on the Web. Web pages are increasingly dynamic, they are constantly changing. To make sure this content survives for the next generation, it must be captured in real-time.
Web archiving is the process of collecting portions of the World Wide Web, preserving the collections in an archival format, and then serving the archives for access and use.
Events
Web Archiving Conference & General Assembly 2018
Web Archiving Week 2017

WIKIPEDIA
The Free Encyclopedia
List of Web archiving initiatives
From Wikipedia, the free encyclopedia
This page contains a list of Web archiving initiatives worldwide. For easier reading, the information is divided in three tables: web archiving initiatives, archived data, and access methods.
This Wikipedia page was originally generated from the results obtained for the research paper A survey on web archiving initiatives[1] published by the Arquivo.pt (the Portuguese web-archive) team.
Contents [hide]
1 Web archiving initiatives
2 Archived data
3 Access methods
4 References
5 External links
Map of Web archiving initiatives worldwide in June, 2014
Web archiving initiatives [edit]
Name
Country
Creation Year
Technologies
Number of Employees
Comments


PANDORA
AUSTRALIA'S WEB ARCHIVE
National Library of Australia and Partners
Search Trove
Browse Subjects
Home
About PANDORA
Blog
Partners
Notification form
Services
Statistics
User survey
Contact us
Other archives
Disclaimer
NLA home page
Search past Commonwealth Government websites in the Library's special web archive collection the Australian Government Web Archive.
Follow PANDORA on Twitter using @NLAPandora
Browse subjects:
Arts
Business & Economy
Defence
Education
Environment
Government & Law
Health
History
Humanities
Indigenous Australians
Industry & Technology
Media
People & Culture
Politics
Sciences
Society & Social Issues
Sports & Recreation
Tourism & Travel
View the complete listing of titles available within the PANDORA Archive or search titles alphabetically
1-9 A B C D E F G H I J K L M N O P Q R S T U V W X-Z
NATIONAL


Search the history of over 341 billion web pages on the Internet.
WayBackMachine
enter URL or keywords
SIGN IN
Search
ABOUT CONTACT BLOG PROJECTS HELP DONATE JOBS VOLUNTEER PEOPLE
Internet Archive is a non-profit library of millions of free books, movies, software, music, websites, and more.
Search
GO
Announcements
6th Annual Aaron Swartz Day & International Hackathon at the Internet Archive – Nov. 10-11, 2018
Bread and Puppet Theater: The Basic Bye-bye Show
Decades of music celebrating Audiovisual Heritage
SEE MORE
Terms of Service Dec 31, 2014

Challenges for web archives

Technical issues

- Characteristics of web resources
- Technologies

Legal issues

- Copyright issues
- Legal deposit laws

Economic environment

- High cost
- Who pay?
- Sustainability of the project

Management issues

- Long-term preservation
- Collaboration
- Acquisition

Specific challenges for the SAC's web archiving initiative

Copyrights issues

Various languages

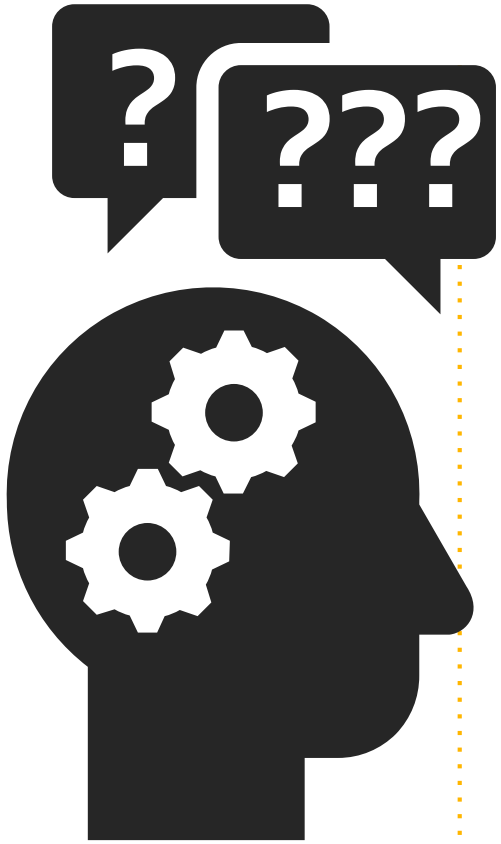
Understanding on web archive

Technology

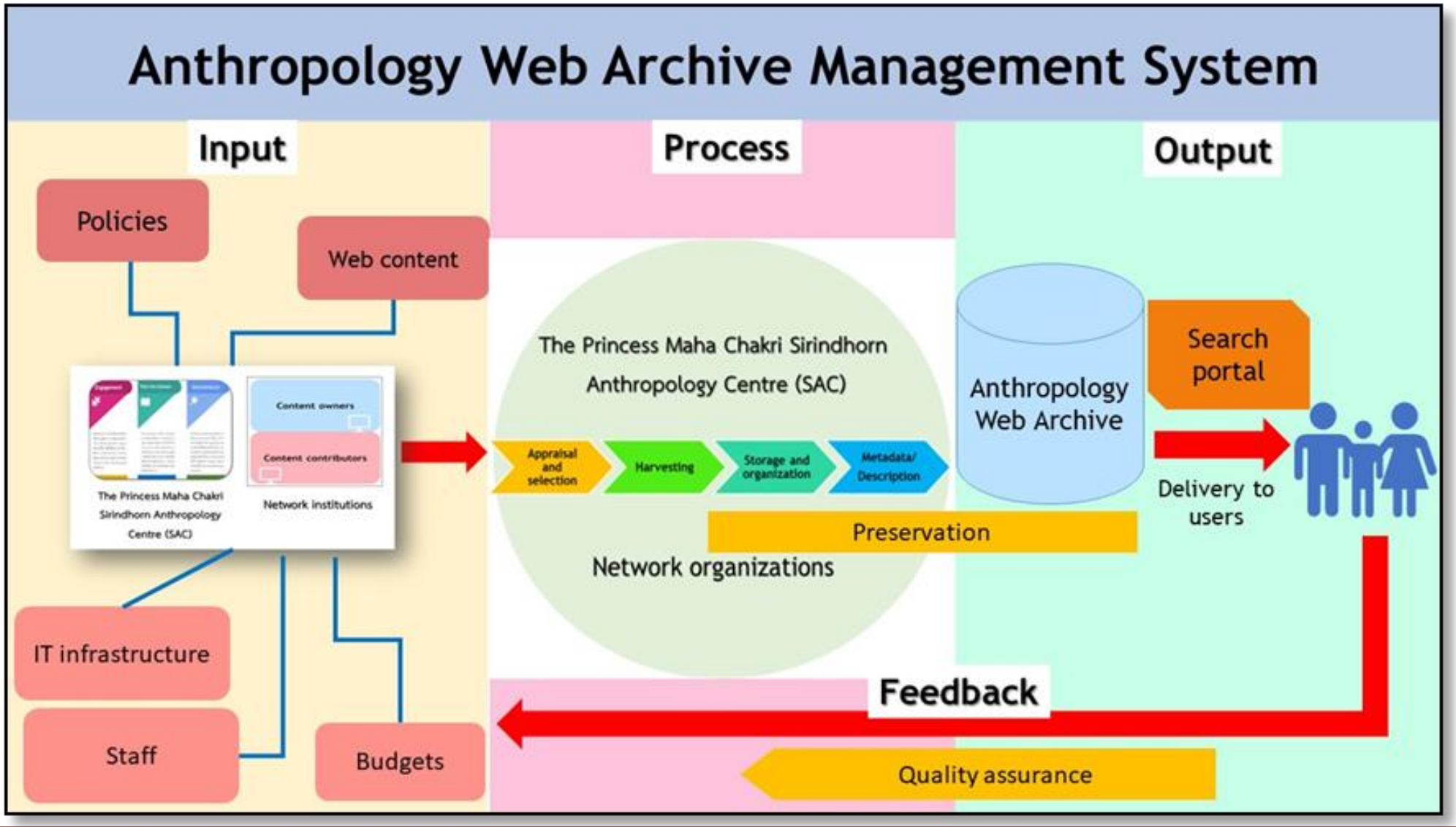
Resources

What is the next step?

1. Web content analysis
2. Literature review
3. Insights on web archiving lifecycle, components, and challenges



Proposed web archive management model for the SAC



Engagement

Communicate with network institutes about the benefits of web archive. Ask for their engagement in the project as content contributors while the SAC is responsible for hosting web archive.

Train the trainers

Upskill the SAC staff and expand their capabilities of web archiving. Provide workshops to the network institutes to gain more understanding of web archiving system, appraisal and selection, documentation, and access

Maintenance

Be the host of web archive and subsidize IT infrastructure for curating and archiving web content. Also, assure the quality of archived web content and provide searching tools as well as tutorial for the public.

The Princess Maha Chakri Sirindhorn Anthropology Centre (SAC)

Content owners

Content contributors

Network institutions

Tentative future plan

2018-2019: Implement SAC's web archiving policies

2020: Develop web archiving program and test with the SAC's digital academic resources

2021: Seek collaboration with the target groups

Q & A

THANK YOU