Bibliography and Data Bank on Tribal Communities in Andhra Pradesh

V. MALLIKARJUNA NAIK
Project Consultant

Bibliography and Data Bank on Tribal Communities in Andhra Pradesh

V. MALLIKARJUNA NAIK
Project Consultant

Contents   **Page No.**

Report on the Digital Library Project of Bibliography and Data Bank on Tribal Communities In Andhra Pradesh

1) Introduction

Tribal Cultural Research & Training Mission (TCR&TM), Tribal Welfare Dept. Visakhapatnam project to develop a framework of virtual repository of learning resources with a digital search facility. Filtered and federated searching is employed to facilitate focused searching so that learners can find out the right resource with least effort and in minimum time. TCR & TM is designed to hold content of any multi languages and provides interface support for leading tribal languages. It is being arranged to provide support for all academic levels including researchers and life-long learners, all disciplines, all popular form of access devices and differently-abled learners. It is being developed to help students to prepare for entrance and competitive examination, to enable people to learn and prepare from best practices from all over the world and to facilitate researchers to perform inter-linked exploration from multiple sources. The pilot project is devising a framework that is being scaled up with respect to content volume and diversity to serve all levels and disciplines of learners. It is being developed at Tribal Welfare Dept., Visakhapatnam.

A) What is Digital Library

A digital library, digital repository, or digital collection, is an online database of digital objects that can include text, still images, audio, video, digital documents, or other digital media formats. Objects can consist of digitized content like print or photographs, as well as originally produced digital content like word/PDF processor files or social media posts. In addition to storing content, digital libraries provide means for organizing, searching, and retrieving the content contained in the collection.

Digital libraries can vary immensely in size and scope, and can be maintained by individuals or organizations. The digital content may be stored locally, or accessed remotely via computer networks. These information retrieval systems are able to exchange information with each other through interoperability and sustainability.

2) The planning processes

When putting together a digital collection, planning is probably the most crucial step of the whole process. It sets the tone for the project flow and helps the team to focus on its goal. One of the first tasks was to decide on a cohesive topic for a digital collection. Our team decided

together on the topic for our DL (digital library). We wanted a topic that we are all familiar with and that we all can contribute to, regardless of the differences in expertise.

We also decided that a topic of interest to both professionals (such as students, research scholars) and laymen would be beneficial for this term project. The collection is therefore intended for use by passion for past and contemporary tribal history and wish to learn more about this particular art, by providing a collection of tribal (books, texts, images and videos), it is the goal of our digital library to give users a better understanding of what kind of work and people make stage and tribal films and musicals possible. The digitized texts illustrate theoretical background and information on the history of documents; images and videos provide a glimpse into artistic and technical aspects.

Our targeted for this digital library includes all those who share a particular interest in the history and background of 20th and 21st century of documents. We decided to include objects that would be to satisfy different interests within our tribal groups. We also decided to include a selection of short tribal videos from these productions to showcase the performance aspect to users, as well as articles, journal's, thesis, annual reports, tribal photos and tribal books to provide more information about each of the end user. One of the goals was to give users a sense of the process of stage data even if they had never seen our website.

3) Digitization process

After we had a plan in place, we had to go through the process of digitizing our items for the collection. All of our images and texts had to be scanned as none were born digital. they would then go over them and make corrections as necessary. We decided that we would use the same specs for each item type to achieve as much consistency as possible. We did not set a limit as to how many items each individual could have. Aside from text and still images, we had decided to include videos in our collection. We found these videos online and had to link them in workable formats for our collection. We all had access to a scanner, so we did not experience problems with the digitization process. And, we all were able to convert our videos from flash video to MPEG files so we could post them on Content DM[2]. Our digitization specs are outlined in Appendix 1.

4) Meta-data acquisition and creation

Due to the different nature of our objects, we decided to use Dublin Core as a metadata standard. Its flexibility gave us the option to decide on mandatory and optional fields as they would make

sense for our specific collection. We used the digital guidelines as basic outline and then discussed which fields we felt necessary to be mandatory and which we could leave optional.

Once this decision was made, we design the web theme and follow the digital guidelines and adapted them so that they would make sense for our collection, i.e. we decided on specifics such as terms for creator and contributors. Each subscribe then assigned metadata to their items. Some items had more metadata elements than others. We decided to publish text files as PDFs, videos as MPEGs, and still images as JPEG files. We also made the distinction between format and type when assigning metadata. For instance, a text file is labeled in the Type field as text but in the format field, it is labeled as a PDF. For subject headings, we decided to use the Library of collecting data subject headings to maintain consistency for our collection.[1] These fields were easily managed in Content DM[2]. Appendix 2 contains a list of mandatory and optional fields. Appendix 3 contains meta-data screenshots.

5) Pros and Cons of Content DM

As with any online management system there are going to be pros and cons of using that individual system. For our group we found that overall Content DM was very user friendly and served us well for our purposes. The system has a User Support Center that allows its users to view various tutorials about both setup and file creation, participate in web-based training, view a variety of resources, and download other programs. This User Support Center is excellent for both first time users and experienced users who may need additional help. This was particularly helpful for our group as all of us had different levels of expertise in regards to software.

Another advantage of using Content DM is that it allows its users to more easily and efficiently organize their digital materials. Upon uploading your files, you can add and edit your metadata very easily. The Administration function of Content DM allows you to either use the metadata fields provided or to change the fields to those you need. Users can also delete fields that are not needed and change the position of fields to organize them to suit your individual needs. We

---

[1] The digital data gathering (DDG) for M&E. However, there exist several constraints that may still play a role in the lack of its adoption. Here, the writer discusses what DDG is and the advantages of using technology for data collection. The writer also discusses some of the limitations that hinder the adoption of technology for data collection.

[2] CONTENT dm is a DAM (Digital Asset Management) service that features tools for creating, displaying, and preserving digital collections. The app is primarily used by historical organizations, museums, and libraries. Also, metadata can be shared through the WorldCat Digital Collection Gate, a premier library database, as well as via other repositories. Furthermore, master files are kept safely in the system's cloud-based archive. This is equipped with heavy security that allows for safekeeping for a long time, while still being conveniently accessible for monitoring and retrieval whenever they're needed.

made heavy use of this feature, changed and organized our fields to match our collection perfectly.

Our group found that Content DM was very efficient in how it allowed its users to download, edit, and organize their digital materials. The system was very user friendly and provided a variety of different resources for further assistance. It was easily to view your collection once it was created as well as viewing other collections that you may have access to. We experienced few downsides to using Content DM. Some group members encountered problems downloading Content DM because not all computer systems are compatible. Furthermore, the full version of Content DM must either be bought or the library or archive must upload their materials to OCLC's server.

Copyright is another point that needs to be considered when choosing a DAMS. Since all our material was published under the educational fair use, we were able to use Content DM and actually upload and publish our collection. Therefore, we would not recommend Content DM if a project includes copyrighted material and the team did not get permission to re-publish these items.

6) Encountered problems and solutions for future projects

We encountered several problems during the planning and creation of our Digital Library. In retrospect, the biggest problem was that we did not decide on a group leader. Failure to do so had an impact on all other steps that followed.

First of all, we had to deal with the fact that we live in different states, which did not give us the option to meet in person. We had to resort to email and chat sessions, and a group leader could have coordinated these meetings better so that they would not be as unorganized and chaotic. We also decided that for the next group project we will exchange phone numbers, instant messenger names etc. to stay in contact if needed as some people do not have the opportunity to constantly check their email.

We experienced that due to different levels of tech expertise, we had problems getting points across via email or while chatting. Oftentimes one group member would not quite understand what another was trying to convey simply because of lack of technical knowledge or lingo. This problem can be solved by having a preliminary meeting before the project starts to determine where the group members stand in respect to technical and other necessary knowledge.

The group leader can also facilitate as moderator in chat sessions, assign functions and responsibilities to group members and act as point of contact in case questions or problems arise. This would have made the work much more efficient.

We believe that, whenever possible, groups members should agree on who is responsible for what step. In our case it was difficult to agree on one person doing the digitization because we do not all live in the same town. However, if team members are all local, we strongly believe that the group leader should name one person responsible for the digitization, another one for meta-data creation, a third one to designing the DL web site and interface, and yet another to prepare a documentation etc. This way each member would feel comfortable because, ideally, he/she works within his/her field of expertise.

The main thing that caused our group a significant amount of trouble was the lack of communication, or rather the courage to mention to the group members how we feel about the process, work done etc. One member, who felt that she was putting in more effort than the other members, did not mention this right away, which caused irritation and dissatisfaction. We believe that having named a group leader in the beginning would have eliminated this problem right away. Members can talk to the leader who they convey the message to the rest of the group and, if necessary, makes a decision in regards to workload, assignments etc. It is most important that all members agree in the beginning to accept the leader's decision once it has been made.

Finally, we believe that we should have set up an internal schedule that dictates which work has to be completed at what date. The advantage is that this would have given us some time to compare our results and not have the feeling of last-minute panic.

In conclusion, our problems were mainly human ones. We did not experience any problems with the technology as we managed to help each other out if questions arose. We learned from this experience that group work requires a lot of effort and preliminary thinking and work before starting the actual project. We also think that it is vital for the group to get to know each other ... this will build stronger ties among the team members and gives them the courage to voice doubts, critique etc.

7) Conclusion

Throughout this class use a lot about digital libraries. We gained experience in planning a project, selecting and digitizing objects, creating meta-data and finally piece it all together. We found that creating the meta-data was one of the most interesting parts of this project. Each of us had to face her own problems, and we realize now, at the end of the some of these problems

could have been eliminated by proper planning. Although we were frustrated at times, we take this as what it should be: a learning experience we can use to improve our next group project and ensure communication and success right at the beginning. Despite our set-backs we think that we worked well together and created a good, interesting and informative Digital Library.

8) Appendixes

a) Appendix 1: Digitization specs

Images

We chose to scan these at this rate in order to have a manageable file size for our end users, while still maintaining the image quality of the original.

Resolution: 300 x 300 dpi and 1600 x 850 dpi

Depth: 24bit color

Hardware: Epson Expression 10000 XL scanner / HP Photosmart C4180 All-in-One / Epson Perfection V700

Software: Adobe Photoshop CS / Gimp

File format: JPEG

Text

Both grayscale and black and white imaging was used to digitize our text objects. For the sheet music, after testing, the grayscale gave better reproduction quality.

Resolution: 300 x 300 dpi

Depth: 8 bit grayscale

Hardware: Epson Expression 10000 XL scanner / HP Photosmart C4180 All-in-One

Software: Adobe Print Shop Pro, also collected and OCRed with Adobe Acrobat 8.0

File format: PDF

Video

These were link downloaded from YouTube and then converted to usable formats. No segmentation was necessary.

Software: OnlineVideoConverter.com to convert flash video to MPEG file, Windows Movie Maker to clean up video quality / FVL to MPEG Converter for Ipod .5.3

File type: mpeg

Variable bit rate

Resolution: 320 x 240 pixels, aspect: 4:3, 30 frames/sec

b) Appendix 2: Meta-data fields

Descriptive meta-data

Mandatory fields

Title

Unique title for each record

AACR2 rules for capitalization punctuation etc.

Remove leading articles (the, a, etc.)

If no title, group member who digitalized item will supply title; in this case do not put title in brackets.

Subjects

All subjects in one field, separated by semicolon and space

Use LC subject headings, or local (SC) when needed

Add terms when appropriate. Decision will be made by individual member.

Publisher

Each record will only have one publisher: University of South Carolina

Date Digital

One Date Digital per record

Date Digital = Date when item was digitized

Follow ISO 8601 standard of YYYY-MM-DD

Type

One type per record

Follow DCMI Type vocabulary: Still Image, Moving Image, Sound, and Text etc.

Format

Only one Format per record

Consolidate Internet Media type (IMT) and subtype into one string

Add file size and content extent when applicable, e.g. image/jpg 2 MB (for a digital photograph) or video/mpg 10 GB ; 34 min., 20 sec. (for digital video)

Do not use bytes. Use KB, MB, GB etc

Source

Enter all sources in one fields, separated by semicolon and space

Use for free-text account of the source of the original item

Include publisher of original item here, plus ISBN or any other original necessary information

Maintain detail from MARC catalog record when possible

Can include URL or finding aid or MARC record in USC's catalog

Description

Enter each Description in one field.

Use for free-text account of intellectual content of the original item, and physical description of original item. Can be quoted or composed my metadata creator.

Is searchable, so think of audience and terminology they would use

Use both Description. Abstract and Description. Table of Contents if applicable. First is

composed by metadata creator, second is found on item.

Website
URL for collection home page	:	aptcrtm.in
Support Us	:	support@aptcrtm.in
Info	:	info@aptcrtm.in

Optional fields

Title. Alternative

Use for subtitles etc.

Enter each Title. Alternative in one field, separated by a semicolon and one space

Follow AACR2 rules for capitalization, punctuation etc.

Remove leading articles (the, a, etc.)

Creator

Enter each Creator in one field, separated by a semicolon and one space

Follow AACR2 rules for capitalization, punctuation etc.

Use LC Authority File for form of corporate and personal names

Enter names as follows: last name, first name. Enter corporate names in full direct form

If no LC established heading exists, group makes decision on name

Can include multiple names or corporate bodies, but use Contributor field for those who played a secondary role in the intellectual content of the item

Can include role of creator, when needed, in parentheses after name. Text creator = author, picture creator = photographer, video creator = director

Contributor

Enter each Contributor in one field, separated by a semicolon and one space

Follow AACR2 rules for capitalization, punctuation etc.

Use LC Authority File for form of corporate and personal names

Use this field for those who played a secondary role in the intellectual content of the item

Can include role of contributor, when needed, in parentheses after name. Text creator = author, picture creator = photographer, video creator = director

Contributing Institution

Only one: Andhra Pradesh TCR & TM (Digital Library and Information)

Language

Enter all Languages in one field

Spell out languages

Separate languages by semicolon and one space

Most prominent language listed first

Relation

Enter each Relation in one field, separated by semicolon and space

Use refinements to specify relations, when necessary

Always include title of Digital Collection in a Relation. Is Part of field

Digitization Specifications

Only one Digitization Specification per record

Include: bit depth, resolution, separate by semicolon and space

Date. Original

Only one Date Original per record

Follow ISO 8601 standard of YYYY-MM-DD. Only use what part of the date is known

Use dates associated with creation/publication of original item

If missing date, use n.d.

If approximate date known, use circa YYYY

For range of dates, put space-dash-space between years.

Resource Identifier

Only one Resource Identifier per record

Use for file name (i.e. phant_01.jpg)

Rights Management

Only one Rights Management per record: Educational Fair Use

Administrative meta-data (optional)

Collection Name
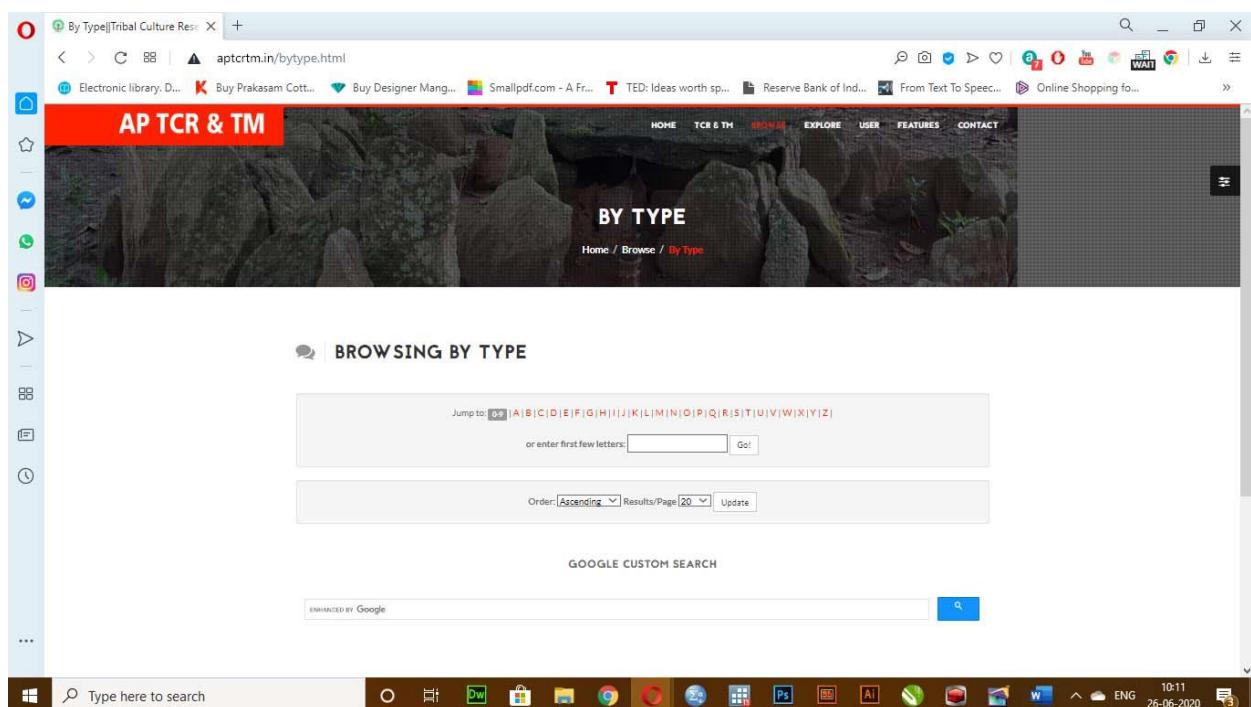
Description of Collection

Number of items

Metadata person

Scanner and software

Date Project Start

c) Appendix 3: Meta-data screenshots

d) Appendix 3: DL screenshot (public user-interface)

This part of the collection can be found at aptcrtm.in