READIT 2018
11th National Conference On Recent Advances in Information Technology

REINVENTING LIBRARIES:
DIGITAL INNOVATIONS & TECHNOLOGIES

8-9 AUGUST 2018
ICGAR, Kalpakkam
ONE LIBRARY. ALL OF INDIA.

1.7CR+ Content II 200+ Languages II 170+ Sources II 30Lakh+ Users
Prof. Partha Pratim Das
ppd@cse.iitkgp.ac.in

Professor, Department of Computer Science & Engineering,
Head, Rajendra Mishra School of Engineering Entrepreneurship,
Professor-in-Charge, IIT Kharagpur Research Park, Kolkata,
Joint Principal Investigator, National Digital Library of India Project
Indian Institute of Technology Kharagpur
NDLI VISION

Build up National Digital Library of India as a National Knowledge and Cultural Asset:

The key driving force for Education, Research, Cultural heritage, Innovation, and knowledge-sharing in India
“IT IS HIGH TIME THAT INDIA SPEAKS FOR ITSELF, TO HIGHLIGHT, OFFER AND SHARE ITS OWN CULTURAL, SPIRITUAL, ACADEMIC AND SCIENTIFIC HERITAGE ON ITS OWN TERMS”

NDLI MISSION

1. To create a 24X7-enabled integrated ubiquitous digital knowledge source
2. To protect and preserve India’s cultural, academic and scientific heritage
"EDUCATION IS THE MANIFESTATION OF THE PERFECTION ALREADY IN MAN"

- Swami Vivekananda
  Teacher, Philosopher, Visionary

HAPPY YOUTH DAY

NDLI MOTTO

1. OPEN
2. INCLUSIVE

WINNER mBillionth South Asia Award 2017: in Learning and Education Category for Android Mobile App
ABOUT NDLI
ISSUES, ARCHITECTURE AND USE MODELS

“Google can bring you back 100,000 answers. A librarian can bring you back the right one.”

- Neil Gaiman
NDLI ISSUES

USER-SIDE

➢ Wide geographic expanse & Large population
➢ Huge number of students
➢ Large number of institutions
➢ Varied linguistic diversity
➢ Severe lack of Teachers

PROVIDER-SIDE

➢ Wealth of digital content
  ● Books and Articles
  ● ETD
  ● Question Papers and Solutions
  ● Video Lectures - MOOCs
  ● Simulations & Animations
  ● NMEICT Projects
  ● Data
  ● ...
➢ No single-window search
➢ Google search uses keyword – no metadata search
➢ Widely varied DL technology
➢ Lack of Interactivity, Vernacular support
➢ Low integration between content and learning system
➢ Weak ecosystem between learners and teachers
USE MODELS

NDLI REPOSITORY

Content Access

Content Contribution

Web Content Harvesting

Institutional Digital Repositories

Web Content

Learning Management System

Educators

Contributing Institutions

Physically Challenged

Researchers

Learners

Professionals
A content is included (metadata ingested) in NDLI if it is expected to have educational value.

**CONTENT AT NDLI**

- Born-digital object
- Digital surrogate of a physical object
- Digital metadata of physical object

**METADATA AT NDLI**

- NDLI does not store contents
- NDLI only ingests metadata for Search & Browse
- Content (Full-text) is delivered from Source
Institutional Digital Repository of Contributing Institutes

Institutions of School & Higher Education, Boards

RANGE OF CONTENTS

Faculty Publications, ETD (Electronic Thesis & Dissertation): DSc-PhD-Masters-Undergrad, Research Projects

Books & Periodicals, Open Access Journals, E-Books & Subscribed E-Resource

Annual Reports, Project Reports, Convocation, Working Papers, Others

Encyclopaedia, Dictionaries Directories Others

Lecture Slides, Videos, Class Notes, Courseware

Term Papers, Assignments, Solutions

Lab Experiments, Manuals, Case Studies

Question Banks (JEE / GATE / NET / CAT), Model Answers

Datasets, Benchmarks, Models, Maps, Software

Manuscripts, Painting, Sculpture, Music, Dance, Drama

Audio & Video Content
CONTENT VIEW ARCHITECTURE

Vertical-Specific Custom Interface and Search

- School Vertical
- Domain Vertical (Medical/Legal/...)
- Competitive Exam Vertical
- Data Vertical
- Application Vertical

Content Baseline

Search

Generic Interface and Search

Browse
# STAKEHOLDERS

<table>
<thead>
<tr>
<th>Stakeholder</th>
<th>Roles and Responsibility</th>
</tr>
</thead>
</table>
| **Government** | 1. Sponsor and facilitator  
                  2. Content Contributor  
                  ● Ministries / Departments  
                  ● R & D Labs                  |
| **Institutions** | 1. Host Institution – IIT Kharagpur  
                 2. Contributing Institution – Supporting IDRs  
                 3. Participating Institution – Providing Users & Feedback |
| Public       | 1. Use and Feedback  
               2. Metadata by Crowd Sourcing  
               3. Content by Crowd Sourcing |
| Industry     | 1. Technology Providers                                       |
| Publishers   | 1. Metadata Provider  
               2. Content Provider (under various licensing schemes) |
CONTRIBUTORS:

CFTI, State and Central Universities, R & D Labs, Govt. Depts, Free Portals, Publishers, etc.
WHO IS THIS FOR?

The NDLI platform is for ALL learners.

SCHOOL

COLLEGE

PROFESSIONALS

LIFELONG LEARNERS
NDLI WEBSITE: Landing page
NDLI WEBSITE: Browse
NDLI WEBSITE: Search Result page

Magnet and Compass - Magnetism, Magnetic Field, Magnets
Source: PhET Interactive Simulations
Classroom | Homework
Abstract: Ever wonder how a compass worked to point you to the Arctic? Explore the interactions between a compass and bar magnet, and then add the earth and find the surprising answer! Vary the magnet's strength View more

Magnets and Electromagnets - Magnetism, Magnetic Field, Electromagnets
Source: PhET Interactive Simulations
Classroom | Homework
Abstract: Explore the interactions between a compass and bar magnet. Discover how you can use a battery and wire to make a magnet! Can you make it a stronger magnet? Can you make the magnetic field reverse?

Magnetism without magnets
Source: IEEE Xplore Digital Library
Author: Ryder, J.D.
UG and PG
Abstract: The earliest discoveries of electromagnetism are briefly recounted. The work of Volta (primary battery), Oersted (generating a magnetic field by a current) and Ampere (solenoid) is described.
MULTI-LINGUAL INTERFACE: BROWSE IN PREFERRED LANGUAGE
METADATA FUNDAMENTALS

Flash-course on data aggregation!
WHAT IS METADATA?

- Metadata is primarily textual information relating to Content.
- It includes information that enables users to identify, discover, search, browse, interpret, or manage Content.
- Includes hyperlinks that direct users to Content on the Source.
- May include an expressive description of the Content.

NOTE: NDLI does not store contents, it only ingests metadata for Search & Browse. The content (Full-text) is delivered from Source.
**METADATA STANDARD**

- Defines common understanding of the semantics of the data
- Ensures correct and proper use and interpretation of the data
- Defined through a set of fields, vocabs (optional), and instructions for fill-up

- Metadata = Data about Data
- Metadata Standard =
  - Content + Intent
  - Content + Context
- Standards proliferate
NDLI METADATA DESIGN CHALLENGES

1. WIDE CATEGORY OF RESOURCES
   Generic metadata or domain specific?

2. OPENNESS OF REPOSITORY
   Closed metadata standard may fail to describe a new resource

3. SCALE IS ENORMOUS
   - Manual annotation is infeasible
   - Automatic annotation guided by crowdsourcing?
NDLI METADATA REQUIREMENT SPECS

To describe any digital resource
- Generic content metadata
  Contributor, Description, Language, Format etc.

To describe domain specific resources
- Educational content metadata
  Educational level, ToC, Type of learning material etc.
- Thesis metadata
  Institution, advisor, degree, researcher
TRANSLATION ISSUES

➢ Variation in Subject classification standards
  ● Dewey Decimal Classification (DDC)
  ● Library of Congress Classification (LCC)
  ● Library of Congress Subject Headings (LCSH)

➢ Mapping terminology for different languages
  ● Translate when equivalent terminology is present
  ● Transliterate otherwise
CHALLENGES IN A MODERN DIGITAL LIBRARY

- How to provide contemporary service – PEOPLE (provider)
- How to build expectations – PEOPLE (user)
- How to protect and preserve contents – IP, MAINTENANCE
- How to provide access to rich contents – MONEY, BUSINESS, OA, RIGHTS
- How to offer best experience with least resources – BEHAVIOR, SEMANTIC SEARCH
- How to stay ahead in competition, survive on expectations – ANALYTICS, STRATEGY
- How to remain sustainable – BUY, BORROW, STEAL? – BUSINESS MODEL
- How to push the horizon (beyond Search & Browse) – COMPUTE, LEARN, TUTOR
- How to ...?
NDLI INITIATIVES & INNOVATIONS
SKILL DEVELOPMENT AND TECH INTEGRATION
LEADERSHIP: IDR WORKSHOPS
PAN-INDIA IDR* WORKSHOPS
*INSTITUTIONAL DIGITAL REPOSITORY

EVENTS: 25
INSTITUTES: 977
INDIVIDUALS: 1576
ONE PLATFORM, ALL OF INDIA
COMMUNITY INNOVATION: NDLI CLUBS
VIRTUAL LIBRARIES ARE MORE THAN A PASSIVE COLLECTION OF BOOKS...

- Like with all other things digital, users expect a lot more customization and interaction from digital libraries - the digital library hence needs to evolve into a bouquet of services instead of just a storage platform.

- Apart from collating data and connecting them meaningfully, the end-success of the library will always lie with its users. NDLI clubs encourages the active involvement of users with the content as well as each other to Learn, Share, Grow.
NDLI CLUBS

- E-learning environment for schools
- Involving all K-12 Stakeholders
- Single point of access for unified academic resources
- Aimed at building active learners among users - NDLI club newsletters, quiz, workshops on themes like STEM, e-commerce, IoT etc.
Clubs can provide insights on learning outcomes and cognitive abilities to better curriculum
Will allow development of concept-based questions modelled on learning behaviour
Student profiles and performance will allow school management to make better decisions
ELEPHANT IN THE ROOM: COPYRIGHTS

- WHAT CAN I SHARE?
- WHAT CAN I REPRODUCE?
- HOW DO I KNOW IF THE CONTENT IS ORIGINAL?
- IS THIS CITABLE?
- HOW CREDIBLE IS THIS INFORMATION?
OPERATIONS LEADERSHIP: INTELLECTUAL PROPERTY
Copyright is a bundle of rights given by the law to:

- Creators of literary, dramatic, musical and artistic works and
- Producers of cinematograph films and sound recordings

Rights are:

- reproduction of the work,
- communication of the work to the public,
- adaptation of the work and,
- translation of the work

Creator / Author is the first Owner

Ownership (Title) can be transferred (assigned)

Copyright Laws are subject to time period and geographical location and the subsequent IP laws of the land
NDLI’S APPROACH TO COPYRIGHTS

- NDLI strives to provide India with its own geographical definition of copyrights at par with international standards
- Copyright Manual for Librarian
- National debate on copyright & OA issues
- Recommendations to Govt. for improved legislation
RightsStatements.org provides 12 standardized rights statements for online cultural heritage.

Our rights statements make it easy to see if and how online cultural heritage works can be reused. Find out more about our statements here.
TECHNOLOGY INNOVATION: METADATA EXTRACTION
METADATA ENVELOPE

DUBLIN CORE (Generic)
NDLI METADATA
LRMI (Educational)
SHODHGANGA (Thesis)

Type
Title
Subject
Language
Description
Date
Author

Board
Difficulty Level
Typical LT
Prerequisite Topic
Type of LM
Pedagogic Objective
Educational Level

Researcher
Keyword
Advisor
Place
Department
Institution
Awarded
Degree
ACQUISITION SCENARIOS

LOCATE CONTENT
- Harvest Institutional IDRs
- Crawl Websites
- In Bulk – from Publishers
- Donated by Source
- Source-supported API
...

ACQUIRE METADATA
- Creation
  - Manual
  - Automated
- Translation
  - Format
  - Standard / Schema

Curation
- Manual
- Assisted
- Ingestion
CHALLENGES IN METADATA ACQUISITION

➢ Different sources follow different norm in annotation
  ● Automation of curation

➢ Errors in sourced data
  ● Error listing
  ● Manual curation

➢ Errors in manual annotation
  ● Review step before actual submission in database
  ● Manual curation

➢ Set up norm/guideline
  ● Controlled vocabulary
NDLI Metadata Standard v 1.0 is an Open Virtual Standard


Schema is categorized into three profiles:

1. **Generic Metadata:**
   - Describes general attributes of the contents
   - Adopted from Dublin Core Metadata Standard

2. **Educational Metadata:**
   - Describes the educational attributes of the resources and helps in enumerating properties of the contents relevant to teaching-learning process
   - Adopted from Learning Resource Metadata Initiative (LRMI)

3. **Thesis Metadata:**
   - Describes dissertation or thesis related metadata fields
   - Adopted from Shodhganga Thesis Metadata Standard

More profiles may be added in future

Uses the **namespace of Dublin Core (dc.)**
NDLI METADATA EXTRACTION TOOLKIT

- Automated metadata extraction workflow
- Syntactic metadata extractor
  - Author name, ISBN, Publisher, dates etc.
- Table of content extractor
- Wikification for keyword extraction
- Learning resource metadata
- Toolchain for metadata extraction

Once completed and tested, the toolkit will be released to public and open-sourced
Workflow
NDLI EXPERIENCE TRACKING

Experience Tracking
- To offer customized search results

Multi-lingual Support
- Reduce cognitive load for native use

Personalization
- Customized UI to suit user grade
EXPERIENCE TRACKING TECHNOLOGY

Record:

1. Partha searched ‘tiger’
2. Partha navigated to Tiger Wiki
3. Partha studied Wiki (3 min)
4. Partha downloaded tiger image (4 images)
5. Partha checked tiger map (2 min)
6. Partha enlarged map at Sunderbans (twice)
7. Partha searched ‘national animal of India’

Infer:
Partha learnt – Sunderbans is a tiger reserve, tiger is the national animal (of India), etc.
... possibly

Experience API (xAPI) / Tin Can API

Connects learning content and learning systems to record and track all types of learning experiences

Learning Record Store (LRS)

Stores learning experiences
INNOVATION IN KNOWLEDGE DISSEMINATION: NATIONAL LICENSING
NATIONAL LICENSING

An initiative by eSS:

- e-ShodhSindhu: Consortium for Higher Education Electronic Resources
  Licensing being negotiated with Publishers
- Institutional License:
  Accessible from within designated institutional network: IP filtered
- NDLI Supported License:
  Accessible if requested from NDLI: Concurrent use-basis
ACCESS AND DISCOVERY INNOVATION: SURROGATOR
- Smart algorithms to identify ‘surrogates’ of a content
- Makes available pre-print copies of academic resources that are otherwise unavailable to learners
- A customizable application that can identify surrogates for access-restricted publications in digital libraries
- Surrogator can make a case for the evolution of specific subject matters through pre-print indexing and open up a whole new realm of research
- Application is being tested through NDLI, can easily be extended to other digital libraries
CATEGORY INNOVATION: CHAVI
DATA-DRIVEN CANCER RESEARCH

- India currently does not have the infrastructure for dedicated bio banks to aid medical research.
- This prevents the Indian medical system to learn of and from trends to prepare for subsequent treatment.
- NDLI has tied up with TATA MEDICAL CENTRE in Kolkata to develop a prototype for bio-banks that will aid cancer research.
CHAVI

COMPREHENSIVE DIGITAL ARCHIVE OF CANCER IMAGING

- Dedicated image banking service
- Will provide annotated images with associated information of cancer patients
- Furthering India’s data banking initiatives
- Library migrating to create a database for bio-banks
AI INNOVATION: CREATING KNOWLEDGE
UNDERSTANDING THE MODERN KNOWLEDGE SEEKER:

● Technology and hence technology-enabled access have led to users becoming virtualized

● As a data aggregator, digital libraries can not only facilitate search and browse, but also bring together content libraries and algorithm libraries to compute big-data solutions remotely and real-time!

● So not only can the library store knowledge for its users, it can also evolve to create/compute knowledge by cross-linking programs with data-sets
THE FUTURE OF REMOTE COMPUTING

NDLI Cloud Computing
- Match AI data sets Library with Software Library Library in NDLI to compute meaningful results
IN SUMMARY

RIGHTS STATEMENTS

CHAVI

AI VERTICAL

NDLI CLUBS

SURROGATOR

RS

AI

Cl

Ch

Su
SO HOW DOES ATOMIC RESEARCH INTEGRATE WITH NDLI?

- Digitization and dissemination of content on atomic research, benefits as well as hazards
- Use the “compute” model for atomic events
- NDLI Atomic Research Club - forum for community research, collaboration, mentorship, identifying budding talent
- You tell me!
Don't limit a child to your own learning, for he was born in another time.

-Rabindranath Tagore
Together, let us

#LearnShareGrow

Use your e-mail address to sign up at: https://ndl.iitkgp.ac.in / www.ndli.gov.in