Institutional Repository using DSpace and Cloud Hosting

YATRIK PATEL/SWAPNIL PATEL INFLIBNET CENTRE

What is Institutional Repository?

"Institutional repositories (IRs) are a development in managing digital objects for effective utilization. IR establishment is a challenge as well as an opportunity for information professionals. It may include a variety of research output of any organization. An IR is a means to ensure that the published work of scholars is available to the academic community even after increases in subscription fees or budget cuts within libraries."

(Source :Wikipedia)

Institutional repositories [are] ... digital collections capturing and preserving the intellectual output of a single or multi-university community. The contents mainly created by institutes members.

It is most essentially for long term preservation.

Institutional Repositories are.

Centred around a university (or academic institution) and contain items which are the **scholarly output of that institution**

- A collection of (digital) objects, in a variety of formats
- Include works of various degrees of scholarly authority and from various stages in the process of scholarly inquiry.
- In addition to published works, an IR may include preprints, theses & dissertations, images, data sets, working papers, course materials, or anything else a contributor deposits
- Typically motivated by a commitment to open access

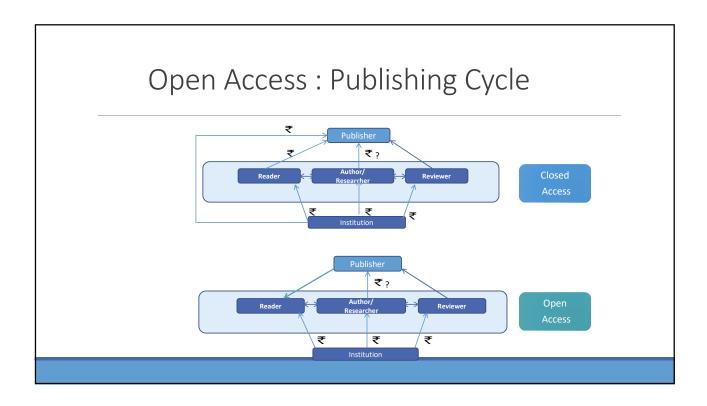
Open Access : Philosophy

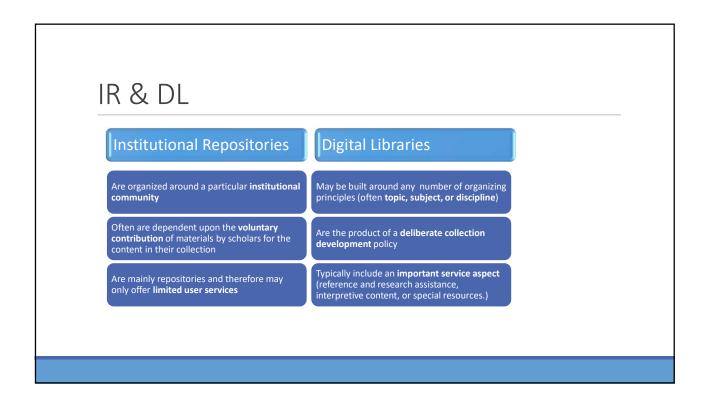
In India Government Funds for approximately 75% of education and 95% of research.

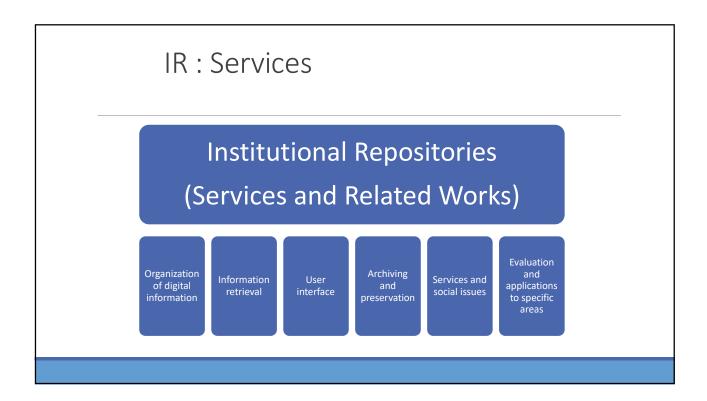
The results of research funded by government should ideally be accessible in <u>open access</u> to the society at large.

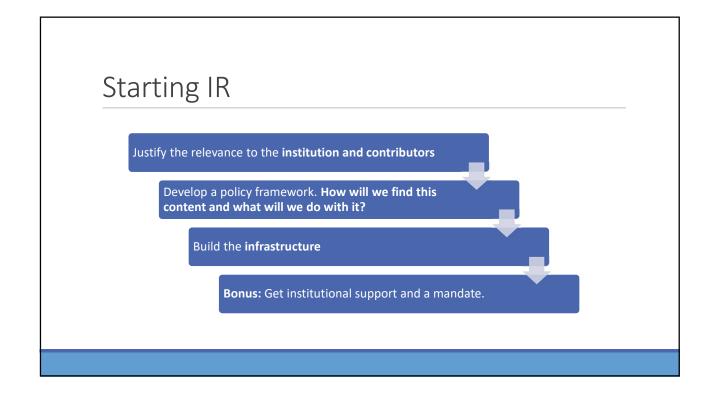
Articles in open access have greater impact in terms of citation rates than those published in subscription-based journals;

Funding agencies like NIH in US & Welcome Trust in UK have already mandated submission of results of research funded by them in open access journals or make them available in open access .









Justify the purpose

Justify the relevance to institution and contributors The justification for a repository must be made to the institution that will own and sustain it.

It is critical to work out a case which is in line with the priorities of the institution.

This means provision of tool to increase visibility, usage and impact of the research output of an institution.

The MIS data derived from a repository will also helpful to present the case before assessment bodies like NAAC

Define the purpose

The primary purpose of any repository is to provide open access to research outputs.

However, it can be also used for encouraging digital publishing initiatives.

The digital preservation can also be a purpose.

Repository services should be developed with a clear idea of the purpose of the repository.

The most successful repository collections are the ones that support the needs of the community

Advantages IR

Institutional repositories collects and curates digital outputs

Institutional repositories opens up the research outputs of the university or institution to the world, along with intensification of the visibility and impact of these outputs as a result

Publicly accessible institutional repositories showcases the university to interested communities – prospective staff, prospective students and other stakeholders

Manages and measures research and academic activities

Advantages IR

Provides a workflow system for collaborative or large-scale projects

Enables and encourages interdisciplinary approaches to research

Facilitates the development and sharing of digital teaching materials and aids

Supports student endeavours, providing access to theses and dissertations and a location for the development of e-portfolio

IR Service Expectations

What is the service's mission?

What kinds of content will you accept?

Who are the key users?

Who are the key stakeholders?

What responsibilities will the library bear versus the content community?

What are your top service priorities?

What are the short-term priorities and long-term priorities?

IR: Policies

Collection

- What types of contents can be submitted to the repository?
- Who will be able to submit in the repository?
- Criteria for determining a collection in the repository. Who regulates, sets, and authorises membership?
- What will be the structure of repository around individual contributors, or by department, research division, etc.?
- How the content will be deposited? (mediated deposit or by contributor)

IR: Policies

Management

- General rights and responsibilities of libraries and those who create collections of digital content.
- The type of metadata to be used.
- Curation and preservation tasks. Access
- Privacy policy for registered users of the system

Access

- Privacy policy for registered users of the system.
- · Possibility of restricted access to content based on request of contributor
- Possibility of providing embargo periods for content

IR: Manpower Requirements

Repository Manager-

 manages the 'human' side of the repository including content policies, advocacy, user training and a liaison with a wide range of institutional departments and external contacts.

Repository Administrator-

 manages the technical implementation, customisation and management of repository software, manages metadata fields and quality, creates usage reports and tracks the preservation issues

Advocacy: Promoting your IR

Profiling:

• This is related to positive branding of repository, it involves the use of brochures, newsletters and web sites that discuss the benefits of repositories.

Pull Approach:

• In this approach the contributor is encouraged by reward for depositing the work. There can be specific incentives for researchers who deposit.

Push Approach:

• The contributor can be demonstrated the positive effects of the repository after submission of content. For example usage statistics for authors whose downloads are very high can be displayed.

Consultation Approach:

- This involves direct communication and consultation with faculty to involve contributors in developing the repository to meet their needs.
- This can be done through surveys, meetings, informal conversations. This can be most effective approach as contributors can campaign their peers (other potential contributors) about the value of the repository.

IR – Key Issues

Faculty buy-in

Submission polices

Intellectual Property issues

Mediated deposit

Metadata

OAI-PMH compliant systems

Specialized staff

Outreach and Liaison services



Cost Effective (Hardware, Software and Maintenance)

Technically simple to install and manage

Robust

Scalable

Open and inter-operable

Modular

User Friendly

Multi-user (Both Aspects User and Administration)

Platform independent

Capable of handling multimedia digital objects

Widely Used Systems



Produced by Berkeley Electronic Press (bepress), focused on maintaining scholarly output. Not open source.



Developed at the University of Southampton (UK). Widely considered to be the least complex of the major repository software platforms.



Developed at Cornell and University of Virginia. Based on a framework known as the Flexible Extensible Digital Object and Repository Framework.



Designed by MIT and Hewlett-Packard to manage the intellectual output of research institutions and provide for long-term preservation.

What is DSpace

DSpace is a platform that

- capture items in any format in text, video, audio, and data.
- It distributes it over the web.
- It indexes digital items, so users can search and retrieve items.
- preserves digital content over the long term.

DSpace is typically used as an institutional repository or digital library. It has three main roles:

- Facilitate the *capture* and *ingest* of materials, including *metadata* about the materials
- Facilitate easy access to the materials, both by listing and searching
- Facilitate the *long term preservation* of the materials

DSpace is a joint project of MIT Libraries and Hewlett-Packard Labs. It is being handled by DuraSpace which is recently merged with LYRASIS.

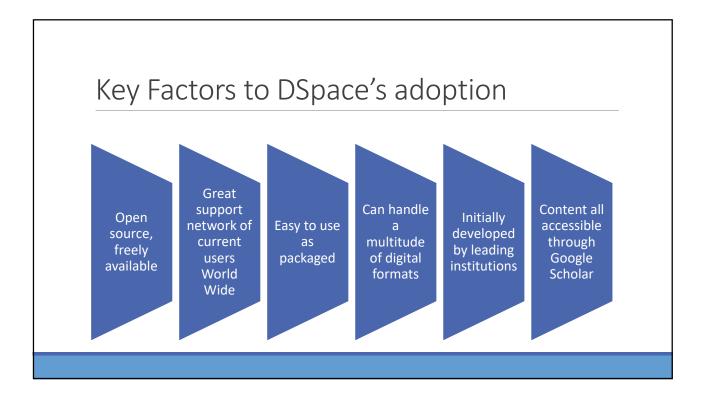
Dspace History

The beginning: 2000

• The DSpace project was initiated in July 2000 as part of the HP-MIT alliance.

Software releases:

- Version 1.0 8th November 2002
- Version 1.1 8th May 2003
- Version 1.2 13th August 2004
- Version 1.3 3rd August 2005 • Version 1.4 – 26th July 2006
- Version 1.5 25th March 2008
- Version 1.6 2nd March 2010
- Version 1.7 17th December 2010 (End of Life JAN 2014)
- Version 1.8 4th November 2011
- Version 3.0 30th November 2012
- Version 3.1 30 Jan 2013
- Version 3.2 -24 July 2013
- Version 4.0 16 December 2013 >> 4.1 on 3rd March 2014
- Version 5.0 January 2015
 Version 6.0 October 2016 >> 6.3 is the latest stable version and DSpace 7 is under development.





DSpace Development Model

Open source software (www.dspace.org)

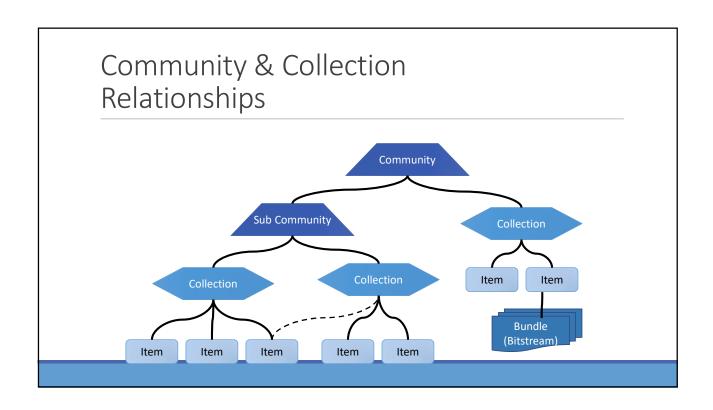
• BSD licence

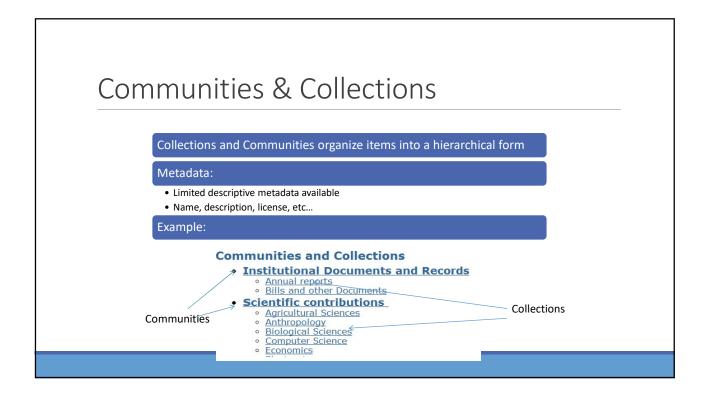
Community development model

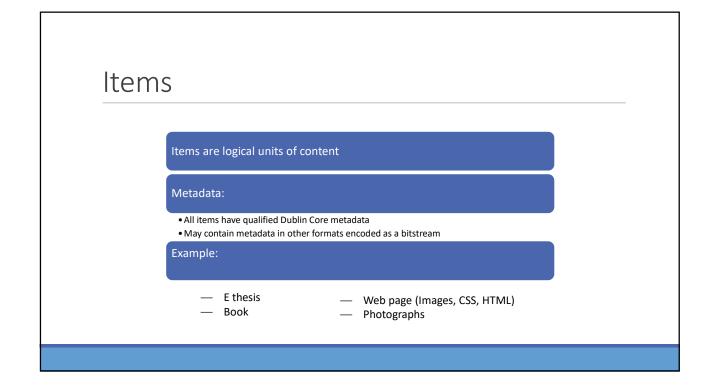
- Source code control repository (SVN)
- Committers
- Community welcome to submit bug reports, patches, feature requests
- Email lists for support

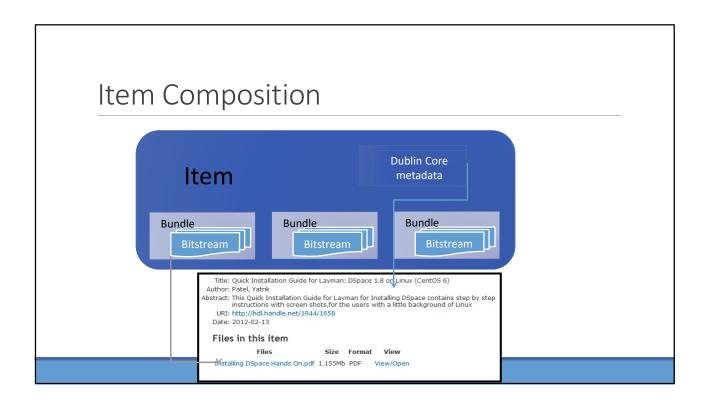
The DSpace Information Model

DSpace Information Model Communities • Research units of the organization Collections (in communities) • Distinct groupings of like items Items (in collections) • Logical content objects • Receive persistent identifier Bitstreams (in items) • Individual files • Receive preservation treatment









Item Metadata

Descriptive

- Qualified Dublin Core
- Non Dublin Core is also supported
- Any other format may be added as a bitstream
 - However, it will not be searchable

Administrative

- Who can access, remove, or modify an item
- Stored in the database, no standard format used

Structural

- Very basic
- What bitstreams are contained in an item
- What collections and communities does an item belong too

Bitstreams

Bitstreams are Individual Digital files

Metadata:

- Limited descriptive metadata available
- name, file format, size, etc...

Example:

- PDF file— Executable program
- Word document
 JPEG picture
 HTML file
 CSS file

Bundles

Bundles group related bitstreams together

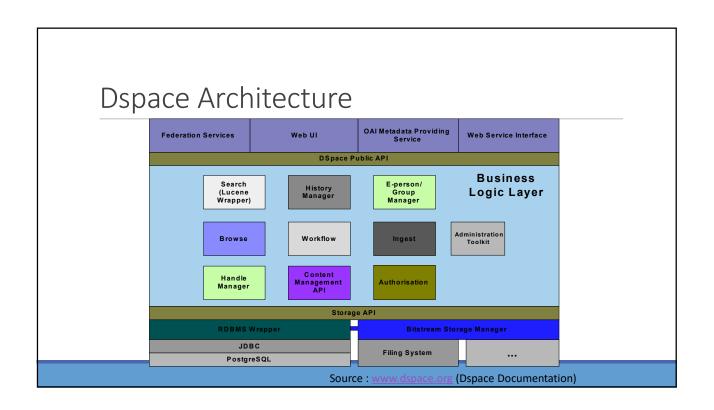
Metadata:

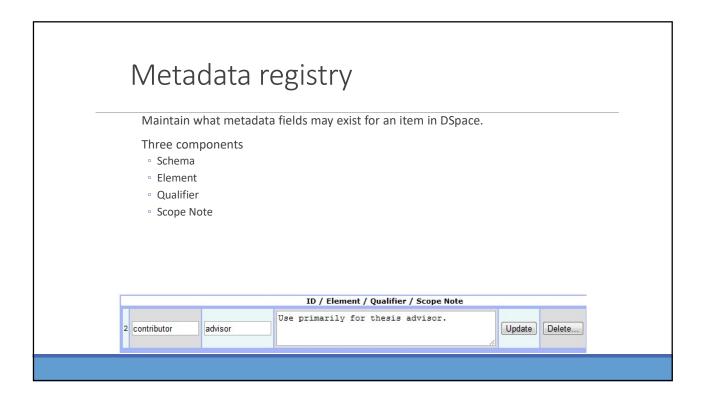
• No metadata

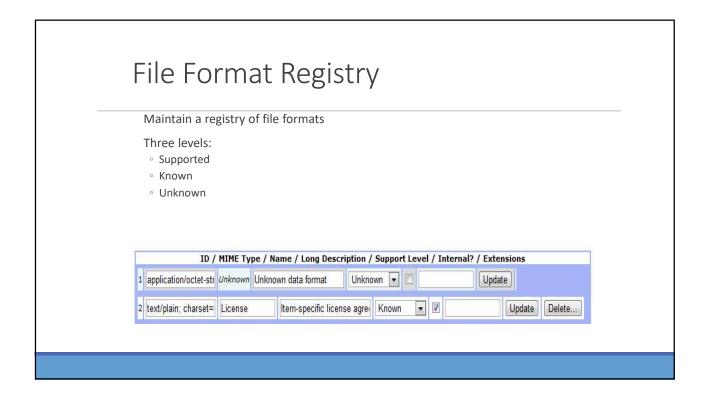
Example:

- HTML files and images that compose a single HTML document may be organized into a bundle
- Typical bundles are:
 - ORIGINAL LICENSE
 - THUMBNAILS CC_LICENSE
 - TEXT

Components & Features of DSpace







E-People

DSpace user accounts are called E-people

If permitted, an e-person may:

- Login to the site
- Sign up to receive notifications about changes to a collection
- Submit new items to collections
- Administer collections/communities
- Administer the DSpace site.

Authorization

The DSpace authorization system enables administrators to give e-people the ability to perform the following operations on an object.

- Add / Remove
 - Enable an e-person to add or remove any object (community, collection, item)
- Collection Administrator
 - Enable an e-person to edit an item's metadata, withdraw items, or map items into the collection.
- Write
 - Enable an e-person to add or remove bitstreams
- Read
 - Enable an e-person to read bitstreams

Ingestion

Ingestion = getting contents into DSpace

Batch import

- Many at a time
- Needs to be in a specific format
 - XML encoded metadata
 - Bitstreams

Web based submission

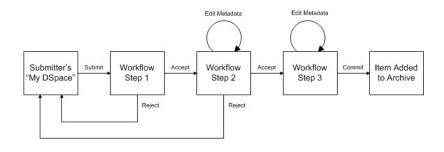
- One at a time
- Workflow processes

Workflow

Step 1: May reject the submission

Step 2: Edit metadata or reject

Step 3: Edit Metadata



Source : <u>www.dspace.org</u> (Dspace Documentation)

Search & Browse

Users may browse any item in DSpace

- Title
- Author
- Date
- Community / Collection
- Subject

Users may search for any item in DSpace based upon any Dublin Core value or a full text search.

Handle System

Provides a persistent identifier

Standard URL's change

- Hardware or software changes
- Political changes
- Network changes

Handles attempt to address these problems by creating a permanent URL independent of the repository.

Example:

• http://hdl.handle.net/1944/225

OAI-PMH

Enables other sites to harvest metadata from a DSpace repository

Collections are exposed as OAI sets

Only Dublin Core metadata is available

Statistics

Analyses the DSpace logs to generate a set of statistics on how DSpace is being used.

Metrics collected:

- Number of items archived
- Number of bitstream views
- Number of item page views
- Number of collection page views
- Number of community page views
- Number of user logins
- Number of searches performed
- Number of license rejections
- Number of OAI Requests

Presented in a by-month form or in-total form.

Areas one can customize

Submission process- one can configure the submission steps to suit organization

Browse and search terms- can set what fields and files you choose to index and display in the browse interface

Database- can choose Postgres or Oracle

Extend DSpace to work with other web services- using Light Network Interface you can pull or push content to/from DSpace

User interface- you can create your own user interface

4

Cloud Service Models

Cloud Software as a Service (SaaS)

• Use provider's applications over a network

Cloud Platform as a Service (PaaS)

• Deploy customer-created applications to a cloud

Cloud Infrastructure as a Service (IaaS)

 Rent processing, storage, network capacity, and other fundamental computing resources

To be considered "cloud" they must be deployed on top of cloud infrastructure that has the key characteristics

/1Ω

Key Characteristics

- Cost savings for resources
- Cost is greatly reduced as initial expense and recurring expenses are much lower than traditional computing
 - Maintenance cost is reduced as a third party maintains everything from running the cloud to storing data
- Platform, Location and Device independency
- Adoptable for all sizes of businesses, in particular small and mid-sized ones

5 Essential Cloud Characteristics On-demand self-service Resource pooling Location independence Rapid elasticity self-service

Cloud service: Potential Benefits

Enhanced Service Accessibility

- Access to Services that are otherwise unavailable
- Access to Services from multiple access devices
- Access to Services from scaled-down devices
- Access to Services from multiple device-types

Other Technical Benefits

- Professionalised backup and recovery
- Scalability
- Collaboration convenience
- Copyright convenience

Financial Benefits

- Lower Investment / up-front cost
- Lower Operational Costs
- Lower IT Staff Costs

CLOUD for Libraries

Libraries have been adopting cloud-based solutions services like electronic journal access management, statistics tracking, digital library hosting and now trend is coming up for hosted library management systems, Institutional Repository, Access Management and many more

....

CLOUD for Libraries

As libraries are having service-oriented mission they are in a position to adopt cloud computing.

Libraries (or librarians) are in constant search of finding proper solution within limited resources, moreover the outreach of service is quite dependent on support of external or internal computing (IT) support staff.

There are very few libraries or which are having IT support staff with expertise on advance IT management.

This situation makes SaaS, PaaS or laaS approach tempting to move towards cloud computing for providing better library services.

Cost Implications (In House Setup)

Sr. No.	Component (Capital)	Amount	Sr. No.	Recurring Components	Rs.		
1	Server	3,50,000	1	Power Consumption Server (800 watt, Avg), will	70,080		
2	Networking/Cabling/Others	50,000		consume 0.8 units in 24 Hours x 365 = 7008 Unit x 10 Rs.Per Unit			
	Total:	4,00,000					
Bandwidth Cost Not			2	AirConditioning Power Requirement 1800 Watts + 200 Watts for lighting and other = 2000 watts per hour will consume 48 Units in 24 Hours x 365 = 17520 Unit x 10 Rs. Per Unit	1,75,200		
			3	Manpower to Administer Server @ Rs. 10,000 per month	1,20,000		
	Counted			Recurring Approx Total (Rounded off) Per Annum	3,65,000		

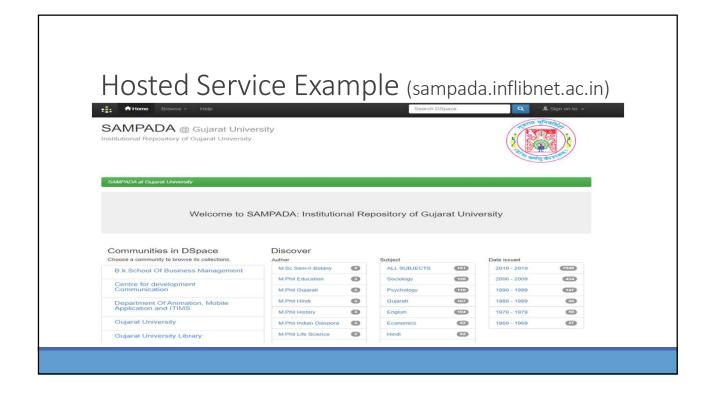
INFLIBNET Provides Hosting Services

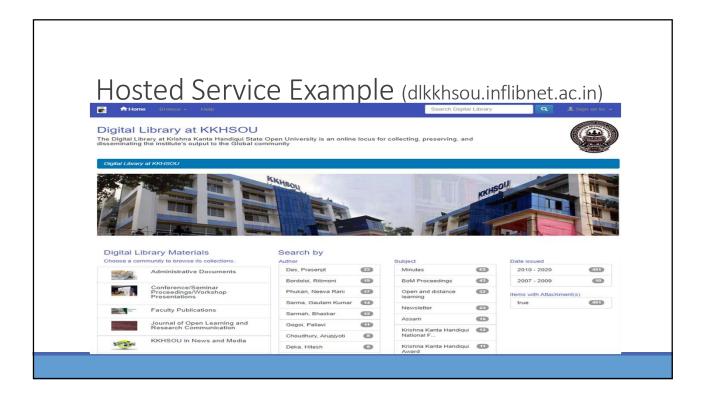
Virtual Machine Hosting	Dedicated Linux Virtual Machine (50GB), with dedicated IP Address and domain name with root credentials. Unrestricted Access on NKN	₹ 42,000.00 (Per Annum)	
Additional Storage Space	50 GB	₹ 10,000.00 (Per Annum)	
Server Hosting	Hardware to be supplied by client with requisite OS Licences, Unrestricted Access on NKN	₹ 75,000.00 (Per Annum Per Rack U)	

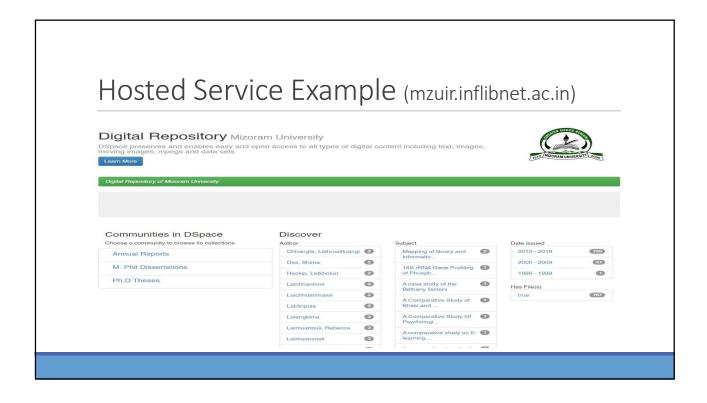
INFLIBNET also helps you to install and maintain...

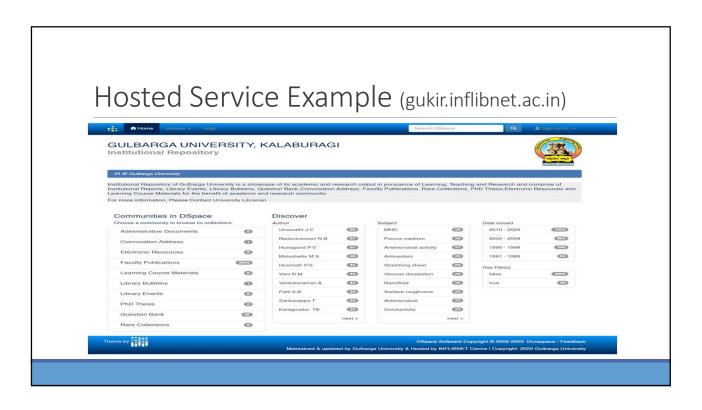
		The Rate Structure is Effective	ve from 19th June,2019
	Service	Deliverables	Rate*
A DSpace Installation Servi (Plain)		Latest stable version of DSpace installation in pre-installed Operating System (Windows/Linux) along with all pre-requisites (PostgreSQL, JDK, Maven, Ant, Tomcat)	₹ 7500.00
В	DSpace Basic Customization (Layout)	Layout Customization (Logo, Sidebar News, Header, Footer, Color-Fonts)	₹ 12500.00
С	DSpace Advance Customization and Data Import	Search Parameters, e-mail templates, OAI Interface, Localization, Branding, Custom Indexes, Custom Item display, Custom Input forms, Facets, Controlled Vocabulary, Data Import from compatible formats etc.	Will be Based on Complexity of Requirement.
D	DSpace Maintenance	Backup (Monthly) Log Maintenance Version upgrade with client's consent (once per annum) Statistics Generation (Monthly)	₹ 20,000.00 (Per Annum)

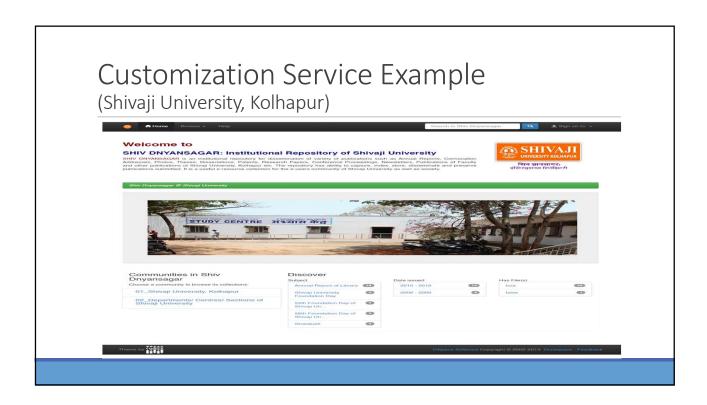












Acknowledgement

http://www.dspace.org

Thank You