# Open Access Landscaping in India: Building Institutional Repositories (IRs) Using 'DSpace'



M.G. Sreekumar Indian Institute of Management Kozhikode mgsree@iimk.ac.in

#### Presentation Plan

- India's Scientific Potential
- Availability of Scholarly Literature
- Prescriptions
- OA and its missions
- OA Efforts in India
- IRs and its Features
- IR Softwares

- Features and Functionality
- DSpace presence in India
- DSpace Promotion in India and the Asian Region
- Training Programmes
- Support Organization

#### Higher Education Scenario in India

- A vast country having one of the largest higher education system in the World
- > 171 Universities, 19 Central Universities
- > 99 Deemed Universities
- > 17000+ Colleges, 2000+ Polytechnics
- > 100,00,000+ (100 Lakhs) Students
- > 4,57,000+ (4.57 Lakhs) Teachers
- > 888,000+ educational institutions
- World class institutions of higher learning such as IISc, IITs, NITs, IIMs, Universities...

#### Scientific Research

- The third largest scientific and technical manpower in the world
- Vision oriented efforts since Independence (1947 +)
- Exclusive Government Departments for Science & Technology, Atomic Energy, Space, Electronics, Oceanography, Biotechnology...
- Over 300 Research Laboratories belonging to CSIR, ICMR, ICAR, ICSSR, DRDO, ISRO...
- Education/Science performed by IISc, IITs, NITs, IIMs...and most of the Medical/ Engineering/Business Schools, Universities and research labs are of international standards

# Scientific Productivity

- Science in India presently shows sings of stagnation
- ❖ Number of Scholarly Articles in SCI 12,000 (approx.)
- ❖ Total number of articles SCI + non-SCI 30,000+
- A large number of articles are left uncaptured by international indexing / abstracting / fulltext databases
- Contributes only a small percent to the International literature

# Availability of Scholarly Information/Literature

- Widening Divide between Information Haves and HaveNots
- Non-availability of international science literature (peer reviewed scholarly content)
- Only a few of the Institutions have endowed infrastructure
- Lack of exposure and lack of a level playing ground to the academic/scientific community
- Adverse influence on India's overall scientific productivity

### Prescriptions

- Do Not Change the Scholarly Communication Practices
- Embrace and Promote Open Access
- Publish the findings in the publication of choice (national / international);
- Also contribute to Open Access Archives
- Consider Publishing in Open Access Publications

# Open Access

Open Access Publications (OAP)

+

Open Access Archives (OAA)

### Examples of OAP

- **\*BMJ**
- Indian Academy of Sciences
- **❖BMC** and PLoS
- High Wire Press: Delayed access
- Selected papers in a journal OA

### Examples of OAA

- Central Archives: arXiv, CogPrints
- ➤ PubMed Central, CiteSeer
- > Eprints@iisc, Idl@drtc, OpenMed@nic
- Though not an OA Archive, "PubMed"s Open Access Philosophy is Laudable
- ➤ Interoperability the Key

#### Drivers of OAP and OAA

- Rising prices of journals
- ARL and SPARC initiatives
- Librarians want OA so they can afford
- Scientists want OA so they could be read and recognized
- People want OA because they pay for the research

### Public Support for OA

- Budapest Open Access Initiative
- US House of Representatives Appropriations Committee [NIH]
- UK House of Commons Committee on S&T publications
- Similar initiatives in Australia (APSR...), Canada, Scotland
- Statement signed by 25 Nobel Laureates in USA
- Berlin Statement
- Bethesda Statement

# Open Access Initiatives

- Projects
- Training Programmes
- Publishing

# Open Access Publishing

- Indian Academy of Sciences
- \*INSA
- IndMed, MedInd, OpenMed
- ❖ MedKnow
- DOAJ Directory lists over 1200 Journals
- ...

# Open Access Archiving Training Programmes

- **\*DRTC**
- \*NCSI, IISc
- **❖INFLIBNET**
- **♦**IIMK
- Individual and other isolated efforts...

# Advantages of Archiving (for researchers)

- Dissemination
- ❖Increased visibility (Google, OAI…)
- More visibility leads to more citations
- Research impact
- Preservation
- Control / Monitoring of one's own Publications

# Advantages of Archiving (for institutions)

- Pooling the Organization's Intellectual Capital
- One Stop Source / Point for the research output of an Institution
- Scope for Introspection / Strategies / Action Plan
- Generation of reports
- Long term preservation

#### IR Software - Desirables

- Key component of an IR is the repository management software
- Several software are now available under open source license
- Comply with OAI metadata harvesting protocol
- Released and publicly available

#### IR Software - Features

- Capture and describe digital material using a workflow
- Provide interface for online submission of research material (Intranet)
- Provide access to this material over the web (metadata and/or full pub)
- Preserve digital material over long period of time
- Expose metadata through OAI-PMH protocol
  - Default: Unqualified Dublin Core
  - Other metadata standards

#### IR Softwares

- ARNO Tilburg University, The Netherlands
- \* CDSware CERN, Geneva, Switzerland
- ❖ DSpace MIT Libraries and the HP Labs, USA
- EPrints University of Southampton, U.K.
- \* FEDORA University of Virginia, USA
- i-Tor Netherlands Institute for Scientific Information Services
- MyCoRe University of Duisburg-Essen, Germany

# Content Categories

#### Published material

- \*Ex.: Journal papers (post-prints), book chapters,
- conference papers

#### Unpublished / gray material

- Ex.: Pre-prints, working papers, minutes, theses and
- dissertations, technical reports, progress/ status
- reports, committee reports, course material,
- presentations, multimedia material, etc.

#### Supporting material

Ex.: Data sets, models, simulations

# Additional Functionality

- Multilingual content/ interface support
  Unicode
- Batch import/export
- Other interoperability protocolsZ39.50, OpenURL
- Other metadata standards and crosswalks
  (e.g. METS, MARC)
- Persistent URL (Identifier)

# **DSpace**

- An Open Source Software BSD License
- MIT Libraries and HP Labs
- Digital Object / Asset management system
- Create, search and retrieve digital objects
- Facilitate preservation of digital objects
- Allows open access and digital archiving
- Allows building Institutional Repositories

# H/W and S/W requirements

- Let's NOT underestimate!
- It's a resource intensive platform
- ❖Based on Java / JSP
- UNIX recommended (Java-based program should run on anything)
- Open source, built on Apache web server and Tomcat Servlet engine
- Uses postgreSQL relational database

# **DSpace Functionality**

- Captures
  - Digital content in any format directly from creators (e.g. researchers, authors)
- Describes
  - Descriptive, technical, rights metadata
  - Persistent identifiers
- OAI-PMH version 2.0 compliant
  - Allow metadata creation

# Content Categories / Publication Types

- Preprints, Articles
- Postprints
- Technical Reports
- Conference Papers
- Theses/Dissertations
- Working Papers
- Datasets
  - -e.g. statistical, geospatial, scientific

#### File Formats

Supported: fully supports the format

Known: recognizes the format, but cannot guarantee full support

Unsupported: cannot recognize a format; these will be listed as "application/octet-stream", aka Unknown

#### Known/supported

TIFF, SGML/XML, PDF

#### Known/unsupported

- MS Word, PowerPoint
- Lotus 1-2-3, WordPerfect

#### Information Model

- Communities
  - Departments, Labs, Research Centers,
     Schools...
- Collections
- Items
- Files (bitstreams)
  - Multiple formats same content
  - Complex objects multiple files

# Special Features

- 100 % Open Source
- International Acceptance
- Modular, Scalable and Componentized Architecture
- Remote Publishing facility
- CNRI 'Handles' support for Persistent URLs
- OAI-PMH Compliance and Interoperability
- Open URL Compliance

#### Search Features

- Fielded
- Boolean
- Exact term
- Proximity
- Wild Cards
- Fuzzy
- Range
- Boosting Terms

# Strengths of DSpace

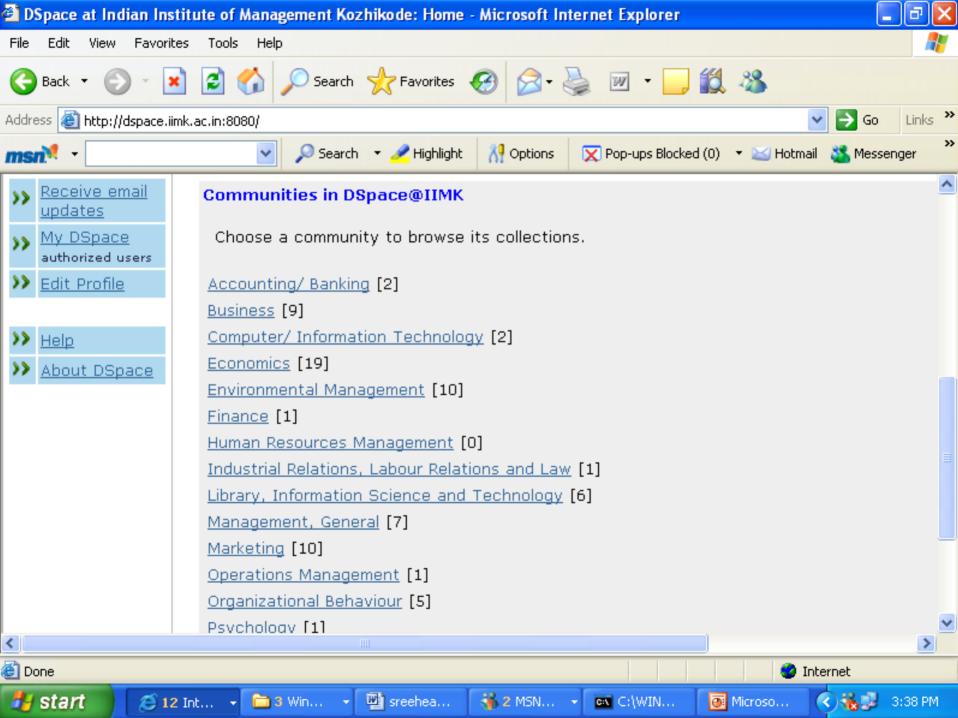
- Communities / collections
- Backed up by MIT and HP
- Simple yet powerful documentation
- Strong workflow support
- Handle-based identifier
- Better articulation of preservation strategy
- Default support for qualified DC
- User (E-)Groups, Lists, User Meets ...

#### **DSpace Based Archives in India**

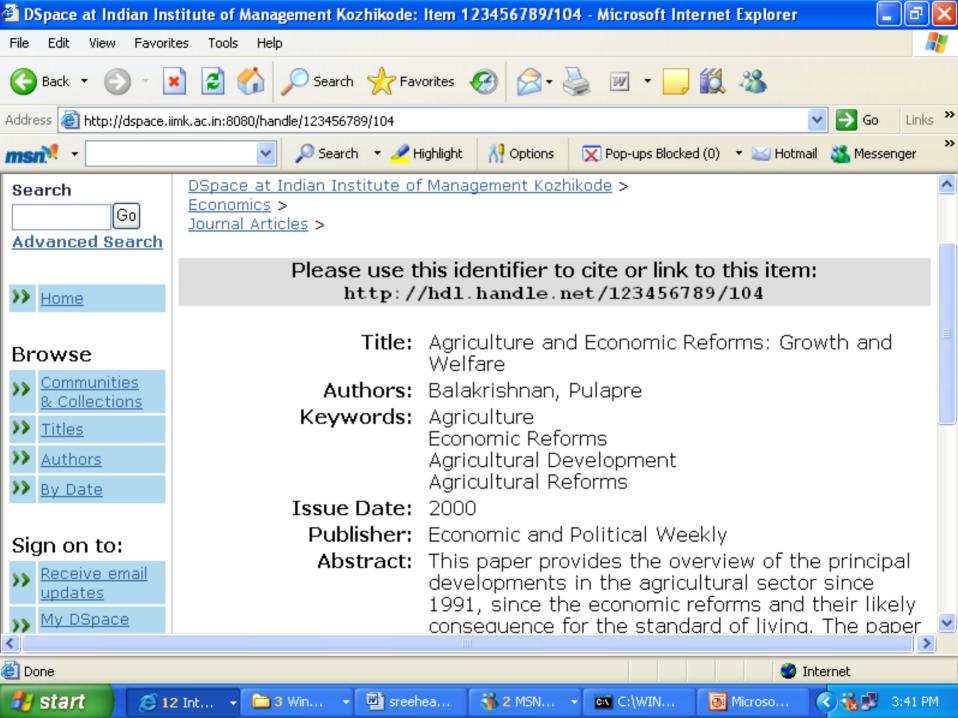
- Indian National Science Academy (INSA)
- National Chemical Laboratory (NCL)
- ETD@IISc, Indian Institute of Science
- ❖ Indian Institute of Technology (IITD), Delhi
- **❖** INFLIBNET
- ❖ ISI, Bangalore
- University of Hyderabad
- National Centre for Radio Astrophysics
- National Institute of Technology (NIT), Rourkela
- LDL: Librarians' Digital Library, DRTC
- Indian Institute of Management Kozhikode (IIMK)

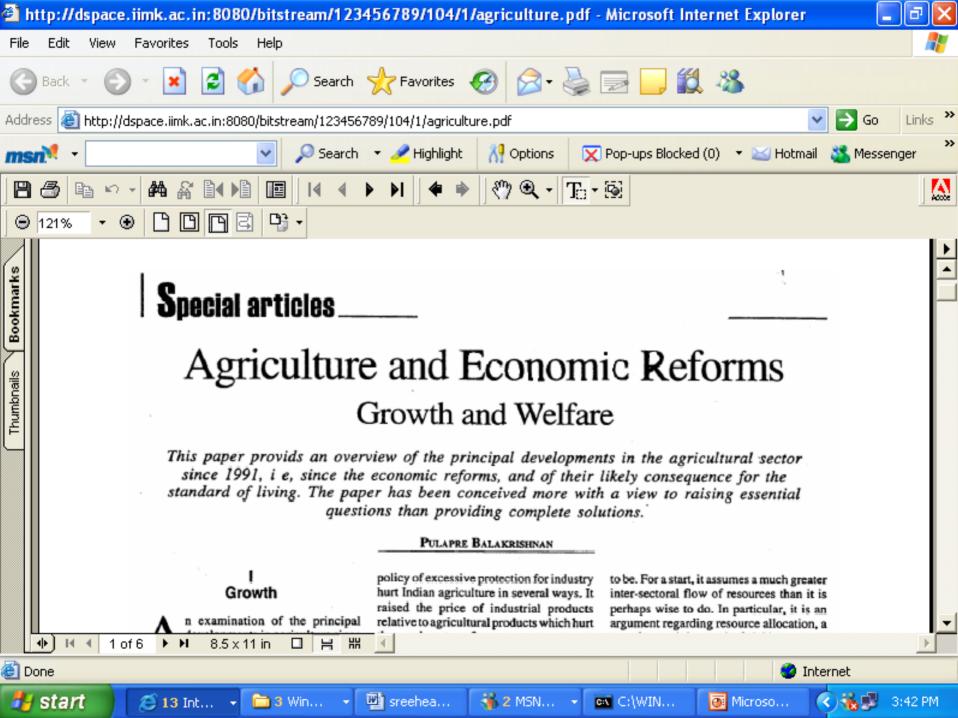


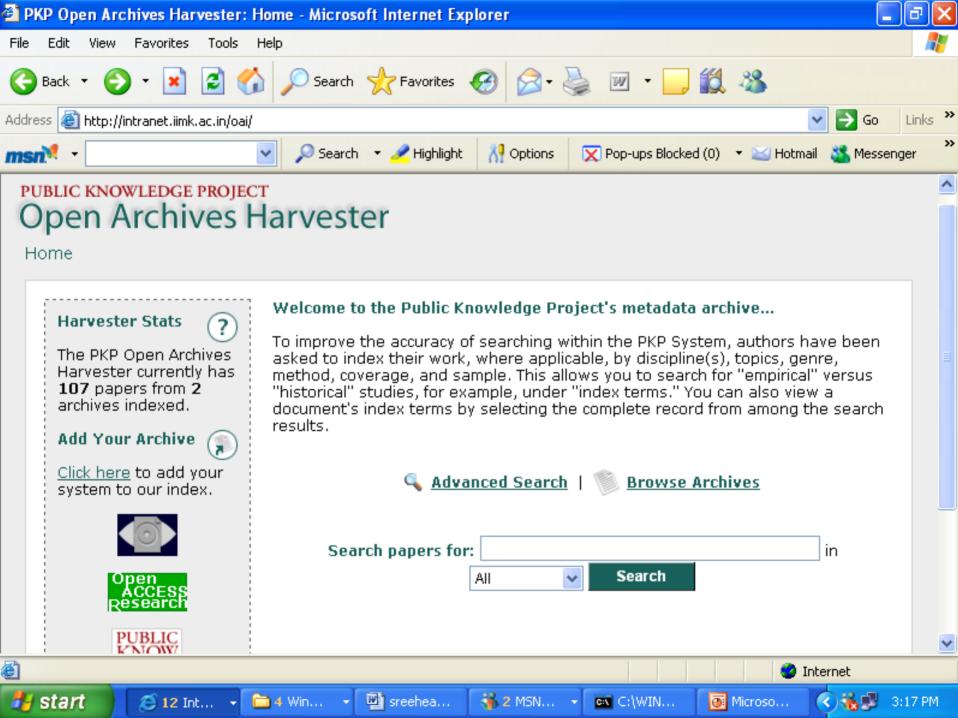


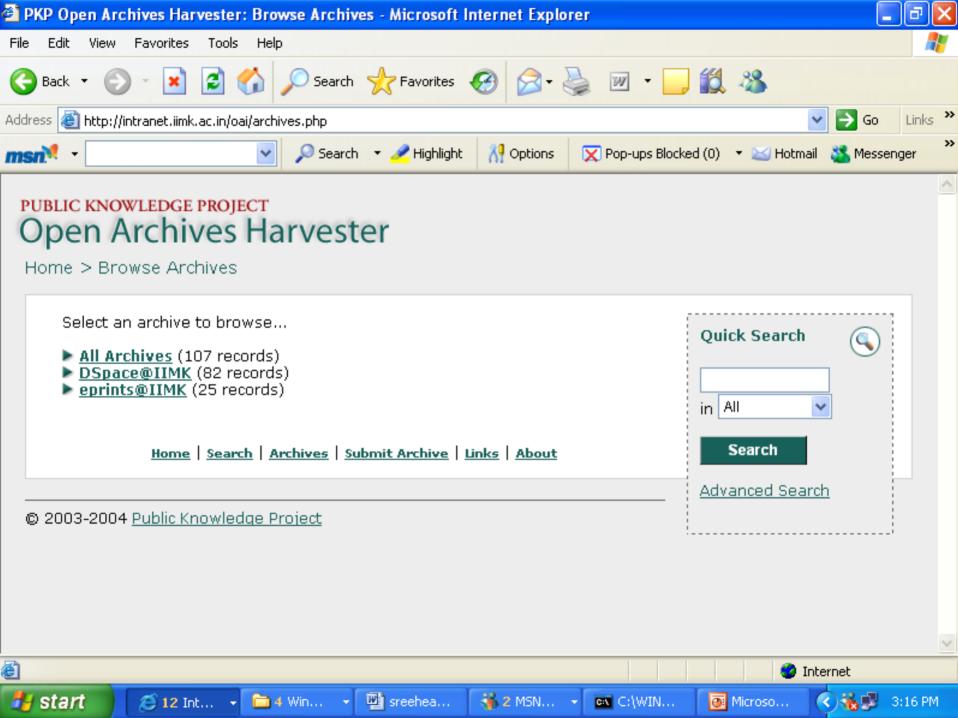


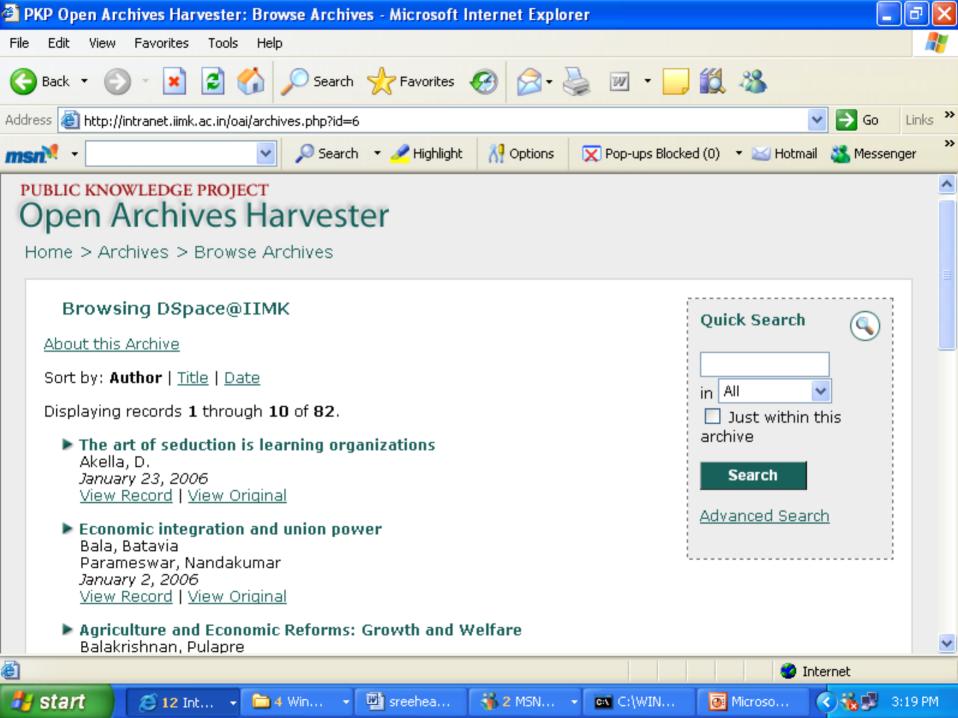


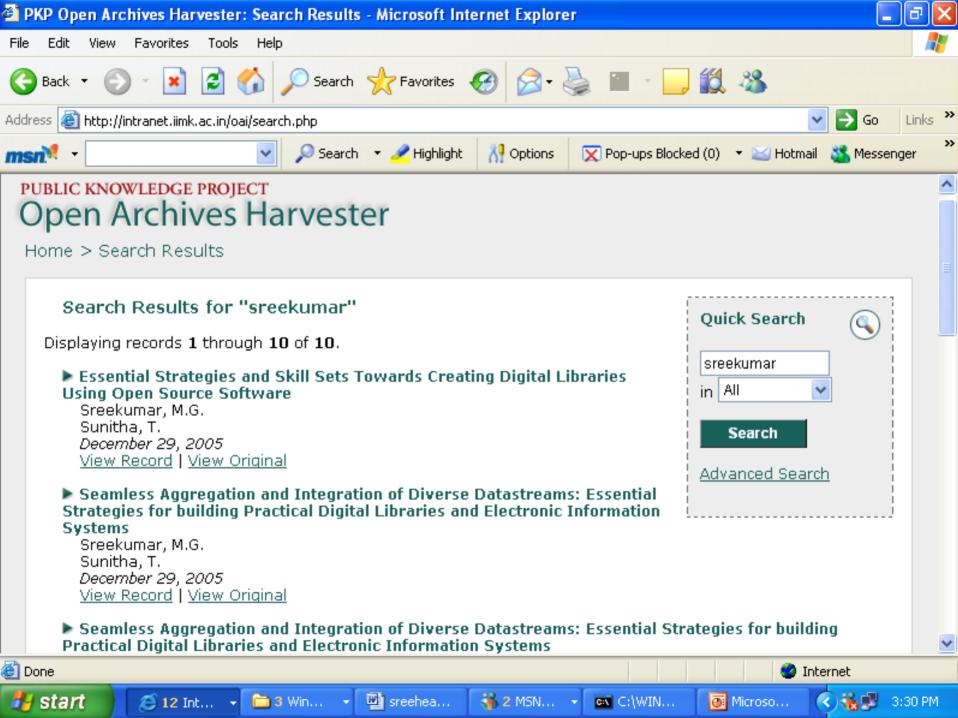








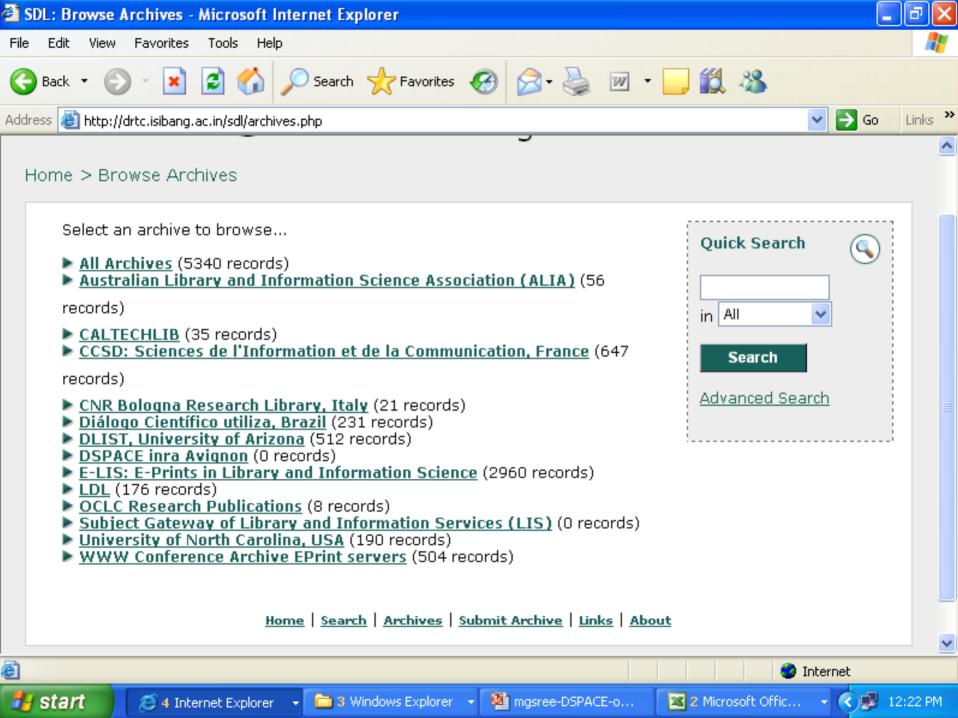












#### **DSpace Promotion in India**

- Conducive Environment
- Growing Interest in Building IRs by the Indian Library / Information / Academic Fraternity
- Scope for Institutional as well as Central Archives
- Growing Acceptance of DSpace
- Training Programmes
- Regional Support Organization (Virtual)

#### **DSpace Training in India**

- Training Programmes
  - DRTC, IIMK, INFLIBNET....
- Institutional Repositories Workshop in Chennai (December 2-4, 2005)
  - MALA-CSI 45 participants (URL)
- E-Publishing Workshop at IIMK (December 12-14, 2005)
  - 15 participants (URL)

Participants of the Chennai Workshop (December 2-4 2005)







## Participants of the IIMK Workshop (December 12-14 2005)



#### DSpace Workshop in Progress



#### DSpace Training in India...

- Forthcoming Training Programmes
  - MHRD Institutions (40 Institutions)
  - ICMR Institutions (20 Institutions)
  - IIMK's E-Publishing Workshop (30 Institutions)
  - Programmes Hosted by other Organizations/Institutions
     Eg.: MANLIBNET, KLA etc.

# DSpace Support Organization for India and Asian Region

- Identify a Core Group for Technical Support
- Identify an Institute as the DSpace Nodal Center for the Region
- Coordinate DSpace Support (Online as well as face-to-face) Actively
- Candidates : DRTC, INFLIBNET, IIMK

#### Acknowledgement

- DSpace.org
- MIT, Boston
- Hewlett Packard
- APSR, Australia
- EPrints.org
- Dr. T.B. Rajashekar (Late) & Team, NCSI
- Prof. A.R.D. Prasad & Team, DRTC
- Prof. Subbiah Arunachalam, MSSRF
- Prof. A. Amudhavalli, University of Madras
- DL Team, IIMK



### Thank you