



DIGITAL LIBRARIES: AN OVERVIEW



Traditional Libraries

- Libraries with the same purpose, functions, and goals
 - Collection development and management
 - Technical Processing
 - Index creation
 - Counter Transactions
 - Reference work
 - Preservation



Five Elements in Various Definitions of DL

- The digital library is not a single entity;
- *The digital library requires technology to link the resources of many;*
- *The linkages between the many digital libraries and information services are transparent to the end users;*
- *Universal access to digital libraries and information services is a goal;*
- *Digital library collections are not limited to document surrogates: they extend to digital artefacts that cannot be represented or distributed in printed formats.*

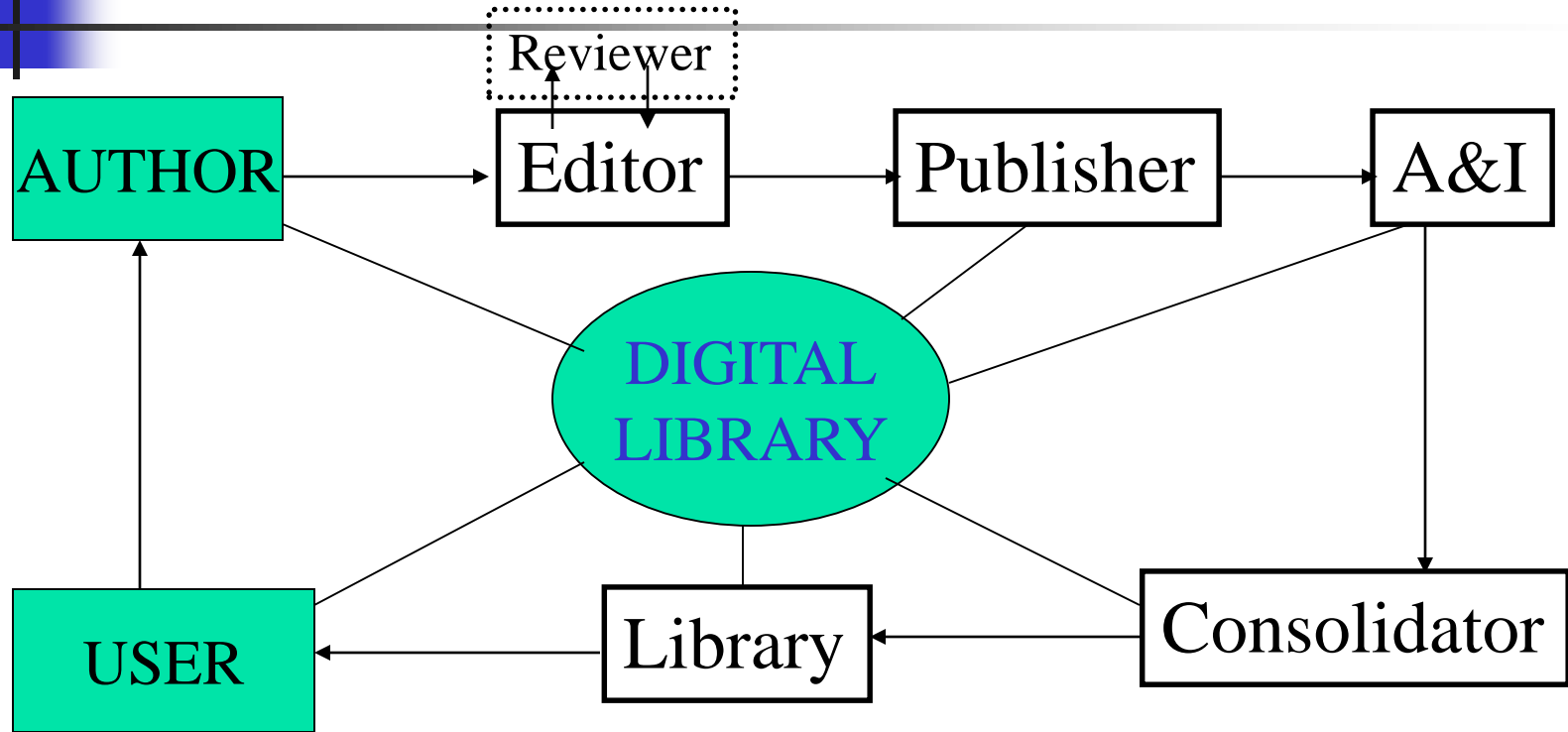
Association of Research Libraries (1995)



Goals of DL

- Focused on digitization technology, metadata schemes, data management techniques, and digital preservation.
- Second-generation digital library
 - exploring new opportunities and developing new competencies.
- Third-generation digital library
 - focusing instead on fully integrating digital material into the library's collections through a modular systems architecture.

Digital Libraries Shorten the Chain



ROLES

READER

AUTHOR

LIBRARIAN

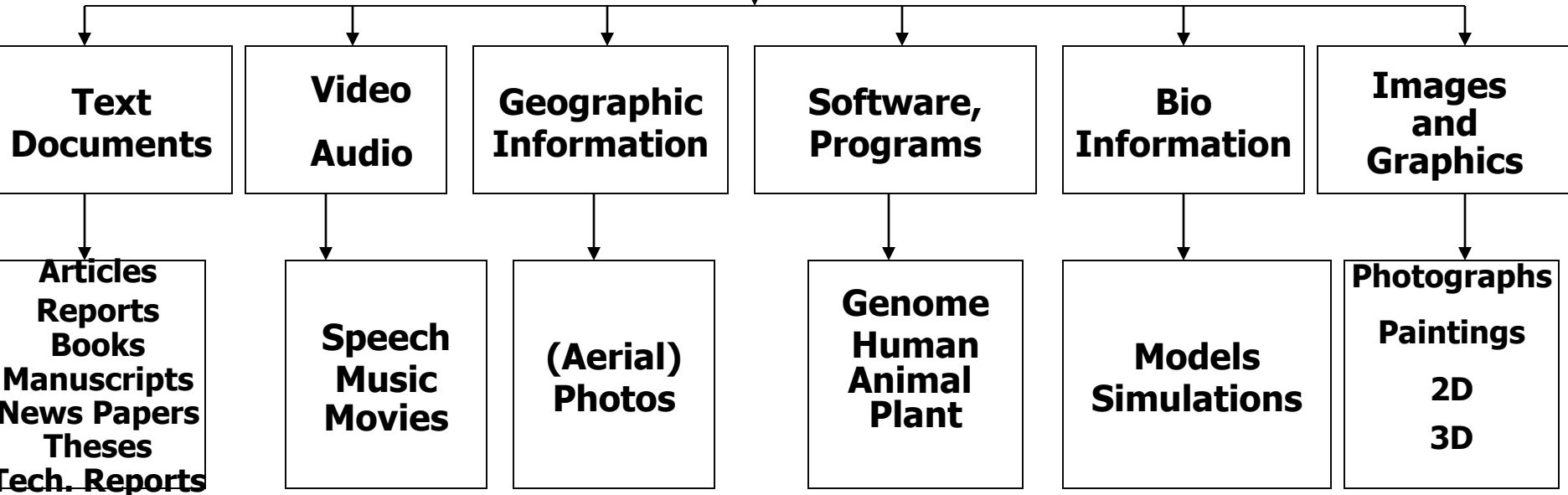
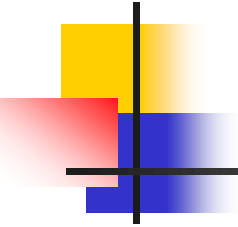
EDITOR

LEARNER

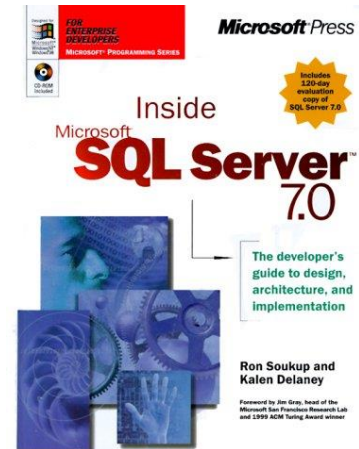
TEACHER

Digital Library Content

Content Types



Content is King



The information content is more important than the systems used for its storage, management and retrieval

Objects should not be “locked” in specific DLs or archives



Types of Digital Collections

- **Digitization**

- Converting paper and other media in existing collections to digital form

- **Acquisition of original digital works**

- Created by publishers and scholars like electronic books, journals, and datasets

- **Access to external materials**

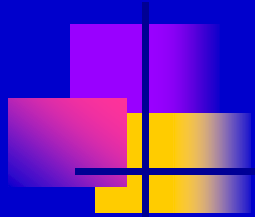
- Like Web sites, other library collections, or publishers' servers



Creating DLs ...

- Six steps
 - Selecting
 - Acquiring
 - Digitization
 - Organizing
 - Archiving
 - Providing Access

Process



Selection of
Books

Identification of
Books

Meta Data

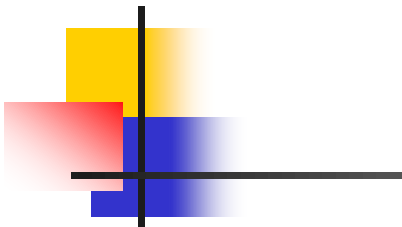
Scanning
Process

Scanning

Image
Processing & QC

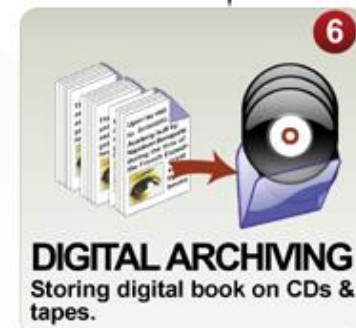
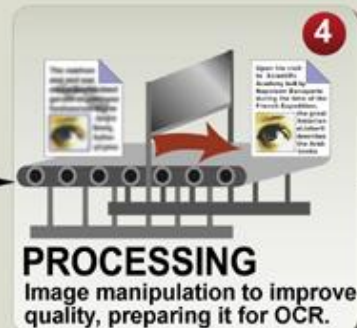
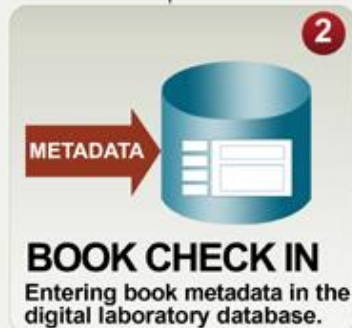
OCR

Publishing



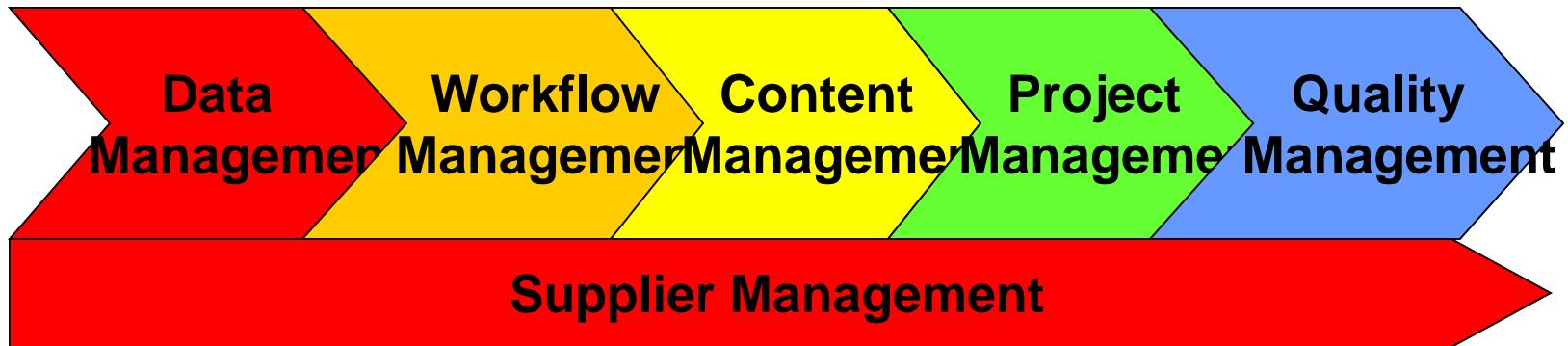


Digital Books workflow



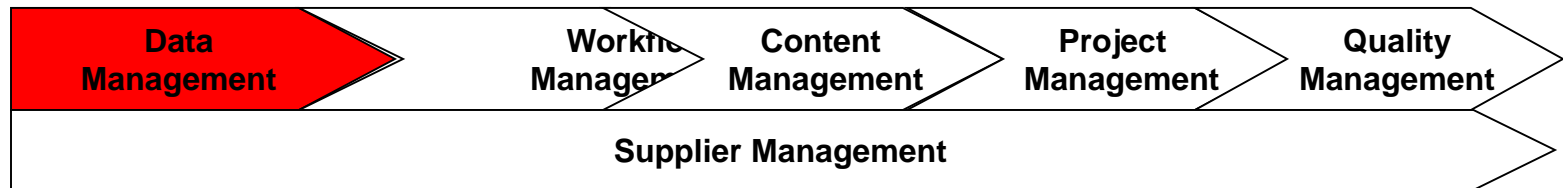


Digital Production Process





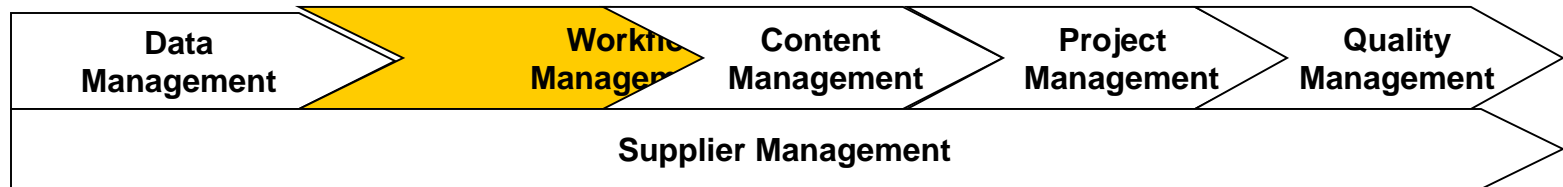
Data Management



- Formats: TEX, PDF,PS
- Metadata and content data
- Structuring (Tagging)
- Media neutrality



Workflow Management



- Processing
- Conversion
- Automatization
- Interfaces - input / output



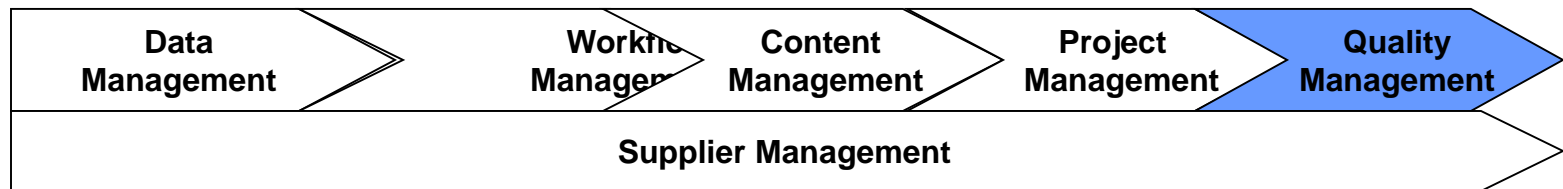
Content Management



- Style files
- Information /Object models
- Archiving

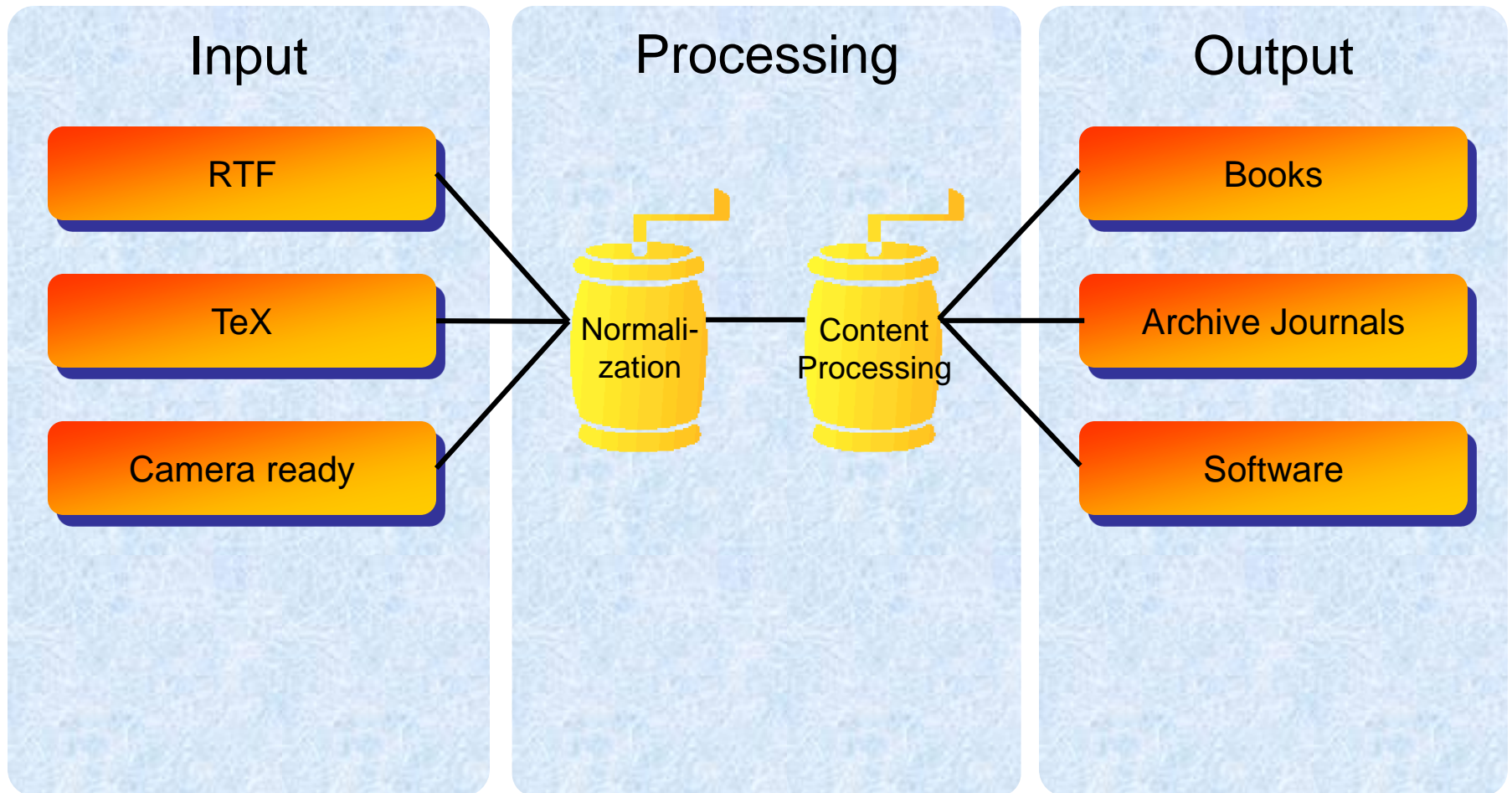


Quality Management



- Data consistency
- Process consistency
- Content consistency

Various Workflows





Possible Delivery Formats

- Pure image formats: TIFF, JPEG
- Open encoded formats: XML, HTML, ASCII, and Unicode
- Hybrid formats: PDF, DjVu – can contain both image and text
- Proprietary formats: Microsoft Word, WordPerfect



Good Principles

- What to digitize?
 - Selection and policy is important
- Collection description is important
 - such as scope, format, restrictions on access, ownership etc



Digitization: Issues

- Copyright
- Access copy and archive copy
- File size
- Storage media(CD, Hard disc...)
- File format (TIFF,JPEG...)



Challenges in Publishing

- Preservation of layout
- Searchability of content and metadata
- Efficient image compression
- Easy browsing of books
- Accommodating low bandwidth user
- Multilingual text support
- Multipaging



Digital Library Standards

- **Common User Interface:**

Data Handling and Interchange:

Graphic Formats – JPEG, TIFF, GIF, PNG, Group 4 Fax, CGM

Structured Documents – SGML, HTML, XML

Moving Pictures/3-D – MPEG, AVI, GIF89A, QuickTime, Real Video, ViviActive, VRML

- **Metadata:**

Resource Description – Dublin Core, WHOIS++ Templates, US-MARC, TEI Headers, Other Open Source and Domain Specific Standards.

Resource Identification – URN, PURL, DOI, SICI

- **Security, Authentication and payment services:**

Emerging e-Commerce Standards.



Indian DL Initiatives: Contents

- Books (out of copyright)
- Scholarly Journals
- Theses
- Institutional E-Prints
- Manuscripts
- Data
- News Papers
- Metadata Level
- Portal and Gateway Services