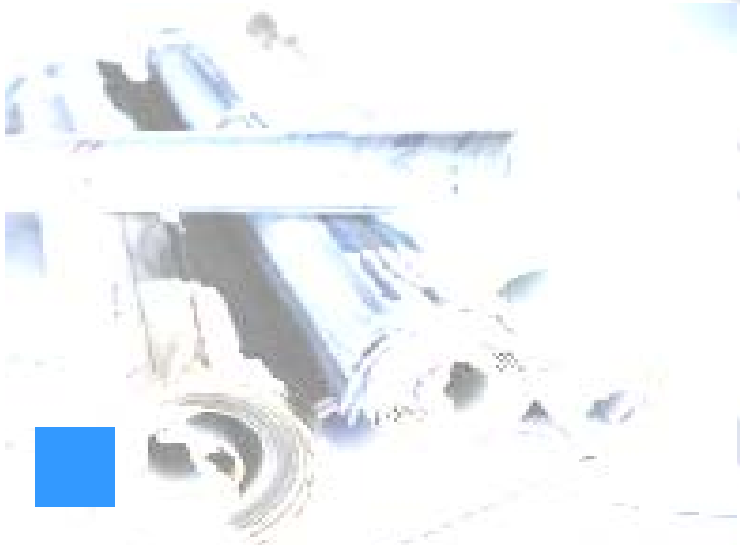


WIPO
WORLD
INTELLECTUAL PROPERTY
ORGANIZATION

Introduction to WIPOScan Software

An overview of available WIPO technical assistance on digitization, such as WIPOScan and detailed modules for digitizing all kinds of industrial property data

Gregory Sadyalunda, Project Manager
Infrastructure Modernization Division

Manila, Philippines
7 – 9 December
2010

Contents

- INTRODUCTION
- SYSTEM OVERVIEW
- DEPLOYMENT CONSIDERATIONS

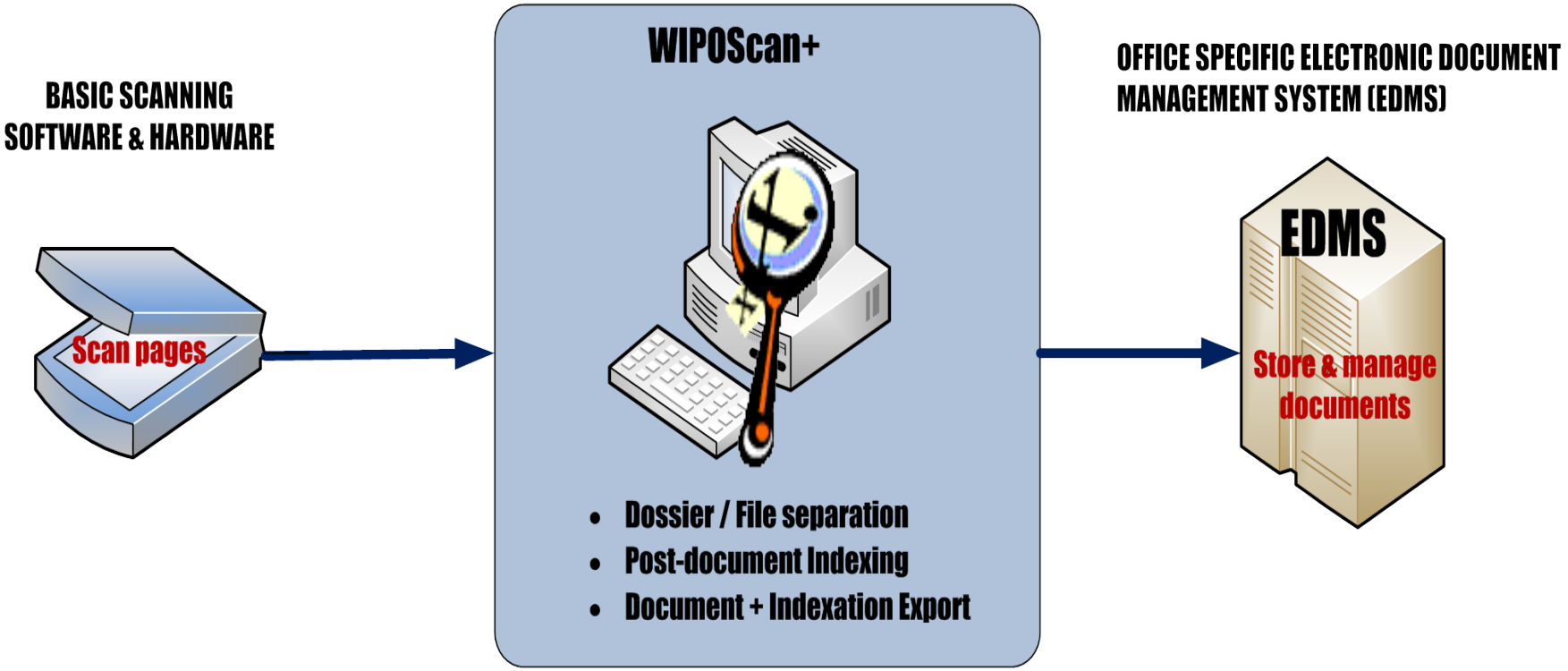
Contents

System Background

- Recognized the need for conversion of paper documents to support new business models / services and data exchange cooperation

- Provides an application that enables the indexing of scanned documents

WIPOScan+ Concept & Scope



What is WIPOScan?

- Tool for business process and backfile scanning & digitization
- Production tool for conversion of printed documents into fully indexed/tagged digital objects
- New Version of WIPOScan launched in 2010
- Capable of scanning documents across different IP domains i.e. Patents, Industrial Designs, Trademarks etc.

Benefits of Digitization

- Preserve the origin
- Enable quick and enhanced access by high structured documents
- Open up new dimensions of new business models, statistics & research
- Provide standardized output formats for data exchange & systems integration
- Reduce cost of paper processing
- Increase user productivity & throughput
- Add value by increasing quality of service

Contents

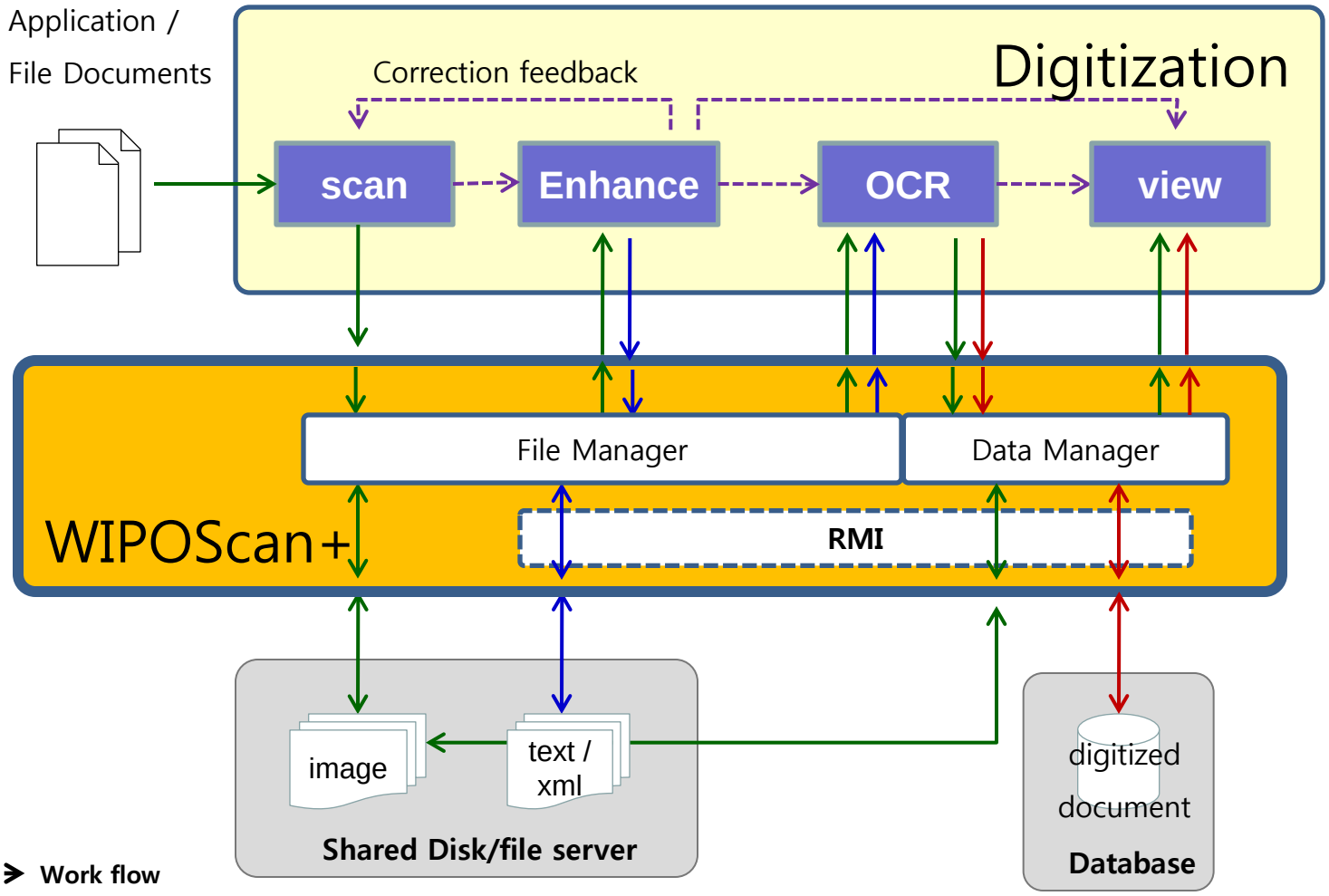
Basic Functions

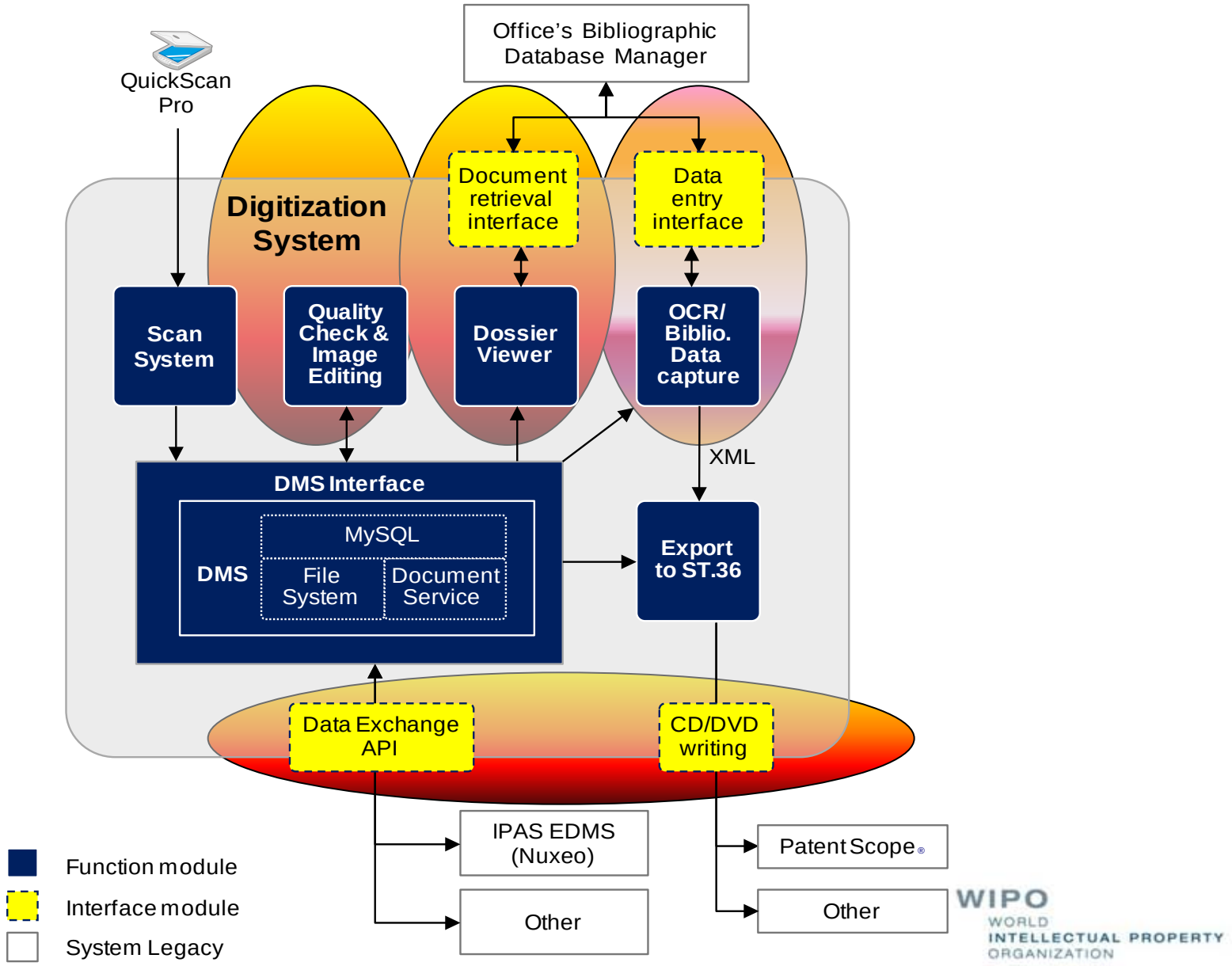
- **File / Dossier separation and indexing**
  - WIPOScan+ separates batch scanned files & indexes them by file/dossier number, document type and document date
- **Document image editing and enhancement**
  - Provides functions for improving the quality of scanned images including spots removal, deskew and dirt removal
- **File/Dossier viewer**
  - View indexed documents and search by document number, type and date
- **Document export**
  - Export scanned documents in zipped TIFF & XML formats

Technologies and Standards

- Java Swing (windows-based) application
- Java Advanced Imaging (JAI) for image enhancement & processing
- Remote Method Invocation (RMI) for DBMS Application Programming Interface (API)
- eXtensible Markup Language (XML) / WIPO ST.36
- Tagged Image File Format (TIFF) G4, 300 dpi
- Portable Document Format (PDF)
- FineReader Optical Character Recognition (OCR) – optional
- MYSQL Database Management System

WIPOScan Architecture




Office's Bibliographic Database Manager
QuickScan Pro
Digitization System
Document retrieval interface
Data entry interface
Scan System
Quality Check & Image Editing
Dossier Viewer
OCR/ Biblio. Data capture
XML
DMS Interface
DMS
MySQL
File System
Document Service
Export to ST.36
Data Exchange API
CD/DVD writing
IPAS EDMS (Nuxeo)
Other
Patent Scope
Other
Function module
Interface module
System Legacy
WIPO
WORLD INTELLECTUAL PROPERTY ORGANIZATION

Hardware and Software Requirements

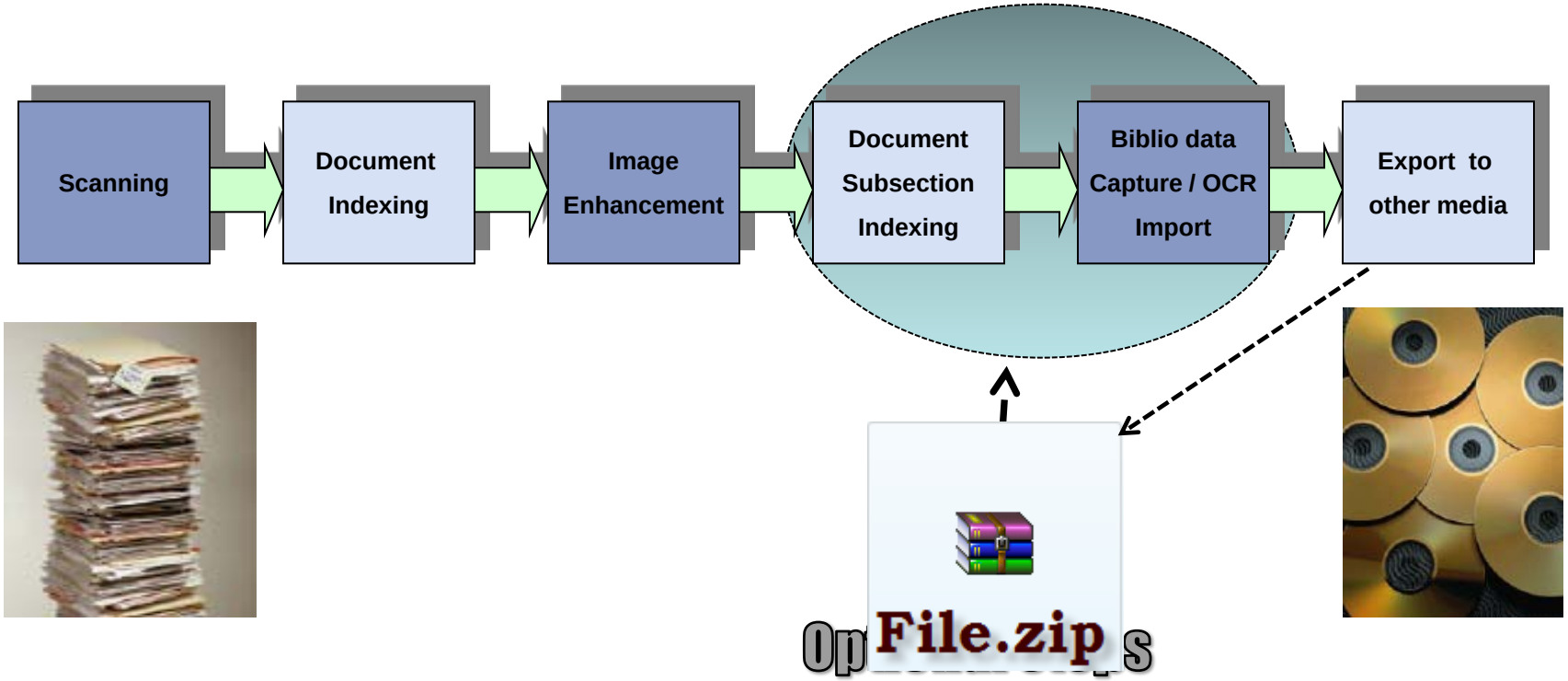
Hardware

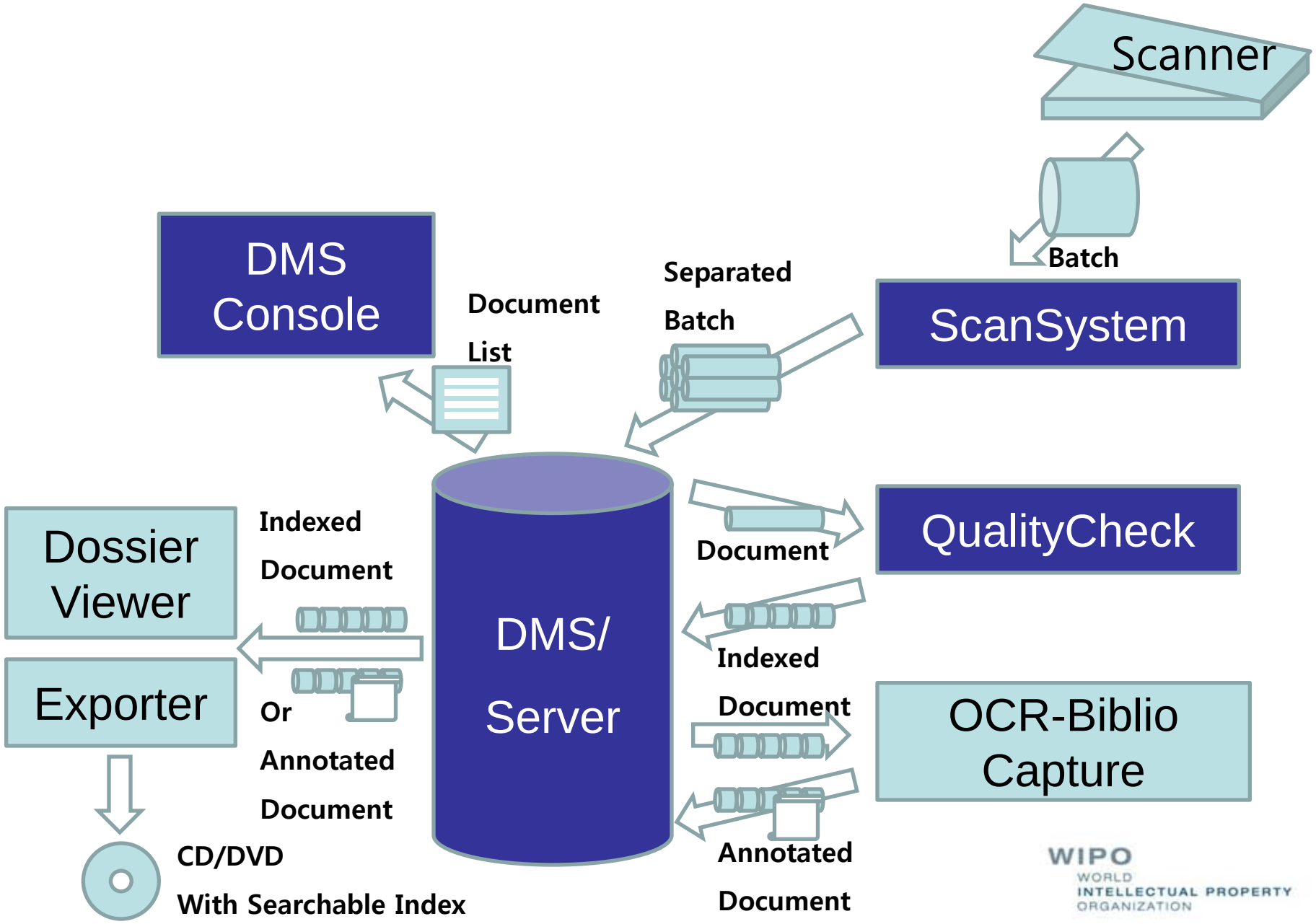
- Minimum Specification
  - CPU : Pentium IV
  - RAM : 2 Gigabyte (GB)
  - HDD : 13 GB Client and 7 GB Server (installation files) / User files storage depends on volumes
- Stand-alone Workstation, Client / Server or WAN environment
- Peripherals
  - Color monitor
  - Scanner and printer
  - CD / DVD drive / writer
  - Network environment

Software

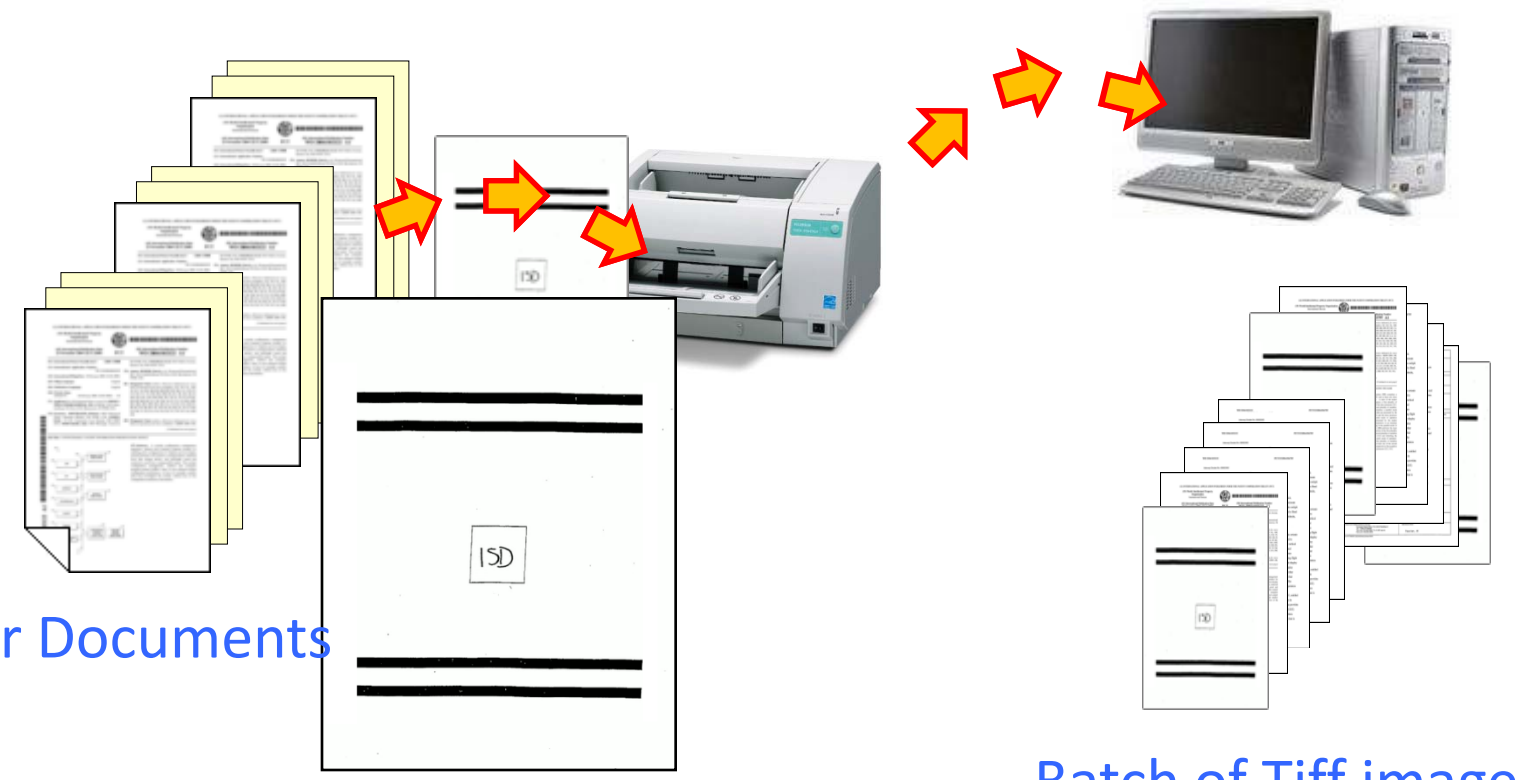
- Required software
  - O/S : Windows XP or higher
  - Scanning tools
  - CD / DVD burning tools
  - Text Editor i.e. Notepad, WordPad etc.
- Optional software
  - Database Management System (Oracle or MS SQL SERVER)
  - FineReader OCR (*current under development*)
- Freeware
  - MYSQL
  - Java Virtual Machine (JVM)
  - Java Editor and compiler (*for further customization and development by the office*)

WIPOScan+ Basic Workflow




Scanner
Batch
DMS Console
Document List
Separated Batch
ScanSystem
Dossier Viewer
Indexed Document
Exporter
Or Annotated Document
DMS/ Server
Document
QualityCheck
Indexed Document
OCR-Biblio Capture
Annotated Document
CD/DVD With Searchable Index

Scanning Document

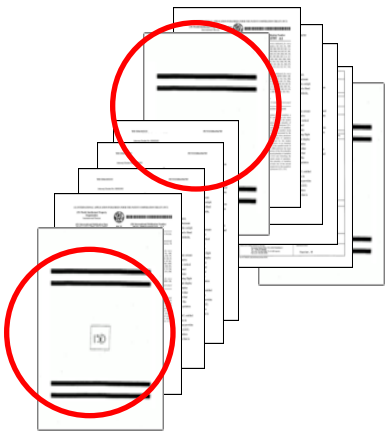


Paper Documents

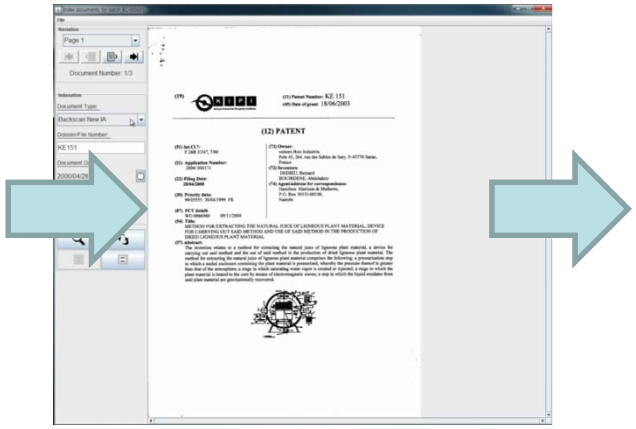
Separator Sheet

Batch of Tiff images

Loading Images

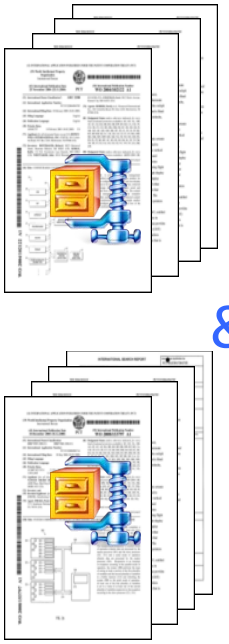


Tiff images



Detect Separator sheet,

Input DocID & type



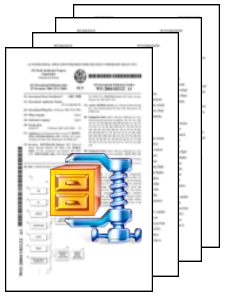
Separated

& Compressed

Image files

Editing Scanned Documents

- Image Quality Improvement (Deskew, etc.)



Document

Image files

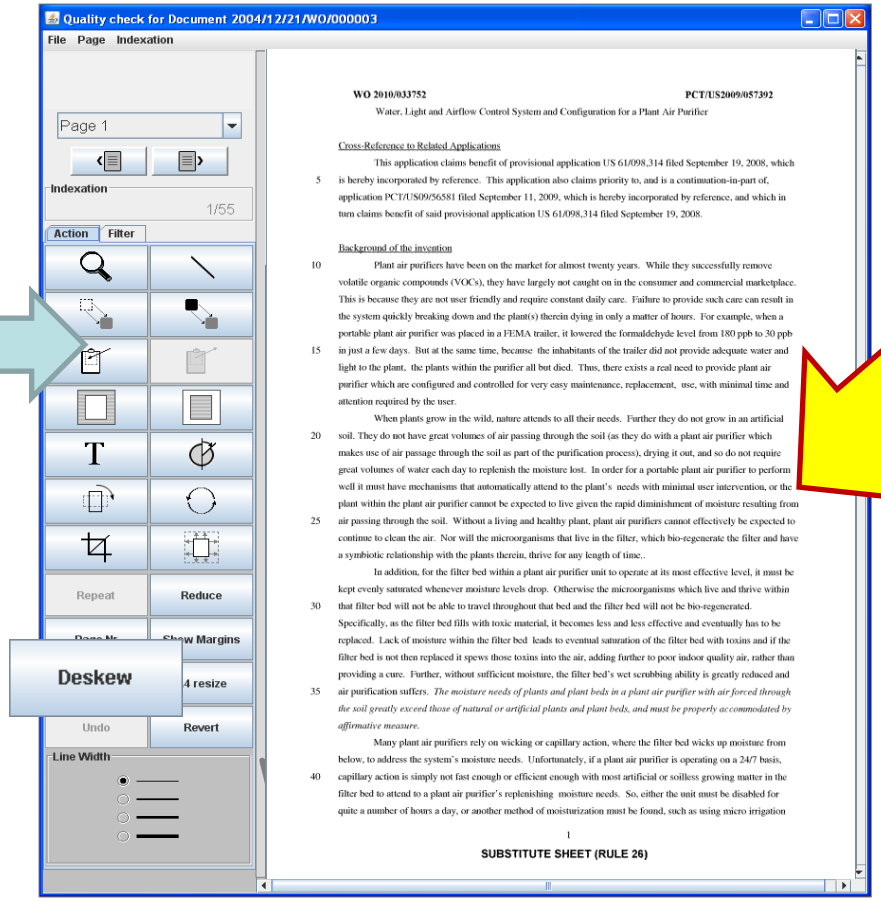
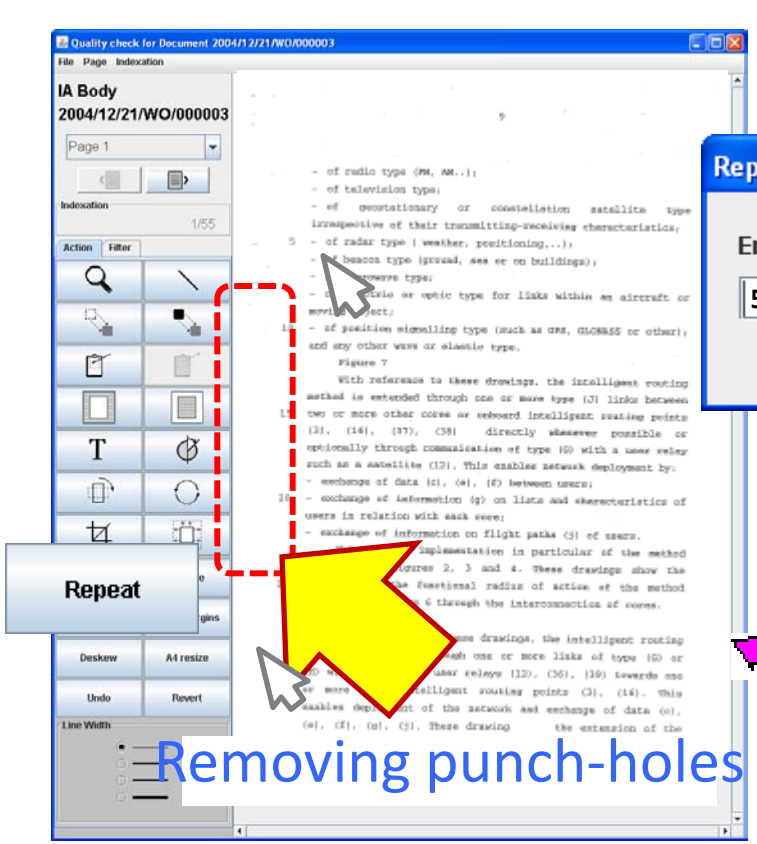


Image Enhancement

WIPO WORLD INTELLECTUAL PROPERTY ORGANIZATION

Editing Scanned Documents

- Repeat over pages

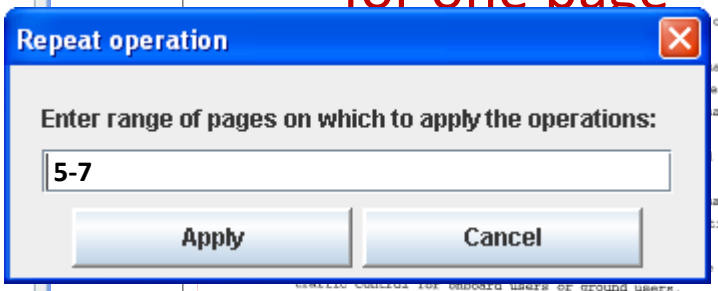


Removing punch-holes

Edit image

Enter the Range

for one page



And More

Image Improvement Functions

As non-restrictive example, a fixed ground user may be a natural person, company, communication relay, computer server, measurement device, radio transmitter, stationary vehicle, airport base, stationary grounded aircraft or any other entity on the ground or below the ground.

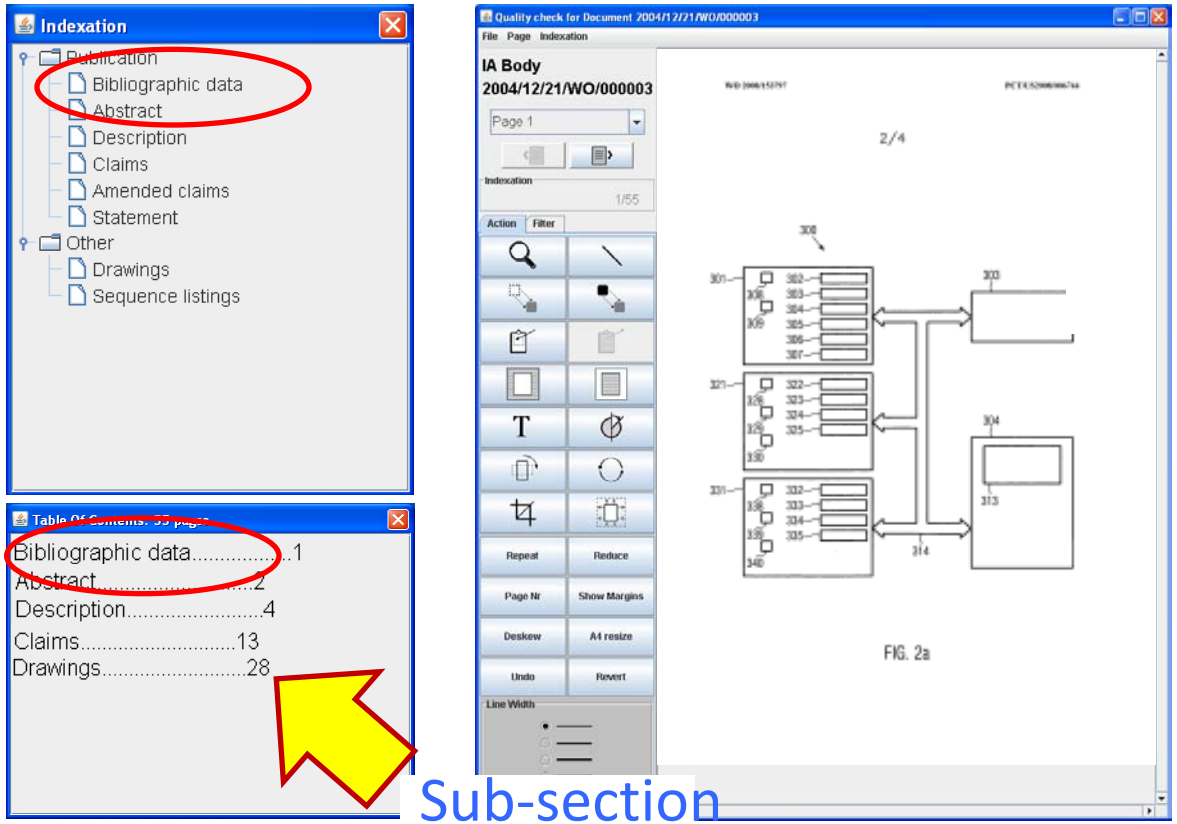
As non-restrictive example, a mobile ground user may be a moving natural person, a moving vehicle, a moving aircraft on the ground or any other moving device on the ground.

As non-restrictive example, a fixed airborne user may be a hot air balloon, airship, helicopter, geostationary

WORLD INTELLECTUAL PROPERTY ORGANIZATION

Editing Scanned Documents

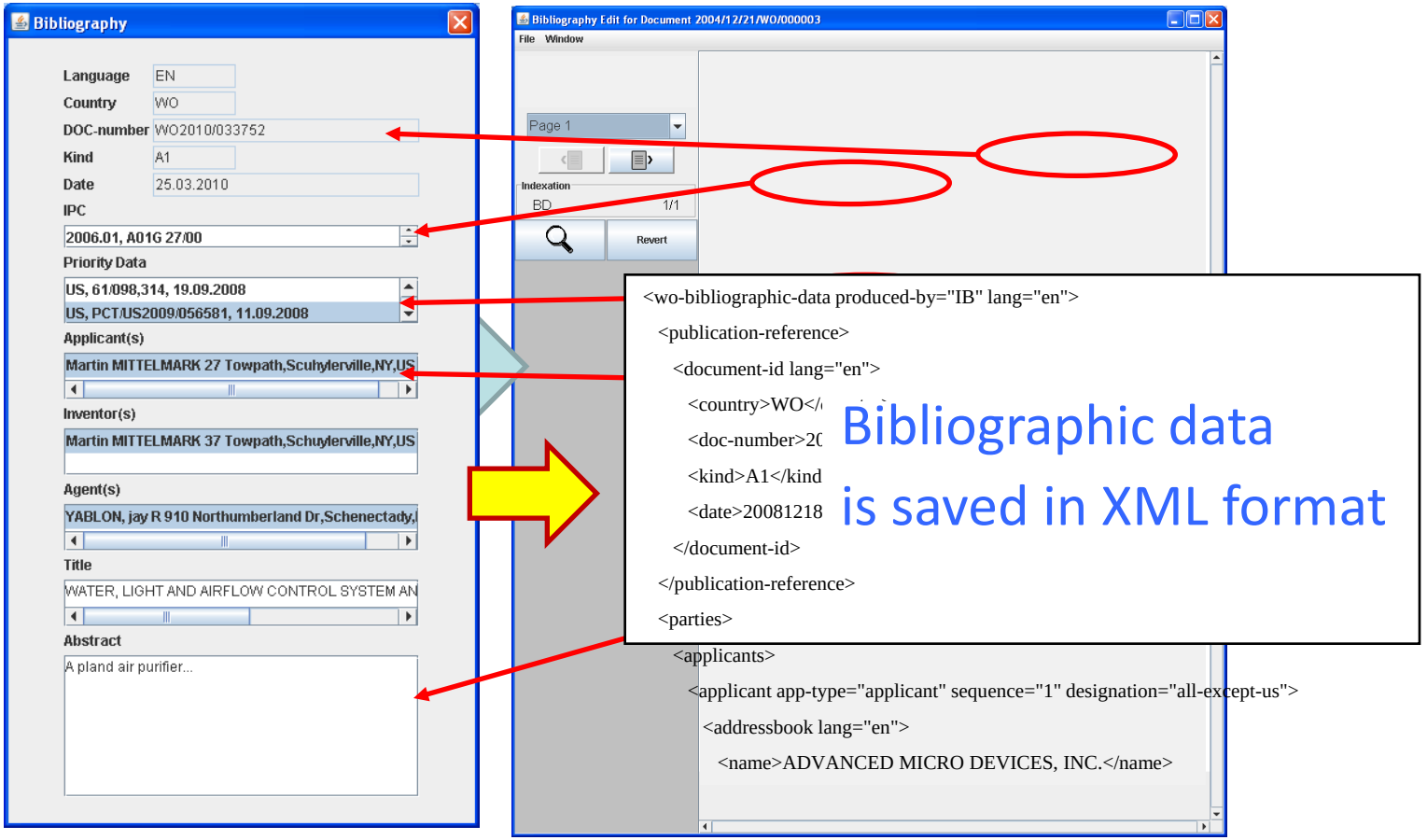
- Index Sub-section



Sub-section Bookmark

WIPO WORLD INTELLECTUAL PROPERTY ORGANIZATION

Generation of Bibliographic data



Contents

- DEPLOYMENT CONSIDERATIONS
  - Deployment Status
  - Deployment Strategy
  - Future Direction

Deployment Strategy

Example Assumptions on costing

- WIPOScan data will be sourced from scanned documents & existing systems (or not perhaps OCR licence for bibliographic data capture)
- Networked solution
- 10 users
- Backlog scanning to be outsourced
- Selection timescale: 2 months
- Implementation timescale: 1 – 4 months

Indicative Costs

- Software licences
- Hardware costs
- Backlog scanning (sample costs from supplier if outsourced)
  - Scanning documents up to A3 - $0.80 per page
  - A4 scanning - $0.50 per page
  - Preparation of documents pre-scanning (unfolding, destapling etc) - $10 per hour
  - Indexing - $5.50 per 1000 keystrokes
- Temporary workers
- Training costs

Please note that these are just some of the basic candidates for costing. The actual costs may be higher / lower depending on:

- Functionality
- Scale of data to be captured/ stored
- Level of access (e.g. remote or local)
- Range of documents and IP domains to be captured
- Number of user licences
- Complexity of solutions
- Implementation timescales

WIPO
WORLD
INTELLECTUAL PROPERTY
ORGANIZATION

Scanning Preparation

Some key questions need answering to determine configuration and cost of solution

Needs

- How many Documents to store?
- Number of users
- Access (remote, local, networked)?
- Business problems to be resolved?
- Type & size of network?
- Who does the backlog scanning?

Determines

- H/w configuration, storage size, h/w costs
- S/w costs
- S/w costs, security features
- Which modules to deploy & OCR licence?
- System configuration
- Implicit or explicit Cost of scanning

- Based on existing implementation templates
- New and unique configuration to specific office
- Local tendering vs. international purchase of software and equipment
- In-house scanning vs. outsourced

Pilot Implementation

- Start small (perhaps registered & published documents) to allow procedures to be developed and tested
- Training of admin + users

Full Implementation

- Take all historic records including born-digital documents (convert to tiff)
- Backlog scanning of all paper based records
- Training of systems administrators and end users
- Implement full network version

Benefits of WIPOScan

- Minimizes storage, retrieval and workflow management
- Cost savings on data entry, filing and personnel management
- Operational efficiencies (minimizes errors, quick retrieval, and is not labor intensive in full operation)
- Customer Service efficiencies
- Reduction in volume of paper and need to photocopy
- Sharing of information quickly and to several individuals at once
- Secure documents electronically minimize loss due to damage or disaster

WIPO
WORLD
INTELLECTUAL PROPERTY
ORGANIZATION

WIPOScan involves the migration of paper and electronic documents or reports onto an electronic storage medium and provides the ability to easily retrieve the information using an indexed search in bibliographic data and abstract. The diagram below shows the five basic components of WIPOScan.

- Determine size of collection
- Determine quality of paper
- Determine requirements for bibliographic data (import from IP Admin sys or Capture/ OCR)
- Organize paper for scanning
- Move docs to scanning point
- Remove duplicate docs/paper
- Prepare docs for scanning

- Any scanning source
- TIFF images 300 dpi
- Batch scanning
- Simplex or duplex mode
- No page limit
- Paper documents are usually labeled, sorted, indexed, placed in folders & filed in cabinets
- *Electronic documents are handled in a similar manner*
- Indexing must allow ease of use & be easily understood
- Indexing include document reference & folder structure

- Document type Indication
- Document section indexing e.g. bibliographic data, description, claims, drawings
- Image cleaning and editing including deskew, removal of dirty marks, spots
- Alignment of margins

- Capture of bibliographic data
- OCR of bibliographic data
- Import of bibliographic data
- Export to IP Admin systems
- Export to external media; data exported in WIPO ST. 36 format

- Documents once brought into the system must be stored
- Uses non-proprietary and widely used storage standards & format i.e. xml, tiff, mysql, pdf, jpg
- Storage devices include hard drives, optical, and tapes
- Retrieval is where an indexing system pays off
- Systems creates searchable CD/DVD capable of bibliographic data search + abstract

Future Direction

- **Cost effective System to:**
  - Lower total cost of ownership (open source)
  - Locally deployed and maintained
  - Reduced training costs and maintenance

- **Smarter IP Office**
  - Interface with EDMS
  - First call for online products / services
  - Providing source code to the IP office for future customizations

Thank You

WIPO
WORLD
INTELLECTUAL PROPERTY
ORGANIZATION