MAGPIE

Metadata Assignment GUI Providing Ingest and Export

Stephanie Elmquist, William Welling (corresponding author), James Creel, Jeremy Huff, Jason Savell, Rincy Mathew, Doug Hahn, and Michael Bolton

ĀĪM | LIBRARIES
TEXAS A&M UNIVERSITY

Magpie

- **Magpies** are <u>birds</u> of the <u>Corvidae</u> (crow) family, including the black and white <u>Eurasian magpie</u>, which is considered one of the most intelligent animals in the world, and the only non-mammal species able to recognize itself in a <u>mirror test</u>.[1] In addition to other members of the genus *Pica*, corvids considered as magpies are in the genera *Cissa*, *Cyanopica* and *Urocissa*.
- MAGPIE is an acronym for "Metadata Assignment GUI Providing Ingest and Export"

Magpie. (n.d.). In *Wikipedia*. Retrieved May 12, 2016, from http://en.wikipedia.org/wiki/Magpie

Objectives

- Intuitive and efficient interface for managing metadata annotation of digital content
- Consume external resources to provide definitive records and suggestions
- Providing single and batch export into Institutional Repository (IR), i.e. DSpace
- Generation and export of archival cataloging spreadsheets for preservation

Use Cases

- Legacy dissertations
- Image collections
- Newspapers and articles
- Etc.

Implementation

- Using A&M developed Weaver framework
- Distributed architecture
  - Scanning workstation
  - ABBYY OCR
  - Authorization Service *Weaver*
  - Webservice *Weaver*
  - User Interface *Weaver*
  - Institutional Repository
  - Archivematica

  * Weaver Webservice Core   * Weaver UI Core
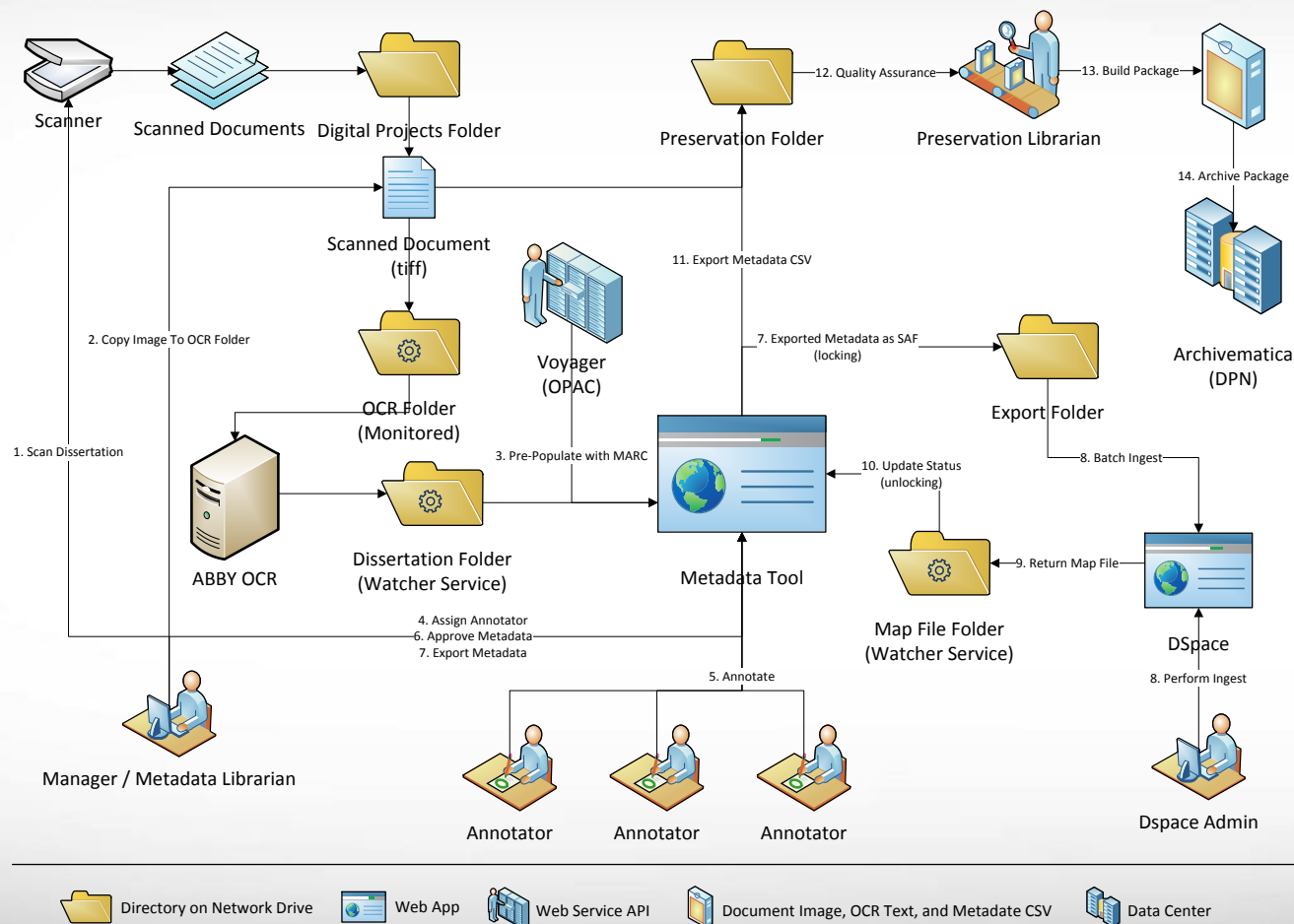
Project Configuration

- External JSON for metadata input fields
  - label, gloss, repeatable, and input type
- Network Directory Watcher (Polling)
  - New project directories
  - Project folder for scanned documents
  - Map folder for DSpace ingest map files
- External Services
  - Voyager for MARC records

Workflow

- Provide Digital Content to Watcher

  – For example, scan/ocr print document

- Automatic annotation from definitive sources

- Assign, Annotate, Review, and Approve annotations

- Export for ingest into IR or directly ingest

- Export preservation spreadsheet

Workflow Diagram

Metadata Tool: Dissertation Project Workflow

Metadata

- Labels
  - Authors
  - Subjects
  - Abstract
  - Publication dates
  - Etc.
- Justification
  - Discovery
    - Indexing
    - Categorization
  - Preservation

Benefits for IR Content

- DSpace is the first IR to which MAGPIE exports
  - via CSV (with supporting tools), SAF, or REST API
- Facilitates controlled vocabulary
- Normalizes metadata
- Allows opportunity for curation of new and legacy metadata
- Custom forms for specific projects

Brief Demonstration

- User management
- Document management
  - Assign
  - Annotate
  - Review
  - Publish
- Export
  - CSV
  - SAF

Future Direction

- Semi-automatic abstract, author, and citation extraction
- Statistical Classification
- Topic Modeling

Thank You.

Any Questions?

Stephanie Elmquist – s-elmquist@library.tamu.edu
William Welling – wwelling@library.tamu.edu
James Creel – jcreel@library.tamu.edu
Jeremy Huff – huff@library.tamu.edu
Jason Savell – jsavell@library.tamu.edu
Rincy Mathew – rmathew@library.tamu.edu
Doug Hahn – dhahn@library.tamu.edu
Michael Bolton – michael.bolton@library.tamu.edu

ĀĪM | **LIBRARIES**
TEXAS A&M UNIVERSITY