WP6- Analysis in the Visual Domain

17/05/2012 – London, IPTC Metadata Conference
@CEPIC Congress

www.arrow-net.eu

It’s all about metadata...

- As you’ve just seen, the ARROW system addresses the issue of Orphan Works but goes beyond it towards a comprehensive Rights Information Infrastructure to search for information on all categories of books

ORPHAN WORKS ARE AN ISSUE OF MISSING OR INCOMPLETE INFORMATION

- ARROW started addressing textual material only in the form of books ARROW plus is addressing visual materials embedded in books with a feasibility study and a pilot implementation

ARROW TAKES IMAGES EMBEDDED IN BOOKS SERIOUSLY AND IS LOOKING FOR A VIABLE SOLUTION INVOLVING THE STAKEHOLDERS COMMUNITIES
...but the real world is complicated

Art books  Travel guides  Books on historical photography

Fiction and non fiction: cover and inside  Children’s books

AND MANY MORE

...and legal issues can make it even worse
What has been done

- D6.1 – Analysis for integration of artists’ data
  - Assessment of the sources of information included in the current ARROW workflow as far as coverage of visual materials is concerned
    - National Libraries catalogues and authority files
    - Books in Print catalogues
    - RROs repertoires
  - Analysis of domain specific sources of information, with the perspective of enhancing the current ARROW workflow
    - visual CMOs
    - picture agencies and libraries
  - First phase - ARROW pilot countries: UK, Germany, France, Spain
  - Second phase - ARROW Plus target countries: Austria, Belgium, Bulgaria, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, the Netherlands, Poland, Portugal

Key findings – NL catalogues and authority files

- Book entries in National libraries catalogues include exceptionally information on images and their author/right holder
  - When book’s subject is a visual author or a specific visual work, then the visual artist also has an authority entry
  - If illustrations are contained in a book the entry includes a general mentioning (such as: „illustr.”) without any further specification
  - Libraries’ information on embedded images not complete
- Completing the information provided by library by hand possible but work and cost intensive
Key findings - BIPs

- Books in Print databases hold “publisher-oriented” information needed to the book supply chain
  - Good metadata about the book as tradable item
    - Structured and standardised way of expressing contributors roles (illustrator, photographer, etc.) and contributors names
    - Structured and standardised way of expressing the types of edition (illustrated and type of visual material included)
    - Few information about rights ownership
    - Collect only information on book cover illustration
  - No information on embedded image inside a book
  - Practices largely vary from country to country and from publisher to publisher
  - Extension to all images embedded possible because publishers should hold all information on the contents, but this is largely depended on “paper” publishing contracts and difficult to track down for older editions

Key findings – RRO and visual CMOs

- RRO/CMO hold relevant information
  - Hold members directories in IPI standard (not public)
  - Rights holder quality is controlled
    - Search for stored author/rightholder data based on search submitted by library search for author/rightholder
    - Name needed
    - Some CMOs have image databases and could use visual recognition software.
Key findings - Picture Agencies

- Picture agencies hold a wealth of information
  - Information about rights holder
  - Elements of licensing history
  - Date of first publication and individual publications that have used the image
  - In some cases orphan works status
- Information may be available in a variety of databases
  - E.g., information about the work is kept in one database, licensing information kept in bookkeeping department; separate database for orphan works
- Integration of these databases necessary for information to be efficiently retrieved in ARROW
- Data is usually held in not accessible databases
- Survey shows circumspect interest in providing information to ARROW

To sum up: good news

- Information that can be retrieved from the actual ARROW workflow
  - A generic statement that a book contains visual material
  - When a visual artist is a "main contributor":
    - Names and names variants including pseudonyms
    - VIAF identifier and NL authority record ID
    - Dates of birth and death
    - Nationality
  - Bibliographic information on the work and its manifestations where visual materials are published:
    - Titles
    - ISBN and NL catalogue ID
    - Publisher and publisher identifiers
    - Country of publication
To sum up: bad news

- Information that cannot be retrieved from the actual ARROW workflow
  - Information on visual artists when they are not listed as main contributors
  - Information on the single image or visual material embedded in the book

Next steps - Feasibility study

- Is it possible to enhance the ARROW workflow by integrating visual material-oriented data sources?
  - Visual CMO databases
    - Members’ Repertoires
    - Additional database to be set up from license information and claims for remuneration rights
    - Setting up of work registry
  - Picture databases
    - Repertoire of represented photographers
    - Open up and make interoperable internal databases
    - Role for visual image recognition technologies
  - Visual authors associations
    - Members directories
Next steps - Piloting

- Piloting of ARROW with image material search
  - France, Germany and UK (tbc)
  - Case studies jointly prepared per countries
  - Preparation of a draft workflow including image
  - Selection of test cases (records of books with increasing level of complexity)
- Technical integration
  - What sources?
  - How can they be queried?
  - What query keys are needed (ISBN, titles, authors names, authors IDs)?
  - What information can be retrieved and in what formats?
  - Messages choreography and workflow

Paola Mazzucchi – Project Coordinator
Associazione Italiana Editori
Paola.mazzucchi@aie.it

FURTHER INFORMATION on WP6
Carola Streul - EVA
c.streul@evartists.org
Sylvie Fodor – CEPIC
S.fodor@cepic.org

ARROW Plus is a Best Practice Network selected under the ICT Policy Support Programme (ICT PSP)
About ARROW Plus

- ARROW plus
  - Start date: 1st April 2011
  - End date: 30th September 2013
  - Full Partners: 26
  - Supporting partners: 10
  - ARROW Plus target countries: Austria, Belgium, Bulgaria, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, the Netherlands, Poland, Portugal

- Objectives of ARROW Plus
  - Objective 1: inclusion of further countries
  - Objective 2: enhancement of the system
  - Objective 3: piloting the inclusion of images