

ARROW and ARROW Plus: activities coordination

Giulia Marangoni, Krakow, 10 December 2013

www.arrow-net.eu

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



Management structure

ARROW Plus activities have been organised in Work packages (WPs) addressing **three main objectives**:

□ **Objective 1: inclusion of further countries**

- } Coordinating national initiatives (FEP) **WP3**
- } Creating conditions for inclusion – design of new registries (MVB) **WP5**
- } Connecting new data sources to ARROW infrastructure (CIN) **WP4**

□ **Objective 2: enhancement of the system**

- } Enhancements (CIN) **WP4**
- } Infrastructure for registries (CIN) **WP4**
- } Validation (UIBK) **WP7**

□ **Objective 3: feasibility study for the inclusion of images**

- } Dedicated efforts, resulting in a feasibility study (EVA) **WP6**

Inclusion of new countries

Inclusion of new countries: TEL



- Inclusion of National Libraries catalogues in TEL
 - } Analysis of local MARC formats and cataloguing rules (UNIMARC, RUSMARC, HUNMARC); ad hoc mappings, when necessary
 - } Fine tuning of matching and clustering algorithms on country basis
 - } Implementation of tools for workflow configurability (matching and clustering preferences) to serve use cases specificities
- **12 new countries** included, for a **total of 16** countries fully integrated in ARROW and ARROW Plus projects

Inclusion of new countries: BiP



- Inclusion of existing Books in Print
 - } Analysis of data coverage (books and e-books) and semantics (publishing and availability status)
 - } Design of query methods (ISBN and work queries) and internal processing to output clusters of manifestations enriched with publishing and commercial status data
 - } Implementation of ad hoc routing mechanisms on the French, Belgian and Dutch workflows, basing on the language of books (Flemish, Dutch and French)
- **6 new countries** (AT, IT, IE, NL, BE, BG) included, for a total of **10 countries fully integrated**
- **6 countries** adopting the **ARROW BiP software** (GR, HU, LT, LV, PL, PT) will be connected to the system by the end of December 2013
 - } To complete the ARROW integration the new registries shall be fed with books in print data

The ARROW BiP software

- The **ARROW BiP Registry open source software** will be released in 6 countries for independent management and maintenance
 - } Portugal, **Poland**, Latvia, Lithuania, Hungary, Greece
- The BiP software includes the functionality for ARROW Integration
- Licensing terms are delivered together with the software package
- The system have been populated with automatic data-feeds (bibliographic and publishers data) to minimize the manual workload for national partners and publishers
- **Next Steps:**
 - } Local deploy of BiP software and finalisation of national sustainability plans
 - } Enrichment of BiP system with commercial data from the book trade
 - } A working group with NCP has been established to discuss and agree possible additional developments to the sw after it's been delivered and beyond the project end

Inclusion of new countries: RRO



- Inclusion of RROs
 - } Analysis of coverage and structure of repertoires (title vs. not - title specific)
 - } Design of query methods and internal processing to answer to user requests
 - } Each RRO defines business rules and according to the key decision elements adopted by the RRO
 - Whether the RHs is known or not and who to contact
 - Whether a licence can be granted or not
 - } Implementation of an alternative workflow (WF-B) for some RROs who license the BiP catalogue
 - } Depending on the RRO and the use cases: answers can be synchronous or can be provided after the manual completion of the diligent search process
- **4 new countries** included (AT, BE, IE, IT) for a total of **8 countries**
- A module for ARROW integration has been implemented in WISE, the IFRRO standard software for RROs

The IFRRO Standard Software



- The development of an RRO software was originally foreseen in ARROW Plus project
- It was brought to the attention of the project management an initiative lead by IFRRO (the International Federation of RROs) with the objective of developing a standard software to be used by RROs – thus completely overlapping the scope of the ARROW RRO registry
 - } the IFRRO Standard software has been chosen as the system ARROW suggests to adopt to the RROs that need to set up a technical infrastructure
 - } Integration with ARROW has been completed and the IFRRO software has been piloted successfully in Ireland

ARROW Plus Performance indicators

□ PERFORMANCE INDICATORS

- } Number of stakeholders association involved
- } Number of data sources made interoperable
- } Number of new registries created

Performance indicator	Expected results (M30)	Actual result (M29)
Number of stakeholders associations involved	32	<50
Number of data sources made interoperable	13-15	24 (expected 28 at M34)
Number of new registries created	3-6	4 (expected 6 at M34)



System enhancements

Enhancements: methodology



- Following up the validation of the system released at the end of the ARROW project (2011), some key enhancements have been identified:
 - } Enhancements in data processing to support a better identification of rightsholders and RHs data enrichment
 - } Enhancements in data processing to support a better determination of the copyright status of a work
- Further enhancements have been identified
 - } While gathering requirements for technical integration of new countries
 - Ex. To manage transliterated names forms within ONIXRS messages
 - } While collecting requirements for emerging use cases
 - To increase the system performance to support high volume of queries
 - To increase the system flexibility, allowing for workflow configurability according to user needs

Identification and rightsholders data enrichment

- In case where not all the contributors names are registered in the appropriate structured fields, names are extracted from the natural language text in statements of responsibility (MARC field 245\$c) to complete list of relevant RHs for ARROW. The process applies to UK, Ireland, Germany, the Netherlands, Belgium, France
 - } Parsing of free text string to extract contributor names from MARC field 245\$c
Preliminary result: **88.2% success**
- Contributors names are matched with VIAF records even in cases where the author authority ID is missing in VIAF (ex. when the national library is not a participant of VIAF, or the contributors names have been extracted from free text strings). Dates of birth/death extracted from VIAF further enrich the ARROW data processing
 - } VIAF integration:
Preliminary result: **82.5% success**

Copyright status determination



- According to the Harmonised copyright rule, when a work is anonymous or with a corporate authorship, terms of protection shall run for 70 years after the work is lawfully made available to the public (i.e. work date of first publication). To manage this case, the oldest publication date is identified and extracted from TEL clusters of manifestations and then processed by the copyright algorithm.
 - } Work date extraction:
Preliminary result: **96.3% success**
 - } Public domain algorithm calculation for Anonymous, Collective or Corporate contributors
Preliminary result: **99.2% success**

Validation

- Validation of system enhancements has been completed and the preliminary results are positive
- Validation of national workflows for all the countries with the collaboration of National Contact Points is ongoing:
 - } General objective: assessment of the ARROW system performance by comparing ARROW results (copyright status and publishing status determination) with the result of the manual diligent search
 - } In depth analysis of matching and clustering process in each country
 - } Estimation of time saved using ARROW based on a real exercise
- Feedback from partners and other stakeholders is being collected via surveys and interviews
- Results of all validation activities will be reported in the final Validation Report D7.2 due at the end of the project

Use cases

Use cases: methodology

- Anytime a new use case emerged during the projects (i.e. someone asks to use the ARROW system)
 - } the use case is brought to the attention of the project Board (in future to the ARROW Association Board), and if approved
 - } a working group is set up with the stakeholders involved in the use case and AIE, finalised to the use case requirement analysis and specifications
 - } requirements are analysed to identify:
 - Enhancements that could be beneficial to the whole ARROW system
 - Customisations that are beneficial only to the specific use case
 - } outcomes of the analysis work allowed AIE to coordinate with ARROW Plus technical partners TEL and CINECA/mEDRA for the design and the implementation of the specific requirements for the use case

Wellcome Trust Library

- Status: **concluded**
- Characteristics
 - } **Multi-national collection**
 - } The value is in **finding the rightsholders that give the permission**
 - } **Manual handling of the requests by the RROs involved**
 - Additional manual search by the RROs
 - Contact with the rightsholders to get the permission
- Enhancements and customisations developed
 - } Improvement in the initial matching
 - } Special **routing mechanism** to allow all requests under this use case to reach the UK RRO for the RRO process, regardless the country of publication of the book
- Lesson from a business viewpoint
 - } ARROW provides **value to the RRO**, which is **innovating its business model**

British Library - WW1 project



- Status: **concluded**
- Characteristics
 - } Multi-national collection on **specific theme**
 - } The value is in identifying different categories of works (by country, by copyright status, by commercial availability) starting with very few information
 - } Manual handling of the requests by the RROs involved
- Customisations developed
 - } New software to identify the country of publication from BL records: considering the Place of publication (i.e. city, town, geographic area) and using Geonames the corresponding country has been added to BL records
 - } Special **routing mechanism** to allow all requests under this use case to reach the UK RRO for the RRO process, regardless the country of publication of the book
- Lesson from a business viewpoint
 - } ARROW provides **value to the final user (BL) and at the same time to the RRO**, which is innovating its business model

French Out of commerce DB (ReLIRE)

- Status: **started in Dec. 2013**
 - } In the last 3 years: preliminary studies carried on with French stakeholders
 - } In 2013: from functional and technical specifications to system implementation
- Characteristics
 - } **Very large database (2 millions records to be processed)**
 - } Search limited to **the identification of out of commerce**
 - } Implementation of the new **OOC French law**
- Enhancements and customisations
 - } Actual reengineering planned to increase **performance** and **flexibility**
 - } Refinement of **clustering** process and works relations
 - } Adoption of specific **out of commerce criteria**
- Lesson from a business viewpoint
 - } ARROW **provides value** to the one actor (the **BNF**) in a broader picture
 - } **All countries benefit from system enhancements**

ARROW in the future

The ARROW Association

- In September 2013 the ARROW Association has been set up as a membership organisation under the Italian law, holding the IP Rights of the ARROW system
 - ┌ The Association is an open room for all stakeholders to continue collaborating for shared purpose
 - ┌ Stakeholders = authors, publishers, libraries, RROs/CMOs, Books in Print databases, EU associations
 - ┌ All project partners will be formally invited to join

- The future of ARROW lays in stakeholders collaboration and the capacity of the association members to identify and collaborate to serve new use cases

The sustainability model

- The value proposition
 - } ARROW facilitates the rights management in large scale digitisation programmes which include in-copyright works
 - } Facilitating = time/cost saving
- Mere maintenance ensured by membership fees
 - } Stakeholders take the responsibility
- ARROW proved to be a flexible system that can serve different projects needs
- Revenues from use depending on the services provided
 - } Open to different models (per service, subscription by the national authority...)

How it works in practice

- A new use case is brought to the attention of the ARROW Association (request directly from user with possible support of National members)
- A working group is set up with the stakeholders involved in the use case, finalised to the use case requirement analysis and specifications
- Requirements are analysed to identify
 - } Enhancements that could be beneficial to the whole ARROW system
 - } Customisations that are beneficial only to the specific use case
- Costs will be assessed according to ARROW business model
- Outcomes of the analysis will feed into the design and the implementation of the specific requirements for the use case

So: the future

- The ARROW Association is very well positioned as provider in large-scale digitisation plans, including orphan works or out-of-commerce
- So, the future depends on two variables
 - } The number of digitisation plans including in-copyright works
 - } Our capacity to confirm our current reputation