

D4.1 Interim report on system maintenance and enhancement

Project Acronym: ARROW Plus
Grant Agreement number: 270942
Project Title: European Registries of Rights Information and Orphan Works

Revision: 3

Authors:

Gabriella Scipione (CINECA)
Nuno Freire (The European Library)
Cinzia Caroli (CINECA)
Nicola Ghirardi (CINECA)
Elda Rrapi (mEDRA)
Giuseppe Trotta (mEDRA)



Project co-funded by the European Commission within the ICT Policy Support Programme		
Dissemination Level		
PU	Public, for wide dissemination	x
CO	Confidential, only for members of the consortium and the Commission Services	
RE	Restricted to a group specified by the consortium	

REVISION HISTORY

Revision	Date	Author	Organisation	Description
Draft 1	2012-03-09	Elda Rrapi	mEDRA	First draft
Draft 2	2012-03-15	Cinzia Caroli	CIN	Added Section 1 and 2
Draft 3	2012-03-23	Giuseppe Trotta	mEDRA	Added Section 6
Rev 1	2012-03-27	Nuno Freire	TEL	Added Section 5
Rev 2	2012-03-28	Gabriella Scipione	CIN	Overall revision of document
Rev 3	2012-05-15	Cinzia Caroli	CIN	Adjustments after Cedro-APEL's Review

Statement of originality:

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

TABLE OF CONTENT

EXECUTIVE SUMMARY	1
Purpose of the document	1
Relationship with other documents	2
Structure of the document	2
1. THE ARROW SYSTEM	3
1.1 The main components	4
1.1.1 Arrow Web Portal Services	4
1.1.2 Arrow DataCentre	4
1.1.3 The Rights Information Infrastructure	4
1.1.4 TEL service	5
1.1.5 The BIP and RRO services	5
1.2 Workflow	5
1.2.1 The TEL process	7
1.2.2 The BIP process	8
1.2.3 The RRO process	8
2. SYSTEM MAINTENANCE	9
3. OVERVIEW OF THE ENHANCEMENTS	10
3.1 Workflow enhancements	11
3.1.1 Management of books that does not find a match in the TEL catalogue	12
3.1.2 More granular classification of the match types	12
3.1.3 Automatic matching validation process for "exact" matches	12
3.1.4 Manual matching validation process for "primary work" matches	13
3.1.5 Automatic matching validation process for "partial" matches	13
3.1.6 Management of books with all proposed matches rejected	13
4. ENHANCEMENTS OF THE WEB PORTAL	14
4.1 ARROW PLUS Intranet	14
4.2 ARROW Intranet archive	15
4.3 User account management	15
4.4 Website public section	15
4.5 Register interest form	16
4.6 Homepage	17

5. ENHANCEMENTS OF THE EUROPEAN LIBRARY SERVICE	18
5.1 EVOLUTIVE MAINTENANCE AND NATIONAL USE CASES	18
5.2 INTEGRATION INTO TEL'S PRODUCTION SYSTEMS AND INTERNAL WORKFLOWS	19
5.2.1 Source code restructuring and documentation	22
5.2.2 Integration with the TEL central repository	22
5.2.3 Integration with the TEL authority file	22
5.2.4 Reimplementation of the matching and clustering indexes based on SOLR	23
5.2.5 Implementation of Work Metadata based on the internal TEL data model	23
5.2.6 Catalogue ingestion internal workflow	23
5.3 IMPROVEMENTS IDENTIFIED IN THE VALIDATION REPORT OF PROJECT ARROW	24
5.3.1 Extraction of work contributors the natural language text in statements of responsibility	24
5.3.2 Matching contributors of works with VIAF records of persons	24
6. ENHANCEMENTS OF THE RII	27
6.1 Workflow extensions	27
6.1.1 Management of "No match" case	27
6.1.2 Management of automatic validation	27
6.1.3 Business process model of the enhanced workflow	27
6.2 Update of ARROW schemas to ONIX-RS 1.0	29
6.2.1 CINECA's activities	30
6.3 FrontEnd web interface	30
6.3.1 Diligent Search Results for different User Profiles	30
6.3.2 Display of ARROW Availability/ARROW Publishing status	32
6.3.3 List of permission sets refined	32
GLOSSARY	33
LIST OF REFERENCES	36

LIST OF FIGURES

Figure 1 ARROW system	3
Figure 2 ARROW workflow diagram.....	6
Figure 3 The TEL system for ARROW before the start of project ARROW Plus.....	20
Figure 4 The current TEL system for ARROW.....	21
Figure 5 Workflow diagram in ARROW before the enhancements.....	28
Figure 6 Workflow diagram in ARROW plus after enhancement	29
Figure 7 FrontEnd Results Page	31

LIST OF TABLES

Table 1 Main activities of the first 13 months	11
Table 2 Summary table of workflow enhancements.....	12

EXECUTIVE SUMMARY

One of the outcomes of the ARROW project was the creation of the Arrow system: an infrastructure allowing a network of databases to be made interoperable through the use of standards in order to collect and elaborate the information about the rights of books.

At the end of the project, the ARROW system went through to the evaluation process, which enabled the identification of key points to be addressed for ARROW enhancement. The results of the evaluation process were presented in the report "D7.2 Validation Report" delivered at the end of the ARROW project.

Moreover between the end of ARROW project and the begin of ARROW Plus project, a national use case showed up, the one of Wellcome Trust, requiring in turn some customisations of the ARROW national workflow.

All these requirements have been merged into Work Package 4 (WP4) of the ARROW Plus project, and implemented as described in this document.

The new features allow to maximise the number of requests that can be processed and, at the same time, minimise the human intervention in the ARROW workflow, to make it more suitable for mass digitisation projects and to improve the user experience when using the system.

The components of the ARROW system that have been involved are:

- the ARROW Web Portal Services
- the TEL service
- the Right Information Infrastructure (RII)
- the BiP and RRO services

The implementation work is still continuing and will continue till the end of the ARROW Plus project, as foreseen in the Work Package 4, in order to integrate new countries in the ARROW workflow, to set up the new BIP and RRO registries as per WP5 technical and functional specifications and to implement the needed customisations for forthcoming use cases.

It is worth noting that CINECA as WP4 leader has also the responsibility to ensure the maintenance of the ARROW system during the project.

PURPOSE OF THE DOCUMENT

This document presents the description of the enhancements implemented in the first 13 months of the ARROW Plus project. It describes briefly the ARROW system as it was before the start of the ARROW Plus project. Moreover it gives a general overview of the enhancements and describes the new features in detail.

RELATIONSHIP WITH OTHER DOCUMENTS

This document is based on the following documents:

- “D7.2 Validation Report” delivered at the end of the ARROW project
- “D2 1 Reviewed ARROW Website_final.pdf” delivered by IFFRO in ARROW Plus project.

The first document, and specifically the section "System enhancement plan", contains the recommendations from the early adopters (pilot libraries, survey participants) and the outcomes of the performance measurement along with the experiences made in the pilot implementations of the ARROW project.

The second document is the Deliverable 2.1 of ARROW Plus, produced by IFFRO. It outlines the requirements for the changes needed in the ARROW website for the transition to ARROW Plus project. The main change is triggered by the change of focus on the dissemination of ARROW into presenting the system and the services facilitated by it, enabling potential users to express interest for making use of ARROW and assisting potential as well as actual users to familiarise themselves with the system.

STRUCTURE OF THE DOCUMENT

The document is composed of 6 chapters.

- Chapter 1 contains a short description of the Arrow system before the enhancements.
- Chapter 2 describes briefly the system maintenance.
- Chapter 3 provides an overview of the enhancements.
- Chapter 4 outlines the enhancements of the Web Portal.
- Chapter 5 presents the enhancements of the TEL service.
- Chapter 6 describes the enhancements of the Right Information Infrastructure.

1. THE ARROW SYSTEM

The ARROW System is a comprehensive service to support any diligent search model adopted by libraries, by facilitating the identification of right holders (authors/publishers) and the identification of the rights status of works with particular concern to orphan and out-of-print works.

ARROW System is made up of the following macro components:

- Arrow Web Portal Services
- The Rights Information Infrastructure (RII)
- The ARROW Work Registry (AWR)
- The Registry of Orphan Works (ROW).

Figure 1 shows a schematic representation of the Arrow system. The results and the information, collected during the RII workflow, form the basis for the AWR system and for the ROW system.

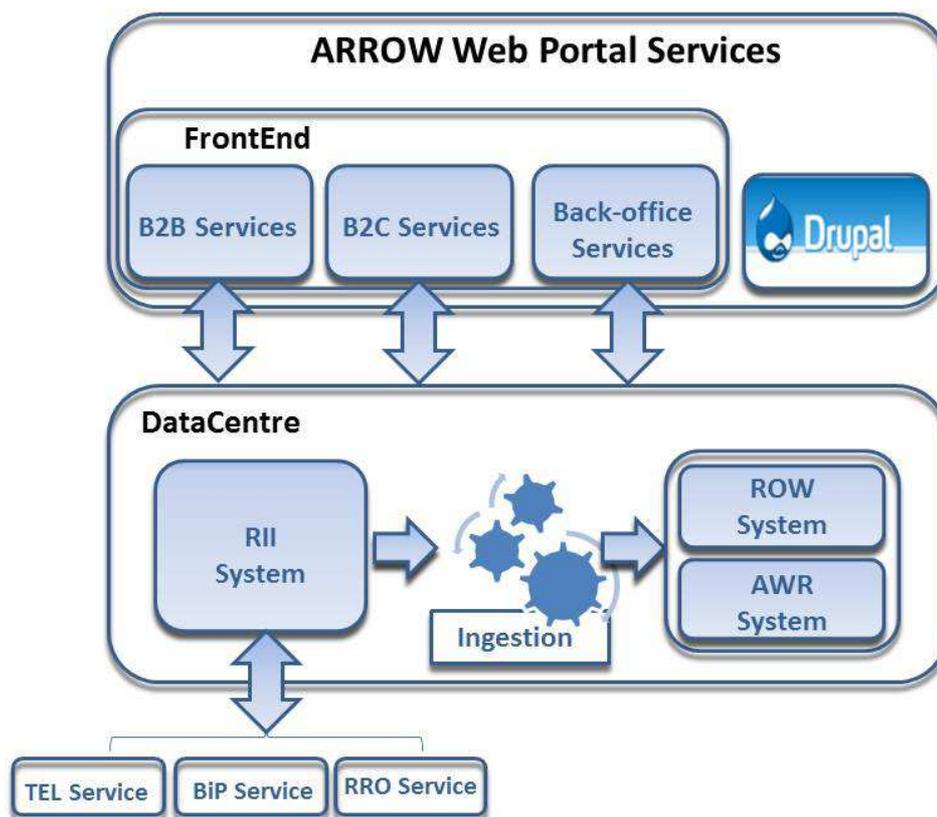


Figure 1 ARROW system

During the ARROW project, it had been defined and implemented a set of messages based on the ONIX format. EDItEUR has designed them in order to be used by various players in the ARROW

workflow and to cover the basic stages in the process, from the initial library request through to the grant or denial of a license.

Since the end of the ARROW project and still now, four pilot countries are included in the ARROW workflow: France, Germany, Spain, UK. For each country it is possible to query the related catalogue of the National Library, of the BIP (Books in Print) and of the RROs (Reproduction Rights Organisation) through the ARROW system.

1.1 THE MAIN COMPONENTS

1.1.1 Arrow Web Portal Services

The Arrow Web Portal Services comprises the following two main components: the FrontEnd and Drupal Content Management System (CMS).

The ARROW FrontEnd is responsible for collecting the initial request from the user, processing it and sending it to the ARROW DataCentre; in other words it represents an interface between the user and the DataCentre. It was also planned that the interaction can bypass the FrontEnd and directly query the ARROW Datacenter Web services (B2B), although at the moment this mode is not used by anyone.

The CMS is a software tool designed to facilitate the management of the web site content. Arrow uses Drupal as CMS, a free and open source CMS distributed under the GNU General Public License. The Website Back Office is conceived as a set of services that allows the ARROW administrators to manage users and their roles, as well as let them publish public and private documents and events.

1.1.2 Arrow DataCentre

The Arrow DataCentre constitutes the back end and performs the business logic of the entire system, including both the RII and the Arrow Work Registry flow. The business logic of the DataCentre is based on a well defined workflow that requires the exchange of information with other external data providers, exposing data via different interfaces and protocols.

1.1.3 The Rights Information Infrastructure

The Rights Information Infrastructure (RII) is the "engine" of the ARROW system.

According to a planned workflow, it queries and retrieves information from a multiplicity of data providers, in multiple formats. It makes the formats interoperable, processes the information and takes decisions on the successive elaboration.

The RII receives the initial request coming from a library. The library asks for the permissions to digitise and to use the manifestation of a work, for instance a book.

After having queried the data providers included in the workflow (TEL/VIAF, Books in Print, RRO) and elaborated the gathered results, the RII provides information on the work rights status.

1.1.4 TEL service

Bibliographic data from the catalogues of Europe's national libraries is one of the key data sources in ARROW. Such bibliographic data is already being aggregated through The European Library (TEL), that's why it has been possible to have a single access point for querying such catalogues and thus to decrease the complexity of the ARROW system.

In ARROW, the European Library system serves two main purposes:

- allow the libraries to identify the bibliographic record describing the manifestation whose rights are to be cleared. The bibliographic record should be identified in the catalogue of The National Library of the respective country of publication;
- identify all other manifestations that potentially share, in part or totally, intellectual work with this manifestation, for further processing in the ARROW workflow.

The TEL service was conceived, designed and implemented by The European Library during the Arrow project with the aim to serve the Arrow system and so it is considered part of the ARROW system.

1.1.5 The BIP and RRO services

The BiP (Books in Print) and RRO (Reproduction Rights Organisation) are the organisations that maintain information necessary to establish the right status of the work. To collect this information, Arrow sends a message (M6Q for the BIP, M7Q for the RRO) to the web service of the BIP/RRO organisations. Asynchronously with respect to the receipt of the message from ARROW, the BIP/RRO systems elaborate the request and send back the response to ARROW, through the messages M6R and M7R.

At the end of the ARROW project, the BIP and RRO of the pilot countries (France, Germany, Spain, UK) were included in the ARROW workflow.

1.2 WORKFLOW

The ARROW RII component is the one in charge to perform the necessary workflow to obtain right status information. The workflow is constituted by three subsequent processes: the TEL process, the BIP process and the RRO process, better described in the following paragraphs.

The initial library request is performed at manifestation level , whereas the response at the end of the workflow is provided at work level. This means that the initial request passes through stages of identification and matching, work and manifestation clustering and the identification of related works and manifestations; each process adds a piece of relevant information towards the identification of the rights status of the work.

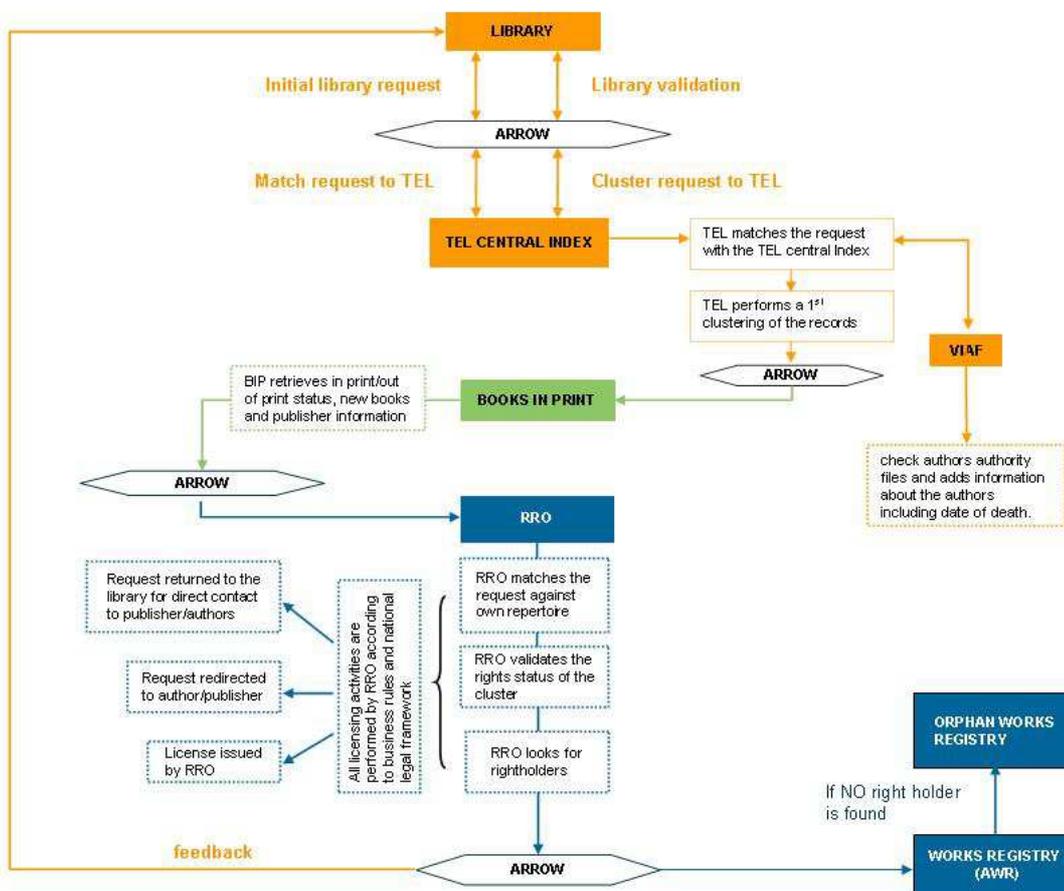


Figure 2 ARROW workflow diagram

At the end of the ARROW workflow, the following pieces of information have been retrieved in the message exchange and stored in the RII repository:

- Work information
- Manifestation information
- Relation between each manifestation and the work they belong to
- Relation between works
- Authors and other contributors information
- Relation between each identified author and the work they have contributed to
- Relation between each piece of information (work, manifestation, author) and the reference source that provided that information (TEL, VIAF, BIPs, RROs)
- A set of so called ARROW Assertions on each work: Copyright Status, Publishing Status and Orphan Status

ARROW Work Registry (AWR) stores and maintains all these pieces of information for every request processed by ARROW.

The Registry of Orphan Works (ROW) is based on a subset of the AWR on the base of specific criteria. It is publicly available to specific categories of users for specific purposes (for more details see ARROW Del 6.4 Rights information infrastructure - final release).

1.2.1 The TEL process

The first stage of the workflow (orange highlighted part of the Figure 2) takes place in the library domain and involves TEL as main actor and VIAF as source for author's information. ARROW exchanges and elaborates the information coming from TEL through two distinct phases:

- Matching (message M2)
- Clustering (message M4)

The Matching process enables the libraries to identify the bibliographic record that describes the manifestation for which permission is requested. The bibliographic record is identified in the catalogue of the National Library of the country where the manifestation was published, by querying the TEL's catalogue that hosts the catalogues of the National Libraries of all the European countries.

The TEL's matching algorithm¹ provides as output a list of bibliographic records that match the metadata of the manifestation of the library.

At this point the ARROW system presents to the user (library) the list of retrieved records and the user has to choose among them the one that has to go on with the ARROW elaboration. This operation was called "manual validation" and the selected record is the "target manifestation".

The Clustering process identifies the records that, besides the target manifestation, potentially share the same work. The records are organised in a main cluster called "PrimaryCluster" and one or more "SecondaryCluster".

The PrimaryCluster contains the manifestations that match title, all contributors and all languages of the target manifestation, while the SecondaryClusters do use less restrictive matching criteria.

At the end of these two phases, the following information is obtained:

¹ The matching algorithm, developed by TEL, takes into consideration the following bibliographic metadata: title, all contributors (regardless the role), all languages, publisher and date of publication. It returns a measure, expressed as a percentage, of the matching of a bibliographic record with the manifestation submitted by the library. On this base, a bibliographic record is said an "exact match", if it has a value of matching greater than a pre-defined and customizable threshold, for example 80%.

- the work to which the original library manifestation belongs;
- a list of manifestations sharing the same work;
- any other related work and the list of respective manifestations;
- a set of authoritative information for each author and other contributor of each work, including preferred and alternative forms of their names, their dates of birth and death, their nationality;
- the copyright status of each work: whether the work is in the public domain or copyrighted, or that this information cannot be certainly asserted.

1.2.2 The BIP process

The second stage (green highlighted part of the Figure 2) takes place in the Books in Print domain and involves the BiP organisations in the countries included in the ARROW system. The data coming from TEL are further elaborated and enriched with the information gathered by the BIP through the message M6.

As output of this stage, the following information is obtained:

- a list of additional manifestations belonging to the work and related works;
- the “in print/out of print” status and the commercial availability of each manifestation belonging to the work and related works;
- the Publishing Status of each work: whether the work is currently active (in print) or currently not active (out of print), or that this information cannot be certainly asserted.

1.2.3 The RRO process

The third stage (blue highlighted part of the Figure 2) takes place in the Reproduction Rights Organisation domain.

After receiving the information from the BIP, ARROW sends to the RRO the library request enriched by all the data at work and manifestation level collected and processed previously. The message M7Q is sent from ARROW to the RRO and the RRO send back the message M7R.

As output of this stage, the following information is obtained:

- a set of information provided by the RRO concerning licensing conditions and reasons supporting the decisions;
- the Orphan Status of the work: whether the work can be considered probably orphan as the rights holders cannot be identified or traced, or not orphan, or that this information cannot be certainly asserted.

2. SYSTEM MAINTENANCE

The Arrow hardware infrastructure is currently being hosted at the CINECA premises in Bologna, with the exception of the external components: the TEL service, the BIP services and the RRO services. CINECA provides both the staging and production environment of the Arrow system and the corresponding databases.

As WP4 leader, CINECA has the responsibility to ensure the maintenance of the ARROW system during the project, which implies the provision of the following services:

- hosting and maintenance
- ordinary maintenance
- integrity and confidentiality of data through the CINECA data centre
- network infrastructure.

3. OVERVIEW OF THE ENHANCEMENTS

At the end of the ARROW project, the ARROW system went through the evaluation process, that produced as results the recommendations from the early adopters (pilot libraries, survey participants), the outcomes of the performance measurement and the identification of key points to be addressed for ARROW enhancement. The results of the evaluation process along with the “System enhancement plan” were presented in the report "D7.2 Validation Report".

Moreover after the Validation Report was delivered, it arose the necessity of additional enhancements stemming from the Wellcome Trust use case.

All these requirements have been merged into Work Package 4 (WP4) of the ARROW Plus project, in order to improve the existing ARROW system.

CINECA, that is also the coordinator of the WP4, has implemented the requested enhancements in collaboration with its subcontractor mEDRA, with Ditech.IT and Beeneer. The technical tasks are tracked by means of task management tool. The enhancements in the library domain have been implemented with the strict collaboration of TEL.

The following table summarises the main activities for the first 13 months of the ARROW Plus project:

Subtasks of Task 4.1	Activities	Date of completion
Frontend new interfaces	<ul style="list-style-type: none"> ■ Diligent Search Results for different User Profiles ■ Diligent Search Partial Results for Basic Users ■ Data entry Interface Enhancements ■ Other updates 	June 2011
Wellcome Trust use case	<ul style="list-style-type: none"> ■ Schema upgrade to 0.27 ■ “No match” management only for UK ■ Automatic validation only for UK 	August 2011
Web Site	<ul style="list-style-type: none"> ■ Structure refinement ■ Intranet archiving ■ User Account management ■ Other updates 	November 2011
Update of ARROW messages schema to ONIX-RS 1.0	<ul style="list-style-type: none"> ■ Schema validation ■ Updating of data model 	February 2012

	<ul style="list-style-type: none"> ■ Update of service component involved 	
Workflow enhancements	<ul style="list-style-type: none"> ■ “No match” management ■ Automatic validation 	March 2012

Table 1 Main activities of the first 13 months

3.1 WORKFLOW ENHANCEMENTS

It is worth describing in more detail the enhancements made to the ARROW workflow. The new features allow to maximise the requests that can be processed and at the same time minimise the human intervention in the ARROW workflow to make it more fit to mass digitisation projects.

Table 2 summarizes the main enhancements in the ARROW workflow that are outlined more in detail in the following paragraphs.

Scenario	Workflow behaviour	System components involved
Management of manifestations that does not find a match in the TEL catalogue (“No match” case)	If the manifestation record submitted to the system has no matching record in TEL, the request is treated as “single manifestation-cluster”, generated by TEL on the basis of the record submitted, and goes straight to the BiP processing phase.	DataCentre, TEL service
More granular classification of the match types	In addition to “exact” and “partial” matches, a new match type called “primary work” has been defined as a partial match sharing with the manifestation record submitted the same intellectual work.	DataCentre, TEL service
Automatic matching validation process for “exact” matches	If the manifestation record submitted has one or more exact match in the TEL catalogue, the one with highest matching probability percentage is automatically selected.	DataCentre
Manual matching validation process for “primary work” matches	If the manifestation record submitted has no “exact” matches but “primary work” matches the system asks the user for a manual validation. Manual validation of proposed matches labelled as “Primary Work match” is still requested.	DataCentre
Automatic matching validation process for “partial” matches	If the manifestation record submitted has neither “exact” matches nor “primary work” matches but “partial” matches they are all automatically rejected	DataCentre

	and the request is treated as a “No match” case.	
Management of manifestations with all proposed matches rejected	If all the proposed matches are rejected, the request is treated as “single manifestation-cluster”, generated by TEL on the basis of the submitted record, and goes straight to the BiP processing phase.	DataCentre, TEL service

Table 2 Summary table of workflow enhancements

3.1.1 Management of books that does not find a match in the TEL catalogue

ARROW Workflow has been extended in order to manage diligent search requests even for books that do not find a match in the TEL catalogue (“No match” case).

If the book record submitted to the system has no matching record in TEL, the request is treated as “single manifestation-cluster”, generated by TEL on the basis of the record submitted, and goes straight to the BiP processing phase. This enables ARROW to gather relevant information in the BiP domain and finally send them to the relevant RRO, instead of aborting the workflow.

Consequently the status "No Match result" does not exist anymore and all requests automatically go through the workflow.

3.1.2 More granular classification of the match types

The matching process has been refined in order to provide a more granular classification of the match types. In addition to “exact” and “partial” matches (already managed by the system), a new match type called "primary work" has been defined as a partial match sharing with the book record submitted the same intellectual work (i.e. title, all contributors and all language matches). This refinement, in addition to the information provided on matching probability percentage, is designed to support the set up of automatic procedures for the matching validation process in ARROW, as described in points 3.1.3, 3.1.4 and 3.1.5.

Consequently when a book record is submitted to the system, the matching algorithm compares it against the TEL catalogue and returns the list of proposed matches classified according to the match type and the matching probability percentage.

In case the user is asked to perform a manual matching validation, in the list of proposed matches, he will find items labelled as "Primary Work match" along with items labelled as “Partial match” if present.

3.1.3 Automatic matching validation process for "exact" matches

ARROW Workflow has been extended in order to manage automatic procedures for the matching validation process in case of "exact" matches.

If the book record submitted has one or more exact match in the TEL catalogue, the one with highest matching probability percentage is automatically selected.

So when you query the ARROW system you do not have to perform the matching validation step and the request automatically goes through the workflow.

3.1.4 Manual matching validation process for "primary work" matches

ARROW Workflow has been extended in order to continue supporting manual procedures for the matching validation process in case of "primary work" matches.

If the book record submitted has no "exact" matches but "primary work" matches the system asks the user for a manual validation. This feature will be available for customisation through the possibility of setting thresholds to support automatic validation (for example if at least one "primary work" match has matching probability percentage higher than 80%, the one with highest matching probability percentage is automatically selected).

Manual validation of proposed matches labelled as "Primary Work match" is still requested.

3.1.5 Automatic matching validation process for "partial" matches

ARROW Workflow has been extended in order to manage automatic procedures for the matching validation process in case of "partial" matches.

If the book record submitted has neither "exact" matches nor "primary work" matches but "partial" matches they are all automatically rejected and the request is treated as a "No match" case.

When you query the ARROW system you do not have to perform the matching validation step and the request automatically goes through the workflow.

3.1.6 Management of books with all proposed matches rejected

ARROW Workflow has been extended in order to manage diligent search requests even for books that does not find a satisfactory match in the TEL catalogue, with the consequence that the user rejects all proposed matches in the manual validation step.

If all the proposed matches are rejected, the request is treated as "single manifestation-cluster", generated by TEL on the basis of the submitted record, and goes straight to the BiP processing phase. When you query the ARROW system the status "Submission rejected" does not exist anymore and all requests automatically go through the workflow.

4. ENHANCEMENTS OF THE WEB PORTAL

The transition of ARROW to ARROW Plus is reflected also on the ARROW website, being one of the main tools for dissemination and official source of information about the system and latest developments.

The requirements for the changes needed in the ARROW website have been outlined in Deliverable 2.1 produced by IFFRO.

The following paragraphs present the work done on the website.

4.1 ARROW PLUS INTRANET

The ARROW Plus Website has a new “ARROW Plus Intranet” area, on the left of the page, with the same structure and permissions of the previous “Menu Intranet” with the ARROW Plus Work Package names.

The ARROW Plus Website has a new role (profile) for ARROW Plus partners. ARROW Plus Partners are able to:

- Access the ARROW Plus Intranet
- Access ARROW FrontEnd
- Access ROW Claiming Service
- Create ARROW Plus General Documentation, ARROW Plus Knowledge Repository, ARROW Plus Work Packages, public private events.
- Edit and Delete just own content belonging to the above mentioned content types.
- Access and Post Comments without approval.
- Maintain and Manage own subscriptions
- Subscribe to post by author, content, content type.

In the “Partners contact List” area of the “ARROW Plus Intranet”:

- it is provided the list of all the users having the role “ARROWPlus_Partner”
- it was removed from display field “ProjectRole”
- it has been added field “ContactSurname” (previously only the name was displayed).

The following new Content Types have been added in the ARROW Plus Website:

- “ARROW Plus General Documentation”
- “ARROW Plus Knowledge Repository”
- “ARROW Plus Work Packages”

For each content type the appropriate Access Control (CRUD) setting is applied to the following user profiles AIE Editor, ArrowPlus_Partner, IFFRO Editor.

The following roles have no access rights on the above content types: Partner, Supporter, National Group Members, ECReviewers.

4.2 ARROW INTRANET ARCHIVE

The ARROW Plus Website has a new “Archived Content” area, on the left of the page, corresponding to the previous “Menu Intranet”. Such area (and all its content) can be edited only by Administrators.

ARROW Partner role has the following restrictions:

- Access (just view) the “Archived Content” area
- Create just public and private events. They can no longer create Content Types such as “General Documentation”, “Knowledge Repository” and “Work Packages”
- Edit and Delete just own public or private events
- Cannot post comments any longer but can view posts and files uploaded to comments
- Cannot maintain and manage own subscriptions (Cannot manage this from “My account area”)
- The “Notification tab” shall be no longer available for Arrow Partners.

4.3 USER ACCOUNT MANAGEMENT

Accounts for the ARROW Plus partners and supporters was created in order to allow access to the website as well as to the provided services (FrontEnd and ROW services)

Country code list was updated with the new countries.

Company name and Company ABR name was updated with new ARROW Plus partners.

Users have been notified of account creation.

4.4 WEBSITE PUBLIC SECTION

The deliverable D2.1 set out the changes to be made to the ARROW public website. The main menu on the top of the page has been redefined as a drop down menu in order to fit the following new elements:

- Presentation of the ARROW system including a demo and/or a guided tour
- ARROW Business model
- Sharing of experiences from using ARROW including the results of the piloting of ARROW in France, Germany, Spain and the UK
- Support: a page with information on support and where to get it for users of ARROW
- Resource page with names of resource persons to contact for users of ARROW requesting help; list of experts to contact; training opportunities; references from those who have used the system with their recommendations; users to contact for further references and information
- Register interest: possible for a non-user of ARROW to register interests in order for ARROW to make contact with potential users.

The new drop down menu has the following structure:

ARROW (heading without the full name, just the word "ARROW")			
ARROW ARROW system ARROW Plus Resources Press FAQ Contact			
ARROW: Sub-menus to be included: <ul style="list-style-type: none"> • Business Model • Organisation • Legal Entity 	ARROWSYSTEM Sub-menus to be included <ul style="list-style-type: none"> • Presenting ARROW system/DEMO • Experience from using ARROW • Support page 	ARROWPlus Sub-menus to be included: <ul style="list-style-type: none"> • What is ARROW Plus • Working Packages • List of partners • Information material 	Resources: Sub-menus to be included: <ul style="list-style-type: none"> • ARROW Plus Deliverables • ARROW Plus Information Material • Arrow Deliverables • Arrow Information Material • Links

The following tooltips was added to the first three sections:

ARROW/Tooltip: Who we are

ARROW system/Tooltip: What is the ARROW system

ARROW Plus/Tooltip: What is the ARROW Plus project

4.5 REGISTER INTEREST FORM

This form enables both ARROW Plus users (authenticated ones) and non-users to submit interests from this area. The submitted content:

- is sent both via e-mail to the relevant address:
- is saved in a site area (Administer -> Content Management -> Webforms) where it can be accessed by site administrators;
- users having AIE Editor Role can manage (view/edit/delete) web form submissions;
- users having IFRRO Editor role can just view the submissions.

The public registration form has a “CAPTCHA” to block eventual spamming.

There is no limit on the number of the requests sent by a user in an established time period.

The registration form has an “Insert text area” and contact information details. The contact information detail contains an email address.

4.6 HOMEPAGE

Home page was updated in order to show the new content of ARROW Plus to Arrow Plus Partners, more specifically:

- ARROWPlus Knowledge Repository
- ARROWPlus Work Packages
- ARROWPlus General Documentation.

Regarding the analogue content of ARROW, it is no longer available to ARROW Partners.

5. ENHANCEMENTS OF THE EUROPEAN LIBRARY SERVICE

This section presents the description of the enhancements implemented by The European Library (TEL) in the first 13 months of the ARROW Plus project.

In this section, we describe the three main lines of work carried out. First we describe the work on the evolutionary maintenance of the TEL system. Then we follow with a description of the work that aimed to achieve a new version of the TEL system for ARROW, which is fully integrated with the TEL systems and regular operations. We finalize with a description of the on-going work in enhancements that were identified during the evaluation of the ARROW system during the earlier project ARROW.

5.1 EVOLUTIVE MAINTENANCE AND NATIONAL USE CASES

The system has been enhanced in order to respond to the evolution of the ARROW infrastructure and to support the requirements emerging from the national use cases of ARROW.

EDItEUR has published the ARROW message schemas as an EDItEUR official standard under the name of ONIX for Rights Information Services (ONIX-RS). The system at The European Library has been updated to support the new specifications of ONIX-RS, together with other ARROW participants, as described in section “6.2 UPDATE OF ARROW SCHEMAS TO ONIX-RS 1.0” .

The national use cases have presented new requirements that had an impact in several parts of the ARROW infrastructure. The following requirements affected the system at TEL and have been implemented:

- Management of books that does not find a match in the TEL catalogue (described in Section 3.1.1)
- More granular classification of the match types (described in Section 3.1.2)
- Management of books with all proposed matches rejected (described in Section 3.1.6)

Some requirements of the national use cases affected only the behaviour of the matching and clustering operations of system at TEL, and did not affect any other ARROW systems. These requirements were related with the particular characteristics of the data in the national bibliographies of France and the United Kingdom. The system was changed to allow the configuration, at a country level, of matching and clustering behaviour as follows:

- Matching criterion for “language of text”: the system can be configured to use only the first “language of text” in the record, or require matching of all languages.
- Handling of missing “country of publication”: when the bibliographic record is missing the data field containing the country of publication, the system may be configured to assume the country to be the one of the national library.

5.2 INTEGRATION INTO TEL’S PRODUCTION SYSTEMS AND INTERNAL WORKFLOWS

During the course of the ARROW project, the requirements for TEL service were uncertain until late in the project. This was due to the existing data quality problems of library catalogues, which could impact the automatic machine processing of the information required for ARROW. At the time, the impact of these data quality problems on the matching and clustering services was uncertain. It was not known which quality could be achieved given the data quality problems, nor the computational power necessary to overcome them.

In this scenario it was necessary to adopt a prototyping software development methodology for the TEL service. The required agility for a prototyping approach could not be achievable with the constraints imposed by the regular operations of the TEL production systems. Therefore the TEL system for ARROW was developed independently of the main production systems of TEL, and it was maintained and operated solely by the software development team of TEL.

ARROW also required the usage of bibliographic data in the highest quality possible. Therefore, TEL had to include in its regular operations, the support for the richer data formats used in libraries (the MARC family of bibliographic formats). The support for these formats had an impact in the internal business processes of TEL involving the data ingestion, and in the systems that support them.

Figure 3 presents the general architecture of the TEL system for ARROW before the start of project ARROW Plus, and Figure 4 presents the new general architecture. The remainder of this section will describe the main actions taken in order to make the TEL service for ARROW, a part of the production systems of TEL, and the organizational changes to accommodate the operation of the TEL service within regular TEL operations.

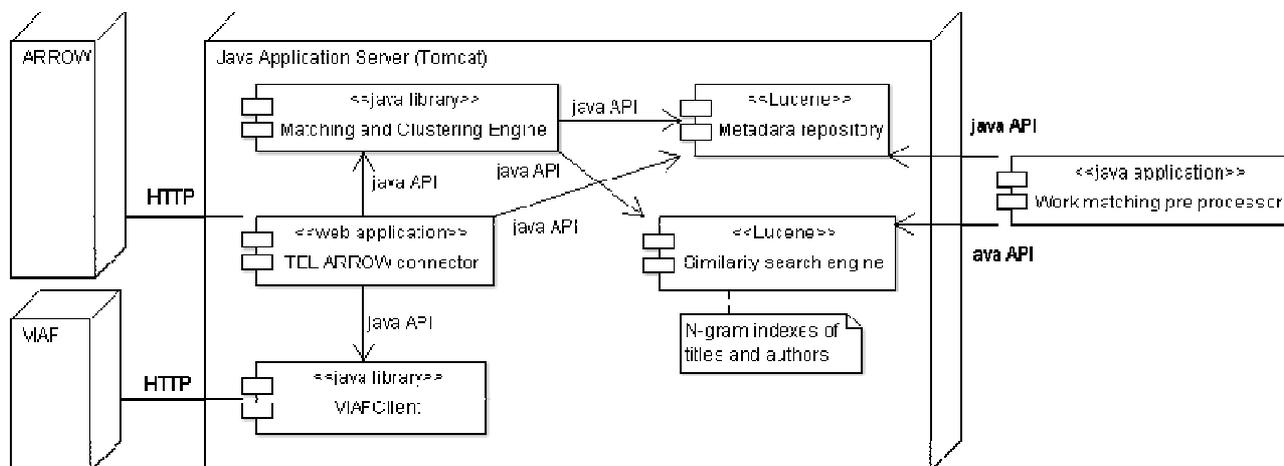


Figure 3 The TEL system for ARROW before the start of project ARROW Plus

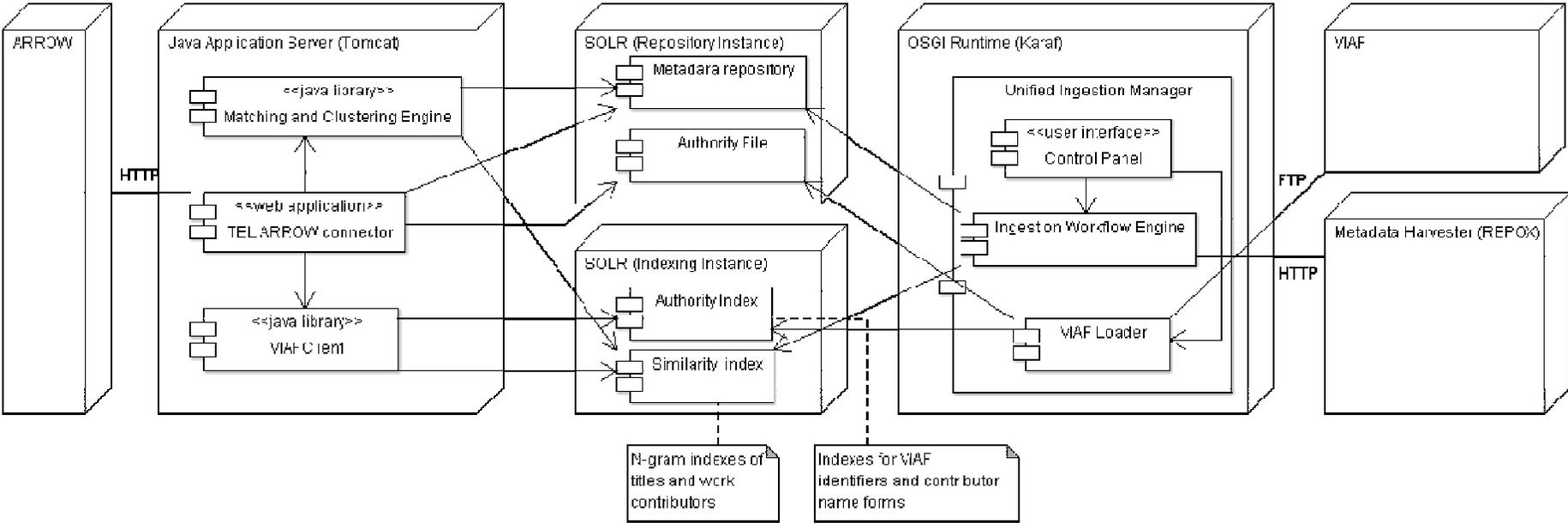


Figure 4 The current TEL system for ARROW

5.2.1 Source code restructuring and documentation

The first issue addressed in order to integrate the ARROW system in TEL's production systems, aimed to restructure the source code to follow the practices in use for production systems.

This task involved the following actions:

- Restructure of the Maven project according to the standard Maven hierarchy in use at TEL.
- Replacement of external libraries with alternative versions supported by TEL's software development team.
- Comprehensive Javadoc documentation.
- Creation of unit tests achieving at least 50% of line coverage.
- Integration with the continuous build tools in use at TEL.

5.2.2 Integration with the TEL central repository

The new system no longer maintains its own storage of bibliographic data and work metadata. The system was changed to use the general repository of bibliographic records used by the production systems of TEL, which is based in SOLR.

This change included the adaptation of the Matching and Clustering Engine and the TEL-ARROW Connector components in order to use the TEL central repository. It also included the replacement of the Work Matching Pre-processor with the Unified Ingestion Manager. This component is used in TEL to manage all the data ingestion operations, such as TEL's main portal, and the ARROW system. The Unified Ingestion Manager was extended with data ingestion workflows particularly designed for the ARROW system.

5.2.3 Integration with the TEL authority file

VIAF was the external data source used to gather more information about work contributors and make it available to the ARROW workflow. In the current version of the TEL system, VIAF maintains its role in the system, but it is used in a different way, in order to enable its use in other operations of the TEL system for ARROW, such as the enhancement described in Section 5.3.2.

In the previous system, VIAF was queried remotely. In the new version the complete data VIAF data set is transferred and ingested into the TEL authority file. The TEL authority file is the system that TEL uses to centrally manage data about entities and concepts, often used in bibliographic metadata.

Specific search indexes for the ARROW system are created from the VIAF data in the Authority Index. These indexes allow particular searches to be performed in VIAF data which could not be performed through the remote searching interface of VIAF.

5.2.4 Reimplementation of the matching and clustering indexes based on SOLR

In the previous version of the TEL system, the Similarity Search Engine was implemented in Lucene. With the intensive use of ARROW, Lucene became a limitation in terms of scalability, since it needs to run in the same machine as the TEL-ARROW Connector.

The new system implements all the similarity searching that supports the matching and clustering requests of ARROW based in SOLR. The indexes are still based in n-gram techniques, but SOLR can be hosted in a dedicated machine, therefore allowed several scalability strategies to be investigated, such as hardware and software configurations, and storage technologies based on solid state drives.

The previous system required, on average, two seconds to perform a manifestation or clustering operation on a single record. The current system is able to process three hundred records in one second. This level of performance allows the clustering of complete national bibliographies to be performed within one or two days, instead of weeks or months as in the previous version.

5.2.5 Implementation of Work Metadata based on the internal TEL data model

The European Library system for ARROW uses, internally, a specific data model for Work Metadata. This data model aims to make the system independent of the data formats used at the national libraries, while fulfilling the requirements of the rights clearance process of ARROW.

In the previous system, the bibliographic data was encoded according to this model directly from the source library data. However, the TEL production systems use an internal data model instead of the original data formats from the libraries. For this reason, the TEL internal data model was adapted to fulfil the data requirements of ARROW, and the internal Work Metadata of the ARROW system is now derived from the internal data model of TEL.

5.2.6 Catalogue ingestion internal workflow

ARROW is the first scenario addressed by The European Library to require the usage of well-structured bibliographic data. All the other services of TEL are target at search and retrieval use cases, which have been based in simpler data models based in the general Dublin Core elements.

The richer data structure of MARC formats is mandatory to fulfil the requirements of ARROW, and these formats were supported by the previous system. With the integration of the ARROW system in the TEL production systems, all data ingestion processes are handled by the Collections Team of TEL, for this reason the tools that support the work of the Collections Team had to be adapted to include the functions of the Work Matching Pre-processor for ingesting MARC formats, and to fulfil the information requirements for the management of the operation regarding the transfer and maintenance of the national library catalogues in TEL.

5.3 IMPROVEMENTS IDENTIFIED IN THE VALIDATION REPORT OF PROJECT ARROW

In the deliverable “D7.2 Validation Report” of the earlier project ARROW, several recommendations were made for improvement of the results of ARROW, and some of these recommendations directly concern the system of The European Library. In this section we present two particular recommendations that are being addressed at the time of writing of this deliverable, and which are planned to be ready by June 2012.

5.3.1 Extraction of work contributors the natural language text in statements of responsibility

The first author is almost always present in the structured data fields of bibliographic records, but in some cases contributors are only present in the statement of responsibility. The statement of responsibility is a data field that contains natural language text, which cannot be easily processed by computers. The consequence of this problem is that primary work and secondary work clusters may be imprecise, since the algorithms are not able to use the complete list of work contributors. The following are some examples, in several languages, of statements of responsibility found in bibliographic records:

- “translated by Alexander Pope, with notes and introduction by the Rev. Theodore Alois Buckley ... and Flaxman's designs”
- “Érnst Teodor Amadej Gofman. Chudožnik Boris Tržemeckij. [Per. I. Tatarinovoj]”
- “Ernst Theodor Amadeus Hoffmann. Mit Holzstichen von Andreas Brylka”
- “Vicente Aleixandre ; estudio previo, selección y notas de Leopoldo de Luis”

In order to make the complete list of contributors available for ARROW, an information extraction process must be executed on the statement of responsibility, which locates the names of all contributors.

At the time of writing of this document, The European Library is analysing the performance of information extraction techniques for named entity recognition.

5.3.2 Matching contributors of works with VIAF records of persons

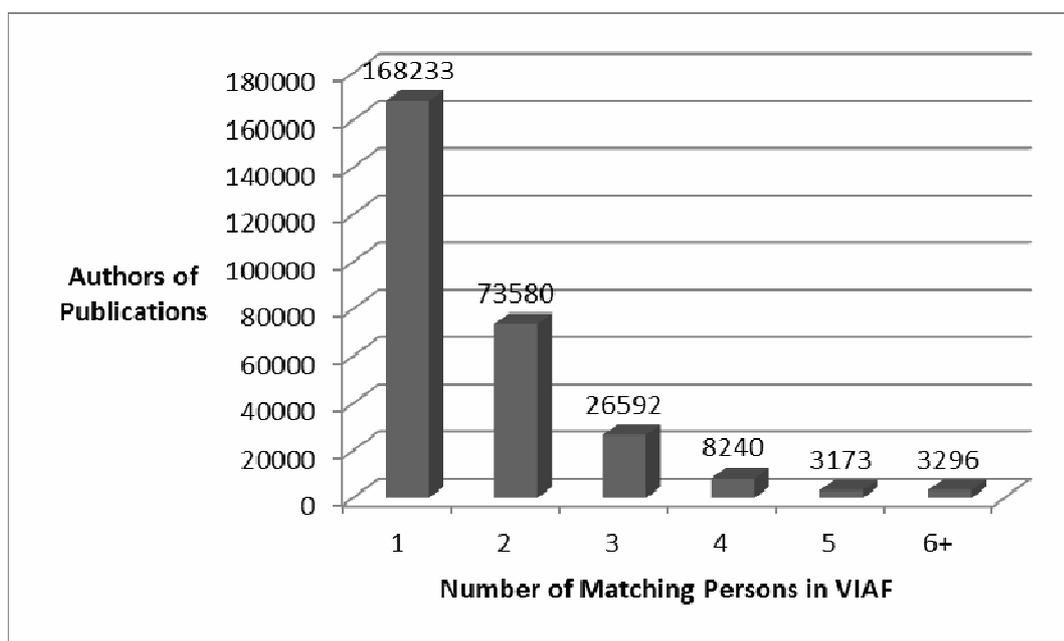
Matching of work contributors in VIAF is already available for some countries (France, Germany and Spain). However, matching the work contributors is currently not supported in countries whose national library is not a participant of VIAF, or the national library cannot send the identifiers of the contributors to The European Library².

In order to support this functionality in every country, the TEL system must support searching, in the VIAF data set, for the contributors’ names, and check for evidence that confirms a match, or

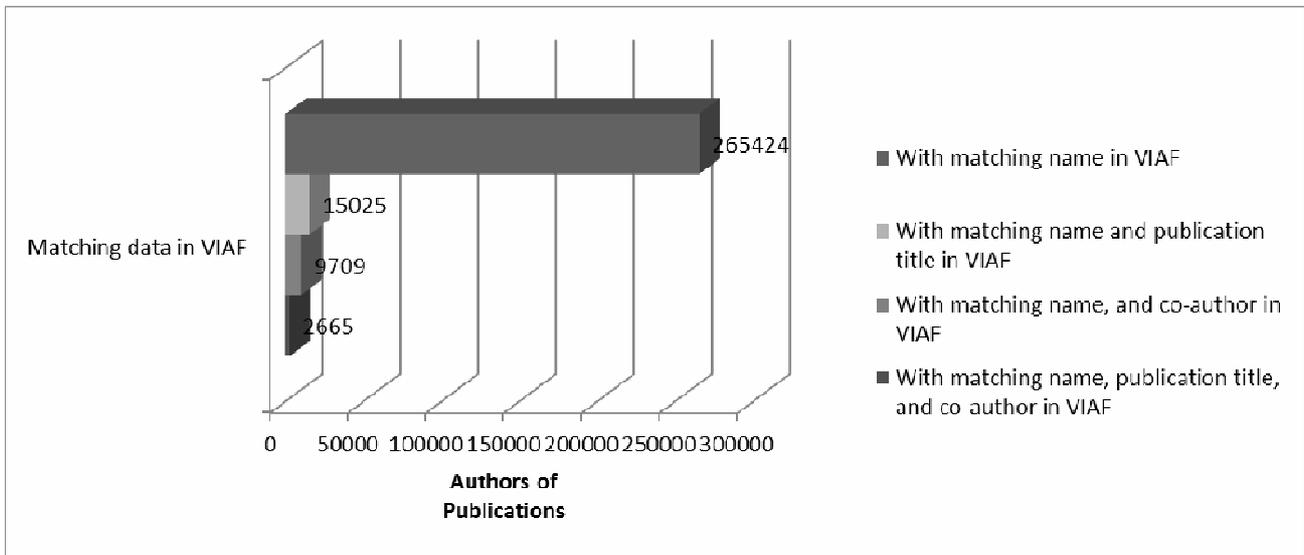
² Due to limitations of the library management systems in use at some libraries, the identifiers of work contributors in the library’s authority file cannot be exported in the bibliographic records, and shared with The European Library.

allows disambiguating between different persons with the same name. Such evidence is taken from information available in the bibliographic records and in VIAF records such as the birth and death dates, known titles by the author, and know co-authors.

Work on this enhancement has started with a study addressing the ambiguity of the contributor's names. The following figure presents how many names of authors in bibliographic records (from a sample of 283,114 contributors) matched names of persons in VIAF, and how ambiguous the name matching can be, by showing the number of distinct VIAF matching records. In total, 59% where not ambiguous, 26% matched two records, 10% matched three records, 3% matched four records, 1% matched five records, and the remaining 1% matched 6 or more records.



We also measured the number of contributors that have a match in VIAF, with just the name in common, with the name and the title of the publication in common, with the name and a co-author in common, or with the name, the title of the publication and a co-author in common. The following figure presents the measured values, and we can observe that in some cases reliable information is available in the data to allow a correct disambiguation of the contributors. However, for the majority of cases, only a matching name was found.



The outcomes of this study indicate that ambiguity in the names of the contributors is high, and in many cases no enough information is available to disambiguate. Since for the rights clearance process of ARROW, the matching of the contributors with VIAF must be very precise, the TEL system will only provide the matching results to ARROW when enough disambiguation information exists.

6. ENHANCEMENTS OF THE RII

6.1 WORKFLOW EXTENSIONS

6.1.1 Management of “No match” case

ARROW Workflow was extended in order to manage library diligent search requests even for manifestations that do not find a match in the TEL catalogue.

No Matches in TEL go on in the BiP processing phase as single manifestation-cluster, generated by TEL on the basis of the target resource record. This enables ARROW to gather relevant information in the BiP domain and finally send them to the relevant RRO, instead of aborting the workflow.

6.1.2 Management of automatic validation

In order to minimise the human intervention in the ARROW workflow to make it fit to mass digitisation projects, the workflow was configured to allow for each submission to choose whether the matching validation step is done by a human being (manual validation) or by ARROW system (auto validation) according to defined parameters (MatchType and matchProbabilityPercentage).

The automatic procedures for the matching validation process in ARROW is:

- if the target resource has one or more exact match, the one with highest matching probability percentage is automatically selected;
- if the target resource has no exact matches but “primary work” matches and at least one of them has matching probability percentage higher than 80%, the one with highest matching probability percentage is automatically selected. Otherwise the system asks the user for a manual validation. If all the proposed matches are rejected, the system treats the manifestations as a “No match” case.
- If the target resource has neither exact matches nor “primary work” matches but “Partial Matches” they are automatically rejected and the manifestation shall be treated as a “No match” case.

6.1.3 Business process model of the enhanced workflow

As a better description of the workflow enhancement, it is useful to consider the activity diagram as it was at the end of the ARROW project and as it is now after the enhancements.

Figure 5 depicts the business process flow performed by the DataCentre component at the end of ARROW project. It is interesting noting that the flow could stop in three cases: “no match” found in the TEL catalogue; all proposed matches rejected by the library; unmanaged national flow for the selected manifestation (i.e. unmanaged country). Furthermore the “primary work” match type was not covered.

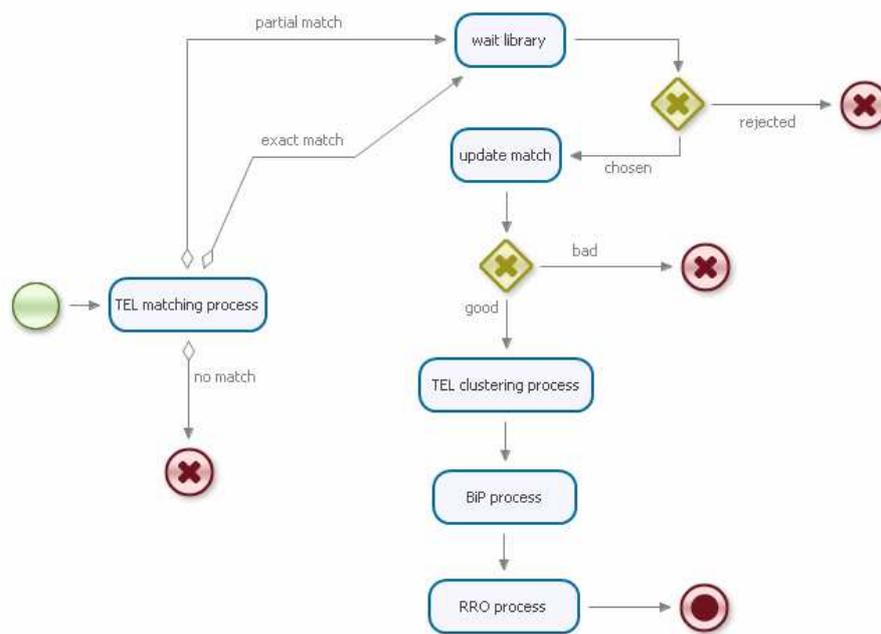


Figure 5 Workflow diagram in ARROW before the enhancements

Figure 6 shows the enhanced workflow: now only the "unmanaged country" case stops the flow. As already explained in the previous sections, the "no match" and "rejected" cases are managed the same way as the "primary work" match type.

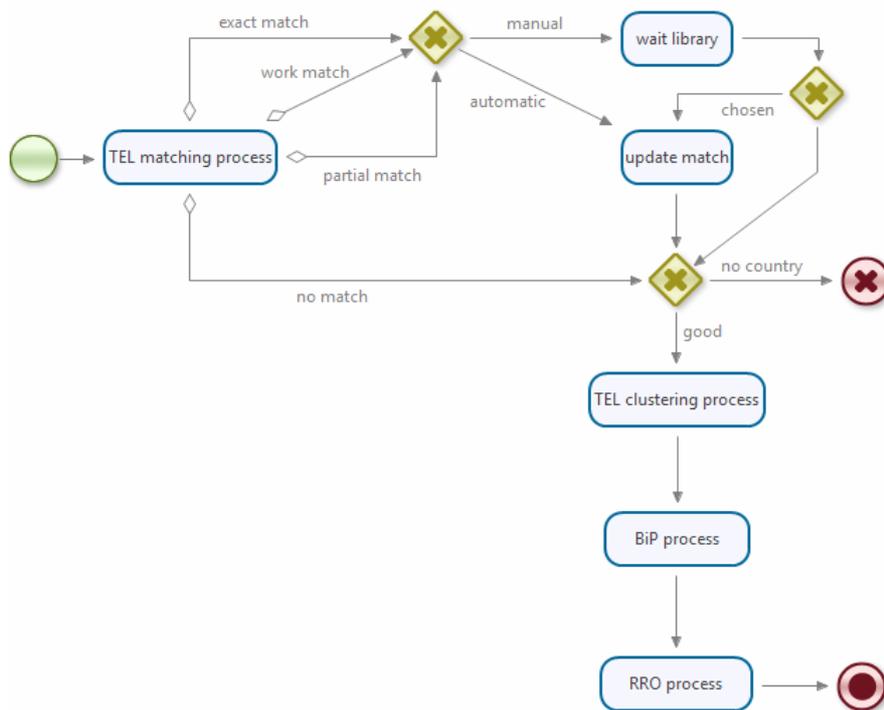


Figure 6 Workflow diagram in ARROW plus after enhancement

6.2 UPDATE OF ARROW SCHEMAS TO ONIX-RS 1.0

EDItEUR has published the ARROW message schemas as an EDItEUR official standard under the name of ONIX for Rights Information Services (ONIX-RS).

The needed activities to upgrade the ARROW system to the new schema have been:

- Notify in advance data providers of update and schedule of work
- Provide data providers with new schema and documentation, including list of changes
- Update of ARROW system (by CINECA)
- Update of TEL
- Update of France: Electre M6 and CFC M7
- Update of Spain: CEDRO M6 and M7
- Update of Germany: VLB M6 and VG WORT M7
- Update of UK: UK RRO M6 and M7
- Review transformation from M4R to M6Q
- Review transformation from M6R to M7Q
- Review of all html view
- Review of Publishing status algorithm and Copyright status algorithm
- Test phase both in staging and the live environment.

6.2.1 CINECA's activities

In detail the activities in charge to CINECA for the upgrade of the ARROW system have been:

- Upgrade data model according to new schema version
- Upgrade the decision handlers of the workflow
- FrontEnd upgrade in order to handle M1Q configurations
- FrontEnd and DataCentre upgrade in order to manage the new schema changes at presentation layer.

6.3 FRONTEND WEB INTERFACE

6.3.1 Diligent Search Results for different User Profiles

ARROW FrontEnd provides diligent search views for two kinds of user profiles:

- Basic FrontEnd User
- Advanced FrontEnd User

Basic FrontEnd User is able to view the diligent search results only upon diligent search completion.

The final Result page is conformant to the following graphic sketch:

Response to the Permission requested

Your Request

Verteidigung der Wolfe, Hans-Magnus Enzensberger (1929-), Suhrkamp Verlag, Frankfurt a.M, 1957
 Permission requested: **Digitize For Preservation**

*Commento: questa parte di info è la terget manifestation quindi è disponibile subito.
 Mentre le successive parti di info si aggiorneranno man mano*

According to the ARROW processing, the Expression embodied in the manifestation you requested is: in Copyright, Commercially Available, and the Ophan Status is Unspecified

*Commento: CopyrightStatus va messo per primo perchè appare con l'M4R,
 PublishingStatus va messo per secondo perchè appare con l'M6R
 OrphanStatus va messo per terzo perchè appare con l'M7R
 insieme alle informazioni sottostanti*

Response to your request

From	To
VG-WORT Germany Contact: Dietmar Gräbner d.graebner@berlinger.cc	CINECA via magnanelli Italy Contact: Giuseppe Trotta, 12345 g.trotta@medra.org

VG-WORT Answer

The permission requested is **Granted** for the following reason: **Rights Holder Mandate Held**

Suggested Action

Contact RRO Specified

Figure 7 FrontEnd Results Page

The system displays in “Response to the Permission Requested” area:

- 1) “Your Request” section with a subset of the Target Resource metadata containing:
 - Title (MARC field 245\$a)
 - List of Contributors (MARC fields 100\$a; 110\$a; 700\$a; 710\$a) and their date of birth and death if present in the MARC file (MARC field 100\$d; 700\$d)
 - Publisher (MARC field 260\$b)
 - Country of Publication (MARC field: check first 044\$c then 044\$a then 008 position 15-17)
 - Year of Publication (MARC field: check first 008 position 7-10 then 008 position 11-14, then 260\$c)
- 2) Work Rights Status information
- 3) “Response to your request” section

The system displays in “Supporting Resources” area:

- Target Expression information
- Selected Manifestation information
- Related Manifestations information
- Related Expression information

Advanced FrontEnd User is able to view partial results containing XML response messages exchanged during ARROW Workflow (TEL, BiP, RRO).

6.3.2 Display of ARROW Availability/ARROW Publishing status

For each manifestation only ARROW Availability status the interface displays:

- ARROW Availability if present,
- ARROW Publishing status if ARROW Availability is not present

6.3.3 List of permission sets refined

The list of permission sets has been refined with the addition of two distinct permission pairs “Digitise for print document supply | Digitise for digital document supply” and “Use the digitised resource for print document supply | Use the digitised resource for digital document supply”.

The permission set chosen is paired with the book record submitted and transmitted to the RRO/CMO in the final step of the workflow.

When you query the ARROW system you can choose a more granular permission, distinguishing on the basis the type of service offered: whether you deliver the document on paper or electronically.

Therefore, in the Query form the Permission Sets shall be:

- Digitise for preservation
- Digitise and make available for use in library
- Digitise and make available on the Internet
- Digitise and make available for academic use
- Digitise for print document supply
- Digitise for digital document supply
- Use the digitised resource to make available for use in library
- Use the digitised resource to make available on the Internet
- Use the digitised resource to make available for academic use
- Use the digitised resource for print document supply
- Use the digitised resource for digital document supply

GLOSSARY

AWR: ARROW Work Registry is a component of ARROW system that stores and maintains all these pieces of information for every request processed by ARROW.

BIP: the Books in Print are the organisations that maintain the information about books in commerce in a certain country or language area.

CAPTCHA: <http://en.wikipedia.org/wiki/CAPTCHA>

DataCentre: the DataCentre is a component of the ARROW system that constitutes the back end and performs the business logic of the entire system, including both the RII and the Arrow Work Registry flow.

EDItEUR: it is the international group coordinating development of the ONIX suite.

Exact match: The TEL's matching algorithm returns a measure of the matching for each record expressed as a percentage. On this base, a bibliographic record is said an "exact match" of the manifestation submitted by the library, if it has a value of matching greater than a pre-defined and customizable threshold, for example 80%. The TEL's matching algorithm takes into consideration the following bibliographic metadata: title, all contributors (regardless the role), all languages, publisher and date of publication.

Expression (or Work): in the ARROW scope a work/expression is defined (as in the ISTC³ definition) by the following metadata: title, language and all contributors (regardless the role). This is the ARROW internal representation of a distinct intellectual or artistic creation.

FrontEnd: the FrontEnd is a component of the ARROW system that is responsible for collecting the initial request from the user, processing it and sending it to the ARROW DataCentre; in other words it represents an interface between the user and the DataCentre.

MARC: it is a family of bibliographic data formats used in library domain to encode and share information about books and other material.

Manifestation: the physical embodiment of an expression. Manifestation represents all the physical objects that bear the same characteristics, in respect to both intellectual content and physical form.

ONIX: the ONIX standards are designed to support computer-to-computer communication between parties involved in creating, distributing, licensing or otherwise making available intellectual property in published form, whether physical or digital. All are expressed in XML.

No match: on the base of the TEL's matching algorithm, no records in the TEL catalogue match with the manifestation submitted by the library.

Partial match: on the base of the TEL's matching algorithm, a bibliographic record is said an "partial match" of the manifestation submitted by the library, if the following fields do match: title and first author only. The record does not share the same intellectual work with the library manifestation.

³ ISTC: International Standard Text Code. <http://www.istc-international.org>
ISTC in practice: <http://www.bic.org.uk/files/pdfs/100702%20bell.pdf>

Permission sets: they are a list of descriptions of permissions. The ARROW users must choose one of these sets to express the permission they want to require for the book.

Pilot countries: they are the countries that during the ARROW project have been made available to users to search for books rights status, that is the BIPs and the RROs in those countries joined the ARROW system in order to provide information. The countries were: France, Germany, Spain, UK.

Primary cluster: on the base of the TEL's clustering algorithm, the Primary cluster contains all the manifestations in the TEL catalogue that match the following fields with the Target manifestation: title, all contributors (regardless the role) and all languages.

Primary work: on the base of the TEL's matching algorithm, a bibliographic record is said a "primary work" of the manifestation submitted by the library, if the following fields do match: title, all contributors (regardless the role) and all languages. The record does share the same intellectual work with the library manifestation.

Related Expression: a related expression (work) found during the calculation of the Secondary Clusters.

Related Manifestation: a manifestation that belongs to the same Target Expression of the Selected Manifestation.

RII: the Rights Information Infrastructure is the "engine" of the ARROW system. According to a planned workflow, it queries and retrieves information from a multiplicity of data providers, processes the information and takes decisions on the successive elaboration.

ROW: Registry of Orphan Works. It is based on a subset of the AWR. It is publicly available to specific categories of users for specific purposes.

RRO: the Reproduction Rights Organisation are the organisations that maintain and manage the information about right status, in a certain country or language area.

Secondary cluster: on the base of the TEL's clustering algorithm, Secondary cluster does match with the Target manifestation applying less restrictive criteria of similarity respect to the ones of the Primary cluster: only the first author must match, and the title similarity is less restrictive. As the Primary Cluster, the Secondary cluster contains all manifestations matching each other for the fields: title, all contributors and all languages.

Single manifestation-cluster: a cluster formed only by the Target Resource. These clusters are created by TEL to allow the ARROW Workflow to proceed if no matches are found in TEL.

Target Expression: the expression (work) at which the Selected Manifestation belongs, as established by the Primary Cluster.

Target Manifestation or Selected Manifestation: the manifestation that drives the diligent search process. It corresponds to the national library catalogue record if there is a match of the Target Resource in the TEL catalogue and is selected; or to the Target Resource itself if such a match doesn't exist.

Target Resource: the library record submitted by a library to ARROW.

TEL: The European Library

Use case: with use case it is indicated a relevant use of the ARROW service in a real project of digitalization.

VIAF: Virtual International Authority File.

Wellcome Trust: Wellcome Trust is a global charity dedicated to achieving improvements in human and animal health. During 2011 the Wellcome Library (that is part of the Wellcome Trust) began a pilot digitisation project using ARROW service to search for the relevant copyright owners.

Work (or Expression): see Expression entry.

LIST OF REFERENCES

- D7 2_ValidationReport.pdf
- D2.1 Reviewed ARROW Website_final.pdf