

5th Meeting of the Member States' Expert Group on digitisation and digital preservation

Luxembourg, 1 October 2009

ARROW

Accessible Registries of Rights Information
and Orphan Works towards Europeana

www.arrow-net.eu

Co-funded by the
Community
programme
eContentplus



Summary

2

- Scope and purpose of the Arrow project
- The work done up to now
- The Arrow workflow
- “Looking for parents and finding other relatives”
- Some (inevitable) comparisons with the Google Settlement

Origins of the project

3

- Inclusion of copyrighted (text) works in the digitisation programme
 - Copyright Subgroup of the HLEG
 - Need to facilitate right negotiation
 - Challenges in **Orphan works** and **Out of print works**
- Innovative initiatives
 - Libreka in Germany
 - Gallica 2 in France
 - eBog.dk in Denmark
 - New initiatives in Spain and Norway
- All need tools for rights information exchange and interoperability

The Arrow consortium and liaison organisations

4

□ Libraries

- *EDL Foundation*
- National libraries in France, Spain, The Netherland, UK, Norway, Slovenia, *Finland*
- University library of Innsbruck

□ Publishers

- **Federation of European Publishers**
- Publishers associations of Italy, Spain, Germany, *UK*

Authors

- *European Writers Council*
- ALCS in UK

□ Reproduction Rights Organisations

- **IFRRO**
- RROs in *Spain, France, Italy, UK, Norway, Denmark, Finland*

□ Technology Developers

- Cineca (University consortium in Italy)
- Numilog (Private company in France)

□ Standard organisations

- ISBN agencies (Italy, Slovenia, Norway, *Finland, Germany*)
- *European DOI agency (mEDRA)*
- Editeur members
- ISTC consortium and ISNI ISO-WG members

The scope: rights information, not rights management

5

- We will support libraries to have rights information (RI) about the books they plan to digitise and/or use
- We'll receive a metadata record
- And we'll give back structured RI

The purpose

6

- We cannot provide an “official” statement that a work is orphan
- We will provide **“the best information currently available”** from the highest number of resources that is possible
- We will support the creation of registries of orphan works
 - ▣ A question not yet answered: one European Registry or a network of interoperable registries?
- Rightholders will be able to claim their books so to challenge the initial statement

The work done up to now

7

- A set of studies on:
 - ▣ The underlying business models for making protected works accessible in digital form to the public
 - ▣ The legal framework
 - ▣ The existing resources to be queried
 - ▣ The applicable standards
- Definition of the ARROW work flow and system architecture
 - ▣ The ARROW work flow is the formalization of what we learned in the studies

The library question

9

- The library provides one bibliographic record of one book
- **+ specification of requested use(s)**
- We are now defining
 - ▣ Metadata schemas accepted (basically Marc-21)
 - ▣ Minimum metadata elements to be provided (i.e.: without them we'll not be able to answer)
 - E.g.: Author, Title, Publisher, Date of publication (?)
 - ▣ Optional metadata elements recommended
 - E.g.: Country of publication, place of publication, etc.
 - ▣ Specification to express use, and relative rights

First phase: querying the libraries

10

- After some manipulations, Arrow forwards the record to TEL
- What we expect in return:
 - *Identification of the manifestation*, i.e.: matching between the input metadata record with what is available in libraries
 - May be complex for pre-'70s books, without ISBNs
 - *Richer metadata record for that manifestation*
 - *First clustering to find other manifestations of the same work, and related works*
 - Collaboration with the FRBR initiative
 - *Rich metadata records for all books grouped*
 - *Information about contributors*
 - Collaboration with VIAF

Second phase: querying the books in print

11

- After some manipulations (e.g. transformation in ONIX) Arrow forwards the record(s) to books in print
- What we expect in return:
 - ▣ *Possibly second clustering to find other manifestations of the same work, and related works*
 - ▣ *In print / out of print status for every record in the cluster*
 - ▣ *Information about publishers*
 - Possible collaboration with PIID

Third phase: querying the RROs repertoires

12

- For out of print works (not books)
- After some manipulations Arrow forwards the record(s) to RROs
- What we expect in return:
 - *Possibly third clustering to find other manifestations of the same work, and related works*
 - *Some specific rights information*
 - E.g.: “we have the mandate to license that book for certain use”; contact details of the author; “we know the author and we informed him/her”; etc.
 - *Possibly terms and conditions of licenses available*

Publication of “apparent” orphan works metadata

13

- Works for which we will not find rightholders information will be published in the registry(ies) of orphan works
- Rightholders are entitled to claim their books in any moment

Finally!: answering the library

14

- All the relevant information collected is passed to the library
- Alternatively:
 - Public domain status
 - With or without rights belonging to publishers on the specific manifestation
 - In print work + details of publisher
 - Out of print work, with known rightholder
 - Relevant clearing center
 - + info on terms of the license available **or** deny of authorisation **or** rightholder contact detail **or** “the request has been passed to the rightholder”
 - Out of print work, with unknown rightholder
 - Last publisher(s) of the work + gathered information about contributors

Implications of the process (the “other relatives”)

15

- In difference phases we found lack of standard instruments to optimise the process
 - In the book identification: lack of ISBNs for old books
 - A working group has been established within the ISBN International to study the problem of retrospective assignment of ISBNs
 - In the work identification: start up of the use of ISTC
 - The project aims at triggering ISTC registrations
 - Contacts established with relevant ISTC Authority and forthcoming ISTC agencies in several European countries
 - In the RH names identification: waiting for the ISNI
 - Contacts established with the ISO Working Group
 - The objective is to piloting the ISNI within the Arrow system
- Strong awareness of relevance of standards to make the system open and effective

A changing environment

16

- The emergence of the Google Settlement influenced our work
 - We need to consider the new arena
 - The BRR broadly overlap our scope (though with a very different approach)
- The analysis of the Settlement database pointed out criticalities
 - ***Matching and clustering*** records is the most difficult part of the job
 - The study on noted Italian authors within the Google database shows error rate of 81% on this respect, which is very worrying
 - The plans about standard identification systems for *matching and clustering* (ISBN, ISTC, ISNI...) is crucial

Comparison with the Google approach and the Book Right Registry

17

- The “Google case” was very useful for Arrow
- In the management of the Settlement database they did all the errors that is possible to do
 - So, we have a comprehensive picture of what to avoid!
 - “*GBS metadata are awful*”,
 - Geoff Nunberg, Google Books: the Metadata Mess, Google Book Settlement Conference, UC Berkeley, August 28, 2009, slides available at <http://people.ischool.berkeley.edu/~nunberg/GBook/GoogBookMetadataSh.pdf>
- They have strong incentive in the Settlement to “kill the parents”
 - The more the “orphans” are, the higher their revenues

Characteristics of the Settlement database

18

- The centralisation of the process results in lack of capacity to exploit existing competences
 - ▣ Bibliographic data, alone, are not sufficient
 - ▣ Stakeholders involvement and transparency of the processes are even more important
- Key problems:
 - ▣ Matching between original metadata records and that existing in other database not afforded properly
 - ▣ Clustering manifestations not afforded properly
 - ▣ Misuse of standards: ONIX, ISBN, ...
 - ▣ Big (technical) confusion about commercial availability
 - Which ONIX field is to be considered?

Relations between Arrow and the BRR

19

- Contacts have been established (already four meetings done, next scheduled during next Frankfurt Bookfair)
- ARROW could serve as an example to BRR (which is very far to be set up)
- The BRR can be seen as a node within the Arrow network
- The current expectations around the Google Settlement (see the DoJ submission to the Court) make the value of collaboration even more evident
 - ▣ It is probable that the Settlement should be discussed to take into consideration also non-US stakeholders interest

Arrow and the BRR

(please note: not Arrow and Google)

20

ARROW	BRR
<ul style="list-style-type: none"> □ Stems from stakeholders agreement 	<ul style="list-style-type: none"> □ Concludes a law suit
<ul style="list-style-type: none"> □ Distributed approach 	<ul style="list-style-type: none"> □ Centralised approach
<ul style="list-style-type: none"> □ Deals with rights information and redirection 	<ul style="list-style-type: none"> □ Deals with rights information and rights management
<ul style="list-style-type: none"> □ Serves a fragmented market and 30 legal frameworks 	<ul style="list-style-type: none"> □ Serves one single market and one legal framework

Arrow and the BRR / 2

21

ARROW	BRR
<ul style="list-style-type: none"> □ Business neutral 	<ul style="list-style-type: none"> □ Serves one business model (though announced as scalable for others)
<ul style="list-style-type: none"> □ Resources: 5 million euro (2,5 mil EU contribution) 	<ul style="list-style-type: none"> □ Resources: 35 millions dollars

- In synthesis: we have to do much more difficult things with much less money
- Our strengths are in the capacity of collaboration (also with them, possibly):
 - The “network” is our only competitive advantage in such a competition

Thank you for your attention

FURTHER INFORMATION

Piero Attanasio

piero.attanasio@aie.it

ARROW

Accessible Registries of Rights Information
and Orphan Works towards Europeana

<http://www.arrow-net.eu>

Co-funded by the
Community
programme
eContentplus

