

**Assises Professionnelles du Livre**

**A l'heure du numérique**

**Paris, 25 November 2009**

**ARROW**

Accessible Registries of Rights Information  
and Orphan Works towards Europeana

**The ARROW Project:  
managing digitisation rights for Europe**

[www.arrow-net.eu](http://www.arrow-net.eu)

Co-funded by the  
Community  
programme  
eContentplus



# Summary

2

- Origin, 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
- Long term vision

# The very first idea (June 2006): The *Rights Information Infrastructure*

3

- Moving the book trade from print to digital does not mean moving from books to files
- In the econtent market consumers do not buy econtent, they buy rights for digital use of content (i.e. works)
  - Rights to access, download, duplicate, distribute, print, integrate in other content, transform, etc.
- The “tangible book” trade was based on very good information infrastructure (ISBN, books in print, national bibliographies, EDI formats, recently ONIX, etc.)
- The long term challenge is to have a similar “**Rights information infrastructure**” (RII) serving the “intangible book” trade

# What we need for the RII

4

	Books	eContent
<b>Identification</b>	Book Identifier = <b>ISBN</b>	Work identifier = <b>ISTC</b> Web actionability = <b>DOI</b>
<b>Trade messages</b>	Communicating data about book transactions = <b>EDI, ONIX</b>	Communicating data about rights transactions = <b>ONIX for licensing terms</b> et al
<b>Metadata syndication</b>	Limited to certain application = <b>MARC, DC, (ONIX)</b>	Crucial for the whole process, requires comprehensive models (e.g. <b>&lt;indec&gt;</b> ) and protocols (e.g. <b>OAI-PMH, ACAP</b> )
<b>Management of secondary (?) rights</b>	Since secondary, delegated to third parties	Uncertain boundaries between secondary and primary rights, require <b>analytic management</b> , and possibly changes in the role of CMOs

# Origins of the project

5

- Inclusion of copyrighted (text) works in the digitisation programme
  - ▣ Copyright Subgroup of the High Level Expert Group formed by the European Commission
  - ▣ Need to facilitate rights negotiation
  - ▣ Challenges in **Orphan works** and **Out of print works**
- Innovative initiatives
  - ▣ **Gallica 2** in France
  - ▣ Libreka in Germany
  - ▣ eBog.dk in Denmark
  - ▣ New initiatives in Spain and Norway
- All need tools for rights information exchange
- Perfect use case for the concept of the RII

# Orphan works: a sad metaphor

6

- A protected work for which it is impossible to find rightholders is defined as ***orphan***
- N.B.: “parents” of these orphans are missing and not dead!
- Therefore, we must have two objectives:
  - ▣ To reduce orphans, discovering the parents (i.e. rightholders)
  - ▣ To take care of orphans, when it is really impossible to find the parents
- Number of orphans grows during wars
  - ▣ A co-operative approach between libraries and rightholders is crucial to reduce orphans
- Orphan works result from unsuccessful ***diligent search*** of rightholders
  - ▣ Diligence is a key factor, and fair-play between actors in the value chain is important
  - ▣ **Tools for facilitating search** are even more important

## The purpose: “to save the parents”

7

- Looking for **rights information** for every book to facilitate diligent search
  - ▣ In the long term: RI may be useful for other purposes
- 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 Arrow consortium and liaison organisations

8

- **Libraries**
  - *EDL Foundation*
  - **National libraries in France**, Germany, Spain, The Netherland, UK, Norway, Slovenia, *Finland, Italy*
  - University library of Innsbruck
- **Publishers**
  - **Federation of European Publishers**
  - Publishers associations of Italy, Spain, Germany, **France**, *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 members
  - ISNI ISO-WG members

## The scope: rights information, not rights management

9

- 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
- We'll not enter any rights management, which is the business of some of our partners
  - ▣ We'll redirect users to the appropriate clearing center, if any exists

# The timeframe: milestones

10

- October 2008: start up
- August 2009: architecture design
- March 2010: First pilot
- March 2011: The system alive

# The work done up to now

11

- 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 from the studies

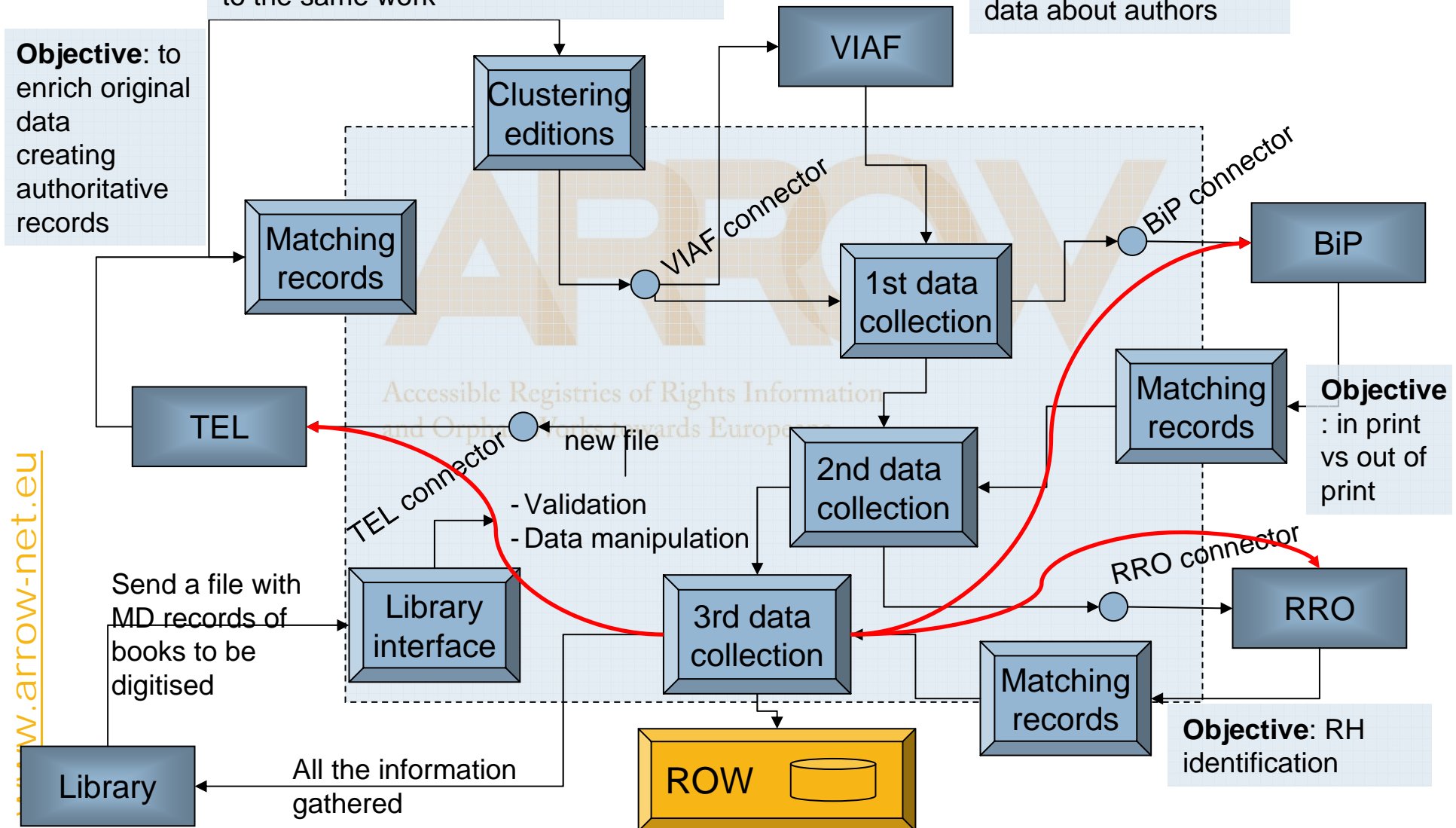
# The ARROW work flow

12

**Objective:** to group all records referred to the same work

**Objective:** to collect data about authors

**Objective:** to enrich original data creating authoritative records



www.arrow-net.eu

# The library question

13

- The library provides one bibliographic record per book
  - **+ specification of requested use(s)**
- We defined
  - Metadata schemas accepted (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, etc.
  - 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

14

- 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

15

- 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

16

- 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

17

- 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

18

- 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”)

19

- 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

## Triggering standard registration and deployment

20

- After the first “matching” (book identification), it is possible to trigger the ISBN registration for old books
  - ▣ So to allow different resources to register and re-use it
- After the “clustering” (work identification), it is possible to trigger the ISTC registration
  - ▣ So to allow different resources to register and re-use it
- After the “name” identification, it is possible to trigger the ISNI registration

# A changing environment

21

- The emergence of the Google Settlement (exactly when we started) influenced our work
  - We needed to consider the new arena
  - The Book Rights Registry 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:

## 1. The (first) “Settlement database”

22

- 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 had strong incentive in the Settlement to “kill the parents”
  - ▣ The more the “orphans” are, the higher their revenues
  - ▣ This is emphasised in the AIE objection to the Settlement (published in [www.aie.it](http://www.aie.it))

## 2. Relations between Arrow and the BRR (which is different than Google)

23

- The relations with the forthcoming Book Rights Registry are friendly
- Contacts have been established: we both recognise the importance of collaboration
- Possible objectives:
  - The Settlement can use the Arrow infrastructure to look for titles / rightsholders in Europe
  - The BRR can be seen as a node of the forthcoming Arrow infrastructure in providing data about books / rightsholders

# Arrow and the BRR

(please note: not Arrow and Google)

24

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

## Arrow and the BRR / 2

25

ARROW	BRR
<ul style="list-style-type: none"><li>□ Business neutral</li></ul>	<ul style="list-style-type: none"><li>□ Serves one business model (though scalable for others)</li></ul>
<ul style="list-style-type: none"><li>□ Resources: <b>5 million euro</b> (2,5 mil EU contribution)</li></ul>	<ul style="list-style-type: none"><li>□ Resources: <b>35 millions dollars</b></li><li>□ <b>Plus additional money</b> to search rightholders</li></ul>

- In synthesis: we have to do much more difficult things with much less money...

# The Amended Settlement (ASA)

## Two aspects related to ARROW

26

1. In principle exclusion of continental European works (though in a highly ineffective way)
  - ▣ Scanning and showing snippets remain a clear infringement of copyright law
  - ▣ However: other use are not permitted via the Settlement
  - ▣ If Google wants to use European books must ask permission
  - ▣ **Can ARROW be the gateway for that?**
2. 60 days between “non commercially available determination” and starting use of the book
  - ▣ The BRR may carry a diligent search in that 60 days
  - ▣ **Can we push the BRR moving toward working on the “European principles”?**
  - ▣ (Though treatment of “orphans” will remain different)

## A long term view: the Rights Information Infrastructure again

27

- We are firmly opposing the Google Settlement
  - ▣ Infringement of the Berne convention
  - ▣ Risk of monopoly in the digital book sector
- We risk to miss a value in the abstract model behind
  - ▣ A business model based on “selling content”, including micro-transactions
  - ▣ Just the opposite of solutions like the “license global”, or similar
- What we lack to save the model while avoiding the risks
  - ▣ Again: an infrastructure to reduce transactional costs
  - ▣ Based on standards to avoid technological or commercial lock in
- Will ARROW help on this?
  - ▣ It will depend on the commitment of many

**Thank you for your attention**

**FURTHER INFORMATION**

Piero Attanasio

[piero.attanasio@aie.it](mailto:piero.attanasio@aie.it)



Accessible Registries of Rights Information  
and Orphan Works towards Europeana

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

Co-funded by the  
Community  
programme  
*eContentplus*

