

# Digitalia

Anno VI, Numero 2 - **2011**

Rivista del digitale nei beni culturali

ICCU-ROMA

# La gestione dei diritti d'autore nelle biblioteche digitali. Il caso di Arrow

**Piero Attanasio**

*Associazione Italiana Editori*

## Il contesto economico giuridico

La gestione dei diritti nei programmi di digitalizzazione su larga scala è un tema che si articola in tre diverse dimensioni: una economica, che attiene ai modelli di investimento istituzionali e – nel caso di iniziative private – commerciali; una giuridica, relativa agli strumenti legali utilizzabili per una gestione efficiente ed equa dei diritti; una infine connessa agli strumenti tecnologici con i quali gestire il processo di autorizzazione e remunerazione. I tre aspetti non possono essere trattati in modo separato, perché profondamente interconnessi. D'altro canto, è necessario tenerli distinti sul piano logico, per maggiore chiarezza di analisi.

In termini economici la questione può essere vista come un classico problema di elevati costi di transazione. Laddove si voglia digitalizzare una collezione ampia di opere, il costo della ricerca degli aventi diritto e della negoziazione individuale con ciascuno di essi può essere ben più alto della remunerazione richiesta dall'avente diritto. Lo ha illustrato con estrema chiarezza Hal Varian<sup>1</sup>: in un mercato efficiente, in genere sia chi vende sia chi compra si impegna a cercare l'altra parte (rispettivamente un compratore per la propria merce e un venditore che soddisfi i propri bisogni). Nel commercio librario, quando gli aventi diritto stanno attivamente commercializzando una certa opera, il costo di ricerca per un acquirente è sufficientemente basso da far sì che milioni di transazioni avvengano in modo semplice, quasi naturale, sul mercato. Nell'ambito dei programmi di digitalizzazione, perché la transazione avvenga, spiega Varian, è necessario che il valore ( $v$ ) attribuito dalla biblioteca al diritto che intende acquisire (es.: digitalizzare e rendere disponibile online quella specifica opera) sia superiore al prezzo che pagherà ( $p$ ) cui va sommato il costo della ricerca ( $c_b$ ). Allo stesso tempo, è necessario che il ricavo per il venditore, dato dal prezzo ( $p$ ), sia sufficientemente maggiore al costo di ricerca di questi ( $c_s$ ) così da consentirgli di coprire i costi marginali di produzione<sup>2</sup>.

Nel caso dei programmi di digitalizzazione su larga scala, la situazione può essere molto problematica. Il valore di ogni singola transazione può essere minimo, mentre è l'insieme ad essere giudicato prezioso: da un'istituzione pubblica in ragione

<sup>1</sup> H. R. Varian, "Copyright term extension and orphan works", *Industrial and Corporate Change*, Vol. 15, 2006, 6, pp. 965-980.

<sup>2</sup> In termini analitici, usando la notazione di Varian, una transazione avverrà se esiste un prezzo  $p^*$  tale per cui per il compratore  $v - p^* - c_b \geq 0$  e per il venditore  $p^* - c_s \geq 0$ .

## Progetti

della sua funzione, da un'azienda privata laddove – come può accadere in questi casi – il valore di un'offerta ampia è significativamente maggiore della somma dei valori di ogni singolo elemento dell'offerta<sup>3</sup>. Varian trova in questa caratteristica la *ratio* economica della legislazione sulle opere orfane. È necessario che vi sia un limite al livello di ricerca che l'acquirente deve fare<sup>4</sup>. Oltre tale livello di ricerca, una norma deve prevedere che l'utilizzo sia autorizzato sulla base di un meccanismo automatico, che alternativamente prevede una remunerazione standard pagata a una società di gestione collettiva<sup>5</sup> o un sistema di limitazione di responsabilità nei confronti dell'eventuale avente diritto che dovesse comparire successivamente: si predetermina un pagamento forfetario che si versa solo in caso di rivendicazione del diritto da parte di un autore o un editore.

Fin qui, nulla di nuovo. L'esercizio può persino apparire una sterile formalizzazione di ciò che è già molto noto per altre vie: ci dice soltanto in modo formale che il costo della ricerca può, in taluni casi, essere così oneroso da inibire questo tipo di progetti. Tuttavia, l'analisi ci consente di individuare alcuni punti chiave dal punto di vista economico, con precise conseguenze sulla valutazione dell'efficacia di possibili soluzioni giuridiche e sull'individuazione degli obiettivi che devono darsi gli sviluppi tecnologici per creare strumenti che servano il mercato in modo efficiente.

Le variabili che entrano in gioco sono tre:

- il valore ( $v$ ) della singola transazione, definito nelle biblioteche digitali come l'utilità percepita da una biblioteca per l'inserimento di una certa opera nella collezione digitale che vuole proporre;
- il prezzo ( $p$ ) chiesto dall'avente diritto per concedere l'autorizzazione all'inserimento della stessa opera nella collezione;
- il costo sostenuto dall'acquirente ( $c_b$ ) e dal venditore / avente diritto ( $c_s$ ) di ricerca e negoziazione per concludere il contratto.

Sia il valore<sup>6</sup> per l'utente sia il prezzo dipendono a loro volta da due fattori: il tipo di opera e l'uso che viene richiesto. Partendo da quest'ultimo, il valore in gioco sarà diverso se l'opera viene digitalizzata per intero e messa a disposizione del

<sup>3</sup> Il caso di Google è paradigmatico: il modello commerciale di Google Books è fondato appunto sull'ampiezza dell'offerta fino all'aspirazione dell'onnicomprendività.

<sup>4</sup> Nelle legislazioni in discussione, inclusa la proposta di Direttiva europea sulle opere orfane, si prevede che la ricerca di un avente diritto debba essere *diligente*, il che già significa che non deve durare all'infinito, e *ragionevole*. *Reasonably diligent search* è anche la definizione nella proposta di regolamentazione in discussione in USA citata da Varian.

<sup>5</sup> È ad esempio questa la soluzione adottata in Canada, uno dei pochi paesi dove la discussione sul tema si è infine tradotta in una legislazione.

<sup>6</sup> In questa analisi il punto di riferimento è appunto il valore per l'utente non un presunto danno che il titolare riceverebbe dall'uso dell'opera, come talvolta si tende a considerare. L'autorizzazione ad utilizzare un'opera protetta è uno dei fattori di produzione del servizio e come tale viene trattato.

## Progetti

pubblico in Internet ad accesso aperto o se si fa un uso più limitato, in termini di porzione di opera messa a disposizione<sup>7</sup> o di modalità di accesso<sup>8</sup>. Una distinzione che diviene sempre più centrale, tuttavia, è appunto quella di categoria di opere. Il valore (e il prezzo) dell'uso di un'opera sono determinati dalla data di pubblicazione (in media, più un'opera è datata minore è il prezzo), dall'essere o no ancora in commercio e dalla natura dell'opera, in quanto – sempre in media – per le opere di saggistica, specie accademiche, l'autore può avere obiettivi non economici alla diffusione dell'opera. Vi sono casi in cui è probabile che l'autore, nell'ipotesi che sia l'unico titolare del diritto, sia disponibile a licenziare gratuitamente l'opera, proprio perché non si attende più un ritorno economico dal suo sfruttamento o non ha mai puntato su questo, essendo più interessato a far conoscere il proprio pensiero.

Anche il costo di ricerca e negoziazione è riconducibile ad alcuni fattori, tra i quali la distinzione tra opere in commercio e opere fuori commercio è cruciale. Se un'opera è ancora in commercio, infatti, almeno l'editore è facilmente individuabile e rintracciabile. Il discrimine tra opere in commercio e fuori commercio è dunque un fattore comune nella determinazione sia del valore dei diritti in gioco sia dei costi di transazione. Non stupisce, allora, che il dibattito sui programmi di digitalizzazione si è spostato dal tema delle opere orfane a quello dei fuori commercio. Si può provare a leggerlo all'interno del modello economico qui richiamato.

La soluzione normativa logicamente conseguente al problema delle opere orfane è quella di creare un regime speciale di gestione dei diritti tale per cui, laddove un utente abbia effettuato una ragionevole ricerca in modo diligente, possa con certezza giuridica utilizzare quell'opera. Vi possono essere soluzioni diverse nel dettaglio<sup>9</sup>, ma il principio base è quello. Il costo della ricerca deve essere comunque sostenuto opera per opera ed è possibile che, *almeno per alcune categorie di opere*, questo sia considerato troppo oneroso.

<sup>7</sup> L'esempio più evidente è quello del *library programme* di Google Books, dove le opere digitalizzate sono utilizzate solo ai fini di indicizzazione e rese accessibili solo per limitatissimi estratti, i cosiddetti *snippets*. È a questa tipologia di uso che Varian fa riferimento in un altro articolo, in cui propone un'applicazione pratica degli stessi principi sopra illustrati: H.R. Varian, *The Google Library Project*, Technical Report, UC-Berkeley, 2006, <http://people.ischool.berkeley.edu/~hal/Papers/2006/google-library.pdf>.

<sup>8</sup> Ad esempio, nel programma norvegese denominato BokHylla ([www.nb.no/bokhylla](http://www.nb.no/bokhylla)), i libri digitalizzati sono leggibili solo a video, senza possibilità di scarico o di stampa, e solo da utenti norvegesi, cfr. V.M. Skarstein, 2009, "Strategies for a Digital National Library", *World Library and Information Congress: 75th IFLA General Conference and Council*, 23-27 August 2009, Milan, [www.ifla.org/files/hq/papers/ifla75/190-skarstein-en.pdf](http://www.ifla.org/files/hq/papers/ifla75/190-skarstein-en.pdf).

<sup>9</sup> Ad esempio la proposta di Direttiva europea proposta nel maggio 2011 dalla Commissione (e in discussione al momento della redazione di questo articolo: novembre 2011) prevede che solo alcune categorie di utenti possano utilizzare le opere, in Canada la stessa possibilità è accessibile a chiunque, a certe condizioni. Una ulteriore distinzione, già richiamata, è tra un regime basato su una eccezione che consente l'uso con remunerazione solo in caso di ricomparsa dell'avente diritto e un regime di licenza che richiede un'autorizzazione concessa da una società di gestione collettiva.

## Progetti

L'esperienza di questi anni sembra indicare che la mera distinzione tra opere orfane e non orfane è di scarsa utilità.

Proviamo allora a estendere il ragionamento, cercando di andare più nel dettaglio (vedi grafico 1). Se consideriamo una certa categoria di opere dell'ingegno (qui si parlerà solo di opere letterarie pubblicate in libri e il cui utilizzo è richiesto in una biblioteca digitale, pubblica o privata), la prima distinzione da fare è (1) tra opere protette e in pubblico dominio. A valle di ciò, tra le opere protette bisogna distinguere tra (2) opere in commercio e fuori commercio. In termini di tutela giuridica in verità non vi è alcuna differenza tra queste due categorie, ma – come visto – lo status commerciale dell'opera è la variabile più importante nel determinare il rapporto tra prezzo dei diritti e costi di transazione che a sua volta può suggerire nuove soluzioni normative volte a risolvere il problema.

Le opere orfane (3) sono un sottoinsieme delle opere fuori commercio, essendo certo, per definizione, che un'opera in commercio ha almeno un avente diritto – l'editore – facilmente individuabile. La suddivisione a questo livello non è binaria: sono tre le categorie in cui è opportuno suddividere le opere fuori commercio: le orfane, quelle con tutti gli aventi diritto noti (e localizzati) e quelli con parte degli aventi diritto noti e parte non conosciuti o non localizzabili. Per comodità, possiamo chiamare quest'ultima categoria come "opere semi-orfane".

Infine, tra le opere fuori commercio con titolari noti è utile un'ulteriore suddivisione, tra opere che hanno titolari attivi che, quando interpellati, rispondono a una richiesta di autorizzazione (positivamente o negativamente, non importa) e quelle con titolari non attivi, che quando interpellati non rispondono. Quest'ultima distinzione attiene ancora al problema dei costi di transazione, in quanto questi non si limitano alla ricerca dell'avente diritto (fattore su cui più spesso ci si concentra) ma anche alla successiva negoziazione individuale.

Se i titolari nemmeno rispondono a una sollecitazione, il costo (in termini di tempo impiegato dalla biblioteca per ottenere tale risposta) può essere di nuovo così alto da rendere non conveniente il rapporto tra valore dell'acquisizione e costo di transazione più remunerazione.

Si può descrivere il dibattito in corso in tutto il mondo sulle possibili soluzioni giuridiche per facilitare i programmi di biblioteche digitali che includano opere protette<sup>10</sup>, constatando che essi possono riguardare, all'interno di una data collezione:

- a) Tutte le opere, senza distinzione in ragione della identificabilità dei titolari dei diritti, quindi inclusi i libri in commercio. È l'unico caso in cui non è necessaria alcuna ricerca preventiva. È inevitabile, tuttavia, che tali programmi – riguardando ap-

<sup>10</sup> Preferisco accantonare, in questo contesto, la problematica relativa alla distinzione tra opere protette e in pubblico dominio, che – sempre pensando a programmi su larga scala – può non essere banale, anche per la coesistenza di diritti sulla forma grafica attribuiti all'editore anche su opere protette, o di diritti su altre opere pure incluse in un libro che si intende digitalizzare (illustrazioni, traduzioni, introduzioni, apparati critici, ecc.).

## Progetti



Figura 1 – Schema delle categorie di opere dell'ingegno

punto qualsiasi opera – abbiano limitazioni in termini di usi concessi, che dovranno essere limitati a sfruttamenti assolutamente secondari, per evitare che il programma entri in competizione con lo sfruttamento ordinario dell'opera<sup>11</sup>. I regimi speciali di gestione dei diritti per le biblioteche digitali, se applicati a tutte le opere, possono assumere forme diverse, ma non possono sfuggire a questa regola<sup>12</sup>.

b) Le opere fuori commercio, per le quali in circostanze diverse sono state proposte soluzioni basate su meccanismi di *opt out*, in cui è il titolare del diritto a dover negare l'autorizzazione e non la biblioteca a doverla ottenere preventivamente. Si deve notare che in questo caso alcuni costi di ricerca sono comunque necessari per la corretta individuazione dello stato commerciale di ogni singola opera.

c) Le sole opere orfane, secondo le linee sopra descritte.

Questi casi sono definiti in ragione di parte dei costi di transazione, relativi alla ricerca del titolare. Tuttavia, come detto, vi è un costo anche nella negoziazione. Un quarto caso può quindi, in linea di principio, delinearsi definito dalla limitazione del regime speciale a:

d) Le opere orfane più le opere con titolari non attivi, pur noti. In questo caso la soluzione legale può basarsi su una gestione collettiva e regole di consenso assenso: il titolare silente viene avvertito dei termini della licenza e, se continua a tacere, il suo consenso si presume<sup>13</sup>.

<sup>11</sup> In termini giuridici, il non rispetto di questa condizione violerebbe il *three steps test* sancito dalla Convenzione di Berna, qualunque sia la forma giuridica attraverso il quale venisse realizzato.

<sup>12</sup> Il caso paradigmatico è il già citato BokHylla in Norvegia. La tecnica utilizzata è quella delle *extended collective licenses*, cosicché la limitazione negli usi è stata definita all'interno di una negoziazione collettiva, ma il discorso non cambierebbe se la tecnica giuridica fosse diversa. In altri paesi del Nord Europa vi sono progetti per introdurre sistemi simili.

<sup>13</sup> Una soluzione simile è compresa nelle modalità di gestione dei diritti in alcune società di gestione collettiva. Il titolare associato conferisce un mandato i cui termini possono essere modificati (in genere entro certi limiti) senza il bisogno del consenso esplicito, ma con l'obbligo per la società di comunicare il tenore della modifica, cui l'associato si può opporre. Nelle biblioteche digitali questa tecnica è stata utilizzata dalla società di gestione tedesca, VG Wort, per dar validità – sia pur limitatamente ai propri membri – all'accordo stipulato con la biblioteca nazionale.

## Progetti

### Infrastrutture per la gestione delle informazioni sui diritti. Il caso di Arrow

Il superamento dell'empasse in cui si trova la gestione dei diritti nei programmi di digitalizzazione va cercata in una combinazione tra innovazione tecnologica e innovazione normativa. Se, come evidenziato dall'analisi, il primo problema è quello della carenza di informazioni sui diritti, e in particolare sulla loro titolarità, una soluzione proposta è quella di creare dei "registri" delle opere dove queste informazioni siano presenti<sup>14</sup>. In termini di innovazione normativa ci si è spinti a proporre che tali registri siano obbligatori e, nella formulazione più estrema, costitutivi del diritto. Com'è noto, ciò contrasta con la Convenzione di Berna, che postula che il diritto nasca in capo all'autore al momento stesso della creazione, senza necessità di alcuna registrazione, il che ha spinto alcuni a proporre una sua modifica<sup>15</sup>. Tuttavia, a parte la scarsa praticabilità tecnico-giuridica della proposta, val la pena di soffermarsi su cosa questo significherebbe nell'equilibrio più generale del diritto d'autore. La norma sull'assenza di registrazione ha una sua *ratio* precisa nella tutela dei creatori come parte debole in un contesto industriale. Sacrificare questa funzione essenziale per risolvere un problema di costi di transazione sembra essere sproporzionato.

Resta l'esigenza di gestire le informazioni sui diritti. Si può allora pensare a "registri" volontari, con funzioni specifiche in questo campo di applicazione o più generici. È in questo contesto che può essere meglio compresa l'esperienza di Arrow, il sistema europeo sviluppato all'interno di un omonimo progetto cofinanziato dalla Commissione europea e che oggi si sta muovendo per divenire una risorsa tendenzialmente paneuropea, grazie a un secondo progetto, che nuovamente gode di finanziamenti comunitari, denominato Arrow Plus.

Arrow è un sistema – già sperimentato con successo in quattro paesi europei: Regno Unito, Francia, Germania e Spagna – per facilitare la ricerca di informazioni sui diritti su opere candidate a entrare in programmi di digitalizzazione su larga scala. Lo ha sviluppato un consorzio molto ampio, coordinato dall'Associazione Italiana Editori, che vede la presenza – tra gli altri – degli organismi europei rappresentativi di tutte le parti in causa: autori, editori, società di gestione collettiva e biblioteche<sup>16</sup>.

<sup>14</sup> Lo stesso Varian, nel corso di una audizione WIPO, insiste sul punto. Cfr. K. Mara, "High Copyright Transaction Costs Cause 'Friction', Google Economist Tells WIPO", *Intellectual Property Watch*, 16 Sept 2010.

<sup>15</sup> È il caso del Rapporto finale del Comité de Sage istituito dalla Commissione europea sulle biblioteche digitali, cfr. *The New Renaissance. Report of The 'Comité Des Sages' - Reflection Group On Bringing Europes Cultural Heritage Online*, Bruxelles, European Commission, 2010, pp. 19-20, [http://ec.europa.eu/culture/documents/report\\_comite\\_des\\_sages.pdf](http://ec.europa.eu/culture/documents/report_comite_des_sages.pdf).

<sup>16</sup> Per notizie complete sulla partnership si veda <http://www.arrow-net.eu>. Qui basti ricordare la presenza – tra Arrow e Arrow Plus – di European Writers Congress, European Visual Artist, Federation of European Publishers, International Federation of Reproduction Rights Organisations, the TEL Office e la Europeana Foundation nonché quella delle biblioteche nazionali di Francia, Germania, Norvegia, Paesi Bassi Regno Unito, Slovenia e Spagna. La forte impronta italiana è sottolineata, oltre che dal coordinamento di AIE, dal fatto che gli sviluppi tecnologici sono prevalentemente affidati al Cineca e dalla presenza di ICCU in Arrow Plus, come coordinatore del National Arrow Group italiano.

## Progetti

Arrow non è però un “registro”, come inteso nel dibattito generale sul tema, o ancor meno una “banca dati dei diritti”, nonostante sia spesso citato come tale<sup>17</sup>. È invece un sistema che integra fonti diverse per recuperare informazioni sui diritti dove queste possono risiedere. Il principio è quello di costruire una infrastruttura, basata su standard riconosciuti internazionalmente, che rende interoperabili fonti diverse per gestire informazioni sui diritti. Secondo l’espressione coniata per l’occasione, si tratta di una *rights information infrastructure* (RII), definizione che intende sottolineare due aspetti che la caratterizzano. Essa si limita a gestire “informazioni sui diritti”, come funzione separata – per quanto logicamente connessa – con quella della gestione dei diritti<sup>18</sup>. Il termine “infrastruttura” intende inoltre sottolineare la *neutralità* della sua natura, nel senso che le informazioni possono essere utilizzate da più soggetti e secondo finalità diverse proprio perché l’infrastruttura non serve un particolare modello di gestione dei diritti.

Dal punto di vista tecnico, le “informazioni sui diritti” sono intese come un insieme di metadati necessari per la gestione dei diritti. Nel contesto delle biblioteche digitali e seguendo la logica prima esposta, queste informazioni comprendono:

- 1) Una precisa identificazione della *manifestazione* che si intende digitalizzare e includere in una collezione online. Poiché può trattarsi di libri editi prima della creazione dell’ISBN, questa identificazione non può essere condotta sulla sola base di un codice standard, ma deve includere il trattamento di metadati bibliografici descrittivi utilizzati in questo caso a scopo identificativo.
- 2) Una precisa identificazione dell’*opera* inclusa nella manifestazione<sup>19</sup>, in quanto

<sup>17</sup> Tra gli altri cfr. *The New Renaissance*, cit., p. 18, che appunto cita Arrow come “rights information database” o – ancor più singolare – la Proposta di direttiva sulle opere orfane della Commissione europea, che in allegato cita Arrow come una delle fonti da utilizzare per la ricerca degli aventi diritto, di nuovo includendolo nella categoria “database”. Opportunamente, il Governo italiano in sede di Consiglio ha proposto un emendamento (<http://register.consilium.europa.eu/pdf/en/11/st15/st15882.en11.pdf>) volto a includere Arrow piuttosto tra le “Sources that integrate multiple databases and registries”.

<sup>18</sup> Sotto questo aspetto una RII non coincide con i “copyright management systems (CMS), [which] are basically databases that contain information about content (works, discrete manifestations of works and related products) and, in most cases, the author and other rightsholders. That information is needed to support the process of authorizing the use of those works by others. A CMS thus usually involves two basic modules, one for the identification of content and rightsholders, the other for licensing” (D.J. Gervais and A. Maurushat, “Fragmented Copyright, Fragmented Management: Proposals to Defrag Copyright Management”, *Canadian Journal of Law and Technology*. Vol. 2, 2003, p. 15-23, corsivo mio). Una RII è limitata al primo dei due moduli.

<sup>19</sup> Ciò richiede una definizione del concetto di *opera*, che in questo caso non coincide con quella del modello FRBR. Poiché il punto di riferimento deve essere la gestione dei diritti, occorre infatti considerare quali aventi diritto sono coinvolti e sotto quali profili. Diviene allora rilevante considerare da un lato le *espressioni* (come definite in FRBR), che hanno esattamente gli stessi titolari dei diritti nonché alcune, ma non tutte, le altre *espressioni* collegate che possano essere considerate in diretta concorrenza con questa.

Semplificando, come sarà più chiaro nel prosieguo della descrizione del funzionamento del sistema, tra le espressioni collegate ci si limita a quelle a stampa e non si considerano le traduzioni

## Progetti

l'oggetto di ogni transazione sui diritti è l'opera dell'ingegno e non una sua singola manifestazione.

- 3) Una precisa identificazione dei soggetti che potenzialmente hanno diritti sull'opera, ovvero autori e altri creatori da un lato (il termine tecnico di *contributor* riassume il concetto) e gli editori dall'altro. Ciò richiama la necessità di accedere ad *authority file* di origine bibliotecaria, a liste altrettanto autorevoli di editori (ad esempio presenti nelle agenzie ISBN) e a repertori di società di gestione collettiva che possono avere liste (oltre che contatti) di autori ed editori con una corretta gestione della relazione tra opera e avente diritto.
- 4) Una corretta determinazione dello status commerciale (essenzialmente in commercio / fuori commercio<sup>20</sup>) a livello di opera, con l'avvertenza che le fonti esistenti rilevano questa informazione invece a livello di manifestazione. In questo caso lo status commerciale è determinato tenendo conto che un'opera è da considerarsi fuori commercio se *tutte* le manifestazioni che la contengono sono fuori commercio<sup>21</sup>.
- 5) L'individuazione dei soggetti che possono concedere (o negare) un'autorizzazione per l'uso richiesto sulla base del contesto legale in cui il programma di digitalizzazione si sviluppa. Da questo punto di vista, Arrow è completamente neutrale rispetto a diverse soluzioni normative possibili. In sostanza – anche se non sempre ciò è semplicissimo – è sufficiente settare i parametri relativi al contesto legale nel sistema per recuperare le informazioni necessarie.

È facile notare come i primi tre sottoinsiemi di informazioni attengono alla "individuazione" dell'avente diritto, il quinto riguarda la sua localizzazione: sono in ge-

come facenti parti della stessa opera. Si può notare come il concetto è molto simile a quello adottato all'interno del Google Settlement laddove si parlava di *principal work* con lo stesso significato.

<sup>20</sup> In verità, nello standard ONIX – generalmente utilizzato dalle banche dati dei libri in commercio – vi sono due campi distinti che possono essere rilevanti: il *publishing status* dove sono registrate le dichiarazioni di messa fuori catalogo di un libro (NB, non dell'opera) e i dati di *commercial availability*, dove sono registrate le informazioni di disponibilità commerciale di un libro in un dato momento. La prima variabile è la più propria in questo contesto, perché evita che si consideri fuori commercio un libro, per esempio, in ristampa o per il quale è in preparazione una nuova edizione. Tuttavia, in termini pratici, la scelta di quali variabili utilizzare deriva dall'effettiva qualità degli stessi nelle specifiche banche dati, cosicché talvolta il dato di *commercial availability* può essere utilizzato – con la necessaria cautela – come proxy o a integrazione del *publishing status*.

<sup>21</sup> Tale definizione corrisponde, operativamente, a quanto stabilito dal Memorandum of Understanding su "Key Principles on the Digitisation and Making Available of Out-of-Commerce Works" ([http://ec.europa.eu/abz\\_en.htm](http://ec.europa.eu/abz_en.htm) siglato nel settembre 2011 (quando Arrow era già stato sviluppato) dalle maggiori organizzazioni rappresentative di utenti e aventi diritto sotto l'egida della Commissione europea, il quale recita: "a work is out of commerce when the whole work, in all its versions and manifestations is no longer commercially available in customary channels of commerce, regardless of the existence of tangible copies of the work in libraries and among the public (including through second hand bookshops or antiquarian bookshops)".

## Progetti

nere queste le due fasi alle quali ci si riferisce quando si parla di “ricerca” nel senso descritto all’inizio di questo articolo. Il quarto sembra atipico, sotto questo profilo, ma è in verità da un lato funzionale all’identificazione degli aventi diritto, perché per le opere fuori commercio possibile che gli autori abbiano riacquisito tutti i diritti, dall’altro lato corrisponde alla logica descritta per cui per i fuori commercio – e solo per questi – si possono ipotizzare meccanismi di gestione particolari.

Per la raccolta ed elaborazione di queste informazioni Arrow interroga e rende interoperabili banche dati di origine diversa, perché i singoli metadati possono essere collocati in luoghi diversi. Si tratta in genere di fonti di dati costruite per ragioni diverse (cataloghi bibliografici o commerciali, repertori di società di gestione), cosicché i dati devono essere rielaborati e combinati così da estrarne il maggior valore ai fini della determinazione dei diritti. La descrizione del funzionamento pratico di Arrow spiega meglio il meccanismo.

Arrow può essere interrogato da una biblioteca sottoponendo un file che contiene i record bibliografici dei libri che si desidera digitalizzare. La prima operazione che il sistema fa è l’identificazione della manifestazione. Ciò avviene attraverso il confronto con TEL – The European Library, il catalogo unico delle biblioteche nazionali europee. Per ciascun record sottoposto viene operato un *matching* con TEL al fine di identificare il libro e di estrarre il record autorevole, prodotto dalla biblioteca nazionale del paese di prima pubblicazione che – da quel punto in poi – sarà utilizzato nel processo. L’operazione è di estrema delicatezza, perché ogni algoritmo di *matching* è costruito su base probabilistica e un eventuale errore si ripercuoterebbe nelle fasi successive. Per questa ragione la biblioteca è chiamata a validare manualmente il *matching*, anche se il sistema è personalizzabile stabilendo che al di sopra di una certa soglia di similitudine tra record originale e record TEL la validazione manuale è superflua.

A partire dal record di TEL, sempre sullo stesso catalogo viene operata una operazione di *clustering*, individuando i record riferiti ad altri libri che – con dati livelli di probabilità – contengono la stessa opera. In verità, Arrow crea due tipi di cluster:

- un *primary cluster* che include tutti i volumi che contengono la stessa *espressione* (per usare un linguaggio FRBR), cioè le diverse edizioni dello stesso testo, con gli stessi *contributors*;
- uno o più *secondary cluster*, che includono le manifestazioni che contengono diverse espressioni della stessa opera, ad esempio una nuova edizione, un’edizione illustrata, una diversa traduzione nella stessa lingua, un’edizione critica, ecc.

È importante tenere distinti i due cluster in quanto gli aventi diritto da ricercare possono essere diversi: se si vuole ad esempio digitalizzare un’edizione illustrata si dovranno contattare anche gli illustratori, cosa non necessaria se si intende invece digitalizzare una diversa espressione della stessa opera.

## Progetti

L'insieme di questa operazione consente di identificare correttamente l'*opera* (che rappresenta la seconda operazione nel processo prima descritto) e, oltre a ciò, di identificare tutte le diverse espressioni dell'*opera* stessa e le manifestazioni di ciascuna espressione.

La terza operazione riguarda l'identificazione degli aventi diritto. Per gli autori e, con un livello di precisione per taluni cataloghi meno precisi, altri contributor che possono avere diritti sull'*opera*, il sistema interroga VIAF - *Virtual International Authority File*, l'iniziativa globale che rende interoperabili i nomi autorevoli delle biblioteche. Da questa fonte è possibile recuperare altri dati, quali la data di morte dell'autore o la sua cittadinanza, rilevanti per la determinazione dello stato giuridico dell'*opera* o per meglio calibrare i successivi procedimenti di ricerca (in caso di un autore di altro paese rispetto a quello di edizione si può ad esempio supporre di estendere la ricerca oltre i confini del paese di prima pubblicazione). Infine, VIAF fornisce anche le varianti del nome di ciascun autore, il che consente di migliorare la ricerca nelle fasi successive.

Fino a questo punto, il processo è interamente svolto su fonti di origine bibliotecaria e i messaggi di interrogazione sono basati sullo standard MARC. Una volta raccolte tutte queste informazioni (i dati descrittivi di tutti i libri dei cluster e quelli degli autori) Arrow interroga la banca dati dei libri in commercio (cd Books in Print - BIP), che tuttavia usano lo standard ONIX. Per questa ragione sono operate alcune operazioni di trasformazione del record MARC in un record ONIX.

Anche in questo caso, il procedimento comprende operazioni di *matching* e *clustering*. Sotto il primo profilo, la presenza nei cataloghi TEL degli ISBN facilita l'identificazione, ma un lavoro sul confronto tra metadati è comunque necessario, per coprire i casi in cui il volume richiesto ha un anno di edizione pre-ISBN o quelli in cui il codice non è riportato nel record (come talvolta accade, almeno in alcuni paesi). Una ricerca aggiuntiva per identificare ulteriori edizioni dell'*opera* si rende altresì necessaria per coprire i casi di edizioni molto recenti (che possono essere assenti dal catalogo della biblioteca nazionale), o in preparazione (i BIP ricevono in genere con largo anticipo, per uso nei canali di vendita, l'annuncio di novità), o in edizione elettronica (che talvolta sfugge ai cataloghi bibliotecari, almeno in questa fase storica).

Le operazioni condotte sui BIP sono poi due:

- Una raccolta di informazioni sui nomi autorevoli sugli editori, così da poter meglio gestire - in particolare - la relazione tra marchio (es. Rizzoli) con il quale il libro può essere stato classificato nel campo editore di MARC e nome dell'editore (nell'es. RCS).
- La determinazione sullo stato commerciale a livello di *opera*.

Per quest'ultimo obiettivo, il sistema controlla tutte le manifestazioni raccolte nelle operazioni precedenti rilevando - se disponibile - lo status commerciale diretta-

## Progetti

mente dalle dichiarazioni degli editori, o inserendolo dalle informazioni sulla disponibilità commerciale. Come detto, perché un'opera sia considerata in commercio è sufficiente che lo sia almeno una manifestazione che la contenga. In linea di principio questa può essere nel *primary cluster* o in uno dei *secondary cluster*. Ciò può implicare scelte diverse, che possono essere personalizzate e che in genere è preferibile gestire con un intervento umano di controllo. Se ad esempio tutte le edizioni del *primary cluster* sono fuori commercio, ma è in commercio un'edizione in un *secondary cluster*, ciò può suggerire di fare controlli, perché il tipo di relazione tra espressioni della stessa opera può essere tale da far considerare le due espressioni pressoché equivalenti sul piano commerciale (digitalizzare l'espressione A genera effetti sullo sfruttamento commerciale dell'espressione B). Tuttavia, è anche possibile che gli algoritmi di *clustering* – per definizione affetti da errori statistici – includano in un *secondary cluster* opere (ad esempio con lo stesso autore e un titolo simile) che non hanno alcuna relazione con quella che si vuole digitalizzare. Un processo meramente automatico non può sempre discriminare correttamente<sup>22</sup>, così che è preferibile aggiungere controlli manuali.

L'ultima fase della procedura consente di ricercare gli aventi diritto, autori ed editori che siano. L'insieme delle informazioni raccolte fino a questo momento è infatti la base per interrogare le banche dati delle società di gestione dei diritti per sapere se gli aventi diritto siano ad esse noti e se abbiano conferito un mandato per la gestione dei diritti richiesti. In quest'ultimo caso, il sistema consente di fornire alla biblioteca richiedente le informazioni per acquisire direttamente la licenza.

Alla biblioteca sono fornite delle risposte dirette alla sua domanda iniziale (es.: è possibile digitalizzare l'opera acquisendo la licenza dalla società x; oppure: non è possibile digitalizzare il libro perché l'opera è ancora in commercio, e così via), assieme a tutte le informazioni raccolte – secondo un principio di massima trasparenza – così che, se lo desidera, possa puntualmente verificarne la correttezza.

### Conclusioni

Il sistema è stato sviluppato per essere estremamente flessibile così da adattarsi a diversi contesti legali. Nella situazione giuridica europea attuale, in attesa della Direttiva sulle opere orfane e del suo recepimento da parte degli stati membri nonché dell'attuazione delle politiche delineate dal *Memorandum of Understanding* sulle opere fuori commercio, può essere utilizzato per la ricerca di mandati individuali espliciti da parte di aventi diritto individuati e localizzati. Il primo caso d'uso in questa direzione è a supporto di un progetto di digitalizzazione lanciato da Wellcome Trust, un'importante fondazione britannica che intende creare una biblioteca digitale tematica con libri sulla genetica.

<sup>22</sup> Un semplice esempio è il seguente: *Tirature 2010* e *Tirature 2011* (l'annuario sull'editoria curato da V. Spinazzola) sono due opere diverse; *Zingarelli 2010* e *Zingarelli 2011* sono due espressioni, molto simili, della stessa opera. Nessun algoritmo automatico può distinguere questi due casi.

## Progetti

In caso di recepimento della Direttiva sulle opere orfane – almeno nel testo noto al momento in cui scrivo (novembre 2011) – il sistema potrà essere utilizzato (e anzi, come detto, la Direttiva auspica che sia usato) come supporto della ricerca diligente degli aventi diritto. Saranno gli stati membri a decidere se la ricerca tramite Arrow possa considerarsi sufficiente per dichiarare, in caso di mancata individuazione degli aventi diritto, un'opera orfana o sia invece necessaria un'ulteriore ricerca. La scelta non potrà che dipendere dalla quantità e qualità dei dati disponibili.

Nel caso italiano ritengo che sarà necessario integrare le procedure automatiche di Arrow con una procedura più generica, in Internet e/o su altre fonti, di ricerca degli autori, in quanto le banche dati di SIAE e AIDRO (le società di gestione collettiva in questo ambito) sono piuttosto limitate in confronto con altri paesi europei<sup>23</sup>. In ogni caso, Arrow consentirebbe di risparmiare molto tempo (e quindi denaro) nella prima fase della ricerca.

La maggiore utilità di Arrow si realizzerebbe, probabilmente, in caso di programmi attuativi del Memorandum sulle opere fuori commercio. In uno schema legale che preveda una concessione automatica dei diritti per le opere fuori commercio, le prime quattro fasi sopra descritte forniscono gran parte della soluzione per determinare, titolo per titolo, se una certa opera ricade nel programma. In questa direzione Arrow sta lavorando in Francia – e vi sono ottime prospettive anche in Germania – in attuazione degli accordi tra biblioteche e rappresentanti degli aventi diritto sulle opere fuori commercio, che tuttavia richiedono alcune innovazioni normative, attese per il 2012<sup>24</sup>. Con una caratteristica ulteriore: in programmi che prevedono meccanismi di licenza legale sui fuori commercio con la possibilità di *opt out* per gli aventi diritto, l'integrazione della fase di ricerca di questi ultimi nello stesso processo utilizzato per la determinazione dei fuori commercio consente di implementare pressoché senza costi aggiuntivi uno schema di notifica individuale agli aventi diritto noti che di fatto limita la forzatura al diritto d'autore derivante dalla previsione di un uso non autorizzato. Non sono più coinvolte tutte le opere fuori commercio, ma solo le opere con aventi diritto sconosciuti o, pur noti e pertanto informati, non attivi.

D'altro canto, poiché in questi casi si prospettano accordi che prevedono una remunerazione per gli aventi diritto, questi dovranno essere in seguito ricercati a fini distributivi. Si tratta allora solo di anticipare il momento della ricerca, con vantaggi sia in termini di maggiore equità del meccanismo legale che si costrui-

<sup>23</sup> La ragione dell'arretratezza è tutta nella scarsa tradizione di gestione di diritti secondari (essenzialmente sulle fotocopie) nel nostro paese. I diritti raccolti annualmente in Italia pro-capite sono circa un decimo che nella media degli altri paesi europei, dove hanno storicamente costituito la base per la ricerca degli autori per poter distribuire i proventi e quindi per la costituzione di ampie banche dati di autori.

<sup>24</sup> Sul caso francese mi permetto di segnalare P. Attanasio, "La digitalizzazione in Francia", *Giornale della Libreria*, aprile 2011, p. 16-18.

## Progetti

sce sia per le biblioteche in quanto una ricerca anticipata consentirebbe di evitare i costi di digitalizzazione per volumi che successivamente devono essere esclusi laddove in seguito un avente diritto decida di esercitare l'*opt out*, richiedendo la rimozione del file.

Una gestione più individuale dei rapporti con gli aventi diritto può consentire altresì di includere nello schema licenze – come le *creative commons* – che consentano agli autori che lo desiderino di autorizzare gratuitamente la digitalizzazione, cosa nient'affatto improbabile per volumi che non sono più sfruttati commercialmente. In una licenza collettiva pura e semplice la remunerazione è sempre dovuta, ed è inevitabilmente determinata su una presunzione "media" di valore dei diritti in gioco. Se invece si includono forme di gestione individuale, perfettamente possibili con strumenti come Arrow, si può prevedere che gli aventi diritto non abbiano solo la scelta tra concedere i diritti alla remunerazione media prestabilita e rifiutare l'autorizzazione, ma anche che possano autorizzare la digitalizzazione gratis, facendo prevalere altri obiettivi, quali la maggiore diffusione e visibilità della propria opera.