

11-22-2016

With Enough Eyeballs All Searches Are Diligent: Mobilizing the Crowd in Copyright Clearance for Mass Digitization

Maurizio Borghi
Bournemouth University

Kris Erickson
University of Glasgow

Marcella Favale
Bournemouth University

Follow this and additional works at: <http://scholarship.kentlaw.iit.edu/ckjip>

 Part of the [Law Commons](#)

Recommended Citation

Maurizio Borghi, Kris Erickson & Marcella Favale, *With Enough Eyeballs All Searches Are Diligent: Mobilizing the Crowd in Copyright Clearance for Mass Digitization*, 16 Chi. -Kent J. Intell. Prop. 135 (2016).
Available at: <http://scholarship.kentlaw.iit.edu/ckjip/vol16/iss1/6>

This Article is brought to you for free and open access by Scholarly Commons @ IIT Chicago-Kent College of Law. It has been accepted for inclusion in Chicago-Kent Journal of Intellectual Property by an authorized editor of Scholarly Commons @ IIT Chicago-Kent College of Law. For more information, please contact jwhipple@kentlaw.iit.edu.

WITH ENOUGH EYEBALLS ALL SEARCHES ARE DILIGENT:
MOBILIZING THE CROWD IN COPYRIGHT CLEARANCE FOR
MASS DIGITIZATION ♦

MAURIZIO BORGHI*

KRIS ERICKSON**

MARCELLA FAVALE***

ABSTRACT

Digitization of 20th Century cultural heritage is severely restricted due to the real or potential subsistence of copyright and related rights. Under the laws on orphan works introduced in many countries, items whose copyright status is uncertain may possibly be lawfully digitized, on condition that a “diligent search” of the copyright owners has been performed. However, carrying out diligent searches on large collections is a lengthy and expensive process, which may discourage institutional users from embarking on large-scale digitization. While the problem of performing diligent searches has been so far approached in a “centralized” manner by individual institutions, the article suggests a de-centralized approach based on crowdsourcing certain phases of the diligent search process. The proposed solution may alleviate the problem of the high costs of diligent search, and may ultimately enable cultural heritage institutions to take full advantage of the orphan works legislation. Suitability of the crowdsourcing solution to the cultural heritage sector is discussed and challenges to implementation are identified.

♦ The authors wish to thank Ruth Towse for her valuable comments on an earlier version of this article. Research for this article has been carried out as part of EnDOW (“Enhancing access to 20th Century cultural heritage through Distributed Orphan Works clearance”), a project funded by the EU Joint Programming Initiative in Cultural Heritage and Global Change. Details on the project are available at <http://diligentsearch.eu>.

* Professor of Law and Director, Centre for Intellectual Property Policy and Management (CIPPM), Bournemouth University.

** Lord Kelvin Adam Smith Fellow in Social Sciences, CREATE, University of Glasgow.

*** Senior Researcher, CIPPM, Bournemouth University.

TABLE OF CONTENTS

INTRODUCTION: DIGITIZATION POLICIES AND THE ISSUE WITH ORPHAN WORKS	136
I. REGULATING THE USE OF ORPHAN WORKS.....	140
A. Ex Post Approach.....	141
B. Ex Ante Approach	143
II. THE CENTRALITY OF DILIGENT SEARCH IN ORPHAN WORKS LEGISLATION.....	143
III. MAPPING THE COST OF DILIGENT SEARCH	149
IV. EXPLORING THE POTENTIAL OF CROWDSOURCING.....	154
A. Economies of Scope	156
B. Economies of Scale	157
V. APPLYING CROWDSOURCING TO SOLVE THE DILIGENT SEARCH PROBLEM.....	159
A. Crowdsourcing Diligent Search for Orphan Works: The Institutional Perspective	161
B. An Online Platform to Enable Diligent Searches.....	164
CONCLUSION.....	165

INTRODUCTION: DIGITIZATION POLICIES AND THE ISSUE WITH ORPHAN
WORKS

Digitization of cultural heritage collections is a key component of a democratic, pluralist and inclusive internet environment. As such, it is an integral part of a broader vision to create a “universal library” comprising all the world’s knowledge, and make it accessible to everyone with an internet connection. This is why conversion of physical collections by libraries, archives and other memory institutions into machine-readable bytes is more than just a technical operation, and has in fact been regarded as a ‘moral imperative’ for our age.¹

1. Statement of Peter Branley, director of technology for the California Digital Library, quoted in Kevin Kelly, *Scan this Book!*, N.Y. TIMES MAGAZINE , April 14, 2006, at 18. See also *The New Renaissance, Report of the ‘Comité des Sages’ on bringing Europe’s cultural heritage online* 2011 14. (“Digitization is more than a technical option, it is a moral obligation”). For a discussion on the

Although large-scale digitization initiatives such as those of the Internet Archive's Collections have been in place as early as in the mid-90s,² it is only in the last fifteen years that mass digitization has achieved prominence in the policy agenda. In 2005, the Google Books project took off, with the goal of digitizing "all the world's books"³ in partnership with major libraries across the globe. One year later, the European Commission announced the decision to promote a joint effort of European cultural institutions to digitize and make available online of their collections.⁴ Europeana, the portal for European cultural heritage, was launched in 2008.⁵ On the other side of the Atlantic, public libraries in the USA partnered in the Digital Public Library of America (DPLA).⁶

Despite the evident connections between all such large-scale digitization projects, the underlying objectives can be radically different. For a market-driven project such as Google Books, the aim of digitization is primarily that of creating *searchable* and *computable* objects, namely a resource to feed search algorithms and, in general, to further the company's mission "to organise the world's information and make it universally accessible and useful".⁷ The unparalleled capacity of a company like Google to exploit the computational potential of digitized books—and of cultural heritage collections in general, as the projects stemming from the Google Cultural Institute now demonstrate⁸—is what makes the strength and uniqueness of Google's model of mass digitization.

The achievements are in fact evident. Google has reportedly digitized over 20 million books, 4 million of which are in copyright.⁹ From a legal perspective, the Google Books project has been developed on the assumption that the operations involved in its model of mass digitization—scanning

compelling nature of digitization *see* MAURIZIO BORGHI & STAVROULA KARAPAPA, COPYRIGHT AND MASS DIGITIZATION 8-15 (2013).

2. INTERNET ARCHIVE BLOGS, <https://blog.archive.org/about> (last visited Apr. 30, 2016).

3. *See* KEN HILLIS, MICHAEL PETIT & KYLIE JARRETT, GOOGLE AND THE CULTURE OF SEARCH 146 (2012).

4. Commission Recommendation of 24 August 2006 on the Digitization and Online Accessibility of Cultural Material and Digital Preservation, O.J. (L 236) 28.

5. EUROPEANA, <http://www.europeana.eu> (last visited April 30, 2016).

6. DIGITAL PUBLIC LIBRARY OF AMERICA, <http://dp.la/info> (last visited Apr. 30, 2016).

7. GOOGLE COMPANY, <https://www.google.co.uk/about/company> (last visited Apr. 30, 2016). For a critical discussion *see* Pamela Samuelson, *Google Book Search and the Future of Books in Cyberspace*, 94 MINN. L. REV. 1308 (2010).

8. *Google Arts & Culture*, GOOGLE, <https://www.google.com/culturalinstitute> (last visited Apr. 30, 2016).

9. Numbers reported in *Petition for a Writ of Certiorari at 1, Authors Guild v. Google Inc.*, 578 U.S. 849 (2016).

books and making only small excerpts available to users in response of search queries—are covered by the fair use defence under US copyright law.¹⁰ Google’s assumption has been eventually upheld by the US Supreme Court, which in April 2016 has put an end to a decade of litigation.¹¹

Digitization in policy-driven projects such as Europeana, however, follow a different model, not only due to lack of a broad and open-ended defense such as the US fair use under European copyright laws¹² but because the very purpose of digitization is different. In line with the public interest mission of cultural institutions, the aim of digitization is primarily to disseminate content to the public at large, and possibly to make it available for creative re-use. Digital libraries are created to be read and used by humans, and not only robots. Whereas Google Books could effectively pursue its mass-digitization model by keeping in-copyright content invisible to users (thereby successfully relying on fair use), digital libraries cannot achieve their public interest goal unless content is made available to users. To date, European cultural institutions have made available around 50 million digital objects through Europeana.¹³ Although these numbers are seemingly impressive, they still represent only a relatively small fraction of the estimated 300 million items held by Europe’s cultural heritage institutions.¹⁴ Most importantly, the corpus of digitized objects comprises only works in which copyright and related rights no longer subsist. This means that a vast majority of *recent* cultural heritage cannot form part of publicly accessible digital libraries.

The problem faced by many cultural institutions is clearing rights in content that may still be protected by copyright, but whose rightsholders either do not (or no longer) actively manage their copyrights, or are simply

10. 17 U.S.C. § 107.

11. Authors Guild v. Google Inc., *cert. denied*, 578 U.S. 849 (2016); see also Authors Guild v. HathiTrust, 755 F.3d 87 (2d Cir. 2014).

12. See Tribunal de grande instance [TGI] [ordinary court of original jurisdiction] Paris 3eme chamber, 2eme section, Dec. 18, 2009, 79 PTCJ 226 (finding that Google infringed the reproduction right in books of French publishers). The decision has been appealed; the case was dismissed in June 2012 based on an agreement between the French Publishers Association (SNE) and Google; see *The National Publishing Union and Google put an end to litigation that pitted them since 2006 on the digitization and indexing of books*, GOOGLE, <http://googlepressfr.blogspot.com/2012/06/le-syndicat-national-de-ledition-sne-et.html> (last visited Apr. 30, 2016). For a comparison between US and Europe’s copyright systems in respect to mass digitization see BORGHI & KARAPAPA, *supra* note 1, at 19-44.

13. Sources – Europeana Collections, EUROPEANA COLLECTIONS, <http://europeana.eu/portal/browse/sources.html> (last visited Apr. 30, 2016).

14. ‘WE TRANSFORM THE WORLD WITH CULTURE’ EUROPEANA STRATEGY 2015-2020 9, available at <http://strategy2020.europeana.eu> (last visited Apr. 30, 2016).

untraceable.¹⁵ To enable digitization and making available of works without rightsholders' consent, *ad hoc* legislative solutions have been adopted in various jurisdictions.¹⁶ In Europe, the policy process initiated with Europeana has led to the adoption of the Orphan Works Directive in 2012, which has introduced a specific copyright exception to the benefit of cultural heritage institutions.¹⁷

The common principle underlying most of the orphan works legislation is the requirement that, before a work is declared to be an orphan, the prospective user should carry out a "diligent search" for the rightsholder(s). To be lawful, this search must meet given conditions that vary from jurisdiction to jurisdiction, and whose standards may be strict or relatively relaxed depending on the legislation.¹⁸ In most cases, however, compliance with this legal requirement involves high costs, or at least costs that may exceed the resources of cultural institutions, especially in times of severe budgetary restrictions. This raises the question of whether orphan works legislation achieve the goal, set by policy makers, to enable wider use and re-use of recent cultural heritage items, in view of developing a truly comprehensive and inclusive "universal library".

This article proposes a solution to the challenge presented by the high costs of diligent search, in order to enable cultural institutions to take full advantage of the legislation. We propose a crowdsourcing platform to enable cultural institutions to distribute the labour costs of diligent search among a large number of institutional patrons or public volunteers. The article is divided into six parts. Part one critically examines the regulatory approaches that can be adopted to address the problem of in-copyright but non-exploitable works. Part two focuses on the diligent search requirement, with a specific focus on the EU legislation on orphan works. Part three addresses the costs involved in right clearance for cultural institutions. Part four introduces the concept of crowdsourcing. Part five discusses how a crowdsourcing model might be applied to solve the problem of diligent search. Part six concludes.

15. Automatic systems to help establishing the public domain status of works have been developed, although they cannot replace individual searches. See, e.g., *The Public Domain Calculator of Europeana OUT OF COPYRIGHT*, <http://outofcopyright.eu> (last visited Apr. 30, 2016).

16. MARCELLA FAVALE, FABIAN HOMBERG, MARTIN KRETSCHMER, DINUSHA MENDIS & DAVIDE SECCHI, COPYRIGHT, AND THE REGULATION OF ORPHAN WORKS: A COMPARATIVE REVIEW OF SEVEN JURISDICTIONS AND A RIGHTS CLEARANCE SIMULATION I (2013).

17. Directive 2012/28/EU of the European Parliament and of the Council of 25 October 2012 on certain permitted uses of orphan works, 2012 O.J. (L 299/5) (hereinafter "Orphan Works Directive").

18. FAVALE, *supra* note 16, at 2.

I. REGULATING THE USE OF ORPHAN WORKS

The problem of in-copyright works that cannot be exploited is inherent in copyright law, which automatically confers the exclusive right on an original work at the time of its creation with no formalities that identify the owner.¹⁹ The long duration of rights vested in a work, which can easily exceed a century from the time the work was created,²⁰ further increases the chances that a work is protected by copyright, even though the author or her heirs (or her possible assignees) no longer actively manage their rights—or even know they own them. When rights are likely to subsist in a work, but no permission can be obtained because the rightsholder is either unknown or untraceable, copyright law does not leave any other option but to desist from using the work, or carry on at the risk of infringing. As a matter of fact, copyright can be enforced to a full extent and at any time, even if the relevant rightsholder has been dormant for years. Clearly, the risk of facing a copyright infringement lawsuit increases exponentially when the use involves a large number of works, possibly bearing several layers of rights.²¹ This is precisely the situation that emerges in large-scale digitization projects on recent cultural heritage.

To address this problem and enable a more efficient use of works that would otherwise remain unexploited, various regulatory solutions have been proposed or adopted. Such solutions can be schematically divided into *ex ante* and *ex post* regulatory systems. Under an *ex ante* system, the would-be user of unexploited works must actively complete certain steps *before* being legally entitled to make certain uses of those works. In an *ex post* system, by contrast, works that qualify as orphans or otherwise unavailable can be immediately used, but specific provisions intervene *after* the use has been made in case the rightsholder intends to exercise his rights in relation to the work. We discuss these two approaches in the following.

19. PAUL GOLDSTEIN & BERNT HUGENHOLTZ, INTERNATIONAL COPYRIGHT PRINCIPLES, LAW, AND PRACTICE, § 1.1 (3rd ed. 2013).

20. In many countries, copyright in literary, dramatic, musical and artistic works lasts for the life of the author plus 70 years. See GOLDSTEIN & HUGENHOLTZ, *supra* note 19, at § 5.3.

21. These include so-called “content copyrights” and “signal copyrights” (*e.g.* rights in sound recording, broadcast, film and typographical arrangement), as well as rights in performance. See Richard Arnold, *Content copyrights and signal copyrights: the case for a rational scheme of protection*, 1 QUEEN MARY J. INTELL. PROP. 272 (2011) (discussion on the interplay of those layers of rights).

A. *Ex Post* Approach

Under *ex post* systems, two kind of provisions can be applied to regulate the use of orphan works. The first kind are measures that reduce the liability for copyright infringement resulting from the use of those works. These can be limitations to remedies available to rightsholders and/or creation of “safe harbours” for qualified users of orphan works. Both measures were included in the proposed US legislation, which did not pass into law.²² Specifically, the legislation limited both monetary and injunctive relief against good faith infringers,²³ and exempted not-for-profit educational and memory institutions from monetary damages if, after receiving “notice of claim of infringement”, they promptly ceased the infringement.²⁴ However, to be eligible for these limitations on remedies, the infringer should have performed a “qualifying search, in good faith, to locate and identify” the copyright owner.²⁵

The second kind of provision that characterizes an *ex post* approach are mechanisms that create a presumption of consent to the use the work in the absence of the permission of the rightsholder. This presumption remains effective unless and until the rightsholder takes active steps to deny consent. Provisions of this kind have been described as “turning copyright on its head”, since they actually alter one of the cornerstones of copyright law, namely that authorial permission is required before (and not after) engaging in a restricted act. This approach found an exemplary application in the Settlement Agreement of the Google Books case.²⁶ The scheme envisaged under the Agreement covered all books already digitized by Google as well as the books that Google would have digitized in the future, and provided rightsholders with the option to claim their rights in a work and either participate in a revenue-sharing scheme or having their books removed from the corpus. Revenues collected on “unclaimed works” would have been used to carry out diligent searches to locate rightsholders and, eventually, would have been used for public interest and charity purposes. The Agreement did

22. See Orphan Works Bill, H.R. 5889, 110th Cong. (2008); and Shawn Bentley Orphan Works Act, S. 2913, 110th Cong. (2008). See also FAVALE, *supra* note 16, at 11-15.

23. Shawn Bentley Orphan Works Act 2008, amending 17 U.S.C. § 514(c)(1)(A) and 514(c)(2).

24. *Id.*, amending 17 U.S.C. § 514(c)(1)(B).

25. *Id.*, amending 17 U.S.C. § 514(b)(1). As we discuss in the following section, such “search” is the key element of *ex ante* systems.

26. See Settlement Agreement, *Authors Guild Inc. v. Google Inc.*, No. 05 CV 8136-JES (S.D.N.Y. Oct. 28 2008); Amended Settlement Agreement, *Authors Guild*, No. 05 CV 8136-JES (S.D.N.Y. Nov. 13, 2009). See also BORGHI & KARAPAPA, *supra* note 1, at 78-80 and 89-90, for a critical analysis.

not receive court approval and remained unapplied.²⁷ However, similar principles inspire the Extended Collective Licensing systems that are applied in various European countries to enable mass digitization of potential orphan works.²⁸ Extended collective licenses are agreements concerning a certain use of works which are concluded between a collective management organization and a user. By operation of the law, the effects of the collective agreement extend also to rightsholders who are not members of the organization. All rightsholders are given the right to claim individual remuneration from the collecting organization. In addition, they have the right to individually prohibit the use of their works under the agreement—or, in other words, a right to opt out. Unless rightsholders have opted out of the scheme, the exercise of their rights will be subject to collective management.

The *ex post* approach to the orphan works problem focuses on removing legal obstacles to the use of those works, in order to maximize public access and re-use. The underlying rationale is straightforward: when a copyright is not actively managed, but someone has an interest in utilizing the work, the general interest is better served by enabling the latter to make use of it—especially when the use is related to a public interest mission. However, as mentioned earlier, the current copyright system may impose critical restrictions to solutions based on an *ex post* approach. In particular, derogations to copyright law's principle of exclusivity are not generally allowed under international obligations.²⁹ Similarly, limitations on remedies cannot deprive the rightsholder of effective means of enforcement.³⁰ For these reasons, the adoption of *ex post* approach to legislation requires close scrutiny and may eventually be subject to compromises.

27. See *Authors Guild, Inc. v. Google, Inc.*, 770 F. Supp. 2d 666 (S.D.N.Y. 2011).

28. Extended Collective Licensing (ECL) scheme are widely used for digitization of cultural heritage in Denmark, Sweden, Finland and Norway. Other European countries have recently introduced ECL or ECL-style systems in their legislation, namely France, Germany, United Kingdom and Slovakia. See Lucie Guibault, *Cultural Heritage Online? Settle it in the Country of Origin of the Work*, 3 JIPITECJ. Intell. Prop. Info. Tech. Elec. Com. L. 173 (2015.), for an analysis of such systems.

29. See in particular Berne Convention for the Protection of Literary and Artistic Works art. 9, Sept. 9, 1886, 25 U.S.T. 1341; Rome Convention for the Protection of Performers, Producers of Phonograms and Broadcasting Organizations art. 11, 13, Oct. 26, 1961, 496 U.N.T.S. 43; WIPO Copyright Treaty art. 8, 10, Dec. 20, 1996, 2186 U.N.T.S. 121; WIPO Performance and Phonograms Treaty art. 7, 10 and 16, Dec. 20, 1996, 2186 U.N.T.S. 203.

30. Agreement on Trade-Related Aspects of Intellectual Property Rights art. 41, 44-46, Jan. 1, 1995, 1869 U.N.T.S. 299.

B. *Ex Ante* Approach

An *ex ante* regulatory approach is based on the principle that a work can only be used without authorization if and only if its rights holder cannot be identified and located after a qualified search. The requirement of a prior search of the copyright owner is the distinctive factor of this approach. As a matter of fact, in most legislation on orphan works, the use of work is premised upon a preventive (unsuccessful) search of the rightholders. The standards of such a search vary considerably and range from the weak provision of Canada (requiring “reasonable efforts to locate the owner of the copyright”³¹) to the more demanding provisions of Japan and the EU, where users have a duty to perform a fully-fledged “diligent search” accompanied by some form of tracking of the steps performed.³²

The aim is primarily to secure fairness to authors and other rights holders, and to avoid the risk that works are used without permission. However, a too strict requirement of preventive search may impose new hurdles and eventually discourage would-be users of orphan works. In other words, the problem with clearing rights, which the legislation is expected to solve, is simply re-introduced by the back door.

Both *ex ante* and *ex post* approaches, as well as the various legal solutions inspired by such approaches, have advantages and disadvantages. No solution can unequivocally be said better than the other from a legal point of view. As a matter of fact, the approaches are not strictly alternative and they can complement each other in various ways. For example, a pure *ex post* system may compromise the legitimate interests of copyright holders, and may be ultimately incompatible with international obligations and other binding legislation,³³ while a system too favored towards *ex ante* provisions may not achieve its purpose and be ultimately ineffective to solve the orphan works problem.

II. THE CENTRALITY OF DILIGENT SEARCH IN ORPHAN WORKS LEGISLATION

Most legislation on orphan works applies an *ex ante* approach or otherwise requires a pre-emptive effort to locate the copyright owner. In this

31. Copyright Act, R.S.C. 1985, c C-42, sec. 77 (Can.).

32. Chosakukenhō [Copyright Act], Law No. 48 of 1970, art. 67; (Japan); Orphan Works Directive 2012/28, art. 3., 2012 O.J. (L 299) 5, 9 (EU). See FAVALE, *supra* note 16, at 49.

33. On this point, with reference to the Slovak ECL system, see Matej Gera, *Extended collective licensing under the new Slovak Copyright Act*, 11 J. INTELL. PROP. L. PRACTICE 170 (2016). See also the pending referral to the Court of Justice of the European Union on the French law on unavailable books of 20th Century (*infra* notes 51-55 and accompanying text).

part, we discuss the EU Directive on Orphan Works, as exemplary of a legislative solution based on a high standard of diligent search. Since the Directive leaves room to EU Member States to adopt national legislative solutions to promote digitization of orphan and out-of-print works, the legislation in some of these states will be briefly discussed.

The orphan works problem was acknowledged in the Green Paper on the Knowledge Economy of 2008 and in the Communication from the Commission 2009.³⁴ In December 2009, the High Level Expert Group on Digital Libraries, within the framework of the i2010 Digital Libraries Initiative,³⁵ recommended the adoption of a common definition of orphan work and of diligent search standards to locate rightholders. A *Memorandum of Understanding on key principles on the digitization and making available of out-of-commerce works* was signed on 20 September 2011 by a number of key stakeholders, including national libraries, research libraries, representatives of creators and publishers.³⁶ Eventually, the European Parliament adopted an Orphan Works Directive on the 25th October 2012.³⁷ The Directive established a new exception to copyright exclusive rights for a number of orphan works.³⁸

However, the new exception in turn is defined by several boundaries, such as defined subject matter, use of the work, users of the work, and ways to perform the diligent search. These limitations, the subject matter is limited to books, journals, newspapers, magazines or other writings, cinematographic or audiovisual works and phonograms,³⁹ and does not

34. *Green Paper on Copyright in the Knowledge Economy*, COM (2008) 466 final, (July 16, 2008); *Communication from the Commission on Copyright in the Knowledge Economy*, COM (2009) 532, final, (Oct. 19., 2009).

35. *Final Report of Digital Libraries: Recommendations and Challenges for the Future*, HLEG (Dec. 2009), <http://www.dlorg.eu/uploads/External%20Publications/HLG%20Final%20Report%202009%20clean.pdf>.

36. *Memorandum of Understanding of the Key Principles on the Digitisation and Making Available of Out-of-Commerce Works*, European Commission for Internal Market and Services, (Sept. 20, 2011), http://ec.europa.eu/internal_market/copyright/docs/copyright-info/20110920-mou_en.pdf>. The MoU was signed by the Association of European Research Libraries (LIBER); the Conference of European National Librarians (CENL); the European Bureau of Library, Information and Documentation Associations (EBLIDA); the European Federation of Journalists (EFJ); the European Publishers Council (EPC); the European Writers' Council (EWC); the European Visual Artists (EVA); Federation of European Publishers (FEP); International Association of Scientific, Technical & Medical Publishers (STM); the International Federation of Reprographic Rights Organizations (IFRRO).

37. See Orphan Works Directive 2012/28, art. 9, 2012 O.J. (L 299) 5, 9 (EU), (establishing that the Directive be implemented by Member States by the end of October 2014).

38. Orphan Works Directive 2012/28, recital 20, 2012 O.J. (L 299) 5, 9 (EU).

39. Orphan Works Directive 2012/28, art. 4, 2012 O.J. (L 299) 5, 9 (EU). The Directive also applies to works and phonograms that have never been published or broadcast, but that have been made available

include photographs. The use of the work has to be non-commercial, except for the amount necessary to cover the cost of digitization.⁴⁰ The diligent search has to be carried out in good faith⁴¹, has to be recorded in a publicly accessible database,⁴² and has to involve consultation of at least a number of databases specified by the Directive in an Annex.⁴³ Users of the work can only be “publicly accessible libraries, educational establishments and museums, as well as [. . .] archives, film or audio heritage institutions and public-service broadcasting organisations, established in the Member States, in order to achieve aims related to their public-interest missions”.⁴⁴ This public interest mission is further specified as “preservation, restoration and the provision of cultural and educational access to works contained in their collections”.⁴⁵ The Directive mandates the establishment of a publicly accessible database for used orphan works, to be managed by the Office for the Harmonisation of the Internal Market (by then renamed EU Intellectual Property Office).⁴⁶ Finally, reappearing rightsholders have a right to put an end to the orphan work status and to receive “fair compensation” for the use that has been made of their works.⁴⁷

Within the European Union, and within the leeway left by the Directive, Member States adopted different solutions for the digitization of orphan and out-of-distribution works. While Scandinavian countries and the Czech

to the public with the consent of the rightholders (Art. 1.3). For these works, special requirements of diligent search apply (Art. 3.3, paragraph 2).

40. Orphan Works Directive 2012/28, art. 6.2, 2012 O.J. (L 299) 5, 9 (EU).

41. Orphan Works Directive 2012/28, art. 3.1, 2012 O.J. (L 299) 5, 9 (EU).

42. Orphan Works Directive 2012/28, art. 3.5, 2012 O.J. (L 299) 5, 9 (EU). The organization has to provide the following information to the competent authority: “a) the results of the diligent searches that the organizations have carried out and which have led to the conclusion that a work or a phonogram is considered an orphan work; (b) the use that the organizations make of orphan works in accordance with this Directive; (c) any change, pursuant to Article 5, of the orphan work status of works and phonograms that the organizations use; (d) the relevant contact information of the organization concerned.”

43. In its Annex the directive indicates the existing databases available to carry out a diligent search, it refers to existing collections of Orphan Works such as ARROW. The ARROW (Accessible Registries of Rights Information and Orphan Works towards Europeana) project was launched in 2007. It aims at developing tools to enable media providers to obtain information on rights holders and the way to obtain licenses on copyright works, in view of implementing the Europeana digital library. See ARROW, <http://www.arrow-net.eu> (last visited Apr. 30, 2016).

44. Orphan Works Directive 2012/28, art. 1.1, 2012 O.J. (L 299) 5, 9 (EU).

45. Orphan Works Directive 2012/28, art. 6.2, 2012 O.J. (L 299) 5, 9 (EU).

46. See Orphan Works Directive 2012/28, art. 3.5(b), 2012 O.J. (L 299) 5, 9 (EU); Regulation 386/2012, 2012 O.J. (L 129) 1 (EU). See also Eur. Union Intell. Prop. Off. Database, <https://eipo.europa.eu/orphanworks> (last visited Apr. 30, 2016). At the time of writing, the database contained around 1,700 entries.

47. Orphan Works Directive 2012/28, art. 6.5, 2012 O.J. (L 299) 5, 9 (EU).

Republic⁴⁸ maintain their Extended Collective Licensing systems, in principle not forbidden by the Directive,⁴⁹ other countries rely on private ordering, with the use of agreements between users and collecting societies,⁵⁰ or on *ad hoc* legislative solutions that comport with the implementation of the Directive.

In France for example, books that are no longer commercially exploited have been regulated by a specific law⁵¹ since September 2012 and by a State Council decree since February 2013.⁵² While the law dictates the general framework, the decree details the licensing system. The French law mandates the establishment of a database of unavailable books (ReLire⁵³) managed by the Bibliothèque Nationale de France. The licensing is handled by an appointed collecting society (SOFIA⁵⁴) which keeps the revenues for

48. Agnieszka Vetulani (European Commission, DG Information Society and Media, Unit E4 – Digital Libraries and public sector information) *The Problem of Orphan Works in the EU: An overview of legislative solutions and main actions in this field* 30-32 (2008).

49. Directive 2012/28/EU of the European Parliament and of the Council of 25 October 2012 on certain permitted uses of orphan works, 2012 O.J (L 299/5) 24: “This Directive is without prejudice to the arrangements in the Member States concerning the management of rights such as extended collective licences, legal presumptions of representation or transfer, collective management or similar arrangements or a combination of them, including for mass digitization.”..”

50. For example, in France the National Audiovisual Institute negotiated agreements with collecting societies of authors, performers and journalists, and with trade unions representing performers and audiovisual directors, in order to use audiovisual Orphan Works. Moreover, the digital library Gallica signed an agreement with the Syndicat National de l’Édition (the French Publishers, Association), the Centre National du Livre (National Book Centre) and the Ministry of Culture and Communication in order to include a relevant number of book references in its database. Also in the Netherlands orphan works are used through contractual arrangement. For example, the EYE Film Institute has negotiated the use of Orphan Works through an extended collective licence for digital exploitation of audiovisual works with the relevant collecting societies, which applies also to non-represented rightholders. An opt-out option is offered to the latter, and the agreement is valid for three years. Other agreements have been signed by the National Library, the Digital Library of Dutch Literature, and the National Archives with relevant collecting societies. *See* KEA EUROPEAN AFF., AUDIOVISUAL ORPHAN WORKS IN EUROPE – NATIONAL SURVEY (May 2011), <http://www.keanet.eu/docs/ORPHAN%20WORKS%20STUDY%20KEA.pdf>.

51. Unavailable books include orphan works and out of print books. *See* Loi n° 2012-287 du 1er mars 2012 relative à l’exploitation numérique des livres indisponibles du XXème siècle, J.O. [OFFICIAL GAZETTE OF FRANCE]. JORF n°0053,2 mars 2012, page 3986, texte n° 1.

52. The Decree of the Conseil d’Etat was issued on the 27 of February 2013. *See* Décret n° 2013-182 du 27 février 2013 portant application des articles L. 134-1 à L. 134-9 du Code de la propriété intellectuelle et relatif à l’exploitation numérique des livres indisponibles du XXe siècle, J.O. [OFFICIAL GAZETTE OF FRANCE] n°0051,1 mars 2013, page 3835, texte n° 41.

53. The database is working as of the 21st of March 2013. *See* RELIRE, <http://relire.bnf.fr> (last visited Apr. 30, 2016). In order to introduce a new work in the database it is necessary to contact the Bibliothèque Nationale de France through a form available on the ReLire website. The database is compiled from a list of literary works as approved by a scientific committee consisting of representatives of authors and editors.

54. Décret du Conseil d’Etat du 27 février 2013, *supra* Note 52. The Collecting society SOFIA has been appointed on the 21st of March 2013 by the Ministry of Culture to issue authorizations to electronically publish non-available books...

prospective reappearing authors for a period of ten years,⁵⁵ after which the works can be used freely by public libraries. Rightsholders can however opt out from the system within six months from the inscription of the work in the database.

However, this law was considered inconsistent with EU law. A Preliminary reference was lodged before the Court of Justice of the European Union to ask whether the exclusive right to authorize reproduction, as defined in article 2 of the Information Society Directive,⁵⁶ precludes legislation that gives a collecting society the right to allow the publication of out-of-print works while allowing the authors to oppose to such practice.⁵⁷ The Court responded in the affirmative. While recognizing that digitization of out-of-print books may be ‘in the cultural interest of consumers and of society as a whole’⁵⁸ the Court concluded that EU law precludes Member States from introducing national legislation that undermines the exclusive nature of the rights of reproduction and communication to the public. It remains to be seen what the full impact of this decision will be on the French law, as well as on other national laws that enable licensing schemes for orphan and out-of-print works.

In the UK, legislation on orphan works has been in force since 2014.⁵⁹ This legislation allows for licensing of every type of work for every type of use. This scheme runs in parallel with Extended Collective Licensing⁶⁰ for some types of works, and with the exception for the digitization of orphan works by public cultural institutions established by the Directive.⁶¹ The UK Orphan Works licensing scheme as well requires that prospective users carry out a good-faith diligent search of rightsholders. Works licensed under this scheme are recorded in a national orphan works register,⁶² maintained by the

55. After this period the revenues are donated to activities supporting authors and artistic creations. CODE DE LA PROPRIÉTÉ INTELLECTUELLE [C.P.I.] [INTELLECTUAL PROPERTY CODE] art., L. 134-9 (Fr.) as modified by the new law.

56. Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society (OJ 2001 L 167, p. 10).

57. Case C-301/15, Marc Soulier, Sara Doko v Premier Ministre, Ministre de la Culture et de la Communication, Judgement of the Court (Third Chamber), 16 November 2016.

58. Case C-301/15, § 45.

59. Enterprise and Regulatory Reform Act 2013, c. 24, § 77.

60. Copyright, Designs and Patents Act 1988, sec. 116A and 116B (introduced by the Enterprise and Regulatory Reform Act 2013).

61. Copyright and Rights in Performances (Certain Permitted Uses of Orphan Works) Regulations 2014, S.I. 2014/2861.

62. *Orphan Work Register*, U.K. INTELL. PROP. OFF., available at <https://www.orphanworkslicensing.service.gov.uk/view-register> (last visited Apr. 30, 2016). At the time of writing, this database contained 320 entries.

UK Intellectual Property Office, while cultural institution relying on the orphan works exception will register their works on the EUIPO database. The Intellectual Property Office sets licensing fees depending on the use (commercial and non-commercial) and the type of work, and it keeps the revenues for 8 years, for prospective reappearing authors. Diligent search requirements are met after the consultation of a number of sources determined for each work by the Intellectual Property Office.⁶³

In Germany, a new legislation on orphan works has been in force since January 2014.⁶⁴ The new law implements the Orphan Works Directive, and therefore permits a number of public institutions to digitize works in their possession for non-commercial purposes. Moreover, out-of-commerce works published before January 1966 can be object of Extended Collective Licensing, provided that they are in the collections of public cultural institutions and are used for non-commercial purposes.⁶⁵ Similarly in Germany an orphan works database is to be established and administered by the German Patent and Trademark Office (DPMA) wherein rightholders can opt out from the scheme within 6 weeks from the publication of their work in the database.

In all Member States, the diligent search requirement introduced by the Orphan Works Directive heavily relies on consultation of existing (and future) databases allowing for a centralised search, such as ARROW,⁶⁶ FORWARD⁶⁷, etc. The establishment and maintenance of such centralised database has proved extremely helpful in easing costs of diligent search.⁶⁸ However, digitization of cultural collections requires right clearance for large numbers of items, and the costs of similar undertakings are still the biggest hurdle on the road to world-sharing cultural heritage.

63. See *Orphan Works Diligent Search Guidance for Applicants*, U.K. INTELL. PROP. OFF. (Sept. 17, 2014), <https://www.gov.uk/government/publications/orphan-works-diligent-search-guidance-for-applicants>.

64. Gesetz über die Wahrnehmung von Urheberrechten und verwandten Schutzrechten (Urheberrechtswahrnehmungsgesetz) [UrhWahrnG] [Law on the Administration of Copyright and Neighboring Rights], Sept. 9, 1965, BGBl. I at 1294, as amended by Gesetz [G], Oct. 1, 2013, BGBl. I at 3728, art. 2, §13d (1).

65. *Id.*

66. *About Arrow*, ARROW, <http://www.arrow-net.eu> (last visited Apr. 30, 2016).

67. FORWARD is a EU funded project which aims at providing an Orphan Works registry on audiovisual works, Framework for a EU-Wide Audiovisual Orphan Works Registry, FORWARD, <http://project-forward.eu/> (last visited Apr. 30, 2016).

68. Findings of a study on four pilot countries using ARROW for diligent search are available on the ARROW web site: Experience from using Arrow, ARROW, <https://www.arrow-net.eu/experience-using-arrow>.

III. MAPPING THE COST OF DILIGENT SEARCH

Prior to the implementation of recent orphan works legislation, a significant amount of empirical research was gathered from memory and cultural institutions about the costs of digitizing works whose rightholders are difficult or impossible to locate. One purpose of this research was to evaluate the cultural and economic benefits of providing legislative solutions to the problem of clearing copyright in mass digitization projects.⁶⁹ An approach used in many of these studies has been to descriptively tally up the costs of rights clearance for holders of different kinds of collections containing copyright material. As a result, policy makers and researchers have been able to obtain insight into the cost of identifying and seeking permission to use a range of different types of works, including printed books, personal papers, photographs, sound recordings and audio-visual works.⁷⁰ For example, a meta-study by the University of Glasgow estimated that on average, institutions reported spending 2.58 hours and £31 GBP per work searching for rightholders in books.⁷¹ Searching for rightholders in magazine and newspaper clippings was found to take 1.52 hours at a cost of £23 per work. Multiplied across a collection consisting of 48,000 individual clippings, diligent search costs in one project were estimated to reach more than 8 years of total work for a single person, an impractical undertaking for most cultural institutions.

This previous literature draws upon a theory of transaction costs to evaluate the appropriate limits to copyright in a cultural heritage setting. Economic analyses suggest that while copyright incentivises the creation of new works, in certain circumstances its presence can increase the cost of making use of old works to the point that it reduces efficiency.⁷² The transaction costs associated with negotiating rights between would-be users

69. See Anna Vuopala, *Assessment of the Orphan works issue and Costs for Rights Clearance*, EUROPEAN COMMISSION: DG INFSO, (May 2010), http://www.ace-film.eu/wp-content/uploads/2010/09/Copyright_anna_report-1.pdf. See also *European Commission Impact Assessment on the cross-border online access to orphan works accompanying the document 'Proposal for a directive of the European Parliament and of the Council on certain permitted uses of orphan works,'* at 48 COM (2011) 289 final, SEC (2011) 616 final (24 May 2011).

70. For a summary of costs in different collections of works, see UK IPO, *COPYRIGHT WORKS: SEEKING THE LOST. CONSULTATION ON IMPLEMENTING A DOMESTIC ORPHAN WORKS LICENSING SCHEME AND THE EU DIRECTIVE ON CERTAIN PERMITTED USES OF ORPHAN WORKS* 91 (Crown 2014).

71. See VICTORIA STOBO, RONAN DEAZLEY & IAN G. ANDERSON *COPYRIGHT AND RISK: SCOPING THE WELLCOME DIGITAL LIBRARY PROJECT* 8 (CREATe Working Paper No. 10 Dec. 2013), <http://www.create.ac.uk/wp-content/uploads/2013/12/CREATe-Working-Paper-No.10.pdf>.

72. David B. Sherman, *Cost and Resource Allocation Under the Orphan Works Act of 2006: Would the Act Reduce Transaction Costs, Allocate Orphan Works Efficiently, and Serve the Goals of Copyright Law*, 12 VA. J. L. & TECH. 19, 19 (2007).

of works may produce a sub-optimal net outcome in terms of availability and use of expressions. In this context, transaction costs generally refer to the “search costs, bargaining costs, and enforcement costs of entering into a transaction”.⁷³ Empirically demonstrating that the costs associated with searching for rightsholders prevents cultural institutions from using orphan works strengthens the argument that the transaction costs imposed by regulation are too high.⁷⁴

It is clear from the evidence accumulated across a range of collections that rights clearance imposes high costs on cultural institutions, with the handling of orphan works a significant contributing factor. However, costs associated with copyright are present throughout the digitization process and not only at the moment of rightsholder search. Digitization by cultural institutions may be characterised as progressing in three phases. Table 1 outlines the steps typically undertaken by cultural institutions when digitizing collections. The steps consist first of the exhibition design phase, in which factors such as the institution’s human resource capabilities and readiness to embark on digitization efforts represent potential barriers to digitization. The second phase, consisting of the identification of works and search for rightsholders, is frequently a focus of research on the role of copyright as a barrier to mass digitization. In the third phase, after rightsholders have been located, there may be additional costs related to fees negotiated or changes to the exhibition mandated by rightsholder requests.

A 2010 study by the University of North Carolina (UNC) at Chapel Hill documenting local efforts to digitise the papers of Thomas E Watson, illustrates typical challenges and outcomes of a digitization project across each of the three phases.⁷⁵ A first source of costs may be encountered before digitization begins, at the stage of conception of the exhibition. Institutions may lack the adequate legal knowledge or resources to plan a project in the first place, and may choose to abandon the project altogether. They may not know where to obtain information about copyright or may lack access to search databases. A survey carried out in three EU Member States after the implementation of the Orphan Works Directive has shown that between 30%

73. See Sonia Baldia, *Transaction Cost Problem in International Intellectual Property Exchange and Innovation Markets*, 34 NW. J. INT’L L. & BUS., 23 (2013).

74. The assertion that transaction costs are excessive was made by the European Commission in support of an exception for use of orphan works, although the authors conceded that there remain significant costs associated with diligent search. See European Commission Impact Assessment 2011, *supra* note 69, at 24.

75. Maggie Dickson, *Due diligence, futile effort: Copyright and the digitization of the Thomas E. Watson papers*, 73 THE AMERICAN ARCHIVIST 626-636 (2010).

and 50% of the databases required to begin a diligent search were not freely available to public.⁷⁶ Project design may therefore be shaped by institutional perception of the costs of using copyright works.

In the case of UNC and similar university-led studies,⁷⁷ archivists benefited from research funding to hire paid researchers as well as the presence of University legal departments to provide advice. Although the Watson heirs had given permission for the UNC archives to digitise materials which remained in copyright, the collection included a range of other artifacts by third-party copyright owners where the author's identity or the date of death were unknown. The first direct cost encountered by researchers was generating metadata about the authors and recipients of the letters and other document in the collection (pre-digitization desk research). The next step undertaken in the clearance process was searching for individual rightsholders, which the researchers conducted on a sample of 3,304 of the identified authors. After four and half months of work and \$8,000 USD in hourly-paid staff costs, the researchers were able to positively identify 51% of the authors and determine that 18% had died prior to 1939, leaving their works in the public domain. A remaining 49% of authors could not be positively identified, making their contributions orphan works.⁷⁸

76. A survey found that less than 30% of the databases required to conduct a diligent search in the UK, Italy and the Netherlands were freely available to public users. See MARCELLA FAVALE, SIMONE SCHROFF & AURA BERTONI, REQUIREMENTS FOR DILIGENT SEARCH IN THE UNITED KINGDOM, THE NETHERLANDS, AND ITALY, REPORT 1 (EnDOW February 2016), available at <http://diligentsearch.eu/resources>.

77. DENISE TROLL COVEY, ACQUIRING COPYRIGHT PERMISSION TO DIGITIZE AND PROVIDE OPEN ACCESS TO BOOKS 58 (Digital Library Federation, Council on Library and Information Resources 2005).

78. Dickson, *supra* note 75, at 628-630.

TABLE 1: Characterising costs of rights clearance in three phases

Project design	Rightsholders search	Exhibition
Cost of exhibition development (calendar time, scheduling, space)	Labour cost of examining works (Dickson, 2010)	Fees paid to rightsholders located by DS (Covey, 2005)
Knowledge costs related to identifying and handling IP	Labour cost of searching for rightsholders / DS (Dickson, 2010)	Fees paid to license orphan works in UK scheme or ECL
PR / reputation costs arising from embarking on infringing activity	Labour cost of corresponding with rightsholders (Covey 2005; Stobo et al, 2016))	Alterations to display of work at request of rightsholders
Subscription fee to database required for DS (Favale et al. 2016)	Material cost of communicating w/ rightsholders (Covey, 2005)	Takedown of work on rightsholders reemergence (Schofield & Urban, 2015)
	Alterations to project design incurred by rightsholders requests	Compensation paid on rightsholders reemergence

In the end, the UNC library decided to rely on a risk-based strategy and proceed with digitization of all works in the collection, even when rightsholders could not be definitively identified. They therefore assumed the cost of dealing with any re-emerging rightsholders (compliance costs),

with impacts spread through the design and exhibition phases of the process (phases 1-3). For example, in order to engage in a risk-based exhibition strategy, the collection would have to be conceived in such a way at the beginning to permit later removal or alteration of individual works. Eventual emergence of a rightsholder could present a major problem for prospective digitisers, because exhibition design (software graphical interfaces) may make it difficult for institution staff to comply with requests.⁷⁹

Collectively, studies conducted in European jurisdictions prior to 2012 indicate that without clear legislative guidance, the presence of orphan works in a collection posed an intractable challenge. The Orphan Works Directive provides some legal clarity to institutions and likely reduces costs in phase 1 by providing institutions a clear path to legally using orphan works with reduced reputational and legal risk. The Directive also addresses some of the sources of transaction costs which may prevent institutions from ever undertaking digitization, notably the bargaining and enforcement costs that may arise when a rightsholder emerges (although empirical evidence suggests this is viewed more as a reputational risk than a frequent occurrence).⁸⁰ However, the Directive does not reduce transaction costs related to rightsholder search, which make up the most significant source of direct costs for institutions. In fact, it imposes additional search costs via the diligent search requirement, by obligating archivists to search on unidentifiable works even when they have ascertained orphan status through the nature of the collection or their initial audit of holdings.⁸¹ Even if cultural institutions are willing to search for individual rightsholders externally (at considerable expense across large collections), the proportion of those with identifiable rightsholders is often reported to be low. For example, a study at the British Library in 2011 reported that following a pre-digitization audit, some 31% of a collection of books remained orphaned.⁸² A study on

79. A survey of results of digitization efforts suggests that takedown requests are rare, but that libraries and archives are generally poorly equipped to deal with them. See Brianna L. Schofield & Jennifer M. Urban *Takedown and Today's Academic Digital Library*, UC BERKELEY PUBLIC LAW RESEARCH PAPER No. 2694731 5 (Nov. 2015), available at SSRN: <http://ssrn.com/abstract=2694731>.

80. See STOBO, *supra* note 71, at 40. Who found that following digitization of a collection of 1.6 million images by the Wellcome Library, only one request to take down an item was received from a re-emerging rightsholder; See SCHOFIELD, *supra* note 79, at 1. A study of libraries in the USA by Schofield & Urban found that requests by copyright owners to take down material remain rare in the digital library sector.

81. See CULTURE AND INTERNATIONAL ECONOMIC LAW 185 (Valentina Vadi & Bruno de Witte eds., 2015).

82. BARBARA STRATTON, SEEKING NEW LANDSCAPES: A RIGHTS CLEARANCE STUDY IN THE CONTEXT OF MASS DIGITISATION OF 140 BOOKS PUBLISHED BETWEEN 1870 AND 2010 10 (British Library 2011).

illustrated posters held by the National Archives in 2010 similarly reported that 13% of the collection remained orphaned after a costly search on the items themselves.⁸³ Prior to the legislative solution offered by the Directive, these works could not be made available at all unless the digitising institution was willing to assume *ex post* costs (including reputational harm or the cost of complying with a removal request). The Orphan Works Directive addresses the tractability of the problem, enabling at minimum a path to legal use of those works that remain unidentified, but it does not solve the problem of transaction costs, since significant costs identified by previous studies were concentrated in phase 2, the rightsholder search phase. In the following sections, we evaluate the potential for crowdsourcing to reduce the costs of diligent search for cultural institutions, in order to more fully benefit from the orphan works legislation.

IV. EXPLORING THE POTENTIAL OF CROWDSOURCING

In copyright systems where diligent search is a requirement to make use of orphan works, could the fruits of digitalization offer a solution to one of its ailments?⁸⁴ While the declining cost of digitally preserving cultural collections has brought ambitions practically within reach, we have seen that the costs associated with handling orphan works render some projects prohibitively expensive. However, other benefits of digital technology may offer a potential solution to this problem. While tools such as scanners and image processing software have reduced the cost of digitising cultural works, so to have digital communication networks and platforms which made it increasingly possible to draw on external help and expertise to perform labour-intensive tasks. In the following section, we draw on recent scholarship on the phenomenon of crowdsourcing to explore its practicality in assisting cultural institutions with diligent search.⁸⁵

Crowdsourcing emerged as a social practice alongside the adoption of networked digital technologies. The term came to prominence in the early 2000s through its usage in *Wired Magazine* to describe new management practices made possible with the online communication.⁸⁶ Although other

83. See VUOPALA, *supra* note 69, at 19.

84. In contrast to digitization, 'digitalisation' refers to the effects of digital technology on society more broadly, in this case its impact on the cultural heritage sector. See Anders Henten & Reza Tadayoni, *Digitalization*, in A HANDBOOK OF CULTURAL ECONOMICS 190-199 (Ruth Towse ed., 2nd ed. 2011).

85. Crowdsourcing has attracted significant academic attention in the past decade. By 2011 there were at least 175 scholarly publications on the subject. See Enrique Estellés-Arolas & Fernando González-Ladrón-de-Guevara, *Towards an integrated crowdsourcing definition*, 38 J. OF INFO. SCI., 189, 200 (2012).

86. Jeff Howe, *The Rise of Crowdsourcing*, 14 WIRED MAGAZINE, 1-4 (2006).

means of gathering intelligence from large numbers of people pre-date the internet (national elections, census-taking), lower cost and immediacy of digital networking led to wider adoption of the practice. Communication networks were not the only ingredient necessary for crowdsourcing; the tools of digital collaboration have also become more accessible and more sophisticated. Examples include tools with open Application Programming Interfaces (APIs) such as Google Maps and MediaWiki (the open source software used by Wikipedia). Another development which has facilitated crowdsourcing is the emergence of online collaboration platforms like GitHub, which allow groups of users to contribute information to collective projects.⁸⁷

Crowdsourcing is distinct from other collective uses of digital technology attracting attention from communication scholars. These neighbouring practices include commons-based peer production,⁸⁸ prosumption,⁸⁹ produsage,⁹⁰ collective intelligence⁹¹ and user-led innovation.⁹² Unlike those co-productive practices in which participants choose their own roles, crowdsourcing is typically initiated for a specific purpose which is often more clearly defined than for collectively-owned projects. In that regard, crowdsourcing reflects its origins as a portmanteau of the words “crowd” and “outsourcing”, the latter denoting a client-supplier relationship.⁹³ In some configurations, crowdsourcing resembles crowdwork, in which large numbers of people perform repetitive “microtasks”.⁹⁴ As a general concept, crowdsourcing refers to the practice of seeking inputs (either in terms of labour, information or resources) from a large number of contributors, typically from an undifferentiated public.

87. Jean-Claude Burgelman, David Osimo & Marc Bogdanowicz, *Science 2.0 (change will happen . . .)*, 15 FIRST MONDAY (2010).

88. See YOCHAI BENKLER, *THE WEALTH OF NETWORKS: HOW SOCIAL PRODUCTION TRANSFORMS MARKETS AND FREEDOM* 59 (Yale U. Press 2006).

89. Georg Ritzer & Nathan Jurgenson, *Production, Consumption, Prosumption The nature of capitalism in the age of the digital 'prosumer'*, 10 JOURNAL OF CONSUMER CULTURE, 13, 36 (2010).

90. Axel Bruns, *From Prosumption to Produsage*, ARC CENTRE OF EXCELLENCE FOR CREATIVE INDUSTRIES AND INNOVATION (2013) <http://snurb.info/files/2014/From%20Prosumption%20to%20Produsage.pdf>.

91. See Henry Jenkins, “*Collective Intelligence vs. The Wisdom of Crowds*,” CONFESSIONS OF AN ACA-FAN: THE OFFICIAL WEBLOG OF HENRY JENKINS (Nov. 27, 2006), http://henryjenkins.org/2006/11/collective_intelligence_vs_the.html.

92. ERIC VON HIPPEL, *DEMOCRATIZING INNOVATION* 4 (MIT Press 2005).

93. ESTELLÉS-AROLAS & GONZÁLEZ-LADRÓN-DE-GUEVARA, *supra* note 85.

94. HOWE, *supra* note 86, at 1 (Indeed, in his original Wired Magazine article, Howe referred to the crowd as “the new pool of cheap labor.”)

Research on crowdsourcing has identified two main sources of cost savings to firms and organisations. The first source of cost savings arises from economies of scope obtained through access to many different perspectives and solutions to a given problem. The second source of savings comes from economies of scale offered by distributing a task among many contributors. In certain cases, these two motivations may combine or overlap in the design of crowdsourcing initiatives.

A. Economies of Scope

Research and development (R&D) activity is essential to product innovation, but is expensive for firms. An additional problem facing innovators is that primary research may not produce protectable intellectual property, making exploitation of discoveries difficult.⁹⁵ Crowdsourcing can reduce the cost of innovation by enabling firms to access a range of innovative ideas from contributors outside of their traditional boundaries. The IP protectability problem remains, but may be offset by the significant cost reductions in hiring, training and equipping staff. There may be additional qualitative benefits: the range of “scope” on a problem offered by a crowd is likely to exceed even the most well-funded R&D department. A practical example of successful pursuit of economies of scope in crowdsourcing is offered by InnoCentive, a crowdsourcing firm that was featured in the original articulation of the concept by Wired Magazine.⁹⁶ InnoCentive is a Massachusetts-based company that mediates between firms posing research challenges and experts from the public offering potential solutions. Fields of expertise include engineering, computer science, math, chemistry, life sciences, physical sciences and business. The firm operates as an intermediary between client organisations who post “challenges” and “problem solvers” who register their interest and expertise with the platform. Solvers are motivated by rewards, which range from tens of thousands to hundreds of thousands of US dollars.⁹⁷

95. See David Teece, *Profiting from Technological Innovation Implications for Integration, Collaboration, Licensing and Public Policy*, 15 RESEARCH POLICY 285, 300-01 (1986).

96. See HOWE, *supra* note 86, at 7.

97. One example of a recent challenge was for an ‘information sharing solution backed by market-based incentives whereby private industry can [. . .]disrupt foreign based individuals, corporations, terrorist groups [. . .] from acquiring US technologies to build nuclear weapons of mass destruction.’ The call, sponsored by The Stimson Centre, an American think tank, was accompanied by an award of \$10,000 USD. See INNOCENTIVE, *Innovative Approaches to Proliferation Prevention: An Industry-Government Information Sharing Platform*, INNOCENTIVE (Apr. 6, 2016), <https://www.innocentive.com/ar/challenge/9933709>.

While InnoCentive is a successful example of crowdsourcing for commercial purposes, not all crowdsourcing initiatives are paid, nor are they exclusively conducted by private firms. Research institutions and universities have sought to benefit from crowdsourcing by involving members of the public in data gathering and data analysis activities. For example, the Planet Hunters initiative, hosted on the Zooniverse.org platform invites “citizen scientists” to flag up patterns observed in data from the Kepler telescope. Contributors must be trained in order to input useful data, requiring human judgment about trends or patterns that may not be detectable by software processing. This project specifically aims to benefit from efficiencies of scope provided by users who exercise judgment and creative thinking in analysing patterns. An organiser of Zooniverse calls it “distributed thinking, not distributed computing”.⁹⁸ The concept of building upon collective intelligence to perform legally binding searches of information has been successfully applied in patent law.⁹⁹ Crowd-sourced systems of prior art searches have been used by patent offices, including the US Patents and Trademark Office and the IP Australian Office. The UK Intellectual Property Office has also run a pilot “peer to patent” experiment aimed at sourcing prior art investigation through observations on patent applications by the research and technology communities through the Internet. The information collected from the public helps patent examiners determine if a patent application is new and inventive. This pilot experience was judged useful and time-saving in the context of UK patent examination.¹⁰⁰ In a similar vein, NGOs that oppose patenting in certain fields, such as biotechnology or information technology, have relied on similar systems to search prior art capable of destroying the novelty of patent applications.¹⁰¹

B. Economies of Scale

Crowdsourcing offers the opportunity to draw on large numbers of people to carry out work beyond the capacity of the local organisation. In order for crowdsourcing to deliver economies of scale of this type, significant planning and forethought is required. Tasks which can be broken

98. See Chris Lintott, ‘Citizen’ Science and ‘Real’ Science, ZOOIVERSE BLOG (Dec. 29, 2010), <http://blog.zooniverse.org/2010/12/29/citizen-science-and-real-science>.

99. Beth S. Noveck, *Peer to Patent: Collective Intelligence, Open Review, and Patent Reform*, 20 HARV. J. LAW TECHNOL. 123, 128 (2006).

100. INTELL. PROP. OFF., PEER TO PATENT PILOT, 17 (2012), <http://webarchive.nationalarchives.gov.uk/20140603093549/http://www.ipo.gov.uk/p2p-report.pdf>.

101. Noveck, *supra* note 99.

up into many smaller, divisible activities are ideal for crowdsourcing. Other considerations include the time-sensitivity of the work (asynchronous or simultaneous) as well as the interchangeability of members of the contributing public (so that tasks not completed by one member can be picked up by another). In order to be “scalable”, the organisation of a crowdsourcing network must be such that the effort of parsing and distributing tasks to workers does not exceed the benefit of doing so. In distributed computing initiatives, this has been called the “computability to content ratio”.¹⁰²

One familiar example of a crowdsourcing initiative which delivers economies of scale is the reCAPTCHA internet plugin. Developed by computer scientist Luis von Ahn with a team from Carnegie Mellon University, the software distinguishes between humans and automated bots on the web by asking visitors to decipher words and images. By deciphering these images, users unwittingly assist in digitising books, linking addresses to photographs in Google Street View and training image recognition algorithms.¹⁰³ Other business models have incorporated large-scale crowdsourcing. Amazon offers a paid system called Mechanical Turk, which mediates between customers and human crowdworkers (called “turkers”) who perform large quantities of “human intelligence tasks” in exchange for small payments. In large-scale crowdsourcing efforts such as these, contributors tend to be anonymous to each other; systems are designed to present minimal barriers to participation, and value to the initiator increases with the size of the network of willing contributors. In the field of cultural heritage, a resourceful initiative by the British Library, the Mechanical Curator, enables quick crowd-sourced classification of images, which in turn are offered for free reutilization to users.¹⁰⁴ Another similar example is the Artstor Arcades project, a crowdsourcing platform designed to catalogue D James Dee photographs.¹⁰⁵

102. David P. Anderson, Jeff Cobb, Eric Korpela, Matt Lebofsky, & Dan Werthimer, *SETI@ home: an experiment in public-resource computing*, 45 COMMUNICATIONS OF THE ACM 56-61 (2002).

103. *Introducing the New reCAPTCHA*, GOOGLE, <https://www.google.com/recaptcha/intro/index.html#creation-of-value> (last visited Apr. 30, 2016).

104. THE BRITISH LIBRARY, *The British Library*. FLICKR, <https://www.flickr.com/photos/britishlibrary> (last visited Apr. 30, 2016).

105. Artstor, *Artstor Arcades: Introducing our New Crowdsourcing Software*, THE ARTSTOR BLOG (Aug. 31, 2015), <https://artstor.wordpress.com/2015/08/31/introducing-artstors-new-crowdsourcing-software>.

V. APPLYING CROWDSOURCING TO SOLVE THE DILIGENT SEARCH PROBLEM

By offering efficiencies of scale and scope, crowdsourcing would appear to present a workable solution to the problem of the costs of diligent search. Large institutional collections regularly reach into the tens of millions of individual works, posing a considerable challenge for rights clearance. Spreading the task of diligently searching for rightsholders across a crowd of contributors would reduce costs for institutions which do not possess adequate resources in terms of staff or funding. The process of diligently searching for rightsholders could also benefit from efficiencies of scope. Handling an unknown work requires knowledge and expertise about its age, provenance and context. Niche communities of enthusiasts may possess knowledge about the provenance or likely ownership of works unavailable to the host institution.

Successfully initiating crowdsourcing is not trivial and presents a number of challenges. One challenge relates to the culture and professional organisation of the initiating institution. In science, the rise of crowdsourcing has offered an opportunity to flatten or democratise access to knowledge production, but at the risk of undermining existing structures governing professional status, allocation of funding and credit.¹⁰⁶ Some of the public critiques of the Zooniverse.org platform relate to the hierarchical structure of the crowdsourcing endeavour and the relative low status of user-contributors.¹⁰⁷ One journalist posed the question, “Isn’t citizen science simply scientists making the public do the boring parts of science for them?”¹⁰⁸ These questions have implications for the structure and organisation of crowdsourcing projects in the digital heritage sector: What is the status of “citizen archivists” or “citizen legal experts” with respect to host institutions? What would the enrolment of patrons in crowdsourcing mean for future public funding in an already resource-constrained sector? Of more immediate and pragmatic concern, what is the level of institutional readiness to embark on planning and constructing crowdsourcing infrastructure, given the knowledge requirements and other maintenance costs?

Should a cultural heritage institution decide to embark on crowdsourcing its diligent search efforts, a second challenge is to locate and

106. See Burgelman, Osimo & Bogdanowicz, *supra* note 87.

107. See Alex Wright, *Managing Scientific Inquiry in a Laboratory the Size of the Web*, N. Y. TIMES, Dec. 27, 2010.

108. Emma, *Zooniverse*. IMPERIAL COLLEGE LONDON FACULTY OF NATURAL SCIENCES STUDENT BLOGS (Jan. 26, 2014), <http://wwwf.imperial.ac.uk/blog/studentblogs/emma13/2014/01/26/zooniverse>.

motivate a significant “crowd” of contributors. One motivation for users to engage in crowdsourcing is that contributors can benefit from “selective incentives” which arise from the process of contribution and are only available to those who contribute actively.¹⁰⁹ These benefits can include knowledge and skill building as well as reputational gains. In commercial contexts, contributors can benefit if their ideas are adopted by the project initiator and it leads to feature changes that benefit the contributor.¹¹⁰ Research has found that this is an important motivator for participation in open source software development.¹¹¹ Therefore, as with large-scale crowdsourcing efforts, it is not always necessary to offer a monetary reward for participation – participants may find private incentives to contribute to public goods.

Such research on incentives in private-collective innovation suggests a number of potential reasons why contributors might engage in diligent search. Patrons of a cultural heritage institution may wish to altruistically support the institution for the public good; curious members of the public or experts may seek to gain knowledge or improve their skills; users or enthusiasts of a particular cultural good may be motivated by their own private interest in improving it (e.g. to attract more interest); finally, users may be motivated purely by private interest in an individual orphan work (e.g. to access or use a digital version of a work). Recognising and enabling such a range of motivations to contribute is likely important in successful crowdsourcing design.

In summary, an analysis of the suitability of crowdsourcing for cultural institutions must take account of the expected benefits in terms of efficiencies of scope and scale, weighted against the costs of setting up an initiative. Empirical research on rights clearance explored in section III suggests that search costs for institutions are currently high, so crowdsourcing offers potential in terms of cost savings. The nature of the task – diligently searching for rightsholders in individual works – appears to be suitable for crowdsourcing, being easily divisible and scalable. Challenges relate to locating and motivating a suitable community of willing contributors, while minimising the cost of managing distributed contributions.

109. Eric von Hippel & Georg von Krogh, *Open source software and the 'private-collective' innovation model: Issues for organization science*, 14 *ORG. SCI.* 209, 215 (2003).

110. Dietmar Harhoff, Joachim Henkel & Eric Von Hippel, *Profiting from voluntary information spillovers: how users benefit by freely revealing their innovations*, 32 *RESEARCH POLICY* 1753, 1767 (2003).

111. See von Hippel & von Krogh, *supra* note 109.

A. Crowdsourcing Diligent Search for Orphan Works: The Institutional Perspective

The potential of crowdsourcing could be harnessed to help ease the costs of diligent search for cultural institutions. While performing centralized diligent searches on cultural items is legally complicated and expensive, a de-centralised mechanism that allows end users to perform independent diligent searches and provide cultural institutions with the results of these searches could offer a solution. To enable such integration between institutions and users, we propose a dedicated platform with specific characteristics intended to facilitate uptake, lower costs to institutions and provide legal clarity while being accessible to non-lawyer contributors.¹¹²

First, in order to be scalable a crowdsourcing solution must take into account the legal requirements for diligent search in every jurisdiction where it was to be used. The European Orphan Works Directive makes it clear that cultural institutions (the users of orphan works) themselves are legally responsible for ensuring that diligent searches are carried out in compliance with the law. The Directive leaves the matter of defining the legal requirement of diligent searches to individual Member States. As discussed in Section II, some member states have adopted closed lists of databases and sources which must be consulted for individual classes of works, while others have adopted more open ended guidelines. In either case, crowd contributors would need to be guided by the platform through the diligent search process in such a way to produce in a legally valid search, according to the national requirements. The multi-jurisdictional aspect of the problem makes the design of such a platform more complex. For instance, a diligent search across different EU Member States would have to go through the consultation of the specific sources and databases that are determined under national implementation of the Orphan Works Directive. In order to be practicable for cultural institutions, the guided diligent searches performed by end-users would require review and approval by the cultural institution holding the item, ensuring that the requirements of national legislation and the best practices of the specific sector are met, but also increasing costs relative to a purely automated system. By automating laborious parts of the processes, and automatically adapting search instances to relevant legislation, the crowdsourcing system could still be expected to reduce costs.

112. The proposal is currently under development in the EU-funded project EnDOW (“Enhancing access to 20th Century cultural heritage through Distributed Orphan Works clearance”) *Diligent Search: Welcome*, <http://diligentsearch.eu>.

A legally compliant and rigorously designed crowdsourcing solution could reduce risks to cultural institutions and help overcome perceived challenges associated with mass digitization. One advantage of a crowdsourcing platform is that it could enable small and medium-sized cultural institutions, with little or no experience in rights clearance, to perform independent diligent searches on their own archives, and eventually take informed decisions about the use of works contained in their collections and their inclusion in digitization programmes. Small and medium-sized cultural institutions have been hitherto left out of large-scale rights-clearance projects. Making the platform accessible to smaller institutions would require care in the interface design of the platform, so that legal requirements were accurately translated into steps that could be easily understood by laypeople as well as professional archivists. A second anticipated benefit of a crowdsourcing solution in large-scale digitisation efforts is that it could increase the likelihood of locating rightsholders and improve the accuracy and quality of diligent searches when compared to efforts by single institutions. This is due to particularities of crowdsourcing discussed in Section IV, in particular the possibility for error-correction by cross-checking one user's work against another's and encouraging accuracy by enabling a transparent record of search activity on a particular item.¹¹³

The principal benefit of the proposed platform for institutions would be to facilitate rights clearance for digitization of library and archive material by distributing a substantial part of the labour costs among end-users. As discussed in Section III, the digitization process comprises many separate but interconnected activities in addition to legally required diligent search, some of which may be suitable to crowdsourcing. For example, a user-led diligent search may result in identification of a rightsholder, so mechanisms will need to ensure that contact with rightsholders and permission requests are managed by the system or referred to a responsible agent of the cultural institution for progression. The platform should be adaptable to a diverse range of institutional contexts, exhibition types and rights clearance strategies. The crowdsourcing platform may also be designed to be responsive to a diversity of non-traditional institutions and mediums where orphan works are increasingly prevalent (such as software and born-digital cultural works).

113. For a detailed discussion of the dynamics by which quality is improved in large-scale collaborative projects, see Eric S. Raymond, *The Cathedral and the Bazaar*, 3 FIRST MONDAY 3 (1998) (These are summarized by the author in the proposition that 'given enough eyeballs, all bugs are shallow').

As discussed in Section IV, crowdsourcing can help gather information that would otherwise be difficult or impossible to collect by a lone organisation. In particular, in the field of cultural artifacts, information necessary to right clearance can be complex and dispersed in various channels. The type of required information for a compliant diligent search includes questions such as: who is the author? When did the author die? To whom have the rights being transferred? The relevance of this information to ascertain the legal status (e. g. ‘in-copyright’, ‘out-of-copyright’, ‘orphan’) varies from jurisdiction to jurisdiction and for categories of works, depending on the copyright duration and regulation governing orphan and out-of-print works. A diligent search normally could not be carried out without specialised legal guidance.

It is suggested that a semi-automated crowdsourcing platform could provide such legal guidance, by collecting, classifying and making available, in clear and user-friendly terms, all diligent search requirement for each jurisdiction where it is to be used. The platform would guide lay-contributors and non-lawyer professional archivists through legally appropriate questions and channels to search for the identity of the author or owner of a work. In the absence of the owner’s identification and location, the work can be declared “orphan”, and it could then be used according to the terms of the particular national legislation. When possible, the platform would search online databases and automate as much as possible the search process to reduce labour costs.

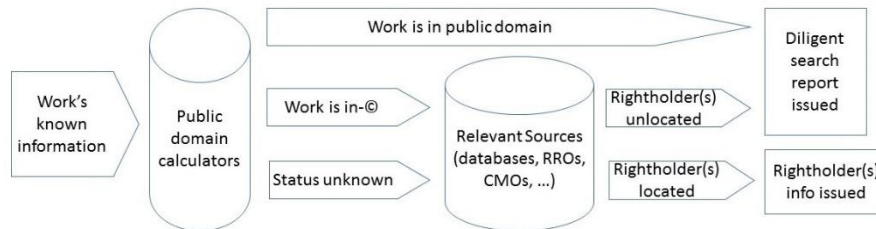
Such a platform would be novel in that it would apply the concept of crowdsourcing to a hitherto unexplored problem (rightsholder search). However, we propose this solution founded on an analysis of the suitability of crowdsourcing to the nature of the task, and the appetite for public participation to be mobilised in the context of the cultural heritage sector. Mass digitization and access to 20th Century cultural heritage will benefit the public and will foster unprecedented circulation and creation of cultural products. Altruistic contribution to a collective good is one possible incentive for the public to collaborate and contribute their labour. As discussed in Section IV, there are also private incentives related to learning, accessing individual works and increasing community status. In fact, collaborative user-generated platforms have already proven successful models for culture and information sharing as demonstrated by initiatives such as Wikimedia Commons. On the other hand, cultural institutions have a strong incentive to participate in a platform that reduces costs of clearing rights, enhances public participation, and eventually empowers sustainable

use and management of the cultural heritage items contained in their collections.

B. An Online Platform to Enable Diligent Searches

The Orphan Works “Clearance” Platform will be an interactive online resource that will enable users to perform accurate and legally valid diligent searches on items contained in the collections of cultural institutions. The platform would be specifically targeted to two typologies of users: 1) members of the public voluntarily contributing with partnering cultural institutions, and 2) small and medium size cultural institutions willing to digitize items contained in their own collections. The platform will act as a one-stop website for diligent searches under the jurisdictions of multiple countries, and will apply to all categories of works eligible under EU and national legislation (literary and artistic works, cinematographic and audiovisual works). Below is a summary flowchart of the functioning of the proposed clearance platform:

FIGURE 2: Workflow of crowdsourcing system for diligent search



At the first step of the process, users will be instructed on how to determine the copyright status of a work by submitting information to existing public domain calculators.¹¹⁴ If the work is in the public domain, the search terminates.¹¹⁵ If the work is still in copyright, or the status could not be determined by means of public domain calculators, the user will be guided towards the identification of the rightholders by means of semi-automated

114. For searches carried out in European countries, the main instrument is the Europeana Connect Public Domain Calculator. *Determining the Copyright Status of Works*, www.outofcopyright.eu.

115. In the case of a complex work, potentially carrying several rights, users will be instructed on clearing *all* possible copyrights in a work (e.g. in case of musical work: lyrics, music, sound recording).

search through the relevant sources established by national laws and regulations.¹¹⁶ These should normally include 1) publicly accessible databases (e.g. library and archive catalogues, WATCH and FOB¹¹⁷), 2) (where available) restricted databases (e.g. ARROW¹¹⁸), 3) direct inquiry to national reproduction rights organisations, collecting societies and authors/publishers associations. The possibility of implementing automated requests by directly connecting to the relevant organizations would be explored in collaboration with cultural institutions and databases.

The process guided by the platform will then lead to three possible results:

1. the work is in the public domain;
2. the work is in-copyright but the rightholders cannot be located; or
3. the work is in-copyright and the rightholders are traceable.

In the first two cases, a Diligent Search Report will be issued and passed on to the cultural institution that possesses the work; the cultural institution would have the final responsibility to validate the search and to determine the status of “orphan”. In the third case, information about the rightholder(s) will be issued to the user who has performed the search and to the cultural institution that possesses the work. The validated searches will be available on the platform, which will act as a repository of rights-cleared material.

This platform, in short, will draw upon diffuse and dispersed information about cultural artefacts with the aim of collecting it into a reliable and legally valid source through semi-automated diligent searches on works in order to determine their copyright status.

CONCLUSION

The problem posed by orphan works is growing and has been identified and addressed by numerous jurisdictions in the wake of expanding mass digitization. Worldwide, orphan works solutions range between ‘pay first and use after’ and ‘use first and pay after’ approaches. More specifically, they include: a) limited liability (use first and pay after -if the owners show up, as in the US); b) Extended Collective Licensing (mass licensing extended to non-represented owners); c) central licensing (central governmental body, or collecting society, which issues the licenses); and d) a mix of the above.

116. In the UK, guidelines for diligent search have been published. UK Intellectual Property Office, *Orphan Works diligent search guidance for applicants*, GOV.UK (Sep. 17, 2014), <https://www.gov.uk/government/publications/orphan-works-diligent-search-guidance-for-applicants>.

117. *Writers Artists and Their Copyright Holders*, HARRY RANSOM CENTER, <http://norman.hrc.utexas.edu/watch> (last visited Apr. 30, 2016).

118. ARROW, <http://www.arrow-net.eu> (last visited Apr. 30, 2016).

After lengthy discussions, Europe chose an intermediate stance for public cultural institutions: ‘search first and use after’.

Diligently searching for the owner of a cultural product may be practical and relatively easy if carried out for a single work, but it becomes prohibitively costly when applied to mass digitization. Public cultural institutions, especially medium and small ones, are unlikely to possess the necessary resources to carry out large numbers of diligent searches in due form, as required by law. This article submits that a solution for mass digitization of cultural works could be found in mobilising the collective expertise and volunteer labour of users of digital cultural heritage. Crowdsourcing has demonstrated promise in other contexts as a way for institutions to solve otherwise unsolvable problems. By providing legal guidance and structure, a centralised platform, such as the one suggested, could effectively be employed to address the difficult issue of diligent search for mass digitization of orphan works.