

# The ArchAIDE Archive: the open-data policy and management of material covered by copyright

Francesca Anichini, Gabriele Gattiglia

Francesca Anichini, University of Pisa, Department of Civilisation and Forms of Knowledge;  
[francesca.anichini@unipi.it](mailto:francesca.anichini@unipi.it)

Gabriele Gattiglia, University of Pisa, Department of Civilisation and Forms of Knowledge;  
[gabriele.gattiglia@unipi.it](mailto:gabriele.gattiglia@unipi.it)

## Abstract

*This paper is focused on one of the less well known aspects of the ArchAIDE project: the open-data policy and management of material covered by copyright. To achieve the correct management of the material that falls under copyright or database protection, the EU directives on Copyright (2001/29/EC) and Database protection (96/9/EC) were analysed. Furthermore, participating in the H2020 open-data pilot, ArchAIDE was committed to creating sustainable outputs where the project held the copyright. This included making the interoperable, multilingual vocabularies, video corpus, 2D and 3D models, and the source code and neural network models created by the project available.*

**Keywords:** IPR MANAGEMENT; COPYRIGHT; OPEN-DATA.

## Introduction

The ArchAIDE project (2016–2019) was a European funded project developed by a consortium spread within five countries (Germany, Israel, Italy, Spain and UK). The project built a system for the automatic recognition of pottery with an innovative app for tablets and smartphones (Gualandi, Gattiglia, and Anichini 2021). This goal has been implemented by developing two distinct neural networks for appearance-based and shape-based recognition, which lay in the creation of a digital comparative collection, incorporating existing digital collections, digitised paper catalogues, and multiple photography campaigns. For this reason, the EU directives on copyright (2001/29/EC) and Database protection (96/9/EC) were analysed to achieve the correct management of the material that falls under copyright or database protection.

## IPR Management

In the European Union, the copyright is governed by the Directive 2001/29/EC, whereas Directive 96/9/EC governs the Database protection.

The Directive 2001/29/EC of 22 May 2001 on harmonising certain aspects of copyright and related rights in the information society, also known as the InfoSoc Directive (Information Society Directive), implements the WIPO Copyright Treaty and harmonises the aspects of copyright law across Europe. Copyright is the legal right granted to an author, composer, playwright, publisher or distributor to exclusive publication, production, sale, or distribution of creative works with enough originality (individual character) to warrant such a right in the

literary, scientific, and artistic domains.<sup>1</sup> Copyright is made of two components: exploitation rights and moral rights.

Exploitation rights are those that can be transferred and licensed and represent the economic value of the copyright. Moral rights are non-transferrable and, in some European legislations, perpetual. They protect the creator and guarantee that attribution must be given. In Europe, this right lasts for 70 years after the death of the latest creator of a published work. In some countries, when the author is not a natural person (e.g. an institution) or the author is anonymous/pseudonymous, this right lasts for 70 years after the first publication. The rights are not absolute but framed by limitations and exceptions to copyright law, including scientific research use. A major limitation on copyright is that it only protects the original expression of ideas and not the underlying ideas themselves (Wright *et al.* 2016: 4).

The Directive 96/9/EC of 11 March 1996 on the legal protection of databases (also known as Database Directive) considers both copyright and *sui generis* right to protect the databases. Article 3.1 provides that ‘databases which, by reason of the selection or arrangement of their contents, constitute the author’s own intellectual creation shall be protected as such by copyright’. The Directive creates *sui generis* protection against unauthorised use or extraction of the facts in the database, although again, not to the facts themselves. Article 7.1 stipulates that ‘Member States shall provide for a right for the maker of a database which shows that there has been qualitatively and/or quantitatively a substantial investment in either the obtaining, verification or presentation of the contents to prevent extraction and/or re-utilisation of the whole or of a substantial part, evaluated qualitatively and/or quantitatively, of the contents of that database.’ *Sui generis* right lasts for 15 years, but each time a database is substantially modified, a new set of rights are created for that database. A *sui generis* database right is considered a property right, comparable to but distinct from copyright, that exists to recognise the investment made in compiling a database, even when this does not involve the creative aspect that is reflected by copyright. *Sui generis* right is created automatically and does not have to be registered to have an effect. The *sui generis* right can be divided into two rights: the extraction right and the re-utilisation right. Owners have the right to object to copying substantial parts of their database, even if data is extracted and reconstructed bit by bit.

The implementation of the ArchAIDE project can be described as scientific research where (i) part of the content is extracted from outside sources (see below), which falls under copyright or database protection; (ii) the content is, when necessary, transformed to fit operational needs; (iii) the content is loaded into a data set or repository; (iv) researchers gain access to the data and analysis tools are applied to the data set; (v) new knowledge is created. As for the outside sources, they can be mainly outlined as: (i) books and journal papers which contain the paper catalogue and are under copyright protection; (ii) database such as the ADS Roman Amphorae digital resource;<sup>2</sup> Ceramalex Hellenistic and Roman Pottery in Alexandria digital resource that is covered by *sui generis* right, whose uses and limitations were stipulated in the ArchAIDE Consortium Agreement.

<sup>1</sup> [https://intellectual-property-helpdesk.ec.europa.eu/regional-helpdesks/european-ip-helpdesk/europe-glossary/glossary-c\\_en](https://intellectual-property-helpdesk.ec.europa.eu/regional-helpdesks/european-ip-helpdesk/europe-glossary/glossary-c_en) (accessed 10/06/2021).

<sup>2</sup> [https://archaeologydataservice.ac.uk/archives/view/amphora\\_ahrb\\_2005/](https://archaeologydataservice.ac.uk/archives/view/amphora_ahrb_2005/) (accessed 23/11/2020).

Both Directives provided specific exceptions for scientific research in Article 5.3.a) of the InfoSoc Directive, Article 6.2.b) and Article 9.b) of the Database Directive.

### ***InfoSoc Directive***

Article 5.3 a) of the InfoSoc Directive provides for an exception to the right of reproduction (Article 2) and the right of communication to the public (Article 3) when the protected work is used ‘for the sole purpose of illustration for teaching or scientific research, as long as the source, including the author’s name, is indicated, unless this turns out to be impossible and to the extent justified by the non-commercial purpose to be achieved’.

As it stands, the text of Article 5.3 a) allows acts of reproduction, communication to the public and distribution consisting in the use of protected works for scientific research. In the Articles or the Recitals of the InfoSoc Directive, there are no indications that the use should be limited to extracts of the protected works (Triaille *et al.* 2014: 59). The Netherlands, Greece and Slovakia did not implement an exception for scientific research in their national legal order. In the absence of a legal provision recognising an exception for scientific research, the researchers were presumably infringing the copyright of the authors of the works used for the scientific research. The United Kingdom law does not use the expression scientific research, considering the insertion of the latter word as redundant. Reproduction is included in all the national provisions, to a greater or lesser extent. Most of the time, this act of reproduction permits digital reproduction, explicitly or not. Communication is referred to in Belgium, France, Luxemburg, and (along with publication) in Italy, making available to the public and transmission in Germany, borrowing in Hungary, summary and quotation, and publication on the Internet in Italy. Acknowledgement (i.e. the indication of the source including the author’s name) is required in Belgium, France, Germany, Luxemburg, Hungary, Italy, Denmark, and the United Kingdom (sufficient acknowledgement). Most of those countries provide for this indication ‘except when this is impossible’ (Triaille *et al.* 2014: 57–58).

Finally, the non-commercial nature of the purpose pursued is implemented in Belgium, Denmark, France, Germany, Hungary, Italy, and the United Kingdom. The non-commercial criterion can be understood as follows: ‘Non-Commercial means not primarily intended for or directed towards commercial advantage or monetary compensation’ as used in the CC licenses (‘Non-Commercial Interpretation – Creative Commons’, n.d.).

Beneficiaries of the research exception are not defined because of the adoption of a functional approach based on the activity of doing research. Hungary, Denmark, Italy, Luxemburg, and the United Kingdom have not included in their domestic law a specific beneficiary. In Germany, the exception authorises the reproduction of protected works ‘for one’s own scientific use’ and the making available ‘exclusively for a specifically limited circle of persons for their personal scientific research to be made available to the public’.

To summarise, the use of works (data, texts, images, etc.) for ArchAIDE purpose will not infringe authors’ exclusive rights if (cumulative conditions):

- The works are used for the sole purpose of scientific research.
- The source, including the author's name, is indicated unless this turns out to be impossible.
- Works are used to the extent justified by the non-commercial purpose to be achieved.
- The use of the works does not conflict with a normal exploitation of the work or other subject/matter and does not unreasonably prejudice the right holder's legitimate interests.

Secondly, Article 5.2 c) of the Infosoc Directive provides for an exception to the reproduction right (Article 2) for certain non-profit making establishments, such as publicly accessible libraries and equivalent institutions, as well as archives: 'c) In respect of specific acts of reproduction made by publicly accessible libraries, educational establishments or museums, or by archives, which are not for direct or indirect economic or commercial advantage.'

### **Database Directive**

As for our purpose, we take into consideration two different articles of the Databases Directive: 6.2.b) related to databases protected by copyright and 9a) for databases protected by sui generis right.

Article 6.2.b) contains the exception to copyright for scientific research, which applies not to the works contained in a database but to the structure of the database itself: 'Member States shall have the option of providing for limitations on the rights set out in Article 5 in the following cases: [...] (b) where there is use for the sole purpose of illustration for teaching or scientific research, as long as the source is indicated and to the extent justified by the non-commercial purpose to be achieved.'

This exception has been implemented in Belgium, Spain, the United Kingdom, and Italy. The requirements of indication of the source and 'non-commercial purpose to be achieved' have been taken up in all these Member States, and the exception always applies to the database as a whole (and not just a part of it). Some Member States have sometimes added conditions for the exercise of this exception: in Italy, the user can 'access and visualise' the database without the author's consent, but 'permanent reproduction [...] shall always be subject to the right holder's authorisation'. The Netherlands, Germany, Poland, Luxembourg, Denmark, and Hungary have not implemented the exception for scientific research to the copyright protection of databases; moreover, they make no explicit reference to databases in their copyright legislations (Triaille *et al.* 2014: 69).

In other words, in the event that process involves an act of copying of the structure of the database, this will not infringe the author's rights if the user can prove that:

- The database is used for the sole purpose of scientific research.
- The sources are indicated.
- The database is used to the extent justified by the non-commercial purpose to be achieved.

It is not explicitly stated in the Database Directive that the exception only applies to published databases (Triaille *et al.* 2014: 71). Article 9b) contains the exception for scientific research to the *sui generis* right contained in the Database Directive: ‘Member States may stipulate that lawful users of a database which is made available to the public in whatever manner may, without the authorisation of its maker, extract or re-utilise a substantial part of its contents: [...] (b) in the case of extraction for the purposes of illustration for teaching or scientific research, as long as the source is indicated and to the extent justified by the non-commercial purpose to be achieved.’

If the database is not freely accessible, users will only be lawful if they can avail themselves of authorisation as granted through the Consortium Agreement. Under Article 9b) of the Database Directive, scientific research does not have to be the sole purpose behind the use of the database. Article 9 provides for an exception to the right of extraction and not to the right of re-utilisation.

The exception to the *sui generis* right for scientific research has been implemented in: Belgium, Spain, the United Kingdom, the Netherlands, France, Germany, Poland, Luxembourg, and Hungary. The requirements of ‘indication of the source’ and ‘non-commercial purpose to be achieved’ have been taken up in all these member states and that the exception always applies to a ‘substantial part of the database’. As for the copyright, the United Kingdom does not specify that the research must be scientific. Finally, the exception has not been implemented in Italy and Denmark (Triaille *et al.* 2014: 79–81).

The extraction of data, for the purpose of ArchAIDE, will not infringe the database maker’s rights when the user is the lawful user of the database and can prove that (cumulative conditions):

- Data are extracted for the purpose of illustration for teaching or scientific research.
- The source is indicated.
- Data are extracted to the extent justified by the non-commercial purpose to be achieved.

Article 9b) of the Database Directive does not include the adjective ‘sole’ purpose of scientific research. As a result, uses done for non-commercial purposes other than scientific research the exception would still cover, i.e. statistical analysis, etc.

## Discussion

Analysing the scientific research exceptions in the InfoSoc Directive and the Database Directive, we may conclude that:

As regards the area of copyright: published works, mentioning the source and the authors’ name of the works, can be used to the extent justified by a non-commercial purpose; the use of the structure of published databases can only be used, mentioning the source, to the extent justified by a non-commercial purpose.

As regards the *sui generis* right: databases can be used, even if scientific research is not its sole purpose, mentioning the source and the authors' name of the works, to the extent justified by a non-commercial purpose.

Participating in the H2020 open data pilot, ArchAIDE was committed to creating sustainable outputs where the project held the copyright. Unfortunately, not all the collected data could be disseminated as open data. The research exceptions allowed by the EU Directives do not mean the ArchAIDE project automatically holds the copyright to the newly digitised or remixed data. Negotiation with copyright holders will be necessary for making these data available outside the project. ArchAIDE is able to demonstrate that paper catalogues, once digitised, can be actively re-used, also many years later from the first publication. This opens the possibility of reaching an agreement with publishers and other data providers for making their resources available in new ways, 'with a tangible benefit (seeing their data in use within the app), thus furthering the long-term discourse around making research data open and accessible' (Anichini *et al.* 2020). Instead, data owned by the project, i.e. multilingual vocabularies, videos created by the project, as well as the 2D and 3D models created from the ADS Roman Amphorae digital resource, were made available as open-data for download.<sup>3</sup> The ArchAIDE archive contains 2D vector drawings in SVG format and interactive 3D models navigable through a 3DHOP 3D viewer that can also be downloaded for 3D printing. These models exemplify an excellent standard of best-practice re-use. When the Roman Amphorae digital resource was deposited in 2005, creating automated 2D and 3D models for training a neural network could not have been a use envisioned. As 2D and 3D models were produced for each type included in the digital resource, it was possible to link the two archives, amplifying their mutual usefulness. The multilingual vocabularies were published from the UoY SPARQL endpoint and are also freely available for download and re-use in other Linked Open-data projects focussed on archaeological pottery. Finally, the ArchAIDE archive contains the video corpus created to both documents and promote the project, including the 30-minute ArchAIDE Documentary, created with footage shot over the course of the project. The ArchAIDE video archive represents a unique record of the project and an unusual record of the experience of implementing a European Commission-funded project with partners working across several countries.

It is also hoped that the thousands of photos taken by the project for training the algorithms and currently not available might result in new comparative collections that could be deposited as open research data into the ArchAIDE archive. Still, in many European countries, copyright on cultural heritage is very restrictive and does not allow us to make available the images of potsherds taken by ArchAIDE partners in national and regional collections. Showing the usefulness of these data within the ArchAIDE application might help convince Cultural Heritage national institutions to move towards more open-data policies. Finally, the source code and neural network models are publicly available as open-source in a GitHub repository<sup>4</sup> to allow re-use and future development by other researchers. Although all the data collected by users are, by definition, private and are not published, and all system components are designed to comply with this privacy statement, the system offers the option to publish the data as open data. Sponsoring the open-data philosophy and EU open data pilot, ArchAIDE

<sup>3</sup> [https://archaeologydataservice.ac.uk/archives/view/archaide\\_2019/](https://archaeologydataservice.ac.uk/archives/view/archaide_2019/) (accessed 26/11/2020).

<sup>4</sup> <https://github.com/mappaLab/archaide-software> (accessed 21/07/2021), <https://github.com/mappaLab> (accessed 30/11/2020).

suggests to the user to share the data with the community, leaving each user the choice to do that or not.

### Acknowledgements

This research was supported by the EU Horizon 2020 grant agreement No. 693548. We thank all the members of the ArchAIDE team ([www.archaide.eu](http://www.archaide.eu)). This contribution was conceived by Francesca Anichini (AF) and Gabriele Gattiglia (GG) together. AF wrote 1. Introduction, 2.1 Infosoc Directive. GG wrote 2.2 Database Directive and 3. Discussion. 2. IPR Management was written by the authors together.

### Bibliography

- Anichini, F., F. Banterle, J. Buxeda i Garrigós, M. Callieri, N. Dershowitz, N. Dubbini, D. Lucendo Diaz, T. Evans, G. Gattiglia, K. Green, M.L. Gualandi, M.A. Hervas, B. Itkin, M. Madrid i Fernandez, E.M. Gascón, M. Remmy, J. Richards, R. Scopigno, L. Vila, L. Wolf, H. Wright and M. Zallocco 2020. Developing the ArchAIDE Application: A Digital Workflow for Identifying, Organising and Sharing Archaeological Pottery Using Automated Image Recognition. *Internet Archaeology*. 18 March 2020. <https://doi.org/10.11141/ia.52.7>.
- Gualandi, M.L., G. Gattiglia and F. Anichini 2021. An Open System for Collection and Automatic Recognition of Pottery through Neural Network Algorithms. *Heritage* 4 (1): 140–59. <https://doi.org/10.3390/heritage4010008> (accessed 13/10/2021).
- Traille J.-P., J. de Meeûs d'Argenteuil and A. de Francquen 2014. *Study on the legal framework of text and data mining (TDM)*. [http://ec.europa.eu/internal\\_market/copyright/docs/studies/1403\\_study2\\_en.pdf](http://ec.europa.eu/internal_market/copyright/docs/studies/1403_study2_en.pdf) (accessed 13/10/2021) <https://doi.org/10.2780/1475>.
- Wright, H., K. Green, F. Schäfer, H.S. Hollander and H.D. Tjalsma 2016. *Archaeology and Intellectual Property Rights: CARARE - Project Deliverable*. <https://pure.knaw.nl/portal/en/publications/archaeology-and-intellectual-property-rights-carare-project-deliv> (accessed 13/10/2021).

### Supplementary Materials

Open Data is available online at [https://archaeologydataservice.ac.uk/archives/view/archaide\\_2019/index.cfm](https://archaeologydataservice.ac.uk/archives/view/archaide_2019/index.cfm) (accessed 13/10/2021).