Archeologia e Calcolatori
34.1, 2023, 135-144
doi 10.19282/ac.34.1.2023.15

IADI: AN OPEN INTERACTIVE ATLAS OF DIGITAL IMAGES FOR THE JOURNAL «ARCHEOLOGIA E CALCOLATORI»

1. Before the Atlas

1.1 *The commitment for open visual resources*

Visual resources are fundamental complements for the understanding of scientific literature. In the Arts and Humanities domain, this is especially true as regards material cultural heritage publications, which are – for their very nature – extremely rich in embedded images (photographs, drawings, charts, etc.). In recent years, journals and editorial platforms have put much effort towards improving the fruition experience of their visual contents on the web, offering solutions for the display of static images drawn from publications, and often of the 3D models described therein.

The journal «Archeologia e Calcolatori» (A&C) was founded in 1990 as an international observatory of research in computing and information technologies applied to archaeological studies. The journal's website, established in the mid-1990s for the promotion of its contents, started providing the open access PDF versions of the articles in 2005 (http://www.archcalc.cnr.it/). It also adhered to the Open Archives Initiative and implemented an OAI-PMH repository to provide the articles' descriptive metadata (Barchesi 2019). Soon afterwards, digital versions of the previous print-only issues were also made available on the website, covering the entire production of the journal.

Since the beginning of the online publication of the articles and the open access commitment, the editorial board has paid attention to the enhancement of digital visual contents of the publications besides the textual ones. For instance, as long as colour tables were being printed at the end of each volume, a specific section of the website hosted a digital gallery of the images contained therein (Fig. 1). The implementation of the gallery with new contents naturally stopped as colour images started being embedded into the PDF files of the respective articles since vol. 20 (2009)[1].

Driven by the constant motivation for facilitating interoperability and reusability of materials and data produced, «Archeologia e Calcolatori» recently committed to implementing its digital ecosystem with new cataloguing, navigation, and provision functionalities for the visual resources, to the benefit

[1] For a historical framework of the online provision of contents of traditional vs digitally-born journals in the archaeological domain, see Fromageot-Laniècpce 2019, 160-211. On the solutions adopted by archaeological journals for viewing visual contents, see Rossi, Paraciani 2021, 339-340.

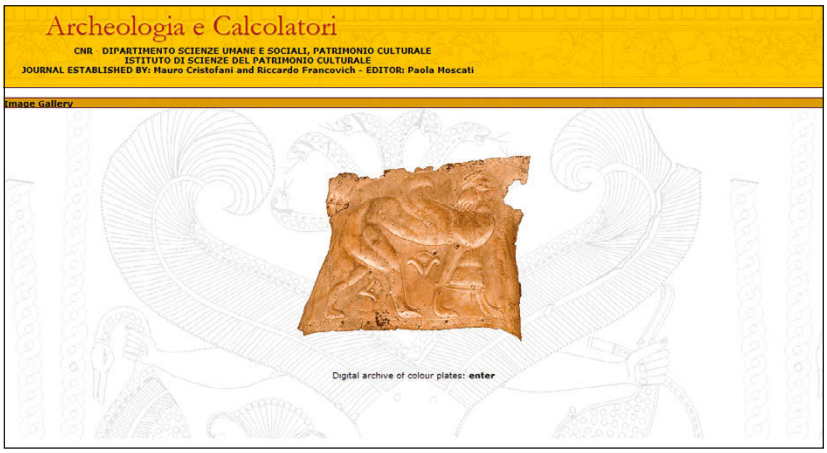

Fig. 1 – Image Gallery access page in A&C website.

of the scientific community interested in the application of IT technologies to archaeology and cultural heritage.

1.2 *Integrated access to A&C resources since 2021*

The opportunity of providing full direct access to the images of the publications as standalone resources, aligning «Archeologia e Calcolatori» with the above-mentioned evolving trends in archaeological publishing fruition, recently gained new impulse by the 30th anniversary of the journal (Moscati 2019), with the decision of turning A&C into a content provider for OpenAIRE, which at the time was expanding its scopes beyond the aggregation of literature metadata towards the harvesting of related data archives.

Indeed, after the repository records of A&C textual resources were aligned to the *OpenAIRE Guidelines for Literature Repositories v3* in 2020[2], the journal turned its efforts towards the management of its visual archive. As a consequence, since the publication of vol. 32.1 in 2021, A&C started providing the figures extracted from its articles and any related 3D models as independent resources through the implementation of its custom digital ecosystem (database, website, and repository) (Rossi, Paraciani 2021).

The database model was enriched with specific interrelated entities for the recording of image and 3D model resources, provided with relevant descriptive attributes. The pages of the public website presenting metadata of

[2] This process was described in Rossi, Paraciani 2021. Afterwards, A&C repository records were further aligned to the *OpenAIRE Guidelines for Literature Repository Managers v4* (https://openaire-guidelines-for-literature-repository-managers.readthedocs.io/en/v4.0.0/).

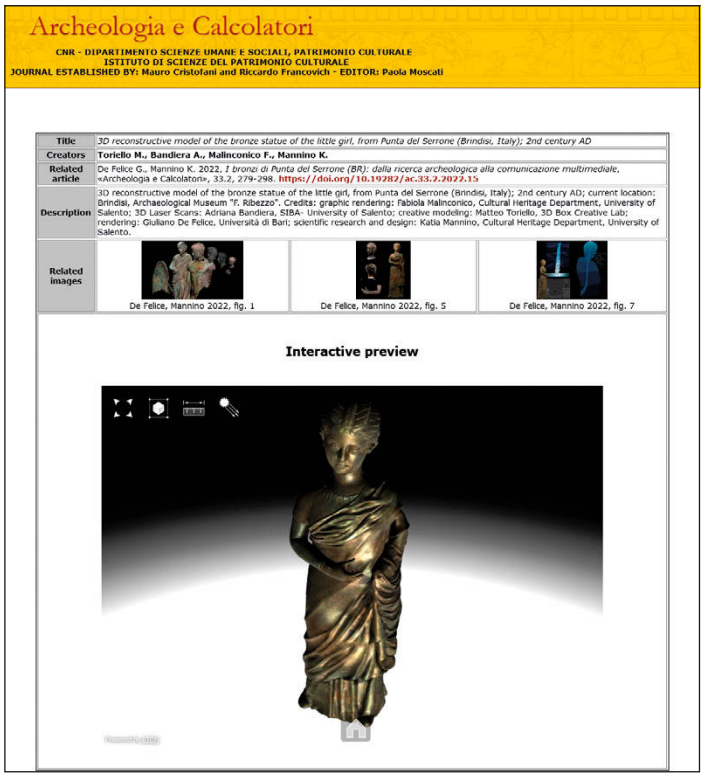

Fig. 2 – A 3D model resource in A&C website, referring to a 2nd c. AD bronze statue of a little girl from Punta del Serrone (Brindisi, Italy) from the article De Felice, Mannino 2022 (http://www.archcalc.cnr.it/journal/resource.php?id=model:25).

individual articles drawn from the database were adapted to include links to the web pages displaying each related visual item and its descriptive metadata (e.g., creators, sources, description, licence, etc.), thus creating a net of interlinked textual and non-textual resources. Not only can images and models be viewed in the website[3], but users can also freely download them (Fig. 2). Images and 3D models have been also exposed in A&C's OAI repository as standalone records, in order to enhance interoperability and reuse of such resources. The inclusion of specific metadata elements for the mutual links between image and model records and with the related literature records

[3] For the visualisation of the 3D models, the website was integrated with the open source, web-based software framework ATON 3.0 (https://doi.org/10.5281/zenodo.4618387).

allows maintaining the connection among resources and clearly expressing the semantics of these relations.

The thorough description of the visual resources getting published every year has involved the collaboration of the authors in providing the metadata, especially as regards specific Dublin Core properties such as: "creators", in case visual records are not (or only partially) original creations of the author of the article; potential "sources", when an image is reused from a previous publication; and "rights", as A&C allows to release visual resources with less restrictive licences than the journal's default CC BY NC ND 4.0 International licence. The inclusion of such descriptive metadata ensures full compliance with the *OpenAIRE Guidelines for Data Archives* (https://guidelines.openaire.eu/en/latest/data/index.html) in view of harvesting and aggregation of A&C visual resources by OpenAIRE. Of course, the synergy with the articles' authors is all the more indispensable when it comes to the provision of additional, standalone contents such as 3D models.

2. A&C_IADI: an Interactive Atlas of Digital Images

2.1 *A legacy through images: 30 years of research in archaeological computing (1990-2020)*

The resources concerned by the work described above currently amount to the ca. 630 images and 25 3D models related to articles of the four journal volumes published since the kick-off of this initiative in 2021 (vols. 31.1, 31.2, 32.1, 32.2). This number will grow as new volumes of «Archeologia e Calcolatori» journal and Supplements will be published. Still, the richness of A&C's archives called for an action of valorisation and sharing of a digital heritage formed by approximately 4000 images spanning the previous thirty-year-long history of the journal (1990-2020). In line with the FAIR principles which have informed the journal's policies through the years, the intent is not only to make those resources available for fruition and reuse, but also to enhance their information potential by annotating them with descriptive metadata.

Therefore, a further initiative has been undertaken to build a new portal presenting 30 years of history of Archaeological Computing through the journal's images: A&C Interactive Atlas of Digital Images (A&C_IADI). The Atlas is one of the projects promoted by «Archeologia e Calcolatori» aimed at enhancing both the multimedia approach – as a dynamic process opposed to a static way of storing and querying documentation and characterised by instantaneity (Orlandi 1999, 3) – and the complex world of computer eidology – which started with the digital representation of the real world (Guidazzoli, Forte 1992) and is now increasingly oriented towards the complex and promising world of virtual reality.

2.2 *Preliminary work on data extraction*

To create a new archive of images and metadata of the volumes published between 1990 and 2020, it was necessary to operate at different levels. Information about titles, authors, and volume numbers of the articles were extracted from A&C database, and captions and images were extracted from the source files of the articles. Indeed, we shall recall that A&C is not a digitally-born journal, and has preserved the tradition of producing printed volumes along with the provision of their digital (PDF) versions on its website. The last phases of the editorial work, including the final proofs' correction, are carried out by means of desktop publishing software. Therefore, contrary to digital-only journals which normally provide HTML or XML versions of their publications, no plain or marked-up text files of the final versions of the full articles of A&C were available for past issues to facilitate the extraction and association procedure of figures and captions. This goal was instead achieved by means of a process carried out by A&C's publisher Edizioni All'Insegna del Giglio in Florence, on the paged, print-ready versions of the articles[4].

The publisher was provided with spreadsheets containing the articles' data which had been retrieved from A&C database, to be implemented with information on images and captions. After converting A&C volumes' files to the latest version of Adobe InDesign, the publisher isolated the captions in the files of each volume by filtering the paragraph styles, and removed segmentations of their texts over multiple lines. A list of unique names of the JPG images contained in exported PDF files was generated as a plain text file, which could then be imported into the spreadsheet, one row per filename. A similar process was adopted to extract all the captions, save them to a plain text file – one row per caption – and then import them into the spreadsheet. This ensured that each row in the spreadsheet had the proper relation between image filenames and captions, and that these were associated with the related article. The workflow applied to the very first issues of the journal differs in that the source files consisted in the digitised versions of the printed articles, which required an OCR conversion of the PDF files into editable text files.

Data retrieved were then imported into the dedicated database table of the new web tool developed for the public presentation of A&C 1990-2020's visual contents.

2.3 *Database implementation*

A&C's Interactive Atlas of Digital Images was conceived as a standalone application with respect to the journal's website, although it accesses the same

[4] We wish to thank Tommaso Ariani, head of the publishing house, for sharing with us the procedure he followed for the retrieval and extraction of data from the original source files of the articles, which implied a careful work on "legacy resources".

underlying database. Since the application is a static archive of the first 30 years of journal publications, the database state is "locked", meaning that it is read-only. We then defined and implemented the database table structure that would hold all Atlas images. Each row in the table represents an image, which is defined by the following main attributes:

– Caption
– Filename
– Figure number
– Article the image belongs to.

The filename simply points to the physical image file in JPEG format, stored on the same server as the web application. The article the image belongs to is identified by a set of specific attributes that are all part of the same table, including its title, the order it appears in the respective volume, and the volume number and type.

It was also necessary to add an attribute holding the actual numeric identifier of the article as found in the DB, to facilitate some queries used by the web application. This task was accomplished by means of a custom command line interface (CLI) tool, which is versioned along with the web application and exposes a few additional commands to manipulate the database table. Indeed, the application does not rely on a migration system such as those found in many modern web frameworks, at least for the time being, considering the simplicity of the underlying structure. The fact that references to foreign tables are minimal reduces the number of joins required to perform "SELECT" queries on it.

2.4 *Web application*

The main web application is written using PHP 8.1 as server-side language, with plans to include features from the recently released 8.2 version (December 2022). Specifically, it implements the widely used Model-View-Controller (MVC) design pattern (Leff, Rayfield 2001) by leveraging individual Symfony components (https://symfony.com/doc) and adopting an Object-Oriented Programming (OOP) style. The reason for choosing PHP as server-side language is twofold: i) previous experience in working with it; ii) possibility to reuse parts of the code for the new version of A&C website, which is currently running in a PHP hosting environment. The frontend is implemented with native rendered views (via the Plates library; https://platesphp.com), with additional client-side JavaScript for some parts – especially those generating content from AJAX calls. In terms of stylesheets, the Atlas relies on the open source minimal CSS library Spectre.css (https://picturepan2.github.io/spectre/), mainly for the basic layout and to ease the implementation of responsive features.

Fig. 3 – A&C_IADI: application's home page, providing access to journal and Supplements' volumes.

Fig. 4 – A&C_IADI: image browsing by a volume's articles, listed by their publication order. The full title of the article is displayed by hovering with the mouse over its card (detail of Limoncelli, Scardozzi 2013 in A&C 24).

Fig. 5 – A&C_IADI: browsing by an article's images (above) and image slideshow with captions visualised on a desktop screen (below). The example refers to Fig. 2 in Limoncelli, Scardozzi 2013, showing some virtually reconstructed monuments of Hierapolis (Phrygia).

The application has a simple structure, in terms of navigation. The home page (Fig. 3) shows card-like buttons in order to browse images by volume (grouped into journal and Supplements issues), while the main navigation bar includes links to the "About" page and the search feature, which is still under implementation. Navigating back and forth between views is facilitated by breadcrumbs on all pages. Clicking on a volume opens a view listing all the articles published in that specific volume. The image representing each article

card is selected randomly by the system from the set of figures related to the article itself (it can vary with each request). Articles are identified by their publication order within the volume and by the short citation. Hovering with the mouse over an article card shows the article's full title (Fig. 4).

After selecting an article, a gallery view opens with all its figures and truncated captions. Clicking on any figure in the gallery opens a slideshow of all images with full captions, implemented with the open source JavaScript library Spotlight.js (https://nextapps-de.github.io/spotlight/) (Fig. 5). At the bottom of the image gallery in the article view the full article citation – formatted according to the journal's style – is included. There is also a "Read article" button that opens the article's page in A&C's website in a new browser tab. A search feature provides the user with the possibility to search for keywords in image captions, which often include terms related to the images' subjects, in addition to filtering by other metadata (e.g., publication year, article title and authors etc.).

3. Future perspectives

In view of enriching the search feature with filters for image categories (cf. https://idai.world/what/images; https://arachne.dainst.org/), terms extracted programmatically from image captions could be used to generate labels for the automated classification of image topics, to provide granular filters for specific searches. We are considering Machine Learning (ML) frameworks and tools for image classification, with a preference for open source software, such as, e.g., the Python framework TensorFlow (Seetala, Birdsong, Reddy 2019) that we started using for some preliminary tests. Similar experiences in archaeology or adjacent fields will be evaluated, possibly to reuse any open existing, pre-trained classification models. The data generated by this classification effort, as well as the already available information associated with the images, could be made interoperable by exposing a dedicated REST API that external services and providers could query automatically to retrieve the data in a structured format (e.g. JSON).

The provision of open visual resources from A&C's website and from the Atlas to Europeana (https://pro.europeana.eu/), by implementing respective repositories with the Europeana Data Model (EDM) format, is also being considered to further increase dissemination and reuse opportunities. This would augment A&C's presence in the Europeana Archaeology collection, currently consisting of more than 900 literature resources (articles) from both the journal and the Supplements (Piergrossi, Rossi 2019).

Nicolò Paraciani, Irene Rossi
Istituto di Scienze del Patrimonio Culturale - CNR
nicolo.paraciani@cnr.it, irene.rossi@cnr.it

REFERENCES

Barchesi C. 2019, *Archeologia e Calcolatori: un'esperienza pionieristica nel mondo dell'Open Access e dell'Open Science*, «Archeologia e Calcolatori», 30, 39-54 (https://doi.org/10.19282/ac.30.2019.04).

De Felice G., Mannino K. 2022, *I bronzi di Punta del Serrone (BR): dalla ricerca archeologica alla comunicazione multimediale*, «Archeologia e Calcolatori», 33.2, 279-298 (https://doi.org/10.19282/ac.33.2.2022.15).

Fromageot-Laniepce V. 2019, *Les pratiques de la recherche en archéologie à l'heure du numérique. L'évolution de la recherche d'information et de la publication de 1955 à nos jours*, «Archeologia e Calcolatori», Suppl. 12 (https://doi.org/10.19282/acs.12.2019).

Guidazzoli A., Forte M. 1992, *Archeologia e tecniche di eidologia informatica*, «Archeologia e Calcolatori», 3, 37-76.

Leff A., Rayfield J.T. 2001, *Web-application development using the Model/View/Controller design pattern*, in *Proceedings Fifth IEEE International Enterprise Distributed Object Computing Conference (Seattle 2001)*, IEEE, 118-127 (https://doi.org/10.1109/EDOC.2001.950428).

Limoncelli M., Scardozzi G. 2013, *Dalla cartografia archeologica digitale al restauro virtuale: recenti casi di studio a Hierapolis di Frigia*, «Archeologia e Calcolatori», 24, 75-99 (http://www.archcalc.cnr.it/journal/id.php?id=663).

Moscati P. (ed.) 2019, *30 anni di «Archeologia e Calcolatori». Tra memoria e progettualità*, «Archeologia e Calcolatori», 30, 9-138 (http://www.archcalc.cnr.it/journal/idyear.php?IDyear=2019-01-01).

Orlandi T. 1999, *Multimedialità e archeologia*, «Archeologia e Calcolatori», 10, 145-157 (http://www.archcalc.cnr.it/journal/id.php?id=272).

Piergrossi A., Rossi I. 2019, *Archeologia e Calcolatori. Accessibilità e diffusione della cultura scientifica*, in Moscati 2019, 75-92 (https://doi.org/10.19282/ac.30.2019.06).

Rossi I., Paraciani N. 2021, *IT applications to archaeology and the OA diamond journals' challenge. Enhancing access and reuse of textual and visual resources*, «Archeologia e Calcolatori», 32.1, 325-347 (https://doi.org/10.19282/ac.32.1.2021.18).

Seetala K., Birdsong W., Reddy Y.B. 2019, *Image Classification Using TensorFlow*, in S. Latifi (ed.), *16th International Conference on Information Technology-New Generations (ITNG, Las Vegas 2019)*, Advances in Intelligent Systems and Computing, AISC 800, Cham, Springer International Publishing, 485-488 (https://doi.org/10.1007/978-3-030-14070-0_67).

ABSTRACT

Scientific literature understanding benefits from visual resources, which is even more evident in the case of material cultural heritage. In recent years, journals and publishing platforms have been increasingly offering extensive access to publications via the contextual provision of visual media, such as images and 3D models. The diamond open access journal 'Archeologia e Calcolatori', founded in 1990, started publishing its articles in 2005 on its website and has always paid attention to giving proper value and presentation to visual contents related to publications. Indeed, it maintained an online image gallery displaying colour plates from volumes until coloured images started being embedded in the articles' PDFs (since 2009). Then, in 2021, the journal added images and 3D models as resources together with publications and displayed them both as standalone content and in relation to articles. However, this later work did not include the previous thirty-year-long history of the journal, since it required close cooperation with authors. Thus a new dedicated web application was specifically developed to present a structured and visually appealing archive of about 4000 images. The paper illustrates this application, entitled A&C_IADI (Interactive Atlas of Digital Images).