Archeologia e Calcolatori
35.2, 2024, 471-482
doi 10.19282/ac.35.2.2024.49

A NEW WEBSITE FOR THE JOURNAL «ARCHEOLOGIA E CALCOLATORI»

1. Introduction

This paper describes the recent development of a new website for the journal «Archeologia e Calcolatori» (A&C), along with the implementation of new features for its OAI-PMH repository and the addition of a JSON REST API to improve data interoperability. The motivation for developing a new website was mainly due to the need to fully exploit the potential offered by the journal's relational database. This latter was introduced to provide a more reliable system for managing publication data and was originally populated by Salvatore Fiorino, migrating from the previous system based on XML and Microsoft Access (see §1.1). The database has since been continuously updated by inserting data via an administration frontend each time a new journal or supplement volume was published. This preliminary work proved to be essential in the developments that followed.

1.1 *Previous versions*

A very early implementation of A&C's website became available on the internet in 1996 at the address http://cisadu2.let.uniroma1.it/iaei/index.htm (Fig. 1), while the PDF versions of the journal's articles were first published online in 2005. At this latter stage, the website was hosted at http://soi.cnr.it/archcalc/ and it relied on OAISistema to provide open access metadata via the OAI-PMH protocol. The website itself consisted of an ASP codebase, which generated pages using data stored in a Microsoft Access database (Barchesi 2005, 2019). This database was also used to prepare XML files for the OAI-PMH repository (which served them statically via a third-party provider).

A major reworking of the website was then performed by Salvatore Fiorino with a complete rewrite of the source code in PHP, mainly motivated by the obsolescence of the (Classic) ASP language (Fiorino 2019). The website kept most of the existing styling – in terms of layout and page organization (Fig. 2) – and still relied on static XML files to generate article views from available data, which was accomplished by using XSLT. All other pages were written essentially as static HTML documents with PHP used mainly as a templating language (for example, to include common sections such as the sidebar menu). This is in contrast with the new version of the website, which uses routing and views to generate pages dynamically.

The OAI-PMH repository was also reimplemented in 2020 to move away from the third-party provider (and static files), using the relational database

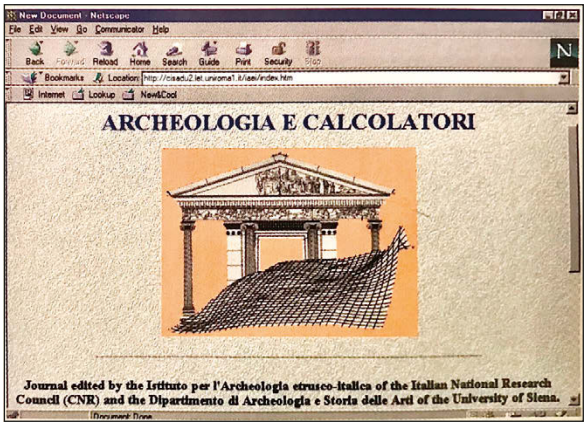

Fig. 1 – The original version of the journal's website, designed upon publication of the Proceedings of the 3rd International Symposium on Computing and Archaeology (Moscati 1996) and hosted by the Centro interdipartimentale di servizi per l'automazione nelle discipline umanistiche of the Sapienza University of Rome (CISADU), directed by Tito Orlandi (image: P. Moscati).

as the main source of data for serving XML responses dynamically according to the protocol's specification. This led to the inclusion of the repository in OpenAIRE as a data provider, via incremental harvesting of the metadata (Rossi, Paraciani 2021).

1.2 *Initial availability of visual resources*

In 2021, new features were added to the website in relation to the fruition of visual resources (Rossi, Paraciani 2021). These consisted of article figures presented both as a gallery of images in the article's page and as individual resources with their own URLs. The same approach was applied to 3D models related to articles, which were displayed in an interactive preview using the ATON framework (Fanini *et al.* 2021) as a web 3D viewer. Where applicable, the models could also be downloaded as GLTF/GLB files with an open access license chosen by the authors.

2. Description of the new website

2.1 *Reasons for technical choices*

When considering the choice of a web development language and framework for the new website, several possibilities were evaluated in terms of performance, package ecosystem and existing knowledge from previous

Fig. 2 – A page from the previous version of the website (volume view).

development work. The possibilities were mostly JavaScript with NodeJS, Python with a web framework (such as Django or Flask), and PHP, either with the Symfony framework or with individual components. This latter option was deemed the best final choice because of already established knowledge of the language and the possibility to reuse code from other projects.

The decision to implement a custom solution almost from scratch instead of relying on open-source CMS software, for example, or – better still – on the Open Journal System (OJS) platform, was taken because of two main reasons: 1) the existence of an already populated database holding 34 years of publications data; 2) the specific custom features that were to be transferred to the new website (mainly the management of visual resources). Reason 1) implied the need to manually insert the data for all volumes and articles into a system like OJS, which has its own database schema and is thus not compatible with the existing one. Reason 2), instead, meant that a CMS or OJS itself would have to be modified at source code level to implement custom features as needed. Since there was no sufficient pre-existing knowledge in this respect, a considerable amount of time would have to be spent in learning a new development environment and uploading data for about 1200 articles.

The fact that the new website is based on Symfony components makes it easily adaptable to the full Symfony framework, which could be beneficial in terms of stability and future updates.

2.2 *Database changes*

Some minor changes to the database structure were required in order to prepare the data for optimal usage by the website backend. A new entity

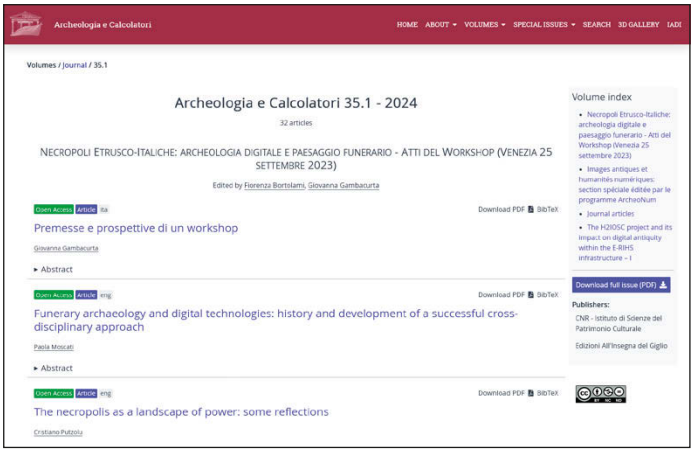

Fig. 3 – Volume view (partial) with linked section titles in the right-hand sidebar.

was added to model sections within volumes and relate them with the corresponding articles and curators. This allowed querying the database to obtain section titles with curators' names directly for a given volume, as well as the list of articles in each section. Previously, displaying section titles and their occurrence within the articles sequence of a volume was difficult due to the absence of a separate entity to hold this kind of data. The new model solves this problem. Sections, when present, are now displayed in the volume's table of contents with links for easier navigation (Fig. 3). These links can also work as permalinks pointing directly to a specific volume section.

A pivot, or relationship, table was also added to relate articles with Pleiades places (https://pleiades.stoa.org/) and implement geographical coverage. The table models a many-to-many relationship since a single article can be related to several Pleiades places and vice versa. This addition proved useful for data retrieved via the REST API as well (see §4.1).

2.3 *Backend implementation*

The website's backend is written in PHP 8.3 (latest stable version at the time of development) and adopts several modern features of the language, such as pattern matching, strict type checks, read-only class properties, etc. As mentioned above, the backend relies on some Symfony components that facilitate the implementation of the common Model-View-Controller (MVC) pattern for web application development and provide useful features in a reliable manner. The codebase adopts an Object-Oriented Programming style, with entities modelled as objects with methods to retrieve data from

the database, which is queried using the Doctrine DBAL library (https://www.doctrine-project.org/projects/dbal.html). This data is itself mapped to so-called Data Transfer Objects (DTOs) that hold information in strictly typed read-only properties 'known' to the code editor (which can then show suggestions while typing), instead of relying on plain associative arrays with arbitrary column names mapped to string indexes.

Since images and their metadata are available in the website's database starting from 2021, in order to include them also for articles before that date, the backend interacts with IADI, the Atlas that hosts all images for the first

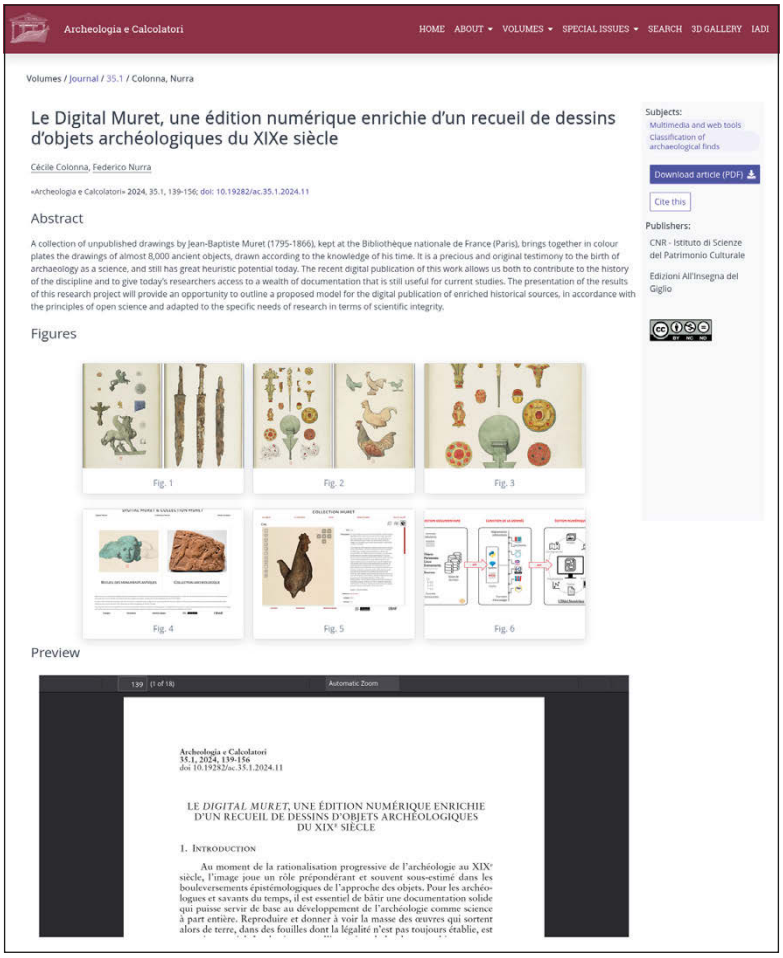

Fig. 4 – Article view with related images and PDF preview.

30 years of publications (Paraciani, Rossi 2023), forming a link that is displayed in the article view and directs to the corresponding page in IADI.

2.4 *User interface (frontend)*

The frontend or user interface presented by the website is generated mostly with views or templates, written using Plates (https://platesphp.com/), a templating library that employs native PHP capabilities. This library adds a 'lightweight' dependency and is easily extensible. In the future, however, the views could be ported to Twig, a templating engine with its own dedicated syntax and features that is perhaps more well-established. Minimal JavaScript is used for interactive features and AJAX calls to the REST API, mainly to retrieve citations for articles, which are available in three different styles: A&C,

Fig. 5 – List of articles by a single author (partial).

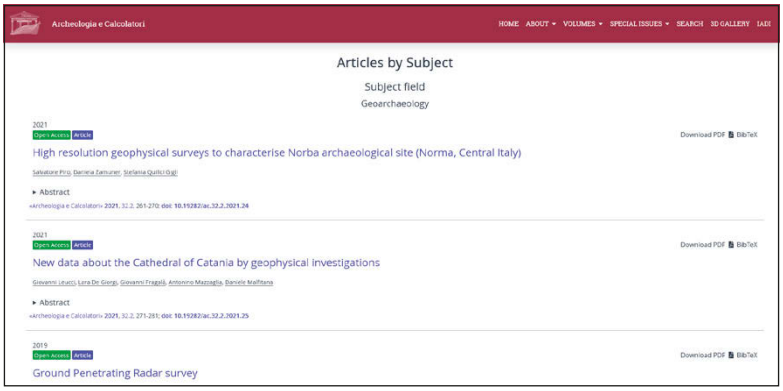

Fig. 6 – Example view of all articles for a given subject (partial).

APA and Harvard (the latter two styles were added in the new version of the website). The citation can also be downloaded in the BibTeX format and then imported in a reference management software of choice, for example. The RIS file format for citations will most probably be added in the future. This feature is available in the article view (Fig. 4), which displays information for a single article along with any related images, 3D models and supplementary materials.

The names of authors are linked to (dynamic) individual pages with the list of all articles by a specific author (Fig. 5). Subjects in the right-hand sidebar of the article view are also linked to the corresponding list of articles (Fig. 6). Static pages, such as 'About', 'Credits and Contacts', 'Policy and Guidelines' etc., are generated from Markdown sources, instead of more complicated templates, which are parsed and rendered as HTML by a single controller. This makes it easier to add new static pages since Markdown is a very simple and minimalistic mark-up language.

2.5 *Search feature*

The search feature has also been redesigned as a structured form with filters for text, author, publication year, publication type (journal or supplement), subject field and computer technology. The text filter allows searching for arbitrary keywords or sentences – enclosed in double quotes – and supports wildcards with the asterisk (*) symbol and the exclusion of words by prepending them with a minus (-) symbol. Keywords separated by a space are interpreted as concatenated by a logical AND operator. The text search is performed in the content of article titles and abstracts. The author filter has an autocomplete feature that shows possible author names during typing, while the subject field and computer technology filters are selections over the list of A&C subject terms. By default, a limit of 15 search results per page are shown, although this value can be parametrized (Fig. 7).

3. REST API implementation and OAI-PMH repository updates

The development of the new website offered the opportunity to add features and improve existing services, specifically the implementation of a REST Application Programming Interface (API)[1] and updates to the journal's OAI-PMH repository.

[1] The Representational State Transfer (REST) paradigm is a software development style which defines seven constraints to be respected when designing and developing web services, although technically it is not limited or bound to the web specifically. Without going into too much detail, we can say that the main requirements of this paradigm are that the client-server communication should be 'stateless' (the result of a client request depends only on the parameters passed to the server and has no effect on the server's state for subsequent requests), cacheable and with a uniform interface centred on resources (Richardson, Amundsen, Ruby 2013).

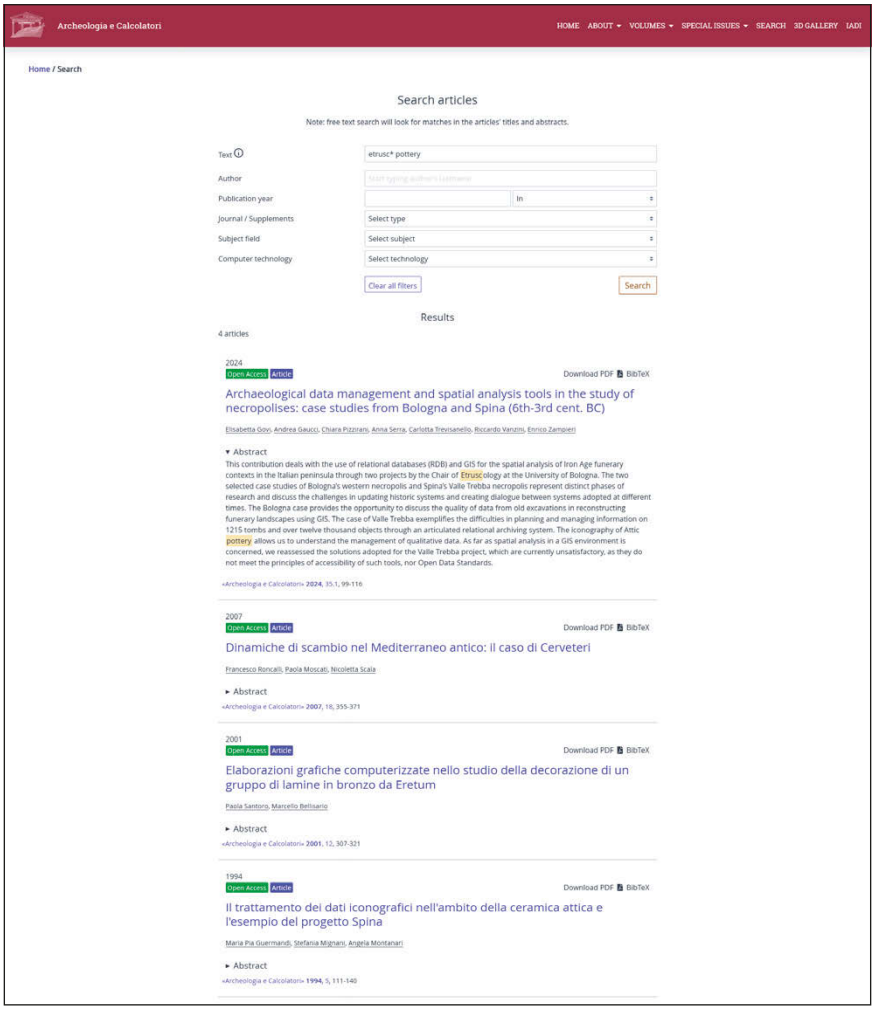

Fig. 7 – The search form with results.

3.1 *The JSON REST API*

Given the increasing reliance on the JavaScript Object Notation (JSON) document format for the exchange of (meta)data among modern web services and applications, typically via a REST API (Serbout, Lauro, Pautasso 2022), it was decided to implement one for A&C, in addition to the OAI-PMH repository. Although the metadata related to A&C publications were already openly accessible via the repository, these are served as XML documents

(required by the protocol), which can be cumbersome to parse for any third-party services that do not rely specifically on the OAI approach. The REST API offers an advantage in this case, since reading JSON documents programmatically is easier compared to XML, given the widespread support for this data exchange format in many programming languages. More specifically, the JSON document can be converted directly into a native data structure and can be manipulated comparatively easily and in a performant way, which is not always the case with XML (Vanura, Kriz 2018).

Currently, the beta version of the REST API consists of the following five endpoints (the last two are not JSON endpoints):

– /api/articles: Retrieves a list of the IDs of all articles found in the journal's database; it can be filtered by geographic place (Pleiades ID);
– /api/articles/{id}: Retrieves metadata for a single article based on its ID;
– /api/articles/range/{firstId}/{lastId}: Retrieves metadata of all articles included in a given range of IDs;
– /api/cite/{id}: Returns the citation for a given article formatted in HTML according to a citation style, controlled by a parameter;
– /api/cite/{id}/bibtex: Returns the citation for a given article as a BibTeX file download.

The REST API is documented with the OpenAPI (formerly 'Swagger') specification (https://www.openapis.org/), which offers a way to generate the documentation automatically along with a graphical testing interface for each endpoint (Fig. 8). The interface shows the HTTP method associated with the endpoint, the formal structure of the endpoint itself (URI with dynamic path parameters) and details about any required and optional parameters[2]. An HTML form allows testing the API directly in the browser. The results are formatted and displayed with colour-coded syntax.

The OpenAPI configuration is itself generated dynamically when a client sends a request to the '/api/openapi' endpoint and is served in JSON format. Configuration parameters are defined using PHP 8 attribute syntax, supported by the Swagger-PHP library (https://zircote.github.io/swagger-php/). Adopting this kind of (informally) standardized description makes it possible to offer a well-known and easily reusable documentation interface, which allows potential users to test the API endpoints directly.

The addition of this API also facilitated interoperability with the ArchaeoHub system, a hub for digital archaeology that is being developed within the framework of the H2IOSC project (Caravale, Moscati, Rossi 2024). In particular, the hub collects article data for topics related to geographic places

[2] Parameters embedded in the URI are called 'path parameters', while those appended to the URI via the HTTP query syntax (following a question mark) are called 'query parameters'.

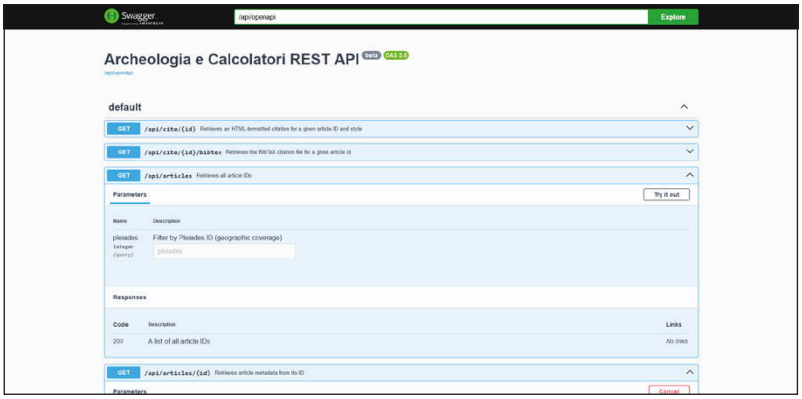

Fig. 8 – The OpenAPI (Swagger) web interface for A&C's REST API.

by filtering relevant article IDs from the REST endpoint, using the Pleiades identification number of the place as a parameter.

3.2 *Updates to the OAI-PMH repository*

The backend code that generates the OAI-PMH repository was entirely rewritten since the existing one suffered from so-called 'technical debt' and wasn't suitable for the new codebase. Since managing and writing XML programmatically with default PHP libraries (specifically the DOM library) is somewhat cumbersome, a more flexible approach based on object serialization was adopted. Each metadata format served by the repository – i.e. Dublin Core, OpenAIRE and DataCite – is modelled as an object class implementing a common interface, which represents the OAI record in that specific format. The object is then serialized to XML based on the values of its (read-only) properties – initialized via database queries – and additional fields encoded in associative arrays. The actual serialization, that is the 'translation' of the object to a well-formed and formatted XML string, is performed with the aid of the Symfony Serializer component. This approach makes it easier to add new metadata formats to the repository, since it is a matter of modelling the format as an object class, implementing the base interface, and defining properties as necessary.

Pagination of records and identifiers was also added using the resumption token method, as required by the protocol. The resumption token is effectively a string that encodes a particular state of the request to the repository and is used to simplify harvesting when the list of records is large. Indeed, each request returns only a subset of records, with subsequent requests resuming from the last offset by means of the resumption token, until the end of the list

is reached. The token string adopts a custom syntax – allowed by the protocol – to encode the state of the request, namely the total number of records in the list, the current offset, the token's expiration date, the values of the set and metadata format parameters, as well as any date-filtering parameters (from and/or until). The resumption token string is parsed and converted to an object with typed properties, which allows the backend to retrieve the proper set of records or identifiers for each request.

Finally, a new schema was adopted for the identifiers of articles, namely 'oai:www.archcalc.cnr.it:article:{id}', which is similar to the identifiers assigned to images and 3D models and is arguably more readable than 'oai:www.archcalc.cnr.it/oai/aec_oaipmh2.php:{id}', used in the previous version.

4. Future developments

An important future development is the conversion of the articles' full texts to structured HTML, which could then be displayed in the web page (article view), instead of the embedded PDF preview. The journal's publisher, Edizioni all'Insegna del Giglio, has already provided the Editorial Board with XML versions of all volumes from 1997 onwards. These XMLs, while being more useful than PDFs to extract content as plain text, were exported automatically from the corresponding PDF documents and are not structured in a way that can be easily converted to HTML as needed for the website. However, some parsing and string-manipulation techniques could be applied to retrieve content from the XMLs and then build the proper HTML documents. These could be then persisted in the database and displayed with the rest of the article's data. The obvious advantage of HTML is that it can be interactive, with links to references embedded in the text, for example, and a browsable table of contents, as is usually provided by many digital journals.

Furthermore, having the content of articles as text in the database would enable searching for keywords or free text in the entire article, instead of just in titles and abstracts.

Another foreseeable development is the use of geographical coverage in the form of Pleiades IDs to, e.g., display an embedded map with relevant places directly in article views and / or search content by geographic place. Finally, the REST API could be improved by adopting a more standardized format for the JSON data. For example, JSON-LD (Linked Data) could be used as a schema, instead of the current custom one, adopting Dublin Core (DC) terms as context for key names, which could be facilitated by the existing DC mapping of database entities used by the OAI-PMH repository.

Nicolò Paraciani
Istituto di Scienze del Patrimonio Culturale - CNR
nicolo.paraciani@cnr.it

REFERENCES

Barchesi C. 2005, *'Archeologia e Calcolatori': nuove strategie per la diffusione di contenuti in rete sulla base dell'OAI-PMH*, «Archeologia e Calcolatori», 16, 225-241.

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).

Caravale A., Moscati P., Rossi I. 2024, *Landscaping and integrating Digital Archaeology and Digital Epigraphy resources: New challenges and future opportunities. Introduction to the Special section*, in A. Caravale, P. Moscati, I. Rossi (eds.), *The H2IOSC Project and Its Impact on Digital Antiquity within the E-RIHS Infrastructure – I*, «Archeologia e Calcolatori», 35.1, 515-520 (https://doi.org/10.19282/ac.35.1.2024.30).

Fanini B., Ferdani D., Demetrescu E., Berto S., d'Annibale E. 2021, *ATON: An open-source framework for creating immersive, collaborative and liquid web-apps for cultural heritage*, «Applied Sciences», 11, 22, 11062. (https://doi.org/10.3390/app112211062).

Fiorino S. 2019, *Lo stato dell'arte dell'innovazione tecnologica per le architetture web: presente e futuro per Archeologia e Calcolatori*, «Archeologia e Calcolatori», 30, 55-74 (https://doi.org/10.19282/ac.30.2019.05).

Moscati P. (ed.)1996, *III International Symposium on Computing and Archaeology (Roma 1995)*, «Archeologia e Calcolatori», 7, 1147-1156 (https://www.archcalc.cnr.it/journal/idyear.php?IDyear=1996-01-01=).

Paraciani N., Rossi I. 2023, *IADI: An open Interactive Atlas of Digital Images for the journal «Archeologia e Calcolatori»*, «Archeologia e Calcolatori», 34.1, 135-144 (https://doi.org/10.19282/ac.34.1.2023.15).

Richardson L., Amundsen M., Ruby S. 2013, *RESTful Web APIs: Services for a Changing World*, O'Reilly Media, Inc.

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).

Serbout S., Lauro F.D., Pautasso C. 2022, *Web APIs structures and data models analysis*, in *2022 IEEE 19th International Conference on Software Architecture Companion (ICSA-C)*, 84-91 (https://doi.org/10.1109/ICSA-C54293.2022.00059).

Vanura J., Kriz P. 2018, *Perfomance evaluation of Java, JavaScript and PHP serialization libraries for XML, JSON and binary formats*, in J.E. Ferreira, G. Spanoudakis, Y. Ma, L.-J. Zhang (eds.), *Services Computing – SCC 2018*, Cham, Springer International Publishing, 166-175 (https://doi.org/10.1007/978-3-319-94376-3_11).

ABSTRACT

Over the course of 35 years since its foundation, the open access journal «Archeologia e Calcolatori» has been providing its publications online via the website, as well as open access metadata since 2005 through the OAI-PMH repository. In 2024, the website underwent a major update and restyling, including the addition of modern interoperability features, most notably the newly implemented JSON REST API. This paper is concerned with the technical description of the design and development of the new website and additional functionalities for «Archeologia e Calcolatori», while also giving a short overview of how the journal's online presence has evolved over time. The exchange of data between the journal's REST API and the Open Archaeology Hub (ArchaeoHub) currently being developed in the context of the H2IOSC project is described as well. Finally, some future perspectives for further improvement are presented.