# SITAR: a new open-data infrastructure for a public archaeology of Rome

Mirella Serlorenzi, Ascanio D'Andrea, Riccardo Montalbano

Mirella Serlorenzi , Ministero della Cultura, SSABAP – SITAR Scientific Director; mirella.serlorenzi@beniculturali.it
Ascanio D'Andrea, Data Manager & GIS expert; ascaniodandrea@gmail.com
Riccardo Montalbano, GIS expert; ricca.montalbano@gmail.com

**Abstract**
*SITAR is a project by the Soprintendenza Speciale Archeologia Belle Arti e Paesaggio di Roma that has changed the relationship between the institution and citizens as regards the knowledge, protection, enhancement and communication of the city's archaeological heritage. The need to open up to external users was taken into account when structuring the technological platform, so that the work carried out by the Soprintendenza could benefit the various communities that make up our society. SITAR – awarded with the Archaeological Heritage Prize 2021 by the European Archaeology Association (EAA) – has become a platform that provides open-data under a CC-BY-SA 4.0 license, characterized by a participatory approach thanks to which different types of users can interact with the institution, directly intervening in the process of creating and reviewing the data.*

**Keywords:** ARCHEOSITAR PROJECT; ARCHAEOLOGICAL OPEN DATA CC-BY-SA 4.0; *LOD*.

## SITAR for a public archaeology of Rome

The health emergency has highlighted the lack of data digitization projects and major infrastructure for their management and communication within public institutions. This delay can be ascribed to a lack of investment in strategic sectors, such as research, technological development and Cultural Heritage. The gravity of the situation can be seen in the failure to consider the extent to which the absence of investments leads to stagnation in Italy's cultural and economic growth, given the collective utility and social function of Cultural Heritage.

One of the problems to be resolved is the absence of infrastructure capable of guaranteeing long-lasting data preservation strategies (*infra*) on the one hand, and on the other the potential to create specific interoperability protocols, creating new shared, open and accessible knowledge without any limitations. We therefore need to rethink the concept of Cultural Heritage, so that archaeology is experienced as an integral part of a local area, whose significance may be of interest, above all, to the communities that have inherited it. New technologies are the tool with which public sector institutions can disseminate data, information and contents extensively, beginning that effort to democratize our heritage that represents both its constitutional fulfilment and the most effective way of safeguarding it.

In this context, already in 2007 the *Soprintendenza Archeologica di Roma* decided to create an archaeological information system: SITAR. The system has evolved over time to become a complex digital platform, structured to host the data on Rome's extraordinary archaeological

heritage. SITAR took its first steps in a fruitful phase of discussion, starting with its participation in two ministerial joint committees that had opened a debate between different institutions, connecting the academic world to that of safeguarding and creating the conditions for adopting shared guidelines that serve the needs and objectives of research, protection and enhancement of heritage (Sassatelli 2011).

The decade that separates us from that debut and the long road travelled demonstrate that SITAR is now among the longest-lived and locally well-rooted platforms; unfortunately, very few digital projects developed over the last decade still exist and operate today. The overview presented in the latest edition of ArcheoFOSS highlighted the difficulties of the medium and long-term survival of projects to digitize and manage archaeological data at various levels (see in this volume Bogdani, Sciacca, table. 3). The burden of keeping a complex system such as SITAR alive over time is onerous and depends on factors that are not always predictable or subject to the will of those who operate and manage it. These include the medium and long-term planning of the objectives to be pursued and the communities to be reached:

- Continuous updating of the technological infrastructure, involving constant evolutionary maintenance with the related costs.
- Data maintenance that also includes storage, where possible prioritizing reliable public partners who can guarantee security and the appropriate dissemination of data.
- The availability of annual grants that may not remain constant or be lower than expected, with the consequent forced rescheduling of planned activities.
- A robust scientific management that oversees the general planning process, considering the variables that arise over time, but simultaneously able to pursue the objectives and the agreed development lines with determination, even in the absence of the necessary conditions.
- The support of a specialized working group, made up of different professionals, who share the project philosophy.
- An awareness that staff numbers are subject to variation, with operational repercussions.
- Contact with universities and the world of freelance professionals, to guarantee the dissemination of the contents and logical of the platform and training on its use.

An information system is thus a living organism that must adapt to the times and change in line with contingent circumstances. Given the difficulties that may arise in the long term, it is advisable to evaluate the existence of well-established projects already in place before embarking on new ones of a similar type. Interoperability protocols now make it possible to create a dialogue between projects of a very different nature. Therefore what is really needed is the launch of collaboration policies between subjects who share fundamental assumptions, such as the provision of open-data and the awareness that the circular reuse of information generates new knowledge.

The following pages focus on the latest developments of the SITAR platform, which have made it possible to meet some of the objectives underlying the project from its early design stages. The solutions available today have made it possible to enhance the collaborative aspects and improve the processes of acquisition and use of SITAR data. Furthermore, a new website has

been created,[1] now also available in English, which provides information on the project at various levels and comprises helpful user support tools, such as tutorials and manuals.

The purpose has always been to provide open data and libre contents to the community, with the aim of opening up a dialogue on fundamental issues such as those related to the protection of Cultural Heritage, the quality of suburban areas, the participatory development of policies for the city's growth (Serlorenzi 2011; Serlorenzi *et al.* 2020). Special users that SITAR has targeted since the beginning include freelance archaeologists working daily in the field on behalf of the *Soprintendenza*. The activities carried out generate archaeological knowledge in the form of scientific documentation, which must be processed and standardized and whose essential data must be made public as quickly as possible to benefit not only the planning of new public works but research in general. In this regard, it was essential to consider the intended social significance of such a complex project managed by a public body (Serlorenzi *et al.* 2020; Figure 5), how to create a dialogue between the institution and users, and the original contributions that may arise from this exchange (Cerami *et al.* 2020).

Here, therefore, it is unnecessary to recall the general outlines of the project;[2] instead, we will illustrate the latest changes that have transformed SITAR into a technologically advanced platform, allowing for the dissemination of high-quality knowledge and improved governance of the local area and its Cultural Heritage, thanks to the involvement of civil society. This approach is also sanctioned by the Faro Convention, which sees the active participation of the community as a new resource, fundamental for the conservation and enhancement of heritage.

*M.S.*

## Three pillar-based infrastructures: data, IT, users

The temporal element in the development and implementation of a public administration Information System is undoubtedly a foundational and structurally congenital value. On the one hand, it may favour a controlled evolution of functions and the reformulation of objectives, whilst on the other it may act as a technological brake given the constant adaptation that such systems require. SITAR is a clear example of longevity both in terms of design and data implementation, having enjoyed an operational continuity that has allowed for a constant renewal over the past decade. Starting from 2018, on the wave of a drive to reformulate the strategic objectives and thus expand the user base, we have embarked upon a significant re-engineering of the system. Four main objectives were taken into consideration and used to guide the process:

- *Scalability*, tailoring the resources on the basis of the actual needs, both in terms of the number of users and the delivery of services.
- *Performance*, increasing the processing capacity with new systems for indexing and request/response management, ensuring an interaction within milliseconds.

[1] https://www.archeositarproject.it/ (accessed 07/08/2021).
[2] For the history of the project, see Serlorenzi 2011; Serlorenzi et al. 2012; Serlorenzi 2013; Serlorenzi and Jovine 2013; Serlorenzi and Leoni 2015; Serlorenzi and Jovine 2017; Serlorenzi, Lamonaca and Picciola 2018; Serlorenzi et al. 2020.

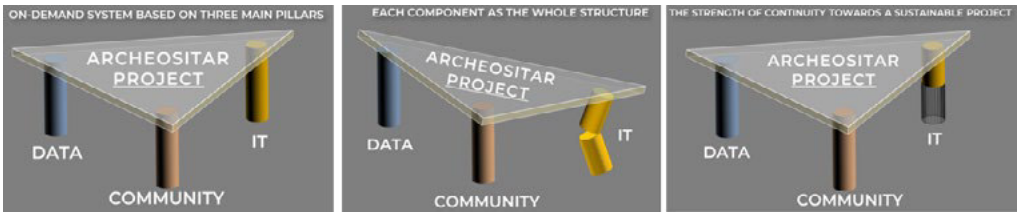


Figure 1: Example of life-cycle of the three main Sitar platform components.

- *Maintenance*, improving scheduled maintenance operations and bug-fixes, guaranteeing the potential for immediate intervention on the individual components of the platform.
- *Interoperability*, with the creation of a data-independent technological system to ensure maximum communication with other existing platforms.

The creation of such a fully 'on-demand' system has greatly optimized the delivery of open-data services, from the outset at the heart of the entire platform. The potential to add/modify 'technologies' and the related exposure of public APIs independent of the data model ensures the system will have a long lifecycle that can easily adapt to the needs dictated by technological evolution. In this regard, it is important to stress that the IT component is just one of the three pillars on which the system is based. At the same level of importance, both quantitative and qualitative, are the 'data' component and that relating to the user experience and more generally to the reference communities (Figure 1).

This is a sound type of structure for SITAR, but more generally it can be considered extensible to all those information systems in the field of Cultural Heritage that aspire to become endemic operational tools within the state sector (and beyond). This is a system of design and operational forces that must necessarily 'work' in unison to ensure full efficiency:

- *The data component:* SITAR reaps the rewards of the innovative, shared and consolidated design of the data structure, at the level of both the conceptual and logical model of the database. The ongoing funding, though not always homogeneous in terms of resources, has made it possible to carry out constant and onerous data entry over time. This process has ensured, and continues to ensure, that the system remains constantly up to date on the state of scientific archaeological knowledge on Rome, a factor that underlies many of the decision-making/administrative and documentation standardization processes within the institution in its relationship with users, whether they are ordinary citizens or archaeological professionals.
- *The IT component:* As described below, the technological investment was tailored to the medium and long-term objectives and not to satisfy specific contingent needs. Thanks in part to prior experience, this is ensuring a lasting lifecycle for the platform. The creation of an on-demand system, with a clear separation of the data structure from the application technologies, now makes it possible to keep maintenance interventions to a minimum. This in turn, albeit with a higher initial investment in terms of both resources and development hours, means that the working group can focus on testing new features to expand the range of services. Once again, the IT component should be seen as a supporting 'pillar' of the structure that, once established, acts as a driving

force for the planned implementations, namely the development of a 3D GIS and the creation of tools for assessing the archaeological potential of the city of Rome.

- *The community component:* This term describes everything related to the user experience and more generally the needs of the project's reference 'communities'. SITAR responds to the need to bring to light Rome's vast documentary archaeological heritage and make it accessible to the public. Over the years there have been discussions, internal to the project and shared through the intensive dissemination of the results, about 'what' and 'who' the project's target audience should be; these discussions have entrenched the belief in the fundamental role played by attention to the system 'outputs'. Creating a tool that satisfies all audiences is impossible, but it is certainly feasible to create one that allows all audiences to use data regardless of individual needs, guaranteeing the possibility of choosing between differentiated services (public APIs, geo-services, direct download or a simple WebGIS interface). The next development will be to further implement the interaction with users, in this case with the specific class of 'insiders', allowing those working in the archaeological sector (companies, individual professionals, etc.) to upload data into the system. This will make it possible to further ensure the continuous updating of the information going beyond the initial objectives of an albeit complex digital archive of Rome.

In this light, whilst on the one hand, in terms of continuity of action, SITAR's longevity among national and international archaeological information systems may be surprising, on the other it allows us to identify some key factors that lay under projects in this sector, also in light of the substantial funds that will shortly be made available for the digitization of Cultural Heritage: a shared conceptual and logical modeling, distributed server technologies, data structures independent of specific products, mapping, analysis and interaction with user communities of reference. An integrated approach to technology and human resources in which to invest in the medium and long term to create endemic and sustainable systems within the Public Administration.

*A.D.A.*

## SITAR cloud-based infrastructure, web application, open-data and open-data services

The SITAR re-engineering started in 2018 and is still in progress. It also led to a general restructuring of the IT infrastructure. Since 2013 the development and implementation of new web applications, combined with a wish to provide additional services and data, had prompted the transition from a self-hosted infrastructure, based on servers located at the *Soprintendenza* Data Centre to one based on virtual machines residing on a server infrastructure managed by the GARR Consortium. The latest evolution (2018–2020) entailed the creation of a new and wholly cloud-based infrastructure, hosted on the GARR network servers and managed through *OpenStack* open-source cloud technology. The GARR IAAS cloud helped in containing the server management and maintenance costs, and offered moreover a guarantee of the long-lasting preservation of the documentary heritage acquired and processed (Fresa and Justrell 2014).

It is important to stress that the new SITAR microservices architecture – based on the separation of the individual components, residing on dedicated machines as shown in the diagram in

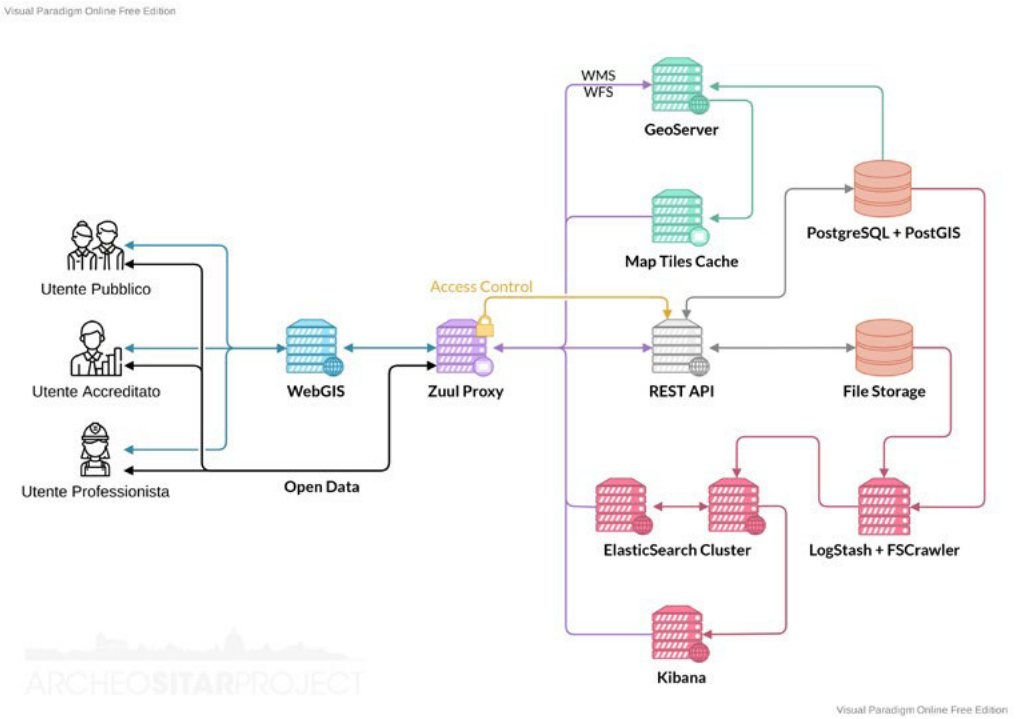


Figure 2: Server and services schema.

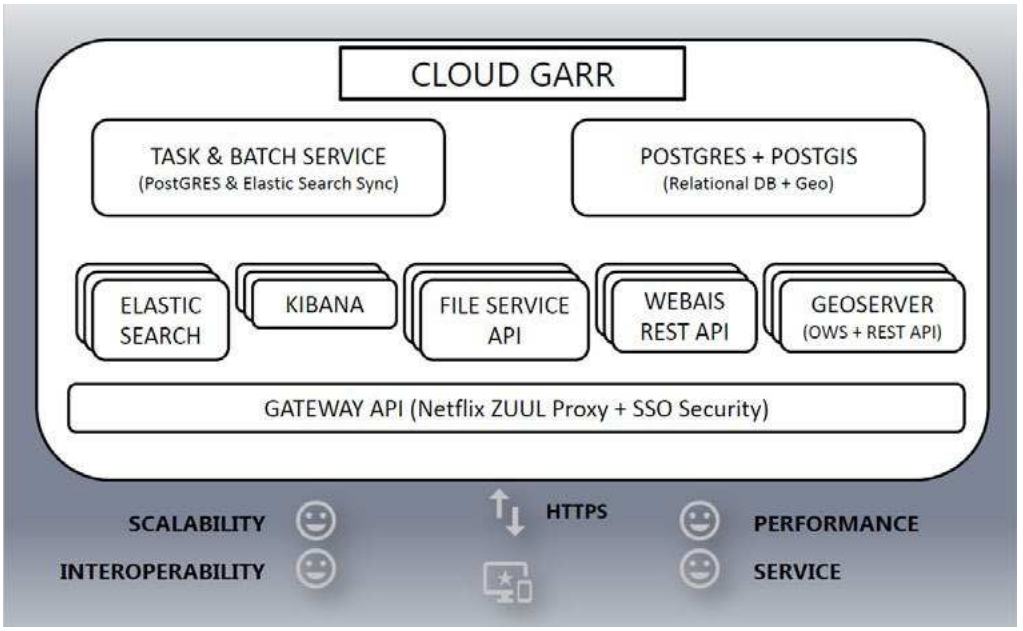


Figure 3: SITAR service scalability.

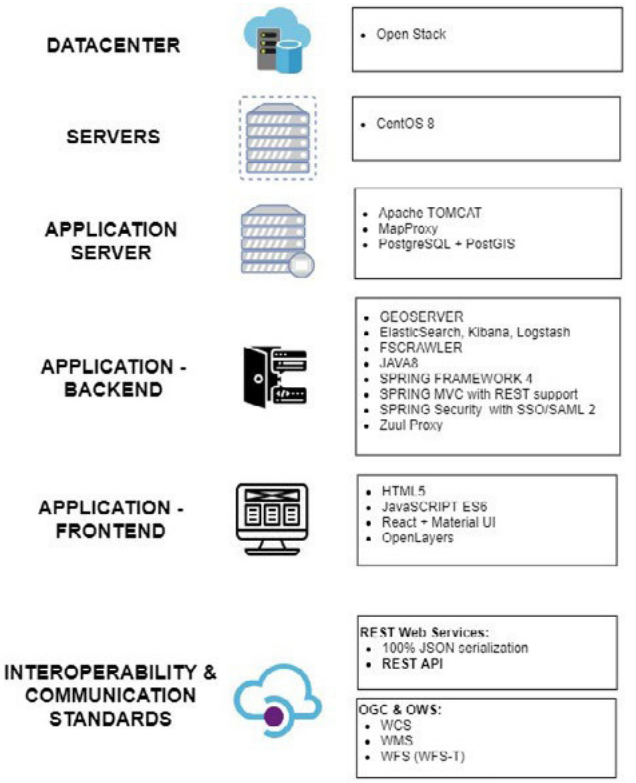


Figure 4: SITAR Tech Stack.

Figure 2 offers countless advantages. The services are distributed independently, resulting in easier maintenance, which can be limited to individual parts of the application. In terms of performance, the interactions between components are more streamlined, the horizontal and vertical scalability can be modulated on demand (Figure 3), and finally the automation of the development phases can be managed more easily, thanks to the implementation of CI/CD (Continuous Integration/Continuous Development) strategies, increasing the frequency and speed of testing and the release of new features. In general, adopting a paradigm based on a distributed cloud-based architecture hosted by an institutional Data Center and the use of open-source solutions (Figure 4), not only helps in being sustainable in economic terms, but also allows for the replicability of the system, which may in the future be cloned and transferred to support academic projects or public sector bodies that share its methodological and conceptual structure.[3]

Compared to the previous application, available online until 2019, the most important novelty of the new SITAR is the merger and integration of the three original *webAIS*, *SIGEDO* and *Ambiente*

[3] A code refactoring is currently under way, aimed at making the application and all its services available to the SOttoSiena project. This is the first step in a long roadmap, launched but not completed, that aims at creating an application that can be made available to other bodies and state institutions free of charge and on the basis of specific memoranda of understanding.

*Tutela* applications within a single environment. Today the WebGIS has a totally renewed interface, rendered more intuitive and enriched with tools that allow users to customize the map layers, search for entities on a geographical basis, filter individual excavation or find, explore their contents, carry out simple measurement operations (areas and distances), download and freely use spatial data and the related attributes in the most commonly used open formats. To allow users to browse the SITAR's rich corpus of documents, a Digital Library was created, using the open-source *ELK* suite (ElasticSearch, Logstash, Kibana). Thanks to the powerful search and indexing engine, users can perform full text and fuzzy searches on the records for the four SITAR entities and on the individual documents linked to them, consisting in texts in pdf format and images in various formats. This is possible thanks to an OCR system that identifies the texts contained within the static documents, which are subsequently indexed. Special filters allow users to refine the results obtained, for example by selecting the logical entity to be searched or the file type (pdf, image, ZIP, etc.).

A need seen as central from the early design stages, but that has now finally been implemented, is the collaborative aspect of the platform. Users now have a new set of tools thanks to which they can actively interact with the database, triggering a participatory procedure aimed at improving the dataset. Through the *request for change* users can send requests to update the data already in the system, report new work in the territory of Rome that has not yet been registered – potentially attaching photographs, videos, documents – or other types of evidence relating to the scope of SITAR and the *Soprintendenza*, about which a user deems it necessary to inform the offices in charge. These reports are subjected to an internal validation process, after which the data is acquired and published.

The new tools implemented also include the 'booklet' function, which allows registered users to create their own selection of SITAR data consisting of individual excavations and/or finds ('bookmarks'), making up a personal library that can also be shared. In terms of data sharing, on the one hand it is possible to publish the records for all entities with a public link on the main social media platforms, while on the other hand, registering with the RSS feed service allows users to receive updated information on the excavations for which updates have been requested, without the need to constantly check them on the WebGIS platform.

A more advanced level of dynamic interaction with the SITAR database is represented by the possibility to transfer the data-entry process to the professional archaeologists who work daily in the city's construction sites. Thanks to the implementation of dedicated back-end tools, to be released in the next months, the SITAR office will be entrusted with the task of identifying the study areas (OI), while the completion of the detailed information, the creation of the records on Archaeological Partition (PA) and the uploading of the accompanying documents will be outsourced to the external professionals, who will be entrusted with the operational privileges for their creation and updating exclusively for the excavations on which they are working. A double mandatory validation system is required for data to be published. A first formal semi-automated verification ensures that the delivery meets prerequisites in terms of georeferencing, attached documents, formats, etc. A second scientific validation of the data is entrusted to the staff members. Once fully operational, this procedure will guarantee real-time updates on developments in the territory of Rome, but above all it will allow the system to self-feed, thus freeing up useful resources to complete the acquisition and processing of all the vast quantity of archive data, still in progress.

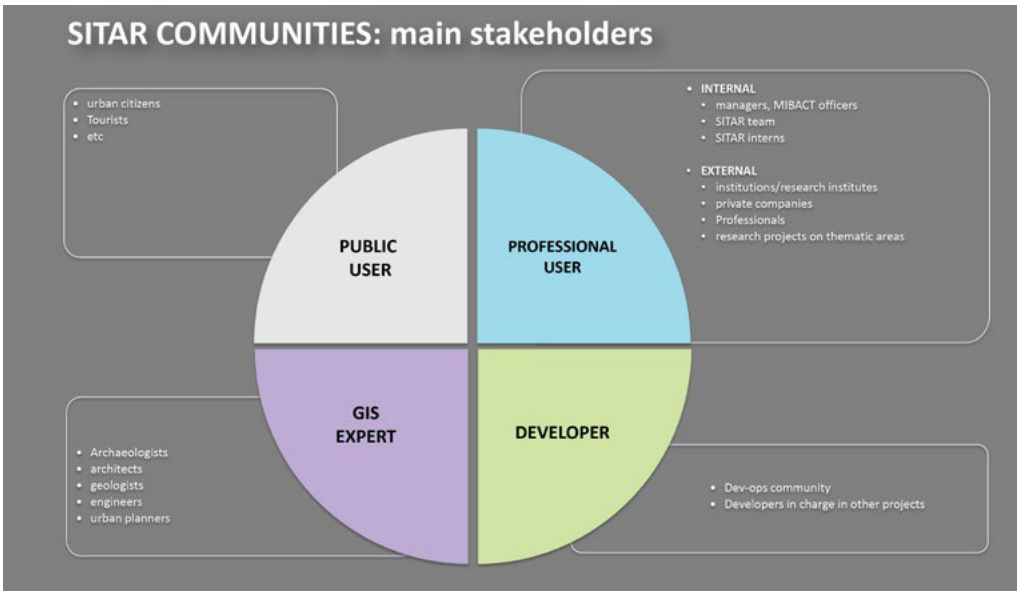


Figure 5: SITAR communities: the main stakeholders.

The SITAR data model has now been recognized on the national and international level, described in numerous publications, aligned with CIDOC CRM through the CRMarcheo extension and shared with the scientific community in linked data format (RDF), through the XLM language.[4] Here, therefore, greater emphasis will be placed on the renewed attention to open-data services, which are now the preferred tool for sharing data. As already mentioned, SITAR has for some time been pursuing a data publication and dissemination strategy in line with the FAIR principles (Serlorenzi 2018). Following the recent developments, SITAR can finally contact its communities of reference (Figure 5) through a help toolkit that facilitates data acquisition by users (Figure 6).

The new application includes a function to export map extracts relating to vector datasets, thanks to which users — using their monitor as a framing area — can freely download the spatial data and with the related attributes with a CC-BY-SA 4.0 license, choosing the format that best suits their needs. The main open GIS formats are available, vector types (Shapefile, GML2, GML3, KML, SVG), raster types (GeoTIFF) and textual types (CSV, GeoJSON). The SITAR platform also delivers geo-data using WMS and WFS protocols, and is thus aligned with the main operating standards and regulations defined by the OGC. At present, only attached documents (scientific reports, photographs, graphic documentation, in general the so-called raw data) are excluded from public use and restricted to *Soprintendenza* staff, who can download and use them if data is related to the area they administer. A geographically-based ownership system has been developed, entailing a dynamic association between excavations and staff through the areas of which they are in charge. The system is able to dynamically assign the new excavations to the officers in charge, allowing them to download the attached documents. By contrast, public users can view a list of available documents, which can be

[4] https://www.archeositarproject.it/wp-content/uploads/2020/10/Mapping-SITAR-towards-CIDOC-CRM-final_12.10.2016-1-1.pdf (accessed 01/07/2021).

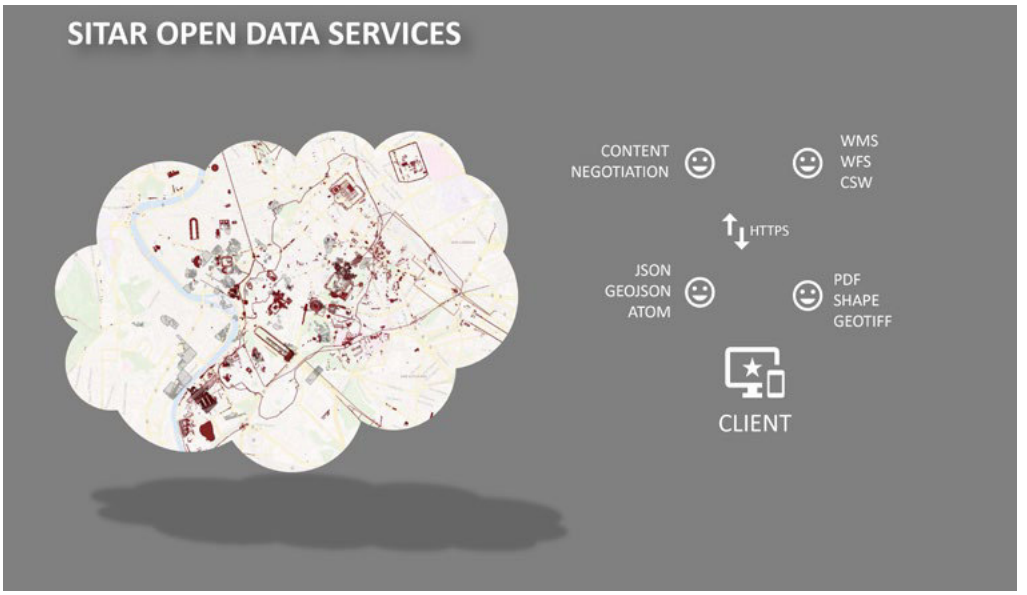


Figure 6: SITAR open data services.

requested to the officers in charge through the integrated request for download procedure. A special management panel allows users to make the request, monitor its status (accepted, in process, rejected) and — upon acceptance — finally download the document.

The need to establish institutional synergies with other public institutions and projects that produce and/or implement datasets in the Cultural Heritage sector required a significant effort in terms of interoperability. Thanks to the latest implementation, SITAR provides a series of specific public APIs to make REST calls both to the Elastic Search engine and to the REST API component.[5] In both cases, the data is returned in JSON format. This technological renewal allows for a further level of data and information services exchange, both inbound and outbound. SITAR will thus be able to receive data from external sources and to configure and provide services in support of other infrastructure and/or projects based on the same interoperability standards.[6] Another stage of the path, now ten years old, has therefore been reached, thanks to which SITAR has become a *trait d'union* between the Public Administration, professionals working on the ground and citizens, making its data available to the various users and communities of reference, who with differentiated skills produce or use data relating to Cultural Heritage.

*R.M.*

[5] https://www.archeositarproject.it/piattaforma/open-data/ (accessed 01/7/2021).
[6] The most recent experiment is a memorandum of understanding that provides for the exchange of data, via API, with the Forma Romae project curated by the Sovrintendenza Capitolina ai Beni Culturali.

## Bibliography

Cerami, F.R., M.L. Scaduto and A. De Tommasi (eds) 2020. *I bacini culturali e la progettazione sociale orientata all'Heritage Making. Tra politiche giovanili, innovazione sociale, diversità culturale*. Firenze: All'Insegna del Giglio.

Fresa, A. and B. Justrell 2015. *A Roadmap for preservation for digital cultural heritage content*. Roma, ICCU - DCH-RP Project. Rome.

Sassatelli, G. 2011. La Seconda Commissione ministeriale per la formazione di un Sistema Informativo Territoriale Archeologico Nazionale. In M. Serlorenzi (ed.) *SITAR. Sistema Informativo Territoriale Archeologico di Roma. Atti del I Convegno (Roma, Palazzo Massimo, 2010)*: 99–101. Roma: Iuno Edizioni.

Serlorenzi, M. 2011 SITAR. Sistema Informativo Archeologico di Roma. In M. Serlorenzi (ed.) *SITAR. Sistema Informativo Territoriale Archeologico di Roma. Atti del I Convegno (Roma, Palazzo Massimo, 2010)*: 9–29. Roma: Iuno Edizioni.

Serlorenzi, M. (ed.) 2013. *ARCHEOFOSS Free, Libre and Open Source Software e Open Format nei processi di ricerca archeologica. Atti del VII Workshop (Roma, 11-13 giugno 2012)*. Archeologia e Calcolatori Supplemento 4. Roma: All'Insegna del Giglio.

Serlorenzi, M. 2018. Accessibilità e diffusione del dato archeologico: l'esperienza del SITAR. In M. Arizza *et al*. (eds) *I dati archeologici. Accessibilità, proprietà, disseminazione (Roma – CNR, 23 maggio 2017)*. *Archeologia e Calcolatori* 29: 31-40.

Serlorenzi, M. and J. Jovine 2013. In M. Serlorenzi and I. Jovine (eds) *SITAR. Sistema Informativo Archeologico di Roma. Potenziale archeologico, pianificazione territoriale e rappresentazione pubblica dei dati. Atti del II convegno (Roma, 9 novembre 2011)*. Roma: Iuno Edizioni.

Serlorenzi, M. and I. Jovine 2017. *Introduzione*. In M. Serlorenzi and I. Jovine (eds) *Atti del IV convegno (Roma, 14 ottobre 2015)*. Archeologia e Calcolatori Supplemento 9, 9-18. Roma: All'Insegna del Giglio.

Serlorenzi, M. and G. Leoni. 2015. SITAR. Sistema Informativo Archeologico di Roma. Il SITAR nella Rete della Ricerca italiana. Verso la conoscenza archeologica condivisa. In M. Serlorenzi and G. Leoni (eds) *Atti del III convegno (Roma, 23-24 maggio 2013)*. Archeologia e Calcolatori Supplemento 7, 199-208. Roma: All'Insegna del Giglio.

Serlorenzi, M., F. Lamonaca, S. Picciola and C. Cordone 2012. Il Sistema Informativo Territoriale Archeologico di Roma: SITAR. *Archeologia e Calcolatori* 23: 31–50.

Serlorenzi, M., F. Lamonaca and S. Picciola 2018. The SITAR Project: Web Platform for Archaeological Knowledge Sharing. In V. Apaydin (ed.) *Shared Knowledge, Shared Power Engaging Local and Indigenous Heritage*: 125–127. Cham: Springer.

Serlorenzi, M., G. Leoni, F. Lamonaca and S. Picciola 2020. Il SITAR e le comunità degli utenti: un'infrastruttura culturale al servizio del patrimonio archeologico di Roma. In F.R. Cerami *et al.* (eds) *I bacini culturali e la progettazione sociale orientata all'Heritage Making. Tra politiche giovanili, innovazione sociale, diversità culturale*: 201–212. Palermo: All'Insegna del Giglio.