CAA2015

KEEP THE REVOLUTION GOING >>>

Proceedings of the 43rd Annual Conference on Computer Applications and Quantitative Methods in Archaeology

edited by

Stefano Campana, Roberto Scopigno, Gabriella Carpentiero and Marianna Cirillo

Volume 1

Archaeopress Archaeology

ARCHAEOPRESS PUBLISHING LTD
Gordon House
276 Banbury Road
Oxford OX2 7ED

www.archaeopress.com

CAA2015

ISBN 978 1 78491 337 3
ISBN 978 1 78491 338 0 (e-Pdf)

CAA2015 is availabe to download from Archaeopress Open Access site

This book is available direct from Archaeopress or from our website www.archaeopress.com

GQBWiki Goes Open

Stefano Costa
(steko@iosa.it, http://orcid.org/0000-0003-1124-3174)

Alessandro Carabia
(alessandrocarabia@gmail.com)

Department of Historical Sciences and Cultural Heritage, University of Siena, Italy

Abstract: GQBWiki is a wiki website for the archaeological research project on the Byzantine city of Gortyn, based on open source software and publicly accessible under the CC-BY-SA license. It has been used since 2005 to record excavation data together with other content ranging from bibliography to reflexive documentation. While the MediaWiki software lacks native capabilities for structured querying, the Semantic MediaWiki extension has been used to provide all the infrastructure necessary for linked open data. GQBWiki is the on-going result of collaborative work and strives to give attribution to all contributors in a transparent way, with all the challenges that a non-traditional publication workflow brings.

Keywords: Multivocality, Collaborative authorship, Open data, Open source

1 GQBWiki

GQBWiki is an online wiki website (http://www.gortinabizantina.it/wiki/) dedicated to the archaeological research project in the Byzantine Quarter near the Pythion shrine in Gortyn (Crete), run by the University of Siena. It has been operational since 2006. While fieldwork at the site started in 2001, it was only in 2005 that we decided to start building a digital archive where the documentation could be collectively created and curated, not limited to excavation data *strictu sensu*, migrating over content from previous relational databases for stratigraphic data. In 2005, choosing a wiki over other available systems seemed to provide strategic advantages, such as being online, always available when and where an Internet connection was available and more generally facilitating the creation of an encyclopaedic resource about the research project. GQBWiki has always been restricted to the research team members until April 2015, so there was no benefit in terms of visibility of the resources that were created and updated. In retrospect, the choice of an online platform brought several 'revolutionary' advantages that took us some time to appreciate their full potential. In this paper, we outline the current status of GQBWiki and what we think we learned in the past 10 years, particularly with respect to our first discussion on the same topic (Zanini and Costa 2006) and a wider overview of the situation for knowledge sharing in the archaeological world (Zanini and Costa 2009).

The research project at GQB has a focus on the Late Antique and Early Byzantine phases of the urban area of Gortyna, and is therefore part of the rather large topic about the end of the ancient Mediterranean city. With this premise in mind, it seemed natural for GQBWiki to become a comprehensive archive where the archaeological record could become part of a hypertext, and could be linked to historical evidence, broader interpretive texts and so on.

On the technical side, GQBWiki is based on the popular MediaWiki software, better known for being used by Wikipedia and other related websites, but also available for use by third parties under the GNU General Public License. Since we adopted it in 2005, MediaWiki has been constantly updated and improved by the Wikimedia Foundation and by other contributors, so far reducing the risk of finding ourselves with an obsolete tool – while other pieces of wiki software were abandoned in the meantime – acknowledging that the maintenance of such a complex tool is well beyond the technical capabilities of a small team, not to mention the increasing need to keep web-based software free from security bugs that may put the privacy of users at risk. MediaWiki is built on the well-known LAMP platform (Linux, Apache, MySQL, PHP) and can be run with no difficulty on any web hosting service, at least in its basic functionality. GQBWiki is used all year round, but it is essential during the fieldwork season. Due to the lack of an Internet connection at the mission house in Agioi Deka near Gortyn, it was only in 2012 that we could work directly on GQBWiki using a commercial mobile broadband Internet provider. In the previous years, we would simply take advantage of software freedom and the flexibility of GNU/Linux systems to install a local wireless network with a web server running MediaWiki on a spare laptop and a local 'clone' of GQBWiki (the online version was put in read-only mode). At the end of each field season, the updated content has been put online again until the next year. This approach does not seem very widespread and in our case in has become obsolete, but in our case it worked well as an alternative to file-based collaboration, where all team members work separately and there is a collation process at the end, while retaining a 'slow' pace as described by Caraher (2015).

The entire content of GQBWiki is in Italian: our team is not international and it would be unnatural to write our documentation in English. Italian is also known by many scholars of the ancient world. In fact, it could be argued that Greek is the language that is actually missing from GQBWiki, because the research project is taking place in Greece, under control by the Greek authorities. The prevalence of the English language on the Web and in academic literature is undebatable (especially in a paper written in English), but its advantages over multilinguism are less clear. The Wikimedia movement

has taken a clear practical stance in favour of multilinguism, with hundreds of Wikipedias in minority languages. Since we adopted the software platform of the Wikimedia movement, it seems appropriate to reflect on this global issue, not just from our privileged point of view and with the concern of visibility and academic value of our work, but also from the perspective of making knowledge available to as many people as possible.

The wiki home page guides both the casual reader and the regular contributors to the various sections of the website, acting as a table of contents for the various areas of interest and the levels of detail. At a glance, the contents range from excavation data to interpretive texts, providing a necessary companion to the final GQB publication, the tone of which will be narrative and holistic rather than enumerative. There are certainly parallels with other similar systems that were built in the same years, like the one developed for Villa Magna (developed by Andrew Dufton and Elizabeth Fentress), both in terms of types of content and of technical solutions.

When comparing GQBWiki and our use of MediaWiki to other 'archaeological information systems', one aspect that should be immediately clear is that we are not proposing wiki systems as the best solution for any archaeology research project, particularly from a technical standpoint. There are limitations that make GQBWiki imperfect, if not in principle at least in practice, and it is important to recognise these limitations. The most substantial limitation is with spatial data (context plans, sections, etc.) that MediaWiki is completely unable to support natively. Looking at this seemingly unacceptable issue from a broader perspective, we can observe that in 'traditional' site archives and archaeological information systems, alphanumeric, graphic and spatial data are managed in the same platform, while interpretation and publication are left on their own. On the other hand, GQBWiki is missing spatial data that is managed through separate software tools, but all other content is part of the same platform. Spatial data consists mainly of context plans that are rendered as static raster images and uploaded to the wiki in batch, and then dynamically loaded in the relevant pages based on the semantic features outlined below.

A wiki page is a free form web page, where a lightweight markup is used instead of HTML to ease authoring. Therefore, any schematisation (such as the requirement that all stratigraphic context records have the same appearance and minimum information) is obtained by means of discipline and templates, not unlike Wikipedia content. There can be as many templates as needed in a wiki page, for formatting parts of content in specific ways (e.g. the well-known 'infobox' in the top right) or more complex tasks.

Wiki systems are by definition multi-user, both technically and socially. The net result is that GQBWiki is an incarnation of written multivocality, probably not of the same kind envisaged by Ian Hodder, but nevertheless stimulating, especially when we consider that all users/members have access to the same total amount of information, both for reading and editing. Users can edit any page as they see fit, fixing small typos or changing the functional interpretation of a deposit. The reality is less radical than what it may seem, though. Each wiki page preserves its own 'history' of edits, providing an overview of who has been adding (or removing) content, when, etc., as anyone familiar with Wikipedia will find normal (we hope that members of an academic audience have a basic understanding of these tools). This allows relevant meta-information to be immediately available, such as the last date when a page was updated – and therefore whether the content is possibly outdated. In a general sense, the page history provides an overview of 'who contributed what' with respect to the page under examination.

The big step we are taking in 2015 is opening GQBWiki to the public, even before there is a print edition of the research project, under a Creative Commons – Attribution – Share-Alike license (again, the same used by Wikipedia). By doing this, we hope to provide a useful digital resource for those working in Mediterranean archaeology, for example by sharing digital images of finds from dated contexts (a very common quest in this field of studies). At the same time making GQBWiki open is a straightforward way to elicit and stimulate feedback about our archive as a whole. Unfortunately, for the moment the ability to edit content is limited to team members and registered users (mainly due to the need to avoid spam): this is perhaps not even considered in most cases when similar digital resources go online, but it seems worth pointing out that it would be very interesting for anyone and especially other scholars to be able to comment on pages and even provide alternate interpretations for site features and finds. We are keen on registering new users on demand, but not with an effortless registration procedure that is standard for modern websites. However, this limitation is in our available time, not in the software.

2 Dealing with limitations, exploring possibilities

A quick numerical summary of GQBWiki shows that, at the time of writing, there are 2089 pages, with 16190 internal links and 27618 single edits. Pages range from stratigraphic units to find records, but it is journal entries that play a central role in the navigation path, rather than the useful but confusing list of stratigraphic units. There are also pages about team members, both as a means of collective memory and as a kind of meta-documentation. GQBWiki contains data about who excavated a certain stratigraphic unit, so in a sense we have become part of the data we create, and made it explicit. There is a category of pages devoted to bibliographic references, usually with extensive notes linking evidence from other sites and regions to GQB and, as noted above, 'incubators' for ideas and written content that will be included in the print publication. Internal links are certainly one of the main strengths of wikis, and GQBWiki makes no exception: looking at the broad categories outlined above, it is important to point out that there is no restriction to links, and any page can point to any number of other pages, regardless of their 'category'.

The consequence of the 'flat nature' of wiki is that in several cases, the content ends up being very raw, not just in a technical sense of 'raw data', but also in terms of human readability and usability: if, for example, on a certain day the archaeologist did not feel like writing more than one sentence in their journal, that will be the content for that day – there are minimum requirements that are directly derived from those of paper recording sheets, but since our methodological toolbox leaned towards using multimedia, the amount of mandatory data has been reduced (Zanini and Costa 2006). The structure of a wiki is only created by adding content and links. Having no predefined structure is stimulating on an intellectual level, because every bit of information has the same theoretical

importance within the documentation system and there is room for both data and discussion of uncertainty, but in practice we need to create lists of pages, entry points and navigation paths that will guide both contributors and readers, keeping in mind that MediaWiki has a very good internal search engine, and that is usually the quickest and most effective way of finding a specific page. Having no separation between structure and data also means that both can be changed by editing wiki pages, and that this can be done at any moment. Following in the steps of Wikipedia, structured information in GQBWiki is stored in lists and 'infobox' templates. If we decide to record a new piece of information in a page, or a category of pages, there is no underlying structure separate from the frontend 'Edit this page' button. Another significant enabler is that MediaWiki markup encourages the kind of copy-and-paste editing made of trial and error (edit, save, review, edit again) that was so beneficial to the early development of the Web in the 1990s.

Again, great advantages come together with limitations: despite being based on a relational database (MySQL), MediaWiki is not a database and there is no native support for retrieving structured information using SQL-like queries. After an initial period of confidence in this 'dictatorship of the unstructured', it became clear that it was impractical to be left without the capability of doing structured queries on our knowledge base. At the same time, the amount of information we already had in place was substantial, and team members were pleased with the general functionality of the wiki, despite a slow learning process. Using Semantic MediaWiki, an extension to the base software package, we added a 'thin ontology' layer to GQBWiki, not with the aim of building a Semantic Web resource, but as the most convenient way of adding typical 'relational' functionality into our wiki. So, we could add dynamic content blocks like 'a gallery of images of the context at the bottom of each context page'. In practice, this works by turning internal wikilinks into 'typed' links: an image page is linked to the page of the item it depicts, conveying both the link and the relationship between these two pages; a page about a stratigraphic unit is linked to another stratigraphic unit by expressing the type of stratigraphic relationship between the two (following the Italian standard of highly descriptive 'physical relationship' as opposed to the British/MoLAS 'earlier than/later than' standard). At a basic level, Semantic MediaWiki usage is equivalent with the creation of a custom ontology that is only valid for the wiki in use, based on properties, but there is a possibility of 'mapping' the internal properties to universal URI-based properties. In the example of the image-item link, the 'depicts' relation becomes a local mirror of the equivalent FOAF property, where FOAF is the 'Friend of a Friend' ontology, one of the earliest and most widespread Semantic Web vocabularies in use. This makes for another case of serendipity: we started using a tool that worked natively on the web, before it was widely acknowledged that it would have been the only sensible choice in just a few years. GQBWiki had unique, clean URLs for every excavation context and find, since the very beginning, even though it was only in more recent years that we understood how this represented a possibility for doing other things, such as linked open data. The idea that external vocabularies (such as Nomisma.org for coins) can be used to link content from GQB to other online archives and catalogues is based on the assumption of an 'open world' of information where there are both internal (wiki)links and external links in a continuum.

3 Collaborative authorship and attribution

Apart from the technical aspects discussed above, there is a second set of problems that are of equal interest and touch on the intrinsic difference of wiki authorship from traditional publication, again from a standpoint where GQBWiki is first of all the recording of a research process, and the archaeological excavation is only one part of that process, as is the digital archive. The material wiki practice of creating content confronts us with problems such as: how do we manage contributions ranging from simple digitisation and data entry of analog records to fully digital stratigraphic data?

In a traditional setting, the path from content creation to publication is more or less linear, from the bottom up, with checks for consistency at each step. With thousands of pages, each one accessible separately, the need for a solid review is even stronger, but the difficulty is in the systematic application of review procedures in a way that is both efficient and quick, otherwise new contributions will stagnate. Therefore, content review happens on an opportunistic basis in GQBWiki, and it is not enforced. In general, the internal review process has worked well for us, but some content is still outdated or missing, and a complete external peer review seems unlikely and we do not expect a substantial amount of feedback even after opening the wiki, as most potential contributors would have their own archives to curate.

Another issue we think we are dealing with is attribution for all the digital work done by supervisors and undergraduate students alike. The approach seen in GQBWiki is taking inspiration from initiatives like Fair Cite (2012), which tackles the problem of 'how best to cite a web-based collaborative project developed in the humanities' and whose names should be included in the citation. At the bottom of each wiki page, a list of all contributors has links to each user page and the suggested citation for a single page contains the URL of a special visualisation showing that list. Furthermore, 'bot' users like the prolific GQBot (controlled by the pywikibot software) give us a chance to reflect upon the contribution of machines to our work, not only as mere tools, but as executors of instructions that we only prepare, for repetitive work like batch uploading of images or importing from databases. Our work is collaborative in this sense, too.

4 Conclusions

After ten years working with GQBWiki we are convinced that the benefits exceed the disadvantages, and that making this body of knowledge open will further increase its value for the wider archaeological community.

A decade could seem a long time span, since most digital works can easily become obsolete even in less time: the truth is that we are collectively used to rapid decay cycles of our digital archives and publications, while traditional paper-based publication has stood the test of time. When, in an academic context, we put our data and studies online, it usually means that we want them to be accessible and we want to ensure them a long life. Being on the Web does not make data automatically linked and open, but as described above GQBWiki is incrementally going in that direction, finding common ground with other existing initiatives in the field of ceramic studies (Gruber and Smith 2015), numismatics (Gruber *et al*. 2014)

and ancient world studies in general (Elliott and Gillies 2009) with a very practical, URI-focused stance, and we hope that GQBWiki URIs will make appearance in linked open data graphs. That said, we also think that a more pronounced focus on the human components of any technological platform is needed, and Web 2.0 is no different in this respect (Shanks and Whitmore 2012). In our experience, a wiki needs to be actively used in order to have a chance to survive, and having a long-term archival of wiki content or any other archaeological data in a 'freezed' form is increasingly unsatisfactory, since the discoverability of such content is not getting better. Other, separate wikis that we started for other research projects are unfortunately not as thriving as the one described in this paper.

So far, GQBWiki is the virtual workplace of our research team: consulted and updated by users all the time from many places in Europe, with huge peaks of activity reached during the excavation campaigns (Carabia 2013). The availability of excavation data, interpretive texts, diaries, pictures and so on, all on the same platform concurrently and without any hierarchical limitation, has been a transformative environment for our work.

We believe that this approach is fruitful at the research team level and can be adopted on a wider basis. In an ideal situation, new studies about specific aspects of archaeological interest (for example, the type of artisanal activity recognised in 8th-century contexts from Byzantine Gortyn) would result not only in a specialist, peer-reviewed publication, but also in the updating of a range of 'wiki pages' about Byzantine craftsmanship, the history of Crete, or the work of Italian archaeologists abroad.

At a global scale, Wikipedia represents the main way of accessing the knowledge landscape for a majority of Internet users. Archaeology is well represented on Wikipedia but expert contributions are scarce, driven by the lack of incentive for academics to contribute and the rarity of collaboration-driven publication among archaeologists (Hadley 2013). For very general topics, Wikipedia is recognised as the right platform and there are known patterns for contributing content, debating contrasting views, accommodating for different types of source material and so on. It is less clear whether more specialist content (for example the chronology of a very specific type of ceramic production – even a minor one – or the calibrated radiocarbon date for an occupation sub-phase in an otherwise settlement) can fit in the Wikipedia notability guidelines. As we have shown, the tools and some of the good practice to develop long-term collaborative platforms are already in place.

Should we start working in a collaborative and incremental fashion, rather than starting from scratch at each new study?

Bibliography

Carabia, A. 2013. Wiki=beta: il modus vivendi di un sistema per documentare la ricerca. *Archeologia e Calcolatori Supp*. 4: 209-13. Firenze, All'Insegna del Giglio.

Caraher, W. 2015. *Industrial Archaeology and Student Resistance*. [Online] Available from: https://mediterraneanworld.wordpress.com/2015/06/26/industrial-archaeology-and-student-resistance/ [accessed: 8 July 2015]

Elliott, T., Gillies, S. 2009. Digital Geography and Classics, «Digital Humanities Quarterly» 3, 1. *Changing the Center of Gravity: Transforming Classical Studies Through Cyberinfrastructure*. [Online] Available from: http://www.digitalhumanities.org/dhq/vol/3/1/000031/000031.html [accessed: 13 November 2015].

Fair Cite 2012. *Current Citation Practices in Academia*. [Online] Available from: https://faircite.wordpress.com/ [accessed: 31 March 2015].

Gruber, E., Heath, S., Meadows, A., Pett, D., Tolle, K., Wigg-Wolf, D. 2014. Semantic Web Technologies Applied to Numismatic Collections. In P. Verhagen (ed.), *Archaeology in the Digital Era: Papers from the 40th Annual Conference of Computer Applications and Quantitative Methods in Archaeology (CAA), Southampton, 26-29 March 2012*: 264-74. Amsterdam, Amsterdam University Press.

Gruber, E., Smith, T. J. 2015. Linked Open Greek Pottery. In F. Giligny, F. Djindjian, L. Costa, P. Moscati, S. Robert (eds.), *Proceedings of the 42nd Annual Conference on Computer Applications and Quantitative Methods in Archaeology. CAA 2014 - 21st Century Archaeology*: 205-14. Oxford, Archaeopress.

Hadley, P. 2013. *WikiProject Archaeology*. [Online] Available from: http://pathadley.net/projects/wikiproject-archaeology/ [accessed: 8 July 2015].

Mediawiki Contributors 2015. *MediaWiki*. [Online] Available from: https://www.mediawiki.org/ [accessed: 8 July 2015]

Semantic Mediawiki Contributors 2015. *Semantic MediaWiki*. [Online] Available from: http://www.semantic-mediawiki.org/ [accessed: 8 July 2015].

Shanks, M., Whitmore, C. 2012. Archaeology 2.0? Review of Archaeology 2.0: New Approaches to Communication and Collaboration. *Internet Archaeology* 32. [Online] Available from: http://dx.doi.org/10.11141/ia.32.7

Zanini, E., Costa, S. 2006. Organizzare il processo conoscitivo nell'indagine archeologica: riflessioni metodologiche ed esperimenti digitali. *Archeologia e Calcolatori* 17: 241-64.

Zanini, E., Costa, S. 2009. Sharing knowledge in archaeology: looking forward the next decade(s). In M. Tsipopoulou (ed.), *Digital Heritage in the new knowledge environment*: 69-72. Athens, Hellenic Ministry of Culture.