Archeologia e Calcolatori
35.2, 2024, 107-118
doi 10.19282/ac.35.2.2024.12

USING KART AND GITHUB FOR VERSIONING AND COLLABORATING WITH SPATIAL DATA IN ARCHAEOLOGICAL RESEARCH

1. Introduction: Open Science, Archaeology, Git and Version Control

One of the many aims of Open Science is to enhance the transparency behind the process of data creation, data manipulation, etc., in order to foster reproducibility and openness. In fact, an Open Science approach to data cannot be limited to just a final product, but should provide means to access and inspect the full creative and methodological process behind data creation (Strupler, Wilkinson 2017). Distributed Version Control Systems (DVCS), and particularly Git, are one of many tools that has potential to enhance Open Science practices (Ram 2013). It has been suggested that Git can help in data management (Marwick 2017), lead to greater accountability and better documentation (Kansa 2012), and provide more opportunity for feedback and collaboration at different stages of archaeological work (Kansa *et al.* 2014). In its simpler form, DVCS allow to inspect snapshots of a file at different stages, thus making the whole creation process fully transparent, but also allowing to roll back changes to a specific snapshot.

Most importantly, DVCS, when coupled with remote repositories, allow for a more efficient collaboration between peers. While these benefits have been highlighted for some time, DVCS are still not widely adopted by archaeologists (Strupler, Wilkinson 2017; Karoune, Plomp 2022), being mostly employed in code-based approaches to archaeological data (Marwick, Birch 2018; Schmidt, Marwick 2020; Batist, Roe 2021, 2023), and, in some cases, for entire archaeological workflows, from fieldwork to publication (Strupler, Wilkinson 2017).

However, Git and DVCS are not without pitfalls. Apart from the steep learning curve and the slow adoption, a well-known limitation of Git is the versioning of binary files. Most of the archaeological workflow is still reliant on binary files, being either word processors, spreadsheets, geospatial data or raster images (Schmidt, Marwick 2020). The lack of a way to efficiently track changes in these point-and-click software means that much of the data creation and the methods are hidden, unless reported into a not-always-ideal final report (Kansa 2012).

2. Version Control for geospatial data

For text files and tabular data the 'issue' of adopting DVCS is overcome by employing plain text formats like Markdown, TXT, CSV. However, GIS is another area in which the graphical nature of the software obscures the research process, and data cleaning, data restructuring and so on remain hidden. Advocating for a code-only solution is not feasible, and intermediate tools should exist to be able to bridge point-and-click with code-based approaches (Strupler, Wilkinson 2017).

Kart (https://kartproject.org/) is a DVCS for geospatial and tabular data, a cross-platform FOSS software launched in 2020 by the company Koordinates. Kart works with many different formats, such as Geopackages, PostGIS, MySQL, and support the most common geospatial data types. It is built on Git, and provides the same Git functionalities but for geospatial data, meaning that a versioned history of datasets, both locally and remotely is available, enabling collaboration and changes tracking. Kart is a command line (CLI) program that bundles both Git and Git-LFS (for large files, e.g. rasters), so having Git installed on the system is not a prerequisite. As Git, Karts allows to track granular changes at the layer level (row and cells), making it possible to clearly inspect commits for spatial dataset and retrieve information in a meaningful way.

Kart also offers a QGIS plugin (https://plugins.qgis.org/plugins/kart/), which provides a convenient interface for common Git operations, and, most importantly, a visual tool for tracking commit history and visualize changes on a 2D map. More importantly, just like Git, one can host the versioned data on a remote repository such as GitHub for a more efficient collaboration.

3. The Assyrian Governance Project

The 'Governance Policies and Political Landscapes in the Southern Levant under the Neo-Assyrian Empire' is a 2-year project based at the University of Turin, founded by the Gerda Henkel foundation (https://www.dipstudistorici.unito.it/do/progetti.pl/Show?_id=0sk8)[1]. The project's goals are to understand Neo-Assyrian imperial strategies, and their effects on population and settlement patterns in the Southern Levant region during the Iron Age. The study area provides an excellent case study to investigate long-term changes of archaeological and political landscapes, since across its history it incorporated multiple socio-political entities, from local kingdoms to supra-regional empires. Moreover, the area is characterized by an extensive amount of high-quality archaeological data, providing enough information for a multi-temporal and multi-scalar analysis.

[1] The project is funded by the Gerda Henkel Foundation. Additional funding has been provided by the Rita Levi Montalcini Grant for the project 'The Empire Strikes Back: The Geography of Governance Strategies in the Assyrian Empire'.

The Southern Levant political landscape and the effect of the Assyrian domination over its provinces and client states have been a matter of debate among scholars. On the two extremes, some scholars highlighted how the so-called *Pax Assyriaca* was instrumental to economic prosperity, settlement expansion and adoption of Assyrian values and customs (Gitin 1995; Finkelstein, Singer-Avitz 2001; Stern 2001; Frahm 2006; Fales 2008), while others claimed that Assyrian conquest resulted in destruction and devastation of annexed provinces, whereas independent regional states flourished (Stager 1996; Na'aman 2003; Avraham, Weiss 2005; Faust 2021). However, it has been widely demonstrated how a single model is not fit to understand Assyrian governance strategies, which have been shown to be nuanced and adaptable depending on the history of the region (Macginnis 2016; Morandi Bonacossi 2018; Düring 2018, 2020; Tyson, Herrmann 2018; Parker 2020). Moreover, these phenomena cannot be disconnected from the larger framework of Iron Age settlement pattern changes (Wilkinson 2003).

Regarding archaeological data, the project is gathering spatial, quantitative, and qualitative data from published archaeological surveys and excavations, and online resources. In particular, datasets available for the project are:

– The Samaria Survey (Finkelstein *et al.* 1997);
– The Manasseh Hill Surveys (Zertal 2004, 2007; Zertal, Mirkam 2016; Zertal, Bar 2017, 2019; Bar, Zertal 2021, 2022);
– The Archaeological Survey of Israel online database (https://survey.antiquities.org.il/), and excavation reports from the «Hadashot Arkheologiyot Journal» (https://www.hadashot-esi.org.il/default_eng.aspx);
– settlement patterns and demographic studies (Broshi, Gophna 1984, 1986; Gophna, Portugali 1988).

In the case of published survey, the process of gathering spatial information included manual georeferencing of survey maps and digitisation of archaeological site points. In case the georeferencing process resulted in high inaccuracies, a manual registration of sites coordinates into a CSV file was carried out. This file has later been imported in QGIS and integrated in the project database, consisting of a single geopackage file. For online databases, site information was stored in CSV format and then imported in the database. Most of the work is carried out in QGIS, with sites information stored in the layer attribute table. While the database is still under refinement, currently it stores around five thousands sites and more than twelve thousands occupational phases (see also Palmisano *et al.* 2019) (Fig. 1).

The project is organized around a workflow aimed to be as transparent as possible, not only in the final results, but also in the different stages of data creation, data cleaning, analyses, etc. The project lives on GitHub, with a homepage (https://github.com/UnitoAssyrianGovernance) and different

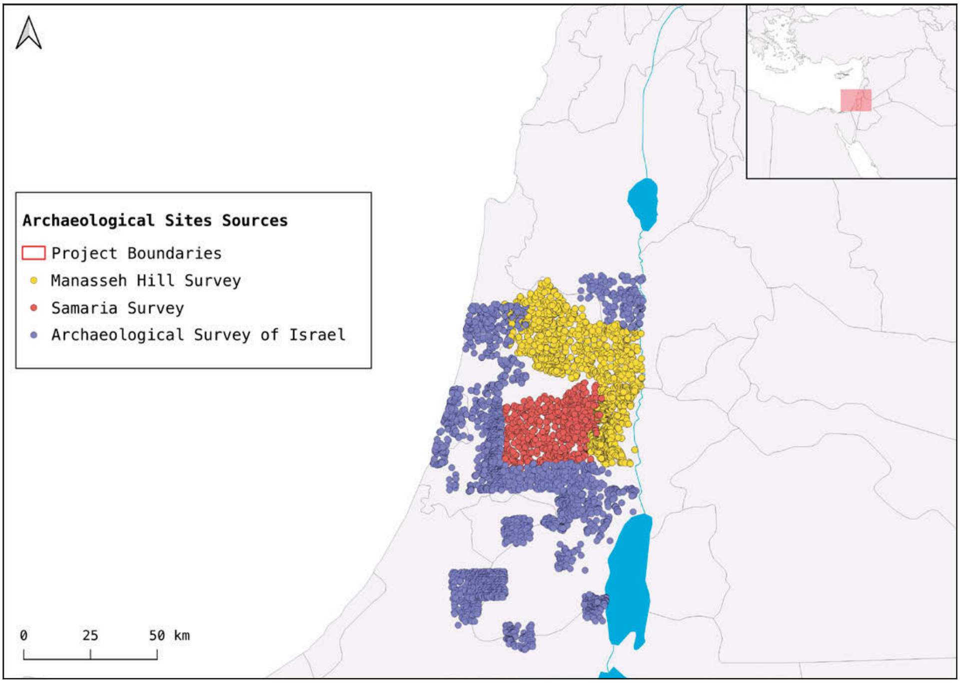

Fig. 1 – Project extent and archaeological sites digitized from different sources.

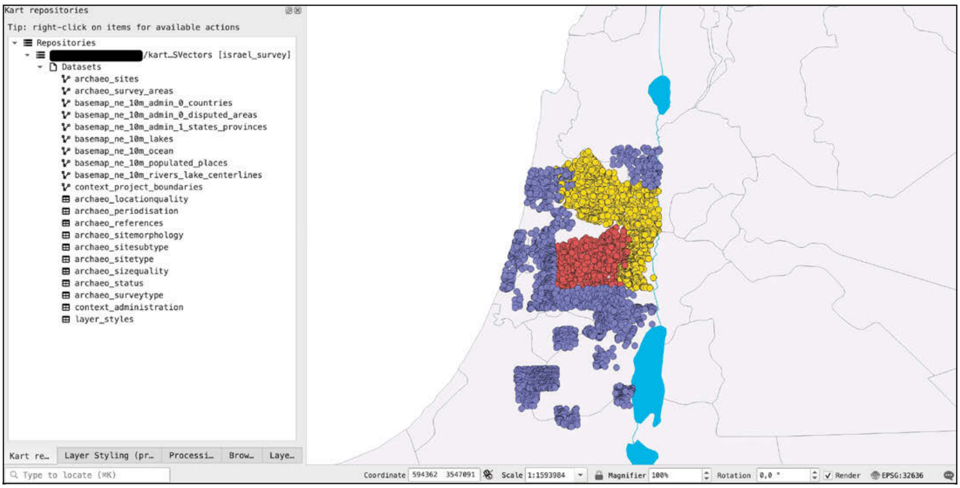

Fig. 2 – Kart QGIS plugin main interface with the list of available layers present in the database.

repositories for any project activity: data, analyses, publications/talks. The use of multiple repositories will allow higher flexibility and granular control in collaborating, accessing, licensing, publishing, and sharing data. Project activities are also tracked leveraging the GitHub 'issues' functionality available on any repositories, and collected into a kanban board (https://github.com/orgs/UnitoAssyrianGovernance/projects/5). The project also offers a public wiki (https://github.com/UnitoAssyrianGovernance/.github/wiki) which hosts most of the project documentation, updated as this proceed. The wiki also gather methods and conventions used inside the project, but offers also a guide on how to use the dataset and how to use Kart, which is the basis of the workflow attached to the paper.

Kart is inserted into the aforementioned workflow to provide transparency and openness to the process of working with geospatial data. In particular, the main aims are versioning vector data, collaborating with project members remotely, and to have a public record of changes made to the spatial database layers (Fig. 2).

3.1 *Kart for remote collaboration*

The project adopts a very common workflow for collaboration, based on Git features such as branching and merging, with the following structure:

– main branch: considered the final copy, where only completed data are merged;
– feature branch: these are the branches worked on daily, with as many branch as many collaborators as possible inside the project.

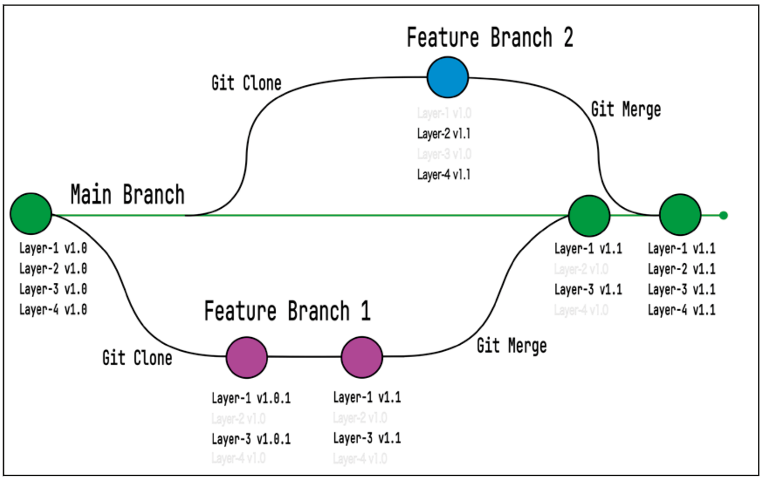

Fig. 3 – Simplified representation of Kart workflow.

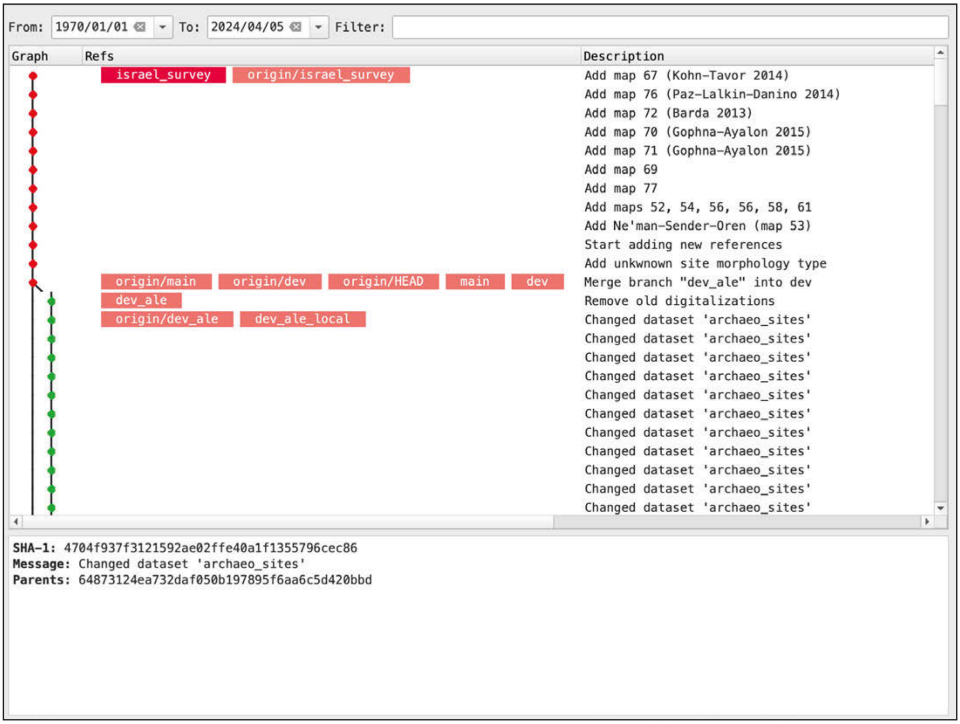

Fig. 4 – Example of Kart QGIS plugin visual commit log.

In this workflow, each project member works on a separate feature branch, edits the layers and pushes the branch to the corresponding GitHub repository when needed. When a specific task has been completed (e.g. the digitization of an entire survey catalog), the feature branch is merged into the main branch (Fig. 3).

3.2 *Kart for version control of spatial data*

The advantage of version control is that each project member is able to track the edits made in GIS, and to easily revert back if mistakes were made. Data cleaning, refactoring of tabular structure, and other changes made during the project can be easily highlighted by looking at the commit history (Fig. 4). This in turn helps to highlight issues, presents the reasoning behind methodological choices and provides a reference beyond the simple final product. In the project framework, Kart was tested on two MacOS machines, and on an Ubuntu-based Linux machine. Since the workflow is rather simple, there were not many issues in the collaboration process. The only complication presented itself when merging data on the same layer from two different branches, which caused an overlapping conflict of the unique primary IDs generated by the geopackage. To solve the issue, one has to resort to the command line, which offer to automatically renumber one of the two conflicting series of IDs[2].

4. Discussion and conclusions

One of the long-standing issues of collaborating on any GIS project is how to handle either simultaneous edits on the same data or how sharing the same data among colleagues after each edit. Using a DVCS based on different working branches is one of the ideal solutions to overcome this problem. Commit messages are also self-explanatory and can clearly define what changes have been made to the dataset even before a collaborator inspects them. Kart allows multiple people to efficiently work on the same dataset without the need for back-and-forth emails, drive uploads, or personal communications for updates on the dataset. Rolling back through commit history to correct mistakes is also generally easier than trying to recover something made during an undocumented change. The graphical user interface provided by the QGIS plugin also allows for a convenient visual change inspection (Fig. 5). This plugin, while still lacking some functionalities of the CLI, if further developed, could effectively bridge the gap between code-based approaches and mouse-driven software, since ideally the use of the terminal could be completely avoided.

Another advantage is related to the way Kart stores the spatial dataset on a remote repository. Data are in fact broken down in a series of SQL-like tables instead of a single file. While this could make things harder to read, it also makes it harder to access spatial information about the sites. This might be relevant in the ongoing discussion about site stewardship related to public data sharing (e.g. Cohen *et al.* 2020; Fisher *et al.* 2021). One of the authors also employed Kart during fieldwork, specifically for versioning daily data and provided a record of data cleaning and processing after collection. This aspect is important, as very few attempts have been made to integrate DVCS in fieldwork activities, highlighting how one missing tool was a way to keep track of GIS/Rasters data (Strupler, Wilkinson 2017), which is exactly the gap that Kart can fill.

However, one must recognize that Kart is still not a widely accessible tool and it presents a steep learning curve (only partially mitigated by the graphical plugin). While being built on Git means that the workflow will be

[2] A tutorial and a dataset to drive step by step any practitioner interested in learning how to use Kart is available at https://unitoassyriangovernance.github.io/kart4arch/ and have also been deposited in Zenodo: https://zenodo.org/records/10962416.

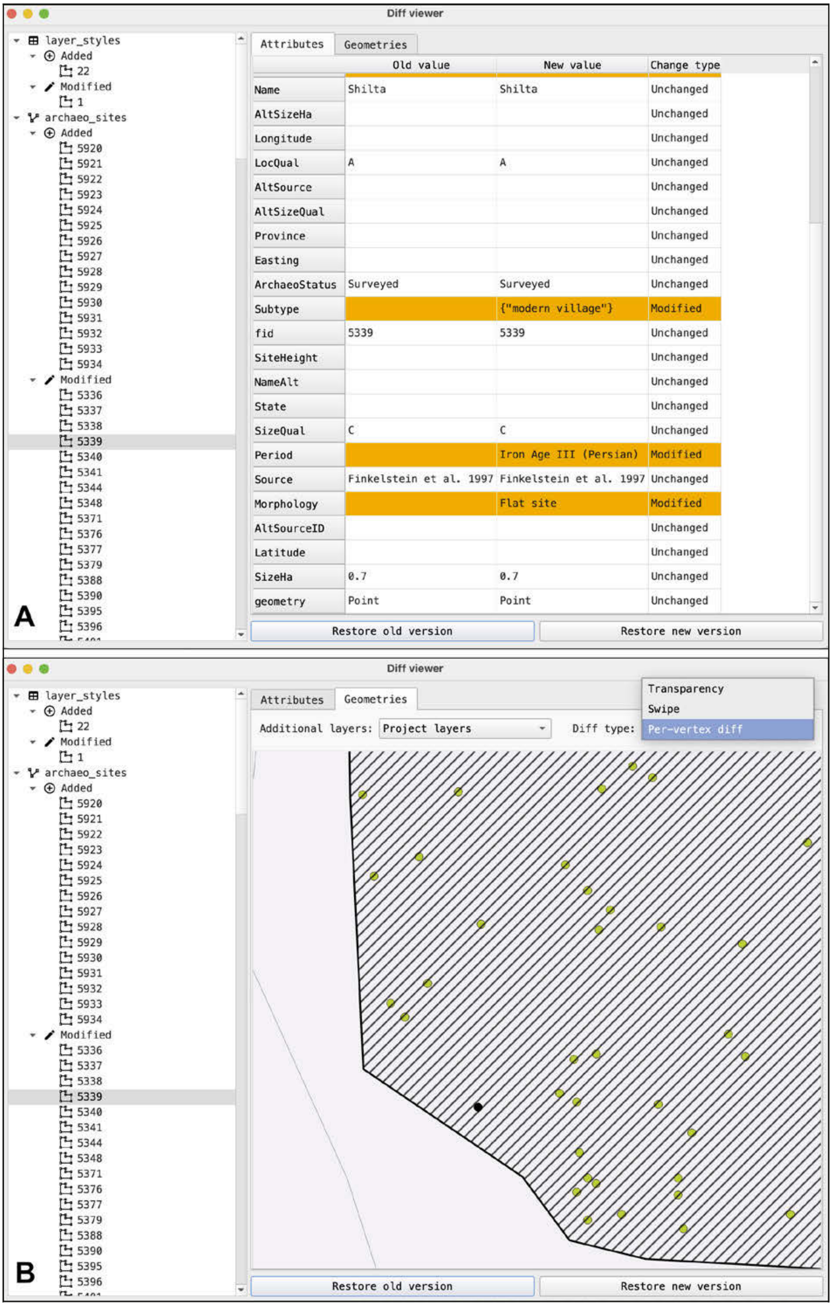

Fig. 5 – Example of Kart QGIS plugin tabular (A) and visual (B) diff.

familiar for Git users, most archaeologists are likely not confident enough with coding tools and terminal applications, and training and resources will be needed if a wider adoption is hoped for. Most of the features are also useful only if a specific workflow is adopted, as incorrect or partial information might hamper the collaboration process and transparency. This workflow is suitable for desk-based work, but it might be considered slower or less efficient in situations such as fieldwork, when time constraints are higher and Internet connection is not always available (although internet access is not a requirement). Another limitation is the lack of easier methods for solving merging conflicts (see above), which still require manual use of the CLI (although according the documentation tools to smooth this process are in development).

However, while being a relatively young and small tool, Kart can fit well in the Open Science practices applied to archaeology. Kart is still under development, and while mature enough in our opinion, it might still face substantial changes before reaching a stable state. Nonetheless, the authors think that it can provide a valuable addition to the current archaeological workflow regarding spatial data management, with our project being a main testing venue for its applicability.

Andrea Titolo*, Alessio Palmisano
Dipartimento di Studi Storici
Università degli Studi di Torino
andrea.titolo@unito.it, alessio.palmisano@unito.it

*Corresponding Author

REFERENCES

Avraham F., Weiss E. 2005, *Judah, Philistia, and the Mediterranean World: Reconstructing the economic system of the seventh century B.C.E.*, «Bulletin of the American Schools of Oriental Research», 338, 71-92 (http://www.jstor.org/stable/25066890).

Bar S., Zertal A. 2021, *The Manasseh Hill Country Survey*, *6. The Eastern Samaria Shoulder, from Nahal Tirzah (Wadi Far'ah) to Ma'ale Ephraim Junction*, Leiden, Brill (https://brill.com/display/title/60190).

Bar S., Zertal A. 2022, *The Manasseh Hill Country Survey*, *7. The South-Eastern Samaria Shoulder, from Wadi Rashash to Wadi 'Aujah*, Leiden, Brill (https://brill.com/display/title/62098).

Batist Z., Roe J. 2021, *Open Archaeology: A survey of collaborative software engineering in archaeological research*, Paper presented at *Computer Applications & Quantitative Methods in Archaeology 2021, Limassol (Virtual)*.

Batist Z., Roe J. 2023, *Open-archaeo: A resource for documenting archaeological software development practices*, «Journal of Open Archaeology Data», 11 (https://doi.org/10.5334/joad.111).

Broshi M., Gophna R. 1984, *The settlements and population of Palestine during the Early Bronze Age II-III*, «Bulletin of the American Schools of Oriental Research», 253, 41-53 (https://doi.org/10.2307/1356938).

Broshi M., Gophna R. 1986, *Middle Bronze Age II Palestine: Its settlements and population*, «Bulletin of the American Schools of Oriental Research», 261, 73-90 (https://doi.org/10.2307/1357066).

Cohen A., Klassen S., Evans D. 2020, *Ethics in archaeological Lidar*, «Journal of Computer Applications in Archaeology», 3, 1, 76-91 (https://doi.org/10.5334/jcaa.48).

Düring B.S. 2018, *Engineering empire: A provincial perspective on the Middle Assyrian empire*, in B.S. Düring, T.D. Stek (eds.), *The Archaeology of Imperial Landscapes: A Comparative Study of Empires in the Ancient Near East and Mediterranean World*, Cambridge, Cambridge University Press, 21-47.

Düring B.S. 2020, *The Imperialisation of Assyria: An Archaeological Approach*, Cambridge University Press (https://doi.org/10.1017/9781108778701).

Fales F.M. 2008, *Canals in the Neo-Assyrian rural landscape: A view from the Habur and middle Euphrates*, in H. Kühne (ed.), *Umwelt und Subsistenz der assyrischen Stadt Dur-Katlimmu am Unteren Habur (BATSH 8)*, Wiesbaden, Harrassowitz Verlag, 181-187.

Faust A. 2021, *The Neo-Assyrian Empire in the Southwest: Imperial Domination and its Consequences*, Oxford, Oxford University Press (https://global.oup.com/academic/product/the-neo-assyrian-empire-in-the-southwest-9780198841630).

Finkelstein I., Lederman Z., Bunimovitz S. 1997, *Highlands of Many Cultures: The Southern Samaria Survey. The Sites*, Tel Aviv, Institute of Archaeology of Tel-Aviv University, Publications Section (https://cris.tau.ac.il/en/publications/highlands-of-many-cultures-the-southern-samaria-survey-the-sites).

Finkelstein I., Singer-Avitz L. 2001, *Ashdod revisited*, «Tel Aviv», 28, 2, 231-259 (https://doi.org/10.1179/tav.2001.2001.2.231).

Fisher M., Fradley M., Flohr P., Rouhani B., Simi F. 2021, *Ethical considerations for remote sensing and open data in relation to the endangered archaeology in the Middle East and North Africa project*, «Archaeological Prospection», 28, 3, 279-292 (https://doi.org/10.1002/arp.1816).

Frahm E. 2006, *Assyria and the South*, in S. Holloway (ed.), *Orientalism, Assyriology and the Bible*, Sheffield, Sheffield Phoenix Press, 74-94 (https://doi.org/10.1002/9781118325216.ch15).

Gitin S. 1995, *Tel Miqne-Ekron in the 7th century B.C.E.: The impact of economic innovation and foreign cultural influences on a Neo-Assyrian vassal city-state*, in S. Gitin (ed.), *Recent Excavations in Israel: A View to the West*, Dubuque, Kendall/Hunt Publ. (https://ixtheo.de/Record/1589006739).

Gophna R., Portugali J. 1988, *Settlement and demographic processes in Israel's coastal plain from the Chalcolithic to the Middle Bronze Age*, «Bulletin of the American Schools of Oriental Research», 269, 11-28 (https://doi.org/10.2307/1356947).

Kansa E. 2012, *Openness and archaeology's information ecosystem*, «World Archaeology», 44, 4, 498-520 (https://doi.org/10.1080/00438243.2012.737575).

Kansa E.C., Kansa S.W., Arbuckle B. 2014, *Publishing and pushing: Mixing models for communicating research data in archaeology*, «International Journal of Digital Curation», 9, 1, 57-70 (https://www.ijdc.net/index.php/ijdc/article/view/9.1.57/341).

Karoune E., Plomp E. 2022, *Removing Barriers to Reproducible Research in Archaeology* (https://doi.org/10.5281/ZENODO.7320029).

Macginnis J. 2016, *The archaeological exploration of the provinces of Assyria*, in J. Macginnis, D. Wicke, T. Greenfield (eds.), *The Provincial Archaeology of the Assyrian Empire*, Cambridge, McDonald Institute for Archaeological Research, 3-20.

Marwick B. 2017, *Computational reproducibility in archaeological research: Basic principles and a case study of their implementation*, «Journal of Archaeological Method and Theory», 24, 2, 424-450 (https://doi.org/10.1007/s10816-015-9272-9).

Marwick B., Birch S.E.P. 2018, *A standard for the scholarly citation of archaeological data as an incentive to data sharing*, «Advances in Archaeological Practice», 6, 2, 125-143 (https://doi.org/10.1017/aap.2018.3).

Morandi Bonacossi D. 2018, *The creation of the Assyrian heartland: New data from the 'Land behind Nineveh'*, in B.S. Düring, T.D. Stek (eds.), *The Archaeology of Imperial Landscapes: A Comparative Study of Empires in the Ancient Near East and Mediterranean World*, Cambridge, Cambridge University Press, 48-85.

Na'aman N. 2003, *Ekron under the Assyrian and Egyptian Empires*, «Bulletin of the American Schools of Oriental Research», 332, 81-91 (http://www.jstor.org/stable/1357809).

Palmisano A., Woodbridge J., Roberts C.N., Bevan A., Fyfe R., Shennan S., Cheddadi R., Greenberg R., Kaniewski D., Langgut D., Leroy S.A.G., Litt T., Miebach A. 2019, *Holocene landscape dynamics and long-term population trends in the Levant*, «Holocene», 29, 5, 708-727 (https://doi.org/10.1177/0959683619826642).

Parker B. 2020, *Re-modeling empire*, in A.L. Boozer, B.S. Düring, B.J. Parker (eds.), *Archaeologies of Empire: Local Participants and Imperial Trajectories*, Albuquerque, University of New Mexico Press, 324-350 (https://doi.org/10.1017/9781316995495.012).

Ram K. 2013, *Git can facilitate greater reproducibility and increased transparency in science*, «Source Code for Biology and Medicine», 8, 1, 1-8 (https://doi.org/10.1186/1751-0473-8-7).

Schmidt S.C., Marwick B. 2020, *Tool-driven revolutions in archaeological science*, «Journal of Computer Applications in Archaeology», 3, 1, 18-32 (https://doi.org/10.5334/jcaa.29).

Stager L. 1996, *Ashkelon and the archaeology of destruction: Kislev 604 BCE*, «Eretz-Israel», 25, 61-74 (https://doi.org/10.1515/9781575065717-018).

Stern E. 2001, *Archaeology of the Land of the Bible*, *II. The Assyrian, Babylonian and Persian Periods (732-332 B.C.E.)*, New York, Doubleday.

Strupler N., Wilkinson T.C. 2017, *Reproducibility in the field: Transparency, version control and collaboration on the project panormos survey*, «Open Archaeology», 3, 1, 279-304 (https://doi.org/10.1515/opar-2017-0019).

Tyson C.W., Herrmann V.R. (eds.) 2018, *Imperial Peripheries in the Neo-Assyrian Period*, Louisville, University Press of Colorado.

Wilkinson T.J. 2003, *Archaeological Landscapes of the Near East*, Tucson, University of Arizona Press.

Zertal A. 2004, *The Manasseh Hill Country Survey*, 1. *The Shechem Syncline*, Leiden, Brill (https://brill.com/display/title/11056).

Zertal A. 2007, *The Manasseh Hill Country Survey*, 2. *The Eastern Valleys and the Fringes of the Desert*, Leiden, Brill (https://brill.com/display/title/13258).

Zertal A., Bar S. 2017, *The Manasseh Hill Country Survey*, 4. *From Nahal Bezeq to the Sartaba*, Leiden, Brill (https://brill.com/display/title/34196).

Zertal A., Bar S. 2019, *The Manasseh Hill Country Survey*, 5. *The Middle Jordan Valley, from Wadi Fasael to Wadi 'Aujah*, Leiden, Brill (https://brill.com/display/title/54817).

Zertal A., Mirkam N. 2016, *The Manasseh Hill Country Survey*, 3. *From Nahal Iron to Nahal Shechem*, Leiden, Brill (https://brill.com/edcollbook/title/31740).

ABSTRACT

Distributed Version Control Systems are one of the common ways through which scientists collaborate and keep track of different versions of their work. Moreover, scientists, programmers, etc., have been using platforms such as GitHub to host and share their resources versioned through Git. While not as widely adopted as in other disciplines, Git has also been used in archaeological research. In fact, DVCS allow scholars to collaborate remotely and offer the transparency necessary to align with Open Science and reproducible research practices. However, Git is highly inefficient when versioning GIS data. Kart, described as «an open source DVCS for geospatial and tabular data built on git», is a software addressing the need for collaboration and finer incorporation of geospatial data, providing also an integration

with QGIS. Kart and code-hosting websites offers unique resources for archaeologists, from collaboration to more efficient workflows. In this paper, an example of how the authors are using Kart, QGIS, and GitHub in the project 'Governance Policies and Political Landscapes in the Southern Levant under the Neo-Assyrian Empire' will be presented. With this case study, the authors hope to provide a solution to the current gap in the workflow of documentation and collaboration among archaeologists using GIS.