

Using Git repos with webhooks in a common search portal for distributed collections of Byzantine seals

Jan Bigalke (<https://orcid.org/0000-0003-0721-9980>), Benedikte Löbbert (<https://orcid.org/0009-0002-6670-3692>), Claes Neufeind (Cologne Center for eHumanities, <https://orcid.org/0000-0002-9377-9492>)

The subject of this contribution is a distributed digital infrastructure for the consistent edition and presentation of Byzantine seals, which we are currently implementing as part of a cooperation project between the University of Cologne and the Sorbonne Université Paris, entitled “Unlocking the Hidden Value of Seals: New Methodologies for Historical Research in Byzantine Studies” (DigiByzSeal). Within the three-year project funded by the ANR and the DFG¹, the Cologne Center of eHumanities (CCeH) is responsible for the technical infrastructure.

Core of the project is the scientific edition of four collections of Byzantine seals, which comprise a total of around 4000 seals. Seals are coin-like objects, usually made of lead, with iconographic depictions, inscriptions and/or monograms on both sides (cf. Sopracasa et al. 2020). Each seal is an object of study and research in its own right, but only serial analysis makes it possible to fully exploit the potential of the artifact as a historical source. However, the surviving seals are scattered across various collections and can therefore only be used as historical sources to a limited extent.

In our presentation, we will focus on decentralized data storage using Git-repositories and how it is possible to make distributed collections centrally searchable without using a complex database system.

The DigiByzSeal project

The aim of the DigiByzSeal-project is to provide a comprehensive corpus for sigillographic research in Byzantine Studies. The lead seals are digitized using Reflectance Transformation Imaging (RTI). RTI is a digital imaging method that supports the legibility, analysis and presentation of seals (cf. Fischer and Makowski, 2017). We use a custom-build RTI dome², which captures 60 images, all with a slightly different lighting generated by 60 LEDs which are evenly distributed within the dome. The seals are photographed from their obverse and reverse sides. A software combines the 60 images to form a single image for each side. It is now possible to dynamically change

¹ Project number 469385434, see <https://gepris.dfg.de/gepris/projekt/469385434?language=en>

² Schaeben (2019)

the light source in a RTI viewer and to apply various filters. This allows for significantly better legibility of the seals, which cannot be achieved by simple photography or even 3D scans.

The seals are then marked up and annotated in XML using the SigiDoc standard (cf. Sopracasa/Filosa, 2020). SigiDoc is a subset of EpiDoc and provides XML-based and TEI-compliant encoding standards for the edition of Byzantine seals: it aims to provide a set of guidelines for the publication of Byzantine seals in digital form.

Decentralized data management and presentation

Each collection within the project has its own Git-repository in which the metadata of the individual seals is stored. Data that is relevant for all collections (e.g. authority files) is stored in a centrally managed repository. This allows each collection to have full control over its own seals, whereas the information that is relevant for all collections is fully accessible and up-to-date at all times.

For data presentation, we decided to use EFES (EpiDoc Front-End Services; see Bodard and Valentinova, 2020) and adapted it to the specific needs of SigiDoc. The EFES platform is based on Kiln³, an open-source multi-platform framework that allows users to create web pages based on XML without having to run complex database applications in the background. The XML data is simply inserted into the EFES file structure and transformed via a set of XSLT scripts. In addition, a Solr index is created, so that the seals can then be searched within the collection-specific EFES instance. Nevertheless, the creators of the XML files retain full control over their data, as the XML files are managed in their own Git-Repository. At the same time, the data is not limited to being presented via EFES, but can also be integrated into other, custom-build presentation platforms.

Central search portal

To allow for a search over all collections, data from the individual collections is cumulated in a central portal, which itself is also based on EFES⁴. With every new commit to one of the individual collections' Git-repository, the new files are automatically pulled into a central EFES instance. For this, we utilize the Jetty web server integrated in EFES by adding a JAVA web servlet to handle the Git-notifications via a defined endpoint. When this endpoint is addressed and a JSON-file is supplied, a git pull is performed for the repository defined in the JSON. Once the pull is successful, the Solr

³ <https://kiln.readthedocs.io/en/latest/>

⁴ <https://sigidoc.raketadesign.com/en/>

endpoint is called by the servlet for a full re-index, so that the new seals can be searched within the central EFES instance.

By using webhooks⁵, new collections can easily be added to the common search portal. The repository to be integrated is cloned into EFES and a webhook is defined in this repository, so that the EFES endpoint is addressed each time a new commit is pushed into the repository. The definition of webhooks is not limited, so that the Git-repository can also be used for other EFES instances or, if desired and available, for completely different infrastructures.

Storing the data in decentralized Git-Repositories has the advantage that the creators of the data retain full control over their collection, while at the same time the collection is automatically made searchable through the central portal page (as long as the collection data adheres to the SigiDoc schema). The strict separation of data and representation platform allows the data to be used in various scenarios. If the EFES instance is no longer accessible, the data can still be searched and retrieved through the central portal. The problems associated with such decentralized data storage, such as delayed availability, are circumvented by using the webhook functionality.

Bibliography:

Bodard, G. & Valentinova Yordanova, P. (2020). *Publication, testing and visualization with EFES: A tool for all stages of the EpiDoc XML editing process*. Studia Universitatis Babeş-Bolyai.

Fischer, F. and Makowski, S. (2017). *Digitalisierung von Siegeln mittels Reflectance Transformation Imaging (RTI)*. *Paginae historiae - Sborník Národního archivu*, 25/1. S. 137-141. Národní archiv. ISSN 1211-9768

Schaeben, Marcel (2019). *RTI-Dome am CCEH: Erster Testlauf erfolgreich absolviert*. <https://cceh.uni-koeln.de/legacy/2019/04/26/rti-dome-am-cceh-erster-testlauf-erfolgreich-absolviert/>

Sopracasa, A and Filosa, M. (2020): *Encoding Byzantine Seals: SigiDoc*. In: Marras, Cristina; Passarotti, Marco; Franzini, Greta; Litta, Eleonora [eds.]: "Atti del IX Convegno Annuale AIUCD. La svolta inevitabile: sfide e prospettive per l'Informatica Umanistica", Università Cattolica del Sacro Cuore, pp.240-245, 2020, DOI: 10.6092/unibo/amsacta/6316.

⁵ <https://docs.github.com/en/webhooks/about-webhooks>

Sopracasa, A., Filosa, M., & Stoyanova, S. (2020). *The Digital Enhancement of a Discipline. Byzantine Sigillography and Digital Humanities*. *magazén*, 1(1), 101-128.

Web Sources:

<https://kiln.readthedocs.io/en/latest/>

<https://docs.github.com/en/webhooks/about-webhooks>

<https://sigidoc.raketadesign.com/en/>