

Project title: **Digitale Geochemische Daten-Infrastruktur (DIGIS) für GEOROC 2.0**

1. Summary

The [GEOROC data base](#) serves the global geoscience research community to study the origin, evolution and chemical differentiation of Earth, its mantle, oceanic crust, volcanic islands, and the continents through geological time. GEOROC provides free access to published geochemical analyses of igneous rocks and minerals of the Earth's crust for major and trace element concentrations, radiogenic and mass-dependent isotope ratios, and geochronological data. In addition, extensive metadata are recorded with respect to geographic location, sample type and analytical methods. This provision of high-quality “long-tail” research data in geochemistry fostering “digital geochemistry” to become a major pillar in geochemical and Earth System Science. The top 2300 publications that cite GEOROC as their data source have collected a total of 96,218 citations. Without the DIGIS-project, the GEOROC data base would be no longer available to the research community.

The [DIGIS project](#) secures proper function of the GEOROC data base by realigned GEOROC with FAIR data principles by establishing full interoperability through vocabulary harmonization and provision of a modern IT infrastructure for GEOROC. A RESTful API has been implemented to allow for data and metadata exchange. Today, the GEOROC data base is being extended and updated regularly with new data and is now exported on a quarterly basis to the EarthChem Portal.

All coding and background information for data migration, API access, IT resources and developments are open source. GEOROC 2.0 will be available by the end of the first funding period (10/2024) via a new website with more flexible, customizable query options for global geochemical data.

We have achieved all our main goals and secured the sustainable preservation of the GEOROC database for the science community.

2. Progress Report

This report presents the achievements of the DIGIS team (WO 362/53-1, WI 3579/5-1, Ho 3987/58-1). As the result of developing and changing focus during the course of the project, there will be different emphasis on individual work packages.

Because the start of the project coincided with the peak of the COVID-19 pandemic, we faced difficulties related to administrative support for initiating the project. This has caused considerable delay in the hiring process of staff funded through the project. Travel and personal meeting with candidates were severely impeded. Further, shortage of qualified IT developers coinciding with the start of the NFDI-Initiative had severely depleted the number of potential candidates.

Because the start of the project was delayed, some work packages could not be fully completed. Moreover, as the project moved forward, new work packages were identified and considered as essential (data repository, data ingest, vocabularies, front end website, data “clean-up”). GEOROC was re-established as a key contributor of curated geochemical data (Fig. 1) and an integral part of the EarthChem Portal hosted at the Lamont-Doherty Earth Observatory. DIGIS was able to fully establish interoperability between GEOROC and the EarthChem portal's data bases. Newly curated GEOROC data are now exported incrementally to the EarthChem XML schema for integration into the EarthChem Portal. DIGIS is also a founding member of the international OneGeochemistry initiative (Lehnert et al, 2022) that seeks to establish international, community-agreed standards for geochemical data to facilitate interoperability between globally distributed data systems. Community engagement, e.g. at international conferences and workshops, is a key component of DIGIS (Klößing et al, 2024).

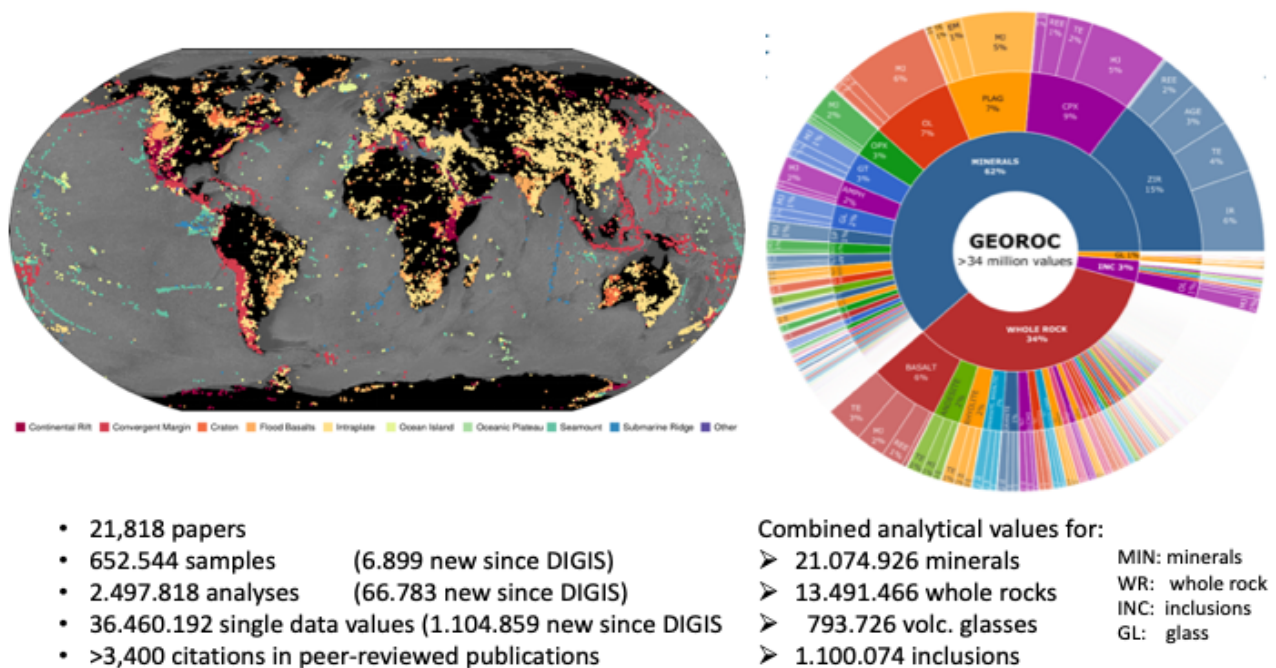


Fig. 1: Overview of data contained in the GEOROC data base

DIGIS has upgraded existing features of GEOROC as well as introduced new functionalities. The project follows open science principles, using free and open access resources and will be making all data and software products available. Further, the existing GEOROC data model has been mapped to the Observations Data Model Version 2 (ODM2, Horsburgh et al, 2016). ODM2 is designed to support interoperability between databases for observational data for physical objects and including time series data (Hsu et al, 2017). The GEOROC 2.0 data model is implemented in the open-source relational database management system PostgreSQL and available on <https://github.com/DIGIS-Georoc/DIGIS-Datamodel> and Zenodo (Klößing and Sturm, 2024, <https://doi.org/10.5281/zenodo.10468217>).

To improve data discovery, GEOROC 2.0 integrates additional persistent universal identifiers DOIs for datasets as well as journal publications, author ORCIDs and the International Generic Sample Number (IGSN). The GEOROC 2.0 database also contains internally controlled and updated lists of vocabularies to describe crucial metadata, such as sample types and locations, lithologies, mineral species, chemical analytes, etc. Wherever possible, vocabularies are based on, respectively harmonized with, published external vocabularies (for minerals: IMA/Mindat, ages: ICS, rocks (Mindat/BGS), units and analytical methods: Richard et al., 2023).

A RESTful API has been implemented to allow for data and metadata exchange. This API forms the basis for querying the new GEOROC 2.0 database (based on the ODM2 structure) and ensures that data can be more readily harvested and exchanged with other data services. A beta-version of the GEOROC 2.0 API, that complies with the GeoSciML data transfer standard for geological data, was released for testing in January 2023.

Already, GEOROC 2.0 supports a number of custom query routes for specialized data use cases. Additional API routes are in development during the remaining months of the DIGIS Project (until 10/2024) following community feedback. A new web interface is in progress for GEOROC 2.0 with more flexible, customizable query options and improved data visualization (currently under development, release by Oct 2024). We also provide detailed documentation tailored to different user groups from all backgrounds (e.g. PDF guides, tutorial videos, exemplary Jupyter Notebooks demonstrating programming solutions for data querying, processing and visualization).

GEOROC 2.0 data are now exported in XML format on a quarterly basis to the EarthChem Portal, where they can be queried jointly with data from five other geochemical databases.

The DIGIS concept for GEOROC 2.0 includes new pathways for data entry using end-to-end text and data mining and facilitating user submission of data and ingest into the GEOROC data base. To this end, the domain-specific GEOROC Data Repository was set up, currently hosted at GRO.data at the Göttingen campus. A MoU is under way with GFZ Data Services where the GEOROC data repository will be hosted in the future. All coding and background information for data migration, API access, IT resources and developments were done predominantly in UNIX, all as open source. A link to GIT-Hub with published documents is found at <https://api-test.georoc.eu/api/v1/docs/index.html>.

3. Publicly Accessible Project Results

As the project progressed and evolved, some work packages were significantly expanded, new work packages were added and some work packages needed to be postponed towards the end of the project and/or reserved for a second funding period.

3.1 “GEOROC 2.0 (Editorial workflow)” and “Web service interfaces of GEOROC 2.0”

Migration of the by then out-dated “vintage” GEOROC data base and its 20 years old IT environment turned out to be more laborious than anticipated and took considerably more time than initially planned. This was largely due to the retirement of IT staff who originally designed and maintained GEOROC and the fact, that the vintage GEOROC IT architecture and set-up was poorly documented. Data clean-up and removal of inconsistent vocabularies were required and abandoned data entries had to be fixed. Moreover, migration and creation of a dedicated data and hardware infrastructure required access to the MPI-data infrastructure at GWDG. The latter involved resolving not-anticipated and time-consuming administrative problems related to data protection issues that were eventually cleared with essential support from J. Nieschulze of the Research and Transfer Dept, Göttingen University.

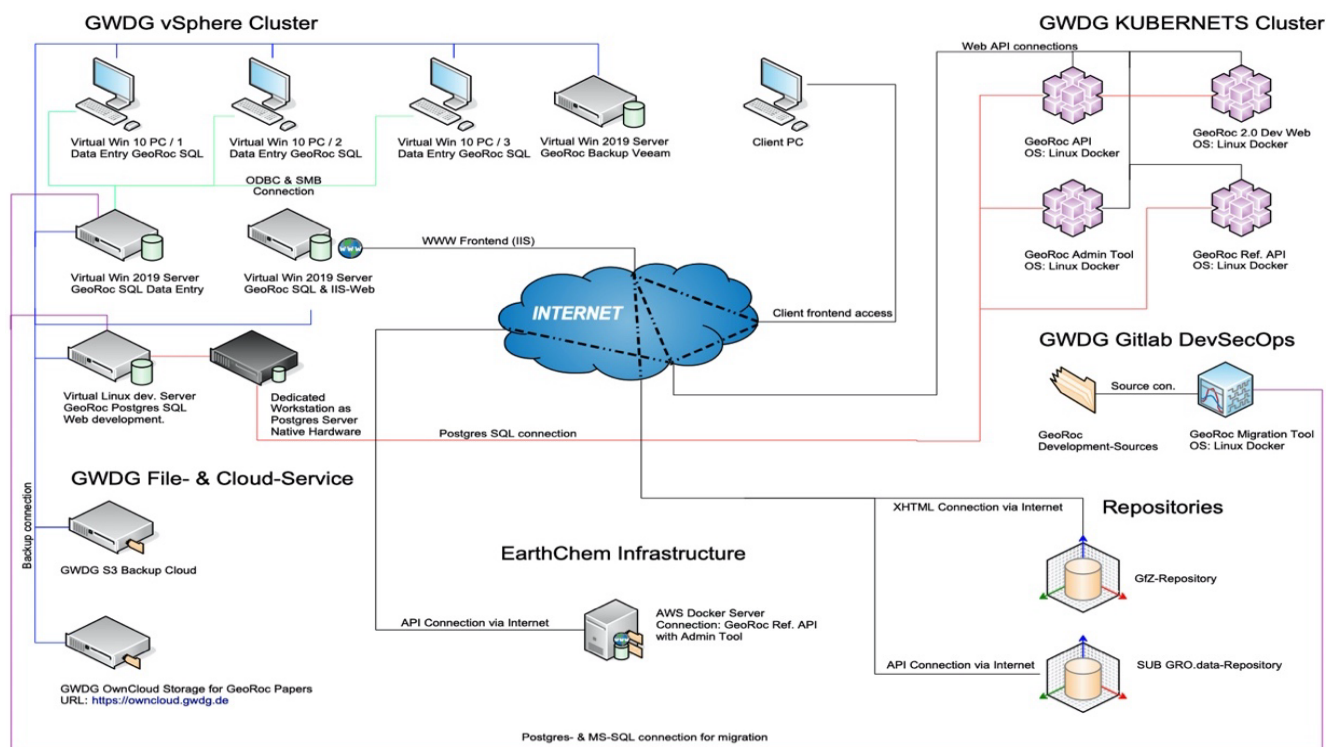


Fig. 2: Overview DIGIS and GEOROC Infrastructure

The new IT infrastructure after migration (Fig. 2) is now based on modern vSphere, Cloud and Kubernetes systems as well as setting up three virtual MS-Windows Servers within the GWDG's VMWare vSphere cluster and the GEOROC data in different data models: (1) GEOROC databases, MS-SQL(geoeroc) & Postgres(ODM2), (2) Webserver IIS 11 with ASP and PHP support, (3) Veeam backup service using GWDG's S3 cloud storage. The creation of a Linux-Apache-Postgres-PHP (LAPP) development server based on Linux within the VMWare vSphere cluster of the GWDG was necessary for running the new database model and also the GEOROC Migration Tool for transferring the vintage GEOROC database into the ODM2 GEOROC 2.0 database.

The GEOROC Migration Tool was developed to run inside a docker container with the advantage of running it on any operating system and different locations inside our Infrastructure. The LAPP Server is also used for several other development steps in the project. For a and reliable infrastructure, we chose to utilize the Kubernetes cluster provided by GWDG: (1) GEOROC API (used for data transfer between the Web interface and the database), (2) GEOROC Admin Tool (Curators tool for entering data in to the new ODM2 database), (3) GEOROC References API (syncing references between different databases), and (4) GEOROC Website 2.0.

The applications developed in the context of this project are deployed on Kubernetes clusters provided by the GWDG. The latter includes a development- and a testing instance of the new DIGIS/ GEOROC API running on two separate clusters. This allows for flexible testing of new features in a live environment while having a stable version available at all times.

The deployment of the new GEOROC website follows the same setup as the API. For the coordination of data entry with EarthChem, an admin database has been set up, which stores information on which research publication is being processed for implementation in to each database to avoid duplication. An API also running in the cluster is used to post new data entries and check whether a data set already exists based on DOI or title and author data.

We also have an instance of a new admin tool that is jointly developed by the EarthChem team for data entry and curation allowing it to be evaluated in terms of fit-for-purpose as part of day-to-day operations. To ensure optimal performance for the new database model we chose to host the GEOROC 2.0 ODM2 Postgres Database on a dedicated workstation for high speed access over I/O performance issues on the GWDG vSphere Servers. For production systems we will have to use dedicated high-speed servers for this purpose.

The GWDG's OwnCloud cluster stores GEOROC papers safely and the GWDG GitLab Infrastructure is used for source code and documentation of the entire project. The codebase for all applications developed in this project, as well as additional documentation is persisted in [GWDG GitLab](#).

Currently the GEOROC database is synchronized with the new data model every three months. In this process the GEOROC data is loaded from the current data model into a transformation script and finally written to the new data model. The pre-compiled files are updated manually in GroData. We are currently establishing collaboration with GFZ Data Services where GEOROC 2.0-curated files will be hosted in the future. When the migration to the new data model is completed, an additional script will export the new data into XML documents. The collection of those documents will then be sent to the EarthChem team where it can be ingested into their database.

The "customized" ODM2 data model for GEOROC 2.0

The vintage GEOROC data model (Lehnert et al., 2000) has been mapped to the Observations Data Model Version 2 (ODM2; Horsburgh et al., 2016). ODM2 is an implementation of the domain-agnostic Observations and Measurements standard (O&M; <https://www.iso.org/standard/82463.html>; Cox 2010) for geosciences that is designed to support interoperability between databases hosting observational data not just from physical objects (rocks and minerals) but also other type of samples

including time series data (Hsu et al., 2017). The data model is centered around sampling features and observations and measurements. For GEOROC 2.0, all basic functionalities of ODM2 are adopted. Only minor additions were necessary to preserve the domain-specific data content of GEOROC 2.0. All legacy data have been preserved from the old data model. New features include the implementation of external identifiers for controlled vocabularies, DOI's for datasets, ORCID's for authors and IGSN for samples. A full description of the new data model is available at <https://github.com/DIGIS-Georoc/DIGIS-Datamodel> and is implemented in the open-source relational database management system PostgreSQL. It is also available on GitHub and Zenodo (Klöcking and Sturm, 2024). To facilitate better data discovery for the user, GEOROC 2.0 now integrates additional persistent universal identifiers such as DOIs for datasets as well as journal publications, author ORCIDs and the International Generic Sample Number (IGSN). Furthermore, time stamps allow for incremental data migration or -export.

Since the existing (vintage) GEOROC data base needs to remain operational for now to feed the existing (old) GEOROC front end website, data entry into GEOROC by curators feeds the old data model via MS-Access over ODBC into the migrated GEOROC MS-SQL database. To enable simultaneous data entry by different curators from different locations three client systems (MS-Windows 10 Pro) were setup. For the moment, newly entered data by curators is still first uploaded (every three months) to the old data base to feed the existing (old) website. In a second step the new data are migrated into the new ODM2 data model to allow API-access and updating the EarthChem library data bases. The development of the new GEOROC 2.0 front end and data entry tool also runs on the new ODM2 data model. By the end of the present funding period (05/2025) work-flows will be more efficient by entering data directly into ODM2 data base, and have a new GEOROC front end up and running.

While the ODM2 model provides interoperability, it imposes a rather deep and complex hierarchy for long-tail data such as ours. For the new frontend with extended tools, ODM2 is rather inefficient, in particular when complex filtering (and thus MSQL requests) by various users may occur simultaneously. For the 2nd funding period we will check basic compliance/compatibility with the new OMS standard (iso 19156:2023) that was released in May 2023 and now supersedes O&M. If the EarthChem team also changes to the OMS standard will we also update from ODM2 to OMS.

Migration of vintage GEOROC content to the new GEOROC 2.0 data model

Migration of the existing vintage GEOROC database to the new data model as described above required the conversion from Microsoft SQL Server 2019 to PostgreSQL (14.9) before mapping the data schemas. For executing the migration schema, a python script was developed that is configurable via environment variables and that can be run directly or in a Docker container. The migration software package is not openly available, but can be accessed by reviewers upon request through a DFG staff moderated communication: https://gitlab.gwdg.de/fe/digis/digis_migration.

The migration was completed after all legal and IT issues had been solved in April 2022 as advertised in <https://www.uni-goettingen.de/en/3240.html?id=6643> and featured as "NEWS" on the homepage of the Geochemical Society. During the migration process, multiple errors and inconsistencies became apparent that had to be corrected. This resulted in a thorough cleanup of the vintage GEOROC database content, including issues such as the removal of incomplete records and duplications, and standardization of vocabulary entries that were not previously regulated via controlled vocabularies. After that, essential changes (such as adding unique identifiers) were applied to prepare for the mapping into the new schema as the final step of the migration. Indexes and materialized views that improve the performance of common database queries were created during the migration. The migrated database is already accessible to the public through the new API (see below). The new search interface based on this improved data model is currently under development and will be available as part of the new GEOROC 2.0 website by October 2024.

Establishing links and data harmonization with EarthChem platform

GEOROC data are exported to the XML format following the EarthChem XML Schema (one file per sample). The use of timestamps allows GEOROC data to be incrementally exported to the EarthChem XML schema followed by integration into the EarthChem Portal. Updates are generated on a quarterly basis and are automatically ingested into the EarthChem Portal where they can be queried jointly with data from five other geochemical databases. The initial data ingest of the DIGIS project in 11/2022, the first after a five-year hiatus before GEOROC transitioned to Göttingen, added data from more than 3800 publications to EarthChem. Data from ~250 publications are now added every three months.

Beyond sharing data between GEOROC and the EarthChem Portal, EarthChem and DIGIS are also collaborating in the development of a new web service that supports submission of data extracted from any previously published peer-reviewed journal article. The latter intends to engage the community in contributing missing data records to the GEOROC and PetDB databases. Building on software originally developed for the Astromaterials Data System, the DIGIS/EarthChem/ PetDB team is planning to develop and implement a browser-based “AdminTool” for data entry that is based on current data entry templates. Another goal is to complete the integration of external vocabularies. This will enable full interoperability between PetDB and GEOROC whilst maintaining a high level of flexibility to implement

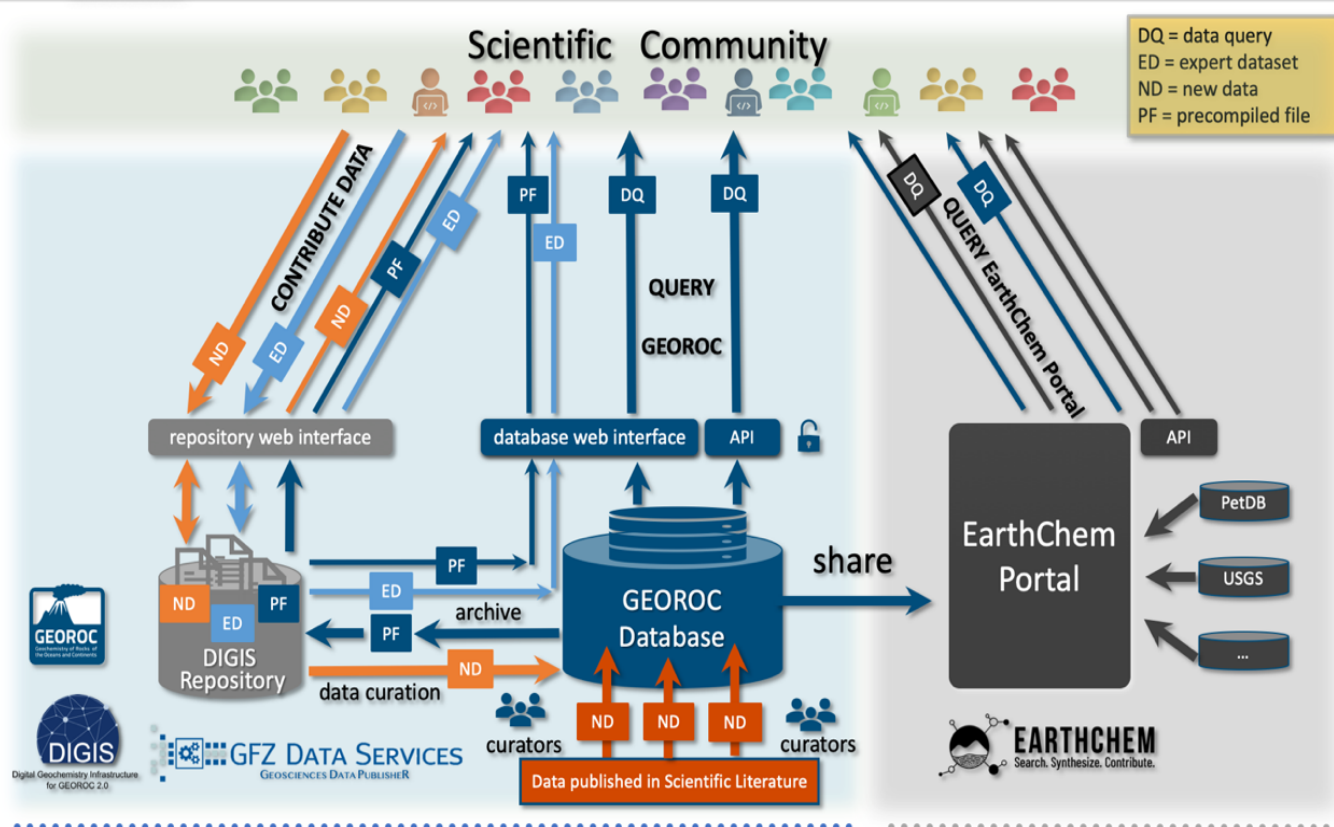


Fig.3: GEOROC and its national and inter-national IT infra-structure environment

changes in nomenclature agreed by the community. With vocabularies harmonized, we plan to develop the “AdminTool” for our DIGIS/GEOROC data entry into the new ODM2 data model. A test version of the “AdminTool” is already in use on the GWDG kubernetes cluster in Göttingen alongside the other infrastructure developed in this project. At the same time, it shares the codebase (via a private GitHub repository) with both EarthChem and Astromat. Each instance of the “AdminTool” has its own, local input database (PostgreSQL), which is consecutively synchronized with the productive databases. In addition, a data curation tool will be designed for data entry into the new ODM2 data model of GEOROC 2.0 for by the end of the first DIGIS funding period. Its full implementation (next funding period) will then allow guest access for trusted members of the geochemical community to assist with data entry. A

dedicated GEOROC curator will oversee editorial handling including checking for completeness and quality of community-entered data. Eventually this approach will fulfil one of our goals to expand and involve the wider geochemistry community for providing input into the GEOROC/PetDB data product via a network of independent curators. This will be one of several important steps to ensure a sustainable future of our project.

Developing Interoperability through API development

RESTful API. DIGIS has implemented data and metadata exchange via a standardized application interface (API). This API will be the basis for any queries of the new GEOROC 2.0 database (based on ODM2) and ensures that data can be more readily harvested and exchanged with other data services. A beta-version of the GEOROC 2.0 API was released for testing in January 2023 (<https://api-test.georoc.eu/api/v1/docs/index.html>). The new DIGIS-API as well as the new frontend is deployed in the GWDG kubernetes cluster of the Göttingen State and University Library. The API provides an endpoint for retrieving sample data in GeoJSON format, for filtering samples, geographical clustering of samples, as well as supporting custom query routes for specialized data use cases. Additional API routes are in development following community feedback and requirements from the frontend design. A Jupyter Notebook available via DIGIS GitHub demonstrates the capabilities of the API at conferences and workshops for performance tests (https://github.com/digis-georoc/Georoc_jupyter). The API's versatility was tested and improved by feedback from participants. To date, nine members of the community have requested/been granted a personal API key for further testing. Shi et al., 2023 employed machine learning to GEOROC datasets, Smith and Asimow (2005) tested links of GEOROC analyses to the new MELTS algorithms. The remaining months of the project will focus on the improvement of performance of the new GEOROC front-end via API.

Improving IT soft- and hardware performance

The new front end and their intended extended filtering and data analysis functions, which are presently in development (see above) requires software and hardware optimization to be fit-for-purpose in terms of searching, filtering and querying results. After optimizing software performance, it became apparent that the GWDG vSphere VMWare Virtual Machines will not provide the performance that will be required to run large SQL queries with an acceptable time frame. The results of our performance test identified a poor I/O performance of the vSphere Cluster as cause for the under-performance of the GWDG vSphere VMWare Virtual Machines, yielding only 38 MB/s of read/write speed of the storage cluster. Our new dedicated hardware (PCI-E V4 NVMe SSD hard drive) is showing a much higher I/O performance with a speed of Instead of 38 MB/s on the GWDG Cluster we achieved up to 4.500 MB/s. This gave us more than ten times faster SQL query performance.

Also, the number of cores and their respective speeds inside a CPU is crucial, because every single query will use one single core of the CPU. The more cores a CPU has, the faster queries can run at the same time without waiting for the next free core. We can handle much more request from the API on dedicated hardware with CPU's how have a lot of cores. Also, the speed that every core runs with is crucial. We determined that a AMD EPYC 9000 Series hardware system will fit our needs best, because it provides a large number of cores at a very high frequency.

The size of main memory is another part that has a large impact on the timing of query answers. An optimized *postgres* database would be able to hold a large data volume in the database cache in the main memory of the computer. To assure that we always can take advantage of the maximum database cache, a main memory of 128GB up to 256GB would be necessary.

In conclusion, our performance testing indicates that in the future we will need dedicated Server Hardware with modern CPU's and very fast PCI-E V4 NVMe SSD hard-drives, supported by a main

memory of 128GB - 256 GB to achieve fast query answers and even quicker search results. During the 2nd funding period, we will further improve speed, performance, and safety and adapt the DIGIS-GEOROC data system to a new dedicated server. We will also establish the dynamic creation of an “interim table” for low-hierarchy index data search (see project proposal).

3. 2. Community Network Building and International Cooperation

Community building and networking with national and international stakeholders is an important aspect of the DIGIS project and has been a major effort particularly for the DIGIS project coordinator and chief scientist (Dr. Marthe Klöcking). A mission statement, database and repository policies, project outputs, news and further project information are available through the DIGIS website at: <http://digis.geo.uni-goettingen.de>. We were able to establish robust links and cooperation with the international digital geochemistry community our DIGIS initiative already established itself internationally as one of the principal players in defining and developing IT infrastructure for geochemistry research (Fig. 3).

The OneGeochemistry Initiative: DIGIS is a founding member of the [OneGeochemistry Initiative](#) (Lehnert et al., 2022). OneGeochemistry seeks to build and maintain international, community-agreed standards for geochemical data to facilitate interoperability between globally distributed data systems and make the data stored in each more findable, accessible and reusable (improve access to and trust in data). Community engagement, e.g. at international conferences and workshops, is a key component of this work (Klöcking et al., 2023). The vocabulary efforts shared between DIGIS and EarthChem will be an important contribution to the OneGeochemistry mission. Through OneGeochemistry, DIGIS is a partner in the geochemistry case study of the [Horizon Europe “WorldFAIR” project](#). Led by CODATA (the Committee on Data of the International Science Council) and RDA (the Research Data Alliance), the project is funded to produce recommendations, interoperability frameworks and guidelines for FAIR data assessment across eleven disciplines.

The EarthChem Portal: GEOROC is a key contributor to the EarthChem Portal (Lamont-Doherty Earth Observatory, USA)). The EarthChem/GEOROC/DIGIS workshop (New York 5-9 Dec 2022) included six participants from Göttingen (+three per ZOOM) and five curators/developers of EarthChem. The meeting resulted in a close coordination of our data bases on (1) data curation, (2) data repository policies, (3) aspects of an indexDB, (4) user statistics, (5) common synchronization service, (6) avoid duplications, (7) joint data entry templates, (8) vocabularies for data and metadata.

Public Outreach and Publications: DIGIS representatives have presented at 18 conferences, convened nine sessions, town hall or splinter meetings, twice led a workshop at the Goldschmidt conference and published three journal articles. Two *OneGeochemistry* exhibition booths at Goldschmidt 2022 and 2023 were jointly organized by DIGIS, EarthChem, AusGeochem. Presentation materials are published on <https://zenodo.org/communities/digis/> and DIGIS publications on <https://uni-goettingen.de/en/resources/643398.html>.

DIGIS is a member of the NFDI₄Earth plenary and representative for NFDI₄Earth at University of Göttingen. DIGIS coordinator M. Klöcking was elected member of the [NFDI₄Earth steering group](#). This activity is enriching FID-Geo (GFZ-Data Services) with curated geochemical data.

3.3 Publications (2020-2024) (OA: open access, DIGIS author in bold)

Chamberlain KJ, Lehnert KA, McIntosh IM, Morgan DJ, **Wörner G** (2021). Time to change the data culture in geochemistry. *Nature Rev Earth and Environment* 2 737–739, <https://doi.org/10.1038/s43017-021-00237-w>, OA

Klöcking M, Sturm A (2024) DIGIS Data Model for GEOROC 2.0, ZENODO, January 8, 2024 | Version 1.0.0. <https://doi.org/10.5281/zenodo.10468217>, OA

Klöcking M, Wyborn L, Lehnert KA, Ware B, Prent AM, Profeta L, Kohlmann F, Noble W, Bruno , Lambart S, Ananuer, H, Barber ND, Becker, H, Brodbeck M, Deng H, Deng K, Elger K, de Souza Franco G, Gao Y, Zhou T (2023) Community recommendations for geochemical data, services and analytical capabilities in the 21st century. *Geochim Cosmochim Acta* 351: 192–205). <https://doi.org/10.1016/j.gca.2023.04.024> OA

Prent AM, Hezel DC, **Klöcking M**, Wyborn L, Farrington R, Elger K, Profeta L, Nixon AL, Lehnert K (2023) Innovating and Networking Global Geochemical Data Resources Through OneGeochemistry. *Elements* 19: 136–137. *Min Soc America*. <https://doi.org/10.2138/gselements.19.3.136> OA

Willbold M, Elliott T (2023) Molybdenum isotope evidence for subduction-modified, recycled mafic oceanic crust in the mantle sources of ocean island basalts from La Palma and Hawaii. *Earth Planet Sci Lett* 621, 118399.

Willbold M, Messling N (2023) Systematic Across-Arc Variations of Molybdenum Isotopes in a Fluid-Dominated Subduction Zone System. *G³* 24, <https://doi.org/10.1029/2023GC011007> OA

We further contributed 36 abstracts at international / national scientific conferences and organized / contributed to eleven conference workshops on digital geochemistry.

4. Long-term continuation of the services developed in the project

The first funding phase (2020 to 2024) has provided the basis for a functional IT data service for GEOROC, which needs to be extended, secured, and augmented by new and innovative data presentation and analysis services during the second funding phase.

The safeguarding of the results so far will be secured by the continuous support from the GZG and SUB (see above). However, at this stage, there are still several challenges towards making the GEOROC data base a fully self-sustained IT data and infrastructure "product". These challenges are described as work packages in the proposal for the next funding phase. Of particular importance here is Task 2-5 "AdminTool". An easy-to-use AdminTool requires (1) harmonized data curation processes aligned between the GEOROC, PetDB, and Astromaterials databases and (2) consistent vocabularies and data quality across synthesis data bases A „user-friendly AdminTool" will then be the essential requirement for establishing a network of curators nationally and internationally for the smooth function and easy workflow of the distributed data curation. In this effort, we will work closely with the EarthChem team at Lamont Doherty Earth Observatory to maintain the close links between our databases. Developing secured data base operation through a distributed network of contributors, it is planned to develop a business model in the next funding period based on advice from the Scientific Advisory Board that we will assemble during the remaining 10 months of the currently funded DIGIS project phase.

Another cornerstone for the sustainable operation of DIGIS/GEOROC is to further strengthen our efforts to establish the "[OneGeochemistry Initiative](#)" as an international institution for safeguarding and defining the smooth and effective operation of an international geochemical data base network. This maintains to be a major focus of our efforts during a second funding phase. To this end, we seek to

actively engage with the community by involving external research groups and by deepening the links between GEOROC, EarthChem, AusGeochem, and other data base systems. This includes educating the community on the importance of best data practices and the value of geochemical data in a broader universal context, such as the UN WorldFAIR data project in which DIGIS is involved.

A key is organizing topical workshops on digital geochemistry, comprising the editorial and curation functionality, data standards, and geochemical research applications. We plan to offer this training at symposia and town hall meetings at international conferences. An already successful measure is the on-site presence of GEOROC staff at key international scientific meetings of the geochemistry community (e.g., Goldschmidt conference, AGU fall meetings 2020 to 2024; see section 1.3 of the proposal) in the form of scientific sessions and information booths, where new features and tools of the database are presented with a special focus on engaging with early career researchers within our community. Engagement of the research community for the GEOROC project and community building / engagement will therefore continue to be continued at a high level.

4.2 Contributors to the project

Daniel Kurzawe	SUB Göttingen, Abt. Forschung und Entwicklung (PI)
Matthias Willbold	GZG, Universität Göttingen (PI)
Gerhard Wörner	GZG, Universität Göttingen (PI)
Leander Kallas	DIGIS project staff at GZG, Universität Göttingen
Marthe Klöcking	DIGIS project staff at GZG (5/2021-10/2023), now University of Münster
Nils Meßling	DIGIS project staff at GZG, Universität Göttingen (part)
Adrian Sturm	DIGIS project staff at SUB Göttingen
Mareen Geestman	SUB Göttingen, frontend design support
Stefan Möller-McNett	IT infrastructure, GZG, Universität Göttingen (GZG permanent staff, part)
Jens Nieschulze	Universität Göttingen, Abteilung Forschung und Transfer (part)
Paul Pestov	SUB Göttingen, frontend design (part)

External Partners

Kirsten Elger,	Deutsches GeoForschungsZentrum GFZ, Potsdam
Kerstin Lehnert	Lamont-Doherty Earth Observatory, Columbia University, New York.
Lucy Profeta	Lamont-Doherty Earth Observatory, Columbia University, New York.
Lesley Wyborn	Australian National University, AuScope Virtual Research Environments
Alexander Prent	One Geochemistry Secretary & Coordinator

Co-authors in joint conference contributions

Manja Luzi-Helbing	Geo-X Netzwerk Berlin / Potsdam und GFZ
Julia Kowalski	University Aachen
Albrecht Hofmann	MPI für Chemie, Mainz

4.3 Community use/re-use of project results.

For more than 25 years, the GEOROC database served the global Geosciences research community to study the origin, evolution, and chemical differentiation of planet Earth, its mantle, oceanic crust, volcanic islands, and the continents through geological time. This provision of curated, high-quality “long-tail” research data in geochemistry has contributed to the rise of “digital geochemistry”. The GEOROC database has developed into a powerful tool and data resource in geochemistry.

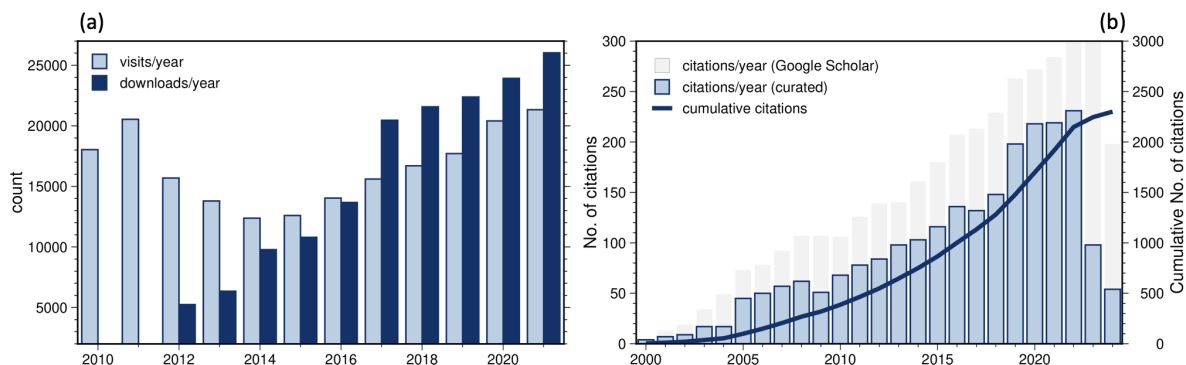


Fig. 4: GEOROC data base use in international digital geochemistry research

Fig. 4: Global use of the GEOROC data base

Since the completion of the GEOROC migration from MPI Mainz to DIGIS (June 26, 2022), access to the website is monitored through Matomo (<https://matomo.org/>) (Fig. 4). So far, we have recorded 27,662 GEOROC website visitors during this period, with accesses from 126 different countries (with the majority of visits by users based in the USA, China, Europe and Australia). The most recent count shows 5000 website visits/month. We observed a steadily increasing number of data downloads from the data base (Fig. 4a), with the most recent count in July 2024 of 1670 downloads of data sets. 97,010 downloads alone of the "Precompiled Files" are recorded from the DIGIS/GEOROC data repository at GRO.data. A more in-depth analysis has shown, that the use of the new map-based data search has doubled in 2023 and in 2024 we observed 2600 "true" visitors with an average of 16 minutes of data search. Between 2000 and 7/2024, GEOROC data were cited in 2300 publications (and four books), increasing from <10 citations in 2000 to >30 citations in 2023 (Fig. 4b). Of these publications, we consider that 2300 articles considerably impacted science as these studies combined were cited 96,218 times. This proves that “big-data” and statistical approaches to global geochemical data from GEOROC to extract new ideas and interpretations in the new field of digital geochemistry.

To name a few selected examples: Only based on large data compilations in geochemistry was it possible to detect secular changes in the composition (Tang et al., 2016), and the production rate of ocean crust and continents through geological time (Condie et al., 2011). More recently, the database also enabled the systematic relation of global crustal thickness to trace element signatures in magmatic settings in modern (Park et al., 2021) and ancient times (Luffi and Ducea, 2022), the formation of crust and copper deposits in convergent plate margins (Wang et al., 2022; Barber et al., 2021), and links between systematics of volcanic activity, magma composition and the origin of magmas (e.g., Oggier et al., 2023; Klein and Müntener, 2023). These representative studies show that data retrieved from GEOROC on the global distribution of the composition of rocks and minerals throughout Earth's history have significantly contributed to new research findings. At the same time, we observe that not all authors follow our guidelines for proper citing GEOROC even when large data sets downloaded from our website are the fundament for the published research. On the new GEOROC frontend we will place and strengthen our requests for proper citations.

5. References

- Barber ND, Edmonds M, Jenner F, Audétat A, Williams H (2021) Amphibole control on copper systematics in arcs: Insights from the analysis of global data sets. *Geochim Cosmochim Acta* 07: 192-211. <https://doi.org/10.1016/j.gca.2021.05.034>
- Condie KC, Bickford ME, Aster RC, Belousova E, Scholl DW (2011) Episodic zircon ages, Hf isotopic composition, and the preservation rate of continental crust. *Bulletin*, 123(5-6), 951-957.
- Klein BZ, Müntener O (2023) MnO/MgO ratios of arc basalts highlight the role of early garnet fractionation. *Geochem Persp Let* 25, 18–24 | <https://doi.org/10.7185/geochemlet.2309>
- Klöcking M, Sturm A (2024) DIGIS Data Model for GEOROC 2.0, ZENODO, January 8, 2024 | Version 1.0.0. <https://doi.org/10.5281/zenodo.10468217>
- Klöcking M, Wyborn L, Lehnert K, Ware B, Prent A, Profeta L, Kohlmann F, Noble W, Bruno I, Lambart S, Ananuer H, Barber N, Becker, H, Brodbeck M, Deng H, Deng K, Elger K, de Souza Franco, G, Gao Y, Zhou T (2023) Community recommendations for geochemical data, services and analytical capabilities in the 21st century. *Geochim Cosmochim Acta* 351: 192–205. <https://doi.org/10.1016/j.gca.2023.04.024>
- Lehnert K, Su, Y, Langmuir CH, Sarbas B, Nohl U (2000) A global geochemical database structure for rocks. *Geochemistry, Geophysics, Geosystems*, 1(5). <https://doi.org/10.1029/1999gc000026>
- Lehnert, K, Klöcking, M, Elger, K, Wyborn, L, Prent, A, ter Maat G, Hezel D (2022) OneGeochemistry Interim Governance. Zenodo. <https://doi.org/10.5281/zenodo.6566075>
- Luffi P, Ducea MN (2022) "Chemical mohometry: Assessing crustal thickness of ancient orogens using geochemical and isotopic data." *RevGeophysics* 60.2: e2021RG000753.
- Oggier F, Widiwijayanti C, Costa F (2023) Integrating global geochemical volcano rock composition with eruption history data sets. *Front. Earth Sci.* 11:1108056. [doi: 10.3389/feart.2023.1108056](https://doi.org/10.3389/feart.2023.1108056)
- Park JW, Campbell IH, Chiaradia M, Hongda H, Lee CT (2021) Crustal magmatic controls on the formation of porphyry copper deposits. *Nat Rev Earth Environ* 2, 542–557 (2021). <https://doi.org/10.1038/s43017-021-00182-8>
- Tang M, Chen K, Rudnick RL (2016) Archean upper crust transition from mafic to felsic marks the onset of plate tectonics. *Science* 351: 372-375. doi: [10.1126/science.aad5513](https://doi.org/10.1126/science.aad5513)
- Wang J, Gleeson M, Smith WD, Ma L, Lei Z, Shi G, Chen L (2022) The factors controlling along-arc and across-arc variations of primitive arc magma compositions: A global perspective. *Front Earth Sci* 10:1055255. [doi: 10.3389/feart.2022.1055255](https://doi.org/10.3389/feart.2022.1055255)