Introduction to the GEOROC database

Marthe Klöcking[*]

March 2023

digis-info@uni-goettingen.de

[*]based on ‘Kompaktkurs 2000’ by Bärbel Sarbas and Mukul Sharma held at the Max Planck Institute for Chemistry, Mainz

Contents

1 Introduction

GEOROC stands for *Geochemistry of Rocks of the Oceans and Continents.* The GEOROC database contains published geochemical analyses of whole rocks, glasses, minerals and inclusions from eleven different geological settings. Figure 1 shows the distribution of these samples across the world. The database provides free access to >34 million individual values of major and trace element concentrations, radiogenic and nonradiogenic isotope ratios as well as analytical ages. In addition to the chemical analyses, extensive additional information (metadata) is stored: geographic location of samples (with latitude and longitude), rock class and type, alteration grade, texture of the sample, information on sampling technique, analytical method and laboratory, measurement accuracy, analysis of reference materials and bibliographic data.

With this global data compilation, GEOROC regularly contributes to the generation of new scientific knowledge in the Earth Sciences as well as other disciplines. For example, the database is used for data discovery in specific geographic regions, development of

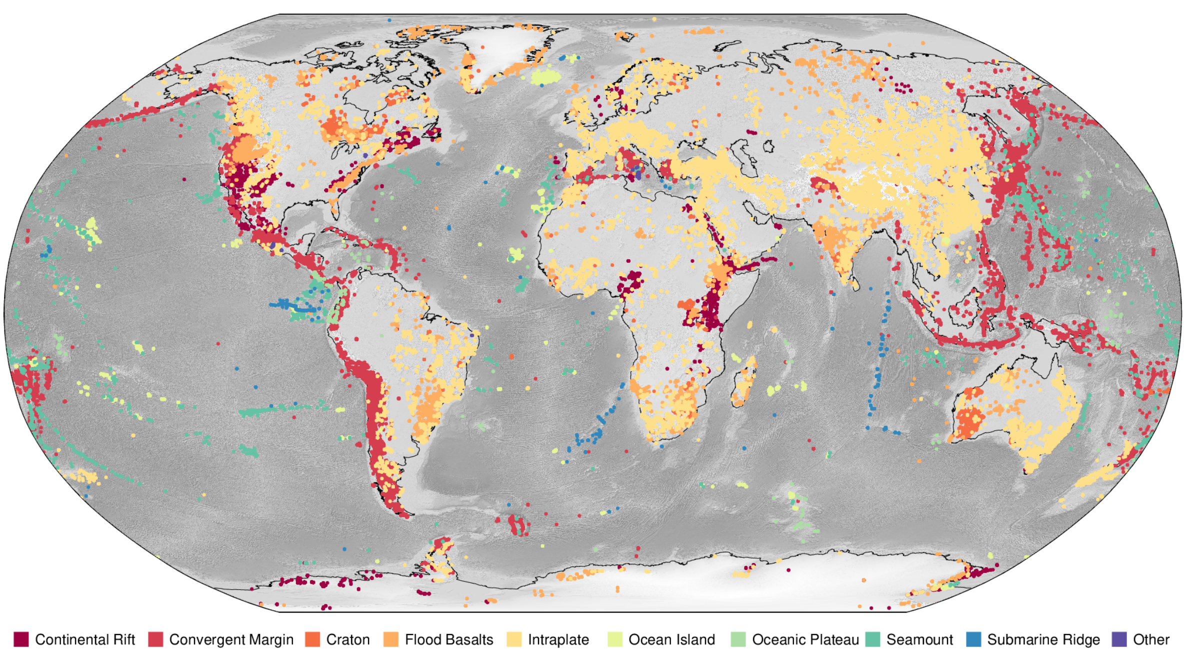


Figure 1: Global distribution of samples compiled in the GEOROC database, coloured by geological setting.

new methods for data analysis and filtering, or to link seemingly unconnected datasets from, e.g., medicine or archaeology to the composition of rocks or minerals. To date, GEOROC has been acknowledged in more than 3,400 peer-reviewed publications, many of which have appeared in high-ranking journals.

Datasets in GEOROC are cross-linked with GeoReM, a database for geological and environmental reference materials at the Max Planck Institute for Chemistry, Mainz (Germany). GEOROC is a key contributor and integral part of the EarthChem platform, hosted by the Interdisciplinary Earth Data Alliance (IEDA) at the Lamont-Doherty Earth Observatory (USA). The GEOROC database is currently maintained by the Digital Geochemical Data Infrastructure (DIGIS) initiative at the University of Göttingen. It can be accessed at `https://georoc.eu`.

1.1 History of the Database

The construction of the GEOROC database under the direction of Bärbel Sarbas began in January 1998 in the former Geochemistry Department of the Max Planck Institute for Chemistry, Mainz. Professor Albrecht Hofmann, then Director of the Geochemistry Department, took on six former employees of the Gmelin Institute (Frankfurt) that had just been discontinued by the Max Planck Society. GEOROC has been available online since the end of 1999. The database structure was developed in close cooperation with the group of PetDB, a petrological database of primarily ocean floor basalts started at a similar time at the Lamont-Doherty Earth Observatory of Columbia University, New York.

The content of the GEOROC database was initially limited to geochemical analyses of ocean island basalts, but was soon expanded to include samples of volcanic rocks from other tectonic environments and the mantle xenoliths they contain. Since 2017, data of all igneous rocks (plutonic and volcanic) are recorded.

After Professor Hofmann retired, the GEOROC group joined the Biogeochemistry department, then the newly Climate Geochemistry department of the Max Planck Institute for Chemistry. However, due to the contractual situation of the former Gmelin employees their number has steadily decreased. The last member of the original GEOROC group, Bärbel Sarbas, retires at the beginning of 2022.

In August 2021, the GEOROC database was formally transferred to the the Georg-August-Universität Göttingen. The new DIGIS group, including Bärbel Sarbas, is continuing to maintain and grow the database. In addition, DIGIS is modernising the database infrastructure and expanding its international connectivity, which will be released under the brand of GEOROC 2.0 in 2024.

MAX-PLANCK-INSTITUT
FÜR CHEMIE

DIGIS
Digital Geochemistry Infrastructure
for GEOROC 2.0

1.2 GEOROC 2.0: Plans for the Future

GEOROC 2.0 is envisioned as a connected platform capable of meeting future challenges of digital data-based research and the evolving needs of the community. The primary objective is to realign the database with current and future demands in digital geochemical research, especially regarding the FAIR data principles (findable, accessible, interoperable, reusable). This service will be provided with future-ready, adaptable IT infrastructure.

To facilitate better data discovery, GEOROC 2.0 will integrate additional universal identifiers such as data DOIs (i.e. where data are published separately from a journal publication), author ORCIDs and the International GeoSample Number (IGSN).

DIGIS will implement an interoperable metadata model as well as data and metadata exchange via standardised application interfaces (APIs). This means that data can be more readily harvested and exchanged with other data services. The API will allow new ways of querying the database from the command line and (user-written) research software. A beta-version of the GEOROC 2.0 API was released for testing in January 2023 and is documented at https://api-test.georoc.eu/api/v1/docs/index.html.

The DIGIS concept for GEOROC 2.0 further includes new pathways for data entry using end-to-end text and data mining and facilitating user submissions of data. To this end, the GEOROC Data Repository was set up in October 2021 which publishes new geochemical data submitted by authors. Note that all data submissions have to be linked to a journal manuscript, however this manuscript does not yet need to be published. A confidential review-link and provisional dataset DOI is provided to authors for the manuscript review process by a scientific journal. All data published within the GEOROC Data Repository will automatically be considered for inclusion into the GEOROC database. A joint submission portal for the DIGIS and EarthChem data systems is currently in development. The two data systems are also collaborating to develop a web service that supports submission of data extracted from any previously published in a peer-reviewed journal article, allowing the community to contribute missing data records to the GEOROC and PetDB databases.

The GEOROC database is built exclusively from analytical data published in peer-reviewed journals. As such, the database relies on the geochemical community as much as the geochemical and other communities rely on GEOROC. Data access and best practises for data reporting and citation are, therefore, fundamental pillars of GEOROC 2.0. DIGIS is one of the founding members of the OneGeochemistry initiative that seeks to build and maintain consensus-driven standards that make geochemistry research data globally findable and accessible, and truly interoperable and reusable to both humans and machines. For more details and updates on GEOROC 2.0 see the DIGIS website: http://digis.geo.uni-goettingen.de.

2 Content of the Database

The GEOROC database currently contains approximately 1.7 million mineral analyses, 500,000 whole rock analyses, 74,000 inclusion analyses, and 50,000 glass analyses (as of March 2023; Fig. 2). These analyses include over 34 million individual values from around 21,000 publications. Table 1 shows how these analyses are distributed over samples of the different geological settings.

Due to GEOROC's initial focus on volcanic rocks, and ocean island basalts in particular, data from volcanic samples are most comprehensive (i.e. almost all published analyses are entered). Plutonic rocks are only recorded for publication dates after 2015, so that a significant portion of analyses across all geological settings may not be included. Currently published works are regularly supplemented.

NOTE: *Submarine Ridge* data are not systematically entered into the GEOROC database. The database PetDB (`https://search.earthchem.org/`) should be used for geochemical data of mid-ocean ridge basalts (MORBs).

Geological Setting	**Whole Rocks**	**Minerals**	**Inclusions**	**Glass**
Archaean Craton	31,130	74,891	282	2
Complex Volcanic Settings	2,850	57,022	1,257	238
Continental Flood Basalt	52,944	129,902	8,537	2,257
Convergent Margin	145,530	410,064	29,013	21,674
Intraplate Volcanics	158,538	820,328	15,932	2,401
Ocean Island	54,181	117,818	11,776	11,457
Ocean-Basin Flood Basalt	1,269	822	42	63
Oceanic Plateau	3,758	6,185	265	503
Rift Volcanics	29,207	109,767	5,454	6,999
Seamount	7,977	18,863	1,221	2,573
Submarine Ridge	1,229	877	11	831

Table 1: Number of analyses collected in GEOROC by geological setting and type of material (as of March 2023).

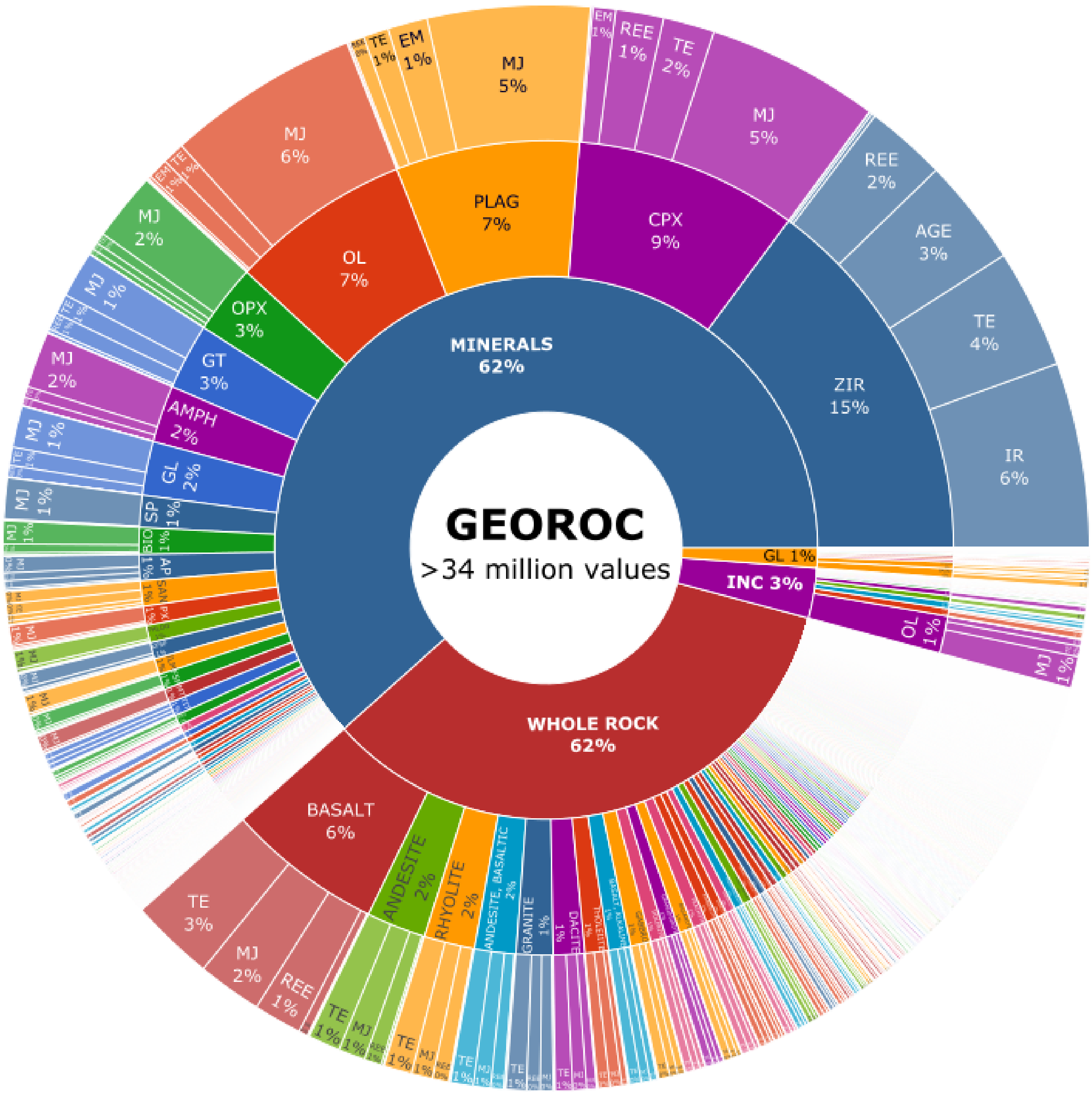


Figure 2: Content of the GEOROC database grouped by material, mineral, lithology and type of analytical data (as of March 2023). Material: GL = glass, INC = inclusions. Analysis type: EM = end-member, IR = radiogenic isotopes, MJ = major element (oxide), REE = rare earth element, TE = trace element.

3 Querying the GEOROC database

Currently, there are five categories via which to start your search of the GEOROC database:

- by Bibliography
- by Geological Setting
- by Geography
- by Chemistry
- by Petrography.

Owing to the large number of data holdings, precompiled search results are also available for direct download. **Precompiled files** are provided for each of the geological settings, lithologies, minerals and for inclusion data.

In addition, GEOROC hosts a number of **expert datasets**. These files are not related to the content of the database, but instead are provided by members of the community as curated compilations of special interest.

3.1 Query by Bibliography

The **bibliographic query** begins with an alphabetical list of authors from which one or more names can be selected (Fig. 3). You can combine this selection with one or more years of publication.

The next window shows the publications of the selected authors currently recorded in the database, regardless of the position of the names in the author order (Fig. 4). These publications can be used as a further selection criterion. In addition, the data tables of the individual publications can also be viewed and downloaded.

The next step is a window listing a set of metadata that can be used to further refine the query (Fig. 5). This list is created dynamically, i.e. only the information that is available for the previously selected samples is shown. The most important criterion,

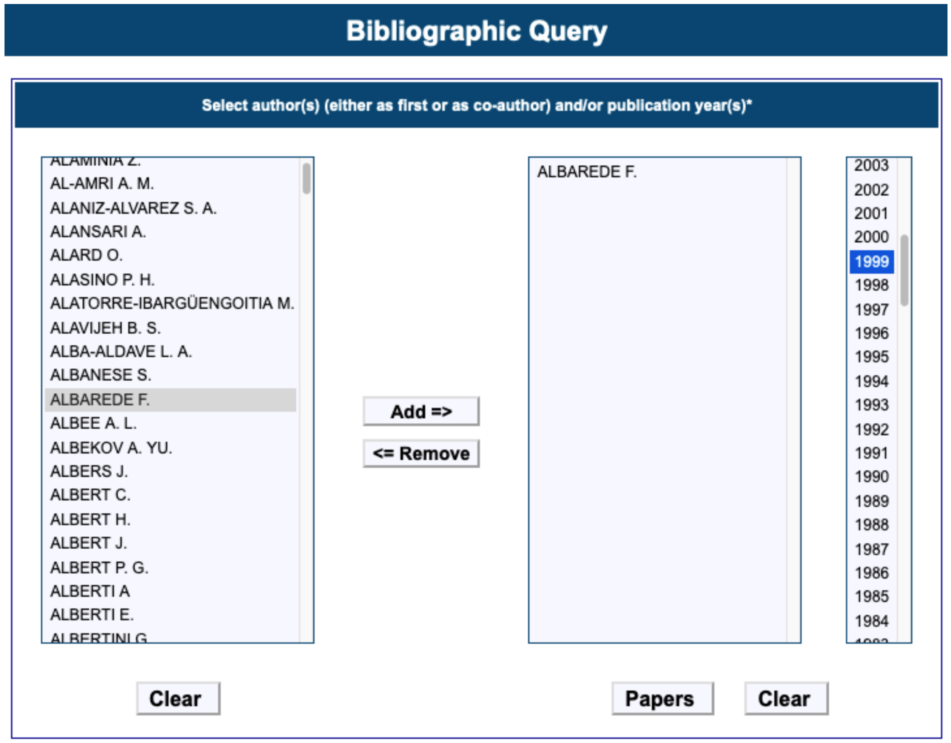


Figure 3: Query by Bibliography: select author(s) and year(s).

and one that should probably always be used, is TYPE OF MATERIAL. This criterion defines the type of material analysed on a sample: whether it is a whole rock analysis or the analysis of a mineral, an inclusion (glass, fused or mineral inclusion) or a volcanic glass. The criterion ROCK TYPE is also important. Here you can limit the search to samples of volcanic rocks, for example.

As an example, if you have selected the criteria TYPE OF MATERIAL and ROCK TYPE, two lists with the corresponding entries open in the next window (Fig. 6). By default, the criteria of the two lists are linked with an AND operator. However, an OR operator can also be selected. Their different functions are best explained using an example. If the search is limited to samples from a particular publication, a query with the criteria 'volcanic rock' AND 'whole rock' leads to the whole rock analyses of volcanic rocks from this publication. In contrast, selecting 'volcanic rock' OR 'whole rock' will select all volcanic analyses (i.e. whole rock, glass, mineral and inclusions analyses) plus all available whole rock analyses from all existing rock types.

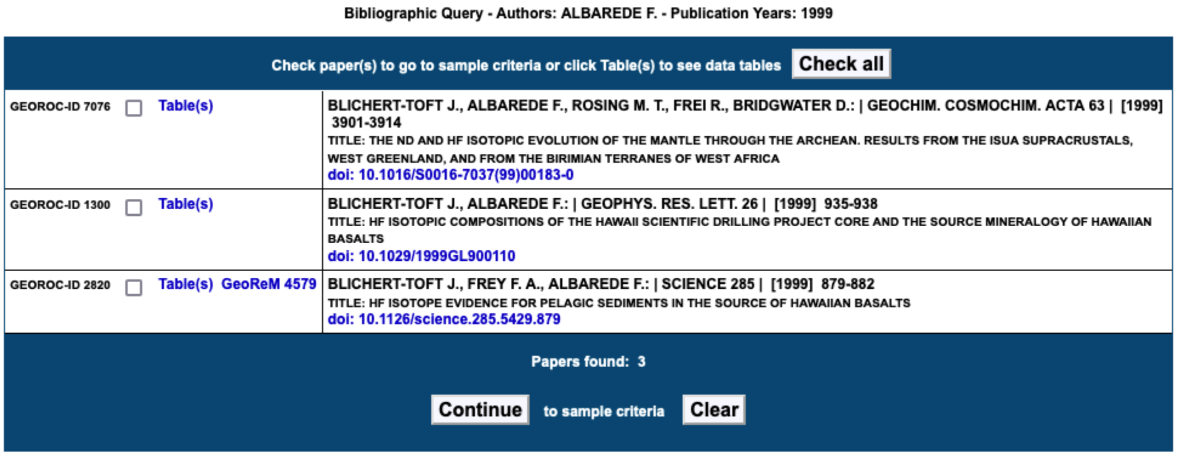


Figure 4: Query by Bibliography: select publication(s).

After refining the search using the sample metadata, the chemical elements and isotope ratios available for the selected samples are listed (Fig. 7). These are divided into groups: major elements, trace elements with the rare earth elements as a separate list, radiogenic isotope ratios, isotope (ratios) of the U-decay series, stable isotope ratios, noble gases and the analytically determined age. Entire groups or individual entries can be selected. Furthermore, on this page you can specify whether the analytical method and laboratory should also be displayed.

In the next window, the metadata is selected that should be included in the search result in addition to the chemical analyses (Fig. 8). Standard fields, which are always

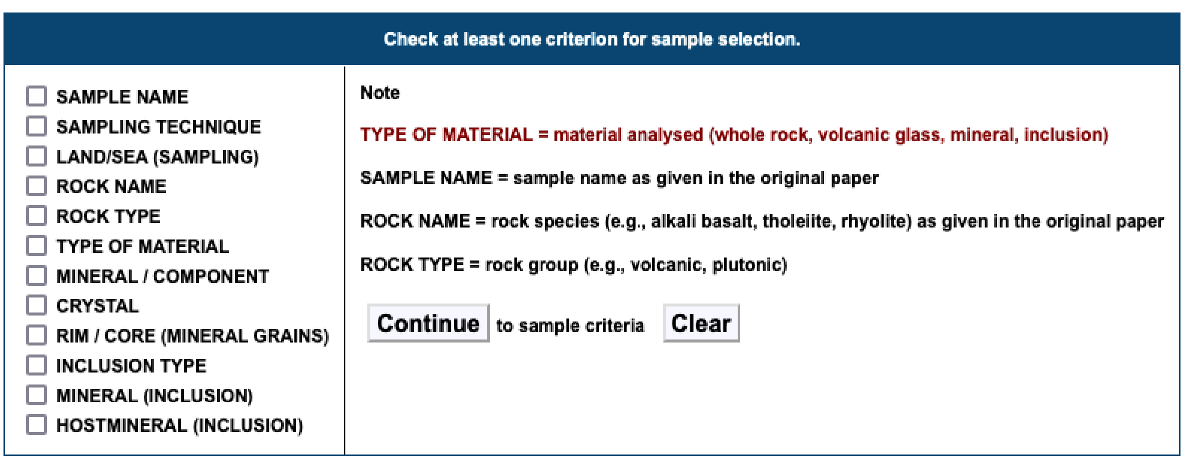


Figure 5: Select which metadata to use for further refining the query.

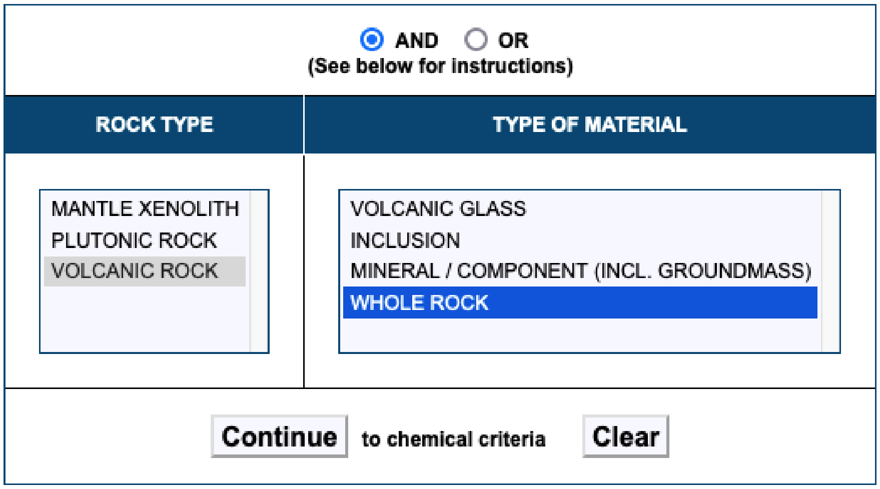


Figure 6: Further refine query by TYPE OF MATERIAL and ROCK TYPE.

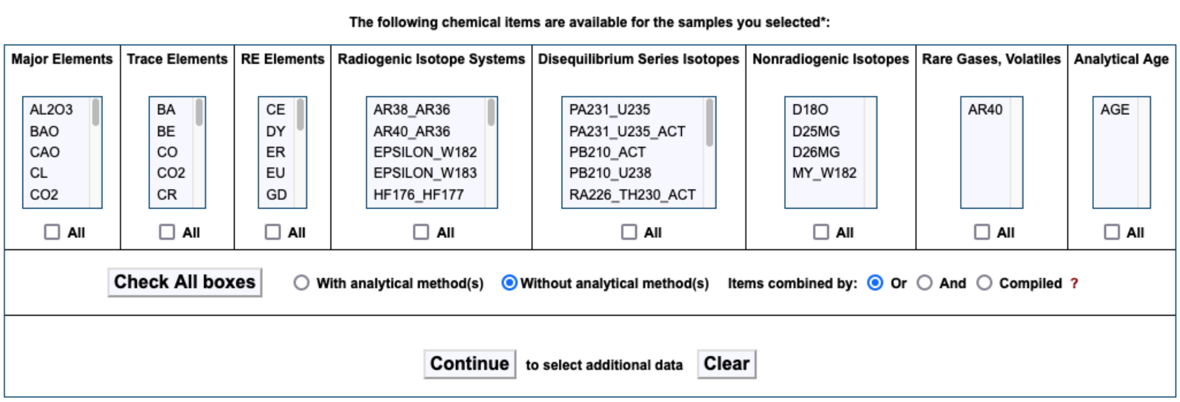


Figure 7: Select which analytes to display.

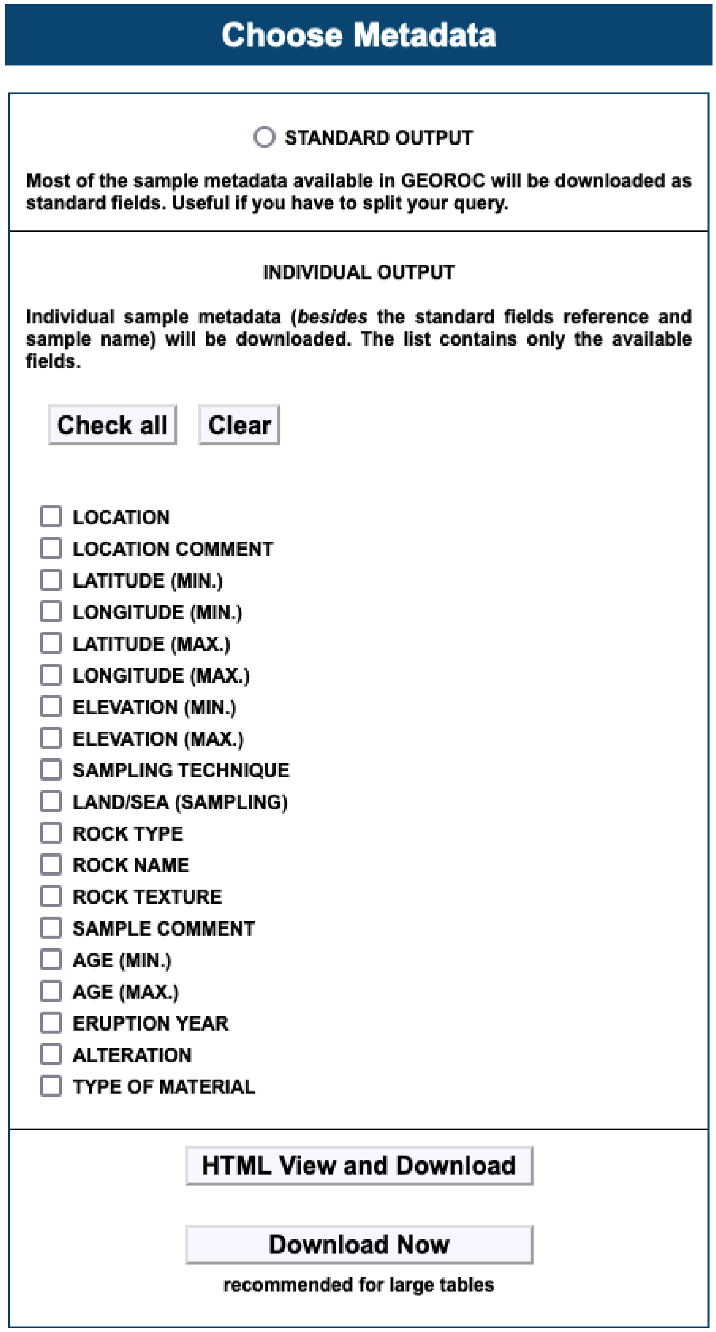


Figure 8: Choose metadata to include in the result along with the chemical analyses.

displayed, are the sample name, a unique sample ID and the reference. You may also choose 'standard output' which displays most of the metadata available in GEOROC and ensures a common download format.

The result of the query can be downloaded directly or viewed as an HTML table before

CITATION	SAMPLE NAME	UNIQUE_ID	LOCATION	ROCK NAME	RB(PPM)	SR(PPM)
HUGHES S. S. (1986) [1827] Standards/Precision(Info)	EV01	7839-EV01	PARANA / SOUTHERN PARANA PROVINCE	BASALT	33	260 \ 242
HUGHES S. S. (1986) [1827] Standards/Precision(Info)	G-215	7839-G-215	PARANA / SOUTHERN PARANA PROVINCE	ANDESITE, BASALTIC	40	150
HUGHES S. S. (1986) [1827] Standards/Precision(Info)	G-402	7839-G-402	PARANA / SOUTHERN PARANA PROVINCE	BASALT	17	210
HUGHES S. S. (1986) [1827] Standards/Precision(Info)	G-648	7839-G-648	PARANA / SOUTHERN PARANA PROVINCE	BASALT	18	120
HUGHES S. S. (1986) [1827] Standards/Precision(Info)	IO2	7839-IO2	PARANA / SOUTHERN PARANA PROVINCE	ANDESITE, BASALTIC	48	200
HUGHES S. S. (1986) [1827] Standards/Precision(Info)	MO1	7839-MO1	PARANA / SOUTHERN PARANA PROVINCE	BASALT	34	220
HUGHES S. S. (1986) [1827] Standards/Precision(Info)	MO2	7839-MO2	PARANA / SOUTHERN PARANA PROVINCE	BASALT	40	110
HUGHES S. S. (1986) [1827] Standards/Precision(Info)	MO3	7839-MO3	PARANA / SOUTHERN PARANA PROVINCE	ANDESITE, BASALTIC	14	260
HUGHES S. S. (1986) [1827] Standards/Precision(Info)	MRO3	7839-MRO3	PARANA / SOUTHERN PARANA PROVINCE	ANDESITE, BASALTIC	73	170
HUGHES S. S. (1986) [1827] Standards/Precision(Info)	RS16	7839-RS16	PARANA / SOUTHERN PARANA PROVINCE	BASALT	18	190
HUGHES S. S. (1986) [1827] Standards/Precision(Info)	RS20	7839-RS20	PARANA / SOUTHERN PARANA PROVINCE	BASALT	19	120
HUGHES S. S. (1986) [1827] Standards/Precision(Info)	RS26	7839-RS26	PARANA / SOUTHERN PARANA PROVINCE	BASALT	22	290
HUGHES S. S. (1986) [1827] Standards/Precision(Info)	RS30	7839-RS30	PARANA / SOUTHERN PARANA PROVINCE	RHYODACITE	140	140
HUGHES S. S. (1986) [1827] Standards/Precision(Info)	RS37	7839-RS37	PARANA / SOUTHERN PARANA PROVINCE	BASALT	22	230
HUGHES S. S. (1986) [1827] Standards/Precision(Info)	RS50	7839-RS50	PARANA / SOUTHERN PARANA PROVINCE	RHYOLITE	210	50
HUGHES S. S. (1986) [1827] Standards/Precision(Info)	RS60	7839-RS60	PARANA / SOUTHERN PARANA PROVINCE	BASALT	25	220
HUGHES S. S. (1986) [1827] Standards/Precision(Info)	RS65	7839-RS65	PARANA / SOUTHERN PARANA PROVINCE	RHYODACITE	120	150 \ 161
HUGHES S. S. (1986) [1827] Standards/Precision(Info)	RS68	7839-RS68	PARANA / SOUTHERN PARANA PROVINCE	BASALT	19	470
HUGHES S. S. (1986) [1827] Standards/Precision(Info)	RS69	7839-RS69	PARANA / SOUTHERN PARANA PROVINCE	ANDESITE, BASALTIC	69	290
CITATION	SAMPLE NAME	UNIQUE_ID	LOCATION	ROCK NAME	RB(PPM)	SR(PPM)
HUGHES S. S. (1986) [1827] Standards/Precision(Info)	RS70	7839-RS70	PARANA / SOUTHERN PARANA PROVINCE	ANDESITE, BASALTIC	67	110
HUGHES S. S. (1986) [1827] Standards/Precision(Info)	RS71	7839-RS71	PARANA / SOUTHERN PARANA PROVINCE	ANDESITE, BASALTIC	33	300 \ 214
HUGHES S. S. (1986) [1827] Standards/Precision(Info)	RS73	7839-RS73	PARANA / SOUTHERN PARANA PROVINCE	BASALT	24	470
HUGHES S. S. (1986) [1827] Standards/Precision(Info)	RS76	7839-RS76	PARANA / SOUTHERN PARANA PROVINCE	BASALT	35	130
HUGHES S. S. (1986) [1827] Standards/Precision(Info)	RS80	7839-RS80	PARANA / SOUTHERN PARANA PROVINCE	ANDESITE, BASALTIC	26	240

(Alternative values are separated by an **)

[GEOROC-ID 1827 Tables] HUGHES S. S., SCHMITT R. A., WANG Y. L., WASSERBURG G. J.: | GEOCHEM. J. 20 [1986] 173-189
TRACE ELEMENT AND SR-ND ISOTOPIC CONSTRAINTS ON THE COMPOSITIONS OF LITHOSPHERIC PRIMARY SOURCES OF SERRA GERAL CONTINENTAL FLOOD BASALTS, SOUTHERN BRAZIL | doi: 10.2343/geochemj.20.173

Figure 9: Example of a query result displayed as an HTML table.

download (Fig. 9). This table contains the reference in short form, the sample name, further metadata selected by the user and the selected analysis data including the respective unit. In GEOROC, concentrations are always entered in the unit given in the original reference. This can result in a query result for an item containing two or more columns with the concentrations in each unit.

If element concentrations or isotope ratios have been determined several times in a work on a sample, these values are given in one field, separated by a forward slash (/). It is then up to the user to decide whether to use one of these values or to calculate an average. In addition to the reference, there is also the 'Standards/Precision' link. This link takes you to the values measured in the respective work on reference samples and to information on the measurement accuracy. The complete references are listed below the table.

3.2 Query by Geological Setting

The **geological query** starts with a list of different geological settings to which the individual localities are assigned. Localities of a geological setting are described in GEOROC by certain location types. Table 2 shows some examples.

The sample selection in the locality-based query starts with the most general hierarchy and can then be further restricted to samples from an island, volcano, or sometimes even to samples from a rock series, a rift zone or a specific volcanic stage. An example for the Hawaiian Islands is shown in Fig. 10. The level of detail supported within the search depends on the respective locality. In the case of the Hawaiian Islands, it makes sense to search for samples from individual volcanoes, since all samples recorded in the database can be assigned to a volcano. In the case of oceanic islands such as Iceland or the Canary Islands, a significant proportion of the samples cannot be assigned to any

Geological Setting	Location Type	Example
Ocean Island	ocean-island group	Hawaiian Islands
	ocean island	Hawaiian
	volcano	Kīlauea
	flow/formation/series	Puna Basalt
	rift zone	East Rift Zone
	volcanic stage	Shield Stage
Oceanic Plateau	plateau	Manihiki Plateau
	plateau subunit	High Plateau
	site	Site 317
	flow/formation/series	
Continental Flood Basalt	province	Paraná
	province subunit	Southern Paraná Basin
	flow/formation/series	Guatá - Bom Jardim Series

Table 2: Examples of location types within geological settings.

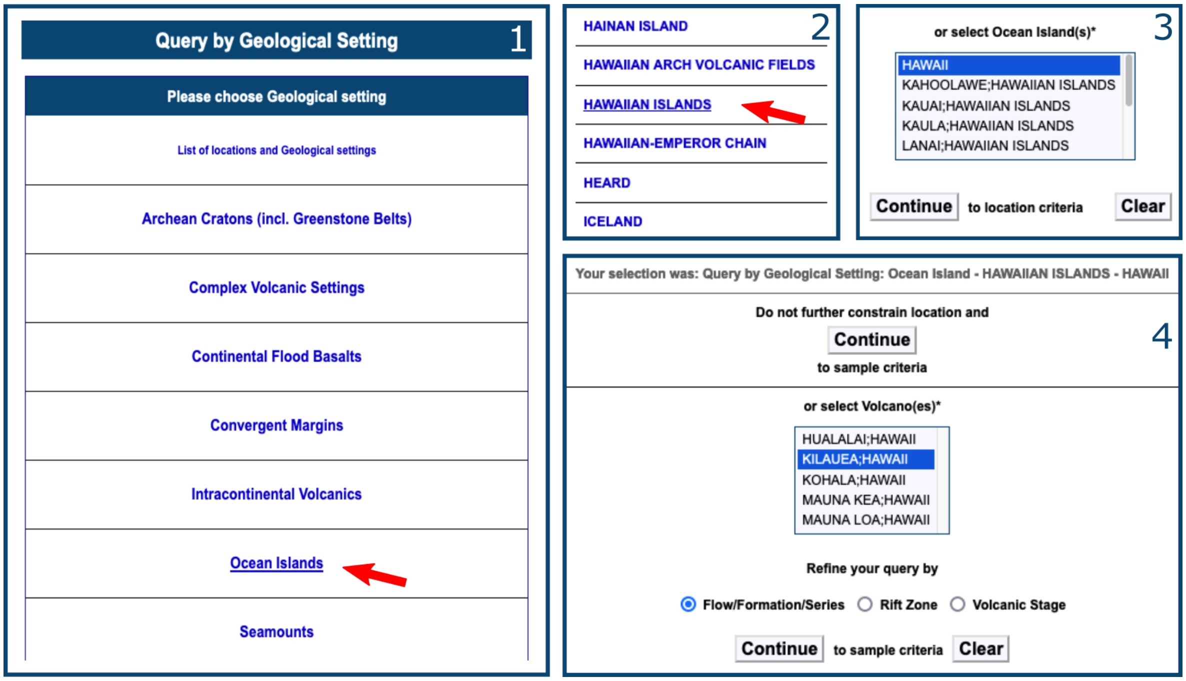


Figure 10: Query by Geological Setting: steps 1 through 4 selecting geological setting and location type.

particular volcano. It is therefore advisable to go straight from the island level to the sample selection and possibly make further restrictions there.

After selecting the desired geological setting and location type, the query continues with sample and metadata selection as described for the bibliographic query above (Sec. 3).

3.3 Query by Geography

If you are interested in a certain geographic region that transcends several geological settings, you can define your region of interest by specifying latitude and longitude (Fig. 11). If the coordinates are known, they can be directly entered into the respective fields for minimum and maximum latitude/longitude. Make sure to enter the values in decimal degrees.

Figure 11: Query by Geography: set coordinates of region of interest.

Alternatively, a rectangular region can be defined on an interactive map (Fig. 12). The map shows the distribution of sample locations available in the database. At low zoom levels, multiple locations are grouped (indicated by round, coloured circles). At high zoom, individual locations are shown and a pop-up box displays additional information about the samples at this location. Several options are available as the base map, including satellite and topographic maps.

In both cases, after confirming your selection this search route takes you straight to the sample and metadata selection as described for the bibliographic query above (Sec. 3).

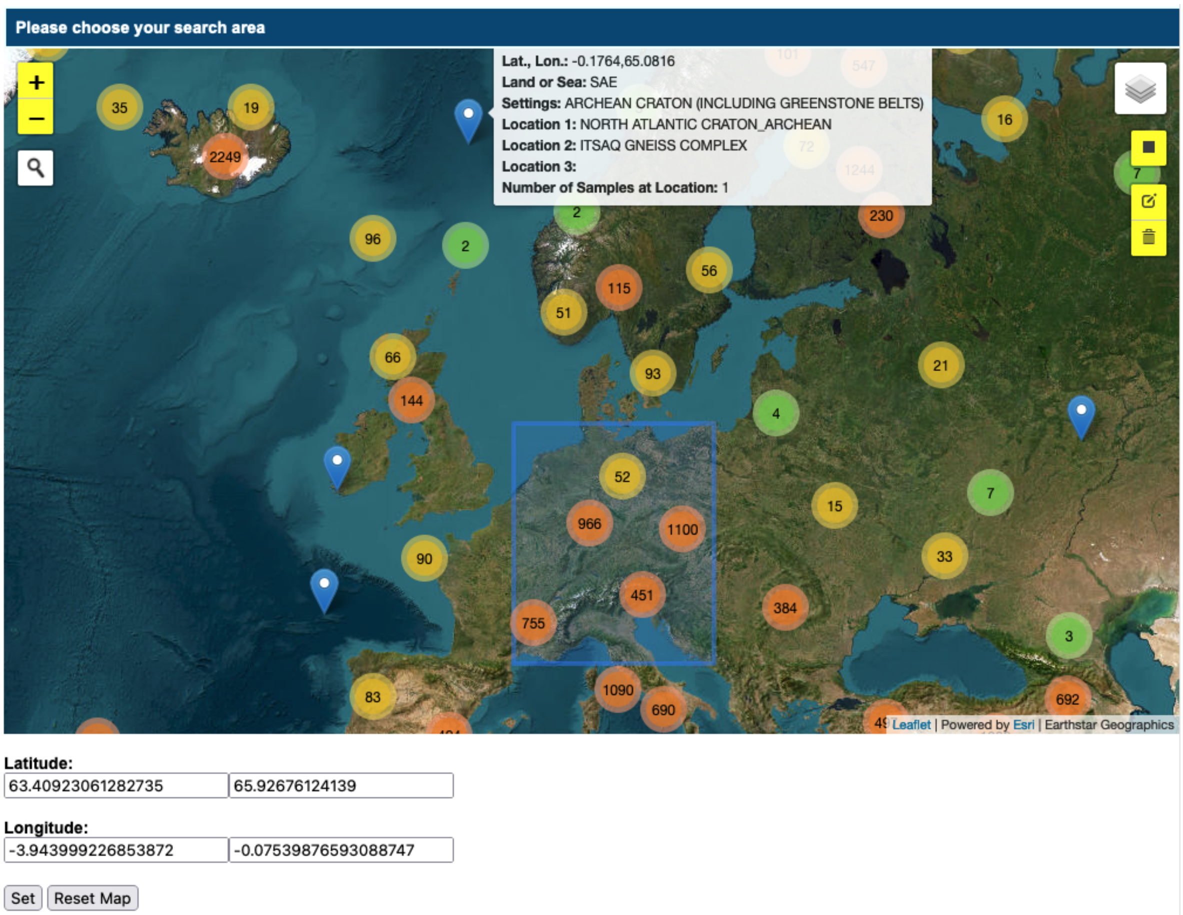


Figure 12: Query by Geography: select region of interest on interactive map showing sample distribution.

3.4 Query by Petrography

The **petrographic query** in GEOROC starts with rock or sample data. You can either select rocks of a rock type, search directly for a specific rock name or for a specific sample or group of samples (Fig. 13).

If you select a rock type, for example *Volcanic Rocks*, the next window shows a list of rock names of this type. If you move the cursor over these names, a sub-menu with additional rock names that fall within these groups may open (Fig. 14). You can either select a general name in the list on the left or one of the specific entries on the right.

NOTE: rock names in GEOROC always correspond to those of the original publication.

When entering a rock name in the field provided, it is recommended to only enter a

Figure 13: Query by Petrography: select rock type or specific sample(s).

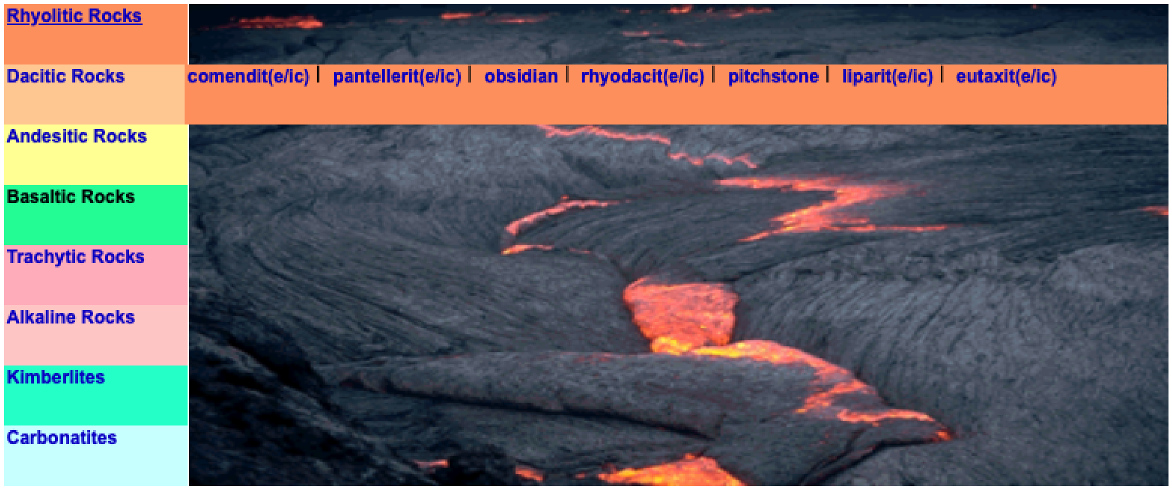
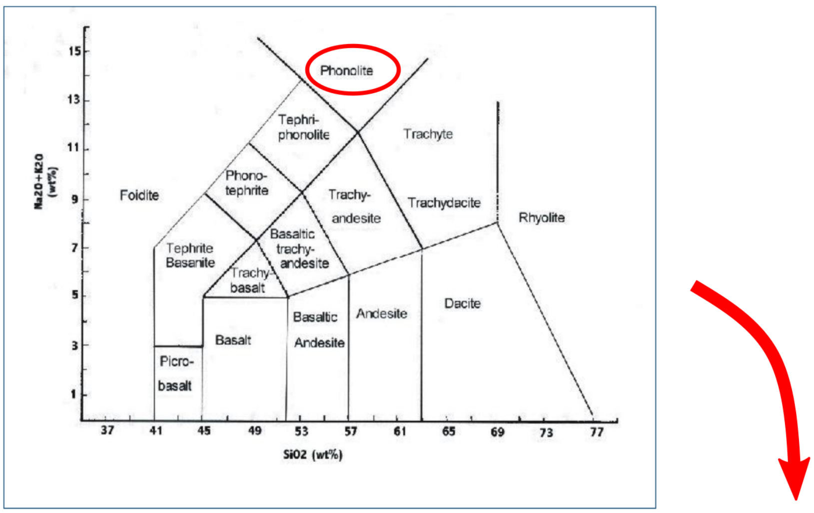


Figure 14: Query by Petrography: select rock name.

part of the name (a substring), for example *andesit* instead of *andesite*. The result will then contain samples with the designation *andesite*, but also others such as *basalt, andesitic* or *andesite, boninitic*. The query for a rock name is primarily recommended for exotic rock types. Searches for names like *basalt* should be avoided if possible since the result will likely contain thousands of values.

Another option in the petrographic query is to search by sample name (Fig. 13). Here, too, you can enter an exact sample name or a substring, for example *ML* to select a specific group of samples from Mauna Loa volcano. The next window shows a list of all samples with the letter combination *ML* in the sample name. From this list (which will not only contain Mauna Loa samples) the desired samples can be selected.

Samples of volcanic rocks can also be selected using the SiO_2 – (Na_2O+K_2O) classification diagram (TAS diagram). It will be particularly apparent here that the rock name of the sample stored in the database corresponds to that of the original reference and may occasionally differ significantly from the name assigned based on the TAS diagram. Figure 15 shows an example for *phonolite*.



after LE BAS et al. [1985]

CITATION	SAMPLE NAME	UNIQUE_ID	ROCK NAME	ND143_ND144
WEIS D. (1993) [1008] Standards/Precision(Info)	LVLK121		PHONOLITE	0.51265 ± 8e-5(2σ)
WEIS D. (1993) [1008] Standards/Precision(Info)	LVLK81	3019-LVLK 81	PHONOLITE	0.51256 ± 2e-5(2σ)
WEIS D. (1993) [1008] Standards/Precision(Info)	LVLK83		PHONOLITE	0.51256 ± 3e-5(2σ)
WEIS D. (1993) [1008] Standards/Precision(Info)	LVLK86		PHONOLITE	0.51251 ± 2e-5(2σ)
DEVEY C. W. (1990) [1457] Standards/Precision(Info)	4-124		BASALT	0.51261 ± 2.3e-5(2σ)
DEVEY C. W. (1990) [1457] Standards/Precision(Info)	4-126		BASALT	0.512607 ± 2.3e-5(2σ)
DEVEY C. W. (1990) [1457] Standards/Precision(Info)	4-2	35082-4-2	BASALT	0.512582 ± 3.1e-5(2σ)
DEVEY C. W. (1990) [1457] Standards/Precision(Info)	4-3	35082-4-3	BASALT	0.512612 ± 2.2e-5(2σ)
MARQUES L. S. (1999) [1498] Standards/Precision(Info)	MV-2	5648-MV-2	PHONOLITE, PERALKALINE	0.512879 ± 3.9e-5(2σ)
MARQUES L. S. (1999) [1498] Standards/Precision(Info)	TR-88	1455-TR-88	PHONOLITE	
MARQUES L. S. (1999) [1498] Standards/Precision(Info)	TR-97		PHONOLITE	
MARQUES L. S. (1999) [1498] Standards/Precision(Info)	TR-98	1455-TR-98	PHONOLITE	0.512826 ± 4.2e-5(2σ)
LE ROEX A. P. (1998) [2421] Standards/Precision(Info)	OKJ-24		TINGUAITE	
LEBRON M. C. (1994) [2633] Standards/Precision(Info)	RU-20		TRACHYTE	0.512831

Figure 15: Query by Petrography: select samples based on TAS diagram.

3.5 Query by Chemistry

The **chemical query** is divided into two parts: measured concentrations (or isotopic ratios) and normalised trace element values can be queried. Several literature sources are available for the various reservoirs used for normalisation (Fig. 16).

The query for measured values begins with a list of all available chemical elements and isotopic ratios (Fig. 17). If you choose normalisation instead, you will only see a lists of trace elements and rare earth elements. In both cases, a publication period and the analysed material must be selected on this page. The query is then carried out for the samples of a particular geological setting. If frequently measured element concentrations or isotope ratios are requested, it is advisable to split the search over several publication periods. The individual result files can be easily combined again if

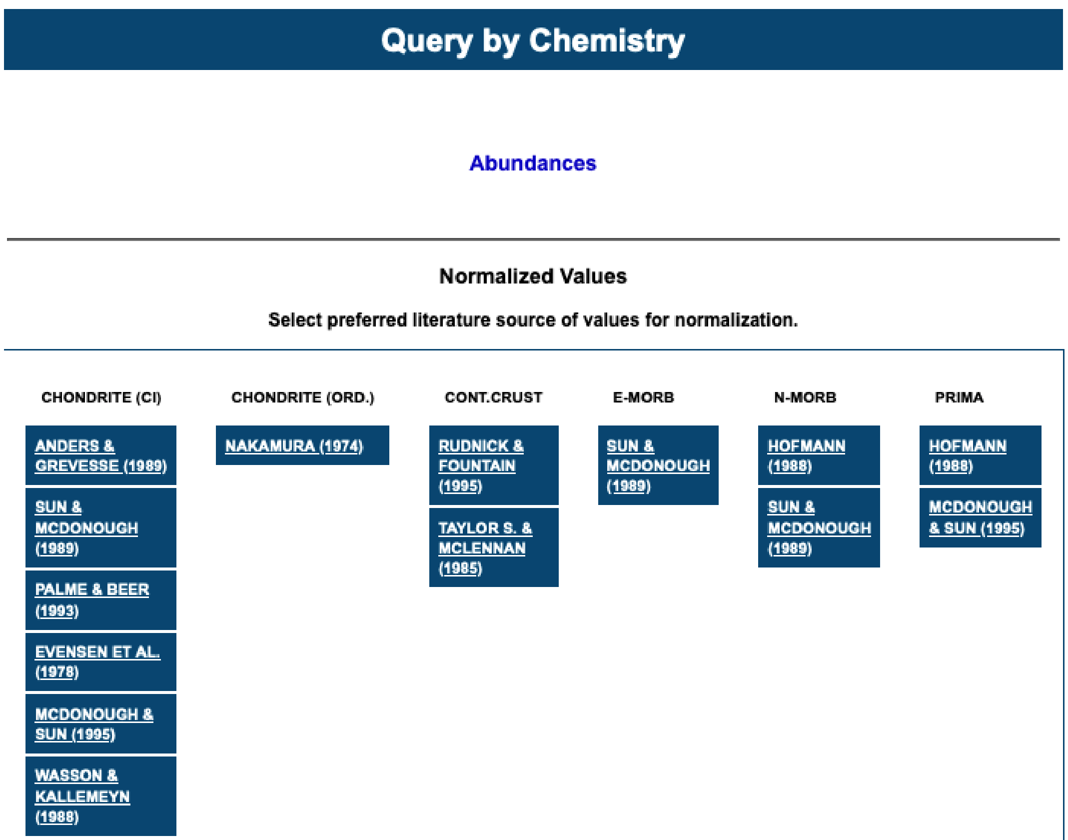


Figure 16: Query by Chemistry: search measured or normalised concentrations.

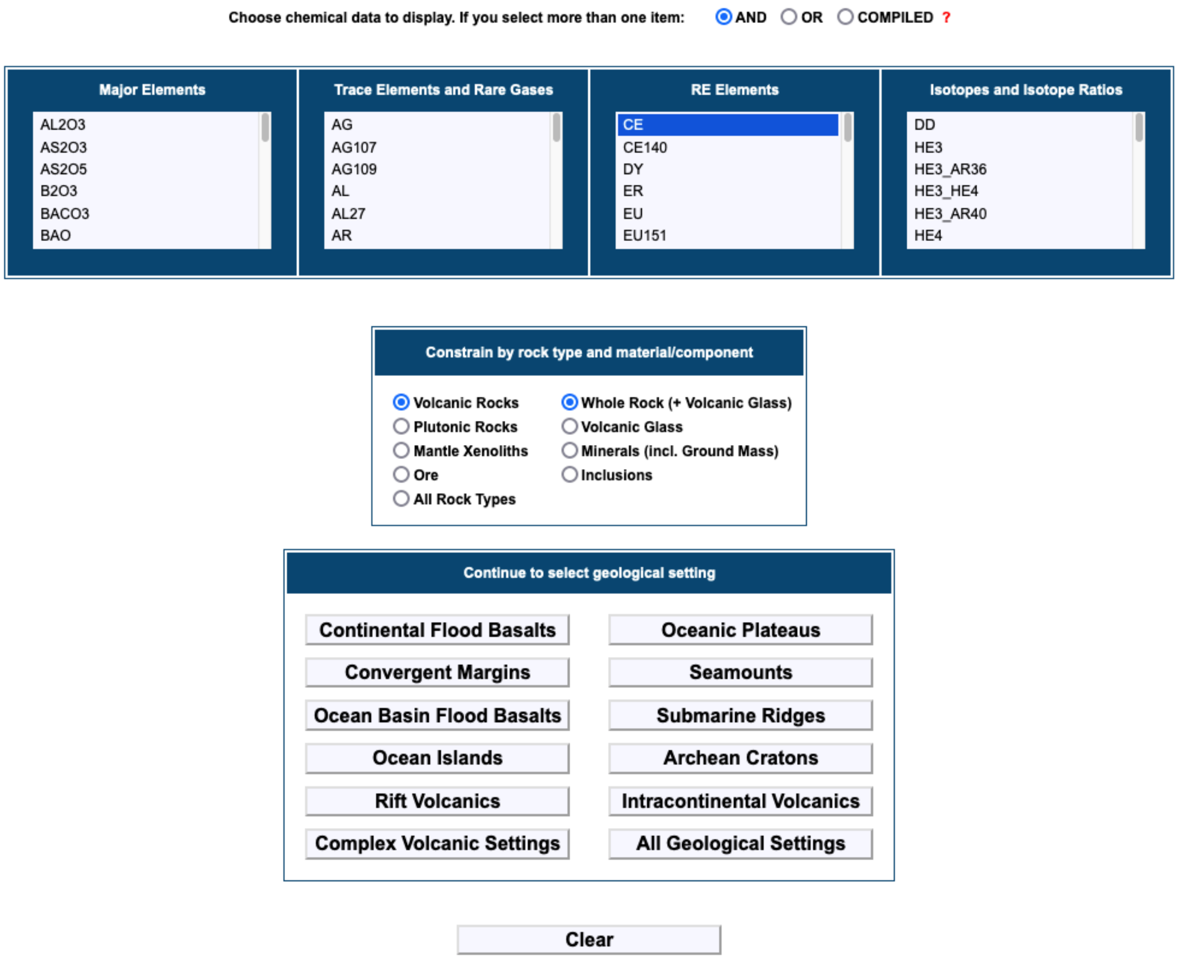


Figure 17: Query by Chemistry: refine search by choosing analytes, rock types and geological setting.

the same or the standard output format is chosen (see Sec. 3 and Fig. 8).

In the next window, all locations of samples are listed for which the selected elements have been measured or are recorded in the database. The location selection is hierarchical again. Before the search result is displayed, you only have to select the metadata that is to be displayed with the result (see Sec. 3).