

Geochemical database of Japanese islands for data-driven science: problem and solution of published domestic data

*Satoru Haraguchi¹, Kenta Ueki², Kenta Yoshida¹, Tatsu Kuwatani¹, Hikaru Iwamori¹

1. Japan Agency for Marine-Earth Science and Technology, 2. Earthquake Research Institute, University of Tokyo

Geochemical database of Japanese islands for data-driven science: problem and solution of published domestic data

Haraguchi, S., Ueki, K.¹, Yoshida, K., Kuwatani, T. & Iwamori, H.

Japan Agency for Marine-earth Science and Technology

¹ Earthquake Research Institute, The University of Tokyo

Recent progresses in the earth science have introjected a data-driven approach, which deals with a huge amount of high-dimensional data. Gathering several kinds of data, e.g. major element composition, trace element composition and isotopic ratio, with GPS-position data can provide a better understanding of the geological phenomena. For such kind of big data science, several databases covering geochemical, geochronological, and petrological data are constructed and available on the internet, such as PetDB, SedDB and GeoRoc. These databases are constructed by collecting a number of published data. However, the existing international databases are based on the collection of studies in the international publication of societies and projects. The data in the publication of non-international journals and domestic institutes are rare in the international datasets. As a result, compositional data of rocks from Japan arc are not fully covered by these databases. Therefore, we construct a geochemical database for those published in the Japanese language and provided in domestic journals.

We collected literature including geochemical data of Japanese islands published from 1980's to present. During the data compilation, we took special care of position data. Position data are crucial for understanding geographical distribution of geochemical components, and also important to estimate geo-neutrino flux from the crust. Increasing availability of handy GPS logger provides easy access to the standardized position data. However, many of the published geochemical data still lack GPS-position data, and of course, old publications provided position information only by map images. We use Google Earth for reading position data from map images. The geological map of papers including sampling points is projected on Google Earth. Latitude, longitude and altitude of the sampling points on the map are obtained based on the coordinate system on the Google Earth.

The articles published in 1980's and 1990's are usually provided as scanned PDF of a printed paper, and thus, tables are provided as images. Furthermore, some journals still provide tables as images. Despite the developed OCR (optical character recognition) technique, the transformation of image tables to numerical data requires checking by eyes of a human. In addition, pdf files provided by some publishers (e.g. J-stage) are protected and data in the file are not directly available for computer-based treatments. To overcome these problems, we have edited this new geochemical database.

Our database covers several rock types including volcanic rocks, plutonic rocks, sedimentary rocks, metamorphic rocks, and so on, and from Hokkaido to Kyushu Island (Figure 1). The main purpose of our database is “basement characteristics” of the Japanese islands: therefore, we pay great attention to collecting data mainly plutonic, metamorphic and sedimentary rocks. One of the important points of our database is covering rock types other than the volcanic rocks, because many parts of the existing geochemical database is focused on volcanic rocks, and metamorphic and sedimentary rock data are rare.

The compiled database is going to be published under the appropriate science commons license, although the idea of *science commons* has not been widely accepted among the geoscientists (Watanabe & Noguchi, 2010). We will provide information about our database on the website (<http://dsap.jamstec.go.jp>).

This database will be used to multiple purposes, such as multivariable statistical analyses (Iwamori et al., 2017), estimateon of the average crustal compositeon (e.g. Togashi et al., 2000), geographyc statistical analyses and the estimateon of the crustal geo-neutrino flux (Enomoto et al., 2007).

Keywords: position data, published geochemical data, Basement of Japanese islands, geochemical database

