

Geochemical database of Japanese islands based on published domestic data: standardization of metadata and chemical data

*Satoru Haraguchi¹, Kenta Ueki^{1,2}, Tatsu Kuwatani^{1,3}, Kenta Yoshida¹, Mika Mohammad¹, Shunsuke Horiuchi¹, Hikaru Iwamori^{1,4}

1. Japan Agency for Marine-Earth Science and Technology, 2. Earthquake Research Institute, University of Tokyo, 3. PRESTO, Japan Science and Technology Agency (JST), 4. Department of Earth and Planetary Sciences, Tokyo Institute of Technology

We reported the database "DODAI" which compiled geochemistry data published in the domestic journals/bulltins and its significance for the data-driven science in JPGU-AGU2017 (Haraguchi et al., 2017). We continued data collection after the report and compiled chemical data of 5818 samples from 224 articles at present. We are reporting problems in conjunction with "the unification of the format" found out through these data collection.

The first is the change of major method used in analysis of the chemical composition. For the major element compositions, "wet chemical analysis" was mainly used for bulk chemistry until the 1970s, but after the 1980s, "X-ray Fluorescence (XRF) method" spread out rapidly. In the wet chemistry, iron was separately reported as Fe²⁺ (FeO) and Fe³⁺ (Fe₂O₃), while XRF treat it as all FeO or Fe₂O₃. Some papers still continue to determine Fe²⁺/Fe³⁺ by wet chemical analysis, combined with other composition measured by XRF and so on. This change of analytical method needs to be taken into account for constructing a geochemical database.

The second is the various assemblages of analyzed trace elements are existences the published data sets. Mass spectrometry such as ICP-MS and neutron activation analyses such as INAA are commonly used for trace element analysis as well as XRF. Although XRF can carry out analysis for major and trace element together, many laboratories and researchers employ various analytical elements and methods depending on their purpose. Generally, ICP-MS and INAA can analyze a number of trace elements synchronously, but there are few laboratories constantly reporting all possible elements as a standard dataset, resulting in a little number of the "complete" dataset. Therefore, depending on the assemblage of elements chosen for multivariate analysis, the number of the available samples may drastically decrease.

The third is that the descriptive terms of geology have been changed by the time and research field where a study was conducted. For example, "Accretionary prism" represented by Shimanto belt was rapidly recognized after 1970s. Therefore, the geological feature of the relevant unit was interpreted based on the idea of accretionary prism recently, while very old papers have provided the description based on other concepts. In addition, the volcanic rocks, originated from the subducted ocean plate, in accretionary belts are sometimes called "the greenstone" due to its signature of weak metamorphism, but descriptions of "volcanic rock" and "metamorphic rock" are used together by researchers in different research fields. Some researchers have struggled to unify these descriptions (e.g., Seamless Digital Geological Map Unified Legend by GSJ AIST, 2015). These differences in in descriptive terms should be carefully treated in constructing a geochemical database.

To compare datasets collected from different papers, standardized description and unified data format are required. In this report, we will show the examples of the abovementioned problems found in our constructed geochemical database and discuss the possible solutions.

Keywords: Geochemical database, Basement rock of Japan, Data-driven science, Statistical analysis, Geocoding