小規模データベースの公開と保存
How to make the data sets in "dark long tail" open and preserve?

*家森 俊彦[1]
*Toshihiko Iyemori[1]

1.京都大学大学院理学研究科付属地磁気世界資料解析センター
1.Data Analysis Center for Geomagnetism and Space Magnetism, Graduate School of Science, Kyoto University

In data analysis, we often encounter the difficulty by lack of data and try to find additional data sets by asking researchers in the same research community. Sometimes, we can reach the data set suitable to fill the gap of data or we find unexpected data set which is very useful. However, in most cases, we cannot find the data. We know that there are a huge number of datasets—mainly obtained on a research project basis—that are not registered to active data centres, and hence are 'dark' to many of us. These datasets are typically built by small research groups for a limited period, and data are not open for public. Although they exist only for a limited period, such data are very important and useful if the location of observation site is highly unique, or if other observations are not available.
One way to make such data sets open from the 'dark long tail' is to register metadata that describe the observations in as much detail as possible. An example of this in practice is IUGONET (Interuniversity Upper atmosphere Global Observation NETwork), which has a common database of metadata and forms a virtual data centre of distributed databases at several institutions. This data system includes databases from the 'dark long tail', as well as large well-known databases. Another way is to use university repositories. However, in this case, we need a common method to find and retrieve the data set.

キーワード：データ公開、データ保存、小規模データベース
Keywords: open data, data preservation, small data set