

## PRACTICE PAPER

# Unified Geomagnetic Database from Different Observation Networks for Geomagnetic Hazard Assessment Tasks

Michael Dobrovolsky, Dmitriy Kudin, Roman Krasnoperov

Geophysical Center RAS, Moscow, RU

Corresponding author: Michael Dobrovolsky ([m.dobrovolsky@gcras.ru](mailto:m.dobrovolsky@gcras.ru))

---

This paper presents the results of the creation of a geomagnetic data storage system that combines raw observation data from different geomagnetic observation networks along with derived indices and indicators of geomagnetic activity. Geomagnetic data, provided by observation networks, undergo a series of data quality validation procedures. The implemented instruments and procedures facilitate the creation of a uniform database from multiple sources.

---

**Keywords:** database; geomagnetism; geomagnetic indices; geomagnetic activity; data quality assessment; space weather

---

## Introduction

Observations of the Earth's magnetic field parameters are essential for fundamental research of solar-terrestrial interactions. Another significant application is supporting the stable operation of various technical systems and industrial facilities and assessing the risks of their exposure to extreme space weather events. Disturbances in the Earth's magnetosphere and ionosphere have been known to cause failures in radio transmission infrastructure and satellite navigation systems, as well as damage of industrial structures, such as power and communication lines, power stations, oil and gas pipelines, railroad equipment.

Therefore, evaluation of possible characteristics of geomagnetic disturbances and geomagnetically induced currents (GICs) for a given region is an important task. A study done by Pirjola (2002) describes the occurrence of erroneous traffic signals on railways caused by geomagnetically induced currents. Other studies have reported the effects of GIC's on electric power grids (Belakhovsky et al. 2019), trans-oceanic cables (Lanzerotti et al. 1995) and pipelines (Pulkkinen et al. 2001).

To determine the various parameters of geomagnetic activity for given territories, access to long time series of geomagnetic measurements of sufficiently good quality is required. Geomagnetic observatories and stations are included in various observational networks such as INTERMAGNET, SuperMAG, IMAGE, etc. They have different spatial coverage and provide data of varying quality in diverse formats with different orientations of magnetometric instruments. These aspects complicate combining data into a single database in a uniform format. This article is dedicated to solving this problem. It describes a unified database that integrates data from different geomagnetic observation networks and brings them into a uniform format that is convenient for further analysis.

In addition to direct observational data other significant source of information on the Earth's magnetosphere are indices of geomagnetic activity (e.g., Kp, Dst, AE, SYM, etc.). It should be noted that these global indices of geomagnetic activity are insufficient to estimate GICs. It is necessary to construct regional indices that more accurately take into account the geomagnetic field disturbances in the regions of interest (Viljanen, Wintoft & Wik 2015). To solve this problem, it is necessary to integrate into a single database information from multiple observational networks located within the studied region. This requires adequate tools to aggregate observational data into a unified data storage. In addition to trivial conversion from multiple data formats used in different repositories, for certain networks it is also necessary to transform the reference orientation of magnetic observations, which are orthogonal components of the magnetic field vector.

Estimation of possible GIC's values requires construction of regional indicators of geomagnetic activity that are based on geomagnetic observations. It is also necessary to select from various indicators those that allow for the best estimation of emerging currents. To facilitate the calculation of such indicators using initial data it is convenient to organize data storage in a relational database instead of file storage systems that are still used in many repositories of geomagnetic data. In this article we present solutions for solving the abovementioned problems.

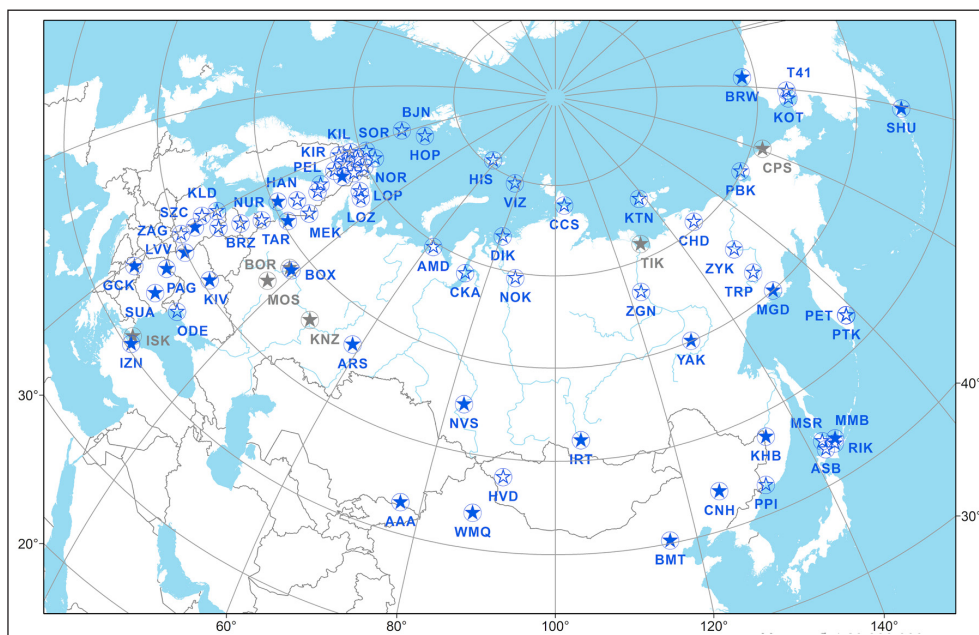
## Data Sources

For more convenient interaction with data on the Earth's magnetic field, it is important to have access to a uniform database of initial geomagnetic data (magnetograms) that consist of measurements of the magnetic field vector orthogonal components. Magnetometers, installed at geomagnetic observatories or stations, can be oriented with reference to geographic (X – North; Y – East; Z – Vertical) or geomagnetic (H – Magnetic North; E – Magnetic East; Z – Vertical) coordinate systems. The latter depends on the location of the geomagnetic pole that changes with time. Data from magnetic observatories, where full magnetic field vector components are determined, can be easily transformed between the systems. Magnetic stations provide only variations of magnetic vector components, and in this case transformation of observations requires additional calculations. To facilitate further data processing and analysis of data both from observatories and stations it is convenient to provide storage in a geographic reference frame that is stable in time.

The spatial coverage of the database described in this paper is the territory of Russia and neighboring countries.

The main source of observatory geomagnetic data is the international network of geomagnetic observatories INTERMAGNET (INTERMAGNET 2020). We selected 27 observatories from this network located in Russia and neighboring countries. Data ingested before being quality controlled is considered 'preliminary'. These data include missing values and disturbances caused by man-made interference. Observational data that went through a baseline correction procedure, removing disturbances, and filling data gaps, are considered 'definitive data'. If available, the use of the definitive data is preferred. For the selected observatories of the INTERMAGNET network, the definitive data from 1991 to 2019 were obtained.

The SuperMAG network was chosen as the main source of data for magnetic stations that are located in the Northern regions (Gjerloev 2012; SuperMAG 2020). We selected 76 SuperMAG stations located in Russia and neighboring countries. A map of selected stations shown in **Figure 1** is published on a GIS server and is available on the Internet as a cartographic web service (Map of Geomagnetic Observatories and Stations 2018). For the selected stations data for the entire available observation period from 1980 to 2019 were used.



**Figure 1:** Map of selected SuperMAG stations shown with hollow stars. INTERMAGNET stations are shown with filled stars.

SuperMAG data was also supplemented by data from other geomagnetic observation networks. From the IMAGE network (IMAGE 2020; Tanskanen 2009), data for 12 stations was selected. The selected time interval was 2007–2018 (2010–2015 for certain stations).

The observation network of the Arctic and Antarctic Research Institute (AARI) provided data for 8 Russian polar stations for the period from 2007 to early 2018. From observation network of the Pushkov Institute of Terrestrial Magnetism, Ionosphere and Radiowave Propagation of the Russian Academy of Sciences (IZMIRAN), data for Kaliningrad (KLD) (2013–2016) and Moscow (MOS) (2009–2011) were included into the database.

As a source of data for Russian observatories and stations, the Russian-Ukrainian Geomagnetic Data Center (RUGDC 2020) was also used. It is the part of the Analytical Geomagnetic Data Center, Geophysical Center of the Russian Academy of Sciences (AGDC GC RAS 2020).

In addition to the direct observational data, geomagnetic activity indices calculated on their basis were collected: AE, AU, AL, AO indices for 1980–2019; ASY-D, ASY-H, SYM-D, SYM-H indices for 1981–2019; Kp-index for the years 2001–2019. Data were taken from the Kyoto World Data Center for Geomagnetism (WDC Kyoto 2020). Also, for the selected observatories, data on the K-index for 2005–2019 were used. K-index data for the Borok Observatory (BOX) were downloaded from the observatory web site and the World Data Center for Solar-Terrestrial Physics (WDC for STP 2019).

## Database Structure

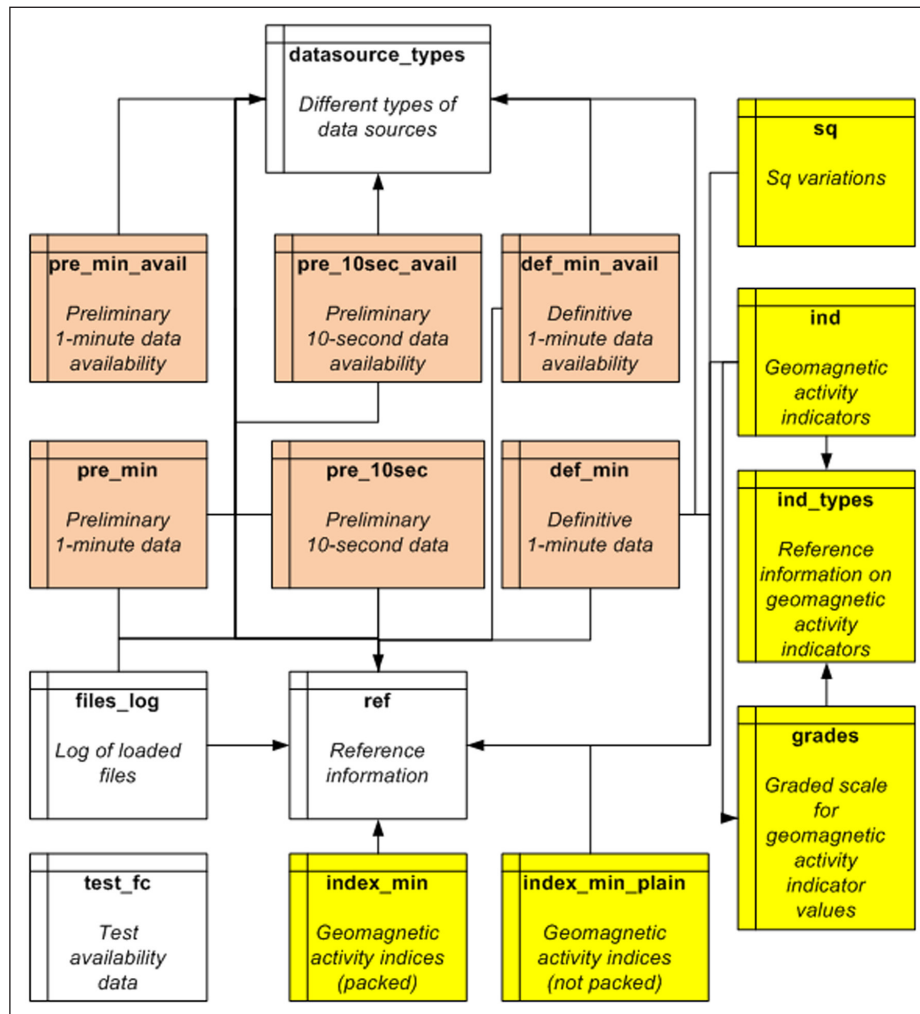
The database was created in two versions: file storage of initial data and MySQL relational database. The standard IAGA-2002 file format (IAGA-2002 2016) and XYZ magnetic field component orientation were chosen as the data format. Files obtained from different geomagnetic observation networks in different formats and various orientations were converted to IAGA-2002 file format and, when necessary, the magnetic field component variations were converted to the XYZ coordinate system. For data conversion the freely available Geomag Algorithms library (Geomag Algorithms 2020) was used. The functionality of reading the necessary data formats and converting the coordinate system to XYZ was added to the library. After conversion, the data were loaded to the MySQL relational database.

The structure of the developed relational database rests upon one of the data storage systems described by Gvishiani et al. (2016). But in contrast to its predecessor versions it has significant alterations that allow storing data from different observation networks. Evaluation of the completeness of observational data was also performed. Initial geomagnetic measurements are stored along with geomagnetic activity indices and indicators. Geomagnetic activity indices were obtained from external sources and geomagnetic activity indicators were calculated using initial measurements. The database structure is shown in **Figure 2**.

The database consists of the following tables:

- ref—reference information,
- datasource\_types—types of data sources,
- pre\_min—preliminary 1-minute data,
- pre\_10sec—preliminary 10-second data,
- def\_min—definitive 1-minute data,
- pre\_min\_avail—time intervals of availability of preliminary 1-minute data,
- pre\_10sec\_avail—time intervals of availability of preliminary 10-second data,
- def\_min\_avail—time intervals of availability of definitive 1-minute data,
- index\_min—indices of geomagnetic activity in packed binary format,
- index\_min\_plain—indices of geomagnetic activity without packing,
- ind—indicators of geomagnetic activity,
- ind\_types—types of indicators of geomagnetic activity,
- grades—graded scales of geomagnetic activity indicators,
- sq—values of solar quiet variation,
- files\_log—source data files loaded into the database,
- test\_fc—test information about the number of missing values in the data.

The data itself is contained in the tables pre\_min, pre\_10sec and def\_min. To increase the query speed, data is stored in an hourly-packed binary format. An important element in organizing data storage is the data-source field, which stores the data source code. This allows to store data from different sources for the same observatory or station. At the same time, it is possible to get access to all data for an observatory or station



**Figure 2:** Database structure. Orange—tables of initial data, yellow—indices of geomagnetic activity calculated on their basis, white—auxiliary tables.

obtained from different sources or to request data only from one specific source. This is the unique feature of the proposed data storage system. The data sources themselves are stored in the `datasource_types` table.

Along with the initial observational data, the database also hosts indices and indicators of geomagnetic activity based on these data. To store indices, which are ready-made from other data sources, the `index_min` and `index_min_plain` tables are used to store packed and unpacked data, respectively. To store the indicators of geomagnetic activity, which are calculated in the system itself, the `ind` table is used. Indicator types are stored in the `ind_types` table. The following types of geomagnetic activity indicators are calculated and stored in the database: amplitude, rate of change  $dB/dt$ , the measure of anomalousness (Soloviev, Agayan & Bogoutdinov 2016). For each indicator and each observatory or station, an individual scale of geomagnetic activity is calculated using the available data. The scales calculated in this way are stored in the `grades` table. The presence of such scales makes it possible to single out moments of geomagnetic activity uniformly and simultaneously at different observation points and for different indicators.

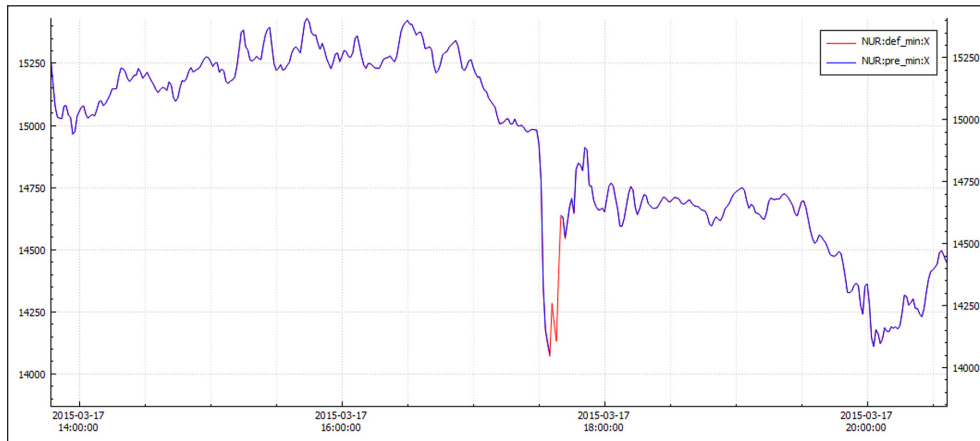
Since we do not have permission to provide online access to data from all used observational networks yet, only the online service for determining the availability of data stored in the database is now available (RSF4 2020). In the future, after obtaining the necessary permissions, it is planned to organize online access to the database similar to RUGDC (2020).

## Data Quality Assessment

To assess the quality of data from the different observation networks integrated into the database, the data sources have been ranked according to the information on processing and verification of the geomagnetic observations. The evaluation of the data source was based on the following criteria:

**Table 1:** Evaluation of the data source quality.

Data source name	Computer processing	Expert processing	Data requirements	Data verification	Score
IMAGE	Yes	Yes	Yes	Internal	4
INTERMAGNET	Yes	Yes	Yes	External	5
SuperMAG	Yes	No	No	No	1
GC RAS	Yes	No	No	No	1
IZMIRAN	No	No	No	No	0



**Figure 3:** Comparison of data from the horizontal X component of the Nurmijärvi Observatory (NUR) from SuperMAG (blue) and INTERMAGNET (red).

1. Use of automatic algorithms to control and correct data quality (computer processing–1 point).
2. Carrying out manual verification of data by qualified specialists (expert processing–1 point).
3. Availability of a document describing the recommendations on data quality (data requirements–1 point).
4. Data verification by an independent expert (data verification–1–2 points).

The values of the criteria for the INTERMAGNET network are obtained from INTERMAGNET (2020) and St-Louis et al. (2012). The INTERMAGNET data is evaluated in three stages by independent experts from around the world. The data quality requirements of the IMAGE network are described by Viljanen & Hakkinen (1997). **Table 1** shows the data sources used to create the database, with their evaluation according to the above criteria.

Lack of expert processing and verification of data may lead to both omissions of abnormal perturbations in the data and removal of periods containing natural field changes. **Figure 3** shows a comparison between the data provided by SuperMAG and the definitive INTERMAGNET data for the horizontal X component of the Nurmijärvi (NUR) Observatory. As can be seen from the figure, there is a gap in SuperMAG data from 17:35 to 17:40 March 17, 2015. This gap is most likely associated with the false detection of man-made disturbances by the automatic SuperMAG algorithm.

## Conclusion

When working with geomagnetic data it is necessary to have data in a uniform format and in a uniform coordinate system. However, existing geomagnetic observation networks have different spatial coverage, use different data formats and coordinate systems. The quality of data provided by different networks also varies considerably. Therefore, we have attempted to create a uniform geomagnetic database that combines data from different observation networks in a single format and coordinate system. It also must be possible to differentiate data from the different networks so that the quality of the data provided by them can be taken into account.

The process of developing the database provided a set of tools needed for its creation. Specifically: tools for converting the data to a unified storage format, performing the necessary coordinate transformations, loading the obtained data into a relational database, calculation of geomagnetic activity indicators for further analysis on the basis of initial data.

For the purposes of our project we created the geomagnetic database for the territory of Russia and neighboring states. Since we did not study individual events, but performed a retrospective statistical analysis, such a database was adequate to the tasks to be solved. At the same time, nothing prevents to create similar database for any other territory or even for the whole world. The created database will contain all necessary information for further analysis of geomagnetic activity in the studied region.

The database for the territory of Russia and neighboring countries was used in the framework of the Russian Science Foundation project (17-77-20034) to produce maps of the possible magnitude of geomagnetically induced currents for the territory of the country.

## Acknowledgements

This work employed facilities and data provided by the Shared Research Facility “Analytical Geomagnetic Data Center” of the Geophysical Center of RAS (AGDC GC RAS 2020).

We thank the institutes who maintain the IMAGE Magnetometer Array.

The results presented in this paper rely on data collected at magnetic observatories. We thank the national institutes that support them and INTERMAGNET for promoting high standards of magnetic observatory practice (INTERMAGNET 2020).

We are grateful to all the Partners, which provide geomagnetic observations and the results of the preliminary analysis to the RUGDC.

We gratefully acknowledge the SuperMAG collaborators (SuperMAG: Acknowledgement 2020).

The indices used in this paper were provided by the WDC for Geomagnetism, Kyoto (WDC Kyoto 2020).

We would like to thank the World Data Center for Solar-Terrestrial Physics for the data on K-index.

## Funding Information

This work was supported by the grant of the Russian Science Foundation no. 17-77-20034.

## Competing Interests

The authors have no competing interests to declare.

## References

- AGDC GC RAS—Analytical Geomagnetic Data Center, Geophysical Center of the Russian Academy of Sciences.** 2020. Available at <http://ckp.gcras.ru/> [Last accessed 27 April 2020].
- Belakhovsky, V, Pilipenko, V, Engebretson, M, Sakharov, Y and Selivanov, V.** 2019. Impulsive disturbances of the geomagnetic field as a cause of induced currents of electric power lines. *Journal of Space Weather and Space Climate*, 9: A18. DOI: <https://doi.org/10.1051/swsc/2019015>
- Geomag Algorithms.** 2020. Available at <https://github.com/usgs/geomag-algorithms> [Last accessed 5 June 2020].
- Gjerloev, J.** 2012. The SuperMAG data processing technique. *Journal of Geophysical Research*, 117: A09213. DOI: <https://doi.org/10.1029/2012JA017683>
- Gvishiani, A, Soloviev, A, Krasnoperov, R and Lukianova, R.** 2016. Automated Hardware and Software System for Monitoring the Earth’s Magnetic Environment. *Data Science Journal*, 15: 18. DOI: <https://doi.org/10.5334/dsj-2016-018>
- IAGA-2002 Exchange Format.** 2016. Available at <https://www.ngdc.noaa.gov/IAGA/vdat/IAGA2002/iaga2002format.html> [Last accessed 5 June 2020].
- IMAGE Magnetometer network.** 2020. Available at <http://space.fmi.fi/image/> [Last accessed 27 April 2020].
- INTERMAGNET—International Real-time Magnetic Observatory Network.** 2020. Available at <https://intermagnet.org/> [Last accessed 27 April 2020].
- Lanzerotti, L, Medford, L, MacLennan, C and Thomson, D.** 1995. Studies of large-scale Earth potentials across oceanic distances. *AT&T Technical Journal*, 74(3): 73–84. DOI: <https://doi.org/10.1002/j.1538-7305.1995.tb00185.x>
- Map of Geomagnetic Observatories and Stations.** 2018. Available at [http://gis.gcras.ru:6080/arcgis/rest/services/RSF4/Geomagnetic\\_observatories\\_stations/MapServer](http://gis.gcras.ru:6080/arcgis/rest/services/RSF4/Geomagnetic_observatories_stations/MapServer) [Last accessed 27 April 2020].

- Pirjola, R.** 2002. Review On The Calculation Of Surface Electric And Magnetic Fields And Of Geomagnetically Induced Currents In Ground-Based Technological Systems. *Surveys in Geophysics*, 23: 71–90. DOI: <https://doi.org/10.1023/A:1014816009303>
- Pulkkinen, A, Viljanen, A, Pajunpää, K and Pirjola, R.** 2001. Recordings and occurrence of geomagnetically induced currents in the Finnish natural gas pipeline network. *Journal of Applied Geophysics*, 48(4): 219–231. DOI: [https://doi.org/10.1016/S0926-9851\(01\)00108-2](https://doi.org/10.1016/S0926-9851(01)00108-2)
- RSF4 DB Data availability.** 2020. Available at <http://geomag.gcras.ru/rsf4db.html> [Last accessed 5 June 2020].
- RUGDC–Russian-Ukrainian Geomagnetic Data Center.** 2020. Available at <http://geomag.gcras.ru/> [Last accessed 27 April 2020].
- Soloviev, A, Agayan, S and Bogoutdinov, SH.** 2016. Estimation of geomagnetic activity using measure of anomalousness. *Annals of Geophysics*, 59(6): G0653. DOI: <https://doi.org/10.4401/ag-7116>
- St-Louis, BJ, Sauter, EA, Trigg, DF, et al.** 2012. *INTERMAGNET Technical Reference Manual*. Edinburgh: INTERMAGNET.
- SuperMAG.** 2020. Available at <http://supermag.jhuapl.edu/> [Last accessed 27 April 2020].
- SuperMAG: Acknowledgement.** 2020. Available at <http://supermag.jhuapl.edu/info/?page=acknowledgement> [Last accessed 5 June 2020].
- Tanskanen, E.** 2009. A comprehensive high-throughput analysis of substorms observed by IMAGE magnetometer network: Years 1993–2003 examined. *Journal of Geophysical Research*, 114(A5): A05204. DOI: <https://doi.org/10.1029/2008JA013682>
- Viljanen, A and Hakkinen, L.** 1997. IMAGE magnetometer network. In: Satellite-Ground Based Coordination Sourcebook. In Lockwood, M, Wild, MN and Opgenoorth, HJ (eds.), *Eur. Space Agency Spec. Publ., ESA-SP 1198*, 111–117.
- Viljanen, A, Wintoft, P and Wik, M.** 2015. Regional estimation of geomagnetically induced currents based on the local magnetic or electric field. *J. Space Weather Space Clim.*, 5: A24. DOI: <https://doi.org/10.1051/swsc/2015022>
- WDC Kyoto–World Data Center for Geomagnetism, Kyoto.** 2020. Available at <http://wdc.kugi.kyoto-u.ac.jp/wdc/Sec3.html> [Last accessed 27 April 2020].
- WDC for STP–World Data Center for Solar-Terrestrial Physics.** 2019. Available at <http://www.wdcb.ru/stp/data.html> [Last accessed 27 April 2020].

**How to cite this article:** Dobrovolsky, M, Kudin, D and Krasnoperov, R. 2020. Unified Geomagnetic Database from Different Observation Networks for Geomagnetic Hazard Assessment Tasks. *Data Science Journal*, 19: 34, pp. 1–7. DOI: <https://doi.org/10.5334/dsj-2020-034>

**Submitted:** 09 May 2020

**Accepted:** 30 July 2020

**Published:** 17 August 2020

**Copyright:** © 2020 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.



*Data Science Journal* is a peer-reviewed open access journal published by Ubiquity Press.

