CDF data archive and integrated data analysis platform for ERG-related ground data developed by ERG Science Center (ERG-SC)

T. Hori[*1], Y. Miyashita[*1], Y. Miyoshi[*1], K. Seki[*1], T. Segawa[*1], Y.-M. Tanaka[*2,*3], K. Keika[*1], M. Shoji[*1], I. Shinohara[*4], K. Shiokawa[*1], Y. Otsuka[*1], S. Abe[*5], A. Yoshikawa[*5], K. Yumoto[*5], Y. Obana[*6], N. Nishitani[*1], A. S. Yukimatu[*2,*3], T. Nagatsuma[*7], M. Kunitake[*7], K. Hosokawa[*8], Y. Ogawa[*2,*3], K. T. Murata[*7], M. Nosé[*9], H. Kawano[*5], and T. Sakanoi[*10]

Abstract

The Exploration of energization and Radiation in Geospace project Science Center (ERG-SC) has developed the science data file archive and integrated data analysis tool for the ground network observation data of the ERG project. We have organized data design consortiums to develop and elaborate the standard metadata and data variable sets for each type of the ERG-related ground data being archived as Common Data Format (CDF) files. The integrated data analysis software for the project has also been developed on the basis of the Space Physics Environment Data Analysis Software (SPEDAS) which works in concert with the CDF data file repository. The software tools, which are provided as plug-in libraries for SPEDAS, are made available to the international science community so that scientists and students are ready to proceed to integrated studies combining the ground data with other satellite and simulation data seamlessly. The integrated data analysis software can make scientific activities more productive and help the ERG project achieve the scientific goals.

Keywords: ERG project, ground network, data archive, integrated analysis software

---

[*1] Solar-Terrestrial Environment Laboratory, Nagoya University, Japan.
[*2] National Institute of Polar Research, Japan.
[*3] SOKENDAI (The Graduate University for Advanced Studies), Japan.
[*4] Institute of Space and Astronautical Science, Japan Aerospace Exploration Agency, Japan.
[*5] International Center for Space Weather Science and Education, Kyushu University, Japan.
[*6] Osaka Electro-Communication University, Japan.
[*7] National Institute of Information and Communications Technology, Japan.
[*8] University of Electro-Communications, Japan.
[*9] Data Analysis Center for Geomagnetism and Space Magnetism, Graduate School of Science, Kyoto University, Japan.
[*10] Planetary Plasma and Atmospheric Research Center, Tohoku University, Japan.

1. INTRODUCTION

The Exploration of energization and Radiation in Geospace (ERG) project is a science project which seeks to explore dynamics of the Earth's inner magnetosphere with a special focus on how the radiation belt particles are generated and then lost by various interactions between particles and electromagnetic waves in the course of space storms. To pursue its scientific goals, the ERG project proceeds as collaborative efforts of three research elements: ERG satellite observation, ground-based network observation, and simulation/integrated studies. The satellite is going to provide in situ measurement of particles and electric/magnetic fields in the inner magnetosphere, while the world-wide network of ground-based observation consisting of, for example, ground magnetometers, ionospheric radars, optical instruments, and radio wave receivers, will supplement the single-point observation of the satellite with the global picture of ionospheric currents, electric field, and particle precipitations. Both observations will be compared from various aspects with numerical simulations/ modeling that reveal important physical processes as well as their spatial structures. Close collaboration of these three efforts, leading to integrated studies on dynamic phenomena in geospace, is a key to comprehensive understanding of relativistic electron generation/loss resulting from cross-energy and cross-regional coupling in which different plasma populations dynamically interact with one another as well as background electric/magnetic fields. A preliminary concept study for the ERG project was reported previously [1] and a more recent, detailed review of the project was also published [2].

It is essential to combine effectively the three different kinds of data for better understanding of the dynamics of the inner magnetosphere. Moreover, facilitating integrated studies combining them in the international science community is crucial to maximizing scientific output of the ERG

project. Taking into consideration this concept, the ERG Science Center (ERG-SC) has launched and started developing key elements of a science center. After the preliminary development of science center functions at Solar-Terrestrial Environment Laboratory (STEL), Nagoya University, ERG-SC has been established in 2013 as a joint research center of space science operated by the Institute of Space and Astronautical Science (ISAS), Japan Aerospace Exploration Agency (JAXA) and STEL.

Figure 1 gives a schematic diagram of the ERG-SC framework by which we bundles the various data and provides a useful data analysis system for the science community, and collaborates with other projects. As stated above, the ERG project data consist of ERG-satellite data, ERG-related ground data, and ERG-related simulation/modeling data. Hereinafter, the latter two data resources are referred to as ERG-ground data and ERG-simulation data, respectively. All of these project data are archived in the science data archive at ERG-SC and are made available to worldwide researchers and students for their scientific activities. Also developed and then released by ERG-SC is integrated data analysis software which provides data users with a unified way to access online, read, manipulate, and visualize various kinds of the project data seamlessly.

2. ERG-GROUND DATA

As evident from the schematic diagram in Figure 1, the science data archive and integrated data analysis tools are the core components to promote the ERG science. Since the ERG-satellite data are present only after the satellite launch in 2016 and the archive of the ERG-simulation data are still in the design phase, first we have developed the core components with the ERG-ground data. So far the ground-based observational data listed in Table 1 have been incorporated into the integrated data analysis platform.
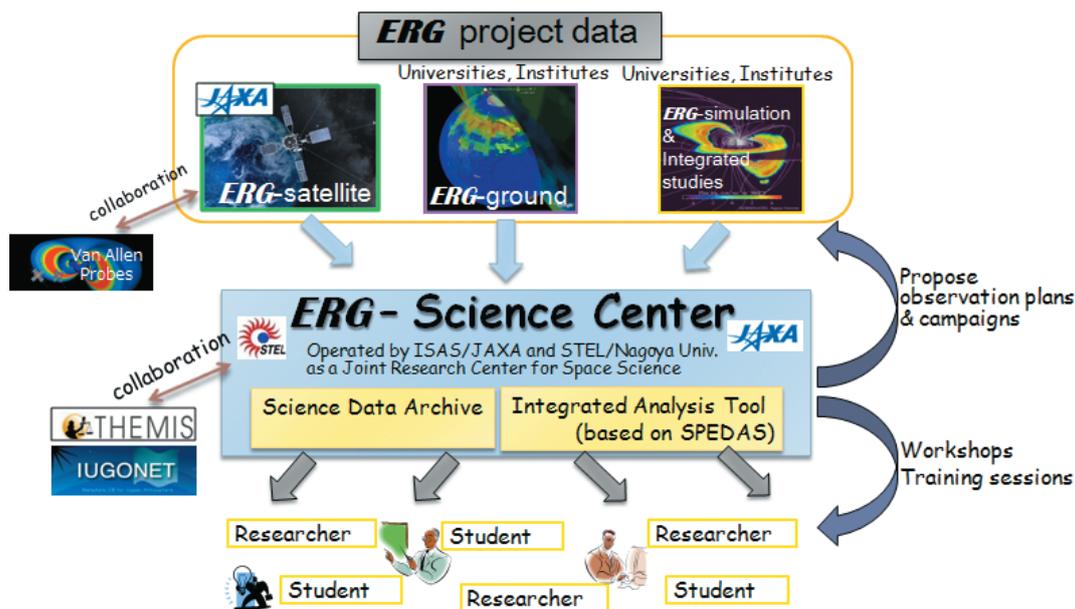
Figure 1. A schematic of ERG science center framework. All science data of the ERG project are archived by and released from ERG-SC. ERG-SC collaborates with other projects in terms of both technical development and scientific coordination. ERG-SC also promotes the science activities of the ERG project through planning of observation schedule and campaigns as well as holding scientific workshops and training sessions for the science community.

Table 1. The list of ERG-ground data archived in the CDF data repository of ERG-SC.

Data name	Observed quantity	Time resolution	Observation site	Data provider
210MM magnetometer chain data	DC geomagnetic field vector	1 sec 1 min 1 h	Japan, Russia, Indonesia, Australia, etc. 33 sites for 1 min/1 hour data, 5 sites for 1 sec data	A. Yoshikawa, Kyushu Univ.
STEL magnetometer data	DC/AC geomagnetic field vector	1 sec 1 min 1 h 64 Hz (AC)	Japan, Russia, Indonesia, Canada, New Zealand. 4 sites for 1 sec data, 6 sites for 1 min data, 5 sites for 64 Hz	K. Shiokawa, STEL, Y. Obana, OECU
MAGDAS magnetometer data	Geomagnetic field vector	1 sec	Asia, Pacific, North America, Australia, Africa, etc. 1 sec data for 16 sites	A. Yoshikawa, Kyushu Univ.
SuperDARN common time fitacf data	Ionospheric plasma convection velocity	~1 min	World-wide, mid. to high latitudes	SuperDARN PI groups*
OMTI all-sky imager data	Aurora / airglow image	~1-2 min	Japan, Canada, Russia, Indonesia, Australia, Antarctica, Hawaii, Norway (15 sites)	K. Shiokawa, STEL

*The list of SuperDARN PIs can be found at the Virginia Tech SuperDARN website at http://vt.superdarn.org/tiki-index.php?page=Radar+Overview.

Figures 2 to 5 provide location maps of the observation sites where the observational data listed in Table 1 have been being recorded. The left and right maps of Figure 2 show the observation sites of 210 Magnetic Meridian (210MM) magnetometer chain data[3] and those of the STEL magnetometer data[4] respectively. 3-dimensional vector geomagnetic field data with 1 minute
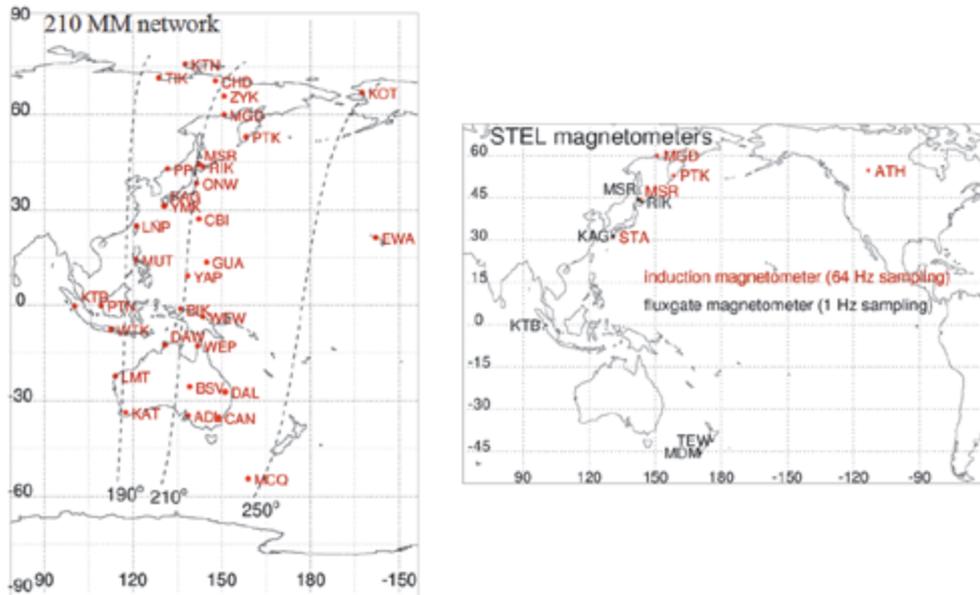
Figure 2. The location maps of magnetometer stations for (left) 210MM geomagnetic field observation and (right) STEL magnetometer observation. Both maps are drawn in the geographical coordinate. Dashed curves superposed on the 210MM map show 190, 210, and 250 geomagnetic meridians as of January 1, 2014 based on the International Geomagnetic Reference Field-11 (IGRF-11) [14].

resolution obtained by fluxgate magnetometers are available for all of the stations and those with 1 sec resolution are available for the selected 5 stations (CAN, KAG, KTB, MSR, and RIK) as ERG-ground data. The STEL magnetometer data are obtained by fluxgate magnetometers with 1 second resolution (in operation at black points on the map) and/or induction magnetometers with a sampling rate of 64 Hz (at red points on the map). Both fluxgate and induction magnetometer data are provided as the ERG-ground data. Detailed information on the 210MM and STEL magnetometer data can also be obtained at the websites at, http://stdb2.stelab. nagoya-u.ac.jp/mm210/, and http://stdb2.stelab. nagoya-u.ac.jp/magne/, respectively.

Figure 3 is a location map for the selected 16 magnetometer stations belonging to the MAGnetic Data Acquisition System (MAGDAS) [5]. The vector geomagnetic field data with 1 second resolution from them are available as ERG-ground data for the science community. Combining data from all of the three magnetometer networks together allows for an extensive spatial coverage of geomagnetic field
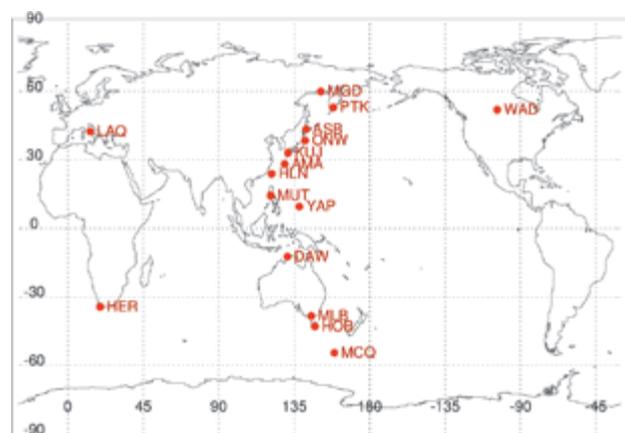

Figure 3. The location map of 16 selected MAGDAS magnetometer stations from which 1-sec data are provided for the ERG-SC CDF data file repository.

observation. These ground magnetic field data give us the crucial information on the global distribution of ionospheric currents and low-frequency plasma wave activities which reflect dynamics of the Earth's magnetosphere and ionosphere. One can obtain the comprehensive, up-to-date information on the MAGDAS project and its data from the MAGDAS website at http://magdas.icswse.kyushu-u.ac.jp/.

Two kinds of 2-dimensional data have joined the ERG-ground data and been made available from ERG-SC. One is all-sky airglow and aurora image data obtained by the Optical Mesosphere Thermosphere Imagers (OMTI) system [6]. The all-sky imagers have been deployed at both domestic
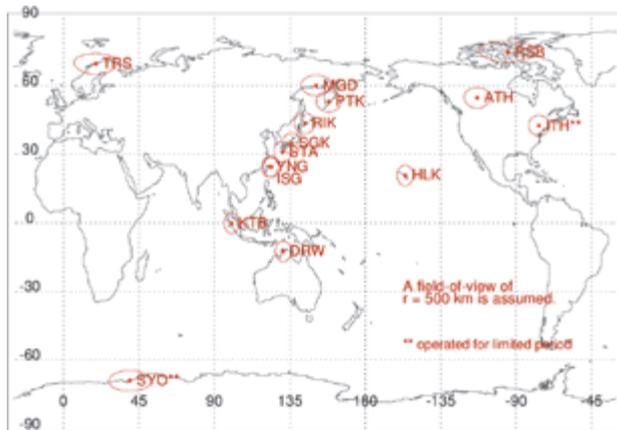

Figure 4. The location map of OMTI all-sky imagers with a rough field of view for each imager. Here the field of view is illustrated simply as a circle with a radius of 500 km whose center corresponds to each imager location. The cameras at Ithaca (ITH) and Syowa (SYO) were operated only for June 28, 2006 to April 16, 2007, and March 1, 2011 to October 1, 2011, respectively. The camera at Yonaguni (YNG) ceased to operate on May 7, 2013 and then it was moved to Ishigaki (ISG) and resumed the operation on April 22, 2014.

and oversea sites as shown in Figure 4, and been producing 2-dimensional optical image data typically with ~1-2 minute resolution. Detailed information on them, such as the observation period for each imager, can also be obtained from OMTI website at http://stdb2.stelab.nagoya-u.ac.jp/omti/.

The other one is the ionospheric radar data obtained by the high-frequency (HF) ionospheric radars joining the Super Dual Auroral Radar Network (SuperDARN) [7]. Their fields of view are illustrated over the globe in Figure 5. Over 30 radars of SuperDARN are currently in operation and provide us with the fitacf data (echo power, Doppler velocity, etc. derived by fitting the raw data) obtained during the common time mode, which are available from the ERG-SC repository.

3. STANDARDIZED DATA FILE FORMAT FOR THE ERG-GROUND DATA

The primary tasks of ERG-SC include development of (1) standardized data file format and (2) program codes to import data to the processing platform of data analysis software. Each standardized data file format was developed by designing metadata elements and data variables so
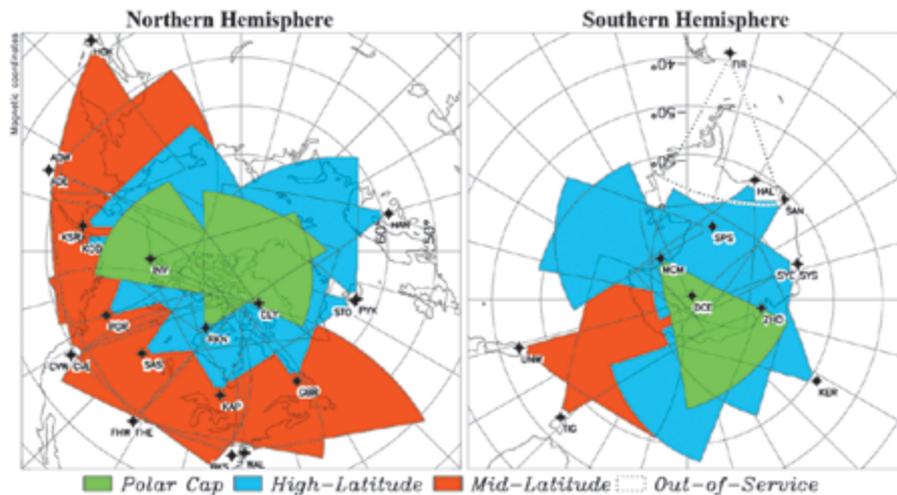

Figure 5. Fields of view scanned by the SuperDARN radars currently in operation in the (left) northern and (right) southern hemispheres. The fields of view are categorized into polar cap radars (green), high-latitude radars (light blue), and mid-latitude radars (red). This radar map is available online from the Virginia Tech SuperDARN website at http://vt.superdarn.org/tiki-index.php?page=Radar+Overview.

that it could be commonly used for all data of each instrument type. The present paper describes details of the resultant products in the following section.

First we describe the concept of the standardized data file formats as well as how to develop them. We started to develop the data file format by organizing a data design consortium for each instrument type of the ERG-ground data. Various topics related to the data archive were discussed in the consortiums, such as suitable metadata/data variable sets, data policy, data file format, practical issues on data processing. First the consortiums for geomagnetic data and ionospheric radar data started off and the similar consortium for all-sky image data followed later on. Those consortiums consisted of the ERG-SC members, the instrument groups represented by principal investigators (PIs), and some researchers who have extensive experience in analyzing the particular types of data. Considering the ideas and suggestions from the consortiums, the ERG-SC members took care of the actual development of the data archive and the associated program coding and data processing.

Discussions in the consortiums were primarily devoted to detailed designs of how data of the particular categories should be archived with the comprehensive but concise metadata and data variables. As a result, we adopted the Common Data Format (CDF) as the data file format (http://cdf.gsfc.nasa.gov/). CDF is a self-describing data format that has been released and maintained by National Aeronautical Space Administration (NASA) / Goddard Space Flight Center (GSFC) and widely been used to archive many satellite data in the solar terrestrial physics community. This file format can contain metadata and various types of data variable in a single file. Data files in CDF are endian-independent and can be read by various programming languages and commercial software on all of major operating systems because the library codes to read CDF files are made available for them by NASA/GSFC. Therefore, providing data files in CDF allows for a ubiquitous access framework with which to analyze the data on the end-user side.

The next task was to design what metadata and data variables are needed to well accommodate the scientific data and metadata of each observation. As for the metadata, there has already existed the standard metadata list released from the CDF website, called the International Solar-Terrestrial Physics/Inter-Agency Consultative Group (ISTP/IACG) CDF guidelines, which are suitable to describe the general information on various data and observation. We adopted them as the base of metadata for our data files and further sought to supplement them with some specific information necessary for actual data analyses. Naturally these data analysis-oriented metadata are different for different types of observational data. Thus we discussed what kind of information, parameter, description would be needed for each data and summarized them as the self-sufficient set of metadata which could be commonly used for each type of data. In contrast to the metadata, the design of data variables is more or less straightforward. Basically we designed data variables for CDF files as they were stored in the original data files. In addition to data quantities readily used for research, we included some engineering data associated with the instrument operation so that data users can use them to check the validity of the observed data to some degree by themselves.

In the present paper, we show the typical set of metadata and data variables for ground magnetometer data in Tables 2a and 2b and for all-sky image data in Tables 3a and 3b, respectively. As for radar data, the metadata/data variable set for the SuperDARN data has already been provided in a separate publication [8].

Tables 2a and 2b show the metadata (also referred to as the global attributes in the CDF framework) and data variable lists, respectively, for 1-minute resolution geomagnetic field data of the 210MM magnetometer chain. The metadata set

consists of the ISTP/IACG-standard metadata and those added by ERG-SC. The former are labeled with asterisks in Table 2a. They provide general information on the data set, such as the observation project name, brief description of the observation, instrument(s), data, and contact information of the PI. The data policy statement is stored in the "Rules_of_use" or "Text" attribute. The latter, namely, ERG-SC original metadata include more detailed information which is useful or further necessary for actual data analyses. For example, "Geographic_latitude", "Geographic_longitude", and "Elevation" provide the exact location of a magnetometer and data users can easily know the location in geomagnetic coordinates by referring to "Geomagnetic_latitude", "Geomagnetic_longitude", and "L_value". We also added "Data_sampling_type" and "Time_calibration_method" so that users can know how the raw data were sampled and averaged and how accurate the given time labels for observed data are. "Known_problem" provides a quick summary about, for instance, unavailable time periods, data caveats, and other problems on data to be shared with users. These metadata sets have been used commonly for CDF data files of the 210MM, STEL magnetometer, and MAGDAS data.

Data variables storing the observed physical quantities obtained by the 210MM magnetometer observation are listed in Table 2b. Since the 210MM magnetometer observation produces 1 minute averaged values and 1 hour averaged values, two sets of data variables with suffixes of "_1min" and "_1h" are prepared to accommodate both data separately. For CDF files, time labels for observation times of data are usually stored in the CDF_EPOCH format, which is an 8-byte floating point value giving the elapsed time in millisecond since 0AD (January 1st, 0000, 00:00:00 UTC). In addition to the CDF_EPOCH time values, we also included those expressed as numbers of year, month, day, hour, minute, second, and millisecond as separate variables ("time_1min" and "time_1h") so that data users can obtain the data time more straightforwardly just in case they do not have any program code to convert CDF_EPOCH values to the ordinary date and time values. As for the observed geomagnetic field, 3-dimensional field variations and their square root of sum of squares are stored in separate data variables (hdz_1min and f_1min for the 1 min data, for example). This is because some geomagnetic field observations have used separate instruments to measure the field variation and absolute intensity and thereby the square root of sum of squares of the field variation may not be necessarily equal to the observed absolute intensity. A set of separate variables can accommodate those independent observed values as they are. If only field variations are measured by a single instrument, which is the case for the 210MM, STEL magnetometer, and MAGDAS data, then their square root of sum of squares should be stored in a data variable for the absolute intensity. Thus our data variable set was designed to support both types of geomagnetic field observation data.

Table 2a. The metadata list for the 210MM geomagnetic field data designed by ERG-SC. The metadata consists of those recommended by the ISTP/IACG CDF guidelines (labeled with asterisks "*", see http://spdf.gsfc.nasa.gov/sp_use_of_cdf.html) and those added by ERG-SC.

List of global attributes (metadata) in 210MM 1-min data CDF file

Attribute Name	Description
Project*	Name of the project holding the ownership of data. "210 MM Magnetic Observation" is set.
Discipline*	Science discipline/subdiscipline in the ISTP/IACG standards. "Space Physics>Magnetospheric Science" is set.

Attribute Name	Description
Source_name*	Name of the mission or investigation under which data are obtained.
Data_type*	Type of the data in CDF file. "1min>1 min Resolution" is set.
Descriptor*	Name of instrument that collects data. "210MM>210 Magnetic Meridian Ground-Based Magnetometer Network" is set.
Data_version*	Version number of data stored in CDF file.
Title*	Title for the data set. "210 MM Ground-Based Magnetometer Network 1 min Resolution Data" is set.
Text*	Description for the data set.
Generated_by*	The generating data center/group. "Solar-Terrestrial Environment Laboratory, Nagoya University" is set.
Generation_date*	Date on which a data file was created.
Mods*	The history of modifications made to the CDF data set.
ADID_ref*	The control authority identifier, currently being left blank.
Logical_file_id*	Name of the CDF file using the ISTP naming convention. "MM210_1MIN_(STN)_ (YYYYMMDD)_V(??)" is set. STN is 3-letter station code, YYYYMMDD is the date in the 8-digit format, and "??" corresponds to the version number of a data file.
Logical_source*	Source_name, data_type, and descriptor information, originally used by NASA/ CDAWeb. "MM210_1MIN_(STN)" is set.
Logical_source_description*	The full words associated with the Logical_source.
PI_name*	Name of the principal investigator (PI) of an instrument.
PI_affiliation*	Affiliations of the PI.
Mission_group*	Data source name originally used in NASA/CDAWeb. "210MM" is set.
Instrument_type*	Type of instrument generating the data set. Following the CDAWeb naming convection, "Ground-Based Magnetometers, Riometers, Sounders" is set.
Rules_of_use*	Rules of the road on using the data is set.
Link_text*	Text describing on-line data available at PI or Co-Investigator (CoI) web sites, to be combined with the following two attributes.
Link_title*	Title of the PI or CoI website.
Http_link*	URL of the PI or CoI websites.
Time_resolution*	Typical time resolution of the data set.
Text_supplement*	Some additional descriptions on the data set, if any.
Station_name	The full spelling of station name.
Station_code	3-letter station code.
Geographic_coordinates	What kind of geographic coordinate system is used for the following latitude and longitude. Usually either of geodetic or geographic.
Geographic_latitude	Geographic or geodetic latitude of the location of station.
Geographic_longitude	Geographic or geodetic longitude of the location of station.
Elevation	Elevation of the location of station.
Geomagnetic_coordinates	What kind of geomagnetic coordinate system and what epoch and time are used for the following geomagnetic coordinate values.
Geomagnetic_latitude	Geomagnetic latitude of the location of station.
Geomagnetic_longitude	Geomagnetic longitude of the location of station.
L_value	(Dipole) L value of the location of station.
Magnetometer_type	Type of magnetometer instrument. Usually "Fluxgate" or "Induction".
Number_of_channels	Number of channels of magnetometer.
Data_coordinates	Coordinate system used for observed magnetic field vector. "other" means that an instrument-specific coordinate system is used.

Attribute Name	Description
Data_type_2	Type of observed vector values, e.g., Absolute field, Variation.
Data_resolution	Typical resolution of data values.
Data_sampling_type	What kind of time bin for averaging or sampling is used to obtain data values.
D_unit	Physical unit of the D-component of the geomagnetic field.
D_conversion_factor	Factor used to convert the unit of D-component from degree to nT.
Time_calibration_method	Time recording method for observation, e.g., "GPS".
Observation_start	The start date of instrument operation.
Observation_end	The termination date of instrument operation. This is left blank if the instrument is currently in operation.
Known_problems	Known problems regarding the data set, such as data glitches and gaps.
K9_limit	Threshold of K-index for K = 9. (for WDC data)

Table 2b. The list of data variables stored in CDF files for 210MM 1-min data.

List of data variables in 210MM 1-min data CDF file

Variable Name	Description
epoch_1min	Time labels for the observed 1-min data in the CDF_EPOCH format.
time_1min	Time labels for the observed data as an array of year, month, day, hour, minute, second, and millisecond.
hdz_1min	1-min values of 3-component vector magnetic field.
f_1min	Absolute value of 1 min magnetic field.
time_cal_1min *	Legacy variable for storing time pulses (every 1 min, 1 hour, 24 hour) from the time signal generator to check the clock inside the data logger attached to the magnetometer.
epoch_1h	Time labels for the observed 1-hour data in the CDF_EPOCH format.
time_1h	Time labels for the observed data as an array of year, month, day, hour, minute, second, and millisecond.
hdz_1h	1-hour values of 3-component vector magnetic field.
f_1h	Absolute value of 1-hour magnetic field.
time_cal_1h *	Legacy variable for storing time pulses (every 1 min, 1 hour, 24 hour) from the time signal generator to check the clock inside the data logger attached to the magnetometer.

* Meaningful values are stored only for magnetometer systems of the old type with an internal clock synchronized with the standard radio transmissions. For those with the GPS-synchronized clock which no longer register the time pulses, this data variable is padded by the fill value or simply not included in CDF data files.

Tables 3a and 3b show the metadata and data variable lists, respectively, designed for a CDF file of the OMTI all-sky image data. Here only the metadata added by ERG-SC are shown in Table 3a; those of ISTP/IACG guidelines are common to all types of data. There are several newly added metadata that do not appear in the metadata list for the geomagnetic field data files.

They include "Camera_ID", "Filter_ID", "Emission", "Wavelength", "Bandwidth", "Exposure_time", "Pixel_size", and so on, which are important information specific to the optical measurement.

Data variables are more or less complicated as compared to those for the geomagnetic field data, as seen in Table 3b. Variable "image_raw" is the one holding the 2-dimensional image data. Variables

"exposure_time" and "sensitivity" have the same names as those in the metadata part. We designed the data variable list to include them also as data variables because they could vary from image to image stored in a single data file and have to be stored in data files as a time series value. We suppose that the metadata elements of the same names store their typical values that data users can quickly refer to.

So far this metadata and data variable set has been used for the OMTI all-sky image data archived by ERG-SC. With CDF files of the same design, currently one of our collaborators, the Inter-university Upper atmosphere Global Observation NETwork (IUGONET) project [9],[10] is working on the archive of the all-sky image data taken in Antarctica, Iceland, and Norway [11] by the National Institute of Polar Research. Although these data are archived by the IUGONET project, they can be accessed by the integrated data analysis software of ERG-SC described later in Section 5.

Table 3a. The list of the metadata added by ERG-SC for CDF files of OMTI all-sky image data.

List of global attributes (metadata) in OMTI all-sky image data CDF file except those suggested by the ISTP/IACG CDF guidelines

Attribute Name	Description
Station_name	The full spelling of station name.
Station_code	3-letter station code.
Geographic_coordinates	What kind of geographic coordinate system is used for the following latitude and longitude. Usually either of geodetic or geographic.
Geographic_latitude	Geographic or geodetic latitude of the location of station.
Geographic_longitude	Geographic or geodetic longitude of the location of station.
Elevation	Elevation of the location of station.
Geomagnetic_coordinates	What kind of geomagnetic coordinate system and what epoch and time are used for the following geomagnetic coordinate values .
Geomagnetic_latitude	Geomagnetic latitude of the location of station.
Geomagnetic_longitude	Geomagnetic longitude of the location of station.
L_value	(Dipole) L value of the location of station.
Camera_ID*	ID of camera used to obtain data.
Filter_ID*	ID of filter in camera used to obtain data.
Emission*	Emission that the camera observed with the filter, such as "OI".
Wavelength*	Wavelength of emission observed by camera.
Bandwidth*	Band width around the central wavelength for the filter.
Exposure_time*	Exposure time for obtaining each image.
Cadence*	Effective time resolution with which individual images are obtained.
Pixels*	Pixel size of each image, such as "512x512 pixels".
Dynamic_range*	Dynamic range of raw count obtained by the camera device, such as "16 bit".
Typical_sensitivity*	Typical value of conversion coefficient from raw count to the physical unit, such as Rayleigh.
Unit*	Unit of raw data, such as "count".
Image_coordinates*	Coordinate system for raw image data.
Data_sampling_type	What kind of time bin for averaging or sampling is used to obtain data values.
Time_calibration_method	Time recording method for observation, e.g., "GPS".
Observation_start	The start date of instrument operation.

Attribute Name	Description
Observation_end	The termination date of instrument operation. This is left blank if the instrument is currently in operation.
Collaborator*	Institute(s) in collaboration for the observation.
Calibration_information*	Detailed information on calibration and conversion to the physical unit.
Known_problems	Known problems regarding the data set, such as data glitches and gaps.

(*) Attributes included only in all-sky image data CDF files, and not included in the geomagnetic field data CDF files.

Table 3b. The list of data variables included in CDF files of OMTI all-sky image data.

List of data variables in OMTI all-sky image data CDF file

Variable Name	Description
epoch_image	Time labels for the observed image data in the CDF_EPOCH format.
time_image	Time labels for the observed data as an array of year, month, day, hour, minute, second, and millisecond.
exposure_time	Exposure time in sec for which each pixel count is integrated.
image_raw	Raw pixel counts for each 2-dimensional image.
sensitivity	Conversion coefficients from raw count to a physical unit, such as Rayleigh.
index_column_image	Original column number for each pixel stored in image_raw.
index_row_image	Original row number for each pixel stored in image_raw.
epoch_cloud	Time labels for the cloud record in the CDF_EPOCH format.
time_cloud	Time labels for the cloud record as an array of year, month, day, hour, minute, second, and millisecond.
cloud	Sky condition data for each hour stored as the following integer. 0: clear sky with stars, 1: a few clouds and many starts, 2: many clouds and a few stars, 3: overcast or rain, 4: no observation, 5: under investigation, 9: no information.

4. CDF FILE REPOSITORY FOR ERG-GROUND DATA

In this section we briefly describe how we have been operating our data repository storing the CDF data files. ERG-SC has developed a data repository with a Linux workstation which is responsible for generating CDF data files of ERG-ground data as well as making them available via the plain HTTP access to data users. The ERG-SC CDF repository server collects the original data files from their respective data servers run by their PI groups on a regular basis. Then the data with the set of metadata are automatically converted to CDF data files and are uploaded to the web space which can be accessed at http://ergsc.stelab.nagoya-u.ac.jp/data/ ergsc/ground/. The CDF file conversion is executed daily or weekly, depending on how often the original data files are retrieved from their data servers. Figure 6 shows a screenshot of the web browser accessing the top directory for the ERG-ground data on the ERG-SC repository. As seen in the Figure, the top directory has three subdirectories, "camera", "geomag", and "radar". CDF data files of the OMTI all-sky image data, all geomagnetic field data, and SuperDARN data are stored in the "camera", "geomag", and "radar" directories, respectively. Since no access restriction is applied, anybody can access all CDF files by just using web browsers or command-line tools such as the GNU Wget.

With permission from the PIs, the ERG-ground data for the entire observation periods, including

Figure 6. An overview of the top directory of the ERG-ground data repository on the ERG-SC website.

before the launch of ERG satellite, have been archived in the same way in the ERG-SC repository. This allows the data for the entire periods to benefit from the useful data analysis platform described in the following section. Further, the data files in CDF with the standardized metadata and data variable sets are all-in-one files carrying all necessary information. Once packed this way with the metadata, they are self-describing by themselves and thus suitable for the long-term archive after the primary research project for them is finished.

In near future, the simulation data generated by various numerical simulations on geospace phenomena will be published from the same website. After the launch of the ERG satellite, its level 2 science data (calibrated, converted to physical unit) and higher level data processed by ERG-SC will be posted as well. Thus this will serve as the comprehensive science data repository for the ERG project from which worldwide researchers can obtain any kind of science data related to the ERG project.

5. INTEGRATED DATA ANALYSIS TOOL

ERG-SC has been developing integrated data analysis software working with the ERG-ground data described in the previous section as well as the other ERG project data. This software has actually been released as a package of plug-in software libraries for the Space Physics Environment

Data Analysis Software (SPEDAS) written in the Interactive Data Language (IDL), a commercial software for data manipulation and visualization. SPEDAS was originally developed by the Time History of Events and Macroscale Interactions during Substorms (THEMIS) project [12] as their own data analysis tools called THEMIS Data Analysis Software (TDAS). Recently, they extended it to a data analysis platform for space physics community-wide use in which other scientific projects could develop program codes to read and visualize their own data with the core routines of SPEDAS and merge them to the SPEDAS source distribution. Detailed descriptions on and the source codes of SPEDAS can be obtained at the SPEDAS website at http://spedas.org/.

Since the SPEDAS core routines are just a set of generic application programming interfaces (APIs) to process data, we have to develop a set of program codes to load and plot each data set of the ERG-ground data by using the core routines, and further customize them to work with our own data repository described in the previous section. Thus the ERG-SC plug-in package consists of data load routines and some original visualization commands based on the SPEDAS core routines. Each data load routine was developed for each data set described in the previous section.

Once data are loaded on the SPEDAS platform, one can visualize them as a time-series plot by using the "tplot" command belonging to the SPEDAS core routines. Because SPEDAS by itself does not provide sufficient routines for visualizing 2-dimensional image-type data, we needed to develop them by ourselves. As a result, most of the ERG-SC original commands are to plot 2-dimensional image type data on a 2-dimensional canvas, which are applied for the all-sky image data and radar data belonging to the ERG-ground data. Figure 7 shows two examples of plots generated by those tools. The left panels are, from top to bottom, time series plots of STEL induction magnetometer
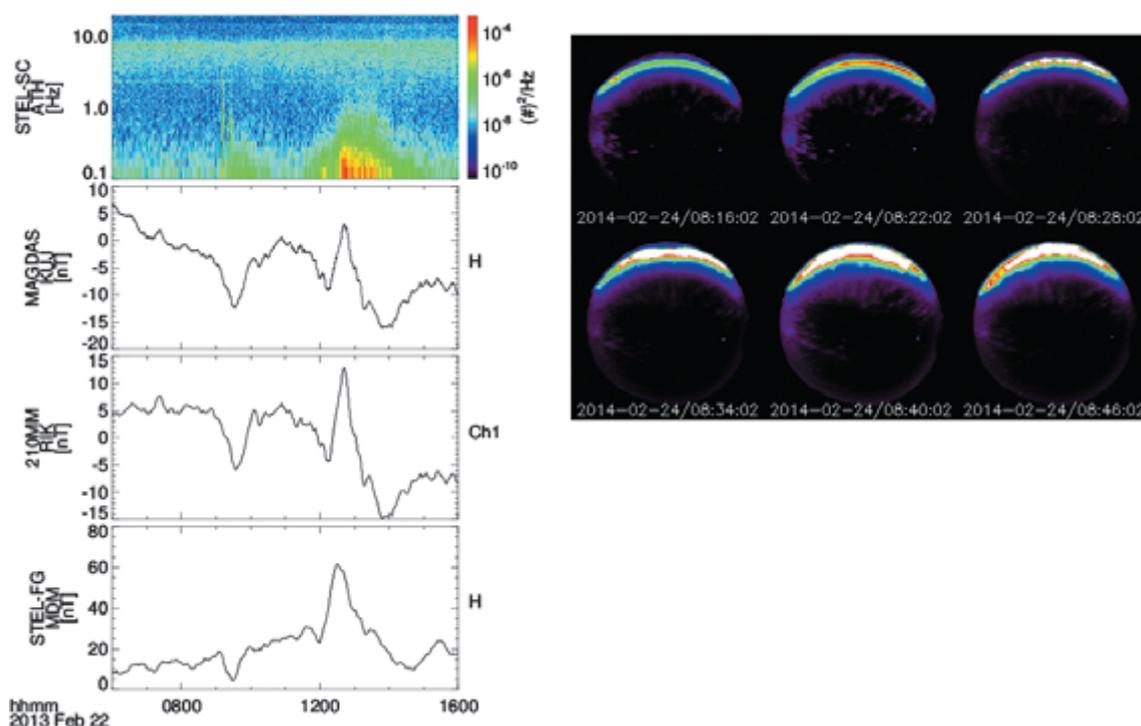
Figure 7. Plot examples for ERG-ground data generated by using the integrated data analysis tools developed by ERG-SC. The left panels are time series plots for the geomagnetic field observations, while the right panels show 2-dimensinal images taken by an OMTI all-sky imager at Athabasca, Canada.

data, MAGDAS data, 210MM data, and STEL fluxgate magnetometer data. The "tplot" command of SPEDAS allows us to easily plot different types of data over the same time axis. The right panels of Figure 7 show 2-dimensional auroral images taken by an all-sky imager of OMTI at Athabasca, Canada (ATH in Figure 4), which is generated by one of the ERG-SC original commands. The other commands can superpose 2-dimensional image type data on the world map to examine their spatial coincidence with other ground data, for instance. The latest version of ERG-SC plug-in libraries can be obtained freely from the plug-in repository on the ERG-SC website at http://ergsc.stelab.nagoya-u.ac.jp/analysis/. They are also merged to SPEDAS basically on a yearly basis.

One of the advantages of using the SPEDAS platform is its generic interface code to load data variables and metadata from CDF files. Once one can prepare his or her own data in CDF, they can readily be read with the useful CDF libraries

implemented in SPEDAS. This is one of the most important benefits when we adopted SPEDAS as the basis of our integrated data analysis tools. Another merit is that its core routines enable us to load different types of scientific data as the common data structure called "tplot variable" and apply various data visualization/manipulation commands for them in the same manner. This means that ERG-ground data can be plotted seamlessly with many other satellite and ground observational data already supported by SPEDAS. SPEDAS-based data load routines also support a function of automatically downloading data files for a specified time interval from remote data servers via the Internet. These routines check the time stamp of data files and retrieves them only if files on the data server are newer or a user does not have them locally. The combination of the common data structure and data downloading function can commonalize programming codes for different types of data and thereby greatly reduce complexity of

coding. Actually one only has to type a series of only a few to several commands to generate each plot in Figure 7. These features benefit not only experts but also newcomers such as graduate students so that they can learn the data analysis tools rather quickly and start their own research without spending too much time in programming just to read data files and plot the data.

6. SUMMARY AND FUTURE PERSPECTIVES

To achieve the scientific goals proposed by the ERG project, ERG-SC has been developing an open data archive of the ERG-ground data in CDF as well as the integrated data analysis platform for the international science community. For the data archive, we designed a set of standard data file structures for the ERG-ground data through discussions in the data design consortiums joined by experts. The integrated data analysis tools have also been developed on the basis of SPEDAS and released for the community to promote integrate studies with the ERG-ground data and other satellite and simulation data.

We will continue to discuss scientific collaboration with other projects and to incorporate new kinds of data that can benefit the ERG science into our open CDF archive - SPEDAS data analysis scheme. Some of the databases of ground magnetometers, radars, and optical imager data developed by the IUGONET project have also been made available in the same manner with SPEDAS [10], contributing to the ERG-ground network. In particular, the collaboration with the Van Allen Probes mission [13] is crucial to advance our understanding of the inner magnetosphere and we need to encourage collaborative studies using ERG-ground data and Van Allen Probes data before the ERG satellite will join the inner magnetosphere observation in 2016. The integrated data analysis platform provided by ERG-SC must be of great help for them, and those efforts definitely lead to good practice for the Japanese space physics community, making us fully prepared to welcome the ERG mission.

REFERENCES

1) Shiokawa, K. *et al.*, ERG – A small-satellite mission to investigate the dynamics of the inner magnetosphere, *Adv. Sp. Res.*, *38*, doi:10.1016/j.asr.2005.05.089, (2006), 1861–1869.

2) Miyoshi, Y. *et al.*, The Energization and Radiation in Geospace (ERG) Project, in*Dynamics of the Earth's Radiation Belts and Inner Magnetosphere, Geophysical Monograph Series, 199*, edited by D. Summers, I. R. Mann, D. N. Baker and M. Shultz, doi:10.1029/2012GM001304, AGU, (2012), 103–116.

3) Yumoto, K. and The 210 MM magnetic observation group, The STEP 210 Magnetic Meridian Network Project, *J. Geomagn. Geoelectr.*, *48*, doi:10.5636/jgg.48.1297, (1996), 1297–1309.

4) Shiokawa, K. *et al.*, The STEL induction magnetometer network for observation of high-frequency geomagnetic pulsations, *Earth, Planets Sp.*, *62*, doi:10.5047/eps.2010.05.003, (2010), 517–524.

5) Yumoto, K. and the MAGDAS group, MAG-

DAS project and its application for space weather, in *Solar influence on the heliosphere and Earth's environment: Recent progress and prospects*, edited by N. Gopalswamy and A. Bhattacharyya, (2006), 399.

6) Shiokawa, K. *et al.*, Development of Optical Mesosphere Thermosphere Imagers (OMTI), *Earth Planets Sp.*, *51*, (1999), 887–896.

7) Greenwald, R. A. *et al.*, DARN/SuperDARN, *Space Sci. Rev.*, *71*, doi:10.1007/BF00751350, (1995), 761–796.

8) Hori, T. *et al.*, An integrated analysis platform merging SuperDARN data within the THEMIS tool developed by ERG-Science Center (ERG-SC), *Adv. Polar Sci.*, *24*, doi:10.3724/SP.J.1085.2013.00069, (2013), 69–77.

9) Hayashi, H. *et al.*, Inter-University upper Atmosphere Global Observation Network (IUGONET), *Data Sci. J.*, *12*, doi:10.2481/dsj.WDS-030, (2013), WDS179–WDS184.

10) Tanaka, Y.-M. *et al.*, Analysis software for upper atmospheric data developed by the IUGONET project and its application to polar science, *Adv. Polar Sci.*, *24*, doi:10.3724/SP.J.1085.2013.00231, (2013), 231–240.

11) Ogawa, Y., T. Motoba, A. Kadokura, Y. Tanaka and K. Hosokawa, Processing and visualization of large amounts of auroral data obtained with All-sky/Narrow field-of-view parallel imagers in Tromso and Longyearbyen (Japanese), *J. Sp. Sci. Info. Jpn.*, *2*, (2013), 51–61.

12) Angelopoulos, V., The THEMIS Mission, *Space Sci. Rev.*, *141*, doi:10.1007/s11214-008-9336-1, (2008), 5–34.

13) Stratton, J. M., R. J. Harvey and G. a. Heyler, Mission Overview for the Radiation Belt Storm Probes Mission, *Space Sci. Rev.*, *179*, doi:10.1007/s11214-012-9933-x, (2013), 29–57.

14) Finlay, C. C. *et al.*, International Geomagnetic Reference Field: the eleventh generation, *Geophys. J. Int.*, *183*, doi:10.1111/j.1365-246X.2010.04804.x, (2010), 1216–1230.