Automatic collection system of meta-data for Solar-Terrestrial Physics (STP) observation data via RSS1.0

Takeshi Murata[1]; Satoshi Ishikura[2]; Eizen Kimura[3]; Takuya Kubo[2]; Kazunori Yamamoto[2]; Iku Shinohara[4]; Mamoru Ishii[5]; Yoshiya Kasahara[6]

[1] CITE, Ehime University; [2] Ehime Univ; [3] CITE, Ehime Univ.; [4] JAXA/ISAS; [5] NICT; [6] Kanazawa Univ.

For the geo-science and the STP (Solar-Terrestrial Physics) studies, various observations have been done by satellites and ground-based observatories up to now. These data are saved and managed at many organizations, but no common procedure and rule to provide and/or share these data files. Researchers have felt difficulty in searching and analyzing such different types of data distributed over the Internet. To support such cross-over analyses of observation data, we have developed the STARS (Solar-Terrestrial data Analysis and Reference System). The STARS consists of client application (STARS-app), the meta-database (STARS-DB), the portal Web service (STARS-WS) and the download agent Web service (STARS DLAgent-WS). The STARS-DB includes directory information, access permission, protocol information to retrieve data files, hierarchy information of mission/team/data and user information. Users of the STARS are able to download observation data files without knowing locations of the files by using the STARS-DB. We have implemented the Portal-WS to retrieve meta-data from the meta-database. One reason we use the Web service is to overcome a variety of firewall restrictions which is getting stricter in recent years. Now it is difficult for the STARS client application to access to the STARS-DB by sending SQL query to obtain meta- data from the STARS-DB. Using the Web service, we succeeded in placing the STARS-DB behind the Portal- WS and prevent from exposing it on the Internet. The STARS accesses to the Portal-WS by sending the SOAP (Simple Object Access Protocol) request over HTTP. Meta-data is received as a SOAP Response. The STARS DLAgent-WS provides clients with data files downloaded from data sites. The data files are provided with a variety of protocols (e.g., FTP, HTTP, FTPS and SFTP). These protocols are individually selected at each site. The clients send a SOAP request with download request messages and receive observation data files as a SOAP Response with DIME-Attachment. By introducing the DLAgent-WS, we overcame the problem that the data management policies of each data site are independent. Another important issue to be overcome is how to collect the meta-data of observation data files. So far, STARS-DB managers have added new records to the meta-database and updated them manually. We have had a lot of troubles to maintain the meta-database because observation data are generated every day and the quantity of data files increases explosively. For that purpose, we have attempted to automate collection of the meta-data. In this research, we adopted the RSS 1.0 (RDF Site Summary) as a format to exchange meta-data in the STP fields. The RSS is an RDF vocabulary that provides a multipurpose extensible meta-data description and is suitable for syndication of meta-data. Most of the data in the present study are described in the CDF (Common Data Format), which is a self- describing data format. We have converted meta-information extracted from the CDF data files into RSS files. The program to generate the RSS files is executed on data site server once a day and the RSS files provide