How CDEX is going to distribute scientific data in its 'Chikyu Laboratory Data Center'

Center for Deep Earth Exploration (CDEX), 19 March 2010

We, CDEX, have distributed scientific data of material curation records, material measurement results, well logging and downhole measurement results in our web site 'Chikyu Laboratory Data Center' [http://sio7.jamstec.go.jp] ('Site sio7' hereafter in this note). For a part of these data, metadata have been also attached so that SEDIS Phase I harvests them to provide data discovery service. This note aims to specify how we are going to (re-) organize theses data in Site sio7 in order to meet our obligation to provide data to the public via the coming SEDIS Phase III.

Snapshots of J-CORES Bulk-Exported files

Currently, contents on our database system J-CORES are exported into a set of a CSV file and various attachment files for each hole periodically and are being distributed in Site sio7. To cope with SEDIS Phase III implementation, we will keep a snapshot of the Bulk-Exported file set for each hole when the moratorium is expired. Each snapshot will have a permanent URI to access the set, and IODP-MI may put DOI for them and may have transparent cache(s) for the snapshots in order to implement SEDIS Phase III. When the data are changed after taking the snapshots, another snapshot will be made with another URI, while the first snapshot will be kept accessible by the original URI.

The snapshots exclude the sampling records; Samples are being taken beyond moratorium period in IODP repositories. We will update this part of Bulk-Exported file set in Site sio7 continuously as ever. The sampling records are in 'Sample column group' in the Bulk-Exported CSV file and files 'curation/sample-codes-upon-requests.csv' and 'curation/lab-sample-codes.csv' (See J-CORES 'Bulk export output specification' [http://sio7.jamstec.go.jp/j-cores/manual/BulkExport/bulk-export.html]).

Contents of J-CORES Bulk-Exported files

We are providing a metadata for a bulk-exported file set for each hole without detailed parameter explanations. Instead, the explanations are given by J-CORES 'Bulk export output specification' so that we expect SEDIS Phase III can serve all the data in the Bulk-Exported file sets (i.e. all the data in the J-CORES database) to the users by its data transformation, parameterized querying, and data aggregation services. J-CORES has some flexibility to store extra numerical data (e.g. 'Extra scalar column group'. See 'Bulk export output specification'). In addition, J-CORES is capable to have various processed depths like core composite depth (CCSF). Depths in such processed scales are calculated and contained in the Bulk-Exported files when such processed depths are stored optionally. We will provide IODP-MI the most recent list of these flexible data with their descriptions.

After SEDIS Phase III implementation, we will inform IODP-MI when we are attempting to modify the Bulk-Export output and when we had any updates for the list of flexible numerical data. If needed in such cases, IODP-MI will be expected to tune SEDIS Phase III in order to keep taking J-CORES data from Site sio7 correctly. By the end of April 2010, we will complete four minor changes in the Bulk-Exported output: (1) implementing IODP Depth Scale Terminology, (2) correcting magnetic susceptibilities in wrong digits, (3) correcting the directory name of age models, and (4) correcting the file name of lithostratigraphic reference. We are thinking about possibilities for a new J-CORES capability to store extra images. For instances, downsized images of cuttings photography, close-up core photography and various borehole images will be loaded into J-CORES by using this function and distributed as a part of Bulk-Exported files sets.

Well logging and downhole measurement results

Site sio7 has well logging and downhole measurements results in various forms. Final forms of these data are in LAS (Log ASCII Standard) format (e.g. [http://sio7.jamstec.go.jp/well_logging/314/C0001D/Standard_log_curves/314-C0001D_gr_gvr.las]) for well logging data and in ASCII files (e.g. [http://sio7.jamstec.go.jp/downhole_measurement/316/temperature/316-C0008C.DH_temperature.txt]) for downhole measurement data. In addition to material curation records and material measurement results, J-CORES will store these final forms of well logging and downhole measurement results as a part of extra numerical data and distributed as Bulk-Exported CSV files. The list of the flexible numerical data in Bulk-Exported file will include these well logging and downhole measurement data (See Section Contents of J-CORES Bulk-Exported files). As mentioned above, when J-CORES has extra image storing function, downsized borehole images will be loaded into J-CORES additionally and distributed as a part of Bulk-Exported files.

The LAS and ASCII files will be removed from Site sio7 once after J-CORES takes care of them. The other data files of well logging and downhole measurement (like DLIS, PDS, and PDF for well logging data; full set of raw measurement results for downhole measurement) will be on Site sio7 as ever. These data will be treated as a group of data files not on J-CORES (See Section Data files not on J-CORES).

According to 'Request for Proposals (RFP) Scientific Earth Drilling Information Service (SEDIS) Phase III' [http://campanian.iodp-mi-sapporo.org/SEDIS/RFP_SEDIS_III_Final.pdf], 'CDEX borehole logging data set metadata is managed by the LDEO Borehole Research Group'. We however decided to provide the matadata for our well logging data by ourselves as mentioned above, rather than LDEO.

Data files not on J-CORES

As mentioned already, J-CORES will have most of the scientific data including material curation records, material measurement results, well logging results and downhole measurement results. Beside J-CORES, we have some other kinds of data files in Site sio7: well logging and downhole measurement results in

various file formats (See Section Well logging and downhole measurement results), axial image sets of X-ray CT scanning, full-size line-scanned images for split core sections, and photography for cuttings samples so far. In the future, we may have other different types of data files additionally.

We will have a metadata to point the index page of such data files for each hole (like [http://sio7.jamstec.go.jp/well_logging/314/C0001D/], [http://sio7.jamstec.go.jp/static.j-cores.data/316/316-C0004C/html/316-C0004C-XRAY.html], [http://sio7.jamstec.go.jp/static.j-cores.data/316/316-C0004C/html/316-C0004C-MSCLI.html] and [http://sio7.jamstec.go.jp/static.j-cores.data/319/319-C0009A/html/319-C0009A-MISC.html]. Access limited for the last one to protect moratorium data. URIs may change and be indicated in the metadata) in order to enable discovering these data by SEDIS Phase I. SEDIS Phase III is not expected to read inside of these data files.

Timeline

CDEX will implement this plan before IODP expeditions in 2010.