Accredited Standards Committee
NCITS, National Committee for Information Technology Standards

Doc: FCW/98w120r5
Date: March 17, 1998
Project:
Reply to: John Scheible

To: Database and mailing coordinators
From: John Scheible

Subject: Common Database Record Interchange Structure

Purpose

Develop a minimal common Interchange format to be exported by the various group databases independent of the database program used. This will allow common tools to be used to generate mailings and HTML code from multiple sources. This revision is the results of a meeting 1998/03/16.

Format (proposed)

The file format is in common separated variable, with text that could possibly contain commas enclosed in quotes. Use of upper and lowercase needs to be consistent to avoid sorting problems. Recommend all lower case except in "title" and "author" fields. Recommend that commas be avoided as some programs (such as word) ignore quotes and will break up the field. Null records are " " or "" or ,,. Some database programs have trouble with "" records (" space " is better). Required indicates a non-null field. The format of all records are shown in Table 1.

Table 1 - Record format

Field	Label	Description	Field format
1	DATABASEID	Identifier of the database	ASCII field
2	COMMITTEE	Committee record is associated with	ASCII field
3	ACTIVITY	General category identifier	ASCII field
4	RECNUMBER	The record number	ASCII field
5	ACRONYM	Official project assigned to (if any)	ASCII field
6	ASSOCIATION	groups documents or drafts	up to 8 characters with sub-directory restrictions
7	MAILINGID	mailing identifier	yyyy_x format (1<=x<=6)
8	SUBMAILING	date of the submailing	yyyy/mm format
9	SHORTNAME	8.3 format filename	8.3 with filename restrictions
10	LONGNAME	long file name, blank if short	up to 64 characters with Joliet filename restrictions
11	TITLE	document subject or draft title	ASCII characters
12	AUTHOR	proposer or technical editor	ASCII characters
13	REVISION	document or draft revision	ASCII characters
14	DATE	date on document or draft	yyyy/mm/dd format
15	PAGES	The number of pages of the PDF file	ACSII numeric value
16	TYPE	type of record	lower case ASCII characters
17	SEQUENCE	ordering within a type	lower case ASCII characters
18	STATUS	status of the record	(see below)
19	STAGE	status of the draft	(see below)
20	ONSITE	indicates file is on the ftp site	"Y" is archived on FTP sire, all other values indicate the file is not available.
21	PROJECTNUM	The draft project number	ACSII numeric value
22	DESIGNATION	The draft BSR number	ACSII numeric value
23	FTPSITE	The FTP URL of the file	ASCII characters in a valid URL format
24-n		database specific	freeform records.

DATABASEID   An identifier consistent between all records of a database (i.e., T10, T11.2, T11.3, etc.).

COMMITTEE   Committee associated with the record.

ACTIVITY   The general activity associated with the record.

RECNUMBER   The number of the record (i.e., T11=98-xxxvn, T10=98-xxxvn or xxxxrnnn)

ACRONYM   The project identifier associated with drafts (i.e., SPC, AL-3, TAPE) or "none". It is up to the database administrator to decide whether documents fall under the a null acronym, a "other" acronym, or the project that they may apply to. *I recommended 4 characters to allow filenames of pppprnnn.*

ASSOCIATION   up to 8 lower case ASCII characters used to group records as defined by the database owner. The 8 characters may be used as a sub-directory to hold appropriate records (i.e., ISO9660 compliant 8.0 format filename). Required for drafts, null for non-drafts.

MAILINGID   An identifier of the mailing to be generated in the format of yyyy_x where x is a sequential number starting with 1 for each plenary mailing. Required.

SUBMAILING   The date of the submailing containing this record in the form yyyy/mm. "none" indicates the record is not to be included in any mailing. A TYPE of MAILING also excludes a record from a mailing. Optional if no interim mailings are done.

SHORTNAME   ISO9660 compliant 8.3 filename with the addition of dash "-". Required.

LONGNAME   If the filename is an 8.3 filename, then this field shall be null. Otherwise, the field includes the Joliet long filename up to 64 characters, and the SHORTNAME field includes a unique tag to identify the file.

TITLE   The title of the draft or document. Recommend length of up to 64 characters. Required.

AUTHOR   The author of the draft or document. Recommended length of up to 25 characters. Required.

REVISION   Revision level. Required.

DATE   Date on the front of the document/draft. To allow better sorting and avoid year 2000 issue, the format is yyyy/mm/dd. The use of / is recommended, but - is acceptable if used consistently. The DATE field is optional (i.e., "" allowed) for non-drafts, but required for drafts.

PAGES   The number of pages printed by the associated PDF file. Optional.

TYPE   lower case ASCII characters used to type records and control placement within a mailing. Required. Types include:
                 | mailing | Excluded from the mailing |
                 | meeting | Meeting announcements, maps, agenda's |
                 | administrative | Administrative files (i.e., jeopardy letters) |
                 | minute | Meeting minutes |
                 | document | Proposals |
                 | draft | Drafts |

SEQUENCE   The ordering within a type field for the mailing. Required.

STATUS   Indicates the status of the version of a draft or document. Currently only used by T10, can be null for other databases. Optional. Types include:
                 | null or "1" | older version |
                 | "2" | latest version |
                 | "9" | obsolete |

STAGE   Used to indicate the status of the particular revision of the draft. Proposed drafts should be considered documents, not drafts. Projects undergoing approval may be considered drafts. Ignored by non-drafts. Acceptable values are:

development draft revision is in active development

letter ballot   draft revision went to letter ballot

comment resolutiondraft revision is undergoing comment resolution

public review draft revision went to public review

in publication draft being published, but not available yet

published   draft revision has been published and is available

obsolete   draft revision is obsolete and not undergoing development

ONSITE   A "Y" indicates the file exists in PDF form on the appropriate FTP site. Any other value will cause the record to be excluded from the mailing.

PROJECTNUM The project number of the draft. Required for drafts, optional for non-drafts.

DESIGNATION The BSR or other appropriate number of the draft. Optional for drafts, not applicable to non-drafts.

FRPURL   The FTP site URL containing the file associated with the record,

FCW / T11.3 plans

I (John Scheible) use 8.3 filenames and carefully control the filenames (drafts are ppprxxx, documents and other files are 98wnnnrx). My database has a minimum of fields which are restricted to set values where applicable. I currently use Symantec Q and A, but am considering changing database programs (should not impact anything). In order to isolate and minimize my database, I plan to export my database, and post-process it to meet the potentially changing requirements of CD-Access and NCITS(see Tools, below). See 98w113rx (currently 98w113r1) for more details in filename conventions. I plan to convert to Roger Cummings system when it has been debugged and proven.

I plan to construct the required fields as follows:

- The DATABASEID and COMMITTEE fields will be a constant "T11.3" or "FCW" depending on the MAILDATE.

- The ACTIVITY and ACRONYM fields will be the associated approved project (4 characters). In the case of unapproved projects, the ACTIVITY and ACRONYM will be "other" until approved.

- The ASSOCIATION field will be agreed upon between myself and CD-access and other interested parties.

- MAILINGID will be 1998_1 starting with the May 1998 T11.3 plenary mailing.

- SUBMAILING will be 1998/mm where mm is the month.

- The RECNUMBER and SHORTNAME field will be 98wnnnrx for documents (nnn=number, x=revision), and nnnnrxxx for drafts (nnnn is the project and xxx is the revision).

- The LONGNAME field will always be " " as I will always use short names (i.e., 8.3 names).

- The REVISION field will be a single digit for documents, with xxx for drafts.

- The STATUS field will be " ".

I use the SUBMAILING, and TYPE fields to include and exclude database records for the mailing (i.e., include records matching SUBMAILING, exclude records with TYPE field of mailing). The month part of SUBMAILING field is also used to create filenames and text within the HTML and text files. The SHORTNAME, AUTHOR, TYPE, SUBMAILING, and PAGES fields are also listed within the HTML. Currently, I am trying to keep a single database for documents and drafts.

I export my database CSV files and run a program to do the following (see 98whtmxx.zip where xx is the mailing month):

- Generate a warning report of any records that are not available in PDF form.

- Generate a text summary of the mailing contents for sending to the reflector.

- Generate a csv file of the mailing contents to generate the cover letter.

- Generate the local HTML code for the main, entire database, and specific mailing.

- Generate the web HTML code for the main, entire database, and specific mailing (different pointers/text).

- Generate the CD-Access common interchange format.

Tools

Contact John Lohmeyer of Symbios (john.lohmeyer@symbios.com) for tools to perform extensive processing of existing CSV (comma separated variables) database files into acceptable common database record interchange structures.

How I would generate HTML

Here is a little discussion on how I would generate the HTML I need for my submailings.

All HTML could optionally sort by meeting, then administrative, then minutes, then documents, then drafts.

First I export the entire database CSV sorted by SHORTNAME, then DATE in descending order. Then I run a check for ONSITE<>"Y" and generate a warning file. Then I generate three sets of linked HTML for both web and local access. I also generate a CSV file for a mailing cover page, and a text file for the mailing announcement. I would then post process to come up with the Common Interchange format. All this is automated. The HTML code type would be:

DATABASE HTML: The Database HTML lists SHORTNAME, TITLE, AUHOR, SUBMAILING, PROJECT and would include based on ONSITE="Y".

MAILING HTML: The individual Mailing HTML lists SHORTNAME, TITLE, AUHOR, PROJECT, PAGES and would sort on:

- ONSITE="Y"
- TYPE <> "mailing" (to exclude the actual records of the mailing)
- SUBMAILING = appropriate date

Web Page HTML: The Web page HTML needs to list several meeting/map/agenda matters, then the latest drafts, then the meetings, administrative, meeting minutes, documents, then non-latest drafts. These could be done with multiple passes and checks of the TYPE field.

Sincerely,

John Scheible
Voice: (512) 823-8208
FAX: (512) 838-8378
Email: jpscheible@aol.com