

# CREATING DISTRIBUTED DATASETS FOR ECAI USING THE TIMEMap DATA PREPARATION TOOLKIT

---

**Ian Johnson**

**Archaeological Computing Laboratory**

**University of Sydney NSW 2006**

**Australia**

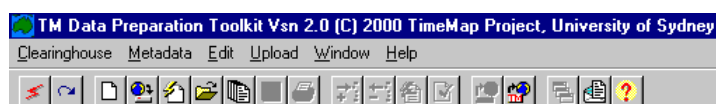
*johnson@acl.archaeology.usyd.edu.au*

*http://www.timemap.net*

## The ECAI metadata clearinghouse

The ECAI metadata clearinghouse serves as a central index to all ECAI web-accessible datasets, and is based on the Dublin Core metadata standard with ECAI extensions. The clearinghouse was conceived in 1998 and fully implemented by a team at the Archaeological Computing Laboratory, University of Sydney, towards the middle of 1999.

The initial implementation of the ECAI clearinghouse included web-based registration of datasets and entry of metadata with standard web browsers, implemented through CGI and javascript. While this system is still operative and will be the tool of choice for many first-time users, we have since developed a Windows-based program TMT (TimeMap Toolkit) with greatly extended functionality:



- Wizard-based creation of metadata files from GIS files or database tables, including creation of TimeMap metadata entries allowing access to data by the TimeMap viewer program (TMView);
- Editing of metadata files on a local drive or network;
- Registration, upload, update and download of metadata files to/from the ECAI Clearinghouse;
- Upload of metadata files to an SQL database server;
- Upload of GIS data files and database tables to a non-spatial SQL database server;
- Inclusion of 'snapshot' images of the data with the metadata for easier assessment of datasets when browsing the clearinghouse.

The increased functionality of the TimeMap toolkit program (TMT) is complemented by new clearinghouse browser functionality in the TimeMap viewer program (TMView). In this paper we will provide an overview of the functions of the toolkit program, and of the new browser functionality in the viewer.

## The TimeMap metadata wizard

The metadata wizard function in TMT takes, as input, a dataset created with GIS software such as MapInfo or ArcView or – for point data only – a database or spreadsheet such as Access, Dbase, Paradox or Excel. The dataset is assumed to consist of a table which contains:

either longitude (X) and latitude (Y) values for point objects (such as site locations);

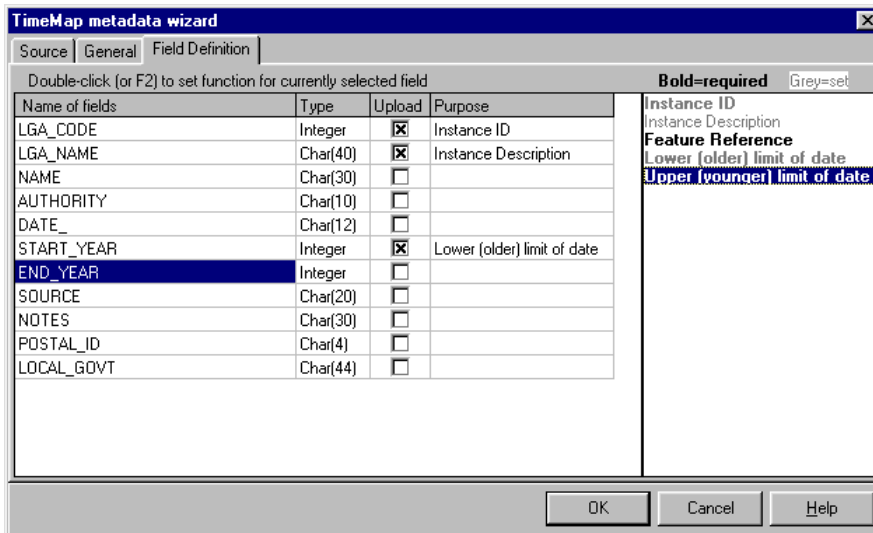
or points, lines or polygons stored as GIS objects in MapInfo TAB or ArcView shapefile format.

The metadata wizard reads the GIS or database table and creates a TimeMap metadata file with the essential data required to allow the TimeMap viewer program to connect to the dataset.

The metadata wizard allows the user to make a number of choices, generally from pulldown lists, to define the way the metadata is set up. A full discussion of the meaning of the terms is provided in the TimeMap Data Preparation Manual on <http://www.timemap.net>:

The screenshot shows the 'Field Definition' tab of the TimeMap metadata wizard. The interface is as follows:

- Table name:** china\_provinces\_4m
- Dataset name:** china provinces 4m
- GIS type:** MapInfo
- Type of map object:** Polygon
- Time Varying:** NoDateFields
- Geographic Coverage:** A box containing the coordinates 53.478680, 135.092682, 73.513115, and 18.157558.
- Link mask:** http://www.acasian.gu.edu.au/county.cgi?id= [ GB\_CODE\_END\_199 ]
- Thumbnail:** Q:\timemap\software\distribution\Resources\SampleData\china\_provinces\_4m\_th.gif
- Sample:** Q:\timemap\software\distribution\Resources\SampleData\china\_provinces\_4m\_sm.gif
- Documentation:** Q:\timemap\software\distribution\Resources\SampleData\china\_provinces\_4m.htm



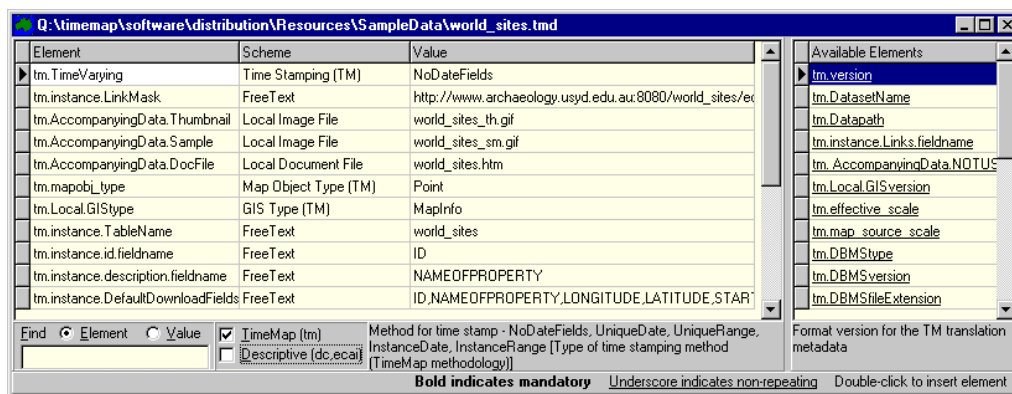
**Fig. 1: Metadata wizard dialogues used to create a metadata file from a GIS dataset**

The wizard carries out a number of consistency checks, such as assuring that lower and upper date limit fields are set if time recording method has been defined as a range of dates.

## Editing the metadata

Once a metadata file has been created with the metadata wizard, TMT provides an editing function. Several files can be open in different editing windows, and the editing window provides pulldown lists of fields, file browsing and verification functions to ensure entry of valid data.

Fig. 2 shows the TimeMap metadata generated by the metadata wizard. TimeMap metadata allows the TimeMap viewer program, TMView, to access the internal structure of the dataset:



**Fig. 2 Typical TimeMap metadata displayed in the TMT editing window**

Fig. 3 shows the descriptive (ECAI/Dublin Core) metadata. Guidelines to filling in this descriptive metadata are provided in the ECAI Metadata Standard (available from <http://www.ecai.org/>).

Element	Scheme	Value
<b>dc.title</b>	FreeText	world sites
dc.creator.CorporateName	FreeText	Archaeological Computing Laboratory
dc.creator.CorporateName.address	FreeText	University of Sydney, NSW 2006, Australia
dc.creator.CorporateName.address	Email address	acl@acl.archaeology.usyd.edu.au
dc.creator.CorporateName.address	URL identifier	http://www.archaeology.usyd.edu.au
<b>dc.subject</b>	FreeText	Cultural heritage sites   world heritage   unesco
<b>dc.description</b>	FreeText	A database of cultural heritage sites assembled by volunteers at the Archae
<b>dc.description</b>	FreeText	Initial data source was UNESCO web site, supplemented by individual work
<b>dc.description</b>	FreeText	Database is now available online through ECAI and can be updated through
<b>dc.publisher</b>	FreeText	Archaeological Computing Laboratory, University of Sydney
dc.publisher.address	FreeText	University of Sydney, NSW 2006, Australia
dc.publisher.address	FreeText	acl@acl.archaeology.usyd.edu.au
dc.publisher.address	FreeText	http://www.archaeology.usyd.edu.au
dc.contributor.PersonalName	FreeText	David Hobson
dc.contributor.PersonalName.address	Email address	hobson@acl.archaeology.usyd.edu.au
<b>dc.date</b>	Dates (ISO 8601)	21-Mar-00
<b>dc.type</b>	Resource Type (DC)	Dataset
<b>dc.format</b>	Resource Format (DC)	TimeMap dataset
dc.identifier	FreeText	http://www.archaeology.usyd.edu.au:8080/world_sites
<b>dc.language</b>	ISO639-2B Short list	eng
<b>dc.coverage.x.min</b>	WGS84 Lat/long	180
<b>dc.coverage.x.max</b>	WGS84 Lat/long	-180
<b>dc.coverage.y.min</b>	WGS84 Lat/long	-90
<b>dc.coverage.y.max</b>	WGS84 Lat/long	90
<b>dc.coverage.t.early</b>	BP years	10000
<b>dc.coverage.t.late</b>	CE (AD) years	2000
dc.coverage.PlaceName	TGN (Getty thesurus)	World
dc.coverage.PeriodName	FreeText	Prehistoric   Historic
dc.coverage.notes	FreeText	Coordinates mostly obtained from Getty Thesaurus of Geographic Names, p
dc.rights	FreeText	This data may be used for non-profit personal and scholarly use

Available Elements
dc.title.alternative
dc.creator.PersonalName
dc.creator.PersonalName.Ac
dc.creator.CorporateName
dc.creator.CorporateName.a
<b>dc.subject</b>
<b>dc.description</b>
<b>dc.publisher</b>
<b>dc.publisher.address</b>
dc.contributor.PersonalName
dc.contributor.PersonalName
dc.contributor.CorporateName
dc.contributor.CorporateName
<b>dc.date</b>
dc.date.accepted
dc.date.acquired
dc.date.available
dc.date.created
dc.date.datagathered
dc.date.issued
dc.date.lastmodified
dc.date.valid
<b>dc.type</b>
dc.type.specific
<b>dc.format</b>
dc.format.specific
dc.format.size
dc.identifier
dc.source
<b>dc.language</b>

Find  Element  Value

TimeMap (tm) Dublin Core Title - the name given to the resource, usually by Creator or Publisher  
 Descriptive (dc.ecai) [Generic unconstrained free text]

Address of publisher, including email

**Bold indicates mandatory**    Underscore indicates non-repeating    Double-click to insert element

**Fig. 3 Typical descriptive (ECAI/Dublin Core) metadata elements**

Metadata elements can be added by double-clicking the element in the righthand panel. Bolded elements are required elements; underlined elements are non-repeating and can only be inserted once into the data.

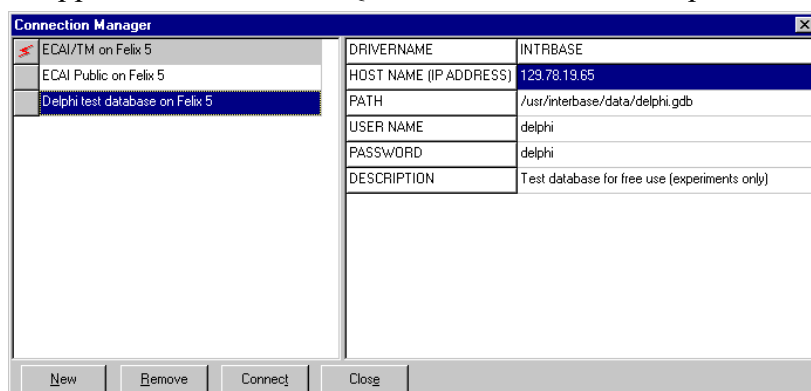
## Registering datasets with the ECAI clearinghouse

TMT allows the user to register a newly created metadata file with the ECAI metadata clearinghouse, or to update the metadata in the clearinghouse if the file has been edited since it was registered or last updated. If the dataset has already been registered, TMT will read the title of the dataset and its location from the clearinghouse, so all or most of the fields on the registration dialogue will be filled in (Fig. 4). If it is a new registration, these fields will be blank:



**Fig. 4: Metadata registration/update dialogue with existing information drawn from ECAI clearinghouse**

TMT uses a special component called the Connection Manager (Fig. 5), which is accessed through the New Connect button. The Connection Manager is a general-purpose tool for defining connections to SQL database servers, and is also used by TMView. Currently (March 2000) it supports Interbase, Sybase, MySQL and any installed ODBC drivers on your computer. Support for Microsoft SQL server is under development.



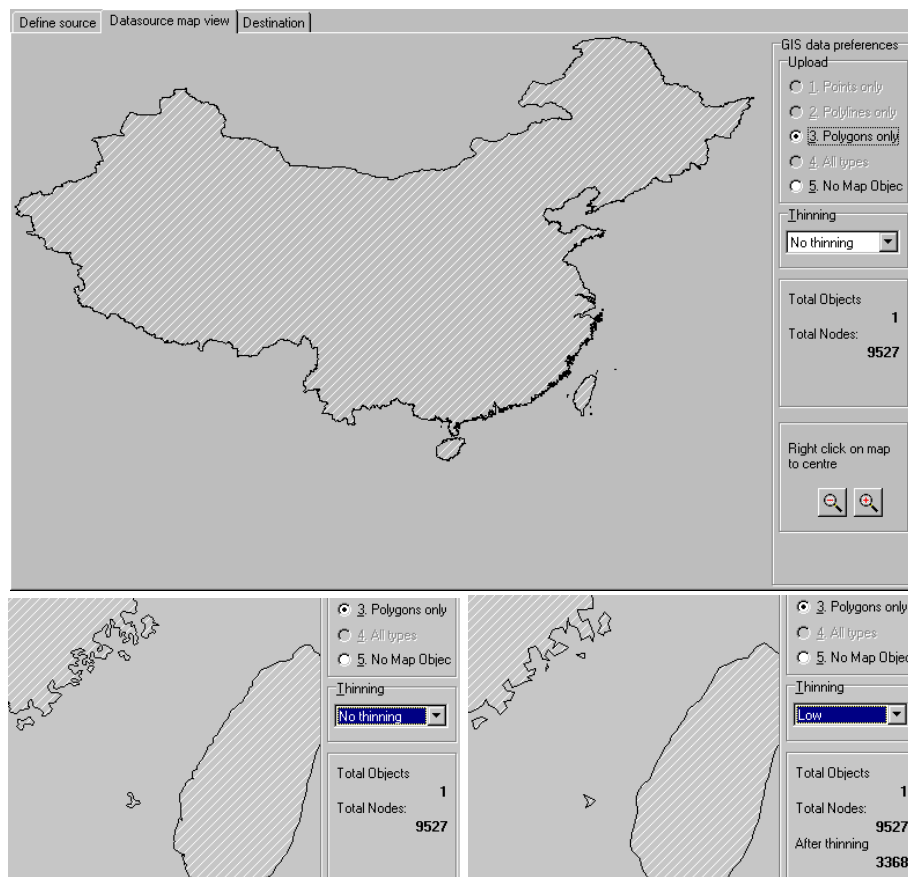
**Fig. 5: The Connection Manager dialogue allows connections to a variety of different SQL database servers**

In addition to registering the dataset with the ECAI clearinghouse, TMT uploads a copy of the metadata table to the SQL server database which will hold the dataset. The copy on the server is read by TMView when it connects to the dataset.

## Uploading data with the DataPump

TMT includes a “DataPump” function which uploads the actual dataset itself from a local GIS or database file to an SQL server database (accessible across the Internet). The DataPump converts local GIS data in MapInfo TAB or ESRI Shapefile format into a form which can be stored in a conventional SQL server database such as MySQL, MSQL or Interbase. The server database does not require spatial data capabilities (which are found only in high-end products), because map features (polylines and polygons) are stored as highly compressed blob fields and indexed through their bounding boxes (the rectangle which encloses them).

The size of the map features uploaded is further reduced by thinning out unnecessary nodes. The effect of different levels of thinning can be visually assessed on a preview map (Fig. 6) prior to upload. The combination of compression methods and thinning can reduce the size of spatial objects stored in the database by as much as two orders of magnitude.

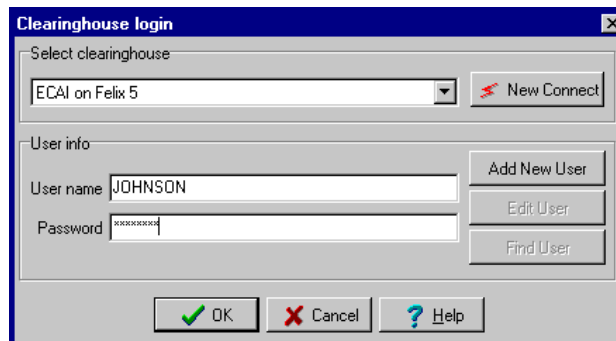


**Fig. 6: Map upload preview illustrating data reduction by thinning of outlines**



## Metadata download and HTML conversions

TMT can connect to the ECAI clearinghouse (Fig. 7), display a list of the datasets registered by the logged-in user (Fig. 8) and download the metadata for any of these datasets to a local file. The local file can be edited and uploaded back to the clearinghouse in the same way as a file created from scratch or with the metadata wizard. This download/upload capability provides a useful backup for local metadata files which might become mislaid or corrupted.



**Fig. 7: Clearinghouse login dialogue (also allows addition of new users)**

Id	Title	Created	Updated	Type
1	Coastline of China (low resolution)	27-Feb-99 20:24	16-Mar-00 01:04	TimeMap Dataset
3	University of Sydney School of Archaeology, Home Page	28-Feb-99 00:55	28-Feb-99 00:55	Web Site (Static)
2	Electronic Cultural Atlas Initiative (ECAI) Home Page	28-Feb-99 00:43	19-Jun-99 09:24	Web Site (Static)
8	Historic photos (20th C) of Sydney including many oblique air photos from 1930s	29-Mar-99 21:28	05-Jan-00 18:40	TimeMap Dataset
9	Chinese Biographical Database (M. Levine)	29-Mar-99 21:45	16-Mar-00 13:40	TimeMap Dataset
62	Railways of the Sydney Region	30-Dec-99 15:20	16-Mar-00 15:11	TimeMap Dataset
16	China counties, modern (ACASIAN)	28-May-99 23:17	05-Jan-00 19:03	TimeMap Dataset
17	Outline of modern China 1:4M	01-Jun-99 16:52	16-Mar-00 00:53	TimeMap Dataset
63	Satellite Image of Central Sydney	30-Dec-99 15:24	05-Jan-00 18:59	TimeMap Dataset
29	Sketch of Sydney Cove, Port Jackson, July 1788		05-Jan-00 18:44	TimeMap Dataset
65	Sydney Cultural Institutions	30-Dec-99 15:35	16-Mar-00 16:31	TimeMap Dataset
21	Water courses of modern China	18-Jun-99 02:16	16-Mar-00 12:22	TimeMap Dataset
26	NADB - National Archaeological Database Reports	04-Jul-99 08:00	04-Jul-99 08:00	Web Site (Database)
30	Grimes's Plan of Sydney, 1800		05-Jan-00 18:44	TimeMap Dataset
31	Meehan's plan of the Town of Sydney, 31st October 1807		05-Jan-00 18:44	TimeMap Dataset
32			10-Aug-99 20:44	
33	Plan of Sydney and Portion of the Harbour, 1866		05-Jan-00 18:45	TimeMap Dataset
81	China Provinces, modern (ACASIAN)	03-Jan-00 16:36	05-Jan-00 19:12	TimeMap Dataset
70	Bibliography of Yunnan Province, 17th-20th Centuries (T. Hahn)	30-Dec-99 15:58	16-Mar-00 13:32	TimeMap Dataset
36	Reconstructed 1788 DEM for central Sydney		05-Jan-00 18:51	TimeMap Dataset
45	Household data for The Rocks 1865		05-Jan-00 19:01	TimeMap Dataset
46	WorldSat composite satellite image of Asia (4km cells)		05-Jan-00 19:07	TimeMap Dataset

**Fig. 8: List of datasets registered with the ECAI clearinghouse by the currently logged-in user**

## Exporting metadata to HTML

TMT allows export of the metadata from a TimeMap metadata table to text-based HTML 4.0 format, as illustrated in Fig. 9. This feature is primarily intended to allow metadata to be incorporated into web pages, but a corresponding import function means it can also be used to copy metadata between datasets or to import metadata from existing web pages.

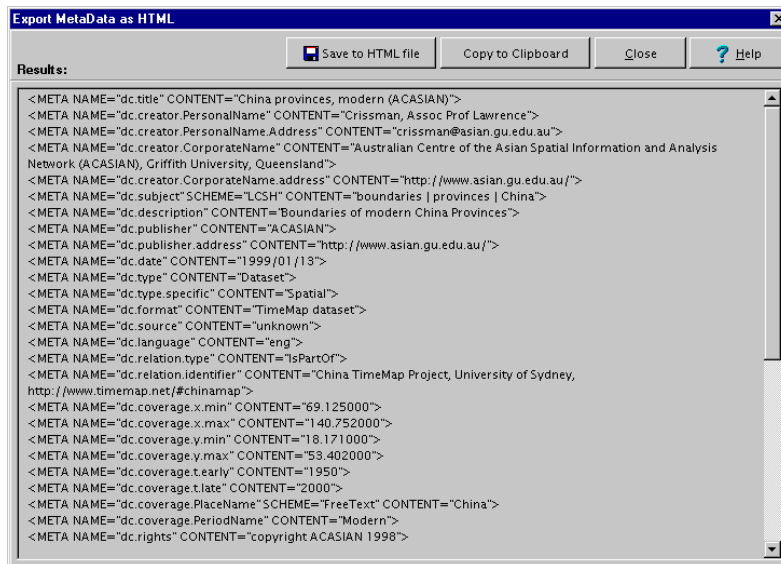


Fig. 9: Metadata export dialogue (Table → HTML 4.0)

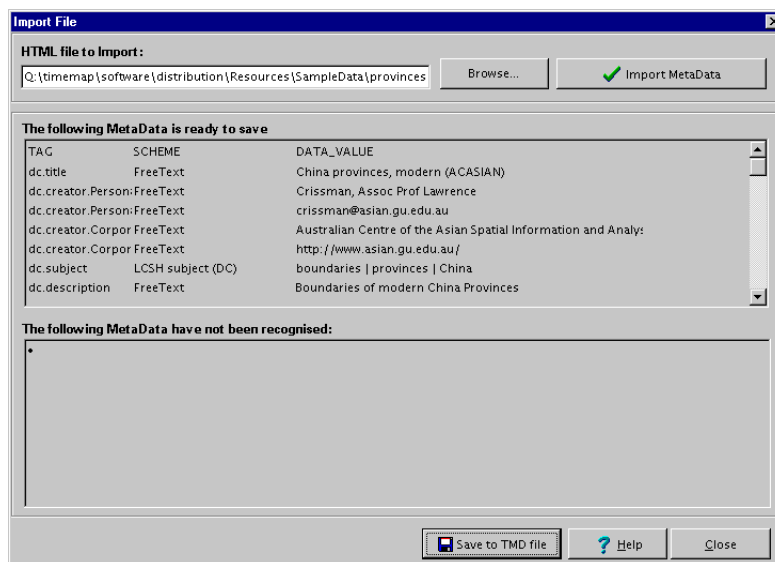
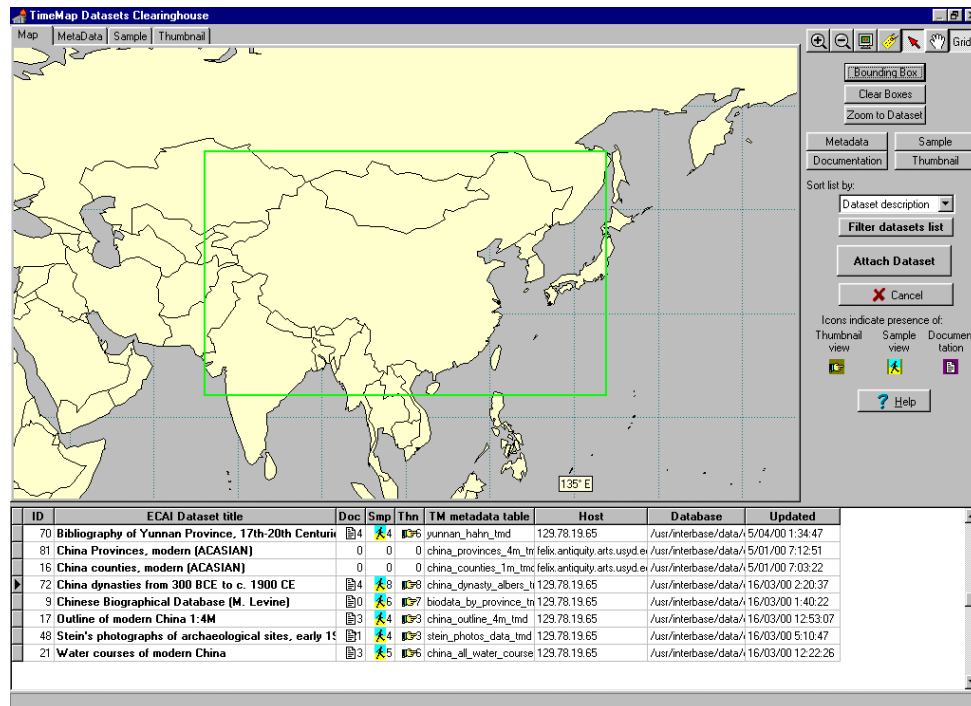


Fig. 10: Metadata import dialogue (HTML → Table)

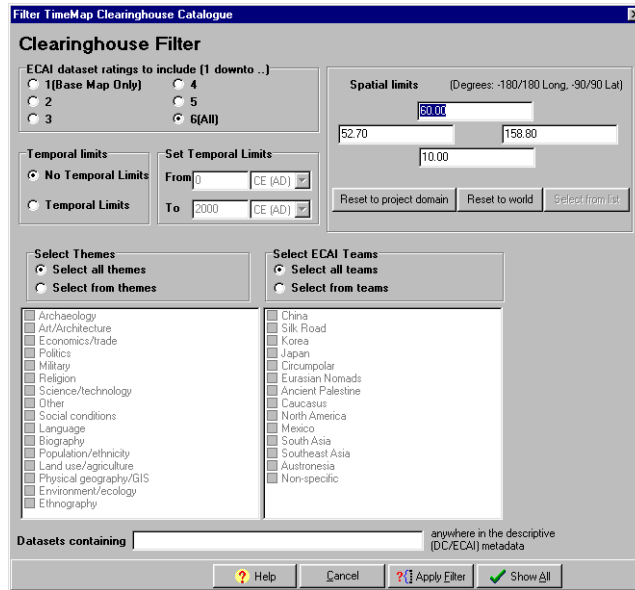
## TMView Clearinghouse browser

The TimeMap viewer program, TMView, includes a revised ECAI Clearinghouse browser which displays a map view on which bounding boxes and snapshot views can be displayed (Fig. 11).

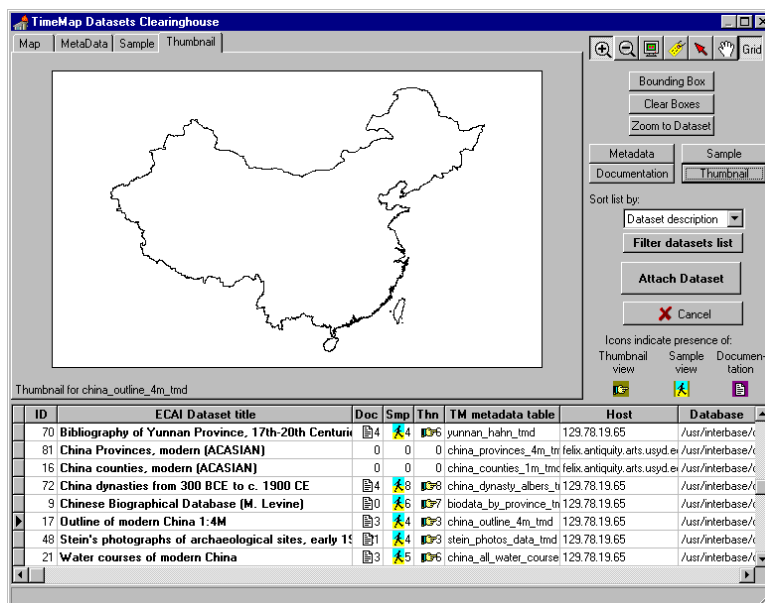


**Fig. 11: TMView Clearinghouse browser showing filtered list of datasets (bottom panel) and bounding box for selected dataset (map panel)**

Datasets displayed in the list of datasets - which appears in the lower panel of the clearinghouse browser - can be filtered according to keyword, spatial and temporal limits, ECAI rating, themes or teams (Fig. 12). Spatial limits are set automatically from the map currently shown in the clearinghouse map browser.



**Fig. 12: TMView Clearinghouse filter dialogue – spatial limits are set automatically from the current project or the clearinghouse browser map (Fig. 11)**



**Figure 13: Clearinghouse browser showing a dataset thumbnail**

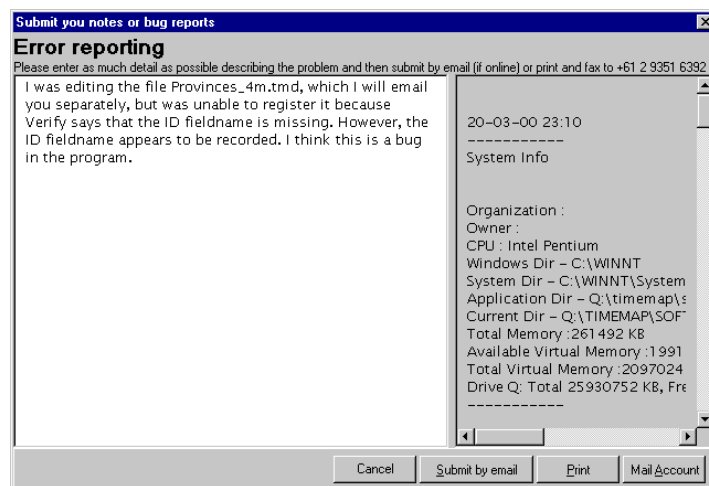
Datasets can be further assessed by displaying a thumbnail or sample view (Fig. 13) of the data. Thumbnails and sample images show, respectively, the entire area or a zoomed-in view of the dataset in question, and are downloaded from the metadata stored with the dataset on the remote server.

Once selected in the browser, datasets can be downloaded for display as a map layer in TMView or other software (datasets are cached locally as MapInfo tables which can be opened independently of TMView if required). During the download process, filtering

can be applied to retrieve a selected set of records (for example, to retrieve records of a particular period from a much larger catalogue).

## Error reporting

Both TMT and TMView have a special function to help users report program errors to the TimeMap development team. The dialogue shown in Fig. 14 allows entry of a description of the problem which is emailed direct to the software developers (or may be printed if the computer is not connected to the Internet). In addition to the user's error report, the program emails extra information (in righthand panel) to assist in debugging the problem, such as the content of the database configuration scripts and version numbers of operating system components.



**Fig. 14: Error reporting dialogue for TMT and TMView programs**

## Conclusion

Over the last few months, TimeMap software development has focussed on the integration of separate data preparation tools (metadata editor and data pump) into a single program – TMT - and the extension of this program with the metadata wizard and improved connectivity. TMT was completed at the end of March 2000 and currently supports MySQL, Interbase, Sybase and ODBC connectivity.

Development efforts are now focussed on debugging of dataset connectivity and SQL queries in the TimeMap viewer program, TMView. Conversion of TMView late in 1999 from a program dependant on MapInfo Professional to a stand-alone program based on the MapX mapping component (also from MapInfo Inc.) have run into some problems associated with the speed of creation of map layers from time-based or SQL selections.

We hope to have these problems fully resolved by the time of the next ECAI meeting in June.

In the meantime, we will be distributing TMT together with a version of TMView without the time selection capabilities, so that ECAI participants can provide feedback on the data preparation process and shortcomings in the software and documentation. Downloads of the TimeMap software and documentation are available from the TimeMap web site (<http://www.TimeMap.net>).