CURRENT ISSUES ON HANDLING GEOREFERENCED RASTER DATA

G. Bitelli - DISTART Dept., University of Bologna, Italy

Commission VI, Working Group 3

KEY WORDS: Raster, databases, GIS, Standards, Metadata, Georeferencing

ABSTRACT

Proposals regarding the adoption of metadata management systems and the use of standard interchange formats are discussed within the framework of an examination of the problems raised by the handling of georeferenced raster data. Such solutions must have the aim of simplifying the sharing and exchanging of raster data, thus contributing to an improvement in the overall quality of processes using the data.

1. INTRODUCTION

Raster databases build all features from grid cells in a matrix: the size of a cell may be coarse or fine, depending on the data source system.

The term "raster data" when applied to geomatics covers a variety of data sources, including:

- scanned images of aerial photography, greyscale typically 8-bits per pixel at between 600 and 3500 DPI, or true color (24-bit RGB);
- digital orthophotos;
- images directly acquired by digital cameras;
- multi-band remote sensed satellite imagery, for instance 7 bands of 10 bits per pixel;
- scanned map images: monochrome images of multiple printing separates at reproduction resolution (up to A0 size at 800 to 2000 dpi), color images of printed maps and charts, typically 8-bits per pixel at between 100 and 800 dpi;
- Digital Elevation Models (DEM): regular grid samples of ground height, often as 16-bit integers or 32-bit floating point values;
- gridded products of analysis procedures performed in a Geographic Information System (GIS).

In recent years there has been a significant increase in the use of the raster format: this has been encouraged both by developments in techniques for the acquisition of data in digital format, and an increase in the performance and capacity of computer systems, from processors, memory devices, peripheric input-output devices, to basic and applications software.

As for hardware systems, the problem of data storage is today still one of the main obstacles to wider adoption of raster data; and is often even seen as the main problem. Yet it can be addressed through various strategies, which can be integrated with each other:

- the availability of new high-capacity systems (very large hard disks, special magnetic tapes, magneto-optical devices, DVDs) with reduced access time;
- adopting image compression techniques, already successfully used experimentally in softcopy photogrammetric image processing and digital map distribution;
- eliminating storage of derived data (for example obtained using analysis modules in a GIS environment), recording only the processing required to generate them.

Similarly, the problem of the availability of quality output devices at acceptable costs is losing importance thanks to technological advances in the sector of raster printers.

It is thus possible to state that other issues, linked to improved handling of geographical data at a general level, are more important.

Data in raster format are ever more widely used, in particular in GIS systems, alongside or as an alternative to data in vector form. Many GIS systems, but also CAD or photogrammetry products, today support both models, with the advantage of being able to exploit the potential offered by each: thus the choice between raster and vector is no longer a radical one. Also in the making of cartographic products, hybrid techniques have been tried for updating and mixing raster and vector overlays: a common phenomenon in many cartographic agencies is that the raster, thanks to scanning map printing separates, constitutes a first step in the transition towards digital methods. Cartographic scanning is, on the one hand, a necessary stage towards creating vector overlays, and, on the other, provides products of great value per se. Indeed, raster cartography has today an important market in many applications: for instance, an interesting use is as a backdrop for applications specifically developed for "in-the-field surveying" or vehicle navigation, e.g. by integration with GPS systems. Among other things this means that map makers have become digital data providers (Woodsford, 1992) and need to consider the problem of making the data available so that they can be used by users on the majority of software systems.

A further reason for the spread of raster data is given by the dissemination of geographical data of this kind over the Internet: protocols developed to prepare documents that can be consulted on the net are in fact always oriented towards the use of raster rather then vector standards.

One of the first applications oriented towards geographical data on the World Wide Web was the interactive consultation of catalogs of remote sensing images: direct access to low resolution previews of images by standard web browsing software can enable the user to immediately evaluate their characteristics and their suitability for a particular application. Similarly, providers have sprung up to offer consultation and on-line downloading of different kinds of maps.

One of the major challenges in competition between GIS software producers is at present centered on obtaining software architectures that make interactive consultation possible of geographical data distributed on the network, with all the procedures and techniques typical of GIS systems.

Given this multiplicity of opportunities for use by many different users of raster datasets, the problem of documenting and standardizing data characteristics assumes increasing importance, so that any user will be able to integrate it immediately within his own system correctly cartographically speaking. There are two ways of obtaining this:

- by developing data exchange standards specifically designed for raster geographic imagery;
- by implementing standard systems for metadata documentation: metadata, or "data about data", constitute complete documentation about a dataset's content.

These two routes are not mutually exclusive and may be developed in parallel, although obviously the former is more specifically suited to direct use of data in specific applications, and the latter to the setting up and consulting of broad spectrum databases. In paragraph 4 we will examine the possibility for integrating them.

Clearly one of the most important and delicate properties of a dataset raster, which per se has an extremely simple structure, consists of its geographical characteristics: a dataset is georeferenced if the data were recorded correctly at the Earth's surface. This implies knowing the position of pixels in relation to a cartographic system and knowing how the latter is linked to the object's real position in the world. While the first property can be expressed by analytical relations, the second requires the introduction of standard conventions and reference values that must be appropriately described and cover a reasonably complex series of case types. Both sorts of information are a part of the set of requirements that are indispensable to define and to guarantee the quality of data.

It must be stressed that although here the accent is placed mainly on applications and raster data of a territorial character, the management of images of metric content concerning other types of object is equally interesting. Think for example of a rectified image of the facade of a historical building subjected to photogrammetric survey. In such cases cartographic data are often non-existent, but it remains important to maintain local georeferencing data to enable immediate integration of the raster with other vector layers relating to the same object and obtained using the same reference system; at the same time, ancillary descriptive data about the survey can provide the support for reusing the images in the future. There exists a large number of information systems that catalog the cultural heritage and they could find a similar type of discussion stimulating.

2. DATA DOCUMENTATION AND METADATA

The problem of raster data documentation and of the development of metadata must be seen within the wider debate at the international level about the dissemination of geographic information (GI); there is an awareness within many bodies that a consistent effort still needs to be made to widen the use and availability of geographical information. There are political, economic, legal and organizational problems, while the value of resources is frequently underestimated: many agencies do not make their information available because they do not realize it has value or because they do not know how to make it circulate, or simply because it is not their job to do this.

Furthermore, many bodies have important datasets, but the significance of their documentation as an aspect of funds invested is not grasped: where data are not well documented, often, over time or because of personnel changes, the availability, content and quality of data become unknown. Moreover, this very often produces unnecessary duplication of data and effort. All the foregoing may be summed up as the lack of a data quality culture.

A dataset may be documented by an internal procedure within the body that originated it, but it would be much better if it were linked to the data exchange format, or if common metadata formats were adopted through the development of full-blown Geographic Metadata Management Systems (in short GMMS).

According to the developing ISO/TC 211 standard (ISO, 1997), metadata could be referred as "data about the content, quality, condition, and other characteristics of the data". Then, a metadata standard could help people determine what data are available, whether they meet their specific needs, how to acquire them, and how to transfer them to a local system. It can also help people who generate geospatial data to share them with others; this could reduce project execution times, reduce costs by minimizing effort duplication and in general improve the quality of processes that use the data. We can expect furthermore that it will help expand the market for GIS data.

There follows a brief description of two of the most important of the many metadata systems developed at the level of international scientific cooperation and within the framework of private and public bodies.

As known, within the USA National Spatial Data Infrastructure framework, the Federal Geographic Data Committee (FGDC) published in 1994 the version 1.0 of "The Content Standards for Digital Geospatial Metadata" (CSDGM): these specifications provide not only a structure for delineating and describing the information necessary to document digital geospatial data resources but also an important set of terminology and definitions (334 different metadata elements with their production rules). The Content Standards is rather complex for the occasional user, but only a subset of the elements is strictly mandatory.

The FGDC standard defines data elements, but not a concrete format to exchange these informations, for the following major sections (Fig. 1):

- 1. Identification Information: basic information about the dataset (title, geographic area covered, currentness, rules for acquiring or using the data, etc.);
- 2. Data Quality Information: assessment of the quality of the dataset (positional and attribute accuracy, completeness, consistency, sources of information, methods used to produce the data, etc.). Recommendations on the information to be reported and tasks to be performed are in the SDTS - Spatial Data Transfer Standard (Dept. of Commerce, 1992);
- 3. Spatial Data Organization Information: mechanism used to represent spatial information in the dataset (method used to represent spatial positions directly and indirectly, number of spatial objects in the data set, etc.);
- 4. Spatial Reference Information: description of the reference frame for coordinates in the dataset (map projections and grid coordinate systems parameters, horizontal and vertical datums, coordinate system resolution, etc.,);
- 5. Entity and Attribute Information: information about the content of the dataset (names and definitions of features, attributes, attribute values and domains, etc.);
- 6. Distribution Information: information about obtaining the dataset (distributor, available formats and media, costs, etc.);
- 7. Metadata Reference Information: information on the

currentness of the metadata information and the responsible party.

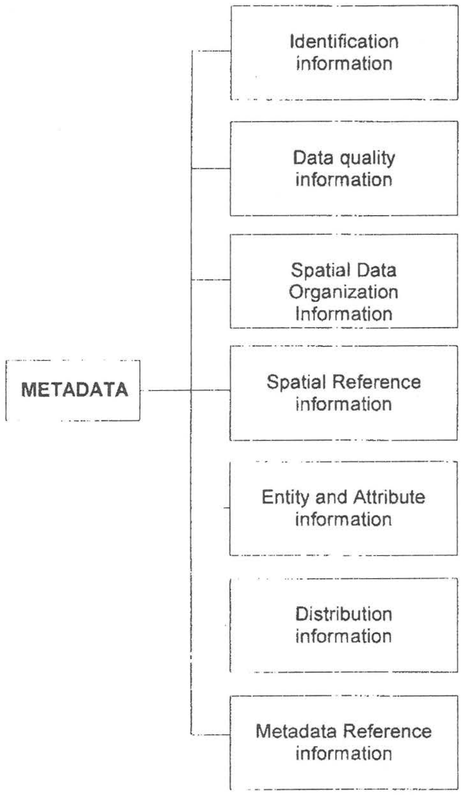

Figure 1 - The major sections of the CSDGM.

Referring to raster dataset description, the elements exclusively referred to the raster format are few, but the user can archive a lot of useful general data, from cartographic reference information to project and sensor characteristics (e.g. for the images deriving from a photogrammetric survey, it could be helpful to store the project figures, the reference for calibration certificate data, etc.), also if the standard is not specifically designed for these.

For implementing metadata standard management, dedicated software can be realized, or templates developed using normal word processing, database or GIS programs running on various hardware platforms. Note that the CSDGM, as its name implies, specifies only the content of the metadata, not its format: this can create difficulties in incorporating data from different organizations who use different metadata tools or template documents into a common clearinghouse. Several implementations of CSDGM metadata generators, functional for general purpose applications, are currently available: figure 2 illustrates as an example some screen copies obtained using the NBII MetaMaker application, developed in US for the National Biological Information Infrastructure (Schneider and White, 1996).
Another interesting GMMS, accessible world-wide, was developed by the European Community Center for Earth Observation (CEO) for its International Directory Network (IDN): it contains metadata describing thousands of datasets of interest to Earth Observation and Global Change data users. Each data set (or collection of data sets) is described within the IDN by an entry called a DIF (Directory Interchange Format).

The DIF contains information describing: Data centre, Personnel contact details, Data sources (e.g. satellite or in-situ sensors), Disciplines, Parameters, Time and geographic area coverage, Potential usage of data, Distribution policy. A simple Microsoft Windows-based metadata generator is also available to provide off-line entries to the directory.

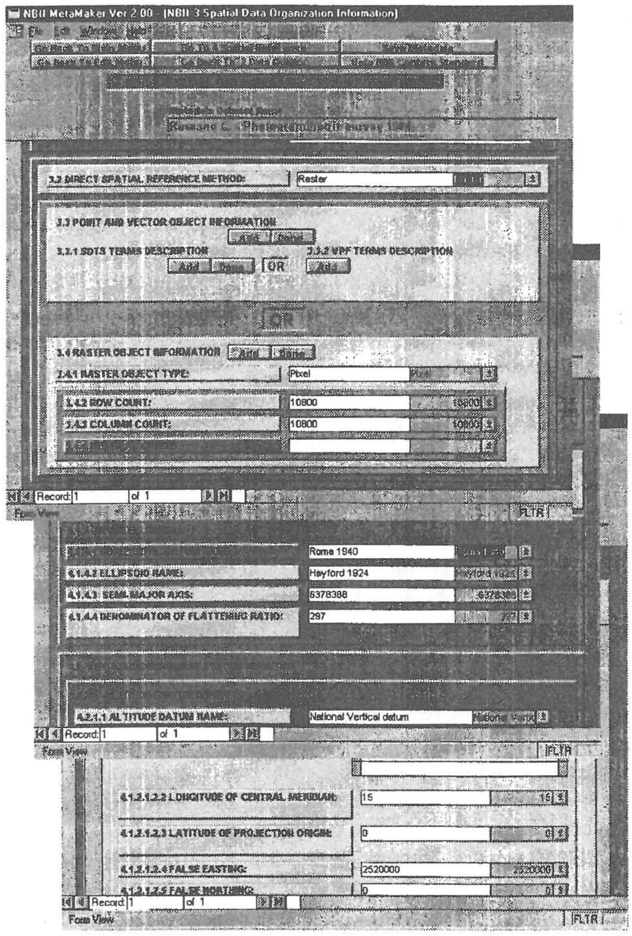

Figure 2 - MetaMaker, an example of a CSGDM compliant metadata tool generator: definition of a raster object (top), of a geodetic model (center) and of the parameters for a projected coordinate system (bottom).

Digital metadata may be of course stored and exchanged in a variety of formats. The most basic is an ASCII text document. this is easy to transfer to other users independent of the hardware/software platform they use. Another common format is Hypertext Markup Language (HTML). HTML provides an attractive way to view metadata using a Internet browser. Recently, there has been also strong interest in creating metadata in Standard Generalized Markup Language (SGML).

3. STANDARD RASTER DATA FORMATS

A lot of interchange standards for raster data are in use today. At the end of 1996 CEC's DGXIII/E Commission reported the following: Digital Data Exchange Specifications (DDES), Fax Groups 3 & 4, Fractal transform coding, Graphic Interchange Format (GIF), Image Interchange Facility (IPI-IIF), Joint Bilevel Image Group (JBIG) standard, Joint Photographic Experts Group (JPEG) standards (including SPIFF), ODA

Raster Graphics Content Architecture (ODA RGCA), Portable Network Graphics (PNG), Photo Compact Disc (Photo CD), Tag Image File Format (TIFF). To these must be added some formats developed by individual producers and now common for tranfer: .BMP, .PCX, .SUN, .TGA, .XBM, etc.

Some comprehensive GIS-oriented standards, currently available or being developed, also provide support for raster data:

- DIGEST (Digital Geographic Information Exchange Standards) is a complex specification, put forward as an ISO standard, developed for exchange of data related to medium and small scale geographic data initially for military applications. At the conceptual level the DIGEST format resembles the US Standard Data Transfer Specification (SDTS);
- NTF (Neutral Transfer Format) was developed by the British Standards Institute (BSI) for interchanging geospatial information within the UK;
- CEN TC287 is under development by the European standardization organization for geographic information: it will define European Norms for Geographic Information by a set of inter-related standards (Reference Model, Spatial schema, Quality, Metadata, Transfer, Position, Query and update, Geographic identifiers);
- ISO TC211/ISO 15046 is also under development: it will define a methodology for creating geographic object, attribute and relationship catalogs and will cover all aspects of GI;
- OGIS (Open Geodata Interoperability Specification) is a standard being developed by Open GIS Consortium, not approved by any international standardization organization. The aim is to provide an application developer to use any geospatial data and any geospatial function or process available on Internet within a single environment and a single workflow (OGIS, 1996).

Thus many generic raster formats exist but they can hardly ever be used to carry within themselves cartographic information about a dataset; in addition, GIS-oriented international standards are not always already released, or turn out to be too complex to be adopted by a single organization which does not deal exclusively in data providing.

Some proprietary raster geographic formats have also emerged, developed by companies producing GIS systems (Intergraph, ESRI, ERDAS, etc.); since their structures are not in the public domain they can only be utilized on the appropriate software platform.

As a result, when images have to be exchanged between different geographic information systems, the georeferencing information can be partially or totally lost, and in any case a complete data exchange operation normally calls for appropriate and by no means easy procedures from an expert user. Frequently, very basic intermediate files are used for this purpose: for example, what is known as the ESRI's Image World File (when accompanying a TIFF file it assumes the .TFW extension) stores in ASCII format the X and Y resolutions, the terms of image rotation, the ground coordinates of the center of the upper left pixel; no information is however kept about the cartographic characteristics of the image or its derivation.

Recently to overcome these problems a format called GeoTIFF has been put forward (Ruth and Ritter, 1995). Born in the early 90s the project was brought forward by a small group of experts from the world of industry, scientific research and government bodies. The aim of the promoters of the GeoTIFF format is to realize a completely documented, public domain, multi-platform standard, with all the characteristics of existing proprietary raster systems but capable of surpassing them in completeness and potential utilizations, with a particular attention to geodetic and cartographic aspects.

The GeoTIFF format is in practice an extension of the TIFF (version 6) format, the most widespread public domain and multi-platform raster interchange format (Aldus, 1992). This format was adopted for a variety of reasons: its ability to support new estensions transparently for the majority of users, because it is capable of supporting different options for compression and tiling, and because it supports the wide variety of image types used in geographic imagery. In 1995 (Ritter and Ruth, 1995) six new TIFF "tags" were defined and made public: they support the information set necessary for georeferencing and geocoding an image; note that the term geocoding in GeoTIFF terminology refers to defining how the object space (e.g. projected coordinates) is referred to points on the earth.

The presence of these "tags" entered into the binary TIFF file does not prevent the file from being correctly read and interpreted, as far as its standard components are concerned, by any type of software which does not support GeoTIFF.

Proceeding to a brief description of the features of the format, a first note of interest concerns the system adopted for raster coordinates. The standard ensures a distinction is maintained between data whose value refers to the cell area (associated to the coordinates of the center of the pixel) and point data "at" a coordinate location (Fig. 3): an example of the first case would be an image acquired through CCD sensors (the sensor aggregates the photons in correspondence with its area), and of the second a digital elevation model (DEM) acquired by actual elevation posting (e.g. through photogrammetric plotting) and thus not referred to mean height values. A similar distinction can be easily found in another sector between different types of spatial interpolators provided by GIS modules.

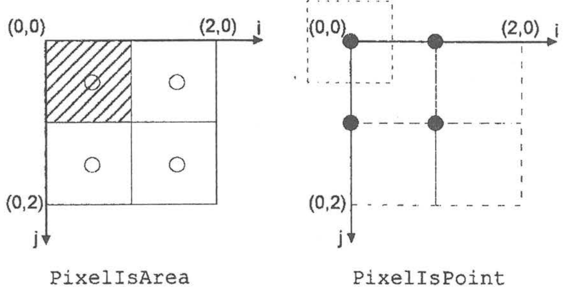

Figure 3 - Types of raster image (example 2x2): data may refer to an area (the value of the pixel refers to the entire cell) or to a point.

This specificity of raster data is often overlooked, with serious consequences at the level of data quality. It is not only a question of the physical meaning to be given to the information: ignoring this problem brings the risk of introducing false data translations, corresponding to half a pixel, which then induce inexact results in the GIS analysis stages or when generating new derived data. Unfortunately it must be observed that this aspect is frequently undervalued and undocumented both by the person providing the data and by those who produce interpolation or modeling software.

The overall process of geo-cartographic definition for an image is described in Figure 4, which shows the six tag names. Each tag requires appropriate parameters. The ModelTiepointTag needs for instance the pixel location (I,J) in raster space and the

corresponding (X,Y) coordinates in object space, the third dimension being not truly managed in the current first release of GeoTIFF. The ModelPixelScaleTag maintains in the object space units the scale values (raster pixel spacing) in X and Y directions. The ModelTransformationTag stores the terms of a transformation matrix between the Image Space and the Object Space (in GeoTIFF terminology Model Space).
The relation between Image Space and raster device space (monitor, printer) is managed through appropriate tags already provided by the TIFF 6 standard. The transformation between Image Space and Object Space is documented by recording the coordinates of control points known in both systems and/or the parameters of bidimensional or tridimensional rototranslation.

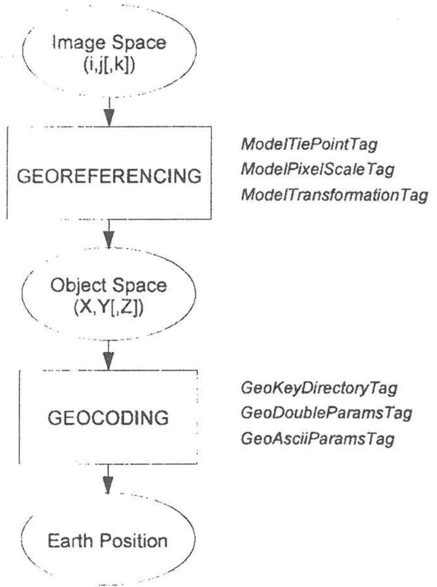

Figure 4 - From image coordinates to the knowledge of position on the earth.

The system even allows a single point of known image and earth coordinates to be entered without a transformation parameter; which indicates how important it is to keep any type of information that is available, even if it insufficient to register the image. Table I shows some frequent cases in georeferencing an image and the use of the tags provided by the standard for documenting the transformation. Note that the last case is not recommended: the best solution is to put an image previously geometrically corrected into a standard projected coordinate system.

case	*tag(s) to use*
1 point known, scale and rotation unknowns	ModelTiePointTag for the point
1 point known, scale factor(s) known, no rotation required	ModelTiePointTag ModelPixelScaleTag
3 non-collinear points known, linearity of transformation unknown	ModelTiePointTag for the three points
affine transformation required	ModelTransformationTag (expressed by a 4x4 matrix)
rubber-sheeting transformation required	ModelTiePointTag for some points

Table I - Common cases in image georeferencing and corresponding GeoTIFF tagging.

Any system that can interpret GeoTIFF can, by reading the georeferencing tags, extract the necessary information to immediately position - if possible - the image in the cartographic system associated with it, then integrate it directly and with maximum precision with the other datasets that may be available for the same area (see example in Figure 5). This certainly ensures a good data fruition and quality standards.
The complex cartographic or geodetic information which make it possible to know by means of what cartographic transformations the data were obtained and with reference to which datum, is maintained by means of a main "metatag" (GeoKeyDirectoryTag) and two accessory tags. In a very smart way these support an information structure obtained through codes organized in a pseudo-hierarchical manner.

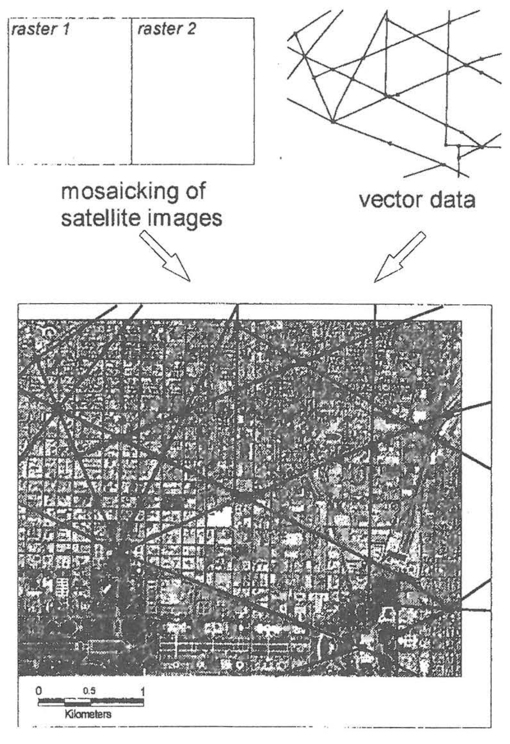

Figure 5 - Georeferenced GeoTIFF images (e.g. from satellite) can be automatically mosaicked and imported into a GIS in the correct place as a layer co-registered with existing layers in vector format (e.g. road network).

The conceptual schema for geocoding information (see Figure 6) and the related numerical codes that identify each element are taken from the second release of the model developed by EPSG (European Petroleum Survey Group) within the framework of the POSC (Petrotechnical Open Software Company), a non-profit organization dedicated to defining standards for the petroleum sector; for the complete list of attributes and values defined by this comprehensive model see (POSC, 1995). Obviously, not all the parameters are always mandatory; when using a coded Projection Coordinate Systems, for instance, it is not necessary to identify the projection datum, the coordinate transformation method with associated parameters, etc.: they are implicitly defined by the first. It is also possible to introduce user-defined datums.
Vertical datums are not implemented yet in GeoTIFF 1.0 but their support is planned for the release 2.0 in order to fully describe DEMs and similar data.

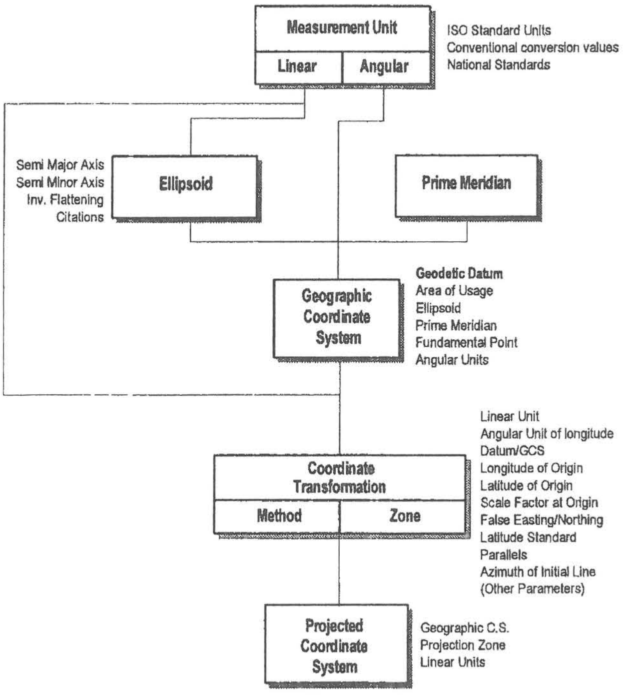

Figure 6 - Conceptual schema for cartographic metadata embedded in GeoTIFF format (ad. from EPSG-POSC model 2.0).

Some large GIS software companies are now providing support for reading and, not always, writing in GeoTIFF; the same is for several important data providers, mainly in the remote sensing sector. Besides a consistent simplification in raster data management for the user (no need of duplicate the data in different formats in order to share them with other users, no need to run translators, quick georeferencing phase, etc,), also for the GIS software producers the adoption of this format can introduce significant saving in software development and testing, eliminating the problems related to the support of a multitude of data drivers.

4. DATA AND METADATA INTEGRATION

GeoTIFF is an example of a multi-platform format that embeds in a single file data and common metadata contents. The latter are compatible with the specifications introduced by (FGDC, 1994) and are in practice related to the cartographic properties handled by section 4 of the CSGDM, but the format is expected to become a superset of FGDC spatial addressing content requirement.

This leads to the idea that adoption of public domain data standard formats like GeoTIFF could make it possible to realize strict integration between practical dataset administration and some GMMSs, enabling interesting solutions for efficient and automated management of a clearinghouse for raster data. Figure 7 presents the general schema of such a project, that can support the basic functionalities of archiving, querying and retrieving of data and metadata.

A data-driven interface thus designed can support two basic information exchanges:

- physical file characteristics and georeferencing information could be automatically transferred from a raster dataset to the GMMS;
- conversely, a raster dataset in a standard georeferenced format can be created by appropriately and automatically embedding in the raw raster file some information, primarily cartographic, derived from the metadata description.

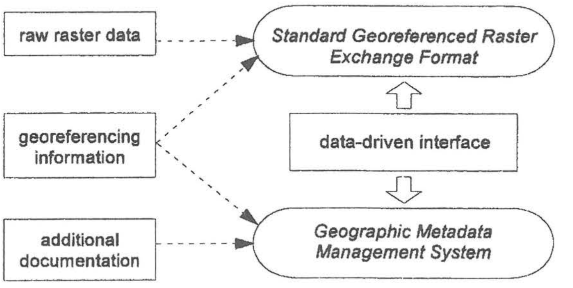

Figure 7: a system for integration of raster data and metadata.

This type of approach would be particularly important for public organizations whose purpose is to manage large cartographic archives in digital form or databases of remote sensing and photogrammetric imagery. Clearly, the great potential of such a project could be achieved if it should be realized in a network environment.

CONCLUSIONS

The emergence of the quality factor in geographical data, together with their ever increasing dissemination in different application environments makes the adoption of advanced forms of standardization essential, both for the transfer of datasets and for the realization of structures describing the datasets. As regards the handling of raster images, several present-day solutions have been illustrated together with a project for possible integration between the two options delineated.

REFERENCES

Aldus Corporation, 1992. TIFF (Tag Image File Format), Revision 6.0.

Dept. of Commerce, 1992. Spatial Data Transfer Standards (SDTS) (Federal Information Processing Standard 173). Department of Commerce, National Institute of Standards and Technology, Washington, D.C.

Federal Geographic Data Committee, 1994. Content Standards for Digital Geospatial Metadata. Federal Geographic Data Committee, Washington, D.C.

International Standard Organization (ISO), 1997. ISO Technical Committee for Geographic Information and Geomatic (ISO/TC 211/PT) Terminology. Secretariat Swedish General Standards Institution, Stockholm.

Open GIS Consortium (OGIS), 1996. The OpenGIS Abstract Specification: An Object Model for Interoperable Geoprocessing, Revision 1. OpenGIS Project Document Number 96-015R1, Wayland.

Petrotechnical Open Software Consortium (POSC), 1995. EPSG geodetic tables 2.0. Available by anonymous ftp from: ftp://ftpmcmc.cr.usgs.gov/release/geotiff/jpl_mirror/tables/

Ritter, N. and M. Ruth, 1995. GeoTIFF Specification 1.0. Available in HTML by anonymous ftp from:
ftp://ftpmcmc.cr.usgs.gov/release/geotiff/jpl_mirror/spec/.

Ruth, M. and N. Ritter, 1995. GeoTIFF Format for Exchange of Raster Data. In: Proceedings GIS/LIS Conference.

Schneider, D. M., and B. White, 1996. User's Manual: NBII MetaMaker Version 2.0. U.S. Geological Survey, Midcontinent Ecological Science Center, Fort Collins, Colorado. Published by the Environm. Management Technical Center, Onalaska, Wisconsin, 19 pp.

Woodsford, P. A., 1992. Map maker and data provider: cartographic agencies must fill both roles. GIS Europe, 1/9, pp. 26-32.