The Venice Lagoon Experimental GIS at the IUAV

Markus M. Hedorfer[1] and Alberta Bianchin[2]

[1] via Ca' Rossa 93,
I-30174 Venezia-Mestre VE, Italy
hedorfer@cidoc.iuav.unive.it
WWW home page: http://www.iuav.unive.it/~hedorfer
[2] Dipartimento di Analisi Economica e Sociale del Territorio (DAEST)
Istituto Universitario di Architettura di Venezia (IUAV), Santa Croce 1957,
I-30135 Venezia VE, Italy
alberta@brezza.iuav.unive.it

Abstract. The construction of a geographical information system on the Venice Lagoon was made particularly difficult due to the predominance of dynamic phenomena and, on the organisational level, to the absence of coordination between competent authorities. Thus, the first step in this work consisted in assessing and classifying existing data and in performing a problem recognition analysis by contacting several public authorities and university departments. Subsequently, data have been entered and manipulated by adopting a strict distinction between static and dynamic data and by employing two additional modelling tools configured as external modules to the main GIS environment. For this purpose, fully automated data exchange procedures have been developed as well. As the experience of the data assessment phase demonstrated that data exchange between organisations is nearly impossible because data are actually produced without any further information, particular attention was devoted to provide for procedures facilitating the metadata production process.

1 Introduction

The *Venice Lagoon Experimental GIS at the IUAV* – or shortly *SISALV*[1] – is a part of a larger research initiative, called *The Venice Lagoon System*, promoted by the Italian Ministry of the University and Scientific and Technological Research that involves several universities and other public and private organisations working on items related to the environmental safeguard and socio-economic renewal of the Venice Lagoon area, including also its largely urbanised mainland. In particular, this work moves from Research Task 4.4. "Optimisation of Base Cartography Supporting Items Related to the Venice Lagoon System"[2]

[1] Original name in Italian: *S*istema *I*nformativo Geografico *S*perimentale presso l'Istituto Universitario di *A*rchitettura di Venezia per la *L*aguna *V*eneta.
[2] Original name in Italian: Linea di Ricerca 4.4 "Ottimizzazione di Basi Cartografiche a Supporto di Tematiche Relative al Sistema Lagunare Veneziano".

carried out by the Department of Spatial Economic and Social Analysis of the Venice University of Architecture (IUAV).

The development of a digital base cartography for the Venice Lagoon – that is, collecting and managing geographical data to be employed by other researchers or by public and private authorities – is made particularly difficult due to the following two reasons.

1. Amphibious territories are highly influenced by dynamic phenomena related to tidal and other periodic and transient factors, so that experiences in traditional digital cartography can hardly be employed.
2. On the organisational level, no systematic coordination between competent public and private institutions and no data infrastructure have been established yet.

The first step in the present research task consisted in achieving the necessary knowledge on how dynamic phenomena act on the territory to be surveyed and on which data can be classified as significant in lagoon-related management purposes. Several interviews with experts in environmental sciences, tidal physics, physical chemistry and marine biology, and with officials of the Venice Water Authority[3], the Department of Ecology of the Veneto Regional Administration and the Consorzio Venezia Nuova (CVN)[4] have been carried out prior to proceed with the data base development. The contacted persons were also asked to describe the data they produce and employ. The main results of this problem recognition analysis can be summarised as follows.

- The only static phenomena in the Venice Lagoon are human constructions, such as buildings or seawalls strengthening vulnerable island limits.
- All other phenomena usually classified as static, such as natural shore lines, must be considered dynamic phenomena as they change in a not foreseeable manner during few years.
- Low water areas, that represent a considerable part of the whole lagoon area, must be considered as periodic dynamic phenomena as they change with the tidal movement.
- The same low water areas – if they are represented using high spatial resolutions (accuracy of about 5 meters or less) – must be considered transient dynamic phenomena as their exact limits change daily in a not foreseeable manner.
- Environmental indicators, such as heavy metal concentrations, temperature, or algae distribution, can be represented only by employing mathematical modelling or simulation techniques.

[3] Magistrato alle Acque di Venezia.

[4] The Consorzio Venezia Nuova (Consortium New Venice) is an organisation of Italian private enterprises that work, under the supervision of the Italian Government, on items related to the Venice Lagoon and maintain a huge data base concerning its environmental and socio-economic survey.

The second item to be analysed concerned the organisational structure of the Venice Lagoon area. The absence, at the present day, of any systematic coordination between the several actors involved this research to observe also the following.

– It will be unavoidable to establish a coordination board common to all public and private institutions working on data concerning the Venice Lagoon.
– The problems caused by the presence of different geographic and other information management software, including self-made applications, must be overcome by establishing appropriate data exchange protocols[5].
– Highly specialised research and administrative organisations, such as the hydrodynamic modelling group at the CNR-ISDGN[6] or the Venice Township Tidal Forecast and Signalling Centre[7], must be enabled to communicate with generalpurpose data management groups.
– Environmental water survey organisations, such as university departments or other research groups at the CNR-ISDGN, must be enabled to carry out parameter sampling sessions by collection additional data in order to allow a correct use of sampled data also by modelling specialists and general-purpose data managers.
– All produced and/or manipulated data must be supplied including the corresponding metadata.

2 The Organisational Model

Many of the purposes defined during the preliminary stages could not be realised as they are beyond the competencies of this research. However, the organisational model applied to this experimental implementation of an inter-disciplinary data production and management system reflects most of the aims to be pursued. A modular logical structure has been adopted by the SISALV, which is composed by the following components.

1. A *main GIS environment* – based on the Arc/Info software package and customised using the Arc/Info Macro Language (AML), small programs written in C/C++, and UNIX Bourne Shell scripts – acting as centre around which are grouped several modular extensions as well as the organiser of the handled geospatial information.
2. A work environment – written in the Prolog programming language for Microsoft Windows platforms – managing the *Sea Transformation [Cellular] Automaton (SeTA)*. This environment acts as stand-alone application and allows – using the cellular automata technique (analysis of cell-neighbourhood

[5] During this research, a possible data exchange protocol for point, line, polygon, raster and alphanumeric data has been developed.
[6] Istituto per lo Studio delle Grandi Masse (ISDGM, literally "Huge Masses Research Centre") at the Italian National Research Council (CNR, Consiglio Nazionale di Ricerca).
[7] Centro di Previsione e Segnalazione delle Maree.

properties and their transformation during a time interval based on specific rules) – to produce simulations on changes of the so-called state of water in the Venice Lagoon.
3. A work environment – written in the Fortran programming language for a Digital Alpha station with a UNIX operating system – managing a *Three-Dimensional Finite-Differences Water Quality Eutrophication [Mathematical] Model* on the Venice Lagoon. This model implementation allows to produce simulations on changes of concentrations of eutrophication-related substances and on growth of macro-algae.
4. A command-line *user interface* – based on the Arc/Info user interface and customised using the AML – supporting users in querying the simulations produced by the automaton and the model, and weighted using cross-reference analysis between simulated and other data held by the GIS.

Finally, this structure has been conceived to be able to extend the SISALV incorporating further modules with specific functionalities.

Integration between the two peripheric work environments and the main environment has been realised through the principle of *data* sharing, while the main module shares *procedures* with the user interface.

1. As both, the cellular automaton and the mathematical model, use cell-based data structures, the Arc/Info Raster Exchange Format (RXF) has been chosen as the common data 'language'. The three work environments deposit all their produced data in a temporary data bank and withdraw from it the data needed for further processing.
2. For each work environment, import and export procedures have been provided in order to obtain a fully automated data exchange system. This allowed also to preserve information on data-georeferencing during its processing by the two peripheric work environments that originally treated cellular data only as indexed matrixes.
3. Since both, user interface and main module, are implemented on the same computer, it was not necessary to define a special communication protocol between these two components. User interface and main environment share the whole procedure library written in the AML, C/C++ and in the Shell programming language.

The functionalities of the main GIS environment and the two actually implemented peripheric work environments (automaton and model) can be described by imagining them as if they would work in a parallel manner.

1. The cellular automaton as well as the mathematic model use *raster* data to represent the initial state (automaton) and the initial conditions (model) before executing the simulation procedures.
2. An automaton *rule* defines the probability with which a single cell will change state from A to B. The mathematic model's *algorithms*, on the contrary, are expressed as mathematical functions using more variables, but with unique results.

3. In either cases, manipulation processes are subdivided in *time intervals*: for each interval are simulated the expected changes according to the automaton rules or the model algorithm until the final state or condition is reached.
4. Output data concern the final cell *states* (automaton) or the *concentrations* of substances characterising the eutrophication degree of the lagoon environment (model). This set of data is finally stored in the main module's simulation library.
5. Among the other data used by the SISALV, morphological data (dry land, lagoon canals, lagoon low water, open sea, fisheries, lagoon delimitation, natural and artificial salt marshes) in topological two-dimensional *vector* format, the digital elevation model stored as a *triangulated irregular network* as well as a cell-based *surface model*, and specific information as surface sediment distribution can be mentioned.
6. A set of automatic and half-automatic procedures, located inside the main module, allows performance of crossed *spatial analysis* between vector, raster, TIN, and surface models.
7. On the user interface side, users are assisted in querying the *simulation library* in order to retrieve information on specific environmental situations that not necessarily must have been stored permanently yet.
8. Subsequently, *user queries* are translated into a series of analytical operations that will be redirected to the procedural nucleus of the main GIS environment.
9. One of the main purposes of the cross-reference analysis complex is the production of *weighted simulations* that can be seen as the result of matrix-arithmetical computing procedures.

3 Data

The major problems encountered during the retrieval and manipulation of existing geospatial data were related to absence of almost any data content, organisation and quality information and to the use of improper geographic data exchange techniques. Two-dimensional data (lines and polygons) have been provided as AutoCAD Drawing Exchange Files (DXF), that did not allow to transfer attribute and topological information. In some cases, the only data content information that has been provided consisted in legend frames directly included inside the drawings. The necessary work to prepare those data prior to their conversion and storage in the SISALV was not negligible. Binary-attribute polygon data[8] have been parsed and corresponding coverages have been built by

[8] Those data where the only polygon attribute consists in a positive or negative response. For example in the CVN data base, dry land perimetral lines define a polygon coverage where each polygon either belongs or not belongs to dry land. On its opposite, fishery perimetral lines are not limited to binary data, but include also additional information on the identity (identification number, name) of each single fishery.

an AML procedure developed purposely for this aim. In other cases also topology consistency needed to be restored by performing manual editing sessions. The following list contains a short description of the treated data sets provided mainly by the CVN but also by the Veneto Regional Administration Information System (SIRV)[9].

- *Main Morphology.* The initial purpose for the management of morphological data was to store all the available data concerning *physically perceivable phenomena* inside a unique coverage by adopting an *attribute*-based organisation of data in opposition to the traditional *layer*-based storage technique. In this context, the geometric concept of metric spaces has been adapted to geography by grouping geographic phenomena into logically, and hence metrically, compatible data features. That means that the original data sets entitled *dry land*, *lagoon canals*, *fisheries*, *lagoon delimitation*, *natural salt marshes*, and *artificial salt marshes* provided by the CVN should be treated as different aspects (attributes) of a unique spatial description (data set). Unfortunately, empirical tests on the provided data (data quality information was not available) demonstrated that the two data sets concerning natural and artificial salt marshes were not metrically compatible (congruent), probably in consequence of the absence of data quality assessment procedures during the data production and updating activities.
- *Natural Salt Marshes* and *Artificial Salt Marshes.* In relation to the main morphology, natural salt marshes resulted metrically compatible with the remaining original data sets. However, the relationships between natural and artificial salt marshes did not correspond to what geographic reasoning would expect: any pair of coordinates representing a point in the lagoon can belong to a polygon declared either as natural or as artificial salt marsh. In other words, the two polygon coverages were not non-intersecting. This observation caused both, the artificial *and* the natural salt marshes, to be excluded from the main morphology.
- *1990 Bathymetry.* Provided by the CVN as ASCII text file containing only x-, y- and z-coordinates of the measured points, conversion and archiving of this data set did not cause any kind of problems. It has been stored as point coverage storing depth (z) values as attributes.
- *Digital Elevation Model.* The digital lagoon elevation model has been produced by using the 1990 bathymetry and by assuming a height of +055m and +1.10m respectively for the perimetral lines of natural salt marshes and dry land[10]. In order to avoid depth-value falsing along lagoon canals due to an insufficient number of measured points, interpolation (triangulation) has been performed in two steps: first, the 1990 bathymetry points that fell inside lagoon canals have been isolated from the rest and interpolated and then suitably densified; second, the quoted points inside the lagoon canals have been interpolated with the other points of the 1990 bathymetry and the vertexes of dry land and natural salt marsh perimeters. The resulting

[9] Sistema Informativo della Regione Veneto.
[10] These values are conventional values determined by the CVN.

triangulated irregular network (TIN) has also been converted to cell-based (raster) Arc/Info surface models with both, floating-point and integer number cell values.

– *Surface Sediment Distribution According to Shepard [9].* This data set, containing surface sediment distribution surveyed by Barillari [1][2][3] and Rosso [3] between 1975 and 1981 applying the triangular classification developed by Shepard in 1954 [9] has been provided by the CVN as layered DXF where single sediment class areas were encoded as closed polylines (g-rings) that did not geometrically share common border lines. Furthermore, areas that fell inside dry land were not excluded from the polygon coverage, as the CVN is in the habit of erasing them from the spatial domain performing only a graphic overlay prior to produce printed copies (paper maps) of the data set. The coverage had thus to be cleaned first by merging congruent, or almost congruent, border lines and then to be corrected to obtain thematically consistent data by erasing physically dry land parts retrieved from the above described main morphology coverage.

– *Seagrass Distribution and Characteristics.* In a very similar way to the surface sediment distribution, seagrass distribution and characteristics, surveyed in 1990 under a direct order of the CVN, had to be cleaned prior to include this data set in the SISALV data base. Its layered structure was even less comprehensible while in the DXF polygon and point data – representing respectively continuous/discontinuous and punctual distribution – were merged together, and as different attribute sets (e.g. type of seagrass population and continuous/discontinuous distribution) were encoded as intersecting, not common border-sharing closed polylines. The original data set has finally been divided into two distinct coverages separating point from polygon data.

– *Simulated Eutrophication Indicators.* As described in the previous paragraph, also the values of simulated concentrations at different depths of *ammonia*, *oxygen*, *phosphorus*, *phytoplankton*, *temperature* and *ulvae* with three different nutrient input scenarios and for fourteen 15-day time slides ranging from 90 to 285 days from the beginning date of the simulation. The mathematical model and its processing have been carried out by an expert group at the University Ca' Foscari at Venice, while its fully automised conversion and inclusion in the SISALV data base has been performed by the authors.

Further data sets (*water embankment* polygons and *place names*) could not be included in the SISALV data base because of their insufficient topological consistency (embankments) and the improper way to hold text data which were provided as graphic annotations splitted into single letters rather than feature attributes.

4 Metadata[11]

In addition to the unquestionable advantage of the use of geospatial metadata, the experience made during the assessment, retrieval, conversion, and manipulation of the data treated by the SISALV demonstrated that metadata represent an absolutely necessary part of the data themselves, especially if various actors are involved in the data handling process. In the end, the lack of metadata – or other comparable information – caused several pre-existing data sets to be in practice untreatable by the GIS: in almost all cases data quality information was completely absent, often no lineage information was given, and in some cases even no data content description was provided by the originator.

On the other hand, metadata absence can often be related to the complexity and the huge amount of information to be organised and stored. After the theoretic discussion on the sense of metadata, the major efforts in the metadata debate should now be concentrated on how to simplify the metadata production procedures. To avoid ambiguity, it must be said that not the content of metadata should be simplified, but an efficient organisation of the information to be encoded would facilitate GIS operators while performing the necessary additional work without sacrificing the quality of metainformation. Considering the large SISALV data base, an optimisation of the metadata compilation process became imperative, and the aims to reach in this context have been defined as follows.

1. *Automation* of metadata production procedures anywhere this is possible.
2. *Acceleration* of manual compilation procedures.
3. Deletion of any *redundancy* in the meta-database.
4. *Compliance* of produced metadata to almost one of the international standards.

Several metadata compilation assisting tools already exist. However, after a careful examination of some of the most accredited programs[12], it must be remarked

[11] A more detailed description of the metadata handling principles in the SISALV is available on the Internet at http://www.iuav.unive.it/~hedorfer/metajca1.html. It is a yet unpublished contribution by the same authors of this article, entitled "Un modèle structurel pour métadonnées – A Structural Model for Metadata" (in French, with an English abstract), held at the third conference "Journées Cassini", Marne-la-Vallée, France, 26.-27.11.1998.

[12] The following public domain software has been considered. *Xtme/Tkme* with *mp* and *cns*, a stand-alone software package for several UNIX and the Microsoft operating systems developed by Peter N. Schweitzer at the United States Geological Survey (USGS). *ESRI document.aml*, an application written in the AML for Arc/Info on UNIX systems by the USGS Environmental Protection Agency (EPA) and later ceased to ESRI, Inc.. *Corpsmet95*, a stand-alone software for Microsoft Windows 95 and NT developed by the Coastal Oceanographics, Inc. for the US Army Corps of Engineers. *NBII MetaMaker*, a stand-alone software for Microsoft Windows 3.1, 95, and NT developed as application-independent Microsoft Access Database by the USGS Biological Resources Division for the National Biological Information Infrastructure (NBII) at the USGS. For further information, see also the *Metadata Tools*

that no tool corresponds to all four points, even if any of the considered programs well fulfils at least one of the defined objectives[13]. Therefore, an own organisational structure of metadata and possibly an own collection of isolated but coordinated procedures had to be developed. Once decided which international standard would have been adopted[14], the first step in the approach to the problem consisted in analysing the metadata production rules and classifying the single metadata elements in relation to the possibility to be completely or partially automised, or for which only the manual compilation activity can be optimised (accelerated). A first subdivision of all elements can be performed calling

- *constant* elements those metadata elements whose content will always be the same independently from the person who compiles the metadata, and
- *variable* elements those metadata elements whose content will change entirely on the discretion of the operator.

This radical distinction in two large classes obviously does not reflect faithfully the nature of all the single metadata elements. Furthermore, each element can be characterised as less or more variable – or, in other words, potentially constant – by relating it to the following two data base conformation aspects.

1. *Additional metadata-related data content.* Geospatial data can be articulated by including some additional information used only for metadata production purposes. Full automation of the corresponding metadata elements will then be possible.
2. *Special metadata production conventions.* The metadata production process can also be developed by establishing some conventions including file naming rules, data organisation principles, and the use of template metadata elements.

Survey maintained by Doug Nebert [6], the overview on *Metadata Tools for Geospatial Data* provided by Hugh Phillips [7], and the description of *Tools for Creation of Formal Metadata* published by Peter Schweitzer [8]

[13] *Automation* of many of the metadata elements is realised by ESRI document.aml by using directly the Arc/Info GIS software, *acceleration* of manual entry is fully realised by ESRI document.aml, Corpsmet95, and NBII MetaMaker by a well-developed graphic user interface (GUI), and partially by Xtme/Tkme which uses a more essential GUI, deletion of *redundancy* has been reached by ESRI document.aml and NBII MetaMaker as they use relational database management systems, and standard*conformance* is guaranteed by the Xtme/Tkme/mp/cns package which refers to the US Federal Geographic Data Committee (FGDC) Content Standards for Digital Geospatial Metadata (CSDGM), by NBII MetaMaker which refers to the NBII sector- specific metadata standard differing slightly from the CSDGM, and partially by Corpsmet95 where problems are encountered while trying to encode all possible element combinations of the CSDGM.

[14] In this work, the *Content Standards for Digital Geospatial Metadata (CSDGM)* [5], provided by the *Federal Geographic Data Committee (FGDC)* at the *United States Geological Survey (USGS)* in 1994, has been adopted.

Metadata elements can hence been grouped into four classes.

1. *Pure constant elements.*
2. *Constant elements by data content.*
3. *Constant elements by convention.*
4. *Pure variable elements.*[15]

Limitedly to the five mandatory CSDGM-sections[16], automation and acceleration procedures have been developed and tested using the *Arc/Info Macro Language* for all geographic data query and analysis operations, the *Bourne-Shell scripting language* for the major part of operating and file system-related operations, and the *C/C++ programming language* for AML output parsing, metadata formatting and assembling operations. At this experimental stage of the metadata compilation structural model, users can execute the available functions from inside the main GIS work environment by typing

```
meta <geo_dataset>
```

from the command line. This will cause all necessary data query and analysis functions to be performed and a new or an updated metadata text-, HTML-, SGML- and DIF-output file to be produced. As variable elements can be compiled at any moment by editing the corresponding source text file (for each variable element a single text file is expected, so that users can decide when and how to produce them) are not compiled, but only retrieved if they have been provided yet, and if not all mandatory elements are present error debug messages are included to the output files. Independently from the main compilation macro, each function can be executed separately as well.

References

1. Barillari A.: Prime notizie sulla distribuzione dei sedimenti superficiali nel bacino centrale della Laguna di Venezia – Individuazione dei vari tipi di sedimenti superficiali e loro distribuzione areale (mappa) quale base necessaria per lo studio sui rapporti tra caratteristiche del fondo e popolamenti bentonici. Atti dell'Istituto di Scienze, Lettere ed Arti, Tomo CXXXVI (1977-78) – Classe di scienze matematiche e naturali, Istituto Veneto di Scienze, Lettere ed Arti, Venezia, Italy (1978)

[15] Considering the United States metadata standard, 87 data elements and 18 compound elements have been declared as pure constant, 10 data and 4 compound elements as constant by data content, 20 data and 59 compound elements as constant by convention, and 95 data and 9 compound elements as pure variable. Notice: the total number of CSDGM elements refers to the representation provided by Susan Stitt [10] and does not consider separately the 22 map projections, the 5 grid coordinate systems, and their corresponding 21 parameters.

[16] These are the following: 1. Identification Information, 7. Metadata Reference Information, 8. Citation Information, 9. Time Period Information, and 10. Contact Information. The remaining five sections have been excluded deferring them to more specialised future research initiatives.

2. Barillari A.: Distribuzione dei sedimenti superficiali nel bacino meridionale della Laguna di Venezia, Definizione dei principali tipi di fondali lagunari e possibili relazioni tra caratteristiche granulometriche e popolamenti bentonici. Atti dell'Istituto Veneto di Scienze, Lettere ed Arti, Tomo CXXXIX (1980-81) – Classe di scienze fisiche, matematiche e naturali, Istituto Veneto di Scienze, Lettere ed Arti, Venezia, Italy (1981)
3. Barillari A., Rosso A.: Prime notizie sulla distribuzione dei sedimenti superficiali del bacino settentrionale della Laguna Veneta, Sedimenti, Carta della distribuzione granulometrica dei sedimenti in Laguna Nord, carta della distribuzione di calcite, dolomite, carbonati totali e percentuale di materia organica. Memorie di Biogeografia Adriatica, Suppl. vol. IX, Venezia, Italy (1975)
4. European Committee for Standardisation: European Standard – Geographic Information – Data Description – Metadata. Draft 287009, European Committee for Standardisation, Bruxelles, Belgium (1996)
5. Federal Geographic Data Committee: Content Standards for Digital Geospatial Metadata. Official version distributed electronically using the Adobe portable document format, Reston, Virginia, USA (1994)
6. Nebert D.: Metadata Tools Survey. Published on the Internet at http://www.fgdc.gov/Metadata/Toollist/metatools797.html, Federal Geographic Data Committee, Reston, Virginia, USA (1998)
7. Phillips H.: Metadata Tools for Geospatial Data. Published on the Internet at http://badger.state.wi.us/agencies/wlib/sco/metatool/mtools.htm, Wisconsin NSDI Clearinghouse, USA (1998)
8. Schweitzer P.: Tools for Creation of Formal Metadata. Published on the Internet at http://geochange.er.usgs.gov/pub/tools/metadata/, US Geological Survey, Reston, Virginia, USA (1998)
9. Shepard F.P.: Nomenclature Based on Sand-Silt-Clay Ratios. Journal of Sediment Petrology, Vol. 24, No. 3 (1954)
10. Stitt S.: The FGDC Content Standards for Digital Geospatial Metadata – An Image Map. Published on the Internet at http://biology.usgs.gov/nbs/meta/meta.html, United States Geological Survey, Biological Resources Division, Reston, Virginia, USA (1998)