Action 8.1

Integration of data and results by means of GIS techniques

Geographic data handling

Baseline guide for EXPAH Teams v. 1.0

Authors

M. P. Bogliolo

INAIL - Research - Dipartimento Installazioni di Produzione e Insediamenti Antropici

1. Introduction

Action 8.1 of the EXPAH Project is devoted to optimize the analysis and presentation of intermediate and final results having a geographic component (i.e. concerning characteristics and phenomena related to location on the Earth or varying spatially across a geographic domain), and to develop a visual system to disseminate final information concerning air quality, exposure risk and possible mitigations to stakeholders and to population.

In this framework, the GIS team will operate to support the geographic aspects of the research, by ensuring a rigorous and uniform use of geospatial data needed as ancillary information or for geographic reference; providing for the accurate geo-referencing of spatial data produced during the project; offering tools to overlay and display spatial data from different sources, and to analyse spatial relationships among them.

To perform these tasks, the team will develop and maintain a Geographic Information System (GIS) to collect, manage and share geospatial information and to allow spatial and geostatistical analysis on the data. This GIS will be the basis for the final platform aimed at presenting the project results. The display and analysis system will be fed by a “geodatabase”, that is a relational database where geographic (geometry) and alphanumeric information will be stored.

The GIS software that will be used at the desktop level to manage, process and analyse spatial datasets is the ESRI ARCGIS platform, with the modules for geostatistical analysis, spatial analysis and 3D analysis. Use of open source software and in particular of GRASS as the desktop software and of POSTGreSQL - POSTGIS as the database will be evaluated as a secondary task.

The web GIS server platform will be ESRI ARGIS Server. Use of open source software and in particular of Map Server will also be evaluated as an alternative.

An effort will be made to exploit OGC (Open Geospatial Consortium) standards, and to conform to the implementing rules of the INSPIRE Directive throughout the Project.

The OGC (Open Geospatial Consortium) is an international organization for the development and implementation of open standards for geospatial content and services, GIS data processing and data sharing.

The INSPIRE (INfrastructure for SPatial InfoRmation in the European Community) Directive (2007/2/EC 14 March 2007), that was transposed in the Italian D.Lgs. n. 32/2010, is aimed to create an "infrastructure for spatial information", to make geographic information more accessible and interoperable across Europe. To do this, every national authority responsible for production and updating of spatial datasets included in the list of the 34 themes of interest ("Human health and safety" is one of them), have to follow a set of common implementing rules that guarantee the uniformity, interoperability and accessibility of data.

In order to make the GIS useful and adequate to the actual needs, a set of requirements and constraints have to be shared among the teams. This is the rationale of the present edition of the Guideline, that is thought as a working tool for the GIS development and for the geographic data interoperability across the project teams. The Guideline includes some of the main concerns that arise when dealing with spatial datasets, and those needing a shared choice. It also includes some basic concepts that will allow a common language to be used when dealing with geospatial information.

As an Appendix, a Questionnaire, addressed to all partners of the Project is included. This latter is aimed at identifying data integration needs and potential uses of the system, to ensure that the designed database and functions will fulfill the requirements.

Since the geodatabase and the tools will be implemented progressively during the project, updated versions of the Guideline will be released, including new items depending on the new issues and needs arising with time.

Due to current problems concerning the setup of a web GIS server, data will be temporarily exchanged via FTP. New or processed data from the GIS team will be put into sub-directories of a folder named "*GIS-data*" located in the EXPAH FTP site. Data rendering on maps will be distributed via *.kmz* files that can be opened by Google Earth users.

For any needs/questions during the project development, the reference contact is Maria Paola Bogliolo at INAIL (DIPIA ex ISPESL): e-mail: mariapaola.bogliolo@ispesl.it or m.bogliolo@inail.it; tel.: +39-06-97893020.

The GIS team will also support spatial data transformation/processing/reformatting to accomplish project needs, as stated in detail in the following chapters.

2. Horizontal spatial reference for maps and spatial datasets

What are we speaking of?

An absolute spatial reference system is the basis that allows heterogeneous geospatial datasets concerning the same portion of the Earth surface to be overlaid and analyzed together.

Geospatial data (vector and raster data sets) can be represented and distributed in either a geographic or cartographic reference system. In the first case, data are not projected on a flat surface and coordinates are angles expressed in degrees (latitude and longitude) (Fig. 1a); in the latter, data are projected on a flat surface and coordinates are distances (using a cartesian reference) in meters (X and Y) (fig. 1b). The reference system is defined by the Datum (dimension and orientation of the reference ellipsoid) and, for projected data, by the map projection as well.

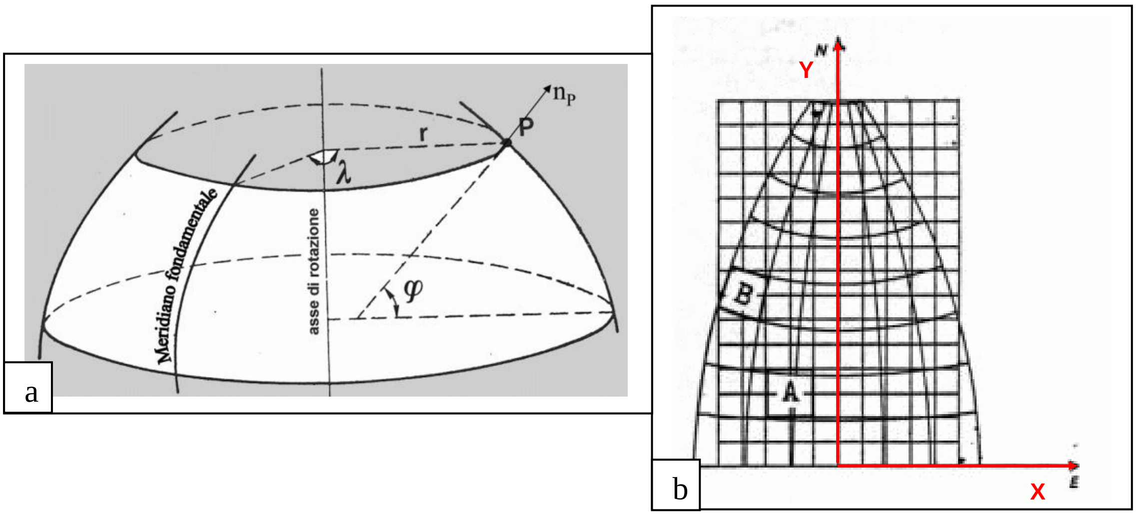

Figure 1. a) geographic reference system (λ = longitude, ϕ = latitude); b) planimetric (or cartographic) reference system

Three different Datums are usually encountered in spatial datasets concerning Italy (WGS84, European Datum - ED 1950, Roma 40) and 2 map projections (UTM and Gauss Boaga), with the WGS84 datum and the UTM projection by far the most (but not the only) used in recent data production. WGS84 is a worldwide valid Datum, and is the reference datum for GPSs.

Errors in associating the reference information (wrong datum) to a dataset can lead to positioning errors greater than 150 m (for the listed datum types). Even if this difference can be considered negligible in meteorological and dispersion modeling, it is not when dealing with ground data positioning (fig. 2).

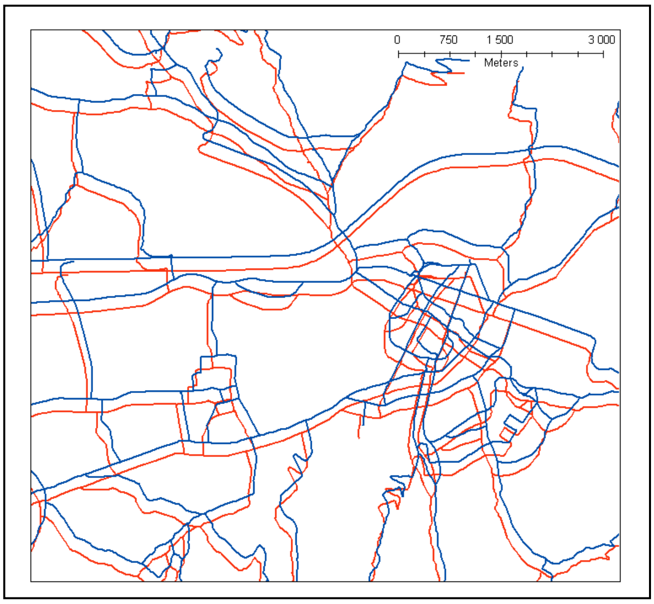

Figure 2. Effects of datum offset. The blue road network is correctly assigned with the WGS84 datum; the red network is the same as the blue one, but a wrong datum (ED50) was assigned to it. The projection is UTM33 for both.

The area of Rome falls on the limit between two zones of the UTM and of the Gauss Boaga projection (UTM32 and 33; Gauss Boaga East and West): consequently both projections are usually found in spatial datasets of the area of Rome (Fig. 3).

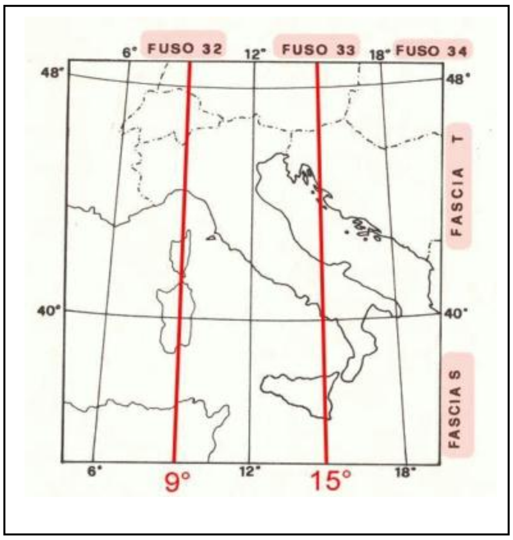

Figure 3. UTM zones for Italy. In red the central meridian of the zone.

Commonly used geospatial data formats can store the georeference information inside the file or in an ancillary file associated to the data. The information however has to be explicitly given in order to be saved in the file.

Guidelines for the project teams

Use of georeferenced coordinates is mandatory for spatial data used and produced in the project. Every type of reference could theoretically be used: many software with GIS functionality are able to transform the reference system of spatial data, also "on the fly". Nevertheless, a common spatial reference system for maps and spatial datasets would be preferable and should be adopted as the official reference in the Project. This will assure data consistency and will also allow a faster overlay of geographic information from different sources, and easier data exchanging.

We suggest the cartographic reference based on the UTM projection in zone 32 (UTM32) with the WGS84 Datum. All teams are invited to point out to the GIS team possible issues coming from the adoption of this reference, since it can be modified in order to achieve a general agreement.

Teams are obviously free to use different coordinate systems for their work, if necessary.

Whatever the coordinate system, teams are requested to specify (or include in the metadata accompanying the dataset) the reference system of the data when exchanging it with other partners or with the GIS team (see also Chapter "Metadata").

Support from the GIS team

In order to guarantee a high geographic quality of the data used in the Project and included in the GIS, we will provide support for all georeferencing issues: in detail:

- we will check the georeferencing quality of the datasets used in the project by analysing their coherence with high quality datasets, helping to solve possible issues;
- we will check the available datasets for the actual reference system, adding the relevant information in the file if lacking or overwriting it if wrong;
- we will be available for transforming all data sets to the official reference of the Project, and from it to other coordinate systems needed for analysis.

Every data will be re-projected to the official reference system before being included in the GIS.

Use of *km* coordinate units is widespread in simulation modeling, but meter is the standard unit of cartographic projections. Meters will be used in the GIS geo-database, but they will be converted in *km* when exporting datasets for simulation modeling. Similarly, *km* will be converted to *m* when importing simulation results in the GIS.

3. Vertical reference for maps and spatial datasets

What are we speaking of?

Similarly to coordinates, also elevations are defined with respect to a reference system. Three types of reference for elevations will be met during the project:

1) Absolute topographic elevation: it is the elevation that can be read on a topographic map. It represents the vertical distance from the geoid reference surface, that is assumed to be coincident with the local sea mean surface level.
2) Absolute geodetic (ellipsoidic) elevation. It represents the vertical distance from a reference ellipsoid (WGS84) surface. It is the elevation provided by GPS receivers and that retrievable from a DEM (Digital elevation Model) obtained by satellite acquisitions (e.g. SRTM, ASTER GeoDEM datasets). Differences with respect to the topographic elevation are about 40-50 m (being the geodetic elevation greater) in Italy (Fig. 4).
3) Relative altitude from the surface: meteorological and dispersion simulation models often use a 3D terrain following grid, where the Z coordinate represents the distance from the Earth surface. The same units are used for vertical sounding and LIDAR profiles.

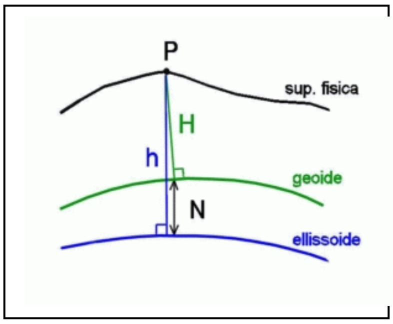

Figure 4. Geodetic (h) vs. topographic (H) elevation.

Guidelines for the project teams

A unique reference system will be used for elevations in the GIS: our choice is to use geodetic elevations referred to the WGS84 ellipsoid. This allows to transfer directly GPS measurements in the database, and to be coherent with the DEM used to represent the land morphology. Teams are requested to inform the GIS team about the type of altitude data provided for the GIS.

Support from the GIS team

Based on the mentioned choice, all elevation data will be converted to geodetic elevation before including them in the geo-database. In detail: elevations deriving from topographic map readings will be transformed into geodetic elevations using the free software CARTLAB1 (produced for the Italian Society of photogrammetry and topography - SIFET). Altitudes in the atmosphere (distances from the surface) will be converted to absolute values by adding the geodetic elevation of the surface; this latter will be retrieved from the DEM in the case of gridded data, and from the geodetic elevation calculated for the measurement station in the case of vertical profiles.

4. Positioning sites of interest

What are we speaking of?

Absolute location of measurement sites and of other sites of interest is needed.

Three methods for performing this task can be followed.

1) GPS measurements. non-professional GPSs provide measurements with an accuracy of 5-15 m. The native datum is WGS84, even if some receivers have the possibility to set a different datum (corresponding transformations are performed internally). Both geographic (latitude and longitude) or cartographic coordinates (metric X and Y) can be acquired, setting the receiver accordingly. GPS receivers also provide elevation, that is referred to the WGS84 ellipsoid (geodetic elevations).

2) Sites of interest can be located on georeferenced images available on the web. One possibility is Google Earth, where geographic coordinates (Lat and Long) can be read in the bottom part of the screen while keeping the cursor on the point of interest. Google Earth images are georeferenced in the WGS84 datum. It is worth to point out that the georeferencing accuracy of Google images can be poor (errors can be above 100 m in some areas) and variable with location. The local georeferencing accuracy can be roughly evaluated by showing the streets map on the images, since its quality is much better. A more accurate geolocation can be obtained on the images available on the "National cartographic portal" of the Ministry for Environment at: http://www.pcn.minambiente.it/PCN/.

3) For sites located along roads and provided with a street address, geocoding is another possibility, for people provided with the necessary GBF (Geography Base File - containing georeferenced street networks with street names, address ranges, intersections etc. as attributes) and software. As an alternative, Google Maps can help in locating the address but coordinates have then to be read on the corresponding point on Google Earth.

Guidelines for the project teams

If providing coordinates of points of interest, the way they were obtain should be specified. Anyone of the three methods described can be used, but GPS measures are preferred.

The following directions should be observed when taking GPS measurements:

a) Record the complete measurement (lat, long, elevation). b) Assure the WGS84 datum is set (usually it is the default). c) Geographic coordinates (lat and long) are preferable (because native): if metric coordinates are taken instead, set the right projection (UTM zone 32). d) Record a number of digits consistent with the instrument accuracy (usually 5-15 m): considering geographic coordinates, if decimal degrees are used, 4 decimal digits have to be provided, because at our latitudes a difference of 0.0001° corresponds to about 11 m in latitude and to about 8 m in longitude; if degrees-minutes-seconds are used, 1 decimal digit has to be included for seconds, since 1″ corresponds to about 31 m in latitudes and about to about 23 m in longitudes. If cartographic (metric) units are recorded, coordinates have to be approximated to the closest integer.

If the coordinates are image-based (method 2), use of the National Cartographic Portal of the Ministry for Environment - PCN (http://www.pcn.minambiente.it/viewer/) is preferable. Method 3 is suggested only if a GBF is available.

Support from the GIS team

The GIS team is available for georeferencing sites of interest for the Project. One of the first two described methods will be used, since the team is not provided with a GBF. Coordinates provided by the other teams will be visually checked by overlying the corresponding points on accurate base maps. Every change will be previously discussed with the involved team.

5. Metadata

What are we speaking of?

Metadata are sets of information describing data and services (mainly web services). Metadata are used as a documentation tool but also to query, find and retrieve data in repositories. An XML schema (ISO 19139) is used to make these documents readily interpretable both by men and machines, making them exchangeable via web.

Metadata are matter of special implementing rules of the INSPIRE Directive (INSPIRE Metadata Implementing Rules: v. 1.2), based on the ISO19115 standard. Creation and updating of geospatial datasets metadata is mandatory for authorities engaged in executing the INSPIRE Directive. Nevertheless, some datasets available from public authorities in Italy are still lacking metadata, because created before the Directive and not yet updated, or because the authority is not assigned with the official task of distributing cartography of interest of the INSPIRE Directive.

In the Project, metadata will be used to document geospatial datasets. An effort will be made to accomplish the INSPIRE rules even if a full compliance is not a goal of the Action. Availability of metadata in the GIS, even in a small subset, reduces the risk of mistakes and helps to make the database efficient and consistent. To make some examples, metadata will allow to record changes made onto the data, to go back to the data source whenever necessary, to compare versions of the same dataset owned by different teams, to assess data reliability.

Guidelines for the project teams

A small set of metadata should accompany every spatial dataset exchanged in the Project, both those produced on purpose and those coming from external sources. We consider mandatory at least the items in bold case of the following list (derived by the INSPIRE implementation rules):

Resource title	Name by which the cited resource is known
Resource abstract	Short summary of the content of the resource
Resource type	Dataset / Series of datasets
Resource locator	URL address for on-line access to the resource (if it exists)
Unique resource identifier	Unique code of the resource (e.g. file name)
Resource language	If text is present
Topic category	Dataset theme using the controlled vocabulary, MD_ScopeCode from ISO 19115
Keyword(s)	see http://www.eionet.europa.eu/gemet/inspire_themes + free text
Geographic bounding box	Coordinates of the outermost points
Spatial representation type	Vector (point, line, polygon) Raster, Grid, etc.
Reference system	Type of coordinates, Datum and projection (if applicable)
Temporal extent	Time period covered by the content of the dataset / series of datasets
Date of publication / last revision / creation	At least one of the 3 cases
Lineage	Summary of the processing history and any other info useful to use and evaluate the quality of the data.
Spatial resolution	Level of detail expressed as equivalent scale (vector data) or ground sampling distance (gridded data)
Limitation on public access	Any special restrictions or limitations on obtaining/accessing the data
Conditions applying to access and use	Free description of terms and conditions and/or link to the URL where they are described
Responsible party	Identification and way to communicate with person(s) and organization associated to the data
Responsible party role	Function performed by the responsible party (e.g. custodian)
Metadata on metadata	Descriptive features and references for metadata

An effort should be made to recover the above listed information also for maps/datasets obtained from external sources.

Further relevant information will be requested by the GIS team for the themes that will be part of the public version of the GIS.

Metadata can be supplied in any digital form.

Support from the GIS team

The GIS team will translate the metadata provided by the teams into standards; will complete the items retrievable from data themselves and will insert the metadata in the geodatabase. These activities could require consultation of the teams responsible for the data.

6. About consistency and updating level of base maps and datasets

What are we speaking of?

Often, datasets of general use (base maps, images and data) are free and can be obtained via web or by informal data exchange (e.g. Administrative limits, coastline, land use). This can result in the lost of metadata information, about source and date of production. As a consequence, often old data are routinely used while updated versions are available. An effort will be made in the framework of the project, to promote the use of uniform, updated and co-registered versions of datasets of general use.

Guidelines for the project teams

One of the aims of the questionnaire included in this guide is to obtain a survey of base maps and other datasets of external production owned by the project teams. If different versions of the same datasets will be found, the teams will be invited to use a unique, "official" version that the GIS team will distribute. Teams are also suggested to point out possible problems/discrepancies among datasets.

Support from the GIS team

The GIS team will carry out an inter-comparison of the possible multiple versions of general-use datasets and base maps owned by the teams. After the analysis and checking of the availability of new versions, these data will be co-registered, cut on the domain(s) of interest and distributed to the teams in a directory called "*GIS-data"* of the FTP site.

7. Data models and data formats

What are we speaking of?

With the term "*data-model*" we indicate the model used to represent in the GIS objects and fields (spatially continuous information) on (or near) the Earth surface and the information associated to them. Three models are used in geographic information science: vector, raster and 3D models.

A set of descriptive attributes is usually associated with objects represented with the vector (points, lines, polygons) model: these can include alphanumeric data, images, pictures, documents, or link to web pages. The whole information is organized in a relational database managed by systems able to store geometric information. Objects can be located in a 2D or 3D space: in the latter case Z dimension (elevation) has to be saved in the geometry.

The Raster model is usually used to store images.

3D models can be further separated in the so-called 2.5D models and the truly 3D models.

2.5 models are used to represent 3D views of surfaces i.e. of a variable (which is assigned with the third dimension) varying continuously on a planimetric domain. The typical example of the variable represented with this model is elevation (DEM).

In 3D models each entity is given 3 spatial coordinates (X Y Z) and one attribute that is the value of the varying field.

In our Project also 4D models have to be accounted for, where a time coordinate is given to each entity along with the three spatial coordinates.

Functions for visualization and analysis of vector, raster and 2.5D models are well developed in GIS. Truly 3D models are not handled by ESRI ARCGIS software while they are handled by open source GIS like GRASS. In ARCGIS 3D information can only be handled and visualized as multiple 2.5 representations (with Z dimension constant). Time dimension is not yet fully supported in GIS: it is handled in ARCGIS, not in GRASS.

Guidelines for the project teams

All formats having georeferencing internal information can be used for data exchange.

Examples of common formats for vector data are: Shapefile (.shp), CAD (.dxf, .dwg), Mapinfo Interchange Format (MIF/MID), Geography Markup Language (.gml). The alphanumeric information can come separately in a DBMS (ACCESS, PostrgreSQL), or in an EXCEL spreadsheet. A common identifier field is required in both the vector file and tabular attributes to link them each other. The third dimension (height) can be accounted for if Z is one of the attributes of the objects.

Examples of Raster formats are: Tiff/Geotiff, ERDAS Image .img, HDF (Hierarchical Data File). 2.5D formats include: TIN (Triangulated Irregular Network) datasets (.tin), ESRI GRID (binary) and ESRI ASCII GRID files. Every raster format can also be represented as a 2.5D surface.

3D raster models can be created in GRASS starting from multiple raster layers (each associated with a Z value) or from vector point data storing the Z dimension (elevation).
Other formats should be accompanied by georeference ancillary information (coordinates of the bounding box, spatial resolution, reference system). Generic grid data formats (ASCII, binary, etc.) should come with information about the internal logical organization of data (e.g. if the first value is the Upper Left, Upper Right, Lower Right or Lower Left of the domain).
Time (and multidimensional data sets) can be accounted for via the NetCDF format or using multiple vector or raster layers with a timestamp attached as an attribute.

Support from the GIS team

The GIS team is collecting spatial information to be included in the GIS. Anyone of the listed models and formats can be used for data exchanging. The FTP site of the EXPAH project is at the moment the only way for an efficient data exchange.
The GIS team is available for any issue regarding conversion of data formats, and to support analysis and visualization of data in their geographic context.

8. Spatial Domains

What are we speaking of?

One of the parameters that have to be defined earlier when planning a new GIS application is the spatial domain of interest. Spatial data sets will be clipped on that domain (domains), or to a larger area depending on the need of preserving boundaries integrity. Accounting for the type of information managed in the EXPAH project, a 4D domain should be defined, including the Z (altitude) and the time dimensions.

Guidelines for the project teams

Teams are invited to interact with the GIS team to define one or many working spatial domains. Likely, a greater spatial domain with low resolution and a smaller domain with high resolution are needed. At the time we are writing, three computational spatial domains have been defined for the meteorological and dispersion models (document *Long range transport of PAHs and definition of computational domains for Rome city simulations*:http://www.ispesl.it/expah/documenti/Expah_4.2_computational_domain_&_minni2005.pdf), having a horizontal resolution of 12, 4 and 1 km respectively. The largest one seems to serve only for modeling tasks: in this case we will evaluate not to include it in the GIS, since it will not be matter of further spatial analysis.

Support from the GIS team

The GIS team is available to clip every larger dataset to the domain(s) of interest.

9. Co-registration issues

What are we speaking of?

When dealing with heterogeneous layers of information sharing the same spatial domain, the quality of any result emerging from a combination of them (obtained by mathematical, logical or geometric operations) depends on the reliability of the spatial relationships assumed to be present among them. If all the datasets are georeferenced with the same accuracy, the spatial relationships among objects or phenomena represented is supposed to be preserved. However, the different models used to represent information can lead to errors that prevent the necessary quality to be assured. While these errors are readily highlighted in a GIS, through the available tools, they are not easily identifiable in simulation models or other software with limited GIS functionality.

This problem is often found when combining vector and raster features, so we will concentrate on this issue in the present version of the Guide.

Guidelines for the project teams

The following problem has been faced when combining gridded and raster datasets (e.g. the 3D simulation grid and the LULC (Land Use-Land Cover) map used in dispersion models).

Usually, the nodes of a grid are dimensionless points (Fig. 5) represented as vectors which attributes refer to a volume that is centered on the node and has a size equal to the grid step along each axis. The extreme coordinates of the grid refer to coordinates of the outermost nodes (ULg and LRg in Fig. 5).

In raster maps, the "picture elements" (cells or pixels) have an extension proportional (by the representation scale) to the extension of a land parcel defined by the dataset resolution (e.g. 250 m). The attribute (e.g. LULC type) of the pixel is assigned to that extension. The extreme coordinates of the raster refer to the extreme corners of the outermost pixels (ULr and LRr in Fig. 5).

In order to have a co-registered (spatially coincident) dataset, each node must fall on the centre of each pixel: as a consequence, the raster map has to be cut to a domain that is apparently larger than the grid horizontal domain: assuming the datasets to have the same resolution (e.g. 250 m) the raster domain is defined by the following coordinates:

UL corner: Xr = Xg-L/2 and Yr = Yg + L/2

LR corner: Xr= Xg+L/2 and Yr = Yg-L/2

where L is the common grid and raster data resolution; Xr and Yr are the raster coordinates; Xg and Yg are the grid coordinates.

On the contrary, imposing the same bounding box (domain) to the two datasets, an offset (of L/2) is produced on each direction.

Support from the GIS team

We will resample the raster maps needed for analysis to make the pixels centers having the same coordinates of the grid nodes.

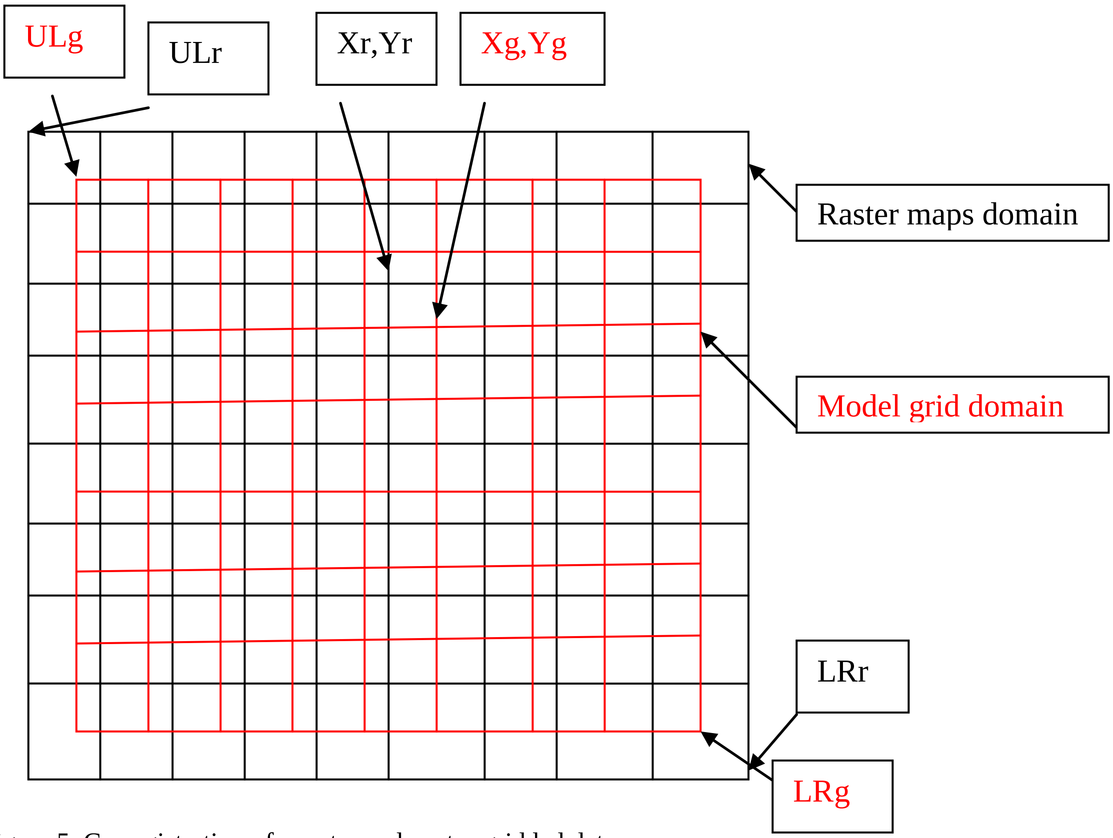

Figure 5. Co-registration of a raster and vector gridded data

10. Datasets in the *GIS-data* directory of the FTP site

The datasets included in the list are available partly under the *GIS-data* directory of the EXPAH FTP site, partly under request. If not specified, spatial data are in the WGS84-UTM32 cartographic projection.

- Polygons of census tracts, linkable with 2001 census data (source: ISTAT 2001);
- Polygons of inhabited settlements, linkable with 2001 census data (source: ISTAT 2001);
- Polygons of administrative districts (municipalities, provinces, regions, municipalities of the Rome area), linkable with 2001 census data (source: ISTAT 2001).

NOTES: for the area of interest the administrative borders are coincident with the 2011 layout.

ARIANET data in the *Borders* folder of the EXPAH FTP site have the following problems:

a) Regions: borders appear to be coincident with data in the Cartographic National Geoportal (PCN: http://www.pcn.minambiente.it/GN/), however, after overlay on the 2006 orthophotos (PCN), ISTAT data show greater accuracy and precision, this make the latter to be preferable.
b) Provinces: same observations as Regions; moreover, borders are not topologically coincident with the Regions dataset.
c) Municipalities: a WGS84 datum is declared, but data appear to be referred to ED50. The dataset is old, as inferred by the absence of the Fiumicino municipality. After datum correction borders are coincident with provinces (and consequently not coincident with Regions).

- Table of alphanumeric census data (source: ISTAT 2001) linkable with census and administrative polygons. Information includes:
  - Population: Resident population: total and sorted by gender, age (16 classes), level of education (only for > 6 years); Workforce: total and sorted by employed and un-employed; Number of resident people moving daily: inside/outside the municipality.
  - Buildings: Number of dwellings: total, and sorted by use (occupied by resident people, by non-resident people, empty); Number of other kinds of dwellings; Number of occupied-by-residents dwellings provided with heating; Number of dwellings with centralized heating; Total area of occupied-by residents dwellings; Number of buildings: total, used and sorted by use; Number of buildings with 1, 2, 3, 4 or more floors.
  - Industry: Number of firms by: type of activity (ATECO 2-digits classification), number of employees (4 classes), type (firm, public entity, no-profit); total number of employees.
- Polygons enclosing the area inside the "Grande Raccordo Anulare" (GRA), the area inside the "Anello ferroviario", the area inside the ZTL; all are linkable with 2001 census data (source: manually adapted from unknown provenience data (to be assessed)).
- Points showing the position of the Meteo Stations of the project (WGS84 UTM33), associated with description data (instruments are listed in a separate table linked through a 1-to-many relationship to the point dataset).

  NOTES: Coordinates for the Tor Vergata station in the original description sheets (.doc) are wrong (the Montelibretti coordinates are reported instead).
- points showing the position of the air quality monitoring stations of the Regione Lazio Authority, as retrieved from the BRACE database (SinaNet: http://www.brace.sinanet.apat.it/web/struttura.html). Main characteristics and the list of sensors are used as attributes. Only active stations (as resulting from the DB) are included.

  NOTES: A comparison with the list included in the document xxx highlighted the following differences: the L.go Preneste station included in the document is reported as dismissed in the BRACE DB; two stations (L.go Montezemolo and L.go Perestrello) are in the BRACE DB but not included in the document. Coordinates retrieved from the BRACE DB are approximated to Seconds: consequently an error of ± 15 m is expected, as confirmed by visual comparison with Google Earth.

- CORINE 2006 LULC map in raster format; resolution 250 m and 100 m, clipped on Italy. Pixels have "integer" coordinates of the center (000, 250, 500 etc. steps for the 250 m resolution; 100, 200, 300 etc. steps for the 100 m resolution), to help an accurate co-registration with a simulation grid having the same coordinate steps (steps can be changed on request).
- CORINE 2006 LULC map in vector format clipped on Italy.
- Spatial representation of the diffuse emission data INCOM2005 at the municipality level for Lazio (provided by ARIANET):
  - global values by pollutant from all the SNAP activities (29 pollutants);
  - values by pollutant and by SNAP macro-activity (2-digits SNAP classes).
- Spatial representation (WGS84, no projection) of point emission ISPRA2005 data for 23 pollutants (provided by ARIANET) and with legend of SNAP activities according to the SNAP97 codes (source: EMEP/CORINAIR 2007 Guidebook).
  NOTE: Data are lacking the source and a date of reference. The datum is unknown (assumed to be WGS84). Positioning precision is poor (about 1 km since coordinates have 2 decimal digits). Points appear to be systematically shifted of about 1 km toward West.
- DEM (Digital Elevation model) with 90 m pixel size: absolute horizontal accuracy is 20 m, absolute vertical accuracy is about 16 m (at the 90% confidence level) (source: SRTM 3-arcsec global DEM in the CGIAR-CSI version 4 - filled gaps)
- DEM with 30 m pixel size: horizontal accuracy is about 26 m at 90% confidence level; vertical accuracy is 17 meters at the 95% confidence level (source: ASTER GDEM 1-arcsec version 2);
- DEM with about 250 m pixel size (but also 500 m and 1 km): vertical accuracy for the global product is between 26 and 30 m (RMSE) at 250 m resolution (source: GMTED2010 database - updating of GTOPO30). Being the data derived from higher resolution data, mean, median, maxima, minima, standard deviation, systematic subsampling and breakline emphasis (stream and ridges are maintained) data are available.

Life+ 2009 EC Project

EXPAH - Population Exposure to PAH

Action 8.1
Integration of data and results by means of GIS techniques

QUESTIONNAIRE

Dear EXPAH partners,

In order to build up a Geographic Information System (GIS) able to address users requests and needs, we need to strictly interact with all partners dealing with geo-spatial data, to know:

1) which spatial data are used/needed/produced;

2) what are the needs in term of spatial analysis, data processing and combination of different types of spatial information.

Indeed, the content of the database of spatial data plays an important role in the overall design of the GIS and will depend on the sources identified, the user profiles and the needed applications.

As the first step of this process we need to know who holds data and information of relevance to GIS; what data and information are used, held or foreseen to be necessary during the project; what formats and software are used to handle spatial data and their attributes; what could be useful in terms of data analysis, processing, availability during the project.

This is the aim of the Questionnaire we are submitting. It also represents the starting point of a connection we will continue throughout the project, with the goal of making our work useful and adequate to the actual needs of the research.

Please be assured that any details you provide us will be held in the strictest confidence.

Thank you,

Maria Paola Bogliolo

Contact details:
INAIL
DIPIA – ex ISPESL
Via Urbana, 167
00184 Rome (Italy)
Tel. +39 06 97893020
Fax: +39 06 97893304
e-mail: mariapaola.bogliolo@ispesl.it
m.bogliolo@inail.it

How to fill in the questionnaire:

- Tick the boxes and write using a word processor or a pen on a printed copy.
- In case of a mistake, simply cross out and continue.
- Please feel free to leave out any questions you do not wish to answer.
- If you feel that a question is not applicable to you, please write: *NA*.
- The third part could require long time to be filled. However, it is the most important for us, so we ask to spend some time on it.
- Please send the compiled form by fax or by e-mail to one of the above addresses.

Section A – Your Organization and Role in the EXPAH Project

1. Organisation: ____

2. Department: ____

3. Reference Email Address: ____

4. Please rank your organisation's activities from 1 to 3 (1=most important 3=least important):

☐ Health ☐ Research ☐ Education ☐ Regulation ☐ Policy
☐ Administration ☐ Consultancy ☐ Trade ☐ Industry ☐ Service
☐ Information/monitoring ☐ Campaigning ☐ Technical support
☐ Other (please specify): ____

5. Which Actions of the EXPAH Project are you involved in? (Insert code(s))

__________ __________ __________ __________ __________ __________ __________

Section B – Information Management

6. Indicate whether you produce/use/need any of the following spatial information in the EXPAH Project

		Produces	Uses	Needs
Land use	Land Use/Land Cover - low detail (<1:50000)	☐	☐	☐
	Transport	☐	☐	☐
	Urban Land use/Plans - high detail (>1:25000)	☐	☐	☐
	Topographic maps	☐	☐	☐
	Remote sensing images	☐	☐	☐
	Other (specify):	☐	☐	☐
	______________________	☐	☐	☐
	______________________	☐	☐	☐
Atmosphere	Meteorological data (2D/3D + Time)	☐	☐	☐
	Pollution data (2D/3D + Time)	☐	☐	☐
	Point and Areal Emissions	☐	☐	☐
	Monitoring network data	☐	☐	☐
	Other (specify):	☐	☐	☐
	______________________	☐	☐	☐
	______________________	☐	☐	☐
Social/economic/political	Health, welfare, education (sensible sites)	☐	☐	☐
	Industrial activities	☐	☐	☐
	Land tenure and property (cadastral)	☐	☐	☐
	Demography and population	☐	☐	☐
	Policies, plans and laws	☐	☐	☐
	Administrative areas	☐	☐	☐
	Geocoded street data	☐	☐	☐
	Other (specify):	☐	☐	☐
	______________________	☐	☐	☐
	______________________	☐	☐	☐
Physical features	Hydrology	☐	☐	☐
	Soils	☐	☐	☐
	Topography (DTM, DEM)	☐	☐	☐
	Urban morphology (3D buildings)	☐	☐	☐
	Vegetation	☐	☐	☐

	Coastline	☐	☐	☐
	Other (specify):	☐	☐	☐
	______________________________	☐	☐	☐
	______________________________	☐	☐	☐
Health	Exposure maps	☐	☐	☐
	Expected cases	☐	☐	☐
	Real cases of disease	☐	☐	☐
	Other (please specify):	☐	☐	☐
	______________________________	☐	☐	☐
	______________________________	☐	☐	☐
Other (specify)	______________________________			
	______________________________			
	______________________________			

7. Do you use any kind of geo-spatial visualization/analysis in your activity for the EXPAH project?

☐ Yes

☐ No (go to question 10)

8. Which tool do you use to analyse your spatial data?

☐ A GIS software: (Specify): ______________________________

☐ A visualization tool embedded in your application software

9. What purpose are you using (or will use) GIS functions for, in the EXPAH Project?

Map production	☐
Modelling	☐
Georeferencing	☐
Geocoding	☐
Data Storage	☐
Planning/Management	☐
Decision Making	☐
Visualisation	☐
Data Analysis	☐
Monitoring	☐
Other (please specify):	☐
________________________	☐

10. What system do you use to archive non-spatial data?

☐ Spreadsheets (Go to question n. 12)

☐ Database Management Systems

11. Which database management system do you use?

IBM DB2	☐
Informix	☐
Microsoft SQL Server	☐
Oracle	☐
PostGres SQL	☐
Microsoft Access	☐
Sybase	☐
MySQL	☐
SQ Lite	☐
Other (please specify): ________________________	☐

12. Please tick the following facilities/approaches you use/will use for the Project:

1-n-variables multitemporal - 3D modelling (X,Y,Z,Time)	☐
1-n-variables 3D modelling (X,Y,Z)	☐

1-n-variables 2D modelling (X,Y)	☐
Multitemporal Geostatistical analysis	☐
3D Geostatistical analysis	☐
2D Geostatistical analysis	☐
Statistical analysis	☐
Monitoring stations	☐
Other (please specify): ______________________	☐
____________________________________	☐
____________________________________	☐

13. Please specify the types of data (both spatial and not) that you will use for the Project (tick all which apply):

☐ Digital photographs ☐ Digital video ☐ Word processor files
☐ Spreadsheet files ☐ Database files ☐ Plots/Graphs
☐ Digital data arrays (n-dimensional) ☐ Raster map data ☐ Vector map data
☐ Point map data ☐ Digital Aerial Photographs ☐ Satellite imagery
☐ Other (please specify): ____

Section C – WEB interface: Functional Requirements

14. If data services will be available on the Web for the Project, would you like to integrate your data with other data using online services?

☐ Yes ☐ No

15. If your data will be available by web access, which types of actions do you would consider as useful?

	For you	**For other Project teams**	**For external users**
Buffering (e.g. zoning)	☐	☐	☐
Overlay of different themes/maps	☐	☐	☐
3D views	☐	☐	☐
Interpolation (e.g. contouring)	☐	☐	☐
Multitemporal animation	☐	☐	☐
Request info about selected layers and/or layers' objects	☐	☐	☐
Geographic queries from attributes	☐	☐	☐
Geolocation (read coordinates of points)	☐	☐	☐
Map navigation (zooming, panning, go to predefined areas)	☐	☐	☐
Build a map legend	☐	☐	☐
Add annotations on the map			
Text	☐	☐	☐
Geometry	☐	☐	☐
Edit (modify/add) spatial features	☐	☐	☐
Edit (modify/add) attributes	☐	☐	☐
Build plots from attributes	☐	☐	☐
Show features on Google Earth	☐	☐	☐
Other (please specify):			
____________________________________	☐	☐	☐
____________________________________	☐	☐	☐
____________________________________	☐	☐	☐
____________________________________	☐	☐	☐

Section D – Your Data in Detail

Use the following table to describe your data in detail. Please copy this form if required

Spatial data

Topic	Format		Georeferenced?	Geographic coverage	Scale or spatial resolution	Use	Way of acquisition	Date of production	Multitemporal	Limitation to use	Metadata
	Vector	**Raster**									
e.g. Land use	1 Shapefile 2 ESRI geoDB 3 MapInfo 4 DXF 5. GML 6. NetCDF 7 Other (specify)	1 IMG 2 TIFF 3 GeoTiff 4 JPEG 5 HDF archive 6 MrSID 7 ECW 8 Other (specify)	1 Yes 2 No	1 International 2. National 3 County 4 District 5 Municipality 6 Local	e.g. 1:10000 or 10 metre	1 Reference 2 Analysis	1 Purchased 2 Free (cost recovery) 3 Informal exchange 4 Produced by you For external data specify the source	Year 000 if unknown	1 Yes (specify rate) 2 No	1 Yes 2 No	1 present 2 absent

Non-spatial data associated to geographic locations

Topic	Format				Can be spatially referenced to:	Way of acquisition	Date of production	Multitemporal	Limitation to use	Metadata
	Database	Spreadsheet	Text	Other digital						
e.g. Pollutant concentration, Wind velocity	1 Access 2 Oracle 3 SQL Server 4 Other (specify)	1 Excel 2 Lotus 3 Other (specify)	1 Word 2 Works 3 Word Perfect 4 PDF 5 Other	1 MPEG 2 AVI 3 TIF 4 BMP 5 Other (specify)	1 Points 2 Lines 3 Polygons 4 Raster cells	1 Purchased 2 Free (cost recovery) 3 Informal exchange 4 Produced by you For external data specify the source	Year 000 if unknown	1 Yes (specify rate) 2 No	1 Yes 2 No	1 present 2 absent