

Open source solutions for access to information in the public domain

Sarah Braccio¹, Fulvio Rinaudo²

¹DITAG, Politecnico di Torino, Torino Italy, sarah.braccio@polito.it

²DITAG, Politecnico di Torino, Torino Italy, fulvio.rinaudo@polito.it

Abstract

Geographic Information Systems (GIS) have always been considered sophisticated tools, designed by researchers and used by experts. Today, thanks to open source software, GIS systems are becoming a low cost instrument for different kinds of users who could take advantage of the simplicity and the effectiveness of GIS themselves. In order to reach this goal, it is necessary to fill the knowledge gap that still exists between developers and ordinary users.

The aim of this study is to identify a simplified design method that could be managed by unskilled users. The research develops the integration of a commercial software commonly used for the database with an open source GIS in order to realise a manual that allows to introduce a novice user to the design of geographic databases and to the issue of a well structured database.

The results achieved confirm the possibility for open source GIS to be integrated with commercial software with satisfactory results. This integrated solution allows users who cannot afford the high cost of commercial software, or who are unable to manage the complexity of the software, to use the GIS tool.

This paper underlines the need to spread open source technologies as tools for querying the archives that must be in the public domain such as one of a cultural nature or, in specific contexts, where low cost and ease of use are needed in developing countries.

1. Introduction

The expression public domain means, in general terms, that set of information to which no public or private entity has the right of exclusivity and thus it can be used by anyone.

The debate on the free dissemination of information has been very heated in the last few years. There have been numerous initiatives funded by the European Commission in order to promote the sharing of information, such as Europeana (<http://www.europeana.eu/portal/>), a European virtual library that enables access to 6 million digital items. International organisations have also been heavily involved in promoting and developing the dissemination of software and information under open access conditions, safe in the knowledge that their free dissemination plays a key role in the field of education, science, cultural heritage and public sector data. This principle may be extended to the free access to information made available free of charge by the rights holders as in the case of open-source software licences. (Koïchiro Matsuura, 2004)

UNESCO's Communication and Information programme approved for the two-year period 2010-2011 (<http://unesdoc.unesco.org/images/0018/001870/187028e.pdf>) indicates as its second priority the strengthening of the capacity in order to ensure universal access to information and knowledge. The programme promotes strategies to contribute to equal and simplified access to scientific information and to increase the use of information technologies open-source type.

The principle underlying the international debate and applied strategies is that every individual must have the possibility of acquiring knowledge and know-how in order to be able to play an active role in the information society and the knowledge economy.

Easy access to software is thus regarded as a key element in a truly inclusive information society.

Unfortunately, the progress in computer science is largely outside of the reach of the user community because of a lack of useful tools to manage these complex systems in an accessible manner.

Easy access to software is thus regarded as a key element in a truly inclusive information society.

Unfortunately, the progress in computer science is largely outside of the reach of the user community because of a lack of useful tools to manage these complex systems in an accessible manner.

GIS Open Source systems have always been criticised for being powerful and versatile tools but that are difficult to use.

However, unskilled users still find it hard to understand the potential of a GIS system because they lack the basic knowledge of Geographical Information Systems which are usually identified exclusively as tools for the production of theme maps.

Sector experts know full well that there is more to it, the problem being how to disseminate the GIS information organisation method in order to make it available to a wider and more varied audience.

Having said that, open source GIS systems are in the public domain and can easily be adapted to specific requirements, which makes them ideal in order to achieve this objective.

This research takes account of the strategies promoted by UNESCO and aims to familiarise unskilled users with the management of information in the public domain. To this end, it proposes a manual which applies a GIS Open Source system to a very simple case of information management.

2. Tools and methodology

The research stems from the choice of the software tools best suited to our requirements.

MS Access (<http://office.microsoft.com/en-gb/access-help/>) has been chosen as the database design software. This is a database storage and management program included in the Microsoft Office suite.

Access is equipped with a simple and extremely flexible interface with user-friendly tools for those who are new to managing their information within a database.

Undoubtedly, the software has some limitations in terms of its "multi-user" performance, especially when the volume of data to be managed and the number of simultaneous connections being written increase dramatically. However, for the purposes of our simple application, its potential is sufficient.

This choice may be criticised by open source "purists" as it is a mixed solution which also comprises

the use of a commercial piece of software.

The choice is due to the fact that in practice we are often having to operate on data banks already designed by MS Access both within the scope of cultural heritage and that of public administrations and, for the time being, it is the easiest database design tool for unskilled users.

The initial hurdle to overcome is that Access is not a geographical database and for this reason it does not allow to directly introduce the geometric component as an entity attribute.

Two desktop applications have been analysed for the choice of the GIS software: QGIS (<http://www.qgis.org/>) and gvSIG. (<http://www.gvsig.org/web/>) The benchmark against which the two Open Source softwares have been compared relates to ease of management, communication effectiveness and integration with Access.

QGIS interacts with a database designed in MS Access in 2 ways.

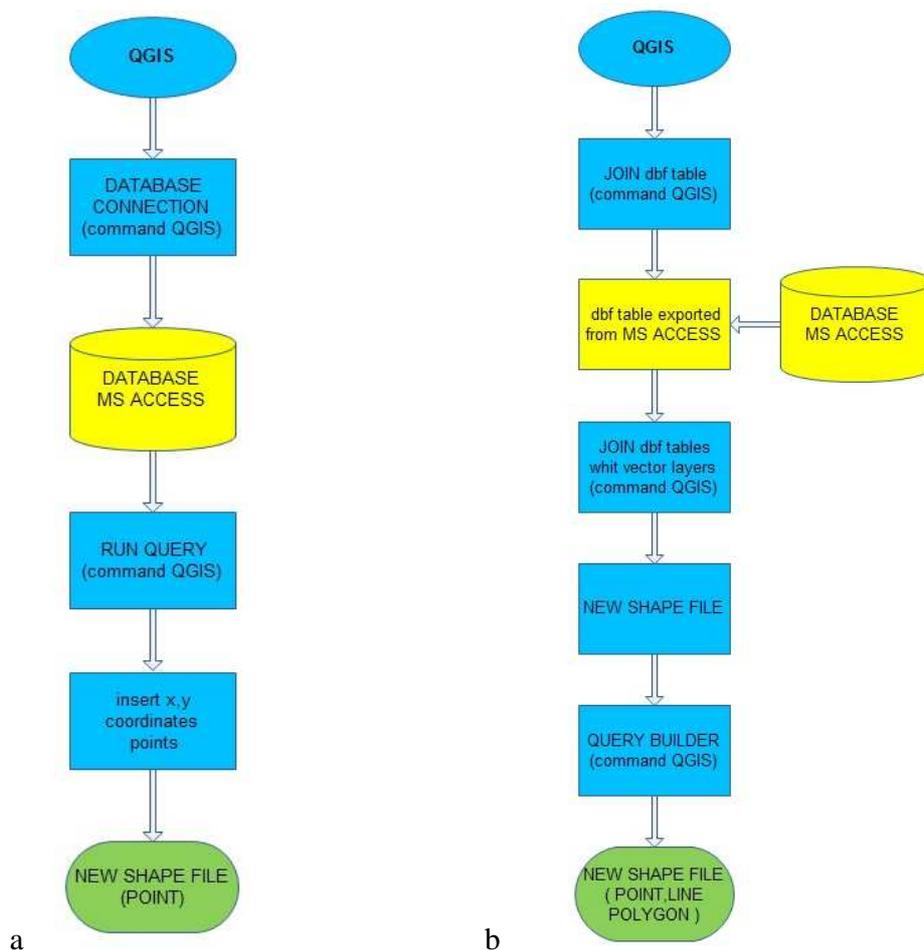


Figure 1. Connection QGIS-Access

1) The first solution envisages the use of the eVis plug-in, which, by means of the DATABASE CONNECTION command (Fig.1 a) enables to connect and view the entire database created in Access and perform queries in SQL language. This type of connection can be used when we are designing a

GIS whose geometric entities are points of x, y coordinates which can be entered as attributes in the corresponding database tables. If we are managing geometric entities other than the point, as in the case of polygons, we may use the centroid technique. The polygon will be linked to its centroid point and, by using the selection by location function, we will be able to give a geometric-spatial reply to our query.

2) The second method (Fig.1 b) envisages exporting from the Access database (mdb format) each single table (dbf format), which will then have to be individually imported in QGis by means of the Join function. The individual table will need to be linked to the corresponding geometric entities and between one another, following the design diagram built in MS Access. At this point we could perform the queries and get the reply shape file.

The gvSIG software connects to the database designed in MS Access by means of only one method (Fig.2).

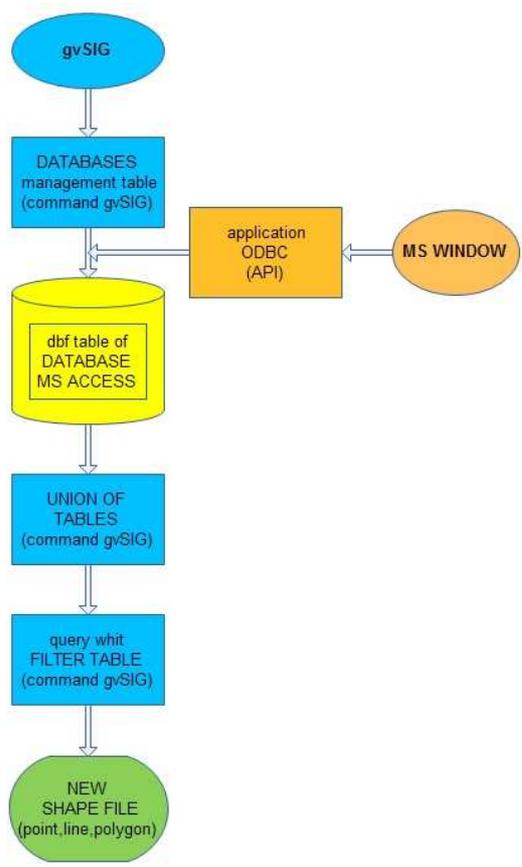


Figure 2. Connection gvSIG-Access

1) In order to connect, it uses the Open Database Connectivity (ODBC) which is a standard API (Application Programming Interface) for connection to Database Management Systems (DBMS). It is

used since it allows programs to connect to the different database types by using more or less the same code. By means of the DATABASES command located within the table management, gvSIG connects to a single table of the MS Access database. It will then be necessary to join each single table with the corresponding geometric entity and subsequently perform the query.

The QGIS software proved to be best suited to integration with Access within the scope of our application which envisages the creation of a GIS system where the main geographical base comprises punctual geometric entities.

The QGIS project is a multi-platform application which runs on all the main Mac OSX and MS Windows versions. It supports numerous database formats and provides a wide range of functions allowing analyses based both on raster and vectorial data. Its use is simple thanks to the modern and straightforward graphic interfaced based on Qt4 graphic libraries.

QGIS has a wide and growing number of GIS functions, provided both by the base program and by its extensions. The potential of this open source project lies in its plug-in architecture which avoids having to deal with a software with a hundred functions as is often the case in commercial GIS systems. Instead, starting from a minimum base, it enables to add or remove functions in relation to the user's specific requirements.

The study continues with the drafting of a manual where unskilled users are guided through a training course in which theoretical explanations are applied step-by-step to a real case relating to the management of the documentation pertaining to the Parks Canada Cultural Heritage (<http://www.pc.gc.ca/eng/culture.aspx>). Thus the guide addresses those who, lacking experience in Geographical Information Systems, have the need to use its potential to manage and query their own data archive.

More specifically, the proposed training allows users to:

- a) Build the databank with the all the information available using MS Access.
- b) Build the geographical base separately with Quantum GIS.
- c) Query and view the results by using the eVis extension of Quantum GIS and the Query tools of MS Access.

3. Application to a case study: Parks Canada National Historic Sites.

The choice of the case study introduced in the guide follows the strategies proposed by UNESCO, which highlights the key role played by archives and by the documentary heritage in the preservation of knowledge.

We are faced with the most classic of cases: the management of a data archive.

In Canada there are approximately 900 National Historic Sites (NHSC), 200 of which are under the jurisdiction of Parks Canada. Most of these do not have any level of documentation as to their belonging to the cultural heritage. The recent introduction of standards and guidelines for the preservation of historical sites in Canada has reaffirmed the value of cultural heritage registration and

documentation. Therefore, it has become necessary to create a documentation method with a view to increasing the efficiency of site registration.

The guidelines state that the registration process must ensure that the value of the heritage being documented is fully captured, thereby providing easy access to information and integration with new technologies.

The documentation of Parks Canada sites is currently being managed through hypertext. Navigation via the World Wide Web makes the information collected directly available to any type of users. Each link also provides the option to save selected information on one's PC through various formats, such as pdf, jpeg and dxf.

Following an overview of the various items mentioned in the table of contents, it is then possible to download the complete NHSC documentation through a single pdf file containing all the information collected.

The objective is to reorganise in a more efficient manner the information already present on a Parks Canada site taken as a sample by introducing the geographical component, so as to provide end users with simplified internal management of the documentation.

The documentation of the sites of cultural significance of the Canadian heritage is set out according to an outline divided into 5 sections, each one being accompanied by sub-chapters (Fig.3).



**CARILLON BARRACKS
NATIONAL HISTORIC SITE OF
CANADA**
50 RUE PRINCIPALE, CARILLON, QUEBEC

TABLE OF CONTENTS

1.0 SCOPE.....4
1.1 INTRODUCTION4
1.2 PURPOSE AND OBJECTIVES.....4
1.3 SITE DESCRIPTION5
1.4 DESCRIPTION OF WORK COMPLETED.....5
1.5 RESPONSIBLE HERITAGE RECORDERS.....6
1.6 TERMS OF REFERENCE FOR THE NATIONAL HISTORIC SITES OF CANADA
BASELINE RECORDING.....7

2.0 WRITTEN RECORD.....8
2.1 AVAILABLE STATEMENT OF COMMEMORATIVE INTENT, STATEMENT OF
SIGNIFICANCE, COMMEMORATIVE INTEGRITY STATEMENT.....8

3.0 PHOTOGRAPHIC RECORD.....10
3.1 LOCATION OF SITE.....11
3.2 AERIAL PHOTOS.....14
3.3 PHOTO KEY PLANS.....16
3.4 DIGITAL RECORD PHOTOGRAPHY.....24
3.5 MONO OR STEREOPHOTOGRAMMETRIC RECORD.....83
3.6 DIGITAL PHOTOGRAPHY FOR RECTIFICATION RECORD.....94
3.7 BLACK AND WHITE CONTACT SHEETS.....100

4.0 GRAPHIC RECORD.....103
4.1 SITE PLAN.....104
4.2 FLOOR PLANS.....105
4.3 DETAILS.....106

APPENDIX A: FIELD NOTES.....107
APPENDIX B: 1977 AS FOUND DRAWING SET.....117

HCD Project No. 489505
Data collected November 2006
Report compiled December 2006

HERITAGE CONSERVATION DIRECTORATE
Real Property Services
Architectural and Engineering Services
PHSCG

Figure 3. Parks Canada Documentation. Input data

The Written Report section concerns all the historical-cultural-architectural information of the building being registered. Through the historical description or historical name of the site it is possible to find further documents giving added value to the building itself.

The Photographic Report contains information ranging from site location, through images of

historical and current areas, to more detailed images of the building. They are in .tiff and .jpeg. format and are hand-made sketches which are then digitised. The digital photos included are divided according to their features (interior, exterior, main floor, etc.).

The Graphical Report section consists of the site plan, in .dxf format, and of technical drawings, or preparatory sketches, annexed to the appendix, or transformed into CAD drawings.

The Appendix section is dedicated to the notes and is an integral part of the documentation useful to gain an understanding of the measurement methodology (preparatory sketches).

The designed database follows the outline defined by Parks Canada. Three main tables have thus been created:

- Project Record;
- Photographic Record;
- Historic Record.

Each table is linked to a corresponding part of the table of contents of the general data collection document:

- Written Record;
- Photographic Record;
- Graphic Record/Appendix.

In the first table (Fig.4) Project Record all the alphanumeric characteristics of the project have been entered as fields.

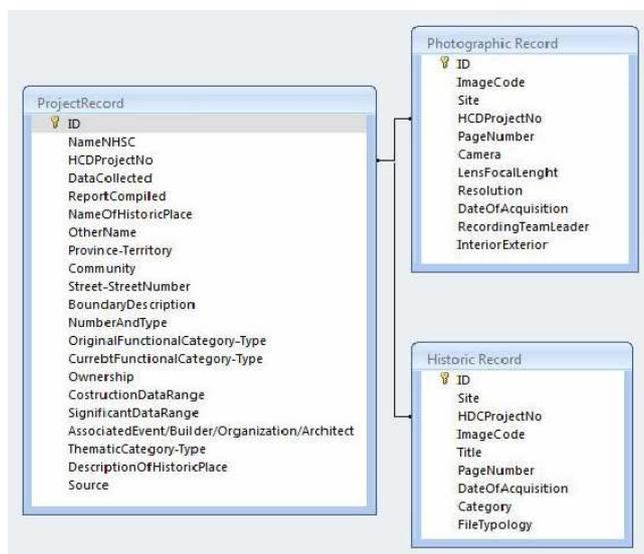


Figure 4. Parks Canada database. Project Record

This first table contains the so-called “HCD Project Number”, which is used as a joining and linking record among the various tables. “HCD Project Number” is also the key record for setting the queries on the newly created database.

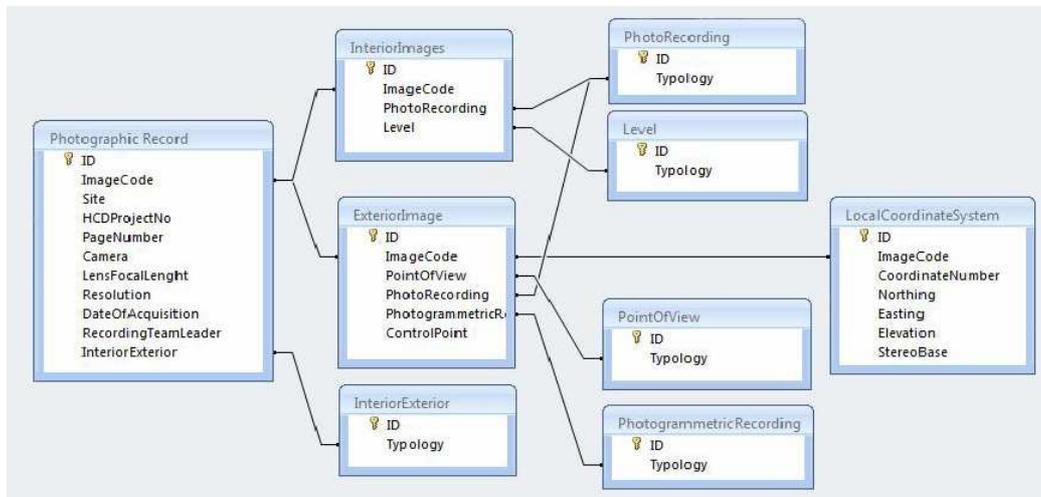


Figure 5. Parks Canada database. Photographic Record

The second table Photographic Record (Fig.5) is structured in a more complex manner.

It is linked to the first table through the definition of InteriorExterior image and to other two tables, Interior and Exterior Image. The relationship between these two tables is One to One. This table contains the image code, it is associated with the HCD Project Number and enables to set the queries.

The third table named Historic Record contains information on the territorial context and the subject.

Both raster files and vectorial files have been used for the geographical database.

The project is divided into two levels. The regional level (Fig.6) consists of documentation on the National Sites and National Parks. It is made up of shape files and a DEM image. The shape files are provided by the Parks Canada association and contain information on the location of the Canadian cultural heritage sites.

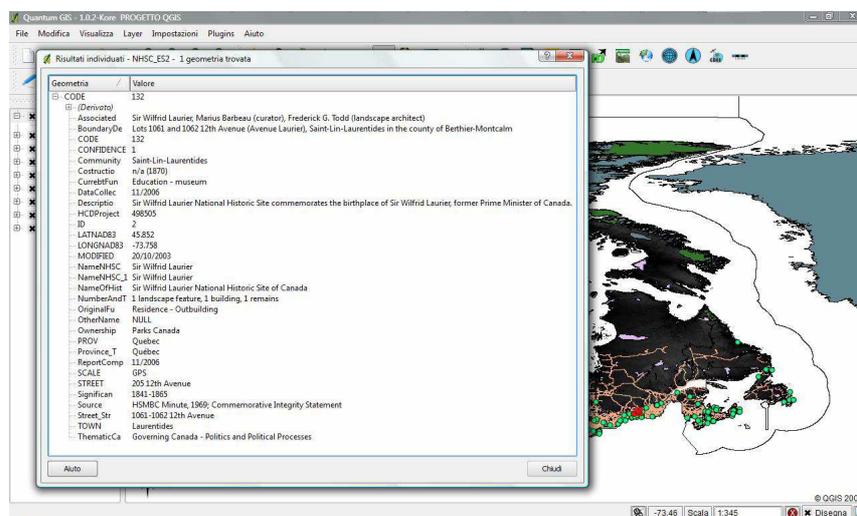


Figure 6. Parks Canada GIS. Regional level.

The local level (Fig7) is linked to the specific site. It is made up of the raster files of the site plans, which are linked to one another through georeferencing. A link to the various images has been included on each one of these by means of a punctual vectorial representation.

The point of contact between the two Systems is represented by the HCD Project Number which allows interaction between the two levels, creating continuous reading of the information collected.

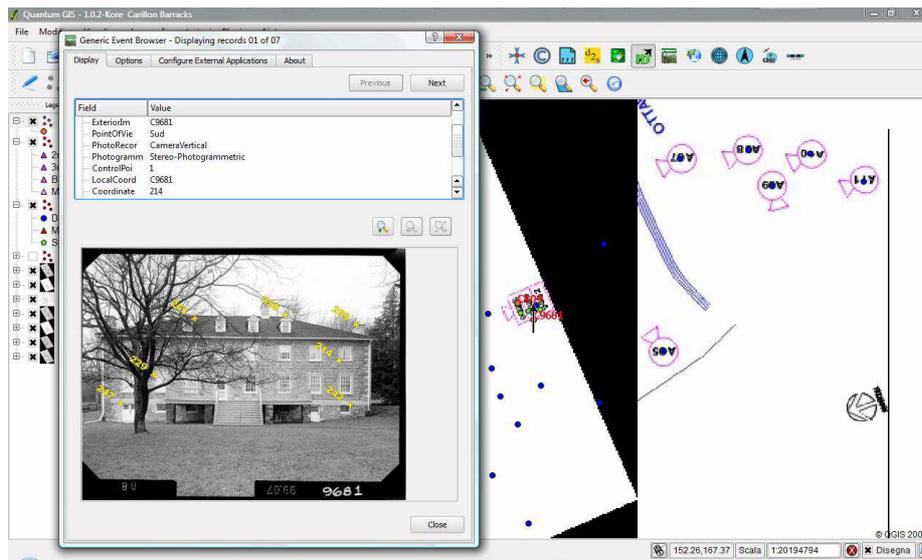


Figure 7. Parks Canada GIS. Local level.

4. Conclusions

The lack of spread of GIS and especially open source systems in certain application areas lies in the conviction that they are difficult tools to use. This study indicates a possible route for getting a wider audience more familiar with using Open Source GIS systems.

The results achieved confirm that open source GIS systems can be satisfactorily integrated with commercial software for simple applications.

For what concerns the effectiveness of the proposed training method, a verification of its application validity is still to be conducted.

It would be necessary to test the manual by proposing it to users from varying background and with different skills.

Among the objectives proposed by UNESCO there is the need to spread information and knowledge through Open Source software in developing countries.

The majority of these countries have so far been unable to take full advantage of ITC progress in terms of access to scientific information, technologies and learning opportunities, thus creating a "digital gap" (Koichiro Matsuura,2004).

The training programme could be tested by administrations also in developing countries where

economic resources are scarce but there is a great need for information.

The technical and scientific knowledge underlying the geographical information systems is complex but the task of those promoting free access to information as a fundamental human right should be to endeavour to bring a wide audience closer to an efficient tool.

The key concepts that should be underscored are:

- 1) information in the public domain should be easily accessible,
- 2) information becomes accessible if it is well organised
- 3) GIS Open Source systems are means of communication that enable the transmission of extremely complex information so that it can be understood by everyone.

If, as highlighted by UNESCO, information plays a key role in terms of development, it is up to the scientific community to provide the technical tools necessary to achieve this objective.

References

Koïchiro Matsuura, 2004, UNESCO's Approach to Open-Access and Public-Domain Information, in Julie M. Esanu and Paul F. Uhlir (eds.), *Open Access and the Public Domain in Digital Data and Information for Science: Proceedings of an International Symposium*, The National Academies Press, Washington, D.C, viewed 30 July 2010, <http://www.nap.edu/catalog.php?record_id=11030#toc>.

The 35th Session of the General Conference. Approved Programme and Budget 2010-2011, 2010, UNESCO document, viewed 30 July 2010, <<http://unesdoc.unesco.org/images/0018/001870/187028e.pdf>>.

Cavallini P., Piccardi S., 2007, Cosa significa Open Source? Cos'è il software libero? Applicazioni ai GIS. , *MondoGIS*, no.59, , 30 July 2010, <http://www.faunalia.com/pdf/Mondogis_2007_59.pdf>

Câmara G., Onsrud H., 2004, Open-Source Geographic Information Systems Software: Myths and Realities, in Julie M. Esanu and Paul F. Uhlir (eds.), *Open Access and the Public Domain in Digital Data and Information for Science: Proceedings of an International Symposium*, The National Academies Press, Washington, D.C, viewed 30 July 2010, <http://www.nap.edu/catalog.php?record_id=11030#toc>.

The Open Source Initiative (OSI) 1998, Public Benefit Corporation, California, viewed 30 July 2010 <<http://www.opensource.org>>

The Open Source Geospatial Foundation ,Delaware, USA, viewed 30 July 2010 <<http://www.osgeo.org/>>

Microsoft Office Support, Microsoft Corporation, 2010, viewed 30 July 2010, <<http://office.microsoft.com/en-gb/access-help/>>.

Gary E. Sherman, 2008, *Desktop GIS: Mapping the Planet with Open Source Tools*, The Pragmatic Bookshelf, Raleigh, North Carolina, Dallas, Texas.

Quantum GIS Project , viewed 30 July 2010, <<http://www.qgis.org/>>.

gvSIG Project, viewed 30 July 2010, <<http://www.gvsig.org/web/>>.

Dassau O., Sherman G., Sutton T., Hugentobler M. e Cavallini P., 2009, Quantum GIS: il desktop GIS più potente, amichevole e libero, GeoMedia,no. 5, pp. 36-39, viewed 30 July 2010, <http://www.faunalia.com/pdf/2009_12_geomedia_qgis.pdf>