The design of an automated workflow for metadata generation

Miguel Manso-Callejo[1], Mónica Wachowicz[2], Miguel Bernabé-Poveda[1]

[1]Dpto. Ingeniería Topográfica y Cartográfica. Universidad Politécnica de Madrid. Campus Sur UPM, Ctra de Valencia Km7.5 CP 28031 Madrid, Spain
{m.manso, ma.bernabe}@upm.es,
[2]Wageningen UR, Centre for Geo-information
P.O Box 47 6700 AA Wageningen, The Netherlands
monica.wachowicz@wur.nl

Abstract. The important role of digital resources relies on whether metadata is available and has been correctly catalogued and indexed so that the user can discover and use geospatial datasets. However, the cost and the error-proneness in the manual metadata creation, the lack of information provided by the producers of geospatial datasets and the lack of experience in cataloguing have motivated us to propose a new workflow for the automated metadata generation for geospatial datasets. This paper describes this workflow based on tasks synchronization that gives support for four metadata functions: discovery, use, evaluate and retrieval of digital geodata. The workflow was implemented using a multi-tier architecture system where the Data, Application and User Tiers can run a single use application as well as web services. The prototype evaluation is discussed in terms of the type of metadata being generated and the type of metadata function being supported by the workflow.

Keywords: metadata, automated generation, workflow, metadata function

1   Introduction

The concept of metadata is hardly new – it was first mentioned in the 1960's in the field of library management, and it was coined by Myers [1]. Metadata are "data about other data" [2][3][4][5], which provide the minimum information required to identify a digital resource. Ercegovac [6] and Milstead [7] state that the metadata describes the attributes of a given resource, whether a bibliographic object, an archive register or inventory, a geospatial object, a visual or museum resource or a software implementation. Moreover, Caplan [8] acknowledges that the concept of metadata is used to avoid the prejudices developed by professionals in the field of information, who are closer than most to the world of libraries: computer technicians, software designers, and system engineers. Finally, metadata are used to describe the context, the quality, the condition or the characteristics of data [1][7] in such a way that users can discover and understand their data sets. For Zeigler *et al.* [2], metadata is "*a hierarchical concept in which metadata are a descriptive abstraction above the data it describes*".

Various experts are in favour of assigning the task of metadata generation to the owners of the geospatial data sets, with the belief that these owners are best suited to provide information about their own data [9][10][11]. In practice, metadata generation has occupied a secondary role within organizations, having been generally created after the production or acquisition of data so far. Unfortunately, the presence of errors is one natural consequence when metadata generation does not occur simultaneously when the actual geospatial data set is compiled [10][11][12][13]. Moreover, the standards are complicated and extensive (e.g. ISO19115 defines more than 400 elements for metadata [14]), and the manual creation of metadata is a monotonous, harsh, and resource-consuming assignment. For this reason, some organizations have considered the generation of metadata as a costly additional burden [15].

However, this view has been criticized in several studies. In the CGIAR-CSI study [16], we find the following statement: *"The creation of metadata to novel data producers might seem burdensome, but the long term advantages are far superior to the disadvantages of the initial burden of implementing a Metadata policy within an organization. The initial expense of documenting data clearly outweighs the potential costs of duplicated or redundant data generation."* Liddy *et al.* [17] suggest that the techniques of automated creation of metadata can produce results of a reasonable quality level. Besides, [10][18][19] argue that metadata created by automated procedures tend to be more efficient, consistent and cheaper than those created manually by individuals. Downey [20] also suggests the development of tools for automated cataloguing such as a workflow, despite of its complexity. Finally, Greenberg *et al.* [21] and Craven [22] propose a combination of automated and manual methods in order to produce quality documentation.

This paper proposes a new workflow capable of maximizing the automation of the generation of metadata for geospatial data sets. Section 2 describes the existing methods developed for the manual and automated generation of metadata for geospatial data sets. Section 3 our proposed workflow for the automated generation of metadata is described by introducing its main tasks and their association with metadata functions. In section 4 we will identify those metadata elements that can be automatically generated by using this workflow and we will also classify each metadata element according to the function it fulfils. Section 5 describes the implementation and the main results, and finally, we conclude in Section 6.

2   Related Work

A vast literature can be found on describing the methods developed for the generation of metadata. Beard [11] proposed five methods for the generation of metadata: a) manually (with a keyboard); b) by extending the stored information with values obtained via consultations (i. e., the geographic identifier based on the geographic extension in a gazetteer); c) automated measurements and observations; d) extracted and calculated; and finally e) inferred from other elements. Moreover, Greenberg [9] identifies two methods of automated metadata generation: extraction and harvesting. In the extraction method, techniques of data mining and indexation are employed for the retrieval of elements or the contents tagging. In the harvesting method, techniques of recollection of existing labelled contents are employed. Later, Greenberg *et al.* [21]

carry out a revision of tools and applications developed for the automated generation of metadata about electronic resources, showing how the use of these automated methods can allow us to direct the effort of human resources towards aspects of a rather intellectual nature. Depending on the degree of automation and the human requisites of the process of creation of metadata, they distinguish between generators and editors in which automated and human processes are integrated.

In contrast, very few proposals of automated workflows are available in the literature. Guy *et al.* [23] propose a workflow based on a totally automated/semiautomatic/manual creation of metadata by the author of the document and of those created by an expert in information management (see Fig. 1). This workflow foresees the following tasks: a) automated, b) automated improved by the author of data, c) automated improved by an author and by an information specialist, d) created manually by an author and improved by an information specialist or e) created by an information specialist.
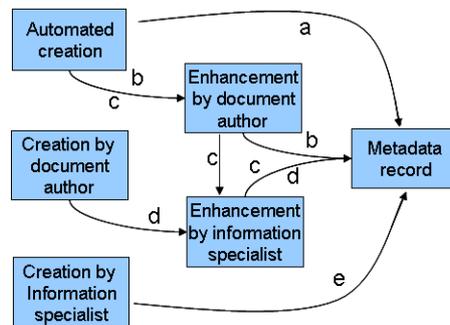

Fig. 1. Proposed tasks for the generation of metadata according to the human interaction in the creation of metadata (Adapted from Guy et al., 2004).
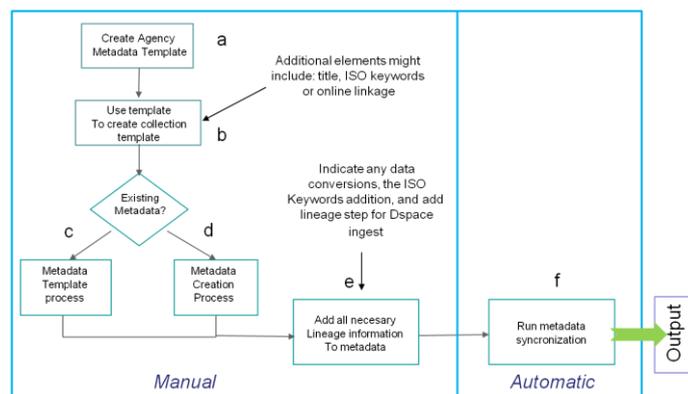

Fig. 2. Example of a workflow process for the generation of metadata (Adapted from Morris et al., 2007, p128).

Morris *et al.* [24] have proposed a *workflow* for the generation of metadata in which the starting point is the definition of a template for the agency or organization, and after

this template is personalized for a given collection of geospatial data sets. Once the template has been defined for a set of data, and if metadata exist, the data is processed in order to adapt it to the template, and it is created if data doesn't exist. Next, all the lineage information is added and finally, a process of synchronization of metadata is applied with a tool of commercial extraction of metadata as shown in Fig. 2. In addition, Hedorfer and Bianchin [25] suggested the use of templates as a device that made the creation of metadata an easier workflow process.

3   The main metadata functions and their respective tasks

Metadata elements have been grouped according to their functions. Some examples include *search, locate, discovery, exploration, evaluation, extract, access, retrieval, transfer, employ, use, exploitation, management, archiving and preservation* [11][26][27][28][29][30][31][32][33]. They have been further classified into discovery, evaluation, retrieval, use and management functions. The proposed workflow consists of a set of tasks which are grouped into four main functions aimed at the automated generation of metadata (Fig. 3). They are one of the following Functions:
-    Discovery: enable users to find/locate geospatial datasets.
-    Use: enables users to explore geospatial datasets.
-    Evaluate: enable users to explore whether a geospatial dataset suits their needs.
-    Retrieval: enable users to know how geospatial datasets can be obtained.

Besides, there is an additional function related to metadata management, in which a metadata template can be used to generate metadata in different standards.

The first task identified and related to three of the metadata functions (i.e., discovery, use, and retrieval functions) is *Metadata Extractor*. This task is designed following the metadata synchronization mechanism previously proposed [24]. The aim is to allow access to the geospatial datasets and obtain all information that may be part of the metadata for providing discovery, retrieval and use of these datasets. The challenge in implementing such a task is to generate the metadata to the wide number of geospatial formats such as raster, vector, grids or databases. Therefore, this task gets hold of all the information implicitly stored by the formats. Some examples include the representation type, format name and version, layers/bands number, characteristics of each one (rows * columns, number of bits per pixel, geometry type, number of elements for each geometry type), geographic extension (BBOX), file size and categories or element types stored in the layers or images. For a few formats, we can extract other items such as the creation date, author name, resources employed, process description and steps, and data access/use restrictions. Additionally, bands statistics can be extracted in one cases or computed in others for the images: max, min, average and typical deviation.

For the discovery metadata function, four tasks have been designed in the workflow. The task *Data Analyzer* uses the information obtained by the metadata extractor to infer the type of content stored in them; for example, a possible way to determine the content type of raster data may consist of evaluating the statistical values max, min, average and typical deviation together with the type of format, number of bands and the data type of every cell to discriminate between multi-spectral image, DEM, aerial image, thematic image and rasterized mapping among others. This task

requires the implementation of handling rules and parameters that are stored in a knowledge database that allows learning from the categories assigned to the data and registered in metadata when the metadata is provided as a template for its update. The metadata generated by this task will serve as important information for the classification of the resources using keywords, which in turn, will improve the discovery function.
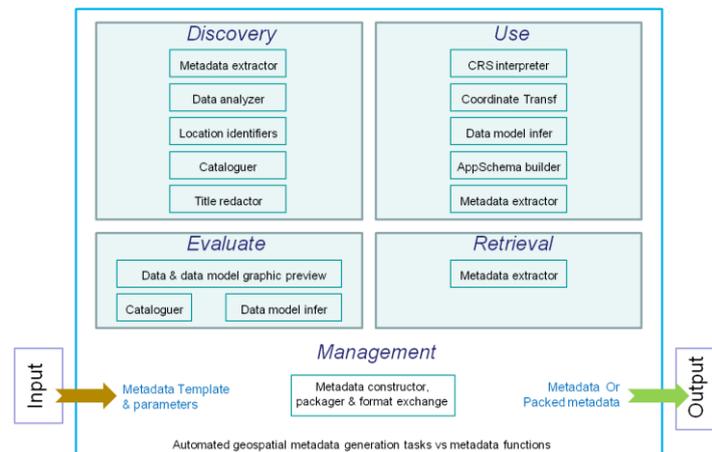

Fig. 3. Overview of the metadata functions and main related tasks of the automated workflow.

Another task related to the discovery function is concerned with obtaining the identifiers of a particular location in order to assist in the cataloguing process associated with keywords such as place theme keywords. The automation of the assignment of this type of keywords is carried out by making simple queries to an external 'inverse' gazetteer or by implementing them internally in the workflow by combining this task and the knowledge database.

The *Cataloguer* task consists of assigning keywords associated with the topic category and place theme based on information obtained by the tasks *Data Analyzer* and *Location Identifiers*. The goal is to provide the automatic cataloguing of resources within the workflow for the discovery metadata function task where a set of thematic keywords is assigned to metadata. These keywords belong to multilingual thesauri to facilitate the searches in the indexing systems and they are related to the subjects identified by the metadata standards (e.g., topicCategory). This task provides storing the keywords in the knowledge database of the workflow, their relationship with the thesaurus information and the themes they are related to. In addition to cataloguing the resource with thematic keywords, this task also provides cataloguing it spatially through location identifiers.

Finally, the *Title Redactor* task summarizes the content of geospatial data set from a sentence expressed in natural language following the grammar rules that are valid for one or several languages. This task is aimed at supporting a number of questions such as What?, When?, Where?, Who?, and What scale?. The result of this task is a concise title that describes the implicit metadata, geographic context, cataloguing and content types.

Another essential task is the *CRS* (Coordinate Reference System) *Interpreter* defined for the use metadata function. It is indispensable to understand and relate the different encodings used by the format owners to describe the CRS. It is also necessary to highlight the variability of used numerical, mnemonic, textual and structured textual encodings used by format owners. The metadata generated within this task allow users to use unambiguously datasets having different spatial reference systems or type of coordinates.

Once the CRS has been identified, it is possible to study whether it is necessary to carry out the task of *Coordinate Transformation* in order to store the geographic extension of a geospatial dataset by means of their geographic latitudes and longitudes with the WGS84 datum. Similarly, it is necessary to highlight the variability of existing CRS when carrying out the coordinate transformations/conversions. The outcomes of this task can support both the functions of discovery and use since they provide the coordinates that define the minimum enclosing rectangle of a data set.

The task *Data Model Infer* aims at a standardized language model in terms of feature types and relationships. This model makes it easy for a user to evaluate and use the attributes and relationships between different types of geographical objects. The result of this task is a UML class model that has been defined and stored in XML interchange format (XMI).

The task defined as an *Application Schema Builder* transforms a data model inferred in the previous task into a GML application schema and stores it in the XML format (XSD). As a result, this task provides users with a standard XML schema that can be used to exploit geospatial datasets from different applications.

For the evaluation function, the automated generation of metadata is associated with the task of *Data and Data Model Graphic Preview* which consists of the production of graphics for allowing users who are interested in a particular type of data or data model to select which one can meet their needs.

Finally, the last task is the *Metadata Constructor, Packager and Format Exchange.* In order to be able to integrate this task into the workflow (e.g., automated generation, metadata wizard assistance, and metadata update), its input should contain the minimum information necessary for allowing data access [23][34]. Currently, the best technical approach of identifying access to local/remote geospatial datasets is through the Universal Resource Identifier (URI). Concerning the output format, it consists of metadata containing information about the file, a set of graphic files and XML coded data models packed together in a pre-defined structure as the one proposed in MEF (Metadata Export Format) of the GeoNetwork project [35]. However, in some cases the output format of this task consists of an XML document with the metadata. It is predicted that the output of a task can lead to the register of metadata in a catalogue that is only able to manage metadata in exchange XML format (ISO 19139, Dublin Core or others), and it is not able to manage the challenge of graphic documents and XML.

4   The synchronization of tasks

The proposed workflow consists of a sequence of connected tasks that represent the flow of work in the automated generation of metadata. Figure 4 illustrates such a synchronization of tasks which have been described in the previous section. In the

synchronization process, the tasks are always related to each other by establishing an order of precedence. For example, the task of *Metadata Extractor* is essential to the workflow since it stars the whole sequence of other six tasks of the workflow, which are *Data Analyzer*, *CRS Interpreter*, *Coordinate Transformation*, *Data Model Infer*, *Data and Data Model Preview*, and the *Constructor metadata*. It is important to point out that Figure 4 does not illustrate the interactions of the tasks with a geospatial dataset due to clarity purposes. These interactions take place during the execution of the tasks: *Metadata Extractor, Data Analyzer, Data and Data Model Preview Generator, and Data Model Infer.*

The tasks can run simultaneously or in parallel, and some of them require the connection with the knowledge database for the identification of CRS, type of data set, keywords, and others. For example, the *Metadata Extractor* retrieves elements which are fundamentally: a) implicit and b) embedded, as they are stored in the geospatial dataset; c) structural, as they are necessary for the use of geospatial dataset; d) many of them are static, remaining unchanged through time, unlike the type of geometry of those phenomena represented; e) others, like the feature count of an entity or the geographic extension, are dynamic, as they may change over time. The *CRS Interpreter* and the *Data Model Infer* generate static, structural metadata elements from the implicit information. The *Data Analyzer*, the *Cataloguer* and the *Title redactor* generate metadata elements: a) dynamic, as they depend on the geospatial dataset's features at any given time, such as the geographic extension and b) subjective, as the procedures of classification and cataloguing generate diverse results depending on who interprets them, or how.
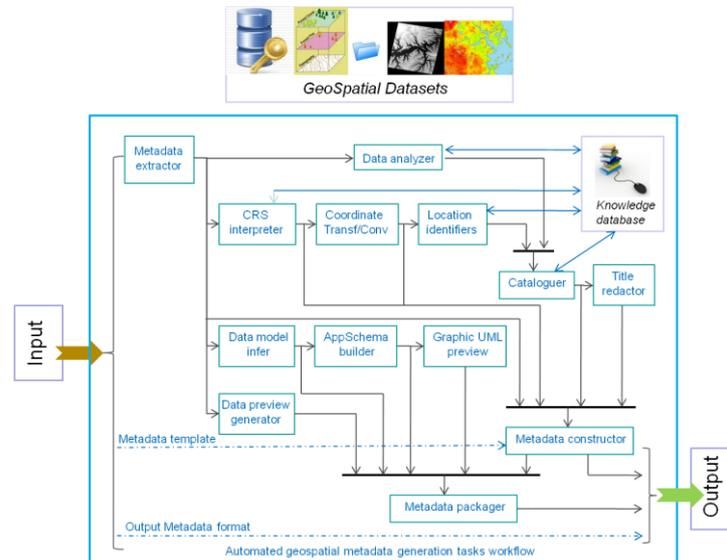

Fig. 4. Overview of the task synchronization of the proposed workflow.

In relation to the tasks of metadata production, the proposed workflow implements the methods of extraction, calculation and inference proposed by Beard [11] in the tasks of *Metadata Extractor, Coordinate Transformation, Location Identifiers, Cataloguer*

and Title Redactor. The workflow does not include alternative procedures for the generation of metadata elements as Balfanz [34] suggests, although several representations of the data model are generated in a parallel way. From the point of view of visualization of metadata, the proposed workflow provides graphics of both the class diagram and data preview. The workflow supports the use of templates of metadata [24][25], or already created metadata to which automatically generated elements can be a contribution. As a result, the automated generation can be used as have been proposed by Guy *et al.* [23] in the life cycle of metadata.

5   Implementation and Results

The workflow was implemented using a multi-tier architecture system where the Data, Application and User Tiers can support a desktop application or a web service. Some of the tasks identified in the workflow have been integrated into a single module of a particular tier and others have been implemented as independent modules (Figure 5).

In the Data Tier, the metadata elements can be extracted, calculated or inferred for a given geospatial dataset, whether in a table format or in text paragraphs. The difficulty lies in the fact that each type of geospatial data (e.g., aerial images, multispectral images, DTM, drawing files and vector layers) is related to different types of metadata, to which we must add the variability of meta-information that can be extracted from each format. Data Tier main modules are based on the libraries such as Java *Mapscript* libraries (MapServer) to draw previews and Java wrapped GDAL/OGR libraries. Besides, there is a set of libraries that allow the access to additional information and the extraction of some graphic formats. The access to these modules has been homogenized by an enclosure that extends the data structures GDAL/OGR.
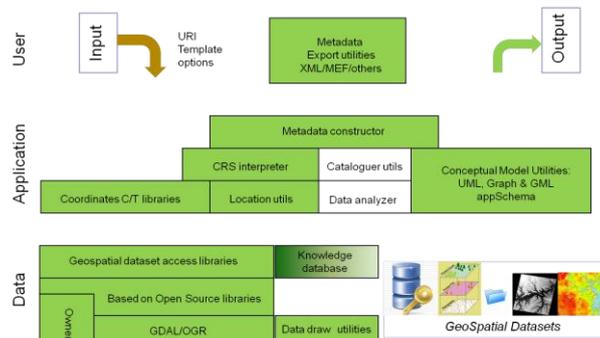

Fig. 5. Overview of the implementation architecture.

In the Application Tier, for example, the Proj4 library and utilities provided by the library GDAL/OGR are used for the transformation of coordinates.

The *CRS interpreter* module based on GDAL/OGR is used to interpret the texts themselves, numbers, mnemonics and other encodings. It needs the information stored knowledge database defined by EPSG. The *Location identifiers* module is based on a world database of place names consulted with the *Mapscript* libraries. Finally the conceptual model utilities module is based on a set of libraries that allow us to infer the

data model of spatial databases or file directories of shapefiles based on primary and foreign keys and attribute names and types. To draw the UML data model of a graphic mode has been used Argo UML libraries. Data analyzer extracts information by methods of image classification, or by text analysis (data mining) to support the cataloguer. This uses that information to inference data content and selects a list of words belonging to multilingual thesauri stored in the Knowledge database, in order to propose keywords capable of cataloguing the geospatial data. Also, a selection of terms that define the 'topicCategory' of the resource will be carried out. Metadata constructor assembles all the information that has been provided, extracted, calculated, inferred and elaborated as described in previous phases.

In the User Tier, the metadata component is developed with export utilities that enable libraries own a complete XML metadata template or create a new metadata. In the current development, it is possible to export in ISO19139 or MEF formats.

Table 1. Overview of the automatically generated metadata elements (The complete list of metadata elements is available [here](#))

MD_Metadata: *Packed*	Metadata element	(G) Extract, compute, infier; (C) dataset dependent; (I) imposed	(R:Raster, V:Vectorial, D:DTM)	Function: (D:Discovery, E:evaluate, A:access, U:use)
	language	G		
	metadataStandardVersion	I		
	dataSetURI	G		
	dateStamp	G		
distributionInfo:transferOptions	unitsOfDistribution	G		A
distributionInfo:transferOptions	transferSize	G		A
distributionInfo:distributionFormat	name	G		A
distributionInfo:distributionFormat	version	G		A
contentInfo:MD_CoverageDescription	contentType	G	R	U
contentInfo:MD_CoverageDescription: dimension: MD_Band	bitsPerValue	C	R	U
contenInfo:MD_ContentInformation	featuresTypes	G		U
spatialRepresentationInfo: MD_GridSpatialRepresentation	numberOfDimensions	C	R	U
axisDimensionProperties:MD_Dimension: resolution	units	C	R	U
spatialRepresentationInfo: MD_VectorSpatialRepresentation	topologyLevel	C	V	U
identification:..: citation: CI_Citation	title	G		D
identification:MD_..: CI_Date	date	G		D
identification:.. MD_BrowseGraphic	filename	G		D
identification:.. MD_LegalConstraints	useConstraints	G		E
identification:MD_Identificacion: descriptiveKeywords: MD_Keywords	keyword	G		D
identification:MD_Identificacion: descriptiveKeywords: MD_Keywords	type	G		D
identification: MD_Identificacion: spatialResolution: MD_Resolution	distance	C	R	D
identification: .. EX_GeographicBoundingBox	westBoundLongitude, etc	G		D
identification: ..:Ex_VerticalExtent	minimumValue	C	D	D
identification: ..:Ex_VerticalExtent	maximumValue	C	D	D
dataQualityInfo: DQ_DataQuality: scope: DQ_Scope	level	G		E
dataQualityInfo: ..: LI_Lineage: source : LI_Source	description	G		E
applicationSchemaInfo: MD_ApplicationSchemaInformation	softwareDevelopmentFile	G		U
referenceSystemInfo: ..: RS_Identifier	codeSpace	G		U
referenceSystemInfo: ..: RS_Identifier	version	G		U
		G:64		D:34
		I:3		E:6
		C:21		A:5
				U:36

Table 1 illustrates some of the results of the implemented workflow. In the first table column we show the class where the metadata belongs (e.g., MD_Metadata), while the

second one identifies the element (e.g., language). The third column classifies the items as "G", which stands for "generated" (extracted, calculated or inferred), "C" with cardinality dependent on the dataset and finally "I", when a fixed value can be set by agreement. The fourth column identifies those elements that can only be applied to certain types of geospatial datasets ("R" raster data, "D" DTM and "V" vector data). The fifth column of the table identifies the function performed by the metadata element ("D" discovery, "E" evaluation, "A" access and "U" use). In order to carry out the classification of the functions performed by the metadata elements, we have used the classification proposed by Danko [36] for metadata packages.

Our proposed workflow for the automated generation of metadata can compile 83 metadata elements for images, 69 for vector data and 68 when it comes to DTM. The final number of elements will depend on the used format and the information it contains. This number of elements can be increased if the digital repository contains more than one band or a layer with 21 more elements identifies for the three classifications (R, V and D), with cardinality higher than or equal to one. There is also another set of elements with cardinality *n* generated by the cataloguer (*keyword*, *theme*, thesaurus name *title*, *date* and *topicCategory*).

Although there are only a few specific metadata elements for DTM (2) and for vector data (3), while many for raster data (17), these data can't be considered decisive, because for vector data we will obtain the UML class diagram that contains the definition and its relations, providing a more elaborate and useful type of information that balances the results. Bearing in mind the functions performed by the metadata elements, 34 of them perform the function of discovery, 6 perform the evaluation, 5 perform the access and 36 perform the use; therefore we can conclude that the functions that benefit most from the automation are *Discovery* and *Use*.

6   Conclusions

The automated generation of metadata responds to the need of providing users with the ability to exploit metadata functions (*discovery*, *evaluation*, *access* and *use*), to provide those who are responsible for the production of geospatial data sets with the ability to manage them. We have analyzed the few workflows described in the bibliography, concluding that from those workflows there is a highest quality when metadata are initially created automatically, having been later complemented by the authors of the resources and finally catalogued by information specialists.

Our new workflow also contributes to reducing time costs and error rates, which in turn can free up the creators of metadata to devote their efforts to more complex tasks instead of the monotonous tasks such as opening several data sets in order to know the coordinates, format and version. Besides, our workflow is compliant with the previously tasks proposed in the literature, mainly the extracting implicit information, making calculations with it and inferring another one. On the other hand, there are several new tasks to this workflow. They can be described as one of the following:
-    the achievement of a preview of geospatial dataset that helps users to e valuate their applicability;
-    the definition of the data model and its representation in standardized languages currently employed in engineering as well as its graphic

presentation, which help both the users and the application developers to evaluate and exploit geospatial data sets;

- the analysis of geographic information in order to classify the information and help in the cataloguing of the resource, and
- the packaging of the information in order to make the interchange possible.

The main scientific contributions can be mainly summarized by the association of metadata elements with the nature of the geospatial data sets (e.g. raster, vector and DTM), their cardinality, the function performed by the element (i.e., *discovery*, *evaluate*, *access* and *use*). This association has shown that those metadata functions that benefit the most from the automation are *discovery* and *use.*

The implementation of the workflow in a multi-tier system can help considerably in the massive generation of basic and initial metadata. Once they have been stored in catalogues, these metadata facilitate the discovery and reutilization of data. At the same time, the creation of quality metadata, once they have been revised and complemented by the creators of data and catalogued by information specialists, can be also facilitated. Our future research work will focus on the analysis of our workflow using the *metagraph* methodology previously proposed by Basu [37] in order to enhance the tasks of *Data Analyzer, Cataloguer and Title Redactor.*

7   References

1. Howe, D.: Free on-line dictionary of computing. http://foldoc.org/index.cgi?Metadata
2. Zeigler, B., Murzy, A., Yilmaz, L.: Artificial Intelligence in Modelling and Simulation. Encyclopedia of Complexity and System Science. Springer-Verlag, Germany, (2006)
3. ANZLIC: Core Metadata Elements for Land and Geographic Directories in Australia and New Zealand. The Australian New Zealand Land Information Council, (1997)
4. Kildow, M.: The value of Metadata. US Fisheries and Wildlife Services, http://www.r1.fws.gov/metadata/meta.html (1996)
5. ANZLIC: ANZLIC Guidelines: Core Metadata Elements Version 1 Report: ANZLIC Working Group on Metadata (1996)
6. Ercegovac, Z.: Introduction. Journal of the American Society for Information Science, vol. 50, n. 13, pp. 1165--1168 (1999)
7. Milstead, J., Feldman, S.: "Metadata: Cataloguing by any other name..." http://www.onlineinc.com/onlinemag/metadata/
8. Caplan, P.: *You call it corn, we call it syntax-independent metadata for document-like objects.* The Public Access Computer Systems Review, Vol. 4, n.6 (1995)
9. Greenberg, J.: Metadata extraction and harvesting: a comparison of two automatic metadata generation applications. Journal of Internet Cataloguing. Vol. 6 n. 4, pp. 59--82 (2004)
10. Kolodney, U.: Metadata requirements for a digital repository to accompany the American Anthropological Society's Anthro Source portal Project.
11. Beard, K.: A Structure for Organizing Metadata Collection. Third International Conference/ Workshop on Integrating GIS and Environmental Modelling (1996)
12. Caplan, P.: *Metadata Fundamentals for All Librarians.* Chicago: American Library Association (2003)
13. Leiden, K. et al., S.: A Review of Human Performance Models for the Prediction of Human Error. National Aeronautics and Space Administration, USA, 125pp (2001)
14. ISO 19115: Geographic Information – Metadata, (2003)

15. Najar, C.: A model-driven approach to management of integrated metadata – spatial data in the context of spatial data infrastructure, (2006)

16. CGIAR-CSI: Metadata Tips: Why Metadata? (2004)
http://www.csi.cgiar.org/metadata/Metadata_Why.asp

17. Liddy, E., Sutton, S., Paik, W., Allen, E., Harwell, S., Monsour, M., Turner, A., Liddy, J.: Breaking the metadata generation bottleneck: preliminary findings. In: Proceedings of the 1st ACM/IEEE-CS Joint Conference on Digital Libraries, Roanoke, Virginia, 464pp (2001)

18. Anderson, J., Pérez, J.: The nature of indexing: how humans and machines analyze messages and texts for retrieval: part I. Information Processing and Management: An International Journal, Vol.37 n.2, pp. 231--254 (2001)

19. Liu, J.: Metadata and its Applications in the Digital Library: Approaches and Practices, Libraries Unlimited, London, pp.143--149 (2007)

20. Downey, D.: What Do Geologists Need to Know about Metadata? (2007)
http://www.searchanddiscovery.net/documents/2007/07030downey/images/downey.pdf

21. Greenberg, J., Spurgin, K., Crystal, A.: Functionalities for automatic metadata generation applications: a survey of metadata experts' opinions. *Int. J. Metadata, Semantics and Ontologies,* Vol.1, n.1 (2006)

22. Craven, T.: DESCRIPTION meta tags in public home and linked pages. LIBRES: Library and Information Science Research Electronic Journal vol.11 (2) (2001)

23. Guy, M., Powell, A., Day, M.: *Improving the Quality of Metadata in Eprint Archives* (2004).
http://www.ariadne.ac.uk/issue38/guy/

24. Morris, S., Nagy, Z., Tuttle, J.: North Carolina Geospatial Data Archiving Project. *NCSU Libraries and North Carolina Center for Geographic Information & Analysis* (2007)

25. Hedorfer, M., Bianchin, A.: The Venice Lagoon Experimental GIS at the IUAV. *Interop99: The 2 International Conference on Interoperating Geographic Information Systems* (1999)

26. Gayatri and Ramachandran, S.: Understanding Metadata. The Icfai Journal of Information Technology (2007)

27. Johnston, P.: Good Practice Guide for Developers of Cultural Heritage Web Services. Research Officer, UKOLN, http://www.ukoln.ac.uk/interop-focus/gpg/Metadata/ (2007)

28. Nebert, D.: Developing Spatial Data Infrastructures: The SDI Cookbook (2004)

29. Gilliland-Swetland, A.: *Setting the Stage*, in Introduction to Metadata: Pathways to Digital Information. Los Angeles: Getty Research Institute (2000)

30. Senso, J. y Rosa Piñero, A.: El concepto de metadato. Algo más que descripción de recursos electrónicos. Ci. Inf., Brasília, v. 32, n. 2, p. 95--106 (2003)

31. Jones, M. y Taylor, G.: Metadata: Spatial Data Handling and Integration Issues. School of Computing Technical Report. Issued: February (2003)

32. Moellering, H. y Brodeur, J. : Towards a North American Profile of the ISO 19115 World Spatial Metadata Standard. *GSDI-9 Conference Proceedings*, 6-10, Chile (2006)

33. Díaz, L., Gould, M., Beltrán, A., Llaves, A. y Granell, C.: Multipurpose Metadata Management in gvSIG. Proceedings of the academic track of the *FOSS4G Conference*, 29, Cape Town, South Africa. ISBN: 978-0-620-42117-1, pp 90--99 (2008)

34. Balfanz, D.: Automated Geodata Analysis and Metadata Generation Visualization and Data Analysis. Proceedings of SPIE Vol.4665 (2002)

35. MEF: Metadata Exchange Format (2007). http://www.fao.org/geonetwork/docs/ch17.html

36. Danko, D.: ISO/TC 211, the Standards in Action. Workshop Implementation (2002)

37. Basu, A., Blanning, R. A formal approach for Workflow Analysis. Information System Research; ABI/INFORM Global; Vol. 11 n.1, pp.17--36 (2000)