CO-027

INTEGRATION OF THE SOUTH AMERICAN BLOCK IN THE AMERICAS GLOBAL MAP PROJECT

HUNT E.(1), MORALES C.(1), LOBOS A.(2), CHACANO Y.(1)

(1) Instituto Geografico Militar, SANTIAGO, CHILE ; (2) Ministerio de Obras Publicas (MOP), SANTIAGO, CHILE

1. ABSTRACT (reduced version)

Introduction and Background

This paper summarises the current status of a data integration program involving Global Map data from South American countries. The Americas Global Map initiative (MGA) was established by the Pan-American Institute for Geography and History (PAIGH) under the terms of its cooperation agreement with the International Steering Committee for Global Mapping (ISCGM). MGA was directed at developing regional coordination on Global Mapping issues; however, MGA is institutionally and economically dependant on PAIGH.

The central objective of MGA was the integration of all existing Global Map datasets into a single, continuous map of the Americas. This paper describes an MGA sub-project involving existing Global Map vector datasets.

Objectives

The work performed achieved the integration of five South American datasets as the first practical step towards completing the South America block and then the Americas Map as a whole. At the same time, the work constituted trials of methodologies previously designed by contributors to MGA.

Methodology and Approach

The Coordinator of the MGA recruited a team of spatial data specialists who downloaded the existing datasets published by ISCGM for Argentina, Brazil, Chile, Ecuador and Uruguay. The sequence of work was as follows:

1. Check the input data for compliance with ISCGM Global Map specifications and for adaptability to the work at hand.
2. Create a workspace in a GIS environment
3. Integrate the data for coastline (line), watercourse (line) and freshwater bodies (polygons).
4. Integration of all other layers except administrative boundaries
5. Integration of administrative boundaries
6. Trials involving the union of certain selected objects across international boundaries

Data integration means that, for each layer, data from different countries was merged into the same database table. It does NOT involve reconciling different versions of the international boundaries.

Results

The MGA team successfully integrated the datasets as planned and completed trials of the methodologies.

Conclusions and Future Plans

Maps of whole continents are of course common; the value of this exercise is that it combines data originating from different institutions in their respective countries, overcoming some compatibility issues while leaving the essential information and its authorship largely intact. The next step of this project will be to add data to the South America block from the countries not yet published by ISCGM.

Note: the above is a reduced version of the abstract submitted in October 2010, serving here as the introduction to the paper. The rest of the paper, below, is being submitted in February 2011.

2. BACKGROUND ABD OBJECTIVES

Global Mapping and ISCGM

Global Mapping projects have been coordinated at world level by the “International Steering Committee for Global Mapping” (ISCGM) since the 1990’s. Its purpose has been to develop and then release to potential users, free of charge, a small scale (mostly 1:1,000,000) digital map of the whole globe. The structure and the purpose of the ISCGM Global Map can be reviewed in the ISCGM web site (www.iscgm.org).

Initially the main partners of the ISCGM were the National Mapping Organisations (NMO) that provided the source material for first version Zero and then Version One of the Global Map. ISCGM joined the Joint

Board of Geospatial Information Societies (JBGIS) and developed agreements with regional organisations, some of these aimed at integrating the national Global Map datasets into regional maps. The first of these regional initiatives was that of the "EuroGlobalMap", developed by the Eurogeographics agency to cover the European continent.

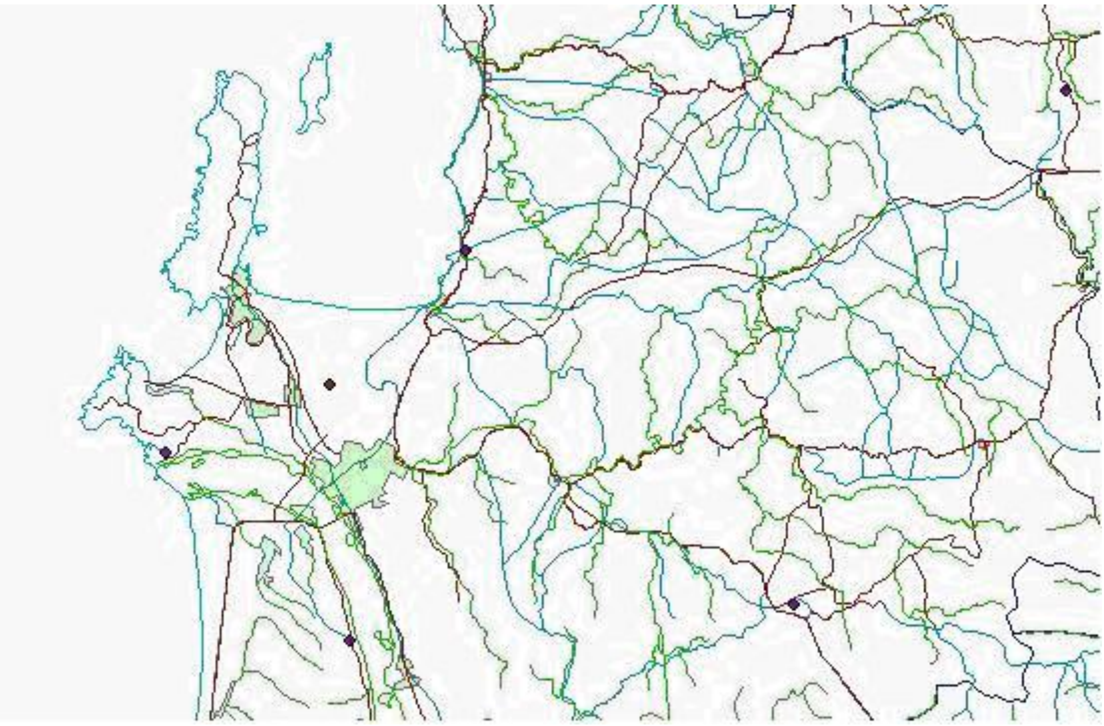

Figure 1. Extract from Global Map of Chile (Concepción area) in GM V.1

PAIGH and MGA

Around the year 2004, ISCGM entered into cooperation agreements with the Pan-American Institute for Geography and History (PAIGH), an international scientific organisation attached to the Organisation of American States (OAS). PAIGH was, and is, the only organisation in Americas region able to push forward a regional integration of Global Map national projects, having the necessary political backing from PAIGH member states, relations with mapping agencies, some limited financial resources and previous involvement in SDI initiatives. To put the PAIGH-ISCGM agreement into concrete form, a PAIGH technical cooperation project was set up with the title "Mapa Global de las Americas - MGA" (Americas Global Map); the first step taken was for representatives from several PAIGH National Sections (in effect, the NMO's supporting PAIGH activities in each member state) to set up an Executive Committee for MGA and define a range of objectives and activities. MGA continued as a technical cooperation project until the end of 2009, while its Executive Committee was, and continues to be, a Working Group of the PAIGH Cartography Commission, thus linking it to the broader scientific programs of PAIGH.

MGA assessed several related ideas for action involving global mapping; the core project was to be the integration of pre-existing Global Map vector datasets already prepared under ISCGM technical specifications. The Global Map does also include data levels – mostly thematic – in Raster structure, but these have been derived from global coverages of remote sensing (satellite) data; from this point on we are concerned only with vector-structure data.

Here we come to the main issue that the sub-project discussed in this paper had to address. Each vector dataset provided by a NMO cooperating with ISCGM has naturally portrayed only the territory within the terrestrial political boundaries of the country where that NMO operates. As ISCGM has received these datasets, it has published them in separate folders and files from its server. In theory, the whole collection is called Global Map Version One. But can we really call this collection of separate maps a "Global Map"?

MGA and a Continental Global Map

MGA intended to develop a methodology for integrating the existing datasets into a single continental map of the American continents. An outline methodology was drafted during 2007 – 2008 and discussed within MGA. It was clear that integrating the whole hemisphere at the same time would be impossible. Eurogeographics had already integrated its EuroGlobalMap by first uniting datasets from a group of adjacent countries, dividing Europe up into several sub-regions. For the Americas, at least three sub-regions or “blocks” would be necessary for preliminary integration, prior to final integration of all the blocks.

The three datasets of North America (Canada, USA and Mexico) have already been integrated into a continental version, although this was done directly for ISCGM, largely outside the PAIGH – MGA initiative. For South America, MGA initiated actual work on data integration of five South American countries as a sub-project of the MGA in 2009. This paper describes the work achieved so far on the South America block and discusses its implications.

Specifications and Standards

The Global Map Specifications issued by the ISCGM have always been the basic standard for the continental map as envisaged by MGA. This specifies a scale of 1:1,000,000 and georeferencing to the WGS 84 Datum. For the Vector data, nine mandatory feature classes and a further nine optional feature classes are described as the content. For each feature class, the graphic symbol type (line, face or point) is described and the attributes to appear in the alphanumeric tables associated with the graphics are listed. Global Map Specifications were issued as Version One from 1998 onwards, with piecemeal revisions being made over the years. ISCGM issued a completely revised version of the specifications as Version Two from October 2009 onwards, however, the South America block sub-project needed to consider Version One as the base specifications because the source data had been developed by the relevant South American NMO’s under Version One.

The question arises whether these specifications can relate only a specific type of project (those linked to ISCGM) or whether these can be considered a potential standard for all small-scale digital mapping intended for non-commercial release.

Some adaptation was necessary for the region and for the specific task of integration. For South America, the relevant reference framework is known as the “Geocentric Reference System for the Americas” – SIRGAS, which is entirely compatible with WGS 84, and is used by the majority of those NMO’s in the region that contribute to ISCGM’s Global Map.

A key decision taken by MGA was to set a rule concerning international boundaries; the international boundaries where two territories, represented in different national datasets, are adjacent, will be left as they are; this means leaving two separate lines in the frontier zone rather than attempting to ‘merge’ the two lines into one. In relation to specifications, the ISCGM specs do not address this because of the way that the national datasets are published.

Objectives of the South America Block Project

Given all of the above, the objectives of the work performed in 2009 – 2010 for the South America block were:

1. Completing data integration as a step forwards towards first of all the integration of the south American block and then, beyond that, potential unification of the entire Americas

2. Testing and developing the integration method itself as a trial

3. Demonstrating that concrete results can be obtained, this being vital for obtaining the institutional support needed for further progress to be made

2. METHODOLOGY AND RESOURCES

MGA Overall Method

The methodology for data integration drafted within the MGA Executive Committee outlined the steps required to integrate data from several countries in a single block. Initially it was not specific to any of the sub-regions, however, it came to be the basis of the work plan for the south America block. The idea was to have available an outline work plan for whichever NMO or other body contributing to PAIGH could be found to take on implementation. This is exactly what happened, except that the work for the South America block was done directly by the MGA Work Group / Executive Committee chair in his own country.

Trans-boundary linkage Method

During 2009 the overall method developed by the MGA was complemented by a proposal for a specific task needed for the area around the “joins”, that is, the areas close to the boundary between two adjacent national datasets. The proposal is for linking objects that need to cross international frontiers; watercourses, roads, railways and the edges of inland water polygons. In summary; the general sequence is as follows:

(a) For a sector of an international boundary (between two nodes on either a three-nation intersection or an intersection with a coastline) copies of the two line elements (one line from the each respective national dataset) of the two adjacent countries are inserted into the same coverage.

(b) A single buffer area extending for 1,000 meters is generated around the two boundary lines, converted to raster structure and ‘cleaned’ of “islands” caught within it where the boundaries separate beyond the buffer edges.

(c) A center line is generated by algorithms down the exact middle of the the buffer zone; this approximates to a mean between the two boundaries.

(d) The two ends of the road or river to be joined should both enter the buffer area; if they do not, it may not be possible to link them.

(e) From here on, the operation on the ends of object are performed manually by the operator, but always taking care to follow a route suggested by the general shape and the location of the new center line, the buffer edges and the highway or river route on each side. The extreme ends of the two lines to join are extended or moved until they meet at a common point located on the center line. As few vertices as possible are moved in the two lines to adjust, and none at all outside the buffer zone.

(f) The union on the center line is a new common point, but the two line elements are not joined into a single segment across the boundary; they must remain as separate records in the database, with their own attributes.

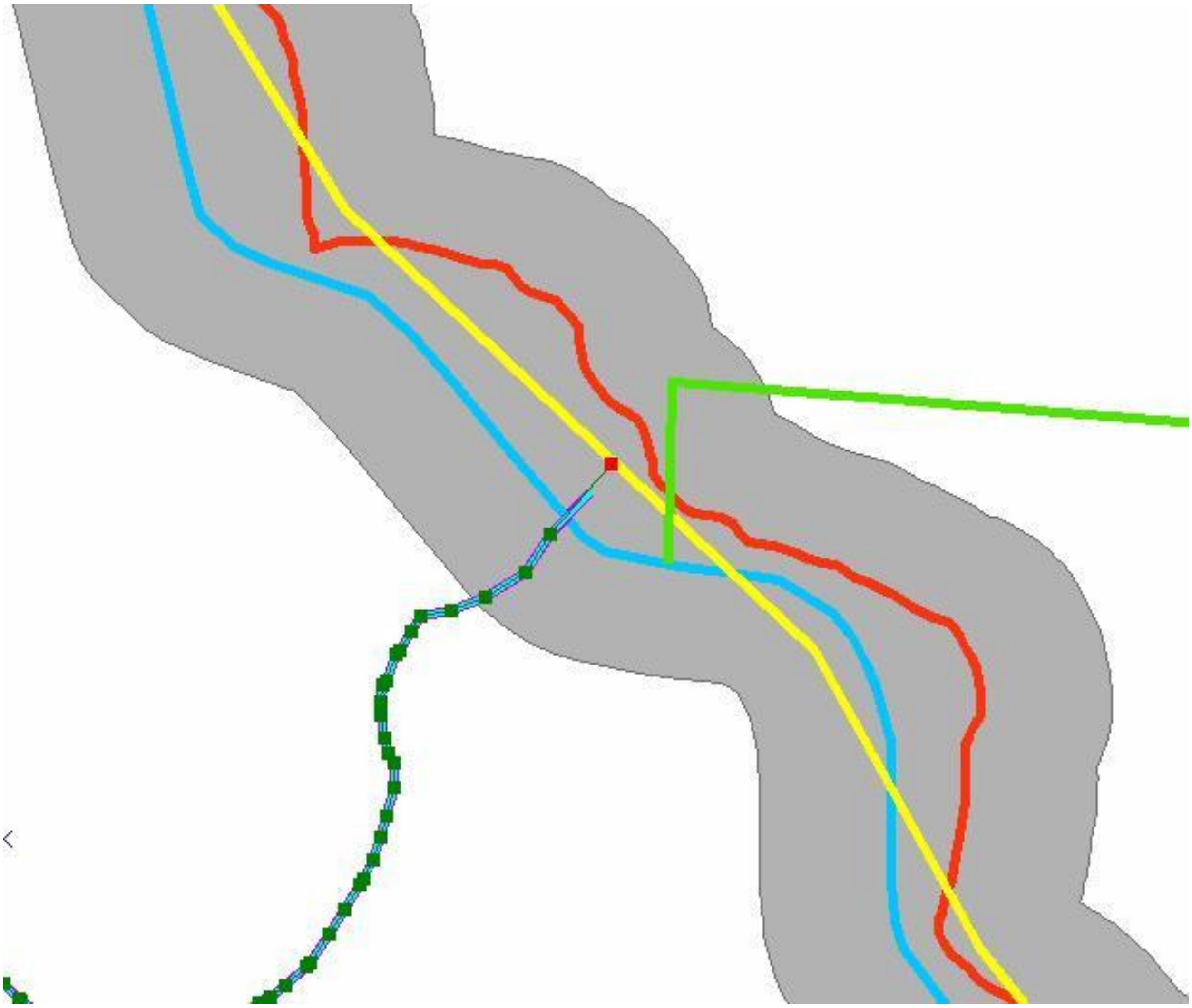

Figure 2. Trial of trans-boundary object union procedure, with existing international boundaries (red and blue), buffer (grey), center line (yellow), and broken object to be joined (green).

The final objective is to eliminate gaps and breaks in the objects in the area close to the international boundary. The objective is NOT to modify the lines making up the international boundaries themselves; these should remain unchanged in the final map. The buffer zone and the center line generated from these boundaries could be published, but only as a technical annex to the main map, to demonstrate how the adjustments were made, without having any official status.

Resources

The team that worked on the South America block during 2009 and 2010 consisted of four persons, all based in Santiago, Chile:

- Team supervisor: On this sub-project, the supervisor set the general objectives and conditions, oversaw the technical aspects and managed the administrative aspects, in addition to being the coordinator of the MGA Work Group.

- Technical Adviser: a professional with prior experience of MGA-related technical research, who provided expert advice, problem-solving and checks on the work of the operators.

- Lead Operator: Performed nearly half of the actual work of data integration, also some detailed planning and reporting in collaboration with the team supervisor

- Assistant Operator: Performed the remainder of the work of data integration.

The Technical Adviser and both Operators are all qualified and experienced professionals. However, they could not perform the tasks on a voluntary basis nor as part of their duties for the institutions where they work, so their work was treated as a service and paid for, per hour, out of the budget assigned by the PAIGH to the MGA as a technical cooperation project. The number of hours they could dedicate to this task was determined by the budget available; in the event, the money available was enough to get the job done, but only just. The data processing was performed by the two operators in ArcGIS 9.2.

Input Data

The raw material for the data integration work consisted of the national vector datasets for five countries in South America which the ISCGM has already published in its web site. It should be emphasized that the authorship of the source data is exercised by the five NMOs that originally provided these datasets to ISCGM. These institutions are:

- Argentina: Instituto Geográfico Militar – IGM (from 2009, renamed as Instituto Geográfico Nacional - IGN)

- Brazil: Instituto Brasileiro de Geografía e Estadística – IBGE

- Chile: Instituto Geográfico Militar de Chile – IGM

- Colombia: Instituto Geográfico Agustín Codazzi – IGAC

- Uruguay: Servicio Geográfico Militar – SGM

The work of data integration described here did not aim to create any new geographic information; rather it integrated existing data whose authorship continues to be held by these five national mapping agencies (or, as Latin American professionals prefer to call them, geographic institutes) and is respected by the MGA team.

The data integration carried out so far has necessarily been restricted to these five countries. Other South American countries have developed datasets for ISCGM; however, while these are not yet published by ISCGM, they are not in the public domain. Moreover, with hindsight we can look back and see that the funding and time available in the recent period would have been insufficient to allow work on the data from any more South American countries outside these five.

Work Sequence

The sequence of work actually performed was based fairly closely on the general methodology defined by the MGA, with many details adapted for practical considerations arising.

The first stage was to inspect all the data downloaded from the ISCGM site and check for these aspects:

- Conformance with ISCGM specifications for the Global Map

- Comparison among the five datasets themselves to determine the degree of homogeneity versus divergence between them

- Compatibility with the integration processes intended for them (in other words, if they had any characteristics that might impede their use in the following stages)

A geo-referenced work space was set up into which the data was to be inserted. The first three feature classes to be inserted were those corresponding to:

- natural water courses (mostly rivers) - line

- coastline (mostly oceanic) - line

- bodies of inland water (mostly freshwater lakes and wide rivers, plus some artificial reservoirs) – face

Collectively, these three features, when combined correctly with the georeferencing, made up the base into which the other features could be inserted.

Integration involves combining all five countries in one feature class into a single database. This meant making the database tables compatible in structure and then merging into a single table. Differences in the tabular structure among the five countries needed to be resolved, so also were problems arising from divergent topologies among the graphic elements.

With the base set up, most of the other feature classes – all those portraying physical objects actually existing in the terrain – were inserted and integrated; these classes were:

- Roads and highways - line
- Trails (mostly unpaved trackways in areas distant from paved roads) – line
- Railways – line
- Ferry routes - line
- Airports (mostly international or regional airports, excluding smaller aerodromes) - point
- Artificial watercourses (mostly irrigation channels) – line
- Settlements (mostly small to medium size towns) – point
- Built-up areas (mostly urbanized areas of larger towns and cities) - face

The final feature classes to be inserted were two classes representing administrative boundaries. The lines in this class are distinguished by their attributes into three types, broadly in accordance with the UN - SALB classification of levels. The first level is the international boundary; the second and third levels are the internal administrative boundaries of the respective nation at the two highest levels. The other feature class is a set of polygons, covering the whole of the territory; one polygon for each of the third-level administrative unit.

At intervals the team technical adviser reviewed the integrated datasets to check that optimum results were being achieved within what could be reasonably expected with the available resources and governing conditions.

Once most of the feature class integration was complete, the lead Operator performed some experiments on a few selected line objects that cross between two of the countries involved, where these have a common boundary and where the pertinent objects could be found in the datasets approaching the international boundary on either side. The project budget was running out at this stage so only a very few highways and watercourses, with no water body polygon edges, could be tested using the trial procedure.

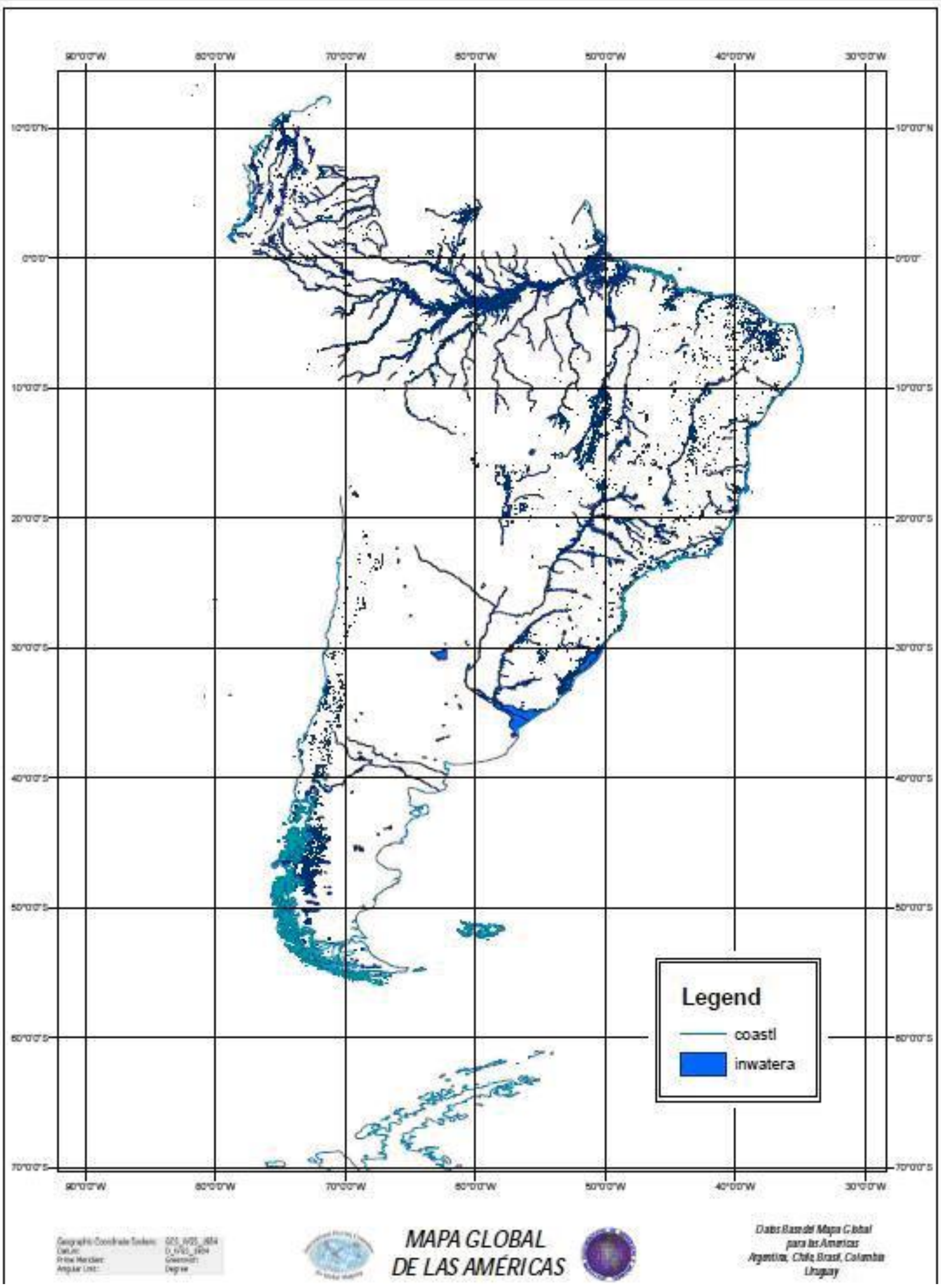

Figure 3. Trial display of integrated coastline and water body datasets

4. RESULTS

Comparison of Five Countries

The assessment of the Global Map material downloaded from ISCGM revealed a large number of differences. The variations on the list of the feature classes is to be expected, given that nine of the Global Map feature classes are optional; several differences were found between the five feature class lists, but always limited to the optional features. Moreover, some differences are inevitable; for example, a landlocked country cannot be expected to present a coastal feature class.

Given the origins of each national dataset as an independent project by each NMO, having total liberty to choose its source data, it is also inevitable that the datasets differ in the scale of the source material used.

IGM – Chile, for example, used an existing coverage at 1:250.000 scale. All the national datasets had been re-scaled to 1:1,000,000, yet the generalisation methods varied.

The national datasets did converge on the GM specifications prior to delivery to the ISCGM, and the ISCGM technical support unit had done some limited processing on these datasets, only enough to ensure that they were sufficiently close to the specifications and compatible with the publication method. Nevertheless, true harmonisation between these datasets was still lacking, as found when actually starting to put them together into single, continent-wide coverages for each feature class. The differences between the source material showed up particularly in differing densities of vertices.

Quality of Five Input Datasets

It is not part of the work of this project to criticise the quality materials received beyond assessing whether is conforms to ISCGM specs and whether it can enter the integration process or not. Nevertheless, a problem was observed with too much frequency to be ignored; this is the inconsistencies in the application of names where the attributes require geographic names. Rather than go into detail, we can only say here that there are inconsistencies within and between the countries reviewed. Some complications may arise from differences (between the authors and the GIS systems holding the data) in terms of the character and symbol sets used. Others may be due to language issues.

Another problem in the database tables was that some records appeared to be blank in the alphanumeric aspect, that is, there were no attributes in the relevant fields, yet the record was attached to a graphic object. These tended to appear in line segments occurring mid-way along a single object such as a river or road, where the segments on either side were correctly attributed.

Variations in the source material also explain differences in the symbology; on displaying the coastlines, the colour code was found to be different between countries; although this aspect would be easy to homogenize.

Experiences of Working with Data Integration

The work actually performed was completed more or less according to the initial expectations, except that it was necessary to dedicate extra time to solving problems of topology in the line feature classes with greater volumes of data. In some of the feature classes, variations between the structures of the tables from each country needed to be resolved in such a way as to enable the ISCGM-mandated attributes (usually the fields to the right of the displayable tables) homogenous all down the single, integrated DB.

Trans Boundary Objects

For all the cases where the union of roads or rivers were attempted, the use of the buffer zone and center line was found to be a considerable help to deciding where to adjust the extreme ends and place the joining node; nevertheless, every case had its particularities which required the operator to exercise careful judgement. The particular situation of each case arise in large measure because of the same differences between the datasets – within the same feature class – already mentioned above for the different countries.

A particular problem arises where a single object has been classed into different feature classes on either side of the border; a trail on one side of the border, a road on the other. These could still be joined at a common point, so long as they are not merged. A particular problem arises where a river that one country represents as a line feature, is, on the other side of the boundary, an elongated polygon belonging to the inland water bodies class. How can a line be joined to a polygon?

5. ANALYSIS AND CONCLUSIONS

Challenges for Data Integration in Latin America

Data integration of national datasets, even those which are theoretically based on a common set of technical specifications – the ISCGM spec – can be achieved but the result does not necessarily produce a homogenous map comparable with the EuroGlobalMap. The cooperation agreements achieved in Latin America within the PAIGH framework are just that, agreements in an alliance of existing national mapping agencies rather than a real infrastructure such as that which Europe possess in the form of the Eurogeographics agency and, more recently, the INSPIRE framework. The funding from PAIGH is of a size suitable for science research projects and trials rather than for concrete regional SDI-building.

Place names are a major issue in many cartographic projects. Possibly future versions of the ISCGM specifications should give more attention to the issue of character sets and languages for names.

Cross-Boundary Unions

The union of objects across boundaries is possible for a single team working on the whole region, but even with the aid of the buffer and center line the final stage still requires human (sometimes subjective) criteria to be applied; this in turn implies an enormous task to unite the thousands of trans-boundary objects (mostly watercourses, many roads, a smaller number of lake polygon edges), with a consequent drain on resources and moreover an increased risk of human error.

Uniting objects across boundaries would be desirable for improving Global Map quality and its usefulness to the intended users; however, if Global Map or any regional map derived from it is used merely for visual display at its intended small scale, most of the gaps and interruptions at the boundary are invisibly small.

International Boundary Issues

To represent international boundaries on a continental map related to the global Map, it is strongly recommended to keep the double line solution. Cartographic purists might object to the idea of divergent interpretations of a single object – an international boundary – co-existing in the same map. However, there are four reasons for doing this:

(a) Most state mapping agencies in Latin America work within a political and legal environment that is very sensitive to the issue of location of international boundaries, obliging surveyors and cartographers to subordinate technical/geospatial criteria to the requirements of national interests and legal requirements for the location and portrayal of international boundaries. For this reason, any international data integration must be careful to leave intact the boundary representations as provided by the national agencies, in order not to forfeit the cooperation of the participating nations.

(b) The differing source scales and data density of the source material found in the national datasets would make it hard to reconcile the line features even in purely technical terms

(c) The Global Map scale, 1:1,000,000 will make the double nature of the line practically invisible for all but the widest divergences of boundaries, so long as users display it at the intended nominal scale without zooming in.

(d) For their practical applications, users may sometimes need to know both versions of the border according to the governments and mapping agencies each side of the frontier.

6. FUTURE PLANS

The next step for the South America data is to complete the block by integrating the data from the rest of this continent. Global Map data for Ecuador, Bolivia, Peru and Venezuela does exist, but as it has not yet been published by ISCGM special permission needs to be obtained to use it. The main Paraguayan NMO has had contact with ISCGM but may not yet have a suitable digital map to use. For Guyana and Surinam, probably alternative sources of data will need to be substituted for data from these countries. A new Technical Cooperation Project has been proposed to the PAIGH and approved for performance during 2011 specifically for the purpose of completing the South America block.

Meanwhile, there are also plans within the MGA Work Group to publish the data integrated so far. Once placed on line, the idea is that the dataset created in the work done so far will attract comments and suggestions for improvements; it will also show the five author institutions how their data looks once integrated and enable their participation in further improvements.

7. REFERENCES

Conference papers and Journal Articles:

- “European geographic data infrastructures: The EuroRegionalMap, EuroGlobalMap projects, and a strategy towards a European Reference data Infrastructure” JP Lagrange, N. Delattre, H. Ursin, C. Luzet, September 2002. GSDI-6 conference.
- “New Directions in Applications for the Global Map” Edwin Hunt, November 2006. GSDI-9 conference.
- “Applications of the Global Map of the Americas” Jean W. Parcher, Anna Lucia Barreto de Freitas, Zachary D. Wilson, Robin H. Gary, 2007. Revista Cartografica Nº 83 - PAIGH
- “Integration of Vector Datasets From Central And South American Nations”, Edwin Hunt, Armando Lobos, June 2009. GSDI-11 conference.

Books:

- “Global Map Specifications Version 1.2”, International Steering Committee for Global Mapping, April 2005.
- “Global Map Specifications Version 2”, International Steering Committee for Global Mapping, October 2009.

Web Sites

- ISCGM: www.iscgm.org
- PAIGH: www.paigh.org.mx
- EuroGlobalMap: www.eurogeographics.org/content/euroglobalmap