Open Geospatial Data: An Assessment of Global Boundary Datasets

Charles Brigham, Steven Gilbert, and Qiyang Xu

World Bank Institute

1818 H St. NW, Washington, DC

cbrigham@worldbank.org
sgilbert@worldbank.org
qxu1@worldbank.org

1. Introduction to GAUL, GADM, and UNSALB

For the World Bank Annual Meeting on September 26th 2011, the Mapping for Results (M4R) team mapped project locations and development indicators for 142 countries and made these maps and data available through the World Bank's external website (The World Bank, 2011). M4R used publicly available administrative boundaries in order to identify the ADM1[1]and ADM2 in which each project activity is located and to map subnational indicator data. This paper uses our insight gained from these activities to analyze the accuracy in terms of whether boundaries are up to date and the usability of publicly available boundary datasets for mapping on a global scale.

1.2 GAUL

The 2009 edition of Global Administrative Unit Layers (GAUL) is a joint initiative of the Food and Agriculture Organization of the United Nations and the European Commission. As of 2009, the GAUL dataset contains national administrative boundary information of 275 countries and territories, among them thirty-eight countries in Table 1 are copyrighted and published by Second Administrative Level Boundaries (UNSALB) (Food and Agriculture Organization 2008).

Table 1. Countries copyrighted and published by UNSALB used within the GAUL dataset (FAO Geonetwork)

Australia	Belgium	Belize	Bolivia	Botswana
Cambodia	Cameroon	Democratic Republic of the Congo	France	Gabon
Germany	Ghana	Guinea	Guinea-Bissau	Jamaica
Liechtenstein	Luxembourg	Madagascar	Malawi	Maldives
Mali	Malta	Mozambique	Namibia	Nepal
Niger	Norway	Portugal	Saint Kitts and Nevis	Saint Lucia

[1] ADM1 signifies the first-order administrative boundaries such as a state, province, or department. ADM2 signifies second-administrative level boundaries such as district, municipality or county.

Sierra Leone	South Africa	Sri Lanka	Timor-Leste	Togo
Trinidad and Tobago	United States of America	Zambia		

1.3 GADM

Global Administrative Areas (GADM) is a global spatial database of administrative boundaries, including national and lower level administrative areas with boundary information for 217 countries. The GADM database was developed in a partnership between the University of California, Berkeley, Museum of Vertebrate Zoology, and the International Rice Research Institute (Global Administrative Areas 2009).

1.4 UNSALB

In 2001, UNSALB was launched as a working platform for the interoperational global boundary dataset of United Nation Geographic Information Group (UNGIWG). UNSALB contains administrative boundaries for 98 countries with 50 still in the process of validation. These boundaries are officially approved by National Mapping Agencies (NMA) of UN member countries (UN Geographic Information Working Group 2011).

2. Accuracy of Boundary Datasets

Accuracy and precision are of primary concern in data representation, processing and geovisualization. In most circumstances, the number of administrative divisions of a country can be found in official country documents and while there is a lack of official up-to-date information on actual shape and boundary of those administrative divisions. For this reason as well as potential boundary issues with disputed areas between regions and countries, our analysis focused mainly on whether boundaries reflected the correct administrative divisions in a country. As a reference for determining the accuracy of each boundary dataset, we used lists of administrative divisions available through statistics bureaus where available or UNSALB's list of historic changes and codes. For the 142 countries analyzed, Figure 1 below shows the number of countries with correct boundaries by publicly available dataset.

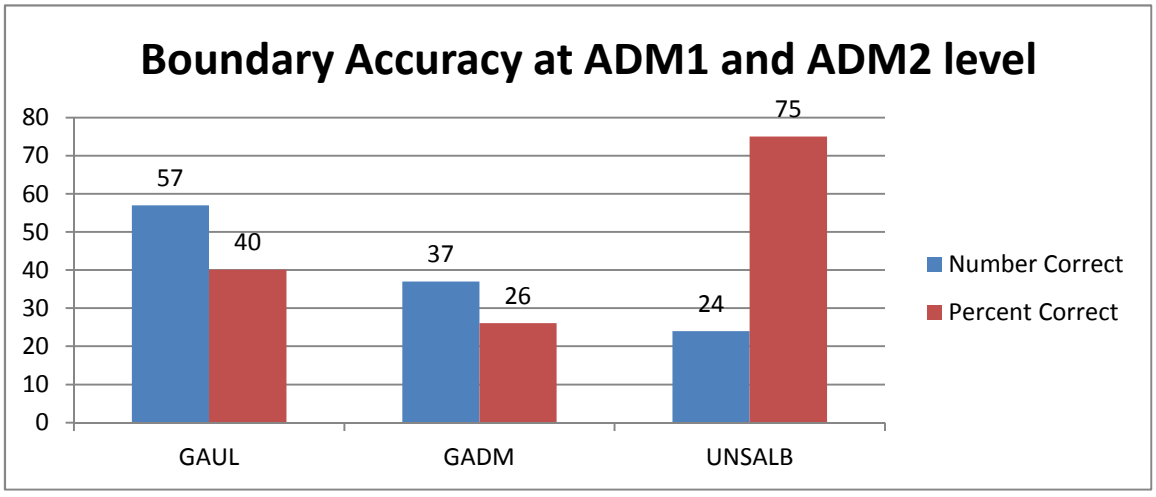

Figure 1. Accuracy of Administrative Boundary Datasets

In terms of accuracy UNSALB consistently contains the most accurate boundaries, most likely due to an update process that involves consultation with national mapping agencies to ensure accuracy. Unfortunately, this strength also inhibits its ability to respond quickly to changes in administrative divisions and to develop boundaries for many countries. During the use of UNSALB by M4R, 75 percent of the boundaries available were found to have the correct administrative divisions, however because UNSALB does not cover many countries, it has fewer correct boundaries than GADM and GAUL. While UNSALB is supposed to represent the ADM2 boundaries of each country, for two countries, Madagascar and Belize, only ADM1 boundaries were included.

The second most accurate of the three boundary sets was GAUL. One obvious reason for this was the inclusion of UNSALB boundaries within the GAUL dataset. However, it should be noted that the GAUL dataset is only updated annually and does not always include the newest boundaries available from UNSALB. For example, Iran GAUL boundaries do not yet include boundaries from UNSALB. Renewal of GAUL dataset as a whole, instead of each individual boundary dataset, may fail to keep pace with frequent changes in ADM boundaries. For example, Rangpur, as a new ADM1 in Bangladesh, is not included in GAUL.

GADM was found to be the least accurate boundary dataset. While GADM lacks the official validation of respective governments, its ability to respond to public feedback makes it possible to rectify problems quickly. However examination of the documentation for the GADM countries indicates little public feedback on data accuracy and no indication whether any provided feedback was incorporated. For example, a user provided feedback that GADM's boundaries for Afghanistan are missing two provinces, Daykundi and Panjshir, and only contained 326 of the country's 398 districts yet these new provinces and districts have not been added to the GADM boundaries. It should be noted that for several countries, we found that the boundaries for GADM contained correct administrative divisions whereas GAUL boundaries were incomplete. For example while GAUL misses most of ADM2 or prefecture level divisions for China, GADM has the correct divisions. Table 2 shows a list of countries where the GADM boundary set contained better boundaries than GAUL.

Table 2. Countries in which GADM boundaries are more accurate than GAUL

At ADM1 level	*At ADM2 Level*
Albania	Angola
Angola	China
Cape Verde	Honduras
Eritrea	Kosovo
Honduras	Macedonia, FYR
Iran, Islamic Rep.	Nigeria
Korea, Rep.	Tajikistan
Kosovo	Tanzania
Mongolia	Thailand
Myanmar	Uganda
Panama	Ukraine
Paraguay	Yemen, Republic of

Samoa	
Tanzania	
Tonga	
Yemen, Republic of	

Given that the much of the World Bank's projects are located in Africa and Latin America, a regional perspective of boundary accuracy is very useful for development practitioners. In our analysis of boundary accuracy we compared the accuracy of each boundary set from a regional perspective, using the six regions recognized by the World Bank (Figure 2)

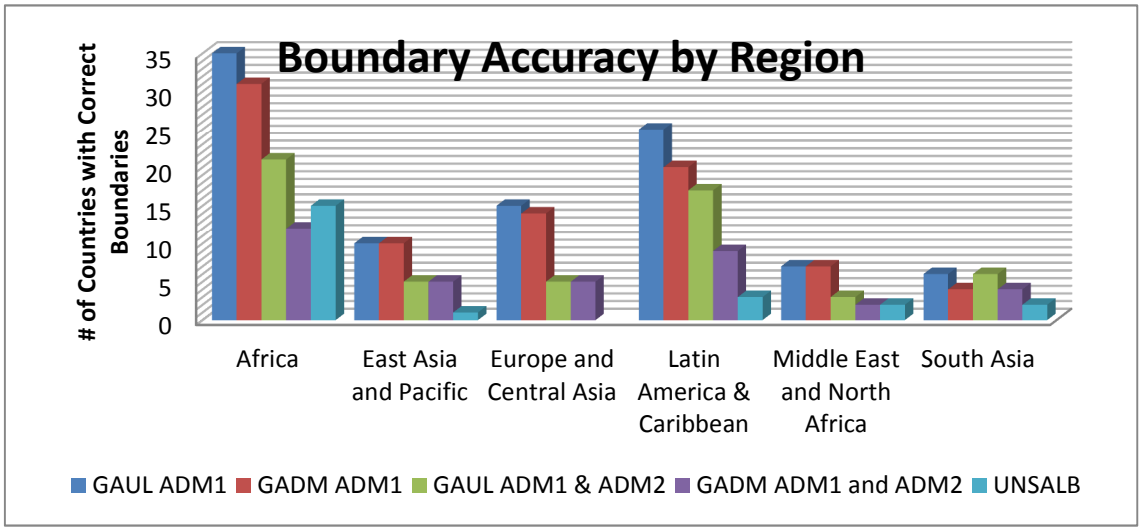

Figure 2. Accuracy of Subnational Boundaries by Region

We found the most profound differences in regional accuracy to be in Latin America and Africa. GAUL was the most accurate in terms of number of countries at both the ADM1 and ADM2 levels in Africa and Latin America & Caribbean regions. In Africa, we found that for 21 countries GAUL had correct boundaries at the ADM1 and ADM2 level, compared with 15 for UNSALB and 12 for GADM. While UNSALB accuracy is highest in Africa, it does not cover many of the countries in the region at all. In Latin America the discrepancy between boundary sets was even greater. GAUL contained correct boundaries for 17 countries, compared to only 9 correct boundaries in GADM and 3 for UNSALB.

For 28 countries in Table 3 none of the publicly available boundary sets had correct boundaries at the ADM1 or ADM2 level. This was often the case with island countries such as Kiribati and the Solomon Islands and with some African countries including Chad, Rwanda, and Sudan.

Table 3. Countries with no accurate boundaries

Afghanistan	Kazakhstan	Moldova	Solomon Islands
Bangladesh	Kiribati	Morocco	St. Lucia
Chad	Kyrgyz Republic	Palau	Sudan
Chile	Libya	Romania	Turkey

Congo, Republic of	Macedonia, FYR	Rwanda	Turkmenistan
Djibouti	Malaysia	Senegal	Uganda
Egypt, Arab Rep.	Mauritius	Seychelles	Vietnam

3. Usability

In addition to boundary accuracy, experience using these boundary datasets provides insight into the usability of each boundary dataset. We found GAUL to be the most usable for mapping on a global scale. Its combination of accuracy and completeness makes it more usable than GADM or UNSALB. A particularly useful aspect of GAUL is the inclusion of unincorporated territories and disputed areas on an individual national-level. For example disputed territories such as Aksai Chin and Arunachal Pradesh on the border between China and India are not excluded from the dataset, but rather included as their own territories giving the user the ability to determine whether to include them or not.

The GADM boundary set is available on a per country basis and as a complete world set without registration, making it easily accessible, but several factors make GADM difficult to use. Unlike the UNSALB and GAUL datasets, historical boundaries are not available through GADM. Additionally, GADM's usability is hindered by inconsistent naming conventions for its administrative divisions that often contrast with official government sources. For example, in Turkmenistan several of the ADM1s are named incorrectly and administrative division names in Albania, Chad, and Uruguay use accents and characters that make joining and processing data difficult. Finally, since the GADM dataset does not support multi-part polygons for countries and administrative divisions consisting of islands or separate areas it make joining disparate sources and geo-processing tasks such as area calculation, and division-based indicator data analysis more difficult to perform.

While UNSALB usually has the most accurate boundaries, it has several pitfalls that affect usability. First, the UNSALB website only provides ADM2 boundaries downloadable on a per country basis, which is inconvenient when boundaries are needed on a global scale. Additionally, while the process of official validation by member countries ensures that boundaries are accurate, it limits the geographic scope of UNSALB, since boundaries for many countries have not yet been validated, are pending, or simply have never been captured.

4. Conclusion

Through comparison of GAUL, GADM and UNSALB boundary datasets we found that each dataset has advantages and drawbacks in terms of accuracy and usability, but overall GAUL was the best dataset due to the accuracy and completeness of the dataset. While UNSALB boundaries have the highest rate of accuracy because of validation with national mapping agencies, it is limited in geographic scope. Although GADM has a global scale, many of the boundaries are outdated and it is unclear whether GADM organizers have utilized public feedback to update boundary sets GADM. For local level mapping exercises boundary selection should be selected on a case by case basis, but given GAUL's superior accuracy and completeness, in particular in the developing countries of Africa and Latin America, makes it best suited for mapping subnational data on a global scale.

5. References

Global Administrative Areas (GADM), 2009. Available From:
<http://www.gadm.org/about>. [21 November 2011]

Food and Agriculture Organization, 2008. *Global Administrative Unit Layers (GAUL).* Available From:
<http://www.fao.org/geonetwork/srv/en/metadata.show?id=12691>. [21 November 2011]

UN Geographic Information Working Group, 2011. *SALB: Second Level Administrative Boundaries*. Available From:
<http://www.unsalb.org/>. [21 November 2011]

The World Bank, 2011. *Mapping for Results.* Available From:
<http://maps.worldbank.org/> [21 November 2011]

6. Biographies

Charles Brigham has worked in Geography for institutions such as NASA, UN, World Bank, CARICOM, and for many countries. Charles' expertise is in International Development, Census and Statistics, Geospatial Infrastructure, Innovative Geospatial technologies, Imagery Analysis, and Humanitarian affairs. He currently works for the World Bank Institute, Innovation Practice

Steven Gilbert has worked as a GIS and data analyst with the Mapping for Results initiative in the Innovation Practice of the World Bank Institute since May 2010. Prior to joining the World Bank Institute, Steve earned his master's degree in Latin American Studies from Georgetown University.

Qiyang Xu is a GIS Consultant with the Sustainable Development Network Information Services unit at the World Bank. She holds a Master of Arts degree in Geography from University of Maryland, and a Bachelor of Science in Geographic Information Systems from the Zhejiang University, China