Contract Report: CR-2005/05 SAC
2005.04.07

Guidelines for Data Content Standards for Africa

Authors:
Antony K Cooper
Bongani Majeke

PREPARED BY:
CSIR
Centre for Logistics and Decision Support (CLDS)
Satellite Applications Centre (SAC)
Unit of Water, Environment and Forestry Technology (Environmentek)
PO Box 395
PRETORIA
0001
Tel: +27 12 841 4121
Fax+27 12 841 3037

This project was initiated by the EROS Data Center of the United States Geological Survey (USGS/EROS) and EIS-Africa, with funding from the United States Agency for International Development (US AID) and the CSIR.

<table>
<tr><td colspan="2">DOCUMENT RETRIEVAL PAGE</td><td>Report No:
: CR-2005/05 SAC</td></tr>
<tr><td colspan="3">Title: Guidelines for Data Content Standards for Africa</td></tr>
<tr><td colspan="3">Authors: Antony K Cooper and Bongani Majeke</td></tr>
<tr><td>Client:
USGS/EROS and EIS-Africa</td><td>Date:
2005.04.07</td><td>Distribution:
Public</td></tr>
<tr><td>Project No:
KSDSE00</td><td>OE2:
05300</td><td>ISBN:</td></tr>
<tr><td colspan="3">Abstract:

This project was to develop guidelines for spatial data content standards in Africa. It focused on standards for data dictionaries, feature catalogues and classification, and also looked at standards for feature instances. The project deliberately did not address metadata, as much work has already been done on metadata in Africa.

With funding from the United States Agency for International Development (US AID), the project was initiated by the EROS Data Center of the United States Geological Survey (USGS/EROS) with EIS-Africa, and executed by the CSIR.</td></tr>
<tr><td colspan="3">Keywords:
Standard, Data content, Geographical information, Data dictionary, Feature catalogue, Classification, Feature instance, Africa</td></tr>
<tr><td colspan="3">Related documents:</td></tr>
<tr><td>Signatures:

Andrew Terhorst
Programme Manager</td><td>

Engela van Heerden
Info Centre</td><td>

Mothibi Ramusi
Centre Manager</td></tr>
<tr><td colspan="3">NOTE: This document is for public distribution.</td></tr>
</table>

EXECUTIVE SUMMARY

With funding from the United States Agency for International Development (US AID), the EROS Data Center of the United States Geological Survey (USGS/EROS) initiated a project with EIS-Africa to develop guidelines for spatial data content standards in Africa. The project was executed by the CSIR (which has contributed additional funding), and completed during March 2005.

In general, data content standards include:

- Documentation specifying the information in a data set, such as:
  - Data dictionaries, feature catalogues and classification (feature types, attribute types, attribute domain, feature relationships)
  - Feature instances (unique, definitive versions of features)
  - Metadata (including data quality)
  - Data organization (eg: XML, GML)
  - Reference models

- Formal description of a model, for example using UML. Such formal descriptions are (hopefully) embedded in the data content standards.

With the short time limits for this project and its small size, the project focused on standards for data dictionaries, feature catalogues and classification, and data organisation for them, using automated tools where possible. The project also looked at standards for feature instances. The project deliberately did not address metadata, as much work has already been done on metadata in Africa.

We identified and assessed 170 candidate standards from various sources (eg: ISO/TC 211, Open Geospatial Consortium, Federal Geographic Data Committee, Zimbabwe and Standards South Africa). From these we selected several to compare their feature types (classes). The project included developing guidelines for data content standards in Africa.

TABLE OF CONTENTS

Guidelines for data content standards for Africa

30 March 2005

1. Introduction to the project

Early in 2004, the EROS Data Center of the United States Geological Survey (USGS/EROS) initiated a project with EIS-Africa to develop guidelines for data content standards for geospatial data in Africa. The project was funded by the United States Agency for International Development (US AID). The project was awarded to the CSIR (which has contributed additional funding for the project), and was executed jointly by three of its units: the Satellite Applications Centre (SAC), the Centre for Logistics and Decision Support (CLDS), and the Unit of Water, Environment and Forestry Technology (Environmentek).

In general, data content standards include:

- Documentation specifying the information in a data set, such as:
  - Data dictionaries, feature catalogues and classification (feature types, attribute types, attribute domains, and feature relationships);
  - Feature instances (unique, definitive versions of features);
  - Metadata (including data quality);
  - Data organization (eg: using the eXtensible Markup Language (XML) [XML 2004] or the Geography Markup Language (GML) [ISO 19136]); and
  - Reference models.
- Formal description of a model, for example using the Unified Modelling Language (UML) [UML 2003]. Such formal descriptions are (hopefully) embedded in the data content standards.

There are already many initiatives addressing metadata for geospatial data in Africa, and with the short duration of this project and its small size, its sponsors felt that the focus should be on other data content standards for geospatial data. This project focuses primarily on standards for data dictionaries, feature catalogues and classification, and the data organisation for them, using automated tools where possible. As appropriate, the project will also investigate standards for feature instances.

To disseminate information about the project, we have used the *SDI Africa Discussion List* of the Global Spatial Data Infrastructure Association (GSDI), and we would like to thank GSDI for allowing us to do so. See: http://www.gsdi.org/gsdiForums/forum_topics.asp?FID=8. We have also used the EIS Africa mailing list to promote the project, as well as the list of delegates at the 5th African Association of Remote Sensing of the Environment (AARSE) Conference, held in Nairobi, Kenya, in October 2004. Details of the project have been included in presentations made at the following meetings:

- FAGIS/GISSA (the Free State branch of the Geo-Information Society of South Africa) Regional Conference, held in Bloemfontein, South Africa, on Thursday 26 August 2004;
- AARSE Conference, in a pre-conference workshop on standards on Sunday 17 October 2004 and in a plenary session on standards on Thursday 21 October 2004 [Cooper 2004a, 2004b];
- Global Land Cover Network (GLCN) Workshop held by the Food and Agriculture Organisation (FAO) in Pretoria, South Africa, on Monday 13 December 2004 [Cooper 2004c]; and

- Africa Water Information Clearing House (AWICH) Workshop on Geo-Water Information Development and Management, held by the United Nations Economic Commission for Africa (UN ECA) in Pretoria, South Africa, on Thursday 3 March 2005 [Cooper 2005].

The results of this project will provide useful inputs for the Working Group on Standards, established by the Executive Working Group of the Geoinformation Sub Committee of the Committee on Development Information (EWG-CODI-Geo) of the UN ECA in August 2004. We also plan to make a presentation of the results at the next CODI Geo meeting, in Addis Ababa, Ethiopia, in April 2005.

2. Introduction to data content standards

An expansion of formal definitions for *data content standard*, *application schema*, *conceptual schema*, and *universe of discourse* [GOS 2004], provides the following definition of a data content standard:

> *A standard that specifies what information is contained within a geospatial data set, and provides a formal description of a model that defines the concepts of a view of the real or hypothetical world that includes everything of interest, for data required by one or more applications.*

From this definition, data content standards include:

- **Documentation specifying the information in a data set**. This includes:

  - **Data dictionaries, feature catalogues and classification**. These define the types of geographical features (ie: the *classes* or *feature types*) one would find in a data set, together with their *attributes* (types and domains) and other peculiarities, enabling users to have a shared understanding of the contents of the data set. They would also include the conceptual relationships between feature types (eg: that an instance of the feature type '*bridge*' can carry an instance of the feature type '*road*' over an instance of the feature type '*river*'). It is critical that data dictionaries, feature catalogues and classifications include proper definitions to differentiate between their feature types, and not depend merely on the label (name) attached to the feature type. A *data dictionary* is an unstructured collection of feature types, while a *feature catalogue* is a structured collection of feature types (eg: as a hierarchical classification), and hence easier to use. For our purposes, the terms *feature catalogue* and *classification* are synonyms. Typically, a feature catalogue would be constructed from a data dictionary, perhaps as a *profile* (subset) of the data dictionary.

  - **Feature instances**. The instance of a feature in a data set represents a discrete phenomenon in the real (or imaginary) world – that is, something specific out there that is modelled in the data set. The instance normally has coordinates and may be portrayed on a map by a particular graphic symbol. Standards for feature instances specify the unique, definitive versions of features (eg: SANS 1876 *Feature instance identification standard* – however, this was the only standard of this type that we were able to find).

  - **Metadata**. "*Data about data*", metadata includes the various types of information that describe the structure and content of a data set, and how to access and use the data. Metadata encompasses issues such as data quality (positional accuracy, attribute accuracy, completeness, logical consistency, currency and lineage), data schemas, spatial referencing systems, constraints on the use of data, and contact details for those responsible for the data set. Metadata enables data discovery, determining fitness for use, data access and data transfer.

- **Spatial representation**. This addresses how locations are represented, either through coordinates (given within the context of a spatial referencing system) or geographical identifiers (eg: the name of a town).

- **Data organization**. This encompasses the logical description of the data set, using formal languages such as the eXtensible Markup Language (XML) or its implementation for geographical information, ISO 19136, *Geography Markup Language* (GML).

- **Reference models**. These provide conceptual descriptions of data sets, using standards such as ISO/IEC 10746, *Open Distributed Processing – Reference model* (RM-ODP) with its five Viewpoints (Enterprise, Information, Computation, Engineering and Technological).

o **Formal description of a model**, for example using the Unified Modelling Language (UML). Such formal descriptions are (hopefully) embedded in the data content standards and hence for the purposes of this project, we need only record their presence or absence and use them to understand the content of the standards, rather than analyse the models per se.

It was not realistic or necessary to include all of these components of data content standards in this project, given the short duration of this project and its small size, as well as the differing data regimes across Africa, national security concerns regarding access to data, and other issues. There are already many initiatives addressing metadata for geospatial data in Africa, such as the USGS/EROS's Environmental Monitoring and Information System (EMIS) Project for Africa, and the metadata training workshops arranged by EIS-Africa and others. Hence, the project's sponsors felt that the focus should be on other data content standards for geospatial data. This project focuses primarily on standards for data dictionaries, feature catalogues and classification, and the data organisation for them, using automated tools where possible. As appropriate, the project will also investigate standards for feature instances.

3. User requirements

Unfortunately, while we have had a very good response to the project at the meetings where it has been presented, only a few people across Africa have asked to be put on the project's mailing list and to date, we have not received any input from them on the user requirements for data content standards. Hence, these are based on our own experiences.

The standards in general are so important not only in geographical information systems (GIS), but in all aspects of life. It is imperative that there must be consistency in the way things are done globally: that's primarily the reason why there are standards that guide us. Hence, this project was initiated, not only for Africa, but also for the entire world to benefit. One can imagine how life can be if each country or continent is doing things its own way with no standards that have to be met.

Referring to land cover mapping as an example, which is being undertaken by many countries, the Land Cover Classification System (LCCS) standard [Di Gregorio & Jansen 2000], developed by the Food and Agriculture Organization of the United Nations (FAO), is being adopted by many land cover and land use classifiers as the guide in classification of the land cover maps. LCCS provides a comprehensive data dictionary of feature types for land cover, and provides the flexibility to customise the structure of one's land cover classification to suit one's specific requirements, without compromising the ability to exchange the data with others using a differently structured land cover classification.

Such flexibility and interoperability are key requirements for data content standards. Other requirements include: complying with any policies and standards that have been adopted locally and globally; the accuracy of the data; understanding the limitations of the data; fulfilling any reporting mechanisms; and being understandable and accessible (low barriers to entry).

4. Possible data content standards

4.1. Standards generating bodies

There are many standards generating bodies around the world, some of which have an international scope while others have a more localised scope; some are open organisations developing standards through consensus, while others are subscription-based industry standards groups. The two largest open standards generating bodies are the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC), who develop standards through their Technical Committees (TCs) and their Joint Technical Committee (JTC 1) for Information Technology.

Of particular relevance to the geographical information community in Africa are the following standards generating bodies:

- *ISO/TC 211*, *Geographic Information/Geomatics*, which has developed the ISO 19100 series of standards for digital geographical information, such as ISO 19115:2003, *Geographic information – Metadata*. See: http://www.isotc211.org/

- *Open Geospatial Consortium, Inc (OGC)*, an industry standards group that works very closely with ISO/TC 211. OGC's focus is on creating open and extensible software application programming interfaces for geographic information systems and other mainstream technologies. They make their adopted specifications available to the public at no cost. See: http://www.opengeospatial.org/

- *Federal Geographic Data Committee (FGDC)*, which was established to promote the coordinated use, sharing and dissemination of geospatial data across the United States of America. It developed standards that have been widely used around the world, some of which have been the pre-cursors to some ISO 19100 standards. The prime example is the Content Standard for Digital Geospatial Metadata (CSDGM). See: http://www.fgdc.gov/

- At the continental level, the *African Regional Organization for Standardization (ARSO)* is meant to promote standards and coordinate standards efforts, but it would appear that it is not very active at this stage, and that it is not considering standards for geographical information. See: http://www.arso-oran.org/

- The *Southern African Development Community (SADC) Cooperation in Standardization (SADCSTAN)* appears to be the only regional standards generating body in Africa active in standards for geographical information. See: http://www.sadcstan.co.za/. SADCSTAN was established to:
  - Promote regional cooperation in the development of harmonized standards and technical regulations;
  - Facilitate the exchange of information on existing standards, draft standards and technical regulations among members; and
  - Facilitate the adoption of regional standards by member states [SADCSTAN 2005].

  The SADCSTAN standards for geographical information that are currently under consideration are those from Standards South Africa's SC 71E (see below), which were placed on SADCSTAN's list of projects for harmonization in April 2004.

At the national level, there are several African countries that are active, or are becoming active, in the development of standards for geographical information, with Kenya, Mauritius, Morocco, South Africa, Tanzania and Zimbabwe being members of ISO/TC 211. To date though, only South Africa has been active in the work of ISO/TC 211.

- South African National Standards (SANS) are the responsibility of Standards South Africa, which overprints many ISO and IEC standards, and also develops local standards and profiles of international standards, through its 450+ TCs and SCs. See: http://www.stansa.co.za/. Its committees include:
  - TC 71, Information Technology, the local committee for ISO/IEC JTC 1, but it also has several SCs that are the local committees for other ISO TCs, including:
    - SC 71E, Geographic Information, the local committee for ISO/TC 211, which has four data content standards: SANS 1876, *Feature instance identification standard* (Unique identifiers of feature instances in core data sets); SANS 1877:2004, *A Standard land-cover classification scheme for remote sensing applications in South Africa* (Implementation of FAO's LCCS); SANS 1878, *South African spatial metadata standard* (Profile of ISO 19115); and SANS 1880, *South African Geospatial Data Dictionary (SAGDaD) and Its Application* (Implementation of ISO 19110).

- In Zimbabwe, the Department of the Surveyor General is leading the development of the Zimbabwean National Standards for Digital Geographical Information [DSG 2003]. The emphasis of the standards is on data models, classification, metadata and quality, but they also address several other supporting standards, such as spatial reference systems and data transfer rules.

4.2. The data content standards considered

USGS/EROS provided the project team with a collection of candidate data content standards to assess from various standards generating bodies. In addition, we considered standards being developed by organisations in which the project's team members are active, and looked for sources for such standards on the Web. Standards from the following were considered:

- ISO/TC 211;
- ISO/IEC JTC 1;
- The Open Geospatial Consortium (OGC);

- The United States of America's Federal Geographic Data Committee (FGDC);
- Standards South Africa (StanSA);
- The Zimbabwean Department of the Surveyor-General;
- The Digital Geographic Information Working Group (DGIWG);
- Australia (including Standards Australia and regional standards generating bodies such as the Queensland Spatial Information Infrastructure Council);
- GeoBase;
- Japan;
- Land Information New Zealand;
- The standards generating bodies of some States in the USA (eg: Rhode Island, Indiana and Pennsylvania);
- Property Information Systems Common Exchange Standard (PISCES) and Open Standards Consortium for Real Estate (OSCRE).

We have attempted to look for candidate standards in French, but the only ones that we have been able to identify are those standards developed by ISO/TC 211 that Canada has translated into French, such as ISO 19115, *Geographic information – Metadata (Information géographique – Métadonnées)*.

We have compiled a spreadsheet of the 170 standards we identified (included in Annex A to this document), recording for each standard the following:

- The name of the standard;
- The source of the standard (ie: the relevant standards generating body);
- The official identifier for the standard (if we could find one);
- The date published (if provided);
- The status of the standard (published, draft, etc, if we could determine it);
- Whether or not it is a data content standard;
- Whether or not it is a metadata standard (and hence out of scope for this project);
- Whether or not the standard is relevant for the project (ie: defines a data dictionary, feature catalogue or classification);
- For a few standards, the reasons for relevancy or irrelevancy;
- How accessible the standard is to the public; and
- For a few standards, notes about the standard.

Some of the standards were available only at a price and to access others one needs to subscribe to the standards which in some cases is a long and complex process – as such standards are then not widely available, we did not consider them to be suitable as candidate standards. From the list of standards that were available the ones that were relevant for the project were chosen for detailed comparison with a reference standard.

5. Comparison of feature types

5.1. A reference standard

To facilitate assessing the contents of standards for data dictionaries, feature catalogues and classification schemes, we selected SANS 1880, the *South African Geospatial Data Dictionary (SAGDaD)*, as a reference standard, largely because of our personal experience with it. SAGDaD is the South African implementation of ISO 19110, *Geographic information – Methodology for* *feature cataloguing*. SAGDaD focuses on the core data types likely to be transferred between users, and currently defines 81 feature types and 29 attributes, with the attributes being defined uniquely across all the feature types. One of these attributes is *Enumerated Type*, which refines a feature instance's feature type to provide subtypes or subclasses – the brief for the design of SAGDaD was to use as few feature types as possible.

5.2. The standards compared

The process of comparing the feature types from these standards was more difficult than we had anticipated at the beginning of the project. This is because different countries use terminology differently, and have different legal frameworks and other factors that influence the labels they apply to feature types and the ways they structure their feature types into feature catalogues. The result is that the same feature type is represented differently in different catalogues. Not all the standards had definitions readily available for their feature types. It was also not feasible to trawl through in fine detail all those definitions that were available, to determine the subtle nuances used to create each feature type.

The classification of digital geographical information is a subjective process because people observe different properties in features and require information about the features to different levels of detail [Scheepers *et al* 1986]. There is also a grey area between feature types and attributes: what constitutes a feature type in one data dictionary or feature catalogue, could constitute an attribute of a feature type in another. For example, the data dictionary SANS 1880 has the feature type Measurement Sample Point, for which one of its values for the *Enumerated Type* attribute is *Tide Gauge* (defined as "*a point where the extremities or present level of the tide is measured*"), while a specialist hydrographic feature catalogue might have a feature type exclusively for tide gauges.

Finally, some of the data dictionaries and feature catalogues cover the whole gamut of digital geographical information (though perhaps focusing only on the core feature types), while others focus on a narrow aspect of digital geographical information (though providing feature types to a greater level of detail). Nevertheless, we believe that the resultant comparison is useful, highlighting core feature types.

The feature types of SAGDaD and of the chosen standards were compared in a spreadsheet (included in Annex B to this document), whereby similar feature types from different standards were put in the same row. The following standards were selected for comparison with SANS 1880:

- FGDC Cadastral Data Content Standard, FGDC-STD-003
- FGDC Utilities Data Content Standard, FGDC-STD-010
- FGDC Hydrographic Data Content Standard for Coastal and Inland Waterways
- Zimbabwe Cadastre
- Zimbabwe Transport
- Zimbabwe Building
- Zimbabwe Utilities
- Zimbabwe Land description
- Zimbabwe Administration
- Zimbabwe Water
- Zimbabwe Heights
- Zimbabwe Control Points
- Zimbabwe Annotation and Text

Some of the feature types from the candidate standards matched up well with the corresponding feature types in SANS 1880, but there were those that did not. SANS 1880 provides a general purpose data dictionary principally for core data sets, while some of the others provide detailed feature types for their particular application domain – their detailed feature types that could not be matched to feature types in SANS 1880 have been shown at the end of the spreadsheet. SANS 1880 also contains refinements of its feature types through its *Enumerated Type* attribute, but these attribute values have not been included in the integrated feature catalogue. It is probable that many of those unmatched feature types in the other standards would match one of these feature type and attribute combinations in SANS 1880.

The following table provides a few examples of the different naming of feature types between different standards:

SANS 1880	**FGDC Cadastral**	**FGDC Hydrographic**	**Zimbabwe**
Isoline		*Depth contour*	*Height contour*
Administrative Area		*Administration Area*	*Adm Area*
River		*River*	*River*
Cadastral Property	*Parcel*		*Property Parcel*

6. Advice on data content standards

6.1. Collection criteria for feature types

ISO 19110, *Geographic information – Methodology for feature cataloguing*, does not contain a feature catalogue *per se*; it specifies the structure of a compliant feature catalogue. Unfortunately, it does not address the collection criteria for compiling a feature catalogue – that is, it does not provide any requirements or guidelines for how to identify and define a “good” feature type or feature catalogue. The new version of ISO 19126, *Geographic information – Feature data dictionaries, feature catalogues and registers*, will most likely use a profile (subset) of ISO 19110’s conceptual model for feature catalogues, define the conceptual model for a data dictionary, and specify how these form the basis for establishing and managing registers for data dictionaries and feature catalogues. Nevertheless, it will also not address the collection criteria.

Collection criteria can be very complex and dependent on the field of application, as well as on one’s cultural, linguistic, legal and political framework. Hence, it is likely that one will have to deal with a variety of data dictionaries and feature catalogues, with translation tables between them. Unfortunately, it can be tedious to set up such translation tables. Normally, data will be transferred from a user that has deeper (or the same) knowledge of the data being transferred than the recipient has – an individual feature is more likely to be transferred from the user that would classify the feature more precisely [Cooper 1993]. Hence, such mappings are invariably one-to-one or many-to-one, and can be done automatically.

Unfortunately, it is far too common to categorise things for the wrong reasons – that is, use invalid collection criteria. Typical mistakes identified by Cooper [2003] include:

- **Using a quantitative measure to differentiate feature types**: a small change in the measure that then crosses a threshold will necessitate reclassifying a feature, such as when towns are classified on the basis of their population. Unless there are obvious breaks in the range of numeric values, these feature types could also appear to have artificial boundaries.

- **Overloading a feature type**: this is very common, and occurs when one feature type is used to convey several different meanings (often independent), which could cause confusion or which could render the feature type invalid when one has better knowledge. The identification of a feature type should be based on only the most important set of its characteristics.

- **Assuming there is only one valid categorisation**: rather than overloading a feature type, one should consider whether or not there are two or more perspectives of the same data, and develop a taxonomy for each. An example would be a post office, which one could view as belonging to the feature type '*building*' (for the physical building housing the post office), and one could view as being a '*post office*' (as part of the postal network and/or for the postal services provided), and one could view as being '*government*' (for land use zoning of the erf containing the post office).

- **Categorising the symptoms, not the causes**: it is easiest to start by categorising the effects (symptoms, or superficial aspects) of one's subject of interest, while it is much more useful to categorise based on the causes (fundamentals).

- **Making the categorisation dependent on its encoding**: this occurs typically with an hierarchical classification, where the number of feature types on a level is limited by the way those feature types are encoded (eg: by having a structured code with a single digit for the feature types at each level). There are many classifications that fall into this trap. Needless to say, this places an artificial limit on the number of feature types, or results in superfluous feature types being created. This often manifests itself in the 'need' to have a round number of feature types.

- **Assuming there is a perfect categorisation**: it is very easy to get into "analysis paralysis" trying to develop the perfect standard or taxonomy – and then once it is released into the real world, discover its shortcomings. Hence, one should expect to revise whatever categorisation is developed, especially based on feedback from users. It is best to test the proposed categorisation as quickly as possible in the real world.

6.2. Developing data content standards

African users of digital geographical information should not be mere passive receptors of technologies (such as standards) from other parts of the world, but should be active contributors to the development of such technologies. They need to play active roles in planning and developing standards, to ensure that the relevant standards are appropriate for African conditions (including being viable and affordable), and that they meet the needs of Africa.

A study initiated by the German Institute for Standardization (DIN), found that companies actively involved in standards development incur lower costs when the standards are implemented [DIN 2000]. They get early access to current technologies and thinking (insider knowledge), are able to assert their interests in the standardization process (getting desired contents included and undesired contents excluded), and lower the economic risk and costs of their research and development. Some key findings include:

- The savings from internal (company) standards are even greater than those from industry-wide standards;
- The positive effect of standards on communication between departments within the organisations was rated significantly higher than the effect on production costs;
- The longer the lifespan of the products, the greater the relevance of the standards, particularly industry-wide standards; and
- Using standards reduces liability risk.

Involvement in international standardisation efforts provides access to current technologies and thinking, and the possibilities to take technical or managerial leadership roles in national and international standards generating bodies. However, it can be expensive to attend the meetings of international standards generating bodies, particularly as they rarely meet in Africa (ISO/TC 211 met in Cape Town, South Africa, in March 2000), and funding is not always readily available. Indeed, it has not been uncommon for South Africans participating in ISO/TC 211 meetings to pay some of their own costs out of their own pockets.

Local standards need a massive investment to support their implementation, because of the small local market available to support the standard. Local vendors have limited resources, and don't necessarily have access to the source code for the products they sell, making it very difficult for them to get their products adapted to support local standards.

However, it is quite feasible to develop local profiles of international data content standards, as the products that support the international standard will automatically support the profile, as it is a straight subset. Local profiles make the international standards easier to use, as they cater for local conditions. It should also be feasible to develop local implementations of international data content standards, and the products that support the international standard should have the 'hooks' to accommodate the local implementation, which should be available in a flat file or a simple data base format. Again, a local implementation caters for local conditions.

In addition, African experts need to get involved in the activities of standards generating bodies (such as ISO/TC 211), to influence the development of international standards to ensure they meet Africa's needs – that way, the commercial GISs will support Africa's needs off the shelf, without needing to develop local standards. Participation in ISO/TC 211 can be done through national bodies (six African countries are members of ISO/TC 211, or through the 23 Class A Liaisons to ISO/TC 211, such as the International Association of Geodesy (IAG), the International Cartographic Association (ICA), the International Federation of Surveyors (FIG), the International Society for Photogrammetry and Remote Sensing (ISPRS), or the International Steering Committee for Global Mapping (ISCGM). It is hoped that regional organisations in Africa will also enter into liaison agreements with ISO/TC 211, to provide greater input from Africa into the development of standards for geographical information.

Participation in the activities of ISO/TC 211 can be done successfully via email – a very good example within ISO/TC 211 is that of the Czech Republic, which has made useful contributions to many standards without being represented at ISO/TC 211 meetings. However, such participation still requires much effort, as there is a high barrier of entry in the complexity of standards and one needs to dedicate the time to read the drafts of standards and supporting documentation, so that one can comment usefully on them – unfortunately, one needs to take care over the comments one provides on a standard to ensure that they are comprehensible and valid, because of the pressure under which the Editing Committee for the standard works when dealing with the comments.

Ideally, those African experts interested in standards and able to contribute to their development will be able to find the resources to attend meetings of standards generating bodies. However, it must be pointed out that many experts who have participated in ISO/TC 211 meetings, for example, have had to take leave and/or pay some of the costs out of their own pocket to do so.

6.3. The perfect standard

There are some who believe that a 'perfect' standard can be developed by a project team purely from a theoretical basis and using their skills and experience. However, no matter how much diligence is applied to writing a new standard, invariably many problems will be discovered as soon as anyone attempts to use the standard. Hence, it is critical that the standard be tested as early as possible in the development cycle, to validate the approach being taken and to highlight key issues that the developers might have overlooked.

It is also all too easy to assume that once a standard has been approved and implemented and the relevant training given, it will be used properly by those who should use it. Indeed, it would appear that this assumption is the case in some organisations. All implementations need to be followed by assessments of their implementation, to identify and understand:

- Non-compliance – and hence, possible remedial actions to ensure compliance, such as revising the standard;
- Implementation problems, such as software incompatibilities, missing data, onerous requirements, conflicts with other standards, or errors or ambiguities in the standard;
- Costs of implementing the standard – and hence, whether or not the standard is economically viable;
- Additional training and/or training materials needed; and
- Further standards that need to be developed.

7. Recommendations on which standards to use

ISO 19110 has some limitations, particularly relating to providing mechanisms to accommodate explicitly cultural and linguistic adaptability (CLA), that is, "*the ability for a product, while keeping its portability and interoperability properties, to:*

- *be internationalized, that is, be adapted to the special characteristics of natural languages and the commonly accepted rules for their use, or of cultures in a given geographic region; and*
- *take into account fully the needs of any category of user*" [ISO/IEC TR 11017].

Such adaptability is particularly useful in multi-lingual environments, which would apply to most, if not all, countries in Africa.

Nevertheless, we recommend that any feature catalogues used should conform to ISO 19110. Such a feature catalogue can still have feature types with labels (names) in multiple languages, implemented as aliases. Currently, the limitations to ISO 19110 are unlikely to affect most users of feature catalogues – indeed, at this stage, few feature catalogues will use the optional constructs provided by ISO 19110, such as feature operations and feature associations. In addition, ISO 19110 will continue to be maintained and enhanced by ISO/TC 211.

Unfortunately, it is not possible to recommend one, definitive data dictionary or feature catalogue to be used for all digital geographical information by all users across Africa under all circumstances. It is probably more effective to use a widely used feature catalogue that meets most of one's needs (adding more detailed feature types for in-house use, if necessary), than to try seek out that arcane feature catalogue that matches one's needs perfectly. The data dictionary or feature catalogue one should use depends on a number of factors:

(1) Is one using one's GIS in a narrow field, where one requires detailed feature types? Under such circumstances, it is best to use a specialist feature catalogue for that domain.

(2) Is one using one's GIS for general work, using mainly core data sets? Under such circumstances, it is best to use a general purpose data dictionary (eg: SANS 1880) likely to be used by one's data providers, though structured as a feature catalogue.

(3) Does one have one dominant client for one's data? Under such circumstances, one should use one's client's data dictionary or feature catalogue.

(4) Is no suitable data dictionary or feature catalogue readily available? Unfortunately, developing a data dictionary or feature catalogue can be very demanding (of expertise and time), while the user needs one immediately. Invariably, this results in a "quick fix" feature catalogue being developed, with subsequent long term problems with its use. Under such circumstances, one should network with one's peers to find a suitable data dictionary or feature catalogue.

(5) Are there too many data dictionaries and feature catalogues available? This is probably fairly common, with the added problems of there not being one data dictionary or feature catalogue that really meets one's needs, and there being mismatches between the various candidates (even Annex B shows this, with standards from only three sources). Under such circumstances, one should network with one's peers to get these data dictionaries and feature catalogues harmonized.

Unfortunately, developing or harmonising data dictionaries and feature catalogues and can a time consuming and tedious process.

8. Conclusions

We have provided an introduction to data content standards for digital geographical information in this report, and summarised our assessment of the 170 standards that we identified as candidates for further examination. Of these, we selected 14 standards containing data dictionaries or feature catalogues, and compared their feature types. We have also provided some advice and recommendations on data content standards (particularly for data dictionaries and feature catalogues) for Africa.

We trust that this report provides some useful guidelines for data content standards for Africa, and that it can be used as a basis for bringing some degree of harmonization to the data dictionaries and feature catalogues used in Africa, and to disseminating the better data dictionaries and feature catalogues to new users of digital geographical information. We also trust that this project will serve as a basis for future research on data content standards, not only in Africa.

9. Acknowledgements

We would like to acknowledge the support of US AID, USGS/EROS, EIS-Africa and the CSIR, that made this project possible. We would also like to acknowledge the initiative of Kate Lance of USGS/EROS who initiated the project and provided us with much insight and input, as well as the contributions of Doug Nebert (FGDC), Sives Govender (EIS-Africa), Andrew Terhorst (CSIR Satellite Applications Centre), Peter Schmitz (CSIR Centre for Logistics and Decision Support) and all the others who have provided us with inputs.

10. References

[Cooper 1993] Cooper AK, June 1993, *Standards for exchanging digital geo-referenced information*, unpublished MSc thesis, University of Pretoria, South Africa, 247 pp.

[Cooper 2003] Cooper AK, 2 October 2003, *Thoughts on categorising bloodstain patterns*, unpublished CSIR report (CSIR ***i***comtek document number: 0442-0001-701-A1) submitted to the European Network of Forensic Science Institutes (ENFSI), as part of the consideration to establish an Expert Working Group for Bloodstain Pattern Analysis (BPA), 3pp.

[Cooper 2004a] Cooper AK, 17 October 2004, *Status of the Standardization of Geographical Information in Africa*, Geoinformation Standards ISO Workshop, 5th African Association of Remote Sensing of the Environment (AARSE) Conference, Nairobi, Kenya.

[Cooper 2004b] Cooper AK, 21 October 2004, *Status of the Standardization of Geographical Information in Africa*, 5th African Association of Remote Sensing of the Environment (AARSE) Conference, Nairobi, Kenya.

[Cooper 2004c] Cooper AK, 13 December 2004, *Standards and their relation to the Land Cover Classification System (LCCS)*, Global Land Cover Network (GLCN) Workshop, Pretoria, South Africa.

[Cooper 2005] Cooper AK, 3 March 2005, *Metadata standards and related topics*, Africa Water Information Clearing House (AWICH) Workshop on Geo-Water Information Development and Management, Pretoria, South Africa.

[Di Gregorio & Jansen 2000] Di Gregorio A & Jansen LJM, 2000, *Land Cover Classification System (LCCS): Classification Concepts and User Manual*, Food and Agriculture Organization of the United Nations, ISBN: 92-5-10421.

[DIN 2000] DIN, April 2000, "*Economic benefits of standardization: Summary of results*", Beuth Verlag, 39pp.

[DSG 2003] Department of the Surveyor General, Zimbabwe, 2003, *National Standard for Digital Geographical Information, Version 1.0*, 93pp.

[GOS 2004] Geospatial One Stop, July 2004, *Information Technology – Geographic Information Framework Data Content Standards: Part 0 – Base Standard [draft]*, Information Technology Industry Council.

[ISO 19110] ISO 19110:2005, *Geographic information – Methodology for feature cataloguing*.

[ISO 19115] ISO 19115:2003, *Geographic information – Metadata*.

[ISO 19126] ISO 19126, January 2004, Geographic information – *Feature data dictionaries, feature catalogues and registers [Committee Draft]*.

[ISO 19136] ISO 19136, February 2004, *Geography Markup Language [Committee Draft]*.

[ISO/IEC 10746] ISO/IEC 10746, *Open Distributed Processing – Reference model*.

[ISO/IEC 11017] ISO/IEC TR 11017:1998, *Information technology -- Framework for internationalization*.

[SADCSTAN 2005] "*SADCSTAN brochure*", undated, 3 pages. Downloaded from the SADCSTAN Web site: http://www.sadcstan.co.za/pdf/SADCSTAN%20Brochure.pdf on 17 February 2005.

[SANS 1876] SANS 1876, 2005, *Feature instance identification standard [Draft South African Standard]*, Standards South Africa.

[SANS 1880], SANS 1880, May 2003, *South African Geospatial Data Dictionary (SAGDaD) and Its Application [Committee Draft]*, Standards South Africa.

[Scheepers *et al* 1986] Scheepers CF, Van Biljon WR & Cooper AK, September 1986, *Guidelines to set up a classification for geographical information*, NRIMS CSIR Internal Report I723, 12 pp.

[UML 2003] Unified Modeling Language™ (UML®), Version 1.5, March 2003, Object Management Group™ (OMG™). Available on 28 March 2005 from: http://www.omg.org/technology/documents/formal/uml.htm

[XML 2004] Extensible Markup Language (XML) 1.0 (Third Edition), 4 February 2004, World Wide Web Consortium (W3C). Available on 28 Marc h 2005 from: http://www.w3.org/TR/2004/REC-xml-20040204/