# OCS Coding Structures

## Information Management Team Report

Version 11 March 2011a

### Controlled lists

Lists 1 to 5 are contained in File: **IM-standards-lists-20110310a.xlsx**  
  
**1. CGIAR Centers and Challenge Programs**  
List of names, acronyms, slugs and logo files of CGIAR Centers and Challenge Programs with associated one-character codes as proposed by the Finance team. Currently in use in CGMap.  
  
**2. Countries**  
List of world countries, as derived from a combination of current UN and ISO standards, and including specific codes for use in maps and online information systems ( id, name, slug, iso\_alpha\_2, iso\_alpha\_3, iso\_numeric, polygon, latitude, longitude, subregion\_id, region\_id).   
  
Additional background information and rationale for the composite standard list is provided below  
  
**3. Regions**  
World regions based on UN Datastats standards.  
  
**4. MTP projects**  
Extracted from CGMap, this list includes: MTP period, Center, Project code, Project titles (starting from 2009-2011 MTP period)  
  
**5. Currencies**  
List based on ISO 4217 standard including a 3-digit alpha currency code   
  
**6. Location codes**   
It is proposed to use the UN LOCODE maintained by the United Nations Centre for Trade Facilitation and Electronic Business. UN/LOCODE (UNITED NATIONS CODE FOR TRADE AND TRANSPORT LOCATIONS) is the UN standard list of place names used in trade and transport sectors. This standard should be used in OCS to provide standardized names of places, as a complement to the country standards. The database can be requested from http://unece.org/cefact/codesfortrade/codes\_index.htm  
  
Institutions: Donors and Research partners; CG bodies  
File: **merged-partners-donors-reference-20110310a.xlsx**  
  
The final deliverable is a merged list of unique institution names from the two separate lists originally provided for donors and partners.

This deliverable is the result of the merger of two separate source lists of institutions to provide better consistency in the terminology and classification.  
  
At the time of detailed system design, it should be assessed whether the lists can definitely be merged and maintained as one, with appropriate search-and-browse interfaces to retrieve the appropriate name.  
  
The merged list is structured as follows:

* Organization Main Entry: authorised full name
* Acronym: authorised acronym
* Variant Names: in local language, old names, other names
* Country: of HQ (or independently operating office), based on CG standard labelling
* ISO Country code
* Type: common type proposed for OCS

An additional column is used to identify entries that are CG bodies, in preparation for a separate controlled list.  
  
**Methodology**  
Two lists were compiled starting from existing sources (the CGIAR Financial Information Systems for donors, lists in use in Centers for partners). Entries were standardised, duplicates removed, cataloguing guidelines identified on the basis of international standards and practices in the CGIAR.    
  
An additional round of validation was carried out on the merged partner and donor list: as of date, there are **2499** entries marked as unique for any final single list of institution names.

The volume of entries that were researched, standardised and cleaned was the major challenge encountered by the cataloguers. Here are some statistics:

Donors

|  |  |
| --- | --- |
| Sources | No. of entries |
| CGIAR FIS Members | 767 |
| CGIAR FIS Non-Members | 751 |
| Total entries in source lists | 1518 |
| **Total entries in OCS donor list (after “cleaning”)** | **881** |

Partners

At the onset of the project, the number of partner names that would be obtained from the Centers and analysed by the cataloguer was unpredictable. Therefore, priority was given to analysing and cleaning the lists provided by the Centers who will implement OCS in 2011.

|  |  |
| --- | --- |
| List owner | No. of entries |
| Bioversity | 640 |
| CIP | 358 |
| GRISP (incl. IRRI, AfricaRice, CIAT Rice) | 906 |
| ICARDA | 267 |
| IFPRI | 199 |
| WorldFish | 906 |
| Total entries in source lists | 3276 |
| **Total entries in OCS partner list (after “cleaning”)** | **1893** |

The remaining lists contain 1216 entries that need to be cleaned, analysed and catalogued. Finalising this non-priority lists is estimated to require 25 working days (which could not be completed in this Phase).

|  |  |
| --- | --- |
| List owner | Left to do |
| CIAT | 331 |
| ICRISAT | 486 |
| IITA | 165 |
| ILRI | 148 |
| IWMI | 86 |
| Total entries in source lists | 1216 |

A **consultation** was carried out in December 2010 on a series of specific aspects of the classification and level of detail of the name entries. On the basis of the feedback received, we introduced new types for the classification, and decided to cover a sample of offices of governments and development organizations that operate independently as donors.  
  
A **common set of classification types** were identified, defined and applied to both partner and donor lists (see the Cataloguing guidelines section below): the Type describes the nature of the organization and is intended for global analysis of donors and partners across projects and programs in OCS.   
  
The proposed types were derived starting from:

* types identified and defined during an ICT-KM pilot project in 2008 (based on the types defined by the Science Council);
* analysis of the FIS classification;
* analysis of the classification types currently in use in the Centers that submitted their partner lists.

The December consultation highlighted some classification issues (e.g. missing types, generic definitions) that led us to:

1. introduce new types and   
2. reclassify some entries to test the applicability of, for example, IO (International Organization) and CSO (Civil Society Organization) originally bundled with NGOs. These specific issues are discussed in the section “**Important notes**” below.  
  
**Guidelines** were defined on how to select and standardise names, variants and acronyms. As for the Country field in the lists, the country of the organization’s headquarters was entered.   
  
A separate **sample of regional or country offices** which operate independently as donors is provided. This sample was derived from suggestions provided by the survey respondents during the December consultation. From this sample, cataloguing guidelines were identified.

### Important notes about the status of the institution name list

**One common list of institutions in OCS**: the merged list of donor and partner institutions sets the ground for the creation of a single list for common terminology and classification purposes. The feasibility of managing one list instead of two should be finally assessed during the system design phase, because it may have dependencies on the software functionality offered.   
  
In preparation for final implementation, it should also be assessed whether institutions should be identified specifically as donors and/or partners in the controlled list on the basis of their function in the CGIAR business processes. This additional property (function of the institution) may be needed in preparation for the formulation of donor codes. In case it is not needed, e.g. an institution’s function depends exclusively on the context of use (e.g. donor in a specific grant, partner in a specific project), every entry in the controlled list should have a code, as a property of the record, regardless of their function in specific contexts.  
  
**A separate controlled list for CG bodies**: names of Centers and Programs are found in the names list, partner and donor list, and codes are provided in the finance documentation. Considering that in an internal system like OCS, CG bodies(t Centers, Programs and Central Offices) will need to be mentioned in different contexts, from internal documentation to public communications, we propose to create a separate controlled list for such names, possibly with additional properties that would bring together the different coding and standardisation requirements, such as:

* full name
* acronym
* short name
* legal name
* donor code
* Web site
* logo

Given the potential frequency of use in business documentation, different cataloguing guidelines may be developed for the CG body controlled list.  
  
**Independently Operating Offices**: the consultation confirmed that some organizations have national or regional offices that operate as donors independently from headquarters. Therefore, this level of detail should be captured in the Donor list. A sample of organizations that operate in such a way was analysed and catalogued in order to derive the guidelines for entering decentralised offices in addition to headquarters in the Donor list. This sample is delivered separately from the merged list. Additional entries should be entered in order to thoroughly test and validate the guidelines. As validation of the structure and cataloguing guidelines for this sample is pending, the sample has not yet been added to the merged list of institutions.  
File: **DONORS-Independent-Offices\_sample\_20110128a.xlsx**  
  
**Parallel Center-specific classification of partners**: we have observed a wide variety of types and classification styles across the Centers  which have provided their current lists. While the global classification by type in OCS is intended for common reporting purposes at the CRP level primarily, we foresee that Centers may want to maintain their own local classification of partners. Should this requirement emerge clearly during the design and specification stage of OCS, our recommendation is that this should be possible in OCS with a cautionary note on the overall costs of maintaining a double classification and possibly a duplicated curator role (a central and a Center-based one). What must be enforced is  the quality control process on the entry and modification of new and existing names and acronyms, so as to maintain a standardised terminology across OCS and the CGIAR System.  
  
**Type International Organization (IO)**: after the December consultation, it was clear that this type was not sufficiently specific and could cause confusion with other types like Development Organization (DO) and Non-profit Organization (NGO). Therefore, we reclassified the very few instances of IO (less than 10 in the Donor list): there are still 7 IO entries, for which more research into their nature should be carried out.  
  
**Type Civil Society Organization (NGO-CSO)**: during the consultation, a respondent suggested that CSO’s should be kept separate from NGO’s. We introduced a temporary type NGO-CSO to track the occurrence of organizations that may be classified as CSO. Seventeen organizations have been classified as NGO-CSO: this temporary type is currently assigned to women’s associations, federations of unions or farmers’ unions, religious organizations.  
  
Since we have not found a suitable definition for CSO, we recommend that the applicability and definition of this type be reassessed in due course.

### Cataloguing and classification guidelines for donors and partners

File: **Donor and Partner Cataloguing Guidelines for OCS**

This document lays out the guidelines and examples of how to select, validate and enter names of organization, and how to classify names by type. While it documents the key criteria by which the lists were developed, it is intended as the basic reference manual for the proposed curation role.

### CGIAR Standard for Country Names and Codes

The purpose of country name and code standardization is to:

1. provide syntactic interoperability (data exchange, portability and reusability) compatible and aligned with the well-maintained and extensively used standards implemented by international organizations, governments, the United Nations, and private companies;
2. make available the required representations of the CGIAR’s work in locations of geographical interest  and of the CGIAR’s dealings with countries.

To this end, the CGIAR uses the [ISO 3166 standard for country codes](http://www.iso.org/iso/country_codes/background_on_iso_3166/what_is_iso_3166.htm).  This standard, as shown by the most widely-used implementations (see a [selection of implementations](http://www.iso.org/iso/country_codes/background_on_iso_3166/implementation_of_iso_3166-1.htm)), is [adaptable and customizable](http://www.iso.org/iso/country_codes/background_on_iso_3166/customizing_iso_3166-1.htm) to fit an implementer’s needs by offering space for coding and labeling.  
  
The motivations for maintaining the CG Countries list as a composite of ISO/UN standards are to:

1. retain the historical geographic significance of CGIAR information and records;

Note: this enables the CG to maintain an historical record of information for statistical use and information management (see a [World bank document found via search for Serbia](http://econ.worldbank.org/external/default/main?pagePK=64165259&theSitePK=469372&piPK=64165421&menuPK=64166093&entityID=000009265_3961003073937);  the document is from 1992, before Serbia became an independent country, yet it maintains the geographic historical record and lineage).  The United Nations maintains the ISO 3166 names/number codes and extends the standard to retain historical information on changes, additions, and codes not in use;

2. provide the ability to label country names for the context of use (publications, information systems, diffusion and representation of information, statistical compilation);

Note: this enables the CG to rely on codes while the labeling of country names can be customized to best fit the context and use (e.g., politically neutral names, short name labels, exonyms).  The CGIAR has already varied from the ISO 3166 name for PALESTINIAN TERRITORY, OCCUPIED in the operational context of where we are performing agricultural research.  The name we use, West Bank and Gaza, is based on the geographic area and is widely used (see [World Bank country page](http://web.worldbank.org/WBSITE/EXTERNAL/COUNTRIES/MENAEXT/WESTBANKGAZAEXTN/0,,menuPK:294370%7EpagePK:141159%7EpiPK:141110%7EtheSitePK:294365,00.html)) in the international community.

### Proposed curation function

We propose that a curation function be defined and made operational under the auspices of and mandated by the Consortium Office. Such a function is fundamental to guarantee continuity, maintenance and update of the controlled lists and information management standards, including but not limited to the donors and partners. It is fundamental that an ultimate authority makes executive decisions on the official terminology, classification and update of the information management standards in OCS.  
  
The function may be assigned to a geographically distributed team operating according to the same mandate and functions.  
  
The team in charge of curation should be responsible for managing and overseeing the development and maintenance of the information management standards, and in particular for:  
●     rules and processes needed to maintain and update the standards;  
●     regular maintenance of entries and classification;  
●     validation of new requests and data entry as per the cataloguing guidelines;  
●     regular revision of the classification applicability;  
●     future expansion and redefinition of the standards.  
  
In the upcoming design and implementation phases of OCS, the curator should also participate in defining the configuration of OCS for the part regarding the centralised management and update of the standards.

### Recommended next steps

What has been achieved during the OCS Coding Structures project has to be considered as the first major step towards the full setup of a comprehensive body of information standards that are useful and usable in the context of the CGIAR business practices. The lists were cleaned and the names classified to the best of our knowledge and research expertise[[1]](#footnote-1). The standards now constitute a living body of knowledge that needs to be further refined and continuously maintained in preparation for implementation in OCS and the setup of operational maintenance processes.   
  
In order to maintain and further refine this body of knowledge, it is proposed that **the Steering Committee endorses the creation of an interim curation function** to complete a number of preparatory tasks and fully participate in the system design phase in defining requirements for common information standard management in OCS.  
  
In particular, the following tasks should be assigned to an interim curation team under the auspices and supervision of the Consortium Office:

**1. Partners and donors**

* the remaining lists of partners submitted by CIAT, ICRISAT, IITA, ILRI, IWMI should be cleaned, classified and made available in the main list for implementation in OCS;
* it should be assessed whether institutions should be identified specifically as donors and/or partners in the controlled list on the basis of their function in the CGIAR business processes. This additional property (function of the institution) may be needed in preparation for the formulation of donor codes, and subsequently as one of the requirement for a common list of institution names;
* validation of the level of detail needed for independently operating offices is needed for this group of entries to be added to the official list.

**2. Donor codes**

Formulation of donor codes is needed as part of the preparatory phase: this should be achieved in collaboration with the Finance theme team.

**3. CG bodies**

The structure and cataloguing rules to be used in a separate controlled list of names of CGIAR Centers and Programs should be reviewed and further detailed according to the different contexts of use (e.g. finance, project management, codes, etc).   
  
**4. CG country labelling standard**While an executive decision is needed on the requirement to maintain a composite standard list of countries that fit the coding and labelling requirements outlined above, updates to the composite standard should be applied over time.

**5. Cataloguing guidelines**

The cataloguing guidelines should be maintained up-to-date and expanded on the basis of emerging requirements (e.g. CG bodies, independently operation offices).

**6. Resolution of IO and NGO-CSO type assignment**

Applicability of the IO and CSO definitions and types should be assessed before implementation.

**7. Quality review before creation of a single OCS list for institutions**

Another round of quality control should be carried out to ensure that all invalid duplicates are removed, CG bodies are managed separately, any additional property for priority display is included.

**8. Centralised list management functionalities in OCS**

Definition of OCS functionalities for centralised management and update of the standards and the common institution name list.  
  
To carry out the tasks above, it will be necessary to hire a cataloguing expert for 3 months or until the start of the OCS design phase, at which time this function should be transitioned to the Information Standards team of the OCS Support Unit.

1. The merged list holds 12 entries with no Type assigned. [↑](#footnote-ref-1)