A Joint Initiative for Harmonization and Validation of Land Cover Datasets

Martin Herold, *Student Member, IEEE*, Curtis E. Woodcock, Antonio di Gregorio, Philippe Mayaux, Alan S. Belward, John Latham, and Christiane C. Schmullius

***Abstract*—An international initiative aimed at the harmonization and validation of existing and future land cover datasets is needed to support operational earth observation of land. The goal is to overcome current limitations of land cover datasets with respect to their compatibility and comparability and unknown accuracy. These limitations significantly hinder a variety of applications. Key entities in this effort are the Land Cover Implementation Team of Global Observation of Forest Cover/Global Observation of Land Dynamics, the Global Land Cover Network, and the CEOS Group on Calibration and Validation. In their recent efforts, they have explored and provided the methodological and organizational resources to foster such an international cooperation. The approaches described in this paper include an introduction of the UN Land Cover Classification System as a common land cover language and a basis for legend translation. All actors involved in land cover mapping are invited to participate in this initiative.**

***Index Terms*—Global Land Cover Network (GLCN), interoperability, land cover, Land Cover Classification System (LCCS).**

I. Introduction

SEVERAL global and regional datasets have been derived in response to the need for information about land cover and land cover dynamics. Their development was driven by different national or international initiatives; the subsequent mapping standards adopted reflect the varied interests, requirements, and methodologies of the originating programs. Available data products include IGBP DISCOVER, the MODIS land cover product, University of Maryland (UMD) land cover product, GLC 2000, CORINE-land cover 1990 and 2000, AFRICOVER, and the MODIS continuous fields products. Although efforts in harmonization and validation are mentioned in nearly all related mapping projects, as well as in many other circumstances, there is only limited compatibility and comparability between these different maps and their thematic legends—they basically exist as independent datasets. Despite indications of existing uncertainties in successful validation exercises [1], [2], they often are not or only marginally validated [3]. This hinders their use, especially considering the original purpose of some these datasets to serve a large number of applications and the problems of insufficient validation become particularly apparent in efforts to analyze land cover changes.

For example, the report on the implementation plan of the global observing systems for climate supporting the work of the United Nations Framework Convention on Climate Change (UNFCCC) emphasizes how this lack of homogeneous observations limits our capacity to monitor terrestrial changes relevant to climate as well as our ability to investigate the causes of land-surface changes [4]. Seeking solutions to overcome these limitations would increase the value of these datasets, in particular, by providing a consistent legend for upcoming global land cover mapping initiatives, such as GLOBCOVER. Indeed, the GCOS report [4] recommends establishing an international mechanism to prepare and issue regulatory and guidance material relating to terrestrial observing systems—a recommendation that the governments participating in the Conference of the Parties to the UNFCCC have endorsed (COP 9/Decision 11), in particular, by tasking the Global Terrestrial Observing Systems (GTOS). Both, the GCOS implementation plan [4] and the Group on Earth Observation [5] implementation plan task Food and Agricultural Organization (FAO)-GTOS and Global Observation of Forest Cover/Global Observation of Land Dynamics (GOFC-GOLD) to move forward in this arena.

This paper outlines a strategy and first experiences for an initiative to foster harmonization and validation of all existing and upcoming global land cover datasets. A coordinated international effort and comprehensive consensus building are essential for such a task to be successful. The general approach is to combine experience and resources from all actors involved in global earth observations of land including space agencies such as the European Space Agency (ESA) and the National Aeronautics and Space Administration (NASA), the FAO, GTOS and the Global Land Cover Network (GLCN), the Committee Earth Observation Satellites (CEOS) with the group on calibration and validation (Cal/Val), and the European Joint Research Center (JRC). All of these institutions, as well as the users of land cover datasets, will benefit from such a combined effort. The current initiative is supported by the Land Cover Implementation Team (LC-IT) of the organization GOFC-GOLD. More specifically, we describe the background for international cooperation in earth observations of land, a general framework for land cover harmonization, introduce the UN Land Cover Classification System (LCCS), and outline the validation strategy as part of this joint initiative.

Manuscript received November 2, 2004; revised May 22, 2005. This work was supported in part by the European Space Agency (ESA) through the GOFC-GOLD land cover project office.

M. Herold and C. C. Schmullius are with the GOFC-GOLD Land Cover Project Office, Department of Geography Friedrich-Schiller University, Loebdergraben 32, 07743 Jena, Germany (e-mail: m.h@uni-jena.de).

C. E. Woodcock is with the Department of Geography, Boston University, Boston, MA 02215 USA.

A. di Gregorio is with the Istituto Agronomico per L'Oltremare (IAO), Florence 50131, Italy.

P. Mayaux and A. S. Belward are with Global Vegetation Monitoring Division of the Joint Research Center (JRC) of the European Commission, 21020 Ispra, Italy (e-mail: alan.belward@jrc.i).

J. Latham is with the Food and Agricultural Organization of the United Nations (FAO), Rome 00100, Italy.

Digital Object Identifier 10.1109/TGRS.2006.871219

II. Framework for International Cooperation

The need for harmonized and validated earth observation products is endorsed in several international conventions and treaties, i.e., in UNCED's Agenda 21, the World Summit on Sustainable Development (WSSD) in Johannesburg, South Africa, in 2002 and the related Group of Earth Observations (GEO) formed in 2003 [5], [6]. However, land cover data interoperability or compatibility would be a dream for the users but can be a nightmare for the data developers [7]. Full comparability is certainly a long term goal but has to start and evolve from an international consensus building effort. In the domain of land observations, the organization GOFC-GOLD and its LC-IT provides an appropriate body to support international initiatives for harmonization and validation of land cover products [8], [9]. GOFC-GOLD developed as prototype activity of the Integrated Global Observation Strategy (IGOS) and is now an established panel of the GTOS, the Secretatriat of which is hosted by the FAO in Rome.

Fig. 1 highlights the role of GOFC-GOLD to support communication and cooperation with key actors involved in global terrestrial earth observations. The goal is to strengthen cooperation and coordination among global observing systems and research programs for integrated global observations of the environment including the endorsement of data interoperability and data sharing [6] by learning from problems obvious in previous international space cooperation [10], [11]. The framework to foster a joint initiative on harmonization and validation of global and regional land cover dataset was outlined in two GOFC-GOLD workshops in Jena, Germany, and Rome, Italy, in 2004 [12], [13]. Harmonization and validation are parallel efforts that complement and profit from each other. The multifaceted (though individually limited) experiences and resources need to be combined to approach the issues as one joint project incorporating all existing global land cover datasets.

III. Framing Harmonization

Harmonization in the context of land cover characterization can be understood as a process whereby the similarities between existing definitions of land cover are emphasized, and inconsistencies reduced. The ultimate goal is to bring various land cover datasets in "harmony,"thus allowing direct comparison between them. This process follows a "bottom up" perspective. Beginning from a state of divergence in land cover datasets it seeks compatibility and comparability. Harmonization does not necessarily eliminate all differences, but should eliminate major inconsistencies. In other words, one product's forests should not be another's woodland. However, it should be recognized that different land cover products could characterize forests to different levels of detail. Standardization, in contrast, is a "top down" process, and is, therefore, far more rigid. It requires common definitions and standards to derive land cover information and should eliminate all inconsistencies—and differences—between the datasets [14].

Versatile international harmonization experiences exist, e.g., in fields of economy, legal issues, medicine, and environmental statistics [14], [15]. One prominent example is the world wide harmonization of soil maps. The FAO classification system was used as the standard classification system and common language to harmonize existing national and regional soil maps and their legends [16]. In general, heterogeneity in land cover maps results from different methods and standards used to create them and has multiple facets. They include syntactic issues (e.g., logical data models: vector/raster), schematic heterogeneity (e.g., database models, spatial reference systems, cartographic standards including variable minimum mapping units and mixed units), and semantic aspects [17].

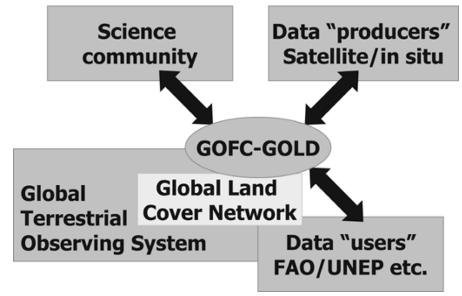

Fig. 1. Role of GOFC-GOLD as platform for international communication and cooperation.

The latter referring to naming and cognitive conceptualizations of land cover legends. The reasons for data heterogeneity and a lack of harmonization are versatile. There is often confusion between the terms "classification system" and "legend," especially in treatment of mixed units. A classification is an abstract representation of the situation in the field using well-defined diagnostic criteria to order and arrange objects into groups or sets on the basis of their characteristics and relationships, i.e., in terms of factors like percent cover and height. A thematic legend is developed from a classification system for a specific mapping purpose (Fig. 2). Many land cover legends are derived without underlying classification systems, or at least use different systems, and, thus, lack compatibility. These maps reflect the requirements of different national and international mapping agencies or specific applications in their legends and mapping approaches.

Sometimes, legends show inconsistencies in definitions, such as overlap or gaps between thematic classes that could have been avoided if a comprehensive classification system had been used. In remote sensing applications, land cover definitions often result from spectral and/or temporal classes distinguishable in the image data that are then transferred into thematic categories (e.g., as in unsupervised classifications). The related mapping products often contain a list of "producer" defined classes that are more or less suitable for users. Also, land cover legends differ due to the unique requirements of specific applications or models. Harmonization does not cause compromises of the applications requirements, but standardization could negate the utility of a legend for specific uses. Hence, there can be tradeoffs between the requirements of specific applications, the quality of dataset, and the level of standardization.

Despite these known differences and problems, there is, as yet, no internationally accepted land cover classification system. Thus, one essential component to complete harmonization and validation between different land cover dataset is to

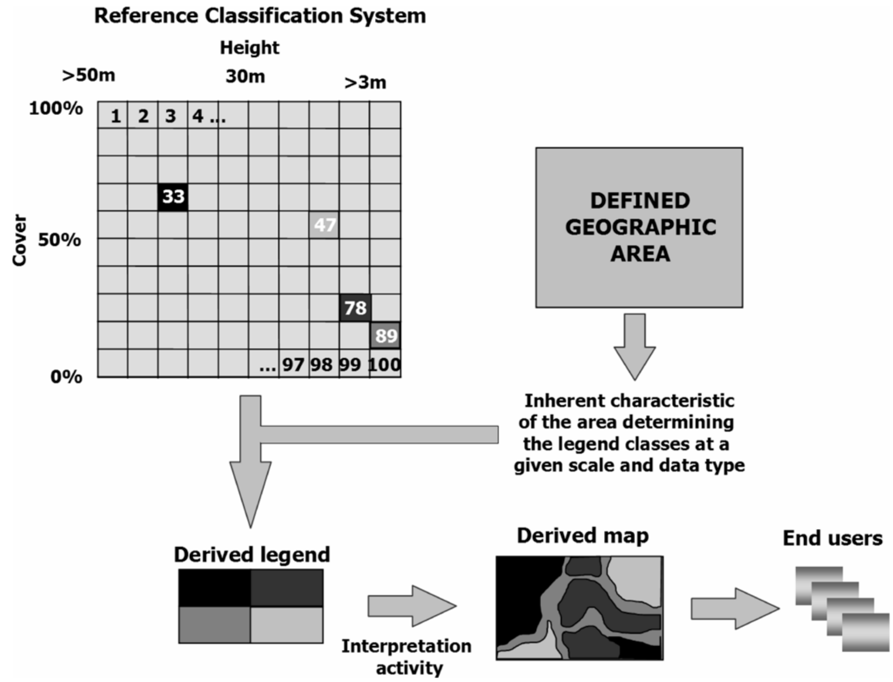

Fig. 2. Process of developing of land cover legends from a classification system.

find a common language to translate between the different legends (see Section IV). Similar to the experience of soil sciences, a harmonization of land cover should first try to harmonize the parameters used for description of land cover (e.g., classifiers); then, if these are applied to various systems and legends, one can always harmonize the individual criteria used to create land cover categories of whatever name in whatever language.

Common problems exist in the definitions of land cover types. A prominent example is the definition of forest. The discrimination between woodlands and shrublands is usually based on physiognomic aspects like the minimum height and percent cover of trees. The IGBP legend defines a forest with a percent tree cover >60% and a height exceeding 2 m. GLC2000 uses percent tree cover >15% and height exceeding 3 m to discriminate shrubs and trees. This difference is one of the major reasons for the disagreement between the forested areas represented in existing global land cover maps (Fig. 3). Although there is a fair amount of agreement (green areas), GLC2000 shows the most forest (in blue) since the requirements for percent cover of trees is the lowest of the various legends. This difference is most prominent in transition zones between tropical forests to savannahs and boreal forests to tundra vegetation. This example highlights just one class definition problem, but many others exist. The different definitions of forest directly affect the mapping of other classes such as shrublands. In fact, the IGBP legend contains savannah classes that include different densities of tree cover, e.g., woody savannah (30%–60%), and savannah (10%–30%). The canopy density differences provide some flexibility in comparing maps with different legends. For example, the DISCover- GLC2000 comparison (Fig. 3) of forested areas could have used all of the IGBP tree covered classes, not just the ones actually named forest. This highlights the importance of a common language and a set of commonly agreed classifiers to approach harmonization. Legend developments and comparisons should not be based on a list of names. The definition of the generic land cover characteristics (e.g., basic life forms such as trees, shrubs, herbaceous vegetation, bare areas, etc., and their density), using a common classification system will be independent of any local or regional naming conventions. Similarly, unclear and inconsistent definitions and discriminations remain for other land types such as urbanized areas or human settlements. In the same context, their are often unclear separations between land cover and land use terms (e.g., pasture versus grasslands versus herbaceous vegetation).

Harmonization efforts to overcome the inconsistencies between land cover dataset were initiated by the United Nations Environmental Program (UNEP) and FAO in the early 1990s, in parallel with the increased use of GIS and spatial analysis. The main objective of this initiative was a response to the need for harmonized and standardized collection of data mentioned in several international conventions. The first expert meeting on harmonizing land cover and land use maps was hosted by UNEP/FAO in November 1993 in Geneva, Switzerland [18]. The meeting was framed by a strong requirement for a single internationally accepted land cover and land use legend. However, it became obvious that too much standardization reduces application, relevance, and versatility. It became apparent that it was more important to standardize terminology than categories. The development of a prototype international land cover reference system was fostered, through an iterative international working group process which eventually evolved to become the FAO/UNEP LCCS. Building on these experiences, the GOFC-GOLD Land Cover Implementation Team in tandem with the team of Global Land Cover Network of FAO/UNEP started to coordinate and foster the international harmonization effort and development of an implementation strategy during two workshops in March and July 2004 in Jena, Germany, and Rome, Italy.

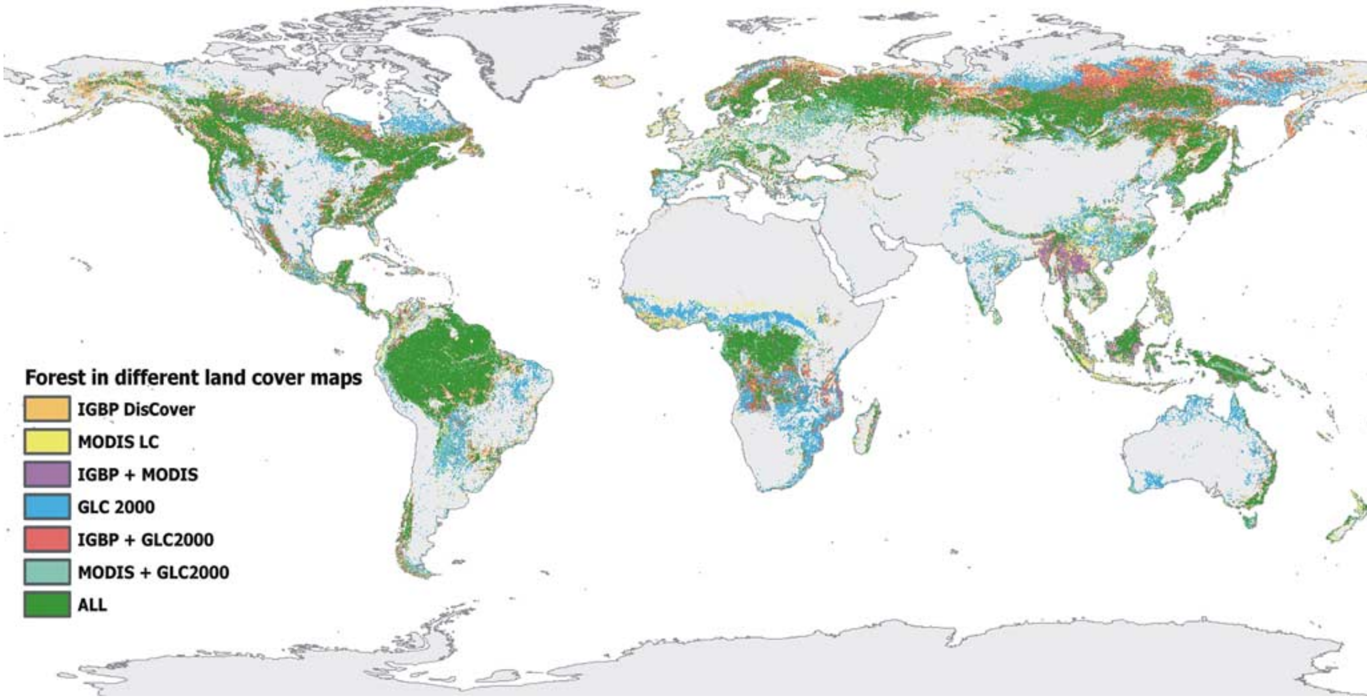

Fig. 3. Forest areas as represented in three different global land cover maps: IGBP Discover (classes 1–5), MODIS land cover (classes 1–5), and GLC2000 (classes 1–8).

A parallel development that might be of interest in the harmonization efforts is scientific progress in the field of "Interoperability in GIS." GIS research dealing with categorical ontologies and semantic interoperabilities has developed methods to measure and resolve terminological and conceptual incompatibilities in spatial datasets. There are approaches to measure semantic similarities between land cover legends using natural language processing or concept lattices, tools for integration of heterogeneous ontologies, and change detection analysis in various land cover datasets [17], [19]–[22]. These approaches can assist in the comparison and translation of legends, but do not approach the actual problem of heterogeneous mapping standards. The implementation and the development of frameworks that impact operational mapping will only be successful if they are conform to strategic decisions on the political level and pushed through the international bodies and agencies responsible for such tasks.

IV. UN Land Cover Classification System (LCCS)

The previous consensus building efforts and experiences of GOFC-GOLD have resulted in general agreement that the UN LCCS provides a valuable universal land cover language for building land cover legends and comparing existing legends. LCCS was created by FAO and UNEP in response to a need for harmonized and standardized collection of land cover data, availability of land cover data for a wide range of applications and users, and comparison and correlation of land cover classes. LCCS is an *a priori* classification system. It represents a worldwide reference system for land cover able to combine high flexibility (ability to describe land cover features all over the world at any scale or level of detail) with an absolute level of standardization of class definitions between different users [23]. The first version of LCCS was available in 2000; a new updated and improved version (LCCS-2) was released in November 2004 and is available from the GLCN webpage: http://www.glcn-lccs.org.

LCCS allows a dynamic creation of classes via a dynamic combination of land cover diagnostic attributes called classifiers without obliging users to use a predefined list of names. In its basic dichotomous categorization levels, LCCS distinguishes eight major land cover groups in its dichotomous phase (Fig. 4). These classes can be further described in the modular-hierarchical phase where the set of classifiers (e.g., vegetation life form, density, height, leaf type, longevity, etc.) and their hierarchical arrangement are tailored to the major land cover type. A resultant suite of more than 200 000 classes of land cover are potentially derivable using a combination of these classifiers. Further definition of a land cover class can be achieved by adding attributes. Two types of attributes that form separate levels in the classification are distinguished: environmental attributes (e.g., climate, land form, soils/lithology, and erosion) and specific attributes (e.g., floristic composition and crop type). Through the addition of these environmental classifiers, the resultant number of potential classes increases exponentially.

A central component in creating land cover legends from LCCS-2 is the incorporation in the software of a standardized codified syntax to deal with the representation of mixed unit classes. The mixed unit concept is an important component of the "cartographic standards" of a map and is directly linked with the scale and the minimum mapable area (MMA) concept. This standard is applied in LCCS when passing from the abstraction of the classification (by definition not related to scale) to a specific legend. Several kinds of mixed units exist (Fig. 5). The mixture $\mathbf{A}/\mathbf{B}$ between two classes represents a cartographic generalization, since, due to the scale, the extent of the features is

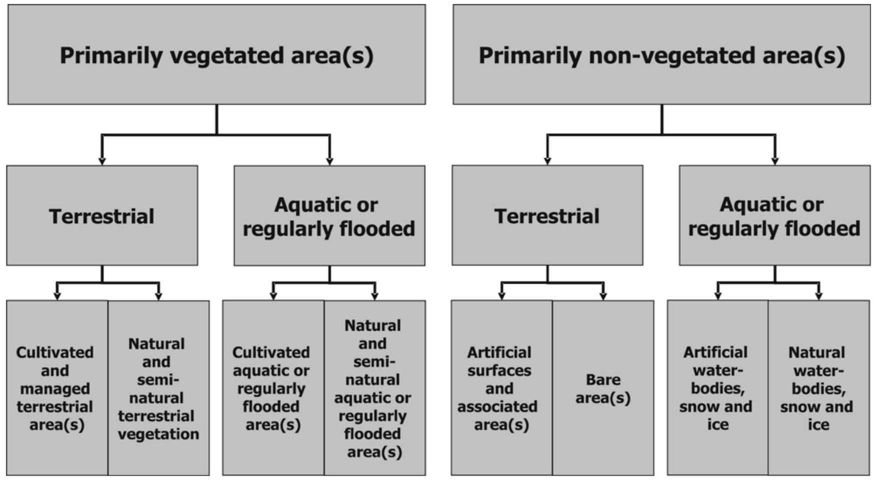

Fig. 4. LCCS-2 user interface at the initial dichotomous classification phase.

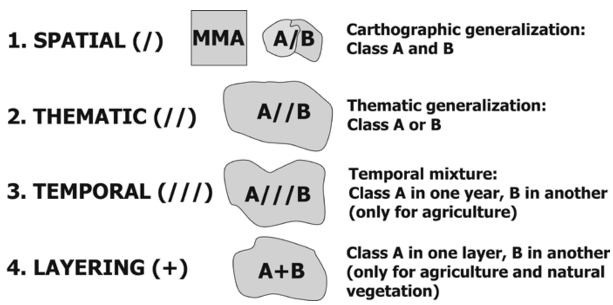

Fig. 5. Mixed unit concept within LCCS 2 (MMA = minimum mapable area).

smaller than the MMA. It means that both features A and B exist, but in this specific polygon they cannot be represented singularly. The sequence of the classes in the code means that feature A is predominant (larger area extent in the polygon) as compared to B (B covers at least 20% of the polygon area). A cartographic mixed unit can be done with more than two classes. $\mathbf{A}//\mathbf{B}$ is a thematic generalization. It means the area can have the feature A or the feature B, i.e., for classes like "snow and ice." $\mathbf{A}///\mathbf{B}$: is a time related mixture. It is used only for agricultural classes and means the area has the class A one year and the other year the class B. Another type of mixed units $\mathbf{A}+\mathbf{B}$ exists if different layers of vegetation are present, e.g., to describe areas with agro-forestry.

The LCCS software program contains three different Modules. Land cover categories are defined in the classification module by the combination of a set of independent classifiers. They are hierarchically arranged and can be linked with environmental and specific discipline related attributes. The legend module contains storage of land cover classes. This module allows the export of data in commonly used file formats. It also supports users incorporating their own user- defined names to the already provided standard names. Finally, there is a translator module which can be used to compare and correlate classifications and/or legends. The software and the manual are available for download from the LCCS web site. LCCS is in the process of being considered into ISO TC 211 (International Organization for Standardization—Technical Committee) to obtain the status on an international standard. This will strengthen LCCS as basis for the evolution of a standardized approach to land cover classification.

V. Implementing Harmonization

The implementation strategy for harmonization of land cover datasets developed in consensus with actors involved in land cover mapping during the GOFC-GOLD/GLCN workshops has several components. Currently, the implementation strategy focuses on "land cover."

A. Identification and Supply of Harmonization Resources

An essential resource for harmonization is LCCS. During the GOFC-GOLD/GLCN workshops LCCS has been tested to determine if it is as appropriate classification system to provide a common language and legend translation device. GOFC-GOLD in cooperation with GLCN and endorsed through GTOS recommend LCCS to space agencies and other actors involved in land cover mapping as a standard for land cover legend generation and as an exploratory tool for comparing and contrasting different legends. The software and documentation for LCCS are freely available on the web. GOFC-GOLD and its partners are supporting the UN initiative of the Global Land Cover Network to raise awareness and foster the use of LCCS and harmonized products in workshops, tutorials, and on demand for specific users or projects. As part of that process, GOFC-GOLD and its partners will work toward international consensus for the classifier thresholds and the hierarchy of these thresholds used in LCCS.

B. Development and Documentation of Harmonization Experiences for Existing Datasets

Exercises are underway to evaluate the resources and provide better understanding on how to harmonize the variety of existing land cover datasets and their legends. The first step is to develop legend translation protocols for the different existing legends into LCCS. The translation of the IGBP legend, for example, into LCCS, is shown in Table I. The IGBP legend was translated based on the IGBP class definitions. The comparison shows general differences in terminology (forest versus trees); different types of thematic and cartographic land cover mixtures

TABLE I
TRANSLATION OF THE IGBP LEGEND INTO LCCS

IGBP class	LCCLabel	LCC-Code	LCCLevel	IGBP description	Translation Notes
Evergreen Needleleaf Forests	Needleleaved Evergreen Trees	20092	A3A10B2XXD2E1	Lands dominated by woody vegetation with a percent cover > 60 % and height exceeding 2 meters. Almost all trees remain green all year. Canopy is never without green foliage.	LCCS limits tree height classification to > 3 m
Evergreen Broadleaf Forests	Broadleaved Evergreen Trees	20089	A3A10B2XXD1E1	Lands dominated by woody vegetation with a percent cover > 60 % and height exceeding 2 meters. Almost all trees and shrubs remain green year round. Canopy is never without green foliage.	LCCS limits tree height classification to > 3 m
Deciduous Needleleaf Forests	Needleleaved Deciduous Trees	20093	A3A10B2XXD2E2	Lands dominated by woody vegetation with a percent cover > 60 % and height exceeding 2 meters. Constists of seasonal needleleaf tree communities with an annual cycle of leaf-on and leaf-off periods.	LCCS limits tree height classification to > 3 m
Deciduous Broadleaf Forests	Broadleaved Deciduous Trees	20090	A3A10B2XXD1E2	Lands dominated by woody vegetation with a percent cover > 60 % and height exceeding 2 meters. Constists of broadleaf tree communities with an annual cycle of leaf-on and leaf-off periods.	LCCS limits tree height classification to > 3 m
Mixed Forests	Broadleaved Closed Trees / Needleleaved Closed Trees	20088 / 20091	A3A10B2XXD1 / A3A10B2XXD2	Lands dominated by woody vegetation with a percent cover > 60 % and height exceeding 2 meters. Consists of tree communities with interspersed mixtures or mosaics of the other four forest types. None of the forest types exceeds 60 % of landscape.	LCCS limits tree height classification to > 3 m Tricky IGBP definition, since none of the other 4 forest types should exceed a 60 % coverage. However, often not all those types are present so that one type may exceed this threshold easily.
Closed Shrublands	Closed Medium High Shrubland (Thicket)	20018-13476	A4A10B3-B9	Lands with woody vegetation less than 2 meters tall and with shrub canopy cover > 60 %. The shrub foliage can be either evergreen or deciduous.	LCCS defines shrubland between 5 m and 0.3 m
Open Shrublands	Open Medium High Shrubs (Shrubland)	20022-13476	A4A11B3-B9	Lands with woody vegetation less than 2 meters tall and with shrub canopy cover between 10 - 60 %. The shrub foliage can be either evergreen or deciduous.	LCCS defines shrubland between 5 m and 0.3 m
Woody Savannas	((70-60) - 40%) Woodland with Herbaceous Layer	20317-1	A3A11B2XXXXXXF2 F4F7G4-A12	Lands with herbaceous and other understory systems, and with forest canopy cover between 30 - 60 %. The forest cover height exceeds 2 meters.	LCCS limits tree height classification to > 3 m Lower treshold LCCS vs. IGBP: 40 vs. 30 % "and other understory systems" cannot be specified within LCCS
Savannas	(40 - (20-10)%) Woodland with Herbaceous Layer	20317-3012	A3A11B2XXXXXXF2 F4F7G4-A13	Lands with herbaceous and other understory systems, and with forest canopy cover between 10 - 30 %. The forest cover height exceeds 2 meters.	Upper treshold LCCS vs. IGBP: 40 vs. 30 % "and other understory systems" cannot be specified within LCCS
Grasslands	Herbaceous Closed to Open Vegetation	21453	A2A20	Lands with herbaceous types of cover. Tree and shrub cover is less than 10 %.	
Permanent Wetlands	Natural And Semi-Natural Aquatic or Regularly Flooded Vegetation	0007	A24	Lands with a pemanent mixture of water and herbaceous or woody vegetation. The vegetation can be present in either salt, brackish, or fresh water.	
Croplands	Herbaceous Crop(s)	10025	A3	Lands covered with temporary crops followed by harvest and a bare soil period (e.g., single and multiple cropping systems). Note that perennial woody crops will be classified as the appropriate forest or shrub land cover type.	
Urban and Built-Up Lands	Built Up Area(s)	5001	A1	Land covered by buildings and other man-made structures.	
Cropland/Natural Vegetation Mosaics	Cultivated and Managed Terrestrial Area(s) / Closed to Open Trees / Closed to Open Shrubland (Thicket) // Herbaceous Closed to Open Vegetation	0003 / 21445 / 21449 // 21453	A11 / A3A20 / A4A20 // A2A20	Lands with a mosaic of croplands, forests, shrubland, and grasslands in which no one component comprises more than 60 % of the landscape.	It was impossible to define a spatially mixed class with more than 3 parts, therefore the last class was added as thematic mix. An alternative could be to define it as layer. Both remedies are not actually correct.
Snow and Ice	Snow // Ice	8005 // 8008	A2 // A3	Lands under snow/ice cover throughout the year.	
Barren	Bare Area(s)	0011	B16	Lands with exposed soil, sand, rocks, or snow and never has more than 10 % vegetated cover during any time of the year.	LCCS specifies < 4 % vegetation cover for bare areas, snow is actually not included
Water Bodies	Artificial Waterbodies // Natural Waterbodies	7001 // 8001	A1 // A1	Oceans, seas, lakes, reservoirs, and rivers. Can be either fresh or salt-water bodies.	

Notes: The LCC-Code is unique numerical identifier for each LCCS category based on any possible combination of classifiers. The LCC-code was intended for use in GIS to allow consistent and transparent coding of LCCS classes. The LCCLevel describes a Boolean formula showing each classifier used for this class. For example the first IGBP class "Evergreen Needleleaf Forests" has a formula of "A3A10B2XXD2E1." The first level "A3" stands for trees; "A10" for a cover of >65%; "B2" for a possible height of "3–30 m;" XX for a blank, hence, an unused classifier (in this case spatial distribution); "D2" for needle-leaved; and "E1" for evergreen. The use of "/" or "//" show cartographic or thematic mixture between individual land cover classes (see Fig. 5). For information, please refer to the LCCS-1 manual available at: http://www.africover.org/download/manuals/LCCSMS20.pdf

(e.g., woody savannah versus woodland with herbaceous layer). The IGBP class name and category descriptions are maintained in the translation. Specific translation issues and problems are documented, e.g., some inconsistencies remain like tree density of 60% (IGBP) versus 65% (LCCS) and the minimum tree height of 2 m (IGBP), which does not match the 3-m definition of a tree within LCCS. However, it can be assumed that such differences cause only small inconsistencies among global maps.

Several ongoing case studies fostered through GOFC-GOLD are currently being completed to also gain harmonization exercises between global and regional datasets in specific test sites, i.e., in Siberia, Mongolia, Thuringia/Germany, and others. Translations, comparative analysis, and evaluation of the translated products are based on local expertise and in situ data. Based on these experiences, it has become apparent that every legend can be translated into the LCCS framework if sufficient information about the category definitions is available from the original interpreter.

Despite successful examples for legend translation, this step only resolves some inconsistencies between semantics of land cover datasets. The process of legend translation also highlights differences between legends and shows which classes can be harmonized (and, thus, directly compared) and where legends show inconsistencies. Different classifier definitions or thresholds (e.g., density and minimum height of trees) remain problematic and cannot be resolved through legend translation after the maps have been completed. Common methods to overcome these problems usually results in a loss of detail or information. Additional information might be needed to overcome these inconsistencies and to compare different datasets. In this context, vegetation continuous field products can also play an important role in harmonization of global and regional land cover products when used in a parametric manner and based on ordinal classification. Given a clear definition of the basic vegetation traits (e.g., trees), these products could be incorporated into existing classifications and used to explain and overcome some of the definition differences. A case study example on exploring the synergy between existing global mapping products is presented in this issue [24].

The translation provides only one step in the harmonization process. Other issues still remain, e.g., from cartographic standards, or from unresolved differences in classifier definitions. For existing maps, harmonization will, in most cases, not remove all inconsistencies. Limitations are expected for *a posteriori* harmonization or retro-fitted intercomparison of existing legends even after a successful translation. Hence, harmonization exercises for current land cover datasets should mainly focus on outlining these inconsistencies and try to improve understanding of how and why these datasets are not directly comparable. The current case studies will provide and document some translations (e.g., IGBP, CORINE, IPCC, etc.) including limitations and capabilities of the harmonization mechanisms and the use of LCCS. This information is helpful for future projects and the update of LCCS. Future harmonization efforts should also pay particular attention to these results for upcoming land cover mapping efforts.

C. Influence on Future Mapping Projects

Harmonization efforts will be most effective when applied in the development phase of projects producing land cover datasets. In fact, the whole initiative can only be successful if the harmonization efforts influence operational land cover data collection. Mapping projects are encouraged to profit from resources and harmonization experiences, especially in terms of identified problems and inconsistencies in existing legends. The harmonization strategy is endorsed through the international panels (e.g., GTOS) and the space agencies. There is a motivation for developers of land cover products, especially in the science community, and more particularly in the national development and planning agencies, to seek comparability and develop their legends from a standard classification scheme like LCCS. This should be especially true for land cover products that are derived for a variety of purposes and not for a specific user or application.

Some land cover products are already LCCS compatible e.g., GLC2000. The United Nations Global Land Cover Network as well as GOFC-GOLD is in process of communicating with active mapping projects to implement the harmonization objectives. One major effort is the new ESA initiative GLOBCOVER launched in 2004. This project is an international collaboration between ESA, FAO and UNEP—through GLCN, JRC, IGBP, and GOFC-GOLD with the objective to produce a global land cover map for the year 2005, based on fine-resolution (300 m) MERIS data, a sensor on the ENVISAT satellite. This new product is intended to update and to complement the other existing comparable global products, such as the global land cover map for the year 2000 (GLC 2000).

VI. Validation of Global Land Cover Maps

Validation is important both during the production and after the completion of land cover maps. Validation exercises have been completed to varying degrees for individual land cover datasets. Until recently, the IGBP DISCOVER product was the only global products that can be considered thoroughly validated [1]. Further progress in this arena is presented by Mayaux *et al.* [2] that provide the analysis of a thorough validation of the GLC2000 global dataset based on LCCS. The accuracies of other global datasets have been indicated by initial validation efforts based on limited resources [25]. There were further verification and comparisons in specific case studies [3], [26]. These studies have certainly improved our understanding of accuracy and comparability between these datasets. However, a joint accuracy assessment, statistically robust and consistent for the whole globe, is still missing to date.

Validation of global land cover datasets is a challenging effort due to the high frequency of mixed pixels at the resolutions currently used for mapping, difficulty in precise geolocation of map products and reference materials, and logistical difficulties associated with reference data collection. Hence, a coordinated international effort is required for such a task. The methodological background for validation of global land cover datasets has been provided by the group on calibration and validation (Cal/Val) of the Committee Earth Observation Satellites. Previous validation experiences and the scientific state of the art for robust accuracy assessment of regional and global land cover datasets was recently summarized in "best practice" document [27]. A set of core analysis methods exist for accuracy assessment, which should be routinely adopted as a baseline for reporting map accuracy. The recommended approach is to develop a "living" dataset of validation sites that could be used to verify any new global land cover map. Previous efforts have resulted in several land cover test site datasets [28]. They are currently the only source of validation sites and provide important information for current verification processes. However, for a joint initiative to validate all existing and future global land cover datasets a new set of validation sites is needed to provide a statistically robust, consistent, harmonized, updated, and accessible reference database. The intent is to select the sites in such a way that they are not associated with any specific land cover map and that they may retain statistically rigor when used in conjunction with a variety of maps. The validation will be based on high-resolution satellite data like Landsat ETM and others. For example, global Landsat TM/ETM mosaics exist for the years 1990 and 2000. Continued observations on this scale are essential to maintain the keep the reference database "living" and up to date. Such a dataset will allow continued assessment of the accuracy and validity of datasets even after many years of their production. For the image interpretation task, GOFC-GOLD maintains a world wide system of regional networks with local remote sensing/validation experts. Moreover FAO and UNEP through GLCN are implementing a significant outreach program to support capacities for land cover mapping production and validation at national and regional scales.

A key component for a joint validation is the harmonization framework. A common language and understanding of semantic differences between existing land cover dataset is essential for comparative analyses of accuracy. Thus, the harmonization strategy described earlier and the validation are parallel efforts. For example, in situ or reference data for accuracy assessment need to be acquired in a standardized way to be of value and comparable for a variety of land cover datasets and legends on different scales. Hence, the individual validation site interpretations will be generic descriptions of land cover characteristics and independent of any land cover legend in LCCS language. LCCS will assign the right land cover class based on the different implemented legend translations. This makes the validation process transparent, consistent, and applicable to any land cover map compatible with LCCS. The resulting comparative validation will move forward the degree of harmonization and interoperability of land cover datasets. A joint accuracy assessment increases the comparability of datasets and understanding of quality and possible problems. An understanding of joint dataset uncertainties is essential for synergetic use and the combination and aggregation of different land cover information. Such information is important to users and any application and will improve the knowledge available for future mapping efforts.

VII. Conclusion

The outline presented here is an invitation to entities involved in global earth observations of land to participate and benefit from a combined effort to bring global land cover datasets to an improved level of standardization which will enhance their ultimate utilization. The political framework, the organizations for international cooperation, and the methodological resources to support a joint harmonization and validation initiative for land cover datasets is fast developing, as exemplified by UNFCCC related decisions in this direction. It is now up to the individual members of the community to provide their share in this initiative. Previous efforts have suffered from a lack of funding for harmonization and validation. This new initiative offers the framework to overcome these problems.

- The space agencies and major users of global land cover datasets should support the initiative with resources and funding. Considering the investments in earth observation and development of the datasets, the required resources are quite small, particularly if the effort is a true international cooperation with several potential sources of funding.
- Producers of land cover datasets, such as national/international mapping programs, land cover facilities, and the scientific community are encouraged to follow the guidelines for harmonization of land cover datasets and to engage in the standards advocated by the GLCN. This suggests some changes in current practice. The flexibility of the harmonization mechanisms, however, respects the previous investment in existing data. Basically, current data acquisitions can be continued using existing data systems and many available nomenclatures. Changes such as using LCCS standards should be discussed and considered to permit separate data sets to be combined more easily than at present, possibly with minimal effort for data originators and without significant information loss. For example, there could be an effort for joint reprocessing of global datasets with harmonized legends or, once validated, a combination of the different maps that is more accurately tailored toward specific applications. The earth observation data exist to map land cover changes on global scales over the last 20 years.
- An essential requirement for success of this initiative is continuity in satellite observations. This includes both global land cover mapping sensors (e.g., MODIS, MERIS, etc.) and higher spatial resolution systems (such as Landsat) for continuous accuracy assessment and wall to wall inventories.
- The science community and development planning agencies should further explore methods for dealing with inconsistent land cover semantics and pursue *a posteriori* harmonization.

Again, a framework for this joint initiative is in place and the earth observation community is encouraged to participate and contribute in this important step toward operational mapping of land. So far, the efforts are focused on land cover, approaches to land use classification should be explored as well and such efforts could follow a similar avenue to the one presented here.

References

[1] J. Scepan, "Thematic validation of high-resolution global land-cover data sets," *Photogramm. Eng. Remote Sens.*, vol. 65, pp. 1051–1060, 1999.

[2] P. Mayaux, H. Eva, J. Gallego, A. Strahler, M. Herold, A. Shefali, S. Naumov, E. De Miranda, C. Di Bella, D. Johansson, C. Ordoyne, I. Kopin, and A. Belward, "Validation of the Global Land Cover 2000 Map," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 7, pp. 1728–1739, Jul. 2006.

[3] R. Latifovic and I. Olthof, "Accuracy assessment using sub-pixel fractional error matrices of global land cover products derived from satellite data," *Remote Sens. Environ.*, vol. 90, pp. 153–165, 2004.

[4] (2004, Oct.) GCOS Implementation Plan for the Global Observing System for Climate in Support of the UNFCCC. GCOS-92, WMO Tech. Document 1219, Geneva, Switzerland. [Online]. Available: http://www.wmo.ch/web/gcos/gcoshome.html

[5] (2005) The Global Earth Observation System of Systems (GEOSS) 10-Year Implementation Plan and Reference Document. GEO-Group on Earth Observation. [Online]. Available: http://earthobservations.org

[6] G. W. Withee, D. B. Smith, and M. B. Hales, "Progress in multilateral earth observations: CEOS, IGOS and the ad hoc group on earth observations," *Space Policy*, vol. 20, pp. 37–43, 2004.

[7] R. Laurini, "Spatial multidabase topological continuity and indexing: a step toward seamless GIS data interoperability," *Int. J. Geograph. Inf. Sci.*, vol. 12, no. 4, pp. 373–402, 1998.

[8] J. R. G. Townshend and C. O. Justice, "Toward operational monitoring of terrestrial systems by moderate-resolution remote sensing," *Remote Sens. Environ.*, vol. 83, no. 1–2, pp. 351–359, 2002.

[9] J. T. Morisette, J. L. Privette, A. Strahler, P. Mayaux, and C. O. Justice, "An approach for the Validation of Global Land Cover Products through the Committee on Earth Observing Satellites," in *Remote Sensing and GIS Accuracy Assessment Lunetta*, J. G. Lyon, Ed. Boca Raton, FL: CRC Press, 2003.

[10] E. Sadeh, "A failure of international space cooperation: the international earth observing system," *Space Policy*, vol. 18, pp. 135–150, 2002.

[11] R. Ohlemacher, "The earth observation summit: heralding a new era for international cooperation," *Space Policy*, vol. 19, pp. 277–281, 2003.

[12] M. Herold, K. Neumann, and C. Schmullius. (2004, Mar.) Report on the Land Cover Implementation Team Meeting. GOFC-GOLD Rep. Ser. 18, Jena, Germany. [Online]. Available: http://www.fao.org/gtos/gofc-gold/series.html. 2004

[13] M. Herold and C. Schmullius. (2004, Jul.) Report on the Harmonization of Global and Regional Land Cover Products, Workshop Rep. FAO. GOFC-GOLD Rep. Ser. 20, Rome, Italy. [Online]. Available: http://www.fao.org/gtos/gofc-gold/series.html.2004

[14] M. Köhl, B. Traub, and R. Päivinen, "Harmonization and standardization in multi-national environmental statistics—mission impossible?," *Environ. Monitoring Assessment*, vol. 63, no. 2, pp. 361–380, 2000.

[15] A. Annoni and P. Smits, "Main problems in building European environmental spatial data," *Int. J. Remote Sens.*, vol. 24, no. 20, pp. 3887–3902, 2003.

[16] (1998) World Reference Base for Soil Resources. FAO/ISRIC/ISSS, World Soil Resources Rep., 84. [Online]. Available: http://www.fao.org/docrep/W8594E/W8594E00.htm

[17] Y. Bishr, "Overcoming the semantic and other barriers to GIS interoperability," *Int. J. Geograph. Inf. Syst.*, vol. 12, no. 4, pp. 299–314, 1998.

[18] "Report of the UNEP/FAO Expert Meeting on Harmonizing Land Cover and Land Use Classifications—Geneva, 23–25 Nov. 1993," FAO/UNEP, GEMS Rep. Ser. 25, 1994.

[19] M. Kavouras and M. Kokla, "A method for the formalization and integration of geographical categorizations," *Int. J. Geograph. Inf. Sci.*, vol. 16, no. 5, pp. 439–453, 2002.

[20] O. Ahlqvist, J. Keukelaar, and K. Oukbir, "Rough and fuzzy geographical data integration," *Int. J. Geograph. Inf. Sci.*, vol. 17, no. 3, pp. 223–234, 2003.
[21] C.-C. Feng and D. M. Flewelling, "Assessment of semantic similarity between land use/land cover classification systems," *Comput., Environ., Urban Syst.*, vol. 28, pp. 229–246, 2004.
[22] A. J. Comber, P. Fisher, and R. Wadsworth, "Integrating land cover data with different ontologies: identifying change from inconsistency," *Int. J. Geograph. Inf. Sci.*, vol. 18, no. 7, pp. 691–708, 2004.
[23] L. J. M. Jansen and A. Di Gregorio, "Parametric land cover and land-use classification as tools for environmental change detection," *Agricult., Ecosyst., Environ.*, vol. 91, pp. 89–100, 2002.
[24] L. See and S. Fritz, "Toward a global hybrid land cover map for the year 2000," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 7, pp. 1740–1746, Jul. 2006.
[25] (2003) Validation of the Consistent Year V003 MODIS Land Cover Product. [Online]. Available: http://geography.bu.edu/landcover/index.html
[26] M. Hansen and B. Reed, "Comparison of IGBP DISCover and University of Maryland 1 km global land cover classifications," *Int. J. Remote Sens.*, vol. 21, pp. 1365–1374, 2000.
[27] A. Strahler, L. Boschetti, G. M. Foody, M. A. Fiedl, M. C. Hansen, M. Herold, P. Mayaux, J. T. Morisette, S. V. Stehman, and C. Woodcock, "Global Land Cover Validation: Recommendations for Evaluation and Accuracy Assessment of Global Land Cover Maps," Report of Committee of Earth Observation Satellites (CEOS-WGCV).
[28] J. Morisette, J. L. Privette, and C. Justice, "A framework for global land product validation," *Remote Sens. Environ.*, vol. 83, pp. 77–96, 2002.

Martin Herold (S'98) was born in Leipzig, Germany, in 1975. He received the the graduate degree (Diplom in Geography) from the Friedrich Schiller University of Jena, Jena, Germany, and the Bauhaus University of Weimar, Weimar, Germany, in 2000, and the Ph.D. degree from the Department of Geography, University of California, Santa Barbara, in 2004.

He is currently coordinating the ESA GOFC GOLD Land Cover project office at the Friedrich Schiller University Jena. Earlier in his career, his interests were in multifrequency, polarimetric and interferometric SAR-data analysis for land surface parameter derivation, and modeling. He joined the Remote Sensing Research Unit, University of California Santa Barbara, in 2000, where his research has focused on the remote sensing of urban areas and the analysis and modeling of urban growth and land use change processes. His most recent interests are in international coordination and cooperation toward operational terrestrial observations with specific emphasis on the harmonization and validation of land cover datasets.

Dr. Herold is a member the German Society of Photogrammetry and Remote Sensing (DGPF) and the Thuringian Geographical Union (TGG).

Curtis E. Woodcock received the B.A., M.A., and Ph.D. degrees in geography from the University of California, Santa Barbara.

He is currently a Professor with the Department of Geography and Environment, Boston University, Boston, MA. His interests are in the remote sensing of forests and the monitoring of environmental change. He has previously served on the Landsat Science Team. He is currently a member of NASA's Land Cover and Land Use Change Science Team, NASA's Land Measurement Team, NEESPI Science Team, and the GOFC/GOLD Land Cover Characteristics and Changes Implementation Team.

Antonio Di Gregorio was born in Teramo, Italy, in 1951. He received the Doctorate of Biology (Botany, Zoology, Ecology) from the University of Studies, Florence, Italy, in 1977, and the Diploma in rural survey and rangelands ecology at the ITC, Enschede, The Netherlands, in 1981.

His professional activity is mainly focused on the combined use of remote-sensing and GIS techniques for the assessment, mapping, and monitoring of natural resources and green biomass estimation for grazing capacity and ecological monitoring. For more than 20 years, he has been involved in the implementation and execution of UN Projects all over the world. He is the main author of the Land Cover Classification System (LCCS) that has become the FAO/UNEP official Land Cover Classification, as well as a worldwide standard.

Philippe Mayaux was born in 1963 in Belgium. He received the Engineer in Forestry degree in 1986, the M.S. degree in land planning in 1988, and the Ph.D. degree in 1999 from the University of Louvain, Louvain, Belgium.

In the early 1990s, he worked for three years in field projects in Africa. He has worked for over 15 years in land cover mapping and tropical ecosystem monitoring from satellite imagery. He coordinated the African TREES project's activities, produced the Global Land Cover 2000 Map of Africa and Australia and is involved in many projects dealing with biodiversity and land cover changes, in particular, in Central Africa. His research focused also on the scaling issues in remote sensing and on validation of regional to global scale land cover products.

Dr. Mayaux is member of the Scientific and Technical Board of the GOFC-GOLD project.

Alan S. Belward received the B.Sc. degree in plant biology from the University of Newcastle upon Tyne, U.K., in 1981, and the M.Phil. and Ph.D. degrees in the remote sensing studies of vegetation from the School of Agriculture Food and Environment, Cranfield University, U.K., in 1986 and 1993, respectively.

He is with the European Commission at the Joint Research Centre, Institute for Environment and Sustainability, Ispra, Italy, where he is head of the Global Environment Monitoring Unit. He co-chaired the International Geosphere Biosphere Program's Land Cover Working Group during its work to create the first satellite derived global land cover data set in the 1990s and was the Chairperson of the G7 Committee for Earth Observing Satellites Working Group on Calibration and Validation from 1996 to 2000. He has been Vice Chairperson of the UN-sponsored Global Terrestrial Observing System's science panel dealing with Global Observations of Forest Cover since its creation in 2000, and he has chaired the UN Global Climate Observing System's Terrestrial Observing Panel for Climate since January 2001.

John Latham is Remote Sensing Officer of the Environment and Natural Resources Service of the Food and Agriculture Organization of the United Nations (FAO), Rome, Italy. He is the Program Director of the Global Terrestrial Observing System (GTOS), an international program for the observation, modeling, and analysis of terrestrial ecosystems to support sustainable development. He is also Project Coordinator of the Global Land Cover Network (GLCN), a program to increase the availability of reliable and standardized information on land cover and its changes at the global, regional, and national levels. During the past 20 years, he has been involved in the research, development, application, technology transfer, and capacity building programs in the use of information management systems based on the use of geographical information system and remotely sensed data from earth observation satellites as related to environmental and natural resources management programs in developing and developed countries in the Middle East, Africa, Asia, Latin America, and Europe.

Christiane C. Schmullius received the M.A. degree in geography/remote sensing from the University of California, Santa Barbara, as a Fulbright and ITT Fellow, in 1986, and the Ph.D. degree in radar remote sensing of agricultural crops from the Institute of Space Sciences, Free University Berlin, Berlin, Germany, in 1991.

From 1991 to 2000, she was group leader "Geomatics" at the German Aerospace Center, DLR. In 1994, she coordinated the SIR-C/X-SAR Science Team and in 2000 became DLR's Project Scientist for the Shuttle Radar Topography Mission (SRTM). She has coordinated several EU projects in Siberia, Russia, especially on radar interferometry applications for climate change parameter mapping. Since 2000, she has been a Professor of remote sensing at the Friedrich-Schiller University, Jena, Germany.