



Convention on Biological Diversity

Distr.
GENERAL

UNEP/CBD/NP/COP-MOP/1/INF/8
18 September 2014

ORIGINAL: ENGLISH

CONFERENCE OF THE PARTIES TO THE CONVENTION
ON BIOLOGICAL DIVERSITY SERVING AS THE
MEETING OF THE PARTIES TO THE NAGOYA
PROTOCOL ON ACCESS TO GENETIC RESOURCES
AND THE FAIR AND EQUITABLE SHARING OF
BENEFITS ARISING FROM THEIR UTILIZATION

First meeting

Pyeongchang, Republic of Korea, 13-17 October 2014

Item 8 of the provisional agenda*

IMPLEMENTING THE NAGOYA PROTOCOL IN MICROBIOLOGY: GAINING TRUST, BUILDING TRUST

Note by the Executive Secretary

1. The Executive Secretary is circulating herewith, for the information of participants in the first meeting of the Conference of the Parties serving as the meeting of the Parties to the Nagoya Protocol, a document on implementing the Nagoya Protocol in Microbiology and the Transparent User-friendly System of Transfer prepared by the World Federation for Culture Collections (WFCC).
2. The document is being circulated in the form and language in which it was received by the Secretariat.

* UNEP/CBD/NP/COP-MOP/1/1.



WORLD FEDERATION FOR CULTURE COLLECTION

Brussels, 15 September 2014

Dr Braulio Ferreira de Souza Dias

Executive Secretary of the Convention on Biological Diversity
United Nations Environment Programme
413 Saint-Jacques Street, Suite 800
H2Y 1N9 Montreal, Quebec
Canada

Subject: Submission by WFCC of an information document for MOP-1

Dear Sir,

I have the pleasure to submit an information document on behalf of the World Federation for Culture Collections. It is the result of collaborative efforts and dialogue between experts of different fields and your staff.

I am particularly proud to announce that WFCC can propose a cost-efficient, simple, fast and multiple users - multiple purposes system that may be connected to the ABS-CHM. The breakthrough is the development of the Global Catalogue of Microorganisms by the team of Dr Juncal MA, the World Data Centre for Microorganisms.

WFCC members will present the outcomes of their initiative at three side events during MOP-1. These side-events will be complementary and at the same time independent, so that people attending one side-event can have a full view of the proposal but from different perspectives. During these events, among others, we will show:

- how to connect the information system of WFCC to the CHM,
- how to implement the concepts of the EU regulation on the Nagoya Protocol, at a global level,
- how to keep a flexible system adaptable to unexpected situations.

The Culture Collections community thanks you and your team for the collaborative spirit in which WFCC members and your staff are cooperating.

I hope to have again the privilege to meet you and your team in Korea.

Yours sincerely,

Philippe Desmeth
President

President

Philippe Desmeth
International Co-operation Manager
Belgian Co-ordinated Collections of
Microorganisms
c/o Belgian Science Policy Office
Avenue Louise, 231
Brussels B-1050
Belgium
Tel: +32 2 238 3713
Fax: +32 2 230 5912
Email: philippe.desmeth@belspo.be

Vice-President

Dr Ken-Ichiro Suzuki
Director
NITE BRC
National Institute of Technology and
Evaluation
2-5-8, Kazusakamatari, Kisarazu-shi,
Chiba 1292-0818, Japan
Fax: + 81 438 52 2329
Email: suzuki-ken-ichiro@nite.go.jp

Secretary

Dr Marizeth Groenewald
Curator
CBS-KNAW Fungal Biodiversity Centre
Uppsalalaan, 8
NL-3584 CT Utrecht, The Netherlands
Fax: + 31 30 2512097
E-mail: m.groenewald@cbs.knaw.nl

Treasurer

Dr Vera Bussas
Leibniz-Institut Deutsche Sammlung von
Mikro-organismen und Zellkulturen GmbH
7B Inhoffenstrasse,
38124 Braunschweig, Germany
Fax: +49 531 2616 225
E-mail: yew@dsMZ.de

WDCM-Director

Dr. Juncal Ma
Institute of Microbiology
Chinese Academy of Sciences
No 1 West Beichen Road,
Chaoyang District
Beijing 100101, CHINA
Tel.: +86 10 64807422
Fax: +86 10 64807426
E-mail: ma@im.ac.cn

Editor Newsletter

Dr Ipek Kurtboke
Senior Lecturer in Environmental
Microbiology
Faculty of Science, Health, Education and
Engineering
University of the Sunshine Coast
Maroochydore DC
Queensland 4558, Australia
Fax: +61 07 5430 2881
E-mail: ikurtboke@usc.edu.au

WFCC Home Page: <http://www.wfcc.info/>

Executive Board Members: Dr. Mohamed Amar (Morocco); Dr. Kyria Boundy-Mills (USA); Dr. Vera Bussas (Germany); Dr. Manuela Da Silva (Brazil); Philippe Desmeth (Belgium); Prof. Xiuzhu Dong (PR. China); Dr. Marizeth Groenewald (The Netherlands); Dr. Takashi Itoh (Japan); Dr. Ipek Kurtboke (Australia); Dr. Juncal MA (PR China); Dr. Kevin McCluskey (USA); Dr. G.S. Prasad (India); Dr. Oleg Stupar (Russia); Dr. Ken-Ichiro Suzuki (Japan)

/...

Implementing the Nagoya Protocol in microbiology Gaining trust, building TRUST

Note of the World Federation for Culture Collections

EXECUTIVE SUMMARY

The World Federation of Culture Collections (WFCC) has introduced a monitoring and tracking system for the utilisation of microbial genetic resources. This is the Global Catalogue of Microorganisms. It implements key provisions of the Nagoya Protocol and provides required transparency, legal certainty while lowering transaction costs and reducing administrative and governance burdens. This innovation is supported by a suite of supporting tools grouped into TRUST; namely, the MOSAICC code of conduct, MOSAICS integrated conveyance system, a ‘microbial commons’ material exchange concept and NIEMA, a system implementing this concept of “microbial commons” within Asia.

I. INTRODUCTION

1. Since the infancy of microbiology, scientists have struggled to ensure long-term *ex situ* conservation of living microbial material and enable their sustainable use. This specialised work, essential for microbiological research, is performed by culture collections. Culture Collections are infrastructures specialised in long term conservation of microbial resources and management of related data and information. Their mission is to provide facilitated access to technically and legally fit-for-use microbiological resources of consistent quality and related data. These facilities, hosted in private or public research institutions, are established all around the world and most of them are registered in the World Data Centre for Micro-organisms ([WDCM](#)) of the World Federation for Culture Collections ([WFCC](#)).

2. The new EU Regulation on ABS recognizes the important role of culture collections, as well as other *ex situ* conservation facilities such as botanical gardens and museums. It does this through its inclusion of the concept of "registered collections". These are collections accredited against specified criteria and thus trusted sources for legally sound biological raw material.

3. Culture collections themselves have also taken several initiatives to translate the ABS principles into practice. The outcomes of these initiatives comprise a set of coherent and complementary tools to implement the ABS concept:

- The [Global Catalogue of Microorganisms \(GCM\)](#), the data management system, the backbone of a global system linking the microbial genetic resources to all kind of related data (**see annex 1**). The flow of information generated by GCM will be connected to the ABS Clearing House Mechanism via machine-processed link (**see annex 2**).
- The [TRUST](#) guidelines providing for administrative and technical procedures. GCM and TRUST constitute the general management system facilitating access to microbial genetic resources (**see annex 3**). TRUST combines the Code of Conduct [MOSAICC](#) (Micro-organisms Sustainable use and Access regulation International Code of Conduct) and the [MOSAICS](#) Integrated Conveyance System.
- The design of “**microbial commons**”¹ for the exchange of (micro) biological material which would provide basic common use principles for access to both material and information.

¹ This development is complementary to national regulations on ABS and to existing IPR laws, as it will constitute a demarcated space where material and information are relatively freely accessible provided that the outputs are injected back into this open space, to be shared again. Inside this space access and benefit-sharing are “commonly shared”. Outside this demarcated space, access and benefit-sharing will be ruled through ordinary national and international laws, including IPR and specific CBD

- The **NIEMA** system is implementing this concept of "microbial commons" within the Asian Consortium for the Conservation and Sustainable Use of Microbial Resources (ACM). (see annex 4)
- The innovative concept of “**bundle of rights**”².

Although most have been developed before the Nagoya Protocol, these tools have been reviewed and refined to be relevant solutions for microbiologists implementing the Access and Benefit Sharing concept. The World Federation for Culture Collections and its members work towards the development of a global system incorporating these developments, combining adapted legal concepts and bioinformatics.

II. GAINING TRUST, BUILDING TRUST

4. TRUST stands for TRansparent User-friendly System of Transfer. It aims at managing the incidence of the CBD and Nagoya Protocol on the scientific, technical and administrative activities of culture collections and, more generally, incorporating the Nagoya Protocol into the daily life of microbiologists.

5. Several institutions specialized in microbiology, including culture collections, have decided to join forces. Because trust is a prerequisite for lasting cooperation in science and for socio-economic development and because trust can be attained partly through a transparent system of transfer of microbial material, they have decided to coordinate their efforts, to build TRUST, literally and practically. They are aiming at adjusting their tools to each other and at providing a cost-efficient, simple, fast and multiple users - multiple purposes global system.

6. Through the World Data Centre for Micro-organisms (WDCM)³ databases system and the assignment of Globally Unique Identifiers (GUIDs) specific to microbial items, WFCC contributes to building a transparent, safe and sustainable handling system of *ex situ* microbial diversity worldwide. A major programme has been launched by WDCM under the aegis of WFCC: the **Global Catalogue of Micro-organisms (GCM)**. GCM is a powerful scientific tool as well as a way to build safe, ethical and socio-economically balanced ABS processes at global level.

inspired regulations. A practical development similar to this model is the NIEMA system.

See <http://www.thecommonsjournal.org/index.php/ijc/article/view/215/144>

² The innovative concept of “bundle of rights” is a dynamic model of ownership management moving away from the static concept of ownership towards a flexible allotment of rights. Ownership constitutes a “bundle” of use and decision rights that are attributed to a number of stakeholders / economic agents. It is a set of operational and collective choice rights defining respectively who decides upon the use that one can make of a resource, and who decides upon the future exercise of the rights on the resource. Such scheme allows multi-ownership of a gradual level of use and decision rights. These rights can begin with basic access rights, encompassing research delivering outputs to the public domain, distribution on to third parties, exploitation rights to develop intellectual property and its ownership which may include reach through rights. The application of the “bundle of rights” makes possible the enforcement of the “sovereign rights of States over their natural resources” without prejudice to private rights. Unambiguous allotment of rights in advance will facilitate rightful benefit sharing “at the end of the pipe”. Dedeurwaerdere Tom. *Understanding ownership in the knowledge economy: the concept of the bundle of rights*. BCCM News Edition 18 - Autumn 2005.

³ WFCC has developed a pioneering database system by registering its members through a unique acronym and numerical identifier in its official directory. It also helps its members to catalogue their microbiological resources. This system is managed by the World Data Centre for Micro-organisms (WDCM). Combining the WDCM system and the use of electronic markers called “Globally Unique Identifiers (GUIDs)” set up a robust system to organise transfers of (micro) biological items, tracking the flow of resources and related information. This system also facilitates the application of ABS since it can potentially retrieve all kinds of information about microbiological resources, including information related to the location and movements of the resource. The WDCM portal acts as an information broker between all online catalogue entries of the culture collections. See <http://www.wdcm.org> and http://bccm.belspo.be/documents/files/projects/mosaics/ics_report.pdf

7. TRUST is a modular system having as backbone the Global Catalogue of Microorganisms and making use of the expertise gained by MOSAICC, MOSAICS, and other initiatives. The TRUST system comprises four elements:

- i. Updated MOSAICC features with administrative **workflows** adapted to the structure of the Nagoya Protocol and improved in light of past experience. **The flow of information will be connected to the ABS Clearing House Mechanism⁴** via a machine-processed link from a defined IRCC⁵-CHM field to a defined IRCC-GCM field (see annex 2).
- ii. **Refined Material Accession Agreement (MAA) and Material Transfer Agreement (MTA) models** with standardized definitions. When considering the concept of "registered collections" as defined by the EU regulation implementing the Nagoya Protocol, culture collections community has responded by developing an efficient information and material handling strategy to deal with administrative handling of transfers of material, especially the compulsory Prior Informed Consent. Thus the concept of registered collections as future trusted sources for ABS-compliant genetic resources is imbedded into the culture collections community at the outset. Consequently it is of primary importance that:
 - a. Every microbial genetic resource "entering" a collection is covered by a PIC obtained at the time of its isolation from *in situ* conditions or after corrective administrative action.
 - b. Every microbial genetic resource having entered a collection with the appropriate initial PIC may be distributed, accompanied by the original PIC, without any additional PIC procedure set by the country of origin or the country of use.
- iii. **An automated powerful integrated data management and processing system** able to provide for any information related to microbial material: **the ground breaking Global Catalogue of Microorganisms (GCM)**. The CGM provides users with the ability to trace the possession, location, transmission and use of uniquely-identified microbial strains, including country of origin, existence of PIC and MAT, the creation of derived patents and all associated scientific publications. This system already includes more than 60 collections from more than 30 countries and information on nearly 290.000 strains from 41.000 species.
- iv. **Cooperative structure** wherein culture collections
 - a. make use of the latest ICT technology to develop the necessary identification and tracking system, primarily for scientific purposes but also for any other bona fide ends.
 - b. conduct and facilitate research in genomics and functional genomics, thus develop capacities of storage and processing of genomic, transcriptomic and metabolomic information. These compiled data improve definite characterization of microbial resources.
 - c. conduct their efforts in networks, in conformity with NP provisions on Technology Transfer, collaboration and cooperation.

III. Conclusion

8. The best way to achieve ABS with effective socio-economic benefits is to build on existing procedures, to make the appropriate linkages between the various actors, and provide for the necessary incentives to the users so that ABS is effectively more beneficial to all and does not require coercive measures or penalties.

⁴ <http://absch.cbd.int>

⁵ Internationally Recognised Certificate of Compliance

9. Conducted increasingly in networks, and in conformity with the Nagoya Protocol provisions on technology transfer, collaboration and cooperation, the efforts of the culture collections community are already bearing fruits via GCM. It sets up a global network of trusted microbiological resources centers. These efforts foster cumulative research by linking scientific publication, microbial material and data via referencing of microbial strains through the use of GUIDs such as strain number. This integrated information platform also contributes to bio-economy by facilitating production of added value and socio-economic profit out of microbiological material that can be traced back to its original provider.

10. The TRUST programme and the complementary initiatives are executed with the voluntary contribution of experts and the participation of other invited contributors. It is based on a major pioneering development in biological diversity management: the [Global Catalogue of Microorganisms](#), backbone of efficient exploration, conservation and exploitation of microbiological diversity and, from a broader perspective, an essential ICT part in microbiology.

11. These developments made in the microbiological field can be adapted to other fields of research, also enhancing interdisciplinary study of biological diversity. They are aiming at providing a cost-efficient, simple, fast and multiple users - multiple purposes global system.

This document has been prepared with the contribution of the following experts: Philippe Desmeth (1), Juncal Ma (2), Linhuan Wu (3), Geoff Burton (4), Katsuhito Ando (5), Rie Funabiki (6), Chris Lyal (7), Lyle Glowka (8), Mamadou Gueye (9), Tae-Eun Jin (10), Tom Dedeurwaerdere (11), Ken-Ichiro Suzuki (12).

- (1) President, World Federation for Culture Collections, c/o Belgian Coordinated Collections of Micro-organisms, Belgian Science Policy Office, avenue Louise, 231 1050 Brussels, Belgium. Email: philippe.desmeth@belspo.be;
- (2) Director, World Data Center of Microorganisms, Institute of Microbiology, Chinese Academy of Sciences, 1-3 West Beichen Road, Chaoyang District, Beijing 100101, P. R. China. E-mail: ma@im.ac.cn
- (3) Senior scientist; ICT, World Data Center of Microorganisms, Institute of Microbiology, Chinese Academy of Sciences, 1-3 West Beichen Road, Chaoyang District, Beijing 100101, P. R. China. E-mail: wulh@im.ac.cn
- (4) Adjunct Senior Fellow, United Nations University Institute of Advanced Studies of Sustainability, 5-53-70 Jingumae Shibuya-ku, Tokyo 150-8925 Japan, Tel: +81 3 5467-1212. E-mail: geoff@jeanshannon.com or burton@ias.unu.edu
- (5) Acting Director-General, Biological Resource Center, National Institute of Technology and Evaluation (NITE), 2-5-8 Kazusakamatari, Kisarazu, Chiba 292-0818 Japan. Email: ando-katsuhiko@nite.go.jp
- (6) Chief, Office for Global Initiatives, Biological Resource Center, National Institute of Technology and Evaluation (NITE), 2-5-8 Kazusakamatari, Kisarazu, Chiba 292-0818 Japan. Email: funabiki-rie@nite.go.jp
- (7) Department of Life Sciences, The Natural History Museum, Cromwell Road London SW 7 5BD, UK; Email: C.lyal@nhm.ac.uk
- (8) Executive Coordinator, Convention on Migratory Species UNEP, Abu Dhabi Office, c/o Environment Agency - Abu Dhabi, P.O. Box 45553, Abu Dhabi, United Arab Emirates Email: cmsoffice.ae@cms.int
- (9) Professor, University of Sine Saloum Kaolack (USSK), Member of the Academy of Sciences, Senegal Email: mgmgueye@gmail.com
- (10) Department of ABS research support, Korea Research Institute of Bioscience and Biotechnology, 125 Gwahak-ro, Yuseong, Daejeon, Korea; E-mail: tejin@kribb.re.kr
- (11) Department of Philosophy of Law, Université catholique de Louvain, Belgium. Email: tom.dedeurwaerdere@uclouvain.be
- (12) Director, Biological Resource Center (NBRC), National Institute of Technology and Evaluation 2-5-8 Kazusakamatari, Kisarazu-shi, Chiba, 292-0818 Japan Email: suzuki-ken-ichiro@nite.go.jp

ANNEX 1

Global catalogue of microorganisms (GCM): a comprehensive database and information retrieval, analysis, and visualization system for microbial resources

The increasing demands on culture collections for authenticated, reliable biological material and associated information have paralleled the growth of biotechnology.

In the WFCC guidelines, it is pointed out that collections should publish online or printed catalogue to disseminate information of strains to promote scientific and industrial usage of holdings.

However, according to the statistics, only one-sixth of collections registered in [CCINFO](#) have their online catalogue, which greatly hinders the visibility and hence the accessibility of strains.

The World Data Centre for Microorganisms has constructed a data management system and a global catalogue to help organize, unveil and explore the data resources of its member collections.

The WFCC Global Catalogue of Microorganisms (GCM) is expected to be a robust, reliable and user-friendly system to help culture collections to manage, disseminate and share the information related to their holdings. It also provides a uniform interface for the scientific and industrial communities to access the comprehensive microbial resource information. The development of specific softwares started with the straininfo.net initiative, now further extended in GCM.

DATABASE

Open Access

Global catalogue of microorganisms (GCM): a comprehensive database and information retrieval, analysis, and visualization system for microbial resources

Linhuan Wu^{1,2}, Qinglan Sun^{1,2}, Hideaki Sugawara⁴, Song Yang^{1,2}, Yuguang Zhou¹, Kevin McCluskey⁵, Alexander Vasilenko⁶, Ken-Ichiro Suzuki⁷, Moriya Ohkuma⁸, Yeonhee Lee⁹, Vincent Robert¹⁰, Supawadee Ingsriswang¹¹, François Guissart³, Desmeth Philippe^{3*} and Juncai Ma^{1,2*}

Abstract

Background: Throughout the long history of industrial and academic research, many microbes have been isolated, characterized and preserved (whenever possible) in culture collections. With the steady accumulation in observational data of biodiversity as well as microbial sequencing data, bio-resource centers have to function as data and information repositories to serve academia, industry, and regulators on behalf of and for the general public. Hence, the World Data Centre for Microorganisms (WDCM) started to take its responsibility for constructing an effective information environment that would promote and sustain microbial research data activities, and bridge the gaps currently present within and outside the microbiology communities.

Description: Strain catalogue information was collected from collections by online submission. We developed tools for automatic extraction of strain numbers and species names from various sources, including Genbank, Pubmed, and SwissProt. These new tools connect strain catalogue information with the corresponding nucleotide and protein sequences, as well as to genome sequence and references citing a particular strain. All information has been processed and compiled in order to create a comprehensive database of microbial resources, and was named Global Catalogue of Microorganisms (GCM). The current version of GCM contains information of over 273,933 strains, which includes

43,436 bacterial, fungal and archaea species from 52 collections in 25 countries and regions.

A number of online analysis and statistical tools have been integrated, together with advanced search functions, which should greatly facilitate the exploration of the content of GCM.

Conclusion: A comprehensive dynamic database of microbial resources has been created, which unveils the resources preserved in culture collections especially for those whose informatics infrastructures are still under development, which should foster cumulative research, facilitating the activities of microbiologists world-wide, who work in both public and industrial research centres. This database is available from <http://gcm.wfcc.info>.

Keywords: Microbial resources, Data management, Data sharing

* Correspondence: Philippe.DESMETH@belspo.be; ma@im.ac.cn

³Belgian Coordinated Collections of Micro-organisms Programme, Belgian Science Policy Office, avenue Louise, 231 1050 Brussels, Belgium

¹Institute of Microbiology, Chinese Academy of Sciences, Beijing, China

Full list of author information is available at the end of the article

Background

Microbial culture collections play an important and essential role in collecting, maintaining, and distributing quality assured living microbial strains. The World Federation for Culture Collections (WFCC) is a Multi-disciplinary Commission of the International Union of Biological Sciences (IUBS) and a Federation within the International Union of Microbiological Societies (IUMS). The WFCC promotes the interests of culture collections, develops shared resources, and organizes the International Conference on Culture Collections every three years. As one of its longstanding activities, the WFCC participated in the development of the WFCC World Data Centre for Microorganisms (WDCM) in the late 1960s [1]. With additional input from the United Nations Educational, Scientific and Cultural Organization Microbial Resources Centers (MIRCEN) project, the WDCM was maintained as the WFCC-MIRCEN WDCM and became accessible as an internet page in 1997. The WDCM serves as the data center of the WFCC and provides an important information resource for all microbiological activities. Additionally, the WDCM acts as a coordination center for data activities among WFCC members. As one of the main databases in WDCM, CCINFO (Culture Collection INFORMATION database) lists 652 culture collections from 70 countries maintain more than 1.9 million strains. (<http://www.wfcc.info/ccinfo/>, accessed 12/3/2013).

Increasing demands on culture collections for authenticated, reliable biological material and associated information were accompanied by the growth of biotechnology and basic science. The WFCC guidelines recommend that every collection publish an online or printed catalogue regularly, both to disseminate information about strains and to promote scientific and industrial usage of materials held in their collection. However, according to the available statistics, fewer than one-sixth of collections registered in CCINFO post their catalogue online and this greatly hinders the visibility and hence the accessibility of strains in these collections without public electronic catalogs.

To help all collections establish an online catalog, the WDCM has constructed a data management system and a global catalogue to organize, make public, and explore the data resources of its member collections. This data management system, called the WFCC Global Catalogue of Microorganisms (GCM) is a scalable, reliable, dynamic and user-friendly system that helps culture collections manage, disseminate and share the information related to their holdings. It also provides a uniform interface for the scientific and industrial communities to access the comprehensive microbial resource information.

Construction and content

Data

sources

The Global Catalogue of Microorganisms database contains information from a variety of sources:

- Information provided by culture collection staff
- Data from public data sources such as the US National Library of Medicine (PubMed) and the Patent database
- Links to external databases
- Tools for bioinformatics analysis including a search engine to enhance exploration of GCM data.

By the end of August 2013, the GCM contains strain information from 52 collections (Table 1) located in 25 different countries and regions. While the project is still in its construction phase, preliminary statistics describing the participating collections are unique and informative (Table 2).

The GCM implements the WDCM Minimum Data Sets (MDS) and Recommended Data Sets (RDS) based on widely applied standards such as the OECD Best Practice Guidelines for Biological Resource Centres [2], the Microbial Information Network Europe (MINE) [3], as well as the Common Access to Biological Resources and Information (CABRI) [4]. A detailed description, together with examples of 15 WDCM MDS items can be found at <http://gcm.wfcc.info/datastandards/index.jsp> (last accessed 12/3/2013). To build the GCM, each participating collection transferred their catalogue information by one of several pathways. Some collections sent Excel or XML files while others provided direct access to their database files. WDCM integrated the data into a global dataset, processed the data to identify relationships among collections (for example strains held in multiple collections), and published the strain information on the GCM web page (<http://gcm.wfcc.info>). Because not all collections use the same data schema, some of the data items provided by culture collection staff were manually reclassified by GCM staff to allow for an easier integration of catalogue information.

Publications concerning strains are collected from PubMed using both strain number and species name for keyword queries. Nucleotide sequences are extracted from GenBank [5], protein sequence data are collected from UniProt [6], and information about protein 3D structure are extracted from the PDB database [7]. Genome sequencing information is collected from NCBI Microbial Genomes Resources (NCBI).

Organization of data

The GCM database contains the following fields for each strain entry: strain number, other collection numbers, name, organism type, history of deposition, date of

Table 1 Participant list of GCM collections

Acronym	Full name	Country
BCC	BIOTEC Culture Collection	Thailand
BCCM/DCG	BCCM Diatom Collection Gent	Belgium
BCCM/IHEM	Belgian Coordinated Collections of Microorganisms / IHEM Fungi collection	Belgium
BCCM/LMBP	Belgian Coordinated Collections of Microorganisms / LMBP Plasmid Collection	Belgium
BCCM/LMG	Belgian Coordinated Collections of Microorganisms/ LMG Bacteria Collection	Belgium
BCCM/MUCL	Mycotheque de l'Universite catholique de Louvain	Belgium
BCCM/ULC	BCCM/ULC Culture Collection of (sub)polar cyanobacteria	Belgium
BCRC	Bioresource Collection and Research Center	Chinese Taipei
BIM	Belarusian Collection of non-pathogenic microorganisms	Belarus
CBS	Centraalbureau voor Schimmelcultures, Filamentous fungi and Yeast Collection	Netherlands
CCAP	Culture Collection of Algae and Protozoa	U.K.
CCARM	Culture Collection of Antimicrobial Resistant Microorganisms	Korea
CCCryo	Culture Collection of Cryophilic Algae	Germany
CECT	Coleccion Espanola de Cultivos Tipo	Spain
CGMCC	China General Microbiological Culture Collectio Center	China
CIP	The Collection of the Institut Pasteur	France
CIRM-CF	Centre International de Ressources Microbiennes - Champignons Filamenteux	France
CIRM-CFBP	Centre International de Ressources Microbiennes - Levures (CLBP)	France
CIRM-Levures	Centre International de Ressources Microbiennes - Levures	France
CM-CNRG	Coleccion de Microorganismos del Centro Nacional de Recursos Geneticos	Mexico
CVCM	Centro Venezolano de Colecciones de Microorganismos	Venezuela
CWU-MACC	Herbarium of Kharkov University (CWU) – Micro Algae Cultures Collection	Ukraine
DMic	Medical importance fungi culture collection	Argentina
DSMZ	Leibniz-Institut DSMZ-Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH	Germany
FACHB	Freshwater Algae Culture Collection, Chinese Academy of Sciences	China
FGSC	Fungal Genetics Stock Center	USA
Fiocruz-CLIOC	Coleção de Leishmania do Instituto Oswaldo Cruz	Brazil
GDMCC	Guangdong Culture Collection Centre of Microbiology	China
HPKTCC	Helicobacter pylori Korean Type Culture Collection	Korea
IMI(CABI)	CABI Genetic Resource Collection	U.K.
ITDI	Industrial Technology Development Institute Microbial Culture Collection	Philippines
ITM	Belgian Coordinated Collections of Microorganisms Mycobacterial Culture Collection	Belgium
JCM	Japan Collection of Microorganisms	Japan
KCTC	KCTC Korean Collection for Type Cultures	Korea
KEMB	Korea national Environmental Microorganisms Bank	Korea
KMMCC	Korea Marine Microalgae Culture Center	Korea
LEF	Korea Lichen & Allied Bioresource Center	Korea
LIPIMC	Lembaga Ilmu Pengetahuan Indonesia , Indonesian Institute for Sciences	Indonesia
MCC-MNH	Microbial Culture Collection - Museum of Natural History, Museum of Natural History (MNH)	Philippines
NBRC	NITE Biological Resource Center	Japan
PNCM	Philippine National Collection of Microorganisms	Philippines
PVGB	Plant Virus GenBank	Korea
TISTR	TISTR Culture Collection, Bangkok MIRCEN	Thailand
UCCAA	Ukrainian Collection of Cholera Aetiological Agents O1 and non O1 serogroups	Ukraine

Table 1 Participant list of GCM collections (Continued)

UCDFST	Phaff Yeast Culture Collection	USA
UL	The UNILAB Clinical Culture Collection, United Laboratories	Philippines
UMinho-MUM	Micoteca da Universidade do Minho	Portugal
UOA/HCPF	UOA/HCPF University of Athens/Hellenic Collection of Pathogenic Fungi	Greece
UPCC	Natural Sciences Research Institute Culture Collection	Philippines
UPMC	MICROBIAL CULTURE COLLECTION UNIT	MALAYSIA
VKM	All-Russian Collection of Microorganisms	Russia
VTCC	Vietnam Type Culture Collection	Vietnam

isolation, isolation sources, geographic origin, status, optimal temperature for growth, minimum temperature for growth, maximum temperature for growth, medium, application, and published citations to the use of the strain. In addition to these WFCC MDS entries, the GCM contains extensive citation, patent, and gene or genome information related to each strain. All of this information is available from the strain information page for each strain. A schema of the data flow of GCM is shown in Figure 1.

Strains belonging to the same species as well as sub-species are automatically associated to form a species page (Figure 2). A taxonomic tree of species 2000 [8] is generated to serve as a reference for taxonomic identification. Type strains, indicated by their collections are listed on species page. Data on individual strains are organized by culture collections location, type of strain, isolation sources, and genus and species as well. As a

result, all data can be retrieved through the browse option provided in the web server according to these properties.

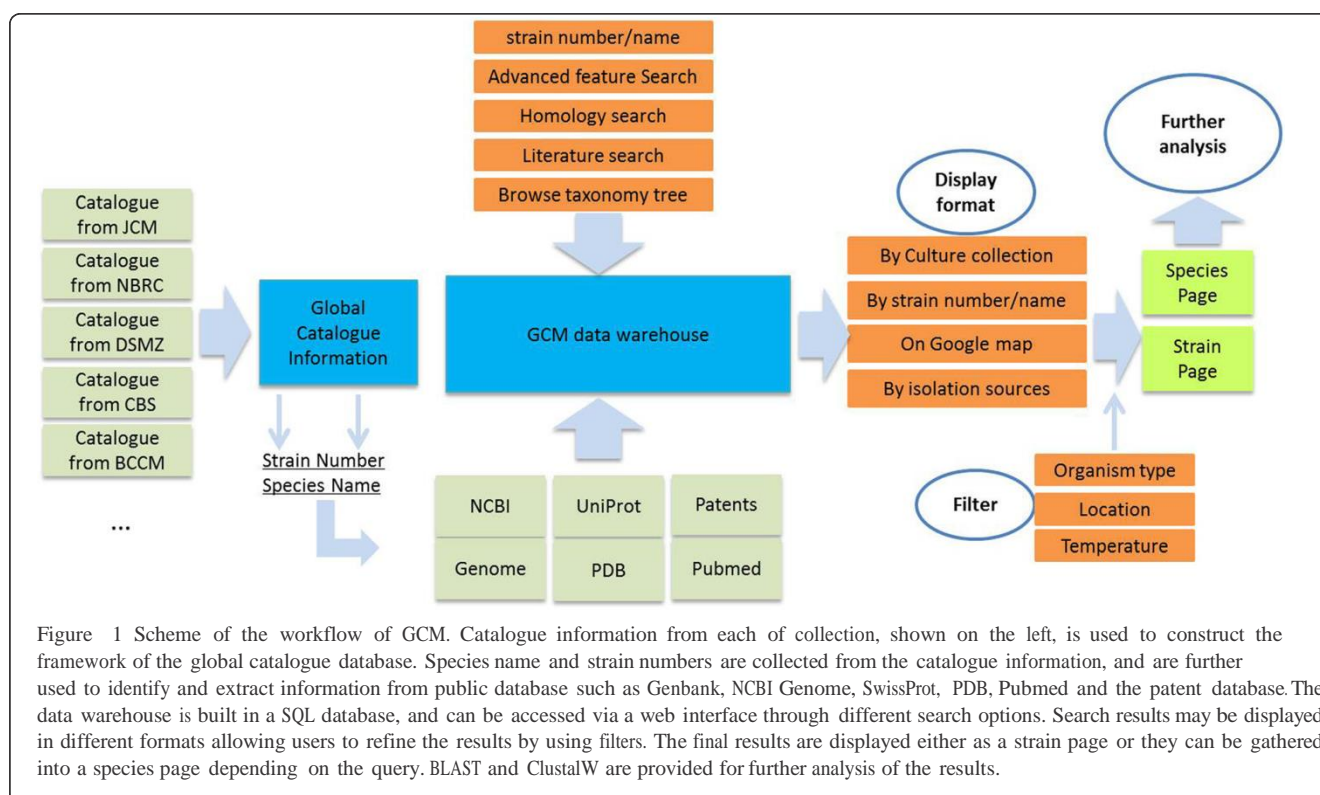
Metagenome and Microbes Environmental Ontology (Hiroshi Mori [9]) which is an ontology about microbial environment was used for text mining of values of isolation sources. The text contained in this data item was automatically compared with the terms of MEO and then sorted into 13 different categories such as soil, microbial-mat/Biofilm, or host-associated, among others (Table 3). For the values that could not be automatically assigned to a specific category, manual curation is required. Data concerning environmental habitats of the isolates can provide important information about the diversity of organism types that are related with certain isolation source types.

About 48% of the strains have geographic information and these strains are from 164 different countries or

Table 2 Summary of GCM strain data

Organism type	Species number	Strain number	Type strain	Sequences	Publications	Patents
Antibody	7	33	0	0	0	0
Phage	181	239	0	0	1	0
Virus	33	296	0	0	0	0
Cyanobacteria	134	287	0	178	0	0
Protozoa	236	754	0	0	0	0
Actinomycetes	842	1490	0	271	192	9
Archaea	1410	3273	1165	2176	1573	48
Microalgae	1820	5495	4	2	1	1
Plasmid	2030	2030	0	0	5	9
Yeast	3668	34907	4796	54773	2089	98
Bacteria	13714	101395	14233	29304	10975	268
Fungi	18537	121548	29916	94348	1960	65
Diatom	19	242	0	0	0	0
Mycobacteria	50	214	0	0	0	0
Other/Rotifera	755	1730	0	0	0	0
Total	43436	273933	50114	181052	16796	498

Information was as submitted by individual collections. Sequence, publication, and patent data were extracted from Genbank, Pubmed, Patent database using the strain numbers.



regions. Data on the geographic origin of isolates (Table 4) is complementary to the habitat and can provide useful information on relative biodiversity and sampling efforts for different countries and regions. These data will ultimately be integrated into the Global Biodiversity Information Facility (GBIF) database through planned activities of GCM (Éamonn [10]).

Data quality control

Because original catalogue data are sometimes non-validated, quality control measures are necessary before data can be published in GCM online. The most frequent quality problem is the misspelling of species name or non-standard naming of species. For example, “Absidiapsychrophilia” was wrongly spelled as “Absidiapsychrophila” in certain collections. In such cases, GCM uses standard microbial nomenclature databases to perform a quality check of its taxonomic data. Databases include the List of Prokaryotic names with Standing in Nomenclature (LPSN) [11], “Species 2000”, and NCBI taxonomy [12] for bacteria and archaea, MycoBank [13] for fungi and yeast.

A programming script was written (in the Java™ language) to automatically compare species names between the GCM catalogue and the nomenclature databases cited above. The comparison showed that from the 36,340 different archaea, bacteria, fungi and microalgae contained in GCM, 2188 could not be found in any of the nomenclature databases above. The average

mismatching is 6% (Table 5). When conflicts are identified, GCM sends the results of these comparisons to curators at the relevant collections to allow them to edit their catalogue information online. When mismatches occur, the system provides the probably correct species name based on character string similarity. Following such comparison, the majority of spelling mistake is corrected.

The second type of problems with the quality of information is related to data content. For example, some “Escherichia coli” strains were wrongly assigned as “Fungi” in the host collection databases. The GCM system collects and compares the lists of differences in the description of cultures in one collection with cultures of the same strains in other collections.

History information was used to do the quality check for species name as well. Totally 12147 strains contain detailed history information in GCM. The system listed all of species name and compared with their history species name in other collections. The result indicated that among 12147 strains, 1746 strains had different species name with their history strains. Further analysis on the result showed that, among the mismatch, 267 belonged to misspelling problems such as “Candida viswannathii” was wrongly spelled to “Candida viswanathii”. However, the left were mistakes or name changes occur during the strain transfer between collections.

Divergent results are forwarded to the curators of the respective collections for corrections. Performing such

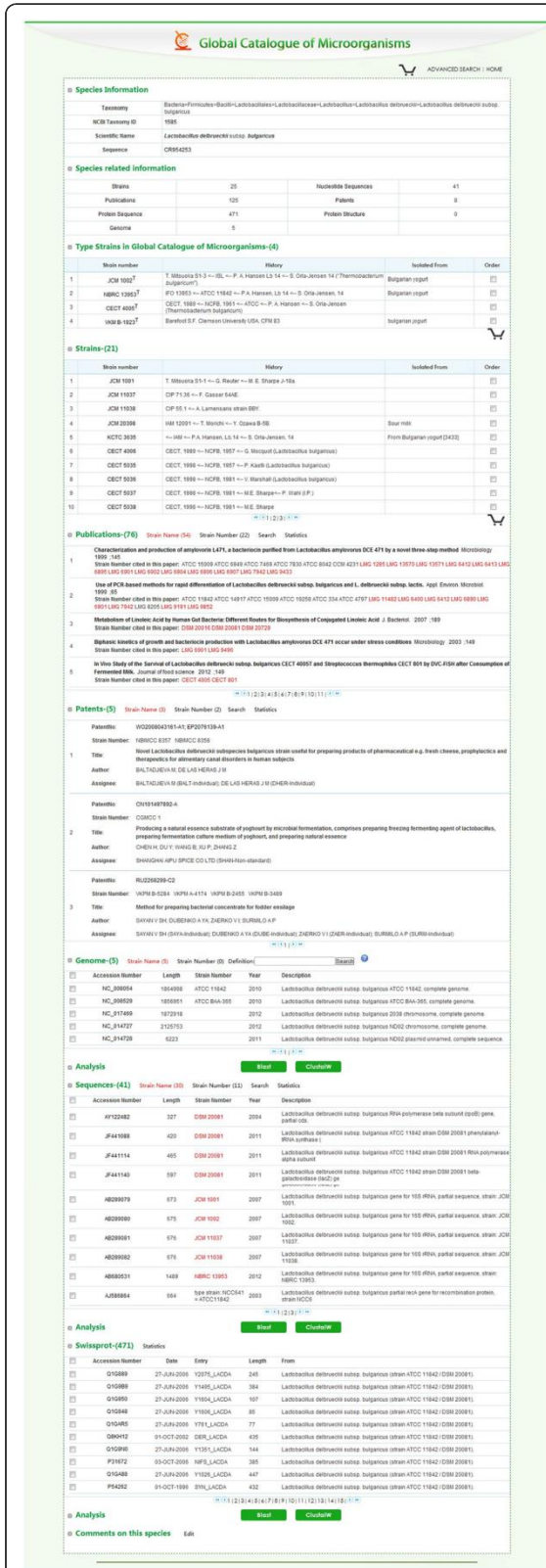


Figure 2 Example of Species page of *Lactobacillus delbrueckii* subsp. *bulgaricus*.

controls for all fields of the database greatly assist collections in correcting existing mistakes.

Utility

Interface and web tools

The database homepage contains a world map which indicates the countries and regions that have already joined the GCM project. Statistics and graphics indicate the continuing acquisition of data into the GCM. A simplified search interface allows the querying of the database by using the strain number and species name. In addition, a variety of tools have been implemented to enhance its use. The main web tools that were integrated into the GCM are the following:

Advanced search

Three query options are available in the advanced search section. Users may search strains within a range of values for one or several properties, including cultivation temperature, substrate, or application, before retrieving the retrieve corresponding results.

Since GCM maintains nucleotide sequences data associated with individual strains, a sequence alignment tool based on the Basic Local Alignment Search Tool (BLASTN) [14] is included. Results are ordered by similarity.

Bibliographic and patents queries are also possible and allow users to search by keywords in titles, abstracts of articles or patents. Search results are listed as strain numbers, strain names, publication abstracts and titles and can be exported in text file format.

With the advanced search tools, the system can perform the following searches

- Searching for type strains for some taxa in certain culture collections
- Searching for strains with specific characteristics in the list of Culture Collection (CC) or Biological Resource Center (BRC), such as range of growth temperature, transfer history, collected location and others
- Searching for strains with specific properties
- Searching strains isolated from various substrates, including sludge or wastewater, soils, sediment, fermentation products. Results are listed in table format, with the type of organism type used as column name;
- Searching strains with particular protein coding genes

Results are listed by strain number, species name, culture collections, and isolation sources. A few filter windows are provided in the result page to allow users to refine the results by collections, growth temperature, isolation sources or organism type.

Table 3 Isolation sources of Strains sorted by type of organism

Isolation source type	Fungi	Bacteria	Yeasts	Actinomycetes	Archaea	Phage	Microalgae	Total
Sludge/Wastewater	1	1091	6	-	9	-	2	1109
Soil	1708	3468	484	264	95	1	1	6021
Sediment	4	46	17	-	14	-	-	81
Fermentation products	123	358	327	-	1	-	-	809
Plant-associated	405	314	644	-	1	-	2	1366
Host-associated	139	480	180	-	3	-	-	802
Human-associated	18	11167	55	-	15	-	2	11257
Water	4	398	50	-	48	-	-	500
Microbial-mat/Biofilm	-	-	-	-	1	-	-	1
Air	6	20	29	-	1	-	-	56
Genetic engineering strain	22698	-	-	-	-	-	-	-
Food	193	83	69	-	2	-	-	347
Others	135	728	76	-	25	-	1	965
Total	2736	18153	1937	264	215	1	8	23314

Species tree viewer

A species 2000 taxonomy tree is used for the organization of strain information. Species names are used to map between GCM data and species2000 name (<http://www.sp2000.org/>), and then a taxonomic tree containing the number of strains for each genus is constructed. User can then browse the taxonomy tree itself, or search a species name within it.

Map viewer

While geographic origins of strains are usually provided as rural location, national park or cities, GCM can automatically translate such locations into more precise

information of longitude and latitude. Strains are then displayed on a map using the Google maps API. In some cases, the location information is a more specific place such as a university or an institute, which could not be translated directly into longitude and latitude values. In such cases, manual annotation by the administrator of GCM will then use the value of the located city as an approximation. An example strain information page is displayed in Figure 3.

Data analysis

A variety analysis tools are also employed on both the strain information and species page. The BLAST program (Altschul SF [13]) was used for sequence homology searches within the database. For sequences related to the same strain or species, the ClustalW [15] program is provided to perform multiple sequence alignment analysis.

Data update and management

To provide the greatest benefit to partner collections, a database management function was provided to GCM participating collections (Figure 4). After registration with the GCM project and filling out a metadata form, a user account will be given to the collection. Curators can then either export catalogue information in batch or add strain information individually. The system automatically records every operation, including updates, additions or deletions and after approval by the administrators in charge, the updated records are published online.

Discussion and conclusion

A large amount of microbial resources are preserved as living strains in collections, however, information describing these strains is often unavailable. Each culture

Table 4 Top 20 countries from which strains were collected

Order	Country	Counts	Order	Country	Counts
1	Japan	8248	11	China	3429
2	France	8070	12	India	2907
3	United States	7701	13	Russian Federation	2872
4	Netherlands	6709	14	South Africa	2419
5	Korea	6270	15	Italy	2009
6	Germany	6051	16	Canada	1848
7	Thailand	5894	17	VietNam	1818
8	United Kingdom	5717	18	Sweden	1786
9	Belgium	5177	19	Australia	1695
10	Spain	3869	20	Switzerland	1466
Total				85955	

114,578 of 273,933 strains contain information regarding their geographic origins. The strains were collected from 164 countries and regions, of which, 85,955 strains were collected from only 20 countries. This takes up approximately 74% of total strains, which indicates a relatively high sampling effort in these countries.

Table 5 Result summary of species name check

Organism type	Species names	Un-matched species name	Percentage of un-match
Archaea	1399	32	2.30%
Microalgae	1457	360	24.70%
Fungi	20719	698	3.40%
Bacteria	12855	1098	8.50%
Total	36430	2188	6.00%

This table provides comparative results of the species names within GCM with public microbial nomenclature database. Species2000, NCBI taxonomy, LPSN and Mycobank were used as reference databases. The average percentage of unmatched names is 6%, while the archaea and fungi showed lower than average percentage of unmatched names. The percentage of unmatched names is relatively high for microalgae, possibly due to the irregular naming for microalgae.

Strain Information

Strain Number:	JCM 1002 ^T (Original site)
Name:	<i>Lactobacillus delbrueckii</i> subsp. <i>bulgaricus</i> (Orla-Jensen 1919) Weiss <i>et al.</i> 1984
Other Collection Numbers:	ATCC 11842; BCRC 10696; CCM 7190; CCUG 21450; CCUG 41390; CECT 4005; CIP 101027; DSM 20081; IAM 12472; IFO 13953; KCTC 3635; LMG 6901; NBRC 13953; NCFB 1489; NCIMB 11778; NRIC 1688; VKM B-1923; VTT E-96662
Organism Type:	Bacteria
History Of Deposit:	T. Mitsuoka S1-3 <-- ISL <-- P. A. Hansen Lb 14 <-- S. Orla-Jensen 14 (" <i>Thermobacterium bulgaricum</i> ").
Isolated From:	Bulgarian yogurt
Type Strain:	Type strain
Optimum Temperature For Growth:	37°C
Literature:	131883
Author:	(Orla-Jensen 1919) Weiss <i>et al.</i> 1984
Add to shopping cart:	

Publications-(15)

1	<i>Lactobacillus equicursoris</i> sp. nov., isolated from the faeces of a thoroughbred racehorse. <i>Int J Syst Evol Microbiol</i> 2010 ,Volume1
2	<i>Lactobacillus capillatus</i> sp. nov., a motile bacterium isolated from stinky tofu brine. <i>Int J Syst Evol Microbiol</i> 2008 ,Volume11
3	Weissellicin 110, a Newly Discovered Bacteriocin from <i>Weissella cibaria</i> 110, Isolated from Pla-a-Som, a Fermented Fish Product from Thailand <i>Appl. Environ. Microbiol.</i> 2007 ,Volume7
4	Structural and Functional Differences in Two Cyclic Bacteriocins with the Same Sequences Produced by <i>Lactobacilli</i> <i>Appl. Environ. Microbiol.</i> 2004 ,Volume5
5	<i>Lactobacillus thermotolerans</i> sp. nov., a novel thermotolerant species isolated from chicken faeces. <i>Int J Syst Evol Microbiol</i> 2003 ,Volume1
6	An In Vitro Study of the Probiotic Potential of a Bile-Salt-Hydrolyzing <i>Lactobacillus fermentum</i> Strain, and Determination of Its Cholesterol-Lowering Properties <i>Appl. Environ. Microbiol.</i> 2003 ,Volume8
7	Cholesterol Assimilation by Lactic Acid Bacteria and Bifidobacteria Isolated from the Human Gut <i>Appl. Environ. Microbiol.</i> 2002 ,Volume9
8	Identification of the bacterial microflora in dairy products by temporal temperature gradient gel electrophoresis. <i>Appl. Environ. Microbiol.</i> 2002 ,Volume8
9	Cholic Acid Is Accumulated Spontaneously, Driven by Membrane pH, in Many <i>Lactobacilli</i> <i>Journal of bacteriology</i> 2000 ,Volume8
10	Phylogenetic analysis of the genus <i>Thermoactinomyces</i> based on 16S rDNA sequences <i>Int J Syst Evol Microbiol</i> 2000 ,Volume3
11	Rapid identification of 11 human intestinal <i>Lactobacillus</i> species by multiplex PCR assays using group- and species-specific primers derived from the 16S-23S rRNA intergenic spacer region and its flanking 23S rRNA. <i>FEMS Microbiol Lett</i> 2000 ,Volume2
12	<i>Lactobacillus paralimentarius</i> sp. nov., isolated from sourdough. <i>IJSEM</i> 1999 ,Volume4
13	Identification of and Hydrogen Peroxide Production by Fecal and Vaginal <i>Lactobacilli</i> Isolated from Japanese Women and Newborn Infants <i>J. Clin. Microbiol.</i> 1999 ,Volume9

Figure 3 Example of strain information of *Lactobacillus delbrueckii* subsp. *bulgaricus*.

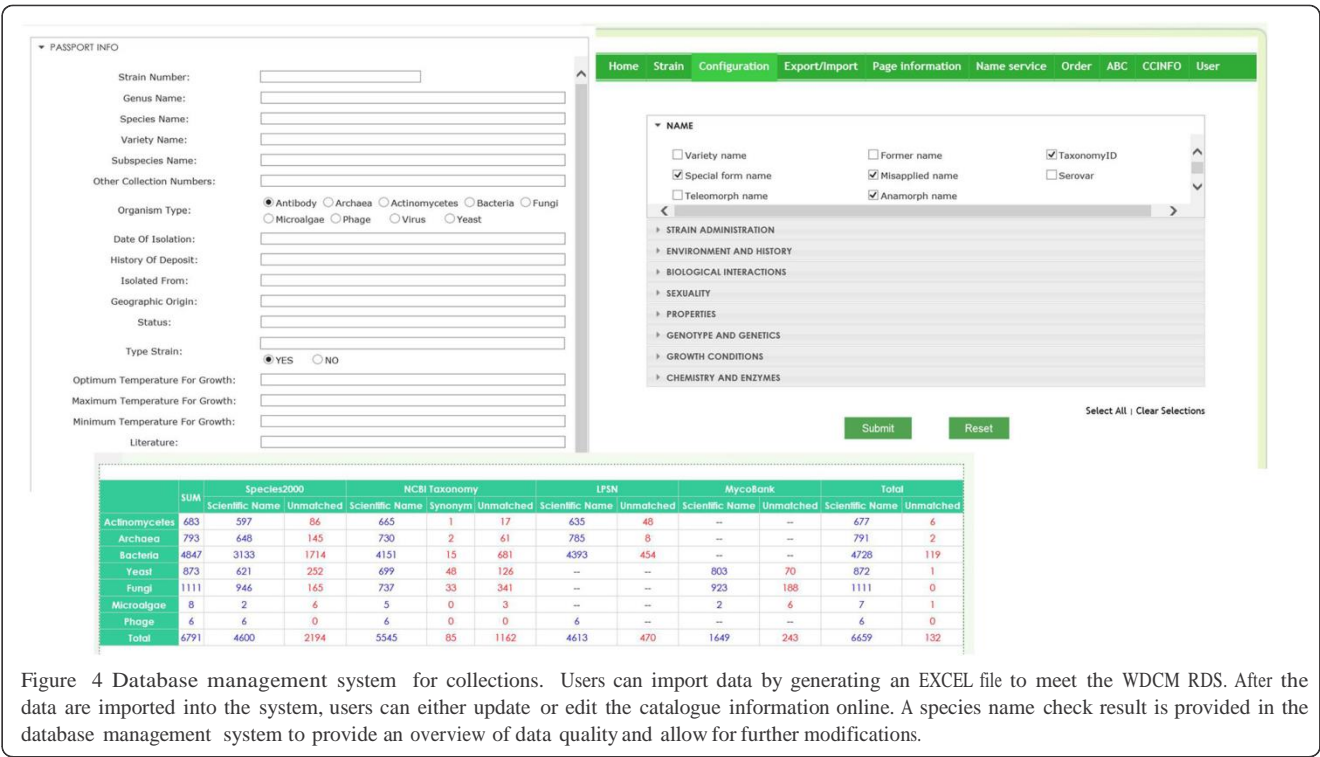


Figure 4 Database management system for collections. Users can import data by generating an EXCEL file to meet the WDCM RDS. After the data are imported into the system, users can either update or edit the catalogue information online. A species name check result is provided in the database management system to provide an overview of data quality and allow for further modifications.

collection is independently responsible for the maintenance of data associated with their microbes, there is presently no enforced data harmonization and information sharing mechanism is available. Such situation hinders both the efficient management of collections and the ability to explore statistics about world microbial resources. Therefore, there is great demand for developing a mechanism for digital, online resource sharing, which provides a fundamental tool for best practices in information management.

The major target group for such system are culture collections staff, as well as academic and industrial microbiologists. We believe that GCM will assist collections, which lack the required human resources and information technology, to publish their stock information in an efficient and standardized way that is most useful for scientific and industrial communities. Database queries via a user-friendly and web-based interface should greatly promote the sharing and use of microbial resources.

While this project is still in its early stage, we are confident that it will continue to grow with the further addition of data, analytical tools and other functionalities. In the future, additional database management tools will be provided to allow more culture collections to share their data via GCM. These tools will lead to the increased availability of accessible data pertaining to microbial strains held in public collections and their utilization for bioindustry, medicine, and research. As it

grows, GCM will incorporate information related to enzymatic and metabolic pathways using developing genomics and bioinformatics tools. Ultimately, GCM is a comprehensive data platform on microbial resources that is available to the public.

Availability and requirements

The GCM database runs on a platform with both Java and MySQL server. Catalogue information gathered from associated collections is centralized within WDCM servers, which is hosted at the Institute of Microbiology, of the Chinese Academy of Sciences.

The Blast program is used for the sequence homology search in the database (BLASTN 2.2.25). Multiple sequence alignments are performed using the ClustalW program (version2.1). GCM is available at <http://gcm.wfcc.info>.

Competing interest

The authors declare that they have no competing interest.

Authors' contributions

LW designed database and web services. QS integrated data resources from Pubmed and Genbank. SY developed the database, made the webpages and wrote the database queries. HS designed and supervised construction of the database. YZ, KM, AV, SKI, MO, YL, VR, SI and FG provided the catalogue information and participated in the design of the database. LW and KM wrote the manuscript. DP and JM designed the database and supervised construction of the project. All authors read and approved the final manuscript.

Acknowledgement

GCM project was initiated by WDCM and approved by the WFCB board. WDCM acknowledged the contributions of all participating collections to the

GCM project. At the time of writing this article 52 collections from 25 countries have already joined the effort.

Author details

¹Institute of Microbiology, Chinese Academy of Sciences, Beijing, China. ²World Data Centre for Microorganisms(WDCM), Beijing, China. ³Belgian Coordinated Collections of Micro-organisms Programme, Belgian Science Policy Office, avenue Louise, 231 1050 Brussels, Belgium. ⁴National Institute of Genetics, Yata, Mishima 411-8540 Japan. ⁵Fungal Genetics Stock Center, University of Missouri, Kansas City, Missouri, USA. ⁶G.K. Skryabin Institute of Biochemistry and Physiology of Microorganisms RAS, Pushchino, Moscow region, Russia. ⁷Culture Collection Division Biological Resource Center (NBRC), National Institute of Technology and Evaluation (NITE), 2-5-8 Kazusakamatari, Kisarazu-shi, Chiba 292-0818 Japan. ⁸Japan Collection of Microorganisms/Microbe Divion, RIKEN BioResource Center, Koyadai 3-1-1, Tsukuba, Ibaraki 305-0074 Japan. ⁹Seoul Women's University/Korea National Research Resource Center, Seoul, Korea. ¹⁰CBS-KNAW, Fungal Biodiversity Centre, Uppsalalaan 8, Utrecht, The Netherlands. ¹¹Bioresources Technology Unit, National Center for Genetic Engineering and Biotechnology, 113 Thailand Science Park, KlongLuang, Prathumthani 12120 Thailand.

Received: 4 September 2013 Accepted: 27 December 2013
Published: 30 December 2013

References

1. Satoru Miyazaki HS: Networking of biological resource centers: WDCM Experiences. *Data Science Journal* 2002, 1(2):102–107.
2. OECD Best Practice Guidelines for Biological Resource Centres; 2007. <http://www.oecd.org/health/biotech/oecdbestpracticeguidelinesforbiologicalresourcecentres.htm>.
3. Gams W, Hennebert GL, Stalpers JA, Janssens D, Schipper MA, Smith J, Yarrow D, Hawksworth DL: Structuring strain data for storage and retrieval of information on fungi and yeasts in MINE, the microbial information network Europe. *J Gen Microbiol* 1998, 134:1667–1689.
4. CABRI: Common Access to Biological Resources and Information (CABRI), GUIDELINES FOR CATALOGUE PRODUCTION; 1998. <http://www.cabri.org/guidelines/catalogue/CPcover.html>.
5. Benson DA CM, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW: GenBank. *Nucleic Acids Res* 2013, 41(D1):D36–D42.
6. Consortium TU: Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic Acids Res* 2012(40):71–75.
7. Rose PW, Beran B, Bi C, Bluhm WF, Dimitropoulos D, Goodsell DS, Prlic A, Quesada M, Quinn GB, Westbrook JD, Young J, Yukich B, Zardecki C, Berman HM, Bourne PE: The RCSB protein data bank: redesigned web site and web services. *Nucleic Acids Res* 2011, 39:D392–401.
8. Cachuela-Palacio M: Towards an index of all known species: the Catalogue of Life, its rationale, design and use. *Integrative Zoology* 2006, 1(1):418–421.
9. Hiroshi M: Metagenome and Microbes Environmental Ontology; 2013. <http://bioportal.bioontology.org/ontologies/ME>.
10. Eamonn OT: Meeting report: hackathon-workshop on Darwin Core and MiX standards alignment. *Stand Genomic Sci* 2012, 7(1):166–170.
11. Euzéby JP: List of bacterial names with standing in nomenclature: aFolder available on the internet. *Int J Syst Evol Microbiol* 1997, 47(2):590–592.
12. Federhen S: The NCBI Taxonomy database. *Nucleic Acids Res* 2012, 40(1):D136–D143.
13. Crous PW, Walter G, Stalpers JA, Vincent R, Gerrit S: MycoBank: an online initiative to launch mycology into the 21st century. *Stud Mycol* 2004, 50:19–22.
14. Altschul SF GW, Miller W, Myers EW, Lipman DJ: Basic local alignment search tool. *Journal of molecular biology* 1990, 215:403–410.
15. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG: ClustalW and ClustalX version 2. *Bioinformatics* 2007, 23(21):2947–2948.

doi:10.1186/1471-2164-14-933

Cite this article as: Wu et al.: Global catalogue of microorganisms (gcm): a comprehensive database and information retrieval, analysis, and visualization system for microbial resources. *BMC Genomics* 2013 14:933.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit



Annex 2

Linking the Access and Benefit Sharing Clearing House and user sector databases: a test case to maximise legal certainty and information provision using the Global Catalogue of Microorganisms

Introduction

With the coming into force of the Nagoya Protocol, and the development of Access and Benefit Sharing legislation by Parties, there are two overarching needs: providers need to know what happens to genetic resources that are accessed, and users need to have legal certainty when those resources are utilized. These needs need to be met transparently and in a manner providing confidence to both providers and users. They also need to be met simply and cost-effectively, and in a manner minimising additional administrative burdens for all stakeholders.

An obvious means of addressing the need is to seek to link data systems in which providing countries store the relevant information with those in which user communities store theirs, so that relevant information can be made available in a timely manner to all stakeholders.

Under Article 14 of the Nagoya Protocol the Access and Benefit-Sharing Clearing House (ABSCH) has been established as a part of the Convention on Biological Diversity's Clearing House Mechanism (CHM). Its purpose is to provide information to Parties and others as set out in Articles 14 and 17. Two elements of the ABSCH that are of particular relevance here are the *Internationally-Recognised Certificate of Compliance* (IRCC), developed as required under Article 17 paragraph 2, and the *Checkpoint Communiqué* (CPC), developed to facilitate Checkpoints making reports on utilization under Article 17 paragraph 1. Following advice by an Expert Group (UNEP/CBD/ICNP/1/2), both of these documents have been built with the optional facility to embed links to external references, in particular taxonomic name(s) (e.g. as served by the *Catalogue of Life*) of the species covered by an IRCC, and to specimens when held elsewhere (e.g. as available through the *Global Biodiversity Information Facility*). The Expert Group cautioned that there were possible risks of the external information changing as part of usual scientific practice, and suggested the potential of such links be investigated. To date such an investigation has not been undertaken.

The development of the WFCC Global Catalogue of Microorganisms (GCM) provides an opportunity to test the utility of the linkages available from the IRCC and CPC to an external data system. A major criterion of the link between data fields in this and the ABSCH proposed below is that it avoids possible loss of information following scientific activity (e.g. change in taxonomic name or identification, but instead allows a persistent association with information about genetic resources once accessed and entered into the Catalogue.

The GCM is a robust, reliable and user-friendly system being developed to help Culture Collections manage, disseminate and share information related to their holdings. It also provides a uniform interface for the scientific and industrial communities to access comprehensive microbial resource information. It holds records for microbiological strains held by Culture Collections, assigning each a Strain Number. Use of this strain number with the associated Analyzer of Bio-resource Citations (ABC) allows users to check citations of the strain among patents, scientific papers, sequences and microbial resources.

It should be noted that the ABSCH, while functional, is still under development, and the mechanisms of the IRCC and Checkpoint Communiqué are still to be tested in practice; feedback from Parties and others on their utility and advice for further development can be expected. Similarly, the GCM and the ABC are in a relatively early stage in development. Thus all data systems being considered are in a state where workflows can be developed and improvements sought.

The proposal below is designed to investigate the practicality and benefits of linking the IRCC (and CPC) to external records covering Genetic Resources. It is written specifically in the context of the GCM and its

anticipated operation, but it may serve as a model for application to other systems. The proposal is not intended to be prescriptive, or dictate how a Party might choose to use a system. It is intended to provide an option to support Parties and users to their mutual benefit.

Requirements of a reciprocal links between the ABSCH and the GCM

Desirable overarching properties of the system are:

- Associated records must be persistent and not subject to change;
- low cost;
- rapid;
- requires minimal human intervention and activity;
- reduces risk of error through transcription;
- provides high legal surety for users of a strain;
- provides comprehensive information for the CNA on what has happened to material that has left the country, based on simple queries;
- provides comprehensive information for a Checkpoint in a user country, based on simple queries;
- builds on existing properties of the ABSCH and GCM and minimises novel developments of properties and changes in workflow;
- provides a model for other systems.

More specifically, the emplacement of reciprocal links between the two systems should enable:

- Competent National Authorities to be able to search the GCM using one or more references (links) on the IRCC that will facilitate their (i) understanding of what GR have been accessed under a given permit or its equivalent, in a way which accommodates any subsequent changes in scientific name or other identification assigned to those GRs; and (ii) tracking the place where accessed GRs are held *ex situ*, including any third party transfer.
- Users of the GCM (or even databases linked to the GCM) to discover rapidly whether a given GR was accessed legally (i.e. is covered by an IRCC) and who to contact in the Providing Country in the case of proposed change of use or other issue requiring understanding of the original PIC and MAT.

A desirable property of a system would be for it to provide to the CNA information about any utilization of the GR.

Constraints

There are two workflows to consider that have an impact on the proposal.

1. The generation of the Internationally Recognised certificate of Compliance.

The collector of the GR seeking access is granted Prior Informed Consent by Competent National Authority and Mutually Agreed Terms are agreed with the provider according to the legal framework of the country of origin. These two elements comprise the permit or its equivalent. Following this and access (collecting) the collector may leave the providing country with the GR, and the CNA may publish the permit or its equivalent on the ABSCH, whereupon it constitutes the Internationally Recognised Certificate of Compliance. The ABSCH sends an e-mail to the User and the CNA confirming the IRCC has been issued; this serves to inform the user of the unique identifier (GUID) of the IRCC.

2. The generation of a record of a strain in the CCM.

The strain is the most appropriate level to track in the GCM; a species may comprise multiple strains with different genetic constitution and different properties, while a specimen has little meaning in a microbiological collection, where a sample may comprise many thousands of specimens. Strains may of course be separated into different samples, and these samples transferred to different collections. Consequently the GCM initially creates a record when a strain is accessed into a Culture Collection; this can be linked to further records of the same strain as it is shared with another Culture Collection. The process then is for a Culture Collection to notify the

GCM when a strain is accessed, and provide with that information additional data regarding that strain. The GCM then creates a record and issues a GUID which can be used to refer to that strain in further activities such as citing in publications or patents.

There are several constraints arising from these workflows that need to be taken into account.

The IRCC is likely to be generated before the genetic resources (microbiological strains) have (i) been isolated from the samples collected *in situ*, (ii) been identified and accessed into a culture collection and (iii) a record of that strain created in the GCM. As a result of this, the record of the strain in the GCM cannot be provided to the CNA of the provider country for inclusion in the IRCC on first publication of the IRCC.

The IRCC cannot be changed once published on the ABSCH. However, an amended version can be created by the CNA, linked to the first version.

Multiple Strains may be isolated from a single sample or set of samples covered by a single IRCC, and these may be identified and placed in culture collections separately, leading to many different GUIDs being issued by the GCM over time. Currently the IRCC and CPC have only a single field for external links. Developing additional fields in the IRCC would presumably not be a technical challenge. However, if the model required the CNA to repeatedly update the IRCC on receipt over time of many different strain records (as is very likely) this could pose a problem.

On a technical level, while records in the ABSCH and in the GCM are assigned Globally-Unique Identifiers (GUIDS) there is not as yet a reliable system of resolving these GUIDS operating independently of either system. Consequently a means of linking records by use of APIs or persistent URLs needs to be developed. This is a technical issue not addressed in this paper (although it will be a part of any test project), and not expected to be unduly problematic.

Any system will require transfer of information between the actors, be they individual, institutional or data management systems. Explicit above are the IRCC number and the Strain reference: if the Culture Collection does not have the IRCC number, for example, it will be unable to inform the GCM that a given strain was collected under that IRCC. However, these two data items are not the only ones that will be required. In order that any system functions successfully, the subject biological resources (genetic resources) must be accompanied by a record containing a 'minimum data set'. Without prejudice to confidentiality (which may have to be managed) this minimum data set should be sufficient to: (i) enable links between the IRCC and the GR covered by that IRCC to be maintained and developed as appropriate; (ii) provide the information required when a report is made on utilization which would cause the generation of a Checkpoint Communiqué; (ii) provide additional information that national legislation might require when a report is made on utilization (note that under the EU Regulation¹, for example, considerably more detail may be required than is mandatory on the IRCC); (iii) provide sufficient information for the recipient to effectively manage the legal context of the GR (again, the requirements of the EU Regulation might be considered here). As well as these elements the minimum data set will include information about the origin and biology of the GR that are of importance scientifically. This minimum data set may be embodied in the requirements of a standard Material Transfer Agreements, as developed under TRUST².

¹ Regulation (EU) No 511/2014 of the European Parliament and of the Council of 16 April 2014 on compliance measures for users from the Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization in the Union

² TRUST - Transparent Users-Friendly System Of Transfer, Based on MOSAICC - Micro-Organisms Sustainable use and Access regulation International Code of Conduct

Models for linking the IRCC and the GCM

Five models to establish a link were considered and assessed against the requirements above:

1. Creation of a record in the GCM for the IRCC number and the pre-issue of this record, prior to population, to the ABSCH for addition to the IRCC as appropriate;
2. CNA notified of Strain GUIDs once issued;
3. Pre-issue of GCM strain GUIDs to the user for provision to the CNA when MAT is sought;
4. Using the GCM as the *de facto* search system instead of the ABSCH;
5. Pre-issue of an “Access Record” by the GCM.

These are not all reproduced here, but are available on request. The solution which delivers the greatest number of benefits and most closely meets the requirements is the first of these, and has been selected for further development. It delivers the simplest workflow of the five models considered and involves the least amount of human intervention (and consequently the least amount of time by the CNA, Checkpoint and users).

Proposed Solution – Creation of ‘IRCC record’ in the GCM and pre-issue to ABSCH

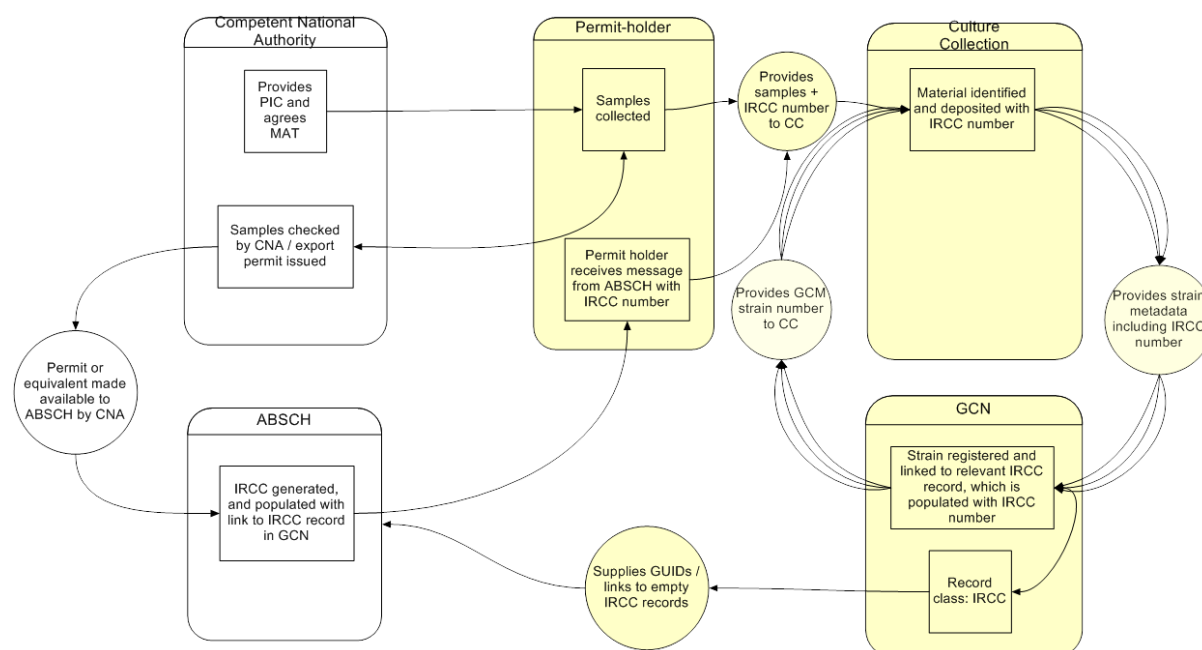


Fig.1. Information flow for proposed solution

Proposal

1. The GCM creates a set of unpopulated records to hold the IRCC numbers / links once these are created in the ABSCH, each accessible externally, either by a GUID or another means (‘GCM IRCC records’).
2. A set of GCM IRCC record identifiers is provided to the Secretariat for use in the ABSCH.
3. Each time an IRCC is issued for a microbiological GR the CNA can select a GCM IRCC record for inclusion on the IRCC. The provision mechanism needs to be developed; ideally it would be an autopopulation of a link box if the subject of the IRCC is selected as ‘Microorganism’, but it could be selected from a list elsewhere in the ABSCH, provided that once selected it is no longer available for use on another IRCC.
 - a. In this model it is important that the ultimate GCM record of any strain isolated from samples taken includes the IRCC number and can correctly associate this with the relevant GCM IRCC record that

has been identified in the IRCC. There appear to be two means of achieving this in an automated manner:

- i. The e-mail sent by the ABSCH to the user when the IRCC is published also includes the identifier of the GCM IRCC record; both the IRCC number and this identifier become part of the minimum data set supplied with the material and any strains isolated from it.
 - ii. The GCM contains a means, when the IRCC number is provided, of automatically looking up the relevant field in that IRCC for the GCM IRCC record and populating that with the IRCC number. Note: If the link to the GCM IRCC record is entered on the IRCC in the free-text field describing the GR this may cause additional work in creating the link with the GCM.
4. Each GCM IRCC record is empty until populated by a submission from a Culture Collection. When a strain record is supplied by a CC with the IRCC number this is attached to the appropriate GCM IRCC record, populating it with the IRCC number and a child record of the strain GUID and minimum data set.
 5. A subsequent submission of a strain with the same IRCC number (or GCM IRCC record identifier) will attach the new record to the existing GCM IRCC record. This will be the case for every subsequent submission.
 6. A search of the CGM from the ABSCH using the link to the GCM IRCC record would retrieve information about all of the strains so far identified and recorded in the GCM that were originally collected under the same IRCC

There could be development of tailored reports that would automatically be generated by a search of the GCM IRCC records. Much of the content of the GCM is unnecessarily complex for the needs of CNAs or Checkpoints. Consequently, a means of generating simplified reports containing only the information they might require should be developed. Such reports might include:

- Any patents issued
- Any publications citing the strain(s)
- Where the strains are housed.

There could also be the possibility of such reports being generated automatically for provider countries or user countries for those records which related to them, either whenever a new report was received or at predetermined intervals. This might constitute an 'automated Checkpoint'.

This model has the potential to be applied to other sectors, not just microbiology.

Acronyms and Glossary

ABC	<i>Analyzer of Bio-resource Citations</i> - A platform which supports the researchers on checking the citations among papers, sequences and microbial resources, in order to evaluate the contribution of each cultural collection. According to the statistics acquisition by WFCC-MIRCEN World Data Center for Microorganisms, less than one hundred culture collections have been covered in ABC platform. http://abc.wfcc.info/
ABSCH	<i>Access and Benefit-Sharing Clearing House</i>
CNA	<i>Competent National Authority</i> – The body or individual in a country authorised to sign ABS agreements.
GCM	<i>Global Catalogue of Microorganisms</i>
GUID	<i>Globally Unique Identifier</i>
IRCC	<i>Internationally-Recognised Certificate of Compliance</i> – A record generated when the Competent National Authority of a Providing Country publishes a permit or equivalent (e.g. PIC and MAT) on the ABS Clearing House. This is given a unique identifier by the Clearing House (IRCC Number) and provides legal surety of the genetic resources covered. It may also be used to simplify reporting.

ANNEX 3

TRUST TRANSPARENT USER-FRIENDLY SYSTEM OF TRANSFER

Introduction

TRUST stands for TRansparent User-friendly System of Transfer, for Science & Technology. It aims at managing the incidence of the CBD and Nagoya Protocol on the scientific, technical and administrative activities of culture collections and, more generally, incorporating the Nagoya Protocol into the daily life of microbiologists.

TRUST proposes a cost-efficient, simple, fast and multiple users - multiple purposes global system. TRUST is based on outcomes and experience built up during the [MOSAICC1](#) and [MOSAICS2](#) projects.

The administrative flow of TRUST combined with the data processing of the Global Catalogue for Microorganisms (**annex 1**) form the system implementing the Nagoya Protocol in the culture collections community. Linkage between the Global Catalogue of Microorganisms and the ABS Clearing House Mechanism is done via an automatic machine driven process described in **annex 2**.

The present annex is the executive summary of the TRUST guidelines. The full version of the TRUST guidelines will be available at <http://bccm.belspo.be/projects>.

1 MOSAICC - Micro-Organisms Sustainable use and Access regulation International Code of Conduct.

- [MOSAICC](#) is a voluntary Code of Conduct, a tool to support the implementation of the Convention on Biological Diversity at the microbial level, in accordance with other rules of international and national laws.
- Access to microbial genetic resources (MGRs) is a prerequisite for the advancement of microbiology. Monitoring the transfer of MGRs is necessary to identify the individuals or groups that are entitled to be scientifically or financially rewarded for their contribution to the conservation and sustainable use of the MGRs. MOSAICC combines the need for easy transfer of MGRs and the need to monitor the transfer of MGRs.
- MOSAICC recommendations facilitate access to MGRs and help partners to make appropriate agreements when transferring MGRs, in the framework of the CBD and other international and national laws. It is a major pioneering initiative setting up guidelines for sound management of MGRs. Although predating the negotiated Bonn Guidelines MOSAICC contains all its elements. MOSAICC is now updated via the TRUST initiative as described in this document.

2 MOSAICS - Micro-Organisms Sustainable use and Access management Integrated Conveyance System

- [MOSAICS](#), combined with MOSAICC, is the forerunner of TRUST. It was funded by Directorate General Research of the European Commission under the Sixth Framework Program.
- Already in 1999, the MOSAICC project had identified three necessary features for a system for management of access to and transfer of microbiological resources, to implement coherently the CBD provisions on ABS.
- MOSAICS central objective is the development of such an integrated conveyance system that:
 - has reliable tools to evaluate the economic value of microbiological resources (EVA)
 - disposes of validated model documents with standard provisions to enable tracking via simple procedure, widely applied by microbiologists (ADAM)
 - combines valuation and tracking in one system for trading of microbiological resources, with balanced benefit sharing for those that are entitled to be rewarded for the services and products they provide to society (ICS)

Executive summary

The best way to achieve ABS with effective socio-economic benefits is to build on existing procedures, to make the appropriate linkages between the various actors, and provide for the necessary incentives to the users so that ABS is effectively more beneficial to all and does not require coercive measures or penalties.

TRUST is looking for adjusting existing tools to each other and for providing a cost-efficient, simple, fast and multiple users - multiple purposes global system.

TRUST is a modular system having as backbone the Global Catalogue of Microorganisms and making use of the expertise gained by MOSAICC, MOSAICS, and other initiatives. The TRUST system comprises four elements:

- i. Updated MOSAICC features with administrative **workflows** adapted to the structure of the Nagoya Protocol and improved in light of past experience. **The flow of information will be connected to the ABS Clearing House Mechanism³** via a simple machine-processed link from a defined IRCC⁴-CHM field to a defined IRCC-GCM field.
- ii. **Refined Material Accession Agreement (MAA) and Material Transfer Agreement (MTA) models** with standardized definitions detailed in a two-tier glossary system⁵ to optimize sectorial implementation of national laws implementing the Nagoya Protocol.
When considering the concept of "registered collections" as defined by the EU regulation implementing the Nagoya Protocol, culture collections community has responded by developing an efficient information and material handling strategy to deal with administrative handling of transfers of material, especially the compulsory Prior Informed Consent. Thus the concept of registered collections as future trusted sources for ABS-compliant genetic resources is imbedded into the culture collections community at the outset. Consequently it is of primary importance that:
 - a. Every microbial genetic resource "entering" a collection is covered by a PIC obtained at the time of its isolation from *in situ* conditions or after corrective administrative action.
 - b. Every microbial genetic resource having entered a collection with the appropriate initial PIC may be distributed, accompanied by the original PIC, without any additional PIC procedure set by the country of origin or the country of use.
- iii. **An automated powerful integrated data management and processing system** able to provide for any information related to microbial material: **the groundbreaking Global Catalogue of Microorganisms (GCM)**. The CGM provides users with the ability to trace the possession, location, transmission and use of uniquely-identified microbial strains, including country of origin, existence of PIC and MAT, the creation of derived patents and all associated scientific publications. This system already includes more than 60 collections from more than 30 countries and information on nearly 290.000 strains from 41.000 species.

³ <http://absch.cbd.int>

⁴ Internationally Recognised Certificate of Compliance

⁵ The terms used in the Nagoya Protocol and the CBD are conceptual definitions. These definitions, such as "derivative", are by essence umbrella terms covering all sectors of application. It is illusive to expect a satisfactory operative text with such broad ranging definitions because the conceptual definitions used in the text of the NP will be subject to different interpretation, depending on the sector of application.

Therefore these conceptual definitions must be declined into operational definitions. In other words, the general definitions used in the NP and eventually translated into national laws should be connected to subsets of operational definitions specific for each sector. TRUST suggests that stakeholders define subsets of operative terms apposite for each conceptual definition. National laws using the terms defined in the NP will make reference to these specific glossaries designed sector by sector. This will produce a two tier glossary system for disambiguation. It will minimize the degree of uncertainty and leave almost no margin for diverging interpretations.

- iv. **Cooperative structure** wherein culture collections
 - a. make use of the latest ICT technology to develop the necessary identification and tracking system, primarily for scientific purposes but also for any other bona fide ends.
 - b. conduct and facilitate research in genomics and functional genomics, thus develop capacities of storage and processing of genomic, transcriptomic and metabolomic information. These compiled data improve definite characterization of microbial resources.
 - c. conduct their efforts in networks, in conformity with NP provisions on Technology Transfer, collaboration and cooperation.

LOGICAL FLOW CHART - Operating the Nagoya Protocol in microbiology

The Logical Flow Chart of TRUST is divided in 4 parts, based on the analysis of the Nagoya Protocol provisions in light of usual practices in microbiology.

I. FRAMEWORK - Read the script

First, one must define what should we strive for and how. Translate the general objective of the protocol into operational objectives for the sector; transcribe the terms of the CBD and NP in a relevant glossary for microbiology; identify the most profitable benefits for microbiology.

II. WORK - Act in consequence

Then, one must define what it takes to succeed. The parts to develop and assemble where grouped in two major blocks:

- a) Technical measures. Technical tools developed initially for scientific purposes can be exploited to manage access and use of microbial material;
- b) Legal measures. Administrative and juridical procedures set to manage contractual arrangements rooted in private laws also organise the rights and duties of stakeholders in ABS matters.

III. NETWORK - Join forces and coordinate

Furthermore, life sciences technology and bioinformatics increase at a speed exceeding the pace of legislation design and update. Cooperation in both technical and legal fields, cooperation between lawyers and life scientists is crucial because complementary. NP article 22, paragraphs 4 and 5, points the importance of **capacity building** in both technical field and legal field. Any kind of legal or scientific collaboration imbeds some continuous training, awareness-raising and technology transfer, in other words capacity building *sensu largo*. The objective is to foster the capacity building process.

IV. WORKOUTS - Imagine solutions and improvements

Fourth, the system should include a cyclic improvement system, assessing the system efficiency, especially the cost/benefit ratio, and encourage innovation and creativity.

Figure 1. Logical flow: Reading the Nagoya

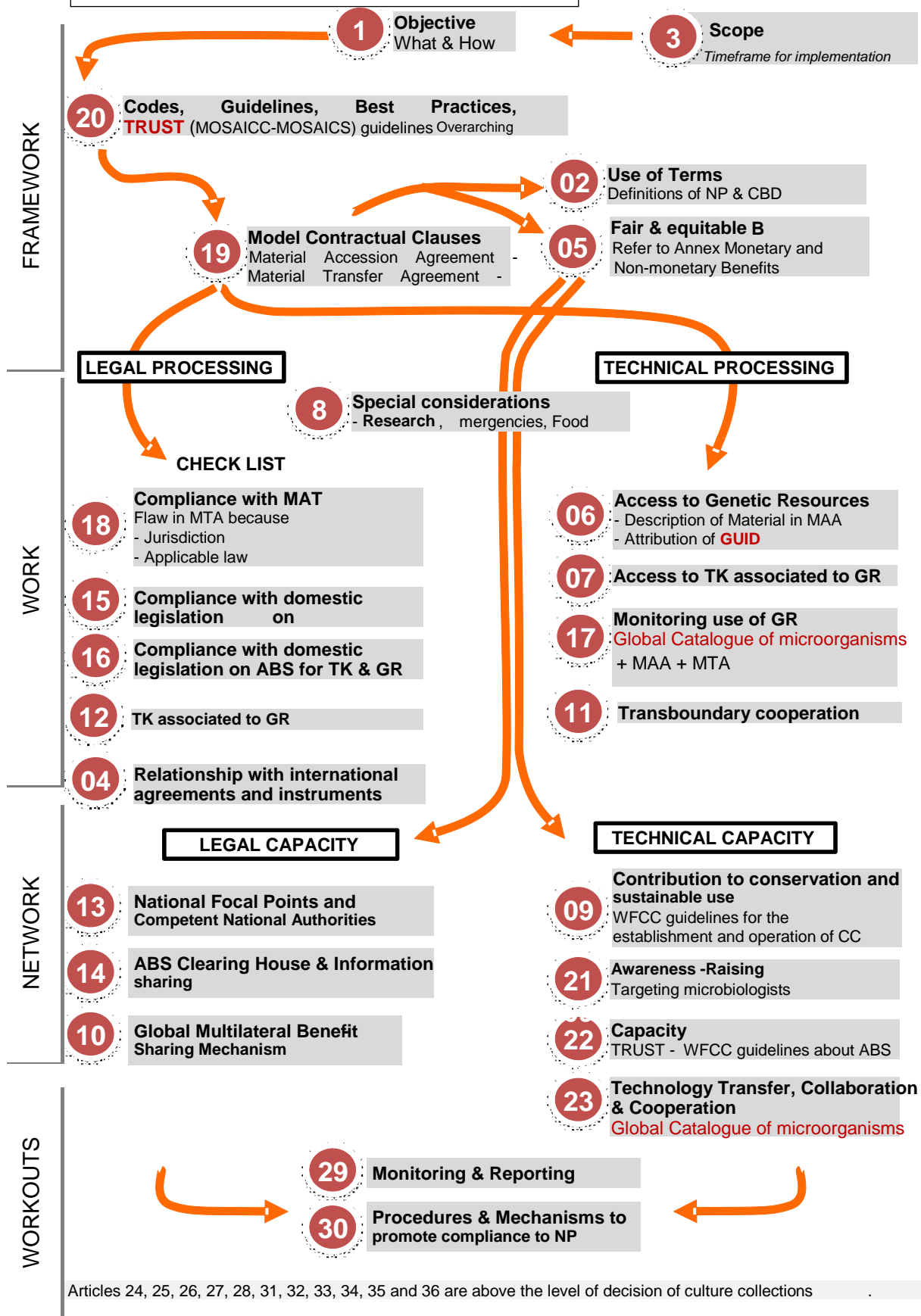
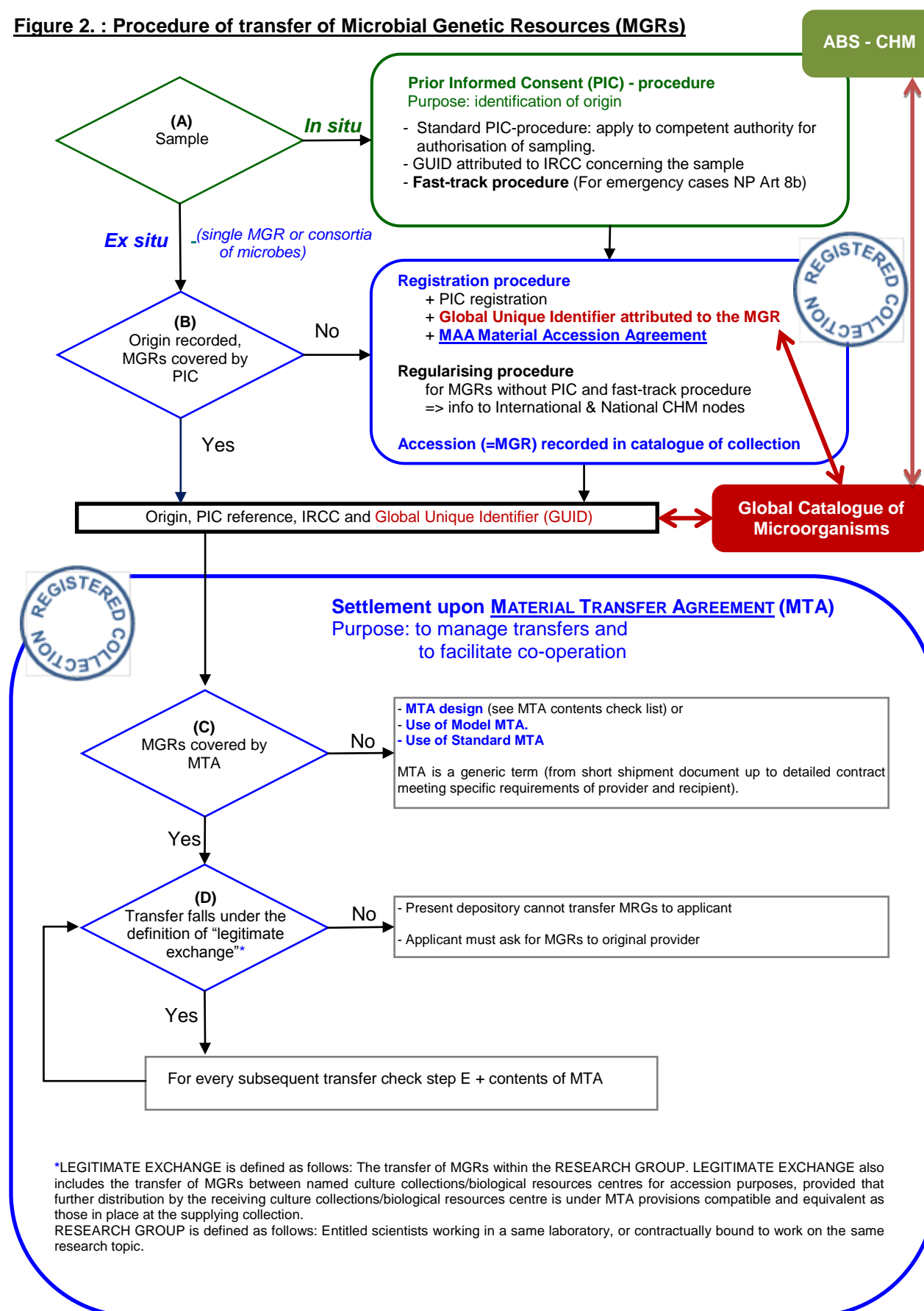


Figure 2. : Procedure of transfer of Microbial Genetic Resources (MGRs)



Annex 4

NIEMA Network of International Exchange of Microbes under ACM

ACM

The Asian Consortium for the Conservation and Sustainable Use of Microbial Resources (ACM) was established in 2004. Currently 23 organizations in 13 countries (Cambodia, China, India, Indonesia, Japan, Korea, Lao, Malaysia, Mongolia, Myanmar, Philippines, Thailand and Vietnam) are the members of the ACM. The objective of the ACM is to promote collaboration among governments or public organizations in Asian countries for the purpose of enhancing conservation and sustainable use of microbial resources in Asia. An ACM task force is developing a new scheme, which is "[Network of International Exchange of Microbes in Asia: NIEMA](#)", proposing a legitimate and streamlined way of transferring and utilizing microbial resources in line with the Nagoya Protocol.

NIEMA System

Having recognized the importance of Microbial Resource Centres (MRCs) in the development of microbiology, a voluntary scheme named "Network of International Exchange of Microbes under ACM (NIEMA)" in which microbes are transferred transnationally and utilized in a streamlined way in line with the Convention on Biological Diversity (CBD) and the Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from Their Utilization to the Convention on Biological Diversity (NP) has been developed by a Task Force of Management of Material Transfer (MMT-TF) in Asian Consortium for the Conservation and sustainable Use of Microbial Resources (ACM).

The basis of the NIEMA is the registration of MRCs that declare their adoption of a common policy (NIEMA Code of Conduct) which is in compliance with the provisions of the CBD and the NP. The registration is limited to MRCs only. The NIEMA covers;

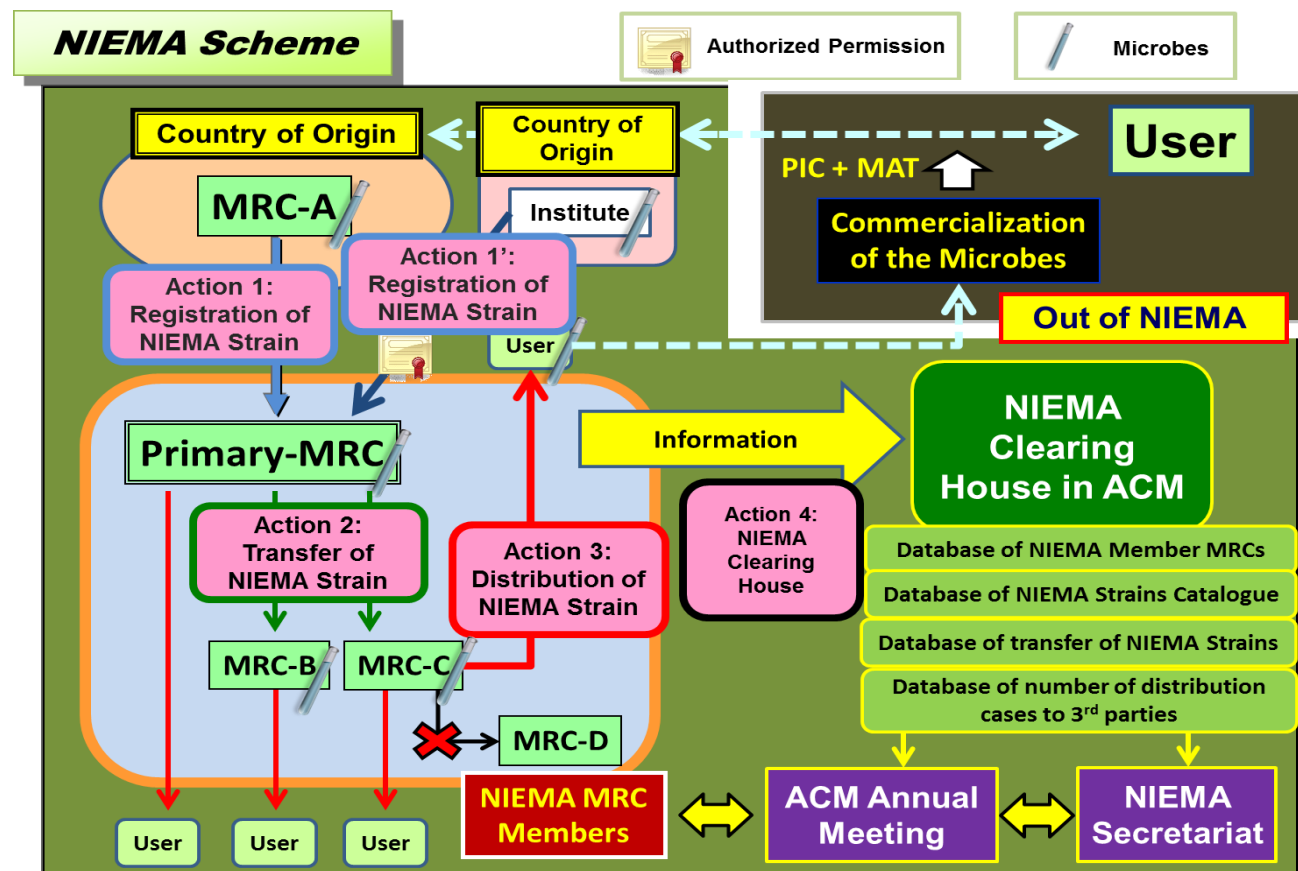
1. Registration of microbial strains into the NIEMA system
2. Exchange of the microbial strains between NIEMA MRCs
3. Distribution of the NIEMA strains from NIEMA MRCs to third parties

NIEMA strains which are exchanged among NIEMA MRCs and distributed to third parties are utilized only for non-commercial research purposes. A NIEMA Clearing-House stores and makes available any data of transfer information of NIEMA microbial strains between NIEMA MRCs.

NIEMA Scheme

NIEMA covers:

1. Registration of microbial strains into the NIEMA system
2. Exchange of the microbial strains between NIEMA MRCs
3. Distribution of the NIEMA strains from NIEMA MRCs to third parties



NIEMA strains which are exchanged among NIEMA MRCs and distributed to third parties are utilized only for non-commercial research purposes. A NIEMA Clearing-House stores and makes available any data of transfer information of NIEMA microbial strains between NIEMA MRCs.

/...