

# Microbiological Common Language (MCL): a standard for electronic information exchange in the Microbial Commons

Bert Verslyppe<sup>a,b</sup>, Renzo Kottmann<sup>c</sup>, Wim De Smet<sup>a</sup>, Bernard De Baets<sup>d</sup>,  
Paul De Vos<sup>a,e</sup>, Peter Dawyndt<sup>b,\*</sup>

<sup>a</sup> Laboratory of Microbiology, Department of Biochemistry and Microbiology, Ghent University, K.L. Ledeganckstraat 35, 9000 Ghent, Belgium

<sup>b</sup> Department of Applied Mathematics and Computer Science, Ghent University, Krijgslaan 281, 9000 Ghent, Belgium

<sup>c</sup> Microbial Genomics Group, MPI for Marine Microbiology, Celsiusstr. 1, D-28359 Bremen, Germany

<sup>d</sup> Department of Applied Mathematics, Biometrics and Process Control, Ghent University, Coupure Links 653, 9000 Ghent, Belgium

<sup>e</sup> BCCM<sup>TM</sup>/LMG Bacteria Collection, Ghent University, K.L. Ledeganckstraat 35, 9000 Ghent, Belgium

Received 14 October 2009; accepted 12 February 2010

Available online 6 March 2010

## Abstract

Although Biological Resource Centers (BRCs) traditionally have open catalogs of their holdings, it is quite cumbersome to access meta-information about microorganisms electronically due to the variety of access methods used by those catalogs. Therefore, we propose Microbiological Common Language (MCL), aimed at standardizing the electronic exchange of meta-information about microorganisms. Its application ranges from representing the online catalog of a single collection to accessing the results of StrainInfo integration and ad hoc use in other contexts. The abstract model of the standard precisely defines the elements of the standard, which enables implementation using a variety of representation technologies. Currently, XML and RDF/XML implementations are readily available. MCL is an open standard, and therefore greatly encourages input from the microbiological community.

© 2010 Elsevier Masson SAS. All rights reserved.

**Keywords:** Microbiological Common Language (MCL); XML (Extensible Markup Language) format; Microbial meta-information; Electronic information exchange; Strain integration; StrainInfo

## 1. Introduction

During the last two decades, the advent of modern information technologies has changed the way information is generated, distributed and shared. The low marginal cost of publishing new data and metadata has caused an exponential growth of publicly available information upon which new scientific niches using this data have originated. Scientists working in these niches often perform experiments completely *in silico*, investigating relations between different parameters or datasets. The success of such

experiments greatly depends on the availability and interoperability of the data. In this context, availability covers the intrinsic openness (licence) and access possibilities (e.g. online vs. offline, web services vs. web pages), but also quality (a minimal quality is necessary in order to have useful data).

It is common practice to capture scientific results and insights in literature, and this is a central aspect of scientific culture. In microbiology, the scientific legacy also includes the actual biological material used for research, and therefore it is equally important to store the material itself for later reuse. This task is carried out by Biological Resource Centers (BRCs). Using techniques such as lyophilization and cryopreservation, BRCs store and preserve authentic biological material in the long term. BRCs then distribute this material to interested researchers, allowing validation of scientific work and possibly forming a basis for further developments and

\* Corresponding author.

E-mail addresses: bert.verslyppe@ugent.be (B. Verslyppe), rkottman@mpi-bremen.de (R. Kottmann), wim.desmet@ugent.be (W. De Smet), bernard.debaets@ugent.be (B. De Baets), paul.devos@ugent.be (P. De Vos), peter.dawyndt@ugent.be (P. Dawyndt).

applications. Upon deposit in a BRC, the material is assigned a so-called strain number which is used to identify it internally, as well as in other resources (such as publications and databases). Next to the material itself, BRCs store additional meta-level information necessary for archiving, cataloging and documenting the collection. This information is required when retrieving specific isolates from the BRC (e.g. isolates belonging to a certain species).

BRCs usually have open catalogs, as these are the billboards of their holdings. Although traditionally distributed in print, most catalogs are nowadays available online through polished web applications with advanced search functionality. The actual level of detail contained within catalogs depends on the governance of the individual BRCs. Some catalogs only state minimal information such as the strain number itself, a species name and basic information about the depositor, whereas other BRCs offer detailed information on the origin of the culture, its history, its isolation, references to scientific literature and sequence information. A rich catalog is an incentive for researchers to retrieve material from that BRC. Even though most BRCs have put some effort into making their catalogs available online, there is no uniform access to this meta-information. It is virtually impossible to perform electronic processing, as the information does not have a uniform structure and is not made available in structured formats.

The Microbial Commons designates various initiatives that attempt to move towards a globally distributed research infrastructure, based on facilitated access to biological materials, related data and information for basic research, education and commercial use purposes (Dawyndt et al., 2006; Dedeurwaerdere et al., 2007). Therefore, part of a true Microbial Commons is seeking a way of electronically exchanging information about microorganisms, leveraging a lingua franca for disseminating microbial information and meta-information in online environments. There have been several standardization (Gams et al., 1988; Stalpers et al., 1990) and integration (CABRI: Commons access to biological resources, <http://www.cabri.org/>) attempts in the past, but these have failed to set up a standardized, open infrastructure that allows electronic processing. Although relatively new, StrainInfo (Van Brabant et al., 2006) has quickly become the de facto reference for information and meta-information on microorganisms. StrainInfo functions as a world-wide, virtual catalog integrating the information from BRC catalogs with related information. Its novel approach is its true integration by which equivalent strain numbers that represent the same strain are merged into single records (Dawyndt et al., 2005). Each record corresponds to a single strain and all information known about that strain is integrated and summarized on a corresponding passport page. However, this information still needs to be made available in a common electronic format. Therefore, next to the legal and governance aspects of a Microbial Commons, a common language for the exchange of microbial information is essential to lift microbial research to the next level of comparative in silico analysis.

## 2. Microbiological Common Language (MCL)

The MCL has been designed to facilitate the electronic exchange of information on microorganisms. It allows capturing information available in BRC catalogs in an electronic form. The standard is viewed as a framework for the rich description of microbial material ranging from information on the sampling and isolation process and availability in BRCs to the biochemical properties of the microbial material being described. However, there is no clear boundary on what is reasonably seen as being “the fields required to fully describe microbial material”. Therefore, the standard consists of an abstract model which establishes a generic framework, together with a basic set of terms that facilitate capturing everyday BRC catalog contents. This allows translating BRC catalogs into this new format using transparent mapping from existing fields. Moreover, the framework also allows the use of specialized terms which come, as a consequence of their specificity, from other standards.

The standard has been heavily influenced by the Microbial Information Network in Europe (MINE) vocabulary (Gams et al., 1988; Stalpers et al., 1990), and implements the recommendations put forward by the OECD (OECD, 2007). The MINE standards were early attempts to standardize BRC databases and rigorously define data fields and their application range. However, as MINE also defines the underlying database structure, this imposes a heavy burden on implementing the standard within a BRC. Therefore, MCL is conceived as a communication standard rather than an internal database scheme. Moreover, the abstract model of the standard is loosely coupled with actual implementations. This means that the standard meticulously defines the name, definition and structure of its elements, but not how to syntactically format the data. The advantage of this approach is that the standard does not depend on a particular representation technology and that it can be easily ported to new data transfer systems. For practical usage, however, it is necessary to agree upon a particular implementation, and therefore concrete usage is discussed in Section 4.

Along with the loose coupling of abstract model and implementation, MCL is also novel in its approach to structuring microbial information. MCL uses the natural workflow from sampling and isolation to a description of the original isolate and subsequent deposits in BRCs. Traditional standards do not model these steps and use a flat list of properties and their values instead. The model is visualized in Fig. 1 and contains seven cornerstone “entities” that represent classes of objects that can be described using the standard. By convention, entity names start with capital letters. The logical microbiological workflow starts with a physical Sample taken from the environment, followed by an Isolation process which results in a Culture. Cultures can then be subsequently Deposited in BRCs to gain world-wide distribution. Moreover, cultures have associated Publications and BRCs often list recommended growth Mediums. These processes and results are independently described by the standard.

For each entity, several “terms” are defined that can be used to describe its properties. Entities correspond to “things”, whereas terms correspond to “properties” of those things. By

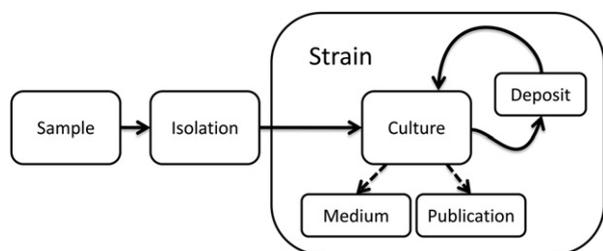


Fig. 1. The abstract model of MCL follows the logical flow from sampling to subsequent deposits in BRCs.

convention, the names of terms start with a lowercase letter. A few examples of the MCL terms can be found in Table 1. Although most terms have a self-explanatory meaning, all terms are precisely defined by the online language reference (<http://www.straininfo.net/projects/mcl/reference/>). This document contains the exact definition for all terms and is the authoritative description of term usage.

Although the concepts Culture and Strain intuitively may appear similar, they are on different levels in Fig. 1. The distinction has been defined by Staley and Krieg (1984) and is adopted by StrainInfo (Van Brabant et al., 2006). Staley and Krieg (1984) see a bacterial culture as a population of bacterial cells of a strain, instantiated at a given place at a given time, e.g. in a test tube, on an agar plate or in a cryopreserved or lyophilized state intended for long preservation. In practice, Cultures are associated with strain numbers and, after deposit in another BRC, each issue of a new strain number yields a new Culture. A strain is seen as all descendants of a single isolation in pure culture, usually the succession of cultures ultimately derived from a single colony. In consequence, a Strain entity contains all underlying Culture entities and is associated with the integration result of all information known about the strain.

Table 1  
Some examples of terms used in the MCL standard.

Term	Definition [MINE]	Entity
strain Number	A strain number used to identify a culture or strain [STN, ACCN]	Culture, strain
other Strain Number	Equivalent strain number. Not to be used as an identifier for cultures [OCC]	Culture
species Name	Species name of the organism [SP]	Culture, strain
type Strain Of	Name of taxon for which strain is type strain [AT]	Culture, strain
sample Location Country	Country where sample was taken [LOC]	Sample
sample Habitat	Textual description of habitat where sample was taken [ISOFR]	Sample
isolation Date	Date of isolation [ISOL]	Isolation

For each term, a short definition and an indication (in square brackets) of a corresponding MINE field is given. The last column indicates the corresponding entities. For the full and authoritative definition of the terms, we refer to the language reference.

### 3. Relation to and compatibility with other standards

#### 3.1. Relation to existing standards

The MINE project performed pioneering work in structuring information on microorganisms. Originated in the late eighties, the original fungal (Gams et al., 1988) and later bacterial (Stalpers et al., 1990) MINE standards defined a database scheme along with an extensive list of fields conceivable at that time. Many of the MCL terms are based on MINE fields, and if applicable, the related field(s) are indicated in the language reference document. MINE defines a Minimum Data Set (MDS) of fields to be used for the exchange of information between the national nodes of the MINE network and in printed catalogs. As an extension of the initial goals of the MINE framework, CABRI was the first initiative for integrating multiple BRC catalogs into a single virtual catalog. Next to the standardization of information validation and dissemination, and the linking between catalogs, it also aims to guarantee BRC quality by the creation and enforcement of quality management guidelines and standards. The database includes 28 catalogs covering bacteria, archaea, fungi, yeasts, plasmids, phages, DNA probes, animal and human cell lines and plant cells and viruses. The search engine, which incorporates queries on species name and strain numbers, is able to search through all catalogs at once or through individual catalogs. The system includes a uniform front-end for making the catalogs of multiple BRCs accessible as a one-stop shop. The database model of CABRI was heavily influenced by MINE and many of the CABRI terms exactly follow the MINE format and syntax. CABRI also adopted the notion of an MDS, but additionally defines a Recommend Data Set (RDS) and a Full Data Set (FDS).

MINE (and thus also CABRI) fields often contain multiple pieces of information, which are frequently demarcated using a peculiar syntax. In consequence, current BRC catalogs still contain fields that combine several distinct pieces of information (MINE uses the term “subfields”) into a single field. This poses problems for automatic processing, as these fields ideally need to be split into their atomic components. To overcome this problem, terms that are specifically designed to match these legacy fields were added to MCL. These terms are most often marked as deprecated, but were nevertheless added to allow BRCs to export semantically rich files without the initial need for manual data curation. An example of this type of curation is splitting a free text sample description into a separate habitat, sampling country and place field, a task which is difficult to carry out electronically. Moreover, the language uses terms from the PRISM standard (Publishing Requirements for Industry Standard Metadata, <http://www.primstandard.org/>) for the fine-grained description of publications. Dublin Core (Dublin Core Metadata Element set, <http://dublincore.org/documents/dces/>) is used to include a “human understandable” (and thus not normalized) citation (see Guidelines in Dublin Core: <http://dublincore.org/documents/dc-citation-guidelines/>). However, this by no means signifies that normalization is not necessary. Data normalization is one of the goals of MCL. These fields enable

postponing the normalization work to a later stage, which in practice means that StrainInfo will have to develop means to normalize the legacy data and thus improve the recoverability and retrievability of data from BRCs.

### 3.2. Compatibility with genomic standards

Genomic and post-genomic analysis of microbial strains has become more routine and widespread in recent years. The genomic community became aware that a standardized rich set of meta-information accompanying DNA sequence data is of importance to better facilitate comparative genomics studies. The Minimum Information about a Genome Sequence (MIGS) checklist published in 2008 (Field et al., 2008) captures the complete context of a genome sequence and defines the minimum information needed to better facilitate comparative studies. This information includes geographic and spatial details of the original sample, information on cultured organisms and details of the experimental techniques used. MIGS is also implemented in the Genomic Contextual Data Markup Language (GCDML) (Kottmann et al., 2008) using XML.

Especially in the description of cultured organisms, there is quite some overlap between MIGS and MCL. To align and establish a strong bond between both standards, MCL adopts the same modular approach as already used by GCDML (see Fig. 1). In turn, this allows GCDML to incorporate the MCL terms which overlap MIGS. This transparent alignment of MCL and GCDML is technically possible due to the capacities of XML to consistently mix vocabularies. This would not be possible with the formats defined by MINE and CABRI.

## 4. Usage and implementations

Although the previous section describes the overall structure of the proposed standard, we have not yet described its practical application. As the structure of the standard intuitively maps to the XML Document Object Model (DOM), XML has been chosen as the default representation technology. It is equally possible to represent the data using RDF: the RDF/XML representation is highly similar to the plain XML representation. Each distinct use case of the standard has a corresponding “schema”. Schemata enforce strict formatting, term order and also define obligatory fields. Documents can be automatically validated against the schema, which facilitates error detection and reliable electronic processing. Therefore, a distinct schema is developed for each particular use case, in order to meet the demands of the application at hand. A full reference of some implementations, including links to the corresponding XML and RDF/XML schemata, can be found at the MCL project home page (see <http://www.straininfo.net/projects/mcl>).

An important use case is the representation of BRC catalogs. This is implemented by the “catalog” schema, which contains specific guidelines for exporting BRC catalogs. The schema defines the “minimal set” of required (obligatory) terms and introduces a header which contains meta-information on the

export itself. The exports are used by StrainInfo in its new synchronization procedure as shown in Fig. 2. Currently, seven BRCs have already adopted this format and actively synchronize with StrainInfo by regularly uploading an MCL export of their catalog (see Table 2). In this procedure, BRCs push their updates to StrainInfo instead of StrainInfo pulling the catalog data at regular time intervals. This allows timely and accurate updates and enables BRCs to control their presence in the StrainInfo index. At the same time, this also solves the scalability problems of the pull paradigm. As BRCs indexed using the pull paradigm require the development and continuous maintenance of custom parsers required for screen-scraping their catalogs, this eliminates the need for time-consuming manual interventions when performing index updates. Moreover, by possibly offering the files available for download separately, BRCs can allow external resources to electronically access their catalogs independently of StrainInfo.

Another use case illustrated in Fig. 2 is the export of the StrainInfo integration itself. StrainInfo receives information on cultures from multiple resources and performs integration at strain level. This integration process bundles all information known about a given strain, i.e. all cultures and all culture-related information. The integration process consists of linking equivalent cultures, the necessary error detection and correction and the semantic aggregation of information attached to the cultures. Although the resulting information is made available to the end-user through so-called strain passport pages, well-formatted MCL documents are necessary to be able to further process these integration results automatically. The strain export (defined by the “strain” schema) populates a full Strain element which includes the integration result as well as the original data (Cultures). A stripped-down example is shown in Fig. 3. The example displays a strain listing three equivalent strain numbers, species and type strain information, and meta-information on the original biological sample and isolation. The example does not contain the original Culture elements. Note the usage of the **mcl** namespace to indicate that the terms are defined in the MCL context. The full declaration of the namespace is defined in the Strain element attribute. This document is intended to be consumed by downstream applications performing further analysis on the strain.

Furthermore, ad hoc use of the standard is accepted, but terms should be correctly formatted and employed. A typical example of this is in tables, where one or more MCL columns can be mixed with user-defined columns. By using MCL terms instead of free-form names as column headers, the column contents are strictly defined, and by consequence, this enables integration with other resources. For example, particular custom table exports in StrainInfo use certain MCL terms as column names.

## 5. Results and discussion

MCL is intended to be broadly applicable in situations where microbial material is referenced or used and therefore has been designed to be interoperable with existing and future standards. Its use ranges from identifying particular cultures (with a minimal Culture element) to the full description of a strain (i.e.

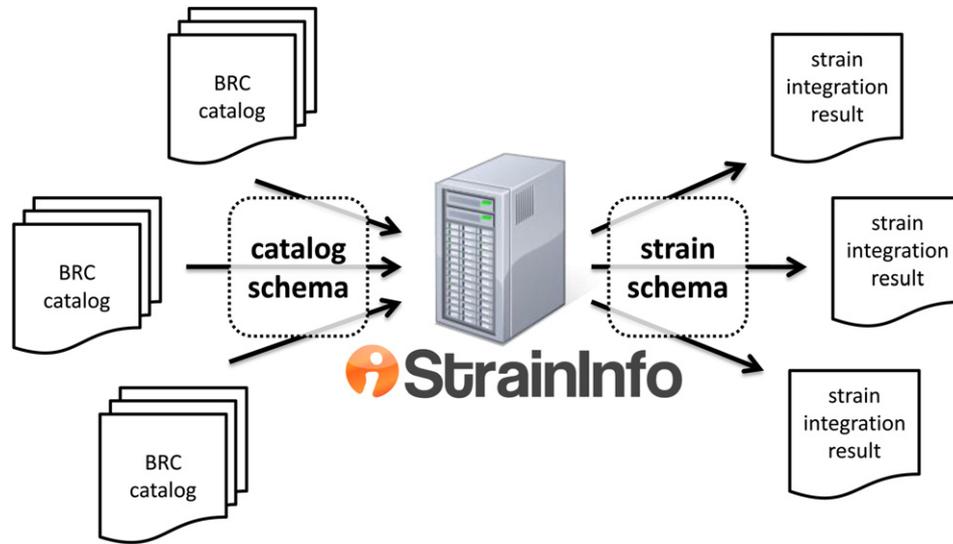


Fig. 2. Two distinct use cases of MCL. At the left-hand side, MCL is used to synchronize BRC catalogs with StrainInfo (using the catalog schema). This information is integrated by StrainInfo, and the integration results are made available in MCL documents formatted according to the strain schema (right-hand side).

the integrated view of StrainInfo). Using the standard uniquely for referencing cultures (i.e. without appending additional meta-information) already allows interesting applications. For example, publishers could use the standard to list all cultures mentioned or used in their publications. This allows the accurate and complete indexing of publications and thus the reliable listing of all publications applicable to a certain strain or species. Analogously, GenBank/EMBL/DDBJ flat files could use MCL terms to link the sequence to the corresponding culture (strain number) from which it was taken. Adding more meta-information to the cultures increases the possibilities of

downstream applications. For example, in order to be able to list all strains isolated from samples taken from a particular geographical region, sample descriptions that contain detailed geospatial meta-information are necessary. Moreover, this information needs to be available in a format suitable for computational processing, as it must be combined with a gazetteer to determine whether it is part of the geographical region. Building semantic search functions subsumes computational access to detailed meta-information along with the availability of supporting ontologies and reasoning technologies.

The standard intentionally contains only a basic set of entities and terms, that more or less correspond to meta-information available in BRC catalogs. It is expected that more specialized vocabularies will make use of or naturally extend MCL to

Table 2  
List of early adopters (alphabetical order). These BRCs synchronize with StrainInfo using the MCL standard.

Acronym	WDCM no.	BRC name	Country
CECT	412	Spanish Type Culture Collection	Spain
CFBP	639	Collection Francaise des Bacteries Phytopathogenes	France
CIP	759	Collection de L'Institut Pasteur	France
CIRM-Levures	788	Centre International de Ressources Microbiennes – Levures	France
CNCTC	130	Czech National Collection of Type Cultures	Czech Republic
IHEM	642	BCCM <sup>TM</sup> /IHEM Biomedical Fungi and Yeasts Collection	Belgium
LMG	296	BCCM <sup>TM</sup> /LMG Bacteria Collection	Belgium
MUCL	308	BCCM <sup>TM</sup> /MUCL (Agro)Industrial Fungi & Yeasts Collection	Belgium
PCC	481	Pasteur Culture Collection of Cyanobacteria	France
VKM	342	All-Russian Collection of Microorganisms	Russia
VTT	139	VTT Culture Collection	Finland

```
<mcl:Strain xmlns:mcl="http://www.straininfo.net/ns/mcl/2.0/">
  <mcl:strainNumber>CCUG 27413</mcl:strainNumber>
  <mcl:strainNumber>CIP 103790</mcl:strainNumber>
  <mcl:strainNumber>DSM 4216</mcl:strainNumber>
  <mcl:speciesName>Bacillus smithii</mcl:speciesName>
  <mcl:typeStrainOf>Bacillus smithii</mcl:typeStrainOf>
  <mcl:Sample>
    <mcl:sampleLocationCountry>USA</mcl:sampleLocationCountry>
    <mcl:sampleHabitat>cheese</mcl:sampleHabitat>
  </mcl:Sample>
  <mcl:Isolation>
    <mcl:isolationDate>1947</mcl:isolationDate>
  </mcl:Isolation>
</mcl:Strain>
```

Fig. 3. Example of usage of XML implementation of the standard. The example contains export of a strain as integrated by StrainInfo. For clarity, detailed culture meta-information has been omitted. Terms are defined in the **mcl** namespace to indicate that the terms are defined in the MCL context. The full definition of the namespace is included as a Strain element attribute.

model more complex microbial information. The standard is envisioned to form a framework for the description of rich, semantic microbial information. Nevertheless, it is foreseen that the set needs to be augmented with new terms if proven necessary. However, this may be problematic, as existing terms cannot change their definition and the structure of the language has to remain unaltered. To allow future extension, the MCL namespace deliberately contains a version number. This version number reflects revisions of the standard and, if new terms are added to new versions, can also be used to distinguish between different implementation levels.

The standard does not solve the issue of globally unique identifiers (GUIDs) (Van Brabant et al., 2008). Strain numbers as traditionally used by microbiologists have a number of problems making them unusable as identifiers in a global environment. Strain numbers are not globally unique due to overlaps (for example equal acronyms). Sometimes strain numbers are recycled (and used in two different contexts) or do not have enough discriminatory power to be distinguished. This is often the case for strain numbers locally assigned by individual researchers, which have, for example, a form such as “A1”, “R-4” or “F 16”. There is no uniform form and often different syntactic variants of the same strain number are in use.

To overcome this problem, StrainInfo has introduced the “cultureId” (Dawyndt et al., 2005). For each occurrence of a strain number in a separate context (i.e. strain), a distinct “cultureId” is allocated. This allows discriminating between the different contexts in which a strain number is used and thus making true unambiguous references. External objects should preferably be linked to “cultureIds” instead of strain numbers to ensure long-term validity of the references. Nevertheless, “cultureIds” are not intended to replace strain numbers, but rather resolve them in a global context. References should be made with “cultureIds”, whereas it is expected that end users will continue to make use of strain numbers.

Obviously, a resolution mechanism is necessary to easily translate between strain numbers and “cultureIds”. A large fraction of BRC strain numbers can be unambiguously resolved to their corresponding “cultureIds”. However, in cases of ambiguous strain numbers (e.g. researcher numbers or recycled strain numbers), automatic resolution needs extra information to distinguish between different contexts. In this case, additional contextual information can be provided (e.g. species names or equivalent strain numbers) to determine the correct strain number instance. Automatic resolving is possible using the Resolver web service from StrainInfo (see <http://www.straininfo.net/docs/Webservices>).

## 6. Future directions

MCL is envisioned as an open, fundamental infrastructure for the microbiological community of the future. Next to the accessibility of biological material, a cornerstone of the Microbial Commons is the free distribution of the corresponding meta-information. StrainInfo used its unique position to create a language to fulfill its own practical needs, and this is an

opportunity to formalize the technological basis of universal electronic information exchange in the Microbial Commons. Only via wide participation by the community will MCL be able to grow beyond its initial practical scope and form the basis of a new era in microbiology where new insights might be gained from the vast amount of scientific knowledge readily available. Therefore, the microbiological community is encouraged to adopt the standard by creating and consuming data in this format and to provide feedback on the language structure, on existing or missing terms or on governance issues. Standards become truly valuable when there is a community consensus on their structure and a vision of universal usage materialized in the practical adoption of correctly formatted documents.

It is expected that an increasing number of BRCs will export their catalog in MCL format and StrainInfo will use the standard for all its web services and exports. The standard will continue to evolve as a consequence of community feedback and the practical experience of an increasing number of MCL documents becoming available. Extensions that include mappings to other data standards and addition of specialized terms (if necessary) will be carried out in modular case studies.

## Acknowledgements

The authors wish to thank Wim Gillis for his contributions to the development of the StrainInfo platform. The development of StrainInfo and the MCL standard are funded by the Belgian Science Policy (contract IUAP VI/06). We also want to express our gratitude to Claudine Vereecke for her valuable input from the BRC perspective and to the MCL early adopters who have made their catalog available in XML and provided feedback on the language structure.

## References

- Dawyndt, P., Dedeurwaerdere, T., Swings, J., 2006. Exploring and exploiting microbiological commons: contributions of bioinformatics and intellectual property rights in sharing biological information. *Int. Soc. Sci. J.* 188, 249–258.
- Dawyndt, P., Vancanneyt, M., De Meyer, H., Swings, J., 2005. Knowledge accumulation and resolution of data inconsistencies during the integration of microbial information sources. *IEEE Trans. Knowl. Data Eng.* 17, 1111–1126.
- Dedeurwaerdere, T., Desmeth, P., Green, P., et al., 2007. Microbiological resources, who owns what? Building the microbial commons. In: Stackebrandt, E. (Ed.), *Connections between Collections*. WFCC and DSMZ, Goslar, pp. 129–130.
- Field, D., Garrity, G., Gray, T., Morrison, N., Selengut, J., Sterk, P., Tatusova, T., Thomson, N., et al., 2008. The Minimum Information about a Genome Sequence (MIGS) specification. *Nat. Biotechnol.* 26, 541–547.
- Gams, W., Hennebert, G.L., Stalpers, J.A., Janssens, D., Schipper, M.A., Smith, J., Yarrow, D., Hawksworth, D.L., 1988. Structuring strain data for storage and retrieval of information on fungi and yeasts in MINE, the Microbial Information Network Europe. *J. Gen. Microbiol.* 134, 1667–1689.
- Kottmann, R., Gray, T., Murphy, S., Kagan, L., Kravitz, S., Lombardot, T., Field, D., Glöckner, F.O., 2008. A standard MIGS/MIMS compliant XML schema: toward the development of the Genomic Contextual Data Markup Language (GCDML). *OMICS* 12, 115–121.
- OECD Best Practice Guidelines for Biological Resource Centres, 2007. OECD.

- Staley, J.T., Krieg, N.R., 1984. Classification of prokaryotic organisms: an overview. In: Krieg, N.R., Holt, J.G. (Eds.), *Bergey's Manual of Systematic Bacteriology*, vol. 1. Williams & Wilkins Co., Baltimore, pp. 1–4.
- Stalpers, J., Kracht, M., Janssens, D., Deley, J., Vandertoorn, J., Smith, J., Claus, D., Hippe, H., 1990. Structuring strain data for storage and retrieval of information on bacteria in MINE, the Microbial Information Network Europe. *Syst. Appl. Microbiol.* 13, 92–103.
- Van Brabant, B., Dawyndt, P., De Baets, B., De Vos, P., 2006. A knuckles-and-nodes approach to the integration of microbiological resource data. *Lect. Notes Comput. Sci.* 4277, 740–750.
- Van Brabant, B., Gray, T., Verslyppe, B., Kyrpides, N., Dietrich, K., Glöckner, F.O., Cole, J., Farris, R., et al., 2008. Laying the foundation for a Genomic Rosetta Stone: creating information hubs through the use of consensus identifiers. *OMICS* 12, 123–1277.