



Priority actions to consolidate the Pt-mBRCN as MIRRI-PT Node

-Abridge Version-

ABSTRACT: In Portugal, there are more than 16 culture collections (CCs) with a rough estimation of more than 25,000 holdings, normally associated with taxonomic experts, covering different taxonomic groups. These CCs are linked with research or service activities, either in academia or governmental laboratories. The Portuguese landscape on this field goes from older CCs (e.g., the fungal collection EAM-INIAV funded in 1948 and hosted in the National Agronomic Station; the Portuguese Yeast Culture Collection-PYCC funded in 1952) to more recent ones; from more diverse taxonomic holdings to thematic ones; and with different organization and quality management levels. Furthermore, CCs size is variable, ranging from larger ones with about 5,000 preserved strains to smaller ones with around 300 preserved strains or less. This heterogeneity and fragmentation represents a constraint to underpinning life science and biotechnology developments. Associated with this scenario, Portuguese CCs lack digital documentation and online presence, i.e., e-catalogues. Only 5 CCs, partners in this project, have individual e-catalogues: PYCC <http://pycc.bio-aware.com/>, MUM <http://www.micoteca.deb.uminho.pt/>, ACOI <http://acoi.ci.uc.pt/>, LEGE <https://lege.ciimar.up.pt/> and, more recently, UCCCB <http://www.uc.pt/en/uid/ucccb>. Since 2017, the CCs have been working together in the Portuguese microBiological Resource Centre Network (PT-mBRCN, <https://mbrcn.pt/>) to have a common voice, share experience and foster cooperation to offer services that none single CC are not able to do. To accelerate biodiscovery and the development of innovative solutions for different bioindustries, preserved microorganisms are important assets that need to be available to the public with the guaranty of high quality and under the necessary legal framework. To improve the current national panorama, and in alignment with the work developed at European level through the Microbial Resource Research Infrastructure (MIRRI, www.mirri.org), this project aims to unite the microbial Portuguese CCs and foster the link between the collection community and microorganisms users via a common portal in a one-stop site. To build this PT-mBRCNNetwork, the future Portuguese MIRRI node (MIRRI-PT), a unique e-catalogue with at least 15,000 entries and e-services needs to be established. Simultaneously, quality control and quality assurance needs to be put in place within the network in order to define common procedures based in a quality management system, ensuring that the microbiological resources and associated data have high quality. To pursue these aims, a non-CC partner needs to be outsourcing (or partner, such it is CCG/Centro de Computação Gráfica in the INFRADEV-3 IS_MIRRI21 EU project) with an outstanding record of accomplishments on developing platforms that support information and operations as needed for Pt-mBRCN/MIRRI-PT node. The CCs' partners have a deep knowledge on microbial taxonomy and preservation techniques related with their holdings. The major challenge of this project is to find common standard procedures and develop a new strategy to operate as a single entity, collectively delivering biological materials and associated information to the user communities, as well as other relevant services. With this approach, the project expects to enlarge its membership with new data providers and associate/collaborator CCs, and to be data provider to other databases such as the European MIRRI e-catalogue, GBIF, Global Catalogue of Microorganisms, Life Watch, among others.

BACKGROUND: The majority of organisms in an ecological niche are microorganisms and are of critical importance for the sustainability of life on our planet. The abundance and diversity of microorganisms is breathtaking, as emphasized in multiple studies. For example, current estimates of the total number of bacterial cells on Earth ($4-6 \times 10^{30}$) outnumber the estimated number of stars in the universe by several orders of magnitude (10^{21}). Microorganisms, being either beneficial or detrimental, play essential ecological roles, interacting with plants and animals, and control vital global biogeochemical and nutrient cycles. Humans have been making direct use of microorganisms since the dawn of humankind – originally associated with the production of fermented foods and beverages, but during the past half



century as producers of antimicrobial agents and enzymes for application in various areas of modern-day biotechnology. Their importance in food production and agriculture, specially their role in soil quality, was recently highlighted in the FAO publication contributing to the Sustainable Development Goals from the UN, namely on zero hunger and secure food (Obj. 2).

Moreover, microorganisms are everywhere and, with about 150 years of microbiological cumulative studies, our knowledge about them is still limited and fragmented. Less than 1% of microbial diversity is known. This undiscovered diversity is predicted to host a large amount of new genetic material of bioeconomical importance. Only a small fraction of microbial diversity is preserved in *ex-situ* Culture Collections (CCs). Public service CCs have been performing this function for over 130 years and Europe has played a pioneer's role by creating the first public CCs in the late 1800's at the German University in Prague. Currently, more than 3 million microbial strains are available in the 771 collections listed in the WDCM. There are 12,170 strains in 7 Portuguese CCs registered in the WDCM, which represents a low fraction of the preserved microbial diversity. The CCs, and the Microbiological Resource Centres (mBRCs), which represent the most advanced and up-to-date concept CCs according to the OECD definition, are repositories that provide high quality microbial resources and information. mBRCs differ from CCs as they have implemented Best Practice Guidelines on BRC and quality-assurance through certification or accreditation according to dedicated standard norms, such as ISO 9001, ISO 17025 or the most recent one on biobanks ISO 20387. In addition, CCs provide the optimal environment for long-term maintenance, collect and store data and information about these microorganisms, conduct research on their holdings, frequently in collaboration with scientists at other institutions, and provide support and services related to microbial resources to their users. The CCs' curators are experts on identification/classification, cultivation, and preservation of different taxa; on regulatory, biosafety/biosecurity, access and benefit sharing (ABS; e.g., Nagoya Protocol, EU 511/2014), among many other legal issues such as international shipping of microbial resources.

Europe, due to its long tradition of preserving microbial diversity, has the foremost archive of strains maintained in CCs. Within the corporate members of European Culture Collections' Organisation (ECCO) it is estimated that around half a million strains are preserved. Nevertheless, the supply of these resources, associated data and service provision are highly fragmented among Member States. The Portuguese landscape is similar, which constitutes a large barrier for users who must consult different information systems with heterogeneous data and contact several CCs or mBRCs to access the wide range of resources. It is also well-known that there are unregistered exchanges of resources between scientists, leading to a loss of traceability, significant genetic drift and unregulated transfers of genetic material which probably do not follow legal provisions (transport, biosafety, biosecurity, ABS, etc.). In addition, it was shown during the FP7 EU funded EMbaRC project that only a small proportion of the studied bacterial strains are actually deposited in CCs (less than 1% in 2008), leading to non-referencing and poor standard conservation of a huge proportion of genetic material.

Collaboration and alignment between the European CCs started in the mid-1980's with several EU funding projects and the activities that have been developed under the ECCO umbrella. Such constant efforts on networking and a progressive integration of services over the years led to the preparatory phase of the Microbial Resource Research Infrastructure (MIRRI) from 2012 to 2016. Currently, MIRRI is preparing to be a legal entity (ERIC) and Portugal, as a Country Member, has the responsibility to host its European headquarters at University of Minho (UM). Following this trend, Portuguese CCs have been working together and the PT-mBRCN was successfully launched on December 2017, during the Congress of Microbiology and Biotechnology held in Porto (Portugal), with the motto of "bringing together national BRCs to support life sciences and biotechnology".

PLAN AND METHODS: The main objective of the current project is to launch a one-stop site platform where the Portuguese e-catalogue of microorganisms is available to deliver live strains and their associated information to the user communities and related services. Common procedures will be



established and quality control/quality assurance (QC/QA) will also be adhered. With this project MIRRI-PT will become the national node of the European MIRRI-ERIC and, in addition, MIRRI-PT will be data provider or linked with different Research Infrastructures (RIs) in the ESFRI Biological and Medical Sciences (BMS) Landscape and other initiatives such as GBIF, Global Catalogue of Microorganisms, Tree of Life, Encyclopedia of Life, among others.

To achieve this objective we propose a project structured as shown on the flow-diagram (Fig. 1). The interaction and data flow between the five tasks are main research plan and methods of this proposal and are summarised below. The CCs are services dedicated to supporting a wide variety of microbiological work and cover different taxonomic groups: archaea, eubacteria (including cyanobacteria), yeasts, filamentous fungi, and micro-algae. This coverage implies substantial differences in terms of methods of identification and long-term preservation. In addition, the CCs hold different strains that can be ex-type strains, reference strains, or strains without a sound and complete identification, including absence of DNA sequences identification. The data generated and the internal procedures, including QC/QA management system, are also different in scope and approach in each CC. Finally, for a CC to operate properly, after controlling the microbial identification and long-term preservation of its holdings, it needs to implement a quality management system (QMS) in order to control all internal procedures and external relationship with its users. A data management system associated to a QA/QC is strongly recommended by the WFCC guidelines, the OCDE best practices, in compliance with regional, national and international legislation for the sector. QC/QA is essential to ensure that CCs provide authenticated products and have the best services. This also allows the continuous improvement and a faster detection and resolution of issues regarding methods, collaborators or equipment. Improved customer satisfaction and overall increased laboratory activities are all the result of implementing a functional and successful QMS. In absence of a certified QMS for a third party, such as the ones imposed by ISO standards, and in the present partners MIRRI-PT consortium only UM-MUM is certified under ISO 9001 since 2011, this means that best practices must be implemented and QC/QA developed using common procedures, including internal audits in order to move forward for an independent certification/accreditation system based on the ISO 9001, ISO 17025 or ISO 20387. In addition, databases are also heterogeneous within CCs, in terms of architecture, language used or, in the Portuguese case, the majority of CCs use excel sheets to perform these activities. Out of 11 CCs, only 5 CCs have so far portals and databases that from the beginning of this project can be data providers to the common e-catalogue. The other ones they will use the PT-mBRCN platform to archive and deliver their data.

The platform will be the core of the Portuguese Microbial Resource Research Infrastructure. The main goal is to integrate information from different catalogues and repositories, facilitating the access to microbial resources. The platform will be able to integrate information from different existing data sources (catalogues) and provide a component to allow the creation of catalogue, in case of institutions that do not have digital catalogues of microorganisms. Each institution will be able to select the information to integrate and select who will have permissions (profiles) to visualize each type of information. In order to integrate the different catalogues (data sources) the platform will provide an integration layer (API) that will be developed during the project. To implement this integration layer, it will be adopted a multidisciplinary ontology to storing and searching for information when process large amounts of data. The platform will also support the acquisition/order of microorganisms and the stock management process. To guarantee the security of the microorganisms acquisition the platform will implement security mechanisms, ensuring data privacy, legal and legislative considerations and the EU General Data Protection Regulation (GDPR).

TASK 1 - CULTURE COLLECTIONS (STRAINS, DATA & PROCEDURES): In a decentralised research infrastructure, harmonised criteria, such as standard operation procedures (SOPs), are essential.



To increase the quality of the preserved biological resources (task 2) supplied to the scientific, industrial and services communities (task 4), MIRRI-PT CCs will implement common procedures respecting the national and international legal frameworks (task 3). For this, the principles of the OECD Best Practice Guidelines in combination with ISO 9001, ISO 17025, ISO 20387 and, eventually, ISO 21710 on “Microbial Resource Data”, will be considered. Each CC will be enquired about the available information of the strains. The minimum data set (MDS) contains the required data to generate the public catalogue of strains.

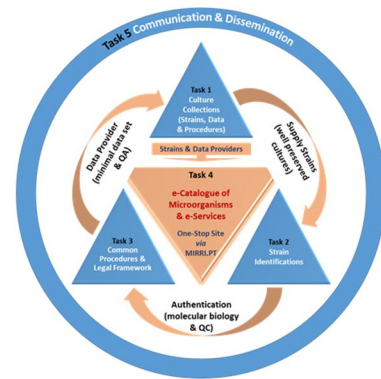


Fig. 1-Conceptual flow-diagram of the project.

The long-term preservation methods used in each CC will be surveyed, and a strategy for PLAN - DO - CHECK - ACT will be developed by each CC to requalify their holdings, confirm identifications (task 2) and provide data for task 4. The main deliverable of task 1 will be a gap analysis that will originate in-depth understanding of each CC in terms of procedures, QMS levels yet to be put-in-place and identification of needs that will be addressed in task 2 (authentication and identification of holdings) and task 3 developing common procedures within a legal framework.

TASK 2 – STRAIN IDENTIFICATIONS: Within CCs’ project holdings, some microorganisms, particularly the oldest ones, are not completely identified with molecular biology methods. This task aims to provide conditions to speed the process and requalify these holdings. In this task, each CC will be able to sequence at least 150 strains, in an expected 1600 strains, that were never molecularly identified or to confirm their identification using complementary housekeeping genes. Priority will be given to species not yet covered in the PT-mBRCN e-catalogue.

TASK 3 – COMMON PROCEDURES & LEGAL FRAMEWORK: The feasibility of implementing common SOPs and quality policies will be scrutinised through an exercise of internal audit (self-evaluation) in compliance with a QM requirements checklist. The results will be verified and completed by a team of 2 experts that will visit these CCs originating a detailed and confidential gap analysis for each CC audited. Key issues will be listed and possible solutions discussed. The output of this exercise will be a guide for implementation of QM in CCs. To provide structured and QC/QA validated strain data for the e-Catalogue of Microorganisms (task 4), the data set will be organised following: i) the requirements of Minimum Data Set/Full Data Set discussed in task 1; ii) the digital format defined in the task 4 to integrate, curate, deploy metadata (following FAIR criteria) and dedicated e-services in association with Mycobank, Yeast-id, Yeast-IP, Fungene, FungalDC, etc. In addition, this data will be linked to databases focused on macro-biodiversity and provide information for the construction of the European Open Science Cloud (EOSC-Life project) and LifeWatch-ERIC. At the end of the project, the deployment of a 15,000 strains dataset is envisioned as a result of the requalification/identification performed in task 2 along with the already available information from well characterised strains within CCs. The main deliverable of this task is the implementation of a common QM, including SOPs, legal requirements and data management in CCs. This QM will be published as a guide for implementation of a QMS in other CCs supporting the future MIRRI-PT consortium enlargement.

TASK 4 – e-CATALOGUE OF MICROORGANISMS & e-SERVICES: This task will allow the development of the MIRRI-PT platform. The effort will be divided in four subtasks: (1) Develop the platform backoffice; (2) Develop the platform frontoffice; (3) Develop the integration layers; (4) Implement the security layer.



Portuguese
microBiological
Resource Center
Network



The platform backoffice will implement functionalities to manage internally the microbial resource database, in a simple and intuitive way. The backoffice will allow to: Structure microorganism information (Minimum data Set/full data Set provided by tasks 2 & 3) and collections repository/catalogue; Manage microorganisms stock, Research information (papers, support to programs and projects), Events, Documentation, Links, News and Profiles and Users; Manage the process to select and acquire a microorganism, structure information for microorganism acquisition and manage microorganism orders (requests and answers). The platform frontoffice will implement functionalities to allow external visualization of the information. The frontend will allow the advanced search for the information microorganisms; Execution of the acquisition process; Visualization of the platform information (events, documentation, links, etc.); Downloads, Reports creation. This visualization component can be accessed in multi-devices (pc, tablets and smartphones) adapting the information according to the device. Legal and legislative considerations (provide by task 3) and the EU GDPR have to be also considered in the platform development.

TASK 5 – COMMUNICATION & DISSEMINATION: General Communication strategy: In order to share the information, the partners will develop and deploy information material that provide detailed information on the new results that will be generated by the project and how to efficiently implement them. The communication material includes not only the use of the site and the logo, but also a leaflet, a roll up and a newsletter that will be produced every six months. Press releases at local/national level aim to reach also the general public. The dissemination strategy has as central objective the transfer of knowledge and results with the aim to enable others to use and take up results, maximizing the impact of the funded research. It will comprise the presentation of results in the main conferences in the area, and in many other *fora*. Activities for lay citizens are also envisaged.

BUDGET FOR 36 MONTHS REQUIRED TO ESTABLISH THE PT-mBRCN (MIRRI-PT NODE): A global budget exercise to make the PT-mBRCN operational is presented, where the Human Resources and Equipment are the most key needs for an appropriated operation of Culture Collections. In addition, the informatics platform is also another key element in all structure which can be developed by CCG company. A rough estimation is presented in the fellow table:

Project title: Portuguese microBiological Resource Center Network (PT-mBRCN)													
Duration: 36 months													
Description	MUM	CBMA	CCG (outsourcing)	LEGE	LRIV (Aç)	IHMT	IVDP	INIAV	CIMOCC	PYCC	ACOI	UCCB	TOTAL
Partners	1	1	2	3	4	5	6	7	8	9	10	10	TOTAL
Human resources	140000.00	70000.00	160000.00	70000.00	70000.00	70000.00	70000.00	70000.00	70000.00	70000.00	70000.00	70000.00	1000000.00
Consumables	25000.00	25000.00	5000.00	25000.00	25000.00	25000.00	25000.00	25000.00	25000.00	25000.00	25000.00	25000.00	280000.00
Equipment*	750000.00	50000.00	60000.00	50000.00	50000.00	50000.00	50000.00	50000.00	50000.00	50000.00	50000.00	50000.00	1310000.00
Meetings	15000.00	6000.00	6000.00	6000.00	6000.00	6000.00	6000.00	6000.00	6000.00	6000.00	6000.00	6000.00	81000.00
Dissemination	5000.00	5000.00	5000.00	10000.00	5000.00	5000.00	5000.00	5000.00	5000.00	5000.00	5000.00	5000.00	65000.00
Total	930000.00	151000.00	231000.00	151000.00	151000.00	151000.00	151000.00	151000.00	151000.00	151000.00	151000.00	151000.00	2736000.00
Overheads	232500.00	37750.00	57750.00	37750.00	37750.00	37750.00	37750.00	37750.00	37750.00	37750.00	37750.00	37750.00	684000.00
TOTAL	1162500.00	188750.00	288750.00	188750.00	188750.00	188750.00	188750.00	188750.00	188750.00	188750.00	188750.00	188750.00	3420000.00

*The high budget required by the partner MUM for equipment is to allow the implementation on this CC the IDA for patents. The HR for MUM is also justified by na extra PhD to take care of patent deposits