euCanSHare

*An EU-Canada joint infrastructure for next-generation multi-Study **H**eart research*

List of participants:

Participant No.	Participant organisation name	Acronym	Country
1 (Coordinators)	Universitat Pompeu Fabra	UPF	ES
2	Research Institute of the McGill University Health Centre	MUHC	CA
3	Medical University of Hamburg-Eppendorf	UKE	DE
4 (SME)	Lynkeus Srl	LYN	IT
5	Erasmus Medical Centre Rotterdam	EMC	NL
6	Queen Mary University of London	QMUL	UK
7	McGill University	MCGILL	CA
8	Katholieke Universiteit Leuven	KUL	BE
9	Barcelona Supercomputing Centre	BSC	ES
10	Centre for Genomic Regulation	CRG	ES
11 (Society)	European Society of Cardiology	ESC	FR
12	Biobanking and Biomolecular Resources Research Infrastructure	BBMRI	AT
13	McMaster University, Hamilton Health Sciences	MCM	CA
14	National Institute for Health and Welfare	THL	FI
15	Medical University of Greifswald	UMG	DE
16 (SME)	Nostrum BioDiscovery	NBD	ES

Table of contents

1. Excellence

Cardiovascular research in the age of personalised medicine: Despite continuous advances in diagnosis and treatment, cardiovascular diseases (CVDs) remain the main cause of mortality worldwide, accounting for about a third of annual deaths (www.who.int/cardiovascular_diseases). Furthermore, they greatly reduce the quality of life of cardiovascular patients, who are estimated at a staggering figure of 85 million in Europe alone[1]. CVDs also challenge the financial stability of modern healthcare systems and have a negative impact on economic growth. While reduction in morbidity and mortality has been in fact noticed in acute coronary artery disease, there remains a high and increasing burden of many CVDs. Possibly the best illustration of the remaining needs are the unyielding/persistent high mortality rates in patients with atrial fibrillation even when evidence-based therapy is applied[2]. At the same time, while a reduction in novel drugs reaching the market has been observed in several disease areas, it is particularly salient in the cardiovascular domain[3]. In this context, personalised medicine approaches are urgently needed in cardiovascular research to improve risk assessment and early diagnosis, as well as for treatment personalisation and drug development. In 2014, our consortium member the European Society of Cardiology (ESC) put together a roadmap for new research in cardiovascular personalised medicine[4], calling for integrative data-driven approaches that will link molecular, imaging, functional and clinical data (cf. Figure 1). However, the ESC concluded their paper by emphasising that "*such integration presents a formidable challenge to data storage, management and analysis, IT capacity, and accessibility*".

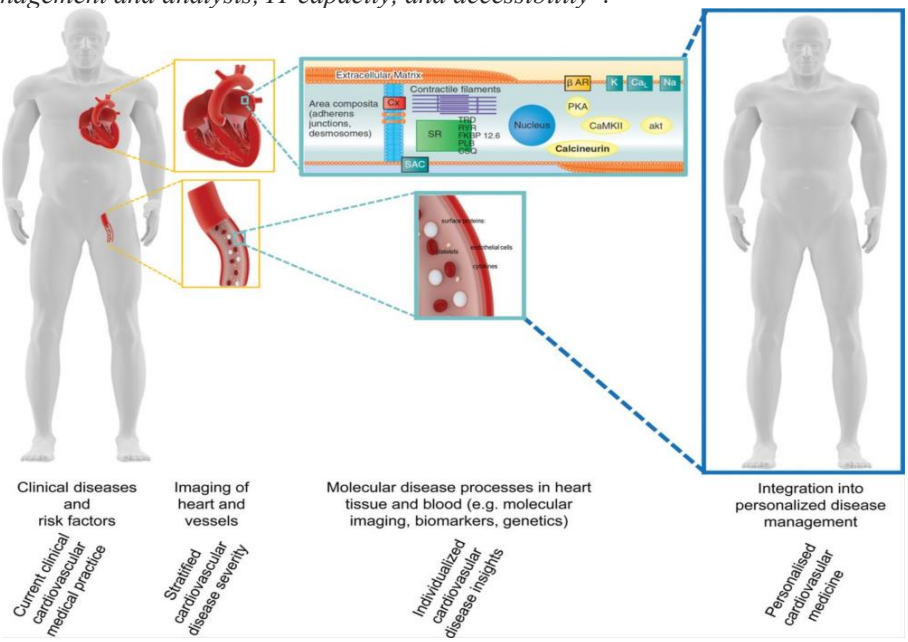


Figure 1 – The ESC roadmap from current clinical to personalised cardiovascular medicine, showing the need for enhanced integration of molecular, imaging, biomarker and genetic data, in addition to classical clinical and lifestyle/environmental factors.

European and Canadian expertise in multi-cohort cardiovascular research: For a long time, cardiovascular research had been conducted based on single cohorts such as the well-known Framingham Heart Study (www.framinghamheartstudy.org). These individual cohorts have greatly advanced the knowledge on cardiovascular risk factors. However, they are also limited to specific populations, geographical areas and even data types. Consequently, multi-cohort approaches have been proposed to allow researchers to investigate heterogeneous determinants and biomarkers of CVD in more comprehensive samples, and to uncover new biomedical knowledge with increased sample size, cohort utilisation, geographical coverage, and data richness[5].

---

[1] http://www.ehnheart.org/cvd-statistics/cvd-statistics-2017.html

[2] Kirchhof P. et al. Improving outcomes in patients with atrial fibrillation: Rationale and design of the Early treatment of Atrial fibrillation for Stroke prevention Trial (EAST), *American Heart Journal*, 2013, vol. 166 (pg. 442-448)

[3] Mullard A. FDA drug approvals. *Nature Reviews Drug Discovery*, 2013, vol. 12 (pg. 87-90)

[4] ESC European Affairs Workshop on Personalised Medicine: "The continuum of personalised cardiovascular medicine: a position paper of the European Society of Cardiology." *European Heart Journal* 35.46 (2014): 3250-3257.

[5] Anand, S. S., Tu, J. V., Awadalla, P., Black, S., Boileau, C., Busseuil, D., et al. (2016). Rationale, design, and methods for Canadian alliance for healthy hearts and minds cohort study (CAHHM)–a Pan Canadian cohort study. *BMC public health*, *16*(1), 650.

In this domain, Europe and Canada, thanks to efforts by several members of this consortium (UKE, THL, MCM, MUHC, MCGILL, LYN, QMUL, UMG), have been at the forefront of multi-cohort approaches in cardiovascular research. For example, THL was the Data Coordinator of the MONICA WHO project (www.thl.fi/monica) (1980 - 2002: >20 cohorts) for investigating risk factors of CVD worldwide, as well as of the MORGAM project (www.thl.fi/morgam) (started in 1998: >30 cohorts) for investigating risk factors and genetics in cardiovascular research. UKE was the Scientific Coordinator of the BiomarCaRE project (www.biomarcare.eu) (2011 - 2016, >40 cohorts), which focused on the investigation of cardiovascular biomarkers across Europe. In Canada, MCM is currently leading the Canadian Alliance for Healthy Hearts and Minds (CAHHM: www.cahhm.mcmaster.ca), with support from MUHC and MCGILL, for integrating eight Canadian cohorts covering several regions of the Canadian territory, with special attention dedicated to cardiac imaging for investigating subclinical cardiovascular disease and early diagnosis. All of these multi-cohort initiatives have focused mostly on data integration, harmonisation and exploitation, resulting in important new findings and fruitful collaborations within the consortiu.

However, despite these results, these initiatives did not develop the much-needed IT infrastructures to enable scalability and sustainability beyond the duration of the funded projects. In other words, they have not been implemented to be FAIR (Findable, Accessible, Interoperable and Reusable). Specifically:

- There are no catalogues that to enable easy access to information on available data, their precise characteristics and potential for cardiovascular research studies.
- Access request mechanisms remain highly traditional, manual and lengthy, which greatly reduce the much-needed efficiency in data-driven biomedical research.
- Each addition of a new cohort to the ensemble represents a tedious process, which discourages data exchange and enrichment.
- Re-usability of the data for cardiovascular personalised medicine is often reduced when the funding is terminated.
- More importantly, none of these initiatives have taken advantage of emerging infrastructures (ELIXIR, EGA, BBMRI, euro-BioImaging) or legal frameworks (GA4GH: Global Alliance for Genomics and Health; GDPR: General Data Protection Regulation), which reduces current security and trust.

euCanSHare – A joint EU-Canada effort to establish the computational, infrastructural, legal and scientific frameworks for next-generation data sharing and analysis in the cardiovascular domain:
euCanSHare emerges as the next natural step for consolidating previous efforts by the consortium and for enhancing multi-cohort cardiovascular research in the age of personalised medicine. Specifically, the project will focus on integration, interoperability and exploitation of established infrastructures (including ELIXIR, EGA, BBMRI, euro-BioImaging in Europe; Maelstrom in Canada), IT solutions developed by consortium members (*e.g.* Opal, Mica, Rocket, FASTR, Square^2 – see Concept Section), and major data sources from Europe and Canada (MORGAM, BiomarCaRE, CAHHM, UK Biobank), within a one-stop-shop web-portal comprising extensive services and functionalities for increased efficiency and data sharing in cardiovascular research like never before.

1.1 Objectives

euCanSHare is structured around five clear and quantifiable objectives that have been carefully defined to be achievable within the duration of the project and to make tangible impact in the domain of personalised cardiovascular medicine:

Objective 1: Build the first centralised, sustainable and "FAIR" platform for cardiovascular data sharing and analysis

euCanSHare will provide to the cardiac research community, from academia, industry, and public health, the first centralised, secure and easy-to-use platform to leverage European and Canadian cardiovascular research data and technologies, improve data discoverability, integration and co-analysis capacities, and lead to cutting-edge collaborative research in the domain of cardiovascular personalised medicine. Because it will build on solid expertise and technology by the consortium (see Table 2), the first version of the platform will be assembled mid-way through the project for extensive testing, while the final release will be fully operational by the end of the project. Ultimately, the euCanSHare platform and its associated legal framework will realise the promise of the FAIR principles, *i.e.* euCanSHare's integrated cardiovascular data will be:

1) more easily **Findable** thanks to the largest and most detailed cardiovascular data catalogue ever assembled;
2) more **Accessible** thanks to an integrated system of Data Access Committees (DACs) and new technologies based on the blockchain technology linked to state-of-the-art infrastructures (ELIXIR, EGA, euro-BioImaging, BBMRI);
3) **Interoperable** by leveraging emerging solutions for data exchange, harmonisation and quality control;

4) highly **Re-usable** thanks to open-source data analysis tools that will allow new knowledge discovery and disease quantification with validation enabled on rich and large multi-country datasets, integrating -omics, imaging, environmental and other clinical data.

Objective 2: Leverage European and Canadian expertise for building a highly comprehensive, robust, secure and scalable cardiovascular data platform

euCanSHare will leverage unique expertise in the field of data sharing, multi-study cardiovascular research and personalised cardiovascular medicine from both sides of the Atlantic. Specifically:

- BSC (coordinators of ELIXIR-ES), CRG (managers of the European Genome-phenome Archive: EGA), EMC (managers of the euro-BioImaging infrastructure) and BBMRI, will assemble a data management system linking these established infrastructures for integrating heterogenous cardiovascular data.
- THL (managers of MORGAM), UKE (coordinators of BiomarCaRE) and QMUL (leaders of the cardiac imaging of the UK Biobank) will provide state-of-the-art cardiovascular cohorts in Europe, as well as unique expertise in cardiovascular data integration, harmonisation and analysis.
- UPF will make available its cardiovascular data analysis platform Rocket developed in various cardiac-related EU projects (NEUBIAS, VP2HF, CardioFunXion), enriched with automated quality control (UMG), bioinformatics (UKE, UMG) and advanced machine learning capacities.
- From Canada, MCM and MUHC will contribute with the Canadian Alliance for Healthy Hearts and Minds (CAHMM) and its eight cohorts distributed across the Canadian territory.
- MCGILL will participate through its world-renowned Centre of Genomics and Policy (CGP), which hosts the Regulatory and Ethics Workstream of the Global Alliance for Genomics and Health (GA4GH).
- Finally, MUHC will contribute with its unique multi-study data management tools from the Maelstrom Project (www.maelstrom-research.org), will be adapted and offered open-source within euCanSHare for comprehensive cardiovascular cataloguing and data harmonisation.

This unique blend of expertise and complementary between the EU and Canadian partners will result in an unprecedented platform covering the whole spectrum of infrastructures, tools and functionalities that are needed to enable a great leap forward in multi-cohort cardiovascular research and personalised cardiovascular medicine.

Objective 3: Demonstrate the power of the euCanSHare platform for cardiovascular research in personalised medicine, integrating biomarker validation, knowledge discovery, risk assessment, public health and industrial research, through multiple inter-sectorial validation studies

To enable full adoption of the platform, its design, development and implementation will be guided by several pilot-test studies that will iteratively assess all the capacities of the platform, from cohort browsing, access applications, data access and study-specific computational workflowing (integrating harmonisation and quality control, omics, imaging, clinical and other cardiovascular variables/end-points). These studies will include:

1) A multi-country study to extract new knowledge on diabetic cardiomyopathy, a pathology/comorbidity which remains incompletely understood and poorly diagnosed in clinical practice;
2) A new risk prediction model integrating for the first time genetics and imaging from large samples for an earlier and more precise identification of individuals at risk of myocardial infarction and stroke;
3) A multi-country public health study to compare risk estimates and population attributable fractions (*i.e.* the fraction of all cases of a particular disease in a population that is attributable to a specific risk factor) across countries and regions (Europe and Canada).
4) An industrial study to assess and evaluate the extent to which biotech companies can exploit the platform for target discovery in preparation of clinical assays.

Through these different studies, it will be possible to assess the performance and usage of the data sharing and analytics platform in different settings, and constantly improve it based on the feedback of the end-users. This will ensure that the platform continuous developments will be guided by requirements of clinical researchers, epidemiologists and industrial developers, thus providing additional trust in the capabilities of the platform.

Objective 4: Develop the legal framework, as well as innovative solutions that will enable responsible data sharing and enhanced Open Science within euCanSHare and the cardiovascular research community

The consortium will perform a detailed ethical and legal interoperability analysis for identifying and addressing the complex ethical and legal issues that inevitably arise in an international research project of the scope of euCanSHare, not only due to the international nature of the project, but also to the diverse characteristics of the cohorts and the coexistence of different regulatory frameworks. They will put together the overall legal framework of euCanSHare by addressing the relevant commonalities, mismatches, and gaps between the European (GDPR)

and Canadian (federal) laws, regulations, and policies. euCanSHare will perform a comparative analysis of existing informed consent forms and information brochures to assess their interoperability, and identify a list of consent items essential for data use within the euCanSHare. Last but not least, the consortium will engage with participating cohorts to assess their motivations and barriers to data sharing and Open Science. Innovative approaches based on blockchain and Smart Contracts will be tested to facilitate data access and sharing processes by partially or fully self-executing/self-enforcing access requirements and data use conditions, while ensuring explicit tracking of academic credits.

Objective 5: Disseminate the euCanSHare platform at large to create the largest community of data contributors as well as users

The dissemination, outreach and exploitation of euCanSHare will be of the highest quality and depth to ensure the platform becomes the reference in data sharing, management and analysis in the cardiovascular domain in Europe, Canada and beyond. Intensive awareness campaigns and dissemination events will be coordinated by our consortium member the European Society of Cardiology (ESC) through its extensive channels and networks, in collaboration with major players in the field of data infrastructures and data sharing within the consortium in both Europe and Canada (BBMRI, ELIXIR/EGA, euro-BioImaging, Maelstrom, MCGILL, MCM). The ultimate goal is to promote the extensive infrastructural and personalised medicine research capacities of the euCanSHare platform, attracting diverse stakeholders for enriching its catalogue with additional cohorts beyond the initial set of datasets (see Table 1), as well as for forming a large community/network of users from the academic, industrial (*e.g.* pharmaceutical) and healthcare sectors to support its exploitation and scalability. To list a few, these dissemination activities will include large-scale live demonstrations at the ESC Annual Congresses (>35,000 participants), dedicated workshops with hands-on sessions both for academic and industrial stakeholders, outreach activities to external cardiovascular initiatives (*e.g.* BigData@Heart IMI project, see letter of interest in the Annex) and Knowledge-to-Action seminars for discussing impact on cardiovascular practice and cardiovascular health.

1.2 Relation to the work programme

The euCanSHare project is addressing the topic SC1-BHC-05-2018: "International flagship collaboration with Canada for human data storage, integration and sharing to enable personalised medicine approaches", as follows:

Scope (text of the call)	How it will be addressed in euCanSHare
"This programme aims to enhance and standardise data deposition, curation and exchange procedures thus ensuring better data reuse and increased benefit to the scientific communities worldwide".	The platform will integrate major infrastructures and solutions (ELIXIR, EGA, Opal, Mica) to enhance and standardise data deposition, harmonisation and sharing procedures in the cardiovascular domain.
"The selected projects should build on the data quality metrics, standards and access policies developed by major international initiatives".	euCanSHare will build on the data quality metrics, standards and access policies developed by the MORGAM, BiomarCaRE and CAHHM projects, in close link with ELIXIR and BBMRI.
"Projects should develop approaches that integrate data from disparate sources".	euCanSHare will implement solutions for integrating more than 35 cohorts from disparate sources, including -omics, cardiac imaging and clinical data.
"Data models that guarantee the interoperability of human health research data from different repositories and integrate different types of –omics data and, where relevant, clinical research and lifestyle data".	Building on the experience of our consortium with past multi-cohort cardiovascular studies (MONICA, MORGAM, BiomarCaRE, CAHHM), data models will be implemented to guarantee interoperability of multi-type cardiac research data (from -omics to imaging).
"The data models should take into account sex/gender differences where relevant".	Thanks to the number and large size of the cohorts in euCanSHare, sex aspects in CVD will be explicitly addressed as part of the project and its analysis tools.
"The projects should build on existing research infrastructures such as –omics repositories, biobanks and registries".	euCanSHare will integrate services and repositories from ELIXIR for the centralised web-portal, EGA for the -omics data, euro-BioImaging for cardiac imaging, and BBMRI for bio-samples.
"Reference architecture for data and process interoperability".	euCanSHare's references architecture is designed to enable interoperability of data, but also of the different tools and services (*e.g.* Opal, Mica, Rocket, DACs).
"Technologies and methodologies for data harvesting,	euCanSHare's technologies for data harvesting, access,

data access, data transfers, and archiving complex datasets".	transfers and archiving will be based on the EGA, Maelstrom, euro-BioImaging and MyHealth-MyData.
"Bioinformatics toolbox to support the analysis and management of data on diseases from a personalised medicine standpoint".	A comprehensive bioinformatics tools will be developed for the processing of cardiac -omics data, also supporting co-analysis of -omics and imaging data.
"International ethical and legal governance model for a research data management and storage infrastructure".	MCGILL and KUL will prepare an international ethical and legal governance model for responsible research data management and storage in euCanSHare.
"An associated data management plan compliant with the required level of data security and privacy that is aligned with the recent recommendations of the OECD Council on Health Data Governance".	BSC (ELIXIR) and CRG (EGA), with support from MCGILL as part of the OECD Working Group of the Health Data Governance, will build a highly compliant data management plan.
"This topic raises important issues of data sharing, privacy protection, informational right to self-determination and data security, which should be addressed from a legal, ethical as well as a social sciences perspective".	MCGILL and KUL will perform in-depth comparative analysis and put together a legal framework for addressing the relevant commonalities, mismatches, and gaps between the European (GDPR) and Canadian (federal) laws, regulations, and policies.
"It is important that proposals enable sustainable, collaborative projects".	Sustainability will be enabled by using established infrastructures such as ELIXIR, EGA and BBMRI.
"… ensure cross-references with existing infrastructures (e.g., BBMRI-ERIC, ELIXIR) and other on-going initiatives (e.g., International Consortium for Personalised Medicine, European Open Science Cloud: EOSC, IHEC, etc.)".	Cross-references will be explicitly implemented between the ELIXIR/EGA, euro-BioImaging and BBMRI. Also, BSC and BBMRI are part of the EOSCpilot project (www.eoscpilot.eu) and will thus support the implementation of the EOSC guidelines.
"The proposals should take stock of the BBMRI-ERIC Code of Conduct for using personal data in health research".	BBMRI and the ESC are collaborating for integrating the BBMRI-ERIC Code of Conduct for using personal data in cardiovascular research.
"A multidisciplinary approach, i.e., involving clinicians, biologists, bioinformaticians, etc., is considered a key aspect of successful proposals".	The euCanSHare consortium includes clinicians (QMUL, MCM), bioinformaticians (BSC, UKE, UMG), imaging experts (UPF, QMUL, EMC), technologists (UPF, CRG, BSC), data sharing experts (MCGILL, KUL, MUHC, THL), SMEs (LYN, NBD) and cardiovascular researchers (QMUL, MCM, UKE).
"Due to the specific challenge of this topic, in addition to the minimum number of participants set out in the General Annexes, proposals shall include at least one participant from Canada".	euCanSHare involves three Canadian partners, one bioethics/legal team (MCGILL), data sharing experts (MUHC) and cardiovascular researchers (MCM: leaders of CAHHM).

1.3 Concept and methodology

(a) Concept

euCanSHare's platform: The project will build upon established multi-cohort initiatives in cardiology (MORGAM, BiomarCaRE, CAHHM), as well as emerging comprehensive databases such as the UK Biobank and the Hamburg City Health Study (see Table 1). The platform will also integrate a range of established technologies developed by our consortium members (cf. Table 2) to offer comprehensive computing infrastructures to its users, encompassing software (for cataloguing, harmonisation, co-analysis and storage of study-specific and harmonised data); methods (supporting complex data integration models); high-performance computing and compute cloud (for data processing and storage); training and user support resources (*e.g.* guidelines, tutorials and workshops informing users on methods, resources and policy tools offered), and a central web portal (allowing secure and user-friendly access to the platform catalogues and functionalities). Figure 2 represents the main components and functionalities of the platform, which include:

1. **A cohort browser & catalogue** for enabling researchers to easily find information on relevant cohorts, their characteristics, list and description of variables, data access policies/procedures, and harmonisation potential across cohorts. For this purpose, our multi-disciplinary experts (THL, MUHC, MCM, EMC,

UKE, QMUL) will build the largest and most comprehensive multi-cohort catalogue ever put together in the cardiovascular domain. The Mica advanced cataloguing software developed by MUHC within its Maelstrom Research programme (www.obiba.org/pages/products/mica) will be leveraged to integrate detailed information on available lifestyle/environmental factors, physical measures, and -omics data. It will also be extended to also support cardiac imaging metadata (imaging modalities, protocols, parameters and biomarkers), as well as to provide information on cohort-specific access policies and consent requirements. Metadata fields will be initially populated with information obtained from participating cohorts in Table 1 and extended with new cohorts through the different awareness campaigns and outreach activities. The final Mica-powered euCanSHare platform will include Maelstrom's powerful search engine for allowing investigators to quickly find the information, variables and data they need for implementing cardiovascular research projects.

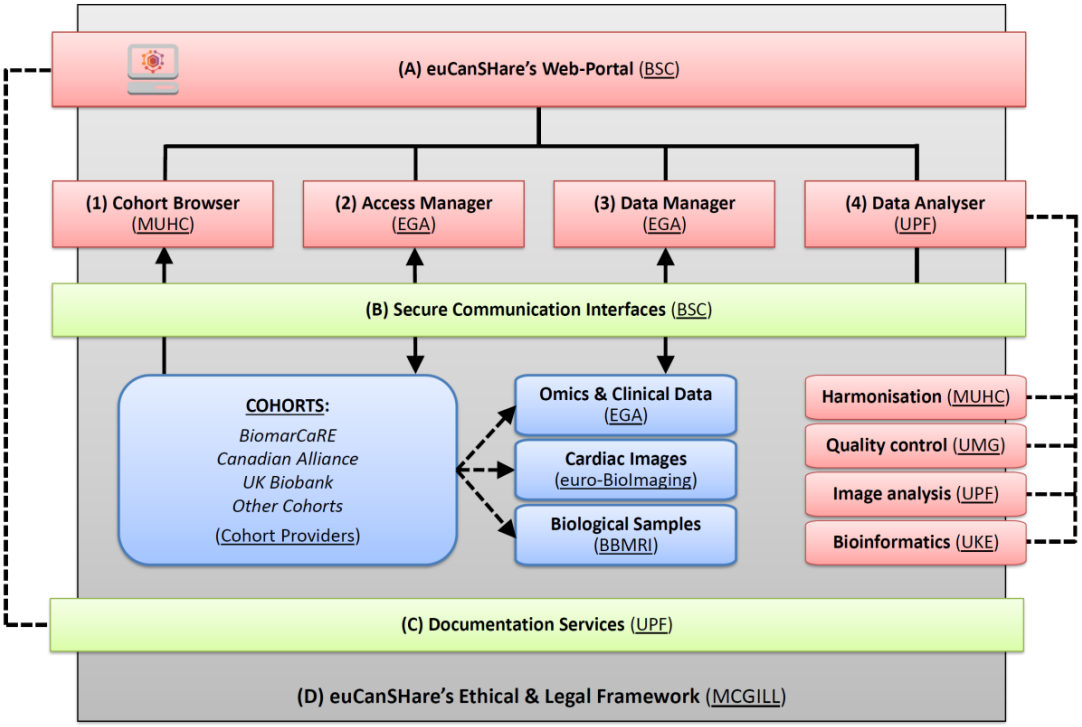


Figure 2 – *Schematic diagram showing the main components and interactions within the centralised euCanSHare platform (A). It will provide functionalities for (1) comprehensive cohort cataloguing and browsing, (2) secure access management, (3) data deposition and management, and (4) advanced data analysis including harmonisation, image analysis and bioinformatics capabilities. Integration and interoperability will be enforced based on technologies developed at BSC and CRG as part of ELIXIR and EGA, respectively (B). A documentation service (C) will be assembled by all consortium members throughout the project to facilitate the use of the platform beyond the duration of the project, in particular for data cataloguing and deposition. Note that detailed comparative analysis of legal frameworks in Europe and Canada will be performed by world-renowned experts from both sides of the Atlantic to inform the shared automation of all services and datasets (D).*

2. **An access manager:** This key component will be dedicated to ease the process of requesting access to the different cohorts and acquiring access credentials. The task will leverage strategies already being assayed in the EGA infrastructure, namely the Data Access Committees (DACs: www.ebi.ac.uk/ega/dacs). Through MCGILL, one of the original drivers of the Global Alliance for Genomics and Health (GA4GH), this sub-portal will disseminate access policies and procedures aligned to GA4GH using the Automatable Discovery and Access Matrix (or ADA-M: www.github.com/ga4gh/ADA-M). The Access Manager component will be provided as a simple user interface for researchers to apply for data access, as well as for cohort owners to facilitate the procedure of granting and managing granted credentials. For selected cohorts, the possibility of automatic credentials assignment based on applications and policies metadata will be explored through a blockchain technology developed by our SME member LYN in the MyHealth-MyData EU project (www.myhealthmydata.eu).
3. **A data manager:** This component will integrate (i) established infrastructures (ELIXIR, EGA, BBMRI, euro-BioImaging) for managing storage of the heterogenous cardiac data, (ii) tools/procedures for uploading new cohorts into the platform, and (iii) APIs for securely connecting the data to the relevant

users and data processing environments (*e.g.* ELIXIR). Metadata across the different repositories will be linked (-omics, clinical, imaging and bio-samples) to assure the transversal consistency of the specifications, thus presenting the heterogeneous cardiac data to the users in an integrated manner. The Data Manager will build on euCanSHare's Data Management Plan (DMP), which will include a comprehensive analysis of the nature of data to be handled, the requirements for interoperability, long-term storage and security issues, as well as compliance with FAIR and EOSC data principles.[6]

4. **A data analyser:** This subportal of euCanSHare will provide state-of-the-art tools for enabling modular, reproducible and cost-effective analysis of multi-source cardiac data, after access is granted. This will include data harmonisation (Opal by MUHC: www.obiba.org/pages/products/opal), data quality control (Square[2] by UMG[7]), cardiac data analysis (Rocket developed by UPF during VP2HF and CardioFunXion EU projects: FATSR by EMC: www.fastr.readthedocs.io) and selected bioinformatics capabilities (UKE, UMG). These tools will be made interoperable within customised pipelines through the well-established workflowing environment Galaxy (www.usegalaxy.org) and the Docker container technology (www.docker.com). euCanSHare's data-analyser will constitute a major incentive for cohort owners to join the platform as it will provide them with state-of-the-art capabilities to harmonise and analyse data across disparate sources, in particular for those that are experts in one domain (*e.g.* -omics) but lack the tools for integrating other complimentary cardiac data (*e.g.* imaging).

To ensure interoperability of these capacities, the platform will include four transversal components:

A. **The centralised web-portal,** which will constitute the entry point and main interface to the platform, through which the users will have access to the four main functionalities described above. This base framework will build largely on MuGVRE (vre.multiscalegenomics.eu), a cloud-based infrastructure developed by BSC. Presented to the multiscale genomics community in November 2017, MuGVRE has received more than 1,000 analysis runs in its first three months of existence, which shows its attractiveness and potential. To enable its sustainability, euCanSHare's web-portal will be hosted by BSC at the same ELIXIR compute equipment hosting MuGVRE.

B. **A secure communication interface**, which will integrate state-of-the-art protocols and technologies for ensuring interoperability and communication between components, secure data distribution, and cloud-based execution of tools. It will include easy-to-use modules for authentication (based on KeyCloak) and secure data management (Oauth2 protocol for all encrypted data transfers). A robust protocol to plug-in tools (data browsers, visualisers, or analysers) on top of the main infrastructure has already been developed by BSC[8]. The framework will provide advanced execution scheduling (using COMPSs, NextFlow), and OpenGrid Engine), allowing to derive tools execution to remote cloud infrastructures (through the OCCI protocol) and also to HPC environments within ELIXIR and the EGA.

C. **A documentation service** that will take the user step-by-step through all the different components and functionalities of the platform. This will ensure that future researchers and developers can easily leverage the platform not only for accessing information and cohorts, but also to deposit their own datasets and tools well beyond the duration of the project. Each document and user manual will be easily accessible through the web-portal similarly to the online Help Service of MuGVRE (see[9]).

D. **An ethical and legal framework** that will provide direct and easy-to-follow guidance for enabling compliant data transactions and sharing operations. This will facilitate responsible data sharing and increase trust in the processes and infrastructure. Cultural and institutional barriers to data sharing will be extensively studied so that the resulting platform implementation lowers concerns for organisation and individual researchers to share their data, while guaranteeing legal and ethical compliance as they do so.

---

[6] Mons, B., Neylon, C., Velterop, J., Dumontier, M., da Silva Santos, L.O.B. and Wilkinson, M.D., 2017. Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud. *Information Services & Use*, *37*(1), pp.49-56.

[7] Schmidt, C.O., Krabbe, C., Schössow, J., Albers, M., Radke, D. and Henke, J., 2017. Square2-A Web Application for Data Monitoring in Epidemiological and Clinical Studies. *Studies in health technology and informatics*, *235*, pp.549-553.

[8] http://www.multiscalegenomics.eu/MuGVRE/integration-of-tools/

[9] https://vre.multiscalegenomics.eu/help/upload.php

Use case – Cardiovascular research before euCanSHare: Steffen is a clinical researcher working in the field of cardiology and currently seeking to study the impact of Diabetes Mellitus (DM) on the onset of cardiovascular disease. The pathology is known as diabetic cardiomyopathy (DM-CM) and while it can be asymptomatic at early stage, it increases the risk of heart failure and mortality. It is currently challenging to diagnose DM-CM with conventional clinical approaches, yet, at subclinical stage, correct diagnosis and early targeted treatment approaches may prevent downstream morbidity and mortality. Thus, new knowledge and biomarkers are greatly needed for preventing, diagnosing and treating heart failure in at-risk diabetic patients. For investigating DM-CM, Steffen has access to a local population cohort of 5,000 individuals. However, only a limited number of cases are found in the database with diabetic cardiomyopathy, which reduces the statistical power of the research. At the same time, the data is not heterogeneous enough to study the impact of environment, ethnicity and genetics. He realises that a multi-cohort approach is needed for this study, but due to the lack of a centralised and rich catalogue of cohorts in the cardiovascular domain, Steffen finds it difficult to survey additional cohorts of relevance to the study. He decides to contact a number of cohort owners in Europe through e-mails but access to information on data suitability and data requests using this method is a lengthy and tedious process. He finds out about the euCanSHare platform and registers as a new user.

Using euCanSHare: Steffen queries the platform's catalogue for the cohorts that meet his specific criteria for inclusion in this study, *i.e.* DM-CM patients with imaging, genetic, clinical and environmental data. The system recommends the UK Biobank, the Hamburg City Health Study, and the Canadian Alliance CAHHM, among others. Steffen quickly instantiates a study-specific Data Access Application copying forward required information from previous studies before sending it in one click to eight different cohorts of interest. He immediately receives responses through the system that his research profile and project fit the access policies of three cohorts, for which access is granted, with directions for accessing the data on EGA's server. With these three data sets, he is able to run preliminary descriptive statistics to further refine his data search and add few study end-points. The other cohorts require additional steps, including completing forms that are automatically forwarded to the Data Access Committees through EGA's system. Subsequently, Steffen finds the computational tools (harmonisation, radiomics, bioinformatics, machine learning) that are relevant to the characteristics of the study and builds computational workflows on the platform. This leads to the estimation of new multi-factorial biomarkers that are descriptive of diabetic cardiomyopathy in a large heterogeneous sample, which can be further validated for early detection and patient screening. Moreover, the effects of environmental, ethnical, genetic and lifestyle differences between Canada and different regions of Europe are uncovered and disseminated to the cardiovascular research community through the Journal of Cardiovascular Diabetology and the ESC's distribution channels. Finally, he initiates a discussion with partners in the euCanSHare Network to assess follow-up studies and projects, as well as potential impact on new healthcare practice and policy strategies for a more holistic approach to managing diabetes and cardiovascular disease.

euCanSHare's preliminary list of cohorts, infrastructures and services: Our strategy in euCanSHare is to integrate in an initial phase a reasonably diverse set of cohorts (35 cohorts) into the catalogue and data repositories, based on four different sources, namely:

1) The MORGAM data (www.thl.fi/publications/morgam/cohorts/full/contents.htm), managed by THL;
2) BiomarCaRE cohorts (www.biomarcare.eu), coordinated by UKE;
3) The Canadian Alliance CAHHM (www.cahhm.mcmaster.ca) coordinated by MCM;
4) Other major cohorts not yet integrated into these networks such the UK Biobank (QMUL) or the highly rich Study of Health in Pomerania – SHIP (UMG).

We opted for a reduced initial list of 35 cohorts (approx. 20% from Canada) to ensure high quality data cataloguing and deposition, while allocating the resources to the development and provision of the tools and processes that will allow cohort owners to integrate their new cohorts into euCanSHare during the project and beyond. However, this initial list detailed in Table 1 below provides a great diversity of cohorts in terms of their characteristics, sample sizes, geographical distribution and data types, ensuring that the implementation of the catalogues and data management plans take into account this diversity and complexity from the start.

Another strength of the consortium for realising the promise of euCanSHare consists in the range of established tools and datasets, as well as expertise and experience accumulated by its members over the years (see Table 2 below). This puts the consortium in a favourable position to robustly implement and disseminate a fully functional and interoperable platform for sustainable multi-cohort cardiovascular research. As such, the project's methodology and resources will be dedicated to the optimal integration and interoperability of the different datasets (Table 1) and tools (Table 2) into the euCanSHare's platform, by developing specific solutions for ensuing alignment with the FAIR and EOSC principles, interoperability at all levels of the platform, and compliance with the ethical and legal frameworks (GDPR, GA4GH) – see Methodology Section and WPs for more details.

***Table 1** – List of initial cohorts to be integrated as part of the euCanSHare project (in light blue the European cohorts, in light red the Canadian cohorts), for a total of about one million records.*

Cohort name	Country	Partner	Size	Socio-demographics	Bio-samples	Omics	Cardiac imaging	Lifestyle data	Environmental data	Physical measurements	Clinical outcomes
Hamburg City Health Study	DE	UKE	>45,000	x	x	x	x	x	x	x	
UKE Clinical Cohort Studies	DE	UKE	>10,000	x	x	x	x	x	x	x	x
Study of Health in Pomerania	DE	UMG	>8,000	x	x	x	x	x	x	x	x
UK Biobank	UK	QMUL	500,000	x	x	x	x	x	x	x	x
Barts BioResource	UK	QMUL	>10,000	x	x	x	x	x	x	x	x
Alpha-Tocopherol Beta-Carotene Prevention	FI	THL	29,133	x	x	x		x		x	x
Caerphilly Prospective Study	UK	THL	2,171	x	x			x		x	x
The ESTHER Study	DE	THL	4,971	x	x			x		x	x
Estonian Genome Centre University of Tartu	EE	THL	4,791	x	x			x	x	x	x
The National FINRISK Study	FI	THL	38,333	x	x	x		x	x	x	x
DAN-MONICA	DK	THL	7,582	x	x	x		x	x	x	x
Kooperative Region Augsburg	DE	THL	17,264	x	x	x		x		x	x
Moli-Sani Project	IT	THL	24,325	x	x	x		x	x	x	x
MONICA Brianza	IT	THL	4,932	x	x	x		x	x	x	x
MONICA Catalonia	ES	THL	5,505	x	x			x	x	x	x
MONICA Friuli	IT	THL	5,510	x	x			x	x	x	x
MONICA Kaunas	LT	THL	4,485	x				x	x	x	x
MONICA Newcastle	AU	THL	5,873	x				x	x	x	
MONICA Northern Sweden	SE	THL	12,013	x	x			x	x	x	x
MONICA Novosibirsk	RU	THL	11,458	x				x	x	x	x
MONICA Warsaw	PL	THL	5,577	x				x	x	x	x
Prospective Study of Myocardial Infarction	UK & FR	THL	10,600	x	x	x		x		x	x
The Tromsø Study	NO	THL	31,847	x	x			x	x	x	x
AtheroGene	DE	UKE	3,800	x	x	x		x		x	x
Kardiologischen Anschlussheil Behandlung	DE	UKE	1,206	x	x					x	x
StenoCardia	DE	UKE	2,000	x	x					x	x
MEDUSA Atherosclerosis Study	IT	THL	8,512	x	x			x	x	x	x
Longitudinal Study of Ageing Northern Ireland	UK	UKE	8,500	x	x	x		x	x	x	x
PAMELA Arterial Pressure Study	IT	THL	2,044	x				x		x	x
British Columbia Generations Projects (BCGP)	CA	CAHMM	486	x	x	x	x	x	x	x	x
Alberta's Tomorrow Project (ATP)	CA	CAHMM	458	x	x	x	x	x	x	x	x
Ontario Health Study (OHS)	CA	CAHMM	2,931	x	x	x	x	x	x	x	x
CARTaGENE	CA	CAHMM	1,418	x	x		x	x	x	x	x
Atlantic Partnership for Tomorrow's Health	CA	CAHMM	854	x	x	x	x	x	x	x	x
Prospective Urban Rural Evaluation (PURE)	CA	CAHMM	1,058	x	x		x	x	x	x	x
Montreal Heart Institute (MHI) Biobank	CA	CAHMM	350	x	x		x	x	x	x	x

Table 2 – *List of existing tools, infrastructures and networks to be integrated within the euCanSHare project*

Tools	Partners	Capabilities
MuGVRE	BSC	Web-portal developed and hosted at ELIXIR by BSC
Mica	MUHC	Data cataloguing and cohort browsing
ELIXIR/EGA	CRG	Multi-cohort data management and storage
Euro-BioImaging	EMC	Image storage and analysis
BBMRI guidelines	BBMRI	Bio-samples deposition and quality assurance
MH-MD	LYN	Blockchain technology for multi-cohort data access management
Opal	MUHC	Data harmonisation and integration
Rocket	UPF	Platform for cardiovascular data analysis
Square2	UMG	Automated data quality control
MORGAM	THL	EU cohorts and expertise on data harmonisation
BiomarCaRE	UKE	EU cohorts and expertise on data harmonisation
CAHHM	MCM, MUHC	Canadian cohorts and expertise on data harmonisation
UK Biobank	QMUL	Large-scale population research (>500,000 cases)
EOSC, FAIR	BSC, CRG, BMMRI	Principles for enhanced cloud-based sharing of human data
GA4GH	MCGILL	Framework for responsible sharing of genomic and health data, used e.g. as the governance framework of CAHHM

Positioning of the project:

The concept of a centralised data platform for cardiovascular research is new and therefore the project will start at **TRL2** (technology concept formulated). However, it will be integrated from existing technologies (*e.g.* Opal, Mica, Square2, EGA, euro-BioImaging), which will be extended, moving from mature but localised and non-fully automated functionalities (**TRL5**) to broader and more robust usage environments for cardiovascular research (**TRL7**). The platform will then undergo all the steps of integration and iterative testing with all types of studies (including for risk assessment, knowledge discovery, biomarker quantification, public health and finally testing by an SME for target discovery – see WP6). This means by the end of the project, the euCanSHare will have been fully integrated, tested, and populated with a large set of cohorts from Europe and Canada (CAHHM, MORGAM, BiomarCaRE), making it ready to be fully deployed for personalised medicine research, thus reaching **TRL9** (system proven in operational environment) by the end of the project.

Given the exceptional expertise of our consortium members, we are reasonably confident that the euCanSHare platform will be fully operational by the end of the project since, as mentioned above, it will be built through the re-use of technologically mature components already in use in public domains and fully operational in other clinical areas. For example, the MuGVRE centralised platform developed by BSC for multi-scale genomic analysis (hosted at ELIXIR-ES), which will be used at the basis framework for assembling the euCanSHare platform and making the different components interoperable, was deployed two years after the start of the MuG project. Note that for some specific tools within the platform, such as the Rocket infrastructure for cardiovascular data processing, developed at UPF during past EU projects (*e.g.* VP2HF), the goal of euCanSHare is to bring them to full maturity as to prepare introduction in the commercial space in the medium term (see Exploitation Section).

Link to national and international initiatives:

Project	Contribution to euCanSHare
EOSCpilot (2017-2018) www.eoscpilot.eu	BSC and BBMRI, members of ESOCpilot, will link euCanSHare's development to the ESOC's principles.
MuG EU project www.vre.multiscalegenomics.eu	BSC will adapt the basis framework, interoperability solutions and documentation services of the MuGVRE platform developed during MuG for assembling euCanSHare.
MORGAM / BiomarCaRE www.thl.fi/morgam www.biomarcare.eu	Through these initiatives, THL and UKE will provide the initial 20 European cohorts, as well as the harmonisation models, for euCanSHare
CAHHM www.cahhm.mcmaster.ca	MCM will provide metadata, data and expertise on seven Canadian cohorts that will included in euCanSHare
ELIXIR-ES www.elixir-europe.org	For hosting the euCanSHare platform (BSC).

EGA, euro-BioImaging, BBMRI www.ega-archive.org www.eurobioimaging.eu www.bbmri-eric.eu	- Hosting the euCanSHare data (-omics, imaging and bio-samples). - Quality assurance according to the relevant guidelines and codes of practice. - Sustainability beyond the duration of the project.
Maelstrom www.maelstrom-research.org	MUHC will offer (1) Mica for building the cardiovascular catalogue and cohort browser; (2) Opal for building the harmonisation database and harmonisation tool for euCanSHare.
NEUBIAS, VP2HF and CardioFunXion www.vp2hf.physense.upf.edu www.upf.edu/web/cardiofunxion	UPF will offer the Rocket platform for cardiovascular data processing: www.upf.edu/web/bia/rocket
MyHealth-MyData www.myhealthmydata.eu	LYN will adjust the blockchain technology for automated secure access / scientific authorship management
BigData@Heart IMI project www.bigdata-heart.eu	Interactions will take place for adding new cohorts into euCanSHare (see Letter of Interest in Annex from Scientific Coordinator of BigData@Heart).
UK Biobank www.ukbiobank.ac.uk	One of the largest population cohorts in Europe. Integration of a subset of the variables most relevant to cardiovascular research (QMUL).
ENGAGE www.euengage.eu	MCGILL and KUL will exploit their work for legal interoperability analysis in ENGAGE.
GA4GH www.ga4gh.org	MCGILL hosts the Regulatory and Ethics Work Stream of GA4GH, which will be leveraged in euCanSHare to build the ethical/legal framework.

(b) Methodology

After presenting the overall concept behind euCanSHare, this section will describe in more detail the specific methodology for the implementation of the platform and the integration of the state-of-the-art functionalities/infrastructures provided by the consortium members (see illustrations in Figure 3).

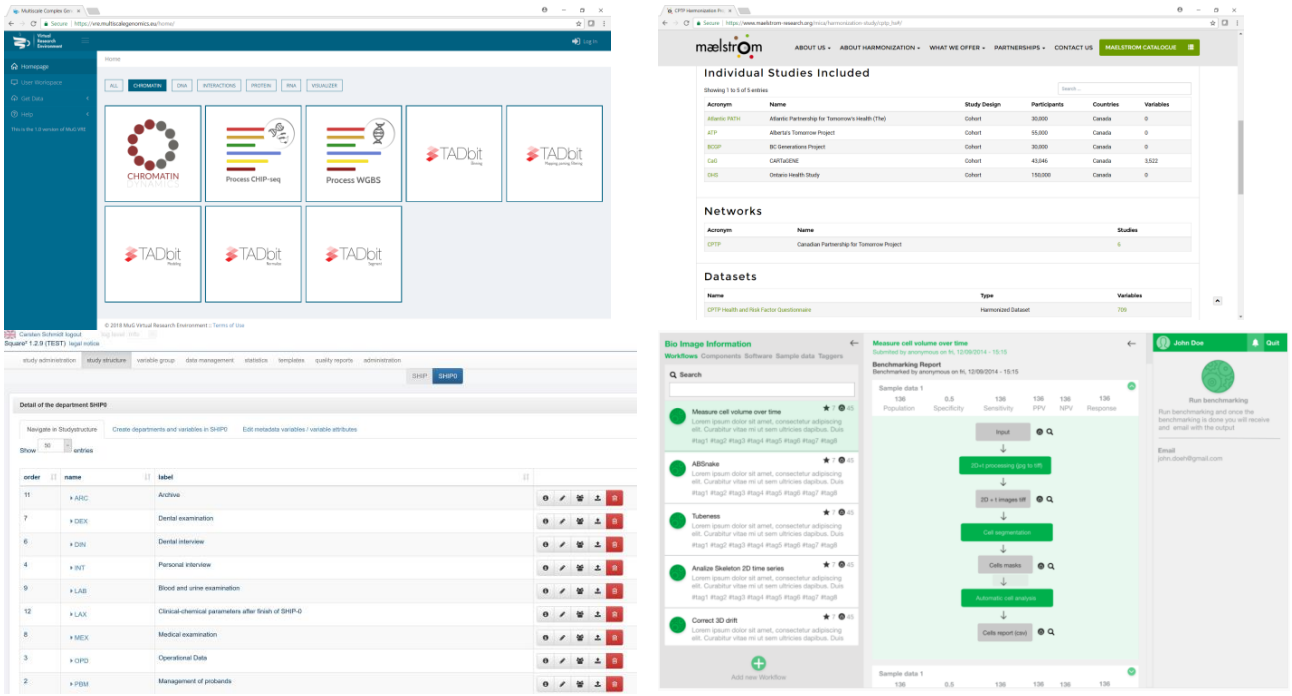


Figure 3 – *Screenshots to illustrate some of the tools that euCanSHare will adapt and integrate for multi-cohort cardiovascular research. On the top-left, the MuGVRE web-portal developed by BSC and hosted at ELIXIR contains several sub-modules for different aspects of genomics data management and analysis, integrated and made interoperable using appropriate interfacing solutions that will be re-used and adapted for integration in the euCanSHare platform. On the top-right, MUHC's web-tool for high-quality multi-cohort cataloguing and browsing will be exploited for building the largest and most comprehensive multi-cohort catalogue and data browser in cardiology. On the bottom-left, the Square[2] tool by UMG for data quality control. On the bottom-right, the Rocket platform for cardiovascular data analysis.*

Developing euCanSHare's ELSI framework: The project will begin in WP1 by building euCanSHare's ethical and legal framework for increasing responsible data sharing and Open Science within euCanSHare, which will also guide the implementation of the platform and the integration of the data. While some of these issues arise from the international nature of the project, and more precisely from the coexistence of different regulatory frameworks, others stem from the unique nature of each population cohort. Our ELSI team led by MCGILL and KUL will address these retrospective and prospective issues to provide a comprehensive overview of potential ethical and legal hurdles, and will work in close collaboration with the Steering Committee to develop a set of minimal common consent, data transactions and access requirements.

Furthermore, the GDPR regulation coming into force in May 2018 provides the European states with a comprehensive framework and guidance for implementing IT systems that deal with personal data. However, the Canadian legal landscape is quite different, since data privacy protections are mainly governed by provincial laws and policies. Thus, one of the main tasks in WP1 will consist of performing a critical analysis of the European (at the Union and country levels) and Canadian (federal and provincial) laws, regulations and policies, and to identify both relevant gaps and commonalities. Gaps will be bridged through the work and expert opinion from the Center of Genomics and Policy (CGP, MCGILL), which hosts the Public Population Project in Genomics and Society (P3G), renowned for its international policy expertise and cohort harmonisation. A set of minimal common data transactions and access requirements will then be defined and ultimately implemented in Smart Contracts, where possible, to automate or semi automate transactions across institutions on the two sides of the Atlantic. Guidelines and international data sharing policies governing the platform will be made available on the portal to make all users aware of the constraints and provisions under which they must operate.

As summarised in Table 1, more than 35 cohorts are expected to be initially integrated in the platform. However, each cohort is governed by a different set of policies and procedures, and access and use of cohorts' data is constrained by the unique terms of the consent form signed by each research participant. To facilitate data sharing, the CGP at MCGILL, in collaboration with P3G, will perform a comparative analysis of existing informed consent forms and information brochures to assess their interoperability, and identify a list of consent items essential for data use within *euCanSHare* (and items that may impede use, such as data storage limitations or international sharing limitations). This analysis will inform whether these datasets can be shared in *euCanSHare*, or if additional steps (*e.g.* formal access approval, re-contact/re-consent, etc.) are required.

Studying incentives for enhancing Open Science in euCanSHare: In addition to the implementation of the platform's legal framework, our multi-disciplinary experts will study innovative approaches for enhancing Open Science, including through emerging solutions such as blockchain technologies and Smart Contracts. The potential of the blockchain technology in healthcare has been recognised by several observers, scholars and institutions around the world including the European Commission[10] and parliament[11] and more recently the United States Congress in its 2018 Joint Economic Committee Report of the President, which indicates blockchain as a potential solution for coordination and portability of medical records[12]. In euCanSHare, we will study how the novel features of blockchain technology, namely decentralised control on access to the data, automated authentication and authorisation mechanisms using Smart Contracts, and heightened safeguards to protect the privacy of the data subjects, will influence the researchers' willingness to share data including biomedical data. The feasibility and adequacy of applying such mechanisms in the context of Open Science in biomedical research will be studied and policy points developed. An expert group coordinated by ESC will be set up to identify challenges and provide recommendations on how to overcome barriers and successfully apply blockchain approaches. We will also analyse the potential of new academic credits for data sharing, such as "data authorship", resulting in new proposals for considerations by the scientific community and funding organisations. Other incentivising mechanisms in terms of the academic credits will be also analysed, such as through "authorship coin" which is one example that can be used to reward data sharing efforts of the data producers (see WP1).

In euCanSHare, we are in favourable position to study the potential of blockchain to increase trust and Open Science as our consortium member LYN is the Scientific Coordinator of the MyHealth-MyData project, one of the main EU projects implementing blockchain for biomedical research (see: www.myhealthmydata.eu). We will re-use the blockchain based authentication and identity management system implementing all the relevant GDPR provisions, as well as a set of data security and protection features (dynamic consent, anonymisation, encryption, etc.). A set of Smart Contracts will be implemented to automatically execute data access control, enforce

[10] https://www.neweurope.eu/article/blockchain-for-europe/

[11] http://www.europarl.europa.eu/RegData/etudes/IDAN/2017/581948/EPRS_IDA(2017)581948_EN.pdf

[12] The 2018 Joint Economic Report: Report of the Joint Economic Committee, Congress of the United States, on the 2018 Economic Report of the President, Union Calendar No. 453, p.198.

institutional policies, user preferences, preconditions and legal requirements. Data access requests made by researchers will be encoded in these Smart Contracts instead of being submitted manually with online forms and executed against each cohorts of interest, drastically reducing manual work to access data and increasing trust. Cohort owners' requirements, encoded as computable parameters, will be checked by the Smart Contract logic, vastly increasing the efficiency of the process. Once the Smart Contract (which "lives" in the blockchain) will be able to validate that all conditions for allowing the data exchange are matched, then a new transaction is triggered, enabling the actual data exchange, at the same time leaving track of the transaction on the blockchain itself.

Note that the majority of research and clinical centres remain bound to traditional data management, in which they serve as data owners and controllers, as trusted third parties. As such, the project will only aim at implementing a pilot study in a pragmatic way, targeting these solutions to the appropriate environment. Variable levels of automation will be implemented in a few selected institutions (*e.g.* QMUL from the EU: also part of MyHealth-MyData and MCGILL from Canada), who have demonstrated capability and interest in this type of innovation, while the remaining institutions will use the more traditional Data Access Committees (DACs system) as managed by the EGA. The project will thus constitute a pipeline of innovation through which a selected group of centres will progress upward in the automation scale, while providing initial evidence for the potential of blockchain for increasing automation, security and data sharing in biomedical research.

Integrating existing infrastructures, technologies and datasets into euCanSHare's centralised platform:
WP2 to WP4 will focus on the building the euCanSHare centralised web-portal, along with it main computational components, *i.e.* cohort browser (and its data catalogue), access manager, data manager and data analyser. The methodology will focus on adapting each component to the cardiovascular datasets in Table 1 and integrating the tools from Table 2 in an interoperable manner. This will be done by re-using the basis framework and IT solutions developed by BSC during the MuG project (ongoing and started in November 2015). In this project, BSC implemented the centralised platform MuGVRE (www.multiscalegenomics.eu/MuGVRE) for multi-scale genomic analysis, which is now hosted at ELIXIR-ES. The MuGVRE experience is a true success story as recently featured in the Digital Single Market website of the European Commission (see[13]). Specifically, thanks to the expertise of BSC/ELIXIR-ES with the availability of robust interoperability solutions, a prototype of the MuGVRE platform was already assembled within 18 months and the full platform was deployed online with two years (November 2017). We expect to replicate the same methology in euCanSHare. Concretely, interoperability of the tools will be maintained through a rich set of metadata allowing the system to associate tools and data in a transparent manner. Easy-to-use modules for authentication (based on KeyCloak) and secure data management (Oauth2 protocol for all encrypted data transfers) will be also integrated. The protocols to plug-in the tools (data browsers, visualisers, or analysis tools) on top of the main infrastructure will re-used from MuGVRE[14], Finally, execution scheduling will be based on a traditional queueing system to handle demand peaks in applications of fixed needs, and an elastic and multi-scale programming model (pyCOMPSs, controlled by the PMES scheduler) for complex workflows requiring distributed or multi-scale executions schemes. In euCanSHare, this will enable to derive tools execution to remote cloud infrastructures (through the OCCI protocol, www.occi-wg.org) and also to HPC environments within ELIXIR, EGA and euro-BioImaging.

euCanSHare's main catalogue: Due to the large number of existing studies across Europe and Canada, new approaches are needed to facilitate the discoverability and interoperability of study data[15], thus maximising the use of resources and multi-study research in cardiology. In practice, the seemingly simple task of locating pre-existing data available for research is in fact a significant challenge; information on studies and on the data they collect is often either unavailable or difficult to find. Direct contact with study staff is then necessary to enquire about data availability and suitability, which results in a time-intensive process for researchers and study managers alike. Large-scale collaborative cardiovascular projects such as BiomarCaRE or BigData@Heart still require considerable resource investments to render the information collected by each study available and to convert heterogeneous data into compatible formats[16]. In WP2 and WP3, euCanSHare will extend and integrate the Mica solution (www.obiba.org/pages/products/mica) developed by MUC for building the largest and most comprehensive easy-to-use multi-cohort catalogue ever put together in the cardiovascular domain. This will help data custodians and network coordinators such as MORGAM, BiomarCaRE and CAHHM to efficiently organise and disseminate

[13] https://ec.europa.eu/digital-single-market/en/news/multiscale-complex-genomics-virtually-unpacking-dna

[14] http://www.multiscalegenomics.eu/MuGVRE/integration-of-tools/

[15] Gregory, A.T., 2014, September. Enhancing Discoverability of Public Health and Epidemiology Research Data. In *EDDI14–6th Annual European DDI User Conference*.

[16] Roger VL , Boerwinkle E, Crapo JD et al. Strategic transformation of population studies: recommendations of the working group on epidemiology and population sciences from the National Heart, Lung, and Blood Advisory Council and Board of External Experts. Am J Epidemiol 2015;181:363–68.

information about their cardiovascular studies and networks without significant technical effort. While Mica already includes modules to add and edit descriptive information pertaining to questionnaires, lifestyle/environmental data and physical/cognitive measures, documentation of cardiac imaging and data ethics/access information are either only partially supported or not supported at all currently.

EMC and MUHC will define metadata fields for the project, as collected by the euCanSHare cohorts and commonly used in cardiovascular research. Selection of the fields will be informed by existing standards adapted to serve the specific needs of the project. A Working Group will be established and convened at consensus meetings to define standard metadata for imaging, omics, epidemiological, clinical, and bio-sample data. EMC (euro-BioImaging) will also develop and implement the models to support cardiac imaging metadata (imaging modalities, protocols, parameters and biomarkers), while MCGILL will focus on cataloguing cohort-specific access policies and consent requirements. Metadata fields will be populated with information obtained from participating cohorts (see Table 1 for the initial cohorts, while that new cohorts will be added through awareness campaigns). The final Mica-powered euCanSHare platform will include Maelstrom's powerful search engine for allowing investigators to quickly find the information, variables and data they need for implementing cardiovascular research projects.

euCanSHare's data management framework: WP3 will establish a Data Management Plan (DMP) that includes the relevant models, tools and infrastructures, with special insight into the (meta)data sharing and distribution within the network, as well as the long term repositories for external stakeholders. euCanSHare will address interoperability issues such as cross-referencing between repositories (ELIXIR/EGA and euro-BioImaging), and the technical solutions for enabling efficient access to data using the technologies developed by ELIXIR, EGA and euro-BioImaging. The compliance with FAIR and EOSC data principles will be analysed from day one and the necessary actions included in the plan. Access policies (in close collaboration with WP1) will be defined and stored in the Automatable Discovery and Access Matrix (ADA-M) introduced by the GA4GH.

Susbequently, the protocols to deposit new cohort raw data into the appropriate repositories will be defined, together with the methods to provide rich metadata to foster a quality re-use of raw data for newly coming research projects. EGA (CRG) and euro-BioImaging (EMC) will implement mechanisms for linking metadata across the different repositories, to assure the transversal consistency of the specification of studies, and to enable presenting them in an integrated way. This task will ensure that request of information is performed according to specific formats as means to check quality control before data deposition. This part of the project will realise the recently published joint-strategy of ELIXIR and euro-Bioimaging to link their infrastructures (imaging and biomolecular)[17]. Furthermore, bio-samples will be deposed according the guidelines and quality assurance defined by BBMRI. Note that EMC will adapt its research infrastructure for euCanSHare to allow for standardised large-scale data storage for cardiac imaging and advanced automated image analysis. The euCanSHare infrastructure will be built around the state-of-the-art open source XNAT software (www.xnat.org), which will facilitate secure storage and management of the cardiac imaging data. Not only raw (DICOM format) images will be stored, but also image-derived data like segmentations and meshes, and other associated quantitative data.

Integrating imaging with -omics, lifestyle and clinical measurements for enabling novel approaches in personalised cardiovascular medicine: For a long time, biobanks focused for a large part on the collection and management of bio-samples, and subsequent analysis of -omics data. But with the exponential growth of medical imaging, more and more efforts are being undertaken to incorporate image phenotypes into biobanks. For cardiovascular research, cardiac imaging -especially cardiovascular magnetic resonance (CMR)- represents a powerful tool to quantify both clinical and subclinical cardiovascular disease, and to investigate novel methods for risk assessment, early diagnosis, and treatment monitoring (*e.g.* by pharmaceutical companies). However, in most multi-cohort cardiovascular efforts in Europe, cardiac imaging was initially overlooked (*e.g.* MONICA, MORGAM, BiomarCaRE). In contrastthe Canadian Alliance CAHMM has since in its design defined cardiac imaging as one of the main pillars of the super-cohort, particularly for studying subclinical signs of cardiovascular disease and the impact of ectopic fat on CVD.

Building on the experience of our Canadian partners (MCM), the UK Biobank (QMUL, leaders of the cardiac imaging of the UK Biobank) and the Hamburg City Health Study (UKE), as well as recent advanced on imaging storage and management through the euro-BioImaging Infrastructure (EMC), euCanSHare will integrate omics, lifestyle and clinical cardiovascular data with cardiac imaging at all the levels of the platforms. The very first multi-cohort cardiac imaging catalogue will be created with detailed information on available images, modalities, sequences and scanning parameters, as well extracted imaging indices and biomarkers for deeper phenotyping of cardiovascular disease. State-of-the-art imaging storage technologies developed by euro-BioImaging will be

[17] https://www.elixir-europe.org/system/files/euro-bioimaging_elixir_image_data_strategy.pdf

adjusted for federated and secure management of imaging cohorts in euCanSHare. EMC's FASTR platform[18] for building image analysis pipeline will be adapted for supporting cardiac imaging data, and will manage the remote execution on high-throughput clusters and cloud environments, enabling heavy and large-scale imaged analyses on large datasets. Using container technologies, the ease of adapting, enriching, sharing and deploying of cardiac imaging pipelines will be enabled within euCanSHare, enabling image analysists and SMEs outside of the consortium to use these technologies in combination with their own. Furthermore, another innovation will consist of integrating the cardiac radiomics pipeline developed by UPF[19] as a new method for more advanced imaging phenotyping of cardiovascular disease (see Ambition Section).

Database harmonisation, quality control and integration in euCanSHare: euCanSHare will also exploit the unique experience of our consortium in data harmonisation and integration of cardiovascular cohorts. The project will leverage state-of-the-art technologies developed by the Maelstrom Research at MUHC (www.maelstrom-research.org), as well as the harmonisation models implemented during the MORGAM and BiomarCaRE projects (>40 cohorts harmonised). However, since it is impossible to perform a single harmonisation that will satisfy all future study requirements, we propose an original solution to facilitate future harmonisations by re-using previous harmonisation efforts in a more systematic manner. Specifically, we will store the harmonisation algorithms in a standardised electronic database such that any harmonisation effort can be easily searched and located in the database and re-used in new multi-cohort research studies, when relevant. With this approach, future harmonisations benefit from previous ones and new harmonisation rules/algorithms are stored to further populate euCanSHare's harmonisation database. This approach will reduce cost and time of future multi-cohort research studies, while providing transparency on harmonisation processes. In this case, the harmonised dataset does not need to be stored; only the harmonisation rules and algorithms are saved in a standardised easy-to-search format and any harmonised data is generated on euCanSHare's cloud by the software in real-time. In euCanSHare, the harmonisation database will be initially populated based on the BiomarCaRE and CAHHM harmonisation experiences, as well as based on use cases that will be investigated to test the platform.

For implementing the proposed iterative harmonisation solution, MUHC's software Opal[20] will be adapted to provide a centralised web-based harmonisation management system allowing study coordinators and data managers to securely store/export a variety of data types and harmonisation rules in different formats using a point-and-click interface. Opal includes functionalities to define variables targeted for harmonisation, develop and implement processing algorithms used to derive common-format data, and efficiently document data harmonisation decision-making. To facilitate algorithm development, Opal also includes a comprehensive JavaScript library of functions commonly used to create harmonised variables. Establishing a secure connection with an R client also allows the use of the R programming language to derive common format variables. Opal then is executed to harmonise, store and display these data under the selected standardised model (*e.g.* BiomarCaRE's or CAHHM's model). Additionally, in euCanSHare, automated data quality control will be enabled through the Square^2 tool recently developed by UMG, which will enable to check for the level of normalisation, ambiguity and overall quality of both new sources slated for integration and the overall set of sources already integrated in the system. The resulting integrated harmonisation system will provide a highly flexible and efficient semi-automated process to on-board and harmonise new databases within the infrastructure.

euCanSHare's multi-scale cardiovascular data analysis platform: WP4 will develop the capacity within euCanSHare to build computational workflows for multi-scale cardiovascular data analysis adapted to different research questions and supporting a wide range of capabilities. The users will be allowed, after having identified the relevant cohorts and obtained permissions to access the data, to process the data and generate results. They will access the data on the secure servers at ELIXIR/EGA and build computational workflows composed of a number of tools comprising data harmonisation between the selected cohorts, image analysis, bioinformatics for analysis multi-omics data, and machine learning/statistical methods to quantify associations and new biomarkers, and to build predictive models of cardiovascular disease. This module will be developed such as it is flexible, allowing for different types of harmonisations, image quantifications and bioinformatic analyses. The content of these tools will be selected thanks to the experience of the consortium member in each of their respective fields (MUHC: harmonisation; UPF, QMUL, EMC: image analysis; UKE, UMG: bioinformatics). The capabilities provided by this

[18] Achterberg, H.C., Koek, M. and Niessen, W.J., 2016. Fastr: A workflow engine for advanced data flows in medical image analysis. *Frontiers in ICT*, *3*, p.15.

[19] Cetin, I., Sanroma, G., Petersen, S.E., Napel, S., Camara, O., Ballester, M.A.G. and Lekadir, K., 2017, September. A Radiomics Approach to Computer-Aided Diagnosis with Cardiac Cine-MRI. In Statistical Atlases and Computational Models of the Heart (pp. 82-90). Springer.

[20] Doiron, D., Marcon, Y., Fortier, I., Burton, P. and Ferretti, V., 2017. Software Application Profile: Opal and Mica: open-source software solutions for epidemiological data management, harmonisation and dissemination. *International journal of epidemiology*.

module will constitute a great incentive for cohort owners to upload their data into the system as they will be able to use the state-of-the-art computational workflows for processing their multi-disciplinary data in a single platform.

For this purpose, UPF will make available the Rocket platform developed during past cardiac related EU projects such as NEUBIAS, CardioFunXion and VP2HF. In this latter project, Rocket was shown for its capability to manage various types of cardiac-related data, including patient demographics, medical signals and images, physiological simulations and statistical results, as well as to enable the connection to cloud-based computational pipelines (*e.g.* bioinformatics). UPF will also incorporate its cardiac radiomics pipeline for deeper phenotyping of cardiac function and myocardial texture in cardiovascular studies. Furthermore, a bioinformatic computational resource will be assembled for analysing the variety of -omics data available in cardiovascular research, specifically genomics, transcriptomics, metabolomics and proteomics data. UKE and UMG, experts in this domain, will survey, analyse, select and test from the large space of existing libraries those methods that are most relevant, including from open-source software such as PLINK2, IMPUTE4, HISAT2, as well as from bioinformatics communities BioPython, Bioconda and, Debian Med. The module will be also connected to the R project[21], which provides a language and an environment for statistical computing. The final bioinformatics methods will integrate computational methods as well as functionalities designed to import, analyse and visualise omics data.

All these methods, including for image analysis, bioinformatics or machine learning, will be integrated to a web-based user interface through the Galaxy framework (www.usegalaxy.org), popular among the genomics community. This will make these analyses more accessible and reproducible, while facilitating navigation and analysis for non-experts. Software will be included in a virtual machine (container), together with a simple wrapper providing the necessary interface to the data infrastructures. Analysis and data will be synchronised with the appropriate protocols (see WP2), such that users may navigate from the data portal to their personal workspace where data and tools will be available for further analysis. The web interface will be extended to enable other researchers beyond the lifetime of this project to add their data analysis methods using the Docker Container technology. For instance, in WP5, our SME member will integrate its PELE environment (Protein Energy Landscape Exploration: www.pele.bsc.es/pele.wt) as one of the use cases for testing the platform (see below).

euCanSHare's iterative testing and feedback loop: In WP5, the euCanSHare platform will be pilot-tested by our clinicals, technical and SME members, both in Europe and Canada, to enable the compiling of feedback and recommendations throughout the project. A number of use cases and test studies with increasing complexity will be executed to adjust the capabilities of the platform throughout the projects and test its capabilities for personalised medicine approaches. Firstly, the verification of the data sharing platform and the harmonisation success will be performed by analysing known correlations and associations of cardiovascular risk factors and biomarkers to phenotypes and clinical outcomes, more specifically those tested in the BiomarCaRE EU project by UKE. Concretely, the tests will be performed by combing BiomarCaRE data with Canadian cohorts from CAHHM and new studies such as the UK Biobank. In a first step, known risk factors will be validated such as lipid status[22] and smoking in combination clinical variables and available endpoints. In a second step, the predictive accuracy of biomarkers such as Troponin[23] and HDL/LDL-Cholesterol will be assessed with respect to the different geographical regions covered in the cohorts. With this task, it will be possible to verify that all parts of the platform are in place and fully functioning.
More advanced testing will follow based on two clinical use cases. Firstly, euCanSHare will exploit improved access, sharing and analytics capabilities to run a large multi-country investigation of diabetic cardiomyopathy. Five to six cohorts which have patients with diabetic cardiomyopathy, imaging, genetic and lifestyle/environmental data will be identified and selected through the Cohort Browser. Computational workflows will be built to analyse the obtained multi-country harmonised dataset, including image analysis tools to derive new deeper phenotypes of DM-CM patients. By nature of this Canadian and European collaboration, it will be possible to uniquely investigate the impact of ethnicity, protective and detrimental lifestyle, environment and context, genetic factors, in diabetic patients with risk of diabetic cardiomyopathy. Secondly, the researchers will evaluate the value of combining imaging and genomics data for personalised medicine approaches, specifically risk prediction of cardiovascular events. European and Canadian longitudinal cohorts with the relevant data, including cardiovascular events such as myocardial infarction and stroke, will be identified through the Cohort Browser. After harmonisation is performed, the image analysis and bioinformatics toolboxes will be used to identify the radiomic features and genes that are

[21] R Core Team. R: A Language and Environment for Statistical Computing. (R Foundation for Statistical Computing, 2017).

[22] Waldeyer, C., Makarova, N., Zeller, T., et al., 2017. Lipoprotein (a) and the risk of cardiovascular disease in the European population: results from the BiomarCaRE consortium. *European Heart Journal*, *38*(32), pp.2490-2498.

[23] Blankenberg, S. et al. 2016. Troponin I and cardiovascular risk prediction in the general population: the BiomarCaRE consortium. *European heart journal*, *37*(30), pp.2428-2437.

associated with cardiovascular risk. Machine learning techniques will be used to integrate these identified variables with known risk factors such as age, hypertension, smoking, etc. The newly obtained predictive model will be validated through a multi-region approach across Europe and Canada and compared against existing tools such as the Framingham Risk Score or QRISK.

To ensure the flexibility of the euCanSHare platform, two other tests will be performed to complement the more traditional cardiovascular research studies described above. The first one will be led by our public health expert THL, who will compare the risk estimates and population attributable fractions (*i.e.* the fraction of all cases of a particular disease in a population that is attributable to a specific risk factor) across countries and continents. Such an analysis would highlight the relative public health importance of the different risk factors in each population. In addition to scientific interest, the results are important for the planning of prevention and control of cardiovascular disease in Europe and Canada. Moreover, our SME partner NBD (Nostrum BioDiscovery) will act as the industrial first demonstrator of euCanSHare during the testing phase, leading a target discovery study, which will provide valuable feedback from a different yet highly important perspective (see WP5).

Through these different multi-form studies (knowledge discovery, risk assessment, public health and target discovery), it will be possible to assess the performance and usage of the data sharing and analytics platform throughout the project, and constantly improve it based on the feedback of end users. This will ensure that the platform continuous developments will be guided by requirements of clinical researchers, epidemiologists and industrial developers. Surveys will be compiled periodically by UKE, while the feedback and recommendations from the different users will communicated to the technical teams working of the platform development.

Methodology for addressing sex/gender issues. Over the past decade, scientists, healthcare providers, public health researchers, and policy makers have made substantial efforts to improve understanding of the sex/gender differences in cardiovascular disease and to recognise the importance of heart disease in women[24]. In this context, euCanSHare will be dedicated to promoting this trend at all stages of the project and platform development. For example, the cohort catalogue and browsers will explicitly present the distribution of men/women in the data when search results are provided to the users, thus ensuring potential bias is removed or corrected for. Furthermore, the data analysis tools will include statistical functionalities to stratify for men and women. For example, it will be possible to see how the introduced cardiac radiomics features differ between men/women and interpret accordingly the results for risk estimation or disease understanding. All the use cases developed as part of the validation campaigns will explicitly integrate sex differences in the analysis, interpretation and dissemination of the results.

1.4 Ambition

(a) Progress beyond the state of the art:

Centralised data platform for multi-cohort cardiovascular research

State-of-the-art: For a long time, cardiovascular population research was performed based on individual studies. One of the first of such studies is the Framingham Heart Study (FHS), which led to a significant body of knowledge on cardiovascular health and disease. However, it covered a population of a single town, Framingham, in Massachusetts. One of the first multi-geography initiatives in the cardiovascular domain is the WHO MONICA study (multinational MONItoring of trends and determinants in CArdiovascular disease), covering a total of 25 countries including 18 in Europe. THL was responsible of the Data Centre of MONICA (www.thl.fi/monica/index.html). MONICA laid the basis for future multi-cohort cardiovascular studies, as well as for international standardisation of the measurement of classic cardiovascular risk factors in the population. It was extended into the MORGAM project (MOnica Risk, Genetics, Archiving and Monograph), integrating cohorts defined by the MONICA risk factor surveys, as well as other similar cohorts which have followed up the individuals for cardiovascular disease and death. Furthermore, MORGAM integrated a wider selection of measurements from the baseline surveys than MONICA, with repeated measurements, genotypes and biomarkers[25]. Recently, it was further extended as part of the BiomarCaRE project, coordinated by UKE, for investigating biomarkers of cardiovascular risk. Currently, the harmonised BiomarCaRE database includes over 30 cohorts in 13

[24] Mosca, L., Barrett-Connor, E. and Wenger, N.K., 2011. Sex/gender differences in cardiovascular disease prevention: what a difference a decade makes. *Circulation*, *124*(19), pp.2145-2154.

[25] Blankenberg, S., Zeller, T., Saarela, O., Havulinna, A.S., Kee, F., Tunstall-Pedoe, H., Kuulasmaa, K., Yarnell, J., Schnabel, R.B., Wild, P.S. and Münzel, T.F., 2010. Contribution of 30 biomarkers to 10-year cardiovascular risk estimation in 2 population cohorts: the MONICA, risk, genetics, archiving, and monograph (MORGAM) biomarker project. *Circulation*, *121*(22), pp.2388-2397.

countries[26] (300,000 persons, with 55,000 deaths, 23,000 incident acute coronary events, 13,000 strokes, 6,500 new cases of heart failure, 7,300 atrial fibrillation cases, and 7,900 new cases of type 2 diabetes). In Canada, the Canadian Alliance for Healthy Hearts and Minds (CAHMM) was assembled by our consortium member MCM in 2015, integrating eight Canadian cohorts covering different geographical areas and ethnical groups of Canada[27].

Progress beyond: euCanSHare will address the lack of IT infrastructures and advanced solutions for enhancing the re-use of data and the sustainability beyond the end of the funded projects. These difficulties are well illustrated with an on-going IMI project entitled BigData@Heart (see letter of interest from the Scientific Coordinator in the Annex), which is investigating heart failure epidemiology across European Union countries. Based on the experience of our UKE and ESC, who are also part of BigData@Heart, a lot of time and effort have been spent on surveying existing data and measurements, organising access to the fragmented cohorts in the EU, and performing yet another round of time consuming and costly data harmonisation. euCanSHare emerges in this context as the natural next-step for consolidating data sharing and multi-cohort research in the domain of cardiovascular personalised medicine. By integrating a number of existing infrastructures and tools (ELIXIR, EGA, BBMRI, euro-BioImaing, Rocket, Maelstrom), BiomarCaRE and CAHMM, as well as emerging existing and large studies (MORGAM, BiomarCaRe, BigData@Heart, CAHHM, UK Biobank), euCanSHare will revolutionise the way multi-cohort studies are conducted in future research. By exploiting the expertise of all our consortium members, robust integration of all of these technologies will be performed, leading to a unique web-portal that will accelerate access to information, facilitate new data harmonisation, integrate highly heterogeneous data such as -omics and cardiac imaging, and increase trust through state-of-the-art data protection solutions and provide new incentives for new cohorts to join the infrastructure beyond the lifetime of the project, by also offering data analysis tools (image analysis, bioinformatics, machine learning).

Integrated cardiac imaging storage and analysis

State-of-the-art: It is now widely accepted that suitable exploitation of -omics data, in particular the human genome, will depend on the availability of phenotypes that can inform about disease manifestations in more individual and finer-grained ways[28]. In the context of personalised medicine, while genetics research is necessary for risk estimation and disease prevention, imaging allows early detection at subclinical stage and thus is important to enable early intervention[29]. For this reason, more and more cohorts are investigating the integration of cardiac imaging, in particular through multi-sequence CMR for detailed quantification of cardiac shape and function. For example, our consortium member QMUL is currently leading the scanning of 100,000 CMR cases in the UK Biobank[30]. Our Canadian consortium member MCM have integrated detailed cardiac imaging to the Canadian Alliance CAHHM for studying, amongst other things, the impact of ectopic fat on CVD[31]. However, integrating cardiac imaging on a larger multi-cohort scale requires suitable solutions for cataloguing, storage and federated analysis. Unlike for -omics, imaging catalogues are lacking and the existing ones mostly include information on the number of participants and available imaging scans (*e.g.* MRI, CT).

Progress beyond: In euCanSHare, euroBioImaging will build the first multi-cohort catalogue of cardiac imaging data, integrating detailed information on the population and imaging modalities, the imaging sequences, protocols and scanning parameters (*e.g.* resolution), and available imaging biomarkers and quantifications for allowing reproducible research in this area. This will bring the quality of the implemented image-data catalogues to the next level and highest standard for future image cataloguing. The raw imaging data will be stored in the euro-BioImaging infrastructure, but as an innovation will be linked to -omics and clinical data at ELIXIR/EGA (WP3). The management of imaging data at euro-BioImaging which currently works for a single imaging study, will be extended to operate in multi-user/multi-centre settings. Finally, the cardiac image analysis tools developed by UPF during past EU projects (euHeart, VP2HF, CardioFunXion) will be made open-source, and will be extended with

---

[26] Zeller, T., Hughes, M., Tuovinen, T., Schillert, A., Conrads-Frank, A., Den Ruijter, H., Schnabel, R.B., Kee, F., Salomaa, V., Siebert, U. and Thorand, B., 2014. BiomarCaRE: rationale and design of the European BiomarCaRE project including 300,000 participants from 13 European countries. *European journal of epidemiology*, *29*(10), pp.777-790.

[27] Anand, S.S., Tu, J.V., Awadalla, P., Black, S., Boileau, C., Busseuil, D., Desai, D., Després, J.P., de Souza, R.J., Dummer, T. and Jacquemont, S., 2016. Rationale, design, and methods for Canadian alliance for healthy hearts and minds cohort study (CAHHM)–a Pan Canadian cohort study. *BMC public health*, *16*(1), p.650.

[28] Delude CM. 2015. Deep phenotyping: The details of disease. *Nature*. 527(7576):S14–15

[29] Wang, T.J., 2011. Assessing the role of circulating, genetic, and imaging biomarkers in cardiovascular risk prediction. *Circulation*, *123*(5), pp.551-565.

[30] 11: Petersen SE et al. 2016. UK Biobank's cardiovascular magnetic resonance protocol. J Cardiovasc Magnetic Resonance. vol. 1;18:8..

[31] Anand, S.S. et al.., 2016. Rationale, design, and methods for Canadian alliance for healthy hearts and minds cohort study (CAHHM)–a Pan Canadian cohort study. *BMC public health*, *16*(1), p.650.

automated quantify control of the images and a recently developed cardiac radiomics pipeline[32] for novel cardiac imaging biomarker discovery.

Bioinformatics toolbox for cardiovascular personalised medicine research

State-of-the-art: Through genome-wide association studies (GWAS), an enormous acceleration in discoveries of novel genetics variants associated with cardiovascular phenotypes have been made[33]. However, there is still a large part of disease variability left unexplained and for most genes the exact disease mechanism still remains unknown. After a decade of GWAS, researchers are moving toward new areas to explain the remaining unknown part of cardiovascular diseases variability. Other omics layers such as the transcriptome[4], epigenome (including methylome), proteome, metabolome, exposome, etc) have potential to improve our understanding and clinical management of CVDs[34,35,36,37]. Taken individually, each of these omics layers can explain a part of the variability, however, current research focuses as well on methods to integrate the multiple omics data[38,39], to unravel the interactions between these different omics layers and to identify relevant biological pathways and network co-regulated molecular patterns.

Progress beyond: In euCanSHare, a bioinformatic computational resource will be assembled specifically for multi-omics cardiovascular research, based on the expertise of our consortium, as well as on existing tools and methods. For example, a new generation of software (*e.g.* PLINK2, IMPUTE4, HISAT2, etc) are currently used, while new software methods are in development[40]. Many bioinformatical communities have also emerged (BioPerl, BioJava, BioPython, Bioconda, Debian Med, etc) for developing and distributing bioinformatics software solutions. One of the most established is the R project[41], which provides a language and an environment for statistical computing including more than 13,000 packages to date. A subcommunity, the Bioconductor project[42], is focusing on software and methods development to analyse omics data, including 1477 packages designed to import, analyse and visualise omics data. For non-experts, navigating through this ocean of omics solutions to analyse their data is not trivial given the variety of omics that are relevant to cardiovascular research. The goal of euCanSHare is thus to analyse and select from this deep space of solutions and libraries those methods that are most relevant to multi-omics cardiovascular research. Subsequently, these methods will be integrated to a web interface within euCanSHare platform through the Galaxy framework and linked to the cloud resources of ELIXIR/EGA to enhance the use of bioinformatics tools in cardiovascular research[43]. The web interface will be designed as to enable other researchers beyond the lifetime of this project to add their bioinformatic methods using the Docker Container technology. This toolbox will increase interactions between different disciplines of cardiovascular research (*e.g.* genomics, proteomics, radiomics), leading to new hypotheses and knowledge.

Knowledge discovery: The diabetic cardiomyopathy use case

State-of-the-art: In recent years, diabetes mellitus (DM) has become an epidemic and now represents one of the most prevalent disorders. Cardiovascular comorbidities are the major cause of mortality and morbidity in diabetic patients, who are 2.5 times more likely to develop heart failure (HF)[44]. The term diabetic cardiomyopathy (DM-CM) has been used to describe myocardial disease in diabetes that is not explained by hypertension, coronary artery disease or other cardiac disease. It has hypothesised that structural, functional and metabolic changes that occur in the myocardium in diabetes[45], as well as other molecular factors[46], may contribute to the development of

[32] Cetin, I., Sanroma, G., Petersen, S.E., Napel, S., Camara, O., Ballester, M.A.G. and Lekadir, K., 2017, September. A Radiomics Approach to Computer-Aided Diagnosis with Cardiac Cine-MRI. In Statistical Atlases and Computational Models of the Heart (pp. 82-90). Springer.

[33] Nikpay M et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. Nat Genet. 2015 Oct;47(10):1121-1130.

[34] Pedrotty, D. M et al.. Transcriptomic Biomarkers of Cardiovascular Disease. Prog. Cardiovasc. Dis. 55, 64–69 (2012).

[35] Bras, A. L. Basic research: Epigenetic map of heart development and disease. *Nature Reviews Cardiology* (2018).

[36] Kullo, I.J. et al.. Early identification of cardiovascular risk using genomics and proteomics. *Nature Reviews Cardiology* 7, 309–317 (2010).

[37] Griffin, J. L. et al.. Metabolomics as a tool for cardiac research. *Nature Reviews Cardiology*. 8, 630–643 (2011).

[38] Antman, E. M. & Loscalzo, J. Precision medicine in cardiology. *Nature Reviews Cardiology*. 13, 591–602 (2016).

[39] Karczewski, K. J. & Snyder, M. P. Integrative omics for health and disease. *Nature Reviews Genet*ics (2018).

[40] de Oliveira Veras, A. A. et al. Chapter 12 - Computational Techniques in Data Integration and Big Data Handling in Omics. *in Omics Technologies and Bio-Engineering* 209–222 (Academic Press, 2018).

[41] R Core Team. R: A Language and Environment for Statistical Computing. (R Foundation for Statistical Computing, 2017).

[42] Huber *et al.* Orchestrating high-throughput genomic analysis with Bioconductor. *Nat. Methods* **12,** 115–121 (2015).

[43] Shachak, A., Shuval, K. & Fine, S. Barriers and enablers to the acceptance of bioinformatics tools: a qualitative study. *J. Medical Library Association*. 95, 454–458 (2007).

[44] Dianati-Maleki, N.et al. 2017. Diabetes Mellitus in Patients with Heart Failure: Bad for All, Worse for Some. *JACC: Heart Failure* 5(1).

[45] Boudina, S. et al. 2010. Diabetic cardiomyopathy, causes and effects. *Reviews in Endocrine and Metabolic Disorders*, *11*(1), pp.31-39.

diabetic cardiomyopathy in affected humans. One particularity of DM-CM is the long latent phase, during which the disease progresses but is completely asymptomatic – yet at subclinical stage correct diagnosis and early targeted treatment approaches may prevent downstream morbidity and mortality. Thus, to contain or reduce the health consequences and the immense economic impact of DM and HF, new clinical knowledge and quantification methods are needed for risk stratification of diabetic cardiomyopathy. Cardiac imaging biomarkers, in particular, are thought to have potential for early diagnosis of this comorbidity[47]. However, current research is limited on the topic and diabetic cardiomyopathy remains incompletely understood and difficult to diagnose at early stage, partly due to sample samples in existing cohorts and biobanks, thus calling for novel multi-cohort research approaches.

Progress beyond: By exploiting heterogeneous and large multi-country cohorts assembled, euCanSHare will more deeply characterise diabetic cardiomyopathy by integrating a wide range of parameters, including genetics, imaging, lifestyle and ethnicity. We will derive and validate the diagnostic accuracy of a comprehensive cardiac imaging signature based on detailed cardiac radiomics and investigate how this cardiac imaging signature is related to both non-modifiable (*e.g.* age, sex, ethnicity, genetic factors, characteristics of DM (such as duration, HbA1c, type and treatments) and modifiable co-variates (*e.g.* protective and detrimental lifestyle or environmental factors). Preliminary unpublished results from the UK Biobank by QMUL and UPF in 100 diabetics and 100 non-diabetics, both groups without heart failure, suggest that cardiac radiomics can identify specific changes that occur in the diabetic heart compared with non-diabetic hearts. euCanSHare will greatly increase the power of the diabetic cardiomyopathy research through combining cohorts with the consequent increase in sample size. By nature of this Canadian and European collaboration, euCanSHare also allows uniquely to investigate the impact of ethnicity, lifestyle, environment and genetics in the genesis, diagnosis and treatment potential of diabetic cardiomyopathy.

Ethical and legal interoperability

State-of-the-art: The General Data Protection Regulation (GDPR) becomes enforceable across Europe in May 2018, which will have a direct impact on data sharing, not only in Europe but also around the world. At the same time, in Canada, data privacy frameworks are divided across provincial jurisdictions, as well as between the public, private, and health sectors. Thus far, little research[48] has examined the GDPR's implications on cross-border data sharing for research purposes[49], or the GDPR's impact on the adequacy of current cross-border transfer schemes (data adequacy, safe harbours, or model clauses). While past bridging attempts have been produced by the European Parliament[50], the GDPR's impact on the data protection equivalency schemes between the EU and Canada is currently unknown.

Progress beyond: Building on previous work in the MyHealthMyData EU project (coordinated by consortium member LYN), which evaluated the implications of the new GDPR on data sharing in a clinical context, euCanSHare will constitute the first impact study of the GDPR on the inter-jurisdictional (Canada-EU) sharing of data for health research. The relevant commonalities, mismatches, and gaps between the European (GDPR) and Canadian (federal) laws, regulations, and policies will be identified. These issues will be analysed, mapped, and bridged through the work and expertise of the Center of Genomics and Policy (CGP, MCGILL), which hosts the Public Population Project in Genomics and Society (P3G), renowned for its international policy expertise and cohort harmonisation. A detailed comparative review of the current legal and ethical landscape in Europe and Canada and will identify data protection challenges for euCanSHare in the context of international data sharing. Through our legal research, we will develop a Points-to-Consider (PtC) document to guide all future collaborative Canada-EU research projects that navigate the GDRP requirements.

Incentives for data sharing and open science

State-of-the-art: Open Science aims to provide broad access to all aspects of research, including datasets (Open data), publications (Open access), software (Open source) and the primary record of research (Open notebook science)[51]. The importance of developing adequate rewarding mechanisms to increase transparency[52] and

---

[46] Bugger, H. and Abel, E.D., 2014. Molecular mechanisms of diabetic cardiomyopathy. *Diabetologia*, *57*(4), pp.660-671.

[47] Miki, T., Yuda, S., Kouzu, H. and Miura, T., 2013. Diabetic cardiomyopathy: pathophysiology and clinical features. *Heart failure reviews*, *18*(2), pp.149-166.

[48] J. Stoddart, B. Chan & Y. Joly, "The European Union's adequacy approach to privacy and international data sharing in health research" (2016) 44:1 *Journal of Law, Medicine & Ethics* 143-155.

[49] Article 29 Data Protection Working Party, Guidelines on Consent under Regulation 2016/679 (28 November 2017) 17/EN WP259.

[50] Directive 95/46/EC of the European Parliament and of the Council on the adequate protection of personal data provided by the Canadian Personal Information Protection and Electronic Documents Act (notified under document number C(2001) 4539); Article 29 Working Party, Opinion 7/2014 on the protection of personal data in Quebec, 4 June 2014.

[51] Chretien, JP *et al.* 2016. 'Make data sharing routine to prepare for public health emergencies', *PLoS Medicine*, 13: e1002109.

incentivise data producers to promote Open Science is underlined in several previous studies, reports and policy statements[53,54]. Nevertheless, empirical evidence has demonstrated that the concerns of researchers about receiving adequate credit and attributes for sharing data are not adequately addressed to date. The lack of incentives for engaging in Open Science could result in data withholding or latency in data sharing by researchers[55]. To mitigate these risks, new platforms (*e.g.* GitHub, DAT, etc) and novel approaches (*e.g.* "Data Authorship") have been suggested[56,57]. The adequacy of such mechanisms in incentivising researchers to engage in Open Science has not yet been studied. Emerging mechanisms have been proposed such as the blockchain technology, which can increase trust, etc. However, the technology has only been implemented in the clinical context for patient data by our consortium member LYN as part of the ongoing MyHealth-MyData EU project.

Progress beyond: In euCanSHare, we will map current Open Science initiatives and their incentives, and critically analyse their strengths and weaknesses. A large-scale survey will be performed to collect first-hand data from researchers regarding their experiences and opinions about data sharing. Subsequently, we will study how the novel features of blockchain technology namely: decentralised control on access to the data, automated authentication and authorisation mechanisms, using Smart Contract, and heightened safeguards to protect the privacy of the data subjects will influence the researchers' willingness to share data including biomedical data. The feasibility and adequacy of applying such mechanisms in the context of Open Science in biomedical research will be studied and policy points developed. In this context, the potential role of blockchain to facilitate enforcement of authorship agreements or reward data sharing, by incorporating explicit tracking of academic credits into Smart Contracts, will be investigated. Since the professional concerns of the researchers regarding receiving credits when sharing data is appeared to be a reason for data withholding, such policies embedded in the blockchain-based platforms for data sharing could be of high value[58]. Finally, an expert group coordinated by ESC will be set up to identify challenges and provide recommendations on how to overcome barriers and successfully apply emerging technology mechanisms such as blockchain.

(b) Innovation potential:

Cardiovascular biomarker discovery

Traditional risk factors including hypertension, diabetes, obesity, and smoking account for a large proportion of the risk of cardiovascular events globally. Most of these traditional risk factors are modifiable and their management is likely to reduce risk of CVD in both primary and secondary prevention settings. However, risk estimates based on these classical risk factors only partially explain CVD incidence in the general population and risk estimates vary across countries. Despite the availability of various global risk assessment scores like the Framingham Score, QRISK and the European Society of Cardiology SCORE, prediction of cardiovascular events is incomplete and a considerable number of patients at risk go unidentified on the basis of traditional risk factors alone. As an illustration, it has been reported that 62 % of patients with MI present with none or only one risk factor[59]. To improve risk estimation above and beyond traditional risk scores, and to improve therapy decision-making and guidance, novel (emerging) biomarkers are of considerable interest[60]. Thus far, a lot of efforts have been put into the identification and implementation of novel biomarkers in the cardiovascular field, not only for proteins but also emerging molecules such as genetic markers, coding and non-coding RNAs, metabolites and lipids[61], as well as more clinically related markers such as cardiac imaging markers[62]. For example, the strongest additional predictive power has been achieved by natriuretic peptides, troponin and C-reactive protein

[52] Bierer, Barbara E et al. 2017. "Data Authorship as an Incentive to Data Sharing." In.: Mass Medical Soc.

[53] Cancer Research UK, Economic and Social Research Council, Medical Research Council, and Wellcome Trust: Expert Advisory Group on Data Access. "Establishing incentives and changing cultures to support data sharing". May 2014.

[54] Fecher, B., Friesike, S. and Hebing, M., 2015. What drives academic data sharing?. *PloS one*, *10*(2), p.e0118053.

[55] Kaye J., C. et al. 2009. 'Data sharing in genomics—re-shaping scientific practice', *Nature Reviews Genetics*, 10: 331-35.

[56] Bierer, Barbara E, Mercè Crosas, and Heather H Pierce. 2017. "Data Authorship as an Incentive to Data Sharing." In.: Mass Medical Soc.

[57] Perkel, Jeffrey. 2016. 'Democratic databases: science on GitHub', Nature, 538: 127-28

[58] J. Soto, D. Houlding 2017. "Blockchains for Data Sharing in Clinical Research: Trust in a Trustless World". Intel Healthcare.

[59] Zeller, T. et al.., 2014. Rationale and design of the European BiomarCaRE project including 300,000 participants from 13 European countries. *European journal of epidemiology*, *29*(10), pp.777-790.

[60] van Holten, T.C. et al.. Circulating biomarkers for predicting cardiovascular disease risk; a systematic review and comprehensive overview of meta-analyses. *PloS one*, *8*(4), p.e62080.

[61] Thum, T. et al. 2015. Long noncoding RNAs and microRNAs in cardiovascular pathophysiology. *Circulation research*, vol. *116*(4).

[62] Wang, T.J., 2011. Assessing the role of circulating, genetic, and imaging biomarkers in cardiovascular risk prediction. *Circulation*, *123*(5), pp. 551-565.

(CRP) as part of the EU MORGAM project[63]. To clinically evaluate new biomarkers, their ability to discriminate patients with a certain condition and a comparison to current standard markers need to be tested in independent and large cohorts and finally, in clinical trials.

Innovation potential: The innovation potential of euCanSHare for biomarker discovery is enormous. Firstly, the platform will enhance like never before the discoverability of relevant datasets for quantifying and validating new cardiovascular biomarkers (including -omics, imaging and circulating). Obtaining information about availability of bio-samples will be particularly useful to perform further measurements of biomarkers. Furthermore, access to advanced and SME-driven data analysis technologies, including bioinformatics, radiomics and machine learning, as well as the ability to enrich the platform with new quantification methods, will offer unprecedented computational capability for cardiovascular biomarker discovery. This will be particularly beneficial for emerging biomarker SMEs or pharmaceutical companies for preparing their assays and monitoring drug response in their trials. Finally, the size and diversity of the data will constitute a major cost-effective resource that will increase the innovation potential related to cardiovascular research, while counting on higher statistical power and reproducibility across populations and geographical regions. An example of such a study will be performed in WP5 to integrate imaging and genetic biomarkers to risk assessment of cardiovascular events such as myocardial infarction.

Multi-repository data integration and management

In the age of personalised medicine, cardiovascular population studies are putting a lot of efforts into the integrative analysis of -omics and cardiac imaging data, particularly for studying subclinical cardiovascular disease[64,65]. However, from a data management point of view, in particular when dealing with multiple cohorts in euCanSHare, this introduces challenges due to the amount of data to manage (in the order of TBs of data) and the diversity of data for providing an integrated view when genomics and imaging data must be both considered. Current solution to these challenges implies the introduction of a key component in large-scale multi-study projects, the Data Coordination Centre (DCC)[66]. The mission of such entity is to define and maintain data representation standards within the project, assure that data components can be accessed in an integrated way, and control data distribution within the consortium, data providers, and consumers, and the appropriate dissemination of the results. The most common DCC design involves a centralised repository of metadata giving access to the different components of projects' data, as employed in the International Cancer Genome Consortium, The Cancer Genome Atlas[67] or the Blueprint Epigenome project[68]. All these projects, however, are in the genomics field and have minimal representation of phenotypic data[69].

Innovation potential: euCanSHare will be the first multi-study project to integrate cardiovascular -omics, imaging and clinical data into a centralised data infrastructure. This will address both the current data fragmentation and the lack of "FAIRness" in the field of cardiology. We will develop IT solutions that will link major repositories, including the European Genome-phenome Archive (EGA: managed by CRG), BBMRI (for BioSamples) and the Euro-BioImaging infrastructure (EMC), thus ensuring the long-term availability, the maintenance of the FAIR principles, and the appropriate handling of security and privacy policies. In addition to -omics, we will enhance the deposition of phenotypical data (normally kept in individual databases) to the EGA repository. The coordination of such repositories will be implemented through mutually linked metadata to allow for the integrated access to heterogenous data belonging to the same sample in a centralised metadata repository. The common knowledge of euCanSHare partners forms an excellent basis to pilot a unified protocol that will enable new and existing studies to submit data to several repositories, while maintaining the cohesion as single studies. This effort constitutes one of the first for such coordination, made possible thanks to the unique blend of IT solutions and infrastructures in our consortium. These technical innovations will benefit the field of cardiology, as well as other clinical domains in need of multi-repository cross-linkage, such as in cancer and brain research.

---

[63] Blankenberg, S. et al. 2010. Contribution of 30 biomarkers to 10-year cardiovascular risk estimation in 2 population cohorts: the MONICA, risk, genetics, archiving, and monograph (MORGAM) biomarker project. *Circulation*, *121*(22), pp.2388-2397.

[64] Tsao, C.W. and Vasan, R.S., 2015. Cohort Profile: The Framingham Heart Study (FHS): overview of milestones in cardiovascular epidemiology. *International Journal of Epidemiology*, *44*(6), pp.1800-1813.

[65] Vargas, J.D. et al.. Common genetic variants and subclinical atherosclerosis: The Multi-Ethnic Study of Atherosclerosis (MESA). *Atherosclerosis*, *245*, pp.230-236.

[66] The 1000 Genomes Project: data management and community access. *Nature Methods*. 9, 459–462 (2012)

[67] https://tcga-data.nci.nih.gov/docs/publications/tcga/

[68] http://blueprint-data.bsc.es

[69] Zhang, J. et al.., 2011. International Cancer Genome Consortium Data Portal—a one-stop shop for cancer genomics data. *Database*, *2011*.

Cardiac radiomics: A new -omics for deeper phenotyping of cardiovascular disease

Current image phenotyping: euCanSHare will integrate cardiac radiomics as new phenotypic approach for deeper characterisation of cardiovascular disease. Radiomics[70] is a systems-scale approach that extracts and analyses large pools of multi-type and multi-scale imaging features that describe a wide range of shape, size, boundary, intensity and textural characteristics of anatomical organs and tissues. It was developed in 2012 in oncology as a tool for estimating cancer diagnoses, treatment response[71] and drug research[72]. In fact, as of today (April 2018), more than 95% of the papers indexed by PubMed (370 papers out of 387) have applied radiomics to oncology. UPF and QMUL were the first to adapt the concept to cardiac imaging based on CMR[73], also leading the estimation of the cardiac radiomic data for the UK Biobank data (first 5,000 CMR scans), which are now available for scientists within the UK Biobank network. They have implemented a portable radiomics pipeline for cardiac quantification using the Docker technology combined with machine learning techniques. Preliminary results have shown that cardiac radiomics have great potential for diagnosing complex cardiovascular diseases such as cardiomyopathies, including left ventricular non-compaction; one of the most challenging congenital CVDs to identify.

Innovation potential: By integrating the cardiac radiomics pipeline into euCanSHare, this will lead to new opportunities for innovations at the research level (better understanding of CVD onset and development), at the SME/industry level (new imaging biomarkers and decision support tools for cardiac assessment, *e.g.* QUIBIM www.quibim.com: see letter of support) and at the clinical level (new approaches for early diagnosis of CVD at the subclinical stage). Given the innovation of radiomics carried out in Oncology (*e.g.* OncoRadiomics SME in Belgium: www.oncoradiomics.com), we expect great innovation potential in cardiology as well, for imaging biomarker discovery, decision support and clinical assays. Another related innovation potential is the power of combining radiomics with other -omics such as genomics. For example, the company SOPHiA GENETICS has recently announced at the the Precision Medicine World Conference 2018 that its genomic testing technology SOPHiA (already deployed in more than 400 hospitals for cancer stratification: www.sophiagenetics.com), has been extended and enhanced with radiomics capabilities[74]. Similar approaches will be promoted by euCanSHare's multi-omics Data Analyser in the cardiovascular domain.

Blockchain innovation for biomedical data sharing

Blockchain medical applications span from health data exchanges and identity management tools to drugs supply chain, from insurance to personalised medicine[75]. Given the specific features of this technology in establishing trust, accountability, traceability, and integrity of data[76], one of the key application envisioned (and being actively developed in the private sector) is the one of health data management and exchange. Blockchain provides direct solutions in health record management introducing a decentralised mechanism for controlling and accessing data (implying also that each involved healthcare organisation manages its own data), while also introducing a time-stamp to the data, as well as robust audit trails mechanism[77]. Compared to traditional databases biomedical/healthcare applications, blockchain provides other clear benefits, beyond the immutable audit trail, including more robust data provenance models (improving both ownership control and traceability of the origin of a specific data asset), increased robustness and availability (thanks to the high level of redundancy provided by the technology), and improved privacy and security (thanks to the usually adopted cryptographic algorithms)[78].

Innovation potential: euCanSHare will innovate based on the architecture implemented in MyHealth-MyData, which is a solution that makes the blockchain (and relevant Smart Contracts) acting as an access-control manager for the data stored at the various repositories or individual local storage. This solution has been developed to keep

---

[70] Lambin, P. et al. 2017. Radiomics: the bridge between medical imaging and personalised medicine. *Nature Reviews Clinical Oncology*, *14*(12), p.749.

[71] Pota, M., Scalco, E., Sanguineti, G., Farneti, A., Cattaneo, G.M., Rizzo, G. and Esposito, M., 2017. Early prediction of radiotherapy-induced parotid shrinkage and toxicity based on CT radiomics and fuzzy classification. *Artificial intelligence in medicine*, *81*, pp.41-53.

[72] Katsila, T., Matsoukas, M.T., Patrinos, G.P. and Kardamakis, D., 2017. Pharmacometabolomics informs quantitative radiomics for glioblastoma diagnostic innovation. *Omics: a journal of integrative biology*, *21*(8), pp.429-439.

[73] Cetin, I. et al. 2017, September. A Radiomics Approach to Computer-Aided Diagnosis with Cardiac Cine-MRI. In *Statistical Atlases and Computational Models of the Heart* (pp. 82-90). Springer, Cham.

[74] https://www.medicaldevice-network.com/news/sophia-genetics-combines-radiomics-genomics-fight-cancer/

[75] Frost & Sullivan Report: Blockchain Technology in Global Healthcare, 2017–2025 - Healthcare Industry Assesses Blockchain Potential to Optimize Healthcare Workflows and Improve Outcome-based Care Delivery Models, June 2017, p.34.

[76] Ribitzky, R. et al. (2018). Pragmatic, Interdisciplinary Perspectives on Blockchain and Distributed Ledger Technology: Paving the Future for Healthcare. Blockchain in Healthcare Today.

[77] Halamka, J. D., & Ekblaw, A. (2017). The potential for blockchain to transform electronic health records. Harvard Business Review, *3*.

[78] Kuo, T. T., Kim, H. E., & Ohno-Machado, L. (2017). Blockchain distributed ledger technologies for biomedical and health care applications. *Journal of the American Medical Informatics Association*, 24(6), 1211-1220.

the overall architecture fast and scalable (which is also the reason for having adopted the Hyperledger Fabric solution, which has demonstrated recently to have reached transactions throughput comparable with the ones of financial third parties as VISA or Mastercard[79]), while at the same time minimising the amount of data being shared over the blockchain, both for privacy and compliancy reasons. This makes blockchain particularly suitable for improving productivity and thus innovation in biomedical research, by automating data access procedures and, if extended to cover actual data transactions, providing researchers with the needed data at the right moment, allowing aggregation of longitudinal health information[80].

2. Impact

2.1 Expected impacts

(a) Impacts mentioned in the work programme:

Intensified sharing, reuse, collaboration and knowledge discovery in the health field, while ensuring legal safety on the use of the data

euCanSHare will promote responsible data sharing in the cardiovascular research field, guided by international principles and policy statements such as the "Framework for Responsible Data Sharing" developed by the Global Alliance for Genomics and Health (GA4GH), and the OECD Recommendation on Health Data Governance. Partners of euCanSHare have been active members of such policy developments in the past (MCGILL, KUL), and their experiences and expertise will assist in developing international ethical and legal governance models for a research data management and storage infrastructure in cardiology. Furthermore, Technical and operational challenges are significant, mostly related to institutional concerns over data privacy and security which, especially in Europe, are very rapidly escalating due to the upcoming GDPR regulations. No clear implementation strategy has been in fact yet identified to guarantee their enforcement in transnational data exchanges. euCanSHare will start by addressing legislative and ethical gaps at the very beginning of the project, analysing and mapping these issues through Center of Genomics and Policy (MCGILL). Cultural and ethical obstacles to data sharing will also be studied through direct interactions with data owners and researchers, which will result in a set of recommendations to incentivise cohort owners to share more data, distributed to all major medical data platforms in Europe and Canada.

We are aware that data sharing among numerous retrospective cohorts is an ethically and legally challenging process, since each cohort is governed by a different set of policies, procedures and consent elements (see previous by MCGILL[81,82,83]). In euCanSHare, the use of the Automatable Discovery and Access Matrix (ADA-M)[84] to capture data sharing requirements of all the cohorts involved in euCanSHare will enable the development of structured metadata profiles tailored to the particular set of regulatory conditions of each cohort, as well as efficient management of that information across the euCanSHare network. This will reduce the costs and burdens of data stewardship and decreases efficient and responsible access to data. Furthermore, the use of the ADA-M and associated Smart Contracts will allow a rapid assessment of the ELSI interoperability of euCanSHare cohorts, and rapid integration of various health and disease data in respect of each cohort's data use restrictions and requirements.

In euCanSHare, a set of technical requirements will be also analysed in privacy preserving technologies such as blockchain and Smart Contracts, distributed authentication and automated access control systems, which will directly lower barriers to data sharing on the technical level. The potential of the blockchain technology in healthcare has been recognised by several observers, scholars and institutions around the world including the European Commission[85] and parliament[86] and more recently the United States Congress in its 2018 Joint Economic

[79] Androulaki, E., Barger, A., Bortnikov, V., Cachin, C., Christidis, K., De Caro, A., ... & Muralidharan, S. (2018). Hyperledger Fabric: A Distributed Operating System for Permissioned Blockchains. arXiv preprint arXiv:1801.10228.
[80] Engelhardt, M. A. (2017). Hitching Healthcare to the Chain: An Introduction to Blockchain Technology in the Healthcare Sector. Technology Innovation Management Review, 7(10).
[81] A.M. Tassé et al., « Retrospective access to data: the ENGAGE consent experience", European J. Human Genetics. 2010 Jul;18(7):741-5.
[82] J. Kaye et al., "Access Governance for Biobanks: the Cases of the BioSHaRE-EU Cohorts", Biopreserv Biobank2016 14:3.
[83] A.M. Tassé et al., « Developing an Ethical and Legal Interoperability Assessment Process for Retrospective Studies » Biopreserv Biobank. 2016 Jun;14(3):249-55.
[84] J.P. Woolley et al., «Responsible Sharing of Biomedical Data and Biospecimens via the 'Automatable Discovery and Access Matrix' (ADA-M) » submitted to Genomics Medicine.
[85] https://www.neweurope.eu/article/blockchain-for-europe/
[86] http://www.europarl.europa.eu/RegData/etudes/IDAN/2017/581948/EPRS_IDA(2017)581948_EN.pdf

Committee Report of the President[87]. euCanSHare will therefore provide a trusted infrastructure for foreign researchers providing a semi-automated workflow for substantial parts, and in some selected cases all of the steps needed to request and obtain access to biomedical research data. Data owners, on the other hand, would see administrative costs lowered as the most trivial and time consuming tasks of preliminary reviewing data access requests forms for compliance, checking their completeness, requesting if needed more information, and finally approving, or not, the request, will be automated thanks to Smart Contracts running on the blockchain integrated in turn with the OpenID system to authenticate users. While some data owners will be willing and ready to automate large parts of their administrative processes, and in some cases all of them, it is expected that not all will be inclined to do so due to technical and organisational limitations. At the same time, as solid demonstrations of these technologies are now commonplace in many sectors, proving their value in biomedical research, even if partially implemented, will drive further the already intense movement to implement them at-scale, generating a network effect that will in turn accelerate data flows across institutions with direct, magnified gains in research productivity.

Integration of various health and disease data in data-intensive fields such as personalised medicine

In 2014, our consortium member the European Society of Cardiology (ESC) put together a roadmap for new research in cardiovascular personalised medicine[88], calling for integrative data-driven approaches that will link molecular, imaging, functional and clinical data. However, the present scenario in the field, with highly fragmented data holders, albeit being a large amount of data of diverse nature, precludes their use in a productive manner. euCanSHare will develop a transnational cardiovascular catalogue that will be unique in its breath, integrating through metadata diverse data types, from genomic and molecular markers to clinical and imaging data, in a centralised location openly accessible to cardiovascular researchers. While more than 30 cohorts from Canada and the EU will be connected initially through the patient catalogue and made semantically interoperable, the number of fully integrated cohorts will grow over time thanks to the scalability of the platform. The Opal tool will increase efficiency of data harmonisation by implementing semi-automated data mapping procedures[89], thus facilitating collaboration across geographically dispersed data and data harmonisers. As it is web-based, it will allow collaborators to easily deposit data onto euCanSHare's secure server and converted them into standardised formats stored in the harmonisation database. These efficiencies will keep growing as the number of data harmonisation efforts accrue in the electronic harmonisation database, steadily reducing the number of unmapped data.

euCanSHare's platform will integrate major multi-study initiatives in the EU and Canada as described in this proposal. Addressing their current limitations and exploiting the accumulated know-how and experience, this integration effort will offer an unprecedented for new research in cardiovascular personalised medicine at a large scale[90]. New hypotheses generated will be more easily validated in available datasets, while increasing samples size for efficient identification of new biomarkers even with low effect sizes. The inclusion of omics and imaging data will offer new possibilities for knowledge discovery on complex cardiovascular diseases. This applies particularly to subgroups in which current evidence is insufficient, such as for diabetic cardiomyopathy. In this regard, novel evidence for comorbidity-related CVD will be gathered (*e.g.* link between diabetes and CVD in WP5). Based on a very large number of available data sources, the most important comorbidities of cardiovascular disease, as well as their influence on cardiovascular outcome and on recurrent cardiovascular symptoms can be evaluated. This new evidence will be used for adjusting guidelines regarding an optimised management of comorbidities[91], which are recognised as one of the main challenges faced by today's medicine[92].

More generally, the potential for personalised medicine will be enabled by the data density in the different cohorts not only for baseline clinical data but also follow-up data, blood biomarkers, genetics, biomaterials and imaging. Each domain has currently a big-data approach allowing for instance genomics, proteomics and metabolomics integrative research. Furthermore, radiomics, the extraction of multiple features from medical images, is a recent new development which brings imaging on the same level as the other-omics technologies. Radiomics tools that extract features from existing imaging repositories will be integrated in the research infrastructure and shared

[87] The 2018 Joint Economic Report: Report of the Joint Economic Committee, Congress of the United States, on the 2018 Economic Report of the President, Union Calendar No. 453, p.198.

[88] ESC European Affairs Workshop on Personalised Medicine: "The continuum of personalised cardiovascular medicine: a position paper of the European Society of Cardiology." *European Heart Journal* 35.46 (2014): 3250-3257.

[89] Doiron, D. et al. Data harmonisation and federated analysis of population-based studies: The BioSHaRE project. Emerging themes in epidemiology. 10. 12. 10.1186/1742-7622-10-12. 2013

[90] Rumsfeld JS et al. Big data analytics to improve cardiovascular care: promise and challenges. Nature Rev. Cardiology 2016 vol.13(6).

[91] UK Department of Health. 2014. Comorbidities: A Framework of Principles for System-Wide Action.

[92] McPhail, S.M., 2016. Multimorbidity in chronic disease: impact on health care resources and costs. *Risk management and healthcare policy*, *9*, p.143.

among the consortium for the specific use cases and to the scientific community. Radiomics is already established as an area of research in oncology, but it's applications in cardiovascular studies are just now starting to appear, pioneered, among others, by two consortium members UPF and QMUL[93]. The potential of identifying clinically relevant correlations that can be diagnosed both at the imaging and biomolecular levels can not only be a substantial step forward in the understanding of cardiovascular diseases, but potentially will lead to new low-cost ways of identifying patients at risk in the early stages of the disease. This will in turn have substantial impact on anticipating diseases trajectories reducing their burden on patients and healthcare systems.

More efficient research through reduced duplication of experimentation

The central goal of euCanSHare is to make cardiovascular data from disparate sources and cohorts FAIR, *i.e.* Findable, Accessible, Interoperable, and Reusable. It is a significant issue that data collected in research studies through previous funding is not easily accessible for secondary use. In euCanSHare, the Canadian and European are bringing distinct yet complimentary solutions that will increase efficient research and reduce duplication like never before in the cardiovascular domain.

Our Canadian partner MUHC has developed the Maelstrom initiative, which has become a world reference in multi-cohort data management. As such, Maelstrom is regularly invited to contribute to multi-cohort initiatives, in Canada (*e.g.* CAHHM: Canadian Alliance of Healthy Hearts and Minds), but also in US-funded projects (*e.g.* IALSA: Integrative Analysis of Longitudinal Studies of Aging and Dementia) and European projects (BioSHaRE.eu: Biobank Standardisation and Harmonisation for Research Excellence in the European Union). In euCanSHare, MUHC is providing its well-established tools for data cataloguing and harmonisation, the so-called Mica and Open tools (see Figure 4). Mica will enable researchers to identify retrospective datasets that are comparable to theirs for study protocol, sample size, date, clinical and genetic targets. The system will require researchers to first submit few key parameters describing the study they intend to carry out, so that similar ones can be identified within the euCanSHare platform. Metadata accompanying such studies will help researchers to clearly understand the re-usability of the available data following the FAIR principles. This way, researchers will be strongly advised to consider existing data sets and accordingly design their own studies only in original and non-repetitive areas, thus clearly discouraging duplication of the experiments. As for Opal, it includes powerful functionalities to define variables targeted for harmonisation, develop, implement and store processing algorithms used to derive common-format data, and efficiently document data harmonisation decision-making. With this tool, researchers can easily search through the harmonisation database and re-use previous efforts when relevant. Any new harmonisation algorithms are, on the other hand, stores such that they can be exploited in future multi-cohort studies. Over time, the harmonisation database becomes comprehensive and future harmonisations required much less time and effort, thus increasing efficiency and reducing duplications.

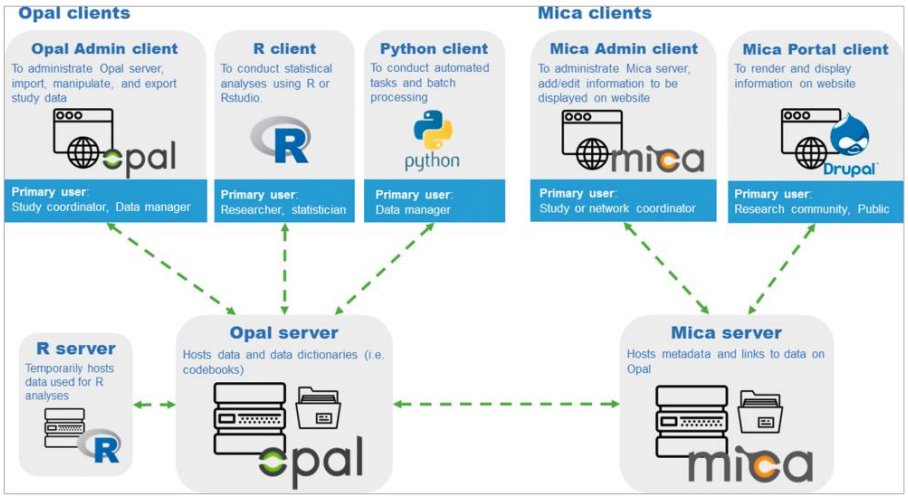


***Figure 4** – The Mica and Opal technologies will be adapted in euCanSHare to increase efficiency of research studies, by enabling easy and comprehensive access to information on previous studies/cohorts, as well as easy re-usability of previous harmonisation efforts in future multi-cohort studies.*

[93] Cetin I. et al. (2018) A Radiomics Approach to Computer-Aided Diagnosis with Cardiac Cine-MRI. In: Pop M. et al. (eds) Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges. STACOM 2017. Lecture Notes in Computer Science, vol 10663.

The European partners will provide a unique Data Analyser module, which will enable researchers to build flexible computational workflows for analysing heterogeneous cardiovascular data. It will integrate data analysis capabilities from different domains, including image analysis, bioinformatics and machine learning, by exploiting results from EU-funded projects such as BiomarCaRe, NEUBIAS, VP2HF or CardioFunXion in which the consortium members THL, UKE, UPF, QMUL and UMG have accumulated technologies and expertise. This module will positively impact the efficiency of future studies, by enabling researchers, particularly those lacking expertise in specific domains, to find multi-disciplinary tools in the same environment and integrate them into computational pipelines for executing novel research studies involving different scales of analysis (*e.g.* genomics and imaging).

Finally, it is worth mentioning the proof-of-concept study that will be implemented to test the feasibility of blockchain technologies to facilitate access requests and their managements, while ensuring the highest level of security. Clear evidence about increased efficiency thanks to blockchain technologies in healthcare is still being collected, but other public European sector have demonstrated drastic improvements in the speed and cost of diverse administrative processes. As an example, the Swedish real estate registry[94], after being automated with blockchain is registering transactions times that are orders of magnitude shorter than previous ones. If comparable gains are assumed in the biomedical research domain, an average data access procedure, including its request, review and approval, which today can take up to 3 months, and typically few weeks, could be shortened to just few hours and in some cases be completed in real time. This would offer researchers a new way to plan for studies and refine their designs, based on an immediate view of what data sources can be available. It would also lower the cost of these studies by allowing easier sampling of distributed data sets and thus early testing of hypothesis, sparing investments on unfruitful experimentations.

A network of research infrastructures and databases in the EU and Canada that build synergies between ongoing activities, contributing to delivering the backbone for new discoveries that address the Societal Challenges delineated in Horizon 2020

euCanShare will put a special effort in exploiting the capabilities of well-stablished public repositories, like EGA (CRG), Euro-BioImaging (EMC) and BBMRI (for bio-samples). A specific task in WP3 will take care of establishing secure data access protocols linking these infrastructures for the first time. This will contribute to set the necessary best practices in the community avoiding the data fragmentation, and enforcing the improvement of the standards in data safety. Furthermore, euCanSHare will constitute a transcontinental collaborative research infrastructure, by connecting on the technical, legal and operational levels some of the largest cardiovascular research networks, including the CAHHM, MORGAN and BiomarCaRE, in addition to other research infrastructures with broader clinical and molecular focus such as EGA, for a total of 30 cohorts and nearly one million cases across Europe and Canada. All these entities actively monitor, curate and expand their data sets as part of their ongoing institutional activities and serve to produce tens of research studies every year. To align these diverse entities, operational flexibility will be implemented within a robust framework that will make it possible for researchers to efficiently access highly distributed resources in the different repositories and for other institutions to keep expanding their cohort data. Our consortium partner BSC, also member of the EOSCpilot EU project (wwweoscpilot.eu), will ensure the multi-repository integration follows the principles of distributed and transparent data access as articulated in the European Open Science Cloud vision.

Thanks to these synergies, euCanSHare will directly respond to the Health, Demographic Change and Wellbeing challenge by increasing the knowledge base in the cardiovascular domain and reducing the burden of cardiac diseases. As mentioned in the Excellence Section, CVDs remain the most prevalent cause of death in the world and genetic factors are one of their leading causes. It has been shown, for instance, that 40% to 60% of coronary artery disease (CAD) risk is indeed due to genetic markers[95]. Modifying even slightly CAD's patients clinical trajectory, with early clinical interventions, will have tremendous socio-economic consequences, considering that is the most frequent cardiovascular condition. Precisely this will be the focus of one of the two internal studies that will work on the extension of current cardiovascular risk models with both genetic and imaging markers (cf. Figure 5). This study will aim at reducing cardiac morbidity by developing tools for identifying patients in preclinical CAD stages to be directed to personalised preventive care. In this population, in fact, standard lifestyle modifications may not deliver the reduction in risk that expected in other demographics and tailored interventions are frequently indicated.

---

[94] https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/using-blockchain-to-improve-data-management-in-the-public-sector

[95] Vinkuyzen AA, Wray NR, Yang J, Goddard ME, Visscher PM. Estimation and partition of heritability in human populations using whole-genome analysis methods. Annu Rev Genet 2013;47:75-95.

The second study will instead address diabetic cardiomyopathy; a condition with rapidly increasing prevalence[96], also with strong genetic makeup. euCanSHare will offer researchers an integrated data and analytics environment designed from the ground up to allow new discoveries in this medical domain, but also to design and test new treatments, in line with the Societal Challenges of H2020. Similarly, all insights into parameter evaluation and ideas regarding the design of future trials will be shared with the scientific community, thereby building the cornerstone for future avenues in cardiovascular research.

NOTE: This tab provides the PROSPECTIVE 10-year ASCVD risk estimate and the EXPECTED AVERAGE risk reduction associated with a preventive intervention		Enter patient values in this column	
Risk Factor	**Units**	**Value**	Acceptable range of values
Sex	M or F	M	M or F
Age	years	70	40-79
Race*	WH or AA	AA	WH or AA
Total Cholesterol	mg/dL	240	130-320
LDL-Cholesterol	mg/dL	170	
HDL-Cholesterol	mg/dL	40	20-100
Treatment with Statin	Y or N	N	Y or N
Systolic Blood Pressure	mm Hg	160	90-200
Treatment for Hypertension	Y or N	N	Y or N
History of Diabetes	Y or N	N	Y or N
Current Smoker (within last year)	Y or N	Y	Y or N
Aspirin Therapy	Y or N	N	Y or N

Figure 5 – *Illustration of current risk assessment tools that mostly traditional risk factors of CVD[110]. With euCanSHare, it will be possible to investigate the integration of genetics and imaging biomarkers into risk assessment for more precise and early identification of at-risk individuals.*

Strengthened position of the EU and Canada in science and more collaboration between academia and industry resulting in more innovation, jobs and growth

Through its consortium, euCanSHare will strengthen existing and yield novel European/Canadian alliances, leading to exchange of knowledge, expertise, data and technologies. Extensions of the already existing translational approaches of consortia like BiomarCaRE in the EU and CAHHM in Canada towards cross-continental multilevel and multidisciplinary approaches will be enabled. In particular, euCanSHare will transform the landscape of Canadian cardiovascular science by offering the potential to dramatically expand the cohorts and population data available within Canada, given that Europe's overall population is roughly twenty times that of Canada.

euCanSHare will contribute substantially to improving innovation capacity and integration of new knowledge of the research and SME partners within and outside of the consortium, *i.e.* the partners that will exploit the platform for generating new innovations of high commercial potential (*e.g.* UPF, LYN, NBD). euCanSHare partners can benefit from the open innovation model that the project will address, acquiring external knowledge from the participating research organisations, while they can increase their competitiveness through their participation in the project and exploit the innovative delivered products (*e.g.* new biomarkers, new blockchain technology). The universities and research organisations of euCanSHare both in Europe and Canada have the necessary potential (infrastructure and knowledge) to be regional drivers of innovations. They are leaders in their field and can develop innovative R&D activities on their own. However, their participation in the consortium will strengthen their innovation potential, since they have the opportunity to collaborate with other disciplines (*e.g.* linking genetics and imaging through UKE-UPF collaboration) in order to bring forward a multi-functional solution. euCanSHare members fully understand that innovation management is a process that requires an understanding of both market and technical problems, with a goal of successfully implementing appropriate creative ideas. Thus, the different aspects of innovation have been embedded in concept, workplan, and exploitation strategy in WP6.

One area where euCanSHare has great innovation potential is for drug and technology development in cardiology, focusing on personalised therapies. Considering the exponential increase in drug development costs that has been observed in the last 30 years[97] (see Figure 6), even marginal reductions in this area will have great overall effects.

[96] Guanghong J. et al. Diabetic Cardiomyopathy: An Update of Mechanisms Contributing to This Clinical Entity. Circulation Research. 2018;122:624-638

[97] DiMasi JA, Grabowski HG, Hansen RW. Innovation in the pharmaceutical industry: new estimates of R&D costs. J Health Economics. 2016;47:20-33.

Due to the cost multiplying effects of personalised therapies development pharmaceutical operational, costs are actually increasing even further. To simply provide a reference, an improvement of 1% in the efficiency of clinical validation functions brought by the euCanSHare platform in the development of a single personalised drug, would yield an expected economic impact in the tens of millions of Euros. The highly heterogeneous euCanSHare datasets that will include genetic, metabolic, proteomic, imaging and clinical records will be leveraged to pinpoint specific patient characteristics in highly-refined cohorts, that can contribute to drug effectiveness, or toxicity, and that should be taken into consideration in the study protocol design for new remedies. The infrastructure for image data storage and large-scale automated image analysis is also attractive to the pharmaceutical industry, as it provides an infrastructure in which imaging can be performed in a standardised fashion for the collection of imaging biomarkers as surrogate endpoints in clinical trials.

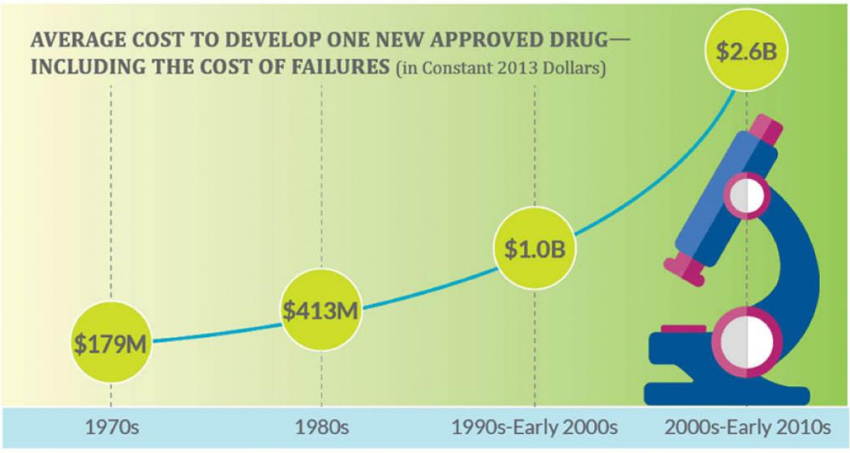


Figure 6 – *Average costs to develop one approved drug (statistics from[98]).*

Furthermore, the characteristics of euCanSHare will support articulated calibration of big data applications. New technologies, in fact, once trained/tested on a given dataset on which they have achieved sufficient accuracy, suffer from lack of extensibility/reproducibility to predict outcomes over populations not used during the development and unseen up to that point. Subsequent reduction in accuracy is frequently severe enough to discourage the utilisation of these systems on large populations. The euCanSHare's Data Analyser module will offer the possibility for external players to upload their tools using Docker containers and subsequently to further test and adjust their technologies based on more diverse and numerous datasets. euCanSHare will then act as an integrated big data calibration system where accuracy on different populations can be tested with minimal technical overhead. Also, both academic centres and SMEs will be able to leverage euCanSHare data for retrospective validation of clinical end points, exploratory statistics, and preliminary hypotheses testing, refinement of sample size and iterative refinement of prospective study designs. Especially within the emerging paradigm of Bayesian frameworks for clinical trials, the extensive data set in euCanSHare will provide rich set of priors on which clinical trial hypotheses can be preliminary tested, saving in patient recruitment cost, length of study and clinical assessments. The cardiovascular data catalogue will be instrumental in guiding researchers to relevant datasets, while providing analytical tools to run statistical analysis within the same environment.

Contribute to the Digital Single Market through piloting IT health research solutions

The ultimate goal of the Digital Single Market is to remove barriers when using online tools and services, thus to increase opportunities. In the biomedical domain, one of the main roadblocks to the utilisation of distributed data repositories is the complexity of accessing datasets under different jurisdictions, control procedures and data owners' institutional requirements. It easily takes a researcher several weeks to months to even know if a give data set will be accessible. A major technological innovation of this project is the integration of automated authentication systems (OpenID) with a blockchain infrastructure and correspondent Smart Contracts to partially automate data access requests and review processes. The system will be integrated with distributed cloud services developed by BSC (ELIXIR-ES) to actually transact data whenever that is allowed by the institution, in full alignment with the components of the European Cloud Initiative, which is the technical support of the concept of Digital Single Market. The same cloud services will in other cases allow to access local data repositories through APIs. While blockchain is gaining mainstream adoption in public and private sectors, its application to this use case is unique. If successful, it will set the foundation for a new and highly efficient way of authenticating, reviewing and allowing data transactions across European countries and Canada under an integrated and open technological framework. The successful implementation so far of this technology in the MyHealth-MyData project is an encouraging precedent in this direction.

The same technology will also be used to address one of the major cultural and behavioural obstacles to data sharing *i.e.* that of transparent and reliable attribution of data authorship among researchers. While this issue will be further investigated during direct interactions with the scientific community, it is already well-established that unrecognised use of data sources and lack of clear acknowledgements of data provenance are major sources of conflicts between scientists, curbing the volume of shared data. Blockchain will directly address this issue by creating a public, non-modifiable and persistent record of data access requests, responses and corresponding transactions, that will be available to the entire community in cases where proper attribution of data authorship has to be clarified and enforced. As a peer-to-peer network, the blockchain will not require central authorities to resolve conflicts. The entire scientific community will be able to observe, by simply querying the public ledger, what data transactions have happened and what institutions and individual researchers have carried them out. This approach has two benefits: on one side it does not require funding for enacting separate control functions, on the other, it leverages public opinion and reputation as natural drivers of compliant behaviour. While it will not be fully integrated to the final euCanSHare platform, pilot-testing the proposed blockchain solution will demonstrate a highly innovative solution with great potential for reinforcing the Digital Single Market.

Further the "Open science" and "Open to the world" priorities and contribute to the Health Research and Innovation Cloud, one of the thematic clouds of the European Open Science Cloud

euCanSHare is committed to addressing the cultural, behavioural and technological barriers to Open Sience. Competition in academic research, issues of trust, intellectual property, lack of robust mechanisms of recognition and reciprocity strongly come into play when data owners set their sharing policies and respond to data access requests. The project will directly investigate during face-to-face workshops with cohort owners the main obstacles, besides technology and data protection concerns, that hinder cross institution and cross-country data flows. These interviews with large numbers of data owners and critical analysis of their environments will allow to rank the effect of various factors and uncover new ones. Based on this analysis, a set of incentives will be designed together with institutions in a targeted and compliant fashion. The project will seek to implement an incentivising model that is self-sustainable, *i.e.* based on a set of mutually beneficial, formalised dynamics between data providers and researchers that lead by itself to more data exchanges. For instance, the method of contributing data in order to obtain data will be studied in detail and prototyped by the end of the project. In addition, the introduction of new types of academic credits building upon recent proposals[98,99] for data sharing, which could include a "data authorship" score and ranking system, will be investigated (see Figure 7). After thoroughly researching these concepts with cohort owners, resulting proposals will be formalised in peer-reviewed papers and communicated to the scientific community and funding organisations.

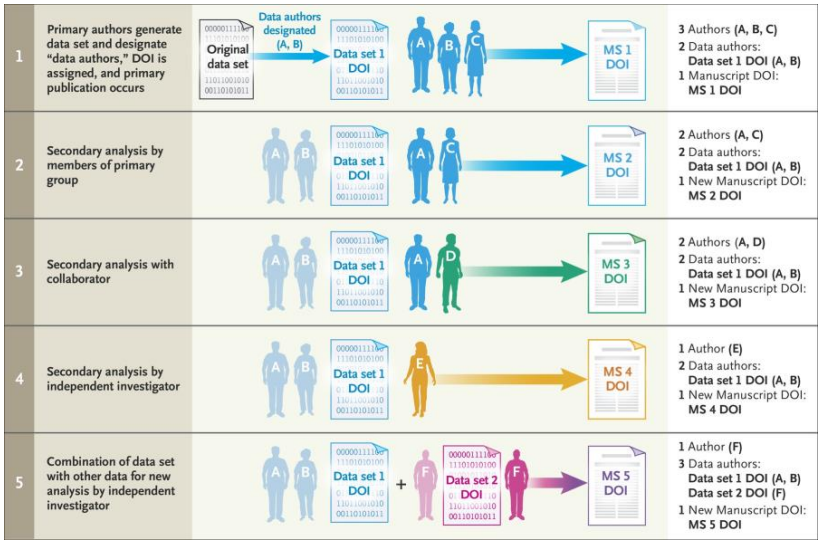


***Figure 7** – Schematic illustration of the blockchain potential for linking the management of data sharing and scientific authorship[99]. Here, an individual researcher (indicated by letters A through F) may be designated and credited as an author, a data author, or both, depending on the person's contribution to the data collection and/or analysis (DOI denotes digital object identifier, and MS manuscript).*

Furthermore, Open Science initiatives are rapidly becoming hampered by the exponential growth of distinct health data repositories that each has their own access methods. Researchers who need to make practical use of these resources will in practice be unable to use them given the impracticality of navigating countless access processes. As such, even if it assembles large and diverse quantities of health data for research, a Health Research and

[98] Barbara E. et al. Data Authorship as an Incentive to Data Sharing. N Engl J Med 376;17. April 27, 20171684

[99] Kiley R1. et al. Data Sharing from Clinical Trials - A Research Funder's Perspective. N Engl J Med. 2017 Nov 16;377(20)

Innovation Cloud will fail to meet one of the core characteristics of cloud computing, namely providing on-demand self-service. euCanSHare's implementation of data access procedures through Smart Contracts is a pioneering work that will overcome this obstacle by restoring the possibility for such on-demand self-service, without sacrificing or artificially flattening diverse local ethical/legal requirements that apply in different cultural and legal contexts. As such, the project will have reverberating impacts throughout scientific research fields, and around the world. euCanSHare will address the issues of incentives in Open Science in WP1, and develop high standards addressing the growing challenges and expectations from access to health data for clinical and research purposes.

The guiding principle of euCanSHare will be to identify and implement strategies that lead to network effects by which the more users participate in the platform, the more the platform itself becomes valuable and thus each participant gets a greater share of benefits. euCanSHare IT architecture is in fact a federation of cloud-based data repositories, orchestrated by standard cloud services that will connect researchers to data and data sources among themselves. The work carried out by BSC within ELIXIR will be instrumental in terms of standardising data access infrastructure not only within the platform but also with external resources to facilitate expansion and seamless interactions with the public. The recent communication regarding priorities and strategic directions of the European Open Science Cloud[100] will be taken into stringent consideration as the details of the architecture are designed and implemented.

(b) Other impacts not mentioned in the work programme:

Impact on healthcare practice and utilisation

For more than 30 years, most European healthcare systems have struggled with a consistent rise of the cost of healthcare – a trend which only recently was broken in some member states as a reflection of the economic crisis 2008/9 and its aftermath[101]. Being the highest burden, CVDs contribute a considerable part of these high healthcare and economic costs, in both Europe[102] and Canada[103]. The potential overall impact of big data applications in healthcare has been estimated in a McKinsy Global Institute report at €90 billion in savings when considering only the reduction of national healthcare expenditure in the EU[104]. As such, new data-driven knowledge, biomarkers and innovations that will lead to improved effectiveness and cost efficiency of prevention and treatment of the most prevalent CVDs will allow to realise important benefits for patients, healthcare systems and societies. Furthermore, high-risk, rising-risk, and low-risk cardiovascular patients require different care models and pathways. With the euCanSHare platform, the respective patient's CV risk score will be fine-tuned, by integrating rich data ranging from genetics, imaging to psycho-social factors, geographic and environmental specifics, behavioural aspects, and others to be identified.

Healthcare systems are the major holders of relevant data both at the phenotypical and the omics level. These data are normally hidden inside producers' databases for obvious privacy reasons. Leveraging highly specialised protocols for data access by the major data repositories, like the access to EGA data from trusted cloud infrastructures (and the establishment of the condition of "trusted" infrastructures), will open the possibility to share and re-use clinical data hold by the healthcare systems honouring the proper data safety and privacy regulations. This will enable to close further the gap between biomedical research and healthcare systems, through trusted intensive interactions. The wide variety of healthcare data collected and their integration will allow the development, validation, and successful application of more comprehensive, diverse and precise personalised medicine approaches than are presently available. These will be implemented into clinical workflows, providing personalised treatment models based on the optimal exploitation of heterogeneous (research and clinical) data available. The implementation of treatment models into clinical workflows for cardiovascular disease as one of the most frequent diseases will be a cornerstone for personalised medicine approaches and may influence treatment decisions in many other diseases. This will apply, for instance, to the management of comorbidities of heart disease, such as diabetes, ischemic stroke or peripheral artery disease.

Impact on drug development for personalised cardiovascular treatments

One of the most notorious success stories in drug development have taken place in the cardiovascular area. Arguably, the cholesterol-lowering agents statins are the single most successful drug class in history. It is worthy to

[100] https://ec.europa.eu/research/openscience/pdf/swd_2018_83_f1_staff_working_paper_en.pdf#view=fit&pagemode=none

[101] Organisation for Economic Co-operation and Development (OECD) Health Statistics 2017.

[102] The European Heart Network. European Cardiovascular Disease Statistics 2017.

[103] 2016 Report on the Health of Canadians. Heart & Stroke Foundation.

[104] When Applying assumptions from the McKinsey Global Institute report "Big Data: The next frontier for innovation, competition, and productivity", June 2011, to the European healthcare sector

mention that, amongst the 1,578 US FDA-approved drugs, 205 are for CVD conditions[105], that is ca. 13%. However, in spite of the arsenal of treatments currently available in the pharmacopeia, there is still high unmet medical need in the CDV area and many of the therapies are not efficacious for all patients. Approved drugs have high percentages of non-responders (a remarkable example is the statin class) and it is still not possible to accurately implement a precision medicine paradigm in the clinic.The development of novel therapies is even worsened by the fact that clinical trials in the CVD area are daunting both in terms of size and duration, becoming a huge undertaking even for the biggest pharmaceutical corporations[106]. The root of the problem is the lack of knowledge on the pathophysiological basis of many CVD diseases, which by definition are usually highly polygenic and extremely dependent on genomic and environmental variables. In this context, euCanSHare will contribute to the emerging paradigm of developing novel, targeted cardiovascular therapies by relying more and more on omics data, which will not only uncover the link between polymorphisms and disease but will also allow to atomise patient populations that will benefit from precision and personalised medicine development[107]. With access to large cohorts, it will become possible to investigate a more complete picture of genetic and environmental variables at play in CVD in order to predict drug response linked to genetic variation, understand clinical efficacy and safety, rationalise the differences between different drugs and predict patient subgroups with defined novel therapeutic options. Fortunately, genomics, transcriptomics, proteomics and metabolomics are rapidly evolving fields of science and are nowadays propelled by a combination of last-generation technologies. Only a combination of these fields, together with clinical data and the support of big IT infrastructures as enabled by euCanSHare, will enable to mine these large volumes of data to gain a deeper knowledge. The platform will additionally enable to uncover the link between so-far unknown polymorphic variants and the onset and progression of prevalent CVD states through radiomics. As illustrated by the use case that will be led by our SME Nortrum Biodiscovery (NBD) in WP5, euCanSHare will provide a new environment for preparing clinical essays *in-silico*, thus reducing the costs and enabling a more targeted approach to drug development.

Impact on citizens and patients

Most of the impacts mentioned in previous sections will mirror patient experience and imply benefits for them. A more advanced and patient-specific stratification model (as expectedly provided by one of the clinical studies) will lead to better prevention, more precise diagnosis, improved treatment and thus less rehabilitation and long-term care. Benefits will be considerable savings (*e.g.* co-payments for medications or while in a hospital), but also, and perhaps more relevant given the great burden of CVDs, in less pain and suffering, improved quality of life, lower disability due to CVDs, and longer lives. The WHO has estimated that about 80% of heart disease and stroke can be prevented[108]. In the USA, it has been estimated that a marginal improvement in risk estimation of cardiovascular events, achieved by integrating evidence from high-quality data, can lead to the prevention of heart attacks and strokes in the order of the million[109].

euCanSHare will also impact the way patients are evaluated by taking into account social, lifestyle, ageing and sex. For example, the sharing of data between Europe and Canada will elucidate the impact of contextual factors on geographic variation in CVD, such as level of access to healthcare services, lifestyle differences (*e.g.* country-specific dietary habits) and their relative impact compared to individual level factors. By translating such results into clinical practice, prevention and treatment will be increasingly tailored to individual groups and patients[110]. The achievement of a long-term research goal to personalise biomarker-guided prediction and prevention will impact the citizens in particular in light of the ever-ageing population both in Europe and Canada. Also, the large datasets will allow to examine gender related questions by stratified analyses according to differences between men and women. Over time, new approaches to the evaluation and treatment of acute coronary syndromes that are more prevalent in women, such as myocardial infarction associated with non-obstructive coronary arteries, spontaneous coronary artery dissection, and stress-induced cardiomyopathy, will be integrated to clinical practice[111].

---

[105] Santos, R. et al. 2017. A comprehensive map of molecular drug targets. *Nature Reviews Drug Discovery*, *16*(1), p.19.

[106] Butler, J. et al. 2015. Trends in characteristics of cardiovascular clinical trials 2001-2012. American heart journal, 170(2), pp.263-272.

[107] Friede, K.A et al. 2018. Future directions in pharmacogenomics discovery in cardiovascular disease. *Pharmcogenomics* 19(5):375-377.

[108] http://www.euro.who.int/en/health-topics/noncommunicable-diseases/cardiovascular-diseases/data-and-statistics

[109] Lloyd-Jones, D.M., et al. 2016. Estimating longitudinal risks and benefits from cardiovascular preventive therapies among medicare patients: The Million Hearts Longitudinal ASCVD risk assessment tool: a special report from the American Heart Association and American College of Cardiology. *Circulation*.

[110] Anand, S.S. et al. 2016. Rationale, design, and methods for Canadian alliance for healthy hearts and minds cohort study (CAHHM)–a Pan Canadian cohort study. *BMC public health*, *16*(1), p.650.

[111] Garcia, M., Mulvagh, S.L., Merz, C.N.B., Buring, J.E. and Manson, J.E., 2016. Cardiovascular disease in women: clinical perspectives. *Circulation research*, *118*(8), pp.1273-1293.

Finally, it is worth noting that ensuring public confidence on data protection and sharing policies is key for future biomedical research. A recent multi-country study[112] has shown that citizens support increased data sharing but their level of trust has decreased, thus calling for new innovations to better protect their health data. This trend has become even more pressing in light of the recent data scandals (*e.g.* Facebook). In this context, euCanSHare's approach of linking well-established infrastructures (*e.g.* ELIXIR/EGA) and reinforcing data protection through emerging solutions based on the GA4GH and GDPR regulations, as well as its commitment to improvement communication with the general public, go in the direction of improved trust and reinforcement of the citizen-research very important relationship.

Impact on public health and policy making

The euCanSHare platform will not only enhance data sharing, but also gather a new knowledge base for public health and policy making in Europe and Canada. For CVDs explored in euCanSHare, it will be possible to place their genesis in the context of the complex relationships of the combined effects of environment (region/country), lifestyle, genetics, and context (*e.g.* immigration) on the health of the respective populations. Such knowledge will improve our understanding of the root causes and probable development trends, population segments, or in the respective regions (Europe or Canada), and will impact the development of new public health actions. The experience of our public health expert THL in the MORGAM project[113,114] shows that public health research can be enhanced though multi-cohort approaches, such as for providing estimates of health risks, potential for disease prevention and the importance of different risk factors in the population or in different population subgroups. The possibility to combine cohorts from many countries has several public health applications: (i) to give a comparison point for individual countries, (ii) inform about the generalisability of the results, and (iii) provide a model for countries where such an evidence base is not yet used extensively. Impact can be further increased when existing datasets are combined with fresh representative survey data. This makes it possible to make projections for future health of the population and the need for health care resources for different scenarios of risk factor development. In this context, the scalability offered by euCanSHare is a significant asset.

The proposed platform will enable the member states of the EU, as well as provinces of Canada, to take advantage of the infrastructures and tools to establish competent initiatives and allow the positive impacts to be realised. As such, euCanSHare will allow to cover almost the complete cycle of the "Knowledge to Action Process" as recommended by the Canadian Institutes of Health Research Roadmap[115] in the "Knowledge Translation" Section. The heterogeneous data and analysis tools provided will enable to validate "old" knowledge, to adapt it to respective local/regional contexts across Europe and Canada, to compare present outcomes with those from other regions, thus resulting in evidence which is highly actionable. "What-If" scenario analyses to identify and explore potential public health policy options will, after their successful implementation, allow both policy makers and population health managers to follow a 'Learning health system' approach for a more targeted and personalised approach for allocating incoming patients to the appropriate care model/pathway. Given this potential impact, a Knowledge-to-Action dissemination is planned as part of the project (organised by ESC) to boost awareness and discuss these capabilities with public agencies across Europe and Canada (see letters of support). It will be discussed, for example, (1) how euCanSHare can support health authorities to substantially improve timely monitoring of the health of small and large populations alike, and the earlier detection and investigation of cardiovascular problems as they emerge; (2) undertake more complex research to predict the development of health threads and identify (new) risks affecting population cardiac health, and thereby enhance prevention, (3) develop and implement better evidence-based sound public health policies, strategies, measures and actions, (4) better analyse, simulate and predict the impact of promotional campaigns towards healthy behaviours in relation to CVD.

Impact on education: Training of future researchers and innovators

The euCanSHare platform and network will represent an exceptional platform for training new researchers, epidemiologists, developers and data managers. In fact, the cross-national intensive training of young researchers across disciplines with the latest methods and skills is one important component of the "Inclusive growth and smart

---

[112] http://www.nextgov.com/ideas/2018/04/citizens-support-increased-data-sharing-and-technology-innovation-enhance-security/147262

[113] Asplund, K., Karvanen, J., Giampaoli, S., Jousilahti, P., Niemelä, M., Broda, G., Cesana, G., Dallongeville, J., Ducimetriere, P., Evans, A. and Ferrières, J., 2009. Relative risks for stroke by age, sex, and population based on follow-up of 18 European populations in the MORGAM Project. *Stroke*, *40*(7), pp.2319-2326.

[114] Müezzinler, A., Mons, U., Gellert, C., Schöttker, B., Jansen, E., Kee, F., O'Doherty, M.G., Kuulasmaa, K., Freedman, N.D., Abnet, C.C. and Wolk, A., 2015. Smoking and All-cause Mortality in Older Adults. *American journal of preventive medicine*, *49*(5), pp.e53-e63.

[115] Canadian Institutes of Health Research's strategic plan for 2014/15-2018/19, Health Research Roadmap II: Capturing Innovation to Produce Better Health and Health Care for Canadians

growth" initiative of the EC. This cross-border network will enable interactions between young and established researchers across Europe and Canada, which also fits well to the Youth-On-The-Move[116] programs. Thus, it will lead to a new generation of researchers well-prepared to face the current and future challenges of data sharing and some of them are expected to become the new leaders in this field. These independent scientists will know to address research challenges using innovative Open Science approaches, including for the use of Open-Cloud based approaches. They will be at the forefront of emerging initiatives including new EU research projects on data-driven health research and will be instrumental in establishing and running new calls (*e.g.* on big data and personalised medicine). These researchers will be highly multi-disciplinary, trained to work with the whole spectrum of data sharing, from cataloguing, interfacing, ethical and legal issues, and data analysis. Even those researchers trained in one domain *(e.g.* bioinformatics), through the interactions within the network, will gain a multi-disciplinary understanding of the heterogeneous factors that need to be considered when investigating CVDs and health using data platforms. Through the different collaborations and outreach activities linking many European and Canadian stakeholders, they will be prepared to become excellent collaborators and generators of new ideas at the interface of academia, industry and clinical practice. With their unique profiles developed as described above, the young researchers will have a high value on the European employment in several sectors such in data management, data science, clinical research and even drug development, where they will fit the increasing demands of employers[117].

(c) Barriers and obstacles to these impacts:

euCanSHare's consortium is highly experienced in the field of data sharing, integrating the best experts from Canada and Europe (MCGILL, MUHC, MCM, ELIXIR/EGA, THL, UKE, QMUL). As such, while we are confident of the high potential and feasibility of the project, we are also aware of inherent obstacles that need to be addressed at different levels (technical, legal and behavioural). These barriers include:

Obstacle	Measure to reduce the barrier
Deep-seated cultural resistance hinders data sharing, regardless of technological solutions.	euCanSHare will produce guidelines for platform implementation to lower these barriers and to deploy incentives targeted at different types of stakeholders.
Concerns over privacy and data security reduce willingness to share data.	Promotion of established trans-governmental initiatives such as ELIXIR, EGA and BBMR. The new EU GDPR regulation is an opportunity for better governance of data sharing.
Data harmonisation costs remain high for researchers.	Promoting the re-use of previous harmonisation efforts, by developing an electronic database where harmonisation algorithms are stored and can be findable.
Divergence between EU and Canada policies and operational models.	EU and Canadian members will focus on the development of shared policies and analyse ethical interoperability from day one.
Cultural resistance to automation and to the deployment of new innovations such as Smart Contracts.	A Dissemination Plan adjusted to raising awareness and promoting the benefits and features of innovative solutions.
Interoperability of tools developed by different institutions.	Major efforts will be dedicated to addressing this issue based on the solutions already demonstrated by BSC in the MuGVRE platform.
The large volumes of available data.	We will fully leverage EU's cloud infrastructures such as ELXIR and EGA.

2.2 Measures to maximise impact

(a) Dissemination and exploitation of results

Dissemination: Given its intended impact, euCanSHare is expected to raise significant interest among academic researchers, private biotech and pharmaceutical companies, public health institutions and even policy makers. Dissemination and exploitation strategy has been therefore assigned to two members, the European Society of Cardiology (ESC) and LYN, who will be working in close collaboration to target different constituencies. ESC will indeed leverage its network of renowned universities, industries, research institutes and affiliated associations in

[116] http://europa.eu/youthonthemove/

[117] European Commission. Directorate-General for Research and Innovation, Socio-economic Sciences and Humanities. "New skills and jobs in Europe: Pathways towards full employment." (2012).

cardiology (140,000 contacts in Europe). LYN as an SME with long experience in data-driven innovation, eHealth and biomedical analytics, as well as of cardiovascular EU projects (Scientific Coordinator of CardioProof: www.cardioproof.com), will take advantage of its existing relations with stakeholders, including research centres and academia, industries, policy makers, patient associations and relevant public and private initiatives including HIMSS Europe and Big Data Value Association (BDVA). Also, as co-chair of the Research Data Alliance (RDA) Healthcare Interest Group[118] and Blockchain Applications in Health Working Group[119], LYN will be leveraging communication channels available within this organisation, such as conferences and internal publications for increasing awareness of the project within the data research community. From the Canadian side, MUHC, MCGILL and CAHHM will make available their extensive channels and collaboration networks (*e.g.* CAHHM, Maelstrom, CGP) to ensure appropriate dissemination to Canadian stakeholders.

To address dissemination and exploitation at best, a dedicated *"stakeholder analysis"* has been planned and will be put in place in the first months of the project to specifically identify relevant subject categories, individuals and entities, along with their different perspectives, needs and expectations, as well as their potential contribution to the project scopes, with the ultimate goal of developing targeted messages, determining most appropriate communication channels and languages, and enact relevant onboarding strategies. At preliminary level, these primary target groups have been identified as reported in the table below, and include (1) *the international community of cardiologists* from public and private healthcare institutions active in data sharing and research on cardiac disease; (2) *biomedical research public institutions and academia*, particularly in the molecular and personalised medicine fields (*e.g.* International Consortium for Personalised Medicine); (3) *biotech and pharma industry*, particularly those active in the development of innovative targeted drugs, as well bioengineering firms.

Table 3 – Preliminary overview of primary stakeholder groups, envisioned effects, relevant messages and communication/dissemination channels to be adopted in euCanSHare.

N	Stakeholder	Intended effects	Message	Targeted C&D methods
1	Cardiologists and medical associations in cardiology, healthcare and data providers	Provide new knowledge, disease phenotypes, real-world evidence and improved guidelines	The clinical implementation of euCanSHare will greatly enhance personalised medicine, CVD modelling and on a broader context cost effectiveness care.	- *Channels:* website, Twitter, Google+, LinkedIn, ResearchGate (2), training seminars and online webinars - *Publications*: Peer reviewed (journal articles, conference proceedings), generalists publications (newsletter, magazine and newspaper articles, press releases, white papers) - *Events:* scientific conferences in cardiology, public health and Big Data events - *C&D materials*: print-based (brochures, ID cards, posters), multimedia (photos, interviews, clips, demos)
2	Biomedical research community (research centres and academia)	Enlargement and diversification of cohorts, integration of new methods and tools, exploitation of platform for knowledge discovery	The clinical implementation of euCanSHare will facilitate data-sharing, improving security and increasing collaborations.	
3	Biotech, bioengineering and data science companies	Provide new resources for calibrating and testing new tools and biomarkers	Detailed mapping of given therapies and clinical outcomes for CVDs will enable adjustment to research and development activities	
	Biopharma industry	Provide new targets for drug development and enlarged cohorts for clinical trial studies		

Besides these main groups, the engagement of which will be of pivotal relevance to the achievement of project objectives, other audiences will be also taken into consideration, with broader scopes of communication of project rationale and mission, as well as general dissemination of project results. This secondary group, described in the table below, includes the (a) *cardiologic patients' community*, particularly the multiplicity of *patients' associations* active in promoting public awareness, prevention and advance diagnosis of cardiac pathologies, (b) *regulatory agencies and healthcare policy makers*, as well as other categories forming public opinion such as *mass media representatives and opinion leaders* (experts, authoritative scientists and authors, institutional representatives) active in the field and (c) *general public* at large.

[118] https://www.rd-alliance.org/groups/health-data.html
[119] https://www.rd-alliance.org/groups/blockchain-applications-health-wg

Table 4 – Preliminary overview of secondary stakeholder groups, envisioned effects, relevant messages and C&D channels to be adopted in the course of the project.

N	Stakeholder	Intended effects	Message	Targeted dissemination methods
A	Patients and patient associations	Raise awareness of euCanSHare, the use of patients' data and its potential impact to improve heart health	euCanSHare result can have a significant impact on outcome and quality of life of CVD patients	- *Channels:* website, Twitter, Google+, LinkedIn, Facebook - *Publications*: generalists publications such as newsletter, magazine and newspaper articles, press releases, white papers - *Events:* public health and Big Data events - *C&D materials*: print-based (brochures, ID cards, posters), multimedia (photos, interviews, clips)
B	Regulatory agencies and healthcare policy makers, mass media representatives, opinion leaders	Increase awareness of euCanSHare potential impact on citizens' quality of life, medical outcomes and public health	The societal impact of euCanSHare in terms of early diagnosis, disease management and relevant costs	
C	General public	Give account of the public funding spent on investment in research, highlighting improvement in data protection	euCanSHare results will improve health outcomes and quality of life	

The euCanSHare project is expected to produce a broad variety of research advancements at both preclinical and clinical levels, derived from the integration of different data sources (current CVD risk factors, genetic data, imaging and other diagnostics) and diversified cohorts in cardiology. Particularly, these studies will include *(1) cardiovascular disease risk assessment studies, (2) studies assessing the relation between diabetes and CVD onset, (3) identification of innovative disease biomarkers, (4) identification of cardiovascular targets and (5) multi-country public health studies*. Therefore, a significant dissemination effort will be put in place in concomitance with the first scientific results and carried on throughout the whole project duration, through the preparation of *peer reviewed publications* in the field and their spreading through *scientific dissemination events (conferences, workshops), website and social channels*, as well as *other publications directed to non-specialised audiences, including press releases, newspaper or magazine articles, and the project newsletter*. In this way, the dissemination strategy will also contribute to the achievement of key project objectives, to raise awareness of the euCanSHare platform. its functionalities and capabilities, increase its reach in data content and promote its use by researchers in the EU and Canada. While the ESC will focus on dissemination within clinical and scientific community affiliated with the society, LYN will focus on extending dissemination activities outside the cardiology sector, particularly in the industry and public health domains, including engagement of institutional subjects, agencies and policy makers, opinion leaders, mass media representatives, patient associations and general public at large, as well as members of industry and academic community in other research sectors.

Peer review articles will be carefully addressed to the most appropriate journals according to their clinical or IT relevance, discipline and impact factor. UPF will ensure the compliance with H2020 rules regarding Open Access to scientific publications, by making freely accessible any scientific publication that will be generated produce. The e-repository, which is the institutional repository of the UPF (www.repositori.upf.edu), meets all the requirements established by the European Commission within the framework of open-access publishing. The bibliographic metadata will be in a standard format determined by the European Commission and include all required information. We will thus not only ensure the acknowledgment of EU funding but also enhance the discoverability of all publications and maximize the impact of euCanSHare's results.

An illustrative list of selected target publications is reported below and will be further expanded throughout the project timeframe:

Publication Name	Publisher	Discipline	Open Access	IF[120]
Canadian Journal of Cardiology (CJC)	Elsevier	Cardiology	Supported ('gold')	4.403
Circulation	Lippincott Williams & Wilkins	Cardiology	Supported ('gold')	19.309
Circulation: Cardiovascular Imaging	Lippincott Williams & Wilkins	Cardiovascular imaging	Supported ('gold')	4.743
Circulation: Genomic and Precision Medicine	Lippincott Williams & Wilkins	Cardiovascular genetics	By default ('green')	4.743
European Heart Journal (EHJ)	Oxford University Press	Cardiology	Supported ('gold')	19.651
European Journal of Public Health	Oxford University Press	Public health	Supported ('gold')	2.431
European Journal of Preventive Cardiology	SAGE Publications	Cardiology	By default ('green')	3.606
Journal of Epidemiology and Community Health	BMJ Publishing Group	Epidemiology, public health	Supported ('gold')	3.608
Journal of the American Heart Association (JAHA)	Wiley Online Library	Cardiology	By default ('green')	4.863
Nature Genetics	Nature Publishing Group	Genetics	Supported ('gold')	27.959
PLOS One	PLOS	Science	By default ('green')	2.806
Scientific Reports	Nature Publishing Group	Natural sciences	By default ('green')	4.259
American Journal of Human Genetics (AJHG)	Cell Press	Genetics	Supported ('gold')	9.025
Journal of Computational Biology	Mary Ann Liebert, Inc. publishers	Bioinformatics	Supported ('gold')	1.032
OMICS: A Journal of Integrative Biology	Mary Ann Liebert, Inc. publishers	Molecular biology	Supported ('gold')	2.723
Journal of Biomedical Informatics	Elsevier	Bioinformatics	Supported ('gold')	2.753
Data Mining and Knowledge Discovery	Springer	Computer science	Supported ('gold')	3.160
International Journal of Epidemiology	Oxford University Press	Epidemiology	Supported ('gold')	7.738
Public Health Genomics	Karger Publishers	Genomics & public policy	Supported ('gold')	2.580
Journal of Grid Computing	Springer	Data sharing, infrastructures	Supported ('gold')	2.766
Biopreservation & Biobanking	International Soc. for Biological and Environmental Repositories	Biobanking	Supported ('gold')	1.698

Whenever possible, the full text of published articles or corresponding accepted manuscript will be also made available on the project website and spread out through concomitant updates on the relevant social media channels upon publication, particularly Twitter, Google+ and ResearchGate. Most importantly, all peer-reviewed publications will be also deposited in Zenodo, the open-access archive funded by EC, the OpenAIRE project and CERN, ensuring public availability of research materials including journal articles, conference proceedings, reports, deliverables and presentations. Thanks to the possibility of granting *Open, Embargoed, Restricted* or *Closed Access*, the platform will represent the chosen digital repository for any technical document pertinent to the project development, along with specific access settings and their evolution over time. Major publications will be also accompanied by production of dedicated *press releases*, which will be distributed to science news agencies and portals to promote media coverage. The *project newsletter*, published at six-month intervals and distributed via mailing list to volunteer subscribers, will be responsible for reporting project achievements to a broader, non-specialised audience, describing them in common language, making large use of images and infographics.

Dissemination of scientific publication and achievements of *euCanSHare* will be also conducted through non-specialised publications, including white papers, newspaper or magazine articles, and press releases to be produced at reaching of major project milestones, and distributed to science news agencies and portals (*EurekAlert!*, *AlphaGalileo, Science News, ScienceDaily, Live Science, New Scientist, Euro Scientist, Wired*) or health (*Digital Health, Fierce Health IT, HER Intelligence, Health Data Management, Health IT Analytics, The Medical Futurist*). Dissemination of *euCanSHare* achievements and updates will be also the focus of the project *Newsletter*, a web

[120] *InCite Journal Citation Report 2017*, Thomson Reuters

magazine produced on a bi-annual basis, to be made available on the website and distributed by e-mail to subscribers. When relevant, *multimedia-based* dissemination materials will be also produced, including *video interviews* to Consortium members, clips and *demos* to be shared through social media, and stored in dedicated archives (*Vimeo*).

As above mentioned, *dissemination events* will also be a major channel for presenting the scientific results and implemented solutions, accounting for academic meetings (conferences, workshops, seminars, symposia, with publication of proceedings and abstracts) organised by ESC and its affiliated organisations, and big public events dedicated to cardiovascular research and data sharing (*e.g.* ESC Annual Congress). To facilitate this, the consortium will design dedicated *presentation templates*, *posters, brochures* or *ID cards*, with project logo and relevant infographics, to be distributed to relevant audiences. Participation to dissemination events will be anticipated by dedicated project news on the project website and related posts on pertinent social media channels (*e.g.* Twitter), as well as spread out in real time with photos and tags. An initial list of annual or bi-annual candidate events pertinent to both categories, to be possibly attended by *euCanSHare* Consortium members, is shown below:

Event Name	Topic	Audience	Partner
Canadian Cardiovascular Congress	Cardiology	Health professionals and researchers cardiovascular-related disciplines	MCM
Big Data Week Conference	Data science, AI, architectures,	Data scientists, researchers, industry	LYN
eHealth Week	Healthcare IT, healthcare data	Institutions, governments, industry, associations, academia	LYN
EuroEcho-Imaging	Advancements in cardiovascular imaging	Cardiologists, paramedics and other healthcare professionals in cardiac imaging	QMUL
European Heart Rhythm Association (EHRA) Congress	The needs of the electrophysiology community	Clinicians, electrophysiologists	ESC
European Society of Cardiology Congress	Latest research findings in cardiovascular medicine	Healthcare professionals and clinical researchers in cardiology	ESC
Canadian Heart Rhythm Society Annual Meeting	Cardiology	Clinicians and other health professionals involved in cardiac arrhythmia	MCM
EuroCMR	Cardiovascular Magnetic Resonance (CMR) research	Cardiologists, radiologists, computer scientists, engineers, industry	QMUL
European Public Health Conference	Public health, health services research, health systems design	Healthcare service providers, governments, researchers, associations	THL
European Data Forum (EDF)	Big data, data economy, data-driven innovation	SME, industry, research, policymakers, community initiatives	BSC
European Big Data Value (BDVA) Summit	Big data technologies, big data economy	Industry, academia, public administration, data owners and users,	LYN
EuroPrevent	Prevention, epidemiology and population science, cardiac rehabilitation and exercise	Healthcare professionals (clinicians, paramedics) and clinical researchers in preventive cardiology	UKE
Frontiers in CardioVascular Biology	Basic cardiovascular science	Healthcare professional and biomedical researchers in cardiology	UKE
Geneva Health Forum (GHF)	Innovative practices in Global Health	NGOs, policy makers, industry, clinicians and academia	LYN
Health 2.0 Europe	Digital health	Industry, academia, policy makers	LYN
Heart Failure Congress	Cardiac pathologies	Clinicians, paramedic and other healthcare professionals in cardiology	ESC
HIMSS Europe	Healthcare IT	Industry, academia, clinicians, policy markers	LYN
Research Data Alliance (RDA) Plenary Meeting	Data sharing, data-driven research	Academia, industry and government	LYN
WIRED Health	Healthcare innovation, Artificial Intelligence in health	Industry, academia and clinicians	LYN

Exploitation initiatives: euCanSHare will implement an exploitation strategy aimed at maximising the future re-use of the platform, integration of new cohorts, commercialisation of specific technologies, and mainstream adoption of euCanSHare as a reference resource in cardiovascular research. These will be planned from M18 onward, relying on the engagement of relevant stakeholders to establish a standalone for-profit data resource for cardiovascular research in the public and private sectors, particularly for knowledge discovery, innovative drug development and clinical trial management. euCanSHare is expected to build bridges between healthcare institutions, research centres and SMEs/companies for the creation of a data market sustaining healthcare

innovation but also creating added value for the economy of Canada and Europe. The activity will have its inception with *exploitation seminars* to take place at annual meetings from M18 onward, where the strategy will be planned with the cooperation of all partners and possible involvement of external partners from industry.

Following the *stakeholder analysis*, which will provide insights on the different users' specific needs, expectations and concerns, the consortium will design and put in place specific campaigns for the onboarding of stakeholders onto the data sharing platform, particularly through *(1) gathering of additional patient cohorts* by direct engagement of healthcare institutions or medical associations, and *(2) inviting research centres, biotech and pharma industry* representatives to test the platform and its functionalities to become potential users. Regarding the former activity, the consortium has planned to *leverage existing cardiovascular multi-cohort consortiums* pertinent to existing research initiatives or medical associations, turning to them for set up possible collaborations and, in ultimate stage, to possibly add their cohorts to the system. Among them, is worthwhile mentioning *(1) Emerging Risk Factors Collaboration (ERFC)*, a consortium led by the University of Cambridge Cardiovascular Epidemiology Unit, that has collated and harmonised individual-participant data (IPD) of >100 prospective studies from >30 countries; *(2) Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE)* consortium, formed to facilitate genome-wide association study meta-analyses and replication opportunities among multiple large and well-phenotyped longitudinal cohort studies in the US; *(3) BigData@Heart*, a five-year project of the Innovative Medicines Initiative (IMI) launched in March 2017 developing a Big Data-driven translational research platform in cardiology, based on a EU public-private consortium consisting of patient networks, learned societies, SMEs, pharmaceutical companies and academia (see letter of support).

For linking to industries, the consortium will rely on the experience of its private sector partners, for instance *Nostrum Biodiscovery (NBD)*, an enterprise collaborating with pharmaceutical and biotech companies in Europe and US, active in development of drugs and molecules of biotechnological interest. By rationalising development process and costs, it will increase research reliability and ultimately maximise market success. Engagement of users will also be promoted through the realisation of *Online Webinars* as efficient tools to explain the rationale of implemented solutions and provide a tutorial of the platform functionalities and features to potential users, meanwhile providing further insights and constantly improving the platform. One of the three workshops organised in WP6 will be dedicated specifically to the industrial exploitation of the platform and many SMEs/companies will be invited to participate. In fact, several companies have already expressed interest (see letters of support in Annex).

The consortium will also timely deal with management of *Intellectual Property Right (IPR)* issues, by conceiving a strategy for a satisfactory protection of both background and forthcoming scientific results, with parallel patent application to the European Patent Office (EPO) and Canadian Intellectual Property Office, meanwhile maintaining knowledge sharing and cross-fertilisation with other parallel research initiatives, to be included into the first *exploitation plan*. Business and commercial exploitation routes will be explored by organising a Strategic Exploitation Seminar by M18. For example, UPF and its New Centre for New Medical Technologies (BCN-MedTech) has experience in transforming EU research results into commercialised technologies for personalising medicine approaches (*e.g.* the spin-off SME Galgo Medical: www.galgomedical.com). BSC has also translated in the past research results into our SME member NBD and euCanSHare will be exploited to consolidate the drug discovery activities. In Canada, MUHC's research led to the creation of Circle Cardiovascular Imaging Inc. (www.circlecvi.com), which has become a leader in cardiovascular functional analysis (see letter of support for the euCanSHare project). Building on these experiences and in close connection with academic and industrial stakeholders, the consortium will identify the most likely scenarios within which to position the project's expected exploitable outcomes. Individual analyses will examine the possible use of outcomes by partners according to their current business lines and isolated capabilities. A global exploitation plan will outline the full potential of the developed solutions in domains other than cardiology and will be used to depict different scenarios where commercial solutions could be developed. Joint plans will investigate the most likely market entry roads bringing together selected groups of partners (*e.g.* UPF and QMUL for the radiomics platform), who will combine their interest and generate strategic partnerships for common purposes. The joint plans will include target customers and markets, a timeline for reaching the market, and a diffusion strategy with regard to the resulting products/services.

Data Management: The euCanSHare consortium will pay particular attention to data management, which will be shared and processed during the project. As explained in "euCanSHare's data management framework" under Methodology section, as well as in WP3, a Data Management Plan (DMP) will be established including a comprehensive analysis of the nature of data to be handled, the requirements for interoperability, long-term storage and security issues, and the relevant models, tools and infrastructures, with special insight into the (meta)data sharing and distribution within the network. The partners BSC (ELIXIR) and CRG (EGA), with support from MCGILL as part of the OECD Working Group of the Health Data Governance, will build this highly compliant DMP, which will be further updated, informed (i) by current legal issues and associated practical requirements, as tackled in WP3, (ii) and by the refinement of the exploitation capacities along the project. Accordingly, the

euCanSHare consortium will seize the opportunity that the new General Data Protection Regulation (GDPR) is applicable from May 2018, to create a Data Governance Framework which will grant researchers access to quality datasets while offering citizens the highest level of trust on how they health data is used. All the infrastructures used during (and after) the project to host the data, specifically EGA, euro-BioImaging and BBMRI, stricly comply with the highest privacy and security standards.

(b) Communication activities

Communication activities will be directed to raising general awareness of the project rationale, mission, principles and goals among the different target audiences, by the differentiated use of languages and channels and the formulation of targeted messages and contents, creating the most appropriate ground for dissemination of results and ultimately to the engagement of stakeholders as part of the exploitation strategy. Key messages will thus be formulated and conveyed through the conception of effective storytelling, linked to the users' personal and/or professional experience, and highlighting the impact of project-delivered innovation in their everyday life.

To achieve these objectives, the following series of activities will be put in place from the project inception, carried out and updated along with the unrolling of project:

- *Stakeholder analysis and identification of relevant target audiences*, starting from the above-mentioned groups, and devoting to detailed research of people, enterprises, associations and public initiatives pertinent to the different groups of stakeholders, to be further consolidated and engaged;
- *Formulation of key messages, differentiated languages and contents* for different audiences starting from the ones previously identified, with special regards to (1) healthcare institutions *data providers*, on one hand, and (2) research and industry *end-users*, focusing on different needs, expectations and possible criticisms, not forgetting a more programmatic communication dedicated to policy makers, institutional and mass media representatives, as well as patients and their associations and general public at large;
- *Design of a coherent project branding*, to be used from the project inception throughout the project duration, creating a clear image to convey the project and make its messages and materials easy recognisable; this includes a logo, infographics, templates (presentations, posters, website, newsletter, etc.) and banner for acknowledgments of EU funding (including EU emblem, standard sentence, name of the project, the programme and grant agreement number) in various formats, to be included in all materials relevant to communication, dissemination, IPR and on equipment, infrastructure and major results;
- *Set up of different communication channels*, including project website, social network accounts (*Twitter, Facebook, Google+, LinkedIn, ResearchGate*), multimedia communities (*Vimeo*), and publication repository (*Zenodo*); project website and social media will be updated on a regular basis with project news and achievements, as well as relevant news in the field, according to specific features and audiences (see table below), further contributing to the growth of dedicated communities of stakeholders;
- Preparation of diversified *communication materials*: *ID-card, brochure, posters*; these will be prepared at project inception for general communication purposes, with focus on project overall vision, mission and goals, and a subsequent series of differentiated materials will contain project updates as well as tailored messages to the different audiences;
- Production of *multimedia*, including *photos* (public events, project meetings, etc.), *video interviews and programmatic clips* describing project mission, goals and achievements, and *demos* (platform, user interface, etc.) to be spread through the dedicated communities (*Vimeo*), website and social media, and possibly other media channels.

To maximise outreach, the Consortium will take advantage of *all potential cross-fertilisation opportunities within the Consortium* (*e.g.*, with regular contacts and active cooperation between LYN and ESC and other partners' communication or marketing officers), as well as with other relevant *associations, foundations and EU-funded initiatives*. These include, for instance, (1) associations like the *International Consortium for Personalised Medicine, the Canadian Cardiovascular Society*, the *British Cardiovascular Society*, the *American Heart Association (AHA)*, the *American College of Cardiology*, the *International Academy of Cardiology (IAC)* and *International Society for Heart Research (ISHR); (2)* EU-funded research projects and initiatives such as *BigData@Heart, CrowdHealth*, *MyHealth-MyData, KONFIDO, IASIS, epSOS, SoBigData; (3)* patients association and other public initiatives such as *the European Heart Network, Heart and Stroke Foundation of Canada, CareChain, Genomes.io, EUPHAnxt, European Patient Forum, Health First Europe, EUPATI, Health Unlocked;* and last but not least, the project will rely on the support and spread from *(4) EU official channels, such as EU_H2020, EU_Health, EU_ScienceHub, European Regional and Local Health Authorities.*

Given the increasing relevance of internet and social media in science & technology communication, the euCanSHare website will remain the reference point of the project communication strategy, constituting a showcase of project rationale, mission, goals and expected results as well as details of its project workplan,

consortium members' profiles and respective roles in the project. The website, besides other sources, will also contain all communication materials (ID-card, brochures, posters, presentations, graphics, photos, multimedia, etc.) produced throughout the project, publications and public deliverables. The website will also contain the coordinator's contacts (UPF) for information requests and setting up of new contacts and will allow the subscription to the project newsletter through a dedicated form. While progressing with the project workplan, the website aims to become a reference for a diversified and active community of stakeholders, by acting in strict coordination with social media channels (Twitter, Google+, Facebook, ResearchGate and LinkedIn), linked through a 'social bar' on the home page, that will contribute in the spreading of updates, including news, public events, major project achievements. A complete list of envisaged web and social media channels, communities and archives, with relevant features, contents and frequency of updates, is reported below:

Name	Type	Frequency	Content
Website	Web	Bi-weekly	Project rationale, mission and goals; project details (workplan, public deliverables, consortium and partners' role); project news and public events; communication materials (brochure, posters, presentations, multimedia); publications (newsletter, white papers, press releases, peer-reviewed publications); subscription form for newsletter mailing list (name, surname, e-mail, profession); link to social media channels, multimedia communities and publication archive.
Google+	Account	Daily	General messages (project mission and goals), project news (internal meetings, participation to institutional events and conferences, publications on media or peer-reviewed journals), video interviews and clips, cross-fertilisation with other relevant EU-funded initiatives, associations and foundations.
Twitter	Account	Daily	
Facebook	Page	Weekly	General messages (project mission and goals), project news (internal meetings, participation to institutional events and conferences, publications on media or peer-reviewed journals), interviews and clips.
LinkedIn	Project page	Bi-weekly	*Major* project news (participation to institutional events and conferences, publications on media or peer-reviewed journals), *occasional* cross-fertilisation with other relevant EU-funded initiatives, associations and foundations, *occasional* spreading of relevant science & tech news about eHealth, Healthcare IT, data science, Artificial Intelligence, Blockchain.
ResearchGate	Project page & research partners' profiles	Una tantum	(1) Project description, initial bibliography, periodic updates at reaching of major project milestones. (1,2) Publications pertinent to project development (directly downloadable in case of open-access publications, or available upon request to the author)
Vimeo	Account	Una tantum	Video interviews taken at public dissemination events, short clips describing project mission, goals and expected results.
Zenodo	Community	Una tantum	Published version or accepted manuscript of peer-reviewed journals, abstracts and conference proceedings (with *Close, Restricted, Embargoed* or *Open Access* according to publisher's copyright permission); deliverables (*Open Access* for public deliverables, *Closed/Embargoed* for confidential deliverables); presentations and technical documents.

Finally, as mentioned in the Impact Section, the euCanSHare platform has great potential as an educational/ medium for many key aspects such as data management, bioethics, cloud infrastructures, computational data analysis (*e.g.* bioinformatics, image analysis), data harmonisation and cardiovascular research. Consequently, thematic educational events will be organised locally (*e.g.* schools, education fair, university open days) as to introduce new students to these domains. The experience of UPF during the CardioFunXion project (showcase for the Rocket platform) has shown that this has a positive impact on young audiences, providing them insight into the future career directions in the biomedical field. Furthermore, a euCanSHare Summer School will be organised by UPF in Barcelona (at year 3), where all the members of the consortium will be invited to provide thematic courses on the different components of the project. The event will be organised in conjunction with the VPH (Virtual Physiological Human) Summer School that UPF organises every year to promote computational approaches in biomedical research (www.upf.edu/web/bcnvph_school). These training events will constitute an additional layer for communicating the euCanSHare achievements to key audiences within the general public.

3. Implementation

The euCanSHare project will be implemented in three main phases:

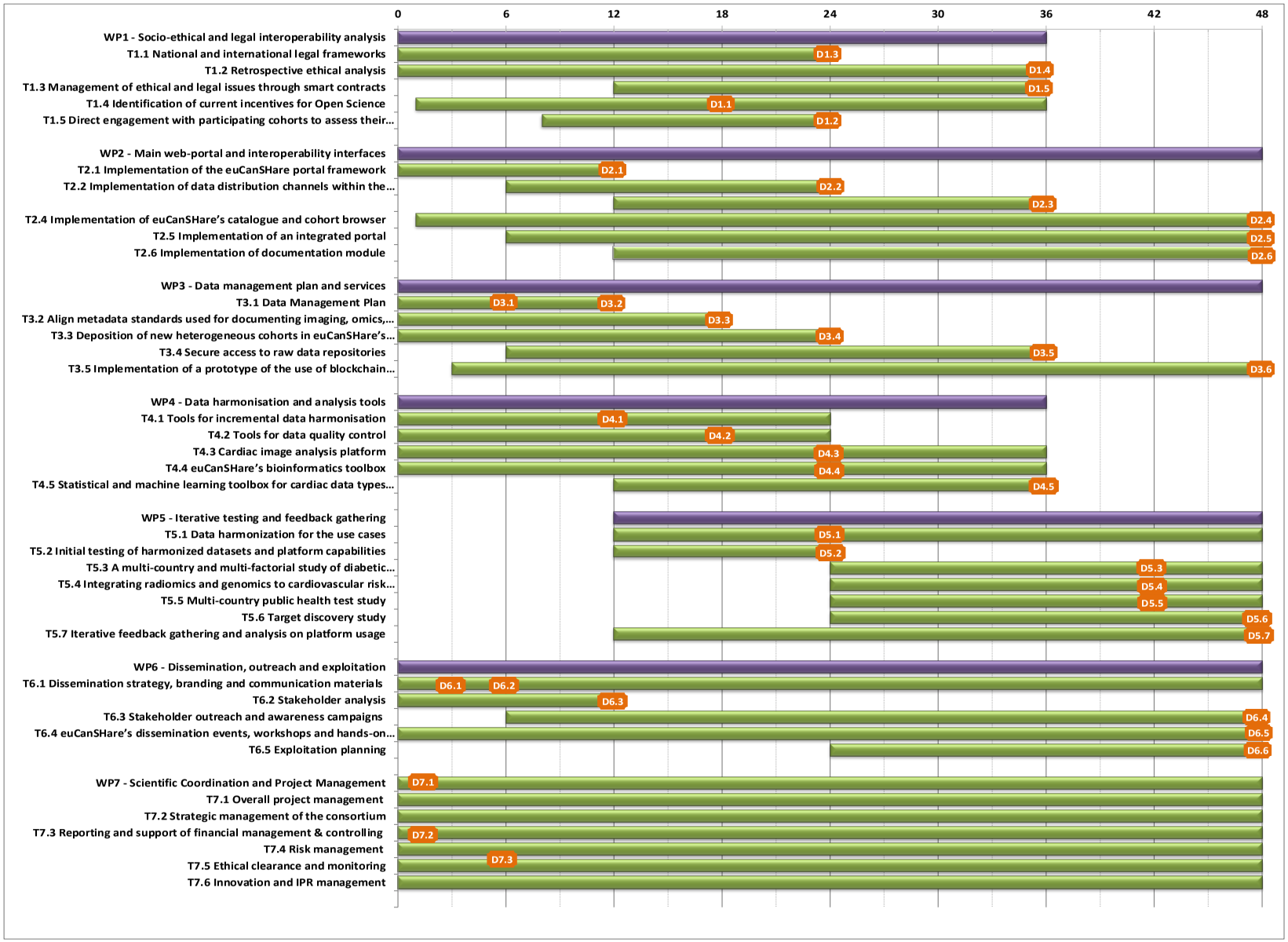
Phase 1 (M1-M12): It will begin by an extensive requirement gathering activity to (1) identify legal and ethical constraints and strategies for bridging codes and regulations across countries in a compliant way, and to (2) put together the overall design and specifications of the euCanSHare platform. Subsequently, the central data and technology environment will be established, covering those aspects not impacted by user level requirements. At the end of this phase, a prototype of euCanSHare's basis framework and data management services will be generated. Furthermore, stakeholders will also be directly engaged to identify a model of incentives that will activate more data exchanges increasing trust and demonstrating direct benefits to future participants.

Phase 2 (M12-M24): The second phase will be focusing on completing the platform integration, including the data system interfaces and help services, as well as the development of user-facing functionalities such as the cardiovascular catalogue, as well as data harmonisation and analysis services. Initial tests of data access technologies will begin and brought to conclusion after extensive feedback gathering from the initial set of users involved. At the end of this phase, the first release of the platform will be deployed within the consortium only.

Phase 3 (M24-M48): The third phase will focus on the execution of four pilot-test studies carried out by public and private participating institutions, together with an iterative refinement and adjustment of the platform capacities based on the compiled feedback. Note that two studies will focus on multi-domain indicators of cardiovascular disease (imaging, genetics, lifestyle, sex) and will aim at extending general understanding and risk prediction. The third study will be carried out by our public health experts THL, while the fourth will target industry-relevant endpoints and will provide highly targeted feedback on the commercial value of the platform. These studies will be slated over the course of the 24 months so that gaps will be identified and required extensions all implemented before the completion of the project. The dissemination and outreach activities will be intensified in this phase (including three workshops with hands-on sessions) to attract new cohorts and users from a range of stakeholders. At the end of Phase 3, the euCanSHare will be hosted at ELIXIR-ES, connected to the EGA, euro-BioImaging and BBMRI infrastructures, and fully deployed online for the cardiovascular research community at large.

Table 5 – List of euCanSHare's work-packages:

WP	Title	Lead No.	Lead Name	PMs	Start	End
WP1	Socio-ethical and legal interoperability analysis	7	MCGILL	65	1	36
WP2	Main web-portal and interoperability interfaces	9	BSC	184.5	1	48
WP3	Data management plan and services	5	EMC	115.5	1	48
WP4	Data harmonisation and analysis tools	1	UPF	149	1	36
WP5	Iterative testing and feedback gathering	3	UKE	179.5	12	48
WP6	Dissemination, outreach and exploitation	11	ESC	63	1	48
WP7	Scientific Coordination and Project Management	1	UPF	61	1	48
				817.5		


0
6
12
18
24
30
36
42
48
WP1 - Socio-ethical and legal interoperability analysis
T1.1 National and international legal frameworks
T1.2 Retrospective ethical analysis
T1.3 Management of ethical and legal issues through smart contracts
T1.4 Identification of current incentives for Open Science
T1.5 Direct engagement with participating cohorts to assess their...
WP2 - Main web-portal and interoperability interfaces
T2.1 Implementation of the euCanSHare portal framework
T2.2 Implementation of data distribution channels within the...
T2.4 Implementation of euCanSHare's catalogue and cohort browser
T2.5 Implementation of an integrated portal
T2.6 Implementation of documentation module
WP3 - Data management plan and services
T3.1 Data Management Plan
T3.2 Align metadata standards used for documenting imaging, omics,...
T3.3 Deposition of new heterogeneous cohorts in euCanSHare's...
T3.4 Secure access to raw data repositories
T3.5 Implementation of a prototype of the use of blockchain...
WP4 - Data harmonisation and analysis tools
T4.1 Tools for incremental data harmonisation
T4.2 Tools for data quality control
T4.3 Cardiac image analysis platform
T4.4 euCanSHare's bioinformatics toolbox
T4.5 Statistical and machine learning toolbox for cardiac data types...
WP5 - Iterative testing and feedback gathering
T5.1 Data harmonization for the use cases
T5.2 Initial testing of harmonized datasets and platform capabilities
T5.3 A multi-country and multi-factorial study of diabetic...
T5.4 Integrating radiomics and genomics to cardiovascular risk...
T5.5 Multi-country public health test study
T5.6 Target discovery study
T5.7 Iterative feedback gathering and analysis on platform usage
WP6 - Dissemination, outreach and exploitation
T6.1 Dissemination strategy, branding and communication materials
T6.2 Stakeholder analysis
T6.3 Stakeholder outreach and awareness campaigns
T6.4 euCanSHare's dissemination events, workshops and hands-on...
T6.5 Exploitation planning
WP7 - Scientific Coordination and Project Management
T7.1 Overall project management
T7.2 Strategic management of the consortium
T7.3 Reporting and support of financial management & controlling
T7.4 Risk management
T7.5 Ethical clearance and monitoring
T7.6 Innovation and IPR management
D1.3
D1.4
D1.5
D1.1
D1.2
D2.1
D2.2
D2.3
D2.4
D2.5
D2.6
D3.1
D3.2
D3.3
D3.4
D3.5
D3.6
D4.1
D4.2
D4.3
D4.4
D4.5
D5.1
D5.2
D5.3
D5.4
D5.5
D5.6
D5.7
D6.1
D6.2
D6.3
D6.4
D6.5
D6.6
D7.1
D7.2
D7.3

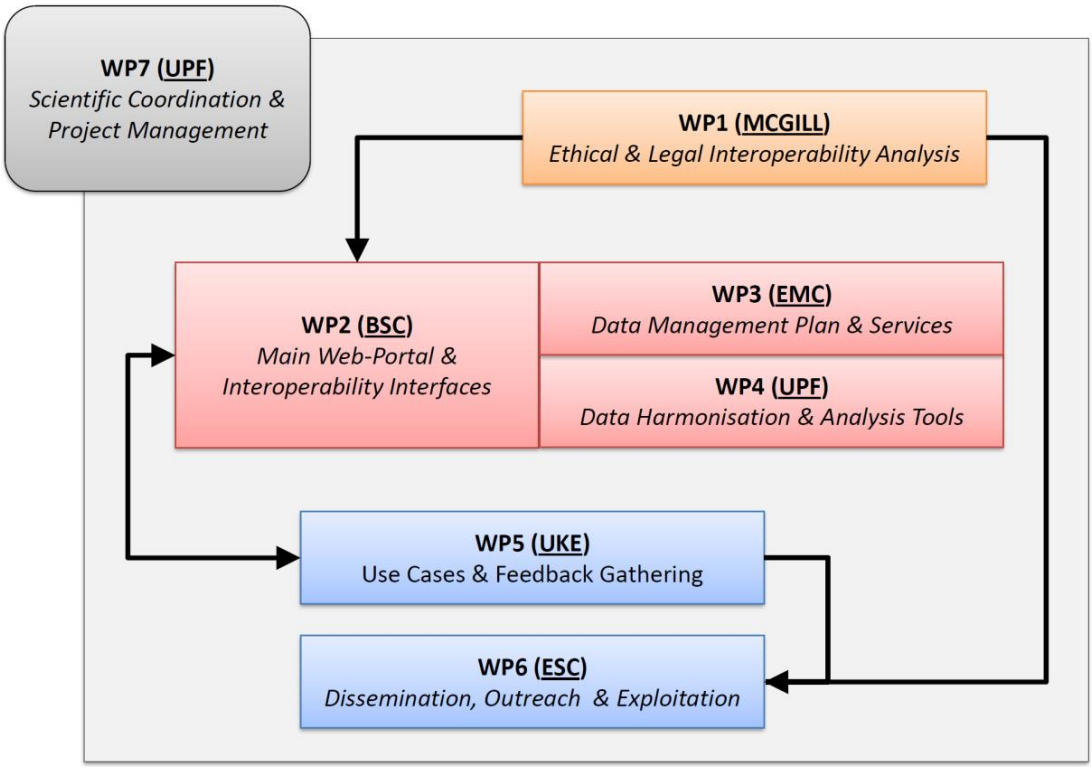


Figure 8 – *Overall structure of the euCanSHare project. It can be seen that the ethical and legal interoperability achieved in WP1 will guide the development of the euCanSHare platform in WPs2-4, while the work on enhancing Open Science will be directly disseminated in WP6 to the research community at large. WP2, WP3 and WP4 will work closely together to implement and integrate all components of the euCanSHare platform, linking services, infrastructures and datasets. There will be an important bidirectional communication between WP2 and WP5, i.e. the platform coordinated in WP2 will be regularly iterated, updated and passed to WP5 for multi-form detailed testing by the multi-disciplinary cardiovascular researchers, who will then gather feedback and specifications that will be passed back to the IT developers for fine-tuning the design and implementation of the platform in WP2 (in link with WP3 and WP4). The platform itself, as well as all research results derived from it in WP5, will be intensively disseminated in WP6, paying particular attention to demonstrating its concrete benefits for the cardiovascular research community from academia, public health and industry, as well as to creating a large community of users, collaborators and data contributors. Finally, all these activities will be coordinated and managed in WP7, ensuring optimal synergies between the IT developers, ethics experts and clinical researchers, as well as between the European and Canadian partners of euCanSHare.*

3.1 Work plan – Work packages, deliverables

WP number	**WP1**			**Period**			M1 – M36			**Leader**			MCGILL			
Title	Socio-ethical and legal interoperability analysis															
Participants	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	UPF	MUHC	UKE	LYN	EMC	QMUL	MCGIL	KUL	BSC	CRG	ESC	BBMRI	MCM	THL	UMG	NBD
Person-months	0.5	0	0.5	12	0.5	0.5	22	20	0	1	6	0.5	0.5	1	0	0

Objectives: WP1 is key to the entire project and is thus placed before all other implementation and integration WPs. It will put together the overall legal framework of euCanSHare by addressing the following objectives:

- Identify the relevant commonalities, mismatches, and gaps between the European (GDPR) and Canadian (federal) laws, regulations, and policies (T1.1).
- Perform a comparative analysis of existing informed consent forms and information brochures to assess their interoperability, and identify a list of consent items essential for data use within euCanSHare (T1.2).

- Develop a set of minimal data transactions, access requirements and data use conditions, for the implementation of a Smart Contract template to automate or semi automate transactions (T1.3).
- Address the issues related to incentives for fostering Open Science among researchers, with the ultimate goal of providing broad access to all aspects of research including datasets (Open data), publications (Open access), Software (Open source) and the primary record of research (Open notebook science) (T1.4).
- Engage with participating cohorts to assess their motivations and barriers (T1.5).

Task 1.1 – National and international legal frameworks
In light of the new GDPR which becomes enforceable in May 2018, this task will identify commonalities, mismatches, and gaps between EU and Canadian laws, regulations, and policies in relation to Canada–EU inter-jurisdictional personal data transfer. This will include an analysis of existing bridging attempts, such as EU regulators' analysis of the adequacy of Canadian data protection law[121], as well as the ongoing regulatory tug-of-war around broad consent to scientific research[122]. These issues will be analysed, mapped, and bridged through the expertise of the Center of Genomics and Policy (CGP, MCGILL). The CGP, in collaboration with other partners in the consortium, will conduct a comparative review of the current legal and ethical landscape in Europe and Canada, and will identify data protection challenges for euCanSHare in the context of international data sharing. This study will directly leverage research conducted by the MyHealth-MyData Scientific Coordinator (LYN) and the relevant legal expertise developed in that consortium. The practical focus of the project applied to GDPR's implementation strategies in distributed biomedical data networks, will be incorporated in the higher analysis as test cases and strategic guidance for euCanSHare's policies. A "Points to Consider" document will be drafted to benefit research projects, policy makers, and lawmakers in navigating these questions.
*Partners: **MCGILL**, KUL, LYN, CRG, BBMRI, THL, QMUL, UKE, MCM, EMC. Duration: months 1 – 24. Deliverables: D1.3.*

Task 1.2 – Retrospective ethical analysis
As summarised in Table 1, more than 35 cohorts will be integrated in the Platform. However, each cohort is governed differently and constrained by the unique terms of the consent form and data access policies. The CGP-P3G will perform a comparative analysis of existing consents to assess possible interoperability, and identify a list of consent items essential for data use (including existing limitations re storage and access). Additional steps (*e.g.*, formal access approval, re-contact/re-consent, etc.) will also be elucidated. This will in turn inform further studies and finally the outreach and implementation strategy for Canadian centres as well. WP1 will work in close collaboration with WP7 (project coordination) for the collection of these materials from participating cohorts. A retrospective assessment method will be used to develop a project-specific ethical and legal interoperability filter[123]. The 75 data use conditions listed in the Automatable Data Access Matrix (ADA-M), developed by the Regulatory and Ethics Working Stream (REWS) of the Global Alliance for Genomics and Health (GA4GH)[124] will be used as a starting point for the identification of data use conditions specific to euCanSHare and the development of an assessment matrix that fosters open science. Subsequently, this retrospective analysis will also inform the implementation of an international blockchain connection (T3.5) to support the maintenance of the *euCanSHare* catalogue. Leveraging the network of European clinical centres (*e.g.* in MyHealth-MyData) that have already implemented blockchain systems for data sharing, this task will evaluate actual value, challenges and overall sustainability in this context.
*Partners: **MCGILL**, LYN, CRG, MCM, THL, QMUL, UKE. Duration: month 1 – 36. Related deliverables: D1.4.*

Task 1.3 – Management of ethical and legal issues through Smart Contracts
In interaction with WP2, this task will develop a set of minimal common data transactions, access requirements and data use conditions, for the implementation of a Smart Contract to automate or semi automate transactions across institutions on the two sides of the Atlantic[125]. To do so, the ADA-M developed by GA4GH will identify comparable consent and access items to streamline cohort selection. This Smart Contract template aims to facilitate

[121] EC Decision 20 December 2001 pursuant to Directive 95/46/EC of the European Parliament and of the Council on the adequate protection of personal data provided by the Canadian Personal Information Protection and Electronic Documents Act (under document C(2001) 4539); Article 29 Working Party, Opinion 7/2014 on the protection of personal data in Quebec, 4 June 2014.

[122] Article 29 Working Party, Guidelines on Consent under Regulation 2016/679, 17/EN WP259, 28 November 2016.

[123] A.M. Tassé, E. Kirby, I. Fortier, "Developing an ethical and legal interoperability assessment filter for retrospective studies", Biopreservation and Biobanking Special Issue on ELSI in BioSHaRE, (2016) 14:3 Biopreservation and Biobanking 249-255.

[124] J.P. Woolley, E. Kirby, B.M. Knoppers, A.M. Tassé, et al., "Responsible Sharing of Biomedical Data and Biospecimens via the 'Automatable Discovery and Access Matrix' (ADA-M)" (submitted).

[125] Ekblaw, A., et al. (2016). *A Case Study for Blockchain in Healthcare: " MedRec " prototype for electronic health records and medical research data MedRec: Using Blockchain for Medical Data Access and Permission Management.*

data access and sharing processes by partially or fully self-executing and self-enforcing access requirements and data use conditions. In addition, explicit tracking of academic credits becomes possible by incorporating them into Smart Contracts. Thereby, when there is an agreement between the data producers and the data users that the data producer should be co-author of the papers resulting from analysis of datasets, it would be possible to put that agreement in a box in the blockchain. The technical implementation will be derived from the MyHealth-MyData project and through a detailed comparative analysis modified to address specifically research environments and international context. The basic computational infrastructure is expected to be minimally adapted.
*Partners involved: **MCGILL**, LYN, CRG. Duration: month 12 – 36. Related deliverables: D1.5.*

Task 1.4 – Identification of current incentives for Open Science
This activity addresses issues related to incentives for fostering Open Science among researchers, with the ultimate goal of providing broad access to all aspects of research including datasets (Open data), publications (Open access), Software (Open source) and the primary record of research (Open notebook science)[126]. The significance of developing adequate rewarding mechanisms and the creation of new platforms (*e.g.* GitHub, Dat.) and novel approaches (*e.g.* "Data Authorship") aim to introduce a broader range of reward in Open Science and increase transparency[127]. First, we will map current Open Science initiatives and their incentives, and critically analyse their strengths and weaknesses. We will use systematic literature review methodology for this study. Second, the incentive mechanisms in the blockchain technology, and the feasibility and adequacy of applying such mechanisms in the context of Open Science in biomedical research will be studied and policy points developed. In particular, we will study how far the use of blockchain technology and innovative incentivising mechanisms in terms of the academic credits such as "Authorship Coin" that could be integrated into Blockchain technologies could address the associated shortcomings with the existing incentives mechanisms and approaches. Third, the ESC will establish an expert group in order to discuss the challenges of incentives with the experts. For this purpose, two workshops will be organised in Brussels in the second and last year of the project. This expert group will consist of senior leaders in cardiovascular research aiming at sharing knowledge in big data approaches and assessing methodologies in data classification. The group will identify challenges and provide recommendations on how to overcome barriers and successfully apply technology mechanisms which will drive excellence in research. Intelligence and feedback will be shared with relevant projects, such as the BigData@Heart project which concerns assembling large cohorts for investigating heart failure and in which ESC and UKE are both partners.
*Partners involved: **KUL**, MCGILL, ESC, LYN. Duration: month 1-36. Related deliverables: D1.1.*

Task 1.5 – Direct engagement with participating cohorts to assess motivations and barriers to data sharing
This task will collect first-hand data from researchers regarding their experiences and opinions about data sharing. In particular, it will be crucial to study how the novel features of blockchain technology namely: decentralised control on access to the data, automated authentication and authorisation mechanisms, using Smart Contract, and heightened safeguards to protect the privacy of the data subjects will influence the researchers willingness to share data including biomedical data. We will use qualitative research methodology (semi-structured interviews) for collecting first-hand data. We will use an inductive, content-driven thematic approach for data analysis, in which common patterns or themes across the data are identified and analysed through a coding process. This coding process involves attributing codes that express the meaning and content of text fragments, and which will identify text fragments that are relevant to answering our research questions.
*Partners involved: **KUL**, MCGILL, ESC, BBMRI, UPF. Duration: month 8-24. Related deliverables: D1.2.*

Deliverables

D1.1 Results of interview study with researchers regarding blockchain technology for data sharing (KUL, M18)
D1.2 Policy "Points to Consider" tool to guide research projects and policy makers (MCGILL, M24)
D1.3 Comparative cross-mapping table detailing which participating cohorts are compliant with euCanSHare requirements (MCGILL, M24)
D1.4 Template smart contact to streamline data access and sharing process (MCGILL, M36)
D1.5 Policy points regarding the incentives mechanisms for data sharing technologies (KUL, M36)

[126] Chretien, J.P. et al. 2016. 'Make data sharing routine to prepare for public health emergencies', *PLoS medicine*, 13: e1002109.

[127] Bierer, B.E. et al. 2016. "Data Authorship as an Incentive to Data Sharing." In.: Mass Medical Soc.; Perkel, Jeffrey. 2016. 'Democratic databases: science on GitHub', *Nature*, 538: 127-28.

WP number	WP2			Period			M1 – M48			Leader			BSC		
Title	Main web-portal and interoperability interfaces														
Participants	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	UPF	MUHC	UKE	LYN	EMC	QMUL	MCGIL	KUL	BSC	CRG	ESC	BBMRI	MCM	THL	UMG	NBD
Person-months	15	12	7.5	12	8	0	0	0	60	30	0	0	10	22	8	0

Objectives: The overall objective of WP2 is to implement the computational infrastructure for euCanSHare, composed by a user web-portal that will provide access to data catalogues, data access committees, and analysis results, and a virtual environment to perform analysis on the data under the appropriate security conditions. WP2 will implement the computational environment for the integration of data, tools and solutions from several sources with both technological and geographical diversity. Specifically, it will prepare a computational infrastructure to offer in a single portal access to the different aspects of the project: i) access to cohort data, and metadata (data browser module), ii) access to data access policies, and facilitation of the procedures for gaining access privileges (DAC module), and iii) access to a data analysis layer honouring the appropriate security regulations (WP1), and data access privileges (Data Analysis module). The infrastructure will be based on previous developments by project partners (Mica, MuGVRE), in established industry standards (cloud managers, container technology, Galaxy, CWL), and accepted protocols (REST for data transfer, OAuth2 for authentication). The infrastructure will be fully virtualised, in a way that local installations to deal of specific data access regulations could be deployed. euCanSHare computational infrastructure will follow in any case the recommendations of GA4GH, ELIXIR, and will assure compatibility in protocols and standards with the EOSC initiative. Specific objectives are:

- Implement a portal framework to offer an integrated access to the components of the environment (T2.1)
- Design and implement data distribution channels within the consortia (T2.2)
- Implement a communication channel to facilitate the interaction with Data Access Committees (T2.3)
- Implement a cohort browser giving access to metadata catalogues (T2.4)
- Implement an integrated environment providing access to analysis tools and data, and to disseminate analysis results, with appropriate security measures (T2.5)

Task 2.1: Implementation of the euCanSHare portal framework

The initial step will be the generation of a base framework for the infrastructure, where the components of the portal can be easily plugged-in. This will allow for the maximum flexibility in the addition of tools or data sources, while maintaining a common structure for the core components. The framework will be largely based on MuGVRE (www.vre.multiscalegenomics.eu), a cloud-based infrastructure developed by BSC. It includes easy-to-use modules for authentication (based on the KeyCloak technology), secure data management (Oauth2 protocol for authentication at all data tranmissions, and encrypted data transfers). A simple protocol to adapt tools (data browsers, visualisers, or analysis tools) on top of the infrastructure is already available and will be exploited for euCanSHare (see[128]). The framework provides advanced execution scheduling (using COMPSs, NextFlow, www.nextflow.io), and OpenGrid Engine, allowing to derive tools execution to remote cloud infrastructures (through the OCCI protocol) and also to HPC environments. To enable its sustainability, the reference infrastructure will be hosted by BSC at the same ELIXIR compute equipment hosting MuGVRE and the EGA (European Genome-phenome Archive). This infrastructure will be also the basis of the technological offer of ELIXIR-ES within ELIXIR. This will be done by following actively the evolution of the EOSC framework. Nextflow binding of the infrastructure is being used by BSC in a EOSCpilot EU project involving re-analysis of EGA data, and MuGVRE itself has been presented at EOSC-DI4R meeting in Brussels (Dec 2017) and in the forthcoming EOSC meeting in Malaga, April 2018.

Partners involved: ***BSC****, CRG. Duration: month 1 – 12. Related deliverables: D2.1.*

Task 2.2: Implementation of data distribution channels within the infrastructure components

euCanSHare will be built by leveraging a series of data and metadata repositories of distributive nature, including EGA (-omics), euro-BioImaging (cardiac imaging), MORGAM/BiomarCaRE (clinical/lifestyle data) and BBMRI (BioSamples). Following the policies and data management plan defined in Task 3.1, this task will design and implement the necessary solutions to assure the efficient data interchange among the components of the

[128] http://www.multiscalegenomics.eu/MuGVRE/integration-of-tools

euCanSHare platform, namely, (meta)data repositories, and providers, public repositories, the access portal, and computational tools (WP4). To this end, firstly requirements will be gathered and analysed through communication with the relevant infrastructures, cohort owners and clinical investigators in the consortium, taking into account the types and heterogeneity of the cardiovascular data, as well as the necessary level of security (authentication, and data encryption). The most appropriate set of technological solutions will be implemented and tested based on guidelines of GA4GH and ELIXIR, as well as already deployed technologies such as FTP servers, Aspera or rest APIs and according to each data type repository (imaging, omics, clinical data, etc). WP1 will inform the task directly from a legal and ethical standpoint.
*Partners involved: **BSC**, CRG, MUHC, UPF, EMC, UKE, THL, MCM, QMUL, UMG. Duration: month 6 – 24. Related deliverables: D2.2.*

Task 2.3: Implementation of a framework to facilitate communication with Data Access Committees
The management of data access is a key issue when analysing of omics identifiable data, and often constitutes a major bottleneck. This task will implement a subportal dedicated to ease the process of acquiring access credentials according to the access rule s of the respective cohorts. The task will leverage strategies already being assayed in the EGA infrastructure (*e.g.* DACs tool). In close collaboration with WP1 (ethics and legal analysis), the subportal will i) disseminate access policies and procedures aligned to Global Alliance for Genomics and Health Standards (DUC: Data Use Condition and ADA-M), ii) provide a simple framework to apply for data access, iii) facilitate the procedure of granting and managing granted credentials by the committees (DACs tools). For selected cohorts, the possibility of automatic credentials assignment based on applications and policies metadata will be explored through the blockchain technology in T2.5.
*Partners involved: **CRG**, MCGILL, KUL. Duration: month 12 – 36. Related deliverables: D2.3.*

Task 2.4: Implementation of euCanSHare's catalogue and cohort browser
MUHC will extend Mica, a freely available open-source software used to create metadata portals for multi-study consortia. In euCanSHare, Mica will be optimised for enabling data custodians or network coordinators to efficiently organise and present information about their studies without significant technical effort. While Mica already includes modules to add and edit descriptive information pertaining to questionnaire data and physical measures, documentation of cardiac imaging data, omics, bio-samples and ELSI information are either only partially supported or not supported at all currently. This task will thus consist of customising the Mica software to allow documenting these new cardiovascular data types for the euCanSHare consortium, and to integrate them in the computational framework designed in T2.1. The metadata standards developed in T3.2 will be used to guide the software customisation and development process. The specific catalogue design will be based upon years of cardiovascular data management in the MORGAM and BiomarCaRE projects by THL and UKE. Our partners will subsequently document in the euCanSHare database the harmonised MORGAM data and, when necessary for more flexible harmonisation with other studies, the relevant original data from the individual cohorts participating in MORGAM and BiomarCaRE according to their respective access rules. This will include cohort baseline and re-examination data on classic cardiovascular risk factors, history of cardiovascular diseases and comorbidities (*e.g.* diabetes), and follow-up data such as on mortality, coronary events, stroke, heart failure and atrial fibrillation. MCM will lead the cataloguing of the Canadian cohorts from the Canadian Alliance (CAHHM).

Additionally, EMC (euro-BioImaging) will enrich the catalogue with additional information about the cardiac imaging data of each cohort, when available. Specifically, this will include information on the imaging modalities used in each cohort (*e.g.* cine-MRI, perfusion MRI, ultrasound), as well as the scan sequences and scan image acquisition parameters used. To this end, EMC will first define the most important imaging metadata from the participating cohorts, the existing literature, and from the image DICOM headers of representative imaging data. Existing tools developed at EMC will be re-used to automatically extract relevant data from the DICOM headers, while addressing heterogeneity in terminology from different image modality vendors and mapping to a harmonised terminology. The MIABIS structure, which is currently available for clinical and sample data, will be extended for imaging biobank information.

Once customisation of Mica metadata cataloguing modules is completed, metadata fields will be populated with information obtained from participating cohorts (*e.g.* BiomarCaRE, Canadian Alliance, UK Biobank). The final Mica-powered euCanSHare platform will include Maelstrom's searchable web-based solution for allowing investigators to quickly find the information and data they need to implement cardiovascular research projects.
*Partners: **THL**, MUHC, EMC, UPF, BSC, UKE, MCM. Duration: month 1 – 48 Related deliverables: D2.4.*

Task 2.5: Implementation of the integrated portal
The different components of euCanSHare infrastructure developed in T2.2 to 2.4 and analysis tools from WP4 will be made available through an integrated portal, deployed on top of the backend infrastructure developed in T2.1. An initial design phase will imply euCanSHare partners acting as data producers or consumers and will define the user experience accessing the portal. Common user interfaces like Galaxy (www.usegalaxy.org) or the MuGVRE

(www.vre.multiscalegenomics.eu) workspaces will be considered as references. The base infrastructure (T2.1) will provide a simple protocol for adding any type of software tool, which be included in a virtual machine (container), and a simple wrapper providing the necessary interfaces to the infrastructure. Analysis and data will be synchronised with the appropriate protocols as described in T2.2, such that users may navigate from the data portal to their personal workspace where data and tools will be available for further analysis. When necessary, data will be accessed following the appropriate secure protocols as defined in T3.5. The deployment of additional tools and interfaces for workflow management (*e.g.* Rocket, see WP4) will be enabled. This will allow users to combine different tools depending on the research questions, including data harmonisation, quality control, image quantification, bioinformatics and machine learning for personalised medicine research in cardiology.
*Partners: **BSC**, CRG, UPF, MUHC, MCM, UKE, THL, UMG. Duration: month 6 – 48. Deliverables: D2.5.*

Task 2.6: Implementation of the documentation module
UPF will assemble and integrate all documentation and user manuals into the web-portal and used as a help service for all users of euCanSHare. Specifically, the documentation will include user-friendly and easy-to-follow guidelines on how to (1) use the catalogue/cohort browser, (2) add cohorts to the catalogue, (3) apply for data access, (4) deposit new data into the repositories; (4) perform data harmonisation using Opal; (5) use the data analysis tools (with examples on how to build the computational workflows or use the bioinformatics methods); and finally (6) deposit the results of the studies back into euCanSHare (at EGA). From a technical point of view, the documentation module will re-use the same methods developed in the MuGVRE web-portal by BSC.
*Partners involved: **UPF**, ALL. Duration: month 12 – 48. Related deliverables: D2.6.*

Deliverables
D2.1 Initial Infrastructure framework and documentation (BSC, M12)
D2.2 Data distribution protocols and interfaces (BSC, M24)
D2.3 Data Access Committee Portal prototype (CRG, M36)
D2.4 Mica-powered catalogue populated with euCanSHare metadata (THL, M48)
D2.5 Integrated euCanSHare computational infrastructure (BSC, M48)
D2.6 Final documentation module (UPF, M48)

WP number	**WP3**			**Period**			M1 – M48			**Leader**			EMC			
Title	Data management plan and services															
Participants	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	UPF	MUHC	UKE	LYN	EMC	QMUL	MCGIL	KUL	BSC	CRG	ESC	BBMRI	MCM	THL	UMG	NBD
Person-months	0.5	8	1	9	30	0	0	0	17	35	0	4	0.5	10	0.5	0

Objectives: This WP will setup the models, tools and infrastructures to coordinate data management within the euCanSHare project, with special insight into the (meta)data sharing and distribution within the consortia, as well as long term repositories for external stakeholders. Data management in euCanSHare is key due to the heterogeneity of data types (-omics, imaging, clinical/lifestyle) and providers (*e.g.* BiomarCaRE, Canadian Alliance), their geographical distribution and the need of maintaining a high standard in secure access to the data. WP3 will establish a Data Coordination Centre for the euCanSHare project, which will 1) enable a global analysis of the data related issues within the consortium, such as data availability, transmission, or long-term storage; 2) address interoperability issues including cross-referencing between repositories (euro-BioImaging, EGA) and phenotype harmonisation between the cohort studies, 3) coordinate with WP2 to implement technical solutions for enabling efficient access to data based on the technologies developed by ELIXIR, EGA and euro-BioImaging, and 4) assure the implementation of FAIR data principles in euCanSHare date repositories. The specific objectives are:

- Define and maintain a Data Management Plan for euCanSHare (T3.1)
- Design the appropriate catalogues for data managed by euCanSHare (T3.2)
- Define and implement data deposition protocols for raw data (T3.3)
- Implement secure access to raw data repositories involved in the project (T3.4)
- Implement a prototype of the use of blockchain technology for clinical data access (T3.5)

Task 3.1 – Data Management Plan
The initial work in WP3 would be the establishment of a Data Management Plan (DMP) that will be published as D3.1. To this end, we will perform a comprehensive analysis of the nature of data to be handled in euCanSHare

(see Table 1). For each type of data (*e.g.* omics, imaging, clinical, etc), the needs for long-term storage, security issues, required metadata, and interoperability requirements will be analysed. The compliance with FAIR data principles will be analysed from day one and the necessary actions included in the plan. Access policies (in close collaboration with WP1) will be defined and stored in DUC or ADA-M. Project's DMP document will be in constant revision as the project proceeds, being formally updated when necessary.
Partners involved: CRG, EMC, BSC, BBMRI, ALL. Duration: month 1 – 12. Related deliverables: D3.1 & D3.2.

Task 3.2 – Align metadata standards for documenting imaging, omics, bio-sample and ELSI information
In collaboration with T2.3 (data cataloguing), EMC and MUHC will define metadata fields for the project, as collected or generated by euCanSHare cohorts and commonly used in cardiovascular research. Selection of the fields will be informed by existing standards adapted to serve the specific needs of the project. A Working Group will be established and convened at consensus meetings to define standard metadata for imaging, omics, epidemiological, clinical, and bio-sample data. It will be led by QMUL and composed by all euCanSHare experts and additional external collaborators including those planning new cohorts or having existing cohorts that are not part of the initial euCanSHare set. The objective of this exercise will be to establish standard attributes to describe different data types and provide means to disseminate this information to the euCanSHare investigators and the broader cardiovascular research community. This will also define the minimum requirement for data provision from cohorts/studies in order to become potentially eligible for contributing and benefitting from euCanSHare, thus maintaining a consistent quality standard. We will concentrate on establishing key common variables across the cohorts in euCanSHare, discussing their scientific merit for cardiovascular research (*e.g.* observer variability, test-rest reproducibility, accuracy, evidence of link to diagnosis, risk and treatment response monitoring). The output will also contain a list of variables that each new prospective cohort is recommended to contain, with guidelines on how the variables are measured and recorded with their accompanying meta-data (*e.g.* details on acquisition and analysis). The experts will have quarterly online meetings using collaboration tools (Google drive) to advance the task and will also have two face to-face meetings at PM18 and PM42 (one in Canada, one in London). The ELSI experts from WP1 will provide key input into the development of metadata fields to document and disseminate information related to the data access requirements and procedures.
*Partners involved: **EMC**, QMUL, UPF, EMC, QMUL, UKE, THL, MCM, BBMRI, BSC, CRG, MCGILL.*
Duration: month 1 – 18. Related deliverables: D3.3.

Task 3.3 – Deposition of new heterogeneous cohorts in euCanSHare's repositories
While WP2 focuses on the documentation of metadata, this task will define the protocols to deposit new cohort raw data into the appropriate repositories following the DMP policies, and the methods to provide rich enough metadata to foster a quality re-use of raw data for newly coming research projects. EMC and CRG will implement mechanisms for linking metadata across the different repositories (imaging, -omics, bio-samples) to assure the transversal consistency of the specification of studies that can be showed in an integrated way. This task will ensure that request of information is performed according to specific formats as means to check quality control before data deposition. This task will realise the recently published joint-strategy of ELIXIR and euro-Bioimaging to link their infrastructures for integrating imaging and biomolecular data[129]. Furthermore, bio-samples will be deposed according the guidelines and quality assurance defined by BBMRI.

At the same time, EMC will adapt its research infrastructure for euCanSHare to allow for standardised large-scale data storage for cardiac imaging and advanced automated image analysis. The euCanSHare infrastructure will be built around the state-of-the-art open source XNAT software (www.xnat.org), which will facilitate secure storage and management of the cardiac imaging data. Not only raw (DICOM format) images will be stored, but also image-derived data like segmentations and meshes, and other associated quantitative data. Images and derived data can be accessed (provided the user is authorised to do so) both via a graphical interface and via a programming interface, facilitating automated batch processing. The euCanSHare infrastructure will enable anonymised image data upload based on the international DICOM standard. While a central imaging archive is most practical for collaborative multi-centre studies, it may not always be feasible due to restrictions on sharing data (*e.g.*, data that must stay within the firewall of an institution). To serve such scenarios, a federated infrastructure for imaging storage will be integrated, where partners host their own XNAT archive. EMC will offer installation scripts (based on Salt by SaltStack) and support to project partners in setting up their own XNAT archives. By registering project metadata in the catalogue (see WP2), these cardiac imaging datasets will become findable for researchers.
*Partners involved: **EMC**, CRG, BSC, BBMRI. Duration: month 1 – 24. Related deliverables: D3.4.*

Task 3.4 – Secure access to raw data repositories

[129] https://www.elixir-europe.org/system/files/euro-bioimaging_elixir_image_data_strategy.pdf

A large set of data managed in euCanSHare requires secure access. Following the policies defined in the DMP (T3.1), a series of repositories (EGA, euro-BioImaging, BBMRI) for long-term storage of raw data will be leveraged. In collaboration with WP2 (T2.3, T2.5), Task 3.5 will implement the necessary technology to gain access to such data, following the rules established in WP1. Secure technology to access EGA repository from cloud facilities is already being developed within ELIXIR. We will extend such technology to the specific needs and repositories used in euCanSHare for supporting -omics, imaging and other phenotypic data, and provide controlled access to both platform components, and external users.
*Partners involved: **CRG**, BSC, MCGILL. Duration: month 6 – 36. Related deliverables: D3.5.*

Task 3.5 – Implementation of a prototype of the use of blockchain technology for automated data access
This task will include two distinct phases stemming from the work carried out in WP1-2. Specifically, research centres which opted to install blockchain-based data access services (QMUL from the EU and MCM from Canada) will work with LYN on the implementation of the "Data Access Control Tool" (DACT), in the form of an institutional Smart Contract exposing to external researchers' criteria and requirements to access a given data set. At the same time, internal researchers involved in the clinical studies (WP5) will be provided with a data access tool instantiating credentials, study details and institutional qualifications, also in the form of a Smart Contract. The blockchain infrastructure developed in the MyHealth-MyData project will be then deployed in a controlled fashion and tested. This will also include extending, based on the work and WP1, legal and ethical clauses governing data access transactions. The basic infrastructure will have been extensively tested as part of the My-Health-MyData project, including formal penetration and hacking challenges. Nevertheless, not only functional testing to check completion of intended transactions will be executed, but also security testing. APIs for direct data access, after authentication and permission, will be developed at least in two institutions (QMUL and MCM). In others, with local data storage, only the data access process will be automated with blockchain. The result of such process is the permission in digital form to access data (or its denial) delivered to the parties.
*Partners involved: **LYN**, QMUL, CRG, MCM. Duration: month 3 – 48. Related deliverables: D3.6.*

Deliverables
D3.1 Data Management Plan v1 (CRG, M6)
D3.2 Data Management Plan v2 (BSC, M12)
D3.3 Recommended metadata standards (EMC, M18)
D3.4 Guidelines and protocol for data deposition (EMC, M24)
D3.5 Report on the implementation of raw data access protocols (CRG, M36)
D3.6 Prototype of blockchain technology (LYN, M48)

WP number	**WP4**			**Period**		M1 – M36				**Leader**		UPF			
Title	Data harmonisation and analysis tools														
Participants	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	UPF	MUHC	UKE	LYN	EMC	QMUL	MCGIL	KUL	BSC	CRG	ESC	BBMRI	MCM	THL	UMG	NBD
Person-months	58	12	24	0	2	2	0	0	2	1	0	0	0	18	30	0

Objectives: WP4 will develop the Data Analyser module of euCanSHare, allowing users, after having identified the relevant cohorts and obtained permissions to access the data, to process the data and generate results. They will access the data on the secure servers at ELIXIR/EGA and build computational workflows composed of a number of tools comprising data harmonisation between the selected cohorts, image analysis, bioinformatics for analysis -omics data, and machine learning/statistical methods to quantify associations and new biomarkers, and to build predictive models of cardiovascular disease. This module will be developed to be flexible, allowing for different types of harmonisations, image quantifications and bioinformatic analyses. The content of these tools has been selected thanks to the experience of the consortium member in each of their respective fields (MUHC: harmonisation; UPF, QMUL, EMC: image analysis; UKE, BSC: bioinformatics). The robustness of the Data Analyser module will be ensured by exploiting existing and validated tools (Opal, Rocket, Square2, Debian). The capabilities provided by this module will constitute a great incentive for cohort owners to upload their data into the system as they will be able to use state-of-the-art computational workflows for processing their multi-disciplinary data (*e.g.* imaging + omics) in a single environment connected to large data. Specific objectives are:

- Make existing tools for cross-cohort harmonisation available to cardiovascular researchers (T4.1).
- Provide data quality control tools to ensure data, including harmonised data, is suitable for subsequent data

quality control (T4.2).

- Provide tools for cardiac image quantification (T4.3)
- Provide bioinformatics toolbox for cardiology (T4.4)
- Provide machine learning toolbox for cardiology (T4.5)
- Enable assembling these different tools into flexible computational workflows for cardiovascular data processing (all tasks, in link with T2.5)

Task 4.1 – Tools for incremental data harmonisation
Through the Maelstrom Research group at the MUHC, tools for data harmonisation will be made available to euCanSHare investigators. This task will implement the standardised electronic database that will store the harmonisation algorithms for cardiovascular cohorts and variables. To this end, MUHC's software Opal will be adapted to provide a centralised web-based harmonisation management system allowing study coordinators and data managers to securely import/export a variety of cardiovascular data types and harmonisation rules in different formats using a point-and-click interface. Opal's functionalities will be made available to define variables targeted for harmonisation, implement and store the conversion algorithms, and efficiently document data harmonisation decision-making. Subsequently, data transformation scripts will be saved and displayed in the catalogue and cohort browser (T2.4) to create a searchable database to enable future researchers to identify past harmonisation algorithms for their multi-cohort studies. Initially populated based on the MORGAM, BiomarCaRE and CAHHM past harmonisations, the harmonisation database will be extended based on the use cases investigated in WP5.
*Partners involved: **MUHC**, THL, CRG. Duration: month 1 – 24. Related deliverables: D4.1*

Task 4.2 – Tools for data quality control
Through UMG, tools for easy accessible and extensive data quality reports will be made available for euCanSHare investigators to obtain broad insights into key data quality aspects such as missing values, extreme values, implausible values, observer or device variability. This will support targeted data selections and analysis plan decisions. The analysis of data quality aspects will be possible before researchers obtain direct access to the data at the individual level. The technical approach draws strongly from the potential of automatised data monitoring processes by improved metadata handling and is currently in use in large scale cohorts such as the SHIP Study; one of the participating cohorts of euCanSHare. The targeted web-tool for automated data quality analysis, Square[2], will be adapted along the PostgreSQL backend to the particular needs of this endeavour. Collaboration within WP 3 will ensure a direct link between Square and the euCanSHare data repositories to enable the necessary study data and metadata use for automated processes. Analyses will be processed asynchronously to improve computational speed. The tool will be applied and extended within the defined use cases in WP5.

Furthermore, the tool will be extended to support data quality control of images to identify for example incorrect coverage of the heart by the cardiac images. To this end, QMUL will score a selected subset of images from the UK Biobank, while UPF and EMC will implement machine learning approaches (*e.g.* support vector machine) to automatically estimate a quality score of the images.
*Partners involved: **UMG**, UPF, QMUL, EMC. Duration: month 1 – 24. Related deliverables: D4.2*

Task 4.3 – Cardiac image analysis platform
UPF and EMC will make available the first open-source cardiac image analysis pipeline, based on the technologies developed during past cardiac EU projects such as euHeart, CardioFunXion and VP2HF. In this latter project, Rocket was shown for its capability to manage any type of cardiac-related data, including patient demographics, medical signals and images, physiological simulations and statistical results, as well as to enable the connection to cloud-based computational pipelines. In this task, the Rocket platform will be used to integrate state-of-the-art cardiac image quantification algorithms available in the literature to produce reproducible cardiac image quantification workflows. Several key capabilities will be offered for researchers, including automated cardiac segmentation (with tools for interactive manual corrections) that will generate estimates of clinical indices of cardiac function such as chamber volumes, ejection fraction, cardiac output, wall thickness, etc. Furthermore, UPF recently developed a tool for radiomics-based cardiac image analysis that will be incorporated to enable deeper phenotyping of cardiac shape and myocardial texture in cardiovascular studies.

To facilitate automated image processing and data analytics, the XNAT-based imaging archived will be linked to high performance and cloud-based computing infrastructures. An open plug-in architecture (FASTR, www.fastr.readthedocs.io, developed by partner EMC) will be provided for testing imaging algorithms that can be processed on top of the images stored in the project image collections. The results of these image analyses cannot only be viewed in the local expert tools, but can also easily be shared with consortium partners by uploading them back onto the XNAT platform. The data transport between XNAT and a FASTR enabled compute environment will be managed by the python library XNATpy (www.xnat.readthedocs.io). FASTR will implement strict provenance mechanisms, to enable future reproducibility of any imaging biomarker computed. FASTR will also

manage the remote execution on high-throughput clusters and cloud environments, enabling heavy and large-scale imaged analyses on large datasets. Using container technologies, the ease of sharing and deploying FASTR pipelines will be further improved.
*Partners involved: **UPF**, EMC, QMUL. Duration: month 1 – 36. Related deliverables: D4.3*

Task 4.4 – euCanSHare's bioinformatics toolbox
In this task, a bioinformatic computational resource will be assembled for analysing the variety of omics data available in cardiovascular research, specifically genomics, transcriptomics, metabolomics and proteomics data. The goal will be survey, analyse, select and test from the large space of existing libraries those methods that are most relevant, including from open-source software such as PLINK2, IMPUTE4, HISAT2, as well as from bioinformatics communities BioPython, Bioconda and Debian Med. The module will be also connected to the R project[130], which provides a language and an environment for statistical computing. UMG's techniques for multi-omics integration (*i.e.* mapping data onto molecular networks such as derived from protein-protein interaction databases) will be also considered[131]. The final bioinformatics methods will integrate computational methods as well as functionalities designed to import, analyse and visualize omics data. Subsequently, these methods will be integrated to a web interface within euCanSHare platform through the Galaxy framework (cf. T2.5) and linked to the cloud resources of ELIXIR/EGA to enhance the use of bioinformatics tools in cardiovascular research. This will make these analyses more accessible and reproducible and will facilitate navigation and analysis for non-experts. The web interface will be extended to enable other researchers beyond the lifetime of this project to add their bioinformatic methods using the Docker Container technology. Linked to the image analysis and machine learning tools (in T4.5. and T4.6, respectively), this toolbox will enable interactions between different disciplines of cardiovascular research (*e.g.* genomics, proteomics, radiomics), leading to new hypotheses and knowledge.
*Partners involved: **UKE**, THL, UMG, BSC. Duration: month 1 – 36. Related deliverables: D4.4*

Task 4.5 – Statistical and machine learning toolbox for cardiac data types integration
In this task, Rocket will act as the platform for integrating a set of machine learning and statistical techniques that are commonly used in cardiovascular research studies, especially assembled to handle different types of variables (-omics, imaging, lifestyle, clinical measurements, ECGs, etc). For example, UPF developed within the VP2HF project methods to automatically build and visualise decision trees from databases of heart failure patients, which were then used to improve clinical guidelines for diagnosis and treatment. Other advanced machine learning techniques in which UPF has experience in cardiology is multiple kernel learning methods for the analysis of cardiac function[132]. A survey will be carried out within the consortium to select the final list of machine learning and statistical techniques to be integrated (*e.g.* multivariate regression, support vector machine). Since Rocket is highly modular and flexible, it will allow the integration of the statistical and machine learning toolbox with the bioinformatics one based on Galaxy (T4.4, T2.5) and the cardiac image analysis platform. Moreover, the resulting infrastructure will allow future researchers to add new methods using Docker technologies. User-friendly interfaces will also be developed using visual analytics principles enabling remote and shared visualisation of clinical data and the outcomes from statistical and machine learning tools. Specifications of these interfaces will be guided by the use cases from WP5 to personalise the available tools to concrete needs of cardiovascular research.
*Partners involved: **UPF**, UKE, THL, UMG. Duration: month 12 – 36. Related deliverables: D4.5.*

Deliverables
D4.1 Opal software integration to euCanSHare (MUHC, M12)
D4.2 Data quality control tool (UMG, M18)
D4.3 Cardiac image analyser (UPF, M24)
D4.4 Bioinformatics toolbox (UKE, M24)
D4.5 Machine learning toolbox (UPF, M36)

[130] R Core Team. R: A Language and Environment for Statistical Computing. (R Foundation for Statistical Computing, 2017).

[131] Amberkar, S.S. and Kaderali, L., 2015. An integrative approach for a network based meta-analysis of viral RNAi screens. *Algorithms for Molecular Biology*, *10*(1), p.6.

[132] Sanchez-Martinez, S. et al. 2017. Characterisation of myocardial motion patterns by unsupervised multiple kernel learning. *Medical image analysis*, *35*, pp.70-82.

WP number	WP5			Period			M12 – M48			Leader			UKE			
Title	Use cases and feedback gathering															
Participants	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	UPF	MUHC	UKE	LYN	EMC	QMUL	MCGIL	KUL	BSC	CRG	ESC	BBMRI	MCM	THL	UMG	NBD
Person-months	24	8	42	0	0	36	0	0	2	1	0	0	10	24,5	8	24

Objectives: This WP will consist of a number of use cases and pilot-test studies with increasing complexity that will enable to adjust the capabilities of the platform throughout the project and test its potential for personalised medicine approaches. This will be done by consortium members with significant experience in data harmonisation and analysis (MUHC: Maelstrom; UKE/THL: MORGAM/BiomarCaRE; QMUL: UK Biobank, MCM: CAHHM). The objectives of this WP are to:

- Test the data harmonisation tools (T5.1).
- Test the platform for biomarker validation (T5.2).
- Test the platform for knowledge discovery in complex cardiac diseases (diabetic cardiomyopathy), taking into account sex and country differences (T5.3).
- Test the platform for improved risk prediction by integrating genetics and imaging (T5.4).
- Test the platform for public health research (T5.5).
- Test the platform for multi-omics target discovery by our SME partner NBD (T5.6).
- Gather feedback from user experience and provide recommendations to the IT team (T5.7).

Task 5.1 – Data harmonisation for the use cases
This task will generate the harmonised data for the tests and use cases performed in T5.2-T5.4. Harmonised variable format specifications developed by BiomarCaRE will be used as the reference formats and thus applied to new European cohorts or Canadian cohorts participating in the use cases when data is available and compatible. To achieve the harmonisation work, a central Opal server will be installed at EGA to host subsets of cohort data. Whenever data sharing policies prevent cohorts from sending data to third parties, servers will be installed locally (*e.g.* using the Local-EGA technology). The Opal software will then be used to input variables targeted for harmonisation (as defined in BiomarCaRE variable format specifications), develop and implement processing algorithms to derive common-format data, and efficiently document data harmonisation decision-making. As a proof-of-concept, only variables relevant to uses case research questions (T5.3 and T5.4) will be harmonised.
*Partners involved: **MUHC**, UKE, QMUL, THL, UMG. Duration: month 12 – 48. Related deliverables: D5.1.*

Task 5.2 – Initial testing of harmonised datasets and platform capabilities
Following initial harmonisation of phenotypic, clinical and biomarker data in T5.1, this task will cover the verification of the data sharing platform and the harmonisation success by analysing known correlations and associations of cardiovascular risk factors and biomarkers to phenotypes and clinical outcomes, more specifically those tested in the BiomarCaRE EU project by UKE based on multiple European cohorts. Here, the tests will be performed by combing BiomarCare data with Canadian cohorts from CAHHM and new studies such as the UK Biobank. In a first step, known risk factors will be validated such as lipid status[133] and smoking in combination clinical variables and available endpoints. In a second step, the predictive accuracy of biomarkers such as Troponin[134] and HDL/LDL-Cholesterol will be assessed with respect to the different geographical regions covered in the cohorts. With this task, it will be possible to verify that all parts of the platform are in place and fully functioning. More advanced testing will be then performed in T4.3 and T4.4 with more complex use cases.
*Partner involved: **UKE**, THL. Duration: month: 12 – 24. Related Deliverables: D5.2.*

Task 5.3 – A multi-country and multi-factorial study of diabetic cardiomyopathy
This task will exploit improved access, sharing and analytics capabilities of euCanSHare to run a large multi-country investigation of diabetic cardiomyopathy. Firstly, five to six cohorts which have patients with diabetic cardiomyopathy, imaging, genetic and lifestyle/environmental data will be identified and selected through the

[133] Waldeyer, C., Makarova, N., Zeller, T., et al., 2017. Lipoprotein (a) and the risk of cardiovascular disease in the European population: results from the BiomarCaRE consortium. *European Heart Journal*, *38*(32), pp.2490-2498.

[134] Blankenberg, S. et al. 2016. Troponin I and cardiovascular risk prediction in the general population: the BiomarCaRE consortium. *European heart journal*, *37*(30), pp.2428-2437.z

Cohort Browser. After access have been requested and granted through the Access Manager, data harmonisation of the cohorts will be performed in T5.1 using the Opal tool by using BiomarCaRE as the standardisation reference. Subsequently, we will build computational workflows to analyse the generated multi-country harmonised dataset as follows. Firstly, we will use the image analysis tools to derive new deeper phenotypes of DM-CM patients and investigate how the cardiovascular image phenotype differs between diabetic patients and non-diabetics. Secondly, we will investigate how these new imaging phenotypes are related to co-variates such as age and sex, as well as to characteristics of diabetes (duration, HbA1c, type of). Subsequently, by nature of this Canadian and European collaboration, we will uniquely investigate the impact of ethnicity, protective and detrimental lifestyle, environment and context, genetic factors, in diabetic patients with diabetic cardiomyopathy. Finally, we will assess the ability of the integrated phenotypes and covariates to predict cardiovascular morbidity and mortality associated with DM and to predict treatment response.
*Partner involved: **OMUL**, UPF, MCM. Duration: month 24 – 48. Related Deliverables: D5.3.*

Task 5.4 – Integrating radiomics and genomics in cardiovascular risk assessment
This task will evaluate the value of imaging and genomics combined data for personalised medicine approaches and risk predictive of cardiovascular events. Longitudinal cohorts with the relevant data, including cardiovascular events such as myocardial infarction and stroke, will be identified and the researchers working in this task will request access through the euCanSHare platform. After harmonisation is performed in T5.1, the image analysis and bioinformatics toolboxes will be used to identify the radiomic features and genes that are associated with cardiovascular risk. Machine learning techniques will be then used to integrate these identified variables with known risk factors such as age, sex, hypertension, smoking, etc. The new predictive model of cardiovascular risk, integrating classical risk factors with advanced imaging phenotypes and genetic factors will be validated through a multi-region approach across Europe and Canada.
*Partner involved: **UKE**, UPF, MCM, THL, UMG. Duration: month 24 – 48. Related Deliverables: D5.4.*

Task 5.5 – Multi-country public health test study
This task will demonstrate the applicability of the data harmonisation, sharing and analytic capabilities of the platform for public health research. First, we will assess the extent to which data on classic cardiovascular risk factors (age, sex, smoking, obesity, blood lipids, blood pressure and diabetes), as well as death and cardiovascular outcomes (acute coronary events and stroke) can be harmonised across the participating cohorts. We will then compare the risk estimates and population attributable fractions (*i.e.* the fraction of all cases of a particular disease in a population that is attributable to a specific risk factor) across countries and continents. The analysis will be weighted to highlight the relative public health importance of the different risk factors in each population. Finally, the results will be assessed for their potential impact on the planning of prevention and control of cardiovascular disease, taking into account the differences within and across Canada (provinces) and Europe (countries).
*Partners involved: **THL**, MCM. Duration: month 24 – 48. Related deliverables: D5.5.*

Task 5.6 – Target discovery study
This task will evaluate the potential of the platform for target discovery/development of new therapies related to CVD. NBD will link its computational tools (entitled PELE: Protein Energy Landscape Exploration: see www.pele.bsc.es/pele.wt) to the platform and analyse the large multi-omics integrated data (genomics, transcriptomics and proteomics). The goal is to localise proteins that are statistically linked to the onset of CVD and to identify their structure in NBD's Protein Data Bank. The identified structure will be thoroughly studied, and a range of personalised variants will be built with the help of homology modelling tools, following hints derived from the genomics studies. A series of simulations will be carried out for exploring the flexibility and molecular recognition of the protein in different settings (wild type, pathological, genomic variants related to population sub-groups, etc). The simulated structures will be then used for *in silico* virtual screening campaigns in search of new chemical entities (NCEs) with activity on the protein, based on compound catalogues such as Zinc15 (www.zinc15.docking.org/substances/home). The prioritised compounds will be tested in *in vitro* biochemical assays (wild type and mutant variants) and in cell-based assays to verify biological activity. The hits found will be optimised for potency and ADMET properties) in several rounds of chemical synthesis. Because of the common difficulty to find suitable proteins, several will be tested using the process described above. As part of this study, a cost/benefit analysis will be conducted by NBD to assess research productivity related to the platform compared to their traditional R&D performance.
*Partners involved: **NBD**, BSC, CGR. Duration: month 24 – 48. Related deliverables: D5.6.*

Task 5.7 – Iterative feedback gathering and analysis on platform usage
This task will assess and evaluate the performance and usage of the data sharing and analytics platform throughout the project, enabling to constantly improve it based on the feedback of end users, as well as to ensure that its continuous developments are guided by requirements of clinical researchers and other end-users (*e.g.* companies). This task will organise surveys at the end of each round of multi-cohort studies (T5.2 -T5.6 and will analyse the

user's feedback in a systematic manner by aggregating the results from the different studies and users. The standard usability measures to be applied are effectiveness (the extent to which the intended goals of use are achieved), efficiency (the resources such as time or mental effort spent to achieve the goals), and user satisfaction (the extent to which the user finds the system acceptable). Finally, UKE will communicate the feedback and recommendations from the different users to the IT teams working of the platform development in WP2-4.
*Partner involved: **UKE**, UPF, QMUL, MUHC, MCM, NBD. Duration: month 12 – 48. Related deliverables: D5.7.*

Deliverables
D5.1 Harmonised data for use cases (MUHC, M24)
D5.2 Results on feasibility testing for classical cardiovascular risk factors and biomarkers (UKE, M24)
D5.3 Results of diabetic cardiomyopathy multi-country study (QMUL, M42)
D5.4 Results of validation of joint improved risk prediction on cardiovascular events (UKE, M42)
D5.5 Results of public health study (THL, M42)
D5.6 Results of target discovery study (NBD, M48)
D5.7 Feedback report on assessment of platform usage (UKE, M48)

WP number	**WP6**			**Period**		M1 – M48			**Leader**		ESC					
Title	**Dissemination, outreach and exploitation**															
Participants	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	UPF	MUHC	UKE	LYN	EMC	QMUL	MCGIL	KUL	BSC	CRG	ESC	BBMRI	MCM	THL	UMG	NBD
Person-months	3	3	1	18	0.5	1	1	4	1	1	18	3	2.5	1	1	4

Objectives: This WP will ensure the platform is first and foremost disseminated to the widest audience of relevant stakeholders, including cohort providers, cardiovascular scientists, biologists and bioinformatians, biotech and pharmaceutical companies, public health researchers and data managers. To this end, concrete dissemination activities will be organised throughout the projects, including (i) live demos and hands-on sessions organised by our partner the European Society of Cardiology (ESC) in Brussels; (ii) three workshops to present and discuss the innovation potential of the platform to the academic, healthcare and industrial communities; (iii) a consensus building campaign for enhancing Open Science such as through the proposed blockchain/Smart Contract approaches; (iv) intensive outreach campaigns to invite new cohorts and boost awareness on the capabilities of the platform. Furthermore, communication activities will be organised through various mechanisms (branding, website, dissemination material), as well as through the different channels of our multi-disciplinary consortium, including the channels of ESC, BBMRI and ELIXIR/EGA in Europe, and MUHC, CAHMM and MCGILL for Canada. This will ensure project mission, rationale and goals *(communication and outreach)*, along with personalised medicine research results generated by the project, are suitable communicated to relevant stakeholder groups, correspondent to as many selected target audiences, to be parallelly addressed in EU and Canada. Furthermore, exploitation activities will target platform scalability and sustainability, as well as commercial exploitation beyond the project timeframe with specific initiatives, taking the best out of achieved results and relying on the integrated network of stakeholders created throughout the project from both Canada and Europe.

Task 6.1 – Dissemination strategy, branding and communication materials
This task will start with the preliminary study phase laying the foundation of the *Communication and Dissemination (C&D) strategy*, to be elaborated in close relation with the concomitant analysis of the stakeholder target groups (T6.2). This also includes the preparation of *branding and preliminary materials* such as project website, logo, infographics, presentation template and banner for EU acknowledgments, social media accounts, and regular press releases. All materials will be made available on the project website at public disposal to facilitate project showcasing by third parties. The C&D strategy plan will result from a thick cooperation between all partners involved, both from the Canadian and European sides. Subsequently, the strategy will be adjusted according to emerging needs and outcomes of different implemented measures, leading to the elaboration of a comprehensive set of C&D materials to be designed in terms of content messages, language and distribution channels according to different target audiences (cardiovascular researchers, SMEs, public health agencies, general public, etc). This material will include newsletters, press releases, white papers, newspaper or magazine articles, as well as peer-reviewed publications. Finally, this task will be also dedicated to the regular update and content curation of web-based communication channels, including the project website and social media accounts (Twitter, Google+, LinkedIn, Facebook, ResearchGate), multimedia communities (Flickr, Vimeo), as well as the Open

Access project material archive (Zenodo). The different community websites of our partners (*e.g.* EGA, ESC, Maelstrom, CAHHM, BiomarCaRE, MORGAM, euro-BioImaging) will be heavily exploited to deploy the project's C&D strategy. Social media accounts will be the primarily diffusion channel for awareness campaigns, project news (meetings, events) and achievements (publications, prototype releases of the euCanSHare platform). Peer-reviewed publications, reports, presentations and deliverables will be promptly uploaded in Zenodo according to the relevant Open Access policy.
*Partners involved: **LYN**, ESC, ALL. Duration: month 1 – 48. Related deliverables: D6.1 & D6.2.*

Task 6.2 – Stakeholder analysis
This task will be aimed at a preliminary analysis of identified target groups of stakeholders in Europe and Canada, including cardiologists and health institutions, biomedical research entities, public agencies and biotech/pharma companies, finalised to an appropriate setting of the C&D strategy from T6.1. The task will not only imply mere identification of categories of stakeholders but will also go in depth into the different professional figures attaining to each category, their respective roles, specific requirements and possible criticalities, together with envisaged *ad hoc* measures to be implemented throughout the project development. Targeted email campaigns and newsletters to a diverse set of mailing lists will be put together early in the project and extended throughout. The task will rely as much as possible on existing contact networks and amplification of messages through related channels of consortium partners (ESC: members, national cardiovascular societies, subspecialty communities and affiliated cardiac societies – over 140,000 contacts; EGA data access committees; BBMRI communities, Maelstrom partnerships: ReACH, MINDMAP, CPTP, CLSA, P3G). Cross-fertilisation with other research initiatives will be pursued such as with BigData@Heart (see letter of interest), the Framingham Heart Study (see letter of support) and other identified projects of relevance (*e.g.* CrowdHealth, KONFIDO, SoBigData, EuroHealthNet1, IC-Health), while maintaining cooperation with other associations and public initiatives around Europe and Canada in the domain of personalised medicine (*e.g.* International Consortium of Personalised Medicine).
*Partners involved: **LYN**, ESC, BBMRI, ALL. Duration: month 1 – 12. Related deliverables: D6.3.*

Task 6.3 – Stakeholder outreach and awareness campaigns
This task will concern the setup of specific campaigns for direct engagement of stakeholders onto the biomedical research data sharing platform, in particular for (1) adding new cohorts beyond the initial set of 35 cohorts (Table 1), (2) attracting new users of the platform including companies and (3) building consensus on the Open Science solutions developed in the project such as the blockchain and Smart Contract for more efficient as well as responsible/transparent data sharing. The task will take base from the preparatory stakeholder analysis carried out in T6.2, and will expand it with new contacts acquired through dissemination activities (mailing list subscribers, social media followers, direct contacts from networking activities at dissemination events) and enact a specific engagement campaign in both Europe and Canada, directly contacting institutional subjects pertinent to different of stakeholders categories to onboard them onto the platform, identifying and addressing their actual needs in terms of data sharing and/or usage, particular concerns or other individual peculiarities. To sustain the enrichment of the platform, the consortium will take care of directly contacting healthcare institutions as well as already established cardiovascular multi-cohort consortia, related to pertinent existing research initiatives or medical associations. These will include for example *(1) the Emerging Risk Factors Collaboration (ERFC)*, a consortium led by the University of Cambridge Cardiovascular Epidemiology Unit, that has collated and harmonised individual-participant data (IPD) of >100 prospective studies from >30 countries; *(2) Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE)* consortium, formed to facilitate genome-wide association study meta-analyses and replication opportunities among multiple large and well-phenotyped longitudinal cohort studies in the US; *(3) BigData@Heart*, a five-year IMI project investigating heart failure based on big data (see letter of interest from the Scientific Coordinator) and (4) other individual cohorts outside the EU and Canada, such as the Framingham Heart Study (FHS – see letter of support) and the Multi-Ethnic Study of Atherosclerosis (MESA), with the double goal of research collaboration and, at later end, possibly integration of their cohorts to the system.
*Partners involved: **ESC**, LYN, UKE, MCM, MCGILL, UPF. Duration: M6 – M48. Related deliverables: D6.4.*

Task 6.4 – euCanSHare's dissemination events, workshops and hands-on sessions
This task will involve the participation to and *ad hoc* organisation of dissemination events (conferences, workshops, seminars) at the presence of the relevant scientific community, industry partners and public institutions. While a number of publicly organised institutional events will be attended by partners to present the project, set up collaboration initiatives and enlarge the stakeholder network, the consortium has also planned the organisation of three project-dedicated workshops (M24, M36, M48) to be hosted within ESC congresses (with an attendance of >30,000 cardiovascular researchers/professionals every year). This will offer substantial opportunities to showcase the platform through hands-on-sessions in which participants will be able to test the functionalities of the euCanSHare platform and provide feedback on their user experience and specifications from their standpoint (*e.g.* epidemiologist, data manager, industrial R&D, public health researcher, etc). While all stakeholders will be invited to participate to the three workshops, each one will have a central theme to promote the platform within specific

contexts. Following this strategy, the first workshop will focus on cardiovascular research and cohort owners, the second on industrial exploitation of the platform, and the last one on the Knowledge-to-Action theme, *i.e.* on exploiting the platform for translating research results into actionable insights, such as through public health initiatives or policy involvement. Specific seminars and discussions will be also integrated to these workshops, in particular on Open Science/Open Cloud initiatives and ethical/legal interoperability.
*Partners involved: **ESC**, NBD, LYN, ALL. Duration: M1 - M48. Related deliverables: D6.5.*

Task 6.5 – Exploitation planning
This task will be devoted to ensuring exploitation as well as sustainability of the platform beyond the project timeframe. This activity will leverage on the preliminary stakeholder analysis carried out in T6.2 and the subsequent onboarding campaign carried out in T6.6, and will be elaborated through a series of dedicated *exploitation seminars*, to be periodically held from M18 onward within annual internal meetings, at the presence of all consortium members as well as some selected representatives of the different stakeholder categories, who will asked to provide insights in regard to their specific needs and issues. The resulting *Exploitation Plan* will be elaborated in close cooperation with all partners, representatives of stakeholders, and with support of relevant national research authorities in Canada and the EU. Also, it will be adjusted in view of additional strategic feedback gathered from external experts from industry and relevant institutions, involved in the project through the workshops from T6.2-T6.4 and convened for exploitation seminars and dedicated TCs from M24 onward. Furthermore, a sustainability plan will be put together by BSC, CRG, EMC and BBMRI, the hosts of the platform and data, who will plan in detail how ELIXIR, EGA, euro-BioImaging and BBMRI will interact beyond the duration of the project. Note that, for example, ELIXIR and euro-BioImaging have already put forward a strategy for such interaction[135], to which euCanSHare will actively contribute. Finally, note that the IPR management will be coordinated in WP7 (project management).
*Partners: **BSC**, NBD, CRG, BBMRI, EMC, UPF and all partners. Duration: month 24 – 48. Deliverables: D6.6.*

Deliverables
D6.1 Project website (LYN, M3)
D6.2 Dissemination strategy (LYN, M6)
D6.3 Dissemination channels (LYN, M12)
D6.4 Awareness campaigns and consensus building (LYN, M48)
D6.5 euCanSHare's dissemination events (ESC, M48)
D6.6 Exploitation plan (BSC, M48)

WP number	**WP7**			**Period**			M1 – M48			**Leader**			UPF			
Title	**Scientific Coordination and Project Management**															
Participants	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	UPF	MUHC	UKE	LYN	EMC	QMUL	MCGIL	KUL	BSC	CRG	ESC	BBMRI	MCM	THL	UMG	NBD
Person-months	45	1	1	2	1	1	3	1	1	1	1	0.5	0.5	1	0.5	0.5

Objectives: The goal of this WP is to establish and guarantee full synergy, motivation, integration and effective interactions among euCanSHare partners with the final aim of delivering successful progresses of the project within the agreed time, cost and quality limits as defined by the project contract signed with the EC and the Consortium Agreement signed among the participants. WP7 will a number of organisational features to accommodate the EU-Canada nature of the project, while ensuring cohesive interactions between both the IT- and data-related components of euCanSHare. This WP will monitor the on-going work and achievements of the tasks, paying special attention to the cohesion between the European and Canadian partners. As such, a Canadian Consortium Leader (CCL: MCGILL) will work closely with the Project Coordinator (UPF) to ensure suitable interactions and decision-making throughout the duration of the project. An Information Technology Committee (ITC) will oversee the consistent implementation and interoperability of the computational tools, as well as optimal communication within this diverse computational team. A Human Data Committee (HDC) will ensure that all data aspects, cross-border legal frameworks, cohort-specific policies, and clinical requirements are constantly taken into

[135] https://www.elixir-europe.org/system/files/euro-bioimaging_elixir_image_data_strategy.pdf

account in the project. The objectives of WP7 are:

- Monitor the ongoing work in order to guarantee the timely achievements of the project goals, as well as the quality and relevance of the outcomes. This is the central task of the WP.
- Take care of all the administrative and financial issues related to the project;
- Optimise partners' interaction within the Consortium, with the aim of making of it a cohesive EU-Canada team committed to the achievement of euCanSHare's objectives.
- Manage the euCanSHare implementation process, through proper allocations of tasks and responsibilities
- Foresee any risks of delay and set up mitigation plans and corrective actions
- Implement the ethical and innovation management strategy, and identify, protect and manage the Intellectual Property (IP) generated by the project and patent applications

Task 7.1 – Overall project management

UPF**,** and more specifically the Project Coordinator (PC), will be responsible for the overall management of the project, together with a project manager that provide support in the coordination, daily management and monitoring activities. A common baseline for task and responsibility allocation will be established. The **Project Manager** (PM) and the **Project Coordinator** (PC) will take measures to guarantee control, validation and verification of project results; ensure that plans are fulfilled and implement necessary corrective actions as described in the paragraph on "Description of project management structure and procedures" under section 3.2. This task will focus on:

- Ensuring the timely completion of the activities in each work package and on monitoring that activities are carried out according to the project plan.
- Providing the Consortium with quality guidelines & procedures, outlining the project's timing, the deliverables production procedures, the self-assessment procedures and the ethical procedures.
- Coordinating the preparation of the reports (official periodic report and half-yearly report) due during the project's life. This task will involve not only the Coordinator and the Project Manager, but also all the other partners.
- Organising the annual project meetings (both physical meetings and teleconferences), and setting up and maintaining the appropriate project communication infrastructure, including a document repository. euCanSHare will setup a dedicated web conferencing system, in order to allow smooth communication between partners and to minimise the needs of physical meeting.

High quality outputs are the basis for effective project implementation and successful exploitation. To ensure all outputs fulfil best possible quality standards, deliverables will be prepared by the responsible partner, controlled by the WP leader (WPL), reviewed by the PC (scientific) and PM (administrative) part and submitted to the EC by the PC. The PC will be responsible for organising the periodic project meetings where each WP status is measured against deliverable and milestone progress. Results and developments will be presented by the WPL. Should any conflict arise, the PC will consolidate arguments, and strive to find the best solution to suit all partners' interests.
*Partners involved: **UPF**, ALL. Duration: month 1 – 48. Related deliverables: D7.1.*

Task 7.2 – Strategic management of the consortium

The coordinator will be the only contact point for the EC and ensure the communication between the Consortium and the EC. He will be supported by the PM for the communication and coordination between consortium members and towards relevant stakeholders. Since euCanSHare consortium gathers European as well as Canadian partners, the monitoring of the project progress and the coordination on the daily work plays an important role.
In addition to the coordinator, there will be a **Canadian Consortium Leader** (Bartha Knoppers, MCGILL) who will represent the Canadian partners in the Executive Board meetings and in other scientific or management meetings. She will be the Canadian main contact point for the PC, she will ensure the reporting of euCanSHare activities to the PC and allow an efficient communication and information flow between European and Canadian consortium members. The consortium will be armed with appropriate data exchange tools.

The monitoring of the project progress will be made through regular meetings and conference calls. In addition to the kick-off meeting and the final meeting of the project, regular project management meetings will be organised, at least one of them in Canada. Besides, throughout the whole project duration, between two project management meetings, we will organise a conference call with WP Leaders which will allow us to continuously monitor the project's progress. These meetings and conference calls will allow to supervise the progress within each WP and foresee any risk of delay and set up any mitigation plans accordingly. With this regular and well-organised monitoring, the smooth running of the project will be guaranteed, and problems that can come out during the project lifespan will be solved as early as possible.
*Partners involved: **UPF**, ALL. Duration: month 1 – 48.*

Task 7.3 – Reporting and support of financial management & controlling
This task will handle all financial matters that should arise during the course of the project, as well as taking care of possibly needed amendments to the Grant Agreement. Following an 18-months reporting schedule, two periodic and a final report will be submitted to the EC. Their compilation will be managed by the PC. euCanSHare will implement internal progress reports to monitor progress. Individual financial records will be regularly compiled in close cooperation with the partners' Grants Offices. While the partners' Grants Offices administer their own resources, the PC coordinates collection and monitoring of periodic cost claims on consortium level, checks declared costs are appropriately justified, provides follow-up of EC payments and collects audit certificates where necessary. Deviations from planned resources use will be discussed with affected partners and brought to attention of the WP leader. Within the first half year, a webinar on H2020 financial management will be offered to familiarise all individuals involved in financial reporting process (scientists and administrators) with the relevant H2020 rules and regulations and related project internal procedures.
*Partners involved: **UPF**, ALL.. Duration: month 1 – 48. Related deliverables: D7.2*

Task 7.4 – Risk management
This task is focused on the identification, assessment, and prioritisation of potential risks with the aim of minimising, monitoring, and controlling the probability and/or the impact of unfortunate events. This will imply sharing the risks' awareness through appropriate communication, leading this way to improved decision-making. The management of risks and corrective actions will be handled for the whole duration of euCanSHare and the risk assessment will be on the agenda in every project meeting and the result of the discussions will subsequently be added to the project-related documents. We will seek to identify new and emerging risks and develop strategies on how to mitigate them. This activity will also enable the Consortium to monitor and control the overall level of risk throughout the project. A detailed description of significant risks and contingency plans is given in Section 3.2.
*Partners involved: **UPF**, ALL. Duration: month 1 – 48. Related deliverables: D7.3*

Task 7.5 – Ethical clearance and monitoring
This task aims at overseeing the alignment to all ethical requirements and ensuring that local, national and international regulations are respected. Events and issues regarding ethics will be reported in the periodic reports and the final report. The PC together with the Ethics and Leal Committee will coordinate the management of ethical issues. This comprises implementation of general rules, collection of copies of the partners' ethical authorisations and liaison on ethical issues with the European Commission and other stakeholders. Two project internal evaluations of ethical issues will be scheduled for M24 (mi-term) and M748 (end of the project).
*Partners involved: **UPF**, ALL. Duration: month 1 – 48. Related deliverables: D7.3*

Task 7.6 – Innovation and IPR management
UPF together with the Innovation and IPR Manager will implement the rules laid out in the Consortium Agreement, amongst others on the policy for publications and on access to data generated. Any innovation related decision-making by the General Assembly will be prepared. A knowledge register, containing information on partners' background IP, knowledge created in the project and associated access rights, will be set up and maintained. Moreover, plans for use of project results will be prepared, the results of which will be included in the final reports. The consortium will timely deal with management of *Intellectual Property Right (IPR)* issues, by conceiving a strategy for a satisfactory protection of both background and forthcoming scientific results, with parallel patent application to the European Patent Office (EPO) and Canadian Intellectual Property Office, meanwhile maintaining knowledge sharing and cross-fertilisation with other parallel research initiatives. Business and commercial exploitation routes will be explored by organising a Strategic Exploitation Seminar by M18. Building on past experiences and in close connection with academic and industrial stakeholders, the consortium will identify the most likely scenarios within which to position the project's expected exploitable outcomes. Individual analyses will examine the possible use of outcomes by partners according to their current business lines and isolated capabilities. A global exploitation plan will outline the full potential of the developed solutions in domains other than cardiology and will be used to depict different scenarios where commercial solutions could be developed. Joint plans will investigate the most likely market entry roads bringing together selected groups of partners (*e.g.* UPF and QMUL for the radiomics platform. The joint plans will include target customers and markets, a timeline for reaching the market, and a diffusion strategy with regard to the resulting products/services. A detailed description of exploitation and management of intellectual property is reported under Section 2.2.
*Partners involved: **UPF**, ALL. Duration: month 1 – 48.*

Deliverables

D7.1 Kick-off meeting report (UPF, M2)

D7.2 Project Handbook (UPF, M2)

D7.3 Quality assurance guidelines and ethical clearance (UPF, M4)

Table 6 – List of euCanSHare's deliverables:

N°	Deliverable name	Lead	Type	Level	Month
D1.1	Results of interview study with researchers regarding blockchain technology for data sharing	KUL	R	PU	M18
D1.2	Policy "Points to Consider" tool to guide research projects, policy makers	MCGILL	R	PU	M24
D1.3	Comparative cross-mapping table detailing which participating cohorts are compliant with euCanSHare requirements	MCGILL	R	PU	M24
D1.4	Template smart contact to streamline data access and sharing process	MCGILL	R	PU	M36
D1.5	Policy points regarding the incentives mechanisms for data sharing technologies	KUL	R	PU	M16
D2.1	Initial Infrastructure framework and documentation	BSC	R	PU	M12
D2.2	Data distribution protocols and interfaces	BSC	P	PU	M24
D2.3	Data Access Committee Portal prototype	CRG	P	PU	M36
D2.4	Mica-powered catalogue populated with euCanSHare metadata	THL	DEC	PU	M48
D2.5	Integrated euCanSHare computational infrastructure	BSC	R	PU	M48
D2.6	Final documentation module	UPF	R	PU	M48
D3.1	Data Management Plan v1	CRG	R	PU	M6
D3.2	Data Management Plan v2	BSC	R	PU	M12
D3.3	Recommended metadata standards	EMC	R	PU	M18
D3.4	Guidelines and protocol for data deposition	EMC	R	PU	M24
D3.5	Report on the implementation of raw data access protocols	CRG	R	PU	M36
D3.6	Prototype of blockchain technology	LYN	DEM	PU	M48
D4.1	Opal software integration to to euCanSHare	MUHC	DEC	PU	M12
D4.2	Data quality control tool	UMG	DEC	PU	M18
D4.3	Cardiac image analyser	UPF	DEC	PU	M24
D4.4	Bioinformatics toolbox	UKE	DEC	PU	M24
D4.5	Machine learning toolbox	UPF	DEC	PU	M36
D5.1	Harmonised data for use cases	MUHC	R	CO	M24
D5.2	Results on feasibility testing for classical cardiovascular risk factors and biomarkers	UKE	R	CO	M24
D5.3	Results of diabetic cardiomyopathy multi-country study	QMUL	R	CO	M42
D5.4	Results of validation of joint improved risk prediction on cardiovascular events	UKE	R	CO	M42
D5.5	Results of public health study	THL	R	PU	M42
D5.6	Results of target discovery study	NBD	R	PU	M48
D5.7	Feedback report on assessment of platform usage	UKE	R	PU	M48
D6.1	Project website	LYN	R	CO	M3
D6.2	Dissemination strategy plan	LYN	R	CO	M6
D6.3	Dissemination lists and channels	LYN	R	PU	M12
D6.4	Awareness campaigns and consensus building	LYN	R	PU	M48
D6.5	euCanSHare's dissemination events	ESC	R	PU	M48
D6.6	Final exploitation plan	BSC	R	CO	M48
D7.1	Kick-off meeting report	UPF	R	CO	M2
D7.2	Project Handbook	UPF	R	CO	M2
D7.3	Quality assurance guidelines and ethical clearance	UPF	R	CO	M4

3.2 Management structure, milestones and procedures

Organisational structure and decision-making: The euCanSHare management structure and procedures are designed to guarantee a smooth and effective cooperation between partners, and to ensure the implementation phase towards the achievement of the overall objectives of the project. The euCanSHare management structure is represented in Figure 9, focusing on four primary managerial tasks:

- **Decision-making**, implemented by the Governing Board.
- **Scientific and technical coordination**, performed by two distinct committees (see below).
- **Operational management**, performed by the Executive Board.
- **Advisory**, through (1) the Dissemination and Users' Board; (2) the Ethical & Legal Manager; (3) the Innovation & IPR Manager; (4) the External Advisory Board and the (5) the Gender Commissioner.

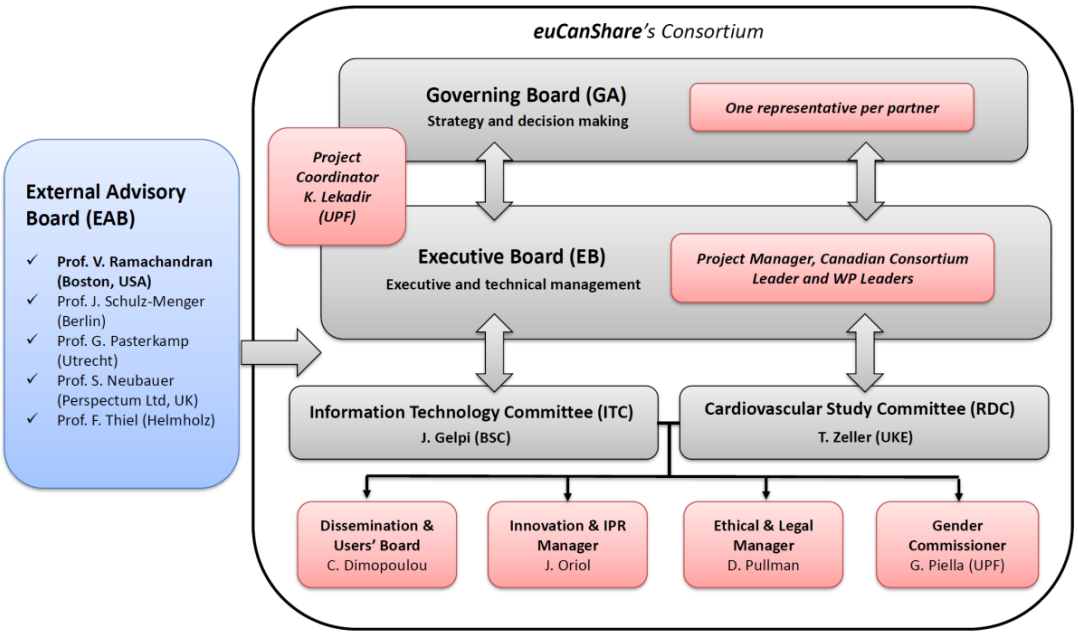


***Figure 9** – euCanSHare's governance structure.*

The **Project Coordinator (PC)**, Karim Lekadir (UPF), will act as the intermediary between the Consortium and the European Commission and will be responsible for communicating all information about events or circumstances likely to affect or delay the implementation of the project (as defined in the GA Article 17). He is ultimately in charge of safe-guarding the interests of the project, ensuring its smooth implementation, monitoring compliance by the partners with their obligations and representing the project towards external stakeholders and networks. The PC will chair the Governing Board and the Executive Board and will be responsible for taking all actions to enable proper decision-making by these management bodies, thus ensuring smooth operation of the project: work plan maintenance, project progress monitoring, identification of potential problems and possible consequences for future research, submission of all required deliverables, progress reports and financial statements to the EC (Article 19 and 20), for transferring payments to the partners (Article 21) and for overseeing any ethical, gender related and societal issue that may arise in the course of project implementation (Article 33 and 34). The PC, with the support of the **Dissemination and Users' Board**, will supervise the communication and dissemination activities of the project.

The PC will be also supported in the administrative issues by the **Project Manager (PM)** in order to make appropriate internal arrangements to ensure the efficient implementation of the project; ensure that the EC administrative guide lines are correctly followed by the participants; coordinate the cost statements and the administrative bodies of the different participants' institutions; collect all the detailed data requested by the Commission for the proper administration of the project; coordinate the preparation of periodic reports; follow-up work plan scheduling; support in the resolution of any administrative or contractual issues within the partnership and with the Commission (Amendments to the Grant Agreement); follow legal issues, among others. The PM will also support the consortium in the Dissemination and Training Activities, developing and implementing plans for the communication and the visibility of the consortium; organising project meetings and a final workshop.

In addition to the project manager, there will also be a **Canadian Consortium Leader**, represented by Bartha Knoppers (MCGILL), with specific responsibilities in the organisational structure and decision-making mechanisms in order to allow a proper and efficient implementation of this EU-Canada consortium. She will ensure on behalf of the Canadian partners that the work is delivered in due time and manage all operational aspects with Canadian members, as well as ensure central coordination of reports and deliverables of the consortium as a whole, supporting the PC and PM, and checking that payments to the respective Canadian partners are according to the terms and conditions established in the Grant Agreement.

The **Governing Board (GB)** is the highest level of management in the project. It is the Consortium's main decision-making and arbitration body, and it is chaired by the project coordinator composed of one representative per partner. The Governing Board will address only the high-level/strategic issues of the project (such as approving re-allocation of the project's budget; discussing and approving requests for major changes proposed to the description of work and proposal of amendment, with subsequent changes in the project work plan; requesting contractual changes to the European Commission; proposing and approving resolutions of critical issues and conflicts, etc.). Each partner has one vote and a relative majority system will be employed, defined in an appropriate Consortium Agreement. Given its nature of high-level decision-making body, assigned also to formal tasks (affected by their relevant bureaucratic procedures), the Governing Board will convene on a yearly basis, but summoned if required at other times.

The project will be scientifically coordinated by two steering committees, with complementary focuses and responsibilities, under the guidance of the Project Coordinator. Both groups will be responsible for setting up and updating place operational frameworks, tools and processes throughout the project.

Information Technology Committee (ITC): Led by BSC (Josep Gelpi), lead of ELIXIR-ES, with support from CRG (EGA), EMC (euro-BioImaging), UKE, UPF and UMG, the ITC will:

- Collect, analyse and document system's requirements (WP2-3).
- Define data management processes in collaboration with the HDC (WP3).
- Defining analytics and statistical requirements for the data analysis platform (WP4).
- Oversee the use cases from statistical and technical standpoints (WP5).
- Organise demonstrations of the computational tools at various intermediate stages and provide quantified updates on their progression (WP6).
- Update the planning of the technical activities based on the feedback received at intermediate validations (WP2-4).

Cardiovascular Study Committee (CDC): This group will be led by UKE (Tanja Zeller), with the participation of THL, MCM, MCM, KUL and QMUL. Its goals are to:

- Ensure data policies and ethical/legal frameworks are strictly respected by the entire consortium (WP1).
- Coordinate the data cataloguing and deposition activities (WP2-3)
- Define requirements for the precise definition of the data analysis capabilities (WP4)
- Oversee the execution of the use cases (WP5), ensuring same standards across all tasks and suitable interactions between the EU and Canadian cohorts.
- Gather and analyse feedback from the different users and communicate recommendations to the ITC.
- Participate to the awareness campaigns, in particularly for attracting new cohorts to the platform.

Finally, the **Executive Board (EB)** acts as the operational body of the Consortium, ensuring day-to-day management and coordination. It is chaired by the Project Coordinator, who is supported by the Project Manager and by all the WPs leaders in the project. Each WP Leader will be responsible for coordinating all activities relating to the objectives and implementation of his/her WP. They will take all operational decisions regarding the WP day-to-day management, in collaboration with the Task Leaders of their WP, on the basis of a detailed definition of milestones, scope and expected results of the WP, as well as being charged with the responsibility of taking all the necessary means to reach the objectives. The WPL will report on the WP activity to the Coordinator. They will also be in charge of addressing and documenting internal risks which may impair progress towards the objectives of the WP and suggesting strategies to anticipate and minimise such internal risks. The WPL will also be responsible for implementing the decisions agreed by the EB, controlling the execution of the project in line with its agreed Work Plan (see Table 3.2a for the Milestones), and monitoring corrective actions.

Advisory boards/committees:

1) The **Dissemination and Users' Board** (ESC: Christina Dimopoulou) will take care of establishing a strong interaction with the relevant stakeholders to promote the usage of the developed tools. The group will be supported

by the BSC, UKE and UPF, as well as by partners with links to industrial players (LYN, NBD). Specifically, the Users' board will:

- Increase the awareness about the euCanSHare platform, the developed technologies and their possible impact on cardiovascular research, as well as innovation in the private sector.
- Facilitate the organisation of workshops and other dissemination events with the involvement of the end-users.
- Enable the creation of a community of end-users.

2) The **Ethical & Legal Advisor (**Prof. Dary Pullmanm, Memorial University, Canada: expert on biomedical ethics) will ensure the ethical clearance of all the project's activities and their adherence to the relevant European regulations. The Committee will be chaired by an independent expert, who will be in charge of supporting the partners, reviewing ethics supporting documents (informed consent forms, information sheets, agreement number, authorisation) to ensure that they are in line with the European and national rules, as well as informing the partners in any update of these. Ethical aspects will encompass the protection of research subjects, including consideration of study risks, burdens and benefits; recruitment, information and consent processes – including frameworks for the withdrawal of consent and data; questions of anonymisation and identifiability in relation to patient-specific models; and the development of a framework for responsible data sharing and re-use – including questions of transparency and accountability of data use. The ethics manager will also respond to any other ethical issues that may arise in the course of the project and may also seek the support of the independent ethics advisers to be selected during the first meeting of the Governing Board, together with the other members of the Committee.

3) The **Innovation & IP Management Committee** (IPC: led by NBD – Josep Oriol) will adopt appropriate innovation management techniques to ensure a successful exploitation of the project results. To this end, this IPC will coordinate different activities that include:

- Periodic brainstorming to evaluate the advancements of the project and recent novelties in the market, thus contributing in the definition of the exploitation strategy.
- The most interesting perspectives coming from the brainstorming sessions will be further explored and discussed, also with the support of the External Advisory Committee.
- Facilitating knowledge exchange among partners, acting as knowledge broker and thus ensuring an appropriate level of trust within the consortium.
- Support for the definition of the IPR's management policies.
- Management of the innovative outcomes and preparing the ground for future exploitation initiatives.
- Conducting preliminary market analysis, which will feed the exploitation plans.
- Facilitating discussions among partners on these topics and organising dedicated strategic exploitation seminars.

4) The **External Advisory Board (EAB)** will be led by Prof. Vasan Ramachandran, Director of the renewed Framingham Heart Study (see letter of confirmation in the Annex). He will be supported by a highly qualified team of international experts in all the fields of relevance to euCanSHare as listed in the table below (note that all these members have confirmed their approval to join the EAB by e-mail):

Name	Institution	Country	Expertise
Prof. Vasan Ramachandran	Boston University School of Medicine	USA	Director of the Framingham Heart Study
Prof. Gerard Pasterkamp	Utrecht Medical Centre	Netherlands	Cardiovascular Epidemiology
Prof. Jeanette Schulz-Menger	Medical University Charité Berlin	Germany	Cardiovascular Imaging
Prof. Fabian Thiel	Helmholz Centre Munich	Germany	Bioinformatics
Prof. Stefan Neubauer	Perspectum Diagnostics Ltd	UK	IPR, biomedical entrepreneurship

If needed, the EAB will invite other experts to join the board (*e.g.* from a patient association or relevant network such as the European Heart Network). The mission of the EAB will be to provide independent advice on the project's scientific activities, as well as feedback and evaluation regarding the technical robustness and usability of the implemented solutions. Its members will annually perform a project review of the project outcomes during the annual internal review meeting.

5) The **Gender Commissioner** (Gemma Piella, UPF) will be responsible for monitoring and ensuring that the consortium addresses gender equality issues, providing requests and suggestions to the Coordinator. In particular, the Gender Commissioner will: 1) track team diversity, mapping the different teams' sex ratios periodically; 2)

encourage women to take on more responsibilities; 3) ensure equal tasks' distribution among male and female partners, and finally 4) supervising equal representation in publications and events participation. Corrective actions will be immediately proposed in case of misrepresentation issues and their implementation monitored.

Coordination and monitoring activities: To secure the achievement of the project objectives, euCanSHare has deemed appropriate to distinguish different procedures for the different activities performed during the project. For managing the overall project activities, including also the coordination of the implementation effort, a classic approach will be adopted, ensuring smooth cooperation, continuous monitoring, periodic assessment of the results and compliance with the reporting obligations to the EC. Coordination and monitoring activities within the consortium include meetings organisation, reporting and financial management. During the 4 years of the project, several meetings will be organised:

- **Project Management Meetings**: These meetings will be organised in order to assess the project progress, foresee any risks of delay, and take necessary measures and decisions (see Figure 10). These meetings will gather the Project Coordinator, Project Manager, Canadian Consortium Leader, at least one representative of each partner and WP Leaders. At these meetings, the progress of each WP will be presented by WP Leaders and involved partners. Nay issues and possible deviation will be discussed. Related contingency plans will be defined and validated during the meetings. Financial issues will also be followed and discussed to allow early identification of any significant deviation from the allocated budget. **At least one of these meetings will be held in Canada**. Due to the time difference between Canada and Europe, we will use a meeting planner in order to schedule meetings across different time zones and find the best time for the participants.
- **Intermediate follow-up meetings (conference calls):** Between physical meetings, conference calls gathering WP Leaders, the PC, the PM and the Canadian Consortium Leader will be organised on a 3-monthly basis. These conference calls will allow for a continuous monitoring of the project's progress, as well as any problem or risk of delay that can occur.

This approach will result in a close cooperation among the Executive Board and the technical committees (HDC and ITC), in order to ensure smooth cooperation among partners and sufficient coordination of the technical efforts. The Governing board will be convened at least once a year, and will mainly serve as assessment and evaluation body, supported by the External advisory board.

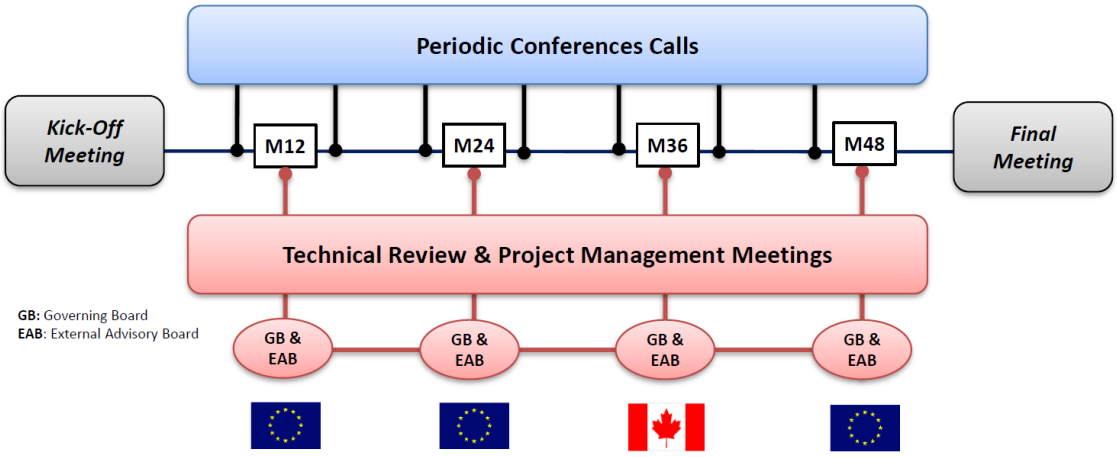


***Figure 10** – Overview of euCanSHare internal communication plan and meetings. One of the four Annual Meetings will take place in Canada.*

Table 7 – List of milestones, including the periodic releases of the euCanSHare platform:

N°	Milestone name	WP	Due date	Means of verification
MS1	euCanSHare's Data Management Plan is drafted	3	M9	The data management plan is published
MS2	Basic framework of the euCanSHare platform is operational – **euCanSHare v1**	1-4	M12	New components can be now added to the computational infrastructure
MS3	The data analyser module is integrated into the webportal	4	M18	The data analyser is preliminarily tested on the available cohorts in WP5

MS4	The ELSI information for all cohorts are collected and harmonised	1,2	M24	The ELSI information are now entered into euCanSHare's catalogue
MS5	euCanSHare's catalogue is implemented	2,3	M24	The catalogue is populated with initial EU and Canadian cohorts from Table 1
MS6	The data is fully stored in EGA, euro-BioImaging and BBMRI	3	M24	Data can be accessed by authorised users
MS7	The first prototype of the euCanShare platform is released – **euCanSHare v2**	1-6	M24	The pilot-test studies in WP5 and the hands-one-sessions in WP6 can be fully initiated
MS8	An update of the euCanSHare platform is released – **euCanSHare v3**	1-6	M36	The researchers in WP5 and demonstrators in WP6 receive an updated version with their initial feedback implemented
MS9	The final euCanSHare platform is released – **euCanSHare v4**	2-4	M48	The platform is fully operational online and can used by third-party researchers
MS10	The multi-country use cases are investigated	5	M48	Results are disseminated in high impact journals

Table 8 – Critical risks for implementation:

Description of risk	WPs	Proposed risk-mitigation measures
New EU data regulations make transfer of information and data more challenging (Medium).	1	We analyse the feasibility of information/data transfer right from the start of WP1 to anticipate such issues.
Platform catalogue is more fragmented than initially thought, *i.e.* variables do not overlap enough between cohorts (Low).	2	We will agree on a minimal set of variables to be harmonised for selecting the eligible cohorts for cardiovascular research.
Some of the protocols/tools cannot easily be integrated (Medium).	2	Alternative open-source tools/protocols will be tested.
Difficulties in the harmonisation of metadata from different data repositories (Medium).	3	We will set specific Working Groups with representatives of data repositories (EGA, BBMRI, euro-BioImaging).
Unbalanced nature of the data in the available cohorts (lack of specific types of data or over representations of others) (High).	3,4	We will define recommendations on data composition and possibilities of integration for building new cohorts.
Image quality control is more challenging than expected (Low).	4	We will switch from quality on the raw data to conventional quality control of the output of the image analysis (clinical indices, imaging biomarkers, radiomics) by identifying irregularities in those results.
The number of machine learning techniques requested by the consortium members for integration is too high (Low).	4	We will select 5-6 major techniques that are most commonly used, such as support vector machine and principal component analysis.
Lack of interest from other cohorts to join the euCanSHare platform (Medium).	3,4	More than 35 cohorts have already agreed by e-mail to join euCanSHare. Extensive outreach efforts to demonstrate the added-value will be initiated.
Interoperability of different tools lower than expected (Medium).	3,4,5	Defining joint standards to make necessary adaptions.
Pooled datasets not large enough for specific subgroups *e.g.* diabetic cardiomyopathy (High)	3,5	We will perform extensive outreach to recruit clinical cohorts with more clinical cases.
Heterogeneity of the imaging data too large for radiomics (Low).	4	We will perform an extensive analyse to detect robust radiomics features for subsequent association studies.
Feasibility use case studies in WP5 (Troponin) do not result in the expected findings (Medium).	5	We will check harmonisation performed.

Definitions of CVD in different cohorts is too heterogeneous (Low).	2-5	Extensive harmonisation and feedback with experts will be performed. Cohorts will be asked to re-define data.
Lack of interest of clinical practitioners in the outcome of the project (Low)	5,6	Although this is not a direct expected outcome of the project, outreach activities dedicated to the clinical community will be intensified.

3.3 Consortium as a whole

Consortium expertise and complementarity within euCanSHare:
Due to the multi-component nature of the euCanSHare project, as well as the challenges to integrate, share and exploit multi-cohort cardiovascular data, we have carefully selected a consortium composed of **13 EU and 3 Canadian experts from a total of nine countries**, covering different geographical regions and with world-renowned experts in areas of direct relevance, while ensuring complementary, non-redundant, and essential contributions to the project. Strategically, we chose a relatively large consortium (13 EU partners, 3 Canadian partners) to cover the whole spectrum of needed expertise to build a comprehensive scalable platform, as well as to create a large network of collaborators that will further extend through each member's own network and collaboration channels, ultimately leading rapidly to a large community of users and critical mass within the euCanSHare platforms.

A number of criteria were considered to select the consortium members during the preparation of this proposal, *i.e.:*

• Promoters of established multi-cohort cardiovascular initiatives, providing access/links to major cardiovascular cohorts: MONICA and MORGAM: **THL**; BiomarCaRE and BigData@Heart: **UKE**; Canadian Alliance CAHHM: **MCM**.
• Developers of the main IT infrastructures in Europe, *i.e.* ELIXIR, EGA, euro-BioImaging and BBMRI → **BSC, CRG, EMC and BBMRI**, respectively.
• Ownership of tools as well as expertise in cloud-based data distribution and transfer (FTP servers, Aspera or REST APIs as part of ELIXIR: **BSC, CRG**) for linking the above infrastructures.
• Experience and access to tools for building rich data catalogues (Mica: **MUHC** from Canada, **THL** from Europe, **EMC** for the imaging catalogue).
• Tools for automated data curation: Data harmonisation tool Opal (**MUHC**); quality control tool Square^2 (**UMG**).
• Ownership of cardiovascular data analysis and machine learning pipelines (the Rocket platform) developed in cardiac-related EU project (euHeart, VP2HF, CardioFuncXion, NEUBIAS): **UPF.**
• Ownership of image analysis and imaging management tools for automatic estimation, storage and management of cardiac imaging indices and radiomics: **UPF, EMC, QMUL.**
• Expertise in bioinformatics tools and cardiac -omics research (**UKE, UMG, NBD, BSC, CRG**).
• Experience with big data in cardiology: **UPF, QMUL** (both as part of the UK Biobank); **UKE, THL** (both in BiomarCaRE >300,000 cases; **UKE** currently part of the BigData@Heart IMI project).
• For pilot-testing of the platform: Expertise in cardiac risk/disease quantification: **UKE, UMG, THL, QMUL, UPF**; cardiovascular epidemiology: **QMUL, MCM, UKE**; public health research in cardiology: **THL;** target discovery and drug development: **NBD** and **BSC**.
• World-renewed experts in biomedical ethics and data protection from both Canada and the EU for interoperability analysis: (1) Centre for Genomics and Policy: **MCGILL** from Canada: promoters of major guidelines such as GA4GH; (2) Centre for Biomedical Ethics and Law at **KUL:** members of GENEBANC and ENGAGE; (3) **LYN**: Coordinators of the MyHealth-MyData project analysing the GDPR framework in healthcare).
• Developers of new solutions for enhancing Open Science through Smart Contracts and blockchain approaches (**MCGILL**, **KUL, LYN**).
• Experts in dissemination and communication in the field of healthcare (**LYN**) and cardiovascular personalised cardiovascular medicine (**ESC)**.

Additionally, an important selection criterion not mentioned above but equally important was to ensure that synergies and past collaborations already exist between the different players to make the consortium tightly knitted and closely interlinked from day one of the project. In this regard, the euCanSHare consortium is remarkable as not only such connections are strong within each side of the Atlantic, but also between the Canadian and EU partners. The EU-Canada past interactions include: QMUL and MUHC witrhin the SCMR (Society of Cardiovascular Magnetic Resonance) Board of Trustess and the UK Biobank CMR Expert Groups, MCM and UKE as part of BiomarCaRE project (2011-2016); MCGILL and KUL within the Ethics Panel of the World Anti-Doping Agency (WADA: chaired by MCGILL); MCGILL, MUHC and BBMRI as part of the BioSHaRE-EU project (2010-2015). The Canadian partners have a long history of collaborations among them: All three institutions have worked together on the build-up of the Canadian Alliance of Healthy Hearts and Minds (CAHHM), as well as in other multi-study epidemiology initiatives in Canada (*e.g.* CPTP: Canadian Partnership for Tomorrow Project).

The European members also have had many collaborations over the last 5 years. QMUL and UPF are co-PIs and close collaborators in the UK Biobank 2964 Study; BSC, CRG and UPF are close collaborators in several national projects in Spain; BSC and CRG have worked together in various EU projects including ELIXIR-ACCELERATE; BSC, BBMRI, CRG and THL were members of the BBMRI-LPC EU project (2013-2017); BSC and BBMRI are part of the EOSCpilot EU project (2017-2018); UKE, THL, UMG have collaborated in the BiomarCaRE project (2011-2016); UKE and ESC are currently co-members of the ongoing BigData@Heart IMI initiative; LYN and QMUL are currently collaborating within the MyHealth-MyData project; NBD is a young spin-off SME of BSC (since 2015); and finally EMC (euro-BioImaging) and BBMRI are collaborators within the BBMRI Network.

Industrial/commercial/exploitation partners: Our first SME partner LYN has long experience in cardiovascular and data management EU projects, for example as the Scientific Coordinator of the CardioProof EU project (www.cardioproof.com) and currently of the MyHealth-MyData project. They will support euCanSHare with all the blockchain related activities, as well as through its long experience with dissemination and exploitation of health and cardiovascular related projects through its links to industries and public agencies. Our second SME partner NBD is specialised in *in-silico* tests for target/drug discovery and thus its customers are mostly pharmaceutical companies who request technological support for preparing clinical trials and identifying through suitable targets. Its role in euCanSHare will be to pilot-test the platform from an industrial perspective and provide feedback and recommendations to enhance the design/development such that the platform becomes a resource for both academia and industry. Through its network of industrial and pharmaceutical clients, NBD will also support the dissemination and exploitation activities, together with LYN, to ensure wide outreach to industrial stakeholders. Finally, it is worth mentioning the important role of the ESC for disseminating and communicating the results of the project, in particular the euCanSHare platform, to a wide range of stakeholders in the cardiology domain (see WP6).

3.4 Resources to be committed

Summary of staff effort and budgets:

Partner	WP1	WP2	WP3	WP4	WP5	WP6	WP7	Total PM
UPF	0,5	15,0	0,5	**58,0**	24,0	3,0	**45,0**	146,0
MUHC	0,0	12,0	8,0	12,0	8,0	3,0	1,0	44,0
UKE	0,5	7,5	1,0	24,0	**42,0**	1,0	1,0	77,0
LYN	12,0	12,0	9,0	0,0	0,0	18,0	2,0	53,0
EMC	0,5	8,0	30,0	2,0	0,0	0,5	1,0	42,0
QMUL	0,5	0,0	0,0	2,0	36,0	1,0	1,0	40,5
MCGILL	**22,0**	0,0	0,0	0,0	0,0	1,0	3,0	26,0
KUL	20,0	0,0	0,0	0,0	0,0	4,0	1,0	25,0
BSC	0,0	**60,0**	17,0	2,0	2,0	1,0	1,0	83,0
CRG	1,0	30,0	**35,0**	1,0	1,0	1,0	1,0	70,0
ESC	6,0	0,0	0,0	0,0	0,0	**18,0**	1,0	25,0
BBMRI	0,5	0,0	4,0	0,0	0,0	3,0	0,5	8,0
MCM	0,5	10,0	0,5	0,0	10,0	2,5	0,5	24,0
THL	1,0	22,0	10,0	18,0	24,5	1,0	1,0	77,5
UMG	0,0	8,0	0,5	30,0	8,0	1,0	0,5	48,0
NBD	0,0	0,0	0,0	0,0	24,0	4,0	0,5	28,5
Total PM	65,0	184,5	115,5	149,0	179,5	63,0	61,0	**817,50**

Budget distribution and roles of euCanSHare members: To implement the centralised multi-functionality web-portal together with its legal framework, as well as to perform all the pilot-test studies and disseminate/exploit the results, the European partners of euCanSHare are requesting a total of **5.3 million Euros** to the European Commission (out of a total Budget of 6.0 million Euros including Canadian contributions). This budget can be considered as relatively modest given the planned achievements; the cardiovascular multi-study catalogue representing alone a great leap for the cardiovascular research community. This is made possible thanks to the availability of a solid technological basis and expertise for assembling the platform (see existing tools and assets in Table 3), which means the consortium will focus on leveraging existing experiences (*e.g.* BiomarCaRE, CAHHM) and technologies for most of the project, and adapting them to the needs of multi-cohort cardiovascular research, including for linking -omics and imaging repositories and enabling interoperability between the infrastructures and tools (EGA, euro-BioImaging, Rocket, etc).

The budget has been subdivided in a highly balanced fashion to ensure appropriate contributions are made from each member of the consortium to the overall design and implementation, as follows:

- Nine EU members out of 13 have been allocated approx. between 350,000 and 600,000 Euros to support their research activities, in this ascending order (UMG, CRG, EMC, BSC, QMUL, LYN, UKE, UPF, THL). Specifically, these members will contribute to three major activities, namely (1) IT integration of the platform, tools and infrastructures (UPF, CRG, BSC, EMC), (2) cardiovascular catalogue development (THL, UKE, EMC, QMUL) and (3) iterative pilot-testing with use case studies (UKE, THL and QMUL).
- The remaining four EU members (BBMRI, NBD, ESC, KUL) have been allocated between 100,000 and 350,000 Euros to support quality assurance of the bio-samples (BBMRI), testing of the platform by an SME (NBD), dissemination of euCanSHare to cardiology related stakeholders (ESC), ethical/legal analysis Open Science activities (KUL).
- UPF has been budgeted an additional 260,000 Euros for the management and communication activities.
- The Canadian partners are applying for an additional 0.64 million Euros (about 10% of the total budget) to the CIHR (Canadian Institute of Health Research) to support three key activities, *i.e.* (1) ethical/legal interoperability analysis (MCGILL), (2) deposition of CAHMM's metadata and data to the platform (MCM) and (3) adapting Maelstrom's data cataloguing and harmonisation tools for euCanSHare (MUHC).

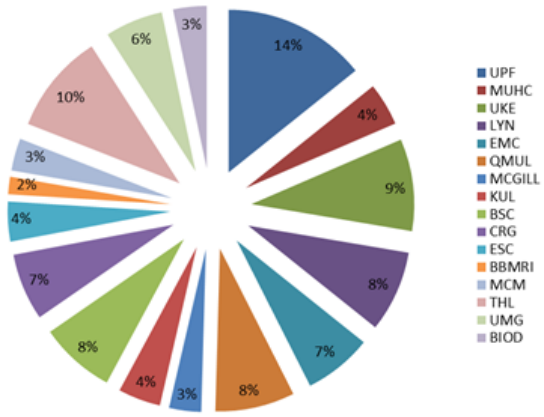


Figure 11 – Budget distribution per partners. It can be seen that the funding is well distributed across partners, with a maximum of 14% allocated to the Scientific Coordinators UPF (*i.e.* 850,000 Euros in total including research and management budget).

'Other direct cost' items (travel, equipment, other goods and services, large research infrastructure):

P11/ESC	Cost (€)	Justification
Travel	16.500	Travel to project meetings, WP meetings and conferences. Travel, subsistence and accommodation for an invited speaker.
Other goods and services	20.000	Production of communication materials, catering for workshops.
Total	36.500	

P12/BBMRI	Cost (€)	Justification
Travel	13.500	Travel to project meetings, WP meetings and conferences
Other goods and services	5.000	conference fees and publications
Total	18.500	