Description of any significant infrastructure and/or major items of technical equipment, relevant to the proposed work (if applicable):

ELIXIR unites Europe's leading life science organisations in managing and safeguarding the increasing volume of data being generated by publicly funded research. It coordinates, integrates and sustains bioinformatics resources across its member states and enables users in academia and industry to access services that are vital for their research. The ELIXIR distributed and virtual infrastructure brings together the most relevant national bioinformatics resources of 22 countries plus EMBL-EBI that are made accessible as ELIXIR services (https://www.elixir-europe.org/services) for Life Science Scientist in Academia and Industry. Currently ELIXIR provides access (via the ELIXIR Nodes) to more than 140 services that are periodically reviewed by external experts and that ELIXIR Nodes have committed to sustain long term. Furthermore, ELIXIR Services are contextualised organised around through alignment with the 5 ELIXIR Platforms (Data, Compute, Tools, Interoperability and Training). Finally, ELIXIR has robust mechanisms to allocate certain Services to Key Service Collections In addition key collection of services have been identified including: (ELIXIR Core Data Resources, ELIXIR Deposition Databases and ELIXIR Recommended Interoperability Resources) being that are recognised by public funders as recommended services for Life Sciences.

EMBL-EBI manages large-scale biological databases, which are available to users via web services available 24/7/365. The demand for rapid access to all publicly available biological databases is constantly growing, as is the volume of biological information held within the databases. To support these needs, EMBL-EBI manages an infrastructure of extensive, high-performance compute infrastructure along with large data-storage farms.

EMBL-EBI's state-of-the-art technical architecture is secure and robust and is distributed in three discrete data centres in different geographical locations to assure long-term security. This gives our data very high protection through redundancy and provides sufficient capacity and reserve to ensure our management of the rising influx of data and compute requests. As of spring 2018 the main compute farm has 34,000 cores (27,000 high throughput and 7,000 high performance) and the installed disk-based storage capacity is above 200 Petabytes. The compute farms are broadly split into three distinct functional areas, namely, (a) input and processing- for curation and processing of data and storage in a variety of ways including flat file and Oracle and MySQL databases; (b) public services- providing services to clients, including web, database and ftp services, served from our off-site Tier III plus data centres; and (c) sequencing- repository of sequencing information, whereby contributing organizations deposit information that is recorded and stored. The internal network has a 100 Gigabit backbone within its data centres and multiple 10 Gigabit connections between data centres and most servers are in these data centres are connected by at least 10 Gigabit networks. EMBL-EBI has two independent 10 Gigabit physical uplink from the data centres to Janet, Internet2 and Geant (the UK, pan-American and pan-European research networks, respectively). The bulk (~80%) of storage deployed at EMBL-EBI is based on scale-out NAS solutions. The remainder of EMBL-EBI's storage is based on SAN and traditional NAS solutions. The internal network has a 100 Gigabit backbone within its data centres and multiple 10 Gigabit connections between data centres and most servers are in these data centres are connected by at least 10 Gigabit networks. EMBL-EBI has two independent 10 Gigabit physical uplink from the data centres to Janet, Internet2 and Geant (the UK, pan-American and pan-European research networks, respectively).

Part of EMBL-EBI's core mission is to maintain large life science archives hosting critical reference data for research. This includes ENSEMBL and genome archives; the EGA for human genomes and UniProt.

The ELIXIR Hub is charged with coordination activities, whilst the Nodes provide the service delivery of the key bioinformatics resources. The largest ELIXIR Node is EMBL-EBI and delivers some of Europe's key bioinformatics services, which are described above.

Further information can be found at: www.ebi.ac.uk

Participant	**2 VIB VZW. VIB (ELIXIR-BE)**	**Belgium**
PI	[redacted] **(M)**	
ELIXIR Node	**ELIXIR-BE**	

Description of the legal entity/department/laboratory/group:
VIB is a life sciences research institute (http://www.vib.be/en/Pages/default.aspx), based in Flanders, Belgium. VIB performs internationally leading fundamental research, and has a strong focus on translating scientific results into pharmaceutical, agricultural and industrial applications. VIB is organized as a decentralized virtual institute composed of specialized Research Centers that each operates in a close partnership with five Flemish universities – Ghent University, KU Leuven, University of Antwerp, Vrije Universiteit Brussel and Hasselt University. VIB unites the expertise of 75 research groups employing more than 1470 scientists from over 60 countries. In addition to the excellently equipped individual research centres, numerous core facilities support the scientists with specialized equipment and know-how, e.g. for microscopic imaging, proteomics, metabolomics, chemical screening, bioinformatics, and many more (http://www.vib.be/en/research/services/Pages/default.aspx). VIB is offering a broad range of training courses for Ph.D. students and postdocs (http://www.vib.be/en/training/Pages/TrainingCategories.aspx) in emerging research fields and cutting-edge technologies. Next to scientific education, VIB also provides personal and professional development coaching and courses to enhance transferable soft skills. VIB was awarded with the 'Excellence in HR' label by the EU, with a strong focus in coaching and mentoring.

Main tasks in the project per WP;

VIB will contribute to the following tasks in **WP1 Expert network:**
- T1.1 Network of data managers and scientific best practice (Participant).
- T1.2 Models for brokering data to ELIXIR Deposition Databases (Participant).

VIB will contribute to the following tasks in **WP2 Training and Capacity Building:**
- T2.1 Identify training needs and solutions in Data Management and Stewardship (Participant).
- T2.2 Develop best practices guidelines and training materials in DMS (Participant).
- T2.3 Capacity Building in Data Management and Stewardship (Participant).

VIB will contribute (**WP Leader**) to the following tasks in **WP3 Common Data Management Toolkit:**
- T3.1 Establish a Starter Toolkit (Participant).
- T3.2 Processes for enriching, maintaining and sustaining the Toolkit strators (Participant).
- T3.3 Access portal to Toolkit tailored to stakeholders (Participant).
- T3.4 Best Practices and training (Participant).

Short Profile of key Staff Members who will be undertaking the work

[redacted]

[redacted]

List of the top 5 most relevant publications and/or products, services (including widely-used datasets or software), or other achievements

1. Van Bel M, Diels T, Vancaester E, Kreft L, Botzki A, Van de Peer Y, Coppens F, Vandepoele K. (2018). PLAZA 4.0: an integrative resource for functional, evolutionary and comparative plant genomics, Nucleic Acids Research, 46(D1):D1190-D1196
2. Miettinen K, Iñigo S, Kreft L, Pollier J, De Bo C, Botzki A, Coppens F, Bak S, Goossens A. (2018) The TriForC database: a comprehensive up-to-date resource of plant triterpene biosynthesis, Nucleic Acids Research, 46(D1):D586-D594
3. Coppens, F., Wuyts, N., Inze, D. and Dhondt, S. (2017). Unlocking the potential of plant phenotyping data through integration and data-driven approaches, Current Opinion in Systems Biology, 4, 58 – 63
4. Kreft, L., Botzki, A., Coppens, F., Vandepoele, K. and Van Bel, M. (2017). PhyD3: a phylogenetic tree viewer with extended phyloxml support for functional genomics data visualization, Bioinformatics (Oxford, England), 33, 2946—2947
5. Van Bel, M. and Coppens, F. (2017). Exploring plant co-expression and gene-gene interactions with CORNET 3.0, Methods in molecular biology (Clifton, N.J.), 1533, 201—212

List of <u>top 5 most relevant</u> relevant previous projects or activities, connected to the subject of this proposal
1. EU H2020 (H2020-INFRADEV-1-2015-1) ELIXIR-EXCELERATE: Fast-track ELIXIR implementation and drive early user exploitation across the life sciences, Grant number: 676559, 2015-2019. 2. EU H2020 (H2020-INFRAEOSC-2018-2) EOSC-Life: Providing an open collaborative space for digital biology in Europe, Grant number: 824087, 2019-2023
Description of any significant infrastructure and/or major items of technical equipment, relevant to the proposed work (if applicable): For almost two decades, the VIB-UGent Center for Plant Systems Biology has been investing to enable the computational analyses as well as hosting e.g. our phenotyping and NGS data. We have an established professional data center within the department comprising a computing cluster (482 nodes, 3TB RAM), mysql servers, web application servers and circa 500TB tiered storage. This infrastructure is maintained and further developed by the dedicated IT team of the Center of Plant Systems Biology, managed by Frederik Coppens. In 2018 the Flemish government invested in the Flemish SuperComputer Center (VSC) to establish a Cloud and Data infrastructure. In collaboration with the VSC team, ELIXIR Belgium is developing services on top of this cloud infrastructure (e.g. usegalaxy.be), ensuring alignment and compliance with best practices at European level, in the framework of ELIXIR and the European Open Science Cloud.

Participant	**3 Swiss Institute of Bioinformatics SIB (ELIXIR-CH)**	**Switzerland**
PI	[redacted] **(F)**	
ELIXIR Node	**ELIXIR-CH**	

Description of the legal entity/department/laboratory/group:

The SIB Swiss Institute of Bioinformatics (www.sib.swiss) is an academic not-for-profit organization whose mission is to lead and coordinate the field of bioinformatics in Switzerland. Its data science experts join forces to advance biological and medical research and enhance health. SIB (i) provides the national and international life science community with a state-of-the-art bioinformatics infrastructure, including services, resources, expertise; and (ii) federates world-class researchers and delivers training in bioinformatics. The institute includes some 70 world-class research and service groups including 800 scientists in the fields of genomics, proteomics, evolution and phylogeny, systems biology, structural biology, text mining and machine learning and personalized health. SIB constitutes the Swiss ELIXIR Node.

Main tasks in the project per WP;

SIB will contribute to the following tasks in **WP1 Expert network:**

- T1.1 Network of data managers and scientific best practice (Participant).

- T1.2 Models for brokering data to ELIXIR Deposition Databases (Participant).

  - **SIB** will contribute with its expertise in data management and stewardship and will support the strengthening of ELIXIR's international network.

SIB will contribute to the following tasks in **WP2 Training and Capacity Building:**

- T2.1 Identify training needs and solutions in Data Management and Stewardship (Participant).

- T2.2 Develop best practices guidelines and training materials in DMS (Participant).

- T2.3 Capacity Building in Data Management and Stewardship (Participant).

- T2.4 Outreach activities to new ELIXIR Members and Communities (**Task Leader**).

- **SIB** Training Group and SIB members have a long lasting expertise in providing training for bioinformatics and data management. SIB will contribute to all tasks in WP2. In particular, SIB will contribute to 1) identifying training gaps, 2) developing new training materials and integrating SIB's existing training materials as part of the ELIXIR wide training efforts, 3) providing training in DMP nationally and internationally, and 4) disseminating DMP knowledge. Participating in WP2 will allow SIB to extend the national guidelines with European ones. Additionally, Swiss researchers will greatly benefit from ELIXIR's expertise and be able to improve their knowledge in DMPs according to European regulations.

SIB will contribute to the following tasks in **WP3 Common Data Management Toolkit:**

- T3.1 Establish a Starter Toolkit (Participant).

- T3.3 Access portal to Toolkit tailored to stakeholders (Participant).

- T3.4 Best Practices and training (Participant).

- **SIB** has a strong experience in the following major aspects of data management and planning:
  - Operating a DMP generation tool (DMP Generator Platform) that follows the recommendations of its national funding agency (Swiss National Science Foundation)
  - Actively supporting scientists with its expertise and providing help in using the DMP Generator Platform
  - Teaching DMP courses to a wide audience of Swiss scientists working in life science and medical fields
  - Specialized infrastructure to store, secure and archive data
- **SIB** will contribute with its knowledge and expertise in order to further develop a common data management toolkit that can be used in several countries and be tailored to national and local needs.

SIB will contribute to the following tasks in **WP5 Demonstrator Projects:**

- T5.2 Implementation of pilots data management plans (Participant).

- **SIB** will focus on DMP-activities related to sensitive human data applying its experience in national and international efforts such as the Swiss Personalized Health Network (SPHN) where data protection and IT security topics are mandatory requirements.

SIB will contribute to the following tasks in **WP7 Federated European Genome-phenome Archives for transnational access of COVID-19 host data:**

- Task 7.2: A technical implementation required for a federated EGA Network (Participant).
- Task 7.3: Coordination of metadata standards for phenotype submission and access of COVID-19 data (Participant).
- Task 7.4: Operational support and maturity model for Federated EGA nodes (Participant).

Short Profile of key Staff Members who will be undertaking the work

[redacted]

[redacted]

List of the top 5 most relevant publications and/or products, services (including widely-used datasets or software), or other achievements

1. Baillie Gerritsen V, Palagi PM, Durinx C.Bioinformatics on a national scale: an example from Switzerland. Brief Bioinform. 2017 Jul 4. doi: 10.1093/bib/bbx073. PMID: 29106442
2. Morgan SL, Palagi PM, Fernandes PL et al. The ELIXIR-EXCELERATE Train-the-Trainer pilot programme: empower researchers to deliver high-quality training. F1000Research 2017, 6:1557 (doi: 10.12688/f1000research.12332.1)
3. Durinx C., McEntyre J, Appel R *et al.* Identifying ELIXIR Core Data Resources [version 2; referees: 2 approved]. *F1000Research* 2017, 5(ELIXIR):2422
4. Ison J, *et al.* Tools and data services registry: a community effort to document bioinformatics resources, *Nucleic Acids Research*, Volume 44, Issue D1, 4 January 2016, Pages D38–D47

List of top 5 most relevant relevant previous projects or activities, connected to the subject of this proposal

1. EU H2020 (H2020-INFRADEV-1-2015-1) ELIXIR-EXCELERATE: Fast-track ELIXIR implementation and drive early user exploitation across the life sciences
2. FAIRplus (802750) This project will develop the guidelines and tools needed to make data FAIR. Through worked examples using IMI and EFPIA data and application and extension of existing methods we will improve the level of discovery, accessibility, interoperability and reusability of selected IMI and EFPIA data (H2020-JTI-IMI2-2017-12-two-stage)
3. Rhapsody (IMI-115881, SERI-16.0097-2) The stated goal of RHAPSODY is to define a molecular taxonomy of type 2 diabetes mellitus (T2D) that will support patient segmentation, inform clinical trial design, and the establishment of regulatory paths for the adoption of novel strategies for diabetes prevention and treatment.
4. Hypo-RESOLVE (IMI-777460, IMI2 - Call 10) The project aims to enhance knowledge about hypoglycaemia by adding to the understanding of the underlying causes of the condition, as well as its predictors and consequences.

5. BEAT-DKD (IMI-115974, IMI2 - Call 5) This project is committed to deliver better stratification of patients and more effective tools for use in innovative clinical trials, with the aim to improve prevention and management of Diabetic kidney disease (DKD) and establish a new paradigm for precision medicine in DKD.

Description of any significant infrastructure and/or major items of technical equipment, relevant to the proposed work (if applicable):

SIB brings together over 800 scientists with a strong expertise in bioinformatics and carrying outstanding science, publishing 1297 peer-reviewed articles over the last 15 years.

World-class databases and resources

SIB develops and maintains world-class databases, including UniProtKB/Swiss-Prot (expertly curated protein knowledgebase providing a high level of annotation; over 400,000 unique visitors per month), neXtProt (humancentric protein knowledge platform), SWISS-MODEL Repository (three-dimensional protein structure models), STRING (protein interaction networks for Systems Biology), SwissRegulon (genome-wide transcription regulatory networks), Eukaryotic Promoter Database (EPD), clipz (database of binding sites of RNA binding proteins), PROSITE (protein families and domains), MyHits (protein sequences and motifs), CleanEx (gene expression data), Bgee (data on gene expression evolution), OpenFlu (influenza database), OrthoDB (the hierarchical catalogue of eukaryotic orthologs), miROrtho (the catalogue of animal microRNA genes), OMA (Orthology MAtrix), ImmunoDB (insect-immune-related gene families), ViralZone (portal to viral UniProtKB entries), UniCarbKB (curated glycan database, etc.

Software tools and platforms

SIB develops and supplies software and platforms for the global life science research community, such as SWISS-MODEL Workspace (protein structure homology modelling), SwissDock (ligand docking) and other drug design tools and databases, ISA (integrative bi-clustering tool), PPA (coupled modularization of several datasets), Melanie (2D gel analysis platform – celebrating 30 years of existence in 2013), MSight (LC-MS imaging and analysis software), DeepView/Swiss-PdbViewer (protein visualization, modelling and analysis - resulted in one of the most cited Swiss scientific publications!), Newick utilities (high-throughput phylogenetic tree processing) and Systems Biology Tools, and more.

Core and computing facilities

Through its bioinformatics core facilities, High-Performance Computing (HPC) centers, and embedded bioinformaticians, SIB provides both computing power and expert data analysis services to (Swiss) life scientists, thus enabling them to perform world-class biomedical research and analyze the resulting data.

SIB groups Vital-IT and Core-IT provide computational infrastructure, development support and bioinformatics expertise to the life science community in the western part of Switzerland – both in academia and industry. SIB also supports the sciCORE center for scientific computing, which is the scientific competence center for the life science community in the Basel area, and closely collaborates with S3IT (Service and Support for Science IT) at the University of Zurich and ITS-SIS (Scientific Information Services) at the ETH Zurich.

Training

SIB has a long-standing experience on training for bioinformatics, data analysis and data management, and organizes annually close to 60 courses, gathering over 1200 participants from all nationalities. The SIB Training group, the Swiss-wide training platform, is in charge of 1) providing teaching, 2) managing the SIB courses portfolio and 3) supporting SIB teachers in organizing training events.

Participant	**4 The University of Cyprus. UCY (ELIXIR-CY)**	**Cyprus**
PI	[redacted] **(M)**	
ELIXIR Node	**ELIXIR-CY (Observer)**	

Description of the legal entity/department/laboratory/group:

The **University of Cyprus (UCY)** is considered to be the leading university and the most active research institution in Cyprus. UCY has implemented a large number of research projects funded by the European Commission, the Cyprus Research Promotion Foundation and several public and private research organizations. UCY excelled in all pillars of the H2020 program. It secured € 13.3 million with a total of 1[illegible] projects from the WIDESPREAD Action, which includes 2 Teaming projects (Phase 2) for the creation of new Centers for Research Excellence. One of these two grants is worth €15 million and was ranked 1st amongst the ICT proposals and 5th in the overall ranking. Moreover, UCY secured €9.3 million from the Excellence pillar with 7 grants from the European Research Council, 34 Marie Sklodowska-Curie Projects, 2 Future and Emerging Technologies Project (FETs) and 2 Research Infrastructure projects. UCY's performance in the pillar "Societal Challenges" was also acknowledged, with 19 projects, 11 of which were in the field of Energy, 3 in the field of Health, 3 in Social Sciences, 1 in Digital Security and 1 in Marine Biology. In addition, UCY secured 5 projects under the "Industrial Development" axis with 4 projects in Information and Communication Technologies and 1 project in Space Action, as well as 1 project under the EURATOM initiative and 2 projects in the horizontal priority "Science with and for the Society". Since the establishment of European Research Council (ERC) in 2007, UCY managed to win 16 ERC projects (5 Starting, 2 Consolidator, 2 Advanced and 6 Proof of Concept Grants). UCY implements 20 Marie Sklodowska-Curie projects. Apart from Individual Fellowships, it participates in ITN and Rise Actions. UCY has successfully implemented 43 Marie Skłodowska-Curie projects.

The **Bioinformatics Research Laboratory** (**BRL**; the first dedicated Bioinformatics group on the island, established in 2005) is involved in Research, Teaching and Training activities relevant to the field. Key contributions of the BRL are in the fields of sequence analysis and protein structural bioinformatics, computational comparative genomics and biomedical text mining. During the last 5 years, the BRL has established a strong research interest towards the characterization and study of proteins and processes related to eukaryotic endomembrane systems, (e.g. nuclear pore complexes, autophagy) using computational and functional genomics approaches. The **Molecular Ecology and Evolution Lab (MEEL)** currently coordinates the recently funded Horizon 2020 Twinning "iBioGen" project, which links the University of Cyprus with three internationally-leading partners in biodiversity genomics and involves a wide range of training and networking activities related to the application of genomic approaches to island biodiversity.

BRL website: http://troodos.biol.ucy.ac.cy

MEEL website: https://annapapadopoulou.weebly.com/

Main tasks in the project per WP;

UCY will contribute to the following tasks in **WP1 Expert network:**

- T1.1 Network of data managers and scientific best practice (Participant).

Short Profile of key Staff Members who will be undertaking the work

[redacted]

[redacted]

List of the <u>top 5 most relevant</u> publications and/or products, services (including widely-used datasets or software), or other achievements

1. Mier P, Paladin L, Tamana S, Petrosian S, Hajdu-Soltisz B, Urbanek A, Gruca A, Plewczynski D, Grynberg M, Bernado P, Gaspari Z, Ouzounis C, Promponas VJ, Kajava AV, Hancock JM, Tosatto S, Dosztanyi Z, Andrade-Navarro MA. Disentangling the complexity of low complexity proteins. Briefings in Bioinformatics (2019) https://doi.org/10.1093/bib/bbz007.
2. Jacomin A.C., Samavedam S., Promponas V., Nezis I.P. iLIR database: A web resource for LIR motif-containing proteins in eukaryotes (2016) Autophagy 12(10):1945-1953.
3. Papadopoulou, A, Taberlet, P, Zinger, L. (2015) Metagenome skimming for phylogenetic community ecology: a new era in biodiversity research. Molecular Ecology 24: 3515-3517.
4. Papanikolaou N, Pavlopoulos GA, Pafilis E, Theodosiou T, Schneider R, Satagopam VP, Ouzounis CA, Eliopoulos AG, Promponas VJ, Iliopoulos I. BioTextQuest+: a knowledge integration platform for literature mining and concept discovery. (2015) Bioinformatics. 31(6):979.
5. Promponas V.J., Iliopoulos I., Ouzounis C.A. Annotation inconsistencies beyond sequence similarity-based function prediction – phylogeny and genome structure. (2015) Standards in Genome Sciences 10:108.

List of <u>top 5 most relevant</u> relevant previous projects or activities, connected to the subject of this proposal

1. EC H2020-WIDESPREAD-05-2017-TwinningPhaseI (AP, coordinator, VP partner) Project title: "iBioGen - Twinning for European Excellence in Island Biodiversity Genomics"
2. EC H2020-WIDESPREAD-04-2017-TeamingPhase1 (VP Partner) Project title: "Integrated Precision Medicine Technologies Research Centre of Excellence"
3. EC FP7 ICT-2009.5.3 (VP Partner) Project title: "GRANATUM: A Social Collaborative Working Space Semantically Interlinking Biomedical Researchers, Knowledge And Data For The Design And Execution of In-Silico Models And Experiments In Cancer Chemoprevention"
4. Cyprus Research Promotion Foundation–PENEK (VP, Coordinator) Project Title: "Development of novel computational genomics web services and methods for the analysis of low complexity regions in amino acid sequences."
5. Cyprus Research Promotion Foundation–HEALTH (VP, Coordinator) Project Title: "Development of an integrated Computer System for Text Mining Biomedical Databases."

Description of any significant infrastructure and/or major items of technical equipment, relevant to the proposed work (if applicable):

N/A

Participant	**5 The Cyprus Foundation for Muscular Dystrophy Research (The Cyprus Institute of Neurology and Genetics). CING (ELIXIR-CY)**	**Cyprus**
PI	[redacted] **(M)**	
ELIXIR Node	**ELIXIR-CY (Observer)**	

Description of the legal entity/department/laboratory/group:

The Cyprus Institute of Neurology and Genetics (CING, www.cing.ac.cy), was established in 1990 as a private, bi-communal, non-profit research, academic and medical organization. The vision of CING is to function as an International Centre of Excellence and a Regional Centre with the mission to provide high level clinical and laboratory services, innovative research and postgraduate education in the areas of Neurology, Genetics, Biomedical and Medical Sciences. CING has organised five specialised Neurology Clinics (A to E), a Clinical Genetics Clinic (Clinical Sciences Sector) and eight Biomedical Departments (Biomedical Sciences Sector), which offer a wide spectrum of specialised diagnostic services. Recent achievements of the CING include: 1. Establishment of the Cyprus School of Molecular Medicine (CSMM) in 2012, an independent Postgraduate School accredited by the Ministry of Education and Culture of Cyprus, 2. Establishment of the 1st Chair of Bioinformatics in Cyprus (EU H2020 ERA Chair grant, BIORISE, 2015-2020), and 3. Current expansion into state-of-the-art premises and new research groups, through a loan granted by the European Investment Bank to create a Translational Medicine Research and Innovation Centre. All activities of the CING are based on a three pillar approach that involves the delivery of state-of-the-art diagnostic services, execution of innovative research programmes as well as participation in postgraduate education. Currently the CING represents the most advanced Medical/Biomedical organisation in Cyprus. It operates as a tertiary centre for Neurological disorders in the region and delivers medical diagnosis, management and treatment to more than 7.000 patients annually. It provides more than 85.000 specialised tests that cover genetics of many disorders, including thalassemia, neurogenetics, cardiovascular genetics, cancer predisposition, as well as biochemical genetics, cytogenetics, molecular virology, electron microscopy and forensics. CING has developed the most competitive biomedical expertise and research infrastructure available in Cyprus, and has provided a very effective platform for bridging and streamlining the activities of the public, private and other non-government organizations working in the Health Sector. In addition CING is very successful in obtaining competitive funding, both at the national and International levels.

The Group of Bioinformatics has been established in 2016 under the support of the EU H2020 ERA Chair grant, BIORISE (Grant Number 669026) with an EU contribution of EUR 2 273 546 for the period

2015-2020. The mission of the Group is to function as a hub of excellence in the areas of applied bioinformatics to early diagnosis, effective prognosis and drug discovery contributing to the concepts of Preventive, Personalized and Precise Medicine. This is accomplished through state-of-the-art bioinformatics research, advanced education in postgraduate level and a bundle of bioinformatics services provided to the other CING Departments/Clinics. The Bioinformatics Group works on a number of powerful in-house servers as well as on SuperComputing facilities available through successful proposals. The Group consists of 4 Postdoctoral Research Fellows, 1 Visiting Scientist, 5 PhD students and the Bioinformatics ERA Chair Holder who is the Head of the Group.

The Bioinformatics Group works in collaboration with all CING Departments and Clinics in order to satisfy urgent needs for innovative computational analyses on a wealth of biomedical data, thus enhancing both the early diagnosis for a large range of diseases and targeted drug discovery. The Bioinformatics Group uses state-of-the-art computational methodologies and techniques to rise to the occasion, such as computational intelligence, machine learning, pattern recognition and data mining, modelling and simulation, network reconstruction and visualization, complex network analysis, text mining and association analysis. Further to these, it employs the most powerful computational ideas and tools to create handy but powerful web databases and applications, number crunching systems and visualizers of multisource information providing unique data integration methods.

We are providing horizontal services to all the Departments and Clinics of CING in terms of large-scale genomics (NGS) and proteomics analysis as well as post-omics analysis in the level of functional analysis and systems analysis as well. Additionally, we are developing relational databases and pipelines for the analysis of medical information. The already developed pipelines within CING are related to: (1) Whole Exome Sequence Preprocessing and Analysis, (2) RNA Seq Preprocessing and Analysis, (3) Microarray Data Preprocessing and Analysis, (4) Proteomics Data Preprocessing and Analysis, (5) Metabolomics Analysis, (6) Metagenomics Analysis, (7) Structural Analysis – Molecular Dynamics - Cheminformatics (8) Functional Analysis – Pathway Analysis, (9) Network Representation and Analysis, (10) Multisource Data Integration and visualization.

The Bioinformatics Group is hosted in a well-structured laboratory space, having the following available computational resources: (a) 11 powerful workstations, (b) Six high performance servers (2 DataBase Servers, 2 Web Servers, 2 High Performance Computing Servers, 144 CPUs in Total and a max RAM of 512 GB) divided in to a three-clustered architecture, and a storage system of 60TB capacity, (c) Access to Super-Computing Facilities, (d) a plethora of open-source bioinformatics tools and in-house built pipelines.

The Group's educational activities include various talks, seminars and workshops as well as the support of a full semester postgraduate course in Bioinformatics for the MSc/PhD programmes of the Cyprus School of Molecular Medicine, supervision/mentoring of MSc and PhD students, as well as specialized trainings and internships.

CING Bioinformatics Group Web Page: http://www.biorisecyprus.com

Main tasks in the project per WP;

CING will contribute to the following tasks in **WP1 Expert network:**
- T1.1 Network of data managers and scientific best practice (Participant).

Short Profile of key Staff Members who will be undertaking the work

[redacted]

[redacted]

[redacted]

List of the top 5 most relevant publications and/or products, services (including widely-used datasets or software), or other achievements

1. Anastasis Oulas, George Minadakis, Margarita Zachariou, and George Spyrou. Selecting variants of unknown significance through network-based gene-association significantly improves risk prediction for disease-control cohorts. Scientific Reports. 2019 Feb 1
2. G. Minadakis, M. Zachariou, A. Oulas, and G. M. Spyrou, PathwayConnector: finding complementary pathways to enhance functional analysis, Bioinformatics, Aug 14 2018.
3. Margarita Zachariou, George Minadakis, Anastasis Oulas, Sotiroula Afxenti, George M. Spyrou. Integrating multi-source information on a single network to detect disease-related clusters of molecular mechanisms. J Proteomics. 2018 Mar 13. pii: S1874-3919(18)30104-0. doi: 10.1016/j.jprot.2018.03.009
4. Anastasis Oulas, George Minadakis, Margarita Zachariou, Kleitos Sokratous, Marilena M. Bourdakou, George M. Spyrou. Systems Bioinformatics: increasing precision of computational diagnostics and therapeutics through network-based approaches. Briefings in Bioinformatics 2017 Nov 27. doi: 10.1093/bib/bbx151.

5. Marilena M. Bourdakou and George M. Spyrou. Informed walks: whispering hints to gene hunters inside networks' jungle. BMC Syst Biol. 2017 Oct 11;11(1):97.

List of top 5 most relevant relevant previous projects or activities, connected to the subject of this proposal

1. Participation in H2020-WIDESPREAD-04-2017-Teaming Phase 1, Grant Agreement (GA) No: 763781, Acronym: IPMT, Title: Integrated Precision Medicine Technologies Research Centre of Excellence
2. Participation in H2020-WIDESPREAD-04-2017-Teaming Phase 1, Grant Agreement (GA) No: 763700, Acronym: AGRICYGEN, Title: Establishment of an Agricultural Genomics Centre in Cyprus

Description of any significant infrastructure and/or major items of technical equipment, relevant to the proposed work (if applicable):

N/A

Participant	**6 Ústav Organické Chemie a Biochemie AV ČR, v.v.i. UOCHB (ELIXIR-CZ)**	**Czech Republic**
PI	[redacted] **(M)**	
ELIXIR Node	**ELIXIR-CZ**	

Description of the legal entity/department/laboratory/group:

The Institute of Organic Chemistry and Biochemistry, Czech Academy of Sciences, v. v. i. (IOCB PRAGUE) - is the Node representative and coordinator of the ELIXIR infrastructure in the Czech Republic and guarantees its organizational background.

The mission of the IOCB PRAGUE is independent basic research (with the purpose to collect new knowledge) in organic chemistry and biochemistry, and molecular biology oriented disciplines related to organic chemistry and biochemistry, with strong aspects of application of the results in practice. The research is oriented mainly towards the following fields: nucleic acid components, proteins, peptides, natural products, synthetic functional molecules and molecular modeling. The aim of the Institute is to reach excellence in the international competition and to keep this position in the long term.

The IOCB PRAGUE creates and maintains selected collections of reference standards for chemical substances and makes them available to the professional public. The IOCB PRAGUE has a long-term experience in service provisions, e.g. the Benchmark Geometry and Energy Database integrated data – http://www.begdb.com/. The creation and administration (1996–2003) of the HIV Protease Structural Database currently fully transferred to the RSCB (Research Collaboratory for Structural Bioinformatics) consortium responsible for the PDB (Protein Data Bank) – http://xpdb.nist.gov/hivsdb/hivsdb.html.

The IOCB PRAGUE will be the provider and administrator of the Small Molecules database as well as the provider of the tools for protein structure analysis, which is a joint project with the EBI and UPOL. For the purpose of ELIXIR project a part of Bioinformatics group is fully dedicated to the running of the RI management, scientific and technical maintenance.

At the technical level, the team will be built up with the vision of collaboration between technical experts from different partners to share and develop new technical approaches to fulfill the needs of the infrastructure. The IOCB PRAGUE team has a strategy of recruiting new specialists to run communication platform of the whole ELIXIR CZ infrastructure with a close collaboration with the IT partner – CESNET.

At the scientific level, the synergy between the participating institutions and different aspects of their expertise should provide complete scientific support and maintenance of the infrastructure. At the IOCB PRAGUE , this support will be provided by the scientific team. and its members.

The strength of the IOCB PRAGUE as the coordinator and leader lies in synergy with partner institutions, which showed their commitment during the construction and preparatory phases of ELIXIR CZ –namely

Masaryk University, Charles University, Institute of Microbiology, Institute of Molecular Genetics and University of South Bohemia.

Main tasks in the project per WP;

UOCHB will contribute to the following tasks in **WP1 Expert network:**
- T1.1 Network of data managers and scientific best practice (Participant).
- T1.2 Models for brokering data to ELIXIR Deposition Databases (Participant).
- T1.3 Business model (Participant).
- T1.4 Sustainable and scalable operating model for harmonised data management in European projects (**Task Leader**).

UOCHB will contribute to the following tasks in **WP2 Training and Capacity Building:**
- T2.2 Develop best practices guidelines and training materials in DMS (Participant).

UOCHB will contribute to the following tasks in **WP3 Common Data Management Toolkit:**
- T3.2 Processes for enriching, maintaining and sustaining the Toolkit strators (Participant).
- T3.3 Access portal to Toolkit tailored to stakeholders (Participant).
- T3.4 Best Practices and training (Participant).

UOCHB will contribute to the following tasks in **WP4 Communications, Industry, International, Impact and Sustainability:**
- T4.4 Implementing an impact assessment toolkit for demonstrating ELIXIR's value, nationally and at European level (Participant).

Short Profile of key Staff Members who will be undertaking the work

[redacted]

[redacted]

List of the top 5 most relevant publications and/or products, services (including widely-used datasets or software), or other achievements

1. Jakubec, D ; Vondrasek, J ; Finn, RD ; 3DPatch: fast 3D structure visualization with residue conservation, BIOINFORMATICS, 35, 2, 332-334, (2019),DOI: 10.1093/bioinformatics/bty464
2. Galgonek, J; Vymetal, J; Jakubec, D; Vondrasek, J: Amino Acid Interaction (INTAA) web server, NUCLEIC ACIDS RESEARCH, 45, W1, W388-392, (2017)
3. Galgonek, J; Hurt, T; Michlikova, V; Onderka, P; Schwarz, J; Vondrasek, J: Advanced SPARQL querying in small molecule databases, JOURNAL OF CHEMINFORMATICS, 8 31, (2016), DOI: 10.1186/s13321-016-0144-4
4. Galgonek, J; Vondrasek, J: On InChI and evaluating the quality of cross-reference links, JOURNAL OF CHEMINFORMATICS , 6, 15,(2014), DOI: 10.1186/1758-2946-6-15
5. Kokoschka, M; Galgonek, J. Vondrasek, J; Hobza, P: Computational methods for the description of pharmacologically relevant platinum complexes - molecular structure and bond dissociation, PHYSICAL CHEMISTRY CHEMICAL PHYSICS, 18, 5, 4051-4062, (2016), DOI: 10.1039/c5cp03391h

List of top 5 most relevant relevant previous projects or activities, connected to the subject of this proposal

1. EU H2020 (H2020-INFRADEV-1-2015-1) ELIXIR-EXCELERATE: Fast-track ELIXIR implementation and drive early user exploitation across the life sciences, Grant number: 676559, 2015-2019.
2. "Data Stewardship Wizard": A Tool Bringing Together Researchers, Data Stewards, and Data Experts around Data Management Planning. Pergl, R., Hooft, R., Suchánek, M., Knaisl, V., and Slifka, J. Submitted to Data Science Journal.

Description of any significant infrastructure and/or major items of technical equipment, relevant to the proposed work (if applicable):

- ELIXIR server (48 cores, 512 GB RAM) and disk (100 TB) at dedicated Node of NGI for national Czech ELIXIR connected to the CESNET infrastructure by 10 Gb/s. Capacity limits up to 500 cores. Know/how for integration into ELIXIR infrastructure.
- Archive of capacity 300TB on HSM system for semi-active data. CESNET connection by two 10G/s channels.
- Dedicated network connection of the ELIXIR CZ Node and its subnodes. Internal network infrastructure upon request.

Participant	**7 HITS GGMBH. HITS (ELIXIR-DE)**	**Germany**
PI	[redacted]**(M)**	
ELIXIR Node	**ELIXIR-DE**	

Description of the legal entity/department/laboratory/group:

The Heidelberg Institute for Theoretical Studies was established 2010 bei SAP co-founder Klaus Tschira (1940-2015) and the Klaus Tschira Foundation as a private, non-profit research institute.
HITS conducts basic research in the natural sciences, mathematics and computer science with a focus on the processing, structuring and analyzing of large amounts of complex data and the development of computational methods and software. The research fields range from molecular biology to astrophysics.

The work group SDBV (Scientific Databases and Visualization) of HITS was founded in 1999 by [redacted] and is led since 2009 by [redacted]. The members unite a wide variety of specialities (biology, biochemistry, chemistry, computer science, computer linguistics). The two main prongs of SDBV's activity are (I) building, populating, and running SABIO-RK, a hand-curated database for FAIR reaction kinetics data, and (ii) being part of the FAIRDOM infrastructure project (lead by [redacted]. FAIRDOM aims at providing management of FAIR Data, Operations, and Models in a Systems Biology and Systems Medicine context. FAIRDOM has its own legal entity, the FAIRDOM Association e.V.

As part of the de.NBI infrastructure and ELIXIR-DE involvement, SDBV runs FAIRDOM-based data management.

Main tasks in the project per WP;

HITS will contribute to the following tasks in **WP1 Expert network:**
- T1.1 Network of data managers and scientific best practice (Participant).

HITS will contribute to the following tasks in **WP2 Training and Capacity Building:**
- T2.1 Identify training needs and solutions in Data Management and Stewardship (Participant).
- T2.2 Develop best practices guidelines and training materials in DMS (Participant).
- T2.3 Capacity Building in Data Management and Stewardship (Participant).
- T2.4 Outreach activities to new ELIXIR Members and Communities (Participant).

HITS will contribute to the following tasks in **WP3 Common Data Management Toolkit:**
- T3.1 Establish a Starter Toolkit (Participant).
- T3.2 Processes for enriching, maintaining and sustaining the Toolkit strators (Participant).
- T3.3 Access portal to Toolkit tailored to stakeholders (Participant).
- T3.4 Best Practices and training (Participant).

HITS will contribute to the following tasks in **WP5 Demonstrator Projects:**
- T5.1 Typology of projects based on the type of resources needed to implement a Ma-DMP (Participant).
- T5.2 Implementation of pilots data management plans (Participant).
- T5.3 Development,implementation and refinement of key performance indicators to monitor the pilots implementation of data management plans (Participant).
- T5.4 Capacity building actions based on pilots outcomes (Participant).

Short Profile of key Staff Members who will be undertaking the work

[redacted]

List of the top 5 most relevant publications and/or products, services (including widely-used datasets or software), or other achievements

http://www.fairdomhub.org commons

http://sabiork.h-its.org literature curated reaktion kinetics data

List of top 5 most relevant relevant previous projects or activities, connected to the subject of this proposal

1. FAIRDOM transnational project (ERANet. SEEK, FAIRDOMHub)
2. de.NBI-SysBio Systems Biology Data Management (BMBF German network For Bioinformatics Infrastructure)
3. Virtual Liver Network data management (BMBF systems biology of the liver)
4. LiSyM data management (BMBF Liver Systems Medicine)
5. NMTrypI data management (EU Grant Agreement 603240)

Description of any significant infrastructure and/or major items of technical equipment, relevant to the proposed work (if applicable):

Servers running the resources described above.

Participant	**8 Danmarks Tekniske Universitet. DTU (ELIXIR-DK)**	**Denmark**
PI	[redacted] **(M)**	
ELIXIR Node	**ELIXIR-DK**	

Description of the legal entity/department/laboratory/group:
For almost two centuries DTU, Technical University of Denmark, has been dedicated to fulfilling the vision of H.C. Ørsted – the father of electromagnetism – who founded the university in 1829 to develop and create value using the natural sciences and the technical sciences to benefit society. Today, DTU is ranked as one of the foremost technical universities in Europe with a profile that increasingly includes the life sciences. At this point the university has over 45% life science activities and operates in close collaboration with the Danish pharma and biotech industries in addition to many international collaborators in this space.

The National Life-Sciences Supercomputing Center in Denmark called "Computerome" is national collaborative computing platform built for processing of very large data sets especially sensitive data from Healthcare. Computerome center serves as the Danish national infrastructure provider for health care and life sciences as well as representing as the Danish infrastructure in ELIXIR and Nordforsk since 2014. Computerome center's infrastructure is tailored specifically towards big data processing with High Throughput Computing (HTC).The center's main infrastructure is an HPDA (High-Performance Data Analytics) platform and the center provides services such as developing, maintaining, optimising (if needed re-design) and scaling HPC application codes.

Computerome hosts computing for both non-sensitive and sensitive research projects. The offering for non-sensitive data is Computerome's HPC system, with configuration aimed for data-driven research. The systems also offers secure cloud computing platform both as PaaS and SaaS variants, with several security profiles including extremely high-security setups for the most sensitive data. Within the cloud platform researchers will be free to create their own "supercomputers on demand", comprising of hundreds of nodes (thousands of CPU cores) and petabyte-scale storage, while maintaining complete freedom to define any environment they require, such as commercial analytical platforms, NGS analysis pipelines, user management schemes, queuing systems etc. The environment is built with certified, high performance and high security components designed for processing of sensitive data.

Main tasks in the project per WP;

DTU will contribute to the following tasks in **WP1 Expert network:**

- T1.1 Network of data managers and scientific best practice (Participant).

DTU will contribute to the following tasks in **WP3 Common Data Management Toolkit:**

- T3.2 Processes for enriching, maintaining and sustaining the Toolkit strators (Participant).

- T3.4 Best Practices and training (Participant).

Short Profile of key Staff Members who will be undertaking the work

[redacted]

List of the top 5 most relevant publications and/or products, services (including widely-used datasets or software), or other achievements

1. Hillion K.H., Kuzmin I., Khodak A., Rasche E., Crusoe M., Peterson H., Ison J., Ménager, H. (2017). Using bio.tools to generate and annotate workbench tool descriptions F1000Research 2017 (article). doi:10.12688/f1000research.12974.1
2. Doppelt-Azeroual, O., Mareuil, F., Deveaud, Kalaš, M., Soranzo, N., van den Beek, M., Grüning, B., Ison, J. and Ménager, H. (2017). ReGaTE: Registration of Galaxy Tools in Elixir GigaScience, doi:10.1093/gigascience/gix022
3. Ménager, H., Kalaš, M., Rapacki, K. and Ison, J. (2016). Using registries to integrate bioinformatics tools and services into workbench environments Int J Softw Tools Technol Transfer, doi:10.1007/s10009-015-0392-z
4. Ison, J. et al. (2015). Tools and data services registry: a community effort to document bioinformatics resources. Nucleic Acids Research. doi: 10.1093/nar/gkv1116
5. Ison, J., Kalaš, M., Jonassen, I., Bolser, D., Uludag, M., McWilliam, H., Malone, J., Lopez, R., Pettifer, S. and Rice, P. (2013). EDAM: an ontology of bioinformatics operations, types of data and identifiers, topics and formats Bioinformatics, doi: 10.1093/bioinformatics/btt113

List of top 5 most relevant relevant previous projects or activities, connected to the subject of this proposal

1. EU H2020 (H2020-INFRADEV-1-2015-1) ELIXIR-EXCELERATE: Fast-track ELIXIR implementation and drive early user exploitation across the life sciences, Grant number: 676559, 2015-2019.
2. The Danish Elixir infrastructure project - the Danish node (case number: 0601-01407B)

Description of any significant infrastructure and/or major items of technical equipment, relevant to the proposed work (if applicable):

- ELIXIR Tools & Data Services Registry (https://bio.tools): registry of software and data resources for the life-sciences
- biotoolsSchema (http://github.com/bio-tools/biotoolsschema): formalised model for life-science software description

- EDAM ontology (http://edamontology.org): ontology for semantic description of life-science software

Participant	**9 Tartu Ulikool. UTARTU (ELIXIR-EE)**	**Estonia**
PI	[redacted] **(F)**	
ELIXIR Node	**ELIXIR-EE**	

Description of the legal entity/department/laboratory/group:

The Estonian ELIXIR Node is organized by University of Tartu with the main focus on maintenance, upgrade and international integration of bioinformatics tools and databases produced by the Estonian research community. The resources have been maintained and constantly updated (e.g. g:Profiler annotations are based on EnsEMBL version 77 covering 196 organisms). Some of the Estonian resources, g:Profiler for gene list characterisations and MEM for large scale gene expression data analysis and queries, have been running on the CSC cloud kindly provided by ELIXIR Finland. Estonian bioinformatics resources are mainly targeted to life science researchers. These services are widely used across the globe.

Members of the Node have been teaching the analysis tools both locally and internationally and plan to provide these trainings to wider communities throughout ELIXIR (including countries not yet in ELIXIR, such as other Baltic States - Latvia and Lithuania). Also, thanks to the support of active training community in ELIXIR, Estonian Node plans to expand the post-graduate trainings (e.g. R for biologists, RNA-seq analysis, primer design). Estonia has been the forerunner of the ICT field with the majority of national infrastructures being digital (eHealth, E-TaxBoard/e-Customs, eGovernment, etc) and thus the education of the infrastructure's developers and maintainers is of high importance in Estonia. With the ELIXIR Hub (hosted by EMBL-EBI) ELIXIR Estonia will drive the virtual community building for ELIXIR software developers and infrastructure operators, thus providing the basis for further coordinated training events and informal knowledge exchange between members of different Nodes. The majority of the Estonian bioinformatics resources have been submitted to ELIXIR Tools and Data Services Registry. This should increase the visibility of the tools and increase the usage over the coming years. The Estonian Node will also contribute to the further development of the Tools and Data Services Registry with already existing link-out software named URLMAP. This will allow the end users to create web links to specific service groups based on a personal list of biological entities.

ELIXIR Estonia is also broadening international collaborations on developing secure data management and analysis solutions. In close collaborations with larger Nodes like ELIXIR Switzerland or ELIXIR UK, the Estonian Node can adopt already proven best practices in all the main areas of ELIXIR.

The Estonian ELIXIR Node will also help to disseminate data produced by Estonian research institutions - Estonian Genome Center at the University of Tartu (training center for BBMRI) and also Estonian Biocentre (closely linked with the University of Tartu) are already having or planning to have a high number of individuals sequenced (plan is to reach 5000 full sequences at 20x coverage and 50,000 individuals genotyped) in the coming years. Thus it is important to work closely with the European Genome-phenome Archive to make the data transition from the producers to the archive as fluent as possible.

Main tasks in the project per WP

UT will contribute to the following tasks in **WP1 Expert network:**

- T1.1 Network of data managers and scientific best practice (Participant).

UT will contribute to the following tasks in **WP2 Training and Capacity Building:**

- T2.1 Identify training needs and solutions in Data Management and Stewardship (Participant).

- T2.2 Develop best practices guidelines and training materials in DMS (Participant).

- T2.3 Capacity Building in Data Management and Stewardship (Participant).

- T2.4 Outreach activities to new ELIXIR Members and Communities (Participant).

UT will contribute to the following tasks in **WP3 Common Data Management Toolkit:**

- T3.1 Establish a Starter Toolkit (Participant).
- T3.2 Processes for enriching, maintaining and sustaining the Toolkit strators (Participant).
- T3.4 Best Practices and training (Participant).

Short Profile of key Staff Members who will be undertaking the work

[redacted]

List of the top 5 most relevant publications and/or products, services (including widely-used datasets or software), or other achievements

1. M. Raghavan, M. DeGiorgio, A. Albrechtsen, I. Moltke, P. Skoglund, T.S. Korneliussen et al, The genetic prehistory of the New World Arctic. Science, 345(6200), 1255832, (2014).
2. L. Kamm, D. Bogdanov, S. Laur, J. Vilo, A new way to protect privacy in large-scale genome-wide association studies. Bioinformatics, 29(7), 886-893, (2013).
3. A. Untergasser, I. Cutcutache, T. Koressaar, J. Ye, B.C. Faircloth, M. Remm, S.G. Rozen, Primer3—new capabilities and interfaces. Nucleic acids research, 40(15), e115-e115 (2012).
4. J. Reimand, T. Arak, J. Vilo, g: Profiler—a web server for functional interpretation of gene lists (2011 update).
   Nucleic acids research, 39(suppl 2), W307-W315 (2011).
5. P. Adler, R. Kolde, M. Kull, A. Tkachenko, H. Peterson, J. Reimand, J. Vilo, Mining for coexpression across hundreds of datasets using novel rank aggregation and visualization methods. Genome biology, 10(12), R139, (2009).

List of top 5 most relevant relevant previous projects or activities, connected to the subject of this proposal

1. EU H2020 (H2020-INFRADEV-1-2015-1) ELIXIR-EXCELERATE: Fast-track ELIXIR implementation and drive early user exploitation across the life sciences, Grant number: 676559, 2015-2019.

Description of any significant infrastructure and/or major items of technical equipment, relevant to the proposed work (if applicable):

The previous flagships of 1TB RAM, 64 core server; 256GB RAM, 16 core; are just recently upgraded by two additional 1TB RAM, 60 core servers. Additionally, there are web servers and ~800TB disk space. Also, University of Tartu has a HPC center running several clusters, with now over 3000 cores, as well as one node with 2TB RAM and 60 cores. University of Tartu is a leading member in Estonian Scientific computing infrastructure initiative.

Participant	**10 Barcelona Supercomputing Center. BSC (ELIXIR-ES)**	**Spain**
PI	[redacted] **(M)**	
ELIXIR Node	**ELIXIR-ES**	

Description of the legal entity/department/laboratory/group:

Barcelona Supercomputing Center (BSC) (www.bsc.es/) was established in 2005 and serves as the Spanish national supercomputing facility. The Center hosts MareNostrum 4, one of the most powerful supercomputers in Europe and its mission is to research, develop and manage information technologies in order to facilitate scientific progress. BSC is recognised as a first-class research center in supercomputing and in scientific fields that demand it, such as Life, Earth and Computer Sciences as well as Engineering. BSC has over 500 staff from more than 40 countries engaged in multidisciplinary scientific collaboration and innovation.

BSC is a hosting member of the PRACE distributed supercomputing infrastructure and an active participant in HiPEAC, the ETP4HPC and other international forums such as BDEC and BDVA. The centre develops technologies for Exascale within the BSC-led Mont-Blanc project, in the DEEP and DEEP-ER projects and the Human Brain Flagship project. BSC has also established joint research centers on Exascale with Intel and IBM.

The **Life Sciences Department** integrates the independent research of senior scientists that work on various aspects of computational biology, ranging from bioinformatics for genomics to computational biochemistry and text mining. This department is the coordinator of the National Spanish Bioinformatics Institute (INB) and, consequently, the Spanish representative of the European Bioinformatics Infrastructure ELIXIR.

Main tasks in the project per WP;

BSC will contribute to the following tasks in **WP1 Expert network:**

- T1.1 Network of data managers and scientific best practice (Participant).

BSC will contribute to the following tasks in **WP2 Training and Capacity Building:**

- T2.1 Identify training needs and solutions in Data Management and Stewardship (Participant).
- T2.2 Develop best practices guidelines and training materials in DMS (Participant).

BSC will contribute to the following tasks in **WP3 Common Data Management Toolkit:**

- T3.1 Establish a Starter Toolkit (Participant).
- T3.2 Processes for enriching, maintaining and sustaining the Toolkit strators (Participant).

BSC will contribute to the following tasks in **WP4 Communications, Industry, International, Impact and Sustainability:**

- T4.2 Operating the ELIXIR Innovation and SME Forum and enhancing implementation of national industry engagement efforts (Participant).

BSC will contribute to the following tasks in **WP5 Demonstrator Projects**:

- T5.1 Typology of projects based on the type of resources needed to implement a Ma-DMP (Participant).
- T5.2 Implementation of pilots data management plans (**Task Leader**).
- T5.3 Development,implementation and refinement of key performance indicators to monitor the pilots implementation of data management plans (Participant).

BSC will contribute to the following tasks in **WP7 Federated European Genome-phenome Archives for transnational access of COVID-19 host data:**

- Task 7.1: Architecture, interfaces, and compliance to support EGA federated network on COVID-19 host data management (Participant).

- Task 7.2: A technical implementation required for a federated EGA Network (Participant).
- Task 7.3: Coordination of metadata standards for phenotype submission and access of COVID-19 data (Participant).
- Task 7.4: Operational support and maturity model for Federated EGA nodes (Task Leader).

Short Profile of key Staff Members who will be undertaking the work

[redacted]

List of the top 5 most relevant publications and/or products, services (including widely-used datasets or software), or other achievements

1. Cañada A, Capella-Gutierrez S, Rabal O, Oyarzabal J, Valencia A, Krallinger M. LimTox: a web tool for applied text mining of adverse event and toxicity associations of compounds, drugs and genes. Nucleic Acids Research. 2017;45(Web Server issue):W484-W489. doi:10.1093/nar/gkx462.
2. Lochmüller H, Badowska DM, Thompson R, Knoers NV, Aartsma-Rus A, Gut I, Wood L, Harmuth T, Durudas A, Graessner H, Schaefer F, Riess O; RD-Connect consortium; NeurOmics consortium; EURenOmics consortium. RD-Connect, NeurOmics and EURenOmics: collaborative European initiative for rare diseases. Eur J Hum Genet. 2018 Feb 27. doi: 10.1038/s41431-018-0115-5.
3. Fernández JM, de la Torre V, Richardson D, Royo R, Puiggrós M, Moncunill V, Fragkogianni S, Clarke L; BLUEPRINT Consortium, Flicek P, Rico D, Torrents D, Carrillo de Santa Pau E, Valencia A. The BLUEPRINT Data Analysis Portal. Cell Syst. 2016 Nov 23;3(5):491-495.e5. doi: 10.1016/j.cels.2016.10.021.
4. Capella-Gutierrez S, de la Iglesia D, Haas J, Lourenco A, Fernandez-Gonzalez JM, Repchevsky D, Dessimoz C, Schwede T, Notredame C, Gelpi JL, Valencia A. Lessons Learned: Recommendations for Establishing Critical Periodic Scientific Benchmarking. bioRxiv 181677. doi: https://doi.org/10.1101/181677.
5. Puente XS, Beà S, Valdés-Mas R, Villamor N, Gutiérrez-Abril J, Martín-Subero JI, Munar M, Rubio-Pérez C, Jares P, Aymerich M, Baumann T, Beekman R, Belver L, Carrio A, Castellano G, Clot G, Colado E, Colomer D, Costa D, Delgado J, Enjuanes A, Estivill X, Ferrando AA, Gelpí JL, González B, González S, González M, Gut M, Hernández-Rivas JM, López-Guerra M, MartínGarcía D, Navarro A, Nicolás P, Orozco M, Payer ÁR, Pinyol M, Pisano DG, Puente DA, Queirós AC, Quesada V, Romeo-Casabona CM, Royo C, Royo R, Rozman M, Russiñol N, Salaverría I, Stamatopoulos K, Stunnenberg HG, Tamborero D, Terol MJ, Valencia A, López-Bigas N, Torrents D, Gut I, López-Guillermo A, López-Otín C, Campo E. Non-coding recurrent mutations in chronic lymphocytic leukaemia. Nature. 2015 Oct 22;526(7574):519-24. doi: 10.1038/nature14666.

List of top 5 most relevant relevant previous projects or activities, connected to the subject of this proposal

1. EU H2020 (H2020-INFRADEV-1-2015-1) **ELIXIR-EXCELERATE** (Finalised); https://www.elixir-europe.org/excelerate
   EC funded project to help ELIXIR coordinate and extend national and international data resources to ensure the delivery of life-science data services.
2. EU H2020 (INFRAEOSC-04-2018) **EOSC-Life** (Active)
   EOSC-Life brings together the 13 Biological and Medical ESFRI research infrastructures for analysing and reusing the prodigious amounts of data produced by life-science.

3. EU H2020 (IMI2-2017-12-02) **FAIR-Plus** (Active)
   Through worked examples using IMI and EFPIA data, this project will develop the guidelines and tools needed to make data FAIR.
4. **Spanish National Bioinformatics Institute** (INB-ISCIII) (Active); www.inab.org
   The INB serves in the coordination, integration and development of Spanish bioinformatics resources in projects in the areas of genomics, proteomics and translational medicine. It has contributed to the creation of a consistent computational infrastructure in the area of bioinformatics, participated in national and international genome projects, and trained bioinformatics users and developers.
5. EU H2020 (SC1-BHC-05-2018) **EUCANCan** (Active); https://eucancan.com/
   EUCANCan will establish the technological and legal framework for facilitating oncology data sharing between different European and Canadian partners.

Description of any significant infrastructure and/or major items of technical equipment, relevant to the proposed work (if applicable):

The BSC hosts the following computational resources:

- **MareNostrum 4** has 48 racks with more than 3,400 nodes with next generation Intel Xeon processors and a central memory of 390 Terabytes. Its peak power is over 12 PetaFlops.
- **MinoTauro** is a heterogeneous cluster where the main computational power is provided by NVIDIA GPUS. This cluster is the second most powerful at BSC and it is available as part of the RES resources and as Tier-1 system at the DECI-PRACE calls. The system provides more than 300 TFlops in total. NVIDIA GPU is a heterogeneous cluster with 2 configurations, the first one with 61 Bull B505 blades and the second with 39 bullx R421-E4 servers.
- **Starlife** has 54 computer nodes, with Intel Xeon chips, with a total of 2,160 cores (expected 140 TFLOPS), 9TB main memory, 1.3PB local storage for fast data mobilization, and 2PB storage under GPFS), connected through OmniPath. Starlife is configured to combine HPC, cloud and fast data mobilization and indexing technologies. Starlife will be complemented by additional hardware resources (Nord3) used mainly for internal testing and development. Nord 3 has 1 rack with 84 nodes with Intel Xeon Sandy Bridge processors with a central memory of 10.7 Terabytes and a peak performance of 27.9 TFlops.

Participant	**11 IT Center for Science Ltd. CSC (ELIXIR-FI)**	**Finland**
PI	[redacted] **(M)**	
ELIXIR Node	**ELIXIR**	

Description of the legal entity/department/laboratory/group:

CSC – IT Center for Science Ltd (http://www.csc.fi), is the Finnish national HPC center providing high-performance and cloud computing, networking and data services to academia, research institutes, public sector and industry. CSC is a non-profit limited company whose shares are fully owned by Finnish Government and Finnish universities. Established in 1971, it currently has over 300 employees, and supports a European-wide customer base of thousands of researchers in disciplines such as biosciences, environmental science, lingustics, chemistry and mathematical modelling. CSC provides Finland's widest selection of scientific software and databases, training and expert support, storage and data services, and Finland's most powerful supercomputing environment, and operates the Finnish national research and education network Funet. CSC is the Finnish ELIXIR node, initiator and coordinator of the collaborative data infrastructure EUDAT and a committed actor in the development of the European Open Science Cloud through leading roles in EOSCpilot and EOSC-hub. As responsible for national e-Infrastructures, CSC is also a member of PRACE and EGI.

Three major national organisations (CSC, Institute for Molecular Medicine FIMM and National Institute for Health and Welfare THL) involved in the biomedical ESFRIs have a consortium agreement since 2012 to build and operate and seek funding for integrated national infrastructure for biomedical science - Biomedinfra.fi comprising ELIXIR, BBMRI and EATRIS. Within this consortium Infrastructure, components for biomedical service providers are delivered via private network solutions that can be used to host tools and build topical data services and software environments needed for personalised medicine. Biomedinfra.fi builds upon existing strengths and track record of Finnish research and enables molecular level genetic diagnostics of, for example, heart diseases, cancer and obesity to improve the health of the aging population, one of the European grand challenges.

Main tasks in the project per WP;

CSC will contribute to the following tasks in **WP1 Expert network:**

- T1.1 Network of data managers and scientific best practice (Participant).

- T1.2 Models for brokering data to ELIXIR Deposition Databases (Participant).

CSC will contribute to the following tasks in **WP2 Training and Capacity Building:**

- T2.1 Identify training needs and solutions in Data Management and Stewardship (Participant).

- T2.2 Develop best practices guidelines and training materials in DMS (Participant).

- T2.3 Capacity Building in Data Management and Stewardship (Participant).

CSC will contribute to the following tasks in **WP3 Common Data Management Toolkit:**

- T3.1 Establish a Starter Toolkit (Participant).

CSC will contribute to the following tasks in **WP7 Federated European Genome-phenome Archives for transnational access of COVID-19 host data:**

- Task 7.1: Architecture, interfaces, and compliance to support EGA federated network on COVID-19 host data management (Participant).
- Task 7.2: A technical implementation required for a federated EGA Network (Participant).
- Task 7.3: Coordination of metadata standards for phenotype submission and access of COVID-19 data (Task Leader).
- Task 7.4: Operational support and maturity model for Federated EGA nodes (Participant).

Short Profile of key Staff Members who will be undertaking the work

[redacted]

[redacted]

List of the top 5 most relevant publications and/or products, services (including widely-used datasets or software), or other achievements

1. M. Linden, T. Nyrönen, I. Lappalainen, Resource Entitlement Management System, Selected papers of TNC2013 Conferencehttp://www.terena.org/publications/tnc2013-proceedings/ (2013)
2. M. Linden, S. Carmody, Data Protection Code of Conduct, Selected papers of TNC2013 Conference. http://www.terena.org/publications/tnc2013-proceedings/ (2013).
3. D. Broeder, B. Jones, D. Kelsey, P. Kershaw, S. Lüders, A. Lyall, T. Nyrönen, R. Wartel, H.J Weyer, Federated Identity Management for Research Collaborations. CERN-OPEN-2012-006: https://cdsweb.cern.ch/record/1442597 (2016)
4. T. Nyrönen, J. Laitinen, O. Tourunen, D. Sternkopf, R. Laurikainen, P. Öster, P. Lehtovuori, T.A. Miettinen, T. Simonen, T. Perheentupa, I. Västrik, O. Kallioniemi, A. Lyall, J. Thornton: Delivering ICT infrastructure for biomedical research, Proceedings of the WICSA/ECSA 2012 Companion Volume 37-44. ACM (2012)

List of top 5 most relevant relevant previous projects or activities, connected to the subject of this proposal

1. EU H2020 (H2020-INFRADEV-1-2015-1) ELIXIR-EXCELERATE: Fast-track ELIXIR implementation and drive early user exploitation across the life sciences, Grant number: 676559, 2015-2019.
2. ELIXIR (http://www.elixir-europe.org) and ELIXIR Finland (http://www.elixir-finland.org) research infrastructure for biological data. CSC developed Infrastructure as a service cloud for organisations and a REMS tool to manage access rights to research data.
3. Corbel (http://www.corbel.eu) and Biomedbridges (http://www.biomedbridges.eu/). ELIXIR coordinates biomedical sciences research infrastructures on the ESFRI roadmap cluster projects to develop the shared e-infrastructure — the technical bridges — to allow data integration in the biological, medical, translational and clinical domains and thus strengthen biomedical resources in Europe. CSC is e-Infrastructure partner participating in Data management and security technologies.
4. GÉANT (http://www.geant.net/About/Pages/home.aspx) is the pan-European research and education network that interconnects Europe's National Research and Education Networks (NRENs). Together Géant connects over 50 million users at 10,000 institutions across Europe, supporting research in areas such as energy, the environment, space and medicine. CSC contributes to the Géant project together with the other Nordic countries via NORDUnet. The CSC main effort has been focused to the campus best practices activity, which aims to address key challenges for European campus networks. This is done by organising working groups and providing an evolving and to-the-point set of best practice documents (BPDs) for the community.

Description of any significant infrastructure and/or major items of technical equipment, relevant to the proposed work (if applicable):

All Finnish research trusts on CSC through one or several of its services: the Finnish University and Research Network (FUNET), supercomputer and computing clusters, storage and data services, software and tools, and training and expert support. CSC has further a role to introduce new technologies and advances within ICT and its application to benefit of Finnish research, which has been accomplished over the years by, for example: connecting Finland to the Internet in 1988; introducing the massively parallel Cray T3E computer in 1996; establishing optical research data network allowing link speeds up to 100 Gbps (2009); building modular freecooling energy efficient data centre (2012); piloting a high density supercomputer prototype with many-core accelerators (2013); and becoming ELIXIR Finland Node (2015). The capabilities of CSC have enabled a number of advancements for Finnish research including: cancer research; galactic dynamo processes; aerosol effects on global warming; analysis of genetic variation in Finland and the Nordic countries, and resolving 3-D structures of viruses and more.

CSC's capabilities have to evolve with the demands of its customer base, the Finnish research system and this plan of intent describes how, with a focus on: Funet the national network backbone and related services; a computing and storage platform; a data centre facility; and a novel cloud service platform. These capabilities also form the platform for additional research infrastructures that CSC coordinates, operates or plans: platform for long term storage and preservation of research data and digital archives, Nordic e-Infrastructure Collaboration (NeIC), European life sciences infrastructure for biological information (ELIXIR), and Partnership for Advanced Computing in Europe (PRACE, and European Open Science Cloud (EOSC).

Participant	**12 Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement INRAE (ELIXIR-FR)**	**France**
PI	[redacted] **(F)**	
ELIXIR Node	**ELIXIR-FR**	

Description of the legal entity/department/laboratory/group:
The "Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement" (INRAE) is the first institute for agronomical research in
Europe and the second worldwide. INRAE develops finalized research for food, agriculture, and the environment. To be able to carry out its research, INRAE is developing significant resources in bioinformatics and in Infrastructures for bioinformatics. INRAE is a major contributor to the French Institute for Bioinformatics (IFB), a French strategic distributed infrastructure gathering the major French institutes working in the field of life science and acting as the french node of ELIXIR since 2014. INRAE is co-coordinating the ELIXIR Plant community since 2017. The URGI platform (https://urgi.versailles.inra.fr/Platform) that belongs to IFB has been involved in many different initiatives, developing and promoting standards and tools supporting open data for plant science, in particular in collaboration with the EMPHASIS infrastructure (Breeding API, Minimal Information About Phenotyping Experiments). Along with INRAE central documentation service, URGI is also very active in the Research Data Alliance, developing proof of concepts with the Wheat Initiative for FAIR data (www.wheatis.org) and helping INRAE's unit to develop and implement Data Management Plans for their projects.

Main tasks in the project per WP;

INRAE will contribute (**WP Leader**) to the following tasks in **WP5 Demonstrator Projects:**

- T5.1 Typology of projects based on the type of resources needed to implement a Ma-DMP (**Task Leader**).

- T5.2 Implementation of pilots data management plans (Participant).

- T5.3 Development,implementation and refinement of key performance indicators to monitor the pilots implementation of data management plans (Participant).

- T5.4 Capacity building actions based on pilots outcomes (Participant).

- **INRAE** will contribute to a demonstrator on Plant data: contribution to the specifications of the DMP: data types and access licence, needs in relation to the DMP Tool kit, needs in terms of implementation during the data life cycle, needs in terms of training
- **INRAE** will contribute to co-developments with WP3 (DMP tool kit) and WP2 (training)

Short Profile of key Staff Members who will be undertaking the work

[redacted]

[redacted]

List of the top 5 most relevant publications and/or products, services (including widely-used datasets or software), or other achievements

1. Amselem J, Cornut G, Choisne N, Alaux M, Alfama-Depauw F, Jamilloux V, Maumus F, Letellier T, Luyten I, Pommier C, Adam-Blondon A-F, Quesneville H (2019) RepetDB: a unified resource for transposable element references. Mobile DNA, 10:6. https://doi.org/10.1186/s13100-019-0150-y
2. Alaux M, Rogers J, Letellier T, Flores R, Alfama F, Pommier C, Mohellibi N, Durand S, Kimmel E, Michotey C, Guerche C, Loaec M, LainéM, Steinbach D, ChouletF, RimbertH, Leroy P, GuilhotN, Salse J, Feuillet C, International Wheat Genome Sequencing Consortium, Paux E, Eversole K, Adam-Blondon A-F, Quesneville H (2018) Linking the International Wheat Genome Sequencing Consortium bread wheat reference genome sequence to wheat genetic and phenomic data. Genome Biology, 19:111. https://doi.org/10.1186/s13059-018-1491-4
3. Ćwiek-Kupczyńska H, Altmann T, Arend D, Arnaud E, Chen D, Cornut G, Fiorani F, Frohmberg W, Junker A, Klukas C, Lange M, Mazurek C, Nafissi A, Neveu P, van Oeveren J, Pommier C, Poorter H, Rocca-Serra P, Sansone SA, Scholz U, van Schriek M, Seren Ü, Usadel B, Weise S, Kersey P, Krajewski P. (2016) Measures for interoperability of phenotypic data: minimum information requirements and formatting. Plant Methods. 2016 Nov 9;12:44. eCollection 2016. https://doi.org/10.1186/s13007-016-0144-4
4. A-F Adam-Blondon, M Alaux, C Pommier, D Cantu, Z-M Cheng, GR Cramer, C Davies, S Delrot, L Deluc, G Di Gaspero, J Grimplet, A Fennell, JP Londo, P Kersey, F Mattivi, S Naithani, P Neveu, M Nikolski, M Pezzotti, BI Reisch, R Töpfer, MA Vivier, D Ware, H Quesneville (2016) Towards an open grapevine information system. Hort Res, 3, 16056. https://doi.org/10.1038/hortres.2016.56
5. Grimplet J, Adam-Blondon A-F, Bert P-F, Bitz O, Cantu D, Davies C, Delrot S, Pezzotti M, Rombauts S, Grant R Cramer GR (2014) The grapevine gene nomenclature system. BMC Genomics, 15 :1077.

List of top 5 most relevant relevant previous projects or activities, connected to the subject of this proposal

1. EU H2020 (H2020-INFRADEV-1-2015-1) ELIXIR-EXCELERATE Fast-track ELIXIR implementation and drive early user exploitation across the life sciences, Grant number: 676559, 2015-2019.
2. Co-leader of the COST action Integrape (CA 17111, www.http://www.integrape.eu/) 2018-2022: data integration to maximize the power of OMICs for grapevine improvement.
3. FP7 TransPlant EU Infrastructure for plant bioinformatics, grant n° 283396.

Description of any significant infrastructure and/or major items of technical equipment, relevant to the proposed work (if applicable):

The URGI platform develops and maintains a modular and interoperable Information System for plant and pest genomics called GnpIS (URL: http://urgi.versailles.inra.fr/gnpis). It is a multispecies centralized information system designed to bridge different kinds of data, accessions, phenotypes, genetic or genomic characterization for species of agronomical interest. The system is used for data from INRAE and its partners. It is the official repository of the International Wheat Genome Sequence Consortium. It is a key resource in the DMP of several important French private-public initiatives on crops (maize, wheat, sunflower, pea, rapeseed, energy crops). The system is being made interoperable with other similar infrastructures at the international level through APIs (e.g. http://www.wheatinitiative.org) and in particular the Breeding API (www.brapi.org). GnpIS offers a virtual access through a web portal (https://urgi.versailles.inra.fr/gnpis/) and different tools for searching: a google like quick search tool, whole portal advanced query tools (Intermine, Biomart, Galaxy) and specific interfaces for queries in each module of the database.

Participant	**13 Centre National de la Recherche Scientifique. CNRS (ELIXIR-FR)**	**France**
PI	[redacted] **(F)**	
ELIXIR Node	**ELIXIR-FR**	

Description of the legal entity/department/laboratory/group:

Founded in 1939, the Centre National de la Recherche Scientifique (National Center for Scientific Research) is a government-funded research organization under the administrative authority of France's Ministry of Research. CNRS research units are spread throughout France, and employ a large body of permanent researchers, engineers, technicians, and administrative staff. The CNRS annual budget represents one-quarter of French public spending on civilian research.

CNRS (http://www.cnrs.fr/index.php) is organized in 1211 laboratories, either intramural or in partnership with universities, other research organizations or industry. As one of the largest fundamental research organisations in Europe, CNRS is involved in all scientific fields and is largely involved in national, European, and international projects. Interdisciplinary programs and actions offer a gateway into new domains of scientific investigation and enable CNRS to address the needs of society and industry.

CNRS has a long experience in the management of European projects. With a large place for collective guidance as part of the management inside the consortium; the organisation will provide a strong administrative support to such projects, with dedicated staff on financial, legal, and administrative issues.

Main tasks in the project per WP:

CNRS will contribute to the following tasks in **WP1 Expert network:**
- T1.1 Network of data managers and scientific best practice (Participant).
    - **CNRS** will contribute to the network of experts that will organize the DMP service across the nodes.

CNRS will contribute to the following tasks in **WP2 Training and Capacity Building:**
- T2.3 Capacity Building in Data Management and Stewardship (Participant).
    - **CNRS** will contribute to co-develop good quality training tools for along with the DMP Service

CNRS will contribute to the following tasks in **WP3 Common Data Management Toolkit:**
- T3.2 Processes for enriching, maintaining and sustaining the Toolkit strators (Participant).
- T3.4 Best Practices and training (Participant).
    - **CNRS** will contribute to the DMP toolkit and to the articulation of its national resources with it (e.g. https://www.cesgo.org, https://dmp.opidor.fr, …)

CNRS will contribute to the following tasks in **WP5 Demonstrator Projects:**
- T5.2 Implementation of pilots data management plans (Participant).
- T5.3 Development,implementation and refinement of key performance indicators to monitor the pilots implementation of data management plans (Participant).
- T5.4 Capacity building actions based on pilots outcomes (Participant).

    - **CNRS** will contribute to the implementation of WP3 toolkits to demonstrators projects.

Short Profile of key Staff Members who will be undertaking the work

[redacted]

[redacted]

[redacted]

List of the top 5 most relevant publications and/or products, services (including widely-used datasets or software), or other achievements

1. Sarah Cohen-Boulakia, Khalid Belhajjame, Olivier Collin, Jérôme Chopard, Christine Froidevaux, Alban Gaignard, Konrad Hinsenh, Pierre Larmande, Yvan Le Bras, Frédéric Lemoine, Fabien Mareuil, Hervé Ménager, Christophe Pradal, Christophe Blanchet.. Scientific workflows for computational reproducibility in the life sciences: Status, challenges and opportunities. Future Generation Computer Systems, Elsevier, 2017, ⟨10.1016/j.future.2017.01.012⟩
2. François Moreews, Olivier Sallou, Hervé Ménager, Yvan Le Bras, Cyril Monjeaud, Olivier Collin, Christophe Blanchet.. BioShaDock: a community driven bioinformatics shared Docker-based tools registry. *F1000Research*, Faculty of 1000, 2015, ⟨10.12688/f1000research.7536.1⟩
3. Damian Smedley, Syed Haider, Steffen Durinck, Luca Pandini, Paolo Provero, et al.. The BioMart community portal: an innovative alternative to large, centralized data repositories. *Nucleic Acids Research*, Oxford University Press, 2015, 43, pp.W589-W598. ⟨10.1093/nar/gkv350⟩
4. Vallenet D, Calteau A, Cruveiller S, Gachet M, Lajus A, Josso A, Mercier J, Renaux A, Rollin J, Rouy Z, Roche D, Scarpelli C, Médigue C. MicroScope in 2017: an expanding and evolving integrated resource for community expertise of microbial genomes. Nucleic Acids Res. 2017 Jan 4;45(D1):D517-D528.
5. Touchon M, Hoede C, Tenaillon O, […], Vallenet D, Médigue C, Rocha EP, Denamur E. Organised genome dynamics in the Escherichia coli species results in highly diverse adaptive paths. PLoS Genet., 2009, 5(1):e1000344.
6. Dominguez Del Angel V, Hjerde E, Sterck L *et al.* Ten steps to get started in Genome Assembly and Annotation [version 1; referees: 2 approved]. *F1000Research* 2018, 7(ELIXIR):148
7. van Rijswijk M, Beirnaert C, Caron C *et al.* The future of metabolomics in ELIXIR [version 2; referees: 3 approved]. *F1000Research* 2017, 6(ELIXIR):1649

List of top 5 most relevant relevant previous projects or activities, connected to the subject of this proposal

1. ELIXIR-EXCELERATE: Fast-track ELIXIR implementation and drive early user exploitation across the life sciences, EU H2020-INFRADEV-1-2015-1, Grant number: 676559, 2015-2019.
2. CYCLONE: Complete Dynamic Multi-cloud Application Management, EU H2020 ICT-07-2014, Grant Agreement Number 644925, 2015-2017
3. MICROME, EU-FP7 7 Collaborative Project, Grant Agreement Number 222886-2

Description of any significant infrastructure and/or major items of technical equipment, relevant to the proposed work (if applicable):

The French national research infrastructure in bioinformatics (Institut Français de Bioinformatique, IFB) was endowed by the "Investment for the Future" initiative with 20 M€ until 2023. The primary mission of IFB is to make available core bioinformatics resources and services to the national life science research community. To achieve this goal, the IFB infrastructure is to:

- Provide core infrastructure of **bioinformatics services** to the French life science community (academic and private partners), namely **tools and IT (Information Technology) facilities** dedicated to the management and analysis of biological data. This IT infrastructure is accessible through a mixte infrastructure: an academic Cloud and an academic Cluster.
- Provide a support to biology, health, agronomy and environment research programs, and access to high-level **expertise and mutualized competences**.
- Provide **training in bioinformatics** for life scientists and bioinformaticians
- Develop a **strategic vision** to maintain French research at the **highest level of expertise** in the analysis of biological data, and provide the scientific community with access to the newest

state-of-the-art technology internationally available in the field of bioinformatics. Be used as a lever to **ambitious national research projects** in life sciences and their applications.

- Ensure the international representation of the French bioinformatics community, especially in the context of the European ELIXIR network

IFB consists in more than 30 bioinformatics platforms providing services to the life science community all over the national territory and in a national hub in charge of providing administrative and technical central support. IFB aims at structuring the French bioinformatics community. It is the French Node of ELIXIR.

Participant	**14 Biomedical Sciences Research Centre Alexander Fleming, BSRCAF (ELIXIR-GR)**	**Greece**
PI	[redacted] **(M)**	
ELIXIR Node	**ELIXIR-GR**	

Description of the legal entity/department/laboratory/group:

The Biomedical Sciences Research Centre "Alexander Fleming" is a governmental not-for-profit organization, devoted to scientific and technological excellence, training and innovation in the biomedical sciences. The Centre was established in 1998 in honor of Nobel laureate Alexander Fleming and is consistently exhibiting top research performance among its peers in Greece, reaching international best practice levels in key academic and research excellence indicators. The Center hosts 21 groups with research focus ranging from gene expression, epigenetics, signal transduction, immunology, cancer research and stem cell biology. Competitive funding amounts each year for 75-85% of the total budget of the Center. This strong position in competitive funding is the result of the center's strategic prioritization of research, which falls tightly within European academic research priorities. The work of BSRCAF's scientists has already led to contributions to global biomedical research, e.g., through the discovery of the effect of Tumor Necrosis Factor (TNF), one of the most important scientific contributions to the treatment of rheumatoid arthritis in recent years. The research groups of the Centre are supported by a number of state of the art core facilities, most of them operating since 2001. Experienced personnel in the core facilities offer maintenance of the instruments, technical training and technical assistance, experimental design depending upon the researcher's knowledge and requirements, experimental troubleshooting and data analysis. At the same time, BSRCAF has been designated as the Greek hub of European Research Infrastructure networks, such as INFRAFRONTIER and ELIXIR, thus playing an important role in materializing EU's mid- to long-term vision for research (see National Roadmap for Research Infrastructures).

BSRCAF's excellence extends into effective innovation, through the operation of a Technology Transfer Office and establishment of a successful spin-off biotechnology company, Biomedcode SA, within its premises. In the field of bioinformatics, BSRCAF's researchers have developed advanced bioinformatic tools for miRNA research and meta-analysis methods for next generation sequencing data.

BSRCAF, the Representing Entity of the Greek ELIXIR Node, is uniquely positioned to coordinate INFRADEV3 activities in Greece. It has considerable in-house expertise on large scale genomic data management and analysis and has developed strong scientific collaborations with all Greek research institutions active in biological and biomedical research.

Main tasks in the project per WP;

BSRC will contribute to the following tasks in **WP5 Demonstrator Projects:**

- T5.1 Typology of projects based on the type of resources needed to implement a Ma-DMP (Participant).

BSRC will contribute to the following tasks in **WP7 Federated European Genome-phenome Archives for transnational access of COVID-19 host data:**

- Task 7.4: Operational support and maturity model for Federated EGA nodes (Participant).

Short Profile of key Staff Members who will be undertaking the work

[redacted]

List of the top 5 most relevant publications and/or products, services (including widely-used datasets or software), or other achievements

1. van Rijswijk M, Beirnaert C, Caron C, Cascante M, Dominguez V, Dunn WB, Ebbels TMD, Giacomoni F, Gonzalez-Beltran A, Hankemeier T, Haug K, Izquierdo-Garcia JL, Jimenez RC, Jourdan F, Kale N, Klapa MI, Kohlbacher O, Koort K, Kultima K, Le Corguillé G, Moreno P, Moschonas NK, Neumann S, O'Donovan C, Reczko M, Rocca-Serra P, Rosato A, Salek RM, Sansone SA, Satagopam V, Schober D, Shimmo R, Spicer RA, Spjuth O, Thévenot EA, Viant MR, Weber RJM, Willighagen EL, Zanetti G, Steinbeck C, The future of metabolomics in ELIXIR, F1000Res. 2017 Sep 6;6. pii: ELIXIR-1649. doi: 10.12688/f1000research.12342.2.
2. Moulos P, Hatzis P, Systematic integration of RNA-Seq statistical algorithms for accurate detection of differential gene expression patterns. Nucleic Acids Res. 2015 Feb 27;43(4):e25. doi: 10.1093/nar/gku1273
3. Paraskevopoulou MD, Georgakilas G, Kostoulas N, Vlachos IS, Vergoulis T, Reczko M, Filippidis C, Dalamagas T, Hatzigeorgiou AG, DIANA-microT web server v5.0: service integration into miRNA functional analysis workflows. Nucleic Acids Res. 2013 (Web Server issue):W169-73. doi: 10.1093/nar/gkt393
4. Vlachos IS, Kostoulas N, Vergoulis T, Georgakilas G, Reczko M, Maragkakis M, Paraskevopoulou MD, Prionidis K, Dalamagas T, Hatzigeorgiou AG, DIANA miRPath v.2.0: investigating the combinatorial effect of microRNAs in pathways, Nucleic Acids Res. 2012 Jul;40(Web Server issue):W498-504. doi: 10.1093/nar/gks494.
5. Reczko M, Maragkakis M, Alexiou P, Grosse I, Hatzigeorgiou AG,Functional microRNA targets in protein coding sequences, Bioinformatics. 2012 Mar 15;28(6):771-6. doi: 10.1093/bioinformatics/bts043.

List of top 5 most relevant previous projects or activities, connected to the subject of this proposal

1. ELIXIR-GR: The Greek Research Infrastructure for Data Management and Analysis in Life Sciences" (MIS 5002780), implemented under the Action "Reinforcement of the Research and Innovation Infrastructure", funded by the Operational Programme "Competitiveness, Entrepreneurship and Innovation" (NSRF 2014-2020) and co-financed by Greece and the European Union (European Regional Development Fund)
2. EU H2020 (H2020-INFRADEV-1-2015-1) ELIXIR-EXCELERATE: Fast-track ELIXIR implementation and drive early user exploitation across the life sciences, Grant number: 676559, 2015-2019.
3. NETWORK OF EXCELLENCE LSHG-CT-2005-005203-MUGEN Integrated Functional Genomics In Mutant Mouse Models As Tools To Investigate The Complexity Of Human Immunological Disease
4. COORDINATION ACTION LSHG-CT-2006-037811-CASIMIR Coordination and Sustainability of International Mouse Informatics Resources
5. COORDINATION ACTION FP7-HEALTH-2007-223592-IDCC The International Data Coordination Centre

Description of any significant infrastructure and/or major items of technical equipment, relevant to the proposed work (if applicable):

The BSRCAF Bioinformatics Facility central computing cluster consisting of 9 servers with a total of 148 CPU cores capable of running 224 threads, a total disk space of 600 TeraByte and total memory of 1670 Gigabyte.

Participant	**15 Athena Research and Innovation Centre. ATHENA (ELIXIR-GR)**	**Greece**
PI	[redacted] **(M)**	
ELIXIR Node	**ELIXIR-GR**	
Description of the legal entity/department/laboratory/group:		

The ATHENA Research and Innovation Center, founded under the auspices of the Greek Ministry of Development in 2001, is a research and technology body focused explicitly on Information and Communication Technologies. ATHENA participates in this project through the Information Management Systems Institute (IMSI). IMSI was founded in 2007, with the mission to conduct research and innovation activities in the area of big data management and large-scale information systems, i.e. the foundations of the data economy. This is apparent in the quantity and quality of scientific publications in these fields, as well as from the success in obtaining competitive funding from EU and national programs. More than 90% of IMIS budget comes from competitive funding. Due to the Greek EC presidency in 2014, IMSI was the co-organizer of the 2014 European Data forum, an annual meeting place for industry, research, policy makers, and community initiatives to discuss the challenges and opportunities of data in Europe, especially in the light of recent developments such as Big Data, Open Data, and Linked Data. The key personnel of IMSI have a long and pertinent experience in participating in and leading EU projects with an emphasis on topics for Big Data management, Cloud technologies and e-Infrastructures, and Open Data.

IMSI provides knowledge management and information systems technologies for large-scale knowledge management, data-intensive applications, and for a wide range of digital resources, and has strong theoretical and practical expertise in the following areas:

- scientific databases and e-infrastructures (large-scale life-science databanks and applications, data preservation technologies, data science technologies for precision medicine),
- data web (knowledge representation and ontologies, linked data, data provenance and preservation, user personalization, social networks),
- scholar data (semantic annotation and retrieval, data curation),
- eGovernment (open data, interoperability, data privacy and anonymization),
- green ICT (biomass data monitoring and management, big data management for large scale, real time water consumption data),
- geospatial data (GIS, spatiotemporal data management, Earth Observation data, spatial data infrastructures, traffic data management).

IMSI is partner of the distributed Greek ELIXIR Node.

IMSI has active collaboration ties with the Greek public administration on technical and policy issues relating to knowledge management, open data, crowdsourcing & citizen engagement, crisis management, and proprietary big data analysis for a number of thematic domains (health, environmental, geo data). Further, IMSI is supporting grassroots and volunteer-driven ICT solutions for open governance and open source software, by regularly organizing open events (e.g. hackathons, open data days, training seminars), contributing in citizen-driven initiatives, and engaging the scientific community with citizen associations

Main tasks in the project per WP;

ATHENA-RIC will contribute to the following tasks in **WP1 Expert network:**
- T1.1 Network of data managers and scientific best practice (Participant).

ATHENA-RIC will contribute to the following tasks in **WP3 Common Data Management Toolkit:**
- T3.2 Processes for enriching, maintaining and sustaining the Toolkit strators (Participant).
- T3.4 Best Practices and training (Participant).

Short Profile of key Staff Members who will be undertaking the work

[redacted]

[redacted]

List of the top 5 most relevant publications and/or products, services (including widely-used datasets or software), or other achievements

1. D. Karagkouni, M.Paraskevopoulou, S. Chatzopoulos, I. Vlachos, S. Tastsoglou, I. Kanellos, D. Papadimitriou, I. Kavakiotis, S. Maniou, G. Skoufos, T. Vergoulis, T. Dalamagas, and A. Hatzigeorgiou. Diana-TarBase v8: a decade-long collection of experimentally supported miRNA-gene interactions. Nucleic Acids Research, 46(D1), 2018.
2. Zagganas, T. Vergoulis, M. Paraskevopoulou, I. Vlachos, S. Skiadopoulos, and T. Dalamagas. BUFET: Boosting the unbiased miRNA functional enrichment analysis using bitsets. BMC Bioinformatics, 18(1), 2017.
3. Maria D. Paraskevopoulou, Ioannis S. Vlachos, Dimitra Karagkouni, Georgios Georgakilas, Ilias Kanellos, Thanasis Vergoulis, Konstantinos Zagganas, Panayiotis Tsanakas, Evangelos Floros, Theodore Dalamagas, and Artemis Hatzigeorgiou. Diana-LncBase v2: Indexing microRNA targets on non-coding transcripts. Nucleic Acids Research, 44(D1), 2016.
4. Olga Gkountouna, Manolis Terrovitis. Anonymizing collections of tree-structured data. 32nd IEEE International Conference on Data Engineering (ICDE), 2016.
5. Manolis Terrovitis, John Liagouris, Nikos Mamoulis, Spiros Skiadopoulos. Privacy Preservation by Disassociation. PVLDB, 5(10), pp 944-955, 2012.

List of top 5 most relevant relevant previous projects or activities, connected to the subject of this proposal

1. ELIXIR-GR: The Greek Research Infrastructure for Data Management and Analysis in Life Sciences (MIS 5002780), implemented under the Action "Reinforcement of the Research and Innovation Infrastructure", funded by the Operational Programme "Competitiveness, Entrepreneurship and Innovation" (NSRF 2014-2020) and co-financed by Greece and the European Union (European Regional Development Fund)
2. Hellenic Republic – Siemens Settlement Agreement: Hellenic Network for Precision Medicine, 2018-2020.
3. EU H2020 (H2020-INFRADEV-1-2015-1) ELIXIR-EXCELERATE: Fast-track ELIXIR implementation and drive early user exploitation across the life sciences, Grant number: 676559, 2015-2019.
4. Program COOPERATION (General Secretariat for Research and Technology, GR) MIKORNA: Exploring the role of MicroRNAs in diseases: a computational, experimentally-verified approach, 2011-2014
5. FP7 Research Infrastructures Program VENUS-C (subcontracted pilot): Targets on the Cloud: a cloud-based microRNA target prediction platform, 2011-2012

Description of any significant infrastructure and/or major items of technical equipment, relevant to the proposed work (if applicable):

N/A

Participant	**16 Centre for Research and Technology-Hellas. CERTH (ELIXIR-GR)**	**Greece**
PI	[redacted] **(M)**	
ELIXIR Node	**ELIXIR-GR**	

Description of the legal entity/department/laboratory/group:

The Centre for Research & Technology Hellas (CERTH) is one of the leading research centres in Greece, established in 2000. CERTH is a legal entity governed by private law with non-profit status, supervised by the General Secretariat for Research and Technology (GSRT) of the Greek Ministry of Education, Research and Religious Affairs. CERTH has important scientific and technological achievements in many areas including: Energy, Environment, Industry, Mechatronics (CPERI), Information & Communication (ITI), Transportation & Sustainable Mobility (HIT), Health, Agro-biotechnology, Smart farming, Safety & Security (INAB), as well as several cross-disciplinary areas. CERTH is essentially a self-supported Research Centre generating an average annual turnover of ~€25 Million from: >30% from bilateral industrial contracts, >60% from competitive projects and, <10% as government institutional funding. More than 700 people work at CERTH, with the majority being scientists. CERTH has received numerous awards and distinctions such as the European Descartes Prize, the European Research Council (ERC) Advanced Grant, the Microsoft International Contest Prize, the Trading Agents Competition Award and many more. It is consistently listed among the Top-20 of the EU's Research Centres in attracting competitive research grants. CERTH has participated successfully in more than 1,000 competitive research projects (with a total budget exceeding 423 M€ and involving more than 1,100 international partner organizations) financed by the European Union (EU), leading industries from the USA, Japan and Europe and the Greek Government via the General Secretariat of Research and Technology. CERTH is partner of the distributed Greek ELIXIR node.

The Biological Computation & Process Laboratory (BCPL – website: bcpl.wordpress.com) was established in 2014 as a business unit of CPERI, with the mission to perform high-performance computational analysis for biological systems engineering and develop applications for energy, environment, security, biochemical processes & synthetic biology. The strategy of BCPL is to establish unique collaborations with leading international industrial corporations, develop strong links with research centres and universities and contribute to the training of young scientists and engineers in state-of-the-art technologies. BCPL has access to compute servers and storage, expertise in multiple programming languages (C/C++/Lisp/Perl/Java/others), ongoing activities in database design (SQL) and data delivery including cloud technology (Azure/others), and algorithm implementations for genomics and systems biology. BCPL has participated in numerous training activities over the years, with future plans to expand on those with appropriate teaching, training and hands-on workshops in emerging areas of research. BCPL is also a member of the MetaSUB and Monumentome international consortia [metagenomics of built environments (MBE)]. BCPL-CPERI-CERTH will participate in INFRADEV3 activities in Greece, coordinated by BSRCAF, in collaboration with other colleagues at CERTH, e.g. ITI & INAB.

Main tasks in the project per WP;

CERTH will contribute to the following tasks in **WP2 Training and Capacity Building:**

- T2.1 Identify training needs and solutions in Data Management and Stewardship (Participant).

Short Profile of key Staff Members who will be undertaking the work

[redacted]

List of the top 5 most relevant publications and/or products, services (including widely-used datasets or software), or other achievements

1. Amann RI, Baichoo S, Blencowe BJ, Bork P, Borodovsky M, Brooksbank C, Chain PSG, Colwell RR, Daffonchio DG, Danchin A, de Lorenzo V, Dorrestein PC, Finn RD, Fraser CM, Gilbert JA, Hallam SJ, Hugenholtz P, Ioannidis JPA, Jansson JK, Kim JF, Klenk H-P, Klotz MG, Knight R, Konstantinidis KT, Kyrpides NC, Mason CE, McHardy AC, Meyer F, **Ouzounis** CA, Patrinos AAN, Podar M, Pollard KS, Ravel J, Muñoz AR, Roberts RJ, Rosselló-Móra R, Sansone S-A, Schloss PD, Schriml LM, Setubal JC, Sorek R, Stevens RL, Tiedje JM, Turjanski A, Tyson GW, Ussery DW, Weinstock GM, White O, Whitman WB, Xenarios I (2019) Towards unrestricted usage of public genomic data. Science 363, 350-352.
2. Hollmann S, Regierer B, D'Elia D, Gruden K, Baebler S, Frohme M, Pfeil J, Sezerman U, Evelo CT, Ehrhart F, Huppertz B, Bongcam-Rudloff E, Trefois C, Gruca A, Duca DA, Colotti G, Merino-Martinez R, **Ouzounis** C, He F, Hunewald O, Kremer A (2018) Standardisation in life-science research – making the case for harmonization to improve communication and sharing of data amongst researchers. ICT Innovations, 489.
3. **Karapiperis** C, Kempf SJ, Quintens R, Azimzadeh O, Vidal VL, Pazzaglia S, Bazyka D, Mastroberardino PG, Scouras ZG, Tapio S, Benotmane MA, **Ouzounis** CA (2016) Brain Radiation Information Data Exchange (BRIDE): integration of experimental data from low-dose ionising radiation research for pathway discovery. BMC Bioinformatics 17, 212.
4. **Kyritsis** KA, Angelis L, **Ouzounis** CA, Vizirianakis IS (2017) Understanding specialized ribosomal protein functions and associated ribosomopathies by navigating across sequence, literature and phenotype information resources. Leveraging Biomedical and Healthcare Data, 35-51.
5. Liechti R, George N, Götz L, El-Gebali S, **Chasapi** A, Crespo I, Xenarios I, Lemberger T (2017) SourceData: a semantic platform for curating and searching figures. Nat Methods 14, 1021-1022.

List of top 5 most relevant relevant previous projects or activities, connected to the subject of this proposal

1. **CHARME** COST project #042/15: "Harmonising standardisation strategies to increase efficiency and competitiveness of European life-science research", http://www.cost.eu/COST_Actions/ca/CA15110. Management Cttee Member, 3/2016-2/2020
2. **Cerebrad** FP7 Collaborative Project #295552: "Cognitive and Cerebrovascular Effects Induced by Low Dose Ionizing Radiation", Euratom (PI), 10/2011-2/2015
3. **Microme** FP7 Collaborative Project #222886-2, EC: "A resource for bacterial metabolism: large scale inference of metabolic flux from genome sequence" (PI-WP leader), 12/2009-12/2013
4. **Enfin** FP7 Network of Excellence #LSHG-CT-2005-518254, EC: "Europe-wide integration of computational approaches in systems biology" (PI), 10/2006-10/2011
5. **BioSapiens** FP6 Network of Excellence, EC: "A European Virtual Institute for Genome Annotation" (PI), 1/2004-1/2009

Description of any significant infrastructure and/or major items of technical equipment, relevant to the proposed work (if applicable):

The Biological Computation and Process Laboratory (BCPL) is equipped with compute servers 32GB RAM, 10TB storage, Gbps network infrastructure, peripherals and has access to very large HPC facilities elsewhere.

Participant	**17 Research Centre for Natural Sciences, TTK (ELIXIR-HU)**	**Hungary**
PI	[redacted] (M)	
ELIXIR Node	**ELIXIR-HU**	

Description of the legal entity/department/laboratory/group:

The Research Centre for Natural Sciences (TTK) is a full-time institution devoted solely to scientific research of the Hungarian research network. We endeavour to undertake a primary role in the cause of public welfare and founding the future, based upon domestic research traditions, by conducting discovery research promising significant achievements on an international scale and valuable scientific results. Focus areas of research in health science and a multi-disciplinary environment provided by the TTK have created optimal conditions for research conducted in the institutes of the research centre, in the following four fields of science: 1) biology (bioinformatics, biophysics, biochemistry, enzymology, genetics, molecular biology, cell biology, interpretation of physiological and pathophysiological processes on cell and- molecular level); 2) chemistry (heterocyclic compounds, carbohydrates, polymers and development of new methods of synthesis); 3) pharmacology and drug safety (analysis of drug effects on molecular level, experimental developments for the production of new, effective and safe drugs); and 4) cognitive neuroscience and psychology (societal-, comparative cross-cultural-, cognitive- and developmental psychology, and psychophysiology). Within the TTK, the Institute of Enzymology pursues research goals stretching across multiple fields of science, resulting in interdisciplinary research using the methods of biology, chemistry physics and informatics at the same time. Structural biological basic research of the institute is directed towards understanding physiological and pathophysiological processes on the scale of molecules and cells. Research topics are continuously extended from structural biology towards system biology to reveal complex biological processes by the utilization of advances in proteomics and bioinformatics. All together four research groups have a primary focus on bioinformatics, the Membrane Protein Bioinformatics Research Group, the Oncology Biomarker Research Group, the Metabolic Drug-interactions Research Group, and the Systems Biology of Reproduction Research Group.

Main tasks in the project per WP;

TTK will contribute to the following tasks in **WP1 Expert network:**
- T1.1 Network of data managers and scientific best practice (Participant).

TTK will contribute to the following tasks in **WP2 Training and Capacity Building:**
- T2.1 Identify training needs and solutions in Data Management and Stewardship (Participant).
- T2.2 Develop best practices guidelines and training materials in DMS (Participant).
- T2.3 Capacity Building in Data Management and Stewardship (Participant).

TTK will contribute to the following tasks in **WP3 Common Data Management Toolkit:**
- T3.2 Processes for enriching, maintaining and sustaining the Toolkit strators (Participant).
- T3.4 Best Practices and training (Participant).

TTK will contribute to the following tasks in **WP5 Demonstrator Projects:**
- T5.1 Typology of projects based on the type of resources needed to implement a Ma-DMP (Participant).
- T5.2 Implementation of pilots data management plans (Participant).

Short Profile of key Staff Members who will be undertaking the work

[redacted]

List of the top 5 most relevant publications and/or products, services (including widely-used datasets or software), or other achievements

1. Ősz Á, Pongor LS, Szirmai D, Győrffy B. A snapshot of 3649 Web-based services published between 1994 and 2017 shows a decrease in availability after 2 years. Brief Bioinform. 2017 Dec 8. doi: 10.1093/bib/bbx159.

2. Lánczky A, Nagy Á, Bottai G, Munkácsy G, Szabó A, Santarpia L, Győrffy B. miRpower: a web-tool to validate survival-associated miRNAs utilizing expression data from 2178 breast cancer patients. Breast Cancer Res Treat. 2016 Dec;160(3):439-446. doi: 10.1007/s10549-016-4013-7.
3. Bojcsuk D, Nagy G, Bálint BL. Inducible super-enhancers are organized based on canonical signal-specific transcription factor binding elements. Nucleic Acids Res. 2017 Apr 20;45(7):3693-3706. doi: 10.1093/nar/gkw1283.
4. Koehler C, Sauter PF, Wawryszyn M, Girona GE, Gupta K, Landry JJ, Fritz MH, Radic K, Hoffmann JE, Chen ZA, Zou J, Tan PS, Galik B, Junttila S, Stolt-Bergner P, Pruneri G, Gyenesei A, Schultz C, Biskup MB, Besir H, Benes V, Rappsilber J, Jechlinger M, Korbel JO, Berger I, Braese S, Lemke EA. Genetic code expansion for multiprotein complex engineering. Nat Methods. 2016 Dec;13(12):997-1000.
5. Nagy G, Czipa E, Steiner L, Nagy T, Pongor S, Nagy L, Barta E. Motif oriented high-resolution analysis of ChIP-seq data reveals the topological order of CTCF and cohesin proteins on DNA. BMC GENOMICS 17:(1) p. 637. (2016)

List of top 5 most relevant relevant previous projects or activities, connected to the subject of this proposal

1. H2020-MSCA-COFUND 754432 ImPRESS: International Interdisciplinary PhD studies in Biomedical Research and Biostatistics, 09/2017 – 08/2022
2. Debrecen Venture Catapult, EFOP 3.6.1 : Data Intensive and Open Science School (2017-2021)
3. FIEK_16, Establishing molecular biomarker research and service centre, 2017-2021, Budget: 9,441,816 EUR
4. EB is node manager of the Hungarian EMBnet between 1994-2012 and is Head of Genomics Department at the NARIC ABC since 2018
5. 2018-1.3.1-VKE-2018-00032, Photoscreen: development of new pathology diagnostic tools, 2019-2021, Budget: 4,235,019 EUR

Description of any significant infrastructure and/or major items of technical equipment, relevant to the proposed work (if applicable):

SGI ICE-X high performance computer cluster with 2800 cores, 31.4 TB RAM and 290TB storage.

Participant	**18 University College Dublin. UCD (ELIXIR-IE)**	**Ireland**
PI	[redacted] **(M)**	
ELIXIR Node	**ELIXIR-IE**	

Description of the legal entity/department/laboratory/group:

University College Dublin, (UCD) is the largest university in Ireland and supports the full range of academic disciplines. UCD has over 30,000 students and is ranked 168th worldwide by the QS World University Rankings 2018 and 73rd in Europe. UCD has made significant investments over the last 5 years in systems biology, computational medicine and personalized medicine. Systems Biology Ireland (SBI) was established in UCD in 2009 and is a multidisciplinary center that combines both wet lab and dry lab techniques in order to design new therapeutic approaches to cancer and inflammatory diseases. The center is located in a newly constructed, state of the art facility that accommodates over 60 researchers, using state of the art techniques in computing, mathematics, statistics, omics technologies, imaging, cell and molecular biology to study biological networks. For this, SBI merges predictive mathematical and computational modelling and bioinformatics with experimental research to gain a systems level, mechanistic understanding of cellular signal transduction networks. The SBI facility boasts top class laboratories, specifically designed to meet the needs of interdisciplinary researchers in this challenging and emerging area of research. Under the leadership of [redacted], SBI has become a flagship program in UCD and Ireland, and an internationally recognized leader in the space with a recent assessment by International Scientific Panel judging it to be "one of the top Systems Biology Institutes in the world". SBI has recently signed a memorandum of understanding with EMBL for the development of an EMBL associated unit in Ireland focusing on the development of personalized systems medicine approaches in cancer. Teaming up with clinical colleagues, this work is now spearheading our strategic program to establish personalized systems medicine approaches in oncology.

Main tasks in the project per WP;

UCD will contribute to the following tasks in WP1 Expert network:

- T1.1 Network of data managers and scientific best practice (Participant).

UCD will contribute to the following tasks in WP2 Training and Capacity Building:

- T2.1 Identify training needs and solutions in Data Management and Stewardship (Participant).
- T2.2 Develop best practices guidelines and training materials in DMS (Participant).
- T2.3 Capacity Building in Data Management and Stewardship (Participant).

UCD will contribute to the following tasks in WP4 Communications, Industry, International, Impact and Sustainability:

- T4.2 Operating the ELIXIR Innovation and SME Forum and enhancing implementation of national industry engagement efforts (Participant).
- T4.4 Implementing an impact assessment toolkit for demonstrating ELIXIR's value, nationally and at European level (Participant).

Short Profile of key Staff Members who will be undertaking the work

[redacted]

List of the top 5 most relevant publications and/or products, services (including widely-used datasets or software), or other achievements

1. Halasz, M., Kholodenko, B. N., Kolch, W., & Santra, T. (2016). Integrating networkreconstruction with mechanistic modelling to predict cancer therapies. Sci Signal, 9(455), ra114.
2. Kolch W, Halasz M, Granovskaya M, Kholodenko BN. (2015) The dynamic control of signal transduction networks in cancer cells. Nature reviews Cancer 15 (9):515-527.
3. Romano D, Nguyen LK, Matallanas D, Halasz M, Doherty C, Kholodenko BN and Kolch W. (2014) Nat Cell Biol. Protein interaction switches coordinate Raf-1 and MST2/Hippo signalling. 16, 673-684.
4. Fey D, Halasz M, Dreidax D, Kennedy SP, Hastings JF, Rauch N, Munoz AG, Pilkington R, Fischer M, Westermann F, Kolch W, Kholodenko BN, Croucher DR. (2015) Signaling pathway models as

biomarkers: Patient-specific simulations of JNK activity predict the survival of neuroblastoma patients. Sci Signal 8 (408): ra130.

5. Campbell, J., Ryan, C. J., Brough, R., Bajrami, I., Pemberton, H. N., Chong, I. Y., Costa-Cabral, S., Frankum, J., Gulati, A., Holme, H. Miller, R., Postel-Vinay, S., Rafiq, R., Wei, W., Williamson, C. T. Quigley, D. A., Tym, J., Al-Lazikani, B., Fenton, T., Natrajan, R., Strauss, S. J., Ashworth, A., Lord, C. J. (2016) Large-Scale Profiling of Kinase Dependencies in Cancer Cell Lines, Cell Reports, 14(10), 2490-2501.

List of top 5 most relevant relevant previous projects or activities, connected to the subject of this proposal

1. **MelPlex.** H2020 Innovative Training Network (ITN), grant no. 642295 (01/12/2014-30/11/2018). The project trained PhD students in research for personalised melanoma diagnostics and therapy. (Walter Kolch) Role: PI and WP Leader. Total value ca. €3.6 Mi, UCD-SBI share €247,675.
2. **ASSET**: Analyzing and Striking the Sensitivities of Embryonal Tumors. EU-FP7 Health, grant no. 259348. (Walter Kolch) Role: Coordinator. Total value ca. €12 Mi, UCD-SBI share €570,694.
3. **PRIMES**: Protein Interaction Machines in Oncogenic EGF Receptor Signaling. EU-FP7 Health, grant no.. (Walter Kolch) Role: Coordinator. Total value ca. €12 Mi, UCD-SBI share €570,694.
4. "Systems Biology Ireland (SBI)" **Centre for Science Engineering and Technology Grant** award. Science Foundation Ireland, Grant no. 06/CE/B1129 (01/06/2009-31/05/2015). This award provided the start-up funding for establishing SBI. (Walter Kolch) Role: Lead PI and Director. Total value ca. €15 Mi.
5. "The contribution of paralog buffering to tumour robustness" **Starting Laureate Award**, Irish Research Council. (Colm Ryan) Role: PI. Total value €400,000, UCD-SBI share €400,000.

Description of any significant infrastructure and/or major items of technical equipment, relevant to the proposed work (if applicable):

The proposed research will have full access to the suite of state-of-the-art technology core facilities at SBI and the adjacent Conway Institute which comprise next generation sequencing (Illumina), proteomics, mass spectrometry (orbitraps, Q-Tof, Triple-Quad), flow cytometry, digital histopathology, and bioinformatics. The research facilities consist of a number of technology platforms maintained and managed by a cohort of expert staff who provides technical support and will support the research activities of the proposed project. The UCD Daedalus Centre provides high end IT research service including storage, servers, and high-performance computing (www.ucd.ie/itservices/researchit/). For very heavy-duty computation we have access to the Irish Centre for High-End Computing (ICHEC), a national high-performance computer centre with two supercomputers and associated services (www.ichec.ie/).

Participant	**19 Weizmann Institute of Science. WEIZMANN (ELIXIR-IL)**	**Israel**
PI	[redacted] **(M)**	
ELIXIR Node	**ELIXIR-IL**	

Description of the legal entity/department/laboratory/group:

WEIZMANN will coordinate the ELIXIR-IL Node as the legal entity. The Weizmann Institute of Science is one of the world's leading multidisciplinary research institutions. The Institute has five faculties, Mathematics and Computer Science, Physics, Chemistry, Biochemistry and Biology, which are divided into 17 scientific departments. The Feinberg Graduate School, the Institute's university arm, trains research students pursuing graduate degrees.

The Weizmann Institute serves as a meeting place for scientists from different disciplines, setting the stage for multidisciplinary collaborations and the emergence of new research fields. To encourage this creative activity, the Institute has created some 50 multidisciplinary research institutes and centers that stimulate activity in a multiplicity of fields. The research infrastructure at the Institute is comparable to that of the top institutes in the world. Examples include the Lorry I. Lokey Pre- Clinical Research Facility and the Nancy and Stephen Grand Israel National Center for Personalized Medicine (G-INCPM).

Numerous international conferences and symposia are held in the Institute, as well as a wide range of other cultural and educational activities aimed to encourage greater interest in science among school and college students as well as the general public.

Today the campus community numbers more than 2,700 including some 1,300 scientists and scientific staff, 1,400 research students and postdoctoral fellows. The Institute has some 250 research groups headed by senior scientists and professors who were either born in Israel or have emigrated from numerous countries around the globe. Each year, Weizmann Institute hosts around 500 scientists from around the globe that visit the Institute or come to work on its campus.

Several faculty members in WIS are leaders in the field of system biology, structural Bioinformatics and tool development. Established in 2012 by the Weizmann Institute of Science, the Nancy & Stephen Grand Israel National Center for Personalized Medicine (G-INCPM) is an advanced research facility providing Israeli academic, medical, and biomedical industry researchers with access and guidance to state-of-the-art genomics, protein profiling, drug discovery and bioinformatics research platforms.

The G-INCPM can:

- Create scalable, robust, cutting-edge infrastructure capable of supporting national activities in the G-INCPM's fields of expertise.
- Provide services and collaboration opportunities with a focus on high-value projects.
- Instil a spirit of innovation, enabling the creation of novel research tools and methods, and developing best practices.
- Adapt cutting-edge research methods developed in Weizmann Institute laboratories and elsewhere, for broad usage.
- Disseminate new research technologies and share relevant knowledge with the biomedical community.

The Israeli Elixir Node will be physically located at the G-INCPM at Weizmann Institute in Rehovot, but will keep an active network that will incorporate scientists and resources for all other universities in Israel. Israel has a unique mixed population structure that is of interest in studying genetic diseases. This includes, for example, diseases that are prevalent in Ashkenazy Jews or in Arab/Bedouin populations.

Main tasks in the project per WP;

WEIZMANN will contribute to the following tasks in **WP1 Expert network:**

- T1.1 Network of data managers and scientific best practice (Participant).

WEIZMANN will contribute to the following tasks in **WP2 Training and Capacity Building:**

- T2.1 Identify training needs and solutions in Data Management and Stewardship (Participant).
- T2.3 Capacity Building in Data Management and Stewardship (Participant).
- T2.4 Outreach activities to new ELIXIR Members and Communities (Participant).

WEIZMANN will contribute to the following tasks in **WP3 Common Data Management Toolkit:**

- T3.2 Processes for enriching, maintaining and sustaining the Toolkit strators (Participant).
- T3.4 Best Practices and training (Participant).

WEIZMANN will contribute to the following tasks in **WP4 Communications, Industry, International, Impact and Sustainability:**

- T4.2 Operating the ELIXIR Innovation and SME Forum and enhancing implementation of national industry engagement efforts (Participant).

- T4.4 Implementing an impact assessment toolkit for demonstrating ELIXIR's value, nationally and at European level (Participant).

WEIZMANN will contribute to the following tasks in **WP5 Demonstrator Projects:**

- T5.1 Typology of projects based on the type of resources needed to implement a Ma-DMP (Participant).

- T5.4 Capacity building actions based on pilots outcomes (Participant).

Short Profile of key Staff Members who will be undertaking the work

[redacted]

List of the top 5 most relevant publications and/or products, services (including widely-used datasets or software), or other achievements

1. Ben-Avraham D, Govindaraju DR, Budagov T, Fradin D, Durda PJ, Liu B, Ott S, Gutman D, Sharvit L, Kaplan R, Bougnères P, Reiner A, Shuldiner AR, CohenP, Barzilai N, Atzmon G. 2017. The GH receptor exon-3 deletion is a marker of male-specific exceptional longevity associated with increased GH sensitivity and taller stature. Sci Adv. Jun16;3(6)
2. Sebastiani P, Gurinovich A, Bae H, Andersen S, Malovini A, Atzmon G, Villa F, Kraja AT, Ben-Avraham D, Barzilai N, Puca A, Perls TT. 2017. Four Genome-Wide Association Studies Identify New Extreme Longevity Variants. J Gerontol A Biol Sci Med Sci. Mar 15. doi: 10.1093/gerona/glx027
3. Ben-Avraham D, Karasik D, Verghese J, Lunetta KL, Smith JA, Eicher JD, Vered R, Deelen J, Arnold AM, Buchman AS, Tanaka T, et al. 2017. The complex genetics of gait speed: genome-wide meta-analysis approach. Aging (Albany NY). 2017 Jan 10;9(1):209-246

4. Zidan J, Ben-Avraham D, Carmi S, Maray T, Friedman E, Atzmon G. 2015. Genotyping of geographically diverse Druze trios reveals substructure and a recent bottleneck. Eur J Hum Genet. Aug;23(8):1093-9
5. Carmi S, Hui K, Kochav E, Liu X, Xue J, Grady F, Guha S, Upadhyay K, Ben-Avraham D, Mukherjee S, Bowen MB, Joseph V, Cruts M, et al. 2014. Whole genome sequencing of an Ashkenazi Jewish panel supports population-targeted personal genomics and illuminates Jewish and European origins. Nat Commun. Sep9;5:4835

List of <u>top 5 most relevant</u> relevant previous projects or activities, connected to the subject of this proposal

1. EU H2020 (H2020-INFRADEV-1-2015-1) ELIXIR-EXCELERATE: Fast-track ELIXIR implementation and drive early user exploitation across the life sciences, Grant number: 676559, 2015-2019.
2. INCPMPM-4283: Relapsed AML- mechanisms and drug efficacy
3. INCPMPM-3646: Hybrid assembly workflow with long and short reads improves discovery of structural variations in evolved bacteria
4. INCPMPM-3914: FuPeXI: prediction of tumor neoantigens from fusion transcripts
5. INCPMPM-4790: Longitudinal transcriptomic patterns through the course of breast cancer disease and therapy in FFPE samples

Description of any significant infrastructure and/or major items of technical equipment, relevant to the proposed work (if applicable):

G-INCPM's IT Platform possesses enabling-technologies that are integrated into a seamless organization with strong ties to academic and clinical researchers, thus creating scientific momentum that extends beyond technological boundaries. It provides resources to support high-end computing across a broad range of domains. Our goal is to enable researchers to leverage our technologies to reduce time to discovery and move their science forward.

IT High level Infrastructure Design:
- SGI UV2000 GPU Supercomputer.
- High volume storage (2.5 petabyte on-line).
- High performance computing cluster (340 CPU, 2.6TB RAM, 20 compute nodes).
- Wide (private cloud) virtual environment (160 CPU core, 1.5 TB RAM).
- Ultra-fast & scalable Backbone network.
- Internal G-INCPM research platforms (Proteomics, Bioinformatics, HTS, Genomics and Medicinal Chemistry).

The team of programmers and bioinformaticians at the Mantoux Institute for Bioinformatics provides computational solutions for the analysis of biological data. The Institute uses state-of-the-art Bioinformatics analysis tools and is responsive to the needs of researchers, evolving with advances in the field. The team develops analysis tools and pipelines for NGS, Proteomics and Chemoinformatics utilizing the computing power afforded by the high performance computing environment built at the GINCPM.

The Crown Institute for Genomics provides next-generation sequencing services using the Illumina technology platform and sample preparation, using in-house and various commercial protocols. The institute offers a modular, albeit all-inclusive service, starting from project design, consultation over sample preparation type, sequencing and data analysis by the Center's bioinformaticians. The Crown Institute for Genomics invests much of its effort in the development of new in-house protocols, adjusting manual protocols to automated ones, in order to generate high throughput, reproducible sample data.

The de Botton Institute for Protein Profiling unit analyses proteins either intact or after enzymatic digestion. The proteins or peptides are then subjected to liquid chromatography coupled to high resolution, high mass accuracy mass spectrometry. Postacquisition, the protein identity and quantity are reconstructed using advanced bioinformatics. Typical experiments include whole protein expression, protein-protein interactions, global analysis of certain post translational modifications (PTMs) and targeted experiments.

The High Throughput Screening (HTS) Platform in the Maurice and Vivienne Wohl Institute for Drug Discovery is capable of screening cellular and/or biochemical assays at rates up to 200,000 chemical entities per screening campaign. This platform has the ability to screen an entire compound library or subsets of compounds in assays with a defined molecular target, such as enzymes, receptors, protein:protein interactions, reporter gene assays, or in phenotypic assays where the molecular target is not known. The facility is equipped to deal with needs ranging from high-throughput screening of chemical compounds on complex cellular readouts, to functional testing of a limited subset of molecules in different concentrations or combinations in a defined model. With the addition of the Medicinal Chemistry Unit, the team is able to support compound characterisation in an iterative manner in order to develop structure-activity relationship (SAR) for newly purchased or designed chemical matter.

Participant	**20 Consiglio Nazionale delle Ricerche. CNR (ELIXIR-IT)**	**Italy**
PI	[redacted] **(M)**	
ELIXIR Node	**ELIXIR-IT**	

Description of the legal entity/department/laboratory/group:

Consiglio Nazionale delle Ricerche (National Research Council, CNR; establ. in 1923), with more than 8000 employees, is the largest multidisciplinary public research institution of Italy. its duty is to carry out, promote, spread, transfer and improve research activities in the main sectors of knowledge growth and of its applications for the scientific, technological, economic and social development of the Country. CNR is distributed all over Italy through a network of 102 institutes aiming at promoting a wide diffusion of its competences throughout the national territory and at facilitating contacts and cooperation with local firms and organisations. The project includes the participation of one CNR Institute, CNR-IBIOM, with expertise in the cloud porting of workflow managers for bioinformatics.

CNR-IBIOM is in charge of the management of the Italian node of ELIXIR, led by CNR and participated by other 22 Institutions through a Joint Research Unit agreement. The Italian ELIXIR Node offers a vast portfolio of high-quality bioinformatics services, tools and data resources for the life science community. Publicly available bioinformatics services are often complemented with downloadable, stand-alone tools used in structural bioinformatics, genomics and proteomics. The Node hosts several flagship data resources and databases. Computational resources, under the form of HPC and cloud computing services, are operational and available to international users. Bioinformatics services are complemented with extensive training activities (courses, workshops) oriented to the international life science user community.

Main tasks in the project per WP;

CNR (ELIXIR-IT) will contribute to the following tasks in WP1 Expert network:

- T1.1 Network of data managers and scientific best practice (Participant).

- T1.3 Business model (**Task Leader**).

- T1.4 Sustainable and scalable operating model for harmonised data management in European projects (Participant).

- **ELIXIR-IT** will lead the collection of information regarding different business models for data management. It will also contribute to the network of data managers and sustainable operating model.

CNR (ELIXIR-IT) will contribute to the following tasks in WP2 Training and Capacity Building:

- T2.1 Identify training needs and solutions in Data Management and Stewardship (Participant).

- T2.2 Develop best practices guidelines and training materials in DMS (Participant).

- T2.3 Capacity Building in Data Management and Stewardship (Participant).

- **ELIXIR-IT** will design, develop, deliver and assess a Pilot Pan-European 2nd level Master in Data science and Data Stewardship and Management for the Life Sciences, building on the experience and expertise of ELIXIR nodes and other WPs. The master will include a set of teaching modules that can be attended as independent training events, plus a period of internship and a final exam for those who wish to get a certification.

CNR (ELIXIR-IT) will contribute to the following tasks in WP3 Common Data Management Toolkit:

- T3.4 Best Practices and training (Participant).

  - **ELIXIR-IT** will contribute to the dissemination of the Common Data Management Toolkit (Task 4) among the Italian research community in close collaboration with WP2 and the **ELIXIR-IT** training platform. At least one dedicated training event will be organised and feedback from researchers will be collected.

CNR (ELIXIR-IT) will contribute to the following tasks in WP4 Communications, Industry, International, Impact and Sustainability:

- T4.2 Operating the ELIXIR Innovation and SME Forum and enhancing implementation of national industry engagement efforts (Participant).

  - **ELIXIR-IT** will facilitate industry engagement, doing local outreach events to understand industry needs, and will support the preparation of a wiki guide on how to engage industry.

- T4.3 Engaging potential new Member countries and enhancing international visibility (Participant).

  - **ELIXIR-IT** will support the preparation of DMP training events, in collaboration with T2.4, scoping industry needs and promoting internship as part of the training

- T4.4 Implementing an impact assessment toolkit for demonstrating ELIXIR's value, nationally and at European level (Participant).

  - **ELIXIR-IT** will understand the needs in terms of demonstrating impact to different local stakeholders and collaborate to co-design the impact assessment toolkit, building on the RI-PATHS handbook.

CNR (ELIXIR-IT) will contribute to the following tasks in WP5 Demonstrator Projects:

- T5.1 Typology of projects based on the type of resources needed to implement a Ma-DMP (Participant).

- T5.2 Implementation of pilots data management plans (Participant).

  - **ELIXIR-IT** will contribute to a demonstrator on epitranscriptomics developing a reference portal for cataloguing and investigating main epitranscriptome modifications. The demonstrator will integrate and connect existing resources, standardizing computational protocols for storing and detecting RNA modifications.

Short Profile of <u>key</u> Staff Members who will be undertaking the work

[redacted]

[redacted]

List of the top 5 most relevant publications and/or products, services (including widely-used datasets or software), or other achievements

1. Ison J, et al. (2016) Tools and data services registry: a community effort to document bioinformatics resources. Nucleic Acids Res. 44(D1):D38-47. doi: 10.1093/nar/gkv1116.
2. Via A, Attwood TK, Fernandes PL, Morgan SL, Schneider MV, Palagi PM, Rustici G and Tractenberg RE (2017) A new pan-European Train-the-Trainer Programme for bioinformatics: Pilot results on feasibility, utility and sustainability of learning. Brief Bioinform. doi: 10.1093/bib/bbx112
3. Morgan SL, Palagi PM, Fernandes PL, Koperlainen E, Dimec J, Marek D, Larcombe L, Rustici G, Attwood TK, Via A (2017) The ELIXIR-EXCELERATE Train-the-Trainer pilot programme: empower researchers to deliver high-quality training F1000Research 2017, 6:1557. doi: 10.12688/f1000research.12332.1.
4. Jiménez RC et al. (2017) Four simple recommendations to encourage best practices in research software. F1000Research. doi: http://dx.doi.org/10.12688/f1000research.11407.1
5. Orchard S, et al (2014) The MIntAct project--IntAct as a common curation platform for 11 molecular interaction databases.Nucleic acids research. 42(Database issue):D358-63. doi: 10.1093/nar/gkt1115.
6. Lo Surdo P, Calderone A, Iannuccelli M, Licata L, Peluso D, Castagnoli L, Cesareni G, Perfetto L. DISNOR: a disease network open resource. Nucleic Acids Res. 2018 Jan 4:46(D1):D527-D534. doi: 10.1093/nar/gkx87
7. Mitchell AL, et al. (2019) InterPro in 2019: improving coverage, classification and access to protein sequence annotations. Nucleic Acids Res. 47 (D1), pp. D351-D360.
8. El-Gebali S.et al. (2019) The Pfam protein families database in 2019. Nucleic Acids Res, 47 (D1), pp. D427-D432.

List of top 5 most relevant relevant previous projects or activities, connected to the subject of this proposal

1. EU H2020 (H2020-INFRADEV-1-2015-1) ELIXIR-EXCELERATE: Fast-track ELIXIR implementation and drive early user exploitation across the life sciences, Grant number: 676559, 2015-2019.

2. EU-JPI ENPADASI, European Nutritional Phenotype Assessment and Data Sharing Initiative (ENPADASI) will deliver the Data Sharing In Nutrition (DASH-IN) infrastructure, 2014-2016
3. EU-H2020 (H2020-INFRASUPP-2014-2) GLOBIS-B - GLOBal Infrastructures for Supporting Biodiversity research (Grant number: 654003), 2015-2018
4. EU-FP7 (ERC-2013-CoG) HomeoGUT - Immune mechanisms that control the homeostasis of the gut and that are deregulated in intestinal pathologies, Grant number; 615735, 2014-2019
5. EU-H2020 (H2020-EINFRA-2014-2) INDIGO DataCloud - INtegrating Distributed data Infrastructures for Global ExplOitation, Grant Number: 653549, 2015-2018

Description of any significant infrastructure and/or major items of technical equipment, relevant to the proposed work (if applicable):
ELIXIR-IT infrastructure will be employed to carry out the task assigned to the Node. In particular the ELIXIR-IT Training and Data platforms, together with the ELIXIR-IT Industry officer, will be involved in the activities of this project.

Participant	**21 University of Luxembourg. UNILU (ELIXIR-LU)**	**Luxembourg**
PI	[redacted] **(M)**	
ELIXIR Node	**ELIXIR-LU**	

Description of the legal entity/department/laboratory/group:
University of Luxembourg hosts the Luxembourg Centre for Systems Biomedicine (LCSB), an interdisciplinary research centre that accelerates biomedical research by closing the link between systems biology and medical research. The Bioinformatics Core at LCSB, led by [redacted] is responsible for the efficient data flow between the experimental groups and the theoretical and medical oriented groups. It focuses on the development of (clinical) data management and storage systems and set up of cost- and time-efficient data analysis pipelines. The group develops new algorithms in various fields including data mining and visualisation to help to understand and interpret the data. Together with the high-performance computing group of the University they are responsible to setup and run large computer and storage facilities. high-performance computing group of the University they are responsible to setup and run large computer and storage facilities.

Main tasks in the project per WP;

UNILU will contribute to the following tasks in **WP1 Expert network:**

- T1.1 Network of data managers and scientific best practice (Participant).

  - Contribute to the development of data management guidance, templates and recommended KPIs

UNILU will contribute to the following tasks in **WP2 Training and Capacity Building:**

- T2.1 Identify training needs and solutions in Data Management and Stewardship (Participant).
- T2.2 Develop best practices guidelines and training materials in DMS (Participant).
- T2.3 Capacity Building in Data Management and Stewardship (Participant).

  - Contribute to the development of training material and courses as well as to deliver trainings for data management of sensitive research data.

UNILU will contribute to the following tasks in **WP3 Common Data Management Toolkit:**

- T3.1 Establish a Starter Toolkit (Participant).
- T3.4 Best Practices and training (**Task Leader**).

  - Contribute to the development of the starter toolkit for sensitive research data (GDPR tools)

- Lead the development of best practices and training of the toolkits.

UNILU will contribute to the following tasks in **WP5 Demonstrator Projects:**

- T5.2 Implementation of pilots data management plans (Participant).

- T5.3 Development,implementation and refinement of key performance indicators to monitor the pilots implementation of data management plans (Participant).

- Contribute to the implementation of data management plans and key performance indicators of the pilots projects involved

UNILU will contribute to the following tasks in **WP7 Federated European Genome-phenome Archives for transnational access of COVID-19 host data:**

- Task 7.4: Operational support and maturity model for Federated EGA nodes (Participant).

Short Profile of key Staff Members who will be undertaking the work

[redacted]

[redacted]

List of the top 5 most relevant publications and/or products, services (including widely-used datasets or software), or other achievements

1. SmartR: an open-source platform for interactive visual analytics for translational research data. Herzinger S, et al. (2017), Bioinformatics 33 (14), 2229-2231; doi: 10.1093/bioinformatics/btx137.
2. Integration and visualization of translational medicine data for better understanding of human diseases. Satagopam V, et al. (2016), Big Data 4 (2), 97-108; doi: 10.1089/big.2015.0057.
3. Making sense of big data in health research: Towards an EU action plan, Auffray C, et al. (2016), Genome medicine 8 (1), 71; doi: 10.1186/s13073-016-0323-y.

4. Provenance-enabled stewardship of human data in the GDPR era, Alper P, et al. (2018), F1000Research. doi: 10.7490/f1000research.1115768.1.
5. MINERVA—a platform for visualization and curation of molecular interaction networks, Gawron, P, et al (2016), npj Systems Biology and Applications. doi: 10.1038/npjsba.2016.20.

List of <u>top 5 most relevant</u> previous projects or activities, connected to the subject of this proposal

1. SYSCID: A Systems medicine approach to chronic inflammatory disease, H2020, 2017-2022
2. SysMedPD: Systems Medicine of Mitochondrial Parkinson's Disease, H2020, 2015-2019
3. bioCOG: Biomarker Development for Postoperative Cognitive Impairment in the Elderly, FP7, 2014-2019
4. FAIRplus: FAIRplus: FAIRification of IMI and EFPIA data, H2020, 2019-2021
5. CHARME: Harmonising standardisation strategies to increase efficiency and competitiveness of European life-science research, COST, 2016-2020

Description of any significant infrastructure and/or major items of technical equipment, relevant to the proposed work (if applicable):

N/A

Participant	**22 DTL Projects Foundation. DTL (ELIXIR-NL)**	**Netherlands**
PI	[redacted] **(F)**	
ELIXIR Node	**ELIXIR-NL**	

Description of the legal entity/department/laboratory/group:

The DTL Projects Foundation ('Stichting DTL Projects', hereafter: 'DTL Projects') is an independent project organisation that supports collaborative projects that fit the scope of the collaborative DTL, the Dutch Techcentre for Lifesciences (www.dtls.nl): Data, Technologies and Learning in the broader life sciences domain. DTL Projects offers a formalised structure to establish collaborative multi-party projects that foster advancement in dealing with data in research and innovation. DTL Projects is the formal host of ELIXIR-NL, the Dutch Node in the pan-European research infrastructure ELIXIR. DTL Projects runs projects facilitating community building, coordination in technology and infrastructure development, standardisation as well as capacity building through training and education. DTL is one of the first organisations where FAIR data and FAIR data stewardship were developed.

Main tasks in the project per WP:

DTL-Projects will contribute to the following tasks in **WP1 Expert network:**

- T1.1 Network of data managers and scientific best practice (Participant).

- T1.4 Sustainable and scalable operating model for harmonised data management in European projects (Participant).

DTL-Projects will contribute (**WP Leader**) to the following tasks in **WP2 Training and Capacity Building:**

- T2.1 Identify training needs and solutions in Data Management and Stewardship (Participant).

- T2.2 Develop best practices guidelines and training materials in DMS (**Task Leader**).

- T2.3 Capacity Building in Data Management and Stewardship (Participant).

- T2.4 Outreach activities to new ELIXIR Members and Communities (Participant).

DTL-Projects will contribute to the following tasks in **WP3 Common Data Management Toolkit:**

- T3.3 Access portal to Toolkit tailored to stakeholders (**Task Leader**).

- T3.4 Best Practices and training (Participant).

DTL-Projects will contribute to the following tasks in **WP4 Communications, Industry, International, Impact and Sustainability:**

- T4.2 Operating the ELIXIR Innovation and SME Forum and enhancing implementation of national industry engagement efforts (Participant).

DTL-Projects will contribute to the following tasks in **WP7 Federated European Genome-phenome Archives for transnational access of COVID-19 host data:**

- Task 7.3: Coordination of metadata standards for phenotype submission and access of COVID-19 data (Participant).

Short Profile of key Staff Members who will be undertaking the work

[redacted]

List of the top 5 most relevant publications and/or products, services (including widely-used datasets or software), or other achievements

1. Bioinformatics in the Netherlands: the value of a nationwide community. Briefings in bioinformatics (2017). van Gelder CWG, Hooft RWW , van Rijswijk MN , van den Berg L , Kok RG, Reinders M, Mons B , Heringa J DOI: 10.1093/bib/bbx087
2. Developing a strategy for computational lab skills training through Software and Data Carpentry: Experiences from the ELIXIR Pilot action. Pawlik A, van Gelder CWG, Nenadic A, Palagi PM, Korpelainen E, Lijnzaad P, Marek D, Sansone SA, Hancock J & Goble C (2017) . F1000 Research (2017). DOI:10.12688/f1000research.11718.1

3. The Dutch Techcentre for Life Sciences: Enabling data-intensive life science research in the Netherlands. Eijssen, L, Evelo, C, Kok, R, Mons, B, Hooft, R. F1000 Research (2016). DOI: 10.12688/f1000research.6009.2
4. The FAIR Guiding Principles for scientific data management and stewardship, Mark Wilkinson, et al. Scientific Data (2016) DOI: 10.1038/sdata.2016.18.
5. "Data Stewardship Wizard": A Tool Bringing Together Researchers, Data Stewards, and Data Experts around Data Management Planning. Pergl, R., Hooft, R., Suchánek, M., Knaisl, V., and Slifka, J. Submitted to Data Science Journal.

List of <u>top 5 most relevant</u> relevant previous projects or activities, connected to the subject of this proposal

1. EU H2020 (H2020-INFRADEV-1-2015-1) **ELIXIR-EXCELERATE:** Fast-track ELIXIR implementation and drive early user exploitation across the life sciences, Grant number: 676559, 2015-2019.
2. Co-development together with partners at ELIXIR-CZ of the **Data Stewardship Wizard**, an activity we undertake as a support to many DTL projects and partners. https://ds-wizard.org/ . Parts of this work have been funded by an ELIXIR implementation study.
3. ZonMw project "Towards a community-endorsed data steward profession description for Life Science research" https://zenodo.org/communities/nl-ds-pd-ls

Description of any significant infrastructure and/or major items of technical equipment, relevant to the proposed work (if applicable):

N/A

Participant	**23 Universitetet i Bergen. UiB (ELIXIR-NO)**	**Norway**
PI	[redacted]**M)**	
ELIXIR Node	**ELIXIR-NO**	

Description of the legal entity/department/laboratory/group:

The University of Bergen (UiB) is a young, modern university with about 14 000 students and 3,400 faculty and staff, making it a medium sized European University. Six faculties cover most of the traditional university disciplines. Within the faculties are included 60 different specialized departments, multi-disciplinary research centers and institutes. The University is engaged in the European Union's Framework programmes for research and technological development and has been designated as a European Research Infrastructure and a Research Training Site in several scientific fields. Since 1997, more than 500 European researchers (professors, senior researchers, post-docs and PhD candidates) have visited Bergen on EU grants, making Bergen one of the most international universities, setting out to attract both established and junior scientists to contribute to research teams and work in multidisciplinary research groups. UiB is currently involved in 105 FP7 projects, 36 of which it coordinates, among them UiB hosts 7 ERC Advanced Grants and 1 ERC Starting Grant and 1 ERC Synergy Grant. In addition, UiB also hosts 3 National Centers of Excellence awarded by the Research Council of Norway in 2012 - Centre for Intervention Science in Maternal and Child Health, Centre for Cancer Biomarkers and Birkeland Centre for Space Science. In 2014, the Research Council of Norway awarded funds to UiB for the development of infrastructures that include Health Registries for Research, Remotely Operated Vehicles for Deep Marine Research, Earth Surface Sediment Laboratory and the Norwegian Nuclear Magnetic Resonance Platform. The Department of Informatics has seven research groups in the areas of algorithms, bioinformatics, machine learning, optimization, visualization, secure and reliable communication, and program development theory. The department was ranked first in Norway with respect to quality of research, in an international evaluation conducted by the Norwegian Research Council in 2012. The department has one ERC advanced grant and one ERC starting grant. It hosts a research-focused center, Computational Biology Unit, consisting of bioinformatics research groups from multiple departments as well as a group focused on infrastructure and service provision. The department has since 2002 coordinated a national technology platform for bioinformatics and is now coordinating the Norwegian Elixir Node.

Main tasks in the project per WP;

UIB will contribute to the following tasks in **WP1 Expert network:**

- T1.1 Network of data managers and scientific best practice (Participant).
- T1.2 Models for brokering data to ELIXIR Deposition Databases (Participant).

UIB will contribute to the following tasks in **WP2 Training and Capacity Building:**

- T2.1 Identify training needs and solutions in Data Management and Stewardship (Participant).
- T2.2 Develop best practices guidelines and training materials in DMS (Participant).
- T2.3 Capacity Building in Data Management and Stewardship (Participant).
- T2.4 Outreach activities to new ELIXIR Members and Communities (Participant).

UIB will contribute to the following tasks in **WP3 Common Data Management Toolkit:**

- T3.1 Establish a Starter Toolkit (Participant).
- T3.2 Processes for enriching, maintaining and sustaining the Toolkit strators (**Task Leader**).
- T3.3 Access portal to Toolkit tailored to stakeholders (Participant).
- T3.4 Best Practices and training (Participant).

UIB will contribute to the following tasks in **WP5 Demonstrator Projects:**

- T5.1 Typology of projects based on the type of resources needed to implement a Ma-DMP (Participant).
- T5.2 Implementation of pilots data management plans (Participant).
- T5.3 Development,implementation and refinement of key performance indicators to monitor the pilots implementation of data management plans (**Task Leader**).
- T5.4 Capacity building actions based on pilots outcomes (Participant).

Short Profile of key Staff Members who will be undertaking the work

[redacted]

	[redacted]

List of the <u>top 5 most relevant</u> publications and/or products, services (including widely-used datasets or software), or other achievements

1. Jonassen, J. F. Collins, D. G. Higgins. Finding flexible patterns in unaligned protein sequences. Protein Science 4, 1587-1595 (1995).
2. A. Brazma, I. Jonassen, J. Vilo, E. Ukkonen, 1998. Predicting Gene Regulatory Elements in Silico on a Genomic Scale. Genome Research 8, 1202-1215 (1998).
3. T. H. Bø, I. Jonassen. New feature subset selection procedures for classification of expression profiles. Genome Biol. 3, research0017.1-0017.11- (2002).
4. C. Stansberg, A.O. Vik-Mo, R. Holdhus, H. Breilid, B. Srebro, K. Petersen, H. A. Jorgensen, I. Jonassen, V. M. Steen. Gene expression profiles in rat brain disclose CNS signature genes and regional patterns of functional specialisation. BMC Genomics. 4;8(1):94 (2007).
5. W. R. Taylor, G. J. Bartlett, V. Chelliah, D. Klose, K. Lin, T. Sheldon, I. Jonassen. Prediction of protein structure from ideal forms. Proteins: Struct, Funct, and Bioinfo. 70, 1610-1619 (2008).

List of <u>top 5 most relevant</u> relevant previous projects or activities, connected to the subject of this proposal

1. EU H2020 (H2020-INFRADEV-1-2015-1) ELIXIR-EXCELERATE: Fast-track ELIXIR implementation and drive early user exploitation across the life sciences, Grant number: 676559, 2015-2019.
2. FUGE national technology platform for bioinformatics (2002-2007 and 2007-2012) – grants from the FUGE program of the Research Council of Norway.
3. ELIXIR.NO - a Norwegian ELIXIR Node. Grant from the Infrastructure program of the Research Council of Norway.
4. Centre for Digital Life Norway – Grant from the Biotek2021 program of the Research Council of Norway, with focus on trans-disciplinary research and data management.
5. Angiotargeting – EU funded project (6th framework program, Life sciences, genomics, and biotechnology for health.

Description of any significant infrastructure and/or major items of technical equipment, relevant to the proposed work (if applicable):

N/A

Participant	**24 Instituto de Engenharia de Sistemas e Computadores, Investigação e Desenvolvimento. INESC-ID (ELIXIR-PT)**	**Portugal**
PI	[redacted] **(M)**	
ELIXIR Node	**ELIXIR-PT**	

Description of the legal entity/department/laboratory/group:

Instituto de Engenharia de Sistemas e Computadores, Investigação e Desenvolvimento em Lisboa (INESC-ID: http://inesc-id.pt) is dedicated to R&D and advanced education in information technology, electronics and telecommunications. INESC-ID is a private non-profit institution, officially declared of public interest. The main shareholder is the Instituto Superior Técnico da Universidade de Lisboa, the largest engineering school in Portugal. In December 2004, INESC-ID was awarded the status of Laboratório Associado by the Ministry of Science, recognising its international relevance and strategic role in the development of the scientific system.

INESC-ID congregates competences from more than one hundred senior researchers in electrical engineering and computer science, dedicated to advancing the state of the art in computers, telecommunications, and

information systems. INESC-ID technology transfer activities span the creation of technology-based start-ups and scientific and technical assistance. INESC-ID activities are structured in five scientific domains, each corresponding to a research lab: Computer Systems and Communication Networks; Embedded Electronic Systems; Energy Systems; Information and Decision Support Systems; Interactive Intelligent Systems. INESC-ID's Information and Decision Support Systems unit (IDSS Lab), a structure comprising 25 PhD researchers, has a vast experience in conducting research projects funded at the European and National level, spanning algorithms for constraint validation and software systems verification, bioinformatics, information search and data mining,information lifecycle management, and knowledge management. The lab has been awarded in 2013 a "Research Line of Excellence" five years grant by the Portuguese Government to foster research in "Large- Scale Data Management in Cloud Environments" and is a main partner of a newly National Research Infrastructure for biological data, as well as in several EU-funded projects. The lab has relevant on-going research collaborations with scientists working domains, especially biologists.
The lab uses and maintains cloud services for large scale computing, based on technologies like Hadoop, Spark, and on deploying several toolboxes for machine learning and simulation in such computing environments. The researchers of INESC-ID have developed in the past several e-infrastructures, in domains spanning from digital humanities to molecular data, in operation today. INESC-ID is a founding member of the biodata.pt consortium that is on the national roadmap for providing a data infrastructure for biological data.

Main tasks in the project per WP:

INESC-ID will contribute to the following tasks in **WP1 Expert network:**

- T1.1 Network of data managers and scientific best practice (Participant).

INESC-ID will contribute to the following tasks in **WP3 Common Data Management Toolkit:**

- T3.2 Processes for enriching, maintaining and sustaining the Toolkit strators (Participant).
- T3.4 Best Practices and training (Participant).

INESC-ID will contribute to the following tasks in **WP5 Demonstrator Projects:**

- T5.2 Implementation of pilots data management plans (Participant).

INESC-ID will contribute to the following tasks in **WP7 Federated European Genome-phenome Archives for transnational access of COVID-19 host data:**

- Task 7.1: Architecture, interfaces, and compliance to support EGA federated network on COVID-19 host data management (Participant).
- Task 7.2: A technical implementation required for a federated EGA Network (Participant).
- Task 7.3: Coordination of metadata standards for phenotype submission and access of COVID-19 data (Participant).
- Task 7.4: Operational support and maturity model for Federated EGA nodes (Participant).

Short Profile of <u>key</u> Staff Members who will be undertaking the work

All staff members have contracts with IST. IST will provide in-kind contributions against payment (article 11), see also 3rd party table on page 155-156.

List of the <u>top 5 most relevant</u> publications and/or products, services (including widely-used datasets or software), or other achievements

1. Cátia Pesquita, João D Ferreira, Francisco M Couto and Mário J. Silva. The Epidemiology Ontology: an ontology for the semantic annotation of epidemiology resources. Journal of Biomedical Semantics, 2014.

2. JD Ferreira, C Pesquita, FM Couto, MJ Silva. Bringing epidemiology into the Semantic Web. Proceedings of the International Conference On Biomedical Ontology (ICBO), 2012.
3. FM Couto, MJ Silva, PM Coutinho. Measuring semantic similarity between Gene Ontology terms. Data & knowledge engineering 61 (1), 137-152, 2007.

List of top 5 most relevant relevant previous projects or activities, connected to the subject of this proposal

1. EU H2020 (H2020-INFRADEV-1-2015-1) ELIXIR-EXCELERATE: Fast-track ELIXIR implementation and drive early user exploitation across the life sciences, Grant number: 676559, 2015-2019.
2. DataStorm "Large-Scale Data Management in Cloud Environments", FCT (Portugal) ProEXCL/EEIESS/0257/2012.
3. Epiwork "Developing the framework for an epidemic forecast infrastructure", FET Proactive Initiative, FP7 231807 (2009-2013).
4. REACTION "Retrieval, Extraction and Aggregation Computing Technology for Integrating and Organizing News", FCT (UTAustin-Portugal) UTA-Est/MAI/0006/2009.

Description of any significant infrastructure and/or major items of technical equipment, relevant to the proposed work (if applicable):

N/A

Participant	**25 Fundação Calouste Gulbenkian. IGC (ELIXIR-PT)**	**Portugal**
PI	[redacted]	
ELIXIR Node	**ELIXIR-PT**	

Description of the legal entity/department/laboratory/group:

Instituto Gulbenkian de Ciência - Fundação Calouste Gulbenkian. The IGC is an international biomedical research and graduate training institute, dedicated to promoting multidisciplinary science of excellence. It was established by the Calouste Gulbenkian Foundation in 1961, and it is still generously supported by the Foundation. Research at the IGC is organism centered, hypothesis driven, integrative and multidisciplinary in approach. The focus is on the genetic bases of development and evolution of complex systems, with a strong theoretical/computational component. There is a strong service tradition, from its inception hosting the first scientific computing center in Portugal, to the current participation of the European Mouse Mutant Archive. The IGC has multiple core facilities that support its own research work as well as the Portuguese Scientific Community at large, including a Bioinformatics Core facility that provides direct user support, engages in collaborative research work, as well as developing data warehousing solutions for the research groups.

The IGC embraces four major missions: 1) to identify, educate and incubate new research leaders, providing state-of-the-art facilities and full financial and intellectual autonomy to pursue research projects; 2) to export new scientific leaders to other research centres and academia, in Portugal and abroad; 3) to provide international graduate teaching and training programmes; and 4) to promote a science-based culture and the values of science in society, as well as the active participation of society in scientific research, through engagement with different communities and stakeholders. The IGC is characterised by its strong interdisciplinary culture. It hosts a diverse set of 37 small and cohesive autonomous research groups, on a rolling five scheme, including about 8 groups devoted to mathematical modelling and computational biology, the largest among Portuguese institutions.

In its role as a host institution with core funding from the Gulbenkian Foundation, the IGC has provided a supportive and stimulating intellectual environment to almost 80 largely externally funded research groups since 1998, half of which have moved on to other research centres, mainly in Portugal. These young scientists have made major contributions to our understanding of auto-immune and infectious diseases, antibiotic resistance, cancer, ageing and genetics. IGC scientists have been very successful in winning external competitive grants, from national and international sources including the Howard Hughes Medical Institute, the Human Frontiers Science Programme, the Bill & Melinda Gates Foundation and the European Research Council.

Main tasks in the project per WP;

IGC will contribute to the following tasks in **WP2 Training and Capacity Building:**
- T2.1 Identify training needs and solutions in Data Management and Stewardship (Participant).
- T2.2 Develop best practices guidelines and training materials in DMS (Participant).
- T2.3 Capacity Building in Data Management and Stewardship (Participant).

IGC will contribute to the following tasks in **WP4 Communications, Industry, International, Impact and Sustainability:**
- T4.2 Operating the ELIXIR Innovation and SME Forum and enhancing implementation of national industry engagement efforts (Participant).
- T4.4 Implementing an impact assessment toolkit for demonstrating ELIXIR's value, nationally and at European level (Participant).

IGC will contribute to the following tasks in **WP5 Demonstrator Projects:**
- T5.2 Implementation of pilots data management plans (Participant).
- T5.4 Capacity building actions based on pilots outcomes (Participant).

IGC will contribute to the following tasks in **WP6 Project Management and Scientific Coordination:**
- T6.4 Excellence in Management (Participant).

IGC will contribute to the following tasks in **WP7 Federated European Genome-phenome Archives for transnational access of COVID-19 host data:**
- Task 7.1: Architecture, interfaces, and compliance to support EGA federated network on COVID-19 host data management (Participant).
- Task 7.2: A technical implementation required for a federated EGA Network (Participant).
- Task 7.3: Coordination of metadata standards for phenotype submission and access of COVID-19 data (Participant).
- Task 7.4: Operational support and maturity model for Federated EGA nodes (Participant).

Short Profile of key Staff Members who will be undertaking the work

[redacted]

[redacted]

List of the top 5 most relevant publications and/or products, services (including widely-used datasets or software), or other achievements

1. Marques I, Almeida P, Alves RJ, Dias MJ, Godinho A, Pereira-Leal JB (2014) Bioinformatics projects
   supporting enquiry-based learning in high schools PLoS Comp. Bio. 10 (1), e1003404.
2. Pereira-Leal JB et al (2014) Comprehensive assessment of the transcriptome of the cork oak Quercus suber) BMC genomics. 5:371.
3. Manuel Corpas, Rafael C Jimenez, Erik Bongcam-Rudloff, Aidan Budd, Michelle D Brazas, Pedro L Fernandes, Bruno Gaeta, Celia van Gelder, Eija Korpelainen, Fran Lewitter, Annette McGrath, Daniel MacLean, Patricia M Palagi, Kristian Rother, Jan Taylor, Allegra Via, Mick Watson, Maria Victoria Schneider, Teresa K Attwood (2014) The GOBLET training portal: a global repository of bioinformatics training materials, courses and trainers. Bioinformatics. 09/2014; DOI: 10.1093/bioinformatics/btu601.
4. Allegra Via, Thomas Blicher, Erik Bongcam-Rudloff, Michelle D Brazas, Cath Brooksbank, Aidan Budd, Javier De Las Rivas, Jacqueline Dreyer, Pedro L Fernandes, Celia van Gelder, Joachim Jacob, Rafael C Jimenez, Jane Loveland, Federico Moran, Nicola Mulder, Tommi Nyrönen, Kristian Rother, Maria Victoria Schneider, Teresa K Attwood (2013) Best practices in bioinformatics training for life scientists. Briefings in bioinformatics. 06/2013; DOI: 10.1093/bib/bbt043.
5. P Fernandes, P Jain, C Moita (2012) Training experimental biologists in bioinformatics. Advances in bioinformatics, 01/2012; DOI: 10.1093/bib/bbt043.

List of top 5 most relevant relevant previous projects or activities, connected to the subject of this proposal

1. EU H2020 (H2020-INFRADEV-1-2015-1) ELIXIR-EXCELERATE: Fast-track ELIXIR implementation and drive early user exploitation across the life sciences, Grant number: 676559, 2015-2019.
2. The IGC has hosted a Bioinformatics core facility that has served the country's scientific community for 20 years, providing support in bioinformatics data analysis, compute servers including a high performance computing center, and in the past mirroring international databases. The Bioinformatics unit, has also participated in international networks such as EMBnet and GOBLET.
3. IGC has pioneered post-graduate education in Bioinformatics in Portugal. In 2001, IGC has partnered with FCUL (Faculdade de Ciências da Universidade de Lisboa) to offer PGBIOINF, a post-graduate course with MSc and PhD titles. Its modular design was fully Bologna compliant and aimed at creating professionals with a broad range of capabilities. The MSc ran continuously for three years in the curricular form and more than 30 students completed the thesis. The collaboration with FCUL was interrupted to give place to a more ambitious doctoral programme: PDBC.

Description of any significant infrastructure and/or major items of technical equipment, relevant to the proposed work (if applicable):

N/A

Participant	**26 Uppsala University, UU (ELIXIR-SE)**	**Sweden**
PI	[redacted]	
ELIXIR Node	**ELIXIR-SE**	

Description of the legal entity/department/laboratory/group:

Uppsala University (UU) is the legal entity for the national bioinformatics infrastructure NBIS (National Bioinformatics Infrastructure Sweden), which is coordinating the Swedish ELIXIR Node.

Uppsala University is the oldest university in Sweden, founded in 1477, with 6,949 employees and 41,470 students. World-class research and high quality education pursued here benefit society and business on a global level. The University is characterised by diversity and breadth, with international frontline research at nine faculties and limitless educational offerings at Bachelor's and Master's levels. Annual turnover of SEK 6.5 billion (€688 million). UU has 2,437 doctoral students (48% women).

NBIS is a distributed national research infrastructure to provide support for life science researchers in Sweden. The organisational structure allows for changes in services over time as new techniques are developed and utilised. NBIS is the Swedish Node and coordinates the Swedish contributions to the ELIXIR infrastructure.
NBIS constitutes the SciLifeLab Bioinformatics Platform and provides project support, infrastructure and tools development, and advanced training.
NBIS has nodes at all major Swedish universities. NBIS provides expertise, methodology and data access within bioinformatics for Swedish research groups and provides the Swedish activities in the European infrastructure ELIXIR.

Main tasks in the project per WP;

UU will contribute (**WP Leader**) to the following tasks in **WP1 Expert network:**
- T1.1 Network of data managers and scientific best practice (**Task Leader**).
- T1.2 Models for brokering data to ELIXIR Deposition Databases (Participant).
- T1.3 Business model (Participant).
- T1.4 Sustainable and scalable operating model for harmonised data management in European projects (Participant).
  - Coordinate the network of experts in data management. Contribute with expertise in the area of data management.

UU will contribute to the following tasks in **WP2 Training and Capacity Building:**
- T2.1 Identify training needs and solutions in Data Management and Stewardship (Participant).
- T2.3 Capacity Building in Data Management and Stewardship (Participant).
  - Contribute with experts for training and capacity building. Host training events in Sweden.

UU will contribute to the following tasks in **WP3 Common Data Management Toolkit:**
- T3.1 Establish a Starter Toolkit (Participant).
- T3.2 Processes for enriching, maintaining and sustaining the Toolkit strators (Participant).
- T3.3 Access portal to Toolkit tailored to stakeholders (Participant).
  - The experienced team of systems developers in Elixir-SE will contribute in creating tools necessary for efficient and user-friendly data management. Much of this work will be conducted in close collaboration with WP1 to match the requirements by the data managers.

UU will contribute to the following tasks in **WP5 Demonstrator Projects:**
- T5.1 Typology of projects based on the type of resources needed to implement a Ma-DMP (Participant).
- T5.2 Implementation of pilots data management plans (Participant).
  - Elixir-SE will contribute to data management for human data, where we have expertise and a significant user community.

UU will contribute to the following tasks in **WP7 Federated European Genome-phenome Archives for transnational access of COVID-19 host data:**
- Task 7.1: Architecture, interfaces, and compliance to support EGA federated network on COVID-19 host data management (Participant).
- Task 7.2: A technical implementation required for a federated EGA Network (Participant).

- Task 7.3: Coordination of metadata standards for phenotype submission and access of COVID-19 data (Participant).
- Task 7.4: Operational support and maturity model for Federated EGA nodes (Participant).

Previous experience relevant to the Tasks:

Elixir-SE has, in the recent EU project ELIXIR-EXCELERATE, contributed to WP9 on human data and to WP10 on capacity building, partly on the local/federated EGA. ELIXIR-SE has also in the Nordic collaboration Tryggve developed systems for handling of human data and contributed to the local/federated EGA, of which one of the first instances will be set-up in Sweden during Q2-2019. Since 2016, ELIXIR-SE has been a data manager and is currently expanding our data management activities.

Short Profile of key Staff Members who will be undertaking the work

[redacted]

List of the top 5 most relevant publications and/or products, services (including widely-used datasets or software), or other achievements

1. Ameur A, Dahlberg J, Olason P, Vezzi F, Karlsson R, Martin M, Viklund J, Kähäri AK, Lundin P, Che H, Thutkawkorapin J, Eisfeldt J, Lampa S, Dahlberg M, Hagberg J, Jareborg N, Liljedahl U, Jonasson I, Johansson Å, Feuk L, Lundeberg J, Syvänen AC, Lundin S, Nilsson D, Nystedt B, Magnusson PK, Gyllensten U. (2017) SweGen: a whole-genome data resource of genetic variability in a cross-section of the Swedish population. Eur J Hum Genet. 25:1253--1260.
2. Jiménez, R. C., Kuzak, M., ... Borg, M., ... Hagberg, J., ... Crouch, S. (2017) Four simple recommendations to encourage best practices in research software. F1000Research 6, 876.
3. Dahlö, M., Scofield, D. G., Schaal, W, Spjuth, O. (2018) Tracking the NGS revolution: managing life science research on shared high-performance computing clusters. GigaScience 7, giy028.
4. Dominguez del Angel H, ... Soler, L., Binzer-Panchal, L., Lantz, H. (2018) Ten steps to get started in genome assembly and annotation. F1000Research 7, 148.
5. Wilkinson, M., ... Persson, B., ... Mons, B. (2016) The FAIR Guiding Principles for scientific data management and stewardship. Nature Sci. Data 3, 160018.

List of top 5 most relevant relevant previous projects or activities, connected to the subject of this proposal

1. EU H2020 (H2020-INFRADEV-1-2015-1) ELIXIR-EXCELERATE: Fast-track ELIXIR implementation and drive early user exploitation across the life sciences. Grant number: 676559, 2015-2019.
2. Tryggve -- Nordic collaboration for sensitive data 2014--2017, grant from NordForsk, https://neic.no/tryggve/
3. Tryggve2 -- Nordic collaboration for sensitive data 2017–2020, grant from NordForsk, https://neic.no/tryggve2/
4. National Bioinformatics Infrastructure Sweden (NBIS) and the Swedish node in Elixir (the European infrastructure for biological information) 2018–2020, grant from the Swedish Research Council, https://nbis.se
5. PhenoMeNal, EU H2020 project 2015--2018 on data processing and analysis pipelines for molecular phenotype data. Grant number 654241.

Description of any significant infrastructure and/or major items of technical equipment, relevant to the proposed work (if applicable):

N/A

Participant	**27 University of Ljubljana. UL (ELIXIR-SI)**	**Slovenia**
PI	[redacted] **(M)**	
ELIXIR Node	**ELIXIR-SI**	

Description of the legal entity/department/laboratory/group:
The University of Ljubljana (http://www.uni-lj.si/eng/; PIC UL: 999923240) is a public autonomous educational, research and artistic higher education institution with a very rich tradition. It was established in 1919 and it encompasses 23 faculties and 3 art academies. It is the largest and the oldest university in Slovenia, with more than 40,000 undergraduate and graduate students attending over 300 different study programmes. In 2016 UL had 286 research groups with over 3,000 registered researchers (548 PhD students) and creates almost half of the research results of Slovenia. From 2008 the UL is committed to respect the principles of the European Charter for Researchers and the Code of Conduct for Recruitment of Researchers. In 2013 the UL was awarded the 'HR Excellence in Research' logo. UL is very active in national research and educational programmes. In 2016 the UL cooperated in 174 national research programmes, 480 basic and 650 applied research projects, 26 post-doc projects, 39 targeted research projects (CRP) and 7 technological platforms. UL is also very active in international and EU R&D programmes: in the period 2007-2016 it was involved in more than 750 European projects, among them 160 FP7/H2020 and 1 ERC grant. Within FP7/H2020 projects the UL cooperated in 20 Marie Curie actions (13 ITN, 2 IAPP, 3 IRSES, 1 CIG, 1 ERG). Currently UL cooperates in 317 EU projects out of which 83 FP7 (12 Marie Curie) and 233 other European projects.

Main tasks in the project per WP;

UL will contribute to the following tasks in WP1 Expert network:

- T1.1 Network of data managers and scientific best practice (Participant).

UL will contribute to the following tasks in **WP2 Training and Capacity Building:**

- T2.1 Identify training needs and solutions in Data Management and Stewardship (Participant).
- T2.2 Develop best practices guidelines and training materials in DMS (Participant).
- T2.3 Capacity Building in Data Management and Stewardship (**Task Leader**).
- T2.4 Outreach activities to new ELIXIR Members and Communities (Participant).

UL will contribute to the following tasks in **WP3 Common Data Management Toolkit:**

- T3.2 Processes for enriching, maintaining and sustaining the Toolkit strators (Participant).
- T3.4 Best Practices and training (Participant).

UL will contribute to the following tasks in **WP4 Communications, Industry, International, Impact and Sustainability**:

- T4.3 Engaging potential new Member countries and enhancing international visibility (Participant).
- T4.4 Implementing an impact assessment toolkit for demonstrating ELIXIR's value, nationally and at European level (Participant).

UL will contribute to the following tasks in **WP5 Demonstrator Projects:**

- T5.2 Implementation of pilots data management plans (Participant).

- T5.4 Capacity building actions based on pilots outcomes (**Task Leader**).

UL will contribute to the following tasks in **WP7 Federated European Genome-phenome Archives for transnational access of COVID-19 host data:**
- Task 7.4: Operational support and maturity model for Federated EGA nodes (Participant).

Short Profile of key Staff Members who will be undertaking the work

[redacted]

List of the top 5 most relevant publications and/or products, services (including widely-used datasets or software), or other achievements

1. Artaza H, Hong NC, Corpas M, Corpuz A, Hooft R, Jimenez RC, Leskošek B, Olivier BG, Stourac J, Svobodanova-Vareckova R, et al. Top 10 metrics for life science software good practices. F1000Research, 5: 1-6, 2016. https://f1000research.com/articles/5-2000/v1.
2. Dominguez del Angel V, Hjerde E, Sterck L, Capella-Guttierez S, Notredame C, Vinnere Pettersson O, Amselem J, Bouri L, Bocs S, Leskošek B, et al. Ten steps to get started in genome assembly and annotation : [version 1; referees: awaiting peer review]. F1000Research, 7: 1-19, 2018. https://f1000research.com/articles/7-148/v1
3. Blagus R, et al. (Leskošek B) Comparison of bibliometric measures for assessing relative importance of researchers. Scientometrics, 105: 1743-1762, 2015.
4. Leskošek B, et al. Lightweight application for generating clinical research information systems: MAGIC. Wiener Klinische Wochenschrift, 127: 228-234, 2015.
5. Dimec J, Leskošek B. Building co-operative bibliographic databases in European bioethics: a contribution from EU new member states. Computer methods and programs in biomedicine, 87 (2): 123-131, 2007.

List of top 5 most relevant relevant previous projects or activities, connected to the subject of this proposal

1. EU H2020 (H2020-INFRADEV-1-2015-1) ELIXIR-EXCELERATE: Fast-track ELIXIR implementation and drive early user exploitation across the life sciences, Grant number: 676559, 2015-2019.
2. CASyM (https://www.casym.eu/)
3. EkoSMART (http://ekosmart.net/en/ekosmart-2/
4. ETHICSWEB (https://www.ethicsweb.eu)
5. BioModUE_PTL (https://unimed.mf.uni-lj.si/uemg/)

Description of any significant infrastructure and/or major items of technical equipment, relevant to the proposed work (if applicable):

N/A

Participant	**28 The University of Manchester. UNIMAN (ELIXIR-UK)**	**United Kingdom**
PI	[redacted] **(F)**	
ELIXIR Node	**ELIXIR-UK**	

Description of the legal entity/department/laboratory/group:

The University of Manchester (UNIMAN) is the largest single-site university in the UK, with research strength across all scientific disciplines. Since 2001, the e-Science Lab in the School of Computer Science have been influential in the development and research in scientific workflow systems, collaborative eScience technologies, semantic and linked data technologies and platforms for reproducible research - developing tools and resources designed for data-driven and computational research. The group pioneered scientific workflow R&D, developing the open source Apache Taverna workflow management system, co-developed the W3C PROV standard for provenance collection and the community-developed Common Workflow Language. UNIMAN have a history of building asset sharing and collaboration platforms for scientists, notably as leaders of the FAIRDOM consortium (fair-dom.org), a pan-national partnership to develop and support a Commons for the stewardship and exchange of data, models, workflows, samples and laboratory SOPs initially for Systems and Synthetic Biology projects. The FAIRDOM-SEEK platform is the basis of the FAIRDOMHub (fairdomhub.org), a public service, supports 130 registered projects, and has been adopted by over 50 other projects for independent installation. The FAIRDOM project supports EU ERANet programmes, EU ESFRI ISBE (Infrastructure for Systems Biology Europe) data management pillar and forms the basis of the IBISBA1.0 asset management platform IBISBAHub. It is also used as the basis for facility management by the Pacific Northwest National Laboratory in the USA as well as the UK Synthetic Biology Centres in Manchester and Edinburgh. The FAIRDOM SEEK platform is the basis of a Common Data Management Toolkit for 6 ELIXIR Nodes and their national data management provision – for example Norway's Center for Digital Life. UNIMAN developed the BioCatalogue for life science web services and the myExperiment the workflow registry for the EOSC Life cluster project's workflow collaboratory (13 RIs in the Life Sciences) and the science workflow platform for the Synthesys+ project, a pillar of the DISSCo ESFRI. myExperiment 2.0 will be based on the SEEK platform. UNIMAN worked on Data Catalogue interoperability in EOSCpilot and lead the BYODs for developing a Capability model for FAIRification of Data sets in the ELIXIR-lead IMI FAIRplus. UNIMAN lead the development of the TeSS, the ELIXIR training portal that aggregates, organises, and disseminates Life science training resources. As founding partners of the UK's Software Sustainability Institute, UNIMAN have worked with ELIXIR on Software best practices and introducing Software and Data Carpentry training and "train the trainers" training to the RI members. UNIMAN are a prime member of the ELIXIR-UK Node, and Goble is Head of Node. UNIMAN are members of the ELIXIR Microbial Biotechnology Community, the Bioschemas and CWL activities and lead the Identifiers WG.

Main tasks in the project per WP;

UNIMAN will contribute to the following tasks in **WP1 Expert network:**

- T1.1 Network of data managers and scientific best practice (Participant).

UNIMAN will contribute to the following tasks in **WP2 Training and Capacity Building:**

- T2.3 Capacity Building in Data Management and Stewardship (Participant).

UNIMAN will contribute to the following tasks in **WP3 Common Data Management Toolkit:**

- T3.1 Establish a Starter Toolkit (Task Leader).
- T3.2 Processes for enriching, maintaining and sustaining the Toolkit strators (Participant).

Short Profile of key Staff Members who will be undertaking the work

[redacted]

[redacted]

List of the top 5 most relevant publications and/or products, services (including widely-used datasets or software), or other achievements

1. Alterovitz G, Dean D, Goble C , et al Enabling precision medicine via standard communication of HTS provenance, analysis, and results (2018), PLOS Biology, https://doi.org/10.1371/journal.pbio.3000099
2. Wolstencroft, K, Krebs O, et al. (2016). FAIRDOMHub: a repository and collaboration environment for sharing systems biology research. Nuc. Acids Res, 45: D404-7.
3. Wolstencroft K, Owen S, et al. (2015). SEEK: a systems biology data and model management platform, BMC Systems Biology, 9(1):1-12.
4. Stanford NJ, Wolstencroft K, et al, (2015) The evolution of standards and data management practices in systems biology, Molecular Systems Biology, Issue 11:12, 2015. 10.15252/msb.20156053.
5. Carole A Goble, Jiten Bhagat, Sergey Aleksejevs, Don Cruickshank, Danius Michaelides, David Newman, Mark Borkum, Sean Bechhofer, Marco Roos, Peter Li, David De Roure: myExperiment: a repository and social network for the sharing of bioinformatics workflows, Nucl. Acids Res. 38 (suppl 2): W677-W682. (2010)

List of top 5 most relevant relevant previous projects or activities, connected to the subject of this proposal

1. H2020-INFRADEV-1-2015-1, 676559, *ELIXIR-EXCELERATE: Fast-track ELIXIR implementation and drive early user exploitation across the life sciences*, 2015-2019, WP lead: interoperability
2. H2020-INFRAIA-2017, 730976, *IBISBA 1.0: Industrial Biotechnology Innovation and Synthetic Biology Accelerator*, total budget €5,000,157, €491,008 to Manchester, 2017-2021, WP lead: data management, workflows platforms and workflow hub.
3. INFRAEOSC-04-2018, H2020-EU.1.4.1.1, 824087, *EOSC-Life, Providing an open collaborative space for digital biology in Europe*, 2019-2023, Overall:€23,745,996 Task lead: Workflowregistry
4. BBSRC BB/M013189/1 *FAIRDOM (DMMCore:) Data and Model Management Core for ERASysAPP & Europe* 2014-2019. £1,015,804 to Manchester, £3m to consortium, Coordinator
5. EP/N006410/1 EPSRC, BBSRC, ESRC, Software Sustainability Institute Phase 2, £4,298,660 full consortium, 2015-2020. (£802,627K to Manchester), Community lead

Description of any significant infrastructure and/or major items of technical equipment, relevant to the proposed work (if applicable):

Software Services

- TeSS (http://tess.elixir-europe.org/) ELIXIR training portal
- myExperiment (myexperiment.org) workflow registry and repository
- FAIRDOM SEEK (seek4science.org) data, model and operations management Commons
- FAIRDOMHub (fairdomhub.org) public Systems Biology project asset Commons
- BioCatalogue (biocatalogue.org) web service registry for Life Sciences
- Apache Taverna (taverna.org.uk) workflow management system
- CWLViewer (view.commonwl.org) CWL Viewer and Gallery

Participant	**29 The University of Cambridge. UCAM (ELIXIR-UK)**	**United Kingdom**
PI	[redacted] **F(Until 28/2/2020)Alexia Cardona, PhD (F) (from 1/3/2020)**	
ELIXIR Node	**ELIXIR-UK**	

Description of the legal entity/department/laboratory/group:

The University of Cambridge is one of the most renowned Research/Higher Education Institute and is frequently ranked amongst the top 5 in international academic rankings such as ARWU and Shanghai Ranking. It has a long-standing history of academic and scientific excellence backed up with rich culture, learning, research and creativity. Many affiliates of University of Cambridge have won Nobel Prizes for their significant advances. University of Cambridge is also a major participant in European projects and is one of the top recipients of FP7 funding. The Bioinformatics Training Programme of the University of Cambridge, headed by [redacted], offers a wide-ranging portfolio of training courses in bioinformatics and data science, at both undergraduate and post-graduate level. In 2018, it comprised 102 postgraduate training events and hosted 83 undergraduate training sessions; 2,600 postgraduates and 760 undergraduates are trained on average each year. UCAM is a member of the ELIXIR UK Node.

Main tasks in the project per WP;

UCAM will contribute to the following tasks in **WP2 Training and Capacity Building:**

- T2.1 Identify training needs and solutions in Data Management and Stewardship (**Task Leader**).
- T2.2 Develop best practices guidelines and training materials in DMS (Participant).
  - In WP2, UCAM will be involved in (i) identifying specific training needs, for data management and stewardship, of both data stewards/experts and researchers T2.1 and (ii) devising a timeline for implementation of activities in T2.2. Activities will involve: bringing together all WPs (through F2F workshops but also regular telecons) to shape activities, setup proper liaisons between WP2 and the other WPs, coordinate discussions, landscape DMS training needs across ELIXIR, produce training needs analysis/timeline/white paper/recommendations etc that will be utilized by the other WP2 tasks and relevant WP3 and WP5 Tasks.

UCAM will contribute to the following tasks in **WP4 Communications, Industry, International, Impact and Sustainability:**

- T4.4 Implementing an impact assessment toolkit for demonstrating ELIXIR's value, nationally and at European level (Participant).
  - In WP4, UCAM will be involved in the development and rolling out of an impact assessment toolkit for demonstrating ELIXIR's value, nationally and at European level. UCAM has been leading the training impact assessment work in ELIXIR and will be bringing in this expertise to shape the activities of Task 4.4

Short Profile of key Staff Members who will be undertaking the work

[redacted]

[redacted]

List of the top 5 most relevant publications and/or products, services (including widely-used datasets or software), or other achievements

1. EU H2020 (H2020-INFRADEV-1-2015-1) ELIXIR-EXCELERATE: Fast-track ELIXIR implementation and drive early user exploitation across the life sciences, Grant number: 676559, 2015-2019.
2. EU H2020 (H2020-INFRAEOSC-2018-2) EOSC-Life: Providing an open collaborative space for digital biology in Europe, Grant number: 824087, 2019-2023
3. EJP RD (H2020-SC1-2018-Single-Stage-RTD) European Joint Programme in Rare Disease, Grant number: (825575), 2019-2024

List of top 5 most relevant relevant previous projects or activities, connected to the subject of this proposal

1. Mulder, N., et al., The development and application of bioinformatics core competencies to improve bioinformatics training and education. PLoS Comput Biol, 2018. 14(2): p. e1005772.
2. Via, A., et al., A new pan-European Train-the-Trainer programme for bioinformatics: pilot results on feasibility, utility and sustainability of learning. Brief Bioinform, 2017.
3. Morgan, S.L., et al., The ELIXIR-EXCELERATE Train-the-Trainer pilot programme: empower researchers to deliver high-quality training. F1000Res, 2017. 6.
4. Larcombe, L., et al., ELIXIR-UK role in bioinformatics training at the national level and across ELIXIR. F1000Res, 2017. 6.
5. Welch, L., et al., Applying, Evaluating and Refining Bioinformatics Core Competencies (An Update from the Curriculum Task Force of ISCB's Education Committee). PLoS Comput Biol, 2016. 12(5): p. e1004943.
6. Schiffthaler, B., et al., Training in High-Throughput Sequencing: Common Guidelines to Enable Material Sharing, Dissemination, and Reusability. PLoS Comput Biol. 2016. 12(6): p. e1004937.

Description of any significant infrastructure and/or major items of technical equipment, relevant to the proposed work (if applicable):

N/A

Participant	**30 The Centre for Genomic Regulation (CRG) (ELIXIR-ES)**	**Spain**
PI	[redacted] **(M)**	
ELIXIR Node	**ELIXIR-ES**	

Description of the legal entity/department/laboratory/group:

The Centre for Genomic Regulation (CRG) is a non-profit foundation created in 2000 as a partnership between the Catalan Government and the Spanish Ministry of Economy, Industry and Competitiveness and affiliated with the Pompeu Fabra University (UPF). Today, with over 400 scientists from 43 countries, CRG is a first-class multidisciplinary research that hosts 30 group leaders, of which ~half hold an ERC grant. Research is organized around 4 programmes: (i) Systems Biology, (ii) Bioinformatics and Genomics, (iii) Gene regulation, Stem Cells and Cancer and (iv) Cell and Developmental Biology. CRG demonstrates proven experience in managing European projects: 71 in FP7, of which 9 were coordinated by our investigators. The CRG hosts 7 scientific and technical cutting-edge core facilities, including the Genomic Unit (one of the largest genomic platforms in Spain) and the Bioinformatics Unit (offering scientific user training, application development and custom-designed data analysis to researchers and the other core facilities). In H2020, 40 projects have already been awarded (including 6 ERCs and 6 projects coordinated by the centre, 4 Innovative Training Networks (ITN) (of which 2 coordinated) and 1 new COFUND (INTREPiD). The centre is also part of pan-European infrastructure networks (e.g. ELIXIR, EuroBioImaging, MuG), as well as large international initiatives (ENCODE, GTEX). In addition, CRG is one of the life-science research centers in Spain that has been awarded with a highly-competitive 'Severo Ochoa' grant of excellence (2017-2021). In 2013, the CRG was awarded the HR Excellence logo by the European Commission to recognize the institute's commitment to improve human resources policies for researchers. The CRG hosts a well-known International PhD Programme (currently training around 100 young researchers - almost 70% from abroad) and a Postdoctoral Programme with highly talented young researchers, and provides mentoring and high-quality training courses, weekly seminars (Data Clubs, Journal Clubs, Lab Meetings, Programme Seminars and Group Leader seminars) and seminars with top invited speakers. The CRG cannot grant PhD titles, but can establish partnerships with Universities to award PhD degrees for experimental work carried out at CRG. The center is committed to the exploitation of CRG research results for public welfare, by developing CRG scientific breakthroughs into novel therapeutic, diagnostic and related products, through patents, licences, research contracts and service and consultancy agreements. Since 2013, the CRG is co-managing the European Genome-phenome Archive (EGA) in partnership with the EBI. The CRG is contributing with expertise in different areas, especially software development and database design. Members of both institutions are part of the EGA Strategic Committee, the EGA Operations Committee, the Helpdesk team and work together to develop new features, the website and any other tool required for the normal EGA operations and for its future

Main tasks in the project per WP;

CRG will contribute to the following tasks in **WP5**

- T5.2
  - The CRG is a core participant in pilot project 4 (WP5, Task 5.2), in particular given its leading role in the EGA, Local EGA, Beacon and Beacon Network. CRG is also a very active member in the creation and adoption of GA4GH standards and in forums about GDPR.

CRG will co-lead WP7 and will contribute to the following tasks in **WP7 Federated European Genome-phenome Archives for transnational access of COVID-19 host data:**

- Task 7.1: Architecture, interfaces, and compliance to support EGA federated network on COVID-19 host data management (Participant).
- Task 7.2: A technical implementation required for a federated EGA Network (Task Leader).
- Task 7.3: Coordination of metadata standards for phenotype submission and access of COVID-19 data (Participant).

- Task 7.4: Operational support and maturity model for Federated EGA nodes (Participant).

Short Profile of key Staff Members who will be undertaking the work

[redacted]

List of the top 5 most relevant publications and/or products, services (including widely-used datasets or software), or other achievements

1. Lappalainen I et al. (Rambla J, co-author) (2015). "The European Genome-phenome Archive of human data consented for biomedical research". Nat Genetics; 47(7):692-5.
2. Dyke SOM et al. (Rambla J, co-author) (2016) "Consent Codes: Upholding Standard Data Use Conditions" PLOS Genetics 12(1): e1005772
3. Fiume M et al. (Rambla J, co-author) (2019). "Federated discovery and sharing of genomic data using Beacons". Nat Biotechnol 37 480-480.
4. Dyke, SOM et al. (Rambla J, co-author) (2018)" Registered access: authorizing data access" Eur J Hum Genet 26 1721-1731

List of top 5 most relevant relevant previous projects or activities, connected to the subject of this proposal

1. ELIXIR EXCELERATE H2020 grant. WP9 co-Lead .
2. IASIS H2020 grant - www.project-iasis.eu
3. Lead of the Life Science datasets EOSCpilot Science Demonstrator
4. Co-Lead of the ELIXIR Europe Human Data committee (https://www.elixir-europe.org/use-cases/human-data)
5. Active member of the Global Alliance for Genomics and Health (http://ga4gh.org), especially in the Beacon project Description

Description of any significant infrastructure and/or major items of technical equipment, relevant to the proposed work (if applicable):

The European Genome-phenome Archive (EGA) https://ega-archive.org/about is a service for permanent archiving and sharing of all types of personally identifiable genetic and phenotypic data resulting from biomedical research projects. The EGA contains exclusive data collected from individuals whose consent agreements authorize data release only for specific research use or to bona fide researchers. The EGA provides the necessary security required to control access and maintain patient confidentiality, while providing access to those researchers and clinicians authorized to view the data. During 2018, CRG, in collaboration with the Barcelona Supercomputing Center (BSC) is setting up a new computing infrastructure for EGA that would host a new set of data discovery functionalities for all EGA users.

Participant	**31 University of Pécs (UP) (ELIXIR-HU)**	**Hungary**
PI	[redacted] **(M)**	

ELIXIR Node	**ELIXIR-HU**

Description of the legal entity/department/laboratory/group:

University of Pécs (UP), established in 1367, is a 25,000 student university in Hungary with 10 faculties (disciplines). The UP's main disciplinary fields are medicine, engineering and business and economics. With its ten faculties and 29 clinics and patient care institutions, the university offers a broad range of training and degree programmes with nearly 2,000 teaching and research staff. UP is considered as the Knowledge Centre of the South Transdanubian Region. One of the main strategic goals of UP is to build fully integrated bioinformatics capacity across its life science research. This is especially critically important at UP as the majority of its research groups are focusing on big biomedical data applications to address key issues in human health. These approaches rely on well-established and state-of-the-art bioinformatics support and solution.

Main tasks in the project per WP;

UP will contribute to the following tasks in WP7 Federated European Genome-phenome Archives for transnational access of COVID-19 host data:

- Task 7.2: A technical implementation required for a federated EGA Network (Participant).
- Task 7.3: Coordination of metadata standards for phenotype submission and access of COVID-19 data (Participant).
  UP will collect local metadata on disease severity, comorbidity, gender and age profiles, ancestry, lifestyle risk factors, and hospitalisation information from COVID-19 patients. These data will then be used to construct a common minimal metadata model that will map across COVID-19 studies in the federated EGA network.
- Task 7.4: Operational support and maturity model for Federated EGA nodes (Participant).
  UP will develop, coordinate and organise the Hungarian operational agreements, standard operating procedures and best practices in order to engage and become operational in the EGA federated network.

Short Profile of <u>key</u> Staff Members who will be undertaking the work

[redacted]

List of <u>top 5 most relevant</u> relevant previous projects or activities, connected to the subject of this proposal

1. Multicenter, non-invasive clinical trial to identify predisposing genetic factors of persons infected with the novel COVID-19 virus and preparation for potential pharmacogenetic therapeutic targets (HUNGEN). Hungarian NKFIH grant, number: 2020-2.1.1-ED-2020-00009, Coordinator: Prof. Dr. Kovács L. Gábor, Scientific leader: Dr. Attila Gyenesei, Received grant: 300 000 EUR, duration: 01.05.2020-30.04.2021
2. Educating Experts of the Future: Developing Bioinformatics and Biostatistics competencies of European Biomedical Students (BECOMING). ERASMUS+ project, number: 2019-1-HU01-KA203-061251, Scientific leader and coordinator: Dr. Attila Gyenesei, Received grant: 259 150 EUR, duration: 01.09.2019-08.30.2022
3. Curriculum Development of Human Clinical Trials for the Next Generation Biomedical Students (CONSCIOUS). ERASMUS+ project, number: 2018-1-HU01-KA203-047811, Coordinator: Prof. Dr.

Kovács L. Gábor, Scientific leader: Dr. Attila Gyenesei, Received grant: 299 755 EUR, duration: 31.01.2018-30.08.2021

4. As technical coordinator and part of the leader team of ELIXIR Hungary, coordinating bioinformatics activities among the members of the nodes in Hungary.
5. Establishing and running Bioinformatics Core Facilities in Turku, Finland, at the Vienna Biocenter, Austria and at University of Pécs, Hungary.

Description of any significant infrastructure and/or major items of technical equipment, relevant to the proposed work (if applicable):

HPE High Performance Computing (HPC) cluster with 42 computing nodes.

4.2. Third parties involved in the project (including use of third party resources)

Participant 6: Ústav Organické Chemie a Biochemie AV ČR, v.v.i. UOCHB (ELIXIR-CZ)

Does the participant plan to subcontract certain tasks (please note that core tasks of the project should not be sub-contracted)	NO
Does the participant envisage that part of its work is performed by linked third parties	YES

The linked third parties involved are:

- **Univerzita Palackého v Olomouci, Palacky University Olomouc (UP)**
- **Masarykova univerzita, Masaryk University, Brno (MU)**
- **České vysoké učení technické v Praze, Czech Technical University, Prague (CTU)**

Univerzita Palackého v Olomouci, Palacky University Olomouc (UP)

1) Organization Profile UP & Link with UOCHB

The Czech ELIXIR Node is represented by the UOCHB as a national hub for distributed research infrastructure for biological data. The Node is created and operated as a distributed infrastructure, where UP is one of the consortium partners. UP signed the ELIXIR CZ Consortium agreement 4. 3. 2015.

2) Contribution in the project

UP will participate in WP1 in the task T1.1 Network of data managers and scientific best practice. UP will provide two experts into the network, and will participate in documentation of best practices in usage of DMP tools within local university without centralised data management.

3) People involved with a short description of their experience (up to 5)

[redacted]

[redacted]

Description

Palacky University Olomouc is the second-oldest university and one of the leading higher education institutions in the Czech Republic. It comprises 8 faculties and 3 research centres. Palacky University Olomouc have made a considerable financial investment into the research infrastructure in the last decade, mainly funded by EU Structural Funds.

5) Relevant Publications (up to 5)

1) de Beer TAP, Berka K, Thornton JM, Laskowski RA: PDBsum additions. Nucleic Acids Res., 42(D1), 292-296, 2014.
2) Pravda L., Sehnal D., Svobodová Vařeková R., Navrátilová V., Toušek D., Berka K., ... & Koča J.. ChannelsDB: database of biomacromolecular tunnels and pores. Nucleic acids research, 46(D1), D399-D405, 201, 2017.
3) Sehnal D., Deshpande M., Svobodová Vařeková R., Mir S., Berka K., Midlik A., Pravda L., Velankar S., Koča J. LiteMol suite: interactive web-based visualization of large-scale macromolecular structure data. Nat. Methods, 14(12), 1121, 2017
4) Juračka J, Šrejber M, Melíková M, Bazgier V, Berka K: MolMeDB: Molecules on Membranes Database. Database, 2019, baz078, 2019.
5) Varadi M, Berrisford J, Deshpande M, Nair SS, Gutmanas A, Berka K, ... , Koča J, ... , Velankar S: PDBe-KB: a community-driven resource for structural and functional annotations. Nucleic Acids Res., gkz853. *2019*.

4) Relevant projects (up to 5)

Ministry of Education, Youth and Sports LM2015047: (2016-2019) ELIXIR-CZ Czech National Infrastructure for Biological Data

Ministry of Education, Youth and Sports LM2018131: (2020-2022) ELIXIR-CZ Czech National Infrastructure for Biological Data

Czech Science Foundation 17-21122S: (2017-2019) MolMeDB - Verified predictions of interactions of low-molecular substances with biological membranes

5) Significant infrastructures (up to 5)

- ELIXIR CZ servers serving dedicated web services https://molmedb.upol.cz, mole.upol.cz, https://decryptor.imtm.cz/ , https://spci.imtm.cz/ and testing for COVID-19 https://covit.imtm.cz.

Masarykova univerzita, Masaryk University, Brno (MU)

1) Organization Profile MU & Link with UOCHB

CEITEC Masaryk University (MU) closely cooperates with UOCHB via ELIXIR project. MU cooperates with UOCHB via ELIXIR CZ project as one of ELIXIR CZ nodes. The Czech ELIXIR Node is represented by the UOCHB as a national hub for distributed research infrastructure for biological data. The Node is created and operated as a distributed infrastructure, where MU is one of the consortium partners. MU signed the ELIXIR CZ Consortium agreement 4. 3. 2015.

2) *Contribution in the project*

MU will participate in WP1 and WP3, specifically in these tasks and the following way:

- T1.1 Network of data managers and scientific best practice: Provide 3 experts into the network. Participate in documentation of best practices. Provide inputs to development of DMP tools (developed in WP3). Use and test tools developed in WP3. Take care, that information generated by users are appropriately structures and annotated.
- T1.2 Models for brokering data to ELIXIR Deposition Databases: Explore existing practices around data brokering (we are directly in PDBe brokering process). Document and resolve best practices.
- T1.4 Sustainable and scalable operating model for harmonised data management in European projects: Based on identified scientific best practice and brokering services, support a design an operating model for providing joint data management.
- T3.3 Access portal to Toolkit tailored to stakeholders: Support the development by feedback and testing.

3) *People involved with a short description of their experience (up to 5)*

[redacted]

Description

Masaryk University is the second-largest public university in the Czech Republic and the leading higher education institution in Moravia. It comprises 9 faculties with over 200 departments, institutes and clinics.

Masaryk University have made a considerable financial investment in research infrastructure in the last couple of years, mainly funded by EU Structural Funds. This effort led to creation of CEITEC – Central European Institute of Technology ("CEITEC"), which is a project focused on establishment of a European centre of excellence in the area of life sciences and advanced materials and technologies, jointly implemented by a consortium of six partners under the coordination of Masaryk University.

4) *Relevant Publications (up to 5)*

1) Laskowski R. A., Jabłońska J., Pravda L., Svobodová Vařeková R., Thornton J. M. PDBsum: Structural summaries of PDB entries. Protein Science, 27(1), 129-134, 2018.
2) Pravda L., Sehnal D., Svobodová Vařeková R., Navrátilová V., Toušek D., Berka K., ... & Koča J.. ChannelsDB: database of biomacromolecular tunnels and pores. Nucleic acids research, 46(D1), D399-D405, 201, 2017.
3) Sehnal D., Deshpande M., Svobodová Vařeková R., Mir S., Berka K., Midlik A., Pravda L., Velankar S., Koča J. LiteMol suite: interactive web-based visualization of large-scale macromolecular structure data. Nat. Methods, 14(12), 1121, 2017
4) Adams, P. D., Aertgeerts, K., ..., Svobodová Vařeková R., ... & Burley, S. K.. Outcome of the first wwPDB/CCDC/D3R ligand validation workshop. Structure, 24(4), 502-508, 2016

5) Sehnal D., Pravda L., Svobodová Vařeková R., Ionescu C.-M., Koča J.: PatternQuery: web application for fast detection of biomacromolecular structural patterns in the entire Protein Data Bank. Nucleic acids research, 44, D38-D47, 2015.

5) Relevant projects (up to 5)

EU H2020 (H2020-INFRADEV-1-2015-1) ELIXIR-EXCELERATE: Fast-track ELIXIR implementation and drive early user exploitation across the life sciences, Grant number: 676559, 2015-2019

6) Significant infrastructures (up to 5)

- ELIXIR server (48 cores, 512 GB RAM) and disk (100 TB) at dedicated Node of NGI for national Czech ELIXIR connected to the CESNET infrastructure by 10 Gb/s. Capacity limits up to 500 cores. Know/how for integration into ELIXIR infrastructure.
- Dedicated network connection of the ELIXIR CZ Node and its subnodes. Internal network infrastructure upon request.

České vysoké učení technické v Praze, Czech Technical University, Prague (CTU)

1) Organization Profile CTU & Link with UOCHB

CTU cooperates with UOCHB via ELIXIR CZ project as one of ELIXIR CZ nodes. The Czech ELIXIR Node is represented by the UOCHB as a national hub for distributed research infrastructure for biological data. The Node is created and operated as a distributed infrastructure, where CTU is one of the consortium partners. CTU signed the ELIXIR CZ Consortium agreement 4. 3. 2015.

2) Contribution in the project

WP3 – active participation on WP3 meetings, participation on the toolkit deliverable plus potential other WP3 tasks coming from the meetings, promotion and implementation of DSW, coordination and harmonisation with local ELIXIR CZ activities

WP2 – knowledge support for training

3) People involved with a short description of their experience (up to 5)

[redacted]

Description

The Czech Technical University in Prague has implemented and is implementing many R & D & I projects with a number of companies in the Czech Republic and around the world. Although the Faculty of Information Technology was established in 2009 and is the youngest faculty of the Czech Technical University, it actively cooperates with a number of key ones industrial and research partners.

4) Relevant Publications (up to 5)

1) Pergl, R., Hooft, R., Suchánek, M., Knaisl, V., & Slifka, J. (2019). "Data Stewardship Wizard": A Tool Bringing Together Researchers, Data Stewards, and Data Experts around Data Management Planning. Data Science Journal, 18, 59. https://doi.org/10.5334/dsj-2019-059
2) Lee, S. et al. (2020). Data Management Planning: How Requirements and Solutions are Beginning to Converge. Data Intelligence, 2, 208–219. https://doi.org/10.1162/dint_a_00043
3) Wittenburg, P. et al. (2019). The FAIR Funder pilot programme to make it easy for funders to require and for grantees to produce FAIR Data. http://arxiv.org/abs/1902.11162
4) Suchánek, M.. Pergl, R. (2018). Data Stewardship Wizard for Open Science. In: Data a znalosti & WIKT. Brno: Vysoké učeni technické v Brně. Fakulta informačnich technologii, p. 121-125. 1. ISBN 978-80-214-5679-2.
5) Suchánek, M.; Slifka, J. Evolvable and Machine-Actionable Modular Reports for Service-Oriented Architecture. In: Enterprise and Organizational Modeling and Simulation. Springer, Cham, 2019. p. 43-59. 1. vol. 366. ISSN 1865-1348. ISBN 978-3-030-35645-3.

5) Relevant projects (up to 5)

- Towards Data Stewardship in ELIXIR: Training & Portal (ELIXIR Implementation Study)
- ELIXIR CZ Research Infrastructure Project (MEYS Grant No. LM2015047)

6) Significant infrastructures (up to 5)

- ELIXIR CZ (member of the Committee)
- GO FAIR projects & Implementation Networks involvement

PMs and Budget distribution

Linked Third parties against payment	PM	Personnel Cost	Other Direct Costs	Indirect Costs (25%)
UP	2	€ 10,920.00	0	€ 2,730.00
MU	5	€ 27,300.00	0	€ 6,825.00
CTU	5.5	€ 30,030.00	0	€ 7,507.50
LTP in total	12.5	€ 68,250.00	0	€ 17,062.50

Activity

Description of activities to be carried out by MU:
WP1:
T1.1 Network of data managers and scientific best practice
T1.2 Models for brokering data to ELIXIR Deposition Databases
T1.4 Sustainable and scalable operating model for harmonised data management in European projects
WP3:
T3.3 Access portal to Toolkit tailored to stakeholders

Description of activities to be carried out by CTU:
WP3 – active participation on WP3 meetings, participation on the toolkit deliverable plus potential other WP3 tasks coming from the meetings, promotion and implementation of DSW, coordination and harmonisation with local ELIXIR CZ activities
WP2 – knowledge support for training

Description of activities to be carried out by UP:

WP1: T1.1 Network of data managers and scientific best practice All participating institutions have their own allocated budget composed of personnel costs and indirect costs.	
*Does the participant envisage the use of contributions in kind provided by third parties (**Articles 11 and 12** of the Model Grant Agreement)*	NO
Does the participant envisage that part of the work is performed by International Partners (Article 14a of the PRIMA Model Grant Agreement)?	NO

Participant 10: Barcelona Supercomputing Center. BSC (ELIXIR-ES)

Does the participant plan to subcontract certain tasks (please note that core tasks of the project should not be sub-contracted)	NO
Does the participant envisage that part of its work is performed by linked third parties	NO
*Does the participant envisage the use of contributions in kind provided by third parties (**Articles 11 and 12** of the Model Grant Agreement)*	**YES**

The third parties making in-kind contributions against payment (Article 11) are Universitat Pompeu Fabra (UPF), and the Institute for Research in Biomedicine (IRB)

- **Universitat Pompeu Fabra (UPF)**

1) *Organization Profile*

Universitat Pompeu Fabra (UPF) is a public, international and research-intensive university that, in just twenty-five years, has earned a place among the best universities in Europe. Awarded with a CEI label (International Excellence Campus) by the Spanish Ministry of Education, the University is present in some of the most influential rankings (e.g. The Times Higher Education ranking (2016) ranked UPF as 2nd best Spanish university and as the 15th in the world among those under 50 years of age). UPF is one of the few European universities that have obtained the Standards of Good Practice certification by the US Forum on Education Abroad. UPF participates in **ELIXIR CONVERGE** through the Research Programme on Biomedical Informatics (GRIB) (http://grib.upf.edu/), which is a joint programme of the UPF and the IMIM (Hospital del Mar Research Institute). GRIB carries out methodological research and technological developments on the application of advanced information technologies and computational methods in the health and life sciences. GRIB has a long record of participation in EU-funded projects and collaboration with pharmaceutical industry in knowledge management, translational bioinformatics and predictive modelling.

2) *Link with BSC*

GRIB is the node for Biomedical Informatics of the Spanish Institute of Bioinformatics (INB) coordinated by the BSC. BSC is partner of IMI2 Project eTRANSAFE coordinated by the GRIB. BSC and GRIB are both partners of IMI Project FAIRplus.

3) *Contribution in the project*

Our contribution is focused on WP5 (Demonstrator projects on Data Management Plans). Based on our expertise on projects related to knowledge management and predictive toxicology modelling, and our current participation in initiatives focused on data FAIRification for omics and toxicology data (e.g. FAIRplus), we will contribute in the design and implementation of data management plans for toxicology data, with particular emphasis in its FAIRification.

4) *People involved with a short description of their experience (up to 5)*

[redacted]

5) *Relevant Publications (up to 5)*

1) Piñero J, Bravo À, Queralt-Rosinach N, Gutiérrez-Sacristán A, Deu-Pons J, Centeno E, García-García J, Sanz F, Furlong LI. DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. Nucleic Acids Res. 2017; 45(D1): D833-D839. doi: 10.1093/nar/gkw943.

2) Sanz F, Pognan F, Steger-Hartmann T, Díaz C; eTOX, Cases M, Pastor M, Marc P, Wichard J, Briggs K, Watson DK, Kleinöder T, Yang C, Amberg A, Beaumont M, Brookes AJ, Brunak S, Cronin MTD, Ecker GF, Escher S, Greene N, Guzmán A, Hersey A, Jacques P, Lammens L, Mestres J, Muster W, Northeved H, Pinches M, Saiz J, Sajot N, Valencia A, van der Lei J, Vermeulen NPE, Vock E, Wolber G, Zamora I. Legacy data sharing to improve drug safety assessment: the eTOX project. Nat Rev Drug Discov. 2017; 16(12):811-812. doi: 10.1038/nrd.2017.177.

3) Piñero J, Furlong LI, Sanz F. In silico models in drug development: where we are. Curr Opin Pharmacol. 2018 Oct;42:111-121. doi: 10.1016/j.coph.2018.08.007.

4) Digles D, Zdrazil B, Neefs JM, Van Vlijmen H, Herhaus C, Caracoti A, Brea J, Roibás B, Loza MI, Queralt-Rosinach N, Furlong LI, Gaulton A, Bartek L, Senger S, Chichester C, Engkvist O, Evelo CT, Franklin NI, Marren D, Ecker GF, Jacoby E. Open PHACTS computational protocols for in silico target validation of cellular phenotypic screens: knowing the knowns. Medchemcomm. 2016 Jun 1;7(6):1237-1244.

5) Queralt-Rosinach N, Piñero J, Bravo À, Sanz F, Furlong LI. DisGeNET-RDF: harnessing the innovative power of the Semantic Web to explore the genetic basis of diseases. Bioinformatics. 2016 Jul 15;32(14):2236-8. doi:10.1093/bioinformatics/btw214.

6) Relevant projects (up to 5)

H2020 ELIXIR-EXCELERATE (2015-19): EXCELERATE drove the development of ELIXIR Platforms, i.e. the common services, and the development of ELIXIR Communities where services are applied in international collaboration. This project builds on EXCELERATE and will align the national data management initiatives. Funded by national roadmaps within ELIXIR Nodes - with the overall ELIXIR Program.

IMI FAIRplus (2019-21) is a public-private partnership (7 Pharma, 3 SME and 12 academic partners) coordinated by ELIXIR that will make data and other assets from past and on-going IMI projects FAIR and publish these resources for reuse by the research community. IMI FAIRplus will develop "FAIR cookbooks" and a "FAIR capability maturation model (FAIR-CMMT)" to provide projects with a clear path towards FAIR data management together with a value driven maturation model to focus efforts on making high value datasets extensively annotated and reusable.

H2020 EU-ToxRisk (2016-21): An integrated European 'flagship' program driving mechanism-based toxicity testing and risk assessment.

IMI2 eTRANSAFE (2017-22) Development of an integrative data infrastructure and innovative computational methods and tools that aim to drastically improve the feasibility and reliability of translational safety assessment during the drug development process. [redacted] is the academic coordinator of the consortium.

H2020 MedBioinformatics (2015-18): Creating medically-driven integrative bioinformatics applications focused on oncology, CNS disorders and their comorbidities. [redacted] was the coordinator of the consortium.

7) Significant infrastructures (up to 5)

The GRIB has access to the infrastructure (hardware and software) required for the execution of the project.

PMs and Budget distribution

In the frame of this proposal, the BSC has 24 PMs assigned, with 142.660 € for personal costs and 9.000 € as Other Direct Costs. PMs and budget will be re-distributed to UPF as follows:

Third party against payment	PM	Personnel Cost	Other Direct Costs	Indirect Costs (25%)

UPF	4	23,776 €	750 €	6,131.50 €
Activity				
UPF will contribute to **Task 5.2** bringing one of the selected use-cases for which a DMP will be assembled and developed. Furthermore, UPF will play an active role within task 5.3 to help refine key performance indicators (KPIs) to measure and monitor demonstrator projects' implementation of DMPs and task 5.4 where UPF will contribute to create the guidelines to develop the DM based on what developed in Task 5.2				

Institute for Research in Biomedicine (IRB)

1) Brief description of the Organization

IRB Barcelona is an independent, non-profit research institution engaged in basic and applied biomedical science. It was founded by the Catalan Government and the University of Barcelona (UB) in October 2005 and it is a member of the CERCA Research Centre Network promoted by the former (www.cerca.cat). Housed within the Barcelona Science Park (PCB) on the University of Barcelona Campus, IRB Barcelona is an integral part of the "Bioregion of Catalonia", a rich landscape of research centres of excellence.

The Institute's missions include conducting multidisciplinary research of excellence at the unique interface between biology, chemistry and medicine, providing high-level training in the biomedical sciences to staff, students and visitors, driving innovation through active technology transfer to the benefit of society, and actively participating in an open dialogue with the public through a series of engagement and education activities.

Headed by [redacted] IRB Barcelona is home to nearly 400 research, technical, support and administrative staff from 30 countries. Its current 26 groups are organised into three Research Programmes, namely Cancer Science, Aging and Metabolism, and Mechanisms of Disease. Well-equipped laboratories are supported by a network of seven state-of-the-art scientific core facilities. These in turn are complemented by platforms and scientific services offered by the PCB (Barcelona Science Park) and the University of Barcelona.

Training the next generation of scientists is one of the main missions of IRB Barcelona. In this regard, 112 students—almost half non-Spanish nationals—are currently working toward their PhD theses at the centre. The final goal of the Institute is to encourage young people to carry out creative scientific projects at the highest standard and to train future leaders for the academic and industrial sectors. IRB Barcelona is also channelling efforts into the consolidation of an institutional postdoctoral programme focused on recruiting and training a highly international postdoctoral community, which currently comprises nearly 100 researchers.

The Institute is a member of the Barcelona Institute of Science and Technology (BIST), a scientific initiative of seven research centres of excellence in Catalonia whose objective is to build a joint scientific project to foster research of excellence in a multidisciplinary environment. BIST is therefore the umbrella institution of a group of highly interdisciplinary research centres of excellence covering a wide range of fields.

IRB Barcelona devotes considerable effort to initiatives aiming to actively engage in an open dialogue with the public and patients' associations. This dialogue is achieved through a wide range of science, outreach, dissemination, and training activities targeting students, teachers, and the general public. The Institute also actively promotes collaboration with top research hospitals, with the aim to foster translational research. Agreements are ongoing with IDIBAPS and VHIR (the research institutes of the Hospital Clínic de Barcelona and Hospital Vall d'Hebrón respectively).

The Institute has put into place an External Advisory Board (EAB) and at Business Advisory Board (BAB), structures that ensure that it delivers on the full promise of its potential. During the period 2011 – 2018, the performance of the Institute was remarkable. Examples are the steady increase in publications in top-ranked journals (>75% in D1, >95% in Q1 - SJR), the award of 17 ERC grants (6 Advanced, 3 Consolidator, 3 Starting and 5 PoC grants), and numerous technology transfer actions in collaboration with the private sector. Furthermore, IRB Barcelona received the "Severo Ochoa" Award of Excellence from the former Spanish Ministry of Science and Innovation in 2011 and in 2015 and the "HR Excellence in Research" logo from the European Commission. Indeed, as part of this HR Excellence in Research recognition from the EU, the Institute has been advancing in several plans and initiatives to continuously improve its human resources policies to ensure that they are in line with the European Charter for Researchers and the Code of Conduct for the Recruitment of Researchers.

2) *Link with BSC*

IRB Barcelona is a Linked Third Party of BSC (art. 14 Annotated Model Grant Agreement V5.2, 26 June 2019).

3) *Contribution in the project*

FAIR organisation of biomolecular simulation information (2nd wave; ES, SE).
Data types & Challenge: long coordinate files containing the evolution of 3D positions of atoms of the system across time; output of a wide variety of analysis tools applied to raw data, metadata describing system setup and the simulation parameters. Simulation data is typically stored locally in an undocumented manner, without any external curation and lacking associated metadata required for their reusability.
Key Deposition databases, standards and Interoperability resources: Some initial attempts in the generation of usable ontologies (UMM, BigNASim).

4) *People involved with a short description of their experience (up to 5)*

[redacted]

[redacted]

5) Relevant Publications (up to 5)

1. A.Arcella, J.Dreyer, E.Ippoliti, I.Ivani, G.Portella, V.Gabelica, P.Carloni and **M.Orozco**. “Structure and dynamics of oligonucleotides in the gas phase”. Angew. Chem. Int. Ed.Eng. (2015), 127, 477-481.
2. G.Portella, M.Terrazas, N.Villegas, C.Gouzález and **M.Orozco**. “Can a denaturant stabilize DNA? Pyridine reverses DNA denaturation in acidic pH”. Angew. Chem. Int. Ed. Eng. (2015), 54, 10488-10491.
3. I.Ivani, P.D. Dans, A.Noy, A.Pérez, I.Faustino, A.Hospital, J.Walther, P.Andrio, R.Goñi, A.Balaceanu, G.Portella, F.Battistini, J.L.Gelpi, C.González, M.Vendruscolo, C.A.Laughton, S.A.Harris, D.A.Case and M.Orozco. “ParmBSC1: a refined force-field for DNA simulations”. Nature Methods., (2016), 13, 55-58. Selected for Faculty of 1000 essentials.
4. R.Collepardo-Guevara, G.Portella, M.Vendruscolo, D.Frenkel, T.Schlick and M.Orozco. “Chromatin unfolding by epigenetic modifications explained by dramatic impairment of internucleosome interactions: a pioneering multiscale computational study”. J.Am.Chem.Soc. (2015), 137, 10205-10215

5. L.Darré, I.Ivani, P.D.Dans, H.Gómez, A.Hospital and **M.Orozco**. "Small details matter: the 2'hydroxyl as a conformational switch in RNA". J.Am.Chem.Soc. (2016), 44, 4052-66.

6) Relevant projects (up to 5)

H2020 European Commision. EU-Center of Excellence in High Performance Computing. BIOEXCEL A center of Excellence in Biosimulation". Amount granted to the MMB Group: 225,000 Euros. 2015-2018

H2020 European Commision. ELIXIR-Excelerate: Fast-track ELIXIR implementation and drive early user exploitation across the life-science. Amount granted to the MMB Group. 158,400 Euros. 2015-2019.

H2020 European Commision. MuG. "Virtual Research Environment for Multidimensional Genomics". Coordinator M.Orozco. Amount granted to the MMB Group: 482,900 Euros. 2015-2018

H2020 European Commision. "BioExcel-2. Centre of Excellence for Computational Biomolecular Research." Principal Investigator M.Orozco. Amount 455,625 Euros. 2019-2021.

RTI2018-096704-B-100. Ministerio de Ciencia, Innovación y Universidades. "Resolución simulación de ácidos nucleicos complejos" Principal Investigator M.Orozco. Amount 338.800 Euros. 2019-2021

7) Significant infrastructures (up to 5)

The lab of [redacted] is fully equipped. Its main components are:

- High Performance Computing (HPC) Cluster with 62 nodes, making a total of 992 Cores and 2 TBytes of memory.
- Hybrid CPU/GPU HPC Cluster with 19 nodes making a total of 90 GPUs cards, 412 cores and 3 TBytes of memory.
- 80 Tbytes of storage.

Besides, [redacted] has access to the Starlife Infrastructure, an IT infrastructure shared between IRB, CRG and BSC. It is composed of:

- 54 nodes making a total of 2160 cores and with different memory configurations: 18 "cloud" nodes making a total of 2,880 TBytes of memory; 33 "thin" nodes making a total of 5,3 TBytes of memory; 3 "fat" nodes making a total of 1 TBytes of memory.
- 2 PBytes of Hard drive (hot storage).
- 4,5 PBytes of tape drive storage (cold storage).

PMs and Budget distribution

In the frame of this proposal, the BSC has 24 PMs assigned, with 142.660 € for personal costs and 9.000 € as Other Direct Costs. PMs and budget will be re-distributes among the third parties as follows:

Third party against payment	PM	Personnel Cost	Other Direct Costs	Indirect Costs (25%)
IRB	4	23,776 €	750€	6,131.50 €

Activity
IRB will contribute to **Task 5.2** which will implement a toolkit It will be assembled to develop and implement aDMP for each of those projects. Furthermore, IRB will play an active role within task 5.3 to help refine key performance indicators (KPIs) to measure and monitor demonstrator projects' implementation of DMPs and task 5.4 where they will contribute to create the guidelines to develop the DM based on what developed in Task 5.2.

Article 12

Some of the work carried out at the Barcelona Supercomputing will be contributed free of charge by Third Party: the Catalan Institution for Research and Advanced Studies (ICREA).
ICREA (Institució Catalana de Recerca i Estudis Avançats) will provide resources (professor/researcher) free of charge to BSC as a third party (Article 12 Grant Agreement). ICREA is a foundation supported by the Catalan Government and guided by a Board of Trustees aiming to recruit top scientists for the Catalan R&D system: scientists capable of leading new research groups, strengthening existing groups, and setting up new lines of research.

Following the rules of ICREA, although the salary costs of [redacted] *are paid by ICREA, he is assigned to physically work at BSC and considered as a member of the BSC. The terms and conditions of this cooperation between ICREA and BSC are reflected in bilateral agreements between the two parties.*

The beneficiary BSC is free to use these resources at will. They are therefore assimilated as "own resources" of the beneficiary, and will be charged to the project without being considered as a receipt. The cost will be declared by the beneficiary and it will be recorded in the accounts of the third party. These accounts will be available for auditing if required. [redacted] *is ICREA Research Professor affiliated with the BSC. He carries out his research at the BSC. Moreover, other ICREA Research Professors can be included in the team in the future.*
It is foreseen a contribution of 1 PM from [redacted] *in the whole action.*

Does the participant envisage that part of the work is performed by International Partners (Article 14a of the PRIMA Model Grant Agreement)?	NO

Participant 17 - Research Centre for Natural Sciences TTK (ELIXIR-HU)

Does the participant plan to subcontract certain tasks (please note that core tasks of the project should not be sub-contracted)	NO
If yes, please describe and justify the tasks to be subcontracted	

*Does the participant envisage that part of its work is performed by linked third parties**	NO
If yes, please describe the third party, the link of the participant to the third party, and describe and justify the foreseen tasks to be performed by the third party:	
Does the participant envisage the use of contributions in kind provided by third parties (Articles 11 and 12 of the General Model Grant Agreement)	YES
If yes, please describe the third party, the link of the participant to the third party, and describe and justify the foreseen tasks to be performed by the third party:	

1. **University of Debrecen (UD)**

Organization

UD was a pioneer among the Hungarian universities in using the most advanced functional genomics techniques in biomedical research. Together with an Illumina HiScan sequencer a high performance computing cluster was installed as well (84 cores, 35TB disk capacity, high memory node etc.). This cluster is dedicated completely to the analysis and storing of next generation sequencing data and will be significantly expanded by the end of 2017. The Department also hosts two proteomics databases. TRANSDAB (http://genomics.dote.hu/wiki/index.php/Main_Page) transglutaminase substrates and interaction partners collection. FuzDB (http://protdyn-database.org/) compiles experimentally observed fuzzy protein complexes, where intrinsic disorder is maintained upon interacting with a partner. During these years, UD implemented the education of Bioinformatics both at MSc and PhD level. Moreover, UD organized several bioinformatic and genomic training courses with the involvement of experts from the EMBL, DKFZ, EMBO, HHMI and relevant commercial players of the field (e.g. GeneSpring). These training activities will be the basis of future bioinformatic training activities.

Main tasks

UD, together with PTE and NARIC, will co-organize the network of data managers and processes to enrich and maintain the common data management toolkit. UD, together with PTE will identify training needs and solutions and develop best practice guidelines and training materials in Data Management and Stewardship. A Demonstrator project will be held at UD or PTE for the implementation of machine-actionable data management plans.

Staff involved

[redacted]

[redacted]

List of publications (top 5)

1. Veto B, Bojcsuk D, Bacquet C, Kiss J, Sipeki S, Martin L, Buday L, Balint BL, Aranyi T: The transcriptional activity of hepatocyte nuclear factor 4 alpha is inhibited via phosphorylation by ERK1/2., PLOS ONE 12: (2) p. e0172020. 2017.
2. Dóra Bojcsuk, Gergely Nagy, Balint L Balint: Inducible super-enhancers are organized based on canonical signal-specific transcription factor binding elements, NUCLEIC ACIDS RESEARCH 45: (7) pp. 3693-3706. 2016.
3. Blaszczyk K, Olejnik A, Nowicka H, Ozgyin L, Chen YL, Chmielewski S, Kostyrko K, Wesoly J, Balint BL, Lee CK, Bluyssen HA: STAT2/IRF9 directs a prolonged ISGF3-like transcriptional response and antiviral activity in the absence of STAT1, BIOCHEMICAL JOURNAL 466: (3) pp. 511-524. 2015.
4. Ozgyin L, Erdos E, Bojcsuk D, Balint BL: Nuclear receptors in transgenerational epigenetic inheritance., PROGRESS IN BIOPHYSICS AND MOLECULAR BIOLOGY 118: (1-2) pp. 34-43. 2015.
5. Daniel B, Nagy G, Hah H, Horvath A, Czimmerer Z, Poliska P, Gyuris T, Keirsse J, Gysemans C, Ginderachter J, Balint B L, Evans R M, Barta E, Nagy L: The Active Enhancer Network Operated by Liganded RXR Supports Angiogenic Activity in Macrophages, GENES & DEVELOPMENT 28: (14) pp. 1562-1577. 2014.

2. University of Pécs (PTE)

Organization

János Szentágothai Research Centre of the University of Pécs is a new research institute established on the basis of modern international science organizational and management normative. It covers all aspects of education, research and innovation at the fields of biomedical, natural and environmental sciences. The infrastructure, instrumentation and expertise of the 22 research groups operating in the building provide an excellent basis to become a well-known, leading research facility in Hungary, as well as in Central-Europe, with an extensive and fruitful collaboration network. The newly establishing bioinformatics group already provides data analysis services and statistical support for pharmaco-, functional- and molecular genomics projects.

Main tasks

UD, together with PTE and NARIC will co-organize the network of data managers and processes to enrich and maintain the common data management toolkit. UD, together with PTE will identify training needs and solutions and develop best practice guidelines and training materials in Data Management and Stewardship. A Demonstrator project will be held at UD or PTE for the implementation of machine-actionable data management plans.

Staff involved (must be employed by the institution)

[redacted]

List of publications (top 5)

1. Gyenesei, Attila ; Semple, Colin A. M. ; Haley, Chris S. ; Wei, Wen-Hua: High throughput analysis of epistasis in genome-wide association studies with BiForce (vol 28, pg 1957, 2012) BIOINFORMATICS 29 : 20 pp. 2667-2668. , 2 p. (2013)
2. Babu, M ; Durga, Devi T ; Makinen, P ; Kaikkonen, M ; Lesch, HP ; Junttila, S ; Laiho, A ; Ghimire, B ; Gyenesei, A ; Yla-Herttuala, S: Differential Promoter Methylation of Macrophage Genes Is Associated With Impaired Vascular Growth in Ischemic Muscles of Hyperlipidemic and Type 2 Diabetic Mice: Genome-Wide Promoter Methylation Study. CIRCULATIONRESEARCH 117 : 3 pp. 289-299. , 11 p. (2015)
3. Koehler, C ; Sauter, PF ; Wawryszyn, M ; Girona, GE ; Gupta, K ; Landry, JJ; Fritz, MH ; Radic, K ; Hoffmann, JE ; Chen, ZA et al. Genetic code expansion for multiprotein complex engineering. NATURE METHODS 13 : 12 pp. 997-1000. , 4 p. (2016)
4. Stomfai, Sarolta ; Gyenesei, Attila ; Ahrens, Wolfgang ; De, Henauw Stefaan ; Gunther, Kathrin ; Marild, Staffan ; Mehlig, Kirsten ; Moreno, Luis A. ; Siani, Alfonso ; Tornaritis, Michael et al.: Non-Invasive Cardiovascular Risk Assessment in Prepubertal Children, ANNALS OF NUTRITION AND METABOLISM 73 : 4 pp. 316-316. , 1 p. (2018)
5. Bauer, Witold ; Veijola, Riitta ; Lempainen, Johanna ; Kiviniemi, Minna ; Härkönen, Taina ; Toppari, Jorma ; Knip, Mikael ; Gyenesei, Attila ; Ilonen, Jorma: Age at Seroconversion, HLA Genotype and Specificity of Autoantibodies in Progression of Islet Autoimmunity in Childhood. JOURNAL OF CLINICAL ENDOCRINOLOGY AND METABOLISM 104 : 10 pp. 4521-4530. , 10 p. (2019)

3. National Agricultural Research and Innovation Centre (NARIC)

Organization

The Agricultural Biotechnology Institute (ABC) of NARIC in Gödöllő always played an important role to contribute to the field of bioinformatics in Hungary. In 1994, ABC was recognized as the Hungarian EMBnet (European Molecular Biology Network) node and [redacted] had been appointed as the node manager. As the national node, we hosted molecular biology databases and provided sequence analysis programs for the Hungarian academic scientists free of charge. Therefore, we also have a long-term experience in setting up and maintaining bioinformatics servers. Recently, we are looking for the arrival and installation of our newest LINUX server cluster, which will offer more than 70TB storage, 56 cores and 500

GB memory. The main genomics related expertise in the ABC is focused on the plant small RNA-seq analysis, bacterial genomics research and genome sequencing and genome based genetic analysis of farm animals.

Main tasks

UD, together with PTE and NARIC will co-organize the network of data managers and processes to enrich and maintain the common data management toolkit. UD, together with PTE will identify training needs and solutions and develop best practice guidelines and training materials in Data Management and Stewardship. Demonstrator project will be held at UD or PTE for the implementation of machine-actionable data management plans.

Staff involved (must be employed by the institution)

[redacted]

List of publications (top 5)

1. Nóra Á Bana, Anna Nyiri, János Nagy, Krisztián Frank, Tibor Nagy, Viktor Stéger, Mátyás Schiller, Péter Lakatos, László Sugár, Péter Horn, Endre Barta, László Orosz: The red deer Cervus elaphus genome CerEla1.0: sequencing, annotating, genes, and chromosomes, MOLECULAR GENETICS AND GENOMICS 293: (3) pp. 665-684. 2018.
2. Nagy G, Czipa E, Steiner L, Nagy T, Pongor S, Nagy L, Barta E: Motif oriented high-resolution analysis of ChIP-seq data reveals the topological order of CTCF and cohesin proteins on DNA., BMC GENOMICS 17: (1) 637. 2016.
3. Baksa I, Nagy T, Barta E, Havelda Z, Varallyay E, Silhavy D, Burgyan J, Szittya G: Identification of Nicotiana benthamiana microRNAs and their targets using high throughput sequencing and degradome analysis., BMC GENOMICS 16: (1) pp. 1025-1045. 2015.
4. Molnar J, Nagy T, Steger V, Toth G, Marincs F, Barta E: Genome sequencing and analysis of Mangalica, a fatty local pig of Hungary., BMC GENOMICS 15: (1) 761. 2014.
5. Daniel B, Nagy G, Hah H, Horvath A, Czimmerer Z, Poliska P, Gyuris T, Keirsse J, Gysemans C, Ginderachter J, Balint B L, Evans R M, Barta E, Nagy L: The Active Enhancer Network Operated by Liganded RXR Supports Angiogenic Activity in Macrophages, GENES & DEVELOPMENT 28: (14) pp. 1562-1577. 2014.

Overview of the budget

Third party	**PM**	**Personnel**	**Other Direct Costs**	**Activity**

NARIC	3	15000 €	1500 €	Co-organize the network of data managers and processes to enrich and maintain the common data management toolkit.
PTE	4	20000 €	1500 €	Co-organize the network of data managers and processes to enrich and maintain the common data management toolkit. Identify training needs and solutions and develop best practice guidelines and training materials in Data Management and Stewardship. Demonstrator project for the implementation of machine-actionable data management plans.
UD	4	20000 €	1500 €	Co-organize the network of data managers and processes to enrich and maintain the common data management toolkit. Identify training needs and solutions and develop best practice guidelines and training materials in Data Management and Stewardship. Demonstrator project for the implementation of machine-actionable data management plans.

Participant 18 - University College Dublin. UCD (ELIXIR-IE)

Does the participant plan to subcontract certain tasks (please note that core tasks of the project should not be sub-contracted)	NO
If yes, please describe and justify the tasks to be subcontracted	
*Does the participant envisage that part of its work is performed by linked third parties**	NO
If yes, please describe the third party, the link of the participant to the third party, and describe and justify the foreseen tasks to be performed by the third party:	
Does the participant envisage the use of contributions in kind provided by third parties (Articles 11 and 12 of the General Model Grant Agreement)	YES

If yes, please describe the third party, the link of the participant to the third party, and describe and justify the foreseen tasks to be performed by the third party:

1. UCC

Organization

University College Cork (UCC) is an internationally competitive, research-led university that plays a key role in the development of Ireland's knowledge-based economy. Our institutional research strategy focuses on creating and supporting world-leading clusters of researchers, building on the research strengths of the University, and is aligned with key Government strategies including the *Strategy for Higher Education to 2030 (Hunt Report)*, and Ireland's Strategy for Research and Development, Science and Technology, *Innovation 2020*.

The UCC Research & Innovation Strategic Plan (2017-2022) details a number of performance targets for research and economic development that seek to position UCC and Ireland as an international innovation development hub.

Main tasks

[redacted] (UCC) will advise the ELIXIR-Ireland CONVERGE grouping at UCD, particularly the data manager to be employed under CONVERGE, on aspects of Task 2 (Tasks 2.1, 2.2 and 2.3).

Staff involved

[redacted]

2019 publications:

Kiniry SJ, Michel AM, Baranov PV. (2019) Computational methods for ribosome profiling data analysis. WIREs RNA. in press. (PDF).

Ning Z, Williams JM, Kumari R, Baranov PV, Moore T. (2019) Opposite Expression Patterns of Spry3 and p75NTR in Cerebellar Vermis Suggest a Male-Specific Mechanism of Autism Pathogenesis. Front Psychiatry. 10:416. (PDF).

Palu CC, Ribeiro-Alves M, Wu Y, Lawlor B, Baranov PV, Kelly B, Walsh P. (2019) Simplicity DiffExpress: a bespoke cloud-based interface for RNA-seq differential expression modeling and analysis. Frontiers in Genetics 10:356. (PDF).

Meydan S, Marks J, Klepacki D, Sharma V, Baranov PV, Firth AE, Margus T, Kefi A, Vazquez-Laslop N, Mankin AS. (2019) Retapamulin-Assisted Ribosome Profiling Reveals the Alternative Bacterial Proteome. Mol Cell pii: S1097-2765(19)30107-8. (PDF).

Gerresheim GK, Bathke J, Michel AM, Andreev DE, Shalamova LA, Rossbach O, Hu P, Glebe D, Fricke M, Marz M, Goesmann A, Kiniry SJ, Baranov PV, Shatsky IN, Niepmann M. (2019) Cellular Gene Expression during Hepatitis C Virus Replication as Revealed by Ribosome Profiling. Int J Mol Sci. 20:1321 (PDF).

Hardy S, Kostantin E, Wang SJ, Hristova T, Galicia-Vazquez G, Baranov PV, Pelletier J, Tremblay ML. (2019) Magnesium-sensitive upstream ORF controls PRL phosphatase expression to mediate energy metabolism. Proc Natl Acad Sci U S A. 116:2925-2934 (PDF).

Kiniry SJ, O'Connor PBF, Michel AM, Baranov PV. (2019) Trips-Viz: a transcriptome browser for exploring Ribo-Seq data. Nucleic Acids Res 47:D847-D852 (PDF)

2. NUI Galway

Organization

National University of Ireland Galway (NUIG) is a university founded in 1845, with over 2000 academic staff over a wide range of disciplines, and over 17,000 students.

Main tasks

[redacted] will contribute to Tasks 2.1, 2.2 and 2.3, in his capacity as training coordinator of the ELIXIR Ireland node.

Staff involved

[redacted]

List of publications

Geeleher P, Nath A, Wang F, Zhang Z, Barbeira AN, Fessler J, Grossman RL, **Seoighe** C, Stephanie Huang R. Cancer expression quantitative trait loci (eQTLs) can be determined from heterogeneous tumor gene expression data by modeling variation in tumor purity. *Genome Biology* 2018 Sep 11;19(1):130.

Seoighe C, Tosh NJ, Greally JM.DNA methylation haplotypes as cancer markers (*Brief Communications Arising*). *Nature Genetics* 2018 Aug;50(8):1062-1063

Yu Y, Ceredig R, **Seoighe** C. A Database of Human Immune Receptor Alleles Recovered from Population Sequencing Data. *Journal of Immunology* 2017 Mar 1;198(5):2202-2210.

Seoighe C, Scally A. Inference of Candidate Germline Mutator Loci in Humans from Genome-Wide Haplotype Data. *PLoS Genetics* 2017 Jan 17;13(1):e1006549

Yang H, **Seoighe** C. Impact of the Choice of Normalization Method on Molecular Cancer Class Discovery Using Nonnegative Matrix Factorization. *PLoS One*. 2016 Oct 14;11(10):e0164880.

Thus, the two third party institution participants, [redacted] (UCC) and [redacted] (NUIG) will act in an advisory capacity only and will have no direct responsibility for the activities, personnel, or other direct costs.

Overview of the budget

Third party	PM	Personnel	Other Direct Costs	Activity
NUI Galway	[redacted]	[redacted] €	[redacted] €	[redacted] (NUIG) will contribute to Tasks 2.1, 2.2 and 2.3, in his capacity as training coordinator of the ELIXIR Ireland node.
UCC	[redacted]	0 €	0 €	[redacted] (UCC) will advise the ELIXIR-Ireland CONVERGE grouping at UCD, particularly the data manager to be employed under CONVERGE, on aspects of Task 2 (Tasks 2.1, 2.2 and 2.3).

Participant 20 - Consiglio Nazionale delle Ricerche. CNR (ELIXIR-IT)

Does the participant plan to subcontract certain tasks (please note that core tasks of the project should not be sub-contracted)	NO

If yes, please describe and justify the tasks to be subcontracted	
*Does the participant envisage that part of its work is performed by linked third parties**	YES

If yes, please describe the third party, the link of the participant to the third party, and describe and justify the foreseen tasks to be performed by the third party:

Organizations

Università degli Studi di Bari "A. Moro" (UniBA) is one of the most prestigious and largest universities in Southern Italy, with about 60,000 students, and is partner of ELIXIR-Italy. UniBA is linked to CNR both via a Joint Research Unit (JRU) representing the Italian ELIXIR node as well as by hosting a section of CNR-IBIOM, the Institute of Biomembranes, Bioenergetics and Molecular Biotechnologies. The group of Bioinformatics and Comparative Genomics of the Department of Biosciences, Biotechnologies and Biopharmaceutics (DBBB) has leading experience in the fields of molecular biology, genomics and bioinformatics. In addition, it has a long standing experience in producing omics data using the most advanced technologies for deep sequencing of nucleic acids and develops bioinformatics resources for life science data analysis. For its activities, the group uses state-of-the-art equipment such as Illumina and Oxford Nanopore sequencing platforms and computational resources for high-performance parallel computing.

Alma Mater Studiorum - Università di Bologna (UniBO) is a university with ca. 86,000 students located in Bologna (Italy) and is a partner of ELIXIR-Italy. UniBO is linked to CNR via a Joint Research Unit (JRU) representing the Italian ELIXIR Node. [redacted] is the local representative in the JRU. His group develops and maintains several resources for the analysis of biological data, in particular protein sequences, structures and variants also in relation to human diseases.

Università degli Studi di Milano (UMIL) is a public teaching and research university, which –with about 65.000 students, 2.400 professors, 135 degree courses, 20 PhD schools and 66 specialization schools– is distinguished by its wide variety of disciplinary fields. It is the only Italian university of LERU, the League of European Research Universities, an association of 21 among the most prominent european research universities. A leading institute in Italy and Europe for scientific productivity, the University of Milan is the largest university in the region and is also an essential resource for the socio-economic context of which it is part. The Department of Bioscience, to which the research unit belongs, has been selected as an Excellence Department by MIUR in 2018 with an ISPD (Indicatore Standardizzato Performance Dipartimentale) of 100/100. The Department has a long and outstanding tradition in developing bioinformatics tools and in providing state-of-the-art bioinformatics services and drives the University's efforts in these fields.

Università degli Studi di Padova (UniPD) is a university with ca. 60,000 students located in Padua (Italy) and is a partner of ELIXIR-Italy. UniPD is linked to CNR both via a Joint Research Unit (JRU) representing the Italian ELIXIR Node as well as by hosting the Padua section of CNR-IN, the Institute of Neuroscience. [redacted] is Deputy Head of Node for ELIXIR-Italy as well as the local representative in the JRU. His group hosts molecular databases, especially for structural bioinformatics, and resources for data management and will therefore be responsible for carrying out the work assigned to UniPD.

Main tasks

UNIBA will contribute to the following tasks in WP5 (Demonstrators):

- T5.2 in WP5 Demonstrator Projects. UNIBA will develop a reference portal for cataloguing and investigating main epitranscriptome modifications. It will integrate and connect existing resources, standardizing computational protocols for storing and detecting RNA modifications. In this context, UNIBA maintains the largest collection of RNA editing events in human and non-human organisms and has developed ad hoc tools for investigating RNA editing and other modifications in deep transcriptome datasets.

UNIBO will contribute as participant to the task T3.4 (Best practices and training) in WP3 (Common Data Management Toolkit). In particular, it will contribute to the dissemination of the Common Data Management Toolkit among the Italian research community in close collaboration with WP2 and the **ELIXIR-IT** training platform.

UMIL will contribute to the dissemination of the Common Data Management Toolkit (Task 4) among the Italian research community in close collaboration with WP2 and the **ELIXIR-IT** training platform.

UNIPD will contribute to the following tasks in WP1 (Expert network):

- T1.1 Network of data managers and scientific best practice (Participant).
- T1.3 Business model (Task Leader).
- T1.4 Sustainable and scalable operating model for harmonised data management in European projects (Participant).

Prof. Silvio Tosatto and his team have been involved in setting up the WP1 Expert network. In particular, as mentioned in the sealed proposal, he will be the co-lead of WP1 and task leader for T1.3.

STAFF INVOLVED

UNIBA: [redacted]

[redacted]

UNIBO: [redacted]

[redacted]

[redacted]

UMIL: [redacted]

[redacted]

UNIPD: [redacted]

[redacted]

List of publications (top 5)

UNIBA

1. Picardi, D'Erchia, Lo Giudice, Pesole. REDIportal: a comprehensive database of A-to-I RNA editing events in humans. NAR 2013, 45 (D1), D750-D757
2. Picardi, Pesole. REDItools: high-throughput RNA editing detection made easy. Bioinformatics 2013, 29 (14), 1813-1814

UNIBO

1. Babbi G, Martelli PL, Casadio R. PhenPath: a tool for characterizing biological functions underlying different phenotypes. BMC Genomics. 2019 Jul 16;20(Suppl 8):548. doi: 10.1186/s12864-019-5868-x. PubMed PMID: 31307376
2. Savojardo C, Bruciaferri N, Tartari G, Martelli PL, Casadio R. DeepMito: accurate prediction of protein submitochondrial localization using convolutional neural networks. Bioinformatics. 2019 Jun 20. pii: btz512. doi: 10.1093/bioinformatics/btz512. PubMed PMID: 31218353.
3. Savojardo C, Martelli PL, Fariselli P, Profiti G, Casadio R. BUSCA: an integrative web server to predict subcellular localization of proteins. Nucleic Acids Res. 2018 Jul 2;46(W1):W459-W466. doi: 10.1093/nar/gky320. PubMed PMID: 29718411
4. Babbi G, Martelli PL, Profiti G, Bovo S, Savojardo C, Casadio R. eDGAR: a database of Disease-Gene Associations with annotated Relationships among genes. BMC Genomics. 2017 Aug 11;18(Suppl 5):554. doi: 10.1186/s12864-017-3911-3. PubMed PMID: 28812536.
5. Profiti G, Martelli PL, Casadio R. The Bologna Annotation Resource (BAR 3.0): improving protein functional annotation. Nucleic Acids Res. 2017 Jul 3;45(W1):W285-W290. doi: 10.1093/nar/gkx330. PubMed PMID: 28453653.

UMIL

1. Ison, Jon, Ienasescu, Hans, Chmura, Piotr, Rydza, Emil, Ménager, Hervé, Kalaš, Matúš, Schwämmle, Veit, Grüning, Björn, Beard, Niall, Lopez, Rodrigo, Duvaud, Severine, Stockinger, Heinz, Persson, Bengt, Vařeková, Radka Svobodová, Raček, Tomáš, Vondrášek, Jiří, Peterson, Hedi, Salumets, Ahto, Jonassen, Inge, Hooft, Rob, Nyrönen, Tommi, Valencia, Alfonso, Capella, Salvador, Gelpi, Josep, Zambelli, Federico, Savakis, Babis, Leskošek, Brane, Rapacki, Kristoffer, Blanchet, Christophe, Jimenez, Rafael, Oliveira, Arlindo, Vriend, Gert, Collin, Olivier, van Helden, Jacques, Longreen, Peter, Brunak, Soren (2019). The bio.tools registry of software tools and data resources for the life sciences. GENOME BIOLOGY, vol. 20, p. 1-4, ISSN: 1474760X, doi: 10.1186/s13059-019-1772-6
2. Salomoni, D., Campos, I., Gaido, L., de Lucas, J. Marco, Solagna, P., Gomes, J., Matyska, L., Fuhrman, P., Hardt, M., Donvito, G., Dutka, L., Plociennik, M., Barbera, R., Blanquer, I., Ceccanti, A., Cetinic, E., David, M., Duma, C., López-Garcia, A., Moltó, G., Orviz, P., Sustr, Z., Viljoen, M., Aguilar, F., Alves, L., Antonacci, M., Antonelli, L. A., Bagnasco, S., Bonvin, A. M. J. J., Bruno, R., Chen, Y., Costa, A., Davidovic, D., Ertl, B., Fargetta, M., Fiore, S., Gallozzi, S., Kurkcuoglu, Z., Lloret, L., Martins, J., Nuzzo, A., Nassisi, P., Palazzo, C., Pina, J., Sciacca, E., Spiga, D., Tangaro, M., Urbaniak, M., Vallero, S., Wegh, B., Zaccolo, V., Zambelli, F., Zok, T. (2018). INDIGO-DataCloud: a Platform to Facilitate Seamless Access to E-Infrastructures. JOURNAL OF GRID COMPUTING, vol. 16, p. 381-408, ISSN: 1570-7873, doi: 10.1007/s10723-018-9453-3
3. F. Zambelli, F. Mastropasqua, E. Picardi, A. M. D'Erchia, G. Pesole, G. Pavesi (2018). RNentropy: an entropy-based tool for the detection of significant variation of gene expression across multiple RNA-Seq experiments. NUCLEIC ACIDS RESEARCH, ISSN: 0305-1048, doi: 10.1093/nar/gky055
4. F. Zambelli, G. Pesole, G. Pavesi (2014). Using Weeder, Pscan, and PscanChIP for the Discovery of Enriched Transcription Factor Binding Site Motifs in Nucleotide Sequences. CURRENT PROTOCOLS IN BIOINFORMATICS, vol. 47, p. 1-31, ISSN: 1934-3396, doi: 10.1002/0471250953.bi0211s47
5. J. Ison, K. Rapacki, H. Ménager, M. Kalaš, E. Rydza, P. Chmura, C. Anthon, N. Beard, K. Berka, D. Bolser, T. Booth, A. Bretaudeau, J. Brezovsky, R. Casadio, G. Cesareni, F. Coppens, M.

Cornell, G. Cuccuru, K. Davidsen, G. Della Vedova, T. Dogan, O. Doppelt-Azeroual, L. Emery, E. Gasteiger, T. Gatter, T. Goldberg, M. Grosjean, B. Grüning, M. Helmer-Citterich, H. Ienasescu, V. Ioannidis, M.C. Jespersen, R. Jimenez, N. Juty, P. Juvan, M. Koch, C. Laibe, J. Li, L. Licata, F. Mareuil, I. Mičetić, R.M. Friborg, S. Moretti, C. Morris, S. Möller, A. Nenadic, H. Peterson, G. Profiti, P. Rice, P. Romano, P. Roncaglia, R. Saidi, A. Schafferhans, V. Schwämmle, C. Smith, M.M. Sperotto, H. Stockinger, R.S. Vařeková, S.C.E. Tosatto, V. De La Torre, P. Uva, A. Via, G. Yachdav, F. Zambelli, G. Vriend, B. Rost, H. Parkinson, P. Løngreen, S. Brunak (2016). Tools and data services registry: a community effort to document bioinformatics resources. NUCLEIC ACIDS RESEARCH, vol. 44, p. D38-D47, ISSN: 0305-1048, doi: 10.1093/nar/gkv1116

UNIPD:

1. InterPro in 2019: Improving coverage, classification and access to protein sequence annotations. Mitchell, A.L., ..., Tosatto, S.C.E., Yong, S.-Y., Finn, R.D. (2019) Nucleic Acids Research, 47 (D1), pp. D351-D360.
2. The Pfam protein families database in 2019. El-Gebali, S., ..., Tosatto, S.C.E., Finn, R.D. (2019) Nucleic Acids Research, 47 (D1), pp. D427-D432.
3. MobiDB 3.0: More annotations for intrinsic disorder, conformational diversity and interactions in proteins. Piovesan, D., ..., Tosatto, S.C.E. (2018) Nucleic Acids Research, 46 (D1), pp. D471-D476.
4. DisProt 7.0: A major update of the database of disordered proteins. Piovesan, D., ..., Tosatto, S.C.E. (2017) Nucleic Acids Research, 45 (D1), pp. D219-D227.
5. Simultaneous quantification of protein order and disorder. Sormanni, P., ..., Tosatto, S.C.E., Vendruscolo, M. (2017) Nature Chemical Biology, 13 (4), pp. 339-342.

Overview of the budget

Third party	PM	Personnel	Other Direct Costs	Activity
UniBa	5.5	35,824 €	2,250 €	Participant in T5.1 and T5.2
UniBo	1	6,513 €	409 €	Participant in T3.4
UMIL	1	6,513 €	409 €	Participant in T3.4
UniPD	9	58,621 €	3,682 €	WP1 co-lead, T1.3 lead, participant in T1.1 and T1.4

*Does the participant envisage the use of contributions in kind provided by third parties (**Articles 11 and 12** of the Model Grant Agreement)*	NO

Participant 23: Universitetet i Bergen. UiB (ELIXIR-NO)

Does the participant plan to subcontract certain tasks (please note that core tasks of the project should not be sub-contracted)?	NO
If yes, please describe and justify the tasks to be subcontracted	
Does the participant envisage that part of its work is performed by linked third parties[61]?	**YES**

The third parties involved are University of Tromsø (UiT), University of Oslo (UiO), Norwegian University of Science and Technology (NTNU) and Norwegian University of Life Sciences (NMBU)

- **University of Tromsø (UiT)**

Organization Profile

UiT - The Arctic University of Norway supports the development of outstanding research, both basic and applied, in all disciplines, but with particular emphasis on cross- and interdisciplinary research efforts focusing on the needs, problems and opportunities of the North. Particular focus is placed on biomarine studies, biomedicine and biotechnology, health and welfare studies, including telemedicine and social medicine. Arctic governance, climate and environmental research in the Arctic and sub-Arctic region in the light of climate change are key areas of interest. Indigenous and ethnic studies, especially on the Sami language and identity, are fields of national responsibility. UiT is a motor in the implementation of the Government's Strategy for the High North. Thus, northern/Arctic studies and research is our privilege and commitment.

Link with UiB

UiT is hosting one of the five ELIXIR Norway nodes which are coordinated by UiB. UiT is a partner, together with UiB, in the ELIXIR-EXCELERATE project coordinated by ELIXIR.

Contribution in the project

UiT will contribute mainly to WP3 (Common Data Management Toolkit) and WP5 (Demonstrator Projects). Based upon the competence and expertise related to management of marine genomic and metagenomic data our main focus will be on the on exploring suitable tools, developing best practises and and implementing data management plans for the marine metagenomics community.

People involved with a short description of their experience (up to 5)

[redacted]

5) ***Relevant Publications (up to 5)***

- Le Donjet T, De Santi C, Klemetsen T, Hjerde E, Willassen NP, Haugen P (2019) Closely-related Photobacterium strains comprise the majority of bacteria in the gut of migrating Atlantic cod (Gadus morhua). Microbiome 17;7(1):64. http://doi.org/10.1186/s40168-019-0681-y.

- Klemetsen T, Raknes IA, Fu J, Agafonov A, Balasundaram SV, Tartari G, Robertsen E, Willassen NP (2018) The MAR databases: development and implementation of databases specific for marine metagenomics. Nucleic Acids Res., 46(D1):D692-D699. http://doi.org/10.1093/nar/gkx1036.
- Tekle, K. M., Gundersen, S., Klepper, K., Bongo, L. A., Raknes, I. A., Li, X, Samy, J.K.A., Zhang, W., Andreetta C., Mulugeta T.D., Kalaš, M., Rye, M.B., Hjerde, E., Fornous, G., Abdulrahman Azab, A., Våge, D.I., Hovig, E., Willasssen, N.P., Drablos, F., Nygård, S., Petersen, K., Jonassen (2018). Norwegian e-Infrastructure for Life Sciences (NeLS). *F1000Research*, 7, ELIXIR–968. http://doi.org/10.12688/f1000research.15119.1
- ten Hoopen, P., Finn, R.D., Bongo, L.A., Corre, E., Fosso, B., Meyer, F., Mitchell, A., Pelletier, E., Pesole, G., Santamaria, M., Willassen, N.P., Cochrane, C. (2017) The metagenomic data life-cycle: standards and best practices, *GigaScience*, Volume 6, Issue 8, https://doi.org/10.1093/gigascience/gix047
- Robertsen EM, Denise H, Mitchell A, Finn RD, Bongo LA, Willassen NP. (2017) ELIXIR pilot action: Marine metagenomics - towards a domain specific set of sustainable services. F1000Res.;6. pii: ELIXIR-70. http://doi.org/10.12688/f1000research.10443.1.

Relevant projects (up to 5)

H2020 ELIXIR-EXCELERATE (2015-19): EXCELERATE drove the development and implementation of ELIXIR Platforms and Communities.

H2020 EMBRIC (2015-2019): a H2020 RI cluster project designed to propose integrated multidisciplinary value chains of services for the exploration of marine bioresources.

ELIXIR.NO (2013-2017) and ELIXIR2 (2017-2021): Developing, implementation and operation of ELIXIR Norway

BioMedData (2019-2023): an infrastructure for data sharing and management.

Significant infrastructures (up to 5)

UiT has access to local, national and international infrastructures and resources necessary for execution of the project.

- **University of Oslo (UiO)**

Organization profile

The University of Oslo is Norway's oldest institution for research and higher education, with 28,000 students and 6,000 employees. The university celebrated its 200th anniversary in 2011. UiO has 8 faculties, 2 museums and several centers. In addition, it has 10 Norwegian Centres of Excellence. UiO has 5 Nobel prize laureates, and is ranked as the world's 62nd university. The Department of Informatics (IFI) was established in 1977. The department has a research environment with 13 research groups, and a wide range of courses and study programmes, with 800 undergraduates, 450 master's students and 220 PhD students. The Department hosts the University Centre for bioinformatics, that hosts the Oslo ELIXIR node.

Link with UiB

UiO is hosting one of the five ELIXIR Norway nodes which are coordinated by UiB. UiO is a partner, together with UiB, in the ELIXIR-EXCELERATE project coordinated by ELIXIR.

Contribution in the project

UiO is central in coordinating activity towards human data together with UiB in WPs 1,2 and 3, and also involved in the demonstrator on GDPR in WP5. UiO will also contribute to Tasks 7.2 and 7.4 in WP7.

People involved with a short description of their experience (up to 5)

[redacted]

Relevant Publications (up to 5)

1: Njølstad PR, Andreassen OA, et al. Roadmap for a precision-medicine initiative in the Nordic region. Nat Genet. 2019 Jun;51(6):924-930.

2: Tekle KM, Gundersen S, et al. Norwegian e-Infrastructure for Life Sciences (NeLS). F1000Res. 2018 Jun 29;7.

3: Simovski B, Vodák D, et al.. GSuite HyperBrowser: integrative analysis of dataset collections across the genome and epigenome. Gigascience. 2017 Jul 1;6(7):1-12. doi: 10.1093/gigascience/gix032.

4: Dominguez-Valentin M, Nakken S, Tubeuf H, Vodak D, Ekstrøm PO, Nissen AM, Morak M, Holinski-Feder E, Martins A, Møller P, Hovig E. Potentially pathogenic germline CHEK2 c.319+2T>A among multiple early-onset cancer families. Fam Cancer. 2018 Jan;17(1):141-153.

5: Møller P, Dominguez-Valentin M, Rødland EA, Hovig E. Causes for Frequent Pathogenic BRCA1 Variants Include Low Penetrance in Fertile Ages, Recurrent De-Novo Mutations and Genetic Drift. Cancers (Basel). 2019 Jan 23;11(2). Pii: E132.

Relevant projects (up to 5)

- H2020 ELIXIR-EXCELERATE (2015-19): EXCELERATE contributed to the development and implementation of distributed computing on human sensitive genomic data (European Genome Phenome Archive (EGA).
- ELIXIR.NO (2013-2017) and ELIXIR2 (2017-2021): Developing, implementation and operation of ELIXIR Norway
- BioMedData (2019-2023): an infrastructure for data sharing and management.
- 1Million Genomes project within the EU. Member of Work group 4, Good genomic practice.

Significant infrastructures (up to 5)

UiO has access to local, national and international infrastructures and resources necessary for execution of the project (including HPC and cloud systems), and specifically to the national infrastructure on human sensitive data, located at UiO, in collaboration with national resources.

- **Norwegian University of Science and Technology (NTNU)**

Organization profile

The Norwegian University of Science and Technology (NTNU) is Norway's largest University with more than 40,000 students. NTNU awards more than 350 PhD degrees yearly within the fields of technology, science, arts and humanities, social sciences and medicine. NTNU encompasses 9 faculties and 55 departments. The annual budget of NTNU is around €820 million. NTNU is a major participant in Horizon 2020, with more than 150 H2020 projects so far. NTNU also took part in more than 130 FP7 projects. Participation in H2020 is a major priority for NTNU; therefore, the university has developed, over many years, a professional administrative project support team consisting of financial, legal and administrative advisors, to ensure the smooth implementation of H2020 projects.

The Faculty of Medicine and Health Sciences (MH) was established 40 years ago and is a centre for both outstanding research and education. It is responsible for one of the world's largest health studies (HUNT) and is home to Nobel Prize winning scientists [redacted] and [redacted]. MH consists of eight departments conducting research on a broad range of disciplines in the biomedical sciences (neuroscience, immunology, public health, medical imaging, cancer, to name a few). The faculty is also currently hosting four outstanding research centres that were established through highly competitive national grants (within neuroscience, ultrasound technology, immunology and genetic epidemiology). Researchers at MH benefit from state-of-the-art infrastructure comprising 13 specialist labs or core facilities run by dedicated and highly trained staff. Together with St Olavs Hospital, MH is responsible for the largest medical school in Norway. This close collaboration also ensures strong ties between basic research and patient-related research.

Link with UiB

NTNU is hosting one of the five ELIXIR Norway nodes which are coordinated by UiB. This includes a shared responsibility for helpdesk activities within bioinformatics, and development and support of relevant tools, pipelines and general infrastructure for bioinformatics within Norway.

Contribution in the project

In the ELIXIR-CONVERGE project NTNU will contribute with competence on data management plans and training (WP1 and WP2), based on NTNUs work within the national infrastructure for data management, BioMedData. NTNU will also contribute with competence and contributions on toolkits for data management (WP3), based on our competence in Galaxy and the implementation of solutions for data processing within NeLS, the Norwegian e-infrastructure for life sciences. NTNU will contribute with competence on relevant tools for processing of data from epitranscriptomics within the demonstrator project on epitranscriptomics (WP5), based on our work on data processing pipelines within ELIXIR Norway helpdesk activities.

People involved with a short description of their experience (up to 5)

[redacted]

[redacted]

Relevant Publications (up to 5)

Jantsch, M.F., Quattrone, A., O'Connell, M., Helm, M., Frye, M., Macias-Gonzales, M., Ohman, M., Ameres, S., Willems, L., Fuks, F., Oulas, A., Vanacova, S., Nielsen, H., Bousquet-Antonelli, C., Motorin, Y., Roignant, J.Y., Balatsos, N., Dinnyes, A., Baranov, P., Kelly, V., Lamm, A., Rechavi, G., Pelizzola, M., Liepins, J., Holodnuka Kholodnyuk, I., Zammit, V., Ayers, D., Drablos, F., Dahl, J.A., Bujnicki, J., Jeronimo, C., Almeida, R., Neagu, M., Costache, M., Bankovic, J., Banovic, B., Kyselovic, J., Valor, L.M., Selbert, S., Pir, P., Demircan, T., Cowling, V., Schafer, M., Rossmanith, W., Lafontaine, D., David, A., Carre, C., Lyko, F., Schaffrath, R., Schwartz, S., Verdel, A., Klungland, A., Purta, E., Timotijevic, G., Cardona, F., Davalos, A., Ballana, E., D, O.C., Ule, J., and Fray, R. 2018. Positioning Europe for the EPITRANSCRIPTOMICS challenge. RNA biology 15(6): 829-831. doi:10.1080/15476286.2018.1460996.

Lizio, M., Harshbarger, J., Abugessaisa, I., Noguchi, S., Kondo, A., Severin, J., Mungall, C., Arenillas, D., Mathelier, A., Medvedeva, Y.A., Lennartsson, A., Drablos, F., Ramilowski, J.A., Rackham, O., Gough, J., Andersson, R., Sandelin, A., Ienasescu, H., Ono, H., Bono, H., Hayashizaki, Y., Carninci, P., Forrest, A.R., Kasukawa, T., and Kawaji, H. 2017. Update of the FANTOM web resource: high resolution transcriptome of diverse cell types in mammals. Nucleic acids research 45(D1): D737-d743. doi:10.1093/nar/gkw995.

Medvedeva, Y.A., Lennartsson, A., Ehsani, R., Kulakovskiy, I.V., Vorontsov, I.E., Panahandeh, P., Khimulya, G., Kasukawa, T., and Drablos, F. 2015. EpiFactors: a comprehensive database of human epigenetic factors and complexes. Database : the journal of biological databases and curation 2015: bav067. doi:10.1093/database/bav067.

Razick, S., Mocnik, R., Thomas, L.F., Ryeng, E., Drablos, F., and Saetrom, P. 2014. The eGenVar data management system--cataloguing and sharing sensitive data and metadata for the life sciences. Database : the journal of biological databases and curation 2014: bau027. doi:10.1093/database/bau027.

Tekle, K.M., Gundersen, S., Klepper, K., Bongo, L.A., Raknes, I.A., Li, X., Zhang, W., Andreetta, C., Mulugeta, T.D., Kalas, M., Rye, M.B., Hjerde, E., Antony Samy, J.K., Fornous, G., Azab, A., Vage, D.I., Hovig, E., Willassen, N.P., Drablos, F., Nygard, S., Petersen, K., and Jonassen, I. 2018. Norwegian e-Infrastructure for Life Sciences (NeLS). F1000Research 7. doi:10.12688/f1000research.15119.1.

Relevant projects (up to 5)

H2020 ELIXIR-EXCELERATE (2015-19): EXCELERATE drove the development and implementation of

ELIXIR Platforms and Communities.

ELIXIR.NO (2013-2017) and ELIXIR2 (2017-2021): Developing, implementation and operation of ELIXIR Norway

BioMedData (2019-2023): an infrastructure for data sharing and management.

Significant infrastructures (up to 5)

NTNU has access to local, national and international infrastructures and resources necessary for execution of the project.

- **Norwegian University of Life Sciences (NMBU)**

Organization profile

The Norwegian University of Life Sciences (NMBU) offers studies and conduct research that meets the big, global challenges regarding the environment, sustainable development, human and animal health, renewable energy sources, food production, and land- and resource management. The university has around 5200 students enrolled in 64 different programs. Participation in the European Research Area is an important strategic priority for NMBU. NMBU participates in a dozen research projects with funding from Horizon 2020. The University is also active in ERA-net, COST, Erasmus + and other European programs. NMBU is ranked 15th on the QS University Ranking's ranking of the world's leading universities in agriculture and forestry.

Link with UiB

NMBU is hosting one of the five ELIXIR Norway nodes which are coordinated by UiB. The NMBU – node has a particular responsibility to develop, maintain and host SalmoBase.org, an integrative genomic data resource for salmonids.

Contribution in the project

NMBU will contribute to WP1 (Expert network) and WP2 (Training and Capacity Building).

People involved with a short description of their experience (up to 5)

[redacted]

Relevant Publications (up to 5)

Helgeland H, Sandve SR, Torgersen JS, Halle MK, Sundvold H, Omholt S, Våge DI (2014) The evolution and functional divergence of the beta-carotene oxygenase gene family in teleost fish-Exemplified by Atlantic salmon. Gene, doi: 10.1016/j.gene.2014.02.042.

Lien S, Koop BF, Sandve SR, et al., (2016) The Atlantic salmon genome provides insights into rediploidization. Nature. doi:10.1038/nature17164.

Samy JKA, Mulugeta TD, Nome T, Sandve SR, Grammes F, Kent MP, Lien S, Våge DI (2017). SalmoBase: an integrated molecular data resource for Salmonid species. BMC Genomics. 6;18(1):482. doi: 10.1186/s12864-017-3877-1.

Tekle KM, GundersenS, Klepper K, et al., (2018) Norwegian e-Infrastructure for Life Sciences (NeLS). F1000Res, 29;7. doi: 10.12688/f1000research.15119.1

Mulugeta TD, Nome T, To TH, Gundappa MK, Macqueen DJ, Våge DI, Sandve SR, Hvidsten TR (2019). SalMotifDB: a tool for analyzing putative transcription factor binding sites in salmonid genomes. BMC Genomics, 2;20(1):694. doi: 10.1186/s12864-019-6051-0.

Relevant projects (up to 5)

ELIXIR.NO (2013-2017) – a Norwegian ELIXIR Node

ELIXIR2 (2017-2021) - a distributed infrastructure for the next generation of life science

BioMedData (2019-2023): an infrastructure for data sharing and management.

Significant infrastructures (up to 5)

NMBU has access to local, national and international infrastructures and resources necessary for execution of the project.

PMs and Budget distribution

In the frame of this proposal, the UiB has 27 PMs assigned, with 207 981€ for personal costs and 11 250€ as Other Direct Costs. PMs and budget will be re-distributed among the third parties as follows:

Third party against payment	PM	Personnel	Other Direct Costs	Activity
UiT	7	53 921 €	2 864 €	All four LTPs will contribute to WP1 and WP2. UiT, UiO and NTNU will in addition contribute to WP3 and WP5. UiT will lead the marine demonstrator in WP5
UiO	9	69 327€	3886€	
NTNU	3	23 109 €	1 227 €	

NMBU	1	7 703 €	409 €	

Does the participant envisage the use of contributions in kind provided by third parties (Articles 11 and 12 of the General Model Grant Agreement)?	NO
If yes, please describe the third party and their contributions	

Participant 24: Instituto de Engenharia de Sistemas e Computadores, Investigação e Desenvolvimento. INESC-ID (ELIXIR-PT)

Does the participant plan to subcontract certain tasks (please note that core tasks of the project should not be sub-contracted)?	NO
If yes, please describe and justify the tasks to be subcontracted	
Does the participant envisage that part of its work is performed by linked third parties[65]?	NO
If yes, please describe the third party, the link of the participant to the third party, and describe and justify the tasks to be performed by the third party and it's budget breakdown	
Does the participant envisage the use of contributions in kind provided by third parties (Articles 11 and 12 of the General Model Grant Agreement)?	**YES**

If yes, please describe the third party and their contributions

The Third Party involved is Instituto Superior Técnico (IST).

INESC-ID is a leading Portuguese Research and Technology Organization, doing research in several areas. Almost all the research staff from INESC-ID come from Portuguese Universities. These researchers work at the INESC-ID premises and all the research work is done on behalf of INESC-ID. The Portuguese University IST, Instituto Superior Técnico, is the main University who ceded, free of charge, in the scope of the Art.12 of the AGA, some professors to INESC-ID, as a proportion of their full time allocation, on the basis of a prior agreement. Under this agreement, IST makes available some human resources to INESC-ID, so that they may participate in the research and development of the latter, according to its own needs and organization. Since the prior agreement between IST and INESC-ID leaves the assignment of the resources to the discretion of the management of INESC-ID, they should not be considered as made available to participate only in this specific project. Under this arrangement, INESC-ID recognizes IST as a Third Party for some of the H2020 Projects in which INESC-ID participates, on a free of charge basis, in the scope of the Article 12 of the Annotated Model Grant Agreement.

The relation between INESC-ID and the University IST) was made explicit in its bylaws: "INESC-ID, to pursue its objectives, will sign specific protocol with the Associate, in order to assure that the human and material resources will be provided by these Associates" (article 23, n. 1). This protocol is effectively signed and is still in use today. For all contractual purposes, INESC-ID assumes full responsibility for the involvement of these researchers in the project, since they are members of its research team.

Staff involved

[redacted]

PMs and Budget distribution

This in-kind contribution in the form of €79,805 for direct personnel costs corresponds to 10.5 PMs.

Third party free of charge	PM	Personnel	Other Direct Costs	Activity
IST	10.5	79,805 €	0 €	IST will contribute human resources according to a prior agreement with INESC-ID

Participant 26: Uppsala University. UU (ELIXIR-SE)

Does the participant plan to subcontract certain tasks (please note that core tasks of the project should not be subcontracted)	NO
If yes, please describe and justify the tasks to be subcontracted	
Does the participant envisage that part of its work is performed by linked third parties	**YES**
• **Stockholm University (SU)**	

Organization profile

Since 1878 Stockholm University has been characterised by openness and innovation. A modern university with a multicultural environment, Stockholm University is one of the world's top 100 higher education institutes. 70,000 students, 1,800 doctoral students and 5,000 staff are active within the areas of science and human science.

Education and research at Stockholm University make a difference. The University contributes to individual and social change through top quality education and outstanding research. Our researchers contribute to the development of public policy and political decision making, and participate in Nobel Prize Committees and international expert bodies.

Link with UU

The National Bioinformatics Infrastructure Sweden (NBIS) constitutes the Swedish Elixir node and has UU as host university. NBIS is a consortium with 11 partner universities. In this project, the contributions will be made from staff at Uppsala University (UU) and staff at Stockholm University (SU).

Contribution in the project

SU staff is engaged in data management (WP1), training (WP2) and systems development (WP3).

People involved with a short description of their experience (up to 5)

[redacted]

Relevant Publications (up to 5)

1. Ameur A, Dahlberg J, Olason P, Vezzi F, Karlsson R, Martin M, Viklund J, Kähäri AK, Lundin P, Che H, Thutkawkorapin J, Eisfeldt J, Lampa S, Dahlberg M, Hagberg J, Jareborg N, Liljedahl U, Jonasson I, Johansson Å, Feuk L, Lundeberg J, Syvänen AC, Lundin S, Nilsson D, Nystedt B, Magnusson PK,

Gyllensten U. (2017) SweGen: a whole-genome data resource of genetic variability in a cross-section of the Swedish population. Eur J Hum Genet. 25:1253--1260.
2. Jiménez, R. C., Kuzak, M., ... Borg, M., ... Hagberg, J., ... Crouch, S. (2017) Four simple recommendations to encourage best practices in research software. F1000Research 6, 876.

Relevant projects (up to 5)

1. EU H2020 (H2020-INFRADEV-1-2015-1) ELIXIR-EXCELERATE: Fast-track ELIXIR implementation and drive early user exploitation across the life sciences, Grant number: 676559, 2015-2019.
2. Tryggve -- Nordic collaboration for sensitive data 2014--2017, grant from NordForsk, https://neic.no/tryggve/
3. Tryggve2 -- Nordic collaboration for sensitive data 2017–2020, grant from NordForsk, https://neic.no/tryggve2/

Significant infrastructures (up to 5)

1. National Bioinformatics Infrastructure Sweden (NBIS) and the Swedish node in Elixir (the European infrastructure for biological information) 2018–2020, grant from the Swedish Research Council, https://nbis.se

PMs and Budget distribution

Third party against payment	PM	Personnel	Other Direct Costs	Activity
SU	8	55 200 €	2 000 €	T1.1 Network of data managers and scientific best practice. T1.3 Business model T2.3 Capacity Building in Data Management and Stewardship. T3.1 Establish a Starter Toolkit. T3.3 Access portal to Toolkit tailored to stakeholders. T5.2 Implementation of pilots data management plans (Participant).

*Does the participant envisage the use of contributions in kind provided by third parties (**Articles 11 and 12** of the Model Grant Agreement)*	**NO**
If yes, please describe the third party and their contributions	

Section 5: Ethics and Security

5.1 Ethics

Wide sharing of knowledge and data drives the progression of science. Shared data allows other researchers to reproduce findings and benchmark quality of experiments. Sharing data so that other researchers can Find, Access and Interoperate – i.e. integrate the data with the outcomes of their own experiments - drives Reuse. It presents an opportunity to build the large aggregated cohorts we need to detect rare signals and manage the many confounding factors in Life Sciences (LS). This project will develop the guidelines and tools needed to develop proper Data Management Plans (DMPs) in LS that promote FAIR. Through worked examples using demonstrator projects that will get access to the ELIXIR Expert Network, as well as to the method process tools and training, to ensure proper DMPs are in place to ensure discovery, accessibility, interoperability and reusability of the project data. In addition, through disseminated guidelines and tailored training for data handlers (in academia, SMEs and industry), data management culture will change and be sustained with datasets reused by industry, academia and SMEs. Our SME Events will enable wide data reuse and foster an innovation ecosystem around these data that power future reuse, knowledge generation and societal benefit.
We expect ELIXIR-CONVERGE to increase the secondary use of previously collected personal data where the consent allows this and/or where the legal and ethical compliance can be ensured. Normally, the staff employed in this project will not act as the 'Data Controller/Processor', they will operate as technical experts advising the data processor on the demonstrator projects how to use ELIXIR guidance and tools to increase the FAIRness of their data. This advice will cover guidance and tools that can be used for the protection of sensitive human data as required by the General Data Protection Regulation (GDPR) as well as good ethical practice in research. It is a major aim of the project to allow FAIR sharing of such sensitive data in compliance with ethics and data protection law and therefore extend ELIXIR activities more towards translational medicine.
WP5 have selected the demonstrator projects incorporated in the proposal and will ensure that expectations are fully accomplished by interacting with demonstrator projects and the Expert Network (WP1). In WP1 the Expert Network will advise and guide the demonstrator projects in the utilisation of ELIXIR guidance, methods, training and toolkits for their particular domain. WP2 will develop the specific training for demonstrator projects, members of the Expert Network and ELIXIR Nodes to apply ELIXIR DMP guidelines, methods, processes and tools in their domains and in the different scenarios (support to individual researchers, Transnational Projects and International initiatives), with the main focus being the pan-european transnational projects. WP3 will define the common data management toolkit that will provide the guidance on how to implement a DMP as well as concrete realisation of the common data management toolkit for the demonstrator projects, tailoring it for the specific scientific domain, data sensitivity (as in Human Data), and scope (Local, European, International). WP4 will ensure that all the project outputs, as well as the success stories generated by the demonstrator projects, are accessible across ELIXIR and beyond (funders, policy makers, prospective ELIXIR Member States), promoting long term sustainability of the ELIXIR Nodes and evaluation the different Business Models that come out of this

project. WP6 will coordinate the project's activities and apply Project Management best practices during the project execution to ensure all benefits are delivered.

Finally, ELSI expertise is ensured via the ELIXIR Node expert in this area as well as via the ELIXIR SAB.

WP6 will deliver a DMP (D6.2), in collaboration with WPs 1, 3 and 5, to provide further evidence of appropriate ethics committees and competent authorities. ELIXIR SAB includes ethics experts who will be consulted on this specific deliverable.

As ELIXIR-CONVERGE will not be handling private data or highly sensitive data, most of the requirements that will apply to projects dealing with research or secondary use data are not relevant here.

5.1.1 EthSR Response

EthSR Comment 1: Protection of personal data

The proposal mainly relies on processing of previously collected data (p. 32, 53, 80-81). Special categories of data (i.e, human genomics data) are going to be processed (p. 53). The proposal also includes a number of workshops (p. 44, 48, 50) and a survey on training needs ((WP2) p. 23), which will possibly involve collection and processing of personal data. Although it is unlikely that any personal data other than contact details will be collected, participants should be provided in writing with details of what personal information will be processed and with other relevant information on processing of their personal data as required by General Data Protection Regulation (Regulation EU 2016/679).

ELIXIR-CONVERGE Response 1:

"The proposal mainly relies on processing of previously collected data (p. 32, 53, 80-81)":
The main goal of ELIXIR-CONVERGE is to enable good Data Management Practice (DMP) for all European life science research. This will be achieved by strengthening DMP operations within national ELIXIR Nodes and provide method, guidance and tools that scale from single national researchers to large, complex European consortia. Thus, the focus will be to ensure that users' independently implement DMP within their projects using European best practice. As a consequence ELIXIR-CONVERGE will not take part in the data processing that users will carry out on their projects but rather provide support and training. ELIXIR-CONVERGE partners will not be data controllers for user projects (users retain control of their own data), similarly it is the users responsibility to get the necessary ethics approvals. The effort allocated to users' projects within ELIXIR-CONVERGE will be devoted to assist (train and support them) them on the use of the DMP toolkit we will be putting forward and to gather their feedback on how it can be improved.

"Special categories of data (i.e, human genomics data) are going to be processed (p. 53)":
The ELIXIR-CONVERGE project partners will perform business-analysis and develop and execute action plans for FAIR DMP with our partners that have already collected data within the research projects. Normally this will not require a data processing agreement (i.e. the user is advised on how to best process the data and project partners will not independently process data on users behalf. There might be exceptions, which will be detailed in D6.2 Project Data Management Plan and regular updates (M3) as well as in D5.1 Categorization of the pilot projects (M12) that will describe the needs in terms of DMP of the demonstrator projects and development of a project categorization from that point of view.

Personal data will be collected by ELIXIR-CONVERGE partners for the execution of the awarded grant. This involves e.g. contact details of members of the different institutions involved in the different WP and tasks. Similarly the planned surveys will study organisational capabilities: business models (D1.2), training needs (D2.1), training assets (D2.2-2.4) and does not foresee collection of personal data except contact details for key personnel (which will likely already be in the public domain, e.g. via facility web pages). Participant data from workshops and other events will record additional personal data such as dietary requirements and gender to comply with reporting requirements. The D6.2 Data Management Plan (M3) will explain how planned data processing is relevant and limited to the purposes of the research project (in accordance with the 'data minimisation principle') and set out the responsible organisation for each of the data collections and their respective Data Protection Officers and data protection measures.

EthSR Comment 2: Third countries

One of the participants is Weizmann Institute of Science from Israel. Moreover, a number of non-EU countries (e.g. Australia, Canada and South Africa) (p. 51), have shown interest in ELIXIR, either in terms of Membership or to enjoy the many mutual benefits of concrete scientific collaborations. Legal framework does allow for non-European Union countries to become full Members and the consortium plans to establish collaboration with a number of non-EU countries, thus there is a probability that the project will involve transfers of personal data between the EU and non-EU states.

ELIXIR-CONVERGE Response 2:

The personal (contact) data that will be shared with non-EU countries to enable the communication required in order to implement the project plan. Similarly the planned surveys will study organisational capabilities: business models (D1.2), training needs (D2.1), training assets (D2.2-2.4). Other than that a collection of personal data is not foreseen except contact details for key personnel (which will likely already be in the public domain, e.g. via facility web pages). Currently, no workshops and events are planned in non-EU countries and so no transfer of additional personal data (dietary requirements, gender) is foreseen.

The Data Management Plan will detail the planned data transfers to third countries within this project.

EthSR Comment 3: Identified Ethics Issues

The project will increase data management capabilities across all ELIXIR Nodes, building on the maturity of the ELIXIR infrastructure and the resources and services provided as well as on the capabilities and strengths of the Nodes in this area to produce a blueprint on how to deliver data management at scale at the European level. It will position the Nodes as the reference when implementing transnational data management plans in life science.

ELIXIR-CONVERGE will not produce primary data, the main objective is to drive publication of existing data assets and enable reuse. It is also expected that the publication of data management and analysis tools and community outreach and training will enable the life science community at large to adopt FAIR data management.

ELIXIR-CONVERGE Response 3:

We agree with this assessment.

As a result of the project, a DMP toolkit that will include Methods, Guidance, and tools will be publically available, following the ELIXIR commitment to Open source, Open Science and Open Data. Life science projects will be able to use the toolkit on their own with none or minimal support from the ELIXIR Data Management Network. ELIXIR-CONVERGE consortium will not be involved in the data processing activities that users' project will carry out to fully implement their DMPs.

EthSR Comment 4: Ethics recommendations

The applicants must clarify what it means by "marine organisms". They refer to them but only in a general way without more details.

ELIXIR-CONVERGE Response 4:

We agree, this is ambiguous and presuppose detailed knowledge of the ELIXIR Marine community. Briefly: the third demonstrator project is aiming to define a common data management plan for the marine metagenomics community. They will consider large metagenomic datasets: whole metagenomes shotgun data, metagenome-assembled genome data, metatranscriptome, metabarcoding data. ELIXIR has supported the development of high quality reference databases for marine microorganisms (prokaryotes and eukaryotes) such as Marine Metagenomics Portal and ITSOne Db (https://elixir-europe.org/communities/marine-metagenomics). The demonstrator will utilise these databases to 1) improve the standardised assignment of organisms in marine research 2) encourage submission of newly curated reference genomes to these resources for community reuse.

Ethics Requirements

EthSR Comment 5: Protection of personal data

The beneficiary must check if special derogations pertaining to the rights of data subjects or the processing of genetic, biometric and/or health data have been established under the national legislation of the country where the research takes place and submit a declaration of compliance with respective national legal framework(s).

ELIXIR-CONVERGE Response 5:

No research data with special derogations or involving genetic, biometric and/or health data will be collected or processed by the ELIXIR-CONVERGE Consortium.

The only personal (contact) data that will be collected and shared with project participants are the contact detail required for the implementation of the project plan and for recontacting key personnel at national facilities in conjunction with surveys. Additionally gender and dietary requirements will be collected for event participants and used to generate summary (aggregate) reports on equal opportunities.

EthSR Comment 6: Protection of personal data

The host institution must confirm that it has appointed a Data Protection Officer (DPO) and the contact details of the DPO are made available to all data subjects involved in the research. For host institutions not required to appoint a DPO under the GDPR a detailed data protection policy for the project must be submitted as a deliverable.

The beneficiary must explain how all of the data they intend to process is relevant and limited to the purposes of the research project (in accordance with the 'data minimisation 'principle). This must be submitted as a deliverable.

A description of the technical and organisational measures that will be implemented to safeguard the rights and freedoms of the data subjects/research participants must be submitted as a deliverable.

A description of the security measures that will be implemented to prevent unauthorised access to personal data or the equipment used for processing must be submitted as a deliverable.

Description of the anonymisation/pseudonymisation techniques that will be implemented must be submitted as a deliverable.

In case personal data are transferred from the EU to a non-EU country or international organisation, confirmation that such transfers are in accordance with Chapter V of the General Data Protection Regulation 2016/679, must be submitted as a deliverable.

In case personal data are transferred from a non-EU country to the EU (or another third state), confirmation that such transfers comply with the laws of the country in which the data was collected must be submitted as a deliverable.

Detailed information on the informed consent procedures in regard to data processing must be submitted as a deliverable.

Templates of the informed consent forms and information sheets (in language and terms intelligible to the participants) must be kept on file.

In case of further processing of previously collected personal data, an explicit confirmation that the beneficiary has lawful basis for the data processing and that the appropriate technical and organisational measures are in place to safeguard the rights of the data subjects must be submitted as a deliverable.

The beneficiary must evaluate the ethics risks related to the data processing activities of the project. This includes also an opinion if data protection impact assessment should be conducted under art.35 General Data Protection Regulation 2016/679. The risk evaluation and the opinion must be submitted as a deliverable.

ELIXIR-CONVERGE Response 6:

"The host institution must confirm that it has appointed a Data Protection Officer (DPO)..."

"The beneficiary must explain how all of the data they intend to process is relevant and limited to the purposes of the research project (in accordance with the 'data minimisation 'principle). This must be submitted as a deliverable."

"A description of the technical and organisational measures..."

Plans for data collection, management and oversight, together with DPO contact details will be provided for those beneficiaries responsible for collecting personal data (see above) in the ELIXIR-CONVERGE Data Management Plan (Deliverable D6.2) to address all these requirements. Although the DMP deliverable is originally due on M06 the consortium will be delivering it by M03 to ensure all these requirements are timely addressed.

"Description of the anonymisation/pseudonymisation techniques that will be implemented must be submitted as a deliverable."
ELIXIR-CONVERGE DoW does not foresee any anonymisation of personal data. Survey reports will be presented as aggregate data (e.g. on geography).

"In case personal data are transferred from the EU to a non-EU country or international organisation, confirmation that such transfers are in accordance with Chapter V of the General Data Protection Regulation 2016/679, must be submitted as a deliverable.
In case personal data are transferred from a non-EU country to the EU (or another third state), confirmation that such transfers comply with the laws of the country in which the data was collected must be submitted as a deliverable."
ELIXIR-CONVERGE Data Management Plan (Deliverable D6.2) will outline the planned transfers. Transfers will be logged and reported by the end of the project. The aforementioned requested confirmation will be provided as part of the ELIXIR-CONVERGE Data Management Plan (D6.2).

"Detailed information on the informed consent procedures in regard to data processing must be submitted as a deliverable.
Templates of the informed consent forms and information sheets (in language and terms intelligible to the participants) must be kept on file."
ELIXIR-CONVERGE will not rely only on informed consent as the legal basis for data processing as per the ELIXIR ELSI Policy. Detailed information on procedures as well as the consent forms and information sheets will be provided as part of the ELIXIR-CONVERGE Data Management Plan (D6.2) and of its regular updates.

"In case of further processing of previously collected personal data, an explicit confirmation that the beneficiary has lawful basis for the data processing and that the appropriate technical and organisational measures are in place to safeguard the rights of the data subjects must be submitted as a deliverable."

Noted. In case of further processing of previously collected personal data, confirmations will be provided as part of the ELIXIR-CONVERGE Data Management Plan (D6.2) or of its regular updates.

"The beneficiary must evaluate the ethics risks related to the data processing activities of the project. This includes also an opinion if data protection impact assessment should be conducted under art.35 General Data Protection Regulation 2016/679. The risk evaluation and the opinion must be submitted as a deliverable."

Noted, this will be submitted as part of DMP (D6.2)

EthSR Comment 7: Third countries
Details on the materials which will be imported to/exported from the EU must be included in the grant agreement before signature kept on file.

ELIXIR-CONVERGE Response 7:
No material will be imported to/exported from the EU. Personal (contact) data that will be shared with non-EU countries (that are involved in ELIXIR-CONVERGE) but will be limited to the contact detail information that will enable the communications required in order to implement the project plan.
Full details on how the consortium will be dealing with the personal (contact) data will be provided as part of deliverable D6.2 (DMP). See above for details

5.2 Security

Please indicate if your project will involve:

- activities or results raising security issues: **NO.**
- 'EU-classified information' as background or results: **NO.**

Annex 1 - Additional information for WP descriptions

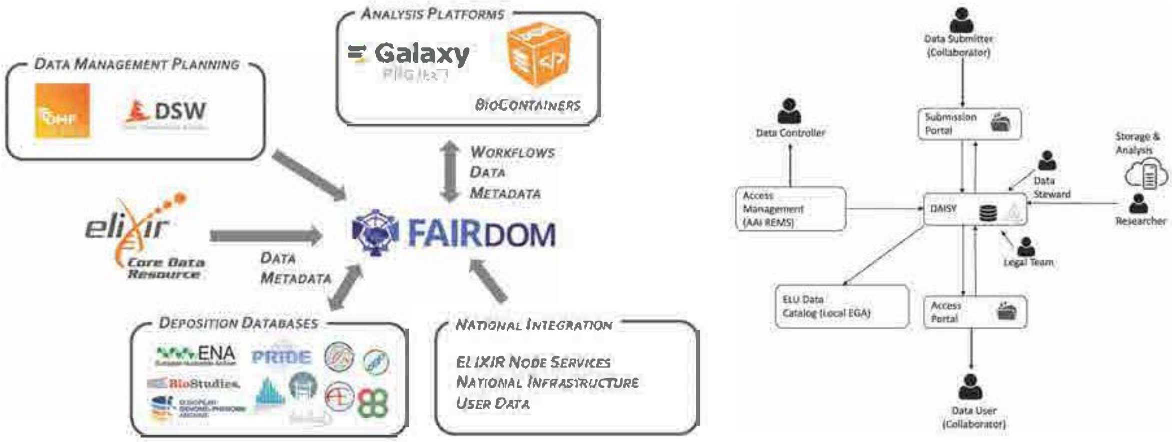

Figure WP3.1: *A Data Management Toolkit covers all aspects of DM planning and implementation: standards, tools, and services. This figure illustrates two examples of interconnected services currently in use in ELIXIR Nodes. The left panel demonstrates an omics oriented toolkit deploying FAIRDOM as hub for data management and Galaxy for data analysis (partners UNIMAN, UIB, HITS, VIB). The right panel showcases a toolkit for sensitive data, covering GDPR requirements, as implemented by UNILU. Similar setups have been implemented in other ELIXIR Nodes (e.g. partners UT, BSC, CSC, SIB, UU).*

ESTIMATED BUDGET FOR THE ACTION

	Estimated eligible[1] costs (per budget category)											EU contribution			Additional information		
	A. Direct personnel costs				B. Direct costs of subcontracting	[C. Direct costs of fin. support]	D. Other direct costs		E. Indirect costs[2]	F. Special unit costs	Total costs	Reimbursement rate %	Maximum EU contribution[3]	Maximum grant amount[4]	Information for indirect costs	Information for auditors	Other information
	A.1 Employees (or equivalent) A.2 Natural persons under direct contract A.3 Seconded persons [A.6 Personnel for providing access to research infrastructure]		A.4 SME owners without salary A.5 Beneficiaries that are natural persons without salary				D.1 Travel D.2 Equipment D.3 Other goods and services D.4 Costs of large research infrastructure	D.5 Costs of internally invoiced goods and services		F.1 Costs for providing trans-national access to research infrastructure[5]					Estimated costs of in-kind contributions not used on premises	Declaration of costs under Point D.4	Estimated costs of beneficiaries/ linked third parties not receiving funding/ international partners
Form of costs[6]	Actual	Unit[7]	Unit[8]		Actual	Actual	Actual	Unit[9]	Flat-rate[10] 25%	Unit[12]							
	a	Total b	No hours	Total c	d	e	f	Total g	h = 0,25 x (a +b+c+f+g +[i1][13]+[i2][13]-n)	Total i1	j = a+b+c +d+[e]+f+g +h+[i1]+[i2]	k	l	m	n	Yes/No	
1. ELIXIR/ EMBL-EBI	1 966 349.19	0.00	0.00	0.00	0.00	0.00	153 080.91	0.00	529 857.53	0.00	2 649 287.63	100.00	2 649 287.63	2 649 287.63	0.00	No	n/a
2. VIB	185 526.00	0.00	0.00	0.00	0.00	0.00	32 000.00	0.00	54 381.50	0.00	271 907.50	100.00	271 907.50	271 907.50	0.00	No	n/a
3. SIB	294 869.86	0.00	0.00	0.00	0.00	0.00	9 900.00	0.00	76 192.47	0.00	380 962.33	100.00	380 962.33	380 962.33	0.00	No	n/a
4. UCY	16 400.00	0.00	0.00	0.00	0.00	0.00	4 500.00	0.00	5 225.00	0.00	26 125.00	100.00	26 125.00	26 125.00	0.00	No	n/a
5. CING	16 400.00	0.00	0.00	0.00	0.00	0.00	4 500.00	0.00	5 225.00	0.00	26 125.00	100.00	26 125.00	26 125.00	0.00	No	n/a
6. UOCHB	46 410.00	0.00	0.00	0.00	0.00	0.00	9 000.00	0.00	13 852.50	0.00	69 262.50	100.00	69 262.50	69 262.50	0.00	No	n/a
- UPOL	10 920.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2 730.00	0.00	13 650.00	100.00	13 650.00	13 650.00	0.00	No	n/a
- MU	27 300.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	6 825.00	0.00	34 125.00	100.00	34 125.00	34 125.00	0.00	No	n/a
- CTU	30 030.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	7 507.50	0.00	37 537.50	100.00	37 537.50	37 537.50	0.00	No	n/a
Σ beneficiary	114 660.00	0.00			0.00	0.00	9 000.00	0.00	30 915.00	0.00	154 575.00		154 575.00	154 575.00	n/a	n/a	0.00
7. HITS	106 050.00	0.00	0.00	0.00	0.00	0.00	9 000.00	0.00	28 762.50	0.00	143 812.50	100.00	143 812.50	143 812.50	0.00	No	n/a
8.DTU	160 000.00	0.00	0.00	0.00	0.00	0.00	9 000.00	0.00	42 250.00	0.00	211 250.00	100.00	211 250.00	211 250.00	0.00	No	n/a
9. UTARTU	104 824.00	0.00	0.00	0.00	0.00	0.00	9 000.00	0.00	28 456.00	0.00	142 280.00	100.00	142 280.00	142 280.00	0.00	No	n/a
10.BSC	237 767.33	0.00	0.00	0.00	0.00	0.00	14 250.00	0.00	63 004.33	0.00	315 021.66	100.00	315 021.66	315 021.66	0.00	No	n/a
11.CSC	258 090.00	0.00	0.00	0.00	0.00	0.00	15 000.00	0.00	68 272.50	0.00	341 362.50	100.00	341 362.50	341 362.50	0.00	No	n/a
12. INRAE	47 508.00	0.00	0.00	0.00	0.00	0.00	27 500.00	0.00	18 752.00	0.00	93 760.00	100.00	93 760.00	93 760.00	0.00	No	n/a
13. CNRS	91 024.00	0.00	0.00	0.00	0.00	0.00	4 500.00	0.00	23 881.00	0.00	119 405.00	100.00	119 405.00	119 405.00	0.00	No	n/a
14. BSRCAF	48 000.00	0.00	0.00	0.00	0.00	0.00	3 900.00	0.00	12 975.00	0.00	64 875.00	100.00	64 875.00	64 875.00	0.00	No	n/a
15. ATHENA RC	24 000.00	0.00	0.00	0.00	0.00	0.00	3 000.00	0.00	6 750.00	0.00	33 750.00	100.00	33 750.00	33 750.00	0.00	No	n/a
16. CERTH	16 000.00	0.00	0.00	0.00	0.00	0.00	3 000.00	0.00	4 750.00	0.00	23 750.00	100.00	23 750.00	23 750.00	0.00	No	n/a
17. TTK	107 000.00	0.00	0.00	0.00	0.00	0.00	9 000.00	0.00	29 000.00	0.00	145 000.00	100.00	145 000.00	145 000.00	0.00	No	n/a
18. UCD	92 000.00	0.00	0.00	0.00	0.00	0.00	9 000.00	0.00	25 250.00	0.00	126 250.00	100.00	126 250.00	126 250.00	0.00	No	n/a
19. WEIZMANN	161 590.00	0.00	0.00	0.00	0.00	0.00	9 000.00	0.00	42 647.50	0.00	213 237.50	100.00	213 237.50	213 237.50	0.00	No	n/a
20. CNR	35 825.00	0.00	0.00	0.00	0.00	0.00	2 250.00	0.00	9 518.75	0.00	47 593.75	100.00	47 593.75	47 593.75	0.00	No	n/a
- UNIBA	35 824.00	0.00	0.00	0.00	0.00	0.00	2 250.00	0.00	9 518.50	0.00	47 592.50	100.00	47 592.50	47 592.50	0.00	No	n/a
- UNIBO	6 513.00	0.00	0.00	0.00	0.00	0.00	409.00	0.00	1 730.50	0.00	8 652.50	100.00	8 652.50	8 652.50	0.00	No	n/a
- UMIL	6 513.00	0.00	0.00	0.00	0.00	0.00	409.00	0.00	1 730.50	0.00	8 652.50	100.00	8 652.50	8 652.50	0.00	No	n/a
- UNIPD	58 621.00	0.00	0.00	0.00	0.00	0.00	3 682.00	0.00	15 575.75	0.00	77 878.75	100.00	77 878.75	77 878.75	0.00	No	n/a
Σ beneficiary	143 296.00	0.00			0.00	0.00	9 000.00	0.00	38 074.00	0.00	190 370.00		190 370.00	190 370.00	n/a	n/a	0.00
21. UNILU	184 704.00	0.00	0.00	0.00	0.00	0.00	10 350.00	0.00	48 763.50	0.00	243 817.50	100.00	243 817.50	243 817.50	0.00	No	n/a
22.DTL PROJECTS	234 400.36	0.00	0.00	0.00	0.00	0.00	32 900.00	0.00	66 825.09	0.00	334 125.45	100.00	334 125.45	334 125.45	0.00	No	n/a

ESTIMATED BUDGET FOR THE ACTION

	Estimated eligible[1] costs (per budget category)											EU contribution			Additional information		
	A. Direct personnel costs				B. Direct costs of subcontracting	[C. Direct costs of fin. support]	D. Other direct costs		E. Indirect costs[2]	F. Special unit costs	Total costs	Reimbursement rate %	Maximum EU contribution[3]	Maximum grant amount[4]	Information for indirect costs	Information for auditors	Other information
	A.1 Employees (or equivalent) A.2 Natural persons under direct contract A.3 Seconded persons [A.6 Personnel for providing access to research infrastructure]		A.4 SME owners without salary A.5 Beneficiaries that are natural persons without salary				D.1 Travel D.2 Equipment D.3 Other goods and services D.4 Costs of large research infrastructure	D.5 Costs of internally invoiced goods and services		F.1 Costs for providing trans-national access to research infrastructure[5]					Estimated costs of in-kind contributions not used on premises	Declaration of costs under Point D.4	Estimated costs of beneficiaries/ linked third parties not receiving funding/ international partners
Form of costs[6]	Actual	Unit[7]	Unit[8]		Actual	Actual	Actual	Unit[9]	Flat-rate[10] 25%	Unit[12]							
	a	Total b	No hours	Total c	d	e	f	Total g	h = 0,25 x (a +b+c+f+g +[i1][13]+[i2][13]-n)	Total i1	j = a+b+c +d+[e]+f+g +h+[i1]+[i2]	k	l	m	n	Yes/No	
23. UiB	53 921.00	0.00	0.00	0.00	0.00	0.00	2 864.00	0.00	14 196.25	0.00	70 981.25	100.00	70 981.25	70 981.25	0.00	No	n/a
- UiT	53 921.00	0.00	0.00	0.00	0.00	0.00	2 864.00	0.00	14 196.25	0.00	70 981.25	100.00	70 981.25	70 981.25	0.00	No	n/a
- UiO	69 327.00	0.00	0.00	0.00	0.00	0.00	3 886.00	0.00	18 303.25	0.00	91 516.25	100.00	91 516.25	91 516.25	0.00	No	n/a
- NTNU	23 109.00	0.00	0.00	0.00	0.00	0.00	1 227.00	0.00	6 084.00	0.00	30 420.00	100.00	30 420.00	30 420.00	0.00	No	n/a
- NMBU	7 703.00	0.00	0.00	0.00	0.00	0.00	409.00	0.00	2 028.00	0.00	10 140.00	100.00	10 140.00	10 140.00	0.00	No	n/a
Σ beneficiary	207 981.00	0.00			0.00	0.00	11 250.00	0.00	54 807.75	0.00	274 038.75		274 038.75	274 038.75	n/a	n/a	0.00
24. INESC ID	95 005.95	0.00	0.00	0.00	0.00	0.00	5 400.00	0.00	25 101.49	0.00	125 507.44	100.00	125 507.44	125 507.44	0.00	No	n/a
25.FCG-IGC	94 433.33	0.00	0.00	0.00	0.00	0.00	5 400.00	0.00	24 958.33	0.00	124 791.66	100.00	124 791.66	124 791.66	0.00	No	n/a
26. UU	234 600.00	0.00	0.00	0.00	0.00	0.00	33 700.00	0.00	67 075.00	0.00	335 375.00	100.00	335 375.00	335 375.00	0.00	No	n/a
- SU	55 200.00	0.00	0.00	0.00	0.00	0.00	2 000.00	0.00	14 300.00	0.00	71 500.00	100.00	71 500.00	71 500.00	0.00	No	n/a
Σ beneficiary	289 800.00	0.00			0.00	0.00	35 700.00	0.00	81 375.00	0.00	406 875.00		406 875.00	406 875.00	n/a	n/a	0.00
27. UL	83 306.06	0.00	0.00	0.00	0.00	0.00	9 900.00	0.00	23 301.52	0.00	116 507.58	100.00	116 507.58	116 507.58	0.00	No	n/a
28. UNIMAN	107 446.00	0.00	0.00	0.00	0.00	0.00	4 500.00	0.00	27 986.50	0.00	139 932.50	100.00	139 932.50	139 932.50	0.00	No	n/a
29. UCAM	70 180.00	0.00	0.00	0.00	0.00	0.00	4 500.00	0.00	18 670.00	0.00	93 350.00	100.00	93 350.00	93 350.00	0.00	No	n/a
30. CRG	190 208.00	0.00	0.00	0.00	0.00	0.00	6 750.00	0.00	49 239.50	0.00	246 197.50	100.00	246 197.50	246 197.50	0.00	No	n/a
31. UP	16 050.00	0.00	0.00	0.00	0.00	0.00	1 350.00	0.00	4 350.00	0.00	21 750.00	100.00	21 750.00	21 750.00	0.00	No	n/a
Σ consortium	5 764 869.08	0.00		0.00	0.00	0.00	475 130.91	0.00	1 560 000.01		7 800 000.00		7 800 000.00	7 800 000.00			0.00

---

[1] See Article 6 for the eligibility conditions.

[2] Indirect costs already covered by an operating grant (received under any EU or Euratom funding programme; see Article 6.5.(b)) are ineligible under the GA. Therefore, a beneficiary/linked third party that receives an operating grant during the action's duration cannot declare indirect costs for the year(s)/reporting period(s) covered by the operating grant, unless it can demonstrate that the operating grant does not cover any costs of the action (see Article 6.2.E).

[3] This is the theoretical amount of EU contribution that the system calculates automatically (by multiplying all the budgeted costs by the reimbursement rate). This theoretical amount is capped by the 'maximum grant amount' (that the Commission/Agency decided to grant for the action) (see Article 5.1).

[4] The 'maximum grant amount' is the maximum grant amount decided by the Commission/Agency. It normally corresponds to the requested grant, but may be lower.

[5] Depending on its type, this specific cost category will or will not cover indirect costs. Specific unit costs that include indirect costs are: costs for energy efficiency measures in buildings, access costs for providing trans-national access to research infrastructure and costs for clinical studies.

[6] See Article 5 for the forms of costs.

[7] Unit : hours worked on the action; costs per unit (hourly rate) : calculated according to the beneficiary's usual accounting practice.

[8] See Annex 2a 'Additional information on the estimated budget' for the details (costs per hour (hourly rate)).

[9] Unit and costs per unit : calculated according to the beneficiary's usual accounting practice.

[10] Flat rate : 25% of eligible direct costs, from which are excluded: direct costs of subcontracting, costs of in-kind contributions not used on premises, direct costs of financial support, and unit costs declared under budget category F if they include indirect costs.

[11] See Annex 2a 'Additional information on the estimated budget' for the details (units, costs per unit).

[12] See Annex 2a 'Additional information on the estimated budget' for the details (units, costs per unit, estimated number of units, etc).

[13] Only specific unit costs that do not include indirect costs.

[14] See Article 9 for beneficiaries not receiving funding.

[15] Only for linked third parties that receive funding.

Digitally sealed by the European Commission
Date: 2020.11.18 11:30:17 CET