EOSC-LIFE: BUILDING A DIGITAL SPACE FOR THE LIFE SCIENCES

FIRST EOSC-LIFE PARTNERS CLOUD DATA DEPLOYMENT CALL

WP1 – **Publishing FAIR RI data resources in the EOSC**
Lead Beneficiary: **ELIXIR / EU-OPENSCREEN**
WP leader: **Helen Parkinson and Phil Gribbon**

Authors: **Helen Parkinson, Andrea Zaliani, Phil Gribbon**
Contributor: **Steven Newhouse**
Version 2: Deadline extension

FIRST EOSC-LIFE PARTNERS

CLOUD DATA DEPLOYMENT CALL

UPDATE: We are pleased to announce that the **WP1 call for proposals deadline has been extended from 5 June to 30 June 2020,** in response to community requests.

Scientific enquiries may be made to WP1 coleads Phil Gribbon and Helen Parkinson. Prospective applicants should refer to the **list of eligible organisations (EOSC-Life beneficiaries and linked third parties)** to determine their eligibility prior to submission.

1. AIMS AND REQUIREMENTS OF THE FIRST CALL

The first call's primary aim is to recruit novel cloud deployments from RIs participating in EOSC-Life. This is the first of two planned calls to be implemented for WP1. A separate planned call from WP3 will have a broader focus and applicants for the WP3 call may also include institutions which are non-beneficiaries of EOSC-Life.

The reasoning for two calls is that there are potentially a large number of implementations per RI with insufficient personnel to deliver these in a single phase. Since the RIs are at different stages of technical maturity, having two calls means RI expertise will accumulate, and training and knowledge exchange will be more effective. Two calls from WP1 will allow us to react to the changing needs of the user communities as their respective scientific domains advance and as other EOSC projects come on line. The second call will launch in Q4 2020.

Proposals submitted under the first call need to be **technical feasible** (a technical feasibility review will be the first step of the evaluation process!) and should meet some or all of the following general criteria:

1. The dataset/service/data resource will be made **newly available** in a cloud instance; the first WP1 call is not intended to support existing cloud-based data/services/resources

2. The proposal involves an **identified user group(s)** (academic or industrial) **outside the submitting organisation** and **preferably beyond the existing user group**. In the ideal case the target user group needs are currently unmet and/or severely constrained

3. Proposals can involve **multiple RIs** and demonstrates the unique nature of EOSC-Life to promote cross-RI collaboration

4. Proposal facilitates cloud access to data/services/resources originating from **new institutional partners or countries which have recently joined/acceded to an RI**

5. The **cloud instances and technologies** proposed are those **supported by WP7** of EOSC-Life

6. The **workflows and/or analytics services** are those **supported by WP2** of EOSC-Life

7. The implementation includes **monitoring** for the collection of user statistics and user feedback

8. The proposal supports **FAIR principles** and implements relevant **FAIR metrics**

2. CONDITIONS FOR PROPOSAL SUBMISSION

1. Submitted by an EOSC-Life consortium partner (EOSC-Life beneficiaries), with a letter of support from the corresponding RI coordination entity(ies)

2. Proposals must have a dataset/service/pipeline on which to build and be technically feasible within the scope of expected funding. This proposal is not intended for "start from scratch" proposals

3. Proposals should have a run time of 12 months from start to completion of planned work

4. Funding per proposal is expected to be 50-75K Euro (approx 8-12 months FTE); smaller proposals will be considered. More funding may be possible for impactful projects

5. Existing resources should be identified in the proposal, for example, available personnel, computational resources or complementary institutional funding

6. Any needs for technology experts or additional computational/data resources should also be identified in the proposal

7. Proposals should include a sustainability plan for the product (supported by WP8)

8. Proposals should use the technologies supported by WP7 and will be allocated appropriate cloud resources as determined by WP7

9. Proposals must adhere to the FAIR principles

10. Any proposal focussed on long-term cloud deployment of Covid-19 pandemic-associated data sets or resources, or which is readily transferable to this application area, would be welcome. However, this is not in any way a necessary condition for

submission of proposals, and the call remit is primarily intended to cover the full range of life science data from the EOSC-Life community

3. ACTIVITIES CONSIDERED AS OUT OF SCOPE

1. Novel dataset generation - the point of the call is to make data available and not to generate new data. Data should be available and accessible at the point of application

2. Large-scale data management-only activities - this call should not replicate activities in a project's data management plan and should have a wide and demonstrable impact and use beyond a single project

3. Data curation of existing datasets - while improvements to datasets are in scope, provided they are newly accessible, this call is not intended to deliver curation resources for datasets

4. Activities that are already funded by other sources. Care should be taken to differentiate applications to this call from other funded projects

5. Proposals which only address the needs of an existing user base and do not have the potential to extend new functionality to additional users or communities

6. Simple extensions or continuations of existing Demonstrator projects

4. PROPOSAL EVALUATION

1. Proposals will use the ARIA tool for submission: Basic template here for proposal which will be formatted into ARIA

2. EOSC-Life project managers will support the administration of the review process

3. Technical feasibility review

   a. Performed by WPs 1, 2 and 7 co-leads

   b. Timeline: 3 weeks

4. Scientific or clinical impact review

   a. Only for projects selected as technically feasible

   b. A group of 5 nominated individuals (external to the project) will be identified by coordinator and WP1 leads

   c. Timeline: 3 weeks

5. Ranking criteria

   a. Technical feasibility - 30%

   b. Scientific impact and excellence - 30%

 c. Potential to deliver new infrastructure/data/services/methods for EOSC-Life - 20%

 d. Quality of training plan for the proposal - 10%

 e. Ability to broaden the EOSC-Life portfolio via diversity, and quality of the proposals - 10%

ANNEX A: WP1 EXAMPLE PROPOSAL AS STRAWMAN

CLOUD-HOSTED AND CONTAINERISED ANALYSIS SERVICE ACCESSING CONTROLLED ACCESS DATA

1. Datasets A-Z are controlled access datasets for which Life Science AAI provides authorisation, access and identification services. The raw data (0.5Pb) is hosted in a controlled access data repository implemented as a relational database hosted on disk and containing the meta data only. The data files are referenced from the relational database and are also controlled access. The meta data is deidentified (personal identifying information has been removed) and user access to the database and data files is tracked for audit and security purposes.

2. The proposed implementation will provide cloud-hosted access to a subset of the data in support of Project Y, which will co-analyse the data with an unpublished dataset. The public data will act as a replication study. This requires a subset of data relevant to Project Y to be identified and for this data to be secured and accessible only to authorised users from Project Y. The data will be periodically updated as new data are received, and some elements of the data may be removed by the owner's request to conform with the rights of the participants to withdraw from the study. Project Y will deploy their workflow within the EOSC-Life provided environment.

3. This proposal will act as a Proof of Concept (PoC) for scenarios in which data are exposed to one or more user groups to enable access by a colocated cloud instance avoiding data egress charges or data streaming challenges. The tasks expected to be addressed are:

 a. Identification and secure population of a controlled access dataset in a cloud instance

 b. Implementation of Life Science AAI in a cloud hosted system

 c. Implementation of a system enabling audit of cloud-hosted data

 d. Accessibility to a pre-existing data analysis workflow

 e. Access to a workflow service from EOSC-Life: this will be used to perform the analysis

 f. Note that no assumptions of legal responsibility for EOSCProviders of services/infrastructure are considered in this strawman proposal, or in the access to compute resources or other infrastructure expected to be provided by EOSC-Life

ANNEX B: RELEVANT INFORMATION FROM DOA

Aims of Work Package 1 (extracted from DoA)

- Development of cloud compatible FAIR-compliant data resources
- Assessment of cloud feasibility of data/data resources and publication these repositories in EOSC for data reuse
- Advance the evolution of RI repository infrastructure for EOSC (sustainability) and the interfaces between the repositories supporting the RI demonstrators and open calls
- Provide controlled access data repositories and deployments (in support of WP4)
- Implementation of FAIR services and WP1 standards for RI data/data resources (delivered in WP6)
- Provide implementation strategies and exemplar implementations for cloud accessibility of RIs archival and added value data resources
- Integrate repository/data access with workflows (supporting WP2)
- Develop a technical sustainability strategy for data/data resources deployed

Open Call activities described in the DoA

In a second phase, we will execute a call for additional RI nominated resources for deployment in WP1. This process will be synchronised with the WP3 calls where the RI data experts will ensure sufficient information is recruited during the call to determine each project's implementation needs and cloud readiness in WP1. If necessary thematic groupings will be refined and extended based on the results of the open calls.

The call will prioritise WP1 interactions with the Training WP9 as the successful open calls represent an important user group. We expect that where an RI nominates a cloud feasible data/data resource(s), it will support an additional use case, e.g. Workflows (WP2).

Budget has been reserved to support RI resources from open calls.