

EUDAT Call for Data Pilots

15th May 2015 <https://www.eudat.eu/eudat-call-data-pilots>

EUDAT offers common data services, supporting multiple research communities as well as individuals, through a geographically distributed, resilient network of 35 European organisations. These shared services and storage resources are distributed across 15 European nations and data is stored alongside some of Europe's most powerful supercomputers. Covering both access and deposit, from informal data sharing to long-term archiving, and addressing identification, discoverability and computability of both long-tail and big data, EUDAT's services address the full lifecycle of research data.

EUDAT invites research communities and research infrastructures to test, integrate and/or further develop, in close collaboration with the EUDAT consortium, one or multiple components of the EUDAT service suite through concrete data collaboration pilots. EUDAT offers data management solutions, storage resources, and consortium expertise to:

- support multi-national research communities and research infrastructures to identify solutions to their data challenges in close collaboration with EUDAT
- work with and support global research data standardization of initiatives such as RDA, and
- further develop existing EUDAT services & identify new ones

Data deposited onto EUDAT resources will be available through different interfaces, including GridFTP and HTTP, and maintained as near as possible to processing facilities required by the community, to facilitate further usage and transmission. In addition EUDAT offers training and consultancy services to support the data pilot teams, visibility and storage resources. Pilots can avail of EUDAT data storage for from the start of the pilot, free of charge, and up to 5PB in total at one, or several, of the 13 EUDAT sites. EUDAT will promote and disseminate the data pilots to a broad range of stakeholders offering considerable visibility opportunities. Solutions building on general research data-management approaches such as proposed by the Research Data Alliance (RDA) will be favourably viewed. **EUDAT will match all effort with a corresponding amount (1 PM Pilot Team = 1 PM EUDAT)**

About EUDAT

EUDAT's vision is to enable European researchers and practitioners from any research discipline to preserve, find, access, and process data in a trusted environment, as part of a Collaborative Data Infrastructure (CDI) conceived as a network of collaborating, cooperating centres, combining the richness of numerous community-specific data repositories with the permanence and persistence of some of Europe's largest scientific data centres.

As a user-driven, service-oriented, trusted, secure and sustainable data infrastructure, EUDAT offers solutions for finding, sharing, storing, replicating, staging and performing computations with primary and secondary research data. Community-specific data repository managers can join the data infrastructure to archive, replicate, process and catalogue data on behalf of their community. While researchers (from academia and industry), citizen scientists, policy makers, and members of the public can share, discover and re-use data via EUDAT.

- **B2SHARE:** a user-friendly, reliable and trustworthy service for researchers and communities to store and share small-scale research data coming from diverse contexts.
- **B2SAFE:** a robust, safe and highly-available data management and replication service allowing community and departmental repositories to replicate and preserve their research data across EUDAT data nodes.

- **B2STAGE:** a reliable, efficient, easy-to-use service to ship large amounts of research data between EUDAT data nodes and workspace areas of high-performance computing systems.
- **B2FIND:** a simple, user-friendly portal for finding research data collections stored in EUDAT data centres and other data repositories.
- **B2DROP:** a personal cloud solution based in the trusted EUDAT CDI domain to store and share datasets in the early state of the research data life cycle.

These core services are supported by multiple underlying components including a PID and a federated authentication and authorization infrastructure. The development, implementation and maintenance of the EUDAT services has been the product of ongoing dialogue with the communities, close collaboration with external experts in four focused working groups (dynamic data, semantic annotation, workflows, and policy & sustainability), and carefully managed pilot projects.

Scope of the Data Pilot Call

Focusing on the core EUDAT services and with a view to collaborative development of future services, EUDAT seeks applications for data pilots in any of the areas listed below

Data synchronisations and exchange

- Sharing and exchange of data between community (own)Cloud storage environments and B2DROP to stimulate cross community collaboration
- Publication of data from a community (own)Cloud storage or B2DROP via B2SHARE
- Further development of the B2DROP service towards an integrated generic data (staging) space for unconsolidated or preliminary data, e.g. those at the pre-publication stage.

Data repository and data sharing

- Establishment of custom instances of the B2SHARE service for efficient community usage (e.g. metadata extension with community specific metadata entries, restriction of access to data objects to specific user-defined groups, etc.)
- Integration of the B2SHARE repository service into community environment and platforms (portals, CMSs, etc.)
- Sharing and publication of large scale datasets (>100TB, 1000+ objects) which require extended upload functionality or integration with B2SAFE, on the one side and/or B2DROP or own data workspace environments, on the other side.

Data replication and preservation

- Replication of large data sets across multiple pan-European sites, with distributed data access and fine-grained authorisation rules
- Implementation of community specific data management policy requirements, for example metadata management nor data curation
- Integration with Community Data repositories based on popular repository packages (e.g. FedoraCommons, CKAN, Dspace, ...)
- Optimization of data access (geo-distance-aware, site-capability-aware access to distributed data, etc.)

Data staging for analysis and processing

- Staging to local computing environments
- Staging to external computing environments (e.g. PRACE, EGI)
- Staging of data on basis of HTTP and RESTful interfaces with extended metadata support
- Integration of B2STAGE within a portal of workflow environment for optimized data movement

- Improvement of data transfer between large-scale data workspaces (e.g. staging areas) and repository services including B2SHARE and B2SAFE

Data discovery and search

- Harvesting of community metadata records in B2FIND
- Development of discipline facets for optimal search
- Federated search across multiple repositories

Data typing and visualization

- Registration of community used data-types in EUDAT Data Type Registry
- Development of the EUDAT hosting framework to provide visualization and conversion services tailored to community specific needs

New services or tools in the area of

- Big Data Analytics
- Semantic web

Who can participate?

The call is open to all European initiatives, infrastructures and research communities. EUDAT services target a broad range of disciplines, in particular those looking for simple solutions to store and/or perform computations with research data, or for a robust data infrastructure to manage, access use, re-use and preserve their data.

How to apply?

Interested applicants are expected to provide a short description of their research community and data requirements, as well as a description of the proposed collaboration pilot which should address at least one of the areas outlined in the scope of the call.

Applicants should contact the EUDAT team for any assistance or information required for the preparation of their proposal at <http://www.eudat.eu/support-request>

Applications should be submitted using the on-line template at <https://www.eudat.eu/eudat-call-data-pilots-application-form> (you may edit at any time until call closure on 9 Oct 2015 at 17:00 CET)

Timing:

Deadline for submission: 9th October 2015 at 17:00 CET

Proposal Evaluation: 10th October – 1st December 2015 (During this period you will be contacted for any further information that is required.)

Notification: 15th December 2015 (All applicants will receive information about the outcome of the evaluation.)

Implementation: 1st January 2016 (Successful applicants will begin implementation with EUDAT.)

Duration: Collaborative data pilot activities are expected to last for 18 months (June 2017)

Assessment & meetings: Data pilots will have 2 interim assessments & representatives will be invited to attend 1 User Forum & 1 event for output & results presentations

Evaluation Criteria & Process

Evaluation & Negotiation will be carried out by a team of EUDAT reviewers and the applications will be categorized on the basis of their estimated level of complexity. Applications will be shortlisted according to the following criteria:

Criterion	Weight	Ranking (1-5)
Technical feasibility of the pilot (within the proposed timeline and resources)	40%	
Involvement of and expected benefits for the research community/discipline targeted	20%	
Added value to the EUDAT infrastructure (in terms of service development and expansion in other scientific research domains and communities)	20%	
Contribution to Open Access	10%	
Solutions building on general research data-management approaches such as proposed by RDA	10%	

Call for Data Pilots Application form

Type of Applicant (required)

- Research Community
- Research Infrastructure
- Other. Please state in the corresponding field opening below

1. Title of the pilot (max 250 characters) (required)

What is the title of your pilot?

2. Purpose of the pilot (max 4000 characters) (required)

What is the purpose of your pilot? Please be as detailed as possible including which organizations will participate to the Pilot and what role they will play.

3. Research community and researcher outline & expected benefits and impact (max 4000 characters) (required)

Describe the community (or communities) you will target with this pilot and what are the expected benefits and impact. Highlight any existing or future cross-disciplinary research aspects.

4. Estimate storage requested (required)

- 10-50 TB
- 50-300 TB
- 300-700 TB

5. Estimate users involved (required)

- Under 100
- 100 to 1000
- 1000 to 10000
- over 10000

- Other. Please state in the corresponding field opening below
6. EUDAT services involved. Select at least one of the following: (required)
- Data synchronisation and exchange
 - Data repository and (long tail) data sharing
 - Data replication and preservation
 - Data staging for analysis and processing
 - Data discovery and search
 - Data typing and visualization
 - New services or tools in the area of Big Data Analytics
 - New services or tools in the area of Semantic web
7. Technology Integration & Service development (required)
What technology and service improvements and integration do you foresee during the pilot implementation? Describe any further services or improved EUDAT service offer you foresee.
8. Building on Research Data Management Approaches (Max 2000 characters) (required)
Outline, as part of your pilot, what solutions will build on general research data-management approaches like the ones proposed by the Research Data Alliance (RDA).
9. Contribution to Open Access (Max 2000 characters) (required)
Please detail any data policies and open access plans currently envisaged for the data involved in your pilot, including details about how will this data be exploited and/or shared/made accessible for verification and reuse. If the data cannot be made available, explain why.
10. Estimated Effort (Max 1000 characters) (required)
What is the expected effort in terms of human resources? EUDAT will match all effort with a corresponding amount (1 PM Pilot Team = 1 PM EUDAT)
11. Duration (required)
Please estimate the time required to complete the pilot. EUDAT estimates a duration of 18 months.
12. Data Storage Site Preference (required)
Please indicate if you have a preferred site for data storage
- CSC
 - STFC
 - JUELICH
 - CINES
 - SURFsara
 - RZG
 - BSC
 - SNIC
 - KIT
 - EPCC
 - CINECA
 - SIGMA2
 - PSNC
 - No preference
13. PILOT CONTACT DETAILS
- First name (required)
 - Last name (required)
 - Email (required)
 - Telephone number (required)
 - Organisation (required)
 - Role in the organisation (required)