



OpenWebSearch.EU
“Piloting a Cooperative Open Web Search
Infrastructure to Support Europe’s Digital
Sovereignty”

**Community Programme:
Call #3 for proposals
– Information for applicants**

Open Web Search 

The Project is funded by the EU under GA 101070014



Funded by
the European Union

Table of Contents

| | |
|---|-----------|
| 1 Introduction | 2 |
| 1.1 OpenWebSearch.eu Project | 2 |
| 1.2 OpenWebSearch.eu Community Programme | 2 |
| 1.3 Overall Procedure | 3 |
| 2 Call Topics | 4 |
| 3 Target Audience | 5 |
| 4 Submission and Evaluation | 6 |
| 4.1 Important Dates | 6 |
| 4.2 Submission Procedure | 6 |
| 4.3 Evaluation | 6 |
| 5 Awarding Procedure and Timeline | 8 |
| 6 Grant Implementation | 9 |
| 6.1 Integration and Capacity Building | 9 |
| 7 Legal Part: Declarations | 11 |
| 8 Contact | 12 |
| 9 Project Background | 13 |

1 Introduction

1.1 OpenWebSearch.eu Project

Web search has become an essential technology and infrastructure, driving not only future innovations but forming a backbone for our digital economy. However, currently a few non-European commercial gatekeepers control Web search, which creates a biased, one-sided information access centred around economic success rather than the needs of individuals or European values and jurisdiction. This one-sided ecosystem puts pressure on the many small contributors to the Web in science, economy, art, culture, media and society to optimize their content to be taken-up by the few gatekeepers and thereby greatly feeding the gatekeepers' monopolies. A vicious cycle, leading to lock-in effects and to a closed search engine market.

To open the search engine markets and to provide a true choice for users in selecting the search engine based on their preferences, OpenWebSearch.eu proposes to develop and pilot the core of a European Open Web Index (OWI) and the foundation of an open and extensible **European Open Web Search and Analysis Infrastructure (OWSAI)**. Our approach is based on four objectives, namely (1) to develop a core suite of search, discovery and analytics services to create, maintain and utilize the OWI; (2) to develop relevant search engine verticals and new search paradigms demonstrating the impact of the OWI; (3) to establish a network of European HPC infrastructure, research and business organisations to pilot the OWSAI based on Europe's values, principles, legislation, ethics and standards and (4) to stimulate an Ecosystem around the OWI.

The envisioned infrastructure will not only contribute to Europe's sovereignty for navigating and searching the web, it will also allow Europe's researchers, innovators and business to systematically tap into the Web as business and innovation resource, without paying huge upfront costs.

More information on the conceptual approach can be found in the "Project Background" section below.

1.2 OpenWebSearch.eu Community Programme

The OpenWebSearch.eu project consortium is eager to identify new project teams to be funded under the OpenWebSearch.eu Community Programme, on-board and integrate them into ongoing and future activities for sustainable Research and Development on Open Web Search. The Community Programme is comprised of activities in support of third-party project initiation, execution and successful completion. Third-party activities are driving forces in the OpenWebSearch.eu Community Programme, with funding provided by the OpenWebSearch.eu project (funded by the EC under the GA 101070014).

OpenWebSearch.eu proposed three Financial Support for Third Parties (FSTP). In spring 2023, we conducted the first FSTP call for proposals, and selected six projects for funding under the OpenWebSearch.eu Community Programme. **The second and third calls are opened from 8th February 2024** inviting researchers and innovators to submit their applications to contribute to the mission of OpenWebSearch.eu project. **Applications will be accepted until 4th of April 2024, 17:00 (CET)**.

The call #3 is dedicated to on-boarding of computing and data centres that are closely related to the project. This will help us to extend the network of infrastructure organizations, and host parts of the OWSAI infrastructure. They should aim at widening and enriching the existing R&D activities as well as suggesting new ones, complementary to the project goals and aims.

Successful applications can request funding between 100,000 and 150,000 EUR in this call, for a funding period of up to 12 months. Major cost categories can include hardware costs, service costs and human resources.

1.3 Overall Procedure

The Community Programme started in 2023 with **openly published call for proposals (#Call 1)**, specified with a clearly defined application procedure. In all calls of the project the applicants asked to submit a short and concise proposal addressing the specified call topics. The call topics are developed within the consortium, partially with the help of external experts. Applications are reviewed by the members of the consortium as well as by the invited external experts chosen beforehand; the procedure is comparable to a review procedure by a programme committee of a scientific conference.

The objective is to select projects that have the greatest potential to maximize the results, outcomes, and impact and success of the OpenWebSearch.eu project. These third-party projects should possess a well-defined research or development component aimed at promoting Open Search as a whole.

2 Call Topics

Call #3 particularly addresses computing and data centers to join our network of infrastructure organisations and host parts of the OWSAI infrastructure. We aim to on-board more data centers as a proof-of-concept for the developed technology and engage in discussions on creating a sustainable future infrastructure.

Projects are expected to implement the following activities:

- Proof-of-concept of federation and cooperation with the other participating data centres, particularly LRZ, IT4I, CSC, DLR and CERN;
- Provision of storage infrastructure using iRoDS and S3 compliant data stores;
- Provision of compute infrastructure for running (parts of) our technology stack, which includes:
 - Apache Storm-based Crawling Infrastructure,
 - Apache Spark-based Preprocessing and Indexing Infrastructure,
 - GPU-capacity for AI-training and inferencing.
- Willingness and legal capacity to align with the project's governance and data policies in contractual form and contribute to the discussions on a joint, federated governance model, sustainability of the infrastructure as well as ethical and legal aspects.

Applicants should clearly describe their own infrastructure which is planned for the third-party project, and, indicate to which quantity will it be made available to the consortium, including storage, CPUs, GPUs and bandwidth. Additional emphasis placed on a green computing strategy and on solid security as related infrastructure will be of advantage.

The OpenWebSearch.EU project consortium is eager to on-board new third-party project teams in OpenWebSearch.eu landscape and integrate them in the future activities for sustainable Research and Development. Therefore, the candidate third-party project should probe the closely related topics addressed in the project and should aim at widening and enriching the existing R&D activities as well as suggesting new ones which are complementary to the project goals and aims.

Activities should involve research, development and, particularly, engineering, and must align with the project's goals. Results achieved must be made available as Open Source and/or Open Data, with documentations or experimental work published through Open Access, aligned with the general practices used in the project.

3 Target Audience

We are targeting universities, public infrastructure providers, data centres, companies (e.g., cloud providers) or research teams with a well-established infrastructure. The applicants should be capable of providing a significant extension to the current OpenWebSearch.eu data and compute infrastructure, and engaging in relevant infrastructure-oriented R&D activities.

Eligible organizations have to be registered in any EU Member State or any of the countries associated with Horizon Europe. Please note that the list of associated countries may change over time, and it is recommended to check the latest list of eligible countries before applying ([List of Participating Countries in Horizon Europe](#)).

Note that applications can also involve teams of different organisations. In this case, one organisation must take the role of main contact point and legally responsible party.

4 Submission and Evaluation

4.1 Important Dates

- Opening date: 8th February 2024
- Closing date: 4th April 2024, 17:00 CET
- Reviewing period: mid April-mid May 2024
- Notification date: June 2024
- Start of projects: June-July 2024
- End of projects (latest): July 2025

4.2 Submission Procedure

The applicants have to submit their appropriately formatted proposal by email to the call management (call3@openwebsearch.eu) by the given deadline. The proposal template is available for the applicants via OpenWebSearch.eu website (under Community Programme).

Note that applicant(s) can submit at most one application per call and multiple submissions of the same project are not accepted. English is the main language for communication with the OpenWebSearch.eu consortium, and all submitted documents must be written in English.

The submission will be acknowledged by the call management, and, only after the confirmation, the proposal can be considered as being submitted. It is advised not to wait till the last moment with submission. If the submission is not confirmed within max. 2 days, and provided the call deadline has not yet passed, you may contact the Grantor, at call3@openwebsearch.eu to request the information and ask for re-submission.

4.3 Evaluation

Topical fit and contribution to the project has been set as the key criteria for the selected third-party projects. Thus, the call addresses specific questions in the project raised by the consortium members. Also, the evaluation procedure strives for a well-balanced mix of project-external assessment and quality control combined with expertise and knowledge on the project-internal demands.

We established the following evaluation procedure. The evaluation will be jointly conducted by the external experts in cooperation with the members of the project consortium.

Proposals will be evaluated according to the following categories and sub-categories:

1. **Relevance** of the resources for the OWSAI in terms of quantity and quality of services offered, and on how the proposed work contributes in the pilot infrastructure of the project;
2. **Experience** in providing scalable infrastructure, and with the technology stack used in the project;
3. **Contribution** to the goals of the OpenWebSearch.eu project, and how well the third-party project contributes to the infrastructure providers strategy;
4. **Support concept and reachability** offered to the OpenWebSearch.eu project;

5. **Expected impact for Open Search** in general (also beyond the pilot infrastructure);
6. **Contributions** in governance and sustainability, on the conceptual level;
7. **Feasibility**, i.e.,
 - whether availability and security can be provided to a sufficient level,
 - whether the planned outcome is realistic,
 - whether the provided resources are appropriate for achieving the planned outputs and outcomes,
 - whether the envisioned target community / ecosystem is addressed,
 - sufficient past expertise of the applicants for carrying out the proposed work.

In total 100 points will be distributed among the categories 1-7. Besides the above listed criteria, formal eligibility criteria in accordance with the grant agreement and European Regulations must be fulfilled, particularly on the target group and official funding criteria.

5 Awarding Procedure and Timeline

The assessment will be conducted, after the primary screening for completeness of the proposal. Following the screening, candidate projects will be evaluated by the review committee. The review board will be comprised with the selected principal investigators of the project and invited external experts. Proposal review will be conducted according to the pre-defined criteria. The evaluation will result in respective points assigned to the sections of the proposal. As a result, a ranked list of all accepted proposals of the call will be made, and the call funding will be distributed to the top candidates of the list. The awarded project summaries will be published online.

The consortium aims for prompt evaluation procedure of approximately within two-months.

6 Grant Implementation

The applicants will be informed by the Business Management Unit (BMU) of the project on positive or negative outcome of the evaluation and selection process (email communication).

The successful applicants will be informed about awarding by means of an electronic funding letter. After signing a legally binding contract (a sub-grant agreement) between University of Passau (the coordinator), BADW-LRZ (the Grantor) and the (main) applicant (the Grantee), and communicating applicant's bank account information, the pre-financing will be transferred to the Grantee by BADW-LRZ on behalf of the consortium.

The Grantees will receive a pre-financing of 50% at the start of the project. The subsequent transfer will 25% of the funding will follow after the successful mid-term review, and the final payment of the 25%, after the approval of the final report.

The grantee is expected to implement the project based on effort, meaning that the promised project results are expected to be delivered within the specified time-frame. If a deliverable cannot be achieved, the grantee must inform the parent consortium in a timely manner and propose mitigation strategies. The proposed strategies will then be negotiated with the OpenWebSerach.eu beneficiaries.

The consortium members selected for the project will review the project deliverables, and the grantee will be notified of any rejection within the time specified in the grant agreement. If the Grantor does not object to the project deliverables within the specified time-frame, their silence may be considered as an implied approval of those deliverables.

An interim project report will be required for the mid-term assessment of the project progress and funding transfer. At the end of the grant implementation, the final results must be presented to the consortium for review. The final payment to the grantee will depend on the positive outcome of the project deliverable review. The grantee will receive the final payment only after acceptance of the final report.

Documents required from the Grantees:

- Completed and signed sub-grant agreement;
- Project reports, including the interim and final reports;
- Financial reports, including budget reports, receipts, and invoices.

6.1 Integration and Capacity Building

The Grantor representing the consortium plans some capacity building measures directed towards seamless integration of the third-parties with the parent OpenWebSearch.eu project.

During implementation, the Grantees will be invited to the consortium meetings and will get access to selected results of the project. If suitable, the consortium will integrate the grantee in the relevant work package and appoint a contact person from the consortium who supports them in promoting the applicant's contributions. During the grant implementation, the grant owners should present their current state of the project at the consortium meeting either in person or virtually or, if the time of the meeting is inappropriate, at a negotiated appointment.

Beyond the financial support, the awarded projects can apply for free services such as accessibility, security or license scans at NLnet¹ or take on business accelerating services from the NGI community².

¹ <http://nlnet.nl/ngi0/review/>

The third-party projects as integral part of OpenWebSearch.eu project's Community Programme can greatly benefit from the NGI support measures and promotional activities such as e.g., NGI catalogue, NGI Communication Task Force and NGI Impact, Exploitation and Sustainability Working Group. The relevant integrative actions within NGI community are already in progress.

² <https://www.ngi.eu/acceleration-services/>

7 Legal Part: Declarations

Note that in order to evaluate your proposal within the third-party call pipeline, the processing of personal data will take place. Personal data submitted within the proposal will be processed in accordance with applicable EU and national data protection law. Such data will be processed by the OpenWebSearch.eu Consortium partners and the external reviewers for the purposes of evaluating the proposed project and implementing, managing and monitoring the evaluation process. After the evaluation process, personal information on rejected proposal will be deleted, with the exception of reporting obligations towards the EU. All personal information from the applicants, particularly the information used in reporting, will be treated as sensitive information according to Article 13.1 of the General Model Grant Agreement.³ Personal information for accepted proposals will be used by the consortium to prepare the sub-grant agreement, which regulates all further use of personal data in the project implementation. In case sub-grant agreement fails, the proposal will be treated as rejected.

In the respective declarations of the proposal you confirm that you have read and understood this information clause concerning processing of the personal data; that you have the legal basis for processing personal data of your team members mentioned in the application; that you will pass the information clause provided under legal part of the proposal to all team members concerned.

Respectively, in the legal context applicants will be asked to acknowledge the such declarations as e.g.:

- You have read and understood the relevant information above. You agree to participate in the OpenWebSearch.eu project;
- You understand that you can withdraw at any time without giving reasons and that you will not be penalised for withdrawing nor will you be questioned on why you have withdrawn;
- You acknowledge that the reviewers and the European Commission and its bodies and agencies may have access to the data collected under the open call;
- The data provided in the application form are true and up-to-date;
- The entity you represent meets the eligibility conditions described in call conditions;
- There is no conflict of interest between the organisation you represent and any of the OpenWebSearch.eu consortium partners or the reviewers;
- You did not make false declarations in supplying the information required, as a condition of participation in the Open Calls and do not fail to supply this information;
- Published results will required to acknowledge the funding from OpenWebSearch.eu (funded by the EC under the GA 101070014) and undergo a peer-review process.

³ https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/common/agr-contr/general-mga_horizon-atom_en.pdf

8 Contact

If you have further questions on the application procedure and third-party call related activities of OpenWebSearch.eu project, feel free to contact us.

Administrative contact for the call #3

call3@openwebsearch.eu

Informal inquiries about the community program

community@openwebsearch.eu

9 Project Background

OpenWebSearch.eu aims to develop and pilot the core for a European Open Web Index (OWI) and the foundation of an open and extensible European Open Web Search and Analysis Infrastructure (OWSAI) by bringing together strong European players, who jointly define, develop and pilot an open technological backbone for cooperative web search. The proposed pilot infrastructure will demonstrate, how search applications and web-based AI data products can be realized through cooperative crawling, analysis, storing and indexing of web content.

Four objectives are pursued by the project:

- **Objective 1** forms the technical core by creating a suite of search, discovery and data analytics services to create, maintain and utilize the OWI. The envisioned infrastructure will demonstrate the feasibility of developing search engine verticals in the long tail and corresponding Web-based data products. While adhering a full open source and open data strategy.
- **Objective 2** aims to develop sensible search engine verticals as demonstrators and for bootstrapping a new search engine and web-data product market. Examples include search verticals for science search and mobile search as well new search technologies, particularly new search paradigms and knowledge representation models (symbolic or sub-symbolic).
- **Objective 3** aims to establish a network of European HPC-infrastructure, research and business organisations for jointly piloting the developed infrastructure while adhering to Europe's values, principles, legislation, ethics and standards. Europe has strong public computing infrastructure and research organisations in its member countries.
- **Objective 4** aims to stimulate an Ecosystem around the Open Web Index (OWI) which consists of innovators, researchers, computing centres, policy and decision makers and developers.

The development is structured along data processing chain for search and discovery systems while considering different stakeholders and services, as shown in the figure 1 below.

CCM - Crawling and Crawl Management (1 + 2) aims to conduct own web crawls and to coordinate independent crawling efforts and to gather information from website-masters to obtain additional information (e.g., provenance data, legal data, license data, website classification). An **Open Website Index** of all accessible websites of the Web plus additional metadata describing the website (e.g., categorization, topics, legal status, license information and compliances etc.) will be developed to coordinate the independent crawlers and provide crawling results.

PPE - Pre-processing and Enrichment (3) will develop **extensible pre-processing and enrichment services** with some base services for pre-processing and cleaning, but also more sophisticated semantic enrichment techniques, like removal of personal data and assessing the information quality automatically. Modules from external contributors will be welcomed.

ISPSA - Indexing, Search Paradigms and Search Architectures (4) builds indexing services for **creating full-text and metadata search indices** for the high-quality web data collections. Conceptually, every index could be maintained by a different organisation / owner in order to have a clear, legal responsible person. With this approach we aim to enable small organisations and even individuals to establish their own brand on searchable content, such that web-users can select – or even define – a search engine completely tailored to their preferences.

SVDPA - Search Verticals, Data Products and Applications (4): A series of vertical search applications (verticals) will be developed that **demonstrate and prove the capabilities and flexibility**

of the **OWSAI**. These verticals are built upon the OWSAI and exploit its various features, in order to facilitate different use cases and serve specific requirements of different end user groups. Thus, verticals not only include specialised search engines and a core managing those search engines, but also data products, smart objects, and particularly AI models. Specific attention will be paid to enhancing user-centred features, such as **advanced trust and transparency models** that open up technical information on the documents, gathering, indexing, and ranking methods.

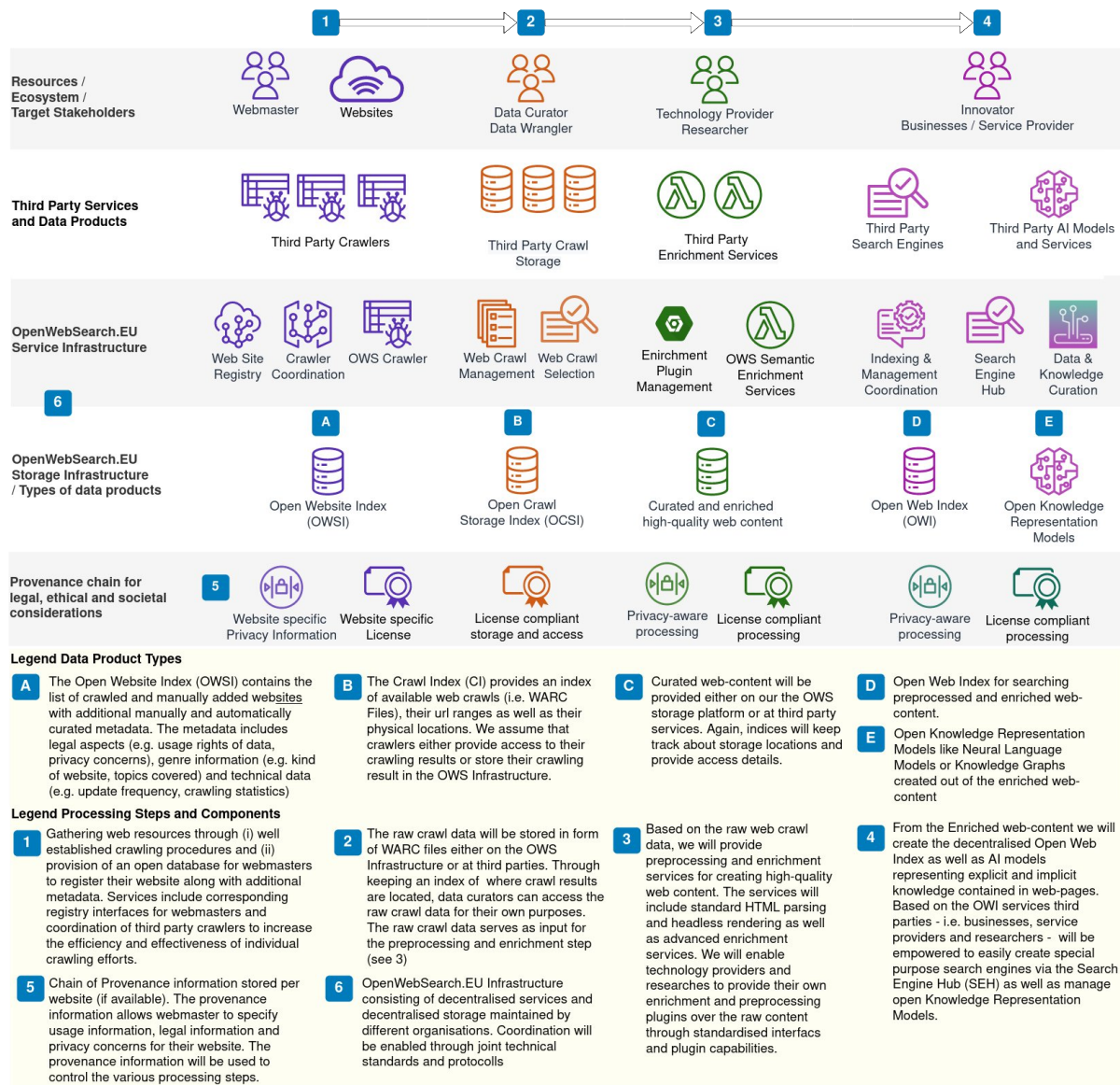


Fig. 1. Data processing chain for search and discovery systems involving different stakeholders and services.

DSPCI - Data Storage, Provenance, and Computing Infrastructure (5+6): Provenance-enabled, distributed, inter-organisational storage facilities will be essential for all the above services. Augmented by computing services for the project, these facilities will **enable FAIR⁴ data and metadata handling**

⁴ M.D. Wilkinson et al. The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3, 160018 (2016)

(and sharing), as well as high-performance data analysis. Provenance-tracking becomes especially important to understand properties like trust, privacy and security. For example, a website that has been registered by a webmaster as being marked as privacy preserving can be indexed by a vertical search engine that focuses on privacy preserving web content or for building privacy-preserving Artificial Intelligence models.