

# A free and open Internet Search Infrastructure

For our society and a prospering digital economy in Europe

## About this document

*“The internet was meant to bring more democracy, economic prosperity and cultural diversity. - It was meant to be a place of transparency, openness and equal rights.”*

According to Andrew Keen’s book “The internet is not the answer,” the initial intentions of the inventors of the internet have not been met as of today in 2018/19; instead, few network-monopolies, primarily from the USA and China, dominate the web. – What about Europe and its dependency on these quasi-monopolies, particularly in searching the internet? The usage of AI and big data techniques is increasing. Thus, the need for a free, transparent, open and economically as well as politically unbiased Internet search is larger than ever. This need is increasingly understood by the public.

This document outlines a concept and general principles for such a free and open Internet search for Europe.

It could serve as a foundation for

- a diversity of search engines,
- a flourishing digital economy in Europe and offer a
- neutral, economically uncompromising access to Internet information for the general public, academia and industry.

Collaborative web indexing itself is not a new idea. The technical foundation as well as the initial software required are available as open source projects. The necessary computational capacities already exist in Europe. Building on these, the approach combines the principles of open source code and distributed computing with transparent public moderation and lean, self-regulatory, non-commercial governance. It is the goal to build a cooperative Internet inventory and navigation capacity in and for Europe, through a concerted action, initiated by a group of committed individuals, academia, public and private-sector computing centres as well as other societal actors.

Seed funding (in-kind contributions and research funds) shall support establishing the core principles through a lean organisational framework and will support the federation of non-commercial computational and bandwidth capacity. Once the web-index and search infrastructure are deployed, the operational power-users of the system may contribute through small cost-recovering/supporting

fees to the hosting and maintenance of this distributed Internet search infrastructure. Public awareness and outreach campaigns will complement the technical and infrastructure-oriented activities of this initiative.

Open Internet Search is neither about money in the first place nor about technology, but rather about cooperation with the common goal of establishing an open search ecosystem.

Access to Internet information must be rendered open, democratic and economically independent in and for Europe.

## Acknowledgements

We would like to thank the many scientists, engineers, senior decision makers, colleagues and friends in research centres, ministries, EU and international bodies, universities, foundations, NGOs etc. for the many fruitful discussions, thoughts, positive support, impulses and support while developing the ideas and writing this paper. This vision of a democratic, free and open access to web-based information could not have been formed without the numerous inspirations we got from the many already existing initiatives in the digital-right, democracy, open-source and free information access communities.

## Authors

---

### **Dr. Stefan Voigt, Lead Author**

Researcher at the German Aerospace Centre (DLR); Founder and Chairman of the Open Search Foundation e.V. (i.Gr)

---

### **Dr. Eva Bilhuber-Galli**

Managing Director HumanFacts AG, Co-Founder of the OSF

---

### **Dr. Wolfgang Sander-Beuermann**

Computer Scientist; Managing Director and Chairman of MetaGer/SUMA e.V.

---

### **Prof. Dr. Alexander Decker**

Professor for Digital Media at the Technische Hochschule Ingolstadt; Co-Founder and Vice Chairman of the OSF

---

### **Dr. Frank Hauser**

Economist at Societe General; Co-Founder and Vice Chairman of the OSF

---

### **Christine Plote**

Managing director Plote Kommunikationsberatung; Co-Founder of the OSF

---

### **Dr. Christine Hungerer**

Medical Doctor; Co-Founder of the OSF

---

### **PD Dr. Sven Hungerer**

Medical Doctor; Co-Founder of the OSF

---

### **Prof. Dr. Nils Jensen**

Professor for Computer Science at the HWA Ostfalia

---

### **Dr. Günther Kohlhammer**

Computer Scientist; former Chief Digital Officer of the European Space Agency (ESA); Co-Founder of the OSF

**Prof. Dr. Dieter Kranzlmüller**

Computer Scientist, Director of Leibniz Supercomputing Centre  
Garching; Co-Founder of the OSF

---

**Hans-Joachim Oettinger**

Managing Director, Auditor and Tax Advisor at Oettinger-Gruppe;  
President DStV-BW e.V.; Vice-President DStV e.V.; Co-Founder of the OSF

---

**Sabine Dietloff**

Managing Director / Tax Advisor Steuerkanzlei Dietloff; Vice president  
LSWB e.V.; Co-Founder of the OSF.

---

**Dr. Martin Mittelbach**

Legal Expert and Administrative Director at Alexander v. Humboldt  
Stiftung

**Table of content**

**1 WHY: Our Vision ..... 4**

1.1 Basic principles of web indexing and search..... 4

1.2 The indispensable need for free and open Internet search ..... 4

**2 HOW: A joint initiative with well-established principles ..... 8**

2.1 Our approach: Together we are strong..... 8

2.2 Our principles: Collaboratively, distributed, open-source and publicly moderated..... 9

2.3 Generating addition value by fusing with European satellite- and geo-data..... 10

**3 WHO: The necessary actors for this non-profit and co-operative approach .....11**

**4 WHEN: The Road map – From concept to deployment ..... 13**

4.1 It’s not about money, it is about inspiring a community to co-create .....13

4.2 Early movers.....13

4.3 The steps.....13

**5 Call-to-Action: Join us and help making it happen .....17**

## 1 **WHY: Our Vision**

### 1.1 **Basic principles of web indexing and search**

The foundation of all Internet search is a well-maintained and up-to-date index of the Internet. Such an index can be considered a well-kept table of content or inventory of the publicly accessible Internet web pages. For this purpose, crawling algorithms are used to automatically and periodically access each web page. Every relevant word on each webpage is then stored in the index: A large database. Beyond the word-index, additional information about the web pages is recorded such as data on servers and hosting, graphics, tables, web links structures, images, geographic information, offered goods, prices, etc.

At the users' side a search front-end is then used to query this index of the web according to the requested search words. The documents, which include the search words, are extracted from the database and based on these. The search engine software "ranks" these results by internal ranking-algorithms and filter parameters. These algorithms typically determine the sequence of results shown to the user. Because the number of results is usually very high, the algorithms mostly determine additionally, which results are displayed to the user and which are not!

Increasingly, parts of the relevant web page content are served back to the user of the search portal. These parts are called the "snippets". They should enable the user to determine, if the underlying webpage might be relevant for the purpose of this search.

It takes substantial computing power and network bandwidth to set-up and maintain such an index. This is the challenge. Usually, this cannot be achieved by a single private or public entity alone. Currently four substantial Internet indices exist: Google (USA), Bing (Microsoft, USA), Yandex (Russia) and Baidu (China). There is no public inventory or index of the Internet accessible. This means, that all requests and responses to and from Internet search engines globally are governed by private/commercial interests, cannot be audited and thus contain unknown level of bias. In addition: All of them are hosted and governed outside of Europe. There are some metasearch engines, which do not have their own web-index. They redirect user requests/queries to one or more of the generic search engines, which then report back the results. Those results are then aggregated and displayed to the user of the metasearch engine.

### 1.2 **The indispensable need for free and open Internet search**

In order to lower or even eliminate bias from Internet searches (commercial or other), it is vital to establish a non-profit, publicly audited, open and distributed search infrastructure for the internet, i.e. a web-inventory/index as well as one or several search front ends. Just as topographic maps, cadastral information and road maps are public goods and accessible basically free of charge for everyone today, the "cadastre of the Internet" and its content needs to be rendered a public, free and open good.

### Imagine ...

- ... you can search and find information in the Internet freely, without exploitation of personal data and without being tracked, profiled or spied on.
- ... your searches in the Internet are not biased by commercial or political influences.
- ... the results shown are generated generically and by transparent algorithms.
- ... you can do business in the Internet without paying high charges to Internet monopolies.
- ... orientation in the Internet as a marketplace, a social sphere and a source of information is a public good. Thus, it is provided freely and in many forms.
- ... the privacy-depriving Internet gold-rush, fuelled by our personal data and information, is transformed into a civilized, collaborative and fair digital marketplace.

### The Reason why

#### **Network-monopolies – Europe's dependency on overseas players!**

The ever-advancing digitization of our daily life and the increasing use of Internet technologies in all spheres of our societies imply that we need to sharpen our awareness of how we deploy those and govern the use of these technologies. We must take great care that internet technology neither has negative influence on our society, decisions and economy nor on our democracy and civil rights.

Derived from the situation described above in 1.1 we observe: More than 90% of all Internet searches in Europe are done via one single search engine. Via scripts and other services, approximately 3 out of 4 web pages constantly transmit personal user data to this search monopoly. In addition, the market share of its mobile operating system on smart phones is immense. Therefore, we see a critical lock-in effect. The result is a giant Internet monopoly. This turns the Internet, previously seen as a force of liberalization, into critical dependencies – inducing an erosion of the fundamental values and rights that our society is built on. Therefore, we face a growing distortion of competition, barriers to free market access, dependencies, loss of privacy rights and accelerating inequality.

Like other fundamental democratic goods there is the need for openness and transparency in the digital sphere:

- Free and unbiased access to public information is a fundamental democratic right.
- Searching, finding and accessing information in the Internet, in an impartial and effective manner, has become a fundamental element for maintaining a sovereign and independent society and economy in Europe.
- Open Internet Search is a fundamental prerequisite for a prospering digital economy and society.

### Our vision

#### **An unbiased, democratic and privacy-respecting digitization**

We strive for an Internet where individuals, civil society, science and economy can roam, orientate themselves, search and find information in a fair, free and open way.

### Our ambition

#### **Building a co-operative, free and open Internet search infrastructure as a basis for a large variety of digital services**

We strive for the Internet as a creative, liberal, personality-respecting as well as data rights respecting space. Societal development, entrepreneurial activity and unbiased access to data, information and markets shall be possible. We strive to inspire all kinds of people –scientists, engineers, politicians, decision makers and active citizens– to join us. Thus, this initiative aims to raise awareness, educate Internet users and create a dialogue that helps to support this vision.

### Benefits

#### **Why you want us to be successful:**

Building such a free and open access to Internet information will:

- help freeing the European economy from Internet monopolies,
- establishing a sovereign European basis for the Industry 4.0 the Internet of things, and machine learning;
- provide the basis for next generation privacy-respecting Internet services and business models,
- allow a good part of the billions of Euros that are currently spent by European enterprises to a single company for improving their search-rank to be saved or spent diversely an on a multitude of platforms, rather than on single ones that dominate the market,
- reduce economic and general biases of Internet search results and the drastic societal consequences thereof and
- set in place a stable and long-term infrastructure for the sovereign development of European society and economy of tomorrow.



## 2 HOW: It's about cooperation

### 2.1 The approach: Together we are strong

#### Potential hurdles

In explaining our vision to others so far, we have heard many arguments already, why we might not succeed. People thought:

- “The current monopoly is too large and too advanced to be challenged.”
- “Somebody must have tried this before.”
- “It is too huge of a task to connect a large number of contributors, like computing and science centres to jointly inventory the Internet and host and publicly share the inventory/index as a basis for a multitude and plurality of Internet search engines.”
- “IT security is too much of an issue.”
- “No one will sponsor CPU time, storage capacity or bandwidth to support this venture.”
- “To start this on a best-effort basis and on no-exchange-of-funds will not work.”

#### Neither a top-down nor a bottom-up-approach will work alone:

The task to set-up and to maintain a public, free and efficient internet search infrastructure for Europe is too large, too strategic and too dynamic to be set in place in a monolithic top-down approach and by typical procurement mechanisms, as it was done e.g. with the European Galileo program.

Any approach of that kind would not be agile enough to be implemented in a fast, cost-effective and efficient way and would suffer from extreme difficulties to keep pace with the rapidly developing Internet technology. It also cannot be left to bottom-up market mechanisms or grass-root initiatives alone. Hence, a collaborative and distributed European approach is needed, which reflects the multitude of European language spaces, cultures and fore example top level domains in the internet.

#### Cooperative approach:

The computational efforts required, call for a synergistic use of existing computing centres: A distributed infrastructure. Setting up computational, network, integrity and data base standards requires a certain critical mass of expertise in high performance computing and network technology, financial and personal resources as well as a good understanding and reflection of European cooperation mechanisms. Thus, this initiative is best being developed and scaled within the European non-profit research/technology sector, such as large-scale science and computing centres, networks and/or associations of those, as well as government institutions and NGOs. In other words: We use existing computing capacity and make use of available capacities of these data and computing centres to jointly index the web. This allows the task to be achieved in a financially viable and relatively short-term way.

### 2.2 Guiding principles: Collaboratively, distributed, open-source, and publicly moderated

Further developing the approach described above, we suggest basic principles, how to set up and operate the necessary:

- Implementing and operating the search infrastructure via **collaborative, distributed computing** and hosting to allow easy sharing of computational, storage and network related efforts.
- Developing and operating an Internet search infrastructure based on **open source** principles and algorithms.
- Managing and curating the thematic basis of the inventory and indexing as well as the search capabilities in an open, **publicly moderated** and unbiased way.

### **Collaborative and distributed computing and hosting**

As mentioned above in 2.1, it is suggested to use a collaborative and distributed European set up. The computational, storage and network capacities to operate a high-performance web-indexing and web-search infrastructure for Europe will be substantial. Hence, the only feasible and economically viable approach is to build the core of it collaboratively: On existing public, and possibly even private sector, IT infrastructure and for a start on best-effort and no-exchange-of-funds basis.

By doing this, the implementation can literally start immediately and as soon as a first basic cooperation scheme and the first version of the algorithms are selected. During the implementation and scaling phase (see details later in this paper) the distributed computational network can be based on large scale research and computing facilities such as CERN, ESA, JRC, ECMWF, etc., as well as a large number of contributing science and computing centres in the European member states. Each of them only contributes a small percentage of their computing/network capacity to the overall system. At a later stage, telecommunication, web-service providers and private sector partners may follow in supporting and contributing to the infrastructure. While some centres will provide public web access to the index and to search front-ends, other centres may limit their contribution to providing processing time, storage or band-width. Each contributing computing facility in this distributed network will operate independent and will remain under full control of their individual computational resources and contribution. The implementation will thus be based on distributed and shared computing principles, redundancy in back-up and storage, mirroring of the latest index version and load-balancing for search- or API-requests. Simple exchange protocols and integrity monitoring procedures will have to be put in place to ensure robust and efficient data exchange among the different peers.

### **Open source**

To ensure best performance, integrity and wide acceptance of the envisaged indexing and search algorithms, their development and maintenance need to be handled in an open-source-approach and under a well-balanced framework for testing, implementation and revision. Each update of the web inventory and search algorithms will undergo careful benchmarking, performance- and integrity-checks, also established and moderated publicly and with support of the scientific community. The Linux operating system can be used as a successful example of how the development and maintenance of an opensource code and software framework works long-term and stable. Based on these opensource principles, users can have full visibility how the underlying web index was generated and how the ranking of their search results and the context searches were used to generate a search result.

### **Publicly moderated and transparent**

The inventory, web-crawling and indexing of Internet content is a technical and mainly automated task. Nevertheless, the storage, thematic moderation, legal curation and maintenance of the up-to-date web-index itself takes substantial publicly moderated organisational and thematic curation efforts. Similar to a librarian, maintaining carefully the inventory and index of all books and their main content in a library, the public and permanently updated inventory of the Internet content needs structuring, thematic aggregation and to some extent also legal supervision. This process will need to be moderated and handled in a transparent and publicly controlled manner. This will ensure that this open Internet inventory, on which Internet searches and all kinds of new commercial services can be based, remains integer, unbiased, free and accessible. The non-profit encyclopaedia Wikipedia can be seen as a model for developing such a concept of public moderation.

### **2.3 Generating additional value by fusing with European satellite- and geo-data**

With the ever-advancing virtualisation and digitization of our society and economy, geo-search capabilities will provide substantial advantage over currently existing Internet search infrastructures, which today often have only limited geo-analytical skills.

Geographic searches, allowing to locate the searched content on a map, location-based services, navigation, routing and geo-spatial orientation based on digital maps, satellite imagery and road network information will be key elements of the Internet of tomorrow. Web based geo-searches and spatial analytical features will furthermore be essential to orientate and to master any activity in the Internet of things. Europe owns very valuable and strategic assets in this respect with its Galileo and Copernicus programs as well as its Sentinel satellite constellations. To ensure that the value added from these assets is going to be generated domestically, it will be essential to provide the required powerful, versatile and state-of-the-art web-based solutions for the best possible exploitation of these programs.

The European Copernicus program for satellite-based Earth Observation not only mobilizes considerable resources, it also provides a frame to spur the integration of European cloud storage and computing facilities. Europe will have to foster the prevalence of open standards and open-source solutions in the public geo-information and Internet search domain to have a chance to build an alternative to proprietary and currently quasi monopolistic competitors.

Appropriate geographic referencing of Internet based information and respective data structures will be incorporated in the European web inventory and search technology, which will then help to synergistically turn the valuable European geo-data streams even better into market-relevant and actionable information. This will support applications in business, ecology and global pressing topics such as energy efficiency, climate protection, natural resource/water management, urban planning, smart cities and many more.

### 3 **WHO: Key actors for this non-profit and co-operative approach**

This approach builds on the readiness of public, private and individual actors to collaborate, synergistically join forces in setting up this Internet inventory and search infrastructure. Different actors are invited to engage in the undertaking. Each of them will cover different, but essential and important roles. These are:

#### **European research and technology centres**

As these centres already have substantial and relevant computing infrastructure, network capacity and expertise in place to initiate the setting-up of such an infrastructure in a fast and cost-effective way, they have the key role in this undertaking. They either will have to take the initiative to prime this approach in a concerted action or will have to be among the early adopters and supporters. If a substantial group of publicly supported computing centres in Europe dedicates only a few percent of their computing, network and storage capacity to this task each, probably no additional investments in computational hardware will be needed during the starting phase.

#### **Government bodies**

Already during the piloting and initialization phase the project will require some public financial support that must be scaled further to sustain this approach in the long run. This financial support will be needed to fund individual contributors to the network and to establish lean governance structures. It will be essential to provide this support in a way that the open-source and publicly moderated character of this Internet search infrastructure is not endangered. This will be crucial for the unbiased nature of the Internet-search-infrastructure. Public bodies may also help substantially with handling legal aspects of this undertaking.

#### **Academia**

Since the Internet inventory and search algorithms, as well as the thematic moderation, will need permanent revision, adaptation and innovation, academia will have a crucial role in supporting and advancing key elements of this distributed infrastructure in dedicated projects and/or via the open-source participation. As the web index (an inventory of the web based on a database) will be open and publicly available for European entities/organisations, it will also stimulate a large number of research and development activities in the academic domain, as currently such comprehensive web indexes are not accessible for research purposes.

#### **Libraries**

As librarians have been safeguarding and curating written scientific and cultural information by maintaining access to literature over centuries, their expertise can substantially help to organize and maintain thematic structures in the index architecture and search semantics. Scientific staff of key libraries should be involved into the project already from early stages.

#### **Private sector**

The private sector will be one of the main beneficiaries of the unbiased and free web search, as it will allow enterprises to be “found” again in the Internet, even without paying astronomic advertising fees to current web monopolies. Furthermore, the private sector will be able to build own new thematic

services and new businesses based on the free and open web-index. Contributions by large IT and network providers will be welcome, if they support the non-profit, open-source and strategic character of setting up this infrastructure. It is important to note that profit and unbiased market opportunities will develop extensively by using this infrastructure once it is in place – not from building it in the first place. The private sector will contribute either in-kind with know-how, computing power or bandwidth or in form of donations.

### **The citizen**

The citizen will also have a key role in this undertaking, as acceptance and ownership will help sustaining this public approach. The timing of the public launch of the infrastructure as well as the level of performance and maturity reached by that time will be important: Launching too early if the performance is not yet convincing is to be avoided as much as overselling. Launching too late puts at risk the motivation, support and important network effect of the community. Since the approach probably requires some evolution in the current search behaviour, a continuous outreach and education of the search engine users will be key for the approach.

## 4 WHEN: The Road map – From concept to deployment

### 4.1 It's not about money, but about inspiring a community to co-create

The great advantage of this approach: it does not require billions of Euros to get started and it is largely based on existing IT infrastructure and organisations. The challenge of setting-up this essential Internet search infrastructure is not a financial one – though of course seed and sustaining funding will be needed. However, it is mainly a matter of communication and organisation to establish a non-profit and collaborative spirit in and around this undertaking. It is about getting the European values and ideas to work in the niche of cooperative and collaborative computing, in order to build this essential infrastructure on existing IT and computing facilities in Europe. In this sense, the Google doctrine is correct: “the competitor is just a few clicks away.” It is a matter of inspiring people to move together on this and to do these “few clicks.”

### 4.2 Early movers

It is evident, that this approach to building a public Internet search infrastructure does not rely on long lasting political processes or –as described– billions of Euros to start. It is being kick-started by a small group of dedicated people working from and within public research and computing centres and jointly in an .org / non-profit-environment. We thrive to spread the word, the work-on-concepts and initial activities in order to reach a critical mass of contributing players. Further, we will need committed personal and resources. The undertaking has already started and is currently being scaled. Thus, it can be further developed technically as well as organisationally. As is done for most open-source projects also for this activity a dedicated foundation has been established in order to source, network and federate first funding and in-kind contributions to the project.

### 4.3 The steps

#### Inception phase (~ 1 year)

First and foremost, it will be important to raise further awareness for the tremendous –technological, societal, economic and even political– importance of such an infrastructure. This will be done by educational and promotion campaigns. Such campaigns will also educate on the enormous threats and biases, which will further grow, if no such free, and open Internet search infrastructure is established in Europe. In parallel, raise 200-400k€/year to fund the basic tasks of a non-profit-organisation (.org) that will further promote the activity. Raise 1,5-2M€/year to support a core team of eight to ten full time experts to work in the non-profit organisation in order to inspire research organisations to further elaborate the concepts and spin the key elements of the architecture together with further partners and players in research, academia, computing centres and even the interested IT industry. During the inception phase, the team will generate the necessary critical mass to implement and roll out the overall approach. It will involve contributing computing centres, kick-start the development of the

algorithms, help defining the overall governance as well as the legal and general organisational framing. Together with the implementation of the general framework, the foundation may be scaled to allow further dedicated funds and contributions to be raised and an overall operational model to be developed.

#### **Pilot phase (~ 1 year)**

Based on the inception phase, the scientists and engineers from the different committed research and computing centres will be designing, developing and testing the architecture, algorithms, computational setting and regulative elements in parallel. First prototypes of all required subsystems will be elaborated and tested. The network will further be developed, and outreach and communication activities will involve further partners.

#### **Implementation phase (~ 1 year)**

It will then be important to free the minimum required resources, either within the different centres or via EC/MS to implement the core of the open and distributed infrastructure and to set up the pilot system in a first operational manner. Swiftly, the basic system of crawler, inventory/indexing and search starts operating in a publicly moderated and fully distributed way.

Outreach and communication will further expand the network of supporting organisations.

#### **Scale-up phase (~ 3 years)**

In this phase the architecture, algorithms and operational models will be taken from pilot and first implementation to scale up-phase. The full scalability of the system will be proven. This enables us to deploy it to more countries and with more partners in Europe, so it can fully be used on operational scale and performance, not only crawling and indexing the EU top-level domains and the .com, .org etc. but possibly scaling to the full internet for the first time. Along with this core development, it will be important to engage the open-source community as well as key European Internet technology experts to help further developing and scaling of the network as well as the infrastructure to the targeted operational level of performance. Proper incentivisation of the developers will have to be set in place to motivate personal contributions of programmers, students, academia and many more.

#### **Operation phase (continuous)**

Raise further awareness and move from concept to operation: Once the core system is implemented and operational, it will be key to swiftly roll out the full performance and gain further partners, support and momentum. This goes together with the full deployment of the infrastructure taking it from the scale-up phase to full performance and long-term sustainability.

A permanent task will be ensuring long term funding, refining the governance and maintaining the integrity, security, and state-of-the-art of all technical and organisational elements of the network and infrastructure. This may also allow initial R&D partners and computing centres to reduce contributions and have dedicated centres, fully devoted to the long-term task of maintaining the infrastructure to take over the core operational tasks.

The table below gives a more detailed view on the different phases and tasks.

#### **Table: Phases and tasks of the initiative**

Phases \ Tasks	Technical	Organisational	Networking and Federation	Result
<b>Inception</b>	Initiate development of algorithms. Arrange for Hardware/ Network resources.	Developing the general structures, organisational model and governance. Form the community. Conceptualize legal and public aspects.	Educate the public. Establish the non-profit as nucleus of the activity. Start experiments, MVP and networking.	Organisational form and governance established. Partners are informed. Further funds are raised Ambassador/ Advisory Group.
<b>Pilot</b>	Prototypes Tests / Pilot operation of indexing and search.	Piloting of public moderation. Establish the foundation	Bring in further partners and institutional support to move from test to first pilot application.	Pilot operation is successful. Foundation is operational. Perspective for long-term success and sustainability is in place.
<b>Implementation</b>	Implementation and scaling to operational set-up. Inventory and indexing at full European scale	Rolling out full scale organisational structures. Operational moderation. Implementation of legal, security sustainability elements	Set up a core of research and computing centres to implement a first fully operational version of systems that moves from piloting to full scaling.	Infrastructural and operational features are implemented and running nominally. Internet Inventory and Index are published continuously.
<b>Scale-Up</b>	Scale to full operation. Scale to all European Countries. First time crawl and inventory all global domains.	Deploy a full implemented European governance of the project. Enter the fully operational production and optional model.	Expand the partnerships such, that the system can be implemented in all EU -Member countries and that the governance and auditing structures are fully scalable to EU level.	Long term sustainability of inventory and search infrastructure is in place.
<b>Operation</b>	Evolutionary maintenance. Continuous improvement of all	Constant monitoring of performance of all organisational aspects. Improving of	Consolidated and maintain the scale and governance of the system such, that	Infrastructure ready for long-term operation and maintenance.



	HW and SW subsystems.	governance and public elements	it may be maintained and operated on a routine and long-term basis.	Cooperative, free and open Internet search.
--	-----------------------	--------------------------------	---	---

### Outlook and challenges

Of course, there will be technical, computational and network related challenges during the development of the distributed search infrastructure. However, it is important to note, that there are no major technological barriers that need to be overcome to reach the goal. The general technical set-up needed for such an operation is relatively basic and straight forward. Again, it is not about money or technology; it is about organizing this distributed search engine and infrastructure.

There will be IT security and integrity challenges, as it might be of interest for some actors to manipulate the indexing or the search algorithms or even attack the hosting and access points. Appropriate exchange protocols, integrity and authenticity checks for data, public source code, public index and redundancy for all parts of the system will have to be set in place to tackle these possible issues.

Ensuring legal integrity will be important, as the crawlers will be inventorying also legally questionable content of the web. Proper, public and commonly accepted procedures for handling such cases need to be put in place, without restricting the freedom and openness of the indexing and search infrastructure.

It will be important to keep the core of the project open-source, since once the engine is fully deployed, it will probably gain a global dimension and impact; it may be extended or copied by other major IT players globally. This may be acceptable and even intended, if it increases openness, performance and diversity in unbiased access to Internet based information. If the latter cannot be ensured, the open auditing may be restricted to selected European players.

Setting-up a self-regulatory and effective governance and licenses for the project will be very important. Open-source licenses (GNU, creative commons, etc.) will be essential to ensure the synergistic and collaborative character of the open source code/software and even the set-up of infrastructure.

## 5 Call-to-Action: Join us and help making it happen

### **Too idealistic, too big, too complex, simply impossible? No!**

This paper provides a conceptual layout on how a free, open and cooperative Internet search infrastructure and respective search front-ends (search engines) can be designed, implemented and maintained cooperatively – in and for Europe, based on the principles of open-source, shared distributed computing and public moderation. It takes an operational nucleus and R&D funds to kick-start the development and implementation.

### **A few final thoughts:**

- It will certainly not be easy, but we are convinced that it can and will have to be done.
- We strongly believe in the power of the scientific and open source community to help realizing this project. Now.
- Let's shape together the future of the Internet we want to live and work with!
- We explicitly abstain from commercializing or trading this approach and therefore need your support.
- We appreciate any form of feedback and active contribution. You can help with advice, practical contributions to the campaign, by offering computing power and bandwidth and/or through financial backing.
- You are invited to support this initiative and get in touch with us for co-operation, ideas, contributions and comments.

Please contact us at: [yourinfo@opensearchfoundation.org](mailto:yourinfo@opensearchfoundation.org)

**Together, for a better net.**