Review Board feedback summary

Deadline for your improved application is **February 11th at 20.00** EET (Eastern European Timezone). Send your revised application to Tiiamari Pennanen at tiiamari.pennanen@uutisraivaaja.fi. The deadline can't be changed.

Address the questions the reviewers are posing the best you can. If they ask for improvements for the visualization, do that. If they ask for better benchmarking, do that. If they ask for a more detailed budget, do that. If the reviewers ask you to consider a new direction for your project, consider that. Give reasoning, if you decide not to. If you, for some reason. After the revisions, your application will be re-reviewed. Best ones will be invited to present their project to the reviewers on 4th of March at the Päivälehti Museum. If you can't travel to Helsinki for the presentation (presenting live is highly recommended), you can alternatively make a video of your project. More instructions will follow about this in the case you will be selected to present. Only a small fraction of revised applications will get a chance to present, so you should really make an effort to be among the best ones by improving your application based on the reviewers' feedback. If you are chosen to present, you'll be notified by 17th of February.

Address these comments and questions when improving your application.

Cover letter

Dear Tiiamari,

Thanks for the feedback regarding our Uutisraivaaja application. We have now crefully considered the feedback and taken it into account in the revised application. Below you can find our brief responses to the reviewer feedback - further details are described in the application text.

kind regards in behalf of the Protomedia team,
Antti Poikola

Feedback on the business model:

- What will be the price of the service for media organizations? That is the main question when thinking about the possibility of someone buying the servic

Thanks. We have now made more explicit plans regarding this in the funding section.

.

- Many managing editors don`t know much about data journalism but might be interested in it if shown interesting examples. You should think more about how you will sell your service to your potential customers.

This is now clarified in the Strategy and business model section. We use as showcases the already established examples on open data analytics and education by the team members.

- Media companies have their own data journalists, but what about freelancers and not tech-savvy students? I think this kind of service would be very handy, but how does it differ from a company selling this kind of service? To be truly revolutionary it should be a public service like a library.

Indeed our model provides a public library of easy-to-use data and code resources, while the expert services tailored to the needs of customers is the key product to be sold. Open data and source code resources will increase the transparency and trustworthiness of the data and analyses and provides additional opportunities for crowdsourcing. We aim to create an enterprise that is sufficiently large to be credible and to achieve advantages of scale that will differentiate it from freelancers. Keeping the project as a private enterprise will ensure agility. Further details are described in the revised application.

Feedback on implementation

- The idea is nice, but the implementation details seem unrealistic

Thanks for the feedback. We have now polished the implementation details and paid further attention on the details of the business plan. It should be emphasized that the key product are the expert resources for open data analytics, while open library of data and code are essential to support the service concept. Our business logic is based on the well-established observation that main revenue in open source projects comes from expert services, not the software itself. In our case, open availability of the project data and code resources will bring added value through increased trustworthiness and impact of journalistic work as described in more detail in the revised application.

.

- What data standardization technology is the team proposing to use? It's not clear from the English diagram or closing paragraph.

Our key product is the support, analytics and education regarding the use of open data streams in journalistic work. This is not limited to the open database maintained by Protomedia but includes further governmental and other data resources. It is not meaningful to try integrate various rich sources of information under a single database structure. Instead, we collect easy-to-use data resources onto a single hub server as a collection that will best support the Protomedia service concept. We will distribute preprocessed data but the final format can vary on a case-by-case basis. A centralized data catalogue will be maintained to keep track on all available data resources. We emphasize flexibility over standardization, as data standards and requirements vary, and stringent requirements for data formats will cause unnecessary friction for data utilization. Moreover, as the analysis pipelines that utilize the data will also be made open these practical demo cases will show how particular data sources can be accessed and utilized. Where possible, the data sets will be integrated into a single database structure and provided through open and standardized API with XML and JSON formats or as ready-to-use CSV files. We will use NoSQL databases such as mongoDB for storing and serving data that will not need many cross-references and SQL databases for data that will benefit from an ability to perform more complicated cross-dataset queries.

Feedback on scope:

- I would redirect it to focus only on Helsinki-area news or even news of smaller geographical area and go focus on some platforms.

To obtain sufficient revenue it is essential not to define the scope too narrowly. In fact, we aim to expand the services at international level in the long term while we agree that the service should have a well-defined scope to succeed in competition. This will be the distribution of trustworthy, high quality open data on society combined with in-depth expertise in utilizing such data in journalistic work from raw data to final conclusions and communication. Building transparent and reproducible analysis pipelines from raw data to analyses and visualizations will create further

opportunities for crowd-sourcing and gamification, which we will however skip in the first stage.

We will intentionally avoid focusing on specific platforms since we are not providing a web service per se. Instead, the expertise and openly downloadable data resources are the key service. We will distribute the data in commonly accepted formats such as CSV and XML and we will guarantee these are easily accessible and downloadable from the server but otherwise the technical implementation details are a side issue from customers' point of view. In the longer term we are investigating opportunities to build completely transparent analysis pipelines and tools - including data and analysis code - in cloud servers. This would bring the service at a new level but will remarkably benefit from practical experiences gathered through a simpler modular approach where the data and analysis code form separate entities that support expert analyses.

Protomedia Uutisraivaaja-kisaan

Uutisraivaajaskaba http://uutisraivaaja.fi/index.html

- Kilpailun kriteerit
- Hakijan tiedot
- Projektin kuvaus
- Visualisointi
- Kohderyhmä

Information of the applicant

Project name: Protomedia
Applicant: Open Knowledge Finland ry. (Y-tunnus: 2520401-8)
Address: Annankatu 9 A 11, 00120 HELSINKI
Tel: +358 40 5655 872 (Leo Lahti); +358 44 337 5439 (Antti Poikola)
Email: leo.lahti@iki.fi (Leo Lahti); antti.poikola@gmail.com (Antti Poikola)

Team: members, ages, titles and employers. List the strengths and weaknesses.

- Antti Poikola, 33, M.Sc.(Tech.), open data expert and data journalism lecturer. Otavan opisto and Hila Open Oy, Open Knowledge Finland ry. founding member (http://about.me/apoikola)
- Leo Lahti, 33, D.Sc.(Tech.), Academy of Finland postdoctoral research fellow and open data advocate; applied data analysis; University of Helsinki, Finland; Wageningen University, Netherlands (http://www.iki.fi/Leo.Lahti); Louhos project (louhos.github.com).
- Juha Yrjölä, 32, software architect; Code4Europe-fellow, Municipality of Helsinki (http://www.codeforeurope.net); Kansan muisti ry parliamentary monitoring organization founding member and chair (http://www.kansanmuisti.fi)
- Joona Lehtomäki, 30, M.Sc., Researcher, Finnish Environmental Institute (SYKE), Kansan muisti ry board member ; Louhos-project (louhos.github.com)
- Teemo Tebest, 28, M.Sc.(Tech.), Data journalist, Yleisradio (http://teelmo.info/)
- Juuso Parkkinen, 28, M.Sc.(Tech.), Researcher; applied data analysis, Aalto-university (http://ouzor.github.com) Louhos-project (louhos.github.com)

The team has strong experience in developing solutions and consulting in open data and data journalism scene in public bodies, companies and academic research as well as civil society. The previous works by the team members have received national awards in Apps4Finland 2011-2012. In the area of data journalism the applicants have collaborated with media companies including Helsingin Sanomat, Yleisradio and Otavamedia. Two team members have an entrepreneurship background and the team has strong national and international networks in information-intensive fields. The team was formed during the Apps4Finland 2012 competition. The versatile backgrounds of the team members bring complementary skills. An investigative journalist is missing from the team. The geographical distribution of team members to Helsinki, Jyväskylä and Wageningen is at the same time a challenge for coordination and an asset for collaborations. The service concept utilizes fully open data and code resources, which is new and innovative but also less well tested in practice.

Project summary

The rise of data-driven journalism is a central factor in the changing media landscape. High quality data journalism is one way to distinguish a media company from the rest in an era where superficial news are easily accessible to everyone. However, to produce high quality journalism and to stand out requires specific resources and skills to preprocess, analyze and visualize data. The expertise of a dedicated Datadesk service is something only the largest media companies can afford to establish themselves. Protomedia brings such expertise available for all media companies regardless of their size and budget.

Our central offering is Datadesk expertise coupled with open access to high quality societal data in an easy-to-use format. The Datadesk will provide analyses, support, and education regarding

the use of publicly available information sources to support data-driven exploration and storytelling. The data created and cleaned through commercial activities will be opened for everyone, accumulating and enriching the societal open data resources. This will facilitate transparency, reproducibility and crowd-sourcing - central factors for producing reliable journalism and key features in Protomedia's data management and process chain from raw data to publication.

Visualization

http://www.slideshare.net/apoikola/protomedia-en
https://dl.dropbox.com/u/2949803/protomedia/protomedia_EN.pdf

Customers

The services are primarily targeted at media companies and journalists who cannot maintain their own Datadesk and need trustworthy technical expertise and infrastructure to carry out data journalistic tasks. Datadesk refers to sufficient technical infrastructure and skills for data analysis and visualization in the context of media and news applications. The accumulating open knowledge base and analysis tools will be also available for researchers, organizations and civil society. This allows transparent, evidence-based analyses and helps to increase the impact of journalistic work and gather valuable feedback from wider audience to improve the data and code sharing infrastructure.

External services

Data from various sources is collected and preprocessed to easy-to-use format. Examples of open data sources that currently require considerable preprocessing before journalistic analysis include, but are not limited to:

- **Governmental organizations' websites** contain rich sets of journalistically relevant content including information on election candidates personal information and funding sources (http://192.49.229.35/). Further data sets will include data on voting results in the Finnish Parliament in 2013 (now available from biomi.org http://biomi.org/eduskunta/eduskunta.html), environmental institutes (http://wwwp2.ymparisto.fi/scripts/oiva.asp), National Institute for Health and Welfare (http://uusi.sotkanet.fi/portal/page/portal/etusivu), Statistics Finland (http://www.stat.fi/) etc. The amount of open government data is rapidly increasing but much of this information is not in a readily usable format and requires considerable preprocessing. During 2012 municipal elections, for instance, we preprocessed and shared the election candidate data from Ministry of Justice (http://louhos.wordpress.com/2012/09/28/datavaalit-oikeusministerion-vaalitulosdata-sorvattu-auki/) and time series from 2004-2012 municipal elections based on open data from Statistics Finland

(http://louhos.wordpress.com/2012/10/05/kunnallisvaalien-vertailukelpoiset-ehdokasdatat-csv-taulukkoina-2004-2008-2012/).

- **Social media feeds** (Facebook, Twitter, etc). The team has created a social media aggregate to collect social media feeds of election candidates in time series (http://www.datavaalit.fi/candidates/social/). This allows focused quantitative and semantic analyses of decision-makers commentaries and ensures the availability of historical data in long term. At the moment the aggregate has half million updates from thousands of candidates in social media.

- **Election machines** Media companies collect information on election candidates' and voters' political views using election machines and often openly share this data. At best, this can make possible the analysis of the political views hundreds of thousands of voters and candidates, as in our previous blog post based on Helsingin Sanomat data, spring 2012 http://blogit.hs.fi/hsnext/helsingin-sanomat-julkaisee-vaalikoneen-tiedot-avoimena-rajapintana. Combining the information across multiple election machines at national and international level and with socio-economical information would enrich these data sources but requires further work.

Other relevant external services:

- **Data catalogues** We will utilize and contribute to relevant data catalogues that provide comprehensive information on available open data, including Helsinki Region Infoshare (http://www.hri.fi/fi), Jyväskylä (http://data.jyvaskyla.fi/) Finland (http://data.suomi.fi), and DataHub Finland (http://fi.thedatahub.org/).

- **Github** Data and source code resources from the project are shared through Github (github.com).

Components that have already been established

The Protomedia technical infrastructure takes advantage of the team members previous code and analysis methodologies and harmonizes these into a single service package. The key componentes include:

Database Protomedia will develop a new database based on www.datavaalit.fi project, which already now provides open data on election candidates and their funding. The project was awarded the in Apps4Finland 2012 competition. The team can also utilize the database and development resources of Kansan muisti ry (http://www.kansanmuisti.fi). The current database resources provide a solid basis for developing Protomedia service concept and sharing infrastructure further.

Source code base Protomedia maintains a comprehensive source code base that covers the complete analysis pipeline from raw data to final visualization. This allows rapid, reproducible and transparent analyses of open data streams based on the Louhos software libraries (http://louhos.github.com/sorvi), double-winner in Apps4Finland 2011. Documented examples are collected in Datawiki (http://louhos.github.com/datawiki/), an Apps4Finland 2012 winner.

Collaboration networks and consulting services The long-term experience of the team members on open data, open source, and open science has created comprehensive national and international collaboration networks. A remarkable portion of the data journalistic consultation services in Finland so far has been provided by the team members. Now the demanded expert services will be made available for a wider audience.

Innovativeness

The central innovation of Protomedia is to accumulate a sustainable open knowledge base of data, code and other material produced in data journalistic applications. This helps to avoid duplication of work, lowering the costs of journalistic work, and to gather feedback from the public. Decreased work load on technical processing side will help the journalists to focus on the actual analysis and journalistic tasks. While the principal assett is the expertise and skills of the working team, the open knowledge base will support the service concept and creates added value through transparency, reproducibility and crowd-sourcing opportunities that have been demonstrated in the teams previous projects (see above). The Protomedia enterprise will revolutionize public data availability through trustworthy, easy-to-use open data libraries and associated analysis resources. Transparency of the data and code is central for ensuring the reliability and trustworthiness of the service and follows good practices developed in data-intensive scientific disciplines. The project will form a public library of open data, and the Protomedia team will possess comprehensive in-depth skills on utilizing these resources as part of a wider set of expertise in the fields of data analysis, data journalism, and consulting. Keeping Protomedia as a private enterprise will ensure agility while at the same time the advantages of scale will differentiate it from freelancers and ensure credibility.

Competitors

Freelance journalists / data analysts can provide similar analysis and education services with potentially reduced fixed costs. The added value provided by Protomedia comes from the volume of accumulated in-house expertise that can be utilized across projects. This will improve efficiency and reduce costs. As the data resources are maintained by Protomedia, this will guarantee the trustworthiness of our service and authenticity of data as well as in-depth knowledge of the available data resources, which is crucial for high quality data-driven journalism. We have here a competitive advantage over independent freelancers. However, we are planning to collaborate with freelancers to help bring together supply and demand for data journalistic expertise.

STT (Suomen tietotoimisto) **"**produces reliable information to help you make your own conclusions and form your own opinions". While STT provides data for journalistic work, our concept is focused on the expertise of utilizing available data sources in journalistic work. In addition, we provide preprocessed societal and governmental raw data that is not available from other sources including STT. In the long term, our service concept can be extended to a cloud/eWork ecosystem where the data, supply, and demand for analysis can meet efficiently.

Infogra.am (http://infogr.am/) *"is an easy to use online service that lets you create, share, discover infographics and online charts."* Various tools are available for data visualization. Protomedia complements these services by providing easy-to-use data on various aspects of society. High-quality data provided by Protomedia is a necessary starting point for further analysis and visualization. Moreover, additional data analysis consultation is often needed and this is also available provided Protomedia.

Tableau Public (public.tableausoftware.com) provides free software-as-a-service concept to visualize and share data through embedded applications. The service has been used by Guardian, Wall Street Journal and Helsingin Sanomat. High-quality data provided by Protomedia is a necessary starting point for further analysis and visualization. Moreover, additional data analysis consultation is often needed and this is also available provided Protomedia.

Sovelto (http://www.sovelto.fi) is a Finnish ICT instructor company in the application development, infrastructure, communications, security and information managements. Protomedia provides data journalism education to complement the data journalitic service concept but does not aim to be a general ICT education supplier. Protomedia focuses on data journalism and open data.

Sourcefabric (http://www.sourcefabric.org/) provides Superdesk software platform for media companies and journalists. *"Open software for managing newsrooms and their content"*. Protomedia is not a software company although we maintain strong open source code base for open data processing purposes. This supports the Protomedia service concept that is primarily targeted to Datadesk education, consultancy, data analysis and other expert services.

FloApps (http://www.floapps.com) provides website development and open data software development services. In contrast, Protomedia provides data, education and expert services. While several companies provide software, the availability of easy-to-use data and analysis services are less widely available. Protomedia provides a service concept targeted to the needs of media companies and journalists.

Collaboration

Open Knowledge Foundation Finland ry. (http://fi.okfn.org/). Protomedia accounting will be handled through OKF Finland ry. OKF Finland promotes the use and availability of open data in Finland. Protomedia service concept supports this objective and can in turn benefit from the

well-established international networks of the foundation.

Kansan muisti ry. (http://www.kansanmuisti.fi) parliamentary monitoring organization - whose board members are included in Protomedia team - has resources and tools for database development and maintenance that can be utilized by Protomedia.

Louhos (http://louhos.github.com) open data analytics project is a community-driven open source project established by Protomedia team members. The project transfers data analysis methodologies from data-intensive scientific disciplines to the domain of open government data and was awarded in Apps4Finland 2011 competition. The methodological resources and scientific expertise of the team can be utilized by Protomedia. Further collaboration with Finnish research groups in applied data analysis and political and social science has been discussed.

Data Hub Finland (fi.thedatahub.org) maintains comprehensive data catalogue for Finnish open data. Protomedia will utilize this valuable service and also contribute to developing it further. The connections with Data Hub maintainers are already well established.

National data openers Protomedia team has wide networks to data openers, which is central for expansion of the service. These include for instance Helsinki Region Infoshare, Statistics Finland, Land Survey Finland - personal connections to these organization are already well established. The team members have also been involved at a regular basis in discussions on opening governmental data sources with certain ministries, the Finnish parliament, Kuntaliitto, Helsinki authorities etc.

Media companies Contacts to main media companies including YLE and HS have been established. While no immediate collaboration plans exist, personal connections and previous collaboration with major national media companies will be a valuable assett for Protomedia.

Contribution to Finnish journalism

With increasing open data streams nationally and internationally, data journalism is facing new opportunities to integrate and utilize data to produce high-quality journalism. A central bottleneck in data-driven journalism has been the lack of resources and skills, in particular associated with data processing and analysis. Protomedia will solve this by bringing the necessary services to the reach of all media actors. Externalization of data processing will reduce the total journalistic workload and centralizing the resources will improve the efficiency and quality of using the resources. A central principle is open sharing of data and analysis code. This will increase the quality of Finnish journalism as the journalists will have more data available, and they can use more of the working times to the actual high-quality journalistic analysis rather than technicalities. This will also allow Finnish media to compete with the international competitors that have become more and more popular among Finnish readers.

Vision

With increasing fluxes of data, superficial reporting is becoming more and more ubiqitous and individual media will have increasing difficulties to distinguish themselves from the mass. On the other hand high-quality investigative journalism has demand and data journalistic resources and expertise can provide tools to support these goals and to provide ingredients for interesting and comprehensive, evidence-based analyses in future journalism and knowledge distribution. Protomedia provides services for media actors of all sizes from individual journalists to large media companies, equalizing the opportunities for high-quality data-driven journalism, first at national and later at international level. The Protomedia enterprise will revolutionize public data availability through trustworthy, easy-to-use open data libraries and associated analysis resources.

Strategy and business plan

Our business logic is based on the well-established observation that main revenue in open source projects comes from expert services, not the software itself. In our case, the open availability of the project resources will bring added value through increasing the trustworthiness and impact of journalistic work. We will pilot the business model using our established network and media company connections. We do not need big initial investments, so we can start with small scale pilots and scale up and tune our business model gradually. In the first phase our customers will be media companies of various sizes. However, we actively follow the development of the open data scene and especially the potential business opportunities there. The kind of Datadesk service we are offering for media companies would probably have demand also outside media industry in both private and public sector. For example, both consulting companies and public organizations prepare tons of reports and surveys that would benefit from easily accessible open data and analysis and visualization tools.

Time schedule, budget, and business logic

Duration (months): 24 months

Annual budget: 158,500 Eur on average, development and costs focus more on the first year, revenues focus more on the second year.

Total budget : 317,000 Eur

.

Money applied from Uutisraivaaja: 240,000 Eur

Other funding: Protomedia is based on Datavaalit project which received Sitra New

Democracy seed funding in 2012. The remaining funds (5000 Eur) will be used in entirety to the development of Protomedia. Funding situation at the moment:

- Received Sitra funding 12,000 Eur
- Selling revenue from data journalism education services: 3800 Eur
- Apps4Finland awards 2012: 1200 Eur
- Used: 12,000 Eur
- Remaining: 5000 Eur

Funding plan during Uutisraivaaja

Protomedia will be funded by providing Datadesk expert services and data journalism education. The major costs come from personnell costs. In addition, we have taken into account office space, server and other ICT resources, marketing and other costs. At the end of the two-year funding period the revenues will cover the personnell and fixed costs.

Budget for Uutisraivaaja funding period:

1. YEAR:

1. YEAR INCOME:		
Service revenues		6,000
Uutisraivaaja funding		165,000
Datavaalit project funding		5,000
TOTAL INCOME		176,000

1. YEAR COSTS:		
Salary + side costs	150,000	
Office space	7,500	
ICT	1,800	
Marketing	7,500	
Other	4,500	
TOTAL COSTS		171,300

2. YEAR:

2. YEAR INCOME:	
Services	66,000
Uutisraivaaja funding	75,000
TOTAL INCOME	141,000

2. YEAR COSTS:	
Salary + side costs	120,000

Office space	6,000	
ICT	3,600	
Marketing	9,500	
Other	6,600	
TOTAL COSTS		145,700

Funding after Uutisraivaaja challenge

With Uutisraivaaja funding we can establish the necessary data and code infrastructure that is needed to support the expert education and consultation services. After that the project will be able to cover its own costs and core functionality can be developed and maintained with the revenues from new customers once the basic infrastructure has been set up.