



**QLectives – Socially Intelligent Systems for Quality
Project no. 231200**

**Instrument: Large-scale integrating project (IP)
Programme: FP7-ICT**

**Deliverable D3.1.3
[Final techno-social “living archive”]**

Submission date: 28-02-2013

Start date of project: 2009-03-01

Duration: 48 months

Organisation name of lead contractor for this deliverable: UniS

Project co-funded by the European Commission within the Seventh Framework Programme (2007-2013)		
Dissemination Level		
PU	Public	x
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Document information

1.1 Author(s)

Author	Organisation	E-mail
Steve Genoud	ETHZ	sgenoud@ethz.ch
Stefano Baliatti	ETHZ	stefano.baliatti@gess.ethz.ch

1.2 Other contributors

Name	Organisation	E-mail
Nigel Gilbert	UniS	N.Gilbert@surrey.ac.uk
Maria Xenitidou	UniS	M.Xenitidou@surrey.ac.uk

1.3 Document history

Version#	Date	Change
V0.1	14 December 2012	Starting version, template
V0.2	18 January 2013	Revised version
V1.0	18 March 2013	Approved version to be submitted to EU

1.4 Document data

Keywords	Quality Collectives, Project summary, Results, Conclusions
Editor address data	n.gilbert@surrey.ac.uk
Delivery date	31/08/2012

1.5 Distribution list

Date	Issue	E-mail
	Consortium members	QLECTIVES@LIST.SURREY.AC.UK
	Project officer Roumen Borissov	Roumen.BORISSOV@ec.europa.eu
	EC archive	INFSO-ICT-231200@ec.europa.eu

QLectives Consortium

This document is part of a research project funded by the ICT Programme of the Commission of the European Communities as grant number ICT-2009-231200.

University of Surrey (Coordinator)

Department of Sociology/Centre for
Research in Social Simulation
Guildford GU2 7XH
Surrey
United Kingdom
Contact person: Prof. Nigel Gilbert
E-mail: n.gilbert@surrey.ac.uk

Technical University of Delft

Department of Software Technology
Delft, 2628 CN
Netherlands
Contact Person: Dr Johan Pouwelse
E-mail: j.a.pouwelse@tudelft.nl

ETH Zurich

Chair of Sociology, in particular
Modelling and Simulation,
Zurich, CH-8092
Switzerland
Contact person: Prof. Dirk Helbing
E-mail: dhelbing@ethz.ch

University of Szeged

MTA-SZTE Research Group on
Artificial Intelligence
Szeged 6720, Hungary
Contact person: Dr Mark Jelasity
E-mail: jelasity@inf.u-szeged.hu

University of Fribourg

Department of Physics
Fribourg 1700
Switzerland
Contact person: Prof. Yi-Cheng Zhang
E-mail: yi-cheng.zhang@unifr.ch

University of Warsaw

Faculty of Psychology
Warsaw 00927, Poland
Contact Person: Prof. Andrzej Nowak
E-mail: nowak@fau.edu

Centre National de la Recherche Scientifique, CNRS

Paris 75006,
France
Contact person : Dr. Camille ROTH
E-mail: camille.roth@polytechnique.edu

Institut für Rundfunktechnik GmbH

Munich 80939
Germany
Contact person: Dr. Christoph Dosch
E-mail: dosch@irt.de

QLectives introduction

QLectives is a project bringing together top social modelers, peer-to-peer engineers and physicists to design and deploy next generation self-organising socially intelligent information systems. The project aims to combine three recent trends within information systems:

- **Social networks** - in which people link to others over the Internet to gain value and facilitate collaboration
- **Peer production** - in which people collectively produce informational products and experiences without traditional hierarchies or market incentives
- **Peer-to-Peer systems** - in which software clients running on user machines distribute media and other information without a central server or administrative control

QLectives aims to bring these together to form Quality Collectives, i.e. functional decentralised communities that self-organise and self-maintain for the benefit of the people who comprise them. We aim to generate theory at the social level, design algorithms and deploy prototypes targeted towards two application domains:

- **QMedia** - an interactive peer-to-peer media distribution system (including live streaming), providing fully distributed social filtering and recommendation for quality
- **QScience** - a distributed platform for scientists allowing them to locate or form new communities and quality reviewing mechanisms, which are transparent and promote quality.

The approach of the QLectives project is unique in that it brings together a highly interdisciplinary team applied to specific real world problems. The project applies a scientific approach to research by formulating theories, applying them to real systems and then performing detailed measurements of system and user behaviour to validate or modify our theories if necessary. The two applications will be based on two existing user communities comprising several thousand people - so-called "Living labs", media sharing community tribler.org; and the scientific collaboration forum EconoPhysics.

Executive Summary

This deliverable, 'D3.1.3 Final techno-social "living archive"', describes a site that offers a place on the Internet where anyone, but mostly social scientists, can share their data. This living archive has been seeded with data from the QLectives project. As part of the deliverable, two "Living Datasets" are also provided. The aim is to provide a place for users to exchange valuable information and encourage individuals and groups to publish new papers.

Contents

1. Introduction	1
2. The Living Archive.....	2
3. Development Attempts.....	2
4. Functionality.....	2
5. Exploration	3
6. Adding a new dataset	4
7. Regular Updates	5
8. Statistics and new trends.....	5
9. Future Development	5
10. Technical aspects	6
11. Datasets Open Data Sources	6
12. QLectives Data.....	7

1. Introduction

Techno-social research produces an important number of datasets. Reproduction of results is crucial to good science. There is a need for the data to be available to other scientists.

The challenge is twofold. First, it is difficult for scientists to share their own data. While many tools exist (for instance FTP servers), they need specialized knowledge to setup. Second, it is difficult to find the data of other scientists. Datasets are scattered on university servers and are difficult to find even with the powerful search engines available today.

This deliverable proposes to offer a place on the Internet where anyone, but mostly social scientists, can share their data. This living archive has been initialised with data from the QLectives project. Also we provide two "Living Datasets", data that will be updated with time.

We endeavor to provide a place for users to exchange valuable information and encourage individuals and groups to publish new papers based on information and data found on the site.

2. The Living Archive

The Living Archive aims to be a place online that both aggregates already available data, and provides an easy way for new data to be offered.

Through the use of available technology it is possible for us to build and maintain a comprehensive platform that truly provides a one-stop place for data. The Living Archive does not store all data on its own servers, but simply acts as a directory having data in its repositories and links it to the relevant source. The goal hereby clearly being, to create and maintain a website that aggregates as many open data sources as possible in order to promote research.

The community has not seen a platform rivaling the concept of Living Archive and we are confident that we will achieve our goals.

3. Development Attempts

Previously we attempted to build the platform with the open source CKAN (<http://ckan.org/>) software. A few weeks into development we realized it was an inefficient process and that we would be better off building it from scratch. When we stopped using CKAN (and stopped spending most of our time fixing CKAN bugs) we had more time to dedicate towards rebuilding the platform the way we wanted it.

4. Functionality

Users have the ability to register new datasets or to collaborate on existing ones (adding ideas, comments, own data visualizations or/and links). We will gradually add more community functions as the project continues.

As the project stands at this point it boasts extensive features giving users the ability to search and interact with datasets within the platform.

Possible new functions include more social network interactions.

5. Exploration

Exploring the Living Archive is quite simple. You can use the search field to look for data linked to your subject of interest. It is also possible to use tags as well.

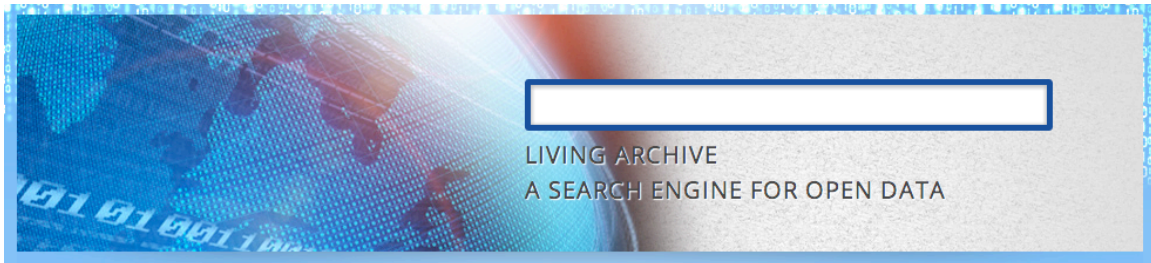
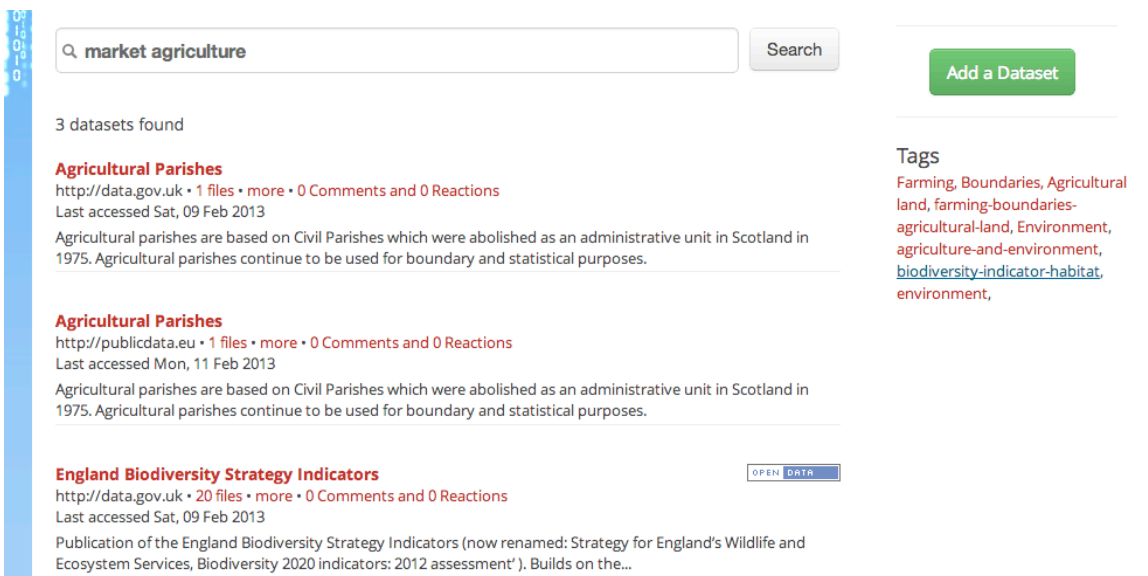


Figure 1: Explore the Living archive with search



Q **market agriculture** Search

Add a Dataset

3 datasets found

Agricultural Parishes
<http://data.gov.uk> • 1 files • more • 0 Comments and 0 Reactions
 Last accessed Sat, 09 Feb 2013
 Agricultural parishes are based on Civil Parishes which were abolished as an administrative unit in Scotland in 1975. Agricultural parishes continue to be used for boundary and statistical purposes.

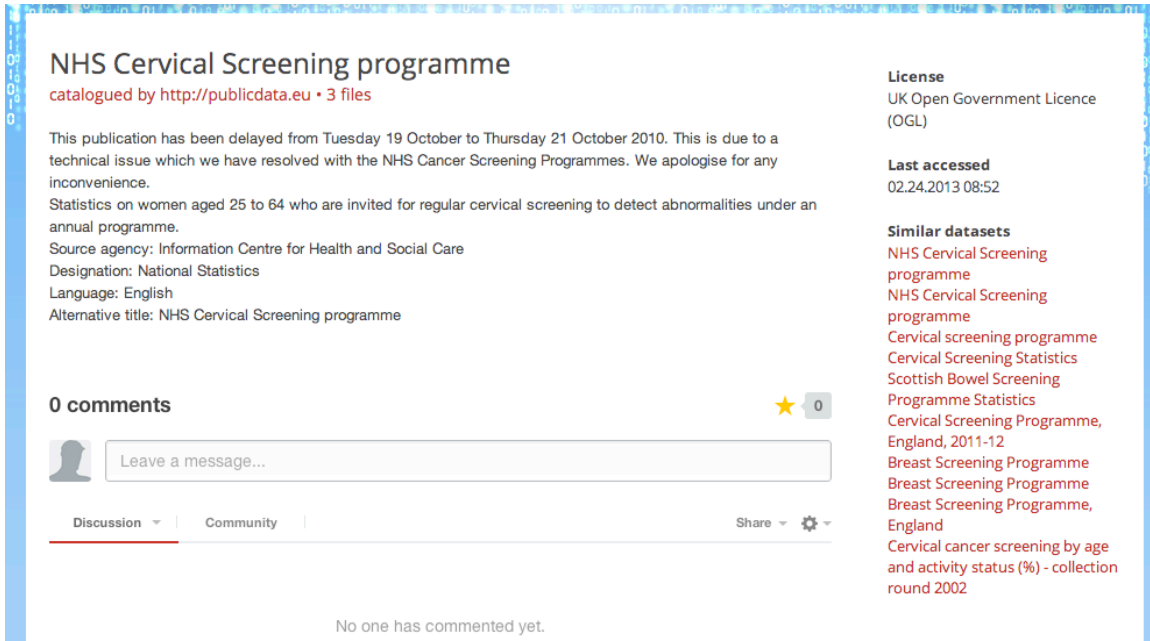
Agricultural Parishes
<http://publicdata.eu> • 1 files • more • 0 Comments and 0 Reactions
 Last accessed Mon, 11 Feb 2013
 Agricultural parishes are based on Civil Parishes which were abolished as an administrative unit in Scotland in 1975. Agricultural parishes continue to be used for boundary and statistical purposes.

England Biodiversity Strategy Indicators OPEN DATA
<http://data.gov.uk> • 20 files • more • 0 Comments and 0 Reactions
 Last accessed Sat, 09 Feb 2013
 Publication of the England Biodiversity Strategy Indicators (now renamed: Strategy for England's Wildlife and Ecosystem Services, Biodiversity 2020 indicators: 2012 assessment'). Builds on the...

Tags
 Farming, Boundaries, Agricultural land, farming-boundaries-agricultural-land, Environment, agriculture-and-environment, biodiversity-indicator-habitat, environment,

Figure 2: Explore the Living archive with search / results

Clicking on any dataset will open a page with metadata and links to the dataset. On this view it is possible to comment on individual datasets.



NHS Cervical Screening programme
catalogued by <http://publicdata.eu> • 3 files

This publication has been delayed from Tuesday 19 October to Thursday 21 October 2010. This is due to a technical issue which we have resolved with the NHS Cancer Screening Programmes. We apologise for any inconvenience.

Statistics on women aged 25 to 64 who are invited for regular cervical screening to detect abnormalities under an annual programme.

Source agency: Information Centre for Health and Social Care
Designation: National Statistics
Language: English
Alternative title: NHS Cervical Screening programme

0 comments ★ 0

Discussion | Community | Share ⚙

No one has commented yet.

License
UK Open Government Licence (OGL)

Last accessed
02.24.2013 08:52

Similar datasets

- NHS Cervical Screening programme
- NHS Cervical Screening programme
- Cervical screening programme
- Cervical Screening Statistics
- Scottish Bowel Screening Programme Statistics
- Cervical Screening Programme, England, 2011-12
- Breast Screening Programme
- Breast Screening Programme
- Breast Screening Programme, England
- Cervical cancer screening by age and activity status (%) - collection round 2002

Figure 3: Comment and Metadata of dataset

Also, the Living Archive has been integrated with Living Science. It offers an automatic link to research articles linked with the topic. This allows the user to explore the topic further, outside of the Living Archive; hence we accomplish part of the objective of encouraging users to research further.

6. Adding a new dataset

Adding a new dataset is easy. After clicking on “Datasets” in the main menu you will see the “Add a Dataset” button. This will bring you to a page where you can edit the basic properties and metadata of your new dataset. It is important when you add a new dataset to spend some time formulating the description of your new dataset. It will make it easier to be found by other users. You can finally edit these fields (and more) by going into the *Settings* menu of a dataset. All your modifications will be saved and can be reverted. This allows anyone to improve the metadata of all the dataset, while being able to fight against vandalism in a similar fashion to what is done within Wikipedia. By allowing this type of editing to be done by users we intend to create a self-regulating community with quality information.

7. Regular Updates

The Living Archive platform is equipped with scripts that regularly update the repositories of datasets. At a later stage users will be able to subscribe and be notified of new updates relevant to their preferences. The scripts also verify the state of the datasets and update them accordingly. This adds to the quality of the platform as well as it shows users that the platform is genuinely committed to providing up to date information; thus encouraging revisits. We believe that, when building a platform such as we are, it is vital to make a good and quality impression on the users. We can guaranty quality by ensuring that these scripts run regularly and can offer our user community quality up to date content.

8. Statistics and new trends

Administratively we collect user queries and use them in order to create statistics and detect new trends. This is invaluable as it also allows us to constantly improve the site based on user interaction with the site. Besides this we keep the statistics to measure our business expectancies on the platform. Constantly analyzing the results shows us if we have a steady and growing user community for the platform.

9. Future Development

The Living Archive is a community driven platform and cannot be in a state where it is “finished”. From a development point of view we can agree on a certain version that we are satisfied with, but we do aim to have an enthusiastic community that maintains and uses data on a regular basis. The purpose of the site is to allow users to find data and the platform should not be seen as a finished product, but truly a platform that is alive in that it is constantly evolving. Evolving not just in the form of technicalities and functionalities, but in that the quality of the datasets is constantly improved by the users themselves. The platform should gain more public awareness as the repositories grow and allow more users to find more information; thus becoming what we aim for it to be, the one stop place for datasets.

10. Technical aspects

We stopped using CKAN because it was starting to get unmaintainable and had many bugs that required a lot of time trying to fix. So we spent that time in a better way by rebuilding the platform. We did it using CakePHP as a framework running on a Apache2, PHP5, MongoDB stack. We chose MongoDB because the schema of our datasets varies a lot and we don't want to constrain it at such an early stage. The search feature now uses an Elasticsearch index, that allows calls via REST API. Elasticsearch integrates perfectly into our stack and can index MongoDB directly. The user is now able to login via a Single Sign On service provided by the Innovation Accelerator and can profit from various other services without registering for multiple accounts.

As for hosting, we use an elastic Linux server in a cloud at Rackspace to be able to scale quickly should the site suddenly have high traffic.

11. Datasets Open Data Sources

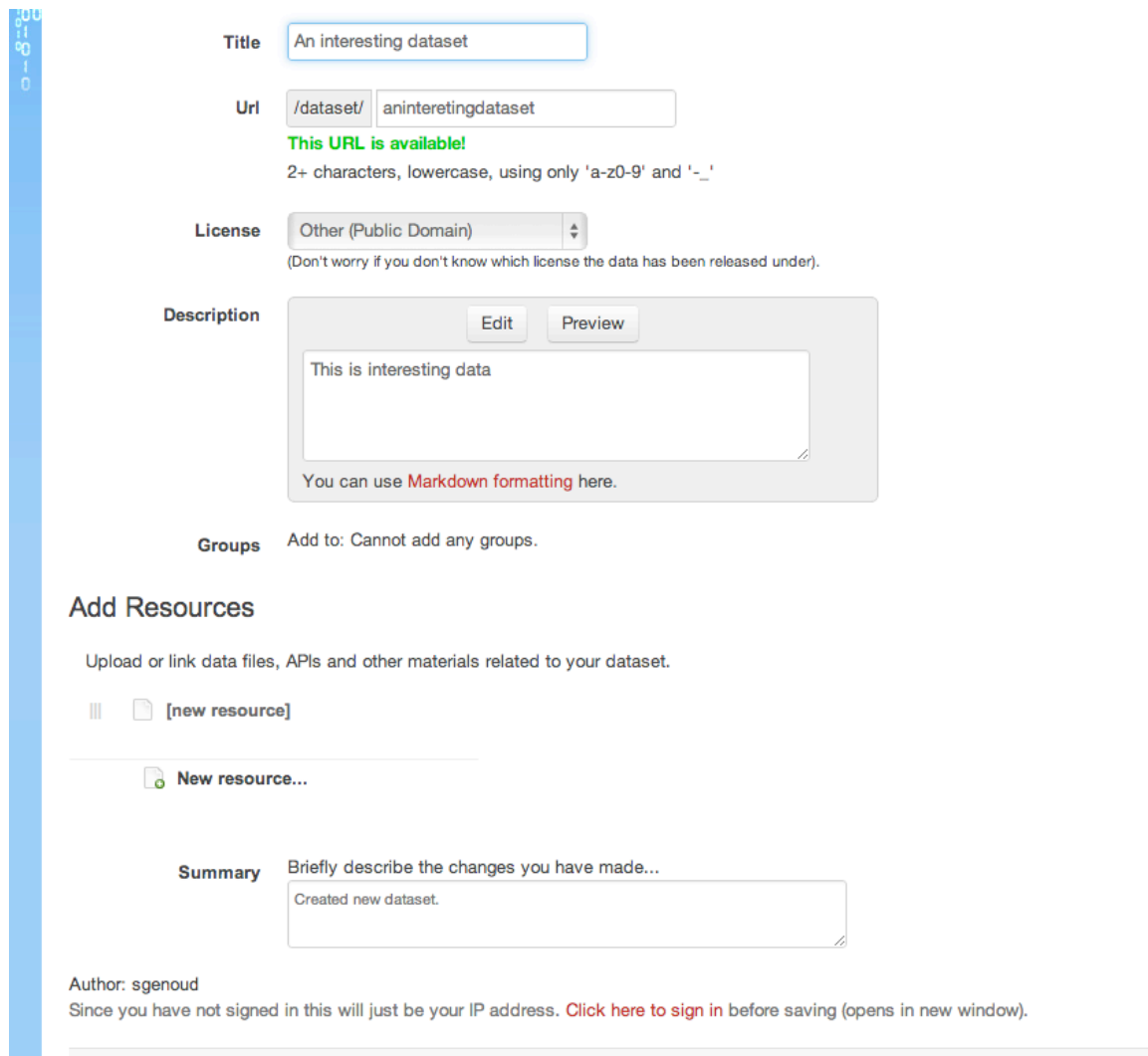
The academic community is not the only one benefiting from more transparent and available data. Governments have started publishing their data under an open license for some years now. First, this is a good source of inspiration for scientists who should share their data more. Then, and most importantly, this open government data can be used to analyze our society. We have therefore gathered open data from different sources. Open data is licensed in such a way, that we are allowed to copy and share it freely. At the end of 2012, our open data repository contained more than 900 datasets. To date (27.2.2013) we have acquired 41,412 Datasets (currently averaging at ca. 100 new datasets a day).

As we offer a "living" archive, we have put in place a system that will retrieve new open data at regular intervals. Social scientists looking for new datasets will therefore be able to regularly go to one place, the Living Archive, to import new data. Later functionalities may include offering email notifications when new datasets relevant to specified interests are found.

12. Q Lectives Data

Living Datasets: QScience

QScience allows users to easily collect and share publication references. Users can upload PDF directly to their own QScience instance, or references can be imported from the main QScience data provider: Living Science (<http://www.livingscience.ethz.ch>).



The screenshot shows a web form for uploading a dataset. The form includes the following fields and sections:

- Title:** A text input field containing "An interesting dataset".
- Url:** A text input field containing "/dataset/ aninterestingdataset". Below it, a green message says "This URL is available!" and a note states "2+ characters, lowercase, using only 'a-z0-9' and '-_'".
- License:** A dropdown menu set to "Other (Public Domain)". A note below says "(Don't worry if you don't know which license the data has been released under)."
- Description:** A text area containing "This is interesting data". Above the text area are "Edit" and "Preview" buttons. Below the text area is a note: "You can use **Markdown** formatting here."
- Groups:** A section with the text "Add to: Cannot add any groups."
- Add Resources:** A section with the text "Upload or link data files, APIs and other materials related to your dataset." Below this is a list of resources, including a "[new resource]" button and a "New resource..." button.
- Summary:** A text area with the prompt "Briefly describe the changes you have made..." and the text "Created new dataset." inside.
- Author:** The text "Author: sgenoud" is displayed.
- Footer:** A note says "Since you have not signed in this will just be your IP address. [Click here to sign in](#) before saving (opens in new window)."

Figure 4: Uploading data

Statistics on user search queries, downloads and ratings are generated by QScience instances and by the Living Science server. This data will be made available to the Living Archive server, and periodically updated.

An extended description of QScience v.4 is available in deliverable D4.2.4 “QScience v4”, due in month 48.

1 comment

★ 0 Stars



Leave a message...

Discussion

Community



Sam Sulaimanov · a month ago

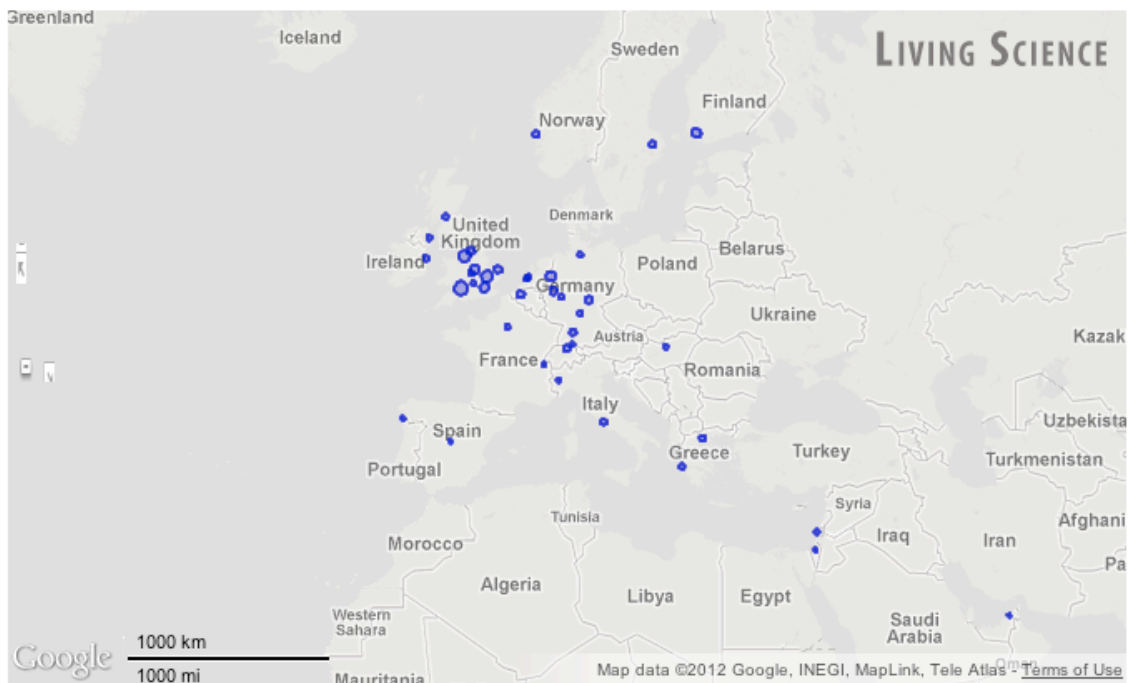
makes me want to build a sensor network to capture own data -

https://cosm.com/explore/air_q...

0 ^ | v · Reply · Share

Powered by **DISQUS**

Related publications



(2008): Twenty-five years of nationwide ambient metals measurement in the United Kingdom: concentration levels and trends.

Brown, Richard J. C., Yardley, Rachel E., Muhunthan, Dharsheni, Butterfield, David M., Williams, Melanie, Woods, Peter T., Brown, Andrew S., Goddard, Sharon L.

(2010): Air Quality Appraisal – Valuing Environmental Limits Interdepartmental Group on Costs and Benefits.

Figure 5: Living Science integration with Living Archive