Web-COSI
Web Communities for
Statistics for Social Innovation
INCREASING TRUST IN COLLECTIVELY
GENERATED STATISTICS

Web-COSI "Web COmmunities for Statistics for Social Innovation"

www.webcosi.eu

SEVENTH FRAMEWORK PROGRAMME
ICT-2013.5.5 Collective Awareness Platforms for Sustainability and Social Innovation
Coordination and support actions (Coordinating actions)

Grant Agreement Number 610422
FP7-ICT-2013-10

Deliverable 3.3

Name of the activity/tool
Wikipedia of progress statistics - redesign of Wikiprogress.stat

Partner in charge OECD

August 2014

This project has received funding from the European Union's Seventh Framework Programme
for research, technological development and demonstration under grant agreement no 610422

Contract Number: 610422

Project Acronym: Web-COSI

Work Package: 3

Deliverable 3.3

Title: Wikipedia of progress statistics -redesign of Wikiprogress.stat

Partner in charge: OECD

Type: Other

Date of delivery: August 2014

Due date of delivery (DoW): August 2014

Author/s of this report:

Salema Gulbahar and Katherine Scrivens (OECD)

Deliverable 3.3

Wiki of progress statistics - redesign of Wikiprogress.stat

Summary

This report describes the decisions and actions taken to develop a new data-sharing portal on the Wikiprogress website: a key deliverable of the Web-COSI CAPS project. The report first describes the reasons the new portal was needed, then sets out the steps taken up to the release of the beta version (made available to the public on Wikiprogress on August 29th), and finally explains the next steps to be taken in the future.

Index

Summary 3

Introduction 5

1. Purpose of the activity 5

2. Set up of the activity and its execution 8

3. The results achieved and the open issues 11

4. Conclusions/Next steps 12

Introduction

The movement for better measures of well-being and social progress that go 'beyond GDP' has been in existence for a decade now. During that time, a large amount of knowledge, research and data has been generated – both in the world of official statistics and government, as well as by more grass-roots, civil society and independent academic organisations. Web-COSI, part of the CAPS2020 initiative, aims to increase public involvement with Beyond GDP statistics, and to find ways in which top-down and bottom-up approaches to measuring progress can be better combined and utilised. Wikiprogress is a Collective Awareness Platform hosted by the OECD, which brings together information and resources related to the measurement of well-being and progress, contributed by an engaged and diverse community of practitioners, researchers and interested non-experts. Improving the data-sharing aspect of Wikiprogress is a central feature of the Web-COSI project, in line with its objectives to improve people's engagement with 'beyond GDP' statistics (ultimately empowering citizens with knowledge to make informed decisions and to be able to monitor well-being and societal progress) and to find ways to better integrate official and non-official measures of well-being and progress.

Wikiprogress has been in existence since 2009, and thanks to its open-source editing software Mediawiki (the same as is used by Wikipedia), it has been able to encourage user contributions in the forms of text articles – with a current archive of over 5000 articles. However, as Wikiprogress is concerned with the *measurement* of well-being and social progress, it has always been the objective of the site to house data, statistics and other statistical resources, as well as text articles. Mediawiki is less suited to the collaborative curation of data and other statistical resources, than it is to text articles. For this reason, statistics on Wikiprogress to date have been housed on a clone of the OECD.stat data warehouse, in a master database entitled Wikiprogress.stat. This solution has been less than ideal, for a number of reasons, including the fact that Wikiprogress.stat uses proprietary rather than open-source software (going against the CAPS ethos); the fact that it is almost impossible for external data providers to upload their own data, thereby relying on OECD staff (and meaning that this part of Wikiprogress is not actually a 'wiki' as it disallows the collaborative creation of content); and the fact that Wikiprogress.stat is not very user-friendly for those who are not used to using OECD.stat.

The new platform based on open-source software, CKAN (by the Open Knowledge Foundation) was developed with a user-friendly interface and functionalities that make it easier to access, share, visualise, upload and interact with information and results (including data and metadata). The platform is open and all data is freely available. The data portal contribute to bring together both official and non-official statistics, into what will eventually be a comprehensive catalogue of data and statistical resources related to the measurement of well-being and progress.

1. Purpose of the activity

1.1 To develop a comprehensive catalogue and 'one-stop-shop' for data and statistical resources related to the monitoring of well-being and societal progress.

The Wikiprogress data portal will become the key reference for progress and wellbeing data and statistical resources such as reports, visualisations and interactive tools and much more.

The portal currently hosts 40 data sets from 32 organisations and will have 60 data sets by the end of 2014, and over 100 by the end of 2015. To date this is the only data portal to host a

comprehensive collection of data and statistical resources to enable one to monitor well-being and societal progress.

1.2 To create a user friendly, open source platform with crowd sourced data and statistical resources.

The portal is intended to be used by researchers, practitioners, policy makers, engaged citizens, app developers, and anyone interested in accessing data on well-being and progress.

In order to engage this wider community, beyond our traditional users, and in particularly citizens, with data on well-being and in the broader conversation on measurement it is key that the core of our work (data and statistics) is accessible and open to all, hence the portal:

- uses open source software, CKAN, which will enable the platform to evolve as developers create and share additional tools and functionalities. A number of governments such as the UK, France and Germany use the CKAN portals to host their data which reflects it's potential.

- has the potential for data and statistical resources to be independently uploaded, shared and discussed by users.

In order for data to be really crowd sourced, we will need to reach a critical mass of dataset and to profile and foster a community of active users. The data visualisation competition, Wikiprogress University and social media campaign will all be used to support this aim.

1.3 To house Open Data.

The site has been developed using open source software, with the aim that the portal will contain increasing amounts of Open Data – data that is *"freely available to everyone to use and republish as they wish, without restrictions from copyright, patents or other mechanisms of control"*[1].

This will require a little time as very few of our current data set are totally 'open'. But we will work with our users and partners to actively encourage and support them to reach this goal.

Organisations will in the first instance be encourage to upload relevant reports, websites, visualisations and other resources that help to make 'beyond GDP' data more accessible to all.

1.4 To be a space to visualize, share and discuss information and results from well-being and societal progress data,

Being able to visualise, tell stories and discuss data is essential if we want to truly improve people's engagement with 'beyond GDP' statistics and to open this discussion up to a broader community (beyond experts).

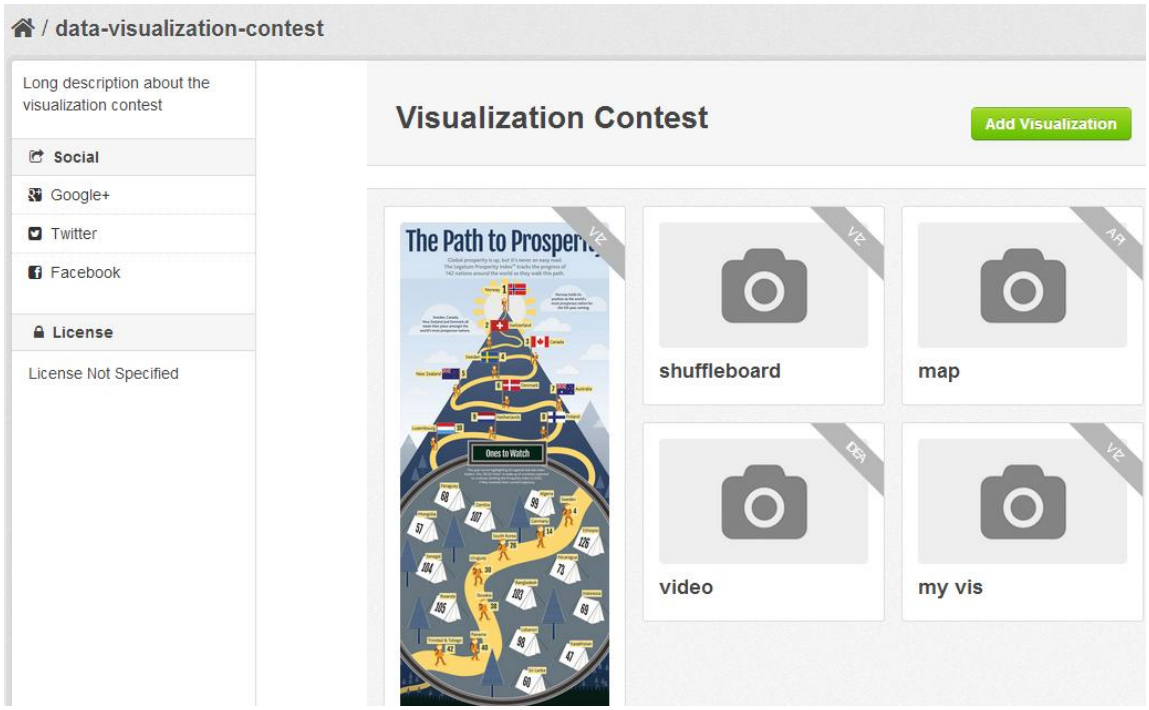
A data visualisation competition and the Wikiprogress University Programme will build upon the resources available in the portal. The competition will require participants to use only data sets that are in the portal or to upload relevant sets that they wish to use – this will introduce potential users

[1] http://en.wikipedia.org/wiki/Open_data

to the portal. The university programme will reach out to students and support them in contributing knowledge and data to the portal.

All the Wikiprogress tools (blogs, Twitter, Facebook and e-Bulletin) and the Wikiprogress platform will be realigned, to feed into the Wikiprogress data portal, which will become a central feature of our activities beyond the Web-COSI project.

A substantial online discussion will take place in the lead up to the 5th OECD World Forum on Statistics and Measuring Progress in October 2015, which will be used to profile the portal and to get feedback from users in order to make the portal as user-friendly and responsive to user needs as possible.



2. Set up of the activity and its execution

2.1 Potential users needs and technical requirements defined

From the initial scoping exercise we knew that the following requirements were essential for the Wikiprogress data portal, much of which DotStat was unable to provide:

- it needed to be more intuitive and easier to use than Wikiprogress.Stat , so that those with little or no knowledge could use the portal,
- be able to upload and share data with no barriers (no permission or formatting of data required) – so data can be crowd sourced,
- a more open source platform that can be changed, improved and built upon,
- a platform that could work with Wikimedia and WordPress and other social media tools, such as Twitter, Facebook and YouTube,
- a portal that uses software that would be relatively cheap to buy, maintain and adapt.

2.2 Platforms and options explored

Various options were explored, against requirements and cost analysis undertaken, below are the three main scenarios:

i) To stay with Wikiprogress.Stat software and attempt to redesign the portal and possibly create a more attractive landing page. This however did not meet most of our needs as outlined above. The DotStat portal is not easy to use, data cannot be crowd sourced and we would not be able to make iterative changes and build upon the portal in the way we would like to.

ii) To work with a partner organisation, that is an expert in managing and visualising data. This was an extremely attractive option as we would not need to be too involved in the details of the structuring and formatting of data. However this was not viable as many organisations had provided us with permission to host data and our ability to integrate a partner's portal with our existing tools and to make changes to the portal would be limited. More importantly it would be far more difficult for us to create a community and monitor our metrics.

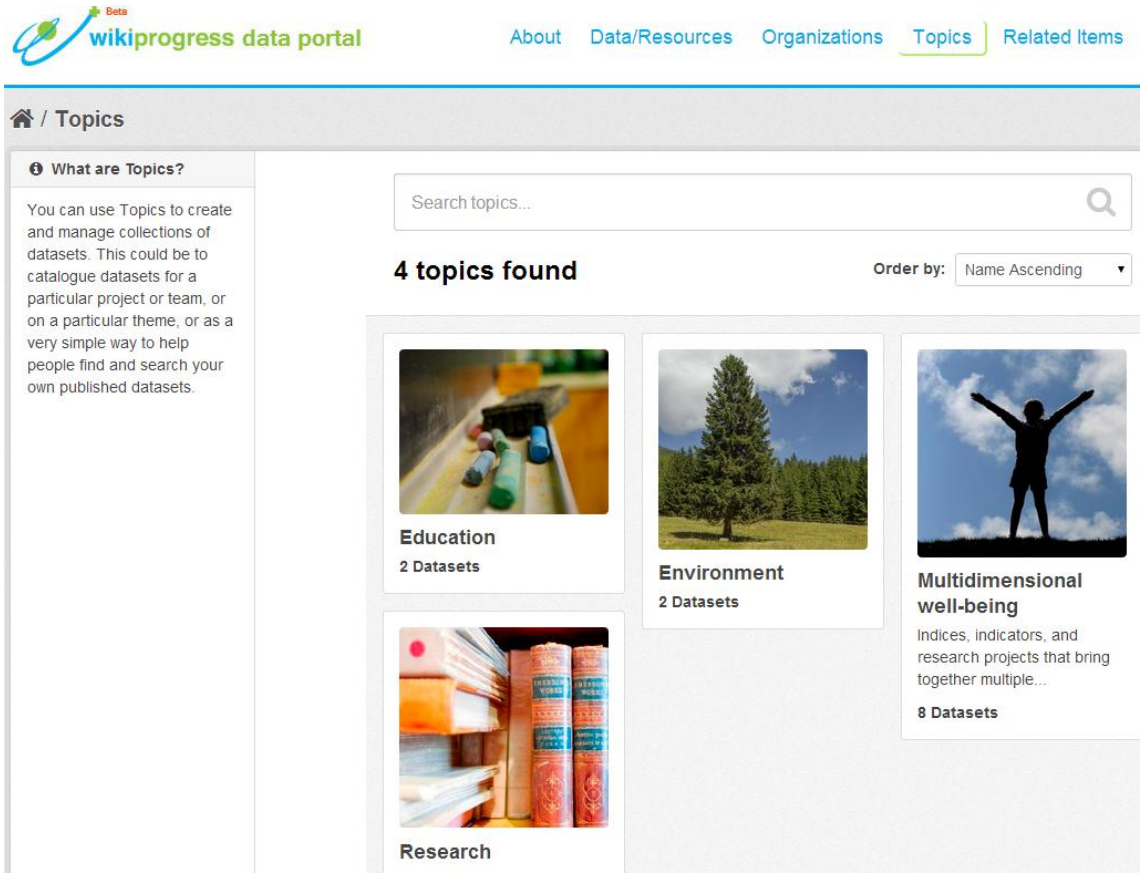
iii) To use an open source data portal, that was becoming increasingly popular way to share data, even by government agencies. This option led us to choose a CKAN portal and the reasons why are outlined below.

2.3 CKAN open-source software chosen

CKAN was chosen because it provided a number of benefits over and above the software that was being used for the Wikiprogress.Stat (DotStat), listed below:

- Altogether much more intuitive and user friendly than DotStat, requires very little to no previous knowledge of data portals.
- Allows discussion and comments on the dataset, thus enabling communities to evolve around data.
- Users can share pages and datasets with their networks through social media and can subscribe for updates and alerts.

- Allows users to upload data in various formats and different structures e. g. html, word, excel, csv, word, pdf etc. The user doesn't need to make any changes to the data they upload, previous DotStat only allowed CSV and excel formats and significant amounts of time was spend formatting and structuring data.
- Allows people to group and categorize datasets by organisations and topics.
- Data can be uploaded very easily and the barrier to uploading significantly reduced, previously we were required to upload and validate new data through an entry gate that could only be validated internally, a time consuming and lengthy process.
- Allows one to create links to visualisations of datasets that already exist, without the need to embed to recreate.
- Allows the uploading of machine-readable data and metadata, including APIs, making it possible to automate the updating of datasets (previously, updates were done manually on Wikiprogress.stat).

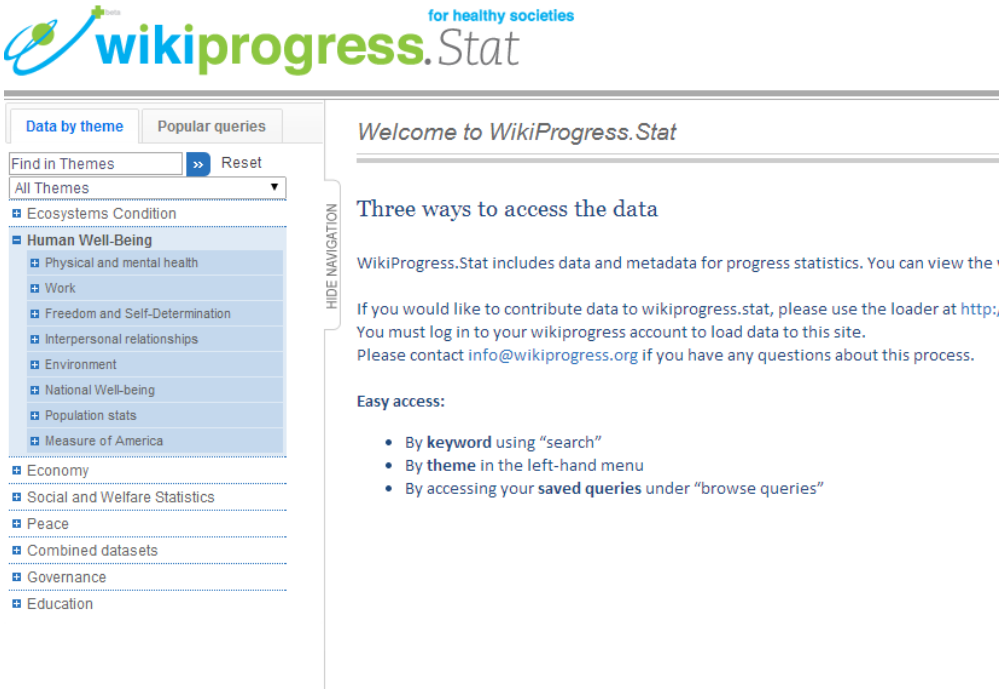


2.4 Customisation and populating of Wikiprogress data

A number of steps have been taken to ensure that the Wikiprogress data responds to our requirements, as follows:

- Key themes have been created.
- A special page has been created for the uploading and displaying of data visualisations for the competitions.
- Extra fields have been added to organisations in order to filter by country, world region and organisation type (e.g. civil society, academia, government, etc.), and location (to feed into an interactive map of initiatives)
- An extension has been created to group datasets and other resources by topic (e.g. education, health, multi-dimensional well-being)
- The functionality has been added to leave comments in datasets and questions in the portal

Below is a screen shot how Wikiprogress.stat used to look, the interface was primarily text and very static. The Wikiprogress data portal is a significant improvement both technically and aesthetically (refer to other screen shots).



3. The results achieved and the open issues

3.1 The Wikiprogress data portal in beta was made public on 29 August 2014 and a link to the portal was placed on the Wikiprogress home page. Between 29 August and 23 September (25 days) the portal has had 40 unique users (have minused two staff), 60% were return visitors and the average sessions duration have been 13:52 minutes , which is extremely high compared to the best articles on Wikiprogress that rarely exceed 8 minutes. This is particularly promising as we have not yet widely communicated the beta portal for user feedback to our community, as we are planning to do this once initial extensions are finalized via an open online discussion on the Wikiprogress platform and continuous feedback from users.

3.2 Comprehensive catalogue for data and statistical resources

The foundation of the data portal has been successfully laid, the main structure and template has been customised to enable one to easily upload and share data with ease and we will aim to have over 100 datasets for monitoring well-being and societal progress uploaded by the end of 2015.

3.3 A user friendly, open source platform with crowd sourced data and statistical resources

The portal has been developed using open sourced software and is extremely easy to use. We will have to profile and reach out to a wider community to get people to upload data in more significant numbers. Currently there are no barriers to uploading data, user are merely required to create an account with an email address, although content will be checked for relevance and to reduce spamming.

3.4 House Open Data.

The portal can host data that is “freely available to everyone to use and republish as they wish, without restrictions from copyright, patents or other mechanisms of control”, as mentioned above the portal itself is designed with this aim in mind. Our challenge will be to encourage a community of active user and partners who are willing to open and share their data.

3.5 Visualize, share and discuss information and results

The Wikiprogress data portal has the functionalities to enable one to visualise, share (via social media tools) and discuss (Disqus tool embedded) data and results. However it is not because we have built the portal that people will come. We are fully aware of the challenges of developing communities and crowdsourcing data and have for this reason designed the our activities over the next 6 months to enable the development of a community that can use data to tell stories via data competitions, student engagement and social media.

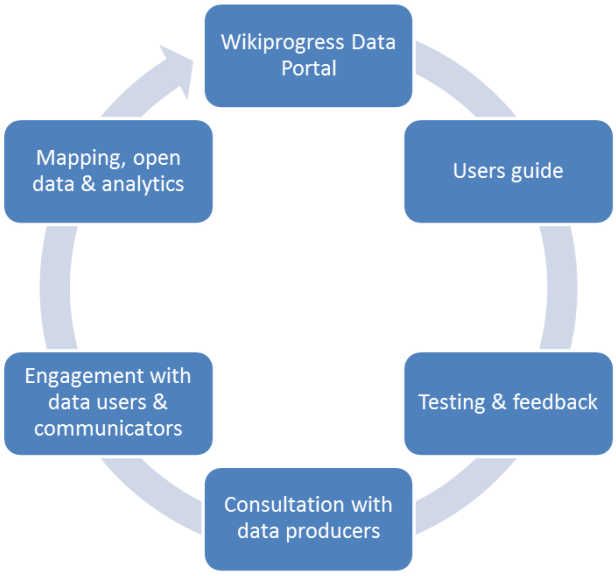
4. Conclusions/Next steps

The launch of the Wikiprogress data portal is really only the beginning of a journey. We will use a lean analytics approach to make continuous, iterative improvement based on user feedback and our own metrics (number of users, return visitors, datasets uploaded, visualisations shared etc.). The portal is a core element in achieving the goal of improving people's engagement with 'beyond GDP' statistics. All the Wikiprogress tools (blogs, Twitter, Facebook and e-Bulletin), activities and the Wikiprogress platform are being realigned, to feed into the Wikiprogress data portal and to create a virtuous cycle of data sharing, use and communication with as wide a community as possible, with the help of Web-COSI partners. Other Web-COSI activities, conducted in cooperation with partners, will help to feed into the data portal by increasing usage, gaining user feedback, and encouraging new users to contribute data and other resources. These include social media campaigns, online discussions, the Wikiprogress University campaign, the data visualisation competitions, and the mapping of initiatives.

OECD Web-COSI activities

Next steps

The following steps should be seen as a feedback loop for the continual improvement of the Wikiprogress data portal and not as a linear one off activity.



1. Users guide

The necessary instruction for using the platform, along with advice, tips and FAQS are being developed to ensure users can make the most of the data portal. A guide for encouraging the sharing of open data will also be developed later.

2. Testing and feedback

A range of data producers, user and communicator will be invited to use the the platform. Most substainal pages on the portal have comment boxes allowing one to provide instant feedback. In October 2014 an online consultaion will be launched on the Wikiprogress platform inviting potential users to provide suggestions and feedback on the data portal and on future Wikiprogress activities. Web-COSI partners will be engaged for feedback, content and to reach out to their networks to promote the portal.

3. Consultation with data producers

Organsiations that produce data for monitoring well-being and social progress will be contacted via email and telephone to both encourage data sharing and to obtain feedback on their experiences with the portal so that continous improvments can be made.

4. Enagagement with data users and communicators.

A data visualisation competition and social media campaign will run alongside the online consultation to reach out to potential user of the portal and data communicators. The competition will require participants to upload any data they choose to use onto the portal, that is not already

available. The Wikiprogress University project and facebook page and will be used to specifically target students and young people and to develop their understanding and interest in well-being statistics and visualisations.

5. Mapping, open data and analytics

An important role of the data portal is to enable one to visualise the well-being data and initiatives easily. A mapping tool will be developed on the portal to enable the geographical mapping of well-being initiatives.

We will be encouraging our community to upload data in .csv and 'clean excel' formats so that the data can be stored in our database. This will allow one to query the data and for it to be published in our API, enabling computer to computer communication, making it possible for application programs to interact with each other and share data.

A analytics report will be developed for the dataportal so that we can set targets, measure activity and make iterative changes in response to demand.