# A Socio-Technical Framework to Enhance the Trust, Reliability and Quality of Citizen Science Data

A. Alabri and J. Hunter

School of ITEE, The University of Queensland

[alabri@itee.uq.edu.au](mailto:alabri@itee.uq.edu.au); [j.hunter@uq.edu.au](mailto:j.hunter@uq.edu.au)

**Abstract**

The number of “citizen science” projects, that involve volunteers, with little or no formal training contributing data and other information to scientific programs, is rapidly growing. The Internet, Web 2.0 and social networking technologies are enabling the establishment of online communities of volunteers who are contributing to projects that range from astronomy [1] to birdwatching [2] and air quality [3]. In particular, the issues of climate change and associated environmental impacts are mobilizing people who want to contribute to the monitoring, management and maintenance of ecosystem health by capturing observational data. Such projects are “democratizing science” in that they enable public citizens to actively participate in scientific programs, and allow them to access and use both their own data and the collective data generated by others. However there are some inherent weaknesses to citizen science, and crowd sourcing more generally – the often limited expertise and anonymity of the contributors can lead to poor quality, misleading or even malicious data being submitted. In this paper we describe a framework that we have developed that allows trust between individuals in an online citizen science community to be measured, inferred and aggregated to generate trust metrics for contributed information and data, based on its provenance and source. Given these measures of trust, the system filters and visualizes search results so that decisions made by scientists and policy makers is based on the most trustworthy of data. The approach also supports user-centric views of trust – enabling the results to be optimized for individual users in an online network. In addition, the system can also be used to motivate contributors by rewarding and recognizing those who are the most reliable and trust-worthy.

**INTRODUCTION**

Examples of citizen science projects have grown dramatically in the last few years. They combine web-based social networks with community-based information systems to harness “collective intelligence” and apply it to a particular scientific problem. In the case of “citizen science” the general public are participating in scientific projects by collecting and contributing data, tags, photos and video – often observations or measurements of biodiversity or environmental indicators. For example, non-scientific communities are actively contributing to the classification of galaxies [1], mapping of threatened species [2] and measurement of air pollution in urban environments [3]. However, these existing projects have also highlighted the problems associated with applying Web-based social networking technologies to Citizen Science projects.

Key features of Web 2.0 and social networking projects are their openness, simple interfaces, low learning curves and dynamic, self-supervised, unregulated nature. However these attractive features are also the source of inherent problems including corrupt, incorrect, misleading or malicious data and information. The issues of trust and data quality are especially significant in citizen science projects since significant and critical decisions are often being made based on the underlying data. In many situations, the community-generated data is being integrated or aggregated with scientific or institutional repositories of authorized data. End users need to know with whom they can share information, from whom they can accept information and which information should be filtered or weighted depending on its reliability and source. Social trust networks provide this information through a combination of user profiles, user rankings of people and information and algorithms for computing and assigning trust.

The aim of this project is firstly to develop capabilities to add trust metrics to citizen science data. Secondly we will evaluate these capabilities with respect to improving the quality of the data being generated and subsequent decisions being made. The implementation, application and evaluation will be carried out in the context of a citizen science case study – the “Coral Watch” project, which is described in detail in the next section.

**BACKGROUND**

**Adding Trust to Social Networks**

In recent years, there have been a number of research projects that have attempted to tackle the problem of adding some measure of trust to information and users within social networking sites [4-6].

**The Coral Watch Case Study**

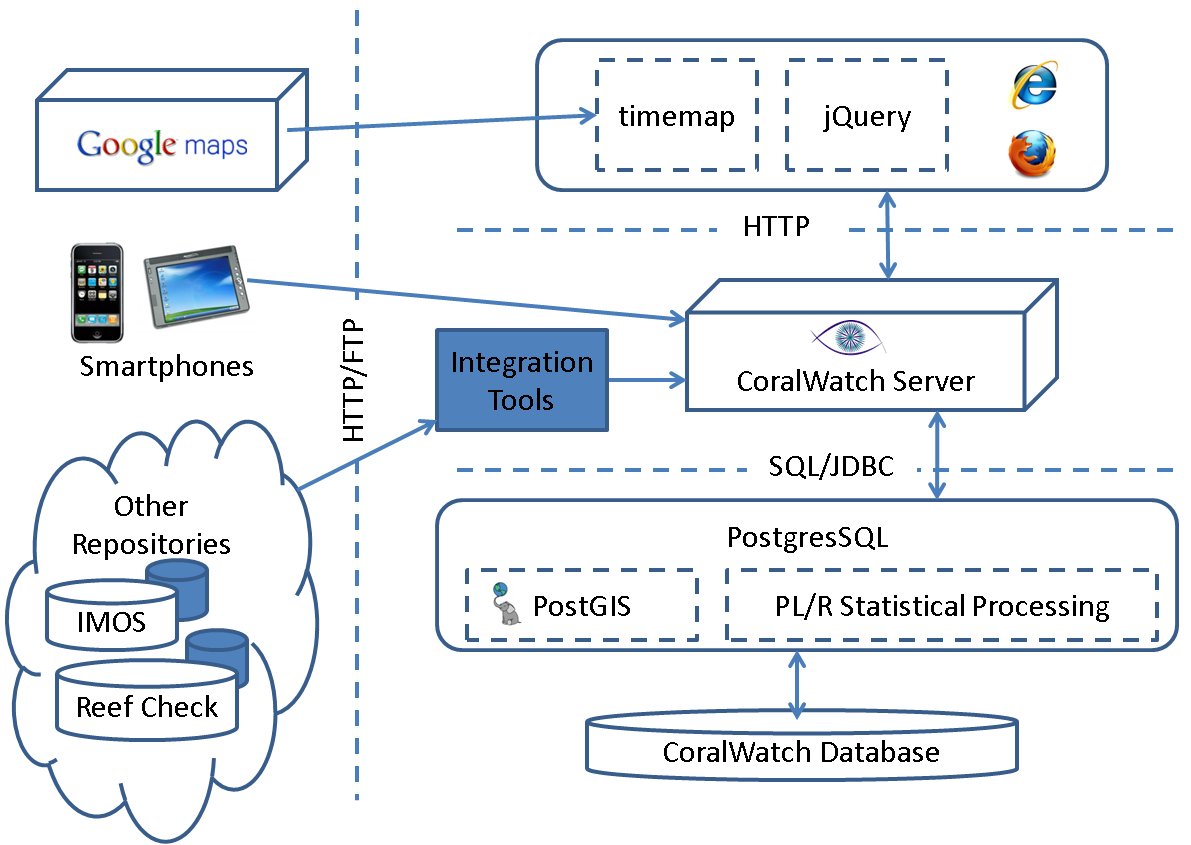
Coral Watch is a citizen science project being managed by the University of Queensland that aims to improve the extent of information on coral bleaching events and coral bleaching trends. Current attempts to monitor coral bleaching often involve costly satellite-born technologies, are restricted to specific locations that researchers are working in and often require sampling of live tissue for physiological analysis. CoralWatch provides simple color charts that can be used by anyone (scientists, tourists, divers, school children) to provide useful monitoring data on coral bleaching on a relatively large scale via an inexpensive, ‘user friendly’ and non-invasive device. As well as collecting monitoring data, the project will educate the public about coral bleaching and its devastating effect on coral reefs.

**IMPLEMENTATION**

In this section we describe the CoralWatch application infrastructure, the user interface of the website and the utilisation of a star rating trust network.

**System Architecture**

The diagram below provides an overview of the system architecture of CoralWatch. The system utilises PostgreSQL object-relational database management system for storing and processing CoralWatch data. PostegreSQL uses PL/R language extension that allows writing R statistical functions (e.g. function for computing trust value of users on CoralWatch) through PostgreSQL functions and aggregate functions. PostGIS is also an OGC compliant plugin that adds support for geographic objects to the PostgreSQL object-relational database [7]. PostGIS offers great geospatial operations such as high speed spatial queries, shape union and difference as well as geometry types such as points, polygons, multipolygons and geometry collections.

****

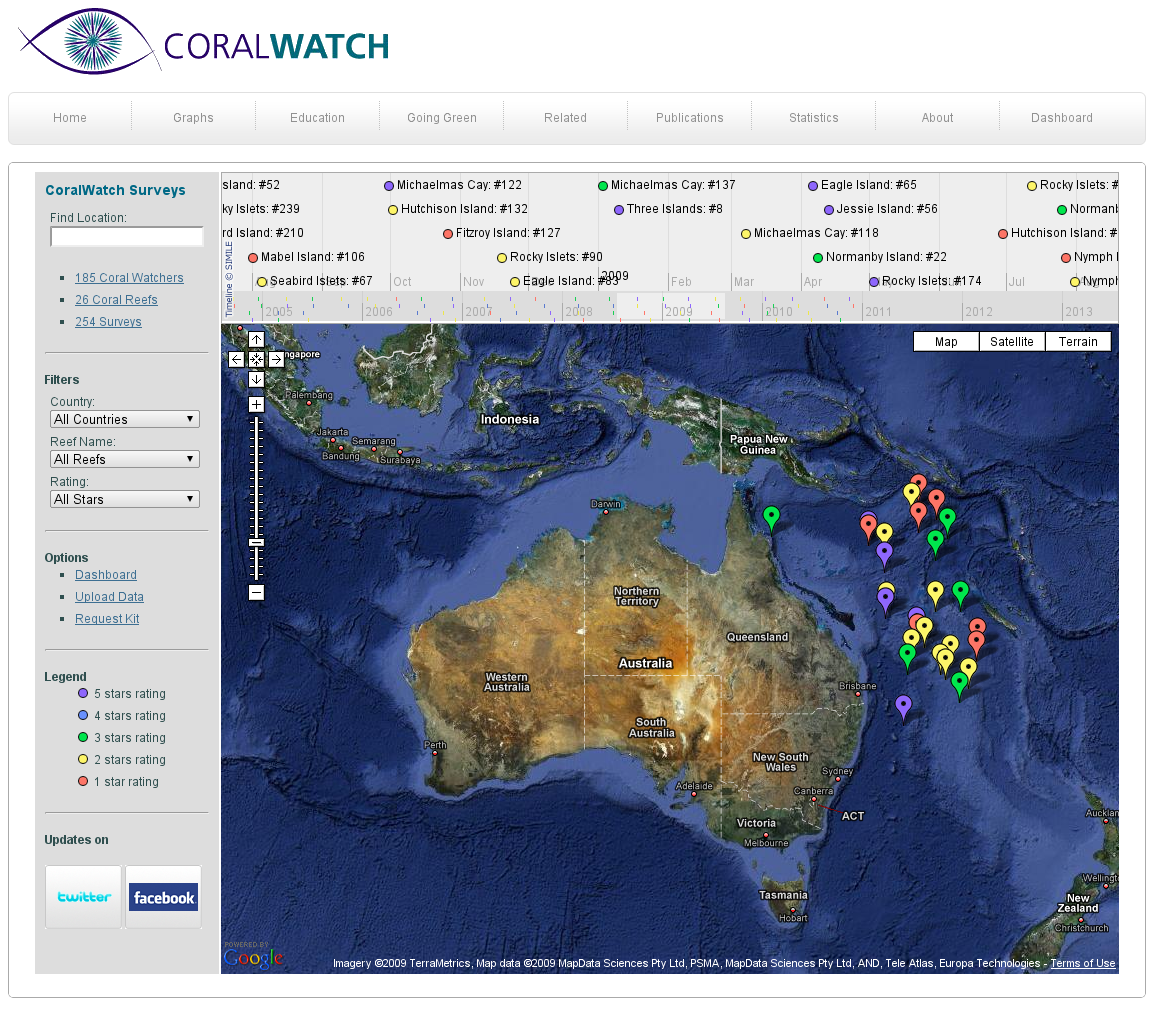
The server component is built using mainly Java and Freemarker programming languages. The server interfaces with third party systems and clients through the following; 1) Web browsers such as Firefox and Internet Explorer. 2) Smartphones. 3) Customised integration tools.

The web browsers are the main interface of the CoralWatch system that allows citizens to upload their data, view surveys and reports, download data and interact with other users. The Smartphone interfaces is used for uploading data from the field. Citizen can collect and submit data as well as photos for their observations through smartphones applications. Utilising the smartphone cmeras, gps data and date and time allows less chances for a citizen to upload incorrect data. The integration tools are a highly customised scripts and programs that are used to harvest data, images and files (e.g. IMOS satellite imagery data) from other repository that provide similar observational data on coral. This data is then used for comparing and correcting the citizen collected data on CoralWatch.

The system utilises Google Maps for the geospatial representation of coral bleaching surveys. The timemap framework allows the integration of spatial and temporal objects into a map (Google Map) and a timeline (SIMILE timeline) simultaneously. This provides a tracking mechanism of bleaching events and speed at which they are happening.

**User Interface**

The screenshot below shows the homepage of CoralWatch website. The main component of the home page is the timemap view of the surveys. The coral bleaching surveys (represented by coloured markers on the map) are layered simultaneously on the map based on the location of the survey and on the timeline based on the date and time in which they were conducted. When the timeline is dragged horizontally to a specific date, it overlays surveys that were conducted around that date on both the time line and the map. The user can click on the surveys (represented by markers) on both the timeline or on the map. This will bring up a balloon showing the survey metadata and data.



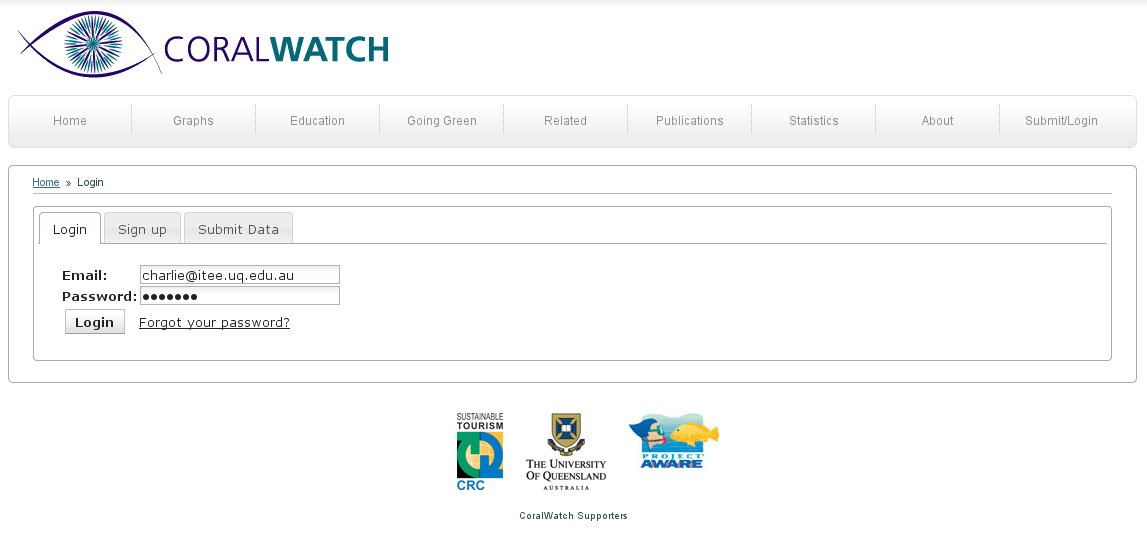
The homepage also provides a sidebar to allow the user to search and interacts with the CoralWatch system through the map. Users can filter surveys based on the value of their rating which is calculated based on a 5 star rating system (see next section).

**Useability**

A user can submit data as registered members with CoralWatch or as a one off submission. One off submission is designed to attract participation from citizen who do not wish to join the CoralWatch program as members. This includes tourists, school and university students.

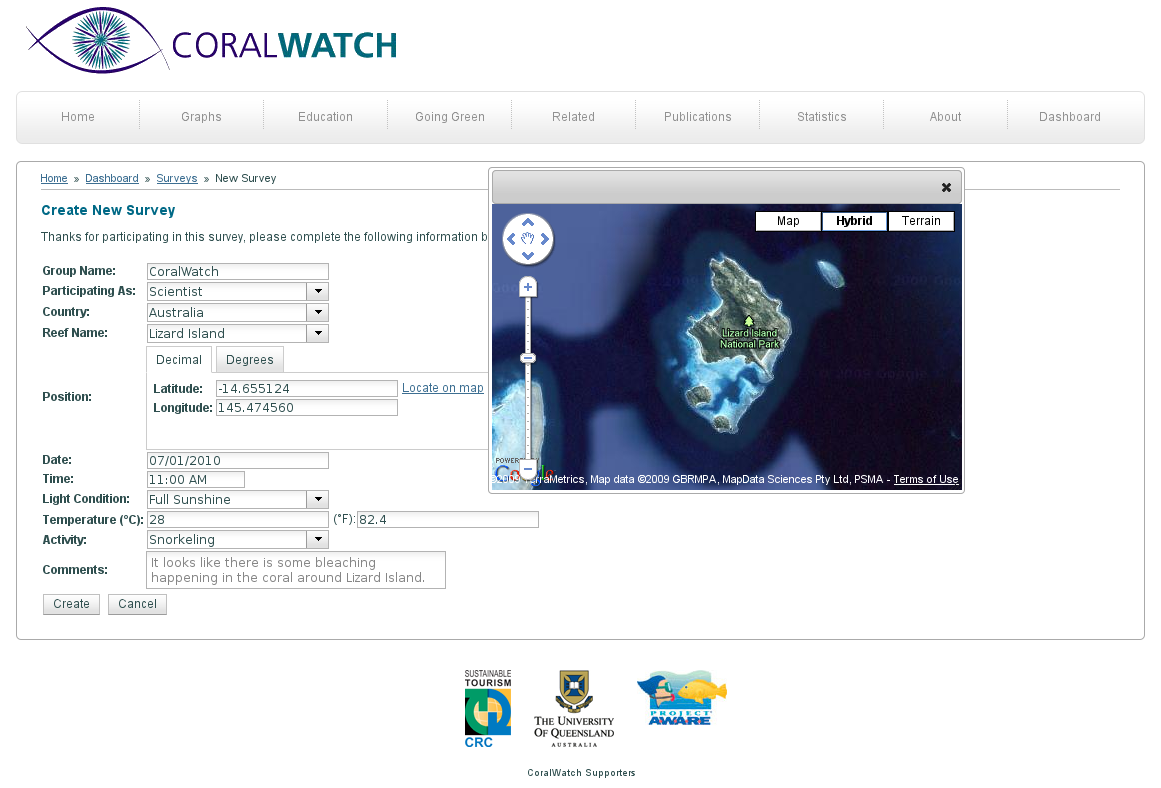
On the login page (See screenshot below), the user will be given three options:

1. Authenticate: this provides and allows the user more options through the website.
2. Sign up: this registers and authenticates a new user.
3. Submit data: this allows the user to submit data without registration but with limited options through the CoralWatch website.

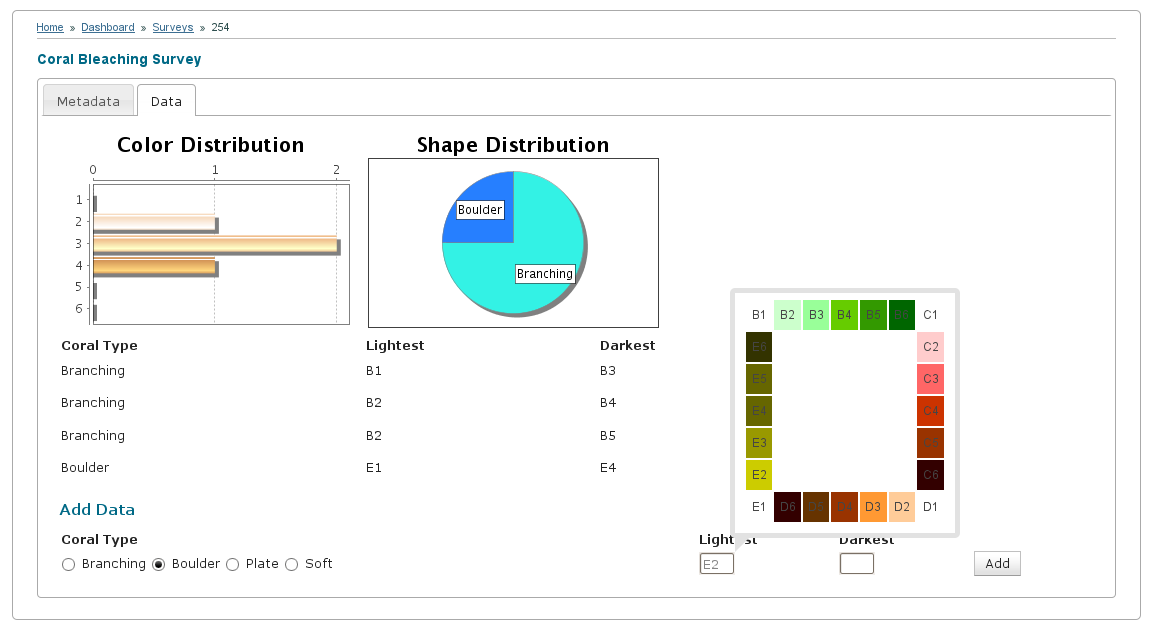


Once the user is authenticated, the user will be presented with a list of options and functionalities. These include changing user settings, viewing existing users, viewing all surveys, creating new surveys and requesting CoralWatch survey kits.

An authenticated user will be able to create a new survey by providing some metadata for the survey first (see the screenshot below). The metadata for a survey include the participant’s expertise and background, the survey location, time and date, the weather condition and water temperature. The users can use a small map widget to locate latitude and longitude of the survey location.



Once the user create a survey, the user will be able to add observations of coral colours to the survey (see the screenshot below). Each observation includes coral types and colour intensity of the coral. The user can use a colour chart widget to record the colour intensity of the coral specie observed. The colour charts are based on the actual colours of bleached and healthy corals. Each colour square corresponds to a concentration of symbionts contained in the coral tissue. The concentration of symbionts is directly linked to the health of the coral. All user have to do is match the colour of the coral with one of the colours in the coral health monitoring chart while in field. The user then record the matching colour codes, along with coral type on the website by creating a survey.



Every time the user submit data for an observation, the data is analysed instantaneously at the server side. The charts generated from the data analysis show the colour distribution across the observed coral. The colour score ranges from 1 to 6 with 1 representing the brightest corals and 6 representing the darkest. It is important to note that not every coral will reach a score of 6, for example the darkest colour observed in some species is a 4. For this reason, you need to analyse grouped data rather than looking at a single coral at a single point in time.

The graphs generated by CoralWatch website show how a particular reef changes over time. These graphs provide an indication of bleaching events of many different reefs at many different times.

**Computing and Representing Trust**

In the CoralWatch context, trust is calculated based on the credibility of participants and the validity of the data submitted by users. Users and surveys are given an accumulative trust value that is between 1 and 5 stars.

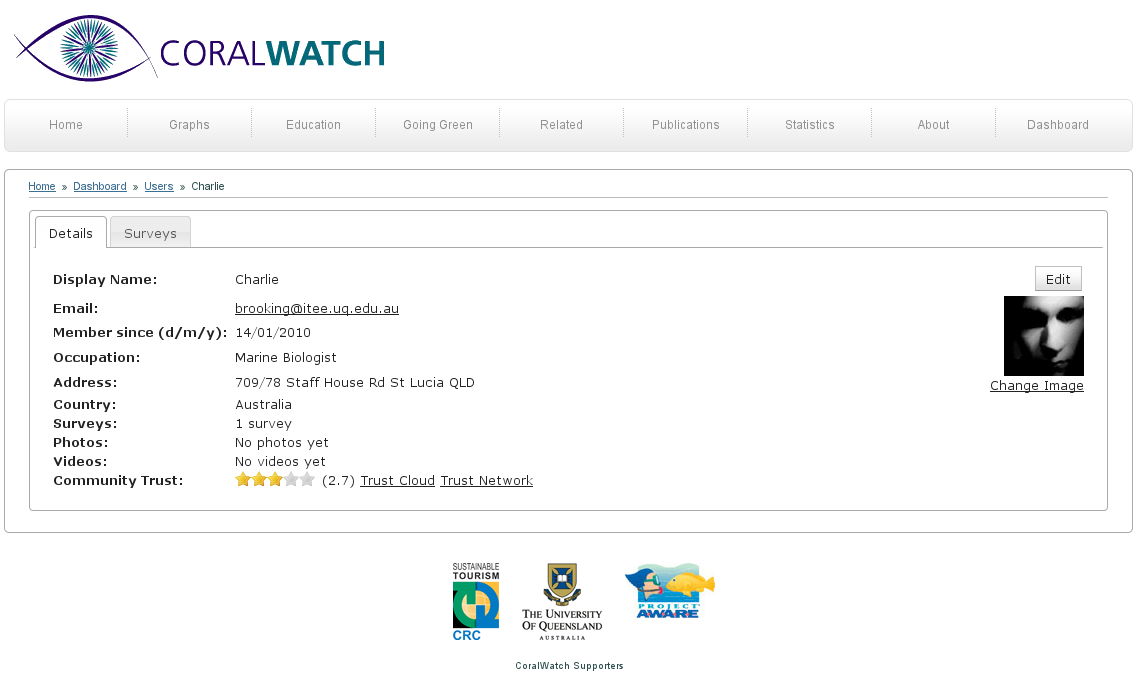
The accumulative trust value of a user is calculated based on the following:

* Average direct trust from other members
* Expertise of the member (e.g. scientist is given higher value than an anonymous user)
* The member’s frequency of participation (number of surveys, images, videos and comments)

The accumulative trust value of a survey is calculated based on the following:

* Average direct rating from other members
* Consistency of the data with other coral monitory data (e.g. reef check)

We use a simple 5 star rating widget to record direct trust between members. Registered members of CoralWatch can give a rating between 1 and 5 stars to other members. The accumulative trust value for a user is shown when viewing a user profile as a star rating of 5 stars. The accumulative trust value of a survey can also be viewed from the survey’s metadata section.



We have implemented two graphical representation of the trust network of CoralWatch members. The first graphical representation is a trust cloud. The trust cloud is simple a word cloud widget showing most trustworthy members in the center of the cloud and the least trustworthy members towards the edge of the cloud (See the screenshot below). This simple visualisation provides an overview of the trust on CoralWatch. It also allows users to identify where they sit in the trust cloud as well as identify other members’ trustworthiness. This visualisation aims to encourage members in a competitive way to enhance community trust on them by participating and providing quality data more regularly.

The second graphical representation of trust network of CoralWatch members is an interactive network visualisation that shows the trust relationship between members.

**EVALUATION**

**DISCUSSION**

**CONCLUSIONS**

**REFERENCES**

[1] Lintott C.J., et al., 2008, "Galaxy Zoo: Morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey", MNRAS, 389, 1179. (available here: [astro-ph](http://uk.arxiv.org/abs/0804.4483)/[UKADS](http://ukads.nottingham.ac.uk/cgi-bin/nph-bib_query?bibcode=2008MNRAS.389.1179L))

[2] **Cooper, C. B., J. Dickinson, T. Phillips, and R. Bonney.** 2007. Citizen science as a tool for conservation in residential ecosystems. *Ecology and Society* **12**(2):11. [online] URL: <http://www.ecologyandsociety.org/vol12/iss2/art11/>.

[3] Mobile Environmental Sensing System Across Grid Environments (MESSAGE)

<http://bioinf.ncl.ac.uk/message/>

[4] J. Golbeck. Trust and Nuanced Profile Similarity in Online Social Networks. *ACM Transactions on the Web.*Vol 3 Issue 4, September, 2009

<http://delivery.acm.org/10.1145/1600000/1594174/a12-golbeck.pdf?key1=1594174&key2=7686351621&coll=ACM&dl=ACM&CFID=22427616&CFTOKEN=67770541>

[5] J. Caverlee, L. Liu, S. Webb, Socialtrust: tamper-resilient trust establishment in online communities, Proceedings of the 8th ACM/IEEE-CS joint conference on Digital libraries

Pittsburgh PA, PA, USA <http://portal.acm.org/citation.cfm?id=1378908>

[6] Caverlee, J., Liu, L., and Webb, S. 2010. The SocialTrust framework for trusted social information management: Architecture and algorithms. *Inf. Sci.* 180, 1 (Jan. 2010), 95-112.

[http://portal.acm.org/citation.cfm?id=1645435.1645491&coll=GUIDE&dl=GUIDE#](http://portal.acm.org/citation.cfm?id=1645435.1645491&coll=GUIDE&dl=GUIDE)

[7] PostGIS http://postgis.refractions.net