Enhancing the Quality and Trust of Citizen Science Data

Abdulmonem Alabri, Jane Hunter
eResearch Lab, School of ITEE
The University of Queensland
Brisbane, Australia
{a.alabri, j.hunter}@uq.edu.au

Abstract - The Internet, Web 2.0 and Social Networking technologies are enabling citizens to actively participate in "citizen science" projects by contributing data to scientific programs. However, the limited expertise of contributors can lead to poor quality or misleading data being submitted. Subsequently, the scientific community often perceive citizen science data as not worthy of being used in serious scientific research. In this paper, we describe a technological framework that combines data quality improvements and trust metrics to enhance the reliability of citizen science data. We describe how trust models provide a simple and effective mechanism for measuring the reliability of community-generated data. We also describe filtering services that remove untrustworthy data, and enable confident re-use of the data. The resulting services are evaluated in the context of the Coral Watch project which uses volunteers to collect data on coral reef bleaching.

Keywords - citizen science; data quality; trust metrics

I. INTRODUCTION

The term *citizen scientist* refers to a volunteer who collects and/or processes data to contribute to scientific research. The number and variety of *citizen science* projects has grown dramatically in recent years. Such projects typically combine web-based social networks with community-based information systems to harness collective intelligence and apply it to a particular scientific problem. Online communities of volunteers are now contributing data to projects that range from astronomy [1] to bird watching [2] and air quality [3]. In particular, the issues of climate change and associated environmental impacts are mobilizing individuals who want to contribute to the monitoring, management and maintenance of ecosystem health by capturing observational data. Citizen science data is often essential for assessing and validating predictive models that involve large-scale spatial and temporal extents. The necessary longitudinal datasets are often incomplete and outdated due to resource limitations with regard to funding and personnel availability. Citizen science can also play an important role in reducing costs associated with research projects and the development of more comprehensive data collection. Furthermore, citizen science programs often lead to increased public awareness of environmental and scientific challenges, civic involvement, fulfillment of academic requirements (in the case of students), and improvement in decision making skills [4].

However, there are some inherent weaknesses to citizen science and crowd sourcing projects. The limited training, knowledge and expertise of contributors and their relative anonymity can lead to poor quality, misleading or even malicious data being submitted [5]. The absence of formal "scientific methods" [6] and the use of non-standardized and poorly designed methods of data collection [7] often lead to incomplete or inaccurate data. Also, the lack of commitment from volunteers in collecting field data [4, 5] can lead to gaps in the data across time and space. Subsequently, these issues have caused many in the scientific community to perceive citizen science data as low quality and not worthy of being considered in serious scientific research [8].

In this paper we propose a technological framework that combines data quality control and trust metrics to provide an indication of the reliability of citizen science data – thus enabling it to be re-used with confidence. The aim of this framework is to provide mechanisms for improving and measuring the quality of citizen science data through both subjective and objective assessments of the data. In addition, tools will be developed that enable trust between individuals in an online citizen science community to be assigned, inferred and aggregated – generating trust metrics for data based on its source. A final objective is to develop querying, visualization and reporting services that take into account the reliability of the data and display it explicitly to users.

A. Hypothesis

A large amount of research has been undertaken into approaches to improve data quality. For example, fairly simple techniques can be applied to validate data input (e.g., syntax, format and values) by checking compliance against schemas. More complex data quality assessment may require comparison with data sets from alternative sources or comparison with historical trends. However these approaches are limited if there are no other sources of comparable data or there is no longitudinal data for trend analysis. An alternative and complementary approach to data quality enhancement services is to exploit social network analysis tools to provide a measure of the trust of the data. A number of different trust models and trust metrics have been developed by researchers in the context of Web 2.0 [9-12]– but to date, none have been applied to citizen science data.

Our hypothesis is that trust and reputation metrics (such as those developed to provide recommender services in

online social networks (e.g., eBay, Netflix) can usefully be applied to citizen science data. Trust models can provide a simple and effective mechanism for filtering unreliable data. Moreover, by combining trust/reputation metrics with data validation services, we can significantly improve the quality and reliability of the community-generated data - enabling its confident re-use by the scientific community.

II. OBJECTIVES

The primary objective of this project is to develop a technological framework for improving the quality and measuring the trust and reliability of citizen science data so it can be confidently re-used by scientists. More specifically the aims are:

- To identify a set of criteria for measuring data quality in citizen science projects;
- To develop a set of services for improving data quality in citizen science projects;
- To evaluate, analyze, refine and optimize these data quality enhancement services – in the context of two exemplary citizen science projects;
- To identify a set of criteria or attributes for measuring trust of citizen science data. For example, these might include:
  - The contributor's role and qualifications (primary student, secondary student, PhD student, volunteer, council worker, scientist);
  - The quality and amount of past data that they have contributed;
  - The extent of training programs completed;
  - Frequency and period of contributing;
  - The contributor's ranking from other members (direct, inferred or calculated using social trust algorithms).
- To survey alternative trust models and algorithms for measuring trust and identify those approaches most applicable to citizen science projects;
- To develop tools for capturing the trust related attributes and for calculating trust within citizen science projects (e.g., the optimum weightings that should be applied to the criteria listed above to determine the most accurate measure of the data's trust);
- To evaluate, analyze, refine and optimize these trust measurement algorithms, tools and services;
- To understand the interactions between data quality and trust metrics – and determine the optimum combination of services for improving the reliability of citizen science data;
- To measure the improvements in data quality that result from using trust metrics to filter or remove untrusted data or untrusted contributors;
- To investigate and identify optimum mechanisms for displaying and communicating the trust, quality of data and reliability of contributors, to other members of the community, especially scientists who are considering re-using the community-generated data.

III. CASE STUDY

This section provides an overview of the CoralWatch citizen science project that we are using as a case study to evaluate the tools and services within our framework.

CoralWatch is a citizen science project managed by the University of Queensland that aims to "*improve the extent of information on coral bleaching events and coral bleaching trends*" [13]. Currently the CoralWatch project has over 1300 members from 80 countries and its members have collected over 29,000 surveys. CoralWatch provides simple color charts (Fig. 1) that can be used by anyone (scientists, tourists, divers, school students) to provide useful monitoring data on coral bleaching on a relatively large scale via an inexpensive, 'user friendly' and non-invasive devices. Data collected through the CoralWatch program includes coral species, coral color, latitude and longitude of the location, reef name, water temperature, data and time and the method by which the data is collected e.g., snorkeling, reef walking or fishing. As well as collecting monitoring data, the project aims to educate the public about the causes and impact of bleaching on coral reefs.

Figure 1. Use of Coral Health Chart in the field

New members register through the CoralWatch website[1]. Once registered, the member can request a DIY Coral Health Monitoring Kit through the website. The kit provides a field guide for recording observations. Each observation includes coral types and color intensity of the coral. The user records the color intensity of the coral species observed by comparing it with a chart. "*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*" [13]. The generates an online survey by recording observations of (species, color, lat, long, etc) along transects and inputting the data to the CoralWatch database via an online data entry page.

A. Current Data Quality

A detailed analysis of a subset of the legacy CoralWatch data (approx. 18560 records, collected between July 2003 and September 2009), was carried out in order to determine the quality of the legacy data. A significant number of errors were identified. Fig. 2 illustrates the distribution of error

types and the extent of errors in the data. Fig. 2 shows that significant errors occurred in the GPS data (~64% of records). Such errors make most of the observations close to useless from a scientific perspective – although the reef name does provide a coarse positioning. There were also a significant number of errors in the volunteers' contact details – making it difficult to attribute errors to individuals, to hold individuals responsible for the data or to contact volunteers to clarify, confirm or correct outlying data. The causes of the majority of the errors were due to:

- Lack of validation and consistency checking;
- Lack of automated metadata/data extraction;
- Lack of user authentication and automatic attribution of data to individuals;
- Absence of a data model;
- Lack of data quality assessment measures;
- Lack of feedback to volunteers on their data;
- Lack of graphing, trend analysis and visualization tools.

By our estimation, over 70% of the errors could be prevented by focusing on new services available via the CoralWatch Portal that focused on the gaps described above.
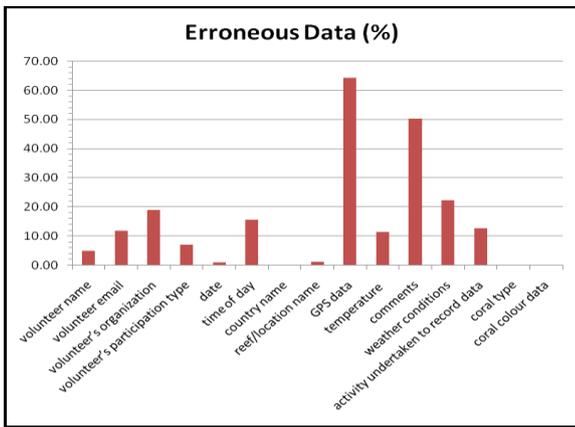

Figure 2. Errors in the legacy CoralWatch data

IV. METHODS

In this section, we describe the different components of the proposed framework (Fig. 3) that will enable the objectives in II to be achieved. Evaluation of these tools and services will be carried out in the context of the Coral Watch project.
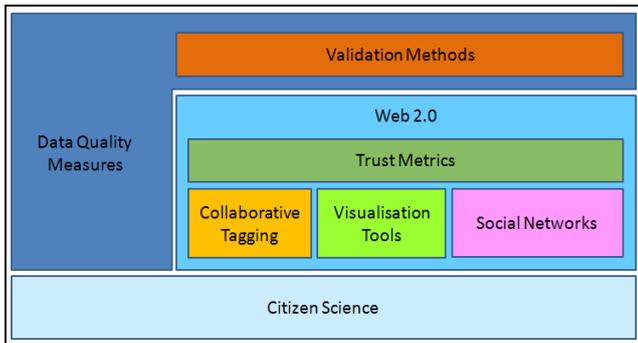

Figure 3. Overview of the proposed framework

A. Data Quality and Data Validation

Wand et al. [14] define data quality as a multidimensional measure of accuracy, completeness, consistency and timeliness. These dimensions can be used to specify whether data is of a high quality by measuring specific deficiency in the mapping of the data from the real-world system state to the information system state. Such dimensions can be used to: develop data quality audit guidelines and procedures for improving the data quality; to guide the data collection process in the field; and to compare the outcomes of different studies.

Currently, most organizations develop data quality measures on an ad hoc basis to solve specific data quality issues where practical and usable data quality metrics are lacking [15]. In many cases these data quality measures are applied as a one-off static process either before or as the data enter the database. This is also apparent in citizen science projects where data quality measures are generally performed during the submission process only. Lee et al. [16] recommend that data quality metrics should be viewed as dynamic, continuous and embedded in an overall data quality improvement process as part of a data collection system.

To achieve data quality improvement in citizen science context, it is necessary to identify the criteria for high quality data. To do this, we employ a data quality measure cycle that includes:

1. Identifying the data quality dimensions
2. Performing data quality measures
3. Analyzing the results and identifying discrepancies
4. Implementing tools that provide necessary actions to improve the quality of data

To identify the data quality dimensions for a citizen science project, we conducted questionnaires and interviews with the stakeholders of the data. Table 1 [17] shows some of the common data quality dimensions which are applicable to citizen science data that were included in our questionnaires.

TABLE I. A SET OF DATA QUALITY DIMENSIONS

Dimensions	Definitions
Accessibility	the extent to which information is available, or easily and quickly retrievable
Appropriate Amount of Information	the extent to which the volume of information is appropriate for the task at hand
Believability	the extent to which information is regarded as true and credible
Completeness	the extent to which information is not missing and is of sufficient breadth and depth for the task at hand
Concise Representation	the extent to which information is compactly represented
Consistent Representation	the extent to which information is presented in the same format
Ease of Manipulation	the extent to which information is easy to manipulate and apply to different tasks
Free-of-Error	the extent to which information is correct and reliable
Interpretability	the extent to which information is in appropriate languages, symbols, and units, and the definitions are clear
Objectivity	the extent to which information is unbiased, unprejudiced, and impartial
Relevancy	the extent to which information is applicable and helpful for the task at hand
Reputation	the extent to which information is highly regarded in terms of its source or content
Security	the extent to which access to information is restricted appropriately to maintain its security
Timeliness	the extent to which the information is sufficiently up-to-date for the task at hand
Understandability	the extent to which information is easily comprehended
Value-Added	the extent to which information is beneficial and provides advantages from its use

In the case of CoralWatch, the syntactic aspects of data quality are easy to measure – and in many cases easy to correct. They include problems with latitude and longitude ranges, spelling errors, invalid temperature, formatting errors.

To reduce syntactic errors, we implemented a metadata and data suggestion/validation process that employs XML Schemas and controlled vocabularies to restrict input to permitted values/ranges and to validate the data. Registered, authenticated members submit their data through a user friendly Web page that performs form validation and checking before the data is saved. For example, country lists and reef names are validated against the GPS data provided by the member. Input data is run through the data quality cycle on submission and the data is assigned a rating value based on the outcome of the quality measure process. If the data does not pass the data quality assessment, it will be marked "unvalidated". This approach also means that the metadata schemas, controlled vocabularies and validation rules may need to be adapted over time.

Checking the syntax of the data is simple compared to checking the validity of the actual values. For example, measuring the quality of the data by comparing it to ground truth, is often very difficult as there is no available ground truth for comparison. In the case of CoralWatch data, it can can be correlated against and compared with related datasets such as ReefCheck data, NOAA satellite data and AIMS bleaching events data. These organizations collect data using other techniques such as sensors, satellite imagery and sea surface temperature to assess the health of coral reef. Hence, these data sets provide an imperfect benchmark against which we may possibly be able to identify outliers or generate a rough indication of data quality.

B. Adding Social Trust Metrics

A considerable amount of research effort has recently been focused on trust, reputation and recommender systems in the context of e-commerce (eBay), social networking sites and social tagging sites [18]. The measured reputation value in these systems is usually a measure of the reliability or quality of users, products (books, films, music), posts, services or user activities. The methods by which these systems calculate and represent a reputation value, varies significantly. For example, online marketplace sites such as eBay and Amazon consider reputation as a single value (represented as number, star or bar rating) that is independent of the context. The information used to calculate the value of reputation is derived from other agents that have interacted previously with the target agent [19]. None of these previous systems has investigated the application of trust metrics to citizen science data [9-12].In this paper we demonstrate how we apply and extend the reputation model developed by Golbeck [18] to calculate reputation within the context of a citizen science project .

1) Calculating Reputation:
Within citizen science projects, trust can be measured by assessing a range of attributes. These include:
- Direct rating between members;

- Inferred ranking or measure of trustworthiness – inferred across the social network using social trust algorithm;
- Direct rating of observations and surveys;
- The contributor's role and qualifications (primary student, secondary student, PhD student, volunteer, council worker, scientist);
- The quality of past data that the volunteer has contributed;
- The extent of training programs that the volunteer have completed;
- Amount of past data contributed;
- Frequency and period of contributing;

In order to calculate the reputation for entities (both users and data) we calculate reputation/trustworthiness by weighting and aggregating a combination of the attributes listed above. Fig. 4 illustrates our model for calculating a unified reputation value for an object `rateeObj` based on the different criteria listed above.

Each time an object is created in the system, the `reputationCalculatorService` is called to create a `reputationProfile` for that object. This contains the `currentReputationValue` for the object which is calculated by executing an `algorithm` in the provided `criterion`. A `reputationProfile` can use more than one `criterion` to derive a `reputationValue`. The `algorithm` can be a function that extracts the different attributes about the `rateeObj` such as:

- `userContributions/totalContributions` or;

- A selection statement such as "if volunteer's role is *scientist* then rating equals *5 stars*.

The collection of `reputationValues` within the criterion is an ordered list from lowest to highest of all the possible reputation values that can be assigned to the `rateeObj` which can be numbers (e.g. {1,2,3,4,5} or strings {bad, fine, good, excellent}). In the case of CoralWatch, we use numbers for the star rating (1-5 stars).
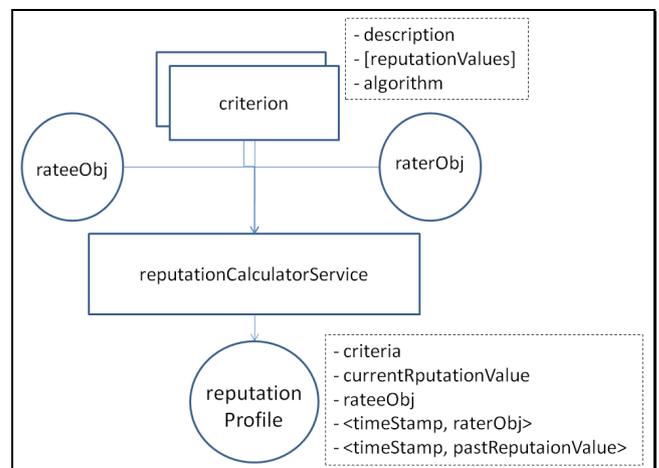

Figure 4. Unified Reputation Calculator Model

The `reputationCalculatorService` also registers all the rater objects `raterObj` (which can be another user in the system or an automatic rating agent) in the `reputationProfile` for a `rateeObj`. An automatic rating agent can be a process that detects the quality of the submitted data and uses `reputationCalculatorService` to evaluate `currentReputationValue` for a user based on its latest submission. It also keeps track of `pastReputationValues` recorded in the `reputationProfile`.

2) Inferencing Reputation: Golbeck [18] used a Recommender System to calculate reputation based on user profile similarity. A collaborative filtering algorithm is used to create a predictive movie rating for a user that will best match the user's attributes. If the user does not have a direct trust value for a person who rated a particular movie, the system will move one step out in the trust network to find connections to users that rated the movie. The process is repeated then a predictive trust value is calculated between user i and user s as in (1). Where t_{ij} is the trust between user i and user j and t_{js} is the trust between user j and user s.

$$t_{is} = \frac{\sum_{j \in adj(i)} t_{ij} t_{js}}{\sum_{j \in adj(i)} t_{ij}} \qquad (1)$$

We use a similar approach for generating inferred trust values for each member of the CoralWatch trust network. In order to calculate an inferred trust value for a particular user, we perform the following steps:

1. Invoke the `reputationCalculatorService` to calculate the reputation of all users for a specific criterion or set of criteria;

2. Compare the current user's reputation for the assessed criteria (e.g. being a scientist and a frequent contributor) against other user's ratings;

3. If no direct connection is found between the current user and the assessed user, the process moves up one step to find neighboring users in the network that have a similar rating for the selected criteria;

4. The process then suggests to the current user how much they can trust other users for a given criterion.

Fig. 5 illustrates the application of this social trust inferencing algorithm to calculate "trustworthiness" between members of the Coral Watch network – some who do and do not directly know each other.
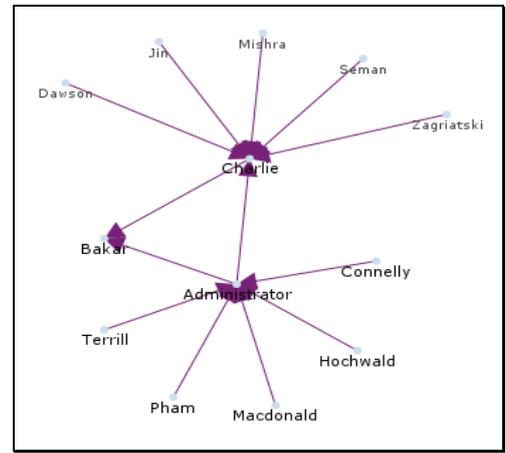

Figure 5.   Visualization of CoralWatch Trust Network

C.   Querying, Visualization and Tagging - Incorporating Data Quality

Once we have calculated trustworthiness metrics for both users and surveys within the CoralWatch project, the next step involves developing filtering, querying and presentation methods that remove, weight or highlight data based on its quality and reliability.

A range of spatio-temporal visualization tools are used to display the data via a GIS visual display (e.g., Google Maps). Through this interface users are able to perform spatio-temporal and keyword-based queries and analysis of the citizen science data. A timeline also allows users to browse, query and animate geospatial observations (by specific volunteers or of specific quality or trustworthiness) over time. Such animations can provide an indication to temporal or seasonal trends in the behavior of eco-systems. Users are also able to specify the level of trust required. For example, users requested the ability to enter queries such as: "Show me all coral watch observations for Masthead Reef between 2007 and 2009 with a ranking of 3 or more stars".

In addition to interactive querying and visualization services, we also developed analysis and reporting tools that produce "Coral Health Reports" which take into account the trust and reliability of the data. For example, users can choose to generate coral health reports for a specific reef using only data that has a 4 star rating or above.

V.   IMPLEMENTATION

A.   System Architecture

The diagram in Fig. 6 provides an overview of the system architecture of the revised CoralWatch Web interface and database that we developed in collaboration with the CoralWatch project managers. The system utilizes the PostgreSQL object-relational database management system for storing and processing CoralWatch data. PostgreSQL uses PL/R language extension that supports R statistical functions (e.g. statistical analysis of coral watch data to determine whether a bleaching event has occurred) through PostgreSQL functions and aggregate functions. PostGIS [20]

also provides geospatial operations such as high speed spatial queries, shape union and difference, geometry types such as points, polygons, multipolygons and geometry collections.

The server component is built using Java and JSP. The server interfaces with third party systems and clients through the following;

1. Web browsers (e.g., Firefox and Internet Explorer).

2. Smartphones.

3. Customized integration tools that correlate CoralWatch data with other data repositories using an ontology such as the Health-e-Reef ontology developed by Allen et al [21].
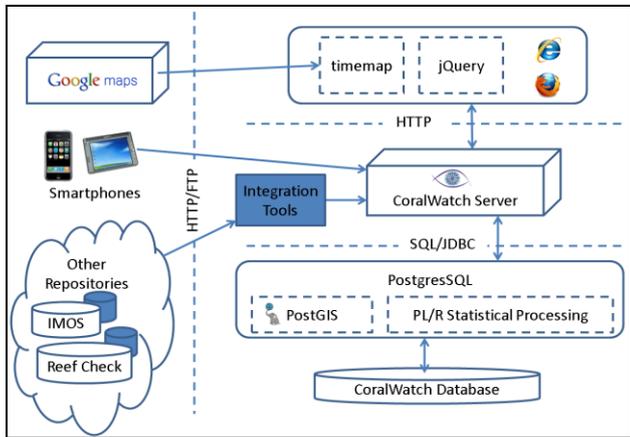

Figure 6.   CoralWatch system architecture.

The Coral Watch portal provides the interface by which users can upload their data, view surveys and reports, download data and interact with other users. A Smartphone interface is currently under development to enable data to be uploaded directly from the field and automatic extraction of GPS data, and date and time.

The system utilizes Google Maps for the geospatial interface to coral bleaching survey data. A SIMILE timeline enables users to specify the temporal. This provides a tracking mechanism of bleaching events and speed at which they are happening.

The integration tools are a set of scripts that harvest data, images and files from other related repositories (e.g. IMOS satellite imagery data, AIMS coral bleaching event registry), and map it to a common observational data model. These datasets are used as a benchmark or "ground truth" to measure the quality of the volunteers' CoralWatch data. It can also be easily displayed as layers through GoogleMaps – on which the CoralWatch surveys can be overlaid.

B.   User Interface for Assigning and Displaying Trust

Users first need to register via the Coral Watch Web site. Registration requires users to enter contact details, current role (secondary student, undergrad, postgrad, post-doc, research fellow, professor, teacher, volunteer), expertise and professional qualifications. Once registered, a user profile is stored and they are assigned a user id and password.

Authenticated users can create a new survey by first entering the metadata for the survey. The metadata includes the survey's location (lat and long), date/time, weather conditions and water temperature. A validation process then checks compliance of the input data against an XML schema and controlled vocabularies. Once the user has created a new survey, they can enter the set of observations of coral species and color (Fig. 7).

Every time the user submits an observation, the data is instantaneously analysed on the server side. The charts generated from the data analysis show the color distribution across the observed coral reef transect. Users can determine whether a bleaching event has occurred on a particular reef by analyzing the change in color over time.

Figure 7.   Submitting data via the CoralWatch Web Entry Form.

Once the survey data has been entered, the next step is to calculate trust metrics for it. To date, we have developed simple tagging tools whereby members of the network can assign trust rankings to other members. The aggregate community trust value on a member is calculated by weighting and aggregating both direct and inferred trust values plus additional attributes (e.g., role, expertise, quality of past data, frequency and extent of past contributions) as described in IV section B. The calculated aggregate trust value is displayed as a 1to 5 star rating in the user's profile (Fig. 8) – this information is visible only to the system administrator. These trust metrics are then also associated with the data uploaded by that member.
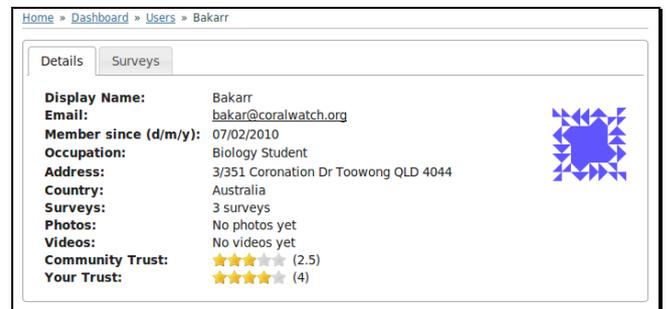

Figure 8.   User profile showing trust as 5 star rating.

The screenshot in Fig. 9 shows the Mapping and Timeline interface to the CoralWatch data.

The coral bleaching surveys (represented by coloured markers on the map) are layered simultaneously on both the map and the timeline above the map. When the timeline is dragged horizontally to a specific date, the surveys that were conducted around that date are displayed on both the map. The user can click on the surveys (represented by markers) on either the timeline or the map to display a balloon that contains the survey metadata and observational data.

The search and browse page also provides a sidebar that allows users to search and filter the CoralWatch data. Users can search on: contributor name, date range, location, species, reef name. They can also filter surveys based on the value of their rating (1-5 stars). For example, the observations in Fig. 9 are colored according to their trust metric – red = 1 star, purple=2 star, yellow=3 star, white = 4 star, green = 5 star.
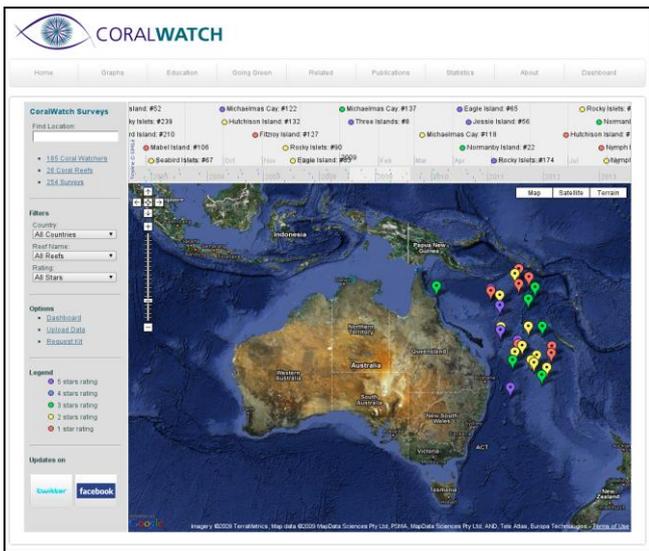

Figure 9.   User Interface of CoralWatch

VI.   EVALUATION

Following the implementation of the revised Coral Watch system described above, evaluations were carried out on different aspects of the system to determine improvements to data quality, the user interface, performance and data reliability:

- We compared the type and distribution of syntactic errors in the submitted data before and after the implementation of the data validation services. As anticipated – through the use of pull-down menus, controlled vocabularies, range checking and XML schema compliance, we reduced the number of syntactic errors in the input data by over 70%.
- A survey of feedback from the users and administrators of the CoralWatch data indicated that the trust metrics associated with individual users should be hidden – so as not to deter volunteers – but that trust metrics associated with specific datasets should be explicit. Poor trust metrics associated with individual volunteers could be used to target online training modules. Good trust metrics could be used to reward, encourage and retain volunteers.

- The response from users to the ranking/tagging tools and the improved filtering, search and browse interfaces – was that these tools were relatively simple to use and greatly improved users' ability to understand the temporal, seasonal and spatial trends in coral bleaching events.
- Deliberate submission of consistently poor data by dummy users was picked up eventually by other members of the network who assigned low rankings to these contributors. But there was a delay period during which the data and the user was unknown and assigned an "average" ranking – which was not sufficient to filter it out.

Additional time will enable further evaluation to be carried out that will:

- Identify the best algorithms, weightings and approaches for measuring trust attributes and for calculating overall trust
- Measure the performance, accuracy, efficiency and scalability of the trust metric tools as the size of the community and the database grows;.
- Monitor changes in the number and frequency of contributing volunteers, the retention of existing volunteers and the attraction of new volunteers.

VII.   FUTURE WORK

In future we would like to investigate the application and viability of Attack Resistance trust metrics [22] in the context of citizen science data. The Attack Resistance trust metric is designed to filter out bogus or malicious users from a social network thus reducing the submission of invalid or deliberately misleading data. A FOAF Role-based Access Control Standard [23] can be adopted to define the relationships between members in a citizen science project. The named relationships will be the basis for certification levels of this approach. A simple relationship model of a trust network is represented by (Fig. 10) with named edges.
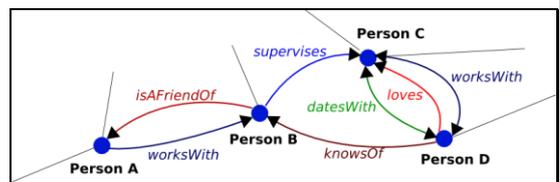

Figure 10. Named Relationships Model of a social network

Each edge between nodes will be assigned a certification level which will be used to calculate the capacities of accounts. Periodic execution of this trust metric will remove any bad nodes (uncertified accounts) within the network. This will ensure that only certified and genuine volunteers remain in the system.

We also plan to extend and evaluate the tagging tools to enable ranking/tagging of geo-located citizen science data. A good example is the approach of Rezel et al.'s [24] that enables users to add tags to data/observations through a mapping interface. For example, users will be able to attach ranking tags and other annotations to specific observations to highlight data quality issues.

VIII. CONCLUSION

Citizen science is democratizing science in that it enables public citizens and the scientific community to work together in monitoring, managing, maintaining and understanding the environment around us. A literature review has revealed that there is a critical the need for a framework to improve the quality and trust of citizen science data – and that there exists a range of technologies from the data quality and social trust fields, that can potentially be combined to maximize the quality and re-use of citizen science data.

Using the CoralWatch project as a case study, we have implemented a system that demonstrates that it is possible to significantly improve the quality of community generated observational data through a set of validation and verification tools. We have also shown that it is possible to calculate a measure of the reliability or trustworthiness of citizen science data using a weighted aggregation of both direct and inferred attributes. By explicitly enabling this metric to be displayed to users, and considered within querying and reporting services, we have enhanced the potential re-use of citizen science data by scientists.

REFERENCES

1. Lintott, C.J., et al., *Galaxy Zoo : Morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey.* Monthly Notices of the Royal Astronomical Society, 2008. **389**(3): p. 1179-1189.

2. Cooper, C.B., et al., *Citizen science as a tool for conservation in residential ecosystems.* Ecology and Society, 2007. **12**(2).

3. (MESSAGE), M.E.S.S.A.G.E. *Project Overview*. 2010 [cited 2010 20/02/2010]; Available from: http://bioinf.ncl.ac.uk/message/?q=node/5.

4. Galloway, A.W.E., M.T. Tudor, and W.M.V. Haegen, *The Reliability of Citizen Science: A Case Study of Oregon White Oak Stand Surveys.* Wildlife Society Bulletin, 2006. **34**(5): p. 1425-1429.

5. Foster-Smith, J. and S.M. Evans, *The value of marine ecological data collected by volunteers.* Biological Conservation, 2003. **113**(2): p. 199-213.

6. Paulos, E., *Designing for Doubt Citizen Science and the Challenge of Change*, in *Engaging Data: First International Forum on the Application and Management of Personal Electronic Information*. 2009: MIT - Cambridge, MA, USA.

7. Silvertown, J., *A new dawn for citizen science.* Trends in Ecology and Evolution, 2009. **24**(9): p. 2832-2842.

8. Delaney, D.G., et al., *Marine invasive species: validation of citizen science and implications for national monitoring networks* Biological Invasions, 2007. **10**(1): p. 117-128.

9. Chandrasekaran, S., et al., *XML-based modeling and simulation: web service technologies and their synergy with simulation*, in *Proceedings of the 34th conference on Winter simulation: exploring new frontiers*. 2002, Winter Simulation Conference: San Diego, California. p. 606-615.

10. Hevner, A.R., et al., *Design Science in Information Systems Research.* MIS Quarterly, 2004. **28**(1): p. 75-106.

11. Liu, Y., A.H. Ngu, and L.Z. Zeng, *QoS computation and policing in dynamic web service selection*, in *Proceedings of the 13th international World Wide Web conference on Alternate track papers & posters*. 2004, ACM: New York, NY, USA. p. 66-73.

12. Nardin, L., et al., *SOARI : A Service Oriented Architecture to Support Agent Reputation Models Interoperability*, in *AAMAS-TRUST*, 2008. p. 292-307.

13. Reid, C., et al., *Coral Reefs and Climate Change: The guide for education and awareness.* 2009, Brisbane: CoralWatch.

14. Wand, Y. and R.Y. Wang, *Anchoring data quality dimensions in ontological foundations.* Commun. ACM, 1996. **39**(11): p. 86-95.

15. Huang, K.-T., Y.W. Lee, and R.Y. Wang, *Quality information and knowledge*. 1999: Prentice Hall PTR. 209.

16. Lee, Y.W., et al., *Process-Embedded Data Integrity.* Journal Database Management, 2004. **15**(1): p. 87-103.

17. Kahn, B.K., D.M. Strong, and R.Y. Wang, *Information quality benchmarks: product and service performance.* Commun. ACM, 2002. **45**(4): p. 184-192.

18. Golbeck, J., *Trust and nuanced profile similarity in online social networks.* ACM Trans. Web, 2009. **3**(4): p. 1-33.

19. Sabater, J. and C. Sierra, *Review on Computational Trust and Reputation Models.* Artificial Intelligence Review, 2005. **24**(1): p. 33-60.

20. PostGIS. *"What is PostGIS?* 2010 11/02/2010]; Available from: http://postgis.refractions.net/.

21. Allen, C., *The Health-e-Reef Project - Eco-Informatics on a Global Scale*, in *eResearch Australasia*. 2008: Melbourne, Australia.

22. Levien, R., *Attack Resistant Trust Metrics*, in *Computing with Social Trust*, J. Golbeck, Editor. 2009, Springer: London. p. 121-132.

23. Grzonkowski, S. and S. Kruk. *D-FOAF: Role-based Access Control Standard*. 2007 [cited 2009 11/10/2009]; Available from: http://www.foafrealm.org/documentation/AccessControl/.

24. Rezel, R. and S. Liang. *SWE-FE: Extending folksonomies to the Sensor Web*. in *Collaborative Technologies and Systems (CTS), 2010 International Symposium on*. 2010. Chicago, IL, USA.