Citizen OBservatory WEB (COBWEB): A Generic Infrastructure Platform to Facilitate the Collection of Citizen Science data for Environmental Monitoring*

Christopher I. Higgins[1], Jamie Williams[2], Didier G. Leibovici[3], Ingo Simonis[4], Mason J. Davis[5], Conor Muldoon[6], Paul van Genuchten[6], Gregory O'Hare[7], Stefan Wiemann[8]

[1]EDINA, University of Edinburgh, chris.higgins@ed.ac.uk

[2]Environment Systems Ltd., jamie.williams@envsys.co.uk

[3] Nottingham Geospatial Institute, University of Nottingham, didier.leibovici@nottingham.ac.uk

[4]Open Geospatial Consortium, isimonis@opengeospatial.org

[5]Welsh Government, Mason.Davis@Wales.GSI.Gov.UK

[6]GeoCat, conor.muldoon@gmail.com; paul.vangenuchten@geocat.net

[7]University College Dublin, Gregory.OHare@ucd.ie

[8]Technische Universität Dresden, stefan.wiemann@tu-dresden.de

Abstract

The mass uptake of internet connected, GPS enabled mobile devices has resulted in a surge of citizens active in making a huge variety of environmental observations. The use and reuse potential of these data is significant but currently compromised by a lack of interoperability. Useable standards either don't exist, are neglected,

DOI: 10.2902/1725-0463.2016.11.art3

poorly understood or tooling is unavailable. Large volumes of data are being created but exist in silos. This is a complex problem requiring sophisticated solutions balanced with the need to present sometimes unsophisticated users with comprehensible and useable software. COBWEB has addressed this challenge by using the UNESCO World Network of Biosphere Reserves as a testbed for researching and developing a generic crowdsourcing infrastructure platform for environmental monitoring. The solution arrived at provides tools for the creation of mobile Applications which generate data compliant with open interoperability standards and facilitate integration with Spatial Data Infrastructures. COBWEB is a research project and the components of the COBWEB platform are at different Technology Readiness Levels. This paper outlines how the overall solution was arrived at, describes the main components developed and points to quality assurance, integration of sensors, interoperability and associated standardisation as key areas requiring further attention.

Keywords: citizen science, crowdsourcing, biological monitoring, standards, Open Geospatial Consortium, environmental governance, spatial data infrastructure, semantics, sensors, access control, privacy, interoperability.

1. INTRODUCTION

New and innovative environmental monitoring and information capabilities can enable effective participation by citizens in environmental monitoring, based on broad stakeholder and user involvement in support of both community and policy priorities (Liu et al, 2014). However, the upsurge in availability of ubiquitous mobile devices has resulted in a fragmented landscape where there are a large, and increasing, number of citizen science type projects collecting data which are often highly specific to those projects (Roy et al, 2012). It is recognized that, in terms of reuse, sharing and integration of data (Newman et al, 2012), significant benefits will be realized once greater cognizance is given to interoperability, the use of standards and leveraging existing investment in data infrastructures.

The EU funded Framework Programme Seven (FP7) Citizen OBservatory WEB (COBWEB) project addresses this challenge. The main objective of the project has been to research and develop an innovative generic infrastructure platform to facilitate the collection of citizen science data for the purpose of improved environmental monitoring, with a particular emphasis on contributing to environmental governance processes. The focus has been on the use of the suite of open interoperability standards that underpin Spatial Data Infrastructures (SDI) initiatives such as the Global Earth Observation System of Systems (GEOSS)1 and Infrastructure for Spatial Information in the European Community (INSPIRE)2. COBWEB demonstrates how advances in mobile and sensor technology combined with the large increase in availability of mobile devices, especially of smartphones, can equip citizens to make observations of use for good environmental governance by a variety of actors.

The project provides open source tooling which can be downloaded3, configured and modified if necessary to enable citizen observatory's – infrastructure which enables citizens to use their own mobile devices to make observations across a wide range of different citizen science scenarios. The platform developed and described below is generic, extensible and powerful enough to accommodate sophisticated requirements arising from a broad range of stakeholders across different sectors, yet sufficiently flexible to enable non-experts to create and use mobile applications which meet their needs.

It should be noted that COBWEB is a research project established to provide some answers, in so far as is possible, to the challenge articulated above. As such, the various components described below are at different levels of technical maturity ranging from Technology Readiness Level (TRL) 4 (validated in laboratory) to TRL 9 (actual system proven in operational environment). The TRL scale (EARTO 2014) is used as a tool for decision making on research and development investments at the EU level and has been developed to enable assessment and comparison of technologies in respect of maturity.

COBWEB provides high TRL tooling for collecting new data and, where possible, the software developed leverages and complements existing well established high TRL open source projects such as those in the OSGeo4 suite. COBWEB has also conducted research (low TRL) into the use of other crowdsourced data, for example, from social media streams and sensors, for both quality assuring and enriching observations (Wiemann et al, 2015). The latter is an example of the benefits of leveraging SDI as standards are critical for relating and combining (the terms fusion and conflation are also used) spatial data from various sources.

The requirements for COBWEB flowed from concentrating on three pilot case study areas: the creation and validation of data products from Earth Observation data, biological monitoring and flooding.

---

[1] https://www.earthobservations.org/geoss.php [accessed 9 June 2016]
[2] http://inspire.ec.europa.eu/ [accessed 9 June 2016]
[3] https://cobwebproject.eu/news/publications [accessed 9 June 2016]
[4] http://www.osgeo.org/uk [accessed 8 June 2016]

COBWEB has worked on these case study areas within the context of the United Nations Educational, Scientific and Cultural Organization's (UNESCO) World Network of Biosphere Reserves (WNBR). Modern Biosphere Reserves can only be designated with explicit support from the local community and are established as areas of high nature conservation value with demonstrably enthusiastic local communities interested in promoting the sustainable development agenda. Consequently, they constitute potentially excellent testbeds for environmental citizen science related projects. The WNBR is being used and evaluated within COBWEB as our testbed environment - where we develop, deploy, test and validate our concept of a citizen observatory. COBWEB utilises Biosphere Reserves in Wales, Germany and Greece to provide a European dimension to this FP7 project and facilitate comparison of different aspects of the infrastructure platform across Europe.

This paper builds upon (Higgins et al, 2016), with further detail on the generic infrastructure platform solution as developed and demonstrated within these Biosphere Reserves.

2. SYSTEM DESIGN THROUGH STAKEHOLDER ENGAGEMENT AND CO-DESIGN

COBWEB commenced with a process of requirements gathering for a system that would enable citizens in the Dyfi Biosphere Reserve area in Wales to collect data in the 3 pilot case study areas as introduced above. Requirements were initially gathered through a process of desk study reviewing existing systems (COBWEB, 2016c) and structured interviews. The latter started with citizen groups located within the Biosphere Reserve and were widened and further refined through broader stakeholder engagement; in particular with groups associated with environmental governance and stewardship.

In parallel, a process of rapid prototyping software development was entered into with the intention of creating software that demonstrated the feasibility of meeting identified requirements and surfacing issues and challenges. Once we had reached the requisite level of software maturity, to further assess the viability of the proposed solution and better understand the needs of the citizen, COBWEB then engaged in a period of structured co-design activity

2.1. Co-Design

Co-design happens where "participants are invited to cooperate with designers, researchers and developers during an innovation process" (Wikipedia, 2016). Seven citizen groups active in the Dyfi Biosphere Reserve in mid-Wales were commissioned (COBWEB, 2016a) through an open tendering process to help gather requirements and validate our concept of a citizen observatory by mobilising citizens (usually volunteers) to go out in the field to collect data (Table 1).

The invitation to tender was issued via the the Dyfi Biosphere Reserve and COBWEB websites and resulted in a strong response from the community within the Biosphere Reserve expressing a desire to work with an unexpectedly (from the project consortiums perspective) rich set of environmental phenomena. This generated a diverse set of requirements and may be interpreted as evidence of the Biosphere Reserve mechanisms successfully working to help meet the sustainable development agenda in the context of a research project which needs to mobilise citizens.

Table 1: Summary of main Co-Design activity in 2015

Citizen Group	Collecting Information on	COBWEB Champion
Cardigan Bay Marine Wildlife Centre	Marine megafauna, e.g. dolphins, porpoise, seals	Aberystwyth University
Coetiroedd Dyfi Woodlands	Variety of different woodland species (flora and fauna)	Ecodyfi
Penparcau Community Forum	Variety of species with a focus on butterflies and larval food plants	Environment Systems Ltd
RSPB (Royal Society for the Protection of Birds)	Habitat (salt marsh and peat bog) reversion processes	Aberystwyth University
Snowdonia National Park Authority	Invasive species (Japanese Knotweed)	Welsh Government
The Outward Bound Trust	Variety of species	Aberystwyth University
Ysgol Bro Hyddgen	Variety of species	University of Edinburgh

Each citizen groups proposal was analysed (COBWEB, 2016b) and refined in respect of what it could offer the project, e.g. research potential, policy implications, technical challenges, ethical implications, etc. The result was a challenging set of requirements to guide system design, based on real user needs (ISO, 2010), and the opportunity to predicate further research on real world requirements.

The seven co-design sub-projects were programme managed by COBWEB partner Ecodyfi (the delivery arm of the Dyfi Biosphere Reserve partnership) with each sub-project allocated a ‘champion’ (Table 1) from within the project. A Steering Group with representation across the COBWEB consortium was established to coordinate and ensure consistency. Throughout the field season the seven co-design partners listed above had, between them, over a thousand people collecting data during fieldtrips of varying duration and frequency that they organised (COBWEB, 2016d). This is a significant number of people considering that the Dyfi Biosphere Reserve area is mostly rural.

Regular workshops, interviews, participation in, and feedback from, the fieldtrips facilitated understanding throughout. Concentrating on the experience gained by a wide variety of users using COBWEB software and access to these users. COBWEB used the field data and the intelligence flowing back from the interactions with the citizens and co-design partners to refine and improve the concept and system design.

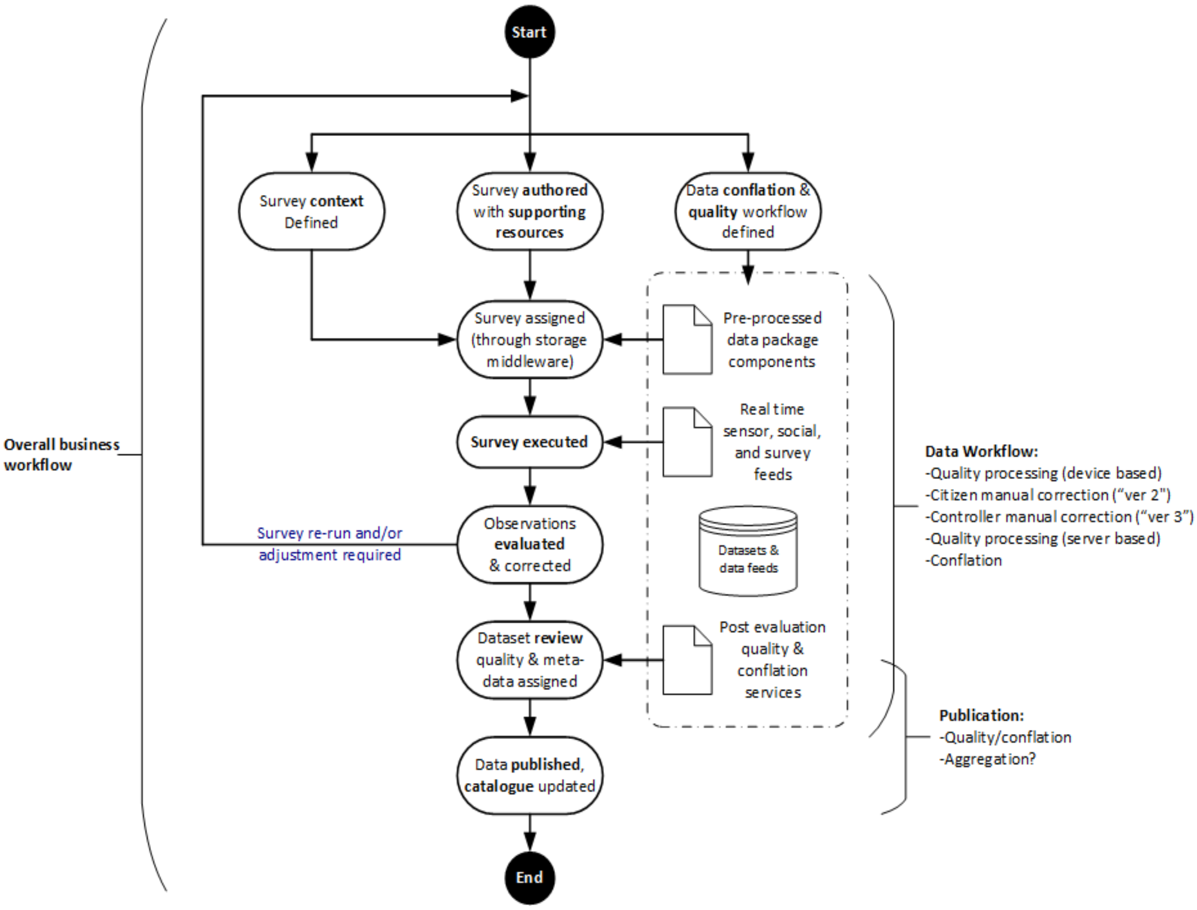
Some lessons learned: the difficulty of enabling communication between scientists and ‘ordinary’ citizens to facilitate dialogue on sometimes obscure scientific concepts should not be underestimated. Interviews with volunteers participating in a project that can claim to be working towards better environmental governance (having Welsh Government as a partner on the project was important) proved to be a strong motivator. A particular challenge of this kind of co-design activity is the need to continually manage expectations - it was emphasised repeatedly to the sub-project leads and volunteers that they were participating in a research project which was testing and validating our concept of a citizen observatory; the components of which are at various TRL’s with research and development continuing in parallel. Significant technical support had to be allocated to respond quickly to prevent disillusionment with failing software.

Participants responded with enthusiasm to software releases where they could see the issues they identified being addressed and their ideas implemented. One of the outcomes has been establishment of an enthusiastic partner network interested in further work. The role of the Biosphere Reserve concept in facilitating sustainable development related activity was widely appreciated, particularly by the seven co-design partner organisations.

3. COBWEB WORKFLOW

As can be seen from Figure 1 (an illustration of the overall workflow underpinning COBWEB), the concept of surveys is central to COBWEB. A survey represents the context for a group of citizens going out into the field with one or more tasks over a period of time in a certain area. It constitutes a set of forms, workflows (business processing rules applying to observations) and datasets, all of which may be adjusted as the needs of a particular survey dictate.

Figure 1: The COBWEB Workflow



COBWEB has found that, depending upon exact circumstances and the amount of control required, several key roles may be identified (Table 2) in order to enact surveys. These roles are of broader applicability though, depending upon circumstances, they may overlap.

COBWEB set out with the aim of being generic; the same software can be used to enable observations on a wide variety of environmental parameters. Surveys are authored by the 'Survey Manager' in order to meet specific requrements. When doing so and if required, the 'Quality Reviewer' liaises with the sponsor to ensure that quality control processes are established that meet their needs and which generates sufficient metadata to enable 'fit for purpose' decisions to be made by consumers, ie, is the data being described by the metadata suitable, appropriate and of sufficient quality for the envisaged use.

Table 2: Key Actors and Roles

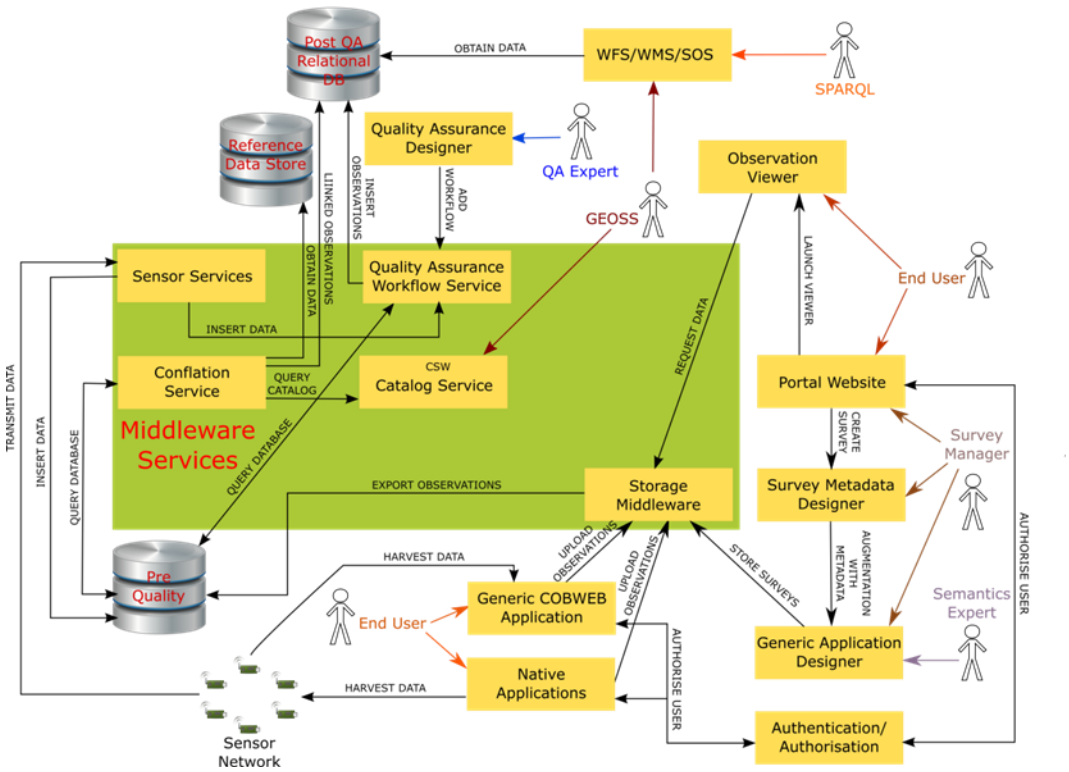
Actor	**Role**
Principal Investigator	Sponsor for a specific crowdsourcing campaign.
Project Coordinator	Lead from the perspective of the community carrying out the task
Administrator	Maintains instance of the COBWEB system. Where the software has been installed to meet the needs of a particular organisation.
Survey Manager	Defines, sets up surveys and invites citizens where appropriate
Quality Reviewer	Configures quality assurance processes in accordance with needs of the sponsor
Citizen	Individuals participating in collecting data. May be known or anonymous (see section 4.2 for different kinds of surveys: public, registered or private)
Consumer	Potentially a large range of actors interested in results
Publisher	Actors with the authority and capability of making crowdsourcing results available
Semantics Expert	Individuals that link information to survey terms to help explain their meaning

Quality control may require diverse input (section 4.6). For example; it may be advantageous or necessary to have data from a variety of sensors; these in turn can be a combination of physical sensors embedded in the environment, cyber sensors leveraging online data, or social sensors leveraging social media streams (see section 4.7). Similarly, it may desirable to configure access to authoritative data from SDI type initiatives such as GEOSS, INSPIRE network services and national SDI's. This may be necessary for conflation purposes (relating and combining spatial data from various sources); as part of quality control or to assist with visualising the results.

4. ARCHITECTURE

The system architecture (Figure 2) necessary to support the above was developed through a combination of rapid prototyping and new software builds informed by requirements derived through co-design activities and other methods of stakeholder engagement. The architecture consists of the following key components: portal website, generic application designer, apps, storage middleware, quality assurance and conflation, sensor networks, whilst also implementing access control and privacy, and open standards. It is a feature of COBWEB that the components we have developed embed within an architecture that uses existing open source projects

Figure 2: An Overview of the COBWEB Architecture, and the High Level Interactions between Components



4.1. Portal Website

Portal websites are the main point of entry to COBWEB instances, providing access to the functionality documented in the sections below. In the simplest case, this is where users login and request to join selected citizen an example is currently integrated with the Dyfi Biosphere Reserve website for demonstration purposes, science surveys. Once users have contributed observations via their mobile device, results are available for visualisation via the portal using Web Map Services (WMS), Web Feature Services (WFS) or Sensor Observation Services (SOS).

4.1.1. GeoNetwork Opensource

At the heart of the portal is the latest version (3.0) of GeoNetwork5, extended under COBWEB to facilitate typical citizen science use cases. GeoNetwork is an open source catalogue application offering resource registration capabilities and the ability to expose those resources for discovery via a number of standards, most notably those of ISO/TC 211 and the Open Geospatial Consortium (OGC), e.g. the OGC Catalogue Services for the Web (CSW) standard (Nebert et al, 2007). Metadata can be exported to other standards such as Dublin core, DCAT, and schema.org.

4.1.2. Survey Discovery

After having been created, surveys can be discovered can by any interested parties via the portal. Making surveys discoverable using accepted and widely used open standards helps avoid waste, duplication, helps recruit volunteers and facilitates research and development in the field.

COBWEB is continuing work on developing an understanding of how existing initiatives in the data sharing domain, e.g., CSW, ISO19139, DCAT , schema.org, may be used or adapted to describe citizen science projects, thereby leveraging substantial pre-existing investment in standards and existing tooling like GeoNetwork. Collaborative work, e.g., creation of a crosswalk (mapping metadata elements from different schemas) to PPSR_CORE (2016), is underway with a variety of citizen science umbrella organisations to better understand how best to meet the citizen science community's needs

4.1.3. Data Discovery

Making citizen science data discoverable online, within communities like GEOSS and INSPIRE, is more challenging than project or survey discovery and there are strong overlaps with the work underway within COBWEB developing a harmonised common data model (see section 6) and assessing data quality (see section 4.6). Questions addressed by COBWEB include:

1. What level of aggregation is used to determine quality: dataset or feature level?

Many user groups are interested in datasets as a whole ('what is the spatial distribution of that species'). Other users may be interested in selected features (meaning abstractions of real world phenomena – buildings, rivers, plants, animals) and want to put questions such as 'when did that particular observation of the

[5] http://geonetwork-opensource.org/ [accessed 23 March 2016]

species of interest occur'. Quality in COBWEB is therefore determined at both the feature and dataset level.

2. How is the quality of datasets and/or features represented?

Quality varies at the feature level, eg, different observations in the same survey may be of different quality, it may be less certain that a particular species has been observed even though the observer is the same. Quality is therefore best represented as part of the data and we recommend storing quality elements as part of the data.

3. Is dataset quality assessed by aggregating the quality level of each of the observations in the dataset?

Some quality elements evaluated at feature level can be aggregated at dataset level for rapid assessment. Some quality only makes sense at the feature level. The ability to visualise quality cartographically may be useful in many circumstances.

4. What schema is being used?

For metadata the iso19139 and iso19157 schemas are used, but they can be exported on request to Dublin Core[6], DCAT[7] (Data Catalog Vocabulary) or schema.org[8]. For datasets a generic SWE4CS model is proposed (see section 6), which allows extension to include any external ontologies.

4.2. Access Control and Privacy

Citizens can participate in COBWEB surveys anonymously (for reasons of privacy and data sharing this can be desirable) or they can register by providing identity information to user management systems integrated with the portal. Whether they do so or not depends on the survey; three kinds of surveys are distinguished in COBWEB:

1. *Public survey*. Any user, anonymous or registered, can join and contribute.

2. *Registered survey*. Only registered users can see the survey in the portal listings and contribute. An example of where this might be required is where the identity of the user making the observation is being used as a proxy for the quality of the observation, e.g., where the observer is expert in the field of study

[6] http://dublincore.org/ [accessed 23 March 2016]
[7] https://www.w3.org/TR/vocab-dcat/ [accessed 23 March 2016]
[8] https://schema.org/ [accessed 23 March 2016]

and recognized as such by the project coordinator and/or the principal investigator.

3. *Private survey*. Invited registered users only

The need for this level of control was identified at the beginning of the design process where initial stakeholder engagement revealed a desire to be able to make and share observations of protected species and consequently, a requirement to be able to control access to sensitive data; for example, species protected under the UK Wildlife and Countryside Act or listed in the Natural Resources Wales Sensitive Species List9.

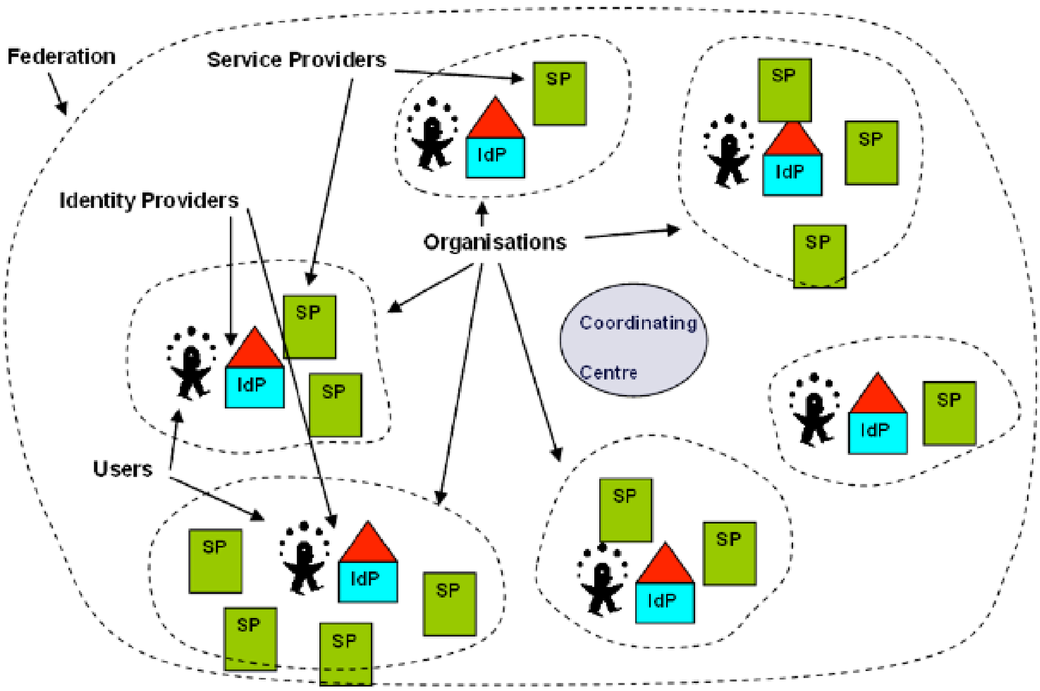
It is not desirable or permitted to make publically available over the web detailed information on the location of these species without any access control. Conversely, these are often the most valuable data for consideration in environmental monitoring, and exactly the kind of information most needed for management and policy purposes.

The need for a reusable, well understood, high TRL, solution to security interoperability (sharing restricted data using standards across organizational boundaries) is a common SDI requirement (Higgins et al, 2012). Throughout the extensive stakeholder engagement conducted by COBWEB, we found this requirement recurring over and over again in citizen science scenarios. The corollary, existing solutions that do not take security (interoperable or not) into account is also commonplace. The private survey concept developed by COBWEB is a powerful feature of the platform.

In addition to data security, concomitant questions of privacy were also a requirement within COBWEB. It is essential to enable users to register using personal information so that decisions concerning what they are authorised to access and contribute towards can be made. In the current citizen science landscape, identity information is also frequently used for quality assurance purposes and verifying observations.

[9] http://www.lrcwalesdat.org/eng/shared_files/help_sensitive_species.php [Accessed 9 June 2016]

Figure 3: Key Roles within a Typical SAML Access Management Federation



Based upon the OASIS Security Assertion Markup Language (SAML) standard, the key advantages of this approach are that it is a proven, high TRL, industry strength solution that allows Single Sign On (SSO) to protected web based resources across administrative domains. Users can login using their familiar organisational credentials (presented via a SAML Identity Provider (IdP)) and then access protected resources (typically OGC web services in SDI scenarios) presented as SAML Service Providers (SP). Users only need to login once (SSO) and can then access multiple protected resources in the federation, providing they are suitably authorised.

Fine grained access control is established based upon attributes of the users and the actions on resources they want to do. The OGC's GeoXACML standard, which extends the OASIS eXtensible Access Control Markup Language (XACML) to enforce geo-specific constraints, is another mature standard with high TRL guarantee. In COBWEB different levels of access are enforced to ensure the privacy of citizens that collected information and also the wellbeing of the observed species.

For the COBWEB framework, this means that surveys can be configured which access both unprotected and protected data sources, e.g., for conflation and/or quality assurance purposes. Private surveys can be setup which enable citizens to collect and share protected data with public authorities in compliance with both data protection and wildlife protection legislation. Public authorities can leverage the full benefits of interoperability, for example, by sharing sensitive data across organisational boundaries with authorised individuals using protected OGC web services, and potentially accessing all citizen sourced data without recourse to non-interoperable, one-off, bespoke security point solutions or lossy mechanisms such as anonymisation, obfuscation, reduction of spatial resolution, etc. The interoperability advantages of this approach are documented more fully in (Higgins et al 2012) and (ARE3NA, 2016).

4.3. Generic Application Designer

For each community, there is at least one 'Survey Manager' whose privileges entitle them to setup and create surveys. COBWEB employs a hybrid App approach enabling survey managers to build custom data collection forms using the generic application designer component of the portal website. Forms designed at the portal can then be synchronised with the generic COBWEB application on individual users mobile devices (Butchart 2013). The generic and extensible application designer supports a wide variety of form elements to cater for a broad range of user requirements.

Similar framework approaches to creating citizen science projects are not uncommon (COBWEB, 2016c) as exemplified by initiatives such as Open Data Kit, Indicia and EpiConnect amongst others. Some of the aspects that help distinguish COBWEB are its flexible and powerful approach to automating quality control using multiple inputs, use of an interoperable security solution and adherence to open standards wherever possible.

4.4. Mobile Applications (Apps)

The generic Application solution offers the citizen the ability to install onto their mobile device, and login using an appropriate identity provider (Figure 3). The citizen then has the ability to either contribute to a directly available survey (public or registered) or a private survey which they have been invited to by the Survey Manager. The citizen will then be presented with the form designed by the Survey Manager, allowing them to participate in data collection with or without network coverage.

In addition to this, the generic capability described above is complemented by functionality allowing the cacheing of high quality basemapping on individual handsets for use in areas of poor or no network coverage. The background mapping used by COBWEB in the UK has been created using a variety of open products (mostly OpenStreetMap and Ordnance Survey Open Data Products) cartographically optimised for zooming in and out on mobile devices in both urban and rural areas (Butchart, 2013). Investigation using Greek data has illustrated the difficulty of replicating this rich open map stack across Europe.

To demonstrate the effective ‘separation of concerns’ in the architecture, and how the COBWEB framework can be used in scenarios where lower level access to inbuilt mobile device functionality is required, a native Application (Figure 2) in the flooding thematic area case study area was also developed. This uses the same interface as the generic application for communicating with the ‘Storage Middleware’.

4.5. Storage Middleware

Storage Middleware receives the observations from the App. As long as Oauth v2 authorisation is supported, the Storage Middleware component provides a generic REST-based API accessed storage compatibility layer on top of a range of cloud based providers (Google Drive, Dropbox, etc.) or physical storage media where local storage is required.

Storage Middleware is a central component of the COBWEB architecture used for managing survey schemas and exporting geospatial observations to the desired encodings, e.g. KML (Keyhole Markup Language), Geopackage, Shapefile, GeoJSON, CSV, etc. By synchronising all stored information with a relational database (the pre-quality PostGIS database in Figure 2) export of data via OGC Web Services (WMS/WFS) is supported.

4.6. Quality Assurance and Conflation

Since the emergence of citizen science as a means to support scientific research, data quality is considered an important issue for data use. Quality assurance accordingly plays an important role for the analysis of data obtained from citizen science (Roy et al, 2012) and for the set-up of citizen science projects (Bonney et al, 2014). It is commonly agreed that an unknown data quality makes citizen science data of limited use. COBWEB’s approach on quality assurance is twofold: first, a number of quality measures are determined based on the variety of data provided by crowdsourcing activities, observations from the co-design projects, sensor feeds and social media; secondly means to express those measures in the metadata of observations are investigated. This also includes an approach to link and conflate observations with relevant external datasets on the Web.

To allow for a customizable and dynamic quality assurance, which remains independent from actual citizen science projects, COBWEB uses a standards based web service chaining approach (the QA Workflow Service in Figure 2), and thus enables survey designers to adapt and adjust the quality assurance process to their needs. This generic approach on QA is necessary, because the relevance of specific quality control processes is often highly use case dependent. The solution extends a pre-existing typology of quality assessment types defined by Goodchild and Li, 2012 to seven categories (or pillars) covering a range of specific quality controls generating quality metadata elements (Meek et al. 2014, Leibovici et al. 2015b).

Table 3: The 7 Pillars of Quality Controls in Citizen Science

Pillar number & name	Pillar description
1.LBS-Positioning	Location, position and accuracy: Location-Based-Services focusing on the position of the user of the targeted feature (if any), local condition or constraints, e.g. authoritative polygon, navigation, routing, etc.
2.Cleaning	Verification, erroneous entries, mistakes, malicious entries: Erroneous, true mistakes, intentional mistakes, removals, corrections are checked for the position and for the attributes. Feedback mechanism can be an important part of this pillar if the mistakes can be corrected.
3.Automatic Validation	Simple checks, topology relations and attribute ranges: Carries further the cleaning aspects by validating potential good contribution. This aim is more positive than with cleaning and may keep as outlier a given captured data rather discarding it.
4.Authoritative Data Comparison	Comparison of submitted observations with authoritative data: Either on attributes or position performs statistical test, (fuzzy) logic rule based test qualifying the data captured or reversely qualifies the authoritative data. Knowledge of the metadata of the authoritative data is paramount.
5.Model-Based Validation	Utilising statistical and behavioural models: Extends pillar 4 testing to modelled data coming e.g. physical models, behavioural models, other user contributed data within the same context. This may use intensively fuzzy logics and interactions with the user within a feedback mechanism of interactive surveying.

	(if some tests will be similar to pillar 4 the outcome in quality elements can be different)
6. Big/Linked Data Analysis	Data mining techniques and utilising social media outputs: Extends pillar 5 testing to using various social media data or related data sources within a linked data framework. Tests are driven by a more correlative paradigm than in previous pillars.
7.Semantic Harmonisation	Conformance enrichment and harmonisation in relation to existing ontologies: Level of discrepancy of the data captured to existing ontology or crowd agreement is transformed into data quality information. In the meantime data transformation to meet harmonisation can take place.

The implementation of the quality assurance process is based on the Business Process Model and Notation[10] (BPMN) in combination with the OGC Web Processing Service (WPS) standard. This combination allows for the definition and dynamic binding of atomic quality control processes, encapsulated by WPS interfaces, to be used in a workflow environment. Therefore, the JBPM[11] suite (workflow editor and workflow engine) has been customized to work with OGC services (Meek et al., 2015) and has been integrated as a component of the COBWEB portal. Thus, each survey manager has the authority to create quality assurance workflows for their particular survey.

Capabilities to link and conflate citizen science observations with external data are designed and implemented to 1) assist the quality assurance process (Wiemann et al. 2015) and 2) infer spatial information from the identified spatial data relations (Wiemann and Bernard,2016, Wiemann, 2016). Since all conflation processes are also offered via the OGC WPS interface, a seamless integration and use by the workflow engine is assured. Currently, identified links of an observation can be attached as additional attributes or stored as RDF (Resource Description Framework) triples in a Linked Data store. Whereas the first option is primarily used to enrich and validate observations, the latter option allows for the reasoning on spatial data relations by means of the Semantic Web, in particular by the

[10] http://www.bpmn.org/ [accessed 23 March 2016]
[11] http://www.jbpm.org/ [accessed 23 March 2016]

application of SPARQL (SPARQL Protocol and RDF Query Language) queries (Wiemann, 2016).

4.7. Sensor Networks

Though data collection from mobile devices is fundamental to the generic COBWEB framework, a variety of sensor platforms, monitoring multiple environmental parameters, within the Dyfi Biosphere Reserve testbed area, have been deployed. These sensor networks have been developed mainly as a result of dialogue between research scientists working on COBWEB and the co-design partners.

There are two ways by which sensor data are incorporated into the storage middleware:

1. *Automated sensing.* The physical sensors embedded in the environment are augmented with telemetry capabilities that enable them to operate in an autonomous manner, without human intervention, in transmitting data to a base station.

2. *Participatory sensing*. Data collection by citizen participating in surveys when in close proximity to suitably equipped sensors. The survey participant's mobile device connects directly to the sensor using short range communication and the citizen physically relays the data back to a position where the data can be uploaded to the COBWEB servers (O'Grady, et al., 2016).

4.7.1. Physical Sensor Platforms

Both participatory and automated approaches are being researched and developed within COBWEB. Some of the Wasp Motes listed in table 4 are being used for participatory sensing. Surveys are configured such that when a citizen comes within range of one of the Wasp Motes, it is discovered by the citizen's mobile device. When an observation is made, the mobile communicates with the sensor via the Bluetooth LE standard and the environmental parameters of interest are simultaneously captured. Such data can be used in quality control or to otherwise supplement the metadata associated with that observation.

Table 4. Sensors Deployed in COBWEB

Sensor type	Measuring
In-Situ sensors	water depth levels, pressure (mBar), pressure (PSI), temperature
Wasp Motes	water depth levels, salinity (g/kg)

Davis Weather stations	various meteorological parameters
phone-based sensors	ambient temperature (C), pressure (mbar), accelerometer readings (radians)

The In-Situ and Davis Weather station sensors are capturing data automatically and transmitting back to the COBWEB servers. Data from the In-Situ sensors are being used in association with habitat reversion at co-design partner RSPB's Yns-Hir reserve. Data from the weather stations are being used in educational scenarios with co-design partner Ysgol Bro Hyddgen and to investigate pollinator scenarios with co-design partner Penparcau Community Forum. In the pollinator scenarios, citizens go out on a regular basis to record observations of butterflies and their larval foodplants around Penparcau in the Dyfi Biosphere Reserve. The data recorded includes plant/butterfly species, plant coverage, butterfly activity, and habitat information.

Weather data (temperature, wind speed etc.), from weather stations, sited in strategic locations, is used in a Twitter alert service to send tweets to citizens following a dedicated Twitter account to encourage participants to go out and collect data when the weather is suitable for butterfly activity. The weather data is also used post-data collection to validate and add value to the observations.

When collecting data within the field, sensors on board the user's mobile device are harnessed to provide additional information. For example, accelerometers values are recorded when the user is taking photographs to enable the line of sight to be determined - useful for various quality assurance purposes. It should be noted that the generic COBWEB App accesses this functionality via a Cordova12 plug-in while the native flooding App obtains the data directly from the sensor.

4.7.2. Cyber Sensors

Cyber sensing refers to the harvesting of online geotagged sensor data from a variety of sources through different web-based API's. In COBWEB, the approach has been demonstrated using data accessed through the Shoothill13 (UK river levels) and Weather Underground14 (network of personal weather stations) API's.

There are two ways (Figure 2) in which sensor data are associated with surveys within COBWEB. In the case of participatory sensed data, sensor data are included as part of the survey data and exported to the pre-quality database via the storage middleware in the usual manner. Otherwise, for both cyber and physical sensors,

[12] https://cordova.apache.org/ [accessed 24 May, 2016]
[13] http://www.shoothill.com/environment-agency-liveapi/ [accessed 20 March, 2016]
[14] https://www.wunderground.com/weather/api/ [accessed 20 March, 2016]

the data are stored directly in the pre-quality database either by sensor nodes transmitting the data in the case of physical sensors or by the harvester in the case of cyber sensors.

4.7.3. Social Sensors

In respect of leveraging social media streams in COBWEB, research and development has concentrated on the use of Twitter and Flickr API's. Data from both social media platforms are acquired by two methods.

Bounding boxes for Biosphere Reserves are used to monitor and capture any social media postings within these areas.

Keywords are used to identify relevant postings worldwide.

Once social media postings have been captured they are stored in a database and are available for viewing at the portal or potentially for use within a Quality Assurance workflow.

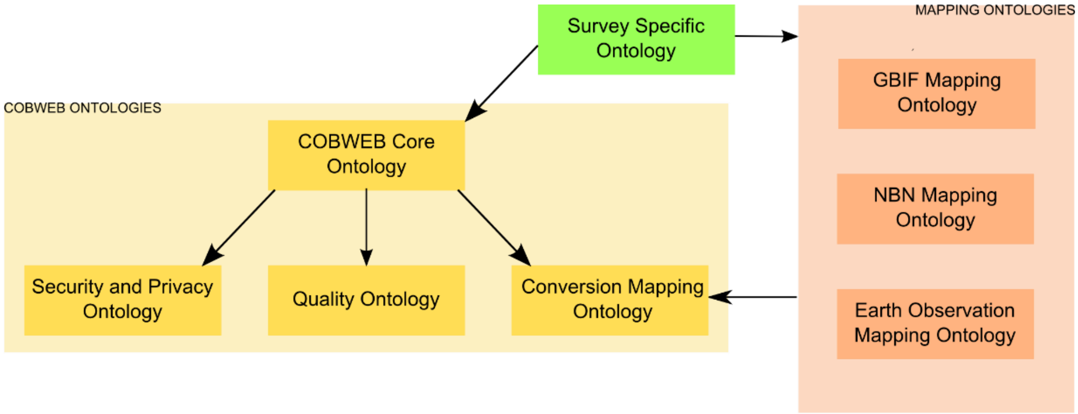
5. SEMANTICS

Semantics is concerned with attaching unambiguous meaning to terms and concepts, typically, using ontologies (formal models describing semantics). In the citizen science domain, as in others, this is very important as understanding exactly what was observed and in which context is essential to making the data comprehensible; both for the immediate use, by others and in the future. The ability to share data, reuse, integrate and compare is often compromised by a lack of semantic interoperability. Semantic Web standards, e.g., from the World Wide Web Consortium (W3C), have been developed to address this problem.

As would be expected in a research project whose main objective is developing a generic infrastructure that aims to satisfy requirements across as broad a range as possible of the multitude of potential citizen science scenarios, semantics is an area of intense research and development interest in COBWEB.

In terms of the COBWEB workflow (Figure 1), semantics can be injected at various stages, but the best place to introduce semantics is when the survey context is defined. At this point, the goals of the project should be clear, and if not, then it is time well spent clarifying exactly what is expected to be observed and to what extent exact definitions can be incorporated. Ideally, the Survey Manager would liaise with the Principal Investigator and possibly also the Quality Reviewer and other experts as appropriate, to incorporate sufficient semantics to ensure that the information generated by the survey is of maximum usefulness to end Consumers.

Figure 4: Ontological References for Specific Surveys



In practice, this means a simple Survey Specific Ontology (Figure 4) is defined, which maps onto the COBWEB model. As there are many areas where semantic descriptions apply, e.g., privacy and security, quality, sensors, conversion mapping, etc., the approach taken has been to split the overall semantic model into a number of ontologies.

The Conversion Mapping Ontology enables the final step; transformation to external formats as required by end consumers. The Global Biodiversity Information Framework (GBIF) 15 and National Biodiversity Network (NBN) 16 examples were high priority targets identified through COBWEB's stakeholder engagement work package.

To conclude and provide an example, semantics are best captured at the point of survey creation. In the example of the Snowdonia National Park Authority co-design sub-project (Table 1), the field/attributes in the survey created which are relevant to biodiversity were mapped using the survey specific ontology (Figure 4) to terms in Darwin Core - a biodiversity standard maintained by TDWG17 and used by many projects globally, including GBIF. The conversion mapping ontology is used to map to, and produce, output compliant with NBN exchange format. The latter is widely used in the UK and the NBN is the UK GBIF node18.

---

[15] http://www.gbif.org/ [accessed 20 March, 2016]
[16] http://nbn.org.uk/ [accessed 20 March, 2016]
[17] http://www.tdwg.org/ [accessed 10 June, 2016]
[18] http://www.gbif.org/participation/participant-list [accessed 10 June 2016]

The COBWEB approach to the complexity inherent in semantics has been to research and develop a solution which is not tied to one particular country or even one particular domain. The aim is to see how far we can get in the current landscape with producing a useable, generic and flexible solution that accommodates different levels of complexity depending upon circumstances.

6. STANDARDS AND SPATIAL DATA INFRASTRUCTURES

COBWEB aims to maximise technical interoperability by the use of standards wherever appropriate - the sections above make multiple references to standards from several standards defining organisations such as, OGC, ISO/TC 211, OASIS, W3C and TDWG.

All the citizen observatory's developed under the FP7 call which funded COBWEB were required to make data collected available within the GEOSS without restriction.

Publishing data into SDI-like initiatives such as GEOSS was addressed using a cooperative approach (based mainly on the processes of the OGC) with the broader geospatial/citizen science community to develop a profile of the relevant OGC standards (swe4citizenscience 2015). This has resulted in progressing a vision of a harmonised common information model applicable to a wide range of citizen science scenarios. The idea is that data compliant with this information model can be discovered and accessed through standardised web interfaces, e.g., OGC web services, and integrated with SDI's as most, if not all, SDI's are based on open geospatial interoperability standards.

Two approaches are currently being explored within COBWEB (Simonis, 2016). First, the definition of an application profile by specialising the Observations and Measurements (O&M, ISO, 2011), SensorML (Botts, 2014) and SweCommon (Robin, 2011) information models. Second, the definition of an O&M, SensorML and SweCommon encodings best practices to allow more efficient reuse of existing components.

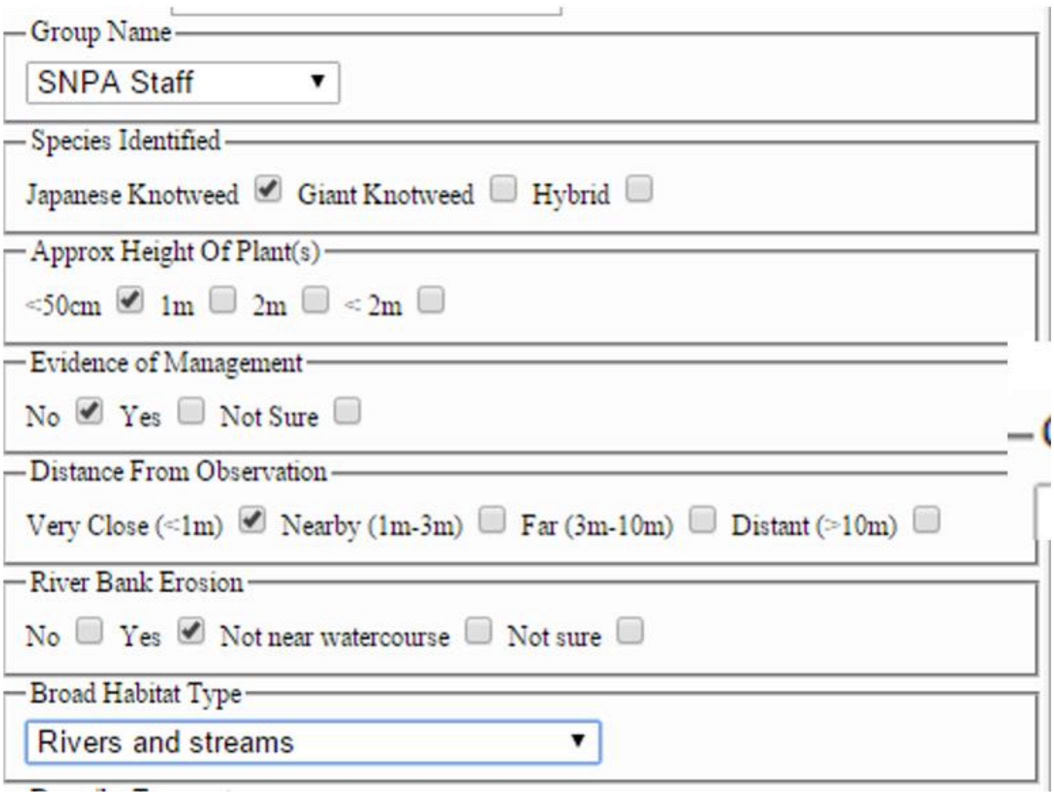
Strong semantics are supported via the mechanism indicated in section 5. Survey specific ontologies are included in the survey definition so that properties of the observed feature are qualified in terms of resolvable URL's. For example, the invasive species survey created by co-design partner Snowdonia National Park (Table 1) includes the form below (Figure 5).

One of the properties of the feature observed is the species name; Japanese Knotweed, Giant Knotweed or Hybrid. To be unambiguous and enable semantic interoperability, a link needs to be provided to an authoritative definition. There are well known and used national resources such as the UK species inventory19, but no single global authoritative definition. COBWEB tried to leverage this problem by defining a temporary SKOS-based reference system that can be replaced once robust and reliant definition endpoints become available. Additionally, the Survey Specific Ontology created references Darwin Core, ontological resources used by GBIF providing the necessary detail on species taxonomy. Still, availability and management of key resources remains an issue that needs to be solved by the Semantic Web community.

Note that this approach is independent of serialisation; the Citizen Science application profile enables encoding as XML (Extensible Markup Language), GeoJSON or linked data. The observations and raw data are of primary interest, various levels of processed or aggregated data may be required at different times depending upon circumstances. All types of observations are supported; including raw observations and those which are derived or result from subsequent quality assurance or conflation processing. A provenance model is included to provide information on the history of each observation.

[19] http://www.nhm.ac.uk/our-science/data/uk-species.html [accessed 10 June 2016]

Figure 5: Data Collection Form Authored for Invasive Species Survey



7. SUMMARY AND CONCLUSIONS

COBWEB is conducting research into the feasibility of creating a common generic framework for mobile device apps for use in citizen science for environmental monitoring, using the UNESCO World Network of Biosphere Reserves as a testbed. We have demonstrated that, in principle, it is possible and brings multiple benefits when starting from a position where the importance of interoperability and flexibility is paramount. This stands in contrast to the current proliferation of Citizen Science Apps generating silos of data.

Specifically, it has been shown that creating a generic solution to automating quality control and assurance which is sufficiently flexible to address the huge range of potential scenarios is beneficial to the reuse of citizen science data. Further development would result in the ability to make very large volumes of data useable, and is an area of potential future research.

Similarly, further research and development is needed to address whether it is possible to create an easily useable framework which is sufficiently flexible to allow a broad range of different kinds of familiar semantic resources to be employed in designing surveys before citizens go into the field. Without this, despite post-processing server-side, continued problems associated with a lack of semantic interoperability may be anticipated.

Despite perceived complexity and proliferation, the use of open interoperability standards still presents the most realistic chance of preventing the waste of resources and reuse opportunities inherent in creating silos of data locked into proprietary solutions. In the citizen science domain, as in other domains that leverage SDI, the lack of a widely adopted solution to security interoperability is again apparent as a major barrier. At the technical level, COBWEB has shown that access management federations work; progressing this relies upon action at the organisational, legal and political level – outwith the scope of COBWEB.

However, standardisation efforts should continue and adherence be required to help realise investment in SDI type initiatives such as GEOSS.

If agreement can be reached on a harmonised common information model, with sufficient community support, most, if not all, crowdsourced, citizen science type data could be made compliant with the developing standardisation efforts initiated under COBWEB. The immediate prize would be a boost to the usefulness of these data by reducing integration costs and enabling the myriad of potential consumers of such data to exploit existing standards based tooling and develop new standards based solutions on top. The long term curation value of holding citizen sourced data at the observation level of granularity compliant with a well thought out and agreed information model is to maximise potential future access and re-use.

The amount of citizen sourced environmental data is increasing dramatically and clearly has the potential to be of use for policy formation and delivery in respect of environmental decision making. At the global scale, towards the end of 2015, the UN adopted the 2030 Agenda for Sustainable Development, this included a set of 17 Sustainable Development Goals20 (SDG) to end poverty, fight inequality and injustice, and tackle climate change by 2030. The importance of geospatial information in monitoring progress with SDG targets is explicitly recognised in clause 76 of the official resolution (United Nations, 2015). It may be argued that citizen sourced information has a potentially important contribution worthy of further research.

[20] https://en.wikipedia.org/wiki/Sustainable_Development_Goals [accessed 20 March 2016]

In this respect, based on experiences in COBWEB, we would recommend increased focus on the UNESCO World Network of Biosphere Reserves (WNBR) as a testbed for sustainability science related research and development. Using co-design processes and consortia such as those employed in COBWEB, WNBR mechanisms can be leveraged to more easily mobilise sufficient numbers of enthusiastic citizens in order to help realise meaningful outcomes based on real world needs.

The COBWEB framework solution described above is complex and significant further work is required to present only the required level of complexity, depending upon circumstances, when required. Integrating all the components (at different TRL's) to create an actual system proven in an operational environment (TRL 9) is beyond the scope of this ongoing research project. At project completion, where applicable, source code will become available21 to the wider community for further use and development.

ACKNOWLEDGEMENTS

This project has received funding from the European Union's Seventh Programme for research, technological development and demonstration under grant agreement No 308513.

REFERENCES

ARE3NA (2016) ARE3NA study on AAA for Data and Services, at https://joinup.ec.europa.eu/asset/are3na-aaa/description and https://ies-svn.jrc.ec.europa.eu/projects/aaa-access-control [accessed 16 Mar 2016]

Bonney, R., Shirk, J.L., Phillips, T.B., Wiggins, A., Ballard, H.L., Miller-Rushing, A.J. and Parrish, J.K. Next Steps for Citizen Science, Science, Vol. 343, Issue 6178, pp. 1436-1437, at http://science.sciencemag.org/content/343/6178/1436.full [accessed 9 June 2016]

Botts, M and A. Robin (2014). SensorML: Model and XML Encoding Standard. OGC 12-000, at https://portal.opengeospatial.org/files/?artifact_id=55939 [accessed 19 March 2016]

Butchart, B., Pope, A., King, M., Hamilton, G., Terzis, P. and M. Koutroumpas (2013). Fieldtrip GB: Creating a customisable mapping and data capture app for the HEFE community, at

[21] https://cobwebproject.eu/news/publications [accessed 9 June 2016]

http://gogeo.blogs.edina.ac.uk/files/2013/04/gisruk2013_submission_85.pdf, [accessed 15 October 2015]

COBWEB (2016a) Citizen Observatory Web Co-Designers, at https://cobwebproject.eu/about/co-designers [accessed 13 Mar 2016]

COBWEB (2016b) D2.1 Annex C: Review of Co-Design Fund sub-project Proposals, at https://cobwebproject.eu/news/publications/deliverables [accessed 13 Mar 2016]

COBWEB (2016c) D4.1 Mobile data collection system specification and design, at https://cobwebproject.eu/sites/default/files/COBWEB%20D4.1%20v2.pdf [accessed 16 Mar 2016]

COBWEB (2016d) Co-Design Reports, at https://cobwebproject.eu/news/publications/co-design-reports [accessed 9 June 2016]

EARTO (2014) The TRL Scale as a Research & Innovation Policy Tool, EARTO Recommendations, at http://www.earto.eu/fileadmin/content/03_Publications/The_TRL_Scale_as_a_R_I_Policy_Tool_-_EARTO_Recommendations_-_Final.pdf [accessed 7 Jane 2016]

Goodchild, M.F. & Li, L. (2012) Assuring the quality of volunteered geographic information. Spatial Statistics. [Online] 1 (0), 110–120. Available from: doi:10.1016/j.spasta.2012.03.002 [Accessed: 26 June 2012].

Higgins, C.I., Williams, J., Leibovici, D.G., Simonis, I., Davis, M.D., Muldoon, C., O'Grady, M. (2016) "Citizen OBservatory WEB (COBWEB): A Generic Infrastructure Platform to Facilitate the Collection of Citizen Science data for Environmental Monitoring", in Berre A. J, Schade, S., Piera, J. (Ed.) ENVIP 2015. Proceedings of the Workshop Environmental Infrastructures and Platforms 2015 - Infrastructures and Platforms for Environmental Crowd Sensing and Big Data, Barcelona, October 28-30 October, 2015, pp. 12-19. CEUR Workshop Proceedings (in press).

Higgins, C.I., Koutroumpas, M., Matheus, A., Seales, A. (2012) Shibboleth Access Management Federations as an Organisational Model for SDI, International Journal of Spatial Data Infrastructures Research, 2012, Vol.7, 107-124.

ISO (2010) Ergonomics of human-system interaction -- Part 210: Human-centred design for interactive systems, ISO 9241-210:2010

ISO (2011) ISO 19156: 2011-Geographic Information: Observations and Measurements.

Leibovici, D.G., Evans, B. Hodges, C. Wiemann, S. Meek, S. Rosser, J. and Jackson, M. 2015 On Data Quality Assurance and Conflation Entanglement in Crowdsourcing for Environmental Studies. International Symposium on Spatial Data Quality (ISSDQ 2015), 29-30 September 2015, La Grande-Motte, France, 4 pages abstract.

Leibovici, D.G. Meek, S. Rosser, J. and Jackson, M. (2015b) DQ in the citizen science project COBWEB: extending the standards. Data Quality DWG, OGC/TC Nottingham, September 2015, U.K.

Leibovici D.G., Williams J., Rosser J.F., Hodges C., Scott D., Chapman C., Higgins C., and Jackson M.J (2016) The COBWEB Quality Assurance System in Practice: Example for an Invasive Species Study. ECSA conference 19-21 May 2016, Berlin. (in press)

Liu, H., Kobernus, M., Broday, D and Bartonova, A. (2014), A conceptual approach to a citizens' observatory - supporting community-based environmental governance. Environmental Health 2014, 13:107

Meek, S., Jackson, M. and Leibovici, D.G. (2014) A flexible framework for assessing the quality of crowdsourced data. 17th AGILE conference on Geographic Information Science, 3-6 June 2014, Castellon, Spain

Meek, S Jackson, M and Leibovici, DG (2015) Addressing the quality assurance challenge for location-based crowd-sourced data through workflow composition of OGC web services. Computers & Geosciences (submitted)

Nebert, D., Whiteside, A. and Vretanos, P (Eds). (2007). OpenGIS Catalogue Services Specification version 2.0.2, Open Geospatial Consortium

Newman, G., Wiggins, A., Crall, A., Graham, E., Newman, S. & Crowston, K. (2012). The future of citizen science: emerging technologies and shifting paradigms. Frontiers in Ecology and the Environment 10, 298–304.

O'Grady, M.J., Muldoon, C., Carr, D., Wan, J., Kroon, B. and O'Hare, G.M., (2016). Intelligent Sensing for Citizen Science. Mobile Networks and Applications, pp.1-11.

PPSR_CORE (2016), Citizen Science Association http://citizenscienceassociation.org/2015/06/11/efforts-underway-to-share-project-metadata/ [Accessed 8 Jan 2016]

Robin, A (2011). OGC SWE Common Data Model Encoding Standard. OGC 08-094r1. http://portal.opengeospatial.org/files/?artifact_id=41157 [Accessed 19 March 2016]

Roy, H.E., Pocock, M.J.O., Preston, C.D., Roy, D.B., Savage, J., Tweddle, J.C. & Robinson, L.D. (2012). Understanding Citizen Science & Environmental Monitoring. Final Report on behalf of UK-EOF. NERC Centre for Ecology & Hydrology and Natural History Museum, at http://www.ukeof.org.uk/our-work/citizen-science [Accessed 8 June 2016]

Simonis, I. (2016). "Standardized Information Models to Optimize Exchange, Reusability and Comparability of Citizen Science Data. A Specialization Approach", in Berre A. J, Schade, S., Piera, J. (Ed.) ENVIP 2015. Proceedings of the Workshop Environmental Infrastructures and Platforms 2015 - Infrastructures and Platforms for Environmental Crowd Sensing and Big Data, Barcelona, October 28-30 October, 2015, pp. 92-101. CEUR Workshop Proceedings (in press).

swe4citizenscience (2015) swe4citizenscience at https://github.com/opengeospatial/swe4citizenscience [Accessed 10 October 2015]

United Nations (2015) Transforming our world: the 2030 Agenda for Sustainable Development. Resolution adopted by the General Assembly on 25 September 2015. A/RES/70/1, at http://www.un.org/ga/search/view_doc.asp?symbol=A/RES/70/1&Lang=E [accessed 20 March 2016]

Wiemann, S., Meek, S.M., Chapman, C. Leibovici, D.G., Jackson, M. and Lars Bernard (2015) Service-based combination of quality assurance and fusion processes for the validation of crowdsourced observations. AGILE 2015, 7-9 June Lisbon, Portugal

Wiemann, S. & Bernard, L. (2015) Spatial data fusion in Spatial Data Infrastructures using Linked Data. International Journal of Geographical Information Science, 30(4), 613–636. doi:10.1080/13658816.2015.1084420

Wiemann, S. (2016) Spatial Data Relations as a Means to Enrich Species Observations from Crowdsourcing, Geospatial Data in a Changing World, Lecture Notes in Geoinformation and Cartography, Springer, DOI 10.1007/978-3-319-33783-8_8

Wikipedia (2016), Participatory design, at https://en.wikipedia.org/wiki/Participatory_design [Accessed: 11 March 2016]