Environmental voluntary groups: Towards curating data for sharing, access and preservation

Mary Anne Kennan
Charles Sturt University (CSU)
Locked Bag 450, Silverwater
NSW 2128 AUSTRALIA
mkennan@csu.edu.au

Kirsty Williamson
CSU & Monash University
P.O.Box 197, Caulfield East,
VIC 3145, AUSTRALIA
kirsty.williamson@monash.edu

Graeme Johanson
Monash University
P.O.Box 197, Caulfield East,
VIC 3145, AUSTRALIA
graeme.johanson@monash.edu

ABSTRACT

This paper reports a recent study which investigated what data were collected by members of an environmental voluntary group (EVG) and how these data were collected, stored, managed and shared. A significant aim was to understand how data management and approaches to data sharing could be improved in order to enhance the contributions of EVGs to research and to science more broadly while also continuing to meet individual and group needs. Interviews were conducted with members of the Australian Plants Society Victoria (ASPV) using a broadly ethnographic approach. Findings indicate that APSV members have a strong interest in conservation biodiversity, and in increasing their own, and society's knowledge and understanding, passions often shared with professional scientists. Yet their data are often poorly managed, creating significant impediments to sharing. The paper explores the major issues of data management and sharing and the resulting impediments to data sharing and information communication. Options for improvement are explored, especially ways to inspire and empower APSV members with skills and technology to contribute to major data repositories so that their valuable data may be preserved and made accessible beyond their immediate Society co-members.

Keywords

Data management, data sharing, data preservation, environmental voluntary groups

INTRODUCTION

Members of environmental voluntary groups (EVGs) collect data which have the potential to contribute to, and support, research and science well beyond the boundaries of their group. Yet few studies, and fewer within the information sciences, have explored data collected by EVG members, or how they are collected, stored, managed and shared. To some extent all volunteer groups share common data dilemmas. This paper reports the findings of a study in one data-collecting EVG, the Australian Plants Society Victoria (APSV). Insights from this study can be useful for repository developers, for example with regard to more structured guidelines and support to facilitate EVG member contributions to data repositories. The study was undertaken in the context of the collection of data by members of APSV, the increasing availability of data repositories, especially those seeking, or depending on, contributions from community groups, and the associated complex and vexed issues of managing data to be used in future for research purposes and to inform the community.

ASIST 2013, November 1-6, 2013, Montreal, Quebec, Canada.

The importance of data is widely recognized. The value of data is seen to increase as they are interconnected, networked, shared and used (Borgman, 2007); as does the usefulness of a good network design to accommodate data (Van House, 2003). The literature speaks of a 'deluge' of scientific and research data and the importance of capturing and managing them (using IT such as repositories) for use beyond the original community, purpose and time (Hey & Trefethen, 2003). For this paper, this 'deluge' largely refers to data collected, created or used for personal and group research purposes. Data sharing is integral to data-intensive science and is a valuable component of current scientific method. It includes the depositing of data and/or metadata in an accessible place and the preservation of data, although it is primarily associated with providing access for use and reuse (Tenopir et al., 2011).

The university and research sectors, particularly, have paid considerable attention to data and repositories (Kennan, Williamson & Johanson, 2012; Reichman, Jones & Schilhauer, 2011). The study reported here, aims to extend this conversation to data and information held outside of 'academic' science by EVGs. The data and information which EVG members collect are, at present, likely to be inaccessible to the research community, or anybody outside those often small EVGs. Indeed, the concept of 'dark data' (Heidorn, 2009) - dark because they are 'invisible' - appears to be appropriate to describe much of the data produced by EVGs. Yet people who conduct 'research in the wild' (outside of the academy and often not as part of

their professional work duties) now and, in the future, have much to contribute to science, research and participative democracy (Callon, Lascoumes & Barthe, 2009). While EVG members collect data to satisfy their own curiosity and interests, there is potential to harness data collected by these 'amateur experts' for more formal scientific purposes (Grove-White, Waterton, Ellis, Vogel, Stevens, & Peacock, 2007) and to generally raise community awareness.

Environmental and biological data sourced from volunteers and voluntary groups make an invaluable contribution to conservation. Volunteers provide detailed data about, for example, an area's biodiversity useful to both professional and volunteer researchers, contribute to biodiversity research, planning and management, and to conservation policy (Grove-White et al., 2007; Atlas of Living Australia, n.d.a). However, volunteers who collect data may not know the repositories and networks through which they might contribute their data; understand, or indeed be interested in, the requirements for making data reusable; or do not feel that their contribution is, or would be, valued (Grove-White et al., 2007). There are other problems. On the one hand, databases and repositories usually require that the data be validated and standardized, especially when recorders are unknown (Grove-White et al., 2007; Atlas of Living Australia, n.d.b). On the other hand, demands for validated and standardized data, possibly uploaded using new and/or complex software tools, can be a disincentive for volunteer data collectors, especially when made without consideration of their motivation and expertise.

In the United Kingdom networks and projects, such as UK Biodiversity Action Planning (Grove-White et al., 2007) and, in Australia, web-based data repository projects, such as "NatureShare", "Bowerbird", and the "Atlas of Living Australia"[1] have been developed for data and information sharing. These repositories offer a range of different open data, information and services. Individuals and organisations can contribute scientific observations and collections and use them for a variety of purposes. For example, researchers and conservation planners can use these data to track the location or spread of endangered or invasive species, ask what lives in a particular location, or where a particular species is found. Data from these repositories can be combined with other data, for example weather or climate data, to track changes over time, and can be interrogated using a variety of mapping and analysis tools (Lasalle, 2013). Volunteers are encouraged to contribute their observations and collections to such repositories thus making a valuable contribution to science.

At the time of writing, few APSV members were contributing. The question arises: "How can volunteers be encouraged and assisted to contribute to shared databases and repositories?" Before this question is answered, we need to understand what kinds of data are collected by volunteers, how their data are collected, stored, shared and managed, and what the data collectors in the EVGs perceive can be done to promote more efficient data storage, management and sharing. The study on which this paper is based addressed all those questions (Kennan et al., 2012). This paper focuses on issues concerned with data storage and management, as well as perceptions about how these can be improved for sharing, curation and preservation.

An important point about terminology needs to be made. There is much debate regarding definitions of terms (Pilerot, 2012). While we have used the term 'data' more often in this paper than the term 'information', we agree with Pilerot and Limberg, (2011) who saw information as including "data and documents that can be regarded as 'informative' and therefore as information" (p.314) when shared .

LITERATURE

This brief literature review first defines volunteers and summarizes their role in collecting data that might be useful if available more broadly to the scientific community. It then defines research data and discusses the importance of data sharing and preservation.

The contribution of volunteers

A volunteer is a person who freely offers to do something useful, or a person who works for an organization without being paid; technically it contrasted in meaning with a 'professional', who engages in a particular activity for money or to earn a living. The term 'amateur' is sometimes used instead of 'volunteer'. For many volunteers the common usage of the term 'amateur' fails to reflect their expertise and dedication (Bell et al., 2008); thus we prefer the term 'volunteer'.

A complex range of motivations - social, cognitive and emotional - encourage volunteers to join EVGs and sustain their commitment. Examples include their passion for their interest, their desire to engage with nature, to interact with people with a common interest, and to increase their personal knowledge and understanding (Kennan et al., 2012). Volunteers collect data in a huge variety of scientific areas, for example, to do with astronomy, water-flows, frog-watching, or butterfly-spotting. The volunteer collector may have no scientific background, but may have a passionate interest in birds, or wildflowers, or conservation of the environment for example (Grove-White et al., 2007). One interest (e.g., in moths) can lead to another (e.g., in host plants). Volunteers participate in data collection as individuals, as members of voluntary groups, and often also as a part of projects organized by professional scientists, commonly known as citizen science. Leaving aside formal citizen science programs, there is potential to harness data collected by 'amateur experts' as a part of their normal EVG activity. In this way, volunteers may eventually contribute to more formal scientific and research purposes,

[1] Atlas of Living Australia http://www.ala.org.au/about-the-atlas/; BowerBird http://www.bowerbird.org.au; NatureShare http://natureshare.org.au/

while also gaining recognition of their needs, contributions and values (Grove-White, et al., 2007).

Data and the importance of data and information sharing

There is a wealth of information on the complexities of research data, defined as data "collected, observed or created for the purposes of analys[why not 'z'?]ing to produce original research results" (Rice, 2009:16). It is important to manage, describe and share research data because they (1) are expensive to collect; (2) may be unique, e.g., represent a snapshot in time or space (Henty, Weaver, Bradbury & Porter, 2008); (3) can be re-used to reproduce and validate original findings, to advance the original research or to open another line of enquiry (Witt, 2009); and (4) can also contribute to answering questions which may require inter-disciplinary problem-solving (Cragin, Palmer, Carlson & Witt, 2010), or which require large-scale data collection beyond the scope of one research team or location. Data sharing is a key element of collaboration (Borgman, 2006) which, along with altruism may motivate researchers to share their data.

Considerations related to data sharing have led to the rise of the open data movement. 'Open data' is an emerging term to define how scientific data may be published and re-used without price or permission barriers. It is related, but not completely analogous, to open access to publications (Murray-Rust, 2008). Despite this movement, there are few obvious explicit and tangible rewards for sharing data and researchers report it as low on the list of their priorities (Henty et al., 2008; Markauskaite, Kennan, Richardson, Aditomo & Hellmers, 2012; Tenopir et al., 2011). Nevertheless, in professional research, funders and journal editors increasingly encourage open data as a grant award pr publication condition (c.f. OECD, 2007).

There are many obstacles to data sharing. Researchers view their data as intellectual capital and report a sense of ownership and responsibility (Pryor, 2009). As Borgman (2006, p. 360) stated; data sharing "is a complex social process involving trust, incentives, disincentives, risks, and intellectual property." A study of controlled-access data repositories finds many reasons for controlling access to data, such as perceptions that data may be misused and that sensitive data needs protecting, along with concerns about intellectual property and attribution (Eschenfelder & Johnson, 2011). These reasons may also resonate with non-sharers of data as a recent study of data sharing practices by scientists found. Many are willing to share their data provided certain conditions are met, including getting credit through citation and notification, provision of secure but flexible infrastructure, and assistance with description and deposit (Tenopir et al., 2011).

There are also considerations from a technical perspective. For data to be shared, at least three sets of criteria need to be met, focused on: 1) accessibility and find-ability; 2) persistence, longevity, and sustainability; and 3) quality. To meet these criteria the data need to be described in machine readable, search-engine findable metadata which include location and access information, and/or stored in repositories which have appropriate technical and organisational infrastructure. Describing and making data accessible are not trivial pursuits (Kowalczyk & Shankar, 2011). Today research data are either digitally generated or digitized from analogue sources such as notes, printed photographs, or specimens (the latter as digital images). The US National Science Board (NSB) defined data as "any information that can be stored in digital form, including text, numbers, images, video or movies, audio, software, algorithms, equations, animations, models, simulations, etc." (Kowalczyk & Shankar, 2011 p. 250). For data to be re-usable it requires description of its context, usually in the form of metadata, which are most often manually created and documented. There are also quality requirements of wholeness, consistency and correctness (Kowalczyk & Shankar, 2011).

Researchers who share data and collaborate with each other tend to have shared disciplinary backgrounds and analytical skills (Borgman, 2006). Likely to be more difficult is sharing data with unknown people via a repository, sharing across disciplines, between professional and volunteer scientists, or between researchers with different motivations and organisational cultures. Inculcating a culture of data sharing requires developing an understanding of those who collect and use data whether professional scientists (Tenopir et al., 2011) or volunteers.

There has been some research on information sharing as a form of information behaviour (Fisher & Julien, 2009), but the field is still emerging (Pilerot, 2012; Wilson, 2010). Nevertheless, there are some parallels with the discussion of data sharing above. Talja and Hansen (2006) discussed information sharing as a form of collaborative information behaviour (Talja & Hansen, 2006). Savolainen (2007) posited that information sharing is a multifaceted phenomenon drawing on social capital manifested through social networks, norms, trust, and mutual understanding. In most cases information sharing is altruistic although, in the case of sensitive information, it tends to be restricted by the assessment of risk by potential sharers. Indeed Wilson (2010) in a literature review found that most work on information sharing incorporates some sense of reciprocity and proposed a number of factors that appear to be common to information sharing: risk, reward or benefit, trust, leadership, and organisational culture. There is thus some overlap with the literature on data sharing.

Information and data sharing is also an area of study in information systems and knowledge management (Widen-Wulff & Ginman, 2004). The emphasis is on how information systems can be designed to enhance collaborative information sharing and to support communities of practice (cf. Lave & Wenger, 1991). Sharing data and information beyond small informal groups through formal systems and repositories is difficult when the information to be shared comes from different contexts,

where the naming systems or classifications used by participants are not shared, and where different information needs to be integrated (Bowker, 2000). Sharing information on this more formal, organized basis, through information systems, repositories and other technological infrastructure, requires researchers and systems developers to make an effort to understand user communities, their cultures and practices (Talja, 2002). It is unclear whether many EVG members understand the possibility of designing a repository or other infrastructure with differing levels of access and use (Denison & Johanson, 2010).

Thus issues associated with data collection and sharing by members of an EVG are complex. For example, data collected for individual or voluntary purposes may or may not be suitable for re-use and sharing for scientific purposes (Palmer, Weber & Cragin, 2011); data need to be described, stored, managed and preserved effectively if they are to be shared; and socio-technical obstacles to data sharing need to be overcome.

This paper will proceed to answer four research questions in relation to the selected EVG, APSV: (1) What are the present data storage and management practices of APSV members? (2) What are the advantages and disadvantages of these practices for the sharing of research data? (3) What are APSV members' present attitudes to sharing research data? (4) What can be done to encourage APSV members to share their data and to empower them with skills and technology to contribute to major data repositories?

RESEARCH PHILOSOPHY AND METHOD

As this research is exploratory, the researchers adopted an interpretivist/constructivist approach to exploring the issues around data and information with EVG members, emphasising their natural setting and seeking to gain deep understanding of the meanings of the actors involved in the social phenomenon under study (Williamson 2013). The researchers' approach in this project was to look for the shared meanings of participants but, at the same time, to take note of the individual meanings emerging in different attitudes and approaches. In other words, the researchers looked for both consensus and dissonance. Within this framework, the method was broadly ethnographic in style, seeking to describe the various behaviours and approaches of the participants involved in the study.

To investigate these issues the researchers required access to an EVG whose members were active collectors of data and who were already actively contributing data beyond internal dissemination channels, or were interested in doing so. Members of a number of groups expressed interest, but the researchers eventually settled on a group, the Australian Plants Society Victoria (ASPV). APSV is a voluntary membership organisation which is a state branch of Australian Native Plants Society Australia (ANPSA). ANPSA has 25 study groups which focus on particular genera (e.g., acacia, correa), and which undertake more scientific activities. Most important in the APSV structure are the district groups to which there is strong local loyalty. APSV (http://www.apsvic.org.au/) has approximately 1,700 members (Hempel, 2007; Walter, 2007). Fifteen members of APSV were interviewed during 2011. Interviewees were purposively selected to include those who were particularly active in data collection in the field, or who had been involved in data/information management and storage in the past. Table 1 presents the demographic profile of the interviewees. Note that participants in each category are broadly representative of the numbers in those categories in the overall membership. Members were mostly from older age groups.

Gender	No.	Age	No.	Length of APSV Membership	No.
male	7	40-49	1	1-5 years	1
female	8	50-59	4	6-10 years	3
		60-69	5	11-15 years	3
		70-79	3	21-25 years	3
		80-89	2	30+ years	5

Table 1: Profile of interviewees

Semi-structured individual interviews, lasting about one hour, were undertaken. The interviews were recorded with the permission of the interviewees and the audio-recordings were transcribed by a trained transcript typist. Although the research method was not 'grounded theory' in itself, the analysis was influenced by the 'constructivist grounded theory' approach of Charmaz (2003). Constructivist grounded theory recognizes that, despite all efforts by researchers to present the views of participants, "the viewer creates the data and ensuing analysis through interaction with the viewed" and therefore the data do not provide a window on an objective reality (273).

The analysis involved initially identifying categories/codes from the interview data and then the merging of these into key themes (Morse, 2008). Quotations from the interviews which illustrate the themes were recorded; resulting in the construction of a 'voice sheet' on each theme, so named because the quotations represent the voices of the participants. As each voice sheet was completed, an overview or summary of the data in that voice sheet was constructed. The themes with illustrative quotes inform the findings, which are presented below.

A follow-up survey of 101 APSV members (about one-tenth of the membership) undertaken early this year (2013), indicated that the qualitative data, collected in 2011, still broadly reflect the situation regarding storage, management, and sharing of data within APSV. These survey data will be included in relevant parts of the 'findings' section.

FINDINGS

To understand the findings in context we provide a further

background to the APSV. The APSV website (APSV, 2013) lists multiple objectives; the one being most relevant to the data collecting activities is to "encourage and facilitate the conservation and study of Australian plants and their habitats". APSV activities for members include monthly meetings where experts (often members of the APSV) share information about specific species and habitats, garden visits, and bush walks. Activities also include the collection of data via surveys of particular habitats or searching for examples of particular species to take observations. Other than meetings, traditional means of dissemination of APSV data and information are via newsletters and magazines and often books, currently in printed form. These forms of dissemination clearly limit access to non-members, i.e., interested individuals, and groups and scientists outside of the APSV.

The reasons given for joining APSV and contributing to activities such as data collection and sharing are many. Expressed motivations include a love of Australian plants, nature and conservation, and the desire to socialize with like-minded people. Many members have a strong interest in increasing their own knowledge and understanding, as well as contributing to society more generally, commitments often shared with professional scientists (Bell et al., 2008).

APSV participants collected a wide range of data, mainly observations. Photographs were the major data type, resulting in the digital camera being the most common data collection tool (used by all participants). Nevertheless, personal notes, such as location, season, time of flowering, habitat, pollinators, growth patterns, and so on, often collected over time, were also important, usually collected with pen and paper. The EVG members made no distinction between the terms 'data' and 'information' in their interviews, using the terms interchangeably.

The findings related to the research questions follow.

Data management approaches: implications for sharing

Personal computers, personal or group websites or databases, CDs, books, magazines, newsletters and notes on paper were the principal ways in which data were stored, managed and, some of the time, shared. These modes are first of all described and then their advantages and disadvantages, especially for sharing data with others, are discussed.

Computers, websites, databases and CDs

All participants stored photos on their computers, except for the oldest participant (in her 80s) who did not have her own computer, although she contributed her photos to her district group's website, as did others. Five participants had their own personal websites, mostly in addition to contributing to their district websites. Personal computers were also used to type up notes, collected in the fields. For example one person used a diary form for her notes.

> "Sometimes it's handwritten out in the fields in case I forget and then I just come home and write it up in a diary form about what was found and what the conditions were." (Interviewee 12)

Another use of computers was for the entry of categorized data, often on spreadsheets (N=11).

> "I've got a spreadsheet which has a list of the different plants … whether it's just come to flower or it's in full flower or it's just nearly finishing." (Interviewee 2)

These categorized data were usually personal to particular participants and were not as a rule shared with others.

> "I was a bit keen on databases so I set this one up ... [with] all the species for Victoria ... [a] list of 4,000 species." (Interviewee 5)

Four participants specifically mentioned having their own databases and one had had a database in the past.

> "It has evolved and it's basically a series of folders and sub folders. I'm steering more and more towards the scientific structure ... it was initially just families. ... because part of the reason for doing this is actually to help me learn." (Interviewee 15)

Five participants mentioned CDs to which they had contributed their data.

> "One thing that we did … was produce a CD of correas with as much as we knew about correas, the different species and the habitats and where they were found. … It's got a photo gallery and … all the correas study newsletters on it." (Interviewee 14)

Books, magazines, newsletters and hand-written notes

A number of books had been written over the years, by individuals and district groups.

> "More than 30 years ago they [Maroondah group] wrote a little book called '500 Australian Plants' … [It] was a recording of individual information [about plants grown in the eastern suburbs]. And that was a very, very popular little book." (Interviewee 7)

This book was being revised and up-dated at the time of the interview.

Magazines and newsletters remain a very important form of communication within APSV and its district and study groups. The Society has a coloured quarterly magazine (printed) in which a wealth of information has appeared over a long period of time. This magazine was mentioned by most participants. An interviewee from the acacia study group reported the following regarding the newsletter of that group.

> "The acacia study group newsletters which go back to the 1960s and which are currently up to … 120 now, and in those newsletters there's a lot of information that has been recorded over the years." (Interviewee 2)

Hand-written notes were very important in the collection of field data by almost all interviewees. They were stored in folders, notebooks, index files, or through annotations (N=10), while others (N=6) saw their memories as storage spaces.

> "I've got two notebooks that I've been entering into over a period of 10 years. ... I don't [enter it onto computer] not for data purposes because you can't then sort of add in on the sidelines that something else looked gorgeous." (Interviewee 1)

Data management: advantages and disadvantages

There was a range of advantages and disadvantages perceived by participants regarding their data management (including storage and preservation) practices. Participants, who had been collecting data in the field for many years, saw their use of computers for storage as a great advantage.

> "Obviously the computer's … made a huge difference in the speed of being able to do things and the amount that you can store really." (Interviewee 14)

Having photos and other data on computer is a first step towards being able to share with others. Indeed participants with their own websites found that they were sharing their data. Interviewee 11 was very active, had more than one website and contributed to other websites as well, as illustrated by the following two quotations.

> "[My] first website … it's a photo hosting site in America … It's a really good data base management system for photos. So just in the description of the plant or the key words, I can put in its color, its shape, its size, its form." (Interviewee 11)

> "I often get requests through my website for people to use photos, in publications …, brochures and things like that." (Interviewee 11)

The district group websites also enabled the sharing of data and information, at least within APSV, but they were not always easy to maintain.

> "We're just getting [our district's website] back on track again. And we had a web master who we gave information to, but we've now set up so that members of the committee will be able to add information themselves." (Interviewee 12)

Another advantage of websites was that there was extra back-up. As several participants admitted, their personal back-up methods for photos on their own computers (e.g., memory sticks and external discs) were often not entirely satisfactory.

> "I don't think we've got a very good system for backup. ... I copy onto just a memory stick. … Yes we need better backup systems." (Interviewee 8)

There was also recognition of the access limitations and looming obsolescence of technology such as CDs; and awareness that increasing volumes of data/information are difficult to manage on an individual or small group level.

A significant problem for modern forms of data management was the lack of IT skills amongst the older membership.

> "A lot of the ones who are the older ones, who have got fabulous information because they've been growing and out there in the bush for so long, and theirs is all stuck up here [in the head] too; fabulous information but they are just not computer literate." (Interviewee 7)

Related was the realization that some members with a wealth of knowledge were dying and taking their knowledge with them.

> "We realized that [we had lost amazing information] at Maroondah group a few years ago when we had a succession of our older members pass away." (Interviewee 8)

The age profile of APSV membership was likely to be influential in the fact that most interviewees were still strongly in favor of print forms of communication (books, magazines and newsletters). Nevertheless, the problem of information becoming dated in books was well recognized, especially given the ongoing changes of botanical names.

> "Just the logistics of producing a book sort of didn't appeal. They're out of date almost instantly, it's changing all the time ... all the family names, everything." (Interviewee 5)

While interviewees enjoyed receiving their magazines and newsletters, the longer-term problems of retrieving information from them was well recognized.

> "The more recent newsletters are on the website … but if they were trying to get information on a particular acacia they wouldn't know where to go, whereas there may well be an article in one of our newsletters from 20 years ago." (Interviewee 2)

Of the 15 interviewees, only the developer of NatureShare had contributed to either it or to the national repository, the Atlas of Living Australia. This was understandable given that both were very new at the time. Nevertheless, the recent survey revealed that few APSV members (N=7), almost two years later, were contributing to either of these repositories – the two options available to them at the time.

Attitudes to data/information sharing

Most data and information were shared by interviewees in meetings with other APSV members; face-to-face networks were tried and trusted. Short articles in magazines and newsletters were widely read. There were some limitations to sharing.

> "[I'm] happy to share data and information ... with other interested people ... but there's a certain protectiveness about it. ... You don't want to generate a lot of interest

from people who are not used to conserving the bushland." (Interviewee 1)

A previous opportunity to share data through the State Government repository (Flora Information System - FIS) had been curtailed when the repository had been sold to private hands. Interviewees had become disillusioned and had refused to contribute further when they found that they had to pay to access datasets to which they originally contributed their own data voluntarily. Moreover, there was resentment that data contributed to FIS seemed to disappear, unacknowledged and unattributed.

> "Well at one stage … I saw myself sending all my observations [into FIS]. I wanted to get it registered somewhere so it could be of use. And it was just so difficult. ... You're not encouraged to do it because you get no feedback and ... you're even wondering ... what's happened?" (Interviewee 5)

The advent of NatureShare (which is freely available to contributors and other users) had begun to change this situation.

> "The NatureShare solution ... allows you to upload all the information that you want." (Interviewee 3)

Nevertheless, relinquishing control over their data was not something all interviewees wanted.

> "I'd be wrong, I'd be lying if I said I didn't [have] a certain amount of pride in [my own website] and it's nice to be recognized because this is your work that's out there." (Interviewee 15)

Even when attempts were made to disseminate electronic data, problems still sometimes arose, e.g., with websites developed by individuals or groups.

> "I do have a number of websites that I produce that nobody can access for some strange reason. We're still trying to work out why Google can't find them." (Interviewee 9)

These problems are likely to be related to inexperience with networks and computers, as alluded to above. At least this interviewee was attempting to use new technology, unlike some older APSV members who were quite often described by interviewees as uninterested in technology:

> "The age group of the members of APSV is generally from 55 or 60 onwards ... so the people there are not really interested in technology and it's very hard to communicate with them via email." (Interviewee 9)

Other barriers arose from the loyalties which members had to their district groups and their preference to contribute their data to those groups using their traditional methods of meeting presentations and newsletter articles, without having to put data into a larger repository with complex metadata requirements. There were other expressions of not wanting to change the way things were done now and not having time for efficient data management and sharing.

Quality control of data was another related issue, one which was mentioned by a number of interviewees as a reason they may not want to add data to a shared repository or website.

> "It depends who has provided it [the data/ information]; some are just people out in the bush, you know ..., and they get it wrong, and they don't know, and so, errors creep in." (Interviewee 7)

Quality control was also influenced by inconsistent naming systems:

> "It's very difficult from our point of view [the constant name changes] because every time there's a name change we've got to relabel everything on our computer … And then you've got to put what it used to be because people know it under the previous name." (Interviewee 14)

DISCUSSION

The findings indicate a range of impediments to data and information sharing, often stemming from data management practices. Nevertheless, there was a willingness to share on the part of most interviewees. While the recent survey indicated little change in data management practices, there was a heartening finding in relation to the question: If available, would you be interested learning more about the use of a repository for your data collections? Of the 27 respondents who answered this question, 25 answered in the affirmative.

The following provides answers to the four research questions and discusses the literature as appropriate.

(1) What are the present data management practices of APSV members?

Sample members used a wide range of management and storage practices, some of them as basic as hand-written text on paper; others as sophisticated as good quality spreadsheets, databases and websites. Computers were universally used for the storage of photos – a good first step. Members are still very much devoted to their printed magazines and newsletters but recognize the difficulties for sharing and long-term preservation of the information in these. At least some of the latter are now being digitized. Most interviewees could articulate the problems in their data management practices, a first step to improvement.

(2) What are the advantages and disadvantages of these practices for the sharing of research data?

Personal data and notes, stored in private homes, are problematic for any kind of sharing or preservation, as acknowledged by several interviewees who had witnessed the loss of valuable data and information as members died.

The use of books, magazines and newsletters are still important to members and certainly provide a central role in information communication amongst the membership of APSV. Their role in wider dissemination is, however,

limited as these forms of sharing provide summaries of analysed data and are generally selective representations of a small amount of the raw data originally collected by the researcher (Latour, 1987), not data that are able to be incorporated into the larger scientific project.

While the common use of computers for the storage of photographic data - and for the development of, and access to, websites and other databases - is an important start for data and information sharing, there are disadvantages in present practices. There are threats to the take-up of future sharing opportunities and to long-term preservation of data when individuals' computers are often not backed up beyond temporary media such as CDs and memory sticks, and when essential metadata is not recorded. Further, although there are commonalities across district groups, members are making their own decisions about what data they collect and store. Interviewee 3 summed up the chaos succinctly:

> "There has not been a good way to store information. Everyone collects their own information and stores it however they can. … We all keep our own information, our own photos, and all that sort of stuff, and there's nowhere for it to go, basically."

The last two years have seen a change in that NatureShare has begun to gain recognition and there is a new repository, BowerBird, about to come online, for a range of environmental data. NatureShare data are already being absorbed into the Atlas of Living Australia and the same will occur with BowerBird.

What is needed now is to change practices so that potential contributions to these repositories are collected and managed to enable the sharing of data more broadly among members and with the scientific community. Wider sharing can be viewed as extension of their local networking. Grove-White, et al., (2007) pointed out that databases and repositories usually require that data be validated and standardized. Changed sharing practice through repositories will also lead to better preservation as repositories are there for the long-term, and preservation issues will be addressed for the entire repository rather than for small, distributed collections.

(3) What are APSV members' present attitudes to sharing research data?

APSV members were aware that some of their data were valuable and were generally willing to share. Interviewee 3 thought that up to 90% of members want to share their data and information "in a useful way, and can't." Yet most interviewees were not attempting to contribute to major data repositories such as NatureShare and the Atlas of Living Australia. They explained their reluctance to share through formal processes, identifying reasons such as difficulties with technology and lack of trust of existing systems. Furthermore, they were aware that sensitive data (such as the location of rare and endangered species) need protecting, and were not sure how the major repositories would manage this. The problem of possible misuse of data has been recognized in the literature (e.g., by Eschenfelder & Johnson, 2011).

The lack of acknowledgement and attribution of their data, which past contributors believed they had experienced, was also a deterrent to data sharing through large repositories. Issues connected to intellectual capital, ownership and intellectual property, and attribution are also discussed in the literature, in a broader context, (Borgman, 2006; Eschenfelder and Johnson, 2011; Pryor, 2009; Tenopir et al., 2011). A major consideration for sharing volunteer data, e.g., of individual observations or small datasets incorporated into larger ones, is that the individual sharer loses control of how the data are used, and who may use them once they are deposited, even if the individual datum is attributed. If shared data are used as a part of a larger data set by professional researchers or for policy purposes, the original contributor may not even be a part of the community that uses or re-uses the data. Thus reciprocity, a factor considered important in the information sharing literature (Savolainen, 2007; Wilson, 2010), is less of a factor in this kind of data sharing. The question regarding providing motivations for data sharing is therefore crucial. How can volunteers be encouraged to be more altruistic and to appreciate the bigger science perspective?

In addition, the time needed for preparing data for deposit and entering them was an issue for APSV members, along with the perception that databases and repositories may not have been developed with the needs and values of EVG contributors in mind. The work of describing data (Kowalczyk & Shankar, 2011), and the lack of incentives for sharing beyond their own community, were concerns in common with professional researchers (Henty et al., 2008; Markauskaite et al., (2012). As Talja (2002) pointed out, sharing information on this more formal, organized basis, through information systems, repositories and other technological infrastructure requires researchers and systems developers to understand user communities, their cultures and practices. Incentives are few. The advent of the new data repositories, such as NatureShare and BowerBird which are responsive to the needs of users (by providing help with identification and social interaction), may change the situation gradually and data sharing may become more common.

The new repositories (if well designed with the volunteer in mind) should overcome the naming issues (Bowker, 2000) by automatically updating and referencing taxonomy changes, and also assist with other quality issues. Apart from the quality control exercised by database moderators, contributors are prompted by forms to provide standardized metadata and where required receive help with identification. Concerns about data quality and description are common, shared with other volunteer scientists (Grove-White et al., 2007) and professional scientists (Borgman, 2007).

(4) What can be done to encourage APSV members to share their data and to empower them with skills and technology to contribute to major data repositories?

Just as the specific needs of professional researchers and their data-sharing needs and practices require further understanding to inform repository developers and encourage data sharing (Talja, 2002; Tenopir et al., 2011), so too work is required alongside volunteers.

Following on from the research reported here an action research project with APSV members will be undertaken. The project aims to develop a deeper understanding of participants' data collecting, sharing and storing practices. Several types of in-field and at home data collection, storage and sharing equipment (smart phones, tablets, laptops and home computers) will be trialled and evaluated. Training packages will be developed and evaluated. They will be focussed on data collection, including the metadata required for contribution to repositories as well the processes for uploading data into NatureShare and Bowerbird. Thus volunteers will learn to use new technologies and also to share their data by uploading them into repositories of global value. Central to this project is its train-the-trainer element. The original participants who will be from different APSV districts will, in time, become the trainers of others. We hope that a 'snowball' process will ensue.

CONCLUSION

APSV members collected a wealth of data 'from the wild' and often spent considerable time in identifying and recording what they had found (Kennan, Williamson & Johanson, 2012). At this point in time, most APSV local groups are parochial in managing and sharing their data, but their underlying conservation philosophy, and willingness to share their data and information, is well demonstrated in Society publications and through their expressed motivations in this study. At present many member data and information practices are basic and there is a risk of long-term destruction of paper-based data and one-off personal and district spreadsheets, databases and websites. At the same time, there is a high level of understanding of the potential benefits to conservation and science of sharing their data beyond their fellow Society members.

Open data shared through biodiversity repositories such as the Atlas of Living Australia, BowerBird and NatureShare are unknown territories for many volunteers. To encourage data sharing, repositories need to more explicitly meet the need of potential volunteer contributors by promoting, along with the potential for sharing (which is well understood), the curation and preservation opportunities such repositories offer, for individual observations and small collections, as they become part of larger, more sustainable collections. An understanding of volunteer needs and practices, and an explicit addressing of their concerns is also likely to see greater future data sharing.

ACKNOWLEDGMENTS

The authors acknowledge, with thanks, the support of the APSV, especially the assistance of Cathy Powers, President, and Russell Best, Research Officer, and the interviewees who gave us their time and their views. The Small Grant funding received by Kennan and Williamson from the Faculty of Education, Charles Sturt University, is also acknowledged with thanks.

REFERENCES

Australian Plants Society Victoria. (2009, Updated 4th July 2013) Website http://www.apsvic.org.au/index.html

Atlas of Living Australia (n.d.a) *Atlas background* http://www.ala.org.au/about-the-atlas/atlas-background/

Atlas of Living Australia (n.d.b) *FieldData and data* http://www.ala.org.au/get-involved/citizen-science/fielddata-software/fielddata-and-data/

Bell, S., Marzano, M., Cent, J., Kobierska, H., Podjed, D., Vandzinskaite, D., et al. (2008). What counts? Volunteers and their organisations in the recording and monitoring of biodiversity. *Biodiversity and Conservation, 17*(14), 3443-3454.

Borgman, C. L. (2006). What can Studies of e-Learning Teach us about Collaboration in e-Research? Some Findings from Digital Library Studies. *Computer Supported Cooperative Work: The Journal of Collaborative Computing, 15*(4), 359-383.

Borgman, C. L. (2007). *Scholarship in the digital age: Information, infrastructure, and the Internet*: The MIT Press.

Bowker, G. C. (2000). Biodiversity datadiversity. *Social Studies of Science, 30*(5), 643-683.

Callon, M., Lascoumes, P., & Barthe, Y. (2009). *Acting in an uncertain world:An essay on technical democracy*. Cambridge, Mass: MIT Press.

Charmaz, K. (2003). Grounded theory: Objectivist and constructivist methods. In N. K. Denzin & Y. Lincoln (Eds.), *Strategies of qualitative inquiry*, (2nd ed., pp. 249-291). Thousand Oaks, CA: Sage.

Cragin, M. H., Palmer, C. L., Carlson, J. R., & Witt, M. (2010). Data sharing, small science and institutional repositories. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 368*(1926), 4023-4038.

Denison, T., Johanson, G. (2010). E-Research infrastructure and community research. *Community informatics*. 6(3). At: http://ci-journal.net/index.php/ciej/article/view/736/621.

Eschenfelder, K. & Johnson, A. (2011) Limits of sharing: Controlled data collections *ASIST Annual Meeting 2011*, October 9-13, New Orleans.

Fisher, K. E., & Julien, H. (2009). Information behavior. *Annual Review of Information Science and Technology, 43*(1), 1-73

Grove-White, R., Waterton, C., Ellis, R., Vogel, J., Stevens, G., & Peacock, B. (2007). *Amateurs as experts: Harnessing new networks for biodiversity*. Lancaster: Lancaster University, Institute for Environment, Philosophy, and Public Policy. http://csec.lancs.ac.uk/docs/Amateurs%20as%20Experts%20Final%20Report.pdf .

Heidorn, P. B. (2009). Shedding Light on the Dark Data in the Long Tail of Science. *Library Trends*, *57*(2), 280-299.

Hempel, P. (2007). Survey of APSV membership.

Henty, M., Weaver, B., Bradbury, S., & Porter, S. (2008). *Investigating data management practices in Australian universities*. Canberra: APSR: Australian Partnership for Sustainable Repositories. Retrieved http://www.apsr.edu.au/investigating_data_management.

Hey, T., & Trefethen, A. (2003). The data deluge: an e-science perspective *Grid computing-making the global infrastructure a reality* (pp. 809-824). New York: Wiley.

Kennan, M. A., Williamson, K. & Johanson, G.. (2012). "Wild Data: Collaborative E-Research and University Libraries." *Australian Academic & Research Libraries* 43(1): 56-79.

Kowalczyk, S., & Shankar, K. (2011). Data Sharing in the Sciences. *Annual Review of Information Science and Technology*, *45*, 247-294Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*: Cambridge, Cambridge University Press.

Lasalle, J. (2013). Personal communication with John Lasalle, Director of the Atlas of Living Australia, 14 February, 2013.

Latour, B. (1987). *Science in Action*. Cambridge, MA: Harvard University Press

Lave, J . & Wenger, E. (1991). *Situated learning. Legitimate peripheral participation.* Cambridge, England: Cambridge University Press

Markauskaite, L., Kennan, M.A., Richardson, J., Aditomo, A., & Hellmers, L. (2012) "Investigating eResearch: collaboration practices and future challenges." In *Collaborative and Distributed E-Research: Innovations in Technologies, Strategies and Applications.* Edited by Juan, A., Daradoumis, T., Roca, M., Grasman, S., & Faulin, J. Information Science Reference, IGI Books, Hershey PA. pp 1-33.

Murray-Rust, P. (2008). Open data in science. *Serials Review*, *34*(1): 52-64.

OECD (2007). *OECD Principles and Guidelines for Access to Research Data from Public Funding*. Paris, OECD.

Palmer, C., Weber, N. & Cragin, M. (2011) The Analytic potential of scientific data: Understanding re-use value. *ASIST Annual Meeting 2011*, October 9-13, New Orleans.

Pilerot, O. (2012). LIS research on information sharing activities–people, places, or information. *Journal of Documentation*, *68*(4), 559-581.

Pilerot, O., & Limberg, L. (2011). Information sharing as a means to reach collective understanding: A study of design scholars' information practices. *Journal of Documentation*, *67*(2), 312-333.

Pryor, G. (2009). Multi-scale data sharing in the life sciences: some lessons for policy makers. *International Journal of Digital Curation*, *4*(3), 72-82.

Reichman, O.J., Jones, M.B., Schildhauer, M.P. (2011). Challenges and Opportunities of Open Data in Ecology. *Science*. 331(6018), 703-705.

Rice, R. (2009). *DISC-UK DataShare Project: Final Report*. Bristol: JISC.'Retrieved' 10 December 2011, from http://ie-repository.jisc.ac.uk/336/1/DataSharefinalreport.pdf.

Savolainen, R. (2007). Motives for giving information in non-work contexts and the expectations of reciprocity. The case of environmental activists. *Proceedings of the American Society for Information Science and Technology*, *44*(1), 1-13.

Talja, S. (2002). Information sharing in academic communities: types and levels of collaboration in information seeking and use. *New Review of Information Behavior Research*, *3*(1), 143-159.

Talja, S., & Hansen, P. (2006). Information sharing. In A. Spink & C. Cole (Eds.), *New directions in human information behavior* (pp. 113-134). Dordrecht: Springer.

Tenopir C, Allard S, Douglass K, Aydinoglu AU, Wu L, et al. (2011) Data Sharing by Scientists: Practices and Perceptions. *PLoS ONE* 6(6): e21101.

Van House, N. A. (2003). Digital libraries and collaborative knowledge construction. In A. P. Bishop, B. P. Buttenfield & N. A. Van House (Eds.), *Digital library use: Social practice in design and evaluation* (pp. 271-295): MIT Press.

Walter, J. (2007). The Story of Arthur Swaby and the Society for Growing Australian Plants: Australian Plants Society Inc.

Widen-Wulff, G., & Ginman, M. (2004). Explaining knowledge sharing in organizations through the dimensions of social capital. *Journal of Information Science*, *30*(5), 448-458.

Williamson, K. 2013. Research concepts. In K. Williamson & G. Johanson (Eds.) *Research methods: Information, systems and contexts* (pp. 3-23). Prahran, VIC: Tilde University Press.

Wilson, T. (2010). Information sharing: an exploration of the literature and some propositions. *Information research*, *15*(4), 12.

Witt, M. (2009). Institutional repositories and research data curation in a distributed environment. *Library Trends*, *57*(2), 191-201.