This is a repository copy of *CURIOS: Connecting Community Heritage through Linked Data*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/92275/

Version: Accepted Version

Article:
Webster, G., Nguyen, H., Beel, D.E. et al. (3 more authors) (2015) CURIOS: Connecting Community Heritage through Linked Data. Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing. 639 - 648. ISSN 978-1-4503-2922-4

https://doi.org/10.1145/2675133.2675247

Reuse
Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher's website.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

CURIOS: Connecting Community Heritage through Linked Data

Gemma Webster Hai H. Nguyen David E. Beel Chris Mellish Claire D. Wallace Jeff Pan
dot.rural RCUK Digital Economy Research Hub
University of Aberdeen
Aberdeen, UK
[gwebster, hai.nguyen, d.e.beel, c.mellish, c.wallace, jeff.z.pan]@abdn.ac.uk

ABSTRACT
The CURIOS project explores how digital archives for rural community heritage groups can be made more sustainable so that volunteer members can maintain a lasting digital presence. It is developing software tools to help remote rural communities to collaboratively maintain and present information about their cultural heritage. The objective is to investigate the use of semantic web/linked data technology to build a general, flexible and "future proof" software platform that could help such projects to develop digital archives and to be sustainable over time. As an interdisciplinary project we aim to synthesise a narrative that draws from both social science and computer science perspectives by critically reflecting upon the novel approach taken and the on-going results that are being produced.

Author Keywords
Cultural Heritage; Digital Archives; Community heritage; Open Linked Data.

ACM Classification Keywords
E.2 [Data]: Data Storage Representations – Linked representations, J.4 [Computer Applications]: Social and Behavioral Science – Sociology, K.4.0 [Computers and Society]: General.

General Terms
Human Factors

MOVING TOWARDS COMMUNITY DIGITAL HERITAGE

Rural areas are characterised by a strong identity of people with place. These identities draw on a repertoire of cultural norms, knowledge, histories, customs and practices which, taken together, construct unique place identities. This cultural distinctiveness is dynamic given traditional cultural practices are reproduced and others introduced as cultural systems evolve and adapt. Forms of cultural expression, such as story-telling, music and song, poetry and literature, dance and drama, together with material objects, artifacts, sites and cultural spaces, are resources for interacting with the past and for experiencing the present. In the collection and transmission of these collections there has been a growing sense that the traditional methods for doing this are failing [13].

In order to address this problem, digital solutions have been sought but this has been a problematic process due to a number of variances. These include the constant changing of file types, software and codes of best practice, as well problems to do with cost and the sheer amounts of 'analogue' data to convert. National institutions have been leading the way in this process, but with the production of such local cultural repertoires, which as Flynn [6] suggests 'are the grassroots activities' where 'control and ownership of the project is essential', there has been a failure to consider the needs of community heritage groups in these processes [7]. Additionally, such groups do not want to be subsumed into national archives, which they do not control, are not sensitive to their needs and are juxtaposed ideologically to the production of their own 'place history'. Following Creswell's [5] claim that such archives represent 'spaces of marginalized memory' CURIOS is therefore seeking a solution, using open linked data, in which a system can be developed that is attuned to the specificity of a local heritage but can also take advantage of already collected materials from elsewhere.

LINKED OPEN DATA: TOWARDS AN ARCHIVE FOR THE FUTURE?

Digital archives can have several limitations, particularly when collected and maintained by different community groups. Firstly, due to community heritage often being gathered by volunteers there is a wide range of computer literacy across individuals as well as organisations. Secondly the data formats between collections and tools for digitising heritage data are not consistent. For example, some groups may choose to use common multipurpose tools such as spreadsheets to maintain their archives. Others may use pre-existing genealogy software or a traditional relational database. Larger organisations such as national institutions may choose other formats to meet their specific requirements such as Key-Value data-stores for performance or Resource Description Framework (RDF) triplestores for integration and reusability.

Due to the wide ranging and different types of data formats, it is not trivial for digitalised cultural heritage to be reused and integrated with each other. Most digital archives, therefore, can only be exploited separately, meaning connecting local cultural heritage with national archives

cannot be done easily. The integration of knowledge from different digital archives where possible, has been done manually by humans. To automate this process, i.e., to integrate/contextualise contents from different archives, it is important to keep data in an open and reusable format [4].

A solution to an open and reusable data format is the Linked Open Data format. The emergence of the Semantic Web [14] has led to several standard formats for representing and interchanging data [2,8]. In 2009, Tim Berners-Lee proposed linked data as a set of best practices using standard web technologies to publish and link data on the web [11]. These practices involved publishing data using Uniform Resource Identifiers (URIs) such as web addresses and relating those using defined relationships (links). The description of data is then stored as RDF triples in the form:

<subject link object>

where subject is a URI, link specifies the type of relationship between the subject and object, and object can be either a literal (i.e., raw data such as numbers, text) or a URI. Using this format, information regarding a subject (e.g., a URI) will be stored as a set of triples in an RDF-supported database, usually referred to as a triplestore.

For example "Angus Macloed lives at 8 Calbost" would be represented in RDF triples as:

<hc:23160 dc:title "Angus Macleod">

<hc:23160 hc:liveat hc:1914>

<hc:1914 dc:title "8 Calbost">

The above example describes two subjects in the Hebridean Connections archive: Angus Macleod (hc:23160) and 8 Calbost (hc:1914) by their titles and the relationship between them, e.g., hc:liveat (hc: and dc: are prefixes to identify namespaces for the vocabulary being used). Data holding in the RDF triples can be retrieved and updated by users via SPARQL, a query language for retrieving and maintaining data stored in RDF format.

So far, there has been a vast increase in the number of linked data datasets becoming available on the web [3]. By making use of linked data, cultural repositories would have the potential for reuse and integration with further related data sources. Using Linked Open Data formats, multiple datasets can be integrated and jointly exploited by SPARQL. There have been several attempts to bring cultural heritage data into the Linked Open Data formats such as in the CultureSampo project [10] and the OpenART project [9].

The CultureSampo project aims to create a cultural heritage archive for the whole nation by providing an infrastructures and a set of tools to publish and annotate contents collectively. The case study used in this project is the Finnish cultural heritage archives. On a smaller scale, the OpenART project brings an important arts research dataset, "The London Art World 1660-1735", to the Linked Open Data format so that contents about the art world during that period can be contextualised and linked to the Tate collection and referred to the relevant contemporary art works.

The CURIOS project also aims to overcome the data heterogeneity problem by keeping cultural heritage data in the Linked Open Data format. Yet unlike above projects, which focus on a specific dataset/case-study, CURIOS aims to not only exploit existing linked datasets but also allow the creation of new ones. To do so, CURIOS provides a set of software and tools to assist users with limited Semantic Technology knowledge in producing and consuming their cultural heritage in the form of linked data.

CASE STUDY – HEBRIDEAN CONNECTIONS

In the past 40 years around 20 'Comainn Eachdraidh' have been established in the Outer Hebrides, a group of islands off the West coast of mainland Scotland. Comainn Eachdraidh is a Gaelic phrase meaning 'Historical Societies' and they are community run groups that began in the 1970's with a very specific political and cultural purpose – to preserve the culture, history and language of the primarily Gaelic regions of Scotland. Comainn Eachdraidh across the Outer Hebrides are often very geographically based, both physically within their community space and with a specific focus on their local community history, meaning each can have very different aims and objectives. Such community heritage practices have been described as a 'messy' endeavour with a wide variety of different formal and informal practices [12,17]. The archives embrace different registers of social memory from tangible to intangible heritage, which have been collected and ordered in a variety of different ways: from the highly 'professional' to the more bespoke and sporadic. As the Comainn Eachdraidh groups are voluntary community archives, they are rooted in local historical values, hence there is often little consistency between groups regarding cataloguing, archiving and content management.

Hebridean Connections (HC), which is a community-managed, online historical resource, was formed due to the driving force of a single member of a Comainn Eachdraidh with a background in IT who saw the benefit of digitising and connecting the different historical catalogues [15]. The idea was proposed to the different Comainn Eachdraidh, with four groups (Bernara, Uig, Pairc and Kinloch) actively involved in securing a Heritage Lottery Fund (HLF) bid that funded the creation of the HC website (www.hebrideanconnections.com/). This initial phase fell into abeyance due to a number of reasons (to be discussed in the following section) but has been resurrected through a second funding phase (which is collaborating with the CURIOS Project) and currently has two paid director/administrative staff funded by the Scottish Government's People's Community Fund (PCF) and has

expanded from the original four Comainn Eachdraidh to ten historical societies.

HC is one example of a community-built digital cultural heritage repository where their long-term future is unclear. The project website, which was launched in 2006, holds some 100,000 records relating to the genealogy, history, archaeology, and cultural traditions of the Outer Hebrides. The collections are currently expanding as the additional six Comainn Eachdraidh are contributing their collections, and the vision is to expand the collections to incorporate all materials held by the remaining Comainn Eachdraidh. The project is enabling local heritage archives to become accessible to international diasporic communities, especially those with genealogical connections back to the islands. The collections fall outside the national and even local (government) institutional frameworks, hence local people are the 'gatekeepers' of their own heritage and are selecting what to commemorate based on their own customs of remembering. Additionally, the website encourages contributions from its users and therefore has the potential to foster reciprocal knowledge exchange across geographical boundaries. This kind of digital archive can have significant social impacts for the communities involved both at a local and global level. The initiative is heavily dependent upon the volunteer efforts of the Comainn Eachdraidh members and their on-going desire to convert their analogue collections into a digital form.

Sustainability and the Problem Case

The long-term future of Hebridean Connections is unclear, with many issues surrounding the current system arising since the initial grant (as mentioned). In particular, there are concerns around how small communities like this maintain such collections when funding periods have finished. This is something the CURIOS project has been trying to address, with its approach picking up from the failings of the HLF phase of HC. The following section will now detail the problem case for CURIOS before moving to detail the solutions proposed.

The initial website was developed by a private development company using proprietary software funded by the initial grant. The proprietary software restricted the use of the system via licenses and only four licenses were able to be purchased. These licenses were applied to four computers each physically located in Comainn Eachdraidh centres. This restriction on data entry made inputting records difficult and four staff were employed (one at each society) on short-term contracts to actually carry out the data entry to the system. The use of paid staff to digitise the physical records added a barrier to the Comainn Eachdraidh in that it stopped volunteers from being able to contribute to the project and restricted the progress of data entry to one person at a time.

Additionally, two new Comainn Eachdraidh groups decided to join the project but there was no additional funding to purchase licences, meaning a member from each of these groups had to physically travel to another Comainn Eachdraidh to use their computer to enter records. Records were also automatically assigned ownership based on licenses, meaning the two new groups records were automatically assigned to the Comainn Eachdraidh computer they were using, but there was no way for this to be changed without purchasing more licenses.

As the project developed, this situation raised the problem that any changes to the system required more financial investment in the software. For community led cultural heritage groups, funds are always scarce yet as the digital collection grew, the Comainn Eachdraidh became more aware of what was possible through digitisation and wanted to expand. The feasibility of this growth taking place within the constraints of the proprietary was not possible; hence when the HLF funding ended so did this phase of the project. The process of digitisation thus created a number of issues for Hebridean Connections:

- How to expand the project remit without additional funding for developers?
- Scalability issues, such as how can more Comainn Eachdraidh collections be integrated in a closed system?
- How can more volunteers be given access to the system?
- How can appropriate training be given to volunteers?
- How can we maintain our cultural archives independently?

The End of the Proprietary Software

After the HLF ran out there was no funding available to employ paid staff to continue the data entry. This caused a significant pause in the digitisation of the Comainn Eachdraidh collections for a number of reasons: the volunteers had not been responsible for the data entry previously; the software being used was no longer suitable as the project had continued to grow with multiple groups wanting to join but with no way for them to enter data; and finally, the new groups had different types of data within their collections that could not be added to the closed database design.

It was at this point that the CURIOS project began to investigate a suitable solution to the problem of how digital archives for rural community heritage groups can be made more sustainable, so that volunteer members can maintain a lasting digital presence. It was during the initial stages of the CURIOS project that more issues with the existing software became apparent. As all the data had been entered into an external database, in order to access and recover the original digitised data Hebridean Connections had to pay the external software company for access to their own data.

The format of the original data within the database was very restrictive and did not lend itself to being expanded. Some

of the data within the database was out of date as it had been given to Hebridean Connections from the Royal Commission on Ancient and Historical Monuments of Scotland (RCAHMS) to enhance the information about important local features of the Outer Hebrides. This data from RCAHMS had been directly copied from their data source but had not been updated since it was copied. Consequently, the database would need to be completely re-designed.

Many of the Comainn Eachdraidh groups that had newly joined the HC project already had some digitized data, but it was within common genealogy software packages such as Brothers Keeper, which use a GEDCOM file format. A GEDCOM file contains genealogical information about individuals and is usually plain text and metadata showing the links between records. GEDCOM has many different proprietary extensions with most genealogy packages allowing the export/import of GEDCOM format but not necessarily the new extensions. This file format in itself caused problems as there was no apparent way to export a GEDCOM file and insert this into the pre-existing database. This meant many Comainn Eachdraidh groups thought they would have to start the digitisation process again.

It was also found that Hebridean Connections now had the only digital versions of some photographs and files. As the Comainn Eachdraidh groups had evolved over time original digital files (not the hard copies) had gone astray with members changing or the physical archives moving locations.

The issues outlined above all helped lead CURIOS to investigate the use of semantic web/linked data technology to build a general, flexible and "future proof" software platform that could help such projects as Hebridean Connections to develop digital archives and to be sustainable over time.

CURIOS

The CURIOS project's aim is to produce a sustainable system that allows a community of users to manage a digital archive of cultural heritage data, or 'cultural repository', releasing them from any specific proprietary software platform. As highlighted in the CrowdMemo project, using 'off the shelf' technologies in new ways can allow for long-term sustainability[1] to achieve this goal, CURIOS has made use of existing open source content management system (CMS) software and Semantic Web standards.

The CURIOS Content Management System

In recent years content management systems have gained popularity on the web by allowing users to build and publish web pages without requiring in-depth knowledge of the underlying web technologies. The CURIOS project has extended the web CMS approach to allow users to manage repositories of linked data. This linked data CMS approach makes use of existing CMS software to retain the usability and scalability of existing tools that are familiar to users, whilst allowing the users to exploit the benefits of linked data.

The Linked Data CMS approach has been implemented as a module for the popular open source web content management system Drupal (http://drupal.org), which we call the CURIOS system. There are three main reasons why Drupal was chosen over other open source CMS systems:

1. Drupal has a strong developer community and a large collection of high-quality and free contributed modules. This makes the system more sustainable and extensible. For example, the users can add new features to their system such as galleries, automatic logout and social media integration easily by installing appropriate modules.
2. The Drupal community has developed an interest in linked data technologies recently, introducing many useful modules to help integrate linked data into Drupal. By using pre-developed modules, the CURIOS project can focus on developing more generic and systematic solutions.
3. Most relevant to the CSCW community is that Drupal has a permissions and roles system that allows simple and effective implementation of collaborative user groups. For each operation, such as creating or updating existing content, there is a corresponding permission whose precondition only allows a special group of users to be able to perform that operation. For instance, if a record is maintained by a historical society, then only users within that historical society can edit or validate the record.

The CURIOS system stores its data in an RDF triplestore and performs SPARQL queries to create, maintain and retrieve the data. However, the complexities of Linked Open Data technologies are hidden away from end-users via a normal CMS user interface. Therefore, CURIOS allows users to consume and, more importantly, produce linked data without users' knowledge of the underneath technology. SPARQL is used as the query language in CURIOS as it is the predominant query language for RDF database format and allows the system to query data from multiple sources and present them as a unified result.

Unlike other approaches combining linked data and CMS, where linked datasets are imported into the CMS internal structures, CURIOS does not maintain the data in the CMS database and all queries and updates are performed directly against the RDF triplestore. As CURIOS does not store the data in Drupal's database, the module is loosely coupled to the dataset. This has the following advantages:

1. The dataset can be exported from the triplestore into different formats (e.g., XML, CSV, JSON…) at anytime without concerning Drupal's operations.

2. The same triplestore and dataset can be reused in other applications. For example, the CURIOS Mobile project also uses the same dataset and triplestore to provide an API so that mobile clients can reuse Hebridean Connections data within a tourist application.

Building the next generation of Hebridean Connections on open source software and web standards has distinct advantages for future development and the use of the system. The Drupal-based system can be maintained by its community of users and can be extended via additional functionality developed by the Drupal open source community, e.g., to support blogging or e-commerce features. This community led maintenance allows for further future development of the cultural repositories as the archives develop.

Validation and User Permissions

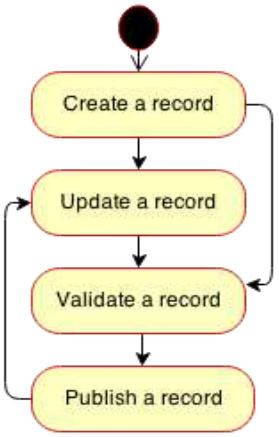

Figure 1: CURIOS record lifecycle

Currently, the record's metadata such as data owner, data maintainer, editing notes and publication status are stored together with the dataset in the RDF triplestore. This approach allows the datasets to be easily reusable in other applications and software without exporting data from Drupal's database. However, as different organisations have different group organisation, as well as policies regarding content creation, maintenance and validation process, CURIOS allows such groups/policies to be implemented in a flexible way by using Drupal permissions and roles system. For example, an "Editor" role allows users to create and update a record maintained by their group, e.g., a Comainn Eachdraidh group. However after being edited, the record will not be available to the public anymore as it needs to be validated by a "Validator", which is also a member of the same group. The lifecycle of a record within the CURIOS system is illustrated in Figure 1.

After a user creates a record, the record's metadata are automatically created by CURIOS: the owner and maintainer of the record is the user's group while the publication status is set to "pending". A user with editing permission can edit the pending record or already published records, given that the maintaining group of that record is also the user's group. However, note that an editor cannot edit the record's metadata and alls record after being edited will become "pending" automatically and hence needs validation. A user with validating permission from the same group can change the publication status of a pending record to "published". Note that in the Hebridean Connection case study, an editor and a validator, although come from the same group, have mutually exclusive permissions. In other words, an editor can only edit the record's content but not the metadata while a validator can only editor the metadata but not the content.

As a group in CURIOS is just a taxonomy-term field in the user's profile, it could be either shared among many users within a group, or made unique for each user. As the Hebridean Connections case study is a multiple group scenarios, we allowed users to shared a group name. In cases where the data is only maintained by group of individuals, we can easily turn the group name into the username, and hence allow a record to be owned and maintained by an individual instead of a group.

Building a CURIOS cultural heritage resource

CURIOS uses a Web Ontology Language (OWL) to describe its dataset classes and properties. Once the ontology is described, it is necessary to provide a separate description of which parts of the data (and the level of detail) are to be managed by the website. This description takes the form of an application-dependent configuration file which is loaded as part of the Drupal module. This file describes the classes, fields, and relationships to be shown in the website and how these relate to the constructs of the ontology [16]. Although the configuration file could be generated automatically, it is a declarative description and can easily be edited by hand.

The configuration file centralises the maintenance of the structure of the CMS with respect to the ontology, e.g., if a new type of page is required, the user can update the configuration and then run the Linked Data CMS mapping. When the CURIOS Drupal module is initialised, it automatically creates a set of Drupal resources and Views based on the configuration file, along with an additional set of pages allowing the linked data to be maintained. Drupal site administrators can then maintain the website generated by the configuration in the same way as a regular Drupal website.

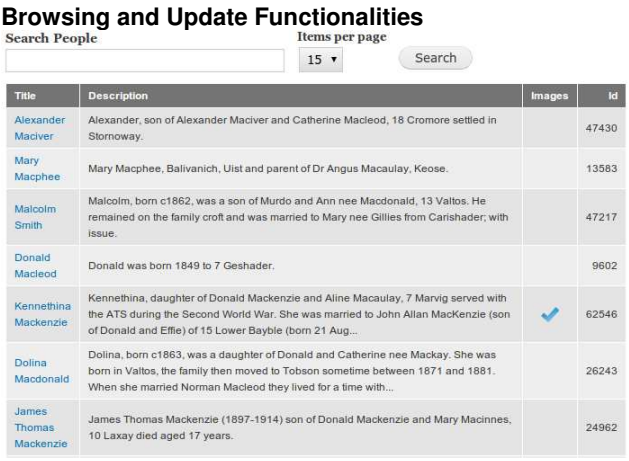

Figure 2: A list of people records

CURIOS allows users to create, update, or delete a record via a user-friendly GUI. For browsing, depending on the parameters such as the URI of a record or the resource types, CURIOS presents data in different ways. For instance, a list of records or details of a record will be displayed depending on whether the record URI is provided (see Figure 2). To support easy navigation between linked individuals, object properties of an RDF individual are presented as hyperlinks to other records as illustrated in Figure 3, instead of the normal text used for data type properties. CURIOS assists users entering data by providing different widgets depending on the data type the user wants to edit (Figure 4).

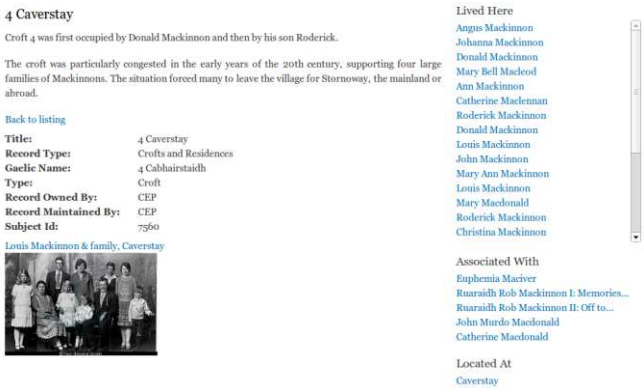

Figure 3: Details of a croft (A Croft is a small rented farm)

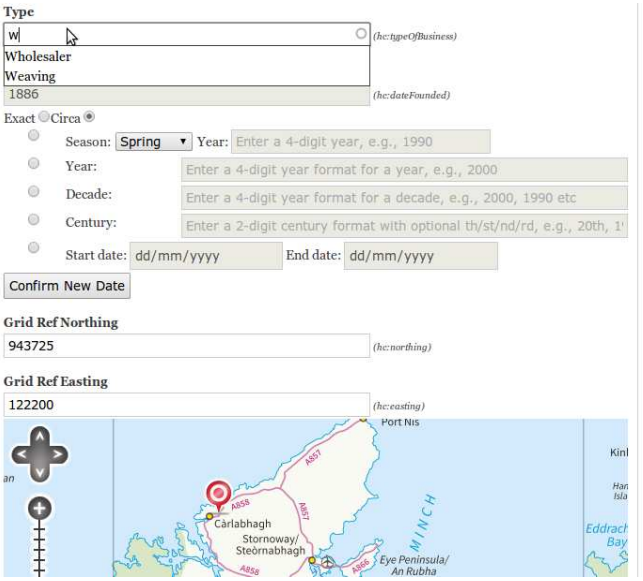

Figure 4: Updating special data types such as dates, geographical coordinates, etc.

For instance,

- Geographical coordinates, a map is displayed to allow users to choose a location rather than to type in the coordinates as text.
- To prevent users from entering incorrect values for some special properties such as an occupation or a type of place, an auto-complete widget is provided.
- It is typical that in the cultural heritage domain, temporal data such as dates are rather vague and not recorded in a consistent format. To facilitate users during data entry process, CURIOS provides a simple treatment to vague dates by introducing the hc:DateRange class which consists of two datetime datatype propertes: hc:dateFrom and hc:dateTo. A user can enter an exact date or a vague date such as a year, a season in a year, a decade, a century, etc, and CURIOS can convert the vague date into an appropriate instance of hc:DateRange. This mechanism separates the presentation of dates in the application-dependent interface from the actual data level.
- To manage object properties (i.e., links) between individuals, CURIOS allows property add and remove operations as presented in Figure 4, which are then mapped onto corresponding SPARQL update queries, e.g., INSERT and DELETE, to insert and remove appropriate triples.

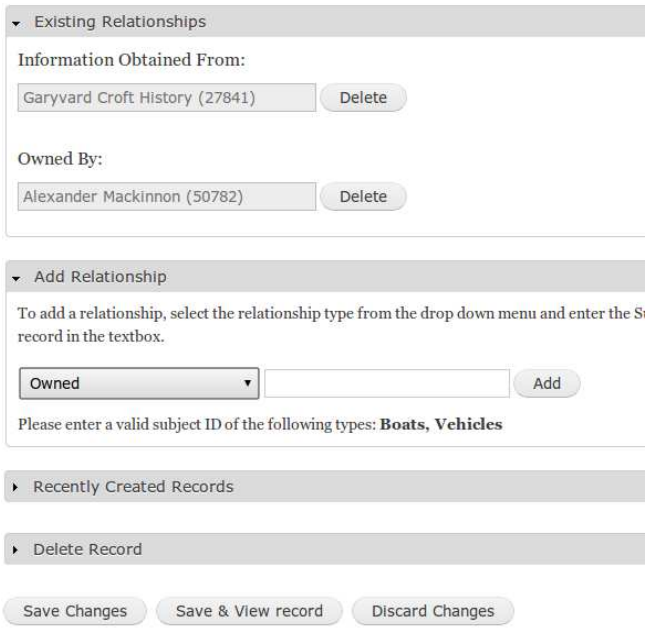

Figure 5: Adding/removing object properties

The CURIOS software has produced a linked data resource for Hebridean Connections documenting different types of information such as people, places, events, boats and businesses. The current instance of the Hebridean Connection data consists of over 700,000 RDF triples, incorporated within a relatively simple OWL ontology. The Drupal website, which was automatically built with the CURIOS tool, is already being used by Hebridean Connections to maintain their data; it is expected that the website will be opened to browsing by the general public shortly.

COLLABORATION

The CURIOS project is currently investigating the changing relationship the Comainn Eachdraidh groups are having with both each other and their collections as they continue the process of digitisation. We are conducting both individual semi-structured interviews with key members of different Comainn Eachdraidh and Hebridean Connections, focus groups/workshops with the different Comainn Eachdraidh groups and ethnographic research.

Through these interviews and focus groups the Comainn Eachdraidh members have repeatedly highlighted why community lead cultural heritage is important to them and how the activities are conducted for the community. Three examples of such points are highlighted in the following quotes:

> "Something about the community, in the *community, and created by the community itself."* (1)

> *"It's the view of the world as seen from Pairc" (3)*

> *"It's about preserving communities history and* culture and Gaelic and pictures and all that, for *the community itself" (9)*

Data Entry

The new CURIOS system has allowed a shift away from employing staff members specifically for data entry to allowing volunteers to manage and complete the data entry themselves. A Comann Eachdraidh has already inputted upwards of 300 new records into the system completely independently. This move towards empowering the volunteers to manage the process of digitising their repositories is essential in order to embed skills, which, in the long term, makes the process more sustainable:

> *"...because I'm now living in the area and saw an* advert for volunteers and time has become available that I wanted to start to take part. So *I've found that I'm now able to link an interest* with being able to learn more and hopefully learn new skills as well*" (13)*

Hebridean Connections Expansion

The new system has also encouraged more Comainn Eachdraidh groups to join the Hebridean Connections project, as the use of open linked data allowed their previous data sources to be cleaned and imported into the new system automatically. This is of key importance within cultural heritage as the sporadic nature of community lead cultural heritage repositories means there is likely to be many different digital archives in different data formats. Without importing the data into a consistent format they will never be able to be easily connected to the wider historical resources.

> *"*And slowly but surely it came round to the first phase of Hebridean Connections where four of the *Comunn Eachdraidh's were invited to take part.* We were quite interested in being one of these but they could only take four on the programme at that time. So the latest, this latest (endeavour), it allows everyone to come in who wants to come in *and we're quite pleased about that otherwise we* might be another four years before we could have joined*" (10)*

Ownership and Collaboration

Through initial evaluations it is becoming apparent that the collaborative nature of the CURIOS tool is allowing Comainn Eachdraidh groups to re-evaluate how they work together and share their collections. In the original proprietary software there was a very restrictive ownership applied to records. Only one group could own a record even if multiple groups had information for that record. This was particularly problematic for records about locations or

significant events such as the First World War. This restrictive setting meant if a group wanted to edit a record owned by another Comainn Eachdraidh they had to contact the group and agree transfer of the ownership to the new group. This was not always done as some Comainn Eachdraidh that had put more information into a record felt they should keep their ownership of their data. An example of this ownership issue was highlighted by a Comann Eachdraidh member that was present during the original digitisation:

> *"Well it's interesting seeing the different styles of* writing from different Comunn Eachdraidh but when I was putting in to the old system it was clear *we were going to have different styles, that's ok* – just conform on cer*tain things and that's all right….It happens quite a lot where people have* moved from one location to another, now, I think if a child is born on Berneray and then they move *away to get married elsewhere, it's often when the married life happens there's mo*re information and my feeling would be that the place where the people were married and had their married life *should own the record, although they'd have to be fed in other information. But then you've got* ministers and people who move around to various p*arishes…"* (7)

This issue was one that became very apparent again in the development of the CURIOS system and had no clear solution. In the new CURIOS system originally the data was designed so that ownership could be assigned to a Comann Eachdraidh but multiple groups could be assigned as maintainers by the owners. However it became apparent when the Comainn Eachdraidh groups met together that they were happy to have a much more collaborative approach with ownership no longer used and maintained by all groups that have added anything to a record. This change was observed during a visit to a training activity being conducted by Hebridean Connection with multiple Commainn Eachdraidh around data entry procedures. The change is discussed in detail in Ethnographic Observation 1.

The discussion outlined in Ethnographic Observation 1 lead to a complicated five step validation/user permission process being removed in favour of a simpler, more open process, where a user can edit a record then a second user can approve the changes. These changes in the Comainn Eachdraidh groups opinions to ownership and sharing of their records poses interesting questions around collaboration between groups through digital archives that need further investigation.

Whilst sitting in on a training session that Hebridean Connections were running to train volunteers to input data, a really interesting discussion opened up around the ownership of digital records. This has been something that has been of constant discussion in the delivery of the new system between CURIOS and HC due to the structure of HC and the autonomy of Comainn Eachdraidh (CE) groups. As independent groups, each have the desire to digitise and publish records from their collections, which has led to a need to collaborate. The discussions therefore centred around the publication of records; who owns them and who has the 'right' to edit such records, especially when an individual may appear in multiple CE. Previously the default choice had been to give the ownership of each record to the CE that first inputs the data and therefore they have the right to edit content and the decision of whether or not to open it up to others. However, the potential 'clunkiness' of this system for others, who may have more and equally valid information for that record becomes very difficult. If they don't have editorial rights, they have to seek permission from the owning CE to add their relevant data and as one pointed question asked 'how do you own a dead person?' especially one that lived in multiple CE areas.

The discussion therefore flowed between the want to retain localised repositories of knowledge and the realisation that in digital form this was not as possible as it used to be. The group wrestled with these issues for some time, representing how issues surrounding value, trust, and 'truth' were being changed through the need to collaborate in a digital archive. Different CEs appeared to have different senses of value towards their records and the digital forms they take. Some saw them as their 'property' directly linked to the people, land and the physical archive of that CE. Therefore, do you trust another group/volunteer to edit a representation of your 'physical' archive in the digital form and to do it in a manner that is accurate or 'truthful' to the record you are representing? Difficult questions but through this discussion, consensus began to develop that, for practicality reasons, a more open system would be better whereby CE would have to trust each other equally (though with some validation still) to edit and maintain each other's records.

Ethnographic Observation 1

FUTURE WORK

The next main developments within the CURIOS project will be the up-coming launch of the Hebridean Connections website to the general public. While investigating the use of the linked data cultural repositories by the public we will also focus on the on-going technical developments of CURIOS and also in-depth investigation of the Comainn Eachdraidh use of the tool. Focusing on:

- Investigation of long-term use by groups and evaluation of whether they change their working practices?

- How do the groups use their data once it's digitised?
- Who does the data entry / is it integrated into part of the collection process?
- Ownership of the records (not in terms of data protection / copyright) – does putting records into a collaborative, open, digital system change how the groups view their ownership or records?

There are also a number of on-going challenges in developing a sustainable user driven cultural repository. Firstly, as new cultural heritage vocabularies and datasets emerge, how can users be encouraged to integrate the Hebridean Connections cultural repository with these datasets and vocabularies? Secondly, the Drupal-based system is currently tailored to the cultural heritage domain. How should the approach be generalised so that it can be released to the open source community and applied to a wider variety of domains on the web? Finally, to investigate how semantic web reasoning technologies can be applied in a scalable way to automatically extract the implicit information from these cultural repositories?

The future technical developments of the system are going to focus on how to automate semantic linking to other data sources such as external cultural heritage repositories or dbpedia. Allow greater data visualisation within CURIOS - how do you visualise linked data in a meaningful way that clearly shows the different sources to the user? Allow users to greater query their data without requiring any specialised knowledge of linked data.

CONCLUSION

Open linked data can help make local cultural repositories sustainable and collective. Linked data allows for collaboration, mutual authoring, distributed responsibilities through community projects and the utilisation of other community or national resources [12]. The CURIOS project is enabling local cultural heritage repositories to become a meaningful identity resource for an international community, who previously had no access to them. By falling outside of national institutional frameworks, local people are the 'gatekeepers' of their own heritage and are selecting what to commemorate based on their own customs of remembering. This kind of digital archive can have, therefore, potentially significant social impacts which need to be better understood. The vision of Hebridean Connections is to expand the collections to incorporate those held by other Comainn Eachdraidh. Additionally, by making use of linked data, there is now the possibility to integrate further sources of data into HC from other historical societies or even national organisations.

ACKNOWLEDGMENTS

We would like to thank Hebridean Connections and the Comainn Eachdraidh for their ongoing commitment to this research. Many thanks to Kim Ross for her assistance in proof-reading however any mistakes held in the report are solely the responsibility of the authors. This work is supported by the Rural Digital Economy Research Hub (EPSRC EP/G066051/1).

REFERENCES

1. Balestrini, M., Bird, J., Marshall, P., Zaro, A., and Rogers, Y. Understanding sustained community engagement. Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI *'14*, ACM Press (2014), 2675–2684.
2. Berners-Lee, T., Hendler, J., and Lassila, O. The Semantic Web. Scientific American May, (2001), 29–37.
3. Bizer, C., Heath, T., and Berners-Lee, T. Linked Data - the story so far. International Journal on Semantic Web and Information Systems, 2009. http://eprints.soton.ac.uk/271285/1/bizer-heath-berners-lee-ijswis-linked-data.pdf.
4. De Boer, V., Wielemaker, J., van Gent, J., et al. Supporting linked data production for cultural heritage institutes: the amsterdam museum case study. *ESWC'12* Proceedings of the 9th international conference on The Semantic Web: research and applications, Springer Berlin Heidelberg (2012), 733–747.
5. Cresswell, T. Value, gleaning and the archive at Maxwell Street, Chicago. Transactions of the Institute of British Geographers 37, 1 (2012), 164–176.
6. Flinn, A. Community Histories, Community Archives: Some Opportunities and Challenges1. Journal of the Society of Archivists 28, 2 (2007), 151–176.
7. Han, K., Shih, P.C., Rosson, M.B., and Carroll, J.M. Enhancing community awareness of and participation in local heritage with a mobile application. Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing - CSCW *'14*, ACM Press (2014), 1144–1155.
8. Hitzler, P., Krotzsch, M., Parsia, B., Patel-Schneider, P., and Rudolph, S. OWL 2 Web Ontology Language Primer (Second Edition). 2012.
9. Van Hooland, S., Verborgh, R., De Wilde, M., Hercher, J., Mannens, E., and Van de Walle, R. Evaluating the success of vocabulary reconciliation for cultural heritage collections. Journal of the American Society for Information Science and Technology 64, 3 (2013), 464–479.
10. Mäkelä, E., Hyvönen, E., and Ruotsalo, T. How to deal with massively heterogeneous cultural heritage data–lessons learned in culturesampo. Semantic Web 3, (2012), 85–109.
11. Manola, F. and Miller, E. RDF Primer. W3C Recommendation, 2004. http://www.w3.org/TR/rdf-primer/.

12. Mellish, C., Wallace, C., Tait, E., Hunter, C., and MacLeod, M. Can Digital Technologies increase Engagement with Community History? Digital Engagement 2011.

13. Nora, P. Realms of memory: rethinking the French past. Volume 1: Conflicts and Divisions. Columbia University Press, 1996.

14. Spector, A.Z. Achieving application requirements. (1990), 19–33.

15. Tait, E., Macleod, M., Beel, D., Wallace, C., Mellish, C., and Taylor, S. Linking to the past: an analysis of community digital heritage initiatives. Aslib *proceedings : New information perspectives 65*, 6, 564–580.

16. Taylor, S., Jekjantuk, N., Mellish, C., and Pan, J.Z. Reasoning Driven Configuration of Linked Data Content Management Systems. JIST, Seoul, Korea, November 28-30, (2013).

17. Wallace, C., Tait, E., Macleod, M., Mellish, C., and Hunter, C. Supporting Digital Humanities Creating Sustainable Digital Community Heritage Resources Using Linked Data. Supporting Digital Humanities: Answering theunaskable Conference, (2011), 17–18.