21st International Congress on Modelling and Simulation, Gold Coast, Australia, 29 Nov to 4 Dec 2015
www.mssanz.org.au/modsim2015

Capturing Data Provenance With A User-Driven Feedback Approach

A. Devaraju[a] **and J. Klump**[a]

[a]*CSIRO Mineral Resources Flagship, PO Box 1130, Bentley, Western Australia, 6102.*
Email: anusuriya.devaraju@csiro.au

Abstract: Various portals have been developed to provide an easy way to discover and access public research data sets from various organizations. Data sets are made available with descriptive metadata based on common (e.g., OGC, CUAHSI, FGDC, INSPIRE, ISO, Dublin Core) or proprietary standards to facilitate better understanding and use of the data sets. Provenance descriptions may be included as part of the metadata and are specified from a data provider's perspective. These can include, for example, different entities and activities involved in a data creation flow, such as sensing platforms, personnel, and data calculation and transformation processes. Moving beyond the provider-centric descriptions, data provenance may be complemented with forward provenance records supplied by data consumers. The records may be gathered via a user-driven feedback approach. The feedback information from data consumers gives valuable insights into application and assessment of published data sets. This might include descriptions about a scientific analysis in which the data sets were used, the corrected version of an actual data set or any discovered issues and suggestions concerning the quality of the published data sets. Data providers might then use this information to handle erroneous data and improve existing metadata, their data collection and processing methods. Contributors can use the feedback channel to share their scientific analyses. Data consumers can learn more about data sets based on other people's experiences, and potentially save time by avoiding the need for interpreting or cleaning data sets. The goals of the study are to capture feedback from data users on published research data sets, link this to actual data sets, and finally support search and discovery of research data using feedback information. This paper reports preliminary results addressing the goals. We provide a summary of current practices on gathering feedback from end-users on research data portals, and discuss their relevance and limitations. Examples from the Earth Science domain on how commentaries from data users might be useful in practice are also included. Then, we present a data model representing key aspects of user feedback. We propose a system architecture to gather and manage feedback from end-users. We describe how the core PROV model may be used to represent the provenance of user feedback information. Technical solutions for linking feedback to existing data portals are also specified.

Keywords: *User feedback, Provenance, Open research data, Linked data, RESTful web service*

1 INTRODUCTION

Numerous collections of research data are available through public repositories on the Web, such as data sets hosted on Australian National Data Service[1], Canada Open Data[2], Data.gov.au[3] and Socrata OpenData[4]. Research data should be made available with relevant metadata to enable a better understanding of the data, and to facilitate data sharing and re-use. Provenance (also known as lineage) is a type of metadata that describes the entities and processes "involved in producing and delivering or otherwise influencing" a data set (Belhajjame et al., 2012). Scientists often require relevant information in order to better interpret and apply long term and unfamiliar research data in their applications. In this context, the provenance of research data has been perceived as essential, as "from it, one can *ascertain the quality of the data based on its ancestral data and derivations*, track back sources of errors [...]" (Simmhan et al., 2005, our emphasis). This importance has been addressed on several occasions, and various treatments of provenance associated with the quality of scientific data sets have been proposed. For a survey of related approaches and applications, see Simmhan et al. (2005); Glavic and Dittrich (2007); Freire et al. (2008); Moreau (2010). While we agree on the role for provenance in data quality assessment, the reality is that information, that is required to support these treatments, is insufficient. Providers usually describe how the data sets conform to their own product specifications, and the descriptions provided are rudimentary. Further, provenance records specified by providers are primarily focused on data creation, such as the source data, instrument, and transformation method associated with the creation of a published data set.

Forward provenance[5] focuses on how a resource is used after it has been created. We argue that moving beyond the provider-centric provenance information, research data should be made available with forward provenance records from users. These records may be obtained via a user-driven feedback approach. According to the Oxford English Dictionary, feedback refers to "information about reactions to a product, a person's performance of a task, etc. which is used as a basis for improvement"[6]. In the context of research data, examples of user feedback are comments, suggestions, content requests, usage and evaluation reports. Note that not all user feedback records are classified as forward provenance information. Examples of forward provenance records are applications and evaluations that users express with regard to published research data sets.

1.1 Motivation

Why does user feedback information matter in the context of research data? Provider-centric provenance metadata might give some basic guidance that would support the user's assessment of data fitness, for example to identify the source and methods involved in data creation. However, feedback information from data consumers gives a better insight into application and assessment of published data sets, such as the descriptions about a scientific analysis in which data sets were used, any issues related to the quality of the published data sets, corrections to data sets, related publications and users providing feedback. Figure 1 shows an example of corrected groundwater chemistry data sets provided by the Geological Survey of South Australia and correction notes produced by researchers (Gray and Bardwell, 2015). Linking the corrected data sets and the supporting documents to the existing data repository[7] can improve the re-usability of the data, reduce the duplication of effort in data handling, and potentially stimulate collaborations among researchers working in similar domains. Data providers can use the feedback information from users to handle erroneous data and improve their data collection and processing methods. For example, an issue tracking component installed as part of the Terrestrial Environmental Observatories (TERENO) data portal[8] is used by TERENO members to report any problems or issues related to data sets made available through the portal. The feedback information from the users is consulted by the data management team to handle erroneous data and improve the existing data processing and inspection methods (Devaraju et al., 2015).

In the context of research data, to the best of the authors' knowledge, little progress has been made to gather and exploit users' views on published data sets. Table 1 shows a list of research data portals in Australia and features available on the portals to elicit feedback on published data sets from end-users. Some portals offer research data sets covering a wide range of disciplines (e.g., RDA and CSIRO), while others are domain-

[1] http://www.ands.org.au/

[2] http://open.canada.ca

[3] http://data.gov.au/

[4] https://opendata.socrata.com/

[5] http://www.w3.org/TR/2013/WD-prov-aq-20130312/#forward-provenance

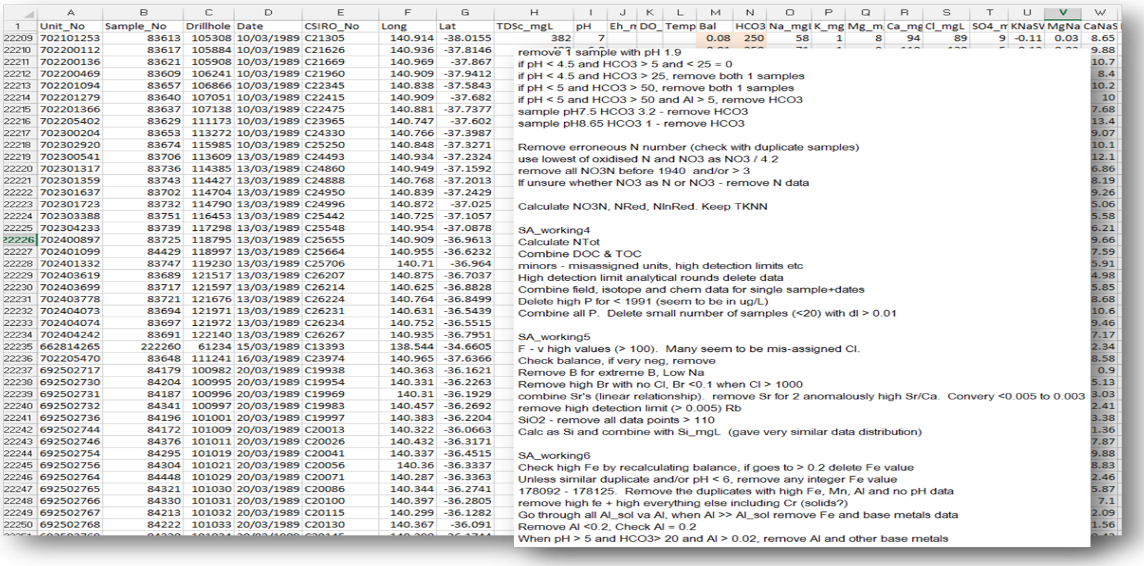
[6] http://www.oxforddictionaries.com

[7] South Australian Resource Information Geoserver: https://sarig.pir.sa.gov.au/MapViewerJS/

[8] http://teodoor.icg.kfa-juelich.de

specific (e.g., AODN, ALA and OzFlux). Most of the portals include email links and general contact forms. These feedback mechanisms are too simplistic and can lose vital context. Further, the commentaries from end users are rarely published. There are some portals that support on-line forums (e.g., the AODN's help forum[9] based on Drupal) and customer feedback service (e.g., the UserVoice service used by ALA). These are great ways to engage with end-users, and the details contributed by data users can be preserved for future use. Despite these advantages, at the present time, the feedback information is not explicitly linked to the source data or the existing metadata, and thereby cannot be discovered easily by other users interested in the same data set.

In the Earth Science domain, there are several data models addressing different aspects of user feedback. Some of the models are too simplistic as they lack key aspects required to represent feedback information, while others are either provider-centric or too complex. The ISO 19115[10] standard has been widely adopted for geographic data discovery, but only includes one concept (e.g., *MD_Usage*) for reporting data usage. ISO 19157[11] focuses on expressing quantitative measures of data quality; only one quality element, *DQ_UsabilityElement*, can be used to specify the suitability of a data set to requirements set by a data provider. A very closely related work is that by the GeoViQua project (Yang et al., 2013), which focused on the quality information of data sets in the Global Earth Observation System of Systems (GEOSS). The project has developed two data models - the User Quality Model (UQM) represents the users' perspective on data quality, whereas the Producer Quality Model (PQM) focuses on the providers' view of data quality. The UQM includes a comprehensive set of mechanisms for recording feedback, but is complex. The data model is based on predefined data structure as it incorporates concepts and relationships from ISO 19115 and 19157 standards. Further, a number of concepts in the model have not been clarified, e.g., distinction between different types of target, and the difference between *itemUnderReview* and *FeebackTarget*. The data model we developed is based on UQM and shares some similarities with it, but simplifies it to allow wider adoption of the model (see subsection 2.1).



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	Unit_No	Sample_No	Drillhole	Date	CSIRO_No	Long	Lat	TDSc_mgL	pH	Eh_n	DO_	Temp	Bal	HCO3	Na_mgl	K_mg	Mg_m	Ca_mg	Cl_mgL	SO4_m	KNaSV	MgNa	CaNaS
22209	702101253	83613	105308	10/03/1989	C21305	140.914	-38.0155	382	7				0.08	250	58	1	8	94	89	9	-0.11	0.03	8.65
22210	702200112	83617	105884	10/03/1989	C21626	140.936	-37.8146																9.88
22211	702200136	83621	105908	10/03/1989	C21669	140.969	-37.867																10.7
22212	702200469	83609	106241	10/03/1989	C21960	140.909	-37.9412																8.4
22213	702201094	83657	106866	10/03/1989	C22345	140.838	-37.5843																10.2
22214	702201279	83640	107051	10/03/1989	C22415	140.909	-37.682																10
22215	702201366	83637	107138	10/03/1989	C22475	140.881	-37.7377																7.68
22216	702205402	83629	111173	10/03/1989	C23965	140.747	-37.602																13.4
22217	702300204	83653	113272	10/03/1989	C24330	140.766	-37.3987																9.07
22218	702302920	83674	115985	10/03/1989	C25250	140.848	-37.3271																10.1
22219	702300541	83706	113609	13/03/1989	C24493	140.934	-37.2324																12.1
22220	702301317	83736	114385	13/03/1989	C24860	140.949	-37.1592																6.86
22221	702301359	83743	114427	13/03/1989	C24888	140.768	-37.2013																8.19
22222	702301637	83702	114704	13/03/1989	C24950	140.839	-37.2429																9.26
22223	702301723	83732	114790	13/03/1989	C24996	140.872	-37.025																5.06
22224	702303388	83751	116453	13/03/1989	C25442	140.725	-37.1057																5.58
22225	702304233	83739	117298	13/03/1989	C25548	140.954	-37.0878																6.21
22226	702400897	83725	118795	13/03/1989	C25655	140.909	-36.9613																9.66
22227	702401099	84429	118997	13/03/1989	C25664	140.955	-36.6232																7.59
22228	702401332	83747	119230	13/03/1989	C25706	140.71	-36.964																5.91
22229	702403619	83689	121517	13/03/1989	C26207	140.875	-36.7037																4.98
22230	702403699	83717	121597	13/03/1989	C26214	140.625	-36.8828																5.85
22231	702403778	83721	121676	13/03/1989	C26224	140.764	-36.8499																8.68
22232	702404073	83694	121971	13/03/1989	C26231	140.631	-36.5439																10.6
22233	702404074	83697	121972	13/03/1989	C26234	140.752	-36.5515																9.46
22234	702404242	83691	122140	13/03/1989	C26267	140.935	-36.7951																7.17
22235	662814265	222260	61234	15/03/1989	C13393	138.544	-34.6605																2.34
22236	702205470	83648	111241	16/03/1989	C23974	140.965	-37.6366																8.58
22237	692502717	84179	100982	20/03/1989	C19938	140.363	-36.1621																0.9
22238	692502730	84204	100995	20/03/1989	C19954	140.331	-36.2263																5.13
22239	692502731	84187	100996	20/03/1989	C19969	140.31	-36.1929																3.03
22240	692502732	84341	100997	20/03/1989	C19983	140.457	-36.2692																2.41
22241	692502736	84196	101001	20/03/1989	C19997	140.383	-36.2204																3.38
22242	692502744	84172	101009	20/03/1989	C20013	140.322	-36.0663																1.36
22243	692502746	84376	101011	20/03/1989	C20026	140.432	-36.3171																7.87
22244	692502754	84295	101019	20/03/1989	C20041	140.337	-36.4515																9.88
22245	692502756	84304	101021	20/03/1989	C20056	140.36	-36.3337																8.83
22246	692502764	84448	101029	20/03/1989	C20071	140.287	-36.3363																2.46
22247	692502765	84321	101030	20/03/1989	C20086	140.344	-36.2741																5.87
22248	692502766	84330	101031	20/03/1989	C20100	140.397	-36.2805																7.1
22249	692502767	84213	101032	20/03/1989	C20115	140.299	-36.1282																2.09
22250	692502768	84222	101033	20/03/1989	C20130	140.367	-36.091																1.56

remove 1 sample with pH 1.9
if pH < 4.5 and HCO3 > 5 and < 25 = 0
if pH < 4.5 and HCO3 > 25, remove both 1 samples
if pH < 5 and HCO3 > 50, remove both 1 samples
if pH < 5 and HCO3 > 50 and Al > 5, remove HCO3
sample pH7.5 HCO3 3.2 - remove HCO3
sample pH8.65 HCO3 1 - remove HCO3

Remove erroneous N number (check with duplicate samples)
use lowest of oxidised N and NO3 as NO3 / 4.2
remove all NO3N before 1940 and/or > 3
If unsure whether NO3 as N or NO3 - remove N data

Calculate NO3N, NRed, NInRed. Keep TKNN

SA_working4
Calculate NTot
Combine DOC & TOC
minors - misassigned units, high detection limits etc
High detection limit analytical rounds delete data
Combine field, isotope and chem data for single sample+dates
Delete high P for < 1991 (seem to be in ug/L)
Combine all P. Delete small number of samples (<20) with dl > 0.01

SA_working5
F - v high values (> 100). Many seem to be mis-assigned Cl.
Check balance, if very neg, remove
Remove B for extreme B, Low Na
Remove high Br with no Cl, Br <0.1 when Cl > 1000
combine Sr's (linear relationship). remove Sr for 2 anomalously high Sr/Ca. Convery <0.005 to 0.003
remove high detection limit (> 0.005) Rb
SiO2 - remove all data points > 110
Calc as Si and combine with Si_mgL (gave very similar data distribution)

SA_working6
Check high Fe by recalculating balance, if goes to > 0.2 delete Fe value
Unless similar duplicate and/or pH < 6, remove any integer Fe value
178092 - 178125. Remove the duplicates with high Fe, Mn, Al and no pH data
remove high fe + high everything else including Cr (solids?)
Go through all Al_sol va Al, when Al >> Al_sol remove Fe and base metals data
Remove Al <0.2, Check Al = 0.2
When pH > 5 and HCO3> 20 and Al > 0.02, remove Al and other base metals

Figure 1. Corrected groundwater chemistry data and a list of changes made to the data.

1.2 Goals and Scope

The goals of the study are to capture user feedback on published research data and then link these to actual data sets, and finally to support search and discovery of research data with this feedback information. In this paper, we report preliminary results to achieve the goals. We present a data model representing key aspects of user

[9] http://portalhelp.aodn.org.au/Portal2_help/?q=forum

[10] http://www.iso.org/iso/catalogue_detail.htm?csnumber=26020

[11] http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=32575

feedback (subsection 2.1). We illustrate a high level architecture of a system to gather and manage feedback information from data consumers (subsection 2.2). We demonstrate the application of the core PROV model to represent the provenance of user feedback information (subsection 2.3). Section 3 concludes the paper with some directions for future work.

Table 1. A list of research data portals and feedback mechanisms.

Research Data Portals	Data Collections	Feedback Mechanism
Research Data Australia (RDA)[12]	Research data	General feedback form, and user contributed tags for data discovery.
CSIRO Data Access Portal[13]	Research data published by CSIRO	Refer to the email of the data collector in the metadata.
TERN Data Discovery Portal[14]	Australia's terrestrial ecosystem data	General contact form
Australian Ocean Data Network Portal (AODN)[15]	Ocean	General contact form and portal help forum.
Atlas of Living Australia (ALA)[16]	Biodiversity	UserVoice feedback portal
OzFlux Data Portal[17]	Flux data	Email link (for all inquiries and assistance).
National Marine Mammal Data Portal[18]	Marine mammal conservation	General feedback form
Urban Research Infrastructure Network.[19]	Urban settlements	Email link for general inquiries, Social media buttons for distribute the link of a data set.

[12] https://researchdata.ands.org.au/
[13] https://data.csiro.au/dap/home?execution=e1s1
[14] http://portal.tern.org.au/
[15] http://portal.aodn.org.au/aodn/
[16] http://www.ala.org.au/
[17] http://data.ozflux.org.au/portal/home.jspx
[18] https://data.marinemammals.gov.au/
[19] http://data.aurin.org.au/

2 PRELIMINARY RESULTS OBTAINED

This section includes a description of the data model representing key aspects of user feedback and the architecture of a system to facilitate the capture and access of feedback data from users.

2.1 User Feedback Representation

Figure 2 shows the relational data model to capture user feedback information. A *collection* of feedback comprises one or more *feedback items*. A feedback item can be described in terms of who, what, when, where and why it is reported by a data user. A collection is targeted at one or more data sets. Any data sets with a Uniform Resource Identifier (URI) can be a valid *feedback target*. A target data set may be associated with *context* descriptions, e.g., related data portal and data creation information. Several tables have been developed to populate controlled vocabularies, e.g., target types, feedback status and feedback types. The *feedback types* indicates the possible intentions of a data consumer to provide feedback, and these are compiled from existing literature (Schneider, 2011; Morales-Ramirez et al., 2014; Pagano and Maalej, 2013) on feedback from software users. Examples are comment (recommendations and references to other related sources), requirement (new feature and content request, shortcomings and discovered issues), clarification request, rating and user experience. *Supplementary files* refer to additional documents supporting feedback from users.

There are some similarities between our data model and the GeoViQua's UQM. For example, the grouping of feedback items into a collection and the relation between a feedback item and its contributor. Nevertheless, our model differs from UQM in several aspects. The UQM is designed using a class-based modeling, whereas we have developed a relational model to represent the feedback concepts. We propose a Linked Data approach to publish the feedback information so that it can be shared between different sources (see subsection 2.2). In UQM, some attributes of the classes are restricted to the concepts defined in PQM[20]. We do not impose such restrictions in our model. UQM allows many different options and one-to-many relations between feedback item and its related classes, e.g., *UserComment*, *Rating*, and *UsageReport*. We have simplified this by treating them as *feedback types*. The *feedback_item* table can be extended with additional tables to include the user rating support for datasets if required. Further, in our approach, the relation between a feedback and its targets have also been clarified. One aspect covered by the UQM that is not fully specified in our model is the summary of feedback data (e.g., *tagCount, domainUsageCount, numberOfPublications and numberOfRatings*.)

[20] The PQM is developed based on ISO 19157.

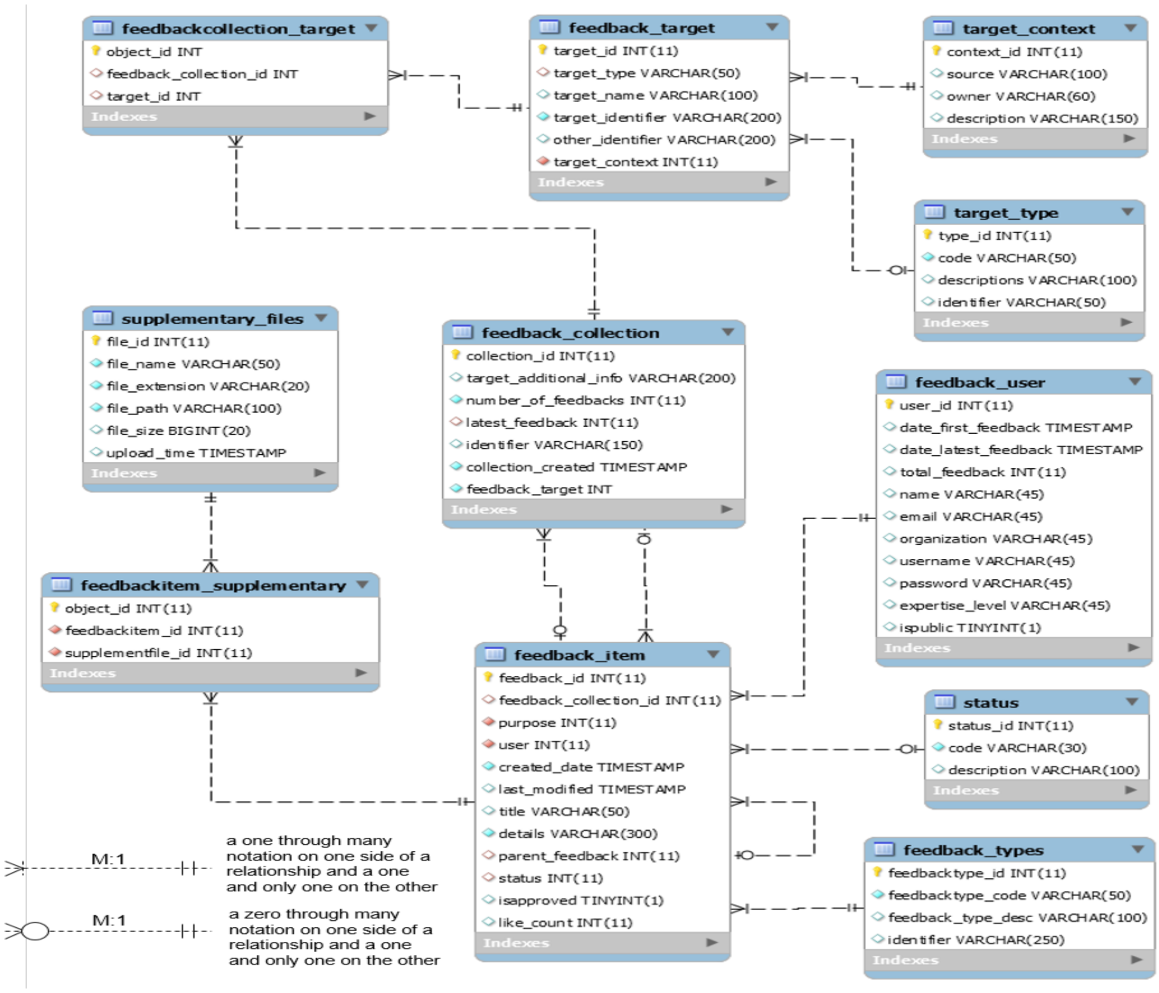


Figure 2. The entity-relationship diagram of user feedback using Crow's Foot notation.

Our data model only specifies a number of instances of feedback and the latest feedback for each feedback collection.

2.2 System Architecture

Figure 3 shows the system architecture of the feedback system. Users contribute feedback via a JavaScript browser plug-in. The plug-in is designed after the ivoviz open-source project[21]. The OAuth[22] framework handles authentication of external users. The CSIRO's active directory is used to authenticate and access information about internal users. The RESTful Web Service retrieves, creates and updates feedback data. The service's requests and responses are specified in JSON notation as it is easy to load and process the data structure within the JavaScript plug-in. Feedback records are stored in a database implemented in MySQL. The D2RQ[23] platform converts records into Resource Description Framework (RDF) graph data format.

2.3 Provenance of User Feedback Information

Feedback records can be shared in a flexible and extensible manner across the Web by adopting the Linked Data approach. The records may be published using several existing specifications, e.g., Dublin Core Metadata Terms[24] and W3C Provenance Ontology (PROV-O)[25]. The PROV-O represents the PROV Data Model in

[21] https://github.com/ivoviz/feedback
[22] http://oauth.net/
[23] http://d2rq.org/
[24] http://dublincore.org/documents/dcmi-terms/
[25] http://www.w3.org/TR/prov-o/

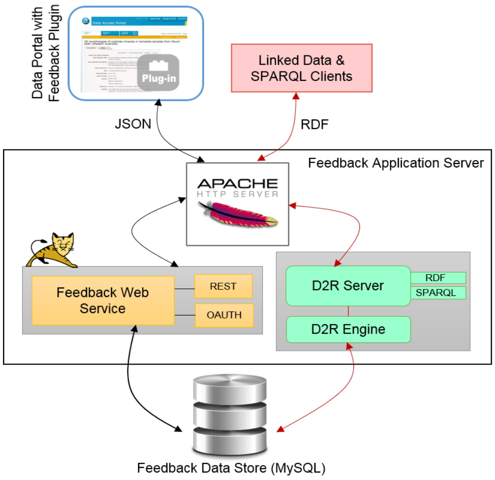
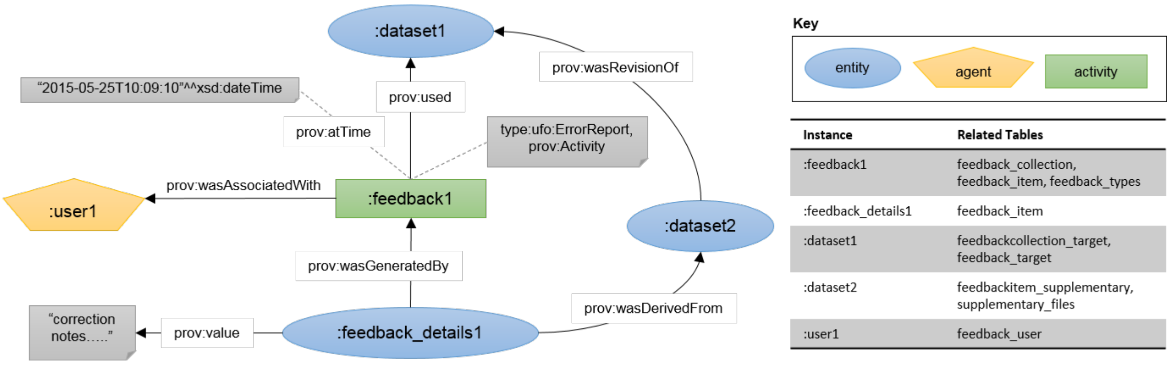

Data Portal with Feedback Plugin
Plug-in
Linked Data & SPARQL Clients
JSON
RDF
Feedback Application Server
APACHE HTTP SERVER
Feedback Web Service
REST
OAUTH
D2R Server
RDF
SPARQL
D2R Engine
Feedback Data Store (MySQL)

Figure 3. Architecture of the feedback system.



Instance	Related Tables
:feedback1	feedback_collection, feedback_item, feedback_types
:feedback_details1	feedback_item
:dataset1	feedbackcollection_target, feedback_target
:dataset2	feedbackitem_supplementary, supplementary_files
:user1	feedback_user

Figure 4. Entities and an agent involved in an error report feedback activity.

OWL2 Web Ontology Language. The provenance ontology is useful to clarify the contributor (*prov:Agent*), the target data set (*prov:Entity*), the feedback activity and its outcomes (*prov:Entity*). Figure 4 illustrates an error report activity that generated feedback details (e.g., correction notes) of a published data set (*dataset1*). The *dataset2*, a corrected version of the published data set, is the related source from which the feedback details were derived.

3 DISCUSSIONS AND CONCLUSIONS

In this paper, we presented a user-centric approach to capture forward provenance information, e.g., application and assessment of published data sets. This structured information may help data producers to enhance the quality of their data. We contrasted this against the provider-centric approaches which focus on data creation and release. We described a data model representing key aspects of user feedback and proposed a system architecture to gather, manage and publish feedback information from data users. The data model is kept at a sufficiently general level to apply it to different use cases. Although the feedback mechanism focuses on datasets, it can be applied to an instrument, a specimen or a project with an identifier. The system architecture provides new capabilities in terms of gathering, managing and publishing user feedback information, by combining a number of open-source technologies. The potential benefit of publishing the feedback records as Linked Data is the interoperability with other systems on the Web. The provenance of feedback information is addressed by incorporating PROV-O concepts.

Our ongoing work focuses on implementing the feedback service and the plug-in, and testing them with the

CSIRO data access portal. The proposed feedback model will also be extended based on the database schema[26] of the JIRA issue tracking system to capture change history and priority levels of feedback. Moderation of feedback information is important, but is not our primary concern at this stage of development. Moderation capability will be added into the system when there are sufficient instances of feedback gathered from end-users.

An important aspect in developing the feedback system is identifying usability features that motivate feedback contributors. We are currently exploring approaches in social media (e.g., up-voting and down-voting, point scoring and sharing buttons) to design a system that will encourage users to contribute their views on published datasets. Minimizing required inputs, handling possible errors and offering privacy controls are also vital to improve the user experience.

ACKNOWLEDGEMENT

The authors would like to thank Josh Vote from the Mineral Resources Flagship, CSIRO for his suggestions on the architecture of the user feedback system.

REFERENCES

Belhajjame, K., H. Deus, D. Garijo, G. Klyne, P. Missier, S. Soiland-Reyes, and S. Zednik (2012). Prov model primer. Technical report, W3C.

Devaraju, A., S. Jirka, R. Kunkel, and J. Sorg (2015). Q-SOS - A sensor observation service for accessing quality descriptions of environmental data. *Special Issue on Open Geospatial Science and Applications, ISPRS International Journal of Geo-Information 4*, 1346–1365.

Freire, J., D. Koop, E. Santos, and C. T. Silva (2008). Provenance for computational tasks: A survey. *Computing in Science and Engineering 10*(3), 11–21.

Glavic, B. and K. R. Dittrich (2007). Data provenance: A categorization of existing approaches. In A. Kemper, H. Schning, T. Rose, M. Jarke, T. Seidl, C. Quix, and C. Brochhaus (Eds.), *BTW*, Volume 103 of *LNI*, pp. 227–241. GI.

Gray, D. J. and N. Bardwell (2015). Hydrogeochemistry from South Australia - Data Release: Accompanying Notes (Working Draft). Technical report, CSIRO.

Morales-Ramirez, I., A. Perini, and R. Guizzardi (2014). Providing foundation for user feedback concepts by extending a communication ontology. In E. Yu, G. Dobbie, M. Jarke, and S. Purao (Eds.), *Conceptual Modeling*, Volume 8824 of *Lecture Notes in Computer Science*, pp. 305–312. Springer International Publishing.

Moreau, L. (2010, November). The foundations for provenance on the web. *Foundations and Trends in Web Science 2*(2–3), 99–241.

Pagano, D. and W. Maalej (2013). User feedback in the appstore: An empirical study. In *Requirements Engineering Conference (RE), 2013 21st IEEE International*, pp. 125–134.

Schneider, K. (2011, Aug). Focusing spontaneous feedback to support system evolution. In *Requirements Engineering Conference (RE), 2011 19th IEEE International*, pp. 165–174.

Simmhan, Y. L., B. Plale, and D. Gannon (2005). A survey of data provenance techniques. Technical Report 612, Computer Science Department, Indiana University. Extended version of SIGMOD Record 2005.

Yang, X., J. D. Blower, L. Bastin, V. Lush, A. Zabala, J. Mas, D. Cornford, P. Daz, and J. Lumsden (2013). An integrated view of data quality in earth observation. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 371*(1983), 20120072.

[26] https://developer.atlassian.com/jiradev/jira-architecture/database-schema