

Integrating Text and Graphics to Present Provenance Information*

Thomas Bouttaz, Alan Eckhardt, Chris Mellish, and Peter Edwards

Computing Science, University of Aberdeen, Aberdeen AB24 5UA, UK
{t.bouttaz,a.eckhardt,c.mellish,p.edwards}@abdn.ac.uk

Abstract. We describe two approaches for the visualisation of provenance - one using natural language generation to produce texts, the other using a graphical approach. Our main contribution is a mechanism using a combination of these modalities.

Keywords: provenance, NLG, visualisation, HCI.

1 Introduction

The presentation of provenance should allow users to intuitively understand such information, without requiring any particular knowledge about the underlying model. Therefore, identifying appropriate means to present provenance information to end-users is an issue that needs to be addressed. In this paper, we discuss two ways to present this information, one based on generated textual descriptions and the other based on a graphical visualisation¹. Our approach was motivated by the needs of multidisciplinary users within a web-based research environment *ourSpaces* [1], but is applicable in other contexts.

*ourSpaces*² enables researchers from different backgrounds to manage their projects by allowing them to communicate and share their research artefacts. The underlying architecture of *ourSpaces* is based on Semantic Web technologies (e.g. OWL, RDF) and at the heart of *ourSpaces* is an OWL representation of the Open Provenance Model [2]. OPM is a generic model and as a result, *ourSpaces* also supports additional domain specific provenance ontologies that are created by extending the concepts defined in the OPM ontology. Several services make use of these ontologies, including *Metadata Access* - provides Java access to a RDF data repository, and *Provenance* - manages provenance data and provides HTTP and Java access.

2 Visualisation Services

Natural Language Visualisation. We have developed a Natural Language Generation (NLG) service that generates short textual descriptions based on

* This work is supported by the UK Economic & Social Research Council (ESRC) under the Digital Social Research programme; award RES-149-25-1075.

¹ The software is open source and available at: <https://github.com/Policygrid-II>

² <http://www.ourspaces.net>

The artifact focus group data: It was used by [the interview of Alan Eckhardt](#) and was generated by [the resource upload process](#).

The process Interview of Alan Eckhardt: It used [the focus group data](#) and generated [the recording of the interview of Alan Eckhardt](#). It was controlled by [Edoardo Pignotti](#) and [Pete Edwards](#), and involved [Alan Eckhardt](#).

Fig. 1. Example of generated texts containing provenance information

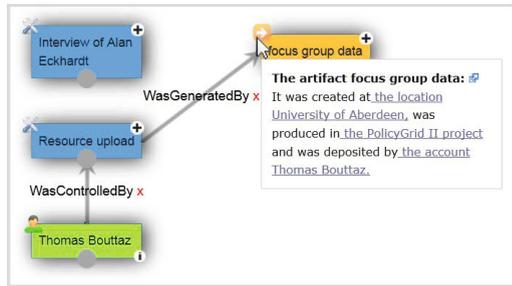


Fig. 2. Integration of textual and graphical visualisations

RDF data. We have implemented two modes for this service: one for generating text about general information regarding an entity (e.g. type, title, date of creation), and one specifically about its provenance (i.e. how it relates to other OPM entities).

The NLG service is composed of two components - *Text Generator* and *Text Formatter*. First, the metadata about an entity is obtained from the *Metadata access* service, which is then used by the *Text Generator* to build a semantic model representing information about the entity. Then the model is transformed into text using the appropriate *Language Specification* files. These XML files are divided into two categories: *Property Language Specifications* contain the linguistic information required to structure the sentence corresponding to a property (e.g. syntactic category, verb tense); *Class Language Specifications* indicate which properties should be used to refer to a particular class.

The *Text Formatter* converts the plain text generated by the *Text Generator* into HTML. Figure 1 shows an example of a text generated by the NLG service. This text contains a description of the provenance of the artefact (*focus group data*) and can be further expanded by clicking on hyperlinks to related entities. This calls the service with the URI of that related entity, appending the resulting description to the original text.

Graphical Visualiser. The *Graphical Visualiser* was developed to visualise provenance information using a graph displaying OPM entities as nodes and OPM causal relationships as directed connections between nodes. It is implemented in HTML and Javascript and divided into two main components - one for communication and one for visualisation. The communication component makes use of the *Provenance* service to query for provenance data, as well to update the provenance graph. The data are then used by the visualisation component to draw the actual graph. The nodes are initially ordered according to timestamp, but the user can control the position of nodes, as well as the level of magnification. Each node has a “plus” icon that loads the provenance of this node when clicked, thus expanding the provenance graph. An example of a graphical visualisation is shown in Figure 2. A user is able to create and edit

provenance information using the graphical visualisation. He/she can create new links between nodes, create new processes and delete links.

Combining Visualisations. We have integrated the NLG service within the graphical presentation, by allowing users to generate textual descriptions of the different entities present in the graph. In this way, a user can easily access more information about entities. Figure 2 shows how a user can generate the description of an entity (*focus group data*) by hovering the mouse pointer over its node.

3 Discussion

In order to evaluate this work, we ran a series of focus groups with potential users of *ourSpaces*. Users were presented with the same provenance information presented in different ways and were asked to discuss the main advantages and drawbacks of each. The participants preferred the terminology to adapt to the context of use (e.g. referring to the agent that controlled an interview process as the *interviewer*, rather than with `wasControlledBy`). This suggests that more adaptation of the underlying representation of provenance information may be required to improve usability. They also felt that some form of temporal information would help them to better understand the provenance presentations. The participants did find that the combination of graphical and textual presentations was useful to better understand provenance.

This work differs from other provenance visualisation services by providing users with two presentation modalities supporting each other, as well as a mechanism to allow users to edit the provenance metadata. In order to further improve this work, we could also implement other “Types” described by [3], such as *Comparison* (comparing the provenance of two artefacts, highlighting their differences) and *Participation* (emphasising the agents involved in the processes) which would be beneficial in the context of a collaborative environment such as *ourSpaces*. Currently, the visualisation conforms with *Timeline* (chronological ordering) and *Result* (focus on the main artefact).

References

1. Edwards, P., Mellish, C., Pignotti, E., Ponnampereuma, K., Bouttaz, T., Eckhardt, A., Pangbourne, K., Philip, L., Farrington, J.: Demo: ourspaces - a provenance enabled virtual research environment. In: 4th International Provenance & Annotation Workshop (2012)
2. Moreau, L., Clifford, B., Freire, J., Futrelle, J., Gil, Y., Groth, P., Kwasnikowska, N., Miles, S., Missier, P., Myers, J., Plale, B., Simmhan, Y., Stephan, E., den Bussche, J.V.: The open provenance model core specification (v1.1). *Future Gener. Comput. Syst.* 27, 743–756 (2011)
3. Kunde, M., Bergmeyer, H., Schreiber, A.: Requirements for a Provenance Visualization Component. In: Freire, J., Koop, D., Moreau, L. (eds.) *IPAW 2008*. LNCS, vol. 5272, pp. 241–252. Springer, Heidelberg (2008)