Social Dendro: Social Network Techniques Applied to Research Data Description

Nelson Pereira(✉), João Rocha da Silva, and Cristina Ribeiro

INESC TEC, Faculdade de Engenharia, Universidade do Porto,
Rua Dr. Roberto Frias, 4200-465 Porto, Portugal
nelsonpereira1991@gmail.com, joaorosilva@gmail.com, mcr@fe.up.pt

Abstract. Research data management has become an integral part of the research workflow. Currently, concern with data appears mainly at the very last stages of projects, rather than being present from the moment of data creation. The goal of this work is to make data easier to find, share and reuse through early metadata production and in-group review. The approach proposed in this paper, Social Dendro, introduces social network concepts such as posts, shares and comments, in Dendro, our research data management platform. The implementation follows the ontology-based architecture of the platform. Results of a preliminary user test have provided insights for future improvements.

Keywords: Data repository · Data curation · Research data management · Social networks · User interfaces · Ontologies

1 Introduction

Research data management (RDM) is a very complex problem involving a multitude of stakeholders and issues ranging from the social to the technical [2]. Furthermore, it is becoming an essential part of research workflows, as funding institutions are either recommending [4] or requiring [5] the inclusion of Data Management Plans (DMPs) in research project proposals.

A series of surveys conducted by the Data Archiving and Networked Services (DANS) showed that large amounts of information created during the research process, from 70 to 90%, is not stored outside of the lab context [3]. It is also clear that researchers need to be involved in the data curation process, because they know much about the data, and such information is crucial for producing the quality metadata required to interpret datasets. However, their engagement relies on the availability (existence) of tools that can handle the details and formalities involved in the practice of data curation, so that their focus can remain on the research activity.

Dendro[1] is a collaborative data management platform currently in use at the University of Porto. Its goal is to provide a friendly, "Dropbox"-like

[1] Link: http://github.com/feup-infolab/dendro.

J. Kamps et al. (Eds.): TPDL 2017, LNCS 10450, pp. 566–571, 2017.
DOI: 10.1007/978-3-319-67008-9_47

interface combined with data description features built over ontologies [7]. Dendro's main focus is the description of the files and folders created by research groups. Dendro is designed to assist researchers during the data production phase and complements existing repository platforms, supporting researchers in the organization and description of the data before they are deposited in repositories such as Zenodo, Figshare, the EUDAT B2Share or CKAN.

In this paper, we present the design and implementation of Social Dendro, an extension that draws inspiration from the Science Repositories 2.0 concept [1]. Data management is regarded there as a social process, involving researchers right from the moment of data creation—covering deposit, reuse and replication of experiments. We carried out a preliminary analysis for best suited ontologies, allowing us to build the data model for Social Dendro and integrate it in Dendro's data model [6]. We then carried out a small usability test covering data description in this collaborative setting.

2 Social Dendro

Being aware of work by others is one of the cornerstones of collaboration systems [8]. Given the collaboration features that are already in place and its graph data model, Dendro is a good base for the implementation of the vision presented in Science 2.0 repositories.

This approach advocates that the sharing and reuse of information should occur as early as possible in the research cycle [1], and goes on to propose the introduction of posts, ratings, comments, and likes as a way to review the research activity and its products. The nurturing of tacit knowledge, which is mostly shared by discussion with other individuals, is also quite important in the context of the research activity, as it can influence the reproducibility of results in experiments. Social Dendro is expected to preserve this very valuable type of knowledge, via social network techniques, during and after the research activity.

2.1 Data Model

Following the principles of data sharing and reuse proposed by the Semantic Web initiative[2], we have identified ontologies and classes that match the concepts of like, share, and comment. The schema.org ontology[3] already has a set of classes for these concepts, namely `SocialMediaPosting`, `ShareAction`, `CommentAction` and `LikeAction`. They were adopted in Social Dendro.

As an example, consider a user filling in descriptor "Creator", from the Dublin Core Terms ontology[4], for a folder. A new `SocialMediaPosting` instance is created as a result (see **1** in Fig. 1).

[2] Link: https://www.w3.org/standards/semanticweb/.
[3] Link: http://schema.org/.
[4] Available at http://bloody-byte.net/rdf/dc_owl2dl.

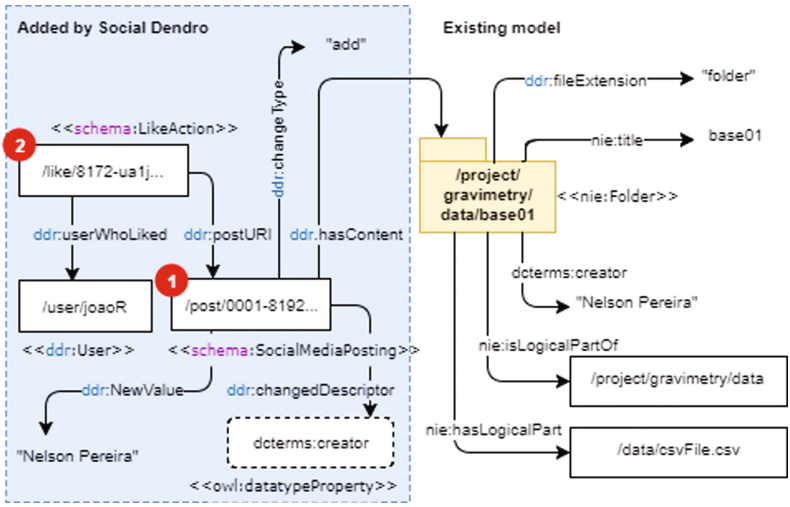

Fig. 1. Social Dendro creates a post as a descriptor is added

In the "Added by Social Dendro" section of Fig. 1, we can see the data changes triggered by this Social Dendro event in the Dendro graph. Properties `newValue`, `changedDescriptor`, `hasContent` and `changeType` are defined in the Dendro ontology, represented here with the `ddr` prefix. The Social Dendro event generates instances for these properties as depicted.

Following this interaction, another user liked the `SocialMediaPosting`. This is represented in the graph as a `LikeAction` instance (see **2**) associated to this `SocialMediaPosting` object, with the properties `userWhoLiked` and `postUri` identifying the user who liked the post and the post that was liked.

3 Usability Tests

The implementation of Social Dendro was validated through usability tests with a set of 8 users. The subjects included researchers, students and software developers. Each experiment consisted in a sequence of real-time interactions between two users in the Dendro platform under the close monitoring of two evaluators.

Dendro is a web application, so we considered the five quality components specified by Jacob Nielsen to test website interfaces: learnability, efficiency, memorability, errors and satisfaction[5]. Memorability was left out due to the short time span of the evaluation.

We requested the participants to fill a short questionnaire before the start of the experiment. The questions are in Table 1; some of them have an open-ended nature, while others ask the participants to rate their experience with a specific concept in a scale of 1 to 5, low to high.

[5] Link: https://www.nngroup.com/articles/usability-101-introduction-to-usability/.

Table 1. Questions of the preliminary questionnaire

ID	Description
QI1	Do you have any previous experience with the Dendro platform?
QI2	Please designate your degree of experience with research data management (1–5)
QI3	What do you usually do with the data from your projects?
QI4	Please rate how frequently you use social networks (1–5)

The results for question **QI1** showed that 6 out of 8 people had previously interacted with Dendro. Note however that, for all participants, this was the first interaction with the Social Dendro extension.

Results for **QI2** showed that 3 out of the 8 individuals acknowledge to have a very low experience level in research data management (level 2). Also 3 in 8 claim to have level 4 experience in this field. The highest level of experience was only associated with 1 of the evaluated individuals. We had therefore a set of users with balance between experts and non-experts in research data management.

The answers to **QI3** reveal that storing research data in personal computers or in external hard drives is still very common, a fact that reveals the need for RDM tools.

Finally, **QI4** shows that most of the participants use social networks very frequently.

3.1 The Tasks for the Usability Tests

The tasks designed for the experiments are abbreviated in Table 2. To simulate a creator and a collaborator interacting on a Dendro project, each task requires two users (**A** and **B**), on two separate computers.

Table 2. Description of the evaluated tasks

ID	Description
TA1	Find the "Social Dendro" section
TA2A	As user A, create a project, add user B as collaborator, add a folder and upload a file
TA2B	As user B, wait to be added to the project, then create a folder, and upload a file
TA3	Add two metadata descriptors to the created folder
TA4	Check the posts generated from the changes
TA5	Like a post
TA6	Comment a post
TA7	Share a post
TA8	Identify the notifications section

Although **TA1** seemed a very simple task, it had an average time for completion of more than 2 min. Most of the users commented that the Social Dendro timeline section was not easily accessible.

Tasks **TA2A** and **TA2B** had the highest average completion time, with values above 5 min. One of the required steps for user A was to add user B as a collaborator. As the input for this step required writing the full user URI of the collaborator, it was observed that it took some users several tries to succeed.

The average completion time for **TA4** was above one minute, perhaps because the default Dendro interface includes an area that shows the recent changes on the metadata for the selected file or folder. The Social Dendro timeline, on the other hand, shows the changes made to resources in all the projects where the user participates and this similarity was confusing. As for the remaining tasks, the average completion times were quite low as expected.

3.2 Post-experiment Questionnaire

To collect feedback and assess user satisfaction with Social Dendro, users were required to answer a post-experiment questionnaire. In some questions they were requested feedback on each of the social components—like, comment, share, as well as on the notifications. In others, they where asked to rate each social component on various properties relevant to the research activity, in a scale of 1 to 5, low to high: the **utility** in the context of RDM; the **visibility of research**; the **quality of metadata**; and the reduction of the **time for adaptation** to Dendro.

Users considered that the like component was the least useful feature, giving it the lowest average degree of utility (3.5) in the context of RDM. As it was also the social component with the highest standard deviation value (1.41), we may conjecture that the usefulness of this component for RDM is less obvious than the others. All base data of this study will be available at our institutional repository.

4 Conclusions and Future Work

Social Dendro adds social features to Dendro, allowing users in a research project to describe data and like, share and comment on each other's descriptors. Its user interface needs improvement to be easily used and understood, as seen by the time some users took to complete some of the tasks in the user experiments. Specifically, the like component was shown to lack usefulness in contrast to the share and comment components.

In research contexts, the ability to give both positive and negative ratings is essential, since criticism and peer review are at the basis of the scientific process. At this stage, we are planning to introduce an upvote/downvote system for descriptors, files, and folders.

The script given to the users about the tasks proved to be confusing at some points, making the process more complex than anticipated. The evaluation was

focused mainly on usability and could not cover the impact of Social Dendro on RDM workflows. Future experiments will address this issue, by surveying researchers as they interact with their own data in real scenarios, as well as introducing a larger set of participants.

Acknowledgements. This work is financed by the ERDF- European Regional Development Fund through the Operational Programme for Competitiveness and Internationalization - COMPETE 2020 Programme and by National Funds through the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia within project POCI-01-0145-FEDER-016736.

References

1. Assante, M., et al.: Science 2.0 repositories: time for a change in scholarly communication. In: D-Lib Magazine 21.1/2 (2015). doi:10.1045/january2015-assante
2. Borgman, C.L.: The conundrum of sharing research data. J. Am. Soc. Inf. Sci. Technol. **63**(6), 1059–1078 (2012). doi:10.1002/asi.22634
3. Dillo, I., Doorn, P.K.: The Dutch data landscape in 32 interviews and a survey. Reporting year: 2011. Data Archiving and Networked Services (DANS) (2011). ISBN: 978-94-90531-10-2
4. European Commission. Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020 (2013)
5. National Science Foundation. Grants.Gov Application Guide A Guide for Preparation and Submission of NSF Applications via Grants.gov (2011)
6. Rocha, J., Ribeiro, C., Lopes, J.C.: Ontology-based multi-domain metadata for research data management using triple stores. In: Proceedings of the 18th International Database Engineering & Applications Symposium (2014). doi:10.1145/2628194.2628234
7. Rocha, J., et al.: The Dendro research data management platform: applying ontologies to long-term preservation in a collaborative environment. In: iPres Conference Proceedings (2014)
8. Yang, S.J.H., Chen, I.Y.L.: A social network-based system for supporting interactive collaboration in knowledge sharing over peer-to-peer network. Int. J. Hum. Comput. Stud. **66**(1), 3650 (2008). doi:10.1016/j.ijhcs.2007.08.005