Research Objects for audio processing: Capturing semantics for reproducibility

Research Objects for Audio Processing: Capturing Semantics for Reproducibility

Kevin R. Page[1], Raúl Palma[2], Piotr Hołubowicz[2], Graham Klyne[1], Stian Soiland-Reyes[3], Daniel Garijo[4], Khalid Belhajjame[3], and Rudolf Mayer[5]

[1] *University of Oxford, Oxford, UK*

[2] *Poznan Supercomputing and Networking Center, Poznań, Poland*

[3] *University of Manchester, Manchester, UK*

[4] *Universidad Politécnica de Madrid, Madrid, Spain*

[5] *Secure Business Austria, Vienna, Austria*

Correspondence should be addressed to Kevin Page, Raúl Palma (`kevin.page@oerc.ox.ac.uk; rpalma@man.poznan.pl`)

ABSTRACT

Earlier work has identified the potential for reuse and reproducibility when applying workflow systems to audio analysis and Music Information Retrieval. In this paper we extend this approach with the introduction of Research Objects to capture semantic information about the use of workflows within the audio research and development process. Once aggregated, the metadata encapsulated in a Research Object can be used to manage and disseminate research output, providing a well structured foundation for meeting the needs of reproducibility. We report on the development and deployment of a software suite that practically applies this notion of Research Objects to capture the semantics surrounding the use of an audio processing workflow, and reflect upon how this might be further integrated with lower level semantics from the audio processing domain.

1. MOTIVATION AND SCOPE

In previous work [9] we analysed numerous software tools from the fields of Music Information Retrieval and audio processing, assessing their strengths and limitations in terms of reusable code and reproducible outputs. We proposed "*improving interoperability* [within MIR] *where the benefits are clear and well scoped, and can be transferred between these different layers of reuse*" and suggested Research Objects (ROs) as a potential technology towards achieving this.

In this paper we report on the adoption, adaptation, and implementation of an RO enabled architecture for an audio processing task. We focus on how the tools aid and enable the capture of research context semantics and support best practices in reuse and repeatability. While some of the tools have been used and developed previously in different domains (e.g. astronomy and bioinformatics), this is the first time such technology has been brought to use in an audio processing context. We demonstrate how a number of different tools, through the common Research Object model and based upon Semantic Web technology, have been brought to bear on this domain, and describe where additional features were developed. At this stage of our work we focus on the semantics of the audio analysis *process* (not the semantics of the audio content itself) and how, through ROs, this can be integrated with the repeatability provided by workflow environments and the preservation capabilities of Digital Libraries.

2. WORKFLOWS

Scientific workflows have become popular among scientists in the last decade as a means for specifying and automating computational experiments [4]. These workflows have dramatically revolutionized the way many scientists conduct their daily experiments as suggested by the increasing number of scientific disciplines that have embraced workflow technology including bioinformatics, biomedical informatics, cheminformatics, geoinformatics, and astronomy. Workflows serve a dual function: first, as detailed documentation of the scientific method used for an experiment (i. e. the input sources and processing steps taken for the computation of a certain data item), and second, as re-usable, executable artefacts for data-intensive analysis. Using a workflow, a scientific experiment is defined as a series of analysis steps which specify the flow of data between them.

In this paper we make use of workflows for a third purpose, contingent upon their utility as described above and when embedded in the research process: as a means to capture a *semantic* encoding of *research process and context*. Workflows are not the only means to garner such information, but their native digital form and structured approach to analysis means that workflow environments are particularly suitable for automatically gathering the details of the research method applied, and providing a skeleton with which wider aspects of the research context can be associated.

In this work we focus on an exemplar workflow for audio genre analysis, originally developed for [8], and shown in the Taverna Workbench [7] in Figure 1. This workflow illustrates both how an audio analysis process can be encoded within a workflow, and several common workflow features:

- Several workflow input ports: an MP3 URL, a Web Service authentication token, and a ground truth file URL;
- Several workflow output ports: the detailed genre classification results (as a file), and another containing classification accuracy;
- a call to a remote web service ("extractRHSOMLib"), passing in links to the MP3 data and an authentication token;
- a call to a local service ("doClassify") which performs the genre classification using the extracted features;
- numerous support stages to process and format input, output, and intermediate stages (e.g. to extract MP3 URLs from a list, or merge and convert feature vectors).

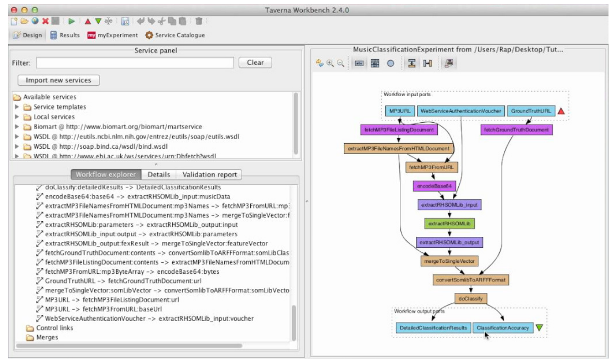

Figure 1: The genre classifier workflow in Taverna Workbench

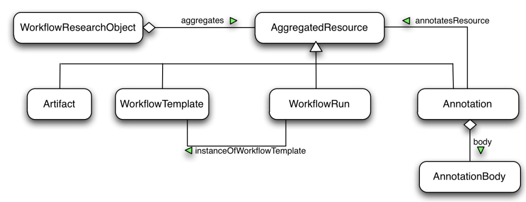

Figure 2: An overview of the Research Object model.

3. RESEARCH OBJECTS (ROS)

Research Objects [1] (ROs) aim at providing support for the description of scientific investigations in a machine readable format. In addition to the scholarly article that reports on the results of the research investigation, a Research Object encapsulates other resources that enable and promote the reuse, interpretation and reproducibility of such investigation results.

A special type of Research Objects are Workflow-Centric Research Objects, targeted towards investigations carried out using scientific workflows. Figure 2 illustrates a coarse-grained view of a workflow-centric research object, which aggregates a number of resources, namely: a workflow template, which defines the workflow; workflow runs obtained by enacting the workflow template other artefacts which can be of different kinds, e.g., a paper that describes the research, datasets used in the experiments, etc.; annotations describing the aforementioned elements and their relationships.

Workflow-centric ROs are encoded using RDF, according to a set of ontologies published by the Wf4Ever project. As with myExperiment packs, Research Objects use the Object Exchange and Reuse (ORE) model[1] to represent aggregation. ORE defines standards for the description and exchange of aggregations of Web resources. Using ORE, a workflow-centric RO is defined as a resource that aggregates other resources, i.e., workflow(s), provenance, other objects and annotations. The RO model also uses the Annotation Ontology[2] to specify annotations, and uses and adapts the PROV-O ontology[3] for specifying workflows, their workflow runs and the evolution of Research Object. As a result, we have developed a family of ontologies for specifying Research Objects, including the RO ontology[4] for specifying Research Object Aggregations, Wfdesc[5] for specifying workflow descriptions, Wfprov[6] for specifying workflow runs, and ROEvo[7] for specifying evolution of research objects.

4. RO-ENABLED SYSTEM COMPONENTS

The Wf4Ever project has developed an RO-centric architecture of services and tools [10]; in this section we introduce the subset of these components which have been deployed in a system[8] that augments the use of audio analysis workflows (section 5).

The strength of this architecture is the use of the RO model for interoperability – an RO information "bus". Some services create and update the semantic structure of the RO itself, other tools and services provide domain and workflow information to annotate the RO and the content within it, while further services automatically analyse ROs to aid the process of preservation. This embodies an approach of "just enough" semantics for interchange between components, using the RO model in its RDF encoding, whilst not overly constraining the RO model. Unlike purely syntactic interoperability, the RO model can accommodate domain specialised semantic extensions as required, without forfeiting schema compatibility between the core RO tools.

The tools and services deployed are:

Taverna workbench a workflow development and execution environment; enhanced here to export workflow runs as provenance data in RO bundles.

[1] http://www.openarchives.org/ore/1.0/
[2] http://www.openannotation.org/
[3] http://www.w3.org/TR/prov-o/
[4] http://purl.org/wf4ever/ro
[5] http://purl.org/wf4ever/wfdesc
[6] http://purl.org/wf4ever/wfprov
[7] http://purl.org/wf4ever/roevo
[8] Source code available at: https://github.com/wf4ever

RODL the Research Object Digital Library, is built upon the dLibra Digital Library and preservation platform and the dArceo long-term source data preservation service. It has been specifically designed to preserve data organised using the RO model, and provides a number of APIs (including the RO API, RO Evolution API, and Notification API) that enable other RO enabled clients and services to retrieve, create, and manipulate the ROs it stores. It also provides query access to RO RDF through a SPARQL endpoint.

Wf-RO is the Workflow to RO transformation service which extracts structured metadata from Taverna t2flow workflow description files and encapsulates it in a workflow bundle (a class of RO).

Quality Service is a component that, when provided with a Minimum Information Model [6], can periodically check that an RO meets the requirements laid down in the MINIM.

RO Portal is a web-based user interface acting as integrated front-end and client to the services provided by RODL, the Quality Service, and Wf-RO.

5. USING ROS FOR AUDIO PROCESSING

In the previous section we introduced the software components that can assist a user when constructing and manipulating Research Objects. In this section we describe in greater detail how these tools can be applied by a user developing an audio processing algorithm within a workflow and RO enabled environment.

5.1. Outline of use case

We begin by outlining a typical series of steps[9] an audio researcher or developer might undertake in the course of an investigation. These illustrate the benefits derived from using the RO-enabled infrastructure described in the previous section, but do not indicate an *exclusive* path through which a user must apply the tools. In general use the software can be combined in many different orders, and at almost any point during the investigation.

[9] A video capture of the software described being used on screen while following this scenario can be accessed at: http://purl.org/wf4ever/music-ro-screencast

1. As a first step, the audio researcher creates a new RO to organise her work on a new genre classifier she is developing. This enables her to track, share, publish, and preserve the output and outcomes from her work at all the stages of her investigation.

2. She begins by aggregating the genre classification workflow (figure 1), along with the local software library dependencies required to execute it (in this case SOM Toolbox and Weka JARs) and basic documentation such that a colleague could download the RO in the future and use the workflow.

3. At this stage the researcher can create a snapshot of the RO that can be shared with colleagues for feedback – perhaps this if the first time she has used this feature extractor service and classifier in combination and wishes to solicit feedback from more experienced lab members. While the snapshot identifier is shared to avoid a 'moving target' for feedback, the researcher continues to work with the Live RO.

4. The Quality Service automatically inspects ROs against a checklist. While the Live RO meets the requirements of a baseline workflow-centric RO (that is, an RO and basic documentation) the researcher knows that to preserve her work to institutional standards she needs to meet the requirements of a "Workflow-centric RO with provenance requirements" and selects the appropriate MINIM so that the Checklist Service can guide her through the necessary steps.

5. The researcher experiments with different parameters for the workflow in Taverna workbench, running the workflow there to generate results. When a workflow has been executed generating particularly significant output, the researcher elects to export this data – the provenance of the workflow run output – from Taverna, and to add it into the RO, where it is automatically linked to workflow references within the RO.

6. Having preserved these successful results, the researcher adds some explanatory and concluding documentation, and a draft outline of a paper writing up the investigation. The RO now meets the checklist items and is verified by the Quality Service. She generates another RO snapshot which she distributes to co-authors.

7. After several iterations of the paper, the researcher uploads the camera-ready copy as has been published. At this point she creates and archive release of the RO which, containing the final version of the paper and results, is identified for long-term preservation.

8. If the researcher, or another member of her lab, wishes to revisit the investigation and workflow, they can use the RO history visualisation to re-trace the steps that led to the publication as captured by the snapshot ROs.

The steps above outline the series of possible stages a researcher might undertake as part of her research, and how the RO-enabled tooling can help. In the rest of this section we go into further detail as to how the process above is achieved within the software through exchange of ROs between components.

5.2. Core RO Management

The RO Portal provides a web-based interface for core RO management functions, which are then enacted by the RODL service, and a front-end to several supplementary services (described below).

When first viewing the RO Portal a researcher can view lists and collections of ROs stored in the RODL. The user can log in to the RO Portal to view a customised or extended set of ROs she can access, for instance only those available to members of her research group. By representing ROs as resources according to Web Architecture design principles, authentication can be implemented using common web-based mechanisms – OpenID and Google in the case of the RO Portal.

As more of the local lab processes are RO-enabled, we expect a greater number of ROs to be reused and branched as the starting point for new investigations (step 8). In our particular scenario, however, the researcher creates a new RO (step 1). The user can give a title hint ("musicStudy") which, where possible, will generate a more human-memorable URI, and a new resource will be created in RODL.

Where a community of users (e.g. within the same lab, or working in the same field) find they use similar organisation for their folder and data – here for grouping datasets, workflows, configuration scripts, and bibliography – an *RO Template* is used as a quick and labour-saving mechanism to initiate the RO structure. The template is not

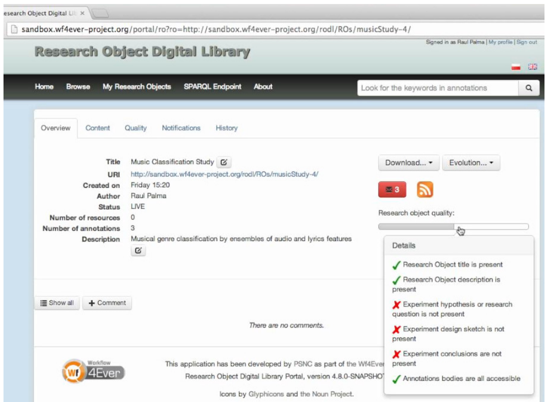

Figure 3: Overview of the music investigation in the RO Portal, also showing the Quality Service evaluation.

a mandatory schema required by the tooling, but a user-driven time saving feature – since ROs are completely semantically interchangeable through the model at the system level, any number of templates can be constructed and applied without impacting interoperability between software components. In our example, the researcher uses the template for a *Workflow-centric RO* (step 2).

On creation, the RO is automatically populated with a baseline amount of metadata that annotates the inherent RO structure and ownership from the user. This metadata is updated as changes to the RO occur, and can be seen RO Portal overview tab (figure 3). At any time this metadata can be downloaded from the RO Portal (know as the "manifest") or as a portable zip file containing both the RO metadata and content[10].

The user can then add any variety of digital resources to their RO. These can be references to remote resources (by URI), or as files that are uploaded into the RO via the Portal. These aggregated ROs are then imported and preserved in the RODL. Aggregated resources can be assigned a type, which enables automated services to process them and further enrich the RO (section 5.3). In our scenario the researcher initially uploads a Taverna workflow (the t2flow file described in section 2), software library files that are required to run the workflow, a diagram of the workflow design (typed as a *Sketch*), and text file documentation on how to run the workflow and the research question it is designed to address[11] (typed as a *Research Question*).

As research progresses (steps 6 & 7), the researcher can modify, delete, or add additional resources through the Portal interface. Our researcher adds workflow run output (section 5.4), conclusions, and the academic paper that results from the work (through drafts to the final camera ready version). To disseminate and preserve the RO at different stages of composition, and of her work, our researcher uses the lifecycle management features described below (section 5.5).

5.3. Automated assistance

Beyond the core RO management functionality described in the previous section, the Portal calls upon the RO supplemental services to automatically enrich the RO as the researchers investigation progresses.

When the user marks the uploaded genre classifier t2flow as a workflow (step 2), the Portal invokes the Wf-RO service which inspects the content of the workflow and from it creates a *workflow bundle*. This is itself a specialised subtype of RO, in which the individual parts of the workflow have been identified and annotated as such, and once recombined within the RO allows the researcher to attach annotations directly to elements within the workflow (e.g. the "mergeToSingleVector" operation in our example, or the input and output ports).

The *Quality Service* is used by the Portal to guide the user towards ensuring the RO is suitable for e.g. long term preservation and, using the metadata extracted by the Wf-RO service, to monitor that services used by the workflow (e.g. the feature extraction service) are functioning and available – and to alert the user to problems if they are not. When our researcher first creates her workflow-centric RO (step 2) it inherits the MINIM for "RO basic requirements", which checks for the presence of five pieces of RO metadata [12] and the accessibility of elements within the RO for further annotation. This checklist is considered the minimum for a usefully shareable and discoverable RO, and is analogous to a well maintained lab book. Selecting the "Workflow-centric RO with provenance requirements" MINIM (step 4) automatically performs more stringent tests associated with the long term preservation and dissemination of significant results, with the checklist additionally requiring

[10] This can be loaded into offline RO tools, such as RO Manager [12]

[11] "Using the features of type SSD provided by the feature extractor service, in combination with the SVM classifier, should yield classification results higher higher than the baseline with the MARSYAS feature set."

[12] title, description, hypothesis or research question, design sketch, and conclusions.

an annotated workflow and example and result data from workflow runs – although much of the metadata required to meet this MINIM can be automatically created when using an RO-enabled workflow system such as Taverna and the Wf-RO service. When viewing the RO in the Portal the "quality" (i.e. it's conformance with the selected MINIM) of the RO is shown (figure 3), and the user can view details of each item in the checklist in the *Quality* tab.

In the *Notifications* tab our user can view the history of changes to the RO, including the addition of aggregated resources; the underlying Notifications API in RODL enables a user to receive messages describing changes to ROs they wish to "watch".

5.4. Workflow runs and provenance

Our researcher downloads the workflow as a t2flow file from her RO and opens it in Taverna Workbench. In Taverna she experiments by passing different inputs to the workflow – alternative collections of MP3 audio data and ground truth sets – and classifier parameters, each time executing the workflow to and viewing the output results (step 5).

Taverna has been extended to export the provenance of an individual workflow run, encoded using the W3C PROV ontology. So our researcher can now export this specific setup and results of a workflow run from Taverna as an *RO bundle* (another specialisation of ROs, serialised within a zip file) and import the RO bundle, through the Portal, into her RO.

On import, the Portal recognises the bundle as an RO and creates a URI for it in its own right – it can now be identified and inspected as per any other RO. But it is also recognised as the output from a workflow run, and the content within the new RO (bundle) is annotated using the provenance export, marking out metadata that captures the input variables, algorithmic output, etc. – this enables automated inspection by supplemental services such as the Quality Service to check to the health of the workflow (e.g. availability of web services and input data). Furthermore the RO containing the workflow run is automatically linked back to the RO the user created to describe the overall investigation, and can be typed as an exemplar input and output or a set of experimental results. On inspection, this allow the Quality Service to verify the RO now meets the requirements of a "Workflow-centric RO with provenance requirements" as a necessary step towards its long-term preservation.

5.5. RO Lifecycle management

When the researcher first creates her RO (step 1), the system considers it a *Live RO*. This is the default state of any RO accessed through the portal.

At various stages in the investigation it is useful for the researcher to share a known, fixed, version of the RO as it was at a specific point in time. The Lifecycle features of the Portal enable her to do so, in turn backed by the RO Lifecycle implementation in RODL (and the RO Lifecycle model).

The most frequent lifecycle activity is the creation of a *Snapshot* (steps 3 and 6) using the *Evolution* button in the Portal *Overview* tab. Each snapshot is given its own URI by RODL so that it, and its content and metadata, can be accessed and viewed unambiguously, both through the RO Portal and via any other RO-enabled tool. A *Release* version (step 7) is created by the user in a similar way to a snapshot, but indicates to the underlying Digital Library infrastructure that this RO will need to be preserved in the longer term – which may invoke a checklist to assess the completeness of the RO to fulfil this preservation tasks (e.g. using a "Workflow-centric RO with provenance requirements").

Crucially, relationships between live, snapshot and archive ROs are created and maintained, and can be viewed by the user in the Portal *History* tab (step 8 and figure 4). This allows our researcher to visualise the relationships between the different stages of her work, as encapsulated by the ROs, perhaps to find an earlier revision from the investigation to start of a new line of enquiry, or to build upon the published work of another researcher – transforming a release or snapshot into a new live RO.

6. RELATED WORK

In this work we have reported on the use of Taverna as a workflow environment, using it with a genre classifier workflow and an online feature extraction service. Within the audio processing domain there are several alternative workflow systems including M2K [5], NEMA [11], and VAMP/Sonic Annotator [2], and numerous frameworks with workflow-like properties – the strengths and limitations of which are described in [9]. As commented in Section 2, workflows are not a precondition for capturing the semantics required for reproducibility and preservation, although they provide a well defined and described digital environment from which to automatically extract this information – and it is in this

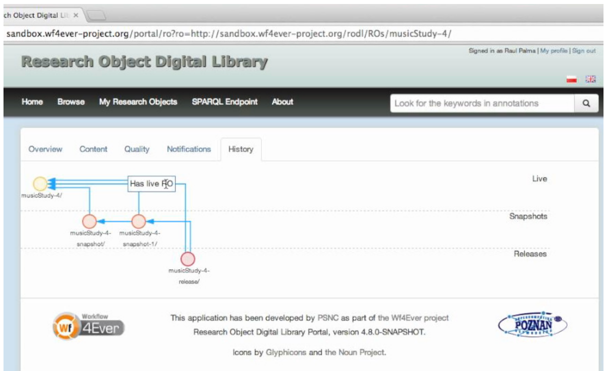

Figure 4: The History tab in the Portal illustrating the evolution relationships between Live ROs, Snapshots, and Releases.

context we have applied and extended Taverna. As such, there is no inherent impediment to RO-enabling any of the above alternative workflow systems, however Sonic Annotator must be considered particularly relevant since it was specifically designed for work in the audio domain, and encodes both workflow and audio processing semantics.

ROs have been conceived as a relatively lightweight "bottom up" approach to capturing and aggregating semantics from multiple sources such that the audio analysis process can be sufficiently well described for the purposes of preservation. *Mayer and Rauber* [8] describe a more formal process for defining digital preservation requirements called *Context Models*, and build a use case and context model using the genre classifier Taverna workflow we have reused in the work presented here (section 2). We envisage these as complementary approaches, with RO-enabled tooling providing a practical means to gather and collate metadata alongside simple checklist Quality Service evaluations, while Context Models can be used to evaluate the metadata to more detailed or rigorous standards (which may be required for e.g. patent applications). The possibility of combining the two models is particularly appealing and practical due to the underlying use of RDF in both.

Finally, we draw attention to other RO-enabled tools which were not included in our scenario due for reasons of brevity, such as RO Manager [12] and myExperiment [3], both of which can be used to view and modify the ROs described herein due to the flexible semantic interoperability provided by ROs.

7. CONCLUSIONS AND FUTURE WORK

The integration of the genre classification workflow into our RO-enabled toolset was in itself a lesson in the perils of reproducibility: in the state it was received the workflow no longer functioned. The audio files were unavailable, the API for the feature extractor had changed, and the access token for the web service had expired. While there were basic installation instructions for the workflow, we had no record of when the workflow had last worked, who had run it, what parameters were used (and why), or what had changed in versions since or before then – and then the original author was found to be absent on leave just at the moment we were working on the software! In this paper we have shown how the use of Research Objects can overcome these problems in the context of developing audio processing algorithms, and described a suite of RO-enabled tools that embody this approach through an initial software implementation.

The semantic interoperability offered by the RO model enables us to flexibly combine user facing applications such as Taverna and the RO Portal with preservation and support services such as RODL, Wf-RO, and the Quality Service. This enables each tool or service to contribute to a common RO whole while retaining specialisation in purpose and interface.

Although ROs provide a firm and compelling foundation, future work is needed to further automate the processes of reproducibility and preservation. Integration with formalised requirements (e.g. Context Models as mentioned above) could enable the "certification" of ROs in situations where a high bar is set for quality – e.g. in national repositories, or in businesses (cf. patents). More investigation is also required into tooling that can take the semantics already captured within an RO and *ensure* its reproducibility – perhaps through use of a virtual machine image, or archiving or all software and data used in a workflow run (although this is clearly not a trivial task). Where version control is used for software code, a web-based interface (i.e. provision of URIs) might enable integration with RO tooling.

A principle benefit of the RO model and approach is that it provides enough structure to formalise the necessities for re-use without being overly prescriptive or exacting as to fully describing every new and unique scenario (though, of course, these can be prescribed and semantically encoded if necessary). Perhaps this is a lesson in practical reproducibility – while there is a baseline of information required for any reproducibly, too much specificity can preclude the benefit, demonstrated here, that comes from the re-use of common tools. While our work to integrate an audio processing workflow into our toolset required the development of additional features, in hindsight may of these have been found to be more generally applicable.

There is, of course, a balance between the utility of generic and domain specific tooling. A more specialised client or portal using ROs for audio processing might integrate audio playback functionality; an RO template could aid in the management of audio files. Given the flexible structure of ROs it is technically possible to aggregate many more elements that can describe the mechanics of audio analysis in much greater detail (e.g. using the Audio Features ontology), and there is obvious utility in defining MINIMs such that a checklist can operate over these semantics – the question then becomes of how and where we can capture the metadata to enable audio processing functionality. Building on the principles of automated assistance we have demonstrated in the paper, and wishing to avoid the extra user burden which makes metadata capture less likely, we believe the best way to achieve this is wherever possible through the RO-enabling of tools that are already beneficial and integrated into a users work. For example, an RO-enabled Sonic Annotator could collect far more semantic detail specific to the audio domain than the more generic workflow environment offered by Taverna.

In this paper we have demonstrated the utility of ROs when applied to the audio processing domain. We believe that, when connected by a common RO infrastructure, the sum of the tools usefulness is greater than their individual parts. We will continue to "RO-enable" software in the audio domain according to this observation, but also hope this demonstration encourages others to do the same.

ACKNOWLEDGEMENTS

This research is supported by the EU Wf4Ever project (FP7-270129). We are especially grateful to our collaborators in the TIMBUS project for presenting the opportunity to develop this scenario, and for their help when using the genre classifier workflow.

8. REFERENCES

[1] K Belhajjame, O Corcho, et al. Workflow-centric research objects: First class citizens in scholarly discourse. In *Proc. Future of Scholarly Communication in the Semantic Web*, 2012.

[2] C. Cannam, M. Sandler, et al. Linked data and you: Bringing music research software into the semantic web. *Journal of New Music Research*, 39(4):313–325, 2010.

[3] D. De Roure et al. The design and realisation of the virtual research environment for social sharing of workflows. *FGCS*, 25(5):561–567, 2009.

[4] E Deelman, D Gannon, et al. Workflows and e-science: An overview of workflow system features and capabilities. *FGCS*, 25(5):528–540, 2009.

[5] J.S. Downie, A.F. Ehmann, and X. Hu. Music-to-Knowledge (M2K): a prototyping and evaluation environment for music digital library research. In *Proc. 5th ACM/IEEE Joint Conference on Digital Libraries*, pages 376–376, 2005.

[6] M Gamble, C Goble, et al. Mim: A minimum information model vocabulary and framework for scientific linked data. In *Proc. 8th Intl. Conference on e-Science*, pages 1–8. IEEE, 2012.

[7] D Hull, K Wolstencroft, et al. Taverna: a tool for building and running workflows of services. *Nucleic acids research*, 34(suppl 2):729–732, 2006.

[8] R. Mayer and A. Rauber. Towards time-resilient MIR processes. In *Proc. Intl. Conference on Music Information Retrieval*, pages 337–342, 2012.

[9] K Page, B Fields, et al. Capturing the workflows of music information retrieval for repeatability and reuse. *Journal of Intelligent Information Systems*, pages 1–25, 2013.

[10] K. Page, R. Palma, P. Houbowicz, et al. From workflows to research objects: an architecture for preserving the semantics of science. In *Proc. of the 2nd Intl. Workshop on Linked Science*, 2012.

[11] K. West, A. Kumar, et al. The Networked Environment for Music Analysis (NEMA). In *Proc. 6th IEEE World Congress on Services*, pages 314–317, 2010.

[12] J. Zhao, G. Kylne, et al. RO-Manager: A Tool for Creating and Manipulating Research Objects to Support Reproducibility and Reuse in Sciences. In *Proc. 2nd Intl. Workshop on Linked Science*, 2012.