Towards Usable and Interoperable Workflow Provenance: Empirical Case Studies Using PML

James R. Michaelis*, Li Ding*, Zhenning Shangguan*, Stephan Zednik*, Rui Huang*,
Paulo Pinheiro da Silva†, Nicholas Del Rio† and Deborah L. McGuinness*
*Tetherless World Constellation, Rensselaer Polytechnic Institute, Troy, NY 12180
†Computer Science Department, University of Texas, El Paso, El Paso, TX, 79968

***Abstract*—In this paper, we describe how a semantic web-based provenance Interlingua called the Proof Markup Language (PML) has been used to encode workflow provenance in a variety of diverse application areas. We highlight some usability and interoperability challenges that arose in the application areas and show how PML was used in the solutions.**

I. Introduction

In scientific research, workflow systems are used to assemble steps (each corresponding to certain tasks) for processing scientific data. Provenance is a well-known and important component in these systems [1]. In particular, access to a workflow system's *data flow* has proven crucial for users to understand, validate, and reproduce its workflows [2], [3].

As workflow systems become more complex and distributed in nature, a number of provenance management challenges are known to emerge [1]. Within the scope of this paper, we focus on two particular challenges: *usability* and *interoperability*. To address the usability challenge, provenance information must be both sufficiently intuitive and expressive for end users to understand. Likewise, for the interoperability challenge, provenance representations must be capable of linking to, integrating, and reusing each other's content for unexpected purposes.

In this paper, we investigate how both challenges can be addressed through a domain independent provenance interlingua called the Proof Markup Language (PML) [4]. PML facilitates generation and sharing of provenance metadata for data derivation within and across intelligent systems, and acts as an enabler of trust by supporting explanations of information sources, assumptions, and learned information. As a critical part of the Inference Web (IW) [5] project, PML has been used in many domains [6], including: information extraction [7], logical reasoning [8], workflow processing [9], semantic eScience [10], and machine learning [11], [12]. Three workflow-based case studies we explore are as follows:

- **Case Study 1, Semantic Provenance Capture in Data Ingest Systems (SPCDIS)**: This project integrates provenance representations into scientific workflows in the fields of solar, solar-terrestrial, and space physics. These workflows include numerous scientific data products annotated by complex domain-specific ontologies. Here, provenance is needed to facilitate querying based on domain-knowledge (for instance, to list which scientific instruments were used to derive a certain type of data product).
- **Case Study 2, Generalized Integrated Learning Architecture (GILA)**: GILA is a multi-agent machine learning platform, which generates a workflow log about how a problem was resolved collaboratively by an ensemble of learning agents. Provenance in this system is implicitly encoded through domain-specific structuring, and needs to be normalized to allow basic querying.
- **Case Study 3, The Third Provenance Challenge (PC3)**: Unlike the former two case studies, this focuses on a workshop aimed at developing interoperable provenance. Here, multiple participants investigated a workflow from an astrometry/photometry-based system. Using individual approaches, everyone had to monitor this workflows execution and export the resulting provenance data for import, integration and querying by the other teams.

The remaining sections are organized as follows. Section 2 briefly reviews PML and shows its applicability in workflow provenance representations. Sections 3 through 5 detail the three case studies on SPCDIS, GILA and PC3 respectively. For each of these, we highlight: (i) examples of usability and interoperability challenges, (ii) how PML was used to address these challenges, and (iii) lessons learned from these efforts. Section 6 discusses related work with PML, and section 7 provides concluding remarks.

II. PML and Workflow Provenance

PML is a Semantic Web based provenance representation, defined through three core OWL ontology modules: the **Provenance module**(namespace: pmlp), which supports annotation of general provenance related entities, (such as agents, data products, and information sources); the **Justification module** (namespace: pmlj), which supports annotating derivation relations (pmlj:InferenceStep) among data products, represented by justification-based concepts (pmlj:NodeSet); and the **Trust module**(namespace: pmlt), which supports annotating complex trust relations on provenance and justification concepts. The modular design of PML facilitates future reuse and extension of these core ontologies.

In tracking workflow provenance, PML can be used to capture data flow by recording: (i) the sequence of operations taken to derive data products, and (ii) descriptions about these operations. Figure 1 depicts a simple workflow covering basic workflow concepts (above) and shows how these concepts are represented by PML (below). The workflow includes a sequence of processes $P_0 \dots P_n$. Each process P_i(denoted by a rectangle and mapped to pmlj:InferenceStep), is defined as an execution of an operation O_i (denoted by a diamond and mapped to pmlj:InferenceRule) by an agent A_i (denoted by a person figure and mapped to pmlp:Person), and takes as input a data product D_i (denoted by an oval, and mapped to pmlp:Information) and derives another data product D_{i+1} as the output.

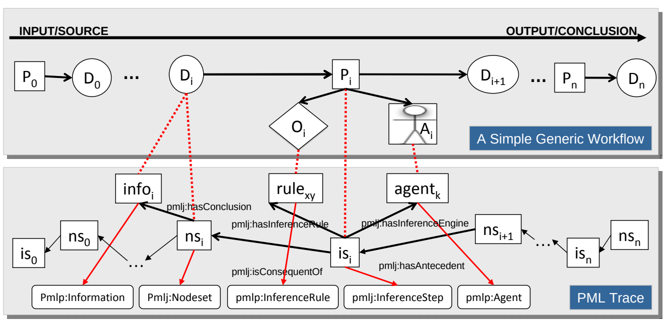

Fig. 1. Representing Execution of Workflow

There are some immediate benefits in representing workflow provenance using PML. Using Semantic Web representation strategies, PML-encoded data can be linked to domain ontologies supporting improved usability, and may be extended by (or mapped to) other provenance models for better interoperability. For this, OWL is used to facilitate linking of domain concepts to PML through constructs such as subclass relations. Likewise, PML is used for direct representation of provenance concepts (like those defined in its provenance module). This combined PML/domain data can in turn be processed by Semantic Web based tools. Examples include: OWL reasoners, SPARQL engines, and Inference Web based tools (such as Probe-It! [13] for PML visualization, and the OWL Instance Validator [14] for checking validity of PML data).

III. Case Study: SPCDIS

Semantic Provenance Capture in Data Ingest Systems (SPCDIS) [15] is a research project aimed at integrating provenance at data generation/ingest time into a data portal managed by the Mauna Loa Solar Observatory (MLSO). In SPCDIS, provenance annotations are being used to incorporate trust and transparency into generated data products. Figure 2 illustrates the Coronal Helium I Imaging Photometer (CHIP) pipeline - an example collaborative scientific workflow for generating scientific images in a distributed environment. The magnified portion of the workflow shows a fragment of the data flow: the *Instrument Capture* process uses certain configuration data under the *Instrument Configuration* category as input and generates some output image-based data products under the *Image File with Header* category.

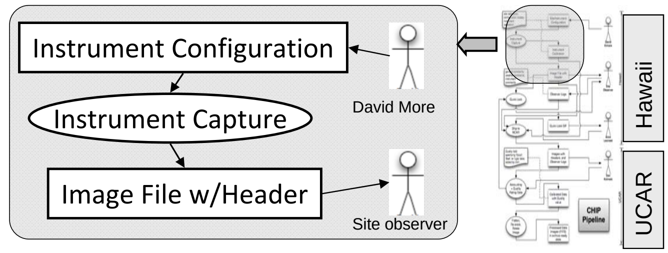

Fig. 2. A Example Fragment of a distributed Workflow from SPCDIS

A. Use of PML

Provenance encoding in SPCDIS faces both usability and interoperability challenges, given the high volumes of data processed by heterogeneous components in diverse locations. We see PML as capable of both of these challenges in the context of this system.

In our introduction, we emphasize that provenance usability depends on its intuitiveness. Here, we consider the idea of intuitiveness from both the perspective of a domain expert and computer scientist – two types of people likely to be collaborating to generate a computer based provenance representation. For instance, consider the query "Which photometers (or more generally, optical instruments) were used to generate the DataImage at a specific URL?" This is rich in domain knowledge, but may not make sense to a non-expert. Likewise, consider a modified version of the query: "Which pmlj:InferenceStep instances $X_0 \dots X_n$ were used to generate the pmlp:Information instance Y?" This would expose more of the representational details than a domain expert needs to see, but captures an abstraction usable by a computer scientist or computer program. By combining domain-dependent concepts with PML, we facilitate its use by individuals with varying degrees of expertise in a target domain.

Likewise, the interoperability challenges faced by SPCDIS stem from its recording of provenance from a series of distinct workflow components with varying terminologies. The issues underlying integrating this heterogeneous provenance are resolved through terminology linking through PML-based concepts.

To carry out the strategy above, we extended the justification and provenance modules of PML through domain-specific concepts from the Virtual Solar Terrestrial Observatory (VSTO) ontology (prefix: vsto) [1]. Figure 3 shows an example of PML provenance data generated for SPCDIS. It conveys the following information: a CSRImage with the name "MLSO CHIP CSR Image" was generated by the execution of a software agent called "CSRImageCapture" via the "CHIP-He-I Continuum Capture" operation (which is a specialized VSTO instrument operation mode) using a sensor (i.e. Photometer) called CHIP. Four different ontologies (namespaces: pmlp, pmlj, vsto, and spcdis) are integrated together in Figure 3: PML contributes domain independent concepts, VSTO contributes a domain ontology and the SPCDIS ontology carries

[1]VSTO ontology: http://vsto.org/forward.htm?forward=ontology

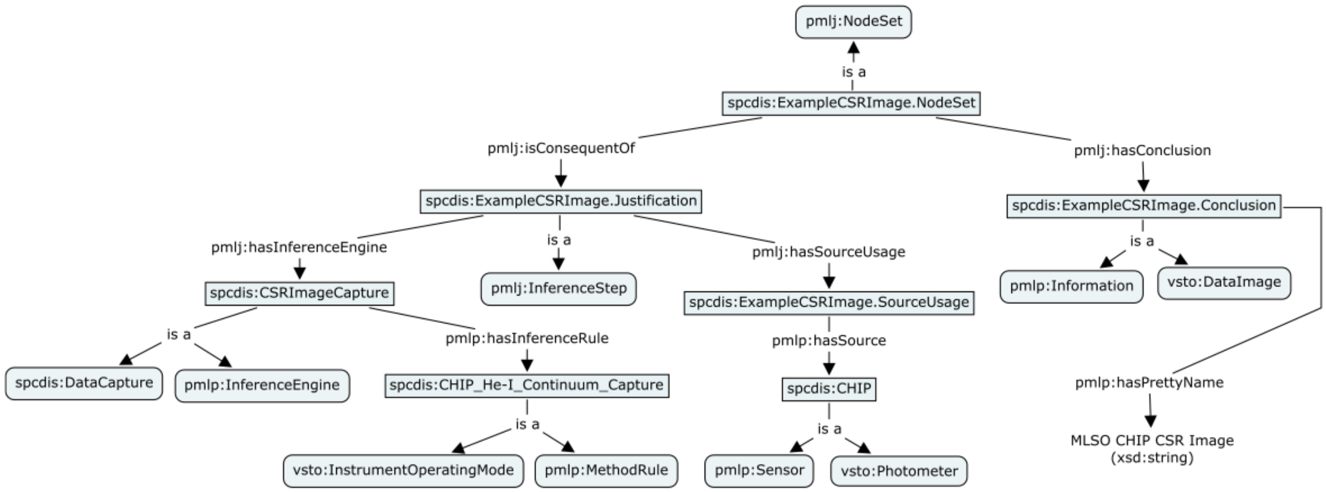

Fig. 3. Generated PML Provenance Metadata. The rounded rectangles denote concepts, and the rectangles represent instances. The edge labels denote properties, with ”a” meaning instantiation of concepts (rdf:type), and ”are” representing sub-class relations (rdfs:subClassOf).

out the integration of concepts that connect PML and VSTO. For example, the class spcdis:DataImage is a subclass of both the pmlp:Information concept in PML and the vsto:DataImage concept.

Here, PML's integration with domain-specific ontologies is necessary to answer the question from the beginning of this section (specifically, to determine which pmlp:Information instances are also of type vsto:DataImage). Below is a SPARQL query for accomplishing this, which leverages our provenance representation strategy:

```
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdf  <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX pmlj: <http://inference-web.org/2.0/pml-justification.owl#>
PREFIX pmlp: <http://inference-web.org/2.0/pml-provenance.owl#>
PREFIX vsto: <http://dataportal.ucar.edu/schemas/vsto.owl#>
PREFIX spcdis: <http://example.com/spcdis.owl#>
PREFIX image: <http://iw.vsto.org/data/mlso/chip/raw/>
SELECT ?photometer
WHERE  { ?image pmlp:hasURL "image:2008_09_03_00_04_07.csr"^^xsd:anyURI .
         ?nodeset pmlj:hasConclusion ?image ;
                  pmlj:isConsequentOf ?step .
         ?step pmlj:hasSourceUsage ?usage .
         ?usage pmlp:hasSource ?source .
         ?source a ?spcdis:Photometer . }
```

B. *Lessons learned*

In this case study, we showed how PML could be used, in conjunction with domain-specific ontologies, to address provenance usability and interoperability challenges. This effort hinged on manual ontology mapping, which did require domain expertise. However, most of the mappings were done simply by linking domain-specific concepts with generic provenance terminology from PML (e.g. pmlp:Information, pmlp:Source, pmlp:InferenceEngine and pmlp:InferenceRule). This helped establish value restrictions on provenance concepts, so we could support appropriate qualified searches (such as the one above that needs only particular kinds of images).

IV. Case Study: GILA

The Generalized Integrated Learning Architecture (GILA) [9] is a multi-agent platform for learning how to solve domain-specific problems. Initially, GILA's agents learn from domain expert generated workflow traces, each consisting of an encoded sample problem with an accompanying solution sequence. Following this, these agents collaborate to solve similar problems. During both steps, all agent knowledge and work (solutions) are recorded to a communal blackboard as a way to facilitate inter-agent communication. The GILA system was tested using the Airspace Control Order (ACO) Scheduling Scenario, which consisted of the following parts: (i) a problem state is submitted to a central scheduler (usually a domain expert) - defined as a list of requests for temporal-spatial airspace allocation, each encoded as an Airspace Control Means (ACM), (ii) the scheduler selects and updates the ACMs one at a time to resolve their temporal-spatial conflicts (generating a new problem state each time to reflect the remainder of the problem). This deconfliction process requires domain expertise in, for instance, prioritizing which ACMs should be changed initially. In this scenario, GILA was compared against novice human participants in playing the role of the scheduler.

A. *Use of PML*

GILA's logs, derived from agent submitted information on the communal blackboard, were used to evaluate its performance. These were structured as RDF graphs, with their semantics preserved by a handful of domain ontologies encoded in OWL. These domain ontologies implicitly covered both the provenance annotations for domain entities and derivation relations among data products. However, the derivation relations were represented using complex domain structuring, such that it was hard to see a clear picture of GILA's data flow. To address this, we had to overcome a usability challenge on provenance normalization - that is, to normalize derivation relations to facilitate intuitive querying. One such query, which could not easily be answered by the original log, was to list all the problem states $P_1 \ldots P_n$ generated before a given ACM deconfliction S was generated.

This challenge was approached in a two-step process [16]. Starting with a set of domain ontologies and a GILA log instance, an *analysis phase* would first be conducted. This would return the following: (i) from the domain ontologies, a list of OWL classes and properties corresponding to PML classes (e.g., pmlp:Agent, pmlp:InferenceRule), and relationships (e.g., pmlj:isConsequentOf), and (ii) from the log instance, a set of RDF based structural relations, not captured by the domain ontologies, which correspond to PML relationships. Following the analysis phase, a *mapping phase* would be conducted, in which PML-based information would be inserted into the log instance.

Figure 4 illustrates how provenance normalization could be applied for the example above. First, in the analysis phase, two domain ontologies - gilcore and gilaco - are inspected to identify the hidden provenance information from the original GILA log. The following domain knowledge is uncovered: (i) ACM deconflictions are represented as instances of the class gilcore:Solution, (ii) each solution S_i has a corresponding problem state P_i, defined as an instance of gilcore:Problem, (iii) the property gilcore:hasProblem is used to link S_i to P_i (in the figure, property names are omitted due to limited space), (iii) the property gilaco:hasSolutionListResolveConflict links P_i to a recursively declared list of instances of gilcore:SolutionListResolveConflict where each list item $list_{i,i-1}$ represents an earlier solution S_{i-1}. This list helps define the context of a current problem, but doesnt explicitly define the solution S_{i-1} used to transform $P_i - 1$ to P_i knowledge required for uncovering the solution generation data flow.

In the mapping phase, PML data is built in the following steps: (i) an OWL ontology is defined for linking the domain ontologies to PML, which asserts gilcore:Solution and gilcore:Problem as subclasses of the OWL class pmlp:Information, (ii) through OWL inference, instances of these two classes will be inferred to be of type pmlp:Information, (iii) through JENA [2] (a Java-based RDF data processing API) and SPARQL, PML data is generated from a GILA log instance which normalizes links between problems and solutions (e.g. from P_i to S_{i-1}).

B. Lessons learned

In this case study, we demonstrated how to use provenance normalization to address usability challenges by generating PML data based on both GILA's domain ontologies and logs. Although non-trivial domain expertise was needed to (i) identify the provenance components in the domain ontologies and log data, and (ii) establish mappings from the domain ontologies to PML, such work usually ended up being a one-time job. Subsequent generation of PML data in the mapping phase could then be automated using off-the-shelf tools and easily be performed.

One of our future goals will involve determining ways in which the analysis phase could be (at least partially) automated. However, as part of this, a set of constraints on how log and domain ontology information can be structured will be required.

[2]http://jena.sourceforge.net/

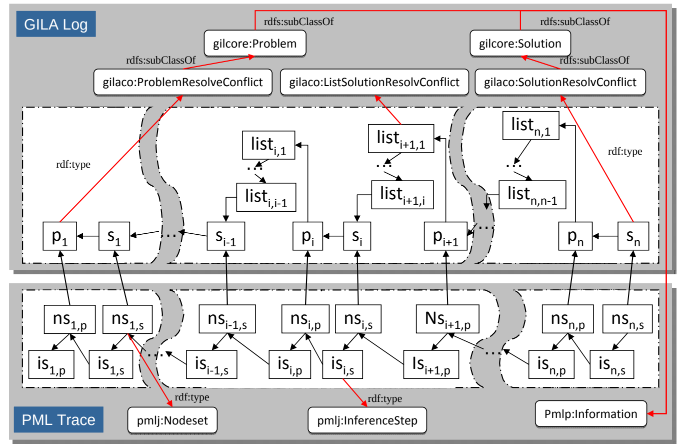

Fig. 4. PML encoding based on GILA log

In general, many complex systems like GILA can record provenance in their workflow logs, but usually use domain-specific terminology and structure. A fair amount of work must be done to individually tailor explanation interfaces for these systems. By normalizing workflow provenance into PML, we can more easily apply general-purpose explainers [13], [5] to various workflows from different domains.

V. Case Study: PC3

In the Third Provenance Challenge (PC3) [3], 15 research groups were asked to use their own approaches to: (i) generate provenance metadata for exposing the execution of a given workflow, (ii) use this metadata to answer a set of provenance-based queries, (iii) export this metadata, and (iv) import metadata from other teams and answer the queries from (ii) using it. A common interchange format, the Open Provenance Model (OPM) [17], was chosen for teams to import and export their provenance metadata. The workflow investigated in this effort was derived from the Pan-STARRS project [4], which processes data on 99% of visible stars in the northern hemisphere, and manages a pipeline for loading domain data in CSV files into a relational database and validating it. Here, control flow was viewed as the sequence of processes executed within the workflow, subject to conditional branching (e.g., if a process fails to complete correctly, halt the workflow otherwise, continue normally). Likewise, dataflow was defined as a sequence of steps by which data would be generated and used by processes.

A. Use of PML

During PC3, two important requirements emerged for us (and many other groups) in encoding provenance capable of answering the queries.

[3]http://twiki.ipaw.info/bin/view/Challenge/ThirdProvenanceChallenge
[4]http://pan-starrs.ifa.hawaii.edu/

First, domain-specific provenance (as with GILA and SPCDIS) was needed to answer many of the queries. One such query, known in PC3 as Core Query 1, reads: ”For a given *detection*, which CSV file(s) *contributed* to it?” Here, two domain-specific concepts are referenced: (i) a detection, which is a type of data handled in the workflow, and (ii) a contribution, which references a data loading sequence carried out by the workflow.

Second, many queries required the control flow of the workflow to be tracked in parallel with the dataflow. We viewed this challenge as consisting of two parts: (i) explicitly distinguishing execution of operations from the operations themselves, and (ii) representing the dependencies among executions of operations. An example of a control-flow based query, Optional Query 2, reads: ”Which pairs of procedures in the workflow could be swapped and the same result still be obtained (given the particular data input)?”

The generation of provenance meeting both these requirements that both we and other groups could answer queries over constituted an interoperability challenge in PC3. To address this, we used Semantic Web technologies to manage and query provenance - both recorded from the system workflow and imported from other teams.

Specifically, we explored storing provenance as RDF data structured around a prototype ontology containing both OPM and PML based concepts [5]. In turn, we were able to export provenance in both the OPM and PML formats for use by other groups (although during PC3, only OPM was used by other groups). In both our exported OPM and PML, we were able to handle the provenance specialization requirement mentioned above. However, some interesting issues emerged with both OPM and PML in control flow tracking.

For OPM, these seemed to emerge from ambiguities in its Process concept definition - which could either be viewed as an operation, or the execution of an operation. Such ambiguity was avoided in PML through the concepts pmlj:InferenceRule (the operation) and pmlj:InferenceStep (its execution). While the names InferenceRule and InferenceStep may be used most often in logical theorem provers, they are applied in any setting where some inference (possibly statistical or process) is used to manipulate information thus they are easily applied in a process setting.

Likewise, PML lacked a mechanism for directly tracking dependencies between operations and their executions. However, OPM did provide a way to track dependencies between instances of its Process concept - through the provided wasTriggeredBy relation. Both PML and OPM are evolving to meet community needs and we might expect co-evolution and potentially inclusion or importing of some representational features from one into the other. A comparison of the OPM and PML models can be found at [18].

B. Lessons Learned

Based on our experiences with PC3, and the other case studies, we feel that Semantic Web technologies are well suited for representing workflow provenance (in particular, for facilitating integration of provenance from heterogeneous sources). In addition, while we found some expressivity limitations in PML, these could easily be fixed by adding/referencing other ontology modules (e.g. for control flow concepts).

[5] http://www.cs.rpi.edu/˜michaj6/provenance/PC3OPM.owl

VI. Related Work

Workflow Provenance Models. There is a diverse literature on workflow provenance models [19]. Although these models differ in certain aspects, they all model some general provenance concepts, including processes, data, and process-data dependencies [1]. Many of them include domain specific concepts required by applications. For instance, Taverna [20] has included bioinformatics ontologies and the VisTrails [21] system adds *workflow description* as a kind of data in tracking user behavior in assembling workflows. PML, as a provenance interlingua, covers these general concepts. It is notable that PML, as an OWL ontology, can be connected to domain concepts (without hard-coding) via ontology mapping (declaring the rdf:type of certain domain data as a subclass of pmlp:Information).

Beyond the basic provenance concepts, some useful concepts like control flow may also be captured by workflow provenance. Furthermore, [22] identified prospective provenance (abstract workflow descriptions) and retrospective provenance (workflow execution logs) in a layered model, and both types are supported by the REDUX [23], Taverna, Pegasus [24], and Karma [25] provenance models. PML core vocabularies only cover the basic provenance concepts in the retrospective provenance because they were designed to only capture generic data derivation processes. However, PML can be extended with workflow specific modules, such as WDO (http://trust.utep.edu/wdo/) and SAWs [26] for capturing prospective provenance.

The Open Provenance Model (OPM) is another general-purposed provenance model. While OPM remains technology agnostic, PML presently provides a family of OWL ontologies with RDF syntax. This brings about a current implementation advantage of PML: it can be seamlessly integrated with domain ontologies and thus support queries involving both domain constraints and generic provenance relations.

Semantic Web Vocabulary for Provenance. There are some existing works on provenance representation in Semantic Web communities. The Dublin Core (DC) ontology (http://dublincore.org/documents/dc-rdf/) offers generic provenance related properties. The Friend of a Friend (FOAF) ontology (http://xmlns.com/foaf/spec/) offers classes and properties for annotating entities involved in provenance, such as people (foaf:person). It is also notable that there are some emerging provenance ontologies [27]. These ontologies have a good overlap with PML, especially its provenance module. However, PML differs from these works based on its justification module which offers support for tracking complex relationships between provenance-based entities.

VII. Conclusion

In this paper, we have shown the usage of PML in representing workflow provenance through three case studies. In addressing these case studies, both usability and interoperability challenges emerged in various forms, which required differing strategies to handle. With this, we give some final words on both the challenges of provenance usability and interoperability.

For the usability challenge, many workflow systems (such as SPCDIS and PC3) will rely upon domain-specific concepts that cannot be expressed using a domain independent representation alone. Likewise, others (GILA) will encode provenance data using a domain-specific log that is not intuitive for a general audience. In our case studies, PML proved effective by (i) helping users answer queries involving both domain specific and independent provenance knowledge, and (ii) helping with normalization of domain-specific provenance relationships.

Likewise, to address the interoperability challenge, PML can be (and has been) easily connected to domain ontologies and other provenance models, including OPM, via ontology mappings (as was done with SPCDIS and GILA) and ontology extensions (like with PC3). Here, It should be emphasized that the interoperability challenge requires the establishment of best practices for information exchange (as well as an effectively designed provenance representation like PML).

To establish best practices for provenance interoperability, we stress the adoption of Semantic Web languages (such as OWL) as a common data exchange medium. Through this, explicit mappings can be established between concepts (for instance, by adding "owl:sameAs" assertions). In addition, this would allow for a wider degree of terminology to be used for concept descriptions (such as the PML terms "pmlp:hasName" and "pmlp:hasFormat").

Acknowledgment

This work is partially supported by NSF award #0524481,HRD-0734825, IARPA award #FA8750-07-2-0031, DARPA award #FA8750-07-D-0185,#55-002001,#FA8650-06-C-7605.

References

[1] S. B. Davidson and J. Freire, "Provenance and scientific workflows: challenges and opportunities," in *SIGMOD Conference*, 2008.

[2] Y. Simmhan, B. Plale, and D. Gannon, "A survey of data provenance in e-science," *SIGMOD Record*, vol. 34, no. 3, pp. 31–36, 2005.

[3] S. Miles, P. T. Groth, M. Branco, and L. Moreau, "The requirements of using provenance in e-science experiments," *J. Grid Comput.*, vol. 5, no. 1, pp. 1–25, 2007.

[4] D. L. McGuinness, L. Ding, P. Pinheiro da Silva, and C. Chang, "Pml 2: A modular explanation interlingua," in *ExaCt*, 2007.

[5] D. L. McGuinness and P. Pinheiro da Silva, "Explaining answers from the semantic web: the inference web approach," *Journal of Web Semantics*, vol. 1, no. 4, pp. 397–413, 2004.

[6] P. Pinheiro da Silva, D. L. McGuinness, N. D. Rio, and L. Ding, "Inference web in action: Lightweight use of the proof markup language," in *International Semantic Web Conference*, 2008.

[7] J. W. Murdock, D. L. McGuinness, P. Pinheiro da Silva, C. A. Welty, and D. A. Ferrucci, "Explaining conclusions from diverse knowledge sources," in *International Semantic Web Conference*, 2006.

[8] P. Pinheiro da Silva, P. J. Hayes, D. L. McGuinness, and R. Fikes, "Ppdr: A proof protocol for deductive reasoning," Knowledge Systems, AI Laboratory, Stanford University, Tech. Rep. KSL-04-04, 2004.

[9] X. S. Zhang, S. Yoon, P. DiBona, D. S. Appling, L. Ding, J. R. Doppa, D. Greeny, J. K. Guo, U. Kuter, G. Levine, R. L. MacTavish, D. McFarlane, J. Michaelis, H. Mostafa, S. Ontanon, C. Parker, J. Radhakrishnan, A. Rebgunsy, B. Shrestha, Z. Song, E. B. Trewhitt, H. Zafar, C. Zhang, D. Corkill, G. DeJong, T. G. Dietterich, S. Kambhampati, V. Lesser, D. L. McGuinness, A. Ram, D. Spearsy, P. Tadepalli, E. T. Whitaker, W.-K. Wong, J. A. Hendler, M. O. Hofmann, and K. Whitebread, "An ensemble learning and problem solving architecture for airspace management," in *IAAI'2009*, 2009.

[10] D. L. McGuinness, "Explaining complex systems," in *Semantic e-Science Workshop co-located with the Association for the Advancement of Artificial Intelligence Conference*, 2007.

[11] D. L. McGuinness, A. Glass, M. Wolverton, and P. Pinheiro da Silva, "Explaining task processing in cognitive assistants that learn," in *Proceedings of the 20th International FLAIRS Conference (FLAIRS-20)*, 2007, pp. 284–289.

[12] A. Glass, D. L. McGuinness, and M. Wolverton, "Toward establishing trust in adaptive agents," in *IUI*, 2008.

[13] N. Del Rio and P. Pinheiro da Silva, "Probe-it! visualization support for provenance," in *ISVC (2)*, 2007, pp. 732–741.

[14] J. Tao, L. Ding, and D. L. McGuinness, "Instance data evaluation for semantic web-based knowledge management systems," in *HICSS*, 2009.

[15] D. L. McGuinness, P. Fox, P. Pinheiro da Silva, S. Zednik, N. D. Rio, L. Ding, P. West, and C. Chang, "Annotating and embedding provenance in science data repositories to enable next generation science applications," in *American Geophysical Union, Fall Meeting (AGU2008), Eos Trans. AGU, 89(53), Fall Meet. Suppl., Abstract IN11C-1052*, 2008.

[16] J. R. Michaelis, L. Ding, and D. L. McGuinness, "Towards the explanation of workflows." in *Proceedings of the IJCAI'09 Workshop on Explanation-Aware Computing.*, 2009.

[17] L. Moreau, B. Plale, S. Miles, C. Goble, P. Missier, R. Barga, Y. Simmhan, J. Futrelle, R. McGrath, J. Myers, P. Paulson, S. Bowers, B. Ludaescher, N. Kwasnikowska, J. V. den Bussche, T. Ellkvist, J. Freire, and P. Groth, "The open provenance model (v1.01)," December 2008. [Online]. Available: http://eprints.ecs.soton.ac.uk/16148/

[18] J. R. Michaelis, S. Zednik, L. Ding, and D. L. McGuinness, "A comparison of the opm and pml provenance models," in *Tetherless World Constellation (RPI) Technical Report*, 2009, pp. TW–2009–21.

[19] J. Freire, D. Koop, E. Santos, and C. T. Silva, "Provenance for computational tasks: A survey," *Computing in Science and Engineering*, vol. 10, no. 3, pp. 11–21, 2008.

[20] J. Zhao, C. A. Goble, R. Stevens, and D. Turi, "Mining taverna's semantic web of provenance," *Concurrency and Computation: Practice and Experience*, vol. 20, no. 5, pp. 463–472, 2008.

[21] J. Freire, C. T. Silva, S. P. Callahan, E. Santos, C. E. Scheidegger, and H. T. Vo, "Managing rapidly-evolving scientific workflows," in *IPAW*, 2006, pp. 10–18.

[22] B. Clifford, I. T. Foster, J.-S. Vöckler, M. Wilde, and Y. Zhao, "Tracking provenance in a virtual data grid," *Concurrency and Computation: Practice and Experience*, vol. 20, no. 5, pp. 565–575, 2008.

[23] R. S. Barga and L. A. Digiampietri, "Automatic capture and efficient storage of e-science experiment provenance," *Concurrency and Computation: Practice and Experience*, vol. 20, no. 5, pp. 419–429, 2008.

[24] J. Kim, E. Deelman, Y. Gil, G. Mehta, and V. Ratnakar, "Provenance trails in the wings/pegasus system," *Concurrency and Computation: Practice and Experience*, vol. 20, no. 5, pp. 587–597, 2008.

[25] Y. L. Simmhan, B. Plale, and D. Gannon, "Karma2: Provenance management for data-driven workflows," *Int. J. Web Service Res.*, vol. 5, no. 2, pp. 1–22, 2008.

[26] A. Gates, P. Pinheiro da Silva, L. Salayandia, O. Ochoa, A. Gandara, and N. Del Rio, "Use of abstraction to support geoscientists' understanding and production of scientific artifacts," in *Geoinformatics: Cyberinfrastructure for the Solid Earth Science*. Cambridge University Press, 2009.

[27] O. Hartig, "Provenance information in the web of data," in *In Proceedings of the Linked Data on the Web (LDOW) Workshop at WWW'09*, 2009.