

Authors:

Victor Tan, U. Southampton
Paul Groth, U. Southampton,
Sheng Jiang, U. Southampton
Simon Miles, U. Southampton
Steve Munroe, U. Southampton
Luc Moreau, U. Southampton

August 31, 2006

WS Provenance Glossary

Status of this Memo

This document provides a glossary of terms intended to be used as a reference to other specification documents pertaining to the data model for process documentation [MGJ⁺06]. It does not define any standards or technical recommendations. Distribution is unlimited.

Copyright Notice

Copyright 2006.

Abstract

This glossary defines a set of terms used in the draft Provenance standard specification documents ([MGJ⁺06], [TMG⁺06a], [MTG⁺06], [TMG⁺06b], [GTM⁺06], [MMG⁺06b], [MMG⁺06a]) for the process documentation data model. The terms described here are intended to be implementation and technology independent, with the intent that they can be analyzed and applied to as many contexts as possible.

Contents

1	Introduction	3
1.1	Goals and Requirements	3
1.1.1	Requirements	3
1.1.2	Non-Requirements	3
2	Provenance Glossary	3

1 Introduction

The provenance of data and information generated within a computer system is the set of documentation that describes the processes that resulted in their creation. These processes can be viewed as sequences of causally related events that commence at some specific point in the past and eventually lead up to the event of creating the data or information of interest. The documentation of this sequence of events is based on a representational model that describes the storage, retrieval and processing of the documentation in a generic and technology independent manner. The data model for process documentation [MGJ⁺06] provides a detailed description of this model. This glossary assembles and defines a set of terms used in the model description. It also acts as a repository of all defined terms for the provenance standardisation documents related to the data model document.

1.1 Goals and Requirements

The goal of this document is to provide a set of defined terms used in all provenance standardization documents related to the data model for process documentation [MGJ⁺06].

1.1.1 Requirements

In meeting this goal, this document must address the following requirement:

- It must provide definitions for all terms used in all provenance standardization documents.

1.1.2 Non-Requirements

This document does not intend to meet the following requirement:

- Provide further clarification and context for the definitions provided. Such material will be provided in the documents from which these terms are collated from.

2 Provenance Glossary

Glossary

Actor

An actor is an entity with a distinct identity capable of undertaking some autonomous action within a provenance system. See also Asserting Actor and Querying Actor.

Actor State P-Assertion

An actor state p-assertion is an assertion made by an actor about its internal state in the context of a specific interaction.

Anonymous Documentation Style

The anonymous documentation style denotes a transformation of a message by which a part of (or the whole of) its contents has been replaced by an anonymous identifier. This identifier hides the actual data without losing the link to them.

Asserter Identity

An asserter identity is an identifier for an actor that asserts p-assertions.

Asserting Actor

An asserting actor is an actor that creates or asserts a p-assertion, which may subsequently be recorded to a provenance store.

Assertion Category Policy

An assertion category policy specifies what categories of p-assertions a service can record.

Composite Documentation Style

A composite documentation style indicates that more than one atomic documentation style has been applied to a message.

Data Accessor

A data accessor is a reference to the location of a p-assertion data item within a p-assertion's content.

Data Staging Policy

A data staging policy specifies whether or not a provenance store is capable of data staging. If so, it identifies which other provenance stores it can send its contents, its data staging capabilities for different classes of recording actors, and whether it is capable of both recording p-assertions into other provenance stores (push-based data staging) or querying other provenance stores (pull-based data staging), or both.

Data Upload Policy

A data upload policy specifies if and how frequently a provenance store has to resolve a reference contained in a p-assertion.

Document Language Mapping

A document language mapping is a definition of how to transform documents formatted in one document language into another document language.

Documentation Style

Documentation style is a representation of the transformation according to which the content of a message is asserted in an interaction p-assertion or the state of an actor is asserted in an actor-state p-assertion.

Documentation Style Policy

A documentation style policy describes the various different kinds of documentation style a recording service offers.

Global P-Assertion Key

A global p-assertion key is used to uniquely identify a p-assertion throughout the provenance system, and consists of an interaction key, a view kind and the local p-assertion identifier for that p-assertion.

Index Management Policy

The index management policy states whether a provenance store is capable of performing indexing and, if it can, the different kinds of indexing offered by the store must be enumerated.

Interaction Context

A set of interaction metadata about an identified interaction.

Interaction Identifier

A value that is unique for a given interaction message sent from one message source to one message sink.

Interaction Key

A globally unique identifier for an interaction.

Interaction Metadata

Provenance-related data about an interaction.

Interaction P-Assertion

An interaction p-assertion is an assertion of the contents of an interaction message by an actor that has sent or received that message; the message must include information that allows it to be identified uniquely.

Interaction Record

An interaction record encapsulates all the p-assertions and identifiers related to one interaction, and is uniquely identified by an interaction key.

Internal Reference Documentation Style

The internal reference documentation style specifies a transformation of a message by which a part of (or the whole of) its contents has been replaced by a global p-assertion key, which refers to another p-assertion that contains the actual data.

Link

A link is a reference to another provenance store.

Local P-Assertion Identifier

A value that is unique for each p-assertion made by one asserting actor about one interaction.

Message Sink

The address to which an interaction message was sent.

Message Source

The address from which an interaction message was sent.

Object Identifier

An object identifier uniquely identifies the objects of an asserted relationship p-assertion.

P-Assertion

A p-assertion is an assertion that is made by an actor and pertains to a process. See also Actor State P-Assertions, Relationship P-Assertion and Interaction P-Assertion.

P-Assertion Category

A p-assertion category is classification of a p-assertion on the basis of the type of information it can record.

P-Assertion Data Item

Part, or all, of a p-assertion.

P-Assertion Data Key

A globally unique identifier for a p-assertion data item.

P-Header

The p-header of an interaction is provenance-related contextual information, sent along with the interaction message.

P-Structure

The p-structure is a common logical structure of the provenance store shared by all actors including asserting, recording, querying and managing actors.

P-Structure Reference

A p-structure reference is a declaration of the p-structure over which a provenance query's entity search will be executed.

Parameter Name

A parameter name identifies an entity's role in a relationship, where that entity is documented by a p-assertion data item.

Process Documentation

The documentation of a process consists of a set of p-assertions made by the actors involved in the process.

Provenance Store Template Policy

A provenance store template policy indicates the ability of the store to accept templates from asserting actors, the languages for producing p-assertions the store supports, and the recording actors it can accept templates from.

Provenance of a piece of data

The provenance of a piece of data is the process that led to that piece of data.

Provenance Query Result Full Relationships

A provenance query result full relationship is the relationship between two p-assertion data items in the provenance of an entity found by a query.

Provenance Query Result Start

A provenance query result start is the p-assertion data key(s) to the process documentation of the entity for which the provenance was found, i.e. the key(s) for the p-assertion data item(s) found by resolving the query data handle.

Provenance Store

A provenance store is a repository dedicated for purpose of storing p-assertions created by asserting actors, and subsequently retrievable by querying actors.

Query Data Handle

A query data handle is a search over the contents of a provenance store in order to find the record of an entity at a given instant that the querying actor wishes to find the provenance of.

Query Store Policy

A query store policy states which provenance stores a querying service has access to.

Querying Actor

A querying actor is an actor that retrieves p-assertions from a provenance store for purposes of answering process documentation related queries.

Recipient Store Policy

A recipient store policy specifies which provenance stores a service can record to.

Reference Documentation Style

The reference documentation style denotes a transformation of a message by which a part of (or the whole of) its contents has been replaced by a reference to the location where the actual contents can be found.

Reference-Digest Documentation Style

The reference-digest documentation style specifies a transformation of a message by which a part of (or the whole of) its contents has been replaced by a reference to the actual location where it can be found and a digest of the substituted data.

Relationship P-Assertion

A relationship p-assertion is an assertion made by an actor, that describes the relationship between two interactions. A relationship p-assertion is directional and the nature of the relationship is application dependent .

Relationship Target

A relationship target is the full set of information about a p-assertion data item that is the subject or object of a relationship passertion.

Relationship Target Filter

A relationship target filter is a mechanism by which a querying actor can scope the provenance query results.

Search Language Policy

A search language policy identifies the search languages the provenance store supports.

Security Signature Checking Policy

A security signature checking policy specifies whether or not a provenance store has the capability to examine and validate the signatures placed in a p-assertion by an asserting actor, as well stating what action is to be taken if a conflict is detected.

Security-encryption Documentation Style

The security-encryption documentation style specifies a transformation of a message by which a part of (or the whole of) its contents has been encrypted.

Security-signing Documentation Style

The security-signing documentation style specifies the transformation of a message by which a part of (or the whole of) its contents has been signed.

Subject Identifier

A subject identifier uniquely identifies a data item or message acting as the subject of an asserted relationship p-assertion.

Tracer

A tracer is a piece of information used to associate an interaction with other, related interactions on the basis of some shared information.

Verbatim Documentation Style

The verbatim documentation style denotes a null transformation applied to the contents of a message.

View

A view is the set of p-assertions asserted by an actor about a specific interaction.

View Kind

Denotes, for a p-assertion, whether the actor making that p-assertion was the sender or the receiver in the interaction to which the p-assertion refers.

References

- [GTM⁺06] Paul Groth, Victor Tan, Steve Munroe, Sheng Jiang, Simon Miles, and Luc Moreau. Process Documentation Recording Protocol. Technical report, University of Southampton, June 2006.
- [MGJ⁺06] Steve Munroe, Paul Groth, Sheng Jiang, Simon Miles, Victor Tan, and Luc Moreau. Data model for Process Documentation. Technical report, University of Southampton, June 2006.
- [MMG⁺06a] Simon Miles, Luc Moreau, Paul Groth, Victor Tan, Steve Munroe, and Sheng Jiang. XPath Profile for the Provenance Query Protocol. Technical report, University of Southampton, June 2006.
- [MMG⁺06b] Simon Miles, Steve Munroe, Paul Groth, Sheng Jiang, Victor Tan, John Ibbotson, and Luc Moreau. Process Documentation Query Protocol. Technical report, University of Southampton, June 2006.
- [MTG⁺06] Steve Munroe, Victor Tan, Paul Groth, Sheng Jiang, Simon Miles, and Luc Moreau. WSRF Data Model Profile for Distributed Provenance. Technical report, University of Southampton, June 2006.
- [TMG⁺06a] Victor Tan, Steve Munroe, Paul Groth, Sheng Jiang, Simon Miles, and Luc Moreau. Basic Transformation Profile for Documentation Style. Technical report, University of Southampton, June 2006.
- [TMG⁺06b] Victor Tan, Steve Munroe, Paul Groth, Sheng Jiang, Simon Miles, and Luc Moreau. A Profile for Non-Repudiable Process Documentation. Technical report, University of Southampton, June 2006.