

Consolidated List of Document 'Topics':

Opening statement:

- * What is 'Success' of this effort?

IMO, this means that all Authors agree on the representation and description of the model to move forward.

To me, anything less would raise a big red flag that there is something critically wrong with the model. (ie: if the authors could not agree on the content, why is that? what's wrong?)

- * Status of document moving forward.

Pat has been pretty clear in the TCG meetings, that the level of RFC comments and level of contention within the author list

really indicates that the document needs to go back as WD. It needs to spend some time (3-4 wks?) for review

and comment in the working group.

+ We've seen in this interop especially, that there are MANY projects which are tied into the Provenance concept (outside of these use cases), it is important for it to be reviewed at each level. (Janet has been compiling a spreadsheet of project and their dependencies, and Provenance comes up much more than I expected, also coming up in many sessions.)

+ this model did not go through WG review at any time (my fault for allowing it to PR without that stage)..

Agenda:

- 0) High level comments on PDF and vodml-xml representations

* These should have the same normative content and be consistent. The pdf provides an avenue to add supportive information and non-normative information, but the model description should be identical (classes, attributes, and descriptions). Any differences will cause confusion for implementations. Further, if the vo-dml 'validates' and the pdf differs from that, how can you claim the model is vo-dml compliant? In my models, the model descriptions are pulled directly from the vo-dml/xml representation with an xslt script (vo-dml2ivoatex.xsl) to ensure they stay are the same.

< Snipping out several comments of inconsistencies between the vo-dml/xml and the document, before I realized this couldn't be correct>

o was looking at <https://volute.g-vo.org/svn/trunk/projects/dm/provenance/vo-dml/>

o these look like modelio generated html;

<https://volute.g-vo.org/svn/trunk/projects/dm/provenance/PR2ModelioDocumentation/IVOAProvDatamodelFeb22/2.html>

https://volute.g-vo.org/svn/trunk/projects/dm/provenance/ProvDM_PR2/?

I don't seem to find current vo-dml/xml and html files.?

< NO URL at this time >

* I see this came up before (scanning mail threads looking for pointer to vo-dml/xml...

+ The official 'model' is the vo-dml/xml

+ The official format for the Standard document which describes the model is PDF. (docStd)

So, it is very important that the PDF document exactly/correctly describes the vo-dml/xml.

- 1) Model change items

o assuming they are not changes from the representation that everyone agreed to proceed with, these are not part of the discussion today.

- 2) Critical items

o seems to be universal agreement that there is 1 primary, deal-breaker item.

- 3) non-normative elements moved to normative

- 4) Vocabularies, Enumerations and free strings

- 5) Typo-s and general organization comments

Agenda: 2019-05-16

- 1) Finish review items (pick up from where left off below.)
- 2) Review final change list going to the editor from the mtgs
- 3) Validation of changes: (misinterpretation of changes, typos, omissions)
- 4) Fill in details for Roadmap/Schedule

=====
Item Lists:

1) Model related items: NOT WITHIN SCOPE OF THE DISCUSSION

- o Multiple entities generating one entity.
 - *AG: would like to have at least a short discussion on that.
- o Section 2.3.2: WasGeneratedBy (same?)
 - *OS: Allow more than one Generation for an entity -- composition (mainly for future extension --> alternative activities/dataflows)
- o Figure 4: AGENT
 - *MCD: is a concrete object, unlike most other objects which are referring to a concrete object that is being related to some other. This object, in particular, overlaps with elements in DatasetMetadata (Contact, Contributor, Creator, Publisher, which refers to a 'Party' == Individual or Organization). There should be a plan to resolve this. Relates to Section 2.4.1
- o Section 2.6:
 - * OS: remove ValueEntityDescription.default
 - * OS: remove ~~(or strengthen)~~ ValueEntityDescription.options (same for ParameterDescription)
- o Section 2.6.2: ValueEntity.value
 - * OS: attribute may go into the main "Entity" class
- o Section 2.7: SORT-OF (See topic 0)
 - * OS: remove duplication of ParameterDescription.name in Parameter.name

2) Critical items:

- o Sections 2.5.4 and 2.7: attribute value "Configuration".
 - * AG: Major point on "configuration" keyword in the usage role - it is definitely required and needed (lots of arguments for that). This is a crucial point to discuss and needs to be settled.
 - * FB:
 - The text we have now is enough to introduce the subtlety of "configuring" using ActivityConfig and using some peculiar entities related to activities with some specific usage role.
 - a couple of reasons for that : we cannot claim to have two different representations of exactly the same thing in the model.
 - Actually after looking at MuseWise implementation I am more convinced that the supposed "two ways" are actually significantly different. The "configuring" entities of MuseWise are real entities
 - + follow-up mail on the topic thread
 - * MCD:
 - Sec2.7 describes the config as entity option.. not sure it is clear that it is not talking about the diagram for this section..
 - an example of each? is this a 'real' distinction? or perception?

RESOLUTION:

UsedDescription.type = "setup" added to Table 15
Section 2.7 text: 'with a given role and type="setup".'
In description: setup: "usage of an Entity as configuration information, see also Section 2.7"

3) non-normative elements moved to normative

Section 2.3.4: WasDerivedFrom

*AG: was not normative, personally I am in favor of keeping it, short discussion and a conclusion how to handle it.

*OS: should be again non-normative

** adds complexity and multiple threads thr

* Normative path open to misuse and does not produce meaningful 'graphical' provenance information? [comment from Markus]

* serves use case for derived entities

RESOLUTION: Keep normative. Can clarify usage thereof if needed in minor point update.

4) Vocabularies, Enumerations and free strings

Section 2.3.1/2: Usage and Generation roles: [MEDIUM]

* FB: it would be nice in the future to have standard vocabulary for that. But defining such vocabulary is a full semantics project attached to provenance. Although it should make our model and implementations more powerful I don't think it can be accomplished in a near future and I think the model can already work with free role vocabulary. Don't delay the PR for that.

RESOLUTION: non-issue; not doing vocabularies

* OS: (see 2.5.3; may be already discussed here)

* OS: "If an activity is deleted, then the corresponding Used relation needs to be removed as well" --> we don't have a "delete" operation, so we can't use that for description.

RESOLUTION: Add appendix section for UML conventions used. (ala Cube, Coords, Meas, etc)

Remove UML related information (delete stuff, *, etc..); can retain Entities may be used by more than one activity.

* OS: We need an implementation of Used.time that shows the intended use case (usage time after start and before stop of the activity)

NOTE FOR IMPLEMENTATIONS: (CTA Implementation does this)

Section 2.4.1: [SMALL]

* MCD: uses 'type' attribute vs Dataset model which has subclasses. Agent.type should be Enumeration in this case?

RESOLUTION: move to enumeration.

* MS:

move the part concerning roles:

« For example: telescope astronomer, observatory, pipeline operator, principal investigator, software engineer, project helpdesk. »

to the section 2.4.4:

« Agents may play a specific role with respect to an activity or an entity (for example: telescope astronomer, observatory, pipeline operator, principal investigator, software engineer, project helpdesk). »

RESOLUTION: move list of example Agent roles to 'roles' paragraph

Section 2.4.4: Agent Roles description [SMALL]

*MCD: does this mean one can use any word, but there is a set of 'preferred' words which should be used if appropriate? I think that could be phrased more clearly.

RESOLUTION: "the field is free text, but if one of the terms in Table X applies, it should be used."

*OS: shorten the list to what we really use in the moment. [observer, program (missing)]

*OS: a role is missing for the science group that is running a (ESO) observation program (see above)

RESOLUTION: stem from data cite; but more focused to astronomy.

OS to provide missing line, otherwise list remains as is, until the time when this moves to vocabulary.

*OS: Agent roles table header: what is "Label"? Remove it?

RESOLUTION: choose the text to use for the role and have only that column.

Table 9 describes what is in an Enumeration in the vo-dml (AgentRole)

Table 9: [SMALL-MEDIUM]

*MS: remove or explain the Label column [DONE above]

Table 9: observer [SMALL-MEDIUM]

*MCD: "should only be used in association with an 'observation' activity" [WITHDRAWN - no issue]

Table 9: contributor

*MCD: the actual creation? or defining the content?

RESOLUTION: Is intended to be open.

Table 9: creator

*MCD: publisher? or software tool.

** OVERLAP WITH DATASET METADATA **

Section 2.5.1: [SMALL]

* MCD: free with 'reserved' list.. see above comment Section 2.4.4.

RESOLUTION: use same text format as before.

* OS: restrict list to the types that are proven to add value (maybe, "observation" only).

RESOLUTION:

1) remove '...', add description.

Section 2.5.4: [SMALL]

* MCD: see 2.4.4

Section 2.7: [SMALL?]

* OS: explain WasConfiguredBy explain WasConfiguredBy.typeOfConfiguration and its type (vocabulary? String?)

RESOLUTION:

Add text for this attribute and type

5) Typo-s and General organizational comments

o OS: The document needs to be self-contained. Have a full textual representation of the model in the PDF. So, any class, any attribute, any relation shall be described in the document. And we need the corresponding vodml.xml *additionally*.

RESOLUTION: Agreed, PDF and vodml content must be in agreement and complete and fully described..

o OS: We should have the complete diagram in the document, not just in the appendix. Since the diagram shown in a1 is not so much more complex than the first one in section 2, we could just replace it.

RESOLUTION:

* Overview diagram full, containing all objects and relations.

* If attributes are omitted, all attributes should be omitted (all or none)

* all diagrams need correct relation image (no double arrow)

o OS: Merge description classes and specializations them into the previous sections for better readability

RESOLUTION:

* Overall organization is at the discretion of the editor.

Section 1.3: First sentence [TRIVIAL]

*MCD: wording of first sentence

"The IVOA Provenance Data Model is structuring and adding metadata to trace the original process followed during the data production for providing astronomical data. "

to

"The IVOA Provenance Data Model is structuring and adding metadata to trace the original process followed in the production of astronomical data. "

RESOLUTION: OK

Section 1.4: Previous efforts

* OS: clarify the role of W3C; see especially last sentence [??]

** not sure is W3C compliant.

RESOLUTION: Remove last sentence re "Compliance"

Figure 3: oddities in various relations in the diagram [TRIVIAL]

Figure 3: caption mentions 'dark yellow' objects, and there are none [TRIVIAL]

RESOLUTION: OK

Figure 4: Entity.location == string [SMALL]

* MCD: description says 'path or geographical location, e.g. a URL' what is the other option? how does one interpret? not anyURI?

RESOLUTION: in description, add another example not a URL.

Figure 4: Collection

* MCD: representation here looks good, but does not match figure 3

RESOLUTION: OK

Section 2.2.2: Activity [TRIVIAL]

* MCD: "The *Activity* class in the model have the attributes given in Table 2." 'have' -> 'has

RESOLUTION: OK

=====

Section 2.2.1 Entity, Collection [??]

*OS: Entity.invalidatedAtTime: either rename this, or make really clear what it means

RESOLUTION: add sentence: "This does not relate to the validity of the entity."

Section 2.2.1/2 Entity/Activity.name

* OS: clarify what this is; make clear (for clients) that it often may be empty

RESOLUTION: Keep it as is.

Section 2.3.2: WasGeneratedBy

*MCD: "a composition, as indicated in Figure 4 by a filled diamond. ...", Explaining the meaning of the UML within the description is odd.

** Already discussed in other context

Section 2.3.3: Roles in Entity-Activity relation [SMALL]

*MCD: "If this is not the case, if the image can only play the same role everywhere, only then it can be an intrinsic property of the entity. " - Can it?

RESOLUTION: remove the sentence

*OS: Don't mention (future) vocabulary (reserved words) for usage/generation roles

RESOLUTION: remove 'future' speculation text.

Section 2.5.1: [SMALL]

* MCD: " can be known before an *Activity* instance is created. " <== ??

RESOLUTION: is OK

Section 2.5.3: [SMALL]++

* MCD: "the role attribute of the related *Used* (respectively *WasGeneratedBy*) instances must point to the role attribute of this *UsageDescription* "

'must point to' ?? I think you're trying to say that instance A must point to a description with matching role, but there is no direct pointing.

RESOLUTION: 'match' instead of 'must point to'

* OS: clarify whether they are linked from *Used*/*WasGeneratedBy* via "role" or "(usage| generation)Description". Remove the other from *Usage*/*WasGeneratedBy* classes.

RESOLUTION: model change.. out of scope.

Section 2.6:

* OS: ~~remove~~ (or strengthen) *ValueEntityDescription.options* (same for *ParameterDescription*)

RESOLUTION: Table 17, items not ivoa.type

string with description 'comma separated list of possible values'

Section 2.6:

- * OS: ~~remove~~ (or strengthen) ValueEntityDescription.options (same for ParameterDescription)

Table 17: [MEDIUM]

* MCD: min, max types (number) are not ivoa type, there is no definition for what that means. The diagrams show 'string', so is inconsistent.

* MCD: options type (list) is not ivoa type.. closest is 'enum'? diagram shows 'string[0..1]', so what would the format be?

RESOLUTION: 'string whose value can be interpreted by the valueType attribute'

Section 2.7.2:

*Parameter.value == string in diagram, (value dependent) here.. with interpretation via 'Parameter.valueType' from corresponding description. This is not clear or consistent.

RESOLUTION: string type.

Typo-s ONLY

Section 2.3: Entity-Activity [TRIVIAL]

* MCD: "Each entity is usually a result from an activity, " ==> 'is usually a result of' or 'usually results from'

Section 2.3.5: [TRIVIAL]

* MCD: "flow of activities as it occurred " => it -> they

*OS: should be again non-normative

Section 2.4.1: [TRIVIAL]

* MCD: "The *Agent* class in the model have the attributes given in Table 5 " => 'have' -> 'has'

* OS: contact e-mail --> contact url (more general)

RESOLUTION: add attribute 'url' type 'anyURI', description "reference URL to the agent"

* OS: Clarify agent roles. "provider", "publisher" are not roles in the creation of an entity

RESOLUTION: retain

Section 2.4.4: Agent Roles description

*OS: Table header: comments --> description; make the description non-trivial

RESOLUTION: these terms are self-describing

Section 2.6:

* MCD: "must then be expose as properly described values. " => 'expose' -> 'exposed'

RESOLUTION: OK

Table 19:

* MCD: min, max types.. same as Table 17

RESOLUTION: same solution

Table 20:

* MCD: ConfigFile.location type is string, but is 'e.g. a URL', which would make it an ivoa.anyURI

RESOLUTION: keep as is.

Appendix A: diagram

* MCD: several oddities in the diagram

Appendix B: changes

* MCD: the list is small, but there have been SIGNIFICANT changes since the last drop.

RESOLUTION: OK

** Check for other instances of 'preferred word' usage maybe not covered above.

