

Init3dbw2 - Provenance metadata to support Semantic Workflows

Initial Analysis:

Item	Information/Response
Date:	12/10/2009
Requirement # unique id <SemCon Ops Initiative>. <analysts initials><requirement number> e.g. Init1dbw1 (eventually linked to Use Cases)	Init3dbw2
Originator/Customer's Name :	Yolanda Gil
Originator/Customer's Company:	USC
Stakeholder Community: Enter appropriate category of stakeholder from Primary Stakeholders: <ul style="list-style-type: none"> • Software and Application designers and architects • Software and Application engineers and developers • Scientific and medical researchers • Medical research protocol designers • Clinical and scientific research data and metadata managers • Clinicians • Patients • Medical research study participants • Broader Stakeholders: caBIG® Community WS NIH projects and related commercial COTS vendors (caEHR, SDO's (HL7, CDISC); International Collaborators (e.g. NCI, cancerGrid, China), Government and regulatory bodies (FDA, CDC, ONC) (link to view SemConOps Stakeholders description). 	cancer/medical researcher
Summary of requirement pre-interview, by Reviewer:	<p>This will require additional information be collected about the dataset, to go along side the data as an object. Additionally, there will be metadata about the individual data fields, that will allow appropriate transformations if needed between codesets, and this higher level data will describe the things that need to have been done to the data "the protocol" for the data. There were 3 usage scenarios described for which having this type of metadata is needed. It will describe the provenance metadata/pre-processing as a workflow at sufficient level of detail to learn about whether the data are suitable for processing for specific purposes.</p> <p>1: Construct, Store and Reuse workflows from existing discoverable data and services Narrative: A cancer researcher searches the grid or internet for data of interest and finds one or more sources of data in genomic repositories that is available for use. He asks the system if there are any services on the grid that accept this kind of data, and expects for the system to tell him about the service, whether the data can be aggregated, if so it does it, if not it tells him why not and offers options for transforming the data so it will be usable. The system detects if its possible to do something additional to the data to make it usable to the service (pre-processing), and if there is a transformation, allows him to find an existing transformation he can reuse, or build and save a new transformation. He can discover what processes or pre-processing was performed on the data he has found. He can assess whether or not the data is really suitable for use based on some known, domain specific information the system has about this service, or the algorithms it uses, and compare the data to these known requirements, the system will tell him if it is a good match or not. The system will let him make changes to the workflow, adding or removing steps. The researcher can run the service with the data he has found and save the results. When he saves the results, enough information about the workflow and the parameters/processes that were performed on the data are captured from the authoring tool to allow discovery and reuse of this workflow by others.</p> <p>2: Discover data of interest: A researcher wants to find data that has already been collected for use with caArray. They are able to find the data and to inspect the system to learn about what type of cells are in the database, what type of pathology is available for the data, etc.</p> <p>3. Aggregate data of interest: A researcher is able to query the system to find data that can be combined with their data. It is able to compare the characteristics of the dataset to ensure that the data are combinable, for example .</p>
Recommended Next Step Enter one: Follow-up interview, Observe, Use Case Template (text), Use Case Model (formalized/UML diagram), Group Discussion, Prototype, Waiting Room	Semantic workflows use cases - this is similar to Anand's requirement.

Interview

Item	Script / Question	Information/Response
------	-------------------	----------------------

1	Hello, my name is NAME. I am calling you today because NCI and caBIG are working toward a new and improved version of the semantic infrastructure to better support integration scenarios . Our first step was to organize requirements collected over the past year. Your organization has expressed a requirement/need for BRIEF STATEMENT OF USER REQUIREMENT. This has been identified as potentially a critical component to support application/data and service integration, and we need more information in order to enable us to meet this requirement. Do you have about 30 minutes to talk about this?	If yes, proceed. If no, note a good time to call back.
2	Do you have any solution integration needs? If so, what are they? Have you envisioned new ways of interacting with existing or new parts of the semantic infrastructure? <i>(prompt to elicit changes/new ways of using the infrastructure)</i>	Yolanda develops workflows and has some use cases that can leverage semantics.
3	Are there any business changes you are assuming we will be able to deal with? <i>(prompt to elicit changes/new ways of using the infrastructure)</i>	envision certain capabilities will exist - suggested that they be captured in pre-conditions
4	Are there any capabilities you are expecting to be available to support your needs? <i>(prompt to elicit expectations/dependencies)</i>	Provenance metadata objects linked to data objects; workflow authoring engine; software developers who can translate domain specific validation rules into algorithms that can be stored and used to query against domain specific provenance instances;
5	Do you use any of the existing software/services? If so, what do you like or dislike about it? <i>(if related to existing capability)</i>	No
6	If this requirement is met, what would be the benefits? If you do not have it, what would be the negative impact? <i>(prompt to elicit benefits/value - will help to prioritize)</i>	Researchers could easily run existing workflows with new data; researchers could find data and experiments that could be performed; researchers could find data and experiments related to their area of interest or data to help support their finding or hypothesis; new scientific discoveries could be made
7	If, for any reason, we were not able to create that solution, do you think there might be another way to solve this issue? Can you think of an alternative solution? <i>(prompt to elicit alternative solutions/workarounds)</i> <i>(to be prepared by the Requirement Analyst)</i>	TBD
8	Would you agree that we can summarize your requirement like this? <i>(Summarize one requirement in 2-3 lines and read back to interviewee for confirmation.)</i>	Create the ability to develop semantic workflows to leverage and build upon services and data available through the grid and internet
9	How important is this requirement to the interviewee? <i>Required: Customer Priority /Annotationment Analyst</i> <i>(Provides concrete assessment of the relative importance for the requirements specification)</i>	Select: 1. Must have 2. Should have 3. Nice to have but not essential 4. Other (describe)
10	On a scale from 1 to 3 with 1 being "not satisfied" to 3 "completely satisfied", how would you rate your overall satisfaction with the product if this requirement was met? (<i>Relative rating/ranking of how satisfied or dissatisfied interviewee would be if this requirement were met/not met</i>)	Select: 1. Not satisfied 2. Mostly satisfied 3. Completely satisfied 4. Other (describe)
11	Are there other requirements that you would like to share with us? I'd be more than happy to call you back another time, or if you have another 10 minutes, please share other issues you can think of. <i>(prompt to elicit any hidden - potentially higher priority requirements if they exist)</i>	(If yes, take notes to use in on a new page with this template; if time not available now, try to make appointment for another call.)
12	Who else should we talk to in order to elicit more information about this need? For specific service enhancement or requirement from Forum entry:	Name of contact(s) ---
13	Can you or someone else give me a step-by-step description of how you would describe the expected performance/behavior of the software in order for you to feel that your requirement is met? <i>(Required: Fit Criterion - will help us create test cases and user acceptance criteria - to be prepared by the Requirement Analyst)</i>	---
14	Forum Link:	---
15	URLs (optional):	Links to pages or applications related to this requirement
16	References (optional):	Links to articles, papers or presentations related to this requirement

Post Interview - ongoing throughout development of use cases:

Item	Description	Information/Response
------	-------------	----------------------

Requirement Type (required)	<p>Analyst's assessment of the most appropriate category/type of requirement (<i>no need to ask interviewee</i>):</p> <ul style="list-style-type: none"> • Functional: Fundamental or essential to the product - describes what the product has to do or what processing is needed • Nonfunctional: properties the functions must have such as performance, usability, training or documentation <ul style="list-style-type: none"> ◦ Project constraint: schedule or budget constraints ◦ Design constraint: impose restrictions on how the product must be designed, such as conformant to ISO 11179, utilizes 21090 or is able to work on a particular type of device ◦ Project driver: business-related forces such as descriptions of stakeholders or purpose of the product/project ◦ Project issue: conditions that will contribute to the success or failure of the project 	Functional and Non-functional Design constraints (new metamodels)
ConOp Initiative(s) <i>Requirements Analyst / Business Analyst</i>	<ul style="list-style-type: none"> • Select most appropriate initiative: (click for descriptions) • Initiative 1 - Distributed, federated metadata repositories and model repositories and operations • Initiative 2 - Automated generation of metadata from line-of-business artifacts • Initiative 3 - Rules management and contracts support (behavioral semantics) • Initiative 4 - Semantics support for W3C service oriented architecture resources • Initiative 5 - HL7 CTS II/ OMG MIF compliant federated terminology services • Initiative 6 - Controlled biomedical terminology, ontology and metadata content • Initiative 7 - Assessment of semantic unification of compositional and derivational models • Initiative 8 - Other 	Init 1, Init 3, Init 4
Use Case Linkage (required) <i>Business Analyst</i>	Which use case(s) is this requirement linked to? (should follow Use Case numbering scheme <SemCon Ops Initiative>.<analysts initials><requirement number>.<use case number>, for example Init1dbw1.1, Init1dbw1.2, Init2dbw2.1, 2.2, etc.	Use case Number(s):TBD
Conflicts / Dependencies (required) <i>Requirements Analyst / Business Analyst</i>	Are there any conflicts with other requirements / use cases?	Yes, there are dependencies on data and analytical services exposing provenance data and relating it to their data objects; dependent on analysts to create rules from domain specific documentation about how data should be curated/processed in order to run the experiment successfully/validly
Next Step (required) <i>Requirement Analyst / Business Analyst</i>	After reviewing the results of the interview, the forum, and all other materials related to this requirement, the analyst should recommend the next step, then attach the Tiny Link (on the Info tab) for this page to the Master List table.	Develop Use Case template (text) and schedule Follow-up interview - Possibly assign to Baris