

Init1bes13 - Bulk load unannotated content to Knowledge Repository

Initial Analysis:

| Item | Information/Response |
|---|--|
| Date: | 01/08/2010 |
| Requirement # unique id <SemConOps Initiative>.<analysts initials><requirement number> e.g. Init1dbw1 (eventually linked to Use Cases) | Init1bes13 |
| Originator/Customer's Name: | Dianne Reeves and Denise Warzel |
| Originator/Customer's Company: | NCI |
| Summary of requirement initial analysis, by Reviewer: (as unambiguously as possible, describe who (List of Actors) is interacting with the system, what the business goal is and how the system might support the actor's ability to achieve their goal) | <p>Business Goal: Streamline bulk load of content to MDR; identify missing content (annotation, components, objects, data types, value domains) , create the new "unique" content. Actor: Metadata Specialist</p> <p>A Metadata Specialist (more specifically Metadata Curator), working at NHLBI and other divisions, wants the ability to bulk/batch load new content from a variety of sources for later curation.</p> <p>So there is a need for a loading tool that will allow:</p> <ul style="list-style-type: none"> - Bulk/batch load contents in parsable formats to MDR (Knowledge repository) - Match loaded content to potential reusable/existing metadata based on descriptions provided in content. Matching content is classified as "used by/using" an existing metadata item. For instance an indication showing it is reused by NHBLI content submitted at so-and-so date (source designation /attribution) - If a perfect match is not possible or to provide more flexibility to user, allow navigation/search /exploration/mapping related capabilities. For instance allow fuzzy matching (e.g. anything but value domain match or a subset permissible values as reported already exists as part of an existing value domain) - If no content is matched, allow creation of new content <p>The loading tool should allow the content to be partially formed and incomplete.</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 | Follow-up interview to Dianne Reeves |

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? | Yes. |
| 2 | What do you do? What are your goals for the next year? Why are you doing this? | Being able provide the content loading capability set by this requirement. To help NCI stakeholders to be able to register their content easily in Knowledge Repository. To streamline the process through this tooling since volume of content is too large and current tooling does not scale up to this challenge. |
| 3 | In interacting with the caBIG infrastructure, 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? (prompt to elicit changes/new ways of using the infrastructure) | The challenging part of this requirement is to be able to identify matched content. This step is time consuming and/or laborous. My expectations is to do this in a almost fully-automated fashion to be able to address a simple question such as what percentage of the content provided has a match in Knowledge Repository (e.g. a new/reusable proportion). |
| 4 | Are there any business changes you are assuming we will be able to deal with? (prompt to elicit changes/new ways of using the infrastructure) | One business change; a staging environment that allows use of full capacity of the required tool but do not publish the results/content until the task of matching/creation is completed. While doing this, the identifier created for content should remain the same even if the content is published later on. |
| 5 | Are there any capabilities you are expecting to be available to support your needs? (prompt to elicit expectations/dependencies) | A content loading tool that will interact with Knowledge Repository that does smart/targetted searches so that it can easily identify missing content and/or create a new one. |
| 6 | Do you use any of the existing software/services? If so, what do you like or dislike about it? (if related to existing capability) | Yes. You use existing tools. The current model to achieve content load as described in the requirement with current tooling is too time consuming. Furthermore current tooling requires a lot manual work that may generate inconsistent/incomplete content because subject to human error/judgement/experience. The more automated this process gets the more consistent content we will be able to generate. |
| 7 | If this requirement is met, what would be the benefits? If you do not have it, what would be the negative impact? (prompt to elicit benefits/value - will help to prioritize) | You won't be able to scale up to content load need from community unless you use extensive human resources. Increasing the number of people for this task will result in more variability due to human factors. If the requirement is not met, the users that will be able to use this tool will be limited to expert metadata curator. The full automation may lead to a wider/broader user community. |
| 8 | 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? (prompt to elicit alternative solutions/workarounds) (to be prepared by the Requirement Analyst) | The current approach is the alternative approach. A curator driven review and matching of content. Again is very time consuming and resource intensive. |

| | | |
|----|--|--|
| 9 | Would you agree that we can summarize your requirement like this? (Summarize one requirement in 2-3 lines and read back to interviewee for confirmation.) | <p>A Metadata Specialist (more specifically Metadata Curator), working at NHLBI and other divisions, wants the ability to bulk/batch load new content from a variety of sources for later curation. Actually we would like expand the use of this tool beyond metadata specialist (e.g. regular users) so that at least initial matching capabilities of loading tool can be done by content owners.</p> <p>So there is a need for a loading tool that will allow:</p> <ul style="list-style-type: none"> - Set a search criteria to restrict the content matching. For instance limiting the match to all caBIG content or project such as caAERS or data standards only. - Bulk/batch load contents in parsable formats (e.g. XLS, CSV, tables in a PDF/Word document) to MDR (Knowledge repository) - Match loaded content to potential reusable/existing metadata based on descriptions provided in content. Matching content is classified as "using" an existing metadata item. For instance an indication showing content is reused by NHLBI content submitted at so-and-so date (source designation/attribution) - If a perfect match is not possible or to provide more flexibility to user, allow navigation/search/exploration/mapping related capabilities. For instance allow fuzzy matching (e.g. anything but value domain match or a subset permissible values as reported already exists as part of an existing value domain) - If no content is matched, allow creation of new content - Generate a reuse report based on the matches/searches. Reuse report should be detailed enough to capture different levels of metadata reuse (e.g. DEC, CDE, VD) <p>The loading tool should allow the content to be partially formed and incomplete. There should mandatory requirements for content (to assess completeness). However during loading process and prior to publishing it the loading should allow to not to enforce mandatory requirements.</p> <p>The loading tool should allow staging the content load without publishing. The public identifier assigned to content should not change when the content is published. The loading tool should prevent any content to be published if they don't meet mandatory requirements for the content. The content in staging phase should be able to shared among a collaborators in the community.</p> |
| 10 | How important is this requirement to the interviewee? <i>Required: Customer Priority /Annotation/Requirement Analyst</i> (Provides concrete assessment of the relative importance for the requirements specification) | 1. Must have |
| 11 | 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> | 3. Completely satisfied |
| 12 | 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. (prompt to elicit any hidden - potentially higher priority requirements if they exist) | No for now. |
| 13 | Who else should we talk to in order to elicit more information about this need? | Robinette Aley (raley@NMDP.ORG) from National Marrow Donor Program is a potential beneficiary of the loading tool described here. In the use case analysis phase, she can be contacted for further input. |
| | For specific service enhancement or requirement from Forum entry: | --- |
| 14 | 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? (Required: Fit Criterion - will help us create test cases and user acceptance criteria - to be prepared by the Requirement Analyst) | <p>For the step "Setting up a search criteria to restrict the content matching.": The tool is expected to match to the content "only" available in the context/project or metadata restricted by the search criteria.</p> <p>For the step "Bulk/batch load contents in parsable formats (e.g. XLS, CSV, tables in a PDF/Word document) to MDR (Knowledge repository)": The tool should be able to parse given format and notify user about a successful /unsuccessful load.</p> <p>For the step "Match loaded content to potential reusable/existing metadata based on descriptions provided in content. Matching content is classified as "using" an existing metadata item.": The tool should demonstrate it can match different levels of metadata. An easy performance/fit test will be reloading existing content in various forms, in which case the tool should have match 100% of the content.</p> <p>For the step "If a perfect match is not possible or to provide more flexibility to user, allow navigation/search /exploration/mapping related capabilities.": The loader should demonstrate it can partially match different levels of metadata. An easy performance/fit test will be reloading existing content with slight variations (e.g. changing VDs, using subset of permissible values), in which case the tool should match all the content to different levels of metadata with exclusion of intended deviations.</p> <p>For the step "If no content is matched, allow creation of new content": The tool will facilitate creation of new content.</p> <p>For the step "Generate a reuse report based on the matches/searches. Reuse report should be detailed enough to capture different levels of metadata reuse (e.g. DEC, CDE, VD)": The tool should generate a report summarizing metadata reuse as different levels (e.g. percentage of VDs matched) and also listing the matched content in detailed form. The report should be printable for reading/revision.</p> <p>For the item "The loading tool should allow the content to be partially formed and incomplete.": There should mandatory requirements for content (to assess completeness). However during loading process and prior to publishing it the loading should allow to not to enforce mandatory requirements.": The tool should allow loading any content even if it requires loosing all the mandatory content requirements.</p> <p>For the item "The loading tool should prevent any content to be published if they don't meet mandatory requirements for the content. ": The tools should give an error if the requirements are not met and user tries to publish the content.</p> <p>For the item "The content in staging phase should be able to shared among a collaborators in the community. ": The user while working on the content, should be able to send a URL or pointer for others to browse/see content under development.</p> |
| 15 | Forum Link: | https://cabig-kc.nci.nih.gov/Vocab/forums/viewtopic.php?f=40&t=112 |
| 16 | URLs (optional): | Links to pages or applications related to this requirement |
| 17 | References (optional): | Links to articles, papers or presentations related to this requirement |

Post Interview - ongoing throughout development of use cases:

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

| | | |
|--|--|---|
| Stakeholder Community: | <p>Enter appropriate category of stakeholder from Primary Stakeholders:</p> <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 NCRI, cancerGrid, China), Government and regulatory bodies (FDA, CDC, ONC) <p>(link to view SemConOps Stakeholders description).</p> | Clinical and scientific research data and metadata managers Ideally any medical informatics professional wanting to load content to Knowledge Repository |
| 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 |
| ConOp Initiative (s) Requirements Analyst /Business Analyst | <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 | Initiative 1 |
| High Level Use Case Summary) Requirements Analyst /Business Analyst | Please write a short descriptive narrative use case; the steps or activities in this use case are usually the things the user wants to accomplish with the system (user/actor's goals). | --- |
| Use Case Linkage (required) Business Analyst | 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): |
| Conflicts / Dependencies (required) Requirements Analyst/ Business Analyst | Are there any conflicts with other requirements / use cases? | Yes OR No - If yes, what and why? |
| Next Step (required) (Requirement Analyst / Business Analyst) | 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. | Use Case Template (text) |