

Init1dbw6.pm8.U0 - Support caB2B to integrate services on caGrid

Contents of this Page

- [Expose service workflow metadata](#)
- [Discovery of analytical steps](#)
- [Storage and access of intermediate data](#)
- [Workflow sharing](#)
- [Define a metadata category](#)
- [Discover services by metadata category](#)
- [Perform operations based on metadata category](#)
- [Discover Related Data based on metadata](#)
- [Query for Data based on permitted valid values](#)

Expose service workflow metadata

Use Case Number	Init1dbw6.pm8.U1
Brief Description	It is commonplace in bioinformatics to string together a number of data and analytical operations in order to produce the desired output. In order for Cancer Researchers to discover which services can be piped together, it is necessary that the designers of the services expose the appropriate metadata.
Actor(s) for this particular use case	Information Modeler
Pre-condition The state of the system before the user interacts with it	A service exists that needs to be annotated.
Post condition The state of the system after the user interacts with it	The service is annotated sufficiently to be discovered and integrated into an analytical pipeline.
Steps to take The step-by-step description of how users will interact with the system to achieve a specific business goal or function	<ol style="list-style-type: none">1. The Information Modeler defines the metadata (inputs, outputs, service functionality, etc.) for the service.2. The service is deployed and accessible to its users3. The metadata is published in a way that can be queried upon and consumed by users of the service
Alternate Flow Things which would prevent the normal flow of the use case	None.
Priority The priority of implementing the use case: High, Medium or Low	High.
Associated Links The brief user stories, each describing the user interacts with the system for the one function only of the use case. There would potentially be a number of user stories that make up the use case.	<ul style="list-style-type: none">• Init1dbw6 - Support caB2B to integrate services on caGrid• Support caB2B Services to integrate data on grid
Fit criterion/Acceptance Criterion How would actor describe the acceptable usage scenarios for the software or service that meets the actor's requirement?	Sufficient metadata needs to be defined so that the service can be discovered and linked into a workflow.

Discovery of analytical steps

Use Case Number	Init1dbw6.pm8.U2
Brief Description	Once metadata about a service is defined and exposed, it must be queriable by users of the service. They must, through consuming of the metadata alone, be able to determine which services can act as consumers of the data the service produces, as well as produces of the data the service consumes. Furthermore, the user must be able to determine that the service is appropriately placed within the workflow.
Actor(s) for this particular use case	Cancer Researcher

Pre-condition The state of the system before the user interacts with it	Service-level metadata is exposed for a number of services that can be linked via a workflow.
Post condition The state of the system after the user interacts with it	The Cancer Researcher knows which services can act as inputs to which other services.
Steps to take The step-by-step description of how users will interact with the system to achieve a specific business goal or function	<ol style="list-style-type: none"> 1. The Cancer Researcher identifies an analytical service that he would like to use. 2. The Cancer Researcher performs a query to identify which services can produce data that can be fed into the analytical service. 3. The Cancer Researcher performs a query to identify which services can accept the data from the selected analytical service. 4. Through the service metadata, the Cancer Researcher can identify which services make sense to pipe data through, and identify any services that are needed for data transformations.
Alternate Flow Things which would prevent the normal flow of the use case	The query could begin with a dataset or data service, and the Cancer Researcher would be identifying all downstream data and analytical services.
Priority The priority of implementing the use case: High, Medium or Low	High.
Associated Links The brief user stories, each describing the user interacts with the system for the one function only of the use case. There would potentially be a number of user stories that make up the use case.	<ul style="list-style-type: none"> • Init1dbw6 - Support caB2B to integrate services on caGrid • Support caB2B Services to integrate data on grid
Fit criterion/Acceptance Criterion How would actor describe the acceptable usage scenarios for the software or service that meets the actor's requirement?	The user must be able to identify services based on input/output types, as well as find the appropriate translation services if needed.

Storage and access of intermediate data

Use Case Number	Init1dbw6.pm8.U3
Brief Description	When services are chained together into bioinformatic pipelines, it is often desirable to be able to store and then later access intermediate results of queries and analytics. These can be used to modify the pipeline as needed, or to share intermediate results with other investigators.
Actor(s) for this particular use case	Cancer Researcher
Pre-condition The state of the system before the user interacts with it	A service that produces data has been identified and is accessible, as well as the mechanism by which intermediate data will be stored.
Post condition The state of the system after the user interacts with it	The results of the service are available via the intermediate data service.
Steps to take The step-by-step description of how users will interact with the system to achieve a specific business goal or function	<ol style="list-style-type: none"> 1. The Cancer Researcher discovers and selects a service for storing the intermediate data. 2. The Cancer Researcher instructs the tool or service to store the results in the intermediate data. 3. The Cancer Researcher invokes the service 4. The results are stored in the intermediate data service and are accessible via service calls
Alternate Flow Things which would prevent the normal flow of the use case	None.
Priority The priority of implementing the use case: High, Medium or Low	Low.
Associated Links The brief user stories, each describing the user interacts with the system for the one function only of the use case. There would potentially be a number of user stories that make up the use case.	<ul style="list-style-type: none"> • Init1dbw6 - Support caB2B to integrate services on caGrid • Support caB2B Services to integrate data on grid
Fit criterion/Acceptance Criterion How would actor describe the acceptable usage scenarios for the software or service that meets the actor's requirement?	Access to the intermediate data must be as seamless as access to any other service, and the data should be secured based on rules that the Cancer Researcher identifies.

Workflow sharing

Use Case Number	Init1dbw6.pm8.U4
Brief Description	Once a user identifies a service workflow of interest, he should be able to share that workflow in a way that makes it easy to encode, share with colleagues, reuse/rerun, modify, and extend.
Actor(s) for this particular use case	Cancer Researcher
Pre-condition The state of the system before the user interacts with it	A set of services of interest has been identified.
Post condition The state of the system after the user interacts with it	The service workflow is stored and shared.
Steps to take The step-by-step description of how users will interact with the system to achieve a specific business goal or function	<ol style="list-style-type: none"> 1. The Cancer Researcher encodes and saves the workflow 2. The Cancer Researcher identifies the other uses that can access the workflow 3. The Cancer Researcher reruns the workflow at a later date 4. The Cancer Researcher modifies the workflow at a later date 5. The Cancer Researcher copies and extends the workflow with additional steps
Alternate Flow Things which would prevent the normal flow of the use case	The steps listed above can be performed in any order any number of times with the exception that the workflow must be encoded and saved first.
Priority The priority of implementing the use case: High, Medium or Low	Low.
Associated Links The brief user stories, each describing the user interacts with the system for the one function only of the use case. There would potentially be a number of user stories that make up the use case.	<ul style="list-style-type: none"> • Init1dbw6 - Support caB2B to integrate services on caGrid • Support caB2B Services to integrate data on grid
Fit criterion/Acceptance Criterion How would actor describe the acceptable usage scenarios for the software or service that meets the actor's requirement?	The workflow must be accessible in much the same way as any other service.

Define a metadata category

Use Case Number	Init1dbw6.pm8.U5
Brief Description	The Metadata Category is the ability to save a particular view of classes and their associations in order to find services that match. For example, a Cancer Researcher may be interested in A->B->C and wants to be able to query services that support those classes and associations.
Actor(s) for this particular use case	Cancer Researcher
Pre-condition The state of the system before the user interacts with it	None.
Post condition The state of the system after the user interacts with it	A Metadata Category has been defined.
Steps to take The step-by-step description of how users will interact with the system to achieve a specific business goal or function	<ol style="list-style-type: none"> 1. The Cancer Researcher discovers a set of classes, attributes, and associations that he is interested in 2. The Cancer Researcher saves these as a Metadata Category in a repository
Alternate Flow Things which would prevent the normal flow of the use case	The Cancer Researcher may want to load, update, delete, or share an existing Metadata Category.
Priority The priority of implementing the use case: High, Medium or Low	Low.
Associated Links The brief user stories, each describing the user interacts with the system for the one function only of the use case. There would potentially be a number of user stories that make up the use case.	<ul style="list-style-type: none"> • Init1dbw6 - Support caB2B to integrate services on caGrid • Support caB2B Services to integrate data on grid

Fit criterion/Acceptance Criterion How would actor describe the acceptable usage scenarios for the software or service that meets the actor's requirement?	None.
--	-------

Discover services by metadata category

Use Case Number	Init1dbw6.pm8.U6
Brief Description	Once a Metadata Category is created, a Cancer Researcher can use it to discover services that support the underlying classes, attributes, and associations.
Actor(s) for this particular use case	Cancer Researcher
Pre-condition The state of the system before the user interacts with it	A Metadata Category has been identified.
Post condition The state of the system after the user interacts with it	A set of services of interest have been identified.
Steps to take The step-by-step description of how users will interact with the system to achieve a specific business goal or function	<ol style="list-style-type: none"> 1. The Cancer Researcher loads the Metadata Category 2. The Cancer Researcher invokes the appropriate function to discover services that support the underlying parts of the Metadata Category 3. The Cancer Researcher selects a subset of the services of interest
Alternate Flow Things which would prevent the normal flow of the use case	None.
Priority The priority of implementing the use case: High, Medium or Low	Low.
Associated Links The brief user stories, each describing the user interacts with the system for the one function only of the use case. There would potentially be a number of user stories that make up the use case.	<ul style="list-style-type: none"> • Init1dbw6 - Support caB2B to integrate services on caGrid • Support caB2B Services to integrate data on grid
Fit criterion/Acceptance Criterion How would actor describe the acceptable usage scenarios for the software or service that meets the actor's requirement?	None.

Perform operations based on metadata category

Use Case Number	Init1dbw6.pm8.U7
Brief Description	Once a user has identified a set of services that support a Metadata Category, he can invoke operations across those services and aggregate the results based upon the classes, attributes, and associations within the Metadata Category.
Actor(s) for this particular use case	Cancer Researcher
Pre-condition The state of the system before the user interacts with it	A set of services of interest has been identified via the Metadata Category.
Post condition The state of the system after the user interacts with it	The results from the cross-service operation are aggregated and presented to the user.
Steps to take The step-by-step description of how users will interact with the system to achieve a specific business goal or function	<ol style="list-style-type: none"> 1. The Cancer Researcher selects the operation (such as "query") he would like to invoke on the selected services 2. The services are invoked and results are returned 3. The results are aggregated based upon the Metadata Category 4. The aggregated results are returned to the user
Alternate Flow Things which would prevent the normal flow of the use case	None.
Priority The priority of implementing the use case: High, Medium or Low	Low.

Associated Links The brief user stories, each describing the user interacts with the system for the one function only of the use case. There would potentially be a number of user stories that make up the use case.	<ul style="list-style-type: none"> • Init1dbw6 - Support caB2B to integrate services on caGrid • Support caB2B Services to integrate data on grid
Fit criterion/Acceptance Criterion How would actor describe the acceptable usage scenarios for the software or service that meets the actor's requirement?	None.

Discover Related Data based on metadata

Use Case Number	Init1dbw6.U8
Brief Description	<p>caB2B relies on establishing/identifying relationships between concepts (entities) and/or properties (attributes) exposed through a service's implementation models (currently registered in caDSR/advertised on caGrid). For instance, we need to be able to identify (or compute) Gene (entity from Model A) and MyFavoriteGene (entity from Model B) are representing identical concepts, even when the entities are not named the same. Furthermore that Gene.entrezGenelid property (attribute from model A) and MyGene.geneEnterid (attribute from model B) are representing the same property/attribute for the given concept.</p> <p>Currently NCI concept mappings stored in the caDSR for the Object Classes and CDEs are leveraged to identify such relationships. CDE mappings are the simplest to map across models because the CDE means that the physical database representation (syntax) is the same in the two models and can be directly aggregated.</p> <p>The new SI should continue to provide services and processes for us to be able to identify such relationships. In accordance with ECCF, the services/processes should at the minimum leverage the traceability (computable) among different levels of models (CIM->PSM) and compliance among model representations (PSM->PSM), caB2B should be able to easily discover this equivalence.</p>
Actor(s) for this particular use case	Cancer Researcher
Pre-condition The state of the system before the user interacts with it	A particular concept of interest has been identified via the terminology browser or metadata browser. Class/entities in models have been associated with common concepts.
Post condition The state of the system after the user interacts with it	The results from the cross-model discovery operation are aggregated and presented to the user.
Steps to take The step-by-step description of how users will interact with the system to achieve a specific business goal or function	<ol style="list-style-type: none"> 1. The Cancer Researcher selects the concept of interest and the operation (such as "discover related models") 2. The SI services are invoked and results are returned 3. The results are aggregated based upon the matching entities and levels of entities in the models. For example, models or portions of models that are conformant at the CIM level are evident, at the PIM level, and at the PSM level and are self-evident in the results. 4. The aggregated results are returned to the user
Alternate Flow Things which would prevent the normal flow of the use case	None.
Priority The priority of implementing the use case: High, Medium or Low	High.
Associated Links The brief user stories, each describing the user interacts with the system for the one function only of the use case. There would potentially be a number of user stories that make up the use case.	<ul style="list-style-type: none"> • Init1dbw6 - Support caB2B to integrate services on caGrid • Support caB2B Services to integrate data on grid
Fit criterion/Acceptance Criterion How would actor describe the acceptable usage scenarios for the software or service that meets the actor's requirement?	Only Models where matching classes, attributes and associations are returned by the operation.

Query for Data based on permitted valid values

Use Case Number	Init1dbw6.U9
Brief Description	<p>caB2B uses value sets (that bind to concepts) to filter/constrain certain data elements in formation of queries to datasets. This requires runtime access (or metadata download prior to runtime) to possible values for specific data elements. Currently value domains that are associated with data elements stored in metadata repository (MDR/caDSR) are fetched and provided in query builders. The enumerated values are provided in pull down menus/lists to ensure the queries are formulated with correct allowable values. Similarly, although not supported by caB2B now,, the allowable ranges (e.g. 0 <= Age <= 130) and units of measure can also be provided to support correct query formulation.</p> <p>The new SI should continue to provide services and processes for us to be able to identify and compute the value set or allowable ranges or units of measure for querying data elements.</p>
Actor(s) for this particular use case	Cancer Researcher

Pre-condition The state of the system before the user interacts with it	The enumerations (value sets) or allowable ranges or units of measures for attributes/data elements/variables in a particular database are available to help form queries of the database.
Post condition The state of the system after the user interacts with it	Data matching the selected enumerations (values) or ranges is retrieved via an SI operation.
Steps to take The step-by-step description of how users will interact with the system to achieve a specific business goal or function	<ol style="list-style-type: none"> 1. The Cancer Researcher selects the entity/class of interest and an operation to retrieve the possible attributes/properties. (such as selecting "Array" and asking getAttributes) 2. The SI service returns the results (such as all the property/attributes associated with Assay in the models/data service). 3. The Cancer Researcher discovers that there is an attribute in the data source that is called Array.type and that the possible values are "Gene Expression" "SNP" and "Exon", and issues a query (such as "query array.type="Gene Expression") 4. The SI services are invoked and results are returned 5. The results are returned to the user
Alternate Flow Things which would prevent the normal flow of the use case	None.
Priority The priority of implementing the use case: High, Medium or Low	High.
Associated Links The brief user stories, each describing the user interacts with the system for the one function only of the use case. There would potentially be a number of user stories that make up the use case.	<ul style="list-style-type: none"> • Init1dbw6 - Support caB2B to integrate services on caGrid • Support caB2B Services to integrate data on grid
Fit criterion/Acceptance Criterion How would actor describe the acceptable usage scenarios for the software or service that meets the actor's requirement?	The researcher is able to discover what the possible or allowable values are for a data field and select/enter more values of interest, results are returned that match only the entered values.