Init6pm23.pm5 - Brain Tumor in silico study - Pathology and Radiology data models

Contents of this Page

- Init6pm23.1 - Agile Metadata Management
- Init6pm23.2 - Modeling and Sharing Analytical Algorithms
- Init6pm23.3 - Modeling Lab and Research Methodology

A Brain Tumor in silico use case consists of determining genetic, gene expression and outcome correlates of high resolution nuclear morphometry in the diffuse gliomas and their relation to MR features using Rembardt and TCGA datasets. This involves integrative analysis involving Pathology, Radiology and molecular data. The following semantic infrastructure use cases fall out of these requirements:

Init6pm23.1 - Agile Metadata Management

Use Case Number	Init6pm23.1
Brief Description	Specific scientific data elements will be shared amongst collaborators, requiring the need for a way to semantically describe the data. However, through the course of the study, new data elements will be added and some data element may change. Therefore, there is a need for an agile modeling approach that does not require significant effort to modify the information model and register the semantic metadata.
Actor(s) for this particular use case	Information Modeler
Pre-condition **The state of the system before the user interacts with it**	An information model is represented in UML, registered in the metadata repository, and in active use.
Post condition **The state of the system after the user interacts with it**	The information model is updated and able to be used in production.
Steps to take **The step-by-step description of how users will interact with the system to achieve a specific business goal or function**	1. The Information Modeler updates the UML model a. Classes and attributes are added b. Classes and attributes are removed c. Classes and attributes are renamed d. Attribute datatypes are changed 2. The new updates are propagated to the metadata repository (e.g. as a newly versioned model, set of changes to an existing model, etc.) 3. The model changes are implemented in the production data and analysis systems 4. The model updates are used in the production systems
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.**	Init6pm23
Fit criterion/Acceptance Criterion **How would actor describe the acceptable usage scenarios for the software or service that meets the actor's requirement?**	The modeling must be able to be performed and updated in a light-weight, Agile environment. Minimally, updates may be made monthly on an iterative basis. They should take no longer than days to define and propagate to the metadata repository.

Init6pm23.2 - Modeling and Sharing Analytical Algorithms

Use Case Number	Init6pm23.2

Brief Description	Data elements are generated using specific algorithms. There needs to be a way to model the features of the algorithm itself and tie it back to the original data. One of the features of the algorithm could be the code of the algorithm itself. It would be ideal if this type of model could be generalized for use in the caBIG analytical community.
Actor(s) for this particular use case	Information Modeler, Software Engineer
Pre-condition **The state of the system before the user interacts with it**	An algorithm exists, it is coded, and its features are known. Outputs from the analytical routine are modeled and registered.
Post condition **The state of the system after the user interacts with it**	A semantically sound description of the algorithm is defined, able to be shared with others, and able to be associated with data that the algorithm created.
Steps to take **The step-by-step description of how users will interact with the system to achieve a specific business goal or function**	1. The Information Modeler imports the standard analytical routine semantic metadata model into his UML model. 2. The Information Modeler extends the analytical routine metadata as needed 3. The Information Modeler associates output parameter data elements to the analytical metadata 4. The Information Modeler registers his model, including input parameters, output parameters, and associated extended analytical routine metadata 5. The Software Engineer generates a grid service using the registered metadata 6. The Software Engineer provides the analytical routine metadata values, which is exposed in the service-level metadata 7. The Software Engineer implements the system, which links the analytical outputs to the analytical routine data.
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**	Medium
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.**	Init6pm23
Fit criterion/Acceptance Criterion **How would actor describe the acceptable usage scenarios for the software or service that meets the actor's requirement?**	A common standard model is highly desirable, though it is necessary for different Information Modelers to extend it.

Init6pm23.3 - Modeling Lab and Research Methodology

Use Case Number	Init6pm23.3
Brief Description	Metadata describing analytical results should link to the methodology used to generate the input data, such as the way that the biological specimens are analyzed and treated. In some cases, there can be clear overlap between the research methodology and the study design, which should be linked if possible.
Actor(s) for this particular use case	Information Modeler, Software Engineer
Pre-condition **The state of the system before the user interacts with it**	Input data to analytical services has been generated using a particular set of steps, which is termed the methodology. A UML model exists for modeling the input parameter data.
Post condition **The state of the system after the user interacts with it**	Research methodology metadata is modeled and associated to input parameter metadata.
Steps to take **The step-by-step description of how users will interact with the system to achieve a specific business goal or function**	1. The Information Modeler imports the standard research methodology semantic metadata model into his UML model. 2. The Information Modeler extends the research methodology semantic metadata as needed 3. The Information Modeler associates input parameter data elements to the research methodology semantic metadata 4. The Information Modeler associates study design data elements to the research methodology semantic metadata 5. The Information Modeler registers his model, including input parameters and associated extended research methodology semantic metadata 6. The Software Engineer generates a grid service using the registered metadata 7. The Software Engineer implements the system, which links the analytical inputs to the research methodology data.
Alternate Flow Things which would prevent the normal flow of the use case	Researcher enters the research methodology data when the input parameter data is entered into the system.

Priority **The priority of implementing the use case: High, Medium or Low**	Medium
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.**	Init6pm23
Fit criterion/Acceptance Criterion **How would actor describe the acceptable usage scenarios for the software or service that meets the actor's requirement?**	A common standard model is highly desirable, though it is necessary for different Information Modelers to extend it.

Init6pm23.4 - Analytical Provenance Tracking

Use Case Number	Init6pm23.4
Brief Description	The ultimate result of the scientific use case is to design a classifier that predicts outcome. It will be necessary to describe the classifier, as well as have that description link back to the data that is used to generate it. In other words, the provenance of the data must be captured.
Actor(s) for this particular use case	Information Modeler
Pre-condition **The state of the system before the user interacts with it**	Information models exist and are registered for analytical input data, output data, analytical algorithm, and research methodology.
Post condition **The state of the system after the user interacts with it**	The inputs, analytical service, and outputs are linked through standard provenance data.
Steps to take **The step-by-step description of how users will interact with the system to achieve a specific business goal or function**	1. The Information Modeler imports the standard provenance semantic metadata model into his UML model. 2. The Information Modeler extends the provenance semantic metadata as needed 3. The Information Modeler associates all data that needs tracking to the provenance metadata 4. The Information Modeler registers his model, including input parameters and associated extended provenance semantic metadata 5. The Software Engineer implements the system, which links the analytical data via provenance data.
Alternate Flow **Things which would prevent the normal flow of the use case**	A Researcher uses provenance to backtrack an analytical flow.
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.**	Init6pm23
Fit criterion/Acceptance Criterion **How would actor describe the acceptable usage scenarios for the software or service that meets the actor's requirement?**	A common standard model is highly desirable, though it is necessary for different Information Modelers to extend it.