DataGrid Application Working Group

WP10 USECASES RELATED TO MEDICAL IMAGING APPLICATIONS

Document identifier:	
Date:	**7/05/2003**
Authors:	**J. Montagnat**
Editors:	
Document status:	

Abstract: This document identifies basic usecases required by many medical imaging applications.

Issue	Date	Comment
1.0	07/05/02	Moved DataGrid-10-TED-345148-0-2 document to this common application usecases format

CONTENT

1.INTRODUCTION

A.OBJECTIVES OF THIS DOCUMENT

This is an internal document of the European DataGrid project [R1]. This document describes the medical data management requirements. This document includes a description of medical images and medical image users, low-level use cases involving basic data manipulations and higher level use cases involving the complete datagrid infrastructure, specification of data related user requirements.

B.APPLICATION AREA

Medical data images management and processing on a grid.

C.APPLICABLE DOCUMENTS AND REFERENCE DOCUMENTS

Applicable documents

[A1] D10.1, Requirements for grid-aware biology applications.
[A2] D10.2, Grid-Aware Biomedical Applications for DataGrid Testbed Assessment
[A3] D10.3, Testbed Assessment for Biomedical Applications

Reference documents

[R1] EDG, http://www.edg.org/

D.TERMINOLOGY

Definitions

Glossary

DICOM	Digital Image and COmmunication in Medicine
MSS	Mass Storage System
RC	Replica Catalog
RM	Replica Manager
SE	Storage Element
WP	Work Package

2.MEDICAL DATA

Medical data is sensitive and should be accessible to authorized users only. This section describes the medical data, the medical user, andthe rights each one has on the data.

A.DATA CLASSIFICATION

Medical data for which a DataGrid is really needed are medical images due to their size. The most established standard for medical image is DICOM3 (Digital Image and COmmunication in Medicine, version 3). DICOM describes both an image format and a client/server protocol to store and retrieve images on/from a medical image server. Recent medical imagers implement the DICOM protocol and are image servers communicating with their clients over a TCP/IP hospital network.

The DICOM file format is made of a header containing metadata plus one or several image(s) in a single file. The metadata contains information on:

3.The patient: name, sex, age...

4.The image acquisition: acquisition device, constructor, acquisition parameters, acquisition date...

5.The image itself: number of images stored, size of images, field of view...

The most sensitive part of the data are patient's metadata. However, medical images should never be accessible to anybody except authorized users in general.

Although the DICOM format only stores a limited amount of metadata, a complete patient file is usually made of other metadata (expert report on images, medical analysis...). In the grid context, the data management system should allow to store this additional information.

Today, medical data are often stored and archived inside each medical image producer (hospitals, clinics, radiology centers...). They are usually disconnected from the outside world to solve security issues. It is not rare that all imagers are not connected to a single network even inside an hospital. Original data produced by medical imagers are stored in read-only DICOM databases. Only transformed images (new images produced by processing original data) are read/write for their owner.

Processing data over a grid raise problems of confidentiality and confidence the user may have in the grid security infrastructure. In particular, some data should not be accessible by non accredited system administrators where the data are replicated. This means that the target for replication of some data should be controlled, or that the data have to be encrypted. Ideally, medical data, including metadata and image data, should not be accessible to any third party that is non accredited (even system administrators). It is clear that this requirement cannot be fully fulfilled with UNIX systems used: encrypted data need to be decrypted for processing. Therefore, a system administrator could intercept the decrypted data on a target machine, or worse, the key used for decryption. To ensure a reasonable confidentiality level, we plan to decouple the sensitive metadata from the image data. The metadata will only be stored on a limited number of trusted sites where administrators are accredited. The image data can be stored on standard grid nodes given that it is encrypted to make its content retrieval impossible for other user and difficult for system administrators.

A.USERS CLASSIFICATION

Medical data may be accessible by 3 user groups:

6.Patients

The patient owns its data. Although the legal issues are not identical in all European countries, the patient is usually the only owner of his data.

7.Physicians

Physicians have a free access to the data of their patients. In hospitals, the authorization unit is a group of physician belonging to a medical department. Every physicians of a given department have access to all data of their patients. However, a physician from a department A has no *a priori* access to another department B data. He can obtain the authorization to access one or several data from department B if a physician from B grants him the access. Most physicians have read-only access to primary data produced in the hospital. They may be able to add some information (*e.g.* medical report) but not to destroy the original data. A restricted set of administrator should have read and write access to the data.

8.Researchers

Medical data may be available, with some restrictions, for research. Researchers that are not physicians should not have access to nominative data. Therefore, part of data headers should be blanked when accessed by this users group.

To each user group is associated different data access rights:

9.Patients

Patients have free read access to their data. They can query the data management system to retrieve data that they own. However, the cannot get any information on other patient's data.

10.Physicians

Physician have free read access to their patient's data. They may delegate their full access right to another physician or limited access right (without nominative data) to a researcher. However, they cannot delegate to other unprivileged users. A physician may query the medical information system to know where are data from a given patient available. However, he may not have direct access to this data if it is in a different department or hospital. The patient, or another physician, may grant him the access.

11.Researchers

Researchers may have access to some blanked images if authorized by a physician who has access right to these images.

Every other grid user (including local system administrators that are not accredited) should not be able to read medical nominative data.

12.LOW-LEVEL USECASES

USE CASE: A NEW DICOM IMAGE IS PRODUCED AND REGISTERED IN THE MEDICAL DATA SYSTEM

Identifier	*UC-10M-1*
Goals in Context	*An imager produces a new DICOM image. This image is registered into the medical data system.*
Actors	***Imager***
Triggers	***Replica Manager registration*** ***Mass Storage System registration***
Included Use Cases	
Specialised Use Cases	
Pre-conditions	***A new image is produced***
Post-conditions	***The image is registered and accessible to any authorized user***
Basic Flow	*The image metadata is stored in a metadata database* *A LFN is associated to the image* *The image is registered in the DataGrid replica manager*
Devious Flow(s)	***The registration fails***
Importance and Frequency	***Each time a new image is produced. Very frequent.***
Additional Requirements	

USE CASE: A PATIENT QUERIES THE MEDICAL DATA SYSTEM TO RETRIEVE ONE OF ITS IMAGES

Identifier	*UC-10M-2*
Goals in Context	*A patient queries the medical data system to know which of its images are available. The query is distrbuted over the metadata site holders to return a global search result. The patient may then pick one of the images found on any data server and retrieve the image for visualization on its console.*
Actors	***Patient***
Triggers	***Metadata query*** ***Data download***
Included Use Cases	
Specialised Use Cases	
Pre-conditions	***The user has proper access rights*** ***A distributed metadata manager is available***
Post-conditions	***The user get a list of LFNs and can download any of them***
Basic Flow	***The user use a high level interface to query the metadata service*** ***The query is distributed over all metadata catalogs*** ***The user get back a list of LFN she owns*** ***The user can upload any of this files through its LFN***
Devious Flow(s)	***The user is not authorized*** ***The query fails on one or several site(s)***
Importance and Frequency	***Frequent.***
Additional Requirements	

USE CASE: A PHYSICIAN QUERIES THE MEDICAL DATA SYSTEM TO RETRIEVE ALL MEDICAL FILES OF A PATIENT

Identifier	*UC-10M-3*
Goals in Context	*A physician want to know the medical history of its patient. He queries the distributed medical information system to obtain a list of files and metadata stored for this patient on all medical sites. He then decide to download a limited set of files for visualization on its console.*
Actors	***Physician***
Triggers	***Metadata query*** ***Data download***
Included Use Cases	
Specialised Use Cases	
Pre-conditions	***The user has proper access rights*** ***A distributed metadata manager is available***
Post-conditions	***The user get a list of LFNs and can download those for which he is authorized***
Basic Flow	***The user use a high level interface to query the metadata service*** ***The query is distributed over all metadata catalogs*** ***The user get back a list of LFN he owns*** ***The user can upload any fil***e he is authorized to access
Devious Flow(s)	***The user is not authorized*** ***The query fails on one or several site(s)***
Importance and Frequency	***Frequent.***
Additional Requirements	

USE CASE: A PHYSICIAN SENDS AN IMAGE TO A COLLEAGUE FOR HIS EXPERT DIAGNOSIS

Identifier	*UC-10M-4*
Goals in Context	*A physician asks a colleague to expertize an image of one of its patient. He delegates his access right to his colleague so that this colleague is able to retrieve the file.*
Actors	***Physicians***
Triggers	***Access Control Service***
Included Use Cases	
Specialised Use Cases	
Pre-conditions	***The physician has access right to the data***
Post-conditions	***The colleague get the same access right***
Basic Flow	***The physician delegates its access rights***
Devious Flow(s)	***The delegation fails*** ***The physician cannot delegate rights to a non accredited user (i.e. a non physician)***
Importance and Frequency	***Medium***
Additional Requirements	

USE CASE: A PHYSICIAN RETRIEVES A PATIENT'S IMAGE, PERFORM A PROCESSING AND STORE THE OUTPUT

Identifier	*UC-10M-5*
Goals in Context	*A physician sends a job to the DataGrid on a patient image. The job output is stored as a new image on a grid storage element. The physician can then query the DataGrid to retrieve and visualize the output.*
Actors	***Physician***
Triggers	***Job submission***
Included Use Cases	
Specialised Use Cases	
Pre-conditions	***The user has proper access rights to some computing resources and the input image(s)***
Post-conditions	***The output is registered and the user is authorized to access it***
Basic Flow	***The user creates a JDL through a high level interface*** ***The job is executed*** ***The output is registered on a storage element*** ***The user is notified of the job termination*** ***The user can download the resulting file(s)***
Devious Flow(s)	***The job fails*** ***The output registraiton fails*** ***The user is not authorized to access the result***
Importance and Frequency	***Basic job submission. Very frequent.***
Additional Requirements	

USE CASE: A RESEARCHER QUERIES THE MEDICAL SYSTEM TO RETRIEVE ALL IMAGES WITH GIVEN PARAMETERS

Identifier	*UC-10M-6*
Goals in Context	*A researcher want to assess a segmentation algorithm of 3D ultrasound sequences of the heart. He make a global query on the distributed information system for all 3D echocardiographic sequences he has access to and submit a job for each sequence. The job output are stored on grid SEs.*
Actors	***Ressearcher***
Triggers	***Metadata search*** ***Job submission***
Included Use Cases	***UC-10M-5***
Specialised Use Cases	
Pre-conditions	***The user is authorized to access some computing resources and some input data***
Post-conditions	***The data are processed and the results stored on SEs***
Basic Flow	***The user query the metadata service*** ***She get a list of LFN back*** ***For each LFN she submit a job*** ***For each job the output is registered*** ***The list of output LFNs is returned to the user*** ***The user can download any of the associated files***
Devious Flow(s)	***Metadata query fails*** ***One or several job submission fail(s)*** One or several output data cannot be registered
Importance and Frequency	***Frequent***
Additional Requirements	

USE CASE: A RESEARCHER WANT TO RETRIEVE ALL IMAGES THAT WHERE OBTAINED FROM AN ORIGINAL IMAGE

Identifier	*UC-10M-7*
Goals in Context	*A researcher want to know which images where produced from a given original image and which algorithms where used to produce them. He queries the medical system and obtain a list of all images processed images available and for each of them, the name of the program that was used for producing the file.*
Actors	***Researcher***
Triggers	***Log search***
Included Use Cases	
Specialised Use Cases	
Pre-conditions	***A computation logging service is available*** ***An algorithm reistration service is available***
Post-conditions	***The user knows a list of LFNs produced and the name of the algorithms used***
Basic Flow	***The user knows an input LFN*** ***The user queries the logging service*** ***A list of LFNs is returned*** ***The program name used for producing each LFN is returned***
Devious Flow(s)	***The query fails*** ***Some LFNs are not returned***
Importance and Frequency	***Low***
Additional Requirements	

USE CASE: A PHYSICIAN WANT TO RETRIEVE THE ORIGINAL IMAGE THAT WAS USED TO OBTAIN A PROCESSED IMAGE

Identifier	*UC-10M-8*
Goals in Context	*A physician studies a processed image and he would like to retrieve the original for checking its content. He queries the medical information system and downloads the original file for visualization.*
Actors	***Physician***
Triggers	***Log search***
Included Use Cases	
Specialised Use Cases	
Pre-conditions	***A computation logging service is available*** ***An algorithm reistration service is available***
Post-conditions	***The user knows the source LFN***
Basic Flow	***The user knows an input LFN*** ***The user queries the logging service*** ***The source LFN is returned*** (even if the source has undergone several transformation stages)
Devious Flow(s)	***The query fails***
Importance and Frequency	***Medium***
Additional Requirements	

USE CASE: A RESEARCHER DELETES AN IMAGE THAT HE EARLIER REGISTERED.

Identifier	*UC-10M-9*
Goals in Context	*A researcher deletes an image that was stored on a SE and that he owns. The image and all its replica are removed from the SEs. The metadata associated to this image are removed from the distributed information system.*
Actors	***Researchers***
Triggers	***Replica Manager*** ***Metadata Service*** ***Storage Element***
Included Use Cases	
Specialised Use Cases	
Pre-conditions	***The user owns the files he want to delete***
Post-conditions	***The file, all its replicas and all associated metadata are removed***
Basic Flow	***The user triggers file deletion*** ***All file replicas are removed*** ***All associated metadata are removed***
Devious Flow(s)	***The user is not authorized*** ***One or several replica is not deleted*** ***Some metadata are not deleted***
Importance and Frequency	***Frequent.***
Additional Requirements	

USE CASE: A PHYSICIAN IS SUSPICIOUS ABOUT AN IMAGE HE IS STUDYING

Identifier	*UC-10M-10*
Goals in Context	*A physician is suspicious about an image he is studying and would like to know where this image has been produced. It should always be possible to track the origin of a medical image (imager source, hospital), even after it has been processed.*
Actors	***Physician***
Triggers	***Log search***
Included Use Cases	***UC-10M-8***
Specialised Use Cases	
Pre-conditions	***A computation logging service is available*** ***An algorithm reistration service is available***
Post-conditions	***The user knows the source LFN***
Basic Flow	***The user knows an input LFN*** ***The user queries the logging service*** ***The source LFN is returned*** (even if the source has undergone several transformation stages) The associated metadata (source hospital, etc) are returned
Devious Flow(s)	***The query fails***
Importance and Frequency	***Medium***
Additional Requirements	

USE CASE: AN ABNORMAL DATA USAGE HAS BEEN DETECTED

Identifier	*UC-10M-11*
Goals in Context	*An abnormal (illegal) data usage has been detected. The system administrators should be able to identify users who accessed and manipulated some given data.*
Actors	***Administrator***
Triggers	***Log search***
Included Use Cases	
Specialised Use Cases	
Pre-conditions	***A logging services of computations is available***
Post-conditions	***The user responsible for invalid usage is identified***
Basic Flow	***The administrator knows the input data used*** The administrator search for computations involving that data The administrator identifies the user responsible for this computation
Devious Flow(s)	***The log search fails***
Importance and Frequency	***Low***
Additional Requirements	

13.HIGH-LEVEL USE CASES

End users want to use the grid middleware to solve complex problems involving image processes of whole databases. The user interface should be simple enough to allow non specialist to use the grid for solving realistic medical problems. Here are few higher level use cases that imply many grid services to respond to user demands.

USE CASE: AUTOMATIC INDEXATION OF NEW IMAGES

Identifier	*UC-10M-12*
Goals in Context	*Each new images imported in the medical information system should be pre-processed to extract some geometric and spectral parameters useful for later queries. In this use case, an image is pushed from an imager to the DICOM server. The image is then automatically sent to the grid for basic processings (texture analysis, intensity analysis, moments extraction...). Various processing may be performed in parallel. The computed parameters are stored together with the image metadata.*
Actors	
Triggers	
Included Use Cases	
Specialised Use Cases	
Pre-conditions	
Post-conditions	
Basic Flow	
Devious Flow(s)	
Importance and Frequency	
Additional Requirements	

USE CASE: ALGORITHM REGISTRATION

Identifier	*UC-10M-13*
Goals in Context	*A researcher wrote a new image processing algorithm he wants to make available to the user community. The medical information system should allow to record algorithm, their version number, the kind of input they accept, and the kind of output they produce. The new algorithm is recorded and its parameters are formally specified.*
Actors	
Triggers	
Included Use Cases	
Specialised Use Cases	
Pre-conditions	
Post-conditions	
Basic Flow	
Devious Flow(s)	
Importance and Frequency	
Additional Requirements	

USE CASE: IMAGE PROCESSING USING A REGISTERED ALGORITHM

Identifier	*UC-10M-14*
Goals in Context	*A medical user studies an image for diagnosis. He select its image, download it for visualization and queries the system for the available algorithms allowing to process this image. He can pick an algorithm, send the image processing to the grid and get as a result the identifier of the image that was produced and that has been stored on some SE. He may then want to download this image for visualization.*
Actors	
Triggers	
Included Use Cases	
Specialised Use Cases	
Pre-conditions	
Post-conditions	
Basic Flow	
Devious Flow(s)	
Importance and Frequency	
Additional Requirements	

USE CASE: COMPLEX QUERY

Identifier	*UC-10M-15*
Goals in Context	*A physician want to compare an image he is studying for similar cases in a medical images database. He queries the system for images with the same kind of acquisition parameters and a similar content. The image is therefore compared to all matching images in the database using a similarity measure algorithm. This causes a large number of similarity measure jobs to be executed on the grid.*
Actors	
Triggers	
Included Use Cases	
Specialised Use Cases	
Pre-conditions	
Post-conditions	
Basic Flow	
Devious Flow(s)	
Importance and Frequency	
Additional Requirements	

USE CASE: PROCESSING RE-EXECUTION

Identifier	*UC-10M-16a*
Goals in Context	*A physician is studying an image that was processed using some image processing algorithm and he would like to re-process the source image using the latest version of the algorithm. He queries the system for the source image and the algorithm that was used for generating the image he is visualizing. He then sends a new processing using the same input image and same algorithm but in its latest version.*
Actors	
Triggers	
Included Use Cases	
Specialised Use Cases	
Pre-conditions	
Post-conditions	
Basic Flow	
Devious Flow(s)	
Importance and Frequency	
Additional Requirements	

USE CASE: PROCESSING RE-EXECUTION

Identifier	*UC-10M-16b*
Goals in Context	*A biologists needs to re-run an experiment in the exact conditions that it was earlier ran (same computing host, same environement, same algorithm and data version).*
Actors	
Triggers	
Included Use Cases	
Specialised Use Cases	
Pre-conditions	
Post-conditions	
Basic Flow	
Devious Flow(s)	
Importance and Frequency	
Additional Requirements	

USE CASE: PIPELINE PROCESSING

Identifier	*UC-10M-17*
Goals in Context	*A pharmacology company want to assess a new drug it is developing. It needs to process a database of images coming from two groups of placebo and a treated population. Statistics are extracted from the two groups of processed images to evaluate the drug efficiency. The company builds a pipeline of processes that should be applied to each image in the control database (a chain of processes to run with inter-dependencies). A macro-job is then submitted that will cause all processes to be executed while dependencies are taken into account. The result is a set of images that are stored for statistics extraction.*
Actors	
Triggers	
Included Use Cases	
Specialised Use Cases	
Pre-conditions	
Post-conditions	
Basic Flow	
Devious Flow(s)	
Importance and Frequency	
Additional Requirements	

USE CASE: INTERACTIVE MODELING

Identifier	*UC-10M-18*
Goals in Context	*A physician want to use a complex organ model to segment an image. The segmentation process is interactive and allow the physician to supervise it progression. Through a graphic interface, the physician can check the algorithm progression and interact to correct some errors.*
Actors	
Triggers	
Included Use Cases	
Specialised Use Cases	
Pre-conditions	
Post-conditions	
Basic Flow	
Devious Flow(s)	
Importance and Frequency	
Additional Requirements	

USE CASE: PARALLEL JOB

Identifier	*UC-10M-19*
Goals in Context	*A researcher developed a simulator of Magnetic Resonance Images. Due to the complex equation of the MR process, the code is parallelized using a message passing interface. The researcher submit a job to create a new simulated image. The job involve deployment over several processes. The input data (the image model) is distributed over the different parallel nodes as well.*
Actors	
Triggers	
Included Use Cases	
Specialised Use Cases	
Pre-conditions	
Post-conditions	
Basic Flow	
Devious Flow(s)	
Importance and Frequency	
Additional Requirements	