





il patrimonio umanistico

Un caso d'uso: le opere di Federico De Roberto

Antonio Calanducci

antonio.calanducci@ct.infn.it

INFN Catania

Conferenza GARR 2009, Napoli 30 Sett - 2 Ott Network Humanitatis - linguaggi, risorse, persone



De Roberto cultural heritage

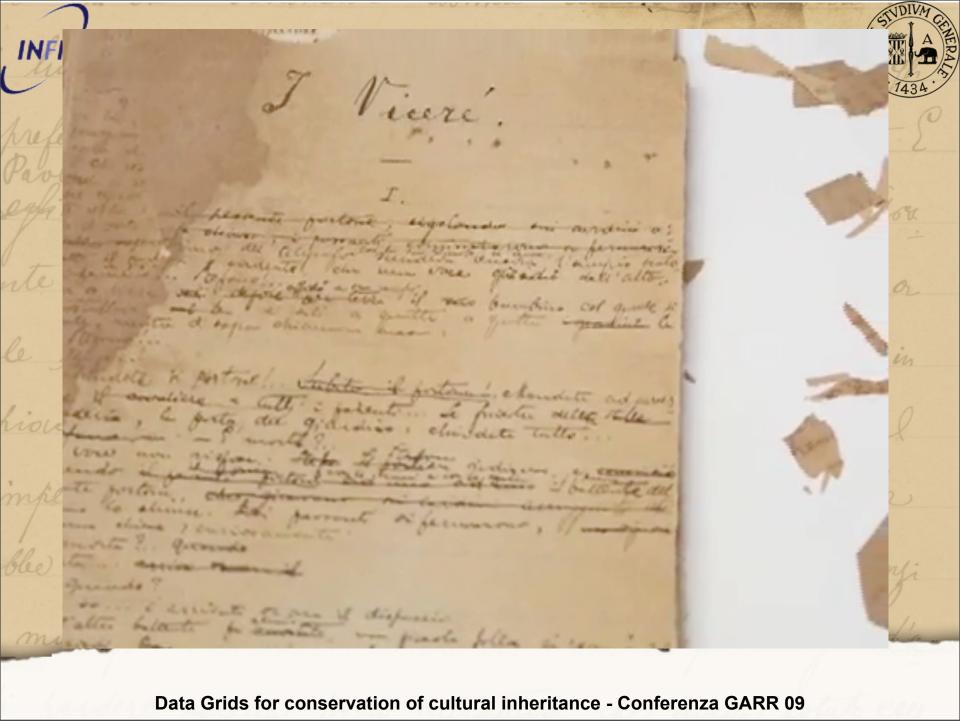


silentiosi, in

non erw an

- De Roberto, an Italian writer of the XIX/XX century, born in Naples, but spending his life in Catania, has left to the humanistic community numerous works
- Those are made up of valuable and hard-to-manage pieces: manuscripts, typescripts, drafts with handwritten corrections, magazines, cuts, sketches, photos, etc.





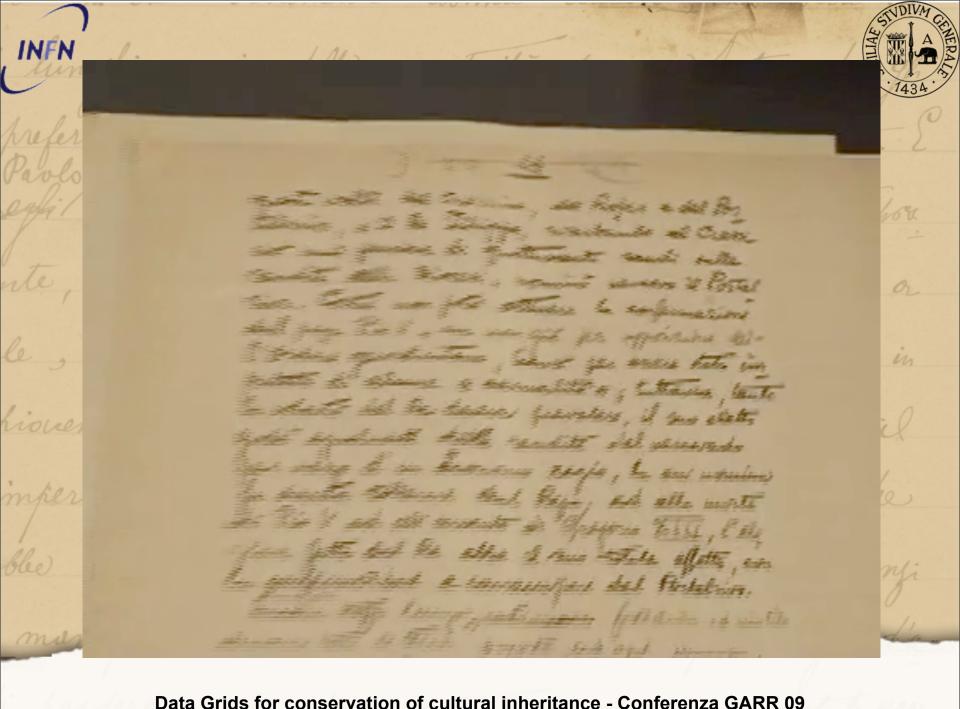


Digitalize to preserve them



- Some sheets are damaged (mold, crumbed pieces) and need physical restoration
- Digitalization to avoid the loss of this works, some of them still unpublished and relevant for the humanistic communities





i Name and a second of the sec

INFN Acquisition stage Jole Che preferirerro quelle and les alle fincialse inseperte. egsille innovaire i suoi giuramenti, con un traspore ute, scompigliondosi i capelli, stringendola fino or le, le sue pour finivans in sorrisi silentissi, in hiovenra estasiata chinanti alla certerror che il infero era aniora molto lontano dal tramonto, che ble stator ameria amuta. Certe volte, mon erw anji

Data Grids for conservation of cultural inheritance - Conferenza GARR 09

Acquisition stage

- Digitalization of manuscripts, typescripts, printed works
 - TIFF Files, one per page, 600 dpi, about 100MB for A3
 - High resolution scans for in-depth examination
 - PDF, one per work, 300 dpi, varying file sizes 40-400MB

liouenra estasiata chinanti alla certertor che il

inpero era aniona molto lontano dal tramonto, che

be stater ancora amata. Certe volte, non erw anji

- Overall examination of works
- 8000 sheets/scans, 3 Terabyte of disk space
- Different physical formats, A3/A4/custom size



Acquisition stage

- Digitalization of manuscripts, typescripts, printed works
 - TIFF Files, one per page, 600 dpi, about 100MB for A3
 - High resolution scans for in-depth examination
 - PDF, one per work, 300 dpi, varying file sizes 40-400MB
 - Overall examination of works
 - 8000 sheets/scans, 3 Terabyte of disk space
 - Different physical formats, A3/A4/custom size

Embedded Metadata

- TIFF with embedded metadata to provide scan physical features and information about the content
 - ImageWidth, ImageHeight, XResolution, FileSize, CreationDate, ModifyDate
 - Description, Keywords, CaptionWriter, Title, Author, Copyright Status, Copyright Notice
- Added with Photoshop after the digitalization phase (Adobe XMP format)

Goals and requirements preferiorers quelle and les alle fonciale inesperte. egsille innovaire i suoi girramenti, con un traspore ute, scompigliondosi i capelli, stringendola fino or le, le sue pour finivano in sorrisi silentivosi, in hiovenra estasiata chinanti alla certertir che il infero era aniora molto lontano dal tramonto, che ble stator ameria amuta. Certe volte, mon erw anji

Goals and requirements



- Make those works accessible to the humanistic communities
 - Always on-line: 24 x 365
 - Available from everywhere
 - Simple and easy-to-use interface for non-expert people

ite, scompigliandosi i capelli, stringendola fino a le, le sue pour finivano in sorrisi silentissi, in hiovenra estasiata dinanti alla certerrer che il inpero era aniona molto Containo dal tramonto, che ble stator anesa amata. Certe volte, mon erw anji



Goals and requirements



- Make those works accessible to the humanistic communities
 - Always on-line: 24 x 365
 - Available from everywhere
 - Simple and easy-to-use interface for non-expert people
- Quickly find the desired document
 - Document organization according the physical and semantic metadata
 - Organization by type/collections
 - · Dynamic filtering of search result sets according the selection of one or more document metadata

Data Grids for conservation of cultural inheritance - Conferenza GARR 09

impero era aniona molto lontoros dal tramonto, che

ble stator ancora amata. Certe volte, mon erw anji



Goals and requirements



- Make those works accessible to the humanistic communities
 - Always on-line: 24 x 365
 - Available from everywhere
 - Simple and easy-to-use interface for non-expert people
- Quickly find the desired document
 - Document organization according the physical and semantic metadata
 - Organization by type/collections
 - Dynamic filtering of search result sets according the selection of one or more document metadata
- Long-term preservation (digital preservation)
 - Multiple copies (replicas) spread in different geographical sites
 - Reliability of storage systems and replica redundancy to achieve secure preservation

Data Management in Grid

Storage Element(SE): front-end server aggregating a set of networked (pool) hard disks providing the illusion of a

big (virtual) disk

- "container" of users' files

generally one SE per site

- mirrored disks to avoid data loss in case of hardware failures
- fine-grained set up of file permissions: owner, group, given lists of users and groups (Access Control Lists -ACLs)
- Keep the mapping file-physical disk of the pool
- File Catalogue: provide a unique virtual file system among several Storage Elements: keep track of which SE (or SEs) contains a given file
 - keep track of replicas
 - mapping file-Storage Element filename



Data Management in Grid



silentioni, in

- Metadata Catalogue: store and organize metadata of files saved on Storage Elements and registered on the File Catalogue
 - metadata organized by "collection" (sort of directory)
 - each collection has its schema, a set of defined attributes:
 - · es: /deroberto/scans/manuscripts
 - o Title: "La lupa"
 - o Author: "Federico De Roberto, Giovanni Verga"
 - o Genre: "Tragedia Lirica"
 - o Pages: 34 dinanti alla certerre che il
 - o FileType: TIFF
 - o surl: srm://infn-se-01.ct.pi2s2.it/dpm/ct.pi2s2.it/home/cometa/generated/ 2008-06-14/filede4d6266-56c4-4d66-95b6-3d69063ef081
 - responsible to answer users' queries against metadata describing files, to find out their physical location for future retrieval

What Data Grids can offer to them preferirerro quelle and les alle finciale insperte. eglille innovaire i suoi girramenti, con un traspore ute, scompigliondosi i capelli, stringendolo fino or le, le sue pour finivans in sorrisi silentissi, in hiovenra estasiata chinanti alla certerror che il infero era aniora molto lontano dal tramonto, che ble stator ameria amuta. Certe volte, mon erw anji

Data Grids for conservation of cultural inheritance - Conferenza GARR 09

What Data Grids can offer to them store the 8000 scans of De Roberto Heritage ----> Data Grid Storage Elements le innovaire i suoi girramenti, con un traspore ute, scompigliandosi i capelli, stringendola fino a le, le sue pour friivano in sorrisi silentiosi, in hiovenra estasiata dinanti alla certerror che il inpero era anera molto Contano dal tramonto, che ble stater ancora amata. Certe volte, mon erw anji

What Data Grids can offer to them

- store the 8000 scans of De Roberto Heritage
 ----> Data Grid Storage Elements
- enable an ubiquitous and 24/24h access to scientists ---> Web Application

hiovense estasiata chinanti alle certestes che il impero ere ancora molto lontorno dal tramonto, che ble stator ancora amata. Certe volte, mon ere anzi manifetti il tima di

le, le sue pour finivano in sorrisi silentissi, in

scomforgliandosi i capelli, stringendola fino a

What Data Grids can offer to them



- store the 8000 scans of De Roberto Heritage
 ----> Data Grid Storage Elements
- enable an ubiquitous and 24/24h access to scientists ---> Web Application
- document organization for a quick search
 ---> Metadata Services

hiovenra estasiata chinanti alla certertor che il impero era ancora molto lontano dal tramonto, che ble stator ancora amata. Certe volte, mon era anzi manifetti il timo di serie solte, mon era anzi

What Data Grids can offer to them



- store the 8000 scans of De Roberto Heritage
 ----> Data Grid Storage Elements
- enable an ubiquitous and 24/24h access to scientists ---> Web Application
- document organization for a quick search
 ---> Metadata Services
- long-term digital preservation of data
 ---> redundancy through Replicas of files on several Storage Elements

Data Grids for conservation of cultural inheritance - Conferenza GARR 09

stator ameria amata. Certe volte, mon erw anji

What Data Grids can offer to them



- store the 8000 scans of De Roberto Heritage
 ----> Data Grid Storage Elements
- enable an ubiquitous and 24/24h access to scientists ---> Web Application
- document organization for a quick search
 ---> Metadata Services
- long-term digital preservation of data
 redundancy through Replicas of files on several Storage Elements
- simple and easy-to-use system for searches, organization, upload and download of digitalized documents on the Grid ---->

What Data Grids can offer to them



- store the 8000 scans of De Roberto Heritage
 ----> Data Grid Storage Elements
- enable an ubiquitous and 24/24h access to scientists ---> Web Application
- document organization for a quick search
 ---> Metadata Services
- long-term digital preservation of data
 ---> redundancy through Replicas of files on several Storage Elements
- simple and easy-to-use system for searches, organization, upload and download of digitalized documents on the Grid ---->



gLibrary features



- INFN-developed tool totally G ite based
- It allows to store, organize, browse & search and retrieve digital assets on a Grid environment with an intuitive front-end
- Digital Assets examples:



Requirements for web 2.0 apps for humanities



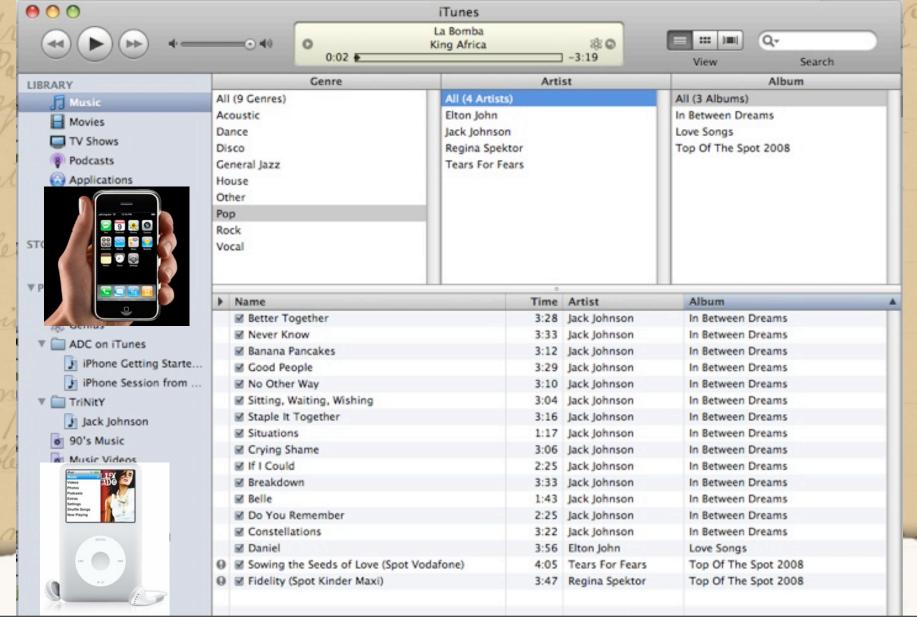
Web 2.0 in eight key concepts

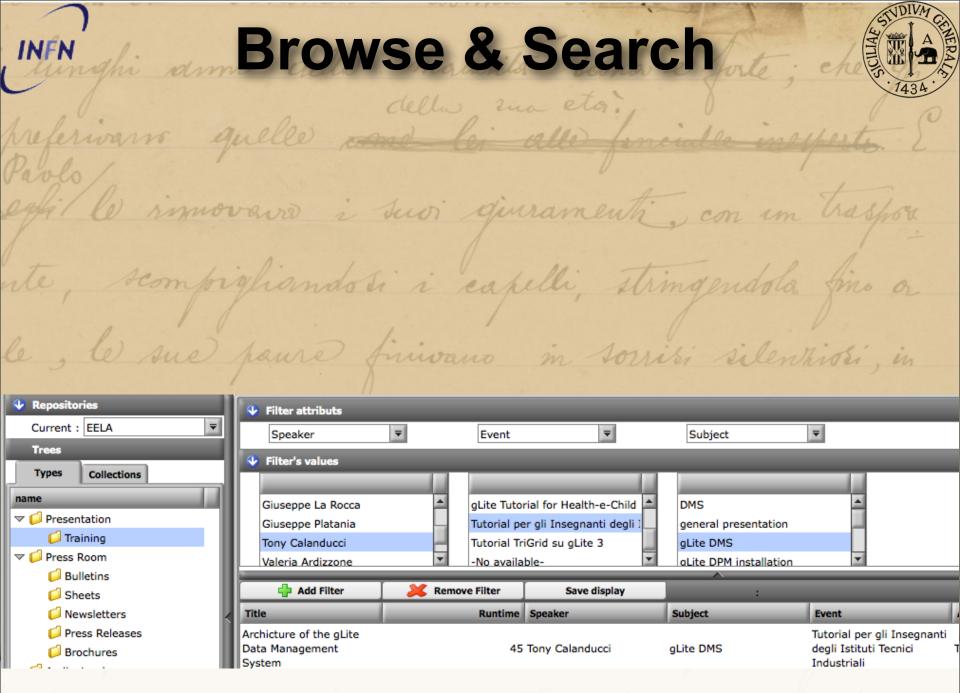
- User Generated Content (UGC)
- semantics
 - formal ontologies
 - social tagging
- collaborative filtering
- RSS feeds
- 5. embedding, reuse, mash-up
- social networks
- page interaction (Ajax, WebApps, widgets)
- 8. interface design



gLibrary as the iTunes for the Grid







Data Grids for conservation of cultural inheritance - Conferenza GARR 09

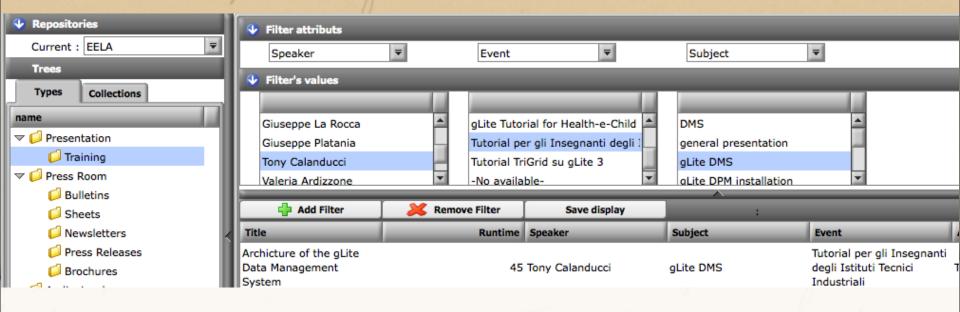


Browse & Search



- Assets can be browsed selecting a type (or collection) and selecting one or more filters
 - attributes of the selected types, chosen from a defined list, used to narrow the result set

scompigliondosi i capelli, stringendola fino ar le sue poure finivano in sorrisi silentiosi, in

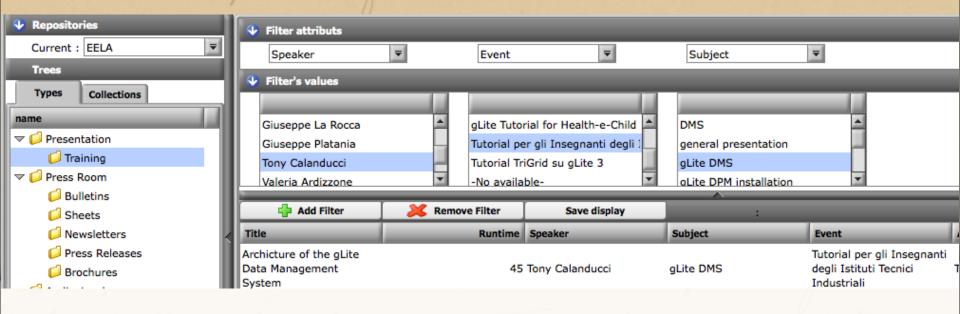




Browse & Search



- Assets can be browsed selecting a type (or collection) and selecting one or more filters
 - attributes of the selected types, chosen from a defined list, used to narrow the result set
- Filter application is cascading and context-sensitive: the selection of a filter value dynamically influences subsequent filter values ("à la iTunes" browsing)
 - Classical search by description and keywords available too



Data Grids for conservation of cultural inheritance - Conferenza GARR 09



Organize assets



"Types" and "Collections" definition by repository

providers/admins:

- Assets are organized by type:
 - a list of specific attributes to describe each kind of asset to be managed by the system
 - hierarchical (a child type shares and extend parent's attributes)
 - queried during searches

EXAMPLE OF TYPES AND ATTRIBUTES' LIST

Туре	Attributes' list	
Audio	Format, Bitrate, Samplerate, Time	
Music Le Sta	(Format, Bitrate, Samplerate, Time), Name, Artist, Album, Genre, Tracknumber, Year, Artwork, Lyric, Rating	
Presentation	Format, NumOfPages	
Training .	(Format, NumOfPages), Title, Runtime, Speaker, Author, Subject, Event, Date, Type	
(Root)	FileName, SubmissionDate, Description, Keywords, LastModificationDate, Size	

Collections Training Bulletins Sheets Newsletters Press Releases Brochures Video Photo 🔻 闻 Audio Music 📁 Film

and/or organized by collection:

- Group together related assets even of different types;
- Useful also to define subsets of assets belonging to the same type
- Multiple category assignment per asset (tagging like)

Collections

☐ M Favorites
☐ Bookmarks
☐ ☐ Playlists

■Summer2006



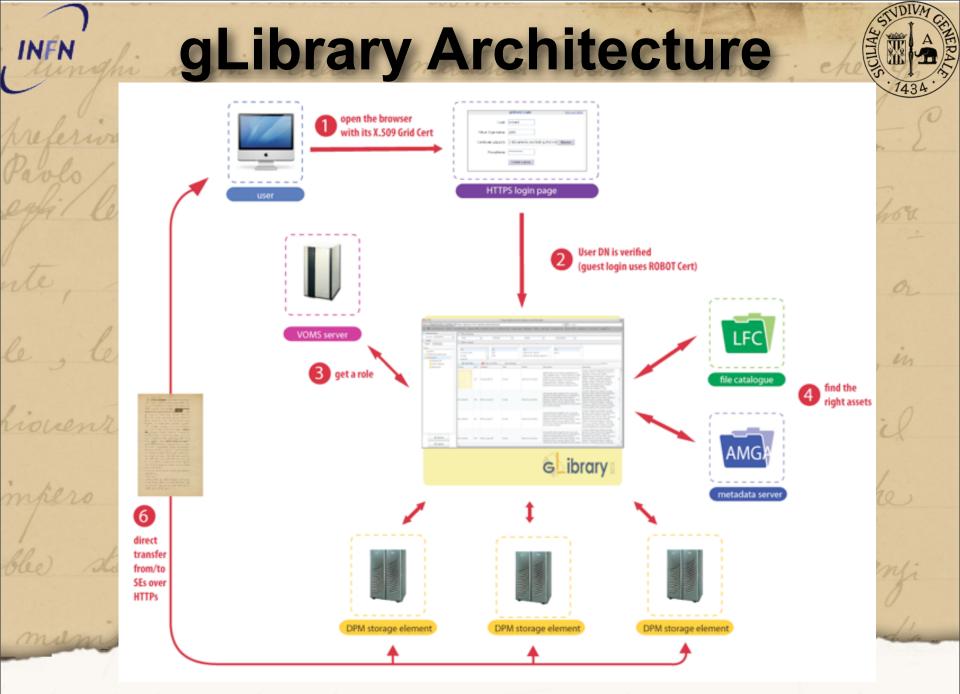
Store & Retrieve

- Users can upload their local assets on one or more (creating replicas) Storage Elements of the Grid
 - Files already on grid SE can be registered in a gLibrary repository by the LFC File Catalogue browser
- Download from SEs to the users' laptop/desktop:
 - selection of a replica link from a list





 Transfers are handled from the browser over HTTP/ HTTPS provided that users have their own X.509 Grid Certificate imported



Technologies used - ch preferioreros quelle and les alle fonciale inseperte. egsille innovaire i suoi girramenti, con un traspore ute, scompigliondosi i capelli, stringendola fino or le, le sue pour finivano in sorrisi silentiosi, in hiovenra estasiata chinanti alla certerror che il infero era aniora molto lontano dal tramonto, che ble stator ameria amuta. Certe volte, mon erw anji

Technologies used



- Web 2.0 standards:
 - Javascript/AJAX/JSON on the client side
 - PHP5 classes to implement business logic on the server side

ute, scompigliandosi i capelli, stringendola fino a le, le sue pour finivano in sorrisi silentissi, in hiovenra estasiata chinanti alla certertor che il inpero em aniona molto lontoros dal tramonto, che ble stator ameria amuta. Certe volte, mon erw anji



Technologies used



- Web 2.0 standards:
 - Javascript/AJAX/JSON on the client side
 - PHP5 classes to implement business logic on the server side
- Grid technologies:
 - Storage Element SRM interface to get the TURLs (Transfer URLs)
 - Transfers handled with GridFTP and X.509 cert auth HTTPS
 - X.509 based Globus Security Infrastructure with the VOMS extensions to handle authentication and authorization (ACL based) on Metadata and Storage Elements
 - All grid services implemented with the EGEE gLite middleware (DPM Storage Elements, AMGA Metadata Catalogue, LFC File Catalogue, VOMS Services)



Technologies used



- Web 2.0 standards:
 - Javascript/AJAX/JSON on the client side
 - PHP5 classes to implement business logic on the server side
- Grid technologies:
 - Storage Element SRM interface to get the TURLs (Transfer URLs)
 - Transfers handled with GridFTP and X.509 cert auth HTTPS
 - X.509 based Globus Security Infrastructure with the VOMS extensions to handle authentication and authorization (ACL based) on Metadata and Storage Elements

Certe volte, non erw an

- All grid services implemented with the EGEE gLite middleware
 (DPM Storage Elements, AMGA Metadata Catalogue, LFC File Catalogue, VOMS Services)
- Other standards:
 - Subset of XMP Metadata
 - Dublin Core Metadata set will be included in the next release
 - Data Grids for conservation of cultural inheritance Conferenza GARR 09



Metadata used in the DR digital library



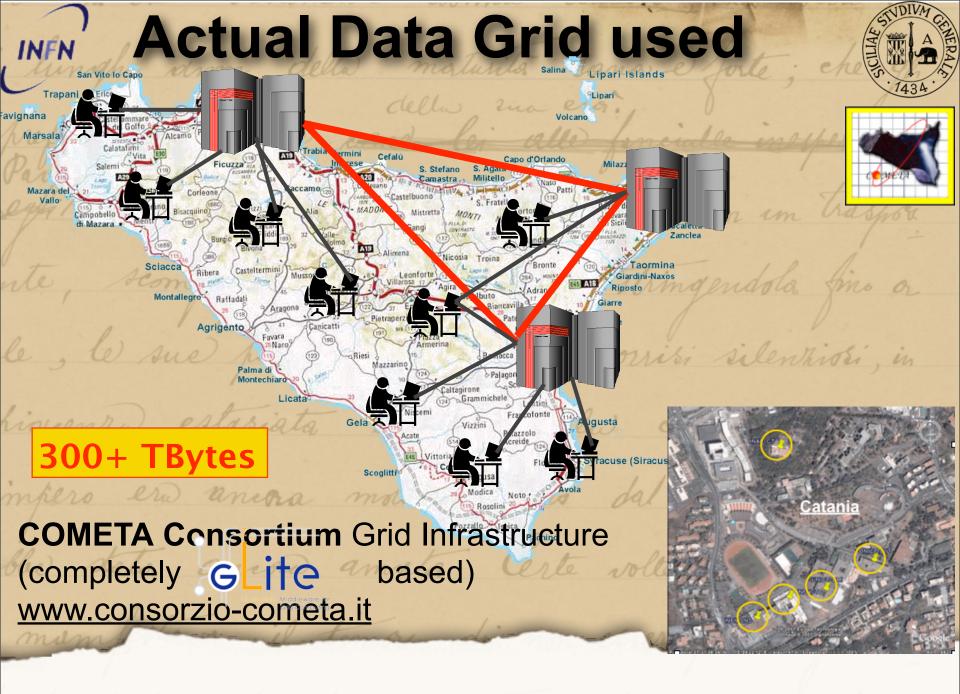
- Types definition for the assets of the DR repository ----->>>
- Attributes definition per type. Es (manoscritti):

Attributo	Valore 2 Suo Guramer		
Title	la lupa		
Author Scompra	federico de roberto, giovanni verga		
Description	manoscritto della tragedia lirica		
Keywords	verismo, federico de roberto, la lupa,		
CaptionWriter	stefania iannizzotto, alessandro		
CopyrightStatus	copyrighted		
PageNum	5		
TotalPages	34		
DocumentGenre	tragedia lirica		
PublicationYear	1916		
Publisher and an	officine tipo-litografiche barravecchia e balestrini		
FileType	PDF		
Resolution	300		
ScanQuality	good		

	Types	Categories			
	name				
 □ Video					
	Pubblicazioni				
۱	Articoli				
	Dattiloscritti				
	📁 Tes	sti a stampa			
	📁 Ma	noscritti			
-	V				

Filter defined per type. Es:

- DocumentGenre
- Title
- FileType
- ScanQuality
- DocumentType
- PublicationYear
- PublicationStatus
- Publisher
- Location





Who can benefit from gLibrary

- Community that has medium/large digital objects repositories to share in a short time
 - files can be taken on their servers if the are reachable from Internet and/or moved to Grid SEs
- Upcoming deployment of new reps:
 - Mediterranean Heritage Repository (*)
 - 250.000 musical scores of ancient Neapolitan musicians
 - digitized documents coming from Sicilian Library (Verga manuscripts)

INFN hrefer

Live Demo



un caso di studio



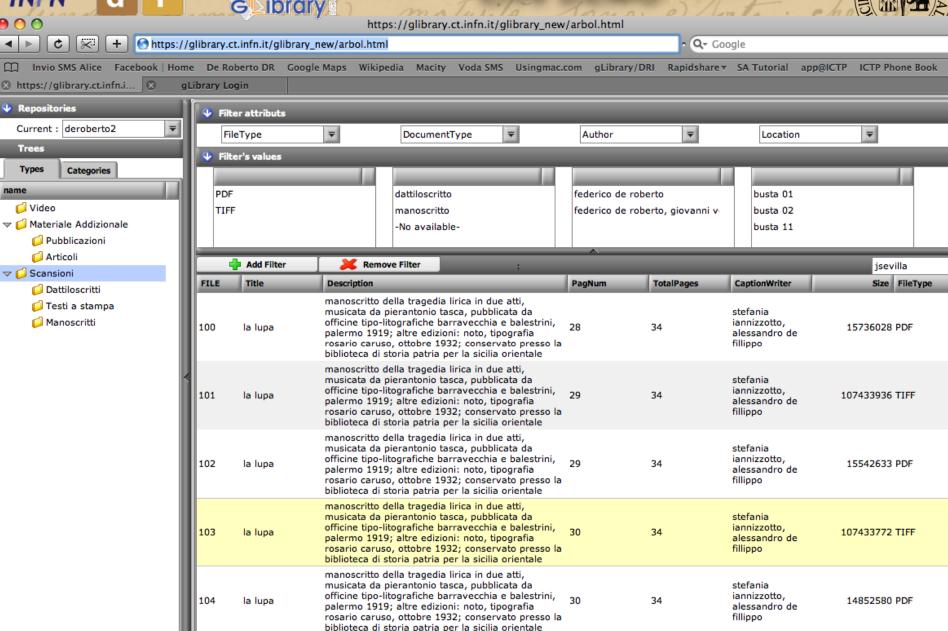


GLibrary with De roberto digital repository



Live DEMO

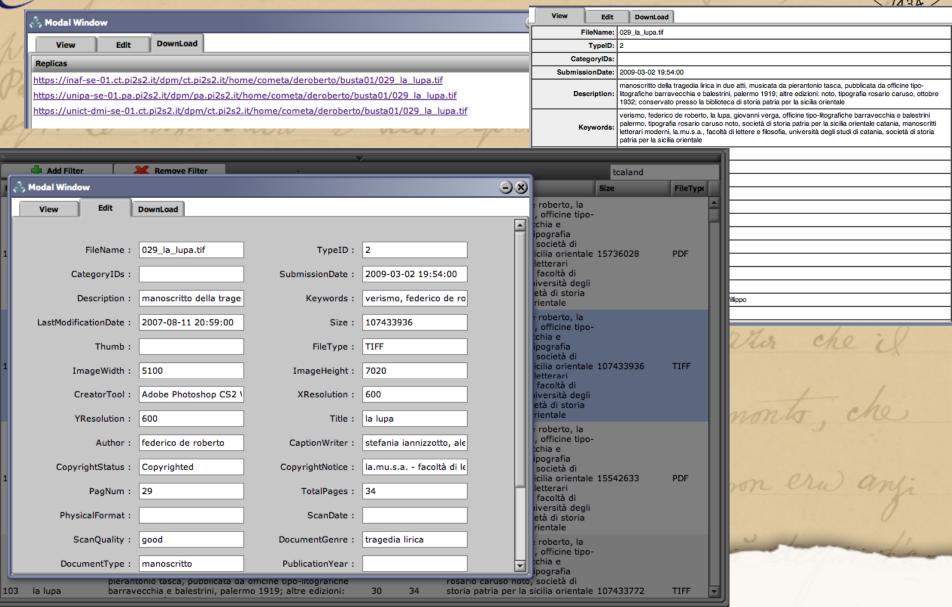


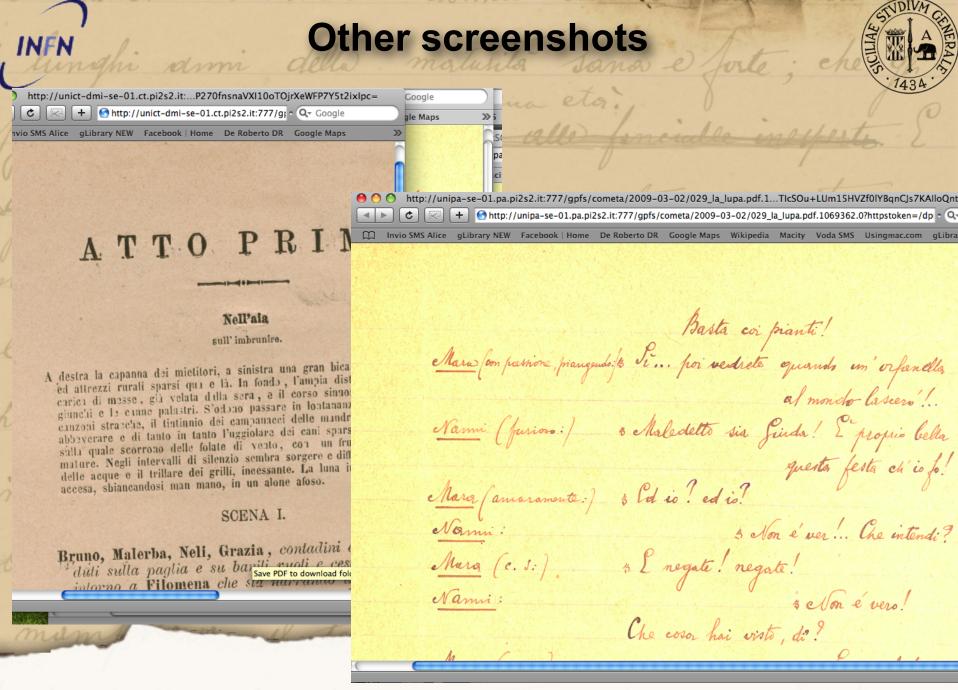


manoscritto della tragedia lirica in due atti,

More screenshots...









References



- Contact: antonio.calanducci@ct.infn.it, glibrary@ct.infn.it
- Prototype of the De Roberto Digital Repository:
 - https://glibrary.ct.infn.it/ i capelli stringendola
- YouTube video:
 - http://www.youtube.com/watch?v=VNN3OnpmUUU
- Previous papers:
 - A.Calanducci, R.Barbera, J.Sevilla, A. De Filippo, M.Saso, S. Iannizzotto, F. De Mattia, F.Vicinanza. "Data Grids for Conservation of Cultural Inheritance", 1st International Workshop on Data Grids for e-Science (DaGreS09) at ACM International Conference on Computing Frontiers, May 18-20, 2009 (http:// www.computingfrontiers.org/2009/)
 - https://glibrary.ct.infn.it/m/DaGRes-editor.pdf
 - A. Calanducci, C. Cherubino, L. N. Ciuffo, D. Scardaci, "A Digital Library Management System for the Grid", Fourth International Workshop on Emerging Technologies for Next-generation GRID (ETNGRID 2007) at 16th IEEE International Workshops on Enabling Technologies: Infrastructures for Collaborative Enterprises (WETICE-2007), GET/INT Paris, France, June 18-20, 2007 (http://etngrid.diit.unict.it/2007/ wex.htmi
 - https://glibrary.ct.infn.it/glibrary/downloads/gLibrary_paper_v2.pdf

