



Secure Cloud Data Storage for Galaxy Workflows

Carsten Friedrich, Catherine Wise, Richard Sinnott

March 2014

DIGITAL PRODUCTIVITY AND SERVICES FLAGSHIP
www.csiro.au

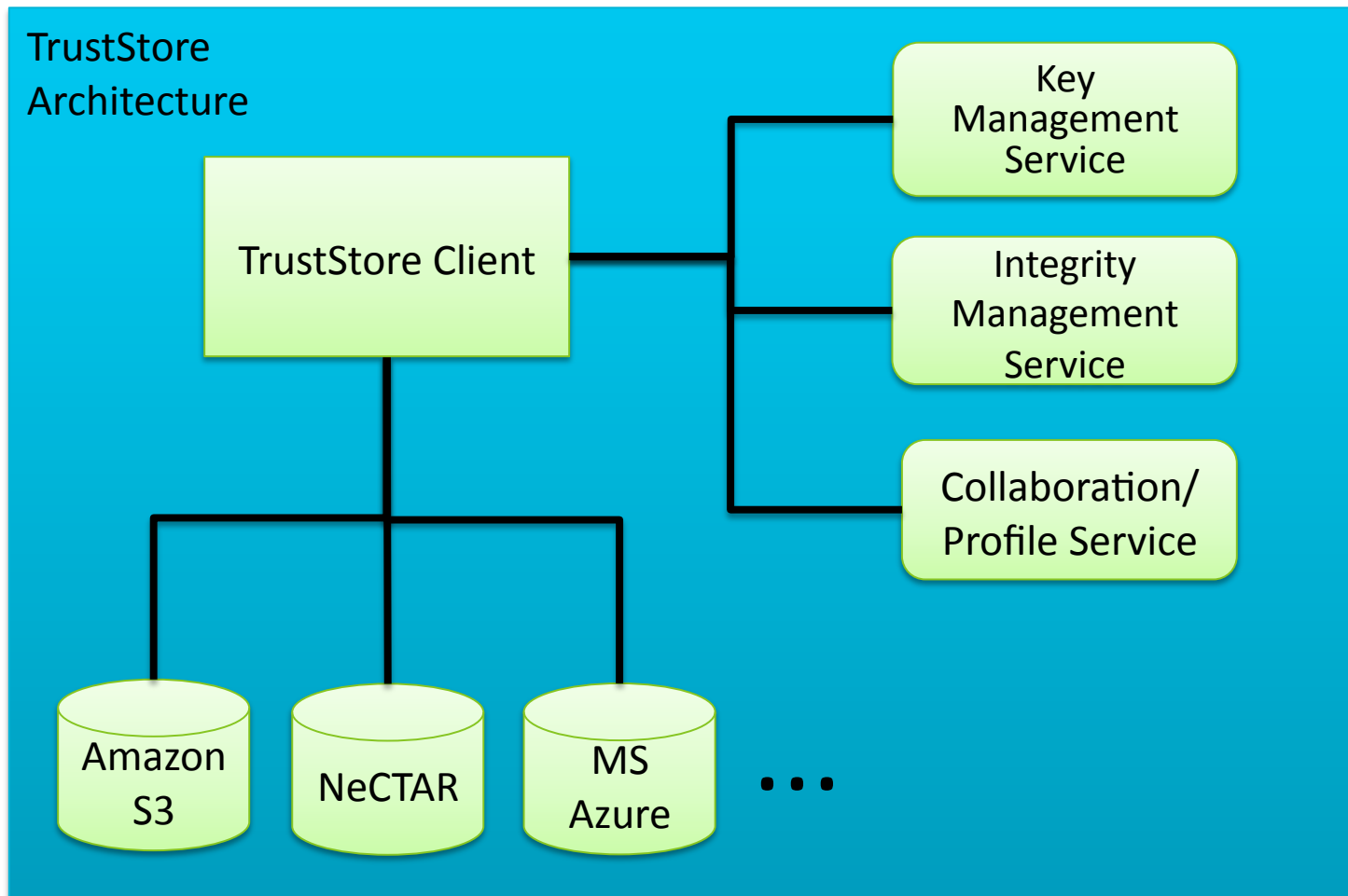


The Problem

- Galaxy workflows often deal with sensitive data
- It is attractive to execute workflows in the cloud
- Clouds are usually not highly trusted

How can we secure sensitive data in untrusted environments for processing in Galaxy?

TrustStore – Secure Data Storage in the Cloud



TrustStore – Store Manager



Name	Owner
Patient X	carsten@csiro
ISDN1	carsten@csiro
galaxy-test	carsten@csiro
galaxy-store2	galaxy-user
test-uploads	galaxy-user
galaxy-store	galaxy-user

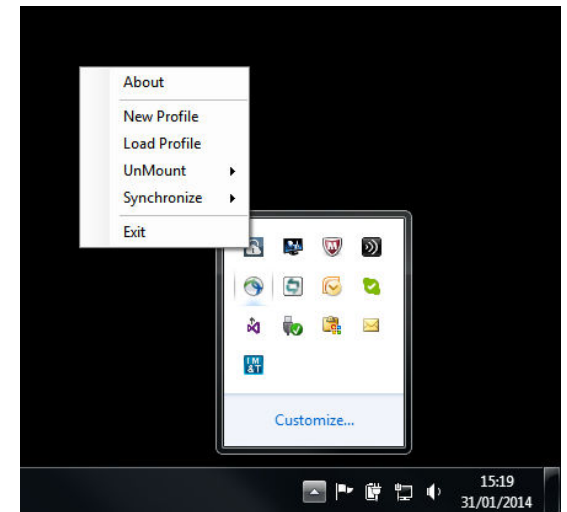
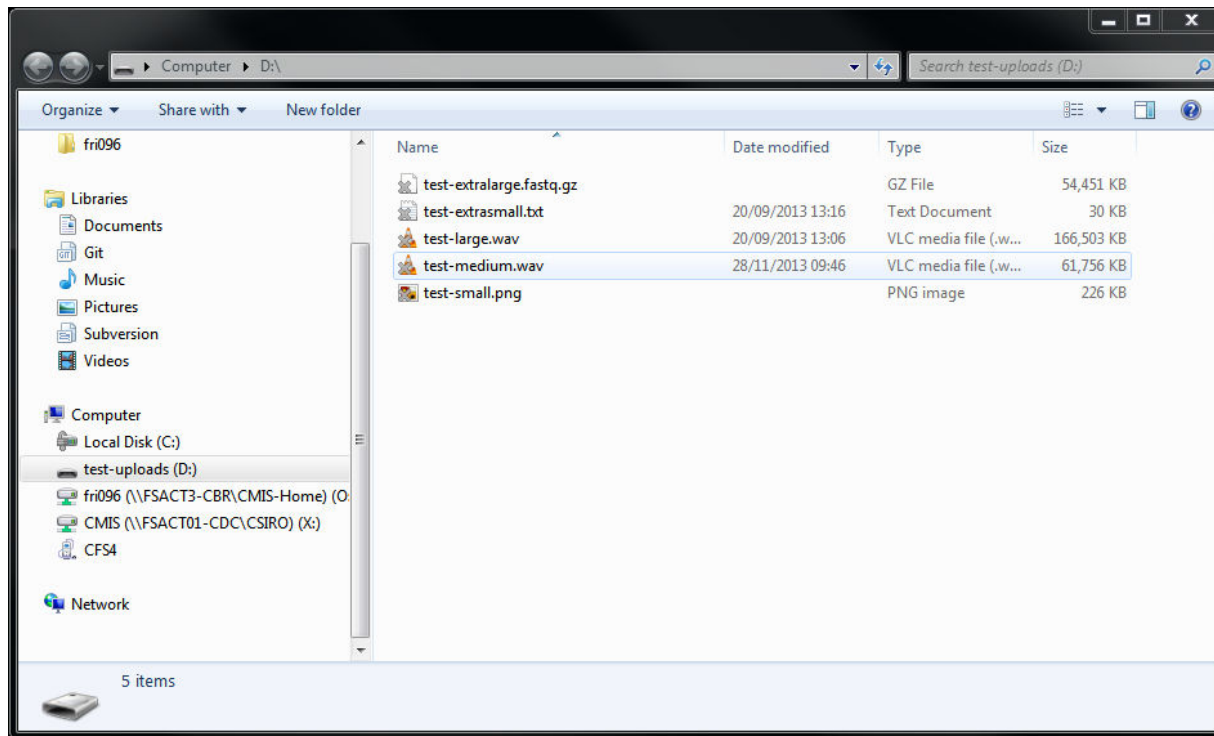
TrustStore – Uploading files



The screenshot shows the TrustStore application window. On the left, there is a 'Virtual Directory' pane with 'Docs' and 'test' folders. The main area displays a table of files with their upload progress.

File	Size	Last Modified	Expiry DateTime	Status	Current Action	Progress
ALA Harvester Architecture.docx	284 KB	08/02/2011 ...	Never	local	Creating integrity signatures	91%
ALA Open Issues.docx	13 KB	08/02/2011 ...	Never	remote		
ALA Open Issues_hw.docx	15 KB	08/02/2011 ...	Never	remote		
ALA Q3.doc	26 KB	08/02/2011 ...	Never	remote		
APA Objectives 2009_2010 (2) (3).doc	97 KB	08/02/2011 ...	Never	remote		
Areas of Work.docx	17 KB	08/02/2011 ...	Never	remote		
cf APA Objectives 2009_2010.docx	72 KB	08/02/2011 ...	Never	remote		
DIAS-B - SOW for ICT Centre 2009 10...	59 KB	08/02/2011 ...	Never	remote		
e00008.txt	1,089 KB	08/02/2011 ...	Never	local	Creating integrity signatures	84%
Glossary of ALA DIAS-B Relationship T...	110 KB	08/02/2011 ...	Never	remote		
IE Lab review agenda - Feb2010 (2).doc	32 KB	08/02/2011 ...	Never	remote		
ingest_workflow_refactor.vdx	93 KB	08/02/2011 ...	Never	local	uploading	0%
Installing Gsearch v 1.1.doc	33 KB	08/02/2011 ...	Never	local	Creating Keys	70%
Lab Review Notes 8 22.docx	21 KB	08/02/2011 ...	Never	local	uploading	0%
Lab Review Notes 8.docx	17 KB	08/02/2011 ...	Never	local	uploading	0%
NeAT 6 Monthly Report 2009-12.doc	51 KB	08/02/2011 ...	Never	local	Creating integrity signatures	80%
NeAT Quarterly Staff Review Report - ...	109 KB	08/02/2011 ...	Never	local	Creating integrity signatures	87%
ProjectPlan.doc	45 KB	08/02/2011 ...	Never	local	Creating integrity signatures	87%
Requirements for Self Provisioning of ...	842 KB	08/02/2011 ...	Never	local	Uploading fragments to SAS Pr...	41%
SOSy Theme1183-ThemeStatement_...	130 KB	08/02/2011 ...	Never	local	Creating Keys	70%
test	3				uploading	0%

TrustStore – Windows Client



TrustStore – Changing Store Access



Edit Store

Name: galaxy-store2

Owner: carsten@csiro

Writers:

- alonie@unimelb.edu.au
- simon.gladman@monash.edu
- rust@agrif.org.au
- rsinnott@unimelb.edu.au
- charlotte.anderson@unimelb.edu.au
- galaxy-user
- glenn.jayaputera@unimelb.edu.au
- carsten@csiro

Readers:

- alonie@unimelb.edu.au
- simon.gladman@monash.edu
- rust@agrif.org.au
- rsinnott@unimelb.edu.au
- charlotte.anderson@unimelb.edu.au
- glenn.jayaputera@unimelb.edu.au
- galaxy-user
- carsten@csiro

Save Cancel

TrustStore – Secure Data Storage For Galaxy Workflows

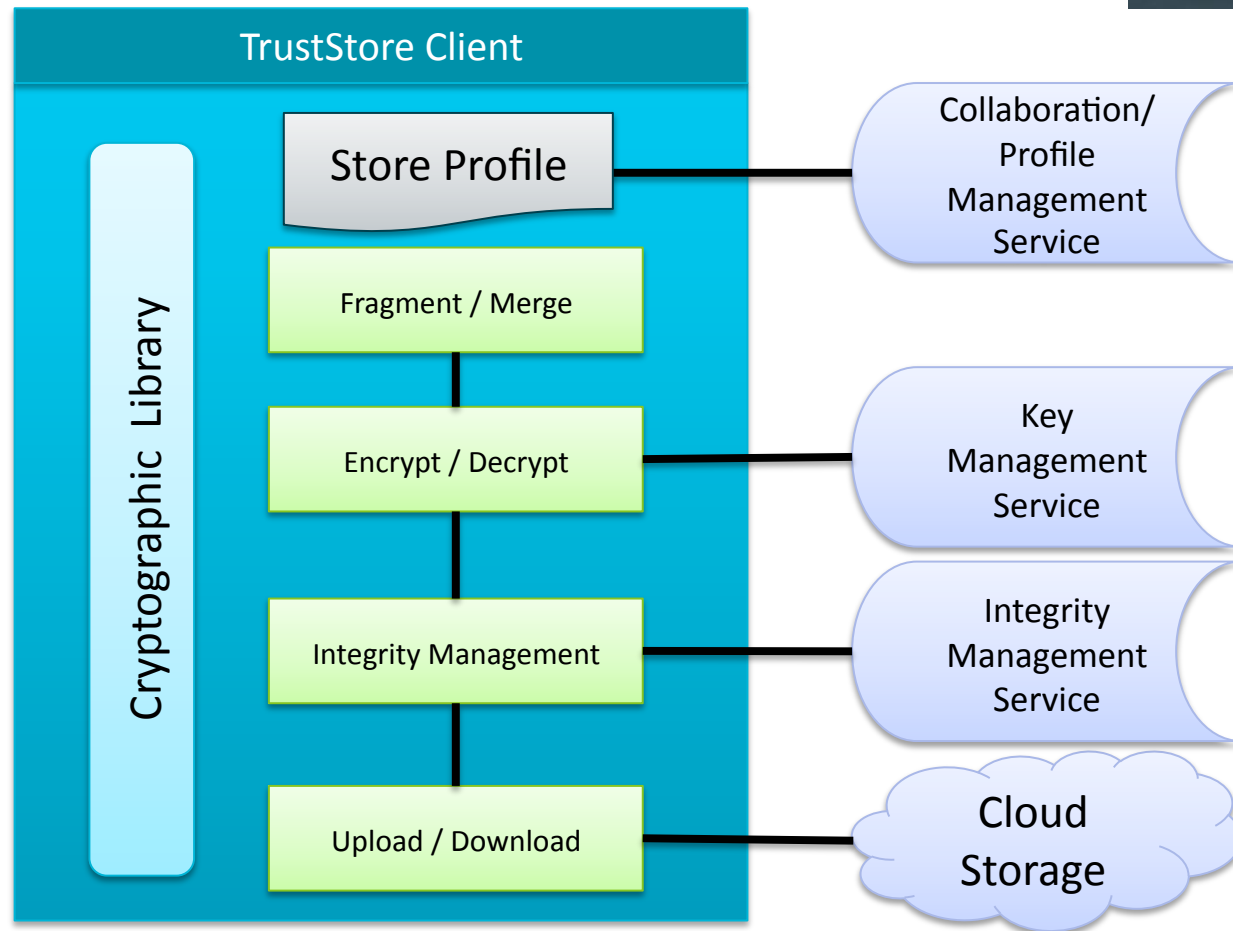


The screenshot displays the Galaxy/GVL web interface in a Firefox browser. The main content area is titled "Import Data from TrustStore (version 1.0.0)". It contains several input fields for configuration:

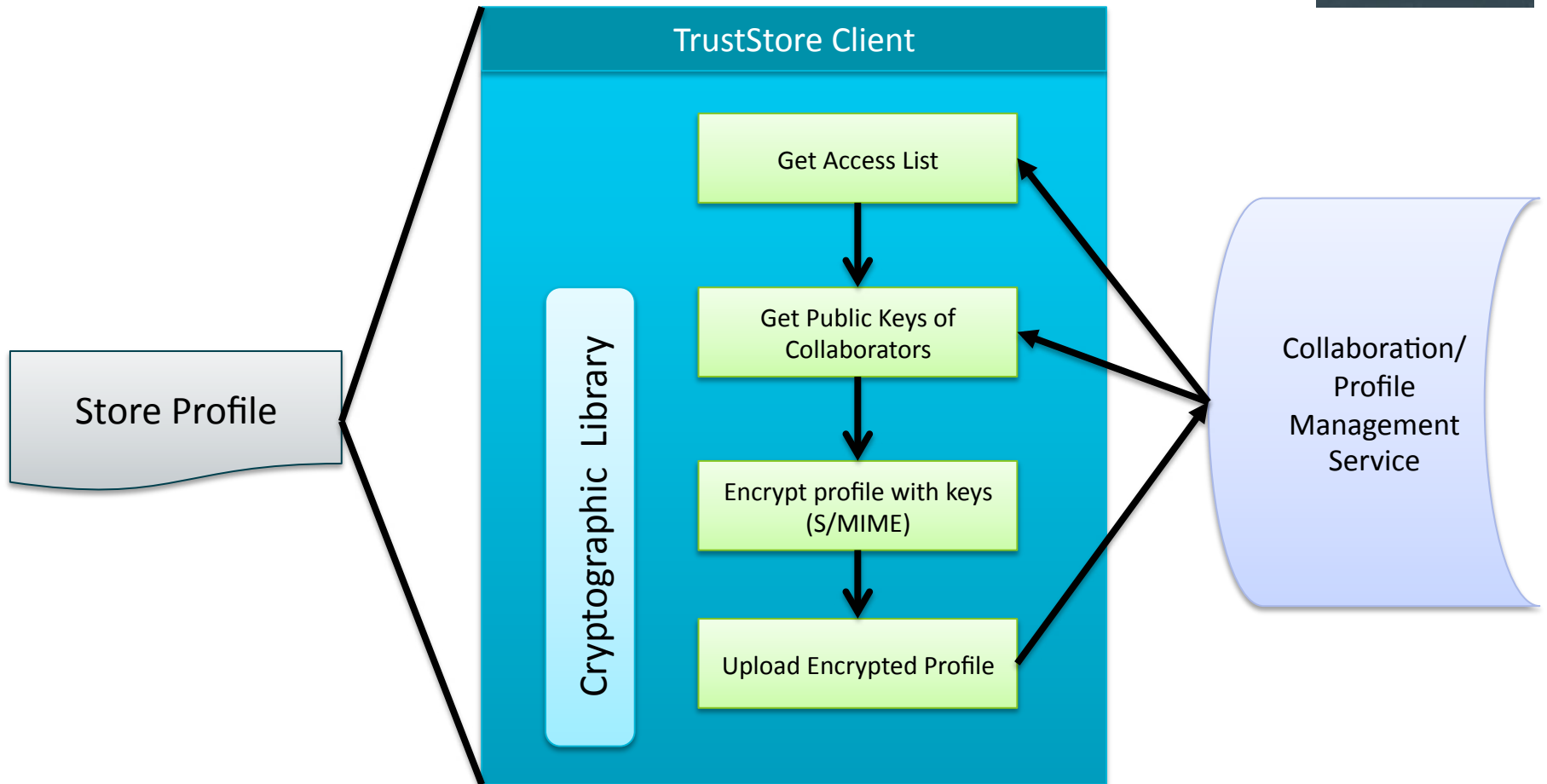
- File Format:** A dropdown menu with "fastq" selected.
- Username:** A text input field containing "user".
- Password:** A text input field containing "pas".
- Store Name:** A text input field containing "galaxy".
- Path in Store:** A text input field containing "/".
- Filename(s):** An empty text input field.

Below the input fields is an "Execute" button. A message below the button states: "This tool lets users import data they have uploaded to the TrustStore system. Upload files to TrustStore using the desktop client." The left sidebar shows a "Tools" section with a search bar and a list of tool categories including "Get Data", "ENCODE Tools", "Text Manipulation", "Filter and Sort", "Join, Subtract and Group", "Convert Formats", "Operate on Genomic Intervals", "Statistics", "Graph/Display Data", "FASTA manipulation", "NGS: QC and manipulation", "NGS: Assembly", "NGS: Indel Analysis", "NGS: RNA Analysis", "NGS: SAM Tools", "NGS: GATK Tools", "NGS: Peak Calling", "ENSEMBL", "NGS: Variant Detection", "NGS: Mapping", and "BED tools". The right sidebar shows a "History" section with "Unnamed history" and "0 bytes".

TrustStore – Client Data Workflow



TrustStore – Collaboration Workflow



Lessons learned

It does work. It does the job. There is room for improvement:

- Galaxy can't keep a secret (It's not even trying):
 - Can't hide passwords in forms.
 - Writes passwords into logs / databases.
 - No streaming support.
 - Can't protect temporary storage.
- Writing Galaxy modules is tricky:
 - Sparse documentation.
 - Sparse error messages.
 - Very fickle library dependency management.
- Implementing as an external data source provider may have been better in hindsight.

Thank you

Carsten Friedrich

Research Team Leader

IE Lab, CSIRO Computational Informatics

Carsten.Friedrich@csiro.au

+61 2 6126 7019

www.csiro.au

