

# caArray

# Microarray data management system

caArray is an open-source, role-based, Web- and programmatically-accessible array data management system. caArray enables the annotation and exchange of array data using a federated model of local installations whose data are shareable across the cancer Biomedical Informatics Grid (caBIG<sup>®</sup>).

As a grid data service, caArray furthers translational cancer research through acquisition, aggregation and dissemination of semantically interoperable array data to support subsequent analysis by tools and services on and off the Grid. As array technology advances and matures, caArray will extend its library of assay management capabilities.

## Features

- Use Web-based forms to submit and annotate data consistent with MIAME guidelines
- Import multiple MAGE-TAB files
- Publish data to the public domain using group-based permission schemes
- Submit and retrieve Affymetrix, GenePix, and Illumina native expression array files
- Submit and retrieve Affymetrix and Illumina native SNP array files



U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES

National Institutes of Health

Biospecimens	Data Sharing	Imaging	Proteomics
Clinical Trials Management	Genome Annotation	Microarrays	Translational Research
	Infrastructure	Pathways	Vocabularies
<ul><li>Data Analysis</li><li>&amp; Statistical Tools</li></ul>			



# **Architecture Overview**

- **Application type:** Web application–remote application with data uploads/ downloads through Web interface
- **System requirements:** While caArray is a Web-based application, it does have a dependent software stack that includes MySQL database, a JBoss middle tier server, optional LDAP server for non-database authentication, and network infrastructure (LAN)

# Installation and Administration

- **Skill sets needed:** System administration and full access rights to the deployment environment; these may include Unix root access as well as database administrator access
- **Infrastructure needed:** The installation checks for existing environment dependencies and either respects them if they exist or acquires the corresponding versions of JBoss and MySQL
- Long-term administration needs: Updates will be provided through GUI installers

### Resources

Tool Overview Page	https://cabig.nci.nih.gov/tools/caArray
Primary Workspace	Integrative Cancer Research (ICR) https://cabig.nci.nih.gov/workspaces/ICR
Knowledge Center	https://cabig-kc.nci.nih.gov/Molecular/KC/ index.php/Main_Page_
caArray Users LISTSERV	https://list.nih.gov/archives/caarray_users-l.html
caBIG <sup>®</sup> Tool Inventory	https://cabig.nci.nih.gov/inventory
NCI Center for Bioinformatics Applications Support	ncicb@pop.nci.nih.gov



### **Key Contributors:**

- 5AM Solutions
- Columbia University
- Jackson Labs
- Lawrence Berkeley
  National Laboratory
- NARTec
- NCI Center for Bioinformatics and Information Technology
- Science Applications International Corporation
- TerpSys
- Washington University

#### Other Life Sciences Distribution Components:

- Cancer Genome-Wide Association Studies (caGWAS)
- caTissue Core
- Clinical Trials Object Data System (CTODS)
- geWorkbench
- National Cancer
  Imaging Archive (NCIA)



NIH Publication No. 09-7437 Printed May 2009

