

caArray

Microarray data management system

caArray is an open-source, role-based, Web and programmatically-accessible array data management system. caArray guides 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™).

As a grid data service, caArray furthers translational cancer research through acquisition, dissemination, and aggregation 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 logical library of assay management.

Features:

- Web-based forms for researcher submission and annotation of data consistent with MIAME guidelines
- Bulk data import of MAGE-TAB files
- Group-based permission scheme including settings for publishing data to the public domain
- Search and navigate features to readily discover and extract data of interest
- Submission and retrieval of Affymetrix, GenePix, and Illumina native expression array files
- Submission and retrieval of Affymetrix and Illumina native SNP array files



The screenshot shows the caArray web interface. At the top, it displays the National Cancer Institute logo and the text 'U.S. National Institutes of Health | www.cancer.gov'. The main header includes 'caArray Array Data Management System'. A 'Welcome to the caArray Data Portal' message is visible. On the right, there is a 'caArray Login' form with fields for 'Username: carrayadmin' and 'Password: password', and a 'Login' button. Below the login form, there are links for 'Register' and 'Forgot Password?'. In the center, there is a 'Browse caArray' section with a dropdown menu for 'Location' set to '(All Locations)'. Below this, a table lists various categories and their counts:

Experiments	3
Observations	3
Array Providers	1
Unique Array Designs	3
Hybridizations	1
Registered Institutions	0
Registered Users	14

To the right of the table is a 'Search caArray' section with fields for 'Keyword:', 'Category: (All Categories)', and 'Location: (All Locations)', along with a 'Search' button.

caArray interface

Categories of Use:

- | | | | |
|--|---|--|---|
| <input type="checkbox"/> Biospecimens | <input checked="" type="checkbox"/> Data Sharing | <input type="checkbox"/> Imaging | <input type="checkbox"/> Proteomics |
| <input type="checkbox"/> Clinical Trials Management | <input type="checkbox"/> Genome Annotation | <input checked="" type="checkbox"/> Microarrays | <input type="checkbox"/> Translational Research |
| <input type="checkbox"/> Data Analysis & Statistical Tools | <input type="checkbox"/> Infrastructure | <input type="checkbox"/> Pathways | <input type="checkbox"/> Vocabularies |



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

Key Contributors:

- 5AM Solutions
- Columbia University
- Jackson Labs
- Lawrence Berkeley National Laboratory
- NARTec
- NCI Center for Bioinformatics
- 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)

Resources

Tool Overview Page	https://cabig.nci.nih.gov/tools/caArray
Primary Workspace	Integrative Cancer Research (ICR)
ICR Workspace Coordinator	Elaine Freund: efreund@3rdmill.com
caArray Users LISTSERV	https://list.nih.gov/archives/caarray_users-l.html
caArray Developers LISTSERV	https://list.nih.gov/archives/caarray_developers-l.html
caBIG™ Tool Inventory	https://cabig.nci.nih.gov/inventory
NCI Center for Bioinformatics Applications Support	ncicb@pop.nci.nih.gov

