

caArray data portal was designed based on the international data standard for microarray data, MAGE-OM. caArray includes web-based data annotation forms that capture MIAME 1.1 level annotations using controlled terminology from MGED ontology. These annotations include information about contacts, protocols, biomaterials, experiments, arrays and array designs. caArray 1.0 supports submission of Affymetrix and GenePix native data files, as well as MAGE-ML import and export. In addition to document based data submission, caArray provides application programming interfaces (APIs) for programmatic access to data: MAGE-OM API can be used for fine grain data retrieval and EJB API via data transfer objects (DTO) for data transfer. End users can access the data files and annotations through the caArray data portal by downloading the native data files, or exporting MAGE-ML. caArray common data elements (CDEs) are stored in the NCICB's shared, publicly accessible metadata repository, caDSR. Currently MGED ontology terms are stored in the caArray database. A connection to the caCore's Enterprise Vocabulary (EVS) services will be available later this year. Working together with caArray database are two open source microarray data analysis tools: *caWorkbench*, a desktop tool for analysis, annotation and visualization of microarray data and *webCGH*, an application that allows users to view DNA copy number measurements relative to genome locations and annotated genome features. Both of these tools also connect to the NCICB's cancer Bioinformatics Infrastructure Objects (caBIO) model, permitting access to a variety of genomic, cancer models, and clinical trials information. Additional tools that retrieve data from the caArray via the MAGE-OM API are being developed by several NCI -designated cancer centers funded by the cancer biomedical informatics grid (caBIG) program.

The caArray database and analysis tools were developed to be consistent with caBIG compatibility guidelines that highlight the use of controlled vocabularies, CDEs, well documented APIs and UML models. caBIG is a new initiative coordinated by NCI in partnership with other members of the cancer research community. caBIG seeks to create a network that links organizations, institutions, and individuals to enable the sharing of cancer research infrastructure, data, and interoperable tools. It is an open-access, open-source activity that promises to expedite progress in cancer research. caArray's compatibility with the caBIG design requirements facilitate the cross silo use of cancer biology information to promote integrated cancer research.

caArray has a n-tier architecture and is built with future extendibility in mind. It utilizes J2EE framework and provides programmatic interfaces as well as a web portal interface for submission and retrieval of microarray data. EJB interface provides transactional API capability with the use of Data Transfer Objects (DTO s) used to transfer actual data. The Web portal built utilizing Struts framework uses the EJB and corresponding DTOs to perform transactions with the backend. An RMI based query api based on MAGE-OM provides fine grain programmatic access to the persisted data. Internally, caArray utilizes an object relational mapping (ORM) tool called OBJ to abstract the Java Messaging actual data source from the application and provide object-based access to data. Netcdf, a binary file format with open API is utilized for storage and fast retrieval of data. This file format stores data in cube matrix format and can be queried on its dimensions. It allows for faster query and retrieval of data compared to database or text files. The application utilizes Service (JMS) for parsing large data files asynchronously.

caArray utilizes JAAS for authentication and authorization. The implementation of the role based access control uses the common NCICB security schema. This configurable security architecture allows for LDAP or RDBMS based authentication and has a concept of groups which allows for sharing of data amongst a consortium of researchers. MAGE-OM also utilizes a common security service to filter objects based on user roles and permission. This implementation is provided via Aspect Oriented Programming (AOP). The Solaris production environment is listed here, information about other configurations that we have tested and verified can be found at <http://caarray.nci.nih.gov/caARRAY/devdoc/caarraydbdocs>.

	DBMS	Application Server
	Sunfire 1280	SunFire 480R
CPU	4 x 900 MHZ (UltraSPARC III)	2 x 900 MHZ (UltraSPARC III)
Memory	8 GB	10 GB
Local Disk	36 GB (Mirrored)	36 GB (Mirrored)
Network Link Speed	Fiber	100 MB (Switched)
OS	Sun Solaris 5.8	Sun Solaris 5.8
Comments	Shared with other NCICB databases. DB: Oracle 8i	App server: JBoss 3.2.3

caArray datasets and open source tools are publicly available, and can be accessed at <http://caArray.nci.nih.gov>; caArray source code is available for local installations at <http://ncicb.nci.nih.gov/download> under an open source license. For more information about caArray database and tools, please contact the NCICB application support ncicb@pop.nci.nih.gov.