Who: The National Cancer Institute
What: The Cancer Biomedical Informatics Grid™ Program (caBIG™)
Goal: Eliminate suffering and death due to cancer by 2015

The National Cancer Institute (NCI) has a lofty goal—to eliminate suffering and death due to cancer by 2015. There are many researchers running disparate projects across the U.S. attempting to solve some small piece of the cancer puzzle. They may not know about or have access to results from other scientists that could be instrumental to solving their particular problem. The NCI decided that it was imperative to bring these scientists together – to combine and leverage their expertise and findings to achieve even greater results. To help them meet this goal, they created the Cancer Biomedical Informatics Grid (caBIG) program.

The strategy the NCI employed was to create a scalable, actively managed organization connected via a biomedical informatics network. They knew they needed to connect domain specific workspaces for individual projects with both syntactic and semantic interoperability. In essence, each researcher would need to have the ability to access and use parts and equipment on another system as if it were in their own environment. To solve the interoperability challenge, NCI turned to the Object Management Group's standard Model-Driven Architecture® (MDA®).

NCI chose to use MDA because it is a set of standards that have worked well in other areas, and they didn't want to have to "reinvent the wheel." The initial interoperability project involved three steps:

1. Analyze what was needed and develop use cases
2. Use UML® to standardize model representations and artifacts, often using class and sequence diagrams
3. Use meta-models to generate code

The advantage to this approach is that MDA allowed NCI to use available tools to automatically generate some of the code, while giving them the flexibility to tailor it to their specific needs. In the future, when the system's requirements change, they can update their models and quickly regenerate the appropriate code. The NCI extended the MDA paradigm to include registered metadata and controlled terminologies. The use of metadata was especially important to eliminate ambiguity in the definitions assigned to particular data classes, attributes, and values. NCI brought together the MDA-generated code with the metadata registry and terminology components to create the caCORE Software Development Kit (SDK; http://ncicb.nci.nih.gov/NCICB/infrastructure/cacoresdk). Using the caCORE SDK, an organization can achieve a "Silver level" of interoperability, as defined by the caBIG Compatibility Guidelines (https://cabig.nci.nih.gov/guidelines_documentation). Silver systems implement object-oriented APIs derived from a UML domain information model, and thus examples of service-oriented architecture (SOA). The next level of compatibility, called "Gold", adds a grid services infrastructure called caGrid. caGrid supports the advertising, discovery, and invocation use cases, and once silver systems are properly registered with caGrid they become Gold.

The NCI caBIG project is entering its third year and has test beds up and available online. The project is open and free for researchers to participate in. NCI application support personnel are available to answer questions about any caBIG system. NCI has been successful by using a flexible approach, knowing that one size would not fit all situations. Using MDA, NCI was able to achieve the interoperability it needed to support the researchers working diligently to find the cure for cancer by 2015.