

# caBIG dotNET and the xl-caBIG Smart Client

Browsing **caBIG** Data using Microsoft Excel

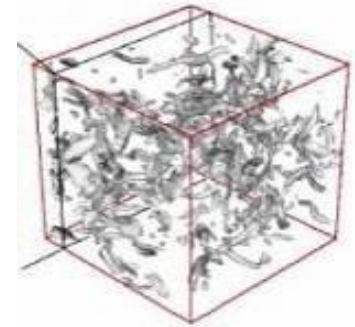
*Tom Macura*  
*University of Cambridge, UK*



# Science Paradigms by Jim Gray

- Thousand years ago:  
science was **empirical**  
describing natural phenomena
- Last few hundred years:  
**theoretical** branch  
using models, generalizations
- Last few decades:  
a **computational** branch  
simulating complex phenomena
- Today:  
**data exploration** (eScience)  
unify theory, experiment, and simulation  
using data management and statistics
  - Data captured by instruments  
Or generated by simulator
  - Processed by software
  - Scientist analyzes database / files

$$\frac{\dot{a}^2}{a} = \frac{4G}{3} \frac{c^2}{a^2}$$



# World Wide Web and eScience

---

- eScience has been inspired and evolved from the World Wide Web
- e.g. **caBIG** (*cancer architecture Biomedical Informatics Grid*) is billed as the “*World Wide Web of Cancer Research*”
- A huge variety of data types in different presentation formats are on the World Wide Web, including:
  - PubMed
  - eBay
  - Nytimes
  - Amazon.com
  - Wikipedia
  - Maps/YellowPages
  - Banking
- All this data is accessible through a single container application: *the Web Browser* which is designed with a variety built-in/plug-in technologies (html, css, jpg/mpeg, JavaScript, Java, Flash, AJAX, XML, PDF, PPT etc.)
- For the computer novice, the Browser is the Internet!

# *eScience Browser*

---

- Is there a common “eScience Browser” for this new “WWW”?
  - No; eScience projects are each developing their own set of specialized tools geared towards limited tasks
- Do we need a common eScience Browser ?
  - It would be beneficial. An eScience Browser would provide a standard interface that is easier for users to learn, and provide generic access to eScience data without additional work from eScience developers.
- The eScience Browser would complement the eScience specialized interfaces
- What would an eScience Browser be like ?

- What would an eScience Browser be like ?

*The common denominator for eScience results are tables. So a eScience Browser would, essentially, be a spreadsheet program with visualization and statistics tools.*

*Microsoft Excel* is the future “web-browser” of eScience

... okay, that’s a nice idea. Can it be realized?

*We developed a prototype, based on [caBIG](#)*



# for Cancer Research

- **caBIG** is a \$60 Million (3yrs) project that will give scientists access to orders of magnitude more related data from researchers around the world.
- What tools do scientists need to browse, query, and analyze this data and make meaningful deductions?
  - Using **Microsoft Excel** and leveraging their familiarity with it's statistical analysis and visualization.
  - Using **Windows Mobile 5.0** to gain on-the-spot hypothesis testing; thus replacing the proverbial "restaurant napkin sketches"
- How could we develop such applications? Using the .NET Framework
  - Visual Studio Tools for Office
  - Compact Framework for Mobile Devices

# caBIG dotNET bridges caGRID and



- 
- caGRID API is written in Java leveraging many Java projects: Globus Toolkit 3.2, OGSA-DAI 5.0, Tomcat, Ant
  - caBIG dotNET is our project bridging the gap between caBIG and .NET
  - It is a C# Client API based on the Hessian Binary Protocol that connects with a corresponding Java Servlet to provide exposition of the caBIG API
  - BSD License; SourceForge Website and CVS Repository <http://xl-cabig-client.sourceforge.net/>

# xl-caBIG Smart Client

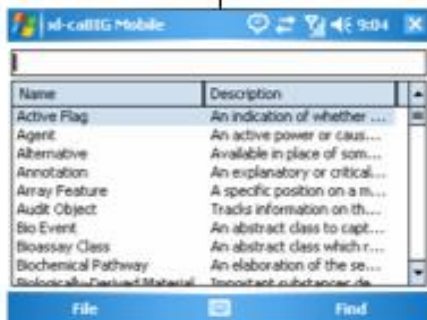


dynamic linking

## caBIG dotNET (C# Client API)



dynamic linking



# xl-caBIG Smart Client MOBILE

# HTML View (JSP/Java)



caucho Hessian Protocol

## caBIG dotNET (Java Servlet)



caucho Hessian Protocol

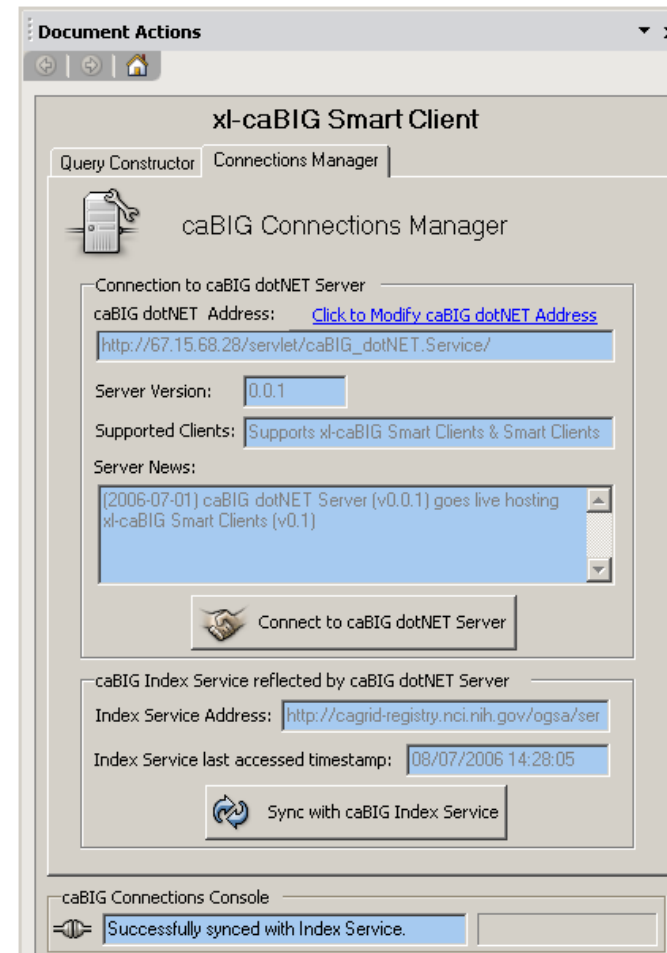


# Console Manager (Java)



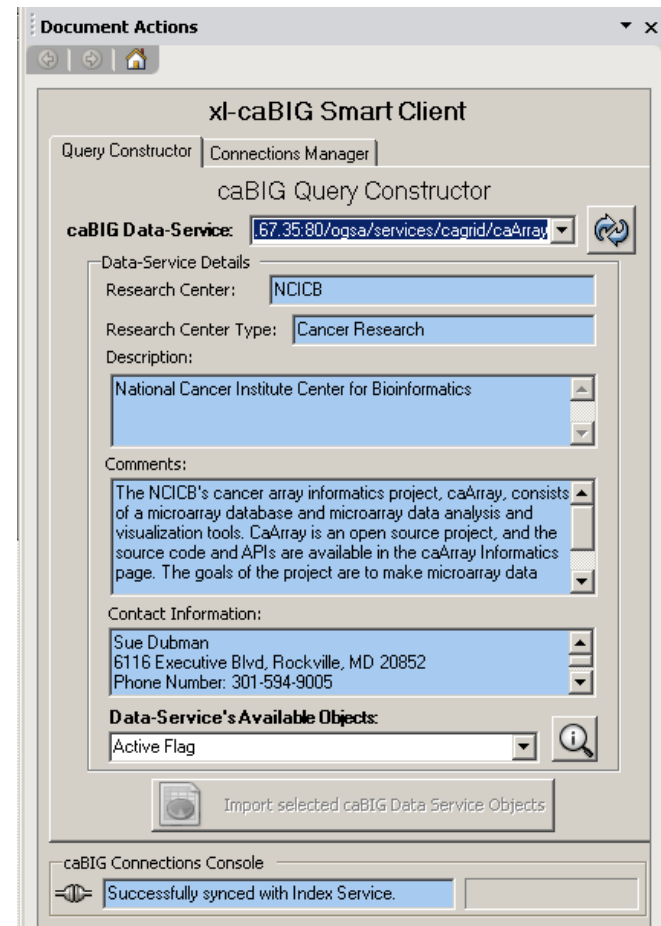
# xl-caBIG Smart Client

- Live demo



# xl-caBIG Smart Client

- Live demo



# xl-caBIG Smart Client

- Live demo

**Domain Object Details**

**Domain Object's Model Name:** Candidate Disease Gene  
**Model Description:** *Description Not Available*

Model Identifiers

**ID:** 2223317      **Short Name:** C19389  
**Class Name:** Target  
**Package Name:** gov.nih.nci.cabio.domain

Object Attributes

	Name	Description	Value Domain
▶	id	Identifier	java.lang.Long
	type	Type	java.lang.String
	name	Name	java.lang.String

EVS Concepts

	Name	Code	Definition
▶	Candidate Disease Gene	C19389	A gene proposed to hav

# xl-caBIG Smart Client

## Planned Features

- The Smart-Client's Document Actions Pane will allow users to graphically compose a query relating multiple Domain Objects.

The image displays three screenshots of the 'Domain Object Details' window, illustrating the planned features for the Smart-Client. Each window shows the details of a specific domain object, including its model name, description, identifiers, class name, package name, and attributes.

**Window 1: Candidate Disease Gene**

- Domain Object's Model Name: Candidate Disease Gene
- Model Description: Description Not Available
- Model Identifiers: ID: 2223317, Short Name: C19389
- Class Name: Target
- Package Name: gov.nih.nci.cabio.domain
- Object Attributes:

Name	Description	Value Domain
id	Identifier	java.lang.Long
type	Type	java.lang.String
name	Name	java.lang.String
- EVS Concepts:

Name	Code	Definition
Candidate Disease Gene	C19389	A gene proposed to hav

**Window 2: Clone Gene Relative Location Object**

- Domain Object's Model Name: Clone Gene Relative Location Obj
- Model Description: Description Not Available
- Model Identifiers: ID: 2223343, Short Name: C37925.C45377
- Class Name: CloneRelativeLocation
- Package Name: gov.nih.nci.cabio.domain
- Object Attributes:

Name	Description	Value Domain
id	Identifier	java.lang.Long
type	Type	Gene Relative Location Type
- EVS Concepts:

Name	Code	Definition
Gene Relative Location Object	C45377	The location of a f
Clone	C37925	Clone; a group of s

**Window 3: Gene Relative Location Object**

- Domain Object's Model Name: Gene Relative Location Object
- Model Description: Description Not Available
- Model Identifiers: ID: 2223344, Short Name: C45377
- Class Name: GeneRelativeLocation
- Package Name: gov.nih.nci.cabio.domain
- Object Attributes:

Name	Description	Value Domain
id	Identifier	java.lang.Long
type	Type	java.lang.String
- EVS Concepts:

Name	Code	Definition
Gene Relative Location Object	C45377	The location of a f

Blue arrows indicate the flow of information from the 'id' attribute in the first window to the 'id' attribute in the second window, and from the 'id' attribute in the second window to the 'id' attribute in the third window.

# xl-caBIG Smart Client

## Planned Features

---

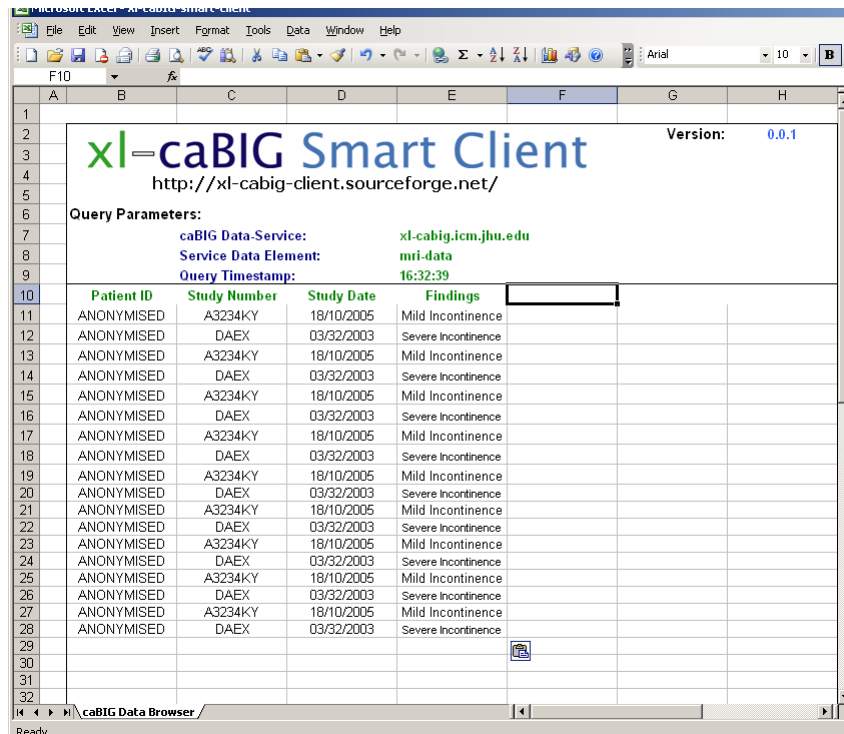
- This query will be transcribed into **caBIG** Common Query Language (XML) and executed against data-services via **caBIG dotNET**

```
<caGridXMLQuery name="caArrayQuery">  
  <criteria name="gov.nih.nci.mageom.domain.Protocol.Protocol">  
    <criteria name="identifier" condition="EQUAL_TO value="P-MEXP-1963"/>  
  </criteria>  
</caGridXMLQuery>
```

# xl-caBIG Smart Client

## Planned Features

- Results of the query (XML) will be interpreted by the Smart Client to populate cells in the Excel Workbook.



The screenshot shows a Microsoft Excel window with the xl-caBIG Smart Client interface overlaid. The interface includes a title bar, a menu bar (File, Edit, View, Insert, Format, Tools, Data, Window, Help), and a toolbar. The main content area displays the following information:

**xl-caBIG Smart Client** Version: 0.0.1  
<http://xl-cabig-client.sourceforge.net/>

**Query Parameters:**

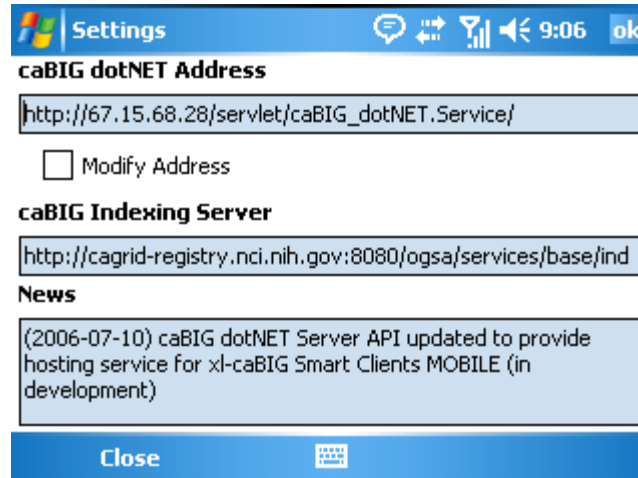
- caBIG Data-Service: xl-cabig.icm.jhu.edu
- Service Data Element: mri-data
- Query Timestamp: 16:32:39

The interface also displays a table with the following columns: Patient ID, Study Number, Study Date, and Findings. The table contains 28 rows of data, alternating between ANONYMISED patients with Study Number A3234KY and DAEX, and Study Dates 18/10/2005 and 03/32/2003. The Findings column lists Mild Incontinence and Severe Incontinence.

Patient ID	Study Number	Study Date	Findings
ANONYMISED	A3234KY	18/10/2005	Mild Incontinence
ANONYMISED	DAEX	03/32/2003	Severe Incontinence
ANONYMISED	A3234KY	18/10/2005	Mild Incontinence
ANONYMISED	DAEX	03/32/2003	Severe Incontinence
ANONYMISED	A3234KY	18/10/2005	Mild Incontinence
ANONYMISED	DAEX	03/32/2003	Severe Incontinence
ANONYMISED	A3234KY	18/10/2005	Mild Incontinence
ANONYMISED	DAEX	03/32/2003	Severe Incontinence
ANONYMISED	A3234KY	18/10/2005	Mild Incontinence
ANONYMISED	DAEX	03/32/2003	Severe Incontinence
ANONYMISED	A3234KY	18/10/2005	Mild Incontinence
ANONYMISED	DAEX	03/32/2003	Severe Incontinence
ANONYMISED	A3234KY	18/10/2005	Mild Incontinence
ANONYMISED	DAEX	03/32/2003	Severe Incontinence
ANONYMISED	A3234KY	18/10/2005	Mild Incontinence
ANONYMISED	DAEX	03/32/2003	Severe Incontinence
ANONYMISED	A3234KY	18/10/2005	Mild Incontinence
ANONYMISED	DAEX	03/32/2003	Severe Incontinence
ANONYMISED	A3234KY	18/10/2005	Mild Incontinence
ANONYMISED	DAEX	03/32/2003	Severe Incontinence

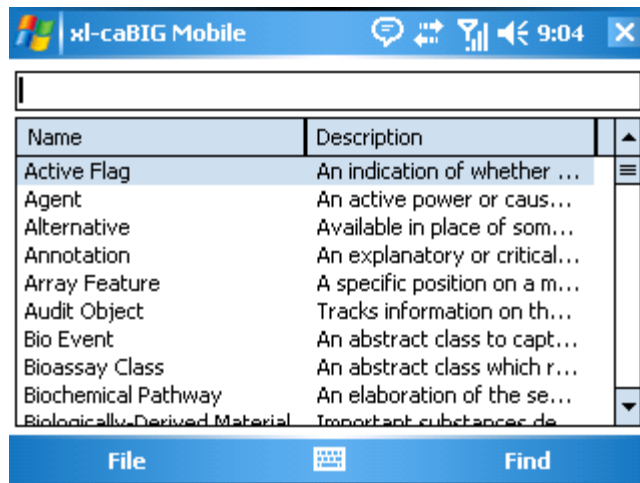
# xl-caBIG Smart Client MOBILE

- xl-caBIG MOBILE uses the Smart Phone's GPRS to connect with the caBIG dotNET service



# xl-caBIG Smart Client MOBILE

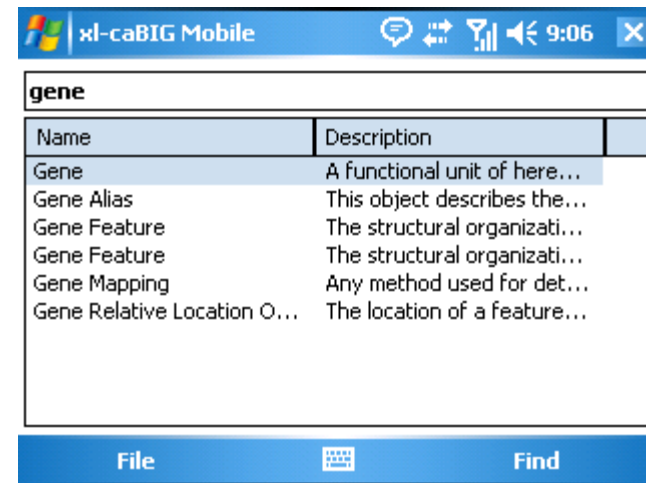
- Scientists use their **Windows 5.0** Mobile Device to browse and search for cancer concepts of interest



The screenshot shows the xl-caBIG Mobile interface. The title bar reads "xl-caBIG Mobile" and includes system icons for help, connectivity, signal strength, and time (9:04). Below the title bar is a search input field. A table with two columns, "Name" and "Description", is displayed. The table contains the following entries:

Name	Description
Active Flag	An indication of whether ...
Agent	An active power or caus...
Alternative	Available in place of som...
Annotation	An explanatory or critical...
Array Feature	A specific position on a m...
Audit Object	Tracks information on th...
Bio Event	An abstract class to capt...
Bioassay Class	An abstract class which r...
Biochemical Pathway	An elaboration of the se...
Biologically Derived Material	Important substances de...

At the bottom of the screen is a blue navigation bar with "File" and "Find" buttons.



The screenshot shows the xl-caBIG Mobile interface with search results for the term "gene". The title bar reads "xl-caBIG Mobile" and includes system icons for help, connectivity, signal strength, and time (9:06). Below the title bar is a search input field containing the text "gene". A table with two columns, "Name" and "Description", is displayed. The table contains the following entries:

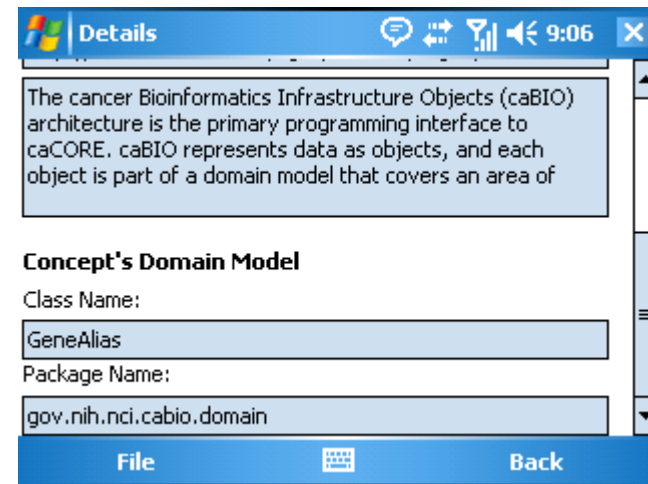
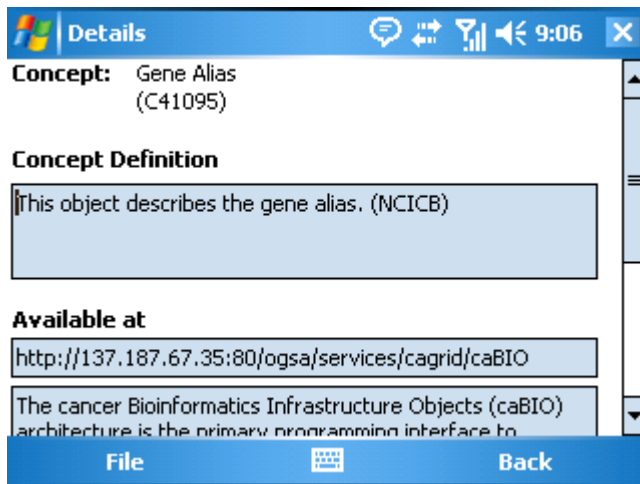
Name	Description
Gene	A functional unit of here...
Gene Alias	This object describes the...
Gene Feature	The structural organizati...
Gene Feature	The structural organizati...
Gene Mapping	Any method used for det...
Gene Relative Location O...	The location of a feature...

At the bottom of the screen is a blue navigation bar with "File" and "Find" buttons.



# xl-caBIG Smart Client MOBILE

- Once users have found a cancer concept they're interested in, **xl-caBIG MOBILE** gives further information about the concept and where on **caBIG** this concept type is available



# xl-caBIG Smart Client MOBILE

---

- xl-caBIG MOBILE running on Cingular 8125 Smart Phone



Microsoft eScience Workshop  
2006

# xl-caBIG Smart Client MOBILE

---

- xl-caBIG MOBILE is easy to use and very portable



# In conclusion what have we accomplished ?

---

- It all began with a, crazy, contrarian idea ...
  - Microsoft Excel should be the eScience Browser
- We developed a prototype as a reference implementation to demonstrate feasibility  
**xl=caBIG Smart Client**
- Are capturing mind-share that our idea is the future!

*the crazy idea ...*

# xl-caBIG vs 20+ other caBIG Tools

---

## 20+ caBIG Tools

- Free
- Open Sourced (Java)
- Developed by Academia (via 20% going to Booz Allen Hamilton)
- Funded by Government
- Domain specific, approach specific – (typically has Cancer/Bio in their names, descriptions)

## Microsoft Excel

- \$499
- Closed Source (C#/C)
- Developed by Software Engineers
- Funded by for-profit Corporation, paid by customers
- Horizontal App, used by millions in dozens of fields.

# Results of our Prototype

---

- Limited to the constraints of a beta version of caGRID, we weren't able to achieve our complete vision. Still,
- **caBIG dotNET** serves as an example of how to bridge the gap between a **JAVA**, Open-Sourced, Government project with a **.NET**, Closed-Source, For-Profit Product
- **xl-caBIG Smart Client** shows how the resulting application is easy-to-install, responsive, has a simple interface, and is useful to a wide-range of domains and approaches.

# Capturing Mind Space

- We've presented at Conferences:
  - Microsoft Faculty Summit 2006 (*Presentation*)
  - caBIG Annual Meeting 2006 (*Poster*)
  - Microsoft eScience Workshop 2006 (*Presentation*)

**Microsoft Excel caBIG Smart Clients**  
<http://xl-cabig-client.sourceforge.net/>

**Johns Hopkins University**  
 R.T. Macura<sup>1</sup>, T.J. Macura<sup>2</sup>, W.K. Macura<sup>3</sup>, K.J. Macura<sup>4</sup>

<sup>1</sup>Knowledge ME Inc; <sup>2</sup>University of Cambridge, UK; <sup>3</sup>University of Maryland, Baltimore County; <sup>4</sup>Johns Hopkins University

**Introduction:** For over a decade, Microsoft Office Excel has been the primary tool used by biomedical scientists for statistically analyzing cancer research data.

In the past, data available to scientists was limited to what was collected in their labs or made available by their collaborators. The cancer Biomedical Informatics Grid (**caBIG**) is revolutionizing past practices by empowering scientists with access to orders of magnitude more related data from researchers around the world. How will investigators synergize **caBIG** information with their own data in order to make meaningful deductions?

We are developing extensions to Excel for accessing **caBIG** data-services. Excel caBIG (xl-caBIG) smart clients will be leveraging scientists' intimate familiarity with Excel by making **caBIG** data accessible to scientists in an intuitive manner.

**xl-caBIG Smart Client Functionality:**

- xl-caBIG's GUIs allow scientists to browse and search the published Index Service in order to find caBIG data-services relevant to their research. Data-services can be filtered by data-service metadata descriptions (e.g. all data-services originating from a particular cancer research center) or based on service-type (e.g. which data-services provide data of a particular type: 'gene').
- After selecting a relevant service-type, users will be able to consume its objects as cells in their xl-caBIG smart-clients workbook. The columns will be service-type fields (e.g. 'patient identifier', 'age', 'gender') and the rows attributes (e.g. 'HU0501321', '42', 'Female').
- Since a portion of data-services are available only to authenticated and authorized users, xl-caBIG smart clients should support User Credential Management via Grid User Management Service (GUMS).



**xl-caBIG Client-Server Architecture:** caGrid 6.5, the current test architecture of caBIG, conforms to Open Grid Services Architecture (OGSA) grid infrastructure standards. caGrid leverages the Globus Toolkit by providing the required core services (e.g. caDSR and EVS), toolkits, and wizards for the development and deployment of community provided services and APIs for building client applications. The Globus Toolkit is written in Java 1.4 as are the caGrid high-level APIs and toolkits.

The xl-caBIG smart-clients are designed with a client-server architecture. We used Visual Studio Tools for Office 2003 to embed CR .NET managed code in Excel 2003 and intermediary middleware to bridge the gap between clients' .NET code and caBIG Java API hooks. xl-caBIG servers are developed in Java as thin wrappers around the caBIG high-level API.

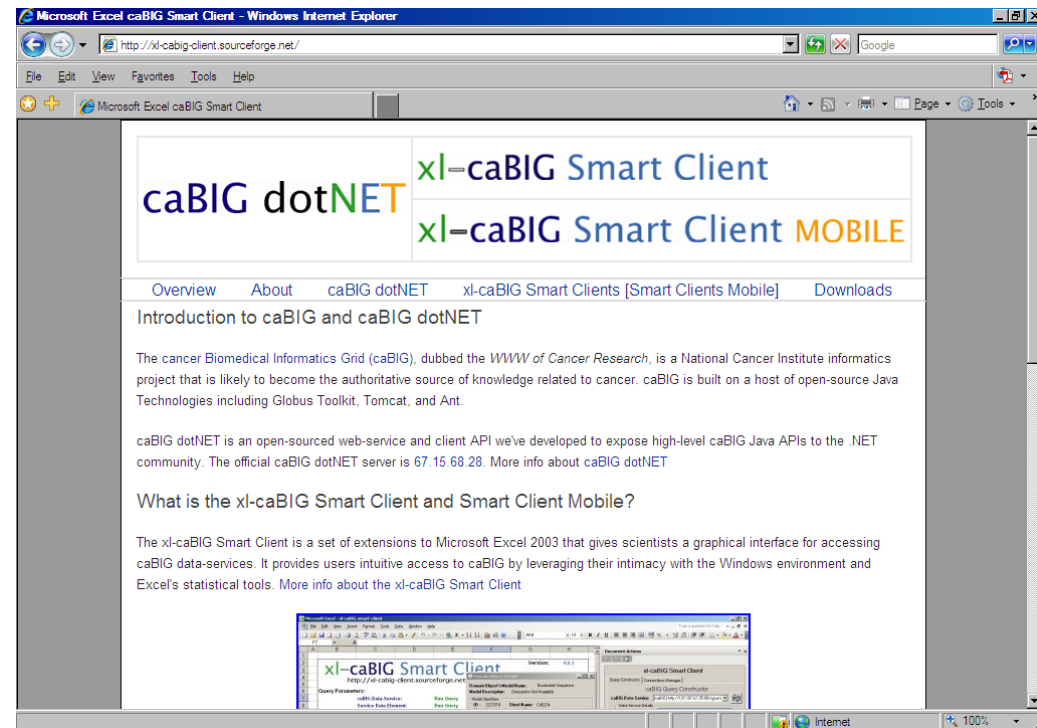
We are using the open-source middleware platform ICE (Internet Communication Engine) from ZeroC (<http://www.zeroc.com/>). ICE is suitable because the client and server can be written in different programming languages including C# and Java. Interfaces, operations, and the types of data that are exchanged between the client and server are defined using SLICE (Specification Language for ICE). The client-server contract defined in SLICE is independent of any specific programming language. SLICE definitions are compiled by ICE into an API of generated code for specific programming languages.

xl-caBIG smart-clients are leveraging .NET's features for managed deployment. Managed .NET code allows updates (in the form of DLLs) to be detected and downloaded from servers but also allows the user to exercise strong control over how the code will be executed.

xl-caBIG servers, because of their strong dependencies on caGrid 6.5 and underlying toolkits, are anticipated to be more difficult to install and configure than the smart-clients. We will be hosting official xl-caBIG servers as the default server that xl-caBIG smart-clients connect with. Advance users with special needs in terms of performance or availability can roll out their own xl-caBIG servers.

# Capturing Mind Space (cont'd)

- We've got a website and host official caBIG dotNET server: <http://xl-cabig-client.sourceforge.net/>
  - HTML pages describing the project
  - PDF copies of all our presentations/proposals
  - CVS repository/FTP server distributing the Smart Clients

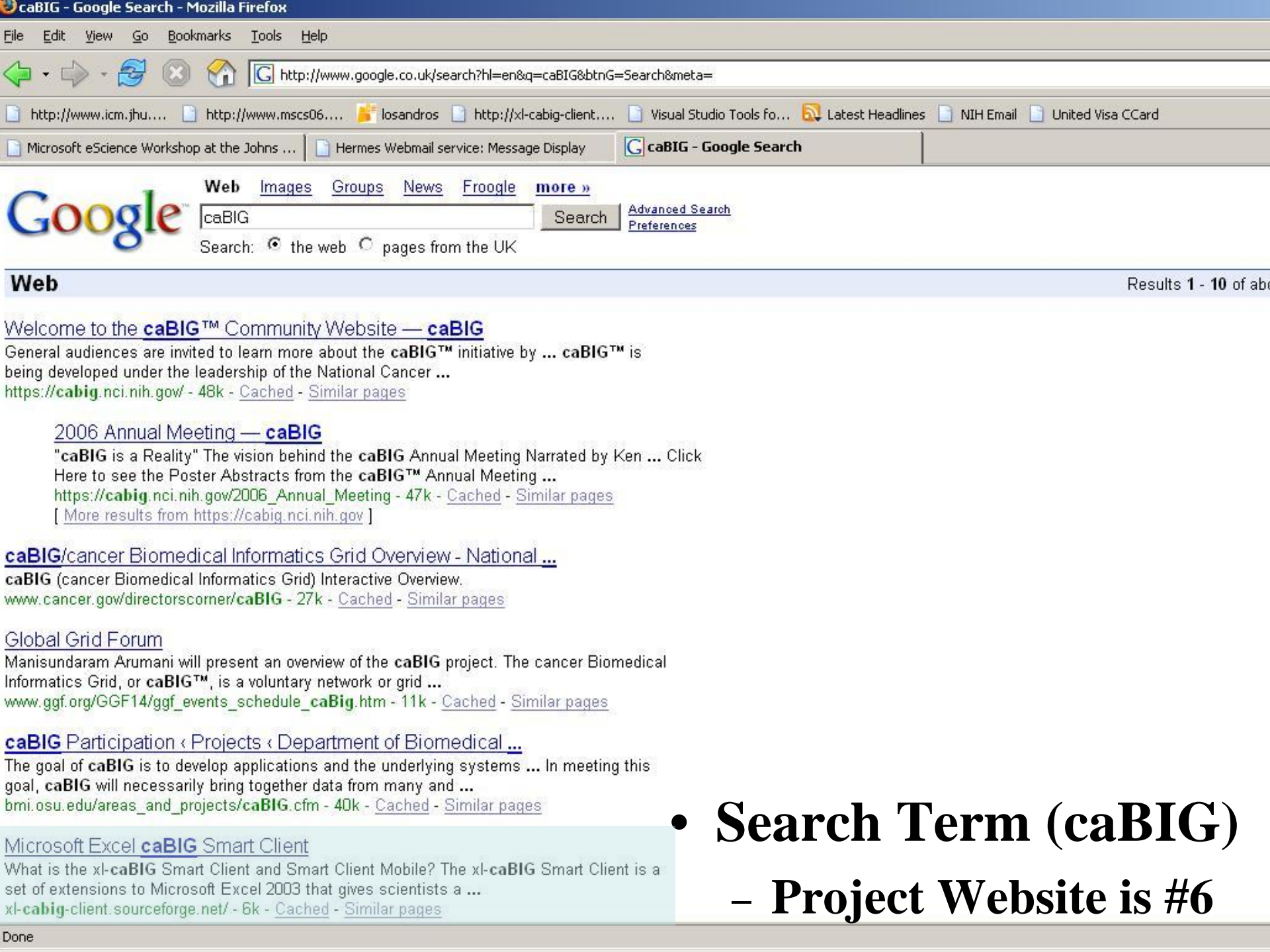




# Capturing Mind Space (cont'd)

---

- Our Website is Influential (according to Microsoft Live and Google)
  - Search Term (xl-caBIG/caBIG dotNET)
    - #1
  - Search Term (Excel caBIG)
    - #1
  - Search Term (Microsoft caBIG)
    - #1
  - Search Term (caBIG client)
    - #1
  - Search Term (caBIG)




[Web](#) [Images](#) [Groups](#) [News](#) [Froogle](#) [more »](#)  
  [Advanced Search](#)  
[Preferences](#)  
 Search:  the web  pages from the UK

**Web** Results 1 - 10 of about 100

[Welcome to the caBIG™ Community Website — caBIG](#)  
 General audiences are invited to learn more about the caBIG™ initiative by ... caBIG™ is being developed under the leadership of the National Cancer ...  
<https://cabig.nci.nih.gov/> - 48k - [Cached](#) - [Similar pages](#)

[2006 Annual Meeting — caBIG](#)  
 "caBIG is a Reality" The vision behind the caBIG Annual Meeting Narrated by Ken ... Click Here to see the Poster Abstracts from the caBIG™ Annual Meeting ...  
[https://cabig.nci.nih.gov/2006\\_Annual\\_Meeting](https://cabig.nci.nih.gov/2006_Annual_Meeting) - 47k - [Cached](#) - [Similar pages](#)  
[\[ More results from https://cabig.nci.nih.gov \]](#)

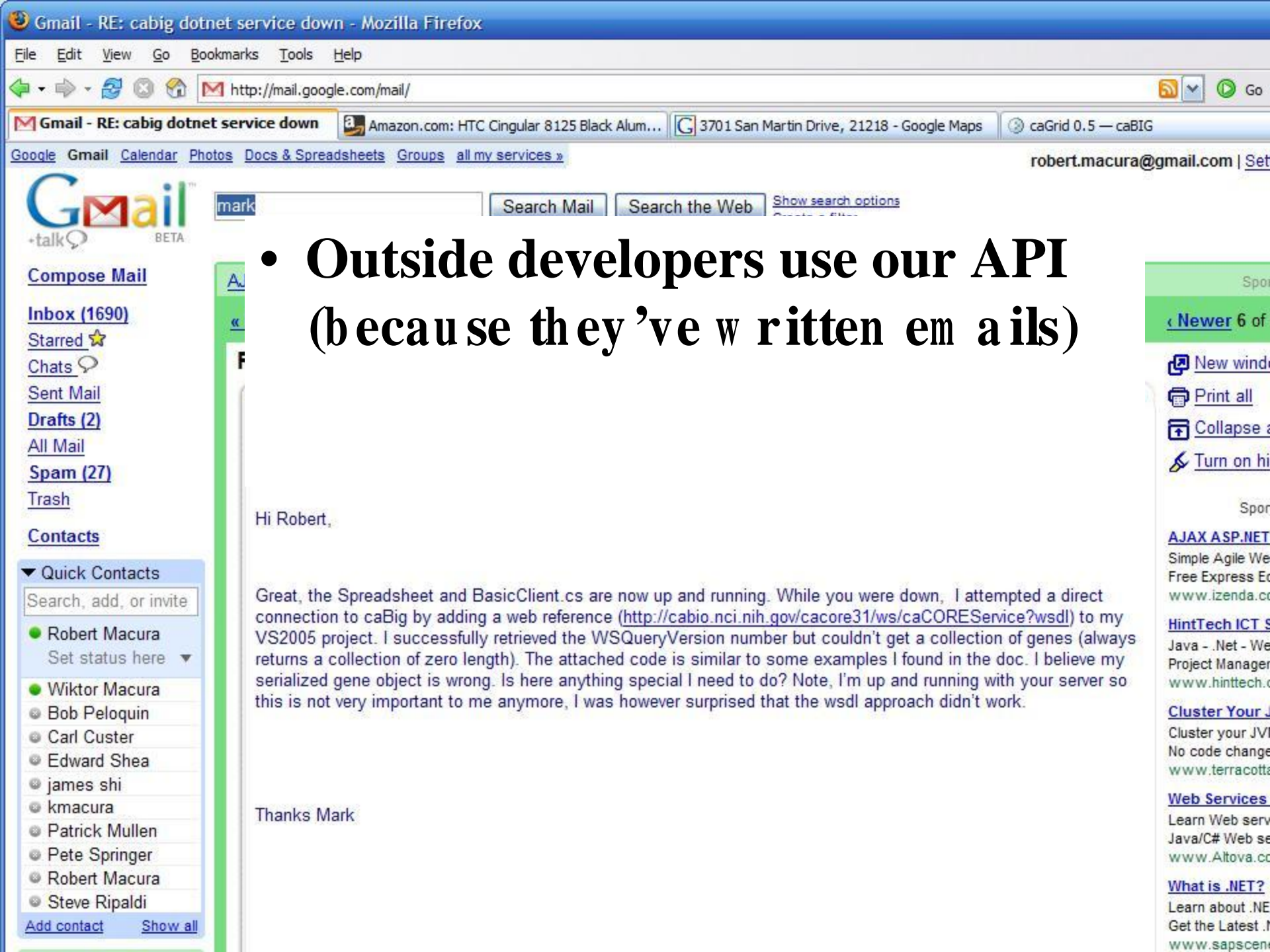
[caBIG/cancer Biomedical Informatics Grid Overview - National ...](#)  
 caBIG (cancer Biomedical Informatics Grid) Interactive Overview.  
[www.cancer.gov/directorscorner/caBIG](http://www.cancer.gov/directorscorner/caBIG) - 27k - [Cached](#) - [Similar pages](#)

[Global Grid Forum](#)  
 Manisundaram Arumani will present an overview of the caBIG project. The cancer Biomedical Informatics Grid, or caBIG™, is a voluntary network or grid ...  
[www.ggf.org/GGF14/ggf\\_events\\_schedule\\_caBig.htm](http://www.ggf.org/GGF14/ggf_events_schedule_caBig.htm) - 11k - [Cached](#) - [Similar pages](#)

[caBIG Participation < Projects < Department of Biomedical ...](#)  
 The goal of caBIG is to develop applications and the underlying systems ... In meeting this goal, caBIG will necessarily bring together data from many and ...  
[bmi.osu.edu/areas\\_and\\_projects/caBIG.cfm](http://bmi.osu.edu/areas_and_projects/caBIG.cfm) - 40k - [Cached](#) - [Similar pages](#)

[Microsoft Excel caBIG Smart Client](#)  
 What is the xl-caBIG Smart Client and Smart Client Mobile? The xl-caBIG Smart Client is a set of extensions to Microsoft Excel 2003 that gives scientists a ...  
[xl-cabig-client.sourceforge.net/](http://xl-cabig-client.sourceforge.net/) - 6k - [Cached](#) - [Similar pages](#)

• **Search Term (caBIG)**  
 – **Project Website is #6**



- **Outside developers use our API (because they've written emails)**

# The Future

---

- First of all, thank you to [Microsoft Research](#), Dan Fay, and Simon Mercer who made this all possible by giving us the initial funding.
- Our immediate goal is to publish a paper, describing the xl-caBIG Project and motivating our view-point that Microsoft Excel is the eScience Browser
- Funding dependent, we want to continue developing our prototype, integrating with caGRID 1.0 and Office 2007 technologies.

# xl-caBIG Project Team

CO-PIs *Katarzyna Macura MD PhD* and  
*Robert Macura MD PhD*



Lead Programmer: *Tom Macura BS BS*  
(caBIG dotNET, Smart Client)



Programmer: *Wiktor Macura*  
(Smart Client MOBILE)



Sponsored by: *Microsoft Research Smart Clients for eScience 2005*

Project website: <http://xl-cabig-client.sourceforge.net/>



*BSD Open-Source License* 