A Web-Based Knowledge Exploitation Toolset for the CBDP

Gaylen W. Drapé
Program Manager
ENSCO, Inc.

Chemical Biological Information Systems Conference
Austin, TX

January 10, 2007

Outline

- Motivation for Knowledge Exploitation Tools
- Conceptual Architecture / Process
- Software Technology Demonstration
- Potential Use Cases
- Implementation Issues
- Summary

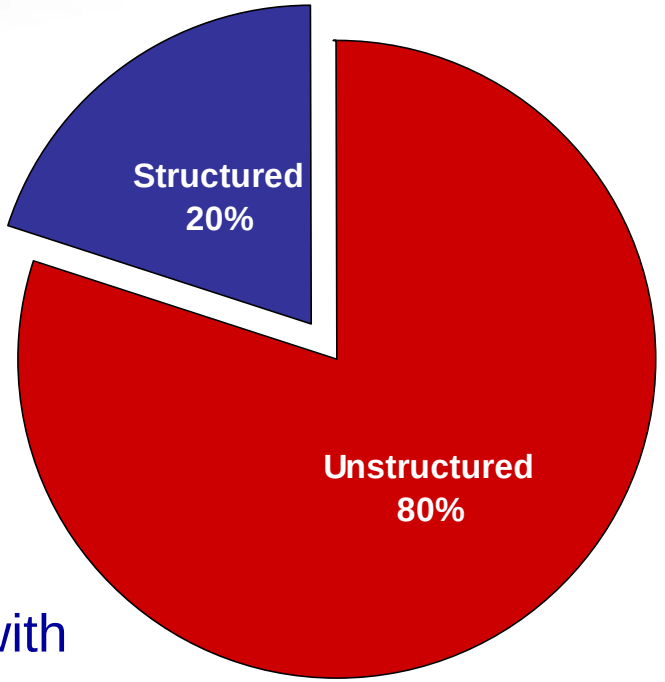
Challenges to Effective CBDP Decision Making

- Multiple participant organizations (government, industry, academia)
- Large volume of data stored in multiple formats and schemas
- Different business processes inherently require different kinds of information
  - S&T development
  - Product development
  - Testing and evaluation
  - Field O&M
- Inconsistent semantic standards increase difficulty of searching, analyzing, and adding to the body of knowledge

Need tools to grow and manage a complex knowledge base

Exploitation of Unstructured Information

- Structured data
  - Rows, columns, tables
  - Relationships predetermined
  - Tells the user “what”
- Unstructured content
  - Text, images, media
  - Structure can be discovered
  - Tells the user “how” and “why”
- Most business processes conducted with unstructured information (see right)



Source: Gartner Group

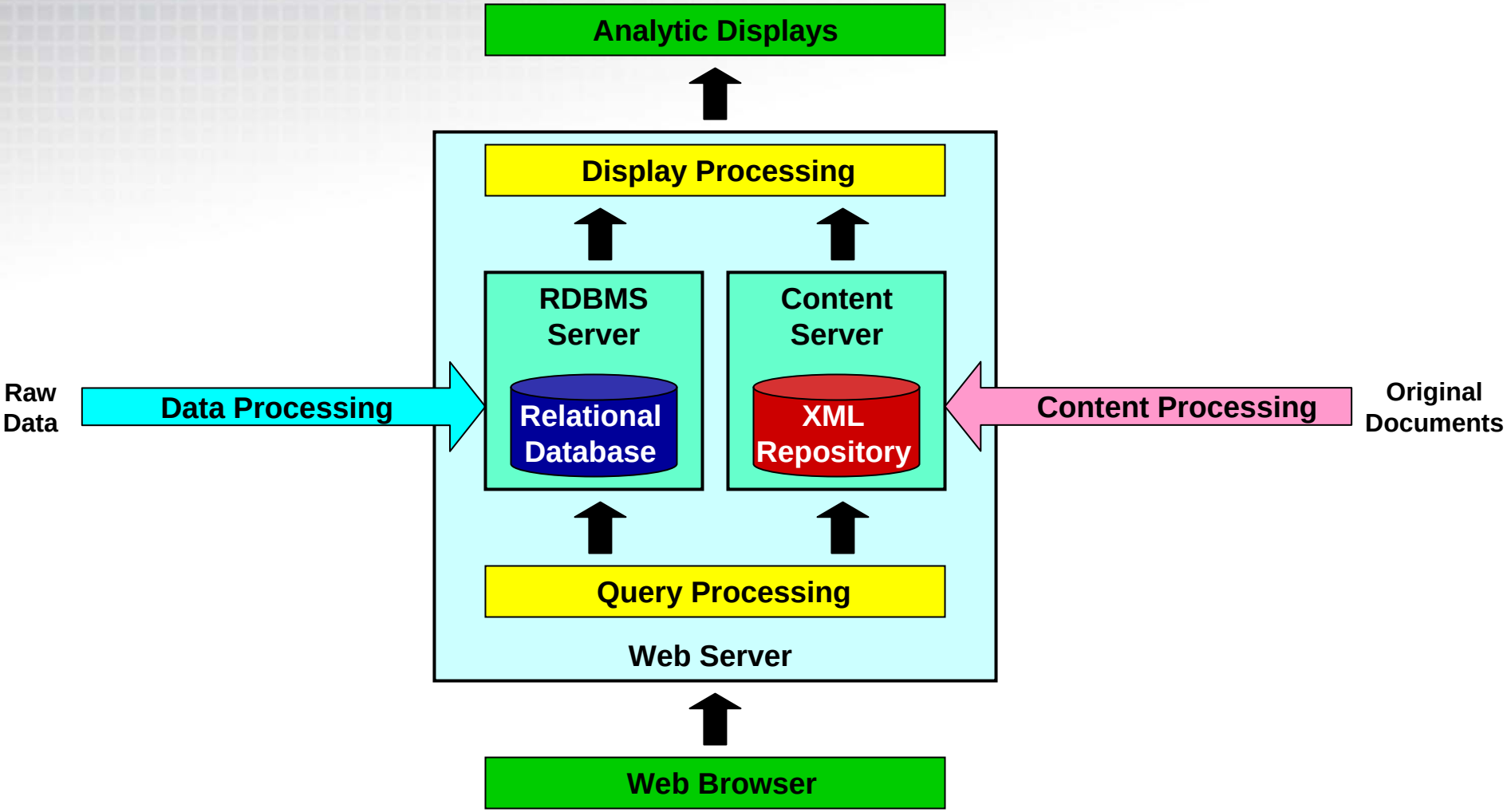
Exploitation of unstructured content is a huge opportunity

Technology Illustration: Biological Knowledge Base

- BioKB was a BAA project funded by DIA and DTRA
- Developed prototype biological data portal for MASINT and Biodefense communities
- Software built with Government-furnished data and content (~1 GB) from multiple sources (see below)

Structured *	Unstructured **
Bioassay protocols Project descriptions	Assay procedures Project deliverable docs Detector market survey DNA Purification survey S&T background info
* spreadsheets	** documents (Web, Word, .pdf, hard copy)

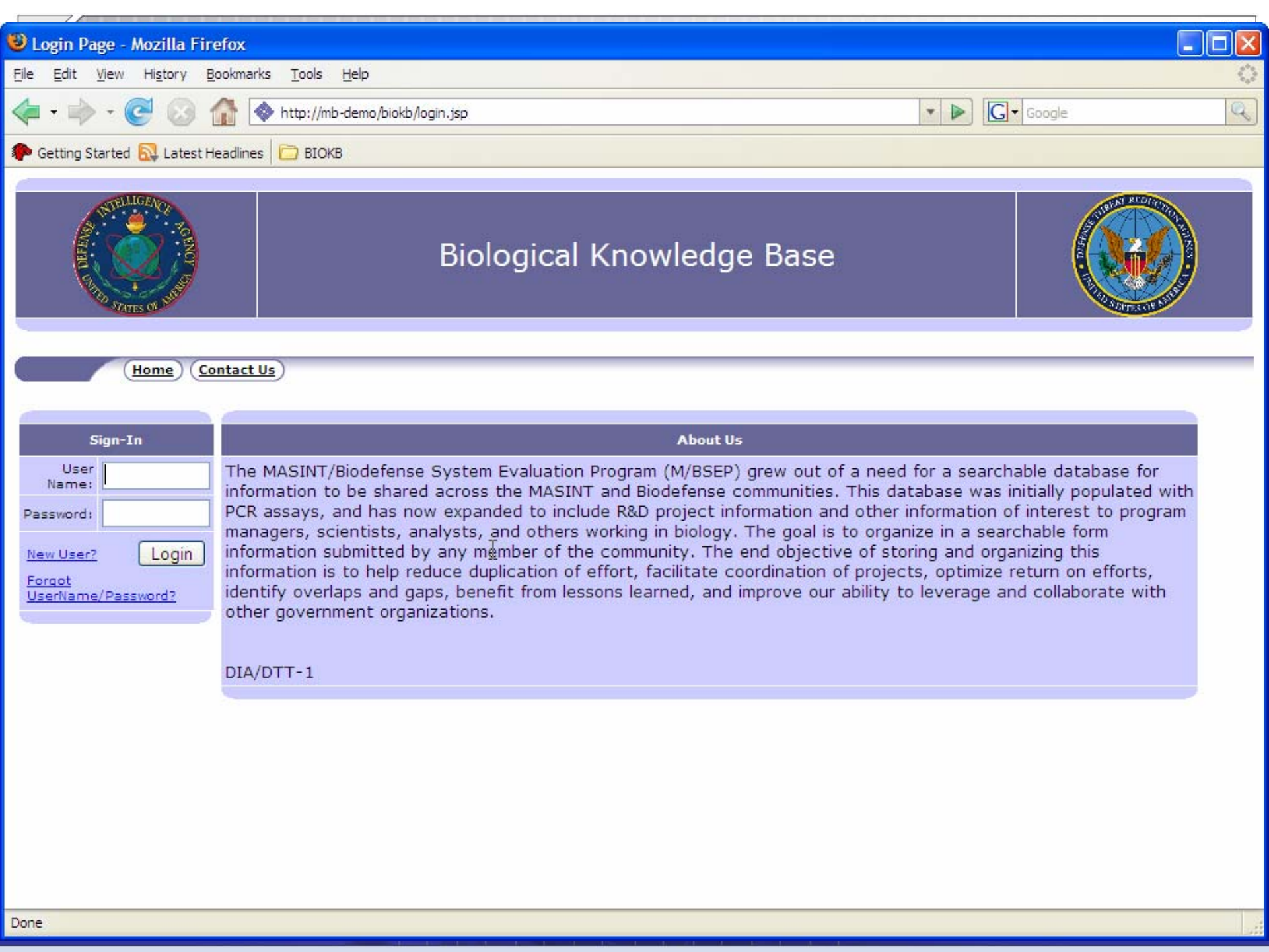
Conceptual Architecture / Processing



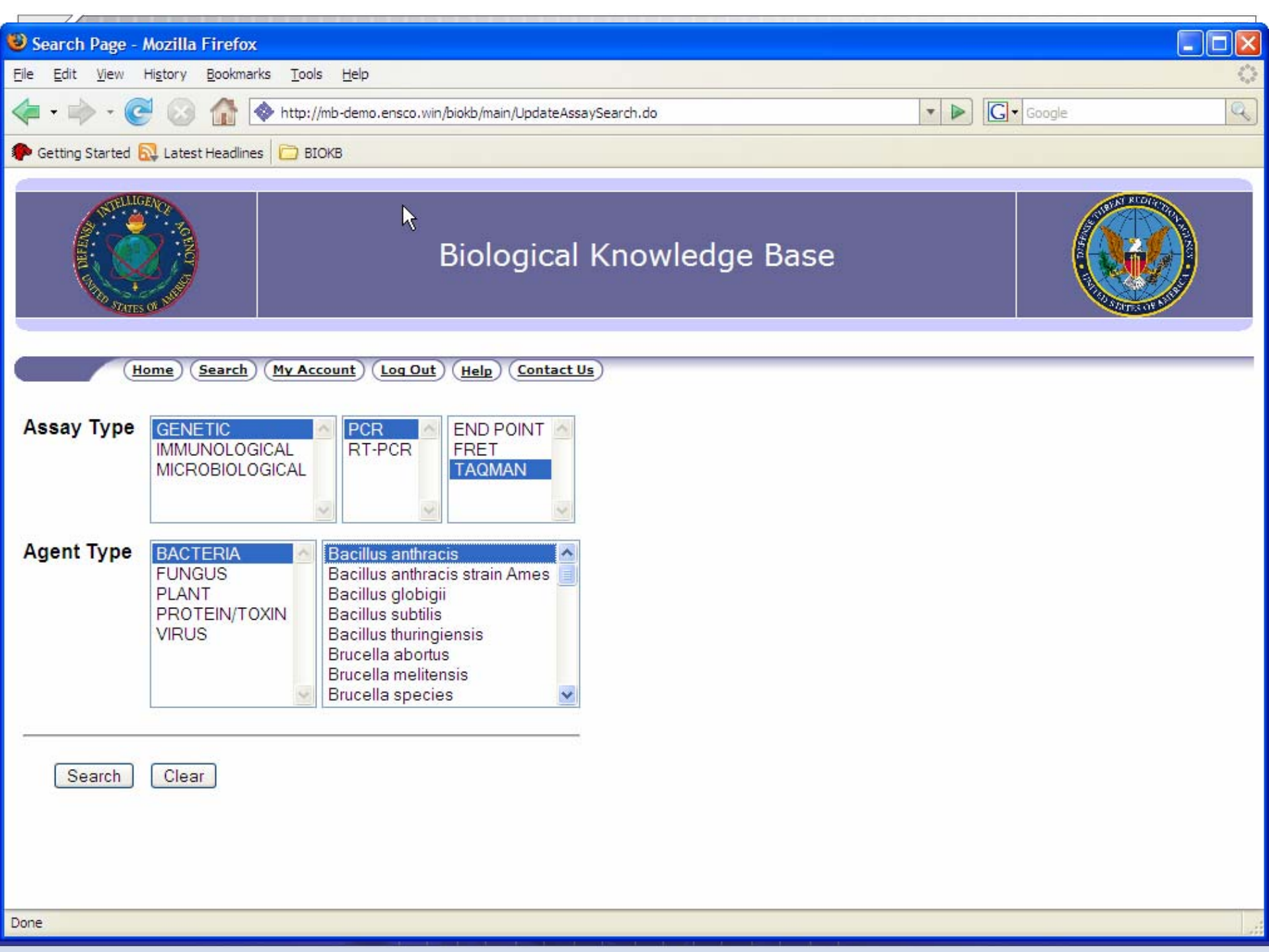
Software Demonstration

Biological Knowledge Base Prototype
Unclassified demo: https://www.biokb.ensco.com

For access, contact: Sandra Mahl – mahl.sandy@ensco.com
(321) 775-7547


Login Page - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://mb-demo/biokb/login.jsp
Google
Getting Started
Latest Headlines
BIOKB
Biological Knowledge Base
Home
Contact Us
Sign-In
User Name:
Password:
New User?
Login
Forgot UserName/Password?
About Us
The MASINT/Biodefense System Evaluation Program (M/BSEP) grew out of a need for a searchable database for information to be shared across the MASINT and Biodefense communities. This database was initially populated with PCR assays, and has now expanded to include R&D project information and other information of interest to program managers, scientists, analysts, and others working in biology. The goal is to organize in a searchable form information submitted by any member of the community. The end objective of storing and organizing this information is to help reduce duplication of effort, facilitate coordination of projects, optimize return on efforts, identify overlaps and gaps, benefit from lessons learned, and improve our ability to leverage and collaborate with other government organizations.
DIA/DTT-1
Done


Home Page - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://mb-demo.ensco.win/biokb/main/home.jsp
Google
Getting Started Latest Headlines BIOKB
Biological Knowledge Base
Home Search My Account Log Out Help Contact Us
Welcome Gaylen
Search Tools
Assay Search
Project Search
MarketSurvey Search
Document Search
Submission Tools
Create Assay For Submission
Submit Assay For Review
Create Project For Submission
Submit Project
Other Information
External Links
Administration
Data Admin
User Admin
Done


Search Page - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://mb-demo.ensco.win/biokb/main/UpdateAssaySearch.do
Google
Getting Started Latest Headlines BIOKB
Biological Knowledge Base
Home Search My Account Log Out Help Contact Us
Assay Type
GENETIC
IMMUNOLOGICAL
MICROBIOLOGICAL
PCR
RT-PCR
END POINT
FRET
TAQMAN
Agent Type
BACTERIA
FUNGUS
PLANT
PROTEIN/TOXIN
VIRUS
Bacillus anthracis
Bacillus anthracis strain Ames
Bacillus globigii
Bacillus subtilis
Bacillus thuringiensis
Brucella abortus
Brucella melitensis
Brucella species
Search
Clear
Done

Search Results - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://mb-demo.ensco.win/biokb/main/UpdateAssaySearch.do

Getting Started Latest Headlines BIOKB

Biological Knowledge Base

Home | Search | My Account | Log Out | Help | Contact Us

Search results for: Assay Type: GENETIC and PCR and TAQMAN and Agent Type: BACTERIA and Agent Name: Bacillus anthracis

Assay ID	Agent Name	Type	POC	Detection Limit	Instrument	Target Gene	Amplicon Length	Last Modified
6	Bacillus anthracis	GENETIC, PCR, TAQMAN	John Doe	100.0 fg	R.A.P.I.D Light Cycler	capA	89.0	2005-02-23
7	Bacillus anthracis	GENETIC, PCR, TAQMAN	John Doe	50.0 fg	R.A.P.I.D Light Cycler	capB	84.0	2005-02-23

New Search

Done

Biological Knowledge Base

Return to Search Results

Oligo Summary for PCR Primer Set # 1

Sequence Type	Melting Temp(°C)	Stock Concentration	Final Concentration	Volume (µL)	Sequence	Length(bases)
Primer A	58.3	0.0 nM	0.0 nM	0.0	CTAGCTAGCTAGCTAGCTAGCTAG	20
Primer A	58.4	0.0 nM	0.0 nM	0.0	CTAGCTAGCTAGCTAGCTAGCTAG	20
Primer B	68.3	0.0 nM	0.0 nM	0.0	CTAGCTAGCTAGCTAGCTAGCTAG	20

Primer #	Amplicon Length (bp)	Detection Limit	Gene	Agent
1	89.0	100.0 fg	capA	Bacillus anthracis of type BACTERIA

Instrument - R.A.P.I.D Light Cycler (PCR)

Reaction Conditions

Component Name	Source	Type	Volume (µL)	Stock Concentration	Final Concentration

Search Home - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://mb-demo.ensco.win/biokb/main/searchHome.jsp

Google

Getting Started Latest Headlines BIOKB

Biological Knowledge Base

- Assay Search
- Project Search
- MarketSurvey Search
- Document Search

Done

Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://mb-demo/biokb/main/documentSearch.jsp

Google

Getting Started Latest Headlines BIOKB

Biological Knowledge Base

Home | Search | My Account | Log Out | Help | Contact Us

viral All search

Using a value from the drop down list (category) as a search term has unpredictable results.

Document Search Document Mining

Searched for **viral** Results **1** - **10** of about 53 (0.23 seconds)

Below are documents including your search term. Displayed with each result is the first instance that contains the search term, a score and the DDMS security and resource information, and a link to the beginning of the document.

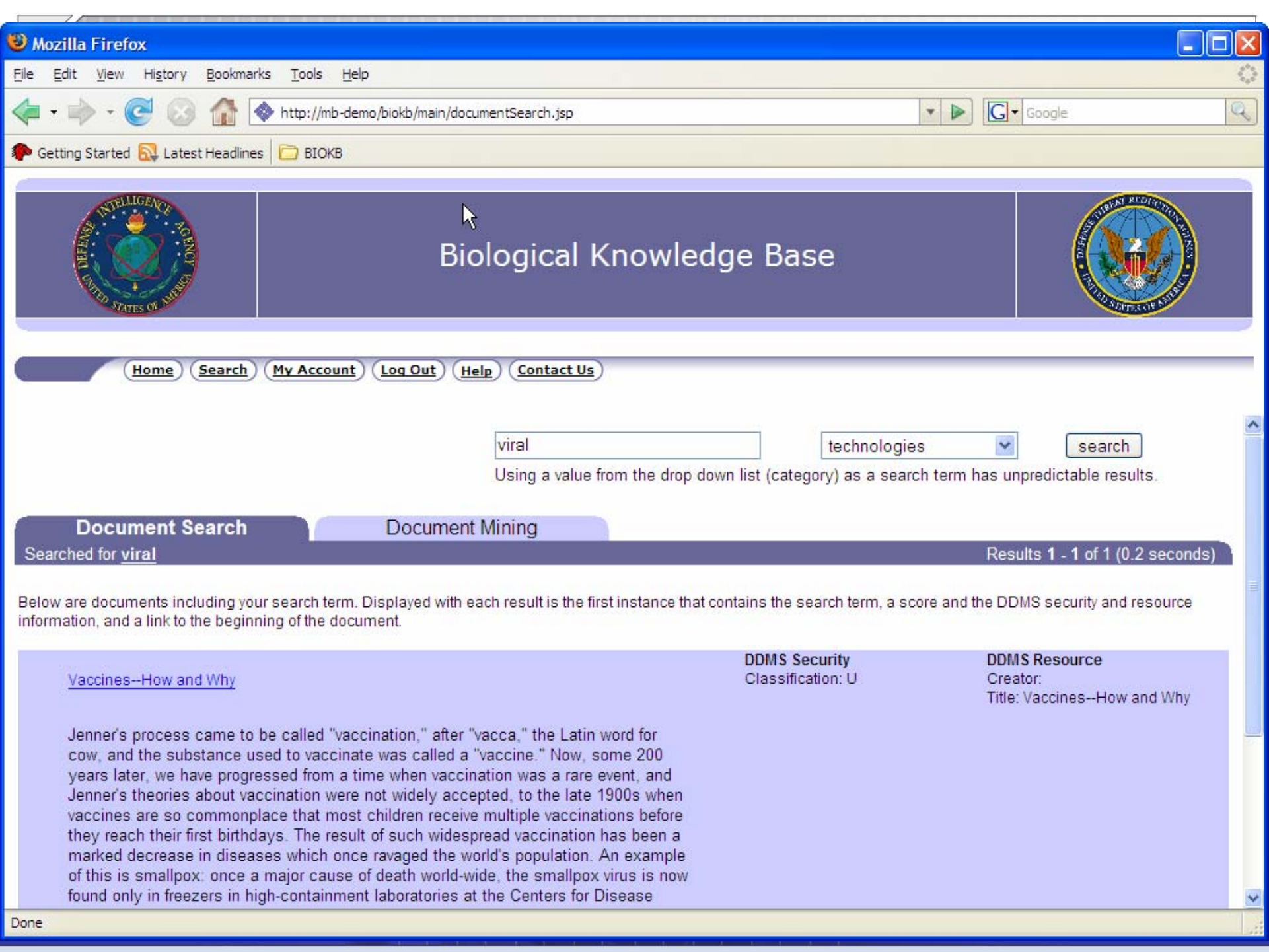
Vaccines--How and Why

DDMS Security
Classification: U

DDMS Resource
Creator:
Title: Vaccines--How and Why

Jenner's process came to be called "vaccination," after "vacca," the Latin word for cow, and the substance used to vaccinate was called a "vaccine." Now, some 200 years later, we have progressed from a time when vaccination was a rare event, and Jenner's theories about vaccination were not widely accepted, to the late 1900s when vaccines are so commonplace that most children receive multiple vaccinations before they reach their first birthdays. The result of such widespread vaccination has been a marked decrease in diseases which once ravaged the world's population. An example of this is smallpox: once a major cause of death world-wide, the smallpox virus is now found only in freezers in high-containment laboratories at the Centers for Disease

Done


Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://mb-demo/biokb/main/documentSearch.jsp
Google
Getting Started Latest Headlines BIOKB
Biological Knowledge Base
Home Search My Account Log Out Help Contact Us
viral
technologies
search
Using a value from the drop down list (category) as a search term has unpredictable results.
Document Search
Document Mining
Searched for viral
Results 1 - 1 of 1 (0.2 seconds)
Below are documents including your search term. Displayed with each result is the first instance that contains the search term, a score and the DDMS security and resource information, and a link to the beginning of the document.
Vaccines--How and Why
DDMS Security
Classification: U
DDMS Resource
Creator:
Title: Vaccines--How and Why
Jenner's process came to be called "vaccination," after "vacca," the Latin word for cow, and the substance used to vaccinate was called a "vaccine." Now, some 200 years later, we have progressed from a time when vaccination was a rare event, and Jenner's theories about vaccination were not widely accepted, to the late 1900s when vaccines are so commonplace that most children receive multiple vaccinations before they reach their first birthdays. The result of such widespread vaccination has been a marked decrease in diseases which once ravaged the world's population. An example of this is smallpox: once a major cause of death world-wide, the smallpox virus is now found only in freezers in high-containment laboratories at the Centers for Disease
Done

Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://mb-demo/biokb/main/documentSearch.jsp

Getting Started Latest Headlines BIOKB

Biological Knowledge Base

Home | Search | My Account | Log Out | Help | Contact Us

viral technologies search

Using a value from the drop down list (category) as a search term has unpredictable results.

Document Search | **Document Mining**

Searched for **viral** Results **1** - **4** of 4 (0.17 seconds)

Each document has been mined for the query term entered. All paragraphs containing the search are displayed.

Document: **Vaccines--How and Why**

fragment 1

DDMS Security
Classification: U

DDMS Resource
Creator:
Title: Vaccines--How and Why

Jenner's process came to be called "vaccination," after "vacca," the Latin word for cow, and the substance used to vaccinate was called a "vaccine." Now, some 200 years later, we have progressed from a time when vaccination was a rare event, and Jenner's theories about vaccination were not widely accepted, to the late 1900s when vaccines are so commonplace that most children receive multiple vaccinations before they reach their first birthdays. The result of such widespread vaccination has been a marked decrease in diseases which once ravaged the world's population. An example of this is smallpox: once a major cause of death world-wide, the smallpox virus is now found only in freezers in high-containment laboratories at the Centers for Disease

Done

Biological Knowledge Base

fragment 2

Classification: U

Creator:
Title: Vaccines--How and Why

Disease causing organisms have at least two distinct effects on the body. The first effect is very obvious: we feel sick, exhibiting symptoms such as fever, nausea, vomiting, diarrhea, rash, and many others. Although the second effect is less obvious, it is this effect that generally leads to eventual recovery from the infection: the disease causing organism induces an immune response in the infected host. As the response increases in strength over time, the **infectious agents** are slowly reduced in number until symptoms disappear and recovery is complete.

How does induction of the immune response occur? The disease causing organisms contain proteins called "antigens" which stimulate the immune response. The resulting immune response is multi-fold and includes the synthesis of proteins called "antibodies." These proteins bind to the disease causing organisms and lead to their eventual destruction. In addition, "memory cells" are produced in an immune response. These are cells which remain in the blood stream, sometimes for the life span of the host, ready to mount a quick protective immune response against subsequent infections with the particular disease causing agent which induced their production. If such an infection were to occur, the memory cells would respond so quickly that the resulting immune response could inactivate the disease causing agents, and symptoms would be prevented. This response is often so rapid that infection doesn't develop - you are immune from infection.

Search Home - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://mb-demo.ensco.win/biokb/main/searchHome.jsp

Google

Getting Started Latest Headlines BIOKB

Biological Knowledge Base

- Assay Search
- Project Search
- MarketSurvey Search
- Document Search

Done

Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://mb-demo/biokb/main/marketSurveySearch.jsp

Google

Getting Started Latest Headlines BIOKB

Biological Knowledge Base

Home | Search | My Account | Log Out | Help | Contact Us

Market Survey Search

Name

Size: Approx the size of a carry-on luggage suitcase

Sensitivity

Signature

Re-Use

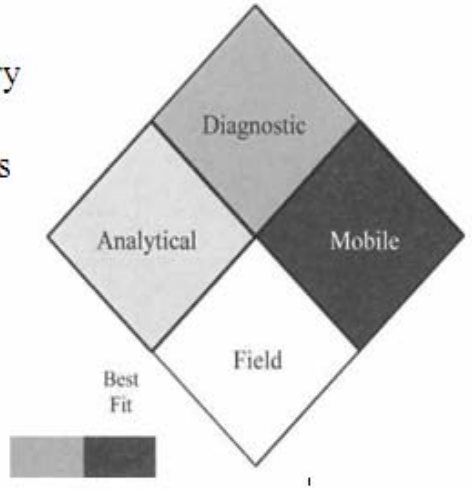
MarketSurvey Type: Detector

Submit Query

Name	Vendor	Size	Sensitivity	Signature	Re-Use
Eppendorf Mastercycler Step*	Eppendorf	Approximately the size of a carry- on luggage suitcase Between 5 and 25 kg Shelf life between 1 and 3 years	Unknown CFU per ml	No sounds produced that cannot be deactivated Between 200-500 BTUS generated	N/A
LD 400 Luminescence Detector	Beckman Coulter, Inc.	Approximately the size of a carry- on luggage suitcase Between 5 and 25 kg NA shelf life	N/A	N/A	Device or system is intended for multiple use NA solutions or buffers used 1 component Tubing cleaning required once a week

Done

Summary of Analysis



MatriCycler Evaluation Criteria Provided by Vendor

Sensitivity:
- NACFUperml

Maturity Gauge:
- A few devices or systems exist (brass board)

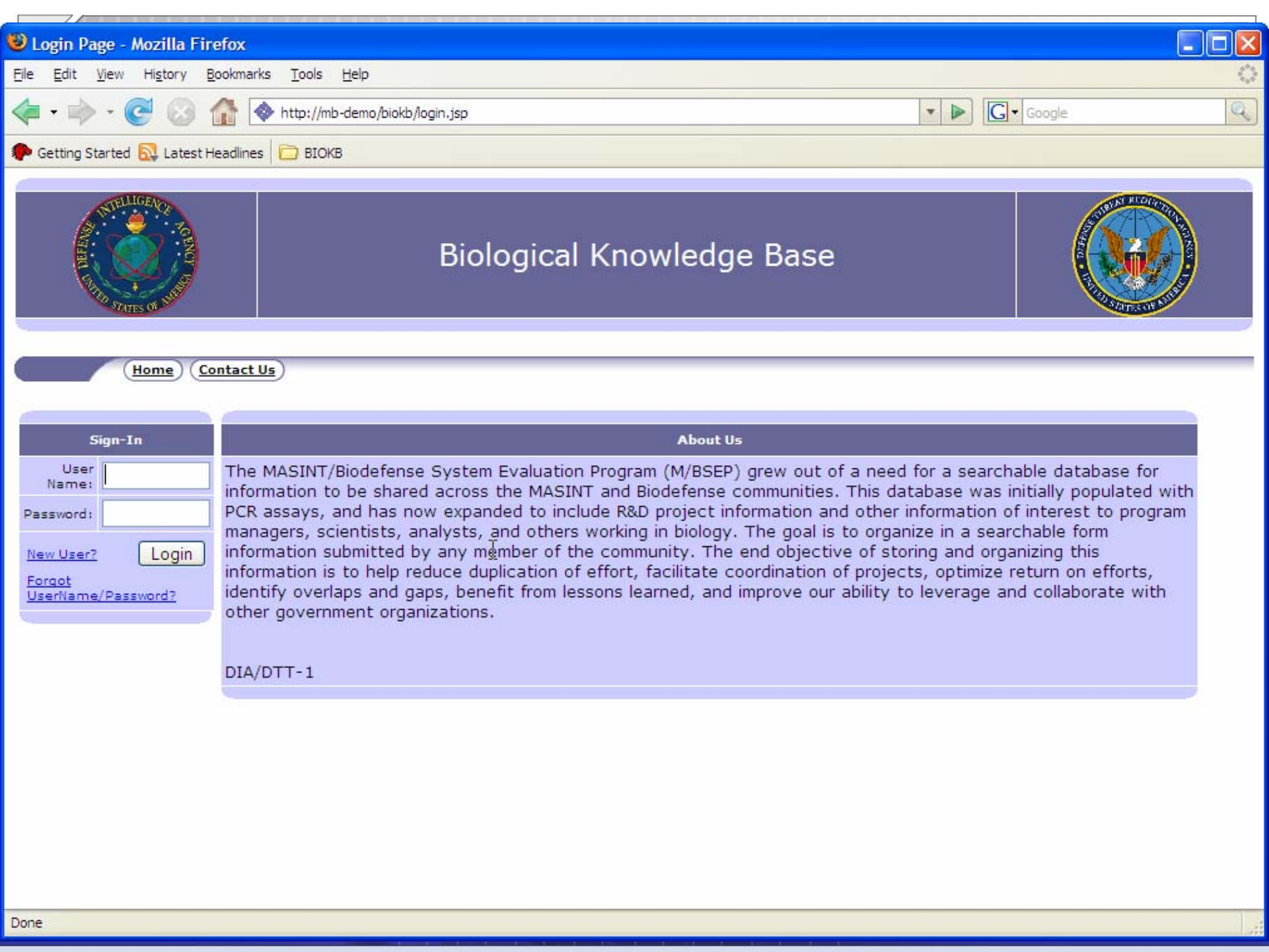
• o 0 O

System requirements:
- System or device has 110V electrical requirement
- The system or device does not require water
- The system or device does not require an external air or gas source
- The system or device does not require an external vacuum source

Re-use:
- Device or system is intended for single use
- NA solution or butler used
- NA components
- NA cleaning

Maintenance:

Ease of use/Utility
- Cannot view results "in real time"
- NA centrifugation steps
- NA vortexing steps
- NA to interpret raw data or call a positive through internal software
- NA detecting multiple biological agents or toxins within the same test
- NA additional pieces of equipment needed
- NA sample purification

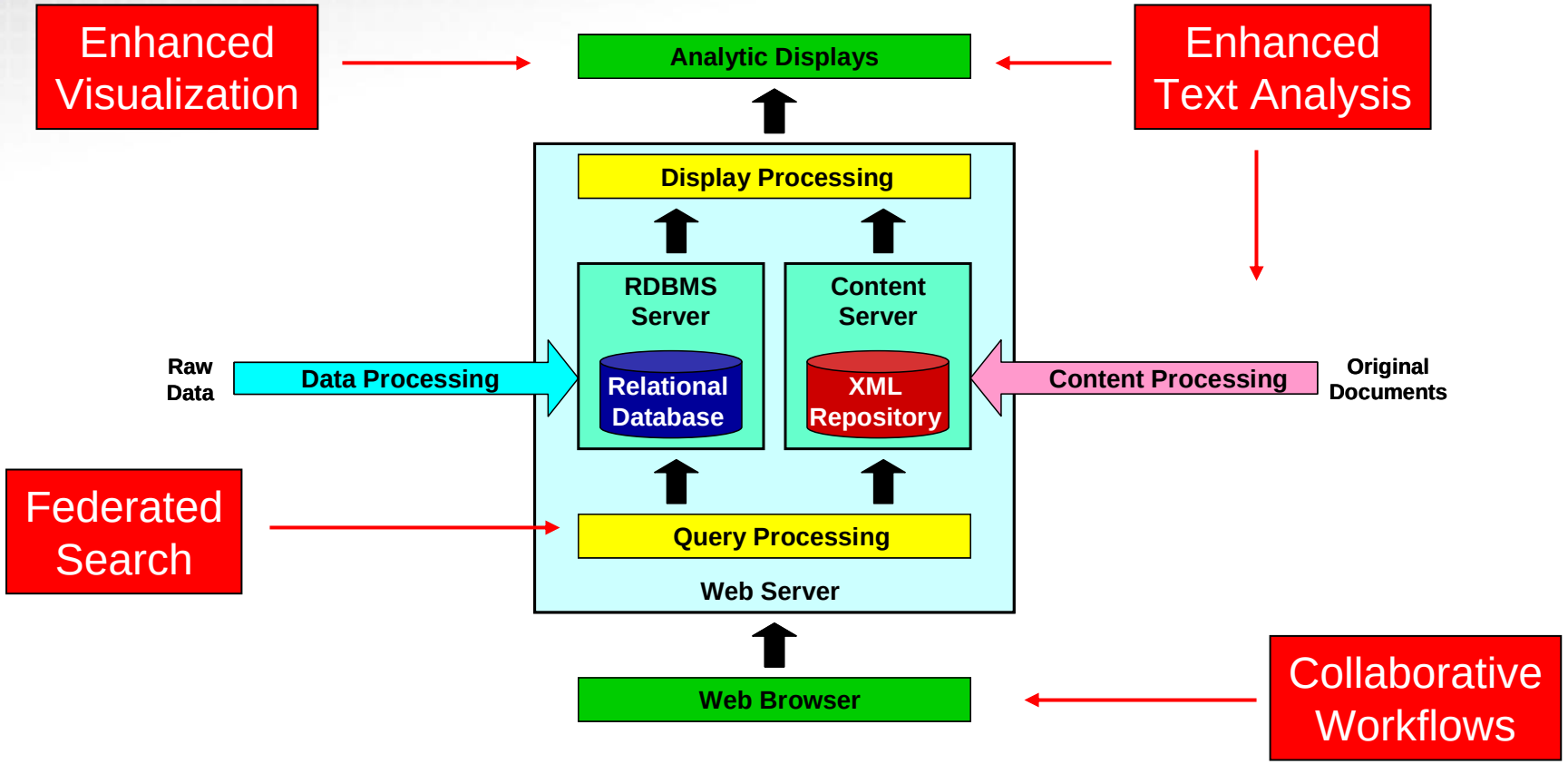

Login Page - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://mb-demo/biokb/login.jsp
Google
Getting Started
Latest Headlines
BIOKB
Biological Knowledge Base
Home
Contact Us
Sign-In
User Name:
Password:
New User?
Login
Forgot UserName/Password?
About Us
The MASINT/Biodefense System Evaluation Program (M/BSEP) grew out of a need for a searchable database for information to be shared across the MASINT and Biodefense communities. This database was initially populated with PCR assays, and has now expanded to include R&D project information and other information of interest to program managers, scientists, analysts, and others working in biology. The goal is to organize in a searchable form information submitted by any member of the community. The end objective of storing and organizing this information is to help reduce duplication of effort, facilitate coordination of projects, optimize return on efforts, identify overlaps and gaps, benefit from lessons learned, and improve our ability to leverage and collaborate with other government organizations.
DIA/DTT-1
Done

End of Demonstration

Business Cases for Unstructured Content Analysis

- Detector Capability Assessment
  - Issue: Difficult to compare and evaluate detector characteristics across multiple dimensions
  - Solution: Analysis of detector documentation, automatically summarize entities & concepts, route among SMEs for comment
  - Benefit: Accelerate creation/adoption of data standards, enable deeper understanding of capability and investment opportunities
- Test Traceability
  - Issue: Test plans & procedures not consistent with standards
  - Solution: Search/retrieve/parse content from existing test DB/documents, then map to elements of standard test process
  - Benefit: Facilitate creation of overarching T&E model for broader application of test data, more efficient use of resources

Potential Architecture Directions



Implementation Issues

Issues we encountered in the BioKB project:

- Accessibility of Documents via Government WANs
- Verification and Validation of input document content
- Certification and Accreditation of COTS technology components
- Organizational barriers to information sharing

Issues are both technical and non-technical in nature

Some ideas:

- Strong governance by Sponsor, e.g. steering group
- Partnership agreements among users/developers/integrators
- Provision of software test / demonstration network

Summary and Concluding Thoughts

- Opportunity for exploitation of unstructured data
  - Increase level of knowledge sharing across CBDP community
  - Answer “tough questions” to speed up development and transition
  - Reduced effort of CBD researchers and analysts
- Maturing industry standards and community-wide data integration make web-based KB tools feasible
- Broad-scale implementation a challenge for technical and non-technical reasons
- Suggest development of focused pilot KB applications to work out implementation issues, demonstrate ROI

Thank You!