

Lessons Learned from the SARS Data Exchange



Presented By

**Patrick Anderson MPH
Surveillance and Statistics Section
Infectious Diseases Branch
Division of Communicable Disease Control
California Department of Health Services
panders1@dhs.ca.gov.**

I couldn't do it alone

**Kate Cummings MPH
Infectious Diseases Branch
Division of Communicable Disease Control
California Department of Health Services**

**Gillian Hamilton MPH
Infectious Diseases Branch
Division of Communicable Disease Control
California Department of Health Services**

**Leah Estberg DVM PhD
Surveillance and Statistics Section
Infectious Diseases Branch
Division of Communicable Disease Control
California Department of Health Services**

**Chris Flanagan
Developer
San Diego, CA
strake@cox.net**

SARS

- Severe Acute Respiratory Syndrome.
- Recently recognized infection.
- Worldwide outbreak 2003.
- Thousands of people became ill.
- Hundreds of people died.
- In the US, there were lab confirmed cases.
 - All had traveled to areas where transmission was occurring.
- There were no deaths in the US.

California

- Had over 150 cases worthy of some follow-up.
- A worst day scenario was upwards of 80 suspect/probable under investigation.
- Daily press reports generated.
- MS[®]Excel used to store data.

Use of Excel as the Data Management Tool

- **Benefits**

- Users were familiar with Excel.
- Easy to extract to other applications.
- Everyone has the application so easy to share document.

- **Difficulties**

- Difficult to enter and manage the information.
- Users unfamiliar with formatting data.
- Daily updates took time.
- Tracking of history data such as travel was cumbersome.
- Difficult to extract numbers for counts.

We Needed
A
Better Way
To Manage
The Data !

- Project team, crossing different disciplines was organized.

- Epidemiologists
- Physicians
- Informaticists
- Data modelers
- Developers



- MS[®] Access data base was developed.
 - Very little built in logic.
 - Most decisions were left to the clinicians and data managers.
 - Tabbed based GUI.
 - Goal to eventually move to SQL-2000.

- Leveraged concepts from:
 - Public Health Conceptual Data Model (PHCDM).
 - HL7 Reference Information Model (RIM).

Microsoft Access - [Case Identification]

File Edit View Insert Format Records Tools Window Help

Type a question for help

Case ID 19 CDC report received [dropdown] Full Name [text] Case status [dropdown]
 CDC ID [text] SARS report date [text] Onset date "Respiratory" [text] Onset date "Fever" [text]
 LHD jurisdiction of case residence [dropdown]

Personal Information PH Contact Info Clinical Criteria Travel History Exposure Observations Lab Results

First name [text] Last name [text] Initials [text] Age [text] Units [dropdown]
 Gender [dropdown] Race [dropdown] Ethnicity [dropdown]
 Address [text] City [text] State [dropdown] Zip [text] 0
 Occupation [dropdown] Place of residence [text] Place of diagnosis [text]

Case: 1 of 8

Search By CDC ID [text] Case ID [text]
 Filter By Case status [dropdown] Assigned jurisdiction [dropdown] Initials [text] resume [button] Close [button] Close Form [button]

Header information
 Tabbed architecture

Search/filter functionality

Microsoft Access - [Case Identification]

File Edit View Insert Format Records Tools Window Help

Type a question for help

Case ID 19 CDC report received [dropdown] Full Name [text] Case status [dropdown]
 CDC ID [text] SARS report date [text] Onset date "Respiratory" [text] Onset date "Fever" [text]
 LHD jurisdiction of case residence [dropdown]

Personal Information PH Contact Info Clinical Criteria Travel History Exposure Observations Lab Results

Overall Lab Assessment [dropdown] Day 22 Day 29

Notes [text]

Pathogen	Status	Assessment	Notes
[dropdown]	[dropdown]	[dropdown]	[text]

Pathogen: 1 of 1

Specimen ID	Specimen type	Lab workup	Test name	Lab	Collection date	Test date	Test result	Filter by lab workup
[text]	[dropdown]	[dropdown]	[text]	[dropdown]	[text]	[text]	[text]	[dropdown]

Notes [text]

Notes [text]

Test: 2 of 2

Case: 1 of 8

Search By CDC ID [text] Case ID [text]
 Filter By Case status [dropdown] Assigned jurisdiction [dropdown] Initials [text]

Close Close Form

Three levels allowed in access

Then What?

- We discovered a CDC web based system was being developed.
- Jurisdictions were given the choice of using the web based system or exchanging data via an Extensible Markup Language (XML) formatted file.
- To use the upload each jurisdiction must pass a four step certification process.
- A jurisdiction could either use the web based interface or upload data from their own in-house system, but not both.

What to Do?



We decided to continue work on an in-house solution and upload data to the CDC.

Problems

- Our data model was slightly different.
 - Greater emphasize on lab workup.
- Data was located across branches.
 - Do to security and disparate systems, access was difficult.
 - A data exchange procedure was needed.
- Los Angeles was designated a jurisdiction.
 - One data entry point to CDC from CA was needed.
 - How to insure secure data transmission.
 - How to insure data integrity at the State level.

Problems

- Lack of internal understanding of XML schemas.
- Lacked the tools necessary to perform XML validation.

Solution

- Adopted the CDC data model.
 - Added fields that were important for our purposes.
- Established a data exchange agreement with the Viral lab.
 - Excel file that would contain the necessary information for input into their system.
- Established a data exchange agreement with Los Angeles county.
 - Data is uploaded to the State database.
 - Any updates are passed back down to LA.
 - ****Problem with the encryption of data****.

Solution

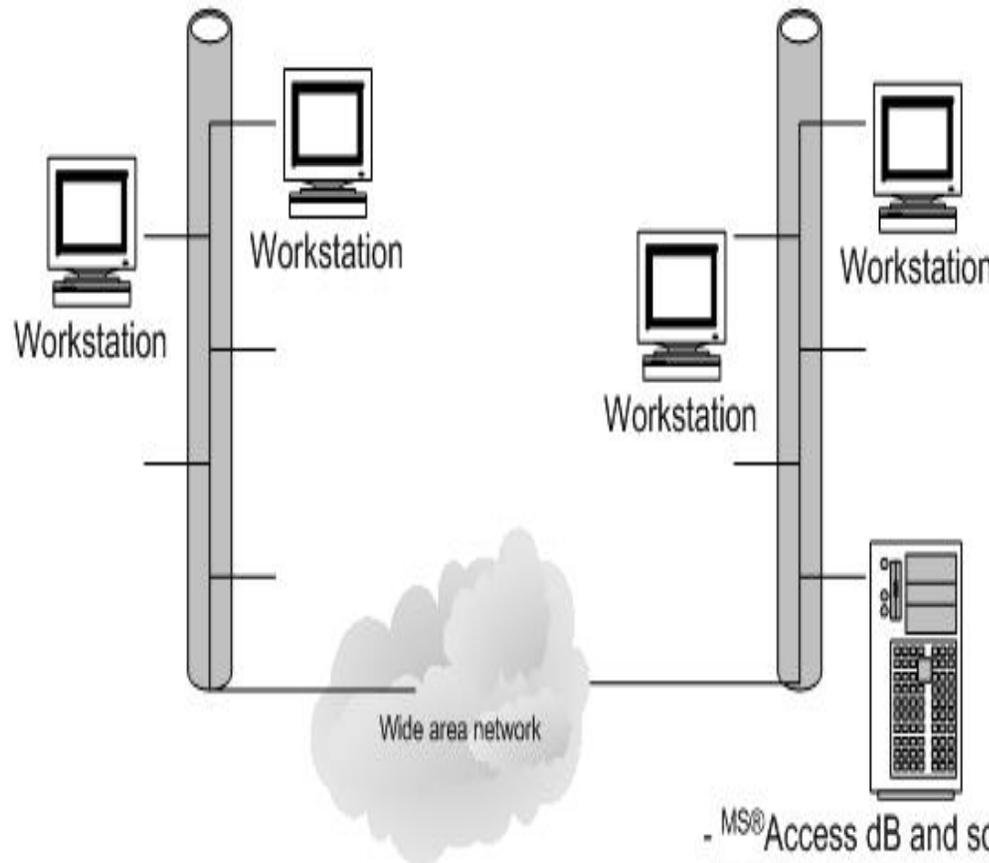
- Downloaded and became familiar with XML Spy to assist in the validation.
- Worked with the SARS certification team and performed internal research to better understand XML schemas and their uses.
 - ****Very important part****

Decided to make it Browser Based

- Utilized Active Local Pages (ALP).
- Developed by newObjects.
 - www.newobjects.com
- Implemented as an Asynchronous Pluggable Protocol
 - Allows MS[®]IE to handle custom URL protocol schemes.
 - <http://msdn.microsoft.com/library/default.asp?url=/workshop/networking/pluggable/overview/overview.asp>
- Extends Internet Explorer's ability to run Active Server Pages locally or over network without the need of a web server.
- Developer license is \$600.00. The system was originally built with the freeware version to determine the feasibility of purchase.

- ALP installed on workstations in the Sacramento office

- All processing done by the workstation



- ALP installed on workstations in the Berkeley office

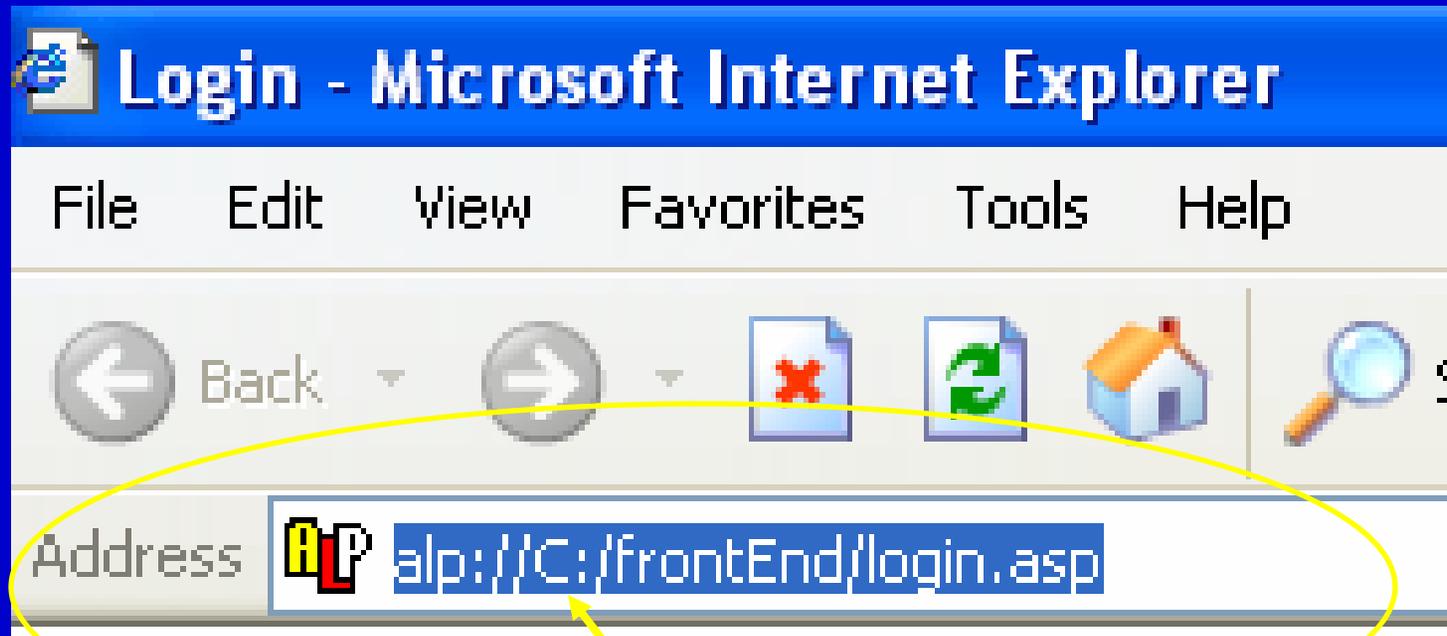
- All processing done by the workstation

- MS[®] Access dB and source code located on a server in Berkeley.

Via ALP, each workstation can access the dB via the browser

- ALP license located in source code folder

- Folder has security placed on it to limit access



Note ALP protocol scheme

Note location of application



SARS Data Management System Version 2.0

California Department of Health Services
Division of Communicable Disease Control

To begin, select from the menu at left
If you experience any difficulties, contact
[technical support](#)

Search	Go
Cases	
Download	Go
Data	
Custom	Go
Download	
Charts	Go
Log	Go
Calendar	Go
technical support	



[Search](#)

[Patient Information](#)

[Classification](#)

[Clinical](#)

[Travel](#)

[P.H. Contact Information](#)

[Epi Risk Factors](#)

[Case to Case Contact Information](#)

[Case to Other Contact Information](#)

[Lab Evaluation](#)

[Notes](#)

[Download](#)

[Custom Download](#)

[Log](#)

Tracking Information:

Local jurisdiction id: 341

Date reported to CDC:

State ID #: CA3

Reporting Jurisdiction

Sacramento

Date reported to state or local health department: 4/7/2004

Submitted By:

Patient Information:

Last Name: dummy First Name: test

Street Address: 999 Zip Code: 12345

City of Residence: sacramento

County of Residence: Sacramento

State of Residence: California

Age at onset: 37 Years

Gender: Male

Ethnicity: Unknown

- Race:
- American Indian/Alaska Native
 - Asian
 - Black
 - Native Hawaiian/Other Pacific Islander
 - White
 - Unknown

Nationality:

Residency: US Resident

Edit

[Search](#)

[Patient](#)

[Information](#)

[Classification](#)

[Clinical](#)

[Travel](#)

[P.H. Contact](#)

[Information](#)

[Epi Risk](#)

[Factors](#)

[Case to Case](#)

[Contact](#)

[Information](#)

[Case to Other](#)

[Contact](#)

[Information](#)

[Lab Evaluation](#)

[Notes](#)

[Download](#)

[Custom](#)

[Download](#)

[Log](#)

Case classification

(using CSTE/CDC definitions)

Initial Classification RUI-1
Classification Date 4/7/2004

This case has been included in the CDC upload and cannot be converted back to being "Under Investigation".

Updated Classification

Classification Date

Update Notes

Classification History

Initial Classification	Updated Classification	Date of Update	Notes
Special Investigation		4/7/2004 2:33:23 PM	
RUI-1	RUI-1	4/7/2004 2:35:15 PM	changed by patrick anderson

- Search [Go](#)
- Cases [Go](#)
- Download Data [Go](#)
- Custom Download [Go](#)
- Charts [Go](#)
- Log [Go](#)
- Calendar [Go](#)
- [technical support](#)

Select the data set you would like to download. The data may be downloaded either in Excel (.xls) as a Comma Separated Value (.csv) file, a Word document (.doc) or as a text file in XML format (.txt).

When prompted, save the files to your hard-drive. If you do not wish to save the file initially, you may open any of the formats. The csv and .txt formats will open in another window while the xls and word formats open within the frame. You will then need to re-run the excel or word documents to save to your hard-drive.

Query Name	File Type	
Case	csv	excel
12/29/03 request	csv	excel
Travel	csv	excel
Demographics	csv	excel
Current Classification	excel only	
Demographics by Status (using initial classifications)	excel only	
Demographics by Status (using updated classifications)	excel only	
Risk Factors by Status (using initial classifications)	excel only	
Risk Factors by Status (using updated classifications)	excel only	
County by Status (using initial classifications)	excel only	
County by Status (using updated classifications)	excel only	
CDC Case Upload *** Can only be submitted via SDN ***	XML only	
VRDL extract	excel	

**xls.
format**

**csv.
format**

**xml.
format**

- [Search](#)
- [Patient Information](#)
- [Classification](#)
- [Clinical](#)
- [Travel](#)
- [P.H. Contact Information](#)
- [Epi Risk Factors](#)
- [Case to Case Contact Information](#)
- [Case to Other Contact Information](#)
- [Lab Evaluation](#)
- [Notes](#)
- [Download](#)
- [Custom Download](#)
- [Log](#)

Select the data set you would like to download. The data may be downloaded either in Excel (.xls) as a Comma Separated Value (.csv) file, a Word document (.doc) or as a text file in XML format (.txt).

When prompted, save the files to your hard-drive. If you do not wish to save the file initially, you may open any of the formats. The csv and txt formats will open in another window while the xls and word formats open within the frame. You will then need to re-run the ex

	File Type	
Case	csv	excel
12/29/03 request	csv	excel
Travel	csv	excel
Demographics	csv	excel
Current Classification	excel only	
Demographics	excel only	
Demographics	excel only	
Risk Factors by	excel only	
Risk Factors by	excel only	
County by Status (using initial classifications)	excel only	
County by Status (using updated classifications)	excel only	
CDC Case Upload *** Can only be submitted via SDN ***	XML only	
VRDL extract	excel	

File Download

Some files can harm your computer. If the file information below looks suspicious, or you do not fully trust the source, do not open or save this file.

File name: xmlUpload.tmp
File type:
From: C

This type of file could harm your computer if it contains malicious code.

Would you like to open the file or save it to your computer?

Always ask before opening this type of file



```
<?xml version="1.0" ?>
- <Import>
  <Sending_Entity>CA</Sending_Entity>
  <Date_File_Created>2004-04-07</Date_File_Created>
  <Data_End_Date>2004-04-07</Data_End_Date>
  <NETSS_SiteID>106</NETSS_SiteID>
- <Case_Data>
  <Agent_Disease>SARS</Agent_Disease>
  <Jurisdiction_CaseID>011</Jurisdiction_CaseID>
  <Internal_CaseID>2</Internal_CaseID>
  <Initial_Classification>R1</Initial_Classification>
  <Updated_Classification>N2</Updated_Classification>
  <Date_Of_Initial_Classification>2004-03-05</Date_Of_Initial_Classification>
  <Date_Of_Updated_Classification>2004-03-19</Date_Of_Updated_Classification>
  <Date_Report_Jurisdiction>2004-02-27</Date_Report_Jurisdiction>
  <Contact_Of_Confirmed_Case>No</Contact_Of_Confirmed_Case>
  <Contact_Of_Possible_Case>No</Contact_Of_Possible_Case>
  <Travel>Yes</Travel>
  <Symptomatic_During_Travel>Unknown</Symptomatic_During_Travel>
  <Submitter>3</Submitter>
</Case_Data>
+ <Case_Data>
+ <Clinical_Information>
+ <Person_Demographics>
+ <Person_Demographics>
+ <Lab_Evaluation>
+ <Lab_Evaluation>
+ <Travel_Destination>
+ <Travel_History>
</Import>
```

We learned some things

Lesson 1

- Both the development and program area teams must be involved in mutual knowledge transfer from the start in order to insure a system that is not only technically but programmatically correct.

Lesson 2

- Compromises must be made so that a system beneficial to all can be developed in a timely manner.

Lesson 3

- The generation (birth) of data within a system must be evaluated and decided upon early in the process in order to avoid potential duplication .

Lesson 4

- An environment must be created that encourages team members from all disciplines to familiarize themselves with standard data exchange technologies such as XML and relational database design.

Lesson 5

- The organizational infrastructure must also mature and develop in order to utilize the technologies indicated in lesson 4.

Lesson 6

- Data exchange standards such as format and security issues must be decided on beforehand.

The End