# The ESSENCE of Disease Surveillance


## Nathaniel Tabernero

Software Engineer

The Johns Hopkins University Applied Physics Laboratory
http://www.jhuapl.edu/


TS-5564

# ESSENCE

## Supporting disease surveillance


Lessons learned from developing a disease surveillance system

# Agenda

Background

Challenges/Solutions

- Data Ingestion
- Detection Algorithms
- User Interface

Technologies

Summary

# Agenda

**Background**

Challenges/Solutions

- Data Ingestion
- Detection Algorithms
- User Interface

Technologies

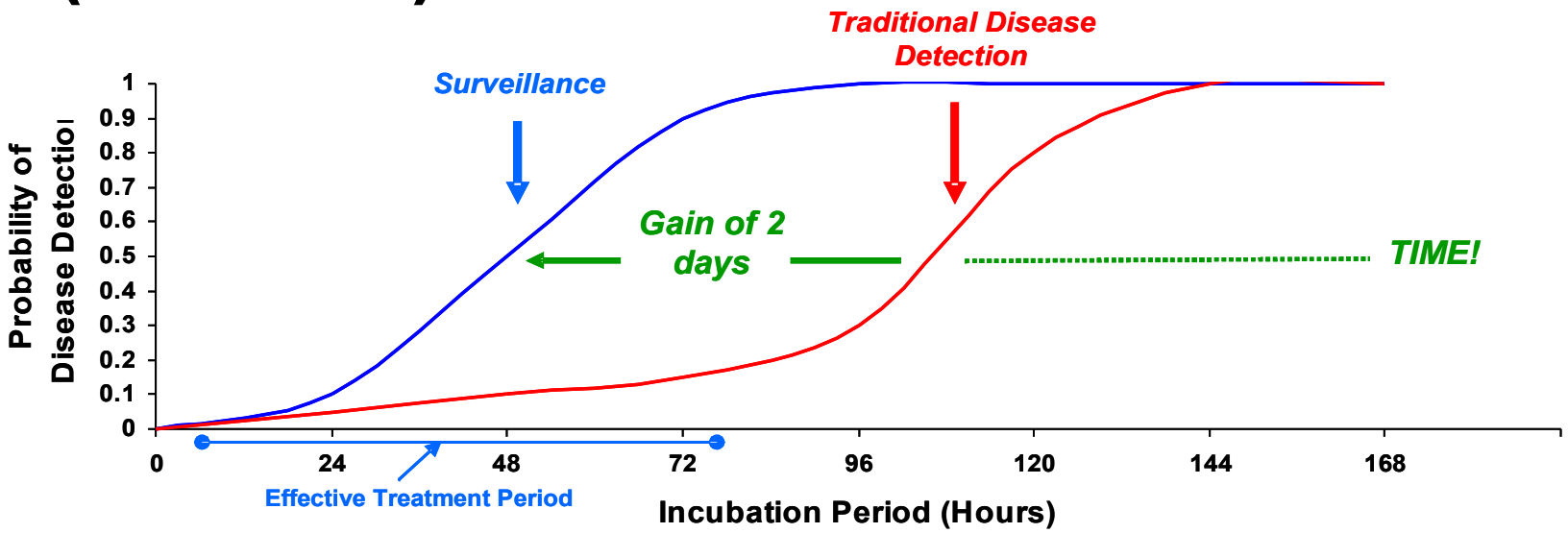
Summary

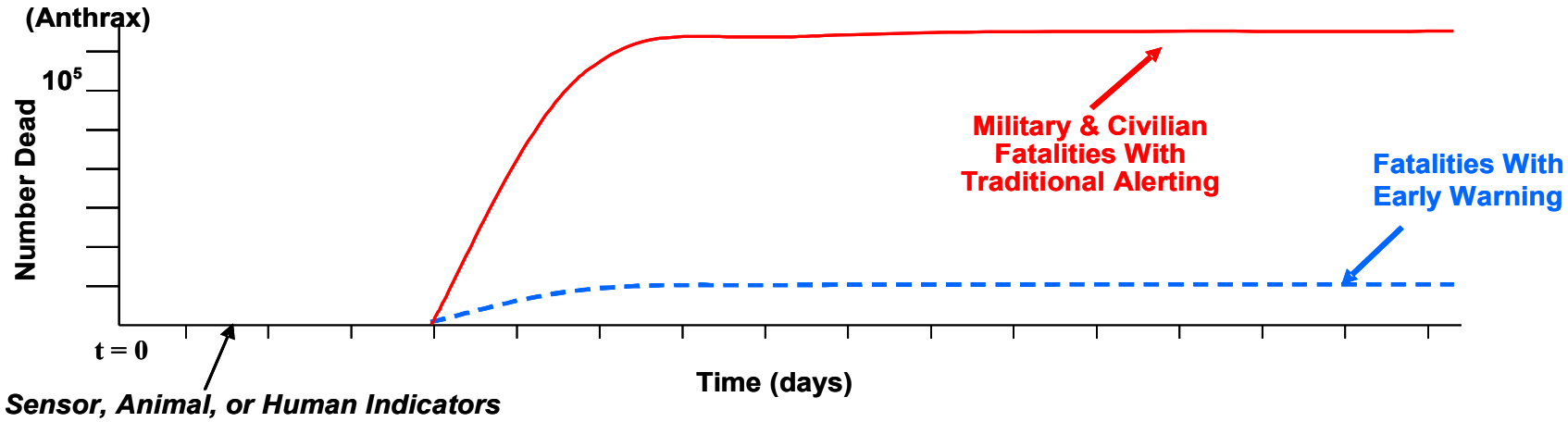
# ESSENCE

## Mission Statement

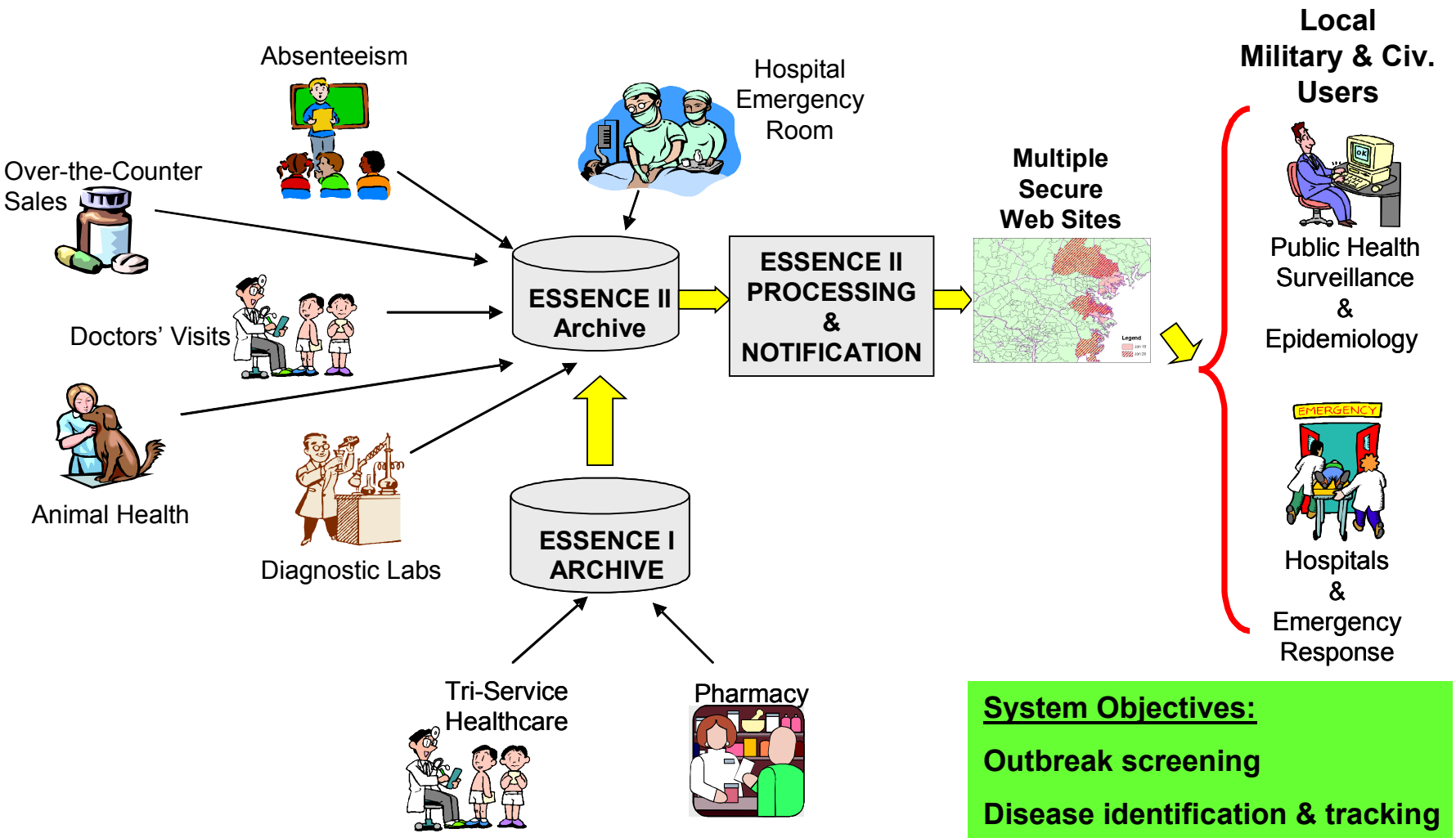
- Electronic Surveillance System for the Early Notification of Community-based Epidemics
  - Provide early warning of abnormal health conditions which may be the result of a Bio-Terrorism or an emerging infectious disease
  - Provide daily medical situational awareness to epidemiologists and health officials

# Motivation for Disease Surveillance (Anthrax)



# ESSENCE Architecture



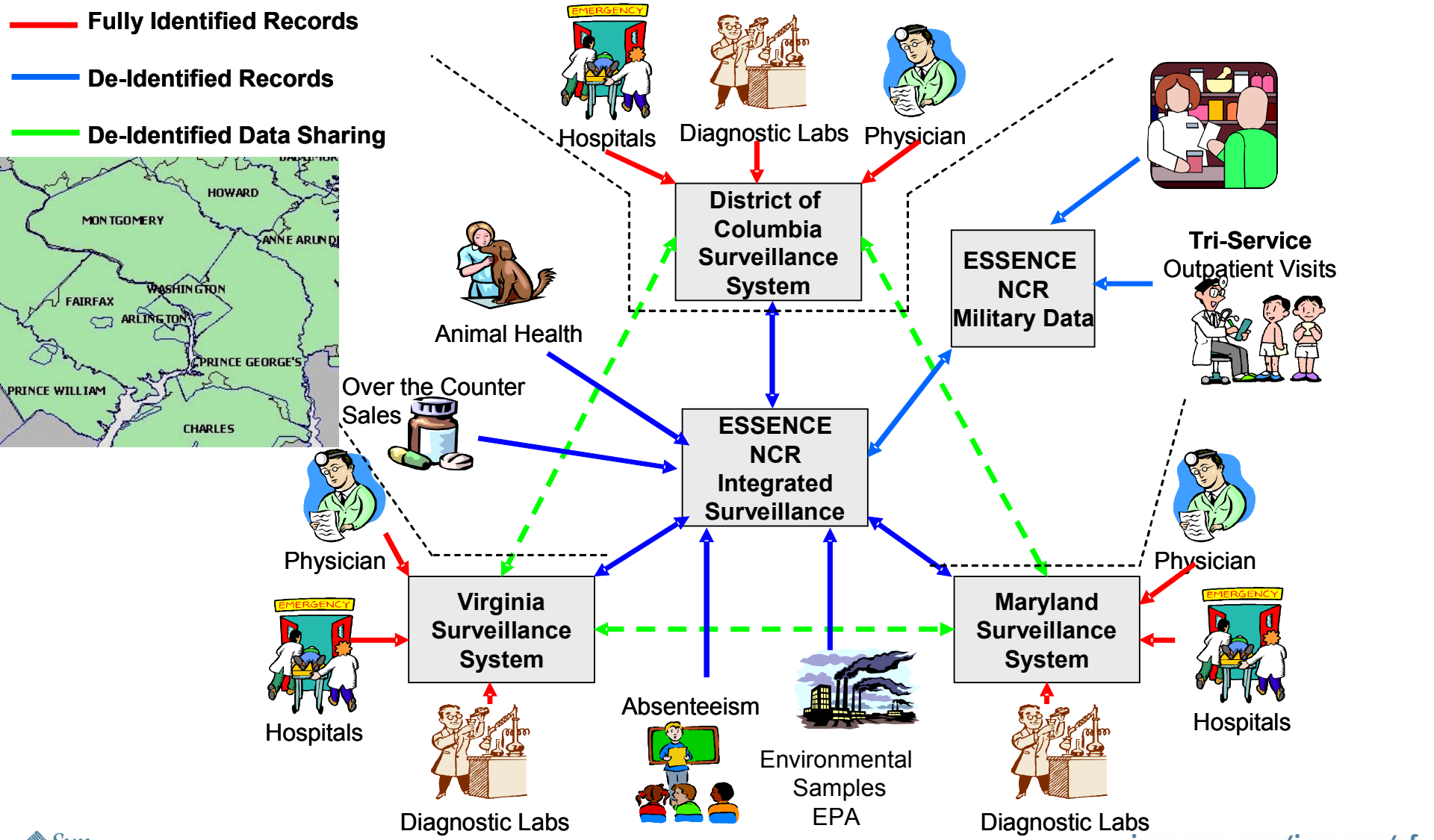

**System Objectives:**

**Outbreak screening**

**Disease identification & tracking**

**Outbreak management**

# National Capital Region Disease Surveillance Network



# Agenda

Background

**Challenges/Solutions**

- **Ingestion—ER Chief Complaint Parser**
- Detection Algorithms
- User Interface

Technologies

Summary

# Emergency Room Chief Complaint

## Challenges and solutions

- Textual description of the patient's reason for visiting the ER (Emergency Room)
- Recorded as part of standard hospital procedure
- Timely—recorded by the triage nurse upon the patient's arrival
- Obtainable in electronic form

# Chief Complaint Textual Artifacts

## Challenges and solutions

- Abbreviations
- Acronyms
- Misspellings
- Negative Context

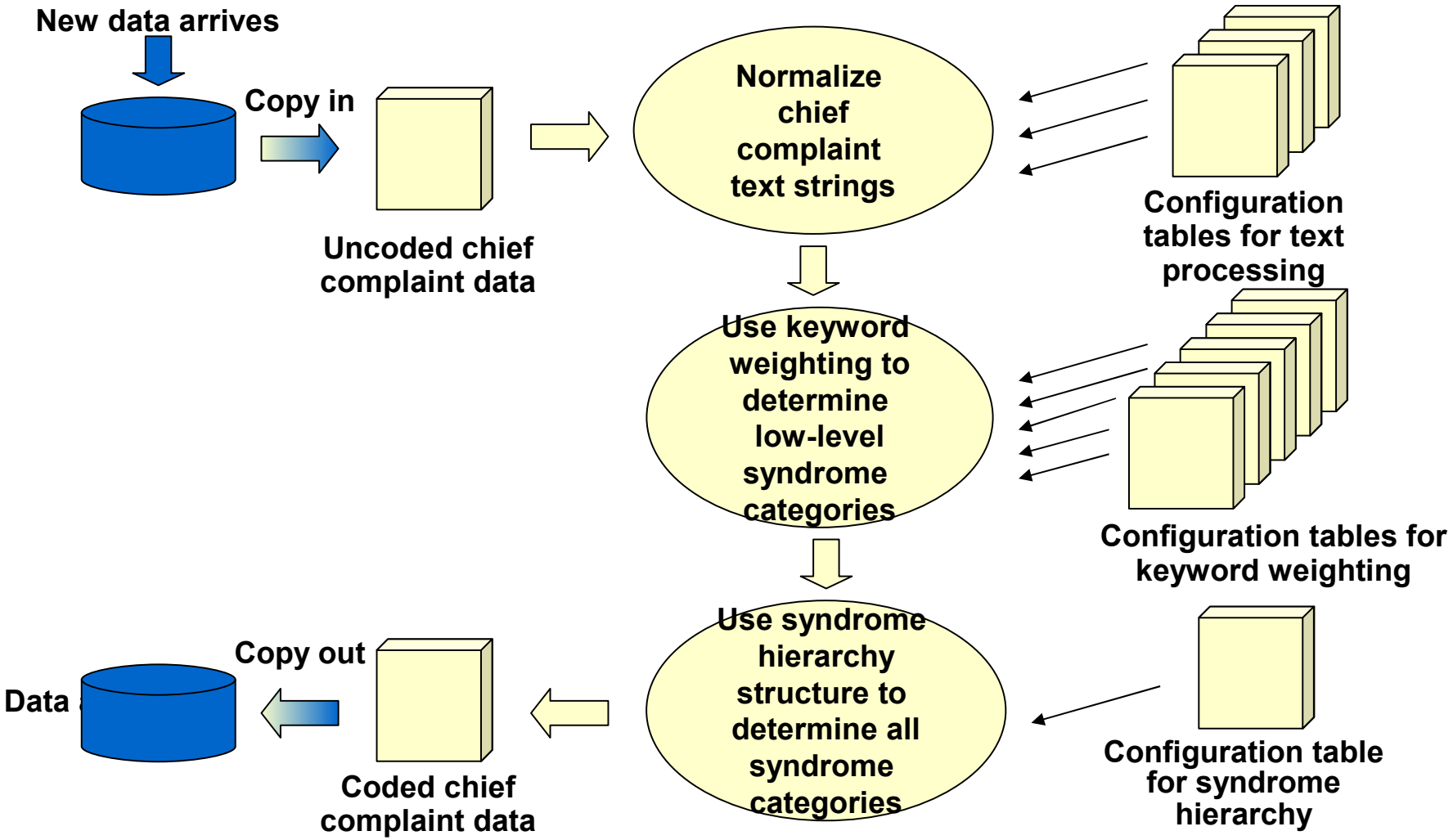
# Chief Complaint Parser's Goal

## Challenges and solutions

- Categorize each CC (Chief Complaint) into syndrome groups for analysis purposes
  - One CC may fall into more than one group

| Syndrome | Example Chief Complaints |
|---|---|
| Gastrointestinal | Coughing/vomiting<br>Ab pain |
| Rash | Spots/welts all over body |
| Respiratory | Coughing/vomiting<br>PNEUMONIA |
| … | … |

# Chief Complaint Text Parsing Process



# Weighted Keyword Matching

## Challenges and solutions

- Keywords are weighted in association with lower level syndrome groups
  - Keywords have both positive and negative weights
  - If a Negative Term such as “NO” or “NOT” precedes than a positive weight is considered negative
  - If the weights’ sum is greater than a configured threshold the CC is associated with the syndrome group

| Group | Keyword | Weight |
|---|---|---|
| Headache | HEADACHE | 10 |
| Headache | HEAD LICE | -4 |

# Weighted Keyword Matching

## Challenges and solutions

- Fuzzy Matching and Pattern Matching
  - Keywords are fuzzy matched to Chief Complaints to one character differences: insertion, deletions, substitutions, and inversions
  - Patterns may be specified for commonly hard-to-spell words

| CC Text | Keyword | Matches? |
|---|---|---|
| Headeche | HEADACHE | Yes |
| Haedeche | HEADACHE | Yes |
| Diarrhea | DIA*A | Yes |
| DIAHRREHA | DIA*A | Yes |

# Fuzzy Matching and Pattern Matching

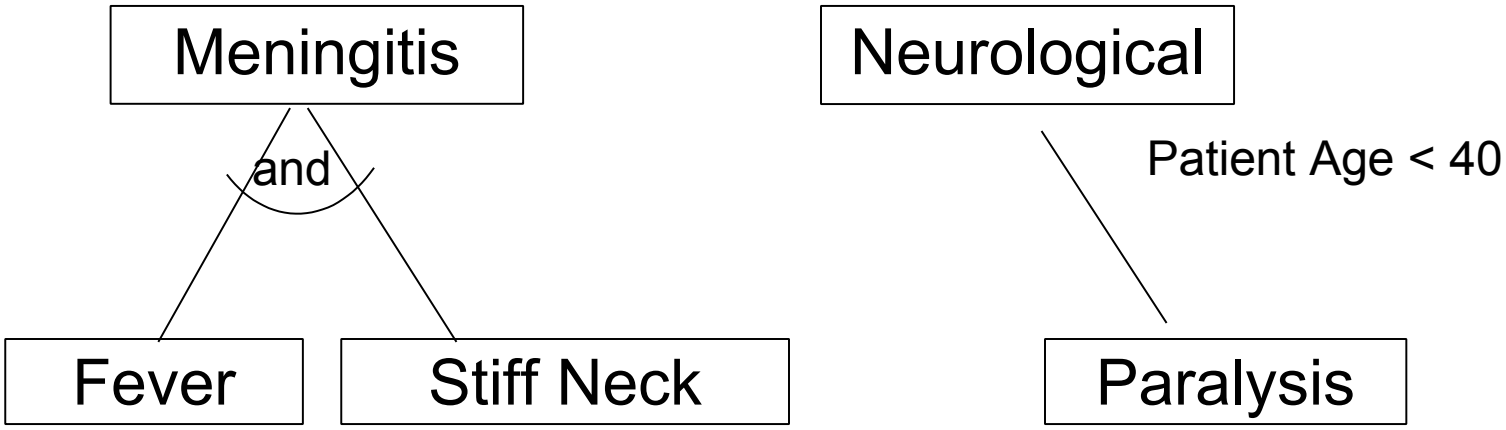
## Challenges and solutions

- Leverage Java™ technology's support of Regular Expressions
- Fuzzy Matching as Regular Expressions
  - Insertion
  - chest = .chest | c.hest | ch.est | che.st | ches.t | chest.
  - Substitution
  - chest = .hest | c.est | ch.st | che.t | ches.
  - Deletion
  - chest = hest | cest | chst | chet | ches
  - Inversion
  - chest = hcest | cehst | chset | chets

# Syndrome Categories

## Challenges and solutions

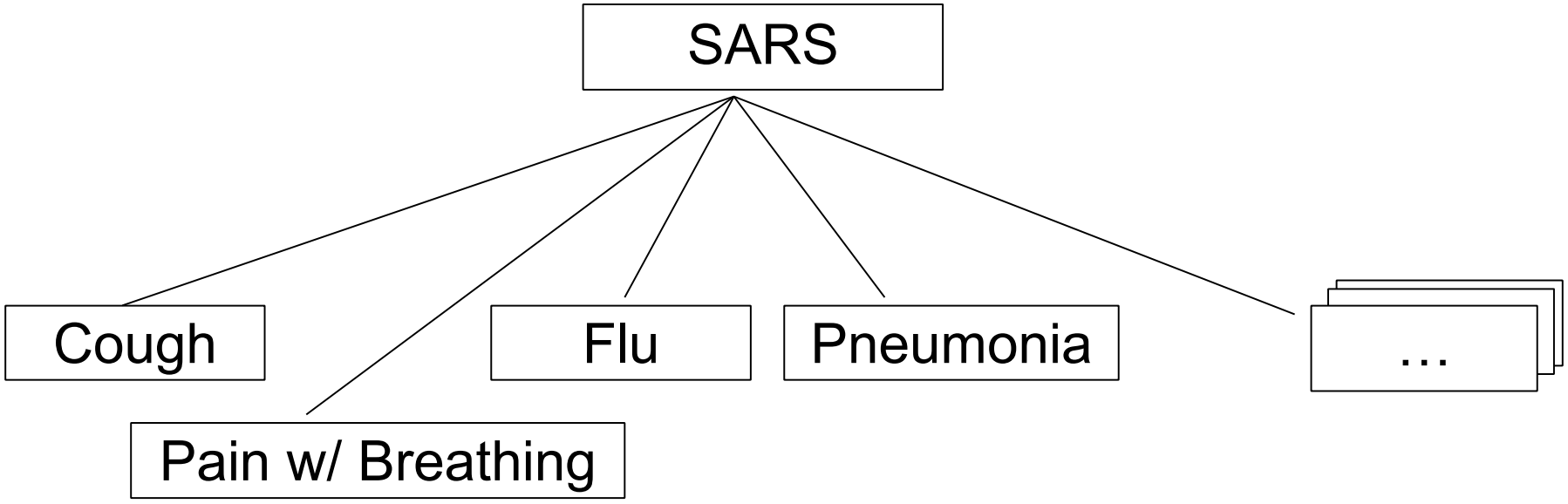
- Top level syndrome categories are defined in terms of lower level groups
  - Defined as logical rules
    - And, Or, Exclusion, Not
    - Simple constraints against Chief Complaint attributes



# Custom Syndrome Categories

## Challenges and solutions

- Users may quickly define custom categories to accommodate surveillance needs



# Query ER Data by Sub Syndrome

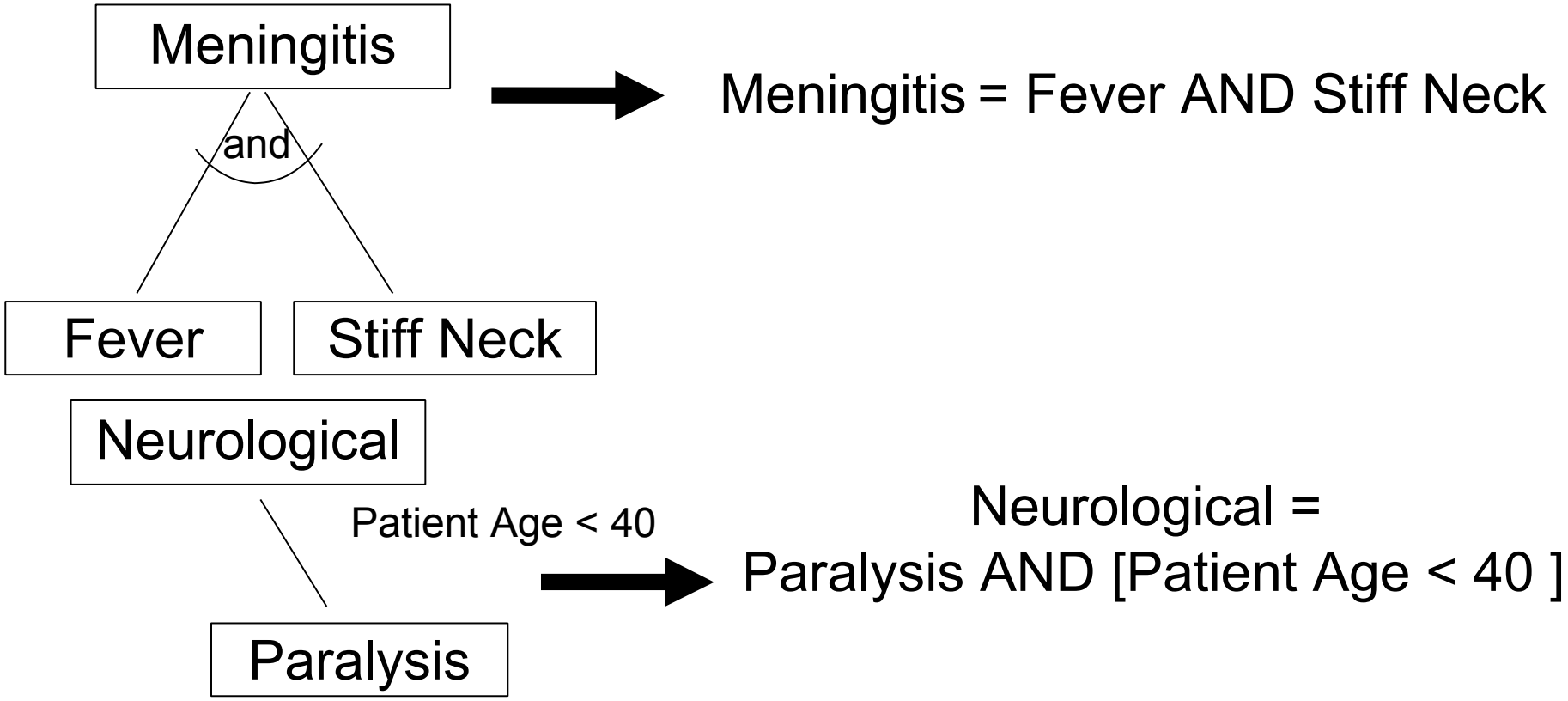
| Current Data Query Selections | | | |
|---|---|---|---|
| **Data Source** | ER by Patient | **Geography System** | Region |
| **Region** | All | **Medical Grouping System** | ChiefComplaintSubSyndromes |

| Next Selections: | | | |
|---|---|---|---|
| **Select ChiefComplaintSubSyndromes:** | All ChiefComplaintSubSyndromess<br>AbdominalCramps<br>AbdominalPain | **Select Detector:** | Regression/EWMA |
| **Select Age Group:** | All Age Groups<br>Unknown<br>0-4 | **Select Sex:** | All Sexs<br>Unknown<br>Male |
| **Select Start Date:** | 02 Feb 05 | **Select End Date:** | 03 May 05 |
| Submit | | | |

# Syndrome Category Rules

## Challenges and solutions

- Represent rules as logical expressions



# Syndrome Category Rules

## Challenges and solutions

- Use Postfix Notation for evaluation
  - Constraints are treated as operands

MaleNeuro =
Paralysis
AND
[Patient Age < 40 ]
AND
[Patient Sex = Male]

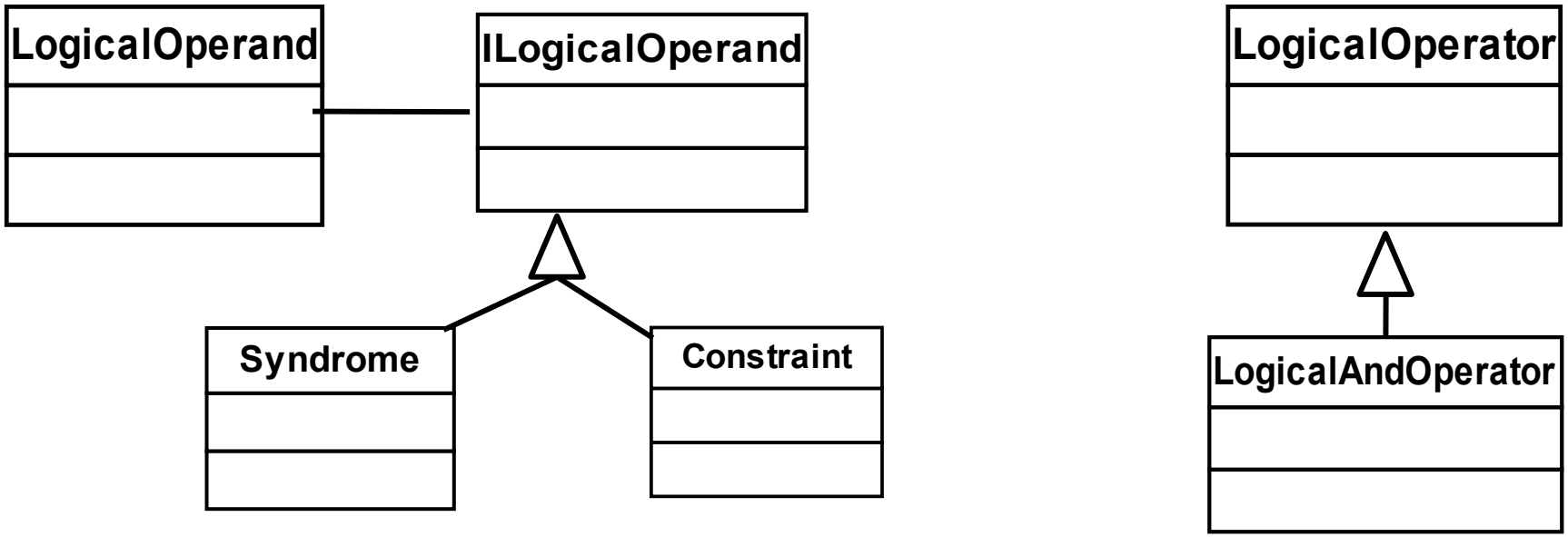
MaleNeuro =
Paralysis
[Patient Age < 40 ]
[Patient Sex = Male]
AND
AND

# Syndrome Category Rules

## Challenges and solutions

- Design Classes to represent and evaluate rules

Neurological = Paralysis AND [Patient Age < 40 ]



# Legacy Chief Complaint Parser

## Challenges and solutions

- Restrictions and limitations
  - Implemented in Microsoft Access
    - Required Windows and Office
  - Cumbersome data input and output
    - Required reading/writing Chief Complaints into tables
    - Forced batch processing
  - Memory constraints and performance make the processing of large data sets difficult
    - 2GB .mdb file limit = ~5000 complaints limit

# Chief Complaint Parser Redesign

## Challenges and solutions

- Redesign goals
  - Abstract data input and output of both reference/ configuration information and Chief Complaints
    - Allow use of arbitrary storage—database tables, CSV, XML
  - Stream processing of Chief Complaints
    - Light-weight process flow
  - Dynamic invocation of processing steps
    - Allow flexibility to add new steps and control order

# Dynamic Setup of Text Normalization

```
# Text Normalizer Classes specifies the ordered list of classes that each
# perform a step within the overall text normalization process.
#
# Each normalization class must implement the ITextNormalizer interface. The
# order that the classes are listed here will define the order of execution.

TN-1 = edu.jhuapl.bsp.ccp.TextNormalization.UpperCaseNormalizer
TN-2 = edu.jhuapl.bsp.ccp.TextNormalization.PunctuationNormalizer
TN-3 = edu.jhuapl.bsp.ccp.TextNormalization.AbbreviationNormalizer
TN-4 = edu.jhuapl.bsp.ccp.TextNormalization.StopWordNormalizer
TextNormalizerClasses = TN-1,TN-2,TN-3,TN-4
```

# TextNormalizer

```
public class TextNormalizer {
  private ArrayList steps;
  public TextNormalizer(…) {
    steps = new ArrayList();
    Iterator it = configuration.getTextNormalizerClasses().iterator();
    while (it.hasNext()) {
      String className = (String)it.next();
      // find class and create instance
      Class c = Class.forName(className);
      Object o = c.newInstance();
      //cast and initialize
      ITextNormalizer tn = (ITextNormalizer)o;
      …
      steps.add(tn);
    }
```

# Agenda

Background

**Challenges/Solutions**

- Data Ingestion
- **Detection Algorithms**
- User Interface

Technologies

Summary

# Finding an Anomaly

## Challenges and solutions

- Detectors use statistical algorithms to find anomalous events in the health data
- To understand what is anomalous behavior we must understand normal behavior

# Finding an Anomaly

## Challenges and solutions

- Temporal detectors model expected counts based on past data and modelling; they quantify the degree of anomaly as a detection level (p-values for a common scale)
- If the detection level meets certain thresholds, it is flagged either a red or yellow alert
- Alerts are displayed to the user and are indicators which may prompt further investigation

# Temporal Alerting Algorithm

## Challenges and solutions

- Multiple algorithm approach
- Adaptive modelling and process control
  - Regression model
    - Controls for weekly, seasonal, and holiday effects
  - Exponential weighting moving average/ Shewhart chart
    - Effective for data or residuals without systematic trends or cycles

# Temporal Alert List

**Simulated Data**

| Region/Syndrome Based Temporal Alerts | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Links | Date | Data Source | Region | Age Group | Sex | Syndrome | Detector | Level | Count | Expected | RareColor | RareLevel | NonZero |
| Time Series | 30Apr05 | ER by Patient | ALEXANDRIA | 18-44 | All | Respiratory | Regression/EWMA | 0.006 | 15 | 7.5 | 33 | 7 | 100 |
| Time Series | 30Apr05 | ER by Patient | ALEXANDRIA | 45-64 | All | Respiratory | Regression/EWMA | 0.044 | 4 | 2.821 | 26 | 25 | 94.521 |
| Time Series | 30Apr05 | ER by Patient | ALEXANDRIA | All | All | Respiratory | Regression/EWMA | 0.003 | 27 | 18.214 | 41 | 6 | 100 |
| Time Series | 30Apr05 | ER by Patient | ARLINGTON | All | All | Respiratory | Regression/EWMA | 0.001 | 51 | 24.321 | 38 | 4 | 100 |
| Time Series | 30Apr05 | ER by Patient | ARLINGTON | 18-44 | All | Respiratory | Regression/EWMA | 0.001 | 27 | 10.964 | 32 | 2 | 100 |
| Time Series | 30Apr05 | ER by Patient | ARLINGTON | 45-64 | All | Respiratory | Regression/EWMA | 0.031 | 7 | 4.179 | 28 | 20 | 98.082 |
| Time Series | 30Apr05 | ER by Patient | ARLINGTON | 65+ | All | Respiratory | Regression/EWMA | 0.027 | 7 | 3.179 | 28 | 18 | 94.521 |
| Time Series | 30Apr05 | ER by Patient | FAIRFAX | All | All | Respiratory | Regression/EWMA | 0.001 | 265 | 124.786 | 52 | 11 | 100 |
| Time Series | 30Apr05 | ER by Patient | FAIRFAX | 18-44 | All | Respiratory | Regression/EWMA | 0.001 | 113 | 46.607 | 42 | 9 | 100 |
| Time Series | 30Apr05 | ER by Patient | FAIRFAX | 45-64 | All | Respiratory | Regression/EWMA | 0.001 | 78 | 20.679 | 36 | 3 | 100 |
| Time Series | 30Apr05 | ER by Patient | FAIRFAX | 65+ | All | Respiratory | Regression/EWMA | 0.001 | 41 | 12.286 | 37 | 3 | 100 |
| Time Series | 30Apr05 | ER by Patient | LOUDOUN | All | All | Respiratory | Regression/EWMA | 0.001 | 41 | 17 | 43 | 4 | 100 |
| Time Series | 30Apr05 | ER by Patient | LOUDOUN | 18-44 | All | Respiratory | Regression/EWMA | 0.001 | 17 | 5.929 | 27 | 3 | 100 |
| Time Series | 30Apr05 | ER by Patient | LOUDOUN | 45-64 | All | Respiratory | Regression/EWMA | 0.012 | 6 | 2.214 | 20 | 11 | 87.671 |
| Time Series | 30Apr05 | ER by Patient | LOUDOUN | 65+ | All | Respiratory | Regression/EWMA | 0.001 | 7 | 1.286 | 26 | 3 | 68.219 |
| Time Series | 30Apr05 | ER by Patient | PRINCE WILLIAM | All | All | Respiratory | Regression/EWMA | 0.001 | 79 | 45.214 | 56 | 5 | 100 |
| Time Series | 30Apr05 | ER by Patient | PRINCE WILLIAM | 18-44 | All | Respiratory | Regression/EWMA | 0.001 | 33 | 17.179 | 37 | 4 | 100 |
| Time Series | 30Apr05 | ER by Patient | PRINCE WILLIAM | 45-64 | All | Respiratory | Regression/EWMA | 0.008 | 13 | 6.357 | 27 | 5 | 99.726 |
| Time Series | 30Apr05 | ER by Patient | PRINCE WILLIAM | 65+ | All | Respiratory | Regression/EWMA | 0.001 | 15 | 3.536 | 30 | 1 | 96.164 |

# Fusion Alerting Algorithm

## Challenges and solutions

- Fusion capability for separate data sources, regions
  - Statically fuses multiple detection levels to discover new alerts
  - Based on output from the Temporal Alerting Algorithm

# Fusion Alert List

**Simulated Data**

| Region/Syndrome Based Temporal Fusion Alerts | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Date | Data Source | Region | Age | Sex | Syndrome | Level | Links |
| − | 26Apr05 | Fusion | WASHINGTON | All | All | Respiratory | 0.001 | |
| | 26Apr05 | Emergency Room Data by Patient Location | WASHINGTON | All | All | Respiratory | 0.076 | Time Series |
| | 26Apr05 | Military Outpatient Visits | WASHINGTON | All | All | Respiratory | 0.035 | Time Series |
| | 26Apr05 | Over-the-Counter Chain 3 | WASHINGTON | All | All | Respiratory | 0.039 | Time Series |
| + | 20Apr05 | Fusion | PRINCE WILLIAM | 65+ | All | Respiratory | 0.008 | |
| + | 19Apr05 | Fusion | WASHINGTON | 18-44 | All | Respiratory | 0.01 | |
| − | 26Apr05 | Fusion | PRINCE WILLIAM | 18-44 | All | Respiratory | 0.025 | |
| | 26Apr05 | Emergency Room Data by Patient Location | PRINCE WILLIAM | 18-44 | All | Respiratory | 0.101 | Time Series |
| | 26Apr05 | Military Outpatient Visits | PRINCE WILLIAM | 18-44 | All | Respiratory | 0.123 | Time Series |
| + | 18Apr05 | Fusion | WASHINGTON | 18-44 | All | Respiratory | 0.026 | |
| + | 21Apr05 | Fusion | PRINCE WILLIAM | 45-64 | All | Respiratory | 0.035 | |
| + | 26Apr05 | Fusion | PRINCE WILLIAM | All | All | Respiratory | 0.039 | |
| + | 26Apr05 | Fusion | WASHINGTON | 18-44 | All | Respiratory | 0.044 | |
| + | 20Apr05 | Fusion | WASHINGTON | 18-44 | All | Respiratory | 0.045 | |
| + | 21Apr05 | Fusion | PRINCE WILLIAM | 65+ | All | Respiratory | 0.05 | |

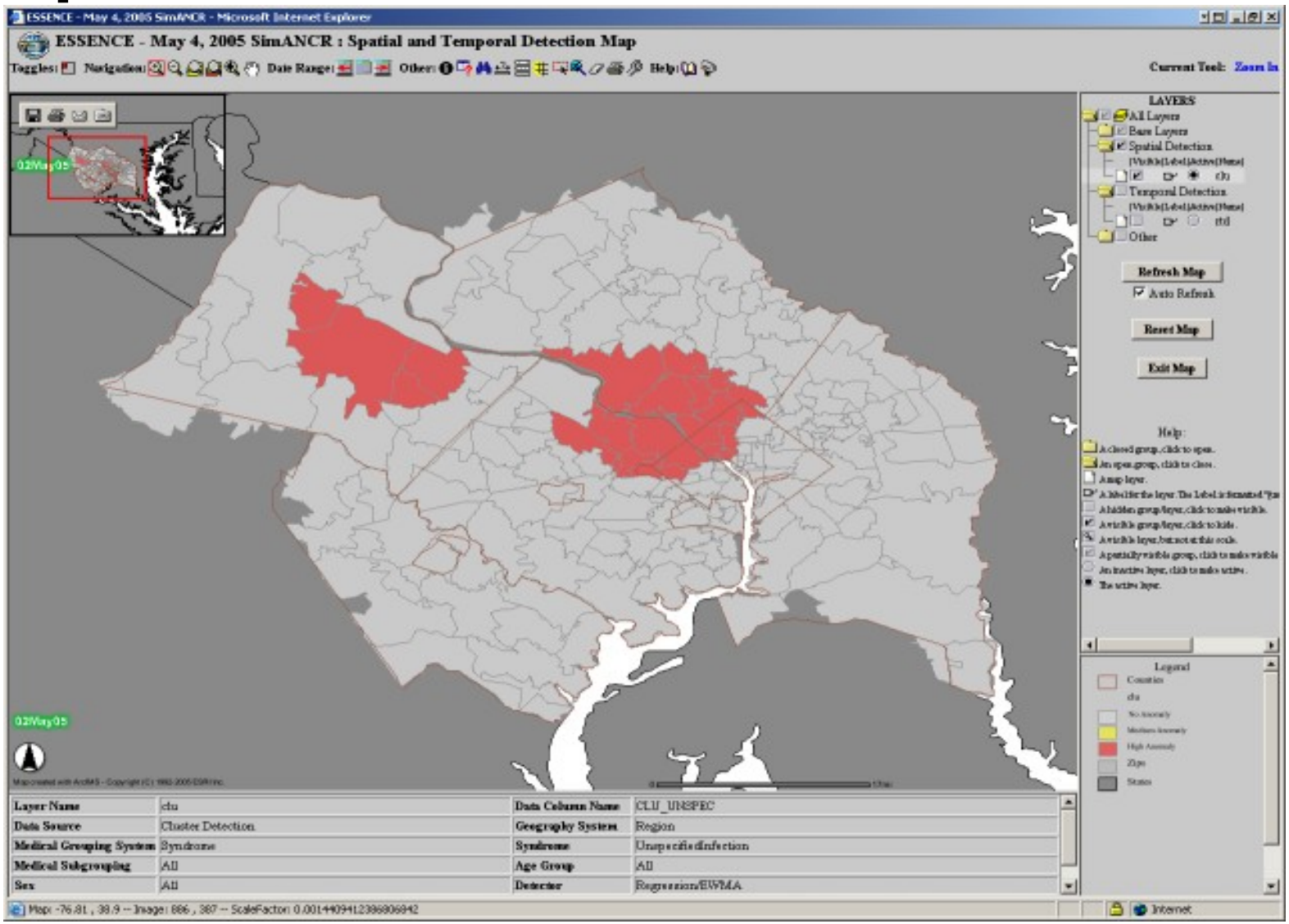
# Spatiotemporal Alerting Algorithms

## Challenges and solutions

- Searches for clusters of cases that are spatially significant relative to expected spatial distribution
- Uses spatiotemporal scan statistics, based on Kulldorff's SaTScan methodology, applied to health surveillance for the National Cancer Institute since 1980s

# Spatial Alerts

**Simulated Data**



# Detector Development

## Challenges and solutions

- Statisticians are the algorithm developers and experts
  - Detection algorithms are prototyped in environments such as MATLAB or other statistical software packages such as SAS, S+…
  - The detector implementation must not depend on those environments

# Detector Development

## Challenges and solutions

- Choices for MATLAB to Java Technology Development
  - Use MATLAB Java technology-based API
    - Only allows executing Java technology from within MATLAB; the reverse is unsupported; also requires MATLAB to be running
  - Compile MATLAB code to C++ and use Java Native Interface (JNI) or command line calls
    - Requires MATLAB binaries
  - Use a Client Server Model
    - Requires a server which would ultimately use one of the above options
  - Re-implement entirely in Java programming language

# Detector Development

## Challenges and solutions

- We chose to re-implement entirely in Java programming language
- JMatLab Java Library
  - MATLAB-like methods
  - Utility methods for manipulating arrays or lists of numbers

```
/**
  * Simulates MATLAB std function. Calculate the standard

  * deviation of the array.
  */
  public static double std (double[] array) {
    …
  }
```

# Detector Development

## Challenges and solutions

- Handling precision
  - MATLAB may produce numbers with very large precision that are not representable as a primitive Java technology double
  - BigInteger and BigDecimal Classes can represent very large numbers or numbers with very large precision

# Detector Development

## Challenges and solutions

- Handling precision
  - Is the extra precision needed? No, the differences were very small
  - However, for testing, having exact matching output ensures correctness
  - Using BigDecimal made the detector run slower; it is used in testing only

# Detector Development

## Challenges and solutions

- Temporal detector is required to run in two modes
- Stand alone scheduled execution
  - Iterates over every combination of—data sources, regions, age groups, syndromes, …
  - Queries data based on the above strata and runs detection; any alerts found are written to the database
- On demand via ESSENCE User Interface
  - A user has queried a specific set of data and requests detection to be run

# TemporalDetectorInterface

## Challenges and solutions

```
public interface TemporalDetectorInterface {

    public void runDetector(TemporalDetectorDataInterface tddi);

    public String getID();
    public String getName();

    public double getRedLevel();
    public void setRedLevel(double _redLevel);

    public double getYellowLevel();
    public void setYellowLevel(double _yellowLevel);
}
```

# TemporalDetectorDataInterface

## Challenges and solutions

```
public interface TemporalDetectorDataInterface {

    public void setCounts(double[] _counts);
    public void setStartDate(java.util.Date _startDate);
    public void setRegressor(String regressorID, double[] _regressor);

    public String[] getAltTexts();
    public double[] getLevels();
    public double[] getExpecteds();
    public double[] getColors();
...
}
```

# Detector Controllers

## Challenges and solutions

- Defer higher level knowledge and management to detector controller
  - Queries each required dataset based on strata
  - Setup each detector run
  - Optimize overall performance through threading, batching, and caching data

# Detector Testing

## Challenges and solutions

- Establish a common testing framework
  - Make use of both test cases created from MATLAB and Java technology development
    - Java language Detectors are tested against MATLAB test cases via MATLAB Java technology-based API
    - Statisticians can independently run tests cases
  - When possible, leverage operational ESSENCE systems and use real data

# Detector Enhancements

## Challenges and solutions

- Improvements originated from software testing and practical application
- Data Dropouts
  - Data providers may miss sending data; that data may be unrecoverable
- Initial Startup
  - New installations may have no historical data

# Agenda

Background

**Challenges/Solutions**

- Data Ingestion
- Detection Algorithms
- **User Interface**

Technologies

Summary

# User Interface Development

## Challenges and solutions

- Multiple data sources
  - ESSENCE handles various data sources
    - ER Chief Complaints
    - Over-the-Counter Drug Sales
    - School Absenteeism
    - Etc.
  - The UI must support each data source consistently

# User Interface Development

## Challenges and solutions

- DataSource Interface
  - UI is generated based on DataSource objects
- DataSource Objects are responsible
  - Querying and retrieving their data
  - Describing their applicable querying parameters
    - – zip code,  hospital name, OTC store name …
  - Maintaining state of the query parameters

# User Interface Design

## Challenges and solutions

- Designing the UI
  - User working groups
  - User surveys
  - User training and exercises
    - Observe and record any problems the user experiences
  - Prototype UI and present to users

# User Interface Design

## Challenges and solutions

- Key user interface areas
  - Detector alert displays
  - Data querying and detailed display
  - Map displays

# User Interface Development

## Challenges and solutions

- User feedback indicated the need for an easier way to visualize alerts
- Users may have many alerts to investigate due to:
  - Multiple data sources
  - Combinations of alerting strata
    - Age group, gender, region

# Temporal Alert List

**Simulated Data**

Region/Syndrome Based Temporal Alerts

| Links | Date | Data Source | Region | Age Group | Sex | Syndrome | Detector | Level | Count | Expected | RareColor | RareLevel | NonZero |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Time Series | 30Apr05 | ER by Patient | ALEXANDRIA | 18-44 | All | Respiratory | Regression/EWMA | 0.006 | 15 | 7.5 | 33 | 7 | 100 |
| Time Series | 30Apr05 | ER by Patient | ALEXANDRIA | 45-64 | All | Respiratory | Regression/EWMA | 0.044 | 4 | 2.821 | 26 | 25 | 94.521 |
| Time Series | 30Apr05 | ER by Patient | ALEXANDRIA | All | All | Respiratory | Regression/EWMA | 0.003 | 27 | 18.214 | 41 | 6 | 100 |
| Time Series | 30Apr05 | ER by Patient | ARLINGTON | All | All | Respiratory | Regression/EWMA | 0.001 | 51 | 24.321 | 38 | 4 | 100 |
| Time Series | 30Apr05 | ER by Patient | ARLINGTON | 18-44 | All | Respiratory | Regression/EWMA | 0.001 | 27 | 10.964 | 32 | 2 | 100 |
| Time Series | 30Apr05 | ER by Patient | ARLINGTON | 45-64 | All | Respiratory | Regression/EWMA | 0.031 | 7 | 4.179 | 28 | 20 | 98.082 |
| Time Series | 30Apr05 | ER by Patient | ARLINGTON | 65+ | All | Respiratory | Regression/EWMA | 0.027 | 7 | 3.179 | 28 | 18 | 94.521 |
| Time Series | 30Apr05 | ER by Patient | FAIRFAX | All | All | Respiratory | Regression/EWMA | 0.001 | 265 | 124.786 | 52 | 11 | 100 |
| Time Series | 30Apr05 | ER by Patient | FAIRFAX | 18-44 | All | Respiratory | Regression/EWMA | 0.001 | 113 | 46.607 | 42 | 9 | 100 |
| Time Series | 30Apr05 | ER by Patient | FAIRFAX | 45-64 | All | Respiratory | Regression/EWMA | 0.001 | 78 | 20.679 | 36 | 3 | 100 |
| Time Series | 30Apr05 | ER by Patient | FAIRFAX | 65+ | All | Respiratory | Regression/EWMA | 0.001 | 41 | 12.286 | 37 | 3 | 100 |
| Time Series | 30Apr05 | ER by Patient | LOUDOUN | All | All | Respiratory | Regression/EWMA | 0.001 | 41 | 17 | 43 | 4 | 100 |
| Time Series | 30Apr05 | ER by Patient | LOUDOUN | 18-44 | All | Respiratory | Regression/EWMA | 0.001 | 17 | 5.929 | 27 | 3 | 100 |
| Time Series | 30Apr05 | ER by Patient | LOUDOUN | 45-64 | All | Respiratory | Regression/EWMA | 0.012 | 6 | 2.214 | 20 | 11 | 87.671 |
| Time Series | 30Apr05 | ER by Patient | LOUDOUN | 65+ | All | Respiratory | Regression/EWMA | 0.001 | 7 | 1.286 | 26 | 3 | 68.219 |
| Time Series | 30Apr05 | ER by Patient | PRINCE WILLIAM | All | All | Respiratory | Regression/EWMA | 0.001 | 79 | 45.214 | 56 | 5 | 100 |
| Time Series | 30Apr05 | ER by Patient | PRINCE WILLIAM | 18-44 | All | Respiratory | Regression/EWMA | 0.001 | 33 | 17.179 | 37 | 4 | 100 |
| Time Series | 30Apr05 | ER by Patient | PRINCE WILLIAM | 45-64 | All | Respiratory | Regression/EWMA | 0.008 | 13 | 6.357 | 27 | 5 | 99.726 |
| Time Series | 30Apr05 | ER by Patient | PRINCE WILLIAM | 65+ | All | Respiratory | Regression/EWMA | 0.001 | 15 | 3.536 | 30 | 1 | 96.164 |

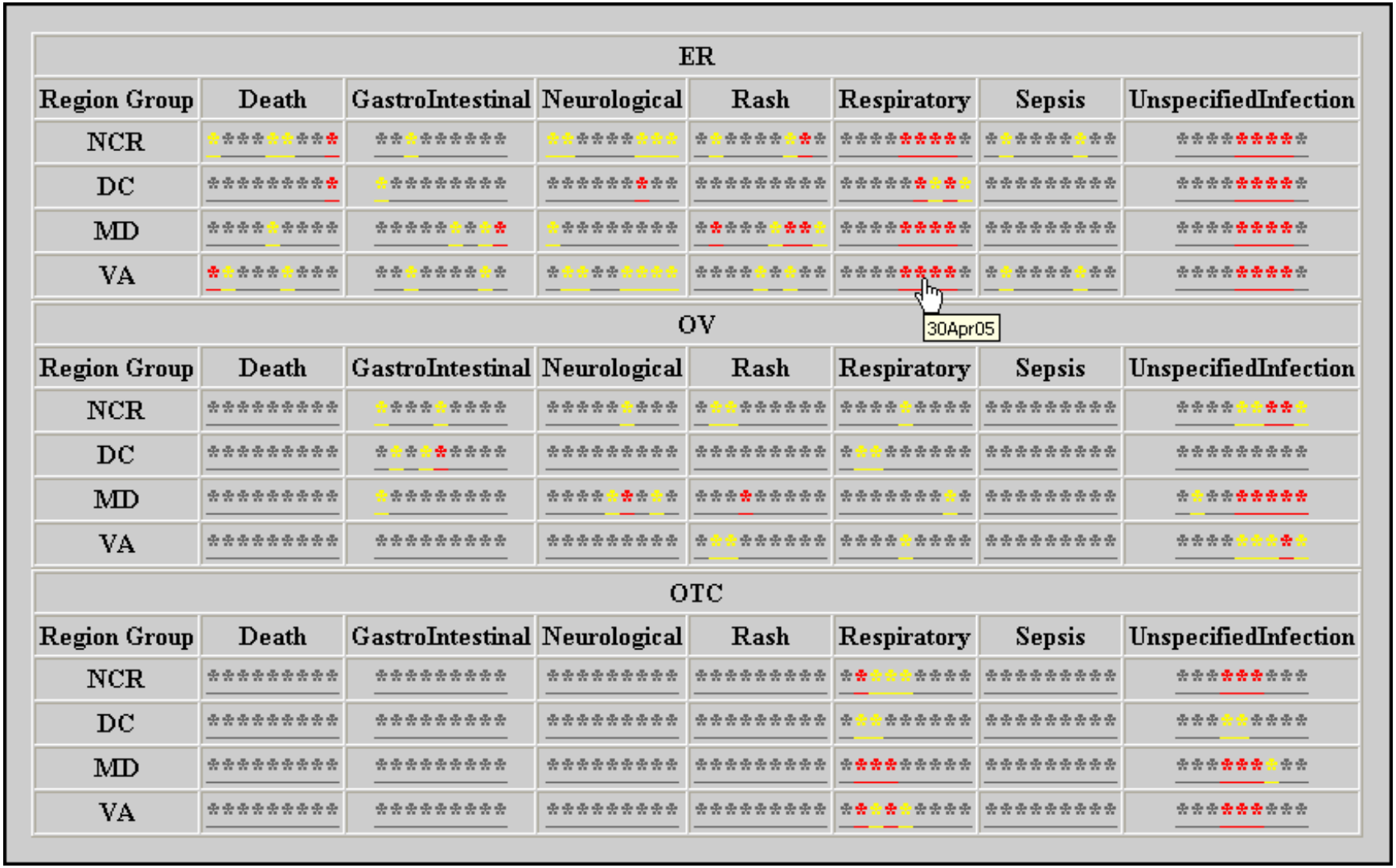
# User Interface Development

## Challenges and solutions

- Summary alert list
  - Concise view of all alerts
  - Provides overall alert status
    - Directed the development of a Summary Detector
  - Allows recognition of patterns

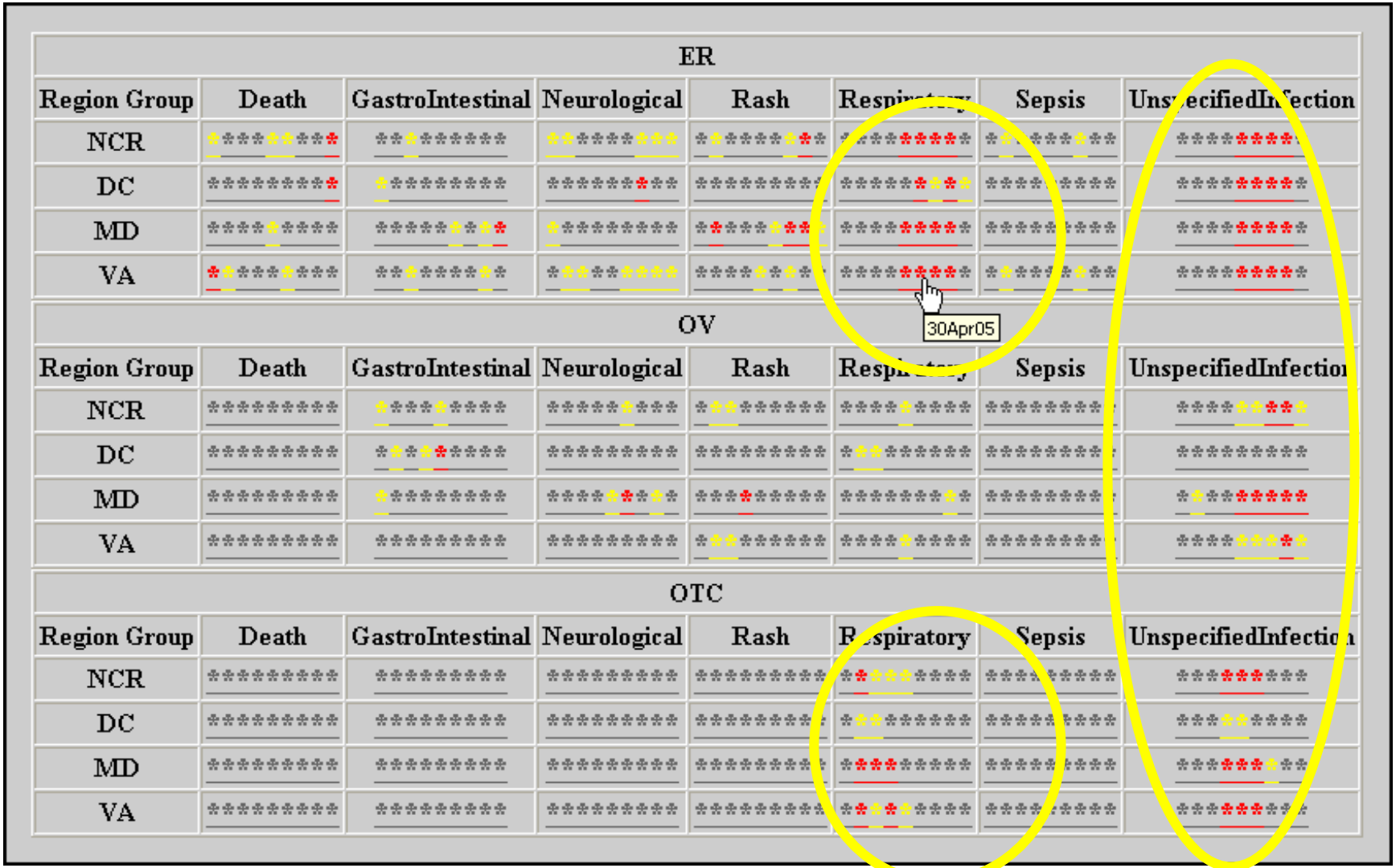
# Summary Alert List

**Simulated Data**



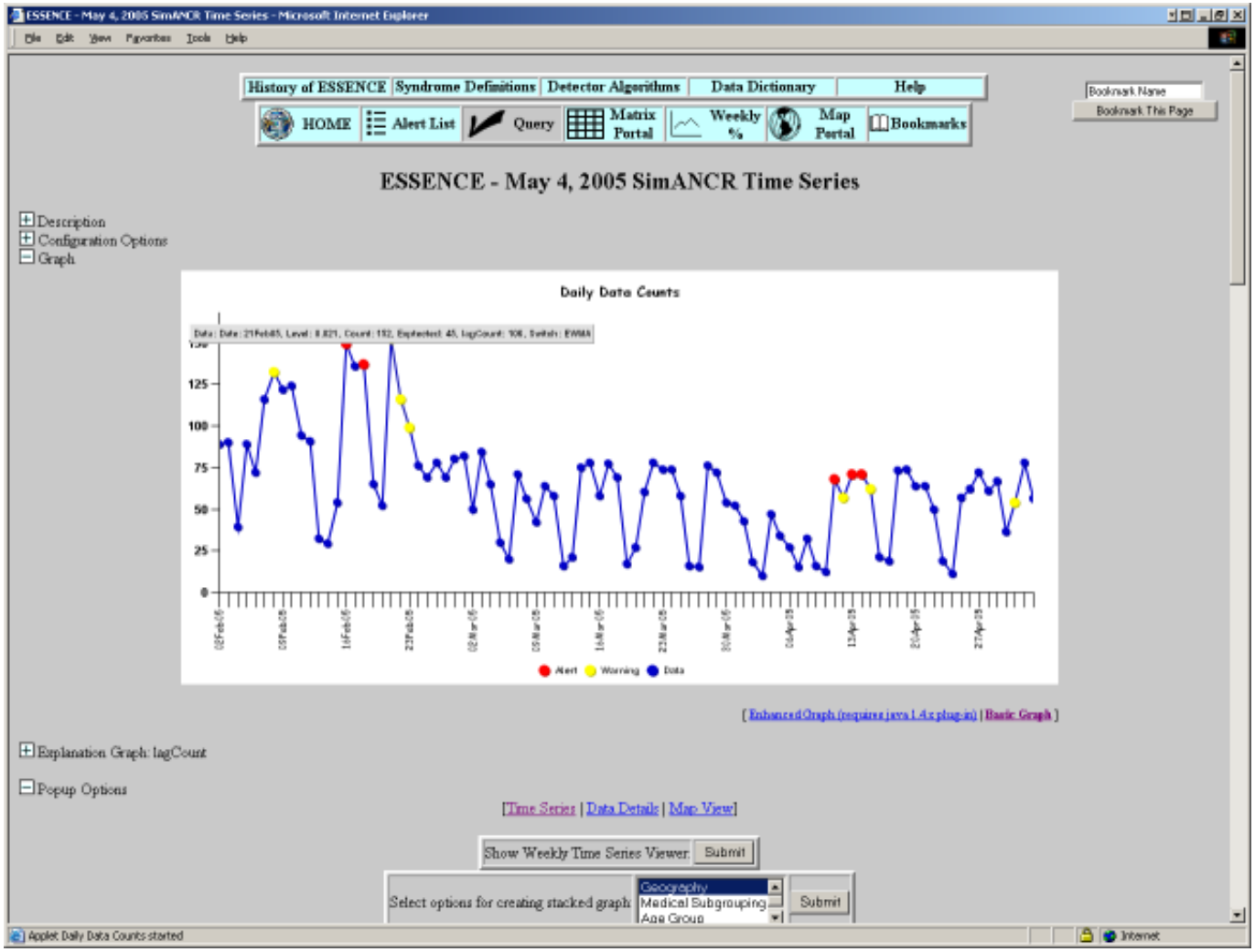
# Summary Alert List

**Simulated Data**



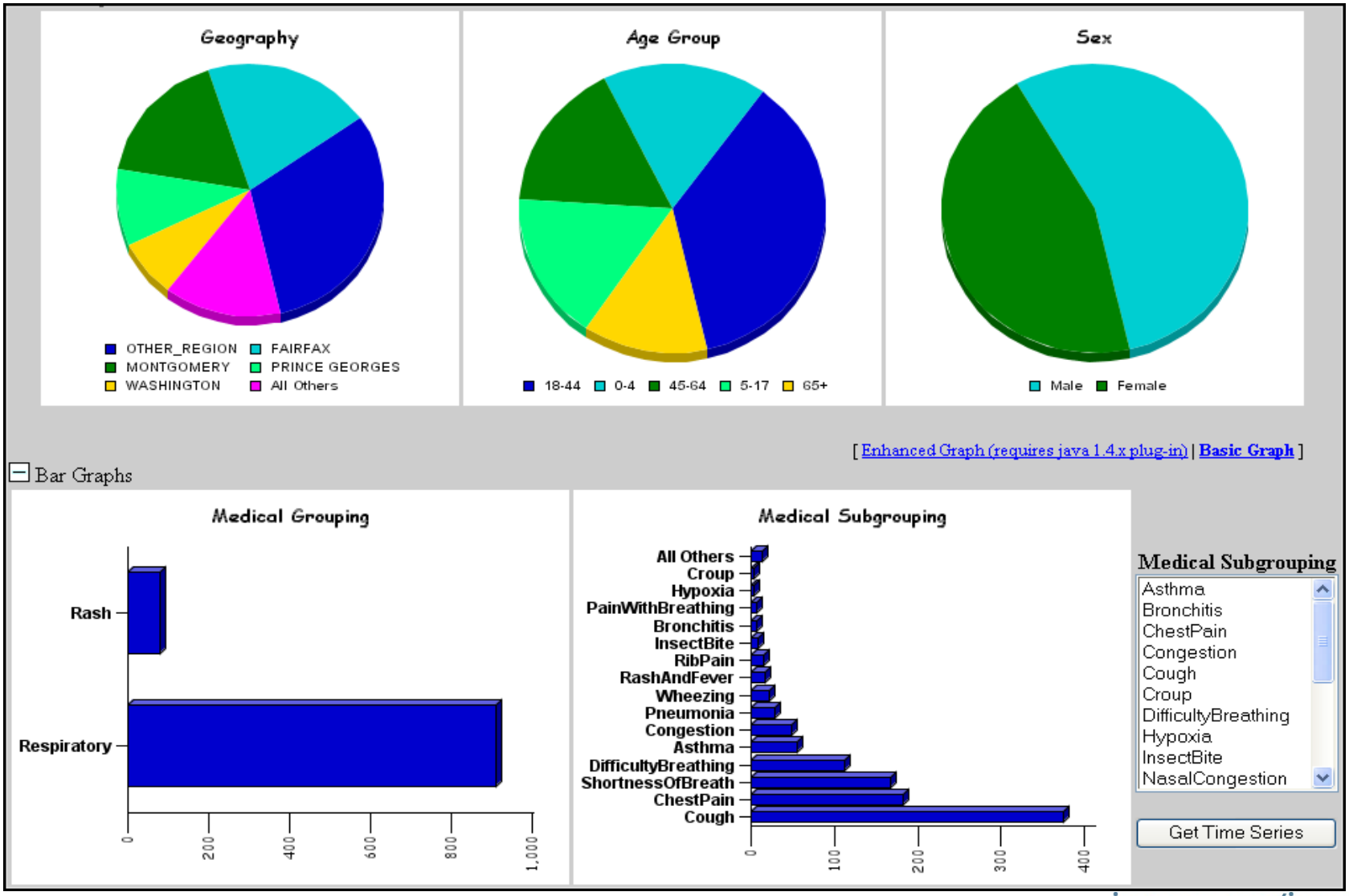
# Time Series

**Simulated Data**



# Data Details

**Simulated Data**



# Agenda

Background

Challenges/Solutions

- Data Ingestion
- Detection Algorithms
- User Interface

**Technologies**

Summary

# Technologies

## Technology summary

- Freely available technologies
  - Java
  - Ant
  - Apache Web Server
  - Apache Tomcat
  - Apache Jakarta Commons
  - Apache Jakarta POI
  - Apache Axis
  - Eclipse
  - GeoTools

# Technologies

## Technology summary

- Commercial technologies
  - ArcIMS
  - Microsoft SQL Server
  - NetCharts

# Agenda

Background

Challenges/Solutions

- Data Ingestion
- Detection Algorithms
- User Interface

Technologies

**Summary**

# Summary

- ESSENCE
  - Provide early warning of abnormal health conditions which may be the result of a Bio-Terrorism or an emerging infectious disease
  - Provide daily medical situational awareness to epidemiologists and health officials

# Summary

- Weighted keyword matching works well for categorizing short phrases/sentences
- Multiple development environments require a common testing framework
- Leverage user feedback to aid in UI design

# ESSENCE Team

- Special thanks to the ESSENCE team
  - Joe Lombardo
  - Sheri Lewis
  - Marty Sikes
  - Raj Ashar
  - Logan Hauenstein
  - Wayne Loschen
  - Carol Sniegoski
  - Nathaniel Tabernero
  - Rich Wojcik
  - Jackie Coberly
  - Brian Feighner
  - Rekha Holtry
  - Steve Babin
  - Howard Burkom
  - Michael Thompson

# Profile of the Applied Physics Laboratory

- Not-for-profit university research and development laboratory
- Division of The Johns Hopkins University founded in 1942
- On-site graduate engineering program in 8 degree fields
- Staffing: 3,600 employees (66% scientists and engineers)
- Annual revenue ~ $680M

# Q&A

# The ESSENCE of Disease Surveillance


## Nathaniel Tabernero

Software Engineer

The Johns Hopkins University Applied Physics Laboratory
http://www.jhuapl.edu/


TS-5564