

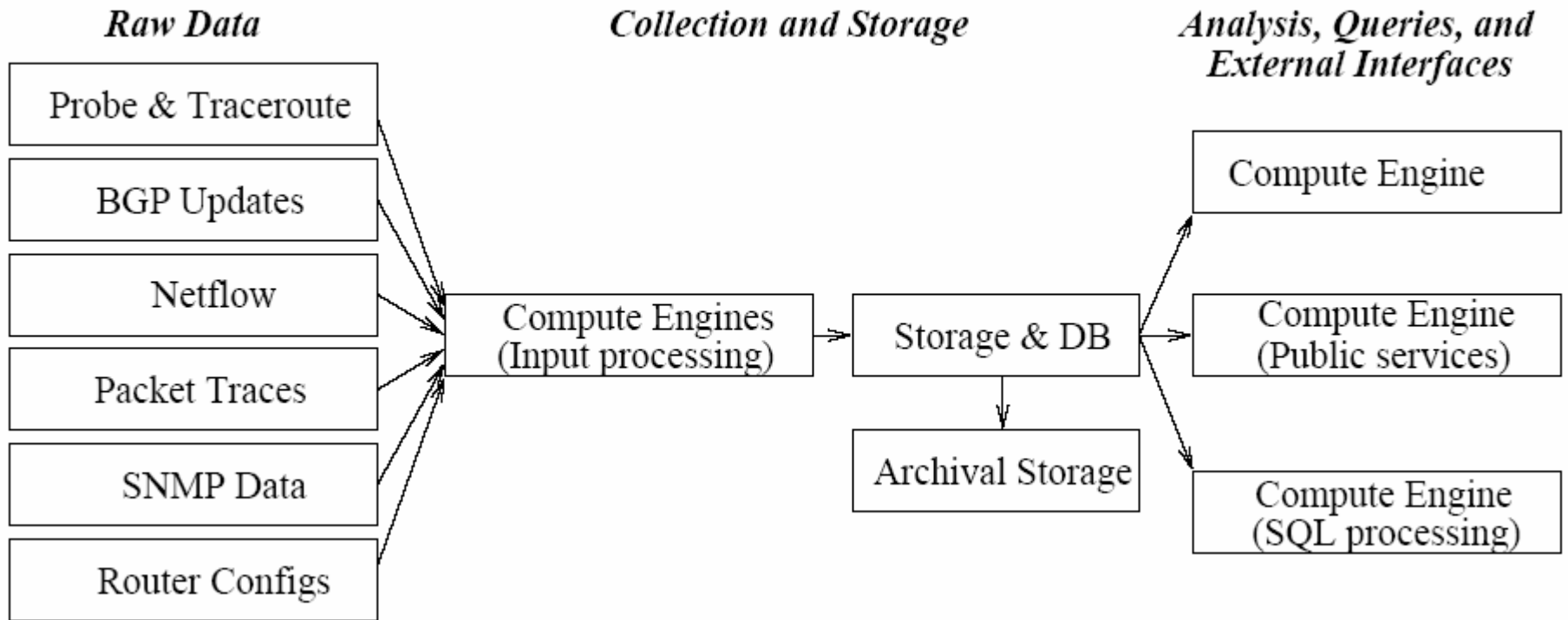
# The Datapositionary

Dave Andersen, CMU

Nick Feamster, Georgia Tech

<http://www.datapositionary.net/>

# Datapository Architecture



- **Separate:** collection, storage, analysis
- **Collection:** abstract type, format, and access method

# A Handy Web Interface

## New Query

Feed	<input type="text" value="fs_net"/>	
Client IP	<input type="text"/>	
Client Port	<input type="text"/>	
Subject	<input type="text" value="contains"/>	<input type="text"/>
To	<input type="text" value="contains"/>	<input type="text"/>
X-Mailer Header	<input type="text" value="contains"/>	<input type="text"/>
From	<input type="text" value="contains"/>	<input type="text"/>
Delivered To	<input type="text" value="contains"/>	<input type="text"/>
Time	<input type="text" value="Last 10 minutes"/>	Start: <input type="text"/> End: <input type="text"/>
Plot scale	<input type="text" value="linear"/>	Binsize (seconds) <input type="text" value="auto"/>
Limit	<input type="text" value="none"/>	
Show	Time <input checked="" type="checkbox"/> Client IP <input checked="" type="checkbox"/> Client Port <input checked="" type="checkbox"/> Subject <input checked="" type="checkbox"/> To <input checked="" type="checkbox"/> X-Mailer Header <input checked="" type="checkbox"/> From <input checked="" type="checkbox"/>	
Action	<input type="text" value="list"/>	<input type="button" value="Submit"/>

# Do Spammers Hijack BGP Routes?

Theory:

1. announce BGP route for mail server
2. Send lots of spam
3. Withdraw route, becoming invisible

Reality? Let's check...

# Selecting Spammers

```
SELECT * from spam
WHERE spam.time > ... AND spam.time < ...
  AND EXISTS
    (SELECT * from bgp
     WHERE bgp.time > spam.time - 20
       AND bgp.time < spam.time + 3600
       AND bgp.prefix = (spam.client_ip &
        ((~0) << (32 - bgp.mask)))
```

# Export Formats

- Web Interface
- XMLRPC
  - Text-based output
  - Programmatic interface
  - Output to Matlab
- (Brief demo)