

EDDY: End to Diagnostic Discovery

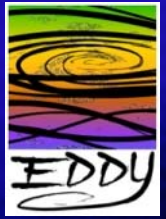
A Backplane for Diagnostic Data

Chas DiFatta

chas@cmu.edu

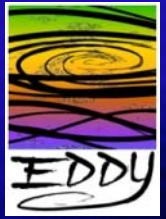
Mark Poepping

poepping@cmu.edu

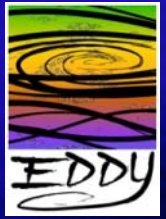


- Internet2 Middleware example
 - Developing multi-site infrastructures (Shib) involving multiple services over transit networks with varying policy
 - Uh, what just happened, who to call?
 - Need diagnostics from all over to handle this
 - Maybe broken or busy or lossy link
 - Maybe duplex mismatch
 - Maybe broken LDAP server
 - Maybe an incorrect router ACL

Problems for Diagnosticians

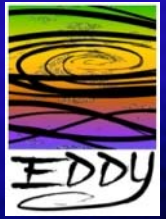


- Limited **creation** of diagnostic data
- Limited **access** to diagnostic data that exists
- **Discovering** value in a growing sea of data
- **Correlating** different diagnostic information
- Providing evidence to confirm or **repudiate** a diagnosis
- Finding **time** to create tools to transfer diagnostic capabilities to less skilled organizations and/or individuals (automate)



Diagnostic tools

- Rarely cross domains (network, application, security, and system)
- Are highly focused to the specific problem set
 - Application
 - Performance
 - Middleware
 - Security
- Have a high investment in setup time
- Are mostly focused for use by highly technical and skilled diagnosticians



Device Management

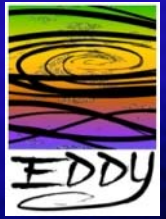
- Track Devices and Configurations
- Summary of Activity

Service Management

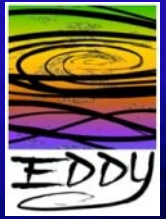
- Uptime and performance parameters
- Weather reporting (local and aggregate)

Activity Logging

- Flow Technologies, Application-specific Logs
- *Some* common logging
 - Syslog, netlogger, but var/log/{cron, maillog, messages, secure,...}
 - SIM products – Security-specific event correlation
 - IBM Common Event Infrastructure (CEI) technology (not 2 years ago)
<http://www-128.ibm.com/developerworks/library-combined/ac-cei/>



- Consolidate events into a simple framework to enable correlation
 - Between infrastructure layers
 - Among application technologies
 - Across administrative domains
- Support event dissemination, data lifecycle, data scaling
- Enable diagnostic tool development platform that leverages existing tools while enabling the next generation (multi-domain analytics)

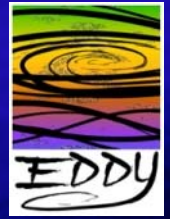


End-to-end Diagnostic DiscoverY

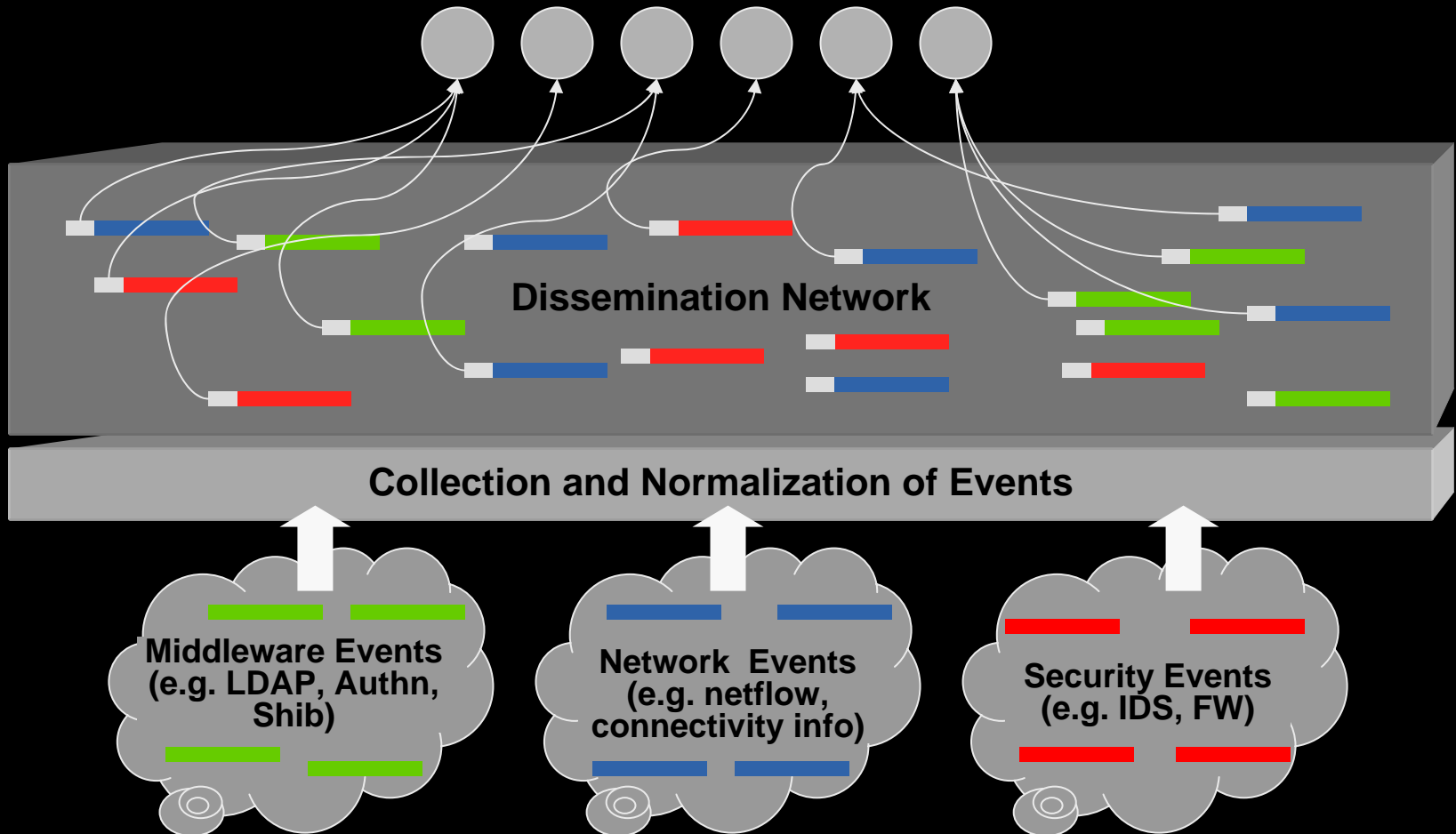
A Diagnostic Backplane to manage data

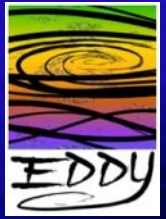
- Common Event Record – schema for the backplane
- Normalization – integrate diagnostic data
- Transport
 - Filtering, duplication, and forwarding
 - Encryption
- Transformation – focus on the important data
- Storage – save what you need
- Analysis – pluggable analytics (left to experts)
- Application – visualization, control
- Extensibility of Agents
- Extensibility of Data

EDDY from 50,000'

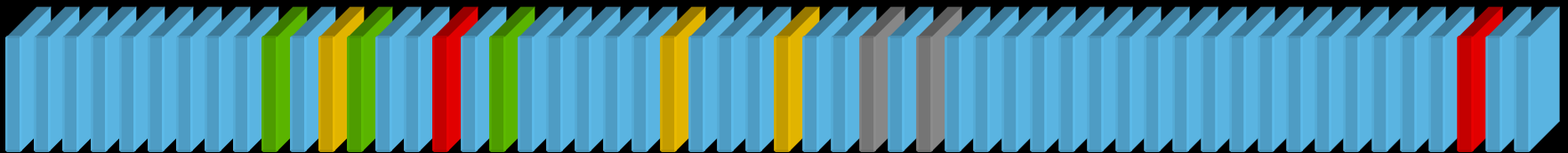


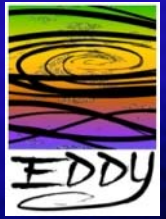
Diagnostic analysis applications



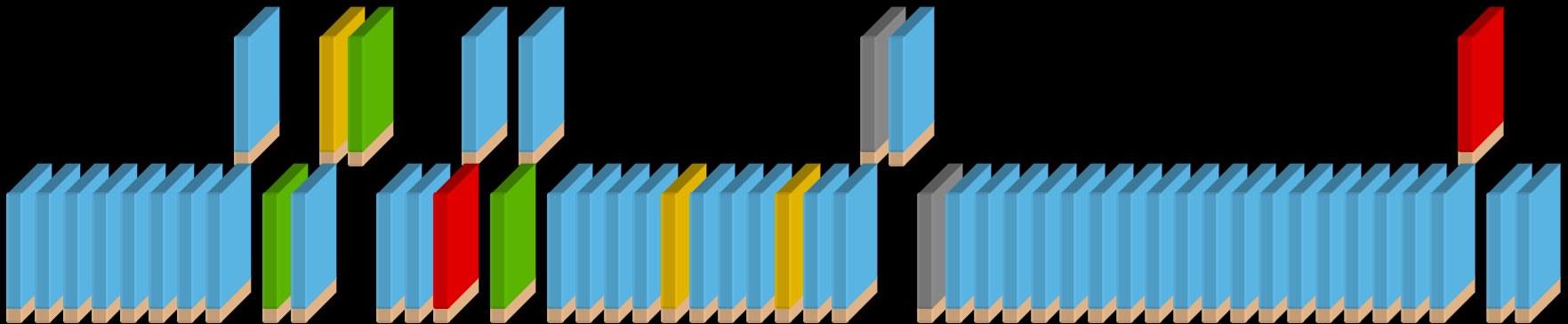


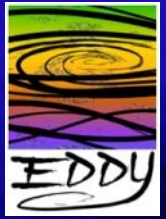
- Events generated all the time all over



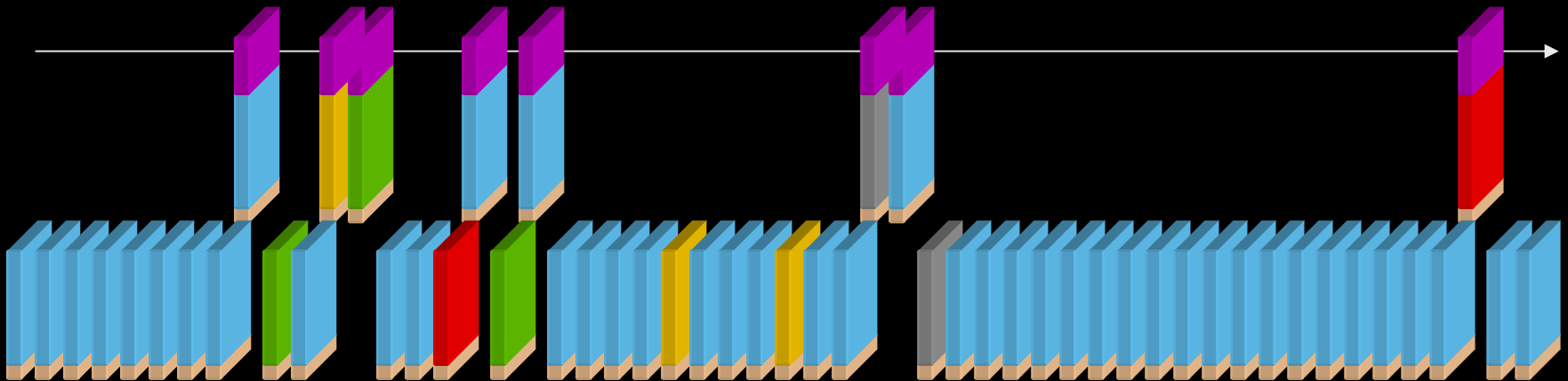


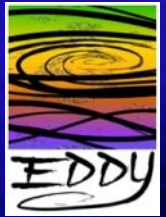
- Events generated all the time all over
- Add a common header to search for certain events





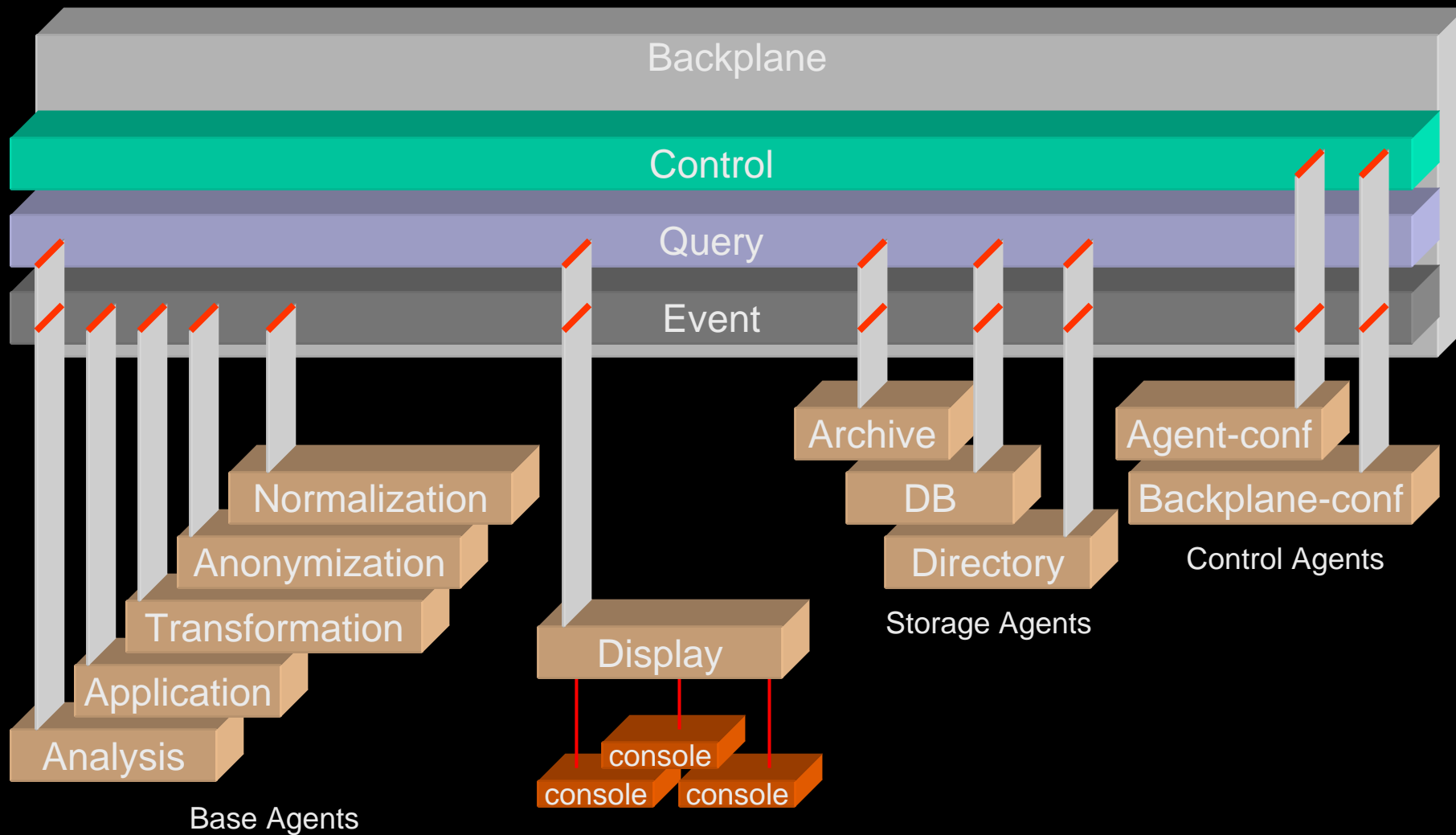
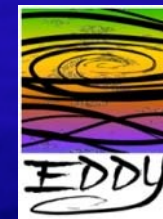
- Events generated all the time all over
- Add a common header to search for certain events
- Enable Analytics to correlate events

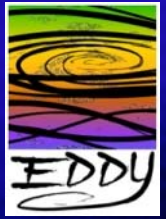




- Backplane
 - Event transport, Agent control, Event Query
- Common Event Record (CER) Structure
 - Raw, Cooked, Analyzed for 'event payload'
 - Event Class Model
 - Network, Application, System, Security, Environmental
- XML formatting for CER elements
 - Shortcut filtering
- Distributed agent processing – pipefitter style
- No change to existing logging infrastructure
- Platform for creation of new tools

Backplane Architecture





Dissemination

- Select events of interest to forward to appropriate analytics
- Control access as necessary

Lifecycle

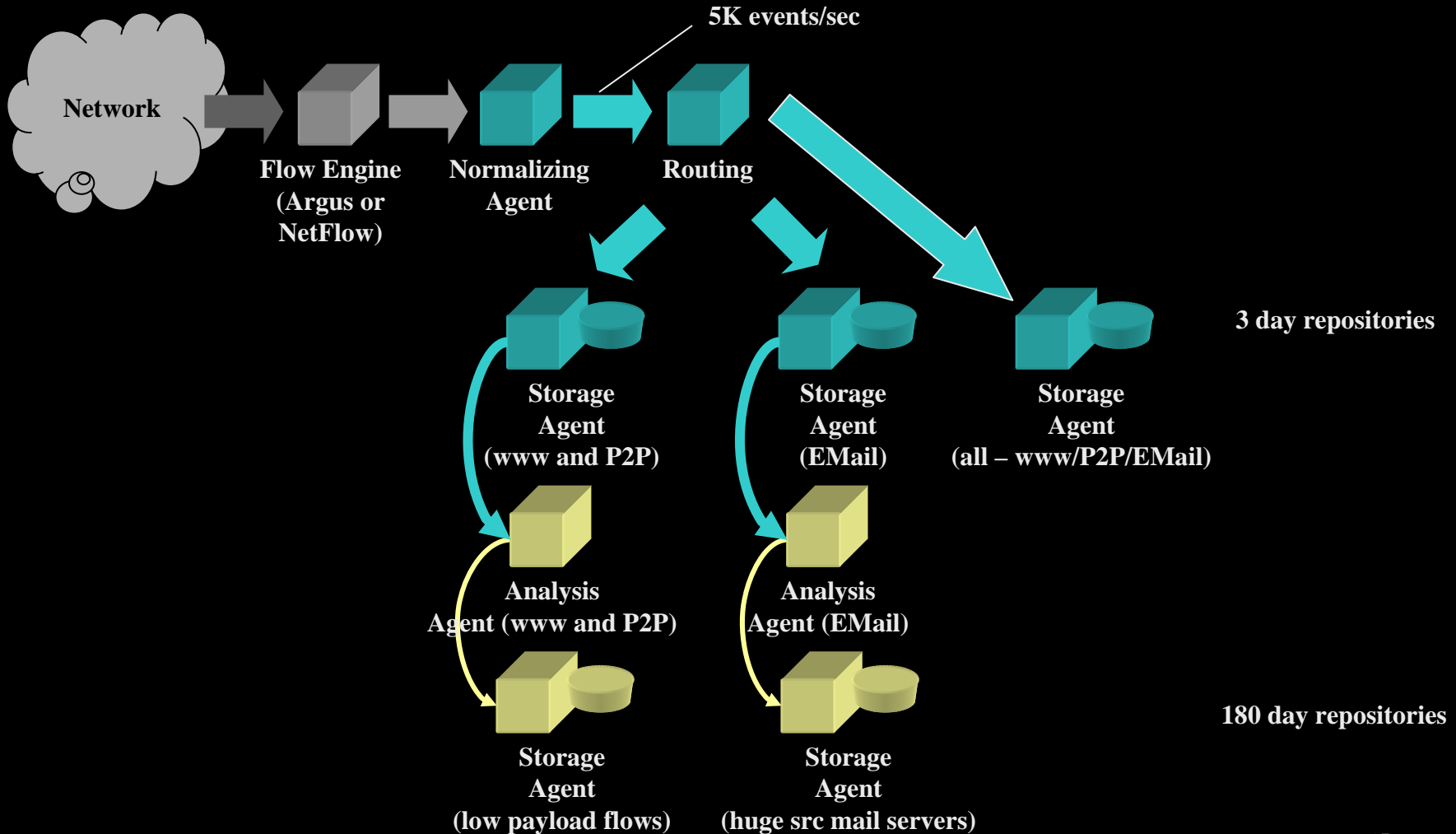
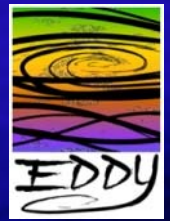
- Keep what you need: summarize, anonymize, eliminate

Scale

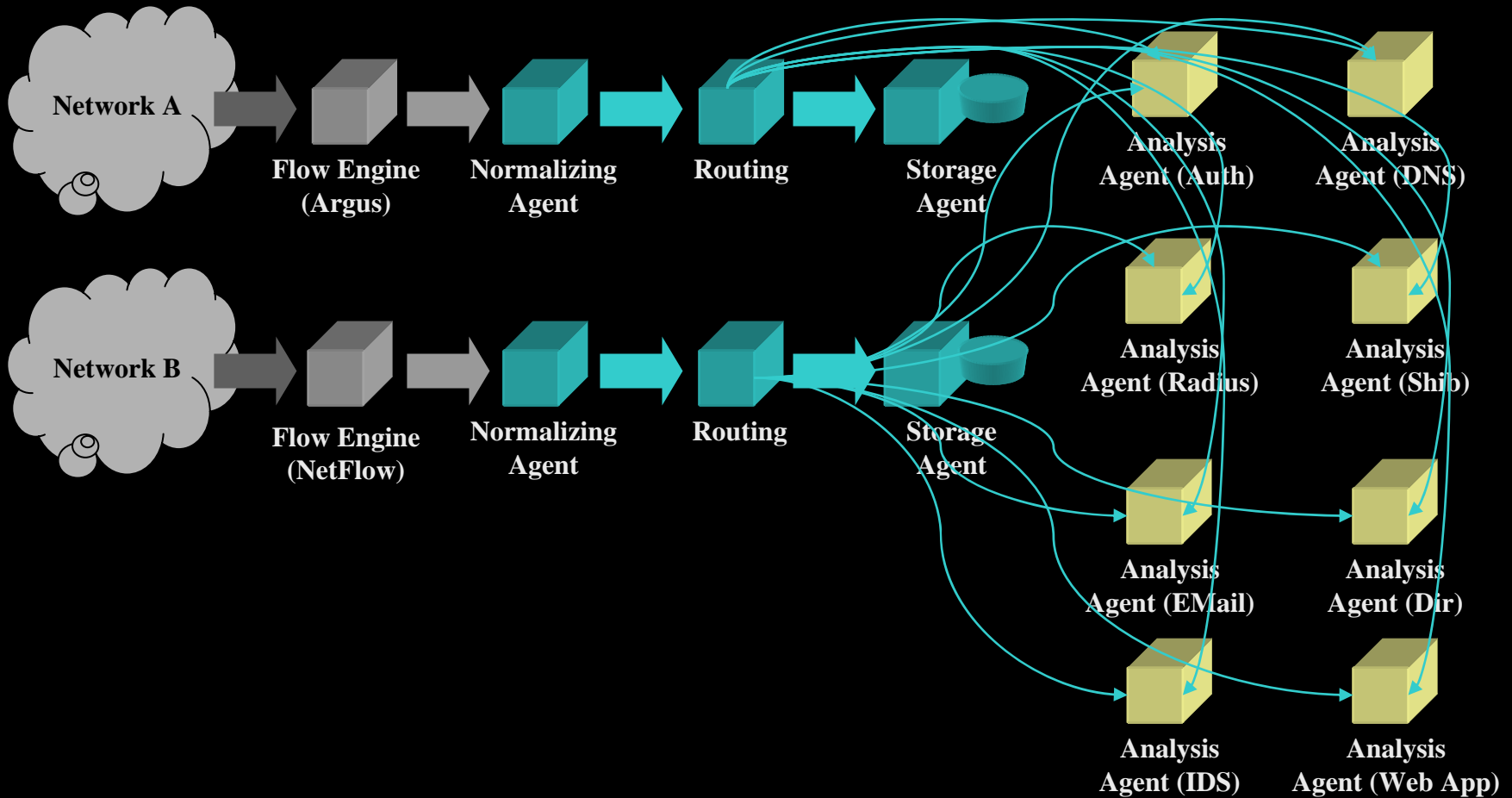
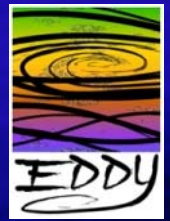
- Transform to expose/copy only what is needed
- Scale to match capabilities, capacities, requirements

All of these based on local policy

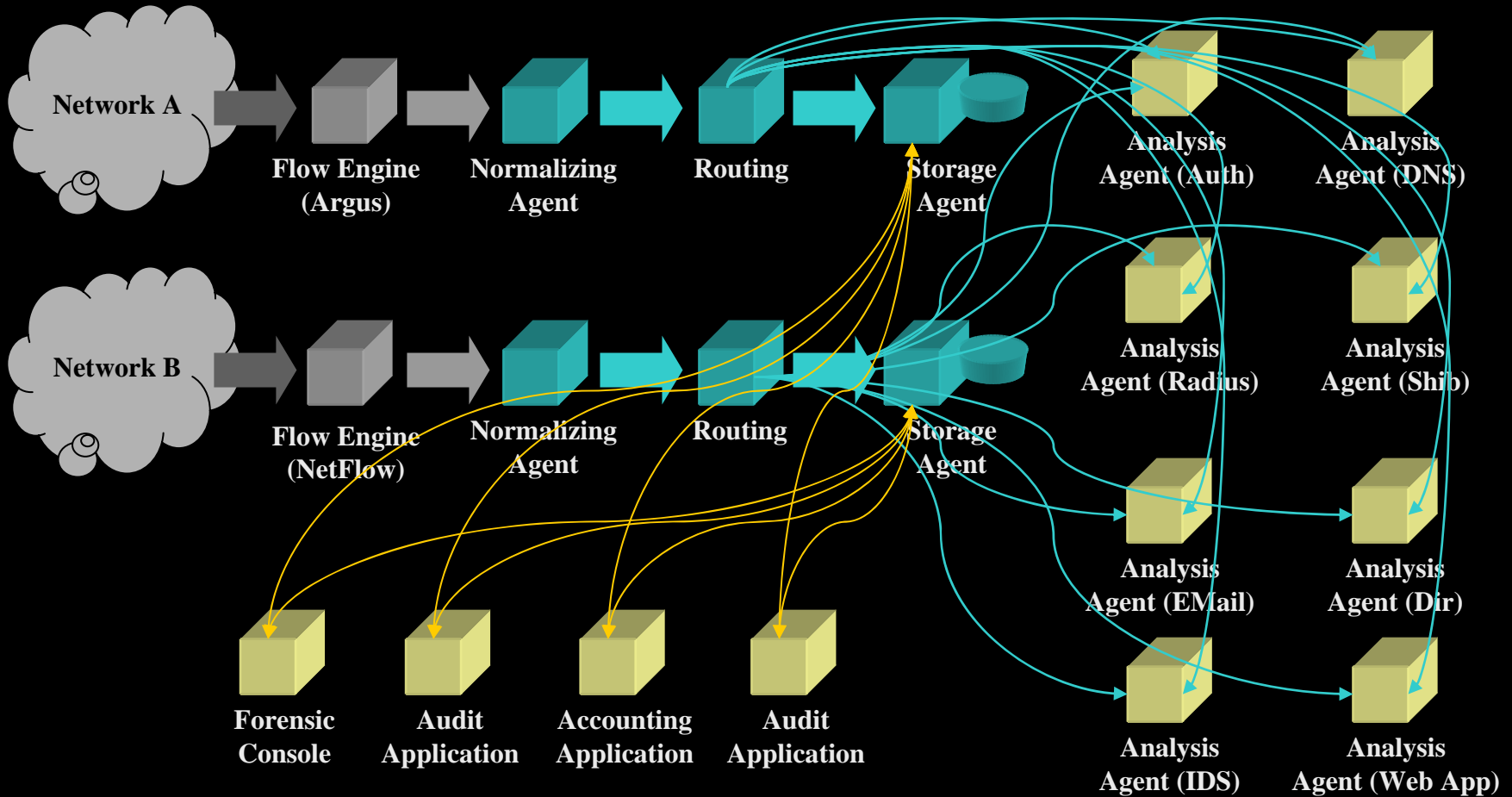
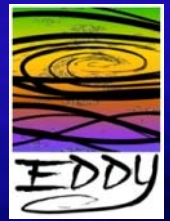
Data Lifecycle

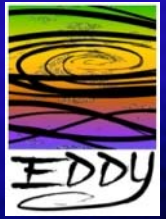


Managing Scale



Enabling Capability



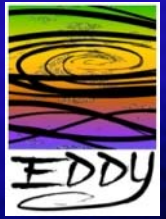


• Early Adopters

- CMU School of Computer Science: Dragnet
- CMU Architecture Department: Intelligent Workplace
- CMU Computing Services
 - Security group: IDS/flow correlation, forensics
 - SysAdmin: activity reporting, diagnostics
 - Network group: traffic accounting, diagnostics

• Collaborators

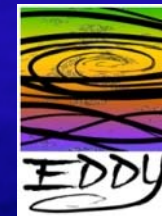
- Internet2: Shibboleth, Lionshare, Signet, E2Epi
- Individuals: Von Welch (Globus); Paul Hill (MIT); Brian Tierney (Netlogger); Kevin Miller, Michael Gettes (Duke)



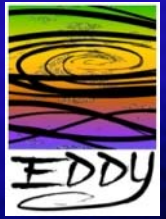
Already said this, but to remind for...

- Limited **creation** of diagnostic data
- Limited **access** to diagnostic data that exists
- **Discovering** value in a growing sea of data
- **Correlating** different diagnostic information
- Providing evidence to confirm or **repudiate** a diagnosis
- Finding **time** to create tools to transfer diagnostic capabilities to less skilled organizations and/or individuals (automate)

EDDY: Helps How?

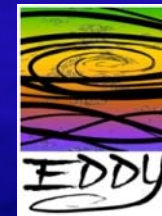


- **Creation**: incentive to improve logs
- **Access**: slicing/dicing enables access control
- **Discovering value**: raise signal-noise; short-cut answers to known questions
- **Correlation**: time is inherent, extensible for other clues
- **Repudiation**: base for affirmative analytics; suitability for audit
- **Tools**: development platform for access to data

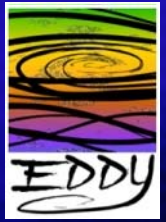


Broad Vision, one day at a time...

- Application/Service Developers
 - Feedback loop for diagnostic instrumentation
 - What's my stuff *really* doing "out there"
- **Diagnosticians**
 - **Flexible data access**
 - **Tools for automation**
- Administrators
 - Data management – scaling, access control, lifecycle
- Help Desks
 - Better information from the user – what their system 'saw'
 - Wide view – general health, trends
- End Users
 - Enable users to help themselves
 - Automate problem notification – view from the edge



- September 2005
 - Common Event Record (CER) specification
 - On the wire specification
 - Java libraries to implement the BackPlane API
 - Transport goal 5K events/sec (400M/day)
 - Normalizers for various log sources
 - Visible examples of early use



- Internet2 – Middleware Area

- Middleware Diagnostics <http://middleware.internet2.edu/e2ed>
- Flywheel support, use cases

- National Science Foundation

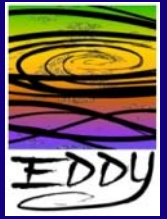
- Grant ANI-0330626
- Development support

- Sun Microsystems

- Agent processing hardware

- Carnegie Mellon

- Development support, Co-location, Administrative support
- Initial customers



EDDY: End to Diagnostic Discovery

A Backplane for Diagnostic Data

Chas DiFatta

chas@cmu.edu

Mark Poepping

poepping@cmu.edu