Thinking about Provenance

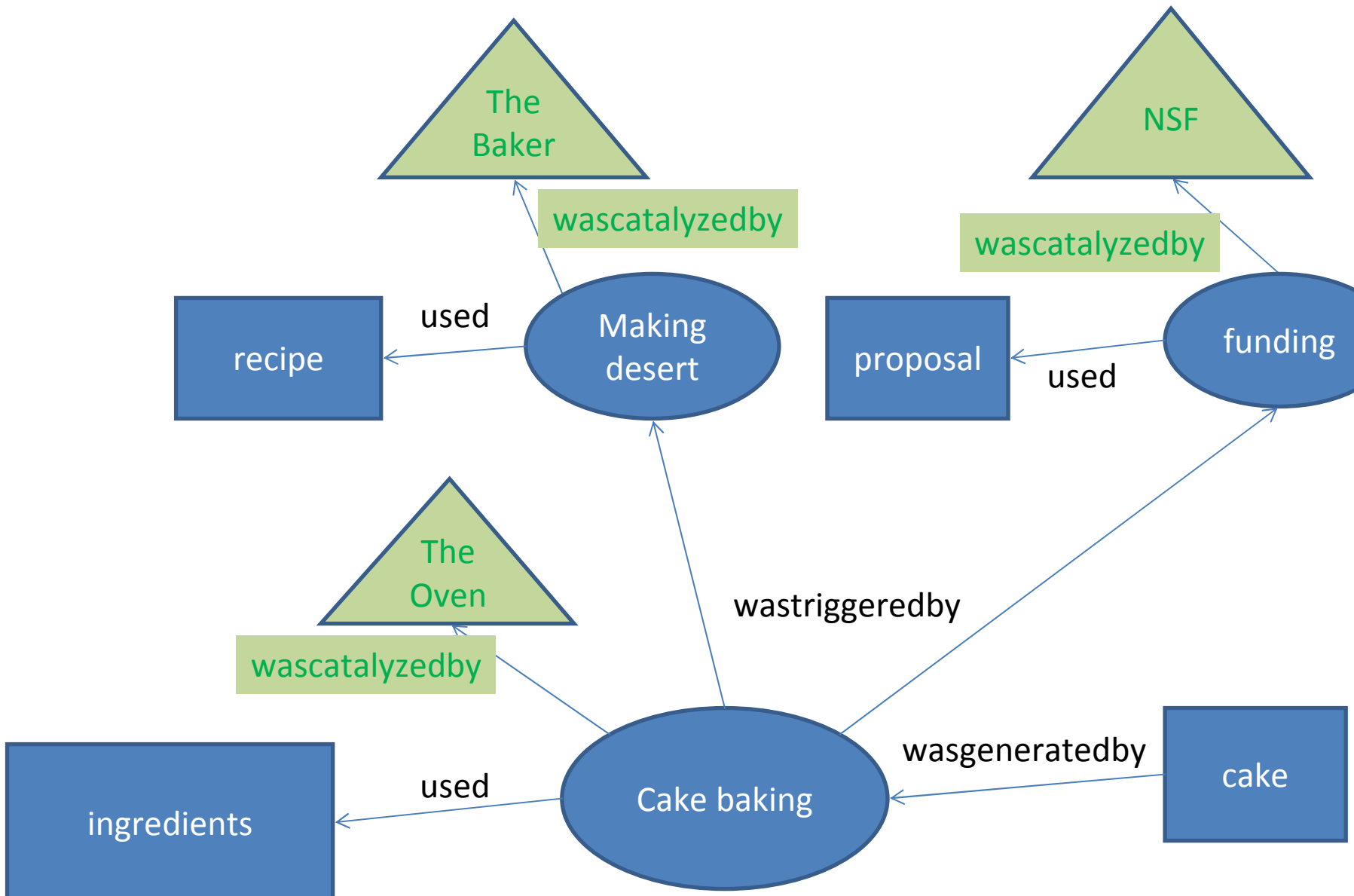Jim Myers, Luigi Marini, Joe Futrelle

PRAGMA 13

What It Is

- A description of what occurred, what caused what.

- A potentially fragmented, incomplete, and conflicting picture assembled from accounts from multiple witnesses

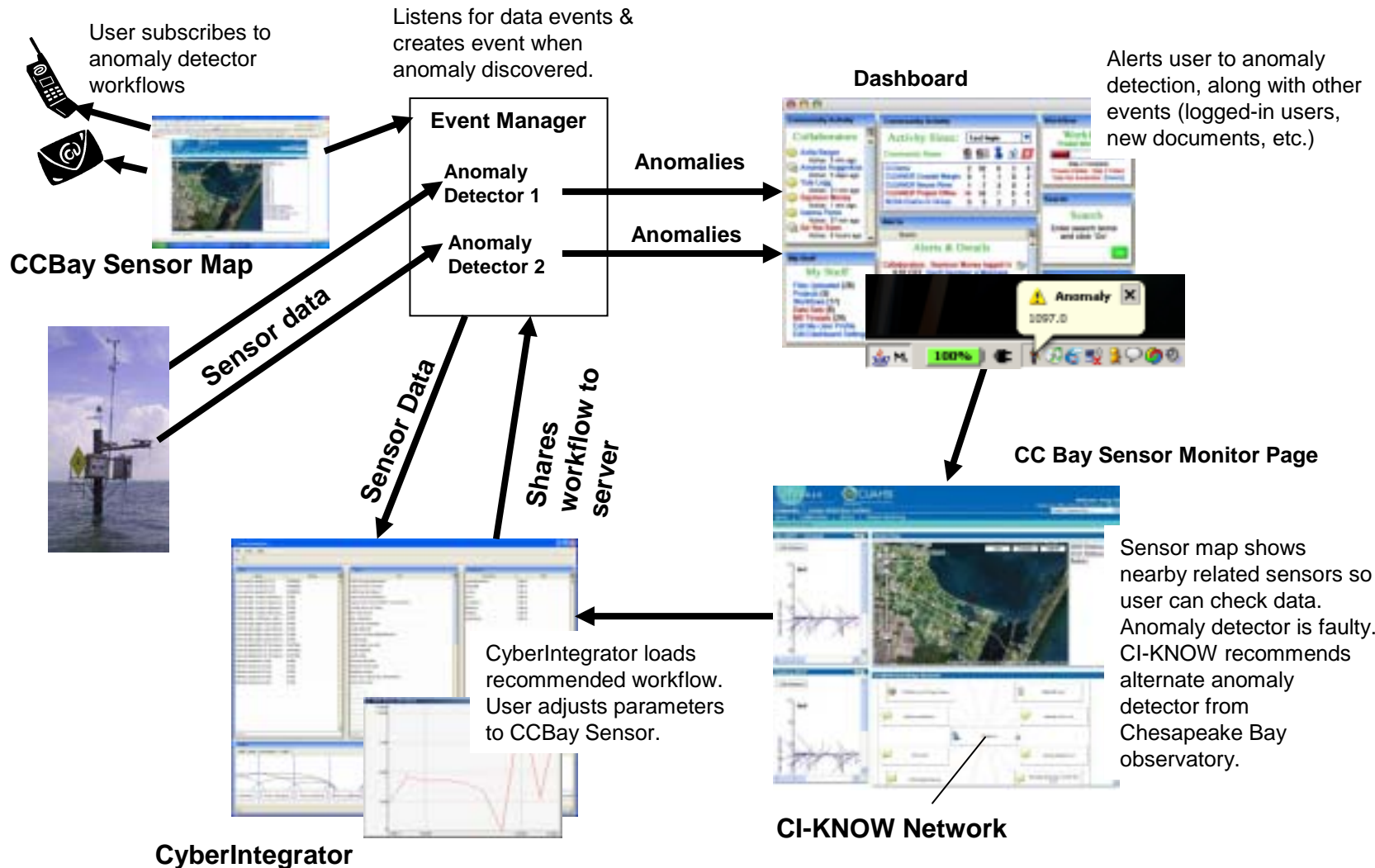- A part of metadata

What It Isn't

- A description of abstract processes
- A description at one level of detail
- A description of one kind of process (computational, intentional, scientific, engineering/mathematical, economic, managerial, etc.)

A Draft Provenance Model – being developed through the International Provenance and Annotation Workshop series

Provenance flows across – data streaming, portals, workflows, and the scientific literature: e.g. the ECID Sensor Anomaly Detection Scenario

User subscribes to anomaly detector workflows

Listens for data events & creates event when anomaly discovered.

Dashboard

Alerts user to anomaly detection, along with other events (logged-in users, new documents, etc.)

Event Manager

Anomaly Detector 1

Anomaly Detector 2

Anomalies

Anomalies

CCBay Sensor Map

Sensor data

Sensor Data

Shares workflow to server

CC Bay Sensor Monitor Page

CyberIntegrator loads recommended workflow. User adjusts parameters to CCBay Sensor.

Sensor map shows nearby related sensors so user can check data. Anomaly detector is faulty. CI-KNOW recommends alternate anomaly detector from Chesapeake Bay observatory.

CI-KNOW Network

CyberIntegrator

NCSA's Provenance Efforts

- Tupelo Semantic Middleware for Provenance and Metadata

- Cybercollaboratory – provenance aware portal

- Cyberintegrator – provenance-aware workflow

- MAEViz – provenance aware app& portal combination for earthquake engineering

- Data preservation – metadata/provenance extraction tools