Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

CHAMPION:

Intelligent Hierarchical Reasoning Agents for Enhanced Decision Support

November 16, 2011

RYAN HOHIMER, FRANK GREITZER, CHRISTINE NOONAN, JANA STRASBURG

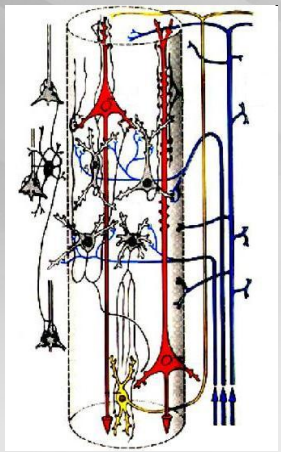
Pacific Northwest National Laboratory
Richland, WA

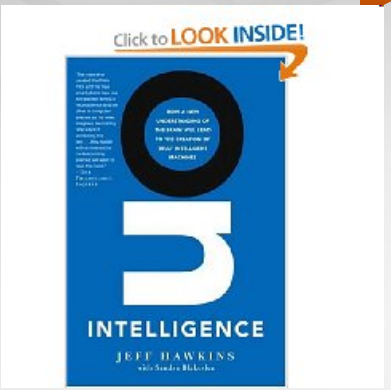
CHAMPION

Columnar Hierarchical Autoassociative Memory Processing In Ontology Networks

- October 2009 – Conceptual Notion
- 2010 – Lab Directed Research & Development (LDRD)
- March 2011 – Patent Application
- April 2011 – Insider Threat Reasoning
- October 2011 – Smart Grid Reasoning
- October 2011 – Nuclear Fuel Cycle Reasoning

Functional bio-mimicry of Neocortex

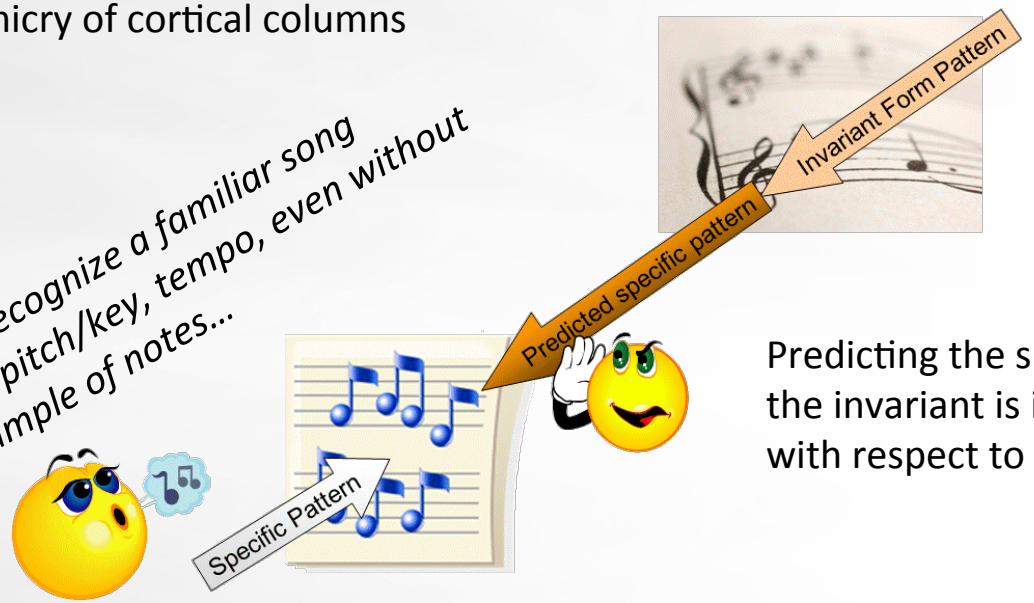




Functional bio-mimicry of cortical columns

- The neocortex excels at pattern recognition:
  - creates and stores *sequences of patterns*
  - stores them in an ***invariant*** form
  - recalls these patterns *auto-associatively*
  - stores them in a *hierarchy*
  - manages different types of memory

Example: We recognize a familiar song regardless of pitch/key, tempo, even without complete sample of notes…

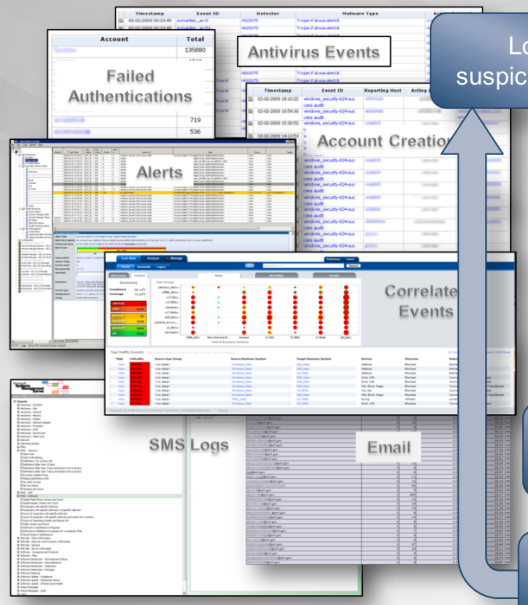


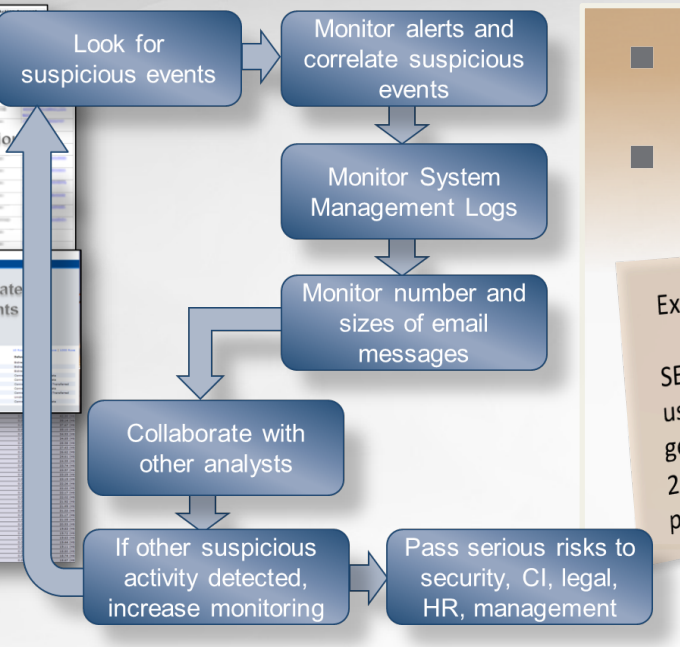
Predicting the specific from the invariant is important with respect to adaptation.

Analysis: Current Practice in Cyber Defense

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle *Since 1965*

- Monitoring of logs
- Anomaly detection
- Incident detection and response
- Forensics

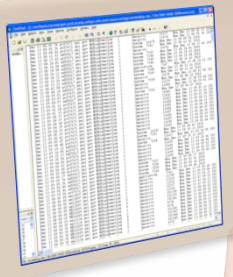




- ***Single-purpose (non integrated) detection systems***
- ***Reactive: Focus on detection after-the-fact***

Example:

SEIM system used at PNNL generates over 2 Billion events per week



Current practice imposes a substantial cognitive burden on analyst to maintain situation awareness by manually correlating and recognizing patterns across multiple classes of data.

Key to improving the analytic process is enhancing the ability to integrate (correlate) information across multiple data domains.

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle *Since 1965*

Conceptual Design

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle *Since 1965*

MODEL–BASED CLASSIFICATION

Data
Directly available information

→

Observation
Inference from data that reflects a specific state

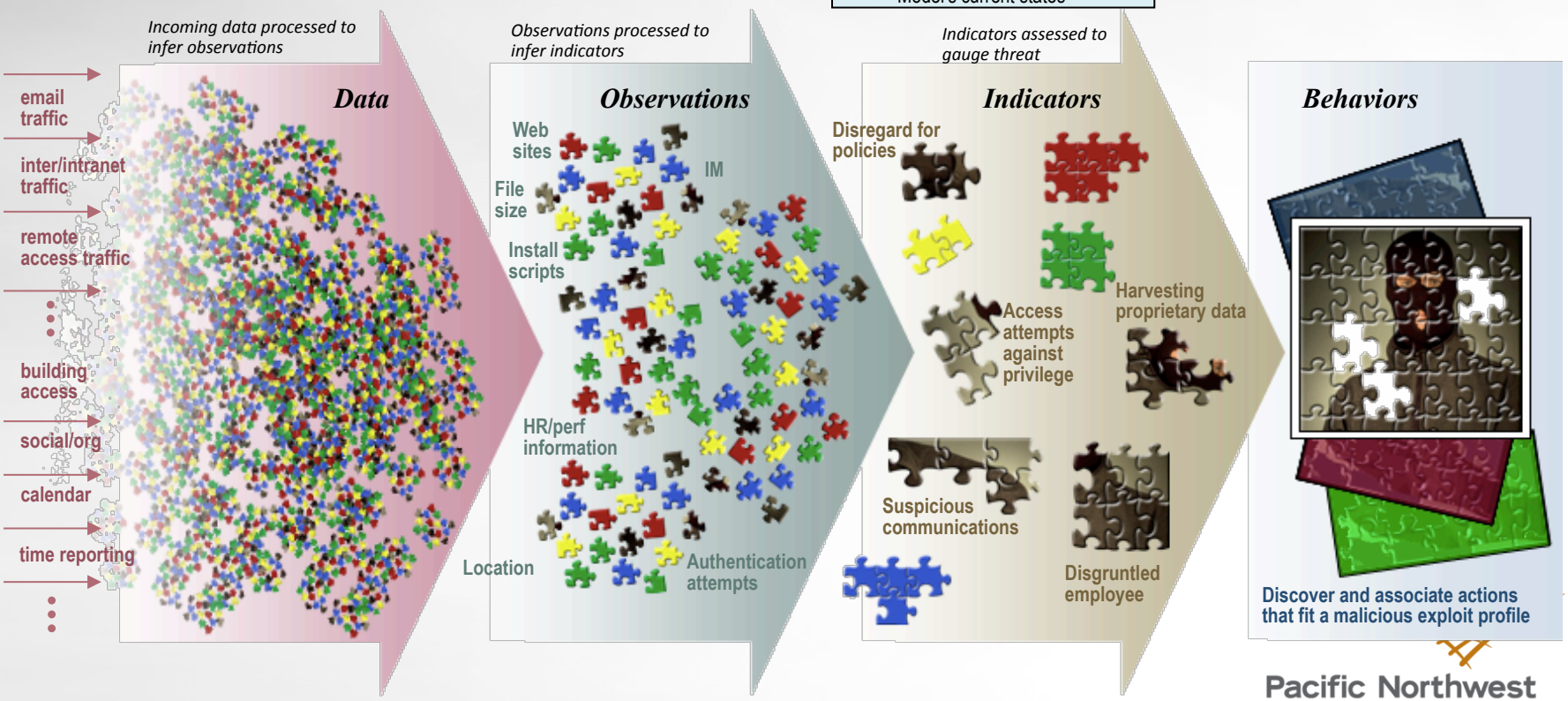
→

Indicator
Action/event as evidence of precursor to inferred behavior

→

Behavior
Sequence of actions associated with a purpose

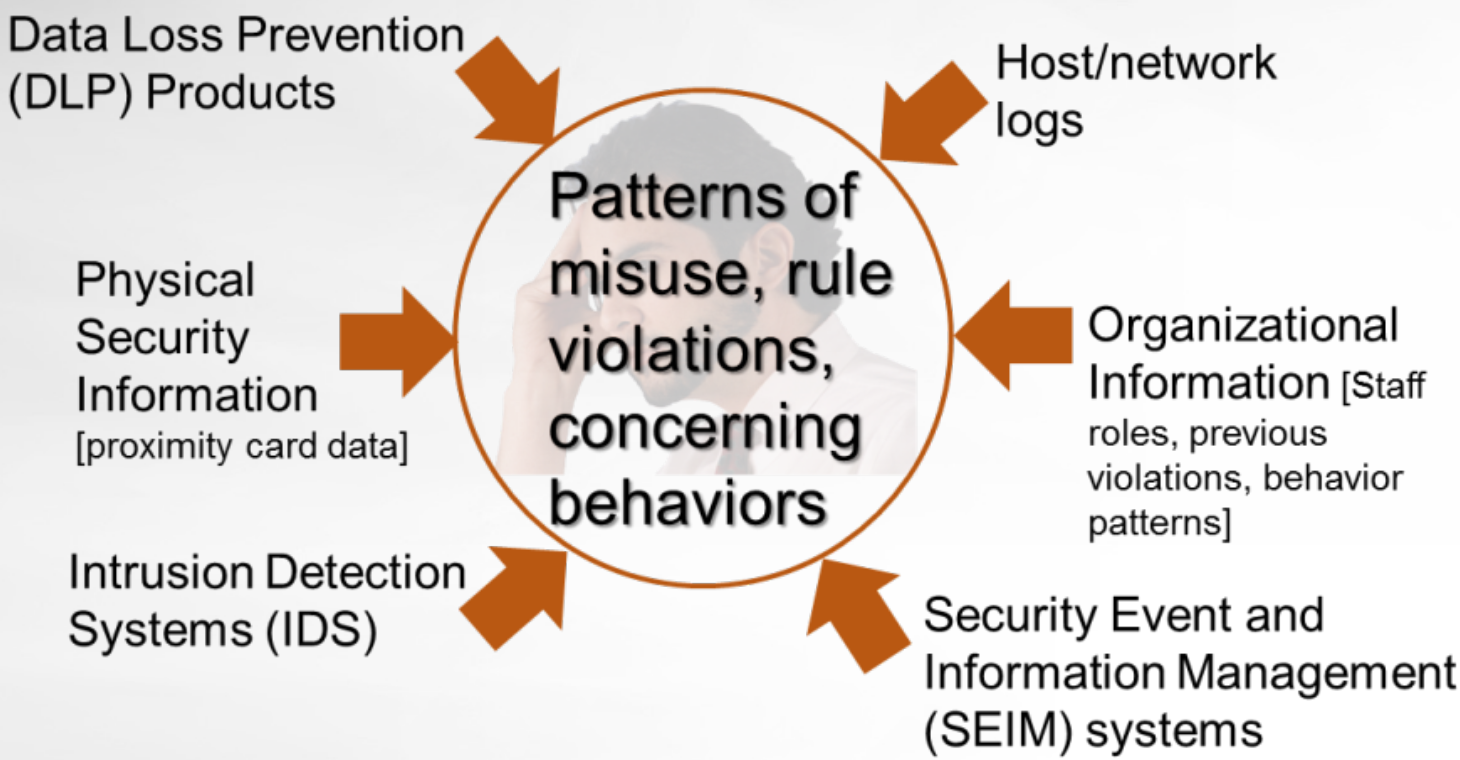
Beliefs
Model's current states



Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle *Since 1965*

Comprehensive Integrated Approach

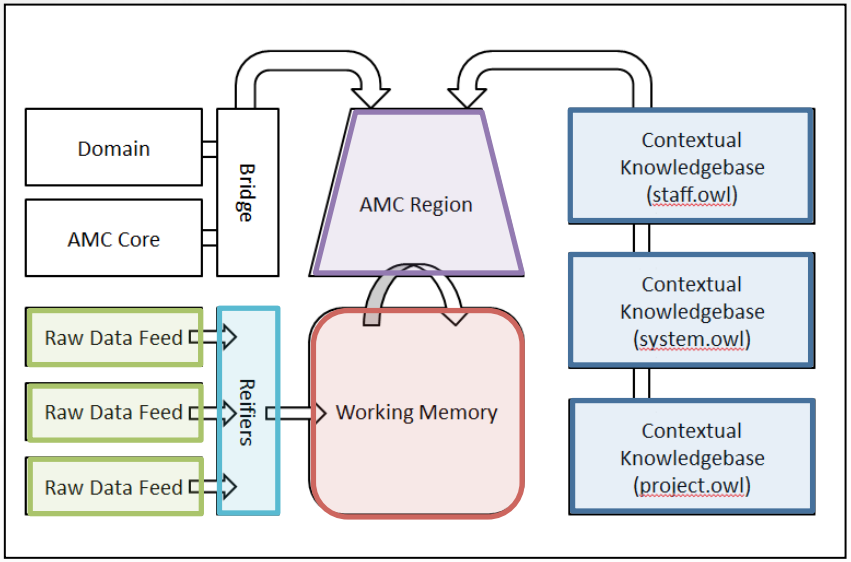
- A comprehensive integrated approach is needed to successfully combat insider threat
- Triage analysis manages "triggers" that indicate ongoing or potential malicious activity



Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

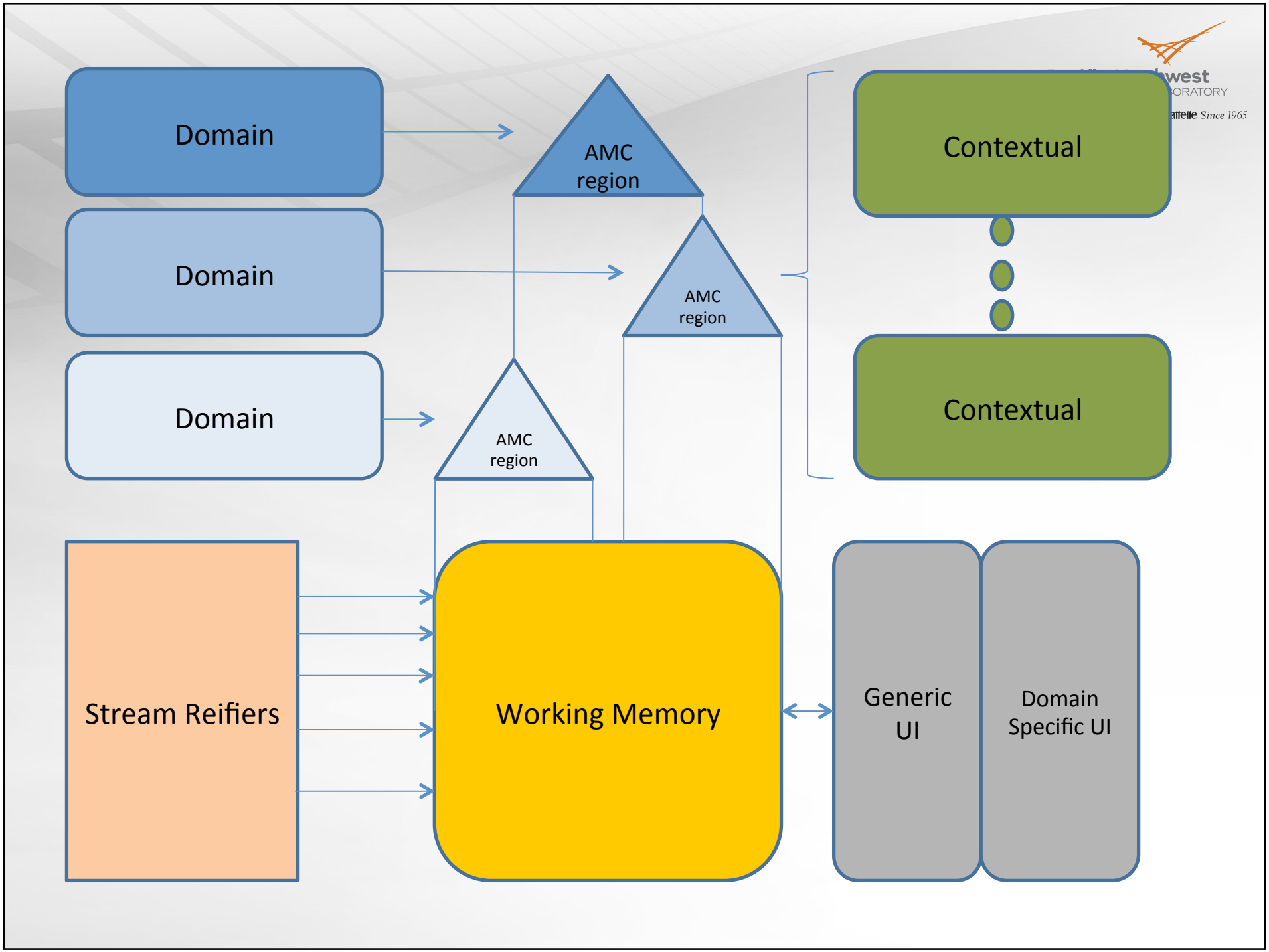
Components of the CHAMPION system

- **Ontologies** represent specialized domain knowledge
- **Reifiers** ingest the *primitive data* as individuals of the types specified in the domain ontologies
- **Auto-associative Memory Columns (AMCs)** interpret the data assertions and infer new assertions
- **Memory** stores the facts asserted from the *primitive data* and the facts inferred by the AMC reasoners

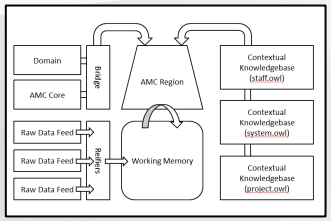


Pacific Northwest NATIONAL LABORATORY

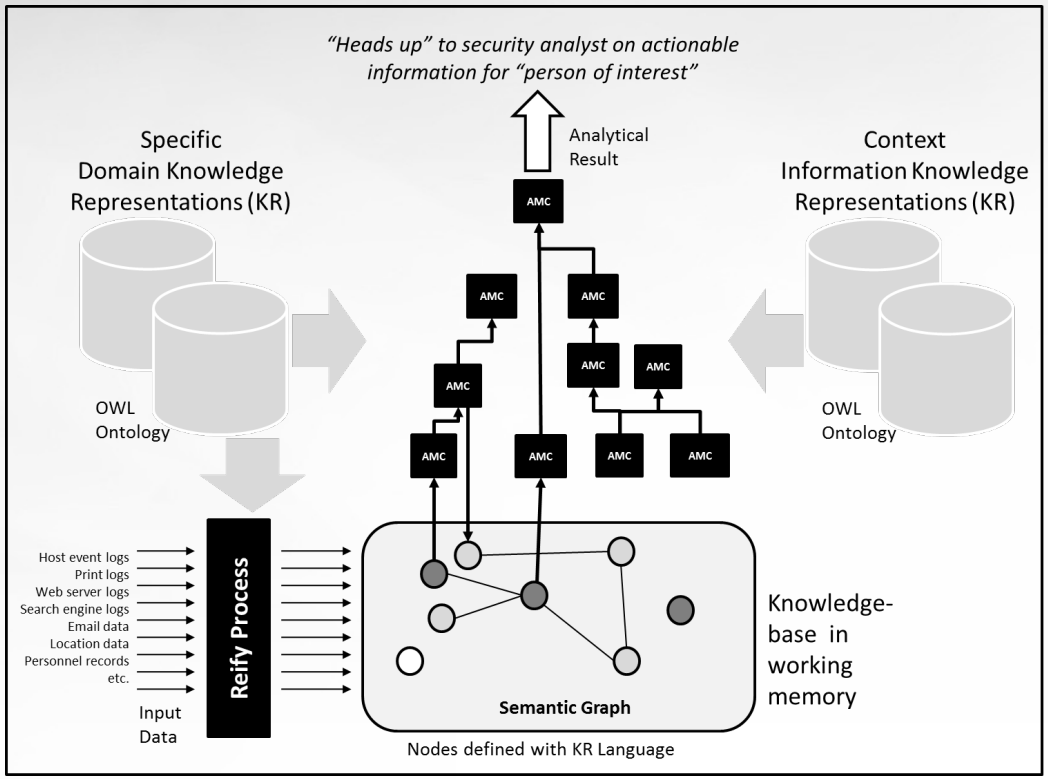
Proudly Operated by Battelle *Since 1965*


Domain
Domain
Domain
AMC region
AMC region
AMC region
Contextual
Contextual
Stream Reifiers
Working Memory
Generic UI
Domain Specific UI

Hierarchical Organization of AMCs



- The hierarchical organization of the AMCs enables integration of the analysis and reasoning process across different domains of data
- Auto-associative memory paradigm
  - AMC dynamically creates semantic graph (in working memory) by abstracting event sequences based on knowledge representations of malicious behavior (expertise captured in domain knowledge ontologies)
  - AMC uses invariant representations of event sequences to predict the next events in an attack sequence.

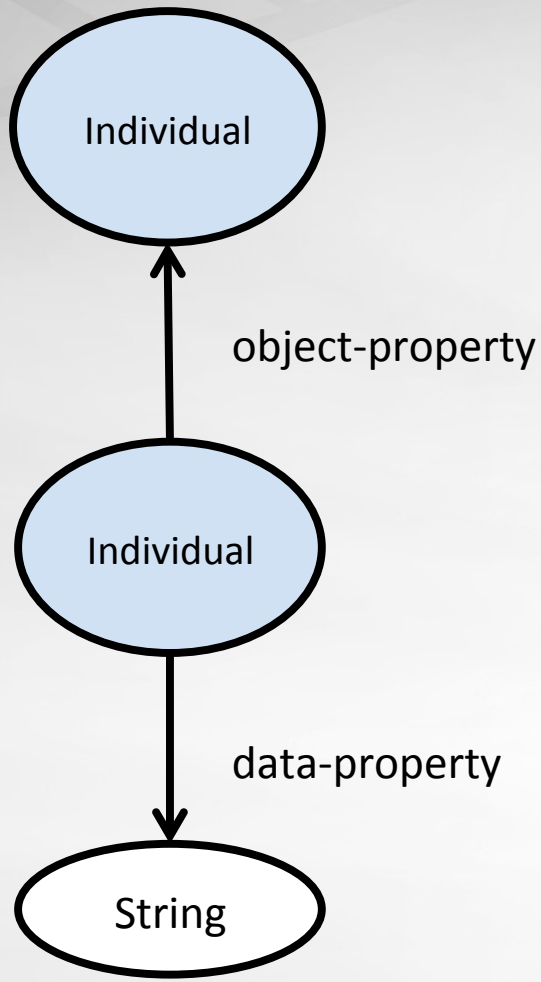


This computational higher level reasoning approach can pull together fragments of existing or available data to provide alerts to ongoing or upcoming potential malicious activity.

Pacific Northwest NATIONAL LABORATORY

Proudly Operated by Battelle *Since 1965*

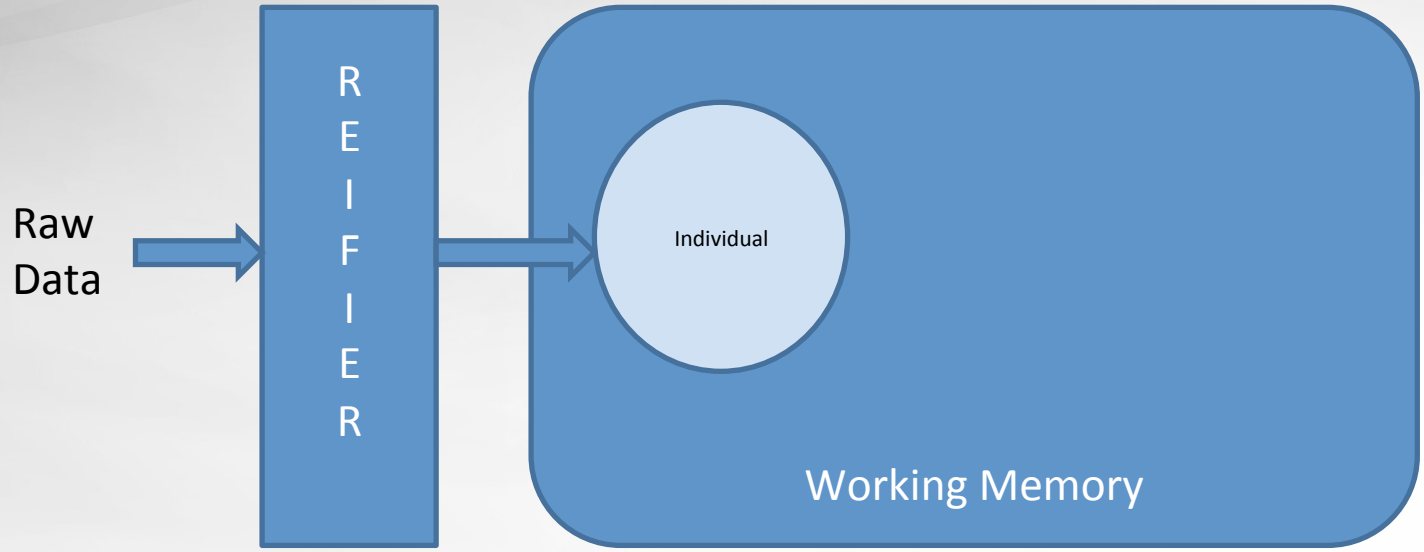
Semantic Graph



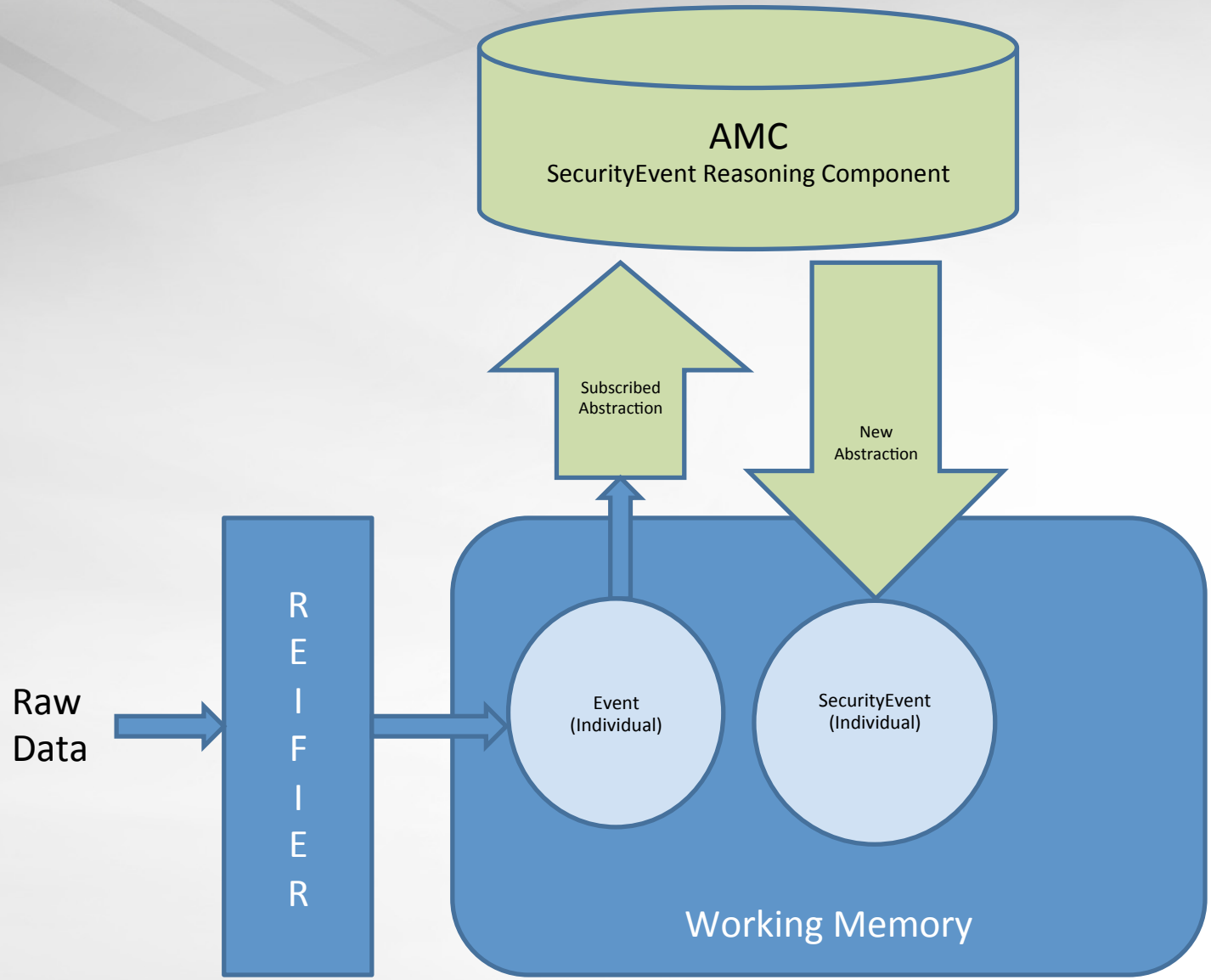
1. Semantic Graphs have nodes and edges.
2. There are two kinds of edges
   a) object-property: connects Individuals to other Individuals

   b) data-property: connects Individuals to data such as integers, strings, booleans, etc

Putting Facts into the Semantic Graph

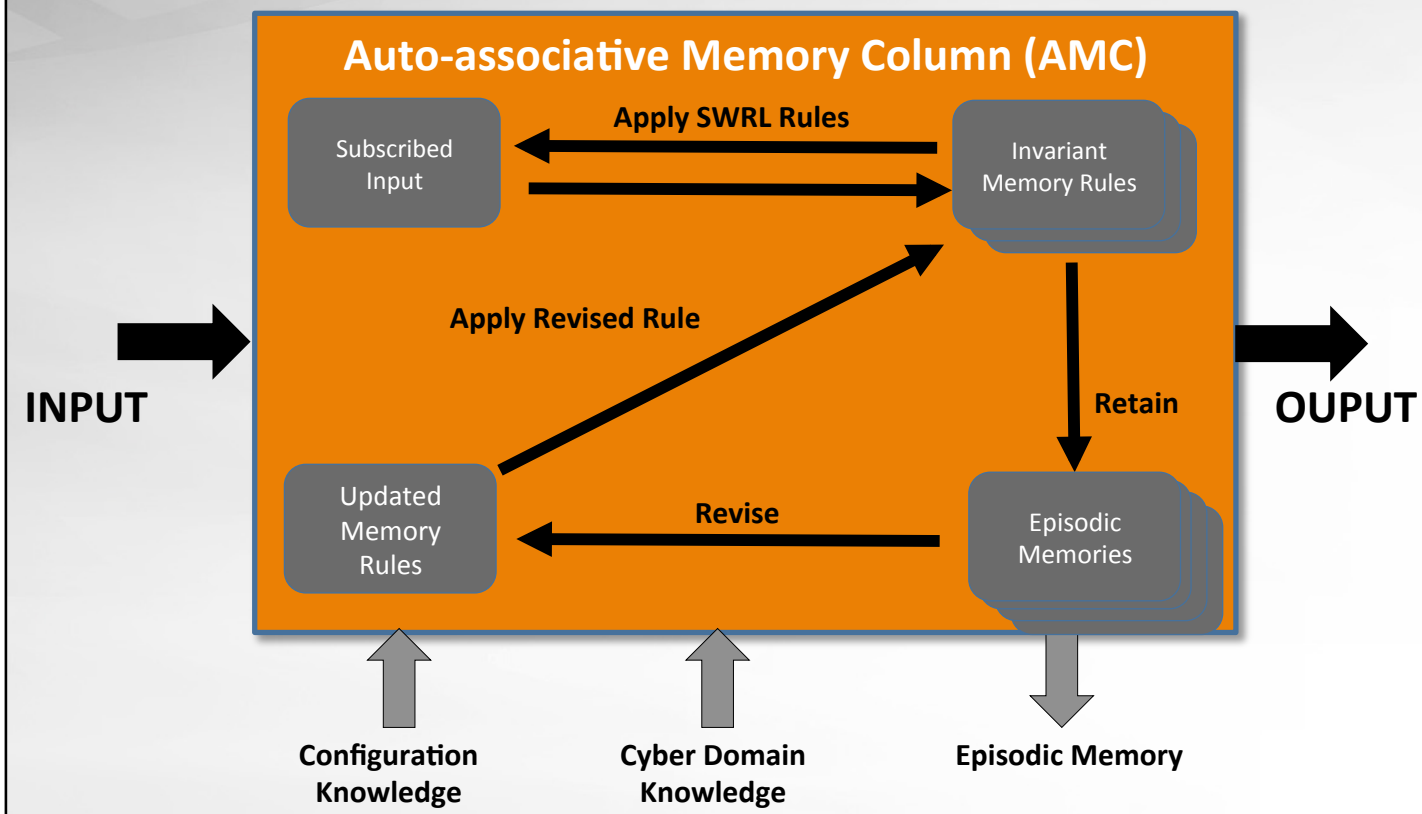


Reasoning about those Facts



Auto-associative Memory

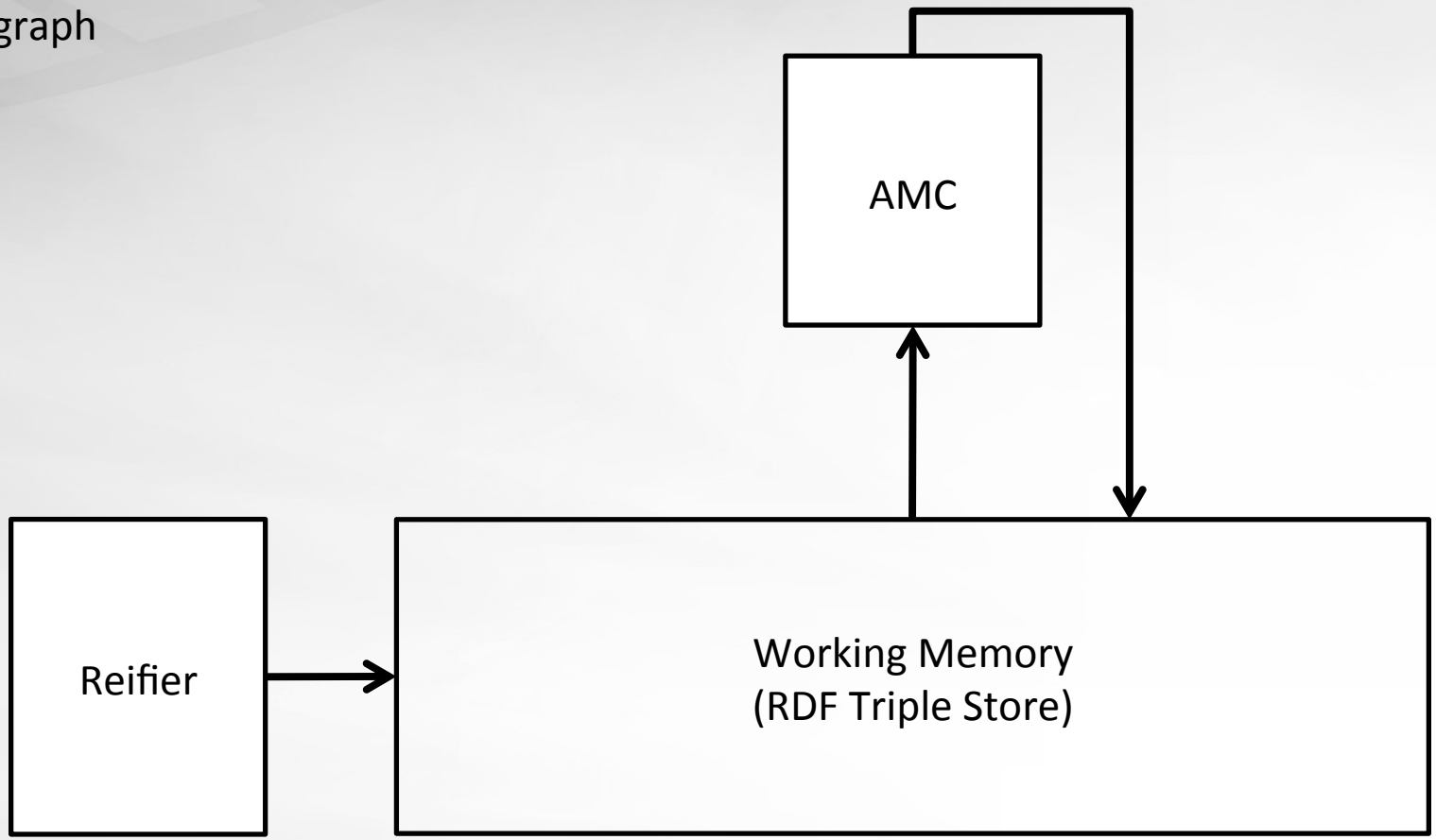
Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965



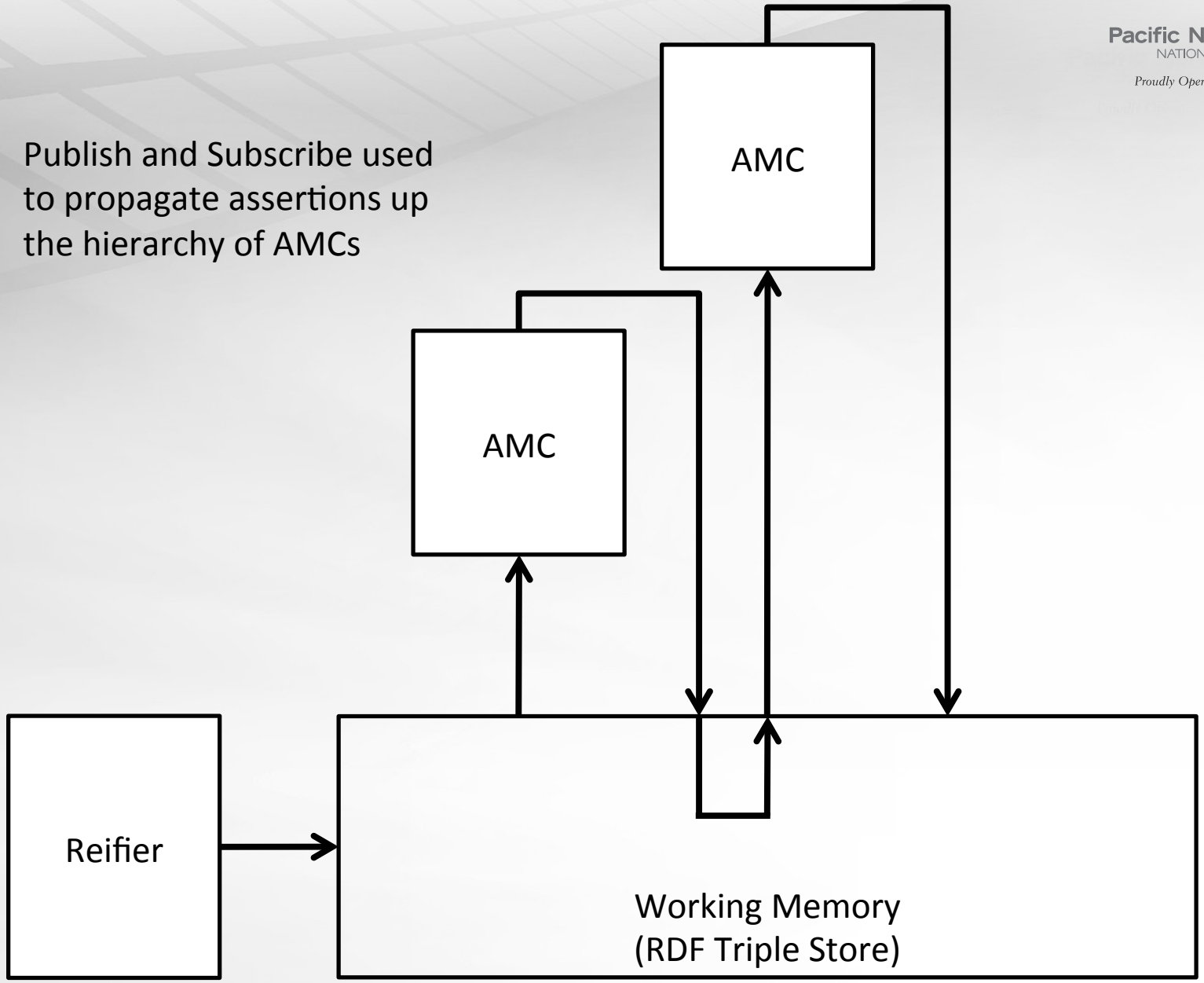
- We use rules* to determine if the input is *recognized.*
- If recognized, retain in memory

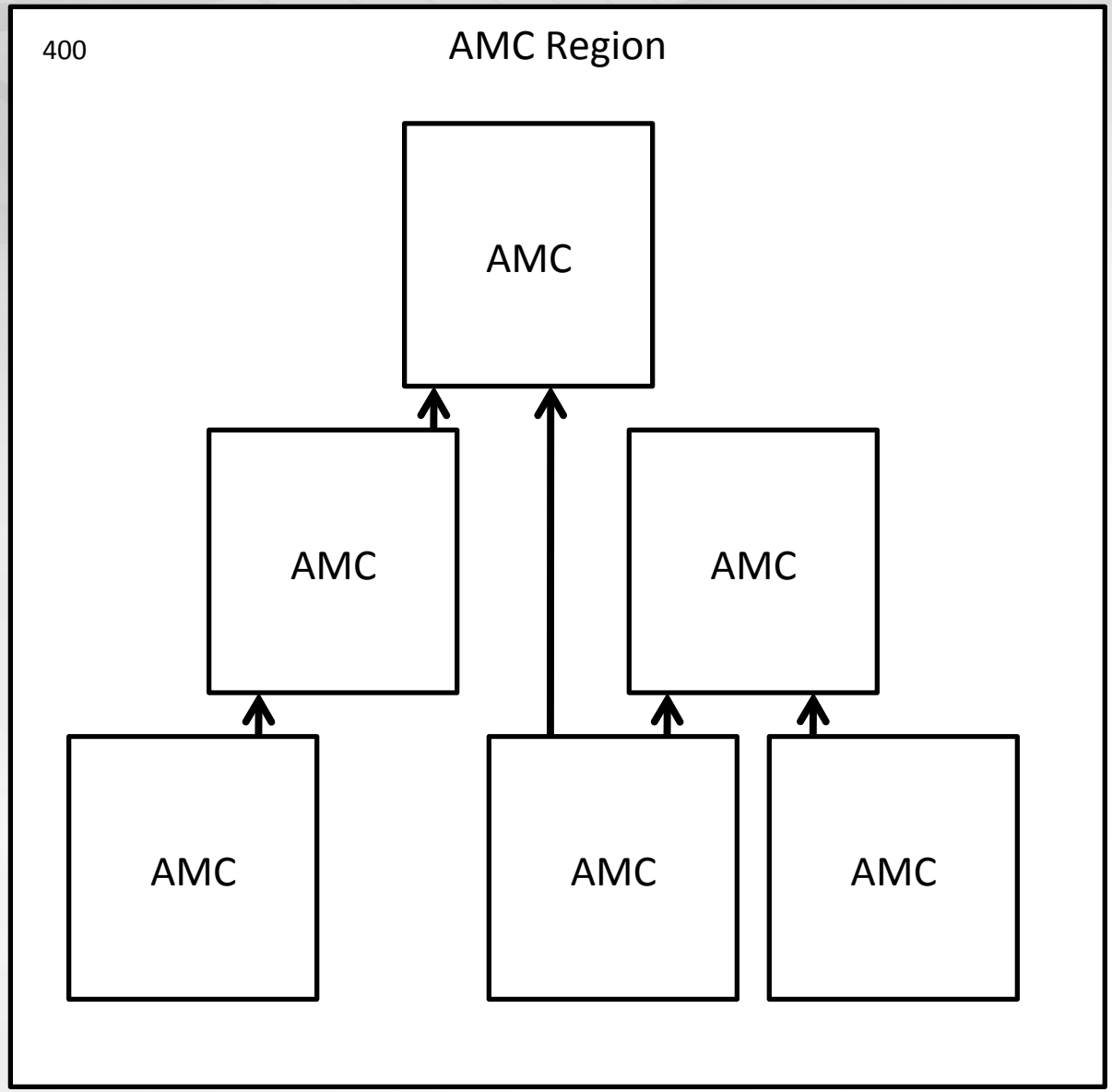
*Semantic Web Rule Language (SWRL)

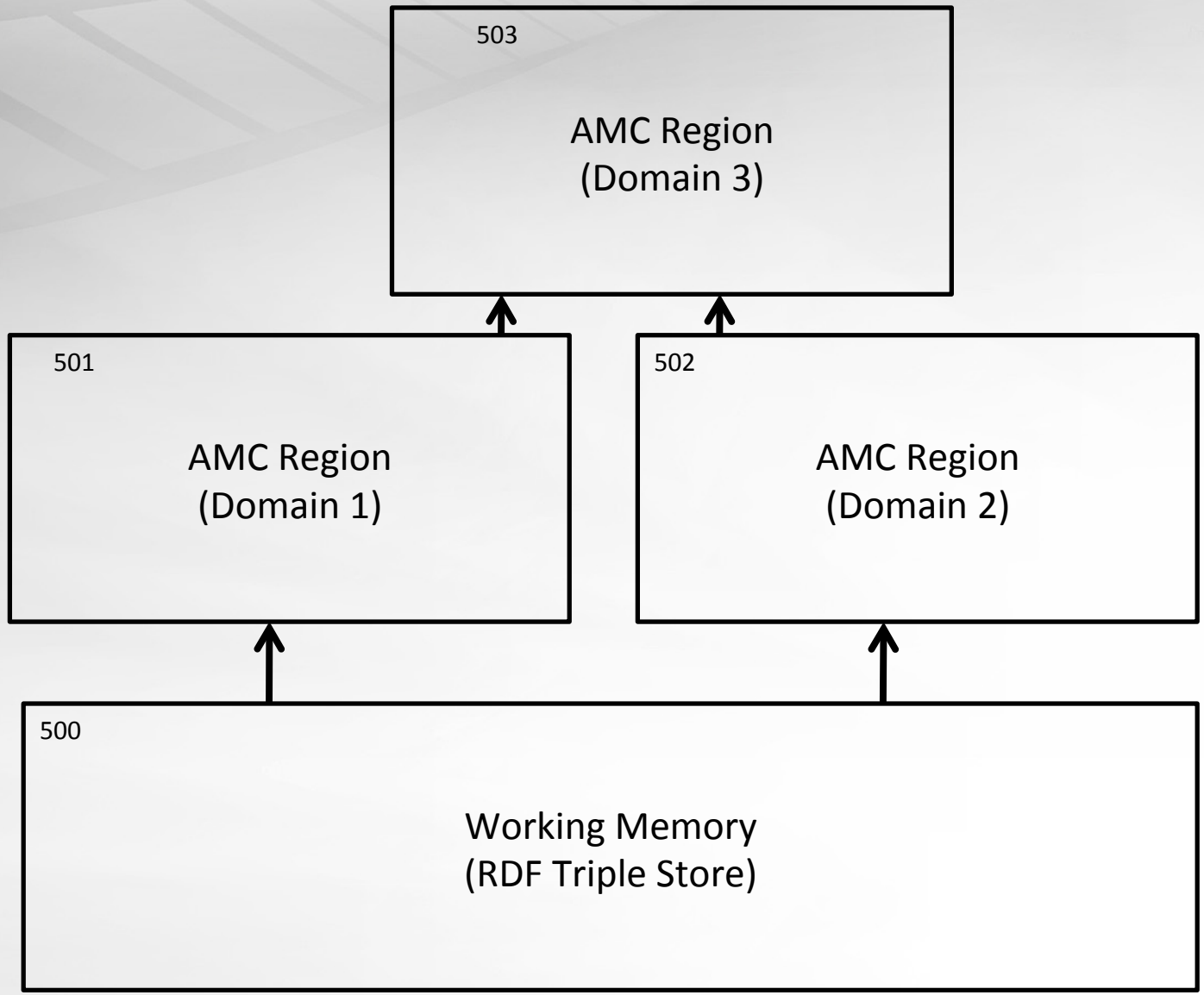
AMCs have a Publish and Subscribe relationship with the semantic graph



Publish and Subscribe used to propagate assertions up the hierarchy of AMCs

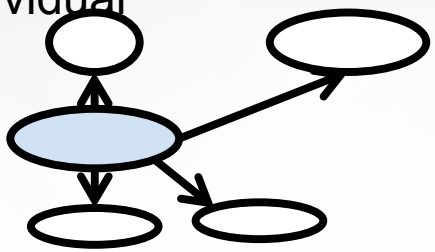




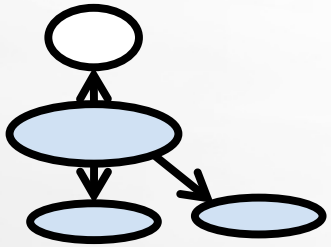

503
AMC Region
(Domain 3)
501
AMC Region
(Domain 1)
502
AMC Region
(Domain 2)
500
Working Memory
(RDF Triple Store)

Two Types of AMCs

- Subsumptive
  - Further classifies an abstraction based on state of individual
  - Asserts additional data properties to the individual

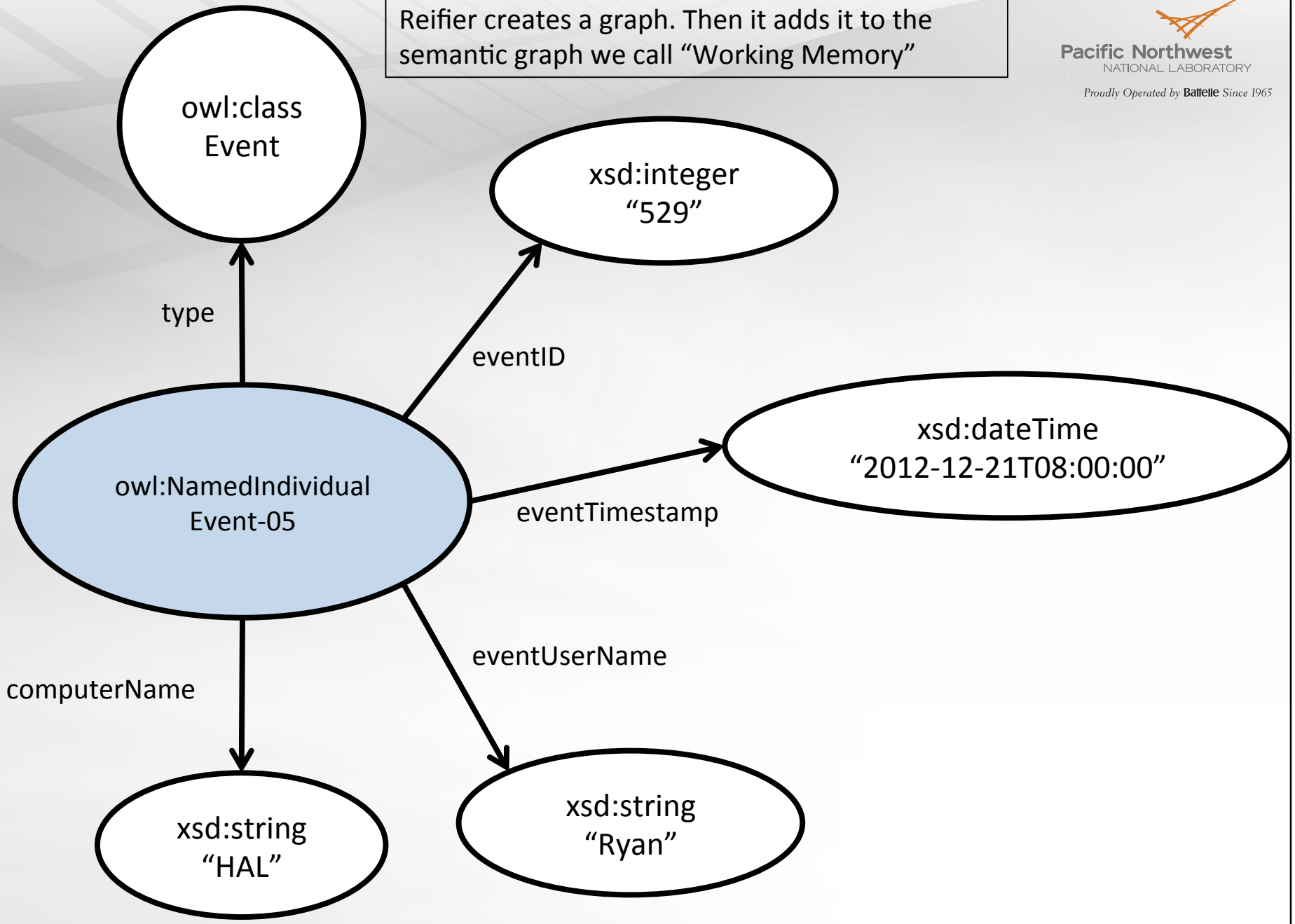


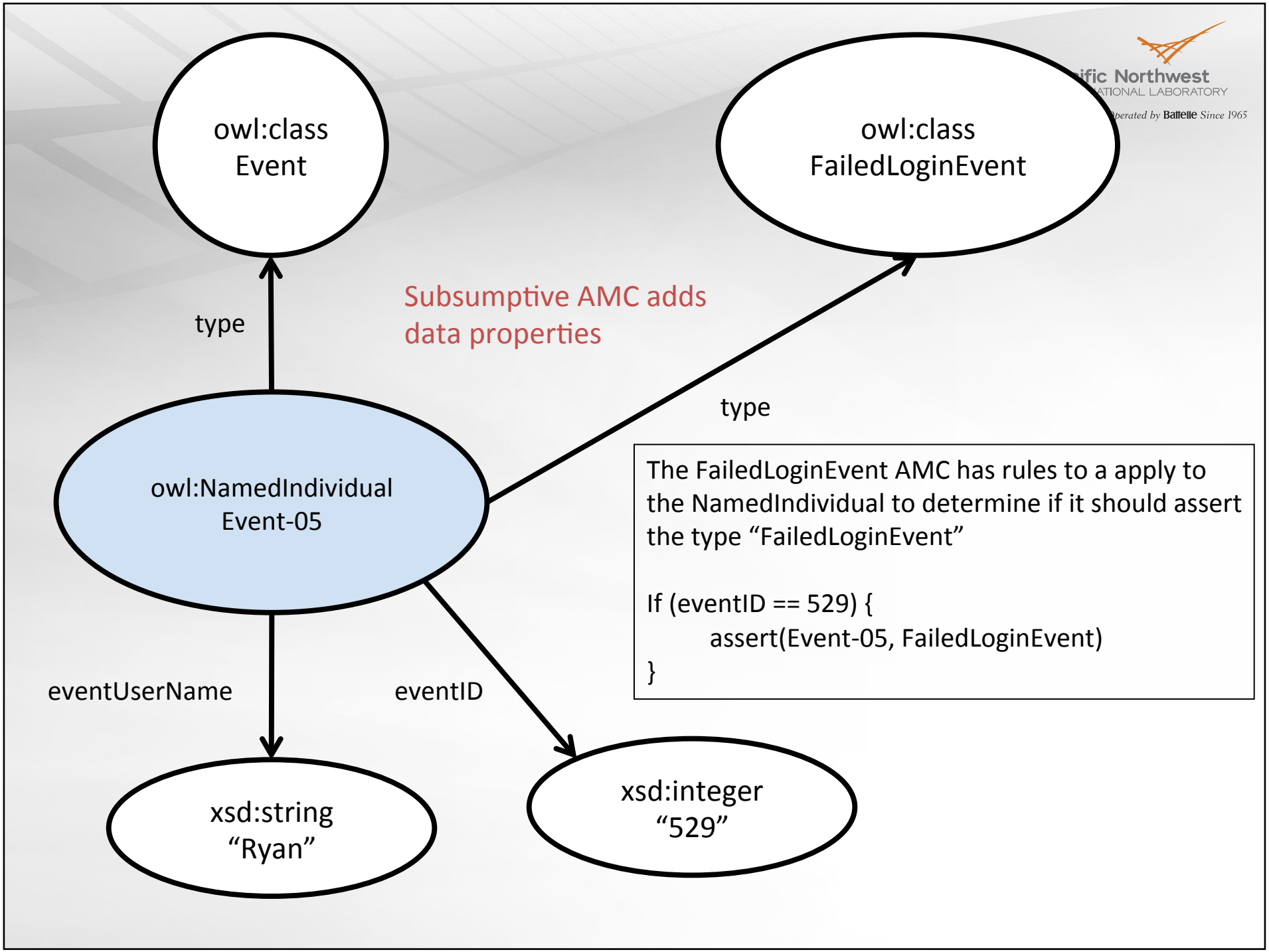
- Composite
  - Asserts a NEW INDIVIDUAL into the semantic graph
  - Asserts object properties between individuals

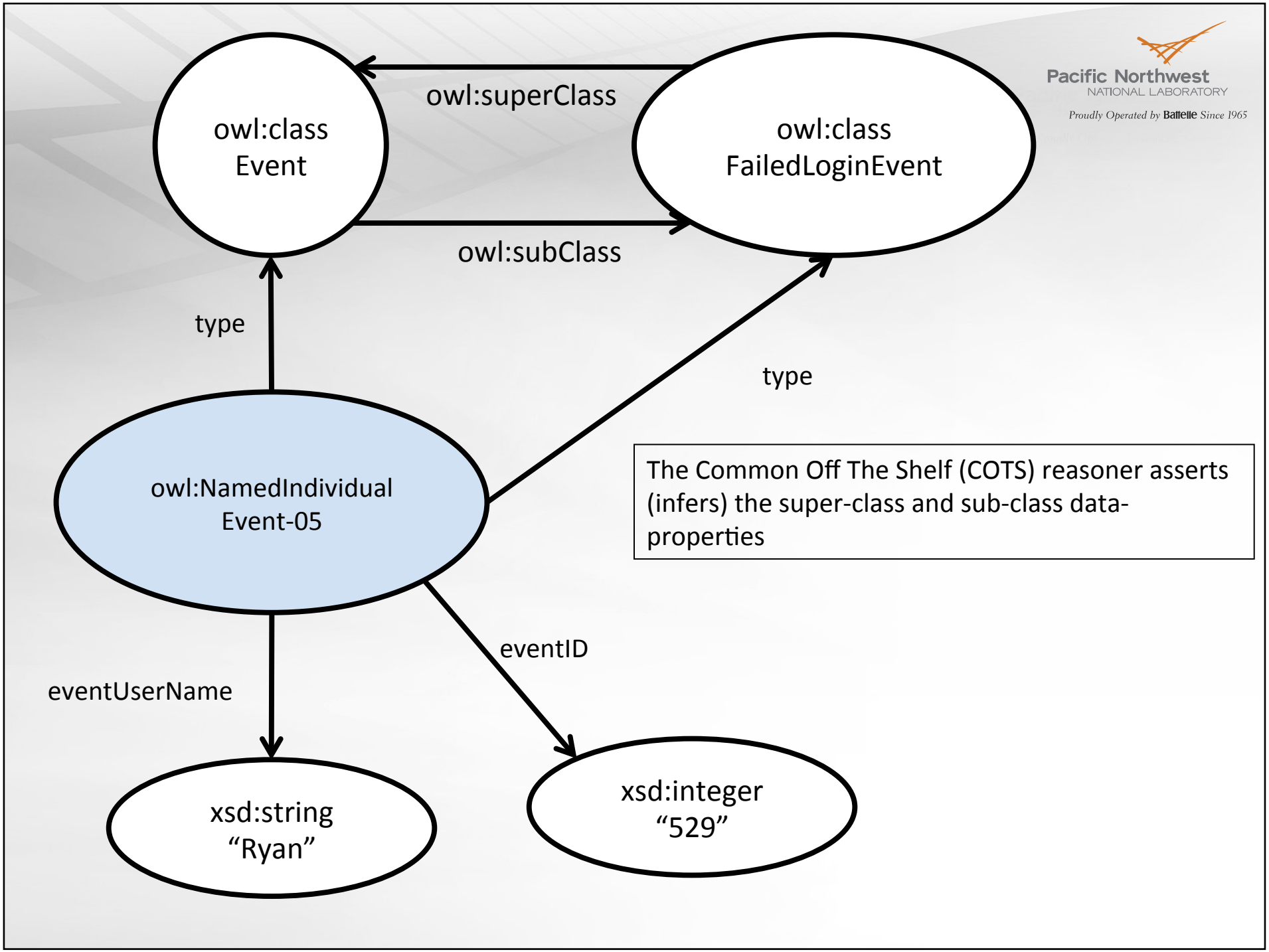


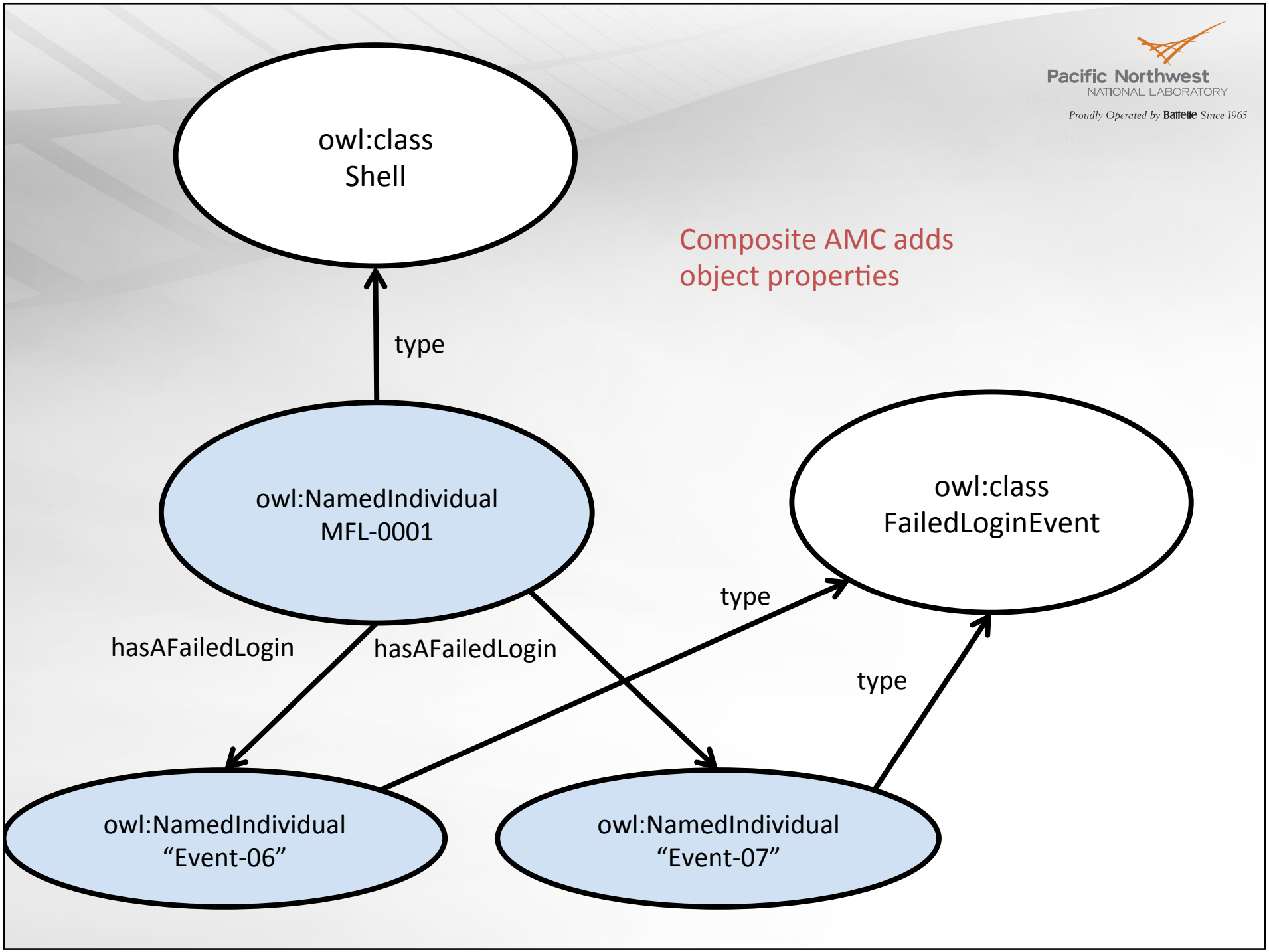
Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

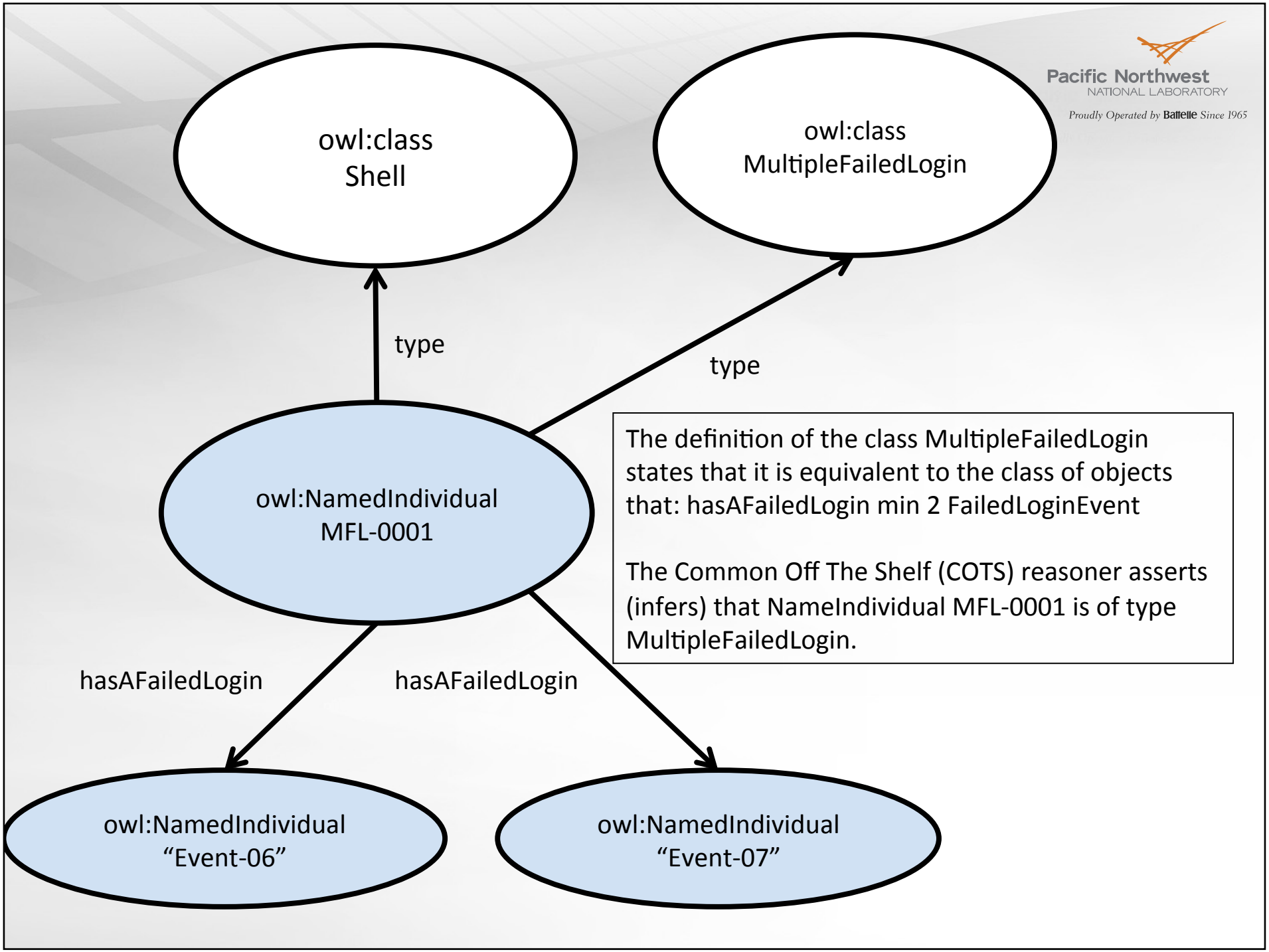
Reifier creates a graph. Then it adds it to the semantic graph we call “Working Memory”



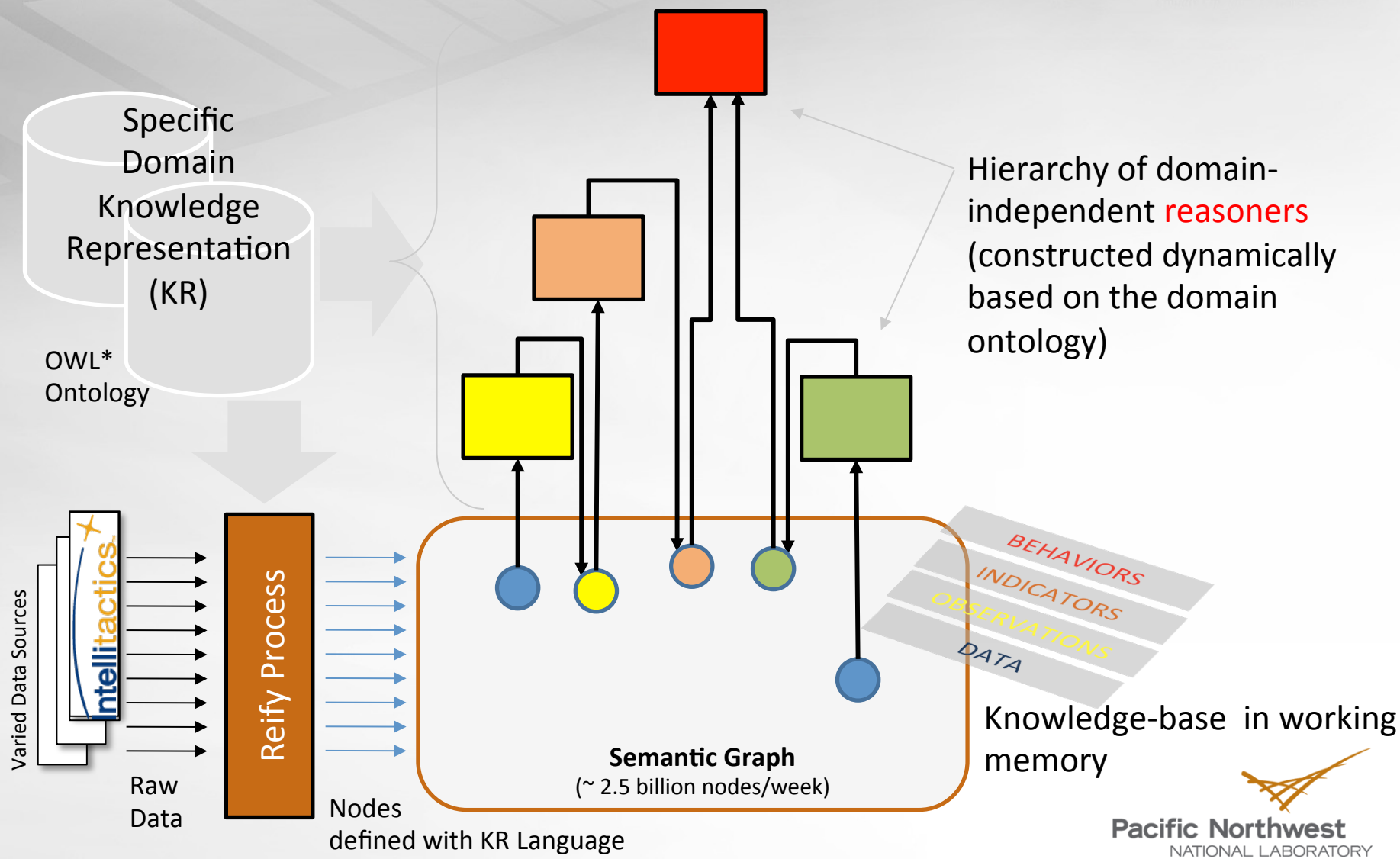

owl:class
Event
owl:class
FailedLoginEvent
type
Subsumptive AMC adds
data properties
type
owl:NamedIndividual
Event-05
The FailedLoginEvent AMC has rules to a apply to the NamedIndividual to determine if it should assert the type "FailedLoginEvent"
If (eventID == 529) {
assert(Event-05, FailedLoginEvent)
}
eventUserName
eventID
xsd:string
"Ryan"
xsd:integer
"529"


owl:class
Event
owl:superClass
owl:class
FailedLoginEvent
owl:subClass
type
type
owl:NamedIndividual
Event-05
The Common Off The Shelf (COTS) reasoner asserts (infers) the super-class and sub-class data-properties
eventID
eventUserName
xsd:string
"Ryan"
xsd:integer
"529"
Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965


owl:class
Shell
Composite AMC adds
object properties
type
owl:NamedIndividual
MFL-0001
owl:class
FailedLoginEvent
type
hasAFailedLogin
hasAFailedLogin
type
owl:NamedIndividual
“Event-06”
owl:NamedIndividual
“Event-07”



CHAMPION Process/Knowledge Flow…

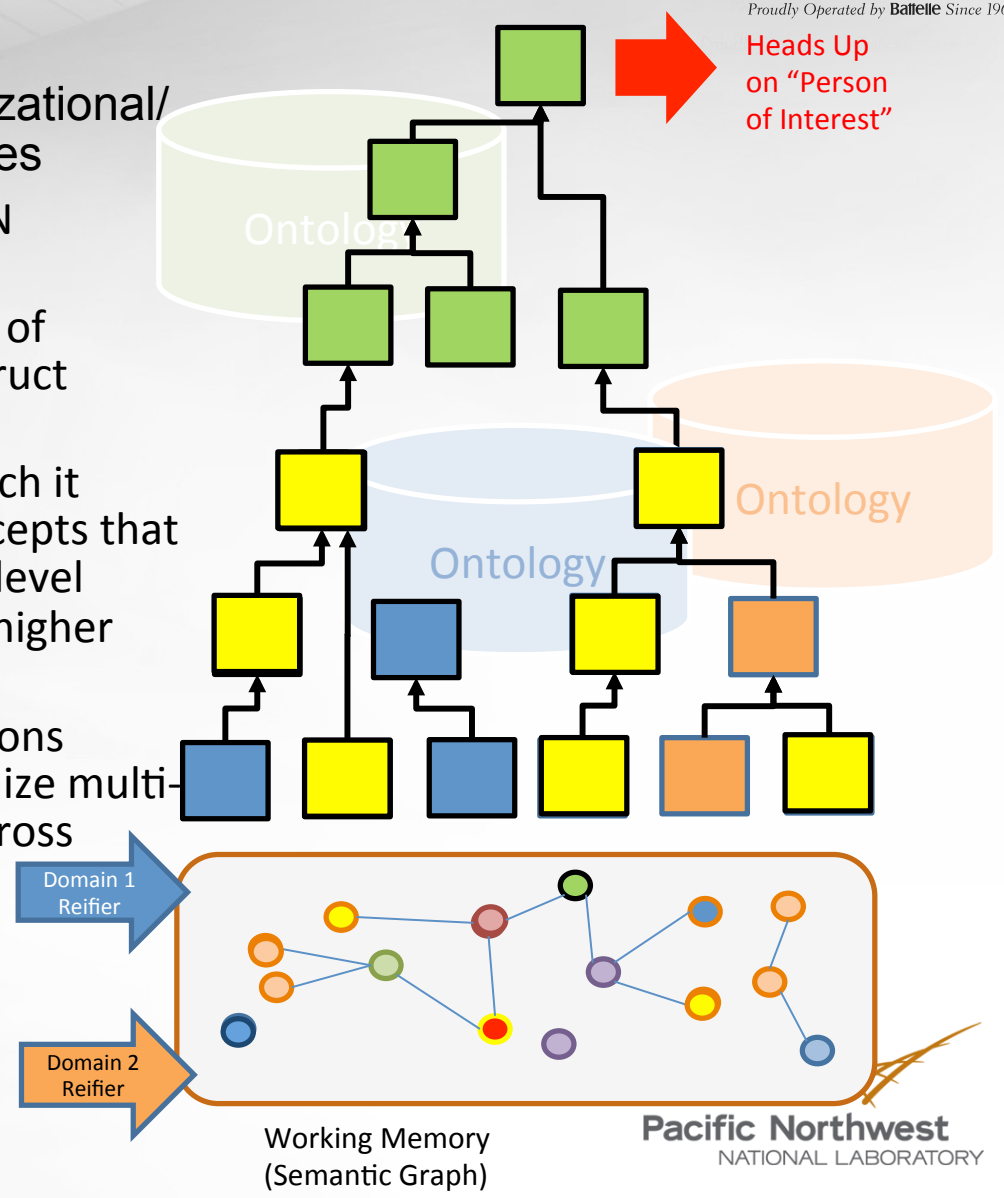


*OWL = Ontology Web Language

We are using Pellet 's OWL-DL Reasoner (Java)

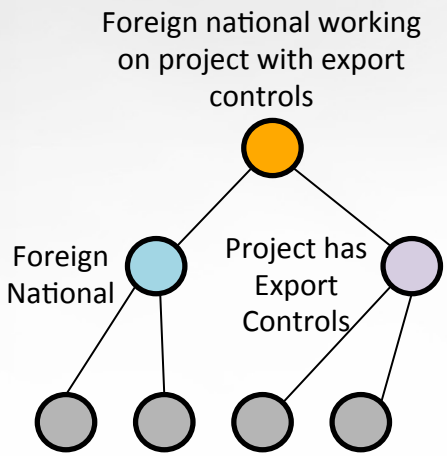
Walk-Through Sample Scenario

- Assume we have cyber and organizational/ behavioral data from various sources
- These data are ingested by CHAMPION processors into Working Memory
- Informed by ontologies, the hierarchy of domain-independent reasoners construct semantic graph
- Each reasoner processes inputs to which it subscribes to recognize "posited" concepts that are activated through previous/lower level inputs as the reasoning progresses to higher levels of abstraction in the hierarchy
- Cross-domain ontological representations enable fusion of information to recognize multi-dimensional patterns that correlate across different data types
- Highest levels of abstraction correspond to behaviors that are deemed "high risk" for insider threat; these are reported as "Persons of Interest" for further monitoring and analysis by security personnel.



Simple Illustration

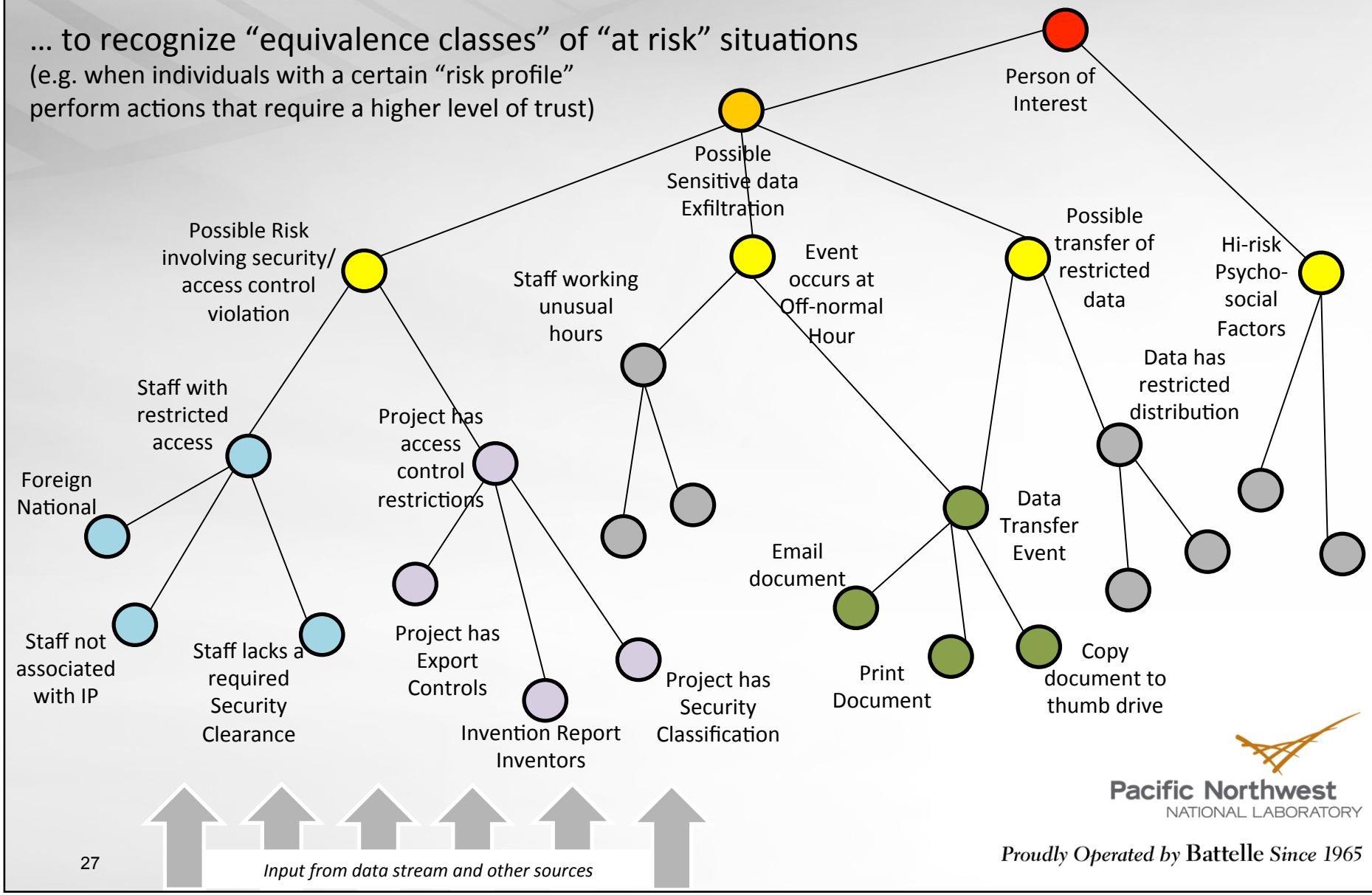
- Cyber/CI Analyst wants to track whether or not there are foreign national staff at a government research Lab who charge some time to a project that has export controls
  - To research this question, the analyst needs to access several different data bases—time reporting, project records, employee records
  - This is a tedious process—a computer can perform this task rapidly and provide the analyst with a list of individuals who meet the criteria
  - This is just the beginning of the analysis process, since there are potentially legitimate reasons for this type of situation
  - Further correlation of cyber or other organizational monitoring data logs may reveal additional situations with increased risk that are associated with this individual—this reflects the sensemaking requirement for recognizing intersecting anomalies.



- This simple scenario may be tweaked in many ways…
  - Find "uncleared" staff working on "classified" project
  - Find staff who access restricted data (e.g., IP data) without authorization
  - etc.

Approach: Configure the System to Reason on an *abstract level* rather than based on exact matches

… to recognize "equivalence classes" of "at risk" situations
(e.g. when individuals with a certain "risk profile" perform actions that require a higher level of trust)



Questions

Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by **Battelle** *Since 1965*

Ryan E. Hohimer
ryan.hohimer@pnnl.gov

Jana Strasburg
jana.strasburg@pnnl.gov