Tools for the Rationalisation & Management of Data Collection – The Semantic Reef.

Trina Myers, Ian Atkinson, Ron Johnstone[2]

1. eResearch Centre, James Cook University, Townsville, Qld
2. Geography, Planning & Environmental Management. University of Queensland, Qld

Data Deluge

- Many non-standardised monitoring efforts – *disparate data* is being generated as we sit here…

- Bottlenecks are developing in the data analysis/data processing phases

Researchers and managers of (say) coral reefs need:
- Less Data &
- More Information!
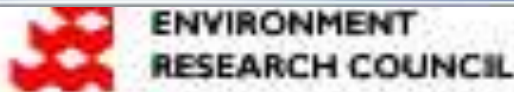
Marine Data Deluge

The Gap

- Current data processing depends on manual manipulation and human intervention
  - Not automatic
  - Slow…

- Existing technologies & tools cannot handle future data analysis requirements
  - Need to explore new technologies
  - Adapt these into new tools

The Challenge

- The challenge is to change current approaches to data analysis through introducing new technologies which will be able to scale

    – Independent of data sources

The Semantic Web

- The Semantic Web allows more efficient ways of managing disparate data

- Making documents **machine readable** not just human readable

The Semantic Web

- Built on Resource Description Format (RDF) triples -
  - Subject - the resource
  - Predicate – linking property
  - Object – value of property
- Each triple can be called a statement
- Anything can be described in a series of triples

Water	has_empirical_formula	H_2O
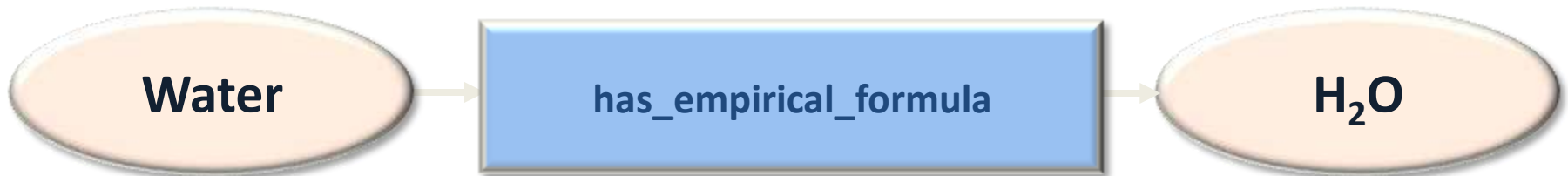
<subject> has a *Property* <predicate> *valued* by <object>

The Semantic Web

- Semantic Web technologies use logic systems to *infer* a conclusion or meaning

Semantic inference can *tame* the Data Deluge

The Ontology

- Ontologies are the foundation of Semantic Web technologies

An ontology is a set of explicit specifications, terms and relationships of a concept.

Formal definitions and axioms are used to constrain the interpretation of the specifications to create, share and re-use "computer-understandable" knowledge.

(Antoniou et al. 2001).
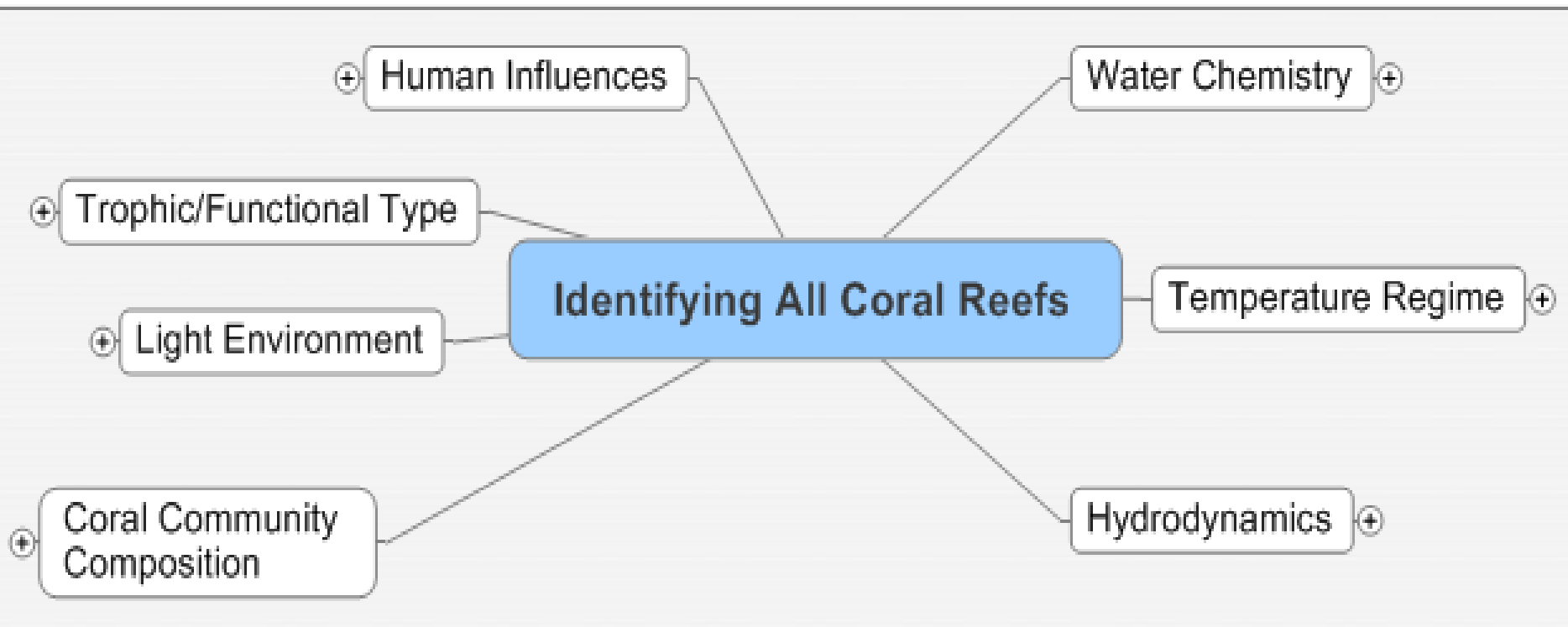
The Semantic Reef Project

- An eco-informatics application focused on automated data processing, problem solving and knowledge discovery for reef management

- The goal is to harness:
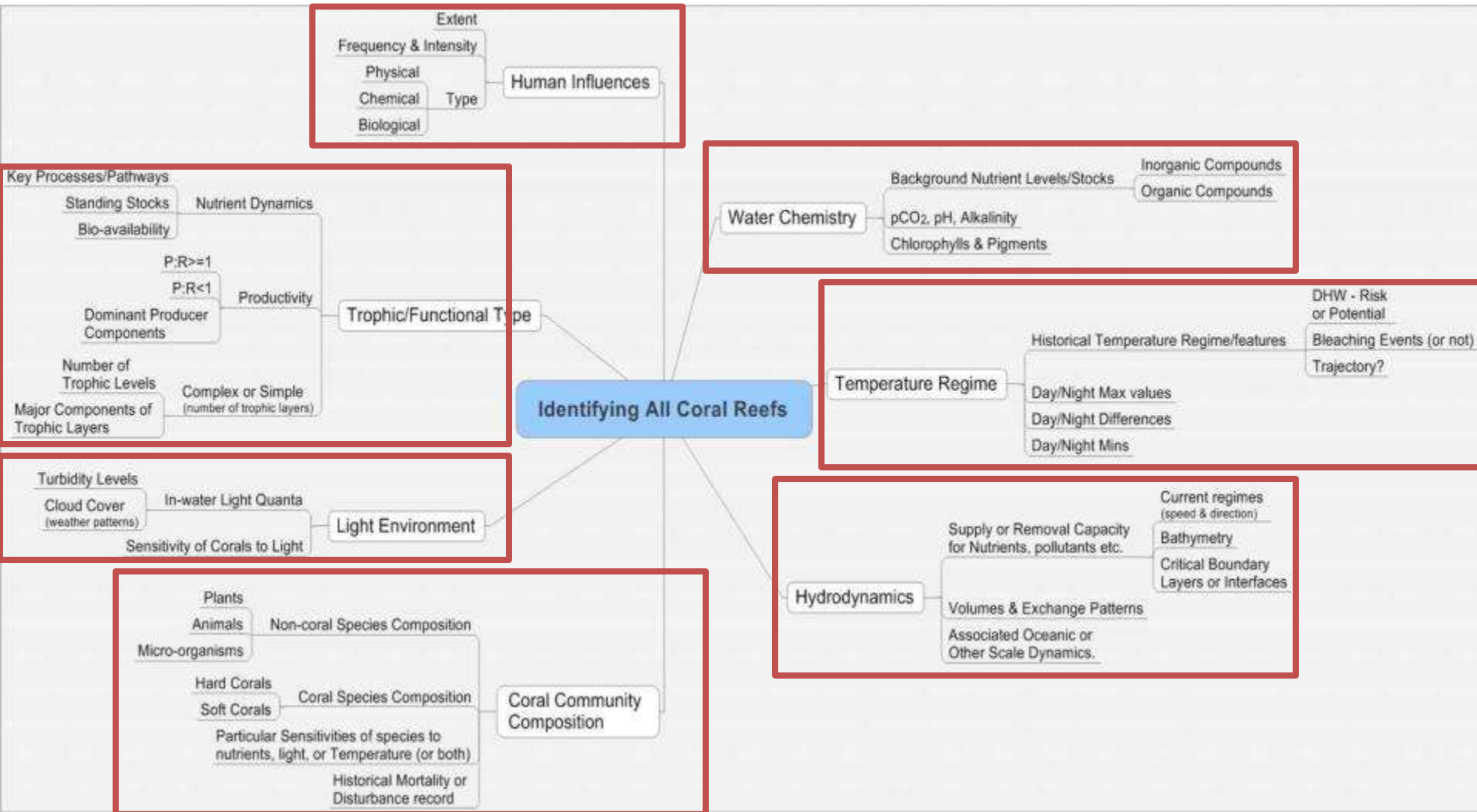 - The Semantic Web
  - Scientific workflows
  - Grid computing

    into a hypothesis-based research tool

Modular Design

- Create a modular Knowledge Base:
  - Domain experts perspective for functions of any coral reef
  - Semantic building blocks for Ontology engineering

Courtesy of Ron Johnstone and Glen Holmes
Centre for Marine Science, University of Queensland

Modular Design of Ontologies

Courtesy of Ron Johnstone and Glen Holmes
Centre for Marine Science, University of Queensland

The Research Question

Can *semantic inference* be implemented in a software hypothesis tool to facilitate research?

– Investigate new semantic technologies

– Investigate the impact of new technology on data integration with a focus on marine data
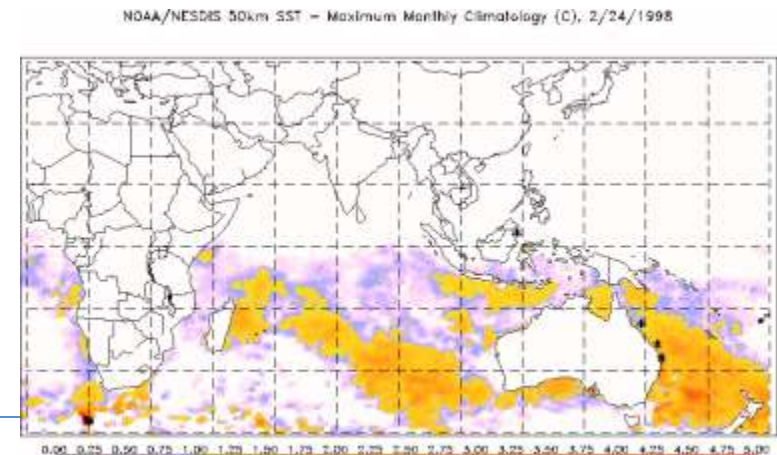
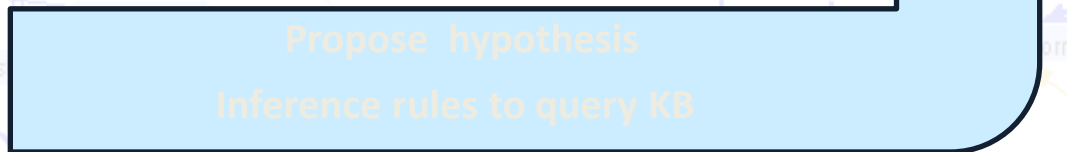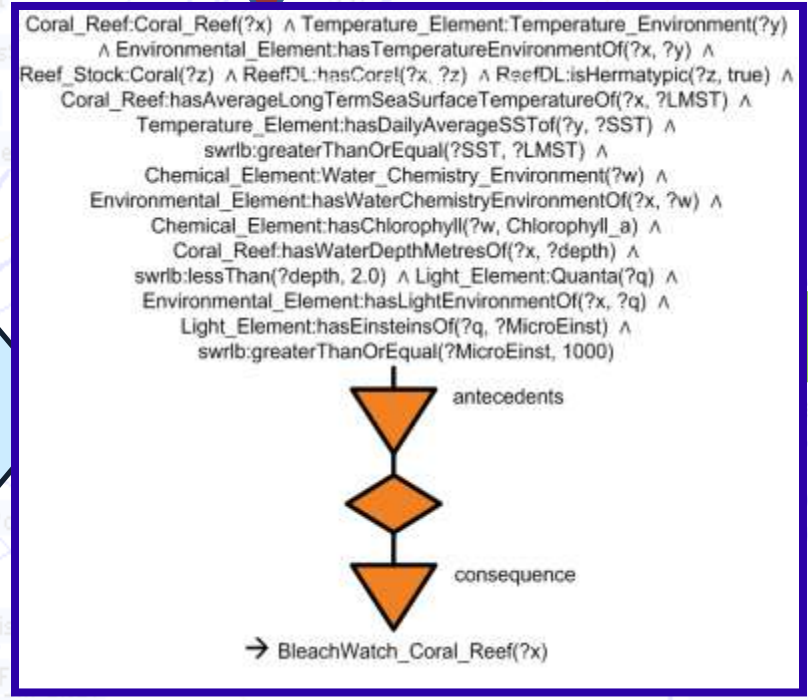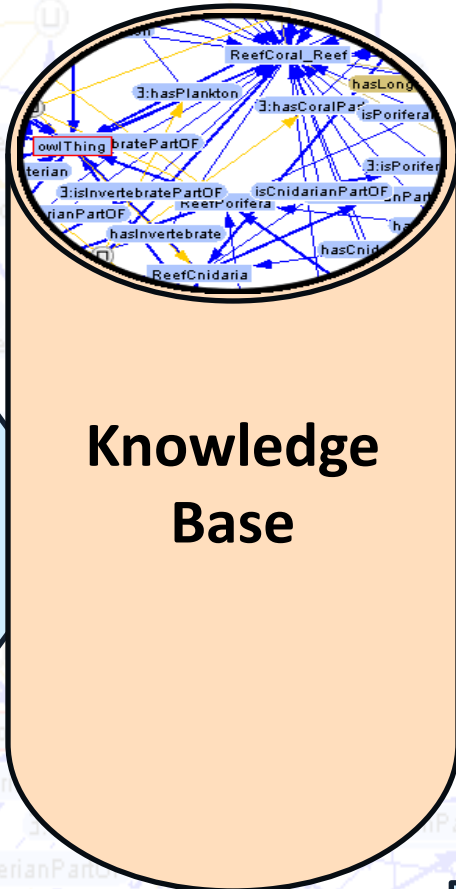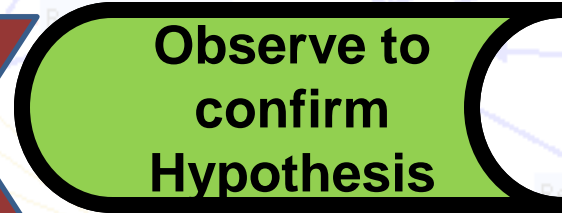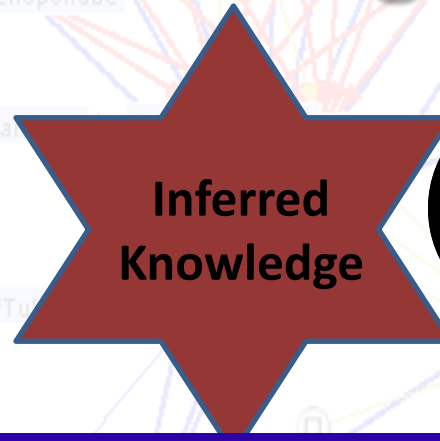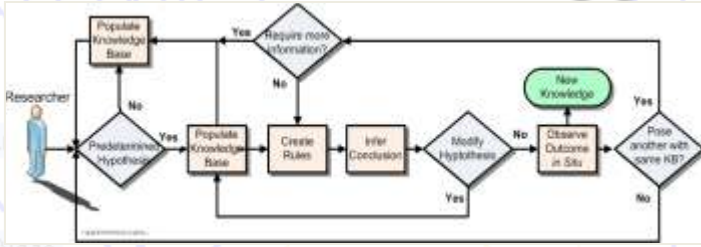Using the Semantic Reef architecture

Validating the Approach
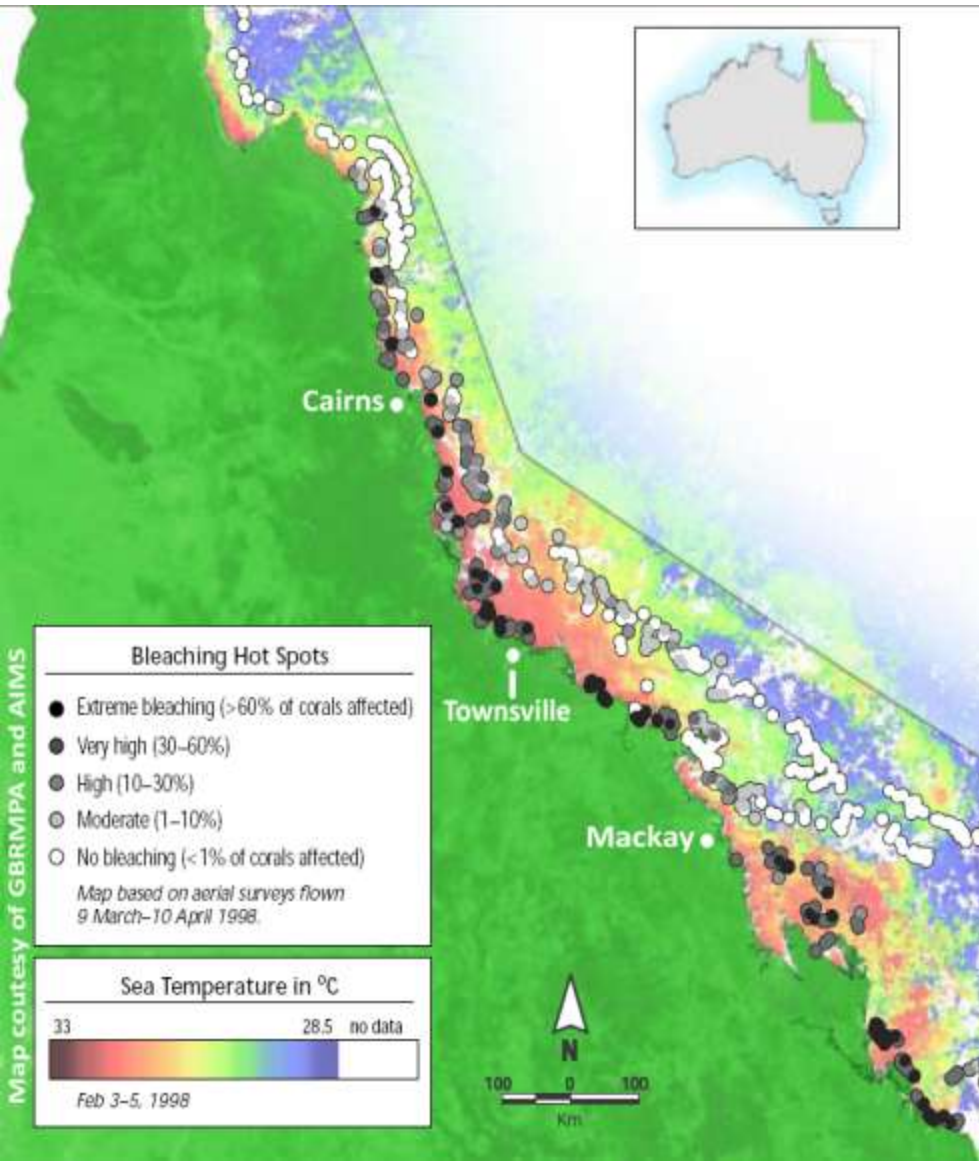
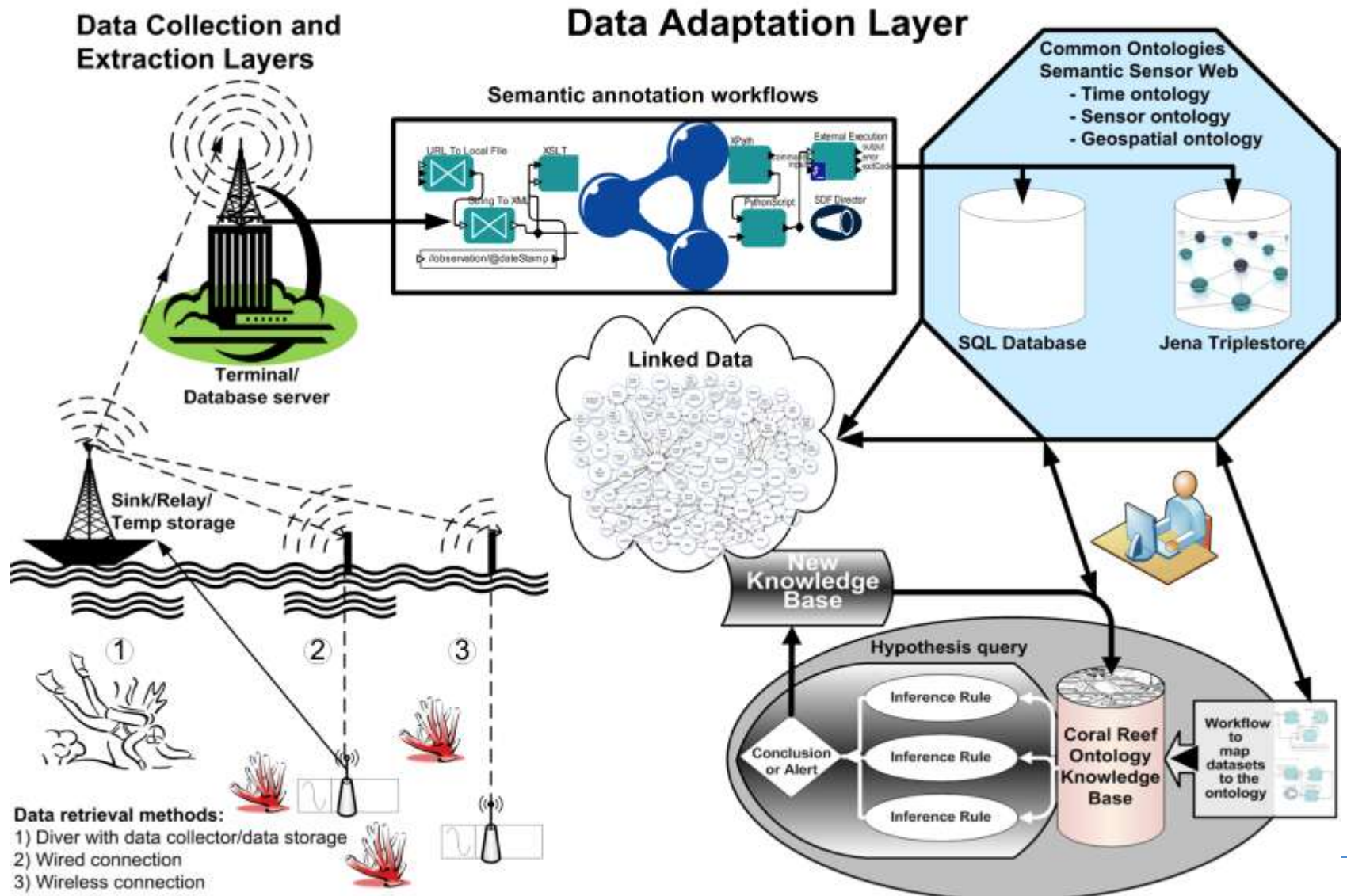Reverse-hypothesis approach to ground-truth the system against historic events

Test case:

Coral Bleaching on the GBR – 1998 and 2002

Flexible Hypothesis Design

Inferred Knowledge

Observe to confirm Hypothesis

Knowledge Base

Coral_Reef:Coral_Reef(?x) ∧ Temperature_Element:Temperature_Environment(?y)
∧ Environmental_Element:hasTemperatureEnvironmentOf(?x, ?y) ∧
Reef_Stock:Coral(?z) ∧ ReefDL:hasCoral(?x, ?z) ∧ ReefDL:isHermatypic(?z, true) ∧
Coral_Reef:hasAverageLongTermSeaSurfaceTemperatureOf(?x, ?LMST) ∧
Temperature_Element:hasDailyAverageSSTof(?y, ?SST) ∧
swrlb:greaterThanOrEqual(?SST, ?LMST) ∧
Chemical_Element:Water_Chemistry_Environment(?w) ∧
Environmental_Element:hasWaterChemistryEnvironmentOf(?x, ?w) ∧
Chemical_Element:hasChlorophyll(?w, Chlorophyll_a) ∧
Coral_Reef:hasWaterDepthMetresOf(?x, ?depth) ∧
swrlb:lessThan(?depth, 2.0) ∧ Light_Element:Quanta(?q) ∧
Environmental_Element:hasLightEnvironmentOf(?x, ?q) ∧
Light_Element:hasEinsteinsOf(?q, ?MicroEinst) ∧
swrlb:greaterThanOrEqual(?MicroEinst, 1000)

antecedents

consequence

→ BleachWatch_Coral_Reef(?x)

Propose hypothesis
Inference rules to query KB

SEMAT and the Semantic Reef

Outcomes

- A new approach to automate data analysis

- A means to automatically extract or disclose unknown relationships and/or phenomena in the data

- Based on the Semantic Web, scientific workflows and grid computing technologies

- Proven to handle disparate data (tested on coral bleaching experiment)

- Integrated with the SEMAT system

Implications and Applications

The Semantic Reef system offers a different approach to the development and execution of observational hypotheses in the marine domain

– Re-usable design for any coral reef worldwide

– Help inform reef Managers in data collection and reef monitoring decisions

– Suitable for other lines of enquiry (medical, business, earth sciences…)

Questions?

Leave Jarrod alone!

Instead e-mail:

Trina.Myers@jcu.edu.au

Ian.Atkinson@jcu.edu.au