A step towards capturing the flux in scientific knowledge
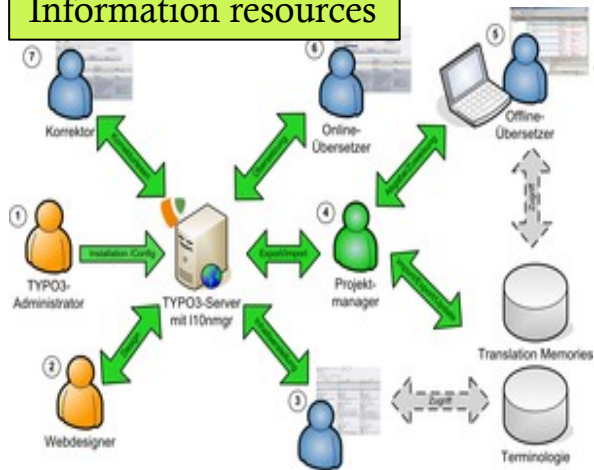
PhD Departmental Seminar
Dept. of Computer Science, University of Auckland

Prashant Gupta
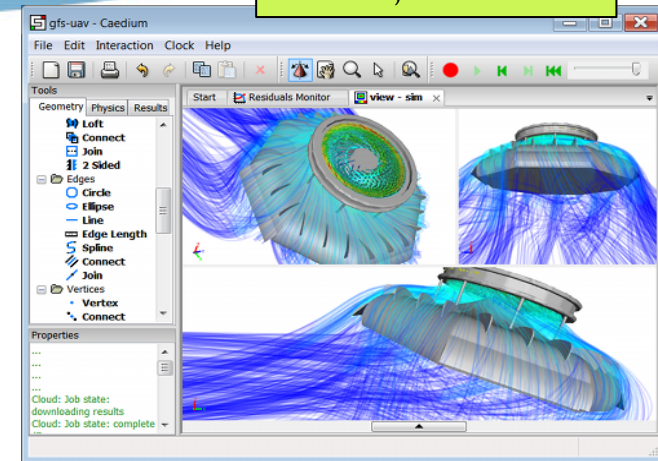p.gupta@auckland.ac.nz

THE UNIVERSITY OF AUCKLAND
NEW ZEALAND

Context of eResearch

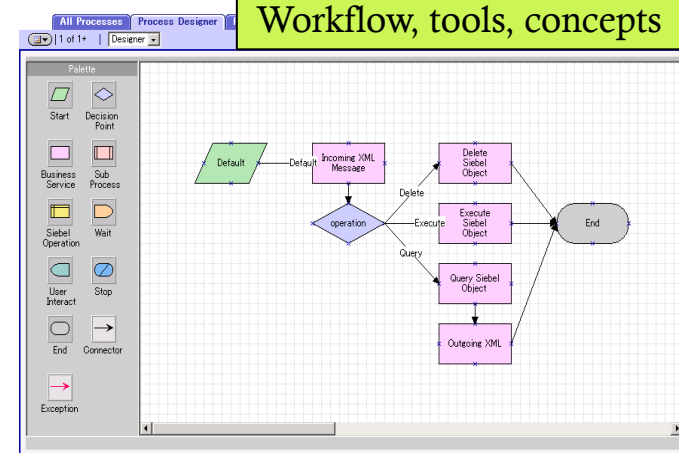
Information resources

Models, simulations

Data, observations, instruments

Collaborations

Workflow, tools, concepts

CHANGE
IS THE ONLY
CONSTANT
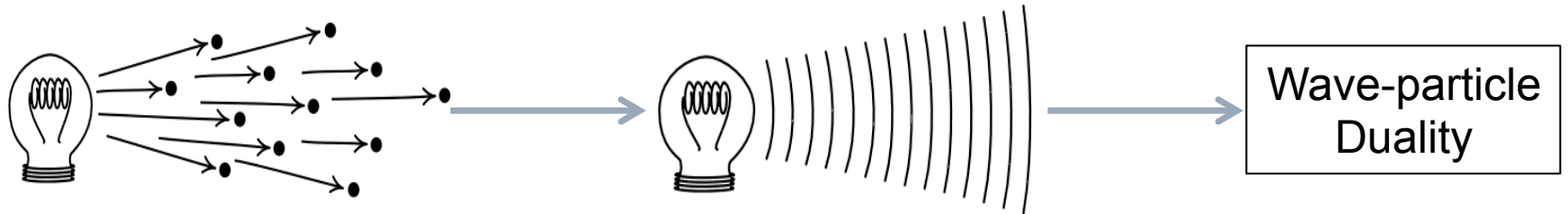
"The flux of things is one ultimate generalization around which we must weave our philosophical system."

--Alfred N. Whitehead, Process and Reality

http://smeitexpo2011.blogspot.co.nz/2010/11/era-of-technological-revolution.html

Conceptual transition in Physical Optics

18th Century – Light as material corpuscles

Early 20th Century – Light as wave particles

Wave-particle Duality

Knowledge Representation

DATABASES


```
<?xml version="1.0"?>
<quiz>
 <question>
  Who was the forty-second
  president of the U.S.A.?
 </question>
 <answer>
  William Jefferson Clinton
 </answer>
 <!-- Note: We need to add
  more questions later.-->
</quiz>
```
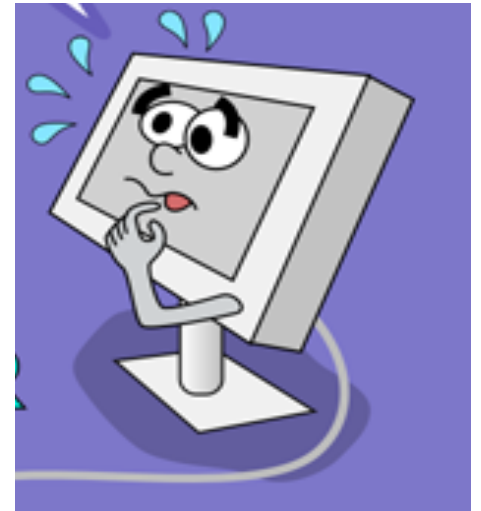
XML

RDF

Ontology

- Share common understanding
- Enable re-use
- Support computational reasoning

The Problem

How our contemporary computational systems deal with the flux (or conceptual change) in scientific knowledge and its implications ?

First aspect

How do ontologies deal with conceptual change

Geological revolution

Before

After

Thagard, P. (1992). Conceptual revolutions. Princeton University Press.

Ontologies – often static in nature

o They do not support continuous revision and refinement

o Problems:
- o Complexity
- o Dependencies

Proposed solution

Formalize representation of conceptual changes and their effects, which would facilitate to automate some aspects of the process of revision

Second aspect

Implications of conceptual change on dependent applications

Semantic Heterogeneity

o Same concepts but different interpretations
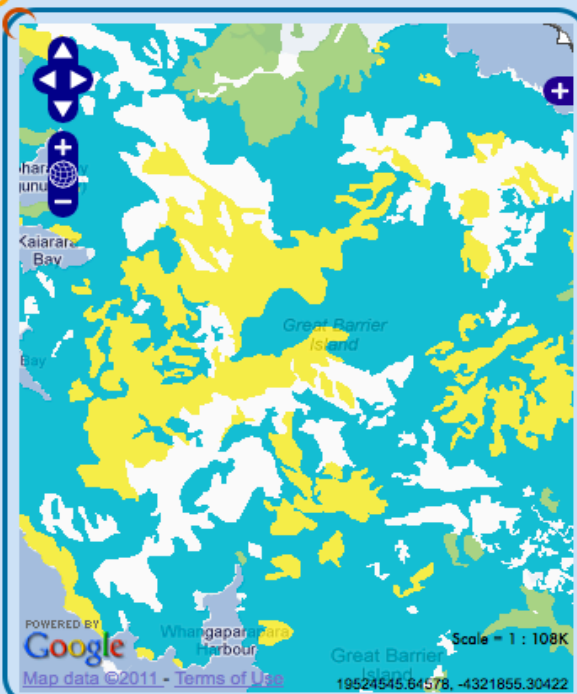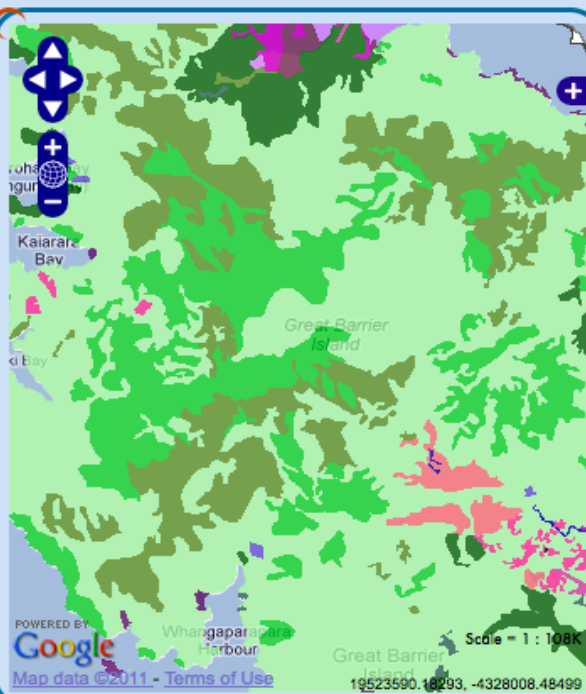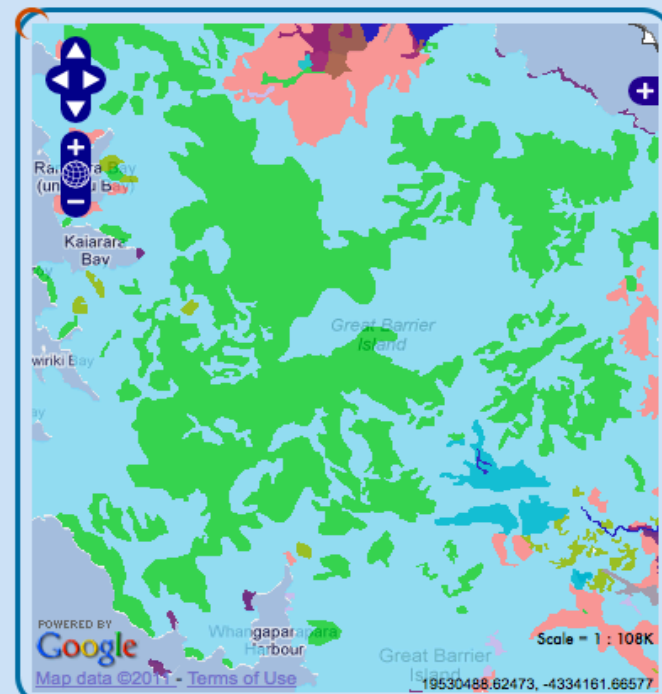
o Diverse concepts but same interpretations

Geographical Maps

o Can we use a new categorical scheme for existing geospatial data?

o How to compare or integrate maps based on different categorical schemas or made at different times or places?

Panel 1 (left map):

Raiapara Bay (un...d Bay)
Kaiarara Bay
...wiriki Bay
...ay
Great Barrier Island
Whangaparapara Harbour
Great Barrier Island
POWERED BY
Google
Map data ©2011 - Terms of Use
Scale = 1 : 108K
_19530488.62473, -4334161.66577

Panel 2 (middle map):

...rohia...ra ...ngun...
...ki Bay
Kaiarara Bay
Great Barrier Island
Whangaparapara Harbour
Great Barrier Island
POWERED BY
Google
Map data ©2011 - Terms of Use
Scale = 1 : 108K
19523590.18293, -4328008.48499

Panel 3 (right map):

...rohara...ra ...gunu...
...Bay
Kaiarara Bay
Great Barrier Island
Whangaparapara Harbour
Great Barrier Island
POWERED BY
Google
Map data ©2011 - Terms of Use
Scale = 1 : 108K
19524545.64578, -4321855.30422

Left legend panel:

Show: ○ Spatial legend, ● Virtual legend, ○ Ontology
... LCDB1 - North Island ...

Class	Equiv. to	Superclass of
BARE_GROUND	BARE_GROUND	Afforestation (im...
COASTAL_SANDS	COASTAL_SANDS	Coastal Sand and...
COASTAL_WETLANDS	COASTAL_WETLANDS	Flaxland Her...
INDIGENOUS_FOREST	INDIGENOUS_FOREST	Broadlea...
INLAND_WATER	INLAND_WATER	Lake and Pond...
INLAND_WETLANDS	INLAND_WETLANDS	Herbaceous Fresh...
MANGROVE	MANGROVE	Mangrove
MINES_DUMPS	MINES_DUMPS	Dump Surfac...
		Major Sh...

Middle legend panel:

Show: ○ Spatial legend, ● Virtual legend, ○ Ontology
... LCDB2 ...

Class	Equiv. to	Subclass of
Afforestation (imaged, post LCDB 1)		BARE_GROUND
Afforestation (not imaged)		Afforestation (imaged, post LCDB 1)
Alpine Grass-/Herbfield	Herbfield	
Alpine Gravel and Rock	Alpine Gravel and Rock	
Broadleaved Indigenous Hardwoods	Broadleaved Indigenous Hardwoods	INDIGENOUS_FOREST
Built-up Area	Built-up Area	URBAN
Coastal Sand and Gravel	Coastal Sand and Gravel	COASTAL_SANDS
Deciduous...	Deciduous...	PLANTED_FOREST

Right legend panel:

Show: ○ Spatial legend, ● Virtual legend, ○ Ontology
... ECOSAT legend ...

Class	Equiv. to			
Bare Ground	Lowland	Mines and Dumps		
	Transport Infrastructure	Alpine Grass...		
Exotic Forest	Deciduous Hardwood			
Herbaceous Vegetation	Inland Wetland	cNon Pastoral	High Producing Exotic Grassland	
	Orchard and Perennial	Short Rotation Cropland	Vineyard	
Indigenous Forest	Indigenous Forest	Major Shelterbelt		M...
	Other Exotic Forest	Pine Forest - Closed Canopy	Pine Open...	
Kanuka	Kanuka			

Proposed solution

Create a category versioning system and have explicit connections of each version with their corresponding applications

Third aspect

How can we capture the flux

Why?

- How we reached the current state of knowledge?

- What factors and processes were involved?

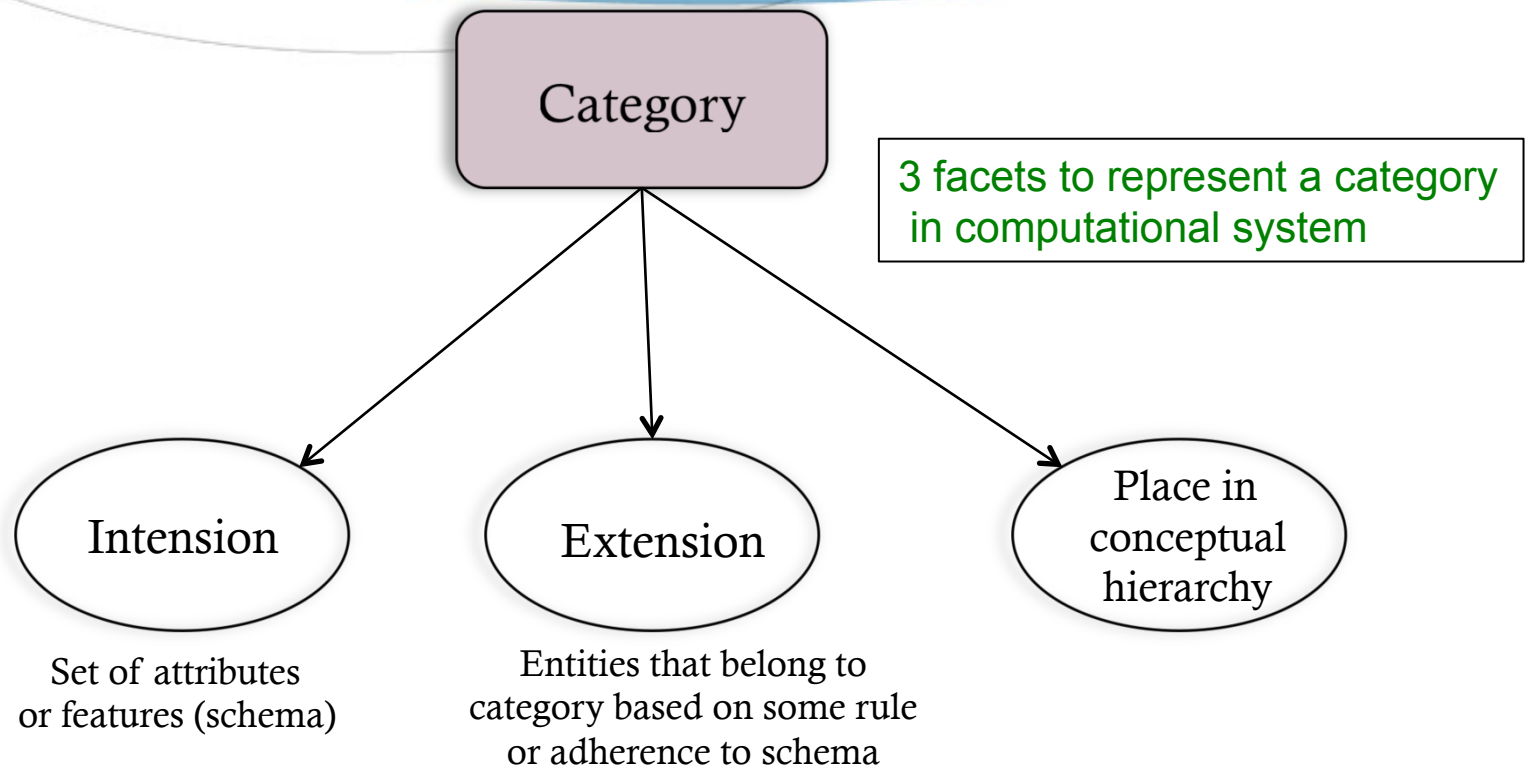- Why it is the way it is (and not some other way)?

What are we missing?

o The source of interpretation behind knowledge formation, i.e. the process of generating knowledge (categories) from raw data

o We argue that we require an approach to represent our scientific knowledge that reflects:

   o The scientific processes involved in its creation and revision

   o The evolution of scientific knowledge over time

Proposed solution

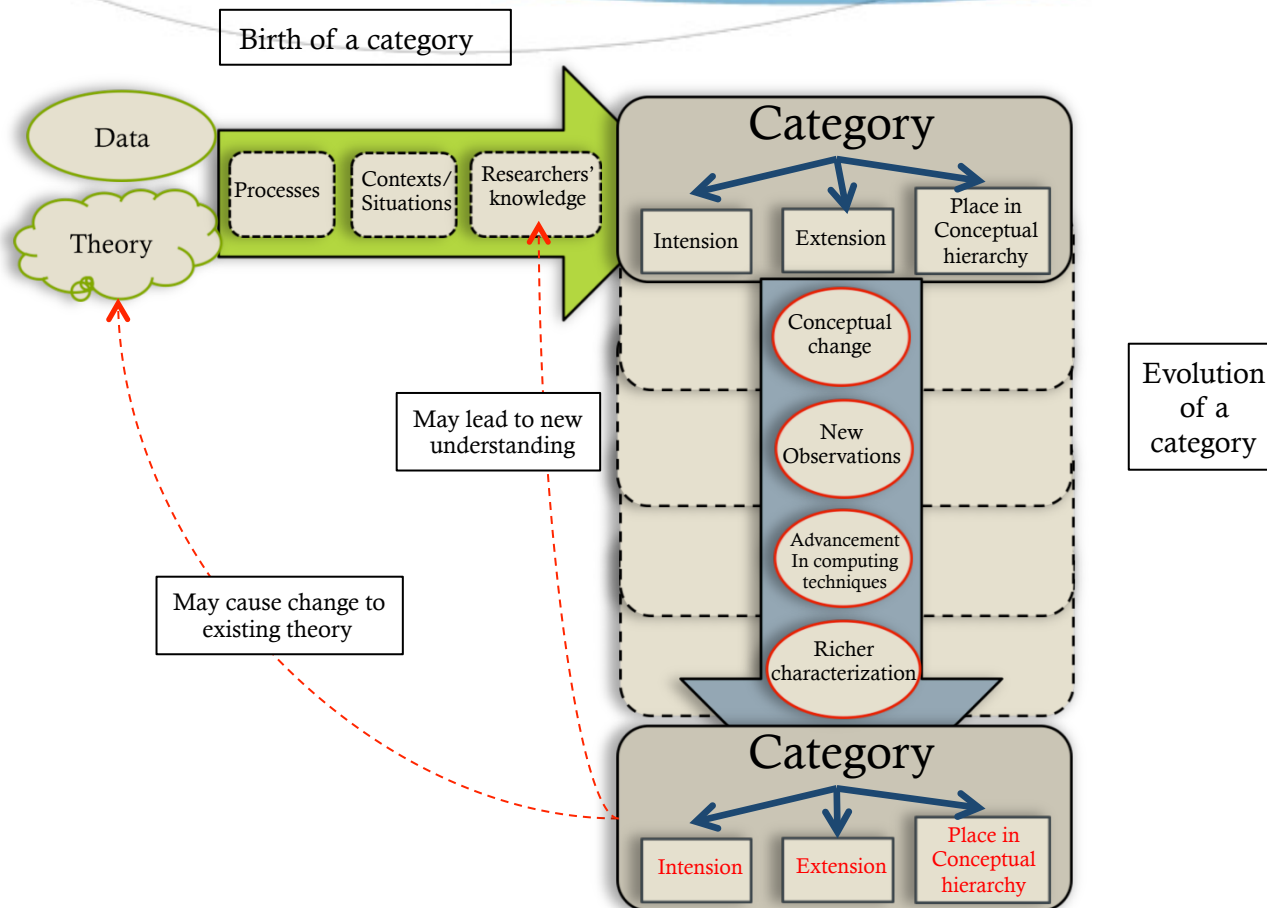Connect categories with the process of science that drives their formation and revision

Category Representation

Category

3 facets to represent a category in computational system

Intension

Set of attributes or features (schema)

Extension

Entities that belong to category based on some rule or adherence to schema

Place in conceptual hierarchy

Does the current representation of a category fully explain its existence and identity and conveys the complete meaning associated with it ?

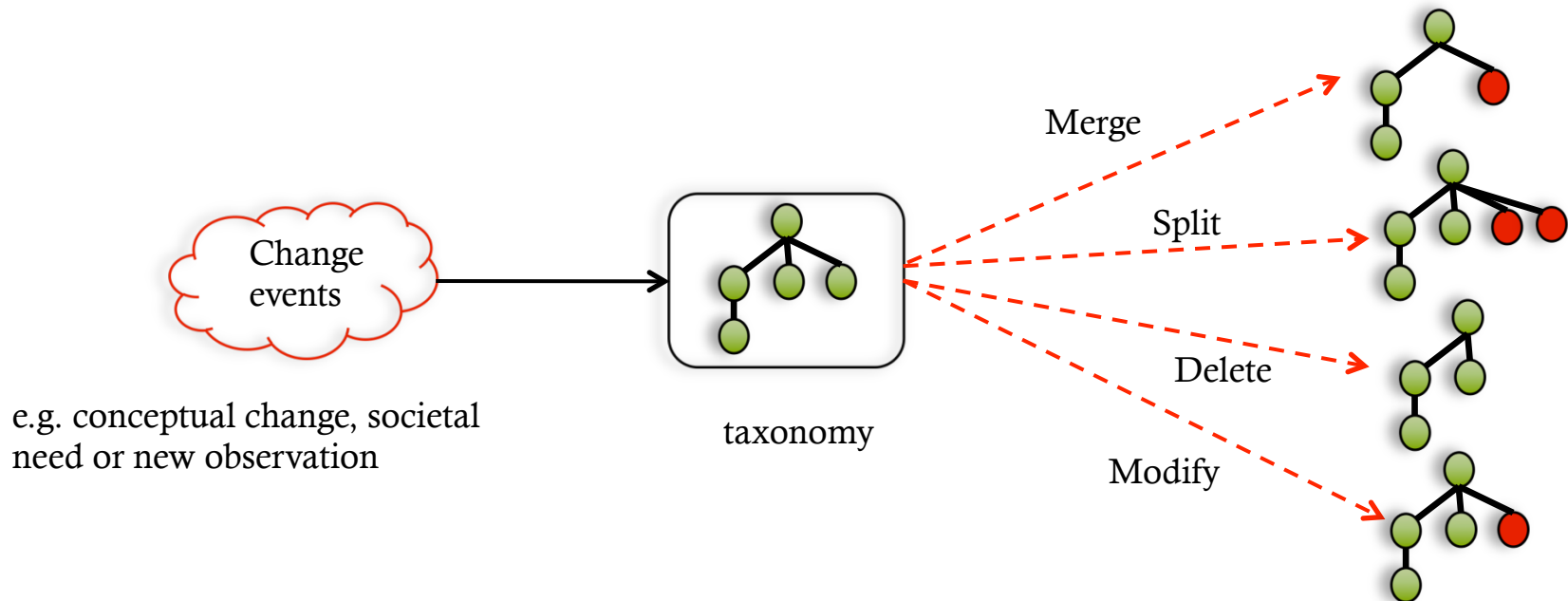# Birth and evolution of a category
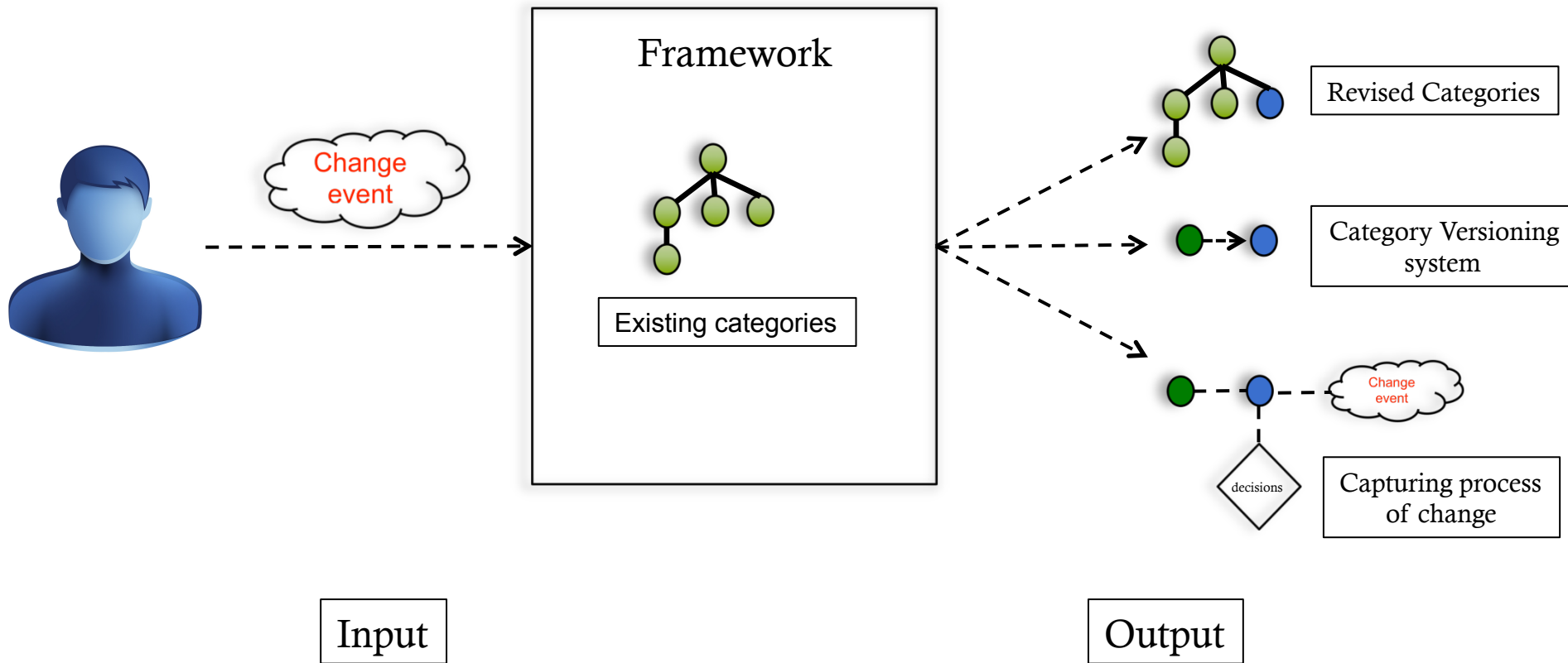
Research Goal

Three aspects of our goal:

1. Facilitating the automation of category revision process - dynamic

2. Connecting categories with the processes that were involved in their revision – Living and more meaningful

3. Create a category versioning system and have explicit connections of each version of categories with their corresponding dataset and applications – Do not lose previous knowledge if there is a conceptual change

Research objectives

o   Build a conceptual model to explore the factors (change events) that may cause changes to category and their corresponding outcomes, i.e. inputs and outputs relating to different kinds of change

Change events

e.g. conceptual change, societal need or new observation

taxonomy

Merge

Split

Delete

Modify

Research Objectives

Evaluation

o Evaluate the framework using datasets and categories that has already gone through some conceptual changes.

   o Good examples available relate to taxonomic revision in biology and landcover mapping in geography

o Evaluate the benefits of the framework by connecting two datasets based on different versions of categories

Questions ??

Thanks to

        Prof Mark Gahegan (Supervisor)

        Prof Gillian Dobbie (Co-supervisor)

        Prof Pat Langley

        Fellow students