

Managing Knowledge in Environmental e-Science

Dr Alun Preece
Computing Science, University of Aberdeen



Some types of scientific knowledge

Publications

- formal/reviewed
- “grey”
- associated artefacts

People

- expert directories
- communities of practice

Projects

- formal/funded
- working groups

Experiment datasets

- formally curated
- raw/pre-processed
- *in vivo* / *in vitro* / *in silico*

Scientific method

- experiment workflow
- knowledge roles:
hypotheses, observations,
predictions, deductions, ...
- Discourse & natural
arguments: proof,
refutation, agreement, ...



Metadata

Metadata is “data about data” - it describes the **resources** available in an information system.

In a knowledge management system, metadata is information about the **resources available in the system**.

Increasingly, the **Web** is providing a uniform platform for sharing information and knowledge resources - things denoted by **URIs**.

Metadata makes those resources **machine-processable** - it facilitates the **finding & fusing** of information on the Web.

As the Web grows in importance to the e-Society, the **value of metadata** becomes greater.



Metadata example: Dublin Core



The **Dublin Core Metadata Initiative** originated with the library community, intended to cover the properties of information artefacts in a library (including digital libraries).

The DC element set spans the contents of an electronic "library card":

title	date
creator	type
subject	format
description	language
publisher	<i>etc</i>

Most information resources in a knowledge management system have these common properties, so the Dublin Core element set has wide applicability to KM....



Managing publications

Reasonably well-explored area

Prominent systems

- EPrints
- DSpace

Key features

- free
- cover peer-reviewed papers as well as "grey" literature
- OAI-compliant
- interoperable

Some issues

- getting people to use the repositories
- metadata standards
- architecture (e.g. peer-to-peer)
- provenance
- quality



KM of people & projects - an example

AKT CS AKTive Space Take a tour through CS AKTive Space

About this page research area/region region/research area

Research area

Radial:

50 miles

Map:

uk-political

Researcher

Top 5 10 20 unlimited

Order by Grant total RAE result

Theory of Computation
 mathematical logic and formal languages
 logics and meanings of programs
 analysis of algorithms and problem computation by abstract devices
 general

Mathematics of Computing
 probability and statistics
 discrete mathematics
 numerical analysis
 general

Information Systems
 information interfaces and presentation
 information systems applications
 information storage and retrieval
database management
 general

Computing Methodologies
 document and text processing
 simulation and modeling



KN Brown
 P Edwards
 AD Preece
 TJF Norman
 JRW Hunter
 PMD Cray
 DH Sleeman
 EB Reiter
 MW Freeston
 PJF Lucas
 CJL Kemp



People+project KM - some issues

Far less standardised than publications

- no equivalent to Dublin Core, though FOAF (Friend-Of-A-Friend) is gaining ground
- several "portal schemas" in substantial use
- little interoperability

CAS is mainly populated by harvesting data

- sites don't provide it
- when they do, it isn't in the right format
- the 90/10 issue is key

Named entity reconciliation is a big problem

- e.g. "Alun Preece" vs "A Preece" vs "A D Preece"

Provenance & information quality (again)



Managing experiment datasets

Sticks & carrots:

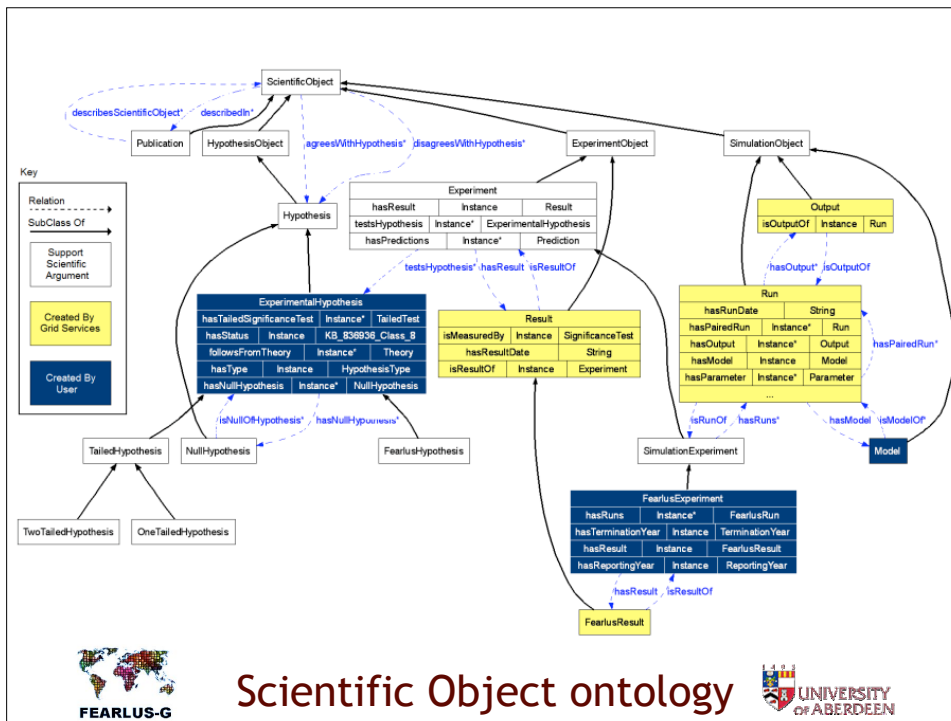
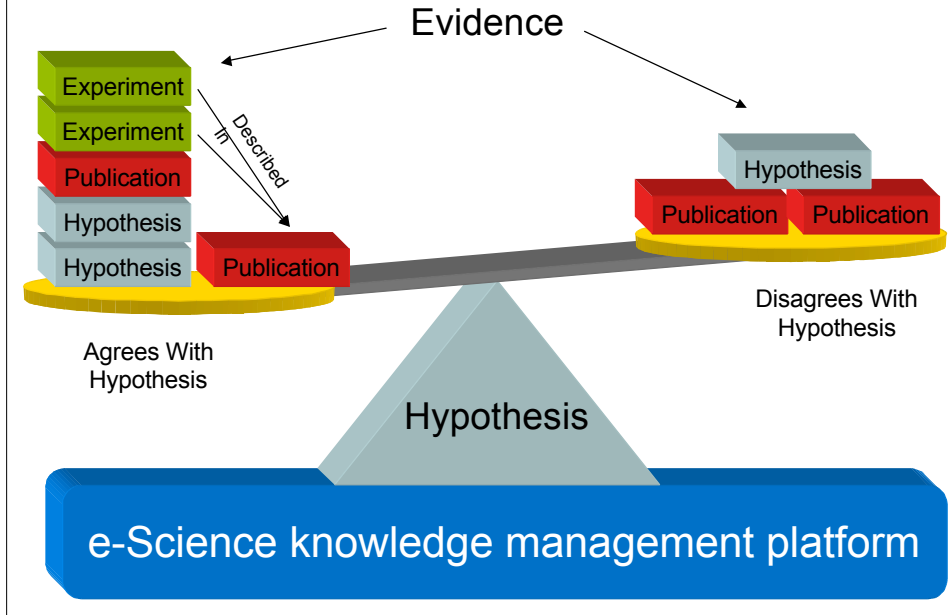
- an increasing number of journals (e.g. in biology) require published datasets
- funding councils are becoming more concerned with reusability of results
- standard data formats are becoming more common, especially those based on XML
- datasets share some metadata characteristics with other published artefacts

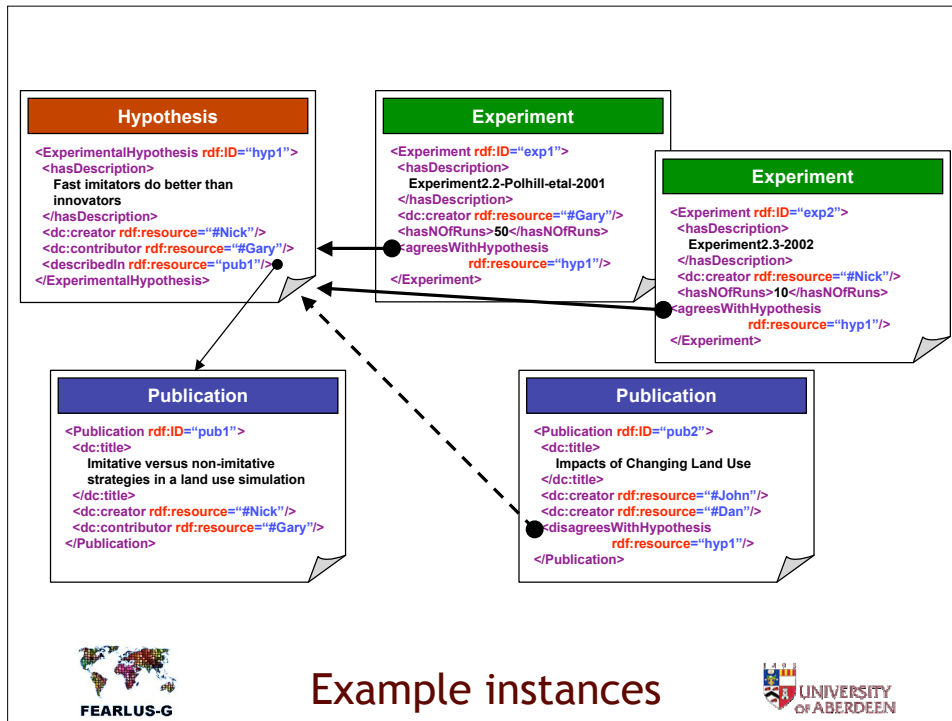
Issues:

- capturing context, for reuse of data
- cost of re-use compared to simply re-generating the data
- inevitably, provenance & information quality



KM & the scientific method





The Semantic Web W3C[®]

“An extension of the current Web in which information is given well-defined meaning, better enabling computers and people to work in cooperation.

“It is the idea of having data on the Web defined and linked in a way that it can be used for more effective discovery, automation, integration, and reuse across various applications.”

Tim Berners-Lee, James Hendler and Ora Lassila
The Semantic Web, Scientific American, May 2001

Logos: FEALUS-G, UNIVERSITY of ABERDEEN

The screenshot shows the Protege 3.0 beta interface. On the left is the 'CLASS BROWSER' with a tree view of OWL classes. The 'INSTANCE BROWSER' in the middle shows a list of instances for the selected class. The 'INDIVIDUAL EDITOR' on the right shows the details for the instance 'SI-II-c-HY-env1-set2.2', including its name, description, and various properties like 'HASREPORTINGYEAR' and 'HASRESOLUTION'.

Tools 1: Protege

The screenshot shows the Longwell repository interface. It features a header with logos for FEARLUS-G, UNIVERSITY of ABERDEEN, and THE MACAULAY INSTITUTE. The main content area displays a list of hypotheses, each with a title, type, status, and a link to 'Add to My Workspace'. On the right side, there is a sidebar with 'Options', 'Restriction' (set to 'limit to 24 results'), and search tools including 'free text search' and 'facet search' (sorted by frequency).

Tools 2: Longwell

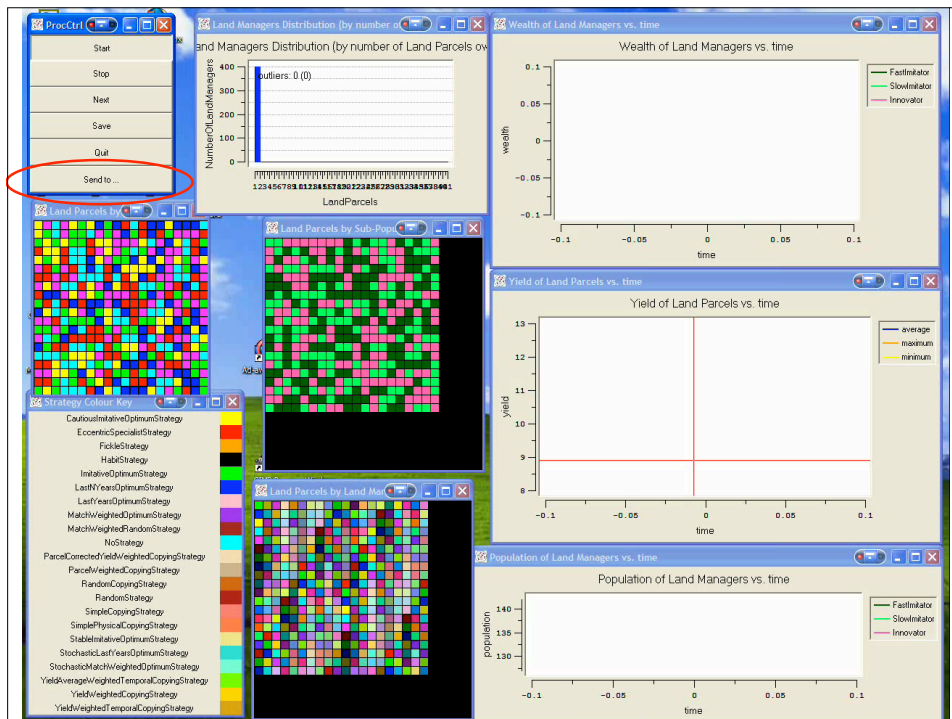
Aims of the Fearlus-G project

To serve FEARLUS, an existing environmental modelling framework, to the scientific community.

- Allow very large-scale experiments to be run, analysed, and repeated.

To promote collaboration by facilitating access to alternative models and comparison of results.

To support training by providing a shared co-laboratory environment for experimentation.





Pilot Semantic Grid Service for Environmental Modelling



[Home](#) [Upload](#) [Search](#) [Experiment](#) [Log Out](#)

Welcome to the FEARLUS-G Web Client

Models Parameters uploaded from: ed

Model Parameters File Name	Date	Time	Description	Status	Options
Fi-v-Si-v-In.model	2004-07-29	16:23:57	Innovator vs. Imitators	Completed	Run View Delete Publish Clone Send
P0-E16u-BET8-H8RvH12R.model	2004-08-20	15:12:57	Innovators vs Imitators	Running...	Run View Delete Publish Clone Send <ul style="list-style-type: none"> ▶ Simulation instance 0 running... ▶ Simulation instance 1 running... ▶ Simulation instance 2 running... ▶ Simulation instance 3 running... ▶ Simulation instance 4 running... ▶ Simulation instance 5 running... ▶ Simulation instance 6 running... ▶ Simulation instance 7 running... ▶ Simulation instance 8 running... ▶ Simulation instance 9 running... ▶ Simulation instance 10 running... ▶ Simulation instance 11 running... ▶ Simulation instance 12 running... ▶ Simulation instance 13 running... ▶ Simulation instance 14 running...

FEARLUS-G

My Workspace



My Projects

[Imitative Versus Non-Imitative Strategies](#)

[+ New Project](#)

My Experiments

Experiments

- | | |
|---|--|
| <input checked="" type="checkbox"/> SI-v-HRYI-c-II-env1-set2.2 | <input checked="" type="checkbox"/> SI-v-II-c-HYI-env1-set2.2 |
| <input checked="" type="checkbox"/> SI-v-II-c-HRYI-env1-set2.2 | <input checked="" type="checkbox"/> II-v-HYI-c-SI-env1-set2.2 |
| <input checked="" type="checkbox"/> SI-v-HYI-c-II-env1-set2.2 | <input checked="" type="checkbox"/> II-v-HRYI-c-HYI-env1-set2.2 |
| <input checked="" type="checkbox"/> SI-v-HRYI-c-HYI-env1-set2.2 | <input checked="" type="checkbox"/> II-v-HRYI-c-SI-env1-set2.2 |
| <input checked="" type="checkbox"/> HYI-v-HRYI-c-SI-env1-set2.2 | <input checked="" type="checkbox"/> HRYI-v-HRYI-c-II-env1-set2.2 |
| <input checked="" type="checkbox"/> II-v-HYI-c-HRYI-env1-set2.2 | <input checked="" type="checkbox"/> SI-v-HYI-c-HRYI-env1-set2.2 |

Experiment Sets

- Experiment2.2-Polhill-et-al-2001

[+ New Experiment](#) [+ New Experiment Set](#)

My Simulation Parameters

Models

- modelo-6-6unix

Environments

- Environment3 Environment2

Subpopulations

- | | | |
|--|--|---|
| <input checked="" type="checkbox"/> OD | <input checked="" type="checkbox"/> OS | <input checked="" type="checkbox"/> HYI |
| <input checked="" type="checkbox"/> ES | <input checked="" type="checkbox"/> II | <input checked="" type="checkbox"/> SI |
| <input checked="" type="checkbox"/> LS | <input checked="" type="checkbox"/> HRYI | |

[+ New Model](#)

Options

My Workspace
 My Repository
[Log Out](#)

FEARLUS-G Options

[Upload a model form file](#)
[Search a model in MyWorkspace](#)

Experiments

Type 1 Experiment

Experiment Set

Experiment2.2-Polhill-etal-2001

Label: Experiment2.2-Polhill-etal-2001

Description: Experiment2.2-Polhill-etal-2001

Comment:

Number of Runs: 60

Significance Level: 0.0010

Single Subject Hypotheses: true

Experiments:

- II-v-HRYI-c-SI-env1-set2.2
- SI-v-HYI-c-II-env1-set2.2
- II-v-HYI-c-HRYI-env1-set2.2
- HYI-v-HRYI-c-SI-env1-set2.2
- SI-v-HRYI-c-II-env1-set2.2
- SI-v-HYI-c-HRYI-env1-set2.2
- II-v-HRYI-c-HYI-env1-set2.2
- SI-v-HRYI-c-HYI-env1-set2.2
- II-v-HYI-c-SI-env1-set2.2
- SI-v-II-c-HYI-env1-set2.2
- SI-v-II-c-HRYI-env1-set2.2
- HYI-v-HRYI-c-II-env1-set2.2


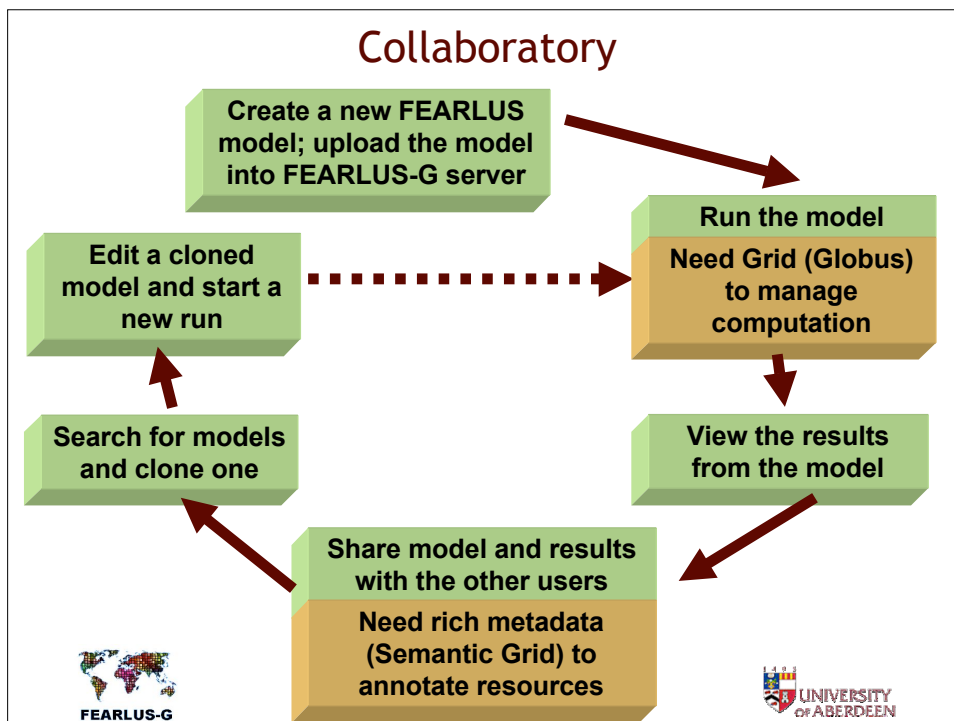
Uses Model: modelo-6-6unix

Termination Year: year201Termination

Subjects: HRYI, HYI, SI, II, Environment1

Hypothesis Type: fearlusType2Hypothesis

Tailed Hypothesis Class: OneTailedHypothesis

Some challenges

Creating & evaluating a usable toolset & methodology.

Integration & interoperability.

Curation: managing provenance & information quality.

Automating e-Science?



Links & credits

See also

- www.csd.abdn.ac.uk/research/fearg
- www.aktors.org

Fearlus-G people

- Pete Edwards (Aberdeen)
- Edoardo Pignotti (Aberdeen)
- Nick Gotts (Macaulay Institute)
- Gary Polhill (Macaulay Institute)



... any questions?

