Knowledge Management in Biology

Phillip Lord
School of Computing Science,
University of Newcastle
(formerly University of Manchester)

What is knowledge management

Knowledge management processes should enable the capture, generation, manipulation, storage, search and dissemination of information.

myGrid and ComparaGRID

- Access to distributed information resources for end users
- myGrid aimed more at “open” access
- ComparaGRID more at data integration.
- myGrid has built a highly usable workflow engine, taverna, providing access to 1000+ services.

View from Biology

- Biology is a knowledge rich discipline.
- The knowledge is highly heterogeneous, hard to structure and, often, incomplete.
- Data models are often developed independently, few standards exist.

Provenance vs “Curation”

- Curation—adding explicatory metadata to the raw information—is highly valued.
- Much of the data has poorly documented origins.
- Older versions rarely preserved.
- myGrid’s workflow approach enables (some) automated generation of provenance information.

Potential Gains from Knowledge Management

- Should increase scalability, reduce error rates, lessen resource requirements
- Enable search and analysis, cross-cutting across different resources, from different providers
- Primarily interested at improving the experimental process
- Essential to enable move from whole genome toward whole systems biology

Constraints

- Coping with legacy data
- Coping with increasing scale
- Coping with change
- Fitting in with existing working practices