NETTAB'2002

On the Use of Agents in a Bioinformatics Grid

Luc Moreau,
University of Southampton

Structure

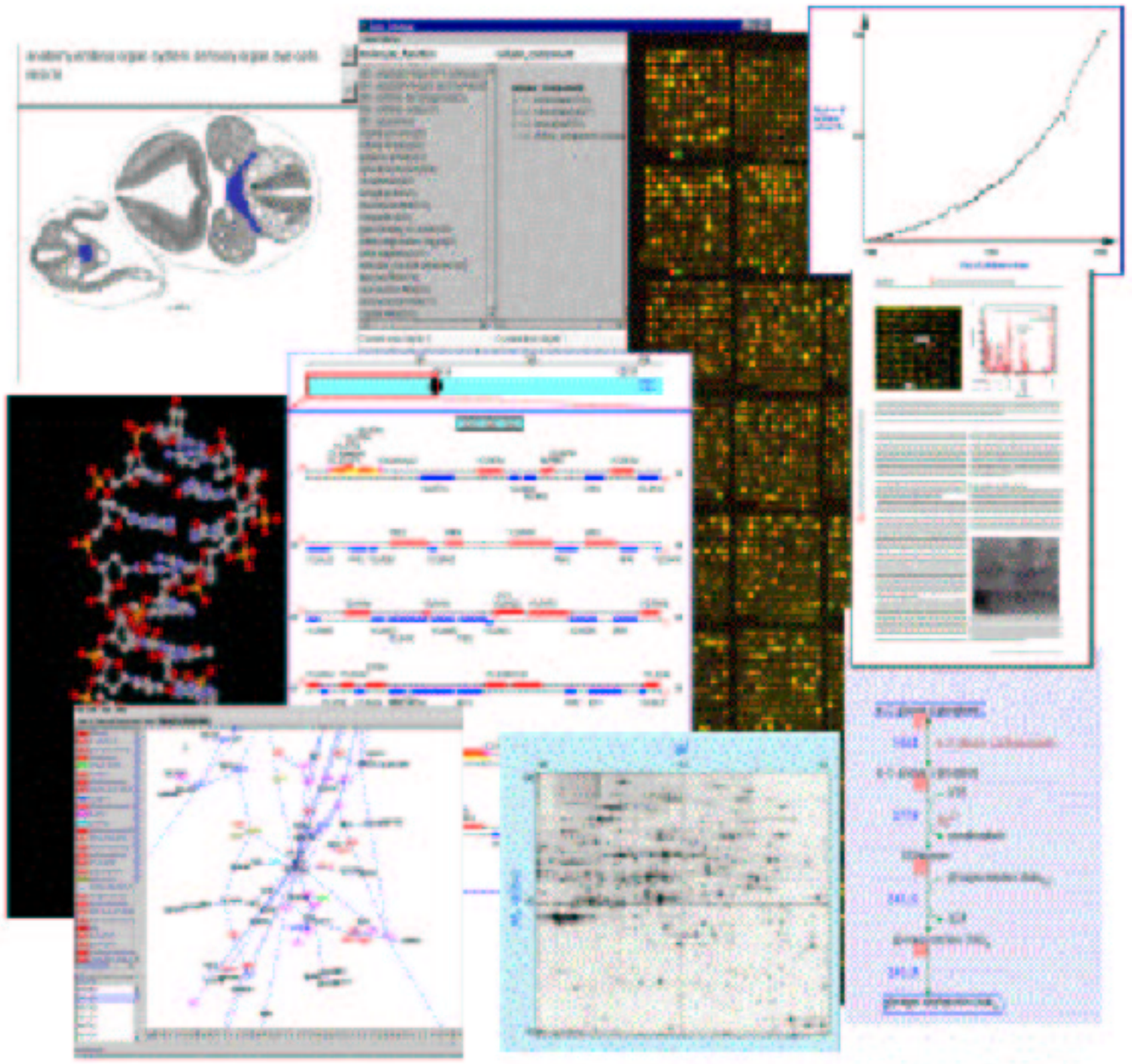
- Background: myGrid
- Architecture
- Use of Agents
- Conclusion

e-Science & Biology

- Biology is a multi-faceted & increasingly multi-disciplinary science
- Bioinformatics is very much an "e-Science"
  - Discovery is done *in silico* on results obtained from experiments using a number of analysis & data resources
- Molecular biology & genomics are the particular focus here

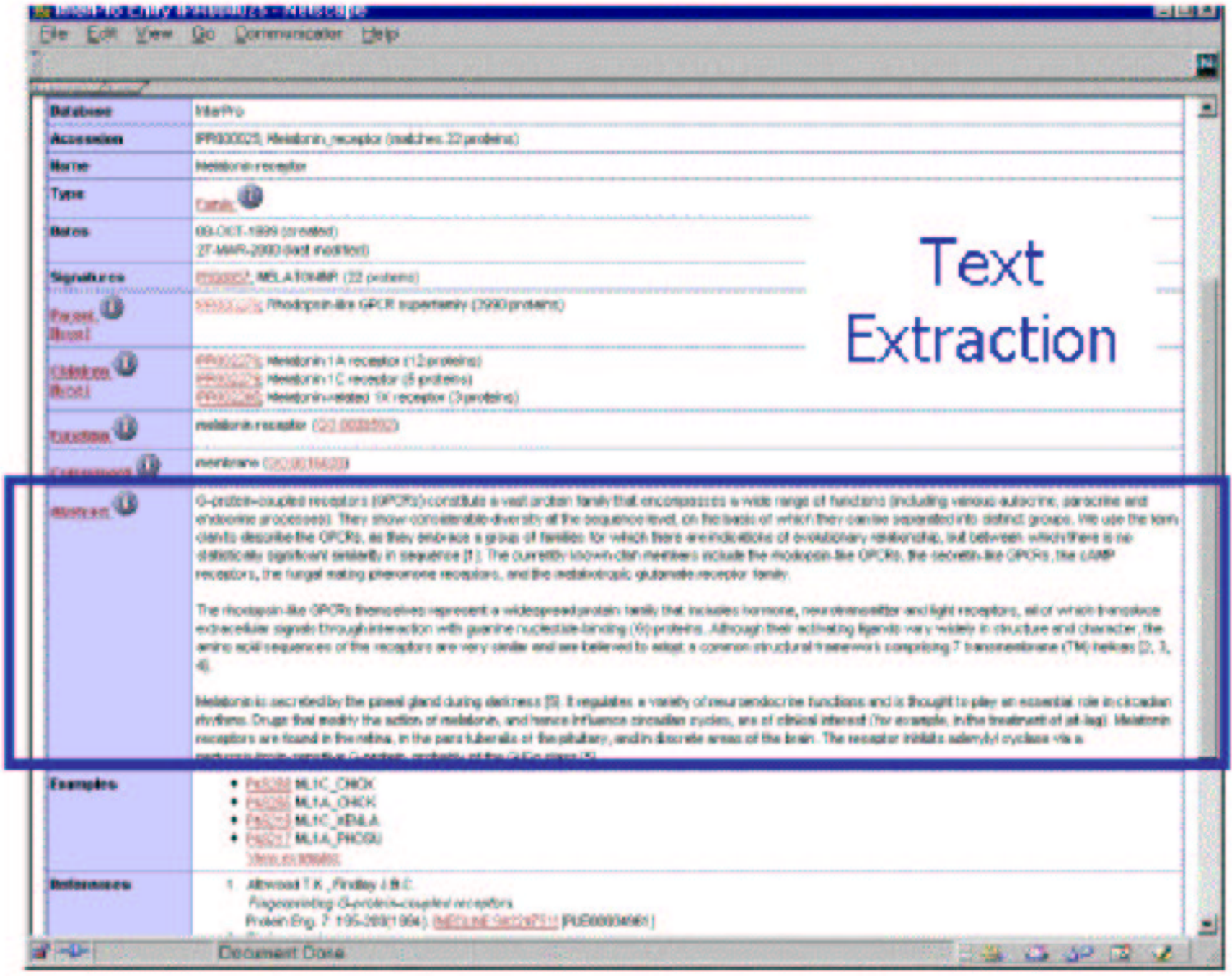
Bioinformatics & Genomics

- Large amounts of data
- Highly heterogeneous
  - Data types
  - Data forms
  - Community
- Highly complex and inter-related
- Volatile



Bioinformatics Data

- Descriptive as well as numeric
- Literature
- Analogy/ knowledge -based



Bioinformatics Analysis

- Different algorithms
  - BLAST, FASTA, pSW
- Different implementations
  - WU-BLAST, NCBI-BLAST
- Different service providers
  - NCBI, EBI, DDBJ

In silico experimentation

- Discovery, interoperation, fusion, sharing
- **Process** is as important as outcome
- Science is dynamic – **change** happens
- Scientific discovery is **personal** & global
- **Provenance** and history

myGrid: facts

- EPSRC funded pilot project
- Generic middleware with application setting
- 36 month period
- Started on 1st October 2001
  - 16 full-time post docs altogether
  - 1 technical project manager
  - 1 system manager
  - 1 secretarial post

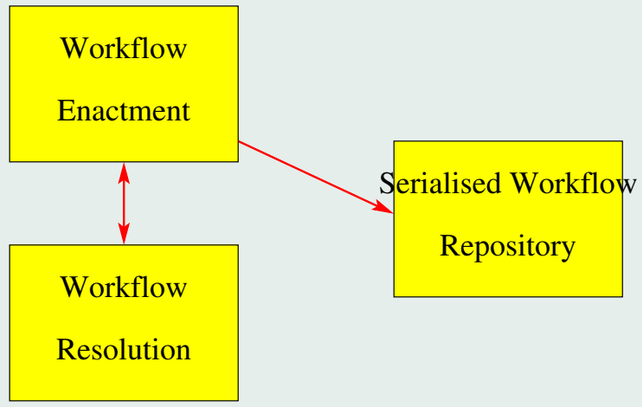
myGrid: facts

- Scientific team
  - Biologists and Bioinformaticians
  - GSK, AZ, Merck KGaA, Manchester, EBI
- Technical Team
  - Manchester, Southampton, Newcastle, Sheffield, EBI, Notthingham
  - IBM, SUN
  - GeneticXchange
  - Network Inference, Epistemics Ltd

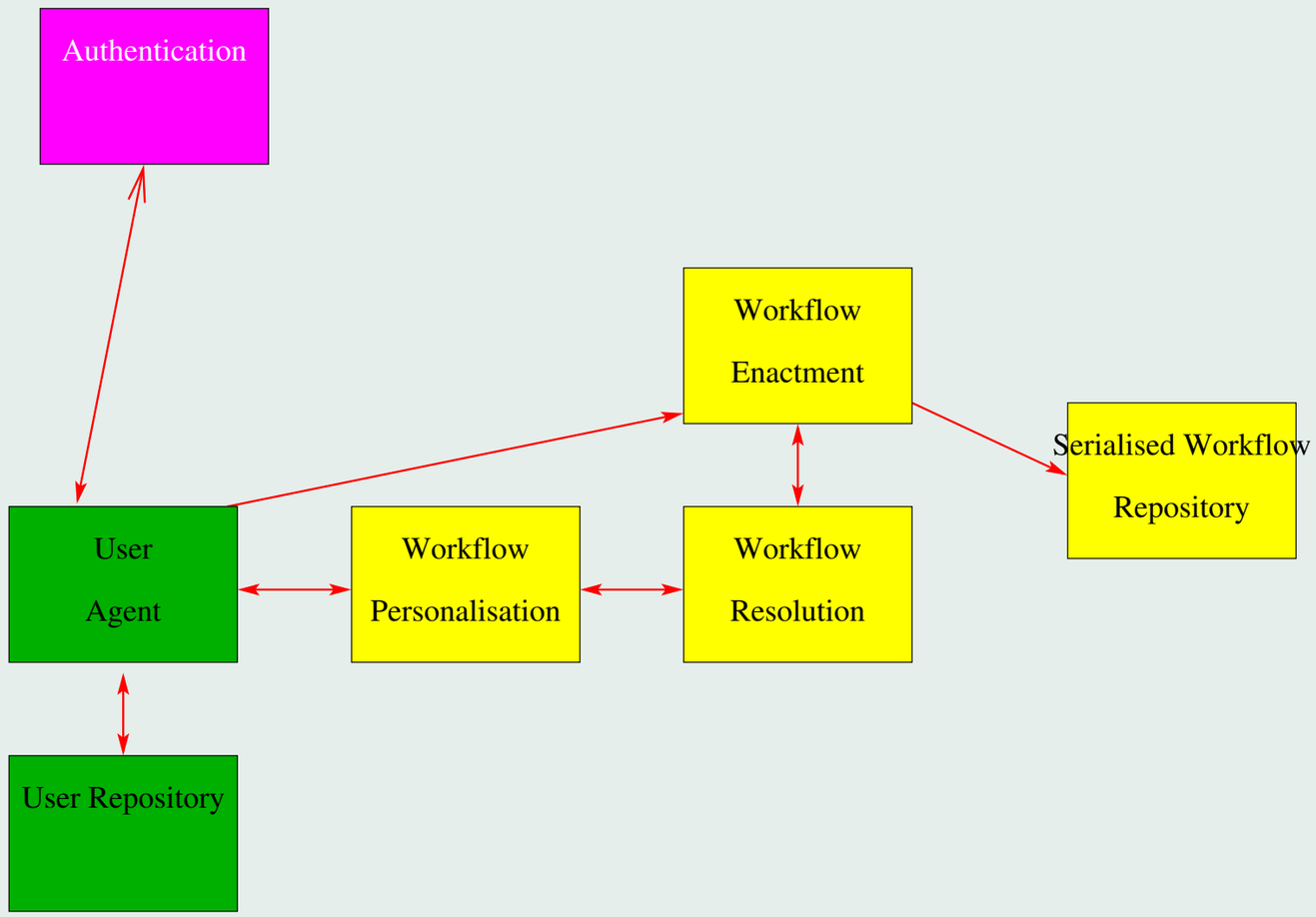
myGrid Outcomes

- e-Scientists
  - Gene function expression analysis using S. cerevisiae
  - Annotation workbench for the PRINTS pattern database
- Developers
  - myGrid-in-a-Box developers kit
  - Re-purposing DAS, AppLab and OpenBSA
  - Integrating ISYS & GlaxoSmithKline platforms

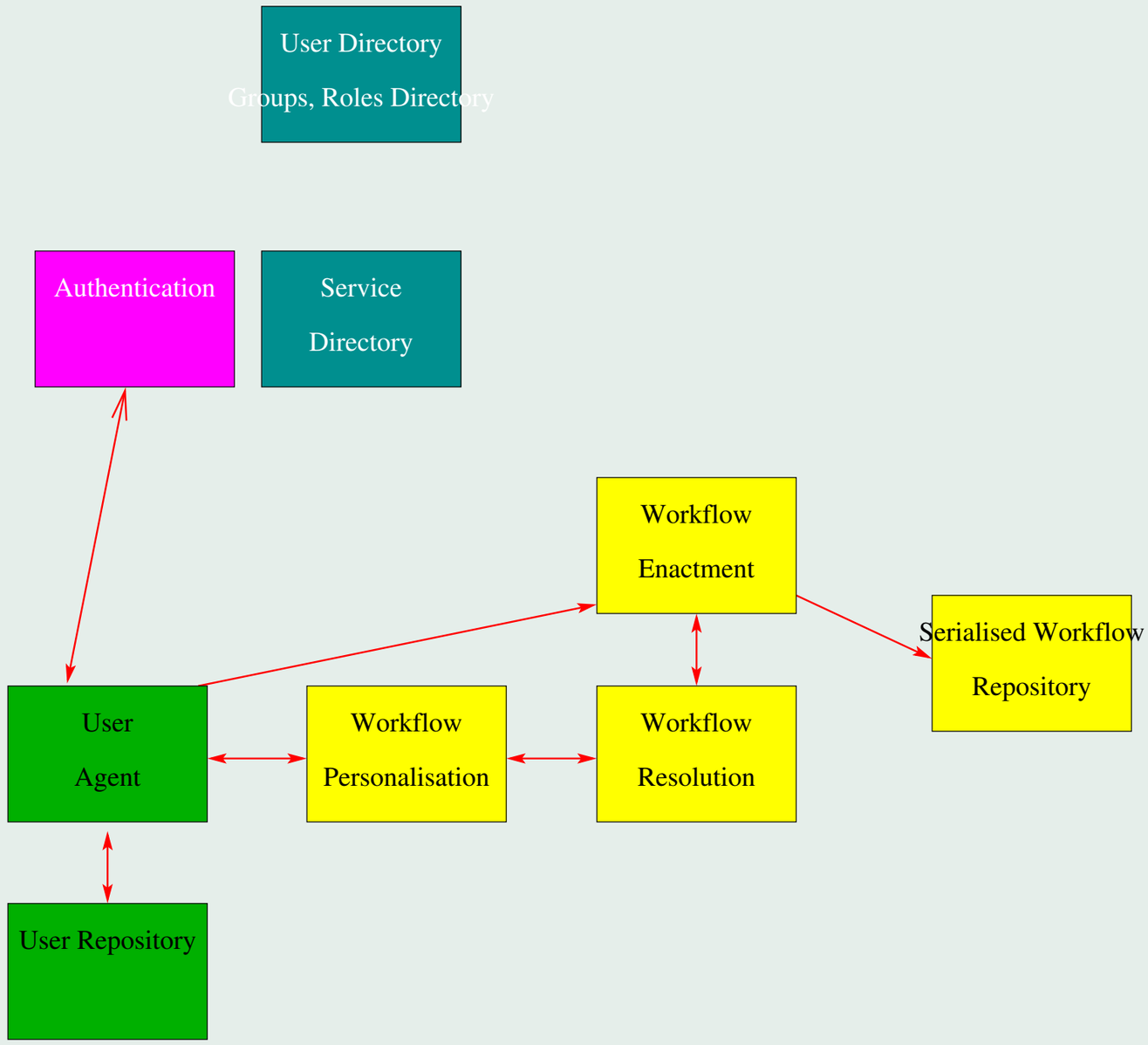
Architecture



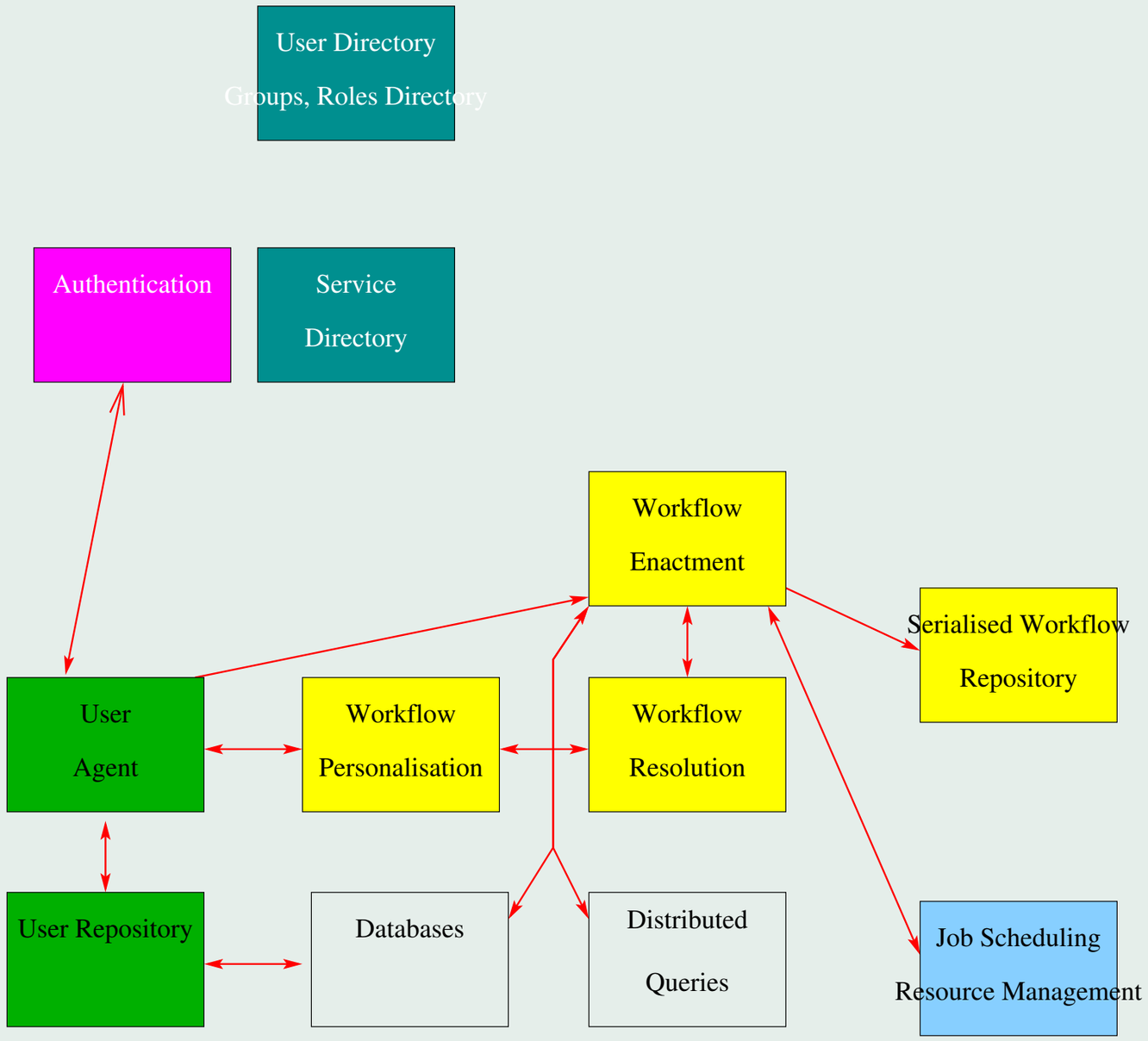
Architecture



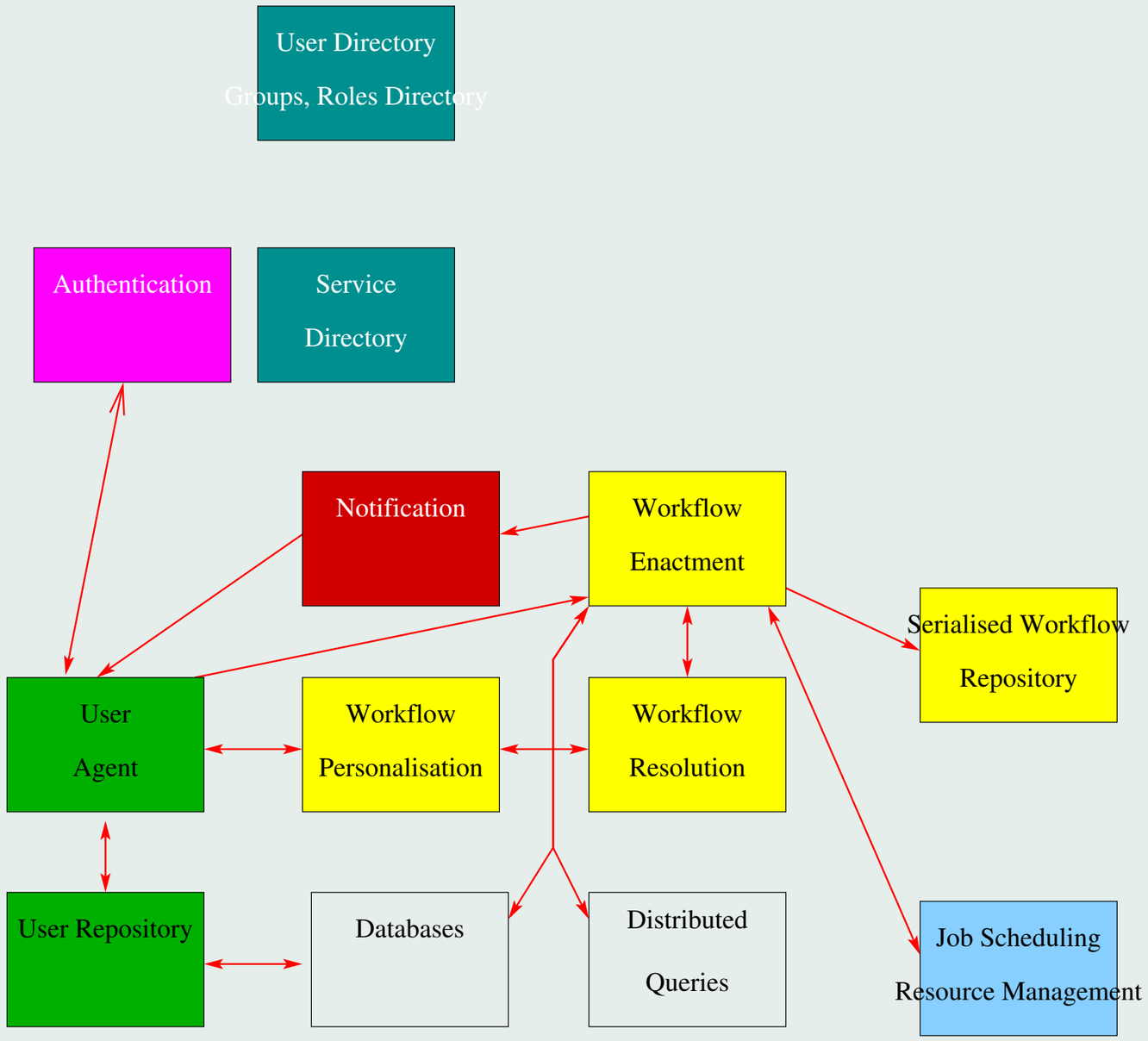
Architecture



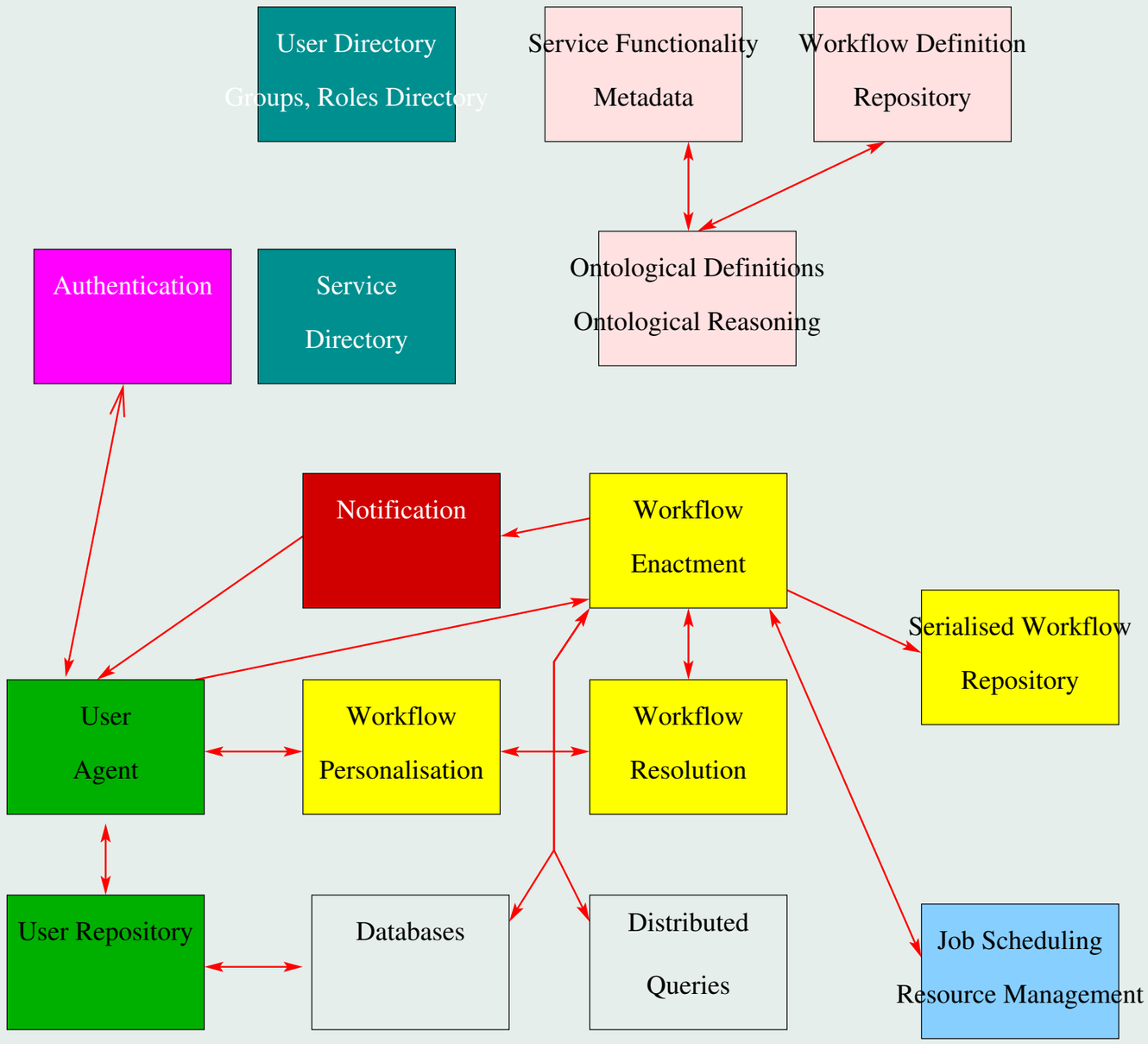
Architecture



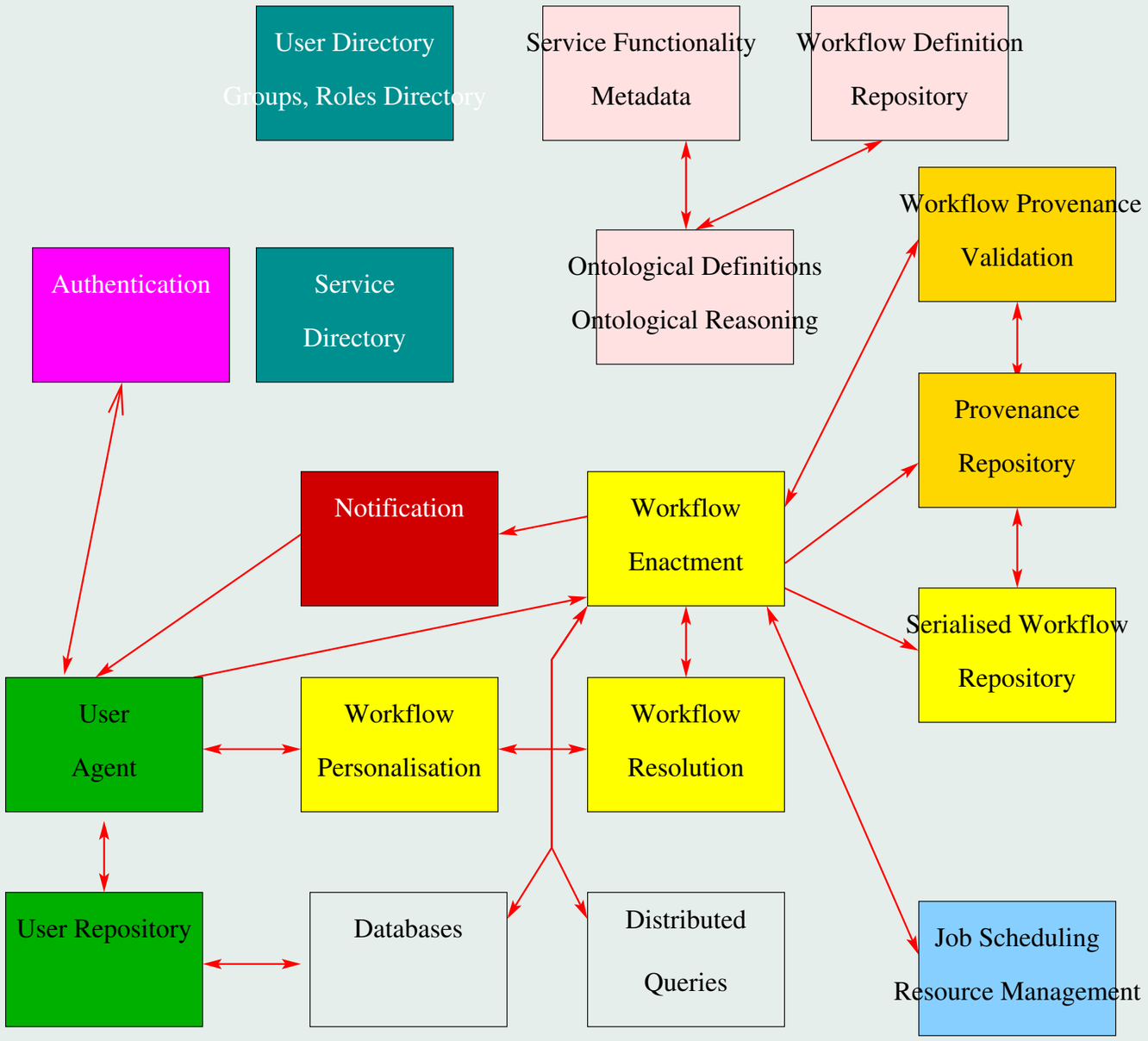
Architecture



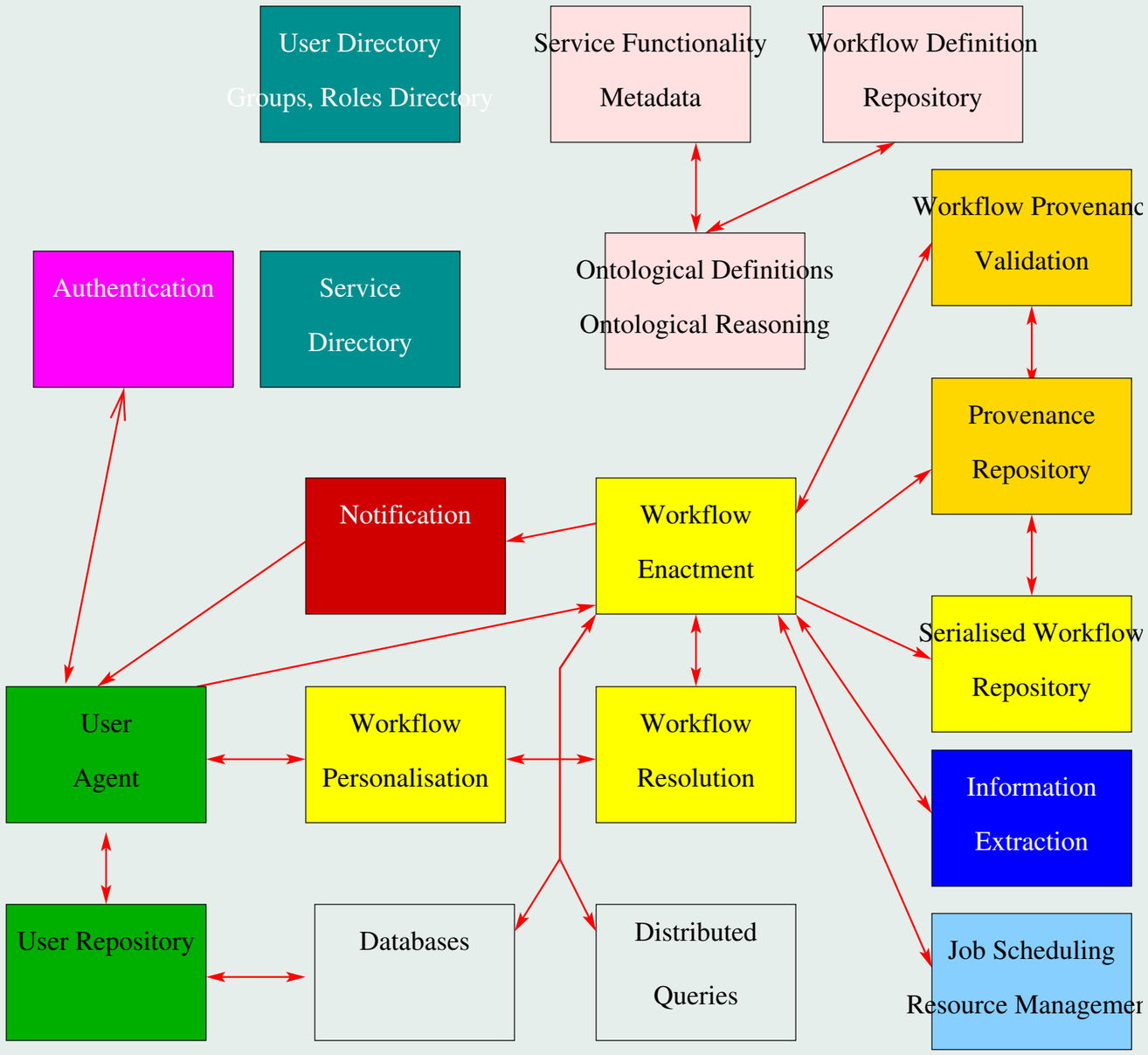
Architecture



Architecture



Architecture



Agents in a Bioinformatics Grid

The bioinformatics domain is characterised by rapid and substantial change over time.

- Change in the resources available to the bioscientist poses problems: new resources can appear, old ones can disappear, and some can simply change.

- Limiting a system to a fixed set of resources (e.g. well-known and highly regarded databases) could impose undesirable constraints.

- Thus, any system intended for application to the bioinformatics domain should be able to cope with this dynamism and openness.

Agents in a Bioinformatics Grid

Agents are the technology to cope with dynamism and openness.

- Agents are flexible, autonomous components designed to undertake overarching strategic goals, while at the same time being able to respond to the uncertainty inherent in the environment.

- Agents provide an appropriate paradigm or abstraction for the design of scalable systems aimed at this kind of problem.

Agents in a Bioinformatics Grid

- So, again another definition of agents?
- No! The focus is not on the definition of what qualifies or not an agent.
- The focus is on techniques that are typical of the agent field.

Agents in a Bioinformatics Grid

Agent-based computing offers useful techniques that can be used in Grid systems, including:

- personalisation,
- communication,
- negotiation.

Agents in a Bioinformatics Grid

Agent-based computing offers useful techniques that can be used in Grid systems, including:

- personalisation,
- communication,
- negotiation.

User Agent

The user agent is an agent in the sense that:

- it represents a user within the myGrid system;
- it can be seen as a personal agent [Maes CACM94];
- Three roles:
  - personalisation
  - contact point
  - “bag man”

User Agent: personalisation

When a workflow is being enacted and a choice of services becomes available:

- the choice should not be made arbitrarily,
- the choice should be based on the priorities and circumstances of the particular user,
- e.g., a user may have greater trust in the ability of one service to produce accurate results than another.

User Agent: personalisation

- The user should not have to be queried each time a service must be chosen;
- these preferences and previous choices can be recorded and acted upon by the user agent to select from each set of options presented to it;

- We call this function personalisation.

User Agent: contact point

The user agent is also a contact point between services within myGrid and the user.

- The user agent is an intermediary able to receive, e.g.:
  - requests from services for the user to enter data or
  - notifications about changes to remote databases.

- These messages can then be forwarded to the user only when the user is able and willing to receive them.

User Agent: the “bag man”

The user can delegate tasks to the user agent,

- such as authenticating itself with a service before use,
- for personalisation of workflows,
- and in general for any tedious and repetitive task.

Agents in a Bioinformatics Grid

Agent-based computing offers useful techniques that can be used in Grid systems, including:

- personalisation,
- communication,
- negotiation.

Agent Communication Language (ACL)

A key requirement of myGrid is the design of a future proof environment in which collaborative distributed bioinformatics applications may be developed.

Bioinformatics is not a green field, and multiple protocols and standards are already supported by the community.

Our methodology is to design a generic architecture able to support multiple existing protocols, languages and standards, and which hopefully will be able to accommodate future developments.

In particular, we want to design an abstract communication architecture that we can map onto concrete communication technologies.

ACL: Web Services

In the eBusiness community, Web Services have emerged as a set of open standards, defined by the World Wide Web consortium, and ubiquitously supported by IT suppliers and users.

Web Services rely on XML, SOAP, WSDL, and UDDI.

Web Services look very appealing for Grid Computing:

- Open Grid Service Architecture (OGSA) which extends Web Services with support for the dynamic life-cycle management of Grid Services. [Foster Kesselman 02]

ACL: Speech Acts

In agent systems, it is common practice to:

- separate intention from content in communicative acts,
- abstract and classify communicative acts according to Searle's speech act theory.

An agent's communications are structured and classified according to a predefined set of "message categorisations", usually referred to as performatives.

That is what we call an "agent communication language" (cf. KQML and FIPA ACL).

ACL: Communication Model

In SoFAR, the Southampton Framework for Agent Research, we have adapted a key concept of the Nexus communication layer [Foster, Kesselman, Tuecke JPDC96] to the world of agents.

Communications between agents take place over a virtual communication link, identified by a startpoint and an endpoint.

An endpoint identifies an agent's ability to receive messages using a specific communication protocol. An endpoint extracts messages from the communication link and passes them onto the agent.

A startpoint is the other end of the communication link, from which messages get sent to an endpoint. Given a startpoint, one can communicate with a remote agent, by activating a performative on the startpoint, passing the message content.

ACL: Agents as Web Services

In [Moreau CCGRID2002, Avila Moreau AgentCities2002], we show how

- agent communication language , and

- startpoint/endpoint communication model

can be mapped onto the communication stack of Web Services.

ACL: Agents as Web Services

Two phases:

1. ACL performatics and message contents can be encoded in SOAP [Moreau:CCGRID2002]
2. Agents can be described using WSDL, registered and discovered in UDDI. [Avila Moreau AgentCities2002]

Hence, agents can be seen and reused any as Web Services.

ACL: Agents as Web Services

- Promising approach, as declarative communication semantics promotes
  - inter-operability,
  - openness,
  - dynamic discovery and reuse of agents.

- Opens the agent world to the Web Services community,

  to enable more complex interactions.

Agents in a Bioinformatics Grid

Agent-based computing offers useful techniques that can be used in Grid systems, including:

- personalisation,
- communication,
- negotiation.

Negotiation Broker

Another application of research from the agent field is in the area of negotiation.

Service users and service providers have differing criteria over the preferable quality and content of the service.

Negotiation Broker: Notification Service

In myGrid, negotiation is particularly useful for notification support.

- Service providers may want to send out notifications concerning improvements to tools, changes to databases or updates concerning the state of enacted workflows, etc.

- Users, services, or agents want to register to receive some subset of these notifications.

- A notification service supports asynchronous messaging and manages message distribution.

Negotiation Broker: Quality of Service

The subjects over which negotiation takes place include the following forms of quality of service.

- The cost of receiving the notification,
- the topic (event category) of the notifications,
- the frequency with which notifications are received, e.g. every time a change occurs, daily, hourly,
- the generality of the change described by the notifications

- the form in which the information in the notification message is supplied,

- the accuracy of information contained within a notification

Here, quality of service refers:

- what a publisher produces, and

- how a publisher produces it.

- A publisher of notifications will be able to produce notifications matching one or more measures of quality of service.

  e.g. a publisher may be able to publish notifications on a particular topic every minute or every hour.

- A consumer of notifications may prefer one measure of quality of service over another.

If demands cannot be met exactly:

- the consumer may choose to negotiate with the publisher to find the next best quality of service that the publisher can provide.

For example,

- the consumer desires notifications weekly,
- the publisher can provide daily or fortnightly notifications,
- the subscriber must find this out from the publisher and then decide between them, or decide not to subscribe at all, based on the subscribers particular priorities.

The notification service must provide notification support for a potentially large and varying number of consumers.

The notification service should have some control over the quality of service agreed upon.

The notification service should limit the interaction between the publisher and consumer so that the knowledge of one by the other is limiting for reasons of privacy.

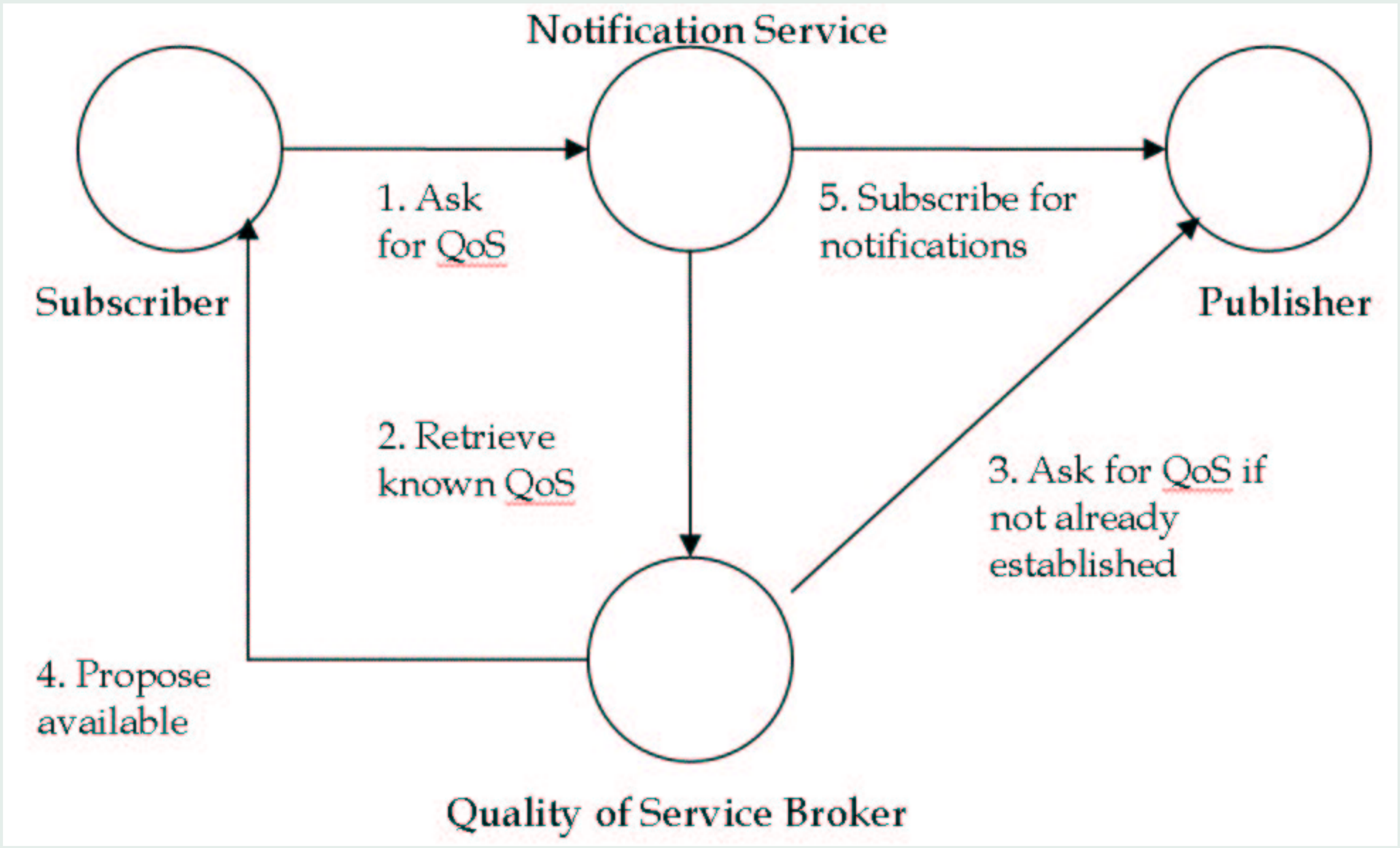
The quality of service broker:

- is an agent conceptually contained within the notification service;

- negotiates on behalf of each consumer wishing to receive notifications of a specified quality, then provides a final proposal to the consumer;

- can negotiate with any of the publishers known to the notification service;

- limit the agreed quality of service to that acceptable to the notification service.


Notification Service
Subscriber
1. Ask
for QoS
5. Subscribe for
notifications
Publisher
2. Retrieve
known QoS
3. Ask for QoS if
not already
established
4. Propose
available
Quality of Service Broker

- we use the concept of pluggable negotiation algorithms, allowing the quality of service broker to select the appropriate protocol for negotiating with a publisher.

Conclusion

- We have presented the myGrid architecture and overview possible use of agents.

- MyGrid aims to provide a personalised environment for the bioscientists, which helps them to automate, repeat and therefore better achieve their experiments.

- Agents are particularly useful in tailoring the myGrid system to the priorities of individual scientists, personalising each step of a workflow and negotiating on their behalf.

Conclusion

It can be seen that:

- with dynamic workflow enactment,
- standardisation of data semantics via ontologies, and
- the many other facilities of myGrid,

agents can make conducting in silico experiments flexible and more easily controlled by the individual or collaborating scientists.

Conclusion

The examples of use of agency we have presented:

- already offer a capability inexistent in current bioinformatics environment,
- still remain rather localised to some specific services (user agent or negotiation over quality of service of notification service), or components such as a communication layer.

Conclusion

Longer term agent-based computing techniques enable individual components to collaborate or to compete with others in the provision of services. For example,

- virtual organisations, where different services come together in some coherent subsystem for a particular purpose

- regulation of open societies of services through the use of norms and electronic institutions.

Use of sophisticated auction mechanisms, or electronic marketplaces, for obtaining the best resources at the least cost to the user.

Final Conclusion

Agent-based computing

is

the technology

to program Grid systems.

www.mygrid.org.uk
www.aboutmygrid.org

AstraZeneca
GlaxoSmithKline
MERCK
Sun microsystems
IBM
EPISTEMICS LTD
knowledge is our business
network inference

EMBL
European Bioinformatics Institute
THE UNIVERSITY of MANCHESTER
University of Southampton
UNIVERSITY OF NEWCASTLE UPON TYNE