



e-Social Science and the NGS

The GENESIS project

Dr Paul Townend
Distributed Systems and Services Group
University of Leeds

p.m.townend@leeds.ac.uk

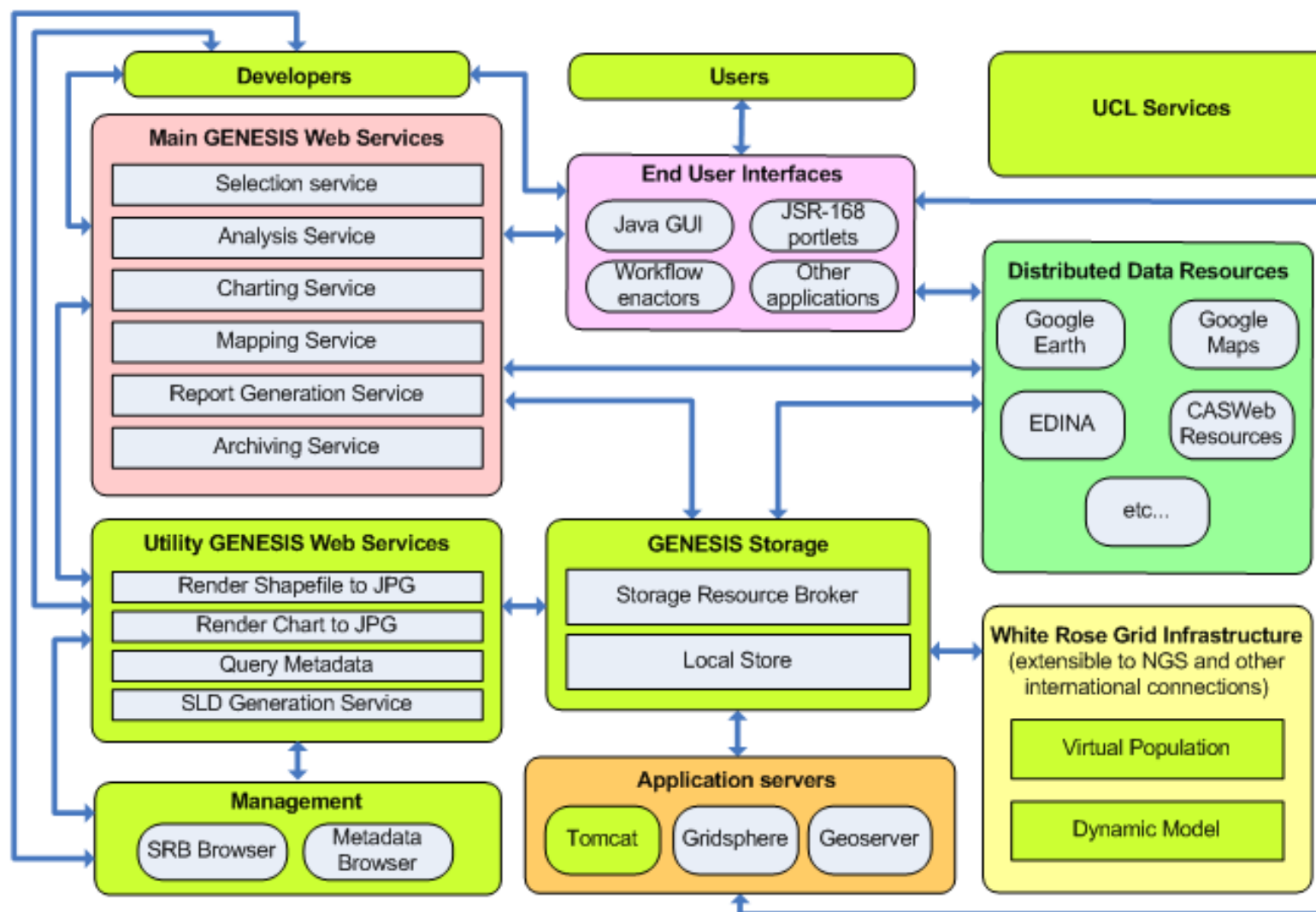


- The **social sciences** are the fields of academic scholarship which explore aspects of human society.
- ***e-Social Science*** is a term which encompasses technological developments and approaches within Social Science. The 'e' can mean '**enabling**'.
- Social Scientists and Computer Scientists are working on tools and research which Social Scientists can take and use to help their research.
- NCESS (www.ncess.ac.uk) is the hub for such activities.



Generative e-Social Science

UNIVERSITY OF LEEDS





Computing challenges for GENESIS

- **Large computational requirements:** population initialisation/reconstruction performed on 30-node Beowulf cluster, and took approximately 2 months for one run.
- **Many data considerations:** size of data, archives, distributed and heterogeneous datasets, security, etc.
- These issues can be directly helped by the **National Grid Service**.



Using NGS compute facilities (1)

- GENESIS population reconstruction was performed on a 30 node Beowulf cluster:
 - Jobs (different segments of a population) were sent to each node in the cluster and processing was performed (a Genetic Algorithm)
 - Each job listened on a specific port, and communicated with a “Master thread”.
- The more nodes we use, the faster we can get final results. We turned to the NGS to help us port our system.



Using NGS compute facilities (2)

- We developed a wrapper script to dynamically ascertain the available machines and ports that our jobs were sent to on the NGS.
- This information is then used by the Master Thread to communicate with the worker threads.
- The result is that instead of using 30 CPUs, we can now use many times that number.
- Recently we used 128 CPUs to reduce the time for one population reconstruction from 2 months to 2 weeks.



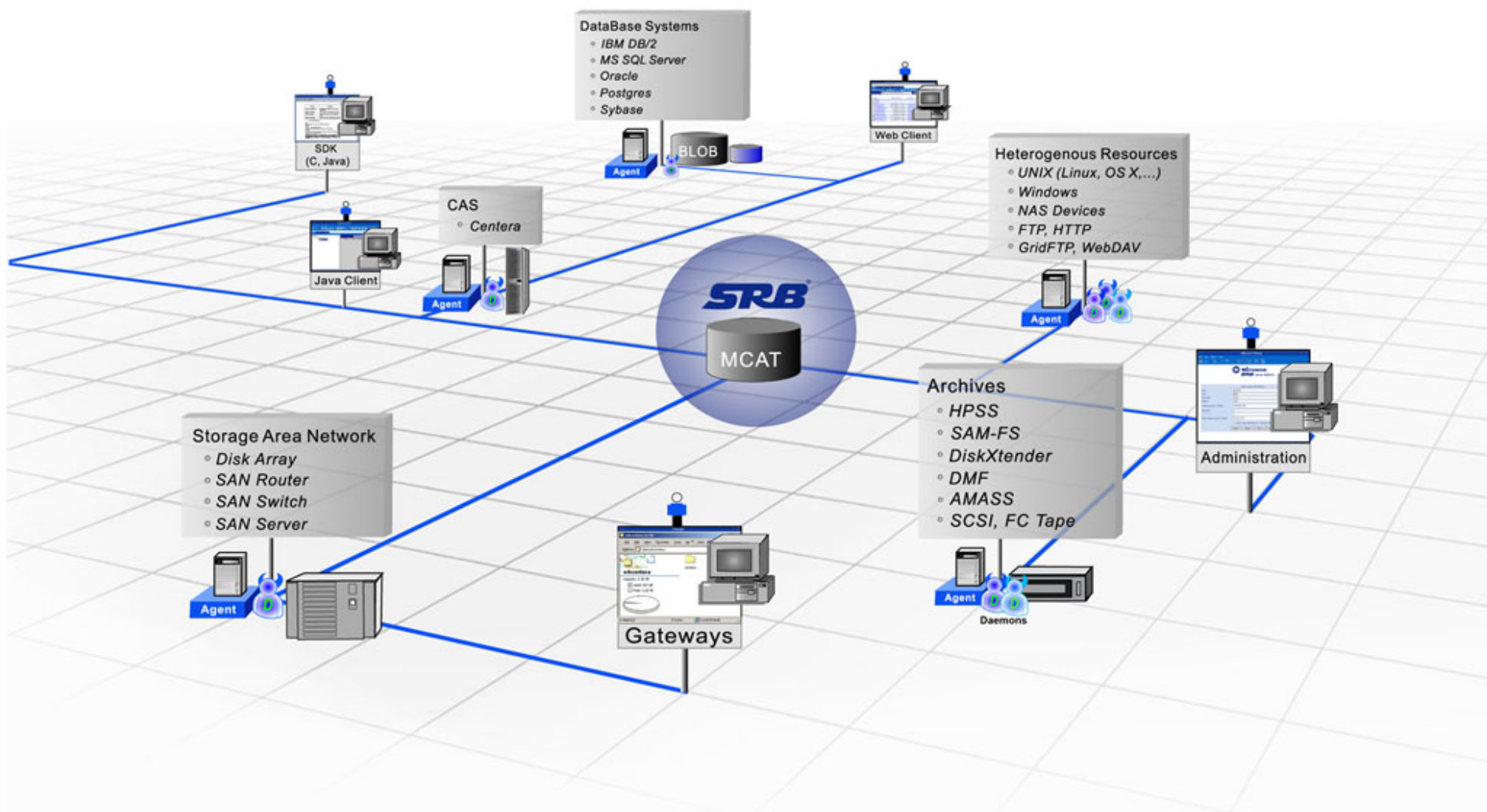
Using NGS data facilities (1)

- GENESIS has many data requirements, as previously discussed.
- To introduce large, distributed storage capacity, we use the NGS Storage Resource Broker (SRB) service.
- SRB is a Data Grid Management System DGMS.
- It is a logical distributed file system based on a client-server architecture, which presents users with a single global logical namespace or file hierarchy.



Using NGS data facilities (2)

UNIVERSITY OF LEEDS





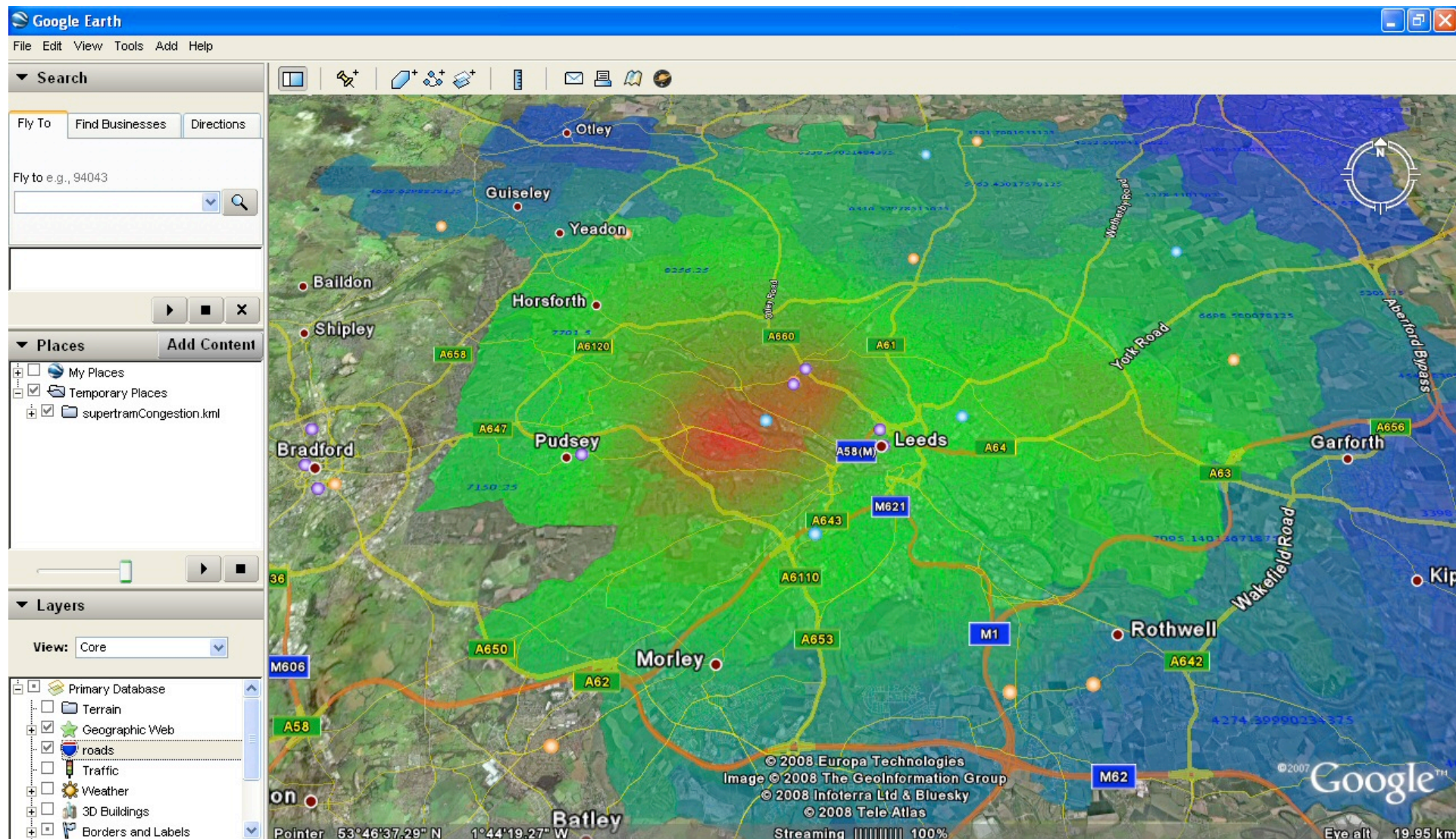
Using NGS data facilities (3)

- Accessing the NGS SRB service has been very straightforward. We use the JARGON Java SRB API in our tools.
- This gives us large amounts of storage space, stored across NGS nodes. We can use SRB security to control access to sensitive datasets.



GENESIS results

UNIVERSITY OF LEEDS



Questions are welcome



UNIVERSITY OF LEEDS

