Over recent years it has become clear that preserving and accessing digital data is increasingly important across a wide range of scientific, artistic and cultural activities. There has been a growing recognition of the need to address the fragility and accessibility of the digital information collected in all aspects of our lives. Access to digital information lies at the heart of the scientific and technical innovation vital for modern economies. Supporting Society’s growing dependence on digital information for its smooth operation provides a real urgency for this task and a wide range of initiatives are already underway at both the national and international level to tackle the many aspects of the end-to-end digital preservation "lifecycle".

Within the UK “curation and preservation” of digital information has been identified by the OST working group on e-Infrastructure as one of six key components the future national e-Infrastructure. The OST working group is charged with mapping out relevant developments, gaps and challenges in digital curation and preservation over the next 10 years. Therefore, it is timely to bring together national and international experts across the full spectrum of the digital lifecycle to assist this process by mapping out the current state of play and future agenda and provide valuable input to the working group. A particular focus of this workshop is the research agenda for digital curation and preservation; what are the major challenges and gaps in current developments, and what needs to be done to tackle them?

Although this is a UK workshop, it is not intended to focus only on UK. International integration is essential and the workshop will include key international experts. The outputs from the workshop will be made available to the wider community and policy bodies, such as the European e-Infrastructure Reflection Group (http://www.e-irq.org/) and the European Strategic Forum on Research Infrastructures (http://www.cordis.lu/era/esfri_home.htm) which, in parallel to the OST working group, are charged with mapping out the requirements for an integrated infrastructure across Europe.

The workshop will consider digital curation and preservation from three perspectives:

**Data Lifecycle Management**

Electronic research is potentially fragile and at risk unless carefully curated. It needs careful management from the moment of its creation and a pro-active policy and strategic approach to its creation and management to secure access, use, and preservation over the longer-term. The cost structure for securing the digital outputs of research also needs to be built in at the beginning if these costs are to be minimised and research investment effectively applied.

Different (and often differently interested) stakeholders become involved with data resources at different stages so there will be many stakeholders and interests in the digital outputs from research over a period of time. Few organisations or individuals have influence over (or even interest in) those resources throughout their entire life. However, the stages of the scholarly knowledge lifecycle, from creation to use, and reuse, are beginning to be actively explored and interlinked in innovative and exciting ways.

With these factors in mind, a life-cycle approach to the integration of research outputs can be of great value in understanding roles and responsibilities, estimating costs, and ensuring effective preservation and curation over time.
Drivers and Barriers

After ten years or so of research into digital preservation, we understand quite a lot about how to prepare digital information for preservation over the long term, and (in today's terms at least) how to manage that preservation process. In particular cases we know a lot about curating data in current use. We expect that digital curation and preservation will be expensive, and will require long term investment and commitment. However, we do not properly understand the costs and many other barriers inhibiting the acceptance of digital preservation as part of the normal operational responsibilities of “memory institutions”, scientific bodies, government and commercial organisations, etc. Some of these barriers are very strong, but they are often social and economic, not primarily technical. The culture of sharing and data re-use varies greatly between (and sometimes within) disciplines.

We also do not understand the down-stream value that can be accrued, or the present-day motivations that should pressure us to bring this about. One of these is the compliance imperative, broadly drawn (which could be described as responsible data management, with legislative backing). But value for the scholarly community has more subtle dimensions, again discipline-dependent.

Not until we understand these factors more clearly will we be able to design policies to drive the needed cultural change, and and make the business and investment cases for digital repositories.

Distributed architectures

Crucial to the preservation and curation of digital data is the storage and management of the data and metadata themselves in what will be, in general, a distributed environment. As the volume of digital information increases, the managing and accessing these data become non-trivial technical challenges in themselves. The integration and linking of different, in general distributed, information is also becoming increasingly important, including:

- linking of data to the information required for its preservation, such as its associated representation information and preservation description information;

- linking information at different stages of the digital lifecycle, e.g. citations and experimental data;

- use of geographically distributed, possibly replicated, data;

- interoperability between disciplines, for example correlating previously unrelated data sets, e.g. disease with genetic profiles
The session will focus on the distributed architecture needed to support the preservation and re-use of scientific and other research data and discuss the technical and scientific challenges that this scenario poses. Discussion of the future research agenda would be expected to cover:

- Preservation issues including OAIS infrastructure
- Certification of Trusted Repositories
- Persistent identifiers
- Provenance – authentication, authenticity, quality
- Interoperability – contemporaneous
- Annotations and Citations
- Mono- and multi-disciplinary use and re-use of digital information

As well as the longer term research agenda we will seek to identify the critical first steps required and key national and international developments.