

*P2P Networking 4 Science

A peer-to-peer (or P2P) computer network is based on the concept of pooling computing power and bandwidth of the participants in the network, eliminating the distinction between servers and clients. In contrast to the strain which transferring large files to multiple users puts on servers, the advantages of this system are that as the storage space, bandwidth, and computing power of the network's computers are pooled, the capacity of the network increases as more members join. Such networks are particularly useful for sharing files, and have attracted controversy for copyright violations, most famously involving the Napster network, set up to share music files.

After the legal problems encountered by Napster, many file sharing networks converted to a fully decentralized system intended to prevent broken links from inducing the collapse of the entire network. The best-known example of this type of file-sharing network is [Gnutella](#), which makes use of servers only to connect peers. Other peer networks, such as [OpenNap](#) have partially adopted the decentralized model while retaining some centralized features, using servers for search functions. New systems of copyright have also begun to be developed to fit the new concept of file sharing. These include the [Creative Commons](#) licenses, designed to cover a range of intellectual property, and the [GNU license](#), which allows users to make modifications to open source computer programs. P2P networks are organized in a variety of ways, but many enforce some type of protocol on members governing the extent of sharing required or rating members on their sharing record and limiting or extending their download privileges accordingly. An alternative approach is to adapt the

P2P model to a closed network for a specific group of people, for example to replace a central office server. An example of this is [On-Systems](#), who sell software enabling the creation of restricted peer networks.

Box 1: Peer-to-Peer Network Resources

The [P2P Foundation](#) maintains a wiki-based site, blog and newsletter with up-to-date news on the progress of P2P as well as some interesting reflective pieces concerned with the implications of P2P for Internet governance.

[P2P Science](#) a site focusing on the applications of P2P networks to science, containing links to software, blogs and discussion forums and events.

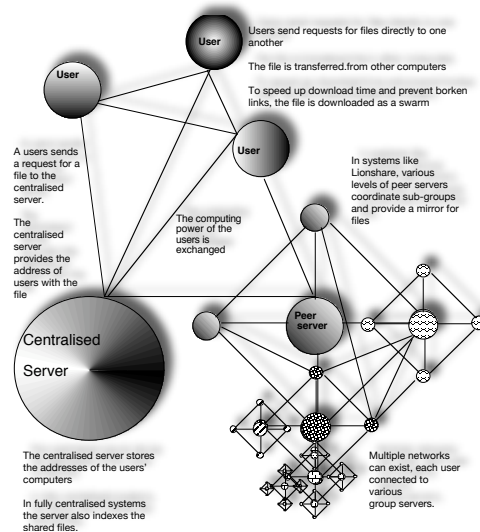
The [Science Creative Commons](#) supplies open source licenses designed to allow scientists to share their work while preventing copyright violations

[CacheLogic](#) a commercial P2P network solutions providers has some useful background and interesting research into P2P issues, including recent surveys showing a shift towards legitimate usage of P2P networks.

Despite these developments, peer-to-peer networks are still controversial and the division between those who advocate them as a form of liberation from the centralized control of the Internet and those who regard them as theft of intellectual property. The controversy surrounding copyright has perhaps delayed the application of its principles outside the business of entertainment. However, the efficacy of using file-sharing protocols such as [BitTor-](#)

[rent](#) that can speed up the transfer of large data sets, such as those involved in the human genome and phenome projects, is evidently applicable to academic collaboration (1).

P2P networks have already been harnessed to identify drug candidates using projects such as Think, a project begun in 2001 as a collaboration between Oxford University's Centre for Computational Drug Discovery, the National Foundation for Cancer Research, and two software companies, the aim of which was for a program running as a screen saver to test the binding interactions of proteins against a bank of small-molecule drug candidates (2). This project led to a variety of similar projects, now coordinated by [Grid.Org](#).



As well as the large data sets involved, bioinformatics a discipline in which the proliferation of available software to analyse large datasets often makes it difficult for researchers to keep up with the latest tools or to select the most appropriate. [Chinook](#), a P2P project sponsored by Genome Canada, aims to unify access to alignment software and facilitate comparisons. To this end, Chinook is a freely available software designed for self-administered

scientific communities of various sizes to run and compare the various programs available for computational biology in a P2P setting. Chinook uses Java client or Perl Engine, but is also available as a web application, Chinook online (3).

Box 2: Peer-to-Peer Software

[Open P2P](#) provides free downloads and documentation for a wide range of P2P software and a host of useful links.

[Linux P2P](#) provides a list of P2P software compatible with Linux and GNU operating systems.

[FolderShare](#) allows the creation of a searchable file-sharing network. Acquired by Microsoft in 2005, it is now available as a freely downloadable plugin.

[Google Desktop](#) provides a similar service, although critics have warned that it will compromise user's privacy, as documents retrieved during searches of local computers will be stored on Google's own servers to enable the search to be run (4).

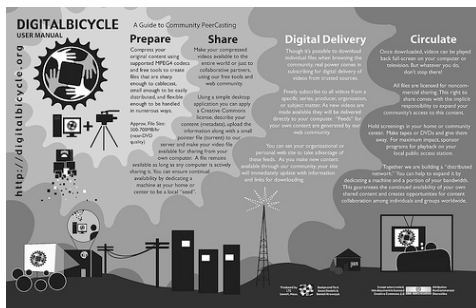
[AllPeers](#) is soon-to-be released free file-sharing software that has been attracting some attention from the P2P community.

[Digital Bicycle](#) is another interesting project in the development stages that plans to combine BitTorrent file sharing protocols with an open-source content management system, [Drupal](#), and RSS syndication.

The utility of file sharing for collating academic information is being increasingly acknowledged. A [UN declaration](#) from the [Civil Society Science Information Working Group](#) stated that peer-to-peer networks should be promoted as a means 'to share scientific knowledge and pre-prints and reprints written by scien-

tific authors who have waived their right to payment'.

One of the first attempts to put this declaration into practice on a multi-institutional level comes from [LionShare](#). LionShare is a project of the Pennsylvania State University and is developing its software in collaboration with the open source authentication project, [Shibboleth](#), MIT's [Open Knowledge Project](#), researchers at Simon Fraser University and the [Peer to Peer Working Group](#). Lionshare includes personal servers to store individuals' files and networking to support peer-to-peer file sharing. To avoid the obvious disadvantage of P2P networks, that the files available are dependent on the particular users connected at a given time, Lionshare uses 'peer servers' to aggregate documents from small and large groups, and to provide a persistent mirror for files to prevent blockages due to disconnections of personal computers. The project also involves building on Gnutella's protocol to develop more advanced search functions. The most ambitious part of the project, however, is to facilitate collaboration between separate academic institutions, using Shibboleth to overcome the traditional barrier of differing end-user identities. Unlike most P2P networks, users of LionShare are required to identify themselves, making violations of copyright standards less likely.



A vision of the future of p2p networks from Digital Bicycle posted to [Flickr](#) on **November 29, 2005** by [daniell](#)

The goal of Lionshare's Connecting Learning Object Repositories parallels previous initiatives, in particular

the work of the Scholarly Publishing and Academic Resources Coalition ([SPARC](#)), which works to promote the concept of institutional repositories, in which academics are encouraged to archive their scholarly output. Like the concept of institutional repositories, Lionshare's protocol shifts the burden of archiving and indexing work towards individual members of the institution. Both systems also allow various levels of groups with different authentication privileges and have the eventual aim of enabling access to resources between institutions. LionShare include in their 2004 white paper the proposal that users will be able to access documents stored in repositories as well as those on peer servers and the user's own computer using a single search function Secure eduSource Communication Layer (ECL). It will be interesting to see how whether the two models can converge, perhaps by evolving short and longer-term storage and retrieval solutions for electronic material.

References

- 1) Church, G. M. "[The Personal Genome Project](#)". Editorial, *Molecular Systems Biology* doi:10.1038/msb4100040 Published online: 13 December 2005.
- 2) MacFarlane, John. "[PCs enlisted to cure cancer](#)". *Nature Medicine* 7, 517 (2001) doi:10.1038/87813.
- 3) Montgomery, Stewart et al. [An application of peer-to-peer technology to the discovery, use and assessment of bioinformatics programs](#) *Nat Meth* 2 (8), 563 (Aug 2005). doi:10.1038/nmeth0805-563
- 4) Steward, Sid. [FolderShare remote computer search: better privacy than Google Desktop?](#) Posted to Orielly Weblog, Feb. 15, 2006 09:21 AM