

## SWCA Initial Research Notes

### Introduction

The South West Creative Archive (working title, SWCA) is a proposed system for easy distribution of large files using peer-to-peer technology (p2p). A peer-to-peer system (aka file trading system) is a method of exchanging data that combines the networked, distributive power of each individual computer, to enable an interconnected mesh of network bandwidth.

### System Description & Overview

This system consists of a p2p client that each user downloads and installs, and a website that provides not only the p2p client, but also a catalogue of files that can be downloaded from the network.

This is a research report as part of a plan for creating a system to unlock the creative archive of the South West by enabling institutions to free their media files for distribution. It is intended that this network consist of a number of interlocking elements with the common factors that they be easy to implement, easy to use and free. It would place the control of content distributed in the hands of individual people and institutions. The system consists of a network of peer-to-peer sites, all using a common system. These are the distributions nodes of the system. Each node would be setup and maintained by any individual or institution wishing to contribute media artifacts for distribution.

Running hand-in-hand with this network would be a website that acts as a catalogue of media artifacts. This would allow people to browse through the archive to search for items of interest and find out how to download them. This catalogue would be moderated to ensure the content is appropriate. Also on this website would be links and assistance for download the software needed either to obtain the media artifacts or to set up a distribution node.

### Peer-to-Peer Distribution Network

P2p technology is a method of exchanging data that combines each individual together to build a distribution system who's whole is greater than the sum of its parts. Each part is considered a node within a larger network. We propose building an easy-to-use wrapper around the existing bittorrent peer-to-peer system that would allow the easy setting up of an area on a computer dedicated to distributing media. This wrapper would remove most of the technical aspects around configuring a computer and its network settings and to begin distribution. This would mean that anyone could easily begin distributing media files to the wider world using an existing connection to the Internet. As the bittorrent system uses spare bandwidth and operates in a manner that means the more requests for a particular file, the easier it is to access.

This system would work as follows: The users who wishes to distribute media artifacts downloads the peer-to-peer node software from the website. They then run this software to set up a node within the distribution network. Once they add films etc. to this node, the cataloging system on the websites lists what has been added. This list is passed on for moderation before being publicly available and added to the catalogue.

### Website

This site would act as the hub of the project. From here software, technical help, information and the content catalogue itself would be hosted. This site would need the following elements:

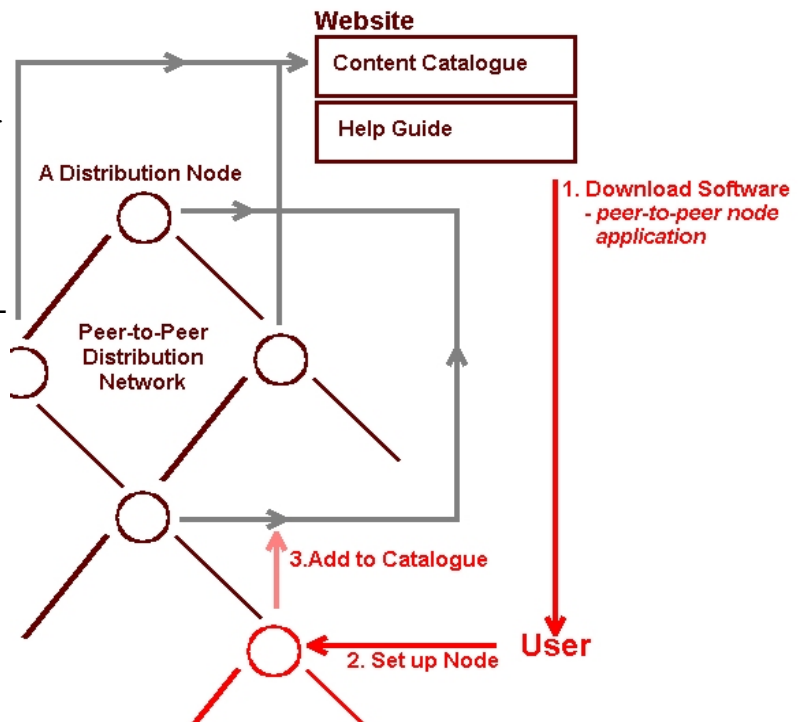
**Content Catalogue:** A system that stores information such as title, introductory text, copyright information, credits, contacts, images, codecs and any other information. A system would need to

be implemented to ensure that entries to the catalogue are vetted to ensure that any inappropriate media artifacts are not added to the system.

**Help Guides:** This section would offer support setting up a distribution node, how to use the content catalogue to both list and find items, how to download media artifacts and any other areas of support.

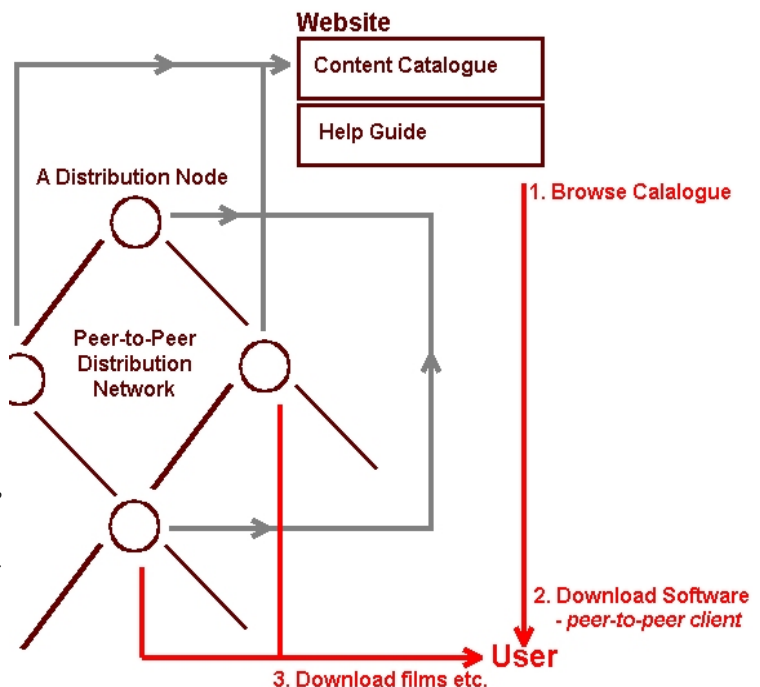
**Links:** This would have links for the downloading of software (such as the distribution node software and the peer-to-peer client), links to the open source code that powers these systems and other pertinent links.

The website would integrate into the system outlined as follows: The user would browse the content catalogue to find media artifacts that are of interest. Once these had been located the user would then need to download the peer-to-peer client to plug-into the network to then obtain the films etc.



### Technological Basis

P2p describes an overarching technological concept rather than an exact protocol. Within p2p technology there are a number of different methods of both transferring and publishing the data. In considering peer to peer systems, a decision needs to be taken in regard whether to use p2p systems or a torrent based p2p system. Torrent based p2p systems use a small text file (the torrent) to 'control' the distribution of a named file. Torrent systems break the file being distributed into smaller pieces, which can easily be transmitted around the network and re-assembled on arrival. The system is an intelligent network in the sense that it boosts demand for rarer pieces of files automatically to keep the load evenly distributed overall. This means that the network of connected



computers can concentrate their collective distributive power on files with a higher demand, meaning a faster download speed and more efficient use of bandwidth than rival systems. Indeed, the torrent based p2p concepts are predicted to be the method that computer networks evolve towards to deliver the bandwidth that current predictions suggest will be required; "...as bandwidth applications grow, the data centers will never be ready to serve 30 million concurrent streams of

data. [For example] Akamai [a major US data backbone provider used by Yahoo!, Microsoft and MTV], with its tens of thousands of servers spread in an intelligent topology, still can't serve more than 150,000 concurrent streams, which is never going to impress the TV network exec used to audiences in the millions...P2P is the solution to delivering not only high quality video but also to audiences that scale in the millions..."<sup>1</sup> The website catalogue system would mirror the torrent files that individual users publish, whilst at the same time incorporating them into a system that allows them to be catalogued for ease of access.

## Case Study – LionShare

A comparable system to the SWCA is the academic peer-to-peer network; LionShare.<sup>2</sup> The project was developed by Pennsylvania State University along with Massachusetts Institute of Technology Open Knowledge Initiative. The system is based on non-torrent technology, specifically the Gnutella model.<sup>3</sup> This is a system that we are familiar with, having documented it for a plugincinema project as well as when writing 'Plug In and Turn On: A Filmmakers Guide to the Internet'.<sup>4</sup> There are obvious attractions to the system, especially its easy of use; as making material available on the p2p network does not require torrents to be created, many of the clients it uses (e.g. GNUcless) are based on the open source GNU license, thus giving a secure path to added development. The system is described as; "...a hybrid model that mixes the Gnutella decentralized peer-to-peer network with a more traditional client-server network. Users of this program are able to upload files to a server where they can be shared continuously, regardless of whether or not the user is online. This network allows for a much smaller than normal sharing community."<sup>5</sup>

In regard to issues of copyright and network security, the LionShare system requires all of its users to be registered. For our purposes - registering every user, even those who just download would be an additional and considerable administrative burden. A non-torrent system that wished to control the content on its network would need to adopt such security measures to ensure that users do not abuse the system. However a torrent based system means that access to the torrents themselves can be used as a form of control rather than registering users, providing a good balance between flexibility and security. We suggest that other methods might be best employed to address the security and copyright issues. It must also be possible that the reason the LionShare system did not use torrents as a basis is that this technology could not be considered 'proved' until 2004, whereas the LionShare basis was formalised in 2003.<sup>6</sup> In addition, the SWCA is intended to extend beyond the academic domain, and therefore needs to be easily expandable.

What is important, is that LionShare recognised the power such technology would have as an educational tool; "The information collected during the study [Penn State's Visual Image User Study (VIUS) - a 26-month project funded by the Andrew W.Mellon Foundation] indicated that current academic collaboration and sharing tools fell short. Specifically, a new application was needed that provided more flexible user-controlled tools and expanded capabilities for the discovery, management, and sharing of multimedia files. The assessment also suggested that peer-to-peer (P2P) technology would offer unique opportunities for expanding and building on the types of tools and services that VIUS participants identified as desirable."<sup>7</sup>

---

1 <http://slashdot.org/article.pl?sid=06/02/25/2115218>

2 <http://lionshare.its.psu.edu/main/> and [http://en.wikipedia.org/wiki/Peer\\_to\\_peer](http://en.wikipedia.org/wiki/Peer_to_peer)

3 <http://www.gnutella.com/>

4 [www.plugincinema.com/plugin/plugin\\_cinema/pffp.htm](http://www.plugincinema.com/plugin/plugin_cinema/pffp.htm) and [http://www.plugincinema.com/plugin/about\\_us/plugin\\_book.htm](http://www.plugincinema.com/plugin/about_us/plugin_book.htm) see appendix iii.

5 <http://lionshare.its.psu.edu/main/info/docspresentation/LionShareWP.pdf>

6 <http://www.suprnova.org/?op=showLong&aID=25> and <http://lionshare.its.psu.edu/main/info/docspresentation/LionShareWP.pdf>

7 <http://lionshare.its.psu.edu/main/info/docspresentation/LionShareWP.pdf>

## System Needs

From our analysis of the construction of this system requires consideration and/or development of the following:

- Consider using torrent technology as the basis for the p2p protocol. This has the following advantages:
  - it allows the control of content (via the torrents) in some areas (via the catalogue)
  - it frees up any constraints on user registration for general users
  - it gives control to those who use the content catalogue to publish.
- The development of a 'safe' version that has a restricted search option and would be limited to legal torrents. It could, for example, be limited to places where torrents are verified such as LegalTorrents.com and the catalogue we create. The system would have several pre-configured versions, e.g. a server version, individual version, university server/individual version, safe version or use an auto detect or wizard to look at this.
- Ports & Firewalls will be an issue especially for university systems. Working along side the university IT teams will give us valuable information as to how the system deals with these issues. It could be that we use assigned ports via specific 'locked' versions or 'Safe' versions to address those issues. An auto-detect function could also handle this issue by searching to connect via pre-prescribed protocols.
- It must be easy to use, both to download and publish content as a client and content provider. This means that we need to gather information on how torrents are published as well as examine possible 'trackerless' torrent publishing.<sup>8</sup>
- It must have the potential to be multi-platform. While this project would concentrate on Windows development, it has to be extendable to other platforms such as Apple Macintosh and Linux/GNU.
- We need to consider how it may relate to other devices such as a PSP or Mobile phones, to ensure that the system can grow beyond this stage.
- Consideration needs to be given to either a tool to enable films to be converted, or into recommendations of media formats for audio/visual works and documents and licenses. This is important as the sharing is just part of the system – users also need to be able to view the files downloaded. This could be complemented by the adding of a media player into the system - open source possibilities include VideoLAN and Mplayer.<sup>9</sup>
- Consideration needs to be given to how future developments, such as the addition of components such as commerce systems and Digital Rights Management (DRM).<sup>10</sup>
- The system development should be open source and use the GNU License, to allow us to both benefit from existing systems and keep the system as a living entity, with scope beyond any development stage that individual groups or companies may undertake.<sup>11</sup>
- The system needs to account for how we are going to incorporate the BMEX server, although it works better with its involvement, as the system has to function without it if needed.

## Software Client Basis

This project proposes a starting point using existing software that has been developed under a license that allows us to build upon their work. There are a number of software clients that use the

---

8 A BitTorrent tracker is the name given to a server which assists in the communication between torrent peers. It is a major critical point in any network, as clients are required to communicate with the tracker to initiate downloads.

9 <http://www.videolan.org/vlc/> and <http://www.mplayerhq.hu> and <http://mediacoder.sourceforge.net/> for file conversion.

10 Possible contenders are DReaM <http://www.sun.com/smi/Press/sunflash/2005-08/sunflash.20050822.2.xml> and OGG-S <http://slashdot.org/articles/03/04/03/2059223.shtml?tid=141>

11 <http://www.gnu.org/licenses/gpl.txt> (Nb. there is a version 3 of this arriving! Draft of GPL 3 is here: <http://gplv3.fsf.org> which considered mobile devices more.)

torrent basis that we could use as the core of the SWCA system – however we need to choose a client that meets the current and possible future needs of the project. This choice was made based on the analysis of the system needs, already outlined that are existing development areas, as well as looking at what additional technology the various clients may currently support. There are over 35 torrent clients. We have examined some of the most widely used and assessed them by the project criteria:

<i>Torrent Client</i>	<b>System Needs (see above)</b>				<b><i>Additional Technical Support</i></b>					
	Multi-Platform	Open Source <sup>12</sup>	Alt. Control Systems <sup>13</sup>	Torrent Search Engine <sup>14</sup>	UPnP Port Mapping <sup>15</sup>	NAT traversal <sup>16</sup>	DHT <sup>17</sup>	Peer exchange <sup>18</sup>	Tracker-less Torrents	RSS <sup>19</sup>
Azureus	Yes	Yes	Yes	Yes	Yes	Some	Yes	Yes	Yes	Yes
BitComet	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
BitTorrent	Yes	Yes	Yes	Yes	Yes	Yes	No	No	Yes	No
Edonkey 2000	Yes	No	Yes	No	No	Yes	Yes	No	Yes	No
Shareaza	No	Yes	Yes	Yes	Yes	Yes	No	No	No	No
µTorrent	No	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes

From this analysis it can be seen that Azureus meets the most criteria established. Azureus<sup>20</sup> has won awards from the development community.<sup>21</sup> It has other advantages in that it is a Java based system, as such, once developed it offers great scope to runs on Linux/GNU, Windows, Macs & Unix via the 'Write once, run anywhere' philosophy of Java<sup>22</sup> and it can be made compatible with Videora, the PSP casting software.<sup>23</sup> In addition there has been another academic project (though not on the same scale as either the SWCA or LionShare that also used Azureus as a basis – this was LocalHost and it was developed as part of a thesis under the Peer-to-Peer Networks and Applications Research Group project at Melbourne University. The architecture and approach being researched by this system (easy publishing and decentralised network storage) also backs the decision to base the client on Azureus.<sup>24</sup>

12 Is the client available under any Free/Libre/Open-Source Software agreements?

13 Related to above – does it have extensions for either/both web and mobile control.

14 Torrent Search Engine - for 'Safe' Version, see above.

15 Universal Plug and Play (UPnP) - "The goals of UPnP are to allow devices to connect seamlessly and to simplify the implementation of networks in the home and corporate environments."  
[http://en.wikipedia.org/wiki/Universal\\_Plug\\_and\\_Play](http://en.wikipedia.org/wiki/Universal_Plug_and_Play)

16 "In computing, NAT traversal refers to a solution to the common problem in TCP/IP networking of establishing connections between hosts in private TCP/IP networks which use NAT (network address translation) devices."  
[http://en.wikipedia.org/wiki/NAT\\_traversal](http://en.wikipedia.org/wiki/NAT_traversal)

17 "Distributed hash tables (DHTs) are a class of decentralized distributed systems that partition ownership of a set of keys among participating nodes, and can efficiently route messages to the unique owner of any given key."  
[http://en.wikipedia.org/wiki/Distributed\\_hash\\_table](http://en.wikipedia.org/wiki/Distributed_hash_table)

18 "Peer exchange (PEX) is another method to gather peers for BitTorrent in addition to trackers and DHT. Peer exchange checks with other peers to see if they know of any other peers."  
[http://en.wikipedia.org/wiki/Peer\\_exchange](http://en.wikipedia.org/wiki/Peer_exchange)

19 RSS (aka Really Simple Syndication) is a family of web feed formats, specified in XML and used for Web syndication.

20 <http://azureus.sourceforge.net/>

21 'Azureus Wins SourceForge Community Awards' <http://www.slyck.com/news.php?story=1151>

22 <http://java.sun.com/features/1997/aug/wora.html>

23 <http://www.videora.com/en-us/Support/Clients/azureus.html>

24 <http://p2p.cs.mu.oz.au/software/Localhost/faq.html>

## Other Development Issues

- The catalogue website, which would possibly be cross-linked to the client, would need to detect if the user is a member of the catalogue website and so aid the publishing of their material. Another consideration is the ongoing issues around the membership and policy of the catalogue website.
- Consideration also needs to be given to copyright of published material. Though individual members of the catalogue website will be responsible for this, the system would benefit from supporting this process via such licenses as creative commons; allowing the files to be linked to the appropriate license.

## Frequently Asked Questions

### **Q. What about people using the system to publish copyrighted material?**

**A.** The content that the system would be responsible for, that which is published via the content catalogue is done using a password protected login that each member of the system has. It is therefore down to each institution or group to decide what to publish and under what copyright status to publish it. All published materials on the content catalogue are therefore accountable. The website will provide support with copyright issues for system members and discussion between the members will also provide a fertile platform for advancing such issues.

### **Q. Will the system be Safe? Will the system make networks insecure?**

**A.** The system will be built in collaboration with the IT departments of the universities and with a wide range of technical input. We are also basing the system on tried and tested existing technology. We are confident this will result in a stable and safe system for use.

### **Q. What will stop users from downloading offensive material?**

**A.** The project is well aware that users such as schools and universities may wish to ensure that users are not abusing the system or exposed to offensive material. The current proposal for how to deal with this issue is to build a 'safe' version of the client software that is locked to download sites (such as the content catalogue we will build) that only have legal material on.

### **Q. Will the system not use too much bandwidth?**

**A.** We do not believe so for two reasons. First is that each client system has controls to allow the user to alter the bandwidth his or her application is currently using. Second the system's use in many areas (see 'Q. What will stop users from downloading offensive material?' above) and so will have limited scope to indulge in the wholesale download of pirated material as some estimate p2p technology is currently doing.<sup>25</sup> In addition as the system is using pre-defined methods of data exchange (see question '*Will the system be Safe? Will the system make networks insecure?*' above) the bandwidth can be monitored and future builds of the software reconfigured to compensate.

---

<sup>25</sup> " CacheLogic suggests that BitTorrent traffic accounts for ~35% of all traffic on the Internet..."  
<http://en.wikipedia.org/wiki/BitTorrent>

## Project Timetable

<i>Stage</i>	<i>Date</i>	<i>Details</i>
A	17/04/2006	Research and Evaluation: Initial report on proposed system to be built.
B	15/05/2006	Institution based Research: Report into client needs and re-evaluation of initial Report.
C	15/05/2006	Pre-alpha phase: Pre-Alpha version of P2P software.
D	12/06/2006	Optimizing Phase and Website testing: Alpha website design and document detailing test results of beta testing.
E	30/06/2006	Release Version Evaluation: Examine existing systems with a view to producing design for release version.
F	30/09/06	Coding Beta Version: Beta Version of P2P software.
G	30/10/06	Full Website Phase: Implementing strategy for final content catalogue membership system.
H	26/02/2007	Coding Release Version: Final version of P2P software.
I	13/04/2007	Final Testing Phase: Website and doc detailing test results of final version testing.

*Report by FluffyLogic / [www.fluffylogic.net](http://www.fluffylogic.net)*