

From steve@cs.uow.edu.au Mon Apr 15 12:50:40 1991
Received: from wraith.cs.uow.edu.au by jatz.aarnet.edu.au with SMTP id AA0577:
(5.65+/IDA-1.3.5 for lmb); Mon, 15 Apr 91 12:50:19 +1000
Received: from jasper.cs.uow.edu.au by wraith.cs.uow.edu.au (5.65b/IDA-1.4.3)
id AA08277; Mon, 15 Apr 91 12:08:18 +1000
(from steve for tjm@avcc.edu.au)
Received: by jasper.cs.uow.edu.au (5.65b/IDA-1.4.3)
id AA00315; Mon, 15 Apr 91 12:08:10 +1000
From: Stephen Cliffe <steve@cs.uow.edu.au>
Message-Id: <9104150208.AA00315@jasper.cs.uow.edu.au>
Subject: AARNet Project Proposal
To: tjm@avcc.edu.au
Date: Mon, 15 Apr 91 12:08:08 GMT

I am including below a copy of a project proposal I wish to submit for consideration under the 1991 AARNet Projects funding.

A formatted copy of this document has been submitted to Professor Ken McKinnon.

I can email you a postscript version of this paper suitable for printing on a Laser printer if you like.

Steve Cliffe,
University of Wollongong.

GA

AARNET PROJECT SUBMISSION

10

Archiving within Australia

Stephen Cliffe

<steve@cs.uow.edu.au>

Department of Computer Science

University of Wollongong

April 1991

Abstract

This document describes the current state of archiving in Australia. It first gives some background on archiving stating the problems facing us before describing some of the solutions that are in place now. Finally, it outlines how both these current activities as well as future developments can be supported by AARNet in the way of funding.

1 Background

1.1 Introduction

There is a vast amount of freely available software and data on various machines spread throughout the network. For this to be of any benefit, a user needs to be able to find the things he is interested in and identify things of potential interest. The problem of finding this valuable information is one that exists worldwide.

In Australia, we are faced with the additional problem that most of this software and data is located overseas. Our network link to the rest of the world is an expensive and limited resource. Because of this, we must ensure that our use of it is as efficient as possible. Recent figures have indicated that 60% of incoming traffic on our overseas link is due to file transfers, so anything we can do to reduce this is going to be very beneficial.

1.2 The Problems

>From the situation outlined above, we can see that we, the AARNet community, are faced with a number of related problems:

1. A user needs to know what software and data is freely available.
2. A user may know that something is available on the network but doesn't know where from.
3. Given that most of this software and data is available from overseas, we need to make sure that we import it as efficiently as possible. This involves some form of co-ordination and co-operation to ensure that something is not imported more than once.

In summary, there are two basic problems. Firstly, the identification and location of information, and secondly, the efficient transfer or distribution of that information.

2 Current activities

Given the problems we face, a number of things are being done.

2.1 Mirroring

In order to try and reduce the amount of traffic on our overseas link, a number of sites in Australia are acting as mirror sites of popular archive sites overseas. This works by a local site maintaining a copy of some or all of the files offered by an archive site overseas. This local mirror site will update it's copy of the original sites files at some quiet time each day.

It is in the best interest for users in Australia to make use of these mirror sites for two reasons. The first is the reduction in traffic on the overseas link, as this frees up valuable bandwidth which can be used for other uses. The second is speed. A file transfer from another Australian site is usually much faster.

This technique works well for large or popular archive sites. For example, Craig Warren of Deakin University mirrors a popular PC archive in the US. His machine serves between 5 and 10 megabytes of data a day. Without his efforts this would mean an extra 5 to 10 megabytes traversing (or trying to traverse!) the US link.

2.2 Caching

On Usenet, a newsgroup called comp.archives was created in an attempt to inform people of the availability and location of new software. Because this is such useful information, it is read by a number of people in Australia. The connectivity of AARNet allows them to immediately fetch a copy of anything they see that interests them. Again, this software will usually be from an overseas site. As a consequence of the ease of access users have to this software through AARNet, it is all too easy for them to fetch these files from overseas rather than to ask if someone else in Australia has already fetched it. The net result is that the same piece of software will end up being brought into the country a number of times.

In an effort to get around this problem, the University of Wollongong is testing a system which will automatically fetch this advertised software. Australian users would be encouraged to fetch the software from Wollongong rather than overseas. The overall idea is to significantly reduce the wasted traffic on the overseas link by caching the software at a local Australian site.

2.3 Locating software

The problem of locating software or data is currently being addressed in a number of unrelated ways. As was indicated above, the newsgroup comp.archives is an attempt to keep people informed

about the availability of new software. Here in Australia we have taken a similar by creating a newsgroup called aus.archives which is used to announce locally (i.e. Australia wide) archived software and data. Using this information, the University of Wollongong maintains a list of Australian archive sites. This list is automatically posted monthly.

At McGill University in Canada, they run a system calledarchie. It is a database of software available to hosts connected to the Internet. The system can be accessed in an

interactive fashion or via electronic mail queries. Craig Warren is currently negotiating with the people at McGill to run the system in Australia.

3 Future Directions

3.1 Archives Working Group

The current activities in the area of archiving are really only interim solutions. What is needed is a more complete solution to the overall problems facing us. At this stage it is not clear what the solution is. To this end, an informal working group has emerged to discuss the problems. The members of this group are:

Stephen Cliffe	University of Wollongong
Mark Prior	University of Adelaide
Craig Warren	Deakin University
Geoff Huston	AARNet

The first three members of the group have been working in the area of archives in Australia since before the inception of AARNet. With the interest and experience of these members in the area and close liaison with AARNet in the form of Geoff Huston, the group should be well able to tackle the problems of archiving facing the AARNet community.

3.2 New Activities

There is something on the horizon which may provide an overall solution to the problems of archiving. It is the Prospero virtual file system being developed at the University of Washington.

It allows users to wander around a virtual filesystem looking for what they want. The appearance, or organisation, of this file system is customizable. When a user has found what he or she wants, they can issue a command and the system will fetch a copy of it for them from wherever it may reside on the network.

The Prospero database will be automatically updated from the articles posted to the comp.archives newsgroup and will be interfaced to thearchie database described earlier. This solves the problem of locating software and data by providing a single, user friendly interface to a database of freely available software.

The system is powerful, distributed system designed to cope with the world wide problem of archiving. It is currently running on systems in 7 countries around the world including here at Wollongong. We are interested in contributing to the development of Prospero in the form of integrating our mirroring and caching systems to it and eventually setting up a number of Prospero servers in Australia for the benefit of AARNet users. This would mean that if someone fetches some files through the Prospero system that has to come from overseas, they will be cached at one of the Australian

Prospero servers and the database updated to ensure that future requests for the those files will be serviced by the local server that caches them.

4 Support

At present, this work on archives in Australia is being carried out at the expense of the institutions involved. In order to continue the ongoing support of the current archiving activities and to enable the support for future work in this area, it is requested that these activities and the working group involved with them be formally recognized by AARNet.

The following funding under the AARNet development projects is sought to support archiving in Australia.

Archive computer systems 3 @ \$15k	\$45,000
Travel and accommodation	<u>\$5,000</u>
	\$50,000

The computer systems, 1 for each of Wollongong, Deakin and Adelaide Universities will be used to support the current activities which are presently being run on general purpose machines at the institutions involved. The systems will also be used for development in the area of archives.

A sum of \$5,000 is also sought to allow the working group to meet to discuss the problems and solutions related to archiving on AARNet.

5 Summary

Through the earlier sections of this document it is hoped that the benefits to the AARNet community of the current activities in the area of archiving are clear. The potential for further benefits to AARNet and it's users through future developments in this area is great.

It is hoped that AARNet will see these benefits and offer its support, mainly in the form of funding to enable the participating institutions to further their work in this area.

