

What's New at the ACCC

The ACCC Post



CONTENTS

What's New at the ACCC

1

Managing Your ACCC Email

2

Statistical Tools of the Trade

8

The Head Crash: Disk Space on Borg

10

The Head Crash: Repeated Infections with a Particular Virus

10

New ACCC personal computer labs

The ACCC's personal computer lab in Chicago Circle Center has moved to the fourth floor of the high rise. CCC 401 has 110 1.6 GHz Pentium 4s running Windows 98, all with 256 Mb RAM and Zip drives; some also have CD-RW/DVD-RW drives. CCC 408 has 39 more of the Windows 98 machines with 15-inch LCD displays rather than CRT monitors and 11 Power Mac G4/867s running Mac OS 9.2.2, with 256 Mb RAM, Zip drives, DVD-RW/CD-RW drives, and 22-inch monitors.

There are new labs in Art and Architecture and in the School of Public Health and Psychiatric Institute. A&A 2312 has 25 Power Mac G4/867s running Mac OS 9.2.2 and 5 1.8 GHz Pentium 4s running Windows 98. (With 22- and 19-inch monitors, respectively.) SPH-PI B34 has 31 1.6 GHz Pentium 4s running Windows 98, with 17-inch monitors, 256 MB RAM, and some with CD-RW drives.

Both CCC 408 and SPH-PI B34 are reservable. For more info, see: <http://www.accc.uic.edu/plabs/>

WebMail retention period for sent email now set to maximum

WebMail, the ACCC's Web-based email utility, automatically keeps copies of the messages you send with it in your **Outbox sent-mail** folder (which is the same **sent-mail** folder that **pine** uses). The default time that WebMail will keep each of these copies has now been set to WebMail's maximum — 999 days. The previous default was 15 days.

This change will not affect you if you have already changed this value yourself. That is, if you set the **Number of Days to Retain Messages in Outbox Folder** WebMail option to something other than 15,

that number is saved in your personal settings and won't be overridden by this global change.

This change means that WebMail won't automatically delete the email that you send with it (or with **pine**), so your **Outbox sent-mail** folder could get really big. To prevent quota problems, you should either regularly delete messages from your **Outbox sent-mail** folder by hand or change the **Number of Days to Retain Messages in Outbox Folder** option back to a shorter time period.

For more information, see the WebMail page: <http://www.accc.uic.edu/software/webmail/>

Tigger upgraded in October

Tigger, the ACCC's general purpose UNIX/email/Web server for UIC faculty and staff, was moved to new hardware with three times the computing power and twice the memory. The new tigger is an IBM pSeries 620 model 6F1 server, with six IBM RS64-IV 668 MHz CPUs and 6 Gb memory.

Norton AntiVirus for Windows XP

Norton AntiVirus (NAV) Corporate Edition 7.61, which supports both the professional and home versions of Windows XP, is now available: <http://www.accc.uic.edu/software/antivirus/>

You must upgrade to version 7.61 if you're running Windows XP. You don't have to upgrade if you're using NAV version 7.51 on Windows 9x, Me, NT, or 2000. (You might want to, though; there's a links to the new version's release notes at the above URL.)

Before installing NAV version 7.61, uninstall all other antivirus programs, including previous versions of NAV. This is especially true if you installed v.7.51 on a Windows XP machine.

System Icons:

The Internet and
the World Wide Web

Apple Macintosh



MS Windows



UNIX

Readership Icons:

Everyone



Novice



Expert

Managing Your ACCC Email

Tech Tips



The inevitability of email quotas is right up there with death and taxes. W disks are groaning under the load. We raise your quotas, they raise the spam. So it's practically inevitable that you'll bump up against your email quota sooner or later (before death and after taxes).

What's not inevitable is how email quotas are handled. Until recently, if your email account was on icarus or tigger and you "went over quota," enough of your oldest email messages were removed from your Inbox so that its size went back below your Inbox quota. The truncated messages were moved to "scratch space" and kept there for about seven days, on the theory that you'd have time to retrieve any messages you wanted to keep for the long run. We even had an Email Tools Web page that would retrieve your truncated email for you.

That looked good to us, in theory at least. We wouldn't actually delete your email without giving you a chance to get it back, and when we did, we would take the oldest messages that presumably you wouldn't want anymore.

The main problem with the scheme is that people didn't always expect this behavior. Getting a large attachment you didn't really want might put you over quota, and then some old mail you really did want would be truncated. In the end, mail truncation just doesn't seem to solve the problems of finite disk space any better than the method that we use on mailserv and that many commercial ISPs use — rejecting new email if your quota is full.

So we've retired the icarus and tigger email truncator — the Grim Reaper, remember him? In his stead, we'll use the same system on icarus and tigger that we use on mailserv.

That is, if you go over your Inbox quota on icarus or tigger, we'll return any new incoming email to its sender with a message saying that their message is being returned because your account is over quota. Your old mail won't be touched, but you won't receive any new mail until you decide which old mail you don't really need and remove that mail from your Inbox. Of course, we'll send you email

warnings when you get close to your quota, so you can take action ahead of time.

Not getting new mail is bad, but at least the people who sent it to you will know that you didn't receive it. And we just can't put 50 pounds of electrons in a 10 pound sack.

What are quotas?

Quotas, in detail, can be complicated, because different numbers apply to different parts of your disk storage, and in different ways. The details are in the next section, but the simple answer is that your Inbox can only grow to about:

- ☐ 7 Mb on icarus
- ☐ 25 Mb on mailserv
- ☐ 10 Mb on tigger

We should be able to expand those limits later on in the year, particularly on mailserv. In the meantime, these quotas will hold a great deal of text, though not tons of attachments.

Quotas: The Not So Simple Answer

Unfortunately, you will need to know a bit about more about how quotas work when the time comes that your email approaches your quota. So complicated or not, here goes.

ACCC server quotas come in pairs:

The soft quota is the smaller quota of the pair; it represents online disk storage space available to you for the long term. But you can go over your soft quota for a while — seven days. That's why it's called the soft quota — because you can exceed it for a while.

The hard quota is the larger of the pair. It represents the absolute maximum online storage available to your account. It's called the hard quota because the server won't allow you to go over it.

Because mailserv is an email-only server, it has only one pair of quotas, which applies to everything that you keep "on the server." On mailserv, that's mostly your email Inbox, which holds your

Mailserv or tigger or both?

Yes, in case you are wondering, faculty and staff can have an account on mailserv for email and an account on tigger for a personal Web page or other computing purposes.

incoming email, and, if you use IMAP, your other IMAP mailboxes.

Icarus and tigger, however, are general-purpose UNIX machines. They're used for email and also for personal Web pages and for other computing and programming tasks.

Because there are two types of tasks people do on icarus and tigger (email and other stuff), there are two pairs of quotas: one pair for the incoming email in your Inbox (and only in your Inbox) and one pair for all other the other files you have on the server, in your "home directory." This includes your personal Web pages and your other IMAP mailboxes.

(You shouldn't keep class-, department-, or unit-related Web pages in your personal disk space on icarus or tigger. If you do, let us know and we'll give you space for them in the departmental Web on www.uic.edu or www2.uic.edu. See "Disk Space

for Official Web Pages": <http://www.uic.edu/depts/acc/webpub/overview.html#1>)

What happens when you go "over quota"?

As the total size of all your stored files on mailserv or the size of your Inbox on icarus or tigger approaches and exceeds your soft quota, and again as it approaches your hard quota, you will receive warning email messages.

You must take these warning email messages seriously.

The consequence for staying over your soft quota for seven days or for reaching your hard quota is that new incoming email messages addressed to your account will be "bounced" — returned to their senders with an error message saying that your account is over quota. You will not begin to receive new incoming email until you drop below your soft quota. Resumption of the delivery of new incoming email is automatic, however, and will occur as soon as you go below your soft quota.

Figure 1: Using the ACCC Quota, Download, and Delete Tool for icarus and tigger

The Quota, Download, and Delete tool works the same for accounts on icarus and tigger. The "Maximum allowed size in Inbox" item gives your soft and hard Inbox quotas (see "Quotas: The Not So Simple Answer") and "Maximum Allowed Disk Storage" gives your hard and soft disk quotas (see "Quotas on icarus and tigger").

If you *only* use POP, the "Combined size of remaining email folders" should be zero, because you won't have any "other mail folders" — mailboxes — on the server. If you use IMAP (including WebMail or **pine**), your IMAP mailboxes — the mailboxes that you keep on the server — *other than your Inbox* are kept in your regular online disk storage and the space they take up applies to your disk quota. That's the "Combined size of remaining email folders." The space taken up by your personal Web pages and all your other files on icarus or tigger also apply to your disk quotas. Your IMAP mailboxes and all your other files are included in "Current home directory usage."

When you're having problems with your Inbox quota, click **breakdown by message** in the "Current Inbox Mailbox Size" row for a list of the largest messages in your Inbox. **Click here to view your ten largest email messages (any mailbox)** might be useful when you're having disk quota problems.

Use the **download inbox**, **delete inbox**, and **breakdown/download/delete individual folders** links to download and/or delete entire mailboxes. See the links in "How to delete mail and what to do if the short answers don't work" on page 7 for instructions on how to delete individual messages.

For more information on using this tool and on using downloaded mailboxes on your personal computer, see "Using the ACCC Quota, Download, and Delete Tool":

<http://www.acc.uic.edu/ecomm/quota/quota.tool.html>

The screenshot shows a web interface for "EMAIL TOOLS" from the ACCC (Academic Computing and Communications Center). The user is logged in as "tigger". The main section is titled "TIGGER Mailbox and Disk Usage Information:" and provides the following data:

Current Inbox Mailbox Size:	276.4 KB <small>(breakdown by message / download inbox / delete inbox)</small>
Combined size of remaining email folders:	2.32 MB <small>(breakdown/download/delete individual folders)</small>
Current home directory usage:	2.42 MB
Maximum allowed size in Inbox: (Soft-quota / Hard-quota)	10 MB / 15 MB
Maximum Allowed Disk Storage: (Soft-quota / Hard-quota)	19.53 MB / 24.41 MB
Click here to view your ten largest email messages (any mailbox)	
Mail last checked:	Wed Apr 3 10:43:29 2002 (help)

Below this table is a "Mail Diagnostics:" section with a message: "-- No Problems were found --". At the bottom of the page, there are links for "Home", "Search", and "Contact Us".

Quota, Download, and Delete Tool

You don't have to wait until you get a warning email message from us to check how much mail you have on your email account; you can do that whenever you want, using the **ACCC Mail Tools Quota, Download, and Delete Tool**. Go to the ACCC home page (or any other ACCC Web page), click the purple **Email** button, select **Email - Check your email quota**, then select your email server and **Quota check and folder download/delete**. After you enter your netid and password, you'll see a screen similar to figure 1, page 3, if your account is on icarus or tigger or figure 2, page 4, if your account is on mailserv.

Quotas on Icarus and Tigger

The **icarus and tigger Inbox soft quota** is the space available for your email Inbox, where your incoming email is kept, in the long term. Nothing bad will happen so long as the size of your Inbox on icarus or tigger stays below your Inbox soft quota. But if the size of your Inbox on icarus or tigger exceeds your Inbox soft quota and stays above it for more than seven days, new incoming email will be returned to its sender with an error message saying that your account is over quota.

The **icarus and tigger Inbox hard quota** is the absolute maximum online storage available for your email Inbox on icarus or tigger. Neither server will allow you to go over your Inbox hard quota. If the size of your Inbox on icarus or tigger ever reaches your Inbox hard quota, new incoming email will be returned to its sender with an error message saying that your account is over quota.

Icarus Inbox quotas:

Soft Quota: 7 MB; Hard Quota: 10 MB
Applies to all email in your Inbox on icarus.

Tigger Inbox quotas:

Soft Quota: 10 MB; Hard Quota: 15 MB
Applies to all email in your Inbox on tigger.

The **icarus and tigger disk quotas**: In addition to the Inbox quotas, icarus and tigger accounts also have another pair of soft and hard quotas that apply to the space taken up by all the other files you keep on the server, including all your IMAP mailboxes other than your Inbox and the files in your personal Web pages.

Icarus disk quotas:

Soft Quota: 10 MB; Hard Quota: 15 MB
Applies to email in all IMAP/WebMail/pine mailboxes other than your Inbox and to all other files you have stored on icarus.

Figure 2: Using the ACCC Quota, Download, and Delete Tool for mailserv

Mailserv has only one pair of disk space quotas (see "Quotas: The Not So Simple Answer"); the space taken up by all your files, including your Inbox and your IMAP mailboxes (the other mailboxes that you keep on the server), applies to that one pair of quotas. (See "Quotas on mailserv.")

"Maximum allowed Disk Storage" is the mailserv soft quota; "Incoming email cut-off size" is the mailserv hard quota. If you *only* use POP, the "Combined size of remaining mail folders" should be zero, because you won't have any "other mail folders" — mailboxes — on the server.

[Click here to view the ten largest of your messages \(any mailbox\)](#) for details on the largest messages you have on mailserv.

Use the [download inbox](#), [delete inbox](#), and [breakdown/download/delete individual folders](#) links to download and/or delete entire mailboxes. See the links in "How to delete mail and what to do if the short answers don't work" on page 7 for instructions on how to delete individual messages.

For more information on using this tool and on using downloaded mailboxes on your personal computer, see "Using the ACCC Quota, Download, and Delete Tool":

<http://www.accc.uic.edu/ecomm/quota/quota.tool.html>

ACCC ACADEMIC COMPUTING and COMMUNICATIONS CENTER

EMAIL TOOLS NETID = address ACCOUNT = mailserv -- logout here

MAILSERV Mailbox and Disk Usage Information:

Note: 1 MB = 1024 KB and 1 KB = 1024 bytes

Current Inbox Mailbox Size:	804.15 KB break-down by message / download inbox / delete inbox
Combined size of remaining email folders:	200.81 KB break-down / download / delete individual folders
Total space used:	1.06 MB
Maximum allowed Disk Storage:	24.41 MB
Incoming email cut-off size:	48.83 MB

[Click here to view your ten largest messages \(any mailbox\)](#)

Mail last checked: Fri Mar 8 16:17:32 2002 (msg)

Mail Diagnostics:
-- No Problems were found --

CSO [Home](#) [Search UIC](#) [Contact UIC](#)

Tigger disk quotas:

Soft Quota: 20 MB; Hard Quota: 25 MB
Applies to email in all IMAP/WebMail/pine mailboxes other than your Inbox and to all other files you have stored on tigger.

For more information on disk quotas on icarus and tigger, see “ACCC Online Disk Space Policy”:
<http://www.accc.uic.edu/policies/spacepol.html>

Quotas on Mailserv

Mailserv is an email-only machine used by faculty and staff members. Since all the disk storage on mailserv is email-related, there is only one pair of soft and hard quotas on mailserv, analogous to the tigger and icarus disk quotas. Mailserv does not have a separate quota for your Inbox.

The mailserv soft quota is the space available for your email Inbox, where your incoming email is kept, and for all the other email mailboxes and files you have on mailserv, in the long term. Nothing bad will happen so long as you keep your online storage on mailserv below your soft quota. But if the size of your email Inbox plus all your other mailboxes and files on mailserv exceeds your soft quota and stays above it for more than seven days, new incoming email will be returned to its sender with an error message saying that your account is over quota.

The mailserv hard quota is the absolute maximum

online storage available for your email Inbox and for all the other email mailboxes and files you have on mailserv. Mailserv will not allow you to go over your hard quota. If the total size of your stored email reaches your hard quota, new incoming email will be returned to its sender with an error message saying that your account is over quota, *and you won't be able to move email messages from one mailbox on*

mailserv to another. (You will, of course, be able to download them to your personal computer.)

Mailserv disk quotas:

Soft Quota: 25 MB; Hard Quota: 50 MB
Applies to the email in your Inbox and in all your other IMAP/WebMail mailboxes, and to all other files you have on mailserv.

Which email applies to your quota?

Only email that is on the server — on icarus, mailserv, or tigger — applies to your quota(s). Exactly which email is on the server depends on how you do email— do you use POP or IMAP or both?

POP vs. IMAP

POP and IMAP are two different ways that email programs manage email that's on a remote server such as icarus, mailserv, or tigger. POP is the older of the two protocols. POP was designed to be used when you always read your email from one computer, using one email program, and it is still the best choice if that is how you do email.

IMAP, on the other hand, was designed to be used when you want to read your email from multiple machines, perhaps using different programs. If this is how you want to do email, then you should use IMAP, and only IMAP. As long as you never use POP, it won't matter how many different programs or machines you use.

So, do you use POP or IMAP or both?

If you don't know whether your email program uses POP or IMAP, then you probably use POP — that's the older method and the default for Eudora and most other personal computer email programs.

You will probably know if you use Eudora or Netscape or Outlook with IMAP, because you will have had to specifically set it up to do so.

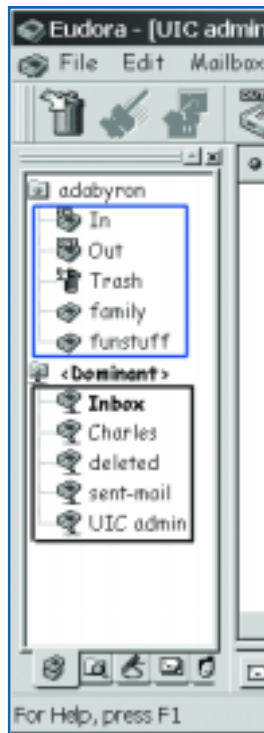
Even if you normally only use POP, if you have ever used the ACCC's email-on-the-Web utility, WebMail, then you've used IMAP, too. WebMail uses IMAP. Likewise if you've used **pine** on icarus or tigger; it's IMAP-compatible.

But rather than guessing, use the ACCC Mail Tools **POP or IMAP?** function to find out for sure. Go to the Mail Tools page, select the server your account is on and **POP or IMAP?**, login with your UIC netid and password, and you'll receive a personalized page that tells you whether you've used POP or IMAP or both to read your email.

Figure 3: Local and IMAP mailboxes in Eudora

The mailboxes in the blue box at the top, **In**, **Out**, **Trash**, **family**, and **funstuff**, are local mailboxes that live on your personal computer; the email in them doesn't apply to your server quotas.

The mailboxes in the black box below the **<Dominant>** are IMAP mailboxes that live on the server. The space taken up by messages in these mailboxes does apply to your server quotas. The **deleted** and **sent mail** mailboxes were most likely created by WebMail.



The **POP or IMAP** Email Tools function can't tell whether you've used **pine**, so if you have, add using IMAP to whatever the POP or IMAP page tells you.

So, now we can answer the important question; which email is on the server?

If you use POP, the email in your local In mailbox that you've downloaded and deleted from the server isn't "on the server" and doesn't apply to any quota. But if you use Leave Mail on Server (don't — the ACCC doesn't support it) or if you mix POP with IMAP, using POP to download a message may not result in its being deleted from the server. (See "How to delete mail and what to do if the short answers don't work" on page 7 for more information.)

If you use a personal computer email program such as Eudora, Netscape, or Outlook, regardless of whether you use POP or IMAP, the email that's in your local mailboxes — ones that live on your personal computer, including the In mailbox, the Out mailbox, and the local Trash mailbox — isn't on the server and doesn't apply to your quotas. If you only use POP (and don't use WebMail), then you will only have local mailboxes.

If you use IMAP (with Eudora or another personal computer email program, or with **pine** or WebMail), opening a message from your Inbox or another IMAP mailbox to read it will *not* delete it from the server. And deleting a message from your Inbox or another mailbox on the server will also delete it from the copy of that mailbox on your personal computer, which won't do for email that you want to keep.

Before You Do Anything

Particularly before you download any email with attachments — if you don't have antivirus software with realtime file protection running, download and install Norton AntiVirus, NAV, now. Run LiveUpdate to get the latest virus definition files and make sure that NAV's Realtime File Protection is running and that it's scanning all file types (page 11): <http://www.accc.uic.edu/software/antivirus/> There are NAV installation instructions for both Macs and Windows in: <http://www.uic.edu/depts/accc/software/nskit/uionline.pdf>

And After You're Done

Use the **ACCC Mail Tools Quota, Download, and Delete** page to check your quota again after you download and delete to make sure you're really back under quota.

Specifically, if you use Eudora with IMAP, then the mailboxes at the top of the Mailbox tab — the ones above <Dominant> — are local mailboxes. (See figure 3, page 5.) The ones below <Dominant>, starting with your Inbox, are IMAP mailboxes that live on the server and all the mail in them applies to a server quota.

Regardless of how you do email, to keep a message indefinitely without it applying to a server quota, you have to transfer it to a local mailbox — one that lives on your personal computer; see figure 3.

Why might you go over quota and what should you do if you do? Three short answers.

① Delete large, unneeded email messages, especially those with attachments.

The quota email warning messages that you receive will list your ten largest email messages and for mailserv will tell you what mailbox they are in. Download these messages if you want to keep them, delete them from the server, and you should be OK.

If a message has an attachment, then *if you know the person who sent you the attachment meant to send it to you* (it's not at all obsessive to ask them) and *only if you really want to keep the attachment*, download the attachment(s) and then delete the message. Just downloading the attachment isn't good enough; that can still leave the attachment on the server. If you want to keep a copy of a message that came with an attachment, forward the message to yourself without the attachment.

Can't find the messages that your email warning message complained about? Want to keep a copy of a message that came with an attachment? There are URLs for Web pages with step-by-step instructions in "How to delete mail and what to do if the short answers don't work" on page 7.

② Haven't checked your email in a while? Log in and take care of your new incoming email.

People often have quota problems when they are away from work or school and don't check their email for an extended period of time. In this case, just logging in and taking care of your new incoming email — downloading it *and* deleting it from the server — might be enough to bring you back under your quota. There are step-by-step instructions in the Web pages listed in "How to delete mail and what to do if the short answers don't work" below.

③ *Consider downloading whole email mailboxes to your personal computer.*

If you don't have just a few really large messages or lots of new incoming email that you can delete from the server, then perhaps the easiest way to solve your quota problems is to download complete copies of email mailboxes — including your Inbox — to your personal computer or to a diskette, and then delete all or most of the messages in those mailboxes from the server.

The **ACCC Mail Tools Quota, Download, and Delete Tool** Web page allows you to download or delete entire mailboxes from the server; see figures 1 and 2 (pages 3 and 4) and, for more information, see the “Using the ACCC Email Quota, Download, and Delete Tool” Web page: <http://www.accc.uic.edu/ecommm/quota/quota.tool.html>

How to delete mail and what to do if the short answers don't work.

Have you tried the short answers above, rechecked your quota, and find that you're still over your quota? Can't find your ten largest messages? Not sure what to do with attachments? Don't worry; the ACCC has a new set of Web pages that give step-by-step instructions on how to clear up quota problems.

Not surprisingly, what you need to do depends on (1) which server your account is on, (2) what program(s) you use when you read your mail, and (3) whether you use POP or IMAP or both.

You most likely know (1) and (2), and for the answer to (3), use the ACCC Mail Tools POP or IMAP page. (See “Which email applies to your quota?” on page 5.) It will tell you whether you've used POP or IMAP or both to read your email. And be sure to add IMAP if you've used **pine**; the POP or IMAP Web page can't tell whether you've used **pine**.

So now that you know the which, the what, and the whether, how do you fix your email quota problem?

Do you use POP properly (on only one computer, don't use WebMail, and don't use Leave Mail on Server)?

Then downloading your new incoming email should fix your problems. There's more in “How to fix icarus, mailserv, or tigger email quota

problems with POP”: <http://www.accc.uic.edu/ecommm/quota/pop.inbox.html>

Do you only use IMAP or IMAP-compatible email programs? That is, do you use Eudora or other personal computer email programs always with IMAP, or do you use WebMail or **pine**, or do you use any combination of these?

Then see the instructions for your email server and whichever IMAP email program you're most comfortable using, in “Fixing Email Quota Problems on ACCC mailserv Accounts”:

<http://www.accc.uic.edu/ecommm/quota/content.s.mailserv.html>

or in “Fixing Email Inbox Quota Problems on ACCC icarus or tigger Accounts”:

<http://www.accc.uic.edu/ecommm/quota/content.s.inbox.html>

Do you use both POP and IMAP, do you use POP with Leave Mail on Server (which the ACCC doesn't support), or would you like to use WebMail?

Then use IMAP to clear up your quota problems, and you should consider switching to using IMAP all the time.

If you want to use Eudora, or even if you use another personal computer email program, for icarus or tigger see: <http://www.accc.uic.edu/ecommm/quota/imap.inbox.html>

or for mailserv: <http://www.accc.uic.edu/ecommm/quota/imap.mailserv.html>

These pages have step-by-step instructions with links to pages that explain how to convert to using IMAP and that have hints on getting started using it.

Or you could use WebMail; with icarus or tigger accounts: <http://www.accc.uic.edu/ecommm/quota/webmail.inbox.html>

or with mailserv accounts: <http://www.accc.uic.edu/ecommm/quota/webmail.mailserv.html>

Or you could use **pine** on tigger or icarus: <http://www.accc.uic.edu/ecommm/quota/pine.inbox.html>

You can even use **pine** on icarus or tigger for mailserv accounts: <http://www.accc.uic.edu/ecommm/quota/pine.mailserv.html>

Comments are welcome; please send them to Judith Grobe Sachs, judygs@uic.edu

Are you going on vacation? Want to know more about the mailbox download and delete tool?

There are tips on how not to go over quota when you're on vacation and on using the ACCC mailbox download and delete Email Tools tool in the online version of this article: <http://www.accc.uic.edu/newsletter/adn33/>

Statistical Tools of the Trade

Tech Tips



Or a Tale of Many Ss

Numbers, numbers, numbers! To make sense of so many numbers, you can use one of the many available statistical programs. The good news is that the ACCC has site licenses for many of the major statistical packages, including SAS, SPSS, STATA, and S-PLUS. (For a list of ACCC-supported packages and links to information about them, see: <http://www.accc.uic.edu/stats/swpages.html>)

Depending on your area of research, you are probably more familiar with one statistical package or another. Social scientists typically prefer SPSS. Biostatisticians typically prefer SAS. But there are many newer packages now that attempt to give you the best of all worlds — a friendly interface, a powerful analytical tool, and a reasonable amount of programming power.

What works for you will depend on a combination of various factors: your general comfort with using software, your programming experience, the size of your data, and the complexity of your analysis. Also the habits and needs of those with whom you share your data may influence your choice of software.

Some people use spreadsheet or database programs for statistical analysis, but that's like using a soup spoon to serve a stew when a serving spoon is available. Statistical packages offer you optimal tools for data management and analysis.

Lets explore what the two major packages, SAS and SPSS, have to offer us.

SPSS

SPSS, formerly "Statistical Package for the Social Sciences," is quite popular not only because it is a valuable analysis tool, but also because it's user-friendly and visually oriented. Open SPSS and in a few minutes you can be using its pull-down menus to run an analysis. SPSS originally emphasized its programming language, but soon became "pull-down" inclined via its windowing environment (which dates back to its DOS implementation!). But the SPSS programming language and syntax is still where the real power of this program resides.

SPSS, like other statistical programs, has evolved into a set of modules. Most of SPSS's modules are available in its main installation media. Some modules that are SPSS's versions of particular algorithms have been packaged as stand-alone products. The installation CD for SPSS contains the following modules: Windows Base, Professional Statistics, Advanced Statistics, Tables, Trends, Categories, Conjoint, Exact Tests, Missing Value, Smart Viewer, and Maps. All of these modules are available under the ACCC's basic SPSS site license. Also available from the ACCC, as stand-alone products, are: Amos, Data Entry, and AnswerTree.

AnswerTree is used for automatically finding statistically significant groups in your data via decision tree algorithms. If you have been around for a while, you might recognize this product under its original name, CHAID. With the new name, a set of new algorithms has been added to this decision tree software. (For more information, see: <http://www.spss.com/answertree/>) "CHAID, for Chi-square Automatic Interaction Detector ..., is an exploratory method used to study the relationship between a dependent variable and a series of predictor variables. CHAID modeling selects a set of predictors and their interactions that optimally predict the dependent measure. The developed model is a classification tree (or data partitioning tree) that shows how major "types" formed from the independent (predictor or splitter) variables differentially predict a criterion or dependent variable." (From "CHAID,"

<http://www.themeasurementgroup.com/Definitions/chaid.htm>)

On the other hand, Amos is used for structural equation modeling and analysis. Formerly known as LISREL, SPSS's version of Amos is a competitive package that "implements the general approach to data analysis known as structural modeling, analysis of covariance structures, or causal modeling. This approach includes as special cases many well known conventional techniques, including the general linear model and common factor analysis." (From: "What is Amos?," <http://www.smallwaters.com/amos/whatis.html>)

This article was written by the ACCC's consultant for statistical software,
Violeta Carrión,
STATS@uic.edu

More detailed information on SPSS products is available from the SPSS site at:

<http://www.spss.com/products/>

SAS

SAS, originally called “Statistical Analysis Software,” has generally been preferred by programmers, primarily due to the power of its programming language and the portability of its code. When you open SAS, you get the somewhat uneasy feeling that the pulldown menus are not going to help. If you have no programming experience, it will take a bit longer to warm up to this powerful tool. But as you get better acquainted with the program, you will come to appreciate its power — one line of SAS code can do many things for your data.

Lately, SAS has been paying attention to the fact that not all its users are programmers and has improved its pulldown menus. Yet its most critical improvement has been with a new add-on product called Enterprise Guide. *This program installs “on top” of your current installation of SAS*, but it can also be individually invoked; running SAS does not imply running Enterprise Guide, but running Enterprise Guide does open SAS.

Enterprise Guide features include a visual display full of helpful pulldown menus, project management tools, and a windowing environment that creates the needed SAS code for you.

Enterprise Guide is included when you purchase Base SAS under the ACCC site license or upon request from qualified UIC SAS user.

SAS has changed a lot across versions, with most of the changes catering to the business community. SAS version 8.2 is the apex of this fact. The version 8.2 installation is noticeably hard drive intensive due to inclusion of a large set of products which are mainly geared towards an “enterprise-level” installation. These are products for a SAS server that might be used at a large company doing e-business, with all the bells and whistles you can imagine. They are a hindrance for the academic user who typically wants to use SAS mainly for its statistical analysis tools.

To complicate matters, SAS does not package its software by the needs of the user, but by the products the site has licensed. That’s of relevance at UIC because the ACCC has upgraded to SAS’s “Gold Tier Academic License.” The good news is

that this license offers a set of new products not previously available under the UIC site license—for example, the SAS data mining product Enterprise Miner. The bad news is that when you purchase SAS from the ACCC, you get a CD folder full of media, containing a bundle of products, many of which might not be useful to you.

To aid you in the selection process of the SAS products you need, we have opted to routinely include with your SAS purchase only those CDs that are typically used for purposes of academic statistical analyses. If you so desire at the time you purchase Base SAS, you can request any of the other products whose installation media we haven’t included. In addition, we have a Web page that will guide you through a custom installation of SAS that is suited to your needs:

<http://www.accc.uic.edu/stats/sas/v8install.html>

From the extensive list of SAS products, the ACCC has licensed the following: Base SAS, SAS/AF, SAS/ACCESS, SAS/ASSIST, SAS/CALC, SAS/CONNECT, SAS/EIS, SAS/ETS, SAS/FSP, SAS/GIS, SAS/GRAPH, SAS/IML, SAS/INSIGHT, SAS/LAB, SAS/OR, SAS/QC, SAS/SHARE, SAS/SPECTRAVIEW, SAS/STAT, Online TUTOR, SAS/IntrNet, SAS/MDDDB, SAS/Secure, SAS/Access to R3, Scalable Performance Data Server, Enterprise Guide, AppDev Studio, SAS/Warehouse Administrator, Enterprise Miner, IT Service Vision, and Enterprise Reporter.

More detailed information on SAS products is available from the SAS site at:

<http://www.sas.com/products/index.html>

We hope you now feel properly introduced to SAS and SPSS. Feel free to contact us for statistical software support matters, suggestions about our Web site, statistical seminars, or any statistical support issue.

Statistically yours,
Violeta Carrión

Stats Support Email: STATS@uic.edu

Support URL: <http://www.accc.uic.edu/stats/>

Webmaster: wwwstats@uic.edu

ACCC Software Sales:

<http://www.accc.uic.edu/stats/software.html>

UICSTATS-L archives:

<http://listserv.uic.edu/archives/uicstats-l.html>

The Head Crash — Q and A

The Head Crash



Disk Space on Borg

Question: Hi, my name is Hillary Cypert, hcyper1. I have an account on the borg system. I am running some heavy duty programs and I have exceeded my disk quota. [See figure 4 below.] Borg says that I should remove 25000 Kb in eight hours. I CAN'T! I need the files I have now and I am probably going to need twice or three times as much in the future. Please stop borg from removing my files and assign me more disk space as soon as possible.

Answer: Hi.

When you login, did you notice the third line of the quota message? [See figure 4 below.]

It is true that you have only 50 Mb of disk space in your home directory. But you also have a directory in the scratch area, `/scratch/hcyper1`. That directory has a quota of 500 Mb [that's half of a gigabyte!], ten times the quota in your home directory.

I took a look in your home directory and one file is the cause of your problem:

```
-rw----- 1 hcyper1 student 75236352
Feb 22 03:59 1013796123.21405.out
```

[That's 75 Mb in the one output file.]

Here's what you should do. Before you run your job, in this case using `bsub`, move your data file and your job file from your home directory to your scratch directory. Then, `cd` to `/scratch/hcyper1` and execute the `bsub` job from scratch.

Not only will you have a soft quota of 500 Mb instead of 50 Mb, you will also see performance gains. The scratch filesystem is striped unlike the home filesystem. A [striped filesystem](#) is one that spans several disks. In the case of scratch, there are fourteen disks comprising it. That means you have much less competition for access to disk heads, because disk writes span several disks. On the other hand, the home filesystem is on a single disk. That means you compete with every other logged-in user for access to the disk.

(Think of borg's filesystem IO like cars travelling on toll roads. Each chunk of data is in a separate car. The road to the scratch filesystem has fourteen lanes, each with its own toll booth. The road to your home directory has only one lane and one booth. Which do you think is faster?)

To ease the urgency of the situation, I've doubled your home directory to 100 Mb. However, I urge you to use scratch for the reasons I've outlined above.

Michael Homa
Systems Group

[Michael is the keeper of our compute server, borg. `bsub` is part of the LSF, Load Sharing Facility, the set of commands that you use on borg to submit and manage jobs. For more information, see "LSF on the Borg System," <http://www.acc.uic.edu/hardware/borg/lsf/>]

Repeated Infections with a Particular Virus

Question: [This question was asked by a departmental "resident expert" who shares the responsibility of maintaining the other machines in their department.] The `VBS.Stages.A` virus has been detected on my PC by Norton AntiVirus (NAV) with ever-increasing frequency over the past few months. I've been running a Norton scan every week, sometimes more if I'm having problems loading databases, very slow response time, and each time Norton has found from a few to 640 (yesterday) infections. The files affected are mostly in the

Figure 4. Quota for Student Accounts on borg

Borg executes the `quota` command for you when you login. This is Hillary Cypert's. The third line says that Hillary is over her soft quota for her home directory. (Her quota is 50 Mb and she's using 75 Mb.) The second line says that Hillary has a 500 Mb — one-half gigabyte — quota in `/scratch`.

Even though Hillary isn't using her scratch space, she has one file in `/scratch`. Everyone has at least one file in `/scratch` — their scratch "home" directory, `/scratch/netid/`

For Hillary, that's `/scratch/hcyper1/`

Disk quotas for hcyper1 (uid 12345):

Filesystem	usage	quota	limit	timeleft	files	quota	limit
<code>/scratch</code>	0	500000	550000	-	1	0	0
<code>/homes/home5</code>	74999	50000	75000	7.8 hours	191	0	0

Windows temporary directory with **.dat** extensions. Norton quarantined them, and because my system is running so slow now, I'm wondering if I should permanently delete them.

Has anyone else experienced this and what solutions can you offer?

Answer: Here is the info from Symantec: "This worm appears as an attachment named **Life_stages.txt.shs**. When you run the attachment it opens a text file in Notepad. The text file describes the male and female stages of life. While you are reading the text file, a script is running in the background. This worm spreads itself using Outlook, ICQ, mIRC, and PIRCH." (From <http://www.symantec.com/avcenter/venc/data/vbs.stages.a.html>) The fix is at: <http://www.symantec.com/avcenter/venc/data/fix.vbs.stages.html>

It appears that something in your office (or an associate's office) is spreading this infection. If you agree that you have NAV on all your machines, and if they are scanned and up to date with their virus definitions, then either you are missing a machine somewhere (someone's laptop?) or someone that you deal with often is giving these viruses to you.

The fact that Stages is old and that it spread pretty far and fast, means it's likely that the ACCC is filtering it from the outside. [We are now, but we weren't when this exchange took place.] So, unless you are using the MS Exchange VirusEngine for mail, someone on campus is probably sending this virus to you.

If I were you, I would (preferably on a weekend):

- * Run Norton Antivirus's LiveUpdate to get the newest NAV virus definitions and scan all my machines with NAV.
- * Run the Stages fix utility on all my machines. [At the Symantec link above.]
- * Set up preventative measures. Depending on your department's infrastructure, this could be default filters on client machines in your office that filter all email attachments with the extension **.shs** to the trash, or server email filters that delete all messages with **.shs** attachments. (The later would obviously have to be done with the help of your clients.)
- * I'd also use this time to run Windows Update on all my machines to prevent other infections and security breaches. [Windows Update is a Microsoft Web site that checks your Windows operating system and Internet Explorer for both necessary and optional updates. Just visit <http://windowsupdate.microsoft.com/> and, with your permission, it will check your machine and offer a list of updates that you may choose to install if you wish. Install every "required security update."]

Or... you could skip the above steps and just take a weekend to install Macs in your department. That's not good for tech job security though. ;-)

As for whether you should delete the quarantined virus files? Of course! Only virus researchers have any reason to keep viruses hanging around. You should have NAV's Realtime File Protection [figure 5] — Microsoft Exchange Realtime Protection if you use Exchange — running all the time. That will keep you from getting infected from a known virus like Stages in the first place.

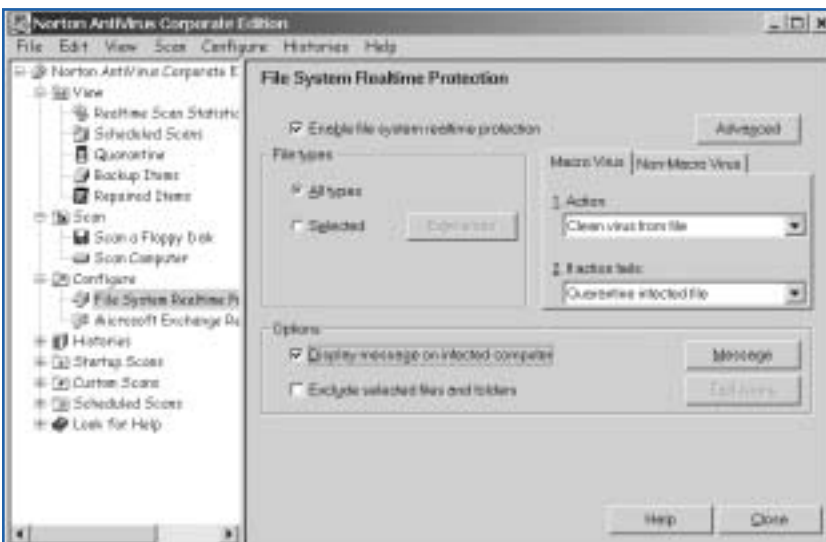
Steven Bandyk, sbandyk@uic.edu
ACCC Micro Repair Supervisor

[Can you tell that Steven is one of the ACCC's resident Mac experts?]

Figure 5: Setting Up Norton Realtime File Protection

If you don't already have Norton AntiVirus, install it (<http://www.accc.uic.edu/software/antivirus/>; for installation instructions, see <http://www.uic.edu/depts/accc/software/nskit/uionline.pdf>.)

The NAV install program turns on Realtime File Protection by default. If you already have NAV installed, select **Start→Programs→Norton AntiVirus→Norton AntiVirus Corporate Edition**, then click **Configure→File System Realtime Protection** in the box on the left. Make sure it's set up as shown.



The A3C Connection

Academic Computing and Communications Center (MC 135)

Room 124 Benjamin Goldberg Research Center

1940 West Taylor Street

Chicago, Illinois 60612-7352

About The A3C Connection

The A3C Connection is published four times per year by the UIC Academic Computing and Communications Center, providing news and information about the use of computers, communications, and networking at UIC. It is edited by Judith Grobe Sachs with help from Bill Mayer and the UIC Office of Publications Services.

Distribution of the *A3C Connection* is free to UIC faculty, staff, and students, and to other universities and not-for-profit organizations. To subscribe, send us your name and address, UIC campus address if possible, including your department name and mail code. To cancel your subscription, send us your address label or a copy of all the information on it.

Contact us by electronic mail at connect@uic.edu; by telephone at the Client Service Office, (312) 413-0003; by US Mail at The A3C Connection, ACCC (MC 135), Room 124 Benjamin Goldberg Research Center, University of Illinois at Chicago, 1940 West Taylor Street, Chicago, Illinois 60612-7352; or by fax at (312) 996-6834.

We welcome any comments, suggestions, complaints, or requests you might have concerning the *A3C Connection*.

The Fine Print

The use of trade, firm, or corporation names in this publication is for the information and convenience of the reader. Such use does not constitute an official endorsement or approval by the University of Illinois of any product or service to the exclusion of others that may be suitable. Trade names that may appear in this publication include the following: Apple, the Apple logo, Mac, Mac logo, and Macintosh (registered trademarks of Apple Computer, Inc.); AIX and AIX/ESA (registered trademarks of IBM Corp.); UNIX (registered trademark of The Open Group); HP and HP-UX (registered trademarks of Hewlett-Packard Corporation); Sun, Solaris, and Java (registered trademarks of Sun Microsystems, Inc.); and Microsoft, Windows, Windows NT, and other Microsoft product names (trademarks or registered trademarks of Microsoft Corporation). All other product names mentioned herein are used for identification purposes only, and may be the trademarks or registered trademarks of their respective companies.

Permission is granted to reprint or adapt all or part of the *A3C Connection* for nonprofit use, provided that full acknowledgment of the source is given.