# Cloud survey

In late 2013 we conducted a survey into cloud usage within the school , this report contains the results and some preliminary conclusions.

The survey was sent out to all staff and all research students (specifically the staff@inf.ed.ac.uk and research-students@inf.ed.ac.uk mailing lists). It generated a total of 53 replies the majority of which were from staff (fig 1)[1]

Breakdown of replies by user type
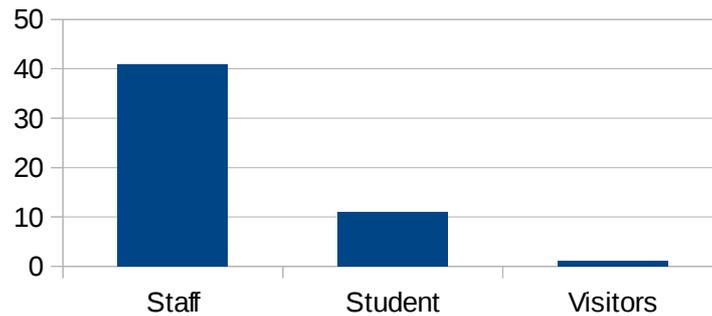
Using roles and capabilities we can get a broad idea of the coverage of responses across the various institutes and groups (fig 2). Note that responders may have multiple affiliations.

---

1   As some staff have the student or degree-phd roles only responders with s<matric> accounts were counted as students.

Breakdown of affiliations of responses

Role

Responses (Responders may have multiple roles/affiliations)
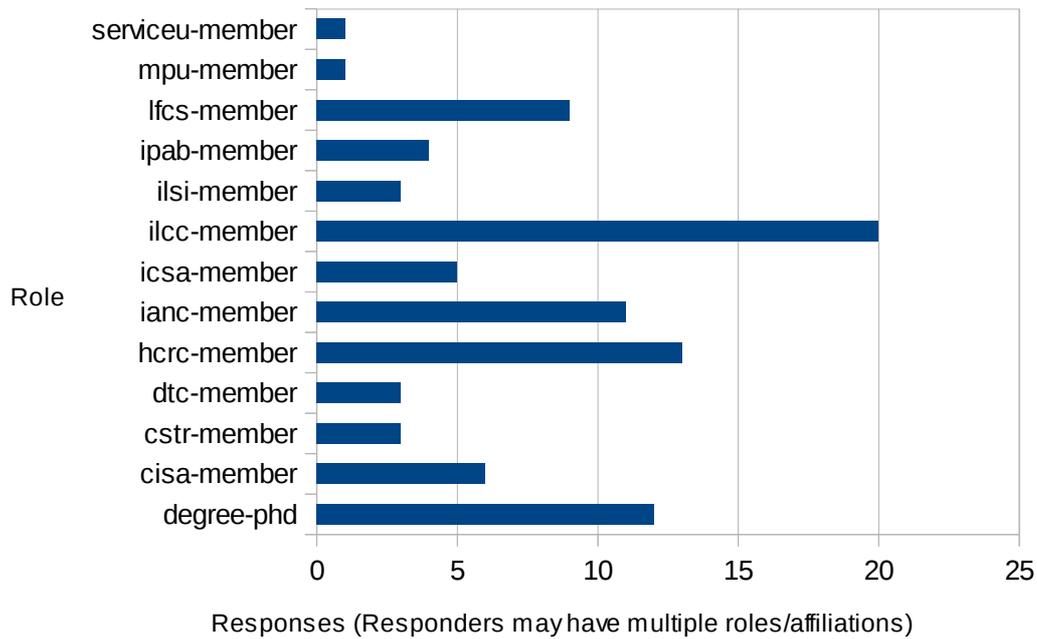
It was hoped to get as broad a response as possible (even if only for some of the questions) and this seems to have been achieved although unfortunately no admin staff and only a limited number of COs seem to have replied.

We also received a query asking if it was appropriate to give a single "Institute response" regarding policies set within an Institute. We hadn't considered this kind of reply and suggested that the responder indicate any organisational policies within the appropriate comments section of the survey.

The results are graphed below. All yes/no/unsure questions are single response returns with "no" being the default option. All pie charts are questions that only allowed single response returns and all other graphs are for questions that allow for multiple responses.
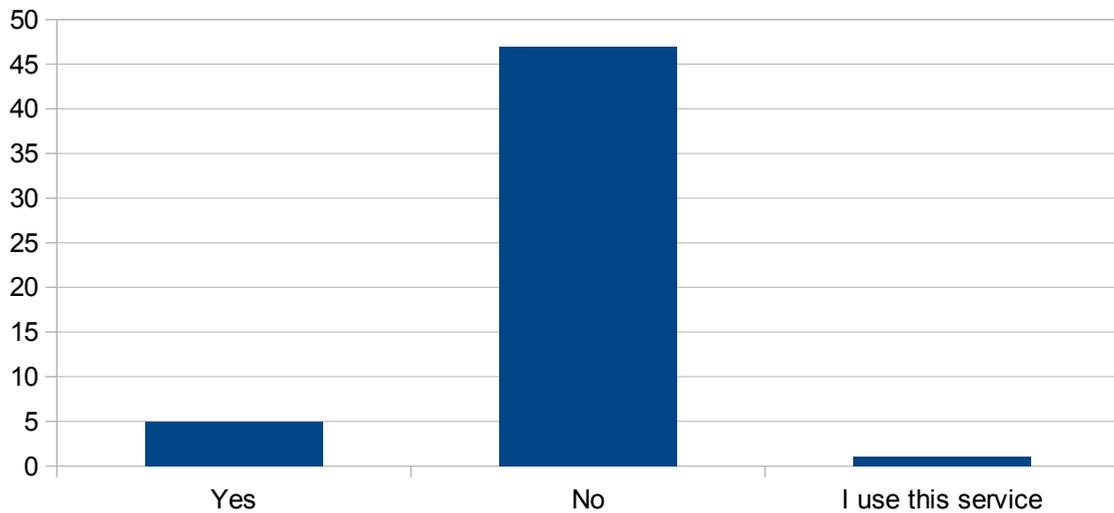
1. Do you currently use VMs to support teaching or research? Please indicate which technology (KVM, VirtualBox, Vmware or other) and what kind of machine hosts the VMs
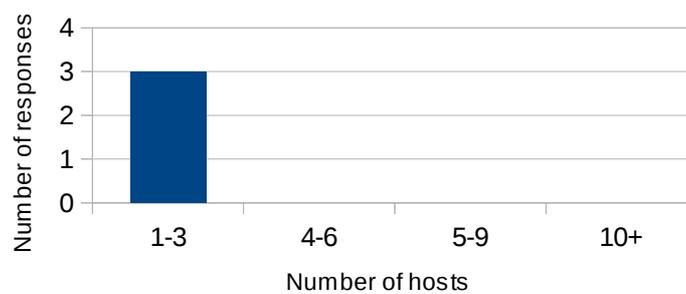
VM Usage by technology and host
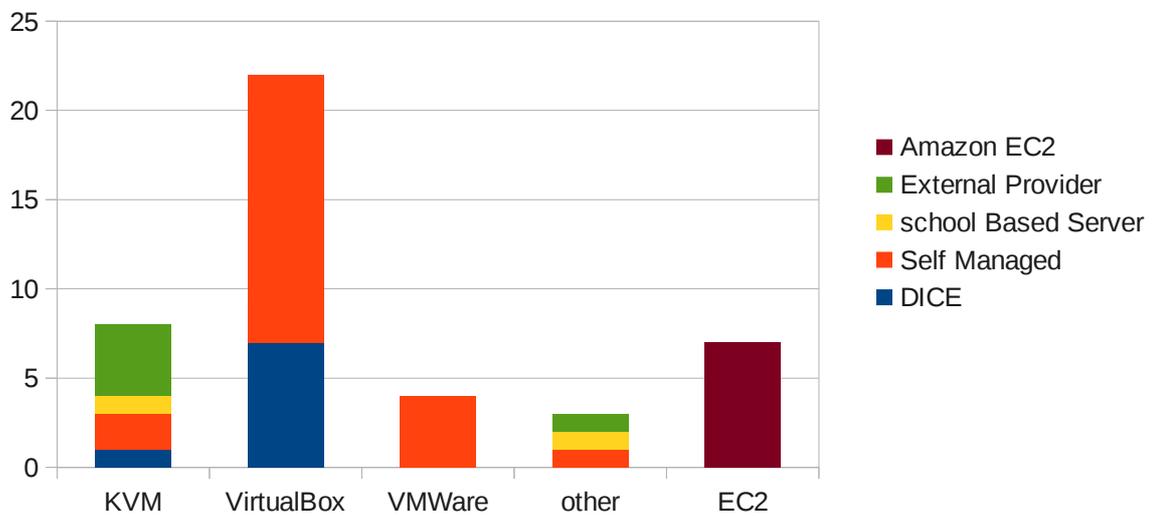
2. IS provide a virtual hosting service for server class machines running on Vmware.
   • Are you aware of this service?

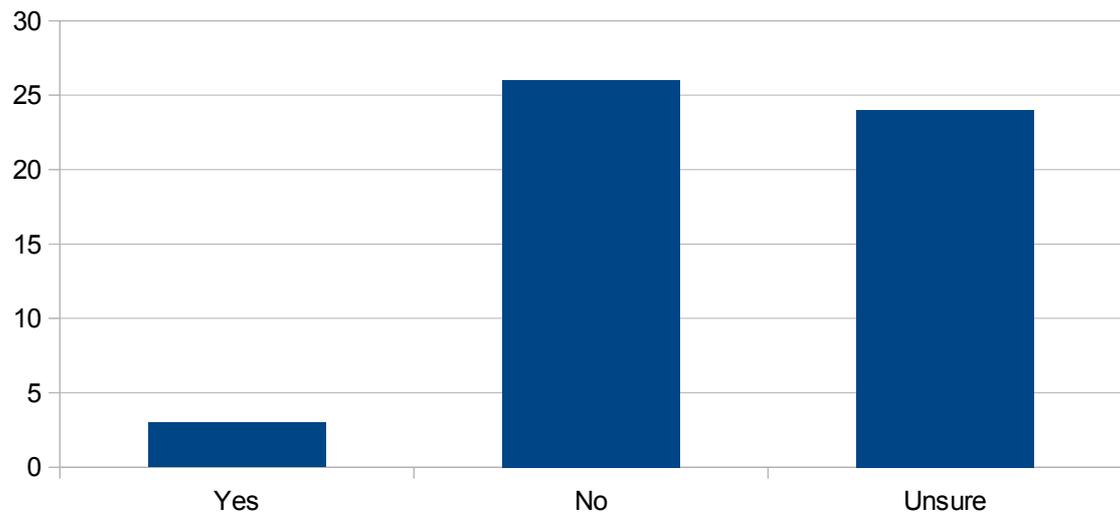## General Awareness of IS service

   • If you use this service how many VMs do you have?

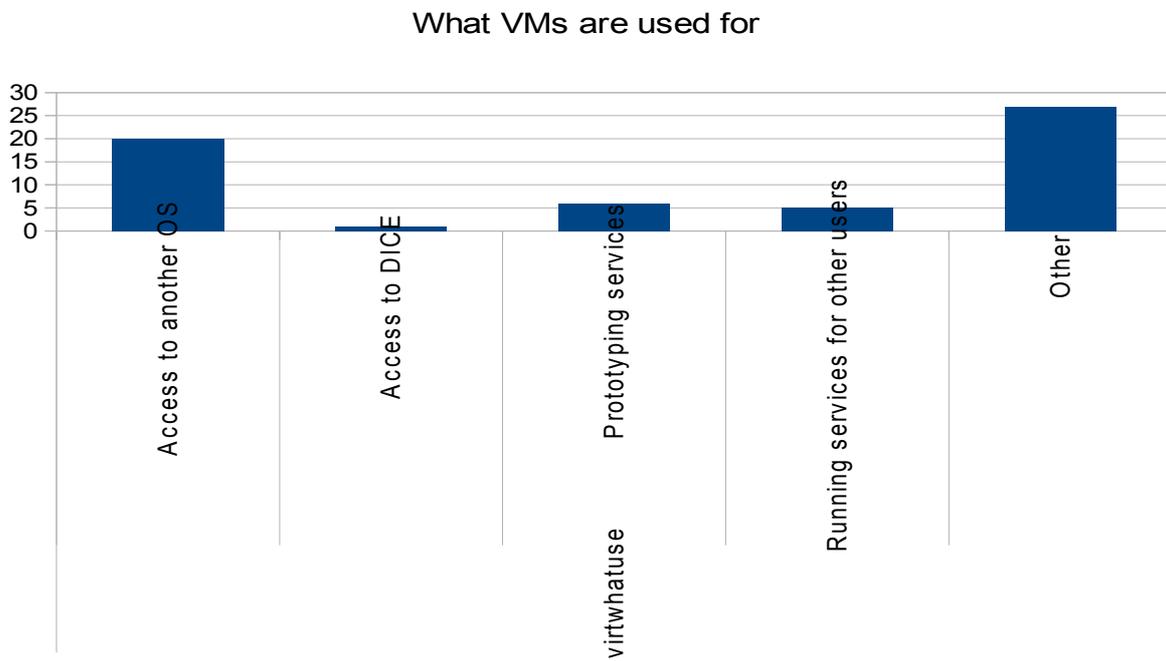## How many IS VM hosts do you currently have

   • Is this a service you are likely to use in the future?

People likely to use the IS service in the future

3. VM use
   • What do you use your VMs for?

What VMs are used for

   • How do you use your VMs?

How VMs are used

   •
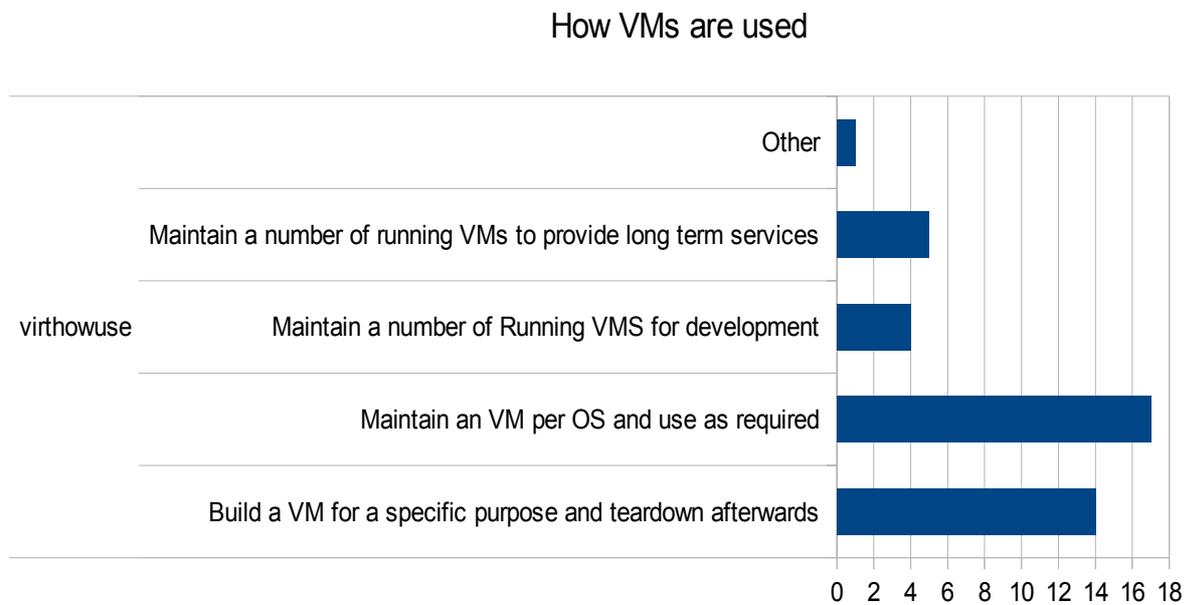
- Which OS do you run on your VMs

Guest OSes
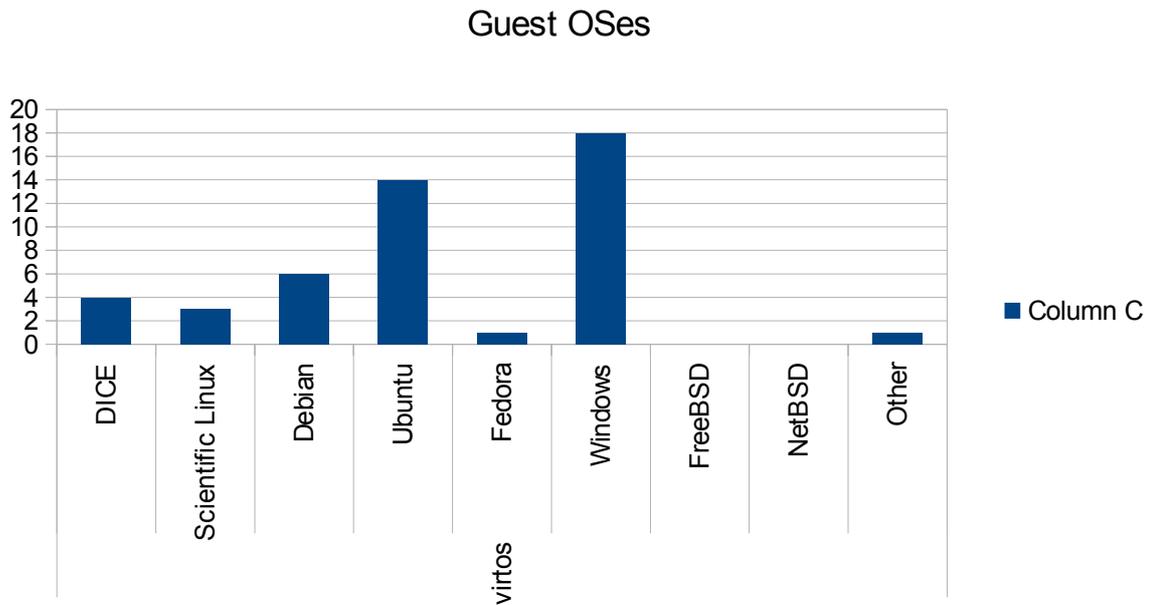
- What, if anything, limits your use of virtual machines?

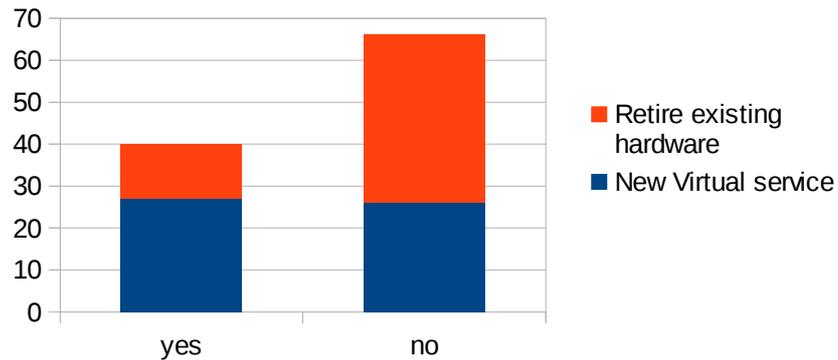## What limits your use of VMs
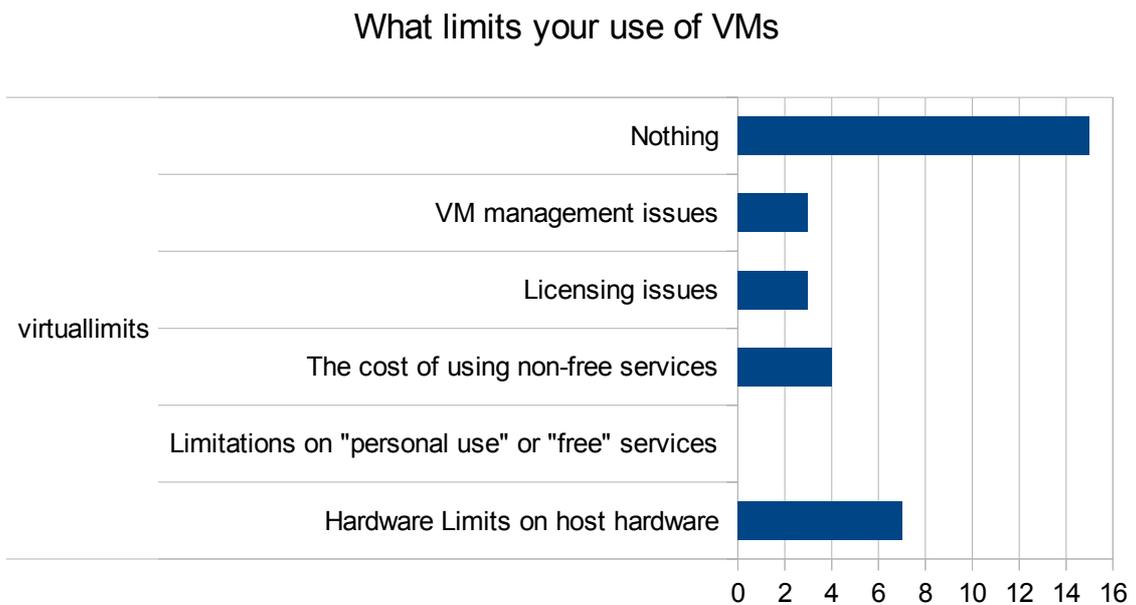
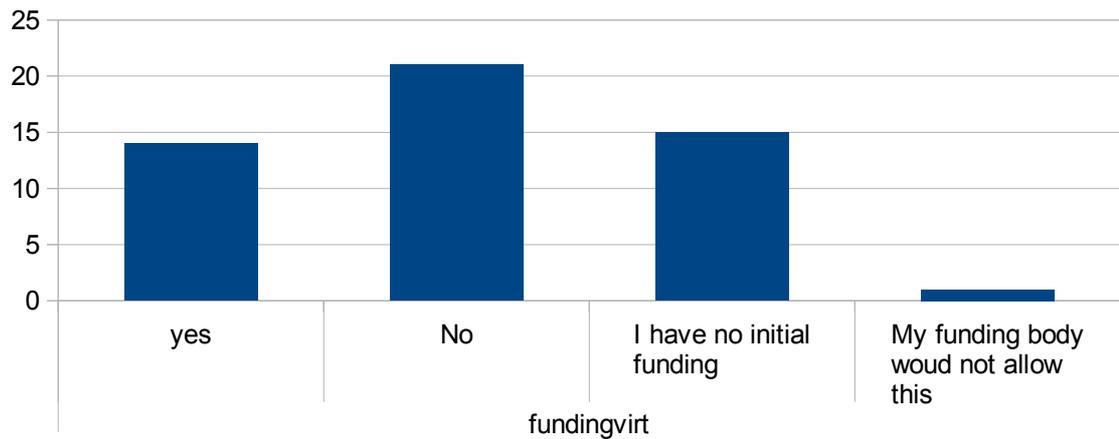4. School Managed Virtualisation service
   - Would you consider using a school managed shared virtualised service?

Interest in setting up new services on VM or retire old hardware

- Would you be willing to pay for a virtual machine rather then buying physical hardware

would you be willing to pay for a virtualisation service rather than buying physical hardware

- Comments

"I like the concept but current university pricing does not feel very competitive - I need to do a detailed breakdown at some point though."

"Personally I've only used virtual machines to quickly try something in Windows. It's not worth paying a monthly charge for that. Very occasionally I want to try something in a clean linux setup, and might try that. Again, the VM is short-lived."

"

I have two completely different use cases ...- Access to software - eg. running Windows - must run directly on laptop- VMs on servers for research"

"Although I'm not currently using VMs I am interested, and I'm certainly pleased to hear about the virtual hosting service. I run a (currently only visible within the University) web service on a research group

owned DICE server. I'd be perfectly happy to run this in a VM instead, and this may actually be a little bit useful. Secondly, I run a large practical course, and I currently allow the students to use any programming language they wish. My one proviso is that their source code compiles and runs on DICE so that I can evaluate it. I currently suggest that if they would like to work from home they may make use of a VM to make sure that they are always DICE compatible. I could imagine making even more use of virtualisation, for example providing a ready-to-go VM which could be cloned."

"Don't really know enough about virtualisation to know what I could do with it."

"I've used an EC2 instance for running machine-learning jobs when all my group's servers were busy."

"I don't know what this is but see something about virtual instead of physical which always makes me scream in horror. I would never accept anything other than physical hardware."

"Only used Amazon EC2 for one project. Otherwise not yet requiring heavy compute power."

"It depends what it was for. I suppose if it were for heavy compute use, I might feel that I wanted my own dedicated server, but if I could be more-or-less guaranteed some level of compute resource on a virtual machine, I would probably be persuadeable."

"I don't know what the School has to offer. It would be good to find out more."
"I use VirtualBox for OS teaching, so students can mess around with OS code."

"I'm counting me running the groups web server and mailing list services on VMs as me supporting research using VMs. Neil."

"I only use Parallels on my Macbook, for testing things. I'd like to have a VM for running a service using Django/Python, set up so that I could mirror it on my laptop for offline development; but it isn't clear that this is currently supported. (It would have to cost nothing.)"

"I have three main use cases: (1) to run things in Windows, including editing Word documents or running prototype software; (2) to experiment with applications and tools without polluting/conflicting with main Linux install; (3) [most important] to support teaching in Computer Security, we have used a range of VMs. Main problem with (3) has been hosting large VM disk images."
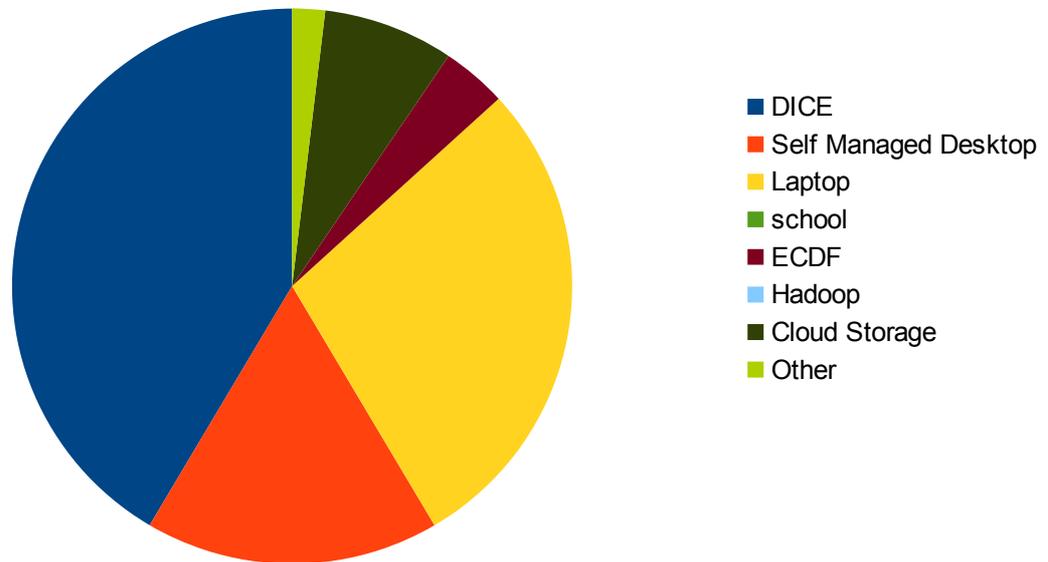
"Use of research partner's VM for production data analysis."

5. Storage
   This section deal with how people are storing data; cloud perception of risks...
   • What do you consider your primary storage for work related data?

Primary storage locations

   • If you did not select DICE Homedir as your primary storage how do you backup your primary storage?

Primary backup method
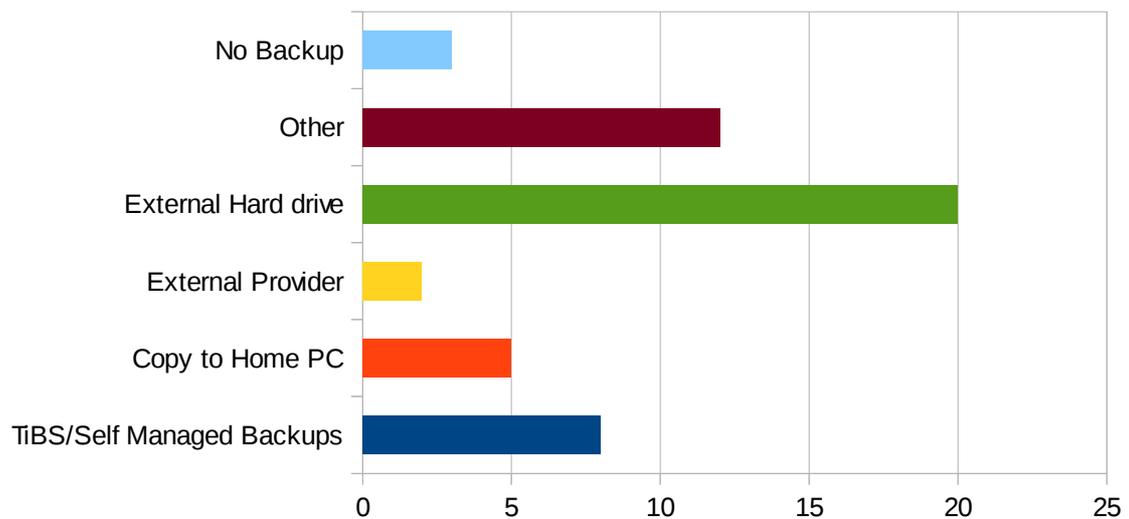
- Where else do you store work data?

Other storage locations

- *Do you use a specific technology to synchronise your data across multiple machines?*

Sync method used between storage and backup

- *If you use an external provider or the cloud do you encrypt files?*

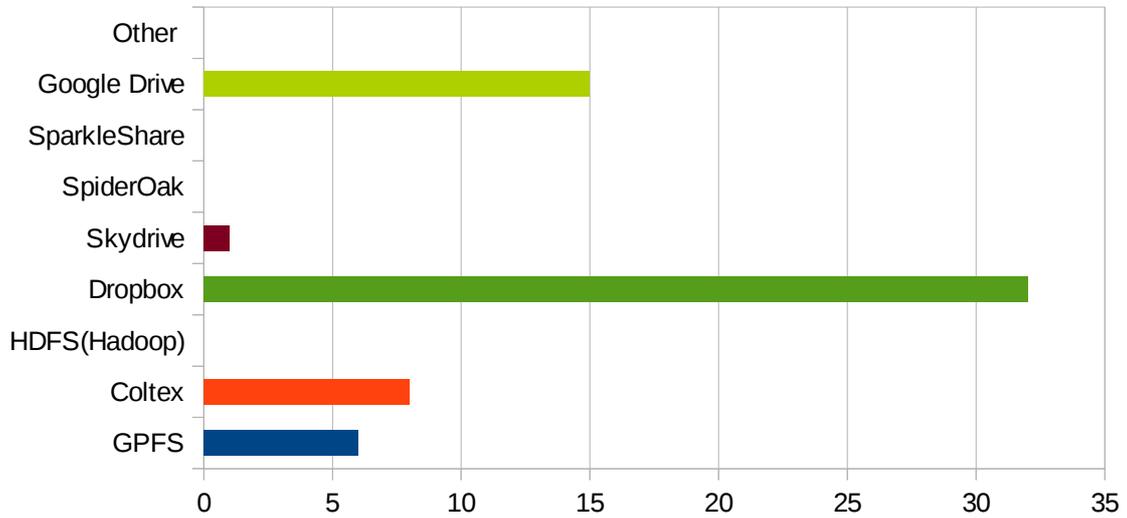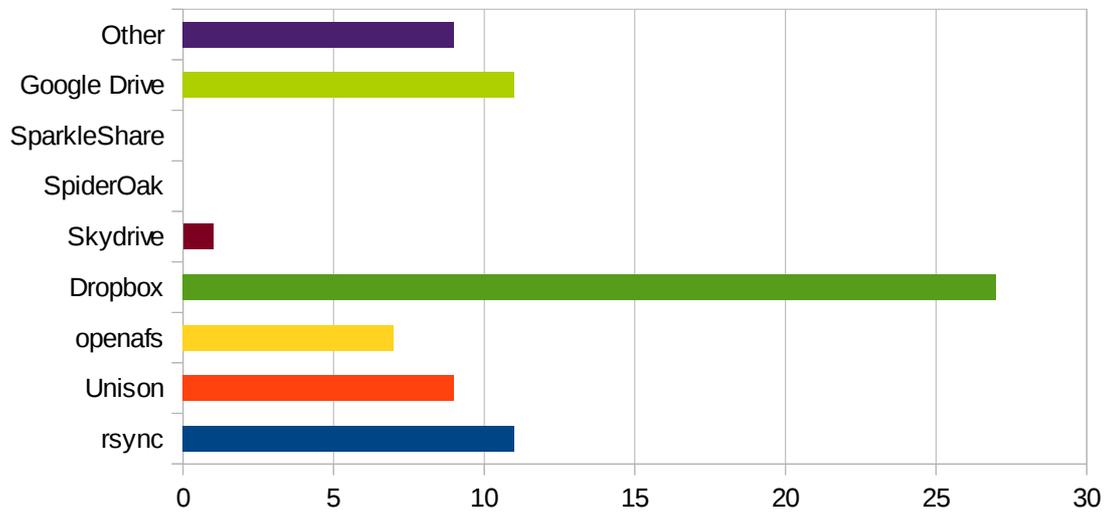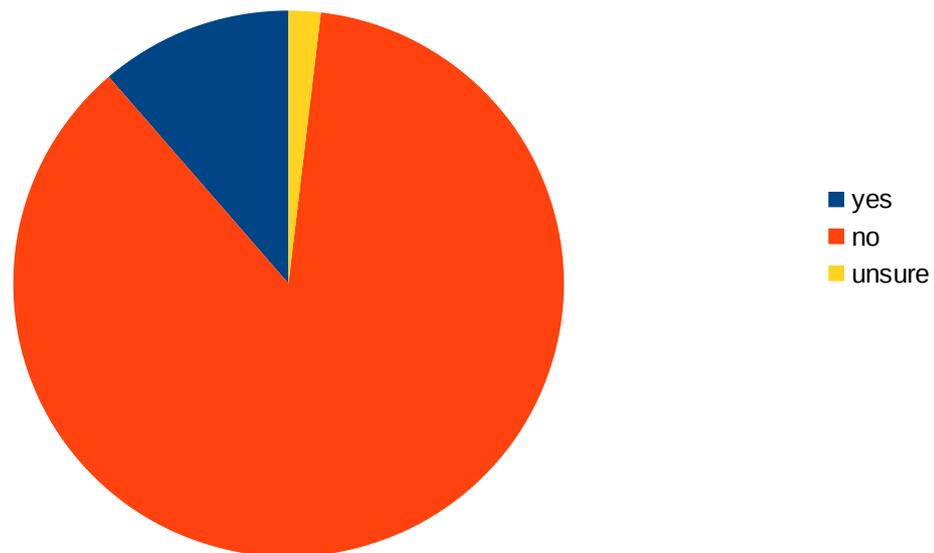# If you use an external provider do you encrypt your files?

Legend:
- yes
- no
- unsure

- Comments

"Auto sync work files and data between three machines with two apple time capsule archives."

"All external disks and laptops use full disk encryption. Backups use rsnapshot. Synchronization with laptop uses unison. Unison in particular I recommend: I've been using it for about 10 years. It's fast, and I've not had any problems. Rsnapshot seems to work well, but is less battle-tested."

"I use Dropbox to share documents with project partners in other schools and at other institutions. We also use ECDF (svn), but Dropbox is easier to use for 'non-technical' people."

"Backup is via TimeMachine to external storage (encrypted), with dual copies at home and at work. I'm not sure that your questions really capture the way in which I (and possibly others) now work - i.e. All my working data is on my laptop, but data which I consider to be "shared" data (projects, teaching information, papers etc.) is in University repositories (ECDF SVN or DICE Git). I do use Dropbox to transfer individual les between my iPad and laptopf, but I don't leave anything there - yes, I know that this probably leaves a copy somewhere on their server - so I have recently been using the University prototype "OwnCloud" service which seems very good (as long as the performance doesn't drop should it be more widely adopted)."

"I use dropbox and googledrive personally, but not for work related data. I find DiCE afs to be pretty satisfactory and the only two reasons I use dropbox and googledrive for personal data (such as photographs) is because it is setup out of the box (for example on my phone and tablet). The other reason is that I am not a permanent member of staff and I recognise that one day I might not be employed here. It just occurred to me that I use github to store source code, both personally and work-related code. I'm not sure if you consider source code to be data"."

"I use ECDF's SVN, which is great for source code and latex (my most important data)."

"My other" is "on my self-managed laptop and my home desktop" but I try with reasonable success to exclude medium and high risk data from being synchronised to these."

"using sshfs to mount directories on different machines"

"My code and papers are on my group's hosted Subversion repository and/or GitHub; papers I've downloaded to read live on Dropbox. This isn't a backup per se, but it does give me some insurance. My actual data mostly isn't very valuable. I do a lot of work on my laptop, frequently in areas with poor connectivity. Hence AFS as a primary working medium doesn't work well for me."

"Just to put myself in the clear here... All identifying participant data for my project ALWAYS remains off the cloud. It is stored in password-protected locations and otherwise handled in accordance with the ethical approval for my project."

"I use dropbox for collaboration with colleagues in other institutions rather than for storage. I also mirror work data on my home machine and laptop. I copy some data from ECDF (Eddie) to DICE, but this is hindered by the small quota of space (10 Gbytes) that I have on DICE."

"Also use DICE (group AFS space) for group data and it works very well. Don't use self-managed backup because I will not select a subset of files on my machine for backup (maybe this requirement has change).... this will take a large amount of time and I will still screw it up. For laptop backup I use two external hard drives, one at work and one at home."

"My coltex use is slightly odd --- I don't run tex there at all, but it's great as a subversion server where I can control shared access and permissions, including with external collaborators."

"GPFS is on ECDF"

"Dropbox uses encryption to transfer and store data but given the latest news it's not clear how secure that actually is."

"Work data is synched between DICE homedir and laptop."

"I use Google Drive and Dropbox for personal stuff, it would be nice (though not essential) to be able to access those natively on DICE, rather than via their web interfaces. (Or some equivalent alternative system)"

"To clarify what I ticked: (1) I store work data on my self-managed desktop and in subversion repositories hosted by ECDF in subversion repositories (not GFPS) and occasionally for sharing/backup on DICE AFS. (2) I access work data on my work desktop, but also other machines including my work laptop, home desktop, mobile devices (for email) (3) I use Dropbox/Google Drive but only for my own data, occasionally for convenience with things like boarding passes for travel (maybe question implies work data, but I do run those services on machines that also have work data on them). I would love to have a Google Drive/Dropbox equivalent that was safe to use for work data."
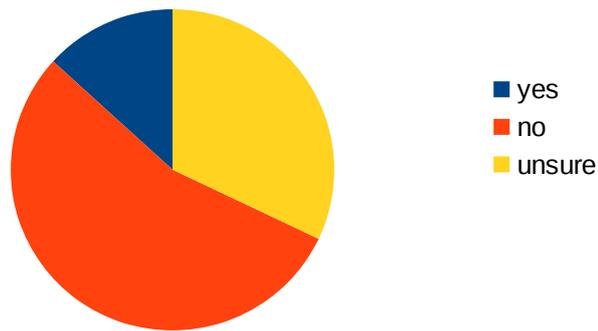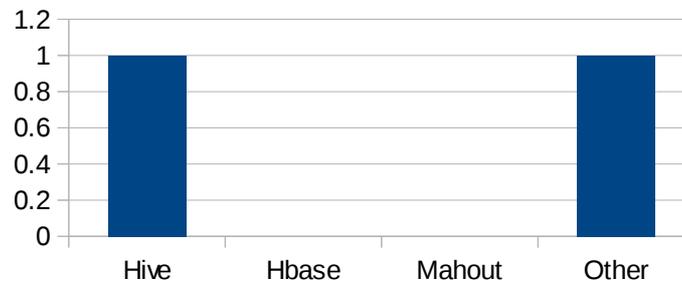
6. Hadoop
   • Do you use Hadoop?

Have you used hadoop
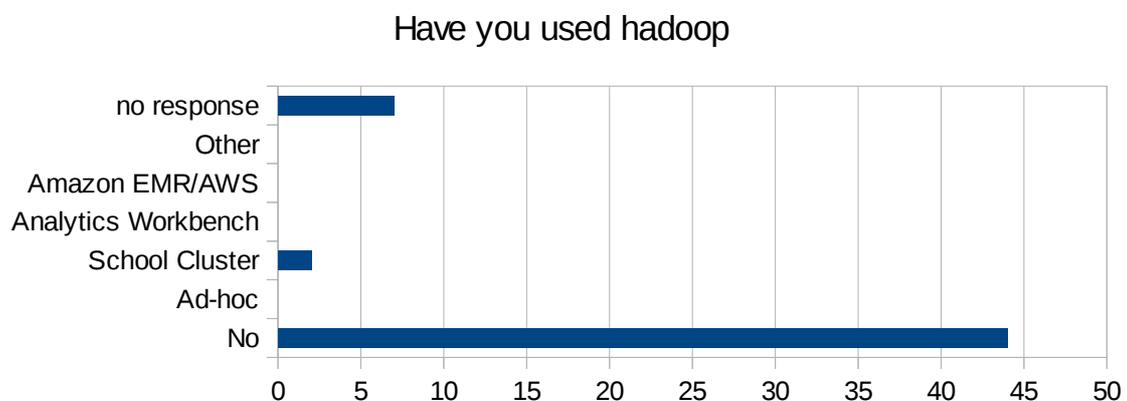
   • *Are you likely to use hadoop in the future?*

Are you likely to use it in the future

- *Have you used any cloud framework that sits on top of hadoop?*

•

Services run on top of hadoop

- *Hadoop Comments*

I've used Amazon EMR and Pig for a Coursera course.
I don't know what this is.

Spark looks very promising: http://spark.incubator.apache.org Some of my MSc students have used the school cluster and found it useful. I would be interested in my PhD students using this where appropriate
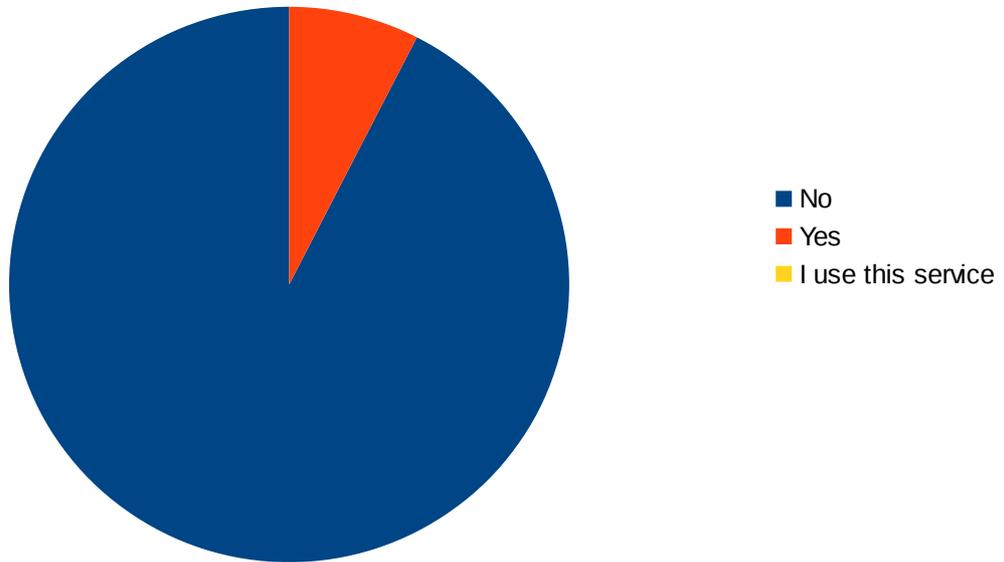
I might use it in future.

I would support the adoption of other distributed computing frameworks, such as Apache Spark (http://spark.incubator.apache.org/)

7. ECDF AWS Pilot
   • Are you aware of the Pilot?

Are you aware of the ECDF Pilot

■ No
■ Yes
■ I use this service

   • Are you likely to use either the pilot or a full service?
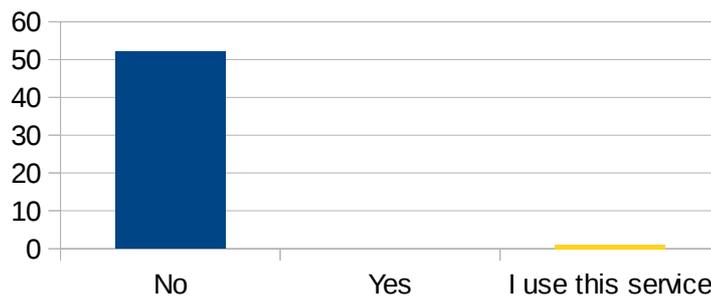
Are you likely to use AWS at ECDF

■ yes
■ no
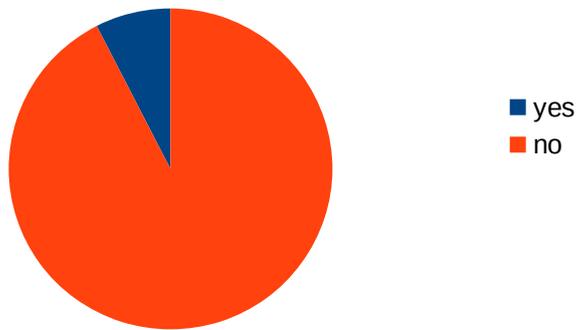
8. Bonfire
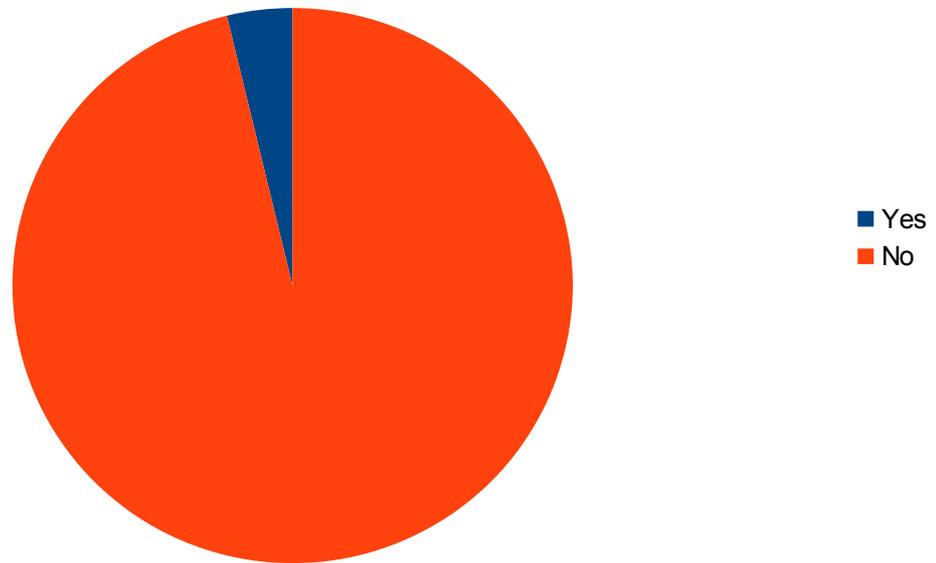   • Are you aware of the BonFIRE service?

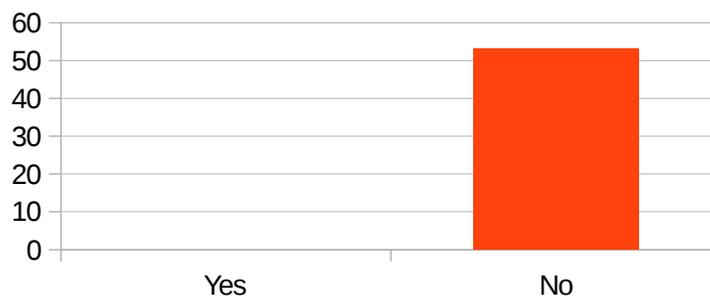## Are you aware of bonfire

• Are you

likely to use it during the free access period?

Are you likely to use Bonfire during the free access period

- Yes
- No

- Are you likely to use the paid for service?

Would you pay to use the bonfire service.

- Bonfire Comments
  I don't know what this is.

9. Platform As A Service (PAAS)
   - Are you currently using Platform as a service type services? (if other then please say what below)

you currently using Platform as a service type servic
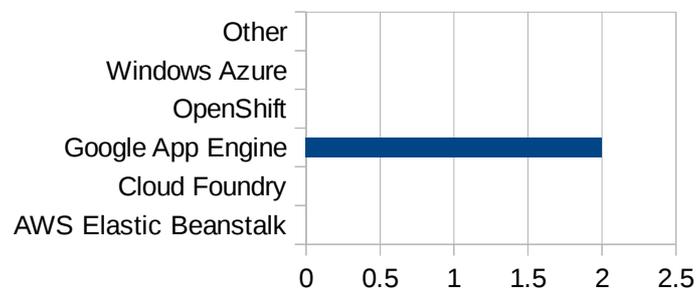
- *Comments*
  "I'd prefer to run my own platform in a virtual machine. However there is certainly some appeal to PaaS."

  "I use Heroku for a side project. It's great."

  "I don't know what this is."

  "Would be unlikely to use a full platform as our needs are very much bespoke... EC2 is a more natural fit for us."

  "I've done some basic experiments with an RA. May do more in future."

10. Software As A Service (SAAS)

- Do you use any SAAS type services?

What SAAS are you using
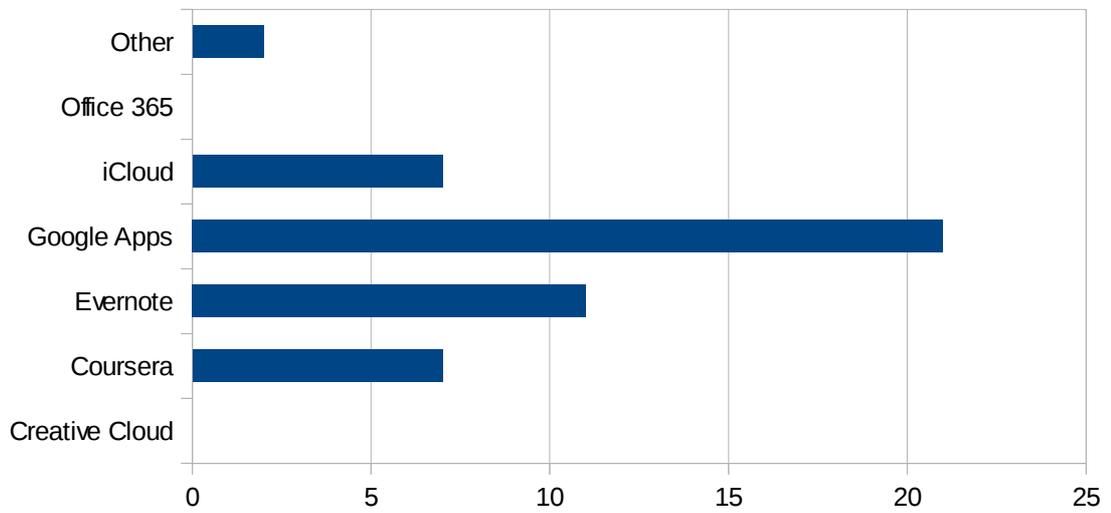
- *Are you likely to use this type of service in the future?*

Are you likely to use SAAS in the future?

- *Comments*
  Coursera? The MOOC? I prefer to have my core applications not depend on the network. I am not happy passing arbitrary data to external services.

11. Remote Desktops
    - Do you use remote desktop software?

Do you use remote desktop software?

- *What do you use it for?*

What do you use it for
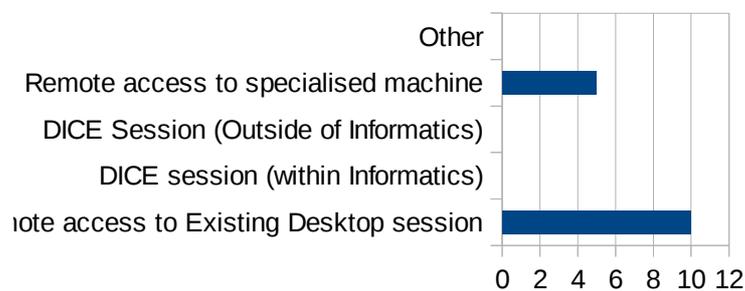
- *Comments*

    "I have been considering an app for remote access to my Mac from my iPad ... "

    "I don't currently use it, but I would."

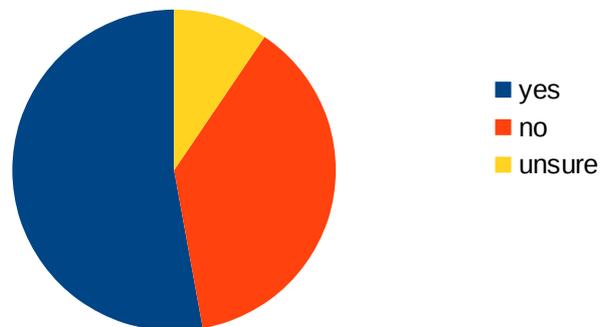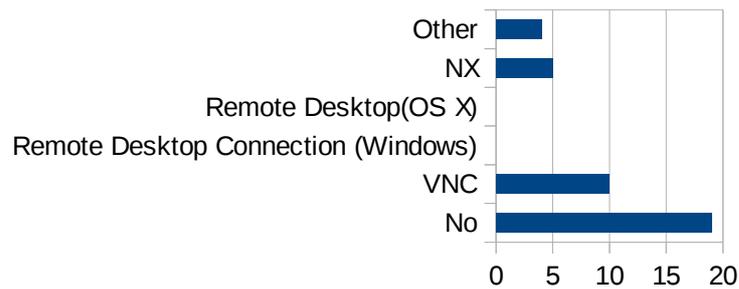   "A few years ago I used remote desktop to access a VirtualBox session from elsewhere in Informatics, but I don't currently need that."

    "I use TeamViewer and Skype screensharing for pair-programming on a side project. Both are flaky."

    "I have no idea what software I use to do this."

    "Call me a dinosaur, but I tend to use ssh and emacs in terminal mode. Occasionally I get adventurous and use tramp in emacs to see files on a remote machine."

    "I used to use that but the tool I used on OS X no longer works."

    "I use x11vnc to get to my actual desktop while remote (usually at home). Recently NX for a new X session, the performance is better, but I miss my actual desktop."

12. Legal Issues Using cloud services
    - When using cloud or other online services how much emphasis do you place on legal terms and conditions?

When using cloud or other online services how much emphasis do you place on legal terms and conditions?
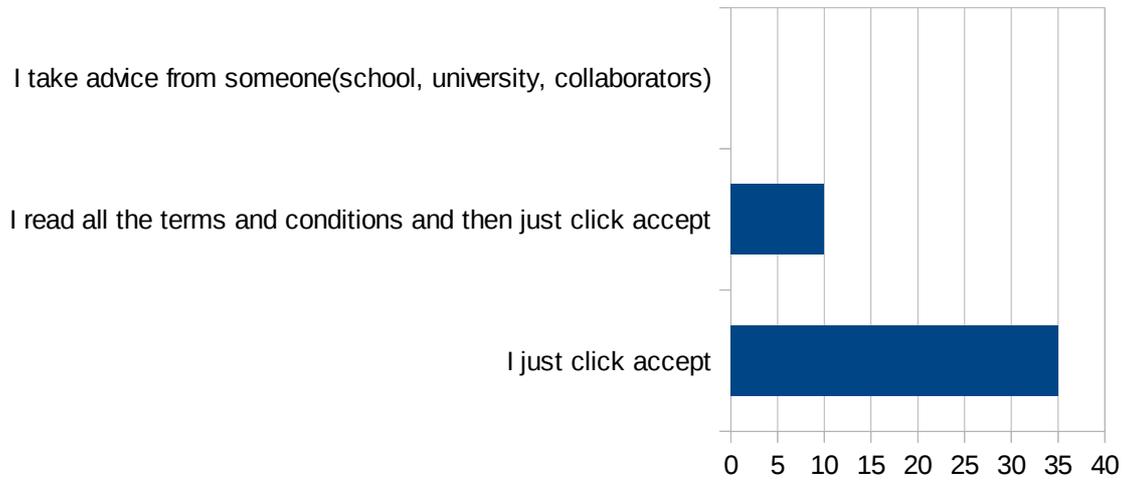
- *Do you consider where your data might be stored when choosing services*

Do you consider where your data might be stored when choosing services

- *Before using a cloud service do you consult the universities policies on Data protection and Intellectual Property*

Before using a cloud service do you consult the university's policies on Data protection and Intellectual Property

■ yes
■ no

- *Comments*

"I do not use any cloud service for data related to teaching and students, only for collaborative research (and my diary on icloud so that I can access it from multiple devices)."

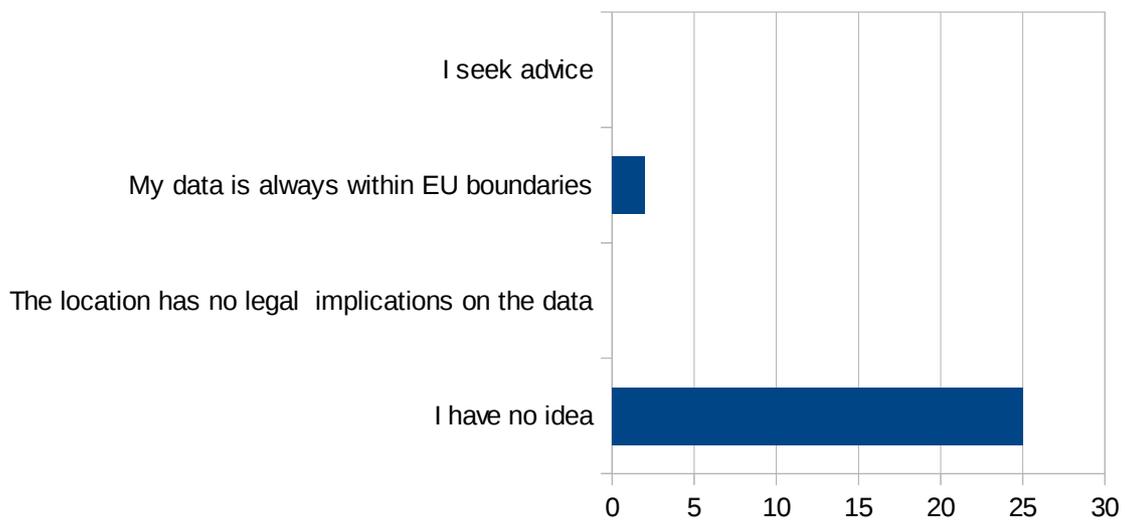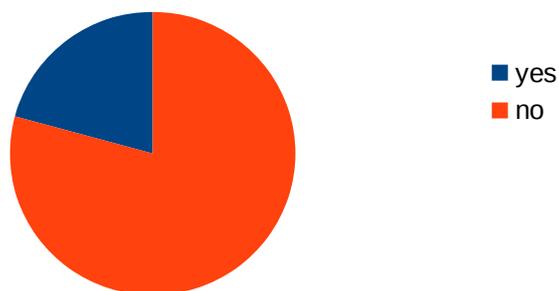"I went through the pain of changing my email work-flow so that University mail doesn't leave the university because of data protection issues. I assume many staff use gmail though."

"I enagage my brain first and don't export data if I consider the data to be sensitive, or the risks to be high."

"I don't have any particularly sensitive data. All work related data is not stored on external cloud but on DiCE afs. My personal data is a bunch of photos etc. that I don't feel are particularly sensitive. If I ever have some sensitive data which I particularly wish to store externally in the cloud I would like to think that I would be more careful."

"I avoid cloud services in general for both work and personal data because of these issues."

"At the moment my use of cloud services is light; in future, I would want advice from school/university/collaborators if using cloud for research work."

"student/sensitive data is stored in AFS and not elsewhere."

"Because I have consulted the university's (sic :-) policies on blah, I don't use cloud services for work stuff. I do use them for my own purposes, and then I just click accept and don't know or care where the data is."

"all the data is on github anyways."

"I have discussed this with my colleagues in Informatics and Education. We concluded that because Dropbox is highly convenient and does not appear to be explicitly forbidden, we will continue to use it for work as long as we take precautions with certain types of data, perhaps choosing to store those sensistive data only on university servers. Again, all identifying participant data for my project ALWAYS remains off the cloud. It is stored in password-protected locations and otherwise handled in accordance with the ethical approval for my project."

"As a matter of principle I don't store sensitive data on cloud services."

"I skim through the conditions, but I don't examine them closely. I'd probably rather not use dropbox, but when one's collaborators use it, it's hard to object."

"I only use cloud services for non-confidential data (e.g., notes about research ideas that would not be meaningful to others, etc). I keep all student data on an encrypted volume (very easy on OS X)."

"I use cloud services personally for data storage and access, but this does not include any sensitive data."

"I suspect I have work-related data on Google Calendar and Google Contacts which is not held within EU boundaries, but does fall under their Safe Harbor" certification. I don't believe I store work email or any medium or high risk personal data or business information on cloud services."

"I only use things like dropbox or googledocs for collaborative projects, essentially for document writing. I prefer svn etc, but sometimes the others want to use cloud mechanisms, which is (AFAIK) easier when you don't all have accounts on the same filespace. I would be very careful about putting confidential data sources outside of UoE filespace."

"I don't use cloud services for work. (Or for personal stuff.) For other random online things, I read T\&Cs if I'm going to be putting anything valuable or confidential there, otherwise just click accept."

"My use of Google Drive, Dropbox is for personal stuff. However photos I take of servers, emergency stop buttons etc, on my phone, do go to my dropbox."

"I am reasonably aware of policies. Generally I try not to store any data which has issues for University wrt DPA/FoI/IP in the cloud."

13. General Comments

"They key thing is to keep up to speed with what people are using and provide appropriate advice on data protection, security etc. We cannot compete with the likes of dropbox so why waste time and money trying? What the limitations are and how to use services like these safety and effectively is important though."

"Speed of change is a problem. It takes a while to learn how to use a service and set something up, then it gets obsoleted by something else. Sorry no solutions! Except to make the management \& admin as simple as possible."

"I wonder whether there's an easy way to provide an encryption wrapper that would make sense? Maybe not."

"I would be strongly, strongly against moving towards SASS for basic applications (at the expense of locally-stored, non-cloud software). Besides fundamentally disagreeing with the idea of software as an unpredictably updated service instead of a finite and purchasable product, some of us can only get anything done when we are NOT ONLINE."

"There was a suggestion at the ANC away day that there should be meetings for ECDF users. I wonder if it might be worth running a workshop for users of cloud services and ECDF within the school. Perhaps there could be short talks from people with experience of using the services, followed by questions. DICE

staff could talk about what's on offer here, and get a sense of people's needs. If a workshop was successful, this could develop into a series. In fact, why not have a regular school computing workshop, where there could be contributed talks and discussion about various practical computing issues? Perhaps get an organising committee formed from spread of people from different institutes and levels (PhDs, postdocs, academic staff) and computing staff? support?"

"Our research is likely to trend more toward big distributed computation and big data."

"When designing this form, make it easy to do a one-click-no to each section."

14. Observations and Conclusions.

- General Observations
    - Virtualisation seems to be a fairly mature technology within Informatics mostly on the desktop via VirtualBox. This seems to be fulfilling peoples needs for desktop type VMs

    - People mainly use it in order to gain access to an additional operating system (windows/Debian based Linux)

    - Awareness of IS services does not seem to be high

    - The newer cloud based services have not made major traction within Informatics

    - As a service dropbox is very popular, probably even more so than we would have thought.

    - Most staff don't let legal implications get in the way of using external services.

    - Are we surprised at the low number of people with no backup?

- Conclusions -Virtualisation
    - We're probably doing enough for virtualisation on the desktop
    - There is demand for a server based VM service (for DICE and self managed VMs)
    - There's sufficient demand for a paid for service if some way can be to run it.

- Conclusions Cloud (hadoop/PAAS/SAAS)
    - There is no widespread demand for a large scale hadoop/cloud type installation in the school but I believe there is interest in maintaining a smaller cluster which could be used to look at emerging technologies (Apache spark)

- Conclusions Storage
    - If we don't want people to use Dropbox we're going to have to point them at something which is as convenient and easy to use.

- We ought to provide some guidance on encrypting files that would mitigate inappropriate material leaking out into the cloud..