

IRU Computational Lab: Intro and Visions

Dr. Andreas Krueger
IRU CompLab day 26/10/2010

Your questions are probably:

1. What will be the **WORKFLOW**,
once the Computational Lab is ready ...
2. What could you already start to
LEARN now (marked everywhere in red)
3. **Visions** of an innovative lab platform
4. Current **news**:
 - Videoconferencing
 - New Hardware, Software, etc.
5. What is already there now. **How to use it.**

ABM = Agent Based Modeling

- SKIN model: **S**imulating **k**nowledge dynamics in innovation **n**etworks, google: SKIN Ahrweiler
- **NetLogo**: google NetLogo → click: “web”
- **Java**: Michel’s implementation of **SKIN** code
- ABM at IRU: Michel Schilperoord, Petra Ahrweiler
- Example workflow: perhaps next IRU day?
- Barry & Tara (CASL) want for the CASL webpage:
 - OpenSource Computer Code, Videos, Pictures

Complex Network Visualisation Example Workflow

1. Publication Data

1. Download data: by **Python** script; from ISI web of science
2. Clean data in **Excel**: manually, semi-automatic
3. Connect-the-dots: **Python** script to generate **Pajek** file
4. Visually inspect, publication pictures [gray]: **Pajek**.
5. next step: 2nd database, [gray-gray] becomes [green-green]

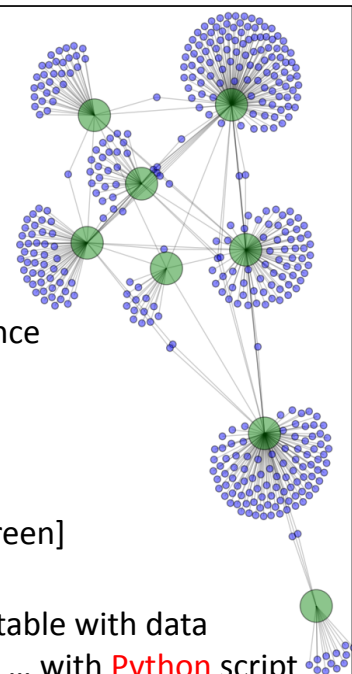
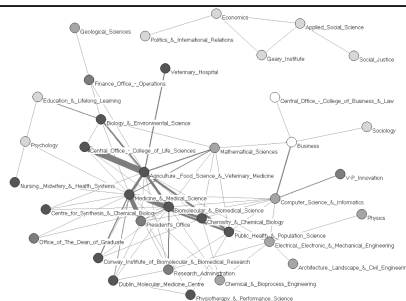
2. UCD Industry contacts [blue]

1. Grants DB: lots of meetings, **privacy** issues ... → **Excel**-table with data
2. Clean, standardize with publication school names, etc. ... with **Python** script
3. Connect [green-blue] *by same school name* ... with **Python** script
4. Restrict to 8 nanobio schools
5. Visualize in nice design with **NetworkX** (Python Library)

Wikipedia & Google: **Pajek/Visone, Python** ← begin to **study manuals & use them**

Privacy & Networks-Introduction: tiny.cc/DrAK

“gray” pic from: arxiv.org/abs/0908.4506



CompLab: A multi-level challenge

in a *very rough* schematics:

1. **Hardware:** Quotes, Choice, Purchase, Setup, etc.
2. (Three) Operating Systems (**OS**) plus Firewall
3. **OS Configuration:** e.g. OS X Server - DHCP, DNS-name the machines, etc.
[still lots to do on that level!]
4. **AAA** = Authentication Authorization Accounting:
Users & Passwords, Access Rights, Home Folders, VPN, etc.
5. **Application** Software (e.g. Word, MindManager, BookPedia, Citation Storage, Project Planning, Python Interpreter, etc.), Database Engine (MySQL), Remote Desktop Services, Parallelization demons, ...
6. DataBase (DB) **architecture**, tables, relations, access rights, etc.
7. **Middleware** code - (loosely) connecting everything: Huge Task
8. **Simulation** code
9. Input **Data**, configuration, policy scenarios, etc
10. Simulation **Runs** (ideal: automatically distributed to idle machines)
11. **Results:** Output, Diagrams, Visualisations, etc.
12. - 23. **Scientific** Hypotheses & Examination Strategies & Publication & Funding

Visions: a possible IRU platform in 2012

- Empirical Data **DB** - all IRU data in one data warehouse
 - e.g. ext: Eurostat, int: nanobio
 - Specified access rights for certain user groups
- **SQL** commands **DB**: simple & complex queries stored for *all* users
 - FROM nanobio.science, nanobio.firm SELECT author, company WHERE company.turnover > 100000
 - Index | Name | SQL query | description | sample output | inventor
- Case **DB**: Scientific Hyptheses (and test results?)
 - Q1→A1, Q2→A2, Q3→A3, ...
 - which data, code, configs, rules, etc
- Workflow **DB**:
 - Analysis and Simulation procedures
 - → SNA → | → ABM → | → SNA → ABM → | → SNA → ABM → SNA → | etc.
- Code Config **DB**:
 - outsource variables & decisions *from code* into DB !
 - more versatile code modules, and even configurable for dummies
- Results Table **DB**: intermediate & final results
- Visualisation **Modules**:
 - Network Pictures, Diagrams, Tables, GIS
- **Dashboard** & Text Output
 - (individually configurable) Dashboard with input, configuration, simulation, analysis, output
 - TeX Report-
 - Generator Machine (“new empirical data ... click click ... → new figure in publication PDF”)

Visions: a possible IRU platform in 2012, ctd.

- Some necessities:
 - SQL server. Consistent DB architecture.
 - **SQL query** formulation knowledge - every scientist!
 - Good choice of *existing* platforms
 - Lots of *new* middleware code
 - Modelling expertise
 - Scientific simulation background
 - Database architecture expertise
 - Programming expertise
 - Usability orientation
 - Database programming
- Tacheles:
 - 1 year minimum
 - 1 scientific **modeler** & data warehouse **architect**: *natural sciences* training, **large programs** track record, simulation experience (and social sciences affinity)
 - 1 coding programmer with computer science education, **fast & reliable coder**; **ideally also with hardware- and OS-knowledge**, and scientific modeling ideas

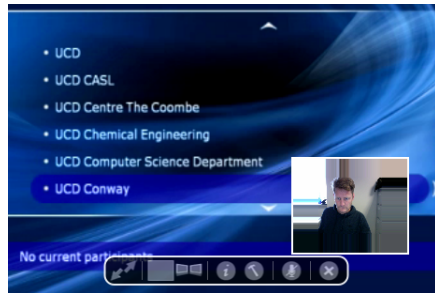
Current News in IRU CompLab

- Video conferencing
- 5 more machines in simulation cluster
- Paragon: NTFS for Mac, HFS for windows
- 2 screens for students
- VMware workstation
- 3 external 2.5inch HDs (instead of USBsticks)
- Dictafon
- SmartPen
- Lightscribe Burner

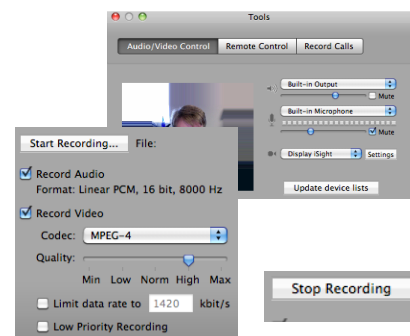
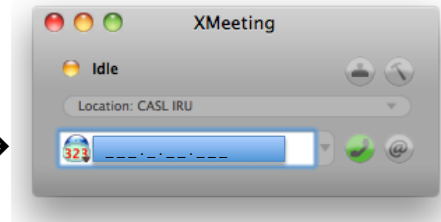
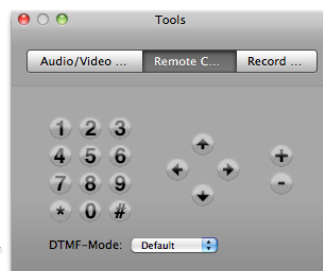


IRU VideoConferencing

- <http://XMEETING.sourceforge.net> (mac) or **NetMeeting** (win): <Windows+R> then **conf.exe** <enter>
- Download, run Installer, copy to Applications-folder
- Start, Click through **all defaults** to define location. IP = _____.____.____.____
- To start a conference, IP = _____.____.____.____ and green button →



use “**Remote Control**” to command the conference server, e.g type “**___#**” to access room “**IRU**”



Conferencing, how to connect:

The Virtual Conference Room is called **IRU** and is **room 213**.

Dialing using an IP address:

Dial _____.____.____.____ and enter roomnumber _ _ _ followed by **#** key.

Dialing using ISDN VC unit or telephone:

Dial **01 4490889**

(adding **+353** if connecting from outside of Ireland)
and then dial shortened version of the GDS number which is

7 _ _ _ followed by **#** sign

Dialing using GDS:

Dial **003530110051 _ _ _**

Important Elements of the CompLab

- IRU LAN
- Structure of our network
- Simulation cluster
- Virtualisation
- Graphics Software
- AAA Server

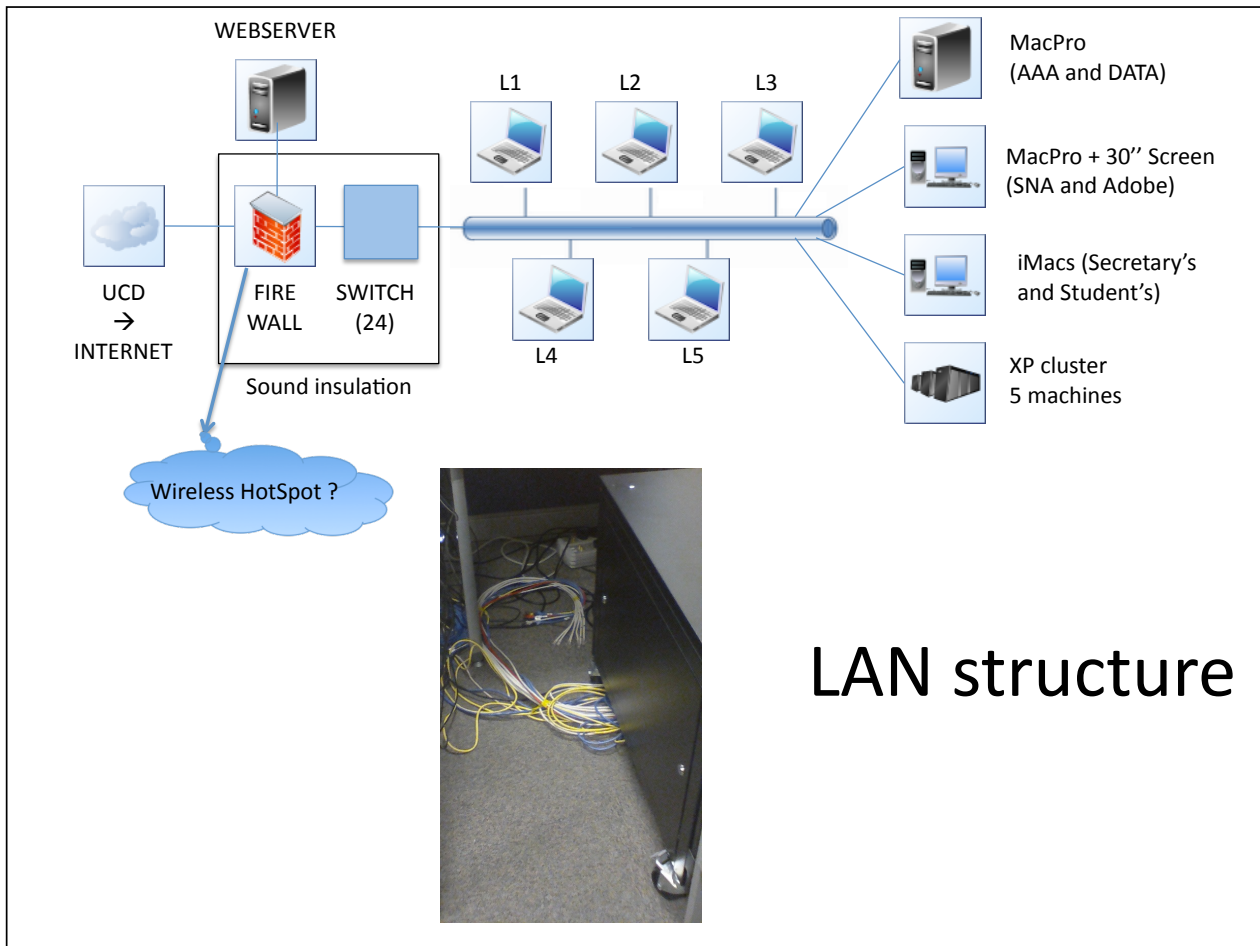
For details, please see past presentations at

AndreasKrueger.de/innovation

We have our own LAN

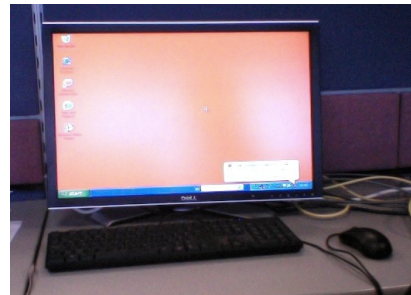
- *All our* machines are now connected to our switch behind our firewall.
- If ever you suspect problems, just use the UCD cable (gray).
- Do NOT connect to *both* networks at the same time (e.g. Ethernet LAN + Wireless UCD) !
 - we were warned that *you* then might power down the whole CASL network!
 - and you would be a security risk (loop around the firewall)
 - → **Switch off Wifi ! Your responsibility!**



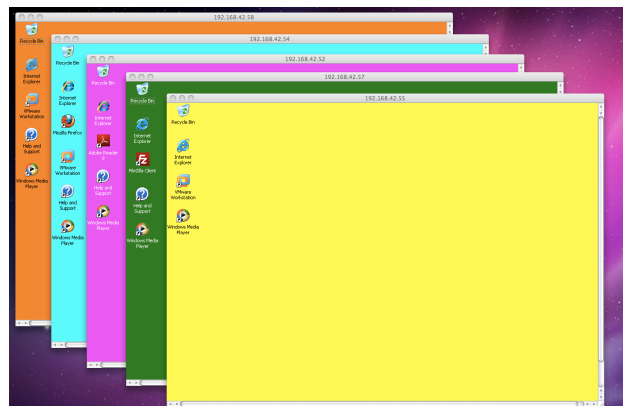
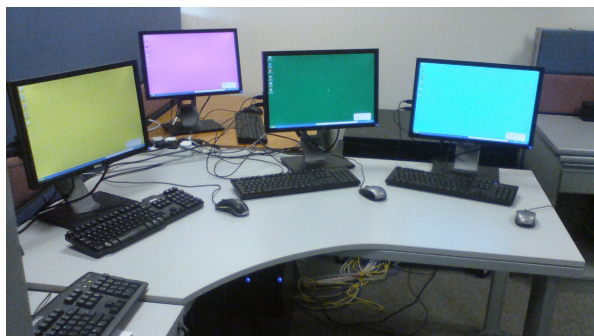


Simulation Cluster

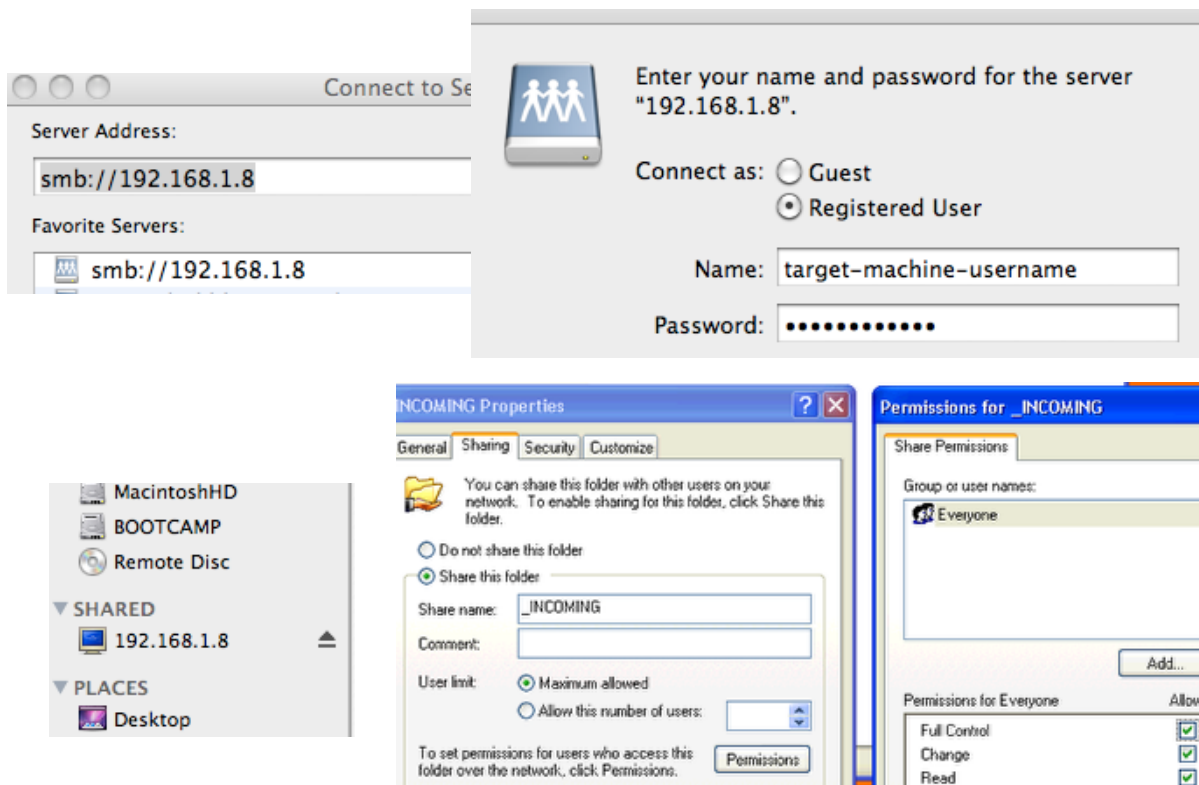
- 5 PCs (Core 2 Duo 2.8 Ghz) Windows XP
- → we can do 5 experiments in parallel
- → we can speed up simulations by factor 5
- → or we can increase the accuracy of the results



- Use "Remote Desktop Connection" to access the machines
tiny.cc/RDCmac Download for Mac
- by their IP-address (192.168.XX.____/____/____/____)
- by ORANGE BLUE RED GREEN YELLOW

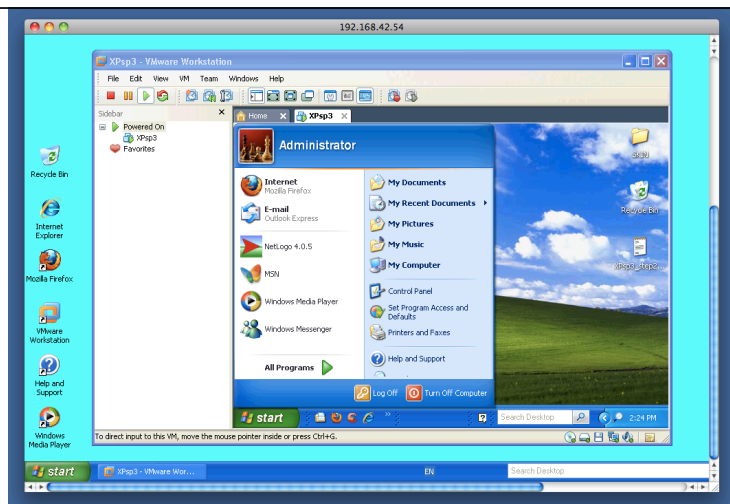


Shared Windows Folders on Mac



Virtualisation

- The XP hardware machines are only “hosts” for vmware “guest” systems
- “Host”= hardware
- “Guest”:
Emulated hardware
- All *your* software is installed into the **guest** images (guest is also also XP)
- Every researcher keeps his own guest image copies
- VMware workstation & VMware player
- **Discussion result (“how to make everything easier”):**
It is a damn good idea to work inside such images
- but not a *necessary* step to do the research



Adobe Software



Creative Suite 5 Master Collection:

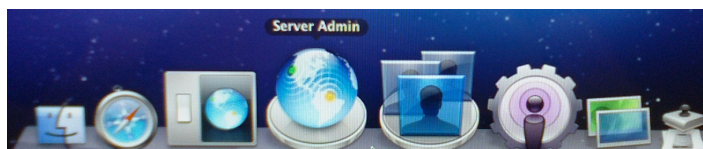
Photoshop, Illustrator, InDesign, FlashCatalyst, FlashProfessional, Dreamweaver, Fireworks, AcrobatPro, Bridge, DeviceCentral, FlashBuilder, Contribute, PremierePro, AfterEffects, Soundbooth, OnLocation, Encore

On MacPro with Samsung
2560 * 1600 pixels = 4 MegaPixels Resolution
→ 1.8 times the Apple Cinema Display Pixel



OS X Server = DATA-Server and AAA

- **AAA:** Authentication, Authorization and Accounting
- **Data storage:** Your home folder centrally → backup.
- RAID data security (2 disks as 1)
- DNS for computer names, internal WikiWiki, mySQL DB, Address book server?
~~iCal server? ...~~
- 2679 pages of manuals ☺
- work in progress ...



What I will do next

- Write documentation manual about the CompLab
- Publish paper about
“An unexpensive and versatile DIY Computational Lab”
- Finish (some of the) above jobs - as far as possible
- KVM switch = **K**eyboard **V**ideo **M**ouse:
only 1 keyboard/mouse/screen for 10 machines
- 24port switch is almost full. Buy new one? Rearrange existing!
- Wireless Access Point (for IRU use & guests)
- Server:
 - ✓ – RAID = Redundant Array of Independent Disks; RAM = 10GB now!
 - AAA: Users and Groups, centralized Login, VPN, etc.
- **We need support with some tasks, students must help!**

Summary – what to learn (**first**, **later/also**)

- (Excel)
- **SQL !** → Structured Query Language
- **Python**, (NetworkX) → programming, scripting
- NetLogo, **SKIN** → ABM simulation
- (Pajek), **Visone**, (gephi) → SNA Tools
- **Java** → programming
- Additional Software for the whole research process

Red = **Strongly suggested**

Green = **Suggested** (choose one of them)