Effective Communication: How to Talk to Researchers about Their Research

Henry Neeman, University of Oklahoma
Director, OU Supercomputing Center for Education & Research (OSCER)
Assistant Vice President, Information Technology - Research Strategy Advisor
Associate Professor, College of Engineering
Adjunct Faculty, School of Computer Science
Advanced Cyberinfrastructure Research & Education Facilitators Virtual Residency Workshop 2017
Monday July 31 2017

Outline

- How to Talk to Researchers: Research Terminology
- Researcher Types
- The Mindset Gap
- Things to Say to a Researcher
- How to Find Researchers
- How to Find Researchers' Projects

How to Talk to Researchers: Research Terminology

Is Oxygen a Metal?

How many of you believe that oxygen is a metal?

Oxygen in Real Life

- Atomic number 8

- Chalcogen

- Key element in life

- Also fire, rust

http://en.wikipedia.org/wiki/Oxygen

Oxygen in Astronomy

- The universe is made of the following:
  - Hydrogen
    - Atomic number 1
    - 75% of all baryonic mass
    - Most stars are made of hydrogen plasma
  - Helium
    - Atomic number 2
    - Noble gas (inert)
    - 24% of total elemental mass
  - Other: ~1%

http://en.wikipedia.org/wiki/Hydrogen
http://en.wikipedia.org/wiki/Helium

Planets etc

What are planets made of?

- Cores of iron, nickel etc
  - Earth's core is 89% iron, 6% nickel, 5% other
- Mantles of silicates

http://en.wikipedia.org/wiki/Planets#Mass

http://en.wikipedia.org/wiki/Earth

Rose Eveleth, "Barns Are Painted Red Because of the Physics of Dying Stars."

http://www.smithsonianmag.com/smart-news/barns-are-painted-red-because-of-the-physics-of-dying-stars-58185724/?utm_source=keywee-facebook.com&utm_medium=socialmedia&utm_campaign=keywee&kwp_0=283306&kwp_4=1091891&kwp_1=506963

So What's a Metal?

- To a chemist, metals have a very specific chemical definition.
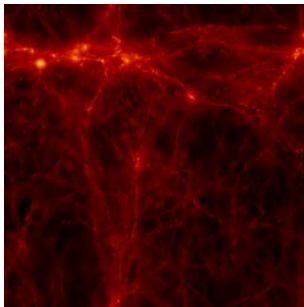
http://user.astro.columbia.edu/~gbryan/Site/IGM_files/gas_density_z0.png

http://images.tutorcircle.com/cms/images/44/periodic-table11.PNG
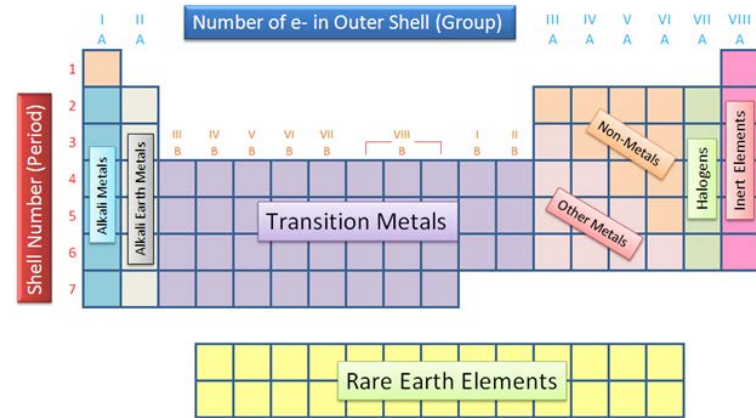
- To an astronomer (especially a cosmologist), metals are anything that isn't hydrogen or helium.

Projection

- What happens if you put a mathematician, a psychologist and a movie producer into a room and ask them to discuss projection?

What Are Fluids?

- <u>Colloquial definition</u>: Liquids.

- <u>Mom's and physician's definition</u>:

  Something you should drink plenty of when you're sick.
  <u>https://www.zocdoc.com/answers/9591/does-drinking-fluids-help-when-you-have-a-cold</u>

What Are Fluids? (cont'd)

- <u>Physical science & engineering definition</u>: Not solids.
  - Computational FLUID Dynamics
    - The most popular fluid studied is air (Earth's atmosphere).
  - "[A] substance, as a liquid or gas, that is capable of flowing a nd thatchanges its shape at a steady rate when acted upon by a forcetending to change its shape." -- dictionary.com
  - Liquids are ***<u>incompressible</u>*** fluids.

Scale

- At quantum scale during femtoseconds, how much does gravity matter?

- How about at cosmological scale over eons?

CS or IT?

- What happens if a domain scientist refers to CS as IT?
- Wait, CS people do research? I thought they were just there to help everyone else with their real research.

Is Simulated Data Actually Data?

- I had a colleague in Chemical Engineering who told me that, if he referred to data from a simulation as "data" in front of his colleagues, he'd be laughed out of the discipline.

Science vs Engineering

- Science is focused on discovery.
- Engineering is focused on design.
- In which case:
    - Is a design project research?
    - Do engineers do science research?
    - What is research about software?

Researcher Types

Researcher Types

- Faculty
  - Tenure-Track Faculty
  - Tenured Faculty
  - Research Faculty
- Staff
  - Postdocs

Tenure-Track Faculty

At research-intensive institutions:

- **<u>Incentive Structure</u>**: I need to publish lots of papers, bring in lots of grant money and graduate lots of students, or I'm fired.

- **<u>Need</u>**: I need stuff to work now and keep working reliably.

- **<u>Timeline</u>**
  - I have 7 years (typical tenure-track duration), **BUT**
  - I have 6 years (the 7th year is finding a job elsewhere if I don't get tenure), **BUT**
  - I have 5 years (the 6th year is when my materials are evaluated), **BUT**
  - I have **<u>4 ½ years</u>**, because it typically takes a journal article about 6 months from being submitted to being published.

Tenured Faculty

At research-intensive institutions:

- **<u>Incentive Structure</u>**: I need to publish lots of papers, bring in lots of grant money and graduate lots of students, or I don't get a raise and I don't get a named chair.
- **<u>Need</u>**: I need stuff to work now and keep working reliably.

Research Faculty

- If I don't bring in grant money, I'm laid off.
- I need to publish a lot to keep bringing in grant money.

Postdocs

- I need to publish a lot or I'll lose my postdoc.
- I need to learn how to get lots of grant money, and even actually get some of my own, so I can get a permanent position.

Students

- My first goal is to graduate.
- Anything that delays graduation costs me money:
  - I may or may not have an assistantship.
  - While I'm in school, I'm giving up that many years of salary and benefits.

Probability of Success

- National Science Foundation, FY2016: 24% overall
  - BIO 26%, CISE 23%, EHR 21%, ENG 20%, GEO 31%, MPS 26%, SBE 24%
  - EPSCoR jurisdictions: Northern Marianas Islands 0% (no PhD-granting), Guam 0%, NV/SD/WY 16%, AL/ND/WV 17%, AR/KS/MS/SC/VT 18%, NE 19%, ID/LA/**OK** 20%, KY 21%, NM/PR 24%, ME 25%, NH 26%, DE 27%, HI/MT 28%, US VI 30%, AK 31%, RI 38%
  - Non-EPSCoR jurisdictions: FL 18%, TX 20%, MO/OH 21%, AZ/UT/VA 22%, IA/IN/MI/NC 23%, GA/NY/TN 24%, CT/IL/NJ/PA 25%, MA 26%, CO/MD 27%, CA/MN/WA 28%, WI 29%, OR 31%, DC 35%
- Funding is governed by the <u>Law of Large Numbers</u>:
  You have to submit lots of proposals to get any funding.
  http://dellweb.bfa.nsf.gov/awdfr3/default.asp

The Mindset Gap

The Mindset Gap

- In the olden days -- say, 10 years ago -- we used to say that our typical new Cyberinfrastructure user came from a Windows desktop or laptop background.

  - Those days are long gone ….

- Nowadays, we say that our typical new user comes from an iOS or Android background.

- How has that changed our job?

Mental Distance

- What's the mental distance between a handheld vs Linux, command line, batch computing?
  - Installing software
    - <u>Handheld</u>: Tap 3 times.
    - <u>Large scale</u>: EasyBuild if you're lucky, configure/make with modest dependencies if you're unlucky, bizarre random weirdness in practice.
  - Installing storage
    - <u>Handheld</u>: Buy a card for $10-50, pop it into the slot, the OS automatically recognizes it and starts using it.
    - <u>Large scale</u>: RFP, bid evaluation, configuration, purchase, deployment, maintenance.

What's the Cost of Storage?

- <u>Handheld</u>: tens or hundreds of dollars (which gets you tens or hundreds of GB).

- <u>Laptop</u>: tens or hundreds of dollars (which gets you TB).

- <u>Large scale</u> (per copy)
  - ~1 PB raw tape: ~$17K
  - ~1 PB raw spinning disk : ~$91K (ultra-cheap version)
  - ~1 PB raw SSD: ~$1M

Enterprise IT
vs
Research Computing:
Natural Enemies,
or Natural Allies?

Enterprise IT vs Research Computing

Enterprise IT: HARDENED

- Secure
- Established technology
- Best practices
- 5 nines: 99.999% uptime = 5.25 **minutes** of downtime per year

Research Computing: SQUISHY

- Fast and flexible (turn on a dime)
- Cutting edge technology (= broken)
- In some cases, no such thing as best practices
- 1 nine: 9x% uptime
  - NSF standard: 95% uptime = 18.25 **days** of downtime per year ("… [$60M NSF-funded] production resources should be unavailable as a result of scheduled and unscheduled maintenance no more than 5% of the time." -- NSF solicitation 17-558)

Enterprise vs Research: Incentives

- Suppose payroll is going out tomorrow, and the payroll system goes down tonight.
  - On payroll day, what happens on the Enterprise IT people who are accountable for the outage?
  - Therefore, what must Enterprise IT people do to stay in business?
- Suppose Research Computing isn't on the cutting edge, and so proposals from the institution are less competitive.
  - Eventually, what will happen to the Research Computing team?
  - Therefore, what must Research Computing people do to stay in business?

Enterprise vs Research: How to Resolve?

- **<u>Research Computing can afford to make mistakes</u>**:
  A system that's mostly up but crashes occasionally is fine.
  - 1 day of HPC downtime = 20-100 lost grad student days
    - 1 grad student = ~$50K/yr fully loaded with fringe + Indirect
      => 100 grad student days = ~$14K productivity loss
      => ~$100-$2000 productivity loss per research group

- **<u>Cost of 5 Nines vs 1 Nine</u>**: At least double, but budgets are fixed -- so the actual cost is cutting computing-intensive and data-intensive research productivity in half.

- **<u>Therefore</u>**: Let the machine go down from time to time, in exchange for having more (but less resilient) resources, to maximize research productivity per year, at the cost of occasional lost days.

Research is the Enterprise Testbed

- Research Computing has only limited best practices.

- But, technologies currently being adopted by Research Computing (e.g., Software Defined Networking) are likely to become enterprise requirements in a few to several years.

- So, let Enterprise IT watch Research Computing make mistakes, and use those observations to develop best practices for Enterprise IT.

Things to Say to a Researcher

Cost

- "This other way of doing it is cheaper than how you're currently doing it."
- "For the same cost, it could be so much better."

Control

- "You get to decide how to use your piece."
- "You can share it with whoever you want."

Effective Communication
ACI-REF Virt Res, Mon July 31 2017

35

Administration

- "Your students won't have to spend their time taking care of this."

How to Find Researchers

Where are the CDS&E Researchers?

1. Go to your institution's website.
2. Click on Academics.
3. Search for departmental websites.
4. On each departmental website, find the list of faculty (the link is usually "Faculty" or "People").
5. Read their research descriptions.

Keywords to Look For

- Computational

- Numerical

- Parallel (especially in CS)

- Informatics

- For Chemistry, look for Physical Chemists and Biochemists.

There are plenty of others – over time you'll develop a feel for it.

Contact Them!

- Contact those faculty.
- Tell them what your role is.
- If it's for a proposal, tell them:
    - what the program is;
    - what the due date is;
    - how much money is on the table.
- Ask them what their computational/storage/network/whatever needs are.

Go to New Faculty Meet-n-Greets

- Does your institution have events for new faculty?

- Go to them!

Visit Them!

- Make an appointment to visit with them.
  - Even better, offer to take them to lunch.
    - If you can get your institution to pay for the lunch, even better.
- Ask them questions:
  - At a high level, what's your research about?
  - What are the computing- and/or data-intensive aspects of your research?
  - If you had an infinitely large, infinitely fast computer, what research would you want to do?

The Intake Interview

Specific, Open-Ended Questions

These are questions whose answers you don't really care about – but they'll lead to useful discussions.

- What language is your software written in?

- Is it parallelized?

- Who wrote it?

- What operating system(s) has it been run on?

- Briefly describe the science problem it's used for.

- Briefly describe the numerical method or algorithm.

Questions cont'd

- How big is the memory footprint when running?
- How many timesteps/iterations do you plan to run per experiment?
- How many such experiments do you plan to run per year?
- Does it have no input, a little bit of input or a lot of input?
- Does it have a little bit of output or a lot of output?
- etc …

How to Find Researchers' Projects

Know Their Research

- If you've already talked to the researchers, you probably have a pretty good idea of who's got big data and/or big compute needs.

- Now you need to find out specifically how much Cyberinfrastructure capacity they need.

- You can always ask, but you'll get more information if you're writing an equipment proposal.

  - "I'm going to get you free goodies. Please send me a one page project summary plus the following details."

Equipment Proposal Questions #1

- How much funding does your research currently have? How much is pending? Planned? From what sources?

- How many faculty, staff, postdocs, grad students and undergrads on your team will be served by this equipment?

- What makes your research transformational?

- What are the broader impacts?

Equipment Proposal Questions #2

- How much of the proposed resource (CPU hours, storage, bandwidth, whatever) do you expect to need over the next N years?

- How did you calculate this amount?

- Please give me a one page summary of your research that incorporates these issues.

  - This is typically straightforward, because faculty often have either a 1 page summary from a grant proposal or a more broad research statement.

MRI/CRI for HPC Cluster Questions #1

- How many CPU core hours or node hours will you need over the next N years?

- How did you determine that?

- Have you benchmarked your code?

  - On what platform?

  - What is the expected performance improvement on the proposed instrument, compared to the platform you benchmarked on?

  - Do you plan to optimize the software? If so, what performance improvement do you anticipate?

    [This only applies to their own homebrew codes.]

http://www.nsf.gov/pubs/2011/nsf11011/nsf11011.jsp

MRI/CRI for HPC Cluster Questions #2

- If the proposal is for a new type of platform (for example, accelerators such as GPUs or Intel Xeon Phi/MIC):
    - Who will be responsible for porting the code to the new platform?
        - If this is either a community code or a commercial code, the porting may already have been done by the developers.
    - Have they committed to do so?
    - What speedup is expected on the new platform?

MRI/CRI for Storage Questions

- How much storage will be needed for this project?
    - If this is a live storage MRI/CRI: What is the maximum amount of storage at a time that will be needed for this project?
    - If this is an archival storage MRI/CRI: What is the total amount of storage needed over the lifetime of the instrument?
- How was that calculated?

Campus CI Questions #1

- What is the expected typical size of each dataset being transferred?

(It would be helpful to know expected growth rate: Are you expecting it to stay roughly the same over the next several years, or to double every two years, or what?)

Campus CI Questions #2

- Where are such datasets originating, and where are they being transferred to?

- Why do such datasets need to be transferred between these endpoints?

(That is, what requirement do these data transfers address for your team's research?)

Campus CI Questions #3

- What is the time window for transferring each such dataset?

- Why does each such dataset need to be transferred during that specific time window?

That is, what's the negative impact of the transfer taking (a) marginally longer and (b) much longer?

- How often do you expect to have such a data transfer need?

ACI-REF Virtual Residency

- Learn how to help researchers use research computing!
- FREE!
- Available onsite and remotely via videoconferencing.
- 2015-16: 128 people from 84 institutions in 37 US states & territories and 3 other countries.
  - 15 Minority Serving Institutions, 17 non-PhD-granting
- 2017: 181 people already signed up from 132 institutions in 45 US states & territories and 3 other countries
  - 26 Minority Serving Institutions, 20 non-PhD-granting
- We still have room for plenty more on videoconferencing (thanks to Oklahoma State U)!

https://www.eventbrite.com/o/aci-ref-virtual-residency-10618773713

Thanks for your attention!

Questions?

hneeman@ou.edu