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
ACI-REF Virtual Residency 2015
Monday June 1 2015

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

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

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

Planets etc

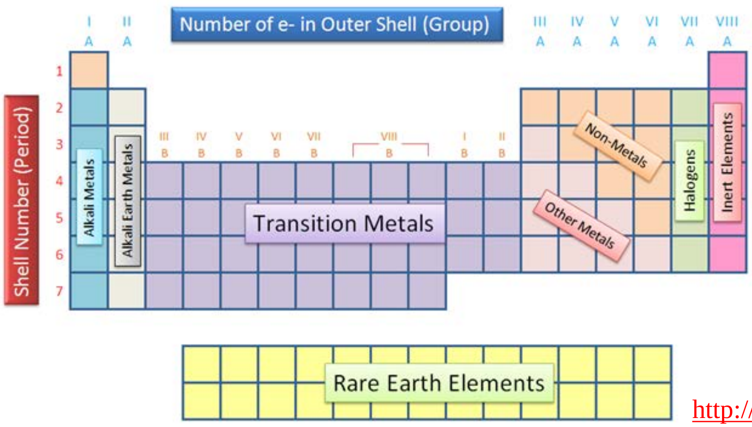
What are planets made of?

- 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

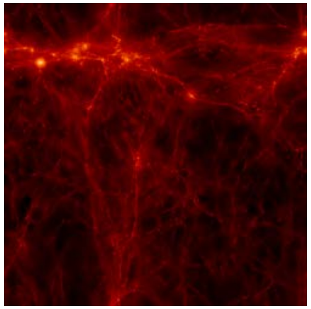
So What's a Metal?

- To a chemist, metals have a very specific chemical definition.



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.



http://user.astro.columbia.edu/~gbryan/Site/IGM_files/gas_density_z0.png

Projection

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

Scale

- At quantum scale over 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 6 months from submitted to published.

Tenured Faculty

At research-intensive institutions:

- **Incentive Structure**: 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.
- **Need**: 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.

Probability of Success

- National Science Foundation, FY2014: 23% overall
  - BIO 27%, CISE 23%, EHR 17%, ENG 18%, GEO 26%, MPS 26%, SBE 22%
    - Office of Director 45%, Office of Polar Programs 75%
  - EPSCoR jurisdictions: American Samoa/Guam 0% (no PhD-granting), MS 12%, WV 14%, ND/ID/AL 15%, AR 16%, SC/KY/KS 17%, MO/NE/LA 18%, PR/WY/NM 19%, OK 20%, VT 21%, NH/IA 22%, HI 23%, NV/DE/SD/ME/AK 24%, MT 25%, USVI 29%, RI 34%
  - Non-EPSCoR jurisdictions: FL/OH 17%, TX 18%, TN/VA 20%, AZ 21%, GA/IN 22%, MI/NC/NJ/NY/OR/PA 24%, IL/MA/MD/WA 25%, CO/CT 26%, CA 27%, WI 28%, MN 30%, DC 36%

Funding is governed by the Law of Large Numbers: you have to submit lots of proposals to get any funding.

- NSF CC-NIE 2013: 87 proposals, 43 grants (49%)
- NSF CC*IIE 2014: 132 proposals, 46 grants (35%)
- NSF MRI 2010: 993 proposals, 941 reviewed, 150 awarded (15%)
- Henry (as PI/Co-PI, 1999-): 52 proposals, 11 awards, 2 pending (22%)

http://dellweb.bfa.nsf.gov/awdfr3/default.asp

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."
- "It's not my place to tell you how to spend your own money."

Administration

- "Your students won't have to spend their time taking care of this."
- "We don't charge for this service."

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
- Calculation
- Modeling
- Simulation
- 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.

How Often?

Try to do this at least every few years, because new faculty come in, and you may not be in the loop about them.

Let's Try It!

Let's give it a try and see whether we can find some.

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?

Specific, Open-Ended Questions

- 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?

Questions cont'd

- Is the code consistently indented?
- Are the variable names meaningful, especially outside the code?
- Is there a preference for numeric literal constants or for named constants?
- Are "inputs" actually input, or are they hardwired at compile time?
- Are arrays allocated statically or dynamically? If dynamically, on the stack or on the heap?

Questions cont'd

- For Fortran 77, are there common blocks (poison)?
- Are there many small variables/objects, or a few big ones?
- Are there many small procedures/methods, or a few big ones?
- What is the format of the input and output? (e.g., plain text, native binary, portable binary etc)
- Does the code input and/or output metadata as well as data?
- Does I/O occur occasionally or frequently?
- Is argument passing consistent or chaotic?

Questions cont'd

- Does the code use non-portable tricks? (e.g., Fortran array dummy/formal arguments of length 1, passing an array of one type but intentionally receiving it as an array of another type)
- Is the code reasonably vanilla, or does it have nonstandard tricks in it?

These Questions Don’t Matter

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

I once interviewed a research team and got almost 9 pages of notes from the first 9 questions.

How to Find Researcher 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?

CC*DNI 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?)

CC*DNI 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?)

CC*DNI 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?

OK Supercomputing Symposium 2015


2003 Keynote:
Peter Freeman
NSF
Computer & Information Science & Engineering
Assistant Director


2004 Keynote:
Sangtae Kim
NSF Shared Cyberinfrastructure
Division Director


2005 Keynote:
Walt Brooks
NASA Advanced Supercomputing
Division Director

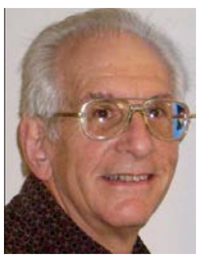

2006 Keynote:
Dan Atkins
Head of NSF's
Office of Cyberinfrastructure


2007 Keynote:
Jay Boisseau
Director
Texas Advanced Computing Center
U. Texas Austin


2008 Keynote:
José Munoz
Deputy Office Director/Senior Scientific Advisor
NSF Office of Cyberinfrastructure


2009 Keynote:
Douglass Post
Chief Scientist
US Dept of Defense
HPC Modernization Program


2010 Keynote:
Horst Simon
Deputy Director
Lawrence Berkeley National Laboratory


2011 Keynote:
Barry Schneider
Program Manager
National Science Foundation


2012 Keynote:
Thom Dunning
Director
National Center for Supercomputing Applications


2013 Keynote:
John Shalf
Dept Head CS
Lawrence Berkeley Lab
CTO, NERSC


2014 Keynote:
Irene Qualters
Division Director
Advanced Cyberinfarstructure Division, NSF


2015 Keynote:
Jim Kurose
NSF
Computer & Information Science & Engineering
Assistant Director

FREE!
Wed Sep 23 2015 @ OU

Reception/Poster Session
Tue Sep 22 2015 @ OU
Symposium
Wed Sep 23 2015 @ OU

Thanks for your attention!

Questions?

hneeman@ou.edu