Effective Communication

Henry Neeman, University of Oklahoma
Director, OU Supercomputing Center for Education & Research (OSCER)
Associate Professor, Gallogly College of Engineering
Adjunct Associate Professor, School of Computer Science
Virtual Residency Introductory Workshop 2023
Monday June 26 2023

Outline

- How to Talk to Researchers: Research Terminology
- Enterprise IT vs Research Computing
- The Mindset Gap
- Researcher Types
- 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, water etc

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

Oxygen in Astronomy

- The universe's visible matter 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% (including oxygen **AND EVERYTHING ELSE**)

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

Planets etc

What are planets made of?

Metals! (In the chemical sense)

- 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

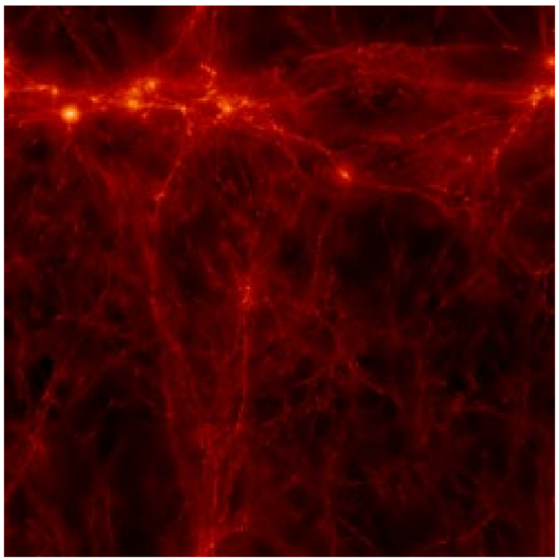
http://en.wikipedia.org/wiki/Earth

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

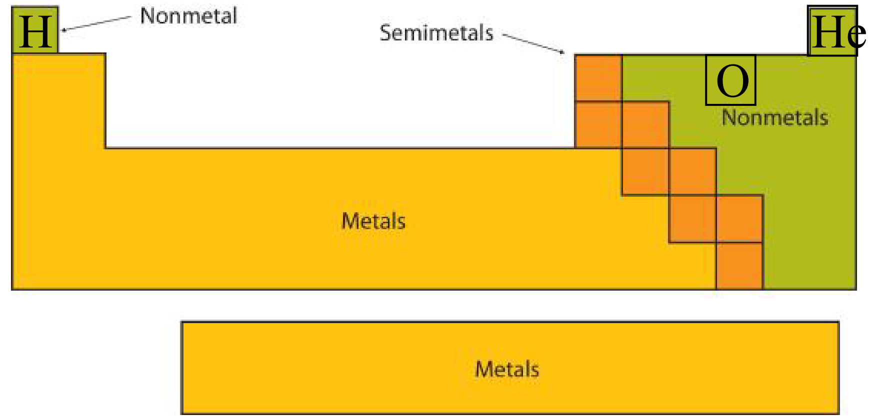
http://www.smithsonianmag.com/smart-news/barns-are-painted-red-because-of-the-physics-of-dying-stars-58185724/?utm_source=keywee-facebook.com&utm_medium=socialmedia&utm_campaign=keywee&kwp_0=283306&kwp_4=1091891&kwp_1=506963

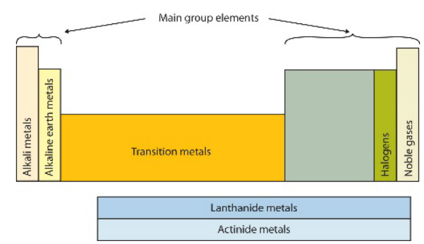
So What's a Metal?

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



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





https://chem.libretexts.org/Courses/College_of_Marin/Marin%3A_CHEM_114_-_Introductory_Chemistry_(Daubenmire)/04%3A_Atoms_and_Elements/4.6%3A_Looking_for_Patterns%3A_The_Periodic_Law_and_the_Periodic_Table

- But, 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 does "projection" mean to a mathematician?
  - What does "projection" mean to a psychologist?
  - What does "projection" mean to a movie producer?

A key skill that you'll develop by experience:

How to tell when people are talking past each other by using the same word to mean different things, or vice versa.

Example: "Who's on first?" comedy routine:

https://www.youtube.com/watch?v=sShMA85pv8M

What Are Fluids?

- Colloquial definition: Liquids.
- Mom's and physician's definition:

  Something you should drink plenty of when you're sick.

https://www.zocdoc.com/answers/9591/does-drinking-fluids-help-when-you-have-a-cold

What Are Fluids? (cont'd)

- Physical science & engineering definition: 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 and that changes its shape at a steady rate when acted upon by a force tending to change its shape." – dictionary.com
  - Liquids are ***incompressible*** fluids.
- Social science definitions
  - Something that has room for interpretation (e.g., gender).
  - Describing a feeling, mood or appearance.
- Finance: An asset convertible into cash.

Database

Which of these is a database?

1. SQL/Oracle/Access/DB2/MongoDB?
   (According to computer scientists)

OR

2. A large collection of files?
   (According to bioscientists)

Data vs Metadata

What's data? What's metadata?

Is code data? To whom?

Scale

- At quantum scale over femtoseconds,
  how much does gravity matter?
- How about at cosmological scale over eons?

What are some other disciplines where gravity **<u>DOES</u>** matter, and some other disciplines where gravity **<u>DOESN'T</u>** matter? For each, **<u>EXPLAIN WHY</u>**. (See this session's Google Doc.)

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?

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 actual research … ?

Is Simulated Data Actually Data?

- I had a colleague in Chemical Engineering who told me that, if he referred to output from a simulation as "data" in front of his colleagues, he'd be laughed out of the discipline.
- A compiler designer considers code to be data – the stuff that everyone else considers data isn't very interesting.
- What are some examples of disciplines with very different definitions of "data" from what we've already discussed?

Do They Know What You Know?

<u>NO, RESEARCHERS DON'T KNOW WHAT YOU KNOW.</u>

- The most normal thing in the world is to assume that the person you're talking to knows what you know.
- The second most normal thing in the world is to be dead wrong about that.

Most People Are Bad at This!

- "The probability two people selected at random will share the same concept about penguins is around 12 percent …" S. Makin, 2023: "People Differ Widely in Their Understanding of Even a Simple Concept Such as the Word 'Penguin.'" *Scientific American*.
  - https://www.scientificamerican.com/article/people-differ-widely-in-their-understanding-of-even-a-simple-concept-such-as-the-word-penguin1/
- "[Researchers] found that only 2 percent of conversations ended at the time both parties desired, and only 30 percent of them finished when [even] one of the pair wanted them to." R. Nuwer, 2021: "People Literally Don't Know When to Shut Up—or Keep Talking—Science Confirms." *Scientific American*. https://www.scientificamerican.com/article/people-literally-dont-know-when-to-shut-up-or-keep-talking-science-confirms/

Thanks to Bev Corwin for these references!

Exercise

1. Go to: https://docs.google.com/document/d/1NKEZ3xnlciu8YEwM4yiPd5kiwwm7u8shW7NXW62_Iws/edit
2. At the bottom, enter a piece of terminology – **a term that's used in many (sub-sub-)disciplines and/or colloquially** – and the name of a research discipline (or sub-discipline or sub-sub-discipline ...) that uses that term, and define what that term means in that (sub-sub-)discipline.

Or, you can add a new definition for a new (sub-sub-)discipline to a term that's already listed.

This way, we can create a shared glossary of terminology as used in various disciplines.

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

Enterprise IT & Research Computing

Enterprise IT: 5 NINES

- Secure
- Established technology
- Best practices
- **5 nines**: 99.999% uptime = 5¼ **minutes** of downtime per year

Research Computing: 1½ NINES

- Fast and flexible (turn on a dime)
- Cutting edge technology (= broken)
- In some cases, no such thing as best practices
- **1½ nines**: 95% uptime = 18¼ **days** (438 hours) downtime per year
  - NSF's standard, from NSF 17-558, 19-587, 20-606, 23-518:

"… [NSF-funded] production resources should be unavailable as a result of scheduled and unscheduled maintenance no more than 5% of the time."

NOTE: OU's supercomputer ~ 99% uptime; OU IT enterprise = 99.995%

Enterprise IT Example

- On Aug 8 2016, Delta Air Lines experienced a power outage in their Atlanta data center that lasted 5 hours.
  - Cost: $150M ($1M for every 2 minutes of downtime)

https://money.cnn.com/2016/09/07/technology/delta-computer-outage-cost/

Enterprise vs Research: Incentives

- Suppose a mission critical capability is needed tomorrow, and the relevant IT system goes down tonight.
  - Tomorrow, what happens to 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 thus 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?

- **Research Computing can afford to make mistakes**: A system that's mostly up but crashes occasionally is fine.
  - 1 24-hour day of HPC downtime = 10-100 lost grad student days
    - 1 grad student = ~$60K/yr fully loaded with fringe+tuition+Indirect
=> 100 grad student days = ~$16K productivity loss
=> ~$300-$3000 productivity loss per research group
- **Cost of 5 Nines vs 1½ Nines**: 5-10x, but budgets are fixed – so the actual cost is cutting computing-intensive/data-intensive research productivity by 80-90%.
- **Therefore**: Let the research machine go down from time to time, as a tradeoff for having bigger (but less resilient) resources, to maximize research productivity per year, at the cost of occasional lost days.

How Much Failure is Normal?

In calendar 2021, on OU's supercomputer:

- ~4.8M jobs ran;
- ~471K jobs failed (8.2% of all jobs);
- of those ~471K failed jobs,
  ~13,724 jobs failed due to server failure
  (2.9% of failed jobs, 0.24% of all jobs,
  and many of these **weren't** hardware fails).

So:

- Batch job failure is **normal**.
- Batch job failure due to server failure is **insignificant**.

```
4,833,261 COMPLETED
  457,092    FAILED
  326,228 CANCELLED
  131,901   TIMEOUT
   13,724 NODE_FAIL
   10,287  REQUEUED
    2,990   RUNNING
      767   PENDING
```

Implications for IT

- Research computing is <u>**LESS EXPENSIVE**</u> than Enterprise IT.
- But, it's also <u>**LESS RESILIENT**</u> (1½ nines vs 5 nines).
- So, when a researcher comes to us for help with a specific capability, we should ask:

"Can this capability tolerate an average of roughly 8 hours of downtime per month?"

- If yes, Research Computing may well be their best bet.
- If no, Enterprise IT is definitely the right way to go.

Research is the Enterprise Testbed

- Research Computing has only limited best practices.
- But, technologies currently being adopted by Research Computing are likely to become enterprise requirements in not-too-many years.
- So, let Enterprise IT watch Research Computing make mistakes, and use those observations to help develop best practices for Enterprise IT.

<u>Example</u>: In 2012, OU Research Computing moved to OU's then-new data center in one week, and was the first team in there.

That helped prove that the data center was ready for enterprise systems – which would have been too risky to move in first.

The Mindset Gap

The Mindset Gap

- In the olden days – say, 15 years ago – we used to say that a typical new Research Computing user came from a Windows or MacOS desktop or laptop background.
  - Those days are long gone ….
- Nowadays, we say that a 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, remote, shared, batch computing?
  - Installing software
    - Handheld: Tap 3 times.
    - Large scale: EasyBuild or Spack if you're lucky, configure/make with lots of dependencies if you're unlucky, bizarre random weirdness in practice.
      - Is it realistic to expect all of our users to be able to do this?
  - Installing storage
    - Handheld: Buy a card for $10-50, pop it into the slot, the OS automatically recognizes it and starts using it – or get new phone!
    - Large scale: RFP, bid evaluation, configuration, purchase, deployment, maintenance, expansion, decommissioning – several months.

What's the Cost of Storage?

- Handheld: tens or hundreds of dollars
  (which gets you tens or hundreds of GB).
- Laptop: tens or hundreds of dollars
  (which gets you TB of spinning disk or GB/TB of SSD).
- Large scale (per copy)
  - ~1 PB usable tape: ~$5K
  - ~1 PB usable spinning disk : ~$47K (ultra-cheap version) => ~8x
  - ~1 PB usable SSD: ~$329K (ultra-cheap version) => ~56x
    - And, for large scale resources, PB are normal!

Why Else is Using Research Computing Hard?

Research Computing is hard because:

1. the mindset gap;
2. Research Computing systems always get harder to use:
   - the storage hierarchy get deeper (e.g., registers/cache/RAM/Optane/flash/spinning disk/tape),
   - parallelism gets more hybrid (e.g., MPI on top of GPUs);
3. more researchers need Research Computing:
   - more disciplines need Research Computing (e.g., life sciences, social sciences, humanities, arts, business, athletics),
   - a higher fraction of users within each discipline need research computing.

Why Are Researchers "That Way"?

Researcher Types

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

What Are Faculty Rewarded For?

Faculty at research-intensive institutions are rewarded for <u>three things</u>:

- <u>bringing in grant money</u>;
- <u>publishing papers</u>;
- <u>graduating students</u>.

Faculty absolutely <u>**AREN'T**</u> rewarded for having good IT.

So they'd strongly prefer <u>**NOT**</u> to pay more than the <u>**ABSOLUTE MINIMUM**</u> for their computing (ideally zero).

(And their mental model for what compute and storage cost is the price of a laptop and some USB hard drives from their local big box store – because that's been their experience so far.)

Tenure-Track (Not Yet Tenured) Faculty

At research-intensive institutions:

- **Incentive Structure**: I need to (a) publish lots of papers, (b) bring in lots of grant money and (c) graduate lots of students, **or I'm fired**.
- **Need**: I need stuff to work **now** and keep working reliably.
- **Timeline**
  - 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 here), **BUT**
  - I have 5 years (the 6th year is when my materials are evaluated), **BUT**
  - I have **4 ½ years**, because it typically takes a journal article about 6 months from submitting it to it getting 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 else:
  - I won't get a raise;
  - I won't get a named chair;
  - I won't get other prestigious outcomes (e.g., elected to the relevant National Academy etc).
- **Need**: I need stuff to work **now** and keep working reliably.

Research Faculty (Non-Tenure-Track)

- If I don't bring in grant money, I'm laid off.
- I need to publish a lot to keep bringing in grant money.
- I need a track record of graduating lots of students, so I can get a tenure track job somewhere.
  - Because I don't want to have to live on "soft money" forever!

Postdocs

- I need to publish a lot or I'll lose my postdoc position.
- I need to learn how to get lots of grant money,
  and 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, scholarship, grant, etc.
  - While I'm in school, I'm giving up
    that many years of salary and benefits.

Research Funding

Academic Research Funding Model

- Most academic research is funded by grants:
  - a <u>random</u> amount of money
  - shows up on a <u>random</u> date
  - and is available for a <u>random</u> duration.
- Grants usually strongly <u>favor</u> researcher salaries and <u>disfavor</u> equipment, services and non-research staff.
  - Researchers don't get much credit for the latter.
  - Recurring charges are very hard to pay long term.
- Some academic research at some institutions is internally funded (for example, via endowment), but this is typically larger and more common at higher-ranked institutions.
  - The problem with the Top 50 is, there's only 50 of them.

Probability of Success #1

- National Science Foundation, federal FY2021: **28% overall**
  - Geosciences (GEO): 42%
  - Mathematical & Physical Sciences (MPS): 34%
  - Computer & Information Science & Engineering (CISE): 28%
    - Office of Advanced Cyberinfrastructure (OAC): 32%
  - Biosciences (BIO): 27%
  - Office of the Director: 25%
  - Engineering (ENG): 24%
  - Technology, Innovation & Partnerships (TIP): 24%
  - Education & Human Resources (EHR): 24%
  - Social, Behavioral & Economic (SBE): 22%
- Funding is governed by the Law of Large Numbers:
  You have to submit lots of proposals to get any funding.
  - So faculty can expect to spend a lot of time writing proposals.

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

Probability of Success #2

- National Science Foundation, federal FY2021: **28% overall**
  - EPSCoR jurisdictions (EPSCoR jurisdiction mean 31.4%): ND 18%, WY 19%, KS 20%, AL/ID/OK 22%, MS/WV 23%, DE/SC/VT 24%, AR/IA/KY/NH/NM 27%, NE/NV 28%, LA/ME 29%, MT 30%, SD 33%, HI/PR 34%, AK/RI 39%, VI 80% (5 proposals), GU 100% (1 proposal)
  - Non-EPSCoR jurisdictions (non-EPSCOR jurisdiction mean 28.6%): FL 21%, MO/OH/TX 24%, VA 25%, GA 26%, MD/MI/TN 27%, AZ/CT/IL/NJ 28%, IN/NC/PA/UT 29%, NY 30%, CA/CO/OR/WI 31%, MA 32%, DC/MN 33%, WA 39%
- Funding is governed by the Law of Large Numbers:

  You have to submit lots of proposals to get any funding.
  - So faculty can expect to spend a lot of time writing proposals.

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

Research Computing Funding Models

Most academic institutions fund research computing via a combination of some or all of:

- internal funds on a regular cadence (annual budget);
- internal funds at random times (Let's buy a supercomputer!);
- grant funds at random times (Let's buy a supercomputer!);
- for public institutions, state appropriations in random years (the legislature giveth and the legislature taketh away);
- usage charges (uncommon);
- "condominium" purchases of researcher-owned compute servers at random times.

Also, some institutions like leasing, others like buying, others do both.

How Should Faculty Spend Their Time?

At OU, we don't ask faculty to write "grant" proposals for time on our supercomputer.

Why?

Faculty have a limited number of hours per year for writing proposals.

We'd much rather they spend that time writing proposals for external research grants, than for internal time on a machine that OU has already paid for.

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."

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 Computational Chemists, 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-intensive and/or data-intensive aspects of your research?
  - Suppose you had an infinitely large, infinitely fast computer. What research would you want to do?

Exercise

1. Go to: https://docs.google.com/document/d/1NKEZ3xnlciu8YEwM4yiPd5kiwwm7u8shW7NXW62_Iws/edit
2. At the bottom, enter a piece of terminology – **a term that's used in many (sub-sub-)disciplines and/or colloquially** – and the name of a research discipline (or sub-discipline or sub-sub-discipline ...) that uses that term, and define what that term means in that (sub-sub-)discipline.

Or, you can add a new definition for a new (sub-sub-)discipline to a term that's already listed.

This way, we can create a shared glossary of terminology as used in various disciplines.

Note that all the terms discussed in this session are already there, so you'll have to add new ones!

Thanks for your attention!

Questions?

hneeman@ou.edu

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?
  - Many small disk I/O transactions, or a few big ones?
- 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 is the intellectual merit of your research?
- What makes your research transformational/innovative?
- What’s the importance/research impact of your research?
- What are the broader impacts?
  - Education/training
  - Underrepresented populations (minorities, women, disabled etc)
  - Economic/social impact

Equipment Proposal Questions #2

- How much of the proposed resource (CPU hours, storage, bandwidth, whatever) do you expect to need over lifetime of the grant (i.e., the next N years)?
- How did you calculate this amount?
- Why is this specific equipment important for your research?
- What if you didn't have this specific equipment?
- 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? How did you extrapolate that?
  - Do you plan to optimize the software? If so, what performance improvement do you anticipate, and why?

    [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 more novel type of platform (for example, accelerators such as GPUs):
  - Who will be responsible for porting the code to the novel 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? How did you determine that?

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?

Thanks for your attention!

Questions?

hneeman@ou.edu