

Present : Ani, Steve, don, anita, mario, jarek, wolfgang,

Example of Sloan easy to identify single point of failure - open its your best people -
Not difficult to anticipate - if value of data is long term consumption - last images stss - still
publishing papers on access to imaging from 2010 (because it was well calibrated, curated,
checked) .

BUying stuff from apps store- example: buy it try it, find dissatisfaction, if I have to pay and it
won't be optimal, I might be disincentivized from buying/investing

We are not ready sociologically for that kind of thing LSST to US and CHile avail but not to rest
of world, there will be pushback morally it is wrong to "not share"

individual will not always be the payer - data centers have a different economic

ispsr example (fee is baked into proposal, and institutions get access via their
membership/subscription) right now there's more than one way for the funding to stream, but
what if when you fund a proposal you also fund curation of the data (& a data management
charge).

In the commons proposal you get credits you can apply for - they give you storage, computing,
data curation points, NEH says the solution need to be compliant) Onus is on data creator

the wealth of networks quote when is it good for the data to be freely available.

being able to freely associate like the person in the hardware - the time to do this is during
phase of invention /curiosity cut it off is when the research has been done and someone has to
monetize it.

inadvertently create a club of people who can get at the data

beautiful thing in astronomy is that barrier of entry has fallen

democratize data gathering democratizes sharing too

agencies really are paying for papers now - when i get up for tenure papers will matter data
wont

- assumptions may be generational when will the review committees accept data citation rather
than or alongside papers?

it's still allowed to "pull" your data.

Clubs don't have incentive to share -

anita: redundancy is beneficial in publishing

derived data is turtles all the way down - the lifetime of what we push is about a year - there's no provenance of the data - they don't cite ea other .

The data citation principles from Force 11 are one way to get at this. For purpose of humans getting credit.

me having a cited doi doesn't buy me much

maybe one way is to incentive reuse, download, citation

how do we encourage and make transparent review of data and code?

new ideas are just as easy to kill - if you threaten a researcher's w/an antagonistic reviewer you risk disincentivizing

data publication alonge - has peer review and accompanying methods paper about how created the data but no science paper

differences in observational science

you can't ask a cheap question in HEP anymore

cancer papers natural language processing carnegie mellon project /darpa
experiments are fairly well described but weak link between hypothesis and experiments and experiments and conclusion do not have hard connections you can see in the NLP, cargo cult science . . .

hauser - harvard experiments w/cotton top tamarin monkeys example - measured looking at the monkeys.

what about the reproducibility projects : contrast between observational interpretations and

signey vanderbilt analogy what about one page wonders aka "sidwipes" -

what if you try to reproduce the veracity of the linkage?

discovering hidden biases and errors

if I want to share something today I don't have anywhere to put it?

what is the long term value of data ?

some data has finite lifetimes that you can measure

princeton - how much does it cost to keep data forever it's a factor of 2 -

VMs for rendering

some data should not be looked at because looking destroys

other data gets better depending with how we develop tech that improves what we can get out of the data

need for low level data is going to become higher as we deal more with noise

in HEP they've been throwing away data for years - "they know it's not what they're looking for"

tool generated metadata -

business models drive behavior and the funder pays model is still dominant

incentives to share data are underutilized/not fully valued ta institutions (altruism won't be a sustaining motivation) - tenure committees etc aren't rewarding creators of beautiful data

when an agency signs off that you did a good job w/sharing could it be built in to incentive by provision of a grad student - what if once you're in (the grant) incentives to share were part of the deal

it's false that astronomy is an observation only science there are comparisons between clinical psychology and observational science like astronomy

What is reproducible science in the context of bits and bytes vs labs or from observations and claims made related to observed data.

Need to be careful about making observations beyond local areas of expertise

But maybe what we can do or n need to do is examine business model - alternatives to whom the agency pays? what they pay for?

is the current way astrophysics data is open good enough for you to do the science you want to do.. It is empirically true that a huge amt of science has been done

minimum level of publication : zero proprietary period. Upon publication all the data is all there

Sometimes papers are serial or part of thesis/dissertation efforts

the gentleman's agreement is that you know a student is working on something for his thesis (you hang back this avoids regulation) but the more trendy a data set the more poachers there are.

bcs NSF provides money for shared facilities, when a proposal is put into use the facility - your proposal for instrument use/facility use is published - intention to do something is published knowledge. Does that go far enough

more than 25% of the papers today use a public database

The minimal circumstance w/careful look or pause to consider the downstream impacts to people just coming into field if every time you give away their starting point.

if somebody can do sharing through a public period why should I have to provide the derived data products. It doesn't necessarily follow that you have to empty your data bank every time you publish.

sociological within astronomy -
mario talked abt incremental cost for forever vs two years (maybe answer should now be "forever")

Princeton slides on cost

we need to incentivize, agencies are going to have to pay too.

When real work is extended to accomplish something there are ways to hold that back or monetize it turn it into present dollars

It's natural to say perpetual care belongs in some specialized environment. We don't bury our relatives in the back yard we put them in the graveyard. There's a specialized place for that .

what would be ideal would be to couple the data curation and migration effort - you don't leave observatory without writing a preservation copy of your data someplace else.

require universities to hold onto data assure availability for x years

certified repositories

if a number of systems

metadata repositories, indexes, data registry

Elsevier grant prop w Carnegie Mellon on this same topic
throw commercial grade software at the problem

cats analogy - different motivators

This one time grant thing is terrible
what abt contingent release based on

what if DMP has a link and final review is not approved in final report if there weren't
accountability

could you offload the data handoff/check?

contingent something being done in a certain framework, but if you use another
method/preferred method you get a reward/incentive

repositories should be bonded.

ESIP data stewardship example shows a discipline successfully adopting a standard practice
and teaching others how to adopt it too.

build layers progressively
understand what you need most: not to lose data

at some time you even become uninterested in your own data because you move on

these are important issues and we should make progress on them - in stepped approach

