What unmet needs (if any) are there around support for data security, data-set protection, controlled access, contract management and tracking, etc.? Who on campus is doing this well? Are there efficiencies to be gained by collaborating on certain aspects or tools? Cf. the ICPSR contract management tools.

What are the next steps in this discussion? Who should participate?

HIPAA workshop at CAS (scientific computing) meeting; lots of questions around compliance and risk management

Notion of security life cycle; IT people think about selecting/implementing security controls, but assessment and monitoring phases (documentation trail for risk management) are just as important

D-Lab: assist researchers with working with data that can't be shared in a public way
Restrictions set up by provider (depends on type: e.g. survey data, vs. from administrative sources - not intended for research but repurposed)

Industry, federal/state/local government, academic institutions

Different sets of incentives around sharing data, different intents, different knowledge about how to protect data

Want to understand what other groups are on campus that can collaborate on this

Technical: protecting end-to-end contact, but also IRB, human subjects, access issues (library involved here), tracking issues (contracts between producer and researcher)

Social welfare - same situation as D-Lab

Most research - child welfare, abuse, getting data sets from federal entities

Independent, separate unique entities - 4 data sets from Australia, 1 from Norway - not just US federal but also global

All have own set of requirements

For faculty, there's not a known place to go to find out what procedures are to acquire data appropriately

PhD candidate was somehow approved a data set that didn't go right for human subjects; data set just appeared one day

Student couldn't find something, contacted data owner directly, and they approved it

Would like to see some controls in place; working on that internally, but if someone else already has it?

D-lab: Getting people directed to the right place, no single process exists that applies to everyone

Physical storage, cold storage for sensitive data - receive it on USB, DVD, CD, have to maintain that master record throughout the research

Every faculty member has at least one data set in their office; we're within compliance, but want a better standard for cold storage

Load working set from the original USB, CD, etc.

Locked box, locked filing cabinet, locked room - for cold storage

Some research projects are for 7 years; have to keep the USB that whole time

For some, master data set can't be in same location as another master set (can't have a CD with a CD) - concern of re-identifying data

Don't know if social welfare should take this on separately, or if this is already being done elsewhere

Almost never have to go back to the original data set; only if there's a catastrophic failure to the working set

Monthly cost for multiple deposit boxes, over 7 years - for something that's not touched

When applying for data sets, have to identify any external entities involved in research

Could be misconstrued if we say "no external entities involved" and "will be stored at the bank"

All groups are sensitive about data, but child, medical, mental health, criminal -- very touchy about who has access

Any corporate name has to be vetted

IST would still be considered on-campus

Faculty member is PI; they're signing their career on that form -- reluctant to give data to a company

PAHMA - developing public access collections browser; lots of material in collection isn't for public consumption (location in the museum of certain objects)

Native American Graves Protection & Repatriation Act - locations of where those materials were recovered is confidential

But people researching the materials need to know where things are

Not clear whether to "fuzzify" data and make it public, or if fuzzy location has to go into repository to start with

Whole chain of custody of materials from accession on; many donations are anonymous

Donors known in database, can't be revealed to public

Valuation of materials -- confidential or at least sensitive

Case-by-case basis for what to do with it

Hallucinogens, endangered plants - don't want that in the public portal

Some images can't be shown to the public
Community feedback about what should happen to it; certain objects have special cultural significance, don't necessarily understand extent of significance

Ceremonial knife that should never be around food - figured they could put a picture of it on the web, but they asked not to, because someone could be eating while viewing it

Depends on regulation - some data sets have one set of requirements, might have extra requirements if combined with other data sets

Data, how you store it, in relation to what other kinds of data - never simple question

Creating data sets inadvertently that have different restrictions

Donations, anonymous, don't want to be anonymous to everyone - difficult to manage programmatically

Some agencies release data one field by field basis, you have to justify need for each field

If something from IRS touches any other data, all subject to IRS rules

Set of shadow systems: different pieces of different fields available depending on their status

Campus has minimum security standards -- people encouraged to go beyond that

Social welfare - has a data set that says data can't be transferred via mobile device, external storage, USB; in request, says that data will be moved via USB during summer sessions; encryption level for USB, it was approved

Even though data owner says you can't do it, school was able to show a level of security that they approved it anyway

If we could create a best practice, most entities would be willing to work with that, exceeds minimum standard

IST has matrix of levels for storage, DMPTool

9 different entities as source of data; put down what all of them have in common, sub-list of all unique's for each one

30-day password change, review of data every quarter to ensure integrity, etc.

Planning to build internal policy that covers everything possible - will use one document to obtain any data set

RIT is working on research data management; CIO agrees this is important, talking with the library about a plan

Non-networked machines

Better formalize nonnetworked machines, by formalizing it, it'll be easier in the long run

Everyone will be doing it the same way

Data rest encryption for anyone who wants to use Berkeley desktop

Bitlocker, FileVault - keys are escrowed (may or may not meet requirements)

UCSF formally recommends encrypted flash drives; would it be worth making one of encrypted drives part of jack standard?

Full disk encryption - only data at rest, other issues around data in transit

Verifyably encrypted data in case laptop is lost

HIPAA - no certification, no way to certify a process as being "good for HIPAA"

Assumption of risk; convince the people signing off on the risk that you're compliant, have mitigated issues

When people were audited, folks in trouble were people who hadn't documented what they did (even if processes were good)

Matter of having a good story around this stuff, basic guidelines, groups set up doing this

"This is how you do it, this is what you point to when you have to tell a story about what you're doing."

Documenting processes means showing accountability

By going through the exercise, you find the gaps, and find out how to close the gaps

Different orgs describe HIPAA difference

Work through those issues centrally, so we can provide better guidelines

NIST is a laundry list, but if your documentation gis based on that laundry list, will go smoothly if audited

If you get audited and there's problems, not the end of the world; a year later and the problems are still there, that's bad

Social welfare: can be blacklisted and lose accreditation for 5-7 years the first time you're audited and not in compliance
Put specifics in request for every data set

Have heard campus might be moving away from Semantech in AV provider -- have to request (not notify, request) -- Semantic was specifically identified in every agreement

They have the right to pull data back if they don’t agree

Want to say something like “industry standard antivirus” to avoid that risk

Security plan for both physical and working st; any deviation, have to request

A good example for why there’s no certificate: e.g. Heartbleed, can quickly become outdated

“maintain reasonable security” - living process that has to change with the time

People want to not be liable, and just follow certification

HIPAA - going to get worse, have funding for audits, increased fines, etc.

Social welfare - going to try to change model where it's not PI, it's auspices of UC Berkeley

Currently in social welfare- being approved to an individual PI

School wants to work with campus, rework that where it's being approved to UC Berkeley, university is owner if faculty member moves on

IRB has resources for investigators - data security policy, guidelines, matrix on website

Talks about different levels of data, security needed for each level

IRB makes sure investigators are following policies when they apply to office

Before investigators can get approval, verify they’re storing data in the way stated as necessary in policy

If investigators do what they say, it'll cover a lot for data security

Make sure number of people who have access to data is limited, encrypt hard drives, etc., code data and store identifiers elsewhere (and encrypt those)

CPHS, extremely sensitive data - contact Leon's department

Starting to put tools out to help people document security practices

If people are struggling with documentation, want to understand what the struggle is

Want to be able to help campus understand how to do this better

Still have questions about what the problems are for the research community

Come up with top 3, top 5 lists of research concerns, follow that to see what services are out there

Sharing elements of data protection plans in aggregate? - difficult to pull information that specific, plus there's privacy/confidentiality measures

Social welfare - data sets can be used by multiple groups, multiple PIs ahve to apply, store multiple copies