Typical sources of VIVO data

back up to

- Structure of the University
  - Departments, Divisions, Groups
  - Different organizations within the campus might have conflicting views of this data
    - For example, a research group that reports to one unit but is funded by another.
- Names and network IDs from an LDAP directory
- Personnel records from Human Resources
  - Must be filtered to remove sensitive data.
- Grants information from the Office of Sponsored Funds
  - Because of data challenges, this data is only harvested a few times each year.
- List of classes from the Registrar's Office
- Publications from a faculty reporting system
  - This also provides many challenges.
    - For example, two co-authors may provide slightly different names for the same journal article, or for the journal itself.
  - If the data comes from a well-curated source (e.g. PubMed), the challenges of data cleaning are greatly reduced, but the challenges of disambiguation are increased.
    - In the on-campus faculty reporting system, it is common practice to use an ID that definitively associates the data with the faculty member.
    - When reading from a public source, are these the same authors or different authors?
      - James L. Fox
      - J. L. Fox
      - J. Leroy Fox
      - James L. Foxx

---

next topic: