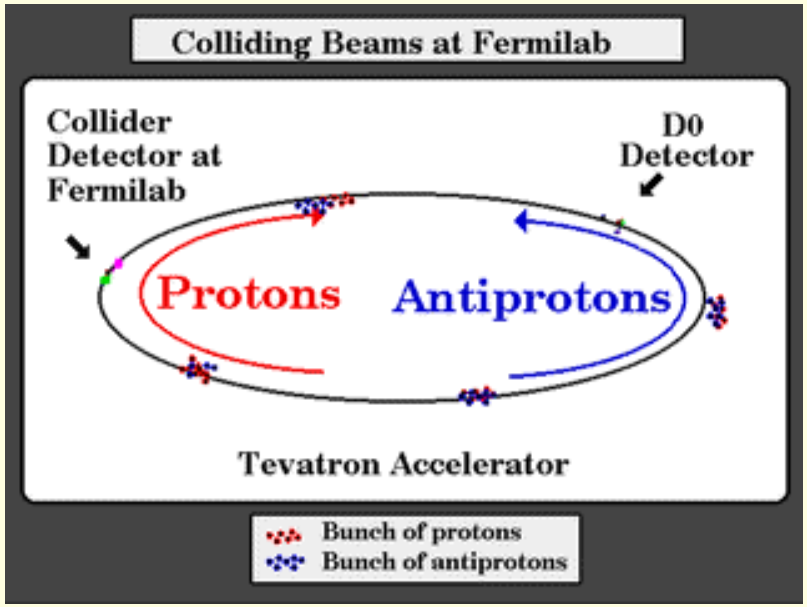
Fermilab

Fermilab Computing Division's Information Management System

Penelope Constanta, Ruth Pordes, Mark Kaletka,
Matt Arena, Joy Hathaway, Marcia Teckenbrock, Jack Schmidt,
Shirley Jones, Norman Ho, Eric Vaandering*

Fermi National Accelerator Laboratory

*Vanderbilt University



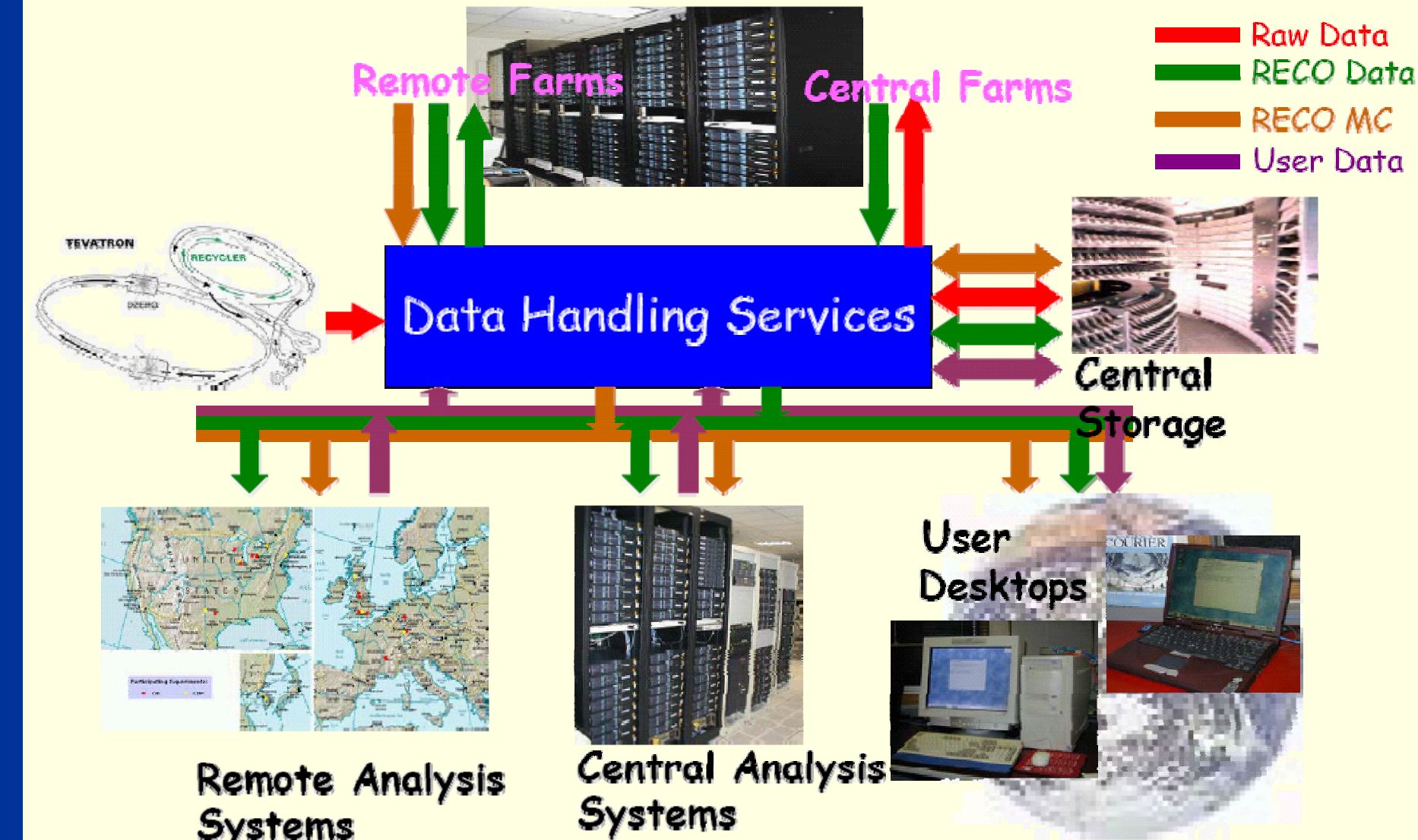
High Energy Physics

- Is accomplished by huge worldwide collaborations
- Produces petabytes of data (3.3 petabyte of data currently in robotic storage at Fermilab)
- Has enormous computing demands that can only be met by using computing resources globally
  - from desktops to physics department servers to big computing centers
- Interchanges data, ideas, people between sites worldwide on a daily basis
- Is leading other sciences in making Grid computing as important and useful for scientific research as the Web (invented in HEP) now is for all of us

Computing Division

- Scientific Computing
  - from raw data handling to storage to analysis
  - Fermilab is one of the leaders in the Grid area. Grids are a way of organizing and interfacing to computing and storage resources so that they can be shared
- Non-Scientific Computing – plain old IT
  - support all of our work at the lab
    - Business Systems
    - Desktops, Servers, Printers, Mail, Backups, Databases, Helpdesk
    - Building controls systems
    - Windows, Unix, Mac systems
  - We strive for all of this to be well managed, with appropriate levels of protections and controls

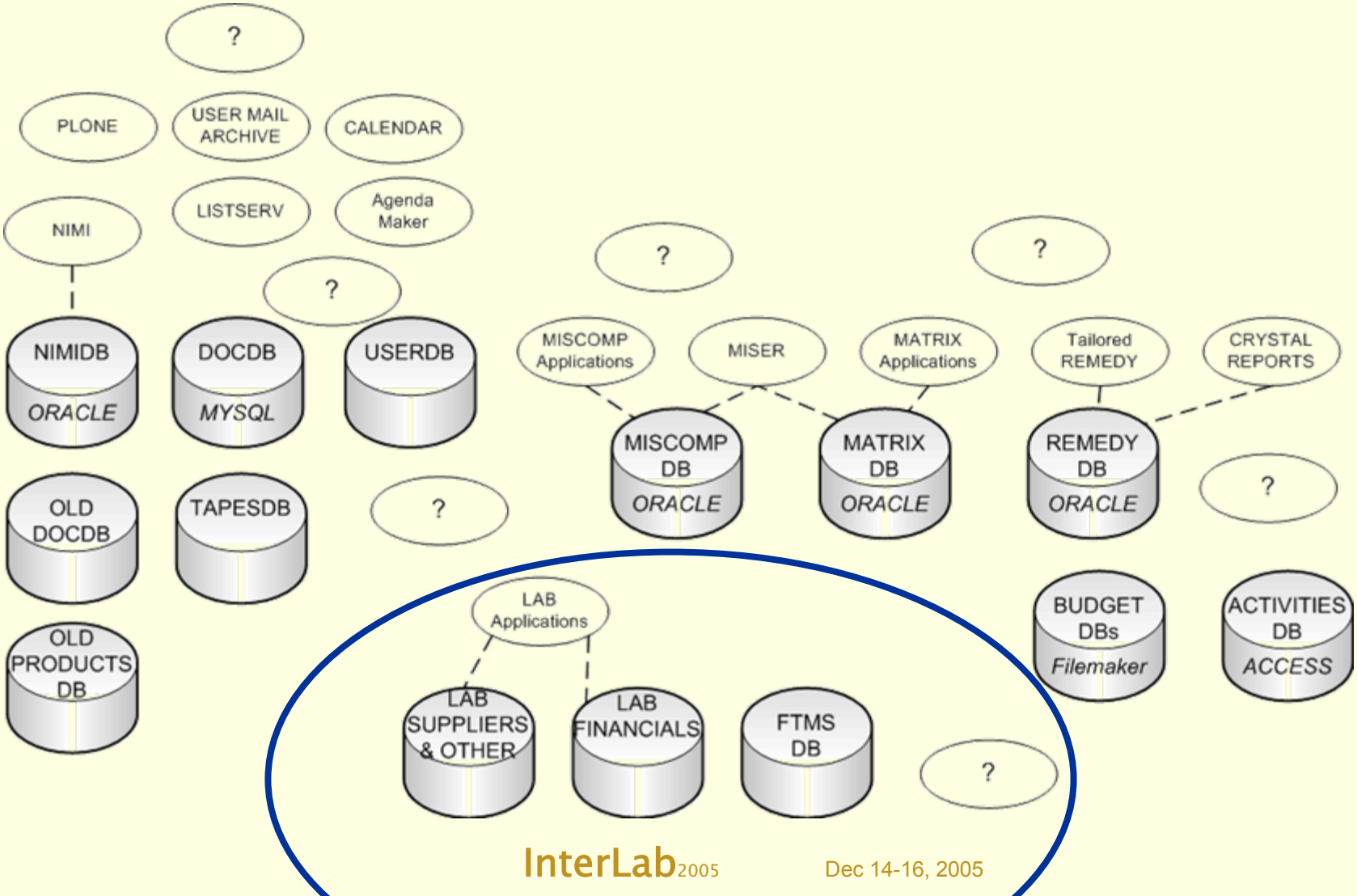
End to End Computing



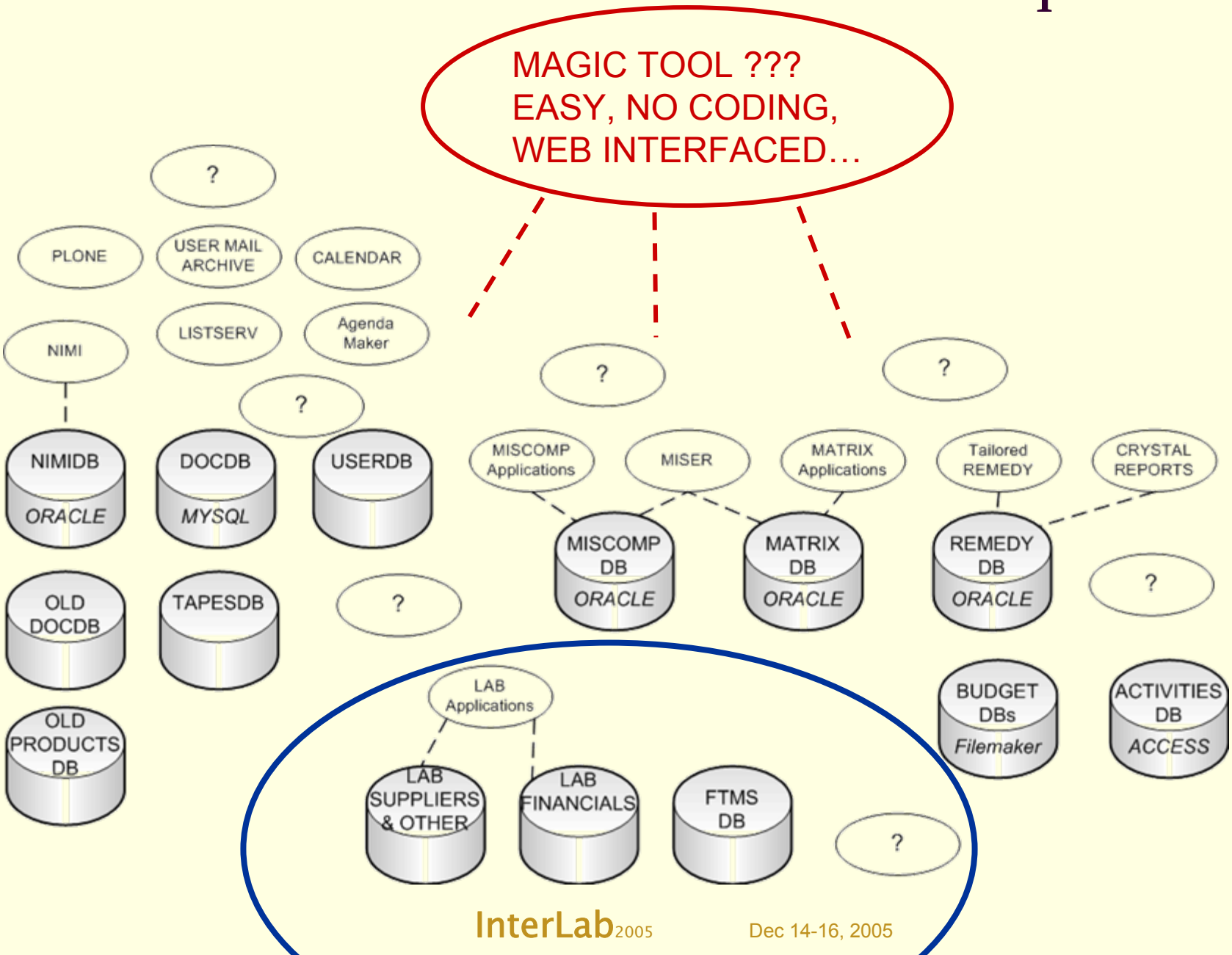
Non-Scientific Computing

- CD personnel use a variety of systems: Unix, Windows, Mac
- There are a lot of computer professionals who are trying a lot of new software packages
  - If they like the package they might use it until the next "cool" package hits the market
  - If they cannot find one easily, they simply develop a piece of software for *"temporary personal use, just to get by in the short term…"* (…nothing more permanent than the temporary)
- So, since most of the times people do not propagate information about their work, we end up with:
  - software that replicate's information
  - software packages that do not comply with the security requirements
  - software that uses obscure tools, that only one person knows how to use

Current CD Infrastructure Systems



What the division head requested



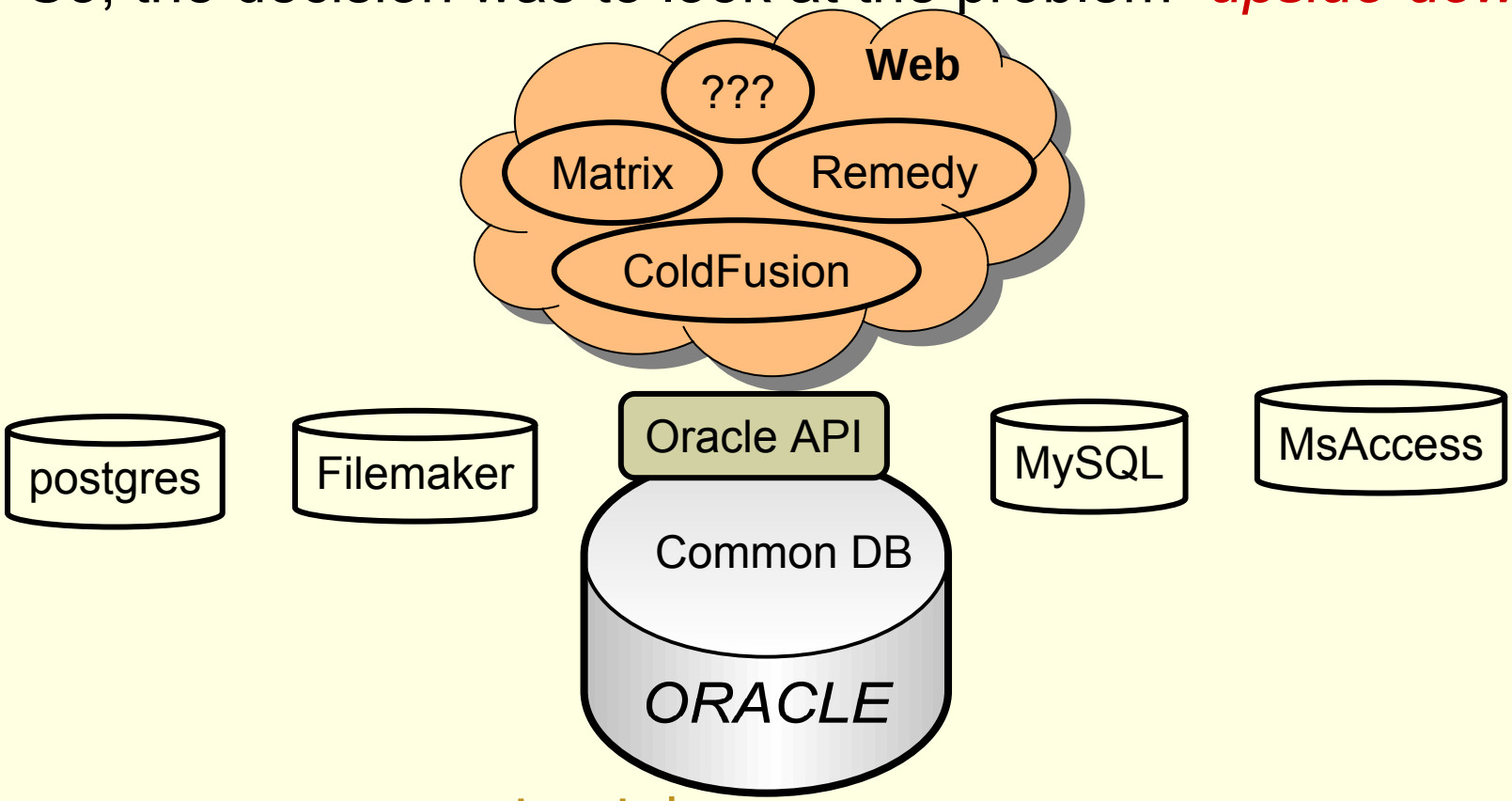
CD Information Management System

Grant name for a practical project

- Evolve, in a pragmatic way, existing workflows towards automation and paperless implementation of all Computing Division processes.
- Provide to the computing division users a uniform and coherent access of all information stored in current and future database repositories, so that the underlying structure is transparent to the user and data from the various individual repositories appear according to its logical use.
- The approach will be to involve the people who run the Division's business, to implement in a practical way and provide sensible and useful results so, as a Division, we will do our business in a professional way and not adhoc (as we are now).

CDIMS – Practical Approach

- Looking for “Magical Tools” that can interface with Oracle, Ms Access, Filemaker, postgres, MySQL, etc was not too fruitful
- So, the decision was to look at the problem *“upside-down”*



Pilot Systems: **Effort Reporting**

- Lab requires employee's to report each month on what activities they worked on.
- CD - Effort Reporting System
  - Uses ColdFusion to interface the Access DB to the web (http://wwwserver2.fnal.gov/cfdocs/effort/)
  - Provides a list of Activities that have a hierarchical structure - "Project manager's" view of the work
- Activity structure is independent of budget structure. Mapping tables ensure correct charging of activity reported effort to appropriate chargeable task
- Provides reports to line managers as well as the Budget Office
  - Maps an activity to a Task Number from the lab's accounting system so that effort can be charged accordingly
    - Allows for default and special Task mappings based on a person or a division group
    - Activities may have restricted access based on a person or a division group

Pilot Systems: **Effort Reporting**

- Activity to Task mapping details:

1. One to One, or Many to One mapping

   Activity a 100% → TaskNumber m

   Activity b 100% → TaskNumber m

2. One to many mapping

   Activity a x% → TaskNumber m

   Activity a y% → TaskNumber n

   Activity a z% → TaskNumber o

   Where x+y+z = 100%

3. Employee dependent Activity to task mapping

   Activity a for Employee x λ% → TaskNumber m

   Activity a for Employee y μ% → TaskNumber n

   Where all effort reported by employee = 100%

Pilot Systems: **Budget Input System**

- A web based system for creating the Computing Divisions budget.
  https://cd-entreport.fnal.gov/budgetinput-fy06prod/budget/login.asp
- Uses asp pages to interface the Access DB to the web.
- The asp pages were created using CodeCharge
- It is coupled with the Effort Reporting system and it feeds a number of Filemaker databases.

Pilot Systems: **DocDB**

- What is a document:
  - User view: Everything is a document
    - Text, figures, photos, sound, video, html, CAD, scanned scribbling on a napkin, …
    - Conference talks, publications, group meetings, vendor quotes, living documents, …
  - Technical view: Metadata + files
    - “Unlimited” number of types of files per document
    - “Unlimited” number of versions per document
      - All versions retained
  - User location: all over the world

http://cd-docdb.fnal.gov

Pilot Systems: **DocDB**

- Underlying technology
  - Metadata: MySQL database
  - Documents are files in a Unix filesystem
    - Actual files are stored, not urls
    - One subdirectory per version of each document
  - Web interface for viewing, adding, modifying and searching documents using Perl and cgi
  - Web interface for administrator functions
    - Administrator does not need to know SQL etc.
  - Public documents are open to the web. Accessing private documents requires a certificate or a username/password

Pilot Systems: **Leave Usage System**

- Provide a paperless way of requesting and approving:
  - vacation, sick, business etc time,
  - *automate the Domestic and Foreign travel processes to a paperless system that will automatically update, push data, notify, and enable electronic signatures.*
  - *Interface with the FTMS database for Foreign travel requests and if appropriate with DocDB*
  - *Interface with the Effort Report system to fill the appropriate percentages for the vacation, sick, etc time off.*
  - *Incorporate the paperless Directorate approval workflow*
- The package uses Oracle and plsql to interface to the web https://appora-dev.fnal.gov/pls/cert/leave_admin.leave_request.html

Pilot Systems: **Security**

- Fermilab is committed to use Kerberized certificates:
  - Leave Usage System can only be access with a valid K-certificate
  - DocDB uses, for the time being, DOE certificates and username/password
  - Effort Reporting System uses only username/password
  - Budget Input System uses only username/password

Final (future) system

- We are working so that:
  - all common information such us personnel, user, equipment etc information will be stored and maintained in one place – the Oracle database
  - all systems will cooperate and feed appropriately each other, for example the Leave Usage will feed the Effort Reporting with the vacation, sick, etc time that a person reported; the budget input will feed the Effort Reporting with any Task changes; the Leave Request will feed DocDB with conference papers so that the traveler can be reimbursed etc.

Penelope Constanta
Fermilab
MS 370
PO BOX 500
Batavia IL 60510

penelope@fnal.gov
(630) 840 8463