To Your Health: Software Development in Genentech Research and Early Development (gRED)

Erik Bierwagen

Genentech


AND NOW FOR
SOMETHING
COMPLETELY
DIFFERENT

Bioinformatics and Computational Biology

- Scientific Software development/engineering

- Big data
- Large, distributed computations
- Statistical analyses
- Algorithmic development

gRED Mission

Develop innovative therapeutics for significant unmet medical needs.

- Oncology
- Immunology
- Metabolism
- Infectious Disease
- Neuroscience

Personalized Medicine

Personalized Medicine

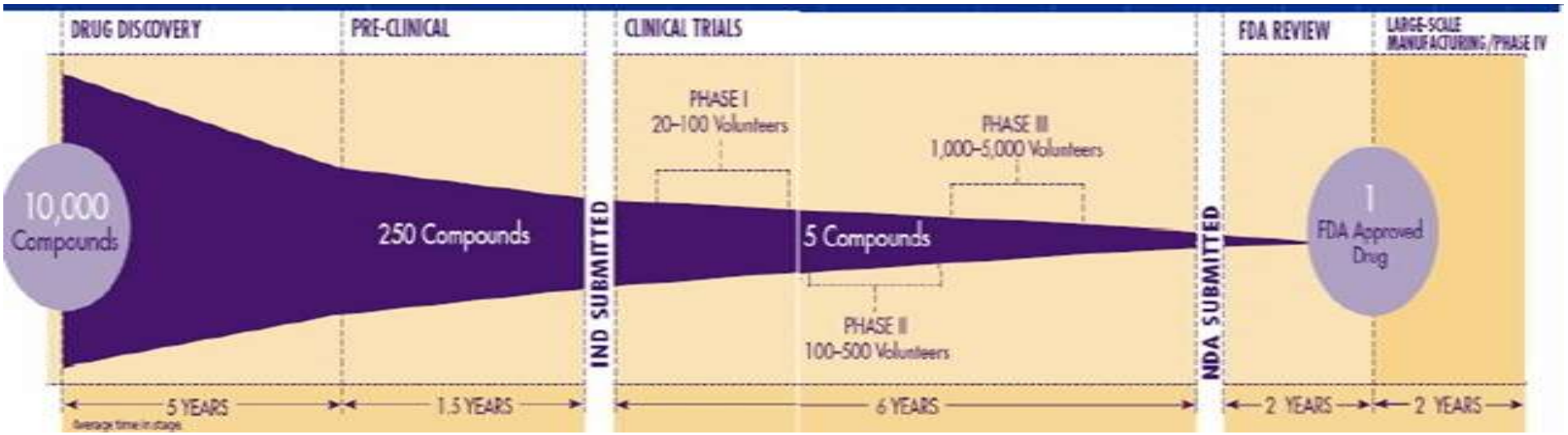
Right Drug to the **Right Person** at the **Right Time**

- Understanding of genetic pathways and protein interactions
- Understanding of genetic variants and their consequences
- Understanding of therapeutics with respect to genetic variants

Overview of Drug Development cycle

Research

gRED



Investigational New Drug (IND): Animal Pharmacology and Toxicology Studies

BIOINFORMATICS & COMPUTATIONAL BIOLOGY

Translational Medicine

- *The translation of non-human research finding, from the laboratory and **from animal studies**, into therapies for patients.*
  - Wikipedia

- Research using animals is critical to our advances in novel therapeutics

How does this fit together?

Animal studies

- Understanding genetic pathways and protein interactions
- Understanding of therapeutics with respect to genetic variants
- Understand toxicological profiles of potential therapeutics before human clinical trials
- Required for FDA IND approval

Animal Electronic Health Records

Handle and treat animals as humanely and ethically as possible

- How?
  - Track breeding of animals (rodents)
    - Control genetics
  - Track clinical information of animals
    - Understand disease response to therapeutics

Health Sciences Software Development

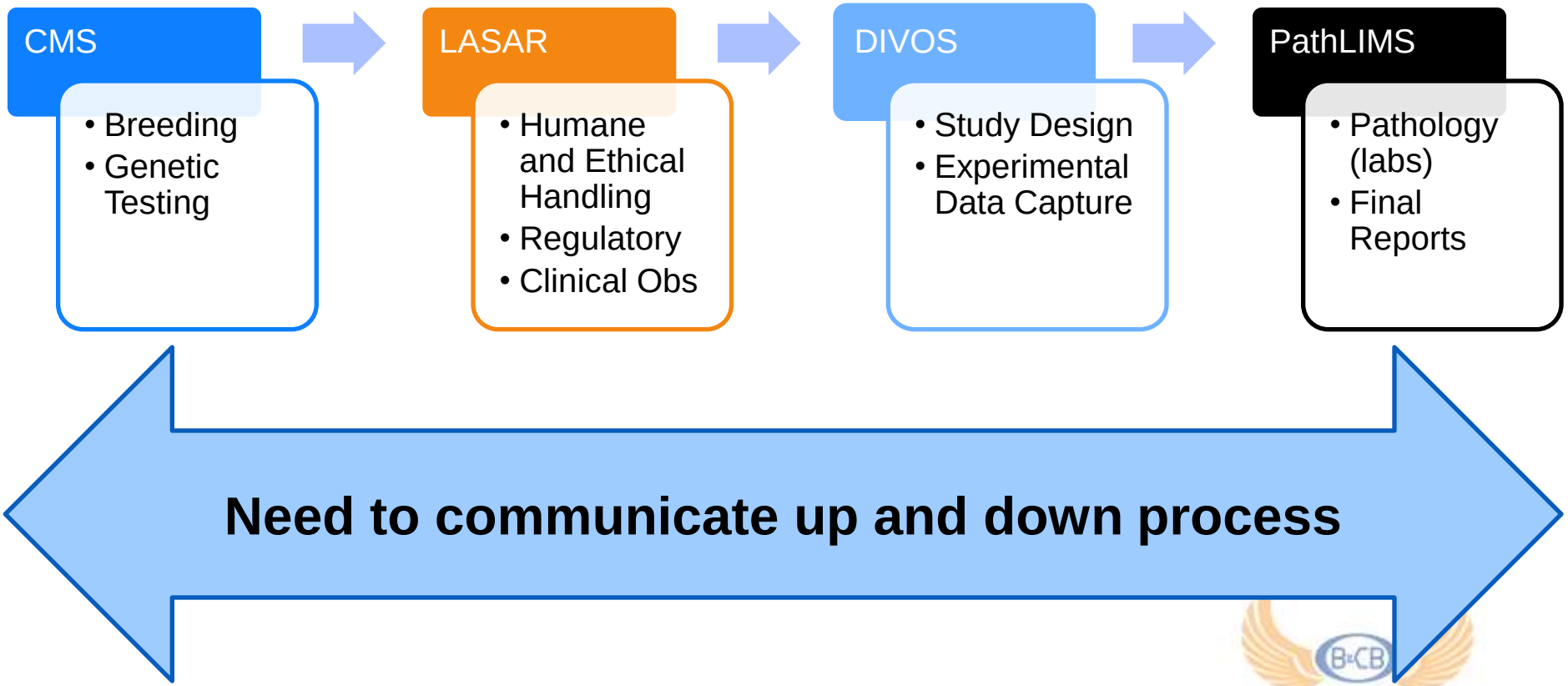
- What do we worry about?
  - Semantics
    - COLD
  - Measurements
    - Error, Units
  - Flexibility
  - Computability
  - Handling data: scientists can focus on science

Landscape

- Have a number of different systems that manage different aspects of the animal lifecycle
- Tuned for different purposes
  - Manage Breeding
  - Manage regulatory information
  - Manage experimental information
  - Manage pathology related information
- Key information captured in each one

Suite of Applications

Each purpose-built to ensure specific operational work gets done:



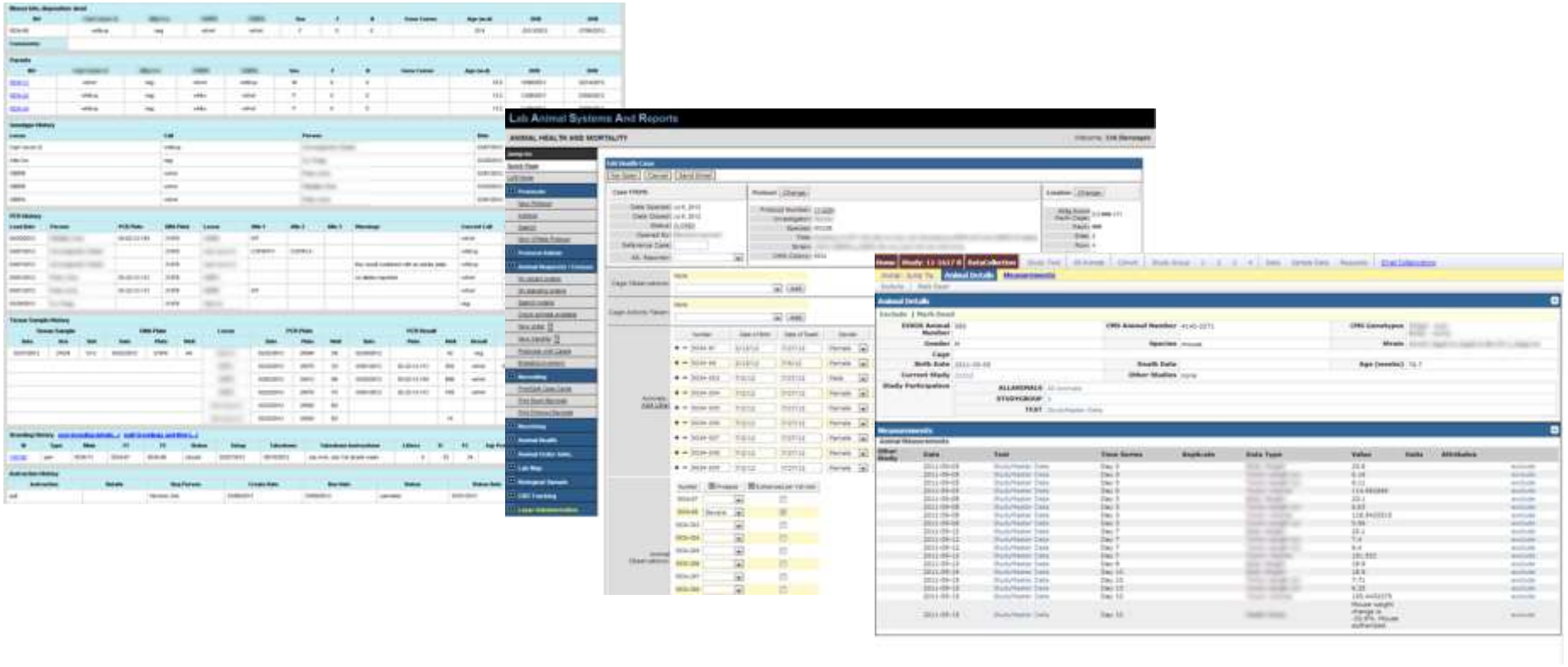
Goals

- Have a unified set of information
- Eliminate redundant data entry
- All systems talk to each other
  - Work in appropriate system

- Be able to assemble a “Health Record” from information in each system
- Compute on the data we gather

How do we think of a Health Record?

- Context specific



CMS LASAR DIVOS

- Connectivity

BIOINFORMATICS & COMPUTATIONAL BIOLOGY

Basic Components of Health Record

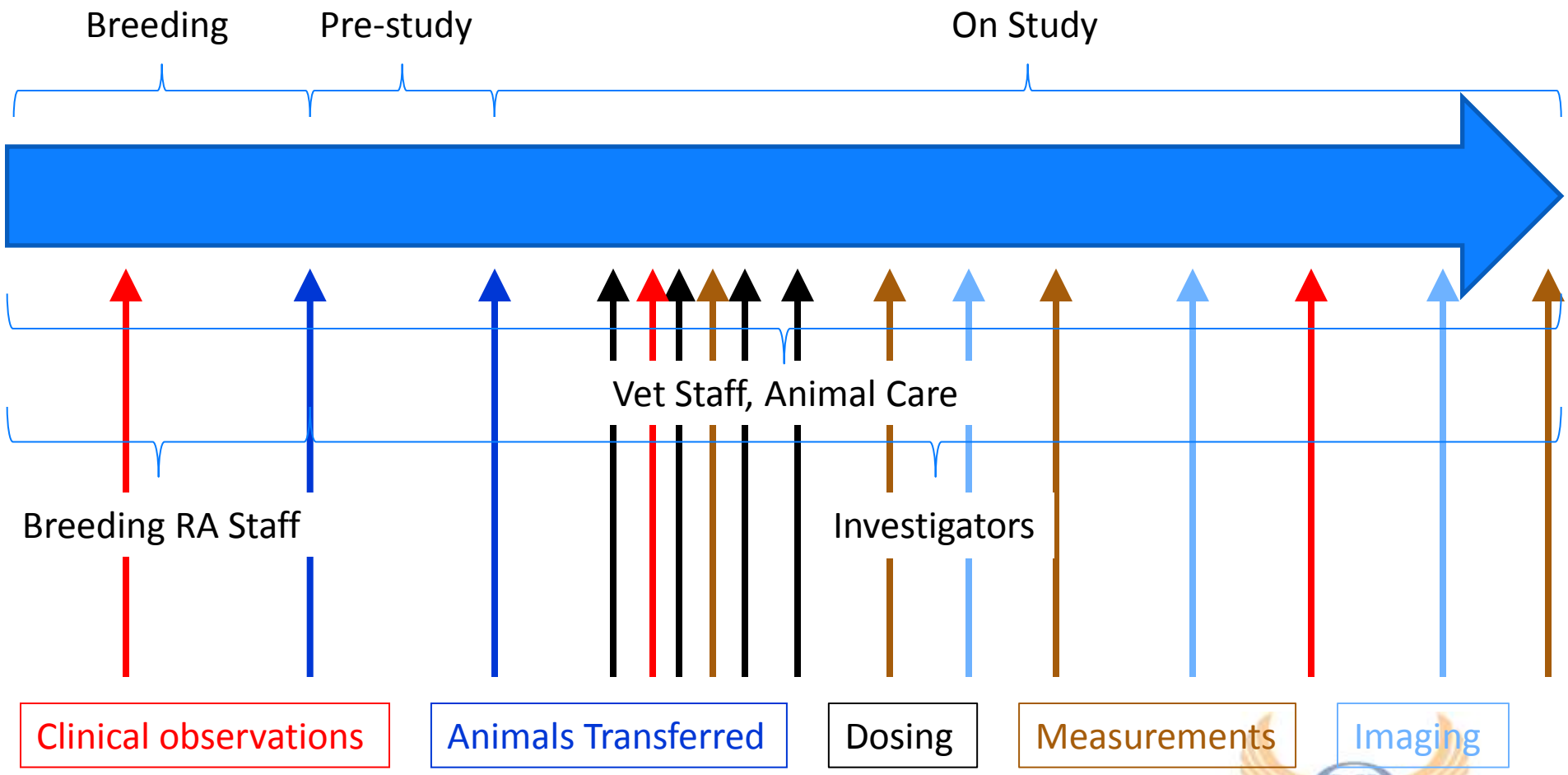
- Animal information: demographics
  - Birth, death dates
  - Strain
- Genetic information
  - Genotypes
  - Pedigree

- Clinical observations
- Location history

- Study information
- Experimental Data

- Clinical information
  - Lab work

Different people, different activities along animal lifespan



Can be several years long!

Challenges

- Ease of data entry
- Easy aggregation
- Communication between systems
- High data quality
- Flexibility of data structures
- Flexible display
- Ease in searching

CMS

LASAR

DIVOS

CMS

- Breeding and colony management
  - Central facility where all physical work performed
  - People managing the colonies/requesting work spread out over multiple buildings/campuses
- Genetic testing: control genetics
  - Samples need to be sent from breeding to central labs
  - Analysis run on machines: need to get data into system

CMS: Ease of data entry

Colony Management: 2 distinct user entry cases

- Work planning
  - Find specific animals
  - Plan work
  - Work with large sets of data at one time
  - At desk
- Work Execution
  - Working in the facility
  - Small amounts of data
  - Tied to physical objects
- Java application
- Mobile

Breeding, Genetic Testing

BIOINFORMATICS & COMPUTATIONAL BIOLOGY

CMS: Ease of data entry

- Mobile Application

- Physical demands
  - Animals live in clean-room environment
  - Need to know where animals are in facilities
    - Multiple buildings across numerous campuses
    - Cages in racks in rooms in buildings

CMS: Ease of Data Entry

- Barcoding
  - Map physical and logical worlds
- Portable
  - PDA/Mobile devices (on 3rd generation device: iOS)
  - Browser based
  - Wireless (challenging!)
- Simplify
  - Processes tuned automate as much as possible, minimize data entry

Breeding, Genetic Testing

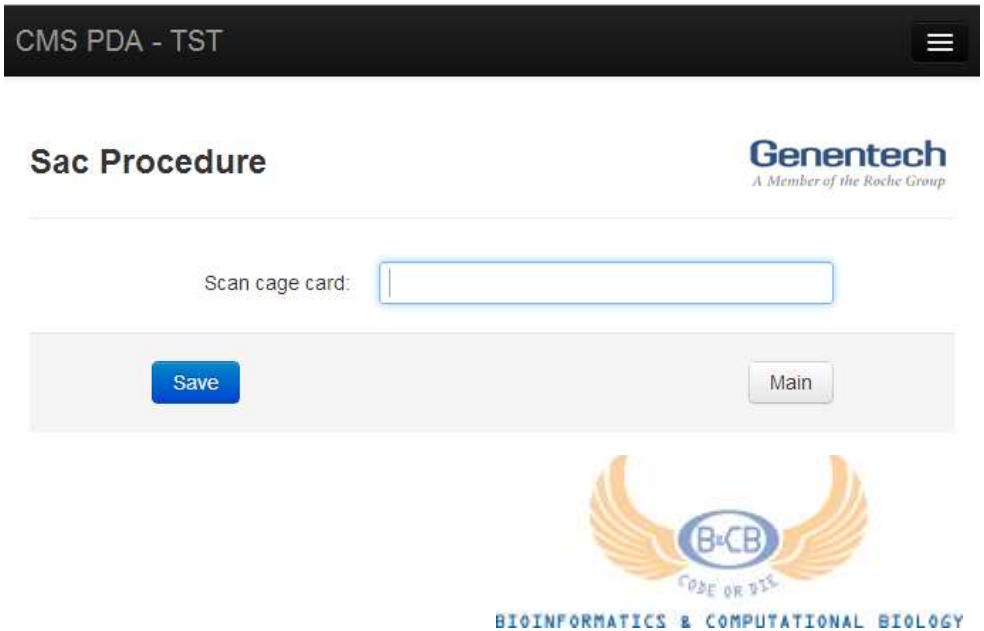
CMS: Ease of Data Entry

- Mobile interface considerations
  - Distinct processes
  - Scan to start process
  - Simplify data entry as much as possible

Breeding, Genetic Testing

CMS Mobile Application

- In transition currently
  - From: fixed device layout
  - To: responsive web design



BIOINFORMATICS & COMPUTATIONAL BIOLOGY

CMS: Ease of aggregation

- Need
  - Manage at many levels
    - Animal
    - Colony
    - Facility
  - Precision
  - Computable information

Data Needs

- High data complexity
- Transactional complexity
- High consistency needs
- ACIDS
- Low data/transactional volume

- RDBMS

CMS: Aggregation Examples

- Real time fecundity
  - Fecundity: measure of the number of children that survive past weaning
- Look for imbalance of genotypes in offspring
  - Counts vs. standard Mendelian ratios
  - aA x aA: ¼ aa + ½ aA + ¼ AA

Basic Components of Health Record

- Animal information: demographics
  - Birth, death dates
  - Strain
- Genetic information
  - Genotypes
  - Pedigree

- Clinical observations
- Location history
- Study information
- Experimental Data
- Clinical information
  - Lab work

LASAR

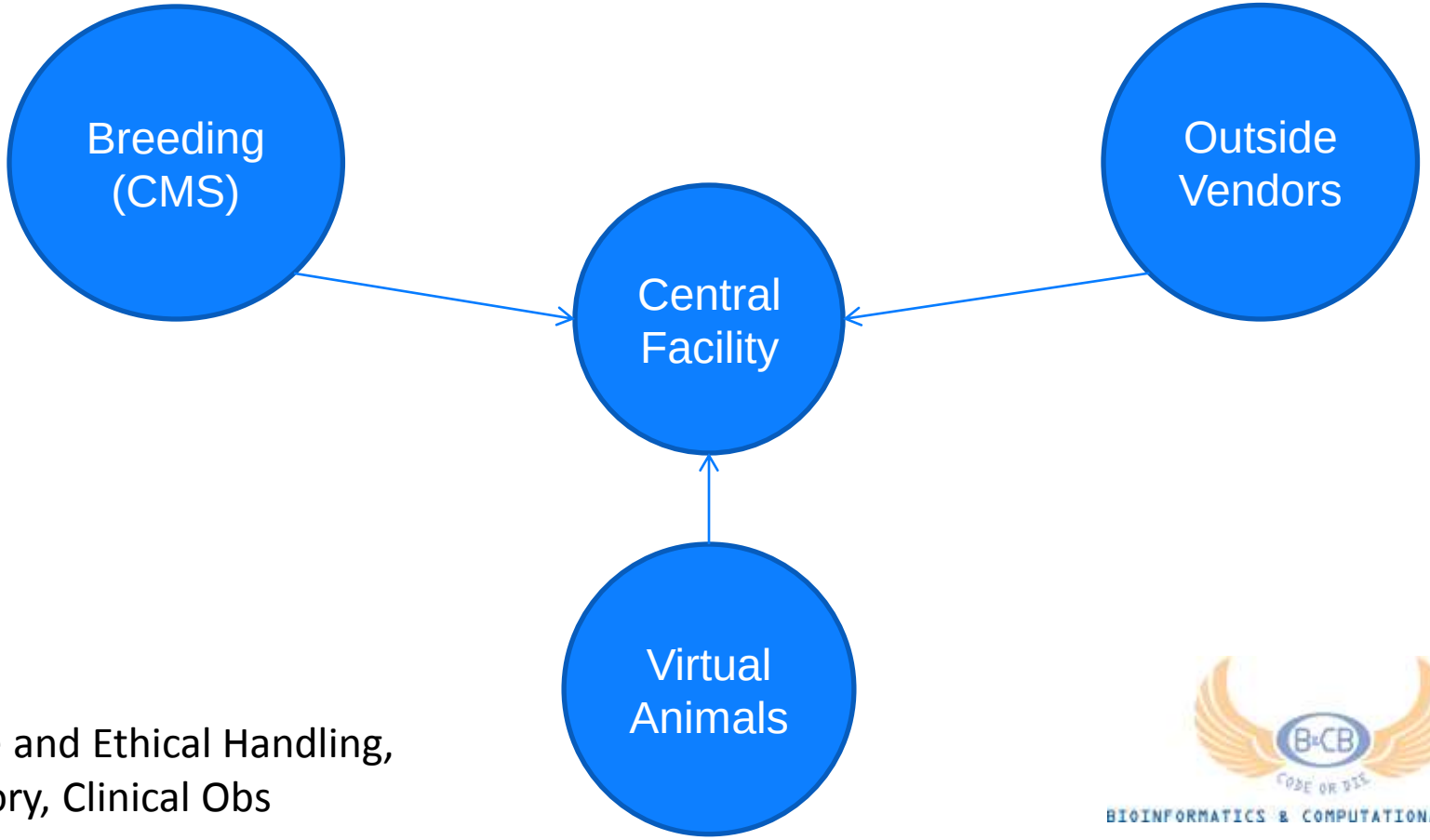
- Humane and Ethical handling of animals
- Regulatory compliance
- Clinical Observations

- All animals are managed by this application
- All animal use covered by IACUC (Inst. Animal Care and Use Committee) protocols

Humane and Ethical Handling,
Regulatory, Clinical Obs

LASAR

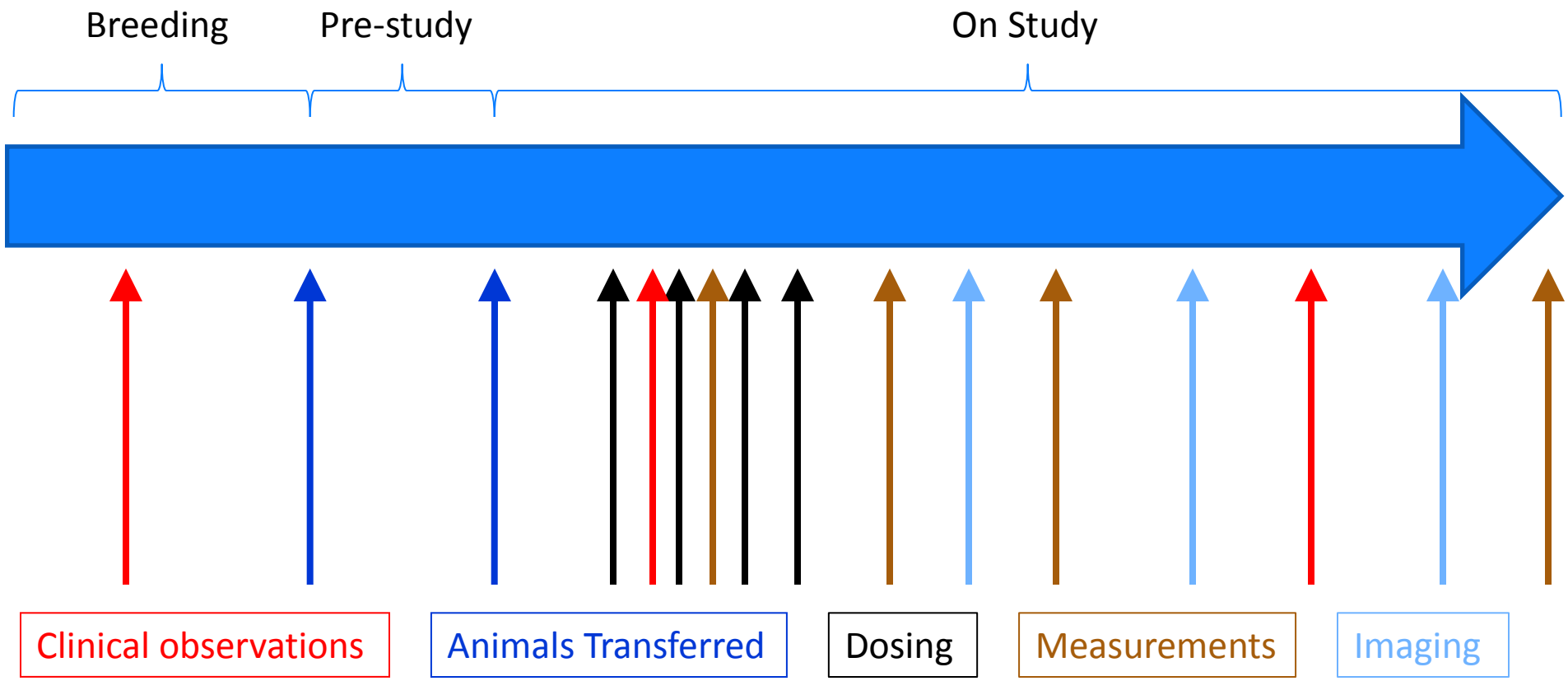
- Many sources of animals



Humane and Ethical Handling, Regulatory, Clinical Obs

BIOINFORMATICS & COMPUTATIONAL BIOLOGY

LASAR

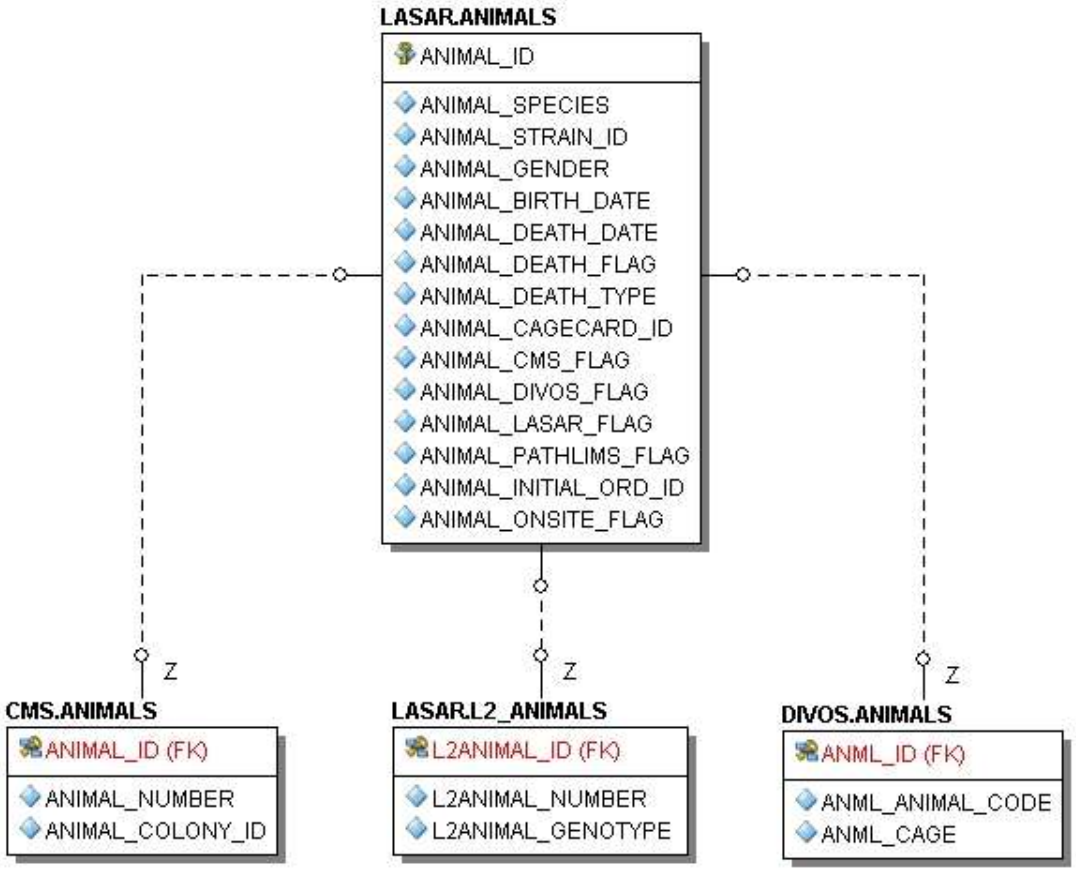


Humane and Ethical Handling,
Regulatory, Clinical Obs

BIOINFORMATICS & COMPUTATIONAL BIOLOGY

LASAR: DB Integration

Single globally unique identifier



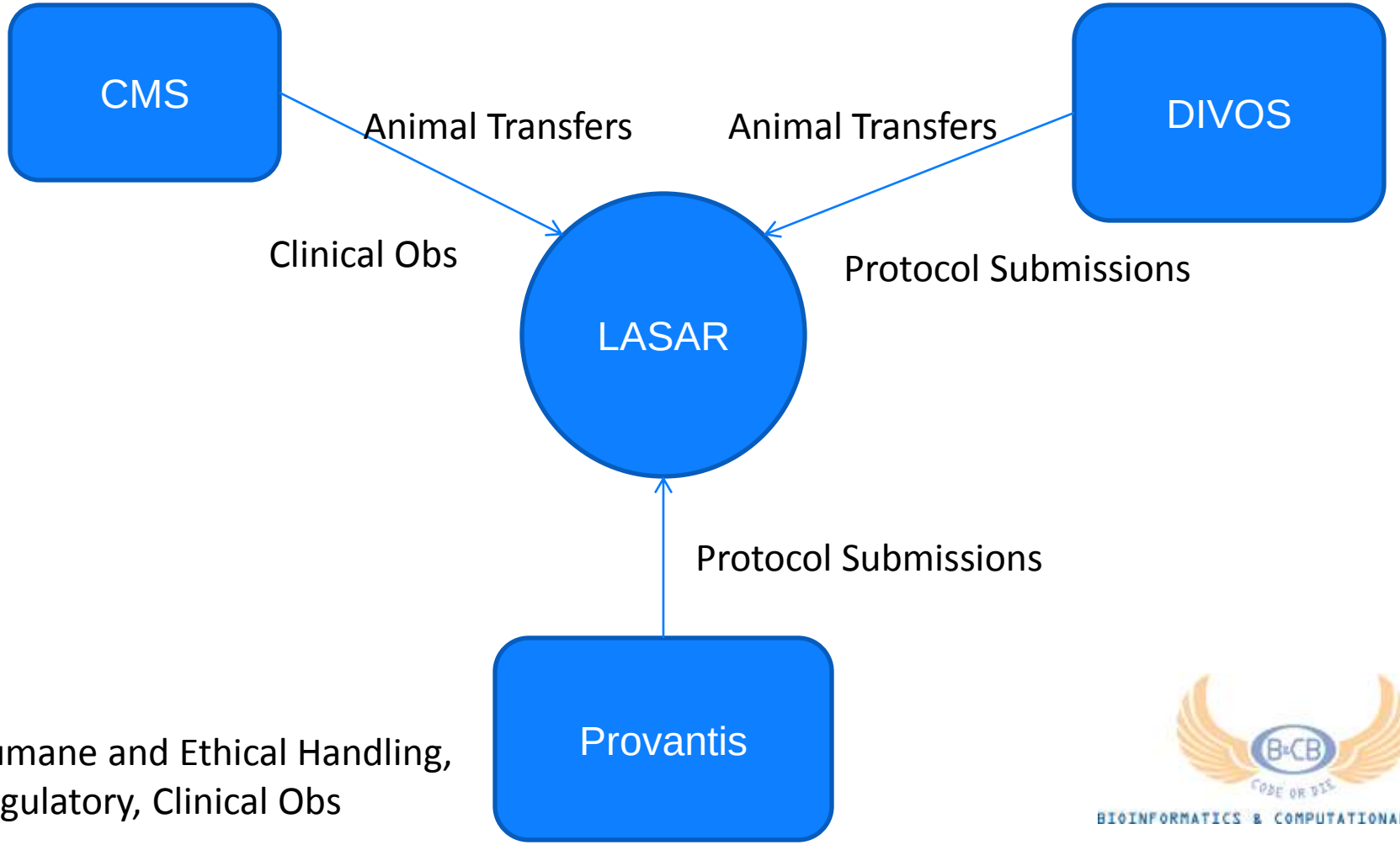
LASAR

- Central point for all animal handling
  - Manage animals coming in and moving around
    - Locations
    - Protocols
  - Superset of functions that other applications use
    - CMS
    - DIVOS
  - Expose services to other applications

Humane and Ethical Handling,
Regulatory, Clinical Obs

LASAR: communications

- Service based



Humane and Ethical Handling, Regulatory, Clinical Obs

Basic Components of Health Record

- Animal information: demographics
  - Birth, death dates
  - Strain
- Genetic information
  - Genotypes
  - Pedigree
- Clinical observations
- Location history
- Study information
- Experimental Data
- Clinical information
  - Lab work

DIVOS

- Animal study design
  - Clinical trial for animals
  - Precise description for plan/execution of study
- Experimental data capture: measurements

- Need flexible system
  - Many (hundreds) of different types of experiments
  - Need to display data in a matter meaningful to class of studies

Experimental Reproducibility

- Describe experiment
  - Pre-conditions (leading up to experiment)
  - Conditions
  - Measurements
  - Values
- Need consistent data semantics
- Critical component of scientific research

In 2012, a study found that 47 out of 53 medical research papers on the subject of cancer were irreproducible.

DIVOS: Flexible data structures

Neurobiology

- Alzheimers Disease
  - Experiments
    - Balance beam
    - Gait test
    - Memory test (maze)
    - Psychological test (open field)
  - Brain imaging
  - Dosing of therapeutics

Oncology

- Pancreatic Cancer
  - Measurements
    - Body weight
    - Tumor size
  - Dosing of therapeutics

Study Design,
Experimental Data Capture

DIVOS: Flexible data structures

- Data needs listed above: RDBMS
- Need for computation: atomize data
- Flexible structures:
  - Entity Attribute Value (EAV) structure
  - Ability to handle complex relationships
- Rigor in data semantics

DIVOS: Flexible display

- Immunology

- Oncology

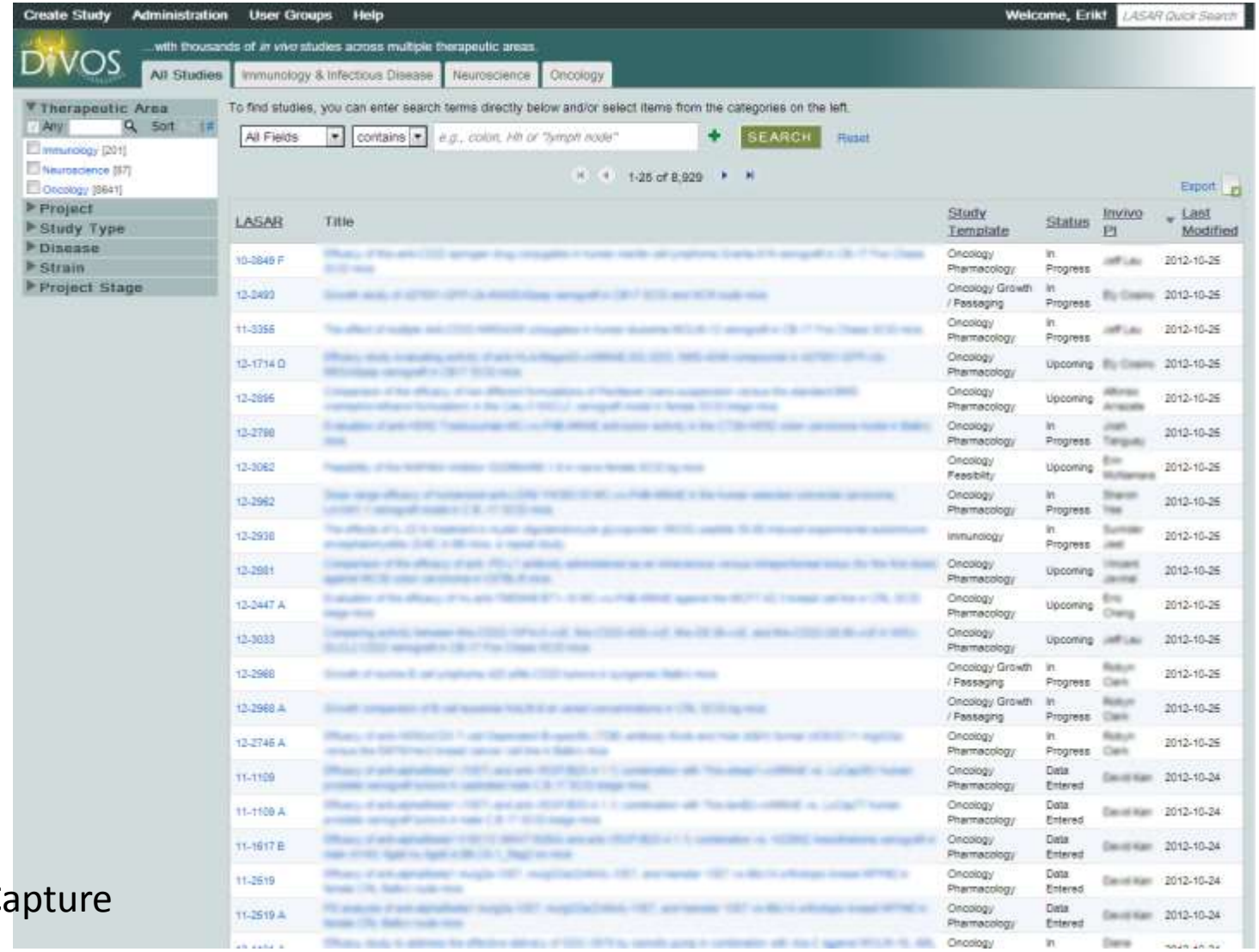
Study Groups

Edit

Grp#	n	Group Description	Test Article	Dose (mg/kg)	Vehicle	Route	Dosing Schedule	Model
1	18	01 - Vehicle (MCT)			MCT	po	oncexUNKNOWN	
2	22	02 - [redacted] @ 1 mg/kg	[redacted]	1	MCT	po	oncexUNKNOWN	
3	20	03 - [redacted] @ 10 mg/kg	[redacted]	10	MCT	po	oncexUNKNOWN	
4	20	04 - [redacted] @ 50 mg/kg	[redacted]	50	MCT	po	oncexUNKNOWN	

DIVOS: Ease of searching

- SOLR with Faceting



Study Design,
Experimental Data Capture

Basic Components of Health Record

- Animal information: demographics
  - Birth, death dates
  - Strain
- Genetic information
  - Genotypes
  - Pedigree
- Clinical observations
- Location history
- Study information
- Experimental Data
- Clinical information
  - Lab work

PathLIMS

- Pathology Labs
- Final Reports

- Currently not explicit link (via Animal ID)
  - Have to infer

Challenges still

- Integrating other systems into this suite
  - PathLIMS
- Samples (blood, tissue)
  - Describe collection strategy
  - Describe complex relationships precisely
  - Homogeneous description
  - Service spanning applications
- Experiments on samples

Lessons Learned

- Work with the right users
- Describe the science as correctly and completely as possible
- “Software development” is
  - Process re-engineering
  - Social re-engineering
  - Software engineering

Acknowledgements

- Doug Garrett
- Dana Caulder
- Wendy Kan
- Michael Vogel
- Jimmy Yu
- Joe Mulvaney
- Pierre Monestie
  - LinkedIn
- Norman Chan
  - Tapjoy
- Dairian Wan
  - Zynga
- Dake Wang

BIOINFORMATICS & COMPUTATIONAL BIOLOGY

Thank You!

BIOINFORMATICS & COMPUTATIONAL BIOLOGY