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

**Erik Bierwagen** 

Genentech





### **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





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



#### **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:





- 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

		-		_	-				A REAL PROPERTY.																		
				-	-				and the second	-																	
-		_	-																								
-																											
44																											
	-		e		an	-			Canal Contract of	Antes .		-															
	100			-	-		811			44		name.															
				***	-					110	100000 0	and a second															
				H	-					100	interesting in the	and a second sec															
COLUMN TWO											Lab Animel Svi	terms And Reod	ette														
TRAP INVITE			120			100				101-0																	
Participant of							-				ANDAR, HEALTH HEE	BACHERIN, TY								serve a	IN DOTUGATO						
<u>-</u>											lant fam	Address of the local sector of the local secto	1000								- 10						
£			-							10000	Witne	Lie lite. Den	er   David Street														
			-				and and					days stream			and the second se					And Street, Taxabase I							
			-1014							-	March Street				Country of					Contract of Contract,							
											A REAL PROPERTY AND INCOME.	Then been	a	10	water and a state of the	10				COMPANY	aire l						
-	-		Contrast 1	And a	-	-	1.000	-		Constant of the	- America		41 A.040		Contraction of the local division of the loc	100				Table on							
00011			1000								tion of the lot of the	Contract in	0		Tree .				Contraction of the local division of the loc	Contract of							
	_		1.000		Contraction of the	-				1.000	COLUMN TWO IS NOT	Selector Dat	(4)		10.00					April 4							
_			1		COLUMN -	-		A				AA, Amore		141	Des Calerar 10	14 C	TRANE P	the second s	and the second second	And Address	Long La	and have a state	a han been	Data Antonio 1 40	and shares		
-								10.000			Contract Name of Contract of Contract		1				and in	Contraction in the	Contraction of the local division of the loc								
-	-										W.BIALDING	Case 104months					1.1.1	And and a local division of the local divisi	County of the local division of the local di								
										-star	D.MARLENS	S. 19.71			341 1,4462		(Access)	TOWN OF THE OWNER									-
- 10 C	-		1000	1000						148	bestodes .	E HERRICE	100				_	TVI III									
the local man											COLLEGE AND A	Lagrance Gar	1.		(A) (446)		14+Dade	a / Narh-Seal									
ince he						-			and in case		DIVISION IN		10 10 mm		And Annual and	The second second		CONTRACTOR AND ADDRESS OF			The Assessed In	august -4140-0070		CPRE Exercit	ABOUT DESIGN		
Add		1846	then into			-	-	- 840	-	And incid	Instantiv 1				and a second second			Second in				and the second					
	1 . 111	manter .			1000	c ) (100	1.281	-		N	Conception of the local division of the loca		• • pass at	3/18/16	1 1000	interest (e)		Case									_
				1.00	-			-	Ballion and	44	Tanking or 1		4 - 3044 44	1000	Marie .	Parate al		THE R. LEWIS CO., LANSING, MICH.	0.00		5445	h Dete ;		Age (Not	101.0466		
				1.00	1.000			-	Brown and a state	and the state			+ + Site 117	1010	100004	1000 147		Correct Made			Other 8	failles your					
								-	-		Partial Jam Jam		A REAL PROPERTY.	-		110 100	ittedy	Participation	ALLANDPALS	Al annals							
											Colden.Acres	Acces	a - part of	- Frank	Partia	10.00			STUDIO COLUMN	(a)							
					1						distinguistions.	Carlos Carlos	R. s = boss on	1010	hirrot	farms (a)			10,01	Southeast Care							
					-						THE REAL PROPERTY AND INCOME.		+ + 1000 AM	1000	Seattles.	Parele (a)	-				_	_	_	_	_	_	_
-	a break a beau	BLI MERINE	ALC: NO. OF CO.	11							mine strengt			THEM.	institut.	Parties Tal	-										
<ul> <li>tet</li> </ul>	-			-	. 14	-	hands	-	Address - Bit	1.00	and the second second		A CONTRACTOR			the lot way	and the second s	An other Designed in	100		and the local division of	in interim	the second se	1000		NAME AND ADDRESS OF	
a	1000	314F 4			-				1.1	11	the second second second		100000	Calls.		10000	marily		- The second sec		the later of	- Bagiltines	Bada Tanto	Value	Parity .	and an and a second second	
a la											H In Ma		• - 369+300	17412	102013	1000 20	1	1011-09-02	middlener in		1			111			-
And optimal in which the		-	marian				1 Barrison	-	-	. Take bet	In summer land		And the		Contract Manager			1011-09-08	Dational Da		6.9			÷			action.
			retest des	1000	and i	1.000	-	1.00		and the second	ALC: UNK DOCTOR		and a second second	122				Disc Birds	Paddinane De	1 1	10		States and states	10446044	1		-
											COLUMN TWO IS NOT		Cold Digerter	- 26	-			101109-08	Rub/terile Ca	4 5	51		and a state of the	8.83			-
											No. of Concession, Name		month   parent	100				301119-10	Publicated in	**	e 0			110.040000	4		0.000
														141	<b>F</b> .			2011/06-12	Party Hellow De				and the second s	10.0			-
													HO-D4	100	10			10100-02	Nutry Sector De	4	. 7		Contraction of the	7.4			C. BALLAN
												100	(ma-34)	142 -	10			3011-09-12	Rub/tena: 24		97			8.4			
												1044	demont	141	10				Contraction of the				and the second	10.0			- second
													The second second		100			100.00.00	multiples of the		1.14			18.1			and and
														181	(D)			1011-09-08	Waterie 24	0a. 70	10.10		The second in	(1)(1)			-
													and a late	12	10			1011-09-08	Buddhees be		0.11		the second second	9.22			
																		0011-00-18	Party Party In	- BA	N 10		The second second	135, 440,257	A		
																								those way	No.		
																			Providence (re		6 (B)			-16-27-6, 745	1.000		-

CMS

Connectivity

LASAR



## **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



## 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



Breeding, Genetic Testing

#### **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





Breeding, Genetic Testing

#### **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



Breeding, Genetic Testing

## **CMS: Ease of Data Entry**

- Barcoding
  - Map physical and logical worlds
- Portable
  - PDA/Mobile devices (on 3<sup>rd</sup> 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**

Home CMS P	DA Logout									
Select a process										
Breeding <u>New Litter</u> <u>Clip/Tail</u> <u>Wean</u> <u>Final Count Update</u> <u>No Toe/Tail Wean</u> <u>Setup</u> <u>Takedown</u> <u>Timed Pregnancy</u> <u>Perinatal Screen</u> <u>Re-Open Breeding</u>	Transfers <u>Exp Transfer</u> <u>Colony Transfer</u> <u>UnTransfer</u> Tech Assistance <u>General Assist</u> <u>Weight Entry</u> <u>Tumor Check</u>									
General Animal Update Cage Move Consolidate House Order iPad	Rack Management Cage Maintenance Rack Transfer Animal Health Link									
iPad Menu	CMS Troubleshooting									



Breeding, Genetic Testing

Dake Wang     1/1       Colony Transfer Procedure       Cage Card Id: CGC504119       Colony:       Investigator:       Protocol: 05-0159       Location:       Current list id:       LST40658       Processing cage:       1/2       Annil # Sex Xfer ND FD       1484-1411       F ● C       Processed and canceled animals:       Annil # SexDone Dead Reprinted       1484-1401       F ♥ ●       ETA date:       01/23/2008	Canceled
Dake Wang     1/1       Colony Transfer Procedure       Cage Card Id: CGC504119       Colony:       Investigator:       Protocol: 05-0159       Location:       Current list id: LST40658       Processing cage: 1/2       Anml # Sex Xfer ND FD       1484-1411 F • C C C       Processed and canceled animals:       Anml # SexDoneDead Reprinted       1484-1400 F • C • C       1484-1401 F • C • C • C	Canceled
Colony Transfer Procedure Colony Transfer Procedure Cage Card Id: CGC504119 Colony: Investigator: Protocol: 05-0159 Location: Current list id: LST40658 Processing cage: 1/2 Annil # Sex Xfer ND FD 1484-1411 F • C C Processed and canceled animals: Annil # SexDone Dead Reprinted 1484-1400 F	Canceled
Colony Transfer Procedure Colony Transfer Procedure Cage Card Id: CGC504119 Colony: Investigator: Protocol: 05-0159 Location: Current list id: LST40658 Processing cage: 1/2 Anml # Sex Xfer ND FD 1484-1411 F ( C C C Processed and canceled animals: Anml # SexDoneDead Reprinted	Canceled
Dake Wang     1/1       Colony Transfer Procedure       Cage Card Id: CGC504119       Colony:       Investigator:       Protocol: 05-0159       Location:       Current list id:       Current list id:       Processing cage:       1/2       Anml #       Sex Xfer       1484-1411	
Colony: Colony: Colony: Investigator: Protocol: 05-0159 Location: Current list id: LST40658 Processing cage: 1/2 Annil # Sex Xfer ND FD	СГ
Dake Wang 1/1 Colony Transfer Procedure Cage Card Id: CGC504119 Colony: Investigator: Protocol: 05-0159 Location: Current list id: 15T40658	NF Attn
Dake Wang 1/1 Colony Transfer Procedure Cage Card Id: CGC504119 Colony: Investigator: Protocol: 05-0159	
Jake Wang 1/1 Colony Transfer Procedure Cage Card Id: CGC504119 Colony: Investigator:	
Colony Transfer Procedure Colony Transfer Procedure Cage Card Id: CGC504119 Colony:	
Colony Transfer Procedure	
Dake Wang 1/1	
	8/08
CMS PDA Log	out: Cti-q

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



## **CMS Mobile Application**

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





TST

Logout

CMS PDA

## **CMS: Ease of aggregation**

#### Need

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



Breeding, Genetic Testing

#### **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



Breeding, Genetic Testing

## **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





- 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





#### Many sources of animals







Regulatory, Clinical Obs

#### **LASAR: DB Integration**

#### Single globally unique identifier







- 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



## **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





- 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



Study Design, Experimental Data Capture

### **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

Study Design, Experimental Data Capture



#### **DIVOS: Flexible display**

#### Immunology

1	Provident Contraction (Contraction)																											
<b>D</b> # 1	Group Descriptio	La Contra	Stimula I? Applier	1 Stram	Genutyp	e Drug Th	erapy												Stin	natus Regis	14231							
rice.	Tosan Di at -3.Styn, aaking Di	Yes	He	SALS-	wt/wt	Test Article	ProteinID (PRO, PUR, CNJ)	SH1 Selt	SHI Lot	Dose 5	loes Vel al.	vicle R	oute	Frequency	# of times	Ocea Order	Dosin Notes	p Notes	1									
	(3=4)										NE. NO	5.1	<u> </u>	100	•													
	100.00	iei :	ves	BALB- C	wtwr	-					1.015 Nor	• 2	N	crce	1970 -				2	0 h	100	1 up	25 ii.	P85	195	٥Ħ	unce	
	1997) ar	Na	Yes	SALD- 9	wt/w†			1	1	0.3	0.025 ho	e p	NC - 1	orce	5				3	0 ħ	er l	1.10	28 st.	PAS	DN.	0 W	0108	
	ip: 15.75	Yes.	Ves	sALD-	wt/wr						1.025 Nor	16 J	N.	unce.	1			1	4	Dħ	212*	Lup	27 st.	PEE :	19.	0.11	0/428	
3	179091. 213 141.1.1.1	No	Yes	SALS-	wt.wt	64,000	PE1	1	1	0.3 mp/sp	0.025 No	e P	N	orce	3			ji.	1	0 h	27	Lug	23 si.	PRS	Di	8 h	ana.	
	ter ta m	Yes	Ves	BALD	wT/WT														į.	0 h	Normal 1	1.00	29 st.	Pets	iN	0.h	01CE	

#### Oncology

Study Groups												
Edit												
Grp#	n	Group Descripti	on	Test Article	Dose (mg/kg)	Vehicle	Route	Dosing Schedule	Model			
1	18	01 - Vehicle (MCT	)			MCT	ро	oncexUNKNOWN				
2	22	02 -	@ 1 mg/kg		1	MCT	ро	oncexUNKNOWN				
3	20	03 -	@ 10 mg/kg		10	MCT	ро	oncexUNKNOWN				
4	20	04 -	@ 50 mg/kg		50	MCT	ро	oncexUNKNOWN				

#### **DIVOS: Ease of searching**

**Create Study** 

Project
 Study Ty
 Disease
 Strain
 Project 1



Study Design, Experimental Data Capture

Administration	User Gro	ups Help	Weld	ome, Eri	A LASA	9 Quick Search
with Prousan	ds of <i>in vivo</i> st	tudies across multiple therapeutic areas.				
All Studies	Immunology	A Infectious Disease Neuroscience Oncology				
tic Area	To find studies	s, you can enter search terms directly below and/or select items from the categories on the left.				
9, 501 1#	All Fields	contains      e.g., colur. All or "Symph roode"     SEARCH Read				
(87) 11]		01 01 1-25 of 8,920 🔸 🗰				Export _
99 9 <del>9</del> 0	LASAR	Title	Study Template	Status	Invivo El	* Last Modified
	10-3846 F	Music Provide TSS arrays the respective tasks all patients base 214 arrays to 18.77 to Date 312 may	Oncology Pharmacology	in. Progress	intim.	2012-10-25
tage	12-2499	South Aris, F. (2007) 207 - 20 Amilia State array of a 1877 2021 and 1021 and 1021	Oncology Growth / Passaging	In Progress	By Copies	2012-10-25
	11-3355	The effect of each product of the theory of the transmission of transmission of the transmission of t	Oncology Pharmacology	in. Progress	inter-	2012-10-25
	12-1714 0	Phase and installing while that in a highly called \$1,223,265,456 are provided at 2012 (24) to Michigan arrival in 2017 (22) with	Oncology Pharmacology	Upcoming	By Coarse	2012-10-25
	12-2895	Empired of the disate of the Menth Strengthene of Facilities and a second strength of the facility of the Strength of the Strengthow of the Strength of the Strength of the St	Oncology Pharmacology	Upcoming	Altrian Actuality	2012-10-25
	12-2798	1 makes if yet (10) "parameter (0) is (10) bits of some site (11) (10) (10) and present takes the max	Oncology Pharmacology	In Progress	int Tegaty	2012-10-25
	12-3062	Parallel, 2 for NATION AND A COMPANY 1 ( 1 - ages from CCC 1), not	Oncology Feesbility	Upcoming	\$10 Wolfsmann	2012-10-25
	12-2962	They may obtain a function get, (b) firstly that an full shift of the long market constant presents, and all compared markets (10, 41, 20) was	Oncology Pharmacology	in Progress	Starter The	2012-10-25
	12-2938	The Africk of L. (2.) Insertion with dependencies groupoint ACC wells 5. If report represent estimates programming the 2.40 to 80 million a report that	Immunology	in. Progress	Summer Josef	2012-10-25
	12-2981	Comprise of the Allers, of p.1, PC, 1 which, advections is a constant or the single-local local to the test of the single-local local loca	Oncology Pharmacology	Upcoming	inset.	2012-10-25
	12-2447 A	Longer and Alexandron and Alexandron and Alexandron and Alexandron and Alexandron (A), 202 hep-th/981000	Oncology Pharmacology	Upcoming	dru Owing	2012-10-25
	12-3033	Computing and any low and the COLD OF and the ACOLD ADD and the COLD ADD ADD ADD ADD ADD ADD ADD ADD ADD A	Oncology Pharmacology	Upcoming	inter	2012-10-25
	12-2968	Small of some 2 propagations (2) alls (22) some i spigerate fights may	Oncology Growth / Passaging	in. Progress	Rature Cent	2012-10-25
	12-2968 A	first separate of all same with the and second states of the first sec	Oncology Growth / Passaging	In Progress	Roture Carlo	2012-10-25
	12-2746 A	Hung, distribution (2): * and human in parely. TOR: which is in the ARM to we of the CP regular sector for CP (Proc.) years and you failed that is not	Oncology Pharmacology	in. Progress	Rature Oah	2012-10-25
	11-1109	Phase if all advantages (1017 and well \$20 \$20 \$10 and well \$20 and \$2	Oncology Pharmacology	Data Entered	Decision.	2012-10-24
	11-1100 A	Many distantiants (MP and an ADP \$10.00 second or 40 Texture) while a local Texture product angularization and (3.7.202 long-res.	Oncology Pharmacology	Data Entered	Deviction	2012-10-24
	11-1617 E	Black of all advances in the local data and an electric of the local interaction of the local states around a new state logic in the local data in the local state and the local data and the local states around a new state logic in the local data in the local state and the local data and the loca	Oncology Pharmacology	Data Entered	Destries	2012-10-24
	11-2519	Phase of a fragmational study (12), required with (12) and marked with the fragmational (12%) a first state of a	Oncology Pharmacology	Data Entered	Carichian	2012-10-24
	11-2519 A	Paragent Part agained range (22 magdin/100, 12) arriverse (22 million ellips rand Within house (2) fails rank the	Oncology Pharmacology	Data Entered	Cover Name	2012-10-24
		states had a strength on the strength of the strength of the strength of the strength of the	Canalana	-	and the second	

## **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



## **Thank You!**

