BUSINESS
TECHNOLOGY

Pfizer Informatics for NGS

Giles Day & Adam Kraut
Pfizer BioTherapeutics R&D – BioTeam, Inc.
XGEN Mar 15th 2010

Rinat and Pfizer

- Small biotech purchased by Pfizer 2006
- Located in South San Francisco
- Antibody Therapeutics
- 454 Sequencing Team co-located
- Approx 100 people

Perennial Informatics Issues

- Capture of knowledge & decisions
  - Why did we do that, who did that, when did they do it
- Integration of structured & un-structured information
- Discovery of relevant information
- Foster collaboration

Overview of NGS Strategy

- New target discovery
  - GWAS
  - In-house & external
- Support of existing therapeutic programs
  - Phage display library analysis
  - Patient stratification

Solutions Come and Go

- In-house relational platforms
- SharePoint
- Semantic Solutions
- Google/Search Devices
- Pathway Platforms

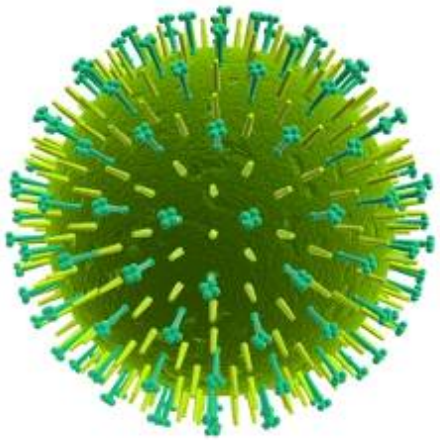
What They Are

- Inflexible
  - Unable to adapt to changing research processes & organisations
  - Can never capture everything
  - Difficult to administer
- Culture Defiant
  - I'm not ready to share
  - What benefit do I get
  - It's not my solution
- Traditional
  - Software Development cycles are too long – even agile ones
  - Too expensive

What They Need To Be

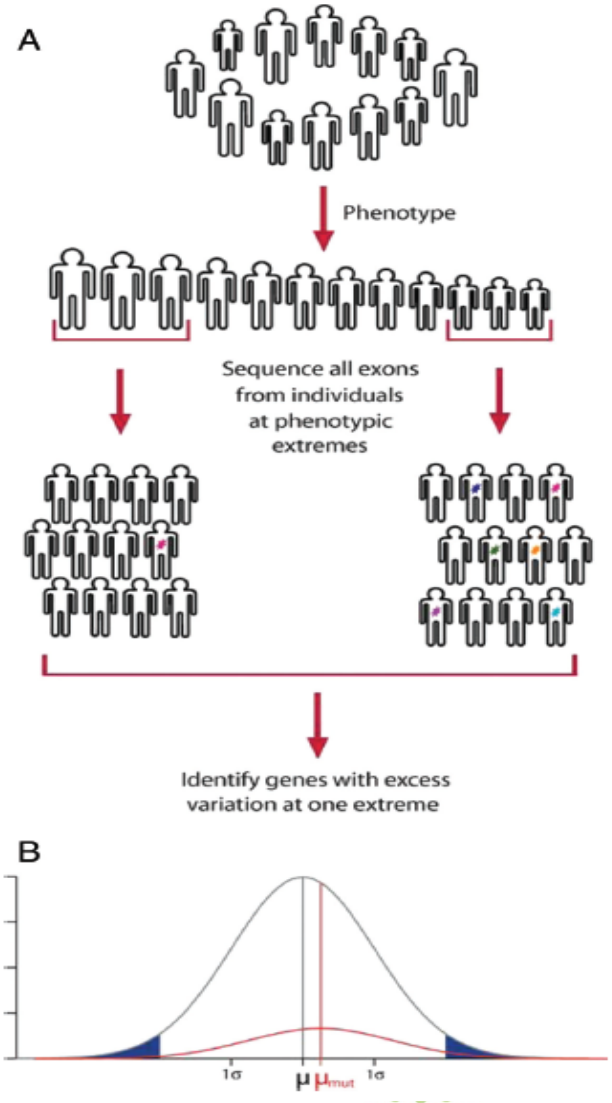
- Useful — return more benefit than cost of construction
- Intuitive — low adoption barrier
- Fast — rapid prototyping
- Flexible — modular, able to retool to new uses
- Talkative — communicates easily with other technologies
- Cheap — design cycle not limited by cost of development
- Open — open source, transparent & easily modified
- Adopted — Community of developers expanding functionality

Informatics must co-evolve with scientific innovation



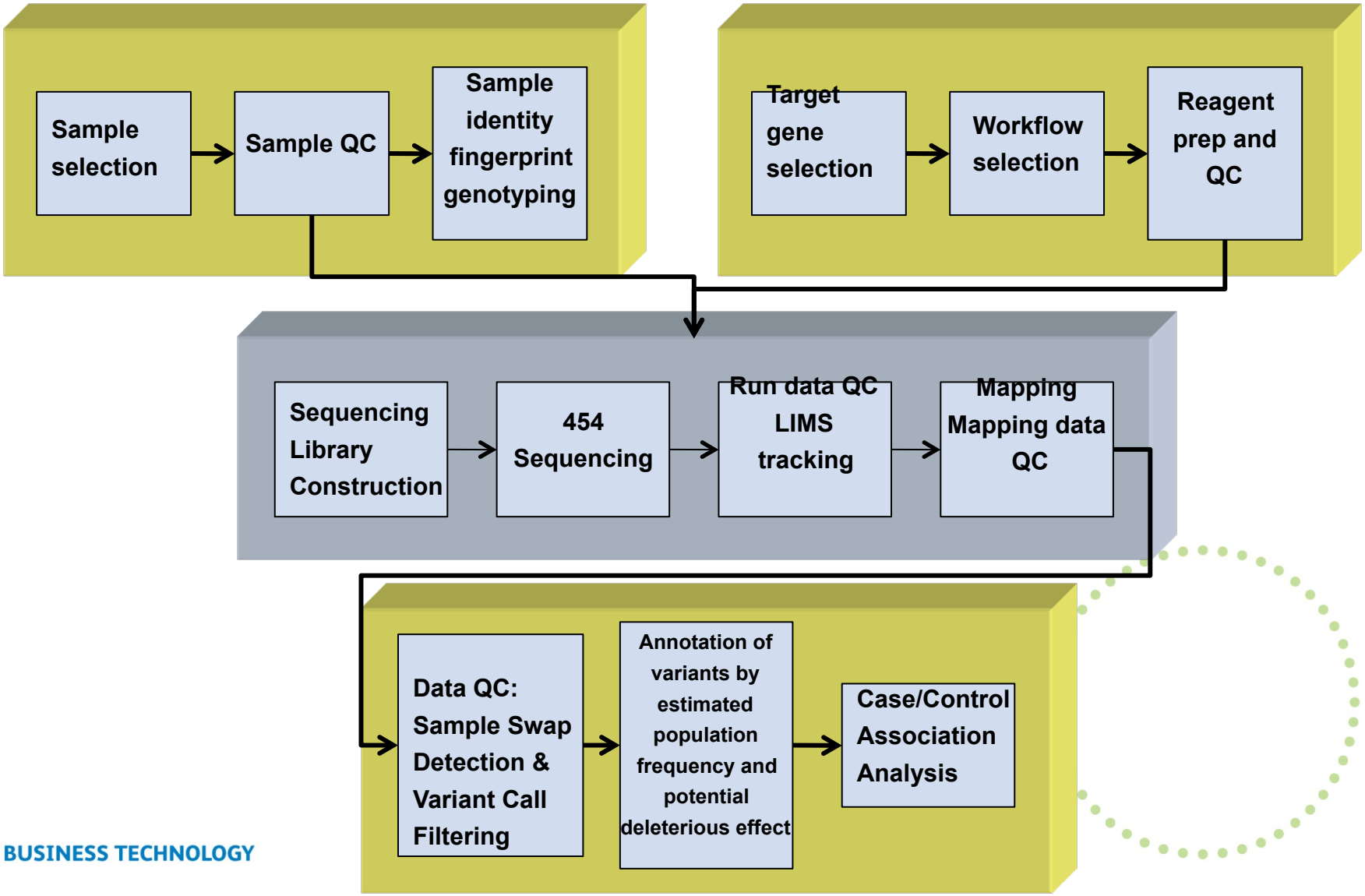
How to Translate Bases to Clinical Outcomes?

- GWAS Approach
  - 100s to 1000s samples
  - Well characterized phenotypes
  - Sequence genes/genomes in extreme
  - phenotypes
  - Sequence deep to discover variants
    - Rare loss / gain of function variants
    - Small (additive) effects of many variants expected
    - Annotation
  - Validation of putative SNPs on larger cohorts
  - Identify drug targets
  - Patient stratification, Dx

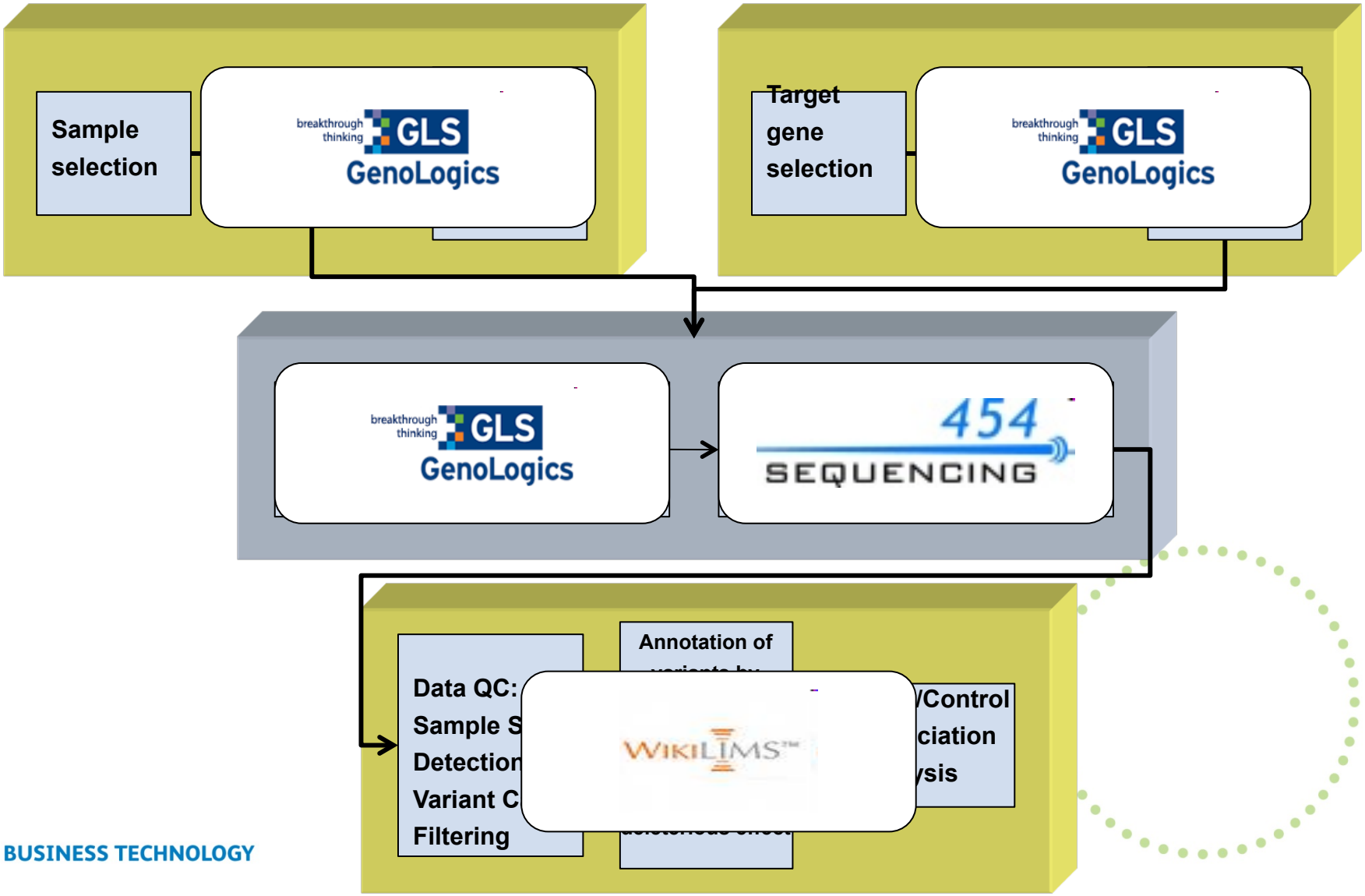


www.pnas.org/cgi/doi/10.1073/pnas.0812824106

Overview of NGS Workflow

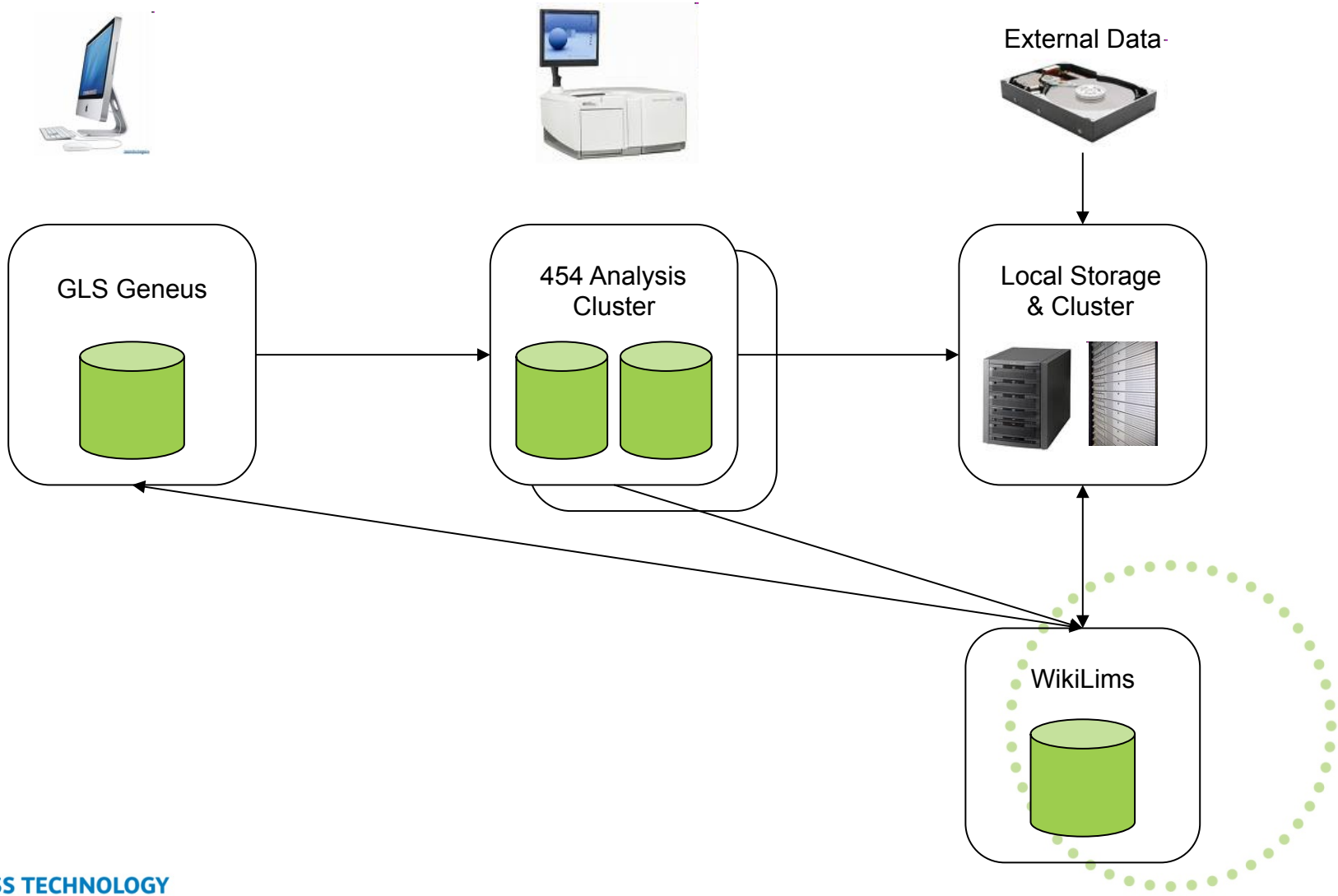


Informatics Solutions

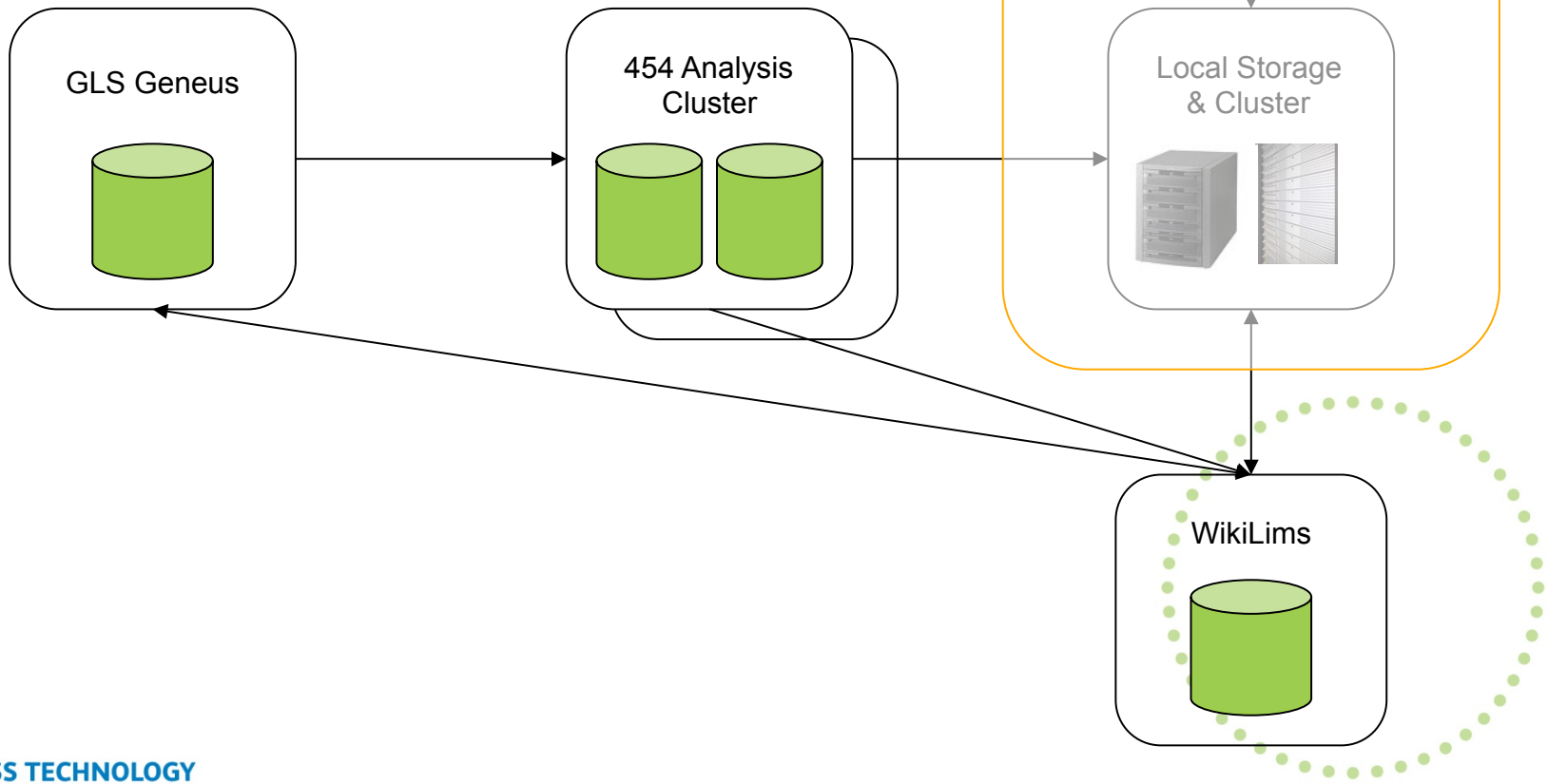


BUSINESS TECHNOLOGY

Informatics Infrastructure



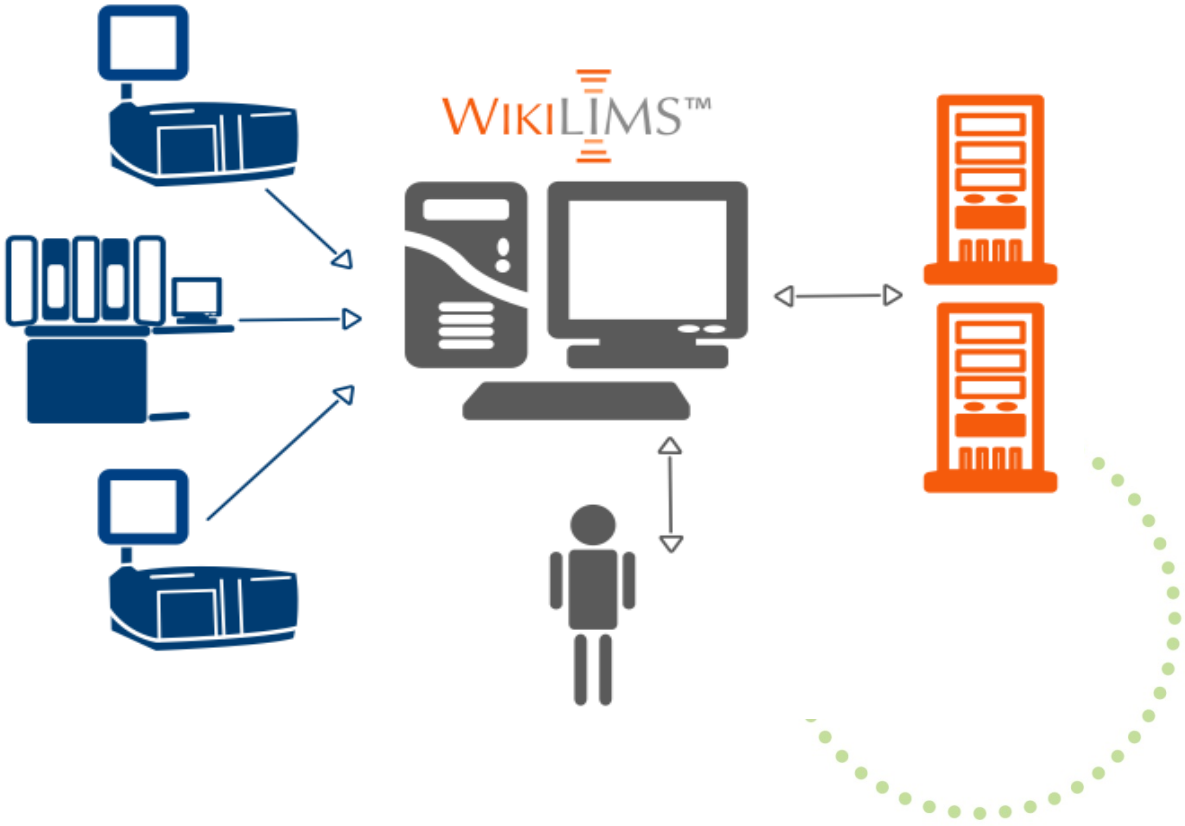
Future Directions



The Details

WikiLIMS

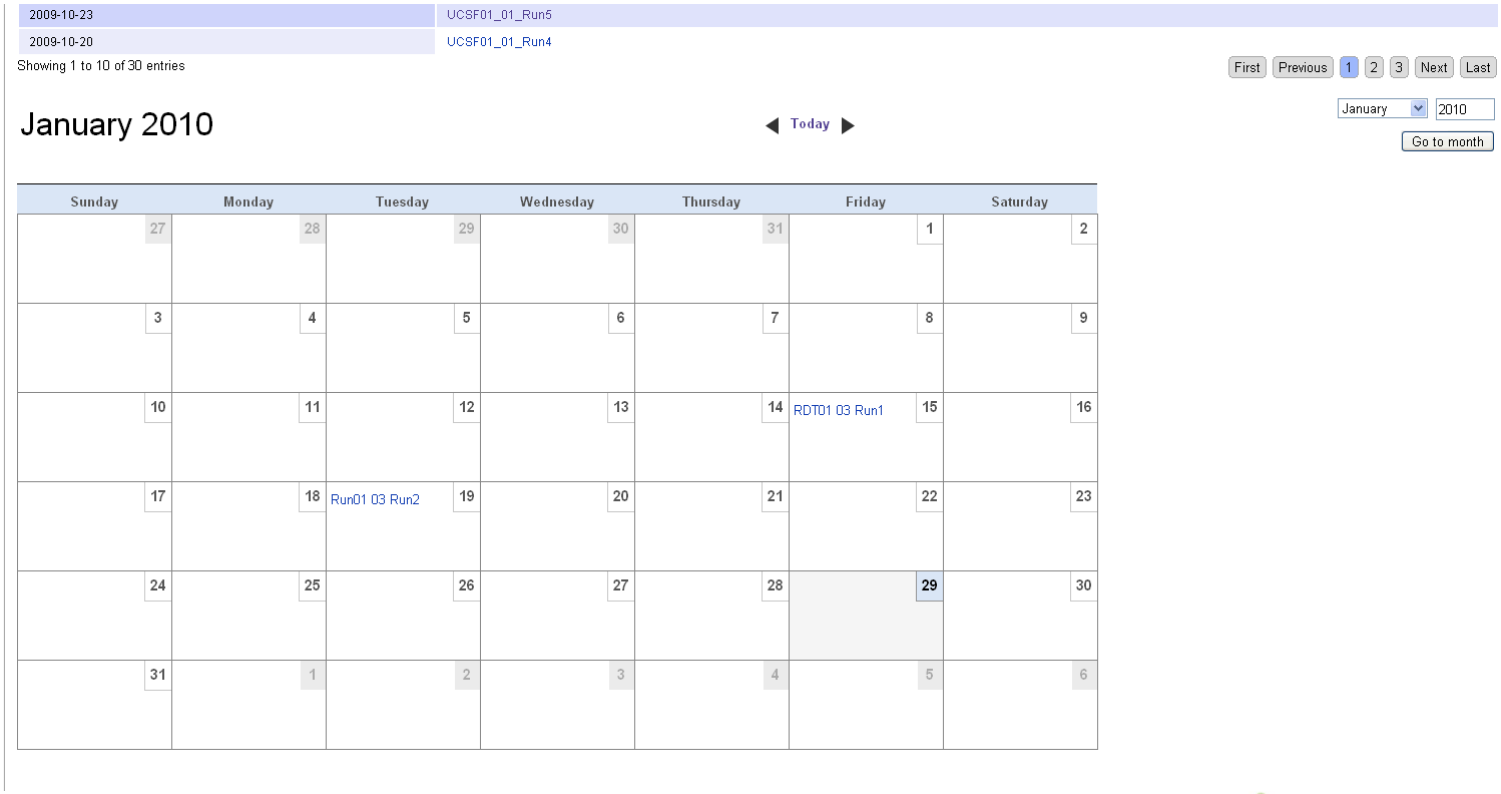
- A system that is quickly constructed and easily revised
- A system that can handle many data types
- A modular and transparent system
- A Web-based system



Simplified tracking information

- Email and RSS notifications for every step in workflow
- Wiki Revision Control explains ***who*** did ***what*** and ***when***
- Lab Managers can revert and undo tasks

Monitoring Sequencing Runs– Calendar view



2009-10-23	UCSF01_01_Run5
2009-10-20	UCSF01_01_Run4

Showing 1 to 10 of 30 entries

First Previous 1 2 3 Next Last

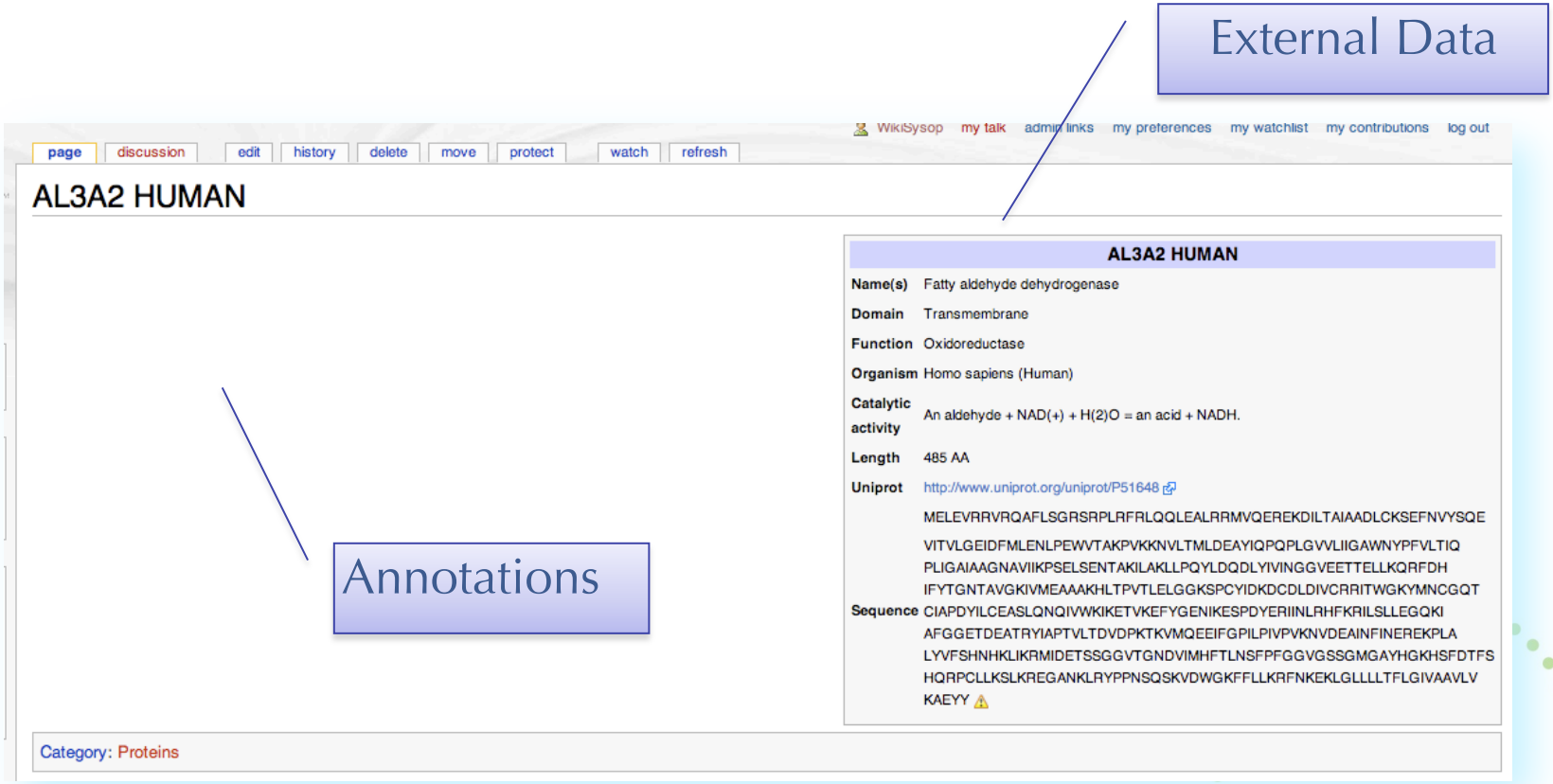
January 2010

◀ Today ▶

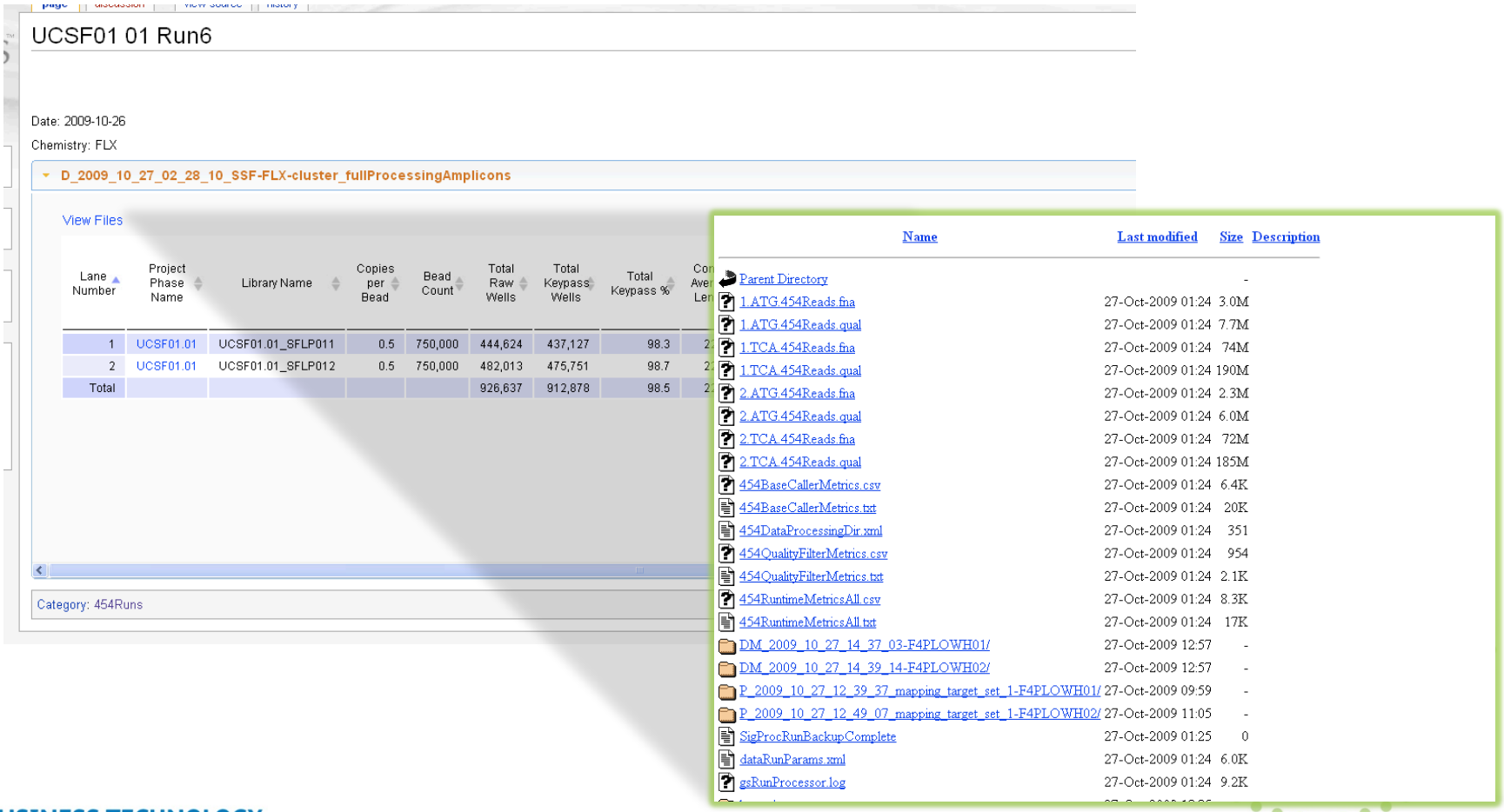
January 2010 Go to month

Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
27	28	29	30	31	1	2
3	4	5	6	7	8	9
10	11	12	13	14	RDT01 03 Run1 15	16
17	18	Run01 03 Run2 19	20	21	22	23
24	25	26	27	28	29	30
31	1	2	3	4	5	6

Intelligent Annotation: Leveraging Web Services



Combining raw data and reporting



Simple Integrations

page | discussion | view source | history

RIN06 02 001

Date: 2010-01-29

Chemistry: Titanium

D_2010_01_30_01_35_18_SSF-FLX-cluster_fullProcessing

View Files

Lane Number	Project Phase Name	Library Name	Copies per Bead	Bead Count	Total Raw Wells	Total Keypass Wells	Total Keypass %	Control Average Length	Control Keypass Wells	Controls with 98% accuracy over 200 bp %	Control Pass Filter %	Control Mixed %	C D
1					995,379	983,282	98.8	443.6	8,427	93.0	94.8	0.8	
2					1,033,895	1,023,075	99.0	440.7	8,417	90.5	91.8	0.8	
Total					2,029,274								

Data retreived from database via 1 line of code

page | discussion | view source | history

View source

for RIN06 02 001

You do not have permission to edit this page, for the following reason:

The action you have requested is limited to users in the group: Users.

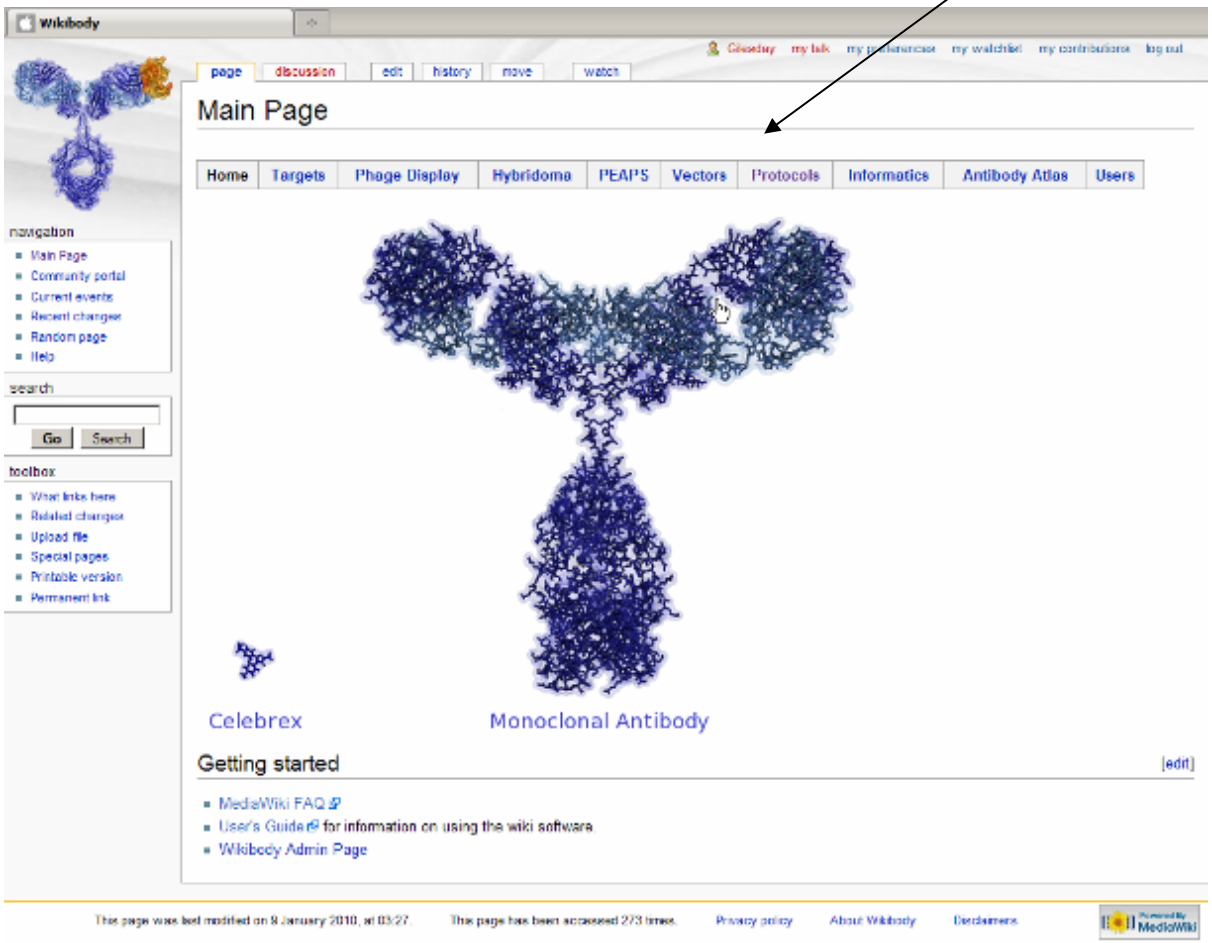
You can view and copy the source of this page:

```
{{454Run|id=33|run_dt=0000-00-00}}

== Primer Design ==

[[Image:primer.design.titanium.png]]

== 6Mer MIDs ==

<PRE>
+---------+----------+
| name    | sequence |
+---------+----------+
| PFMID1  | ACGTGT   |
| PFMID2  | ACTCTC   | **
| PFMID3  | AGATCG   |
| PFMID4  | AGTGAC   | **
```

Getting Started

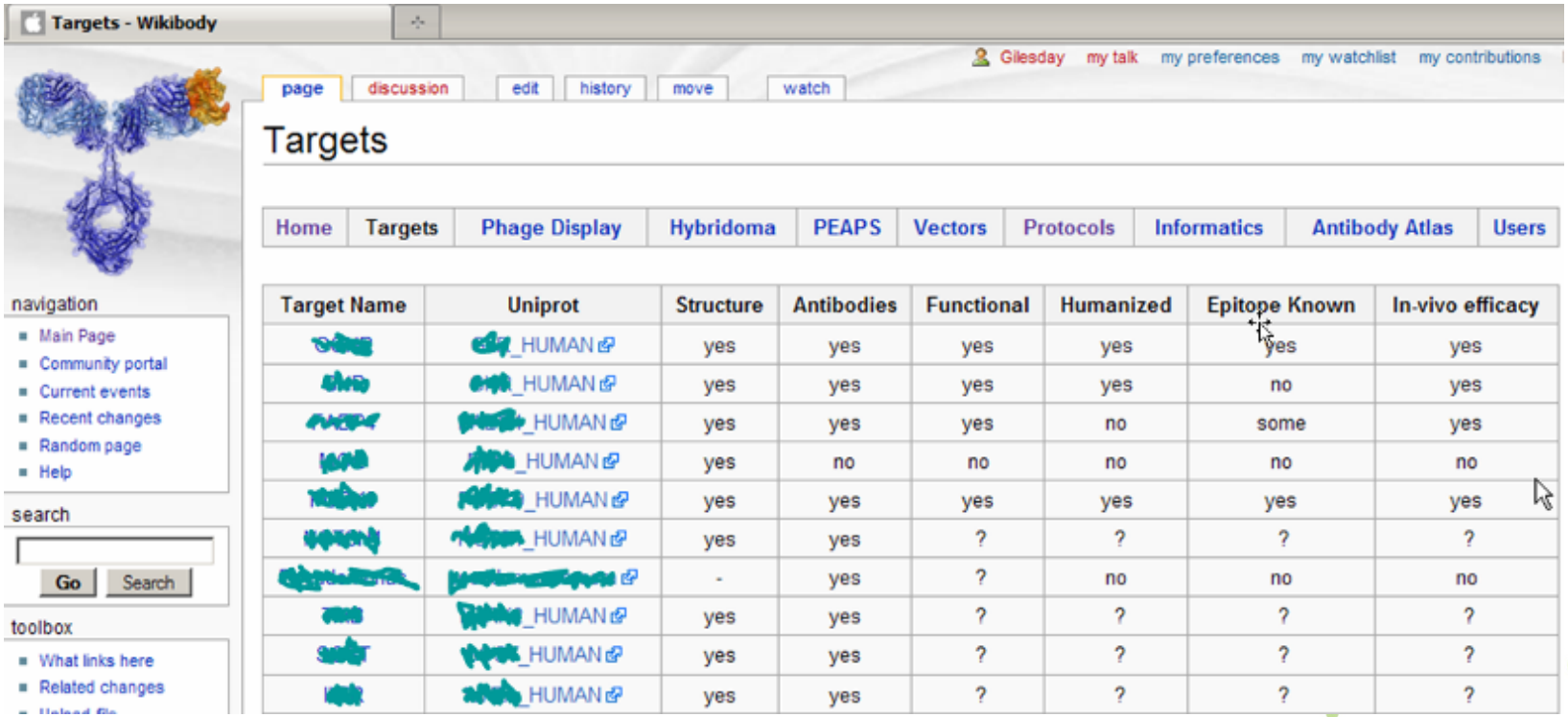
Familiar Wiki layout

Tabs separate concept areas



Supports all areas for Rinat research

Project Overview



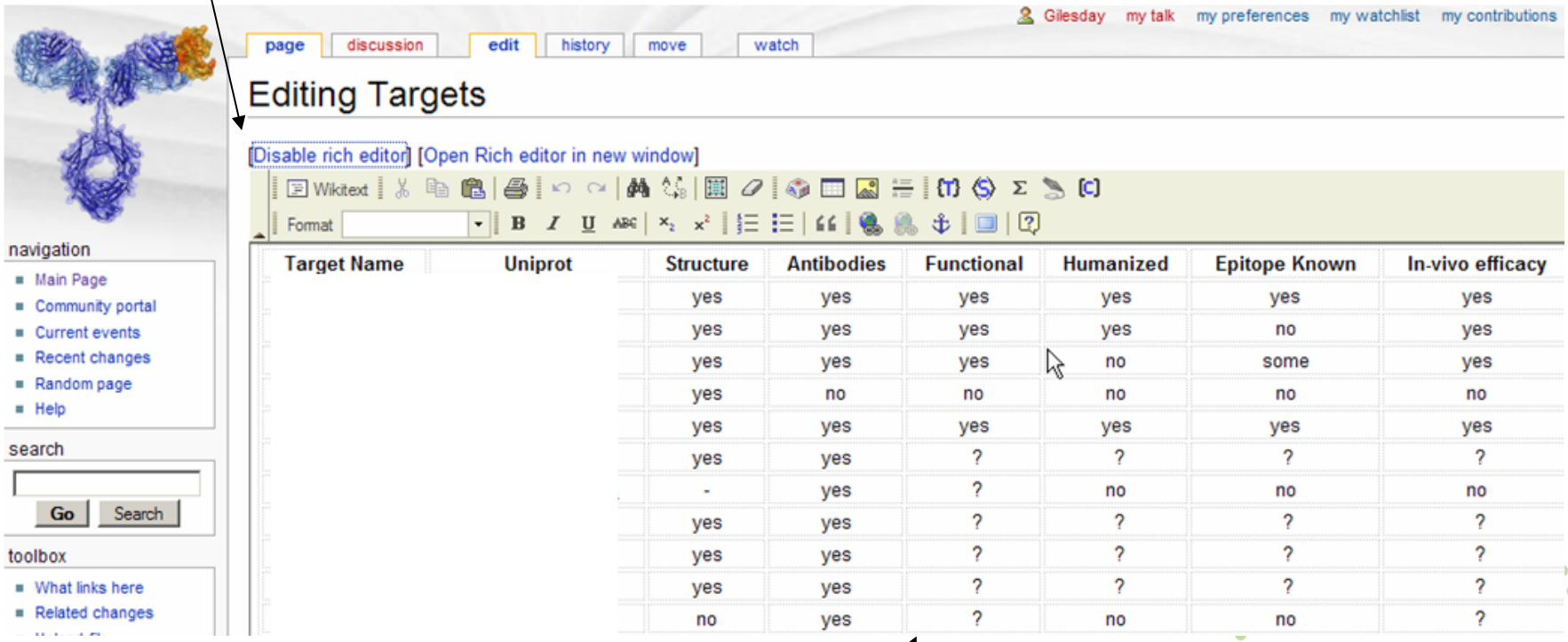
Target Name	Uniprot	Structure	Antibodies	Functional	Humanized	Epitope Known	In-vivo efficacy
[redacted]	[redacted]_HUMAN	yes	yes	yes	yes	yes	yes
[redacted]	[redacted]_HUMAN	yes	yes	yes	yes	no	yes
[redacted]	[redacted]_HUMAN	yes	yes	yes	no	some	yes
[redacted]	[redacted]_HUMAN	yes	no	no	no	no	no
[redacted]	[redacted]_HUMAN	yes	yes	yes	yes	yes	yes
[redacted]	[redacted]_HUMAN	yes	yes	?	?	?	?
[redacted]	[redacted]	-	yes	?	no	no	no
[redacted]	[redacted]_HUMAN	yes	yes	?	?	?	?
[redacted]	[redacted]_HUMAN	yes	yes	?	?	?	?
[redacted]	[redacted]_HUMAN	yes	yes	?	?	?	?

Active projects

Decisions and current status

Super Simple Editing

Rich Editor



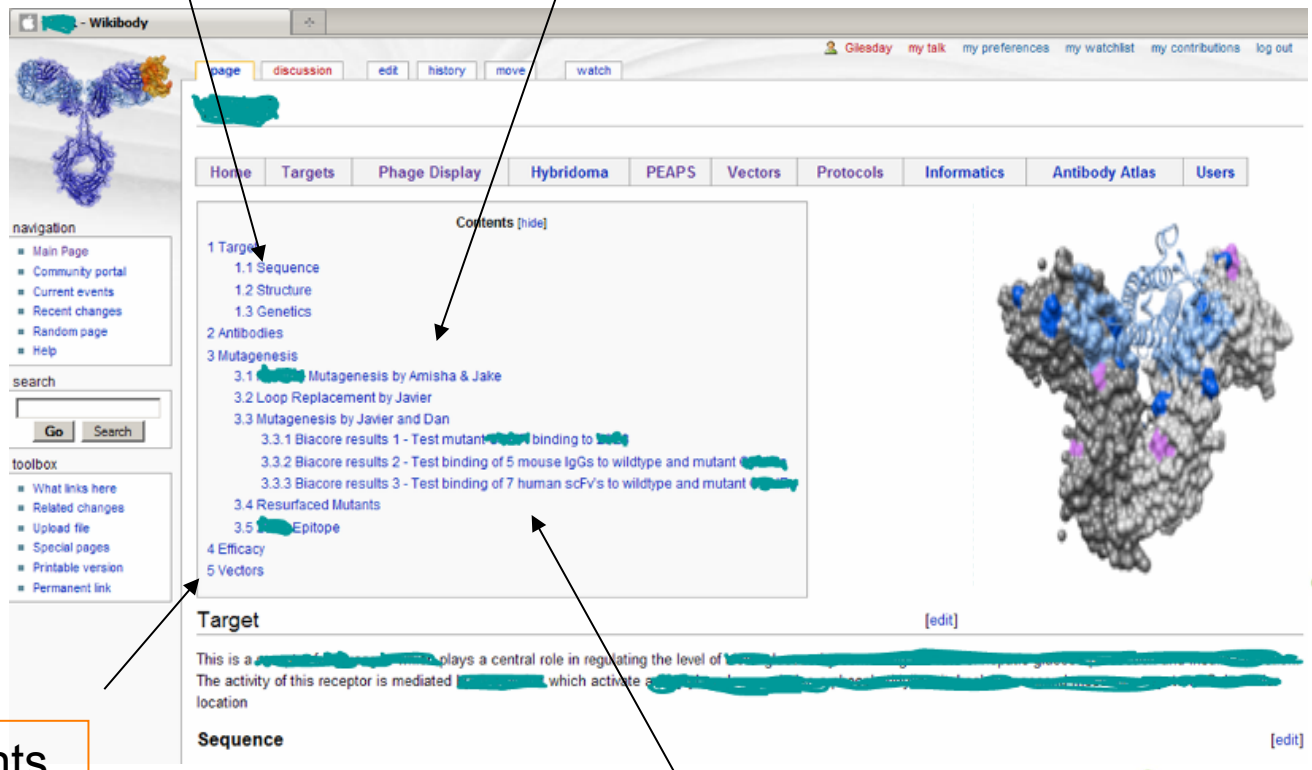
Target Name	Uniprot	Structure	Antibodies	Functional	Humanized	Epitope Known	In-vivo efficacy
		yes	yes	yes	yes	yes	yes
		yes	yes	yes	yes	no	yes
		yes	yes	yes	no	some	yes
		yes	no	no	no	no	no
		yes	yes	yes	yes	yes	yes
		yes	yes	?	?	?	?
		-	yes	?	no	no	no
		yes	yes	?	?	?	?
		yes	yes	?	?	?	?
		yes	yes	?	?	?	?
		no	yes	?	no	no	?

Regular HTML table

Individual Target Page

Public data

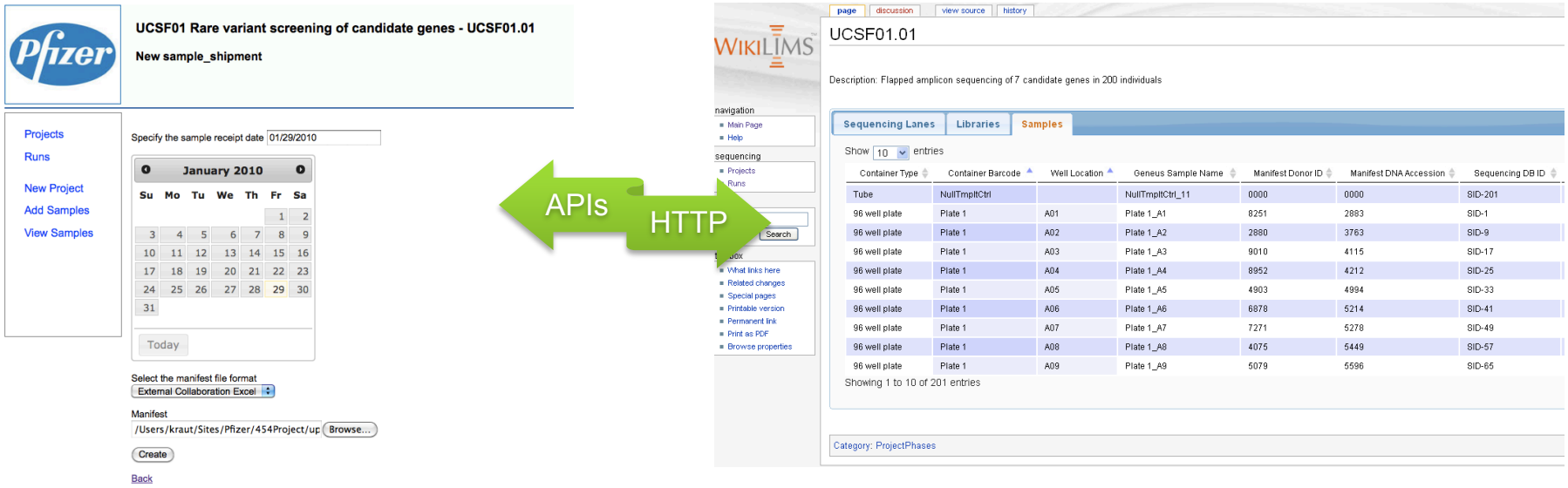
Mutagenesis experiments



Reagents

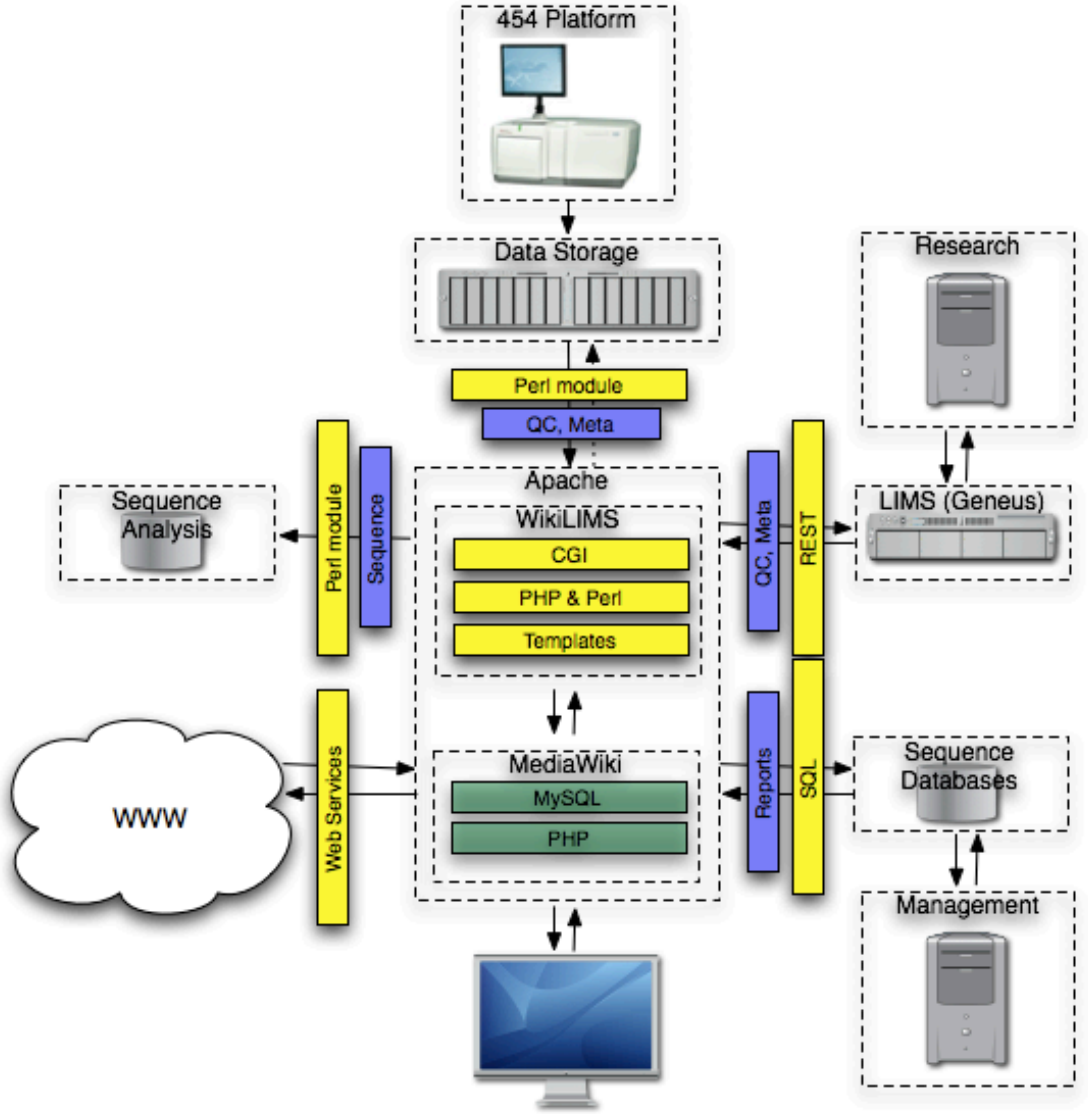
Binding data

“Small things, loosely joined, written fast”



- Bring together diverse systems with loose interfaces
- Design and leverage **APIs** wherever possible
- Avoid all-in-one solutions
- Scale components **independently**

WikiLIMS@Pfizer



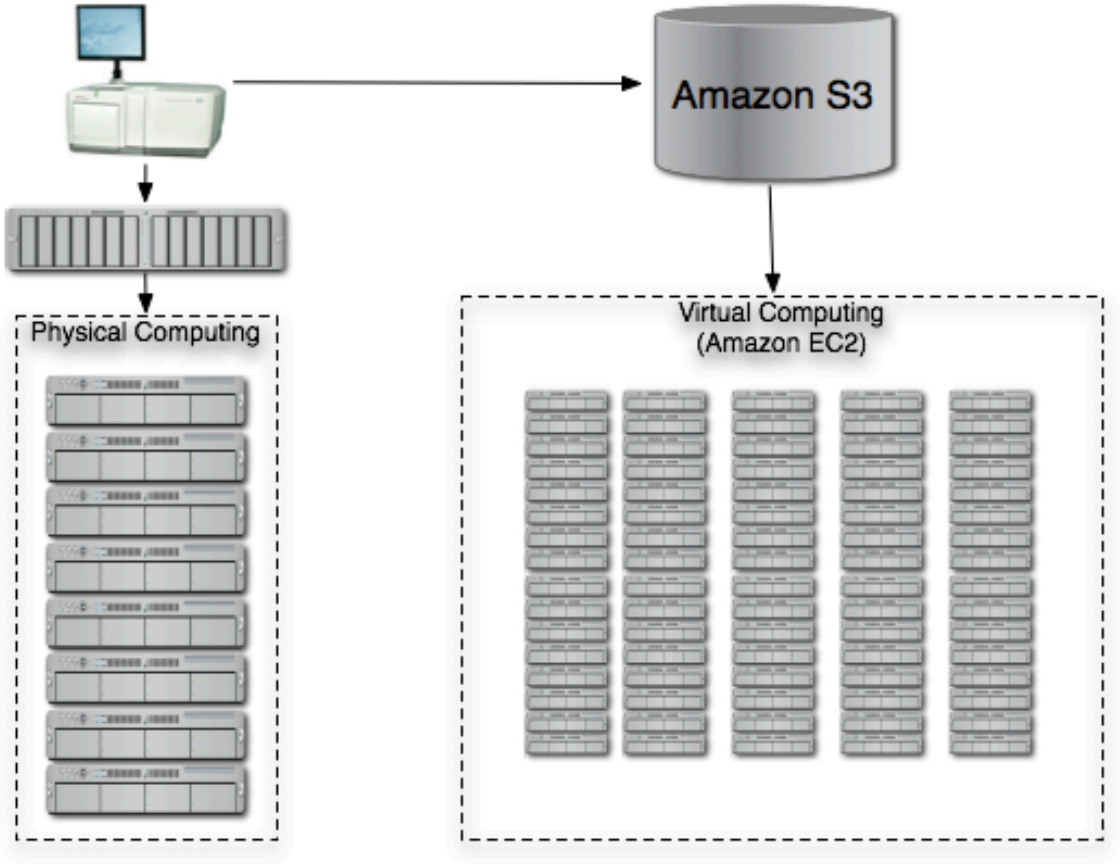
Data Analysis

Infrastructure as a Service (IaaS) for NGS Data

BUSINESS TECHNOLOGY

Balancing Trade-offs…

- Physical computing
- Virtual computing
- Availability
- Performance
- Investment
- Cost savings

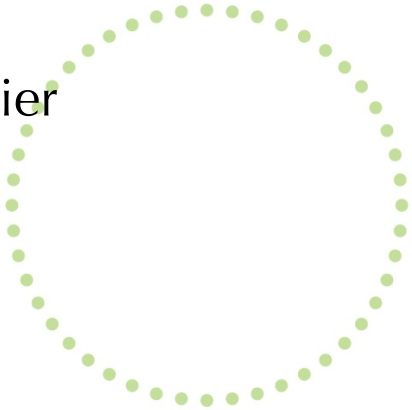


Amazon S3 Bulk Data Ingest/Export

- **How do I forklift 200TB of data into the cloud?**
  - Fast network connectivity is expensive
  - Ship physical disks to Amazon to load into S3
- **What this means:**
  - S3 storage is cheap, deep, and geographically replicated
  - Primary analysis data moved into remote utility storage
  - Data would rarely come back
  - Need to reprocess or analyze?
    - Spin up “cloud” servers and re-analyze in-situ

CloudOpportunity

- Bootstrapping new resources can happen in **minutes** instead of **months**
- Shift capital expenditure to operating expenditure (CapEx/OpEx)
- Primary analysis done on-site or commercial sequencing facility; data moved to remote utility storage
- Easier to distribute and collaborate across multiple sites
- Data would rarely need to move back
- Re-analyze data 'in the cloud' on virtual clusters
- Drop common analysis pipelines into a shared service tier



CloudSobriety

- Security concerns should not be overlooked
  - Regulatory compliance
- Amazon may be the only provider of bulk import/export services
  - What if Amazon goes out of business?
- Life Science tends to be I/O bound; I/O performance in the cloud is still lacking
- IT staff needs to understand 'the cloud' and learn to use it
- Critical to quantify your own internal operating costs

Growth Strategy

- Organic – mirroring Pfizerpedia
- Providing information carrots
- Restricted to Rinat groups
- Open source?

Advantages

- Flexibility
- Cost
- Complexity
- Adoption

Credits

- Jacob Glanville
- Adam Kraut (BioTeam)
- Steve Pitts
- Andrea Rossi

Questions?

- That's it

- Giles Day giles.mr.day@pfizer.com
- Adam Kraut kraut@biotam.net