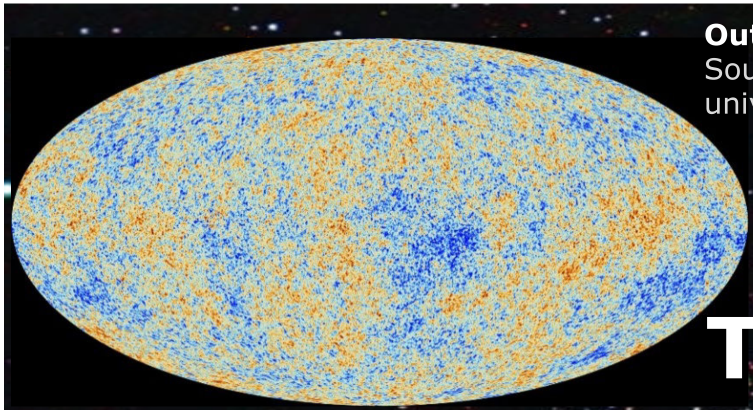


Outer Space:
Source -- Cosmic Microwave Background (CMB): the oldest light in our universe, Planck spacecraft, 2013 (Copyright: ESA and the Planck Corporation)

The Clinical Galaxy

Validation Plan proposal

Galaxy Community Conference 30 June - 2 July 2013 OSLO UiO University of Oslo

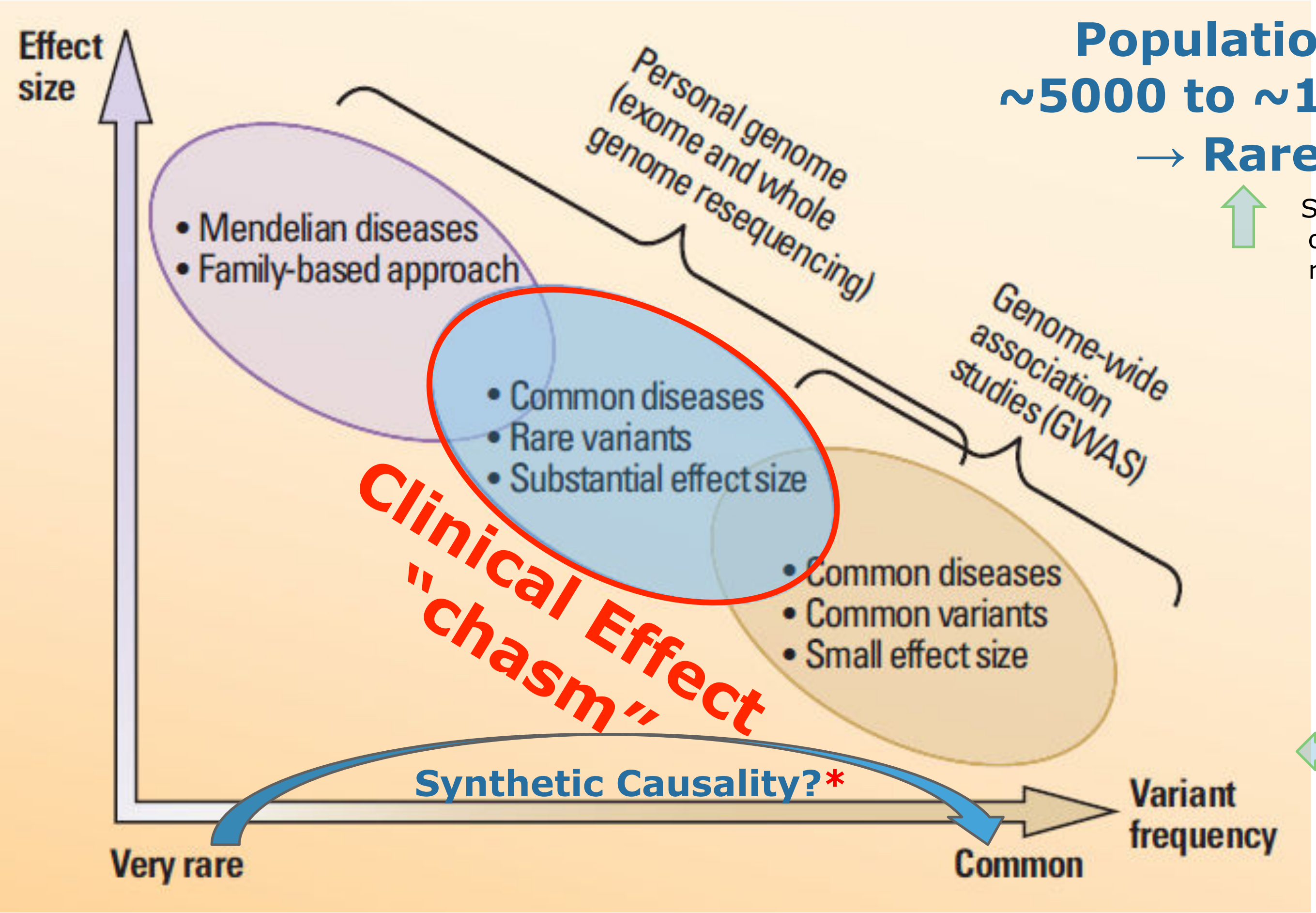
Sanjay Joshi
CTO Life Sciences, **EMC^2 Isilon Storage Division**

Inner Space:
Source -- Bolzer A, et al., "Three-Dimensional Maps of All Chromosomes in Human Male Fibroblast Nuclei and Prometaphase Rosettes.", (2005), PLoS Biol 3(5)

STORY IN 2 PARTS:

- ❑ Rationale
- ❑ Proposal

Effect vs. Variant Freq.



Population Explosion ~5000 to ~1000 years ago → Rare Variants

Source: W. Fu et al. "Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants.", Nature. Nov 28, 2012. doi: 10.1038/nature11690

*Source: Dickson SP, et al, "Rare Variants Create Synthetic Genome-Wide Associations", PLoS Biol 8 (1), 2010: e1000294. doi:10.1371/journal.pbio.1000294

Source: Kaiser, J. "Genetic Influences On Disease Remain Hidden", Science, Vol 338, 23 Nov 2012

1 earth-mole of all human cells?

~7 x 10^{9} humans

range{1, 240} x 10^{12} cells/human*

range{0.07, 16} x 10^{23} human cells

0.7 to 160 x 10^{23} human microbes on earth?

Avogadro Number : 6.023 x 10^{23}

***15-60 picograms/cell**. Source: Phillips KG, et al, "Optical quantification of cellular mass, volume, and density of circulating tumor cells identified in an ovarian cancer patient" Frontiers in Oncology: Cancer Molecular Targets and Therapeutics, July 2012, Vol 2:72

Sample Size

p-values do not work*

In a Population of 7 x 10^9,
with Confidence Interval of 0.025,
Confidence Level of 95%
and 100 degrees (genes) of freedom,

Sample Size = ~15 million

*Sources: du-Prel J-B, et al "Confidence Interval or P-Value?", Dtsch Arztebl Int 2009; 106(19): 335-339

*Ziliak ST and McCloskey DN, "The Cult of Statistical Significance", Section on Statistical Education, JSM 2009

Data Intensive Workflows

Data Center Tradeoffs

Higher Priority

High Availability*, *Archival

CAPEX* vs. *OPEX

Temp-Space*: *Local vs. Scale-Out

(define "writes", understand process)

***CPU + Acceleration* on same node**

***CPU speed control*,**

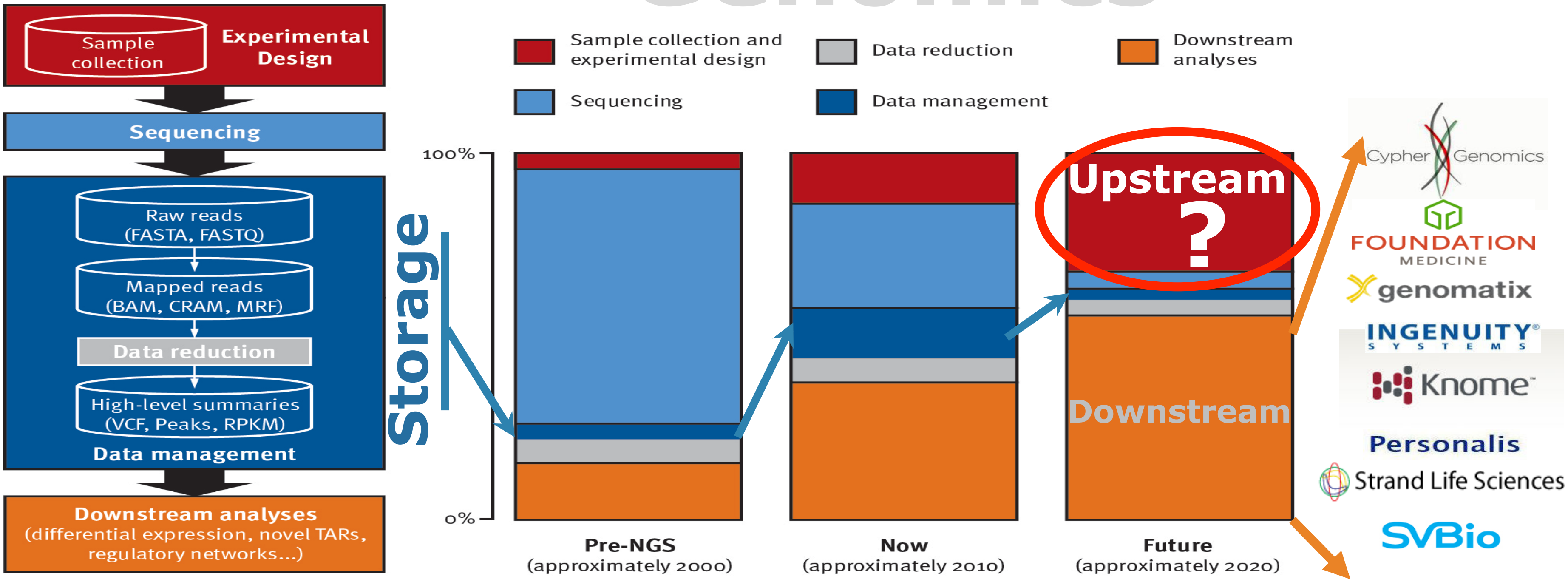
***High-density disk* nodes**

Lower Priority

Derived from:
Stuart Feldman, Google

OPEX

Genomics



Derived from: Sboner A, et al., "The real cost of sequencing: higher than you think!", Genome Biology 2011, 12:125

Storage: Choose Two

Cost

Performance

Speed

Galaxy

Validation Initiative Proposal

Data Management

The Galaxy philosophy:

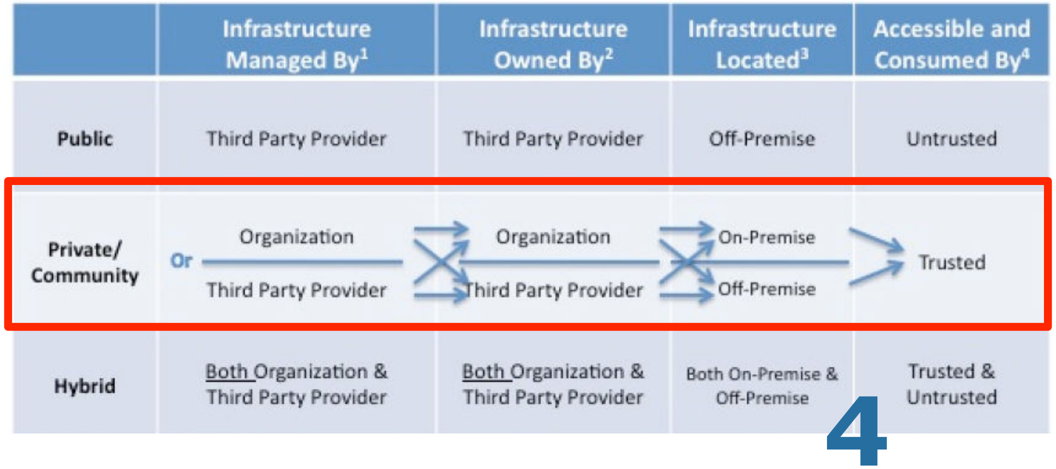
- Data is never overwritten
- Data is never deleted

MOTIVATION

Veracity

Private & Community Cloud?

From: Security guidance for critical areas of focus in Cloud Computing v3.0: http://www.cloudsecurityalliance.org/guidance/csaguide.v30.pdf



	Infrastructure Managed By[1]	Infrastructure Owned By[2]	Infrastructure Located[3]	Accessible and Consumed By[4]
Public	Third Party Provider	Third Party Provider	Off-Premise	Untrusted
Private/ Community	Organization Or Third Party Provider	Organization Third Party Provider	On-Premise Off-Premise	Trusted
Hybrid	Both Organization & Third Party Provider	Both Organization & Third Party Provider	Both On-Premise & Off-Premise	Trusted & Untrusted

79% of respondents are **concerned about Cloud Security***

*Feb 2012

InformationWeek :: reports

15 ***hacks per day in 2012*****

versus 10 in 2011

*** Ponemon Institute Oct 2012*

21M

data breaches in 18-month period • *2010-11 ($2.25M/breach)*

• US Govt. Office of Civil Rights, HHS

£1.79B fines in UK for NHS 2012 breaches

4

position points (GPS) determine identity • •

• • Scientific Reports, 3, Mar 2013

PGP Encrypted De-ID reverse engineered +

+ Data Privacy Lab, CMU, Apr 2013

genome **Y-STR**

Surnames from ancestry data

Science, Jan 2013: 339:6117 pp. 321-324

4M

Common Variants: Trace DNA

Craig DW, et al, (2008), PLoS Genet 4(8)

Regulations

Access Controls

Regulations	Functions	Methods
45CFR 164.312(a) **21CFR 11.10(b)(d)**	**System Access:** Unique Name/Id, Role-based perms, Single Sign-On SAML	FDA, HIPAA AES SAML WS-Trust

Audit Controls

Regulations	Functions	Methods
45CFR 164.312(d) **IETF RFC 4120** **IHE-ITI-TF** **MOL.34960,34968, 34970**	**System Entry:** Enterprise User Auth. (EUA) Cross Enterprise User Auth. (XUA) Logging	FDA, HIPAA ATNA CLIA-CAP

Complied by Joshi S. from various CFR, FDA, HIPAA, IETTF, IHE, ASTM, FIPS and CAP guidelines

De-ID, Re-ID

Regulations	Functions	Methods
45CFR 164.312(d) **IETF RFC 4120** **IHE-ITI-TF**	**PHI:** Code for de-ID, Code for re-ID, Pedigree for Genomics	HIPAA de-ID HIPAA re-ID Pseudo-randomization HL7v3 Pedigree

Data Integrity

Regulations	Functions	Methods
45CFR 164.312(c) **FIPS PUB 180-2 SHA-224** **ASTM Std E1762-95** **MOL.34966**	**Data @rest:** Encryption, Key Management	HIPAA , SHA2, CLIA-CAP, ASTM-Auth

Transmission Security

Regulations	Functions	Methods
45CFR 164.312(d) **FIPS 180-2, 197** **IETF RFC 2246, 3546, 2630, 3852** **MOL.34972**	**Security:** HTTPS (web), Crypto Message Syntax, TLS Compression	FDA, HIPAA SHA2 AES TLS CLIA-CAP

Clinical Reports

MOL.34914, 34929 **MOL.34944, 34952** **MOL.34954**	CLIA-CAP

ISD Life Sciences

EMC²

VALIDATION PROCESS

Story in 4 steps:

❑ Clinical Use Survey

❑ Architecture Review

❑ Quality Systems Review

❑ Validation

STEP 0: Clinical Use Survey

TBD

STEP 1: Architecture Review

Architecture Review

- ❑ **Network Security**
- ❑ **OS Hardening**
- ❑ **Deny all first** **(user/install),**

  allow and use only when needed
- ❑ **Applications config.** **(Galaxy, *PostgreSQL, Apache, …*)**
- ❑ **Physical server access** **(*BIOS, GRUB, SSH, …*)**
- ❑ **Internet access protocols** **(*SSH, sftp, API…*)**

STEP 2: Quality System Review

Quality System Review (QSR)

- ❑ **Logging and Server Management**
- ❑ **Risk Management**

  Business, Functional and Application
- ❑ **Human factors**

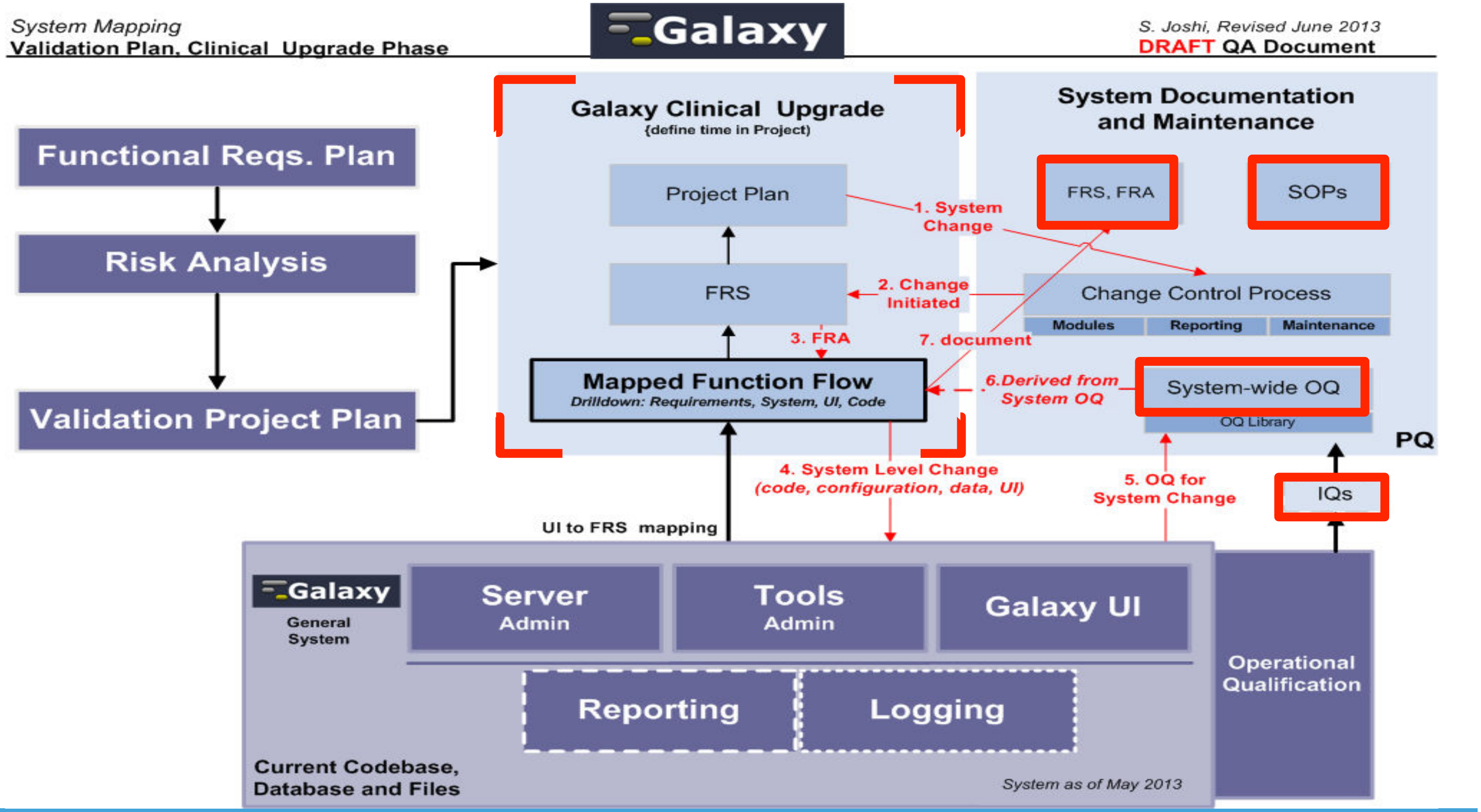
  Training, SOPs, Change Control
- ❑ **Software Development Life Cycle**
- ❑ **IT Processes**

  "The Cloud is your mess managed by someone else"*

**Source: Halamka J., "Life as a Healthcare CIO" blog*

STEP 3: Implement

Galaxy Clinical



System Mapping
Validation Plan, Clinical Upgrade Phase

Galaxy

S. Joshi, Revised June 2013
DRAFT QA Document

Functional Reqs. Plan

Risk Analysis

Validation Project Plan

Galaxy Clinical Upgrade
{define time in Project)

Project Plan

FRS

Mapped Function Flow
Drilldown: Requirements, System, UI, Code

1. System Change

2. Change Initiated

3. FRA

4. System Level Change
(code, configuration, data, UI)

5. OQ for System Change

6.Derived from System OQ

7. document

System Documentation and Maintenance

FRS, FRA

SOPs

Change Control Process

Modules

Reporting

Maintenance

System-wide OQ

OQ Library

PQ

IQs

UI to FRS mapping

Galaxy
General System

Server
Admin

Tools
Admin

Galaxy UI

Reporting

Logging

Current Codebase, Database and Files

System as of May 2013

Operational Qualification

STEP 4: Validate

TBD

The Quantified Patient

- Demographics 🟢
- Genetics 🟠
- Family History 🟡
- Imaging 🟡
- Medications 🟢
- Clinical Narratives 🟡
- Labs 🟢
- Molecular Diagnostics 🟠
- Environment 🟠
- Sensors 🟠

- 🟢 Available at Point of Care
- 🟡 Available Data
- 🟠 Not Available

Source: EMC Healthcare