

Big Data's Starring Role

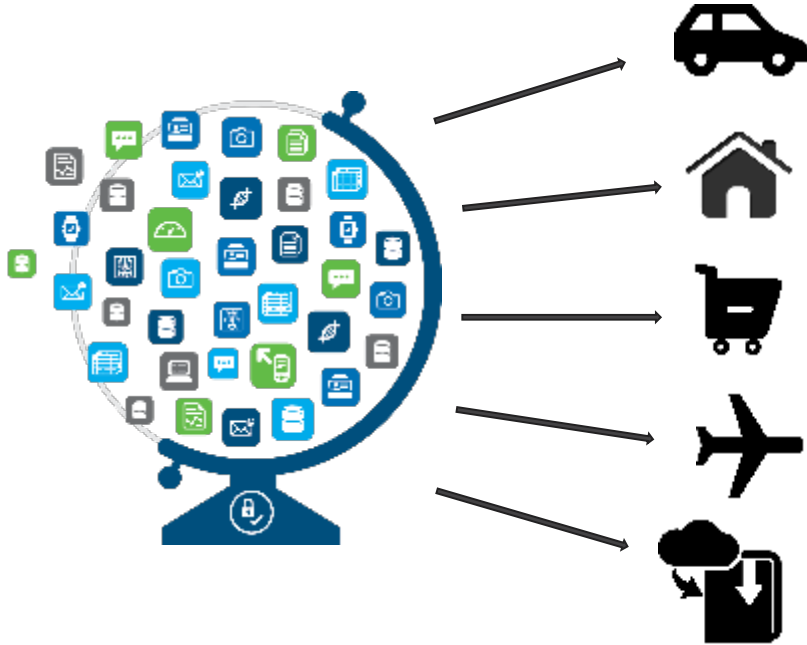
in Next-Generation Information Architecture

Dr. Paul Terry
PHEMI Systems



Data Capital

What is data capital?



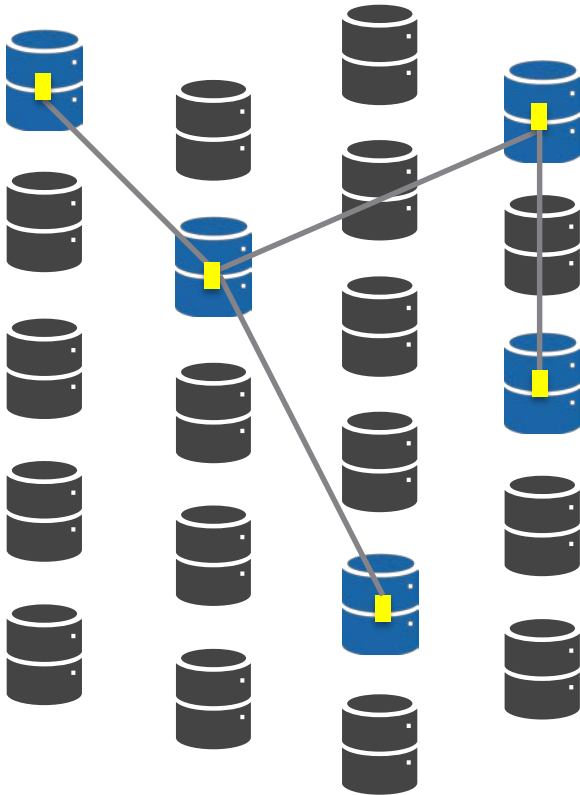
In economics capital goods, real capital, or capital assets are already-produced durable goods or any non-financial assets that are used in production of goods or services

Data capital is defined as recorded information necessary to produce a good or service

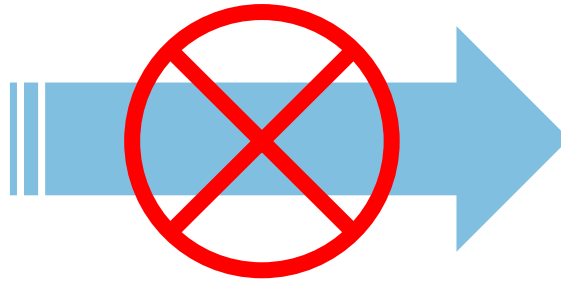
We have a wealth of data in BC that can fuel innovation, improve service delivery, and drive economic growth

The Data Dilemma

Enterprise Data is Locked in Silos



Share and Protect?



Analytics
Applications

Data Available
analysis, insights,
and new services

How do I share data
for secondary use
without compromising privacy,
security, and governance?

Digital Library



Transforming Data into Capital

Gather
Everything



Curate



Services



Gather Everything



X-Ray



Customer Service
Call Audio



Gene Sequence

```
1 // SyntaxHighlighter makes your code
2 // snippets beautiful without tiring
3 // your servers.
4 // http://alexgorbatchev.com
5 var setArray = function(elems) {
6   this.length = 0;
7   push.apply(this, elems);
8   return this;
9 }
```

Code
Fragment



Database



Web Pages



EMR Form



Emails



Documents



Internet
of Things



Social Media



Video



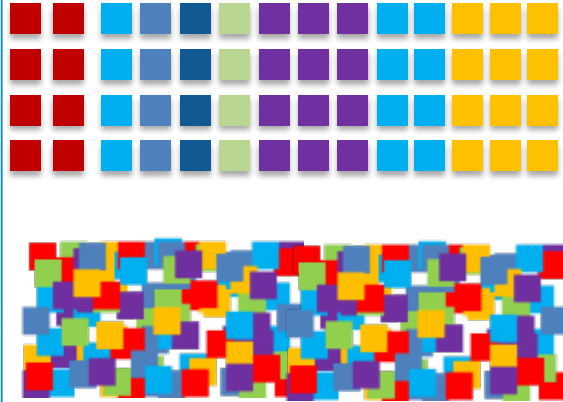
Virtual Machine

Curate

INDEX & CATALOG



EXTRACT INFORMATION



PROTECT

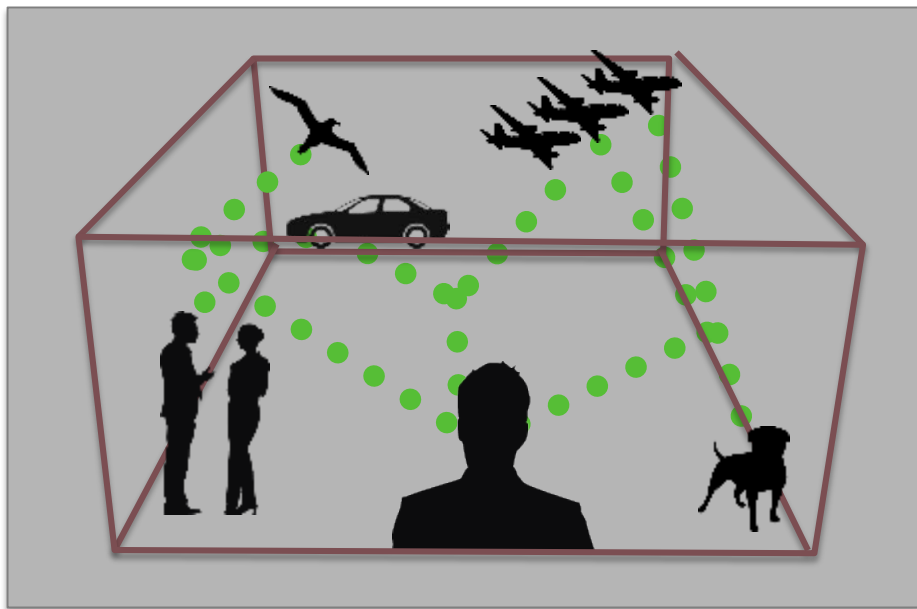


METADATA

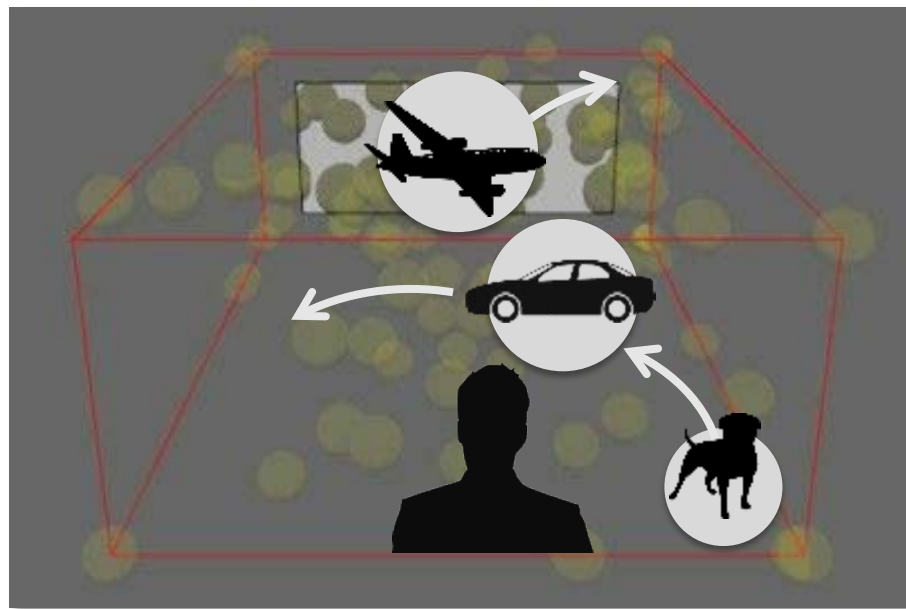
Beyond Traditional Metadata

Metadata for Sound: ATMOS

Sound pre-assigned to channels



Dolby ATMOS: Every element becomes a separate sound object with its own metadata
The system determines where to “play” it



Privacy in the Global Context

OECD Guidelines Governing the Protection of Privacy and Trans-Border Flows of Personal Data

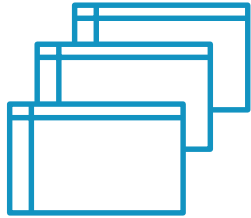


Basic Principles

- Collection Limitation
- Data Quality
- Purpose Specification
- Use Limitation
- Security Safeguards
- Openness
- Individual Participation
- Accountability

Services

**Manufacture
Datasets**



**Data
Science**



Applications



Alerting



BI / Analytics



Search



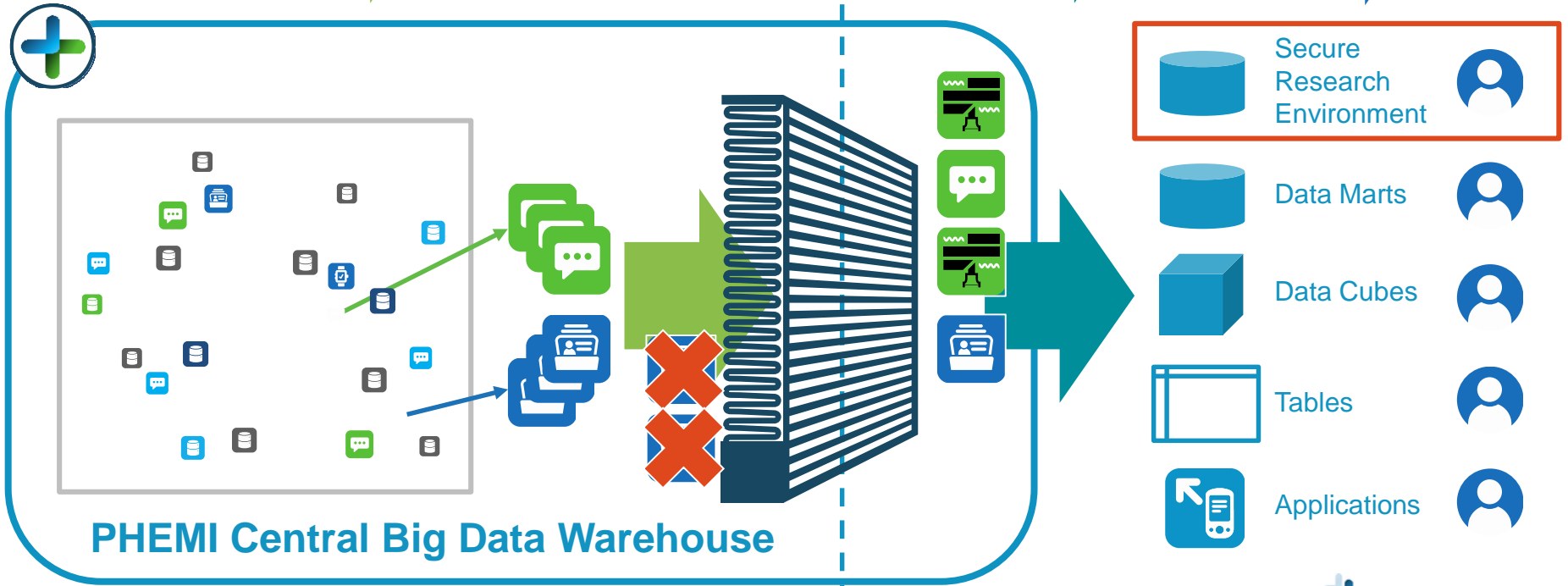
Dashboard



Manufacturing Datasets

Dynamically Manufacture Datasets

Gather → Filter, Transform, and Release → Structure → Users



Filter, Transform & Release

- Automatically mask, or de-identify data tagged as private or restricted
- Apply complex algorithms such as Safe Harbor Guidelines
- Policies for data sharing agreements, consent, compliance
- Structure for analytics, visualization, application



- Allow authorized, but only authorized, access to data
- Consider:
 - User / system identity
 - Where the data is going
 - Attributes of the data

Manufacturing Datasets: a Healthcare Example

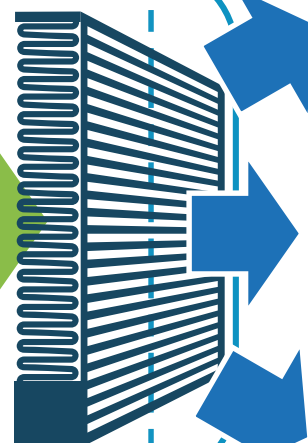
Gather → Filter and Transform → Release



Dataset

Name	PHN	State	DOB	Sex	LDL	HDL	Triglyc	...
Stan	13542	NY	15-05-62	M	2.7	1.1	2.0	
Emil	26534	OK	27-12-47	M	2.8	1.6	1.4	
Carol	94613	TX	03-06-83	F	3.2	1.6	1.6	
Padma	57236	FL	16-09-55	F	1.7	1.3	2.1	
Qing	46478	CA	08-05-79	M	2.9	1.9	1.7	
Madhavi	92741	NV	01-03-68	F	1.9	1.8	1.3	
Jenny	13562	OR	13-09-75	F	3.0	1.5	1.5	
Amir	35253	WA	28-08-55	M	2.9	1.1	1.1	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

PHEMI Central Big Data Warehouse



The right information to the right person with the right context at the right time.

Clinical Information Systems

Registries and Databases

Lab Results and Pathology Reports

Enterprise Data Warehouse

Collaborators

Libraries

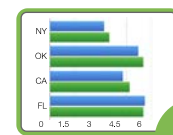
Patient Name: Stan PHN: 13542
 Test Date: 2015-10-09
 Lab: Metrolab
 Your cholesterol level: 2.8 mmol/L
 Target cholesterol level: 1.7 mmol/L
 You are 8% closer to your goal
 Your heart health rating is 6.5

Patient views lab test results

Clinician gets full patient health record
 Patient Name: Stan
 PHN: 13542
 Address: 123 Any Street, Town, NY
 Zip Code: 10029
 DOB: 15-05-62
 Weight: 187
 LDL: 2.7
 HDL: 1.1
 Triglycerides: 2.0
 Total Cholesterol: 4.2

DOB	Sex	LDL	HDL	Triglyc	Total Chol
***62	M	2.7	1.1	2.0	4.2
***47	M	2.8	1.6	1.4	4.7
***83	F	3.2	1.6	1.6	5.1
***55	F	1.7	1.3	2.1	3.5
***79	M	2.9	1.9	1.7	5.1
***68	F	1.9	1.8	1.3	3.96
***75	F	3.0	1.5	1.5	4.8
***55	M	2.9	1.4	1.1	4.5
:	:	:	:	:	:

Researcher accesses report on cholesterol by age and gender, including masked patient data and calculated values



Statistician gets histogram summary, patient cholesterol rates, no PHI

Use Cases

Starring Roles of Big Data Example: Global Insurance Company Uses PHEMI Big Data



Self Serve Data
Catalog

Custom
Applications

Analytics
Tools

Analytics-Ready Data



PHEMI Central
Big Data Warehouse



Claims



Voice
Recording



Customer
Survey



Social Media



Email



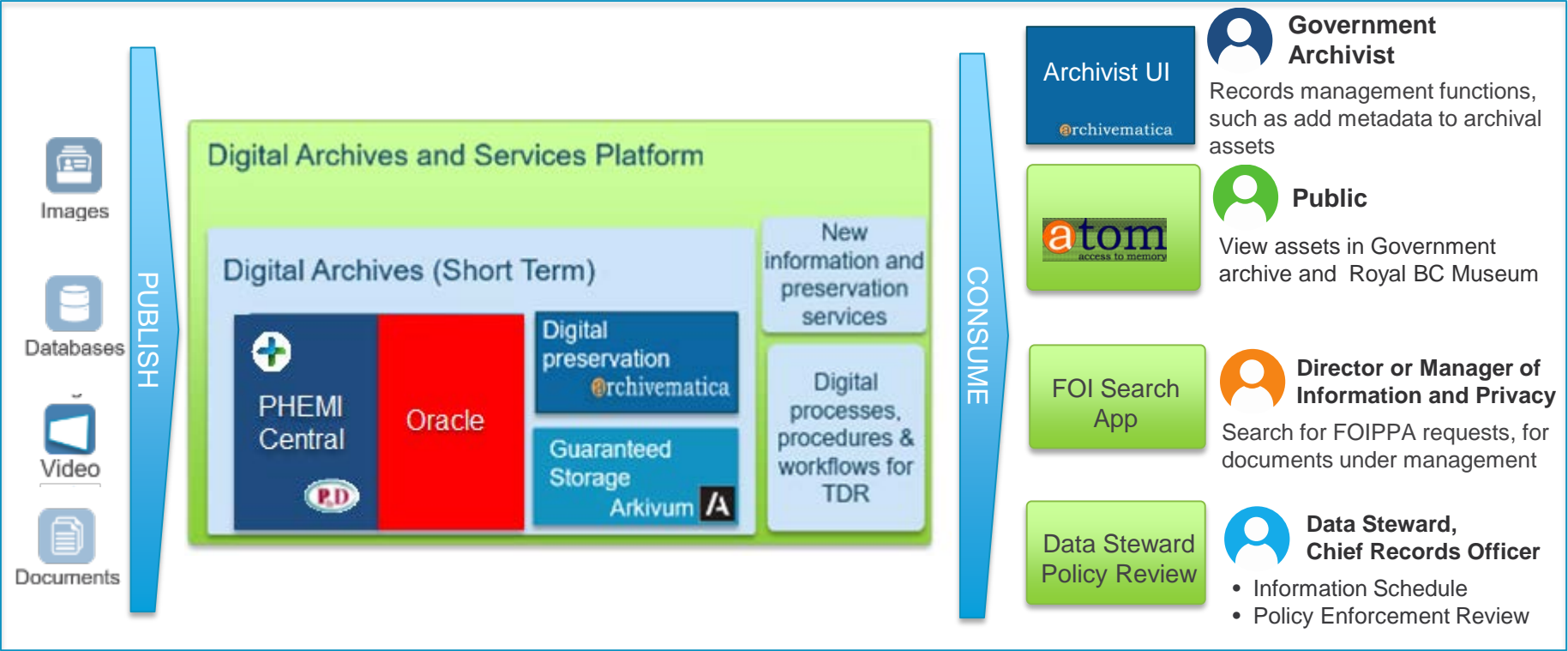
Clickstream



Call Center
Records

- Data Science to improve Enterprise Call Center – Making Data Actionable
 - Sentiment analysis
 - Natural Language Processing (NLP)
- Replace ODS
- Publish Datasets
- Active Archive

Records Management, Digital Archives, and Services



Thank You