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?

Data Available
analysis, insights, and new services
How do I share data for secondary use without compromising privacy, security, and governance?
Digital Library
Managing Digital Assets

Index of a Book

Library Card
Transforming Data into Capital

Gather Everything

Curate

Services
Gather Everything

- X-Ray
- Customer Service Call Audio
- Gene Sequence
- EMR Form
- Emails
- Internet of Things
- Social Media
- Web Pages
- Documents
- Virtual Machine
- Video
- Code Fragment
- Database
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

PHEMI Central Big Data Warehouse

Secure Research Environment

Data Marts

Data Cubes

Tables

Applications
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

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

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Gather</th>
<th>Filter and Transform</th>
<th>Release</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PHEMI Central Big Data Warehouse</strong></td>
<td><strong>Clinical Information Systems</strong></td>
<td><strong>Registries and Databases</strong></td>
<td><strong>Lab Results and Pathology Reports</strong></td>
</tr>
<tr>
<td><strong>Enterprise Data Warehouse</strong></td>
<td><strong>Collaborators</strong></td>
<td><strong>Libraries</strong></td>
<td></td>
</tr>
</tbody>
</table>

**Clinical Information Systems**

- Patient Name: Stan
- Birth Date: 12-06-1980
- Address: 123 Any Street, Town, NY
- Zip Code: 10019
- DOB: 12-06-40
- Weight: 187
- Height: 5'7
- Age: 35
- Medical History: 2.0
- Total Cholesterol: 4.2

**Registries and Databases**

- **Clinical Information Systems**
- **Registries** and **Databases**
- **Lab Results** and **Pathology Reports**

**Enterprise Data Warehouse**

- **Clinician**
  - Gets full patient health record
- **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

**Patient views lab test results**

**Statistician**

- Gets histogram summary, patient cholesterol rates, no PHI

**Clinician**

- Gets full patient health record

**Researcher**

- Accesses report on cholesterol by age and gender, including masked patient data and calculated values

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

**PHEMI**

- Copyright 2016
Use Cases
Starring Roles of Big Data Example:
Global Insurance Company Uses PHEMI Big Data

- 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

Digital Archives and Services Platform

Archivist UI
Records management functions, such as add metadata to archival assets

Public
View assets in Government archive and Royal BC Museum

Director or Manager of Information and Privacy
Search for FOIPPA requests, for documents under management

Data Steward, Chief Records Officer
- Information Schedule
- Policy Enforcement Review

Digital Archives (Short Term)
- PHEMI Central
- Oracle
- Guaranteed Storage
- Digital preservation

New information and preservation services
- Digital processes, procedures & workflows for TDR

Images
Publish
Database
Video
Documents
Consume
Thank You