Know thy (Health) Data…

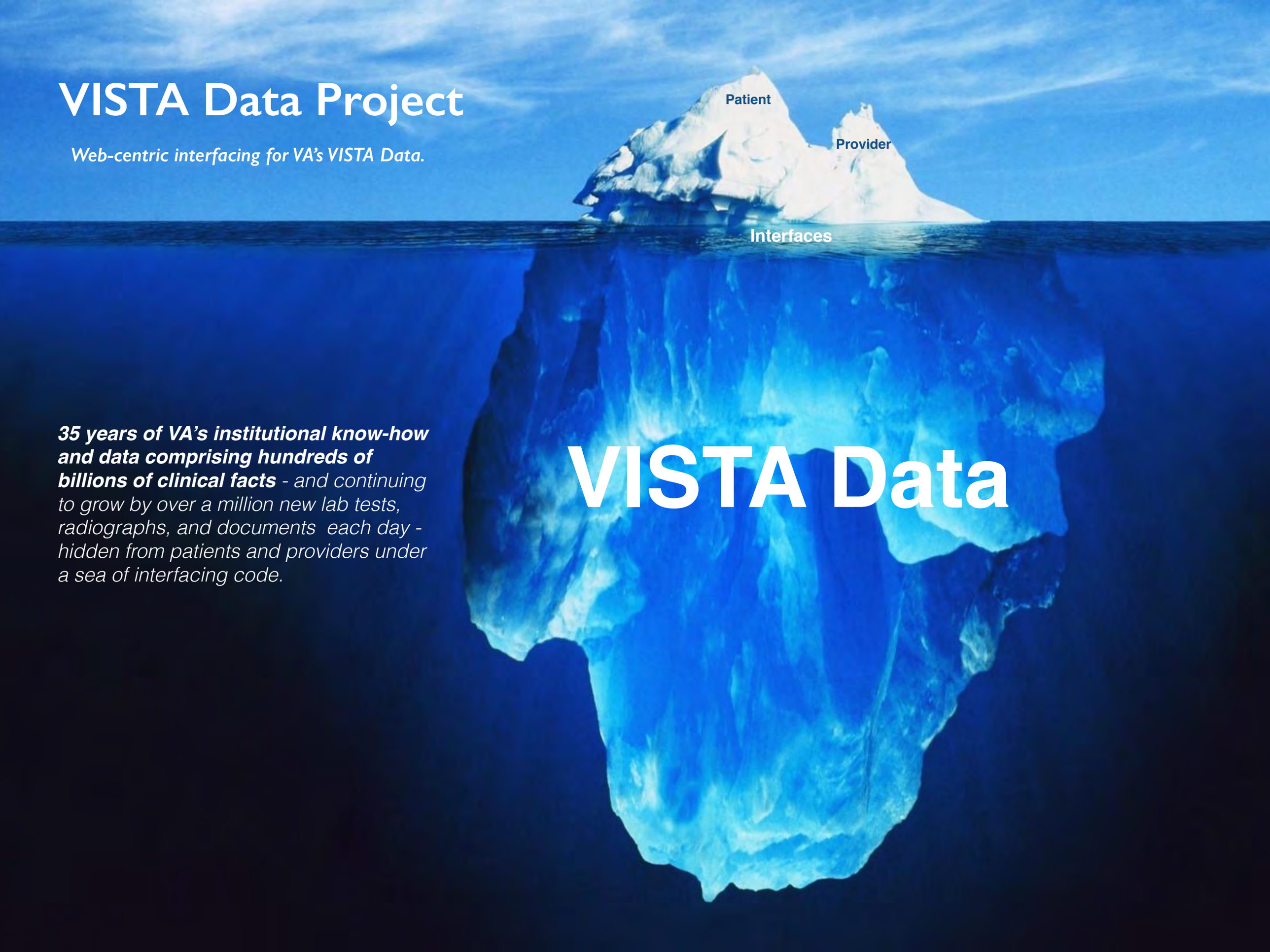
A Linked Data Model Approach to Web-centric Integration

Nov 8, 2016
NHS Interoperability Summit
London, UK

Rafael Richards MD MS
Director, VISTA Data Project
Director, Interagency Metadata Program
Pacific Joint Information Technology Center
VHA Office of Informatics and Analytics
U.S. Department of Veterans Affairs

VISTA Data Project

Web-centric interfacing for VA's VISTA Data.



35 years of VA's institutional know-how and data comprising hundreds of billions of clinical facts *- and continuing to grow by over a million new lab tests, radiographs, and documents each day - hidden from patients and providers under a sea of interfacing code.*

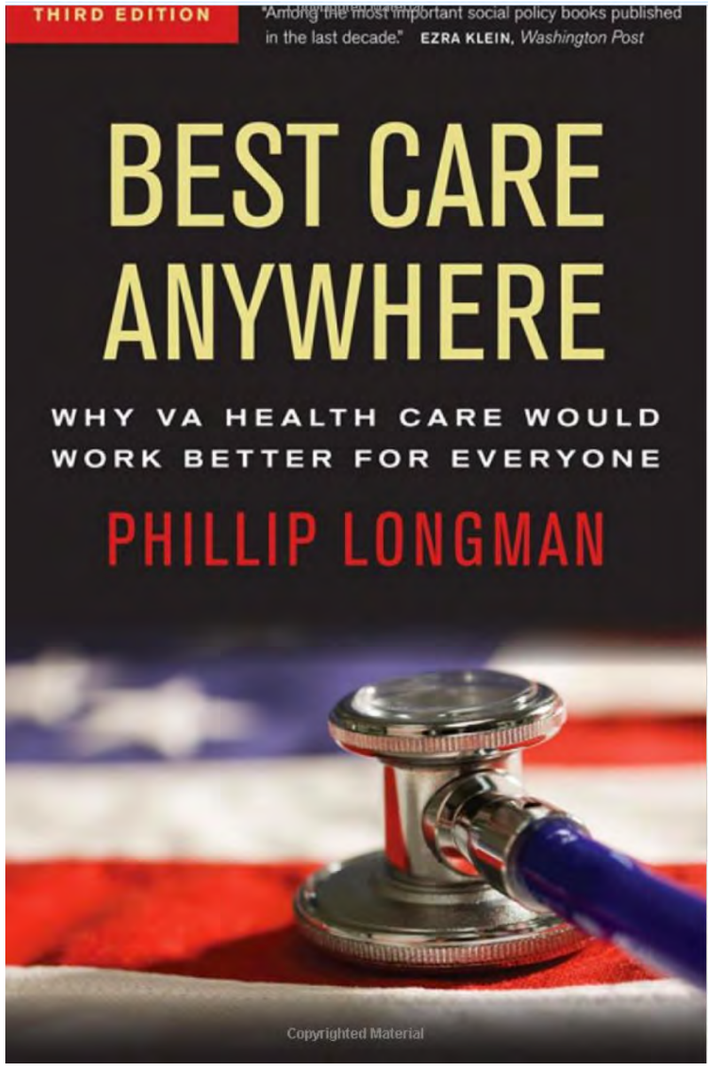
The Largest National health services

http://www.2020health.org/2020health/Publications/Publications-2013/Making-Connections.html

U.S. Veterans Health Administration (VHA)

VHA is an integrated network of 1200 hospitals and clinics supported by a fully digital health information system

U.S. Veterans Health Administration (VHA)

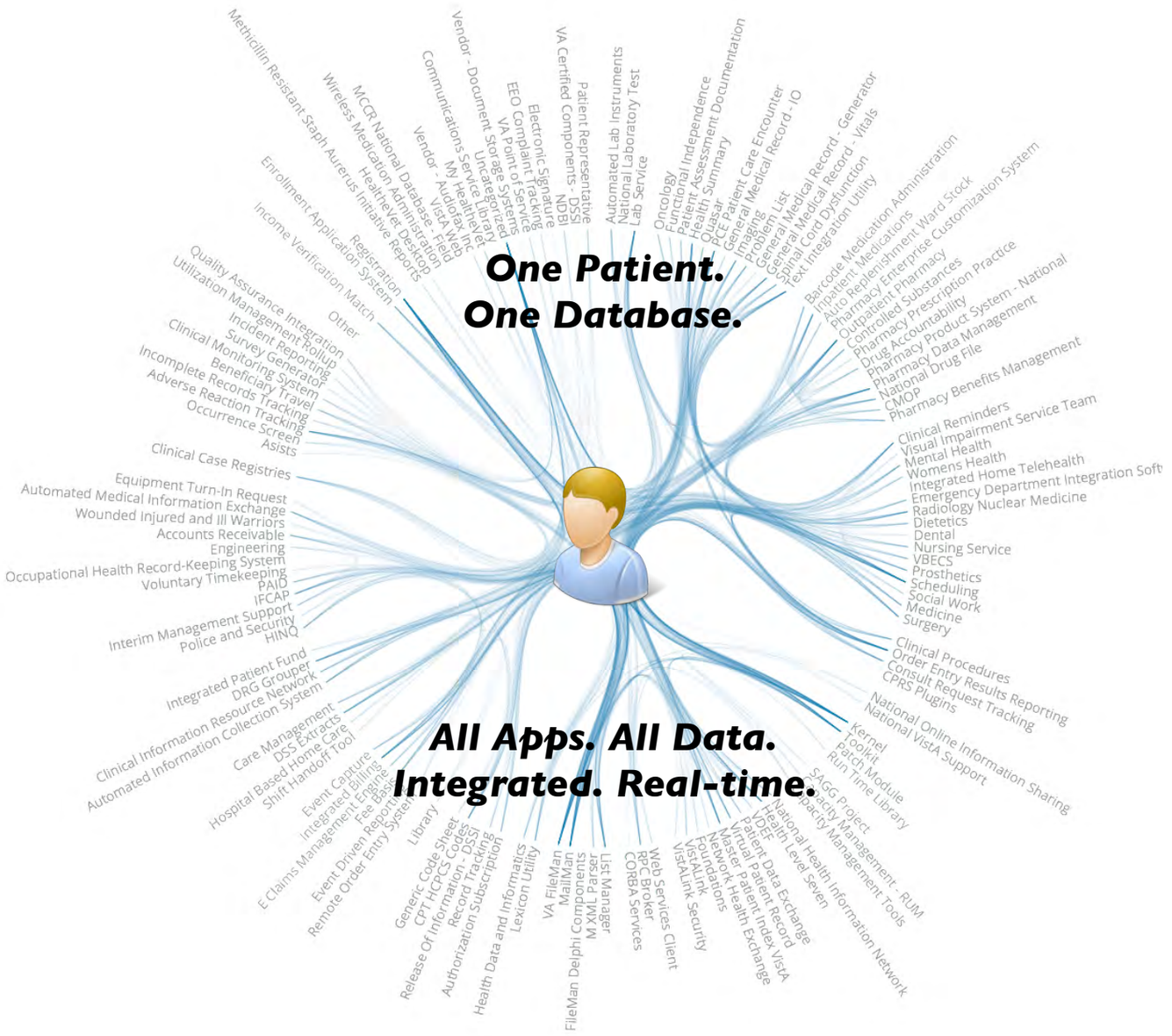


HITECH - U.S. Digital Health

http://www.va.gov/health/IMAGING/docs/Best_Care_Anywhere.pdf

Veterans Information Systems Technology Architecture (VISTA)

A Patient-Centric
Single integrated EHR



https://en.wikipedia.org/wiki/VistA

Patient-centric
Health Record

The data architecture of VISTA consists of over 180 modules for clinical care and administration integrated within a single common multidimensional data engine (MDE).

In VISTA, both business logic (Applications) and data (Database) are managed within the multidimensional data engine. This provides the tight integration of applications to data, and to a single common integrated database.

The integration between VISTA applications (outer ring) and VISTA data (inner circle) is visualized, showing the shared data flow between applications.

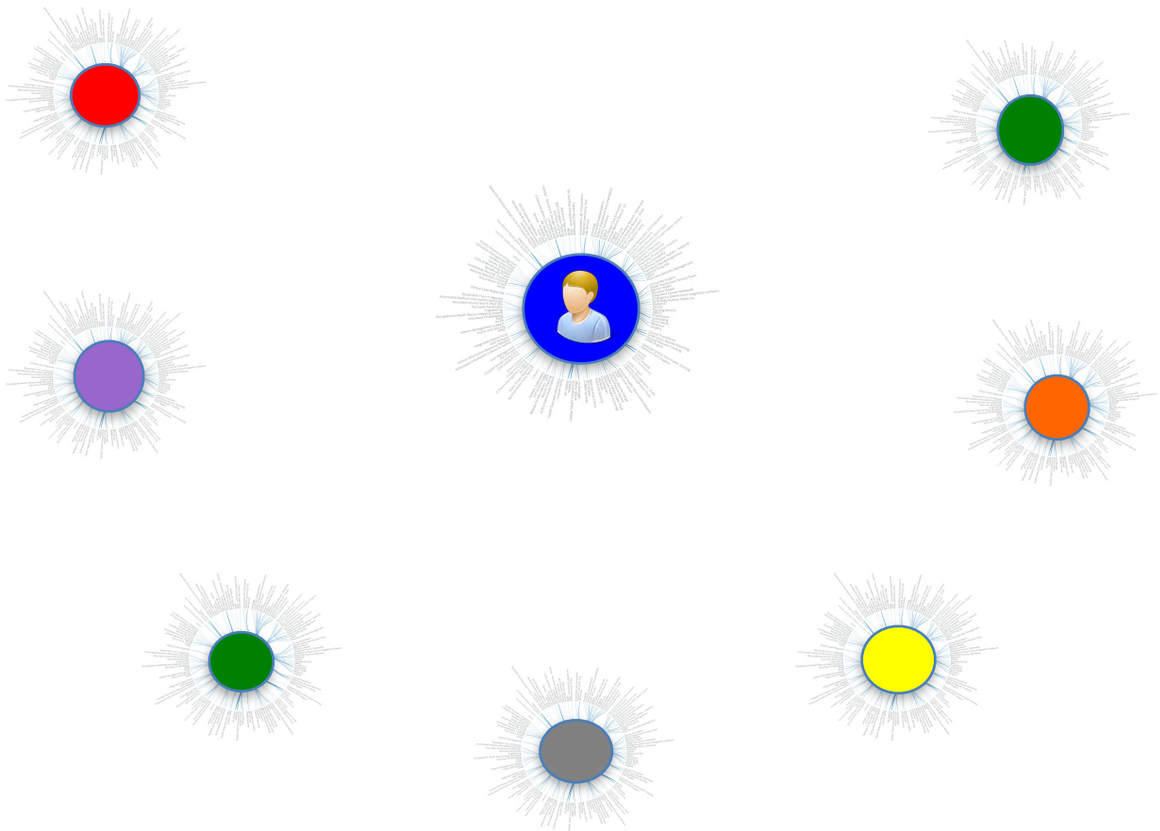
Evolutionary Challenges

Over the past 35 years VISTA has evolved locally to become 131 unique systems. As an enterprise, this fragments patient data and care within VHA.

Externally, it makes it challenging to interface or integrate VHA health data to new technologies and patient care partners outside VA.

Problem: Silo VISTAs

One Patient.
Many Databases.



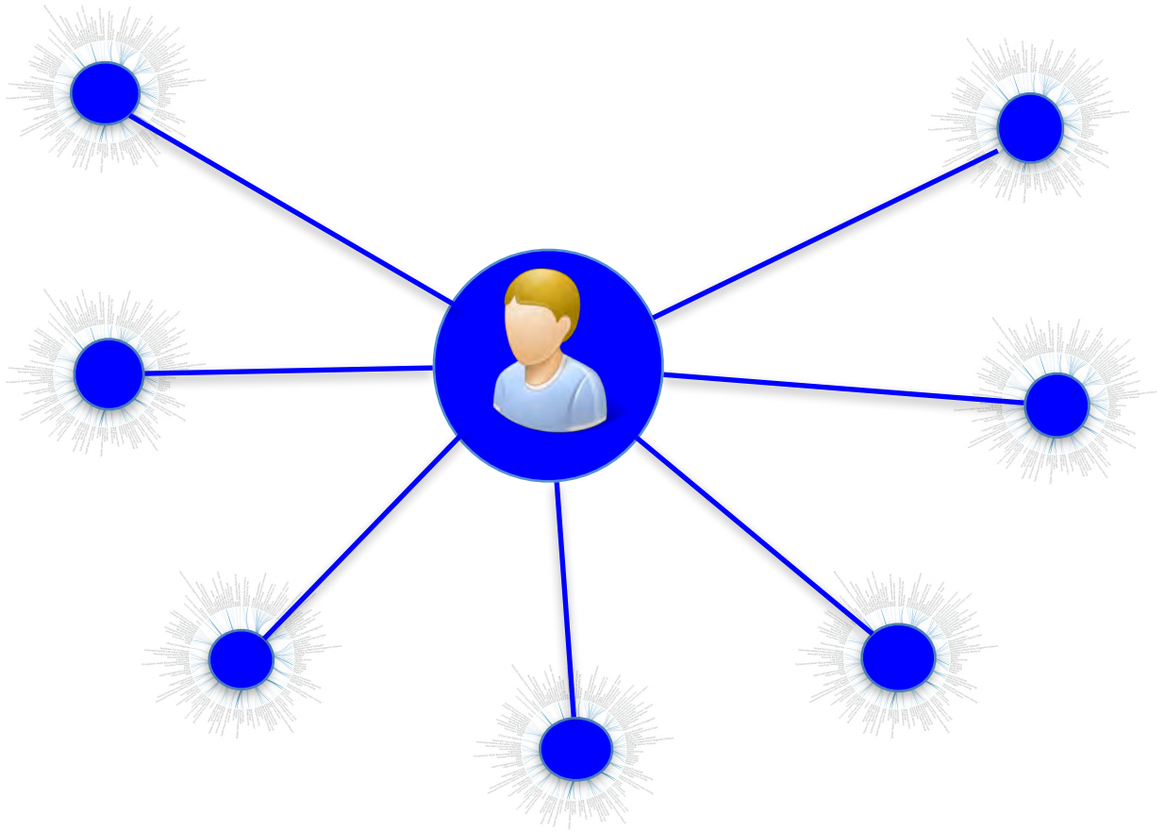
Many VISTAs.
Many Models.
Fragmented Data.

While each VA VISTA system is individually highly integrated, each system has its own distinct data model (shown as different color).

Different data models creates data silos – fragmenting patient data and fragmenting patient care.

Approach: ***Linked Data Model***

One Patient.
One Model.



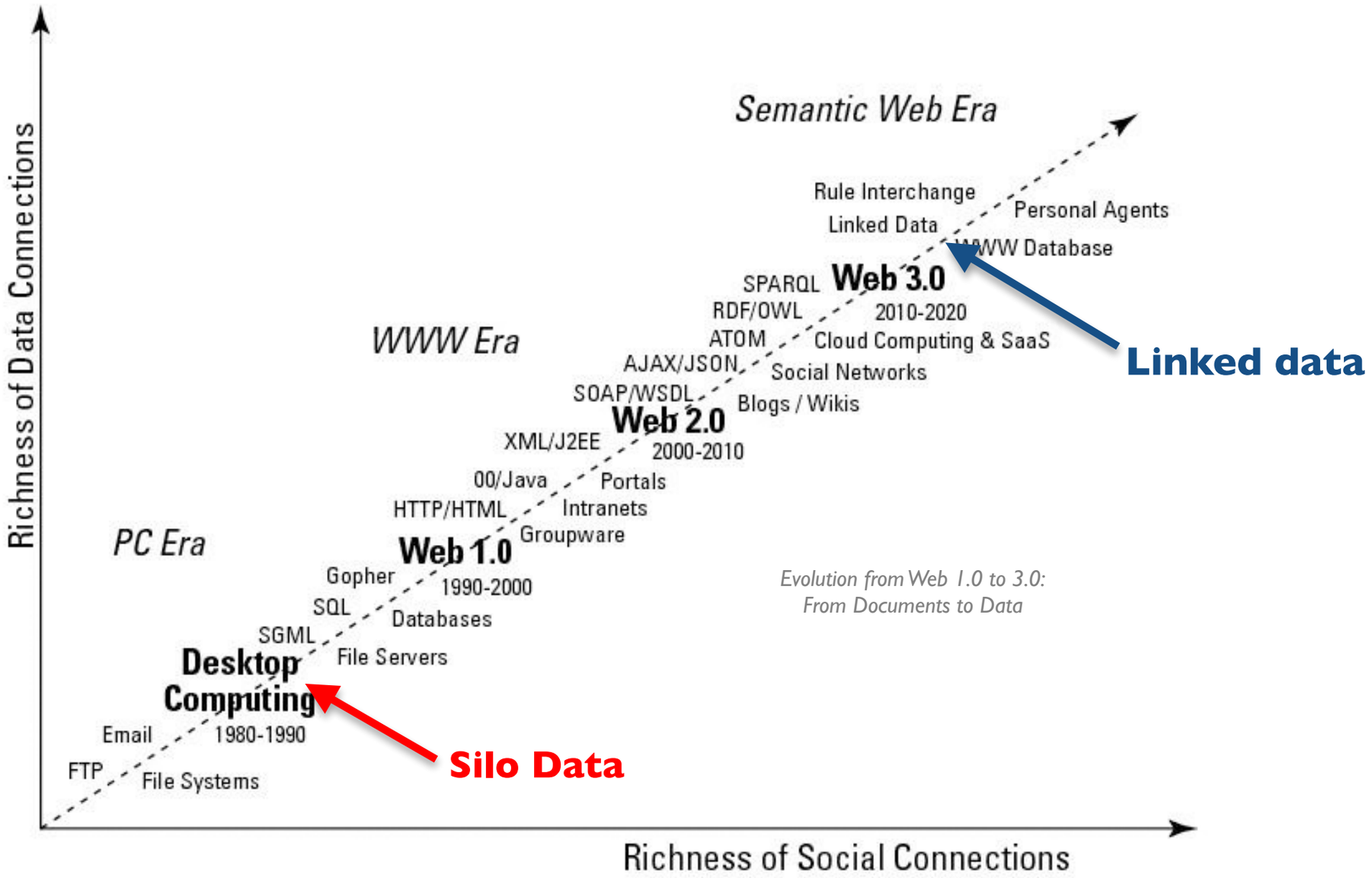
Many VISTAs.
One Model.
Integrated Data.

Exposure and cross-linkage of VISTA's local models to a common web-standard, web-scale Linked Data Model (Master VISTA Data Model) provides an integrated view of all patient data from all VISTA systems.

Linked Data

- *Evolution of Computing: Increased connectivity*
- *What is it Linked Data?*
- *What problems does it solve?*
- *Who uses Linked Data?*
- *Health Data: Many diverse models*
- *Linked Data: Accommodates model diversity*
- *Health Data: PCAST Recommendation*

Evolution of Data: ***Increased Connectivity***



Evolution from Web 1.0 to 3.0:
From Documents to Data

In 1981 VISTA began as the Decentralized Hospital Care Program. This was in a period when databases were isolated within institutional networks (Silo Data). After the HTTP protocol was developed in 1988, the World Wide Web evolved from a document web to a data web. To bring any system into today's world wide data web, it needs to be web data standards compliant. (Linked Data).

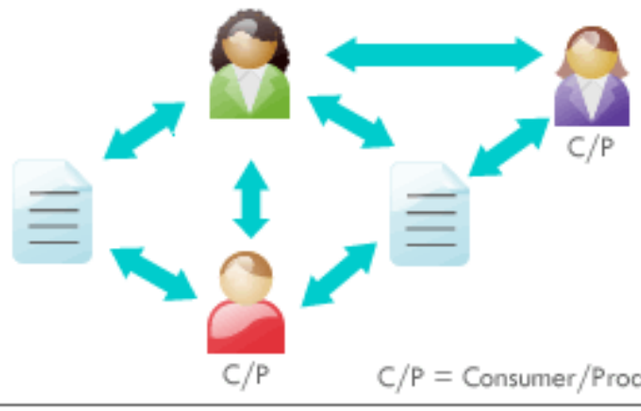
http://en.wikipedia.org/wiki/History_of_the_Internet
http://paanchiweb.blogspot.com/2012/12/getting-essence-of-semantic-web.html

Evolution of the Web: ***From Documents to Data***

Web 1.0

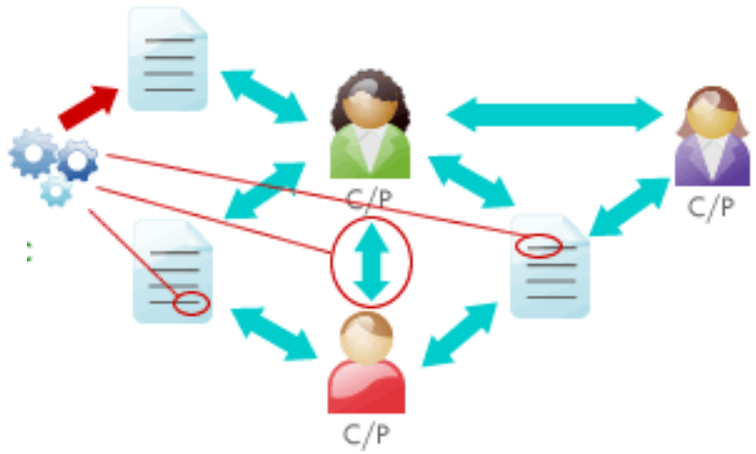
Linked Documents
Document Web (HTML)
Read-only web (**humans** only)

Web 2.0



Linked People
Social Web
Read-write web (**humans** only)

Web 3.0



Linked Data
Semantic Web (RDF)
Read-write web (**machine processable**)

http://tomgruber.org/writing/CollectiveKnowledgeSystems.htm

Linked Data: ***What is it?***

The World Wide Web (W3C) Standard for semantic information integration

HTML (hypertext markup language)
For **humans** to exchange information

enables →

Linked Documents
(Document Web)

RDF (resource description framework)
For **computers** to exchange information

enables →

Linked Data
(Semantic Web)

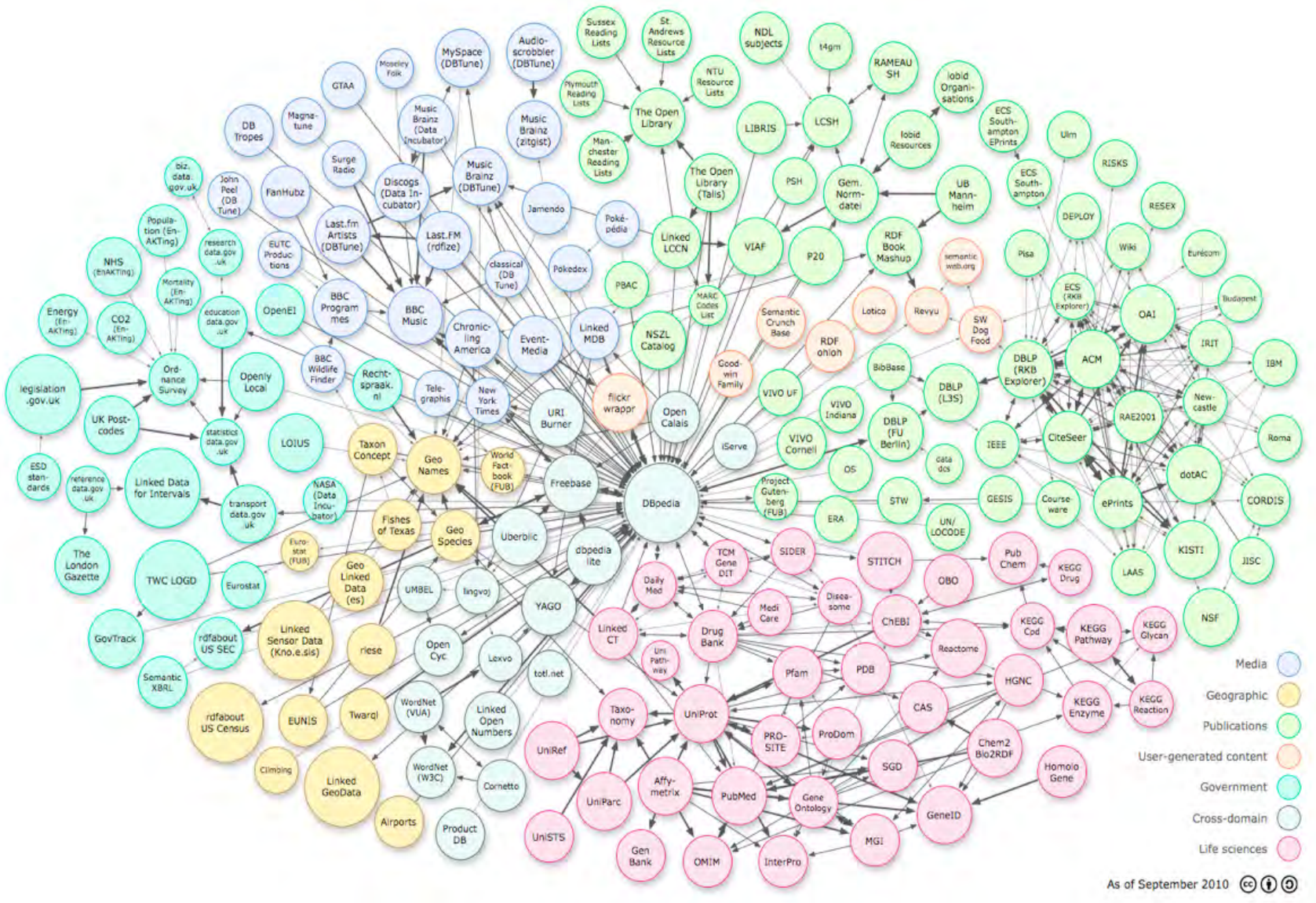
"The Semantic Web [Linked Data] provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries."

Tim Berners-Lee, MIT Professor and Inventor of the World Wide Web

As a W3C standard this supports Internet-scale data integration.

Linked Data: What does it enable?

Web-scale semantic integration of data



Linked Data

This figure shows the Linked Open Data (LOD) cloud, which semantically links hundreds of Linked Data sources including Media, Geographic, Government, and Life Sciences databases.

Each circle represents one data source or database. These are semantically linked to other data sources, creating a single virtual federated internet-scale database.

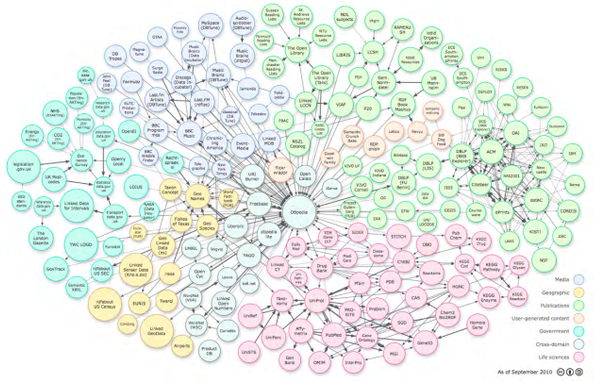
At the center of is DBPedia, the Linked Data version of Wikipedia, which is semantically linked to hundreds of data sources.

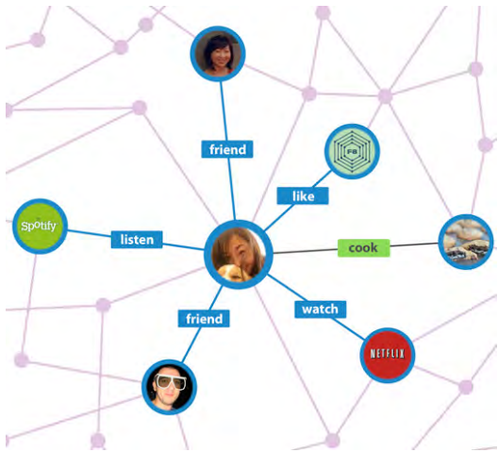
Linked Data: ***Who Uses It?***

The Linked Data approach to ***Internet-scale semantic data integration and search*** *by the world's largest data management organizations such as Google, LinkedIn, Facebook, and IBM Watson.*

Watson
(knowledge graph)

Wikipedia
(knowledge graph)





Facebook
(social graph)

LinkedIn
(professional graph)

Google
(knowledge graph)

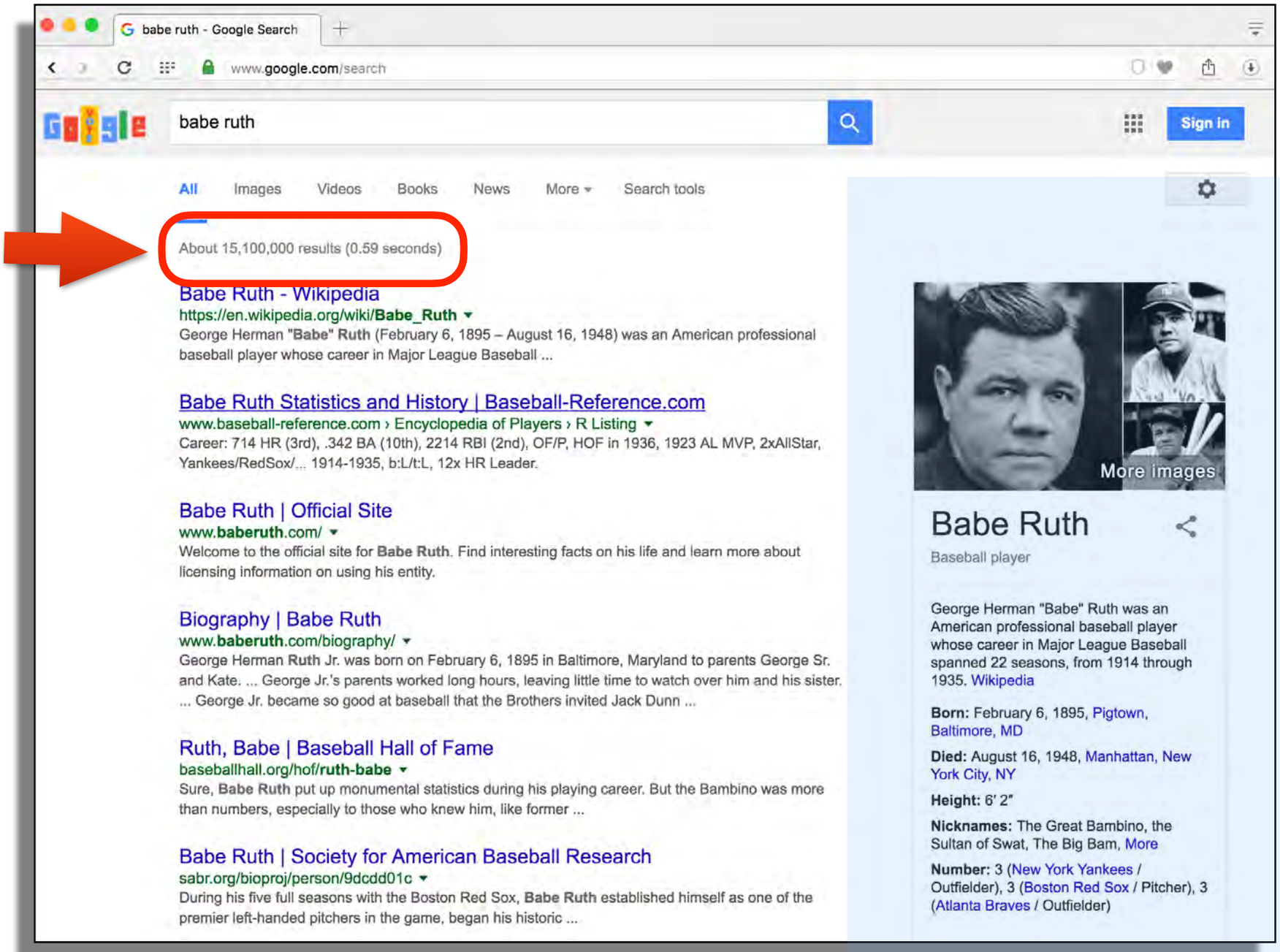
Linked Data: **Semantic Search**

All the major search engines (Google, Yahoo, Microsoft, Yandex…) use the same shared web schema in RDF to index, search, and structure all data on the web, making it all semantically searchable.

From Strings to Things

Google Guess

Over 15 million guesses. String matching. Statistical page rank algorithms. "Black box"



Strings

Things

Google Know

Single, exact, semantic result. Based on on knowledge graph of Linked Data and schema.org (See Linked Data symbol)

U.S. Healthcare: PCAST Recommendations

REPORT TO THE PRESIDENT
REALIZING THE FULL POTENTIAL OF
HEALTH INFORMATION TECHNOLOGY
TO IMPROVE HEALTHCARE
FOR AMERICANS:
THE PATH FORWARD

Executive Office of the President
President's Council of Advisors
on Science and Technology

"The best way to manage and store data for advanced data analytical techniques is to break data down into the smallest individual pieces that make sense to exchange or aggregate. These individual pieces are called "tagged data elements," because each unit of data is accompanied by a mandatory "meta data tag" that describes the attributes, provenance, and required security protections of the data.

The indexing and retrieval of metadata tagged data, across large numbers of geographically diverse locations, is an established, highly developed, technology—the basis of web search engines, for example".

- ***Linked Data (RDF) is the World Wide Web standard for semantic metadata tagging for data on the web, used by all major search engines.***

http://www.whitehouse.gov/sites/default/files/microsites/ostp/pcast-health-it-report.pdf

U.S. Healthcare: PCAST Recommendations

UNITED STATES
DEPARTMENT OF VETERANS AFFAIRS

Search All VA Web Pages | Search | » Open Advanced Search

Home | Veteran Services | Business | About VA | Media Room | Locations | Contact Us

- One-VA TRM Home
- Search One-VA TRM
- Technology/Standard List
- TRM Reports
- Submit Idea to Improve TRM
- TRM Glossary
- Site Map

ONE-VA TECHNICAL REFERENCE MODEL V15.1

TRM **Resource Description Framework (RDF)** ⟵ *(Linked Data)*

General | Decision | Reference | Component | Category | Analysis

General Information

Technologies must be operated and maintained in accordance with Federal and Department security and privacy policies and guidelines. More information on the proper use of the TRM can be found on the TRM Proper Use Tab/Section.

Website: Go to site

Description: The Resource Description Framework (RDF) is a family of World Wide Web Consortium (W3C) specifications. They were originally designed as a metadata data model. It has come to be used as a general method for the conceptual description or modeling of information that is implemented in web resources, using a variety of syntax formats.

Decision: View Decisions

Decision Constraints:

Decision Source: TRM Mgmt Group

Decision Process: One-VA TRM v13.10

Decision Date: 10/25/2013

Introduced By: TRM Request

Standards Body: W3C

http://www.va.gov/TRM/StandardPage.asp?tid=6405#

Linked Data: JSON-LD

The Resource Description Framework (RDF) has many serializations. The form most commonly used for web applications is JSON-LD,

Implementation

Master Linked Data Model

RM Richards MD MS 2016-11-08

VISTA Interfacing: ***Master Linked Data Model***

VA VISTA
Data Project

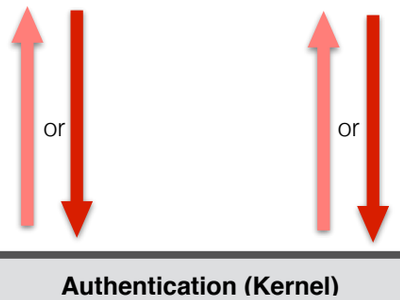
Current
Code-driven VISTA

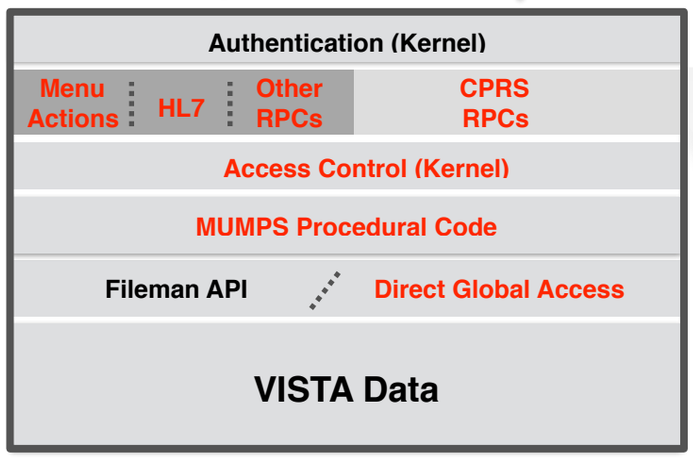
1981-present **1996-present**

Other Clients
(Terminal, HL7, …)

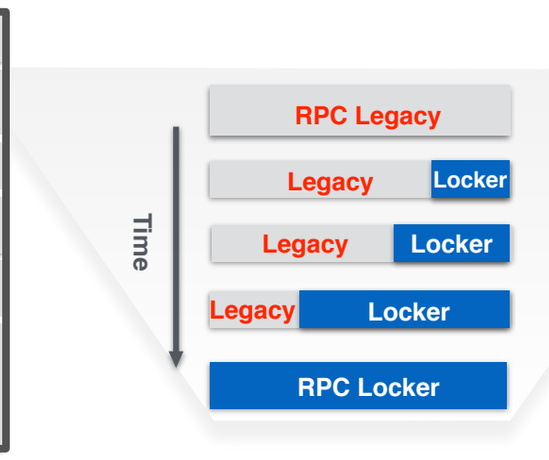
CPRS Clients
(VPR, HMP…)

Thousands of Unique One-way Code-driven Interfaces





Red: Redundant M code
(not required for model-driven VISTA)



Remote Client (RPC)
Isolation and Security Transition

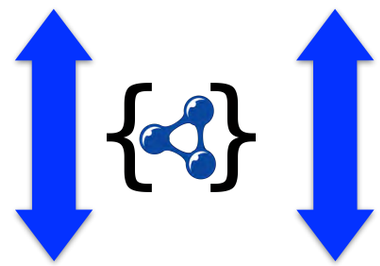
VISTA Data Project
Model-driven VISTA

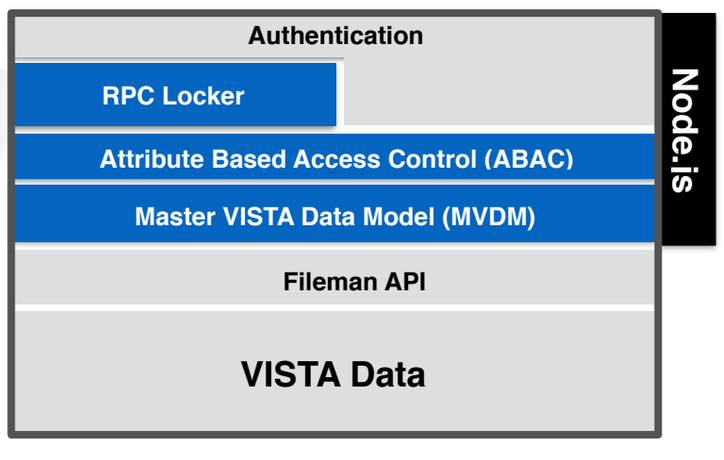
NEW (2016-17)

Existing RPC Clients
Security-enhanced

New Clients & Systems Interop

Single Secure Symmetric Model-driven Interface



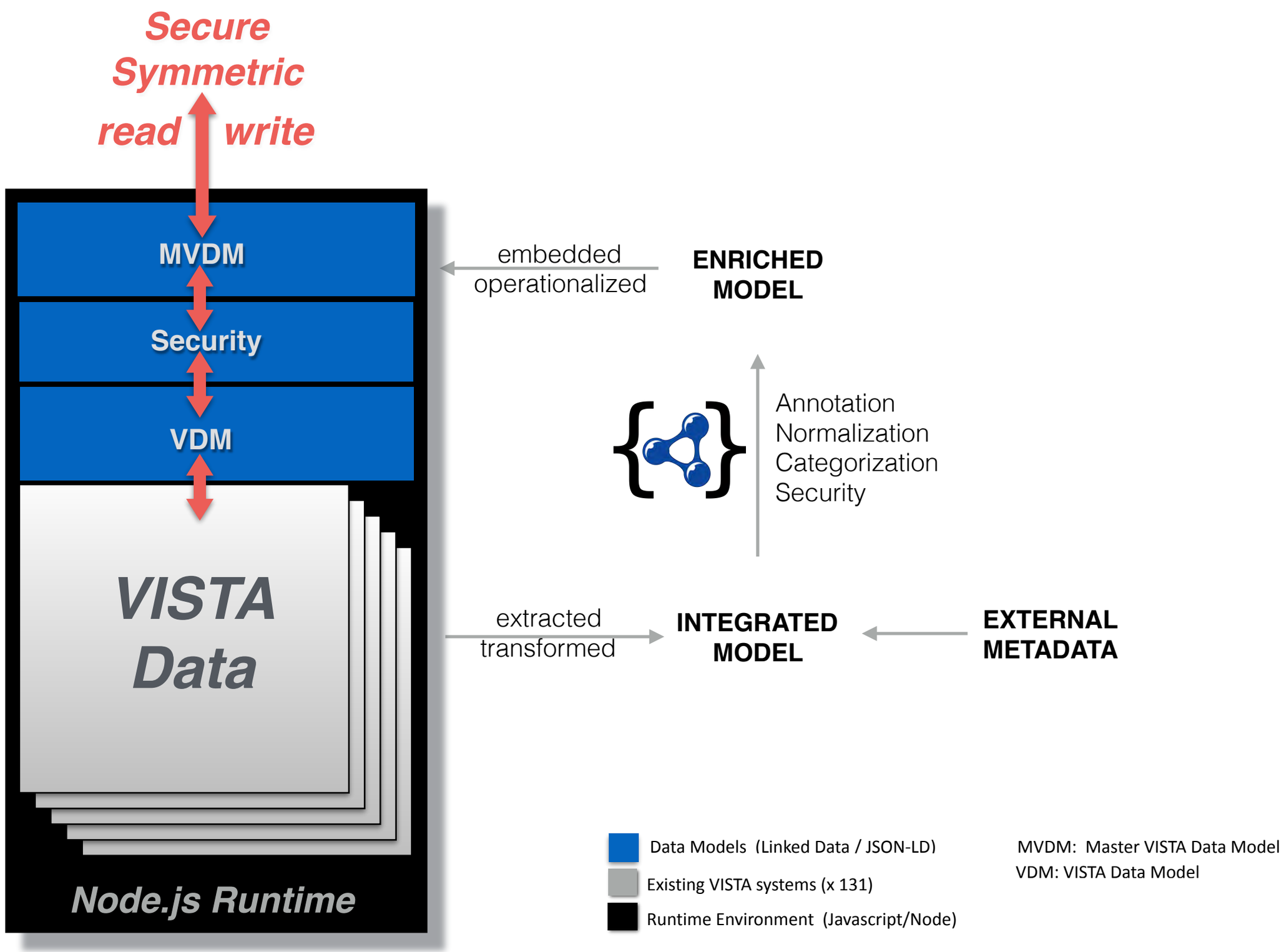


Blue: New, Model-driven modules
(Javascript/Node.js)

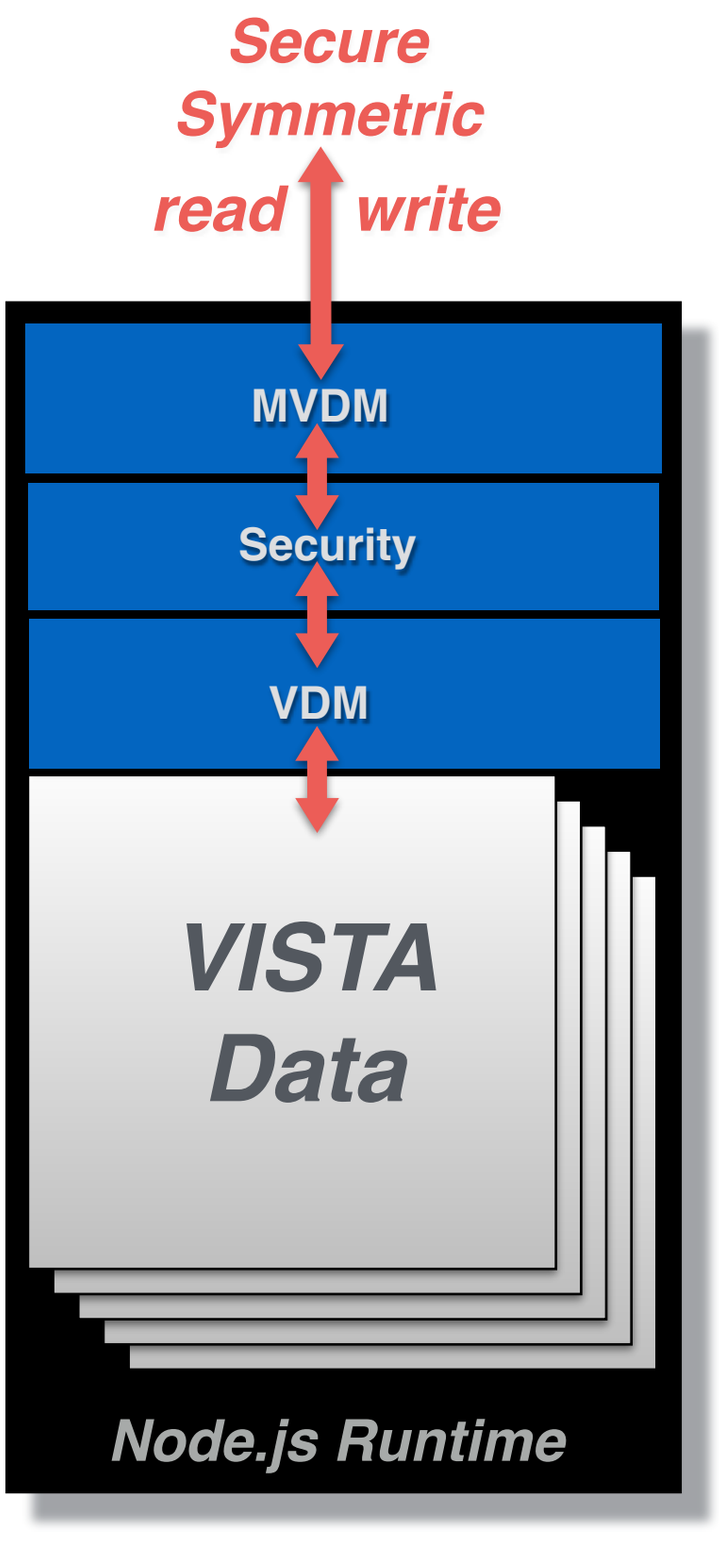
Classic VISTA (MUMPS)

New VISTA (Javascript/Node)

VISTA Interfacing: ***Master Linked Data Model***



VISTA Interfacing: ***Master Linked Data Model***



Linked Data Model
- Industry-standard machine-processable web data model
- Uses schema-backed JSON with Linked Data extensions (JSON-LD)
- All VISTA data models are expressed, processed, and enriched as JSON-LD.

Master VISTA Data Model (MVDM) (x1)
- A subset of VDM that is normalized across all VDMs
- Incorporates all functionality of the Security Model
- Incorporates all functionality of the VDM
- Supports remote secure read-write across all VISTA instances
- Supports Master Data Management across all VISTA instances for any specified data category

Security Model (x1)
- Provides data-centric logical security model for all VA VISTA data.
- Provides data-centric security based on data attributes and categories
- Specifically provides “on-the-data” granular patient-centric data security.

VISTA Data Model (VDM) (x131)
- Represents the full native operational data model of any local VISTA
- Enables comprehensive access to all VISTA data (all 65,000+ data fields)
- Is enriched by additional metadata and logic to support write back
- Provides native symmetric read-write to any local VISTA
- Eliminates need to know anything about VISTA code or internals

VISTA Systems (x131)
- Each contains over 35 years of VA clinical and institutional data

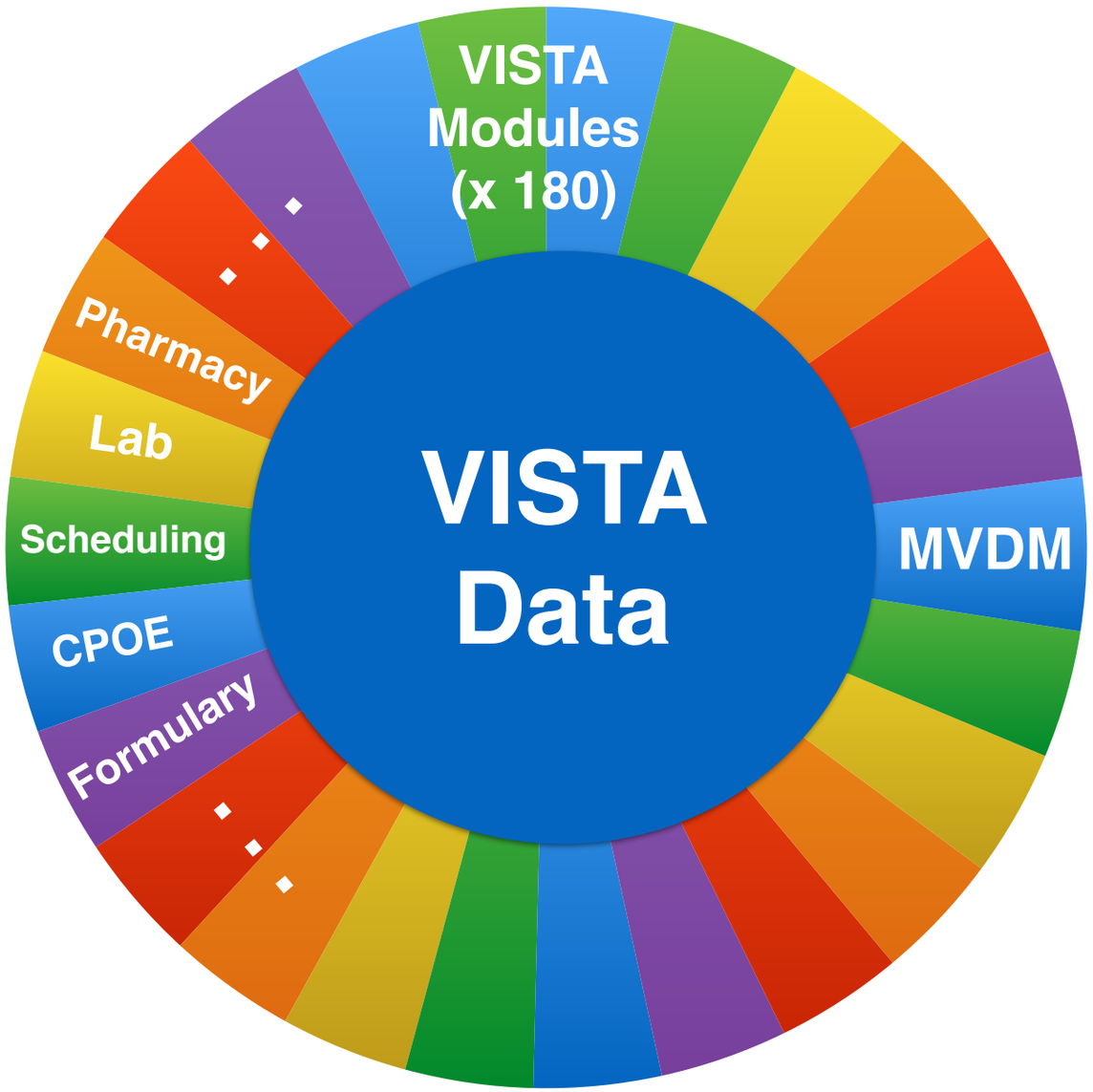
Runtime Environment (Javascript / Node)
- Industry-standard Node.js server-side runtime environment
- All data models and data transformation run in-process, server-side
- All read-write transactions run in-process, server side

Data Models (Linked Data / JSON-LD)

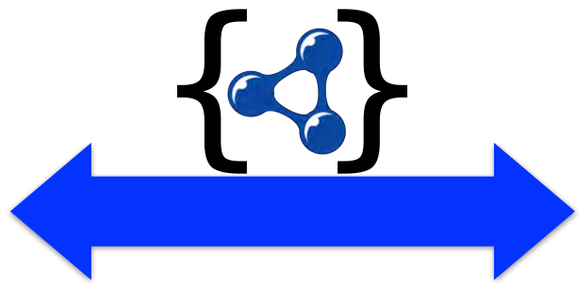
VISTA Interfacing: ***From documents to data***

	Standard "Information Exchange"	Structured, Linked Data Integration
Data Resolution	Document-centric	Data-centric
Data Representation	Documents (XML) Images (Fax,PDF)	RDF (JSON-LD)
Data Storage	Binary (PDFs, scans)	Structured data
Machine Processable	NO	YES
Computable Data	NO	YES
Clinical Decision Support	NO	YES
Supports Analytics	NO	YES
Integrated in Clinical Workflow (Client GUI)	NO (external viewer required)	YES
Integrated into Clinical Health Record (VISTA)	NO (separate data)	YES (data fully integrated)

VISTA Interfacing: ***Master Linked Data Model***



Linked Data Interface



secure symmetric read / write

External Structured Data

All interfacing to VISTA is through one single, secure, symmetric (bidirectional) read-write Master Linked Data Model (MVDM). The read data model is identical to the write data model (i.e. symmetric) providing a single universal structured data read-write mechanism.

VA - Community Care Coordination

VA VISTA
Data Project

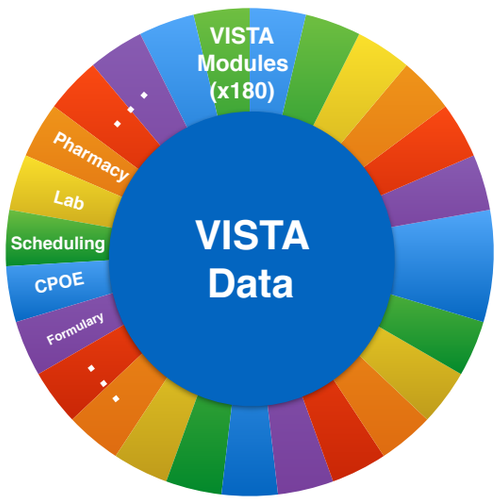
Problem

The "Last Mile". After intake of Community Care data into the VA environment (gateways or repositories) how can structured data be securely integrated the "last mile" to VISTA and leverage all existing, proven, fully-deployed VA clinical workflows, applications, business logic, and clients (including CPRS) so all VA Providers can *immediately* benefit and coordinate care?

VA Care Network

Community Care Network

VISTA (x130)



???

"Last Mile"

"Last Mile" **???**

VA Health Information Gateway

Structured Data

Outbound
Referral
Authorization
Clinical Info
Analytics Data

Inbound
Clinical Info
Admin Info
Analytics Data

Community Health Information Gateway

Community HIE

Community EHR

Structured Health Data Exchange
Clinical Info:
Medical Records
Formulary
Admin Info:
Claims
Referrals
Authorization

CPRS

GUI

VA providers need to see community care data within their existing VA workflows, applications, and clients, including CPRS.

VA Providers

Veteran

Community Providers

VA - Community Care Coordination

VA VISTA
Data Project

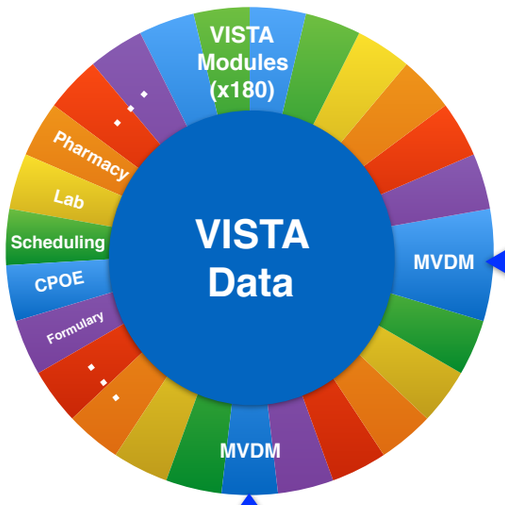
Solution

Master (Linked) Data Model Integration. ***Merge structured Community Care data into all VA VISTA systems through one single secure structured data interface (Master VISTA Data Model), making Community Care data available immediately in computable form in all exiting VA VISTA applications, workflows, and clients (including CPRS).***

VA Care Network

Community Care Network

VISTA (x130)



Linked Data Interface

VA Health Information Gateway / Repository

Structured Data

Outbound
Referral
Authorization
Clinical Info
Analytics Data

Inbound
Clinical Info
Admin Info
Analytics Data

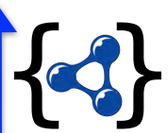
Community Health Information Gateway

Community HIE

Community EHR

Structured Health Data Exchange
Clinical Info:
Medical Records
Formulary
Admin Info:
Claims
Referrals
Authorization

Linked Data Interface



CPRS

Community Care data is integrated in VISTA as structured data and a*vailable immediately in all existing* VA clinical workflows, applications, and clients (including CPRS) to coordinate care.

Care Coordination

GUI

VA Providers

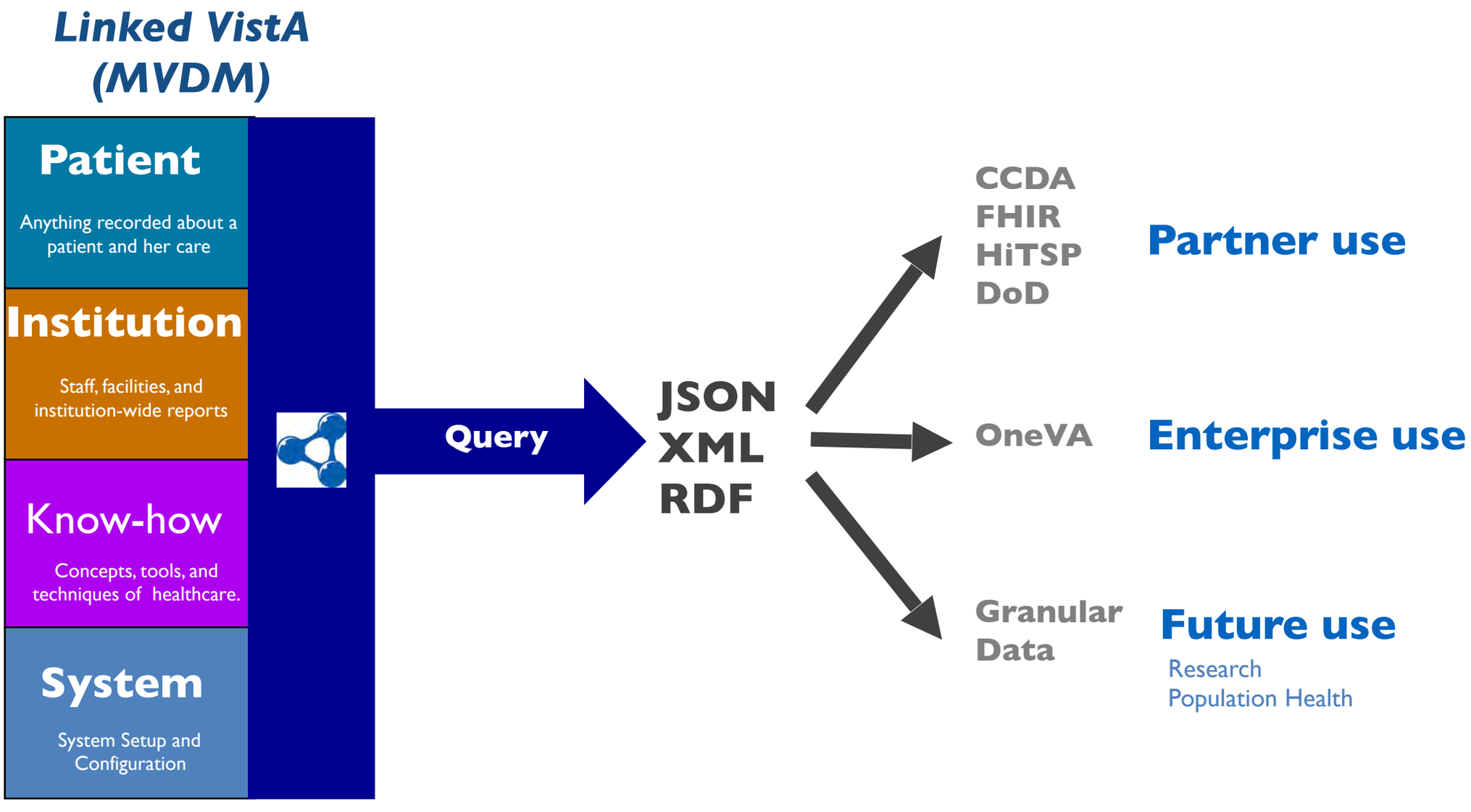
Veteran

Community Providers

Linked VISTA: Future-Enabling Health Data

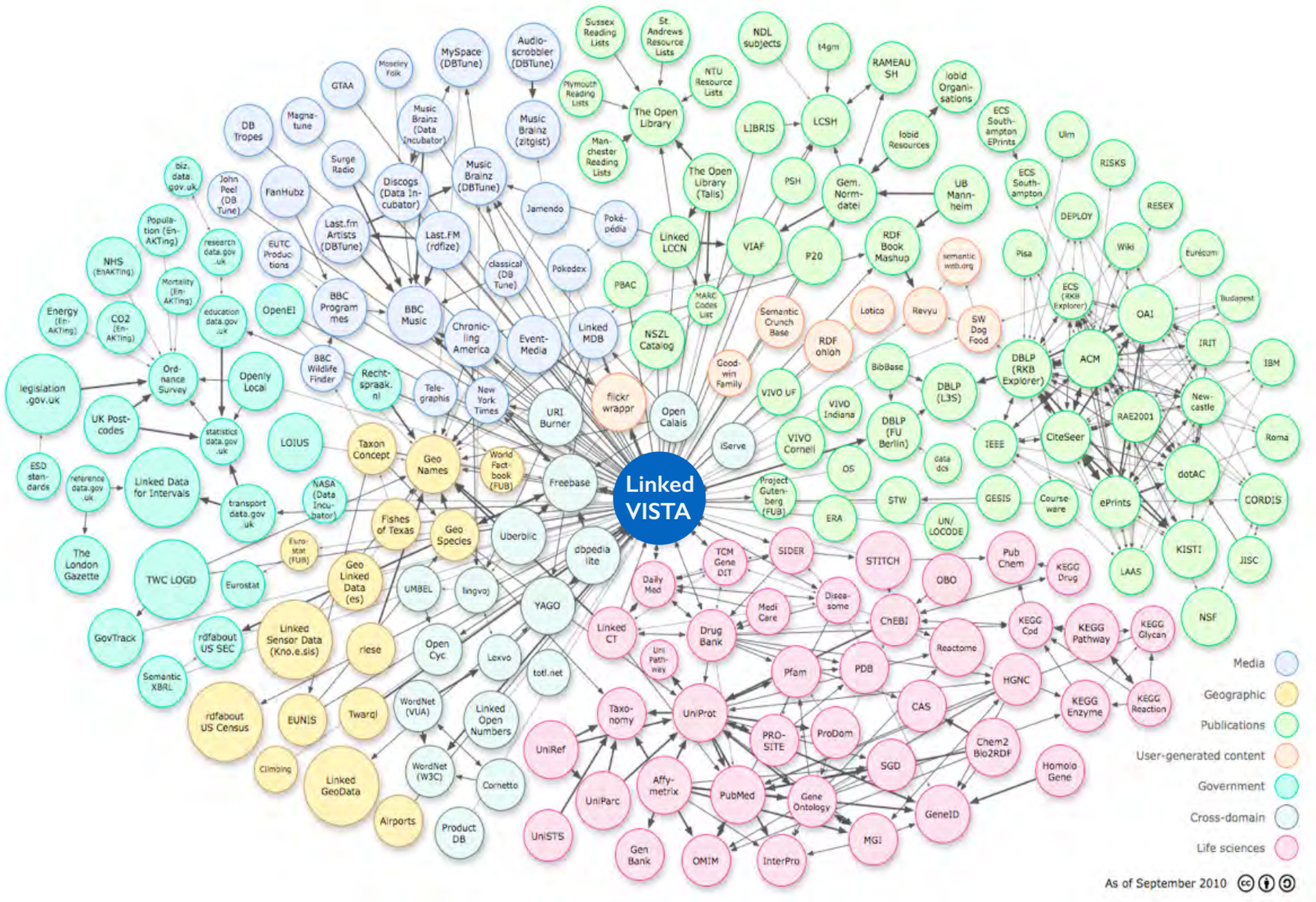
A key benefit of the VISTA master data model (MVDM) is that it can be queried against <u>any</u> VISTA for <u>any</u> data with with <u>one</u> web-standard query.

This would allow any authorized system to securely query authoritative VISTA data in real-time with one standard query interface. The output of these queries can be generated in all web-standard computable to maximize secondary use.



Linked VISTA: ***Enables Web-centric semantic integration***

Managing VistA data as Linked Data supports semantic linkage to thousands of other other Linked Data sources.



Linked VistA

VistA Data - managed as Linked Data can be semantically interlinked with any and all other Linked Data sources.

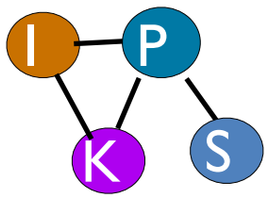
This enables *meshing, enrichment, and augmentation of patient data with any other Linked Data sources, providing an integrated view of all patient data from all locations, clinics, hospitals, or the home.*

Linked Data sources may include patient-generated, mobile, TeleHealth, or any other Internet-enabled device data (Internet of Things).

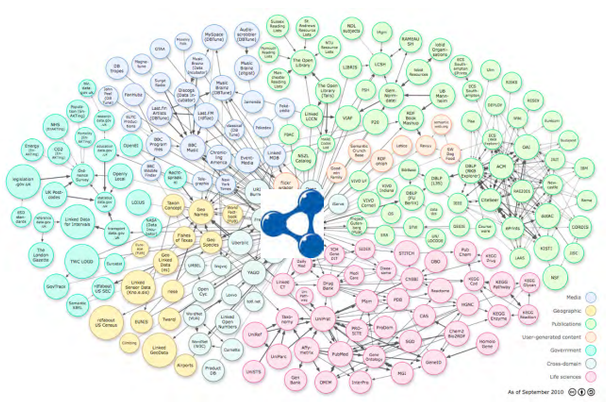
Linked VISTA: ***New Knowledge Discovery***

Linked VISTA can participate in federated queries over unlimited number of other Linked Data sources, enabling meshing, enrichment, and ultimately, new knowledge discovery.

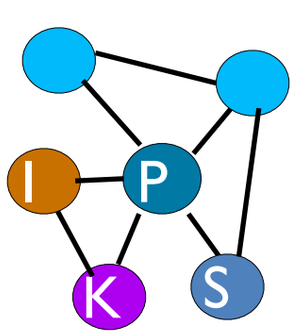
Linked VistA (VistA Model) **+** **Linked Data** (Thousands of sources) **=** **New Knowledge**



+



=

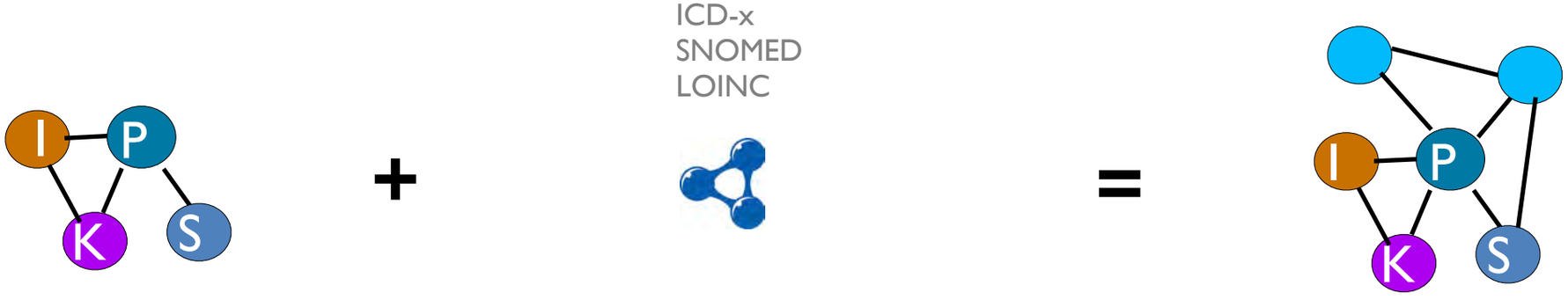


Because data in the VistA Data Model can be represented just like any other Linked Data resources, *one can mesh VistA data directly with unlimited sources of internal or external, public or private life sciences, and other scientific or healthcare related data sources*. This leads to discovery of new relationships between different sources of data - and new knowledge.

Linked VistA: ***Native integration to terminology***

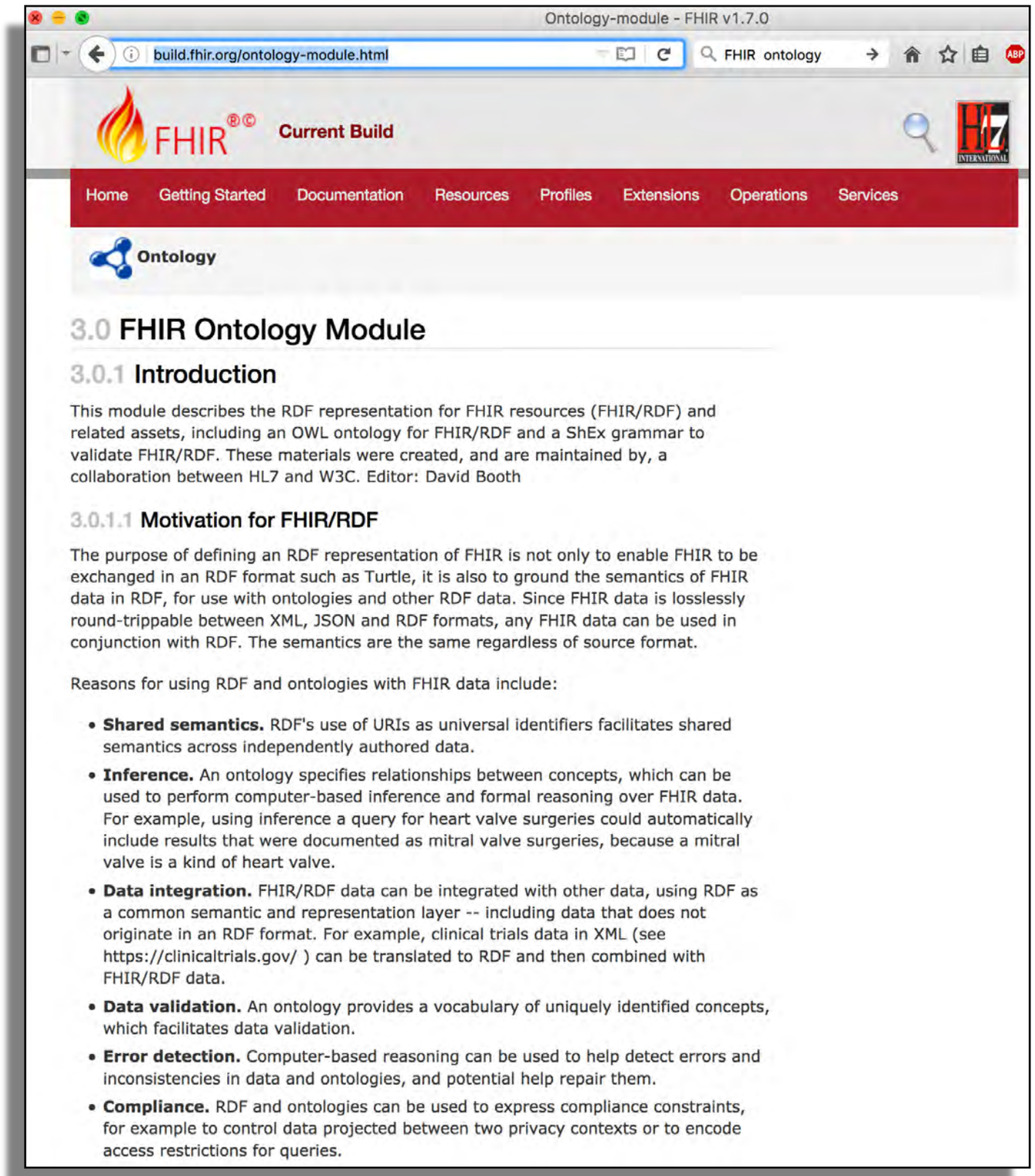
One can merge VistA data directly with any terminology published in Linked Data form.

Linked Data (RDF) **+** **Linked Terms** (RDF) **=** **Data-Terms integrated**



All current major healthcare terminologies including SNOMED, ICD-10, ICD-11, LOINC, RxNORM, and over 350 other terminologies are available and cross-linked as RDF at Bioportal.org.

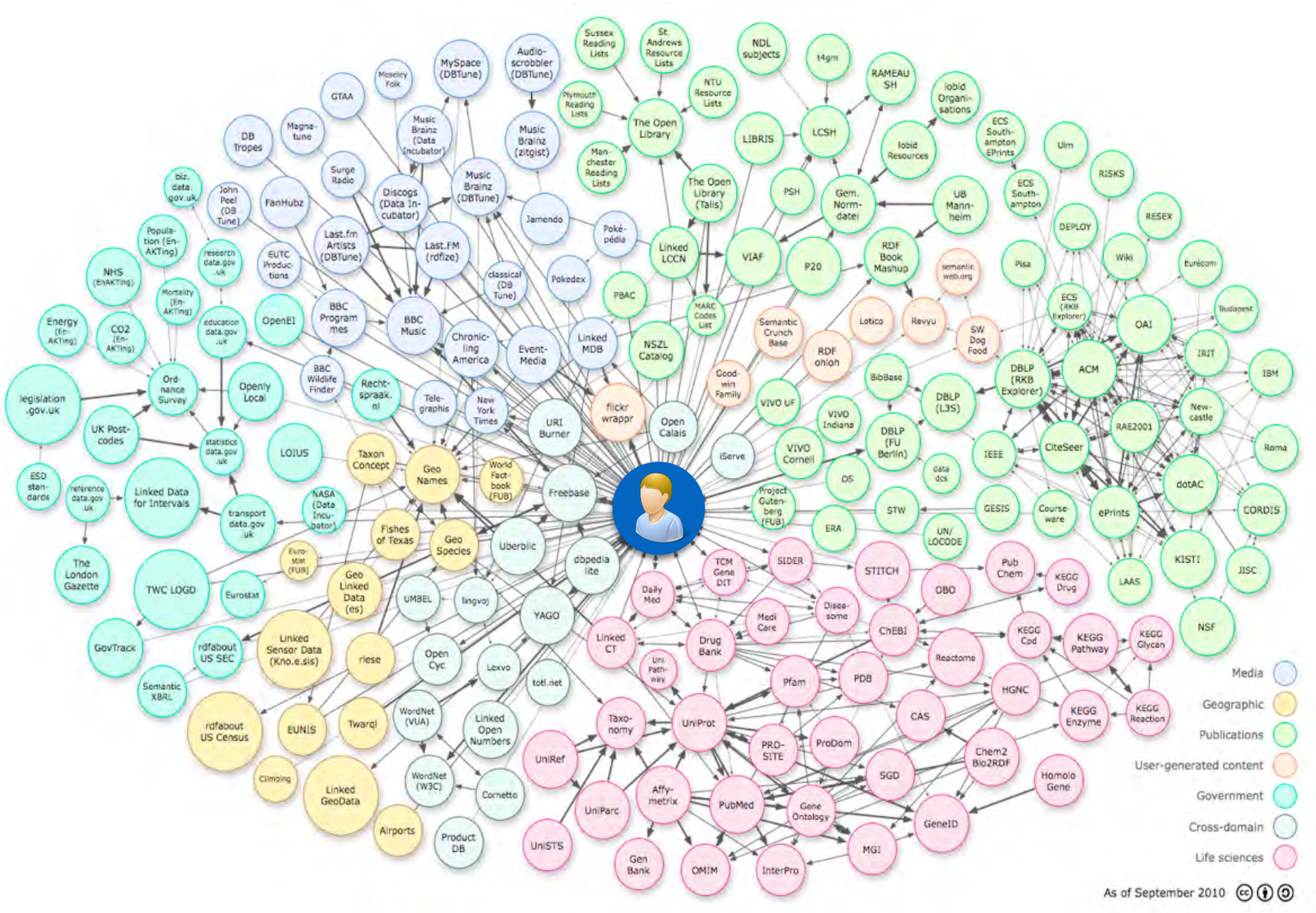
FHIR Ontology Module (RDF)



https://www.hl7.org/fhir/rdf.html#

Linked Data: ***Enables Precision, Personalized Health***

Representing patient data as Linked Data allows semantic interlinkage to thousands of Linked Data sources, enabling personalized, precision health care delivery.



Linked Patient

This enables *meshing, enrichment, and augmentation of patient data with any other Linked Data sources, providing an integrated view of all patient data from all locations, clinics, hospitals, and data sources - a hyper-personalized ecosystem of patient-specific data.*

Contact

Web: *vistadataproject.info*

Github: *github.com/vistadataproject*

Email: *rafael.richards@va.gov*