



**VA**  
HEALTH  
CARE

Defining  
**EXCELLENCE**  
in the 21st Century

# ***Linked VISTA***

## ***A Metadata-Centric Approach to VA Master Data Management***

***February 24, 2015  
Pacific Joint Information Technology Center  
Maui, HI***

***Rafael Richards MD MS***

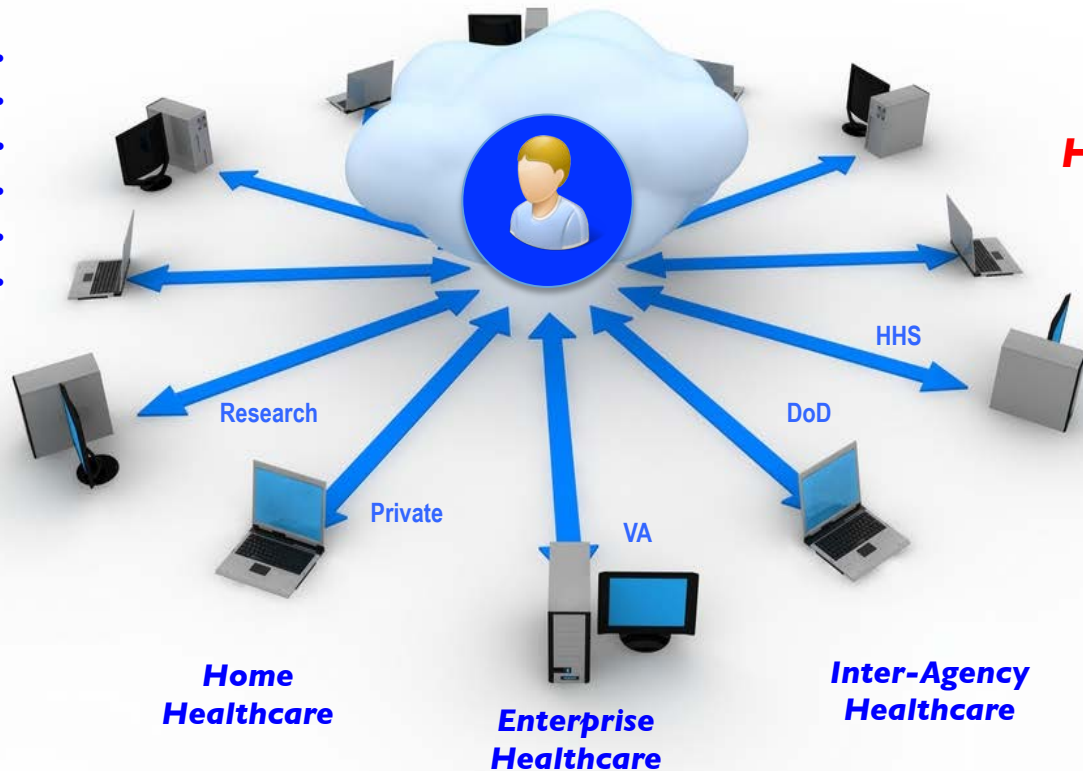
VHA Office of Informatics and Analytics  
U.S. Department of Veterans Affairs



# ***Vision: Linked VISTA***

## ***A Linked Data Approach to Patient-Centric Care***

***One Patient.  
All Data.  
Secure.  
Accessible.  
Integrated.  
Web-scale.***

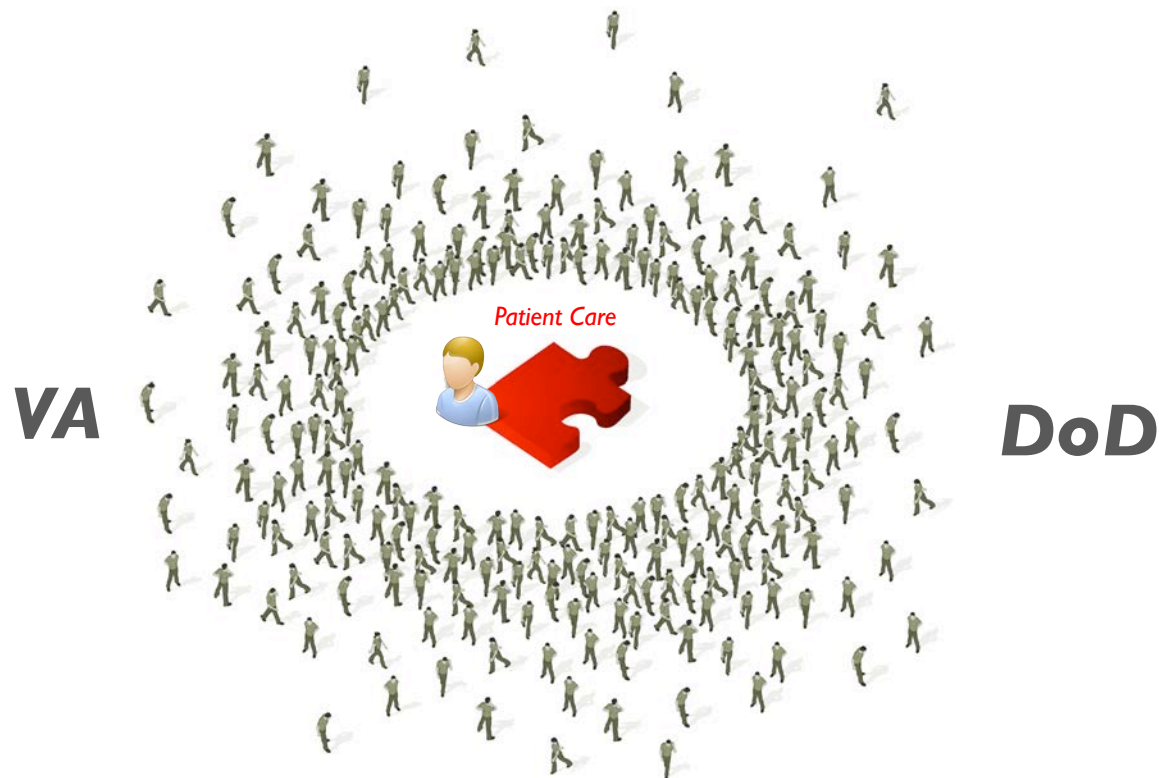


***How do we get here?***



# **Collaboration**

*Required to solve real world problems*



**Common goals**  
**Common vision**

**PJITC**



# Pacific Joint Information Technology Center (PJITC)



## Mission:

“To rapidly research, test, and develop medical solutions and products, through pilots or prototypes that provide mission critical value and actionable information to the Department of Defense, including the Services and the Department of Veterans Affairs.



*“...Supporting the MHS’ medical readiness requirements and IT modernization needs across the continuum of care through **rapid prototyping and advanced concept development**”.*

## About PJITC:

**Pacific JITC provides research and development activities with the goal of improving quality of services primarily within the federal healthcare sector.** The Pacific JITC offers a full suite of services that support systems research, prototyping, development, and testing. The goal of every project within the PJITC is to efficiently develop and transition value added services and solutions to the customer’s organization.

This is primarily accomplished by leveraging PJITC’s Integrated Test and Evaluation Center (ITEC), the first DoD/VA integrated lab where critical systems are virtualized. The ITEC provides an agile computing environment that supports military health and interagency research, development, testing, and evaluation missions.



# **PJITC: DoD-VA Collaboration Areas**

## **Common data management needs:**

- Future-proofing federal health data
- Enterprise view of patient data across all systems
- Enterprise shared data model across systems
- Inter-agency shared data models

## **Common data management technology:**

- Systems derived from DHCP
- DBMS remains identical across all systems\**

## **Intermediate-range goals:**

- DoD: data migration (DHMSM)
- VA: master data management (OneVA)

**There are over 230 DHCP-based systems deployed throughout VA and DoD as of 2015.**

While all have been locally modified at the application layer over the years, they remain **identical in design and function at the database management layer**. Therefore, the largest lever arm that exists to collaboratively manage, modernize, standardize, or migrate the data between VA and DoD systems is through this common, core database management layer. This will be discussed in the context of the VA VISTA systems in this review, but is applicable equally to all DoD CHCS systems.



# ***VISTA Overview***

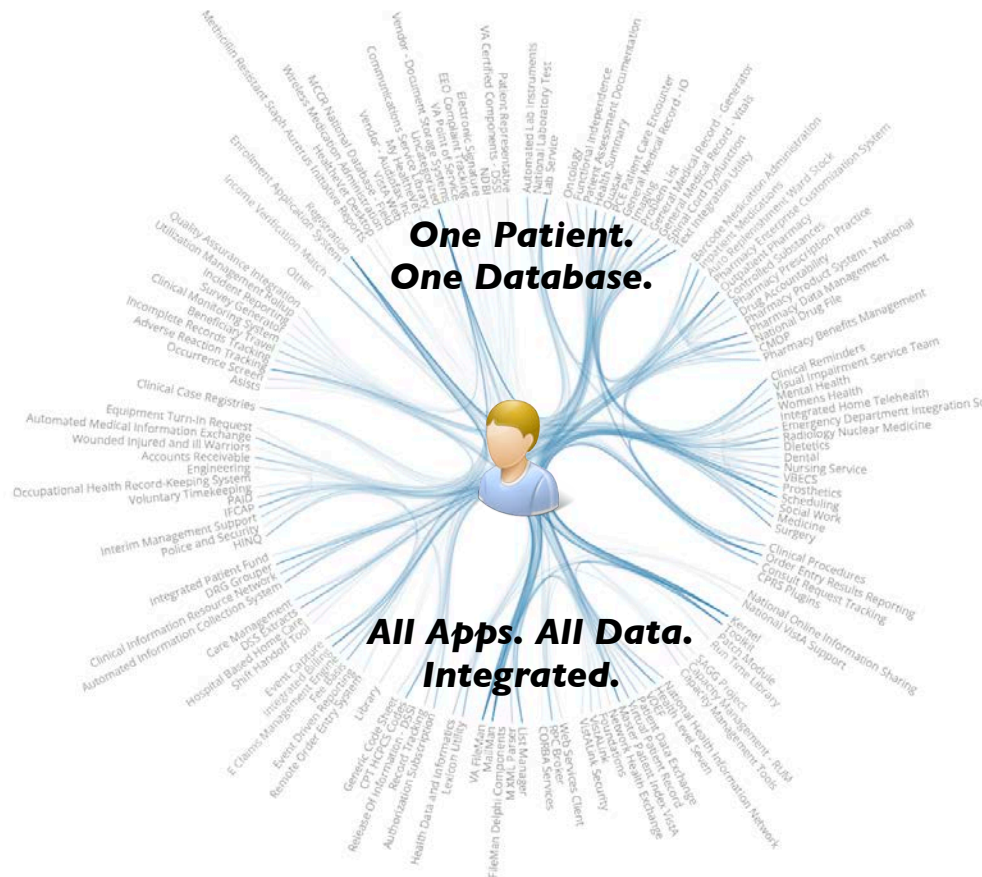
- *VISTA Architecture: Single-integrated EHR*
- *VA Enterprise: One Patient. Many VISTAs*
- *Problem: Isolated VISTAs. Fragmented data.*
- *Approach: Linked VISTAs.*
- *Vision: OneVA.*



## Patient-centric Health Record

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 shared data (inner circle) is visualized, showing the shared data flow between applications.







## ***VA Enterprise: Many VISTAs***

***1200 VA healthcare facilities throughout the US  
Supported by 130 VISTA systems***



***Care coordination for patients within  
VA requires integration of patient  
data across all 130 VISTA systems.***







# ***Problem Statement***

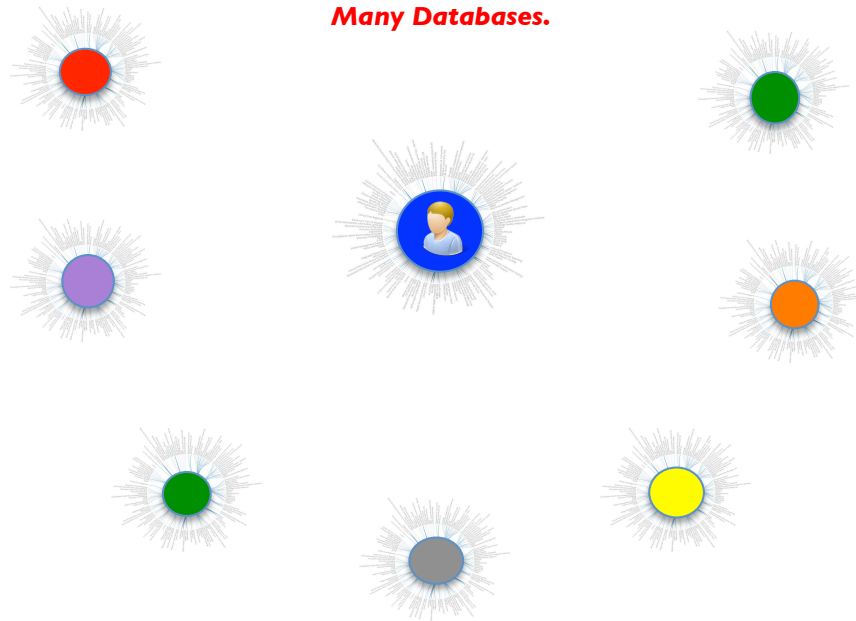
*VA is comprised of 130 unique VISTAs.  
This fragments patient data and care.*



# Problem: Isolated VISTA

## VISTA Silos

**One Patient.  
Many Databases.**



**Many VISTAs.  
Many Models.  
Fragmented Data.**

*While each VAVISTA 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.*



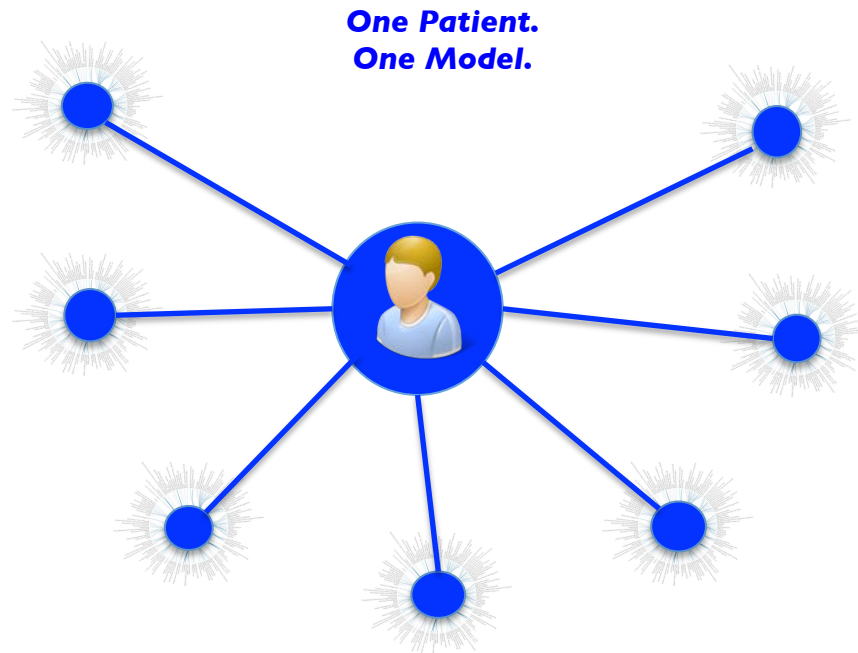
**Common Data Model.**  
**Data Security.**  
**Data Services.**

*This would allow publishing, linking, and exchanging, with patient-centric security, granular data with internal and external systems with international Web standards.*



# Approach: **Linked VISTA**

## **Linked VISTA**



**Many VISTAs.  
One Model.  
Integrated Data.**

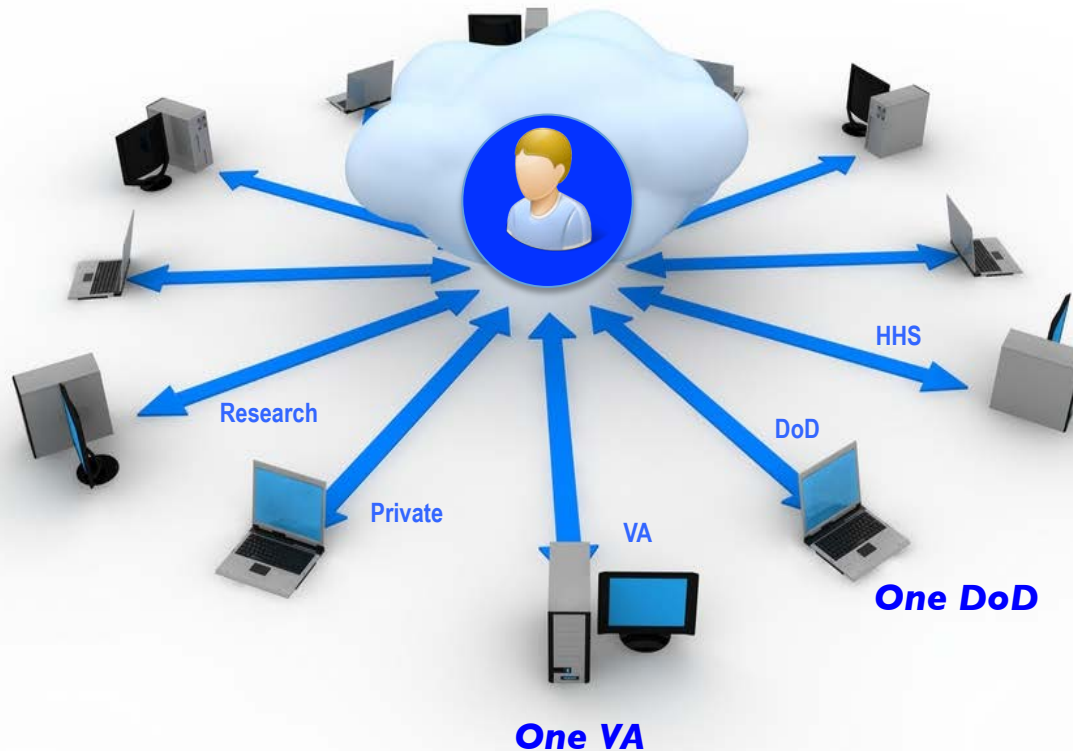
*Exposure and cross-linkage of VISTA's local models to a common VISTA model provides an integrated view of all patient data from all VISTA systems.*

# **OneVA**



## ***Vision: Linked VISTA***

***Web-standards approach to provide  
Patient-Centric Collaborative Care***



***One Patient.  
One Model.  
Integrated.  
Web-scale.***



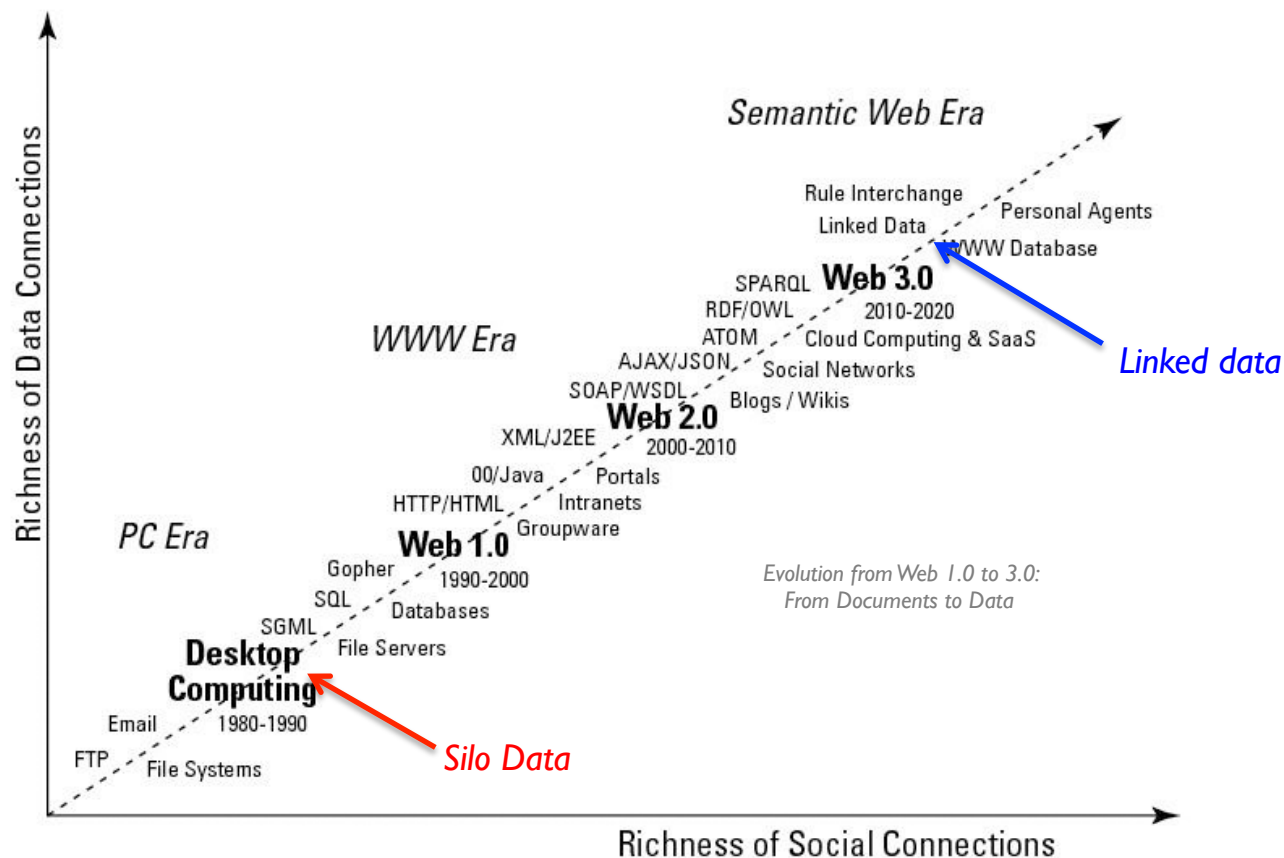
# ***Linked Data***

- *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 management: **Increased Connectivity**

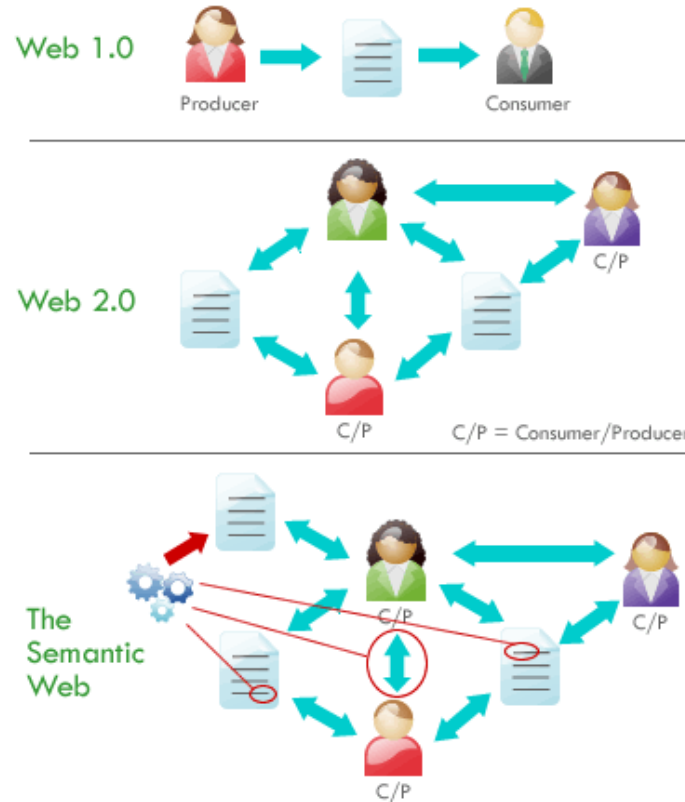


**DHCP was officially launched in 1982 and is the foundation of all VA VISTA and DoD CHCS systems.**

This was in the early PC Era, before the age of inter-networked data ("the internet"), when databases were isolated inside institutional systems (Silo Data: red arrow). Note that the TCP/IP internet protocol had not even been invented at that time. To bring any VISTA system into today's WWW Era, allowing data to be linked to all VISTA systems within VA enterprise (or to any CHCS system in DoD), VISTA needs to be web data-standards capable. This would transform VISTA to a WWW Database that could manage Linked Data (blue arrow).



# Evolution of the Web: From Documents to Data



## Web 1.0: Document Web (HTML)

### **Linked Documents**

Read-only web (**humans** only)

## Web 2.0: Social Web

### **Linked People**

Read-write web (**humans** only)

## Web 3.0: Semantic Web (RDF)

### **Linked Data**

Read-write web (**machine processable**)



## 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

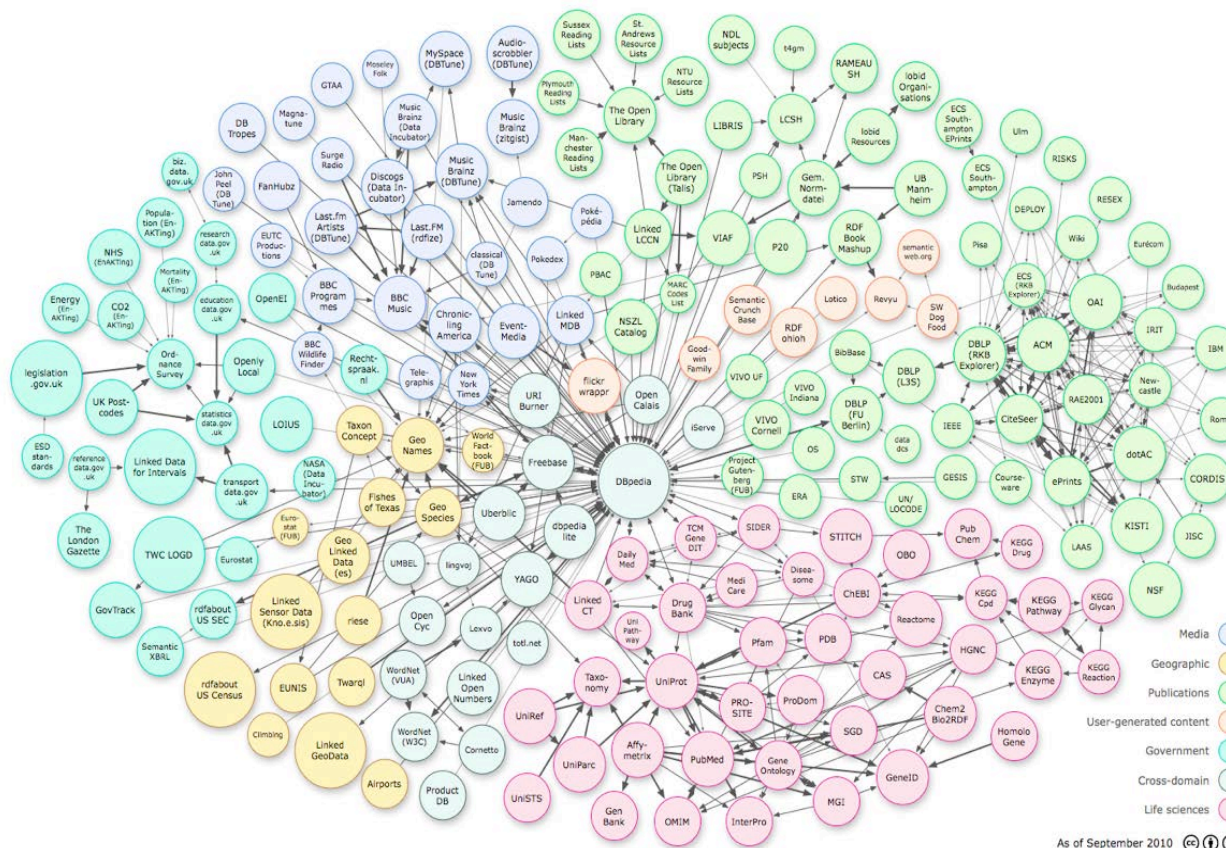


## Linked Data

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.



As of September 2010

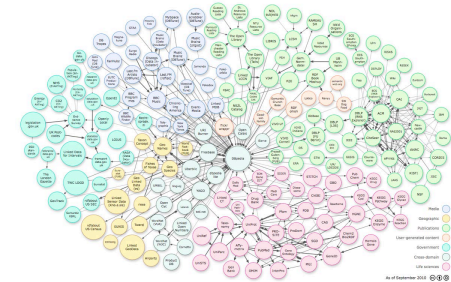


# Linked Data: Who Uses It?

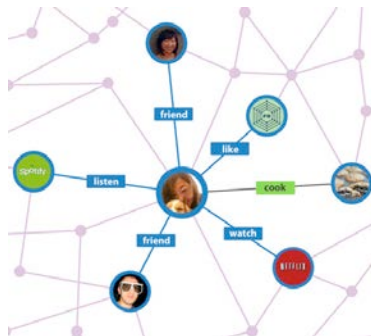
The Linked Data approach to **Internet-scale semantic data integration** is used by the worlds largest organizations such as Google, LinkedIn, Facebook, and IBM Watson.



**Wikipedia** (knowledge graph)



**IBM Watson** (knowledge graph)



**LinkedIn** (professional graph)



**Facebook** (social graph )

**Google Search** (knowledge graph)







## **Health Data: Thousands of diverse models**

**The need for Federal Health sector interoperability is not limited to the hundreds of systems within the VA and DoD.** Over forty percent of all veterans receive care in the Private sector. Therefore, a more general solution is required that can link to the many more systems in the private sector.

**The landscape of healthcare information systems is that of thousands of competing models.** Every EHR in the US has its own model. Each of the four thousand Meaningful Use Certified products has its own model. Each instance of each EHR, even from the same vendor or federal agency, has its own unique model.

**The problem is not just in the number and complexity of models, but that each of these models is rapidly, continuously, and independently changing** in response to the rapid growth of healthcare knowledge, drugs, devices, treatments, regulations, and guidelines.

**Healthcare cannot fit into a static or “one-size-fits-all” model.** There are simply too many models for this to be feasible, and doing so would stifle model innovation and evolution – which is required to keep pace with the changes in medicine.

**We need a universal exchange language that is model-flexible.** This would allow the thousands of models to peacefully co-exist, while still evolving in their respective knowledge domains, and allow them to be incrementally reconciled in a free, open “model marketplace”. The PCAST Report recommended such a “Universal Exchange Language for Healthcare”. **Linked Data fulfills these requirements.**





## **Linked Data: Accommodates diverse models**

A distinguishing feature of Linked Data that makes it an ideal healthcare exchange language is that it is **model-flexible**:

### **1. Allows multiple, diverse models to be used concurrently:**

- Semantically interlinked
- Complimentarily, non-exclusive
- Both standard models and many specialized models

### **2. Enables both model standardization and innovation**

- These are not mutually exclusive goals
- Allows simultaneous use of legacy (diverse) operational models while overlaying with a single (standardized) model for enterprise use and external exchange.

*Currently there are thousands of unique specialized models in healthcare. Within VA and DoD alone there are hundreds of different systems, each with their own data model. Each of these models needs to independently evolve (innovation), while simultaneously resolving to a common enterprise model (standardization). This may appear a mutually exclusive goal, unless there is flexibility in the model to accommodate both standardization and evolution. This requires an adaptive data model.*



# Health Data: 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”.

- ***The Presidential Council of Advisors recommends a metadata-driven approach to interoperability.***
- ***Linked Data is the World Wide Web Consortium standard for metadata.***



# ***VISTA Data Model (VDM)***

*A Path to VISTA Data Management:*

- VISTA's Database
- VISTA's Data Model
- VISTA's Data Model Exposed
- Benefits of Leveraging VISTA's Data Model:
  - Master Metadata Management
  - Centralized Knowledge Management
  - Master Data Definition
  - Patient-Centric Security Model
  - Separating Business logic from Data
  - Query Access



**All Apps.  
All Data.**

**MDE**

**Integrated.  
Real-time.**

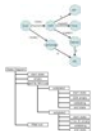
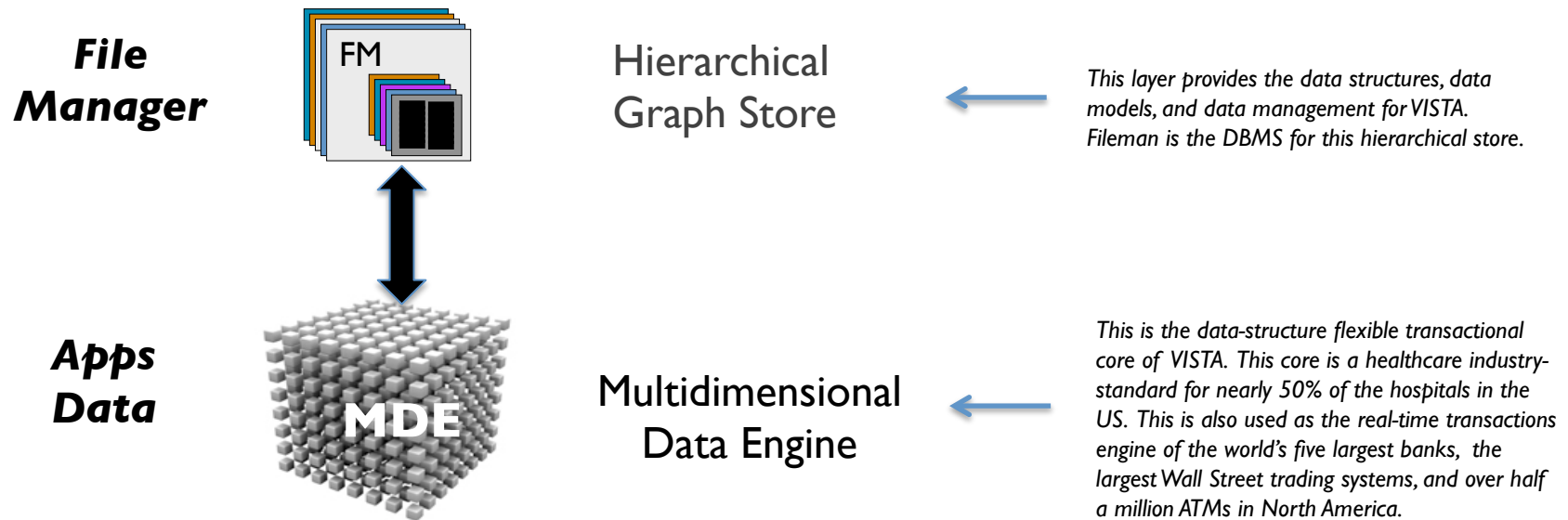
VA uses the same healthcare industry-standard data engine that nearly 50% of hospitals in the US currently use as their core EHR database.

*This same multidimensional data engine is also used as the real-time transaction engine of the world's five largest banks, the largest Wall Street trading systems, and over half a million ATMs in North America.*



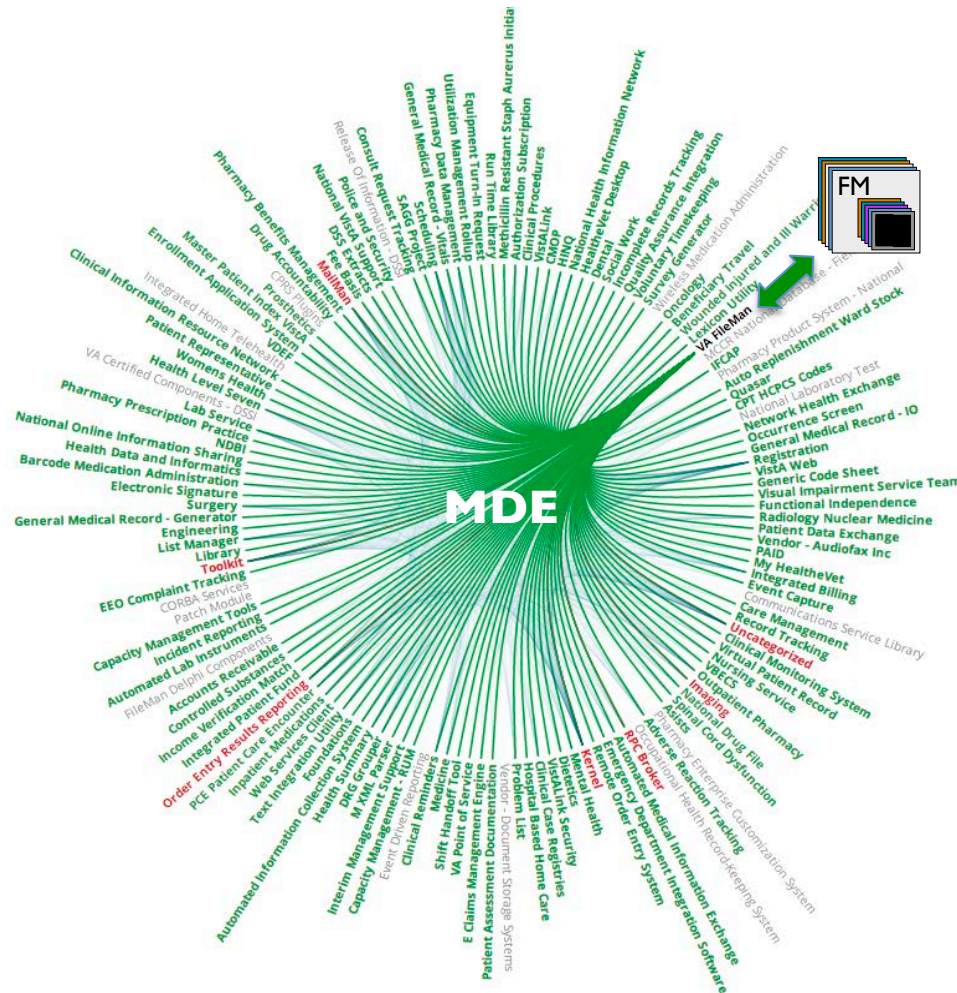
## Review: VISTA's Data Model (VDM)

All real-time transactional operations in VISTA take place within the the multidimensional data engine (MDE). To provide consistent structure and model for the data, a data dictionary driven hierarchical data storage system is overlaid on top of the MDE. All VISTA applications read and write data to this hierarchical store using a file manager called Fileman.



## Review: VISTA's Data Model (VDM)

All VISTA applications read and write data to VISTA's data engine through the NoSQL hierarchical data manager, Fileman (FM). Unlike many NoSQL databases, VISTA's database is not schema-less, but schema-driven. VISTA's schema is self-documenting through data dictionaries in Fileman. *Surprisingly, this fact has not yet been leveraged to expose and modernize the data model.*



**All VISTA data is organized through its data model, managed by Fileman.**

*All VistA applications use Fileman to access, structure, and query all data in VistA.*

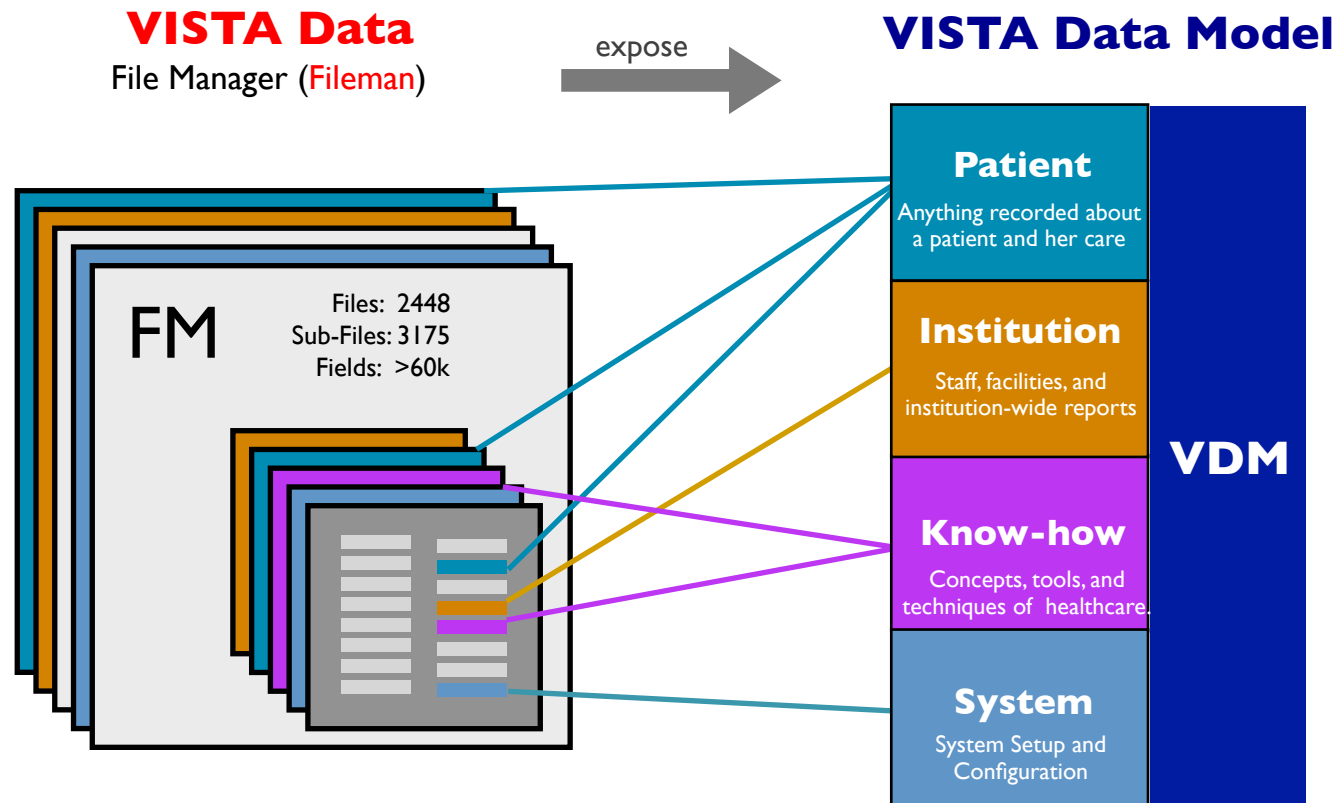
The green lines represent the data flow of all VISTA application to Fileman for all read and write operations to the Multidimensional Data Engine (MDE).





## VDM: *Expose and Leverage the Model*

**The first stage of data modernization is to expose and leverage VISTA's real, live operational data model.** Since this is just metadata, there is no patient data involved. Unlike many NoSQL databases which are schema-less, VISTA's NoSQL database model is self-documenting through Fileman. This allows us to render this in a standard definition format. In this new web-standard medium, data can be sorted, tagged, searched, and organized by data categories such as by patient, institution, know-how, or system information.



Access: **Thousands** of RPCs, API's, HL7  
Model: **Unknown**

Access: **Single** query access  
Model: **Consistent**, Transparent



# VDM: Master Data Management

**A benefit of a VISTA data model allows one to manage data logically across all VISTA application boundaries independent of the source of the data.** This lets one manage data with much flexibility, including logically partitioning and managing the data using metadata tagged categories (such as Patient, Institution, Know-how, and Systems information).

- **One may logically partition data by any class of data**, such as Patient data, Institution data, reusable Know-how, and System configuration data.
- **One may extract and move all patient data from system to system with one operation**, making system configuration migration and patient record movement far more efficient.
- **One may apply security metadata or protocols to any of these logical classes** of data. For example a patient-centric security model for patient data, and thus enforce patient-centric controls on information exchanged.

## VISTA Data Model



## Benefits:

### Patient Data Management

Extract and manage patient data with patient-specific security and metadata, allowing patient-centric controls on data access and exchange.

### Institutional Data Management:

Institution specific data can be exchangeable and centrally manageable

### Knowledge Management:

Common medical concepts, standards, and know-how may be identified and managed as a clearly defined class of VISTA data.

### System Management:

The entire configuration of a system can be identified, extracted, and transported, and inserted from system to system



# VDM: Patient-Centric Data Security

An exposed VISTA data model allows one to tag and partition certain classes of data separately from all other data in VistA. Specifically, this allows one to granularly partition any and all Patient data from all other kinds of data in VistA. This provides true, direct, “on the metal” security on patient data itself.

## VISTA Data Model

- The most important class of data to apply security is the Patient data category. This will allow very granular patient-centric security on the data itself.
- Current VISTA security is based on actions one can take using a legacy menu system - which has no relationship to the data.
- With a data-centric security model, we can specify not just what type of data (“Mental Health Record”...) but whose data (“For patient X”).
- This is much more specific and secure than the prior VistA security model.



### Patient-centric security model

Extract and manage patient data with patient-specific security and metadata, allowing patient-centric controls on data access and exchange.

This is compatible with security notions in Meaningful use Data Exchange: it can suppress even data that exists if there is no access permission.

It is essential to improve precision in data security to permit access to VISTA data securely. Otherwise one will have to reverse-engineer 3300 legacy RPCs and their one-off use of Kernel's menu options for each payload.

*A patient-centric security model is much more appropriate, flexible, and secure as a foundation for patient data security than the current VistA security model. The current VistA security model provides security only indirectly, through legacy controls of a menu system for a legacy roll and scroll terminal interface – which has nothing to do with the type of data at all (!).*



# VDM: Analytics Driving Interoperability

Comprehensive exposure and analytics of the VISTA Data Model will drive enhanced data use and interoperability as well as a major improvement in the structure of the database itself. To address these and other areas, focused reports could be generated from the model including:

Report	Activity		
Inconsistencies between VistA data models	Drive dictionary and code fixes in various centers so that every center is running the same consistent model	➡	<b>Enterprise Data Model</b>
Isolate centrally and locally managed “know-how”	Enables the next generation of enterprise knowledge services that seamlessly synchronize Vistas and other applications	➡	Centralized Knowledge Management
Under-definition in the model	Too many ill-defined string values and not enough nuance (“zip code”, “telephone number”) can be defined, and provided additional metadata (“home”, “work”, “mobile”) leading to a plan for incremental dictionary improvement	➡	<b>Enterprise Data Definition</b>
Key logic performed within FileMan	The barrier between the data store and business logic will be laid bare. This report will encourage the movement of certain types of logic into FileMan and out of less maintainable procedural code.	➡	Clean separation of business logic from data
Overlooked but highly valuable types of patient data	Improve VistA Data mining (for CDW etc.) and interoperability (more comprehensive electronic patient records).	➡	Clinical Research
(Some of this may have been hard-set by the application logic, and overlooked by Fileman or DD)	Without a complete model, how do you know what you’re leaving behind?	➡	<b>Interoperability</b>
VistA model/ FHIR comparison (key types)	Show how a direct from VistA transformation can remove the need for redundant intermediate, hard to maintain procedural code	➡	Accelerate Data Exchange
Isolate patient from other types of data	Enable patient-data access control rather than the crude option/API security now in VistA	➡	<b>Patient-centric Security Model</b>



# ***Linked VISTA***

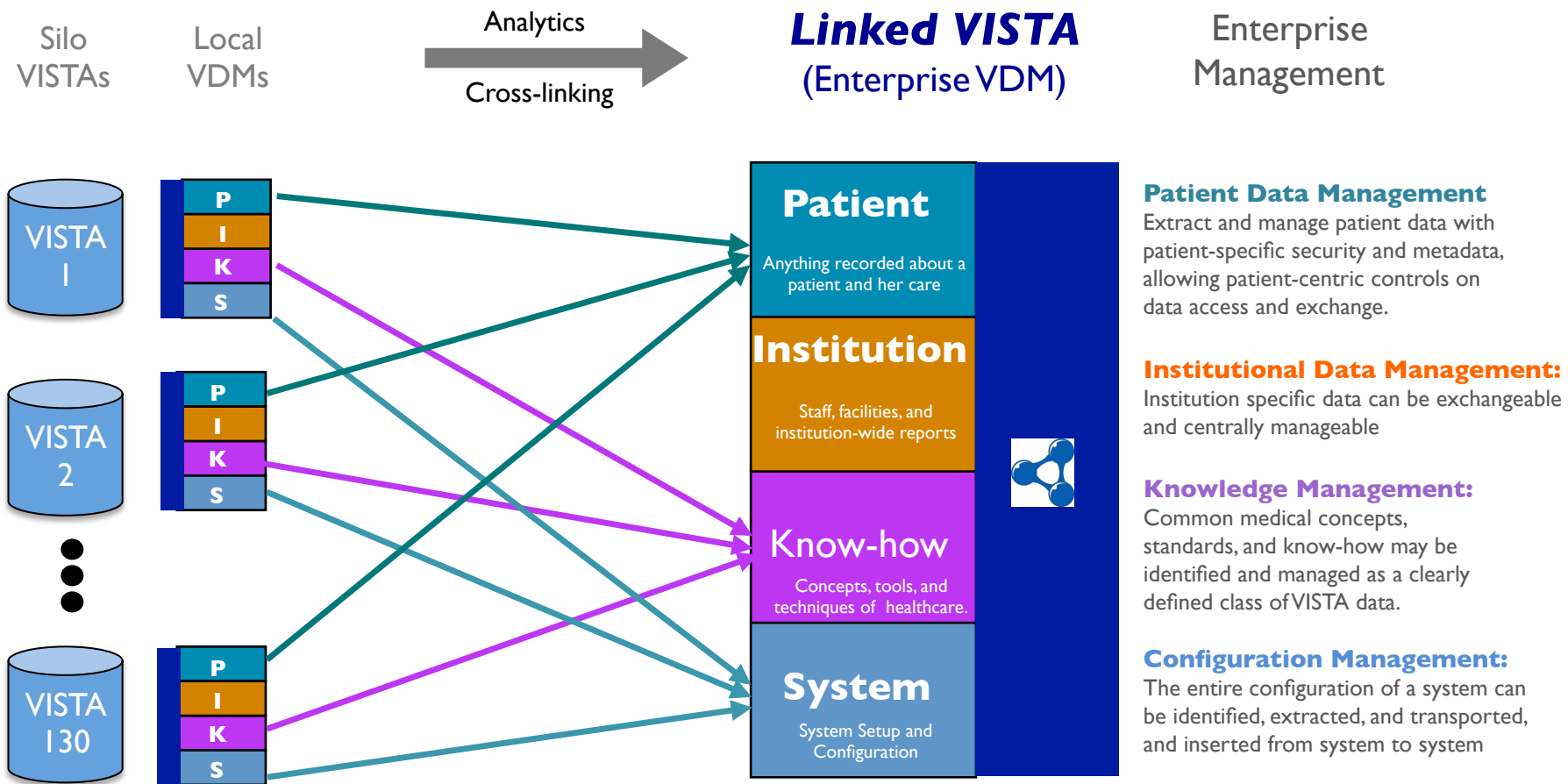
## ***Features***

- *Enterprise Cross-linked VISTA Model*
- *Comprehensive Data Access*
- *Web-standard representation language*
- *Web-scale semantic integration*
- *Knowledge Discovery*
- *Native terminology integration*



## Linked VISTA: The Enterprise Cross-Linked VISTA Data Model

Exposing VistA's Data Model (VDM) and applying analytics allows one to cross-link all local VDMs to an enterprise VISTA data model (Linked VDM), providing the capability for Enterprise query and Enterprise data management. This leverages the capability of Linked Data to create a cross-VISTA (enterprise) data model.





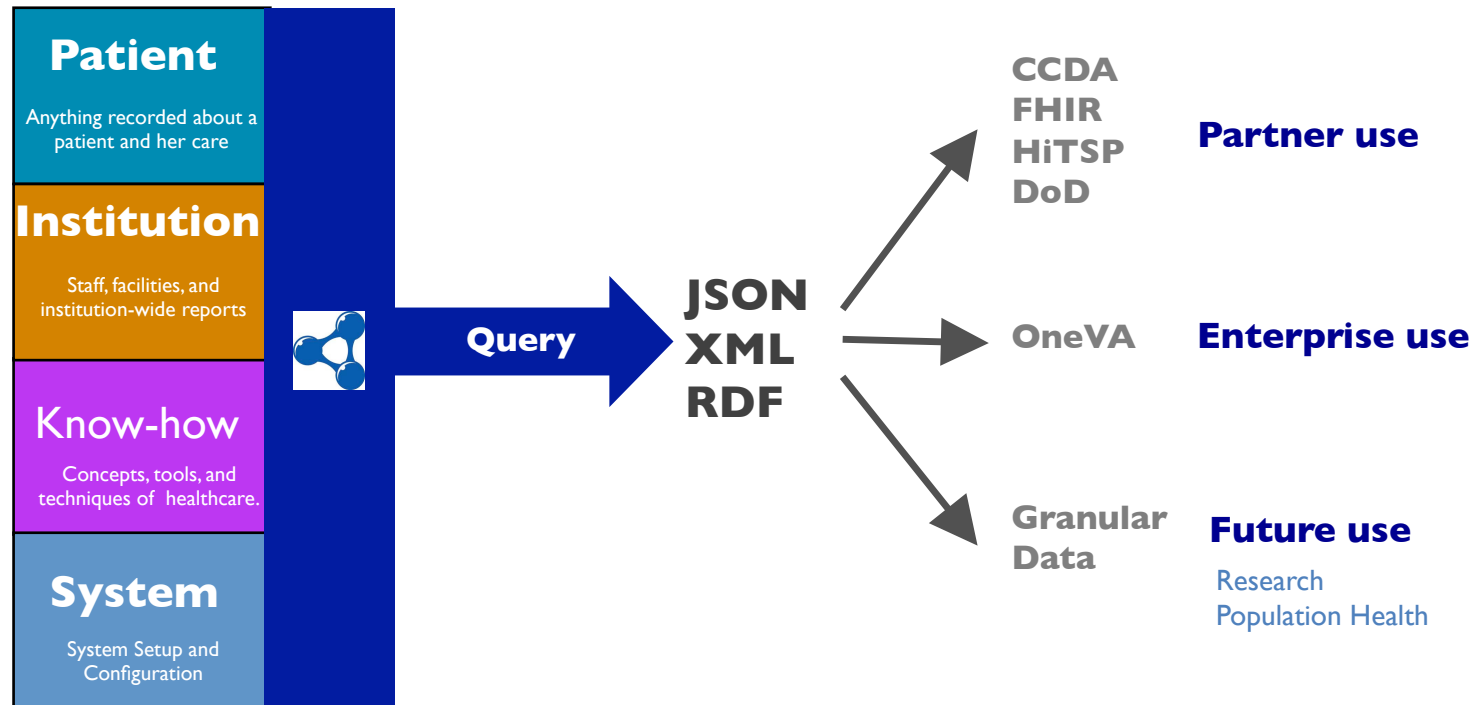


## Linked VISTA: Comprehensive Access

***One benefit of creating a cross-VISTA enterprise data model (Linked VISTA) is that it can be queried against any VISTA for any data with one web-standard query.***

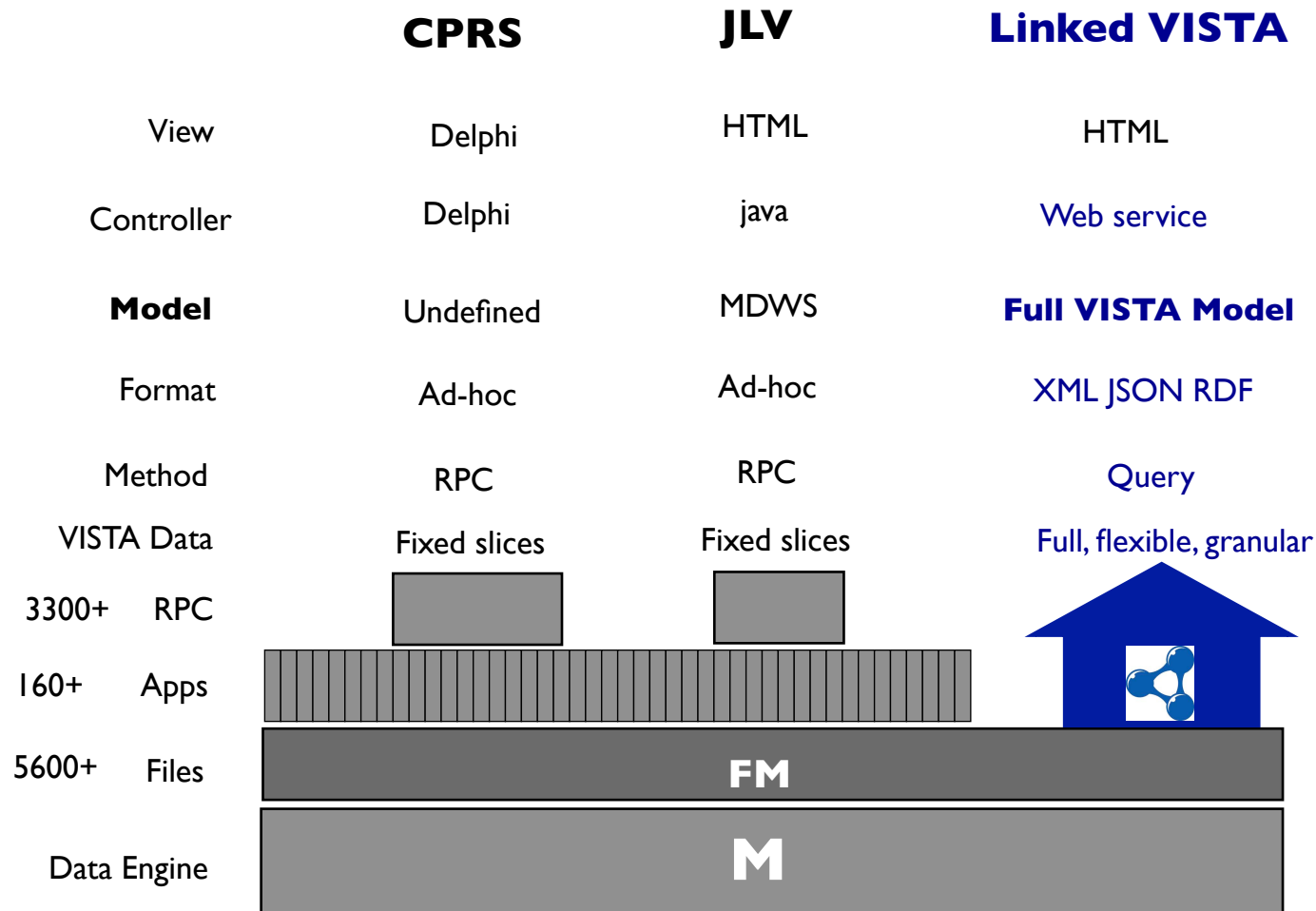
This would allow any authorized system to securely and directly query authoritative VISTA data in real-time with one standard query language. The output of these queries will be in all modern web-standard forms to maximize secondary use.

### Linked VISTA





## Linked VISTA: Comprehensive Access



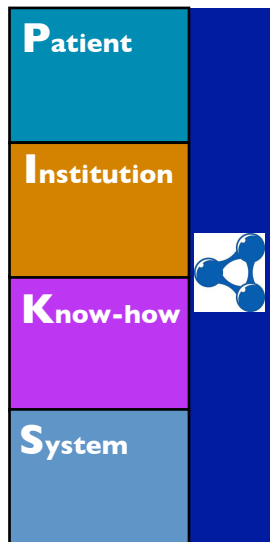
*VISTA Data Model (VDM) can access all data spanning 160+ applications with full granularity because the fully exposed VISTA Data model bridges all applications. No legacy APIs, HL7, RPCs, or MUMPS. Just VISTA data. All of it. Real-time.*



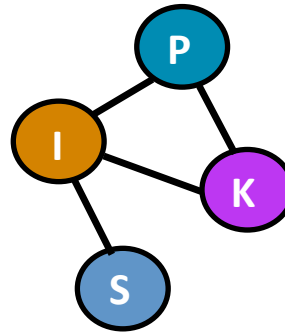
## Linked VISTA: Web-standard Representation

*Use of web-standard representation maximizes data re-use, meshing, and re-mixing with the maximum number of other sources of data for research and patient care.*

### Linked VistA

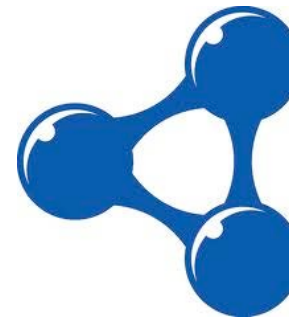


published



Same As

=



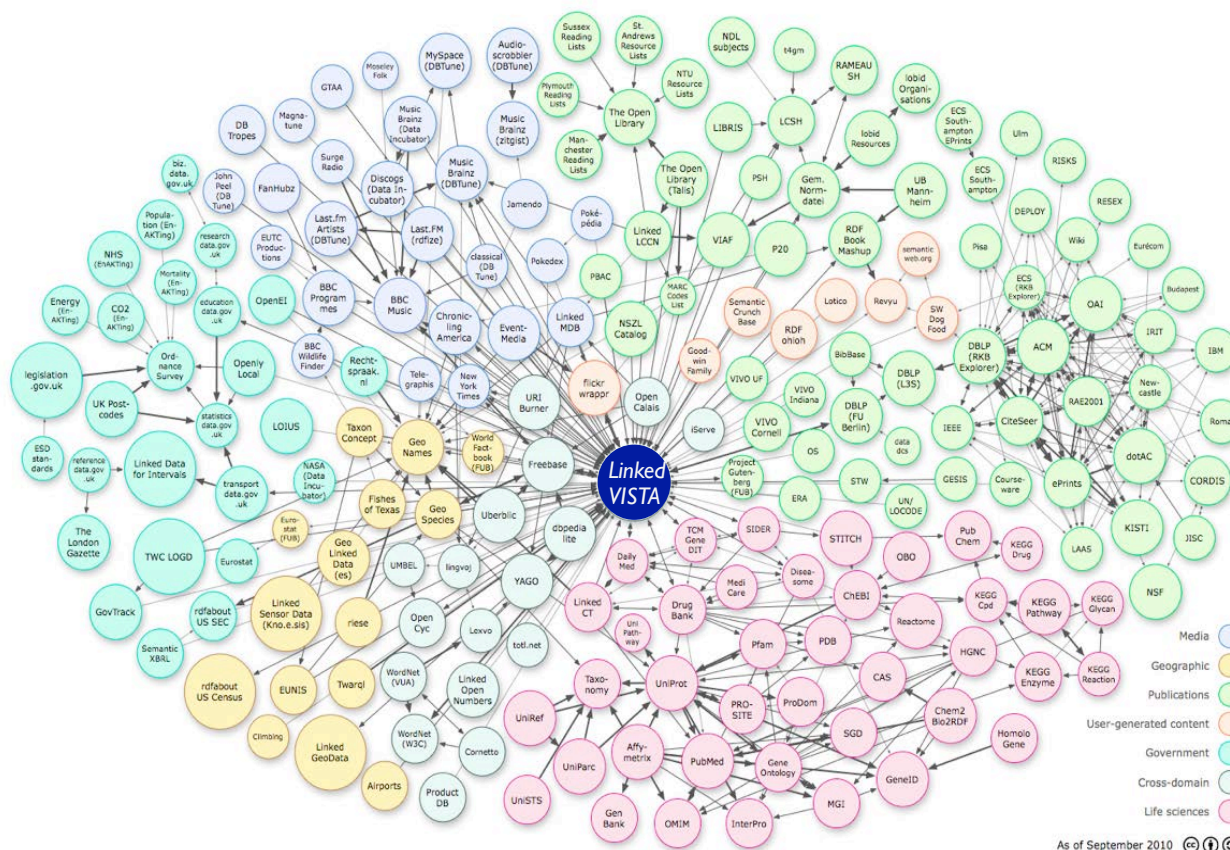
### Linked Data

Granular  
Data-atomic  
Meta-data tagged  
URI-based  
Model-based  
Linkable  
Web Standard



# Linked VISTA: Internet-scale semantic integration

**Representing VistA data in a Linked Data form supports real-time semantic integration with thousands of other linked sources.**



## Linked VistA

VistA Data - represented in standard Linked Data form - can be interlinked with any and all other Linked Data sources.

This will enable *meshing, enrichment, and augmentation* of patient data with all other sources, providing a comprehensive view of all patient data from all locations, clinics, hospitals, or the home.

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

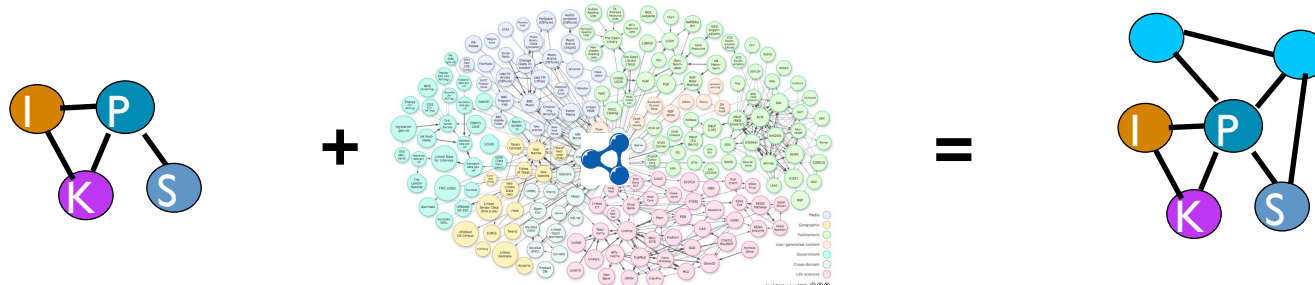
As of September 2010 © ⓘ ⓘ



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



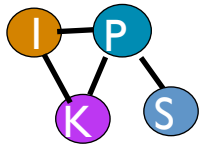
Because data in the VistA Data Model can be represented just like any other Linked Data resources, one can mesh VistA data directly with unlimited sources of internal or external, public or private life sciences, and other scientific or healthcare related data sources. This leads to discovery of new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



+

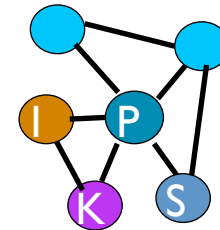
**Linked Terms**  
(RDF)

ICD-x  
SNOMED  
LOINC



=

**Data-Terms integrated**



*All current major healthcare terminologies including SNOMED, ICD-10, ICD-11, LOINC, RxNORM, and over 350 other terminologies (ontologies) at Bioportal.org are represented as RDF.*



# **VA-DoD Use cases**

## ***Applications of Linked Data / RDF***

- *Mandate: Future-proof data access*
- *DoD: Already operates RDF database*
- *VA: Already approves RDF as model*
- *VA-FHIR*
- *VA-DOD (current)*
- *VA-DOD (future)*
- *Linked VA-DoD: RDF databases are capable*
- *An Achievable Vision*
- *Recommendations*
- *Links!*





## ***Mandate: Future-proof Data Access is Required***

***Both VA and DoD currently have patient data that is over 35 years old in their core operational health IT systems.*** This legacy data must be not only universally available today, but available ***for the lifetime of all patients*** in a form that is computable within each agency, and between both agencies.

***Both agencies need to future-proof current and legacy data to meet these Congressional mandates.*** This requires adopting a shared, modern, *model-neutral, internationally recognized, internet-scale standard* for data representation and interoperability.

***Linked Data provides such a PCAST-recommended standard.***



# ***Mandate: Future-proof Data Access is Required***

## ***An Inter-Agency Linked Data strategy would provide:***

- ***Defense Health Agency (DHA) a transition strategy for CHCS data*** from all DHA systems to a common model, allowing uniform, standardized data migration in support of **DHMSM**.
- ***Veterans Health Agency (VHA) an enterprise data model for VistA data*** from all VA systems, providing a single view of patient data in support of **OneVA**.
- ***VHA and DHA a future-looking interoperability strategy*** which will support the **continuous change of OneVA and DHMSM systems** as all these evolve and change over time.

*Note: DoD has taken the lead in transforming CHCS data to Linked Data at the PJITC lab in the TAPS I project.*



# VHA: RDF is an approved technical reference model

UNITED STATES  
DEPARTMENT OF VETERANS AFFAIRS

Search All VA Web Pages  
  
Search

» [Open Advanced Search](#)

HomeVeteran ServicesBusinessAbout VAMedia RoomLocationsContact 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)**

GeneralDecisionReferenceComponentCategoryAnalysis

### 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](#)





# DHA: Operates a Linked Data-capable database

The screenshot shows the Oracle Technology Network website. The top navigation bar includes the Oracle logo, a search bar, and links for Sign In/Register, Help, Country, Communities, I am a..., and I want to... Below this is a secondary navigation bar with links for Products, Solutions, Downloads, Store, Support, Training, Partners, and About, along with an OTN button. The breadcrumb trail reads: Oracle Technology Network > Database > Options > Oracle Spatial and Graph > Overview.

The main content area is titled "Overview" and features a large heading "Oracle Spatial and Graph RDF Semantic Graph" with a red arrow pointing to it. Below the heading, a paragraph states: "RDF Semantic Graph is a W3C standards-based, full-featured graph store in Oracle Database for Linked Data and Social Networks applications."

On the left side, there is a sidebar menu with the following items: Database 12c, Database In-Memory, Multitenant, Options, Application Development, Big Data Appliance, Data Warehousing & Big Data, Database Appliance, Database Cloud, Exadata Database Machine, High Availability, Manageability, Migrations, Security, Unstructured Data, Upgrades, Windows, and Database Technology Index.

On the right side, there are several promotional banners. The top one is for "COLLABORATE15" with a date of "APRIL 12-16, 2015" and a location of "MANDALAY BAY RESORT & CASINO". Below this is a "REGISTRATION NOW OPEN" banner. Further down is a "WEBCAST" banner for "The Future of the Database—Product Introduction" with a "WATCH NOW" button. At the bottom right is a banner for "Get the Latest Oracle Database 12c Tutorials" with a "Plug into the Cloud" button.

At the bottom of the main content area, there are two smaller images. The left one is titled "RDF Semantic and Network Data Model Graph Capabilities in Oracle Spatial and Graph" and the right one is titled "Oracle Spatial and Graph: Benchmarking a Trillion Edges RDF Graph". Both images feature a large white arrow pointing right.

<http://www.oracle.com/technetwork/database/options/spatialandgraph/overview/rdfsemantic-graph-1902016.html>



# Linked VHA-DHA: RDF Databases are capable



## Oracle Spatial and Graph: Benchmarking a Trillion Edges RDF Graph

ORACLE WHITE PAPER | OCTOBER 2014

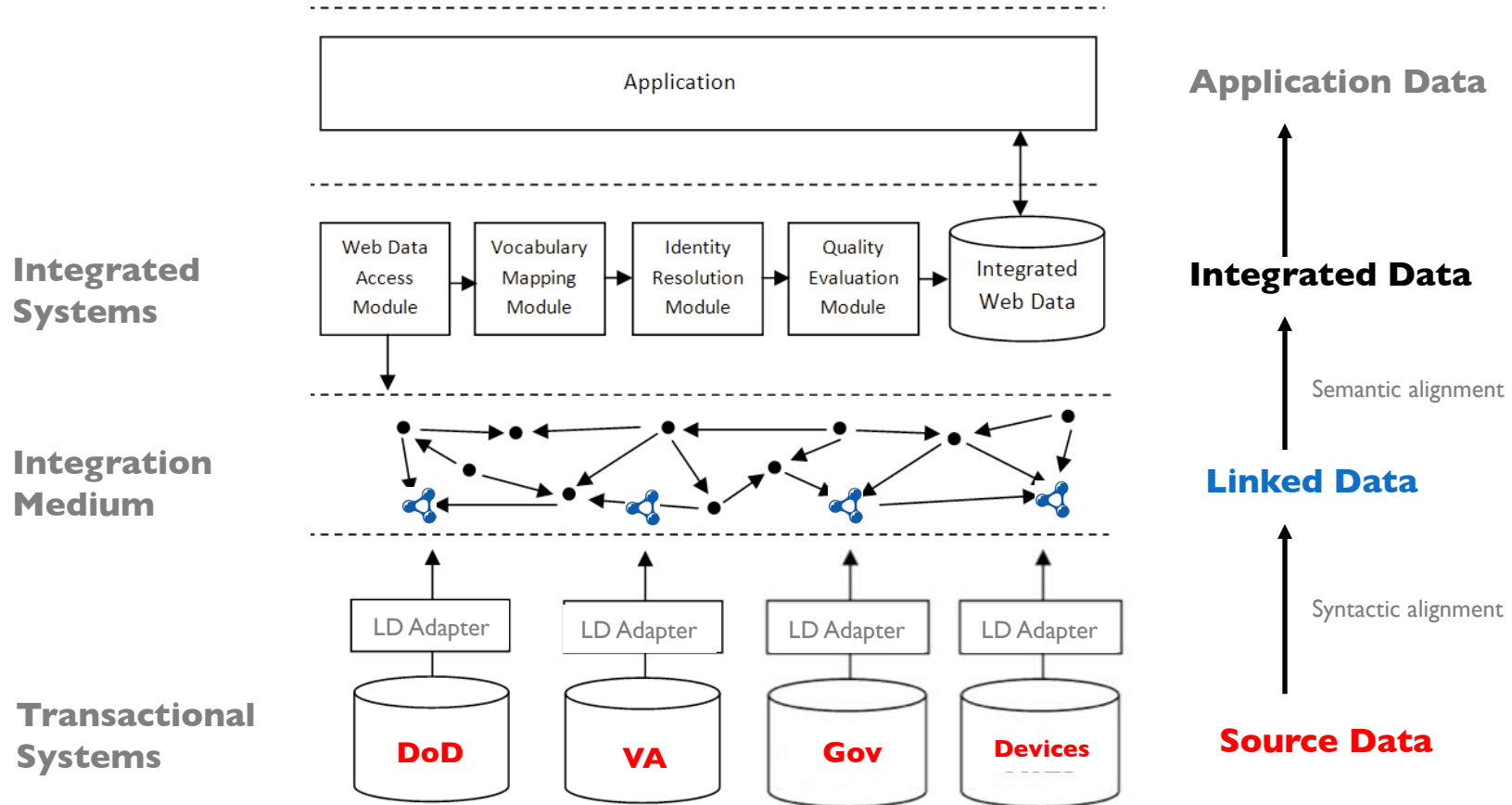
One trillion is a really big number. What could you store with one trillion facts?

- » 1000 tweets for every one of the 1 Billion Twitter users.
- » 770 facts about every one of the 1.3 Billion Facebook users.
- » 10 facts from 107 Billion sensors, located somewhere on the planet.
- » 400 metabolic readings for each of the 2.5 Billion heart beats over an average human life time.
- » 12 facts about every one of the 86 Billion neurons in the human brain.
- » 5 facts about every one of the 200 Billion stars in the Milky Way Galaxy.
- » 7 facts about every one of the 150 Billion galaxies in the universe.
- » 6,350 facts about each of the 158 Million books in the Library of Congress, the largest in the world.
- » 10 facts about each of the 107 Billion people who ever lived

*58,823 facts for each of the 17 million combined VA and DoD patients*  
*3134 facts for each of the 319 million citizens in the USA*

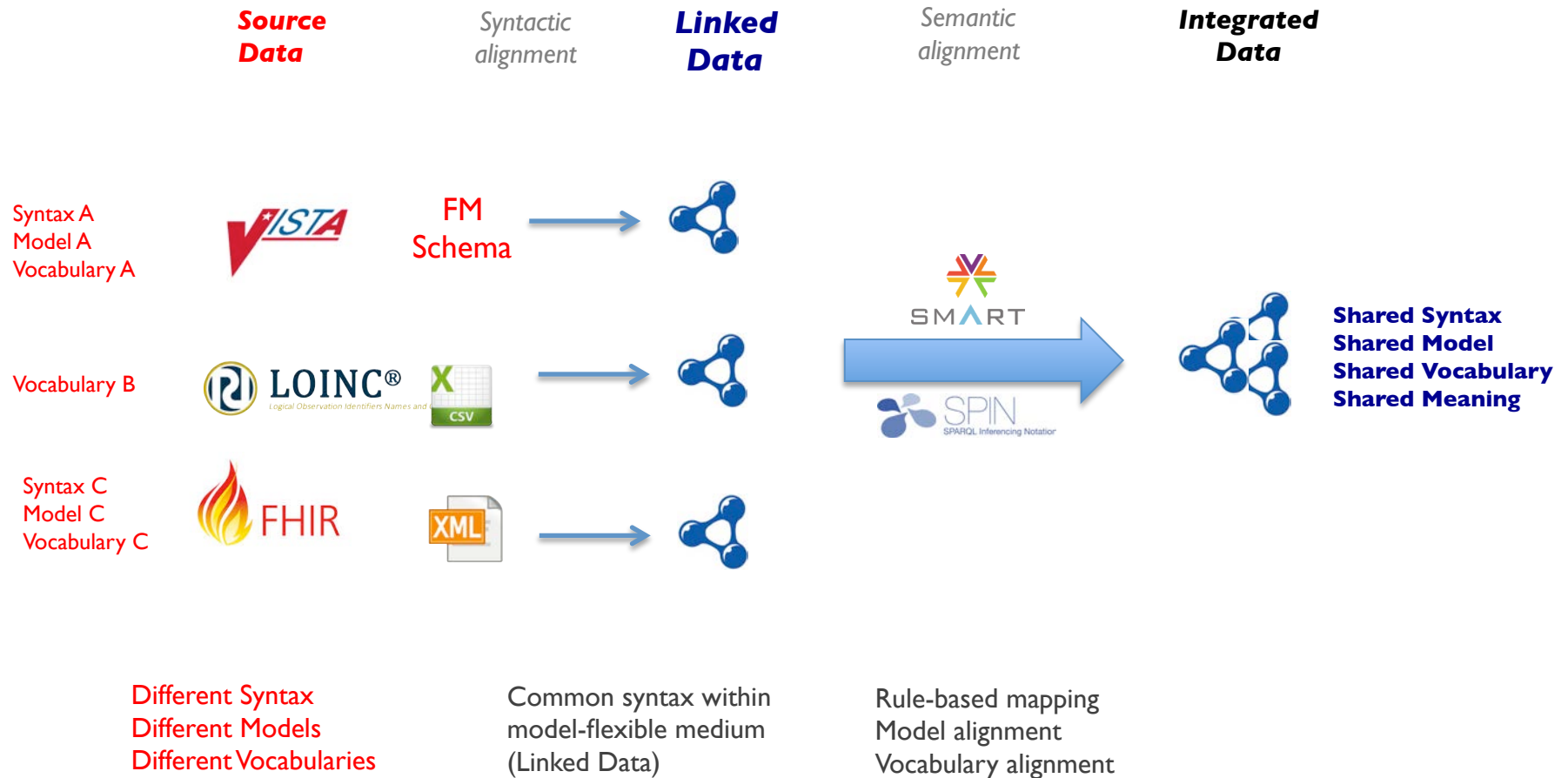


# Linked Data Integration: Semantic Data Integration





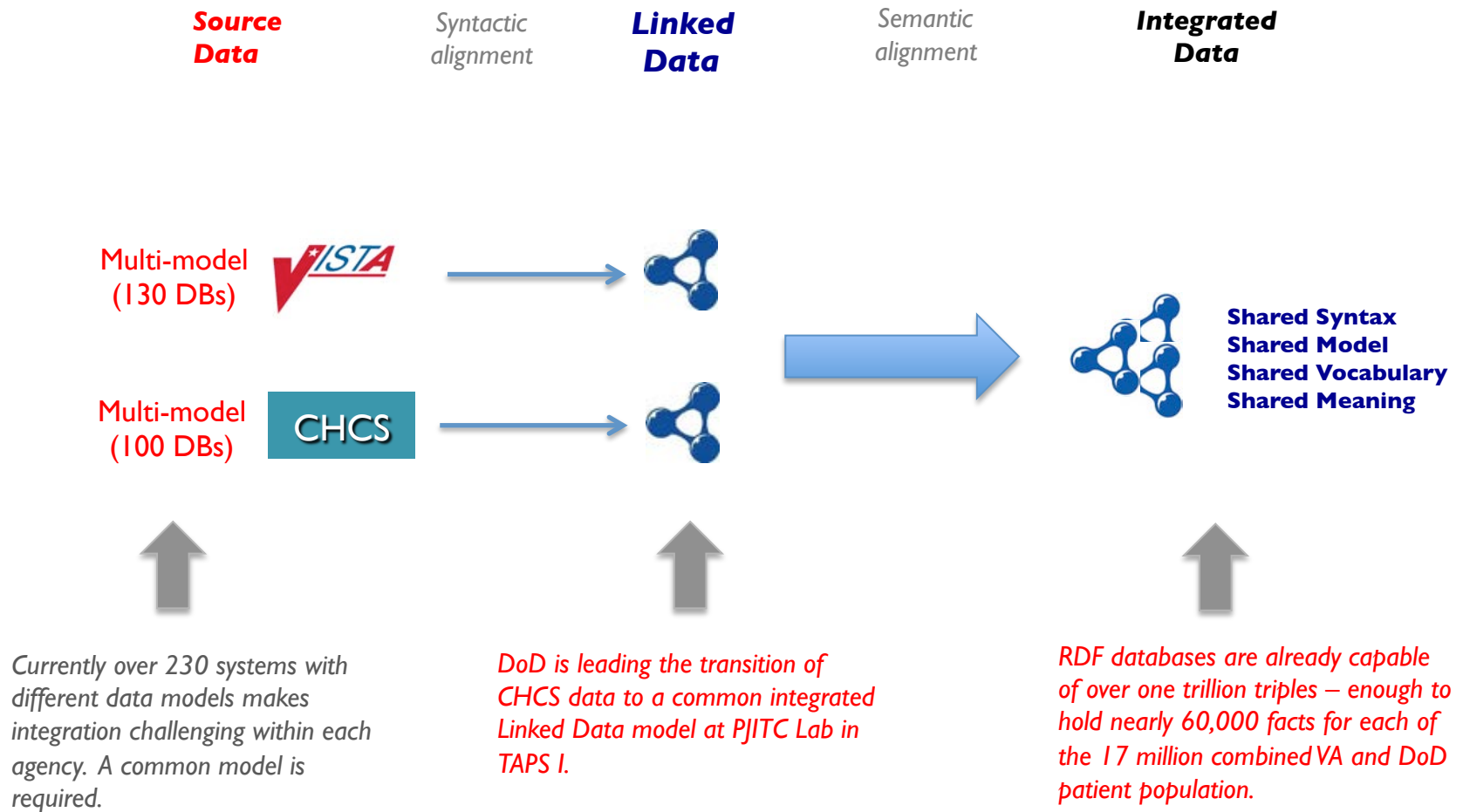
# Linked Data Integration: VA VISTA-FHIR





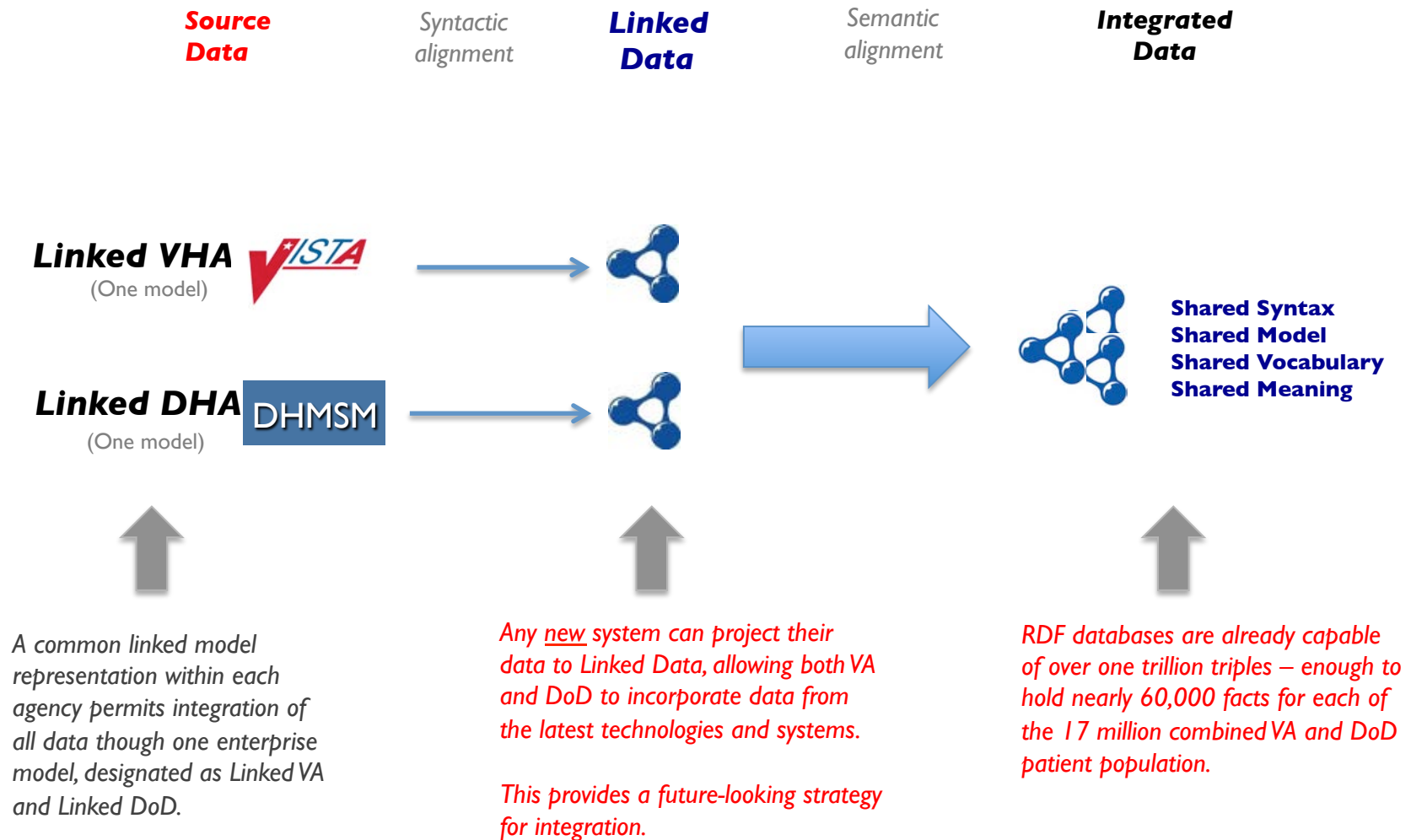


## Linked Data Integration: VHA-DHA (Current)





# Linked Data Integration: VHA-DHA (Future)

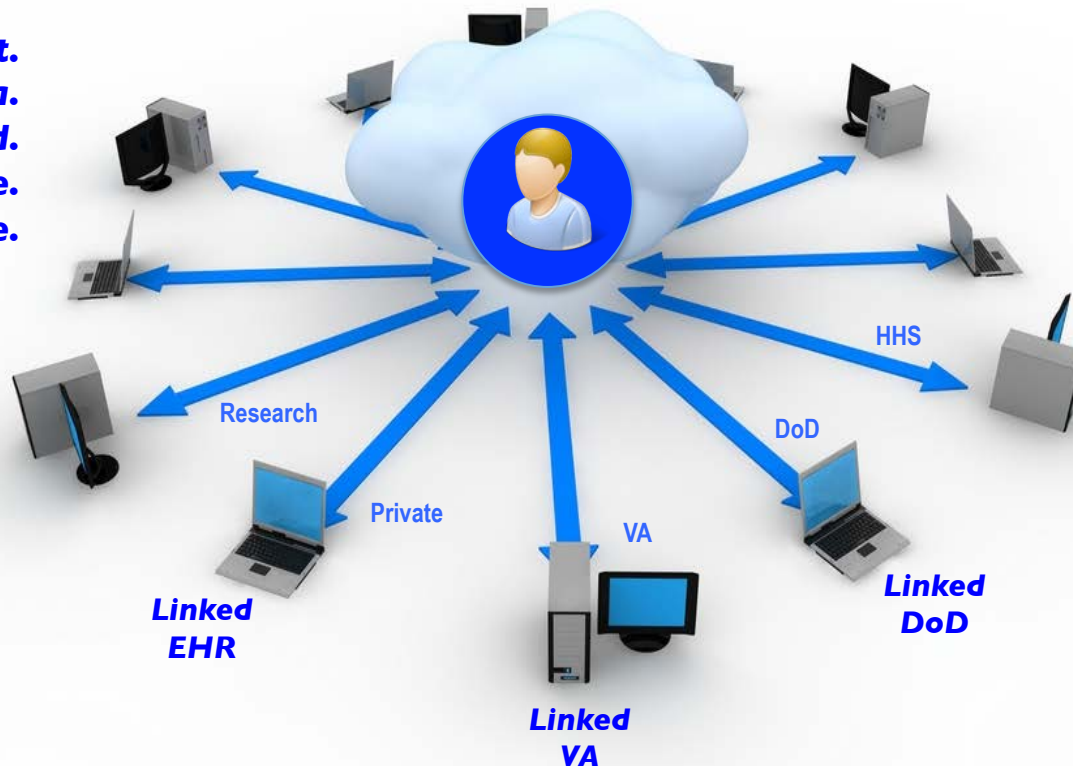




# **Linked VISTA: *An Achievable Vision***

## **Web-Standard approach to Patient-Centric Collaborative Care**

**One Patient.  
All Data.  
Integrated.  
Accessible.  
Web-scale.**



**Linked VA  
Linked DoD**



# ***Reference***

*Linked Data Summary*

*Linked Data Links*



# Linked Data: Accommodates both Standards and Innovation

## Information Models:

An apparent conflict between standardization and innovation

**Standards:** need to remain static in order not to be disruptive for adopters.

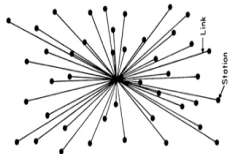
- Static, Brittle
- Centralized
- General (Common Denominator)
- Committee-driven
- Large, “all-or-nothing”, disruptive updates

**VS.**

**Innovation:** requires continuous evolution of thousands of new information models.

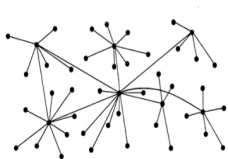
- Adaptive / Evolutionary
- Decentralized
- Highly Specialized, “Best of breed”
- End-user / specialist – driven
- Small, continuous, low-impact updates

## What are the options?



**Centralized, Model-rigid approach:** For exchange of information to occur all models must remain fixed, and data must go through only one central ‘broker’ model..

Our current approach to healthcare data

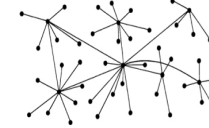
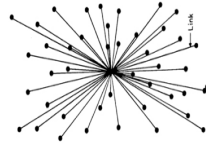


**Decentralized, Model-flexible approach:** Multiple models peacefully co-exist and evolve, mediated by their ability to freely link to any model at all times. In this approach, all models are free to evolve AND are capable of resolving to a common standard model at all times. *The only current technology that supports this granular data-oriented approach is [RDF \(Linked Data\)](http://www.carlsterner.com/research/2009_resilience_and_decentralization.shtml).*

**Linked Data supports both standardization and innovation**



# Data Integration: Legacy vs. Linked Approach



## Architectural Issues

## Current Approach

## Linked Data (RDF)

Data model characteristics	<b>Model-rigid.</b> Only one lowest common denominator model unifies information. Must remain unchanged to orchestrate. <b>Restrictive expression.</b>	<b>Model-flexible.</b> All data models may independently evolve. <b>Maximizes expressivity.</b>
Data model compatibility	<b>No model diversity permitted</b> A one-size-fits-all mega-model	<b>Multiple models peacefully coexist</b> Data model flexible
Data model evolution	<b>Costly and difficult to evolve models.</b> Due to model-rigid architecture.	<b>Cheaper and easier to evolve models.</b> Due to model-flexible architecture.
Data access method	Downloading + Aggregating	<b>Linking + Federating</b>
Scalability: incremental effort required to add new data sources	<b>Common model</b> must be updated	<b>Individual models</b> may be independently and incrementally semantically linked.
Primary Function	<b>Data Syntax and Transport</b>	<b>Data Model</b>
Granularity	<b>Document-centric</b> (message-centric)	<b>Data-centric</b> (individual data elements)
Semantics	<b>Weak semantics.</b> <b>Extrinsic, separate from the data.</b> Depends on an external data model.	<b>Strong semantics.</b> <b>Intrinsic, integrated with the data.</b>



## ***Links: Linked Data***

### ***W3C Linked Data Standard***

<http://www.w3.org/standards/semanticweb/data>

### ***W3C Linked Data Platform***

<http://www.w3.org/TR/ldp/>

### ***W3C Semantic Web Healthcare and Life sciences***

<http://www.w3.org/blog/hcls/>

### ***HL7-RDF Healthcare Standards Work Group***

[http://wiki.hl7.org/index.php?title=RDF\\_for\\_Semantic\\_Interoperability](http://wiki.hl7.org/index.php?title=RDF_for_Semantic_Interoperability)

### ***Semantic Web***

[http://en.wikipedia.org/wiki/Semantic\\_Web](http://en.wikipedia.org/wiki/Semantic_Web)

### ***Linked Data: Tools (>1000)***

<http://www.mkbergman.com/sweet-tools>

### ***Linked Data: Adaptors (100's)***

<http://www.w3.org/wiki/ConverterToRdf>

### ***Linked Data: Roadmap for Healthcare Interoperability***

<http://yosemiteproject.org>





# Linked VISTA: A Semantic EHR

Representation of VistA data as Linked Data creates a **Semantic EHR** – one with well-defined meanings derived from a clear data model in RDF. This supports inferencing, reasoning, knowledge discovery, and semantic interoperability. Contrast with today's **Syntactic EHRs** focused on data structures.

