The PORTAL-DOORS Project and NPDS Cyberinfrastructure

Carl Taswell

21 October 2021

Brain Health Alliance Virtual Institute

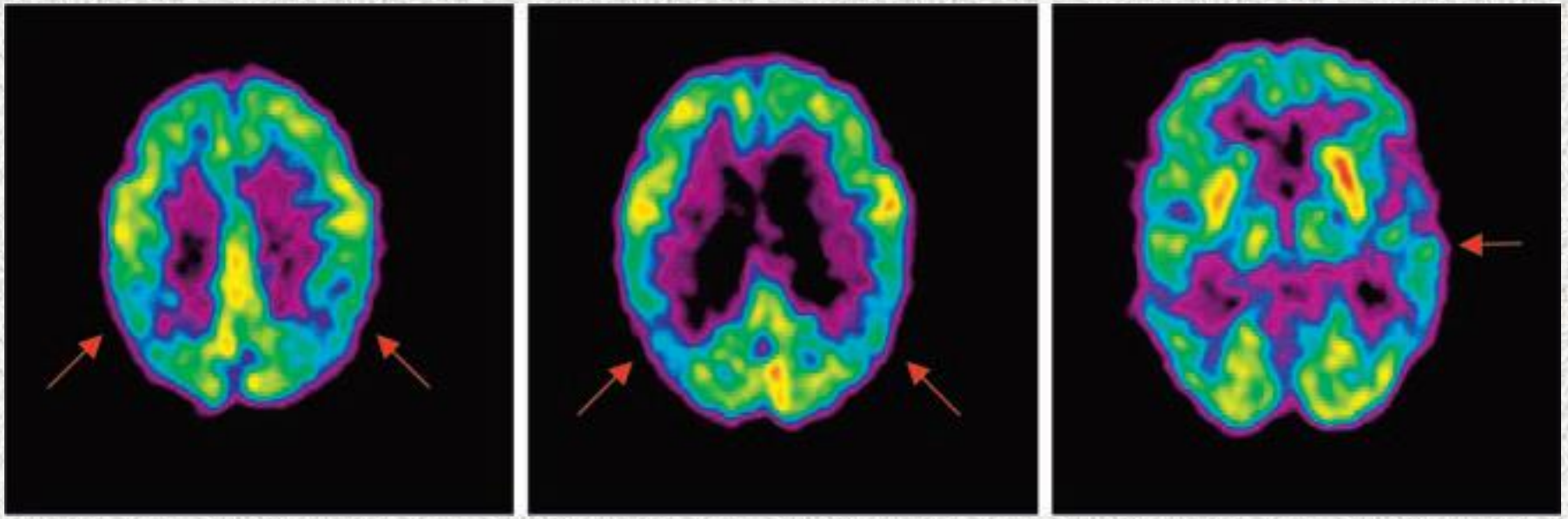
UC San Diego Dept Electrical & Computer Engineering

Brief Overview

- Motivating Problems
  - Medical: Brain imaging for dementias & pharmacogenomic molecular imaging
  - Engineering: Semantic search & data integration to overcome cybersilos & ‘needles in haystacks’
  - Informatics: Automated meta-analysis
- Design Solutions
  - History/review of PORTAL-DOORS Project
  - Distributed systems & architectural styles
  - PORTAL-DOORS vs IRIS-DNS designs
  - Semantic search & HDMM architectural style
- Implementation Goals
  - PDP-DREAM Software for web apps/services
  - Repositories for open source software

Brain Imaging for Dementias & PharmacoGenomic Molecular Imaging

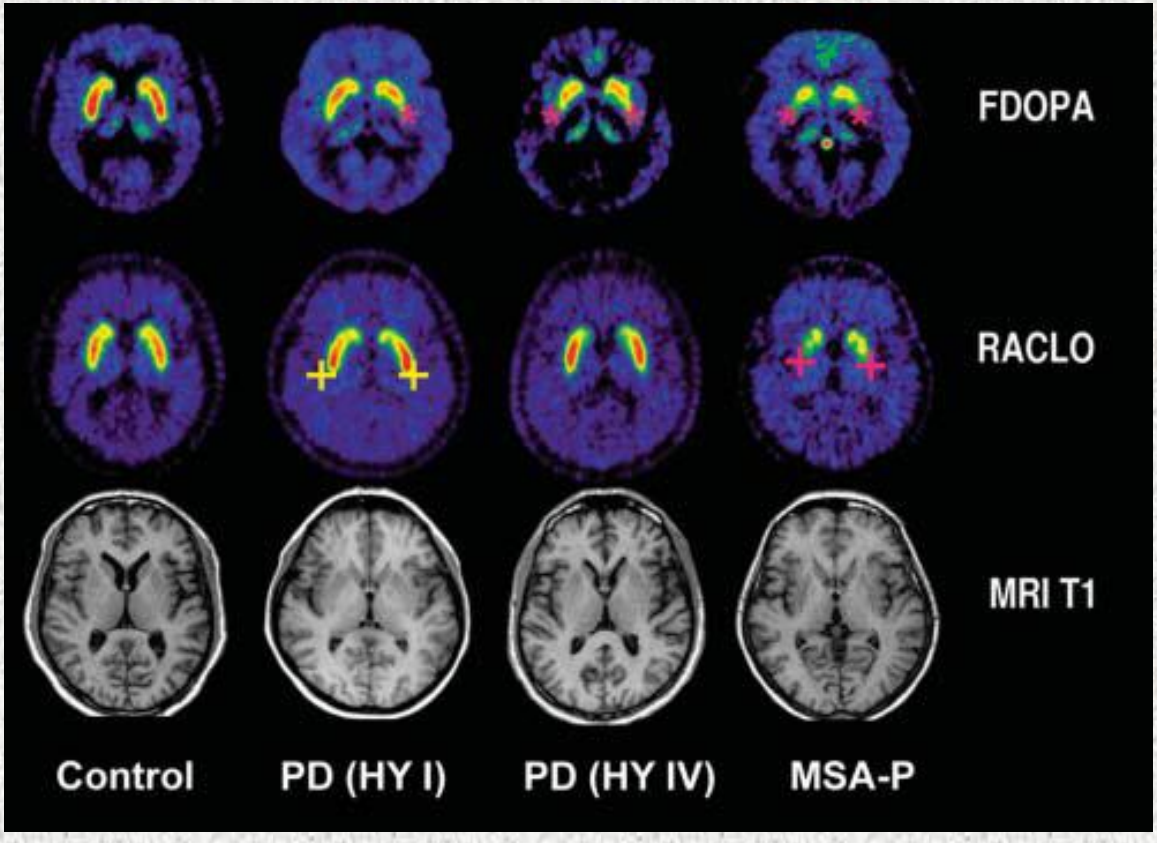
Brain Imaging for Alzheimer's Disease



18F-FDG PET images of early Alzheimer's disease.

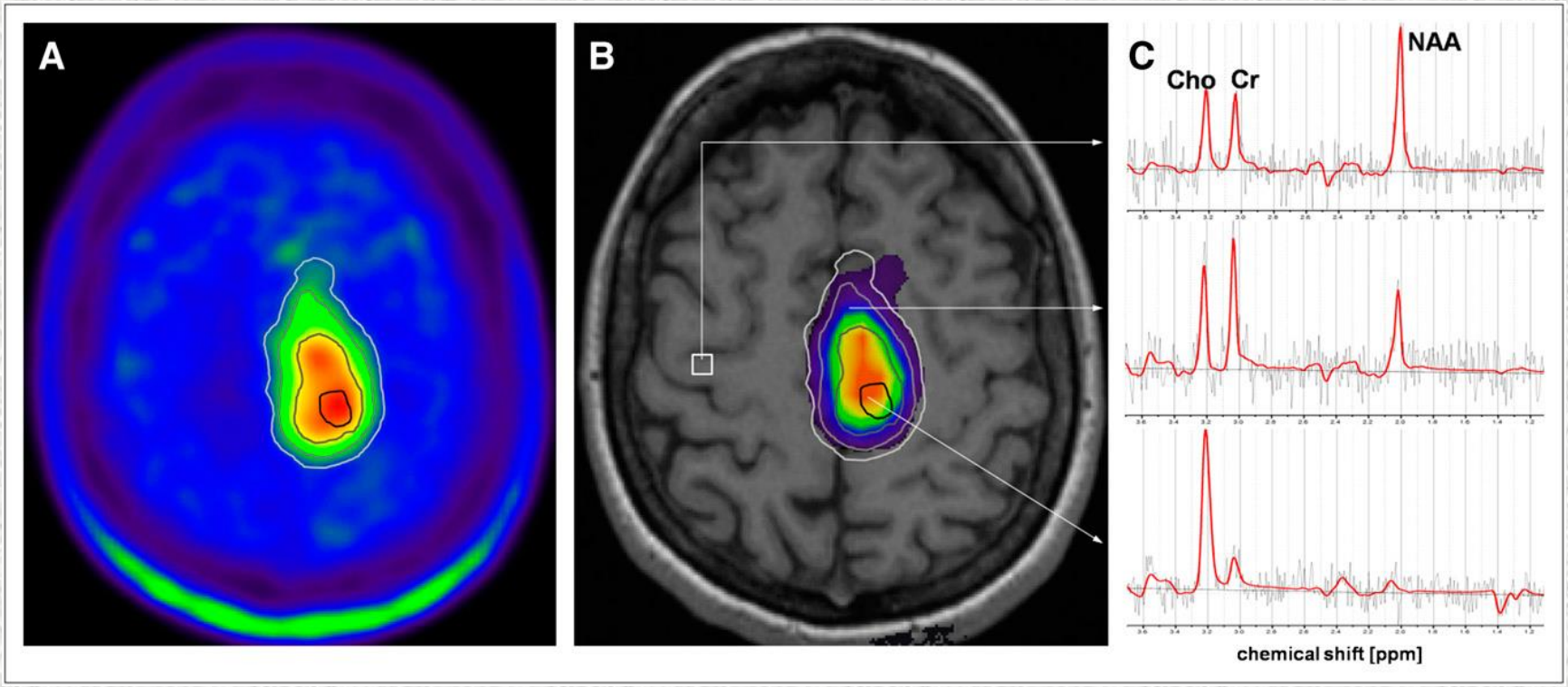
DIAGNOSIS OF NEURODEGENERATIVE DEMENTIAS • Silverman
JOURNAL OF NUCLEAR MEDICINE Vol. 45 No. 4 April 2004

Bi-modal (PET+MRI) Imaging for Parkinson's Disease



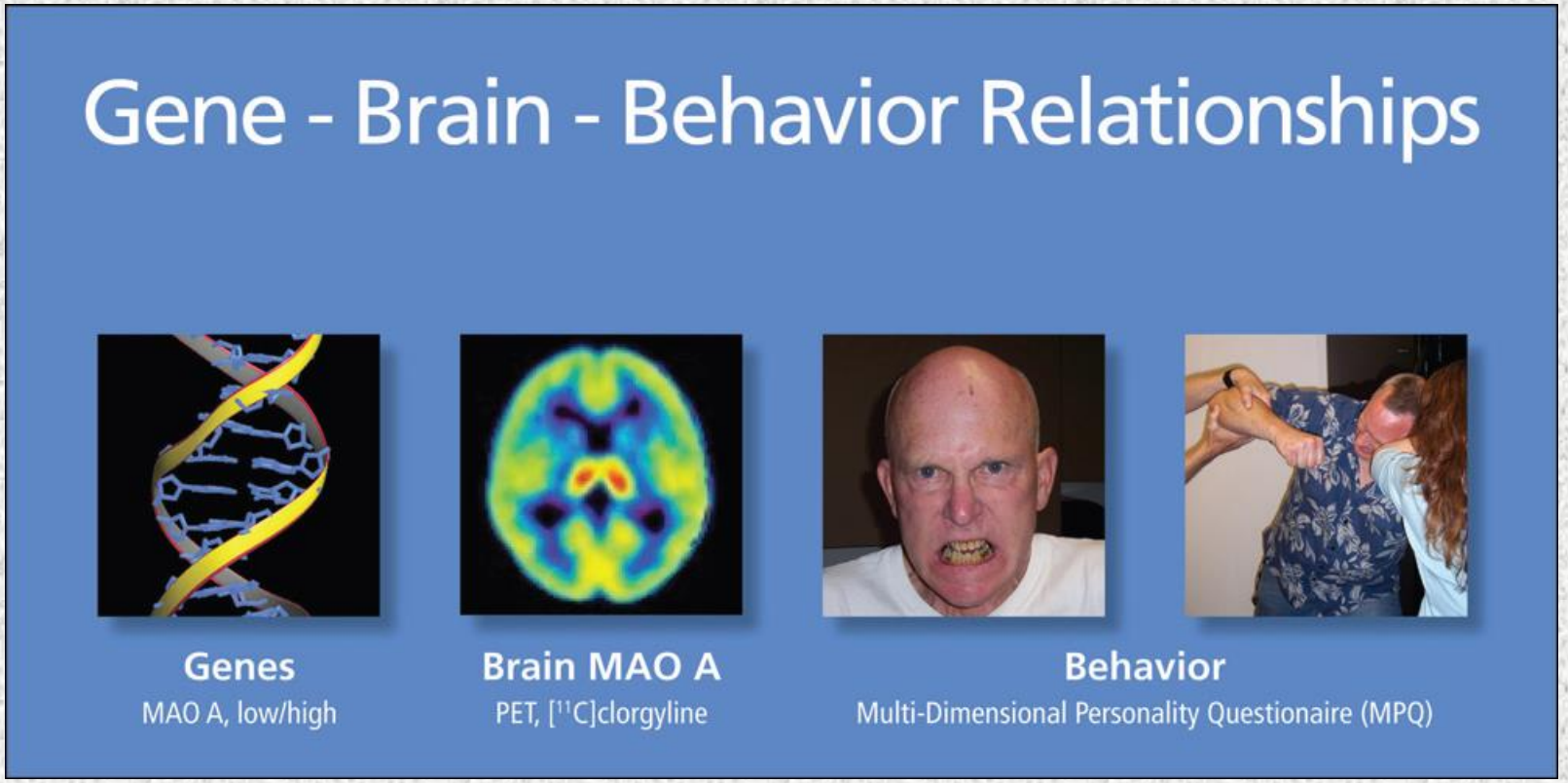
BRAIN RECEPTOR IMAGING • Heiss and Herholz
JOURNAL OF NUCLEAR MEDICINE Vol. 47 No. 2 February 2006

Tri-modal (PET+MRI+MRS) Imaging for Glioblastoma multiforme



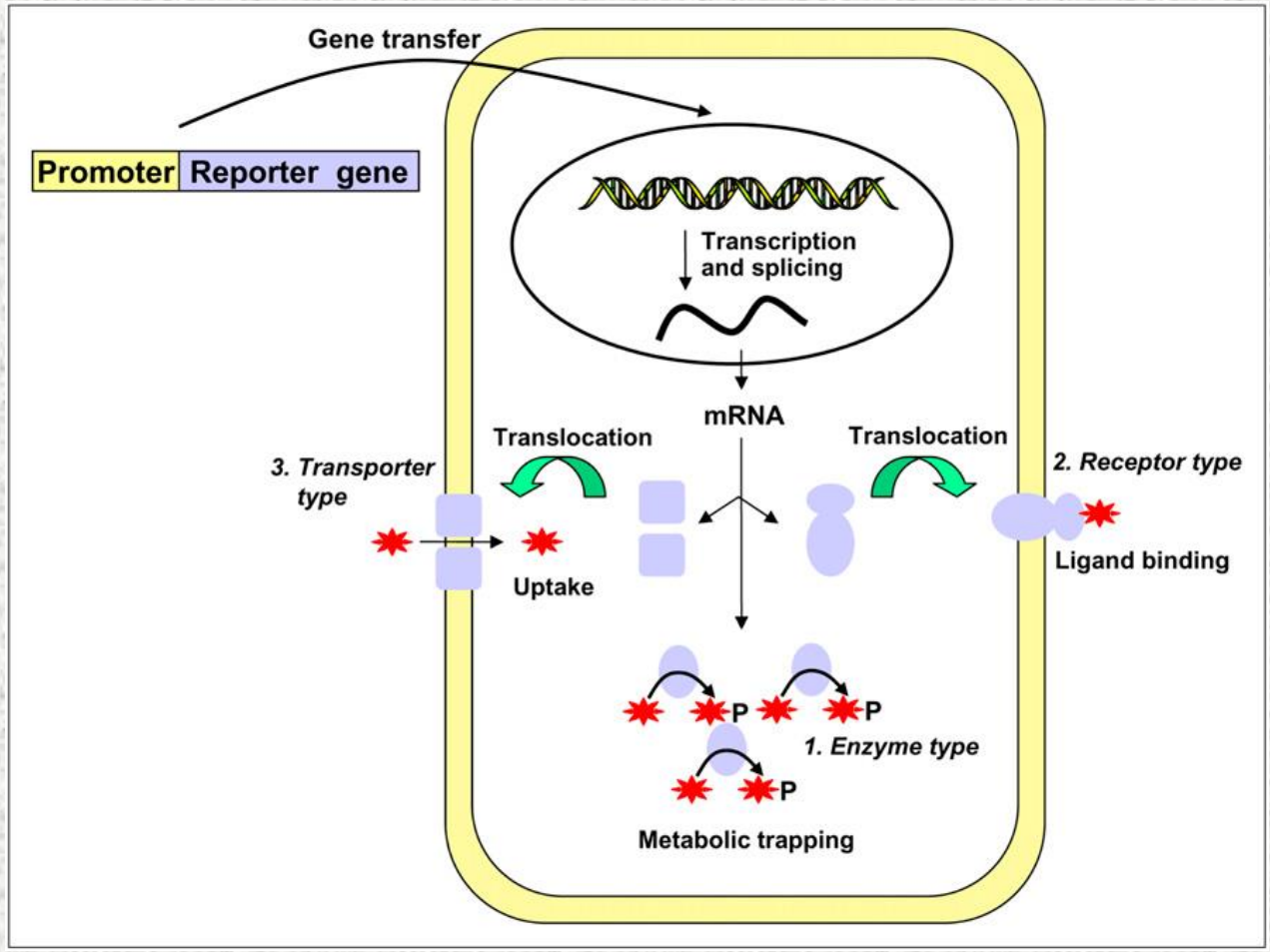
METABOLIC IMAGING OF HUMAN GLIOMAS • Stadlbauer et al.
JOURNAL OF NUCLEAR MEDICINE Vol. 49 No. 5 May 2008

Gene-Brain-Behavior Relationships as a Genotype-Phenotype Correlation Problem



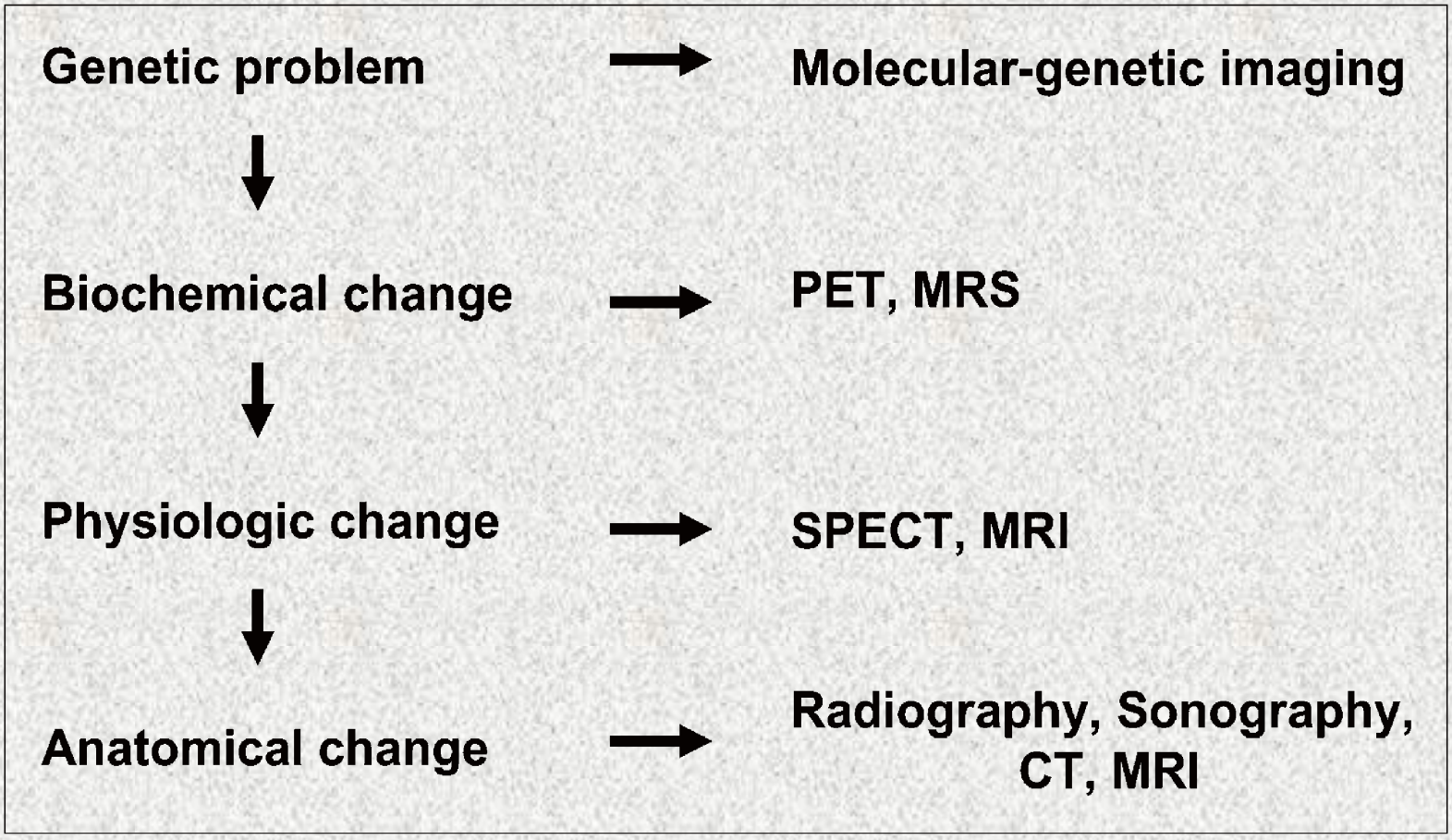
Wagner p15N, Journal of Nuclear Medicine Vol 48 No 7 August 2007

PET/SPECT Molecular Genetic Imaging



IMAGING OF REPORTER GENES • Kang and Chung
JOURNAL OF NUCLEAR MEDICINE Vol. 49 No. 6 (Suppl) June 2008

Molecular Genetic Imaging (MGI) Enables Early Detection



IMAGING OF REPORTER GENES • Kang and Chung
JOURNAL OF NUCLEAR MEDICINE Vol. 49 No. 6 Suppl June 2008

PharmacoGenomic Molecular Imaging

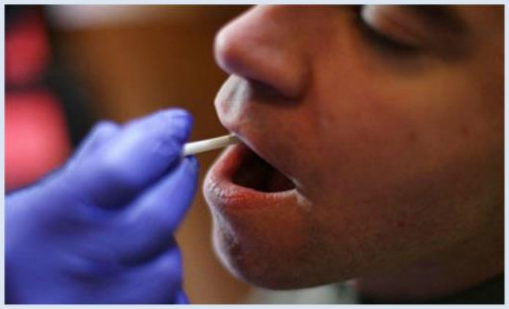
- Term introduced in “PORTAL-DOORS Infrastructure System for Translational Biomedical Informatics on the Semantic Web and Grid” Taswell 2008 AMIA STB
- Concepts elaborated in “Knowledge Engineering for PharmacoGenomic Molecular Imaging of the Brain” Taswell 2009 IEEE SKG
- Updated in “The ManRay Project in Biomedical Informatics for Nuclear Medicine and Pharma-coGenomic Molecular Imaging” 2010 WRSNM

MGI vs PGMI

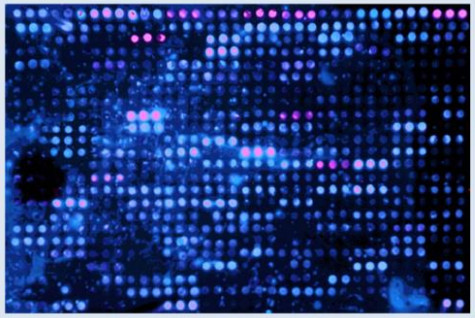
- MGI is not the same as PGMI (by definitions)
- MGI uses molecular genetic techniques to image cells, but does not necessarily involve the study of the pharmacological effects of a drug in relation to the patient's genotype and phenotype
- PGMI may use MGI within the overall study design, but logical converse is not meaningful

PharmacoGenomic Molecular Imaging as a Data Mining "Grand Challenge"

Genetic Testing



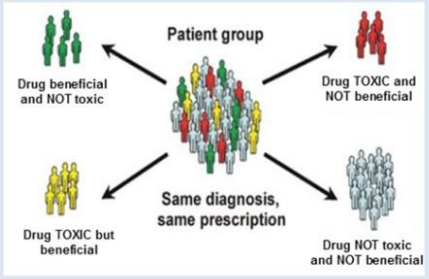
Cheek swab



Gene chip image

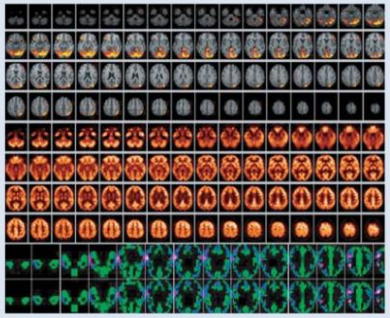
Drug Dosing

Gene-specific drug and dose selection individualized per patient to optimize care and avoid undesirable outcomes shown below for non-individualized treatment



Molecular Imaging

Mobile SPECT scanner



Tri-modal PET/MRI/EEG scans

Gene-specific drug and dose selection individualized per patient to optimize care and avoid the undesirable outcomes of non-individualized treatment approaches

Data Integration/Fusion Problem

- Data with extracted feature sets exists at more than three layers of scale and complexity
  - Genome level with genetic tests for genotypes
  - Brain level with imaging biomarker phenotypes
  - Human level with behavioral syndrome phenotypes
- Genotype-Phenotype Correlation problem further complicated by requirement to track and analyze effectiveness of medical and surgical diagnostic and therapeutic interventions

ManRay Registry in NPDS

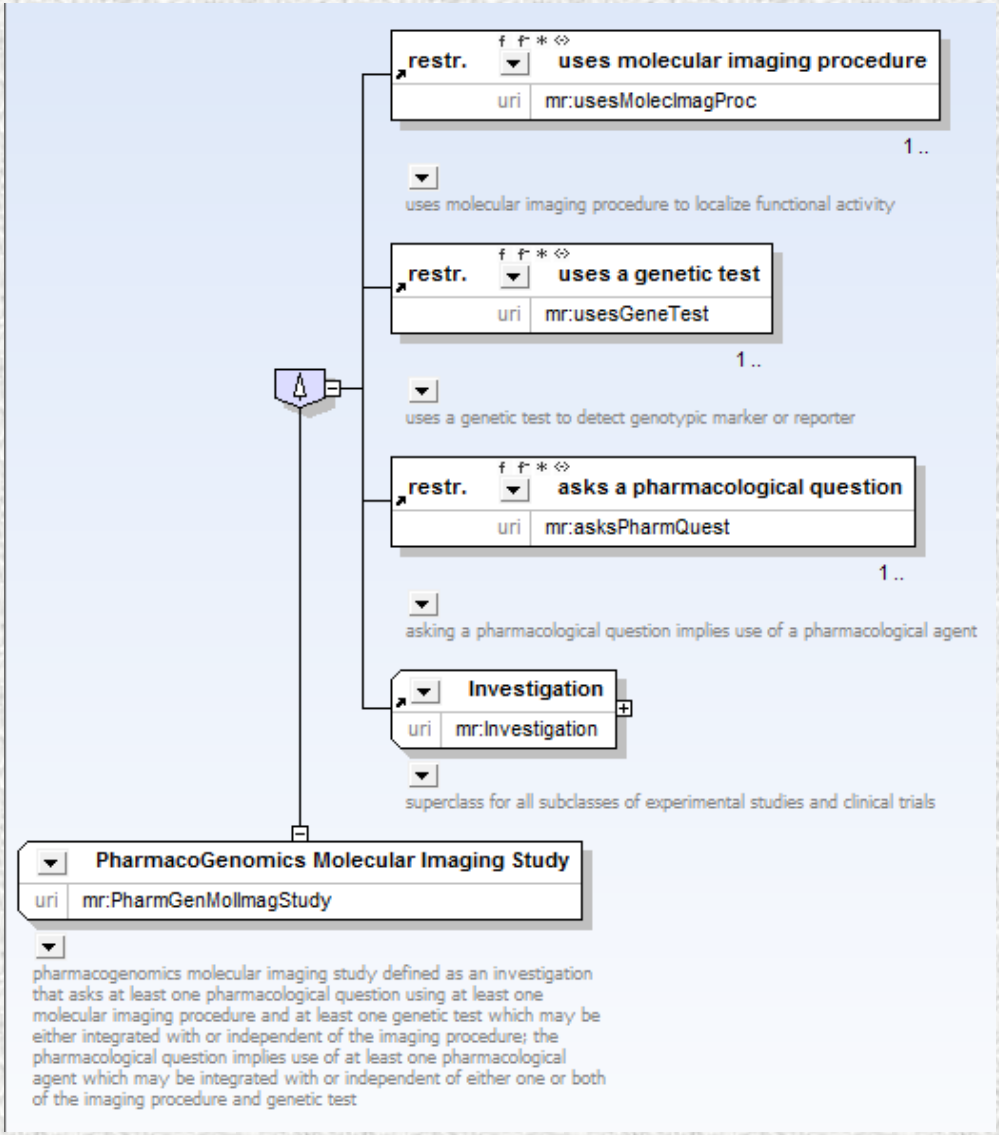
ManRay Registry

	PORTAL Status	DOORS Status	Type	Tag	Name	Nature
›	Valid	Valid	Organization	AMI	Academy of Molecular Imaging	organization for physicians and other professionals in molecular imaging
›	Valid	Valid	Organization	ARR	Academy of Radiology Research	organization for medical imaging research education and advocacy
›	Valid	Valid	OnlineVirtualEntity	AIRO	Advance for Imaging & Radiation Oncology	website for medical imaging and radiation oncology news and information
›	Valid	Valid	Organization	AGHDNM	Allegheny General Hospital Division of Nuclear Medicine	division of nuclear medicine at Allegheny General Hospital
›	Valid	Valid	Organization	AAPM	American Association of Physicists in Medicine	professional group for physicists in medicine
›	Valid	Valid	Organization	ABII	American Board of Imaging Informatics	organization for certification of professionals in imaging informatics
›	Valid	Valid	Organization	ABNM	American Board of Nuclear Medicine	organization for certification of nuclear medicine physicians
›	Valid	Valid	Organization	ABSNM	American Board of Science in Nuclear Medicine	organization for certification of nuclear medicine scientists
›	Valid	Valid	Organization	ACMP	American College of Medical Physics	organization for medical physicists
›	Valid	Valid	Organization	ACNM	American College of Nuclear Medicine	professional association for nuclear medicine

1 2 3 4 5 6 7 8 9 10 ... Page size: 10

178 items in 18 pages

PGMI Class in ManRay Ontology



- Ontology class definitions (such as this example from ManRay ontology) enable new studies
- Drive from question to answer through semantic web of data
- Guided by map of PORTAL-DOORS interconnections

History & Review of the PORTAL-DOORS Project

Nexus-PORTAL-DOORS-Scribe (NPDS)

- A distributed cyberinfrastructure for data repositories to manage online & offline entities
- Read-only API for NPDS services
  - Lexical PORTAL registries
  - Semantic DOORS directories
  - Combined Nexus diristries where a diristry = DIR(ectory) + reg(ISTRY),
- Read-write API for Scribe registrars
- Driving application: automated meta-analysis

2006 Design for PORTAL-DOORS

IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDICINE, VOL. 12, NO. 2, MARCH 2008

DOORS to the Semantic Web and Grid With a PORTAL for Biomedical Computing

Carl Taswell, *Member, IEEE*

***Abstract*—The semantic web remains in the early stages of development. It has not yet achieved the goals envisioned by its founders as a pervasive web of distributed knowledge and intelligence. Success will be attained when a dynamic synergism can be created between people and a sufficient number of infrastructure systems and tools for the semantic web in analogy with those for the original web. The domain name system (DNS), web browsers, and the benefits of publishing web pages motivated many people to register domain names and publish web sites on the original web. An analogous resource label system, semantic search applications, and the benefits of collaborative semantic networks will motivate people to register resource labels and publish resource descriptions on the semantic web. The Domain Ontology Oriented Resource System (DOORS) and Problem Oriented Registry of Tags and Labels (PORTAL) are proposed as infrastructure systems for**

registries are proposed with scientific problem-oriented designs that avoid the engineering-technology-oriented restrictions of existing registries.

Sections II–IV review the background and motivation for DOORS, PORTAL, and BioPORT. Section II explains key concepts of the current semantic web and grid, and summarizes how they are driving the transformation of software architecture from designs based on closed-world computing to those based on open-world computing. Section III reviews the literature and current state-of-the-art in the life sciences web and grid, and summarizes the opinions of leading commentators in the bioinformatics community on existing barriers that impede development. Section IV defines the meaning and scope of biomedical

Manuscript received 10/31/2006; online 8/3/2007; in print 3/5/2008

2009 Update of PORTAL-DOORS

Future Internet **2010**, *2*, 156-189; doi:10.3390/fi2020156

OPEN ACCESS

future internet

ISSN 1999-5903

www.mdpi.com/journal/futureinternet

Article

A Distributed Infrastructure for Metadata about Metadata: The HDMM Architectural Style and PORTAL-DOORS System

Carl Taswell

Global TeleGenetics, Inc., 8 Gilly Flower Street, Ladera Ranch, CA 92694, USA;
E-Mail: ctaswell@computer.org; Tel: 1-949-481-3121

Received: 30 December 2009; in revised form: 12 April 2010 / Accepted: 20 May 2010 / Published: 1 June 2010

Abstract: Both the IRIS-DNS System and the PORTAL-DOORS System share a common architectural style for pervasive metadata networks that operate as distributed metadata management systems with hierarchical authorities for entity registering and attribute publishing. Hierarchical control of metadata redistribution throughout the registry-directory networks constitutes an essential characteristic of this architectural style called Hierarchically Distributed Mobile Metadata (HDMM) with its focus on moving the metadata for *who what where* as fast as possible from servers in response to requests from clients. The novel concept of multilevel *metadata about metadata* has also been defined for the PORTAL-DOORS System with the use of entity, record, infoset, representation and message metadata. Other new features implemented include the use of aliases, priorities and metaresources.

2010 Future Internet 2(2):156-189; doi:10.3390/fi2020156

2016 Update of PORTAL-DOORS

Web Service APIs for Scribe Registrars, Nexus Diristries, PORTAL Registries and DOORS Directories in the NPD System

Adam G. Craig, Seung-Ho Bae, Teja S. Veeramacheneni,
S. Koby Taswell, and Carl Taswell

www.BrainHealthAlliance.org
ctaswell@BrainHealthAlliance.org
8 Gilly Flower Street, Ladera Ranch, CA 92694 USA

Abstract. The Nexus-PORTAL-DOORS System (NPDS) has been designed with the Hierarchically Distributed Mobile Metadata (HDMM) architectural style to provide an infrastructure system for managing both lexical and semantic metadata about both virtual and physical entities. We describe version 0.8 of NPDS, including the separation of concerns between the original Problem-Oriented Registry of Tags And Labels (PORTAL) registries and the Domain Ontology Oriented Resource System (DOORS) directories, the combined registry and directory functionality of Nexus diristries, and the RESTful read-only web service API through which resource representation metadata records can be retrieved from these NPDS servers. We also introduce Scribe registrars with a corresponding RESTful read-write web service API for management of metadata records by both software agents accessing the web services directly and human users accessing them indirectly via web applications.

SWAT4HCLS 2016 (ceur-ws.org/Vol-1795/paper4.pdf)

2019 Update of PORTAL-DOORS

2019 IEEE 11th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)

DREAM Principles and FAIR Metrics from the PORTAL-DOORS Project for the Semantic Web

1st Adam Craig
Brain Health Alliance
Ladera Ranch, California
acraig@bhavi.us

2nd Adarsh Ambati
Brain Health Alliance
Ladera Ranch, California
aambati@bhavi.us

3rd Shiladitya Dutta
Brain Health Alliance
Ladera Ranch, California
sdutta@bhavi.us

4th Pooja Kowshik
Brain Health Alliance
Ladera Ranch, California
pkowshik@bhavi.us

5th Sathvik Nori
Brain Health Alliance
Ladera Ranch, California
snori@bhavi.us

6th S. Koby Taswell
Brain Health Alliance
Ladera Ranch, California
ktaswell@bhavi.us

7th Qiyuan Wu
Brain Health Alliance
Ladera Ranch, California
qwu@bhavi.us

8th Carl Taswell*
Brain Health Alliance
Ladera Ranch, California
ctaswell@bhavi.us

***Abstract*—Articles published in *Scientific Data* by Wilkinson *et al.* argued for the adoption of the Findable, Accessible, Interoperable, and Reusable (FAIR) principles of data management without citing any of the prior work published by Taswell. However, these principles were first proposed and described by Taswell in 2006 as the foundation for work on the PORTAL-DOORS Project (PDP) and the Nexus-PORTAL-DOORS-Scribe (NPDS) cyberinfrastructure, and have been published in numerous conference presentations, journal articles, and patents. This work on PDP and NPDS has been continuously available since 2007 from a publicly accessible web site at www.portaldoors.org,**

set of stakeholders – representing academia, industry, funding agencies, and scholarly publishers – have come together to design and jointly endorse a concise and measurable set of principles that we refer to as the FAIR Data Principles" [1]. While advocacy by more 'stakeholders' for making data findable, accessible, interoperable and reusable represented progress towards the goal of reproducible science, their use of the term 'design' gave the impression that these 'stakeholders' originated these principles. However, all of the FAIR principles

IEEE ECAI 2019 (DOI 10.1109/ECAI46879.2019.9042003)

2020 Update of PORTAL-DOORS

2020 Second International Conference on Transdisciplinary AI (TransAI)

NPDSLINKS: Nexus-PORTAL-DOORS-Scribe Learning Intelligence aNd Knowledge System

1st Shreya Choksi
Brain Health Alliance
Ladera Ranch, CA, USA
schoksi@bhavi.us

2nd Peter Hong
Brain Health Alliance
Ladera Ranch, CA, USA
phong@bhavi.us

3rd Sohyb Mashkoor
Brain Health Alliance
Ladera Ranch, CA, USA
smashkoor@bhavi.us

4th Carl Taswell
Brain Health Alliance
Ladera Ranch, CA, USA
ctaswell@bhavi.us

***Abstract*—With the continuing growth in use of large complex data sets for artificial intelligence applications (AIA), unbiased methods should be established for assuring the validity and reliability of both input data and output results. Advancing such standards will help to reduce problems described with the aphorism 'Garbage In, Garbage Out' (GIGO). This concern remains especially important for AIA tools that execute within the environment of interoperable systems which share, exchange, convert, and/or interchange data and metadata such as the *Nexus-PORTAL-DOORS-Scribe* (NPDS) cyberinfrastructure and its associated *Learning Intelligence aNd Knowledge System* (LINKS) applications. The PORTAL-DOORS Project (PDP) has developed the NPDS cyberinfrastructure with lexical PORTAL registries, semantic DOORS directories, hybrid Nexus diristries,**

The simple intuitive principle implied by that description has remained central to the core foundation of calculating and computing machines from the early history of primitive computers to the present era with the advances of multi-core chip architectures, big data, and artificial intelligence.

Over a century after Babbage made his famous remarks, Army Specialist William Mellin expressed his concern about the inability of computers to think for themselves when interviewed for a 10 November 1957 newspaper article, and explained that "sloppily programmed" inputs inevitably lead to incorrect outputs [3]. The Hammond Times newspaper of Hammond Indiana published Mellin's explanation of this

IEEE TransAI 2020 (DOI 10.1109/TransAI49837.2020.00027)

The Essence of PORTAL-DOORS

- PORTAL-DOORS for the semantic web modeled on the success of IRIS-DNS for the original web
- PORTAL-DOORS designed to address major problems including: spread of misinformation, cybersilos, search engine oligopolies (monopolies?), and continuing transition barriers
- Significant benefits for translational bioinformatics:
  - Distributed registry-directory system
  - Important applications for clinical trials
  - Important applications for complex information systems such as those necessary for the study of brain imaging and gene-brain-behavior relationships

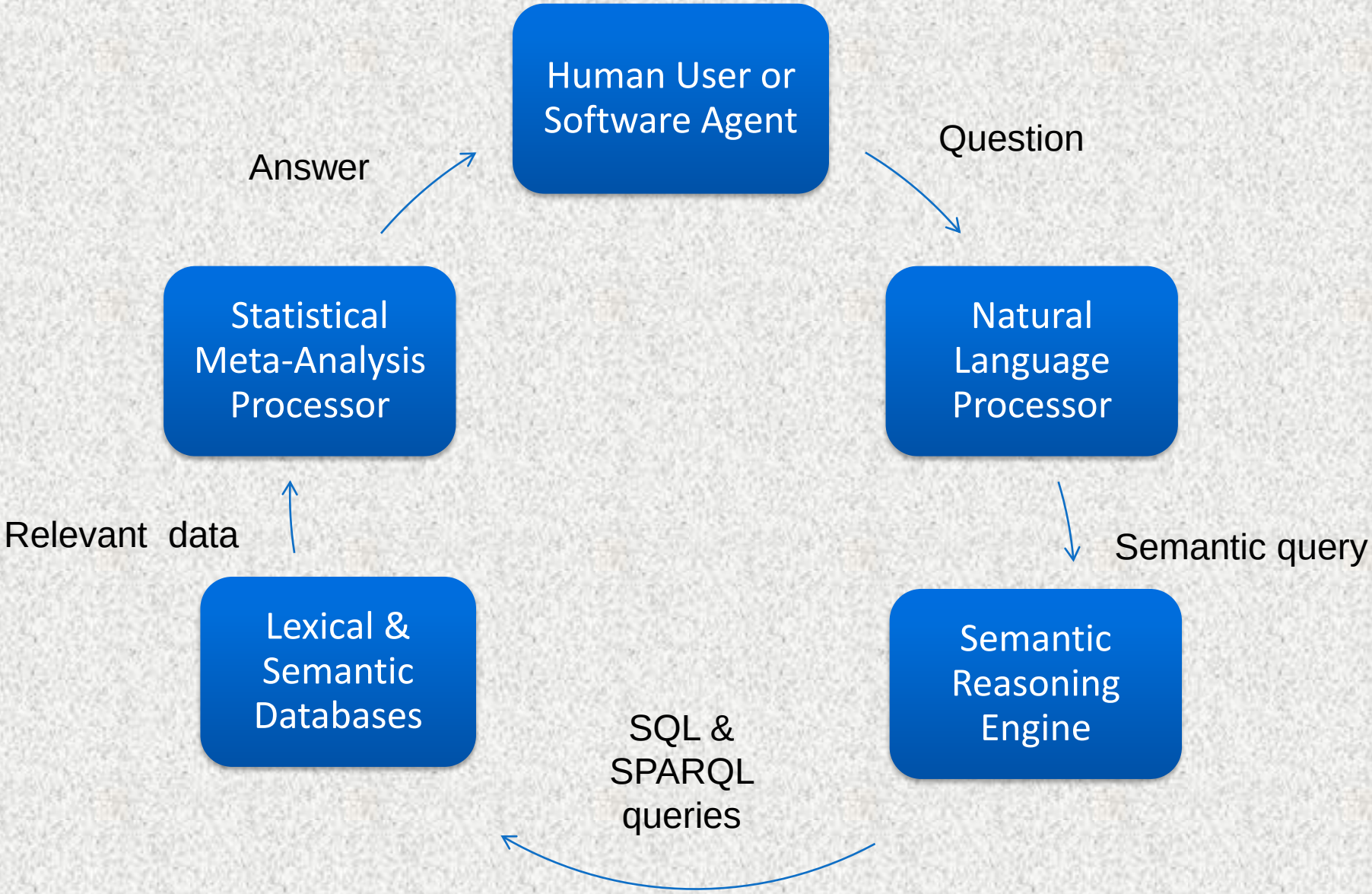
The Problem of Cybersilos

- Example cybersilos? caBIG, caGRID and NCIP built for cancer and oncology by NCI?
- What about brain disease and neurology and psychiatry? Or genetics and all fields of medicine?
- What about chemistry and organic molecules? Or physics and radioisotopes?
- What about plants, animals, and biodiversity? Or all sciences and “astronomy to zoology”?
- Have traditional silos been replaced by cybersilos? Or can we build truly interoperable informatics systems networked without barriers to the freely flowing exchange of information?

Problem: Literature Review

- How do we gather evidence from the clinical trial literature for or against a scientific hypothesis?
- Type the phrase “tauopathy and frontotemporal dementia” into Google Scholar?
  - Then search through “about 14,000 results”.
  - Which ones answer the question?
- Ask a colleague? Query a cybersilo database?
  - Get results with enough breadth and depth?
  - Get results with too much breadth or depth?

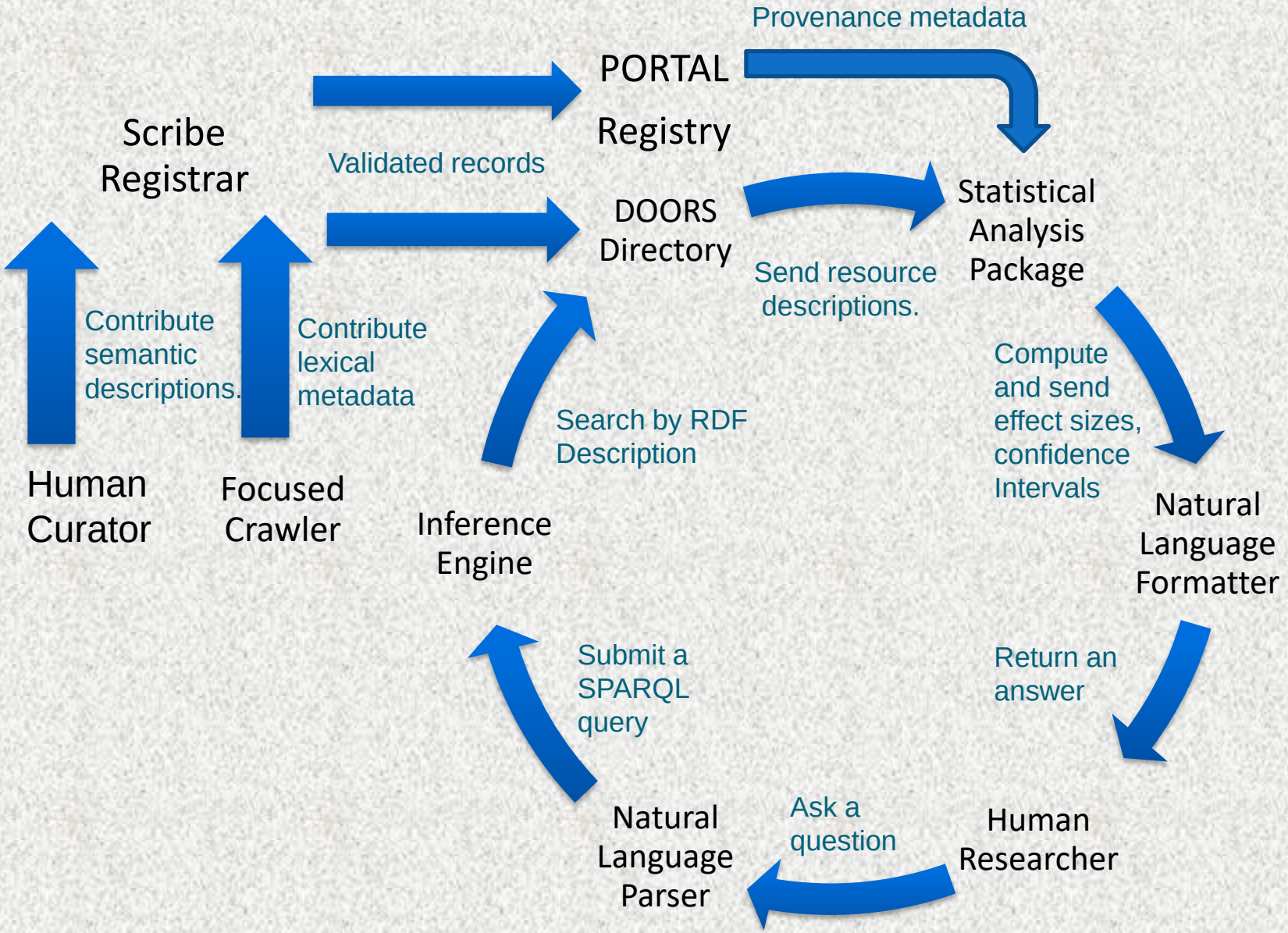
Solution: Automated Meta-Analysis



Search Within/Across Problem Domains

- Beacon – biomedical computing
- BrainWatch – brain health sciences
- CTGaming – clinical telegaming
- Eywa – biodiversity and ecology
- Gaia – biosurveillance and toxicology
- GeneScene – genetic health sciences
- HELPME – health education law policy and medical ethics
- ManRay – nuclear medicine
- Osler – personalized medicine
- SOLOMON – sensory-onset language-onset and motor-onset neurodegenerative disorders

Solution: Automated Meta-Analysis



Maintaining the Integrity of Problem Oriented Domains

- Test metadata records entered in a PORTAL registry for compliance with the concept restrictions imposed by the scope declared for the problem-oriented registry
- Retain the metadata records that pass the compliance tests and delete those that fail
- Alternatively, move records that fail the tests to a more appropriate registry where the records do pass the tests, or else to a catch-all registry that does not impose any concept restrictions

Concept Validating Methods

- A matching algorithm for testing the metadata elements of the PORTAL registry record for the presence of any of the required word stems, terminology term URIs or thesaurus concept URIs
- A matching algorithm for testing the metadata elements of the same registered resource entity at the corresponding DOORS directory record for the presence of any of the required ontology concept URIs
- A workflow algorithm for ordering the sequence of sub-tests within the overall validation test that terminates as early as possible upon first successful pass of the validation test (for the pair of PORTAL and DOORS metadata records for each registered resource entity)

Improved Search Query Efficiency

- Validation tests are simple pass/fail, can be performed with regex matching, and do NOT require any numeric probabilistic analyses
- Problem oriented domains that maintain the integrity of their declared scope can be searched more efficiently because there are NO irrelevant records enlarging the search space
- Optimizing search efficiency becomes increasingly important in direct relation to the increasing scale of the number and size of problem oriented domains

CTGaming PORTAL Registry

GTG CTGaming Registry

CTGaming Resource Records

	PORTAL Status	DOORS Status	Type	Tag	Name	Nature
›	Valid	Valid	Publication	ALCFIIT	A Low-Cost Framework for Individualized Interactive Telerehabilitation	Examines a low-cost framework for upper-limb telerehabilitation
›	Valid	Valid	Publication	AVRRG	Adaptive VR Rehab Games	Development of a VR Game System for Therapeutic Training of Stroke Patients
›	Valid	Valid	Organization	ATA	American Telemedicine Assocation	Quality healthcare through telecommunications technology
›	Valid	Valid	Publication	AXARM	AXARM	Remote Assistance and Monitoring Rehabilitation Tool Publication
›	Valid	Valid	Organization	BL	Biometrics Ltd.	Unique Solutions for Clinical and Research Applications
›	Valid	Valid	Organization	CAN	Canadian Arthritis Network	Today's Arthritis Research Tomorrow's Cure
›	Invalid	Valid	Organization	EASG	EASe Games	therapeutic gaming software for children with autism spectrum disorders and other children diagnosed with auditory hypersensitivity, hyperacusis, central auditory processing disorder, or sensory integration disorder
›	Valid	Valid	Publication	EVRTS	Evaluation of a Virtual Reality Telerehabilitation System	Technical and patient performance of the Rutgers Ankle virtual reality telerehabilitation system
›	Valid	Valid	Publication	FASTS	Force-Assistant Stroke Telerehabilitation System	A system using force-assistance and virtual reality for stroke telerehabilitation
›	Valid	Valid	Publication	GTNT	Game-based Telerehabilitation for Neurological Trauma	Robot assisted neurological telerehabilitation using games for motivation

1 2 3 4 5 6 Page size: 10 53 items in 6 pages

3 Concepts for CTGaming Registry

- AND (clinical OR medical OR diagnostic OR therapeutic OR health care OR rehabilitation OR intervention) == clinical care concept
- AND (telemedicine OR telecare OR telemonitoring OR remote interaction) == telecomm concept
- AND (telegaming OR videogaming OR games OR simulations) == gaming concept
- Should scope be widened or narrowed for problem oriented domain of CTGaming Registry ???

Distributed Systems & Architectural Styles

Distributed System Architectures

- Main approaches:
  - Client-Server, Publisher-Subscriber
  - Object Distribution vs Distributed Object
  - Distributed vs Networked
  - Centralized vs Decentralized
  - Warehouse vs Federated
  - P2P, Grid, Cloud, SOAs, microservices, etc
- Some references:
  - “Distributed Systems Architecture and Specification” Norman Howe 2010 IEEE CS
  - “Software Architecture: Foundations, Theory and Practice” Taylor et al 2010 Wiley

Definition of Architectural Style

- “an architectural style is a named collection of … design decisions that
  - (1) are applicable in a given development context,
  - (2) constrain … design decisions [for] a particular system within that context, and
  - (3) elicit beneficial qualities in each resulting system.”
- Definition per Taylor et al 2010 page 73
- REST and HDMM as examples of styles

An Aside on REST

- Contrary to common misunderstanding, HTTP + XML/JSON is not a requirement of REST
- REST is an architectural style expressed as a set of six principles imposed as constraints on a specific architectural instantiation
- Application of REST style to the specific instantiation with HTTP + XML/JSON is just one of many possible instantiations of RESTful system architectures

HDMM Architectural Style

- Pervasively distributed and shared infrastructure, content, and control of content.
- A hierarchy of both authoritative and non-authoritative servers enabling global interoperable communication while permitting local control of policies.
- A focus on efficiently moving the metadata for "who what where" from servers in response to requests from clients that access non-authoritative local forwarding and caching servers updated regularly by the authoritative servers.
- A separation of concerns with registries for identification and directories for location.
- A freedom of choice in the selection of identifiers.

Taswell 2009 ISPAN (DOI 10.1109/I-SPAN.2009.73)

PORTAL-DOORS vs IRIS-DNS

Domain Naming & Registering

- Success of DNS due to purposeful avoidance of any requirement for client or user to possess prior knowledge of domain name's governing registry
- DNS extensions for security and multilingualism with DNSSEC and IDNA
- IETF Cross Registry Information Service Protocol (CRISP) Working Group has completed versions of the Internet Registry Information Service (IRIS) Core and several IRIS-dependent protocols for different types of registries
- IRIS-XPC (XML pipelining with chunks) is default transport for IRIS with support for security and multilingualism

Resource Identifying & Linking

- IRIS designed to associate information with labels declared by registry (a replacement for whois protocol)
- Resource Directory Description Language (RDDL) built on XHTML and XLink
- XML Topic Maps (XTM)
- Persistent Uniform Resource Locator (PURL) System (www.purl.org)
- Handle System (www.handle.net)
- None of these systems have been built or modified for the semantic web with use of RDF and OWL

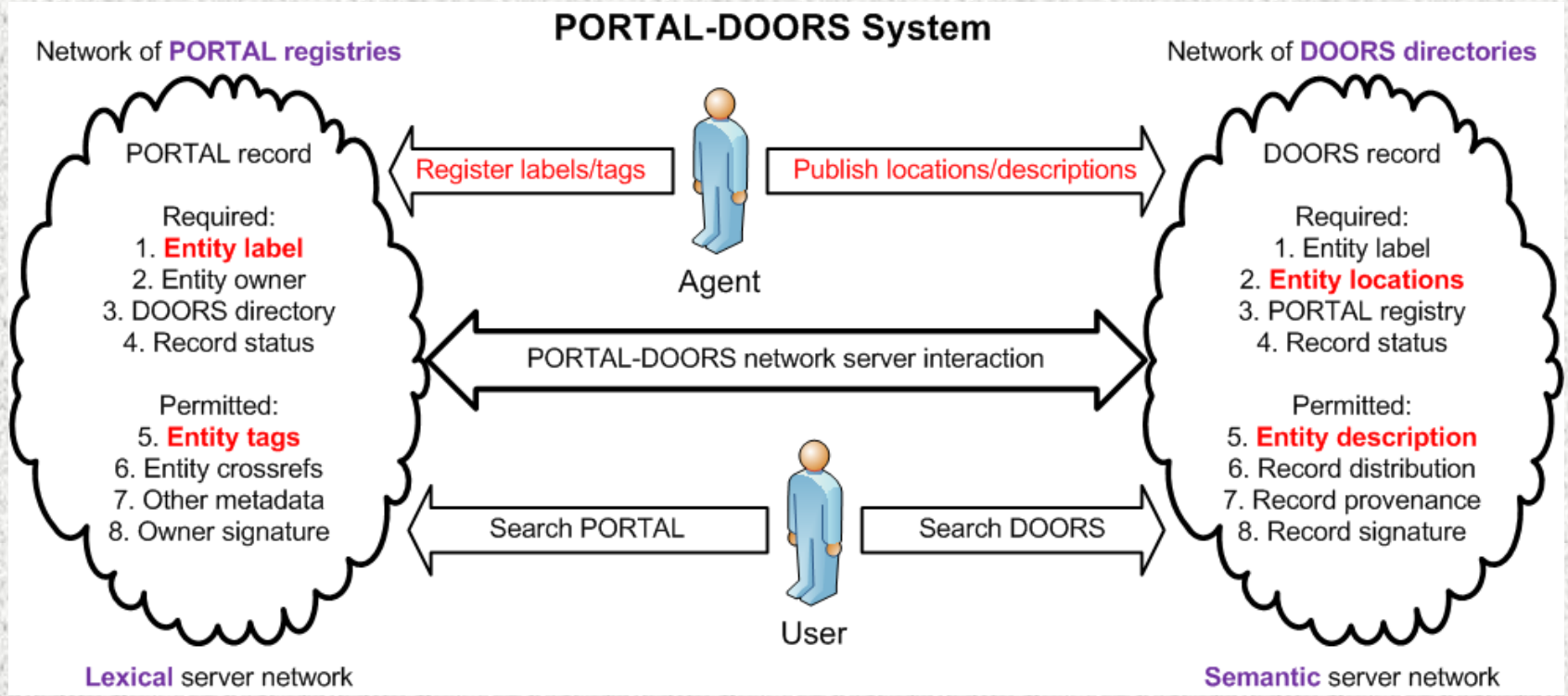
PORTAL-DOORS vs IRIS-DNS

- A paradigm and infrastructure for the semantic web and grid (considered novel when proposed in 2006)
- Problem Oriented Registry of Tags And Labels (PORTAL) for resource entity label and tag registering (designed as an IRIS analogue)
- Domain Ontology Oriented Resource System (DOORS) for resource entity location and description publishing (designed as a DNS analogue
- PORTAL-DOORS for the semantic web modeled on the success of IRIS-DNS for the original web
- PORTAL-DOORS uses an analogous paradigm with resource entity labels instead of domain names
- Taswell 2008 IEEE TITB 12(2):191-204

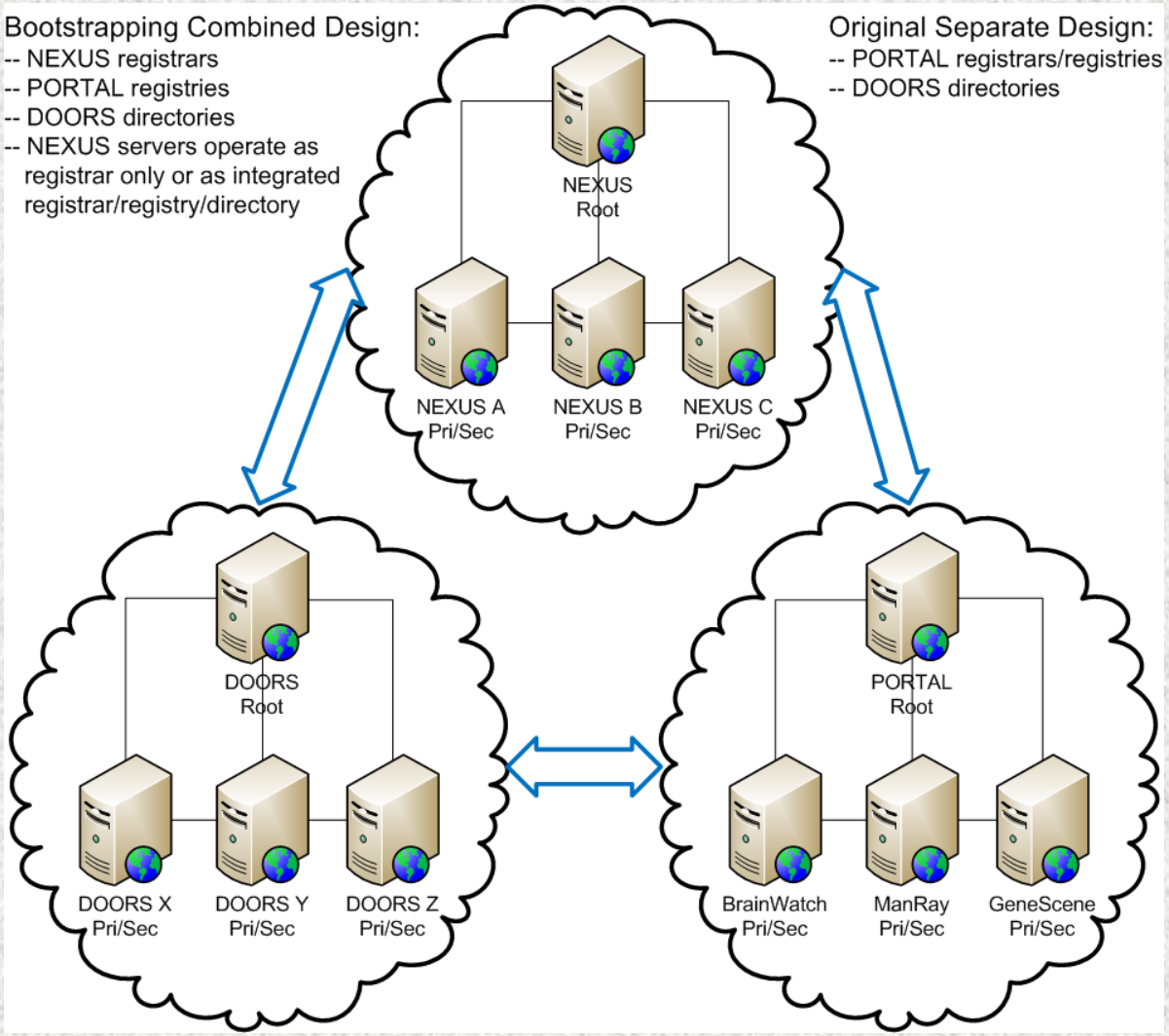
Hierarchically Distributed Mobile Metadata Systems with Entity Registering and Attribute Publishing

	IRIS-DNS	**PORTAL-DOORS**
Registering system	IRIS registries	PORTAL registries
-- Entity registered	domain	resource
-- Identified by	unique name	unique label
Publishing system	DNS directories	DOORS directories
-- Attributes published	address	location and description
-- Specified by	IP number	URIs, URLs, RDF triples

PORTAL-DOORS Data Records



PORTAL-DOORS Server Network



From PORTAL-DOORS to NPDS: This 2010 design has been replaced by the 2016 design with Nexus diristries, PORTAL registries, DOORS directories, and Scribe registrars.

PORTAL-DOORS: Some Key Principles

- Assures globally unique identification of resources while promoting interoperability
- Enables cross-registry and cross-directory searches between different problem domains
- Problem oriented, not technology restricted
- A distributed infrastructure that permits localized control of policies and content
- A hybrid bootstrap and bridge to transition from old lexical web to new semantic web

PORTAL-DOORS: Some Other Points

- Physical infrastructure, content control, and content itself are all distributed and shared
- Analogous to DNS where data records are distributed and mobile with request forwarding and response caching
- Compare Wikipedia where content is centralized but control of content is shared
- Compare Google where infrastructure is distributed but not the control of content

Semantic Search &
The HDMM Conjecture

Semantic System

- A lexical (“dumb”) system is an informatics system in which words are processed as character strings that have no meaning to the processing agent
- A semantic (“smart”) system is one in which words have defined meaning to the agent processing them with logic-based reasoners
- *Semantic search can be very efficient while lexical search can be very inefficient*

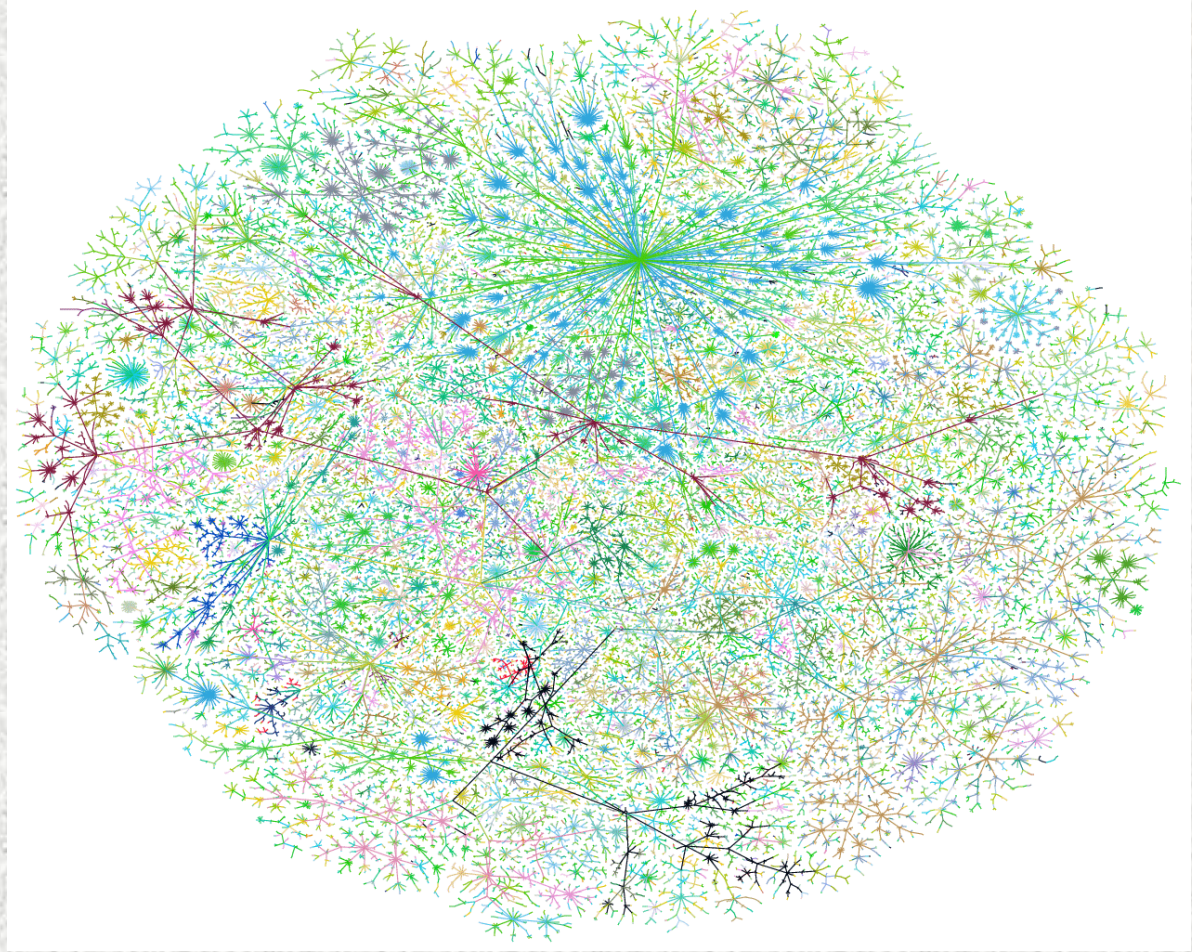
Semantic Search & Applications

- Current web search often yields irrelevant innumerable results that may lose practical usefulness (too costly in time for a person to review)
- Fast accurate delivery of relevant information an important motivating driver for development of semantic web search applications (how to find needle in haystack)
- Translational medical research with drug development, clinical trials, and personalized medicine (including smart search through records linked by genetic pedigree)

Democratization of Search

- PORTAL-DOORS interlinks registries, directories, databases across fields, disciplines, specialties
- PORTAL-DOORS supports mass collaboration via its hierarchical and distributed but decentralized and localizable infrastructure
- PORTAL-DOORS provides a democratized solution to the problems of search engine consolidation
- Mowshowitz and Kumar in Feb 2009 IEEE Computer (p108) discuss the realities and risks of search engines that restrict access to information
- Should a monopoly or some oligarch-led oligopolies control the flow of information in a democracy?

Structure and Function of Networks

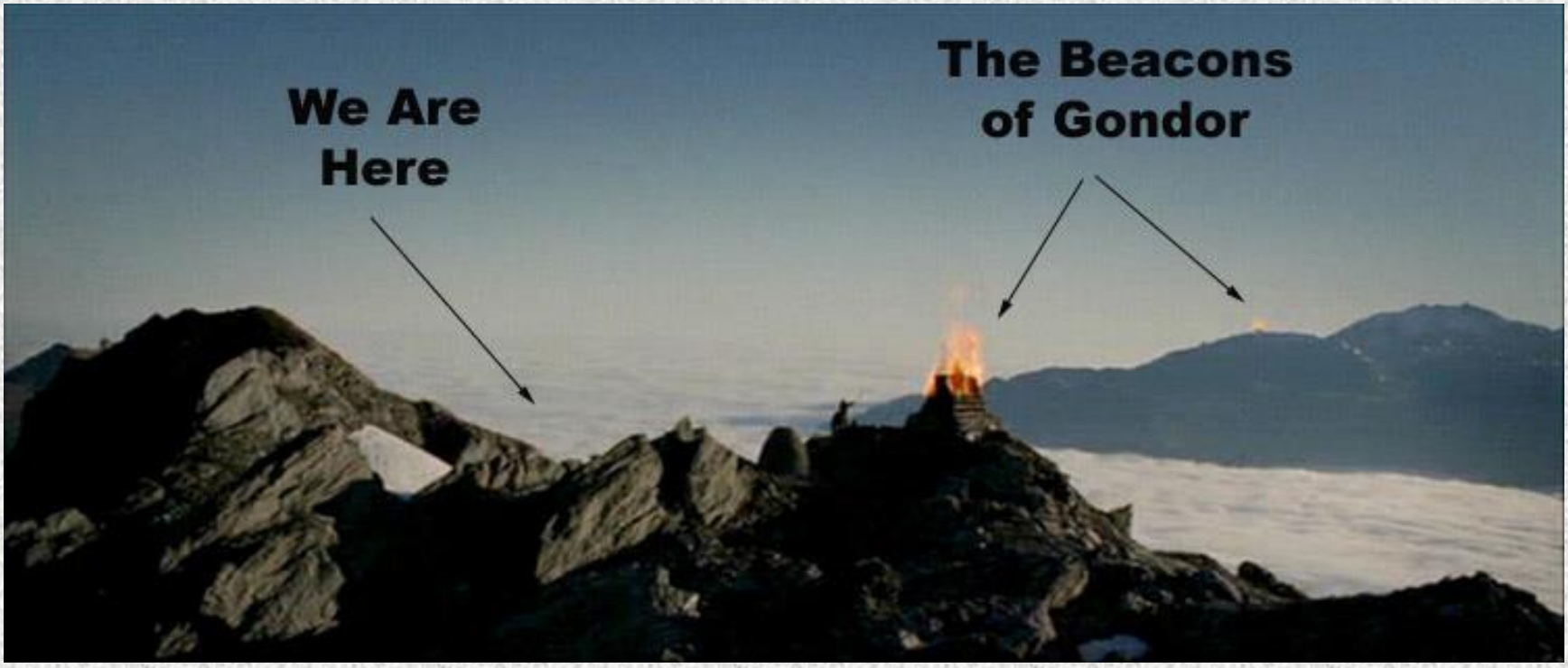


Newman 2003 SIAM Review 45:167 on page 170

Algorithmic Search in Network Graphs

- Random graphs and other dynamic models of network growth
- Node degree distributions, clustering, and preferential attachment
- Search via hierarchical versus peer-to-peer network paths
- Search for best path from known source to known target node → peer-to-peer?
- Search for any path to unknown node for possibly non-existent target → hierarchical?

“Beacons of Gondor” Metaphor



Hierarchical communication networks enable search and discovery of a small item in a very large world when existence and location of item is unknown.

Taswell 2010 Future Internet (DOI 10.3390/fi2020156)

HDMM Conjecture

- Hierarchically Distributed Mobile Metadata (HDMM) architectural style characterizes both IRIS-DNS and PORTAL-DOORS registry-directory *who-what-where* metadata management systems
- Semantic HDMM networks *should scale more effectively and efficiently* than semantic peer-to-peer networks when searching (by various query criteria) for an unknown entity at an unknown location when it may not even exist

DREAM Principles & FAIR Metrics

- DREAM: Discoverable Data with Reproducible Results for Equivalent Entities with Accessible Attributes and Manageable Metadata (= original 2006 PORTAL-DOORS principles + “equivalent entities” principle)
- FAIR: Fair Acknowledgment of Information Records and Fair Attribution to Indexed Reports (with quantitative metrics to evaluate fair citation of published data/papers)
- Some recent references:
  - ECAI 2019 (DOI 10.1109/ECAI46879.2019.9042003)
  - eScience 2019 (DOI 10.1109/eScience.2019.00081)
  - ICSC 2020 (DOI 10.1109/ICSC.2020.00044)
  - TransAI 2020 (DOI 10.1109/TransAI49837.2020.00028)

Next Steps for PORTAL-DOORS Project

- NPDS Cyberinfrastructure implemented with Microsoft C#, SQL Server, and IIS Server
  - Version 9.4.* (current DotNet 6.0 RC previews)
  - Version 10.0.* (November 8 DotNet 6.0 GA release)
- PDP-DREAM Software Library for NPDS
  - github.com/BHAVIUS/PDP-DREAM
- PORTALDOORS Web Apps & Services
  - (npds|www).PORTALDOORS.org
  - (npds|www).PORTALDOORS.net
- NPDSLINKS Web Apps & Services
  - (npds|www).NPDSLINKS.org
  - (npds|www).NPDSLINKS.net

Future Goals for PORTAL-DOORS Project

- NPDS cyberinfrastructure requires interoperability with message exchange independent of server implementation
- NPDS cyberinfrastructure implemented with Microsoft C#, SQL Server, and IIS Server
- NPDS cyberinfrastructure implemented with Python and Django based stack with document-oriented database
- NPDS cyberinfrastructure implemented with JavaScript and Node based stack with document-oriented database
- NPDS cyberinfrastructure implemented with backend database that has native graph-oriented design

For More Info…

- Via email:
  - ctaswell@eng.ucsd.edu
  - ctaswell@brainhealthalliance.org
- Via websites:
  - www.PORTALDOORS.org/NPDS/Site/Papers
  - www.BrainHealthAlliance.org/BhaStemm/Papers

Abstract

The Nexus-PORTAL-DOORS-Scribe (NPDS) cyberinfrastructure provides a ‘who what where’ diristry-registry-directory system for identifying, describing, locating and linking things on the internet, web and grid. PORTAL registries identify resources with unique labels and lexical tags in a manner compatible with the lexical web. DOORS directories specify locations and semantic descriptions for these identified resources in a manner compatible with the semantic web. PORTAL registries and DOORS directories were designed to be analogous to IRIS registries and DNS directories. This original design has been enhanced with Nexus diristries to provide integrated services combining the functions of both PORTAL registries and DOORS directories. The principles for the PORTAL-DOORS Project (PDP) were first proposed and described by Taswell in 2006 as the foundation for work on PDP and the NPDS cyberinfrastructure. This work on PDP and NPDS has been continuously available since 2007 from a publicly accessible web site at www.PORTALDOORS.org. The 2006 PDP principles were renamed the 2019 DREAM principles with the acronym DREAM for "Discoverable Data with Reproducible Results for Equivalent Entities with Accessible Attributes and Manageable Metadata". PDP-DREAM software, available as open source software at Github, provides a comprehensive suite of software for management of the data repositories in the NPDS cyberinfrastructure. A version of PDP-DREAM software has been implemented with Microsoft platform technologies (C#, SQL Server, IIS Server), has been tested on the previews for Net 6, and will be fully validated for compatibility with Net 6 concomitant with its general availability release later in 2021.