GINAS: Global Ingredient Archival Service

TOWARD AN OPEN IMPLEMENTATION OF SUBSTANCE REGISTRATION AND ISO/IDMP 11238

February 2, 2015

NCATS

National Center for Advancing Translational Sciences

GInAS/11238 Collaborators

FDA Larry Callahan, Frank Switzer, Vada Perkins, Yulia Borodina

NIH/NCATS Trung Nguyen, Tyler Peryea, Noel Southall

Health Canada Vik Srivastava

BfArM (Germany) Thomas Balzer

CBG-MEB (Netherlands) Herman Diederik, Ciska Matai

Swiss Medic Philipp Weyermann

EDQM Chris Jarvis

USP Fouad Atouf, Tina Morris, Andrzej Wilk

EMA Sabine Brosch, Ilaria Del Seppia, Telonis Pangiotis

Substance Identification: In Regulatory Practice

- Knowledge of substances is essential for understanding safety and efficacy of medicinal products
- Repeated request of substance information by regulators is inefficient
- Regulators/Industry must understand and be provided with the defining elements and criteria associated with substance identification
- An ISO standard ISO 11238 has recently been developed to describe all substances in medicinal products

ISO 11238:
A global substance standard

- An international standard which
  - » provides a structure for assignment and maintenance of unique identifiers for all substances in medicinal products
  - » sets out the general rules for defining and distinguishing substances
  - » provides a high-level model that structures substances and specified substances for the organization and capturing of data

Need for a global substance database

- Knowledge of substances is essential for understanding safety and efficacy of medicinal products
- Expensive to implement a system based on 11238 on a individual basis may prevent adoption of standard
- Development of a common system would foster rapid convergence of standards
- Global database means better data, less redundancy, less mapping
- Better coordination of regulatory activity and clinical trials (inspections, specifications, drug shortages)
- Enable pharmacovigilence based on global data
- Global marketplace for ingredients requires a global system to monitor the supply chains

Vision of 11238 Implementation

A single global registration system to identify Substances in Medicinal Products providing

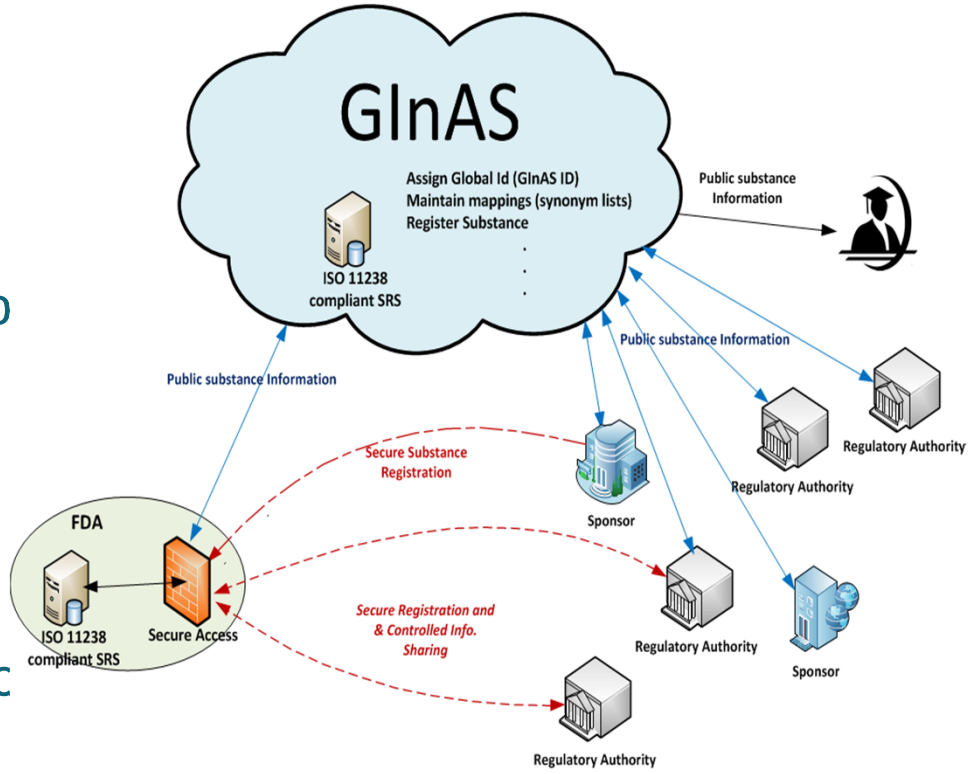
- A global ID for all classes of drug related substances
- A single place for registration and depositing identification, analytical and manufacturing information and relevant biological data
- A freely distributable tool or data system
- Common Messaging to communicate relevant substance information

GInAS Project Goals

- Develop and deploy an information system that can serve as a global repository for definitional, regulatory and scientific information on substances
- Establish a consortium of regulators and other international organizations to manage and govern the repository
- Distribute an information system to both regulators, companies and other interested parties to facilitate registration into the global repository
- **Common System and Common Content results in a Better System and Better Content**

GInAS Stakeholders

- All GInAS participants comply with ISO 11238 for Substance Identification
- Each Regulatory Authority can maintain its own Substance Registration (minimum impact to its regulatory process)
- Regulatory Authorities can establish agreements with each other for substance registration & maintenance if they choose
- GInAS provides a source of public data to stakeholders



GInAS Description

- A freely distributable software system that will implement the ISO 11238 standard and provide a global identifier for substances
- A system capable of registering diverse substances and specified substances
  - » **Chemicals, Proteins, Nucleic Acids, Polymers and Structurally Diverse**
- A central repository of substance definitions and identifiers
- A central repository of information related to substances:
  - » **Physical Properties** (Solubilities, Viscosity, Melting points, Isoelectric Points)
  - » **Specification Information** (Impurities, Properties, Assay)
  - » **Manufacturing Information** (Company, Facility, Starting, Processing, Final Materials, Critical Parameters)
  - » **Pharmacological & Toxicological Information** (Metabolites, LADMER data, Targets, NOAEL)
  - » **Regulatory Data** (Classifications, Authorizations)
- Capable of maintaining public data and confidential data at the element and record level (Roles, User, Depositor can control access)

GInAS Development

- Grew out of a meeting hosted by USP and NIH/NCATS in February 2013
- Will use software developed by NCATS
- Content managed by regulators
- Initial deployment hosted at Health Canada
- US, Canadian, Dutch, German, Swiss regulators and EDQM and USP involved in the development of the system
- Will contain definitional, analytical, manufacturing and biological information (target, metabolites, metabolic enzymes, and transporters)
- Software can also be deployed locally (each regulator can have their own independent system using NCATS software and public data)
- System will be distributed by NCATS with a large set of public domain data and updated periodically

National Center for Advancing Translational Sciences (NCATS)

- Newest IC of the National Institutes of Health
  - » To catalyze the generation of innovative methods and technologies that will enhance the development, testing and implementation of diagnostics and therapeutics across a wide range of human diseases and conditions.
- Highly collaborative across NIH, other government agencies, and with the private sector.
- *Develop*, *demonstrate*, and *disseminate* software for generating, managing, and mining data
- Develop systems in collaboration with domain experts instead of by contract

NCATS Mission

To catalyze the generation of innovative methods and technologies that will enhance the development, testing and implementation of diagnostics and therapeutics across a wide range of human diseases and conditions.

Characteristics of NCATS Initiatives and Programs

- Address significant bottlenecks in the process of translation
- Highly collaborative across NIH, other government agencies, and with the private sector.
- Quick to respond to needs of stakeholders/researchers
- *Develop*, *demonstrate*, and *disseminate* software for generating, managing, and mining data

Why we are working with NCATS

- Standard is Complex
  - No COTS Software Available
  - Unlikely a vendor could develop compliant usable software without a great deal continuous input from substance SMEs
- In-depth expertise; Mutual interest
  - Developers also have core business knowledge
  - Chemistry, Biotechnology
  - Common Interests
  - FDA has developed content that can serve NCATS mission
- NCATS can provide software and content to enhance the FDA system
  - No formal contracting or extensive requirement gathering needed
  - Much more cost-effective
  - Much greater chance of real success. No contract mods. Little bureaucracy
- Software developed can be freely distributed
  - No barrier to implementation

Development Strategy

- Core software developers with intimate domain knowledge and experience with regulatory process
- Iterative development cycle
  - » Functional requirements through in-depth case studies with users/experts
    http://tripod.nih.gov/ginas/data.html
  - » Discussion and feedback from case studies drive design and development
  - » Evaluate implemented features with users through usability studies
  - » ISO 11238 core standard; detailed system description/requirements necessitate a working implementation

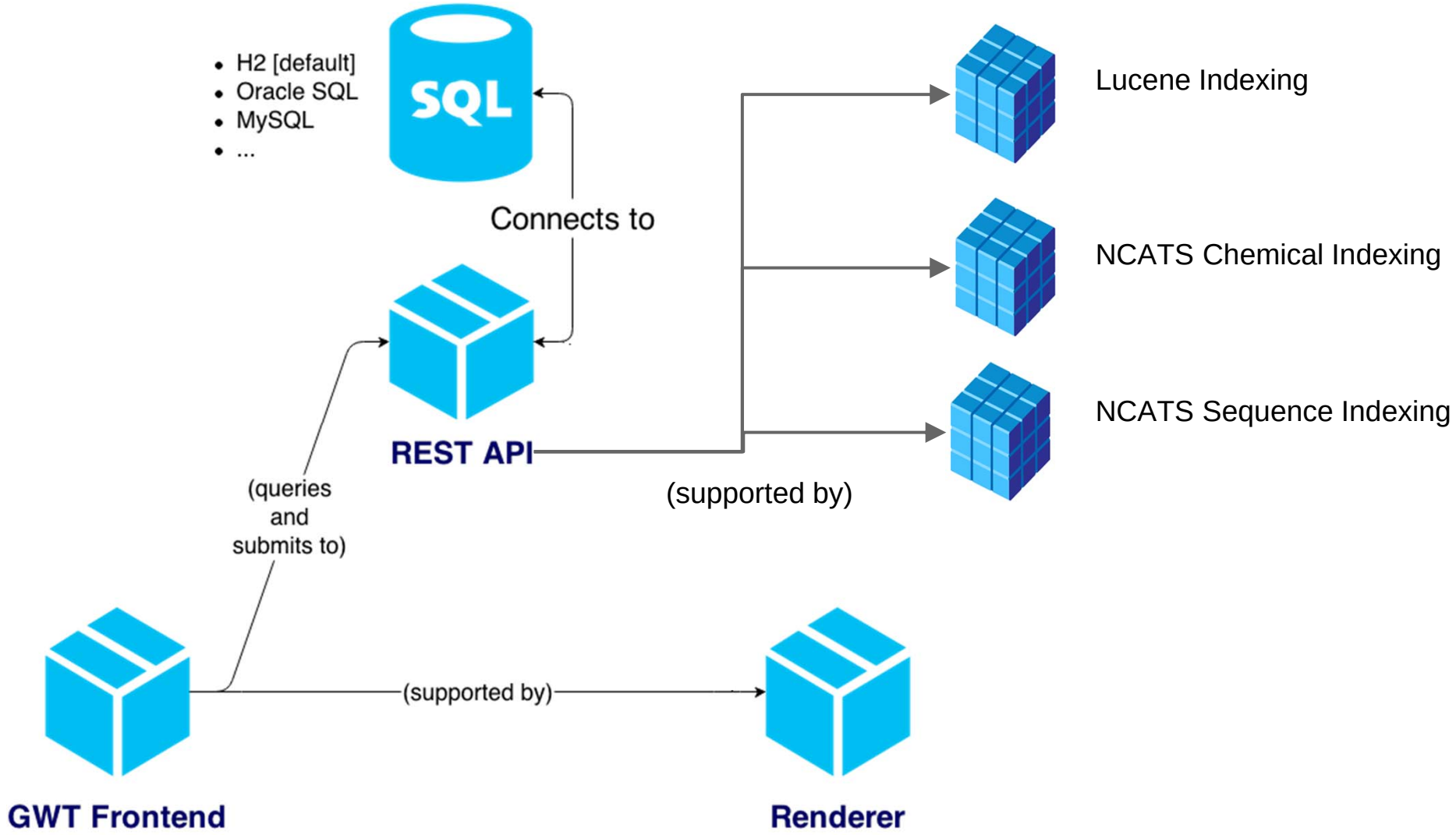
Implementation Goals for Software

- Self-contained and modular
  - » Run entirely on a desktop or access remotely
  - » Freely distributable, predominantly Open Source
- Well-defined data access application programming interface (API)
  - » GInAS or third-party clients / internal business processes
- Fine-grained access control model
  - » Access control for every piece of information
  - » Audit trail of all data fields
- Configurable “business rules”, e.g. standardizing structures
- All data is referenced; primary sources should be retained (e.g., PDF’s, MS spectra, images)
- System will be distributed with a large set of public domain data and updated periodically

Architecture Overview

- Web-based client over RESTful interface
- RESTful API as the lowest common layer
  - » Data models and concepts based on ISO 11238
  - » Support legacy/existing database infrastructure via mapping to common data models
  - » JPA/JDO as data persistence layer for standalone instances; i.e., database vendor neutral.
- JSON as the default data exchange format; MDL's SDF as the molecular format.
- Deployable in J2EE (version 5+) container
- Fine-grained access control
  - » Resource (API), domain (jurisdiction), role (e.g., analyst, administrator, etc.), field- and relationship-visibility (e.g., is the association between a particular name and structure public knowledge?)
- Backend database agnostic (e.g., Oracle, MySQL)
- Structural handling of documents
- Where possible open source is preferable

Architecture Overview



Architecture Overview

- Client-server
- Backend database agnostic (e.g., Oracle, MySQL)
- Standalone server or deployable within a standard web container (e.g., Glassfish, Tomcat)
- Pluggable engines for text, structure, and sequence searching

Reference Substance Data

- A reference substance database is distributed with each software deployment
  - » Initial bootstrap from FDA's public SRS and NCATS data
- Instances can participate in one-way (pull) or two-way (pull/push) updates with each other
  - » Data curation
  - » Conflict resolution