Breakout Session 3: Track B

Public Substance Registration Using the Global Substance Registration System (GSRS)

Dr. Alexander Welsch
Contractor Programmer/Data Manager, NCATS/IFX (Axle)

Public Substance Registration Using the Global Substance Registration System (GSRS)

Development Update

January 2024

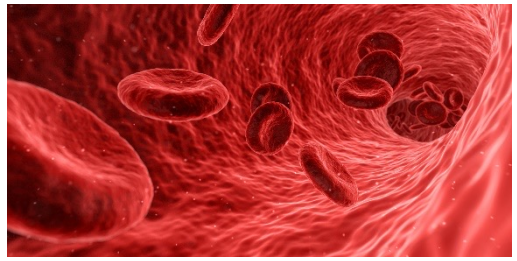
Alex Welsch, NCATS IFX

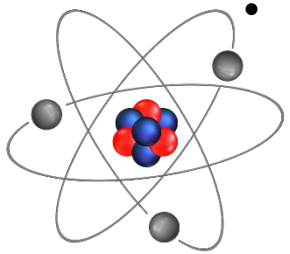
Agenda

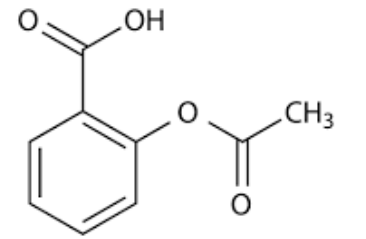
- High-level Overview GSRS
- Introduction to SubstanceReg
- Current State of the STRIDES Initiative
  - Achievements
  - Best Practices
  - Lessons Learned

What is GSRS?

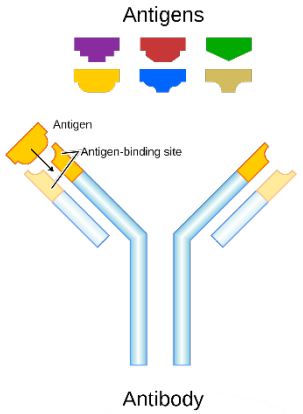
GSRS is an open-source application and database for registering and curating **substance** based on their scientific definitions







Acetylsalicylic acid



What it has:

- 150,000+ substance records
  - Active Ingredients
  - Inactive Ingredients
  - Metabolites
  - …
- Small molecules, polymers, biopolymers, plant parts, tissue parts, vaccines, etc.
- Curated information
  - chemical structures
  - substance *names*
  - database identifiers
  - protein and nucleic acid sequences
  - taxonomic information
- **Unique Ingredient Identifiers (UNIIs)**

Software:

Backend -
Java, Spring Boot

Frontend -
Angular

Global Substance Registration System

- *Collaborating Internationally to define substances at the molecular level that are used in regulated products providing highly curated substance Information globally*

- *Government off-the-shelf software developed by FDA/NIH/NCATS in collaboration*

Core Software

GitHub GSRS SOFTWARE

GSRS Instances/Databases

200,000+ substances

FDA GSRS
GSRS.FDA.GOV

(internal to FDA)
Links to FDA internal systems

Merge New/Curated Substances Into FDA

Public Data Export
150,000+ substances

NIH NCATS GSRS
GSRS.NCATS.NIH.GOV

FDA GSRS
precision.fda.gov

NIH NCATS GSRS
substancereg-dev.ncats.io

read only

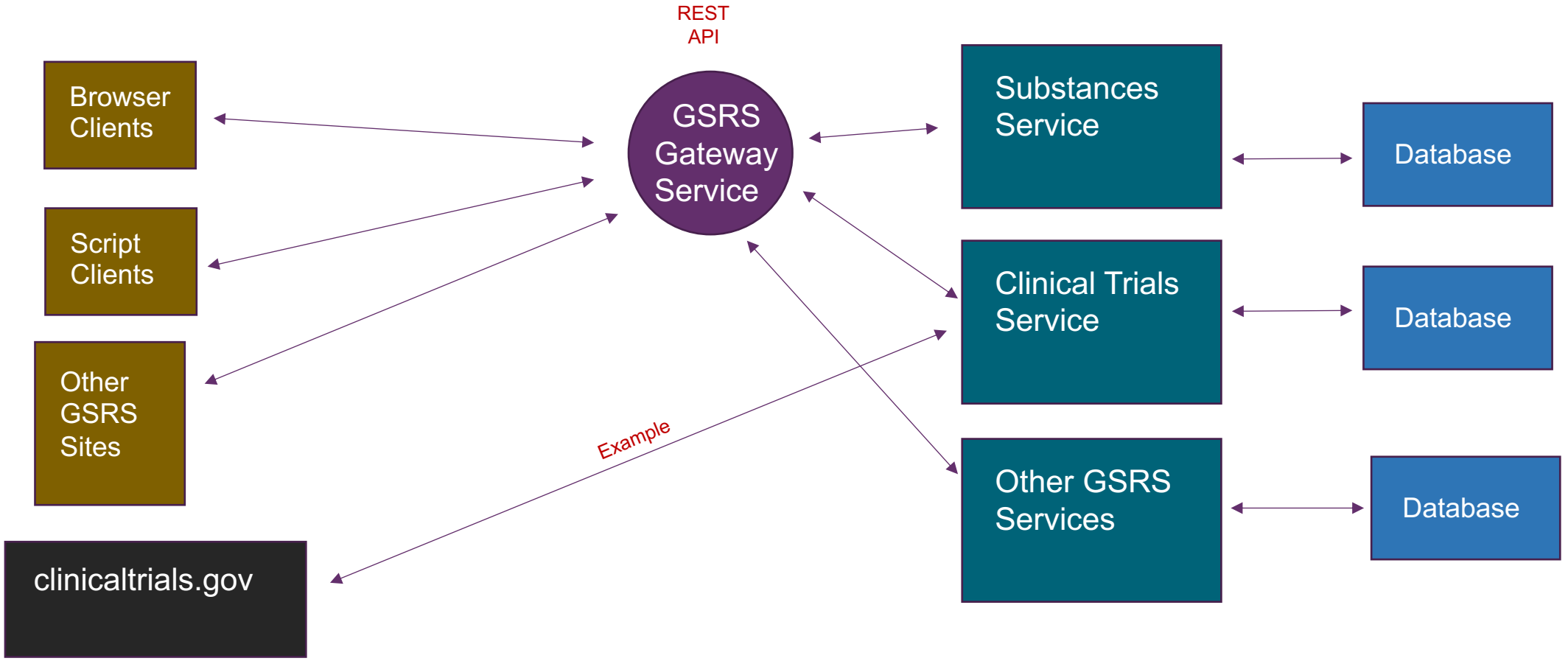
read/write

read/write

View Register Curate

- Academia
- Journals
- Industry

Microservices in GSRS: Modular Network of Applications



Modular Coding Approach to Microservices in GSRS, Example

Make Clinical Trials microservice (Executable via Tomcat)

- Make/Import Clinical Trial Starter modules
  - (imports GSRS Spring Starter modules)
    - Spring Boot Framework

      Jackson (serialization)
      Hibernate (database)
      Lucene (indexing)
      REST Template

    - GSRS base packages

      Users
      Search
      Indexing
      Exports
      Validation
      Event handling

Main Goals of Initial Award

- Development of a robust user management and installation of GSRS on the cloud for registration.
- Work out processes for batch registration of substances into the GSRS from other NIH systems (e.g. ChemId) and Academic Partners.
- Migration of FDA Product Data into the public cloud instance of GSRS.
- Migration and linking of ClinicalTrials.Gov data into cloud instance.
- Migration of EU clinical trial registry data (taken from https://www.clinicaltrialsregister.eu) into cloud instance.
- Migration of Public Adverse Event Data from FDA's FAERS system.

Achievements: SubstanceReg-dev

Global Substance Registration System - GSRS

The main goal of ginas is the production of software, called G-SRS, to assist agencies in registering and documenting information about substances found in medicines. The Global Ingredient Archival System provides a common identifier for all of the substances used in medicinal products, utilizing a consistent definition of substances globally, including active substances under clinical investigation, consistent with the ISO 11238 standard.

Search Substances

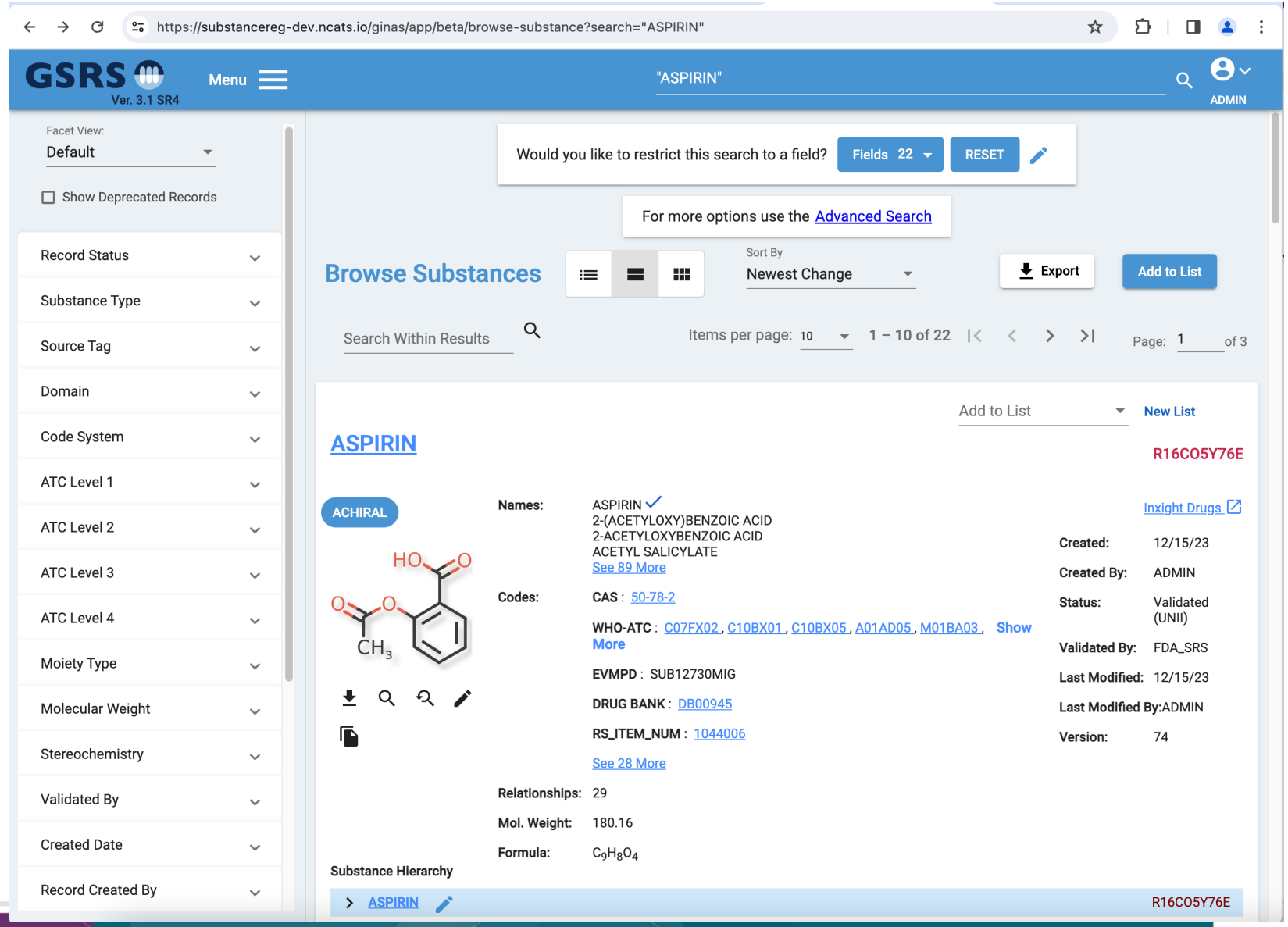
Browse Substances | Structure Search | Sequence Search | Bulk Search

Total substances: 156,829

Chemicals	111,019	Polymers	2,455	Structurally Diverse	27,047
Proteins	6,840	Nucleic Acids	532	Concepts	5,871

https://substancereg-dev.ncats.io/ginas/app/beta

Achievements: SubstanceReg-dev



https://substancereg-dev.ncats.io/ginas/app/beta

Achievements: Data Curation on SubstanceReg-dev

SubstanceReg-dev provided a web-based home for FDA interns and partners.

Previously these partners required a government computer for data curation on FDA systems.

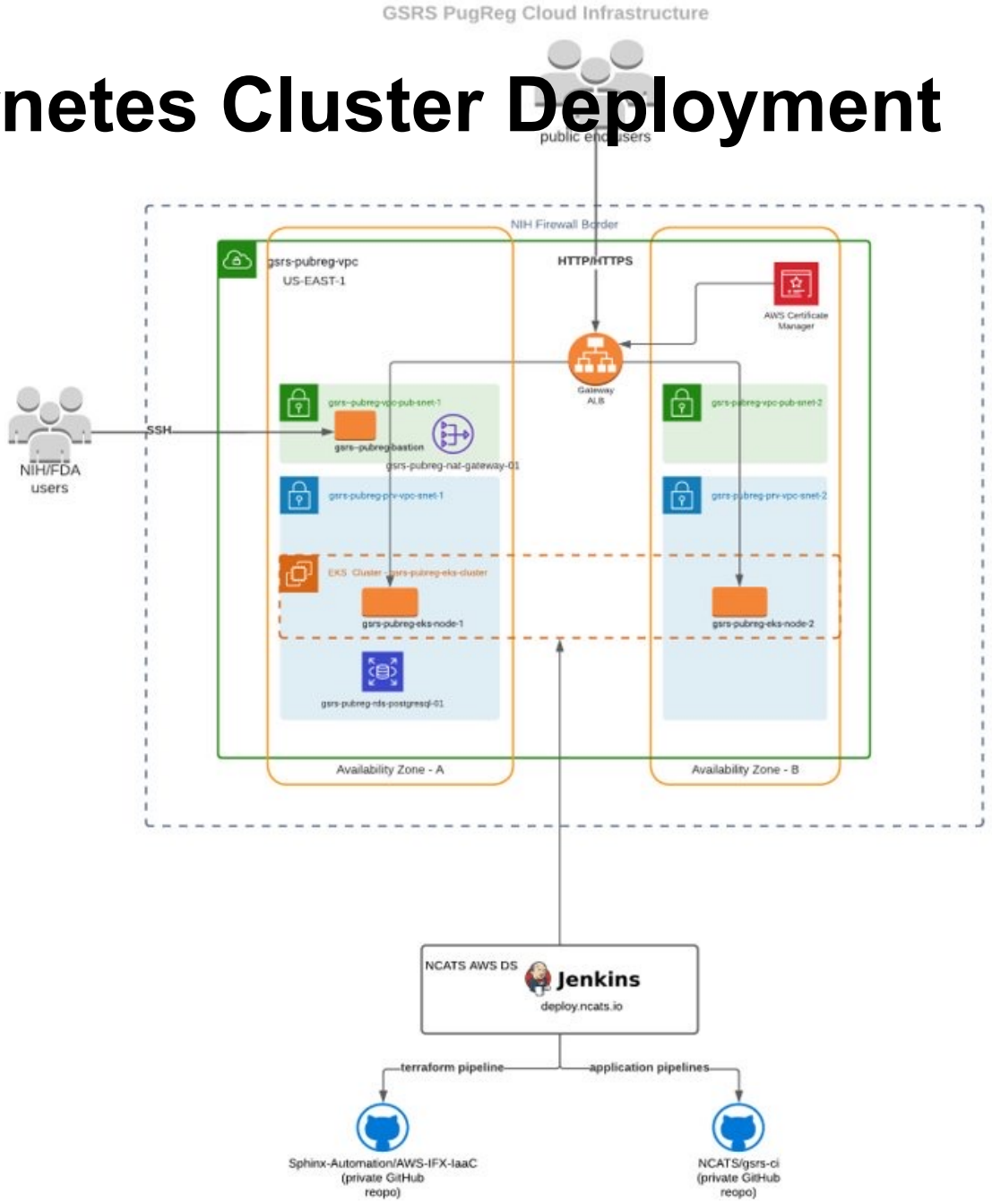
In 2023, seven non-FDA collaborators (6 interns and 1 contractor) created or edited 2267 substances.

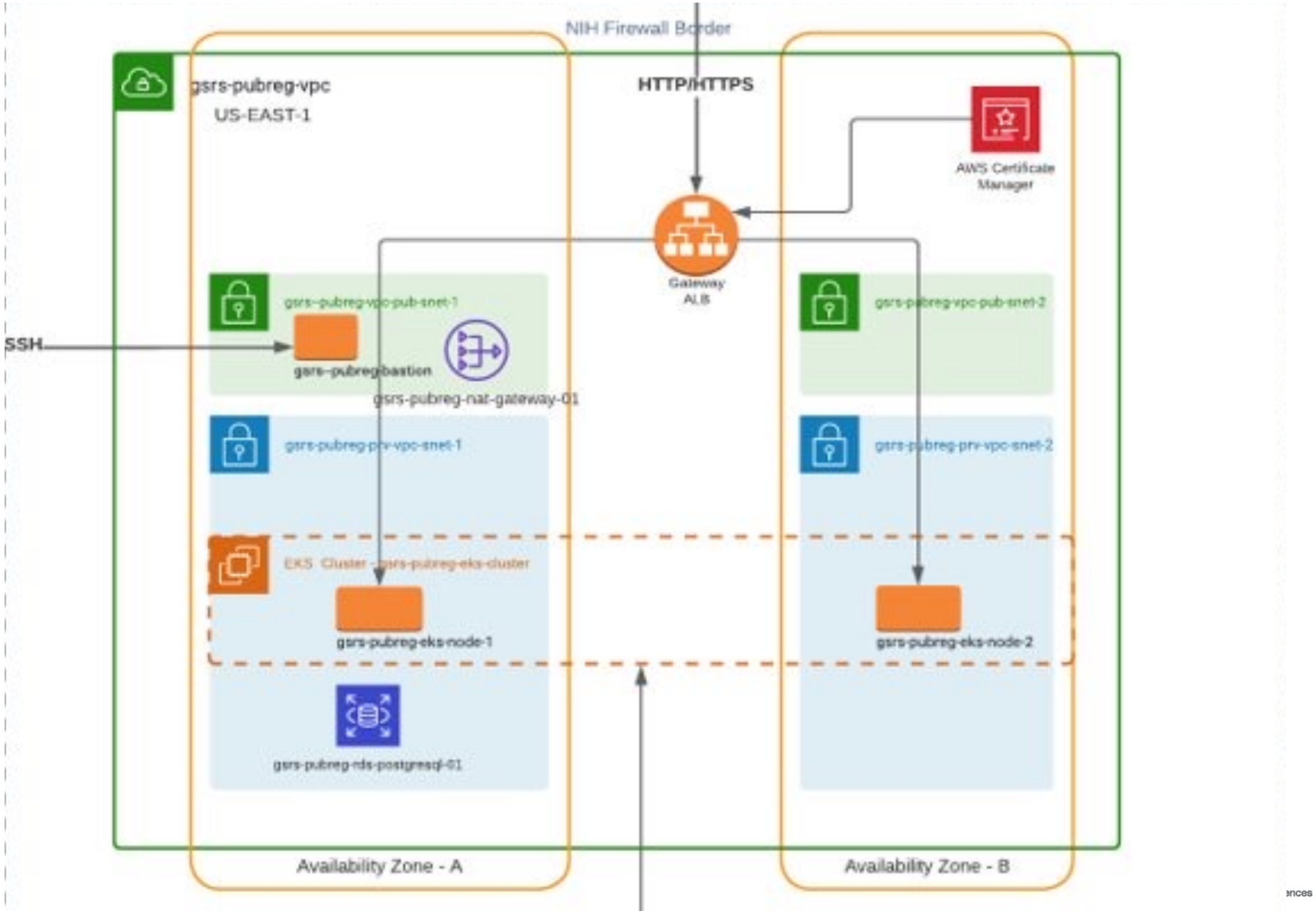
Curation was performed with less overhead and security steps.

Achievements: Test Kubernetes Cluster Deployment

The Test SubstanceReg deployment consists of a Kubernetes cluster. There are “pods” for the Gateway, Frontend and Substances services.

The cluster includes a data volume and a database.




NIH Firewall Border
gsrs-pubreg-vpc
US-EAST-1
HTTP/HTTPS
AWS Certificate Manager
Gateway ALB
SSH
gsrs-pubreg-vpc-pub-snet-1
gsrs-pubreg-bastion
gsrs-pubreg-nat-gateway-01
gsrs-pubreg-vpc-pub-snet-2
gsrs-pubreg-prv-vpc-snet-1
gsrs-pubreg-prv-vpc-snet-2
EKS Cluster - gsrs-pubreg-eks-cluster
gsrs-pubreg-eks-node-1
gsrs-pubreg-eks-node-2
gsrs-pubreg-rds-postgresql-01
Availability Zone - A
Availability Zone - B

Best Practices

- Network separation of concerns
- Automate pipelines
  - Spring boot / Maven
  - Helm Charts
  - Sphinx Automation
  - Kubernetes/Docker

Challenge to overcome

- Difficult to simulate Kubernetes network deployment locally for development and QA

Lessons Learned

- Cultivate good relationships with deployment staff.
- Collaborate on documentation as you go.
- Adopt organizational practices of deployment group/staff.
- Gain access to resources that allow for monitoring progress.
- Streamline data preparation in staging environments:
  - We avoided lengthy/repeat indexing tasks
  - Prevent downtime in production
- Use Git tags and version variables to ensure that applications use the right dependencies.

Thanks to:

FDA
Lawrence Callahan
Tyler Peryea
Frank Switzer
Elaine Johanson
Marlene Kim
Siba Bhattacharyya
Archana Newatia
Ramez Ghazzaoui
Arunasri Nishtala

NCATS/IFX
Lihui Hu
Dammika Amugoda
Mitchell Miller
Alex Welsch
Sarah Stemann
Meghan Mendick
Marian Nkeng
Niko Anderson
Kesandu Nwokolo
Ewy Mathé

NCATS/ITRB
Kanna Bhargav Chevva
Surya Robbi
Sridhar Vuyyuru
Ke Wang

BfArM
Egor Puzanov

USP
Andrzej Wilk
Steve Emrick
Jeff Shick

EMA
Herman Diederik
Panagiotis Telonis

WHO-UMC
Malin Fladvad
Olof Lagerlund

Get Involved:

- Email: ncatsgsrs@mail.nih.gov
- Signup for Newsletter: https://gsrs.ncats.nih.gov/#/
- Join Collaborator Slack: gsrscollaborator.slack.com
- View data on public site: https://gsrs.ncats.nih.gov/ginas/app/beta/
- Get the code: https://github.com/ncats/gsrs3-main-deployment
- View Swagger GSRS API doc: https://gsrs.ncats.nih.gov/#/api

- Stay tuned in next two months, and possibly start contributing to the substances database. The url will be: https://substancereg.ncats.nih.gov

NCATS

COLLABORATE. INNOVATE. ACCELERATE.

ncats.nih.gov @ncats_nih_gov @ncats.nih.gov NIH-NCATS

National Center for Advancing Translational Sciences

Questions