



PrecisionFDA

Transforming the Future of Regulatory
Science

FDA Disclaimer



The views and opinions presented here represent those of the speaker and should not be considered to represent advice or guidance on behalf of the U.S. Food and Drug Administration.

Connecting FDA with External Partners and Experts

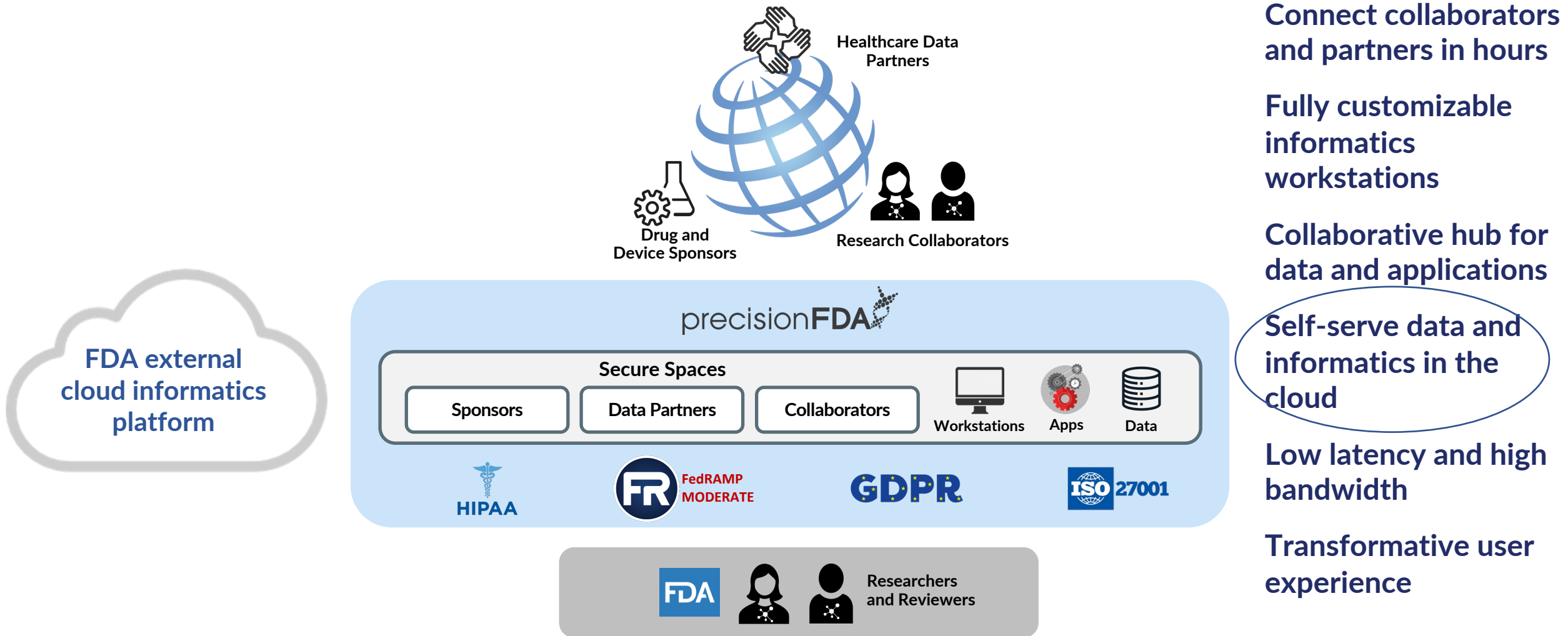


- Secure
- Cloud-Based
- Collaborative
- Multi-omics

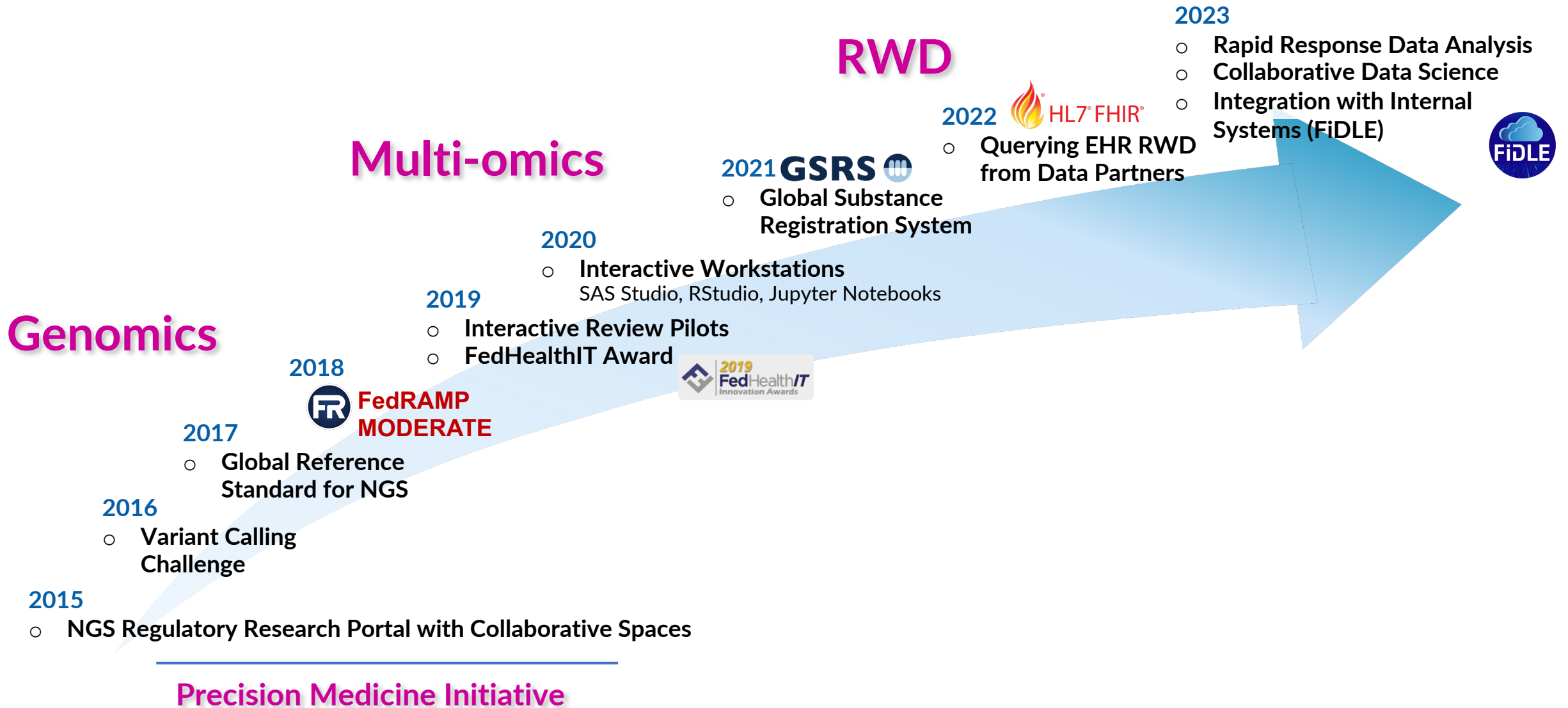
The screenshot displays the precisionFDA Overview page. At the top is a navigation bar with links: Overview, Discussions, Challenges, Experts, My Home, Notes, Comparisons, Spaces, GRS, DaaS, PRISM, and Tools. The main content area is divided into several sections:

- MOST RECENT APPS:** Lists four applications: Cas-OFFinder v3.0.0b3, Trimmomatic_adaptor, Cas-OFFinder Input File Generator, and Cas-OFFinder v2.4.1, each with an update timestamp.
- TOP FEATURED APPS:** Lists three applications: Remote Linux Desktop Through Apache Guacamole, pfa-ttyd, and DNAnexus JupyterLab Server on precisionFDA, each with an update timestamp.
- precisionFDA Challenges:** A section with a banner and a "View All Challenges" link.
- CURRENT CHALLENGES:** Features a "precisionFDA Automated Machine Learning (AutoML) App-a-thon: Democratizing and Demystifying Artificial Intelligence (AI)" challenge, including its dates, duration, and a "View Details" button.
- EXPERT HIGHLIGHT:** Promotes an expert blog post titled "Understanding Risk and Protective Factors in the Veteran Population" by Vinay Pal, PhD, dated Mar 06, 2024, with a "Read Expert Blog Post" link.
- GETTING STARTED:** A sidebar section with links for "Introduction and Next Steps", "Use Case Toolbox", and "Multi-omics App Library", along with a "precisionFDA Mailing List" link.
- LATEST NEWS:** A sidebar section listing recent news items with "View" links.

Providing Secure Computing and Advanced Analytics



Evolving over Time



Engaging Users from Around the World



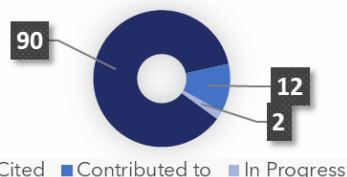
26,584
News subscribers

110 TB
Data stored on the site

3
Leadership Awards

Challenges:
precisionFDA Automated Machine Learning (AutoML) App-a-thon: Democratizing and Demystifying AI closed on April 26th. Results for the App-a-thon will be posted soon.

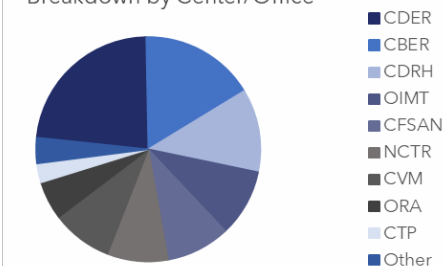
Peer-reviewed Journal Publications



2,887
729,000
■ precisionFDA Twitter ■ FDA LinkedIn

FDA Users

Breakdown by Center/Office



CY24 Goal Progress

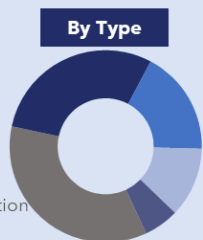


*Meetings occur weekly

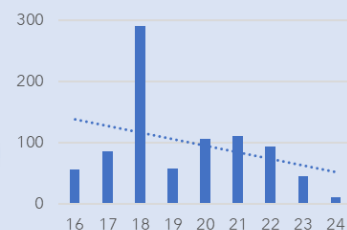
Challenge Participation

859
Total submissions

■ Bioinformatics
■ App/Algorithm Dev
■ AI/ML
■ Data Exploration
■ Workflow Documentation



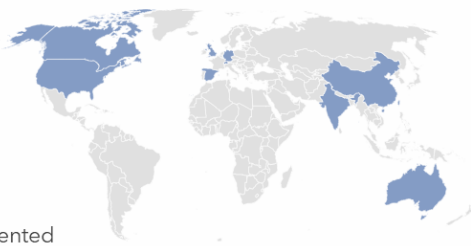
By Year



User Reach

7,776
Users worldwide

54
Countries represented

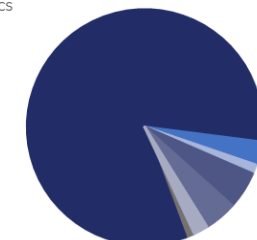


Powered by Bing
© Australian Bureau of Statistics, GeoNames, Microsoft, Navinfo, Open Places, OpenStreetMap, TomTom, Zenrin

716

Public Tools & Apps

■ Chemoinformatics
■ Genomics
■ Metagenomics
■ Proteomics
■ Virology
■ Visualization
■ Workstation
Data Analytics*
Regulatory*



Analysis ongoing*

What's New with precisionFDA:

Space Report functionality has been updated to include Space members and all Discussions within the Space, documenting discussion threads and replies. Additionally, pre-filled app run URLs have been added, allowing users to generate a "one-click" shareable URL that auto-selects inputs for an app or Workstation.

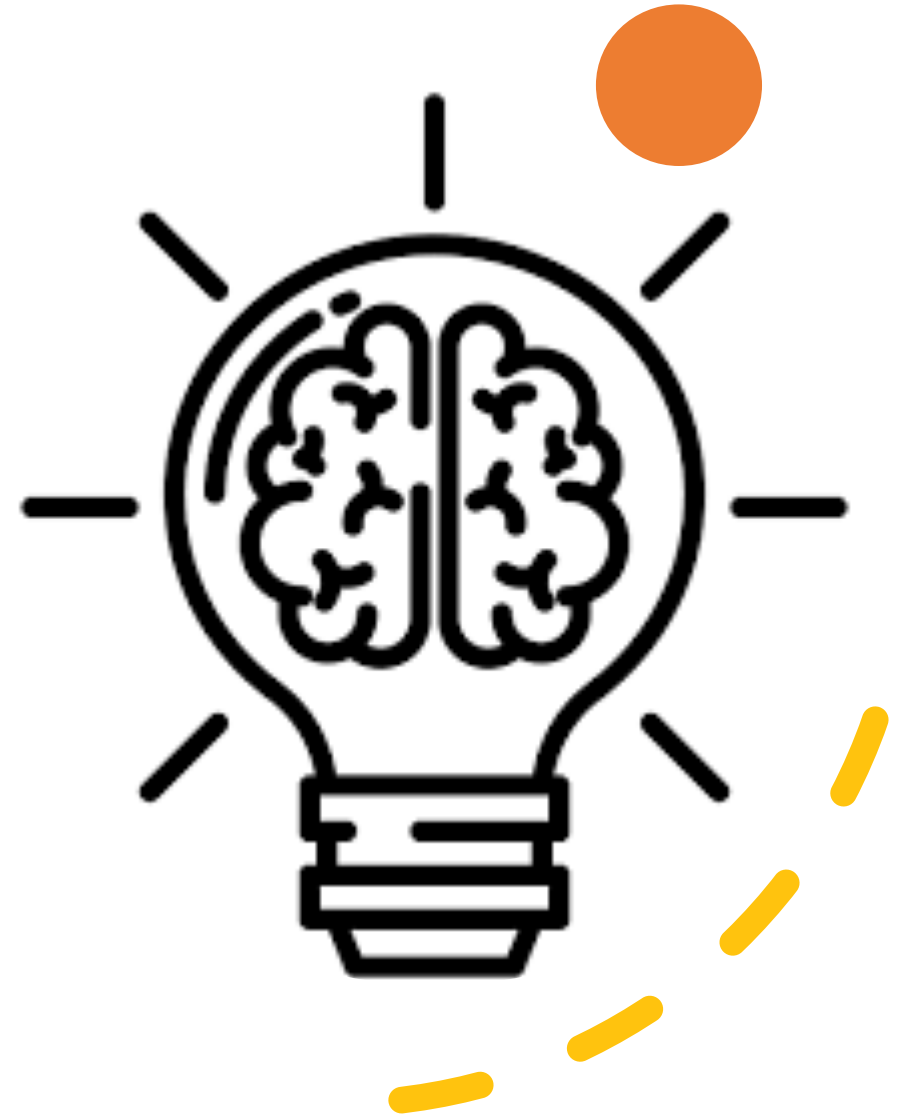
precisionFDA

A secure, collaborative, high-performance computing platform that builds a community of experts around the analysis of biological datasets in order to advance precision medicine.

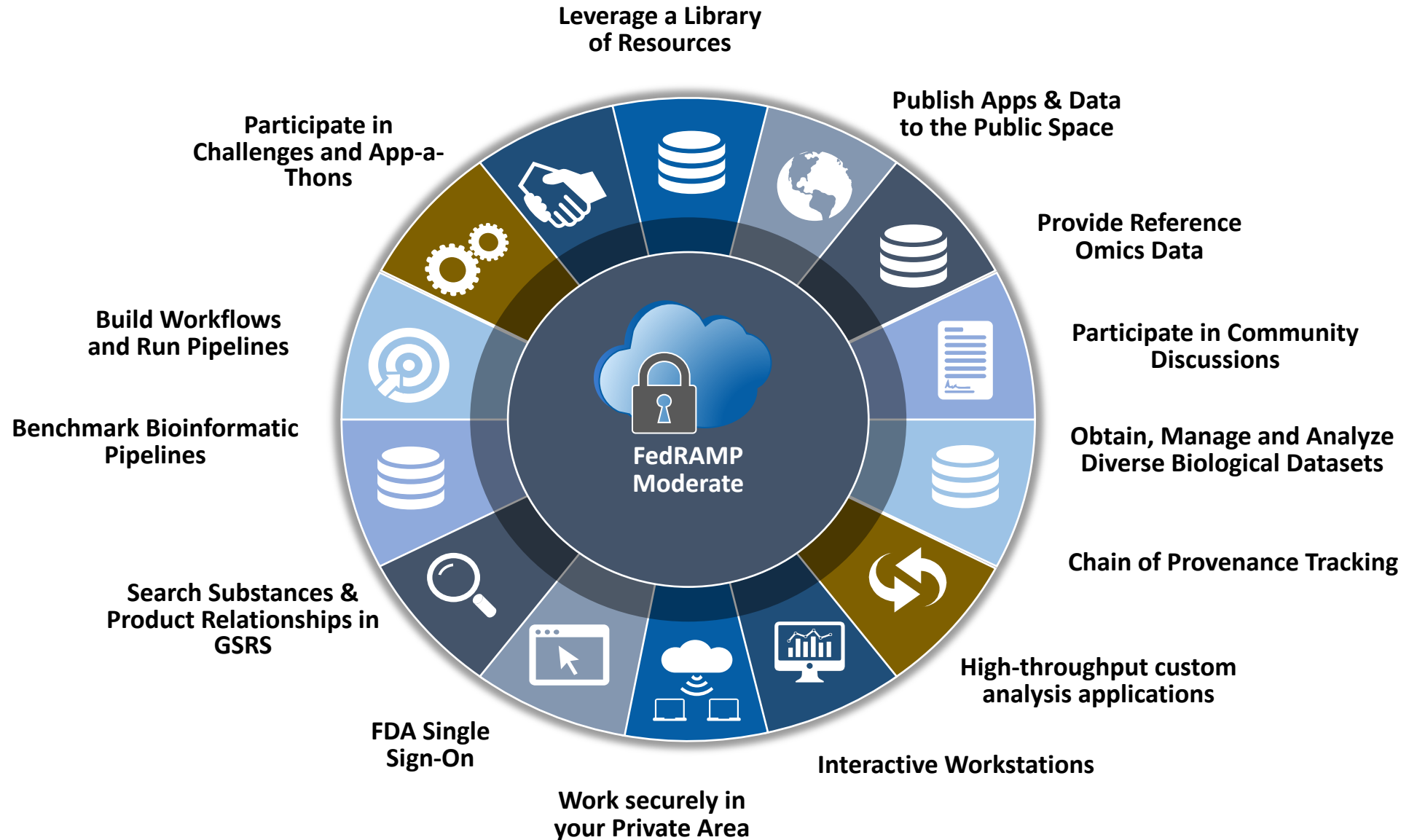
Encouraging Innovation

PrecisionFDA provides a unique environment for encouraging innovation, including:

- **Scalable, on-demand resources** – Providing the platform to manage cloud resources on behalf of users while still providing visibility in cloud usage.
- **Reusable Components** – These are custom-designed tools that are available for all precisionFDA program initiatives and data portals that make it possible to reduce duplication and improve efficiency in development.
- **User Empowerment** - Designed with self service in mind but flexible enough to respond to a wide-range of needs.
- **Collaboration and Community Development** – Opportunities and tools to enable interaction and collaboration between FDA scientists and regulators as well as external partners and experts.



Offering Core Capabilities



Creating Efficiencies through Reusable Components

- **Spaces** – Private, Group, and Public spaces form the foundation for interactive and collaborative activities.

Private Spaces

This space is for an individual user.
Only that user can see what is in their private space.

Shared/Project Spaces

FDA personnel may request a shared space in order to work with one or more other users.

The Public Space

Items placed in the public space are available for all users to see and use.

- **Data Portals** – The ability to quickly and easily build a sub portal within precisionFDA that offers a collection of reusable features with a custom interface for a specific business purpose. Current data portals include the Challenge Framework, Data Analytics as a Service, and PRISM, etc.

The screenshot shows the precisionFDA interface. The top navigation bar includes links for Back Home, Spaces, DAaaS, PRISM, and Tools. The main content area is divided into two sections: 'Reports and Dashboards for Data Consumers' and 'Resources for the DAaaS Team'.

Reports and Dashboards for Data Consumers	
 Infant Formula Pulse Monitor Project	The IF Pulse Monitor reports presents sales and production data of Infant formula across US by state, type of formula and manufacturer.
 Pediatric Analgesics Project	The Pediatric Analgesics Project presents sales and production data of Adult and Pediatric Analgesics across the US.
 Milk Monitoring Project	The Milk Monitoring Project presents sales and production data of Milk across the US.

Resources for the DAaaS Team	
 Team Leads Portal	All the resources required for a data analysis project Team Lead to build and outfit a Data Squad with the data and tools they need to create a project-specific Data Consumer Portal and to publish reports and dashboards to external stakeholders.
 Data Squad Portal	Instructions and hands-on examples for: - Finding and accessing project-specific and general purpose datasets - FIDLE- and precisionFDA-based tools for ETL, cleaning, analysis, and report and dashboard development, and publication

Creating Efficiencies through Reusable Components



- **Workstations** – Custom on-demand sandbox environments that inherit precisionFDA's security and access controls and offer access to specific tools and/or command line interface.
- **User Guided App Development** – Ability to create and manage scripts and deploy Docker containers for running user defined external third-party tools or create their own custom programs to support multi-omics.
- **Complex, Multi-stage Workflows** – Integrating multiple apps to create end-to-end pipeline for multi-omics data analysis.
- **Data and Files Management and Sharing** – A graphical user interface (gui) and application program interface (API) for managing data and files.
- **Databases** – Multi-tenant, persistent structured data storage with access control.
- **Open-Source Tools Library** – A collection of > 100 open-source tools for use by stakeholders such as: KNIME, Jupyter Notebooks, caMicroscope, Somatic and Germline Variant Calling Programs, QA/QC programs, etc.
- **FHIR Stations** – Using the FHIR server on precisionFDA, FHIR Stations uses workstations to lower the barrier of entry into using and learning about HL7 FHIR.



Supporting Program Initiatives and Data Portals

- **Public-Facing Global Substance Registration System (GSRS)** – Enables Industry to Register Unique Ingredient Identifiers (UNII) and access highly defined substance definitions.
- **Common Data Model Harmonization (CDMH)** – Enables regulators to query electronic health records from data partners, regardless of Electronic Health Record format, to support pre- and post market activities.
- **Data Analytics as a Service (DAaaS)** - Provides a dynamic and customizable interface, collaboration spaces, and tools for the DAaaS team to work with external stakeholders (White House, CDC, etc.), collect, prepare and store data, conduct analyses, and run complex reports (e.g., infant formula, analgesics, etc.).
- **PrecisionFDA Regulatory Information Service Module (PRISM)** – Supports a Research Collaboration Agreement (RCA) intended to inform the future of regulatory review in the cloud by providing a customized interface and tools for the Center for Biologics Evaluation and Research (CBER), Center for Drug Evaluation and Research (CDER), Office of Digital Transformation (ODT), and Industry RCA participants.
- **Challenge/Competition Program Portal** - Provides a challenge program and framework for engaging experts to inform regulatory and evolving science.
- **Data/Multi-omics Analytics** - Self-serve tools and information for bioinformaticians and data scientists to enable pipeline development and data analytics.
- **Interactive Regulatory Review** – Secure capabilities and infrastructure for FDA reviewers interacting with industry on submissions with biomedical data sets.

Advancing Biomedical and Regulatory Science via Crowdsourcing

Crowdsourcing Highlights:

- Launched 40 challenges
- Received over 800 total challenge submissions
- Contributed to 12 peer-reviewed journal publications
- Cited by over 90 peer-reviewed journal publications

PrecisionFDA has organized crowdsourcing challenges in which participants **leveraged artificial intelligence (AI) to advance biomedical and regulatory science**, e.g.:

Truth Challenge V2

- **Focus:** Assessed the state-of-the-art of variant calling by benchmarking computational pipeline accuracy on a variety of genomic regions using long and short read sequencing
- **AI Relevant Findings:** Innovative machine learning (ML) and deep learning-based methods performed particularly well, often outperforming traditional statistical approaches on long-read datasets and when combining multiple sequencing technologies, though with some challenges of overfitting to model training data

Multi-omics Enabled Sample Mislabeling Correction Challenge

- **Focus:** Addressed a common issue in research by developing and evaluating computational approaches for identifying and correcting specimen mislabeling
- **AI Relevant Findings:**
 - ML approaches were used by top-performing teams
 - COSMO, an open-source method for identifying and correcting mislabeling, which combines approaches of top-performing teams, was developed in the post-challenge phase

COVID-19 Risk Factor Modeling Challenge

- **Focus:** Enabled preventative medicine through the development and evaluation of AI/ML models for predicting COVID-19-related health outcomes in veterans
- **AI Relevant Findings:**
 - Synthetic health record data is applicable for modeling as a substitute for harder-to access real Electronic Health Record data
 - Models trained on synthetic data showed similar but slightly inflated performance to those trained on real de-identified data

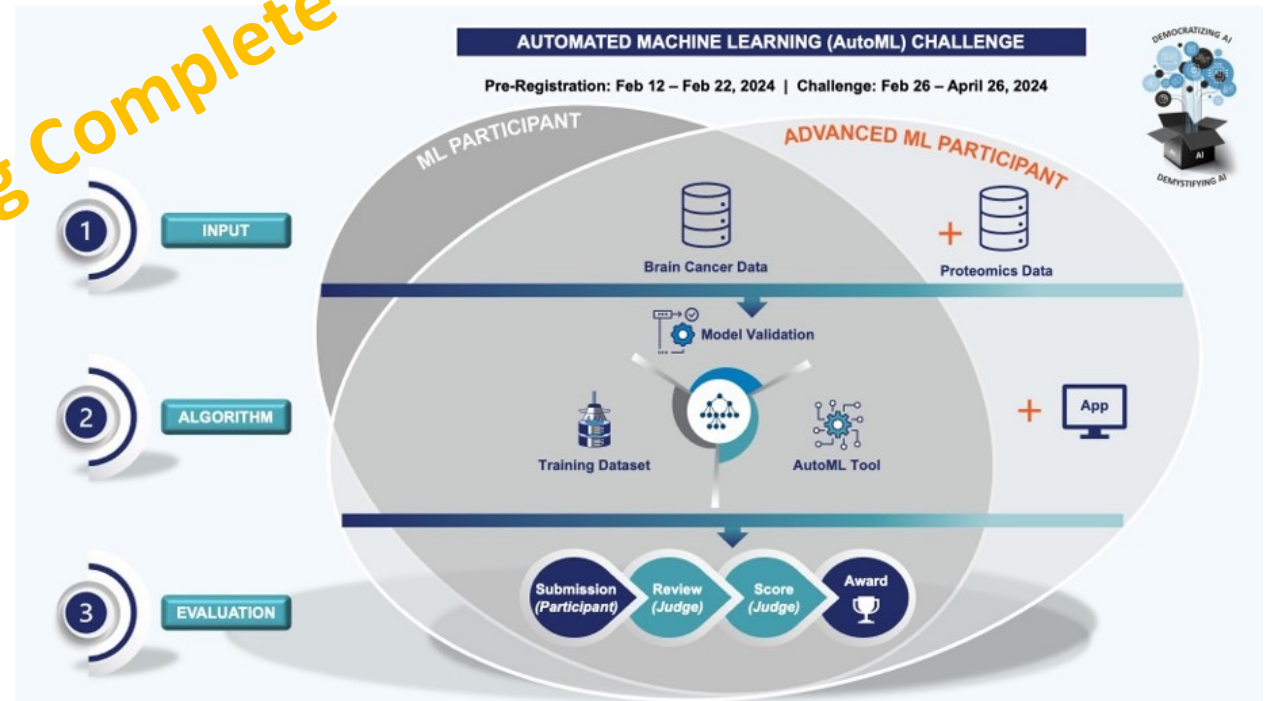
Democratizing and Demystifying Artificial Intelligence (AI)

PrecisionFDA is conducting a **Democratizing and Demystifying AI challenge series** to explore and illustrate how AI can support FDA's mission.

- The series aims to **empower users** to leverage the potential benefits of AI
- Challenges within the series will **consider applications of automated machine learning (AutoML) and generative AI**
- By exploring AI through this series, precisionFDA challenges will help **gather best practices, consider potential applications for FDA, and advance modernization initiatives**

The **precisionFDA AutoML App-a-thon** invited the scientific and analytic communities to use AutoML tools on biomedical datasets to help assess whether AutoML can match or improve the performance of previously developed traditional ML models. In traditional ML workflows, data scientists must have expertise in statistics, feature engineering, data preprocessing, and model selection. With AutoML, these processes are automated, which can potentially reduce time and costs.

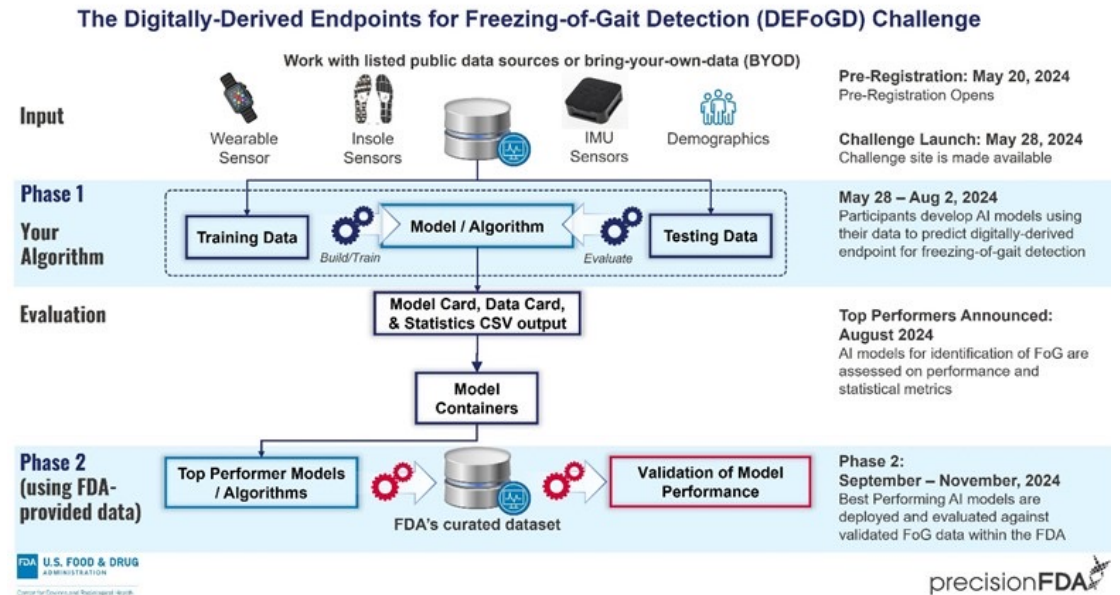
Judging Complete



Developing AI Models to Gain Insights from Digital Health Technology Data

Current Challenge: Digitally-Derived Endpoints for Freezing-of-Gait Detection (DEFoGD)

To better understand the ability of DHTs to become digitally-derived endpoints, **FDA's Digital Health Center of Excellence (DHCOE), the Office of Science and Engineering Laboratories (OSEL), and the Office of Digital Transformation's (ODT's) precisionFDA** are calling on the public to develop artificial intelligence/machine learning (AI/ML) models to identify and predict FoG in Parkinson's patients, from pre-existing datasets, with the models then being validated against an internal FDA dataset. This Challenge can help FDA determine whether the benefits of DHTs can be translated into the healthcare ecosystem and help gather best practices, consider potential applications for FDA, and advance modernization initiatives.



Live!

Continuing to Expand in 2024



Build and Implement PRISM Portal (complete)



Conduct PRISM Initial Use Cases (Regulatory Validation, M11)



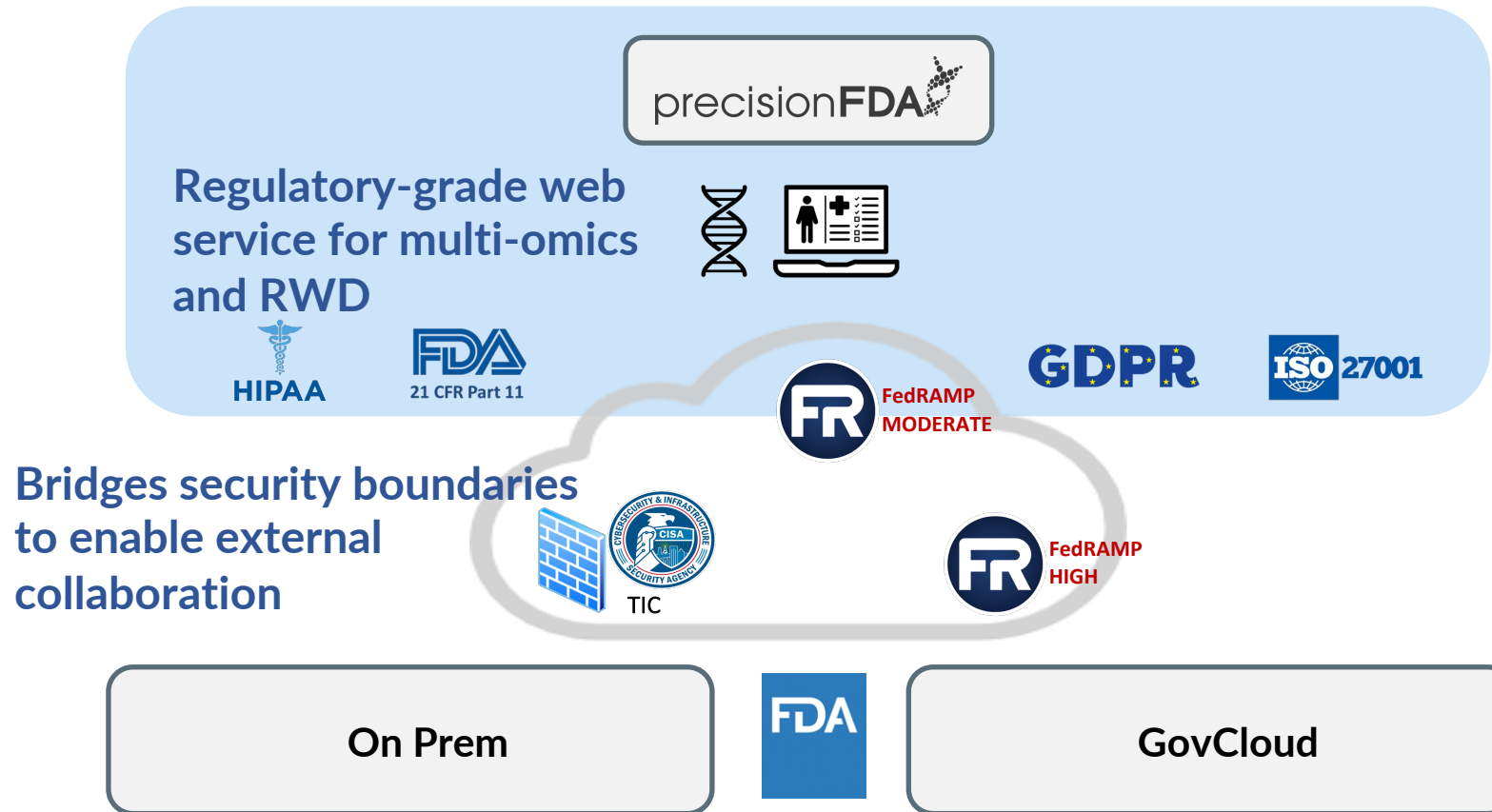
Implement FHIR stations to enable more extensive use of HL7 FHIR



Provide ability to use precisionFDA to explore generative AI large language (complete)



Coming in 2025...



- Conduct PRISM Phase II Use Cases (RWD, Next Gen eCTD integration)
- Implement FISMA High Gov Cloud Environment



Thank you!

For help with precisionFDA, please email precisionFDA@FDA.HHS.gov