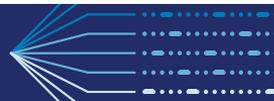# DataForward:

Applied Learning Tracks To Drive FDA Data Science

Raju (Rama) Rayavarapu Ph.D. *Data Science Lead, Talent Working Group Lead*

Ian Edhlund, Ph.D. *Toxicologist, Center for Veterinary Medicines*

- December 2023

Upskilling Data Talent at the FDA

Centers & Offices

Data*Forward* Tracks
Applied Learning Cohorts in Data Topic Tracks

ALT: Beginner

Squad: Intermediate

Data*Forward* Squads
Skilled Data Teams for Critical Agency Needs

Staff are trained and badged in data skill areas

Already badged staff is provided to help Centers achieve critical needs

HIGH VALUE TO CENTERS
Supports Centers' everyday and urgent data needs

FASTER SOLUTIONING
Readily available, internal, highly skilled talent

EFFECTIVE OUTCOMES
Saving time and money with more robust solutions

EQUAL PLAYING FIELD
Access to talent and more opportunities for skilled talent no matter which center

ENDORSED SKILL DEVELOPMENT
Standardized and calibrated learning programs to create a foundational skill level capable of tackling FDA challenges
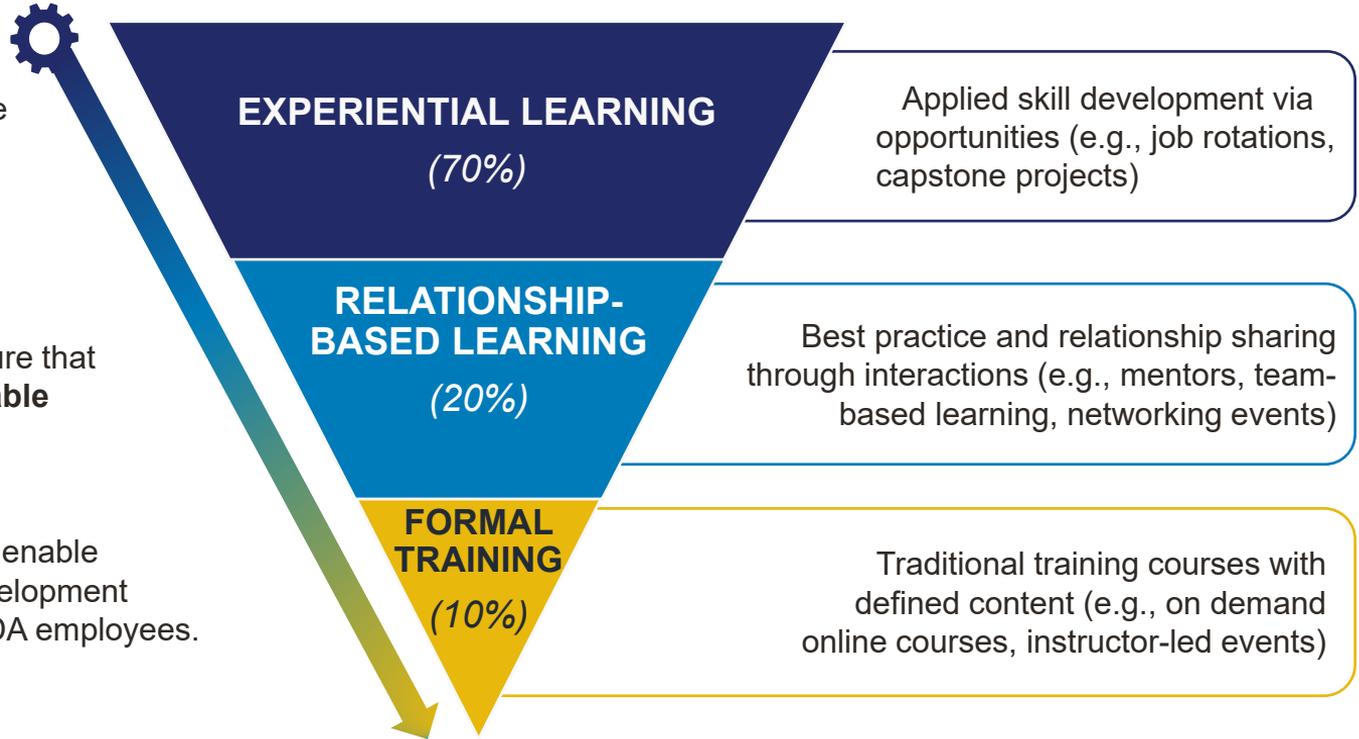
The 70-20-10 Model

Based 2022 trends, data scientists at FDA will utilize the **70-20-10 model** as a baseline framework for FDA's data science community.

This model allows for **unique learning experiences** to ensure that skills are **diverse and applicable** for mission-critical work.

When applied at FDA, this will enable learning opportunities and development **throughout the careers** of FDA employees.

EXPERIENTIAL LEARNING
(70%)

Applied skill development via opportunities (e.g., job rotations, capstone projects)

RELATIONSHIP-BASED LEARNING
(20%)

Best practice and relationship sharing through interactions (e.g., mentors, team-based learning, networking events)

FORMAL TRAINING
(10%)

Traditional training courses with defined content (e.g., on demand online courses, instructor-led events)

Roles and Responsibilities

Track Team Member

Collaborative self-organizing agile team.

Availability: During project working hours

Required Meetings: Weekly Scrum, Sprint Reviews, and Sprint Retrospective

Data Science Mentors

Primary resource for team for guidance and technical questions. Will filter any business questions to product owner.

Availability: During the project working hours and outside of the working hours as their availability allows
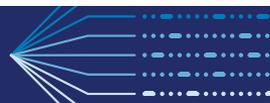
Required Meetings: Sprint Reviews

Product Owner

Resource to define desired business outcomes, provide product vision, and will provide review each sprint.
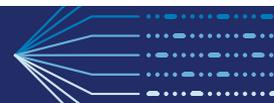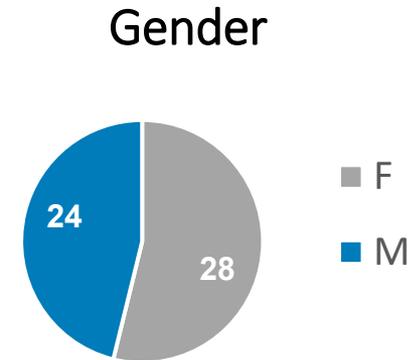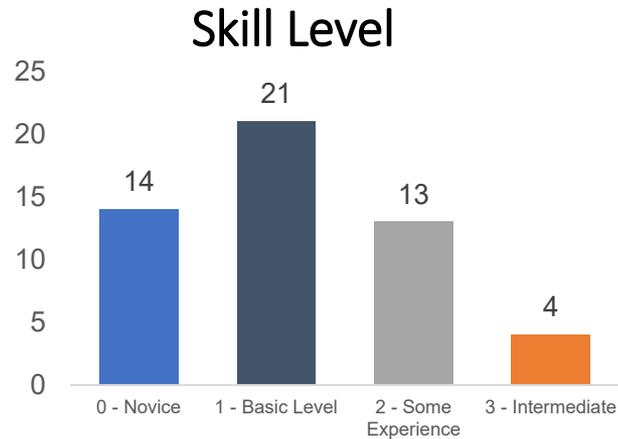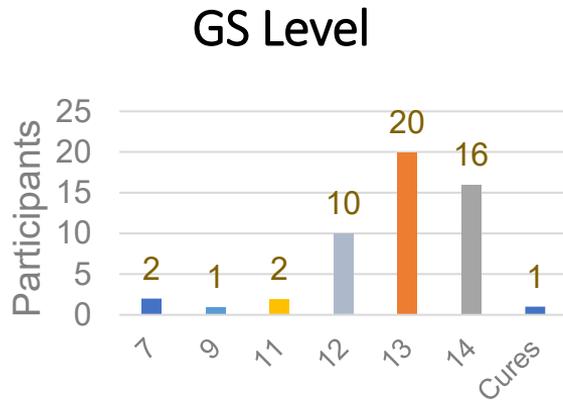
Availability: Limited outside defined meetings

Required Meetings: Sprint Reviews

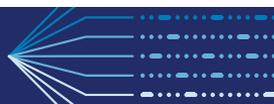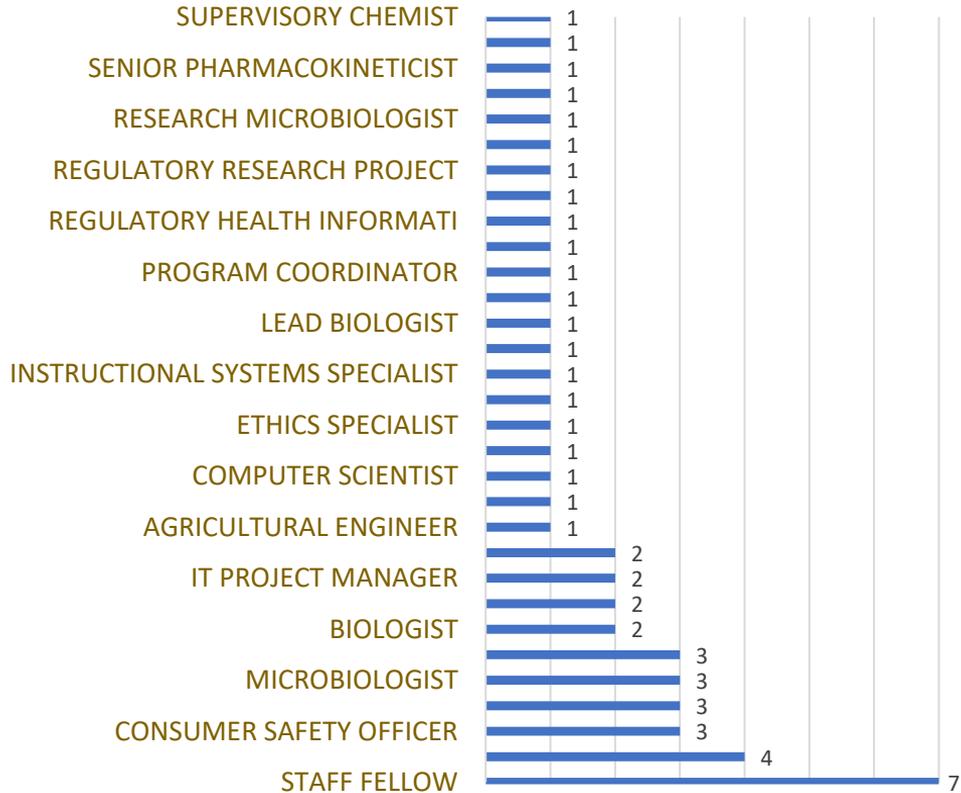# Participants by GS Level, Skill Level, and Gender

Of the 52 ALT participants to date (i.e., Pilot, 1, 2, and 3), the largest number of participants have entered the program with a self-identified **level 1 skillset,** meaning they have some basic coding skills. Participants also have **varying GS Levels**, but the greatest number has been GS-13. The distribution of the participants has been about 50/50 for male and female.

GS Level

GS Level	Participants
7	2
9	1
11	2
12	10
13	20
14	16
Cures	1

Skill Level

Skill Level	Participants
0 - Novice	14
1 - Basic Level	21
2 - Some Experience	13
3 - Intermediate	4

Gender

Gender	Count
F	28
M	24

Participants by Center and Role

Of the 52 ALT participants to date (i.e., Pilot, 1, 2, and 3), the participants have come from **all nine centers** with ORA and CVM having the largest representation. Within the centers, the participant's **job roles vary widely**, but the most common role is Staff Fellow.

Center Distribution

Center	Count
CBER	2
CDER	6
CDRH	5
CFSAN	6
CTP	7
CVM	8
NCTR	2
OC	7
ORA	9

Role	Count
SUPERVISORY CHEMIST	1
SENIOR PHARMACOKINETICIST	1
RESEARCH MICROBIOLOGIST	1
REGULATORY RESEARCH PROJECT	1
REGULATORY HEALTH INFORMATI	1
PROGRAM COORDINATOR	1
LEAD BIOLOGIST	1
INSTRUCTIONAL SYSTEMS SPECIALIST	1
ETHICS SPECIALIST	1
COMPUTER SCIENTIST	1
AGRICULTURAL ENGINEER	1
IT PROJECT MANAGER	2
BIOLOGIST	2
MICROBIOLOGIST	3
CONSUMER SAFETY OFFICER	3
STAFF FELLOW	7

Overview of the ORA/CVM Project

FDA

The Project

The Office of Regulatory Affairs (ORA) receives data from state Labs as part of LFFM (Laboratory Flexible Funding Model) and coordinates with other FDA Centers (i.e., Center for Food Safety and Applied Nutrition (CFSAN) and Center for Veterinary Medicine (CVM)) on surveillance activities. These activities allow the FDA to coordinate with state labs to enhance FDA's capacity to monitor the food supply. The state data is submitted to the FDA for aggregation and analyzed in conjunction with FDA's Compliance Program data to inform FDA risk assessment, surveillance/workplan, and policy development activities. Currently, the task is done manually by each FDA staff/group seeking to use LFFM data and can take significant resources.

The Goal

To use data science tools to normalize, aggregate, and visualize data from LFFM data sets and other internal FDA data. The teams will work on integrating data sets to develop an automated method to ingest data, normalize the data, and generate an end visualization method that can be used by FDA scientists to inform their food monitoring programs.

DATA COLLECTION
- *LFFM Data (State)*
- *Compliance Program Data (FDA internal)*

DATA AGGREGATION
- *Formatting*
- *Cleaning*

DATA ANALYSIS
- *Data Dashboard*
- *SME Review*

SURVEILLANCE PRIOTIZATION / RECOMMENDATION
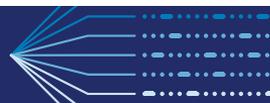
Analysis Workflow
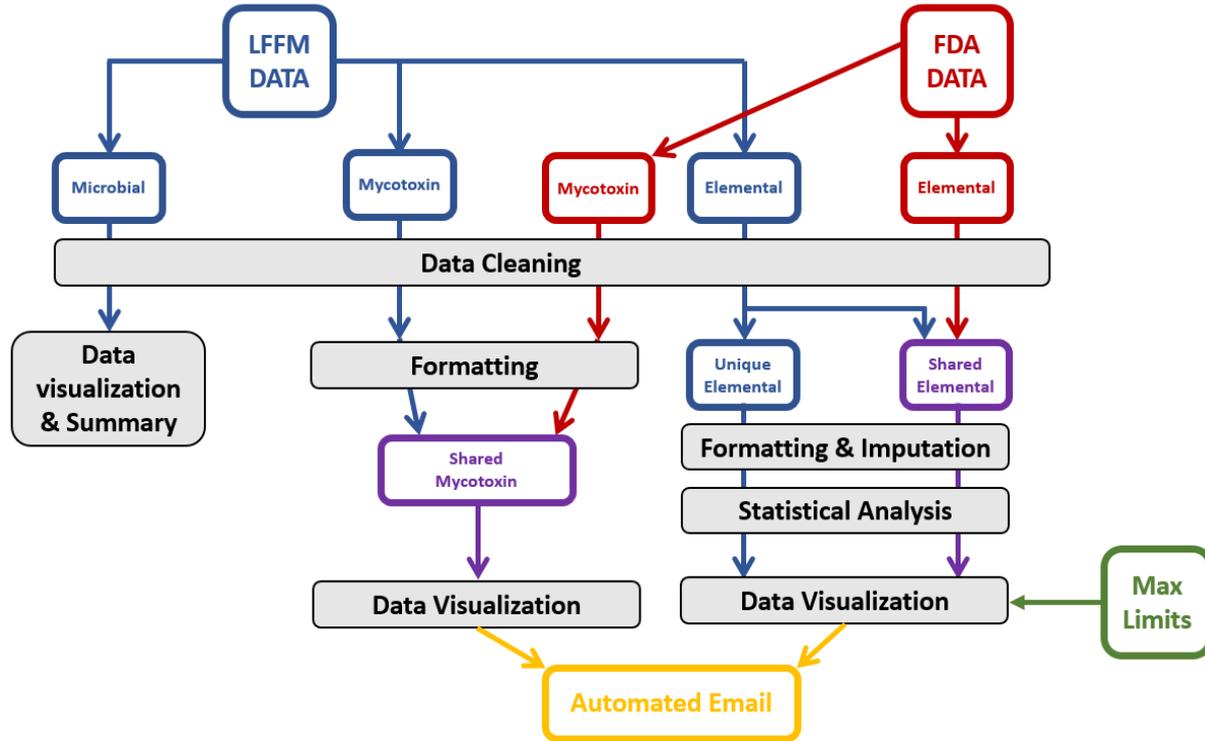
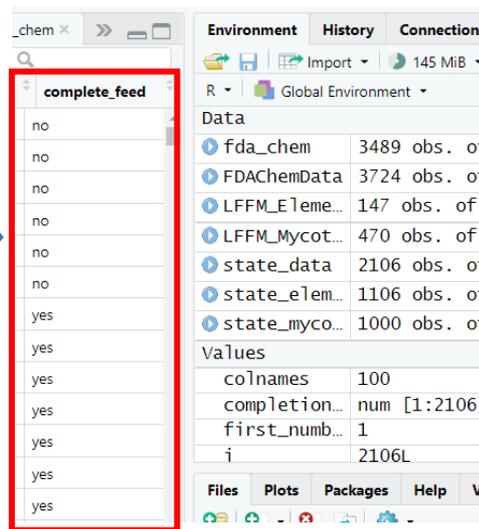# CHEM WORKFLOW: CREATE 'COMPLETE_FEED' COLUMN

Process Flow Guide

Step 7. Create a new 'complete_feed' column and add "**yes**" or "**no**" for the conditions: when feed can be described by the following **terms "69|70M|72"** in the product_code column and/or **"PELLET|PET FOOD|POULTRY FOOD"** in the product_code_desc column, add **"yes"**. When terms **"70|71|72F|72G"** and/or **"RAC"** are found in the same respective columns, add **"no"**.
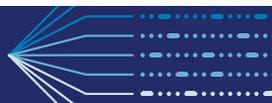
R Studio Code Chunk

```r
```{r}

for (i in 1:nrow(fda_chem)) {
  if (grepl("69|70M|72", fda_chem$product_code[i]) ||
      grepl("PELLET|PET FOOD|POULTRY FOOD", fda_chem$product_code_desc[i])) {
    fda_chem$complete_feed[i] <- "yes"
  } else if (grepl("70|71|72F|72G", fda_chem$product_code[i]) ||
             grepl("RAC", fda_chem$product_code_desc[i])) {
    fda_chem$complete_feed[i] <- "no"
  }
}
```

'complete_feed' column

Successful Completion of ALT 2

THE PROJECT

Normalizing, aggregating, and visualizing LFFM data sets submitted in combination with internal FDA testing data to support **CVM**'s animal food monitoring and sampling activities.

With…	*from…*	*where only…*
12	**6**	**58%**
participants	Centers/Offices	started with some coding skills

PROJECT OUTCOME

This project is continuing as a **Data*Forward* Squad Project** where 5 of the existing ALT 2 team participants will integrate their solutions to create a final product for the CVM team.

FROM THE PRODUCT TEAM

"I wanted to thank you and the team for taking on this project for CVM…. Having the tool, you all developed will help my team move this critical work to the forefront…

Thanks again! Words cannot express how appreciative I am and how impressed I am with the team that worked on this project…

Thanks again and don't be surprised if I tap into you all in the future!"

Linda Benjamin, P
Branch Chief, Science Policy Branch

Objectives and Value to FDA

FDA

DataForward Objectives

Identify project and data needs and help FDA staff become familiar with Data Science at FDA.

Develop the skills and knowledge of tools to resolve FDA issues.

Develop a common vernacular and network connections with FDA data scientists across centers and the agency.

Build and demonstrate a portfolio piece to instill confidence to tackle data challenges in home center or office.

Become comfortable with how data is used across the agency to inform better data systems and data driven decision.

Value to FDA

In house development of data pipelines and increase in institutional memory of built products.

Analysis of a large dataset to identify analysis issues or approaches that are driven by subject matter expertise.

Development of staff and skills in analyzing and presenting data and technical issues coherently.

Acknowledgments

Track Team Member

For the patience, courage and sheer passion it takes to learn something new and help build something great.

Data Science Mentors

For the energy and effort and love that you all show for data, data science, teaching and growing a community that is lucky to have you all.

Product Owners

For being willing to share your work, time, data to help people grow knowledge and community at the FDA.

- This work is driven by the community, for the community to grow the community.