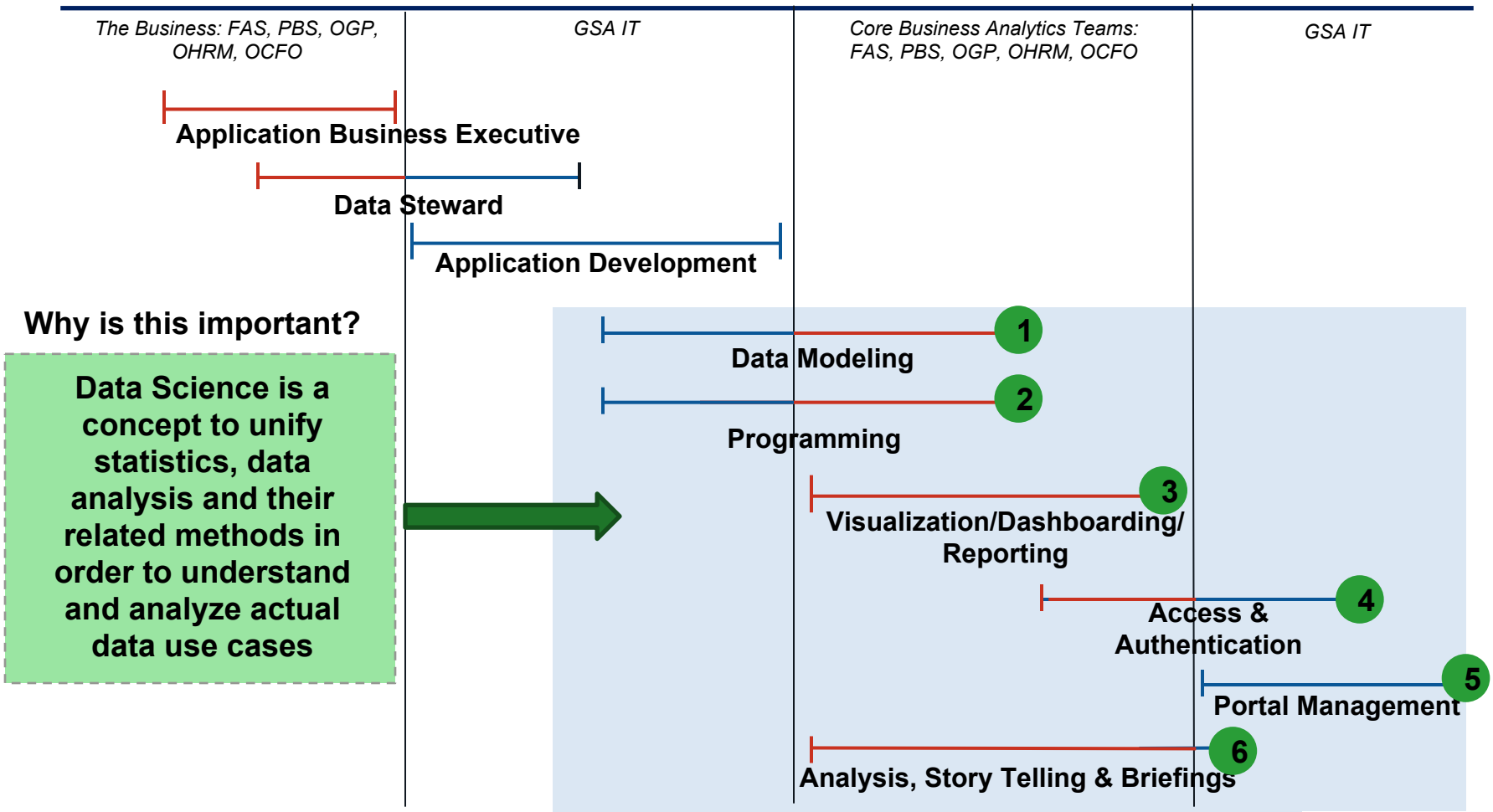
Data Science Tech Talk

10/5/17

GSA

D2D
DATA TO DECISIONS

Creating a *Data Science Center of Excellence* at GSA



Data Science Practitioner Training– Training Courses scheduled to be delivered to agency beginning FY18

Data Science Practitioner Training

13 required courses spread out throughout FY18

	2	3	5	6
Title	**Programming & Data Model**	**Analytics & Reporting**	**Content Management**	**Story Telling & Briefing**
Tools	**R & Python**	**Tableau & MSTR**	**D2D, Drupal**	**OCE**
Description	Leveraging R & Python to run statistical and mathematical scenarios	Using a suite of analytics tools, data is exported to create a high level, detailed dashboards to present & analyze data	Building and tracking content on D2D portal, includes adding and editing existing datasets, dashboards & data models	Internal & external presentations around customer data. Skilled in transforming customer & market sight into materials to help GSA and business line forecast
Curriculum Pathway	**Introduction: Identifying your business problem**	Intermediate R Intro to Stats Data Engineering	Data Modeling Data Visualization	Advanced Tableau Advanced Microstrategy	Human Centered Design
	Beginner Python Beginner R Intermediate Python	Beginner MicroStrategy Beginner Tableau	Intermediate Tableau Intermediate MicroStrategy	D2D Content Management	

GSA ⋯ *Mandatory intro*

**Note: Calendar subject to change*

Data Science training aims to foster proficiency in the power of data; COE certification will greatly enhance skillsets

Training Course	Skills Obtained	Business Value
Programming & Modeling	• Perform basic R/Python programming language functions and code writing • be proficient in statistical analysis, • run R/Python coded programs	**Less reliance on contract support for basic analysis** ✓
Analytics & Reporting	• Wrangle and clean data to prepare for report building • Create dashboard visualizations based on business needs & customer requirements	**Enhancing skill sets of Federal employees** ✓
Content Management	• Create and edit new Reports, Datasets, Documents, and Articles for their Customer	**Increase in self service** ✓
Story Telling & Breifing	• Transform customer and market insights into materials to help GSA and the business lines forecast and anticipate customer needs	**Opening doors to new opportunities** ✓

GSA

In planning stages
Currently active

Intro to Data Science Virtual Desktop (DSVD)

- DSVD is a virtual GSA desktop that can be used to access D2D components, and the D2D Data Warehouse

- GSA's DSVD provides software tools, data, underlying server hardware and network infrastructure.
  - Users need access to a machine.
  - Specific datasets, tools, other content and components are available to the DSVD user by default.
  - Request process in place to obtain approval for non-public content access.

GSA

D2D
DATA TO DECISIONS

DSVD Tools

Developer

- Putty
- Cygwin Shell
- Notepad ++
- WinSCP
- FileZilla
- Oracle Java JDK
- Jasper Studio
- MicroStrategy Developer
- GIT
- AWS Command Line Interface
- Anaconda
- Access to Python
- R
- RStudio

D2D Component Access

- Alfresco
- Jasper
- Python
- D2D Staging Data Warehouse
- Pentaho
- Languages: SQL, Python, R, Java, UNIX Scripting

Database

- MySQL Workbench
- Oracle SQL Developer
- Microsoft SQL Server Management Studio
- Pentaho DI

Upcoming

- November launch
- Production Access
- Access to:
  - JBoss Fuse
  - JBoss DataVirt
  - JBoss BPMS

GSA

D2D
DATA TO DECISIONS

Data Science Curriculum Pathway

Before entering the program:

Prep	Month 1	Month 2	Month 3	Month 4	Month 5	Month 6
	Data Science Overview					Human-centered Design
		Intro to Statistics				Capstone
Python for Data Science	Intro Python	Intermediate Python		Data Engineering	Data Modeling	Data Visualization
Fund. Methods for Data Science in R	Intro R	Intermediate R				
			Beginner MicroStrategy	Intermediate MicroStrategy	Advanced MicroStrategy	
			Beginner Tableau	Intermediate Tableau	Advanced Tableau	
						Beginner Content Mgmt

Programming & Modeling

Analytics & Reporting

Content Management

Career Tools

Electives - Pick 1 or more

Optional

Links

Data Science Practitioner Interest Form

https://docs.google.com/forms/d/e/1FAIpQLSeDQez3GJjHB0vo8Z5yJXstXuBSFBAyQOw1k2E1TbaOno1Ang/viewform

Register for the Oct 20th event

https://open.gsa.gov/events/data-science-codealong

View the class info on GitHub

(work in progress)

https://github.com/GSA/training-pathway-data-practitioner

View the Data Science Practitioner program info

(coming soon)

https://tech.gsa.gov/

Additional questions can be sent to sara.cope@gsa.gov!