

Data Scientist



① Generalist vs Specialist JD Long

=

Generalist

- small organization
- tendency to come out of statistics or quantitative bent of subject matter - astrophysicist, biologist, geographer, economist, psychologist

~ DEVOPS

- R / R4DS

- full range - collection through communication wide but not as deep

- data viz

- Venn Diagram (Drew Conway)

- R - statistics at its core, with programming

Robin

- European vs North American
- ~~the~~ same name
- two very different birds
- migratory = sexual dimorphism - size
- "man with a hat"
- look superficially similar

Specialist

- Pin factory

- org with specialized roles

- MA/PhD esp in CS or Comp Stats

- models ("algorithms")
AI

- narrow but deep

(learn more and more about less and less)

~ IT specialization (extension of IT specialization?)

- Python (widely used programming language)
w/ statistics - DS extensions

subject matter analysts
methodologists

data scientists - pushing both over

{ IS THERE A PLACE FOR A GENERALIST IN A PIN FACTORY? }

2/

So: How DOES A SMALL ORG INCORPORATE DATA SCIENCE?

BC STAYS EXPERIENCE

- LARGELY SKEWS GENERALIST BUT WITH NON-DS TOOLING → SLIDE

- MOVING TO R-CENTRIC MODEL

① FIND "SUPER USERS" (DR. GOTT)

- EMPOWER THEM
TOOLS / PLATFORM
SPACE TO EXPERIMENT / ENVIRONMENT WHERE MISTAKES ARE TOLERATED (ENCOURAGED!)
PRACTICAL APPLICATIONS
- THEY WILL FIND PROBLEMS TO SOLVE
(E.G. IF THEY HAVE BEEN EXPOSED TO PAIN-POINTS IN WORKFLOW)
↳ ALREADY KNOW!

FUNCTIONAL PROGRAMMING

{ PROJECTS TEND TO BE SMALLER SCALE, OFTEN VARIATIONS ON THEMES, SO GENERALIZABLE WITHOUT "MOVING INTO PRODUCTION" }

- DEPLOY ACROSS LINES OF BUSINESS / SUBJECT AREAS

⇒ LOOK FOR EFFICIENCIES

- OPPORTUNITIES TO MOVE TO OAD
- OPPORTUNITIES TO ADD NEW FUNCTIONALITY AND/OR LESS RELIANCE ON TRADITIONAL IT

60
40 2400
[ORDERS OF MAGNITUDE LARGER!]

{ AT THIS SCALE, IT'S HARDER TO BREAK THINGS
- LESS COMPLEXITY

PROPENSITY TO SOLVE HARDWARE & SOFTWARE PROBLEMS
(E.G. HOW TO SET POINTERS IN RPPROFILE)

② Identify "SUPER USERS in TRAINING"

- TRAINING
→ ESP WHEN IN CONTEXT OF OWN WORK
- ~~TRAINING~~ - BENEFIT FROM GROUP TRAINING! (BOTH PROGRAMMING & HARDWARE / SOFTWARE DOCUMENTATION?)
- WILL ALSO BENEFIT FROM INTERNAL PACKAGING → STANDARDIZATION & COLLABORATION

THE
BEGINNING
OF A
HAIRBALL

TANGENT

NOT JUST PEOPLE

- HARDWARE
- AUTHORIZATION TO USE TOOLS
→ DEVELOP OWN PACKAGES
- PRIVACY / SECURITY

LEADERSHIP NEEDS
TO ADVOCATE FOR
THIS - WITHOUT IT,
DS WILL ~~BE~~ FLY SOUTH!

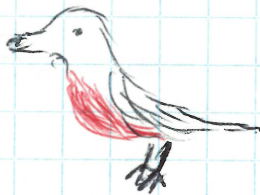


□ SOMEONE NEEDS TO
NAVIGATE THE HAIRBALL SO
THAT THE DS TEAM IS
~~FREE~~ FREE TO WORK

(WHERE TO FROM HERE?)

SYSTEMS THINKING

4/



■ LARGE ORG

- SPECIALIZED → SILOS

PRODUCTION - REGULAR SCHEDULE

- LARGE SCALE

↳ VERY EFFICIENT

• INNOVATIONS ARE INCREMENTAL

↳ ROLE FOR SPECIALIST DS

- SILOS MEAN:

- HARD TO TRANSMIT INNOVATION

- CULTURE IS FOCUSED ON MAINTAINING
"WHAT WORKS" / "IF IT AIN'T BROKE"

- HARD TO SHIFT AS WORLD SHIFTS AROUND YOU

• KODAK VS FUJI

"CHANGING THE ENGINES OF THE PLANE
WHILE IT'S FLYING"

- BETTER METAPHOR - KEEP THE COACHES
RUNNING WHILE YOU ARE EXPERIMENTING
WITH AIRPLANES - OBJECTIVE IS TO
GET FROM A TO B, IN THE MOST
EFFICIENT & EFFECTIVE WAY POSSIBLE