Magenta
Care Continuum

Who is responsible for Data Quality?

Olive Ventures
Portfolio Company

Learning Objectives

1. Understand importance of a healthy relationship between data consumers and data producers

1. How to be a critical data consumer

1. Identify and describe the different roles and responsibilities associated with data products

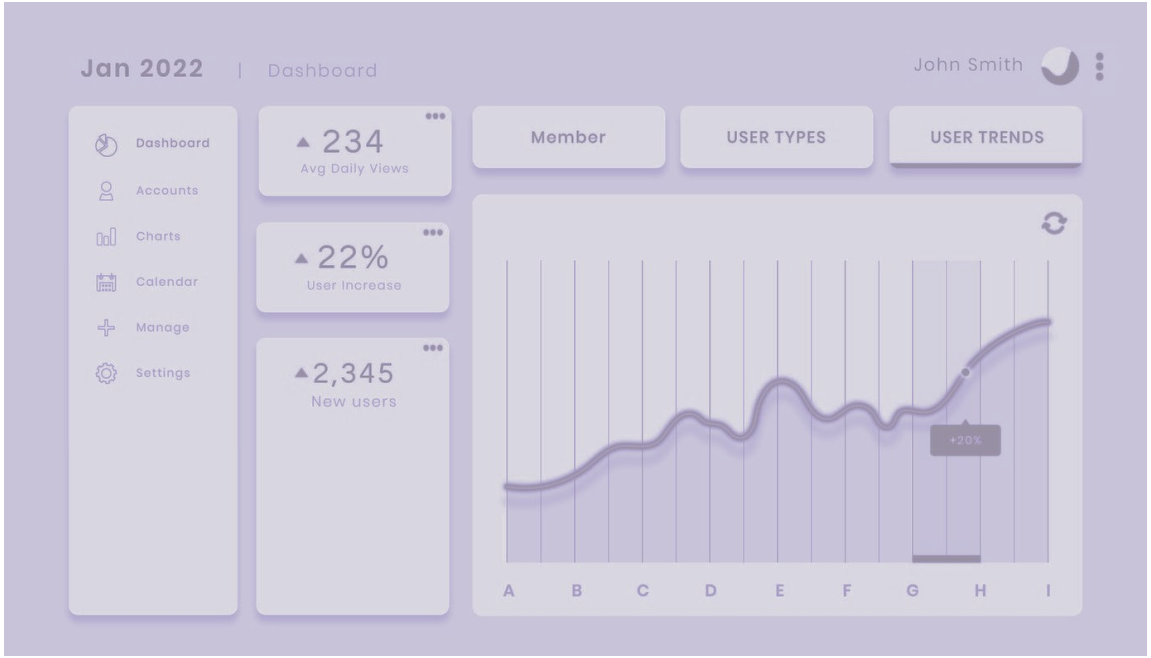
Real-life example

Dashboard Looks Wrong

Now what?

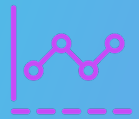
Who do you contact?

How did this happen?



Who Is Involved?

Dashboard User



Data Analyst

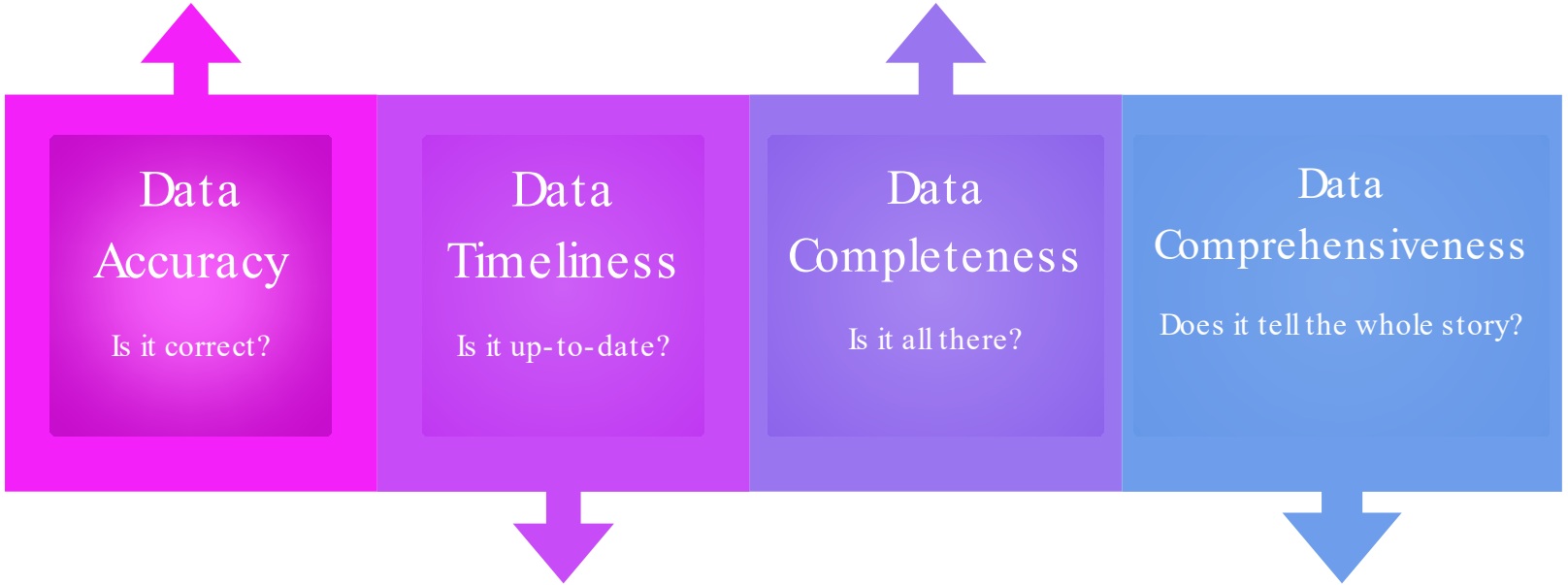
Database Steward

Data Engineer

EHR/ Software Engineer

Domain Expert

What Are We Looking For?



Dashboard User

Initial Questions

Is EVERYTHING wrong?

How long has this been wrong?

Who do I talk to about this?

Solution Efforts

Tracking down the dashboard owner

New Information

Something is wrong!

Dashboard owner (Data Analyst)

Initial Questions

What exactly looks broken?

Is the data up-to-date?

Is my data query correct?

Solution Efforts

Finding the exact metric that seems wrong

New Information

Identified the metric that looks weird

Data Steward- Owners of the data

Initial Questions

Are the database tables up-to-date?

Is the data being used correctly?

Is the data complete?

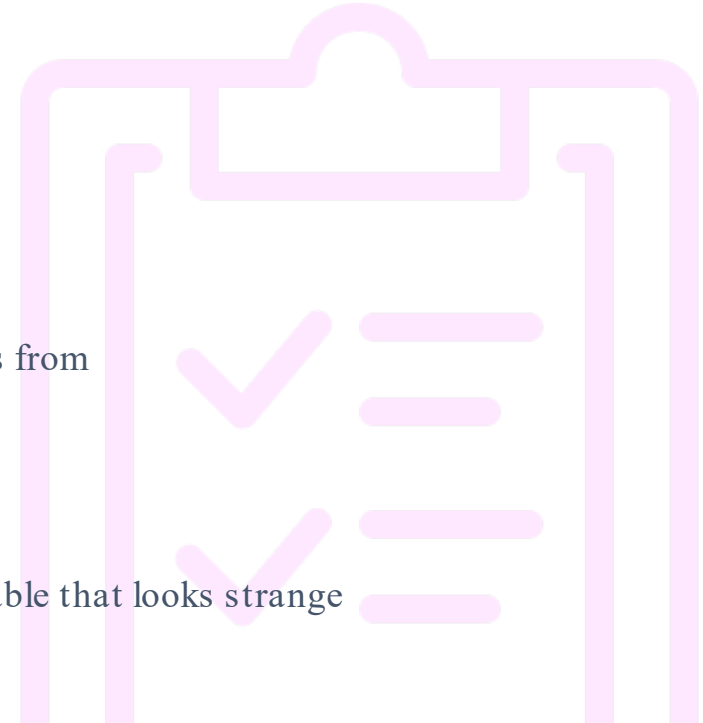
Solution Efforts

Identifying the table the metric comes from

Validates data in tables is accurate

New Information

Pinpoints the exact source field in a table that looks strange



Data Engineers

Initial Questions

Did something happen in the data transformation process?

Did it get into the intended destination tables?

Solution Efforts

Checking their transformation unit tests

Checking volumes of data processed

New Information

Recognizes that the volume of data to that field has decreased

EHR/Software Engineers

Initial Questions

Is there input data still coming in?

Are we storing that data properly?

Solution Efforts

Validating the input and storage of data

New Information

Everything working correctly technically

Data Steward and Domain Expert

Initial Questions

If it's not a technical error (this time!), where is the weirdness coming from?

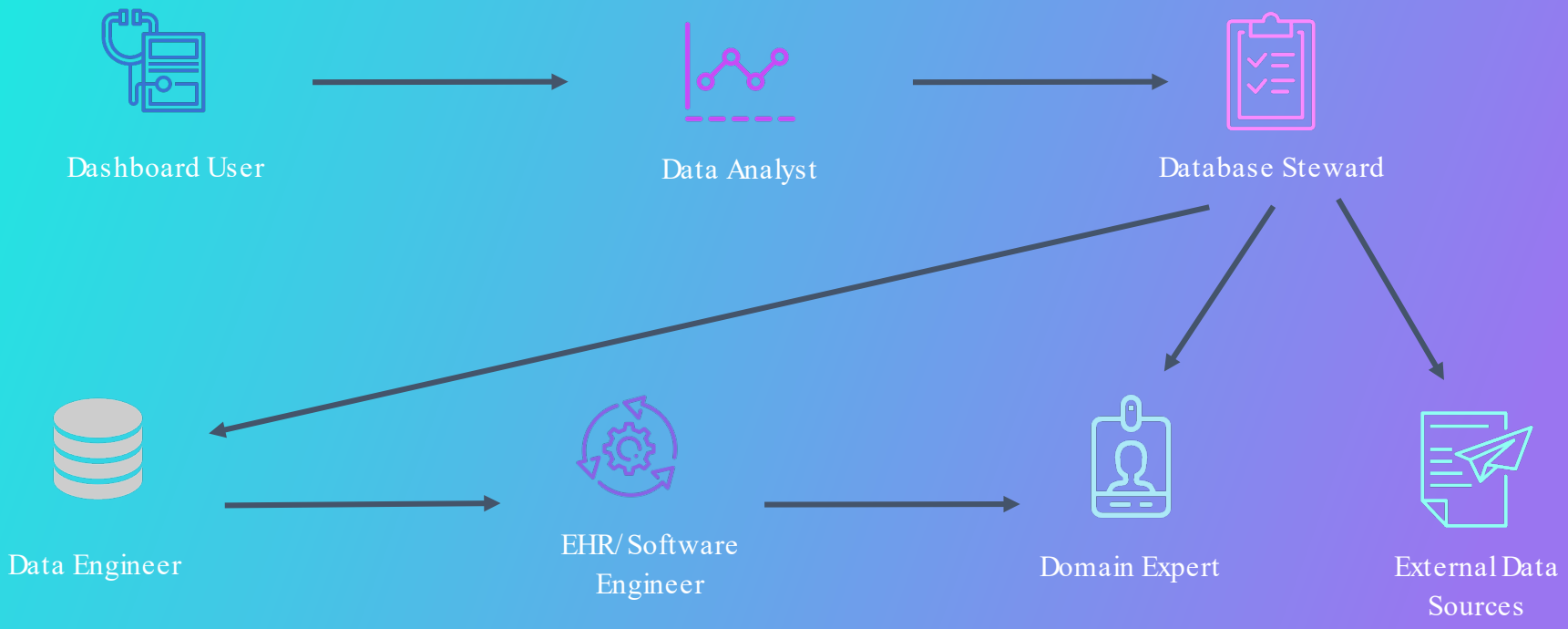
Solution Efforts

Consulting a domain expert

New Information

The communication breakdown about process changes is evident

The Troubleshooting Pathway



Ideally

Define

- A comprehensive list of data elements
  - Special documentation around external data sources (eg EDI)
- Thresholds of acceptability in timeliness and completeness for those elements
- Identify who needs to be notified when there is an anomaly in the data

Measure

- Design tests to check if data meets the defined thresholds of acceptability

Track

- Automate the tests to regularly check for data quality and to notify invested parties the results of the tests

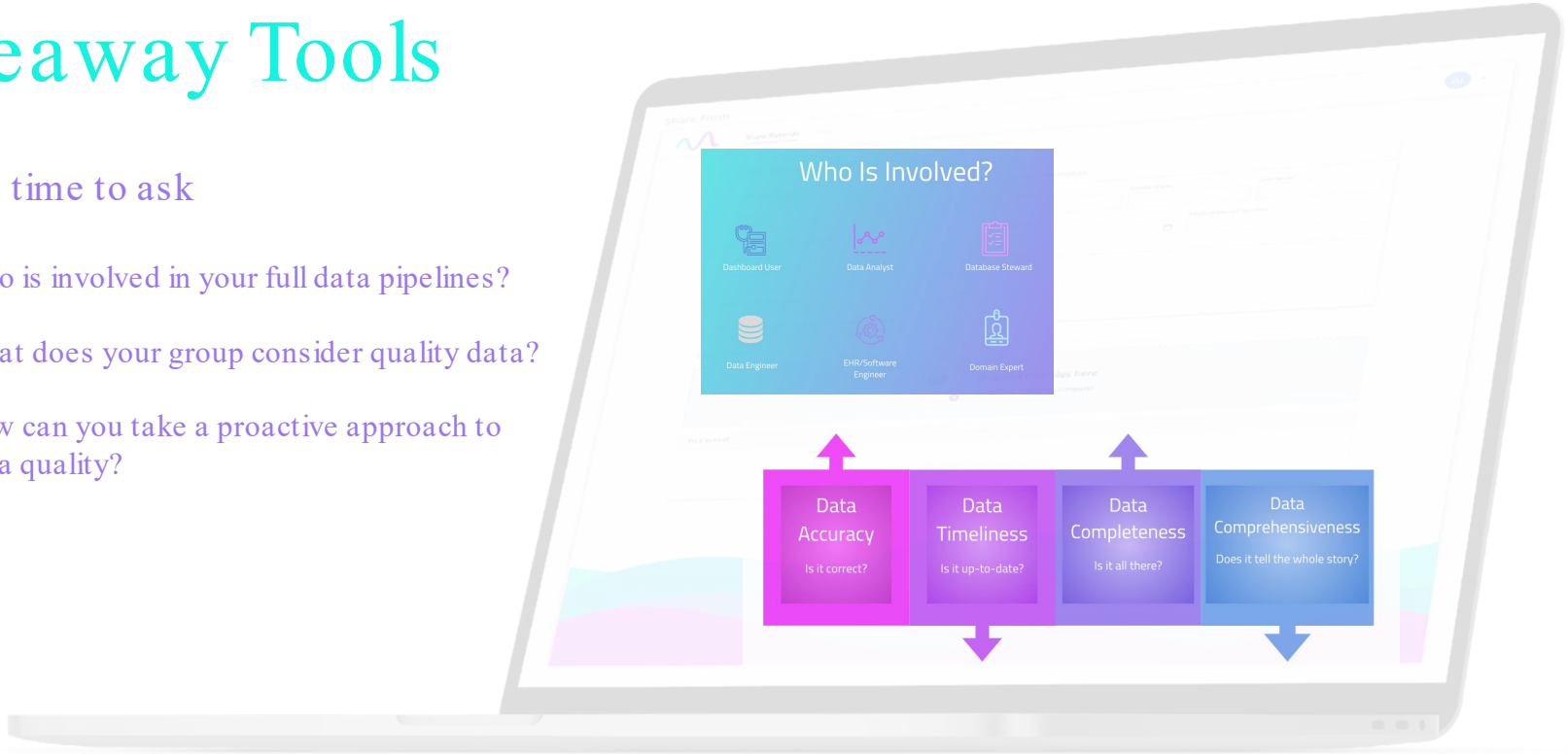
Communicate

- Information is passed along in all directions

Takeaway Tools

Take the time to ask

- Who is involved in your full data pipelines?
- What does your group consider quality data?
- How can you take a proactive approach to data quality?



Thank You!

For further information or questions contact:

Jackie Kokx

Sr Data Scientist

Magenta Care Continuum

jackie.kokx@magentacc.com

517.927.8748

Justin Schaper

COO, Magenta Care Continuum

EiR, Olive Ventures

justin.schaper@magentacc.com

214.676.3533

Magenta

Olive Ventures
Portfolio Company