MODULE 1: DATA FOUNDATIONS

CT ACADEMY | DATA ACTION LAB

1. DATA AWARENESS

DATA FOUNDATIONS

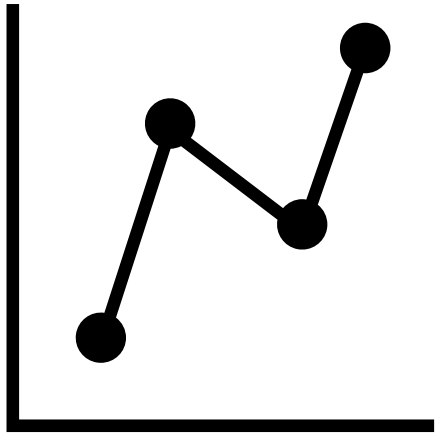
WHAT IS (ARE) DATA?

Should we say, “is data” vs “are data”?

The word data is technically a plural, so “are” is appropriate; the singular is the word **datum** (for a “data point”).

In common usage the word data is used interchangeably (datum has become *passé*): “are” is technically correct but “is” is used all the time!

Being pedantic, if “data” is used as a **mass noun**, then “is” IS appropriate!



PURCH_DOC	AWARD_DATE	SP	ORIGINAL_V	AMENDMENTS	TOTAL_PO_A	NAME
24XXXXXXXX	2015-11-23	TN	24814.30	0.00	24814.30	Canada
24XXXXXXXX	2015-11-23	TN	11327.58	4674.08	16001.66	Canada
24XXXXXXXX	2015-11-23	TN	4860.00	0.00	4860.00	Canada
24XXXXXXXX	2016-05-06	TN	52000.00	0.00	52000.00	Canada
24XXXXXXXX	2014-09-02	TN	23748.68	0.00	23748.68	Canada
24XXXXXXXX	2014-07-24	TN	15943.55	0.00	15943.55	Canada
24XXXXXXXX	2014-07-24	TN	20336.79	0.00	20336.79	Canada
24XXXXXXXX	2014-10-07	TN	29286.40	0.00	29286.40	Canada
24XXXXXXXX	2016-07-28	TN	13800.00	0.00	13800.00	Canada

WHAT IS (ARE) DATA?

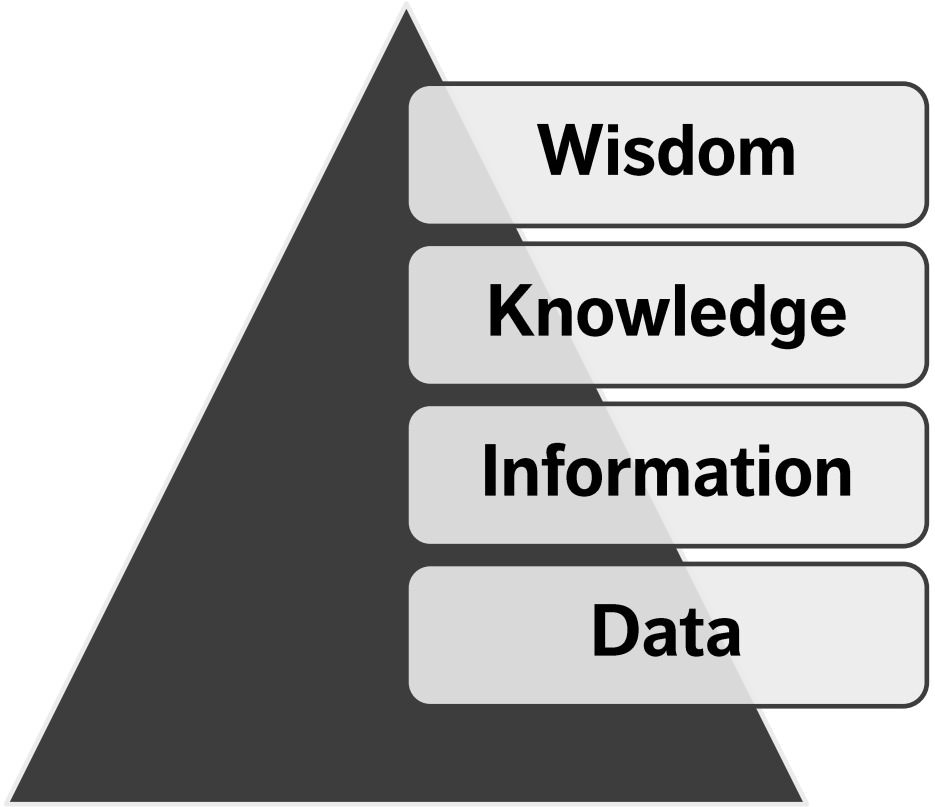
Data can be thought of as **raw "numbers".**

It is often defined as **"a collection of facts from which conclusions may be drawn"**.

Data comes in many different forms and underpins all analyses.

(We will revisit these notions)

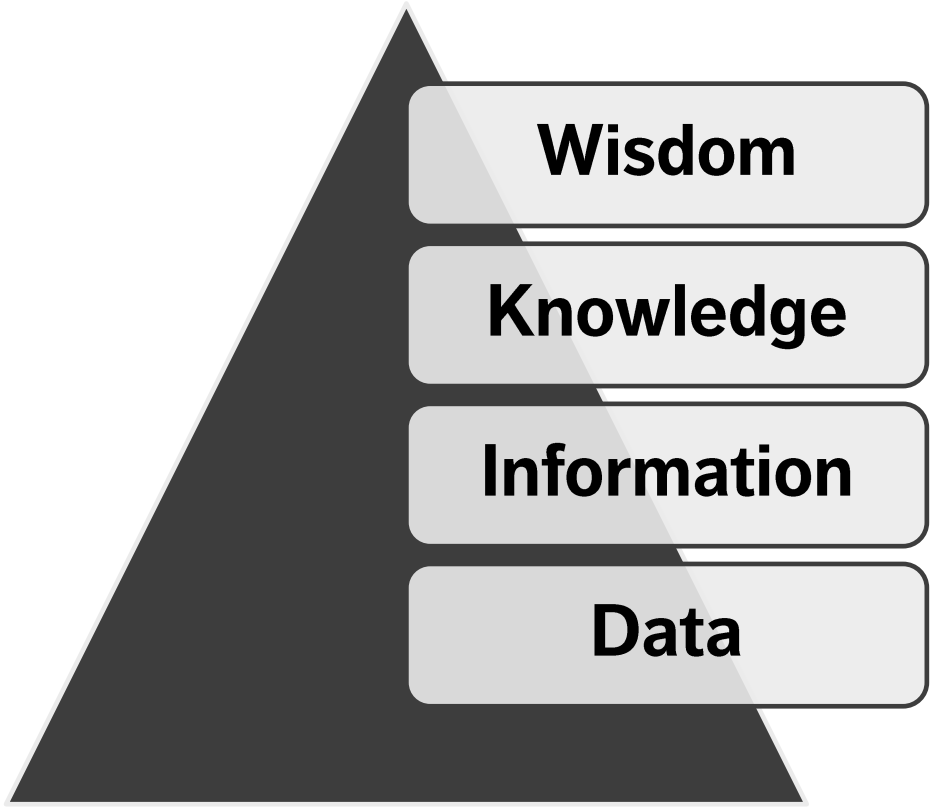
DATA IS A FOUNDATION



DIKW Pyramid

- represents **structure** or functional **relationship** between elements
- we **acquire** data
- **organizing** data gives us information
- using information in **context** yields knowledge
- the correct (and/or incorrect) **application** of information over time makes us wise!

DATA IS A FOUNDATION



DIKW Pyramid Example

- **Data** – individual bank transactions
- **Information** – organizing the transactions into monthly groups (what are my monthly spendings?)
- **Knowledge** – comparing the "spend per month" to a budget
- **Wisdom** – understanding that if I am under budget (over time) I can save up for a vacation!

HOW WE USE DATA

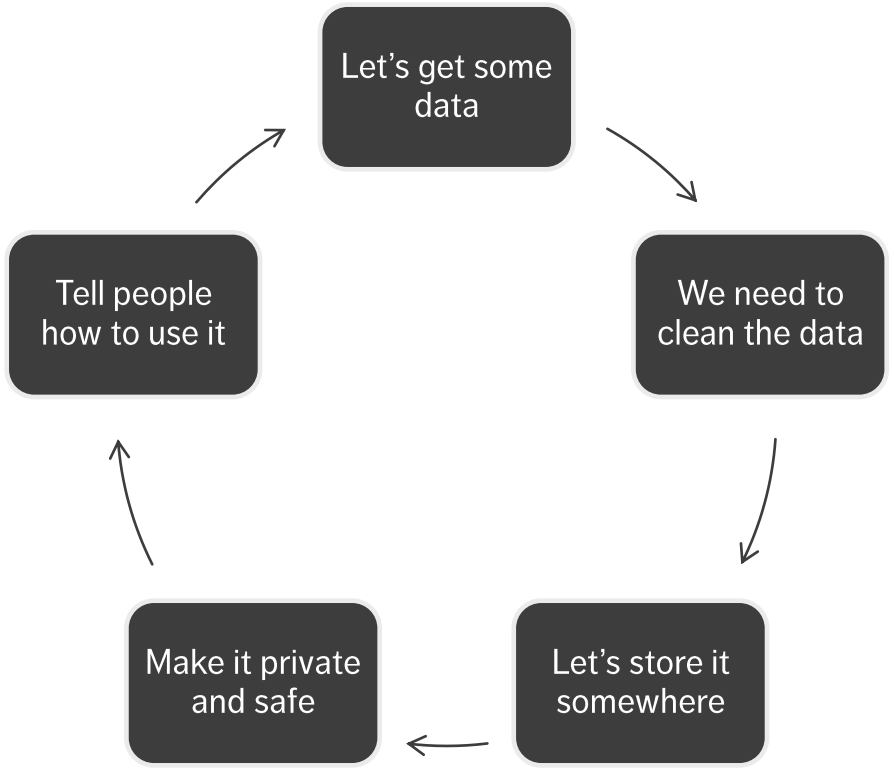
We **analyze** data to tell **stories** that help us make **decisions.**

This requires **data integrity**.

Most departments have **data stewards** who manage the **data lifecycle** (how to acquire, manage, and use data).

Data stewards keep the data in **data assets** and make them available for people to use as they require.

Analysts must **contextualize** the data to help them make effective decisions.



HOW WE USE DATA

What do you do once you have data?

- identify what decisions you, or your team (or your boss) needs to make
- identify the data that you need to make that decision
- get the data, check it to make sure it's ok
- you can then do one of more of the following: **analyze** the data, **visualize** the data, **summarize** the data
- turn the data into **information** and **knowledge**
- make your decisions or provide your outputs to the stakeholders who need them

DATA RELATED TERMINOLOGY

Here are a few common data **buzzwords**:

Term	Description	Example(s)
metadata	data about data	a document that describes what column titles mean in a spreadsheet
data asset	a system or program that stores data	Excel spreadsheets, SAP, PeopleSoft, Access Databases, Web Tables
reference data	data that is common across data assets	list of Provinces, list of countries, list of branches in a department
master data	data that we use to run our business	employee names, transaction amounts
data inventory	a list of data assets	tools like Microsoft Purview maintain lists of data assets
data catalog	descriptions of data	precisely defines words the business uses, e.g., “FTE”, “Headcount”
data model	how data interrelates	“linking” together financial and HR data through a PRI

INTRODUCING ROLES & RESPONSIBILITIES

Who uses data?

- if you are using data, you are a **data consumer**
- if you a responsible for the integrity of the data you are a **data steward**
- **data trustees** are accountable for the data
- If you input data into a system, or acquire it from somewhere and add it to a data asset you are known as a **data contributor**
- if you help to manage the systems in which the data resides, you are a **data custodian**

EXAMPLE : USE OF DATA – HOW MUCH MONEY DO WE HAVE?

We need to see how much money is available in the department ("**free balance**"). We run a report from the relevant data asset, getting the data on what we:

- have spent up to this time (the "**actuals**");
- have committed to spend (the "**commitments**"), and
- think we should have spent (the "**budget**").

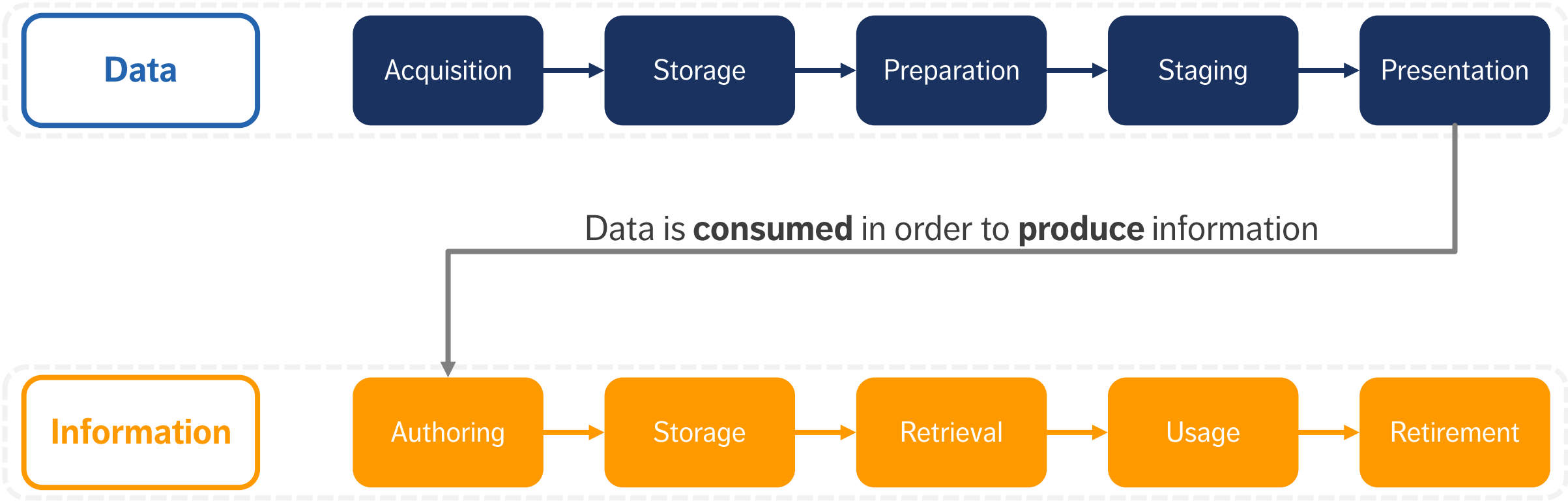
Data stewards check if there are any problems with the data and fix them any such problems.

Finance asks business to **validate the amounts** to see if they are accurate and that nothing is missing.

Finance then applies the following **formula** and stores the result for its financial reporting obligations:

$$\text{Free Balance} = \text{Budget} - (\text{Actuals} + \text{Commitments})$$

DATA AND INFORMATION LIFE CYCLES



DATA LITERACY

To support the data strategy, the GoC requires that GoC employees be **data literate**. Supporting data literacy is the **GoC Data Competency Framework** that this set of courses is built around.

> "Having a data literate workforce is at the core of modernization efforts. This Data Competency Framework is meant to support conversations and aims to advance data literacy by creating a shared understanding and language about data competencies for all federal public servants."

A department's level of data literacy is usually identified through **surveys**; gaps are addressed through **training**, **education**, **mentorship**, and other learning methods.

DATA LITERACY

The *Data Competency Framework* consists of four sections that are divided into three proficiency levels:

Sections:

1. Data Concepts and Culture
2. Data Governance, Collection, and Stewardship
3. Analytics and Evaluation
4. Data Systems and Architecture

Proficiency Levels:

1. Foundational: defining the core level of understanding and awareness
2. Intermediate: putting theory into practice
3. Advanced: applications and enabling others

DATA ROLES & RESPONSIBILITIES

Data roles and responsibilities are “required” by TBS as part of GoC Data Strategy.

They align business operations with **data governance** activities, helping managers and supervisors to define and assign **accountability** & **responsibility** to employees.

Explicit R&R help GoC employees understand how they fit in with their department’s data activities.

Note that one person can be assigned **multiple R&R** at the same time (it is possible to be both a data contributor and a data consumer simultaneously!)

WHERE DO YOU FALL?

Where do you fit? You may have been assigned a specific role (e.g., data steward), but regardless of role assignment it is highly probable that you are a **data consumer**. If you do any of the following you can count yourself as part of that role:

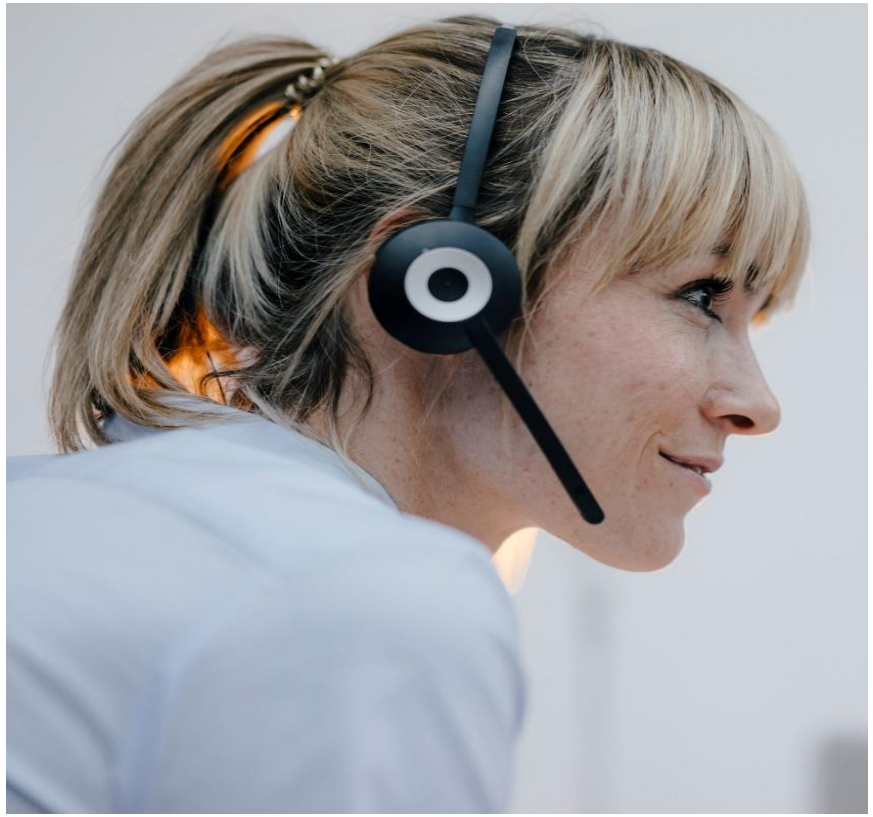
- exporting data from any system;
- create a spreadsheet that people use to make decisions;
- get data from outside of the department and use it in internal or external reports, etc.

You may also be a **data contributor** if you do things like:

- do research and gather data that adds to corporate knowledge;
- enter data into a system (e.g., call center employee entering case data);
- entering overtime into a salary system, etc.

Some roles are **assigned**, some are **inherited**. Other roles: **data trustee**, **data steward**, **data custodian**.

EXAMPLES: R&R



Example: IT Support Technician

You get a call from a data steward wanting you to update a business rule in a database (**data custodian**).

You get the information you need and apply the rule (**data custodian**).

You then export data from the database into a report so you can check that the change was correctly implemented (**data consumer**, **data steward**).

You then update the system with the information and close the ticket (**data contributor**).

EXERCISE

Write down examples of when you are acting in each of the following roles, if applicable:

- data consumer
- data steward
- data trustee
- data contributor
- data custodian.

SUPPLEMENTAL MATERIAL

1. DATA AWARENESS

EXAMPLE: USE OF DATA – MAKING DEPARTMENTS SAFER

The occupational health and safety group at a department wants to make our environment safer.

Every time a health and safety incident occurs, the **details are recorded** (type of incident, when/where/how it happened, etc.).

This data is **tracked** and **analyzed**.

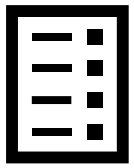
If trends are seen in the data (for example a lot of slips and trips happen at a particular location) then the team **decides to intervene**, and steps are taken to **mitigate the issue** (e.g., coating the floor with a non-slip surface, etc.).

More data is collected after mitigation so that the group can measure "**improvements**".

HIGH LEVEL REVIEW OF DIFFERENT R&R

Data Trustees

- ensure strategic management of assigned data assets as well as compliance with departmental and enterprise data (**related strategies**, **regulations**, **policies**, **directives**, and **standards**)
- executives with business accountability and intermediate level of technical knowledge (typically at **director level**)



Data Stewards

- advise on, enact, and help enforce **data policies** and **standards**
- **operations-focused**
- they have a mix of business and technical background (**branch representatives**)

HIGH LEVEL REVIEW OF DIFFERENT R&R

Data Custodians

- ensure the safe **custody** and **integrity** of hosted data, and safeguard the enterprise data repository
- normally **operations-focused**, with a technical background (**IT-focused**)

Data Contributors

- ensure that the data they provide to the Department (including third-party data) aligns with all technical and business **policies**, **procedures**, and **standards**
- **operations-focused**, with a "**business**" background and some technical expertise for the systems they typically use

HIGH LEVEL REVIEW OF DIFFERENT R&R

Data Consumers

- ensure that usage of data **supports** departmental and government objectives and mandates
- anyone within the organization can play that role, typically with a "**business**" background

EXAMPLES: R&R

Example: Call Center Agent

You take your first call of the day, the information from the call is entered in the call center system (**data contributor**).

A list of calls appears on your screen. You prioritize the callers on the list and select your next call (**data consumer**).

You realized that you entered a wrong piece of data that you can't overwrite, you call your data steward for them to fix it (**data consumer**).

You have a responsibility to review your team call performance for the day and provide feedback to your supervisor. You download a system extract and do the calculations in Excel before forwarding the results (**data contributor**).

EXAMPLES: R&R

Example: Program Director

You are providing a new service as part of the program you run. You review and approve a new database to track the program data (**data trustee**).

Once the database is up and running, you start to review and make decisions on the reports obtained from it (**data consumer**).

Your data steward identifies a major issue that is escalated to you for approval (**data trustee**).

You request that the system be integrated into an existing system (**data trustee**).

You use the reports from the new integrated system to help you administer your program (**data consumer**).