



Data Quality 101: What is Data Quality?

March 29, 2022

Jesse Jorstad, Abt Associates

Melissa Mikiel, ICF

Mindy Mitchell, Homebase



Jesse Jorstad, Abt Associates

he/him

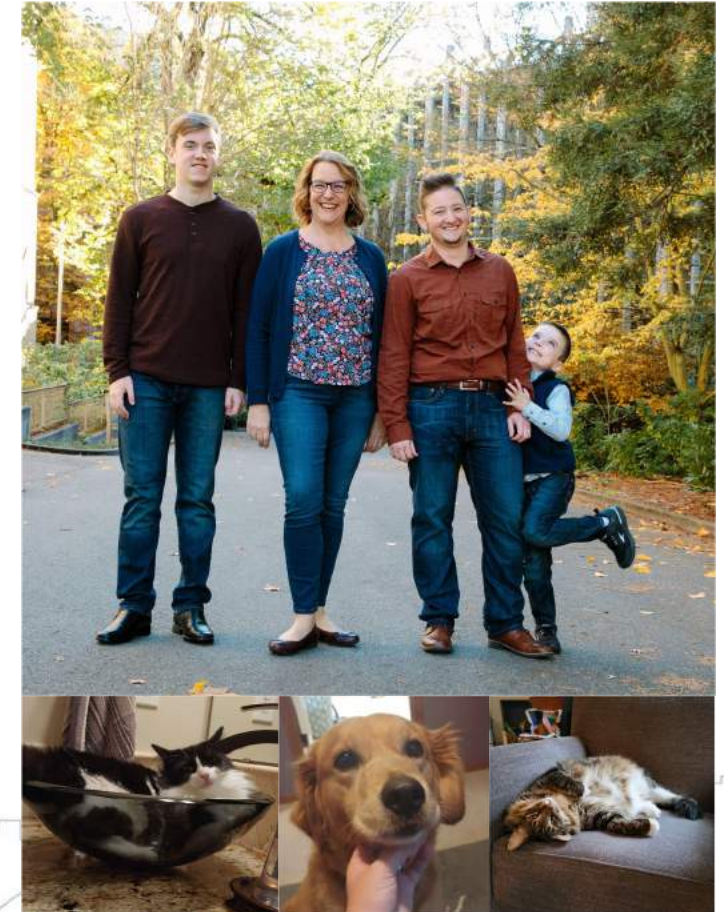
Prior to TA, I worked in homeless housing administration for 12 years:

- *5 in PSH rental admin in non-profit*
- *7 total with the County*
 - *2 in CoC Contract/Planning*
 - *5 in HMIS/Data analytics*

I have been with Abt Associates since August 2020.

I live in Lake Stevens, WA with my spouse, two kids, two cats and a dog.

I enjoy painting and volunteering to support LGBTQIA+ youth and young adults in my community.



Melissa Mikel, ICF

she/her



2010 - 2020

- 10 years as HMIS Lead/System Administrator/Analyst/Data Diva
- Lead/co-lead for federal and local initiatives

July 2020 to Present

- TA with ICF

Live in Fresno, CA with husband, 2 kids, & mini doberman. Oldest daughter is living her dreams in NYC.

Dutch Bros fanatic and love vibrant “wordy” socks

Mindy Mitchell, Homebase

2006: case manager, high-barrier family shelter

2008: CUNY School of Law

2014: NAEH

2021: Homebase

From MS, based in Baltimore, always daydreaming about being in the desert and planning my next western adventure.

Love cats, hiking, rock climbing, birding with my girlfriend, and other stereotypically middle-aged lesbian activities.



Learning Objectives

1. Participants will be introduced to the **core elements, definitions and metrics of data quality**.
1. Participants will become familiar with **HUD's vision and strategy** for data and understanding how data quality fits into that context.
1. Participants will understand the **impact of data quality** on reporting and system planning and ending homelessness in their communities.

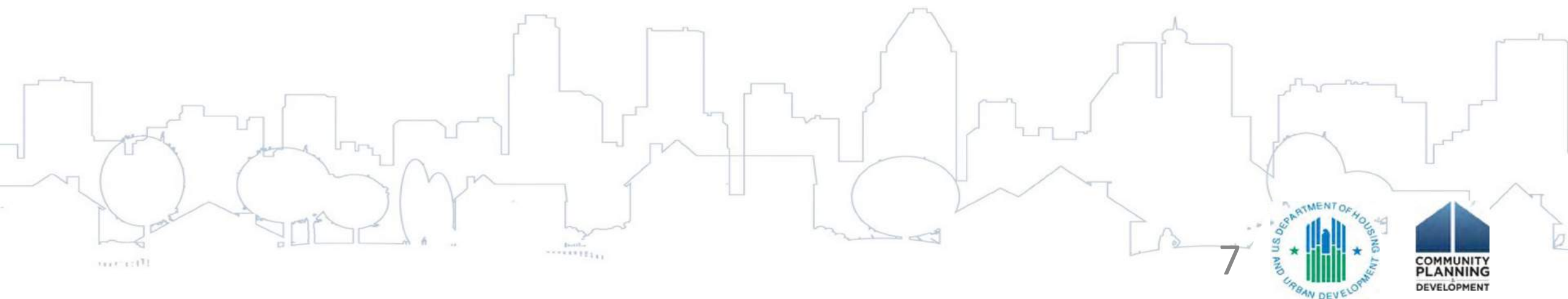
Who's Here Today?

Select all that apply:

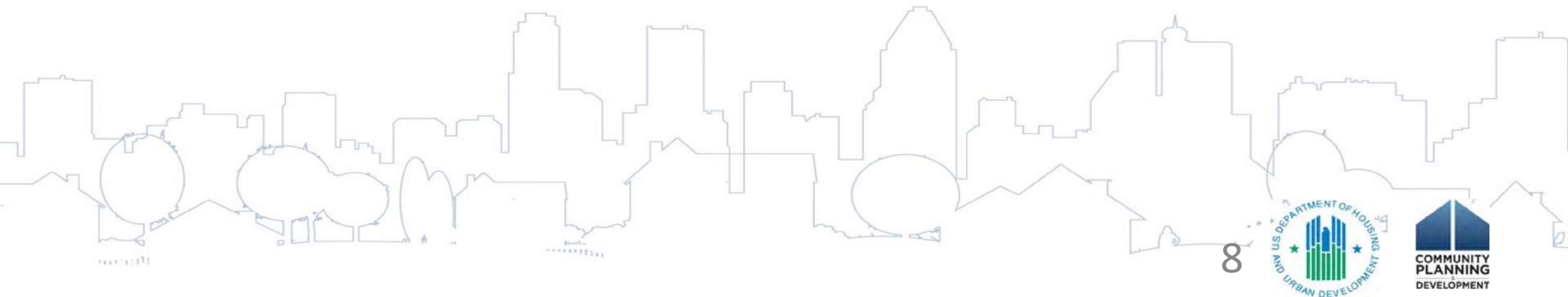
- HMIS Lead/System Administrator
- HMIS Vendor
- HMIS Participating Organization/End User
- Person with Lived Experience
- Government Entity
- Funder
- CoC Lead
- CoC Representative

How long have you been in your current role?

- less than 3 months
- 3 - 6 months
- 6 - 12 months
- 1 - 3 years
- 3 - 5 years
- Over 5 years



What is Data Quality?

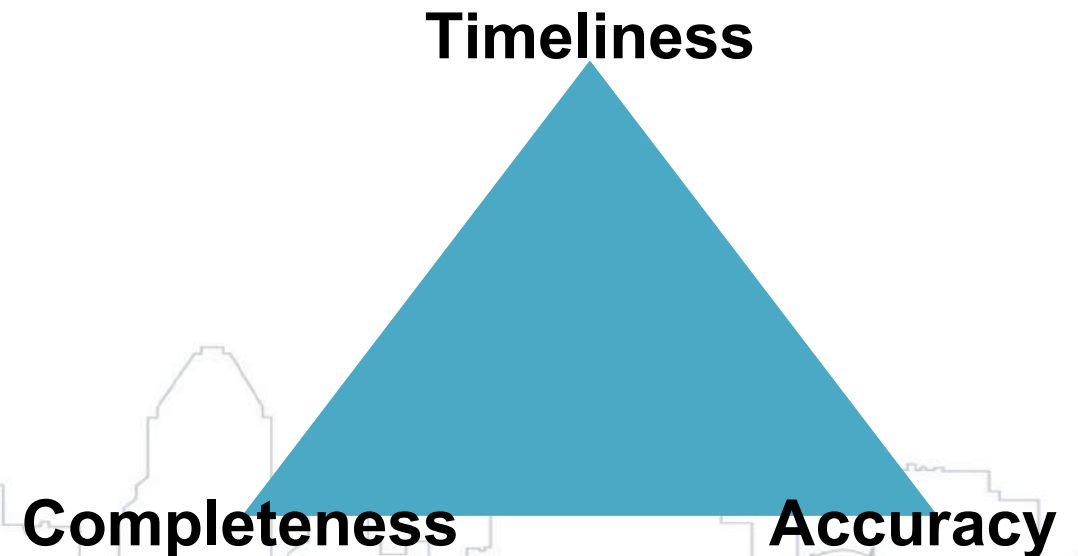


Data Quality Defined

Data Quality refers to the **reliability and comprehensiveness** of your community's data

Components of data quality include:

- Timeliness
- Completeness
- Accuracy



Requirements for Data Quality

2004 HMIS Data and Technical Standards

4.2.2. Data Quality (Baseline Requirement)

- “PPI collected by a CHO must be relevant to the purpose for which it is to be used. To the extent necessary for those purposes, PPI should be **accurate**, **complete** and **timely**.”

[2004 HMIS Data and Technical Standards](#)

Timeliness

What is timeliness? The degree to which the data is collected and available when it is needed.

How do we know if the data are timely? Local CoC should incorporate timeliness expectations into their Data Quality Plan and monitor accordingly. It may be appropriate for a standard to be set by project type.

Are there tools that can help? [HMIS Standard Reporting Terminology Glossary](#) report includes a timeliness measure.

Completeness

What is completeness? The degree to which all required data is known and documented. Coverage and utilization are both forms of completeness.

How do we know if the data are complete?

Are there tools to help? [HMIS Standard Reporting Terminology Glossary](#) report includes a completeness measures.

Levels of Completeness

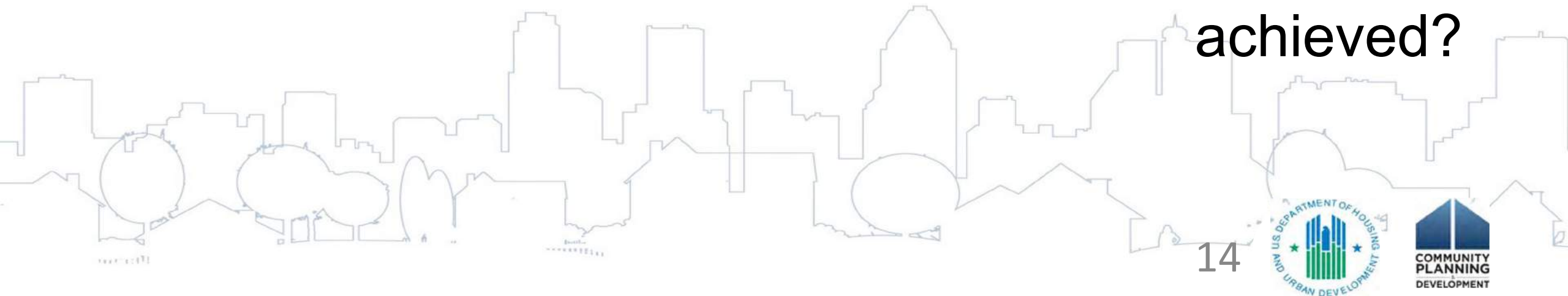
Universal Data Element	ES		TH		PH		SSO		Outreach		Prevention		Rpd Re-hsg	
	Null/ Miss	Unkn/ Dn't Kw/ Ref	Null/ Miss	Unkn/ Dn't Kw/ Ref	Null/ Miss	Unkn/ Dn't Kw/ Ref	Null/ Miss	Unkn/ Dn't Kw/ Ref	Null/ Miss	Unkn/ Dn't Kw/ Ref	Null/ Miss	Unkn/ Dn't Kw/ Ref	Null/ Miss	Unkn/ Dn't Kw/ Ref
1.Name														
2.Social Security Number (SSN)														
3.Date of Birth														
4.Race														
5.Ethnicity														
6.Gender														
7.Veteran Status														
8.Disabling Condition														
9.Residence Prior to Program Entrance														
10.Zip Code of Last Permanent Residence														
11.Homeless Status														
12.Program Entry Date														
13.Program Exit Date														
14.Personal Identifier (Unique ID)														
15.Household Identifier														

<https://www.hudexchange.info/resource/1216/hmis-data-quality-monitoring-tool/>

Data Quality vs Quality Data

“Our completion rate is 95%!”

That is great. Is that the whole story?
Does this mean that data quality has been
achieved?



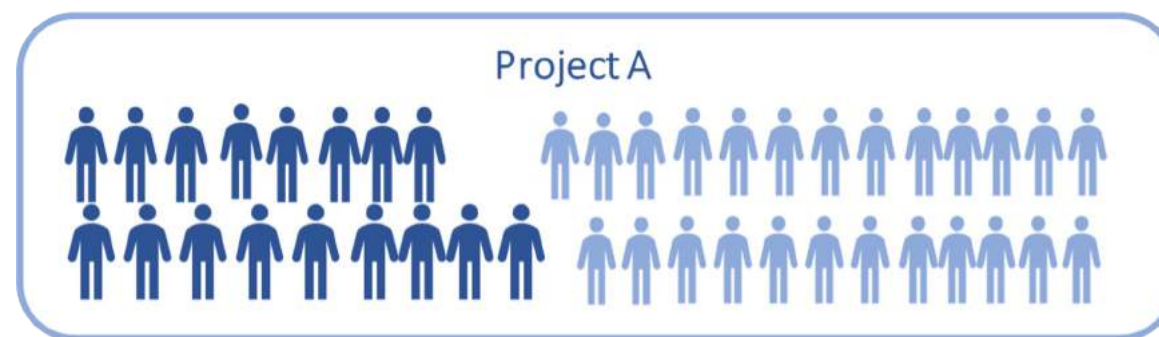
Known Project with Some Data for *All* People Served

In this graphic:

HMIS has the Project Data
Descriptor information

HMIS has all of the data for
some of the people served.

And some of the data for all of
the people served



All data

Some data

This is the scenario that data completeness is designed
to measure.

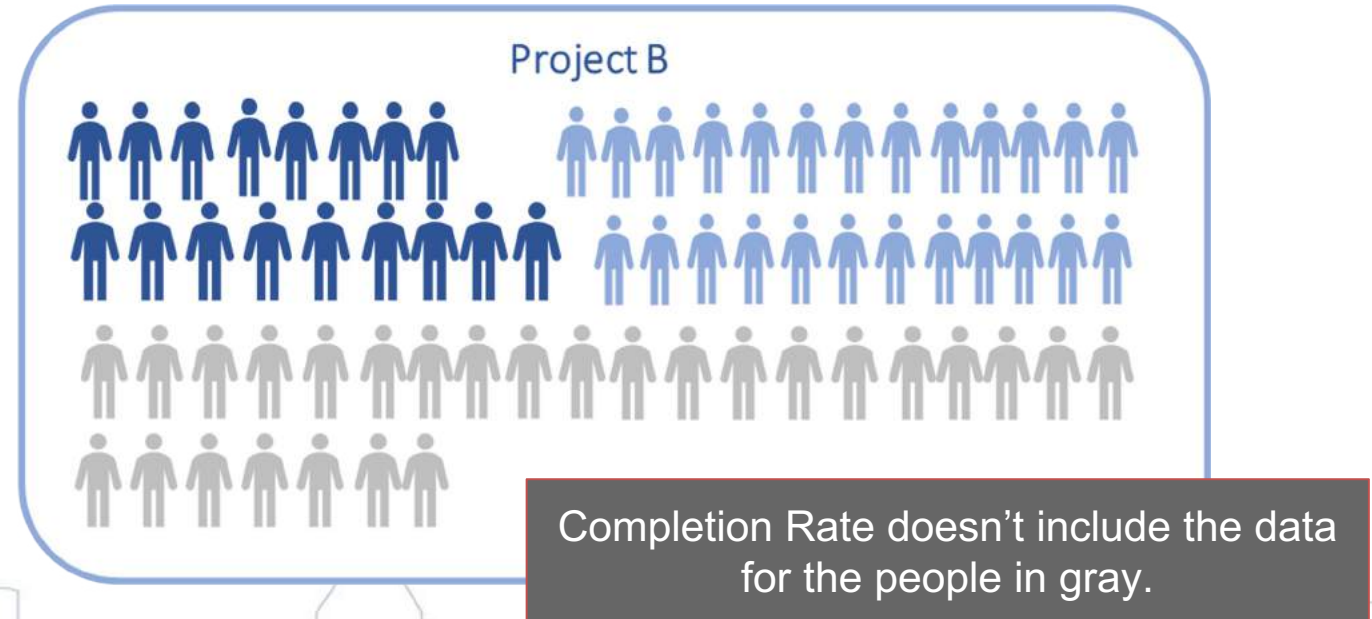
Known Project with Some Data for *Some* People Served

HMIS has the Project Data
Descriptor information

HMIS has all of the data for
some of the people served.

And some of the data for
some of the people served

And has none of the data for
some of the people served



All data

Some data

No data

Known Project with No Data for People Served

HMIS has the Project Data Descriptor information

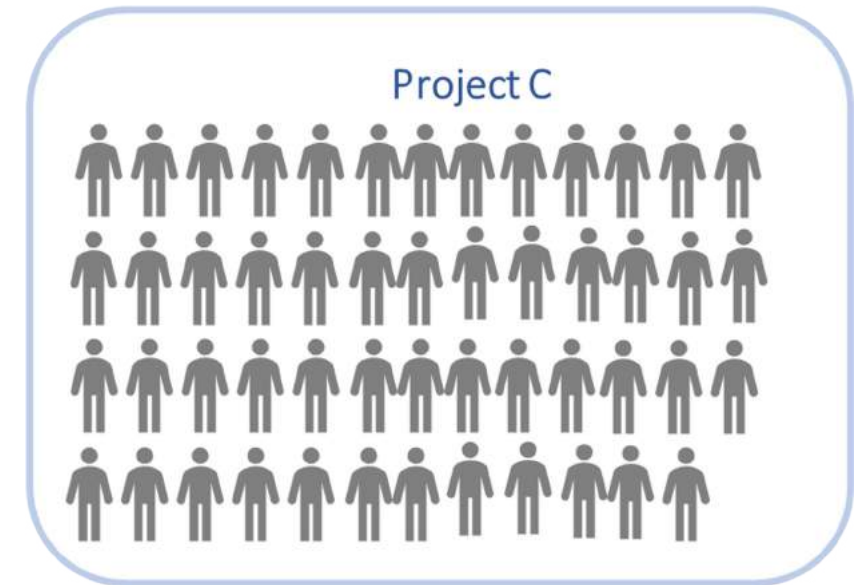
Has none of the data about people served

Either:

Project is not intended to participate in HMIS

OR

Project is not participating but is expected to



No data

Completion Rate doesn't include the data for the people in gray.

Project Not Set Up in HMIS

HMIS is missing the
project descriptor data

Project D

No data

Completion Rate doesn't include the data
for the people in gray.

Accuracy

What is accuracy? The degree to which data reflects the real-world client or service.

How do we know if the data are accurate? Accuracy is dependent on completeness and timeliness but also performing logical checks of your data to uncover where there may be inaccuracies that need attention.

Are there tools to help? [HMIS CSV Data Quality Tool](#) can check your HMIS CSV for logic related issues.

HMIS CSV Data Quality Tool

- Checks CSV to ensure needed files are present
- Checks core data fields to ensure only valid values
- Performs 14 types of logic based data quality checks such as:
 - Missing/Invalid DOB
 - No/Multiple Heads of Household
 - Possible missed exits dates for shelter and street outreach
 - Overlapping Enrollments

You can view the records by each check within the tool via pop up window or click a button to export a list of all records being flagged.

The screenshot displays the HMIS CSV Data Quality Tool interface. On the left, a 'Flagged IDs' window is open, showing a table of enrollment records with overlaps. The table has columns: PersonalID, ProjectID, EnrollmentID, ProjectName, ProjectType, EntryDate, ExitDate, ProjectID-B, and EnrollmentID. The first few rows are:

PersonalID	ProjectID	EnrollmentID	ProjectName	ProjectType	EntryDate	ExitDate	ProjectID-B	EnrollmentID
62156	240	168973	DataLab - Em	1	12/7/2017	12/14/2017	1099	168977
62168	240	168795	DataLab - Em	1	1/11/2017	1/23/2017	1099	168798
62226	1099	169454	DataLab - VA	2	5/4/2019	10/3/2019	240	169591
62226	1099	169454	DataLab - VA	2	5/4/2019	10/3/2019	240	169661
62226	240	169343	DataLab - Em	1	2/11/2019	5/6/2019	1099	169454
62476	240	169373	DataLab - Em	1	3/9/2019	3/25/2019	1099	169390
69679	958	172614	DataLab - Em	1	1/2/2017		958	172615
69720	958	172544_2537	DataLab - Em	1	12/3/2016	5/14/2017	960	176924
73920	960	177928	DataLab - Em	1	4/30/2019	5/6/2019	960	177927
74479	960	178841	DataLab - Em	1	3/31/2020	4/1/2020	960	178840
74489	960	178855	DataLab - Em	1	9/5/2017	9/9/2017	960	178854
76773	242	181474	DataLab - Tre	2	5/1/2020	8/1/2021	240	181420
76774	242	181475	DataLab - Tre	2	5/1/2020	8/24/2021	240	181421
76775	242	181475_2611	DataLab - Tre	2	5/1/2020	8/1/2021	240	181421
76776	242	181476	DataLab - Tre	2	5/1/2020	8/1/2021	240	181422
76777	242	181477	DataLab - Tre	2	5/1/2020	8/1/2021	240	181423
76778	242	181478	DataLab - Tre	2	5/1/2020	8/24/2021	240	181424
76779	242	181479	DataLab - Tre	2	5/1/2020	8/1/2021	240	181425
76780	242	181478_2611	DataLab - Tre	2	5/1/2020	8/1/2021	240	181424
76781	242	181480	DataLab - Tre	2	5/1/2020	8/1/2021	240	181426
76782	242	181480_2611	DataLab - Tre	2	5/1/2020	8/1/2021	240	181426

On the right, the main results panel shows a summary of flagged records. It includes a table with columns: 'Flagged Records', 'Double click a cell in this column to review the underlying records by double clicking the result. Note you with contextual information in addition to the it is likely the entity with the most recent information to', and 'Export All Flagged Records'. The table shows:

Flagged Records	Double click a cell in this column to review the underlying records by double clicking the result. Note you with contextual information in addition to the it is likely the entity with the most recent information to	Export All Flagged Records
276 of 7,045 enrollments		
1 of 7,045 enrollments		
48 of 7,045 enrollments		

A green circle highlights the 'Export IDs' button in the 'Flagged IDs' window. A blue arrow points from the '48 of 7,045 enrollments' row in the main results panel to the 'Export IDs' button.

SNAPS Data Strategy and Data Quality

SNAPS

Data TA Strategy to Improve Data and Performance



SNAPS Data Strategy and Data Quality

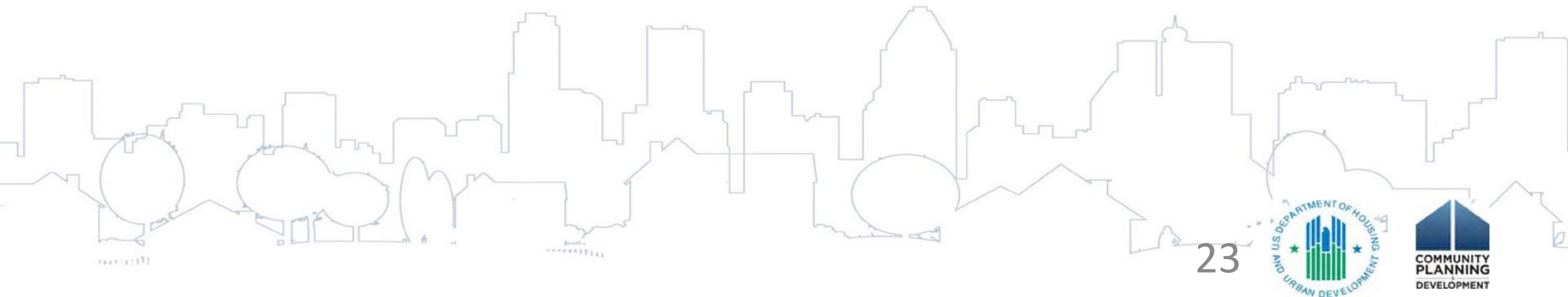
- SNAPS Strategy is intended to be aspirational and not used to monitor projects for compliance
- Focus on ensuring CoCs have data-driven local planning to work towards ending homelessness
- CoCs, HMIS Leads, and Organizations work together to review the strategy and set local goals and performance indicators

<https://files.hudexchange.info/resources/documents/SNAPS-Data-TA-Strategy-to-Improve-Data-and-Performance-Overview.pdf>

SNAPS Data Strategy and Data Quality

3 specific strategies and today, we will highlight Strategy #2, as it focuses on data quality

**DATA SYSTEMS COLLECT ACCURATE
COMPREHENSIVE AND TIMELY DATA**



SNAPS Data Strategy and Data Quality

STRATEGY 2:

Data systems collect accurate, comprehensive and timely data

Characteristic: Bed Coverage across Continuum (funded and unfunded)		
Current status (or near-term)	Majority of CoCs in 3-5 years	Advanced CoCs in 3-5 years
<ul style="list-style-type: none">100% required providers and less than 100% non-required providers contribute to HMIS	<ul style="list-style-type: none">100% all homeless service providers contribute to HMIS	<ul style="list-style-type: none">100% homeless providers and non-homeless service providers contribute to shared data environment
Characteristic: Quality data = Timely Accurate Comprehensive		
Current status (or near-term)	Majority of CoCs in 3-5 years	Advanced CoCs in 3-5 years
<ul style="list-style-type: none">100% completeLess than 100% accurateNo timeliness standard	<ul style="list-style-type: none">100% accurate100% completeProjects directly enter data within 2 hours for crisis response and project start/project exitPSH projects directly enter data within 24 hours	<ul style="list-style-type: none">100% accurateCustomized local data elements are limited to information needed to report, serve and house clients.100% completeTimeliness meets Coordinated Entry placement and referral needs

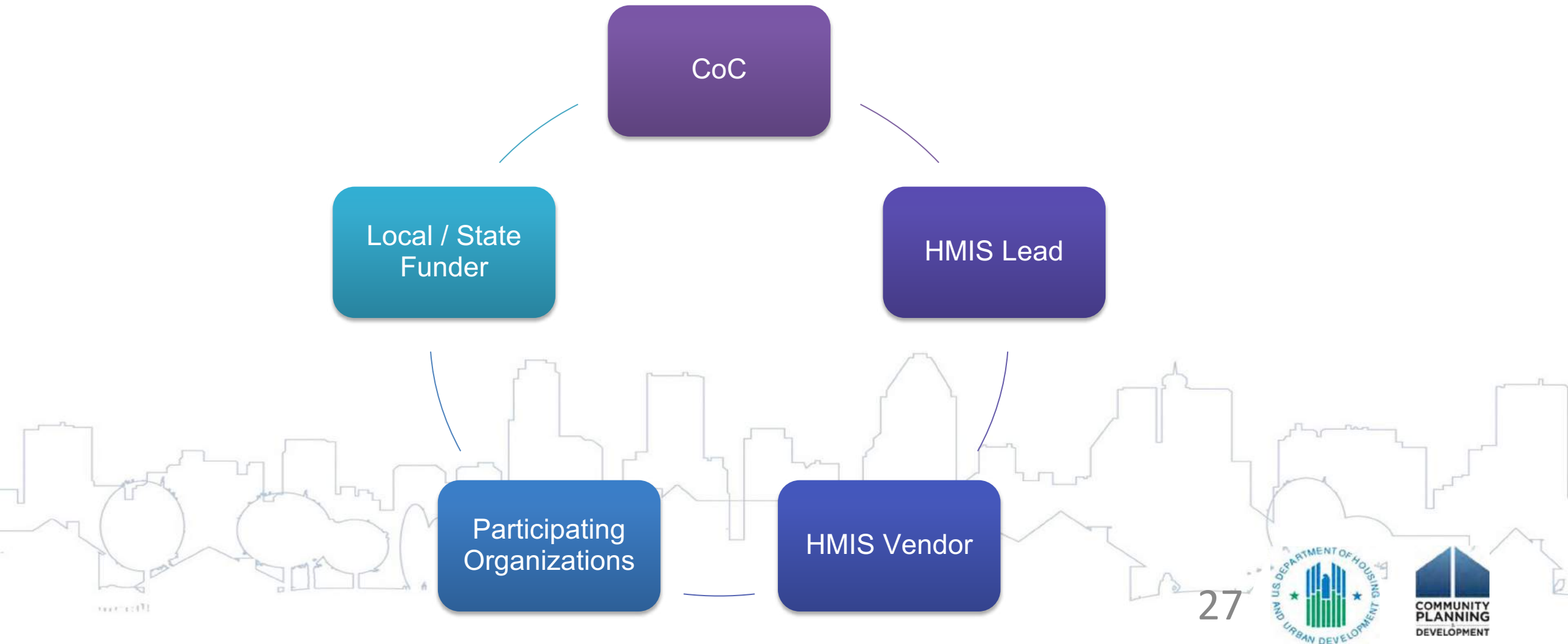
IMPACT OF DATA QUALITY



Why Does Data Quality Matter?

- Reflects the reality of homelessness
- Reports are affected regardless of whether the project is federally funded
- Making data-informed community decisions
- Assessing and prioritizing individuals for the coordinated entry process

Who's Involved in Data Quality?



Who “Bakes” Your Data?

TIMELINESS

OVER-BAKED



- cookies are dry and crumbly
- results in a toasted flavor

NO EGGS



- taste too sweet
- cookies are dry on the outside and doughy on the inside

COMPLETENESS

OVER-CREAMED



- cookies become tough and hard

EVERYTHING MIXED TOGETHER AT ONCE



- results in inconsistent texture and flavor

ACCURACY

CONSISTENCY

Data Quality and Equity Analysis

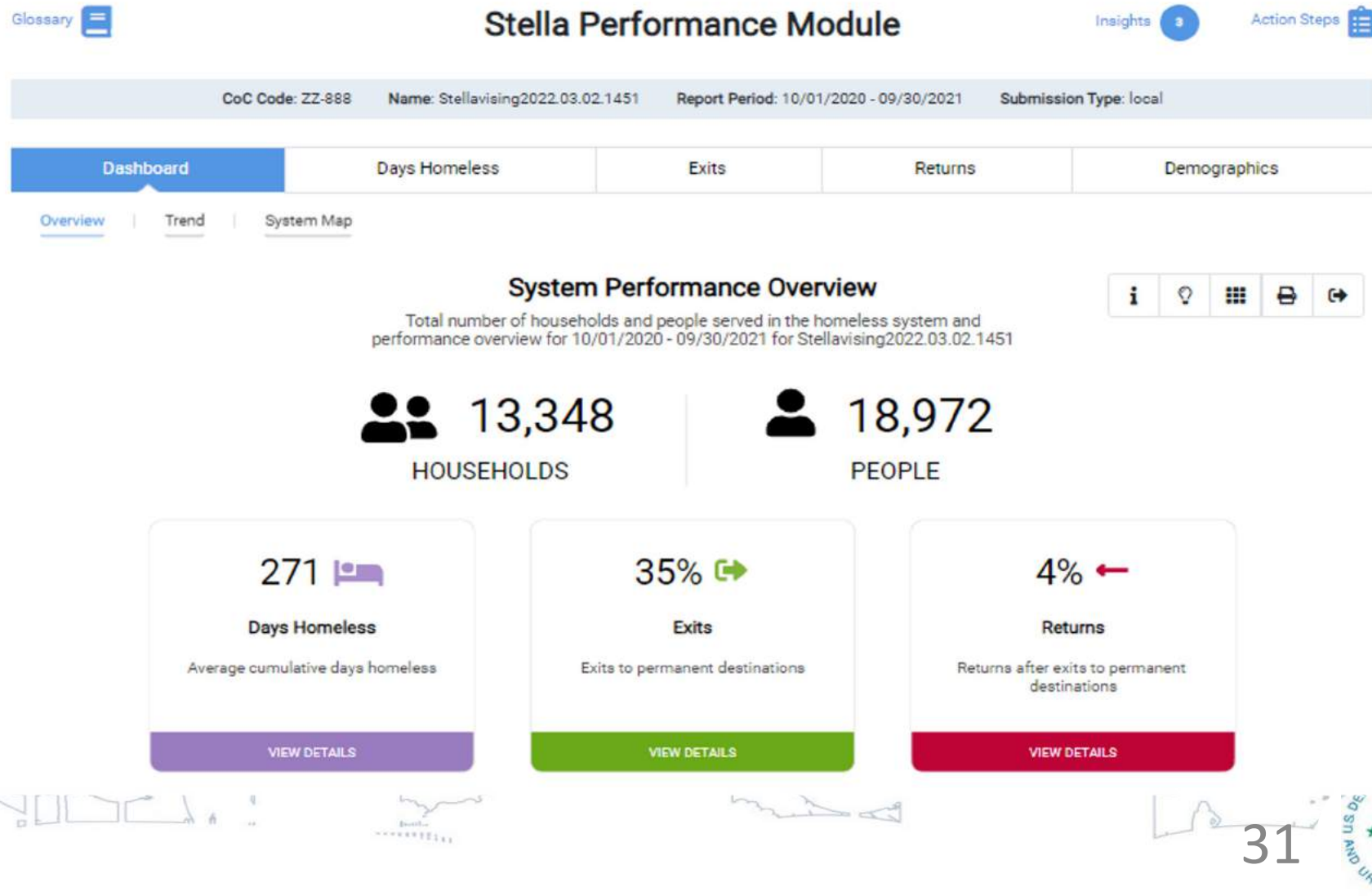
- Improper or inconsistent data collection can impact a community's ability to analyze if their homeless system is functioning equitably or not.
 - Example: Staff observation vs. Client self identification
 - Example: “What’s your sex” vs “Which of these genders best describes how you identify?”
 - Missing date of birth
 - Frequent responses of “Data not collected,” or “Client refused,” equates to missing data.
 - Client can refuse to answer, agency CAN NOT refuse to ask

Save Yourself From The Dreaded LSA Flags

- Project Descriptor Data Elements
 - Project
 - Inventory
 - Grant
- People/Household Data
 - Date of Birth, Social Security Numbers
 - Chronically Homeless
 - Head of Household
 - Improper Entries, Move-Ins, or Exits
 - Project Dates
 - Overlapping Enrollments

<https://files.hudexchange.info/resources/documents/Preparing-for-the-LSA-Guidance-on-Common-Data-Quality-Issues.pdf>

Stella P



THANK YOU!

- Jesse Jostad
 - Jesse_Jorstad@abtassoc.com
- Melissa Mikel
 - Melissa.Mikel@icf.com
- Mindy Mitchell
 - mindy@homebaseccc.org

HUD Certificate of Completion

HUD is offering a certificate of completion for anyone who completes any four of the following six sessions at this NHSDC conference.

- HMIS Project Set-up 201
- HMIS Project Management and Annual Calendar of Expectations
- HMIS Governance 101
- Protecting Data in an HMIS Environment: Privacy, Security, and Confidentiality
- Implementing Effective Contract Negotiation and Relationship Management Strategies 201
- Data Quality 101: What is DQ?

Participants **must** complete the session evaluations for at least four sessions to receive their certificate.

Rate this Session!

