GOMO Data Strategy

Ann-Christine Zinkann, Cynthia Garcia, Alyse Larkin

GOMO Community Workshop 2023

GOMO PROGRAM

GOOS

US contributions since 2017

Deployment locations (floats, drifters, gliders, moored buoys, animals, sail drones) and
lines performed (hydrography, XBTs) since 2017
Latest locations for ship-based weather stations operating since 2017

- Ship-based weather stations (2271)
- XBT lines (18)
- Repeat Hydrography lines (8)
- Profiling floats (2117)
- Surface Drifters / Ice Buoys (4581)
- Glider missions (624)
- Sail drones (37)
- Animal Borne Sensors (45)
- Tropical Moored Buoys (362)
- Coastal/National MB
- Tsunameters (33)

Generated by ocean-ops.org, 2022-06-24
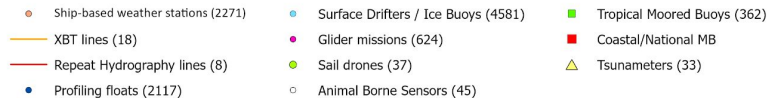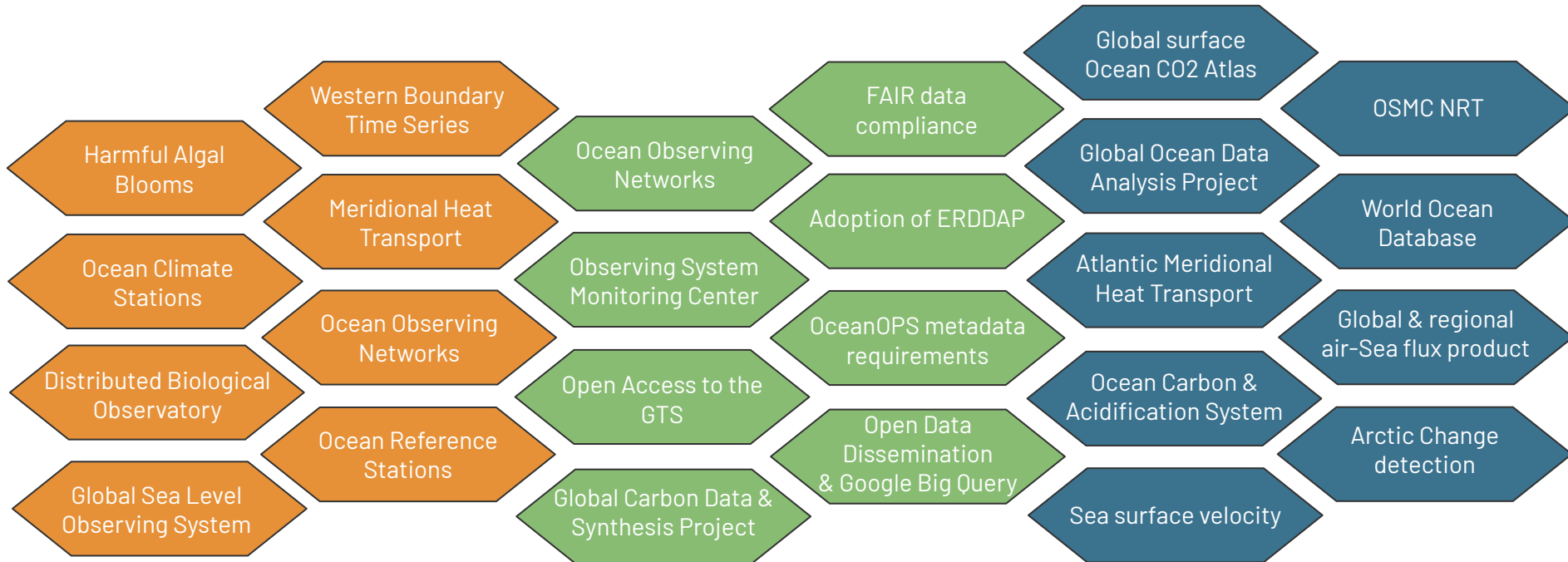Projection: Plate Carree (-150,0000)

- GOMO directly supports over **3000 ocean observing platforms** and the provision of **millions** of measurements of the ocean

- Observational data provided to **data facilities** (DAC's, NCEI etc.)

- Directly ingested and used by weather, ocean and climate **models** (GTS)

SNAPSHOT OF DATA EFFORTS

DATA COLLECTIONS

DATA ACTIVITIES

DATA PRODUCTS

- Harmful Algal Blooms
- Western Boundary Time Series
- Ocean Climate Stations
- Meridional Heat Transport
- Distributed Biological Observatory
- Ocean Observing Networks
- Global Sea Level Observing System
- Ocean Reference Stations

- Ocean Observing Networks
- FAIR data compliance
- Adoption of ERDDAP
- Observing System Monitoring Center
- OceanOPS metadata requirements
- Open Access to the GTS
- Open Data Dissemination & Google Big Query
- Global Carbon Data & Synthesis Project

- Global surface Ocean CO2 Atlas
- OSMC NRT
- Global Ocean Data Analysis Project
- World Ocean Database
- Atlantic Meridional Heat Transport
- Global & regional air-Sea flux product
- Ocean Carbon & Acidification System
- Arctic Change detection
- Sea surface velocity

MOTIVATION & PROGRAM REVIEW

Comprehensive review of the GOMO program (July 11th - 14th, 2022) with a designated focus area on "Information and product development to ensure data access and usability; and to increase the value/impact of ocean observation"

Recommendations include:

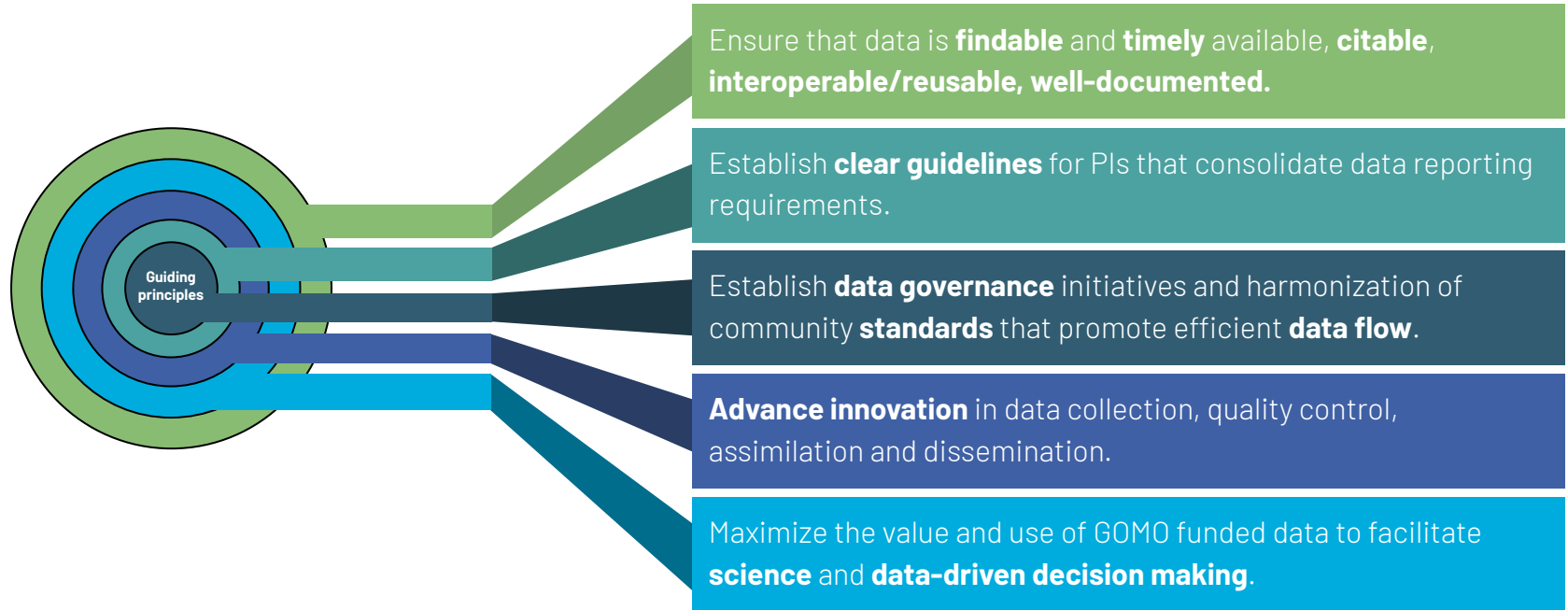|	
Data Strategy	Develop a **coherent** and **long-term** data management strategy and policy
Engagement	Increase engagement in **international data management** & **infrastructure landscape**
Metrics	Establish **metrics** to measure progress in data management
System	Consider establishing a **simple "findable" system**
Existing efforts	Capitalize on efforts & existing expertise in **data quality** across observing systems

DATA STRATEGIC PLAN

- Strategy to describes the **vision, goals**, and **objectives** for establishing the foundation to effectively manage, share, and maximize the value of our data assets

- **Created** and **championed** by the GOMO Data Community

- **Aligns** with GOMO and NOAA's strategic vision and priorities

- Intended to **guide** the development of an implementation plan

- A **living document**, enabling agility in the face of technological and data management innovations
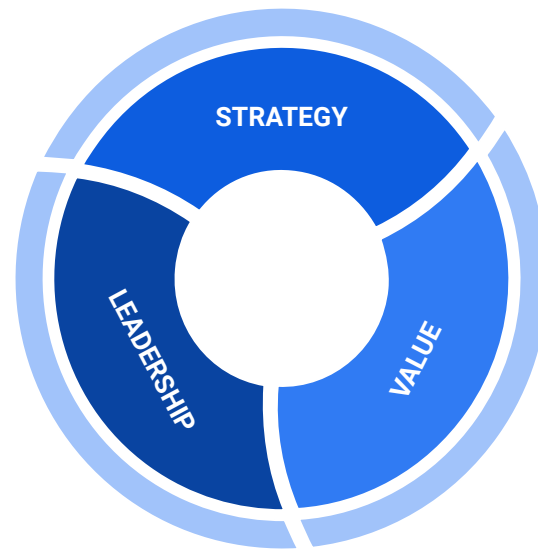
THE VISION

Develop streamlined, robust and integrated data practices that support and harmonize the GOMO data ecosystem.

Guiding principles

Ensure that data is **findable** and **timely** available, **citable, interoperable/reusable, well-documented.**

Establish **clear guidelines** for PIs that consolidate data reporting requirements.

Establish **data governance** initiatives and harmonization of community **standards** that promote efficient **data flow**.

Advance innovation in data collection, quality control, assimilation and dissemination.

Maximize the value and use of GOMO funded data to facilitate **science** and **data-driven decision making**.

STRATEGIC PILLARS

Covers objectives and initiatives surrounding policies and people
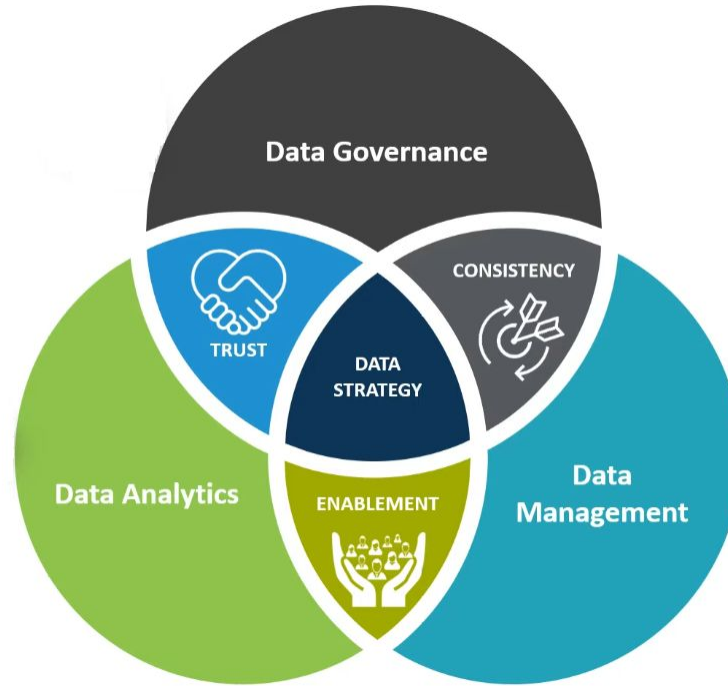
Initiatives to gain insight, make informed decisions, and drive innovation that aligns with the GOMO's vision and objectives.

Activities applying data governance policies to ensure data across GOMO's portfolio are FAIR

DATA MANAGEMENT SURVEY

NOAA
GLOBAL OCEAN
MONITORING & OBSERVING

Department of Commerce | National Oceanic and Atmospheric Administration

DATA MANAGEMENT SURVEY

MOTIVATION

- Understand the current state-of-play of GOMO's existing data management, challenges, and opportunities

- Gather insights and recommendations for informed decision-making

- Contribute towards GOMO's draft data strategy plan

How do you manage your data?

How much are we investing on data management per project?

Are we tracking usage and impact of our data?

Results presented here are from 14 respondents. Survey is still open for input from others.

SURVEY RESULTS

Direct **data collection** that requires data management
  86% of PIs

Processing latency

Real-time	6
Week to 1 month	5
1-3 years	3
6 -12 months	2
Uncertain	1

Existing quality requirements

Mostly yes

Data Stewards

86% Yes
14% No

Key data sources

Research Vessels 24.0%
Ocean observation 28.0%
Ocean buoys / Moorings 20.0%
UxS 8.0%
Model and reanalysis 8.0%
Others 12.0%
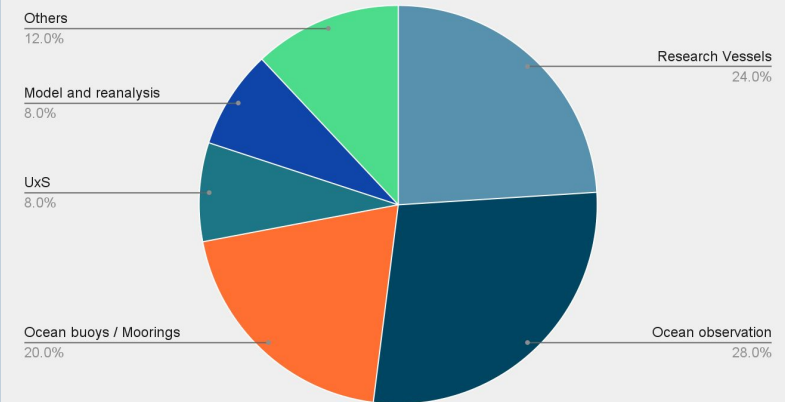
Others: satellite data, submarine telecom cables

Established **metadata standards** that are being followed include WMO standards, ISO 19115, OceanOPs, NCEI NetCDF templates (builds off of ACDD and CF conventions), IOOS metadata profile, COARDS compliant NetCDF, SOCAT, resulting ISO, CF and ACDD metadata are provided with the data and are accessible via ERDDAP

SURVEY RESULTS

DATA MANAGEMENT GAPS

Data standardization and access:

- Need for a clear pathway for upgrading legacy datasets to modern standards.
- The heterogeneity of data formats from different contributors requires substantial efforts to standardize, impacting their reusability and interoperability.
- Ambiguity about compliance with NOAA's requirements in regards to data archival and access.
- Significant delays often occur in the processing of data to an archive and getting a live link, delaying subsequent research publications.

Quality Control and Metadata:

- Many have reported problems with data quality flags and the lack of homogeneity in formats.
- Poor or missing metadata.

Data Management Personnel and Resources:

- Shortage of personnel and resources dedicated to data management.
- Several pointed out the time-consuming process of archiving data with NCEI due to lack of adequate support.
- Concerns raised about the distribution of responsibility and governance structure for the data pipeline.
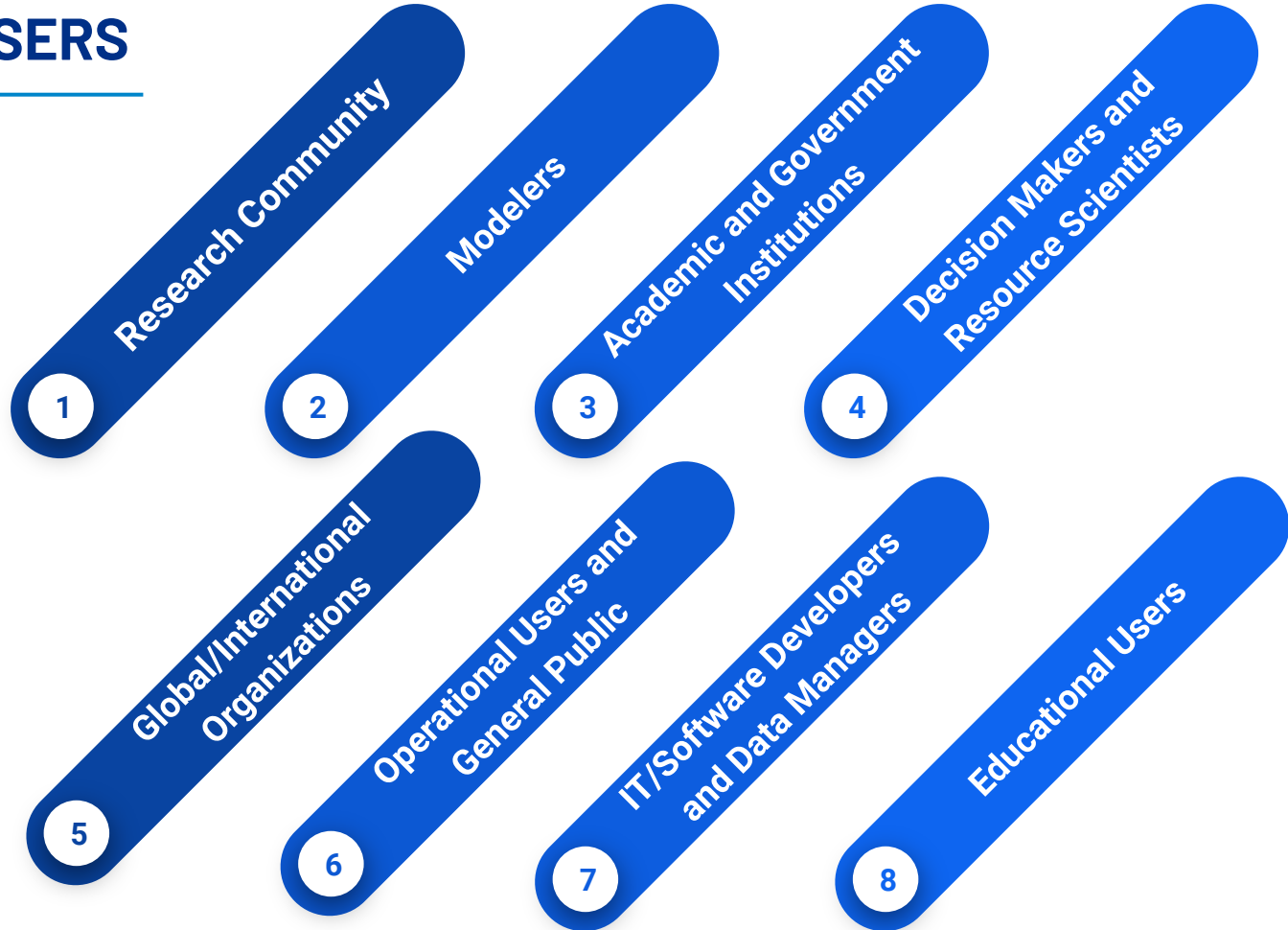
SURVEY RESULTS

DATA MANAGEMENT RECOMMENDATIONS

Funding and Resources	Funding required to for personnel for data management and infrastructure
Standardization	Implementing agreed data and metadata standards
Data Archiving	Streamline data archiving process and standards for updating data sets with DOIs
Community Agreement	Support machine-to-machine data interoperability
Data Access and Visibility	Promote visibility and ease of access, a landing page or hub for various projects
Quality Control	Simplifying the processing pipeline and improve quality control measures
Interoperability	Data Assembly Centers (DACs) to be synchronized

PRIMARY USERS

1 Research Community

2 Modelers

3 Academic and Government Institutions

4 Decision Makers and Resource Scientists

5 Global/International Organizations

6 Operational Users and General Public

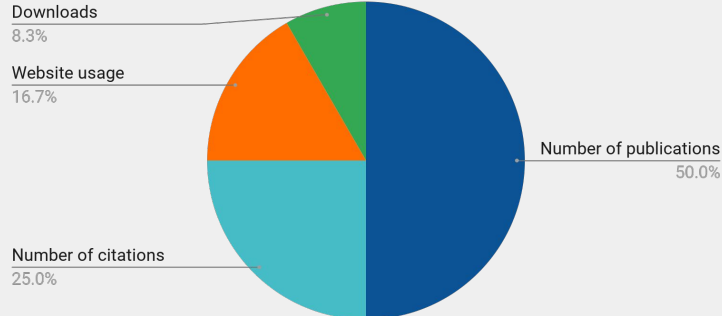**7** IT/Software Developers and Data Managers

8 Educational Users

SURVEY RESULTS

Are you using metrics of success to track impact?
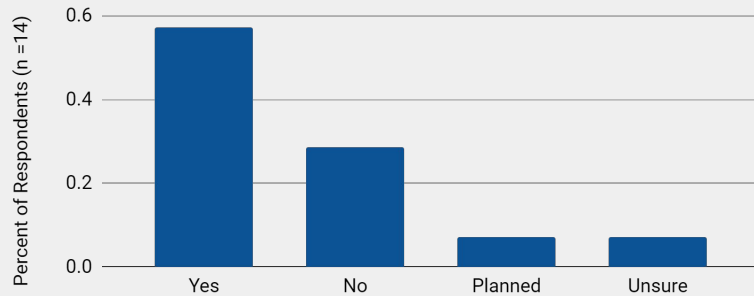
Identify gaps related to measurement of impact:

- Access to data use/download statistics (n = 5)
- Tracking data provenance with persistent identifiers (n = 2)
- Consistent citation of data by users (n = 1)
- Funding (n = 2)

Metrics used:

- Downloads 8.3%
- Website usage 16.7%
- Number of publications 50.0%
- Number of citations 25.0%

Data analytics:

- 2 projects link to or provide web-based analytics products (i.e. dashboards, portals, etc.)
- 3 projects had future plans for online data analysis or visualization software
- 9 projects needed further support (funding, training, personnel) to provide these services

BUDGET AND NEEDS

- Allocation towards data management varies widely. Ranges from **5%** of their budget to as much as **100%**.
- Average percentage of budget allocated to data management is approximately **29%**
- The dollar amount dedicated to data management ranges from **$20,000 to $265,806**, with some projects indicating year-to-year variation.
- Average dollar amount allocated to data management is approximately **$107,076**
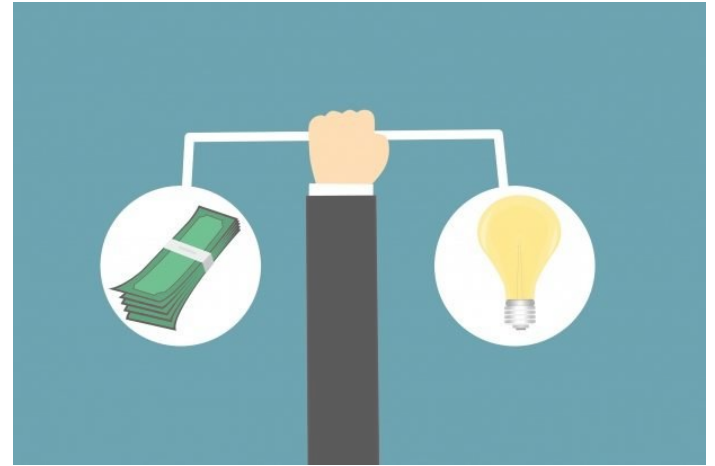
Need for at least one additional FTE, suggesting a general desire for increased personnel to manage data-related tasks.

Need for increased funding to tackle various data management gaps, with a focus on updating infrastructure, software development, and increased personnel.

WORKSHOP KEY OBJECTIVES

- **Co-develop** the **GOMO Data Strategy** with PIs

- Intended to **guide** the development of an implementation plan

- **Advance key sections** of the data strategy and set the stage for breakout sessions

- Identify **metrics, key performance indicators** (KPIs), and **tools** to monitor success and compliance

- Identify the **challenges, gaps, opportunities.** Propose solutions and offer thoughts on how can we bridge said gaps and move forward in a sustainable manner.

- GOMO **Data Governance Group (DGG)** and how to move forward.

Working Sessions

Time		
2:15 – 3:05	**Data strategy and governance** **Group 1 – Great Hall** Facilitators: Kevin O'Brien, Alex Kozyr, Liqing Jiang, Mathew Biddle	**Metrics of success and indicators** **Group 2 – Ellsworth Room** Facilitators: Mathieu Belbeoch, Eugene Burger
3:10 – 4:00	**Metrics of success and indicators** **Group 1 – Great Hall** Facilitators: Mathieu Belbeoch, Kevin O'Brien, Mathew Biddle	**Data strategy and governance** **Group 2 – Ellsworth Room** Facilitators: Eugene Burger, Liqing Jiang, Alex Kozyr

Breakout Assignments

In-Person:

Group 1 (Last name A-N)
Room: Great Hall

Group 2 (Last name O-Z)
Room: Ellsworth Room

Virtual: Randomly assigned

Please stay in the **same group** and **room** for the 2nd working session.
The facilitators will the one switching.
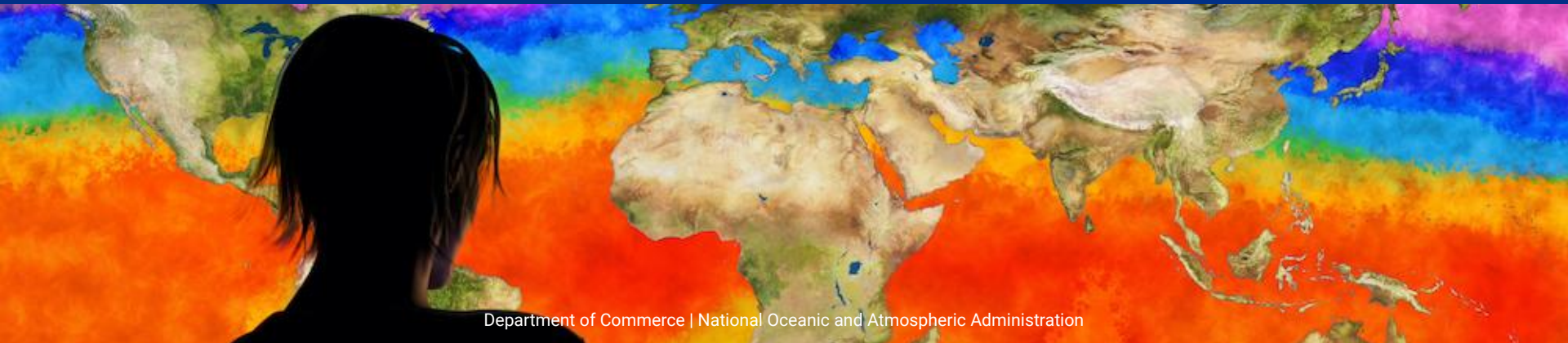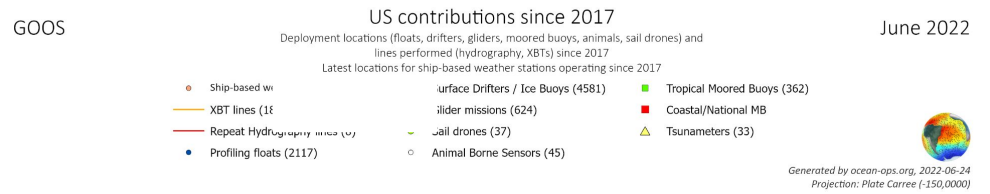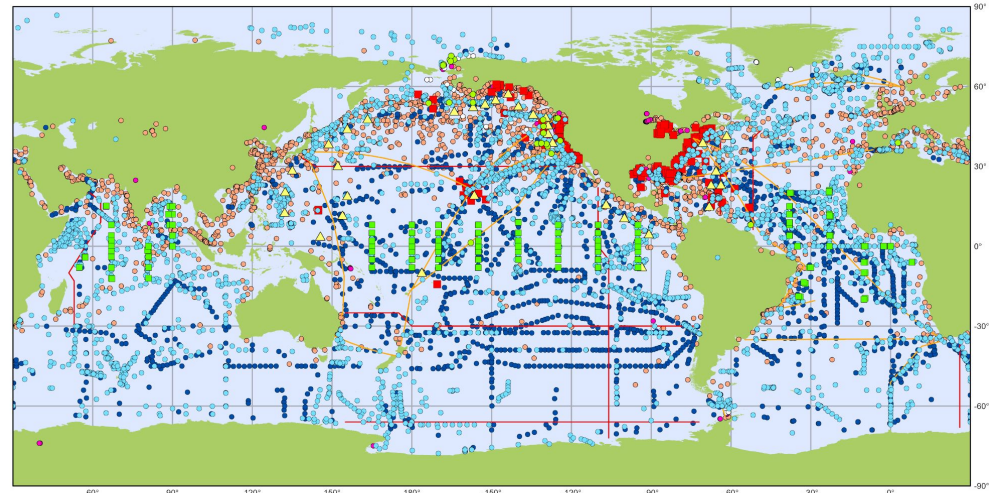
GOMO PROGRAM

GOMO directly supports over **3000 ocean observing platforms** and the provision of **millions** of measurements of the ocean

Data! Data! Data! Data! Metadata! Metadata! Metadata!

"**A coherent and long term data management policy is lacking.** "

"**...what is needed is a Program wide strategic and consistent approach.** "

- From the GOMO 2022 Program review

GOOS

US contributions since 2017
Deployment locations (floats, drifters, gliders, moored buoys, animals, sail drones) and lines performed (hydrography, XBTs) since 2017
Latest locations for ship-based weather stations operating since 2017

June 2022

- Ship-based w...
- XBT lines (18...
- Repeat Hydrography lines (6)
- Profiling floats (2117)
- ...urface Drifters / Ice Buoys (4581)
- ...lider missions (624)
- Sail drones (37)
- Animal Borne Sensors (45)
- Tropical Moored Buoys (362)
- Coastal/National MB
- Tsunameters (33)

Generated by ocean-ops.org, 2022-06-24
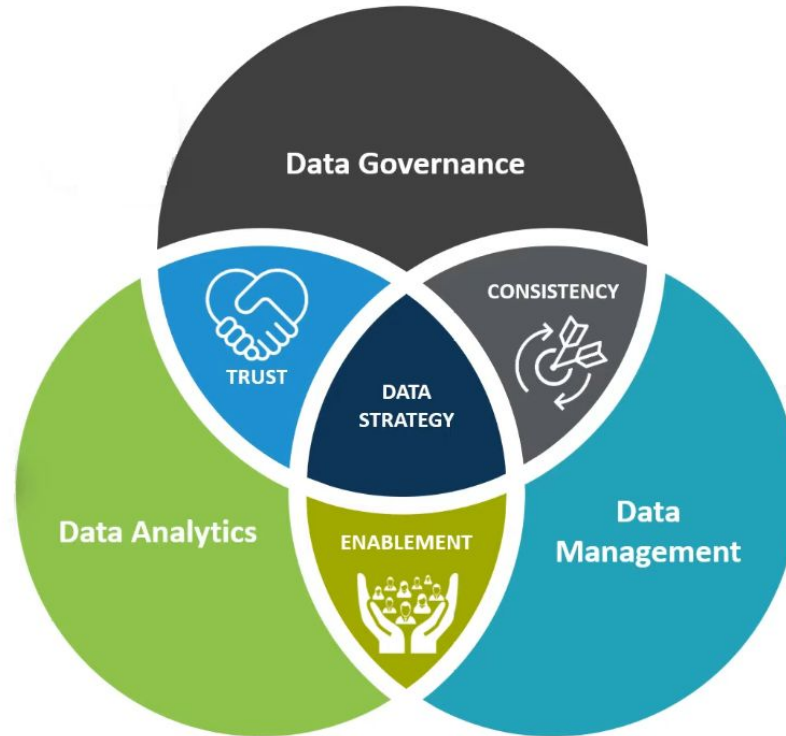*Projection: Plate Carree (-150,0000)*

Data Strategies

- NOAA Data Strategy

- IOC Strategic Plan for Ocean data and Information Management

- WMO Unified Data Policy

- UN Ocean Decade Data and Information Strategy

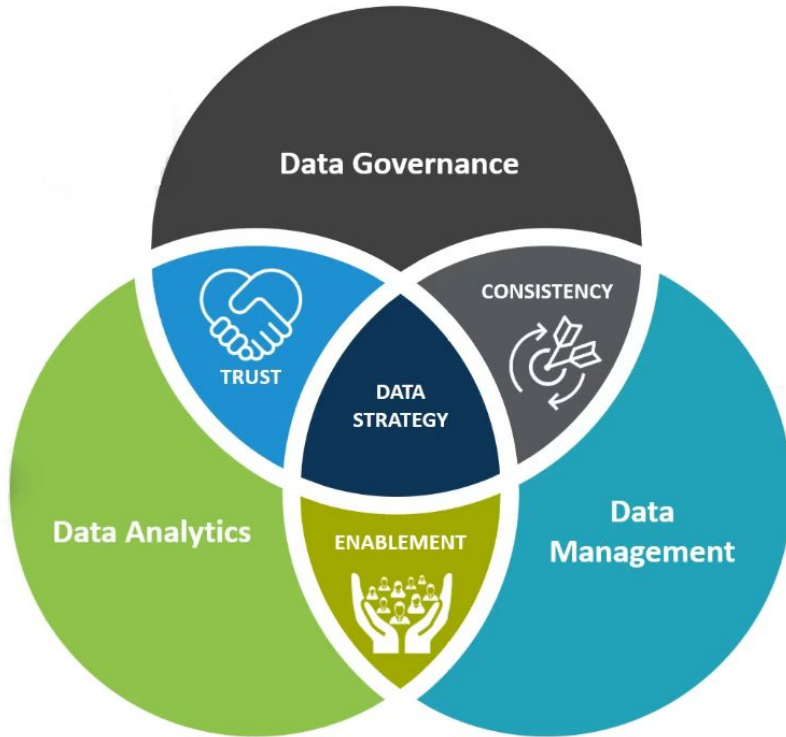- GOOS Observations Coordination Group Cross-Network Data Implementation Strategy
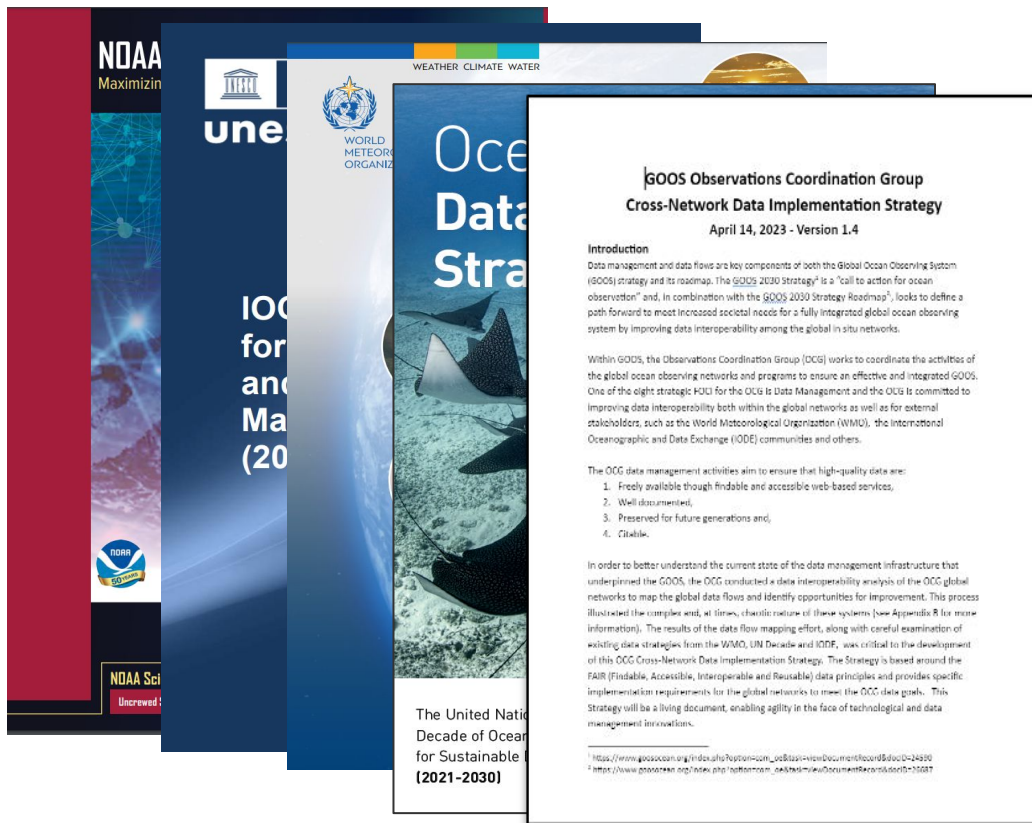
# GOMO Data Strategy

GOMO Data Strategy

Data Governance

- Maximizing discoverability, usability, quality and preservation of GOMO data
- Ensures data is trustworthy and used appropriately (licensing)

Potential Next Steps

- Establish a GOMO Data Governance Group (DGG)
  - Develop Implementation Strategy
  - Foster a data-centric culture through capacity building and communities of practice.
  - Identify metrics, key performance indicators (KPIs), and tools to monitor GOMO data flows

GOMO Data Strategy
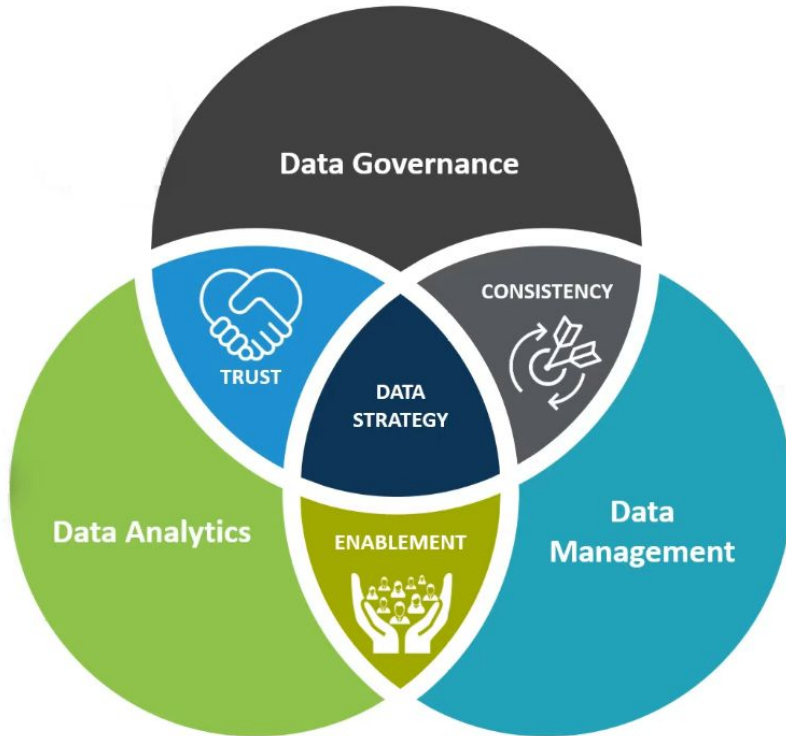
Data Management

- What steps need to be taken in our data systems in order to better comply with FAIR principles?
- In the face of increasing volumes and diversity of data, how do we ensure interoperability of data and data systems?
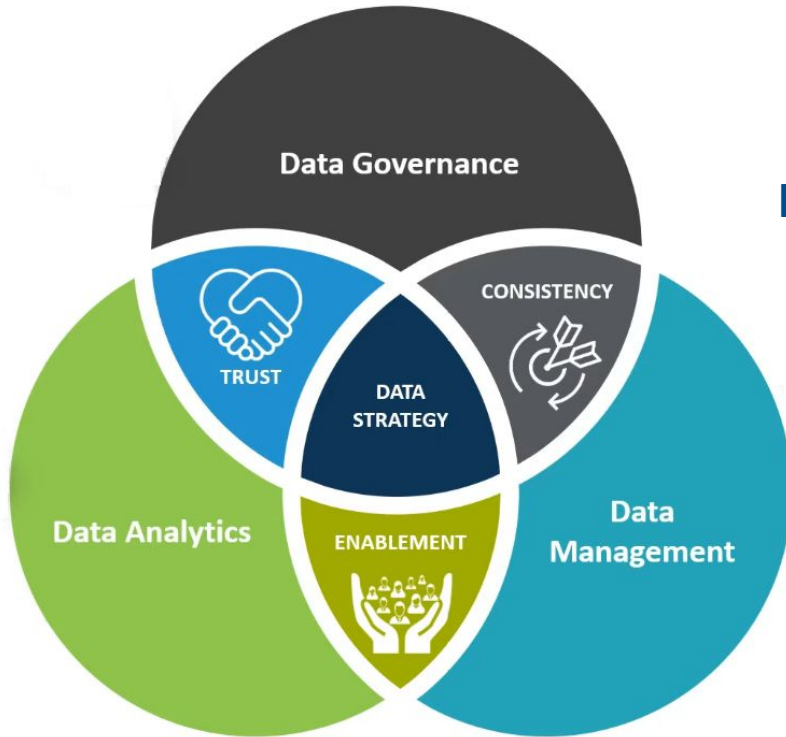
Potential Next Steps

- Ensure Data Management Plans address proper data collection, archiving, and sharing.
- Develop recommended list of data and metadata standards and conventions.
- Assess and leverage on existing services, portals and catalogs.
- Enhance data infrastructure and technologies.
- Harmonize data QC, data validation techniques to ensure the accuracy, reliability of collected data.
- Ensure that best practices are identified and documented.

GOMO Data Strategy

Data Analytics

- Gain insight, make informed decisions, and drive innovation

Potential Next Steps

- Identify key initiatives to leverage data for scientific research, informed decision-making, improved modeling and forecasting.
- Identify and develop needed knowledge products.
- Encourage data-driven collaborations, partnerships, and co-design.
- Develop way to track use of GOMO data and collect feedback from stakeholders

Working Sessions

Time		
2:15 – 3:05	**Data strategy and governance** **Group 1 – Great Hall** Facilitators: Kevin O'Brien, Alex Kozyr, Liqing Jiang, Mathew Biddle	**Metrics of success and indicators** **Group 2 – Ellsworth Room** Facilitators: Mathieu Belbeoch, Eugene Burger
3:10 – 4:00	**Metrics of success and indicators** **Group 1 – Great Hall** Facilitators: Mathieu Belbeoch, Kevin O'Brien, Mathew Biddle	**Data strategy and governance** **Group 2 – Ellsworth Room** Facilitators: Eugene Burger, Liqing Jiang

"**A coherent and long term data management policy is lacking.** "

"**…what is needed is a Program wide strategic and consistent approach.** "
- From the GOMO 2022 Program review

- Maximize value of GOMO data collected
- Ensure trust in GOMO data systems
- Support culture of open science
- Engage in and help lead implementation of the global digital data ecosystem
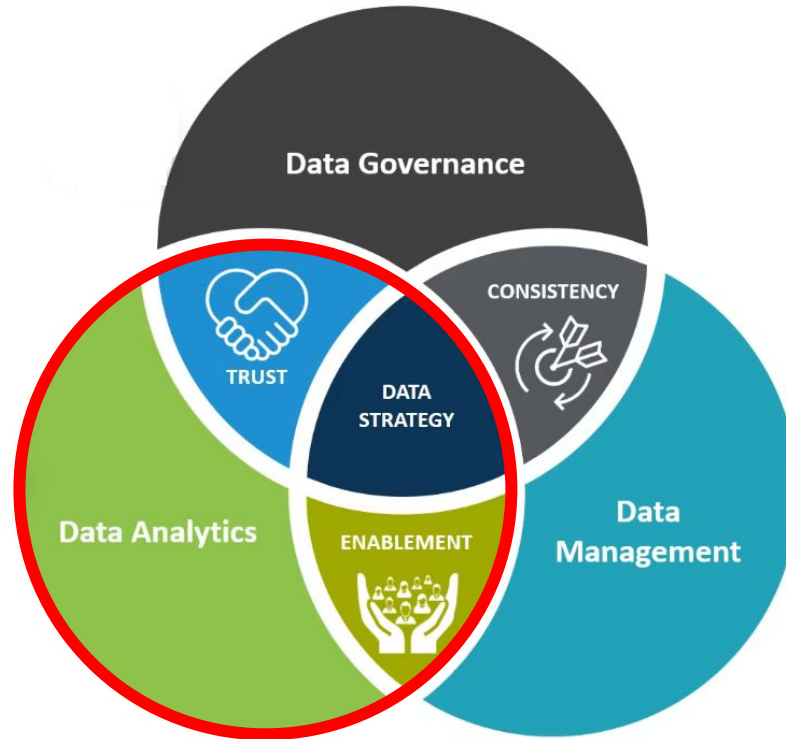
THANK YOU!

Data Management Metrics & Indicators

Mathieu Belbeoch, OceanOPS

GOMO Community Workshop 2023

GOMO Data Strategy

Requirements for Data Analytics

- GOMO components structuration

- Data pathways and responsibility for all components: RT, DM, Metadata, Unique identifiers

- PIDs: persistent identifiers for ships (ICES), platforms/missions/ (WMO/WIGOS), cruises: CRITICAL FOR INTEROPERABILITY – TO BE BOOKED AT VERY START via OceanOPS – and propagated in all data/metadata flows

- Data/Metadata sharing culture – no restriction (or visibility and customers base are reduced)
    - International guidelines (Unified WMO data policy)

    - National/Institutional: NOAA data holdings rules: planning, doc., archiving, citation, access

    - Network, GOOS/Observation Coordination Group (GOMO Data Strategy)

    - International context/emergency

Requirements for Data Analytics

- Monitor Global Data nodes: GTS of WMO, Internet GDACs, EOV/ECV gridded, WOD/WOA ...

  - Mapping and structuration required

- Monitor individual DATA PRODUCTION nodes (often national/institutional)

  - Data production is the key and requires clear workforce: decode/QC/code

  - Customization vs standardization: high vs low cost

  - Operationalization of workflows (timeliness)

  - Integrate / upgrade capabilities to BGC e.g. (not duplicate data centres)

  - Federated data processing capability (workflows, heldesk, github code sharing)

- Metadata will be critical to refine metrics by agency, operator, (sub) network, basin, EOV, etc.

Metrics/Indicators - Examples:

- Availability
- Quantity
- Quality
- Timeliness
- Data uptake

- Used to detect gaps, show success, show progress => improve efficiency

THANK YOU!