Survey Management Challenges

- Panel 1:

Top three challenges organizations are encountering in technology and survey computing

- **Panel 2:**

Challenges and approaches related to Data Governance

Challenges and Approaches Related to Data Governance

This panel will discuss approaches and challenges related to data governance. Because surveys both collect data, and use existing datasets, data governance requirements can have broad impact on survey management and operations. This panel will provide the opportunity for the audience to gain an understanding of data governance requirements, challenges in meeting those requirements, and enablers to being compliant without negatively impacting survey goals. Some of the components of data governance are: Data architecture, Data quality, Data modeling and design, Data storage and operations, Data security, Data integration and interoperability, and Metadata.

Challenges and Approaches Related to Data Governance

Moderator: Jane Shepherd – Westat

Panelists:

- Dan Gillman, BLS
- Karen Davis, RTI
- Ben Reist, NORC
- Dennis Pickett, Westat

Data Governance Done FAIRly

Dan Gillman

Office of Survey Methods Research

US Bureau of Labor Statistics

FedCASIC 2022 – April 5, 2022

Challenges and Approaches Related To Data Governance

Governance vs. Management

- Data Management
  - Primarily about the bits and bytes of data
    - Focus on representation of data
    - Not the meaning
  - IT concerns
    - Software – e.g., which RDBMS to use
    - Hardware – e.g., what server configuration is optimal
    - Extract, Transform, Load (ETL) functions

Governance vs. Management

- Data Governance
  - Primarily about the meaning of data
  - Focus on (among others)
    - Data quality
      - Editing, Analyses, Integration, Harmonization, Standardization
    - Discovery of data and data sets
      - Ability to find relevant data
    - Understanding
      - Well organized and complete documentation
    - Usage
      - Data for for intended use

Governance

- Statistical agencies do a lot
  - Data quality, especially
- But there's more to do
  - Not as advanced
    - Discovery, Understanding, Usage
  - Metadata lacking
    - No standards adopted or followed system-wide
    - Few sophisticated applications
- Need guiding principles

FAIR Principles

- Developed under scientific data community
- Broadly applicable
  - 4 main principles – data and metadata are
    - Findable
    - Accessible
    - Interoperable
    - Reusable
  - 15 guidelines underneath

FAIR Principles

- The principles refer to three types of entities:
  - Data
    - Data acquired through some statistical program
    - Used for further analyses or estimation
  - Metadata
    - Other data used to describe the data above
  - Infrastructure
    - Means to handle all the data and metadata
    - This is more data management than governance

FAIR - Findable

- The first step in (re)using data is to **find** them. Metadata and data should be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets and services.

- F1. (Meta)data are assigned a globally unique and persistent identifier
- F2. Data are described with rich metadata (defined by R1 below)
- F3. Metadata clearly and explicitly include the identifier of the data they describe
- F4. (Meta)data are registered or indexed in a searchable resource

FAIR - Accessible

- Once the user finds the required data, she/he/they need to know how they can be **accessed**, possibly including authentication and authorization.

- A1. (Meta)data are retrievable by their identifier using a standardized communications protocol
- A1.1 The protocol is open, free, and universally implementable
- A1.2 The protocol allows for an authentication and authorization procedure, where necessary
- A2. Metadata are accessible, even when the data are no longer available

FAIR - Interoperable

- The data usually need to be integrated with other data. In addition, the data need to **interoperate** with applications or workflows for analysis, storage, and processing.

- I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (Meta)data use vocabularies that follow FAIR principles
- I3. (Meta)data include qualified references to other (meta)data

FAIR - Reusable

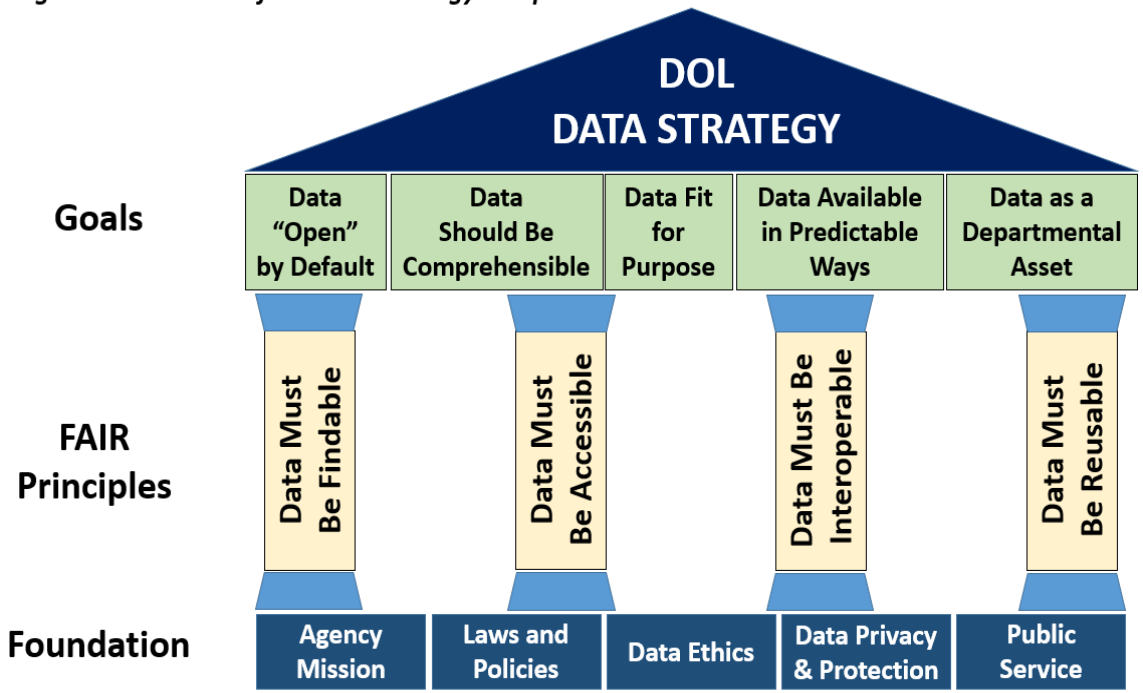
- The ultimate goal of FAIR is to optimize the **reuse** of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.

- R1. (Meta)data are richly described with a plurality of accurate and relevant attributes
- R1.1. (Meta)data are released with a clear and accessible data usage license
- R1.2. (Meta)data are associated with detailed provenance
- R1.3. (Meta)data meet domain-relevant community standards

Draft DOL Data Strategy

- Based on the FAIR principles

Figure 1: Schematic of the Data Strategy Components



Contact Information

Dan Gillman
Office of Survey Methods Research
US Bureau of Labor Statistics
Gillman.Daniel@bls.gov

Management Challenges
Data Governance Panel

FedCASIC April 2022

Karen M. Davis

VP, Research Computing

RTI INTERNATIONAL

Why data governance?

- "Data is the new Oil" – it fuels critical business decisions across the enterprise
- Ensures appropriate data management through the data lifecycle
  - Known provenance
  - Known quality
  - Sets processes, policies, standards, and roles

Data Lifecycle[1]

1 Stobierski, Tim, HBS online Blog, 02 February 2021

What is data governance?

A segment of data management concerned with **establishing a set of policies and procedures** for managing and protecting data assets within an organization.[2]

Oversight of activities that impact/govern data within an organization, including:

- Data architecture
- Data quality
- Data modeling & design
- Data storage and operations
- Data security
- Data integration & interoperability
- Data classification & GDPR compliance
- Meta-data

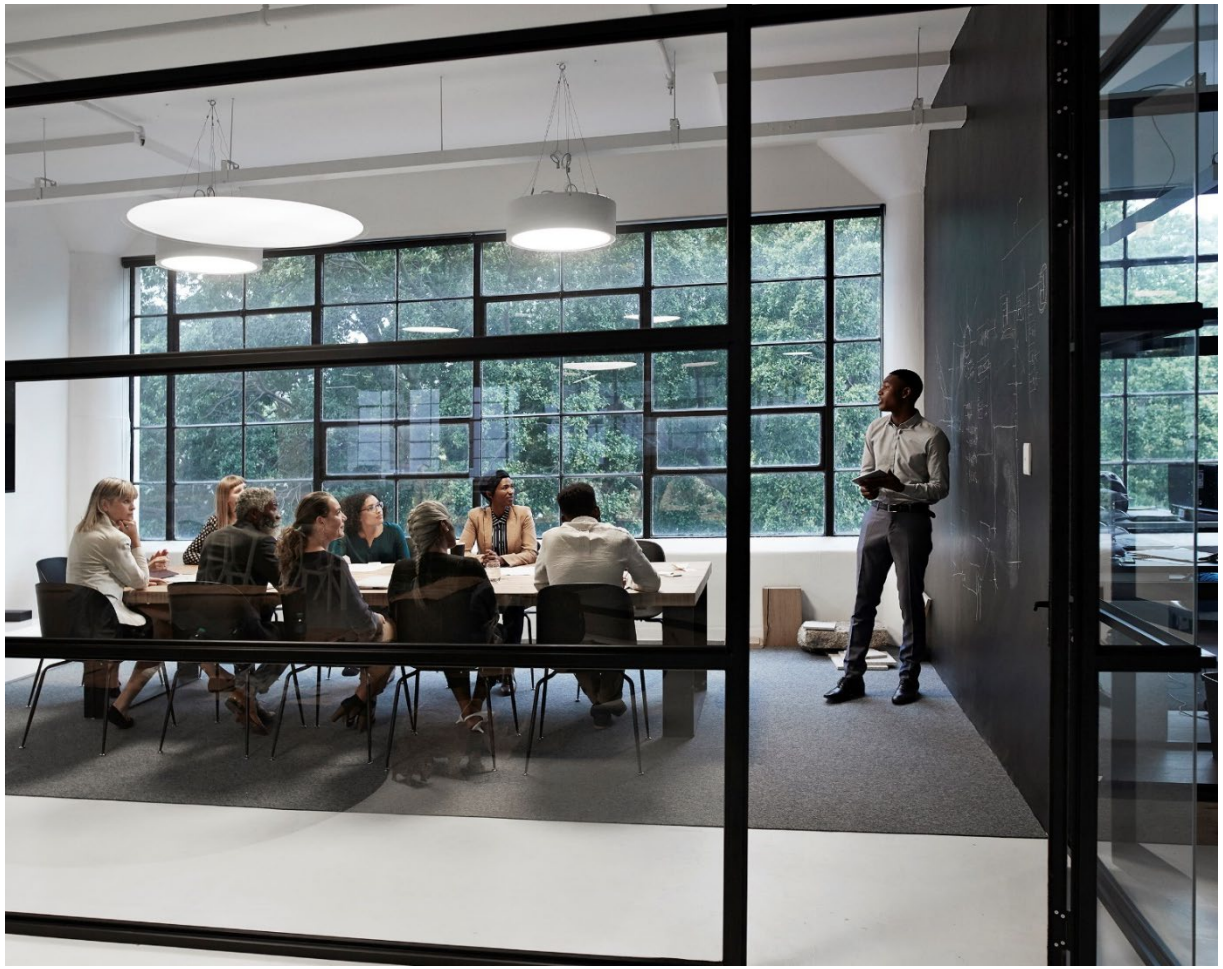
An RTI client summarized this as “understanding the provenance of the data”

2. Data Governance Institute website

Whose Data is it?

- Internal (examples)
  - Financial
  - HR
  - Employees
  - Clients
  - Contracts
  - Partners
- External
  - Project data – often belongs to the client
  - Controlled by external privacy and security requirements, some of which are regulatory in nature

Functions/Roles related to Data Governance



- Data Governance Council
- Risk officer
- Privacy officer
- Data stewards & owners
  - Internally focused
  - Project/client focused

RTI's Journey

- Data Governance Council created more than 4 years ago
  - Fully cross functional team include G&A groups and business unit representation
- Focus is Data Governance+ and has covered
  - Systems and network security requirements
  - Cloud usage, reviews, and approvals
  - Oversight of a variety of initiatives related to security, cloud, networks
  - Other related items such as policies
  - Data Classification initiative
  - GDPR compliance
- RTI also provides Data Governance services to clients in line with Federal Data Strategy and their data modernization initiatives
  - Working with NIH and other federal clients on projects related to making data FAIR (findable, accessible, interoperable, reusable)

Our journey continues

- Ensure appropriate data owners and stewards are assigned and trained in responsibilities
- Classify institute data into appropriate categories – Restricted, Private, or Public, according to risk
- Ensure management of all data following the data lifecycle
- Ensure quality and provenance of data on which business decisions are made
- Results in
  - Knowing where the data resides
  - Knowing how the data is classified and protected appropriately
  - Data classification is kept up to date
  - Data Stewards and owners are aware of their responsibilities
  - Data is managed in accordance with the data life cycle
  - Data is appropriately used in critical business decisions

Challenges and considerations

- Organization/management data vs Survey/client data
- Projects must meet client's security & compliance requirements
  - System security classification (per NIST 800-53) and ATO if required
  - Retention/destruction timeframes
  - Assumption of data quality assurance
  - Roles may be filled by project staff vs organizational titles for corporate
    - E.g. VP, Enterprise Risk, or Privacy officer vs Information Systems security office on project
  - Client data classifications may not match corporate definitions

Thank you

Contact: Karen M. Davis| email: kdavis@rti.org

FAIR Data for U.S. Statistical System

04.5.2022

Ben Reist

FAIR Data Principles

Findable

- For data to be used it needs to be easily discoverable

Accessible

- Once found it needs to be clear how to obtain access

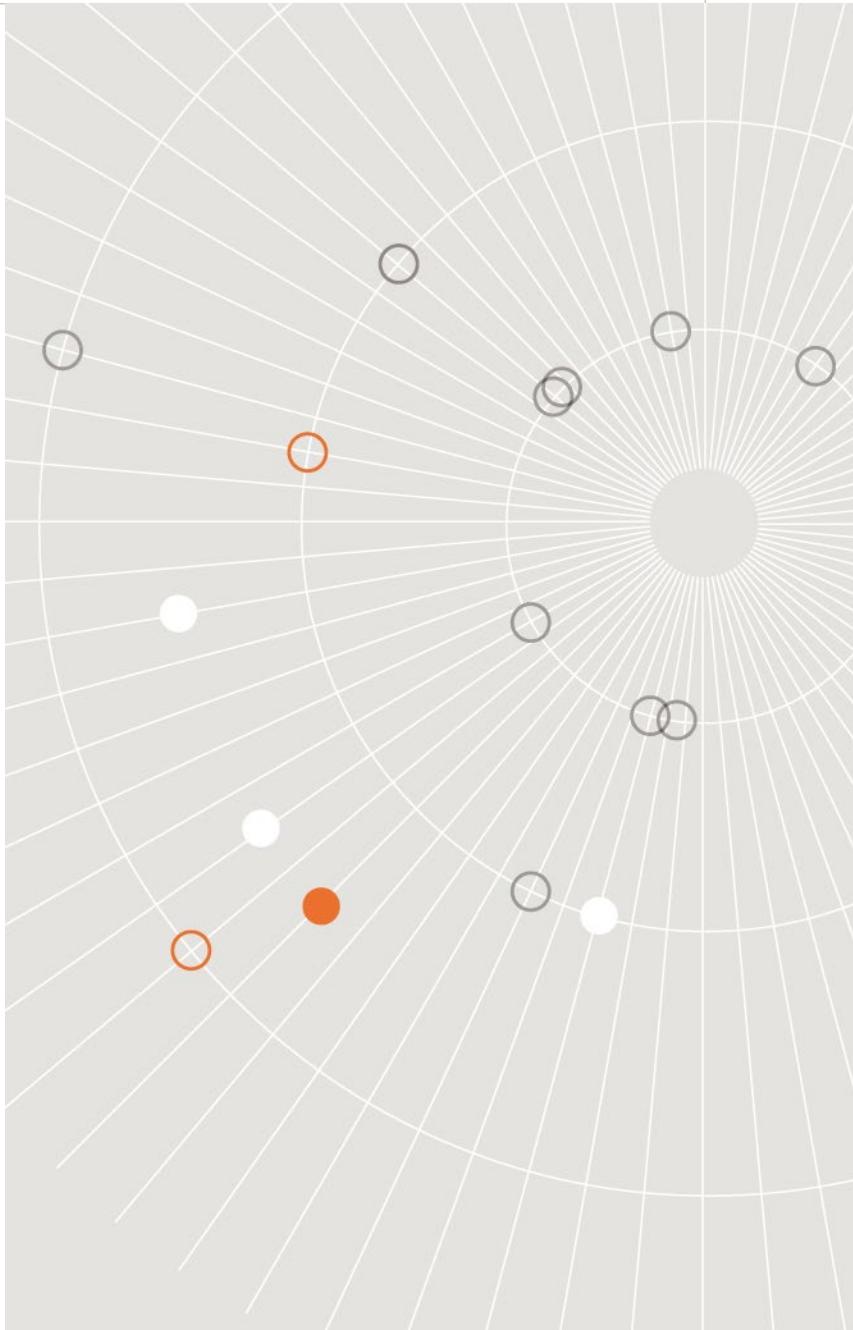
Interoperable

- Needs to be tool agnostic

Reusable

- Overall goal is to promote reuse of data

*Source: https://www.go-fair.org/fair-principles/



Findable

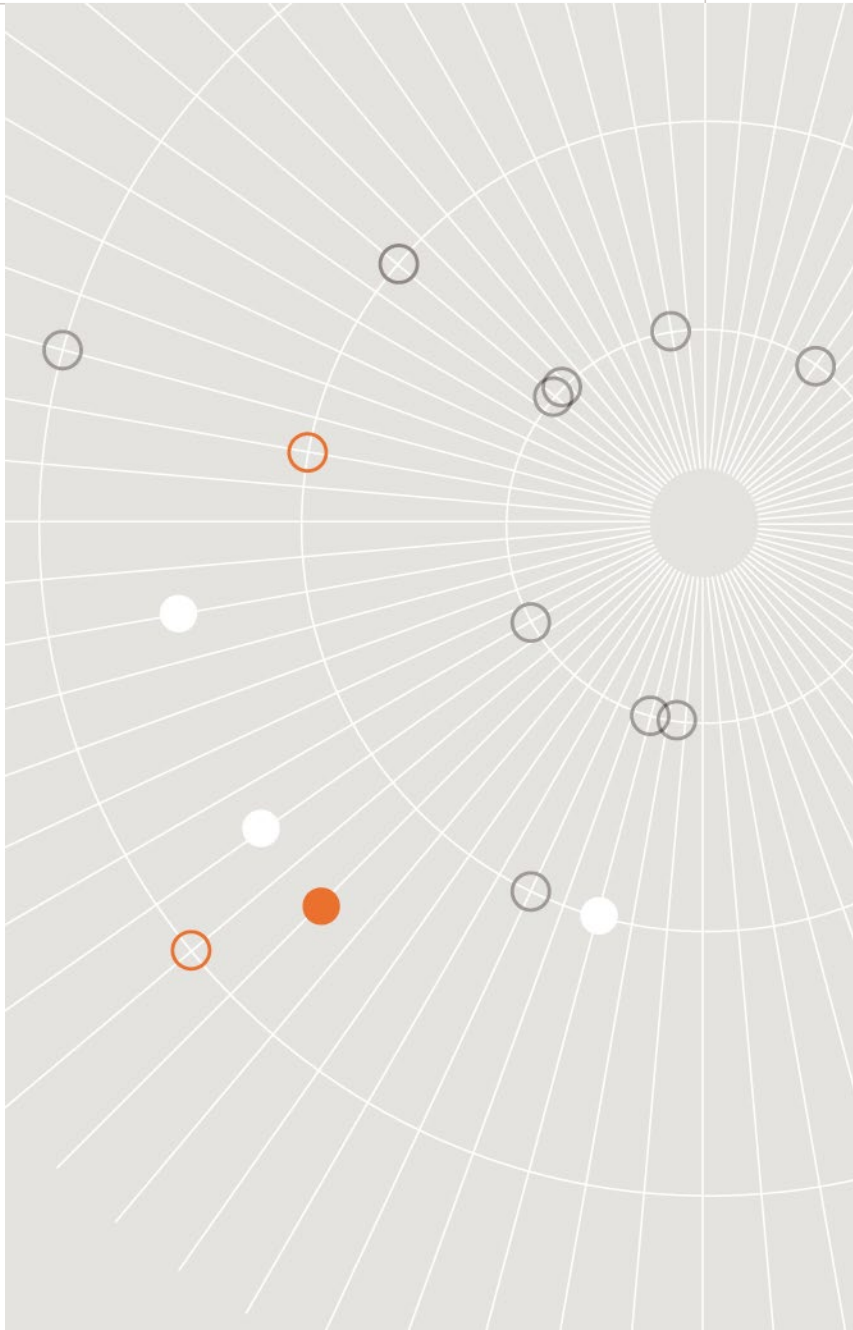
1. Data are assigned a globally unique and persistent identifier

2. Data are described with rich metadata

3. Metadata clearly and explicitly include the identifier of the data they describe

4. Data are registered or indexed in a searchable resource

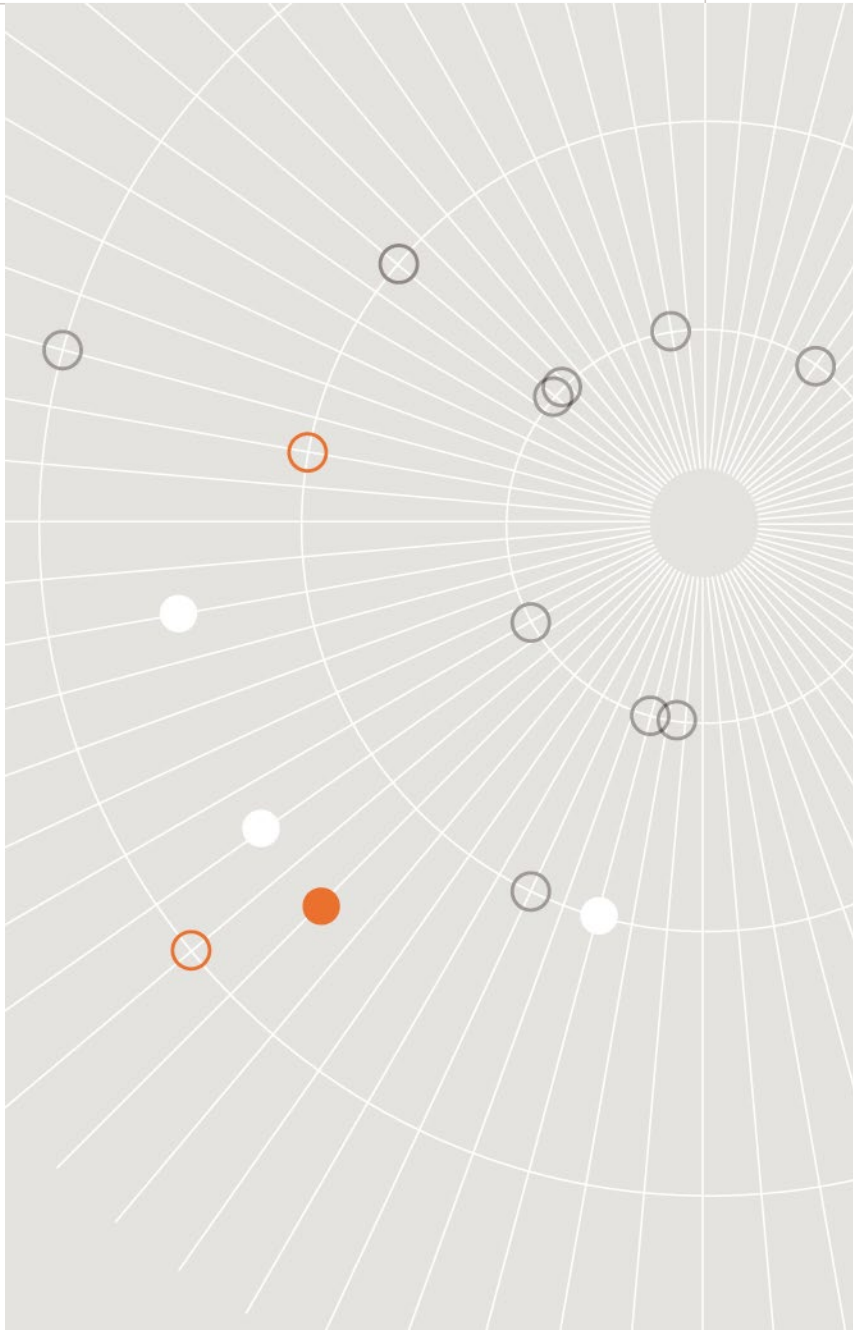
*Source: https://www.go-fair.org/fair-principles/



Accessible

1. Data are retrievable by their identifier using a standardized communications protocol
    1. **The protocol is open, free, and universally implementable**
    2. **The protocol allows for an authentication and authorization procedure, where necessary**

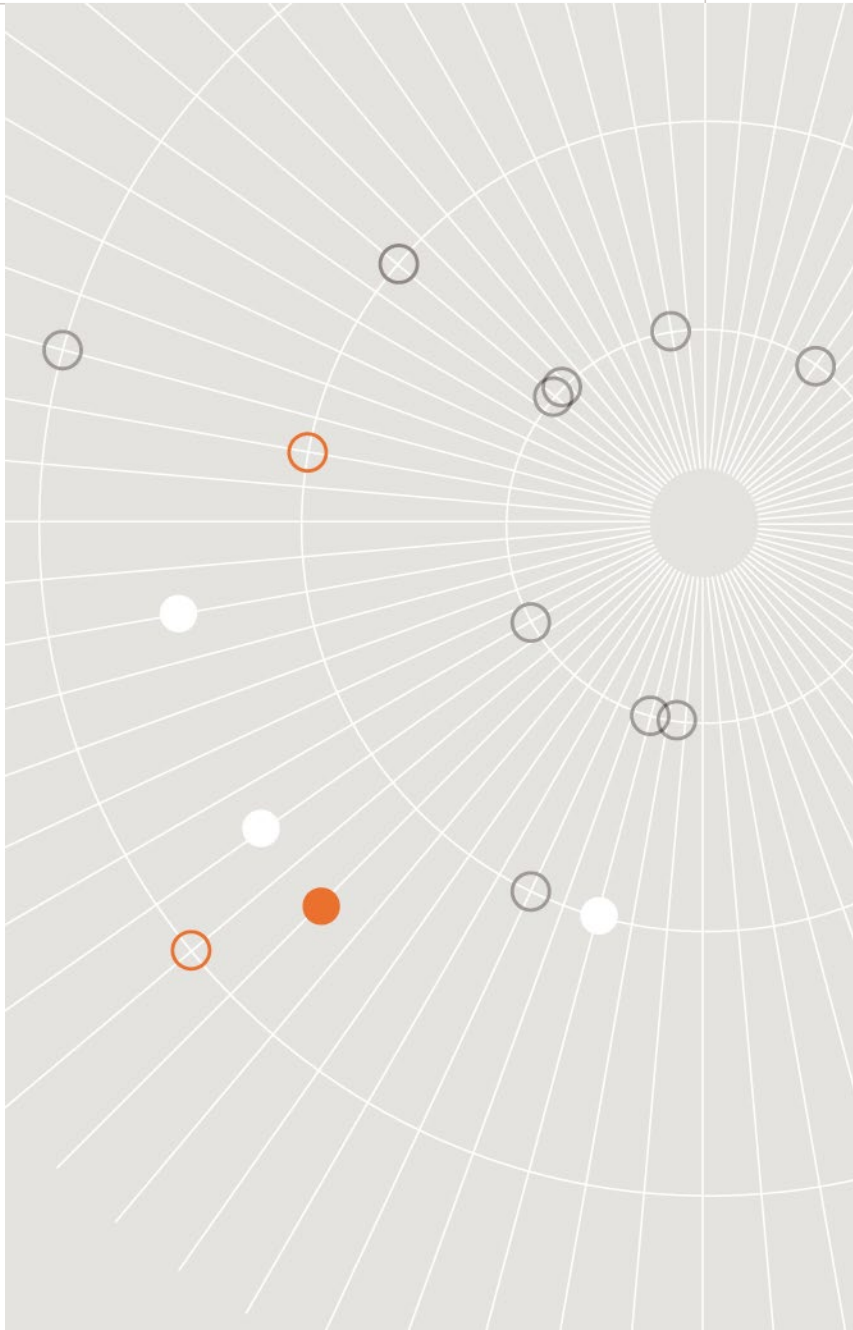
2. Metadata are accessible, even when the data are no longer available



*Source: https://www.go-fair.org/fair-principles/

Interoperable

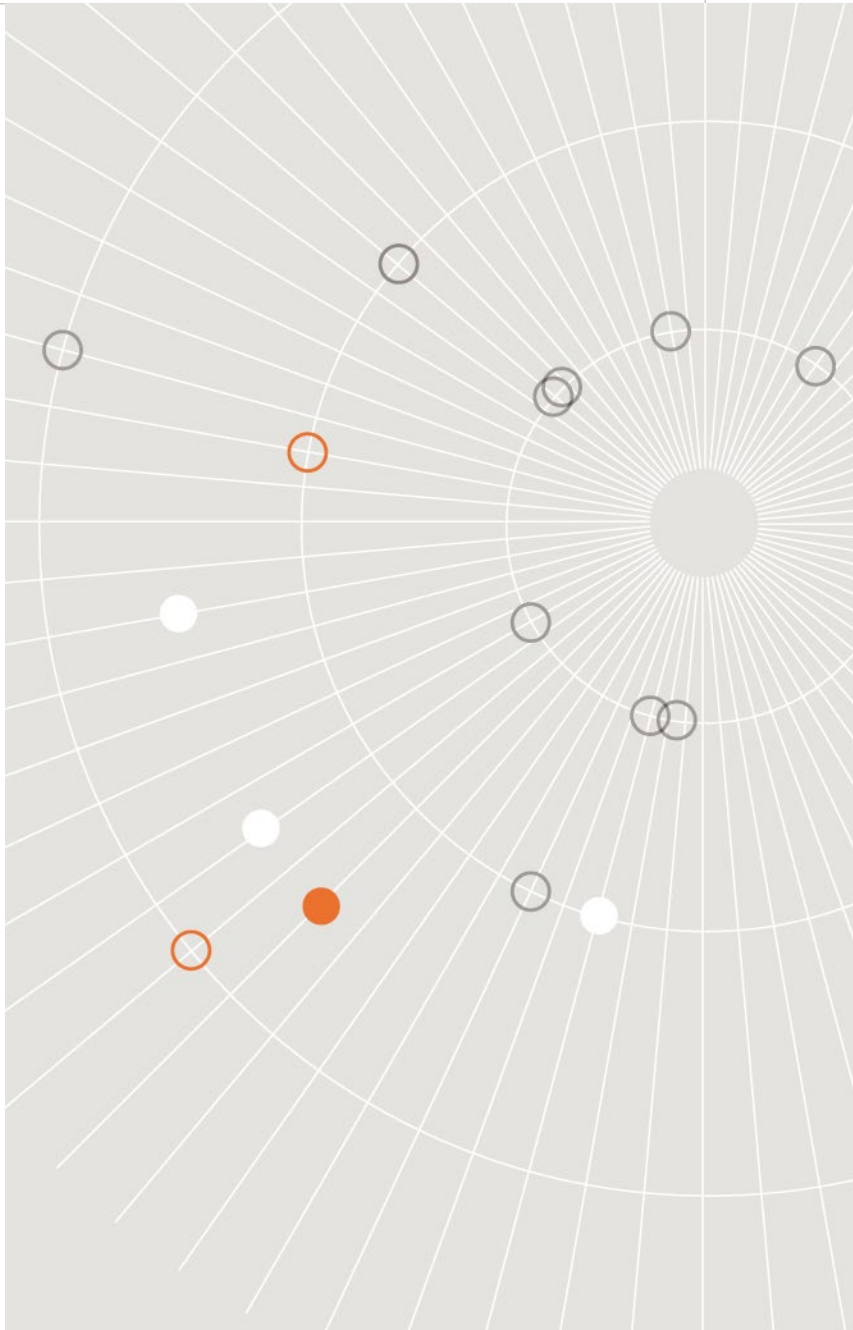
1. Data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
2. Data use vocabularies that follow FAIR principles
    1. Vocabulary used to describe datasets needs to be documented and resolvable using globally unique and persistent identifiers
    2. Documentation needs to be easily findable and accessible by anyone who uses the dataset
3. Data include qualified references to other data



*Source: https://www.go-fair.org/fair-principles/

Reusable

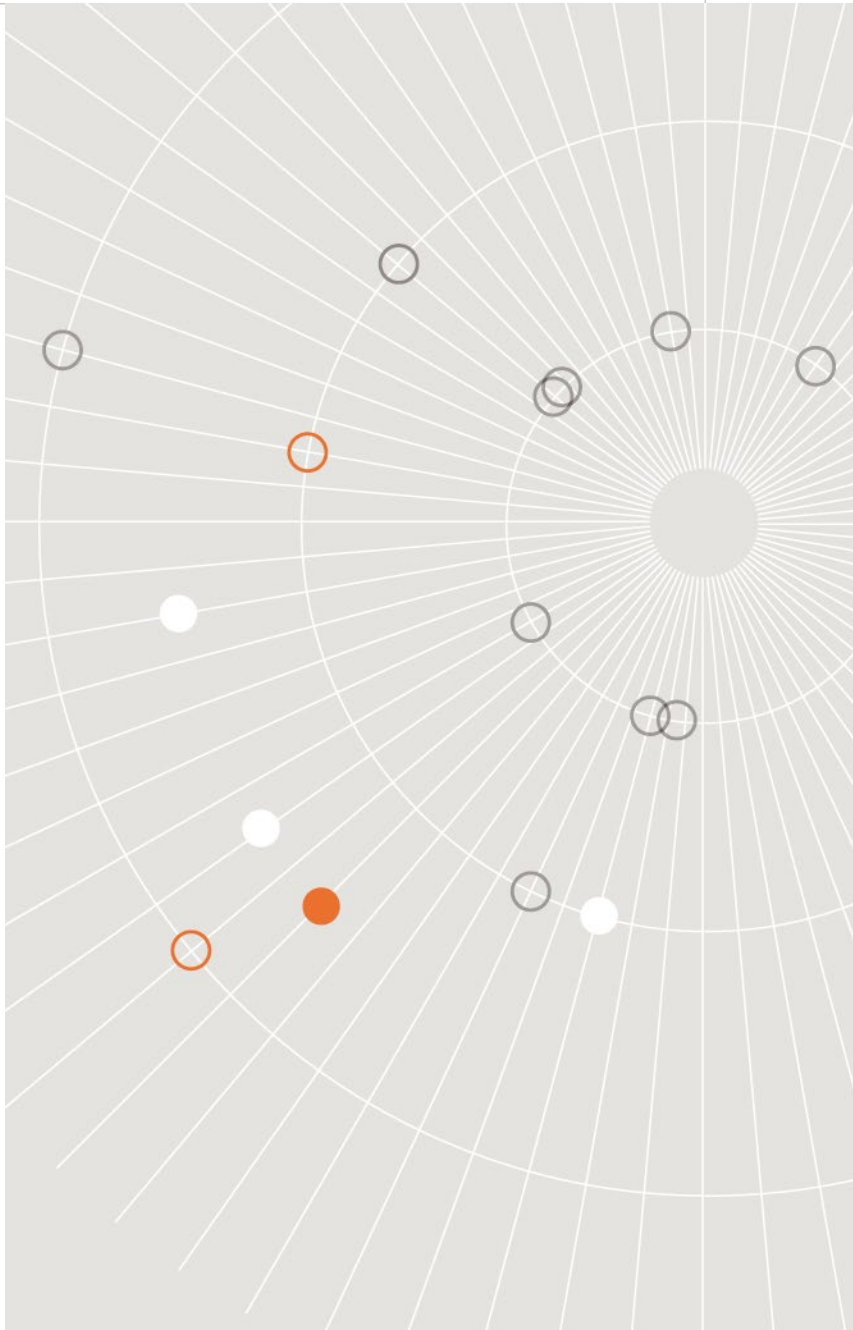
1. Data are richly described with a plurality of accurate and relevant attributes
    1. **Data are released with a clear and accessible data usage license**
    2. **Data are associated with detailed provenance**
    3. **Data meet domain-relevant community standards**



*Source: https://www.go-fair.org/fair-principles/

Challenges for Statistical System

- Decentralized Statistical System
  - **Lack of centralized search tool**
    - **Data.gov**
  - **Difficulty in sharing data across agencies (or within agencies for that matter)**
  - **Users need to know how statistical system is organized in order to know where to look**
- Confidentiality vs. Accessibility
- Useful data not being released despite Evidence Act
- Data collection/generation focus is primary use, not possible secondary use



Thank you.

Ben Reist
Principle Statistician
Reist-ben@norc.org

Research You Can Trust™

NORC at the University of Chicago

Data Governance and Security

Dennis Pickett
Vice President
Chief Information Security Officer (CISO)

Westat®

April 5, 2022

Security and Data Governance, a Perfect Match

- Security and data governance go hand in hand, although just looking at their definitions the relationship not immediately apparent.

- Data governance is a collection of processes, roles, policies, standards, and metrics that ensure the effective and efficient use of information in enabling an organization to achieve its goals.
- Cyber security refers to technologies, processes, and practices designed to protect networks, devices, programs, and data from attack, damage, or unauthorized access.
- They come together when we realize that a key component of both is to define what data assets an organization has, where the data lives, and who can take actions with that data, when they can perform those actions, and under what conditions those actions are allowed.
- This is something Data Governance needs to know, and security must ensure.

Let's Take a Closer Look at the Symbiotic Relationship

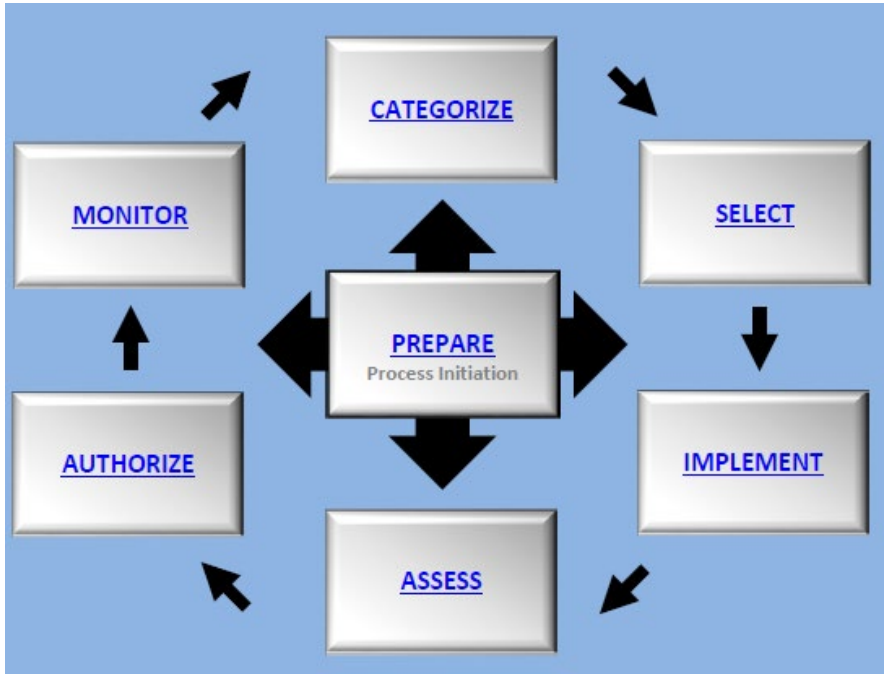
- In both cases the first step is understanding what you have. **Discovery** and **classification** are the processes are fundamental. These processes answer the question of, "What is it that you have, and what is its value?"
- When you know those two things you can take the next steps, and in terms of security those would be to determine a risk level and who should have access. Then security controls can be applied to enforce those rules, and track activity.
- Knowing what and where your highest risk, sometimes termed "sensitive", data is allows an organization to set priorities to remediate and mitigate the greatest risks. It lets you use your resources in the most effective manner.
- Without this initial understanding, trying to implement governance would just be taking guesses at what is and isn't effective. And trying to implement security would lead to one of two probable outcomes; overly restrictive security that locks everything down and impedes business, or too little security where everyone has access to everything.

How Do We Implement Security for Data Governance?

- Pick a security framework and follow the procedures.
  - A framework is a system of standards, guidelines, and best practices to manage risk.
  - You need a framework to implement a complete and mature security program. It ensures quality, consistency, and helps make sure nothing is missed.
  - A framework ultimately is a plan that you will use to identify what assets and data you have, what the risk is for each item, and what security controls are appropriate.
- Frameworks:
  - NIST Risk Management Framework (RMF) - Special Publication 800-37 Rev 2, "Guide for Applying the Risk Management Framework to Federal Information Systems" is the Federal Government security framework. There are "lite" versions of this.
  - ISO/IEC 27001 is an international standard on how to manage information security. Typically used by commercial organizations.
  - HITRUST is a massive undertaking for any organization due to the heavy weight given to documentation and processes.
  - GDPR is a framework of security requirements that global organizations must implement to protect the security and privacy of EU citizens' personal information.
  - Center for Internet Security (CIS) Critical Security Controls – (formerly the SANS Top 20) lists technical security and operational controls that can be applied to any environment. Not a complete framework, focuses on most critical items.

NIST Risk Management Framework (RMF) as an Example

- Prepare by establishing a context and priorities for managing security and privacy risk.
- Categorize the system and the information processed, stored, and transmitted by the system based on an analysis of the impact of loss.
- Select an initial set of controls for the system.
- Implement the controls and describe how the controls are employed.
- Assess the controls to determine if the controls are implemented correctly, operating as intended.
- Authorize the system or common controls based on acceptable risk.
- Monitor the system and the associated controls on an ongoing basis.



How the RMF Categorizes Risk

- FIPS 199 establishes three potential impact levels (Low, Moderate, High) for each of the security objectives (confidentiality, integrity, and availability).
- The impact levels focus on the potential impact and magnitude of harm that the loss of confidentiality, integrity, or availability (CIA).

800-60 Information Type	800-60 Impact Levels			Adjusted Impact Levels		
Debt Collection	*Confidentiality*	*Integrity*	*Availability*	*Confidentiality*	*Integrity*	*Availability*
	Moderate	Low	Low	Moderate	Low	Low
Rationale	Information meeting the definition of Revenue Collection type C.2.5.1, Debt Collection, is utilized by the [REDACTED] project and is stored, processed, and/or transmitted by the [REDACTED] information system.					

800-60 Information Type	800-60 Impact Levels			Adjusted Impact Levels		
Federal Grants (Non-State)	*Confidentiality*	*Integrity*	*Availability*	*Confidentiality*	*Integrity*	*Availability*
	Low	Low	Low	Moderate	Low	Low
Rationale	Information meeting the definition of Federal Financial Assistance type D.23.1, Federal Grants (Non-State), is utilized by the [REDACTED] project and is stored, processed, and/or transmitted by the [REDACTED] information system. Confidentiality objective increased to Moderate from Low because of potential inclusion of PII/IIF and loss of confidentiality because loss of confidentiality could be expected					

How the RMF Selects Security Controls

- The controls are selected based on the risk level.
- How the controls are implemented is documented in the System Security Plan (SSP)

TABLE 3-1: ACCESS CONTROL FAMILY

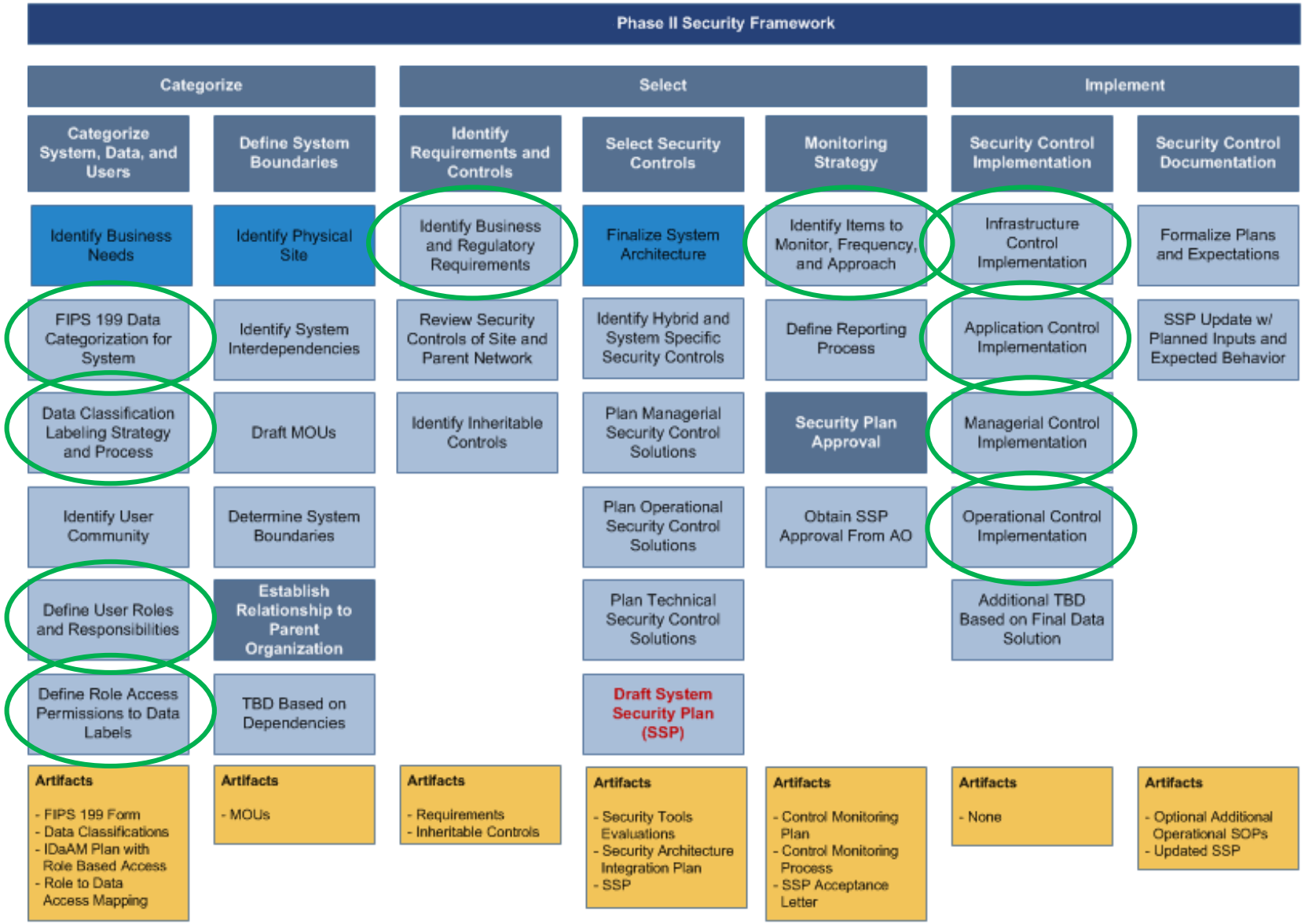
CONTROL NUMBER	CONTROL NAME CONTROL ENHANCEMENT NAME	PRIVACY CONTROL BASELINE	SECURITY CONTROL BASELINES		
			LOW	MOD	HIGH
AC-1	Policy and Procedures	x	x	x	x
AC-2	Account Management		x	x	x
AC-2(1)	AUTOMATED SYSTEM ACCOUNT MANAGEMENT			x	x
AC-2(2)	AUTOMATED TEMPORARY AND EMERGENCY ACCOUNT MANAGEMENT			x	x
AC-2(3)	DISABLE ACCOUNTS			x	x
AC-2(4)	AUTOMATED AUDIT ACTIONS			x	x
AC-2(5)	INACTIVITY LOGOUT			x	x
AC-2(6)	DYNAMIC PRIVILEGE MANAGEMENT				
AC-2(7)	PRIVILEGED USER ACCOUNTS				
AC-2(8)	DYNAMIC ACCOUNT MANAGEMENT				
AC-2(9)	RESTRICTIONS ON USE OF SHARED AND GROUP ACCOUNTS				
AC-2(10)	SHARED AND GROUP ACCOUNT CREDENTIAL CHANGE	W: Incorporated into AC-2k.			
AC-2(11)	USAGE CONDITIONS				x
AC-2(12)	ACCOUNT MONITORING FOR ATYPICAL USAGE				x
AC-2(13)	DISABLE ACCOUNTS FOR HIGH-RISK INDIVIDUALS			x	x
AC-3	Access Enforcement		x	x	x
AC-3(1)	RESTRICTED ACCESS TO PRIVILEGED FUNCTIONS	W: Incorporated into AC-6.			
AC-3(2)	DUAL AUTHORIZATION				
AC-3(3)	MANDATORY ACCESS CONTROL				
AC-3(4)	DISCRETIONARY ACCESS CONTROL				
AC-3(5)	SECURITY-RELEVANT INFORMATION				
AC-3(6)	PROTECTION OF USER AND SYSTEM INFORMATION	W: Incorporated into MP-4 and SC-28			

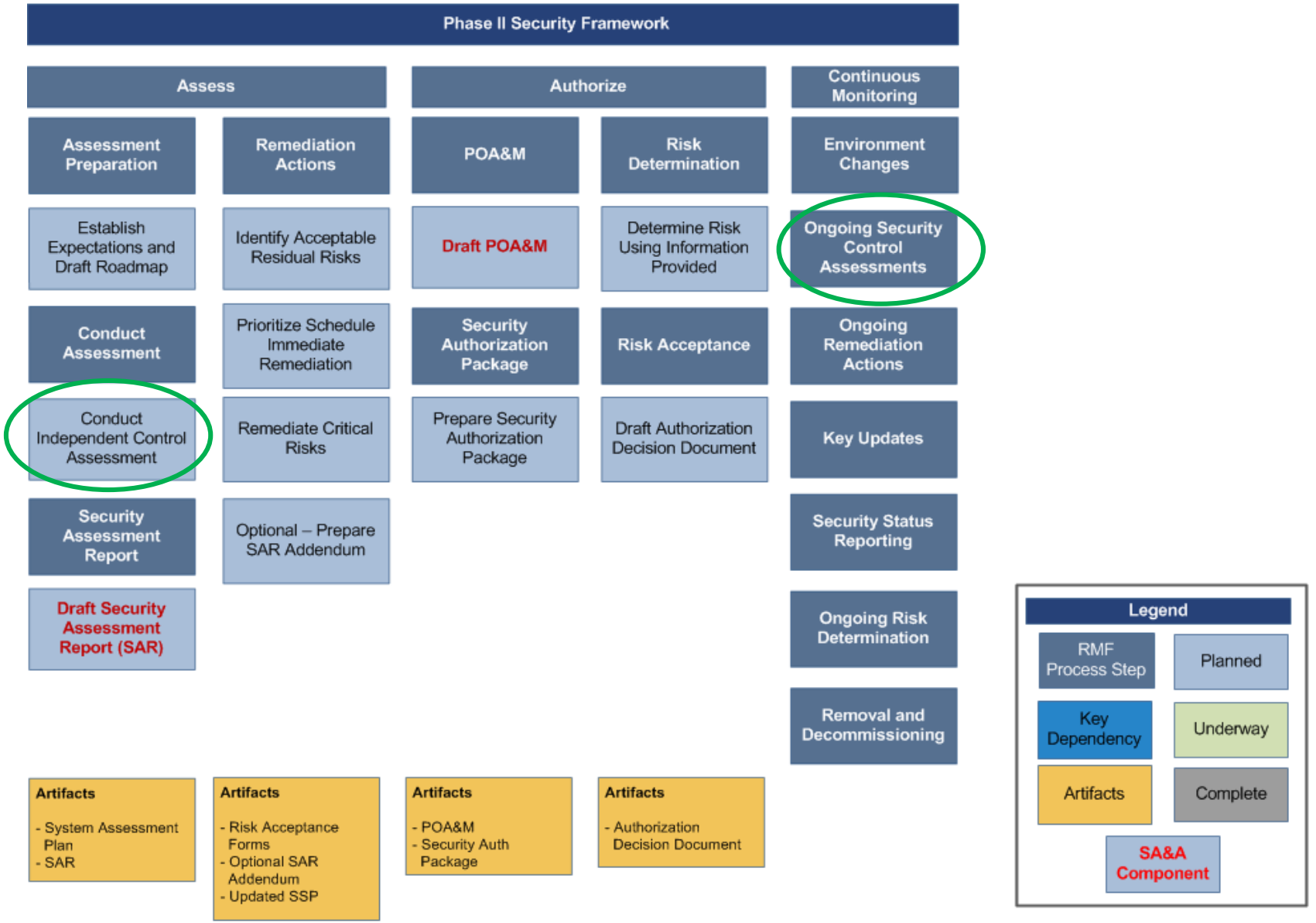
(AC-3) Access Enforcement

Requirements and Controls

Req.ID	800-53 Requirement and Westat/OSEP Overlay(s)	Control
AC-3	The information system enforces approved authorizations for logical access to information and system resources in accordance with applicable access control policies.	POC: [REDACTED] ◊ Designation: System-Specific PDPDCS staff members may access [REDACTED]-IS components according to rights determined by the system owner based on position or user-specific access authorizations. Access to information system functions and information, including privileged functions and security-relevant information, is restricted and limited to the level required for that individual to perform assigned duties. Individual authorizations are tracked in the roster. Enforcement of access authorizations in project-developed/maintained [REDACTED]-IS components is implemented through mechanisms within the component (e.g., application, operating system). Authorization enforcement mechanisms for project-controlled application accounts may include but are not limited to access rights assignments, username and password, other authentication factors, TLS (secure connection), etc. √ Fully Implemented

- Overview of complete program.
- Key intersections with data governance circled.




Phase II Security Framework
Assess
Authorize
Continuous Monitoring
Assessment Preparation
Remediation Actions
POA&M
Risk Determination
Environment Changes
Establish Expectations and Draft Roadmap
Identify Acceptable Residual Risks
Draft POA&M
Determine Risk Using Information Provided
Ongoing Security Control Assessments
Conduct Assessment
Prioritize Schedule Immediate Remediation
Security Authorization Package
Risk Acceptance
Ongoing Remediation Actions
Conduct Independent Control Assessment
Remediate Critical Risks
Prepare Security Authorization Package
Draft Authorization Decision Document
Key Updates
Security Assessment Report
Optional – Prepare SAR Addendum
Security Status Reporting
Draft Security Assessment Report (SAR)
Ongoing Risk Determination
Removal and Decommissioning
Artifacts
- System Assessment Plan
- SAR
Artifacts
- Risk Acceptance Forms
- Optional SAR Addendum
- Updated SSP
Artifacts
- POA&M
- Security Auth Package
Artifacts
- Authorization Decision Document
Legend
RMF Process Step
Planned
Key Dependency
Underway
Artifacts
Complete
SA&A Component

Thank You

Discussion