Issue #28: 💾 Data Management Plan Template

Repository: `CherrelleTucker/codesign-toolkit` **URL:**
https://github.com/CherrelleTucker/codesign-toolkit/issues/28 **Author:** @CherrelleTucker

State: `open` **Labels:** ⚙️ technical-codev, ⚙️ phase-development, 📈 difficulty-intermediate
Assignees: None

Created: 2025-11-14T08:52:32Z **Last Updated in GitHub:** 2025-11-17T05:51:12Z **Worksheet
Version:** 2025-11-17T05:51:22.169Z

💾 Data Management Plan Template

Structured Approach to Managing Earth Observation Data Throughout Solution Lifecycle

Tool Category: Technical Co-Development | **Phase**: Planning & Implementation | **Difficulty**: 📈
Intermediate

> *Establish comprehensive data governance, storage, processing, and sharing protocols to ensure data integrity, accessibility, and compliance throughout your solution's lifecycle.*

📋 Tool Summary Card

Attribute	Value
🎯 **Purpose**	Define data handling procedures, storage requirements, and governance protocols for solution data
⏱️ **Time Required**	3-4 hours initial planning + 1-2 hours stakeholder review + periodic updates
👥 **Participants**	4-6 people: data steward + technical lead + compliance officer + user representative + security lead
📊 **Outputs**	Data management procedures, storage architecture, compliance checklist, data sharing agreements
🔄 **Frequency**	Once per solution, reviewed quarterly, updated when data requirements change
💼 **Materials**	Data inventory, compliance requirements, storage capacity planning, backup procedures

---

🎯 When to Use This Tool

✅ **Essential For:**

- Solutions handling sensitive or regulated Earth observation data
- Multi-organization projects requiring data sharing agreements
- Long-term solutions needing sustainable data storage and access
- Projects with specific compliance or security requirements

⚠️ **Consider Simpler Approaches When:**

- Prototype or proof-of-concept phases using sample data only
- Solutions using exclusively public, open-access datasets
- Single-user applications with minimal data storage needs

⚠️ **Use After:**

- Data sources and types have been identified through user requirements
- Technical architecture decisions are stable
- Compliance and security requirements are understood

📊 Data Management Plan Template Structure

Section 1: Data Overview

Project Information:

```
Solution Name: [Official solution name]
Data Steward: [Name and contact - primary data responsibility]
Technical Lead: [Name and contact - implementation responsibility]
Compliance Officer: [Name and contact - regulatory oversight]
Plan Version: [Version number and date]
Review Schedule: [Quarterly/Semi-annual review dates]
```

Data Summary:

- **Primary Data Types:** [Satellite imagery, sensor data, model outputs, etc.]
- **Data Volume Estimates:** [Current and projected data storage needs]
- **Data Sources:** [External APIs, partner organizations, user uploads, etc.]
- **Data Sensitivity Level:** [Public, internal use, restricted, classified]
- **Retention Requirements:** [How long data must be kept]

Section 2: Data Sources & Collection

Input Data Sources:

Data Type	Source	Format	Volume	Update Frequency
[Satellite imagery]	[NASA/ESA/Commercial]	[GeoTIFF/NetCDF]	[TB/month]	[Daily/Weekly]
[Ground measurements]	[Partner agencies]	[CSV/JSON]	[GB/month]	[Real-time/Daily]
[Model outputs]	[Internal processing]	[HDF5/NetCDF]	[GB/day]	[On-demand]

[User annotations]	[Solution users]	[JSON/Database]	[MB/month]	[Continuous]

Data Collection Procedures:

- **Automated Collection:** [Scheduled processes, API calls, monitoring]
- **Manual Collection:** [User upload procedures, validation steps]
- **Quality Control:** [Data validation rules, error handling]
- **Backup Collection:** [Alternative sources, redundancy plans]

Section 3: Data Storage Architecture

Storage Infrastructure:

Primary Storage:

- **Location:** [Cloud provider, on-premises, hybrid]
- **Capacity:** [Current and projected storage needs]
- **Performance Requirements:** [Access speed, throughput needs]
- **Redundancy:** [Backup copies, geographic distribution]

Storage Tiers:

- **Hot Storage:** [Frequently accessed data - location and criteria]
- **Warm Storage:** [Occasionally accessed data - location and criteria]
- **Cold Storage:** [Archive data - location and criteria]
- **Backup Storage:** [Disaster recovery - location and procedures]

Data Organization:

```
/data
├── /raw                  # Unprocessed input data
│   ├── /satellite        # Raw satellite imagery
│   ├── /ground_truth     # Field measurements
│   └── /external_feeds   # Third-party data sources
├── /processed            # Cleaned and transformed data
│   ├── /level1           # Basic processing
│   ├── /level2           # Analysis-ready data
│   └── /derived          # Computed products
├── /outputs              # Solution-generated products
│   ├── /reports          # Analysis reports
│   ├── /visualizations   # Charts, maps, dashboards
│   └── /exports          # User downloads
└── /metadata             # Data documentation and catalogs
```

🔐 Data Security & Compliance

Security Requirements

Access Control:

- **Authentication:** [Multi-factor, SSO, certificate-based]
- **Authorization:** [Role-based permissions, data classification levels]
- **User Management:** [Account provisioning/deprovisioning procedures]

- **Audit Logging:** [Access tracking, change logs, compliance reporting]

Data Protection:

- **Encryption at Rest:** [Algorithm, key management]
- **Encryption in Transit:** [TLS/SSL protocols, API security]
- **Data Masking:** [PII protection, sensitive data handling]
- **Backup Encryption:** [Backup security measures]

Compliance Requirements

Regulatory Compliance:

- **FISMA:** [Federal security requirements if applicable]
- **NIST Guidelines:** [Cybersecurity framework adherence]
- **Export Control:** [ITAR/EAR restrictions on data sharing]
- **Privacy Regulations:** [PII handling, user consent]

Organizational Policies:

- **Data Classification:** [Internal security levels and handling]
- **Retention Policies:** [How long to keep different data types]
- **Disposal Procedures:** [Secure data deletion methods]
- **Incident Response:** [Data breach notification procedures]

Data Sharing Agreements

Internal Sharing:

- **Team Access:** [Who can access what data and when]
- **Cross-Project Use:** [Reuse policies for other initiatives]
- **Archive Access:** [Long-term data availability procedures]

External Sharing:

- **Partner Organizations:** [Data sharing MOUs and agreements]
- **Public Release:** [Open data publication procedures]
- **User Downloads:** [What users can export and share]
- **Research Collaboration:** [Academic and research partnerships]

⚙️ Data Processing & Quality

Data Processing Pipeline

Processing Stages:

1. **Ingestion Stage:**

   - Data validation and format verification
   - Metadata extraction and catalog updates
   - Initial quality checks and error flagging
   - Staging for further processing

2. **Cleaning Stage:**

   - Missing data handling procedures
   - Outlier detection and treatment

- Format standardization
- Quality flag assignment

3. **Transformation Stage:**

   - Coordinate system conversions
   - Temporal alignment and resampling
   - Data fusion and integration
   - Derived product generation

4. **Validation Stage:**

   - Quality assurance testing
   - User acceptance validation
   - Performance benchmarking
   - Output verification

Processing Documentation:

```
Processing Procedure: [Procedure Name]

Input Requirements:
- Data format: [Required input format]
- Quality criteria: [Minimum quality standards]
- Completeness: [Required data coverage]

Processing Steps:
1. [Step description with parameters]
2. [Step description with parameters]
3. [Step description with parameters]

Output Specifications:
- Format: [Output data format]
- Quality flags: [Quality indicators added]
- Metadata: [Required metadata fields]

Quality Control:
- Validation tests: [Automated quality checks]
- Manual review: [Human verification steps]
- Error handling: [How errors are managed]
```

Data Quality Framework

Quality Dimensions:

- **Accuracy:** [How correct is the data?]
- **Completeness:** [Are there gaps in coverage?]
- **Consistency:** [Are values uniform across datasets?]
- **Timeliness:** [Is data available when needed?]
- **Validity:** [Does data conform to expected formats?]

Quality Metrics:

Quality Aspect	Metric	Target	Measurement Method

Data Completeness	% coverage of area of interest	>95%	Automated spatial analysis
Temporal Coverage	% of expected time series	>90%	Gap analysis scripts
Processing Accuracy	% of successful processing runs	>99%	Pipeline monitoring
User Data Quality	User-reported quality rating	>4.0/5.0	Feedback surveys

🔄 Data Lifecycle Management

Data Retention & Archival

Retention Schedule:

Data Type	Active Period	Archive Period	Disposal Criteria
Raw input data	2 years	10 years	Legal/policy requirements met
Processed data	5 years	Permanent	Historical research value
User outputs	1 year	3 years	User export capability
System logs	1 year	5 years	Audit requirements
Temporary files	30 days	N/A	Processing complete

Archival Procedures:

1. **Data Review:** [Quarterly assessment of data for archival]
2. **Migration Process:** [Moving data to long-term storage]
3. **Access Procedures:** [How to retrieve archived data]
4. **Disposal Process:** [Secure deletion of expired data]

Backup & Recovery

Backup Strategy:

- **Frequency:** [Daily/Weekly/Monthly backup schedules]
- **Scope:** [What data is included in each backup type]
- **Location:** [Geographic distribution of backups]
- **Testing:** [Regular restore testing procedures]

Recovery Procedures:

```
Data Recovery Plan

Recovery Time Objectives (RTO):
- Critical data: [X hours]
- Important data: [Y hours]
- Archive data: [Z days]

Recovery Point Objectives (RPO):
- Critical data: [Maximum X hours of data loss]
```

```
- Important data: [Maximum Y hours of data loss]

Recovery Steps:
1. [Incident assessment and classification]
2. [Backup identification and retrieval]
3. [Data restoration procedures]
4. [Validation and user notification]

Contact Information:
- Primary: [24/7 contact for critical issues]
- Secondary: [Backup contact information]
- Escalation: [Management notification procedures]
```

Performance Monitoring

Storage Performance:

- **Capacity Utilization:** [Current usage vs. available capacity]
- **Growth Trends:** [Data volume increase patterns]
- **Access Patterns:** [Most/least frequently accessed data]
- **Performance Metrics:** [Read/write speeds, response times]

Data Pipeline Performance:

- **Processing Times:** [Duration for different processing stages]
- **Error Rates:** [Frequency and types of processing failures]
- **Throughput:** [Volume of data processed per time period]
- **Resource Utilization:** [CPU, memory, storage usage during processing]

📋 Implementation Checklist

Setup Phase

- ☐ Data storage infrastructure provisioned and configured
- ☐ Security controls implemented and tested
- ☐ Access management system configured
- ☐ Backup and recovery procedures tested
- ☐ Data processing pipeline developed and validated
- ☐ Quality control procedures implemented
- ☐ Compliance requirements verified and documented

Deployment Phase

- ☐ Production data migration completed successfully
- ☐ User access permissions configured and tested
- ☐ Monitoring and alerting systems activated
- ☐ Data sharing agreements executed where needed
- ☐ Staff training on data procedures completed
- ☐ Documentation finalized and distributed

Operations Phase

- ☐ Regular backup verification schedule established
- ☐ Performance monitoring dashboards active
- ☐ Quality control reports generated and reviewed
- ☐ Compliance audit procedures scheduled
- ☐ Data retention policies enforced
- ☐ User feedback collection system operational

Maintenance Phase

- ☐ Quarterly data management plan reviews scheduled
- ☐ Annual compliance assessments planned
- ☐ Storage capacity monitoring and planning active
- ☐ Data archive and disposal procedures operational
- ☐ Incident response procedures tested and documented

🔄 Integration with Other Tools

This Data Management Plan Builds On:

- 📋 **Requirements Definition Canvas** - Data requirements and user needs inform storage and processing decisions
- 🏗️ **System Integration Planning** - Technical architecture determines data infrastructure needs
- 📋 **Solution Implementation Plan** - Implementation timeline includes data migration and setup phases

Information Sources:

- 🗺️ **Stakeholder Mapping Workshop** - Data sharing requirements with different organizations
- 🔍 **Discovery Interview Blueprint** - User data workflows and quality expectations
- ⚡ **Technical Validation Checklist** - Data processing and storage validation requirements

This Data Management Plan Enables:

- 🧪 **User Testing Protocol** - Test data preparation and management during validation
- 📊 **Adoption Monitoring Dashboard** - Data collection for usage analytics and performance metrics
- 🛠️ **Support System Setup** - Data access and troubleshooting procedures for user support

External Tool Integration:

- **Data Catalog Systems:** Metadata management and data discovery
- **Monitoring Platforms:** Storage performance and data pipeline health tracking
- **Backup Solutions:** Automated backup scheduling and verification
- **Compliance Tools:** Audit logging and regulatory reporting systems

---

📚 Source Attribution

Primary Sources:

- **Solution Co-Development Toolkit Narrative** - Data management requirements for user-centered solutions
- **NSITE Solution Project Requirements and Expectations** - Data handling and security standards for Earth observation solutions

Supporting Sources:

- **SERVIR Service Design Tool 2021** - Data sharing and collaboration requirements
- **MSFC Coordination on Solutions Co-Development Toolkit** - Multi-organizational data governance

Methodology Foundation:

- NIST Cybersecurity Framework for data protection standards
- Research Data Management best practices from academic and federal institutions
- Cloud security and data governance frameworks

💬 Community Discussion

Share your data management experience:

- What data quality challenges are most common in Earth observation solutions?
- How do you balance data accessibility with security and compliance requirements?
- What automated tools have been most effective for data pipeline monitoring?
- How do you handle data sharing across organizational boundaries?

Tool improvements:

- What additional data types or sources should be covered in Earth observation data management?
- How do you adapt data management for real-time vs. batch processing requirements?
- What metrics have been most useful for tracking data management effectiveness?

🔧 **Tool Maintainer:** @your-username | 📅 **Last Updated:** [Today's Date] | 📌 **Version:** 1.0