

The Data Steward's Field Guide

A Quick Reference Tool



The Data Steward's Field Guide

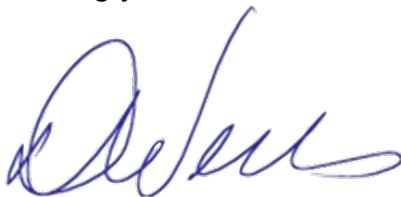
A Quick Reference Tool

The Data Steward's Field Guide is built as a quick, practical reference for anyone with the responsibilities of data stewardship. It brings together essential tools, techniques, and diagnostic aids that stewards can rely on to recognize problems, understand causes, and move toward solutions that strengthen data management and governance.

Use it as a working companion. The diagnostic section helps you start with the symptoms you observe in your data and trace them to their root causes, providing clear options for resolution. The facilitation section offers methods to guide conversations, encourage participation, and foster a shared understanding. Whether you are addressing data quality issues, integration challenges, or governance discussions, the guide is designed to keep stewards effective, focused, and ready to act,

This guide is protected by copyright (© 2026 by Dave Wells). But the purpose is to advance the practice of data stewardship as a profession. The guide is useful only if it is distributed and applied. With that goal in mind, I grant permission to copy, reproduce, and distribute the document as widely as is needed for non-commercial uses to further the practice and disciplines of data stewardship. You may not reproduce for resale or for-profit endeavors—only for internal use in your organization's data stewardship efforts.

Wishing you data stewardship success.



Dave Wells
dwells@infocentric.org

What Does a Data Steward Do?

As a data steward, you’ll find that your responsibilities span across time—looking to the **past**, engaging with the present, and anticipating the future. Each timeframe offers a unique lens through which your work adds value, and each demands a different emphasis on the core activities you perform.

The matrix below outlines five key activities: identifying issues, alleviating symptoms, remediating problems, mitigating risks, and facilitating solutions. While you’ll engage in all of these throughout your role, their relevance and intensity vary depending on whether you’re reflecting on what has already happened, managing current realities, or planning for what lies ahead. For example, identifying issues is essential across all timeframes, but remediation is especially critical in the present. Looking to the future, your focus will shift more toward mitigating risks and facilitating long-term solutions that promote growth and maturity for data governance and data management.

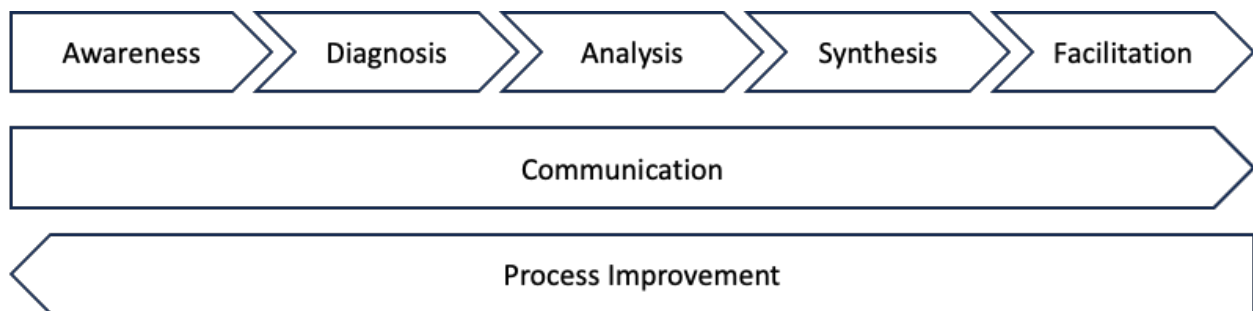
Understanding how your work flexes across time helps you stay proactive, prioritize effectively, and ensures that your stewardship drives both immediate impact and sustained improvement.

	Identify Issues	Alleviate Symptoms	Remediate Problems	Mitigate Risks	Facilitate Solutions
Past	X	X			
Present	X	X	X	X	X
Future	X	X	X	X	X

Key Data Steward Skills

Developing and using right skill set is essential to navigating the complexities of data stewardship. The diagram below outlines the key skills you'll draw on throughout your work, moving from awareness to facilitation in a progression that reflects how problems are surfaced, understood, and ultimately addressed. This isn't a one-time path—it's a cycle you'll revisit often as data challenges evolve.

You begin by building awareness—recognizing when something isn't quite right. From there, you move into diagnosis, where you pinpoint the nature and scope of the issue. Analysis helps to dig deeper into root causes, while synthesis focuses on defining viable solutions. Ultimately, facilitation empowers you to advocate for change and support its effective implementation. Two complementary capabilities strengthen problem solving. Communication spans every stage, making data concepts understandable, managing expectations, and driving cultural and behavioral change. Process improvement sits at the core, using insights from diagnosis, analysis, and synthesis to address root causes and raise the maturity level of data management practices.



- Awareness (recognition of problems and risks)
- Diagnosis (identification of problems and risks)
- Analysis (understanding of root causes)
- Synthesis (definition of solutions)
- Facilitation (project advocacy and process change)
- Communication (data literacy and competencies, behavioral and cultural change)
- Process Improvement (data management maturity)

Facilitation Guide

Data stewards are often called upon to lead or support conversations that influence how data is defined, managed, and used. Whether you're guiding a working session, moderating a discussion, or helping a group reach a decision, effective facilitation is key to creating a productive and inclusive environment.

This guide provides a quick reference to help you prepare for and navigate facilitation responsibilities. It outlines core facilitator goals, highlights the difference between typical and participatory group dynamics, and offers a variety of techniques you can use to encourage engagement, manage disagreement, and support thoughtful decision-making.

Use this section as a practical toolki, something you can return to whenever you need to foster clarity, collaboration, and shared ownership in data-related discussions.

Facilitator Goals	Full Participation	all key stakeholders participate all participants actively engaged
	Mutual Understanding	seeing through the eyes of others appreciation for all perspectives
	Inclusive Solutions	collaboratively developed consensus based
	Shared Responsibility	mutual trust mutual accountability
	Ethical Deliberation	responsible use explicitly considered values and consequences considered

Decision Process

Understanding the three stages of group decision-making is helpful when facilitating decision-making processes. Divergent thinking generates diverse viewpoints and ideas. Open communication with full participation builds mutual understanding among participants. Convergent thinking brings perspectives together to reach inclusive solutions with shared responsibility.



Group Dynamics

Typical Group	Participatory Group
Most vocal, articulate, and confident dominate conversations and are the most heard.	Everyone is heard, not just the loudest and most persistent. Quiet people are encouraged to speak up.
Different points of view are seen as “conflict” that must be resolved or eliminated.	Different points of view are the tools of lively discussion and full understanding.
Questioning is perceived as challenging, refuting, and creating conflict.	Questioning is encouraged as the means by which mutual understanding is achieved.
People stop listening, tune out, or focus on what they’ll say next	Listeners pay attention, confident that they’ll have the opportunity to be heard.
Some people close up and avoid speaking on controversial subjects.	Controversy without conflict is possible. All participants make their positions known.
Side conversations and outside-the-group maneuvers lead to divisiveness.	Out-of-group and behind-the-back discussion is discouraged as a harmful practice.
Minority viewpoints are often overrun by the voice of the majority.	Minority viewpoints are encouraged, heard, and recognized as valuable contributions.
Decisions are driven by the majority viewpoint, the most vocal, or the most powerful.	Decisions are reached when everyone who is affected understands the reasoning.

Facilitation Techniques

Effective facilitation requires more than simply keeping a meeting on track—it involves creating an environment where all voices are heard, ideas are explored, and consensus can emerge. The techniques in this quick reference provide practical tools to support those goals. Each method offers a way to encourage participation, manage group dynamics, or structure thinking so that discussions remain productive and inclusive.

Open Questions	Questions designed to spark discussion by encouraging elaboration, not “yes/no” answers.
Icebreakers	Quick, lighthearted prompts or activities that help participants relax, build rapport, and orient to the collaboration session
Reframing	Shifting the way a problem is presented or understood to open up new perspectives or solution paths.
Restatement	The facilitator paraphrases a participant’s comment to clarify meaning and ensure shared understanding.
Parking Lot	Capturing off-topic or future-relevant ideas in a visible space to keep the discussion focused while respecting those contributions.
Brainstorming	A creativity tool where participants freely generate many ideas without judgment.
Concept Fan / Mind Map	A visual branching diagram that starts with a central idea and radiates outward into related subtopics or solutions.

Facilitation Techniques (continued)

Blue Sky	Encouraging participants to think wildly or without constraints—“anything is possible”—to explore innovative ideas.
Anonymous Index Cards	Participants write ideas privately and submit them without attribution, which can surface candid or less common perspectives.
Red–Yellow–Green	A quick color-coded feedback method. Participants have colored cards that they can place in front of them to provide sentiment feedback without interrupting. Red = Disagree, Yellow = Uncertain/Confused, Green = Agree.
Deferred Judgment	Postponing critique or evaluation of ideas until after generation, to keep creativity flowing without premature filtering.
Role Reversal	Asking participants to argue or think from another stakeholder's perspective or role to broaden empathy and insight.
Simulation	Role-playing or scenario enactments that let participants test ideas or decisions in a dynamic, experiential context.
Facts versus Assumptions	Clarifying what is known versus what is taken for granted or guessed, helping groups ground discussions on evidence and reveal biases
Six Hats	<p>A structured "parallel thinking" model where participants take on six distinct thinking modes (e.g., facts, emotions, creativity, pessimism, optimism, and process control) to explore an issue collaboratively and systematically:</p> <ul style="list-style-type: none"> • White Hat – Facts and Information: Focus on data, evidence, and objective information. • Red Hat – Feelings and Intuition: Express emotions, gut reactions, and intuitive insights. • Black Hat – Caution and Risks: Identify potential problems, weaknesses, and dangers. • Yellow Hat – Benefits and Optimism: Highlight advantages, opportunities, and positive outcomes. • Green Hat – Creativity and Alternatives: Generate new ideas, possibilities, and innovative approaches. • Blue Hat – Process and Control: Manage the thinking process itself—set focus, summarize, and direct the flow.

Diagnostic Guide

This guide is designed to help data stewards diagnose and address common kinds of data dysfunction by organizing them around the core data management disciplines where root causes are most likely to be found. The fourteen core data management disciplines span from basic tasks like naming and defining data to more complex activities such as governing, analyzing, and automating with data.

Core Data Management Processes:

- Data Naming
- Data Definition
- Data Design
- Data Quality Management
- Data Integration & Interoperability
- Data Access
- Metadata Management
- Database & Storage Management
- Systems & Application Management
- Data Governance
- Data Protection
- Data Preparation & Provisioning
- Data Analysis
- Data-Driven Automation

What follows is a set of diagnostic tables—one for each process. Each table includes three columns: (1) Symptoms you might observe in the data, (2) Likely causes linked to that process, and (3) Recommended solutions that address those root issues.

The purpose is to trace visible problems (such as incorrect data names or high data disparity) back to the underlying process issues (such as lack of standards or informal practices), and then take targeted action. An accompanying symptom index ties specific problems back to the relevant processes, helping you navigate this guide based on what you’re seeing in your environment. Begin by browsing the index to find symptoms that you’re experiencing. Then use the numbers in the index to find guidance in the tables.

This is a tool to support your problem-solving, equipping you to not only spot issues, but to understand why they’re happening and how to fix them at the source.

Data Naming

	Symptoms	Causes	Solutions
1	meaningless data names	<input type="checkbox"/> informal naming practices <input type="checkbox"/> lack of naming standards <input type="checkbox"/> standards non-compliance <input type="checkbox"/> data naming in acquired systems	<input type="checkbox"/> data naming taxonomy <input type="checkbox"/> data naming vocabulary <input type="checkbox"/> standard naming structure <input type="checkbox"/> standard abbreviations list <input type="checkbox"/> compliance incentives <input type="checkbox"/> semantic data models
2	non-unique data names		
3	incorrect data names		
4	structureless data names		
5	confusing abbreviations		
6	multiple names and aliases		
7	unnamed data components		
8	hard-to-identify data		
9	high level of data disparity		

Data Definition

	Symptoms	Causes	Solutions
10	lack of data definitions	<input type="checkbox"/> lack of data definition standards <input type="checkbox"/> poor data definition practices <input type="checkbox"/> lack of business participation <input type="checkbox"/> legacy databases <input type="checkbox"/> disparate metadata <input type="checkbox"/> multiple & disparate data catalogs	<input type="checkbox"/> data definition standards <input type="checkbox"/> data definition templates <input type="checkbox"/> business/tech collaboration <input type="checkbox"/> data definition reviews <input type="checkbox"/> metadata repository <input type="checkbox"/> data catalog <input type="checkbox"/> domain-based stewardship <input type="checkbox"/> business glossary <input type="checkbox"/> semantic data models
11	incorrect data definitions		
12	meaningless data definitions		
13	obsolete data definitions		
14	hard to find data definitions		
15	misunderstood data		
16	inappropriate use of data		
17	high level of data disparity		
18	high level of data redundancy		

Data Design

	Symptoms	Causes	Solutions
19	poor structural integrity	<input type="checkbox"/> poor modeling techniques <input type="checkbox"/> wrong choice of model type <input type="checkbox"/> poor business representation <input type="checkbox"/> poorly defined requirements <input type="checkbox"/> excessive detail <input type="checkbox"/> insufficient detail <input type="checkbox"/> process-oriented design <input type="checkbox"/> application-oriented design	<input type="checkbox"/> data model standards <input type="checkbox"/> data modeling guidelines <input type="checkbox"/> normalization guidelines <input type="checkbox"/> data requirements analysis <input type="checkbox"/> multiple levels of modeling (conceptual, logical, physical) <input type="checkbox"/> atomic data guidelines <input type="checkbox"/> data aggregation guidelines <input type="checkbox"/> subject-oriented design <input type="checkbox"/> consumer-oriented design <input type="checkbox"/> domain-based stewardship <input type="checkbox"/> data products and services
20	large change request backlog		
21	failure to meet business needs		
22	spreadsheet proliferation		
23	shadow systems and databases		
24	incorrect reporting		
25	high level of data disparity		
26	inefficient business analysis		
27	high level of data redundancy		
28	difficult-to-use data		
29	hard-to-navigate databases		
30	poor database performance		
31	business rule violations in data		

Data Quality Management

	Symptoms	Causes	Solutions
32	lack of trust	<input type="checkbox"/> poorly defined DQ rules <input type="checkbox"/> missing DQ rules <input type="checkbox"/> lack of data freshness SLAs <input type="checkbox"/> lack of runtime monitoring <input type="checkbox"/> lack of quality measures <input type="checkbox"/> lack of quality reporting <input type="checkbox"/> lack of accountability <input type="checkbox"/> incomplete/incorrect edits <input type="checkbox"/> poor data sampling	<input type="checkbox"/> DQ rules taxonomy <input type="checkbox"/> well-defined DQ rules <input type="checkbox"/> data contracts <input type="checkbox"/> DQ metrics and measures <input type="checkbox"/> published DQ reports <input type="checkbox"/> DQ scorecard <input type="checkbox"/> regular DQ audits <input type="checkbox"/> designated DQ accountability <input type="checkbox"/> DQ tasks in project plans <input type="checkbox"/> observability tools <input type="checkbox"/> data bias detection <input type="checkbox"/> data sampling standards <input type="checkbox"/> fit-for-purpose assessment
33	poor structural integrity		
34	incorrect data		
35	untimely data		
36	stale data		
37	incomplete data		
38	biased or unrepresentative data		
39	difficult-to-use data		
40	incorrect reporting		
41	inefficient business analysis		
42	business rule violations in data		
43	shadow systems and databases		
44	spreadsheet proliferation		

Data Integration & Interoperability

	Symptoms	Causes	Solutions
45	high level of data disparity	<input type="checkbox"/> poor integration architecture <input type="checkbox"/> technology-driven integration <input type="checkbox"/> lack of master data management <input type="checkbox"/> lack of APIs, services & products <input type="checkbox"/> poor integration practices <input type="checkbox"/> missing or wrong data sources <input type="checkbox"/> lack of shared semantics	<input type="checkbox"/> integration reference architecture <input type="checkbox"/> business-driven integration <input type="checkbox"/> data warehousing <input type="checkbox"/> master data management <input type="checkbox"/> APIs, data services & products <input type="checkbox"/> integration best practices <input type="checkbox"/> data sourcing standards & criteria <input type="checkbox"/> semantic model layers
46	overlapping and conflicting data		
47	data lineage gaps		
48	enterprise reporting difficulty		
49	application integration difficulty		
50	complex system interfaces		
51	limited data sharing		
52	spreadsheet proliferation		

Data Access

	Symptoms	Causes	Solutions
53	hard to find needed data	<input type="checkbox"/> missing metadata <input type="checkbox"/> lack of data catalog <input type="checkbox"/> missing or ineffective data curation <input type="checkbox"/> insufficient indexing <input type="checkbox"/> lack of cross-platform technology <input type="checkbox"/> inadequate search capabilities <input type="checkbox"/> poor user interface <input type="checkbox"/> database complexity <input type="checkbox"/> ineffective data security processes <input type="checkbox"/> excessive downtime <input type="checkbox"/> authorization barriers	<input type="checkbox"/> robust metadata <input type="checkbox"/> well-managed data catalog <input type="checkbox"/> data curation best practices <input type="checkbox"/> indexing & search capabilities <input type="checkbox"/> user-friendly tools & interfaces <input type="checkbox"/> data virtualization <input type="checkbox"/> service level accountability <input type="checkbox"/> access authorization procedures <input type="checkbox"/> APIs, data services & products
54	can't access needed data		
55	personal data compromised		
56	data security compromised		
57	sensitive data compromised		
58	access requests delayed or denied		
59	obsolete permissions still active		
60	data not available when needed		
61	hard-to-navigate databases		
62	poor data access performance		
63	spreadsheet proliferation		

Metadata Management

	Symptoms	Causes	Solutions
64	missing documentation	<input type="checkbox"/> casual metadata management <input type="checkbox"/> missing or weak data cataloging <input type="checkbox"/> missing or weak business glossary <input type="checkbox"/> fragmented metadata tools <input type="checkbox"/> lack of documentation standards <input type="checkbox"/> undocumented changes <input type="checkbox"/> no documentation incentives <input type="checkbox"/> no documentation reviews <input type="checkbox"/> no documentation accountability <input type="checkbox"/> "rush to production" projects <input type="checkbox"/> weak AI/ML metadata standards	<input type="checkbox"/> metadata templates & guidelines <input type="checkbox"/> well-managed data catalog <input type="checkbox"/> metadata integration tools <input type="checkbox"/> automated metadata capture <input type="checkbox"/> project documentation standards <input type="checkbox"/> metadata accountability <input type="checkbox"/> metadata tasks in project plans <input type="checkbox"/> incentives and reviews <input type="checkbox"/> AI model metadata standards <input type="checkbox"/> AI features metadata standards
65	incomplete documentation		
66	conflicting documentation		
67	hard to find documentation		
68	outdated documentation		
69	confusing documentation		
70	untraceable data		
71	uncertain business meaning		
72	data lineage gaps		
73	high level of data disparity		
74	high level of data redundancy		
75	AI/ML metadata gaps		

Database & Storage Management

	Symptoms	Causes	Solutions
76	insufficient data storage capacity	<input type="checkbox"/> ineffective storage management <input type="checkbox"/> passive growth management <input type="checkbox"/> ineffective performance tuning <input type="checkbox"/> unscheduled maintenance <input type="checkbox"/> inadequate database connectivity <input type="checkbox"/> insufficient backup and recovery <input type="checkbox"/> cloud misconfigurations	<input type="checkbox"/> continuous capacity planning <input type="checkbox"/> proactive growth management <input type="checkbox"/> performance SLAs <input type="checkbox"/> availability and uptime SLAs <input type="checkbox"/> connection protocol standards <input type="checkbox"/> database connectors <input type="checkbox"/> routine DBMS upgrades <input type="checkbox"/> backup & recovery best practices <input type="checkbox"/> configuration baselines and audits
77	unanticipated growth problems		
78	excessive cloud storage costs		
79	poor query performance		
80	poor update performance		
81	excessive database downtime		
82	unreliable database connections		
83	needed features not implemented		
84	corrupted data can't be repaired		
85	lost data can't be recovered		

Systems & Application Management

	Symptoms	Causes	Solutions
86	complex system interfaces	<input type="checkbox"/> poor data sharing architecture <input type="checkbox"/> lack of data integration <input type="checkbox"/> lack of data interoperability <input type="checkbox"/> poor application design <input type="checkbox"/> "quick fix" maintenance <input type="checkbox"/> "misfit" acquired systems <input type="checkbox"/> inconsistent data formats <input type="checkbox"/> little reuse of data functions <input type="checkbox"/> testing with production data	<input type="checkbox"/> application architecture standards <input type="checkbox"/> semantic data layer <input type="checkbox"/> application architecture standards <input type="checkbox"/> application design reviews <input type="checkbox"/> applied design patterns <input type="checkbox"/> modular & reusable processes <input type="checkbox"/> maintenance & testing standards <input type="checkbox"/> data sharing incentives <input type="checkbox"/> APIs and data products <input type="checkbox"/> managed test data & test cases
87	untraceable data		
88	poor application performance		
89	high level of data redundancy		
90	high level of data disparity		
91	poor data quality		
92	inadequate metadata		
93	data security compromised		
94	personal data compromised		
95	business rule violations in data		

Data Governance

	Symptoms	Causes	Solutions
96	poor data quality	<input type="checkbox"/> lack of data management goals <input type="checkbox"/> poorly defined responsibilities <input type="checkbox"/> unclear or ambiguous authority <input type="checkbox"/> poorly defined accountabilities <input type="checkbox"/> inadequate policy management <input type="checkbox"/> lack of policy compliance <input type="checkbox"/> regulatory complexity <input type="checkbox"/> understaffed data management <input type="checkbox"/> underfunded data management <input type="checkbox"/> lack of AI oversight	<input type="checkbox"/> "data as an asset" culture <input type="checkbox"/> "data as a resource" culture <input type="checkbox"/> focus on data literacy <input type="checkbox"/> designated data ownership <input type="checkbox"/> well-defined responsibilities <input type="checkbox"/> compliance frameworks <input type="checkbox"/> clearly designated authority <input type="checkbox"/> clearly defined accountabilities <input type="checkbox"/> executive data literacy <input type="checkbox"/> executive data sponsorship <input type="checkbox"/> usage and ethics guidelines <input type="checkbox"/> AI policies & AI governance
97	data security compromised		
98	personal data compromised		
99	regulatory non-compliance		
100	disaster recovery uncertainties		
101	territorialism inhibits data sharing		
102	data retention/disposal uncertainty		
103	data consolidation difficulties		
104	data ownership conflicts		
105	need for data standardization		
106	inappropriate use of data		

Data Protection

	Symptoms	Causes	Solutions
107	unauthorized data access	<input type="checkbox"/> weak access controls <input type="checkbox"/> over-privileged user accounts <input type="checkbox"/> poorly defined data classification <input type="checkbox"/> lack of sensitive data discovery <input type="checkbox"/> insufficient logging and monitoring <input type="checkbox"/> weak backup/recovery <input type="checkbox"/> lack of accountability and training <input type="checkbox"/> weak third-party controls <input type="checkbox"/> uncontrolled exports and sharing <input type="checkbox"/> fragmented incident response <input type="checkbox"/> poor password/credential practices	<input type="checkbox"/> role-based access control <input type="checkbox"/> least privilege access policies <input type="checkbox"/> data classification framework <input type="checkbox"/> automate data classification <input type="checkbox"/> automate sensitive data discovery <input type="checkbox"/> audit trails & automated monitoring <input type="checkbox"/> encryption and tokenization <input type="checkbox"/> compliance monitoring <input type="checkbox"/> data loss prevention practices <input type="checkbox"/> centralized incident response <input type="checkbox"/> data-subject rights policies <input type="checkbox"/> data subject rights procedures <input type="checkbox"/> password/credential policies
108	unmonitored data usage		
109	data loss incidents		
110	inappropriate use of data		
111	third party data leakage		
112	regulatory non-compliance		
113	shadow data copies		
114	inability to track data breaches		
115	personal data compromised		
116	sensitive data compromised		
117	data-subject rights compromised		
118	passwords & credentials exposed		

Data Preparation & Provisioning

	Symptoms	Causes	Solutions
119	inconsistent data formats	<input type="checkbox"/> no data prep workflow standards <input type="checkbox"/> poor quality source data <input type="checkbox"/> informal / ad hoc data integration <input type="checkbox"/> unclear data semantics <input type="checkbox"/> lack of data understanding <input type="checkbox"/> metadata deficiencies <input type="checkbox"/> poor documentation practices <input type="checkbox"/> tight coupling to source schemas <input type="checkbox"/> lack of pipeline versioning <input type="checkbox"/> lack of PII handling standards <input type="checkbox"/> biased data filtering & sampling	<input type="checkbox"/> pipeline workflow standards <input type="checkbox"/> data profiling and validation rules <input type="checkbox"/> automated deduplication <input type="checkbox"/> semantic data model <input type="checkbox"/> automated metadata capture <input type="checkbox"/> data observability <input type="checkbox"/> data quality monitoring <input type="checkbox"/> documentation standards <input type="checkbox"/> schema drift detection & handling <input type="checkbox"/> data pipeline versioning <input type="checkbox"/> automate sensitive data discovery <input type="checkbox"/> bias detection methods <input type="checkbox"/> fairness review of prepared data
120	incomplete data		
121	duplicate records		
122	data lineage gaps		
123	data outliers and anomalies		
124	excessive manual effort		
125	undocumented transformations		
126	unverified data sources		
127	schema changes break pipelines		
128	non-reproducible datasets		
129	preparation exposes sensitive data		
130	preparation introduces data bias		

Data Analysis

	Symptoms	Causes	Solutions
131	inconsistent analysis results	<input type="checkbox"/> inconsistent metrics definitions <input type="checkbox"/> missing semantic layer <input type="checkbox"/> excessive use of spreadsheets <input type="checkbox"/> undocumented analysis methods <input type="checkbox"/> limited analytic model governance <input type="checkbox"/> insufficient validation of results <input type="checkbox"/> inadequate self-service literacy <input type="checkbox"/> limited self-service governance	<input type="checkbox"/> business glossary <input type="checkbox"/> semantic layer <input type="checkbox"/> repeatable analysis workflows <input type="checkbox"/> model validation and peer review <input type="checkbox"/> documented analysis methods <input type="checkbox"/> analysis process transparency <input type="checkbox"/> bias detection methods <input type="checkbox"/> focus on data literacy <input type="checkbox"/> self-service policies & standards
132	misinterpreted metrics		
133	conflicting analytical models		
134	excessive manual effort		
135	difficulty reproducing results		
136	lack of trust		
137	data or model bias		
138	unverified data sources		
139	erroneous self-service analytics		

Data-Driven Automation

	Symptoms	Causes	Solutions
140	automation fails if data changes	<ul style="list-style-type: none"> <input type="checkbox"/> schema-dependent scripts <input type="checkbox"/> insufficient testing of workflows <input type="checkbox"/> manual exception handling <input type="checkbox"/> semantic misalignment <input type="checkbox"/> weak or missing AI governance <input type="checkbox"/> lack of performance metrics <input type="checkbox"/> inadequate documentation <input type="checkbox"/> lack of bias detection and controls <input type="checkbox"/> excessive AI agent access <input type="checkbox"/> lack of GenAI verification <input type="checkbox"/> unclear automation ownership 	<ul style="list-style-type: none"> <input type="checkbox"/> schema-flexible design <input type="checkbox"/> workflow testing and validation <input type="checkbox"/> automated triage of exceptions <input type="checkbox"/> semantic layer <input type="checkbox"/> AI/ML model governance <input type="checkbox"/> performance monitoring <input type="checkbox"/> documentation standards <input type="checkbox"/> automated bias detection <input type="checkbox"/> human oversight <input type="checkbox"/> least-privilege AI agent accounts <input type="checkbox"/> verification gates for GenAI <input type="checkbox"/> designated automation owners
141	inconsistent automation results		
142	excessive manual effort		
143	hidden dependencies		
144	data or model bias		
145	decision rule drift		
146	lack of transparency		
147	monitoring and audit difficulties		
148	over-privileged agent access		
149	AI actions lack output verification		
150	decisions without accountability		

Diagnostic Guide: Index of Symptoms

access requests delayed or denied	58
AI actions lack output verification	149
AI/ML metadata gaps	75
application integration difficulty	49
automation fails if data changes.....	140
biased or unrepresentative data.....	38
business rule violations in data	31, 42, 95
can't access needed data	54
complex system interfaces	50, 86
conflicting analytical models	133
conflicting documentation.....	66
confusing abbreviations	5
confusing documentation	69
corrupted data can't be repaired	84
data consolidation difficulties	103
data lineage gaps.....	47, 72, 122
data loss incidents.....	109
data not available when needed	60
data or model bias.....	137, 144
data outliers and anomalies.....	123
data ownership conflicts.....	104
data retention/disposal uncertainty.....	102
data security compromised	56, 93, 97
data-subject rights compromised	117
decision rule drift.....	145
decisions without accountability	150
difficult-to-use data	28, 39
difficulty reproducing results	135
disaster recovery uncertainties	100
duplicate records.....	121
enterprise reporting difficulty	48
erroneous self-service analytics.....	139
excessive cloud storage costs	78
excessive database downtime	81
excessive manual effort	124, 134, 142
failure to meet business needs	21
hard to find data definitions	14
hard to find documentation.....	67
hard to find needed data	53
hard-to-identify data	8
hard-to-navigate databases.....	29, 61
hidden dependencies	143
high level of data disparity.....	9, 17, 25, 45, 73, 90
high level of data redundancy.....	18, 27, 74, 89
inability to track data breaches	114
inadequate metadata	92
inappropriate use of data	16, 106, 110
incomplete data	37, 120
incomplete documentation	65
inconsistent analysis results	131

Diagnostic Guide: Index of Symptoms (continued)

inconsistent automation results	141
inconsistent data formats	119
incorrect data	34
incorrect data definitions	11
incorrect data names	3
incorrect reporting	24, 40
inefficient business analysis	26, 41
insufficient data storage capacity	76
lack of data definitions	10
lack of transparency	146
lack of trust	32, 136
large change request backlog	20
limited data sharing	51
lost data can't be recovered	85
meaningless data definitions	12
meaningless data names	1
misinterpreted metrics	132
missing documentation	64
misunderstood data	15
monitoring and audit difficulties	147
multiple names and aliases	6
need for data standardization	105
needed features not implemented	83
non-reproducible datasets	128
non-unique data names	2
obsolete data definitions	13
obsolete permissions still active	59
outdated documentation	68
over-privileged agent access	148
overlapping and conflicting data	46
passwords & credentials exposed	118
personal data compromised	55, 94, 98, 115
poor application performance	88
poor data access performance	62
poor data quality	91, 96
poor database performance	30
poor query performance	79
poor structural integrity	19, 33
poor update performance	80
preparation exposes sensitive data	129
preparation introduces data bias	130
regulatory non-compliance	99, 112
schema changes break pipelines	127
sensitive data compromised	57, 116
shadow data copies	113
shadow systems and databases	23, 43
spreadsheet proliferation	22, 44, 52, 63
stale data	36
structureless data names	4
territorialism inhibits data sharing	101

Diagnostic Guide: Index of Symptoms (continued)

third party data leakage..... 111
unanticipated growth problems 77
unauthorized data access 107
uncertain business meaning..... 71
undocumented transformations 125
unmonitored data usage..... 108
unnamed data components..... 7
unreliable database connections 82
untimely data 35
untraceable data..... 70, 87
unverified data sources 126, 138