

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



حکمرانی و مدیریت داده‌ها

جلسه ۳

داده‌ها و حکمرانی آن

Data and Its Governance

کاظم فولادی قلعه
دانشکده مهندسی، دانشکدگان فارابی
دانشگاه تهران

<http://courses.fouladi.ir/datagov>

داده‌ها و حکمرانی آن

۱

سیل داده‌ها

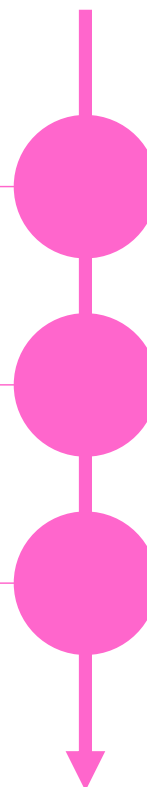
اعصار تطور داده‌ها

THE ERAS OF DATA EVOLUTION

پیش از پیدایش کامپیوتر و پایگاه‌های داده‌ها

پس از پیدایش کامپیوتر و حافظه‌های الکترونیکی پیش از عصر اینترنت

از عصر اینترنت به بعد



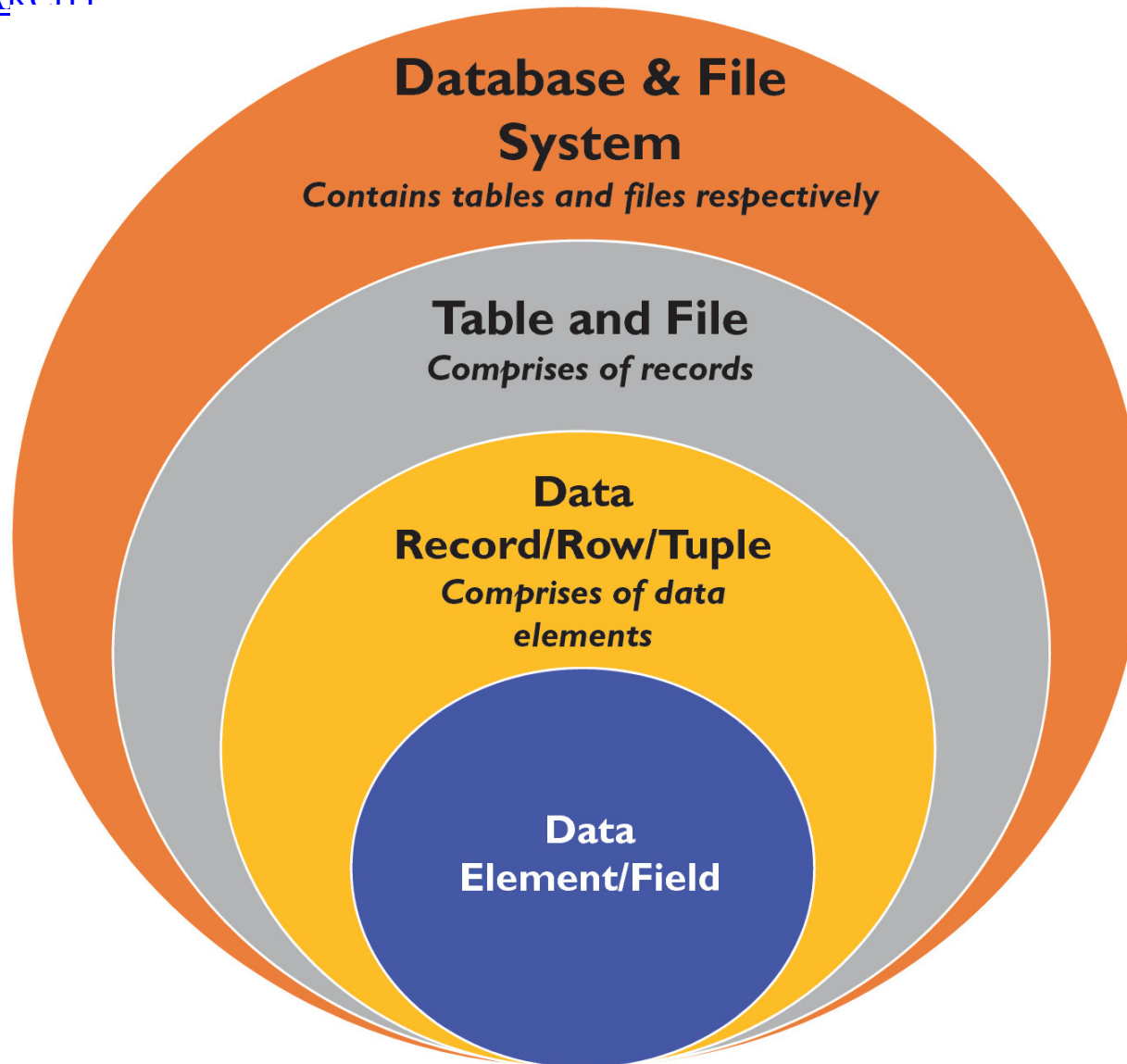
داده‌ها و حکمرانی آن

۲

درباره
داده‌ها

و
سازمان‌دهی
آنها

سلسله‌مراتب داده‌ها

THE DATA HIERARCHY

«دریاچه داده‌ها» در مقابل «انبار داده‌ها»

دریاچه داده‌ها (Data Lake)	در مقابل	انبار داده‌ها (Data Warehouse)
هر داده‌ای (ساخت‌یافته، بدون ساختار و نیمه ساخت‌یافته) را به شکل اصلی خود ذخیره می‌کند	انواع داده‌ها و قالب	داده‌های ساخت‌یافته را در قالبی از پیش تعریف شده ذخیره می‌کند
به سخت‌افزار تخصصی نیاز ندارد	سخت‌افزار	به سخت‌افزار تخصصی نیاز دارد
طرحواره در خواندن بازیابی داده‌ها کندتر است بر اساس استراتژی Extract، Transform، Load یا ETL عمل می‌کند	پردازش	طرحواره روی نوشتن بازیابی داده‌ها سریع‌تر است بر اساس استراتژی Extract، Transform، Load یا ELT عمل می‌کند
بسیار چابک، پیکربندی و پیکربندی مجدد در صورت نیاز به راحتی با تغییرات سازگار می‌شود	چابکی	کمتر چابک، پیکربندی ثابت
استفاده از داده‌ها از پیش تعریف نشده است	استفاده	استفاده از داده‌ها ممکن است به چند هدف عملیاتی از پیش تعریف شده محدود شود. تا زمانی که کاربردها تعریف نشده باشند، داده‌ها در انبار داده‌ها بارگذاری نمی‌شوند
برای ذخیره‌سازی کم‌هزینه طراحی شده است	ذخیره‌سازی	برای حجم داده‌های بزرگ گران است
کمتر بالغ	امنیت	بالغ‌تر
دانشمندان داده و مهندسان داده	کاربران	تحلیل‌گران کسب‌وکار و تحلیل‌گران داده

داده‌ها و حکمرانی آن

۳

داده‌ها
به‌عنوان
دارایی
و
حکمرانی آن

خصوصیات یکتای داده‌ها

UNIQUE PROPERTIES OF DATA



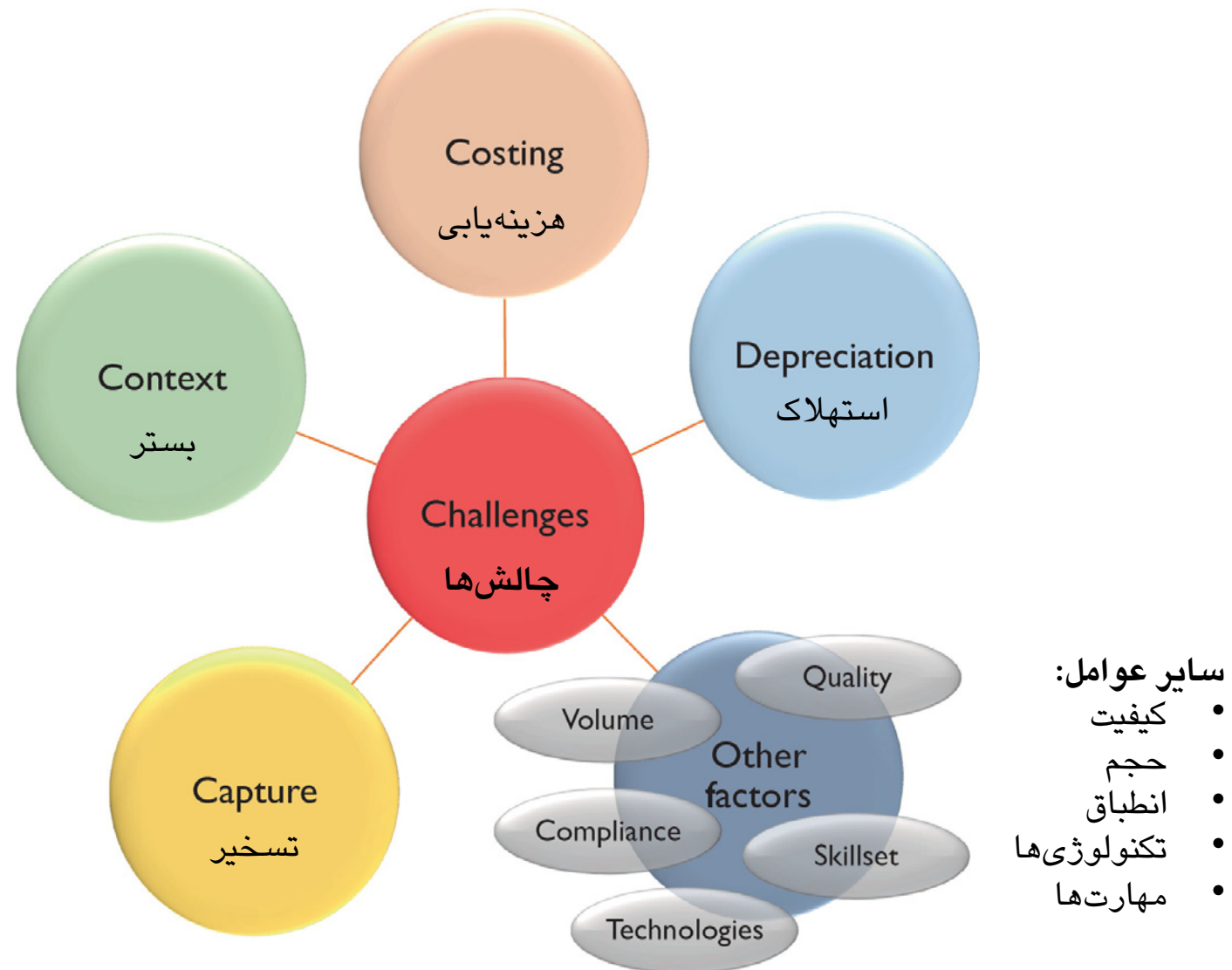
مقایسه خصوصیات دارایی‌های ثابت با دارایی‌های داده‌ای

Fixed assets	Data assets
Tangible	Intangible
Cannot be shared or not easily shared	Easily shared
Cannot be replicated, transformed or deleted easily	Easily replicated, transformed, and deleted
Cannot be copied	Can be copied
Cannot be in multiple places at the same time	Spread across organization
Consumed with use	Not consumed with use
Not accumulative	Accumulative
Non-contextual	Contextual
Rivalrous	Non-rivalrous
Fungible	Non-fungible
Wears out	Does not wear out
Cannot be sold without being given away	Can be sold without being given away

مقایسه خصوصیات دارایی‌های ثابت با دارایی‌های داده‌ای

دارایی‌های داده‌ای	دارایی‌های ثابت
ناملموس	لموس
به‌سادگی به اشتراک گذاشته می‌شود	نمی‌تواند (کلاً / به‌سادگی) به اشتراک گذاشته شود
به‌سادگی کپی، تبدیل یا حذف می‌شود	نمی‌تواند به‌سادگی کپی، تبدیل یا حذف شود
می‌تواند کپی شود	نمی‌تواند کپی شود
در سراسر سازمان پخش می‌شود	نمی‌تواند در یک زمان در چند جا باشد
با استفاده از آن مصرف نمی‌شود	با استفاده از آن مصرف می‌شود
انباشته است	انباشته نیست
وابسته به مضمون است	وابسته به مضمون نیست
غیررقابتی است	رقابتی است
تعویض‌پذیر نیست	تعویض‌پذیر است
فرسوده نمی‌شود	فرسوده می‌شود
بدون واگذاری می‌تواند فروخته شود	بدون واگذاری نمی‌تواند فروخته شود

چالش‌های فهرست کردن داده‌ها در ترازنامه

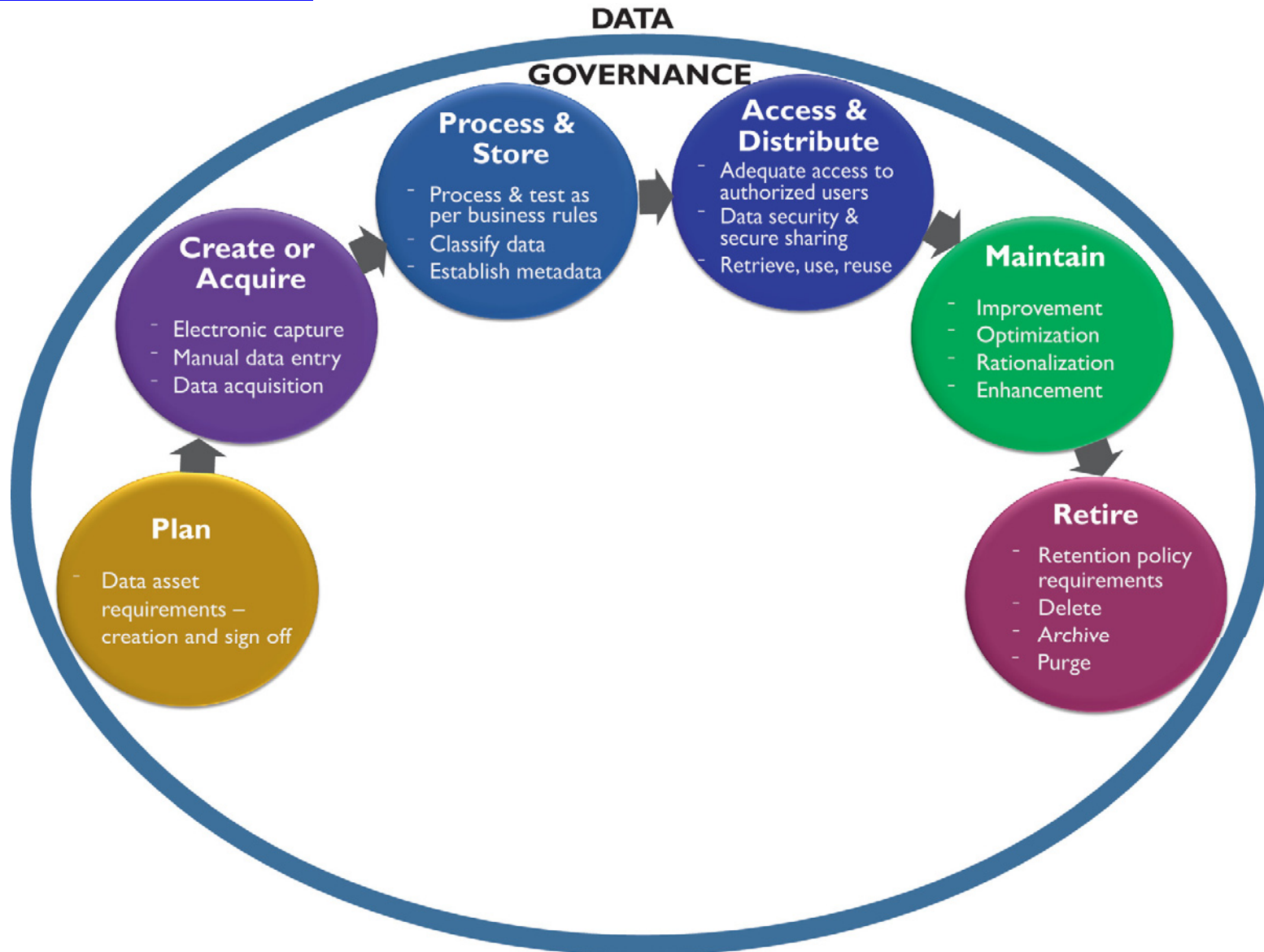
CHALLENGES OF LISTING DATA IN THE BALANCE SHEET

داده‌ها و حکمرانی آن

۴

چرخه
زندگی
دارایی
داده‌ای

چرخه زندگی دارایی داده‌ای

THE DATA ASSET LIFE CYCLE

پردازش و ذخیره

- فرآیند و آزمایش طبق قوانین تجاری
- طبقه بندی داده
- ها ایجاد ابرداده

دسترسی و توزیع

- دسترسی کافی به کاربران مجاز
- امنیت داده و به اشتراک گذاری امن
- بازیابی، استفاده، استفاده مجدد

حفظ کنید

- بهبود
- بهینه سازی
- عقلانی سازی
- افزایش

بازنشسته شوید

- سیاست حفظ و نگهداری الزامات
- حذف کنید
- آرشیو
- پاکسازی

ایجاد یا کسب کنید

- ضبط الکترونیکی
- ورود اطلاعات به صورت دستی
- اکتساب داده

برنامه ریزی کنید

- ایجاد الزامات
- دارایی داده و امضای آن

داده‌ها و حکمرانی آن

۵

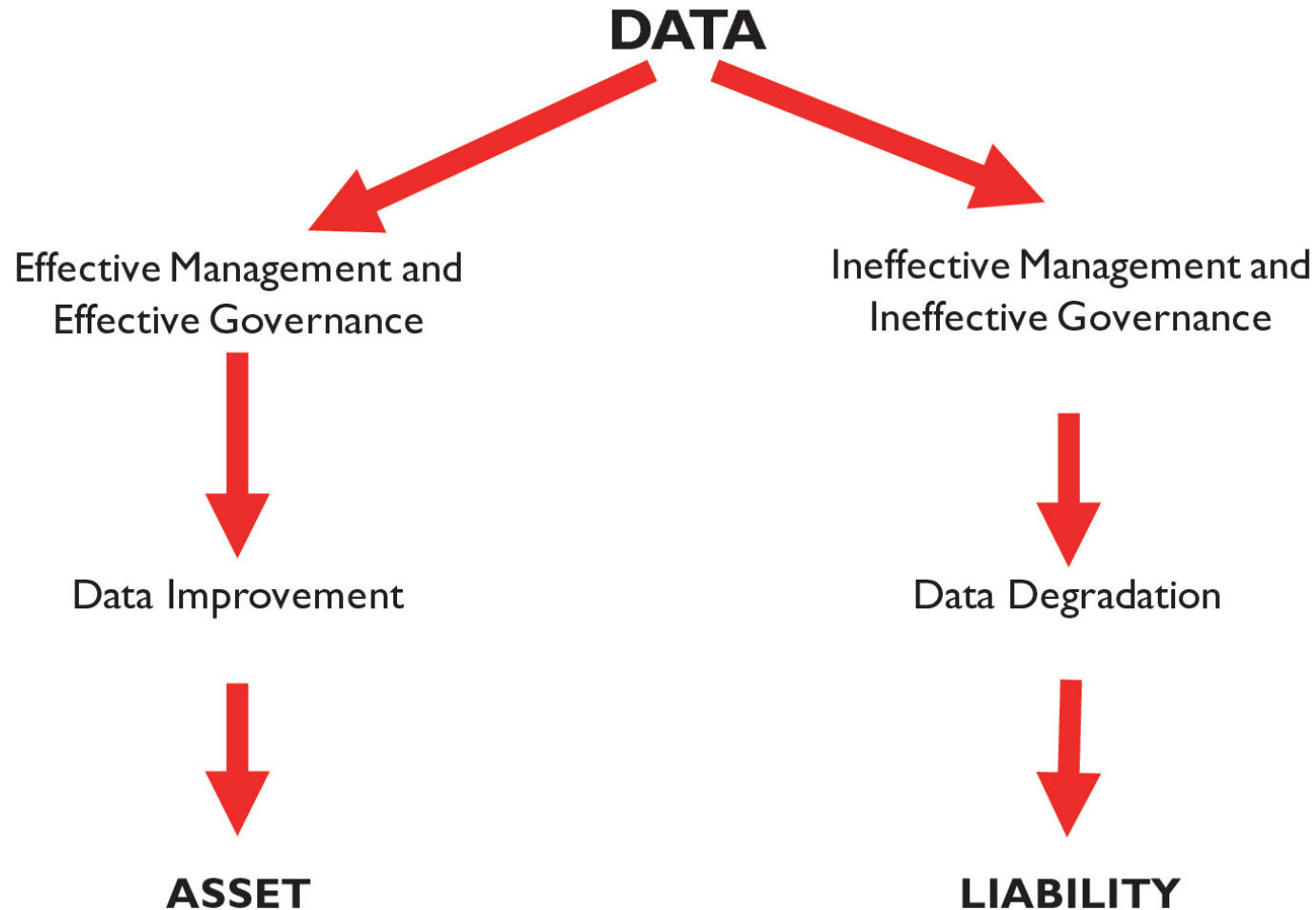
مسائل
معمول در
عدم برخورد
با داده‌ها
به‌عنوان
دارایی

مسائل معمول در عدم برخورد با داده‌ها به عنوان دارایی

COMMON PROBLEMS WITH DATA NOT BEING TREATED AS AN ASSET

While most organizations recognize in principle that their data is an asset, they only pay lip-service to the way the concept is applied in practice resulting in data issues in organizations. In other words, there is a huge gap between preaching and practice.

دارایی داده‌ای در برابر بدهکاری داده‌ای

DATA ASSET VERSUS DATA LIABILITY

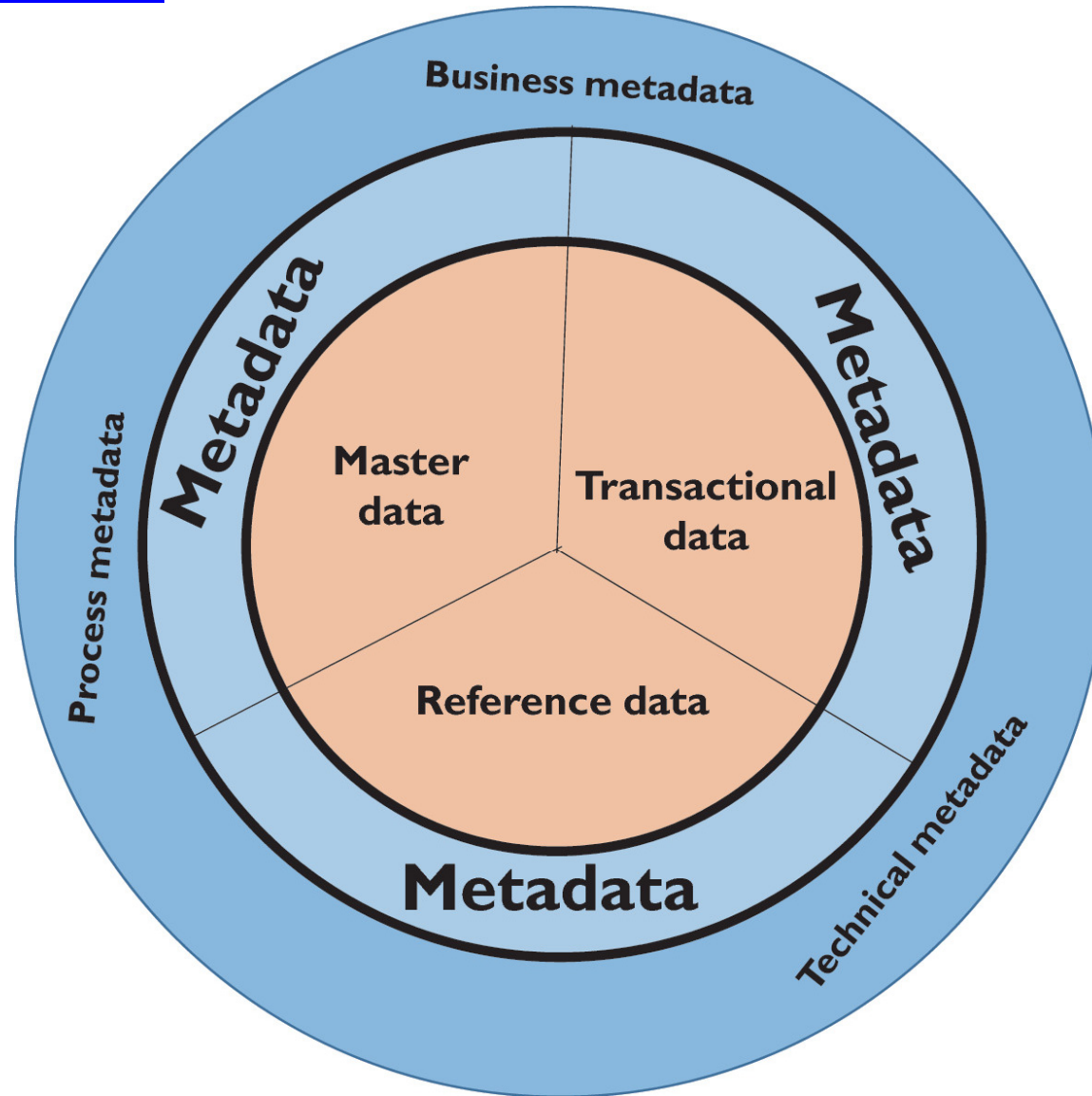
When you do not treat data as an asset, you do not manage and govern data adequately, resulting in degradation of data and it becoming a liability to the organization

داده‌ها و حکمرانی آن

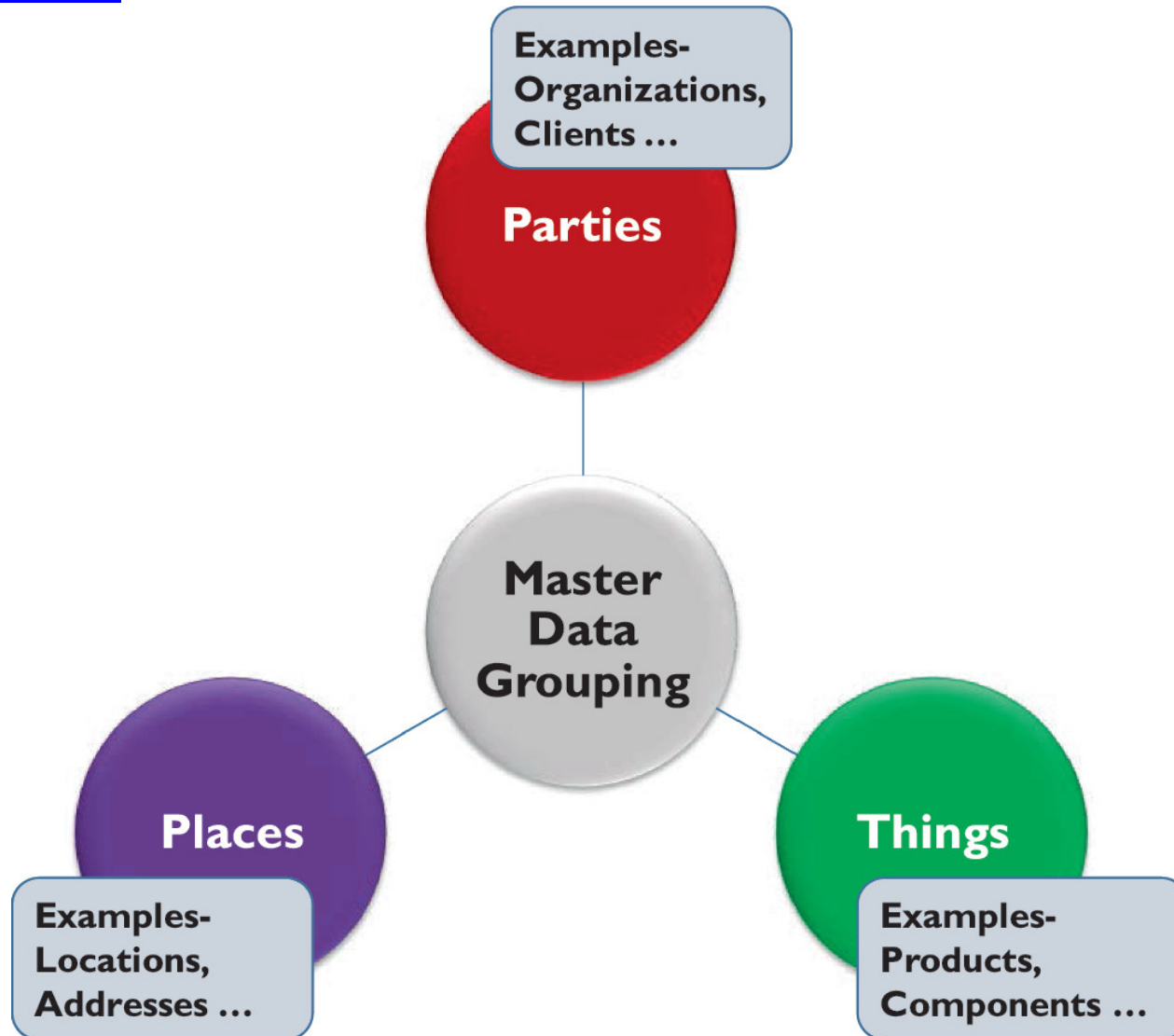
۶

طبقه‌بندی
داده‌ها

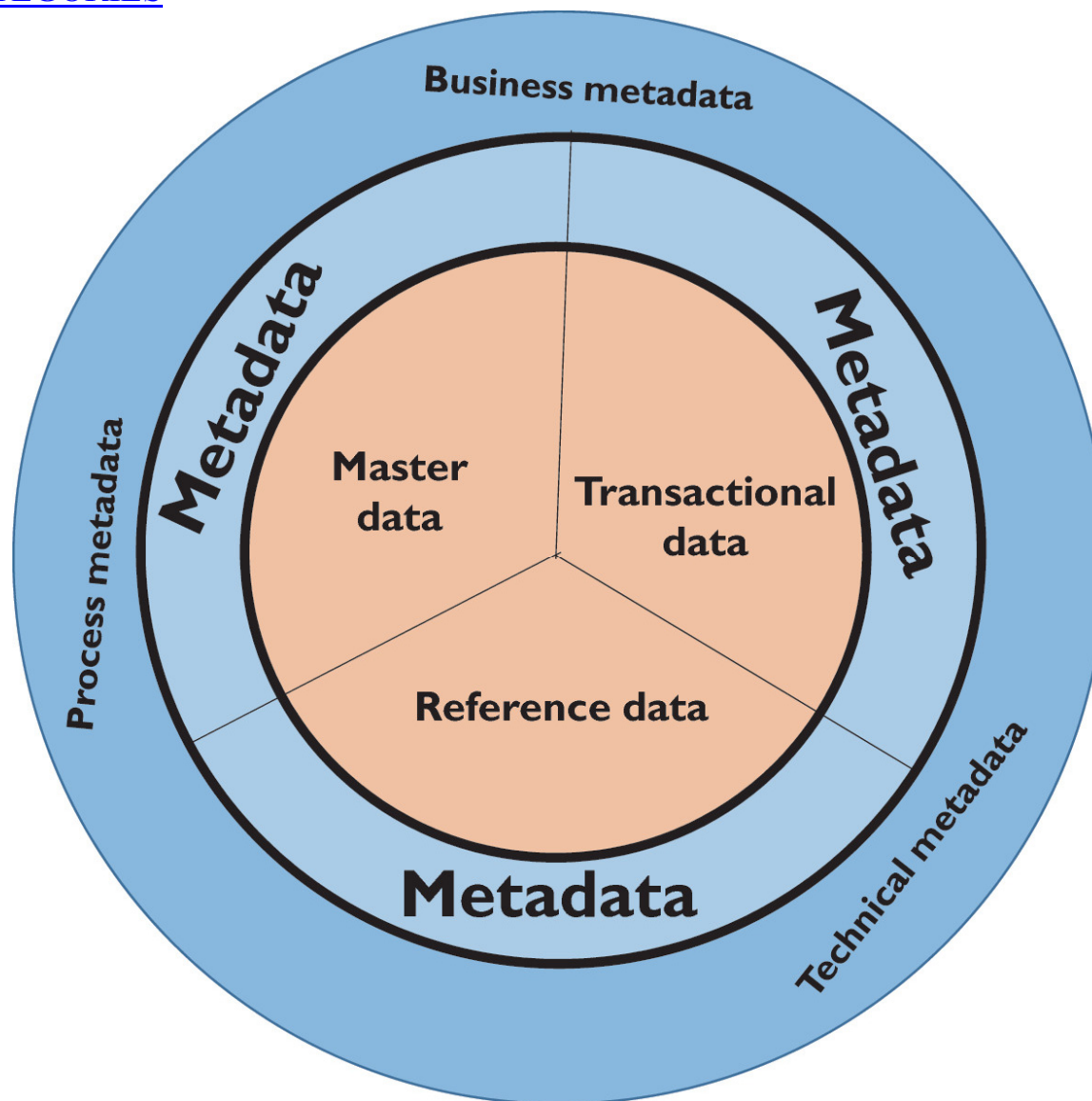
دسته‌های موجودیت داده‌ها

ENTITY—DATA CATEGORIES

گروه‌های داده‌های اصلی

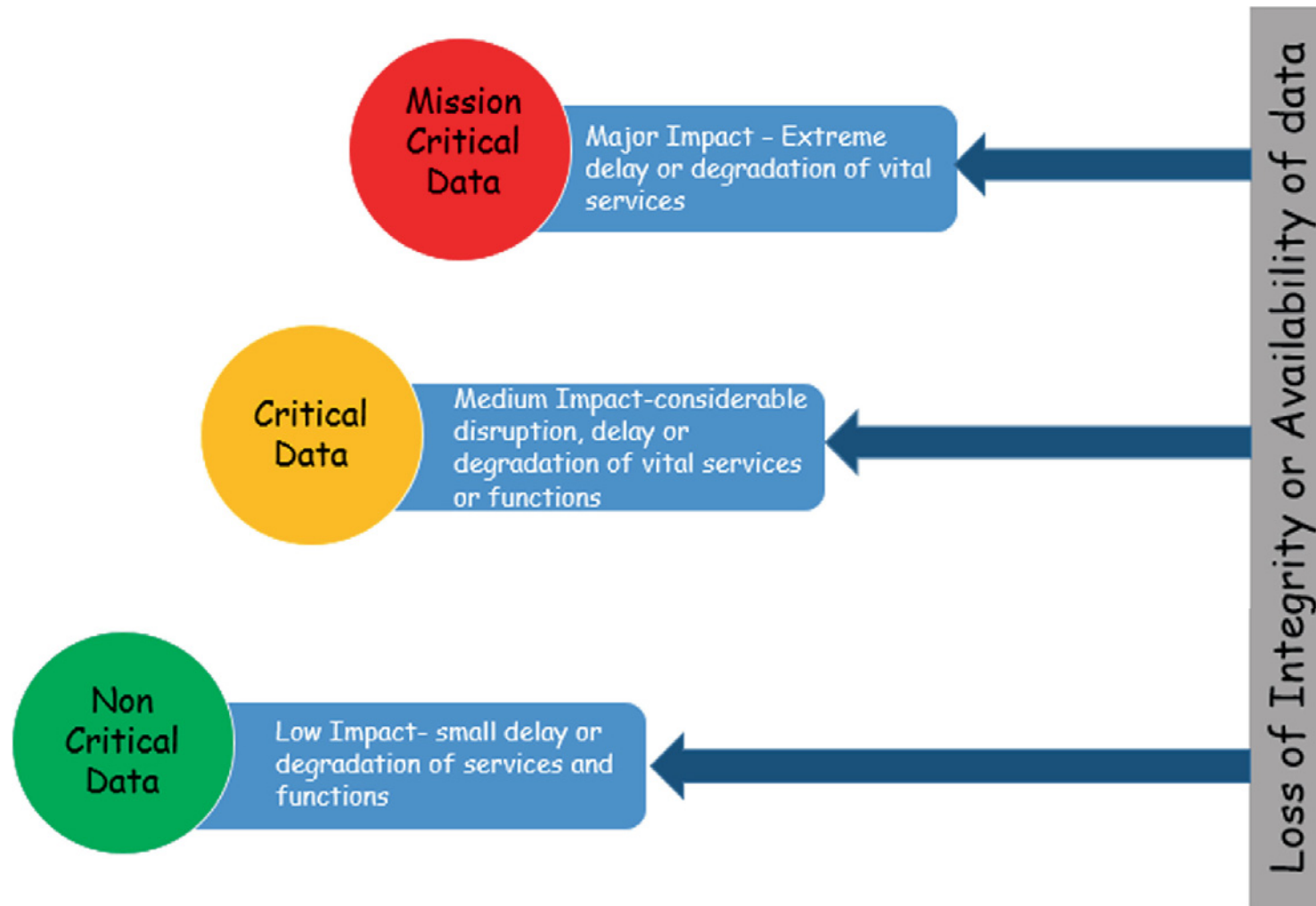
MASTER DATA GROUPS

دسته‌های موجودیت داده‌ها

ENTITY—DATA CATEGORIES

طبقه‌بندی داده‌ها مبتنی بر بحرانی بودن داده‌ها

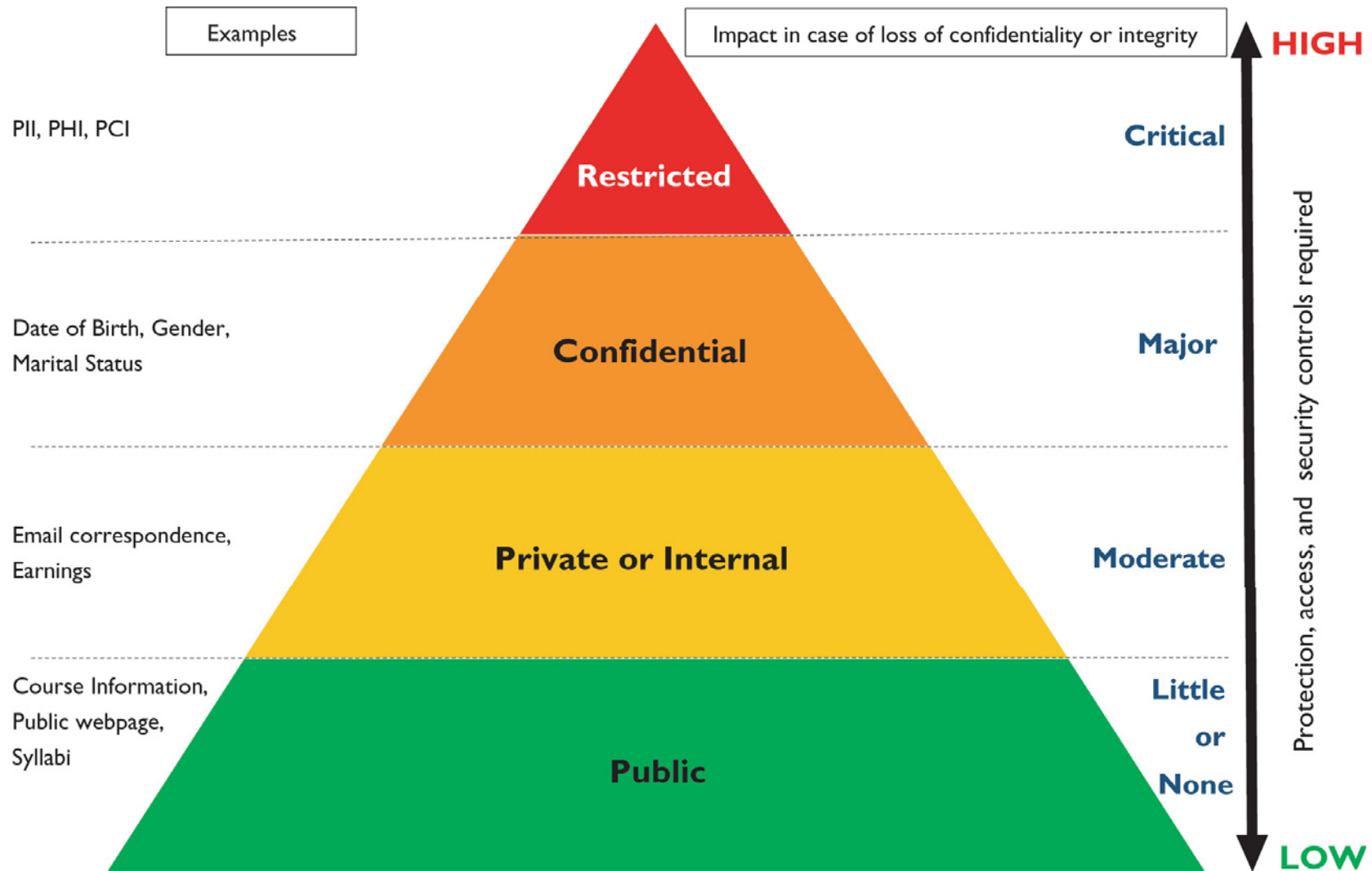
DATA CLASSIFICATION BASED ON DATA CRITICALITY



Data criticality is a reflection of how important is the data to an organization's mission, functions, and processes in terms of integrity and availability, and can be divided into three categories.

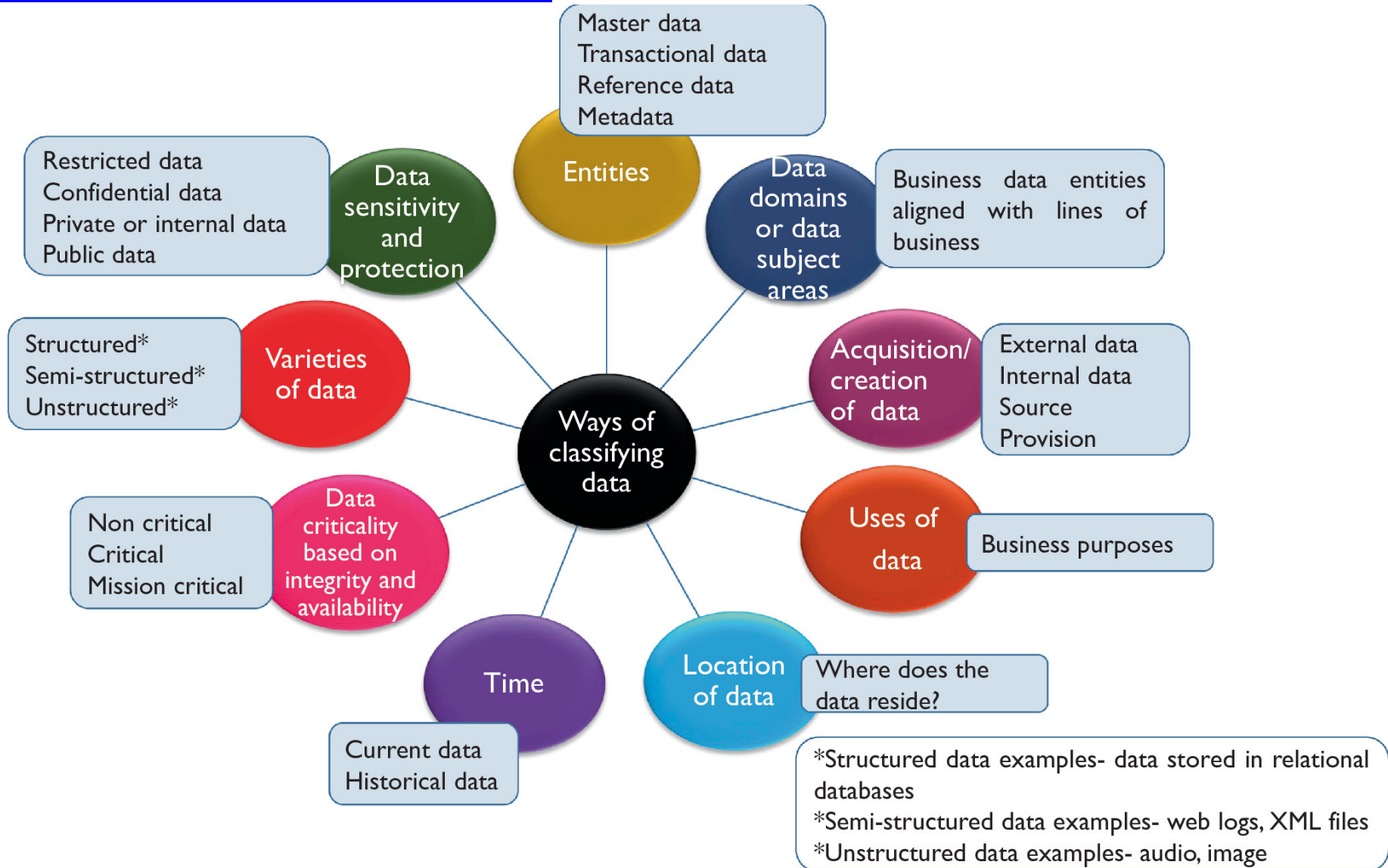
طبقه‌بندی داده‌ها بر اساس حساسیت و حفاظت داده‌ها

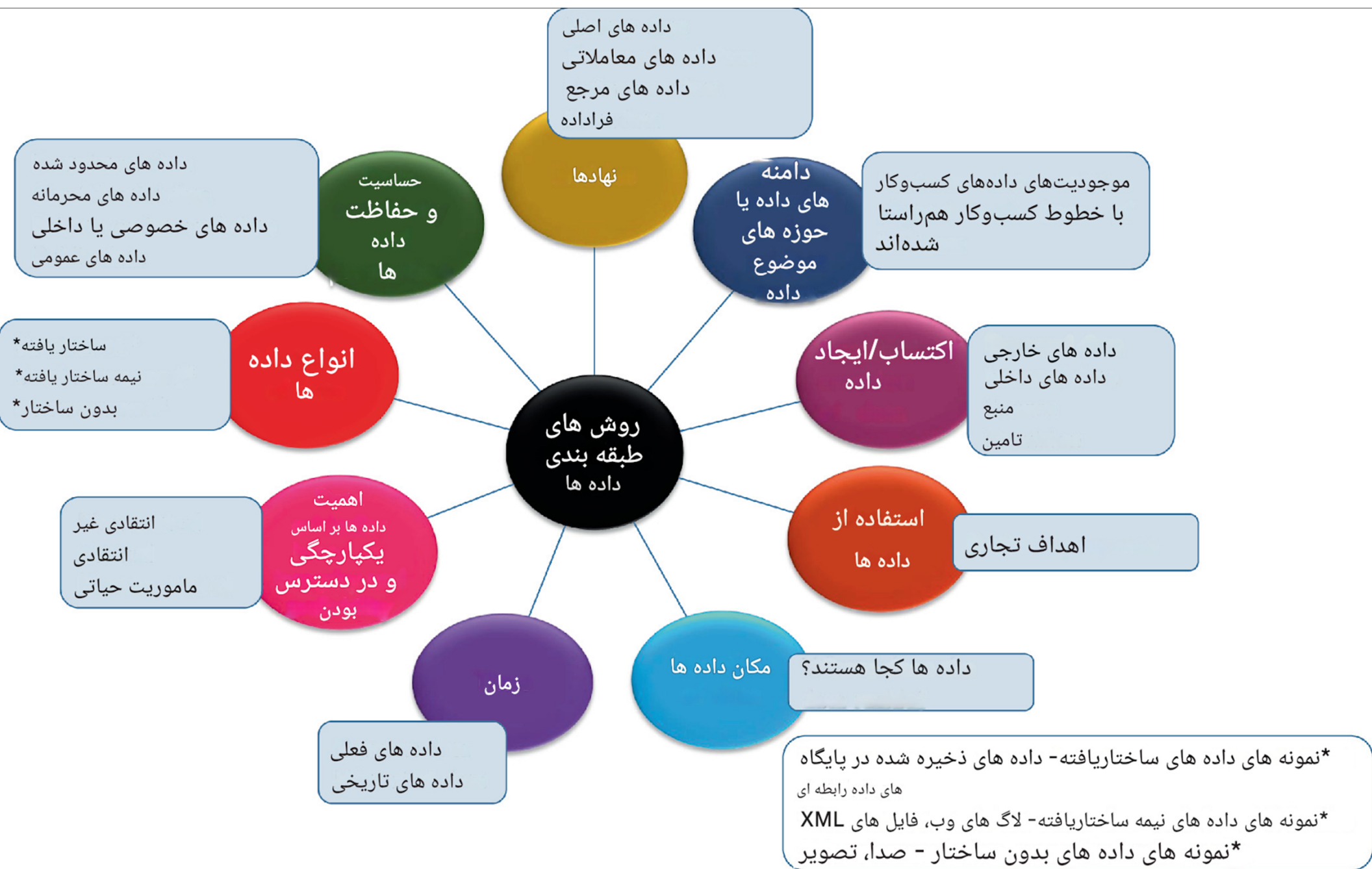
DATA CLASSIFICATION BY DATA SENSITIVITY AND PROTECTION



راه‌های مختلف طبقه‌بندی داده‌ها

DIFFERENT WAYS OF CLASSIFYING DATA





داده‌ها و حکمرانی آن

۷

کیفیت داده‌ها و ابعاد آن

ابعاد کیفیت داده‌ها

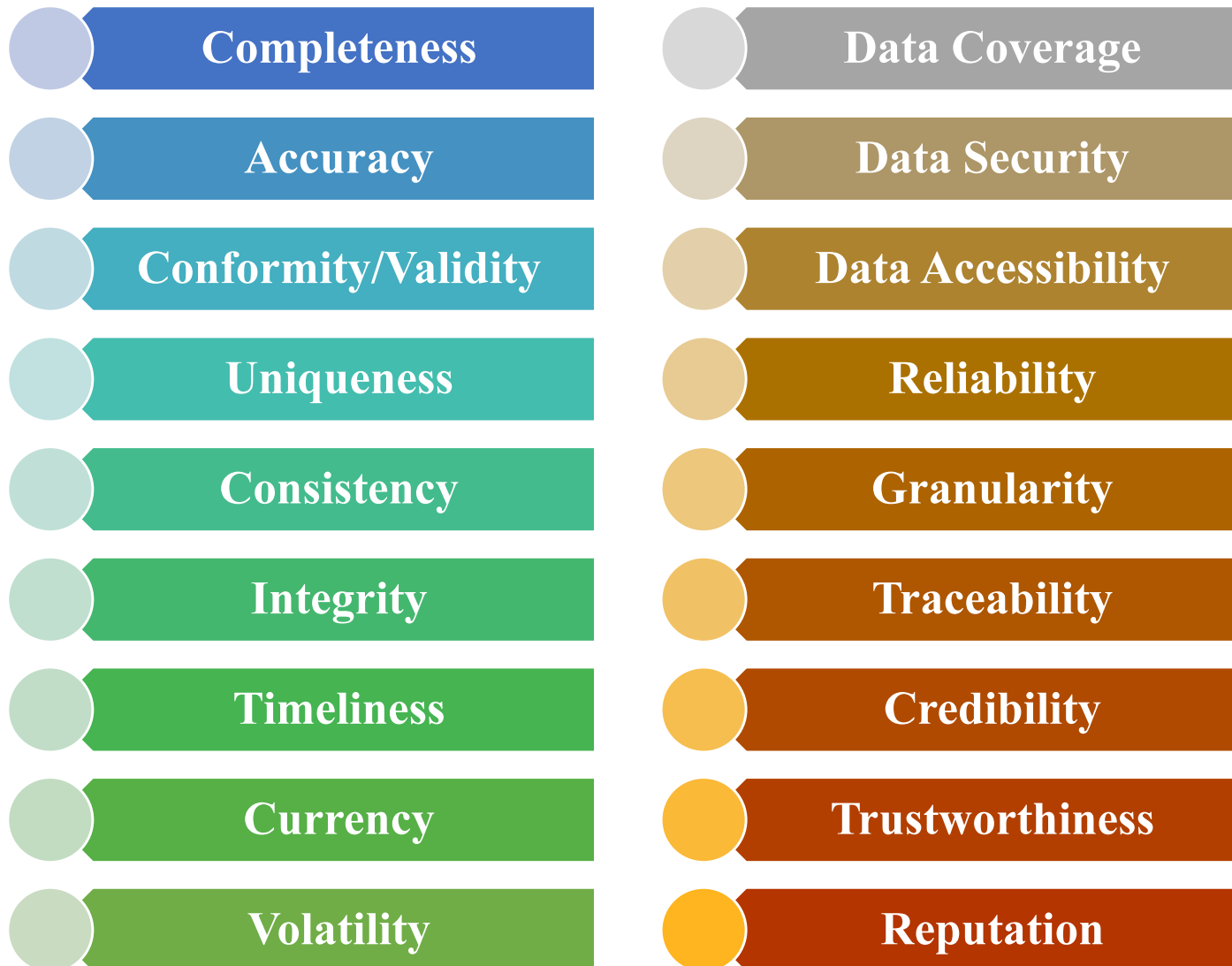
DATA QUALITY DIMENSIONS

Table 2.3 Duplicate record for customer—“Monalisa King”

Customer ID	First name	Last name	Street address	Postal code	State code	Country code	Email ID
CID0000278	Monalisa	Zing	10, DF Ave, SOP	2127	NSW	AU	King12@gmail.com
CID0000290	Mona	Zing	12, O’Neill Ave, Newington	2127	NSW	AU	King12@gmail.com
CID0000357	Lisa	Zing	12, O’Neill Ave, Newington	2127	NSW	AU	King12@gmail.com

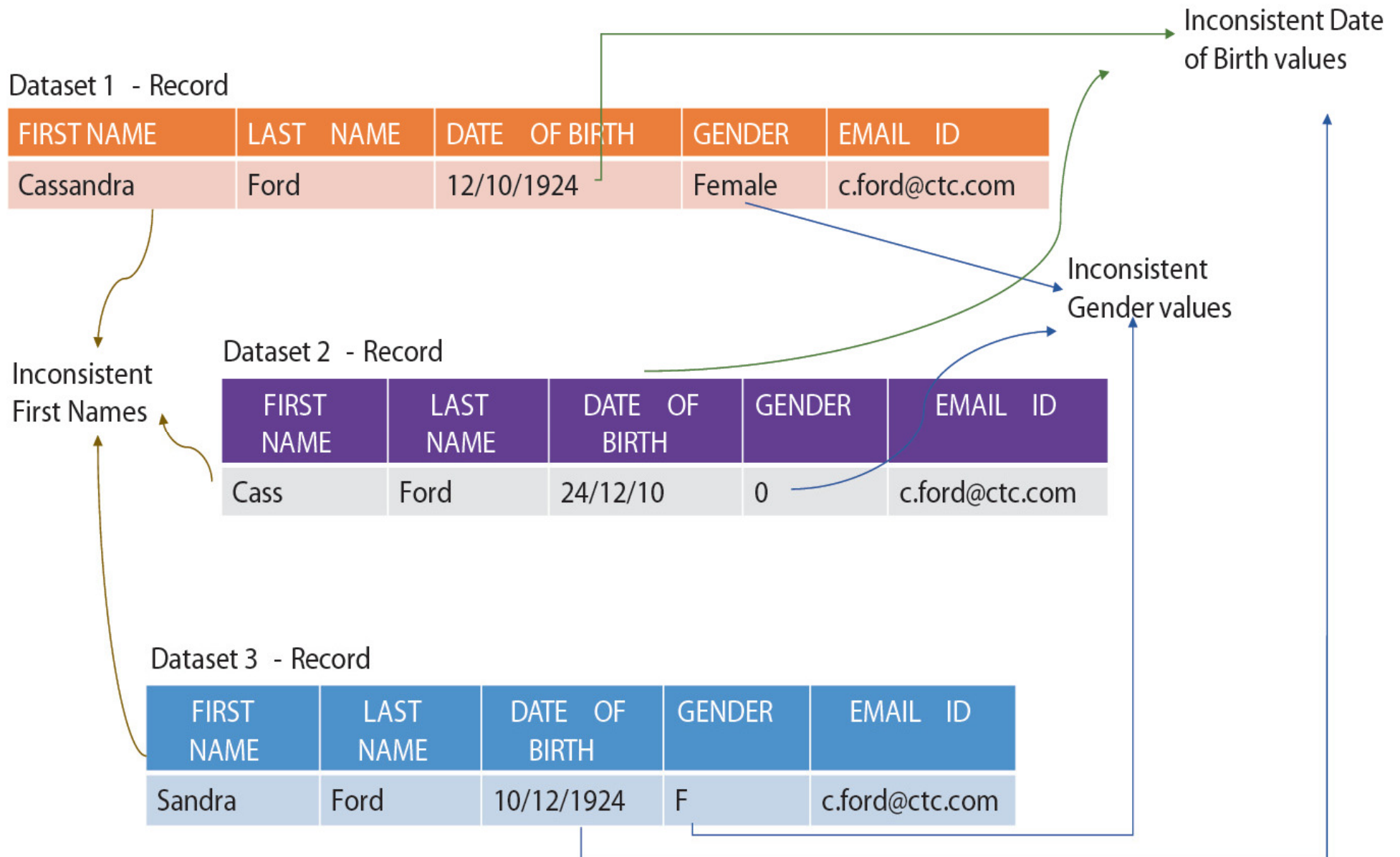


Fig. 2.11 Data inconsistency example

داده‌ها و حکمرانی آن



نیاز به
حکمرانی
داده
خوب

مسائل ناشی از حکمرانی داده ناکارآمد یا نبود حکمرانی داده

ISSUES ARISING FROM INEFFECTIVE DATA GOVERNANCE OR ABSENCE OF DATA GOVERNANCE

Multiple versions of truth

Lack of ownership or accountability of data

Data quality issues

Lack of understanding and use of data

Loss of credibility of data source and lack of trust in data

Difficulty in locating enterprise data

Difficulty in conforming to regulatory requirements

Increasing risk of data breaches and legal and compliance issues

Loss of reputation due to data breaches and legal and compliance issues

داده‌ها و حکمرانی آن

۹

حکمرانی
داده
رسمی /
غیررسمی

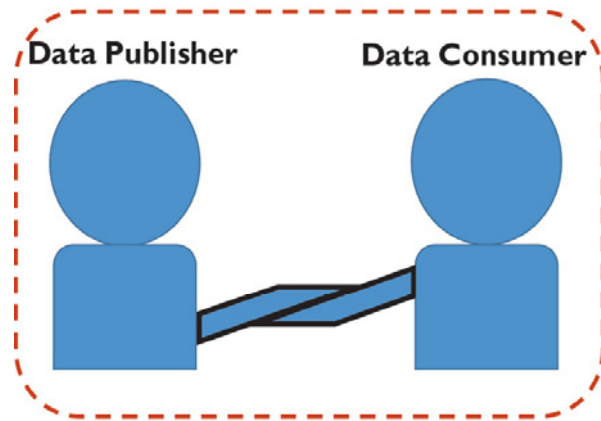
مقایسه حکمرانی داده رسمی و غیررسمی

حکمرانی داده رسمی	حکمرانی داده غیررسمی
یک فرآیند رسمی برای توسعه و تغییر سیاست‌های داده و استانداردهای داده وجود دارد که در سراسر سازمان اجرا و دنبال می‌شود.	خط‌مشی‌های داده و استانداردهای داده به‌صورت تصادفی توسعه‌یافته یا تغییر می‌کنند یا فرآیند توسعه و تغییری در جریان است که به‌طور گسترده اجرا یا دنبال نمی‌شود.
زمان‌بندی جمع‌آوری داده‌ها، تعاریف داده‌ها و/یا فرآیندهای داده تنها پس از در نظر گرفتن تأثیر بر مؤلفه‌های داخلی و خارجی تغییر می‌کنند، تنها به روش‌هایی که با هدف و چشم‌انداز سیستم داده‌های گسترده‌تر سازگار باشد.	برنامه جمع‌آوری داده‌ها، تعاریف داده‌ها، و/یا فرآیندهای داده به شکلی تصادفی و با کمترین توجه به همسویی با سایر تلاش‌های داده‌های آژانس تغییر می‌کنند. این تغییرات گزارش‌دهی مداوم و استفاده از داده‌ها را دشوار می‌کند.
نقش‌ها و مسئولیت‌های داده به‌وضوح تعریف شده است، در خط مشی نوشته شده است و به پایبندی آن نظارت می‌شود. کارکنان جدید در مورد سیاست‌های داده و نقش خاص آنها آموزش دیده‌اند.	نقش داده‌ها و مسئولیت‌های کارکنان آژانس داخلی و خارجی سیال است، اغلب تغییر می‌کند و/یا فقط برای تعداد کمی از کارکنان مشخص است.
اشتراک‌گذاری داده‌های تنظیم‌شده در سطح ابتدایی در میان کارکنان آژانس داخلی و خارجی و در بین آژانس‌های خارجی مجاز است، با طبقه‌بندی داده‌ها و فرآیندهای دسترسی به داده‌ها و مکانیسم‌های تأیید برای اطمینان از اینکه فقط کاربران مجاز مجاز به دسترسی به داده‌ها هستند.	اشتراک‌گذاری غیرقانونی داده در سطح ابتدایی در میان کارکنان آژانس داخلی و خارجی و در بین آژانس‌های خارجی مجاز است.
ورودی فعال ذی‌نفعان	ذی‌نفعان فرصت کمی برای ارائه ورودی در مورد سیستم‌های داده و مسائل مربوط به داده‌ها دارند.
داده‌ها فقط توسط کارکنان با مجوزهای اختصاص داده شده بر اساس موقعیت و آموزش قابل دسترسی است و داده‌های حساس رمزگذاری شده است.	سخت‌گیری محدود یا کمتر در مورد امنیت داده‌ها و دسترسی به داده‌ها
کنترل‌های مؤثر برای علامت‌گذاری و کاهش خطرات و اطمینان از کیفیت بالای داده‌ها	فقدان کنترل‌های مؤثر برای نظارت بر ریسک‌ها یا اطمینان از کیفیت بالای داده‌ها
اطمینان حاصل می‌شود که فرآیندهایی برای دفع منظم و مداوم داده‌های ناخواسته وجود دارد.	داده‌های ناخواسته تمایل به انباشته شدن دارند.
فرآیندهای حل مسئله داده‌ها تعریف و رعایت می‌شوند، و نقش‌ها و مسئولیت‌ها برای اولویت‌بندی، بررسی و حل مسائل داده‌ها به‌صورت به‌موقع وجود دارد.	فرآیند حل مسئله داده‌ها adhoc است. حقوق تصمیم‌گیری، نقش‌ها و مسئولیت‌ها تعریف نشده‌اند یا به‌خوبی تعریف نشده‌اند.

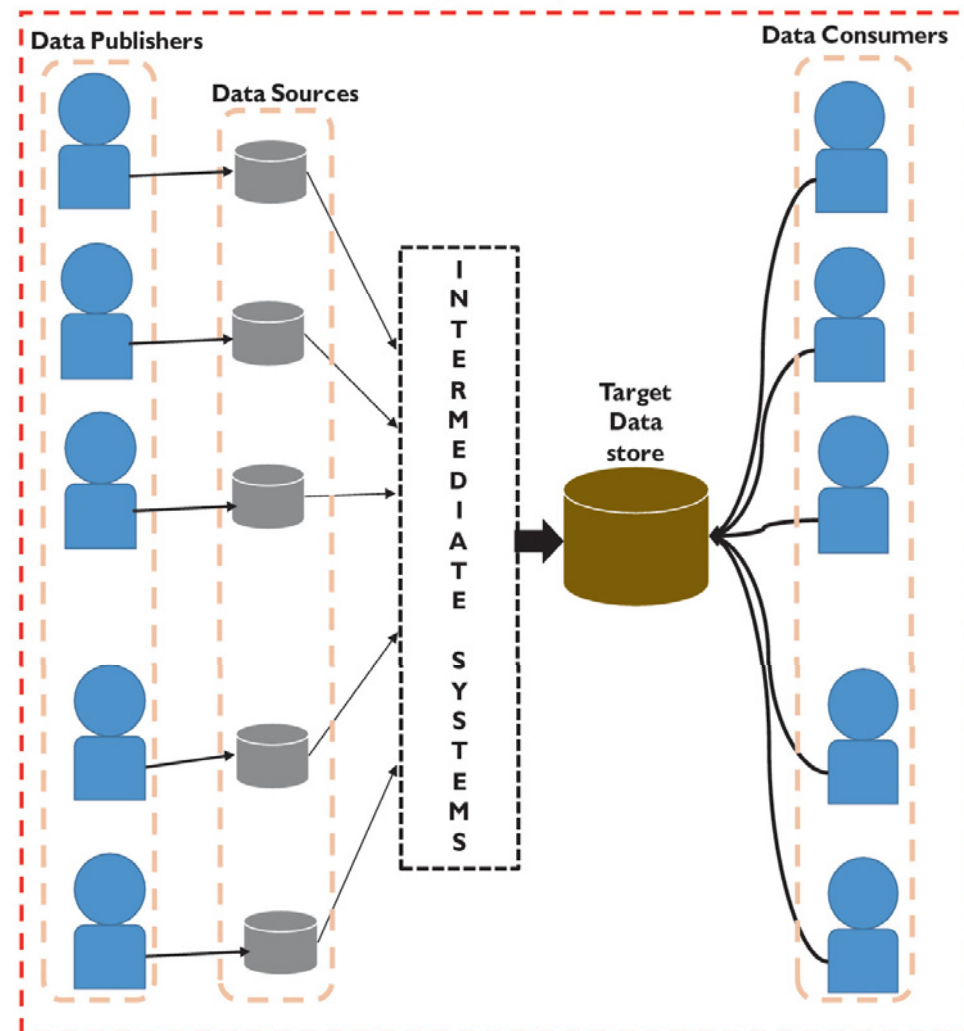
حکمرانی سبک در برابر حکمرانی سنگین

LIGHT VERSUS HEAVY GOVERNANCE

Lighter data governance and data management practices



Heavier and rigorous data governance and data management practices



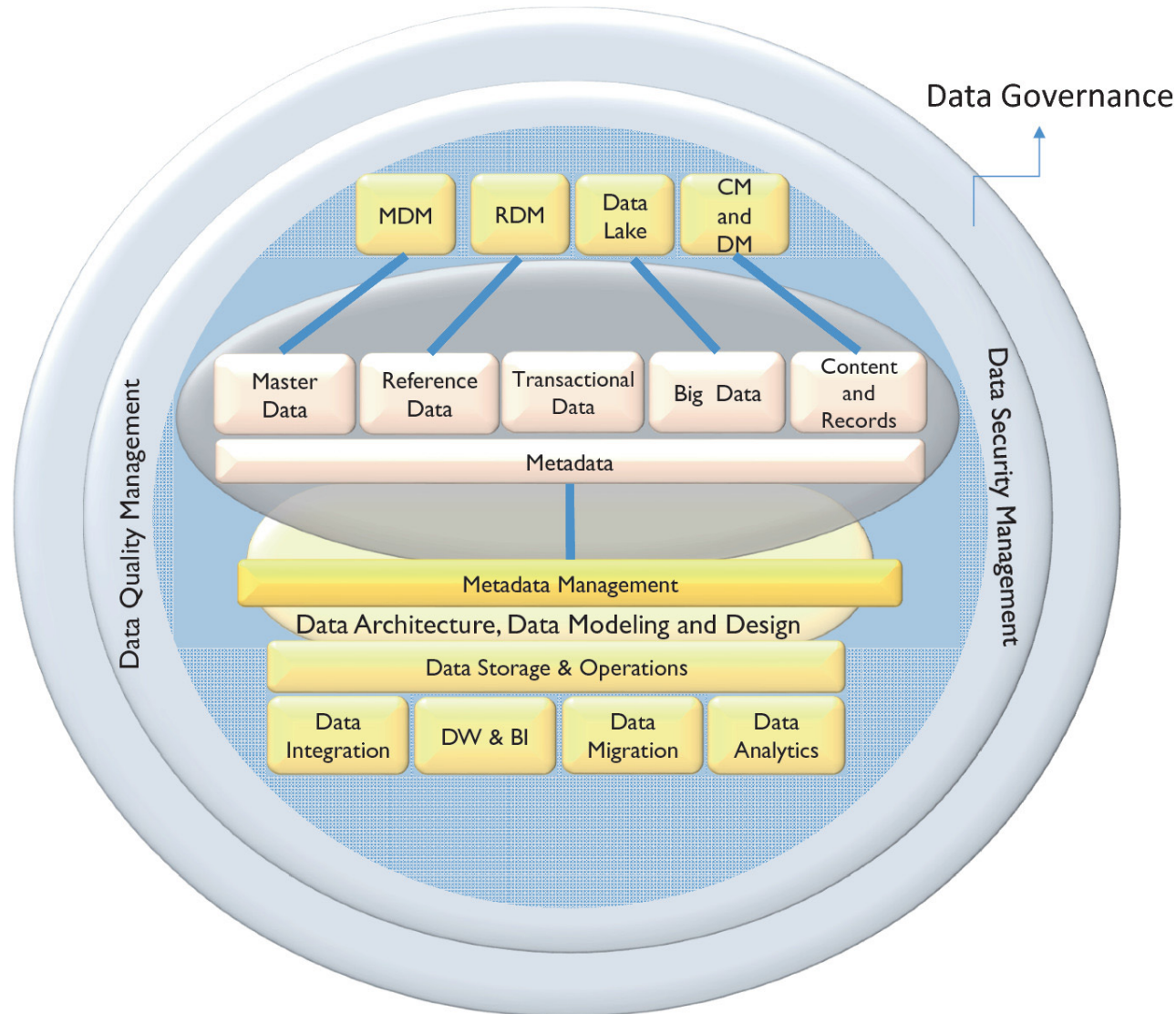
داده‌ها و حکمرانی آن

۱۰

تمایز
حکمرانی
داده
از
مدیریت
داده

حکمرانی داده سعی می‌کند کارکردهای مدیریت داده را جمع کند

DATA GOVERNANCE TYING TOGETHER THE DATA MANAGEMENT FUNCTIONS

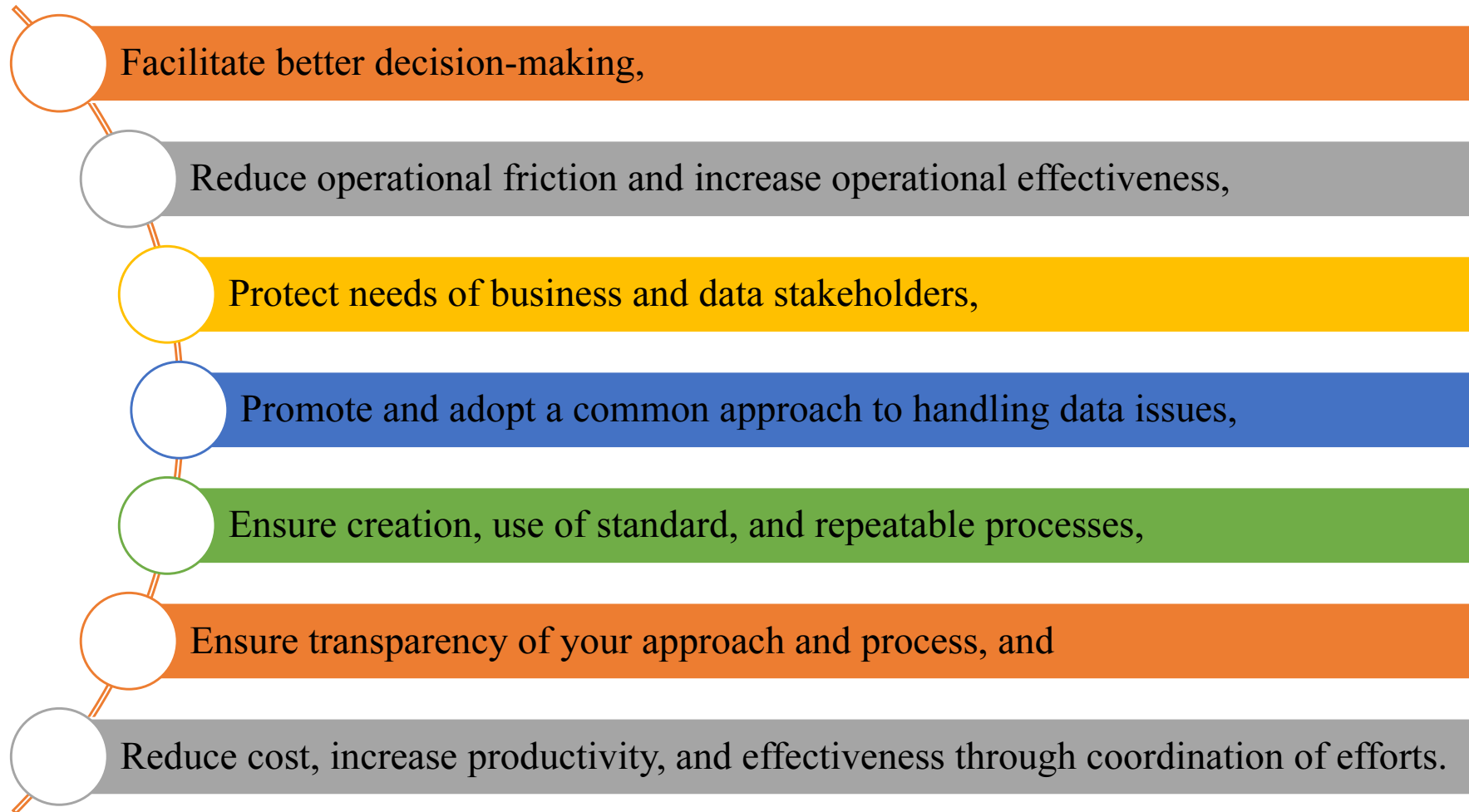


داده‌ها و حکمرانی آن

۱۱

اهداف حکمرانی داده‌ها

اهداف حکمرانی داده‌ها

DATA GOVERNANCE GOALS

داده‌ها و حکمرانی آن

۱۲

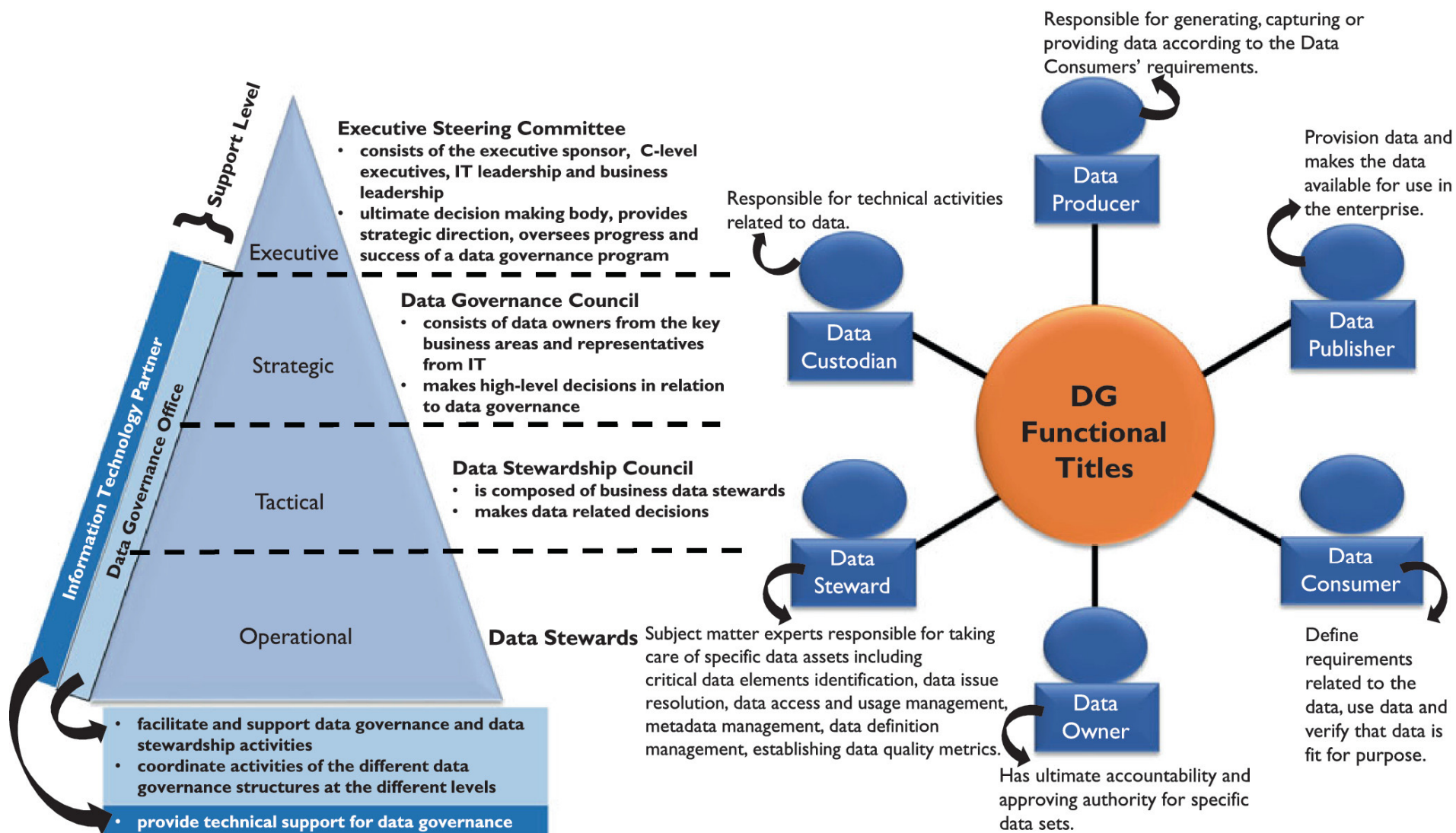
عناصر کلیدی حکمرانی داده‌ها

عناصر کلیدی حکمرانی داده‌ها

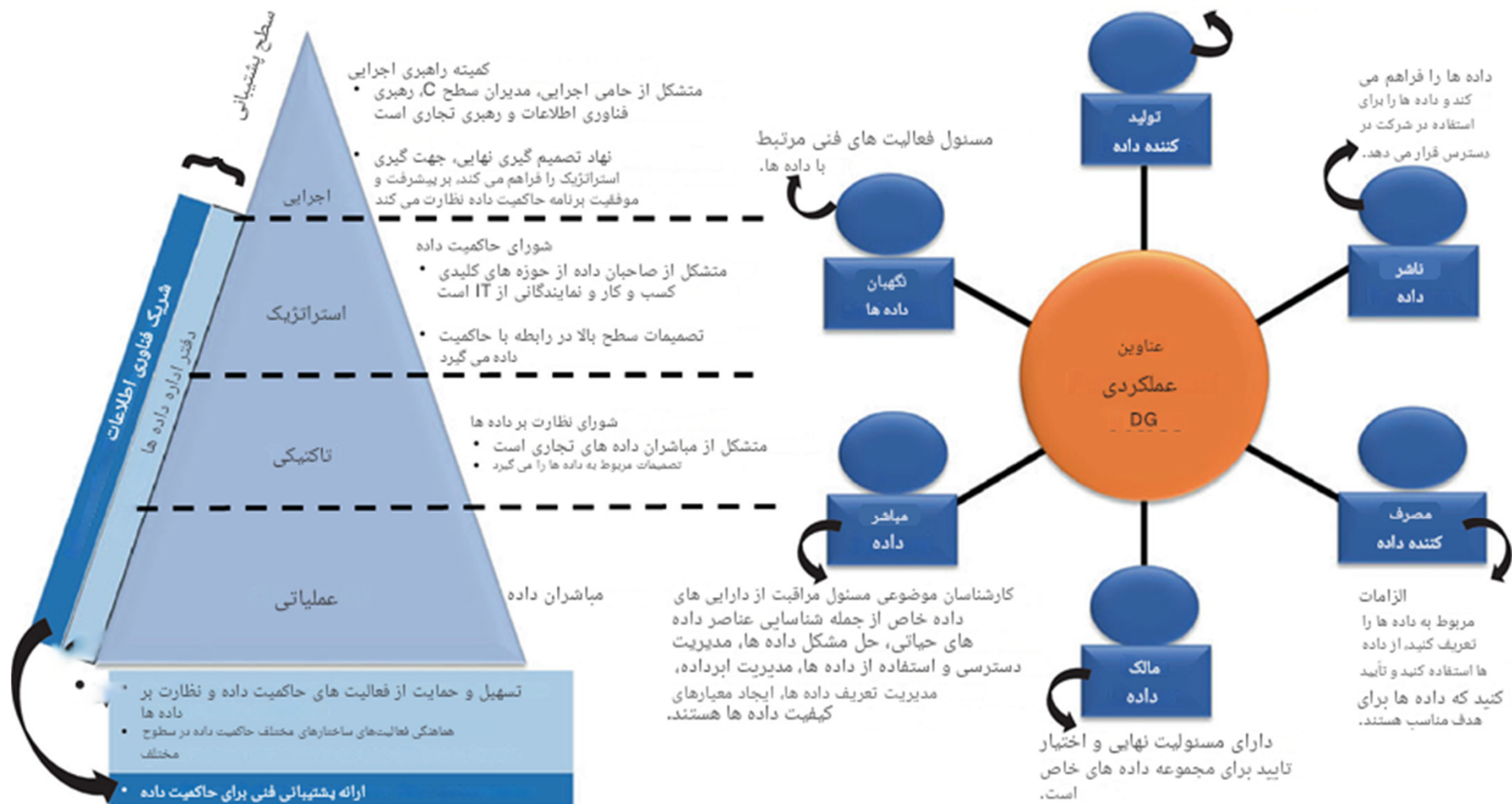
DATA GOVERNANCE—THE KEY ELEMENTS

جنبه «افراد» در حکمرانی داده‌ها

PEOPLE ASPECT OF DATA GOVERNANCE



مسئول تولید، جمع‌آوری یا ارائه داده‌ها بر اساس نیاز مصرف‌کنندگان داده است.



داده‌ها و حکمرانی آن

۱۳

گردانندگان
کلیدی
کسب و کار و
موارد کاربرد
حکمرانی
داده‌ها

محرك‌های كسب‌وکار و موارد کاربرد حکمرانی داده‌ها

DATA GOVERNANCE BUSINESS DRIVERS AND USE CASES





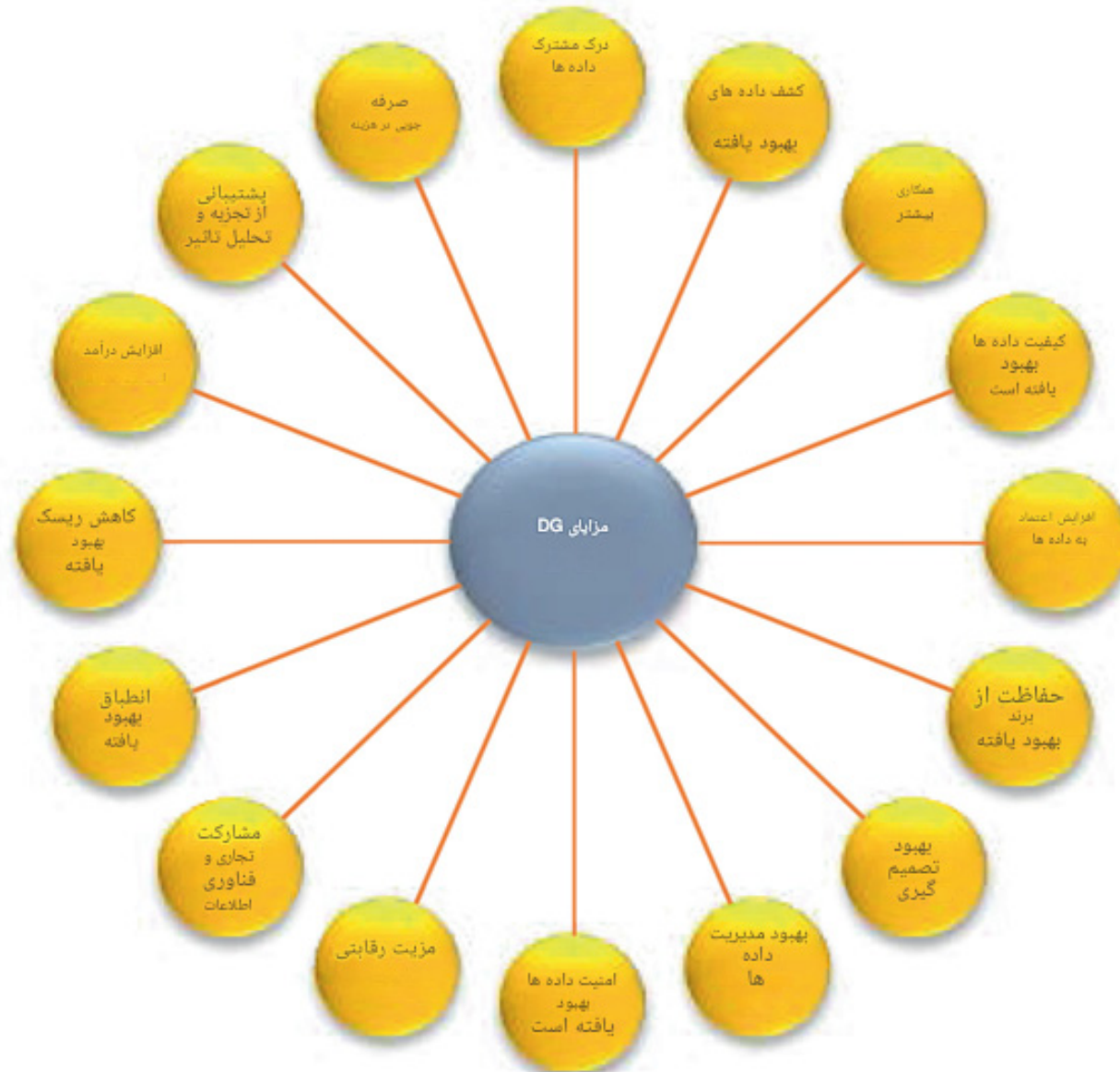
داده‌ها و حکمرانی آن

۱۴

منافع
کلیدی
حکمرانی
داده‌ها

منافع حکمرانی داده‌ها

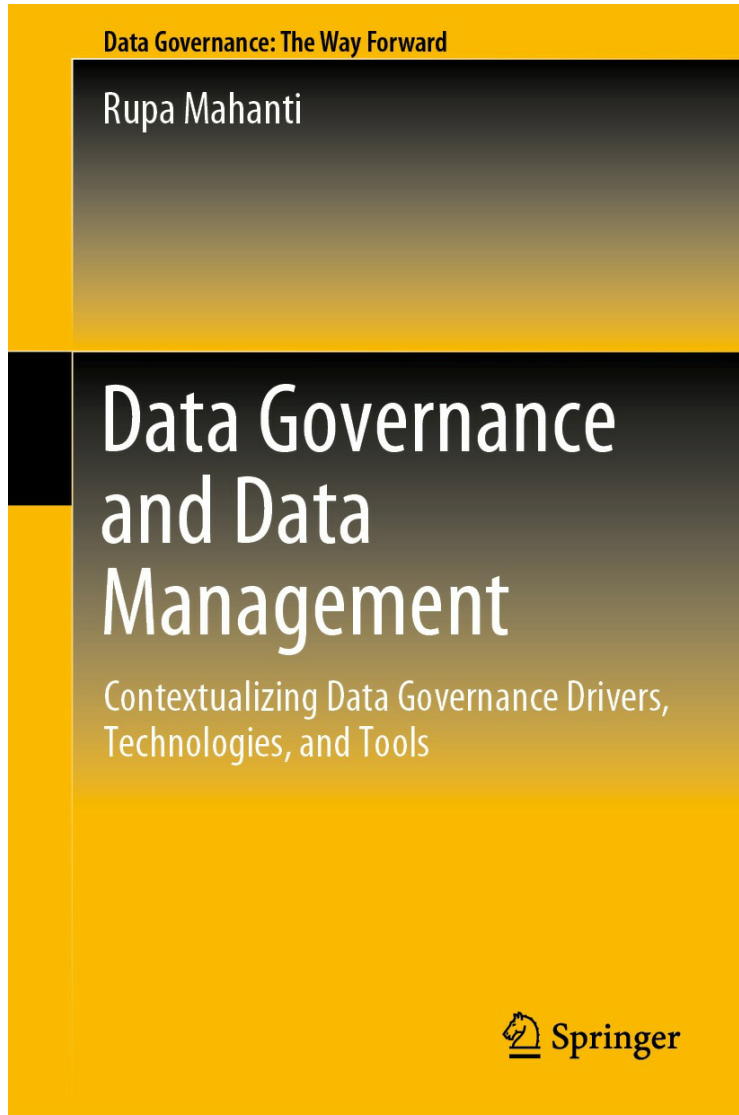




داده‌ها و حکمرانی آن

منابع

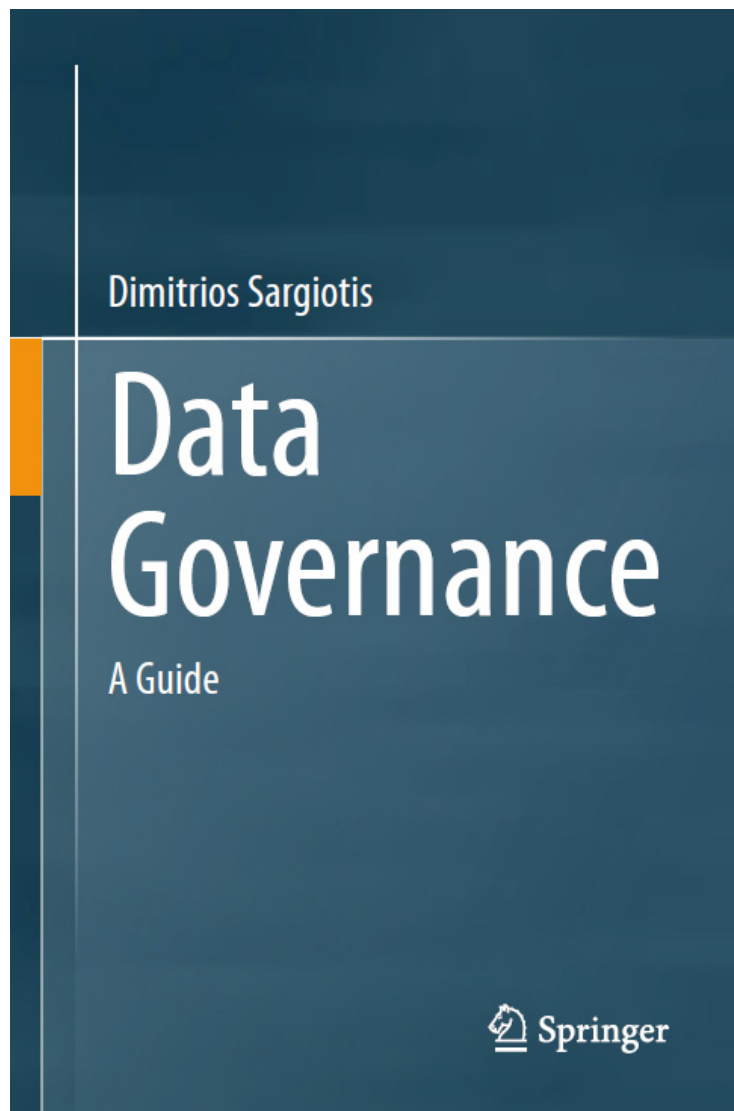
کتاب درس



Rupa Mahanti,
Data Governance and Data Management:
Contextualizing Data Governance Drivers,
Technologies, and Tools,
Springer, 2021.

<https://link.springer.com/book/10.1007/978-981-16-3583-0>

کتاب کمکی

 Springer