

The background of the slide is a dark blue field filled with a grid of small, glowing dots. Some of these dots are bright orange or yellow, while others are a lighter blue. The dots are arranged in a way that suggests a digital or data-driven environment, with some dots appearing more prominent than others. A large, white, diagonal shape cuts across the upper right portion of the slide, creating a clean space for the text.

KEARNEY

# 사용자 데이터 활용을 확장하는 데이터 거버넌스 전략

August 2023

## 01 데이터 활용과 데이터 거버넌스 실행의 한계

### 데이터 투자 및 결과

**2019년 기준 DX 투자 금액은 약 1,000조 이상  
투자의 결과로 생성되는 데이터의 양은 큰 폭으로 증가**

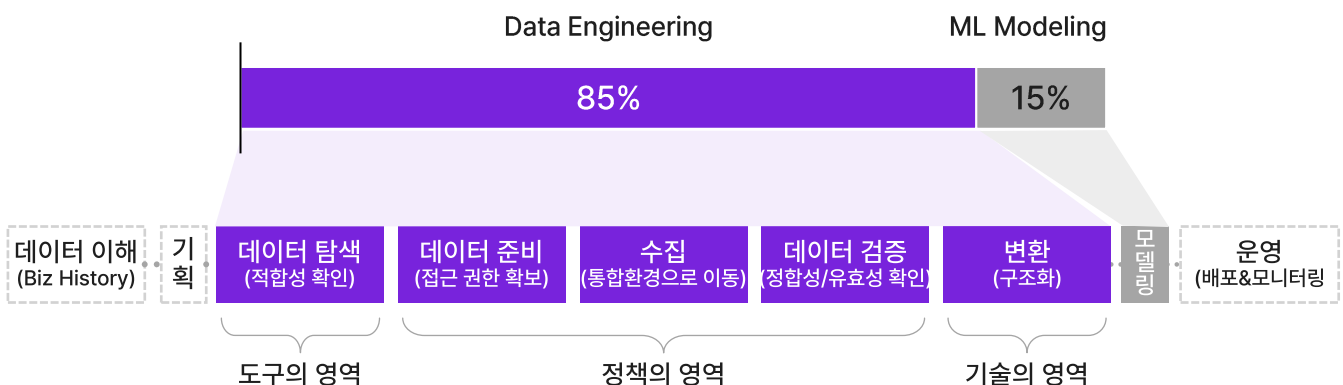
- 2017년 대비 투자 금액은 약 2배, 데이터 양은 무려 6배 증가
- “축적된 데이터로 성과와 혁신이 이뤄질 것”이라는 막연한 기대가 늘어나는 반면, 어떻게 활용해야 하는가에 대한 방법은 아직 찾지 못하고 있어 기업들의 고민 심화

### 데이터 혁신은 왜 어려울까?

**아무리 많은 데이터가 존재하더라도 데이터 엔지니어링이 원활하지  
않는다면, 데이터 기반의 혁신은 요원**

- 최고의 데이터 집단인 Google의 데이터 관련 업무를 세분화해보면, 데이터 엔지니어링은 축적된 데이터를 이해하고 기획한 후에 탐색-준비-수집-검증-변환의 과정을 거쳐, 활용 가능한 데이터로 만드는 업무가 전체의 85%를 차지
- 엔지니어링 되지 않은 데이터는 활용성이 지극히 낮음

### Google같은 Tech 기업의 Data Scientist에게도 Data Engineering은 어려운 과정



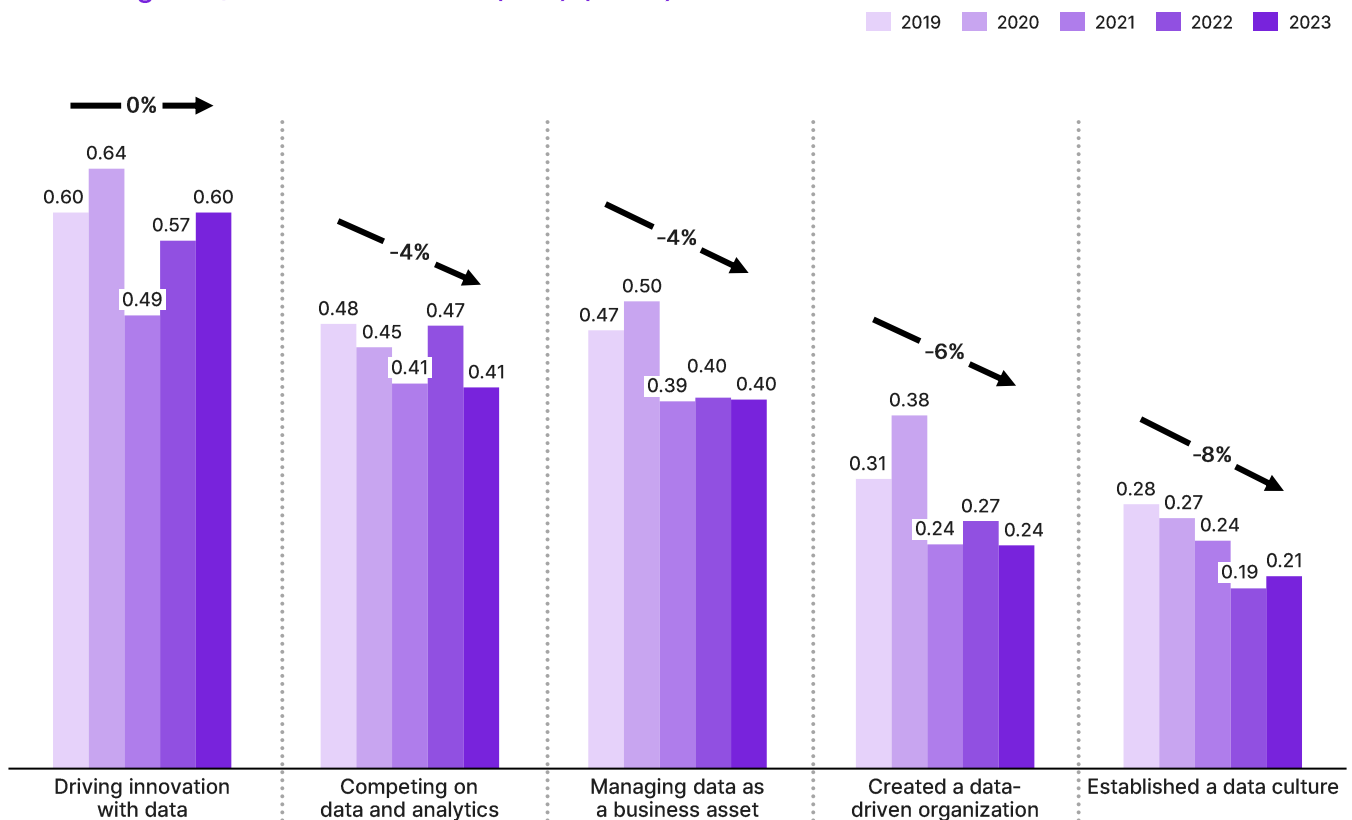
## 체계적인 데이터 혁신 방안은 선도 기업에게도 난제

**Fortune 1000대 기업 데이터 관련 설문 결과,  
과거보다 현재에 데이터 관련 체계적인 활용 및 운영 방안 마련에  
어려움을 겪고 있음**

- 데이터를 활용한 혁신, 데이터 분석을 활용한 경쟁, 데이터를 비즈니스 자산으로 관리하는 정도, 데이터 중심 조직 구축, 데이터 문화 정착 등 설문 문항 전반에서 수용도 지속 하락

### 데이터 혁신 자부하는 선도사들도 체계적 대응 방안 마련 고심

조직 내 Big Data / AI 수용도 설문 응답비율('23.1) (단위 %)



## 데이터 거버넌스의 정의

**데이터의 운영(수집, 처리, 저장, 교환)에 대한 효과적 프레임워크  
제공 및 데이터 사용(분석, 개발)에 대한 규칙과 가이드라인 개발을  
의미(Kearney)**

- 쉽게 설명하면, 데이터를 사용할 때 고민없이 이용할 수 있도록 기술, 절차, 서비스에 대한 구조를 제공하고, 정확한 판단 기준에 따라 관리하는 모든 행위를 뜻함

## 많은 기업들이 데이터 거버넌스를 실패하는 이유

### 사용자가 겪는 문제를 제대로 파악하지 못함

데이터 인프라, 구성원 역량, 사용자 수준을 고려하여 설계 및 실행해야 성공할 수 있음

### 특정 부서를 중심으로 기술과 솔루션에 의존

일반 사용자와 IT/Platform 담당자의 니즈와 역할에 대한 Gap을 인지해야 성공할 수 있음

### 너무 많은 문제를 한번에 해결하려고 함

다양한 조직 간의 이해관계의 조정, 프로세스 및 원칙이 수립되고 문화로 정착해야 성공할 수 있음

## 데이터 사용자들이 겪는 이슈

### 많은 투자와 도구를 도입했음에도 사용자들은 “쓸만한 데이터가 없다”고 토로

#### 데이터 존재 여부 확인 불가

다른 부서가 생성한 데이터가 어디에 어떤 형태로 있는지 찾을 수 없는 경우  
→ 데이터 거버넌스에서는 이를 '**데이터 카탈로그**'를 통해 해결

#### 데이터 의미(목적) 모호

동음이의, 이음동의 이슈 발생으로 데이터 이름만 봐서는 필요한 데이터인지 아닌지 파악할 수 없는 경우  
→ 데이터 거버넌스에서는 이를 '**데이터 메타**'를 통해 해결

#### 데이터 접근 권한 기준 모호

우여곡절 끝에 필요한 데이터 위치와 형태를 확인하였더라도 보안 상의 이유로 공유가 불가한 경우  
→ 데이터 거버넌스에서는 이를 '**데이터 권한관리**'를 통해 해결

#### 통합된 환경에서 제공 불가

DB가 아닌 Document, Excel, 인쇄물 등의 형태로 제공되어 별도의 작업이 필요한 경우  
→ 데이터 거버넌스에서는 이를 '**데이터 통합 관리**'를 통해 해결

#### 데이터 구조 상이

목적에 부합하는 형태로 존재하지 않으며, 가공이 용이한 형태로 제공되지 않는 경우  
→ 데이터 거버넌스에서는 이를 '**데이터 표준, 데이터 서비스**'를 통해 해결

#### 데이터 신뢰도가 낮음

Out-dated 또는 Sanitized 되거나, Logic(산식) 이슈로 인해 사용할 수 없는 경우  
→ 데이터 거버넌스에서는 이를 '**데이터 품질 관리**'를 통해 해결

시스템 및 데이터 투자가 많은 기업일수록 복잡성이 증가하여 더욱 어려움.

## 과거와 현재 데이터 거버넌스의 차이점

과거 ... n개의 **분산된** 데이터 거버넌스

현재 ... 1개의 **통합된** 데이터 거버넌스

- 과거의 데이터 거버넌스는 각 업무 기능의 생산성, 효율성을 위한 개별 시스템의 운영 최적화가 목적
- 반면, 현재의 데이터 거버넌스는 통합된 환경의 운영 최적화와 사용자들의 자유로운 데이터 활용을 목적으로 함. 따라서 다양한 시스템과 사용자의 이해관계를 고려한 역할 재정립 필요. 데이터 정책의 수준과 운영, 변화관리 기술 등에 대한 일련의 활동들을 IT 부서에 국한해서는 안됨. 조직 전체 관점에서 통합적으로 관리하는 것이 중요.

## 데이터 거버넌스의 필요성

**‘때’를 맞출 수 있고, ‘때’를 놓치지 않아야 하기 때문**

### 데이터 관련 업무 능력 향상(업무 활용 및 운영)

- 데이터 탐색, 권한확보, 통합, 변환(구조화 및 가공), 분석 등 활용 전단계의 시간을 단축하고 표준화된 정책에 기반하여 업무를 수행함으로써 효율성 증대
- 데이터 탐색/변환/통합의 시간 단축, Report 개발 시간 단축, 데이터 업무 포화도 분석, 데이터 및 보고서 재활용을 증가

### 데이터 투자, 운영 비용 최적화(인프라 투자 및 인건비)

- 데이터 전략, 정책, 운영 방향성 일원화 및 시스템 간 역할 재정립을 통해 중복된 데이터 및 시스템을 통폐합하여 관리 대상 감소에 따른 투자/운영비용 감소
- 신규 투자 비용(기회비용), 기존 시스템 운영 비용 감소

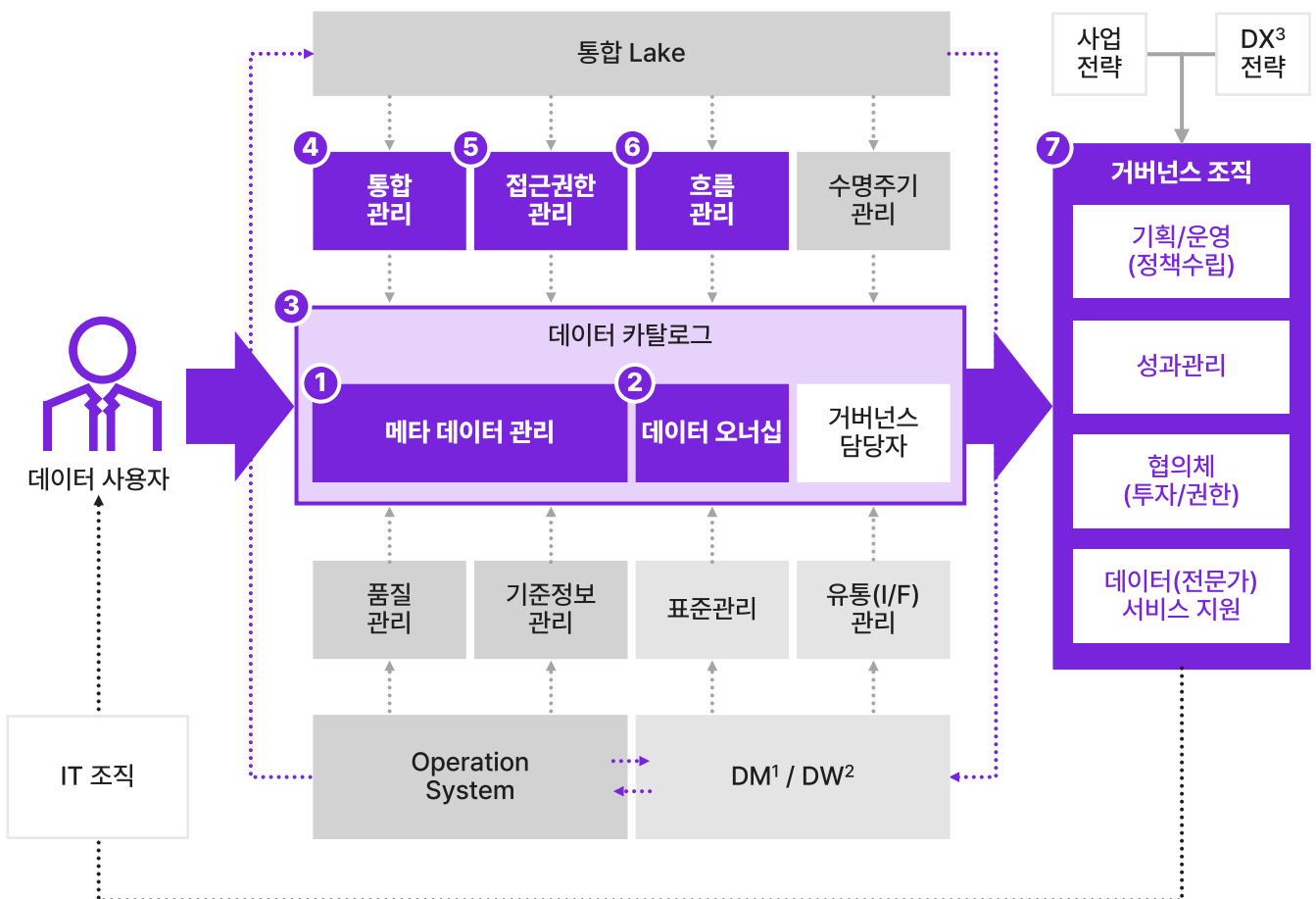
### 데이터 기반 가치 증대(각종 보고서 및 구성원 역량)

- 각종 보고서의 품질, 다양한 부서에서 참여한 보고서의 데이터를 일관되게 관리함으로써 이를 활용하는 구성원의 만족도와 활용역량 증대
- 데이터 신뢰도(일관성, 추적성), 데이터 활용도(사용자수, 시간), 유효 데이터 증가율 및 유효 사용자 증가율에 영향

## 02 데이터 거버넌스를 위한 핵심 요소

활용 관점의  
데이터 거버넌스의  
운영 메커니즘

데이터 거버넌스 운영에서 가장 크게 달라진 점은 데이터 카탈로그를  
통해 데이터가 유통 및 거버닝 된다는 점



## 핵심 성공 요소 7가지

1



### 메타 데이터 관리

데이터의 목적과 의미를 충분히 전달하기

2



### 데이터 오너십 부여

사용자 중심의 생태계를 구성하기

3



### 데이터 카탈로그

사용자들과 만나고, 소통하는 카탈로그 만들기

4



### 데이터 통합관리

레이크를 진흙탕으로 만들지 않기

5



### 접근권한 관리

데이터가 자유롭게 유통될 수 있는 방해요소를 제거하기

6



### 데이터 흐름 관리

생성과 파생을 잘 관리하여 의미 있는 데이터로 만들기

7



### 거버넌스 조직

통제가 아닌 중재와 활성을 위한 조직 설계하기



## 성공 요소 ① 메타 데이터 관리

### 데이터의 목적과 의미를 충분히 전달하기

#### IT Meta

- 시스템 운영을 위해 개발자를 위해 제공되는 정보
- 활용의 상세 설계를 위한 구조와 유효성 확인을 위해 사용
- 기존의 시스템 운영 데이터를 기반으로 자동화가 가능하나 지속 현행화가 필요한 영역

| 목적                 | 제공되는 정보        |            |
|--------------------|----------------|------------|
| 데이터의<br>원천과 경로를 확인 | 데이터 시스템명(물리명)  | 적재 위치(경로)  |
|                    | 소스 DB/System 명 | 소스 Table 명 |
|                    | 적재 방법          | 적재 작업 ID   |
| 데이터의<br>유효성을 확인    | 데이터 적재 주기      |            |
|                    | 데이터 생성 일시      |            |
|                    | 데이터 보관 기간      |            |

#### Business Meta

- 비즈니스 부서의 사용자 및 의사결정자를 위해 제공되는 정보
- 데이터 활용을 위한 적합성, 접근성 확인
- 자동화 하기 어려운 소유자 혹은 사용자들의 집단 지성이 발휘되어야 하는 영역

| 목적  | 제공되는 정보        |              |
|---|----------------|--------------|
| 데이터를<br>빠르게 탐색  | 데이터 명(논리명)     | 데이터 설명       |
|   | 데이터 카테고리(업무영역) | 해시태크         |
|   | 연관 보고서         | ...          |
| 데이터의<br>적합여부를 판단  | 데이터 유형         | 업무주기(발생주기)   |
|   | 주요 데이터 항목      | 주요 데이터 항목 단위 |
|   | 원천 시스템/화면 정보   | 데이터 보안 등급    |
| 담당자를<br>빠르게 확인  | 데이터 소유자        |              |
|   | IT 담당자         |              |
|   | 거버넌스 담당자       |              |
| ↓ ↓ ↓ ↓ ↓   |                |              |
| 비즈니스 메타는 대부분 정적인 데이터<br>동적인 데이터는 데이터 카탈로그에서 별도의 기능을 통해 제공 |                |              |
| 영향 분석      테이블간 비교 분석      워크 플로우      프로파일링              |                |              |



## 성공 요소 ② 데이터 오너십 부여

### 사용자 중심의 생태계를 구성하기

- 그동안 대부분 IT 시스템 담당자가 데이터 오너의 역할을 맡아 데이터를 관리/운영. 하지만, IT 시스템 담당자들이 데이터를 제대로 운영할 수 없는 가장 큰 이유는 데이터에 대한 권한이 없다는 점임. 예를 들어, 어떤 사용자가 데이터를 요청했을 때, IT 시스템 담당자들은 접근 권한을 판단할 수 없을 뿐더러 데이터 접근에 대한 최종 승인은 현업의 담당자들에게 받아야 함. 때문에 데이터 소유자가 IT 시스템 담당자로 지정되어 있는 것이 문제로 지적.

### 이상적 운영을 위해 데이터 소유자에게 부여되어야 할 역할

**데이터 메타:** 비즈니스 메타 작성 및 메타 정보 최신성 유지

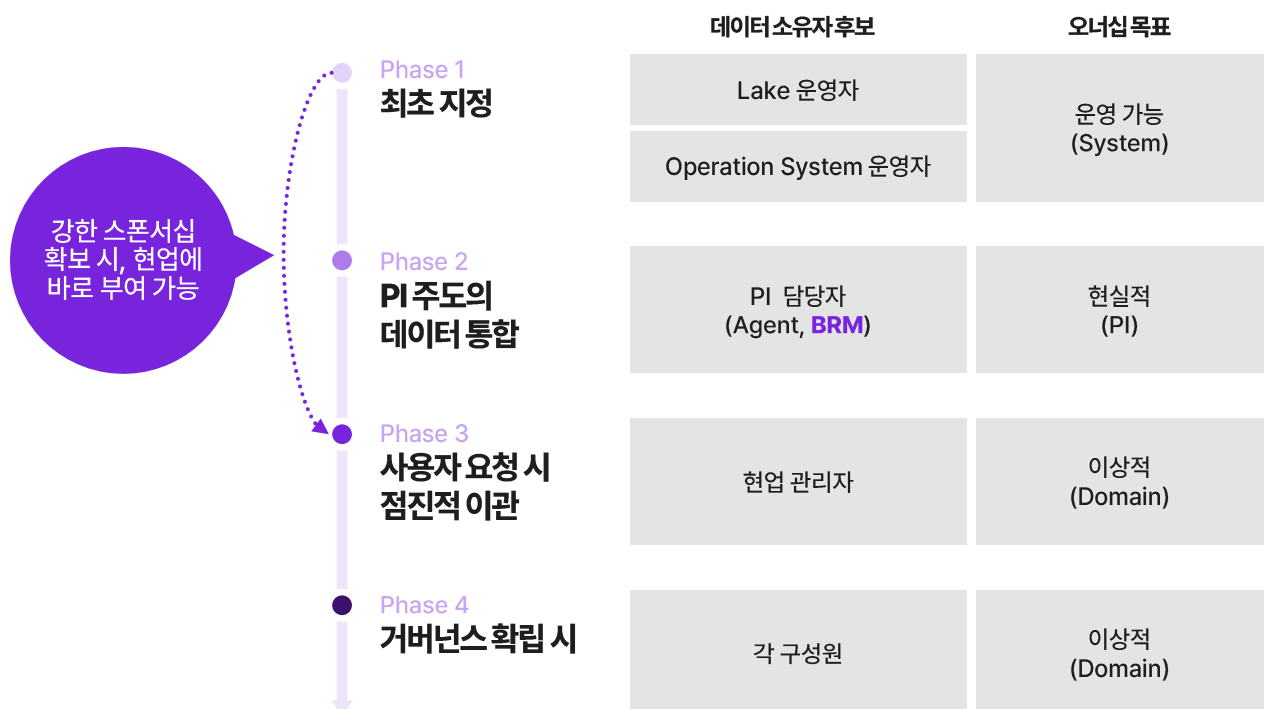
**데이터 권한:** 사용자로부터 요청 받은 권한에 대한 최종 승인

**데이터 통합:** 사용자로부터 요청 받은 통합(Lake로 적재)에 대한 실행

**데이터 품질/표준:** 유통되고 있는 데이터 상의 Logic, 형식의 적정성 판단

- 이러한 역할을 수행하기 위해서는 IT에 대한 기술적 지원 및 막중한 업무를 수행하는 데이터 소유자에 대한 동기부여도 함께 고민 필요.

### 데이터 오너십 부여를 위한 단계적 접근 방안



## 성공 요소 ③ 데이터 카탈로그

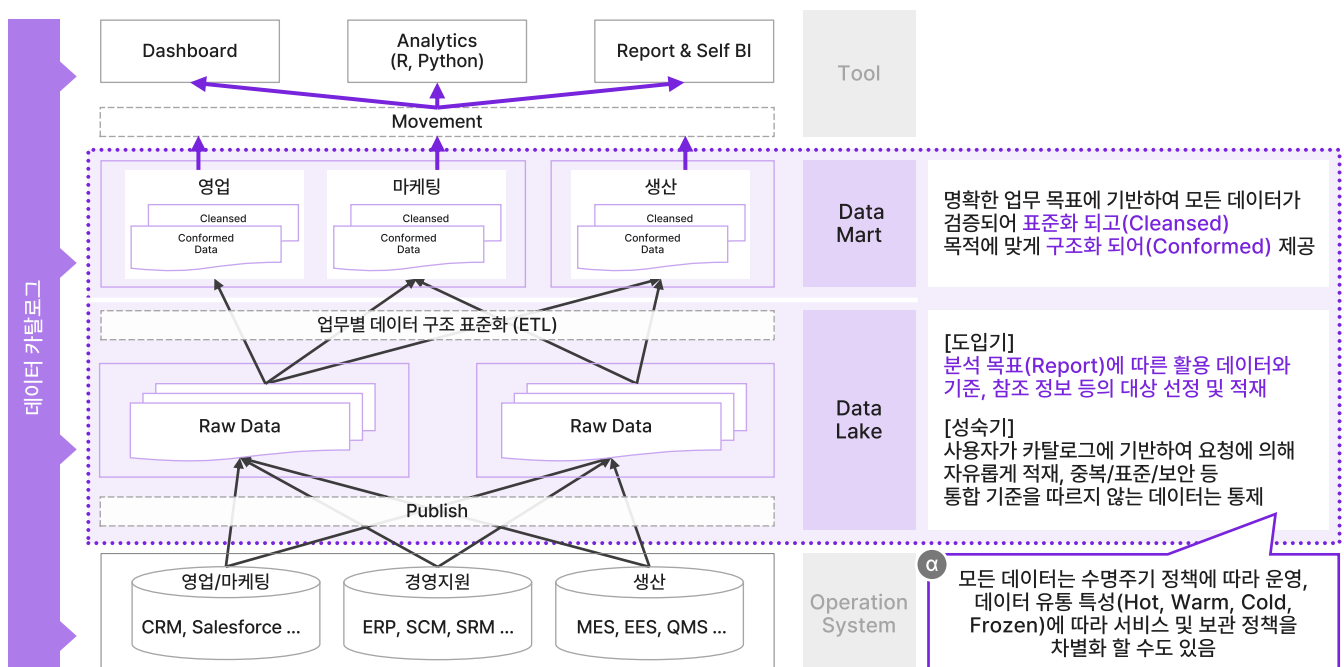
### 사용자들과 만나고, 소통하는 카탈로그 만들기

- 데이터 카탈로그는 업무 데이터에 대한 포털적 성격을 갖고 있으므로 데이터에 대한 정보 제공 뿐 아니라 거버넌스에 대한 전반적인 기능을 함께 제공함으로써 사용자들이 편리하게 데이터를 활용할 수 있는 환경을 조성해야 함.
- 레이크, 데이터 마트, 오퍼레이션 시스템에 있는 데이터들은 활발히 활용되고 유의미한 데이터들이 데이터 카탈로그에 모여야 함. 이렇게 모인 데이터들을 선별적으로 운영해야 하며, 최근에는 R이나 BI, 파이썬과 같은 다양한 툴을 사용하여 사용자들이 동일한 데이터 카탈로그에 접속해서 데이터의 의미를 확인할 수 있도록 지원해야 함.

## 성공 요소 ④ 데이터 통합 관리

### 레이크를 진흙탕으로 만들지 않기

- 오퍼레이션 시스템에 있는 데이터들이 Raw 데이터 형태로 데이터 마트에 구조화된 데이터로 제공되면, 무브먼트 서비스를 통해 대시보드, 애널리틱스 툴로 넘어가는 구조가 가장 이상적인 레이크 활용의 모습임.
- 현재 대부분 양방향으로 데이터가 모이는 이중 구조로 인해 레이크에 중복된 데이터들이 넘쳐나 데이터 탐색 자체가 어려움. 그러므로 레이크에 어떤 데이터를 통합할 것인지, 데이터 마트에는 어떤 데이터를 제공할 것인지를 명확하게 구분 및 정의하여 서비스할 필요.
- 사용자 관점에서 신선도에 따라 데이터 저장 및 서비스하는 정책 수립 필요. 기능적, 비용적, 사업적 요소를 고려하여 정책을 수립하고 거버넌스 관점에서 운영을 해야 실제 사용자가 필요한 데이터를 적시에 제공할 수 있음.



## 성공 요소 ⑤ 접근 권한 관리

### 데이터가 자유롭게 유통될 수 있는 방해요소를 제거하기

- 일반적으로 회사 내 데이터 중 극비 데이터는 1% 미만, public하게 공개해도 되는 데이터도 1% 미만. 즉, 사용하고 있는 99%의 데이터들은 대외비적인 성격을 띠. 이는 같은 회사 직원이라도 다른 부서의 데이터를 쉽게 볼 수 없는 것을 의미. 사업부, 해외법인, 지주사, 협력사, 타 부서 등 사용자의 포지션의 따라 연계하여 데이터 정책이 수립되어야 하는데 실제로는 그렇지 못하고, 데이터 사일로가 발생하는 이유도 여기에 있음. 이와 같은 현상이 벌어지는 원인은 행여나 데이터 관련 문제가 발생했을 때, 책임이 누구에게 있는지 명확하지 않기 때문임.
- 이같은 문제는 데이터 보안 등급을 설정하고 관리함으로써 해결 가능. 먼저, 현업에서 데이터 거버넌스에 대한 등급을 설정 후, 거버넌스 권한 협의체에서 분류 및 검토하여 확정. 일련의 명확한 과정을 통해 검토가 되었기 때문에 만약 유출 문제가 발생하더라도 데이터 제공자가 아닌 유출자에게 그 책임이 집중되도록 제도적 보완이 필요.

## 성공 요소 ⑥ 데이터 흐름관리

### 생성과 파생을 잘 관리하여 의미 있는 데이터로 만들기

#### 생성

- 데이터가 만들어지기까지의 업무상 과정과 사용했던 정보 관리
- 운영자가 이슈에 대해 빠르게 원인을 분석하고, 추가적으로 운영 지원을 할 수 있어야 하기 때문에 생성 데이터 흐름을 관리(거버넌스 관리자, 시스템 담당자)

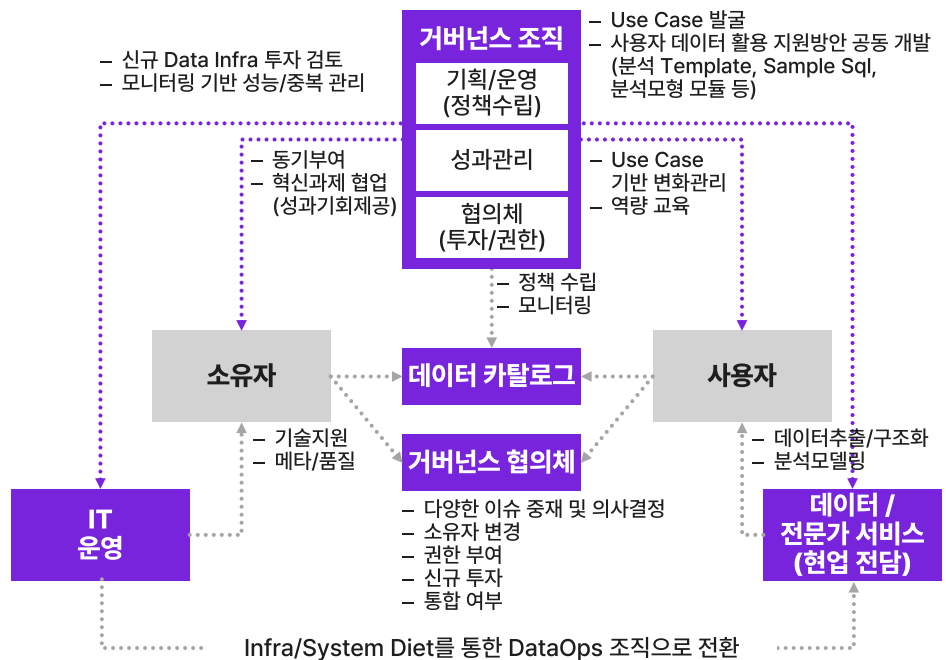
#### 파생

- 데이터가 만들어지고 나서 업무 목적에 따라 다른 구조와 타임 스탬프를 갖고 만들어진 데이터
- 사용자들이 즉시 활용 가능한 데이터를 빠르게 식별하기 위한 것 (일반 사용자, 고급분석가)

## 성공 요소 ⑦ 거버넌스 조직

### 통제가 아닌 중재와 활성을 위한 조직 설계하기

- 데이터 거버넌스는 통제나 정책 수립보다는 실제 사용자들이 데이터를 보다 잘 사용할 수 있게 하는 데이터 추출 구조와 분석 모델링을 어떻게 서비스할 것인가, 다양한 이슈 발생 시 어떻게 중재하고 의사결정 할 것인가, 더 높은 수준으로 활성화하기 위해 변화관리는 어떻게 할 것인가 등에 초점을 맞추고 운영할 필요.
- 거시적인 목표는 IT 시스템 담당자들이 현업의 데이터를 잘 활용하여 **데이터 옴스** 조직으로 전환하는 것.



## 데이터 거버넌스 관점에서 남은 업무에 대한 실행 방안

- ① 품질 관리**  
품질 대상/범위를 명확화하고 품질 이상의 기준과 대응 방안 마련
- ② 기준 정보 관리**  
기준정보(전사/운영 기준, 참조)를 분류, 표준화 대상 선별 및 개선
- ③ 표준 관리**  
대표성을 띄는 표준정책을 기준으로 레이크부터 정의(전 시스템 대상은 무의미)
- ④ 유통(I/F) 관리**  
I/F의 관리 주체를 일원화(중복 적재, 인프라 관리)
- ⑤ 수명 주기 관리**  
데이터를 유형화, Lake내 보관정책~서비스 정책 수립
- ⑥ 기획/운영(정책 수립)**  
각 메커니즘 영역별 프로세스 R&R 정의, 변화관리
- ⑦ 성과관리**  
사업 및 DX 전략 기반 거버넌스 목표 설정, 성과/운영 지표 개발 및 모니터링

## 03 데이터 거버넌스 추진 방안

### 추진 방안 ①

#### 데이터와 디지털에 대한 성숙도에 따라 다른 전략으로 접근

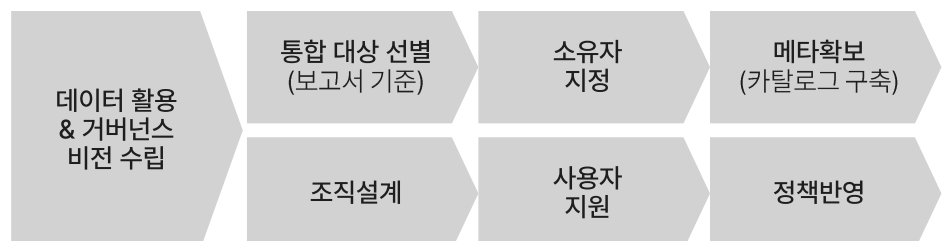
##### 데이터/디지털 성숙도 높은 기업 → “Use Case 기반 Scale-up”

- 업무영역 별로 많은 시스템이 구축된 기업은 목적에 의해 유사한 데이터가 Terminology 차이를 두고 무수히 유통되고 있음
- 다양한 이해관계가 고착화되어 변화시키기 매우 어려움
- Use Case를 통해 데이터 인지(메타), 공유(권한) 및 데이터 중복 생성의 이슈를 드러내어 Top Level 공감대 확보 후 점진적 확장 필요



##### 데이터/디지털 성숙도 낮은 기업 → “Top Down으로 중요 데이터부터”

- 데이터를 어떻게 활용해야 하는지, 데이터와 분석에 대한 기대수준과 데이터 기반 업무 커버리지가 낮아 수작업을 당연하게 생각
- 일하는 방식의 변화를 도모하기 위해 TOP Level의 스폰서십을 확보하고 비전 수립
- 중요 보고서를 기준으로 활용되는 데이터를 선별하고 통합된 환경에 적재, 선별된 데이터는 소유자를 지정하여 메타확보 및 소유자 중심의 운영체제 가동



## 추진 방안 ②

### 막연한 조직의 역할을 기대한 정책 및 프로세스 수립은 위험

- IT 및 Lake 운영자는 플랫폼을 제공해주는 Lessor의 역할, 사용자는 데이터를 소비하는 바이어 역할, 데이터 소유자는 셀러 역할을 수행한다는 관점에서 접근해야 함.

|           | IT/Lake 운영자<br>(Lessor)  | 사용자<br>(Buyer)  | 데이터 소유자<br>(Seller)  |
|-----------|--|---|--|
| 주요<br>역할  | <ul style="list-style-type: none"> <li>- Platform 운영</li> <li>- 데이터 공급자동화 기술 제공</li> <li>- 카탈로그 시스템 운영</li> <li>- Lake 모니터링</li> </ul> | <ul style="list-style-type: none"> <li>- 데이터 탐색 및 소비</li> <li>- 데이터 구조화 및 분석</li> </ul> | <ul style="list-style-type: none"> <li>- 데이터 공급(to Lake)</li> <li>- 데이터 메타관리 (입력/수정)</li> <li>- 품질 Trouble Shooting</li> <li>- 데이터 권한승인</li> <li>- 데이터 표준준수</li> </ul> |
| 현실적<br>제약 | <ul style="list-style-type: none"> <li>- 자동화된 기술 개발이 쉽지 않고 표준화하기 어려움</li> <li>- 사용자, 소유자 변화관리가 어려움</li> </ul>                          | <ul style="list-style-type: none"> <li>- 데이터 구조화 역량 미흡 (연결, 가공 Mapping, 변환)</li> </ul>  | <ul style="list-style-type: none"> <li>- 소유자의 데이터, IT 역량 수준 낮으며, 소유자간 편차 큼</li> <li>- 공급, 품질이슈 대응을 위해서는 기존 시스템 운영과 협업 필요</li> </ul>                                    |
| 지원<br>도구  | <ul style="list-style-type: none"> <li>- 데이터 거버넌스 솔루션</li> </ul>   | <ul style="list-style-type: none"> <li>- 데이터 추출, 구조화를 위한 예시 혹은 서비스 제공</li> </ul>        | <ul style="list-style-type: none"> <li>- IT 전담 지원 조직 (운영 및 서비스 : 중앙 → 현업)</li> </ul>   |

## 추진 방안 ③

### 다양한 사용자의 활용 유형을 고려

- 부서 별로 주요 분석 니즈가 다르고 필요한 데이터 역량의 수준도 차이가 있음.

- 부서 별 주요 분석 니즈

**영업/마케팅·구매/물류 → 실질적 데이터 중심, 정형 리포트/대시보드 분석 필요**

일반 사용자(데이터 구독자) 중심 가공/구조화 데이터 필요

**제조·개발 → 제조·개발 Raw 데이터 활용한 Analytics 및 Simulation 분석**

데이터 분석자 및 데이터 과학자 중심 MES/YMS 등 운영 시스템 Raw 데이터 실시간 분석 환경

**SCM → 다수데이터 간 연계분석**

데이터 분석자 중심으로 다른 부서에서 생생한 데이터들이 모이는 교차점  
데이터 결합/융합하여 표준화하는 작업이 매우 중요함

## 04 데이터 거버넌스 성공을 위한 4가지 제안

1

### 데이터의 소유자로, 소유자들의 관리자로 고위 경영진이 참여하고 독려

데이터는 업무 목적에 따라서 의지와 목표, 비즈니스적 논리가 가미되어 있는 하나의 업무 결과물, 이 결과를 책임지고 있는 고위 경영진의 관리 및 독려 필요.

2

### 정해진 프레임워크에 따라 사용자와 소유자가 자유롭게 데이터를 유통

데이터가 통합될 수는 있지만, 활용하기 어려운 경우가 더 많으므로 자유롭게 데이터가 유통되는 것을 제1방향성으로 여겨야 함.

3

### 통제가 아닌 서비스를 중심으로 하는 데이터 거버넌스 정책 및 제도 운영

자유로운 데이터 유통이 가능 하려면 통제를 목적으로 데이터 거버넌스를 운영하는 것이 아니라 사용자 중심으로 데이터를 서비스한다는 관점으로 정책을 수립하고 제도를 운영할 필요가 있음.

4

### 감시가 아닌 소유자와 사용자를 지원하는 데이터 거버넌스 기술

데이터 유출에 대한 '감시'보다는 데이터 소유자도 사용자도 편리하게 데이터를 이용하고 지원할 수 있는 데이터 거버넌스 기술이 덧붙여졌을 때 성공적으로 운영될 수 있음.





The background is a dark blue field filled with glowing orange and yellow binary digits (0s and 1s) arranged in a grid-like pattern that recedes into the distance, creating a sense of depth. A large, thick white 'Y' shape is superimposed on the image, originating from the top left and branching out towards the bottom right, dividing the composition into several geometric sections.

# KEARNEY

Copyright©2023 A.T. Kearney Korea LLC. All rights reserved.