VIRTUAL DATA LAKES & FEDERATED LEARNING FOR LIFE SCIENCES

October 21, 2022

Accenture Life Sciences

INDEX

1. **Data and Artificial Intelligence transforming Life Sciences - and the corresponding challenges**

2. **How Virtual Data Lakes and Federated Learning can disrupt the data and AI transformation of Life Sciences**

3. **Selected Case Studies using Virtual Data Lakes and Federated Learning**
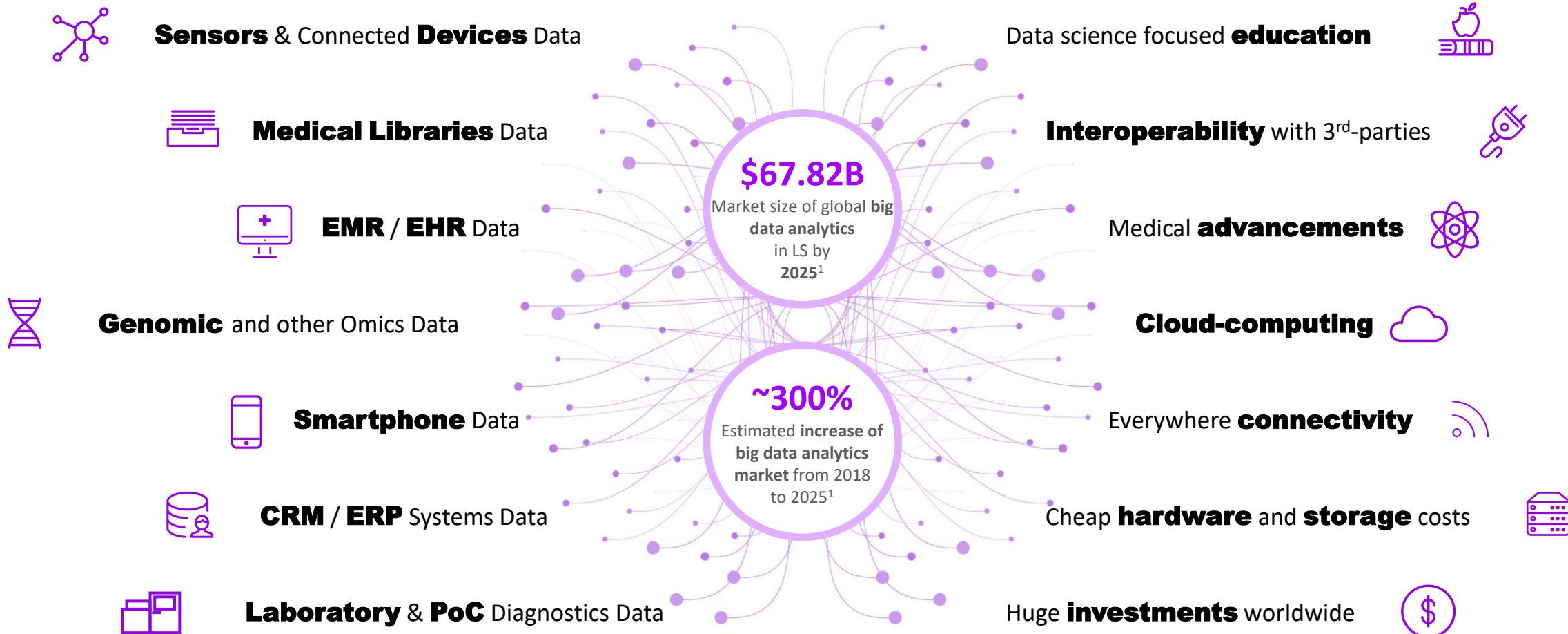
01

Data and Artificial Intelligence transforming Life Sciences - and the corresponding challenges

The Life Sciences (LS) industry is experiencing sustained exponential growth around data and analytics...

DATA SOURCES

Sensors & Connected **Devices** Data

Medical Libraries Data

EMR / EHR Data

Genomic and other Omics Data

Smartphone Data

CRM / ERP Systems Data

Laboratory & **PoC** Diagnostics Data

ENABLERS

Data science focused **education**

Interoperability with 3rd-parties

Medical **advancements**

Cloud-computing

Everywhere **connectivity**

Cheap **hardware** and **storage** costs

Huge **investments** worldwide

$67.82B
Market size of global **big data analytics** in LS by 2025[1]

~300%
Estimated **increase of big data analytics market** from 2018 to 2025[1]

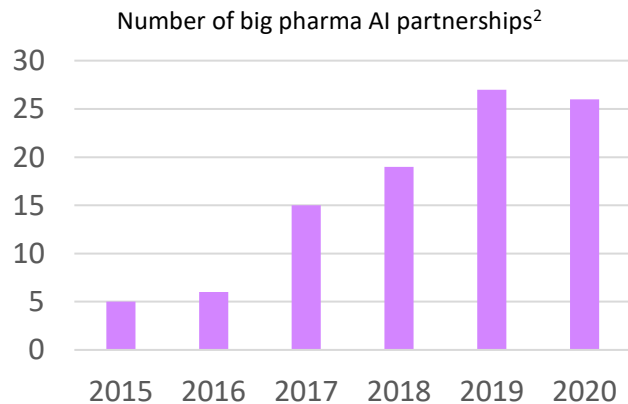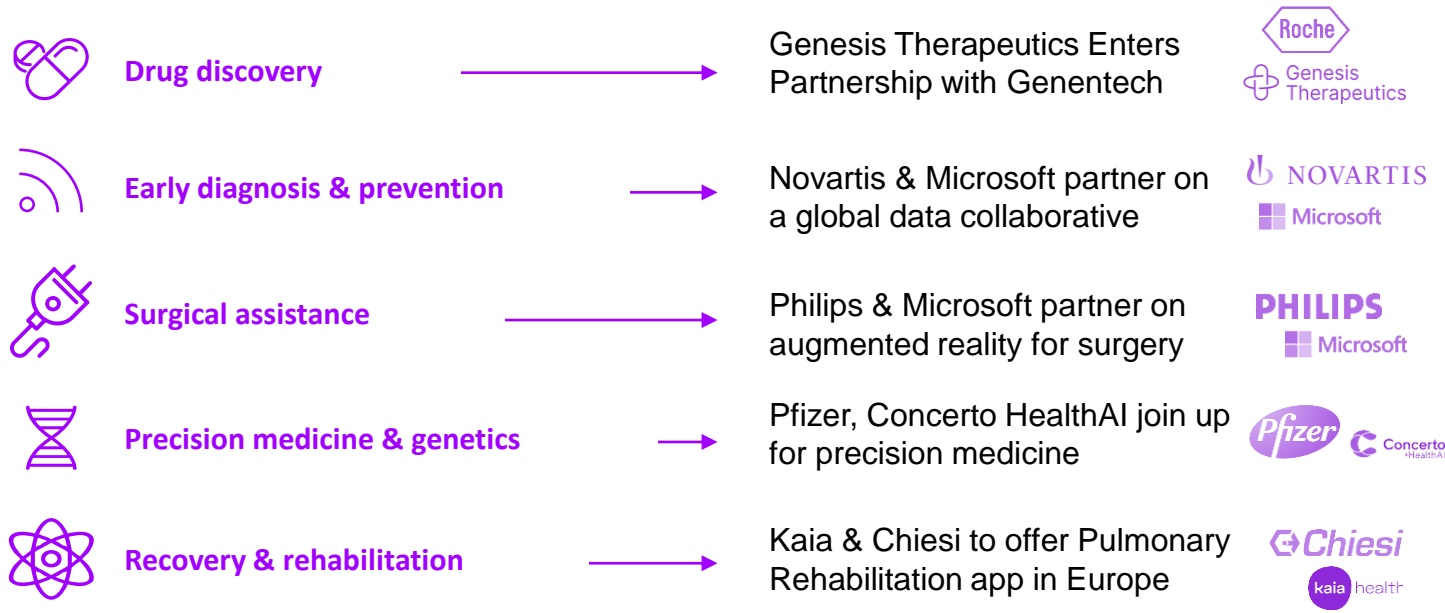[1]alliedmarketresearch

Accenture Life Sciences

...evidenced by multiple partnerships leveraging AI cloud capabilities for some of the most impactful use cases

~86% of life science companies are currently using artificial intelligence technology[1]

The number of big pharma AI partnerships has grown significantly...

Number of big pharma AI partnerships[2]

Year	
2015	5
2016	6
2017	15
2018	19
2019	27
2020	26

...enabling some of the most impactful Artificial Intelligence applications

Drug discovery → Genesis Therapeutics Enters Partnership with Genentech — Roche / Genesis Therapeutics

Early diagnosis & prevention → Novartis & Microsoft partner on a global data collaborative — NOVARTIS / Microsoft

Surgical assistance → Philips & Microsoft partner on augmented reality for surgery — PHILIPS / Microsoft

Precision medicine & genetics → Pfizer, Concerto HealthAI join up for precision medicine — Pfizer / Concerto HealthAI

Recovery & rehabilitation → Kaia & Chiesi to offer Pulmonary Rehabilitation app in Europe — Chiesi / kaia health

[1]life sciencesitnews
[2]GlobalData Pharmaceutical Intelligence Center, December 2022

However, implementation of big centralized cloud platforms comes with significant challenges (1/2)

FINANCIAL

- **High volume, sustained investments** are required for setting up and maintaining data platforms with IaaS vendors

- **Skills and tools required** to move/connect applications to such new platforms **generate significant additional costs**

- **Lack of global interoperability standards** as well as increasingly strict **security requirements further increase costs**

LEGAL & REGULATORY

- Severe limitations due to the **sensitivity of life sciences data** and the **reluctance or inability to share** them

- **Legislative restrictions** limiting **use cases**

- Specific **skills required** to implement **GxP compliant processes** and consent management

TECHNICAL

- Life Sciences **data and platforms** are **highly distributed and fragmented**

- Data **quality and standards** is **heterogeneous** across different sources

- Increasing data volumes and sources result in **data inconsistency** and **performance issues**

BUSINESS

- It is difficult **to demonstrate business of LS value quickly** as the consolidated platform approach is time consuming

- **Business needs are changing rapidly**, and consolidated platforms are not flexible or agile by design
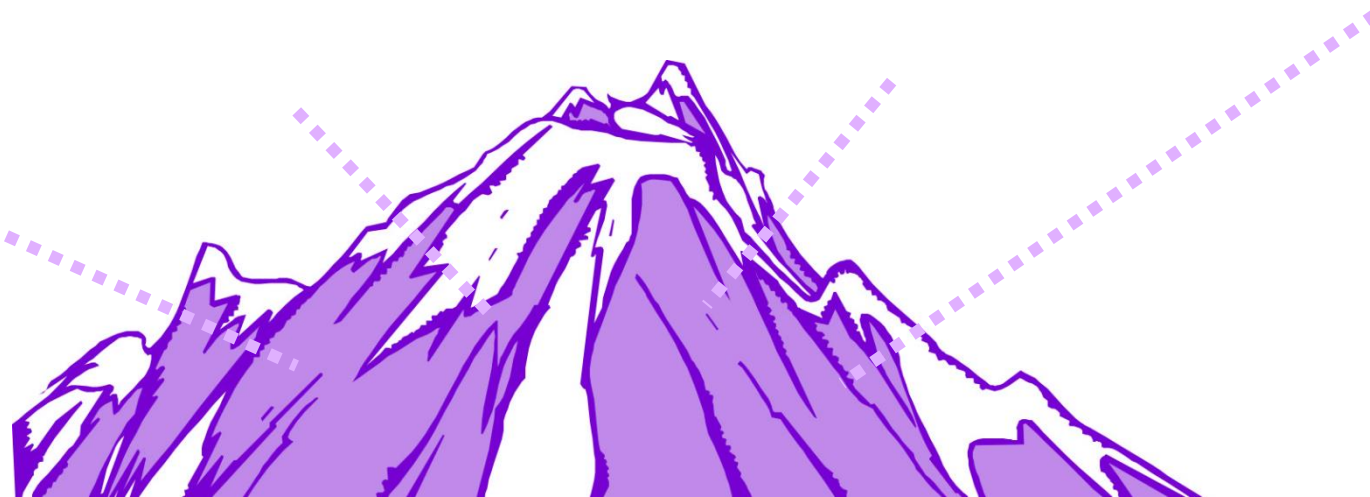
However, implementation of big centralized platforms comes with significant challenges (2/2)

GOVERNANCE

- Data governance needs **acceptance by means of a working communication** between all parties

- Consistency in terms of **data standards and quality mechanisms** is complex and difficult to implement

- **Data governance becomes a political issue**, as this ultimately means distributing, awarding and withdrawing responsibilities

MEDICAL & SCIENTIFIC KNOW-HOW

- Majority of LS organizations **lack in-house teams with track record in** harnessing big data and analytics tools effectively

- Ability to extract **medical value** by leveraging data beyond a narrow treatment/ intervention/ diagnosis focus is not yet mature at most LS clients

CORPORATE & CHANGE MANAGEMENT

- Data and platforms are being seen as a powerful lever, with **business units or entire organizations being reluctant to share them** with other business units or partners

- Data providers mostly anticipate **misinterpretation** or **misuse** of data

- **Guidelines on data sharing** are unclear, inconsistent or inexistent
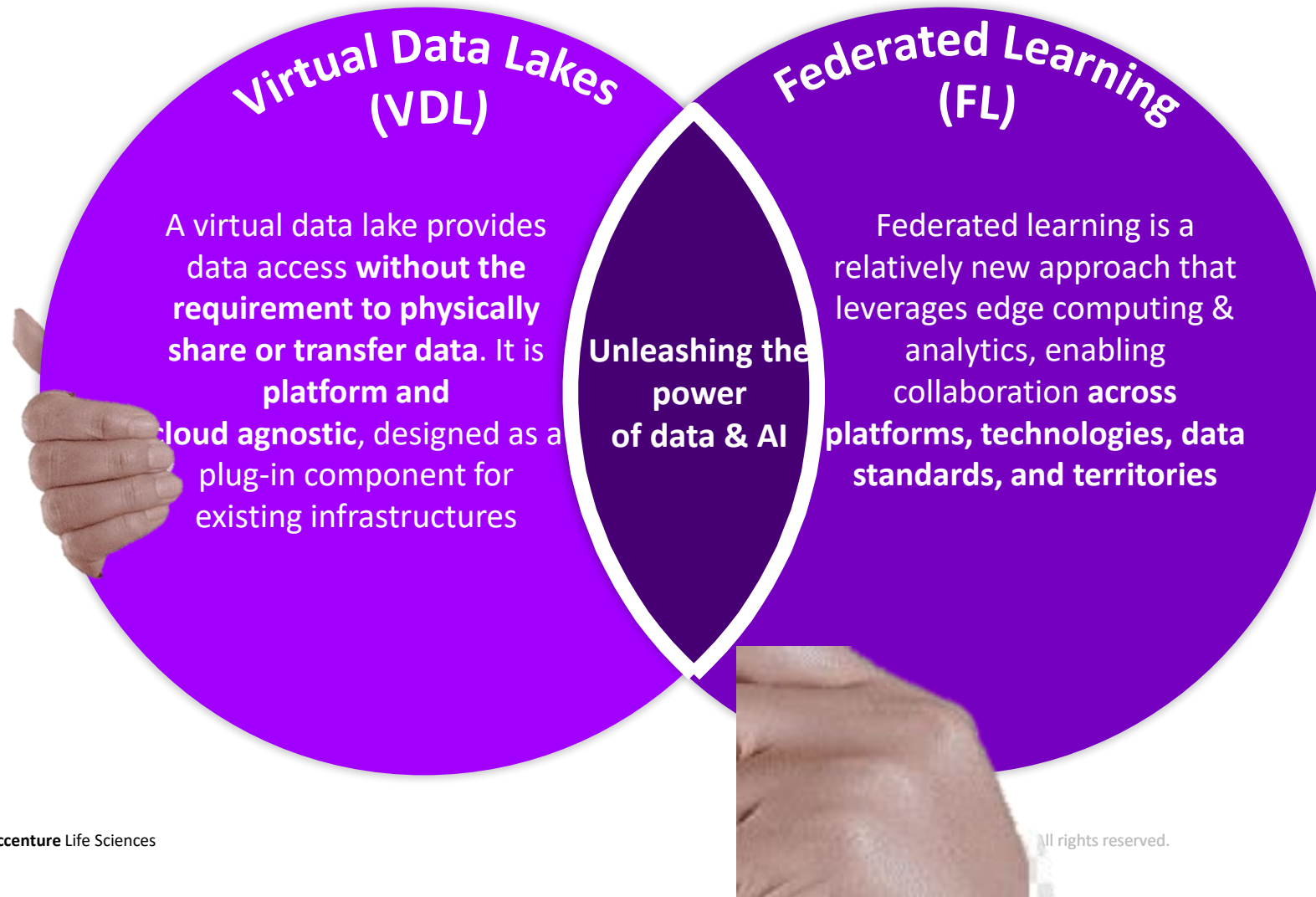
ETHICS

- **Patients** often have no insight into what their data is used for and as such are **reluctant to share them**

- **Unbalanced risks and benefits** derive from the amount and type of data requested compared to the potential impact of its use

- **Unfair data sharing practices** exist and benefit more the data producers than users
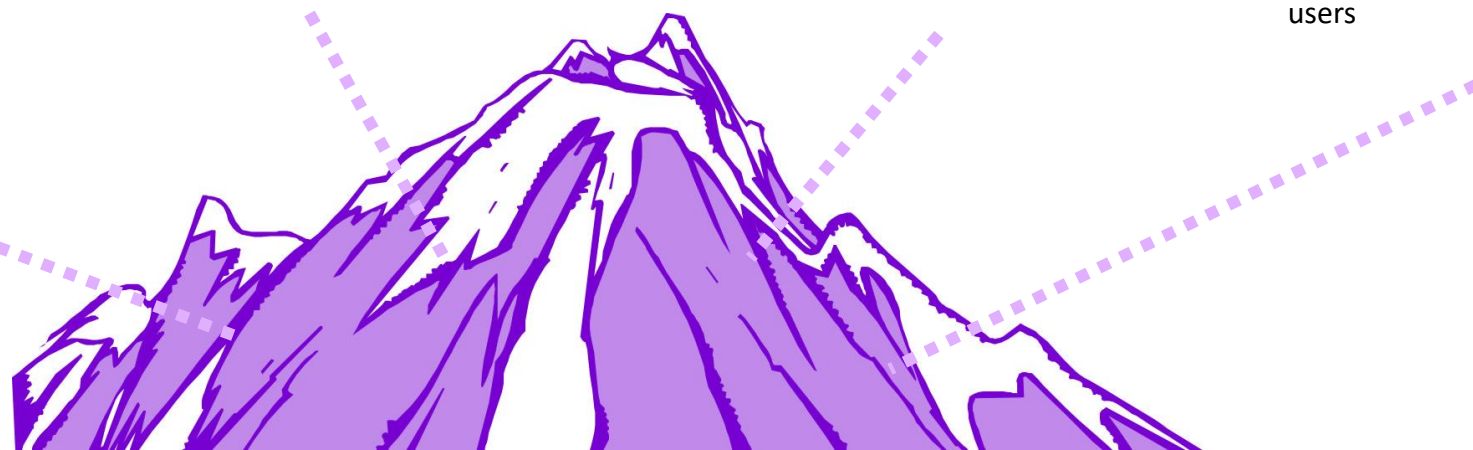
02

How Virtual Data Lakes and Federated Learning can disrupt the data and AI transformation of Life Sciences

Two emerging technologies can help address data and AI related challenges in Life Sciences

Virtual Data Lakes (VDL)

A virtual data lake provides data access **without the requirement to physically share or transfer data**. It is **platform and cloud agnostic**, designed as a plug-in component for existing infrastructures

Unleashing the power of data & AI

Federated Learning (FL)

Federated learning is a relatively new approach that leverages edge computing & analytics, enabling collaboration **across platforms, technologies, data standards, and territories**
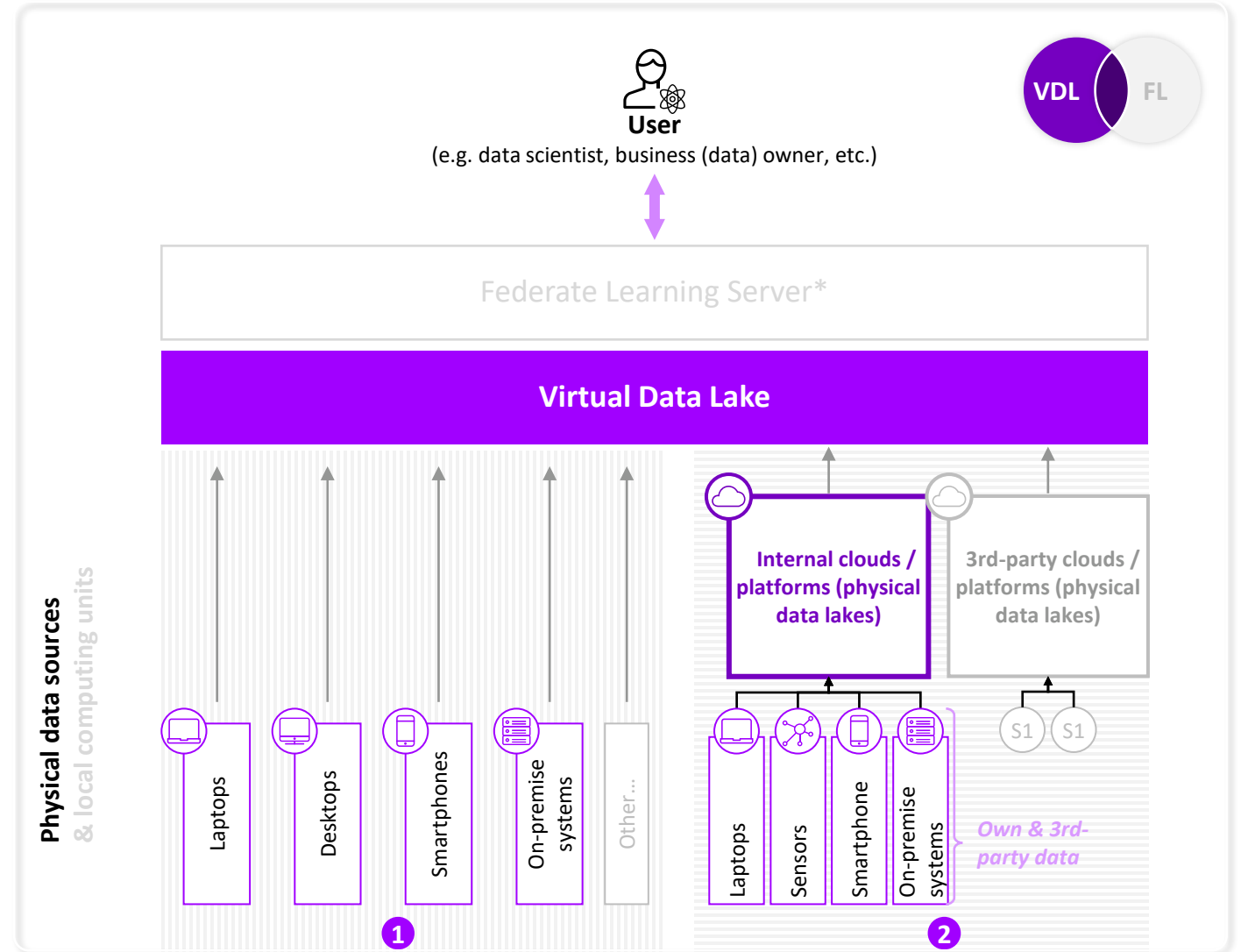
The combination of virtual data lakes and federated learning allow **in-situ access and analysis of data**. Such approach possesses **multiple advantages** including avoiding the overloading of networks with massive data, overcoming the concern of sharing private or sensitive/ proprietary information, etc.

Virtual Data Lakes
High-level overview

A virtual data lake provides **access to disparate data that is physically stored in distributed data sources**/ locations, ideally following a hybrid approach:

1 **Access local data sources**
Access to data that resides in local data internal or 3rd-party data sources such as e.g. on-premise ERP systems, smartphones, etc.

2 **Access centralized cloud data**
Access to data that resides in a centralized internal or 3rd-party cloud which again is consolidating data from various sources into a physical data lake such as e.g. medical sensors, etc.

*See following slide

User
(e.g. data scientist, business (data) owner, etc.)

Federate Learning Server*

Virtual Data Lake

Physical data sources & local computing units

Laptops | Desktops | Smartphones | On-premise systems | Other…

Internal clouds / platforms (physical data lakes)

3rd-party clouds / platforms (physical data lakes)

Laptops | Sensors | Smartphone | On-premise systems

Own & 3rd-party data

VDL   FL

Accenture Life Sciences

Federated Learning
High-level overview

Federated learning is a technological approach to **train a machine learning (ML) model**. Hereby, the **virtual data lake provides the underlying infrastructure** to support this process. At the same time, **data is remaining at its source of origin** (i.e. the physical data source)

Approach

1 Initiation
A ML model is built on a central server for a given research question

4 Sharing of local model
The locally trained model is shared back with the server

2 Transfer
The ML model is transferred to local computing units

5 Aggregation
Local models are aggregated to a model on the central server

3 Local model training
The ML model is trained with local data

6 *Iteration*
The aggregated universal model is again shared back to the local units, repeatedly, until the predictive power of the ML model converges to its optimal performance
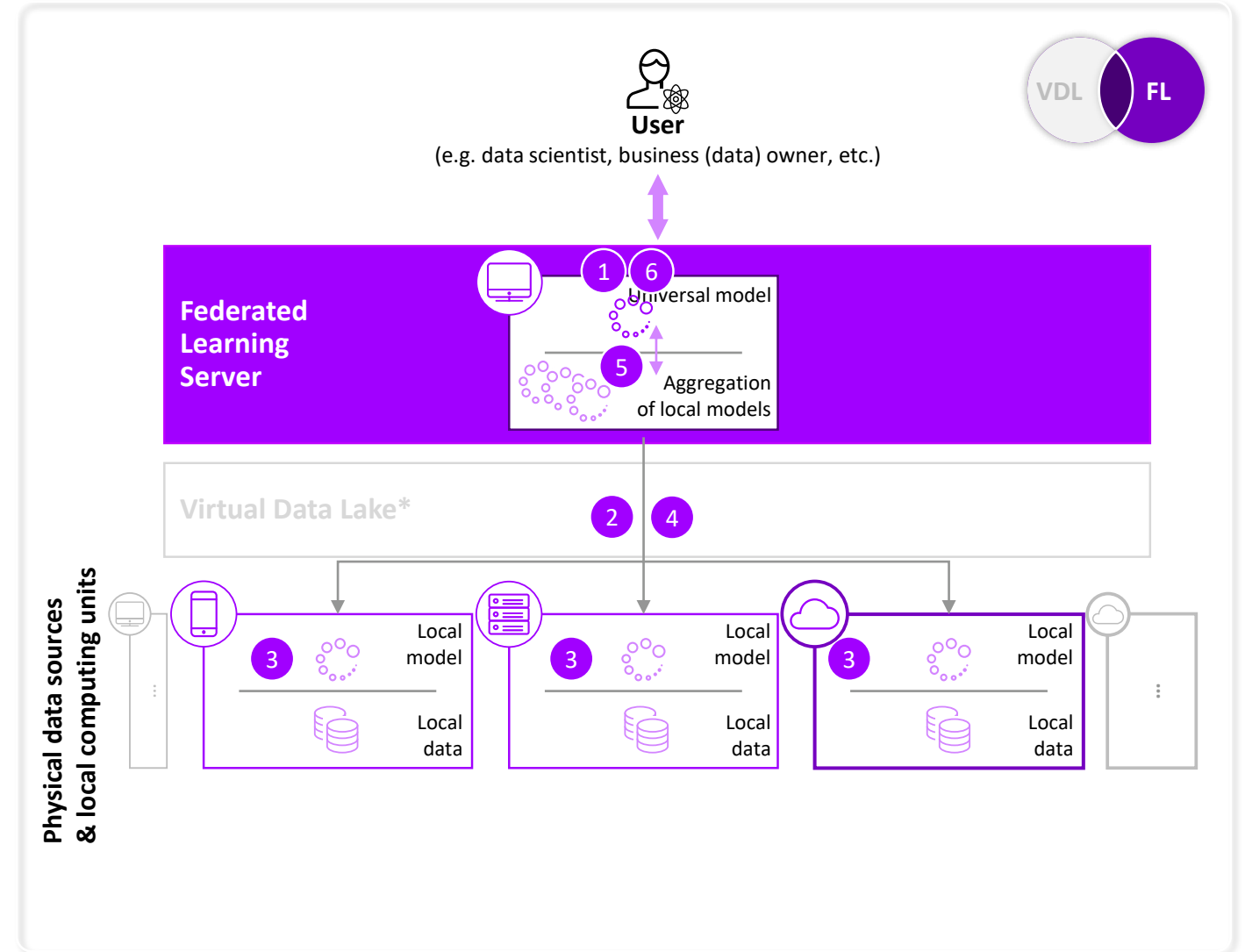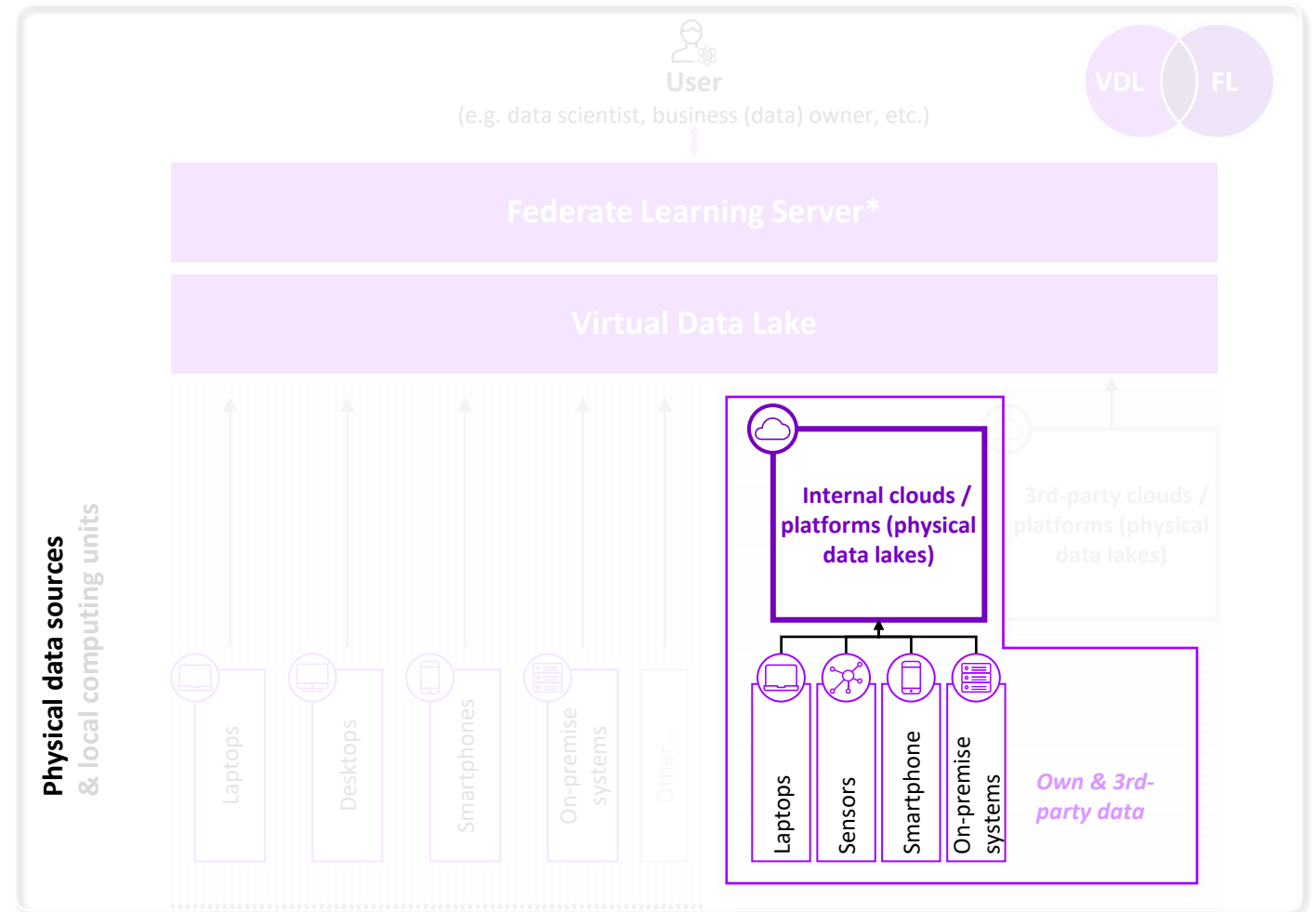
*See previous slide

Over the last few years, Life Sciences companies have been trying to break down data silos by investing in centralized cloud platforms

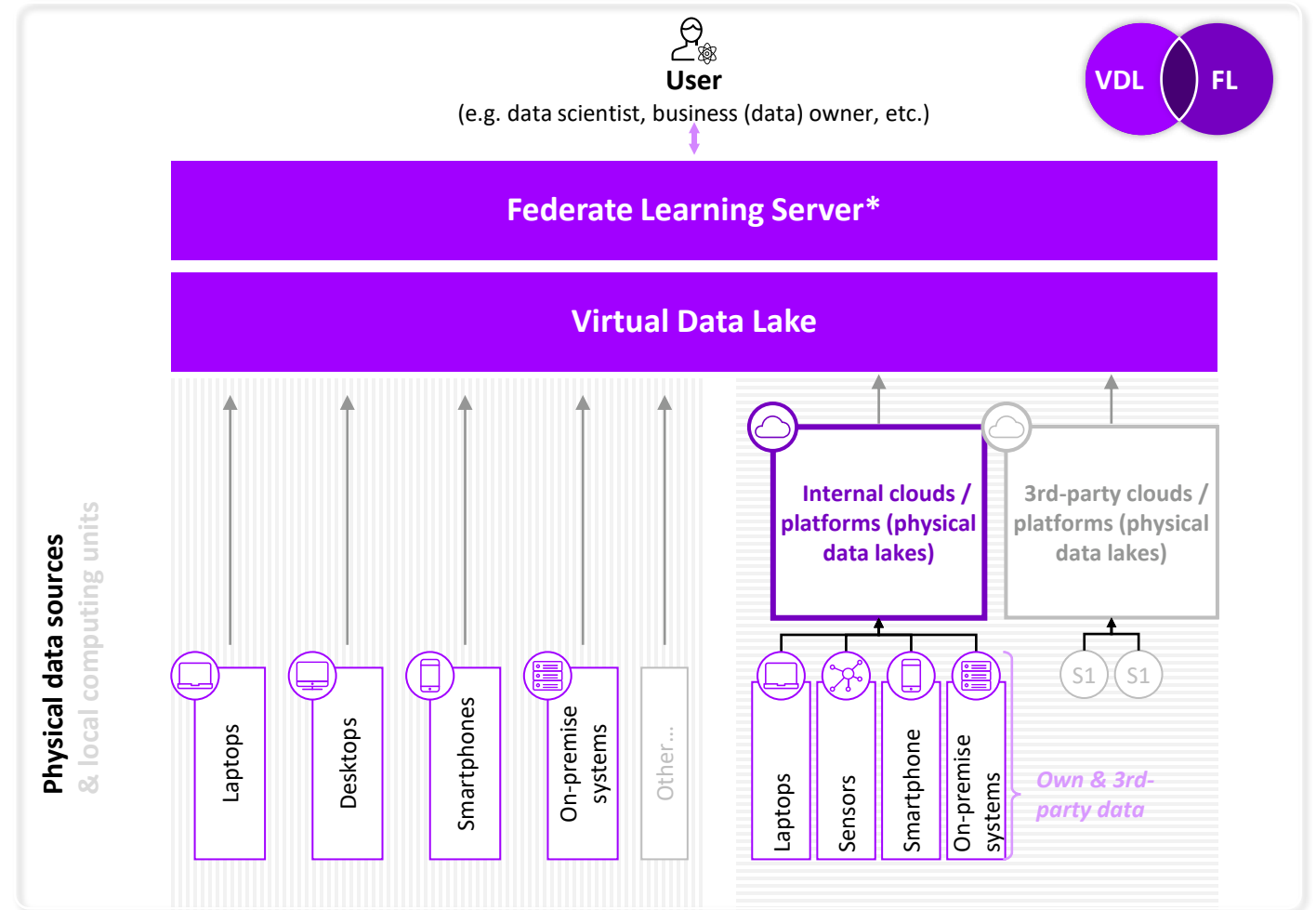**CENTRALIZED DATA PLATFORMS MAKE BUSINESS SENSE WHEN …!**

- Data or AI use case **value is proven**

- **Centralized data ownership** brings added business **value**

- **Regulators require** long-term data storing

- Etc.

However, clients increasigly need to be also able to embrace data silos that do exist across multiple organizations and data owners

RELYING ON DISTRIBUTED DATA MAKES BUSINESS SENSE WHEN …!

- **Data quality** still needs **to be proven**

- Data or AI use case **value** still needs **to be proven**

- Data belongs to 3rd parties that are interested in collaboration but **are not willing to share data or invest efforts in data access**

- Data sits in **territories where legal / technology restrictions** prohibit data transfer

- Etc.

User
(e.g. data scientist, business (data) owner, etc.)

VDL   FL

Federate Learning Server*

Virtual Data Lake

Physical data sources & local computing units

Laptops | Desktops | Smartphones | On-premise systems | Other…

Internal clouds / platforms (physical data lakes)

3rd-party clouds / platforms (physical data lakes)

Laptops | Sensors | Smartphone | On-premise systems

S1   S1

Own & 3rd-party data

Emerging Trend: Federated Learning driven by Virtual Data Lake enabled Data Privacy

Without Data Sharing and Collaboration
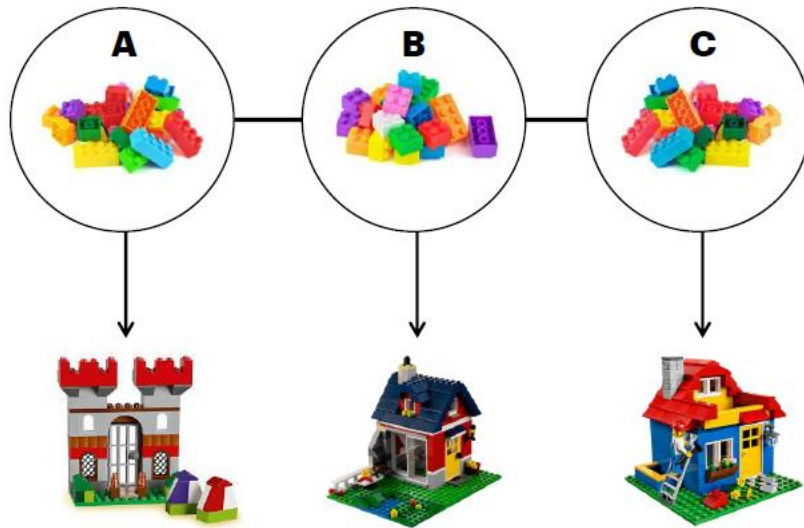
With Data Privacy (using TripleBlind[1], Inpher[1], Fortanix[1] or other vendors) in VDL

Algorithm/Model

[1]Respective company copyrights

Emerging Trend: What's the Magic?
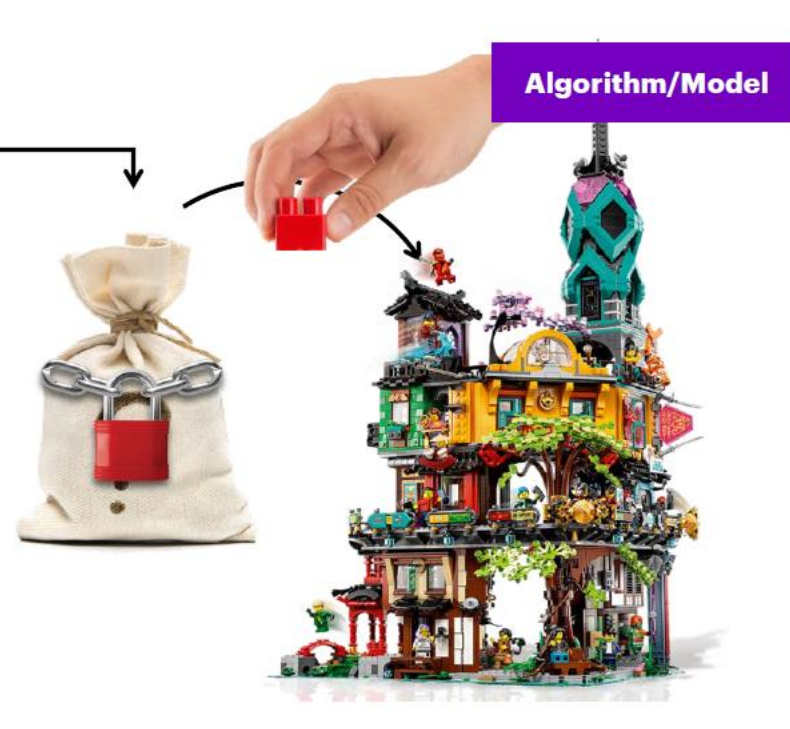
What does this enable?

- **Collaborate** with external & internal partners

- **Expand** access to new data assets at patient level (as relevant)

- **Gain** new and valuable insights

How?

Modular framework that

- Enables **Confidential Computing**[1] and **Privacy Preserving Techniques**[2]

- Enables companies to be **Infrastructure agnostic** and **Cloud ready**

Selected Use Cases

Site Selection & Optimization
HCP & Patient Micro Segmentation
Regulatory Submissions
Adverse Event Prediction
Workforce Understanding

[1] Trusted Execution Environment (Secure Enclaves) [2] Homomorphic Encryption, Secure Multi-Party Computation, Federated Learning, Differential Privacy, etc...

03

Selected Case Studies using Virtual Data Lakes and Federated Learning

Virtual Data Lake (VDL) and Federated Learning (FL) capabilities improve and enable data-driven applications along the entire Life Sciences value chain

| Early R&D | Clinical development & validation | Regulatory approval | Launch | Prevention | Diagnosis | Treatment | Long-term Management |

DATA-DRIVEN APPLICATIONS [1]

- Drug response prediction [2]
- RWD enhanced clinical trial optimization
- RWE enhanced submissions
- Data-driven marketing
- Symptoms assessment tool for patients
- Disease risk prediction
- HCP decision support [2]
- Forecasted treatment prescriptions
- Side effects & symptoms management
- Personalized coaching for well-being
- Outcome based reimbursement

- Digital Biomarker development
- **Clinical trial recruitment [2]**
- **Adverse Event Prediction**
- Chatbot-based support
- Personalized prevention programs
- Automated lab testing
- Advanced diagnostic capabilities [2]
- Increased supply chain transparency
- Adherence support
- Surgical assistance
- Personalized recovery & Rehabilitation

1 list of applications is non-exhaustive
2 selected use cases, further elaborated in deep-dive slides

Clinical trial recruitment - use case overview

⊕ Goal

IMPROVE CLINICAL TRIAL RECRUITMENT FOR PATIENTS AND LS COMPANIES ALIKE
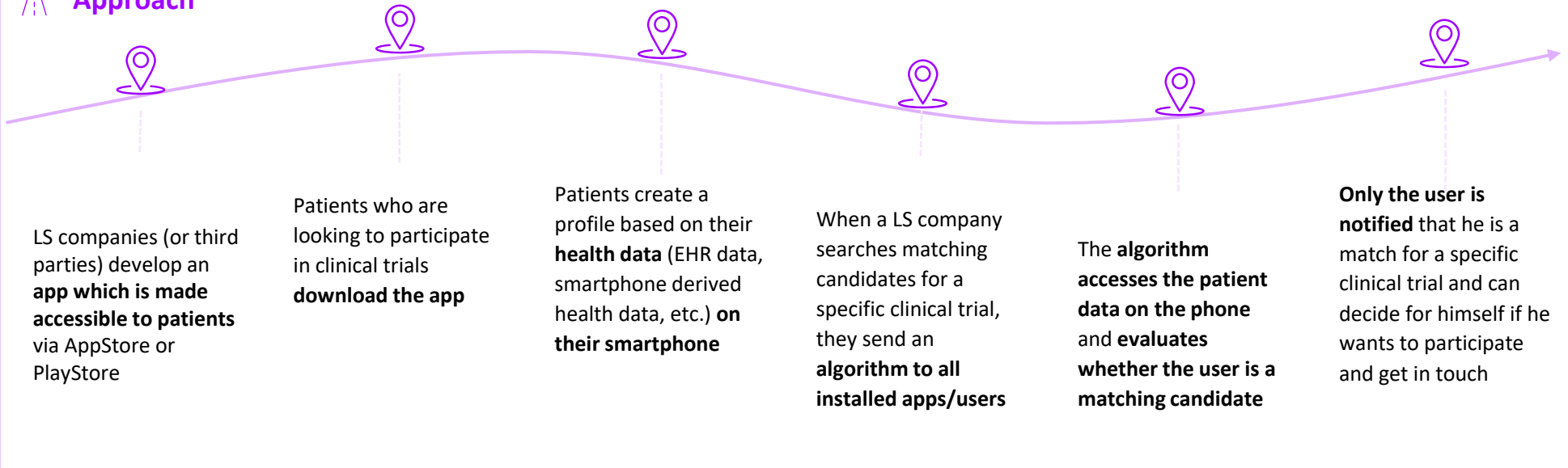
⫟ Approach

LS companies (or third parties) develop an **app which is made accessible to patients** via AppStore or PlayStore

Patients who are looking to participate in clinical trials **download the app**

Patients create a profile based on their **health data** (EHR data, smartphone derived health data, etc.) **on their smartphone**

When a LS company searches matching candidates for a specific clinical trial, they send an **algorithm to all installed apps/users**

The **algorithm accesses the patient data on the phone** and **evaluates whether the user is a matching candidate**

Only the user is notified that he is a match for a specific clinical trial and can decide for himself if he wants to participate and get in touch

◇ Value

For Patients

- Patients do not need to share their data with third parties
- Only the patient knows whether he is a match for a trial or not

For LS Companies

- Direct contact to candidates and faster processes
- More candidates on the platform

⛁ Why VDL & FL

People are reluctant to share their data in a platforms owned by a commercial company which matches candidates to trials
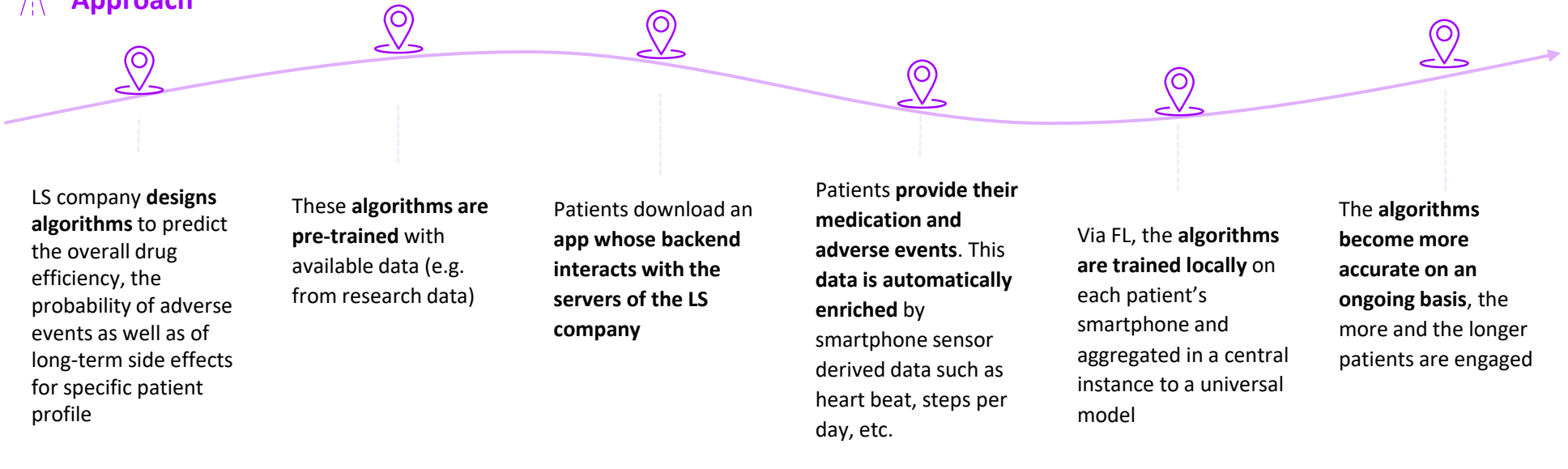
☰ Further limitations

As of now, the data stored on smartphones is limited although the trends of having more and more health data on one's smartphone is favorable for the use case

Pharmacovigilance - use case overview

Goal

PREDICT DRUG EFFICIENCY, ADVERSE EVENTS AND LONG-TERM SIDE EFFECTS FOR **PATIENT ENGAGEMENT**

Approach

LS company **designs algorithms** to predict the overall drug efficiency, the probability of adverse events as well as of long-term side effects for specific patient profile

These **algorithms are pre-trained** with available data (e.g. from research data)

Patients download an **app whose backend interacts with the servers of the LS company**

Patients **provide their medication and adverse events**. This **data is automatically enriched** by smartphone sensor derived data such as heart beat, steps per day, etc.

Via FL, the **algorithms are trained locally** on each patient's smartphone and aggregated in a central instance to a universal model

The **algorithms become more accurate on an ongoing basis**, the more and the longer patients are engaged

Value

For Patients

- Get **truly personalized decision support** on medication (which can be discussed with doctor)

- Monitor **personal tolerance** of drugs

For LS Companies

- Obtain **insights** on drug efficiencies, side effects and long-term side-effects

- Develop **treatment recommendation** tool which can be commercialized to HCP's

Why VDL & FL

- Patients are reluctant to share sensitive data with LS companies
- Contact with patients is not lost after clinical trials

Further limitations

- Patient's trust in the value of the app and in the companies behind the app
- People behavior of not reliably noting adverse events

QA