

Big Data and B2B Platforms: the next big opportunity for Europe



WAVESTONE





WAVESTONE

GRIMALDI | STUDIO
LEGALE



TNO



KOMIS

Final Conference Infrastructure of health databases in T2D

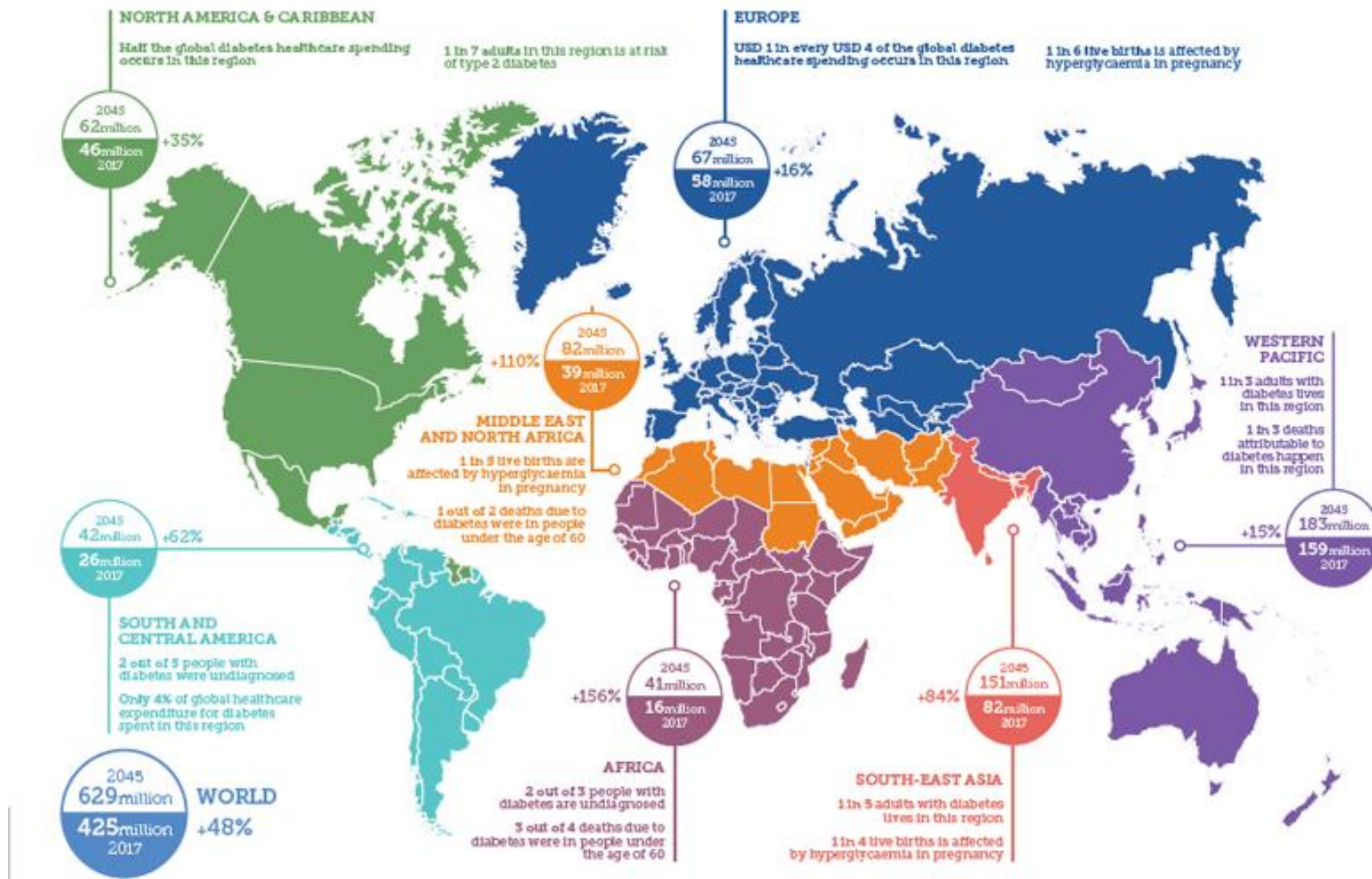
Big Data and B2B platforms: the
next big opportunity for Europe

EASME/COSME/2018/004



26 November 2020

Why an infrastructure of health databases for T2D (projection 2017-2045)?



<https://idf.org/>

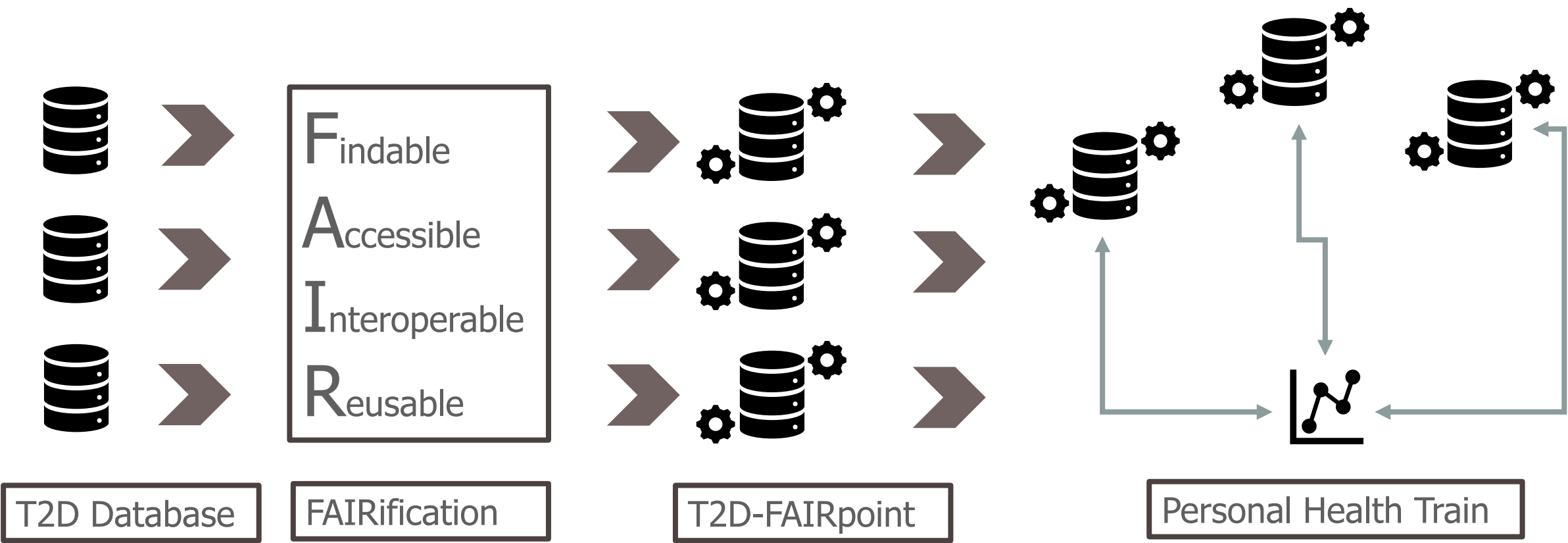
The image shows three tall, cylindrical silos with conical roofs, standing on a grassy hill. The silos are arranged in a row from left to right. The first silo on the left is labeled 'Real world', the middle one is labeled 'Nutrition', and the one on the right is labeled 'Clinical'. The silos have a dark, textured body and a lighter, ribbed conical roof. A blue ladder or access structure is visible on the side of each silo. The background is a clear blue sky, and the foreground is a field of green and brown grass.

Real world

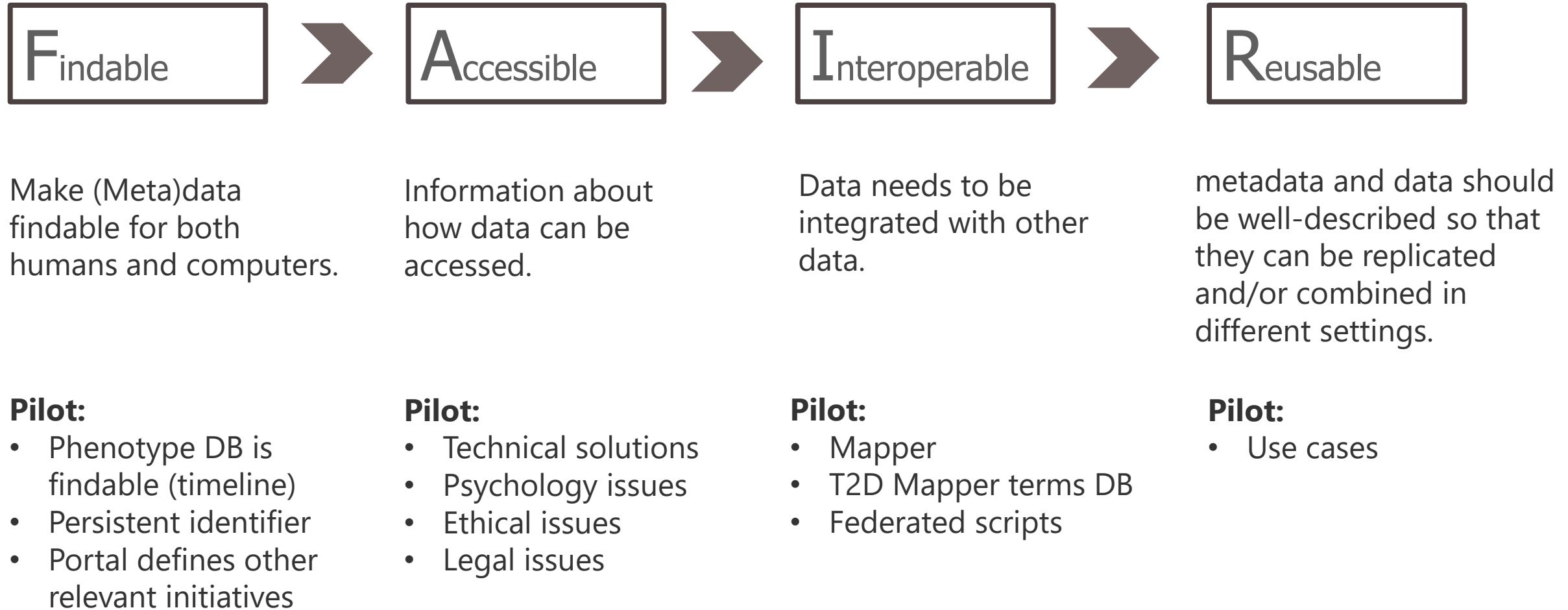
Nutrition

Clinical

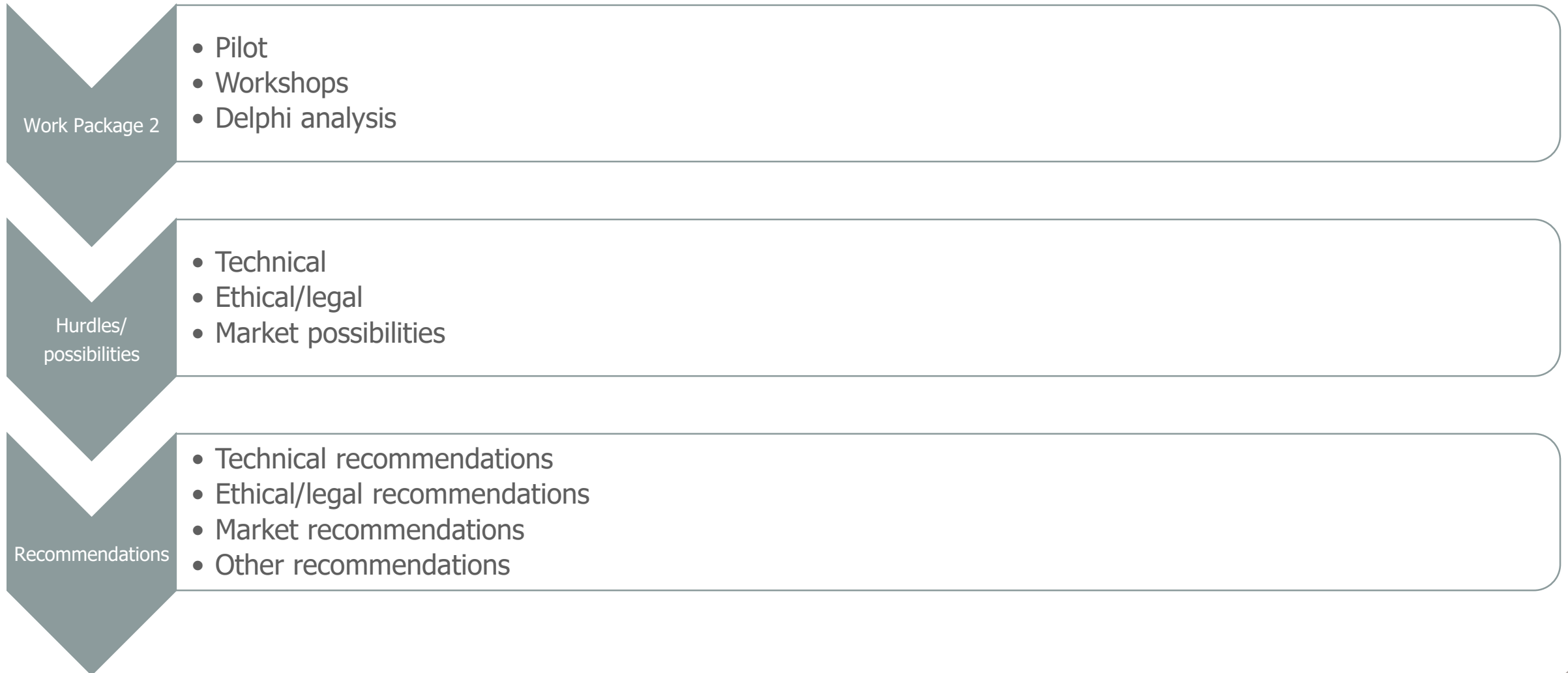
FAIR/PHT APPROACH for T2D databases; the concept



FAIR implementation



Workflow



FINDABLE: An overview of big data initiatives relevant for the pilot

Federated Big Data and B2B Platform

+

← → ↻ 🏠

🔒 🛡️ 🚫 🚧 https://dashin.eu/ft2d/diabetes_data_initiatives


📄 ⋮ 📌 ⭐

🔍 📖 ☰

FT2D Federated Big Data and B2B Platform for Diabetes

Data ▾ Information ▾ Account ▾

https://dashin.eu/ft2d




Metadata: Not available

Data: Not available

Machine readability: Not applicable

FAIR4Health is a H2020 ‘science with and for society’ project. This program will make all data from publicly funded research FAIR. The specific scope of the project is:

1. To support the FAIRification of data, stressing on data quality (certification), their interoperability and reproducibility of research.
2. To generate path finder use cases to demonstrate how data sharing and re-use can generate a ground-




Metadata: Not available

Data: Not available

Machine readability: Not applicable

FAIRplus is a project that aims to increase the discovery, accessibility and reusability of data from selected projects funded by the EU-IMI and internal data from pharmaceutical industry partners. It will also organise training for data scientists in academia, SMEs and pharmaceutical companies to enable wider adoption of best practices in life science data management. FAIRplus collaborates with the aforementioned FAIR4Health project. The SME The Hyve, a company involved in developing open source software for research is involved in the current project.



The European Open Science Cloud for Research (EOSC) pilot project will improve the ability to preserve and reuse data resources and provide an important step towards building a dependable open innovation environment where data from publicly funded research is always open and there are clear incentives and rewards for the sharing of data and resources. The current T2D pilot might in the future be part or be under the umbrella of EOSC. There are a couple of SMEs involved in this project: TRUST-IT, CSC, Surf-sara, all companies that develop IT solutions for research infrastructure.

Remarks: SURF-SARA provides access to [LifeLines](#) that offers rich patient data

FINDABLE: An overview of relevant diabetes initiatives

Federated Big Data and B2B Platform


https://dashin.eu/ft2d/diabetes_initiatives

FT2D Federated Big Data and B2B Platform for Diabetes

Data

Information

Account




Metadata: Not available

Data: Not available

Machine readability: Not applicable

Feel4Diabetes is part of the Global Alliance for Chronic Diseases (GACD) initiative. GACD is a collection of the world's biggest public research agencies that mostly funds joint programs for life-style related diseases like diabetes. The most relevant GACD program is Feel4Diabetes. The Feel4Diabetes




Metadata: Not available

Data: Not available

Machine readability: Not applicable

Healthy diet for a healthy life (JPI-HDHL). Within the JPI-HDHL 26 countries from within and outside Europe are working on a programmed approach to align national research strategies to fund new research in order to facilitate true understanding of the relationship between diet, physical activity and health.



Metadata: Not available

Data: Not available

Machine readability: Not applicable

The goal of the RHAPSODY program is the development of precision therapy and precision prevention for type 2 diabetes. The specific aim is to define a molecular taxonomy of type 2 diabetes mellitus (T2D) that will support patient segmentation, inform clinical trial design, and the establishment of regulatory paths for the adoption of novel strategies for diabetes prevention and treatment. This project is highly relevant for the current pilot since it will develop federated

9

FINDABLE: An overview of health data marketplaces

Federated Big Data and B2B Platform

+

← → ↻ 🏠

🔒 <https://dashin.eu/ft2d/marketplaces>


📄 ⋮ 🛡️ ☆

📁 📖 ABP 🗑️ 📁


☰

FT2D Federated Big Data and B2B Platform for Diabetes


Data ▾ Information ▾ Account ▾




UTILE is a EU-Health Innovation Marketplace. UTILE is an online community where Providers (like research centers, universities and research groups) and Innovation Developers (like big pharma and biotech companies) can meet each other. LifeArc is an SME aimed at the translation of science into practice is involved in UTILE.




The EU MyHealthMyData project is aiming at creating the technical framework for a true information marketplace based on new mechanisms of trust and direct, value-based relationships between EU citizens, hospitals, research centers and business.




The goal of INNOLABS is to leverage cross capacity building between ICT, Health, BIO and Medicine sectors for new emerging industries in personalized health. INNOLABS will select more than 100 innovative project ideas to accelerate and further support monetarily and with innovation services.



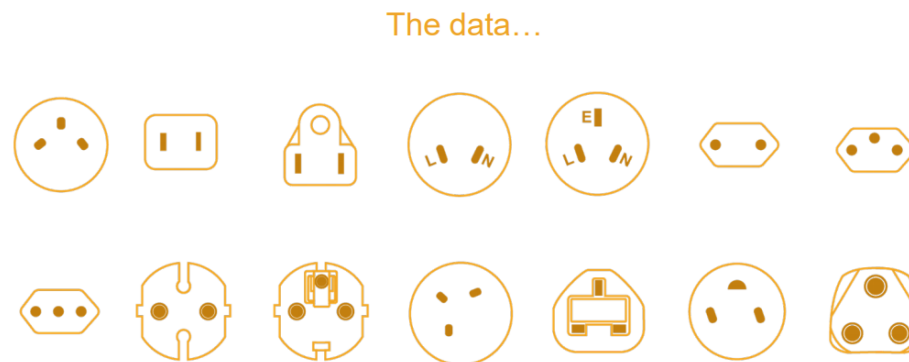
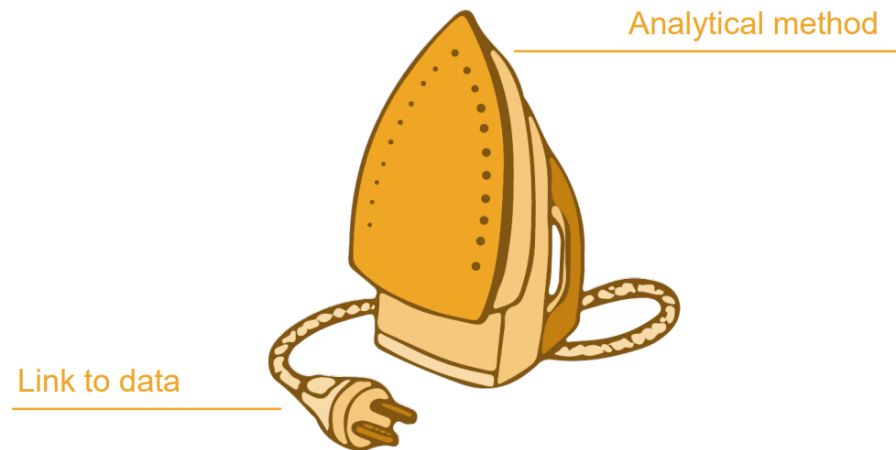
FinnData is an initiative enabled by the Finnish government. FinnData is



Swiss ETH professor Ernst Hafen founded MiData.coop, a platform for the collection



INTEROPERABLE: Ensuring similar semantics (first step in FAIRness)



INTEROPERABLE: Ensuring similar semantics (first step in FAIRness)

Search term to map
to ontology

Ontology Mapper Demo

Search term

Insulin

Enter ontology database (optional)

CHEBI

Search ontologies

Number of ontologies found: 7

Showing top 7

Name	PID	Description
<input type="radio"/> insulin	http://purl.obolibrary.org/obo/CHEBI_145810	A peptide hormone which consists of two polypeptide chains, A- and B- chains which are linked together by disulfide bonds. The amino acid sequence of insulin varies across species and certain segments of the molecule are highly conserved. In most species, the A chain consists of 21 amino acids and the B chain consists of 30 amino acids. In mammals, insulin is synthesised in the pancreas within the beta cells whereas in certain species of fish, the insulin-producing tissue is uniquely located in separate structures called Brockmann bodies.
<input type="radio"/> insulin (human)	http://purl.obolibrary.org/obo/CHEBI_5931	An insulin that is produced in the pancreas and involved in regulating the metabolism of carbohydrates (particularly glucose) and fats. Commonly thought of as a protein, it consists of two peptide chains, one containing 21 amino acid residues and the other containing 30; the chains are joined together by 2 disulfide bonds. Recombinant insulin is identical to human insulin, but is synthesised by inserting the human insulin gene into E. coli,

Add selected ontology

Optional preferred ontology to
be mapped to

Resulting persistent identifiers
associated with search term
(and preferred ontology)

INTEROPERABLE: Ensuring similar semantics

Ontology Mapper Demo

Search term

Enter ontology database (optional)

Search ontologies

Number of ontologies found: 12

Showing top 10

	Name	PID
<input type="radio"/>	insulin	http://purl.obolibrary.org/obo/CHEBI_145810
<input checked="" type="radio"/>	insulin (human)	http://purl.obolibrary.org/obo/CHEBI_5931
<input type="radio"/>	insulin secretagogue	http://purl.obolibrary.org/obo/CHEBI_90415
<input type="radio"/>	insulin dithiol	http://purl.obolibrary.org/obo/CHEBI_5932

Add selected ontology

Preferred term
selected from list
by an expert

Mapped ontologies

Name

PID

glucose

<http://purl.bioontology.org/ontology/CPT/1011517>

Glucose

<http://purl.bioontology.org/ontology/CPT/1011517>

water

<http://purl.bioontology.org/ontology/SNOMEDCT/11713004>

insulin

(human)

http://purl.obolibrary.org/obo/CHEBI_5931

Download ontologies

Term is then saved to
Mapped Ontologies

CSV of terms to persistent identifiers
is available to download directly

INTEROPERABLE: Ontology Mapper Architecture Overview



Name
Creatinine
creatinine_Creatinine.v001
HDL-cholesterol.v004
HDL-cholesterol.v005
HbA1c.v000
HbA1c.v003
HbA1c.v005
BMI (body mass index).v000
BMI (kg/m2)
BodyLength.v000
BodyWeight.v002
Body_Weight

Ontology Mapper



Name	Identifier
Creatinine	Lifelines.LABDATABL.BKR.5.1.1002
HDL Cholesterol	Lifelines.LABDATABL.HDC.5.1.1002
HbA1c	Lifelines.LABDATABL.HB1C.5.1.1002
HbA1c	Lifelines.LABDATABL.HBAC.5.1.1002
Body Mass Index	Lifelines.BEZOEK1.BMI.5.1.1002
Length	Lifelines.BEZOEK1.LENGTE.5.1.1002
Weight	Lifelines.BEZOEK1.GEWICHT.5.1.1002

Name	PID
Creatinine	http://purl.bioontology.org/ontology/MEDLINEP/LUS/C0010294
Creatinine	http://purl.bioontology.org/ontology/MEDLINEP/LUS/C0010294
creatinine_Creatinine.v001	http://purl.bioontology.org/ontology/MEDLINEP/LUS/C0010294
HDL Cholesterol	http://purl.obolibrary.org/obo/ddo.owl#ddo_0000276
HDL-cholesterol.v004	http://purl.obolibrary.org/obo/ddo.owl#ddo_0000276
HDL-cholesterol.v005	http://purl.obolibrary.org/obo/ddo.owl#ddo_0000276
HbA1c	http://purl.obolibrary.org/obo/DDO.owl#DDO_0000243
HbA1c	http://purl.obolibrary.org/obo/DDO.owl#DDO_0000243
HbA1c.v000	http://purl.obolibrary.org/obo/DDO.owl#DDO_0000243
HbA1c.v003	http://purl.obolibrary.org/obo/DDO.owl#DDO_0000243
HbA1c.v005	http://purl.obolibrary.org/obo/DDO.owl#DDO_0000243
Body Mass Index	http://purl.obolibrary.org/obo/CMO_0000105
BMI (body mass index).v000	http://purl.obolibrary.org/obo/CMO_0000105
BMI (kg/m2)	http://purl.obolibrary.org/obo/CMO_0000105
Length	http://purl.obolibrary.org/obo/PATO_0000122
BodyLength.v000	http://purl.obolibrary.org/obo/PATO_0000122
Weight	http://purl.obolibrary.org/obo/PATO_0000128
BodyWeight.v002	http://purl.obolibrary.org/obo/PATO_0000128
Body_Weight	http://purl.obolibrary.org/obo/PATO_0000128

Policies needed to create an infrastructure of T2D health databases

Data access remains an issue and needs policies to change (Findable)

/ Psychological issues: *the data is mine...*

Developing a common and dynamic standard for data exchange, operating according to FAIR principles (Accessibility, Interoperability)

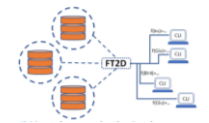
Developing an EU framework for the secondary use of health data (Reusable)

REUSABLE: Identify the technical hurdles to create an infrastructure of T2D databases

Two use cases

- To **extract and exploit data sets** from both, **'omics' profiling and other biomarkers** from key European and national research projects or prospective **cohorts related to T2D**. These data could be used to develop a comprehensive biomarker package that quantifies the relevant processes of the metabolic flexibility system and determines an individual's metabolic health trajectory and predisposition to certain conditions.
- To **combine such data repository** with **Real World Evidence (RWE)** data. RWE is defined as healthcare information that is derived from multiple sources outside the typical clinical research setting.

Use case 1-1



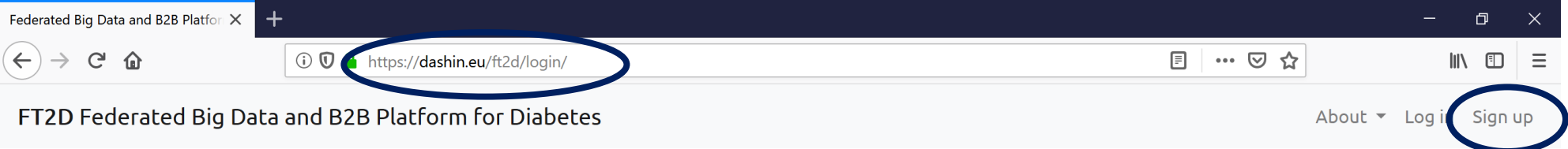
Use case 1-2



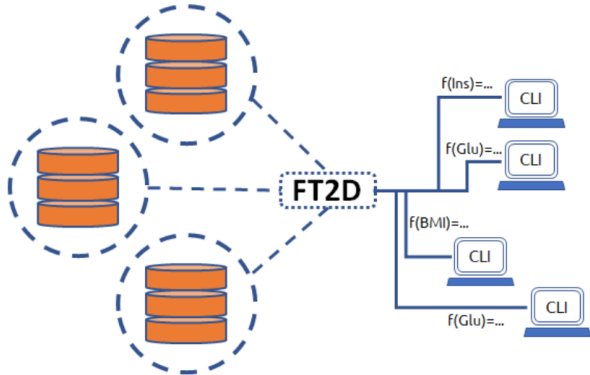
Use case 2



REUSABLE: Portal for the outcome of the pilot



FT2D is a web service that provides privacy-preserved federated machine learning models and data analysis for type 2 diabetes. FT2D makes it possible to train machine learning models on type 2 diabetes related clinical data without having to worry about privacy of the patients.



[Click here](#) to learn more about how it works.

[Click here](#) if you would like to sign up and start using FT2D

<https://dashin.eu/ft2d>

Username:

Password:

REUSABLE: Use case 1: Access to data, use cases and tutorial

Federated Big Data and B2B Platform for Diabetes

OpenShot Video Editor | Download

https://dashin.eu/ft2d/data-tutorial#

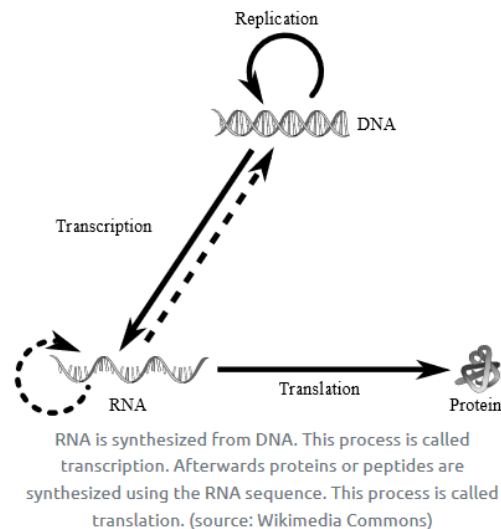
FT2D Federated Big Data and B2B Platform for Diabetes

Data Information Account

Tutorial

In this tutorial, we will try to get some information about β cell dysfunction based on transcriptomics data available in FT2D.

What is transcriptome?



Transcriptome is defined as the collection of RNA (transcripts) collected from individual or population of cells. RNA is synthesized from DNA and it contains the genetic sequence that is necessary to synthesize proteins or peptides.

What are β cells?

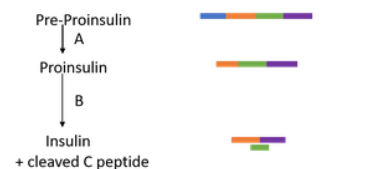


Figure 1: Progression of insulin-like structures. A. The signal peptide of pre-proinsulin is cleaved, forming proinsulin. B) Proinsulin is folded in the ER, then transported to the Golgi apparatus where the C-peptide is cleaved using type I and type II endoproteases to form free C peptide and mature insulin.

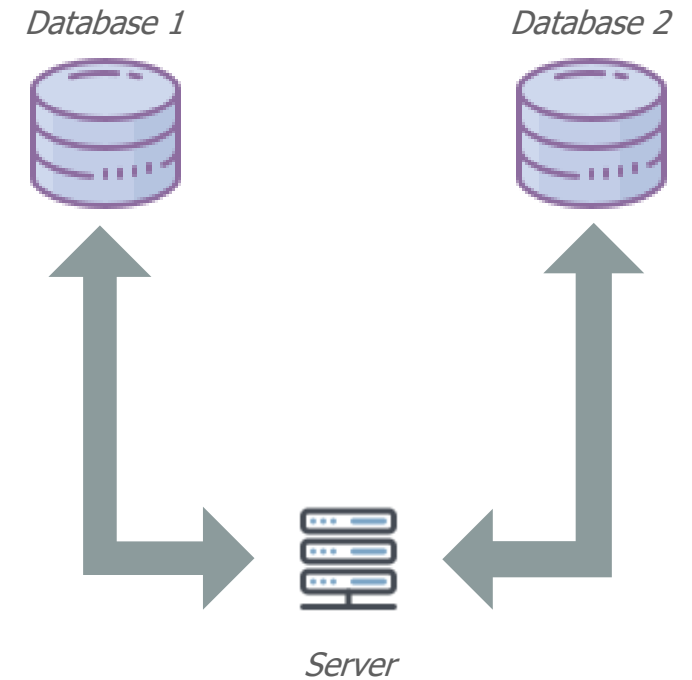
REUSABLE: Prediction of early onset of type 2 diabetes applying a Federated model

- **Task:** Early prediction of Diabetes II onset
- **Model:** Logistic regression – probabilistic model, commonly used for Diabetes II prediction

- - › Federated learning

Branch of Machine Learning that assumes the data to be split among data bases – not centralized.

It works by applying the model locally and combining the local models.



REUSABLE: Use case 1: Federated learning - Lifelines

Lifelines is a large, multigenerational cohort study that includes over 167,000 participants (10%) from the northern population of the Netherlands.

Measurements:

- Anthropometrics
- Clinical measures
- Different proteins (lab measurements)
- Lifestyle (questionnaires) and other self-reporting data

Diagnosis of T2D: Based on questionnaire, plus levels of fasting glucose and HbA1c
2 assessments with 2-11 years in between, most around 4 years

Subjects:

1. Consistently healthy
2. Healthy in the 1st assessment, Diabetic in the 2nd

Resulting clean dataset: 106k subjects, 74 measurements for each



For our case: LifeLines data is artificially split in several parts to simulate different databases, starting with two.

Performance & problems

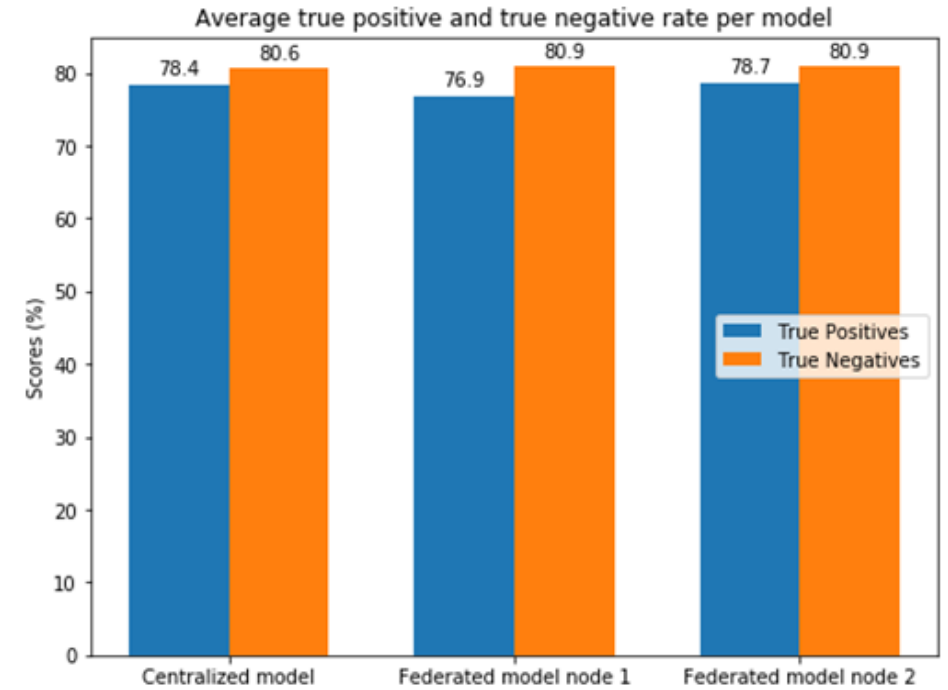
The federated model is able to predict around 76-79% of the future patients in both “databases”, equivalently to the centralized model

Obstacles

Access to the data in a restricted workspace

Not possible to connect to other databases

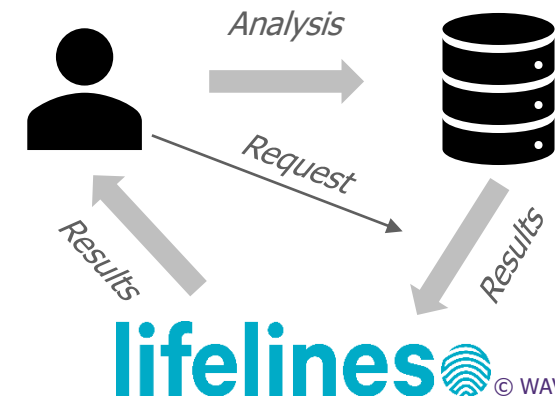
Human readable data, but impossible to automate communication for real Federated Learning applications



IDEAL SET UP

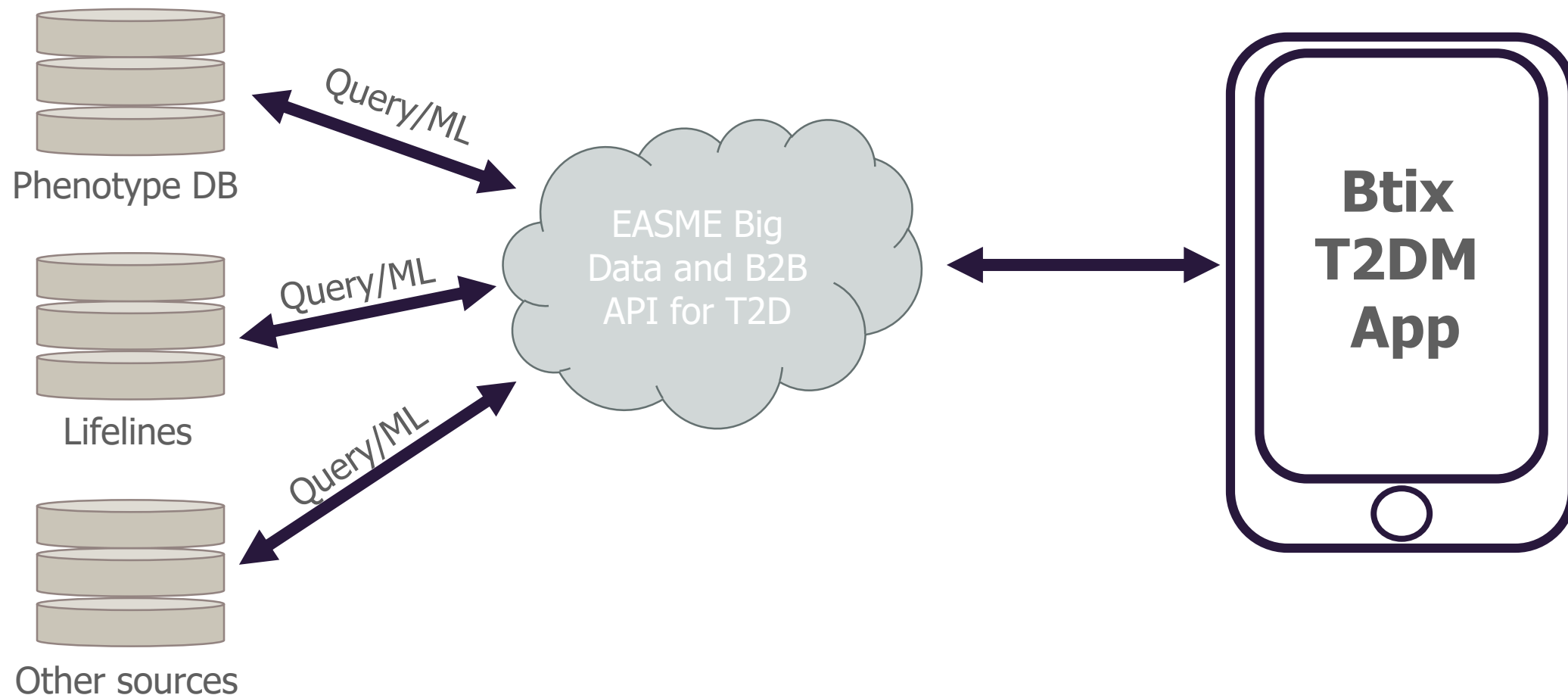


LIFELINES SET UP



REUSABLE: Demo 3 - use case 2 connecting research data & real-world data in an app

Predicting insulin levels using self sampled glucose measurements



REUSABLE: DEMO movie

SHOW TASKBAR

DISPLAY SETTINGS

END SLIDE SHOW

0:16:21

9:29 AM

Use case 1: Insulin prediction

Using the following code, they embed the API to their mobile application

```
getRequest = (args: Object): Promise<Response> => {  
  let requestArguments = [];  
  for(let [key,value] of Object.entries(args)){  
    requestArguments.push(`${key}=${value}`);  
  }  
  let allArguments = requestArguments.join('&');  
  return fetch('https://dashin.eu/easme/api/calculation?study_code=Diclofenac&${allArguments}', {  
    method: 'GET',  
    headers: {  
      authorization: 'Token [redacted]';  
    }  
  })  
}
```

API link

Secret token provided to the client

Next slide

Use case 1: Insulin prediction

```
graph LR  
  PhenotypeDB[Phenotype DB] -- Query/HL --> CAUSE[CAUSE Inc. Data and EDD API for T2D]  
  Lifelines[Lifelines] -- Query/HL --> CAUSE  
  OtherSources[Other sources] -- Query/HL --> CAUSE  
  CAUSE <--> BtixApp[Btix T2DM App]
```

No Notes.

Slide 20 of 41

Suggested policies to resolve the technical obstacles for health databases in T2D?

- Developing clear and transparent rules for data access (Accessibility)
- Privacy-by-design technology is needed to be accepted as GDPR compliant (Accessibility)
 - / Privacy-by-design implementations exist on a continuum. There is no 'one size fits all' solution. The level of security required to access data is essential to determine use-case specific roadmaps
- Data access remains an issue and needs policies to change (Accessibility)
 - / Technical issues: *Insufficient software available for privacy-aware data analysis and federated learning*
- A diabetes data standard should be pushed & data quality be checked (in a framework) (FAIR)
 - / For diabetes eHealth tools the diabetes ICHOM (outcome measure) can be used (www.ichom.org)
 - / Data and metadata over different sources need to refer to the same terms: we made an ontology mapper
 - / Further enabling data portability through technical requirements

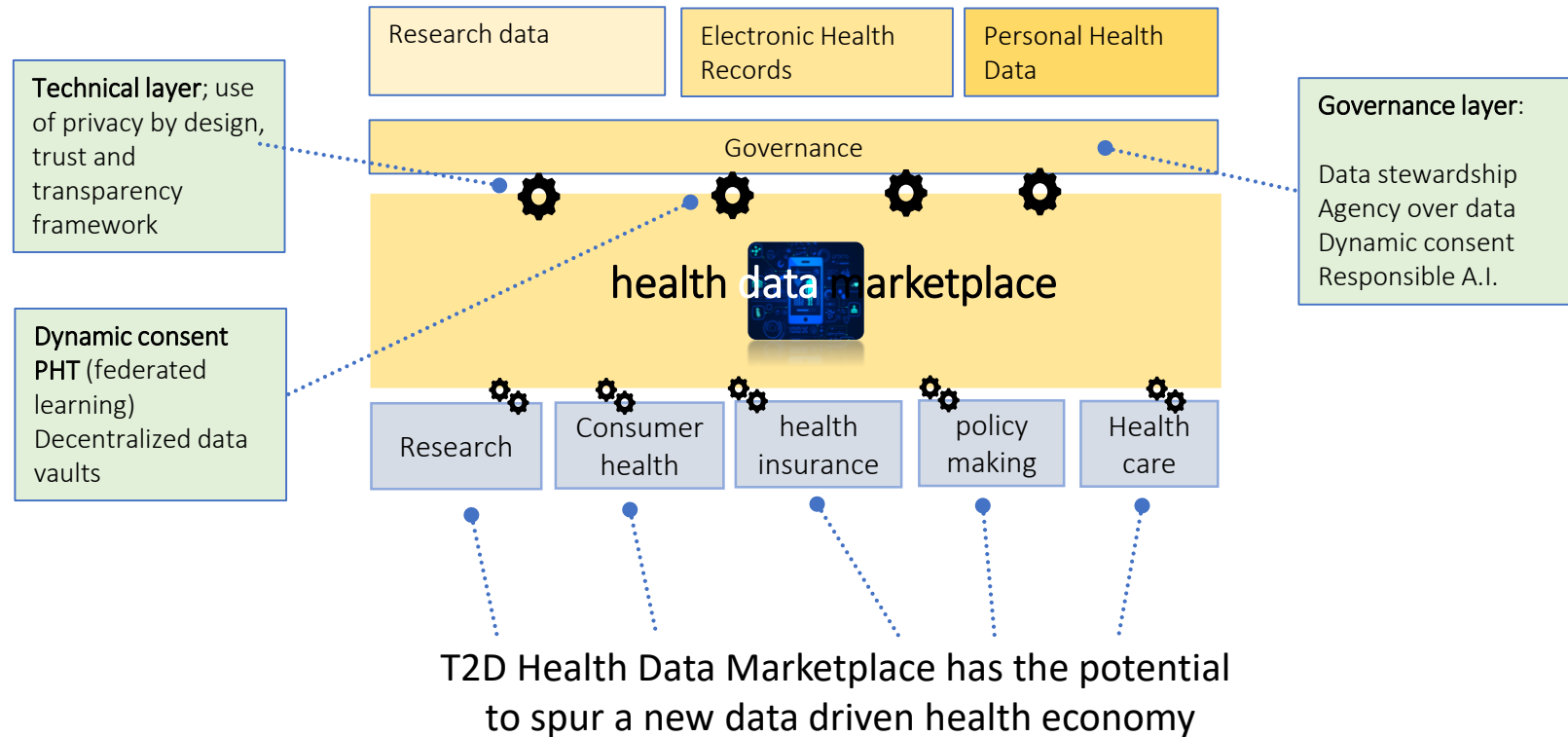
What are the ethical/legal obstacles for health databases in T2D?

DELPHI analysis

- Addressing the fragmentation of liability rules due to diverging national approaches by overseeing their evolution and promoting a coherent approach to enable cross-border data exchanges
- Updating liability rules to meet the challenges of digital transformation
- Creating a 'privacy label' for apps
- Establishing accountability mechanisms for increased transparency

Governance of diabetes data needs to be resolved

- Data stewardship
- Data privacy – trust –transparency
- Security
- Dynamic consent
- Responsible and explainable A.I.
- Agency over data

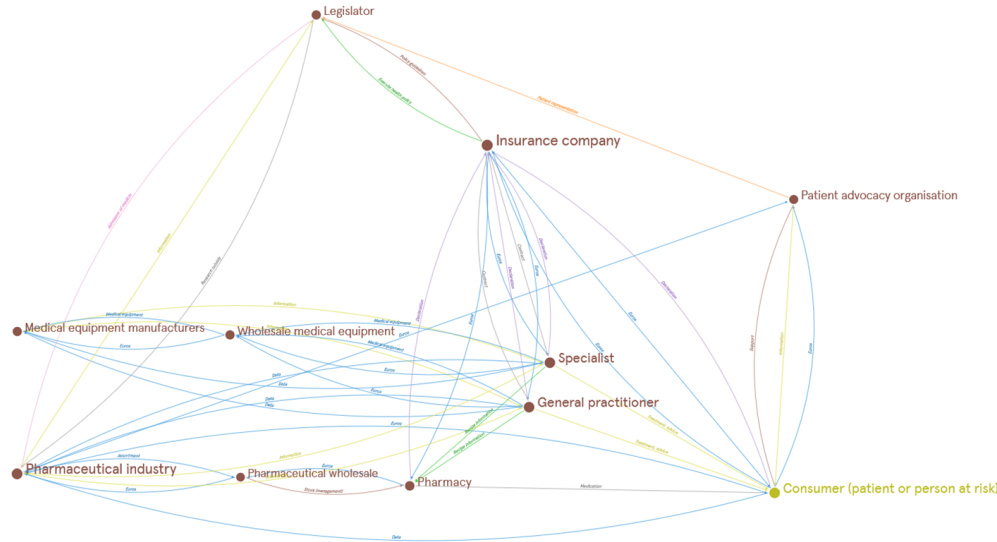


Needed policies to resolve the ethical/legal obstacles for health databases in T2D?

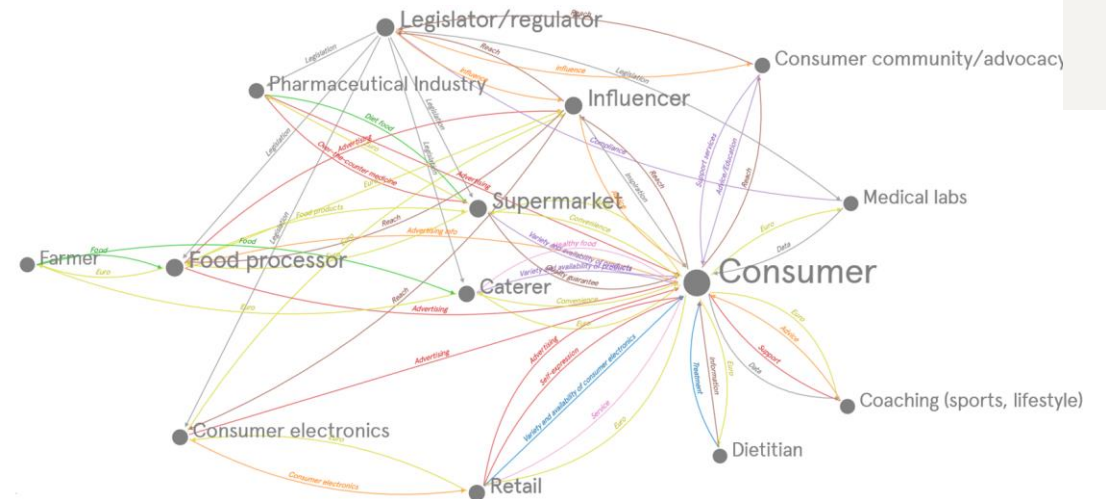
- Data access remains an issue and needs policies to change
 - / Legal issues: *for every data access the contract has to be negotiated and data is often proprietary*
 - / Ethical issues: *Can personal medical data even be used in a commercial setting?*
- Ensuring harmonised implementation of data protection rules.
- Developing guidelines for GDPR-compliant anonymization techniques.

What is the market potential for health databases in T2D?

Medical Value Chain



Health Consumer Value Chain



Relative small market outlook for T2D data solutions (nice markets)

Innovators:

- Pharmaceutical companies
- Insurance companies

Innovators:

- Food processors
- Supermarkets

**Not a direct interest from these stakeholders →
would be different if lifestyle medicine would be reimbursed**

Policies needed to increase the market potential for health databases in T2D?

- Developing business models that support different kinds of stakeholders in accessing the market
- Encouraging national health authorities to develop reimbursement models for health apps
- Supporting research demonstrating the benefits of T2D lifestyle apps in partnership with patient advocacy organisations

Additional recommendations

- Establish a 'European way' for handling data according to European values. This may include that the system that is peer-reviewed and peer-adapted
- Centralizing the organization of handling of health data has advantages, e.g. Therefore, combinations of centralized data-based and control of data by the individual should be explored.
- Establishing education and training policies to prepare data workers with field-specific knowledge, for instance in data science and health studies
- Fostering data literacy skills in healthcare professionals and patients alike
- Supporting the creation of new job profiles acting as facilitators in the digital health ecosystem

The future is bright

Implementation
of
recommendations

Data driven
science on T2D

Fully holistic
data driven
solutions

An official website of the European Union

How do you know?

Cookies

This site uses cookies to offer you a better browsing experience. Find out more on [how we use cookies](#) and [how you can change your settings](#).

[I accept cookies](#) [I refuse cookies](#)

European Commission

English

Search

Home > Press corner > Commission proposes measures to boost data sharing

Available languages: English

Press release | 25 November 2020 | Brussels

Commission proposes measures to boost data sharing and support European data spaces

Page contents

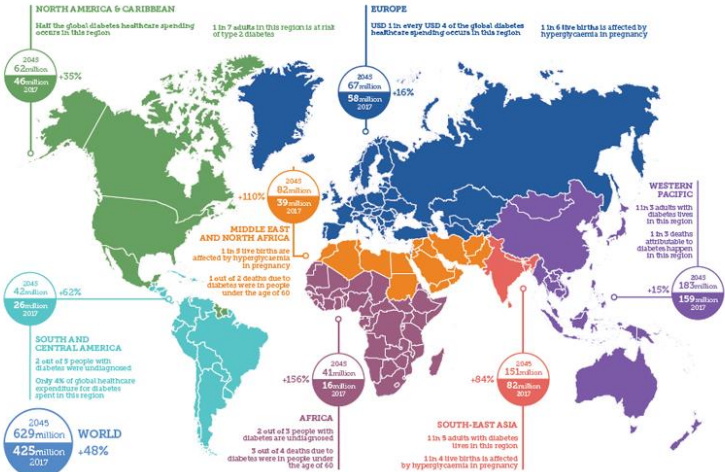
- Top
- Print friendly pdf
- Related media
- Press contact

To better exploit the potential of ever-growing data in a trusted European framework, the Commission today [proposes new EU data governance](#). The Regulation will facilitate data sharing across EU and between sectors to create wealth for society, increase and trust of both citizens and companies regarding their data, an alternative European model to data handling practice of major tech companies.

Europe sets out the rules of the road for its data reuse plan

Natasha Lomas @riptaill / 3:10 PM GMT+1 • November 25, 2020

Comment





WAVESTONE

GRIMALDI | STUDIO
LEGALE



TNO



KOMIS

Thank you for your time

Big Data and B2B platforms: the
next big opportunity for Europe

EASME/COSME/2018/004



7 October 2020