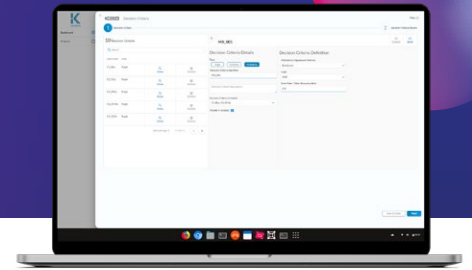


Synthetic Data Generator (SDG)

The unique platform for generating realistic synthetic datasets.



Accessing subject-level data would make a significant contribution to driving innovation across many industries. The use of subject-level data has many applications such as:

- + conducting research
- + testing and validating products
- + building and testing algorithms
- + running simulations
- + business intelligence

However, due to data privacy laws, reluctance and sensitivity around data sharing, it is often too difficult to access data and many organisations have not been able to harness the potential power of this data. An elegant solution to data access issues is to develop synthetic datasets that are derivatives of the subject level data but do not contain any protected health information. These datasets can be shared freely among investigators or those in industry, without raising privacy concerns. The availability of synthetic datasets would fill the void and allow organisations to accelerate the development of new innovative products, services and intellectual property.

Handling complex data scenarios

At the core of our approach is our unique synthetic data generation and simulation platform, **KerusCloud**. **KerusCloud Synthetic Data Generator (SDG)** has broad utility for the generation and exploitation of complex datasets. Leveraging the combined power of analytics and cloud computing,

KerusCloud SDG generates realistic virtual data, the only platform that can handle diverse and complex data collected in disparate data sources. It goes beyond the use of population level statistics as a basis for simulation as it can model the inter-correlation between subject level data such as subgroups and strata, risk factors/covariates and multiple outcomes and data types.

With **KerusCloud SDG**, the user can add in special features such as derived variables, missing data, truncation and censoring. There is the ability to create and compare what-if scenarios that are linked to the source data but contain important variations. This allows **KerusCloud SDG** to generate the most realistic synthetic version of the original data as well as having broad utility for many users.

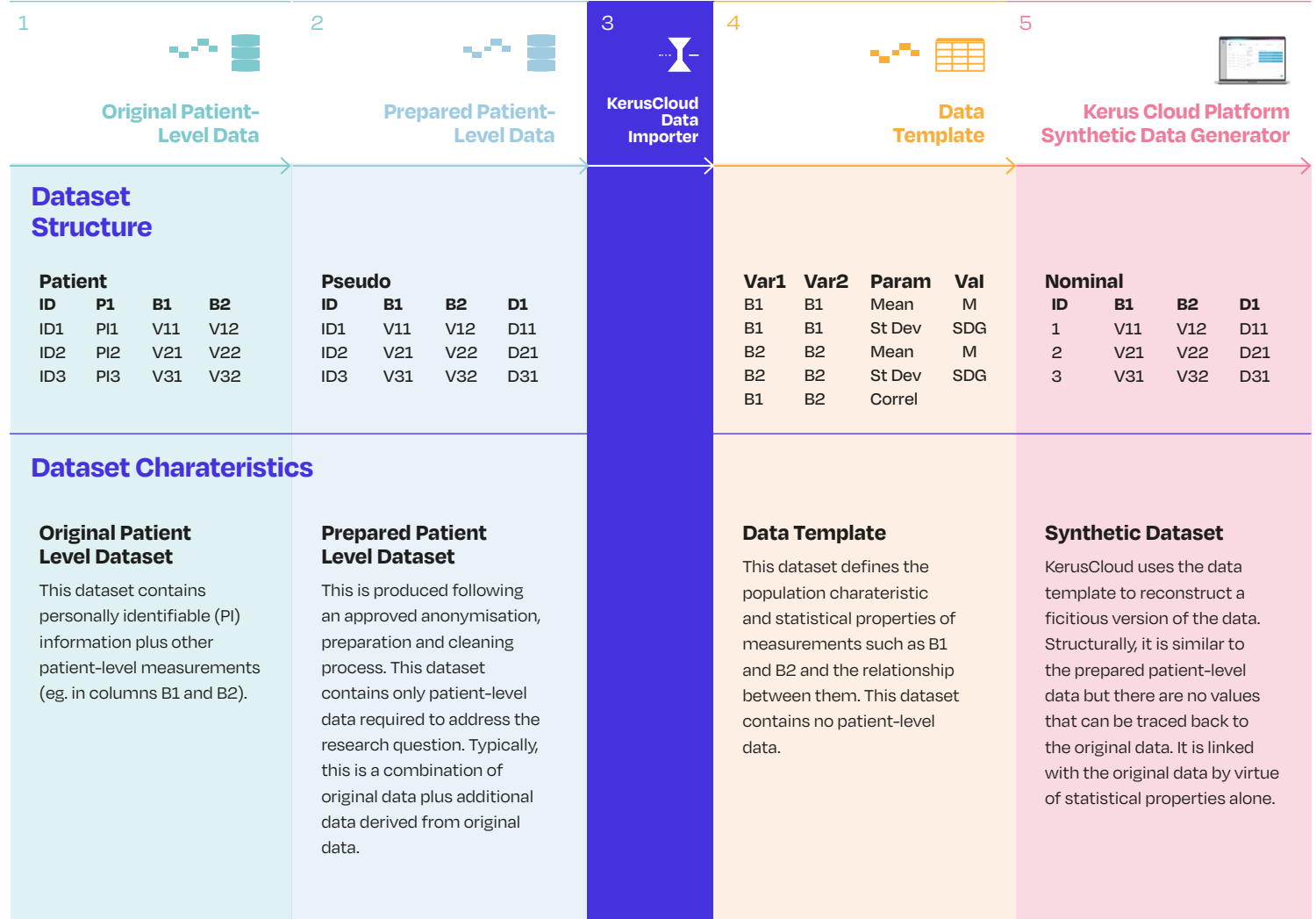
Inputting Data into KerusCloud SDG

There are two ways to input summary statistics into **KerusCloud SDG**. The user can manually define the data structure, input the population characteristics, the data types and the inter-relationships. All of the parameters are defined using standard statistical concepts. There is an alternative approach based on another tool we have developed. The **KerusCloud Data importer (KDi)** is an app which can automatically import subject level data, discover the characteristics and relationships within the data and create templates which allow alternative versions of the data to be generated during simulation. Typically **KDi** works within a private, secure environment so that the original dataset is not transferred.

The outputs created from the **KDi** contain only summary statistics and are structured to facilitate automatic loading into **KerusCloud SDG**.

The four stages of the process to generate synthetic data

Location: Secure Storage or Safe Haven



The four stages of the process to generate synthetic datasets are shown above. Each stage involves an IT system /location, a dataset structure and data characteristics. Once the original patient level data has been identified (stage 1) and prepared (stage 2), they are translated into data templates which are population level summary statistics that describe the features of the data (stage 3). Note, the templates do not contain any patient-level data. These templates are uploaded into the **KerusCloud** platform to randomly generate a synthetic version of the data from stage 2. These data are fictitious.

Download the synthetic data variants

As **KerusCloud SDG** uses the statistical properties of the original data, it can generate multiple unique variants of the source data. The user is able to select the variant that best matches the original data source or to use all the variants. The synthetic data can be downloaded into a CSV file and imported into other packages.

So, don't let data access issues stall your research. Use Exploristics and KerusCloud SDG to generate realistic synthetic data.