



C O R T E X

We clean data

# Our Goal

Connecting and enriching diverse data at scale, rapidly and cost effectively using machine learning algorithms.



Why **clean** data?



Dirty data is the #1 hurdle for data scientists.



**66.7%**

of data scientists say  
cleaning and organizing  
data is their most  
time-consuming task.

**52.3%**

of data scientists cite poor  
quality data as their biggest  
daily obstacle.



What is **digital curation**?



Digital curation is the selection, preservation, maintenance, collection and archiving of **digital assets**.

Digital curation establishes, maintains and **adds value** to repositories of digital data for present and future use.

This has **traditionally** been accomplished by archivists, librarians, scientists, historians, and scholars.



Why do we need digital curation?



In the real world, the **process of preparing data** for any type of analysis, but particularly machine learning, consume the majority of the time.



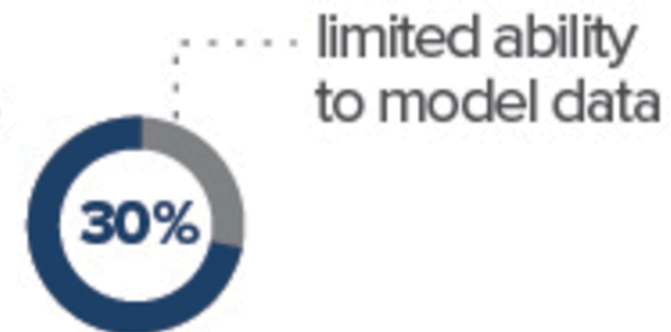
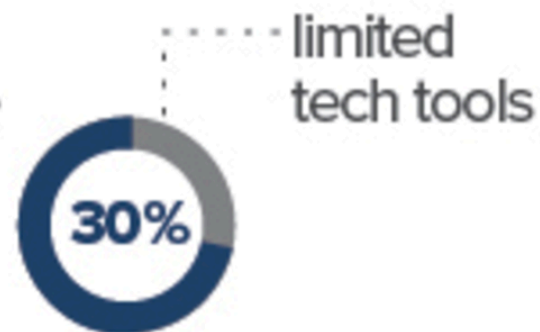
Why built an **AI** for it?



The #2 challenge? Limited resources.



Almost **80%** of data scientists feel **there are not enough data scientists** working in the field.



Costs models and studies on digital curation costs constantly identify humans resources as the **most costly component** of digital curation.



The biggest cost associated with the initial setup of a data warehouse for machine learning applications is often **cleaning** up the data.

So some data problems never go away...they just scale.



**What** Does That Mean?



Analysts can **breakthrough** the data quality and preparation bottlenecks that prevent fast, accurate analysis.



Who needs CORTEX?



Companies and data scientists spending months in **cleaning** and **curating** large data sets for the preparation of running machine learning algorithms.



Where is **CORTEX** now?



We are working with biological and chemical datasets to transform the traditional methods of drug discovery.



# C O R T E X

Outcomes based on evidence, analysis, prediction and execution.

