Data science process models

The goal of the following few sections is to become familiar with several process models being used today. The intent and general flow for all of these models is very much the same. In this course we will use the process of **design thinking**, but it is the consistent application of a process in practice that is important not the exact process itself. There are a number of reasons for choosing the **design thinking** process, but the most important is that it is being applied in a cross-disciplinary way—that is outside of data science. Historically, learning from other disciplines has been a hallmark of scientific achievement.

Important:*This course is not focused on adhering to the specifics of a process model or a way of thinking. The focus is to incorporate best practices into your workflow and in general to scaffold these practices.*

OSEMN

Before the first stage of the OSEMN process a data scientist is required to **identify the business opportunities**. This is an important step in any data science process and perhaps one that deserves its own stage in the process.

According to the OSEMN framework, the elements of data science are:

- Obtaining data
- Scrubbing data
- Exploring data
- Modeling data
- Interpreting data

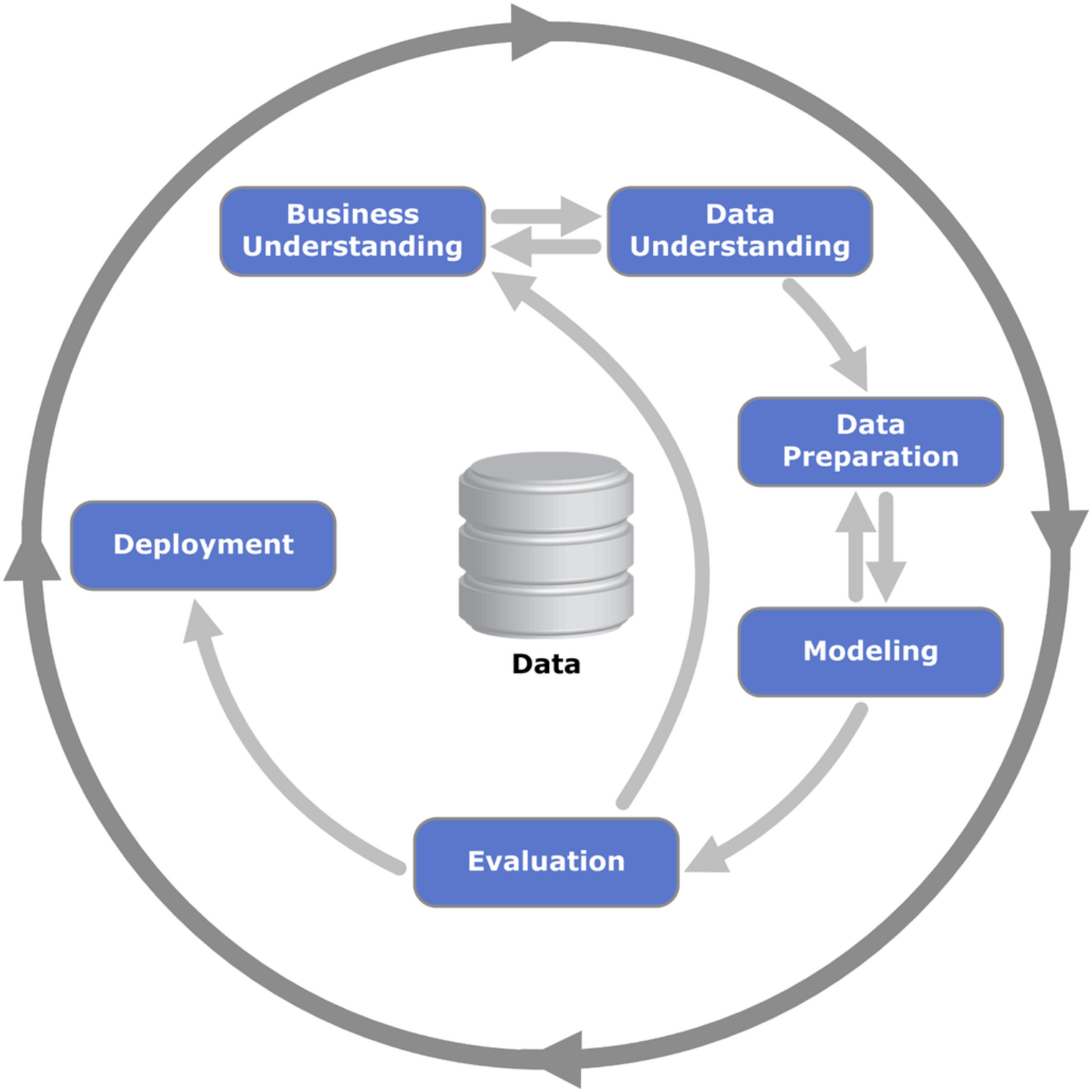
and yes the acronym OSEMN is pronounced as *Awesome*.

OSEMN additional resources

- Towards Data Science blog post
- Medium blog post OSEMN

CRISP-DM

The CRoss-Industry Standard Process for Data Mining (CRISP-DM) is a process model that uses an open standard. The project has been around since 1996 and there are a lot of similarities to the OSEMN process.



Process diagram of the phases of CRISP-DM

It is encouraged to use a nonlinear approach to phase transition, giving this process model a lot of room for variations. The arrows in the diagram indicate the common transitions between stages. There are some known problems with CRISP-DM, but it is not difficult to modify (integrate the business opportunities, add integration etc.) the process to be effective.

CRISP-DM additional resources

- IBM's overview of CRISP-DM

- IBM's CRISP-DM project tool
- Data Science Central blog post on CRISP-DM