

managing the process

Marco Morales

marco.morales@columbia.edu

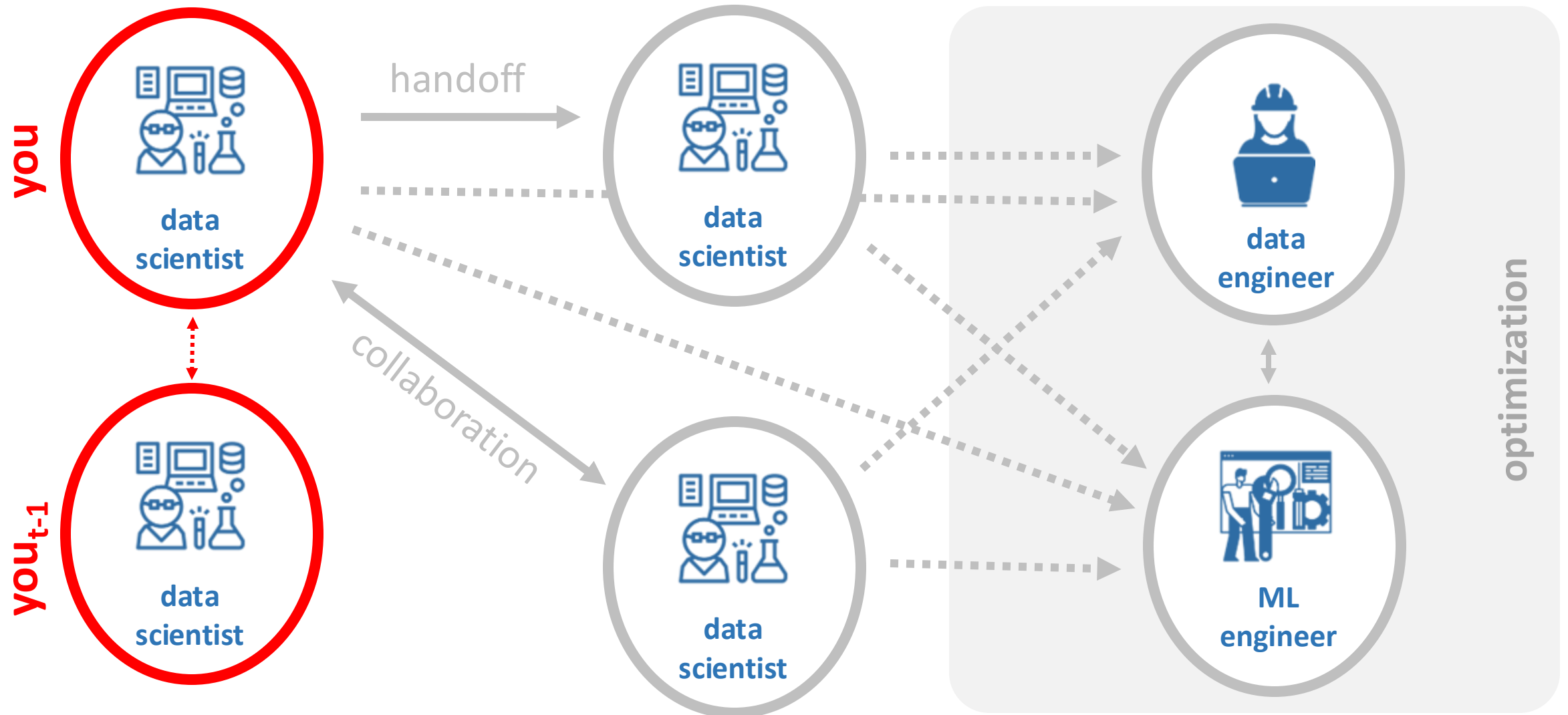
Nana Yaw Essuman

ne2388@columbia.edu

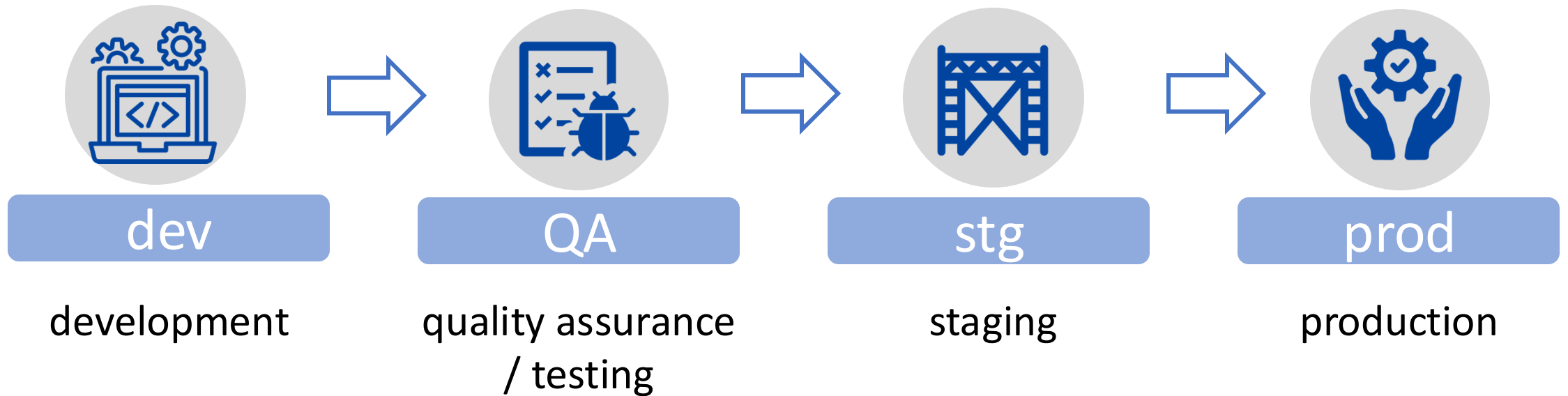
GR5069: Applied Data Science
for Social Scientists

Spring 2025
Columbia University

recap: workflow collaboration in Data Science



recap: working environments



recap: operational concepts in Data Science



portability

anyone should be able
to **pick up where you
left off** from any
machine



replicability

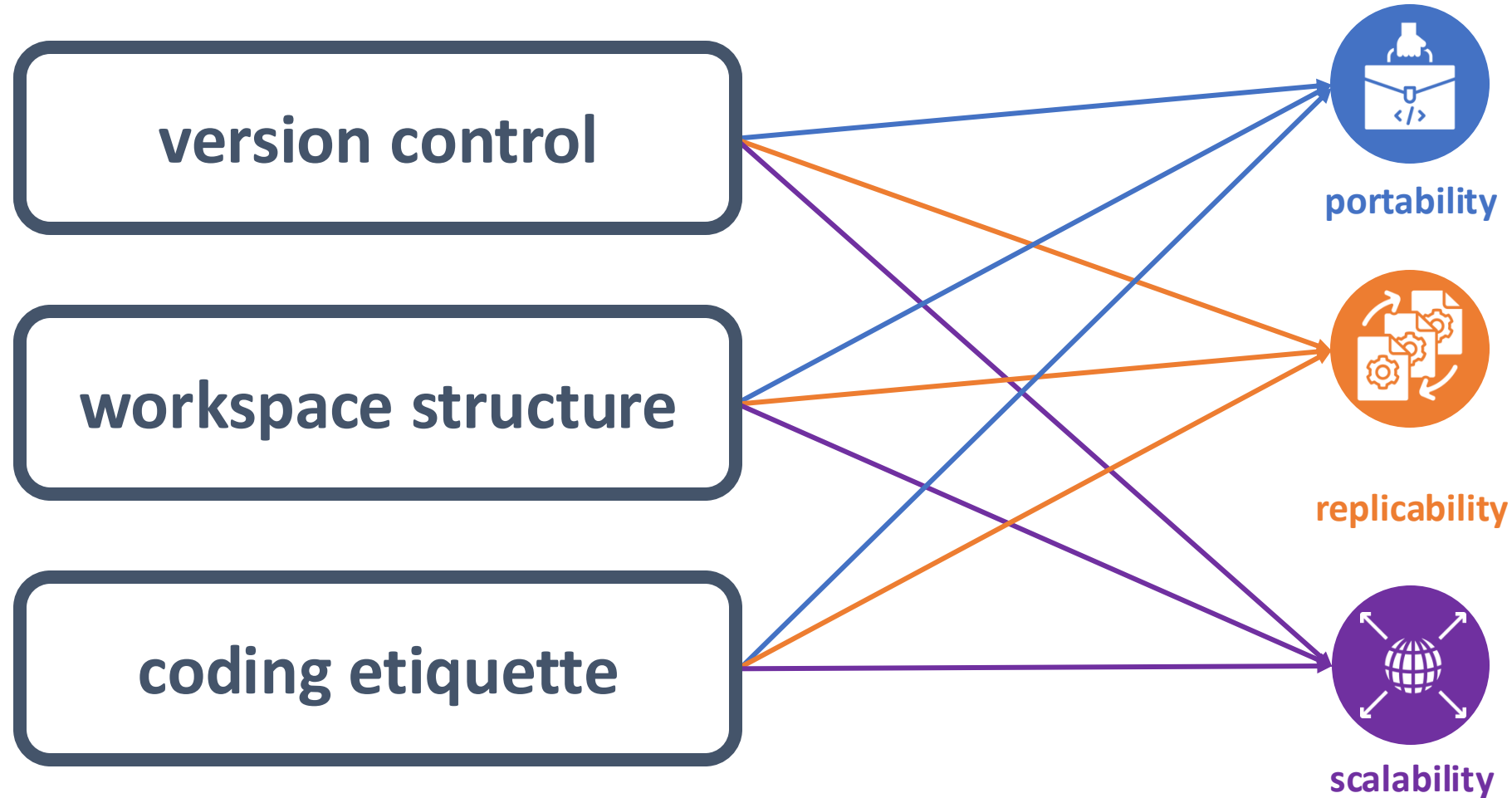
anyone should be able
to arrive at your **same
results**



scalability

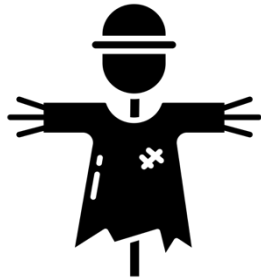
your prototype should
also work for **larger
data sets** and/or be on
the path of **automation**

all pieces to support operational concepts



our focus today:

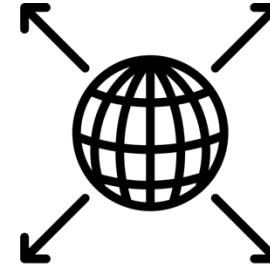
recap: iteration to build Data Products



start small
(MVP)



fail fast

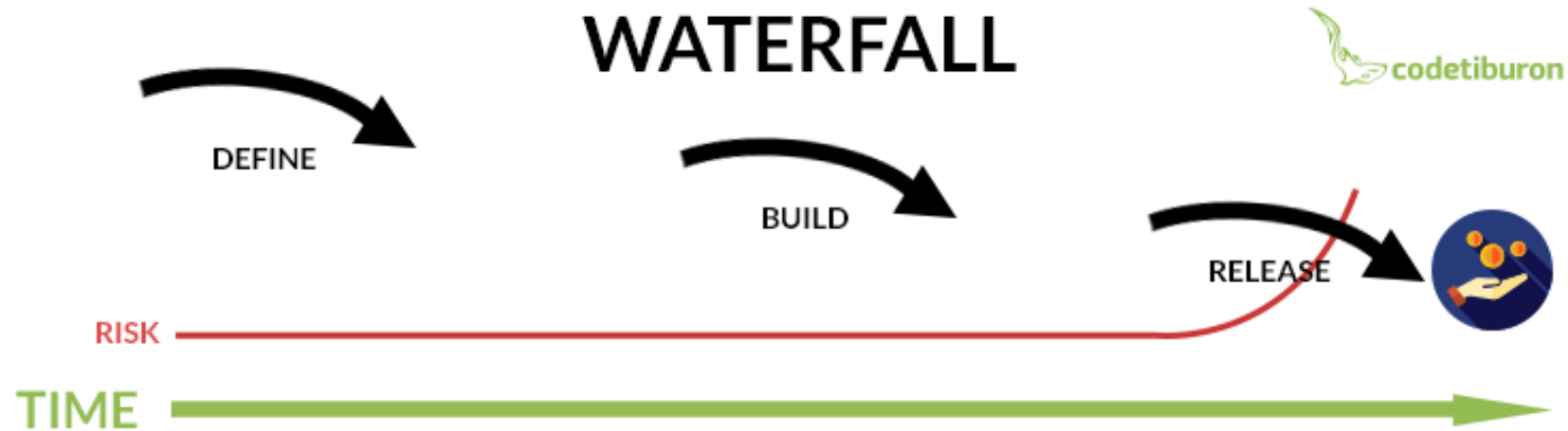


scale up



iterate

a tale of two approaches...



for our purposes, **waterfall** is

➡➡➡ **sequential** approach to manage **development processes**

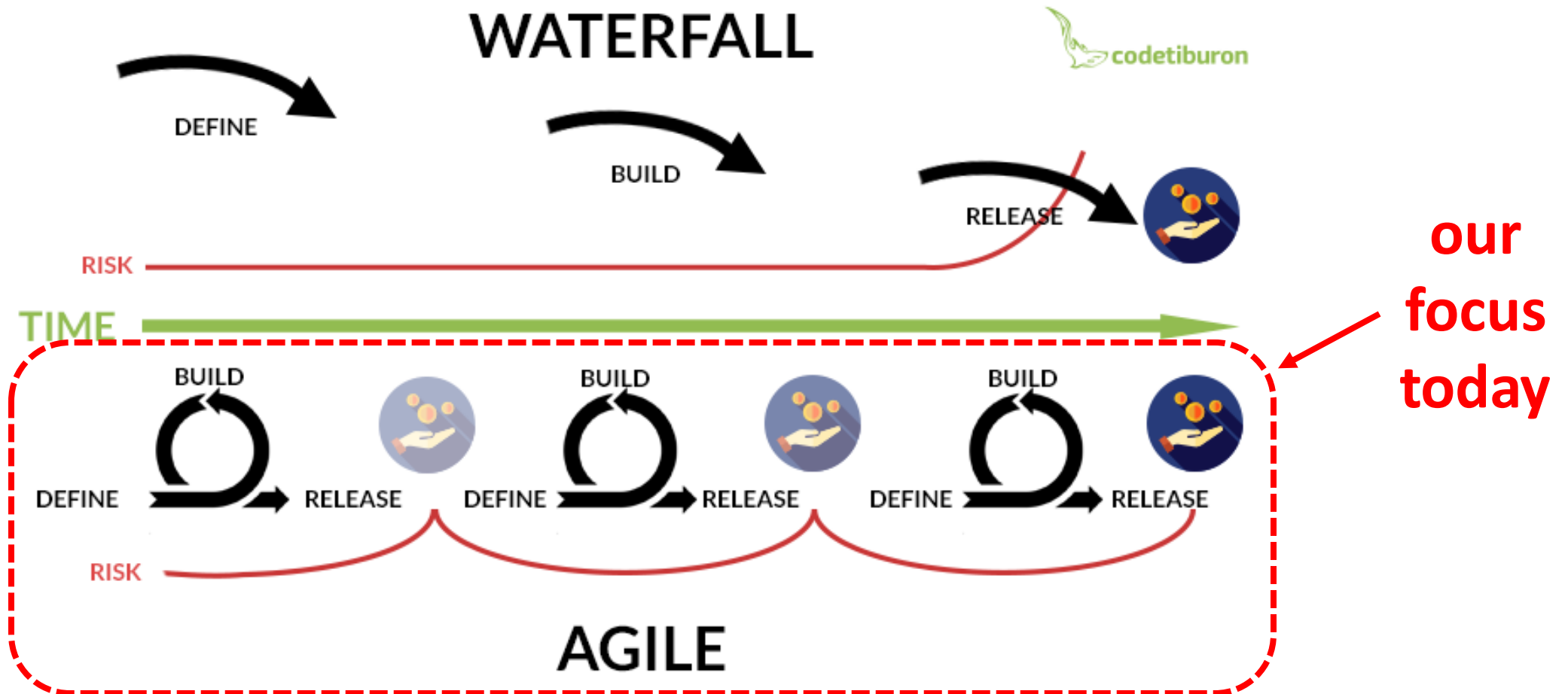


where completing a stage is a **prerequisite** to start the next one



aiming to release a **single final product**

a tale of two approaches...



for our purposes, **Agile** is



iterative approach to manage **development processes**



focused on **continuous releases**

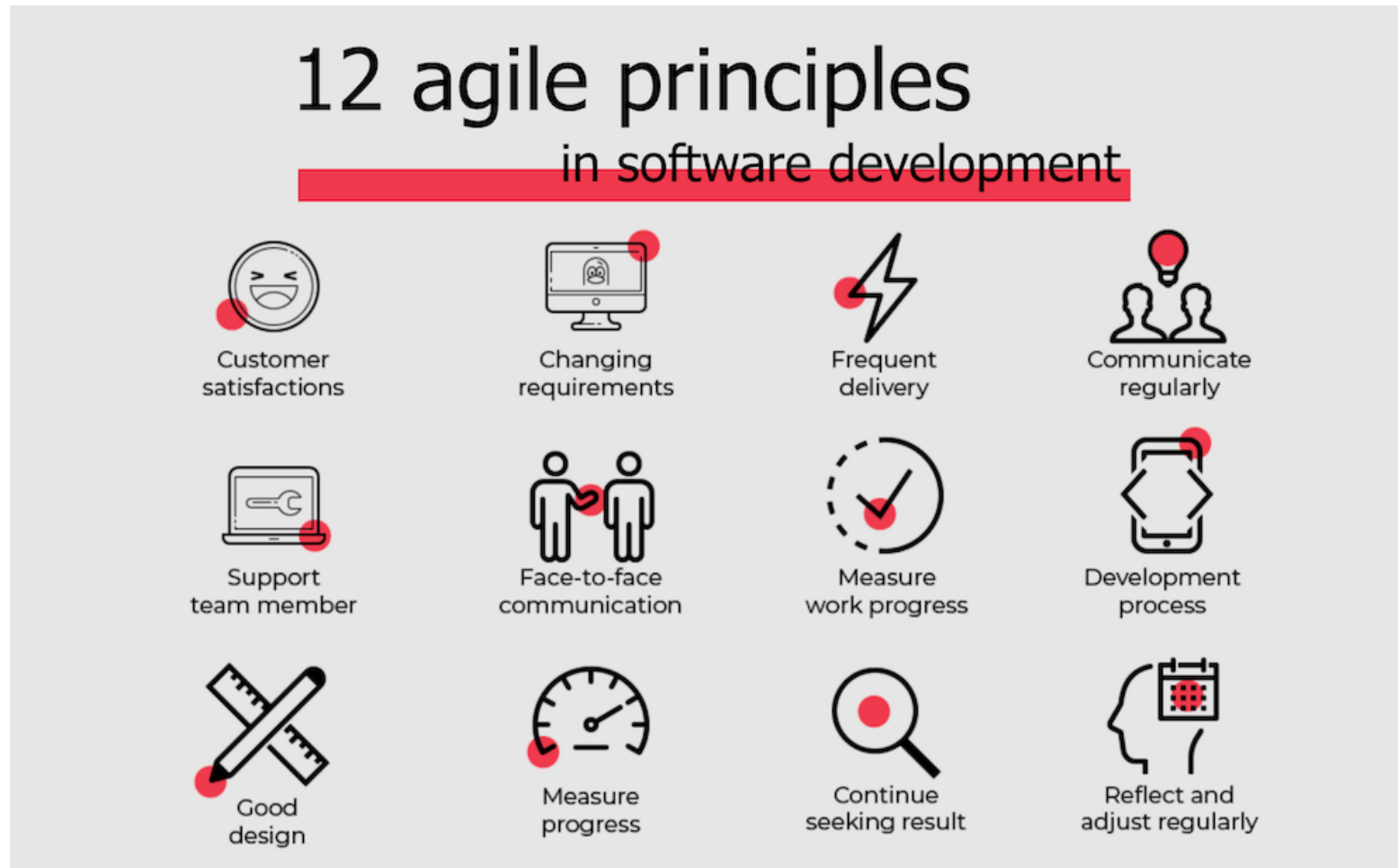


that incorporate **adjustments** at every iteration
based on **user feedback**



enabling **planning for unplanned work**

Agile as a **dynamic** and **user-centric** approach



with a little help from my (Agile) friends!



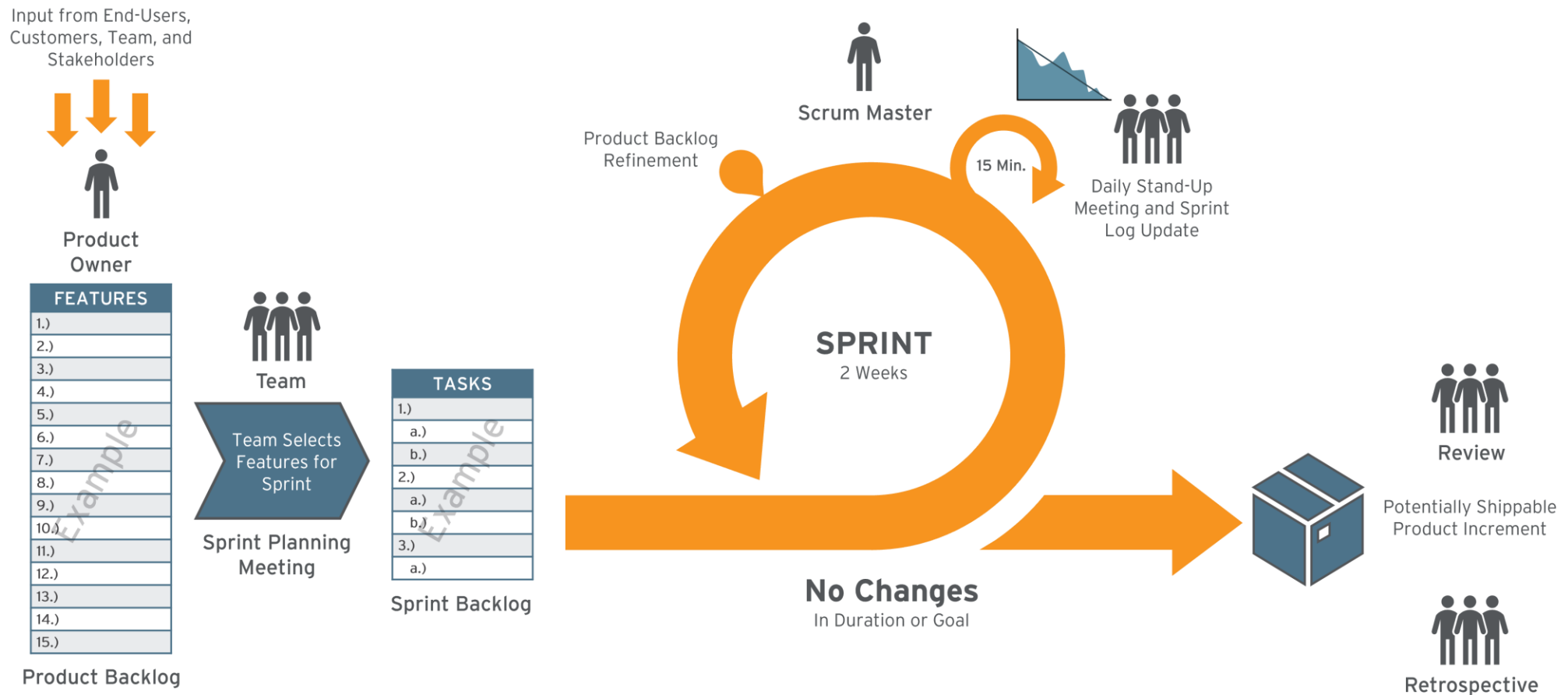
scrum



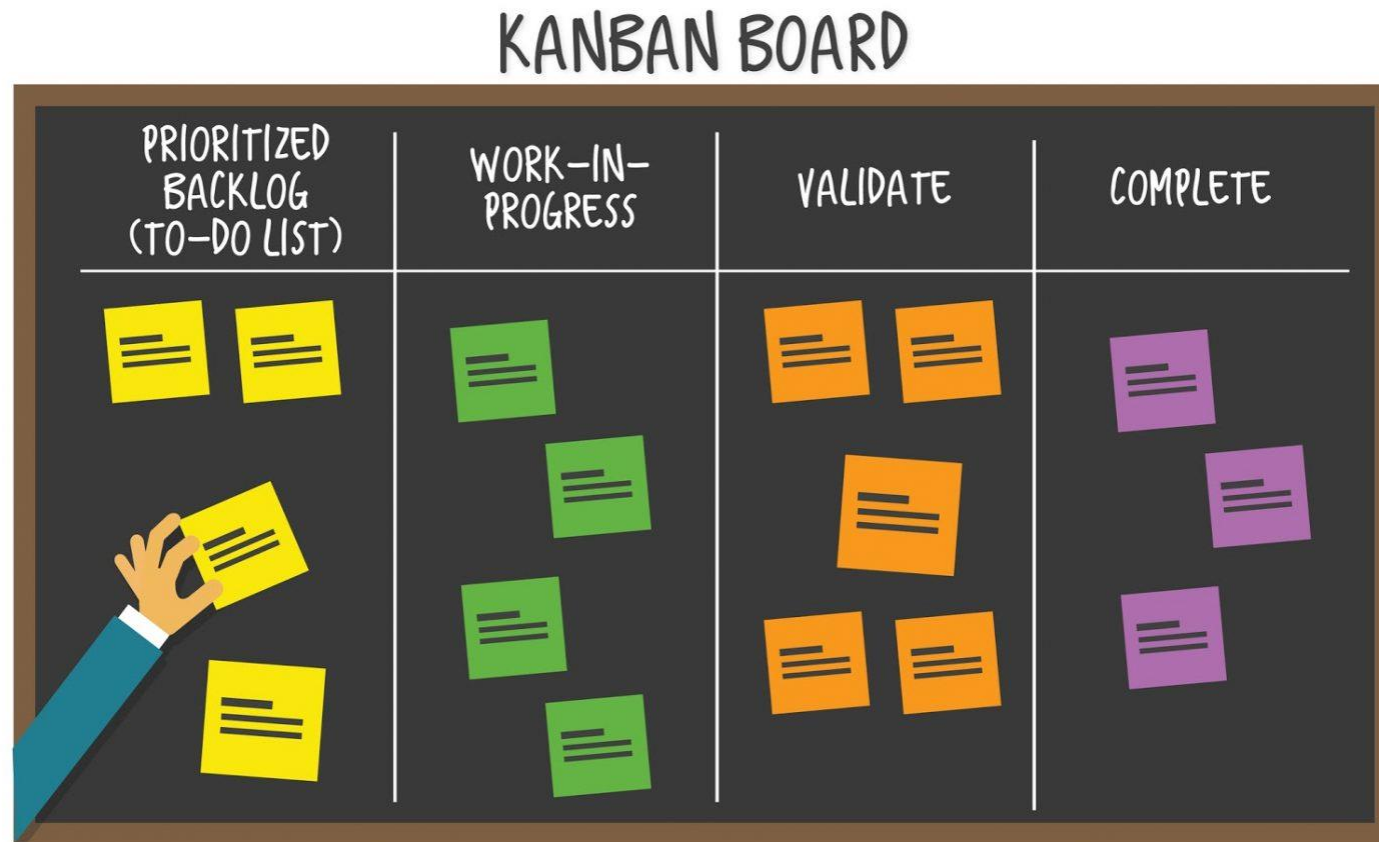
Kanban

two commonly used **methods** to **implement** the Agile philosophy to **build** and **evolve** digital solutions

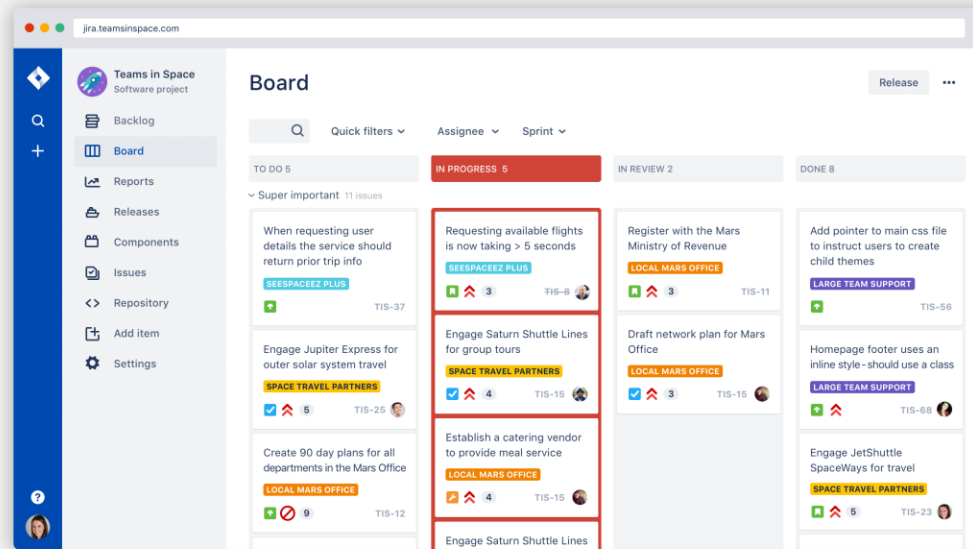
scrum as a structured method to be Agile



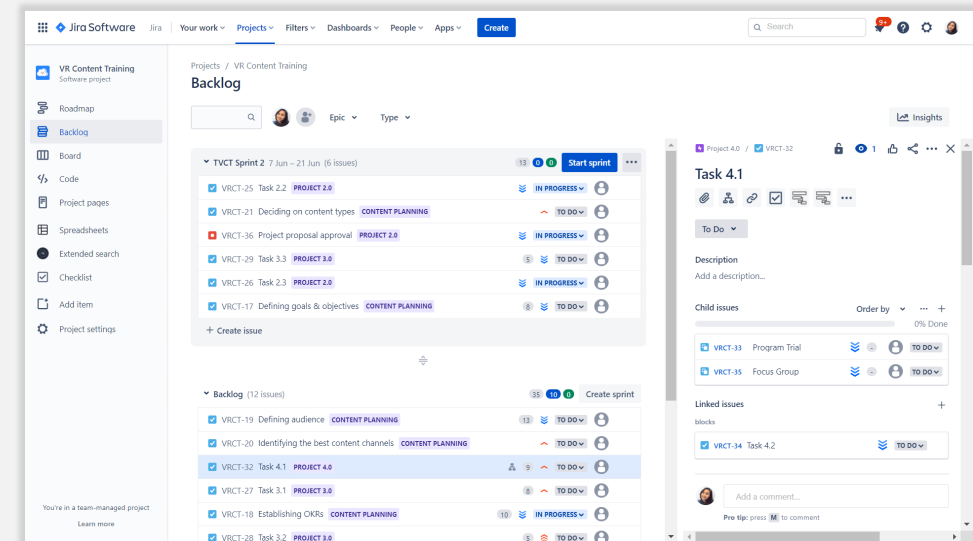
for everything else, there's **kanban** cards



plenty of (software) tools to be Agile



Kanban board



SCRUM backlog



a key concept to grasp: MVPs!



for our purposes, an **MVP** is



a **version** of a product



with just **enough features** to be usable

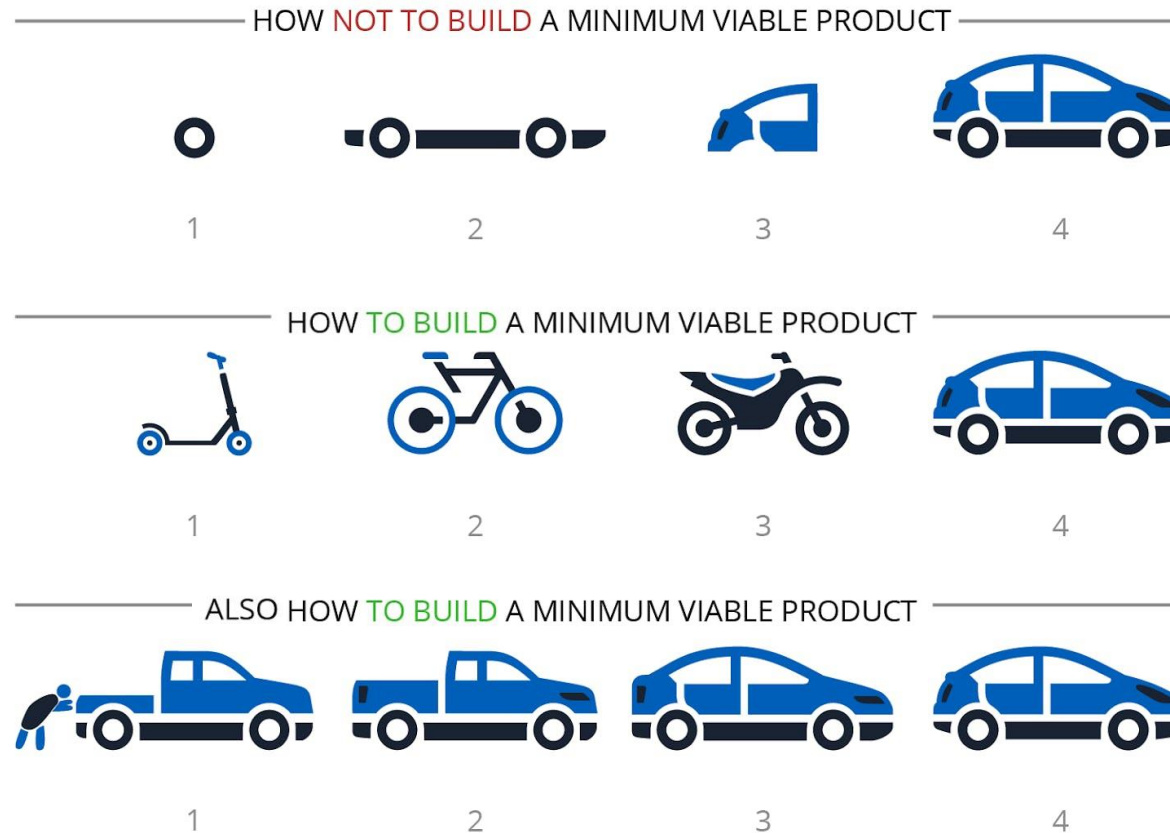


that enables collecting **feedback**



with **minimal effort**

important to clearly understand MVPs!



building MVPs has useful advantages



safe way to **validate** solution



early **user feedback**



lower **upfront costs**

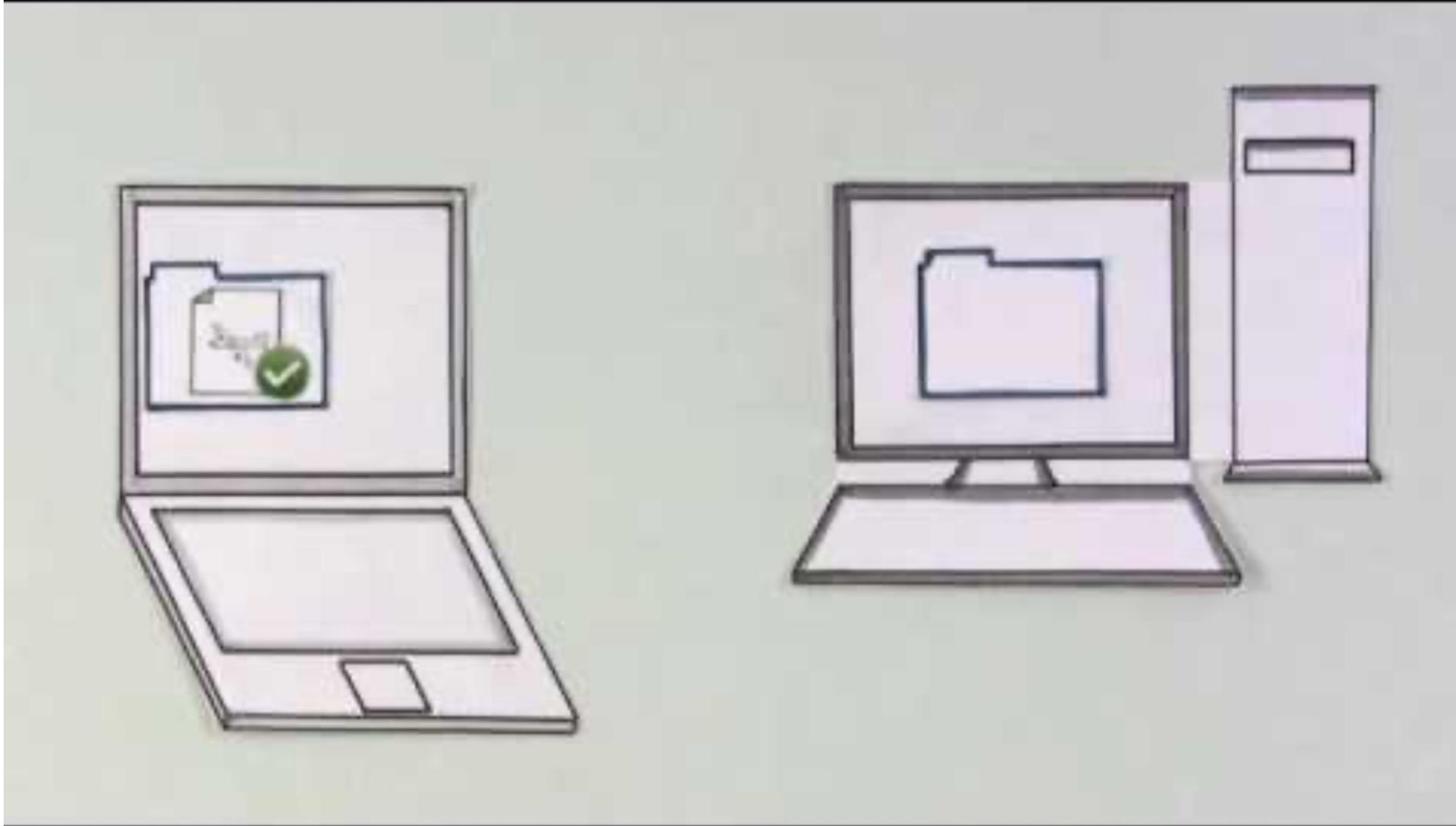


easier iterative development



faster **time-to-market***

from the hall of fame of MVPs: **Dropbox**



managing the process

Marco Morales

marco.morales@columbia.edu

Nana Yaw Essuman

ne2388@columbia.edu

GR5069: Applied Data Science
for Social Scientists

Spring 2025
Columbia University