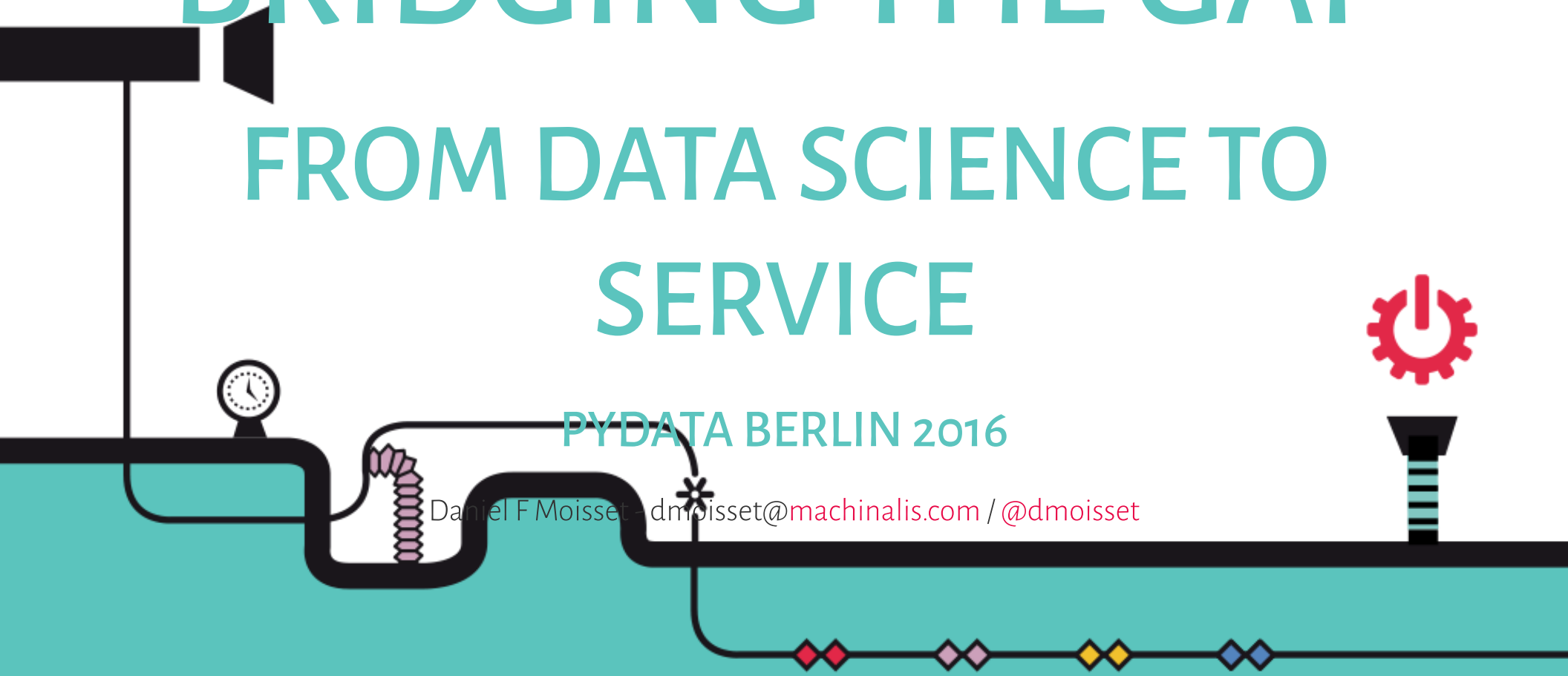
BRIDGING THE GAP

FROM DATA SCIENCE TO SERVICE



PYDATA BERLIN 2016

Daniel F Moisset - dmoisset@machinalis.com / @dmoisset

ABOUT ME

Hi! I'm Daniel Moisset!

My name is French. I'm from Argentina. I live in the UK.

I work at Machinalis

A special thanks to Marcos Spontón who also works there and inspired most of this talk.

WARNING: THIS IS NOT A TECH TALK!

In other words:

THIS TALK IS NOT ABOUT ALGORITHMS, MODELS, TOOLS, OR USE CASES

In event different words:

THIS TALK IS ABOUT PEOPLE

SO, RAISIN BREAD

By jeffreyw (Mmm...raisin bread) [CC BY 2.0], via Wikimedia Commons

Machine Learning development is like the raisins in a raisin bread… you need the bread first. But, it's just a few tiny raisins but without it you would just have plain bread

— I don't really know who, but I love the analogy

WHO WANTS RAISIN BREAD

Different organizations use your services:

1. Large companies with a live product and data, but without enough expertise/manpower in DS: «we'd like to add some raisins to our bread»
2. Small start-up, with maybe just a prototype, that want to get to production-ready scalable MVP: «We want some bread». And «it should have raisins now/at some point in the future»

IS THAT WHAT THEY ACTUALLY NEED?

"All the cool kids are doing it" is not good enough reason.

Raisin cookies that look like chocolate chip cookies are the main reason I have trust issues

— Seen on the internet

PART I: COMMUNICATING WITH THE CUSTOMER

IT'S NOT JUST SOFTWARE DEVELOPMENT!

It also has a heavy R&D component

- Higher uncertainty
- Results are probabilistic

THERE'S A PAPER ABOUT IT ≠ A PRODUCT

The distance may not be something coverable today.

MODELS ARE AN ASSET

- Investing time on it is not a “necessary evil”
- What's produced on a modelling phase is a critical component
- A model emerges from the client data and constraints, so it is unique to the client and an advantage over competitors.

MACHINE LEARNING ≠ CLAIRVOYANCE

- Garbage in, Garbage out
- The solution may not be clear; you may be unsure of what problem is more important; but your business goal should be clear. Data Science will not make it clear for you.

A PICTURE IS WORTH A THOUSAND WORDS

- Visualize your proposal.
- Be minimalistic.
- Use off the shelf tools for a proposal.

PART II: PROVIDING THE SERVICE

THE SERVICE IS THE END, DATA SCIENCE IS THE MEANS

Do not fall in love with the challenge

JUST OUT OF THE BOX MAY BE ENOUGH

You should always be asking:

1. Have I already covered the expectations?
2. Will an improved result here actually improve value?

MEASURE TWICE, CUT ONCE

Get a look at the object of analysis *before* starting work. Has it desirable qualities?

1. Manageable size?
2. It's in an accesible representation?
3. Does it have a reasonable distribution?
4. ...

INVOLVE THE PO

Validate your assumptions with a person familiar with your domain

1. Are there contradictions between your assumptions and their knowledge?
2. Are there contradictions between the data you already have and their knowledge?

Keep learning about the business side, encourage your business counterpart to learn to talk with Data Scientists.

PART OF YOUR SERVICE IS NOT DS

Make sure you use the right tools and people in each area

PART III: WORKING AS A TEAM

SHARE INFORMATION

Basic descriptive statistics should be shared with all involved, even the non DS. People in a team must be aware of what's important and what's not.

SHARE UNCERTAINTY

There are a lot of tradeoffs to make regarding milestones and deadlines. People can plan better (and have contingency plans) if they now what parts of the project have higher risks.

IT'S OK TO BUILD FLIMSY CODE, AS LONG AS IT'S NOT SOFTWARE

- code: programming text that runs on a computer
- software: programming text that is part of a deliverable.

There are differences:

- code does not necessarily need tests.
- code does not necessarily need to follow other processes.
- sometimes the outputs of your code is deliverable and may have to be treated specially.

THE DISCUSSION IS JUST BEGINNING

I'D LOVE TO HEAR ABOUT WHAT YOU'VE LEARNED ELSEWHERE

THANKS!

ANY QUESTIONS?

You can find me at twitter (@dmoisset) or by email (dmoisset@machinalis.com)