

Agile Business Science Framework - Create a ROI-Driven Data Science Practice

Favio Vázquez*

Data Science is an amazing field of research that is under active development both from the academia and the industry. One of the saddest facts in the data-world is that most data science projects in organizations fail. Here I'll present a new iteration of an agile framework called Business Science Problem Framework to implement data science in a way that enables decision making to follow a systematic process that connects the models you create to Return On Investment (ROI) and show the value that your improvements bring to the business.

The Problem Definition

Doing data science for business is not easy for several reasons. One of them is that most people have a partial definition, or description, of what data science actually is and what it means to be a good data scientist for solving real problems.

Because of that I'll start this article with my definition (or description) of how data science should be defined for a business:

Data science is the resolution to business problems through mathematics, programming and the scientific method that involves the creation of hypotheses, experiments and tests through the analysis of data and the generation of predictive models. It is responsible for transforming these problems into well-posed questions that can also respond to the initial hypothesis in a creative way finding the optimal threshold that maximizes the expected profit or savings. It must also include the effective communication of the results obtained and how the solution adds value to the business.

I'll explain my definition step by step below so stick with me.

Modeling is the process of understanding the "reality", the world around us, but creating a higher level prototype that will describe the things we are seeing, hearing and feeling, but it's a representative thing, not the "actual" or "real" thing. This is what we actually do in science and data science is no exception.

What I'm saying here is that data science is very much linked to the business, but it is a science in the end. A lot of people can disagree with me in the part that data science is a science. But keep your mind open and read this carefully. I think it could be very useful that we define data science as a science because if that's the case, every project in data science should be at least:

Reproducible: Necessary for making easy to test other's work and analysis.

- Fallible: Data Science and Science are not looking for the truth, they look for knowledge, so every project can be substituted or improved in the future, no solution is the ultimate. solution.
- Collaborative: The data scientist doesn't exist alone, he needs a team (even a virtual team, like collaborating in an open source project), this team will make things possible for creating intelligence and solutions. Collaboration is a big part of science, and data science should not be an exception.
- Creative: Most of what data scientists do is new research, new approaches or takes on different solutions, so their environment should be very creative and easy to work. Creativity is crucial in science, is

* Data Science Course Instructor, Business Science University, Pennsylvania Area.

the only way we can find solutions to hard and complex problems.

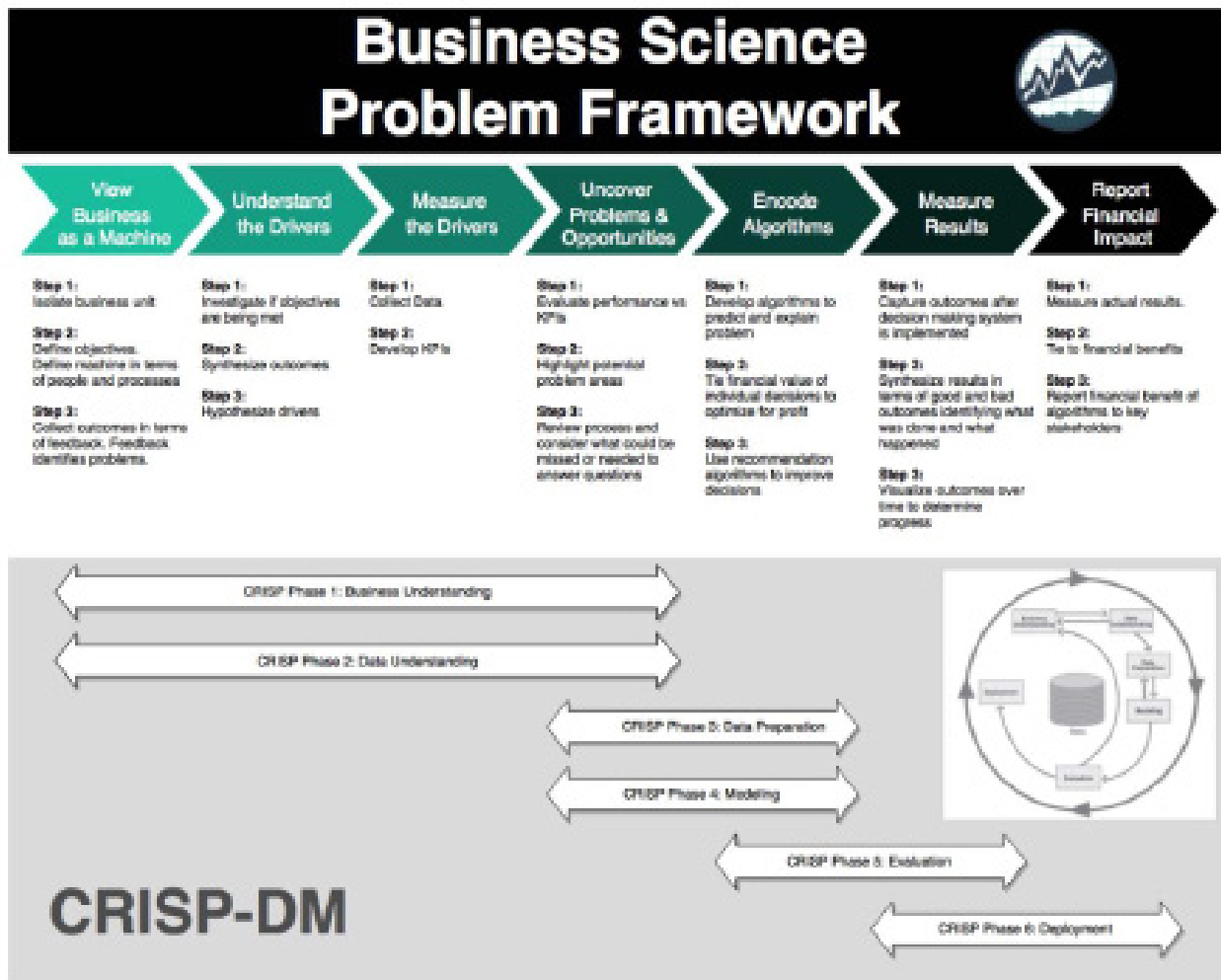
- Compliant to regulations: Right now there are a lot of regulations on science, not that much on data science, but there will be more in the future. Is important that the projects we are building can be aware of these different types of regulations so we can create a clean and acceptable solution to the problems.

If we don't follow those basic principles it would be very hard to conduct a proper data science practice. Data

science should be implemented in a way that enables decision making to follow a systematic process.

Data science isn't about software, knowing how to program, or being able to read data from different databases. Is about solving problems. An analogy would be saying that physics isn't about calculus, moving objects, algebra; it's about studying nature, understanding it and modeling it.

Business Science Problem Framework (BSPF)



<http://www.business-science.io/bspf.html> (Downloadable PDF)

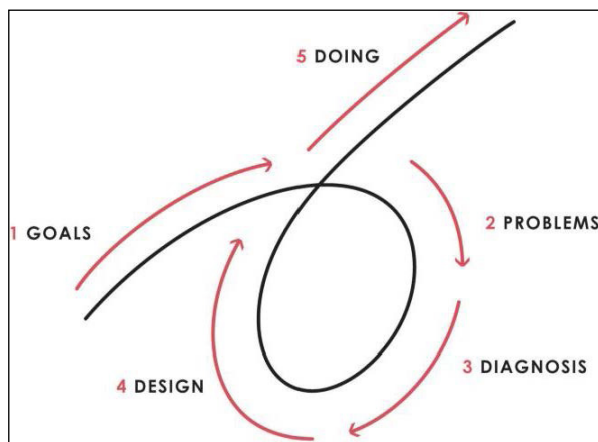
The Business Science Problem Framework is an agile data science process developed by Business Science. Business Science, founded by Matt Dancho, empowers

data scientists by providing software, education, training, and developing the areas of business & finance. It is a proven system conceived after years of work, trial and

error. Business Science say that clients loved the BSPF because it put a clear path forward and they loved it because it systemized their problem-solving method, making results more repeatable.

Many of the philosophies for creating the framework comes from the writings of Ray Dalio (author of Principles) along with Business Science’s experience using the BSPF with clients. Beyond, high level and detailed, it’s a proven framework.

Let’s start with the principles before going to the BSPF. Take a look at this image:



<http://jareddees.com/5-step-process-ray-dalio/>

This is one of the schemas that the author presents us. He tell us that this is the process of actually achieving your goals, and that failure is an important step in the journey. This is an iterative process, meaning that you will have to go through it over and over again; forever. You’ll always have new goals and you’ll also face new problems, but if you learn from them, by first recognizing them and creating a good design you can solve these problems and achieve your goals.

Agility in Data Science

If we go back to data science, and we apply this to solving business problems, the first thing we have to do is recognize that it has to be an iterative process. Agility is a word for doing that in the IT world, and that’s why I’m calling the BSPF an agile framework.

Agility is fundamental to business’ ability to successfully build systems in a world where it’s difficult to predict the future—James Kobielus.

We need agility to adapt. And if we want to go further, and beyond our common sense and intuition, we need to do it in a systematic way and then we can solve complex business problems.

The process of an Agile Data Science Workflow proposed by Russell Journey is an amazing way of understanding how and why data science together with agility helps us going beyond, seeing more and solving problems in a creative way.

We are used to jump to conclusions really fast, not analyzing every side of things. We are used to see what our eyes are seeing and “trusting our gut”.

Sadly, the common sense that reigns in our culture is Aristotelian and Medieval (Études d’histoire de la pensée scientifique—Alexandre Koyré). That means that intuition fails a lot of times when trying to understand the world, also this “common sense” comes sometimes with judgement, something that creates a bias in the way we see things.

Going and seeing beyond in this context means going a step forward, putting your judgment, common sense and intuition aside and really analyzing a situation. We should be doing this for everything that happens around us, question ourselves if the things you are doing, thinking and perceiving are actually correct. This is something very close to the Cartesian Doubt.

The manifesto for Agile Data Science (we should put agile data science workflow here) leads us to this. Iterating, over and over again, rethinking the business process and needs, experimenting a lot, listening what the data has to say, understanding and encouraging the business to understand that the data’s opinion must always be included in product discussions, finding a critical path to solve the problem and then organizing the team around completing it, and going further, letting the models solve the problems, of course using our expertise to help them, but not biasing them.

I need to emphasize that this is an agile framework, not that data science is being agile. This is following the words of Dave Thomas one of the creators of the Manifesto for Agile Software Development,

- You aren’t an agile programmer—you’re a programmer who programs with agility.

- You don't work on an agile team—your team exhibits agility.
- You don't use agile tools—you use tools that enhance your agility.

I'll add:

You are not an agile data scientist - you're a data scientist following a framework with agility.

So after studying the process proposed by Russell and Matt, I found a way of combining them, creating a system that will skyrocket your productivity as a data scientist and adding much more value. I realized that all of the steps that Russell proposed were already a part of the BSPF in some way, I'll make them all clear here.

Understanding The Agile Business Science Framework (ABSF)

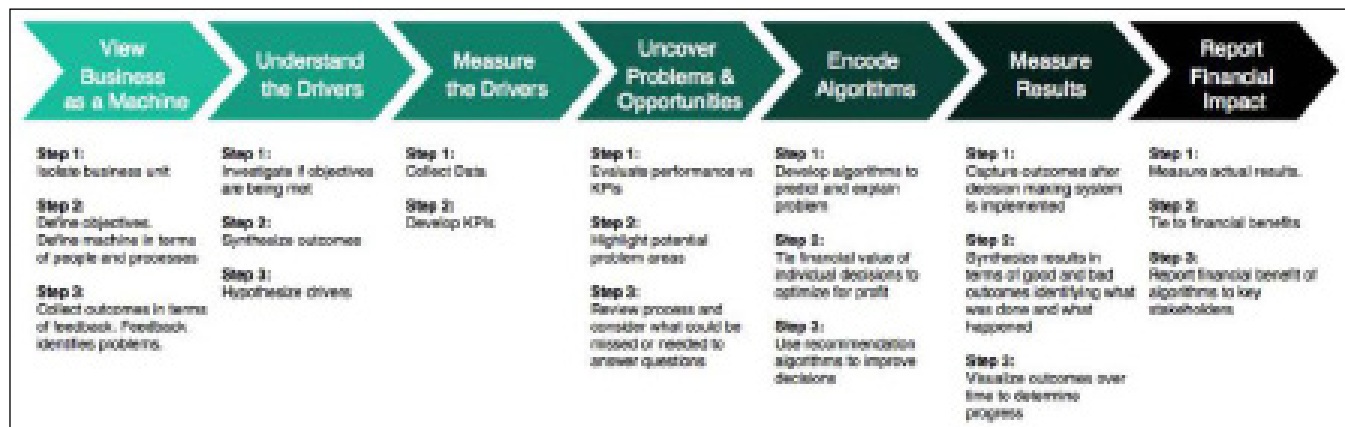
The BSPF is split into a top and bottom section. The top half contains details of what to investigate while the

bottom half contains high level stages of the project. The two sections are integrated, meaning they work together to provide a complete program for managing a data science project in a business context.

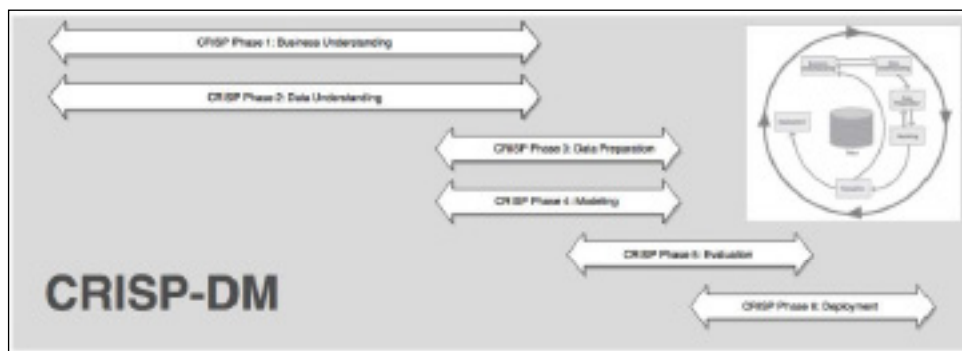
The BSPF has seven phases that are detailed with specific actions focused on understanding the problem and tying the results to Return On Investment (ROI), which is what the organization is keenly focused on:

- View The Business As A Machine
- Understand The Drivers
- Measure The Drivers
- Uncover Problems and Opportunities
- Encode Algorithms
- Measure Results
- Report Financial Impact

And in the bottom there are the six phases of CRISP-DM that are high-level steps for any data science problem:



- Business Understanding
- Data Understanding
- Data Preparation
- Modeling
- Evaluation
- Deployment

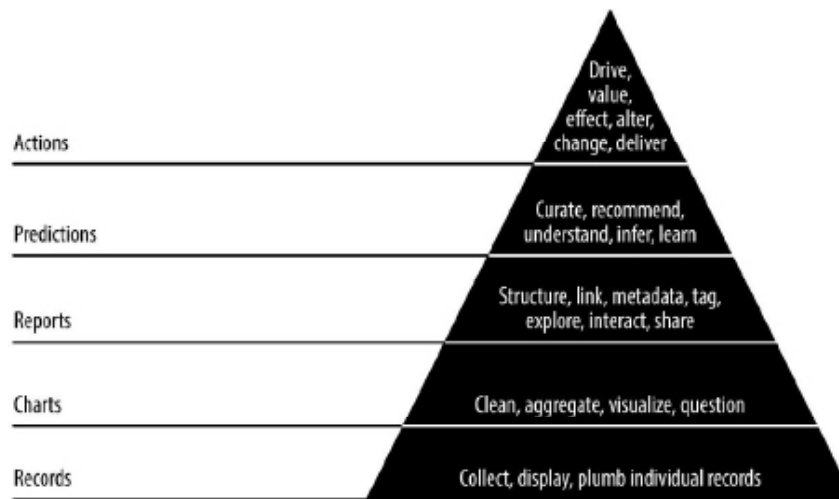


If you follow this agile methodology and framework is much more likely you'll succeed in your practice.

But what about some of the steps that Russell mention?

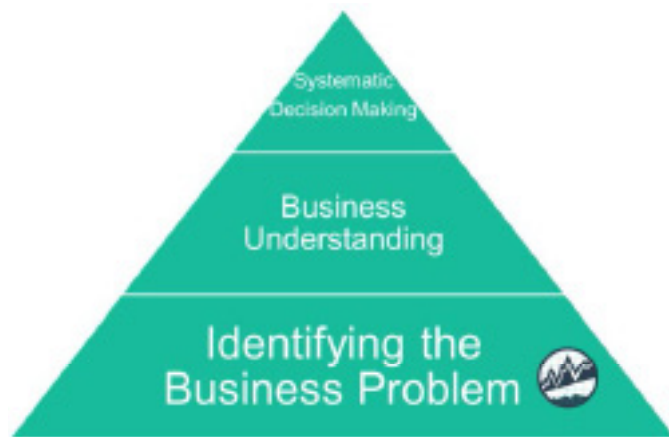
In his book, Russell talks about something called the “data-value pyramid”. It expresses the increasing amount

of value created when refining raw data into tables and charts, followed by reports, then predictions, all of which is intended to enable new actions or improve existing ones:



<https://www.oreilly.com/ideas/a-manifesto-for-agile-data-science>

But wait! Matt also created a pyramid in his methodology. The “Systematic Decision Making Pyramid”:



<http://www.business-science.io/business/2018/06/19/business-science-problem-framework.html>

Agile Data Science Workflow & BSPF Align

If we take a closer look of both pyramids they are actually saying the same things, and have the same hierarchy.

The data-value pyramid gives structure to our work. The pyramid is something to keep in mind, not a rule to be followed. Sometimes you skip steps, sometimes you work backward, making it an interactive process. And the systematic decision making pyramid tells us that we need

to understand the business. And, before we can understand the business, we need to identify the business problem to then being able to achieve systematic decision making, but it is also an interactive process, you’ll have to go back and forth sometimes.

Both frameworks are the same in one way or another, they are both proven systems that comes from years of work related to data, consulting, teaching and more. I say we need to fully understand both for enabling data science to create intelligence through AI. That’s a big sentence.

A Recipe For Artificial General Intelligence

The frameworks presented here are a piece of the story. The full scope of this data and ROI-driven mindset extends to the combination of big data, artificial intelligence, data science and the ABSF. The recipe to create intelligence is not that hard in a high level. This is what I propose we need to achieve it:

Artificial General Intelligence = AI + Big Data + Data Science + Agile Business Science Framework

I’m talking about Artificial General Intelligence (AGI) as the main goal of the data revolution. AGI are general-

purpose systems with intelligence comparable to that of the human mind (or maybe beyond humans). This would be the an amazing solution for solving a business problem.

We need Big Data as a Catalyst to get to AGI, because with more data, plus new ways of analyzing data, plus better software and hardware, we can create better models and better understanding. We need the current state of AI, very close to Deep Learning, Deep Reinforcement Learning and its surroundings for modeling the world, we need Data Science as the controller and science behind this problem solving machine and the Agile Business Science Framework that will enable us to adapt to changes and solve complex business problems in a systematic way.

But what about the ROI part? For this Matt points to the Expected Value Framework (EVF), a framework that connects the models you create to ROI. This framework is tied to the ABSF.

The Bottom Line

The bottom line here is that, in business, the costs associated with false positives and false negatives are rarely equal. In fact, in many cases false negatives are

much more costly. A false positive for your research will lead you to believe that your hypothesis is true, when in fact it isn't. And a with false negative you get a negative result, when you should have got a positive result.

The Agile Business Science Framework combined with the EVF allow us to find the optimal threshold that maximizes the expected profit or savings of the business problem. By iteratively calculating the savings generated at different thresholds, we can see which threshold optimizes the targeting approach and will also put our assumptions in check conducting sensitivity analysis, testing the effect of model assumptions on expected profit.

Next Steps

I joined Business Science a little while ago for helping create courses and content for their University. The first course so far, Data Science For Business (DS4B 201 / HR 201) Course has a student satisfaction rating of 9.1/10, and students are learning how to apply data science to business using R code, the Business Science Problem Framework, and more. I'm creating the Python counterpart course with the Agile Data Science Framework right now so if you are interested in that please let me know!