

Data Science and Physics today
*an industry perspective on the relevance, links and
history.*

Frederico Muñoz | Chief Architect | IBM Technical Expert Council (SPGI) *Data Science in (Astro)particle Physics and the bridge to industry* — LIP Lisbon | March 2018

Hello!

!

[285727?s=460&v=4] | <https://avatars0.githubusercontent.com/u/285727?s=460&v=4>

✉ [<frederico.munoz@pt.ibm.com>](mailto:frederico.munoz@pt.ibm.com)

in <https://www.linkedin.com/in/fsmunoz/>

🐦 <https://twitter.com/fredericomunoz>

🔗 <https://github.com/fsmunoz>

About the Technical Expert Council

The Technical Expert Council for Spain, Portugal, Greece and Israel is an affiliate of the IBM Academy of Technology.

[aotlogo 100x100] | https://researcher.watson.ibm.com/researcher/images/aotlogo_100x100.png

Data Science

... what is it good for?

!

(absolutely... something!)

But what it is?

!

plenty of definitions

!

At its core, data science is applying the scientific method to solve business problems.

— Seth Dobrin & Jean-François Puget (IBM)

!

You can further expand on the definition by understanding that we solve those **business problems** using **artificial intelligence** to create **predictions and prescriptions** and to **optimize processes**.

!

Analytics

- «The scientific process of transforming data into insight for making better decisions» (INFORMS)

The Analytics Maturity Model

- What happened? *Descriptive*
- Why did it happen? *Diagnostic*
- What will happen? *Predictive*
- What should I do? *Prescriptive*

!

Data Science includes analytics... but it is not (just) analytics.

So...what does it look like?

!

Let's hear Alix explaining what she does as a Data Scientist at IBM

!

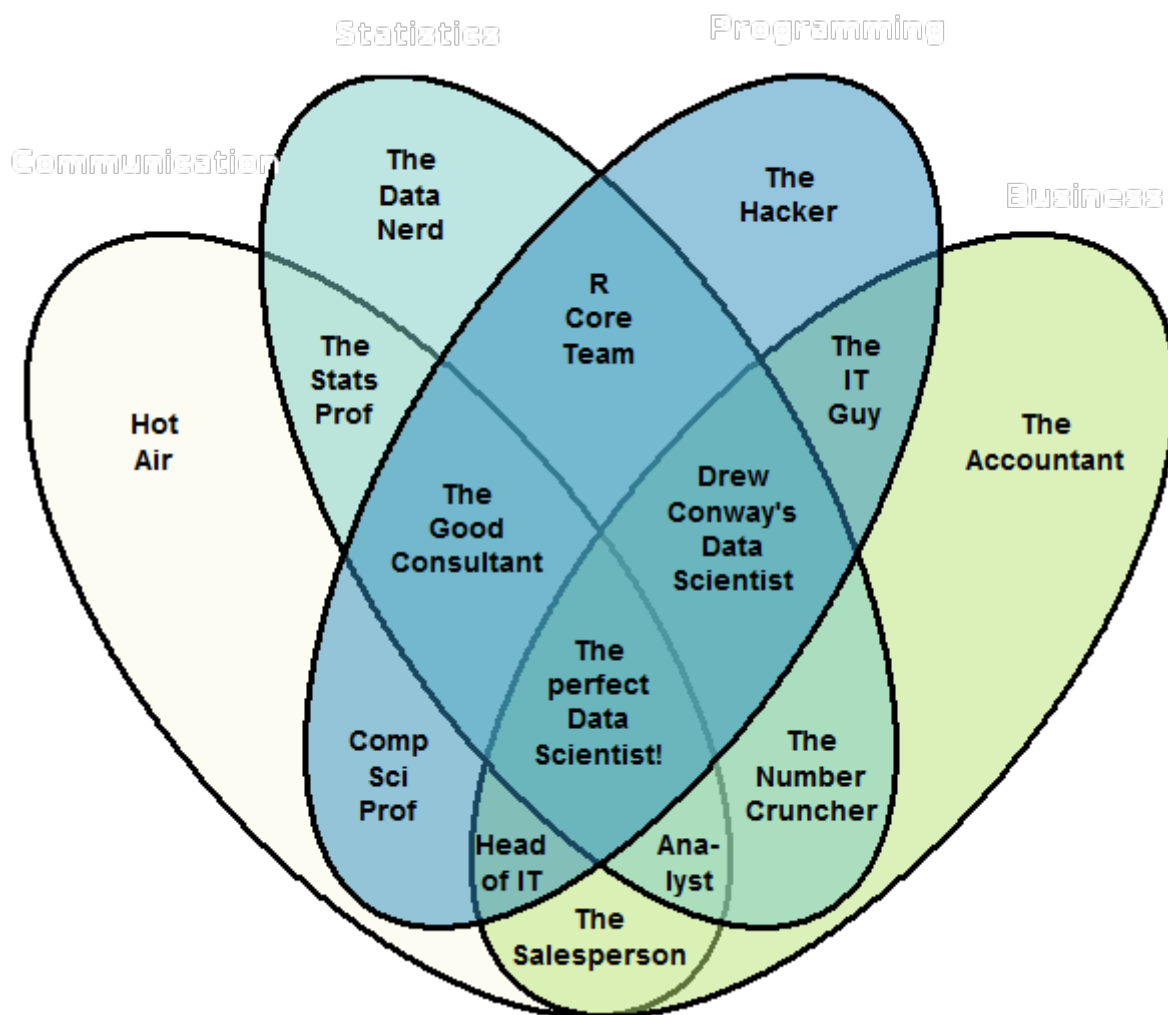
► <https://www.youtube.com/watch?v=7DnVAmbEBfk> (YouTube video)

Data Science

A new frontier in Analytics?

multiple domains, multiple skills

!



(source: Stephan Kolassa on StackExchange)

but... what about Physics?

Is a background in Physics helpful?

!

(spoiler warning: the answer is "yes" regardless of the question)

!

This is one "laundry list" of what IBM looks for in a Data Scientist

!

- Training as a scientist, with an MS or PhD
- Expertise in machine learning and statistics, with an emphasis on decision optimization
- Expertise in R, Python, or Scala
- Ability to transform and manage large data sets
- Proven ability to apply the skills above to real-world business problems
- Ability to evaluate model performance and tune it accordingly

A Physics background prepares for all of them

!

(do focus on the programming though)

!

Crucially, Data Science is about...

Data

...

Data

Lots of data

and making sense out of data

!

sometimes more data ...

...doesn't necessarily help

!

(especially when you don't know what to do with it)

But what's important is to keep asking.

!

I am just a child who has never grown up. I still keep asking these 'how' and 'why' questions. Occasionally, I find an answer.

— Stephen Hawking

!

*...Data science employs techniques and theories drawn from many fields within the broad areas of mathematics, statistics, chemometrics, information science, and computer science, including **signal processing, probability models, machine learning, statistical learning, data mining, database, data engineering, pattern recognition and learning, visualization, predictive analytics, uncertainty modeling, data warehousing, data compression, computer programming, artificial intelligence, and high performance computing...***

most of that should sound awfully familiar to Physics graduates.

which explains why IBM has such long relation with Physics.

...and Physics Engineering...

... and Artificial Intelligence...

...and Science in general.

Five IBM physicists have received the Nobel Prize in Physics

!

- Leo Esaki in 1973 for his work in semiconductors.
- Gerd Binnig and Heinrich Rohrer in 1986 for the scanning tunneling microscope.
- Georg Bednorz and Alex Mueller in 1987 for research in superconductivity.

And not just Nobel prizes

Many essential scientific breakthroughs were born from IBM Research through the decades.

Quantum tunneling

1958: Leo Esaki's discovery of the semiconductor junction, called the Esaki diode, finds wide use in electronics applications

Fractal Geometry

1967 - IBM, researcher **Benoît Mandelbrot**.

Nanotechnology

1981: Gerd Binnig and Heinrich Rohrer invent the scanning tunneling microscope, revolutionizing our ability to manipulate solid surfaces the size of atoms.

Quantum teleportation

1993: An international group of six scientists, including IBM Fellow Charles H. Bennett, confirmed the intuitions of the majority of science fiction writers by showing that perfect teleportation is indeed possible in principle, but only if the original is destroyed.

Deep Blue Chess

1997: First computer to defeat human World Chess Champion, Garry Kasparov.

IBM Blue Gene

2004: Supercomputer to observe protein folding and gene development.

... and many others

1947 Magnetic Core Memory 1957 Landauer Formalism - Conductance must come in Quantized Units 1958 Quantum Tunnelling 1960 Thin Film Heads 1966 Tunable Lasers 1966 Two-Dimensional Electron Gas (2DEG) 1967 Josephson Junctions 1968 DRAM - 1 Transistor RAM 1974 Dennard Scaling (aka Why Moore's Law also speeds up transistors in Lay Terms) 1978 Scanning Tunneling Microscope (1986 Nobel Prize Winner) 1982 Thermodynamics of Computation 1983 High Temperature Superconductors (1987 Nobel Prize Winner) 1990 Moving Atoms 1991 RFID 1993 Quantum Teleportation 1993 Seminal Contributions to the Theoretical Foundation of Quantum Information Processing 1994 High-Speed Silicon-Germanium Electronics 1997 GMR - Giant Magnetoresistive Heads 1998 Copper Interconnect 2002 SOI: Silicon on Insulator 2002 Theory of Nanoscale Material 2007 High-K Gate Dielectric 2008 Racetrack Memory 2008 Cooling 3D Chips 2011 Non-Planar Devices 2012 Holey Optochip - 1 Terabit per Second Optical Bus 2013 Millimeter Wave

more recently, and in the field of Artificial Intelligence

Watson Jeopardy! Challenge

!

- First computer to defeat TV game show Jeopardy! champions.
- Research teams are working to **adapt Watson to other information-intensive fields**, such as telecommunications, financial services and government.

From that we have built something special

Watson Services

!

- Visual Recognition: <https://visual-recognition-demo.ng.bluemix.net/>
- Conversation: <https://conversation-demo.ng.bluemix.net/>
- Speech to text: <https://speech-to-text-demo.ng.bluemix.net/>
- Natural Language Classifier: <https://natural-language-classifier-demo.ng.bluemix.net/>

!

- Natural Language Understanding: <https://natural-language-understanding-demo.ng.bluemix.net/>
- Personality Insights: <https://personality-insights-demo.ng.bluemix.net/>
- Tone Analyser: <https://tone-analyzer-demo.ng.bluemix.net/>

Watson Analytics

Watson Analytics is a smart data analysis and visualization service you can use to quickly *discover patterns and meaning* in your data – all on your own. With guided data discovery, automated predictive analytics and cognitive capabilities such as natural language dialogue, you can interact with data conversationally to get answers you understand.

<https://watson.analytics.ibmcloud.com/>

!

!

- Watson Explorer
- Watson for Oncology
- Watson for Genomics
- ...

a whole portfolio of Data Science, AI and Machine Learning solutions.

...integrated and working together and with researchers

IBM Watson Data Platform

!

- collaboration
- integration
- discovery
- openness
- full lifecycle

(from getting the data to showing it)

on premisses

!

IBM POWER8 and POWER9 processors with NVIDIA Tesla P100 GPU

«... close to ideal scaling with new distributed deep learning software which achieved record communication overhead and 95% scaling efficiency on the Caffe deep learning framework over 256 NVIDIA GPUs in 64 IBM Power systems...»

<https://www.ibm.com/blogs/research/2017/08/distributed-deep-learning/>

or in the IBM Cloud

!

- Data governance
- Data preparation
- Data analysis
- Model creation
- Building apps

!

!

!

We use this to build interesting things

!

an example

Adam Cox, Watson Data Platform, IBM.

"Former astro-particle experimental physicist (neutrino physics and direct dark matter detection). Experience in all aspects of constructing large hardware and software projects, such as system design and integration, hardware design, hardware programming, commissioning and calibration, software engineering, data management and statistical analysis. As a Developer Advocate I build interesting data science projects for citizen scientists utilizing IBM Cloud technologies."

SETI @ IBM Cloud

A citizen scientist project to apply deep learning to improve the state of the art in the search for extraterrestrial intelligence (SETI) research.

!

Deep Neural Networks have been trained to classify simulated radio-telescope signals with 95% accuracy.

<https://medium.com/ibm-watson-data-lab/using-artificial-intelligence-to-search-for-extraterrestrial-intelligence-ec19169e01af>

Cleaning noise it's a huge parte of Data Science

!

...Nyquist noise... Brownian motion... stochastic processes...

Rings a bell?

Data Scientist: The Sexiest Job of the 21st Century

LinkedIn

Jonathan Goldman, PhD in Physics from Stanford "...was intrigued by the linking he did see going on and by the richness of the user profiles. It all made for messy data and unwieldy analysis, but as he began exploring people's connections, he started to see possibilities. He began forming theories, testing hunches, and finding patterns that allowed him to predict whose networks a given profile would land in..."

<https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century>

We have many active projects in these and other fields

in fact, we have a Cognite Asset Factory

0

right next to you

!

wink wink, nod nod

Back to Physics...

!

and on a more personal note.

My own background has helped me in Data Science

!

- Analyse many different types of data sources in search for relations
- Observe social relations and apply scientific principles to the subject matter
- Discover new material evidence and cross-check it with the existing scientific consensus
- Select appropriate models depending on the data and the goals
- Critically review the discovery process and suggest improvements

!

Nothing?

!

!

... but when faced with being limited in certain areas...

!

I don't know. It was in some way my life, you see.

— Niels Bohr

**We've been here for a while
and helped to achieve some rather
important things**

some (extremely) big
some (extremely) small
but always shaping the future
Quantum Computing

!

In the summer of 1981, IBM and MIT organized a landmark event called the First Conference on the Physics of Computation.

It took place at Endicott House, a French-style mansion not far from the MIT campus.

!

Bennett and others realized that some kinds of computations that are exponentially time consuming, or even impossible, could be efficiently performed with the help of quantum phenomena. A quantum computer would store information in quantum bits, or qubits.

!

Nature is quantum, goddamn it! So if we want to simulate it, we need a quantum computer!

— Richard Feynman

!

► <https://www.youtube.com/watch?v=o-FyH2A7Ed0> (YouTube video)

IBM Q

An industry-first initiative to build commercially available universal quantum computers for business and science.

!

- 20 qubit available, 50 qubit developed

- QISkit: open to **anyone** for development.
- IBM Q Network: advancing quantum computing together

These are all reasons behind IBM's motto

THINK

Thank you!

Some final links

Community

Tools, algorithms and approaches are increasingly more open and social.

- Cognitive Class: Build Data Science and Cognitive Computing skills *for free* today <https://cognitiveclass.ai>
- Data Scientist Workbench: virtual lab with Data Science tools ready to explore and put to use <https://datascientistworkbench.com/>

!

- Data Science Experience: Learn, create and collaborate <https://datascience.ibm.com/>
- IBM Code: Code patterns, tech talks, open source projects, developer advocates, dynamic communities, upcoming events. <https://developer.ibm.com/code/>
- IBM Cloud: integrate all IBM services with your solution, including Watson & Analytics services <https://console.bluemix.net/catalog/>

References

- Fractal animation by Alexandre Tavernier (<http://coolfractalanimations.blogspot.pt/2014/05/mandelbrot-set-animation-color-changes.html>)
- IBM AI Research: https://researcher.watson.ibm.com/researcher/view_page.php?id=6813
- Neurons image by Michelle Kuykendal and Gareth Givanasen (Georgia Tech's NeuroLab): <https://www.youtube.com/watch?v=yy994HpFudc>
- Spring cloud time lapse by Harrison Rowntree (<https://www.youtube.com/watch?v=Qu7mcKZgqv0>)