

The zero to one of a data team

First 18 months in Prisa Media Data & Analytics

INMA - november, 2022



The First 18 Months of Data & Analytics Department in Prisa Media



EL PAÍS
EL PERIÓDICO GLOBAL

as **CincoDías**

SER ...



José Meroño

**Data & Analytics Director
(2019-2022)**

Independent Consultant
jmmerono@gmail.com

2019 Q2	2019 Q3	2019 Q4	2020 Q1	2020 Q2	2020 Q3	2020 Q4	2021 ...
---------	---------	---------	---------	---------	---------	---------	----------



We started with a Vision for the Data division - Why we are creating this division? and we had three main strategic objectives

Be at the base of the subscription model creation

Be an active part of advertising marketing based on data

Democratize information in the company, promoting self-consumption

2019 Q2

2019 Q3

2019 Q4

2020 Q1

2020 Q2

2020 Q3

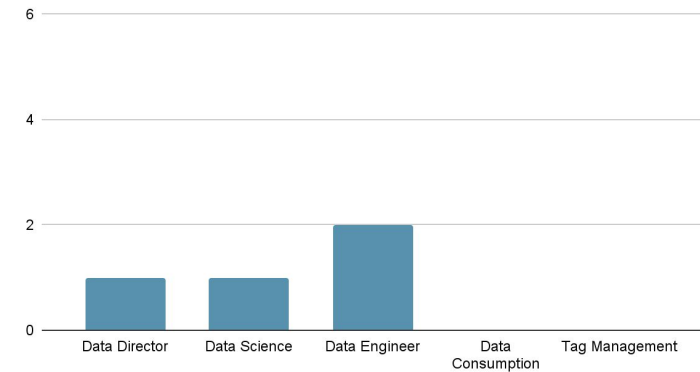
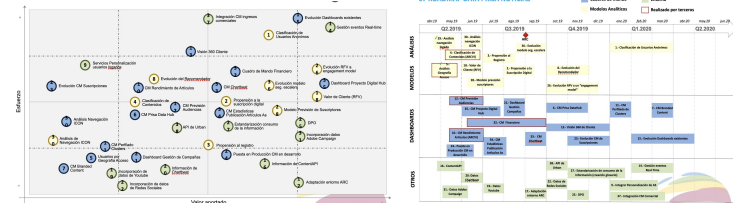
2020 Q4

2021 ...

With the vision in mind, we launched data & analytics area, performing an analysis of the current situation and establishing the desired future situation

- AS-IS Analysis
 - Strong data team (1 internal + 2 external) whose great developments were not sufficiently used by the business
 - Some established data processes and tactical initiatives (8 dashboards and advanced analytics)
 - Departmental and limited usage of the data. Lack of visibility of the analysis performed
 - Lack of corporate data environment
 - Lack of visibility for future initiatives
- TO-BE Analysis
 - Establishment of the desired data architecture
- First contact with business areas and creation of matrix of initiatives by value and effort and prioritization of initiatives (roadmap creation)

Data & Analytics Team - 4 people



2019 Q2

2019 Q3

2019 Q4

2020 Q1

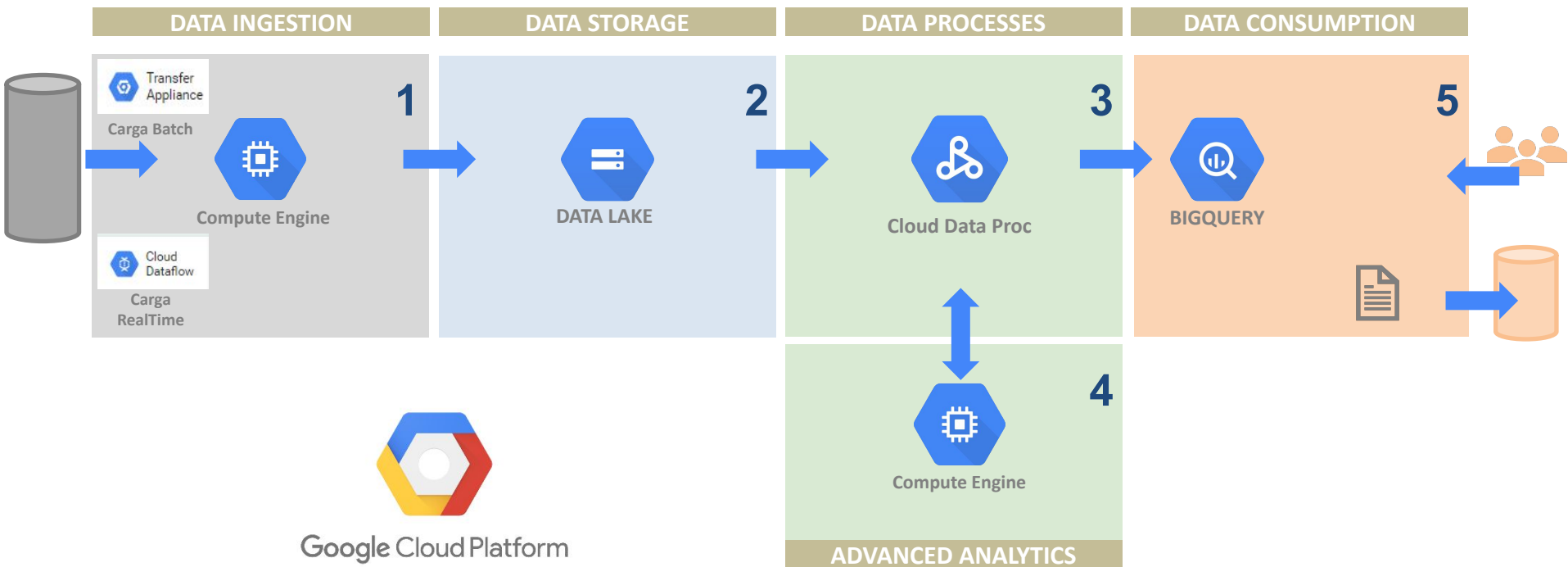
2020 Q2

2020 Q3

2020 Q4

2021 ...

Data architecture composed of 5 functional areas was established that was going to be deployed in Google Cloud Platform, continuing with what has already been done



2019 Q2

2019 Q3

2019 Q4

2020 Q1

2020 Q2

2020 Q3

2020 Q4

2021 ...

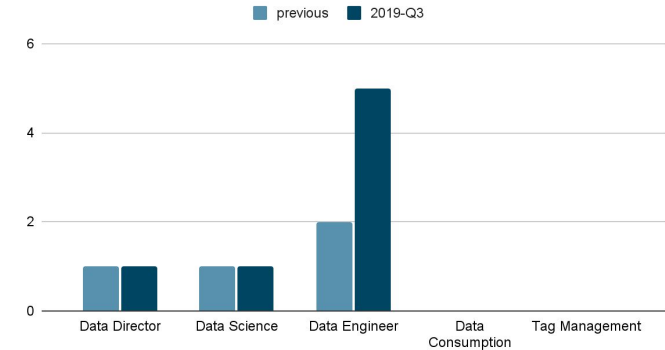


We began to establish the internal processes. The team began to increase, initially focusing on data ingestion and storage and a common data dictionary began to be defined

TEAM & PROCESSES BUILDING

- Two internal data engineers and one external data engineer / data architect were added to the team
- A common data dictionary begins to be defined with the business having clear and global kpis, dimensions and data relations rules
- Data committee with key stakeholders (IT, business, etc.) and other internal processes established to review initiatives delivered and prioritize the new ones
- Data architecture decisions (No ETL and use spark/scala instead, storage raw data in parquet)
- Source systems integration (CRM, campaign, articles production, social networks profiles, etc.) and launch of first global reports with navigation and logged navigation data by geography

Data & Analytics Team - 7 people



2019 Q2

2019 Q3

2019 Q4

2020 Q1

2020 Q2

2020 Q3

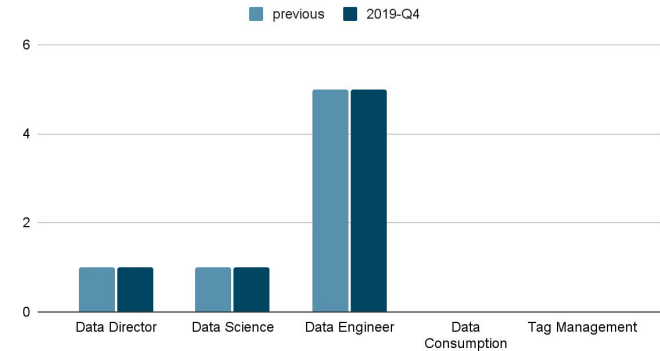
2020 Q4

2021 ...

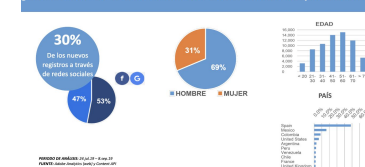
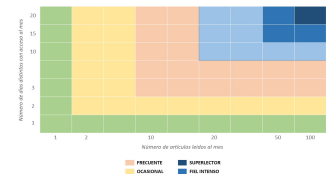
After 6 months, the first focus was to establish a data lake with a governed data loading processes. The main source systems (digital navigation, content/articles, CRM and campaign) were integrated

- Governed environment: Joint work with IT to manage data loads, schedule, jobs performance, etc.
- First data quality issues appeared and data conflicts with other data providers as they don't use same rules
- Visualization proof of concept to consume the data in real time and to share dashboards with third parties - Looker is selected as best option in our environment
- Analysis of impact of different model subscriptions scenarios (subscribers, readers and page views)
- Consumption and therefore, the cost, begins to pick up
- Automation and improvement of the efficiency of data loads and transformation processes
- Start working with advertising (second objective of the vision)

Data & Analytics Team - 7 people



¿Qué resultados hemos obtenido?: Conocimiento del perfil social



2019 Q2

2019 Q3

2019 Q4

2020 Q1

2020 Q2

2020 Q3

2020 Q4

2021 ...

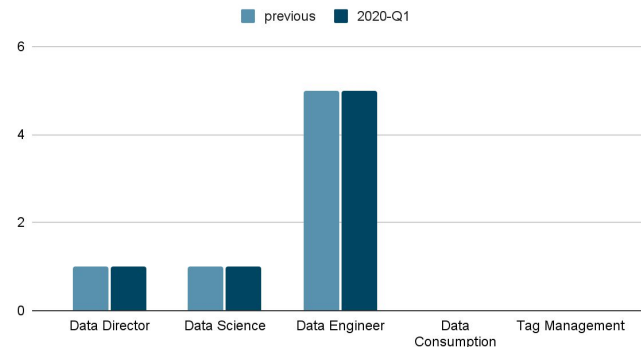


With the Data Lake in place, it was time to stabilize and prepare it to be exploited on a large scale by business users consumption and by advanced analytical models

VISUALIZATION TOOL - DATA DICTIONARY

- Load and transformation processes redesigned in order to make them more efficient
- Real time load included for key events
- Data dictionary available for business in the visualisation tool (to facilitate self-consumption)
- Digital subscription model launched and a dashboard in real time with digital subscription kpis in newsroom gave visibility to the area
- Semantic layer between the data and visualisation created - single point of business rules
- Dashboard migration to the new visualization tool
- Access to dashboard were re-designed by functions and areas, creating roles and profiles

Data & Analytics Team - 7 people



2019 Q2

2019 Q3

2019 Q4

2020 Q1

2020 Q2

2020 Q3

2020 Q4

2021 ...

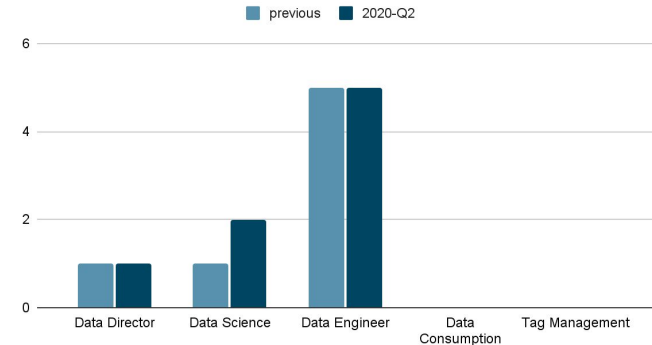


The robust Data Lake and the new visualization tool allowed us to quickly create hundred of dashboards and dozens of analytical models to be used for the business users

DATA CONSUMPTION EXPLOSION

- Explosion of new dashboards, reaching quickly +100 dashboards: subscription, advertising, customer, product, marketing, newsroom, etc.
- Advanced Analytic team reinforced with one senior data science. A new data scientist (the future head of advanced analytics) was hired
- New advanced analytical models created: propensity to subscription and churn models
- New engagement index for readers created. It was based on recency, frequency and value approach
- The number of source systems, processes and dashboards was increased by adding new analyzes requested by the areas

Data & Analytics Team - 8 people



2019 Q2

2019 Q3

2019 Q4

2020 Q1

2020 Q2

2020 Q3

2020 Q4

2021 ...

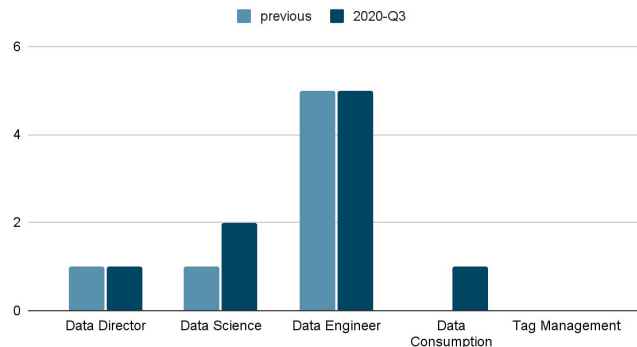


With hundreds of dashboards and dozens business users accessing data, the data consumption team was reinforced and cost due to data consumption was increased

COST INCREASE - CONTROL OF SUCCESS

- New data consumption admin improved the dashboards and made the queries more efficient
- The higher data consumption translated into a higher cost associated with its storage and processing
- Some required analysis were not used much and they were eliminated
- Very good results when using the data in selection of target audience in subscription campaign - advanced analytics boost
- The environment was maturing
 - Customer 360 vision were created for every reader: subscribers, logged users and anonymous users
 - Customer profile included in some analytics
 - Advanced models reinforced with new data

Data & Analytics Team - 10 people



2019 Q2

2019 Q3

2019 Q4

2020 Q1

2020 Q2

2020 Q3

2020 Q4

2021 ...

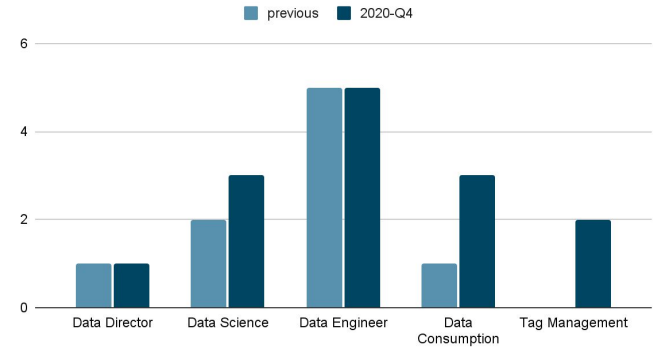


With Data division already consolidated, two existing teams were included in this area: audience & reporting and tagging management

DATA FUNCTION INTEGRATION

- The audience and reporting team joined the visualisation team and double reporting and different figures problems were avoided. Since then, a unique data dictionary were used
- The advanced analytics were reinforced with a new junior data science. (Despite the high turnover of these profiles in this period, only one person left the project)
- The total cost was stabilized by end of this quarter
- New tag management designed to be used for the whole company
- Official reporting migrated to the visualisation tool
- Mid-term roadmap for data evolution defined

Data & Analytics Team - 14 people



- # TO INFINITY AND BEYOND

```

graph TD
    NoemiRamirez[Noemi Ramirez  
CCO] --- JoseMeroño[José Meroño  
Director Data]
    JoseMeroño --- PabloMartin[Pablo Martín  
Data Architect]
    JoseMeroño --- IreneGervasio[Irene Gervasio  
Aspirante]
    JoseMeroño --- RodrigoVelcampl[Rodrigo Velcampl  
Marketing]
    JoseMeroño --- AniaBenitez[Ania Benítez  
Data Scientist]
    JoseMeroño --- NoemPozo[Noem Pozo  
Aspirante]

    PabloMartin --- DataEngineer1[Data Engineer]
    PabloMartin --- DataEngineer2[Data Engineer]
    PabloMartin --- DataEngineer3[Data Engineer]

    IreneGervasio --- DataArchitect1[Data Architect]
    IreneGervasio --- DataArchitect2[Data Architect]

    RodrigoVelcampl --- DataEngineer4[Data Engineer]
    RodrigoVelcampl --- DataEngineer5[Data Engineer]

    AniaBenitez --- DataScientist1[Data Scientist]
    AniaBenitez --- DataScientist2[Data Scientist]

    NoemPozo --- DataEngineer6[Data Engineer]
    NoemPozo --- DataEngineer7[Data Engineer]
  
```

2019 Q2	2019 Q3	2019 Q4	2020 Q1	2020 Q2	2020 Q3	2020 Q4	2021 ...
---------	---------	---------	---------	---------	---------	---------	----------



Some figures about we have achieve in this period

- **25** source systems
- **600TB** storage data (**+130TB** available in big query)
- **+1200** daily queries, **+44TB** processed
- **+100** business users, **57%** active in the last 7 days
- **+250** dashboards, **77%** accessed in the last 30 days
- **82%** business queries in dashboards **< 10 seconds**

2019 Q2	2019 Q3	2019 Q4	2020 Q1	2020 Q2	2020 Q3	2020 Q4	2021 ...
---------	---------	---------	---------	---------	---------	---------	----------



Lessons Learned

What we did well

- **Think big, start small.** Establish the target solution and data architecture and build it according to the needs
- **Try early, fail fast.** Deliver data and insights to the business even if the solution is not complete
- **Share everything.** Share to different departments data and insights delivered

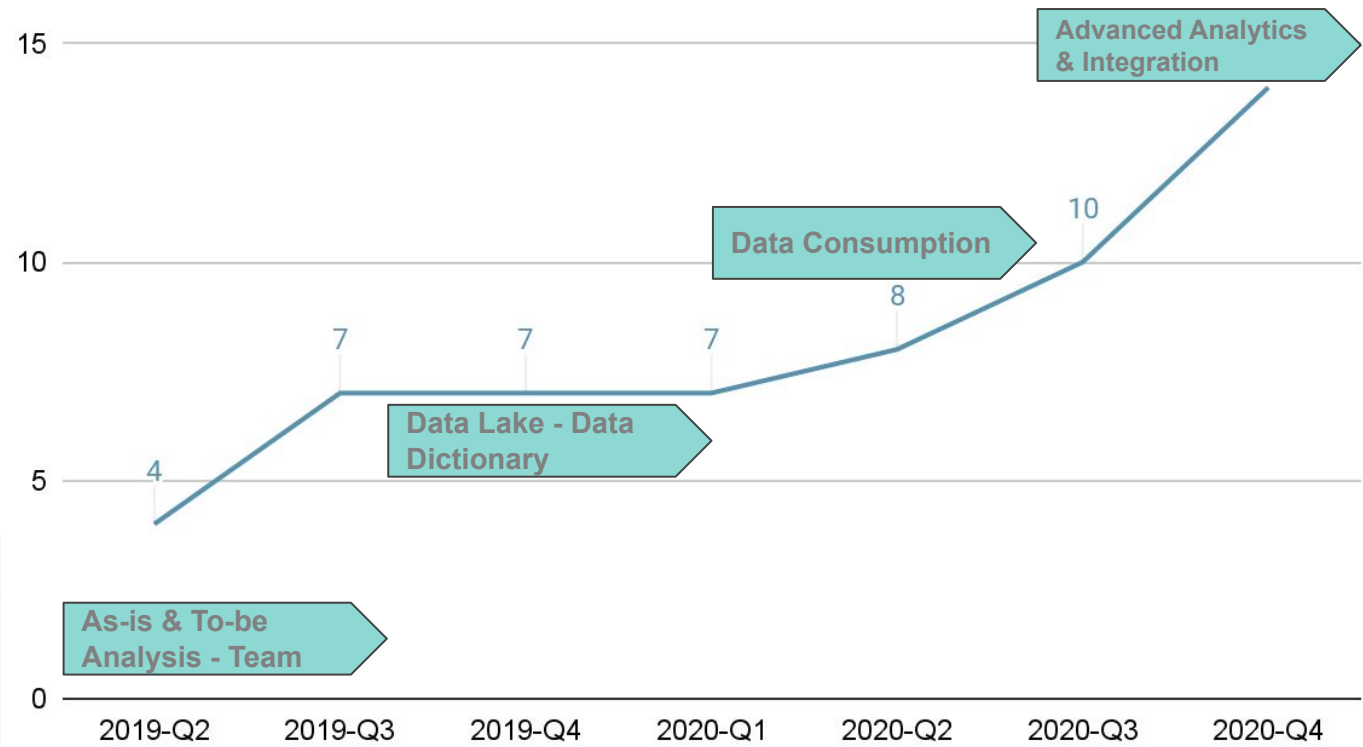
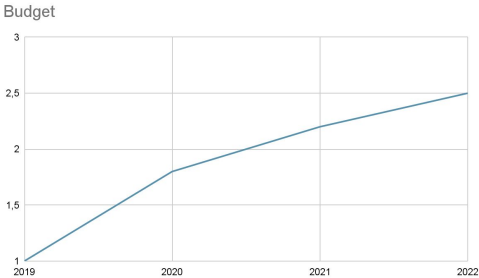
What we did not
so well

- **Beware of the cost.** Always keep an eye on costs
- **One data to rule them all.** Every team or person providing insights and results would be in continuous communication with data department

2019 Q2	2019 Q3	2019 Q4	2020 Q1	2020 Q2	2020 Q3	2020 Q4	2021 ...
---------	---------	---------	---------	---------	---------	---------	----------



Headcounts Data Team





Thank you all

Now Q&A