



# Data Analytics

## Grado en Matemáticas e Informática

Ernestina Menasalvas Ruiz  
Alejandro Rodríguez González  
Antonio Díaz Honrrubia



# Data Scientist

## MODERN DATA SCIENTIST

Data Scientist, the sexiest job of the 21st century, requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

### MATH & STATISTICS

- ☆ Machine learning
- ☆ Statistical modeling
- ☆ Experiment design
- ☆ Bayesian inference
- ☆ Supervised learning: decision trees, random forests, logistic regression
- ☆ Unsupervised learning: clustering, dimensionality reduction
- ☆ Optimization: gradient descent and variants

### DOMAIN KNOWLEDGE & SOFT SKILLS

- ☆ Passionate about the business
- ☆ Curious about data
- ☆ Influence without authority
- ☆ Hacker mindset
- ☆ Problem solver
- ☆ Strategic, proactive, creative, innovative and collaborative



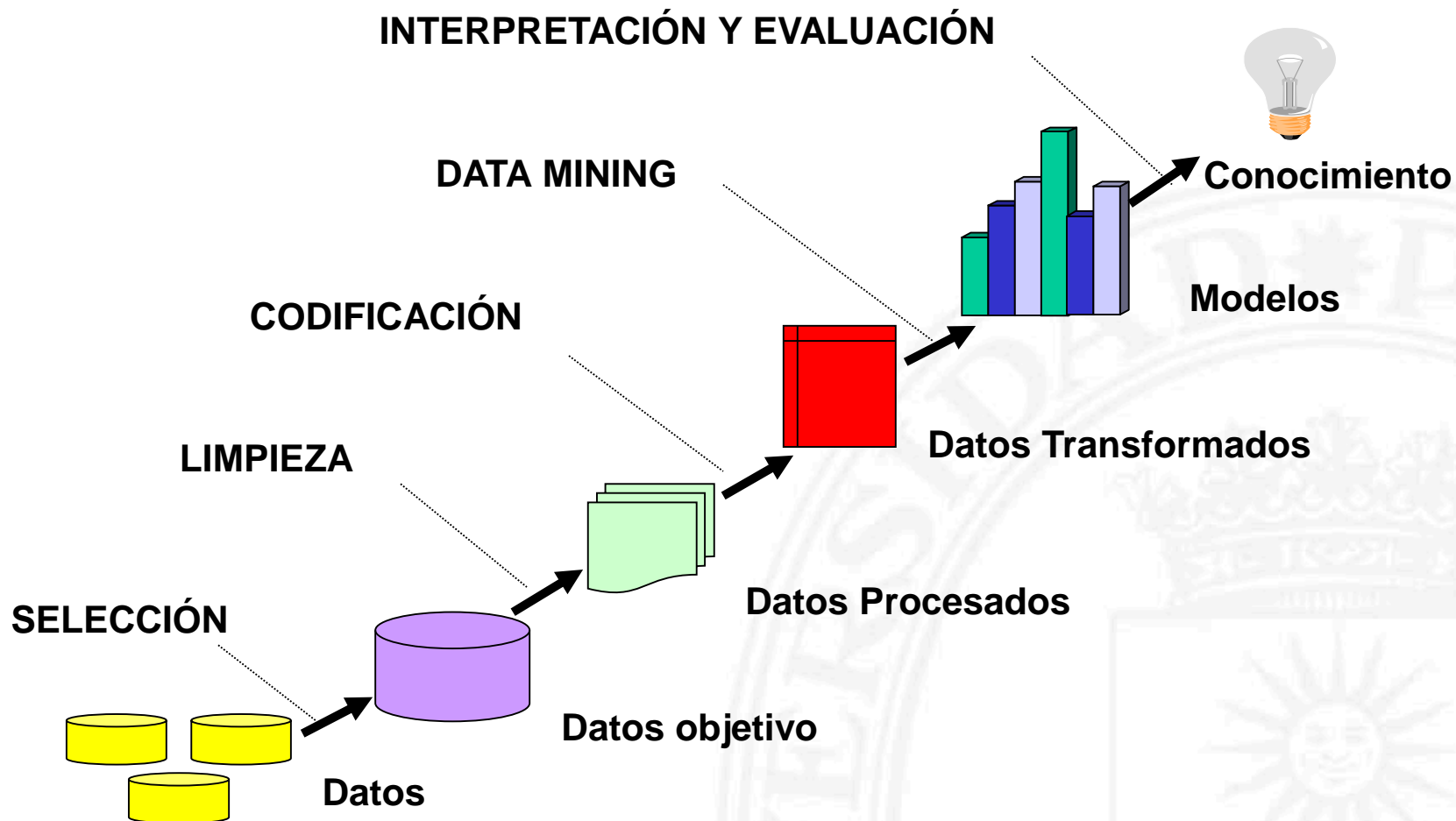
### PROGRAMMING & DATABASE

- ☆ Computer science fundamentals
- ☆ Scripting language e.g. Python
- ☆ Statistical computing packages, e.g., R
- ☆ Databases: SQL and NoSQL
- ☆ Relational algebra
- ☆ Parallel databases and parallel query processing
- ☆ MapReduce concepts
- ☆ Hadoop and Hive/Pig
- ☆ Custom reducers
- ☆ Experience with xaaS like AWS

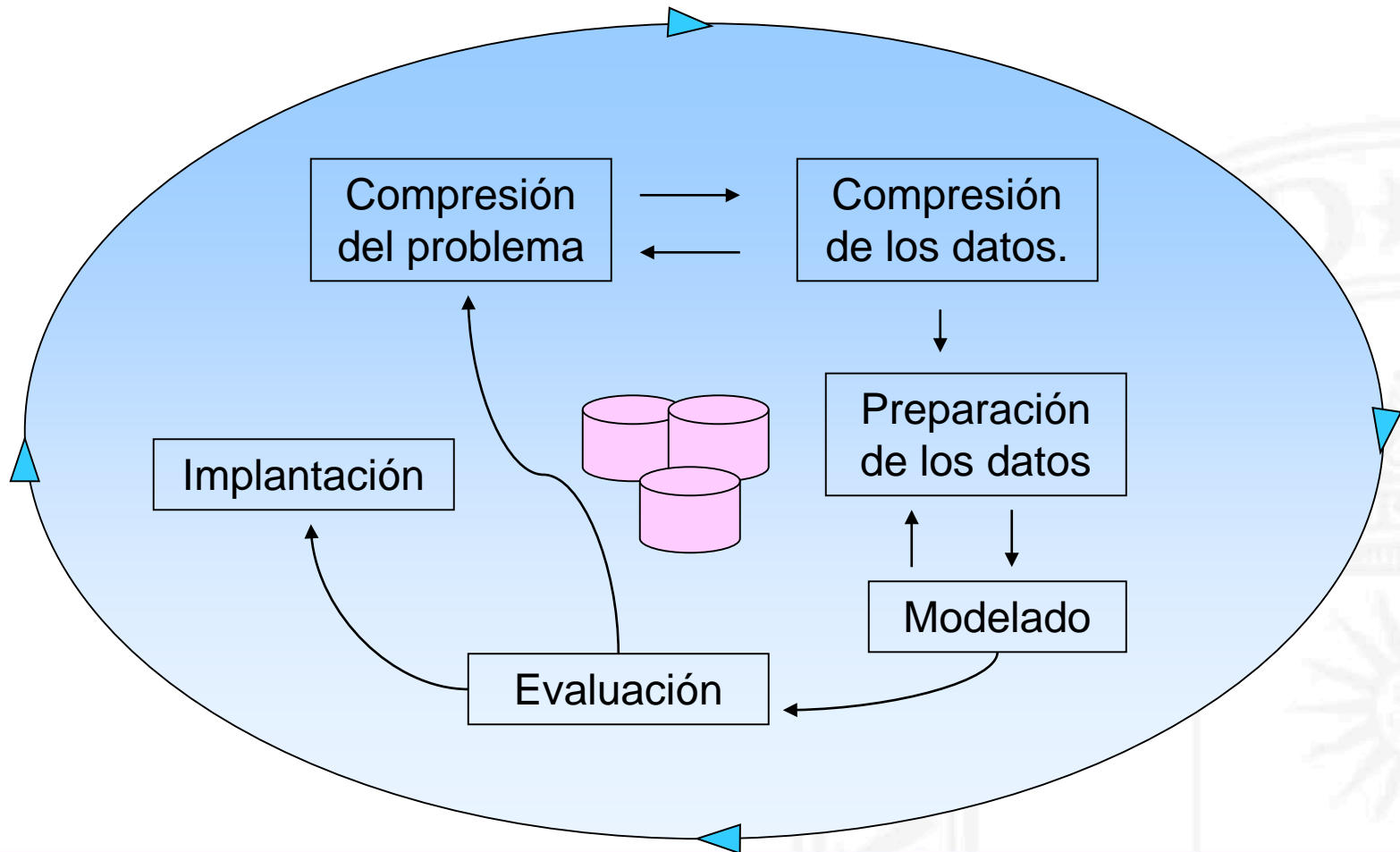
### COMMUNICATION & VISUALIZATION

- ☆ Able to engage with senior management
- ☆ Story telling skills
- ☆ Translate data-driven insights into decisions and actions
- ☆ Visual art design
- ☆ R packages like ggplot or lattice
- ☆ Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau

# El Proceso de KDD



# Estándar de proyecto de Data Mining: Crisp-DM ([www.crisp-dm.org](http://www.crisp-dm.org))



## Temario

1. Introducción. El valor de los datos: variedad, velocidad, volumen.
  - 1.1. Big Data
  - 1.2. Data Mining
2. Procesos de Data Mining
  - 2.1. Proceso de KDD
  - 2.2. CRISP-DM: fases
  - 2.3. El preproceso
3. Tipos de problemas de data mining
  - 3.1. Clasificación
  - 3.2. Clustering
  - 3.3. Asociación
4. Evaluación de los resultados: técnicas de evaluación



## Funcionamiento asignatura

### Clases:

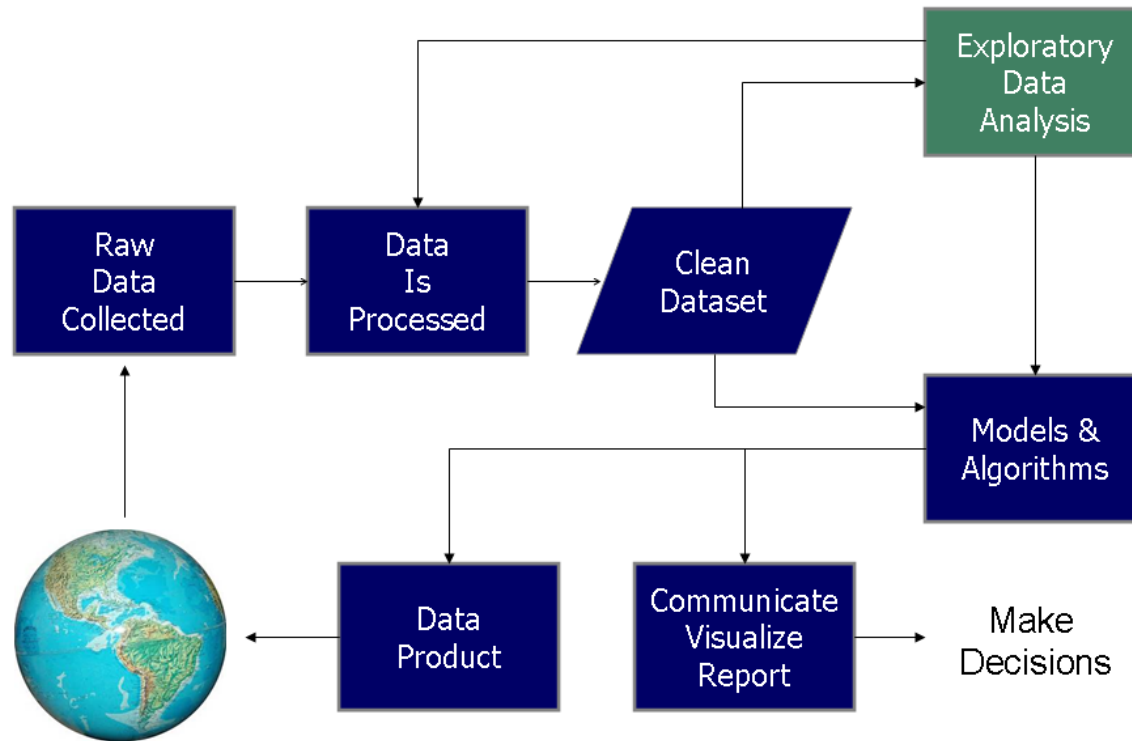
- Teóricas
- Prácticas: Herramientas (Knime, SPSS, IBM Cloud, ...)
  - Prácticas en grupos (2-3)

**Evaluación:** práctica + examen



# Data Science

## Data Science Process





# ¿Dudas, preguntas?

