Agència
per a la Qualitat
del Sistema Universitari
de Catalunya

www.aqu.cat

# Big (and Small) Data meets Quality Assurance

Anna Prades, Maribel Quirós, Maria Giné, Lorena Bernáldez

10th European Quality Assurance Forum

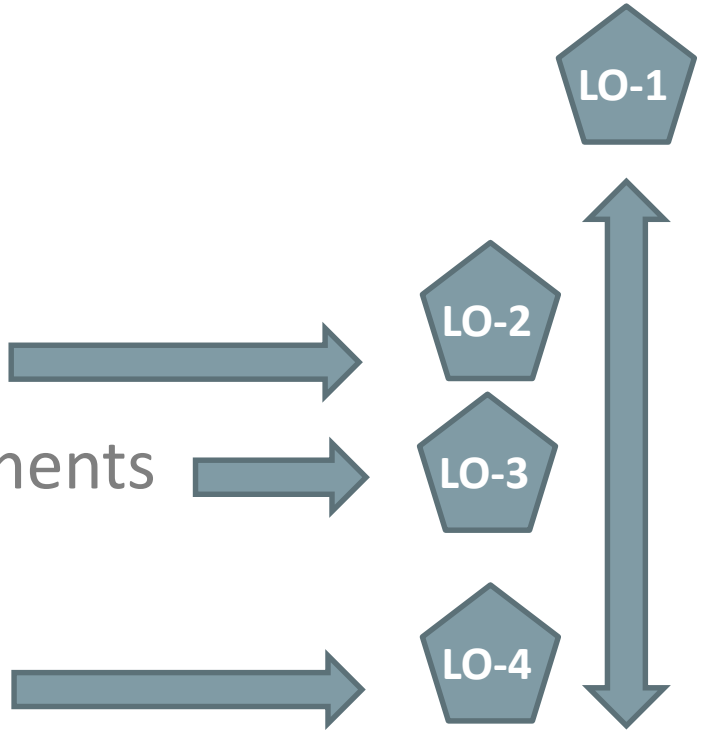
21th November, London

#aqu

# CONTRACT: Intended Learning Outcomes

| | |
|---|---|
| LO-1 | • Recognise the potential of data analysis available for quality assurance as well as the limitations inherent to the “excess” available data |
| LO-2 | • Appraise an evaluation model of research and teaching university departments based on statistical data |
| LO-3 | • Identify the human resources needed to develop small-scale project related data |
| LO-4 | • Reflect on the suitability of introducing a more quantitative analysis of programs and institutions |

- **Context**
- Why do we need data?
- Case study
- Human resources & skills requirements
- The future
- Discussion



# Context

Agència
per a la Qualitat
del Sistema Universitari
de **Catalunya**

The purpose of AQU Catalunya is **the assessment, accreditation and certification of quality in the sphere of the universities and higher education centers of Catalunya.**

ENQA eqar ISO 9001 ISO 27001

INQAAHE
INTERNATIONAL NETWORK FOR QUALITY ASSURANCE AGENCIES IN HIGHER EDUCATION

# Context

Low budget since 2010

Less resources

# Context: Types of Data

## UNEIX

Sources: universitiy pre-enrollment ddbb, research centers ddbb (CERCA), universities microdata (academic results) + teaching data + AQU surveys

## Graduates employment survey (Ba, Ma and PhD)

- Editions: 2001, 2005, 2008, 2011 and 2014 Population: 3 years after graduations (public + private universities)
- 153 variables (n= +/- 50,000)

## Graduates satisfaction survey

- 1st edition: 2014
- Set of 22 common questions

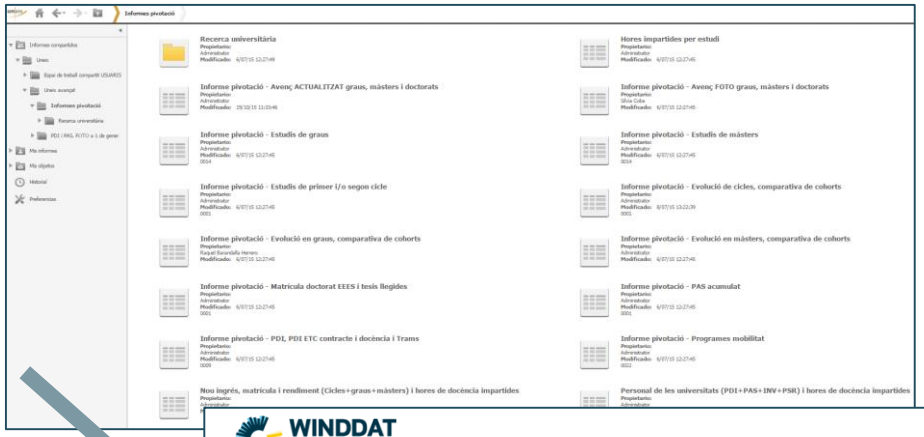
## Employers survey

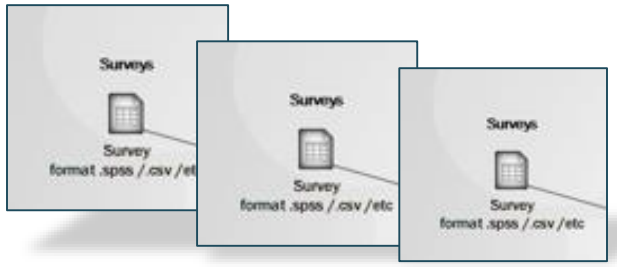
- 1st edition: 2014
- 1000 employers (from Health and Education to SMEs)

## Accreditation reports

Up to date 150 assessment reports (SQL database)

# Context: Uneix + Surveys -> winddat.aqu.cat






WINDDAT
indicadors docents per al desenvolupament i l'anàlisi de les titulacions

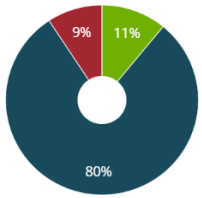
CATALONIA | UNIVERSITIES | COURSES | CONTACT | ENGLISH

13-14

All universities | Public | Private

**237,424 Students**

53% | 47%




All faculties | Main faculty | Affiliated

| | |
|---|---|
| Students who come from abroad | 7,528 |
| Students who go abroad | 6,790 |
| Number of theses defended | 2,250 |
| New students | 64,381 |
| Graduates | 44,128 |

| | Female | Male | Total |
|---|---|---|---|
| Pre-Bologna courses | 11,633 43 | 14,342 78 | 25,975 121 |
| Bachelors | 103,605 26,501 | 85,866 23,563 | 189,471 50,064 |
| Masters | 11,519 7,757 | 10,459 6,439 | 21,978 14,196 |
| | 126,757 34,301 | 110,667 30,080 | 237,424 64,381 |

**Enrolled students** New students

## Graduates

Total 44,128 Female 57% Male 43%

| | Female | Male | Total |
|---|---|---|---|
| Pre-Bologna courses | 5,103 | 5,657 | 10,760 |
| Bachelors | 14,404 | 8,903 | 23,307 |
| Masters | 5,605 | 4,456 | 10,061 |
| | 25,112 | 19,016 | 44,128 |

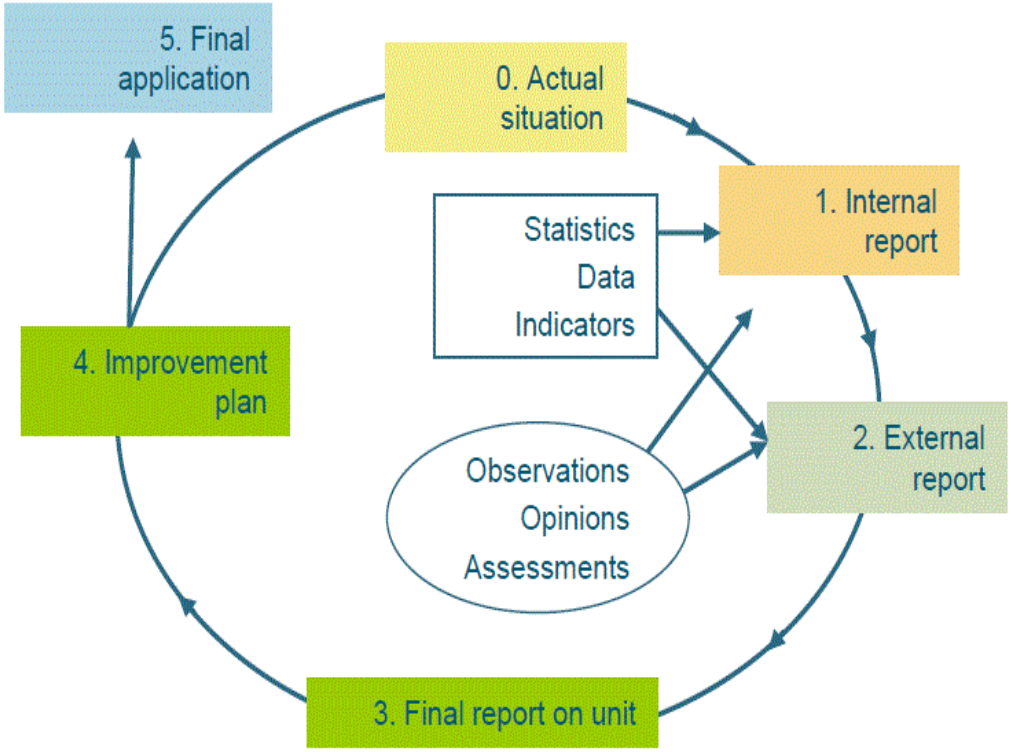
## Staff

Teaching staff and researchers: Women 40% Man 60%
Administration and services staff: Women 65% Man 35%

| | Teaching staff and researchers | Administration and services staff |
|---|---|---|
| Women | 7,743 | 7,088 |
| Man | 11,544 | 3,800 |
| | 19,287 | 10,888 |

- Context
- **Why do we need data?**
- Case study
- Human resources & skills requirements
- The future
- Discussion

# Data and QA: Love at first sight




"The quality assessment "spiral" is a circular and cyclical process which begins with collecting and systematizing information".

Westerheejden, 1996: 274

# Why do we need to use data?


**Standard 1.7**

Institutions should ensure that they collect, analyse and use relevant information for the effective management of their programmes and other activities

Quality culture

Data driven decision making


Turning data into **actionable** insights

**It's cool!!!**

# Starting point: information overloading

- Context
- Why do we need data?
- **Case study**
- Human resources & skills requirements
- The future
- Discussion

# Case Study

| Subject area | Sub-area | № Departments |
|---|---|---|
| Humanities | Fine Arts | 6 |
| | Philology | 22 |
| | Philosophy | 5 |
| | Geography and History | 21 |
| Social Sciences | Political Sciences and Sociology | 5 |
| | Law | 17 |
| | Economics and Business | 20 |
| | Education | 19 |
| | Journalism and Communication and Library Science | 7 |
| | Psychology | 11 |
| Experimental Sciences | Biology | 6 |
| | Geology | 8 |
| | Mathematics and Physics | 19 |
| | Chemistry | 10 |
| Health Sciences | Biomedicine | 21 |
| | Pharmacy | 4 |
| | Nursing | 7 |
| | Medicine and Surgery | 16 |
| | Veterinary Science | 4 |
| Engineering & Architecture | Architecture and Civil Engineering | 12 |
| | Mechanical and Production Engineering | 28 |
| | Agricultural Engineering | 5 |
| | ICT Engineering | 14 |
| Total | | **287** |

# Case Study

| Type | Code | Description |
|---|---|---|
| Research | RIND01 | Sum of TOTAL income generated in 4 years by PhD-holders / number of PhD-holders |
| | RIND02 | Sum of income from EUROPEAN SOURCES generated in 4 years by PhD-holders / number of PhD-holders |
| | RIND03 | Sum of income from STATE SOURCES generated in 4 years by PhD-holders / number of PhD-holders |
| | RIND04 | Sum of income from CATALAN SOURCES generated in 4 years by PhD-holders / number of PhD-holders |
| | RIND05 | Sum of income from CONTRACTS AND AGREEMENTS generated in 4 years by PhD-holders / number of PhD-holders |
| | RIND06 | Sum of income from COMPETITIVE FUNDING SOURCES generated in 4 years by PhD-holders / number of PhD-holders |
| | RIND07 | Sum of income from NON-COMPETITIVE FUNDING SOURCES generated in 4 years by PhD-holders / number of PhD-holders |
| | RIND08 | Current research premiums from regional government / no. people with current research premiums in the category Principal Investigator |
| | RIND09 | Sum of TOTAL income generated in 4 years by: 2-FULL PROFESSOR, 3-SENIOR LECTURER, 16-UNIVERSITY SCHOOL FULL PROFESSOR |
| | RIND10 | Sum of TOTAL income generated in 4 years by ALL OTHER TEACHING STAFF (not categories 2,3 or,16, above) |

## Fundraising

- Income generated
- Competitive vs not competitive
- National vs International

## Research awards / distinctions

By number of PhDs

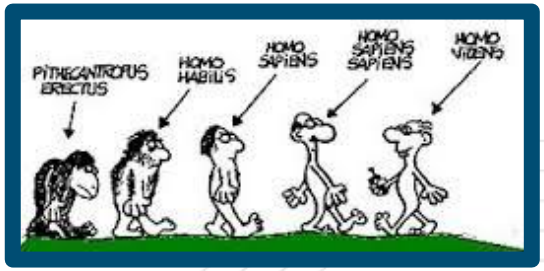
# Case Study

| Type | Code | Description |
|---|---|---|
| Teaching | PIND02 | Full-time teaching staff (A+B) with net teaching capacity > 0 / teaching staff (permanent + adjunct) * 100 |
| | PIND03 | Part-time teaching staff (A+B) with net teaching capacity > 0 / teaching staff (permanent + adjunct) * 100 |
| | PIND05 | Permanent teaching staff >= 60 years / total permanent teaching staff * 100 |
| | PIND06 | Permanent teaching staff >= 45 years and < 60 years / total permanent teaching staff * 100 |
| | PIND07 | Permanent teaching staff >= 35 years and 45 < years / total permanent teaching staff * 100 |
| | PIND08 | Permanent teaching staff < 35 years / total permanent teaching staff * 100 |
| | PIND10 | Six-yearly research premiums from regional government * 6 / five-yearly teaching premiums from regional government * 5 |
| | PIND11 | Total teaching and research staff + International researchers / Total teaching staff * 100 |
| | PIND12 | Total teaching and research staff + EU researchers - 15 + USA & CANADA / Total international * 100 |
| | PIND14 | % permanent teaching staff in (23 Special Services and 22 Services) / teaching staff (permanent + adjunct) |
| | PIND15 | % net teaching capacity of permanent teaching staff / potential teaching capacity of permanent teaching staff |
| | PIND17 | % PhD-holders / total teaching staff |
| | PIND18 | Age indicator (sum of all groups - over-60) / over 60 |
| | PIND19 | Distribution of classroom teaching hours; % according to teaching staff category |

## Teachers

- Academic categories
- Permanent vs non permanent
- Age
- National vs International

# Case Study: the Result





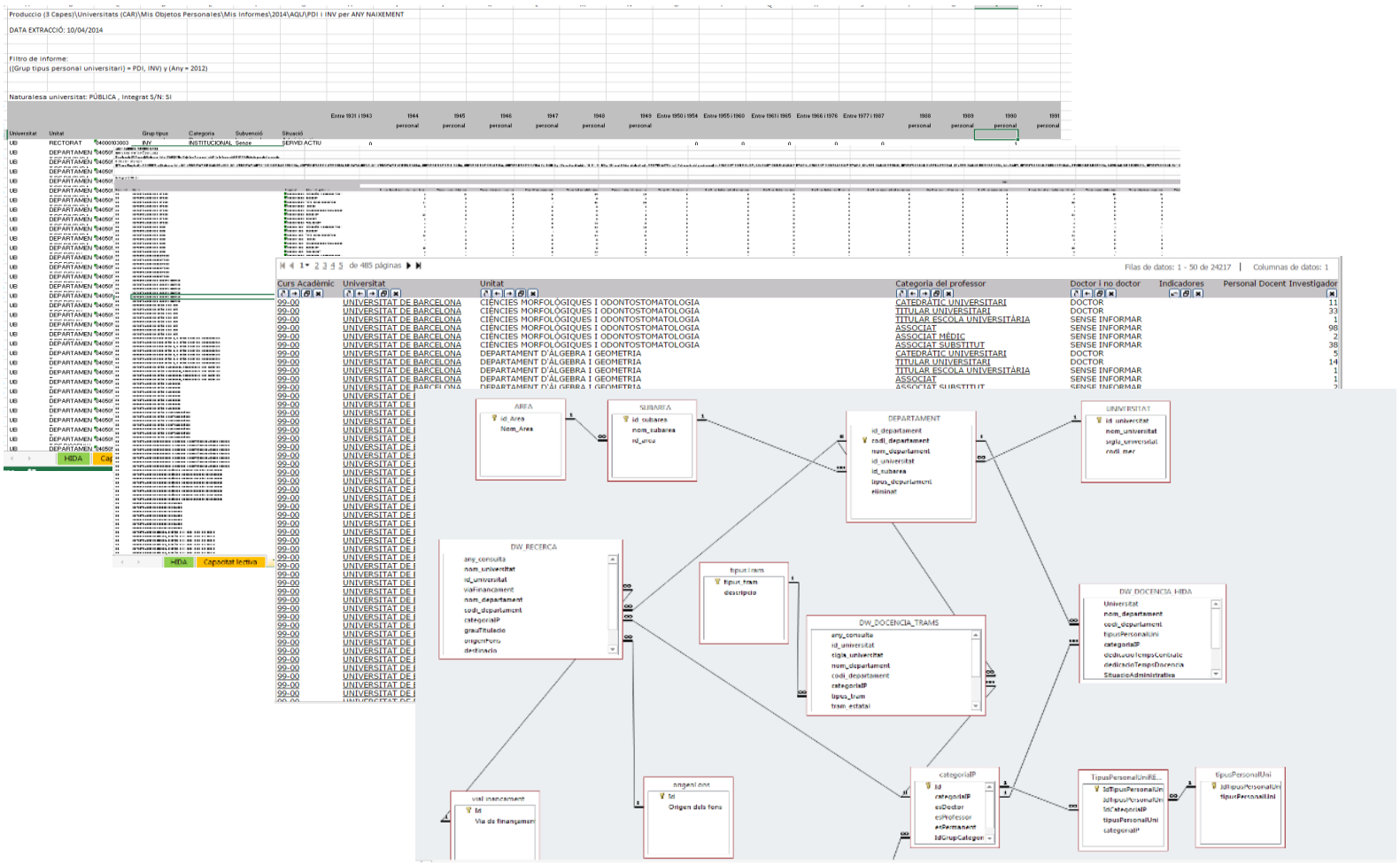


Universities

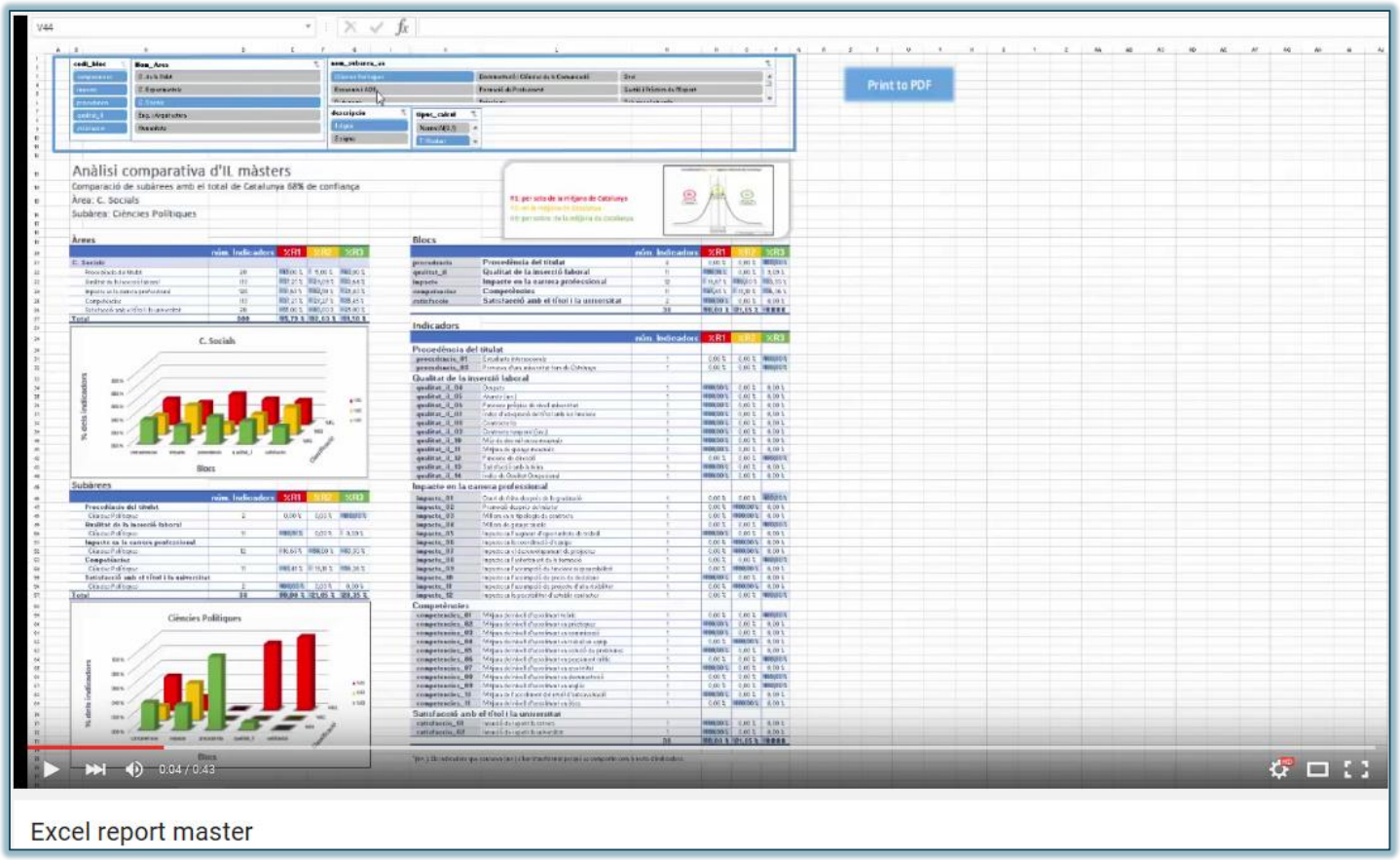
| | Nº Dept. | Nº Ind. | % Q4 | % Q3 | % Q2 | % Q1 |
|---|---|---|---|---|---|---|
| Uni 6 | 106 | 2014 | 23,44 % | 21,80 % | 29,79 % | 24,98 % |
| Uni 1 | 57 | 1073 | 19,38 % | 25,63 % | 29,54 % | 25,44 % |
| Uni 2 | 42 | 798 | 25,69 % | 27,57 % | 24,94 % | 21,80 % |
| Uni 3 | 8 | 152 | 19,08 % | 5,92 % | 10,53 % | 64,47 % |
| Uni 4 | 24 | 456 | 41,23 % | 23,03 % | 21,05 % | 14,69 % |
| Uni 7 | 26 | 494 | 35,22 % | 25,91 % | 22,47 % | 16,40 % |
| Uni 5 | 24 | 456 | 28,07 % | 17,32 % | 32,02 % | 22,59 % |
| **Total** | **287** | **5443** | **25,79 %** | **23,06 %** | **27,28 %** | **23,87 %** |

Quartiles classification of the previous indicators

# Case Study: the backoffice



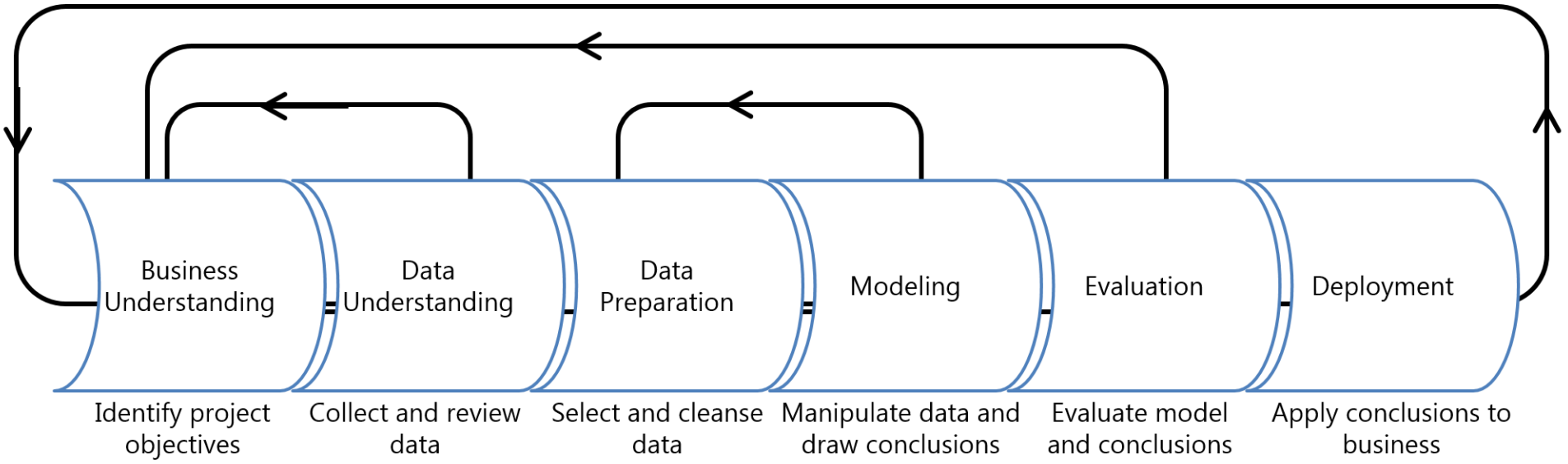
# An extrapolation of the Case Study: Masters employment indicators




Excel report master

https://youtu.be/1fadu1fvf28

# The data science process

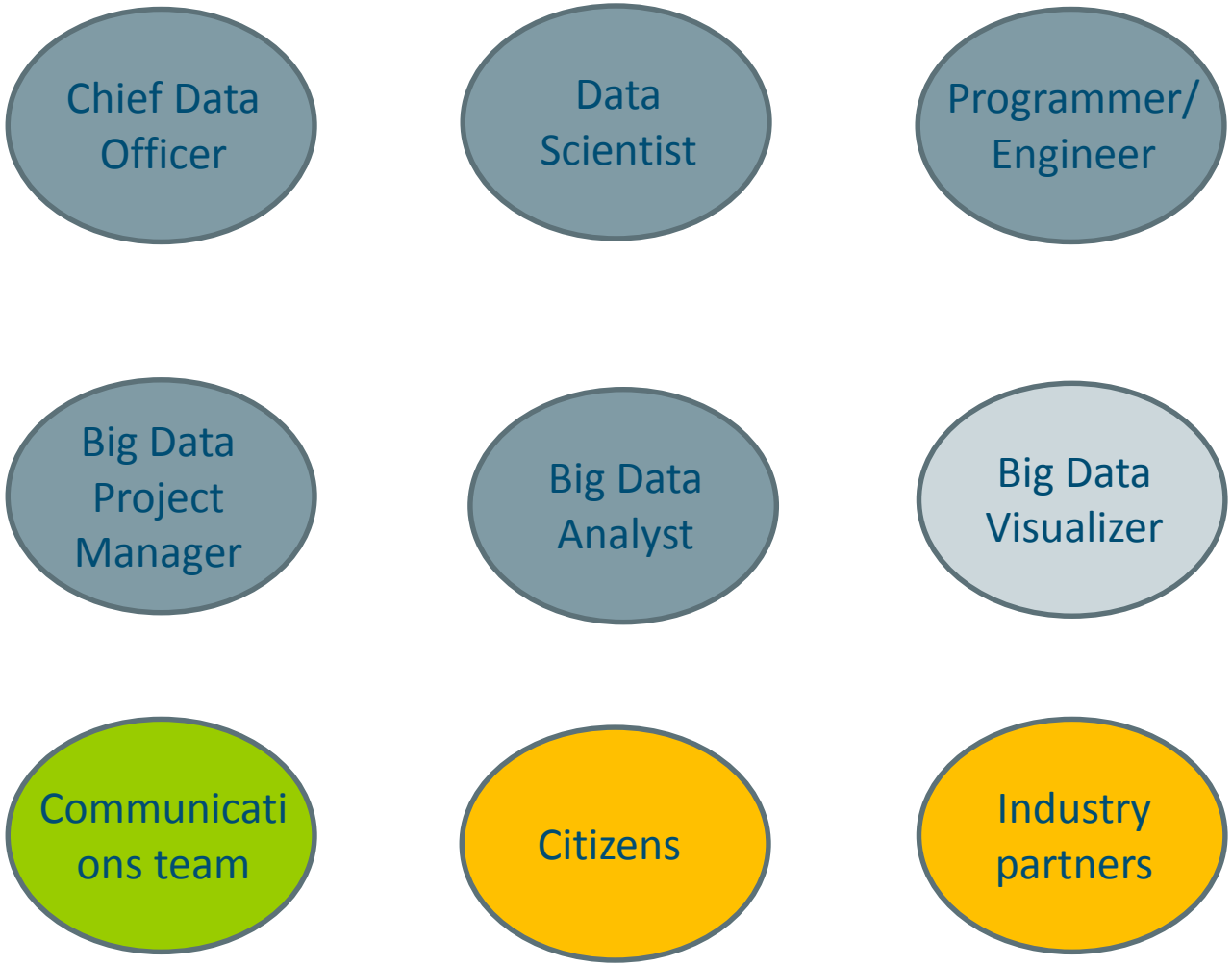



CRISP DM(2000)

Image from Data Science and Machine Learning Essentials (edX), by Cyntia Rudin and Stephen F Elston

- Context
- Why do we need data?
- Case Study
- **Human resources & skills requirements**
- The future
- Discussion

# HR: Your All-Star Team (by GovLoop)

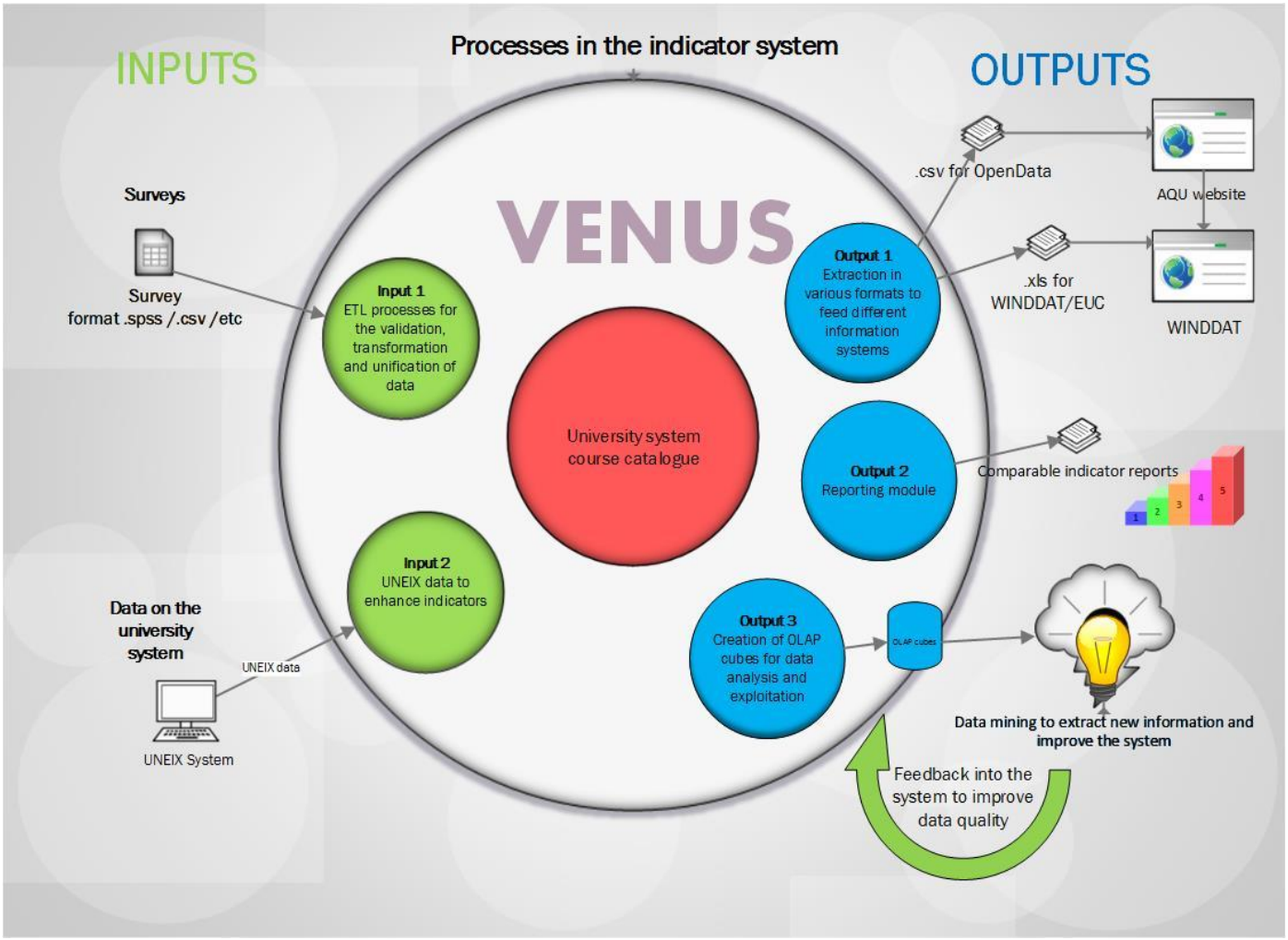


- Context
- Why do we need data?
- Case Study
- Human resources & skills requirements
- **The future**
- Discussion

# The future: a preliminary list

- Predictive models of professional success
- Detect good practices regarding skills gap
- Real-time teaching staff renewal (upcoming retirements)
- Institutional KPI for teaching and research
- Predictive model of accreditation results: a self-assessment tool
- Possible use of existing data to simplify the assessment procedures

# The future


Processes in the indicator system
INPUTS
OUTPUTS
VENUS
Surveys
Survey
format .spss /.csv /etc
Input 1
ETL processes for the validation, transformation and unification of data
University system course catalogue
Input 2
UNEIX data to enhance indicators
Data on the university system
UNEIX data
UNEIX System
Output 1
Extraction in various formats to feed different information systems
.csv for OpenData
AQU website
.xls for WINDDAT/EUC
WINDDAT
Output 2
Reporting module
Comparable indicator reports
Output 3
Creation of OLAP cubes for data analysis and exploitation
OLAP cubes
Data mining to extract new information and improve the system
Feedback into the system to improve data quality

# Thank you for your attention

http://www.aqu.cat

aprades@aqu.cat

mquiros@aqu.cat

@aqucatalunya