



# Data Journalism Workflow

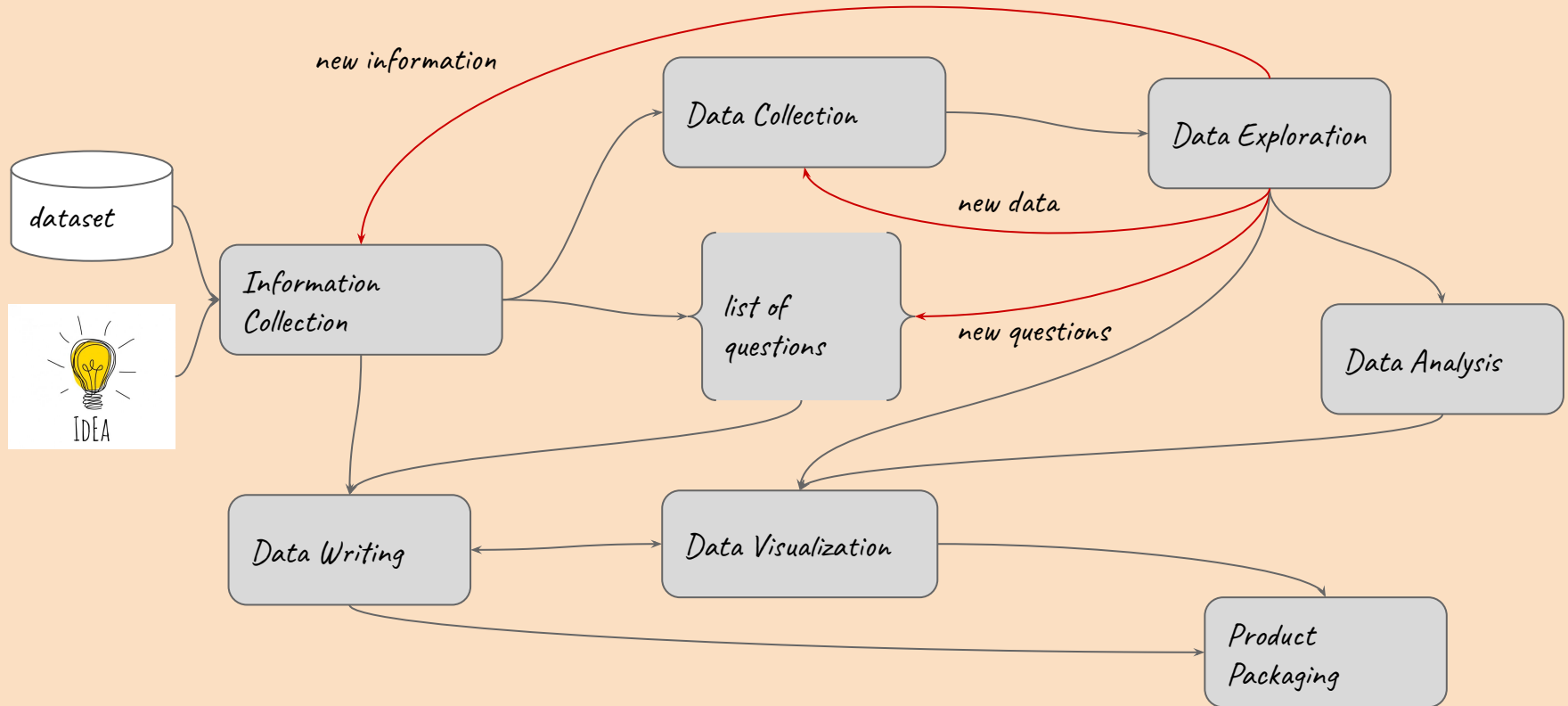
Project a DJ case - 1<sup>^</sup> Part

# Overview

---

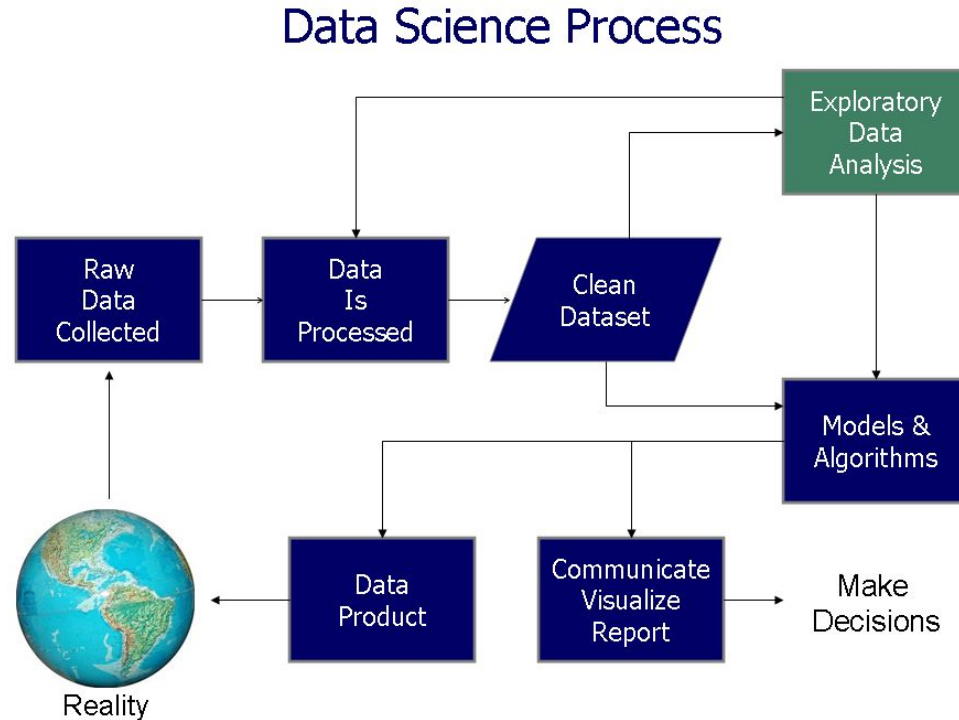
- Data Journalism Workflow
  - Women's Football
- Data Journalism Team

# Data Journalism Workflow

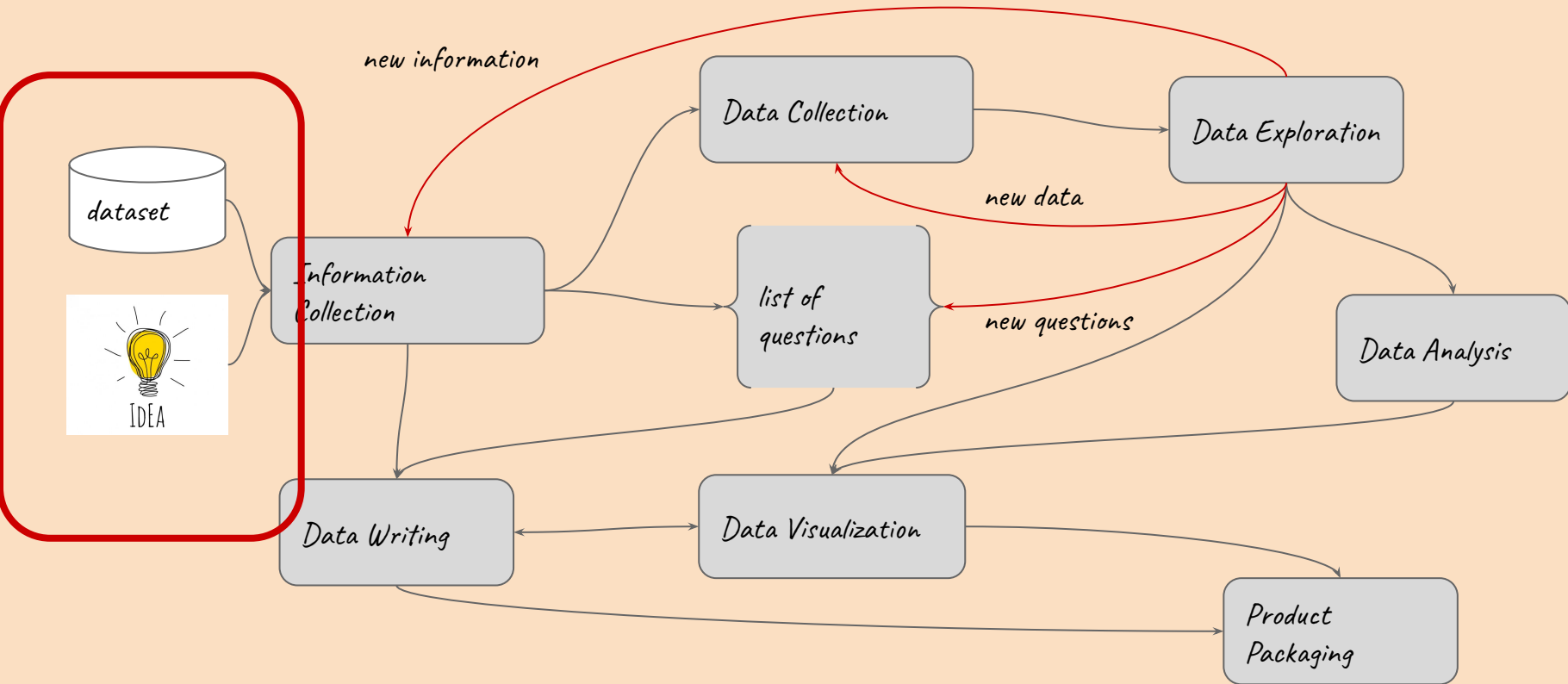


# Data Science Process Flowchart

---



# Start



# The initial two possible Scenarios

---

*Data journalism begins in one of two ways:*

- 1. a question that needs data, (journalistic)*
- 2. an interesting dataset that needs questioning (data scientist)*

# An interesting dataset

**GED**I VISUAL

L'ITALIA DELLE SLOT 2

## QUANTO GIOCANO GLI ITALIANI

Dal Superenalotto al Gratta&Vinci,  
dal Bingo alle Slot: cerca quanto  
pesa la febbre del gioco nel tuo  
comune

**CERCA IL TUO COMUNE**

A stylized illustration in black, white, and red. A man in a white shirt, black tie, and black suit jacket sits at a black table. He wears red-tinted sunglasses and holds a red smartphone in his left hand. On the table in front of him are several stacks of red coins and white banknotes, along with two dice. To his right is a black slot machine with a screen showing three red cherries. The background features a large red shape on the left and a black shape on the right.

# Idea - Topic selection

---





# Idea - Topic Selection

---



## Agriculture

Agricultural output, Agricultural policy, Fisheries, Sustainable agriculture



## Development

Development resource flows, Official development assistance (ODA)



## Economy

Corporate sector, Foreign direct investment (FDI), GDP and spending, Household accounts, International trade, Leading indicators, National income, Prices, Productivity



## Education

Education attainment, Education resources, International student assessment (PISA), Students, Teachers, Youth and the labour market



## Energy

Energy, Transport



## Environment

Air and climate, Biodiversity, Environmental policy, Forest, Materials, Waste, Water



## Finance

Conversion rates, Insurance, Interest rates, Monetary aggregates, Pensions



## Government

General government, Tax



## Health

Health care use, Health equipment, Health resources, Health risks, Health status



## Innovation and Technology

Broadband access, Entrepreneurship, Industry, Information and communication technology (ICT), Research and development (R&D)



## Jobs

Benefits and wages, Earnings and wages, Employment, Unemployment



## Society

Demography, Inequality, Migration, Population by region, Social protection

# Use Case for the course

---

Women's  
football



[wafi.iit.cnr.it/lab/dj/calciofemminile/](http://wafi.iit.cnr.it/lab/dj/calciofemminile/)  
[alod83.github.io/calciofemminile/](https://alod83.github.io/calciofemminile/)

INTRODUZIONE

IN MINORANZA

UN GIOCO DA DILETTANTI

SE L'INTERESSE CRESCE



L'ALTRA METÀ DEL CALCIO

# Audience selection

---

General audience

Experts of the topic

Children

Students

Women/Men

...

# Audience selection - Women's soccer

---

General audience

Experts of the topic

Children

Students

Women/Men

...



# Journal selection

---

Newspaper  
Magazine  
Journal  
...

**CATEGORY**

Women/Men  
Sport  
Economy  
Politics  
Science  
...

**TOPIC**

# Journal selection - Women's soccer

---

Newspaper

Magazine

Journal

...

CATEGORY

Culture

Sport

Economy

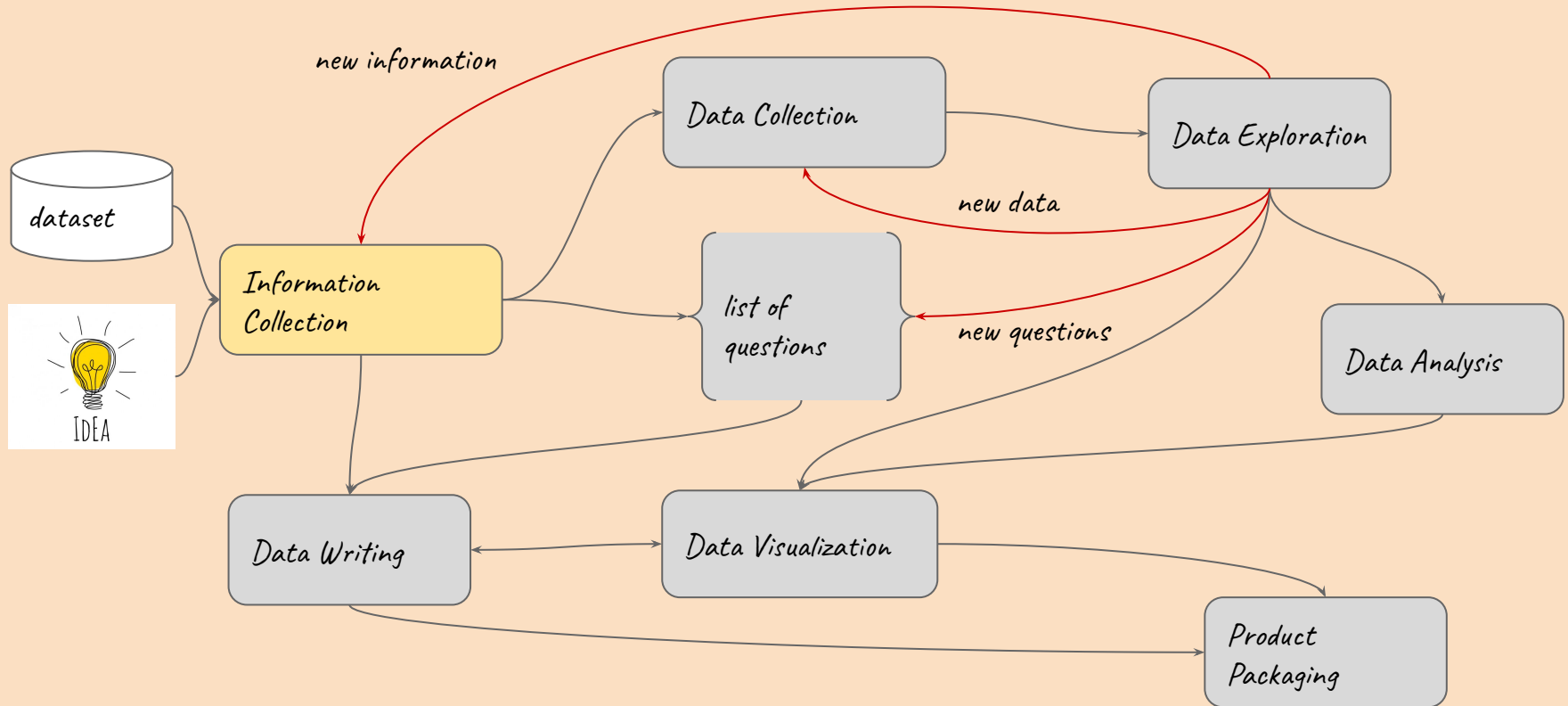
Politics

Science

...

TOPIC

# Information Collection





# Get Information

---

- **Study of previous works on the same subject**
  - Crucial to identify if the story is good
- Goal - Understand the domain
- The interviews to experts can help
- Google & Google & Google



# Major news agencies

---



[Prime pagine dei quotidiani nazionali](#)

# News clustered by a theme

---

## Do prejudices still exist?

[Belloli frase contro ragazze](#) [repubblica.it](#) 14/05/2015

[Le donne non capiscono di tattica - Collovati](#) [gazzettaFanNews](#) 20/02/2019

[sergio-vessicchio-telecronista-insulta-guardalinee-donne-schifo](#) [Gazzetta dello sport](#) 25-03-2019

## In the world women's soccer is exploding, what is the situation in Italy?

[Europa stadi pieni ma in Italia il movimento stenta a decollare](#) [sportbusinessmanagment.it](#) 01/09/2012

[The womens football boom](#) [clublinefootball.com](#) 20/03/2017

[Quanto-vale-calcio-femminile-italia-europa-confronto/](#) [Calcio e Finanza](#) 18/12/2017

[L'ascesa del calcio femminile in italia](#) [sportbusinessmanagment.it](#) 01/07/2018

[Inter sbarca calcio femminile](#) [gazzetta.it](#) 23/10/2018

[Mondiale calcio femminile sorteggio girone Italia](#) [skysportnews](#) 08/12/2018

[Atletico Madrid Vs Barcellona più di 61.000 spettatori](#) [repubblica](#) 18/03/2019 (cont..)

# News clustered by a theme

---

## How does the football government behave?

[Alla Juve i diritti del Cuneo Femminile](#) La Stampa 01/07/2017

[Il calcio femminile passa dalla lega dilettanti alla FIGC](#) Il sole 24 ore 16/04/2018

## What are the difficulties a female soccer player encounters?

[Perché lo sport non ama le donne](#) Repubblica 20/07/2016

[Il professionismo in Italia storia di una discriminazione](#) Sport alla rovescia 07/03/2017

[Quanto-e-difficile-diventare-calciatrici-in-italia-intervista-a-regina-baresi](#) 21/12/2017

[Le-campionesse-che-non-possono-diventare-professioniste](#) Il Fatto Quotidiano del 28/03/2018

[Juventus-women-le-sfide-per-essere-calciatrici-oggi-in-italia](#) Il sole24ore 20/03/2018

# Legge 23 marzo 1981, n.91

---

*Norme in materia di rapporti tra società e sportivi professionisti*

art. 2.

Professionalismo sportivo

Ai fini dell'applicazione della presente legge, sono sportivi professionisti gli atleti, gli allenatori, i direttori tecnico-sportivi ed i preparatori atletici, che esercitano l'attività sportiva a titolo oneroso con carattere di continuità nell'ambito delle discipline regolamentate dal CONI e che conseguono la qualificazione dalle federazioni sportive nazionali, secondo le norme emanate dalle federazioni stesse, con l'osservanza delle direttive stabilite dal CONI per la distinzione dell'attività dilettantistica da quella professionistica.

The law banning women soccer players from becoming professionals is almost 40 years old

# Interviews

---



# Interviews

---

- Select a range of **topic experts**
  - contact each expert separately
  - contacts could be done by email, social network
- Prepare in advance the **list of questions**
- Interviews can be done:
  - by **email** - simple text
  - by **telephone** - record audio
  - **face-to-face** - record video/audio
  - by **videoconference** - record video/audio

# Who is a topic expert?

---

A famous people

An institution representant

A colleague/friend/parent

A professor/researcher/teacher

A journalist

...



# Interviews - Women's soccer

---

[Carolina Morace](#) - AC Milan coach

Luisa Alessio - Triestina ex player

Missed interviews

- Sara Gama - Juventus player
- Laura Perrotta - CONI
- Patrizia Recandio, responsible of *Sviluppo Calcio femminile* – FIGC



# Large scale interviews

---

- Interview a great number of people, e.g. people interested/expert in the field
- Interviews can be done through
  - polls (e.g. Google Forms)
  - face-to-face
- Spread the poll to forums, mailing lists, friends, ...
  - many times the polls are meaningless

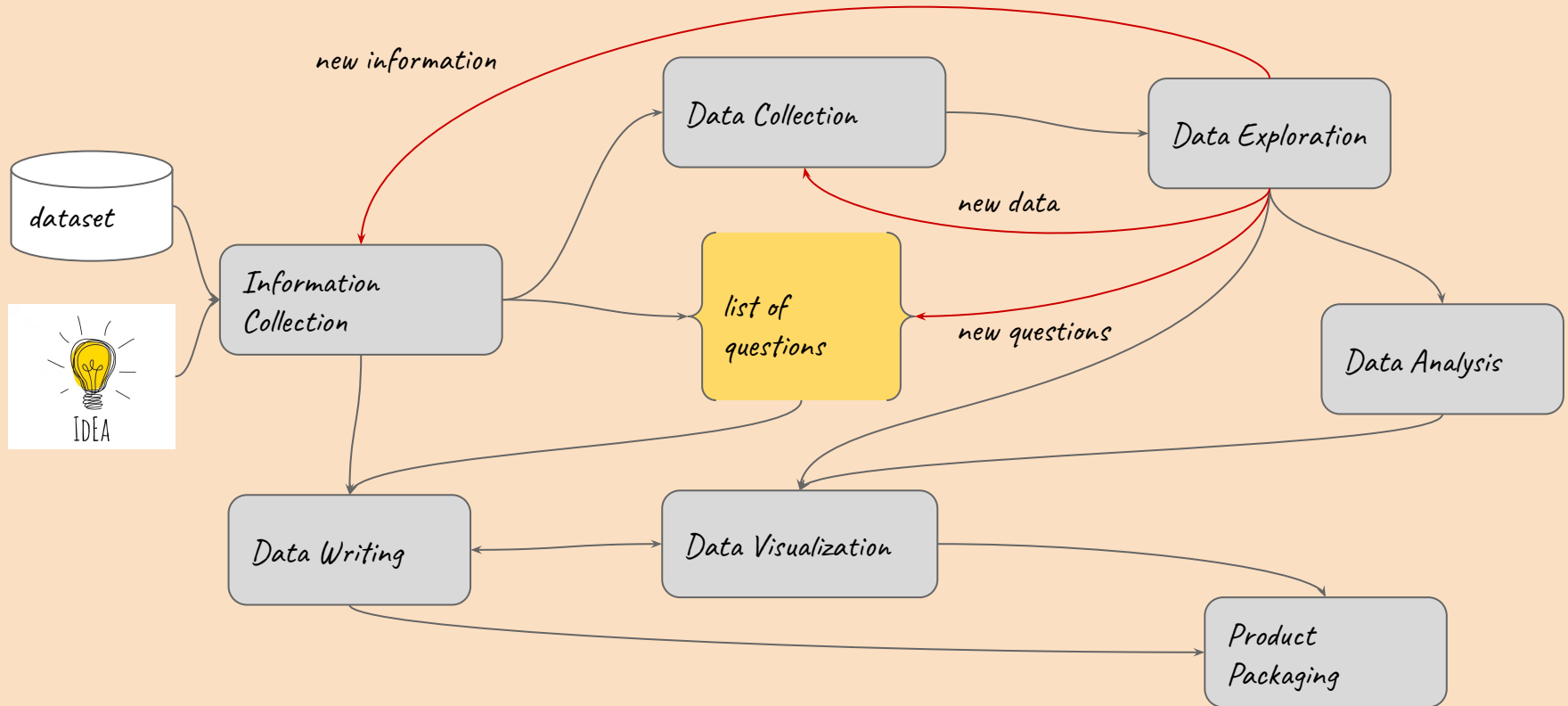
# The role of interviews

---

- Understand the domain of knowledge
- Discover interesting stories to tell
- Verify the final job
- Include the interviews inside the story

# List of questions

---



# List of Questions

---

- Create a list of questions that you want to investigate
- This list can change continuously
- The final list will be the chapters of our story

# List of questions

---

Define which questions the story should answer

- general questions
- specific questions

Each question should be **short** and **clear**.  
The number of questions should be **limited**.  
**New questions** could arise during the investigation.

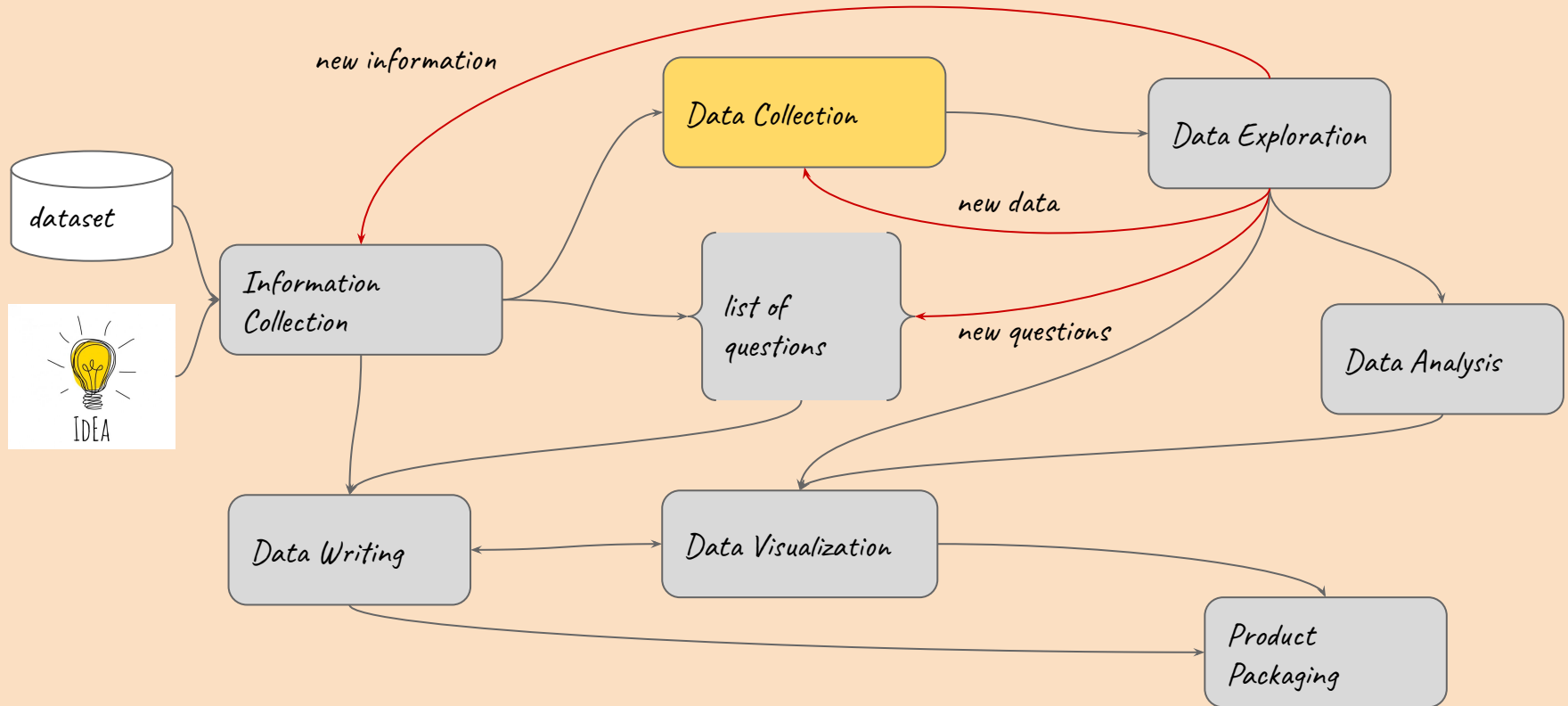


# List of questions - Women's soccer

---

1. What is the situation in Italy with respect to women's soccer? In Europe and in the world?
2. How widespread is women's soccer in Italy?
3. Are there still discrimination against women's soccer?
4. How does the soccer government behave?
5. What are the difficulties a female soccer player encounters?

# Data Collection





# Researching Data

---

- Most important role of DJ
- Time consuming, tedious and boring



# Data collection

---

WHERE?

- Open Data
- Web crawling through API
  - free/for a fee
- Web scraping
  - manage copyrights
- PDF/other documents

# Open Datasets

---

User generated data: [Wikipedia](#), [DBpedia](#), [Wikidata](#)

Open Government Data: [data.gov](#), [data.gov.uk](#), [data.gov.it](#), ...

Statistics Institutes: [istat](#), [eurostat](#), [worldbank](#), [oecd](#), ...

Open Data aggregators: [awesome dataset](#), [cooldataset](#), ...

Open Data search engines: [Google Data Search](#), [Google trends](#)

# Data Collection - Women's soccer

---

Source	Description	Format	Data extraction technique
<a href="http://calciofemminileitaliano.it">calciofemminileitaliano.it</a> , <a href="http://tuttocampo.it">tuttocampo.it</a>	Women's football teams A and B	HTML	Web scraping with python
<a href="http://football.it">football.it</a> femminile, <a href="http://tuttoCampo.it">tuttoCampo.it</a> , <a href="http://it.soccerway.com">it.soccerway.com</a> , <a href="http://calciofemminileitaliano.it">calciofemminileitaliano.it</a>	Women and Men's Players Data (2017-2018, 2018-2019 A,B)	HTML	Web scraping with python
<a href="http://Istat.it">Istat.it</a>	Sports practiced by gender and age 2006, 2015	CSV	Download
<a href="http://fifa.com">fifa.com</a> <a href="http://worldbank.org">worldbank.org</a>	Country FIFA ranking VS GDP VS Gap Gender	HTML table CSV	Copy and paste of the table into a spreadsheet

# Data Collection - Women's soccer

---

Source	Description	Format	Data extraction technique
<a href="#">FIFA report 2017</a>	Number of players under 18, Number of female players, Total number of football clubs, Number of clubs with both male and female teams, Number of clubs only with female team	PDF	<a href="#">Tabula software</a> tabula py
<a href="#">UEFA report 2017</a>	Average distance to local club, Total no of qualified female referees, Do they referee men's matches?, Top level where women referee matches, Av. attendance at national team matches, No of women at managerial level or above, teams budget, ...	PDF	Converted in XLS, XLS->CSV python scraper