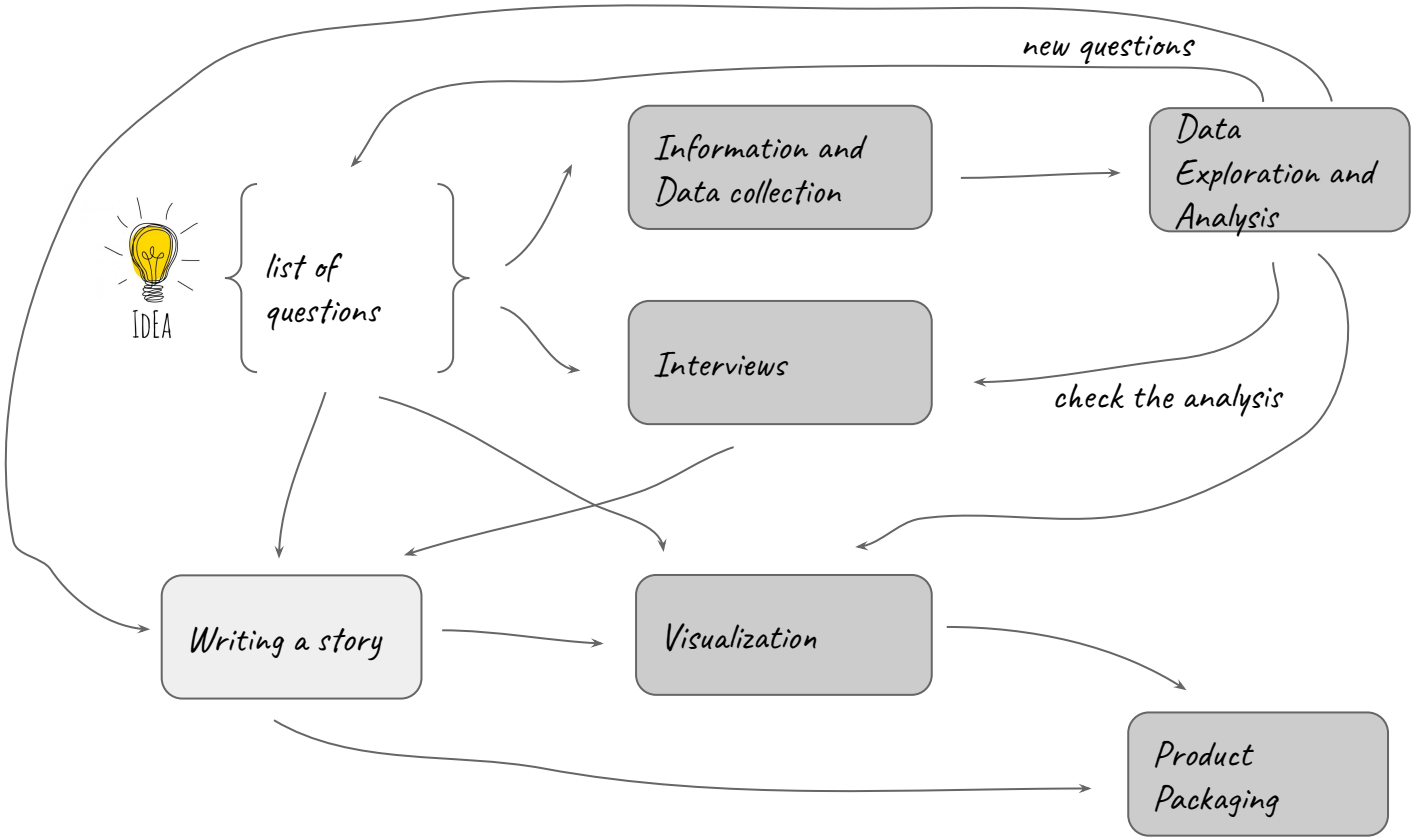
Data Journalism Workflow

InfoUma 2019-20 Andrea Marchetti

Data Journalism Workflow



Idea & List of Questions

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

Idea - Topic selection

Society

Science and Technology

Climate change

Sports

Health

Economy

Politics

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

Major news agencies

To check the **current agenda**

Prime pagine giornali nazionali

List of Questions

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

Interviews

The role of interviews

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

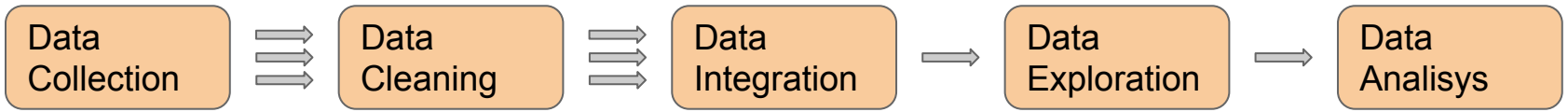
Information & Data Collection

Get Information

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

Researching Data

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



Writing a story

Writing

- Numbers without context are just numbers, **words give context**
- Declare the sources, explain the charts, wrap a story around the numbers, ensure the numbers are clearly understood
- Enter real stories not just facts
- Sometimes words are enough

StoryBoard

- The List of questions will be our storyboard
- Try to guess the visualization useful for the story
- The story it’s the centre

What are you trying to say?

- Stories are aimed at the general public not academics.
- Imagine you are are telling the story to your grandmother.
- Your job is to bridge the gap between the data and the user
- Sometimes it’s sufficent to present open data with simple visualization
  - Example: Art Market for Dummies (Askmedia.fr).

What is or isn’t a story

How to choose what is interesting?

This is the million dollar question

Some questions can help:

- Is it straightforward?
- Can you explain it simply to someone who has never heard of it before?
- Is it newsworthy?
- How much time for producing this work?
- …

What is or isn’t a story

Some questions can help:

- Will it be past its sell-by-date when it is published
- What is the best possible way to tell this story?
- How simple can I make it?

A great storyteller James Cameron from the Guardian said “The main thing is simplicity”

James Cameron and the importance of the story (simonrogers.net)

Where

- GeoLocation of data is very important, it allows different sources of data to be integrated and “mashed up” to create a new story.
- Space and time are natural ways to mash up data.
  - The gun ownership and gun homicides murder map of the world (The Guardian)

The source

- Where did the data come from? **The source**
- The most important W
- **Transparency** about the source is critical.
  - How do you know if the information is correct? We need to use data from reliable sources. We need to declare which modifications have been carried out in order to have reproducible results
- Do not blindly trust the accuracy of supplied data

Causation Vs Correlation

- Data in the right hands can be extremely powerful and should a key element of any decision
- But too often data can be misunderstood
- Correlation is, “A statistical measure (expressed as a number) that describes the size and direction of a relationship between two or more variables
- causation “Indicates that one event is the result of the occurrence of the other event

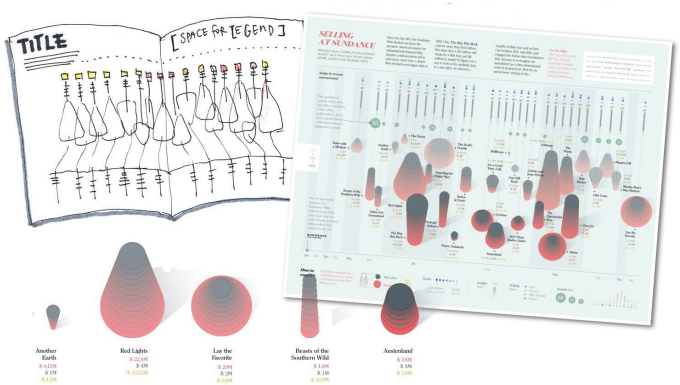
smoking is correlated with alcoholism, but doesn’t cause alcoholism.

smoking causes an increase in the risk of developing lung cancer

Visualization

Designing and visualising

- Designers **plan** and **create** visualisations with the help of journalists and developers
- Sketch a visualization before realize it



Product packaging

The article will be realized as a web application

Final check

- check the results with domain experts
- check the sources and the rights

Newsroom composition and organization

inside a data team

What kind of teams are there?

1. Lone rangers

- one person does everything
- possible using all the tools of the DDJ ecosystem: OpenRefine, Datawrapper, Tableau, Google Fusion Tables, CartoDB.

2) Two-person teams

- The Guardian US's two person DDJ team created the award winning Guide to gay rights in the US.

3) Small scale team (less than 25)

- Able to produce innovative projects quickly
- Part of the newsdesk
- Example: 'Flooding and Flood Zones' map Hurricane Sandy (WNYC)

4) Large team (more than 25)

- Has a deliberate strategy to create a new kind of online journalism
- They help with finding ways to tell the story better

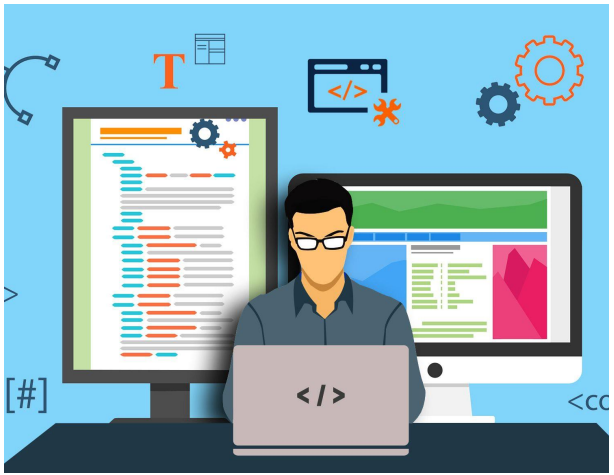
DJ Newsroom roles

1. **data finder:** find relevant data sets
2. **analyst**: carry out good analysis
3. **data visualizer**: create appealing visualizations
4. **designer:** adapt the story to the target users
5. **journalist: identify** the breaking news and then **interview** and **write it up**

Data Journalism fundamental roles

JOURNALIST

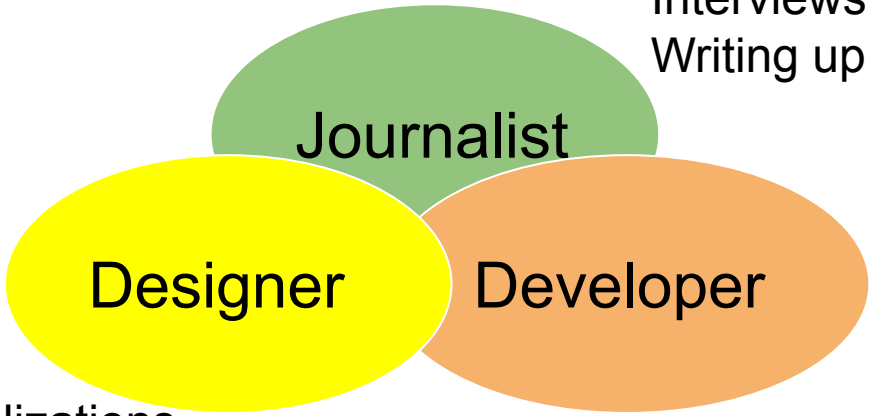
WEB DESIGNER



DEVELOPER (data scientist)

Small DJ Newsroom

Ideas & Questions
Interviews
Writing up the story



Web Layout
Graphics & Visualizations
Photo/Video/Audio

Data Harvesting coding
Data Visualization coding
Web Publish

Coordination

- The journalist is the coordinator
- Journalist, designer and developer must cooperate

Caverage with other courses