Data Journalism

Introduction

InfoUma 2019-20 Andrea Marchetti

Overview

- What Data Journalism is

- Newsroom composition and organization

- How to transform numbers into stories

what data journalism is

What Data Journalism is

- Telling stories using numbers
- Using **visual technologies** to aid communication of the story
- **Technologies** are not important, the focus is on the **story**
- Journalism is **reporting facts** in a way that **people can understand**
- As a data journalist, your role is **to bring data to life**

The origins of DJ

Data has always been part of news work.

The <u>Wall Street Journal</u> was created from a daily Dow Jones afternoon letter about the stock exchange (link).
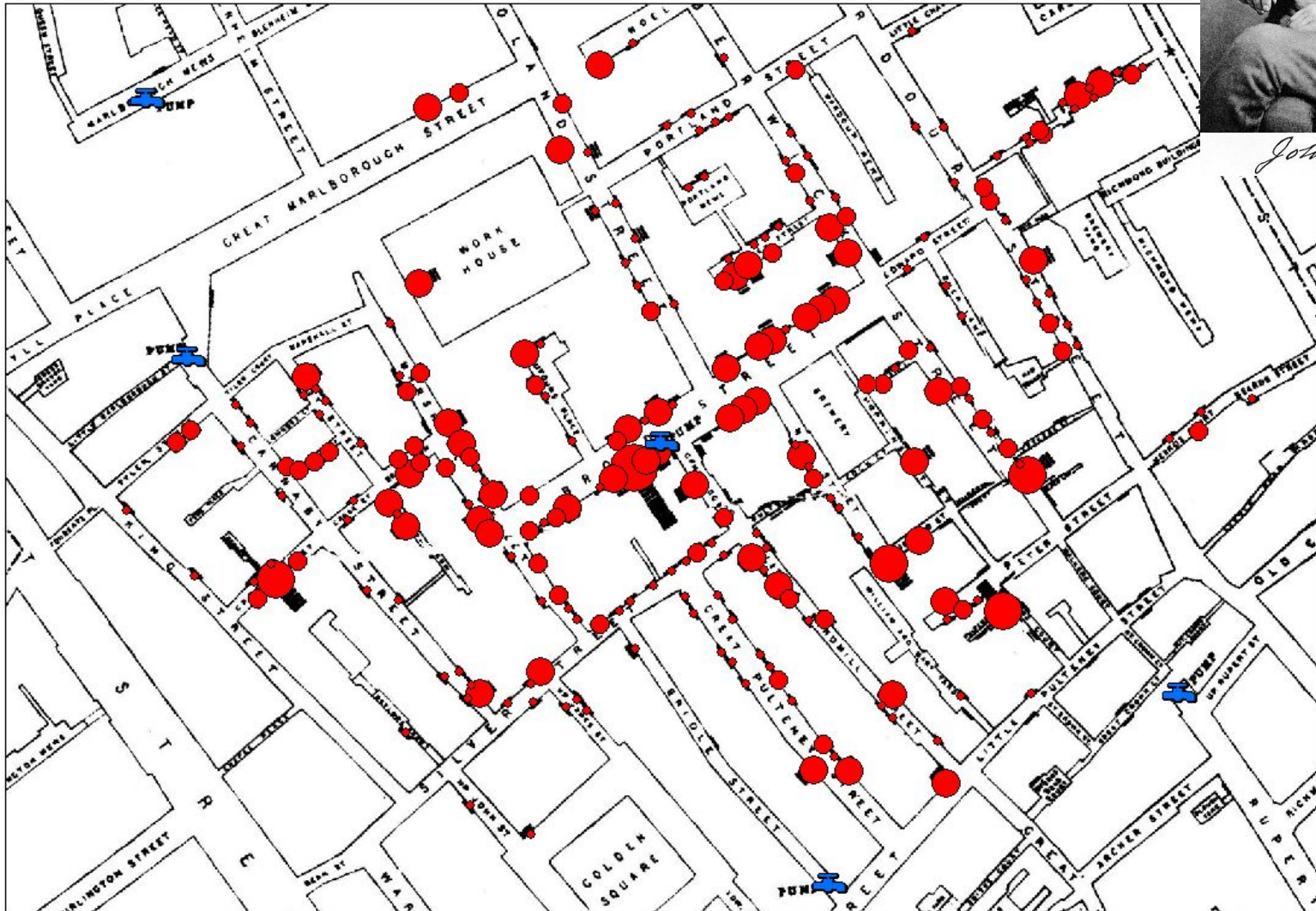
In <u>Sports reporting</u>, data has always been an essential aspect.

<u>John Snow's map of cholera outbreaks</u> from nineteenth century London changed how we saw a disease - and gave data journalists a model of how to work today (link).
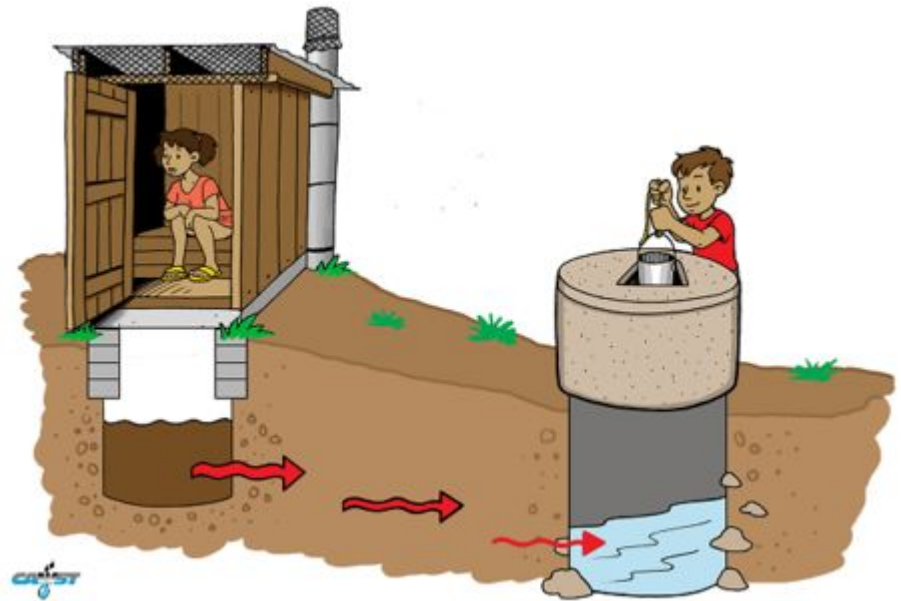
1850' John Snow cholera's map
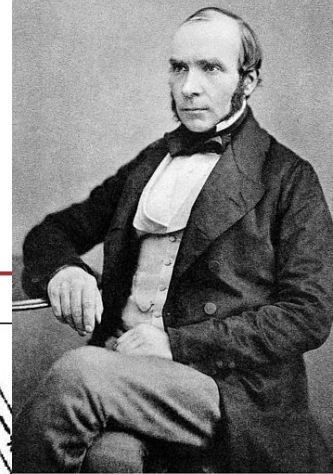
Broadwick Street, showing the John Snow memorial and pub

Air borne Vs water borne deasease

Why Data Journalism is so popular?

- Many governments around the world have published thousands of **open data.**

- on May 2013 Obama started the Open Data Policy: Transparency + Participation + Collaboration
  - **data.gov**
  - **data.gov.uk**
  - **dati.gov.it**

- Many **visualization tools** are available
  often easy to use (Ms excel, tableau, open refine)

Data Journalism Awards

The best work by data journalists is selected during this annual conference

[data journalism awards](#)

- Started in 2012
- The 2019 edition launched in November 2018 closes on 7 April 2019
- **last year 630** works were submitted from **58** countries
- The shortlist announcement was on 13 April 2018 at the [International Journalism Festival](#) in Perugia

Data Journalism Awards [Categories](#)

1-Investigation of the year (large/small newsroom). Best investigative journalism project which sheds light on an important topic of public interest. The entry will be judged based on quality and innovation of data use and analysis, the content created and its presentation, and its impact.

2-Best use of data in a breaking news story, within the first 36 hours. Best data-based journalism around a breaking news story within the first hours of the story breaking.

3-Best individual portfolio. Based on quality of content, frequency and variety of subjects covered by one person.

4. **Award for student and young data journalist** of the year: This award honors outstanding work done by new and young talent, done while still a student or early in their professional career. Awarded to the **best portfolio of data-driven work by an individual aged 27 or younger at the time the work was created**, based on

Data Journalism Awards - Categories

5. **Best data journalism team portfolio: (Large newsroom)** Best portfolio of work by a data journalism team, based on quality and innovation of data use and analysis, content, frequency of updates, range of coverage and variety of approaches.

6. **Best data journalism team portfolio: (Small newsroom)** Best portfolio of work by a data journalism team from an organization with 30 or fewer editorial staff, based on quality and innovation of data use and analysis, content, frequency of updates, range of coverage and variety of approaches.

7 **Data visualisation of the year (large/small newsroom).** Best interactive or static visualisation based on one or several datasets. May be self-contained or combined with a story, but must accomplish a journalistic purpose and use data significantly.

8. **Public Choice Award**: An award for excellence in data journalism voted on by members of the public from the 2018 Shortlist.

Data Journalism Awards - Categories

9. Award for innovation in data journalism: Awarded to the entry that demonstrates the most innovative approach to collecting, analyzing or examining data or to the most innovative approach to presenting data-driven journalism, to shed light on a topic of public interest

10 News data app of the year (large/small newsroom). Best data journalism app. Interactivity is essential and the project should provide both explanation of the topic and user engagement.

11 Open data award. Using freedom of information and/or other levers to make crucial databases open and accessible for re-use and for creating data-based stories.

12 Data journalism website of the year. Best data-based journalism website, based on quality of content, frequency and variety of subjects covered.

Winners of the Data Journalism Awards 2019

1. **Investigation of the year**: <u>Hurricane Maria's Dead (US)</u>

2. **The best use of data in a breaking news story, within first 36 hours**: <u>Indonesia Plane Crash (USA)</u>

3. **Best individual portfolio:** <u>Eva Belmonte (SPAIN)</u>

4. **Student and young data journalist of the year**: <u>Dada Lyndell (Russia)</u>

5. **Best large data journalism team portfolio:** <u>La Nacion (Argentina)</u>

6. **Best small data journalism team portfolio:** <u>Hate Crime Watch (India)</u>

7. **Data Visualization of the year:** <u>The Race to Save the River Ganges (USA)</u>

8. **Innovation in data journalism:** <u>Radmesser (Germany)</u>

9. **News data app:** <u>The Myth of the Criminal Immigrant (USA)</u>

10. **Open data award:** <u>OCCRP Data (Bosnia Herzegovina)</u>

11. **Data journalism website of the year:** <u>The Pudding (USA)</u>

12. **Public choice:** <u>This is how thin you have to be, to walk the catwalk (Netherlands)</u>

Homework

Ognuno studia uno dei 12 vincitori e crea delle slides dove racconta brevemente:

- il fatto dell'articolo
- la menzione della giuria
- i dati visualizzati
- le visualizzazioni realizzate
- le interviste
- osservazioni interessanti

Winners of the Data Journalism Awards 2018

1. Easy Money by The Globe and Mail (**Canada**) won the **investigation of the year**.

2. The search operation of the submarine Ara San Juan by La Nacion (Argentina) won the **the best use of data in a breaking news story, within first 36 hours**.

3. Patrick Stotz from SPIEGEL ONLINE (**Germany**) won the **Best individual portfolio**.

4. Marie-Louise Timcke from the Berliner Morgenpost (**Germany**) won **student and young data journalist of the year**.

Winners of the Data Journalism Awards 2018

1. <u>Caixin VisLab</u> (China) won the **Best large data journalism team**.

2. <u>InfoTimes</u> from Egypt and <u>Monitor de Victimas</u> from Venezuela both won the **Best small data journalism team**

3. <u>Life in the camps</u> by Reuters (**UK**) won the **Data visualisation of the year**.

4. <u>Monitor da Violencia</u> by G1 (Brazil) won the **public choice**.

Winners of the Data Journalism Awards 2018

1. Hidden Spy Planes by BuzzFeed News (**US**), won the **innovation in data journalism**.

2. The Atlas Of Redistricting by FiveThirtyEight (**US**) won the **News data app of the year**.

3. Follow the Money by Postmedia (**Canada**) won the **Open Data**.
   The jury also awarded an Honorable Mention in this category for Yudivian Almeida from Postdata.club (Cuba).

4. #SaferRoadsPH by Rappler (Philippines) won the **data journalism website of the year**.
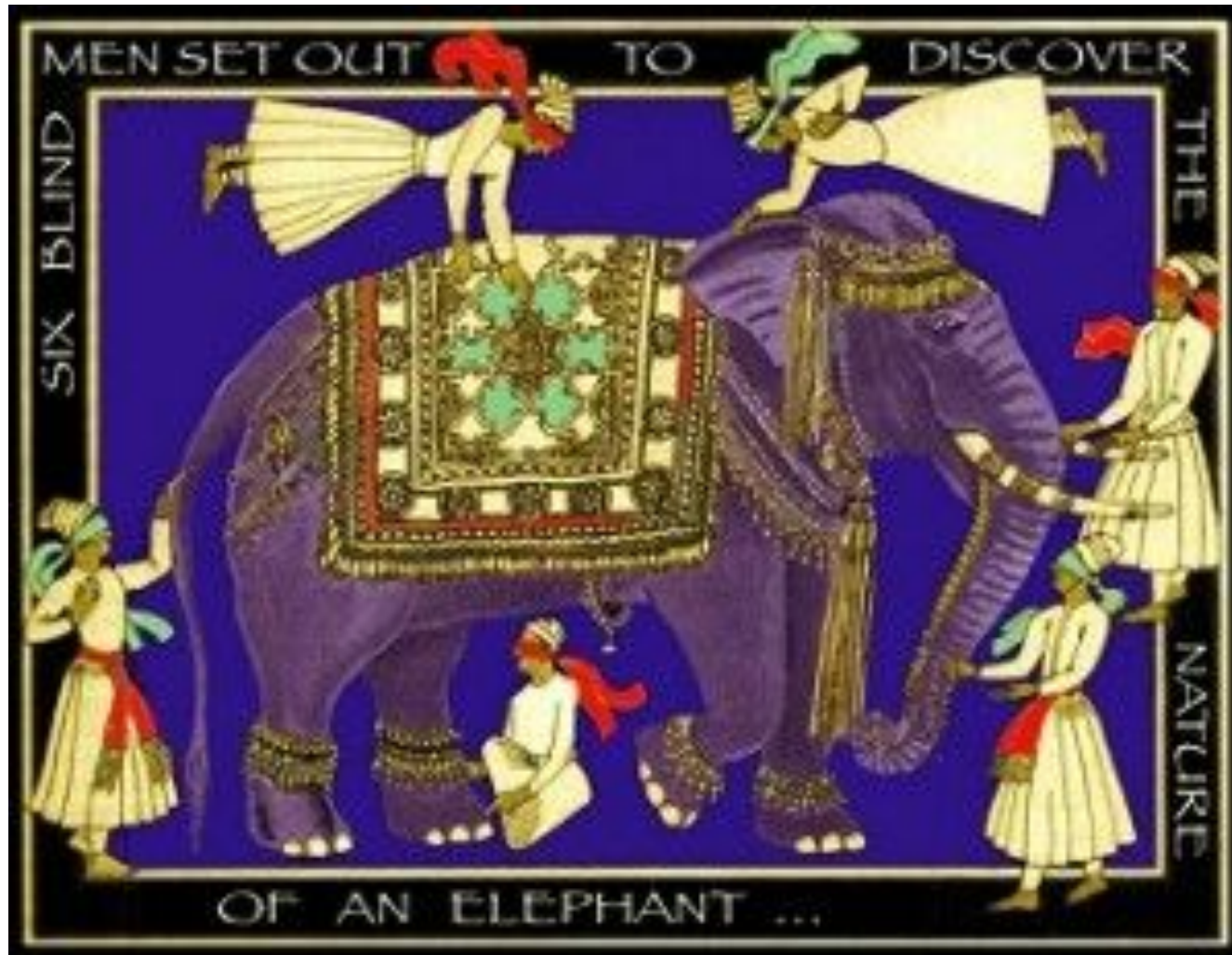
Gambling

Newsroom composition and organization
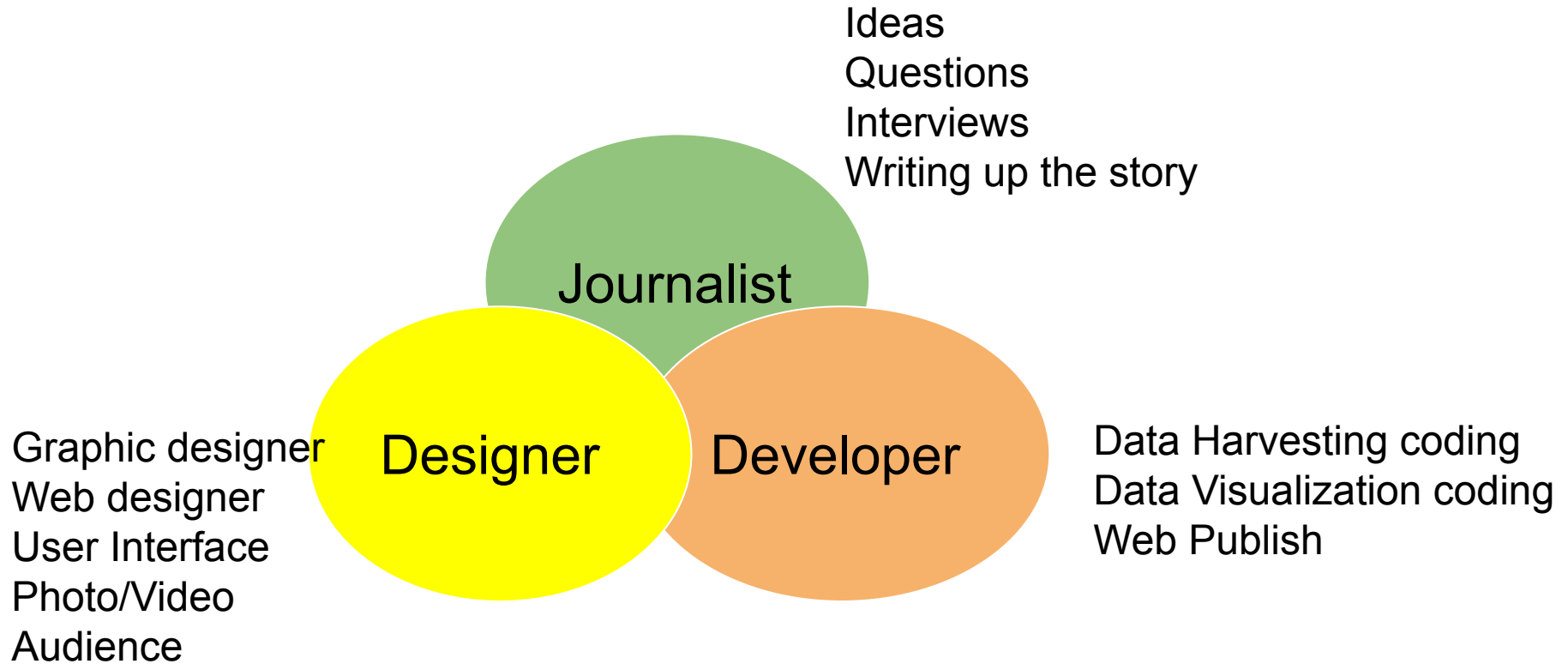
inside a data team

Everyone sees Data Journalism from their own perspective

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**
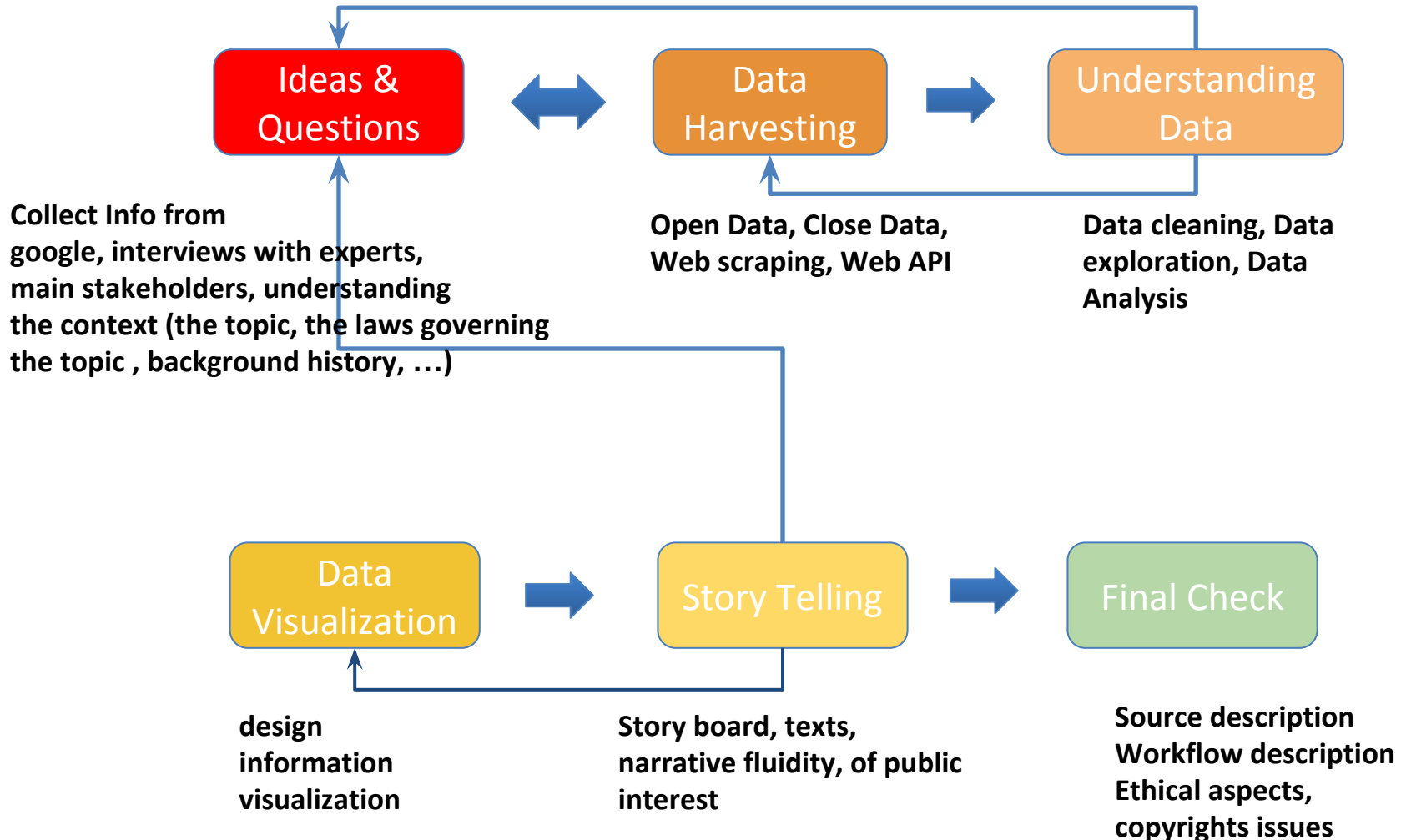
Small DJ Newsroom

Ideas
Questions
Interviews
Writing up the story

Journalist

Graphic designer
Web designer
User Interface
Photo/Video
Audience

Designer

Developer

Data Harvesting coding
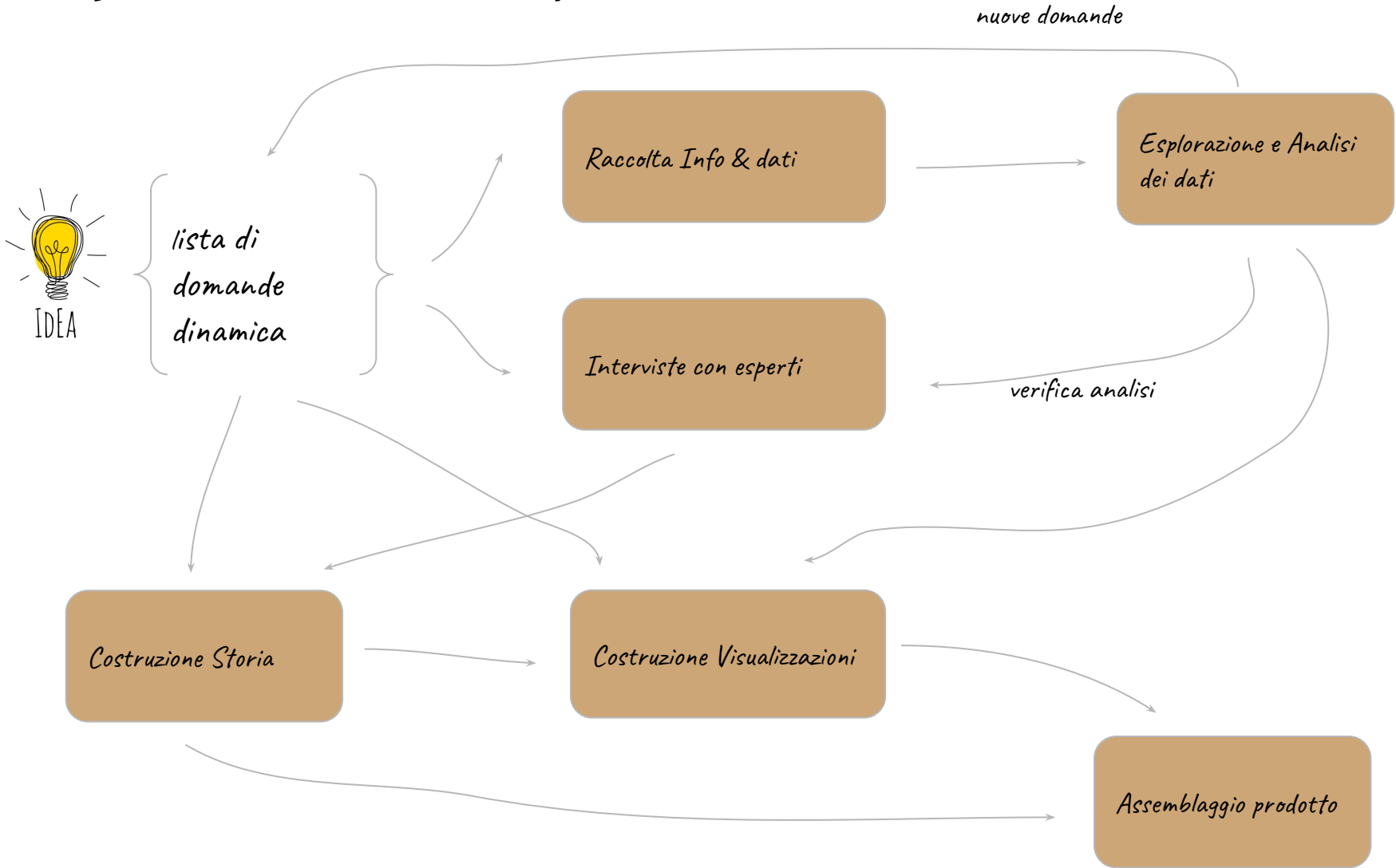Data Visualization coding
Web Publish

Golden Rules to organize the work

- You don't need to immediately compete with the NYT or The Guardian.
- **Visualisation is not always needed**.
- **Start small**: Argentina's La Nacion, the best data journalism site in South America, started without a programmer and by using free software.
- It's not about how many developers you have, but what **you do with them**
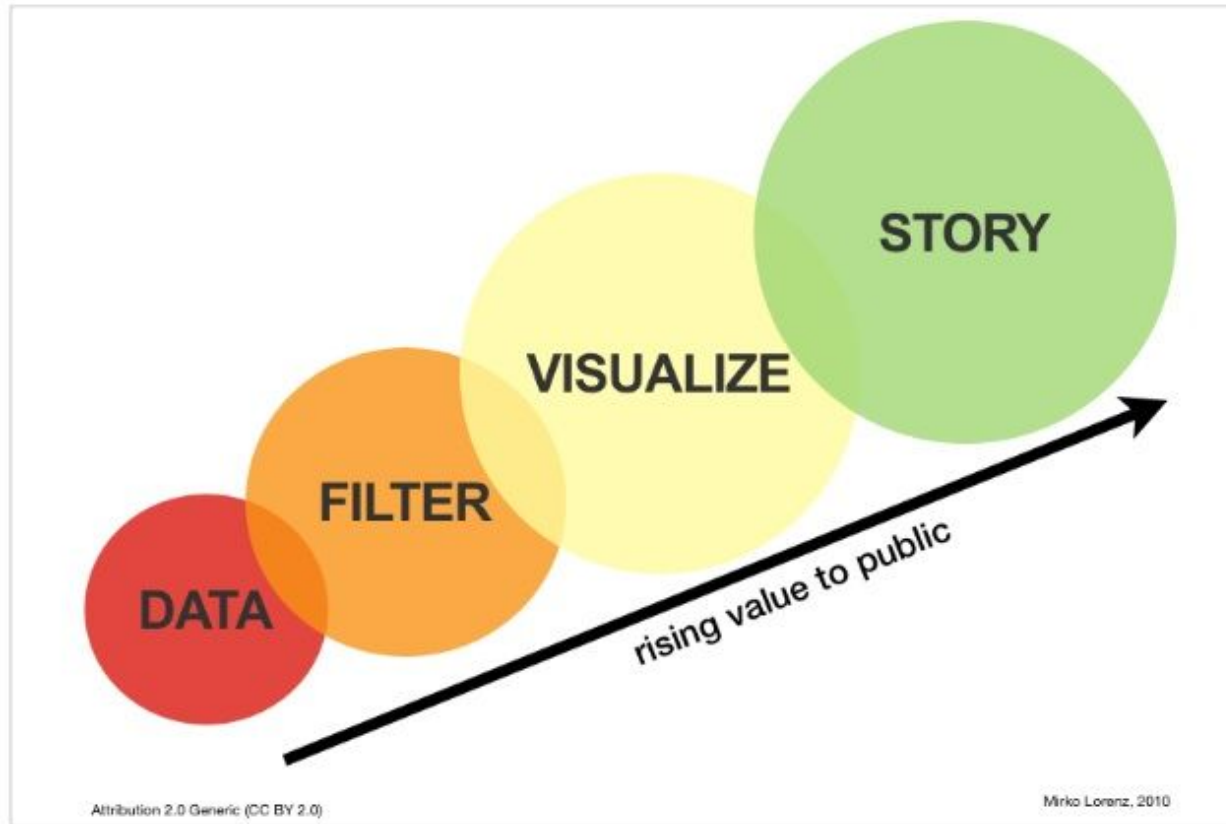
Data Jornalism Workflow

Ideas & Questions ⟷ Data Harvesting → Understanding Data

Collect Info from google, interviews with experts, main stakeholders, understanding the context (the topic, the laws governing the topic , background history, …)

Open Data, Close Data, Web scraping, Web API

Data cleaning, Data exploration, Data Analysis

Data Visualization → Story Telling → Final Check

design information visualization

Story board, texts, narrative fluidity, of public interest

Source description Workflow description Ethical aspects, copyrights issues

Data Journalism Workflow

nuove domande

Raccolta Info & dati

Esplorazione e Analisi dei dati

IdEA

lista di domande dinamica

Interviste con esperti

verifica analisi

Costruzione Storia

Costruzione Visualizzazioni

Assemblaggio prodotto

Data Driven Journalism

Es. wikileaks

The Main Tasks

1. Researching data

2. Writing

3. Coding

4. Designing and visualising

Researching Data

- Most data is organised in a spreadsheet package (Ms Excel, [Google Sheets](), [Apple Numbers](), …)
- Selection of the data is a **key part of what we do**, to tell the story better
- The **fewer NUMBERS you can use to tell the story**, the better
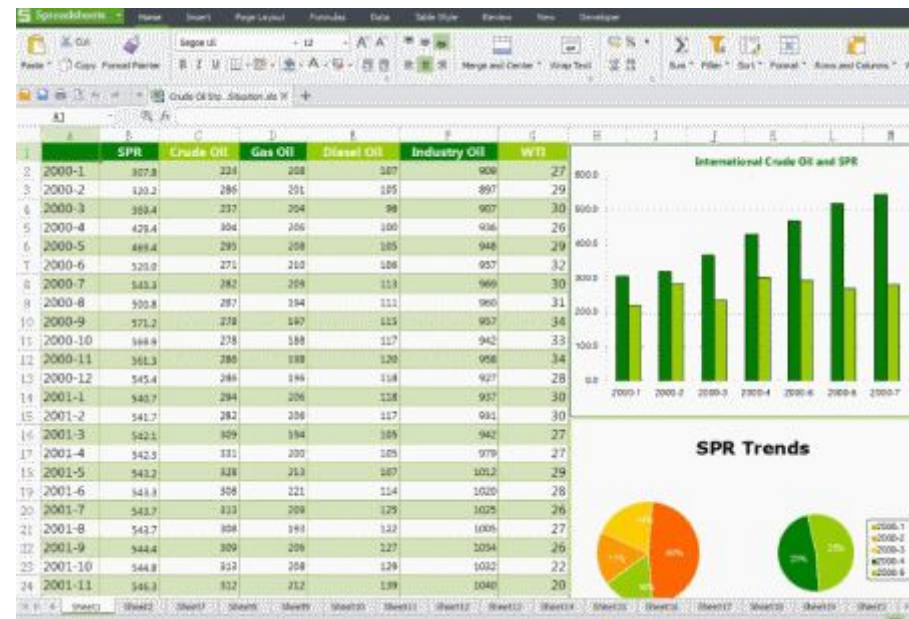
Researching Data

- Most important role of DJ
- Time consuming, tedious and boring
- Data understanding
- Data cleaning (openrefine)
- Data integration
- Data exploration
- Data Analysis

- Crucial to identify if the story is good for news desk
- **Study of previous works on the same subject**
- The interviews to expert can help

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

- Example: Yearly guide to public spending by each government department (The Guardian publishes this guide every year)
- Sometimes words are enough

Coding

- Coders (programmers), can help with the research (by scraping data) and data interactive visualisation (GoogleMaps, D3,…)
- Coordinate who can write and code
- Example: Connected China (Reuters)

Designing and visualising

- Designers **plan** and **create** visualisations
  - (datawrapper, tableau public, GoogleFusion Tables, CartoDB)
- Example: 99% vs 1% (The Guardian)
- DJ Designer
  - Graphic Designer
  - Web Designer
  - User Interface Designer
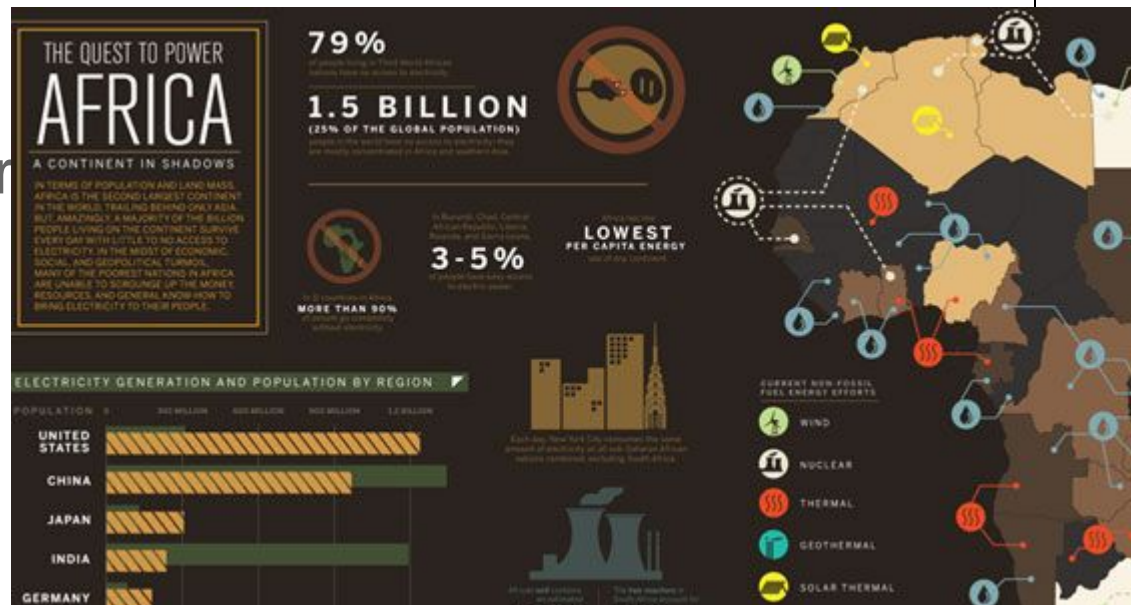
What kind of teams are there?

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
- Part of the newsdesk
- Example: 2012 Olympic Experience (New York Times)

how to transform numbers into stories

time

The initial two possible Scenarios

Quote: *Data journalism begins in one of two ways:*

1. **either you have a question that needs data,** (journalistic)
2. *or* **an interesting dataset that needs questioning** *(data scientist)*

Whichever it is, the **compilation** *(*collection and integration*) of data is what defines it as an act of data journalism.*

-Paul Bradshaw

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).

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
```

The role of interviews

- Understand the domain of knowledge

- Discover interesting stories to tell

- Verify the final job

- Include the interviews inside the story

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)

Data Journalists Gallery

Simon Rogers

Data journalist of **The Guardian**,

Author of '[Facts are Sacred](#)',

Editor of [guardian.co.uk/data](#), an online data resource which publishes hundreds of raw datasets and encourages its users to visualise and analyse them – and probably the world's most popular data journalism website.

Paul Bradshaw

Online journalist and blogger,

Manages the Online Journalism Blog

Co-founder of Help Me Investigate, an
investigative journalism

Author of the Online Journalism Handbook

Alberto Cairo

Information designer and professor

Teaches Visual Journalism at the School of Communication of the University of Miami

Author of The Functional Art book