



## **Module Handbook**

7AAVDM12  
Data Journalism  
2016/17

Convenor:  
Tommaso Venturini and Liliana Bounegru

Credits: 15

## 1. Module Description

In the last few years, the practices of journalism have been shaken by the advent of a new series of digital techniques of investigation and storytelling. On the one hand, the growing availability of digital data opens new ways of enquiring on actions of institutions, companies and individuals. On the other, the advent of new methods of computation and visualization allows new possibility of interactions with the news and exploration of their sources. The ensemble of these, tentative yet promising, evolutions goes under the name of 'data journalism' and will constitute the object of this module.

In this module, students will learn to assume a critical attitude towards data and reflect on the renewed importance of journalistic mediation at the time of the digital networks. Students will also be trained in the fundamentals of data manipulation and experiment several conceptual and technical tools for the cleaning, analysis and representation of digital traces. The importance of embedding findings in an interesting narration will also be discussed with particular reference to two series of techniques for the exploration and representation of data. The first derives from the tradition of visual statistics; the second from the more recent development of visual network analysis.

### Module Aims

- To understand the changing role and the renewed responsibilities of journalists in the age of digital media.
- To learn how to find, extract and refine a set of digital data.
- To become capable to extract meaningful findings from digital data and to turn them into interesting journalistic stories.

### Learning Outcomes

At the end of the module the student will be expected to:

- develop a critical reflection on data and their possible uses in journalism;
- possess the practical skills required to clean and analyse small sets of data;
- understand the principles of visual semiotics and use them to represent diagrams;
- recognise the importance of 'data narration' and be able to produce simple 'data stories'.

### Teaching pattern

In this module the traditional teaching pattern of lectures and seminars will be enriched by a one-day 'immersion' in the situation of a data newsroom. Working in small groups and under the guidance of the module convenors, the students will develop a journalistic project of their choice, analysing and narrating a set of data that they will have collected and cleaned beforehand.

## 2. Module Tutors

Convenor: Tommaso Venturini (and Liliana Bounegru)  
207 Digital Humanities Building (Drury Lane)  
[tommaso.venturini@kcl.ac.uk](mailto:tommaso.venturini@kcl.ac.uk)  
Office Hours: Thursday 11am-13pm

## 3. Module Location and Times

Lecture: Thursday, 14-15 / room S-2.23 Strand Building  
Seminars: Group A: Thursday, 15-16 / room 1.04 (formerly 4A) Chesham Building  
Group B: Thursday, 16-17 / room 1.04 (formerly 4A) Chesham Building  
Group C: Thursday 17-18 / room S-3.19 Strand Building

## 4. Readings

This module employs a ‘flipped classroom’ approach, with most conceptual contents delivered outside the classroom through the readings that will be proposed in preparation of each lesson. This means that the attentive consideration of the suggested material is a necessary condition to participate in the module, particularly by watching the required videos from the online course: <http://learno.net/courses/doing-journalism-with-data-first-steps-skills-and-tools>

Each week has required readings. Please read these texts before arriving to class that week. Alongside the required readings, there are recommended readings. Read as many of these as you find interesting. Take time to look them up. You are expected to read widely on this module, and in your assessment you must refer to a range of relevant readings.

## 5. KEATS

KEATS is essential to this module. All readings, lecture slides, website links and other digital materials will be found on this module’s KEATS page. Your tutor(s) will post information to KEATS throughout both semesters. It is therefore important that you check KEATS regularly.

## 6. Email Communication Policy

King’s communicates to students through their official KCL email address. It is therefore vital that you check your King’s email regularly. King’s staff will only reply to student emails if they are sent through the student’s King’s email address. Please do not use your personal email when emailing your tutors or module convenors.

## 7. Student Participation and Attendance

Active student participation in this module is crucial. Students are expected to contribute to seminars, to develop and articulate their ideas, and to define their position in relevant debates. Students are expected to attend all lectures and seminars across both semesters. If you are unable to attend a session for any reason, please contact the convenor/tutor straight away.

## 8. Weekly Schedule

|        |              |  |
|--------|--------------|--|
| Week 1 | 29 September | <b>Introduction to data storytelling and data journalism</b> |
|--------|--------------|--|

### Lecture: Thu 29 Sep - 14-15: S-2.23 Strand building

- Data storytelling (data do not speak for themselves)
  - The importance of being able to extract stories from data
  - What is data storytelling for and what are its applications
  - Looking at journalism to learn how to tell stories with data
- Journalism and data journalism
  - What journalism is for (the lesson of Walter Lippmann)
  - Types of data journalism: formats, skills and composition of a data newsroom
  - Historical examples of data journalism (eg Detroit riots, Colour of Money, etc.)

### Required reading & homework:

- Watch module 1 (sections 1.1 to 1.3) of Doing Journalism with Data MOOC:  
<http://learno.net/courses/doing-journalism-with-data-first-steps-skills-and-tools>

### Recommended reading:

- <http://viz.wtf>
- <http://datajournalismhandbook.org/1.0/en/introduction.html>
- <http://www.holovaty.com/writing/fundamental-change/>
- <https://www.theguardian.com/news/datablog/2010/oct/01/data-journalism-how-to-guide>
- <https://www.propublica.org/nerds/item/data-driven-journalisms-secrets>
- [http://datajournalismhandbook.org/1.0/en/in\\_the\\_newsroom\\_3.html](http://datajournalismhandbook.org/1.0/en/in_the_newsroom_3.html)
- [http://datajournalismhandbook.org/1.0/en/delivering\\_data\\_2.html](http://datajournalismhandbook.org/1.0/en/delivering_data_2.html)

|        |           |                                      |
|--------|-----------|--------------------------------------|
| Week 2 | 6 October | <b>What makes a good data story?</b> |
|--------|-----------|--------------------------------------|

### Lecture: Thu 6 Oct - 14-15: S-2.23 Strand building

- Avoiding data porn
  - dissecting a bad data story in participation with students
- Showcase and discussion of interesting data stories
  - Story and story angle models from sociology and journalism
  - The “digital methods” angle: telling stories about the politics of digital platforms at the same time as telling stories about how issues unfold on digital platforms, algorithmic accountability angle, issue angle
  - Stories behind the data projects: tools, data, techniques, skills
- Introduction to the case study of this module
  - Amnesty International and Amnesty around human rights violations (cf. <https://wiki.digitalmethods.net/Dmi/SummerSchool2016AmnestyDecoders>).

### Seminar: Thu 6 Oct - 15-16 & 16-17: 1.04 (formerly 4A) Chesham bldg / 17-18: S-3.19 Strand bldg

- Intro by each student (why they joined the course and what they want to get from it)
- Introduction to the course approach (flipped classroom and active participation)
- Group formation
- What counts as a story

### Required reading & homework:

- Watch MOOC module 2 “Finding data to support stories” available here :  
<http://learno.net/courses/doing-journalism-with-data-first-steps-skills-and-tools>
- Find a data story and be prepared to discuss it in class next week.

### Recommended reading:

- <https://source.opennews.org/en-US/learning/finding-stories-structure-data/>  
<http://www.tampabay.com/news/business/how-we-identified-americas-50-worst-charities/2124085>
- [http://datajournalismhandbook.org/1.0/en/introduction\\_3.html](http://datajournalismhandbook.org/1.0/en/introduction_3.html)
- <http://www.knightfoundation.org/blogs/knightblog/2016/6/16/winners-2016-data-journalism-awards-show-growing-quality-promise-field/>
- <http://www.vanityfair.com/news/2016/08/donald-trump-and-the-limits-of-data-journalism>

•

|        |            |                        |
|--------|------------|------------------------|
| Week 3 | 13 October | <b>Collecting data</b> |
|--------|------------|------------------------|

**Lecture: Thu 13 Oct - 14-15: S-2.23 Strand building**

- Collecting data (data-driven VS question-driven approaches)
- Overview of possible data sources
- Overview of methods and interface to data access:
  - Querying
  - Scraping
  - API calls
  - Asking institutions for data (e.g. FOIA requests)
  - ...
- Laws and ethics of data use

**Seminar: Thu 13 Oct - 15-16 & 16-17: 1.04 (formerly 4A) Chesham bldg / 17-18: S-3.19 Strand bldg**

- Groups formation revision
- Discussion of the data stories found by the students
- Help in getting the data collection started

**Required reading & homework:**

- Find and collect an interesting dataset relevant for the Module's case study
- Watch MOOC module 4 "Dealing with messy data" available here :  
<http://learno.net/courses/doing-journalism-with-data-first-steps-skills-and-tools>
- Follow the lesson 1 of Trifacta Wrangle tutorial  
<https://www.trifacta.com/support/articles/article/896169-lesson-1-introduction-to-trifacta-wrangler/>

**Recommended reading:**

- <https://github.com/veltman/learninglunches/tree/master/scraping>
- <http://www.tandfonline.com/doi/abs/10.1080/1369118X.2012.678878>
- [http://datajournalismhandbook.org/1.0/en/getting\\_data.html](http://datajournalismhandbook.org/1.0/en/getting_data.html)

|        |                       |
|--------|-----------------------|
| Week 4 | No lecture or seminar |
|--------|-----------------------|

|        |            |                      |
|--------|------------|----------------------|
| Week 5 | 27 Octobre | <b>Cleaning Data</b> |
|--------|------------|----------------------|

**Lecture: Thu 27 Oct - 14-15: S-2.23 Strand building**

- Keeping a project diary and preparing a data codebook
- Using Trifacta Wrangler (<https://www.trifacta.com/products/wrangler/>)
- Removing duplicates
- Checking for wrong, inconsistent or improbable data points (how to look for problems in data)

**Seminar: Thu 27 Oct - 15-16 & 16-17: 1.04 (formerly 4A) Chesham bldg / 17-18: S-3.19 Strand bldg**

- Revision of the data collected by the students
- Presentation of a data cleaning exercise
- Help in getting the data cleaning started

**Required reading & homework:**

- Carry out the out the data cleaning exercises
- Watch MOOC module 3 “Finding story ideas with data analysis” available here: <http://learno.net/courses/doing-journalism-with-data-first-steps-skills-and-tools>
- Follow the lesson 2A and 2D of Trifacta Wrangle tutorial: <https://www.trifacta.com/support/articles/article/897162-lesson-2-getting-started/>
- Follow the lesson 3 of Trifacta Wrangle: <https://www.trifacta.com/support/articles/article/897165-lesson-3-basic-operations/>

**Recommended reading:**

- <https://github.com/propublica/guides/blob/master/data-bulletproofing.md>
- <https://github.com/veltmann/learninglunches/tree/master/datahygiene>
- <https://infoactive.co/data-design/ch08.html>
- [http://datajournalismhandbook.org/1.0/en/understanding\\_data\\_0.html](http://datajournalismhandbook.org/1.0/en/understanding_data_0.html)
- [http://datajournalismhandbook.org/1.0/en/understanding\\_data\\_2.html](http://datajournalismhandbook.org/1.0/en/understanding_data_2.html)
- <https://www.propublica.org/nerds/item/using-google-refine-for-data-cleaning>
- <http://davidhuynh.net/spaces/nicar2011/tutorial.pdf>

|              |                       |
|--------------|-----------------------|
| Reading Week | No lecture or seminar |
|--------------|-----------------------|

|        |             |                      |
|--------|-------------|----------------------|
| Week 6 | 10 November | <b>Data analysis</b> |
|--------|-------------|----------------------|

**Lecture: Thu 10 Nov - 14-15: S-2.23 Strand building**

- Basic statistical concepts in spreadsheets, pivot tables
- Learn to recognise patterns and anomalies (trends and outliers)
- Basic descriptive statistics: percentages and percent change, measurement units, standard deviation, regression analysis...

**Seminar: Thu 10 Nov - 15-16 & 16-17: 1.04 (formerly 4A) Chesham bldg / 17-18: S-3.19 Strand bldg**

- Revision of data cleaning (both the exercise and the data found by the students)
- Presentation of a data analysis exercise
- Help in getting the data analysis started

**Required reading & homework:**

- Carry out the out the data analysis exercises
- Watch MOOC module 5 “Telling stories with visualisation” available here: <http://learno.net/courses/doing-journalism-with-data-first-steps-skills-and-tools>
- Learn how to create charts in Excel here: <https://support.office.com/en-US/article/Create-a-chart-from-start-to-finish-a745775f-98d9-4c63-bfa8-9c00cd03ff0c>

**Recommended reading:**

- <http://kpg.github.io/spring-2015/classes/git-and-excel/>
- [http://datajournalismhandbook.org/1.0/en/understanding\\_data\\_5.html](http://datajournalismhandbook.org/1.0/en/understanding_data_5.html)
- <https://source.opennews.org/en-US/learning/finding-stories-structure-data/>

|        |                       |
|--------|-----------------------|
| Week 7 | No lecture or seminar |
|--------|-----------------------|

|        |             |                                 |
|--------|-------------|---------------------------------|
| Week 8 | 24 November | <b>Turning data into images</b> |
|--------|-------------|---------------------------------|

**Lecture: Thu 24 Nov - 14-15: S-2.23 Strand building**

- Principles of visual semiotics
- Visualisation patterns (bar charts, line charts, pie charts, scatter plots, networks...)
- Avoiding rookie mistakes and misleading visualisations
- Use Excel charts for exploratory data analysis
- Showing missing data and errors

**Seminar: Thu 24 Nov - 15-16 & 16-17: 1.04 (formerly 4A) Chesham bldg / 17-18: S-3.19 Strand bldg**

- Presentation of a data visualisation exercise
- Help in getting the data visualisation started
- Pitch a data story
- Getting ready for the workshop

**Required reading & homework:**

- Prepare the workshop

**Recommended reading:**

- <https://hbr.org/2013/03/power-of-visualizations-aha-moment>
- <http://www.ericson.net/content/2011/10/when-maps-shouldnt-be-maps/>
- <https://github.com/veltmann/learninglunches/tree/master/maps>
- <https://lh3.googleusercontent.com/-7VHZu1TnnOg/UjdiN9p5NqI/AAAAAAAAAKjg/2kEWW2-TNIA/w1280-h920-no/WntrM6p.gif>
- <https://infoactive.co/data-design/ch14.html>
- <http://vis.stanford.edu/files/2010-Narrative-InfoVis.pdf>
- <http://datajournalism.stanford.edu/>
- [http://guides.library.duke.edu/datavis/vis\\_types](http://guides.library.duke.edu/datavis/vis_types)
- <https://source.opennews.org/en-US/learning/visual-literacy-age-data/>

|        |              |                          |
|--------|--------------|--------------------------|
| Week 9 | 1-2 Decembre | <b>Hands on workshop</b> |
|--------|--------------|--------------------------|

**Seminar: Thu 1 Dec - 15-16 & 16-17: 1.04 (formerly 4A) Chesham bldg / 17-18: S-3.19 Strand bldg**

- Preparation for the workshop

**Data newsrooms one-day workshop: Fri 2 Dec - Strand Anatomy Museum**

**Required reading & homework:**

- Finalise the project carried out during the workshop and get ready to present (and defend) them.

**Recommended reading:**

- [http://datajournalismhandbook.org/1.0/en/in\\_the\\_newsroom\\_6.html](http://datajournalismhandbook.org/1.0/en/in_the_newsroom_6.html)
- Venturini, Tommaso, Anders Munk, and Axel Meunier. 2016. *Data-Sprint: A Public Approach to Digital Research*. Edited by Celia Lury, et al.. *Interdisciplinary Research Methods (Forthcoming)*. Available here [http://www.tommasoventurini.it/wp/wp-content/uploads/2016/08/Venturini\\_Munk\\_Jacomy\\_2016-DataSprints.pdf](http://www.tommasoventurini.it/wp/wp-content/uploads/2016/08/Venturini_Munk_Jacomy_2016-DataSprints.pdf)

|         |            |                   |
|---------|------------|-------------------|
| Week 10 | 8 Decembre | <b>Conclusion</b> |
|---------|------------|-------------------|

**Lecture: Thu 8 Dec - 14-15: S-2.23 Strand building**

- Presentation, discussion and feedback on workshop results
- Course wrapping up

**9. Schedule Summary**

|              |              |   |
|--------------|--------------|---|
| Week 1       | 29 September | Introduction to data storytelling and data journalism |
| Week 2       | 6 October    | What makes a good data story?                         |
| Week 3       | 13 October   | Collecting data                                       |
| Week 4       |              | No lecture or seminar                                 |
| Week 5       | 27 October   | Cleaning data   |
| Reading Week |              | No lecture or seminar                                 |
| Week 6       | 10 November  | Data analysis   |
| Week 7       |              | No lecture or seminar                                 |
| Week 8       | 24 November  | Turning data into images                              |
| Week 9       | 1-2 November | Hands on workshop                                     |
| Week 10      | 8 December   | Conclusion  |

**10. Assessment**

**4,000 words project essay (100% of the final grade):**

Drawing on the project you carried out through the module and in particular in the final hands on workshop, write a 4,000 words journalistic story.

Essays must be submitted via KEATS.

**11. Essay Deadlines**

Project essay 1 (4,000 words): To be decided

**12. Essay Marking Criteria**

Marking criteria for Level 5 modules are set out in the Faculty of Arts & Humanities’

Undergraduate Generic Marking Criteria framework. See full details at

<http://www.kcl.ac.uk/artshums/study/handbook/sguides/assessment/ugmarkcrit.pdf>.

**80 – 89 Outstanding (First / Grade A+)**

Highly thoughtful answer informed by wider reading, showing clarity of thought, personal insight and originality.

- Understanding: Thorough understanding of the relevant material and issues with evidence of evaluation in the discussion. Independent and critical evaluation.



- Depth of Knowledge: Full range of sources used and applied in a highly focused manner.
- Structure: Clear and fluent style. Very well focused and structured.

**70 – 79 Excellent** (First / Grade A)

Thoughtful answer informed by wider reading showing clarity of thought and personal insight.

- Understanding: Thorough understanding of the relevant material and issues demonstrating insight and a good level of evaluation.
- Depth of Knowledge: Comprehensive range of relevant literature, evidence is used to support arguments, awareness of wider issues.
- Structure: Clear, logical and integrated presentation.

**60 – 69 Good** (Upper 2<sup>nd</sup> / Grade B)

Good understanding of basic principles and relevant evidence, with a coherent and logical argument.

- Understanding: Good understanding of the relevant material and issues with development of analytical thought.
- Depth of Knowledge: Good use of relevant literature.
- Structure: Coherent, well organised and logical presentation.

**50 – 59 Fair** (Lower 2<sup>nd</sup> / Grade C)

Sound understanding demonstrated with some analysis.

- Understanding: Sound understanding of the relevant material and issues is demonstrated, evaluative thought is apparent in some areas.
- Depth of Knowledge: Appropriate reading is demonstrated to support the discussion.
- Structure: Clearly presented but little development.

**40 – 49 Pass** (Third / Grade D)

- Understanding: General knowledge demonstrated but the work is mainly descriptive.
- Depth of Knowledge: Sparse coverage of basic literature / material. Low quality in a number of areas and poor range of reading.
- Structure: Adequate presentation. Some unclear sections.

**33 – 39 Fail**

- Understanding: Some knowledge but does not focus on the question or is very limited. Descriptive work with little recognisable analysis.
- Depth of Knowledge: Inappropriate literature/material used in assignment. Key texts missing.
- Structure: Disorganised/unclear presentation with loose ends.