

# Best Practice for Data-Journalism Training

Felix Irmer  
Leipzig University

Lessons Learned and Reflections from the Teachers & Trainers' Day at the European Investigative Journalism Conference & Dataharvest 2018.

Teachers & Trainers' Day Organization: Nils Mulvad, John Bones, Nicolas Kayser-Bril and Brant Houston. Institutional Support: Journalismfund.eu & The Investigative Journalism Education Consortium.

## Contents

<b>About this Report</b>	<b>1</b>
<b>Surveying the field</b>	<b>1</b>
<b>Curriculum Design</b>	<b>2</b>
General Considerations . . . . .	2
Conceptual Questions . . . . .	2
Sample Curricula . . . . .	2
American University, Washington D.C. . . . .	2
Gothenburg University, Sweden . . . . .	2
Leipzig University, Germany . . . . .	2
Danish School of Media and Journalism, Denmark . . . . .	3
Nicolas Kayser-Bril, Data-Journalism Trainer	3
Collaborations . . . . .	3
<b>Conducting Trainings</b>	<b>3</b>
Pre-Training Arrangements . . . . .	3
Technical and Staff Preparations . . . . .	3
Room Structure . . . . .	3
Announcing the Training . . . . .	3
Assessing Participants' Skill-level . . . . .	4
Choosing your Dataset . . . . .	4
Conducting a Training Session . . . . .	4
How to get started . . . . .	4
Best-Practice for Teaching . . . . .	4
Strong Focus on Applicability . . . . .	5
Make it a Fun Experience . . . . .	5
Group-Work . . . . .	5
Different skill-levels in a group . . . . .	5
Working in a 2nd language . . . . .	6
Post-Training . . . . .	6
Keeping in Touch . . . . .	6
Training Performance Evaluation . . . . .	6
<b>Resources</b>	<b>7</b>
Surveying the Field . . . . .	7
Teaching . . . . .	7
Workshop Presentations . . . . .	7

## About this Report

The Teachers & Trainers' Day at the European Investigative Journalism Conference & Dataharvest 2018 brought together data journalism trainers and lecturers from Europe and the United States. This report summarizes structural and practical best-practice for data-journalism training that emerged during the one-day seminar. Findings have been aggregated throughout the day across workshop presentations, knowledge exchanges and discussions of the approximately 30 participants.

The sessions of the Teachers & Trainers' Day were organized by Nils Mulvad, John Bones, Nicolas Kayser-Bril and Brant Houston.

Journalismfund.eu provided the platform for the seminar by allowing the Teachers & Trainers' Day to take place as a pre-event to the European Investigative Journalism Conference & Dataharvest 2018.

## Surveying the field

- A multitude of training formats and topical foci exist and have been exchanged during the seminar. First studies have been undertaken to list and structure existing training offers and their structures.
- Houston and Mulvad (2018), Heravi (2018) and Berret and Phillips (2016) have done assessments of the data journalism education field with a focus on academic training.
- Bradshaw (2018) comments on data-journalism training styles in relation to varying composition of student groups and training programs.
- At the seminar, it was agreed that the field could profit from a more extensive mapping of training programs and their contents.

## Curriculum Design

### General Considerations

- Most trainings take place at university level (undergraduate and graduate level), at journalism conferences or in a professional settings such as a newsroom.
- It's important to realize that there is not one training that fits everyone. Different groups and settings require different training approaches.
- Generally, do not assume that students have a high affinity with numbers. As an example, consider that nine out of ten journalism students in the Nordic countries have studied a humanities degree before entering journalism school.

### Conceptual Questions

- The required data-training also depends on the planned implementation of data-journalism at the journalist's workplace. Today, many newsrooms are interested in digital and data journalism, but the degree of its implementation varies tremendously throughout the industry. In some settings, elaborate data teams will be build up and in others, the trained data-journalist will very likely remain a lone wolf.
- There are different approaches to teaching data journalism. Before setting up a curriculum it's important to think about which approach is the most appropriate for the goal of your training and the people you will be teaching.
- Most training institutions start data training on a spreadsheet level and move on to cover all steps that are required to produce a data-story. More specialized approaches focus on one of the steps, most commonly coding and the connected mindset. A challenge for many trainers in this context is to determine when advanced data training should start and which components it should include.
- Skills that are taught in the context of data-journalism training include information gathering, spreadsheets, databases, statistics, mapping, coding for scraping, data processing.
- After a data training, participants should be able to think critically about data, verify sources, bulletproof data, interpret results correctly and also translate them into stories and graphics.
- When training journalist to work at the intersection with developers, it can make sense to also include contents focusing on cultural differences and IT production logics. Vice versa, it can make sense to prepare coders for journalistic norms and the work environment to be expected. Examples in this context include the fact that a journalist will aim to produce a final product without striving for publishable beta versions in-between.
- There is a natural difference between academic training and practical training. University students are usually being taught a mix of statistics/quantitative methods combined with an overview of available tools and approaches for data-story production. For practitioners, the trainings focus on the latter and usually exclude the more abstract and academic parts.
- Programming itself is not required for every data story and quite often there is the need for very basic data skills. For a writer, a spreadsheet can already be associated with a different kind of thinking.
- Entry-level courses aim to build a general understanding of data and numbers with most trainers starting with Excel Some trainers focus solely on more advanced techniques.
- Among the more specialized classes, data visualization classes tend to pose a challenge to more traditional journalism educators. Cooperations with external visualization specialists are popular. Visual design and maps usually resonate well with students and deserve attention in terms of a curriculum.

### Sample Curricula

**American University, Washington D.C..** Data-journalism courses are offered as mixed classes for undergraduate and graduate students. Undergraduate students need to take an analytics class beforehand (Excel & numbers), graduate students have to go through an Excel bootcamp as preparation. Partnering with computer science and environmental sciences departments allows for specialized courses in python, data science, GIS and remote sensing.

**Gothenburg University, Sweden.** The data journalism education is based on a modular scheme and progresses through the following stages: Collect data, verify data, clean data, query data, find story, communicate story.

**Leipzig University, Germany.** A new MSc Journalism combines modules from the department of journalism with classes from computer science and sociology to provide a mix of skills needed by journalists in a digital and data-rich environment. As part of the computer science education, students take classes in "Algorithms and Data Structures", "Modelling and Programming" as well as two more electives incl. "Principles of Visualization" or "Text Mining".

### **Danish School of Media and Journalism, Denmark.**

As part of a journalism degree scheme, students get trained two times in basic and advanced data-journalism. Contents for the basic data-journalism training include Excel, data-visualization, statistics and survey design.

**Nicolas Kayser-Bril, Data-Journalism Trainer.** Nicolas has recently switched to only offering trainings in python programming under the label of “data journalism”. Given a challenging environment for data education, the focus on coding allowed him to create the right expectations among management as well as participants.

### **Collaborations**

- Establishing cooperations with practitioners is common and has proven especially popular among universities. The main challenge in this context is usually to identify trainers with the needed knowledge in combination with a critical data mindset.
- Cooperation with publishers and media houses also occur frequently. By arranging such a partnership beforehand, trainers can be recited more easily and a platform for the first datastories of the participants can be offered. Especially when training foreign journalists, e.g. in Bangladesh, links to local outlets have helped trainers and participants.
- In an academic setting, collaborations help to combine journalism and data-skills with specific topical or methodological expertise from other departments. In this context, it has also proven effective to integrate data journalistic tools and methods in not related courses to raise the awareness among students for the capacities and problem solving abilities of data journalism.

### **Conducting Trainings**

#### **Pre-Training Arrangements**

##### **Technical and Staff Preparations.**

- Make sure that all technical requirements are met in advance of any training.
- When coordinating a team of trainers make sure that everyone involved knows what teaching style is expected and what the focus of the training will be. Approach instructors ideally multiple times and be available for questions and clarifications. Ensure that everyone follows the same concept.
- Additional training staff can enhance the training experience considerably for the instructor as well as the

students. Ideally, one or more co-teacher(s) can constantly walk around in the classroom to check on participants and help with assignments and an IT technician can be on call for network and software issues.

##### **Room Structure.**

- Put some thought into the classroom setup. Ideally, the room setup and structure should allow the instructor to interact easily with students and move around in-between participants to check on their progress.
- An ideal room setup allows to focus the attention of all participants. Elaborate setups include outward facing workstations and a space for instruction in the center. In such a setup, students turn inwards for instructions and turn outward to work on assignments.
- Among the most popular table setups is a “U-shape”. Traditional classroom setups with all students facing the teacher are not considered very effective.
- Being able to see the screens of students at all the time is desirable.

##### **Announcing the Training.**

- When announcing training session in conference programs, clearly indicate if a session will consist of hands-on training or a traditional lecture/demo. If possible, avoid mixing these two formats.
- If technical skills are taught, session descriptions need to be transparent and understandable for people without tech-affinity. Too technical language is usually not understood by practitioners without data experience.
- Be aware that the naming of training programs and sessions can have a direct effect on attracted participants and their motivation. Especially when training experienced professionals, more advanced-sounding names have proven helpful (Consider: “Data Journalism ACADEMY” vs. “Introduction to Data Journalism I”).
- Especially when conducting in-house trainings, participants need to be focused on the workshops. Interaction with regular everyday tasks and work contacts should be minimized where possible. Necessary arrangements need to be agreed upon before starting the training.
- Voluntary participants come with much higher motivation. In contrast, students need to be kept engaged and involved. Especially when the training is part of a degree scheme or with large groups being required to take the training, do not assume that students understand data and are keen on doing elaborate data investigations.

### Assessing Participants' Skill-level.

- Usually, it is hard to get a realistic picture of the group to be trained in advance to the training. The skills of participants and their self-perception very often do not align. You will have to rely on your own skill assessment of future participants most of the time.
- Depending on your audience, you can consider assigning pre-training online modules. However, experiences with these are mixed and often exercises do not get completed. In academic setting and long-term courses, e-learning has proven successful for some.
- Inquiring for participants' skills via a pre-training online surveys has yielded mixed results since students overestimate their ability or even lie (especially common with professionals).
- Most common is a skill assessment on the spot and during the course. As a first step, participants can be asked to introduce themselves and also tell something about their previous computer experiences (e.g. tasks they often do with computers, like to do, are most proud of, ...)

### Choosing your Dataset.

- Data sets should be matched to the goal of the training.
- You should know your training data set in and out. If you do not know the training data set well, it's not a good training data set. Never do improvised data training on the fly.
- Be aware of the fact that too narrow datasets only allow for a few approaches.
- Choosing a manifold data-set that you can adjust to your audience saves you the need of dealing with new data sets constantly. The chosen data-set should allow you to customize it depending on where and to whom you are giving a training. An example for such a dataset would be the foreign direct investments of a country split by at least country and sector.
- Questioning data and sources should be included in your exercises: Your data-set ideally also contains flaws in order to encourage a critical mindset.
- When working with practitioners, a data set can be chosen that enables participants to walk away from the training with a story that they can publish.

### Conducting a Training Session

#### How to get started.

- Before kicking things off make sure that all participants know what to expect of the training including the structure of the day.
- Consider your audience and choose the introduction accordingly. When training whole newsrooms be aware of the different hierarchy levels in the room. In such as setting, you have the opportunity to impact the overall attitude towards data journalism.
- When facing skeptical or unmotivated groups, it can help to provide some perspective by illustrating use cases for data-journalism skills at the very beginning. Impressive and engaging stories can be shown before diving into more technical parts. With university students, it can make sense to review data-journalism job postings with them to get a feeling for the digital skills expected by future employers.
- It's recommended to start the training on a basic level to allow yourself to get a feeling for the skill-level of participants and adjust the following contents if needed.
- Especially when training practitioners you should progress fairly quickly from lecturing to the first practical tasks. Also younger students can be quite eager to apply knowledge quickly.

#### Best-Practice for Teaching.

- The ideal scenario consists of 20% own talking and 80% of the times participants talking and applying skills. Data journalism training often runs the risk of reversing this ratio.
- When teaching data-journalism, make things as tangible and as simple as possible. Don't try to impress the audience with your technical skills.
- Ideally, all case studies and assignments should engage students with especially the exercises allowing them to take ownership of the data and practice the learned skills. Immediate results help students to learn concepts.
- The right pace is critical. Often, when an instructor thinks she or he is going slow, the instructor is not going slow enough. The goal is not to finish the planned lesson plan, but to impart the maximum knowledge possible.
- Three skills learned thoroughly are better than six learned incompletely.

- Having students following along, simply mimicking the instructors clicks without understanding the purpose is not learning.

### **Strong Focus on Applicability.**

- Make data journalism tangible: Show engaging stories and best-case examples and maintain a journalistic story production focus from the start. Create a strong connection between data storytelling and the concepts/techniques to be learned.
- A clear focus should be placed on case-studies, how the learned contents can be used to tell stories with data.
- Try to make concepts tangible for students by working with data that everyone can relate to. Ask for participants' height or individual living expenses and work with them in demo calculations (e.g. to explain standard deviation).
- Practical example: Journalistic thinking about data can be practiced with an assignment that requires the formulation of possible headlines and story-ideas after having looked at a dataset.
- Practical example: In order to practice the transfer from data analysis to storytelling, assignments can require reducing numbers in reporting and converting quantitative insights into textual images and comparisons that are easily understood by readers.
- Practical example: Students can be asked to convert a text into variables to make them understand machine readability and associated data issues.
- Practical example: After learning python and how to program a scraper immediately progress to scraping data.

### **Make it a Fun Experience.**

- Playful and competitive element can help to focus students and add some excitement especially for younger groups. Participants can be asked to formulate questions for the rest of the group. Whoever presents the best working solution gets rewarded.
- Practical example: At the American University in Washington, D.C., "Query Cash" banknotes are used to reward students for good questions in so-called query-challenges.
- Practical example: Learned contents can be reinforced in creative ways. At the University of Texas at Austin, a parody of Justin Bieber's song "Love Yourself" retitled "Learn Excel" was used to conclude an excel

class. At the American University, Washington D.C., a "SQL Song" has been composed to the tune of Doe, a deer.

- Practical example: At the Danish School of Media and Journalism, an ASCII Music video done in Excel is used before diving into data journalism.

### **Group-Work.**

- In theory, the more often a one-on-one training situation can be created, the better. Given common group sizes, often compromises will have to be made.
- Pairing up people at multiple stages of the session has proven a very successful measure for breaking up solo learning experiences. Students can work together to apply learned concepts or focus on new contents. At the same time, the teacher has a chance to walk around and monitor overall progress and potential issues that deserve attention.
- Practical example: Let students decode HTML code (without knowledge) and have them decide which parts to scrape.
- Practical example: Post-it notes can also be used to break up the solo learning experience. In one class, green and red post-it notes are handed out to students to allow them to signal how they can follow along. As a result, a group feeling for the pace and contents can emerge and co-trainers can immediately support people running into difficulties.

### **Different skill-levels in a group.**

- Remember to cater to all participants with their respective skill-levels.
- Be aware that especially in groups with mixed skill-sets it can be difficult to find a right balance between instruction and exercises.
- It is okay that people fall behind, especially in more advanced classes. In this case, more advanced students can be asked to use their skills to help others.
- Do not get lost in advanced (interesting) questions that only a fraction of participants can follow. Refer to the respective follow-up class or deal with advanced questions individually.
- Rich datasets can help you to keep even more advanced-students busy while keeping them in the same environment as the rest of the group.
- Visualization exercise: Having students sketch graphical concepts and possible data visualizations on paper is a great workaround to different levels of graphic design software knowledge and practice.

### **Working in a 2nd language.**

- When teaching abroad, pay particular attention to the teaching culture that is common and adjust your training accordingly. Talk to people who know the customs at the teaching destination.
- Be aware of the time needed to process a second language. Details and explanations sometimes get lost and things take longer. In such a setting, a “show more, tell less” approach can be helpful.
- Also pay attention to different functionalities and shortcuts of software depending on the OS and language settings.

### **Post-Training**

#### **Keeping in Touch.**

- It is advisable to make some time available for a good discussion of potential data and stories to be pursued following the training. After completing the data training, a quick application of the learned skills by the participants is key.
- Often it is difficult for students to find the right data for their first project. Ideally, instructors should point training participants to suitable data sets for their next projects and be available for basic support in the first weeks after the training.
- To foster application, it has proven successful to keep in touch with the respective group. Groups on social media provide a good platform to interact with participants and briefly answer any post-training questions.
- At the conclusion of your training, participants can be referred to websites with further training materials or information on how to specialize a certain skill.
- Courses materials and tip sheets can be printed on very bright paper (e.g. in neon colors). This way, handouts stick out and get found easily in a post-workshop setting. (This applies especially to journalists with desks covered with stacks of papers.)
- After students complete the training they usually either go on and become data-journalists or end up in roles that do not require data-skills frequently. In the latter case, it is still considered important that they have received a general data literacy and data-journalism process knowledge training.

### **Training Performance Evaluation.**

- Possible training success indicators include data story output by participants as well as the frequency with which they use data skills in their future assignments.
- However, usually resources are limited and little evaluation in the form of surveys or other structured analysis is done in the post-training phase.
- It is common to practice an informal success evaluation by keeping in touch with participants via e.g. groups on social media or the previously used electronic learning platform.
- With regards to technical skills, a test can be taken as final assignment of the training to ensure that participant reached a certain skill-level. Such a graded assignment and a course certificate can also generally help to engage people through the training session.

## Resources

### Surveying the Field

- Brant Houston, Nils Mulvad (2018). Training Data Journalism in Europe. Kaas & Mulvad.
- Bahareh R. Heravi (2018). 3WS of Data Journalism Education. Journalism Practice, DOI: 10.1080/17512786.2018.1463167
- Google News Lab (2017). Data Journalism in 2017: The Current State and Challenges Facing the Field Today. <https://newslab.withgoogle.com/assets/docs/data-journalism-in-2017.pdf>

### Teaching

- Paul Bradshaw (2018). Data Journalism Teaching, Fast and Slow. Asia Pacific Media Educator, DOI: 10.1177/1326365X18769395
- Charles Berret, Cheryl Phillips, C. (2016). Teaching data and computational journalism. New York, NY: Columbia Journalism School.
- Jenny Rogers (2007). Adults Learning. Berkshire, United Kingdom: Open University Press.

### Workshop Presentations

- Helena Bengtsson. Oral presentation on practical training experiences.

- John Bones. To train the trainers.
- Crina Boros. On a DDJ Trainer's Mental Health.
- Anuska Delic. The Granular Approach.
- Jan Dyberg. "How far shall we go? Which elements do we need to train? What is needed as a minimum – and what is for the needs?"
- Sandra Foresti. Share assignments and exercises.
- Brant Houston and Jennifer LaFleur. Data Science + Journalism.
- Felix Irmer. The new Master of Science in Journalism at Leipzig University.
- Nicolas Kayser-Bril. Oral presentation on training experiences.
- Jennifer LaFleur. Training and Teaching Data Journalism.
- Nils Mulvad. Status of training datajournalism in Europe.
- Jonathan Stoneman. Oral presentation on practical training experiences.