Transparency International
Open Data Training

Overview

School of Data proposes to organise a 5 day data training for the members of the Transparency International Network followed by 3 months of intensive support from the School of Data team. The initial data training will take place in the second or third week of October 2017 followed by support through January/February 2018 (depending on December availability).

The training should be made up of three main components:

1. **Pre-training Assessment:** this aims to gain more clarity on each participant's data skills and what they expect from the training. This will also aim to understand the various projects each participant works on and how they foresee using their data skills on these projects or future projects. This will help tailor the structure of the training. This assessment can be done through a pre-assessment survey and/or 30-minute calls with each participant. This assessment phase is critical for us to understand the actual level of the participants and their data literacy needs. While many people will self-report advanced skills and an interest in learning R and/or Python, in our experience, these are not necessarily the skills/tools that they would most benefit from.
2. **In-person Training:** This will aim to kick-off each participant's data skills development and will employ a combination of discussions, hands-on exercises, labs and final projects aimed at developing enough foundation for participants to pursue more learning. A detailed plan of the training can be found below. The overall objective of this training will be to introduce new concepts to the training participants that they will then continue to develop these skills through their practical application in the final stage of the programme.
3. **Post-Training Support:** The aims of the post training support will be to provide online learning resources, mentorship or advice to participants for a fixed period after the training. Due to the number of expected trainees, we will organise participants into groups of around four to five, based on skill level. Following the training, we will work with the different groups to design a data project or investigation that they will work on together. The idea of this learn by doing approach is that the School of Data trainer will be able to help the trainees see the practical application of their newly acquired skills outside of a classroom environment. Furthermore, one of the most important skills that participants will need to learn is how to solve data problems as they emerge in the actual implementation of a project, outside of an artificial environment.

The offline training will aim at exposing participants to relevant data skills needed to continue their personal learning after the five days. This will be achieved through the following:

- **Overview of the data process:** this will be done through the School of Data Pipeline framework which aims to walk participants through the steps/phases of identifying a question or challenge to developing the needed insight or solution for intended users.
- **Four-stage Learning Model:** this model is adapted from the African Leadership University Learning Model which aims to take learners through structured facilitation, self-learning, peer-learning and facilitated group learning. Below is a breakdown of this approach:
    - **Discovery stage:** this is where a question/problem is introduced. For instance, the class will start out with a question about "How many hospitals in Dar es Salaam provide dialysis treatment?". The facilitator and the student will discuss various methods to answer this question including assessing whether this is the right question to ask, what datasets to use, where to find and get this data etc. Discovery stage will involve in-class sessions which could be half day or full-day workshops depending on the specific concept being presented.
    - **Self-learning stage:** students are given links to relevant resources and guided questions to explore the topic further. The goal is to have students spend time exploring solutions on their own.
    - **Peer-learning stage:** students will work in groups of 3 - 5 to share discoveries in tackling specific questions. This creates an opportunity for students who have been able to tackle the challenge explain their learning, and also create an opportunity for students still working through challenges to have questions answered and concepts explained.
    - **Facilitated group-learning stage:** this stage involves regrouping with the entire class to have a facilitated discussions on solutions and tools to the challenge. The facilitator will have answers to the questions but will present an opportunity for students to explain their approach to answer questions. Emphasis will be on individual/group presentations during this stage.

In-Person Training Schedule

Duration	Monday	Tuesday	Wednesday	Thursday	Friday
1 hour (pre-class tasks)	Readings, overview, tools setup	Readings, overview, tools setup	Readings, overview, tools setup	Readings, overview, tools setup	Readings, overview, tools setup
2 Hours	Concept/Discussion (Discovery)	Concept/Discussion (Discovery)	Concept/Discussion (Discovery)	Concept/Discussion (Discovery)	Concept/Discussion (Discovery)
15 mins	Break	Break	Break	Break	Break
1 Hour	Facilitated Lab	Facilitated Lab	Facilitated Lab	Facilitated Lab	Facilitated Lab

	(Discovery)	(Discovery)	(Discovery)	(Discovery)	(Discovery)
1 hour	Lunch	Lunch	Lunch	Lunch	Lunch
1 hour	In-class Exercise (self-learning)	In-class Exercise (self-learning)	In-class Exercise (self-learning)	In-class Exercise (self-learning)	In-class Exercise (self-learning)
1 hour	In-class Exercise (peer-learning)	In-class Exercise (peer-learning)	In-class Exercise (peer-learning)	In-class Exercise (peer-learning)	In-class Exercise (peer-learning)
15 mins	Break	Break	Break	Break	Break
2 hours	Presentation & Discussion of Exercise (facilitated-group learning)	Presentation & Discussion of Exercise (facilitated-group learning)	Presentation & Discussion of Exercise (facilitated-group learning)	Presentation & Discussion of Exercise (facilitated-group learning)	Questions & Review

Skills

The School of Data Pipeline will be used as the framework for data training. This provide learners the opportunity to observe and engage with an example data process starting from a defined question through to developing useful insights or solutions. This breadth ensures that learners have the threshold knowledge and skills to dive deeper in specific areas of weakness or passion during and after the practicum. The specifics of each session will be informed by the pre-training assessment to be carried out. However, School of Data tentatively proposes the following high-level outline:

Session 1: Getting Familiar With The Data User's Environment & Challenge (Define)

- This will incorporate concepts and resources from user-centered design that are valuable to the data process. Learners should have an appreciation of the data user's challenge and environment.
- Learners will be provided resources to help set up the necessary tools for subsequent sessions
- Tools: Human-Centered Design

Session 2: Introduction to Programming and Getting Familiar With The Data (*Find*, *Get*)

- Learners will learn to setup up at least one programming environment (R or Python) on each learner's laptop as well as the basics syntax of R or Python necessary to traverse the data pipeline on a basic level.

- Learners will learn various approaches to identifying the relevant datasets for a specific need, getting the data into the most appropriate tool and answering questions that will provide more insights into the potential use and limitations of the data. Learners will also work with various tools to scrape data from PDFs and websites for their work.
- Tools: Google Advanced Search, Open Data Portals, Spreadsheets, Tabula, R/Python, GitHub

Session 3 : Putting The Data To Work (Verify, *Clean & Analyse*)

- Learners will explore different technique to get data into a format that is easy to use for reproducible analysis. Basics concepts such as use of summary statistics and visualisations will be explore with appropriate tools for data cleaning. Concepts such as dealing with missing values will also be discussed.
- Learners will also explore various analysis techniques to gain initial insights into their proposed problem.

Session 4: Bringing The Data To Life (*Present*)

- This involves techniques to communicate and share insights with the intended audience through visualisations, stories and platforms. This session will explore the basics of data visualisations including best practices and tools to easily deploy insights.

Session 5: Advanced Data Skills (Advanced Visualisation, Documentation & Maintenance)

- This session will involve learning advanced skills for working with data such as creating interactive data visualisations and maps. Learners will also learn how to deploy their solutions on online platforms such as GitHub to maintain and deploy insights via various platforms. These will include exploration of simple websites, documentation. The choice of advanced skills to be taught in the session will also be based on learnings from pre-assessment conversations with learners and skills learners will find useful after going through the first 3 sessions.
- Tools: GitHub, Shiny, Plotly, Tableau, Carto

Follow Up Support Structure

Participants in the training will be broken up into groups of four to five based on their skill level and area of interest. They will work together throughout the training in order to better learn each others respective areas of expertise, interests and strengths. On the final day of the in person training, facilitated time will be dedicated to support the groups to determine a data investigation and advocacy project that they will collectively work on over the coming 10 weeks. The School of Data trainer will spend time with each group, helping them to refine their project to ensure that it will allow them to effectively explore the concepts and skills that were discussed over the course of the workshop.

The final output of the offline training will be a set of exploratory data projects/investigations and a set of tasks to be completed in the weeks following the training.The trainer will agree a weekly meeting time with each group.

During the subsequent weeks, the School of Data trainer will advise and support the groups to implement their data project through weekly meetings of between two and four hours. Certain weeks will be dedicated to project work where the trainer will help the groups overcome challenges that they encountered and advise them on next steps in the project. Other weeks will be dedicated exposing and practicing new skills that were not covered in the offline training. The trainer will be available for up to four 2-4 hour sessions per week, dedicating the remainder of their time to session preparation.

Sample Data Trainings

School of Data has been providing data training to civil society organisations and journalists since 2012. The example training and training resources listed below represents a small sample of the trainings and content that we have developed and delivered.

- Data Journalism in Sudan : In 2016, SChool of Data partnered with the World Bank to deliver a nine week data journalism training programme to 40 journalists in Sudan. The programme was delivered over 9 months and the training resources can be found on the programme's website.
- DCLI Professional Practicum: The DCLI Professional Practicum is funded by the Data Collaborative for Local Impact (DCLI) program, and specifically administered through Data Zetu. The practicum aims to expose selected fellows to the use of data for decision making through individual program work, work with a host institution that is a leader in data analytics and evidence-based decision making, a joint project and structured training. The training is a four session programme over the course of 3 weeks designed to provide fellows with foundational technical skills to effectively work with data. A full breakdown of the course materials can be found on the programme's github page.
- Code Club with the UK Cabinet Office: School of Data delivered a series of 90 minutes training sessions at the UK Cabinet Office throughout 10 weeks in Late 2015. The Cabinet Office recruited a selected group 20 team members from across their staff to attend for the full 10 weeks training programme. Each session was 90 minutes and included the following elements:
    - Short technical tutorials, for example, a presentation OpenRefine or an introduction to Python
    - Walk through of homework from the previous session
    - Presentation of homework for the following week
    - Show and tell: Participants share what they have been working on
    - Homework, presentations, tutorials and all other materials will be shared on a designated Github repo of the Cabinet Office.
- Advocacy Assembly Video Courses: We developed three video courses for Small Media's advocacy assembly programme.
    - The Data Gathering for beginners course is designed for human rights activists and journalists who would like to use data to support their advocacy

work or tell stories. It includes a gentle introduction to some basic data concepts and will also introduce you to various ways you can source data by using data portals, searching data online or collecting your own data through surveys. You'll also learn about verifying data so you can be sure it's reliable. You will also dive in some common misconceptions and learn how to avoid traps like calculating the wrong average or asking leading questions in your surveys.
    - The cleaning and analysing data course is designed for human rights activists and journalists who would like to use data to support their advocacy work or tell stories. Starting with how to do basic math in a spreadsheet this course covers the most important tools and methods for analysing data and aims to leave you with the practical skills to start exploring your own data.
    - The data scraping course is designed for human rights activists and journalists who would like to use data to support their advocacy work or to tell stories. You will learn the fundamental concepts of scraping and discover how to use free and easy-to-use tools to scrape data from web pages (using Google Sheets and a web browser extension called Web Scraper), social networks like Twitter and Instagram (using a web service called IFTTT) and PDF files (using both a web service called Abby Fine Reader and Tabula, a free application you can download to your computer made by journalists).
- Online Courses and Training Materials: In addition to the training resource, the school of data course page has a selection of learning content that we have developed over the years.