Facilitator pack

The Alan Turing Institute

Data Study Group

9 - 13 December 2019

Introduction

This booklet is a set of brief reading materials in preparation for the Alan Turing Institute Data Study Group facilitator training day.

Where and When

The DSG facilitator training for the December 2019 data study group will run across two days at the **Alan Turing Institute** from **Thursday 5th - Friday 6th December from 10 am**.

What we plan to cover:

- **What a facilitator** is, and isn't - setting expectations appropriately
- **Pluses and deltas** - a conversation with previous facilitators about what has worked well in the past and what they would do differently going forwards
- **Tech logistics** - how team members will access the data and computing facilities
- **Team building** - how to ensure everyone in your group is meeting their potential
- **Project write up** - what you need to keep an eye on through the week so that a representative report is delivered on Friday morning
- **Presentation preparation** - this should be a subset of bullet points or figures from the report and provide a brief overview of your adventures through the week
- **How to access to the data**!!

Facilitator Training Agenda

Day 1 – Thursday 5th December

10:00 am – Meet at The Alan Turing Institute

10:15 am – Introductions

10:30 am- What is a DSG, and what is a facilitator? (Jules Manser, DSG project manager)

10:45 am – What does it mean to facilitate (Ben Murton, Head of Researcher Development and Training)

1:00 pm – Lunch with science leads and DSG organising team.

2:00 pm – Tech overview (Dr James Robinson, Senior Research Software Engineer)

Run analyses in a secure environment

Tools available

How to save outputs and images in HackMD

4:00 pm – Closing circle

Day 2 – Friday 6th December

10:00 am – Meet at The Alan Turing Institute

10:15 am – What does success look like?

10:30 am – Reports and Presentations (Jules Manser)

10:45 am – Communication via slack

11:00 am – Recommended workflows

Lessons Learned

Kan Ban boards

1:00 pm – Lunch with PIs and DSG organising team.

2:00 pm – How to access the data and safe haven. (Support from Ben Walden)

3:00 pm – Folder Structures and accessing the data (1 more time!)

3:30pm – Closing circle

Some House-Keeping

Facilitator Bios

Please email datastudygroup@turing.ac.uk with a photo and brief biography so we can add you to this document - you can see examples from all our previous facilitators here too: https://docs.google.com/document/d/1Uw4ZdQehNCWVt0o2ihXCJEhrP0kLN6VyxwwSoGMjEHw/edit#

We will then add this information to the DSG Delegates Handbook.

Facilitator	Challenge	Principal Researcher
Dan Saattrup Nielsen	WWF	Kasra Hosseini
Laura Merritt	Dstl - Anthrax and nerve agent detector	Andrew Dowsey
Sam Blakeman	Dstl - Bright-field image segmentation	Jeremy Pike
Ondrej Bohdal	SenSat	Wen Xiao
Adriaan Hilbers	Agile Datum	Jack Roberts
Fazl Barez	The National Archives	David Beaven

The Meta-Facilitators (See last page)

You can consider us a facilitator for the facilitators. We are focused on enabling your success (which will cascade down to the success of your team).

The 'meta-facilitators' will be available by email, on slack and phone in advance of the data study group and throughout the week itself.

Catherine - 07508 071 851 - clawrence@turing.ac.uk
Jules - 07852 401773 - jmanser@turing.ac.uk
Marya - 07757796985 - mbazzi@turing.ac.uk
Alvaro - 07511398261-

We hope you will also be resources for each other. During the training day we will encourage you to use slack so that you can engage with each other quickly and easily.

The members of the Data Study Group team are also available to answer questions, solve problems and put out fires. You can liaise with them directly, or through Jules/Marya/Catherine/Alvaro

Slack - Good communication is the key to every successful marriage, and DSG.

- Due to the confidentiality of some of the challenges, it is imperative that **no discussions on the actual challenge data, code or insights** be conducted on the Slack channel.
- Do use the Slack channel to:
    - organise in your groups
    - ask the executive team for additional help including IT support
    - share interesting resources
    - ask someone to come and find you to discuss a particular challenge question in person
        - People coming from different teams is totally fine but do not discuss the challenges in any part of the Slack workspace (including private channels and direct messages).
- Everyone is able to leave the channel at any time they wish.
- There will be channels for each of the challenges:
    - **E.g. dec19-WWF**
- Other Useful Slack Channels
    - **dec19-facilitators** private channel for delicate discussions
    - **presentation-qs** for questions about the presentation
    - **report-qs** for questions about the reports
    - **introductions** to share a little more about yourself and virtually meet other participants
    - **requests** for...well...requests!
    - **help** for help!
    - **social** to help coordinate any breaks from the work
    - **opportunities** to share other events or other opportunities that DSG community members may be interested in
    - **pictures** to share pictures *if all people in the photographs consent to the picture being shared*

What is & what is not a Facilitator

Facilitate *(verb) via [wiktionary]*

1. *To make easy or easier.*
2. *To help bring about.*
3. *To preside over (a meeting, a seminar).*

Collaborative working doesn't just happen. It takes hard work to effectively manage the communications between individual researchers, particularly when they are from different fields of study, at different points in their research training (first year PhD students to faculty to professional data scientists and every stage in between) and when you are only together for 5 days.

Your role as a data study group facilitator is ***to bring out the best in the team members and oversee the interactions within the team.*** We want researchers to have a good time during their visit to the Turing, to feel that their expertise was appreciated and harnessed, and to feel proud of the group's contribution as a whole. This will involve ***ensuring that interpersonal communications are clear***, ***work conducted is documented and included in the report (even if it doesn't give a conclusive result or is incomplete) and that the project is managed in a transparent way.*** Know the code of conduct (https://bit.ly/33jABHx) and who to contact if anyone reports a violation.

The best facilitators make it seem like they did nothing at all. Don't be fooled. ***You are key to the success of the data study group challenge. But also, you are not alone.*** *We have a dedicated members of the team (Marya, Catherine, Jules & Alvaro), who are available all week and for you to talk to. Their pictures are in the appendix.*

DSG Facilitator Activities	Not a DSG Facilitators Responsibility
Help the group move the project forward smoothly & on time. Keep their focus on the question! Work towards the goals agreed at the beginning.	***Being the boss!*** If there are two ideas for analyses, then it is not the facilitator's role to pick between them. ***Responsibility for the scientific quality of the final report.*** It is certainly not on your shoulders to deliver a definitive answer to the challenge all by yourself, this responsibility is shared by the whole team.
Harness all the different skills in a group. Celebrate diversity and be creative in finding synergies in the team.	***Project secretary, scribe or data cleaner.*** Those are all tasks that must be completed (and you're in charge of keeping track of their progress) but they do not (and should not) be completed by you.
Cheerlead for your team! Do your best to make sure everyone has a good time, and encourage shy people, or more nervous participants.	***Grading participants on their participation***. So long as they are not breaking the code of conduct, and that the group is giving them opportunities to participate, let them crack on.

Task Timeline

In Advance

1. **Attend the training day!**
2. Don't forget to write your facilitator bio

During the DSG Week

1. ***Get to know your challenge team.*** Focus on making sure they build connections so that they feel comfortable working together.
2. ***Brainstorm with your team the tasks that need to be completed*** and what success looks like for each of them. Document each new task as it arises as a github issue, and use the project board to clearly communicate what everyone is working on. ***Revisit the tasks list regularly.***
3. ***Check in with your team regularly throughout the week***. Try to bring everyone together twice a day, no more and no less.
4. ***Daily meeting with meta facilitator support team*** so that they can support any difficulties before they get out of hand.
5. ***For the report,*** think about your audience. The executive summary is for the (in the best case scenario) the CEO of the company to read when someone from upper management returns after the data study group week singing all your praises. https://hackmd.io/OQ6tj3lrQzapB0HXTxoPPw
6. ***Presentation = a celebration of all your work!*** It is 20 minutes - make sure that everyone's work is represented, but that doesn't mean that everyone needs to speak for an equal amount of time

Day 1 (The Start)

□ Morning Presentations
□ Assemble Team for introductions
□ In-depth introduction to task by Challenge Owner
□ Access the data
□ Group Brainstorming
□ Small group work

Major Milestone: Access! (Every member of the team can access the raw data)

Day 2 - 4 (The Working Days)

□ Morning small group work
□ 1st 30-minute group meeting just prior to lunch
□ Afternoon small group work
□ **2nd 30-minute group meeting @6pm**

Major Milestone: Progress! (Based on looking only at the HackMD & Presentation)

Day 5 (The Presentation)

□ Present!

Major Milestone: Finalising the report!

Lessons Learned

There once was a challenge with overly ambitious challenge owners…

Sometimes, challenge owners can a little over-excited about their project. And sometimes, they forget that the participants are seeing the dataset, the safe haven, and each other for the first time, so it may take a few hours to get into the swing of things. The most important thing for a facilitator to achieve is to make sure the team has the support and environment to do their best work around the project and have a positive experience. If you find that the challenge owners are being a little too 'active' in voicing their opinions during group meetings, or if you sense they are `guiding' the project too much, politely distract them with something else. Some have even suggested asking them to take notes on the whiteboards. And if you feel you need back-up, go the last page of this document and meet your muscle.

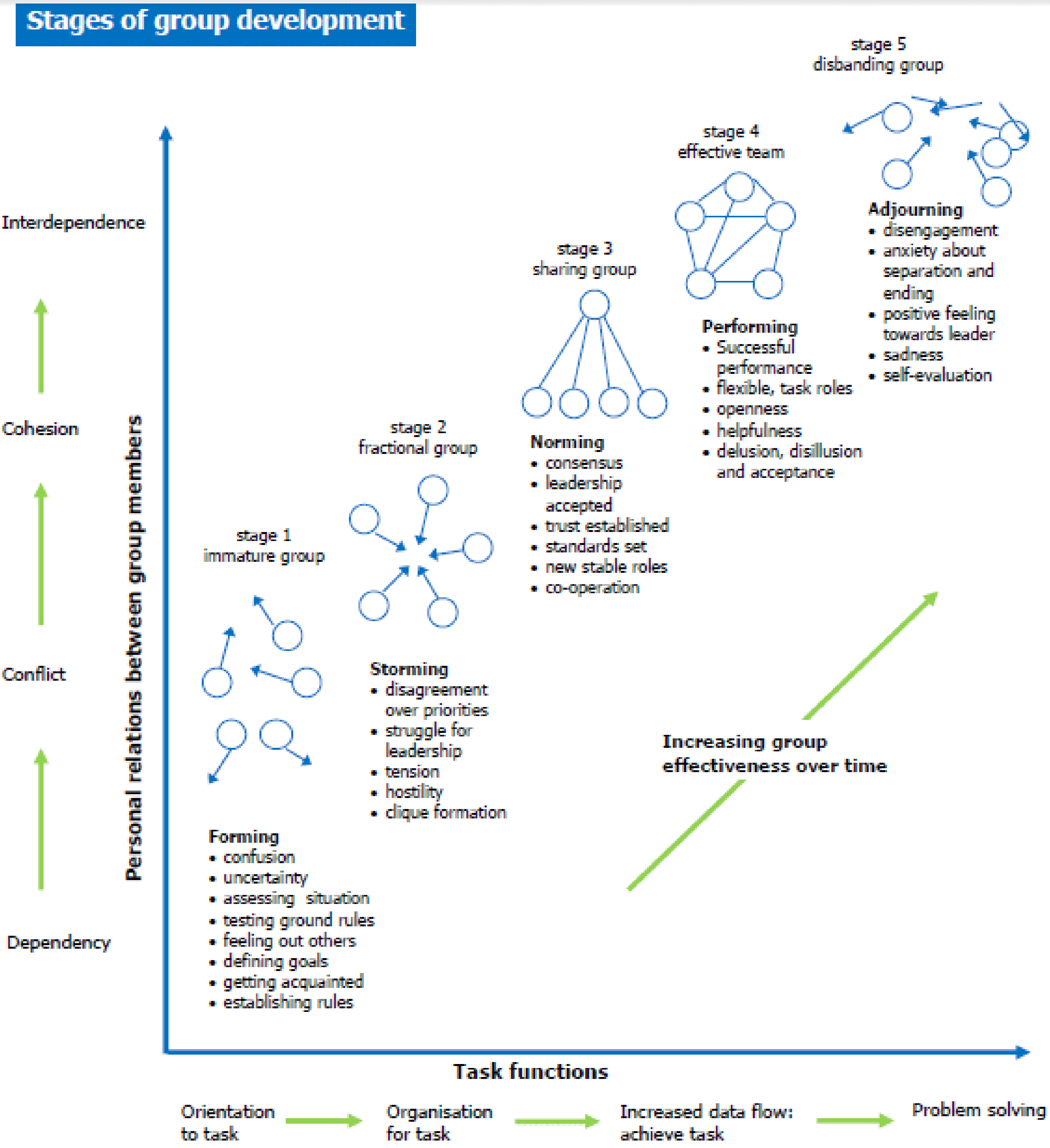
There once was a challenge without data… And it wasn't on purpose.

A challenge owner changed their mind about the data they were sending, downgrading it to a subset, then a curated sample, then an open data set, then nothing at all. (Footnote: This was part of our learning experience as a DSG committee and should not happen again!) We were eventually asked to synthesise our own data by "imagining" how their business might work, and hence what the data generating process might look like…. That is not what the DSG is about, so we flatly refused. We instead suggested some alternative tasks during the week, some of which they accepted. The key to success = mentioning that the organisation had spent a lot of money for access to the data science resource that we were offering, and we must take advantage of it: both for their sake and for the sake of the participants. Each day brought fresh challenges managing a changing cast of company representatives, which required being tactful but firm, and by staying true to our suggestions and producing a report of ideas to trial when the data was available, we actually found that the challenge owners were delighted with the outcome.

There once was a challenge where coding together didn't make everyone best friends…

Whilst it is amazing to hear about the life-long friendships and marriages that have come out of the DSG (currently both zero), the purpose of the week is to make sure that everyone is comfortable, engaged and productive. The two things that facilitators tend to worry about are lone wolves and boredom. Your job, and ours as the organisers, is to try to set a supportive and inclusive tone from the beginning. However, if people want to work on their own particular idea, as long as they aren't disrupting the rest of the group, and continue to engage in the town-halls, let them do it. Similarly, if you're losing team members to boredom, try being creative in finding paths forwards that are both engaging and relevant to the project. It's fine to let individuals work on some exploratory or piloting work, even if you don't think it will work. The goal of the DSG is to have fun! If in doubt just let challenge members work on what they're excited by. There's a nice infographic of team dynamics on the next page for all the social science geeks out there.

Bruce Tuckman's Forming, Storming, Norming, and Performing Model



***TLDR:** Forming a team takes time, and members often go through recognizable stages as they change from being a collection of strangers to a united group with common goals.*

There is only one rule in DSG Club

"Save everything you create in the shared folder, including the report (should already be in there), code (should primarily be in the gitlab, but feel free to duplicate), any outputs you need exported for the presentation, the presentation itself, and any manipulated versions of the dataset."

OR

"SAVE EVERYTHING IN THE SHARE FOLDER. PLEASE."

N.B. Don't worry, we'll show you where this magic folder is on the training day, we just need you to internalise this rule.

Final Report Considerations

After previous DSGs, we haven't always published the reports. We have a model now (since Dec 2017), whereby we require Challenge Owners to agree to publication of a report, before they even sign up. We won't accept any challenges were the partner doesn't want to publish the results. This means that DSG participants can evidence their participation through named authorship of the report, which will be published on the Turing website. However, it also means that if any analysis isn't written up properly in the report, it will not make any sense, and therefore will likely end up being cut from the final version 😫😫.

Please note: In previous DSGs facilitators have been asked to nominate a handful of participants to finalize reports. However as of December 2019 challenge PIs will be finalizing reports.

There will be further guidance on how to work on reports during the DSG week on the facilitator training day, but to get you in the right frame of mind, have a look at the examples below, and the latest ones on the Turing website.

Writing a Scientific report blog

https://www.turing.ac.uk/blog/what-data-scientific-report

Examples of previous reports

https://www.turing.ac.uk/research/publications/data-study-group-final-report-inmarsat
https://www.turing.ac.uk/research/publications/data-study-group-final-report-codecheck
https://www.turing.ac.uk/research/publications/data-study-group-final-report-dstl

Template report

We use HackMD for collaborative report writing. This template will be available:
https://hackmd.io/OQ6tj3lrQzapB0HXTxoPPw (a version with some brief example text)
https://hackmd.io/0YAuFQBqSYCbn8i4onaaxw# (a version with a bit more guidance notes)

If you haven't used MarkDown before - you can find a really great markdown cheat sheet at: https://github.com/adam-p/markdown-here/wiki/Markdown-Cheatsheet

Your Support Network

Jules Manser - Data Study Group Operations Manager
Jules is a member of the Partnerships team at the Turing and has played a lead role in delivering Data Study Groups working closely with Sebastian since their inception in 2016.

Alvaro Cabrejas Egea – Science Liaison
Alvaro is an Enrichment student at The Alan Turing Institute. His research areas are reinforcement learning and traffic prediction and control. He is interested in economic and agent games in networks, machine learning and time series analysis.

Marya Bazzi – Principle Researcher

Marya is a Turing Research Fellow. Before joining the Turing, Marya was a postdoctoral scholar at Oxford University and Head of Analytics at a London-based FinTech. Marya is particularly interested in developing research projects at the interface between academia and industry.

Daisy Parry- Data Study Group Administrator
Daisy has recently joined The Alan Turing Institute and works closely with Jules to deliver the Data Study Group events.

Annex 1. Report Writing Checklist

Layout

- Should be using the Turing TeX template and Turing corporate design
- Where colours are used, it should be made sure that greyscale prints are still readable

Tables and Figures

- Tables, Figures should be in manuscript (not in appendix), have full descriptive captions, legends if necessary.
- All tables, figures should be referenced in-text.
- Table captions should state: what the table is about; what are rows; what are columns
- Figure captions should state: what the figure is about; all figure elements should be explained in the caption
- For tables/figures in panels, e.g., grouped visualization, scatter plot matrix, etc. Caption should state meaning of panel ordering, explain rows, columns, etc.

Must-have sections, section structure

- There should be an executive summary, right at the start. It should be technically correct, but readable by a non-data-scientist.
- There should be a section which explains in technical detail how the data scientific approaches were obtained from the domain questions, and what these are.
- A section which explains the data tables that were available: for each table, sample size, rows (samples), columns (variables).
- Optimally, there is a section on exploratory data analysis (EDA).
- A longer section on limitations. It is a good idea to split this into "data limitations"
- A section on future work – what would the group suggest to do if follow-up happens?
- Further sections (usually in the middle) should describe the different approaches. It is a good idea to not split or order these by individuals, but by scientific approach.

Acknowledgments and crediting

- All contributors should be acknowledged with a short description of their contributions.
- The challenge PI should also be credited with preparing and guiding the challenge, even if not present during the week (assuming the challenge PI indeed contributed in this way).