# Checklist data delivery for IDH

It takes about 3 weeks to do data quality monitoring and data delivery for a case.

During data collection (week 1):

* Prepare a Lumen dashboard, check for outliers or odd numbers on a daily basis. Inform the data collection supervisor on the specific data points (including info on the submitter)
* Inform Joy/Galih that there will be a dataset ready for uploading to the portal at the end of next week (week that delivery is done).
* End of the week: ask data collection supervisor to write down all (qualitative) field observations. Any information that is relevant for the IDH team to know about the context. Think about the company they are working with, harvest, weather, access to finance etc. Any good things or complains that are shared by farmers can be written down.

Data cleaning (week 2):

Preparation:

* Download the dataset from Flow (Data 🡪 Exports 🡪 New export 🡪 Data analysis export)
* Download the survey form from Flow
* Make a new case folder in the 5200 Farmfit map on Github
* Upload the raw data with structure *flowexportnumber*\_*sdmcompany\_crop*\_raw\_data.xlsx
* Copy the data delivery R script in the case folder and name it data\_cleaning\_*sdmcompany*.R

R script:

* Adapt the data delivery script to your case
* Run the full code without the outlier function and produce the file productivity\_check.xlsx (see end of script). Please read the full template carefully. Every step is explained.
* Check the productivity\_check.xlsx file. Use the filter options and arrange columns from high to low, low to high etc. to detect outliers manually. Also use the intake form and check what the ranges should roughly be (note: intake form data provide an indication, contextual factor can make that numbers deviate). Highlight the cells you have questions about or that are obvious outliers.
  + Consult with the data collection supervisor on the cells you highlighted. Find out why numbers are odd. Make changes to the data where possible (e.g. if a mistake is obvious) or delete observations if they make no sense at all. Note that you always need to be able to clarify why an observation is in the raw data and not in the clean data, so be careful with deleting.
  + While waiting for the input, you can start with the data delivery report. Find the template [here](https://docs.google.com/document/d/1EBfApeeL36HL2aQp0_XbgGt-_dPjJidWbVAZwx8B6ME/edit). Make a copy for own use! The template describes what information is needed. Open folders of earlier cases to see what type of information was shared before (e.g. Landmark Millers)
* Once you have made contextual adjustments, you can run the full code again and produce the cleaned, anonymized data file.
* Produce the file with personal information (see end script).
* Upload the calculation sheet and adapt it to your case (if it deviates from the normal one). This can be found in the folder Analytics-Guild > IDH > Templates and input files for data delivery

**Data delivery (end of week 2)**

* Share what you have prepared so far with Jildemarie for review.
* Share data on MS teams (contact Jildemarie for this):
  + 1. Cleaned anonymized data file
  + 2. Personal information sheet
  + 3. Calculation overview (contextualized)
* Prepare a data delivery report. Find the template [here](https://docs.google.com/document/d/1EBfApeeL36HL2aQp0_XbgGt-_dPjJidWbVAZwx8B6ME/edit). Make a copy for own use! Incorporate the field notes of the data collection supervisor.
* Share data with Joy and team for upload to the portal
* Share an email to the SDM team:
  + Proposal for a call in Week 3 to discuss the data delivery. Invite the data collection supervisor
  + Data delivery report

**Week 3:**

* Data delivery call
* Be available for questions from the SDM team