Data sources and cleaning needed

Most of the data comes from the three reports below. The first report can be run as two reports (one for students and one for supervisors) if needed. The last one is the complex one that needs a lot of cleaning. I am waiting to see if we will have anything online to report on for students (might have be an overall number without unique attendance or college for now).

**Attendance - Full list including event, session and user with College**

* Any sessions starting with ‘HDR Supervisor’ or ‘HDR Supervision’ counted for HDR Supervisor Development Program. All others for REST.
* Session date is the unique count of sessions (no session ID in system)
* Training type for all is ‘seminar’ (synchronous)

**Any signed form filter by user and or form**

* Reports on online induction completion, counted towards REST
* Training type for all is ‘online induction’

**Self-Record Development Activity Report**

* All other HDR Supervisor Development activities, including online and external training we count for program points.
* Training Category is what is available to us in the system – does not translate to training type. Needs to be mapped to training type based on title. This is where a lot of manual manipulation comes in .
* Mapping of training types based on Title:
  + ‘HDR Supervisor Core…’ = ‘online core’
  + ‘Epigeum Module:…’ or ‘Module:…’ = ‘online module’ (5 points)
  + ‘Seminar:…’ = ‘online seminar’ (10 points)
  + ‘Workshop:…’ = ‘online workshop’ (15 points)
  + ‘Extra Core attendance’ = ‘seminar’ (10 points)
  + ‘Delivery of…’ = ‘workshop’ (15 points)
  + Everything else = ‘other’
* Some entries should be mapped to seminar or workshop but were not available in the system at the time of entry. So technically should be counted, but are only a handful so could stay under other.
* Some people don’t have colleges mapped to them. This will need to be addressed as needed – might be fixed by the time of next report.