



Data Managers and R programming, The Unlikely Pair

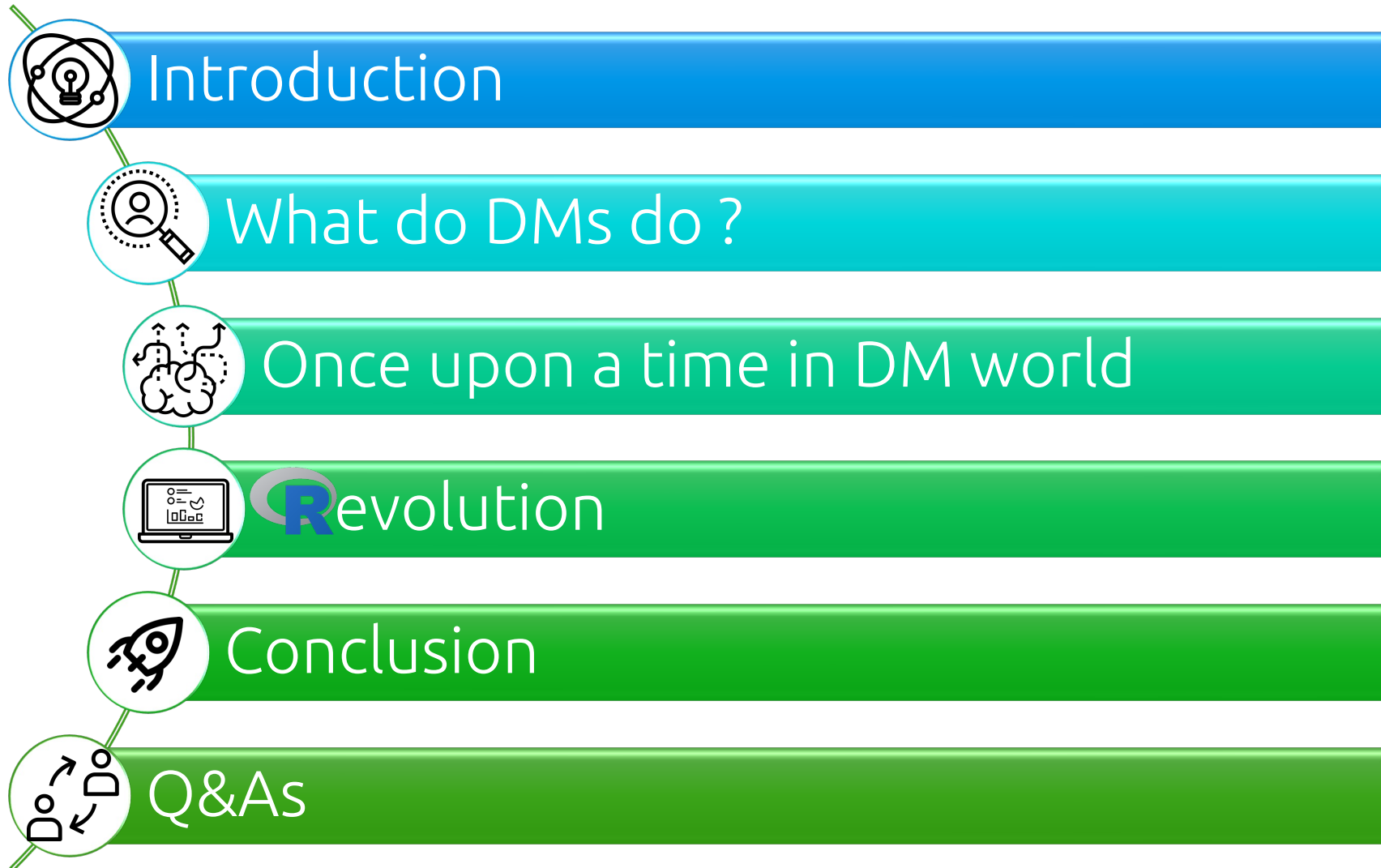
Anaëlle (Bonetta) Perretti

Senior Project Data Manager

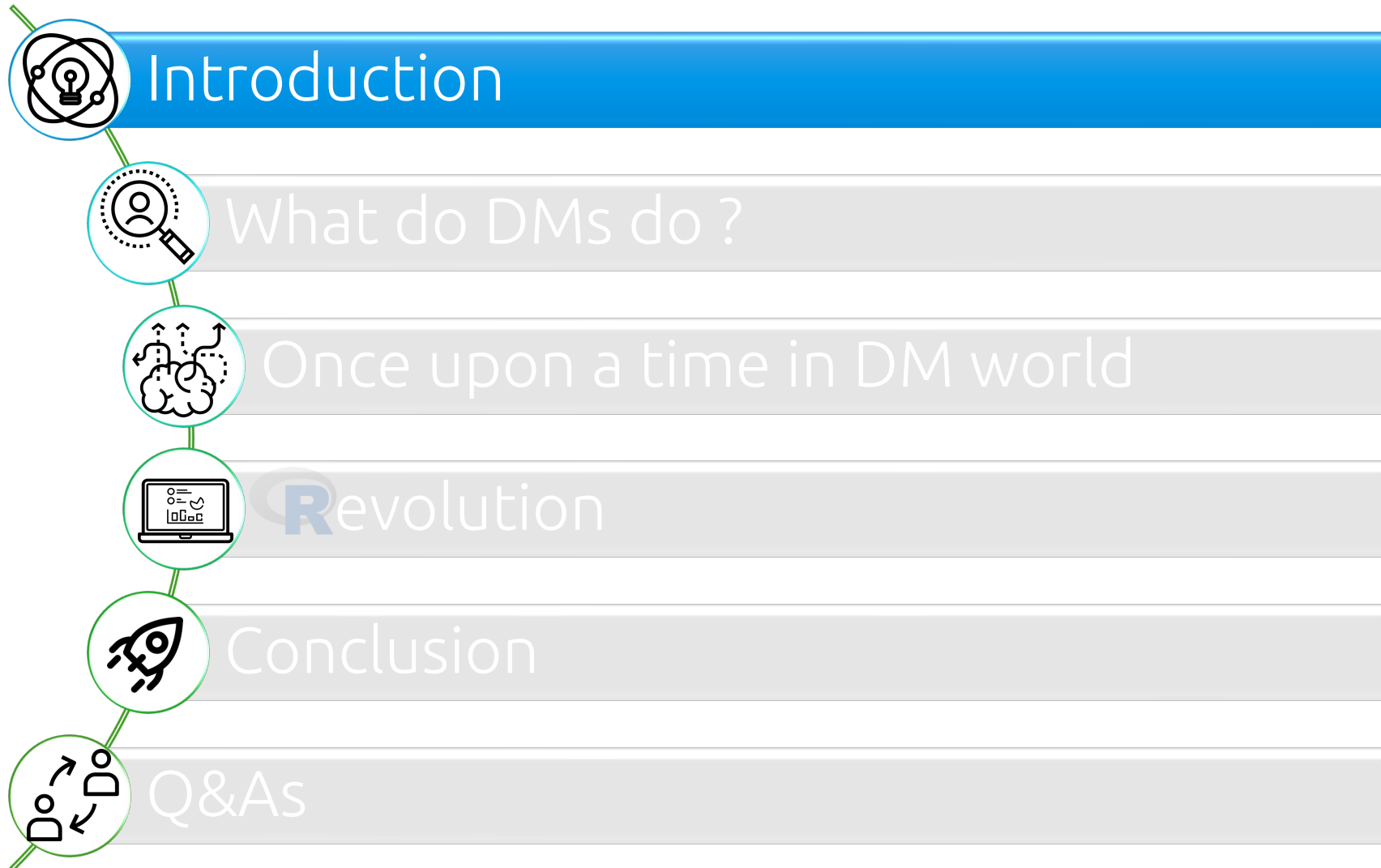
Phuse Connect EU 2023 – Paper DH08



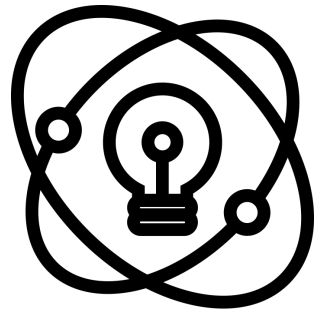
Agenda



Agenda

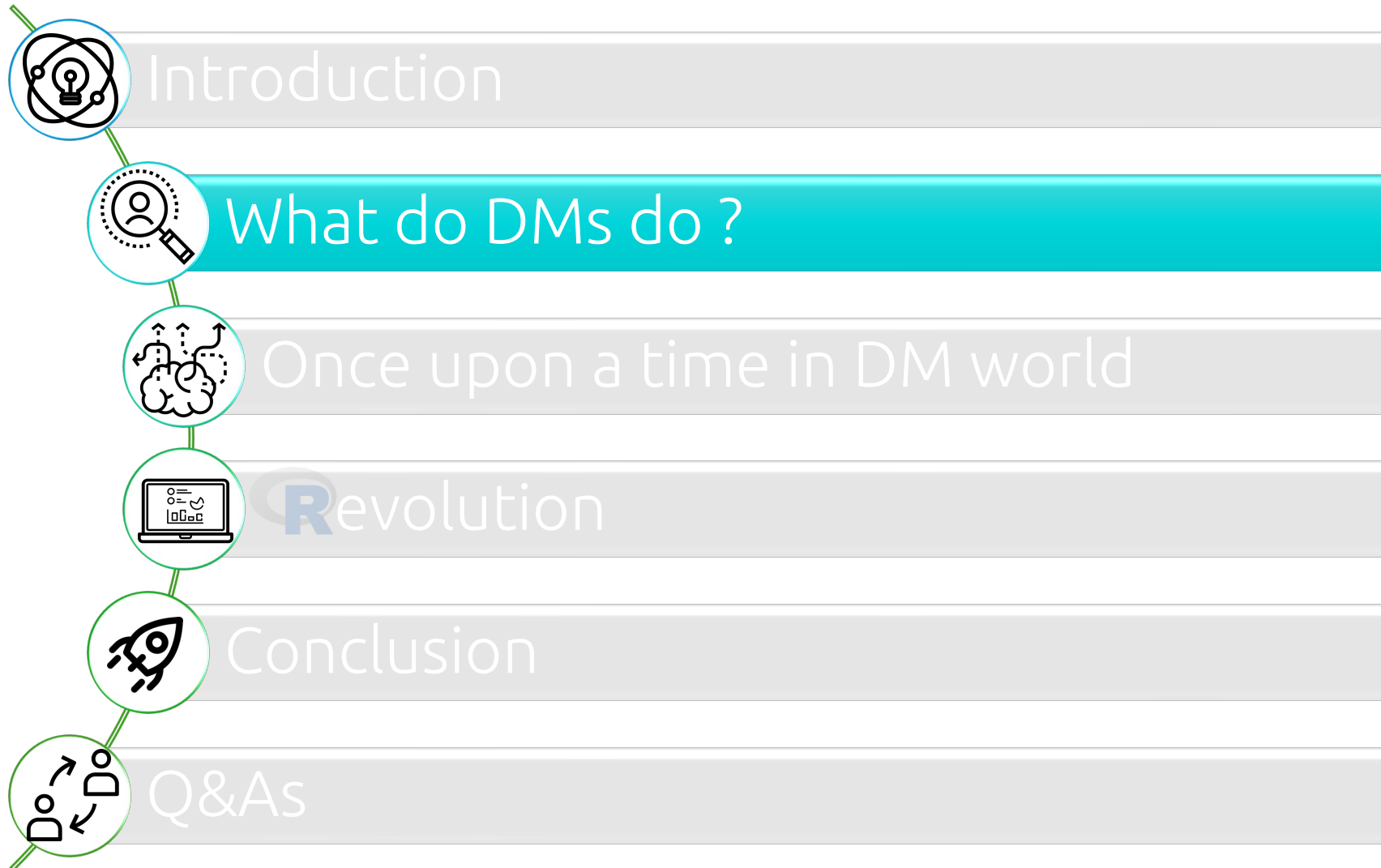


Introduction

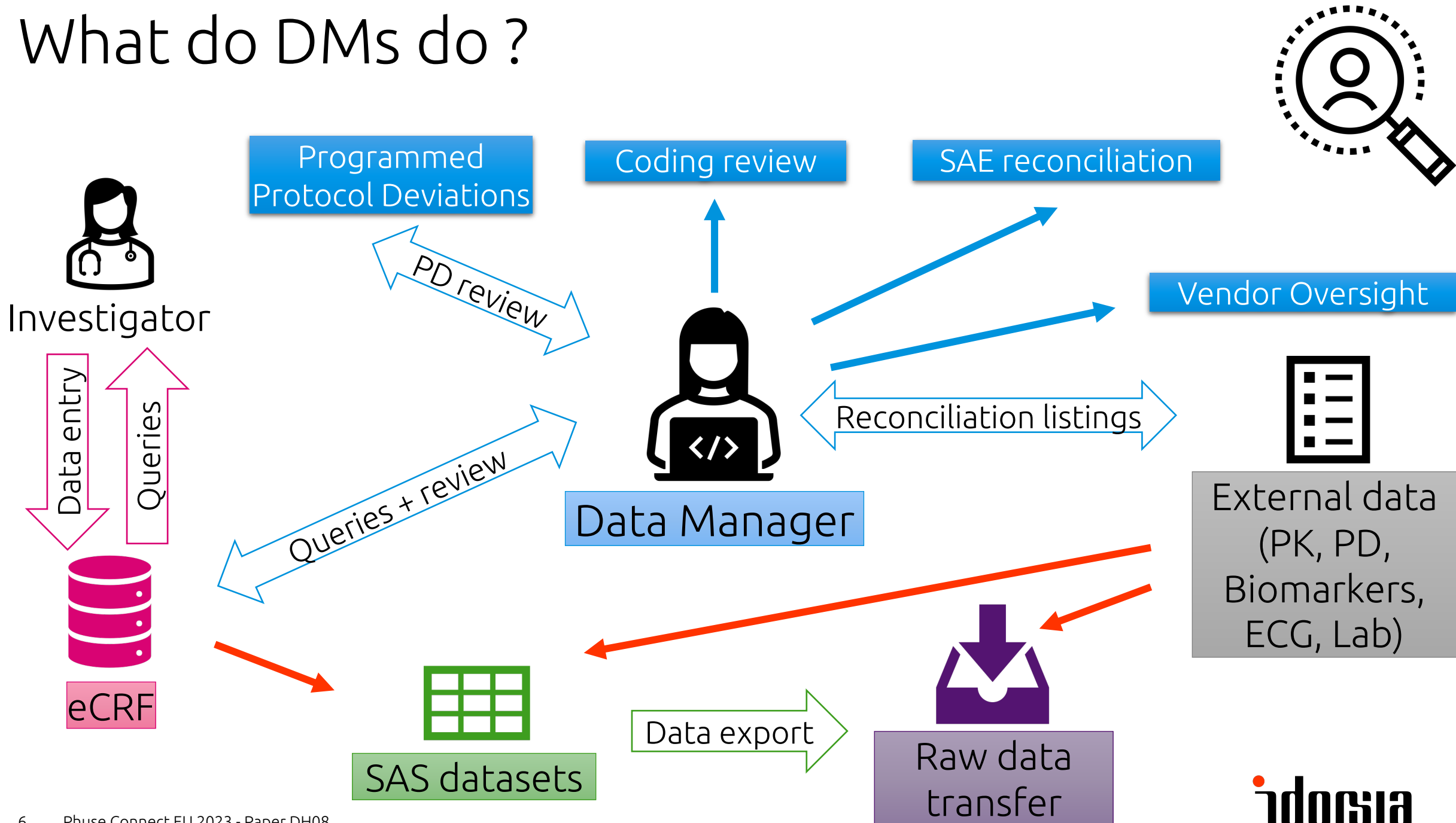


- Main Data Manager task : Data Cleaning
- Important portion implies manual tasks
- Tool needed to optimize time and to reproduce verifications
- Internal training provided by team of Data Scientist
- R programming applied to Data Management activities

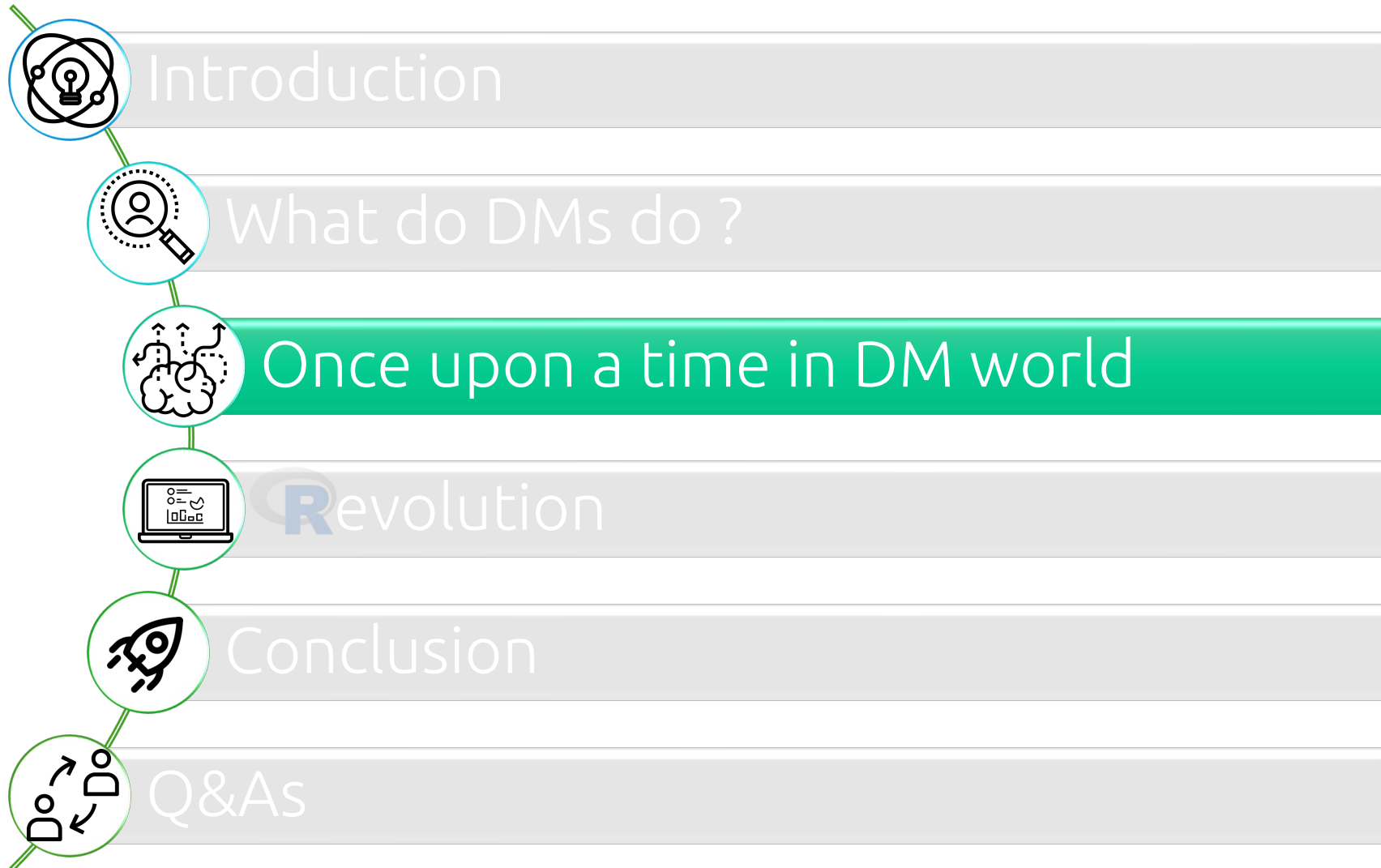
Agenda



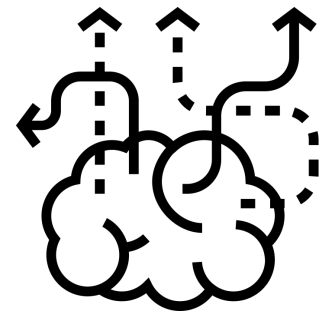
What do DMs do ?



Agenda



Once upon a time in DM world



- Preparation of weekly Data Management status in 2 steps
- Manual filtering in Cleaning Patient Tracker excel listing

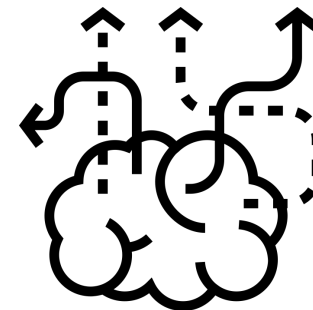
A	C	D	K	O	Z	AH	AI
Country/Total	Number of screened patients	Number of randomized patients	Number of missing visits/records over 21 days	Number of missing eCRF pages for more than 21 days	Number of outstanding queries	Number of open queries for more than 14 days	Percentage of SDV completed
Total	2000	1900	2	25	25	25	94.8



Data Management Metrics and Cleaning Status (cut-off: 03-October-2023)

Summary of Data entry	Comment
Missing visits for more than 21 days	2 visits
Missing pages for more than 21 days	25 pages
Visits not SDVed	5.2 %
Queries	25 open for more than 14 days. 25 outstanding queries
Closed queries	191137 queries in total

Once upon a time in DM world



- Preparation of weekly Data Management status
- Manual filtering in Query overview excel listing

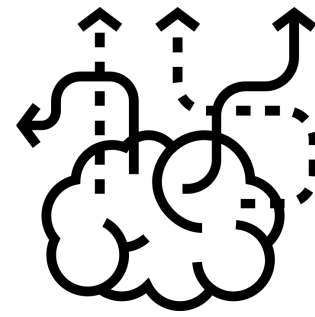
A	F	G
Country	Closed by DM	Resolved by site
Total	6485	184652



Data Management Metrics and Cleaning Status (cut-off: 03-October-2023)

Summary of Data entry	Comment
Missing visits for more than 21 days	2 visits
Missing pages for more than 21 days	25 pages
Visits not SDVed	5.2 %
Queries	25 open for more than 14 days. 25 outstanding queries
Closed queries	191137 queries in total

Once upon a time in DM world



- Spot-check for data cleaning, multiple Informed Consents (ICFs)
- SAS dataset extracted in excel
- Manual filtering in the ICF dataset

Excel listing

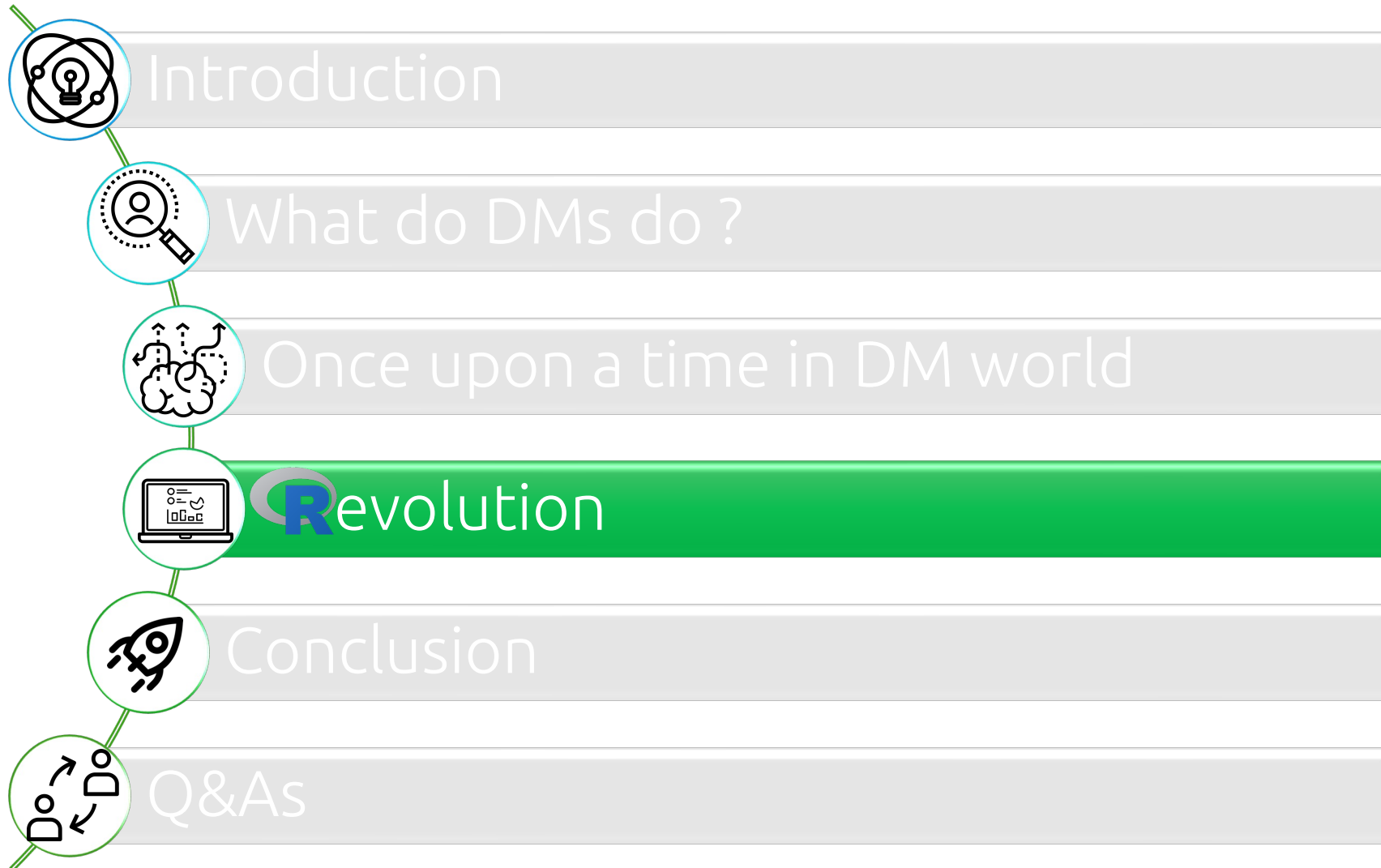
subjid	protvers
X1	4
X2	4
X3	4
X4	4
X1	4
X2	4
X3	4
X4	4



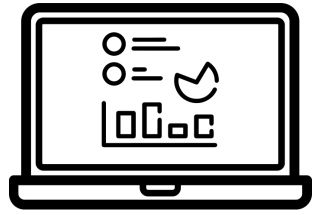
Pivot table

Row Labels	Count of subjid
X1	2
X2	2
X3	2
X4	2

Agenda

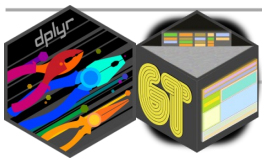






- Tool requested from Data Management end
 - Being able to access live data
 - Manipulating data for all previous tasks mentioned earlier
 - Autonomy to crosscheck several datasets
 - Avoid repeating manual tasks
-
- Internal R training delivered by Idorsia Data Scientist
 - Several weeks of intensive training for non-programmers

- Weekly Data Management status
- Automated generation of the table
- Mainly dplyr and gt packages



Data Management Metrics and Cleaning Status

Date	Nbr screened	Nbr randomized	Nbr Clinical Events	Nbr Injections	Missing data		Queries		
					Missing visits > 21 days	Missing pages > 21 days	% SDV	Nbr Open queries > 14 days	Nbr Outstanding queries
2023-10-04	51	93	4	25	2	25	94.8	25	25

Refreshed on 2023-10-04

Data Management Metrics and Cleaning Status (cut-off: 03-October-2023)

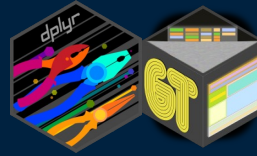
Summary of Data entry	Comment
Missing visits for more than 21 days	2 visits
Missing pages for more than 21 days	25 pages
Visits not SDVed	5.2 %
Queries	25 open for more than 14 days. 25 outstanding queries
Closed queries	191137 queries in total



```
# Preparing table with total numbers in metrics
Total_numbers_p <-
  CPT_p2 |>
  select(
    number_of_screened_patients,
    number_of_randomized_patients,
    number_of_clinical_event,
    number_of_self_injection_performed,
    number_of_missing_visits_records_over_21_days,
    number_of_missing_e_crf_pages_for_more_than_21_days,
    percentage_of_sdv_completed,
    number_of_open_queries_for_more_than_14_days,
    number_of_outstanding_queries
  ) |> tail(1) |>
  rename(
    "Nbr_screened"=number_of_screened_patients,
    "Nbr_randomized"=number_of_randomized_patients,
    "Nbr_Clinical_Events"=number_of_clinical_event,
    "Nbr_Injections"=number_of_self_injection_performed,
    "Missing_visits > 21 days"=number_of_missing_visits_records_over_21_days,
    "Missing_pages > 21 days"=number_of_missing_e_crf_pages_for_more_than_21_days,
    "% SDV"=percentage_of_sdv_completed,
    "Nbr_Open_queries > 14 days"=number_of_open_queries_for_more_than_14_days,
    "Nbr_Outstanding_queries"=number_of_outstanding_queries
  ) |>
  mutate(Date = today()) |>
  select(Date,everything())
#Creating GT table
Total_numbers_p |> gt() |>
  tab_header(title = md("Data Management Metrics and Cleaning Status")) |>
  tab_source_note(source_note = paste("Refreshed on", Sys.Date())) |>
  tab_spanner(label = "Missing data", columns = starts_with("Missing")) |>
  tab_spanner(label = "Queries", columns = 9:10)
```

Preparing table with total numbers in metrics

```
Total_numbers_p <-
  CPT_p2 |>
  select(
    number_of_screened_patients,
    number_of_randomized_patients,
    number_of_clinical_event,
    number_of_self_injection_performed,
    number_of_missing_visits_records_over_21_days,
    number_of_missing_e_crf_pages_for_more_than_21_days,
    percentage_of_sdv_completed,
    number_of_open_queries_for_more_than_14_days,
    number_of_outstanding_queries
  ) |> tail(1) |>
  rename(
    "Nbr screened"=number_of_screened_patients,
    "Nbr randomized"=number_of_randomized_patients,
    "Nbr Clinical Events"=number_of_clinical_event,
    "Nbr Injections"=number_of_self_injection_performed,
    "Missing visits > 21 days"=number_of_missing_visits_records_over_21_days,
    "Missing pages > 21 days"=number_of_missing_e_crf_pages_for_more_than_21_days,
    "% SDV"=percentage_of_sdv_completed,
    "Nbr Open queries > 14 days"=number_of_open_queries_for_more_than_14_days,
    "Nbr Outstanding queries"=number_of_outstanding_queries
  ) |>
  mutate(Date = today()) |>
  select(Date,everything())
#Creating GT table
Total_numbers_p |> gt() |>
  tab_header(title = md("Data Management Metrics and Cleaning Status")) |>
  tab_source_note(source_note = paste("Refreshed on", Sys.Date())) |>
  tab_spanner(label = "Missing data", columns = starts_with("Missing")) |>
  tab_spanner(label = "Queries", columns = 9:10)
```



Data Management Metrics and Cleaning Status (cut-off: 03-October-2023)

Summary of Data entry	Comment
Missing visits for more than 21 days	2 visits
Missing pages for more than 21 days	25 pages
Visits not SDVed	5.2 %
Queries	25 open for more than 14 days. 25 outstanding queries
Closed queries	191137 queries in total

- Script refreshed on demand
- Available on shared folders

```
#Selecting maximum number of closed by DM queries
Closed_sum_p2 <-
  Closed_table_p2 |>
  select(closed_by_dm) |> max()
#Selecting maximum number of resolved by site queries
Resolved_sum_p2 <-
  Closed_table_p2 |>
  select(resolved_by_site) |> max()

#Calculating total number of closed queries
Total_p2 = Closed_sum_p2 + Resolved_sum_p2

#Creating total queries table with all variables
Total_queries_p2 <-
  tibble::tribble(~"Total closed by DM",~"Total resolved by site",~"Total closed queries",
                  Closed_sum_p2, Resolved_sum_p2, Total_p2)
#Creating GT table
Total_queries_p2 |> gt() |>
  tab_header(title = md("Cumulative number of closed queries")) |>
  tab_source_note(source_note = paste("Refreshed on", Sys.Date())) |>
  tab_spanner(label = "Query category", columns = 1:2)
```

Data Management Metrics and Cleaning Status (cut-off: 03-October-2023)

Summary of Data entry	Comment
Missing visits for more than 21 days	2 visits
Missing pages for more than 21 days	25 pages
Visits not SDVed	5.2 %
Queries	25 open for more than 14 days. 25 outstanding queries
Closed queries	191137 queries in total

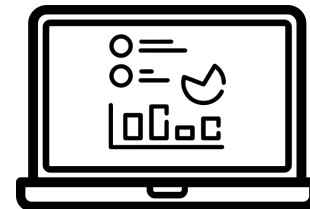
- Finalization of the weekly metrics
- Script refreshed upon demand
- Available on shared folders

Cumulative number of closed queries

Query category

Total closed by DM	Total resolved by site	Total closed queries
6485	184652	191137

Refreshed on 2023-10-04



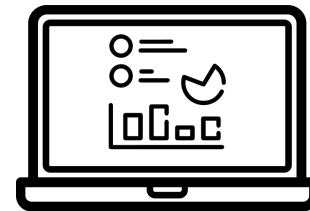
- Answering ad hoc questions
- Searching for patients with multiple ICFs

```
#Grouping data by patient number, counting n consents  
Count_ICF_p2 <- ICF_p2 |>  
  group_by(subjid) |>  
  summarise("Count of ICF entered" = n()) |>  
  filter("Count of ICF entered">1) |>  
  rename("Subject number"=subjid)
```



	Subject number	Count of ICF entered
1	X1	2
2	X2	2
3	X3	2
4	X4	2

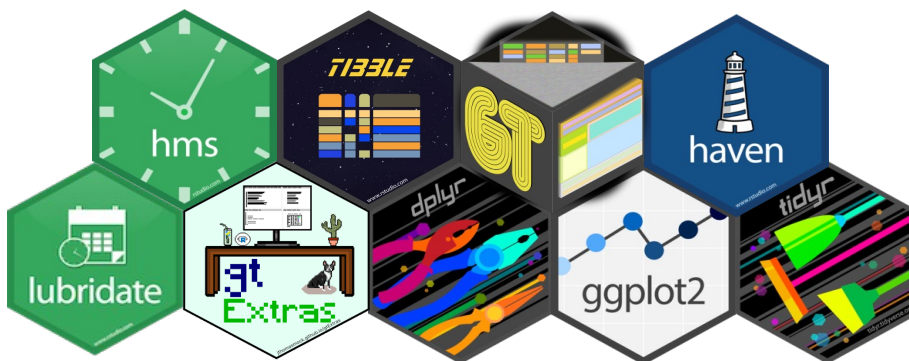
Revolution



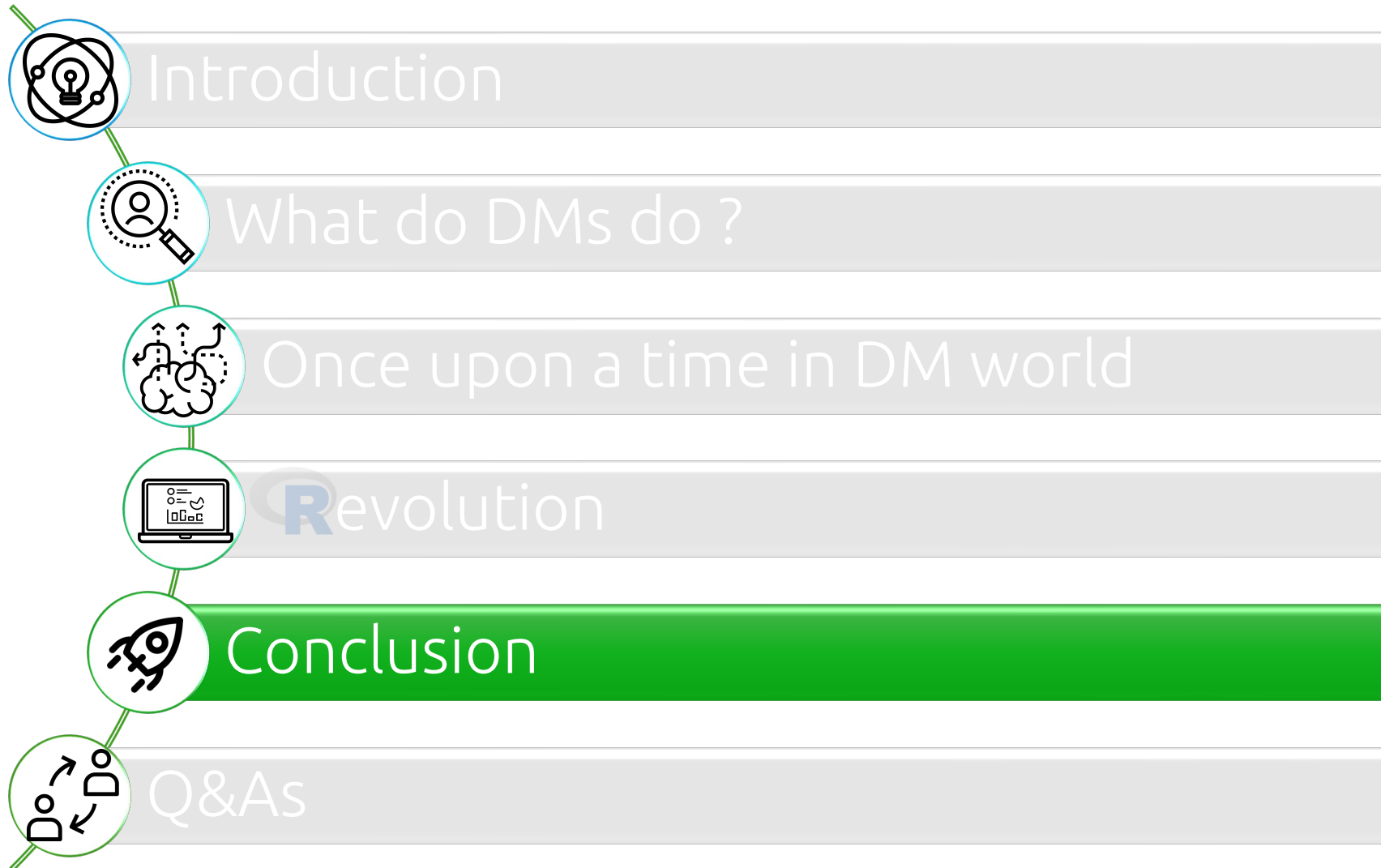
- Reprogramming Protocol Deviations
 - Providing metrics on demand
 - Creating summary tables for newsletters sent to sites
 - Risk Based Monitoring
 - Monthly oversight of vendor activities
-
- Creation of adjudication dashboard with  quarto®



Adjudication status.html



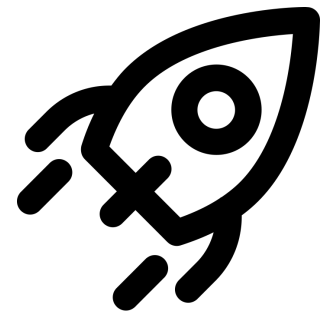
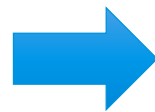
Agenda



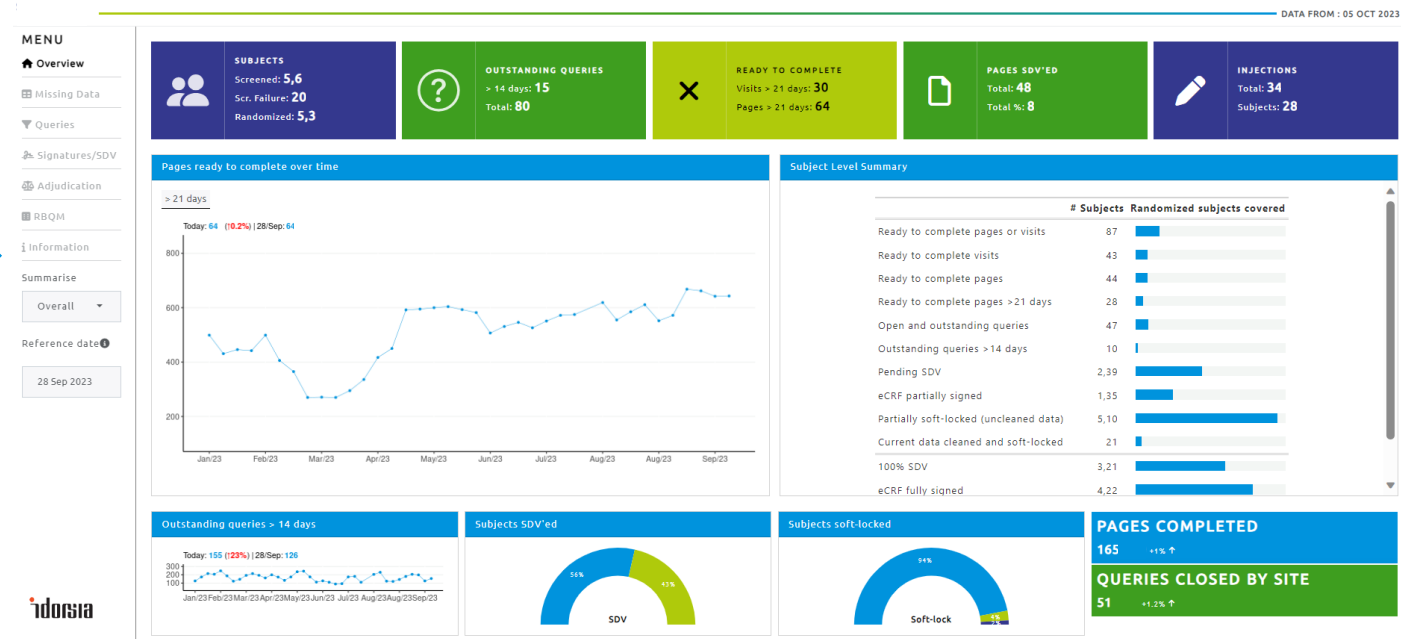
Conclusion

- Introducing a new mindset and way of thinking
- Powerful tool for Data Managers
- Providing daily cleaning solutions
- Optimizing time
- Highly customizable

➤ Ongoing work: *Shiny* app



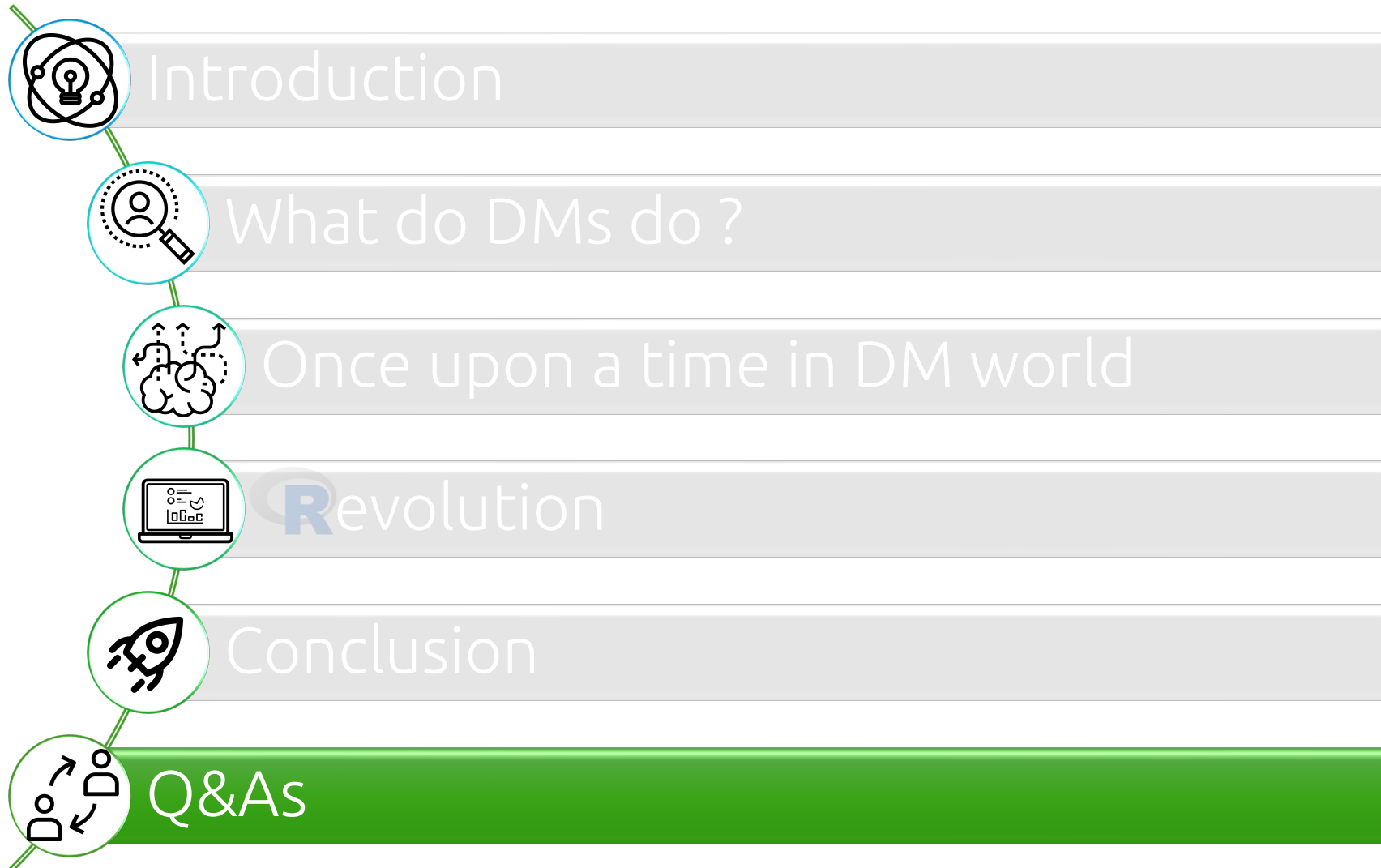
Data Management dashboard



idorsia

idorsia

Agenda





Thank you

Any questions ?

anaelle.perretti@idorsia.com

For more details, please have a look at my paper DH08

