TALL SOM FORTELLER

Creating a paradata pipeline

AND HOW THE PIPELINE FACILITATED ABILITY TO MAKE DATA-DRIVEN ADJUSTMENTS TO IMPROVE DATA COLLECTIONS.

Statistisk sentralbyrå
Statistics Norway

Agenda

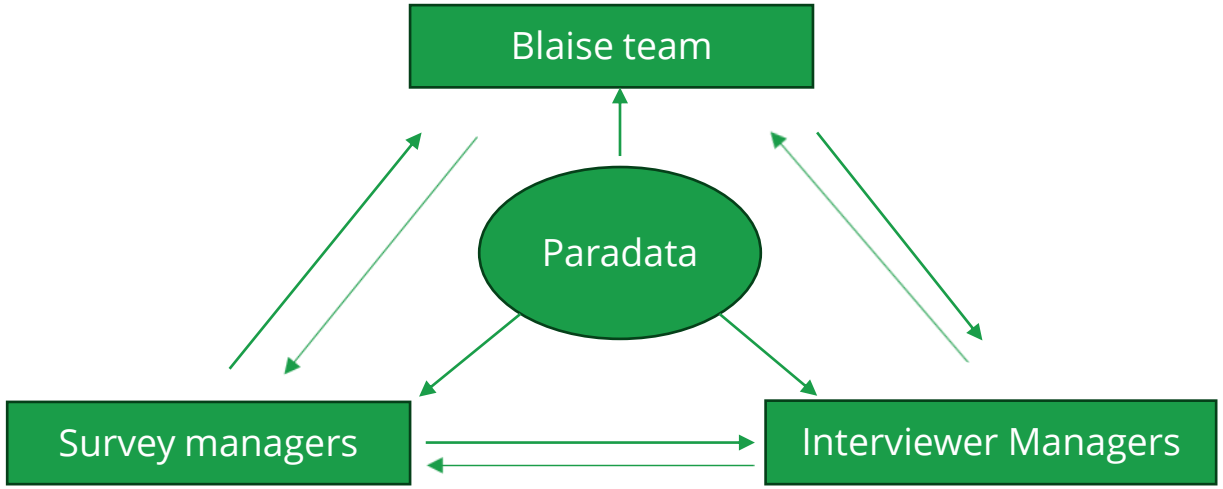
- Paradata
- Why create a data pipeline for paradata?
- Pipeline
- Paradata parser
- Example of how colleagues use paradata
- Next steps

Paradata

- **Dial History and Audit Trail Data**
  - Improving survey questionnaires by understanding respondent behavior in surveys.
  - Dynamic management of surveys by monitoring and adjusting for non-response bias.

Data collection department

- Blaise team (technical team)
- Survey Project Managers
  - Manage communication plan
  - Planning interview resources
  - Contribute to developing survey questionnaires
- CATI managers
  - Provide interview resources
  - Manage interviewers

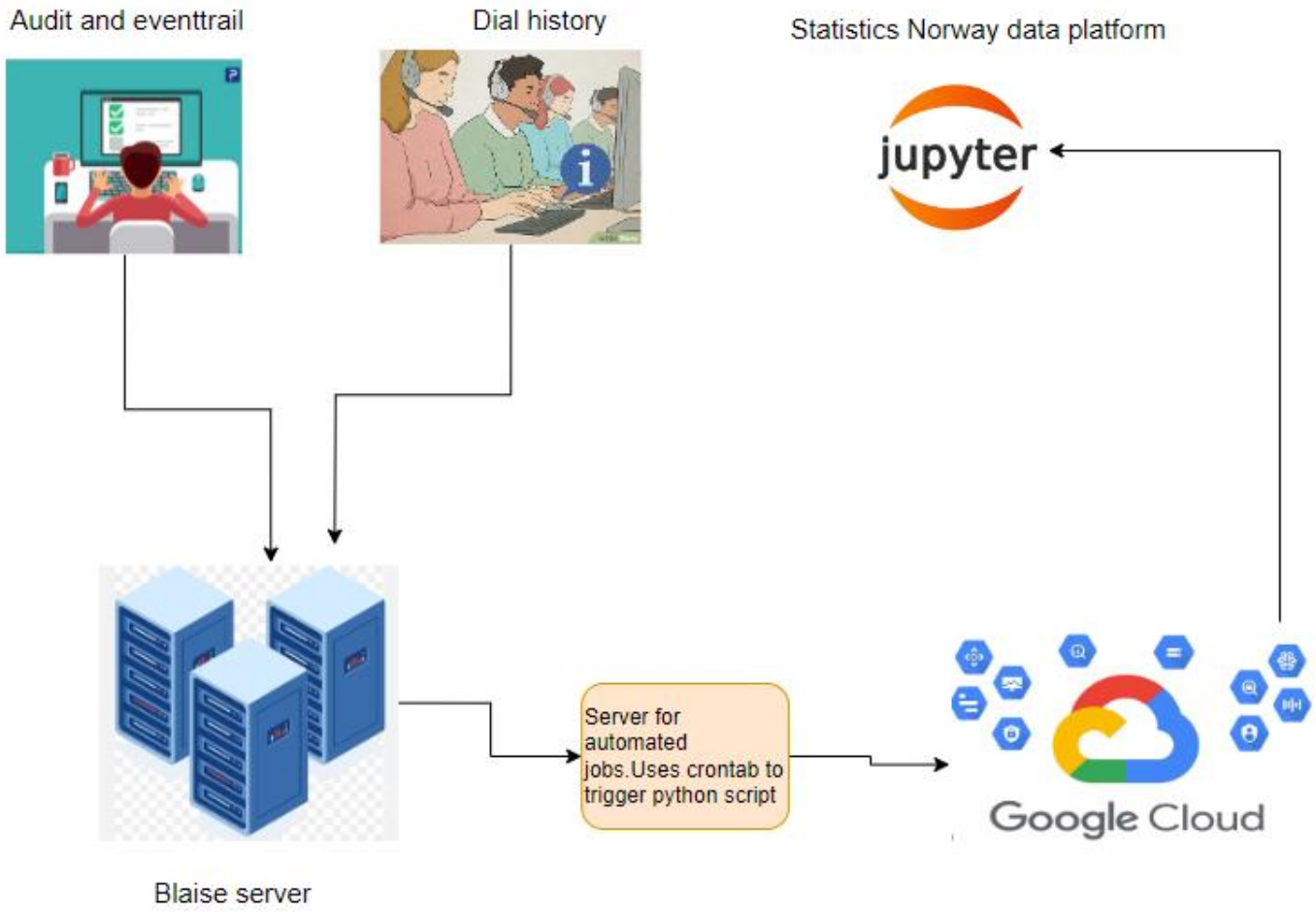


Why create a data pipeline?

- Limit access to databases
- Server capacity – slow loading for interviewers
- Colleagues with no coding proficiency should be able to access results over data collection

Pipeline

1. Export Data from Blaise to GCP
2. Parse and transfer data from Source bucket to Production bucket in GCP
3. Developing reports and programs in JupyterLab
4. Sharing daily results to internal webpage.



Audit trail parser

- <Tag Attribute="Attribute Value">
- Audit trail data from SQL database:

TimeStamp	Content	KeyValue	SessionId	InstrumentId
2023-02-28 08:14:40	<StartSessionEvent Width="414" Height="707" Device="Browser" Browser="Safari" Language="en" Platform="HTML" OS="Mozilla/5.0 (iPhone)" />	123aaa	{abc-12345}	{defg-678910-abc}
2023-02-28 08:15:59	<UpdatePageEvent LayoutSetName="SSB_Small_Touch" PageIndex="3" />	123aaa	{abc-12345}	{defg-678910-abc}
2023-02-28 08:15:59	<EnterFieldEvent FieldName="skjema. Tilfreds" AnswerStatus="Empty" />	123aaa	{abc-12345}	{defg-678910-abc}
2023-02-28 08:16:00	<LeaveFieldEvent FieldName="skjema. Tilfreds" Value="6" AnswerStatus="Response" />	123aaa	{abc-12345}	{defg-678910-abc}
2023-02-28 08:16:01	<ActionEvent Action="NextField()" ControlID="la_2kaba_7" />	123aaa	{abc-12345}	{defg-678910-abc}

Audit trail parser

Width	Height	Device	Browser	Language	Platform	OS	event	KeyValue	TimeStamp
414	707	Browser	Safari	en	HTML	Mozilla/5.0 (iPhone)	StartSessionEvent	123aaa	2023-02-28 08:14:40
							UpdatePageEvent	123aaa	2023-02-28 08:15:59
							EnterFieldEvent	123aaa	2023-02-28 08:15:59
							LeaveFieldEvent	123aaa	2023-02-28 08:16:00
							ActionEvent	123aaa	2023-02-28 08:16:01

SessionId	InstrumentId	LayoutSetName	PageIndex	FieldName	AnswerStatus	Value	Action	ControlID
{abc-12345}	{defg-678910-abc}							
{abc-12345}	{defg-678910-abc}	SSB_Small_Touch	3					
{abc-12345}	{defg-678910-abc}			skjema.Tilfreds	Empty			
{abc-12345}	{defg-678910-abc}			skjema.Tilfreds	Response	6		
{abc-12345}	{defg-678910-abc}						NextField()	la_2kaba_7

Audit trail Fill values – supporting module

Width	Height	Device	Browser	Language	Platform	OS	event	KeyValue	TimeStamp
414	707	Browser	Safari	en	HTML	Mozilla/5.0 (iPhone)	StartSessionEvent	123AAA	2023-02-28 08:14:40
							UpdatePageEvent	123AAA	2023-02-28 08:15:59
							EnterFieldEvent	123AAA	2023-02-28 08:15:59
							LeaveFieldEvent	123AAA	2023-02-28 08:16:00
							ActionEvent	123AAA	2023-02-28 08:16:01

SessionId	InstrumentId	LayoutSetName	PageIndex	FieldName	AnswerStatus	Value	Action	ControlID	VariableName	diff_time
{abc-12345}	{defg-678910-abc}									NaN
{abc-12345}	{defg-678910-abc}	SSB_Small_Touch	3	skjema.Tilfreds					Tilfreds	21
{abc-12345}	{defg-678910-abc}		3	skjema.Tilfreds	Empty				Tilfreds	0
{abc-12345}	{defg-678910-abc}		3	skjema.Tilfreds	Response	6			Tilfreds	1
{abc-12345}	{defg-678910-abc}		3	skjema.Tilfreds			NextField()	la_2kaba_7	Tilfreds	1

Developing reports

- JupyterLab in Statistics Norway Cloud based platform DAPLA
- Query parsed paradata
- Supporting programs
- Develop and run reports:
  - Call history (With sample information, non-response, number of call per respondent)
  - Fieldname (Non-response, Previous Page, Errors, Response time, change of response)
  - Response time / interview time – respondent level

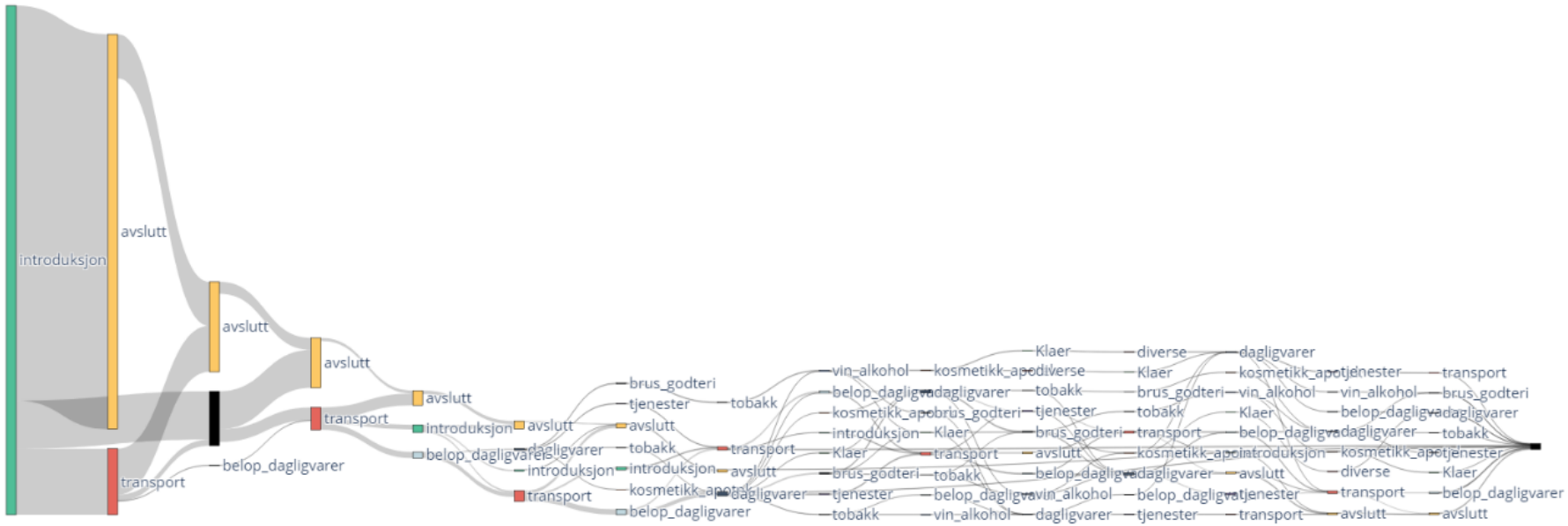
Paradata use – Survey project managers

- Sharing daily results to internal web page
- Monitoring data collection
- Initiate measures and monitor effectiveness

Statistisk sentralbyrå
Statistics Norway

Paradata use – Survey project managers

- Survey flow in questionnaire



Statistisk sentralbyrå
Statistics Norway

Paradata use – Survey Methodologists

- Paradata can be beneficial addition to qualitative data, e.g. user testing or focus groups
- In pilot surveys, paradata can be used to identify users experiencing issues with survey

Next steps

- Paradata reports – improving how we analyze paradata, understanding what we need to look at
- Interactive dashboard instead of static web page for the daily results

TALL SOM FORTELLER

Questions

Statistisk sentralbyrå
Statistics Norway