Dapresy Pro Gavagai integration





Index

1 – Introduction
2 Process
3 Gavagai
3.1 File format in initial upload5
3.2 Topic names
3.3 Only pinned topics are imported to Dapresy Pro6
3.4 Groups are not imported to Dapresy Pro6
3.5 Sentiment data6
4 Dapresy Pro7
4.1 Create an External data service job7
4.2 Run a job manually9
4.3 Run a job via import scheduler12
4.4 Data format13



1 - Introduction

This document is a guide to creation and use of the integration between Dapresy Pro and the Gavagai platform. It does not cover creation of models or projects in Gavagai. To use this integration, you will need to have an account with Gavagai. Please contact Gavagai to facilitate this: <u>http://gavagai.se/</u>

The integration between Dapresy Pro and Gavagai enables machine learning based text analysis on open ended comments which can be visualized in Dapresy Pro.

If you would like to know more about this integration, please contact Dapresy Global Support at: support@dapresy.com and they will be able to assist you.

Best Regards,

Dapresy Team

Email: support@dapresy.com



2 Process

Below you see an overview of the process for setting up the connections between Dapresy Pro and Gavagai.

- Create a project in Gavagai and setup the model to be used for text analyzes
 - One project per open ended comment to be analyzed needs to be created. If multiple languages are available create one Gavagai project per open ended comment and language combination
- Create a project in Dapresy Pro
 - Create an "External data service" job in the project. You have to create one job per Gavagai project
 - The jobs can now be run manually on selected data batches and/or run automatically when new data is imported via the Import scheduler

When running the External data service job the following main steps are done:

- Data (open ended comments) is exported from Dapresy
- All previous imported data is deleted in the Gavagai project
- New data is imported to the Gavagai project
- The open-ended comments are analyzed in Gavagai
- Analyzed data is exported from Gavagai
- The analyzed data is transposed into a stacked data format
- The stacked data is imported to Dapresy Pro as a new Stacked data batch
- The newly imported data batch is activated automatically.

In the following chapters you find more details about how to setup the projects in Gavagai and the External data service jobs in Dapresy Pro.



3 Gavagai

When creating and managing the project in Gavagai there are a few things you need to be aware of:

- The needed file format in the initial data import to the Gavagai project
- The Topic names, what happens if the name is changed?
- Which Topics are exported from Gavagai?
- How Grouped Topics are handled
- Sentiment data is exported per Topic and not per comment

Read more about the points above in the following chapters.

3.1 File format in initial upload

When creating a project in Gavagai the initial data file you import shall contain the following 3 columns only, the order of the columns needs to be followed as well.

- BatchID
- RespondentID
- OpenText

Here we see an example of a data file to be imported to Gavagai for setting up the project. The BatchID column shall be empty. The RespondentID column will include the respondent ids and the Open text will include the comment to be analyzed.

	А	В	C
1	BatchId	RespondentID	Opentext
2		1	Bed was to hard. The lunch was great!
3		2	good selection of wine in the restaurant :)
4		3	the room was to warm. The sauna was closed which was a big disappointment
5		4	The room was nice, good size
6		5	very friendly staff
7		6	The car parking is to expensive
~		-	.1.1

3.2 Topic names

A Topic does not have any special identifier in the Gavagai database so when a topic is imported to Dapresy Pro the Topics are imported as "strings" which are recoded to a categorical variable during the automatic import. Due to that you will have to update the topic name in both Gavagai and in Dapresy Pro in case of having the need for updating a Topic name. If you miss to change name in Dapresy Pro you will end up in a situation like below:



Example: In Gavagai you have the topics Rooms, Restaurant, Bar, Reception and Car Parking. The Topic Rooms is changed to Room in Gavagai only so in Dapresy Pro you end up with the following topics:

- Rooms
- Restaurant
- Bar
- Reception
- Car Parking
- Room

If this situation occurs you can use the data recoding function so all data in the Rooms answer options is recoded into Room and then you can Inactivate Room<u>s</u> so it does not show up in any report user interfaces.

3.3 Only pinned topics are imported to Dapresy Pro

In Gavagai a Topic can be pinned or not pinned. Only the pinned Topics are imported to Dapresy Pro, this gives you control over the topics imported.

3.4 Groups are not imported to Dapresy Pro

If you have been creating Groups of Topics in Gavagai these will not be imported to Dapresy Pro. It is only the topics and the relevant sentiment data that becomes imported.

If a Group is pinned all the topics in this group are imported to Dapresy Pro.

If you want to view the result of a group of Topics these Groups can be created as Grouped answer or by computing new variables in Dapresy Pro.

3.5 Sentiment data

In Gavagai you can either get a Sentiment result for the whole comment or a Sentiment result for each topic. In the integration between Dapresy Pro and Gavagai each topic will have a sentiment. The sentiment for the whole comment is not imported to Dapresy Pro



4 Dapresy Pro

To connect a project in Dapresy Pro to one or several projects in Gavagai Explorer you have to create an External Data service job per Gavagai project. Each job will manage data for one Open ended comment variable and one language.

4.1 Create an External data service job

The External data service jobs are created in the "External data service page" which can be found in the ribbon menu as shown in the image below. If you don't see the External data service page contact Dapresy Helpdesk to get access to the page.

Here we see how to enter the External Data service page.

Start	Project Settin	gs Data	Questions a	& Answers	New Variabl	es/Values	Reporting Objects
Import data	Activate data	Generate reports	Delete Meta data	Delete Case data	Export Case data	Data Cleaning/ Recoding	
Schedule im	ports	Generate	Meta Data		Case Data		
External data	a services	ices					
Import XML							

The External data service page lists all your created jobs as shown below. You can see the Last run time and Last result and potential number of failed batches.

Here we see the External data service page. In this example two jobs have been created.

External data services	5				
Name	Туре	Status	Last run time	Last result	Failed batches
Swedish data	Gavagai	Active	2018-06-20 11:55	Successful	0
Danish data	Gavagai	Active		Not run	0

To create a new job click "Create" in top of the page and the view for creating and managing the new job appears.

Here we see the screen for creating an External data service job for Gavagai.



Comment 1 -Swedish	n data		×
Create / Edit Run Logs			
Name Status Select external service provider	Comment 1 -Swedish data Active Gavagai EXTERNAL SERVICE PROVIDER SETTINGS	~	
	Username Password Gavagai project ID Select variables to be analysed Select data filter	tp@dapresy.com 12527 SGS_OpencommentOPEN Do you have any further comments that could help us? Language Guest Form - Language Guest Form × 10 - Swedish	•
	Generate "Missing" Topic 🕐	V	
	Save		

Follow the steps below to create a new job:

- Enter the name of the new job
- Set status to Active by ticking the check box to be able to run the job later on (only active jobs can be run)
- Select Gavagai as the External data service provider
- Enter the Username and Password of your Gavagai account
- Enter the project ID in Gavagai which the current job shall be connected to (see further down how to find the project id in Gavagai Explorer)
- Select the Open ended comment to be sent to Gavagai for evaluation
- If you have multi languages in your data the job must be filtered by a language variable so only comments belonging to a specific language are sent to Gavagai, use the data filter to achieve that.
- Select if comments that that didn't get any Topic in the analysis will get a system generated Topic named "Missing" or not. See explanation further down.



Here we see how to find the project id in Gavagai Explorer. You find it in the URL when you are working in the project in Gavagai, in this example the project id is 12433.

.V	Gavagai Explorer Simple Lang 🗙 🕂	
	🛈 🎤 🔒 https://explorer.gavagai.se/projects/12433/edit	

The option for generate missing topic will affect the results, like percentage shares, as when results are calculated in Dapresy Pro only respondents which have an answer (a topic) in a question are included in the base size in the calculation. If the "Generate missing topic" option is ticked all comments that do not generate any Topic will get a system generated topic named "Missing" which means that all respondents with a comment are included in the base when, for example, calculating percentage results.

If the option is not ticked the respondents with a comment which did not generate any topic are not included in relevant base values.

4.2 Run a job manually

Once created a Gavagai job it can be run manually for one or multiple data batches in your Dapresy Project. To do that click the Run/Re-run button shown in the image below.

Create new						
lame	Туре	Status	Last run time	Last result	Failed batches	
C		Q	Q Q	~		Run / Re-run
Comment 1 -Swedis lata	h Gavagai	Active	2018-06-28 06:43	Successful	36	/ C 🔍 🖻

Here we see how to enter the Run mode.

In the Run mode you select which data batches you want to send to Gavagai for analysis. Select the desired data batches and clicks the Run button to start the process. You can run one or multiple data batches at the same time.

Here we see the Run mode, select the batches to be run in the most left column and hit Run to start the process.



Create	/ Edit Run	Logs						
Select	Data Batch	File Name	File Format	Batch type	Import date	Respondents	Import tag	Status
	Q)	٩	·	~	٩	Q)	Q,	*
•	29.0	170111.sav	SPSS	New Data	2017-01-15 16:01	696		Not run
	28.0	imported_20170111130125_	SPSS	New Data	2017-01-15 16:01	380		Not run
	24.0	Data File 2016 I	SPSS	New Data	2017-01-12 09:55	302115		Not run
	21.0	170109.sav	SPSS	New Data	2017-01-11 13:01	28		Not run
	15.0	Data File 2015	SPSS	New Data	2016-12-21 16:33	279781		Not run
	14.0	Data File 2014	SPSS	New Data	2016-12-21 16:31	283133		Not run
Showing	541 to 547 of 54	7 entries			Previous	1 51 52 53 54 55 N	ext Sho	w 10 🗸 entries
Run								

When starting the job the status is shown in the bottom of the screen as shown in the image below. The job is completed when the new data set with analyzed data has been imported and activated in Dapresy Pro, as successful message is shown when the process is completed.

The time the process takes is dependent on the data volume sent to Gavagai but also on the size of the project in Dapresy Pro as the newly imported data batch is activated during the last step of process. The more data a project contains the longer time the data activation takes.

Here we see an example of the status of a running process.

0	Processing data batches: 421 18:14:28 Starting gavagal external data service process 18:14:28 Getting data batch 421 from Dapresy Pro 18:14:29 Importing data batch 421 into Gavagal 18:14:29 Polling Gavagal project status 18:14:37 Processing step 1 of 11: Polling Gavagal exploration status 18:14:38 Processing step 2 of 11: Polling Gavagai exploration status 18:14:38 Processing step 3 of 11: Starting Gavagai project exploration	
Rur		

...



Note; you can also always check the status of currently running jobs and previous run jobs in the Log tab as shown below.

Here we see the Log tab which can be used to check the status and any potential error messages if the process fails.

Create / Edit Run Logs	S		
Time	 Event Type 	Message	
	۹.		Q,
2018-06-28 18:14:46	Info	Processing step 4 of 11: Polling Gavagai exploration status	
2018-06-28 18:14:38	Info	Processing step 3 of 11: Starting Gavagai project exploration	
2018-06-28 18:14:38	Info	Processing step 2 of 11: Polling Gavagal exploration status	
2018-06-28 18:14:37	Info	Processing step 1 of 11: Polling Gavagai project status	
2018-06-28 18:14:29	Info	Polling Gavagai project status	
2018-06-28 18:14:29	Info	Importing data batch 421 into Gavagai	
2018-06-28 18:14:28	Info	Getting data batch 421 from Dapresy Pro	
2018-06-28 18:14:28	Starting	Starting gavagai external data service process	
2018-06-28 06:57:27	Successful	Gavagai process finished succesfully	
2018-06-28 06:46:45	Info	Processing step 11 of 11: Activating imported stacked data batches 761 in Dapresy Pro	
howing 11 to 20 of 910 entries		Previous 1 2 3 4 5 91 Next Show 10	✓ entrie

Note: if the process gets stuck in the step "Processing step 6 of 11: Polling Gavagai report generation status" for a long time Gavagai has failed in the report generation step, you can always enter the Gavagai platform to view the current status of the report generation in the "Download report area" (see image further down). If the generation gets stuck on 0% or 99% it will probably not be completed, then you have to "Delete" this report in Gavagai and the process will be canceled.

Here we see the Download area in Gavagai. It shows the status of the current report generation, when the progress percentage is 100% the generation in Gavagai is completed and data will be imported to Dapresy Pro. If it gets stuck in 0% or 99% it is an issue in Gavagai and you will have to delete the report generation by clicking the Delete icon.



DOWNLOAD REPORTS

You can create a report from the 'Save as...' drop down meny in the Explore screen and it will appear here. To see the exact creation date, hover over the started at time.

id 23505, started 2 days ago, of type EXCEL, is IN_PROGRESS (0%) î id 23411, started 4 days ago, of type EXCEL, is FINISHED î id 23397, started 4 days ago, of type EXCEL, is FINISHED î id 23392, started 4 days ago, of type EXCEL, is FINISHED î id 23388, started 4 days ago, of type EXCEL, is FINISHED î

Still in progress? Please reload the page to see if the reports are done.

As described above the data that is coming back from Gavagai is imported as a new Stacked data set in Dapresy Pro. These data sets are displayed in the Activate data page in same way as all other data sets.

Here we see an example of data batches in the Data activation page that have been imported via the Gavagai integration. In the image below the name of a data batches is "Swedish test – 51 – 598.csv". "Swedish test" comes from the External data service job name and 598 is the id of the data batch sent to Gavagi. 51 is an internal id number used for tracking errors etc.

✓ 599.0	Swedish test - 51 - 598.csv	Excel / CSV	Stacked Data	2018-06-25 15:58	137	Edit
✔ 600.0	Swedish test - 52 - 565.csv	Excel / CSV	Stacked Data	2018-06-25 16:30	160	Edit
✔ 601.0	Swedish test - 52 - 566.csv	Excel / CSV	Stacked Data	2018-06-25 16:30	164	Edit
✔ 602.0	Swedish test - 52 - 568.csv	Excel / CSV	Stacked Data	2018-06-25 16:30	172	Edit
✔ 603.0	Swedish test - 52 - 569.csv	Excel / CSV	Stacked Data	2018-06-25 16:30	91	Edit
✔ 604.0	Swedish test - 52 - 570.csv	Excel / CSV	Stacked Data	2018-06-25 16:30	106	Edit

4.3 Run a job via import scheduler

The Gavagai jobs can be run automatically on data batches imported via the Import scheduler. If you have created Gavagai jobs, and these have been set to Active, these can be selected when creating an Import scheduler job. The selected jobs will be run automatically as the last step in the import process after the survey data has been imported and activated.

Here we see the option for selecting the Gavagai jobs to be run automatically when importing data via the Import scheduler.



Schedule name: data Next scheduled import: 2018-06-27 10:00 Import tag: <no tag=""> Import type: Single import Import type: Single import Oate: 08/27/2018 Current server time: 2018-06-28 18:25:00 Time: 10:00 Update to existing dataset(s): Select mapping variable Stacked data to be connected to existing data sets Data options: Import Metadata Activate Data Append dataset to active data Delete inactive data Activate Latest Update HFilter Report pre-generation: Generate reports Data mapping: <none></none></no>	EDIT SELECTED IMPORT SCHEDULE		
Next scheduled import 2018-06-27 10:00 Import tag: Import Import type: Single import Date: 06/27/2013 Current server time: 2018-06-28 18:25:00 Time: 10:00 Update to existing dataset(s): select mapping variable Select mapping variable import Metadata Import Metadata Activate Data Append dataset to active data Delete inactive data Activate Latest 1 v datasets Filter by selected tag Delete input variable data Update HFilter Report pre-generation: Generate reports v Data napping: Import v External data services: import import	Schedule name:	data	✓ Active
Import tag: <no tag=""> Import type: Single import Date: 06/27/2018 Current server time: 2018-06-28 10:00 Update to existing dataset(s): Select mapping variable Stacked data to be connected to existing data sets Data options: Import Metadata V Activate Data Delete inactive data Delete inactive data Delete input variable data Update HFilter Report pre-generation: Cenerate reports Data mapping: <none></none></no>	Next scheduled import:	2018-06-27 10:00	
Import type: Single import Run import on save Date: 06/27/2018 Current server time: 2018-06-28 18:25:00 Time: 10:00 Import Update to existing dataset(s): Select mapping variable Import Stacked data to be connected to existing data sets Import Metadata Import Metadata Activate Data Append dataset to active data Filter by selected tag Delete inactive data Activate Latest Import datasets Delete inactive data Activate Latest Import datasets Delete input variable Import datasets Filter by selected tag Delete input variable Import datasets Import datasets Delete input variable data Import datasets Import datasets Delete input variable data Import datasets Import datasets Delete input variable data Import datasets Import datasets Data mapping: Import second Import datasets External data services: Import connect Import dataset	Import tag:	<no tag=""> v Edit tags</no>	
Date: 06/27/2018 Time: 10:00 v Update to existing dataset(s): Select mapping variable Stacked data to be connected to existing data sets Data options: Import Metadata Vactivate Data Oblete inactive data Delete input variable data Update HFilter Report pre-generation: Generate reports Data mapping: <none></none>	Import type:	Single import V Run import on save	
Report pre-generation: Generate reports Data mapping: <none> External data services: × Comment 1 -Swedish data</none>	Data options:	Date: 06/27/2018 Current server time: 2018-06-28 18:25:00 Time: 10:00 • Update to existing dataset(s): Select mapping variable • Stacked data to be connected to existing data sets • • Import Metadata • Activate Data • Delete inactive data • Activate Latest 1 Delete input variable data • Update HEilter	Filter by selected tag
Data mapping: <none> External data services:</none>	Report pre-generation:	Generate reports	
External data services: Comment 1 -Swedish data	Data mapping:	<none> v</none>	
	External data services:	Comment 1 -Swedish data	
Data source: FTP Server	Data source:	FTP Server	Retry import if unsuccessfull

4.4 Data format

As described above the data imported to Dapresy Pro is imported as stacked data as one comment, for one respondent, might generate multiple Topics and each topic gets a positive, skeptical or negative sentiment.

Here we see an example of the data sent to Gavagai (left table) and the stacked data that imported to Dapresy Pro (right table). As shown the comments for some respondents, like id 1 and 3, generated multiple topics and due to that the stacked data format is needed as each topic has related sentiment data



Respondentid	Comment
1	Bed was to hard. The lunch was great!
2	good selection of wine in the restaurant :)
3	the room was to warm. The sauna was closed which was a big disappointment
4	The room was nice, good size
5	very friendly staff
6	The car parking is to expensive
7	nothing

Analsyed data imported to Dapresy Pro					
Respondentid	Topic	Positive sentiment	Skeptical sentiment	Negative sentiment	
1	Bed			1	
1	Lunch	1			
2	Drinks			1	
3	Room			1	
3	Sauna			2	
4	Room	1.5			
5	Staff	1			
6	Car parking		1		
7	Missing		1		

When a job is run the first time the Topic and Sentiment variables becomes new variables in the Dapresy Pro project. The Topic variable is imported as a categorical variable and the three Sentiment variables are imported as Numeric variables.

The default question code and question text of the Topic and Sentiment variable are based on the code of the open ended comment sent to Gavagai. If the question code of the open ended comment was "Q34" the following four new questions will appear in your project:

Code	Question text	Explanation
Q34Topic	Q1Topic	The Topic variable
Q34Negativity	Q1Negativity	The negative Sentiment variable
Q34Positivity	Q1Positivity	The positive sentiment variable
Q34Skepticism	Q1Skepticism	The skeptical sentiment variable

The question texts of these new variables can then be updated in the "Question page" as always.