



# The Data Tales

research consortium

background information on

## STORY MINING

digital narrative identification

### Coordinator:

Dr. ir. Marlies van Steenbergem | Lector Digital Smart Services | Hogeschool Utrecht |  
+31 6 51382452 | [marlies.vansteenbergen@hu.nl](mailto:marlies.vansteenbergen@hu.nl)

### Assistant coordinator:

Drs. Philippine Waisvisz | PhD researcher & project manager | Hogeschool Utrecht |  
+31 6 146 72238 | [philippine.waisvisz@hu.nl](mailto:philippine.waisvisz@hu.nl)

version: 14 december 2017



# Story Mining

## The data-basis for identity economics

Story Mining is a data-based technique to assess how identity influences decision making.

'A story', the development of an interaction between a protagonist and, for example, an organisation, is an expression of the identity of the protagonist. Economists like George Akerlof recognize the importance of these stories for the purpose of analyzing decision making. However, how to model narratives for the purpose of analysis and policy design is still an unsolved riddle.

Our consortium seeks to construct a method for modelling narratives. We extract the four main elements of story telling (actors, emotions, motivations and actions) from a variety of data. We put them together and construct the story. Thus we can, for example, determine why people decide to buy a product or why employees engage in fraudulent behaviour. Knowing how actors and processes interact we can assess the root cause for (non)performance. These insights yield opportunities for interventions.

### A bird's view on story mining

A narrative requires

- 1- fact-based description of the actor(s)
- 2- emotional experience of the actor(s)
- 3- motivation of the actor(s)
- 4- actions and development over time.

These elements can be assessed with data science technology:

1 - Factual description of the actor: e.g. age, gender, interests, location. These are drawn from available data. For this purpose general data science techniques are applied, mostly handling structured data (i.e. data in excell-sheets).

To identify the narrative components emotion (#2) and motivation (#3) we take digital communication and apply known text mining techniques to unstructured data (e.g. email an chat). There is some tooling available for emotion mining, but motivation mining we we have been developing from scratch.

You may be familiar with sentiment mining, which labels texts as positive or negative. We do the same, but we add these classifications:

- Emotion mining will label text with [Ekman's model](#).
- Motivation mining will label texts according to the full motivational model of [Heckhausen](#) (as introduced by [Tania Singer and Dennis Snower at the INET conference in Paris](#)).

For each actor, we put all data mining results from the story elements #1, #2 and #3 into a virtual box ('a case') Then we load all cases into process mining software to see how these cases correlate to actions and development over time. This generates a story line suitable for analysis.

Process mining builds on the principle that actions are registered in a system, labeled with an actor, an action and a timestamp. The software puts all data in a chronological order. With that information the software builds an exact picture of how a process flows. The output is an animation showing how cases run from beginning to end and how actor's emotions and motivations develop throughout their journey. For example, a case may keep its caring (motivation) label throughtout the journey while emotions vary from fear via anger to happiness.

*See page 4 for an illustration of process mining and page 5 for a simple overview of our research model.*

**Economic identity:** the utility that a person attaches to certain matters whilst acting in an economic interaction.  
**Decision making:** the conscious or unconscious process resulting in the selection of a course of



## Pilot findings

Results from our pilots show that it is possible to recognize motivation in text. We can link specific words and linguistic features to specific motivations. For example: the use of 'you' is associated with Anger. The use of 'we' and inversed word order (e.g. a question) correlate with Care. This can yield very interesting, unexpected insights. We have found for example that – even when the subject is 'study grants' – the dominant motivation is Care, not Consumption (i.e. acquisitiveness).

## Applications

Story mining will be able to model, map and visualize economic identities. It can do so in the context of various processes, even within very large institutions. It might potentially be applied to social developments. It can be applied to governance, social, marketing, management, and many other areas. The output can provide information to leaders, managers and other stakeholders as to any (un)intended side effects of their policies and communication. For example, story mining might indicate that certain areas lack 'Care' and stimulate 'Aggression' and 'Power'.

Users will be able to assess the root causes of problems, so they can seek sustainable solutions. It will help them to create healthy processes and foster sustainable, economic identities.

## Consortium The Data Tales

Story mining requires a variety of (big) data techniques and knowledge of governance, behavioural and institutional economics. The consortium covers all areas of the required disciplines and include leaders in their fields:

- **Behavioural and institutional economics:** [Groningen University](#), Faculty of Economics and Business, Department of Innovation Management & Strategy, Thijs Broekhuizen
- **Governance, law, ethics:** [Saxion University of Applied Sciences](#), Project Group Governance, Willeke Slingerland
- **Text mining, e-humanities:** [VU Amsterdam](#), Computational Lexicology, Piek Vossen
- **Process mining, responsible data science:** [TU Eindhoven](#), Faculty of Mathematics and Computer Science, Marwan Hassani
- **Process mining, text mining, marketing:** [Utrecht University of Applied Sciences](#), Research Centre for Innovation in Business and Communication, Marlies van Steenbergen

Organisations sharing data and process information (Dutch):

- Healthcare provider [SAZ](#)
- [DUO](#), grant office of the Dutch Ministry of Education
- [Utrecht University of Applied Sciences](#), marketing department
- [Aegon](#), insurance
- [Live Chat Service](#), digital helpdesk service provider

We are cooperating with excellent SME consultancy firms specialized in the use of data analysis and/or data driven process management in practice : [Customer Pulse](#), [Data Kitchen](#), [O&I](#).

## Contact

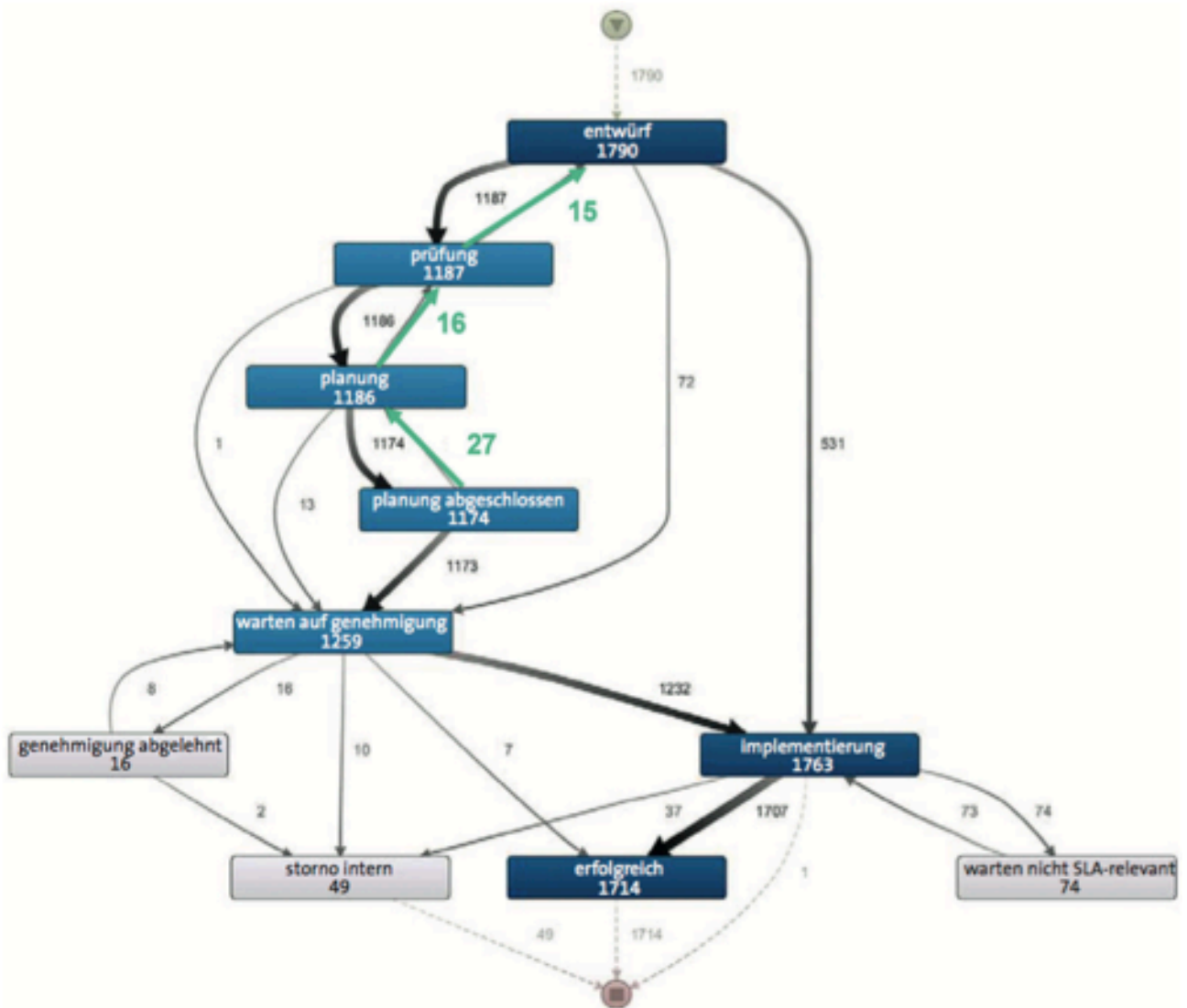
Dr. ir. Marlies van Steenbergen    +31 6 51382452    marlies.vansteenbergen@hu.nl  
Drs. Philippine Waisvisz        +31 6 146 72238    philippine.waisvisz@hu.nl

## PROCESS MINING

The picture below is an example of a process map, generated with process mining.

Process mining collects all actions registered in an organisation's systems. Firstly it organises them according to 'actor', 'action', and 'time'. Next it builds a process map that visualizes which process the actors follow. Using the factor time, process mining creates a series of snapshots, chronologically revealing steps in the course of events. The most exciting visualisation rapidly presents all those snapshots as a revealing animation.

In the picture below, the blocks represent registered steps in the process. The lines show the flow of actions. The numbers indicate a frequency (but could also be used to indicate duration).



Source: fluxicon.com

## RESEARCH MODEL

Based on their identity, people undertake actions: this shows in their communication and in their behaviour. Communication provides us with digital research data as input for text mining. Behaviour leads to data for process mining. When we analyse the aggregated data we can construct the actor's identities and stories.

