

TimeFlows: Visualizing Process Chronologies from Vast Collections of Heterogeneous Information Objects

MAX LONYSA MULLER*, Nationaal Archief, The Netherlands

ERIK SAAMAN, Nationaal Archief, The Netherlands

JAN MARTIJN E. M. VAN DER WERF, Utrecht University, The Netherlands

CHARLES JEURGENS, Universiteit van Amsterdam, The Netherlands

HAJO A. REIJERS, Utrecht University, The Netherlands

In many fact-finding investigations, notably parliamentary inquiries, process chronologies are created to reconstruct how a controversial policy or decision came into existence. Current approaches, like timelines, lack the expressiveness to represent the variety of relations in which historic events may link to the overall chronology. This obfuscates the nature of the interdependence among the events, and the texts from which they are distilled. Based on explorative interviews with expert analysts, we propose an extended, rich set of relationships. We describe how these can be visualized as *TimeFlows*. We provide an example of such a visualization by illustrating the Childcare Benefits Scandal – an affair that deeply affected Dutch politics in recent years. This work extends the scope of existing process discovery research into the direction of unveiling non-repetitive processes from unstructured information objects.

Additional Key Words and Phrases: Timeline, Process Chronology, TimeFlow, Parliamentary Inquiry.

1 INTRODUCTION

In various situations, it is worthwhile to determine retrospectively how a process has unfolded over time. The established discipline of *process mining* provides algorithms for discovering *work processes* from historic data. Such processes are cyclic in nature and can be discovered on the basis of events generated by IT systems. However, there are also processes worth reconstructing that are non-repetitive and lack a log of structured events.

Consider a *parliamentary inquiry*, which is an important use case for the work presented in the paper. A parliamentary inquiry is initiated to gain knowledge on a particular topic when a matter of national importance is at stake. The objective is to reconstruct the process that led to severe consequences, either for the government itself or for the citizens involved. The process in question, once reconstructed, is typically represented in the form of a *process chronology*, e.g. a timeline, a textual summary, or a fact sheet of events. The purpose of a process chronology is to capture what relevant events happened at different moments in time and how these events relate to one another.

The generation of a process chronology, either in the context of a parliamentary inquiry or any other sort of fact-finding investigation, may need to take place on the basis of a wide range of information objects. While such objects may be structured, it is often the case that relevant events are included in emails, text messages, policy briefs, reports, and other unstructured information objects. The analysis of such events is time-consuming as it is mostly done by hand and relies heavily on domain knowledge. A major challenge for these analysts is to present the process of interest in a concise and intuitive way. The objective of this paper is threefold: (1) to conceptualize the the events and texts of a process chronology in a meaningful way, (2) to provide an intuitive visualization thereof, and (3) to set an agenda for process chronology research paving the way for

*Corresponding Author

Authors' addresses: Max Lonysa Muller, max.muller@nationaalarchief.nl, Nationaal Archief, Den Haag, The Netherlands; Erik Saaman, erik.saaman@nationaalarchief.nl, Nationaal Archief, Den Haag, The Netherlands; Jan Martijn E. M. van der Werf, j.m.e.m.vanderwerf@uu.nl, Utrecht University, Utrecht, The Netherlands; Charles Jeurgens, k.j.p.f.m.jeurgens@uva.nl, Universiteit van Amsterdam, Amsterdam, The Netherlands; Hajo A. Reijers, h.a.reijers@uu.nl, Utrecht University, Utrecht, The Netherlands.

automation. We view this work as a first step of a broader ambition to study requirements, design, and the automated generation of interactive process chronologies.

Our proposal links to earlier work, which deals with the generation of *timelines* on the basis of a set of textual documents as input. The Timeline Summarization (TLS) line of research aims at automatically identifying key dates of major events from a corpus of texts along with short descriptions of what happened on these dates [16]. Although applicable in settings where the datasets are small and homogeneous, a drawback of these linear timelines is that they quickly become difficult to interpret, e.g. when the amount of dates grows too large or the textual topics are heterogeneous [21].

More recently, in an attempt to break away from the simple, linear structure provided by TLS, *timelines with graph-based representations* (TGRs) were proposed. These structures provide a more intricate structure for the representation of related events (e.g. Figure 1). Pivotal work in this area was carried out in the early 2010s by Shahaff and others [14]. A recent survey on the topic is provided in [8]. A drawback of TGRs, however, is that they center around one type of relation that ties the various events together. In reality, a wide range of relationships, e.g. causality, topicality, time order, etc., may be required to show how events contribute to the overall chronology of the process. In addition, the TGRs fail to relate to the original information objects that they are based on.

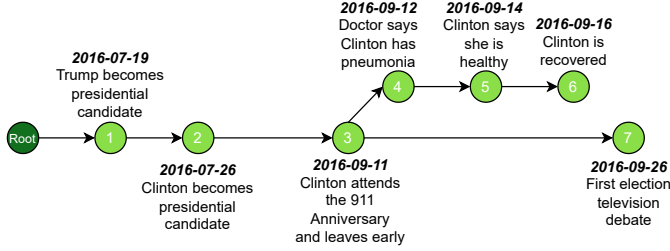


Fig. 1. This US elections TGR [9] links events through the *Subject Relation*.

In this paper, we propose *TimeFlows* as a way to visually represent process chronologies. They extend TGRs to represent collections of related events. Our proposal improves upon the state of the art in that it (1) provides flexibility for the inclusion of a wide range of relationships between events and (2) includes references to the constitutive information objects for the presented events. TimeFlows are applicable within a wide range of applications, including historical and forensic research. Their objective is to come up with a concise yet intuitive reconstruction of a process based on a heterogeneous and potentially large set of information objects.

This paper is structured as follows. In Section 2, we delve deeper into related work. Section 3 outlines our research method. It describes our approach to identify the types of relationships that are important for a process chronology through a set of structured interviews with expert analysts. Then, in Section 4, we (1) present the results obtained by exploratively coding the interviews, (2) motivate our choice of relationships to be included, and (3) provide a Process Delivery Diagram, which depicts how manual constructions of process chronologies takes place. In Section 5, we present a model for process chronologies and describe how TimeFlows can be used to visualize these. We illustrate our proposal by presenting a TimeFlow for the Dutch Childcare Benefits Scandal¹. We conclude the article with an agenda for further research (Section 6).

¹See: https://en.wikipedia.org/wiki/Dutch_childcare_benefits_scandal

2 LITERATURE REVIEW

Different models for TGRs have been proposed [6] [1], all of which aim to represent what we have coined *process chronologies*. Several algorithms have been devised to automatically generate such graphs, as highlighted in a recent overview [8]. It includes the approaches described in [14] and [20].

The authors of [6] anchor their definition of a *narrative map* (a synonym of a temporal graph) by means of a weighted, directed, acyclic graph. Edges between events are made solely on the basis of topical or thematic coherence. The degree of coherence determines the weight assigned the edges.

In [1], the authors define yet another synonym for a temporal graph, namely a *Story Graph*. It is considerably more elaborate and complicated than the aforementioned definition. The authors define notions such as *Related Events* (by means of a user-defined coherence threshold), *Topic Theme*, and *Timeline Path*. The authors of [9] graphically depict four different types of structures to characterize a story. These consist of (a) a flat structure, (b) a timeline structure, (c) a graph structure, and (d) a tree structure. In all of these are graphs the nodes indicate events, while the edges denote a topical coherence among these events. As noted in [8], some frameworks allow for forms of interactivity. For instance, in [14], the authors enable the functionality of zoomability of the events, to inspect them at finer or coarser levels of granularity. Moreover, the authors of [15] propose the introduction of a *personalized coverage function* that can be employed to display temporal graphs that fit the users' needs and preferences, and suit their own queries.

What these approaches have in common, is that they all assume there is only one type of edge - a topical or thematic connection between events. Yet recent research has shown that analysts make different types of connections between events [7]. For instance, they also construct connections on the basis of causality, entities, or specific domain knowledge. In the present paper, we seek to find out which types of relations are most relevant to analysts, and propose a visualization technique process chronologies that captures the variety of relevant relations among events, called *TimeFlows*. Moreover, we investigate how to integrate the documents or texts - more broadly: information objects - on which these events are based within our conceptualization as well.

3 RESEARCH METHOD

We aim at grounding our proposal in data obtained through explorative, semi-structured interviews. These interviews are conducted with people having experience with manually constructing process chronologies in their professional work in their own, specific contexts. By analyzing the answers they have provided us, we seek to determine what kind of challenges analysts face when manually constructing process chronologies, in particular with respect to the relations they wish to establish between relevant events.

We collect data by recording interviews with four interviewee groups. These interviews are transcribed and exploratively coded. Based on the results, we construct a Process Delivery Diagram (PDD) in Section 3 and present our proposal for TimeFlows in Section 4.

3.1 Interviewee Groups

We selected interviewee groups on the basis of their professional experience with creating process chronologies. All participants have been actively involved in the creation of process chronologies based on unstructured information objects; three of them had an affinity with parliamentary inquiries. We conducted four interviews for our data collection:

- (I1) Four researchers of a Dutch Ministry that have experience in creating chronologies for Parliament.

- (I2) An interview with a junior council advisor who assisted a minister before and during the hearings for a parliamentary inquiries.
- (I3) An investigative journalist who works for a well-known Dutch publication platform.
- (I4) Two national government officials that have experience as a coordinator of several parliamentary inquiry research staff.

These interviewee groups are referred to as (I1) through (I4) throughout this paper. When we quote someone from a particular group, it is implied that their views correspond to the consensus opinion of all individuals comprising that particular group.

3.2 Semi-Structured Interviews

The interviews were conducted in Dutch and focused on the approach of the different interviewee groups to create process chronologies. In particular, we asked about the different types of *relationships* the interviewees made between and among information objects and events, in order to be able to create meaningful chronologies.

The questions consisted of the following five sections. We list them below, and describe what kind of questions correspond to each section:

- (1) *Professional Information*. Interviewees were asked about job title, typical workday activities, and years of experience.
- (2) *Activities and Process*. These questions pertained to what activities their occupation consisted of. We also asked them specifically whether their job entailed making chronologies for parliamentary inquiries or other fact-finding investigations. If that was the case, we probed them about these activities, and asked how they conducted them.
- (3) *Textual Analysis*. In this section of the interview, we asked participants to describe the different steps they go through to discover and analyse texts they need to read in order to be able to create chronologies or policy document reconstructions.
- (4) *Relationships between Texts*. Questions within this section pertained to the relationships professionals make between texts while conducting research.
- (5) *Comparison between Texts and Events*. By asking these questions, we aimed at gaining an understanding of how the interviewees interpreted the level of granularity at which events correspond to texts.
- (6) *Operations*. This section would delve into the groups' tactics to single out particular passages in texts – like summarizing, highlighting, or splitting.

Overall, these questions were aimed at comprehending how the analysts carry out their day-to-day activities. We were specifically interested in what tactics and techniques the interviewees employ to create process chronologies, and what kinds of relations bore the most relevance to their work.

The following relationships were discussed during the interview. A substantial portion of them were based on the work in [7]. Other relationships were also incorporated in the interviews. The ones that form additions to the work of the authors of [7] are marked with an asterisk (*). In this context, A and B refer to distinct events.

- Temporal (A took place before B).
- Similarity (A is similar to B).
- Entity (A is about the same entity – a person, organization, or place – as B)
- Citation (A refers to B by means of a hyperlink to another document)
- Correspondence* (person P mentions A in a message to person Q, whom subsequently refers to B in a returned or forwarded message)
- Subject (A falls under the same theme or topic as B)

- Causal (A leads to B)
- Social* (A and B are discussed among persons $P_1, \dots, P_n$)
- Speculative (A is connected to B)
- Domain knowledge (A is related to B through external domain knowledge X, which is not found or represented in the texts themselves)
- Categorical/Ontological* (A is filed under the same category or folder as B, for instance in a DMS or a file structure)

The interviewees were asked to (i) state which relationships they employed during their efforts to manufacture chronologies, and (ii) name the top three of most important relations (in no particular order). Sometimes, they would mention just two important relationships; on other occasions they named four of them.

4 CONSTRUCTING PROCESS CHRONOLOGIES

In this section, we present the results of exploratively coding the interviews with the four interviewee groups. Based on the interviews, we constructed a Process Delivery Diagram (PDD) that describes how professionals currently construct chronologies. A PDD combines a process model and a conceptual model to indicate which concepts are created in which activities [18]. Activities in a PDD can have unordered subactivities to express that there is no predefined sequence in which these need to be carried out. In Section 4.1, we describe how process chronologies are currently constructed, in Section 4.2, we discuss what relations are used, both implicitly and explicitly, to extract concepts – such as events and entities – from the collected information objects.

4.1 Situationalized Process Chronologies

During the interviews, we asked participants their current method of constructing chronologies. The answers of the participants have been coded to discover the different activities and sub activities. This resulted in the PDD shown in Fig. 2. Each activity is supported by interviewee groups, indicated in parentheses.

The first activity is *Object Collection* (I1), which starts either by developing a research question for an information inquiry or by an external inquiry (I3). For example, the parliament defines an inquiry, and then propagates the request to the different departments (I1, I4). Typically, TimeFlows are requested to derive possible interpretations, for example to clarify the subjective images entities (including actors) can have in specific matters (I4). Based on the research question, an initial collection of *Information Objects* is retrieved from a plethora of sources, including Document Management Systems (I1) and open sources (I3). An *Information Object* can range from the agenda and minutes of meetings (I2), memos and email traffic (both within and between organisations) (I1), but can also include public sources such as news papers or social media (I3). The *Information Object* includes metadata about the object itself, such as creation data, version information and author information. The main challenge in this activity is that many information objects are unstructured, or poorly archived (I4). For example, topics can be discussed in meetings, but to discover the records carrying this information can be painstaking (I1). Another problem is versioning, for example the detection and selection of near-duplicate information objects (I4).

The next activity is *Object Analysis*. In this activity, all retrieved *Information Objects* are reviewed on relevance. This review is mostly a manual task ranging from scanning the object and its metadata to a careful read and interpretation (I1, I4). An initial scan is used to determine the relevance of each object (I1). Relevant objects are then analysed to extract the most important *concepts*, such as *events*, *entities*, *subjects*, and *temporal expressions* (I1, I3).

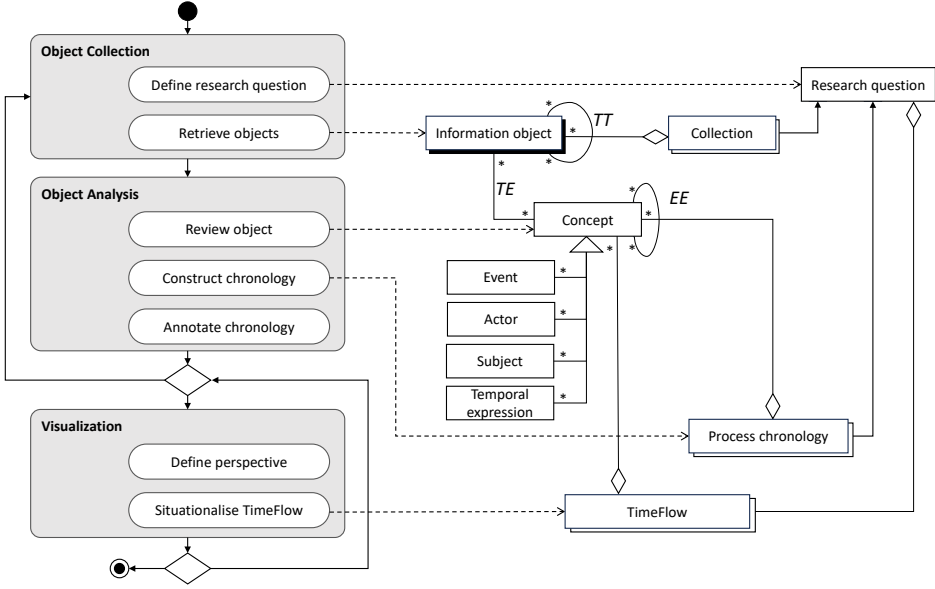


Fig. 2. Process Delivery Diagram of constructing TimeFlows based on the participants. The relations *TT*, *TE* and *EE* refer to the relations specified in Table 1. Together, these relations form a taxonomy.

The extracted concepts are used to construct a *chronology*: a large model containing all events extracted from the relevant objects. Each event in the chronology is annotated with the relevant concepts, containing everything that could be relevant to it (I1), ranging from a single sentence of a document to complete documents (I4). In rare cases, sub-chronologies are created to describe how specific documents were established (I2).

The *Object Analysis* activity is typically a manual task, and thus very laborious and time consuming (I1, I4). As one of the participants stated: “it is like looking for a needle in an exponentially growing haystack” (I4). Additionally, objects as policy documents require a thorough analysis to fully grasp the implications of the document (I3). As such, the object analysis is subjective to the researchers who develop the chronology (I2).

The activities are very iterative, switching between the activities *Object Collection* and *Object Analysis*. Based on the chronology, new objects are retrieved to fill in gaps, to clarify events, or to provide an interpretation (I1). For example, if in the agenda of a meeting specific documents are mentioned, these are retrieved and added to the collection for analysis (I2).

Currently, a *Chronology* is described in a single document that contains all events and the concepts related to it. Such chronology documents easily grow to several hundreds of pages (I4). As a consequence, it is infeasible to fully understand the events, and thus the process described in the document. Instead, a specific perspective is created to construct an overview, for example to brief someone (I1, I2), or to gain a better understanding of the topic at hand (I4). For this activity, each event is manually screened and added, based on a set of criteria that is often left implicit and subject-dependent. For example, if an event could not have been known by a stakeholder at the time, then the event is left out (I2). Different perspectives result in situationalized representations of the chronology. For example, a different representation is used for story telling (I2) than for understanding a certain topic (I4). In the former situation, the representation mainly describes what happened, including important milestones and involved entities, whereas the latter may

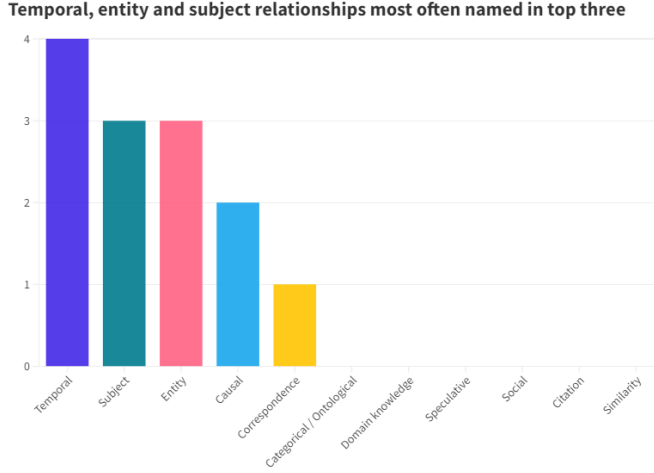


Fig. 3. Number of interviewee groups that mention the respective relationships between events as one of the top three most important ones

be a timeline to identify gaps, to refine specific concepts (I4). As one of the participants noted: a chronology is never static (I4).

4.2 Identified Relations Used in Document Analysis

As became apparent from the interviews, *Object Analysis* is the most laborious and time-consuming activity. All information objects need to be reviewed on relevance. Moreover, important concepts, such as events and entities, ought to be extracted. This is a manual task, as their interpretation is far from trivial. It requires in-depth knowledge, which is often tacit (I3, I4). As the participants indicated (I1, I3, I4), the concept extraction and chronology creation is a subjective task, where the researcher uses their tacit knowledge to decide the relevance of objects and the concepts it contains or discusses.

Equally important are the relations that explain the link between information objects and concepts. We distinguish three levels of relations:

- (1) *Between Information Objects*, i.e., relationships between different information objects (represented by level *TT* in Fig. 2).
- (2) *Between Information Objects and Concepts*, i.e., relationships that connect information objects to concepts (represented by level *TE* in Fig. 2).
- (3) *Between Concepts*, i.e., relationships between different concepts, such as events and entities (represented by level *EE* in Fig. 2).

As shown in Section 2, different types of relations can be identified, such as information objects that describe a certain event, or entities that participated in a meeting described by the information object. Therefore, we also asked the participants which relationships they use during the chronology construction. We extracted eight relations that are commonly used by the participants to obtain concepts from information objects and relate them.

Each of the participants created a top 3 of relation types. The results are shown in Fig. 3: *Temporal*, *Subject*, and *Entity* relationships were scored highest. The *Causal* relationship and *Correspondence* relations were mentioned less frequently. We choose to include all five of them within our framework.

Our coding revealed three additional relations that were described by the participants, but not mentioned explicitly. For example, all participants indicated they extracted the concepts from information objects. In essence, this means that each information object is linked to several concepts that are extracted from it. We refer to this relation as the *Consists Of* relation. This is the first of the three additional relations.

Two other relations were also identified while coding the interviews. In one of the interviews (I1), the participants explained that they considered email conversations, and added information based on the sender/receiver relation: “If we look at the most important message [of a chronology], we also look at policy documents, but certainly at email conversations.” His colleague added: “Yes, so we look at who has responded to whom, and who has responded to that.” Similarly, in another interview (I4), the participants indicated that they used the metadata of policy documents to derive the involved entities, and the dates on which these entities were involved. This analysis resulted in two additional relations: *Succession* and *Reference To*. The former indicates which objects directly follow each other, e.g. if an email is a direct reply to another email or if it is a forwarded message. The latter represents document referrals, such as documents discussed in meetings, or email messages with an attachment.

We categorize all eight relations within three groups: A, B, and C, according to the types of relationships they afford (see Table 1). In this upcoming section, we explain how these relations are incorporated within our conceptualization of process chronologies, and subsequently visualized within *TimeFlows*.

5 VISUALIZING TIMEFLOWS

In this section, we conceptualize process chronologies to construct TimeFlows from collections of information objects. As several participants in the interviews mentioned, visualising process chronologies has many different purposes, ranging from understanding the role of entities in the process (I2) to exploratory studies to identify gaps (I4). The field of Visual Analytics “combines automated analysis techniques with interactive visualizations for an effective understanding, reasoning and decision making on the basis of very large and complex data sets” [5]. Thus, visualizing TimeFlows can be considered as an application of Visual Analytics, as depicted in Fig. 4. From the collection of information objects, a process chronology is extracted. This chronology is then visualized as a TimeFlow to gain new knowledge on the topic at hand. In the remainder of this section, we present our proposal to visualize process chronologies as TimeFlows, and illustrate its capabilities by presenting a TimeFlow for the Dutch Childcare Benefits Scandal.

Table 1. Relations identified by the participants and their use over the different levels.

		TT	EE	TE
Group A	Temporal	x	x	x
	Subject	x	x	x
	Entity	x	x	x
	Causal	x	x	x
	Correspondence	x	x	x
Group B	Succession	x		
	References To	x		
Group C	Consists Of			x

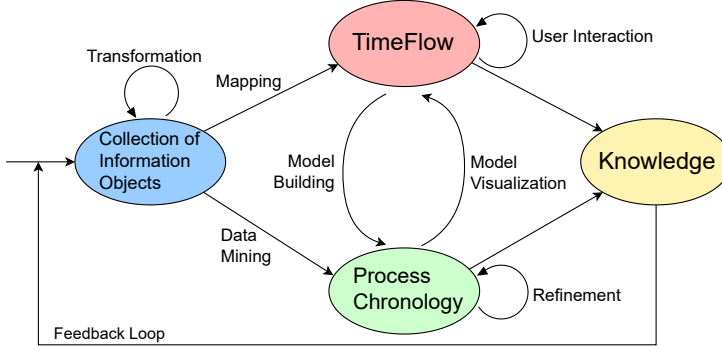


Fig. 4. The Visual Analytics feedback loop [5] applied to process chronologies.

5.1 TimeFlows to Visualize Process Chronologies

As presented in the previous section, a process chronology is a list of all annotated events extracted from a collection of information objects. Every event is annotated with the concepts it is related to, including entities and other events. Moreover, the events are enriched with the information objects that constitute them. A TimeFlow is a visual representation of a process chronology, to study a specific perspective. The following general design principles guide the visualization of TimeFlows:

- (1) We delineate *information objects* on the one hand, and *events* on the other. In their simplest form, events can be seen as “things that happen”. Events leave traces in the form of information objects that can be viewed as manifestations of past events [3]. Events are depicted using white rounded rectangles, while information objects are shown in light grey rectangles with sharp corners.
- (2) Relations are shown by means of arrows. The arrows stemming from the relations in group A are distinguished from one another through the use of both icons and colors. The relations in group B and C are all shown in black, but have different stroke styles (solid, dashed, or dotted).
- (3) The progression of time is indicated by means of arrows from left to right between events, originating from relations in group A. Choosing to depict the flow of time from left to right stems from old conventions ingrained within Western culture [10]. Moreover, events are numbered according to their temporal order, which also makes it easy to refer to them.
- (4) The relationships depicted by the arrows in Group A are explained through the use of highlighted words in the information objects and events. They have the same color as the arrows of the respective relations, and indicate which words are of relevance to the relation.

The relations described in point 2 above are chosen and motivated in Section 4.2. Below, we describe the design considerations for each individual relationship.

- **Temporal-Semantic Relationship.**  We relate events and information objects with one another that bear the same *temporal expressions*. These are words that indicate a particular moment in time [17]. For instance, if one object contains the phrase “On the first of September, we had a meeting,” and another object contains the fragment “Things went well during the meeting on 01/09, we’ll meet again soon,” then these objects are linked through the Temporal-Semantic Relationship by means of a light orange arrow with a clock icon. Other examples of temporal expressions are “tomorrow”, “next week”, “Halloween”, and “The Summer holidays”. (Levels: TT, EE, and TE.)

- 
Subject Relationship.  Events and information objects are linked with one another with a purple arrow and a book icon if the texts they comprise are related by subject. (Levels: TT, EE, and TE.)
- 
Entity Relationship.  An Entity Relationship is made when the same concept, e.g. a name of a person, an organization, or a place, is mentioned in both of the events or objects. Those objects that belong to a particular event and mention the same concept, are assigned an Entity relationship by means of an orange arrow and a stick figure icon. (Levels: TT, EE, and TE.)
- 
Causal Relationship.  This relationship is visualized with a turquoise arrow and an icon of Newton's Cradle on top. It depicts which events are causally related to another, and is also drawn between information objects and events when words that indicate a causal relationship are present. (Levels: TT, EE, and TE.)
- 
Correspondence Relationship.  This relationship is depicted with a dark blue arrow and an icon of a letter on top. It is drawn when the correspondent of one object is also recognized as the correspondent of another object. (Levels: TT, EE, and TE.)
- Succession Relationship.**  The *Succession Relationship* is a black, filled arrow with its head pointing towards to an information object that is a response or forward of the original text. (Level: TT)
- References To Relationship.**  The *References To Relationship* is depicted by means of a shortly-spaced dotted black arrow. It ends at the information object that is referenced, and begins at the object that does the referencing. It is employed when, for instance, an e-mail contains an attachment. (Level: TT)
- Consists Of Relationship.**  The *Consists Of Relationship* is shown by means of a widely-spaced dotted black arrow, originating from the event and pointing towards the different information object the event consists of. (Level: TE)

As intended by the Visual Analytics framework, it is the user that constructs TimeFlows themselves, based on the information objects they have collected and their own perspective on the information objects they wish to study.

5.2 An Illustrative Example

We illustrate the proposal of TimeFlows to represent process chronologies by depicting a sequence of events and accompanying documents pertaining to what is known in the Netherlands as the “Childcare Benefits Scandal”. This political affair has its origins in the year 2004 (when the Childcare Act came into effect), and concerned false allegations of fraud made by the Tax Authorities. It is estimated that between the years 2005 and 2019, this organization wrongfully accused more than 25.000 parents of making fraudulent childcare benefits claims ², driving families into severe financial hardship. After a parliamentary inquiry was conducted in 2021, the third cabinet Rutte resigned over the affair.

This TimeFlow visualization is grounded in true events and real texts. A key document was the “Memo Palmen” [11, 12]. The other texts we used are, in order of appearance: the Dutch Childcare

²See <https://www.trouw.nl/politiek/wie-wist-wat-in-de-toeslagenaffaire-de-kluwen-van-hoofdrospelers-ontward-b721c834/>

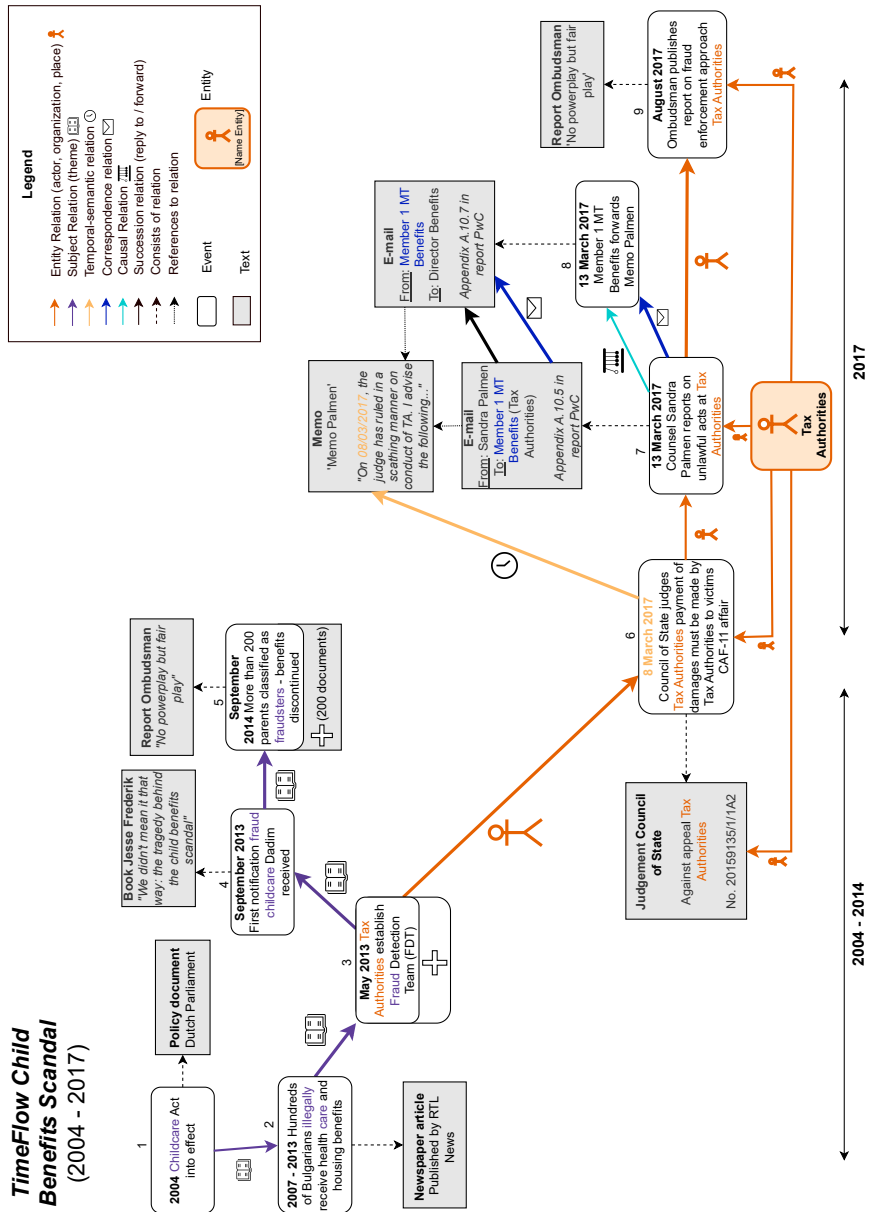


Fig. 5. A TimeFlow of the Childcare Benefits Scandal

Act of 2004 [2], a news article on the ‘Bulgarian Fraud’³, the book on the affair as a whole by Jesse Frederik [4], the report by the Ombudsman [19], and the judgement of the Council of State [13]. We make the following general observations of this TimeFlow:

- It consists of nine events. These are labelled by their temporal order.
- All events are distilled of one or more texts, shown with their respective Consists Of relations. In total, nine texts are explicitly shown. The existence of 200 additional texts is implied through the depiction of the + sign shown underneath Event 5.
- Note that Event 5 and Event 9 both consist of the report by the Ombudsman.

Moreover, additional information can be highlighted in such a TimeFlow. The current visualisation shows different examples of relations between events and information objects:

- A Temporal-Semantic relation is indicated between Event 6 and the Memo Palmen, as they both refer to the date 08-03-2017 (TE).
- Subject relations are drawn between Event 1 and Event 2, Event 2 and Event 3, Event 3 and Event 4, and between Event 4 and Event 5 (EE). For instance, the Subject relation between the first two events is justified because they contain the related words “Childcare” and “care”.
- Entity relations are depicted between Event 3 and Event 6, Event 6 and Event 7, and between Event 7 and Event 9 (EE). We also single out the Tax Authorities as an entity concept, and include relations to the judgement of the Council of State (TE) and to Events 6, 7, and 9 (EE).
- A Causal relation links events Events 7 and 8, because the act of forwarding the email - performed by Member 1 of the MT - is a direct causal consequence of Event 7, in which Sarah Palmen shares her findings (EE).
- Events 7 and 8 are also connected via the Correspondence relation, because the latter consists of an email that is forwarded from an email Event 7 consists of (EE).
- A Succession relation is shown between the Sarah Palmen’s email, and the forwarded email sent by Member 1 of the MT of the Tax Authorities (TT).
- Both Palmen’s and MT Member 1’s emails contain Palmen’s report as an attachment - thus both of these emails point to the report with the References To relation (TT).

As the events in this example show, multiple relation types are needed to understand and study what happened. As shown in Section 2, these relations are not shown in current visualisation techniques, such as the TGR in Figure 1. Moreover, the constitutive documents clarify on what basis the events have been generated – something that is lacking in previous conceptualizations of timelines with graph-based representations (TGRs).

6 FURTHER RESEARCH: CHALLENGES

In this paper, we argue that the visual analytics framework (see Figure 4) can be applied to study process chronologies using TimeFlows. However, to realize this, many challenges still need to be addressed. These challenges originate from the problems described by the interviewees.

Data Mining The future software ought to be able to automatically mine important concepts of heterogeneous document sets. Relevant challenges in this context include *Named Entity Recognition and Linking* (NER and NEL), *Temporal Expression Recognition, Normalization, and Linking* (TERNL), *Topic Modelling* (TM), and *Event Extraction* (EvEx).

Mapping The relations between the properties of the input data (information objects) and the visualizations (TimeFlows) are to be studied. A challenge lies in mapping more heterogeneous data

³See for example <https://www.rtlnieuws.nl/nieuws/nederland/artikel/1173981/uitspraak-bulgarenfraude-hoe-zat-het-ook-alweer>

sources to proper visualizations, including tabular data in Excel sheets, meeting minutes, and voice memos.

Model Visualization In order that the TimeFlows are generated in a visually intuitive manner, research must be carried out on peoples' preferred shapes and colours, the positioning of the events, information objections, and relations, and the amount of relations shown so as to balance clarity on the one hand, yet avoid information overload on the other.

User Interaction Here, the challenge lies in understanding how users can be supported in developing effective TimeFlows, e.g., the interaction with TimeFlows to enable users to obtain the information objects that bear relevance to each event, to move concepts and entities on the screen, to 'to focus on' or 'abstract away' specific events and investigate them at different levels of granularity.

Automated Model Building Future TimeFlows software could allow users to study different perspectives on sequences of events and their information objects. They may single out a perspective or combine different perspectives to study a research question from different angles. To support this activity, generative techniques need to be developed that propose multiple perspectives on the collection of information objects. These perspectives can help in the understanding of the process and the events manifested within the information objects.

7 CONCLUSION

In this paper, we introduce *process chronologies* as non-repetitive processes, which lack an event log. Based on interviews with experts, we have conceptualized different relations between the concepts and unstructured information objects that underlie them. To convey the events in a non-repetitive process, we propose *TimeFlows* to visualize the relation between information objects and the concepts manifested in these objects. TimeFlows allow users to study non-repetitive processes from different perspectives. Although our analysis was conducted on explorative interviews with domain experts, we believe that an holistic approach that visual analytics with TimeFlows offers great potential aid the understanding of such processes. However, to fully support the understanding and study of non-repetitive processes, many challenges need to be overcome. We therefore propose a research agenda to guide future work in this direction, which is aimed at automating the time-consuming manual work for process chronology creation. We view the present work as a broader ambition to develop interactive software that allows analysts to inspect large amounts of heterogeneous data sources to retrospectively study how non-repetitive processes unfold over time.

REFERENCES

- [1] Jeffery Ansah, Lin Liu, Wei Kang, Selasie Kwashie, Jixue Li, and Jiuyong Li. 2019. A graph is worth a thousand words: Telling event stories using timeline summarization graphs. In *The World Wide Web Conference*. 2565–2571.
- [2] Dutch Government. 2004. Wet Kinderopvang. *Wettenbank* (Oct. 2004).
- [3] Michael T Ferejohn. 2013. *Formal Causes: Definition, Explanation, and Primacy in Socratic and Aristotelian Thought*. OUP Oxford.
- [4] J. Frederik. 2021. *Zo hadden we het niet bedoeld: De tragedie achter de toeslagenaffaire*. De Correspondent Uitgevers. 1–400 pages.
- [5] D. Keim, G. Andrienko, J.-D. Fekete, C. Görg, J. Kohlhammer, and G. Melancon. 2008. Visual Analytics: Definition, Process, and Challenges. *Information Visualization LNCS*, 4950 (Sept. 2008), 154–175.
- [6] Brian Felipe Keith Norambuena and Tanushree Mitra. 2021. Narrative Maps: An Algorithmic Approach to Represent and Extract Information Narratives. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW3, Article 228 (jan 2021), 33 pages.
- [7] B. F. Keith Norambuena, T. Mitra, and C. North. 2022. Design guidelines for narrative maps in sensemaking tasks. *Information Visualization* 21, 3 (March 2022), 220–245.
- [8] B. F. Keith Norambuena, T. Mitra, and C. North. 2023. A Survey on Event-based News Narrative Extraction. *Comput. Surveys* 55, 14s (July 2023), 1–39.

- [9] B. Liu, F. X. Han, D. Niu, L. Kong, K. Lai, and Y. Xu. 2020. Story Forest: Extracting Events and Telling Stories from Breaking News. *ACM Transactions on Knowledge Discovery from Data* 14, 3 (May 2020), 1–28.
- [10] R. Núñez, K. Cooperrider, D. Doan, and J. Wassmann. 2012. Contours of time: Topographic construals of past, present, and future in the Yupno valley of Papua New Guinea. *Cognition* 124, 1 (July 2012), 25–35.
- [11] PricewaterhouseCoopers. 2021. Appendices bij PwC rapport Reconstructie en tijdlijn van het ‘memo-Palmen’. *Report PwC* (Sept. 2021), 1–301.
- [12] PricewaterhouseCoopers. 2021. Reconstructie en tijdlijn van het ‘memo-Palmen’. *Report PwC* (Sept. 2021), 1–66.
- [13] Raad van State (Council of State). 2017. Uitspraak 201509135/1/A2. <https://www.raadvanstate.nl/> (2017), 1–8.
- [14] Dafna Shahaf and Carlos Guestrin. 2011. Connecting the Dots between News Articles. In *IJCAI 2011*. IJCAI/AAAI, 2734–2739.
- [15] Dafna Shahaf, Carlos Guestrin, and Eric Horvitz. 2012. Trains of thought: generating information maps. In *WWW 2012*. ACM, 899–908.
- [16] Julius Steen and Katja Markert. 2019. Abstractive Timeline Summarization. In *Proceedings of the 2nd Workshop on New Frontiers in Summarization*. Association for Computational Linguistics, Hong Kong, China, 21–31.
- [17] N. UzZaman and J. F. Allen. 2010. Extracting Events and Temporal Expressions from Text. *2010 IEEE ICSC* (Sept. 2010), 1–8.
- [18] Inge van de Weerd, Sjaak Brinkkemper, Jurriaan Souer, and Johan Versendaal. 2006. A situational implementation method for web-based content management system-applications: method engineering and validation in practice. *Softw. Process. Improv. Pract.* 11, 5 (2006), 521–538.
- [19] W. van den Berg, M. Alhadjri, and M. Mulder. 2017. Geen Powerplay, maar Fair Play. *De Nationale Ombudsman* (Aug. 2017).
- [20] Shize Xu, Shanshan Wang, and Yan Zhang. 2013. Summarizing Complex Events: a Cross-Modal Solution of Storylines Extraction and Reconstruction. In *EMNLP*. Association for Computational Linguistics, Seattle, Washington, USA, 1281–1291.
- [21] Yi Yu, Adam Jatowt, Antoine Doucet, Kazunari Sugiyama, and Masatoshi Yoshikawa. 2021. Multi-TimeLine Summarization (MTLS): Improving Timeline Summarization by Generating Multiple Summaries. In *IJCNLP*. Association for Computational Linguistics, Online, 377–387.