

# Design of a tool for the visualization of historical data

Diseño de una herramienta para la visualización de datos históricos

► Ricardo Cedeño Montaña MSc  
Humboldt-Universität zu Berlin, Germany  
Institut für Kulturwissenschaft  
cemontar@hu-berlin.de

---

## Abstract

In the sciences the combination of different types of analysis, textual and visual, contributes to the legibility of the results of the research. Techniques such as data visualization and information graphics have a long history in the natural and social sciences. Recently, researchers in the humanities have got interested in producing visualizations to synthesize facts and trends. The purpose is to create novel observations and analysis of historical and cultural data. This paper presents the use of programming and markup languages to design a custom and open tool for the creation of one particular type of historical data visualization: chronologies.

**Keywords:** data visualization, chronology, XML, processing

---

## Introduction

Data visualizations are not just graphics that explain a text in a condensate way, rather they are non-verbal forms of argumentation with their own signs, symbols, and syntactic operations. In recent years, research in the humanities have started to produce diagrams and visualizations not just as rhetoric objects but as the site for the production of new analysis of historical and cultural data (Burdick, Drucker, Lunenfeld, Presner, & Schnapp, 2012).

Proprietary and open source programs for diagrams and data visualization, such as Many Eyes, Tableau, or Gephi, offer commonplace results often arriving to similar outputs regardless of the problem or the data. Traditional graphic editors only allow to work with static data. If the data changes, it is hard to modify or update the visualization because they are not programming packages. While working with proprietary software, researchers and designers using such programs must accept on faith the program's results and its prescriptive outputs because there is no way to intervene from outside the way how the program process and render the data. The code is locked behind an interface and the researcher should focus exclusively on the data set.

This paper will present the use of programming and markup languages to design a custom and open tool for the creation of one particular type of historical data visualization: chronologies. This tool offers one prescriptive output too. But as the source codes are free for study and modification, and they are thoroughly documented, anyone interested in can adapt both the data structure and the program itself to reflect the subtleties of their own visualization problems.

This tool has two parts. The first focus on the gathering and registering of historical data and works independently. The second focus on the processing and arrangement of the historical records in order to output a synthetic visual product. The aim is the creation of compelling visuals with the potential to reach both audiences: specialists within the humanities and people beyond the academy. One requirement in the design of this tool was the use of programming and markup languages under GNU GPL; another was to be cross-platform.

The paper will be divided in two parts. Part one will describe the application of the tool to one particular topic of technical media history: formats for the storage of moving images. This problem of visualization addresses when and how the physical features of storage media for moving images have changed. Part two will comprise the set of tools develop to produce the visual outputs. I will explain the data structure used for recording the historical data and the structure of the program to produce the chronologies. In this part, I will also discuss details about the code and its functioning.

## 1. Chronologies

A chronology is a traditional tool of historians to concisely order along a time line a series of events of a subject. It can be a written register of events in strict temporal order from the oldest to the newest. It can be too a series of images, texts, graphics localized atop of a time-line. In either way, a chronology is modeled after the manner in which we, humans, experienced the pass of time.

Within the research project on the cultural and technical history of portable media for amateur movie and video production carried out at Institute of Cultural History and Theory at the Humboldt-Universität zu Berlin, the chronological format was chosen to illustrate the physical changes of storage media formats used in portable equipment for moving image storage.

These chronologies of historical data mix three familiar formats: the time line, the bar chart, and the zoom of a detail. Along the time line, each chronology of portable media uses a bar-chart to synthesize in a graphic shape some details of the media formats. This sequential and graphic ordering of the formats grants their analysis by contrast because many characteristics of several of them are simultaneously visible.

In a close view, each chronology portraits detail information about each storage media format. Information such as year of invention, format gauge and support material are presented for each format. At a more distant view, the overall tendency towards the reduction of the material supports for storage media is observable. Thus, this chronological format to visually present the technical history of these media formats allows what Edward R. Tufte calls: micro and macro readings (Tufte, 1990, p. 37).

Three chronologies conform the core of the analysis of the technical history of the portable media. Each corresponds to one of the media storage supports, film, video, and digital video. These visualizations are based on historical and technical data about



historical conditions for portable media. Such condition expresses in the reduction of the physical size of film formats as well as on the compression algorithms of digital encoding video.

Each of the chronologies orders several media formats from the earliest working film format to the most recent digital formats. Each format is localized on the time line and a set of its characteristics is mapped out. On the left side of each chronology, a bar shows the width in millimeters of each media format and the color of the bar represents either its support material or its recording method. On the right side of each chronology, and extracted from the time line, there appears a small bunch of media formats to look closely at their characteristics and to compare then against each other, see illustration 1. In the analogue video and digital video chronologies, this procedure serves for an additional specific comparison of the storage density of the tape surface and of the data rate transmission, respectively, see illustrations 2 and 3.

In this particular application these chronologies act like the exploded view in engineering drawings. This is no metaphor nor analogy. It is rather a characterization. The history of portable media is exploded along the common time-axis where its components are presented following their sequence in time. Certainly, one might argue that not all the components of the history are presented as in an engineering drawing. But here the explosion works as an analytic not as a projective tool. Therefore, a set of the components is isolated and picked for closer examination. In this case the several media formats located along a vertical time-axis compose the set. The procedure is taken one step further and for each format a set of its features is exploded too along its own axis. In the case of the film formats, the gauge, support material, and format name are shown on the horizontal-axis. As the analysis deepens, a zoom into the details of a set of formats is necessary. Again in the case of film, nine formats are picked from the time-axis for comparison. Their arrange is typological as their gauges and shapes are contrasted to highlight the reduction of their physical characteristics.

Each chronology contains a large number of formats. The list is as comprehensible as possible. The bulk of the data comes from previous media historical and media archaeological accounts such as (Abramson & Sterling, 2003; Kattelle, 2000; Watkinson, 1994; Zielinski, 2010) and the technical details about particular formats comes primarily from the standard reports of organizations such as SMPTE, ISO, and ITU. The specific information of each format was recorded in a structured data file and a custom made program processed the data and produced each chronology.

## 2. Data structure and programming

The process to produce the chronologies consists of two phases. In the first, the historical and technical data is recorded and structured in an XML file and in the second phase all records are processed with a Processing sketch that outputs the selected data along a time line either for the screen or the paper. The separation between the two phases allows to update the data and change the organization of the records regardless of their visual output. On the other hand, it allows to process the same set of records in different ways without altering the actual data. Both, the XML data structure and the

Processing sketch, were designed as flexible as possible to provide a free tool that can be used for other topics.

What follows describes a general purpose tool for chronologies.

### XML Data Structure

XML is a markup language. As HTML, it consists of tags but differently from it, they have to be customized. It was designed to describe data and XML files are simple text files that can be opened and edited with a simple text editor.

As part of the tool for chronologies, the XML file provides a static data set of records. Each chronology must have at least one XML file containing all records. Illustration 4 shows the XSD schema components and structure the above XML file conforms to.

A record using that XML structure in the XML file looks:

```
<items>
  <item>
    <name>Item 00</name>
    <date>
      <year>1870</year>
    </date>
    <info>Some information</info>
  </item>
</items>
```

Where items represents any collection or subject in plural and item represents the repetitive and singular elements that make up said collection or subject. An example of application will be: film formats as the collection and format as its elements, legal policies and policy will be another example. Each item must contain at least two records: a name and a year. The info record is optional, customizable, could store any type of data, and could be repeated as many times as necessary.

### Processing Sketch

The processing sketch takes as input the records stored in an XML file conforming to the aforementioned structure and outputs a

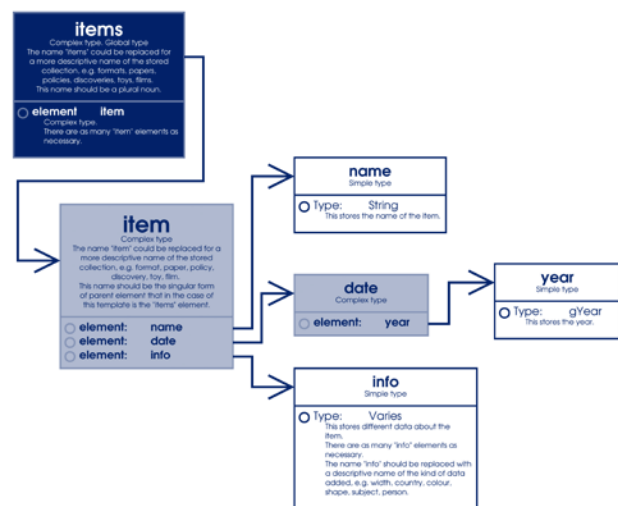


Illustration 4: XSD schema

PDF document with a centered time line, the records located along it, and a number of labels for identification. The sketch has two parts. In the first part, the user should input the value for a number of the variables including: the size of the paper, the span of the time period covered, the dimensions for the time line, a typeface for texts, and the location of the XML file. To label the chronology, she can input a title, a subtitle, an author, and a year. The second part consists of six sequential procedures that select and arrange each record, and render them on the time line. In table 1, the unnumbered part corresponds to the first part of the sketch, i.e., the inputs and the output; and the numbered part corresponds to the second part of the sketch and its six procedures: yearsWithItem, yearPositions, getXMLRecords, locateDataInTimeline, timeline, and labels. Illustration 5 shows in the upper side the input modules a user needs to fill in for the tool to work and below the six procedures it uses to produce a chronology. The outputs and inputs for each module and each procedure are depicted by the arrows.

Table 1: PDL for the timeline sketch file

| Procedure 1 Timeline sketch. Structure |  |
|--|--|
| Require                                | pageWidth := page dimension in mm            |
| Require                                | pageHeight := page dimension in mm           |
| Require                                | begin := initial year                        |
| Require                                | end := final year                            |
| Require                                | timespan := end-begin                        |
| Require                                | timeLineWidth := width of the time line      |
| Require                                | leading := space between each year           |
| Require                                | typeface := font for texts in the timeline   |
| Require                                | XML Data := location of the XML file         |
| Ensure                                 | PDF document                                 |
| 1                                      | function setup( )                            |
| 2                                      | yearsWithItem(XML Data)                      |
| 3                                      | return array yearsWithItem                   |
| 4                                      | yearPositions(timespan)                      |
| 5                                      | return dictionary yearPositions              |
| 6                                      | getXMLRecords(XML Data)                      |
| 7                                      | locateDataInTimeline(name, year, posY, info) |
| 8                                      | timeline(begin, end, yearsWithItem)          |
| 9                                      | return timeline                              |
| 10                                     | labels( )                                    |
| 11                                     | return labels                                |
| 12                                     | end function                                 |

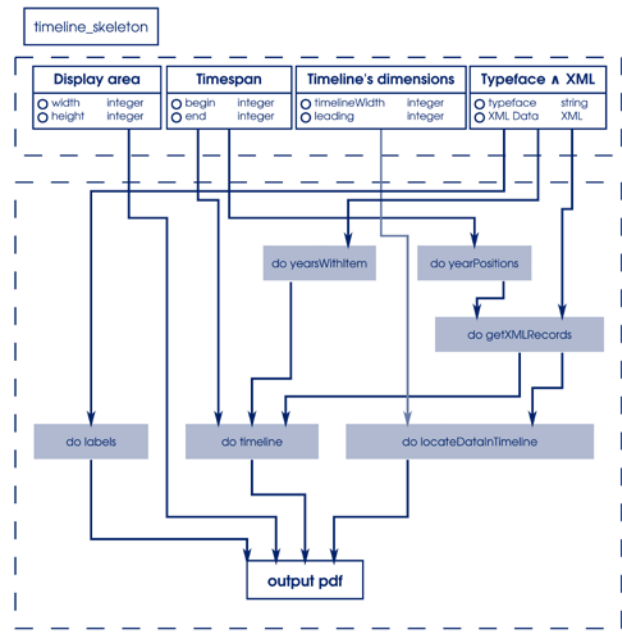


Illustration 5: PDL flow diagram of timeline processing sketch

### Acknowledgments

I'm grateful to the DAAD for providing me the funding for this research. This tool was developed at the Institute of Cultural History and Theory at the Humboldt-Universität zu Berlin and its parts are available under a GNU GPL v3 license at <http://portablemedia.hu-berlin.de>

### References

Abramson, A., & Sterling, C. H. (2003). The History of Television, 1942 to 2000. Jefferson, North Carolina: McFarland.

Burdick, A., Drucker, J., Lunenfeld, P., Presner, T., & Schnapp, J. (2012). Digital Humanities. Cambridge, MA: MIT Press.

Kattelle, A. (2000). Home movies : a history of the American industry, 1897-1979. Nashua, NH: Transition Pub.

Tufte, E. R. (1990). Envisioning information. Cheshire, Connecticut: Graphics Press. Retrieved from <http://books.google.de/books?id=r21HAAAAMAAJ>

Watkinson, J. (1994). The Digital Video Tape Recorder. Oxford, UK: Focal Press.

Zielinski, S. (2010). Zur Geschichte des Videorecorders (10th ed.). Potsdam, Germany: Polzer.