

Challenges for time as digital data

In digital humanities, there always is a gap between digitally stored data and the collected data. The gap is due to the (structural) changes that data needs to undergo in order to be stored within a digital structure. This gap may be small, in the case of natively digital data such as a message on Twitter: a tweet can be stored close to its 'original' format, but it still loses a lot of its frame of reference (the potential audience at a specific point in time, the actual audience, the potential triggers of the message etc.). In digital humanities this distance may become so large that some researchers argue, the term *data* should be largely abandoned and replaced with *capta*. *Capta*, the taken, in contrast to *data*, the given, should emphasise the interpretative and observer dependent nature of data in the humanities [4]. This problem relates to all kinds of data, whether categorical, quantitative, spatial or temporal. I will however focus only on the last type. Time and temporal expressions are particularly susceptible to multiple modes of interpretation and (unintended) modifications due to limitations in digital data structures, but also due to the ambiguous and subjective nature of time itself.

What is time

Within the context of time in computing, two competing philosophies of time are generally discussed [7, 11]. Bergson's notion of time as duration [2] and Bachelard's description of time as a sequence of discontinuous instants [1]. It is however more a difference in focus than in conviction, which lead the two thinkers to emphasise different aspects. Bergson discusses time in relation to consciousness. He distances experienced (*concrete duration*) from mathematical time (*abstract time*): duration extends continuously from past to present, is therefore unique and can not be abstracted. Duration is "a qualitative multiplicity, with no likeness to number" [2]. This model of time is unusable for Bachelard. For a philosopher of science, time needs to allow itself to be quantified; only what can be expressed in numbers counts as scientific [see 8]. Bachelard argues that duration is only perceived through the awareness of individual instants separated by nothingness. A useful side effect of this view is that instants potentially can be counted and time be expressed in numbers.

Other aspects of time

Quantified time is one of mankind's greatest innovations [see 3] and dominates most of our immediate daily life. It is how meetings are scheduled, trains are run, leisure time is coordinated and money is made. But quantified time can be in heavy contrast to *experienced time*. An hour at work can feel longer (or shorter) than an hour elsewhere [6]. And when we remember back in time, it is usually not that we remember a certain number of days or months that passed, but a certain association (a context, an activity, a person, a place; for an application of this in computing see [16]).

Quantified time and experienced time are the two aspects I presented that Bachelard and Bergson highlighted. For a more complete (but still very limited) picture of what time is, we also need to include time as a *context*. And in contrast to Bachelard, who explicitly distances his theory of time from Bergson's account, those three aspects may in fact flow into each other. A date (quantified time), may trigger an association (contextual time) and may feel very long ago, or not long at all (experienced time).

Time in computing

In computing as a theoretical model aspects of time are not really a concern. For the conceptual model of a computer, in the form of a Turing machine, there exist only two forms of time: immediacy and infinity. Either, a programme will halt at a certain state immediately, or it will keep running forever. Duration comes into play when a computer is materialised through hardware and a programme instruction ultimately is executed by an electronic circuit. Computations performed in the physical world take time and engineers, at least in the early days of computing, needed to take care that computations could actually be performed within one running cycle of the computer. Those cycles are, still today, coordinated and synchronised by a clock. This clock is more like a metronome rather than a time piece. Clocks in the form of real-time clocks have been introduced in mainframe computing during the 1960ies, along with the data formats for storing temporal information. The main reason for computers to keep track of real world time was rather mundane: users were typically charged per second of CPU time and time information was recorded for accounting purposes. Time once again needed to be quantifiable and it is this numerical model of time that is still at the heart of time as digital data.

Time as a digital space

The most prevalent model for time as digital data is Unix time [18], which counts the seconds elapsed according to Universal Time from a seemingly arbitrary starting point (1 January 1970, the Unix epoch). Earlier dates are represented as negative numbers and the amount of time this format encompasses depends on how many bits are allocated for its encoding. Some fear another 'millennium bug' on 19 January 2038, when the number of seconds since the Unix epoch exceeds the number which can be stored in 32 bits. Foresightful programmers therefore use 64 bits, which will last us until the year 292471210647.

Evidently, the digital model of time is that of time as a quantifiable entity, a coordinate system for locating events in time. Digital time is a one dimensional space, both mathematically and conceptually, and the issues around representing physical space digitally apply almost equally to time as well.

A time or place, if expressed in numbers, always needs to be defined *precisely*, even if it might not be known with great accuracy. Humanities researchers using GIS (Geographic Information Systems) employ a workaround by limiting the number of digits behind the floating point when a coordinate is intended to be imprecise [14]. In time, there is a similar tendency of storing, for example, only the year number when the precise date is unknown. However, the standard data structure for time always needs to store a precise date after all. Such dates therefore get encoded as 1 January at 00:00 and, reading only from the date, one can only guess the intended precision of the date.

When working with geographic space one needs to be aware of borders which might have changed over time, historic countries or multiple names for places referring to the same entity (e.g. United States, USA, the US, etc.). Also whether or not a country is, in fact, recognised as a country depends heavily on one's political view. Such historic and political perspectives are likewise embedded in digital time, even if they might not be evident at first. Having a western origin, digital time follows the Gregorian calendar and Universal Time. This is however just one of 30 calendars (according to Reingold and Dershowitz [15]) of which several are in use today and many are historically significant. In the same way as changes in political geography, historic changes in chronology need to be taken into account when working with dates. A date in 1927 caused a developer great headache, until a commentator on the online platform StackOverflow realised that the developer was located in Shanghai, where in 1927 the clocks were turned back five minutes in the night of the 31st December [10].

Digital time is 'exact' quantified time. As such it is used and indispensable as a structuring framework, for coordinating future plans equally as for understanding historic events. It however puts an emphasis on precision and time as an unambiguous location. But not only in the humanities, time is often more than that.

A richer digital time

The question for digital humanities therefore is, how are the numerous aspects of time represented in digital data? How to capture temporal information which can't be expressed as a quantity, and how to prevent losing data in the process of converting time to a quantitative format?

In existing datasets, temporal information is all too often stored according to the digital data structure: as a precise value or range of values with defined start and end points. This forces the author of the dataset to put an emphasis on a time span or point in time, regardless of his (subjective) certainty about an event. A few exceptional solutions allow to define a level of confidence with regard to dates (Neatline [12]) or specifying upper and lower bounds of certainty (CIDOC-CRM[9], SIMILE[17]). Still, they do so by translating those levels of confidence essentially to a numerical value, which does not carry any further meaning.

Rather than having a recorded level of confidence, one needs to know the circumstances that informed the dating. It is important to capture the context, the reasoning behind dates. And it is essential to do so in a machine readable format. Explanations in natural language contained in comment fields can not be used for further reasoning or visualisation without elaborate and error-prone reprocessing.

An example is linking a work of art to the biography of an artist, a relation which is relatively simple to establish even in existing datasets. More far-ranging connections allow for outlining the temporal situation and relevance further, such as linking to biographies of related persons (commissioners, collectors, pictured, etc.) or events (exhibitions, transactions, etc.).

Relational databases allow – and linked data formats even enforce – making connections and establishing contexts in this way. However, it is not a question of the technical abilities of a system, but of the mindset and conviction of its user.

Even capturing the subjective experiential aspects of time in digital systems is not necessarily impossible. A step in this direction comes from researchers at the University of Virginia [5, 13].

Time is not only a specification, but a *place* which can be explored. The tools are there for digital time to be more than just a value. What we need is a change in mindset in how we think about time as digital data and, subsequently, the establishment of a consensual methodology for recording and representing multiple aspects of time in digital humanities.

References

- [1] Bachelard, G. 1963. *La dialectique de la durée*. Les Presses universitaires de France.
- [2] Bergson, H. 1950. *Time and Free will: An Essay on the Immediate Data of Consciousness*. George Allen & Unwin Ltd.
- [3] Crosby, A.W. 1997. *The Measure of Reality: Quantification in Western Europe, 1250-1600*. Cambridge University Press.
- [4] Drucker, J. 2011. Humanities approaches to graphical display. *Digital Humanities Quarterly*. 5, 1 (2011).

- [5] Drucker, J. and Nowviskie, B.P. 2003. Temporal Modelling: Conceptualization and Visualization of Temporal Relations for Humanities Scholarship. *ACH/ALLC 2003 Conference*. (2003), 26–28.
- [6] Hammond, C. 2012. *Time Warped: Unlocking the Mysteries of Time Perception*. Canongate Books.
- [7] Harper, R., Whitworth, E. and Page, R. 2012. Fixity: Identity, Time and Durée on Facebook. *Association of Internet Researchers*. (Sep. 2012).
- [8] Hashizume, K. 2009. Bachelard's theory of time: Missing link between science and art. *Aesthetics*. 13, (2009).
- [9] How to implement CRM Time in RDF: 2011. http://www.cidoc-crm.org/docs/How_to_%20implement%20CRM_Time_in%20RDF.pdf. Accessed: 2013-11-06.
- [10] java - Why is subtracting these two times (in 1927) giving a strange result? - Stack Overflow: 2011. <http://stackoverflow.com/questions/6841333/why-is-subtracting-these-two-times-in-1927-giving-a-strange-result/6841479#6841479>. Accessed: 2013-11-27.
- [11] Lindley, S., Corish, R., Ferreira, P., Simbelis, V. and Vaara, E. 2012. Changing Perspectives of Time in HCI. *Workshop CHI 2013*. (2012).
- [12] Neatline.org | About: 2012. <http://neatline.org/about/>. Accessed: 2013-03-05.
- [13] Nowviskie, B.P. and Adviser-Mcgann, J.J. 2004. *Speculative computing: instruments for interpretive scholarship*. University of Virginia Charlottesville, VA, USA.
- [14] Re: Displaying imprecise date and location information - best practice and examples?: 2013. <https://www.jiscmail.ac.uk/cgi-bin/webadmin?A2=ind1311&L=mcg&D=0&P=5304>. Accessed: 2013-11-27.
- [15] Reingold, E.M. and Dershowitz, N. 1989. *Calendrical Calculations*. Department of Computer Science, University of Illinois at Urbana-Champaign.
- [16] Ringel, M., Cutrell, E., Dumais, S. and Horvitz, E. 2003. Milestones in time: The value of landmarks in retrieving information from personal stores. *Human-Computer Interaction, Interact '03*. G.W.M. Rauterberg, M. Menozzi, and J. Wesson, eds. 184–191.
- [17] Timeline Event Display - SIMILE Widgets: 2010. http://simile-widgets.org/wiki/Timeline_Event_Display. Accessed: 2013-05-09.
- [18] Unix time - Wikipedia, the free encyclopedia: http://en.wikipedia.org/wiki/Unix_time. Accessed: 2013-11-27.