# Towards a Unified Theory of Time-Varying Data*

Benjamin Merlin Bumpus[1], James Fairbanks[1], Martti Karvonen[2] Wilmer Leal[1], Frédéric Simard[3]

[1]University of Florida, Computer & Information Science & Engineering, Florida, USA.
[2]University of Ottawa, Department of Mathematics, Canada.
[3]University of Ottawa, School of Electrical Engineering and Computer Science, Canada.

In this talk we will present our answer to the question: '*how does one build a robust and general theory of temporal data?*'. Our approach is twofold; first we distill the lessons learned from temporal graph theory into the following set of *desiderata* for any mature theory of temporal data: **(D1)** (**Categories of Temporal Data**) Any theory of temporal data should define not only time-varying data, but also appropriate morphisms thereof. **(D2)** (**Cumulative and Persistent Perspectives**) In contrast to being a mere sequence, temporal data should explicitly record whether it is to be viewed cumulatively or persistently. Furthermore there should be methods of conversion between these two viewpoints. **(D3)** (**Systematic 'Temporalization'**) Any theory of temporal data should come equipped with systematic ways of obtaining temporal analogues of notions relating to static data. **(D4)** (**Object Agnosticism**) Theories of temporal data should be object agnostic and applicable to any kinds of data originating from given underlying dynamics. **(D5)** (**Sampling**) Since temporal data naturally arises from some underlying dynamical system, any theory of temporal data should be seamlessly interoperable with theories of dynamical systems. Our second main contribution is to introduce *categories of narratives*, an object-agnostic theory of time-varying objects which satisfies the desiderata mentioned above:

**(D1)** There has been no formal treatment of the notion of morphisms of temporal graphs and this is true regardless of which definition of temporal graphs one considers and which specific assumptions one makes on their internal structure. This is a serious impediment to the generalization of the ideas of temporal graphs to other time-varying structures since any such general theory should be invariant under isomorphisms. Thus we distill our first desideratum **(D1)**: theories of temporal data should not only concern themselves with what time-varying data is, but also with what an appropriate notion of *morphism* of temporal data should be.

Narratives, our definition of time-varying data, are stated in terms of certain kinds of sheaves. This immediately addresses desideratum **(D1)** since it automatically equips us with a suitable and well-studied notion of a morphism of temporal data, namely *morphisms of sheaves*. Then, by instantiating narratives on graphs, we define *categories* of temporal graphs as a special case of the broader theory.

**(D2)** Our second desideratum is born from observing that all current definitions of temporal graphs are equivalent to mere sequences of graphs (snapshots) without explicit mention of how each snapshot is related to the next. To understand the importance of this observation, we must first note that in any theory of temporal graphs, one always finds great use in relating time-varying structure to its older and more thoroughly studied *static* counterpart. For instance any temporal graph is more or less explicitly assumed to come equipped with an *underlying static graph*. This is a graph consisting of all those vertices and edges that were ever seen to appear over the course of time and it should be thought of as the result of *accumulating* data into a static representation. Rather than being presented as part and parcel of the temporal structure, the underlying static graphs are presented as the result of carrying out a computation – that of taking unions of snapshots – involving input temporal graphs. The implicitness of this representation has two drawbacks. The first is that it does not allow for vertices or edges to *merge* or *divide* over time; these are very natural operations that one should expect of time-varying graphs in the 'wild' (think for example of cell division or acquisitions or merges of companies). The second drawback of the implicitness of the computation of the underlying static graph is that it conceals another very natural static structure that always accompanies any given temporal graph, we call it the *persistence graph*. This is the static graph consisting of all those vertices and edges which persisted throughout the entire life-span of the temporal graph. We distill this general pattern into desideratum **(D2)**: temporal data should come

*Full-paper version: *Towards a Unified Theory of Time-varying Data*. Benjamin Merlin Bumpus, James Fairbanks, Martti Karvonen, Wilmer Leal, Frédéric Simard. arXiv:2402.00206, 2024.

explicitly equipped with either a *cumulative* or a *persistent* perspective which records which information we should be keeping track of over intervals of time.

Thanks to categorical duality, our narratives satisfy desideratum **(D2)** in the most natural way possible: sheaves encode the persistence model while co-sheaves (the dual of a sheaf) encode the accumulation model. As we will show, while these two perspectives give rise to equivalences on certain subcategories of temporal graphs, in general, when one passes to arbitrary categories of temporal objects – such as temporal groups, for example – this equivalence weakens to an *adjunction*. In particular our results imply that in general there is the potential for a loss of information when one passes from one perspective (the persistent one, say) to another (the cumulative one) and back again. This observation, which has so far been ignored, is of great practical relevance since it means that one must take a great deal of care when collecting temporal data since the choices of mathematical representations may not be interchangeable.

**(D3)** Another common theme arising in temporal graph theory is the relationship between properties of static graphs and their temporal analogues. At first glance, one might naïvely think that static properties can be canonically lifted to the temporal setting by simply defining them in terms of underlying static graphs. However, this approach completely forgets the temporal structure and is thus of no use in generalizing notions such as for example connectivity or distance where temporal information is crucial to the intended application. Moreover, the lack of a systematic procedure for '*temporalizing*' notions from static graph theory is more than an aesthetic obstacle. It fuels the proliferation of myriads of subtly different temporal analogues of static properties. For instance should a temporal coloring be a coloring of the underlying static graph? What about the underlying persistence graph? Or should it instead be a sequence of colorings? And should the colorings in this sequence be somehow related? Rather than accepting this proliferation as a mere consequence of the greater expressiveness of temporal data, we sublime these issues into desideratum **(D3)**: any theory of temporal data should come equipped with a systematic way of 'temporalizing' notions from traditional, static mathematics.

We show how our theories of narratives satisfies desideratum **(D3)**. We do so systematically by leveraging two simple, but effective *functors*: the change of *temporal resolution* functor and the *change of base* functor. The first allows us to modify narratives by rescaling time, while the second allows us to change the kind of data involved in the narrative (e.g. passing from temporal simplicial complexes to temporal graphs). Using these tools, we provide a general way for temporalizing static notions.

**(D4)** Temporal graphs are definitely ubiquitous forms of temporal data, but they are by far not the only kind of temporal data one could attach, or sample from an underlying dynamical system. Thus Desideratum **(D4)** is evident: to further our understanding of data which changes with time, we cannot develop case by case theories of temporal graphs, temporal simplicial complexes, temporal groups etc., but instead we require a general theory of temporal data that encompasses all of these examples as specific instances and which allows us to relate different kinds of temporal data to each other.

Our theory of narratives addresses part of Desideratum **(D4)** almost out of the box: our category theoretic formalism is object agnostic and can be thus applied to mathematical objects coming from any such category thereof. We observe through elementary constructions that there are *change of base* functors which allow one to convert temporal data of one kind into temporal data of another. Furthermore, we observe that, when combined with the adjunction, these simple data conversions can rapidly lead to complex relationships between various kinds of temporal data.

**(D5)** As we mentioned earlier, our philosophical contention is that on its own data is not temporal; it is through originating from an underlying dynamical system that its temporal nature is distilled. This link can and should be made explicit. But until now the development of such a general theory is impeded by a great mathematical and linguistic divide between the communities which study dynamics axiomatically (e.g. the study of differential equations, automata etc.) and those who study data (e.g. the study of time series, temporal graphs etc.). Thus we distill our last Desideratum **(D5)**: any theory of temporal data should be seamlessly interoperable with theories of dynamical systems from which the data can arise.

This desideratum is ambitious enough to fuel a research program and it thus beyond the scope of a single paper. However, for any such theory to be developed, one first needs to place both the theory of dynamical systems and the theory of temporal data on the same mathematical and linguistic footing. This is precisely how our theory of narratives addresses Desideratum **(D5)**: since both narratives (our model of temporal data) and Schultz, Spivak and Vasilakopoulou's *interval sheaves* (a general formalism for studying dynamical systems) are defined in terms of sheaves on categories of intervals, we have bridged a significant linguistic divide between the study of data and dynamics. We expect this to be a very fruitful line of further research in the years to come.