I GOT A LETTER FROM MY PAST SELF

(Un)managed Change and Provenance

Rhiannon Bettivia
Simmons University
USA
bettivia@simmons.edu
0000-0003-4593-562X

Yi-Yun Cheng
Rutgers University
USA
yiyun.cheng@rutgers.edu
0000-0001-6123-7595

Michael R. Gryk
UCONN Health
USA
gryk@uchc.edu
0000-0002-3483-8384

Abstract – Significant properties (sigProps) research often focuses on the preservation targets. Yet research consistently shows that what is significant about an object is not necessarily inherent to objects. Simultaneously, sigProps research does not adequately attend to temporality. Time is built into the concept of sigProps: they are about what ideally should not change over time. This paper centers temporality in relation to sigProps to explore challenging case studies.

Keywords – provenance, managed change, identity, temporality

Conference Topics – Sustainability: Real and Imagined, From Theory to Practice

I. INTRODUCTION

Calvin: My past self is corresponding with my future self.

Hobbes: Too bad you can't write back.

--Watterson, 1995

Digital preservation recognizes that long-term preservation entails managed change. Managing change is necessary to ensure that users understand the overarching conceptual object as one and the same over time [18]. The need to imagine and plan for the future is one of the inherent challenges of digital preservation: digital preservationists must think like futurists [17]. Yet the relationship between identity and change is a quotidian concern. The cartoon character Calvin, of Watterson's Calvin and Hobbes series, constantly engages in time travel wherein he interacts with his future and past selves (Fig. 1). This comedic device points to the very real ways in which a person is, at different points in their life, both the same person and a fundamentally different person.

The challenges of identifying that which must change over time has impacts on digital preservation work across disciplinary spaces. In this short paper, we explore two research themes:

- Theme 1: In what ways is Past Calvin the same and different than Future Calvin?
- Theme 2: How do the nuances that distinguish people over time change when applied to physical and digital objects?

These themes have practical applications for digital preservation. Significant properties (sigProps) are "[t]he characteristics of an Information Object that must be maintained over time…" [9]. The concept of sigProps is both crucial and challenging: the need is acknowledged but the practice is hard. SigProps refer generally to the properties of a conceptual object that are required for its ability to establish its authority in the world. SigProps hinge on two key aspects: objects and time. In this paper, we focus on the temporal aspects and provenance in order to advance the scholarly conversation around the wicked problem of sigProps.

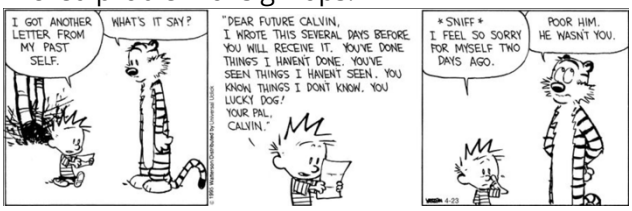

Fig. 1. Calvin and Hobbes, [April 20, 1995]

iPRES 2023: The 19th International Conference on Digital Preservation, Champaign-Urbana, IL, US.

https://hdl.handle.net/2142/121126

II. Literature Review

A. Temporal Provenance

SigProps support inherent change over time. Documenting these changes is part of telling the stories of objects, or provenance. Literature on temporal provenance focuses primarily on the e-sciences domains. Temporal provenance in scientific data is framed as (1) an ordered process based on causal relationships; (2) independent time slices; (3) circular processes.

Provenance models usually express time in an ordered fashion. For instance, in the Open Provenance Model (OPM), a second sequential process can only be initiated after a first process has occurred [15]. This suggests that the processes are directional, forming a directed acyclic, provenance graph.

In defining temporal provenance, Chen et al. discussed the potential of partial ordering of provenance graphs, and how one might be able to partition events into distinct time slices [5]. Similarly, Beheshti et al. proposed the Temporal Provenance Model (TPM) that puts time at the core in provenance documentation, as opposed to other event- or object-oriented provenance models [1]. In the TPM, time in provenance is captured not as a causal event, but as individual time-stamps to allow for versioning control of the same data objects. In this sense, time is an independent variable that partitions data objects into snapshots.

McPhillips et al. developed YesWorkflow, a scientific workflow management system built on the foundational concepts of retrospective and prospective provenance [13]. While retrospective provenance documents the execution, or past occurrences of a program, prospective provenance records the scripts, or the forward-looking recipes that enable a program to run. Here, the concepts of prospective and retrospective provenance are treated in a non-linear, circular fashion that supports a more nuanced approach to time in documentation.

Discussions of the temporal dimensions of provenance often center on metadata documentation, not on the data objects per se. Further investigation is needed on the use of temporal provenance to understand how data objects evolve over time.

B. Necessary Change

The Digital Preservation Coalition defines digital preservation as "...the series of managed activities..." [8]. In discussing artifactual objects, Owens [16] writes, "... what makes Mount Vernon *Mount Vernon*? Like all physical objects, it is changing at every moment" (p. 16). What are the sigProps of objects that are constantly changing? Historical contiguity is maintained through changes that comport with physical changes already happening: preservationists, digital or physical, roll with the changes that are going to come and make conservation decisions accordingly.

The question here, what makes a thing *that thing*, is central to the foundational understandings of the field of digital preservation. Thibodeau (2002) contributes terminological structure to the idea of the things that are the preservation targets: *that thing* is a conceptual object, supported by a pyramid of logical and physical objects. Preservationists make changes that can alter, re-order, substitute, or otherwise move the logical and physical pieces, while the top-level conceptual object must remain the same for the user in question. This approach mirrors models like the Functional Requirements for Bibliographic Records (FRBR), where the overarching conceptual work has various manifestations, expressions, and items that represent it [10]. The PREMIS metadata model also mirrors this structural approach to delineating *that thing* with its top-level intellectual entity object type [17].

Because of the foundational approaches digital preservation takes to *that thing* and managed change over time, it is a field that is poised to make broader impacts on issues at the intersection of the identity of objects and time. The following section employs case studies, biochemical research samples and video game franchises, to explore the themes stated at the outset.

III. Cases

A. Biochemical Research Samples

There is a renewed push to adopt persistent unique identifiers for samples in the natural sciences [4]. Biochemical samples are often altered, degraded or consumed in the process of a study, introducing the question of whether a persistent identifier is warranted for objects which themselves are not persistent.

In a biochemical laboratory, these ephemeral samples are typically given local identifiers, for instance with controlled experiments on multiple samples which vary in the concentration of a reagent or some other preparation step. This local identifier fulfills two simultaneous purposes: (1) it identifies the physical sample which is part of the experimental workflow and (2) it identifies the significant attributes of this particular sample with respect to the other samples which will be part of the study. In the latter case, a sigProp of the sample is its provenance - what it contains, how it was prepared, how it was treated, how it was stored, as well as temporal issues such as how long it has been since it was treated. Each of these concerns manifest itself on both the physical and concept level. It might be of importance whether a sample was stored at 4°C or at -20°. Alternatively, it might matter that a sample was stored in the 3rd floor freezer because there was a power outage in that room.

All of this is compounded by the fact that biochemical samples degrade over time. Samples age just as Calvin does, yet often on a timescale where the controlled variation between samples may be smaller than the variation within a single sample over time. This leads to some particularly tangled provenance stories when one wants to document the provenance of a sample and the methodology of an experiment in sufficient detail that it can be reproduced by others.

B. Super Mario

The previous case looked at the mechanics of organic change and the implications for identifying biochemical research samples over time. This section explores a socio-cultural example of the same phenomenon in the evolution of popular media figures over time, drawing from the work of McDonough and the Preserving Virtual Worlds grants [2,11,12]. *That thing* is Super Mario (Fig. 2), the Nintendo character who features in many media, starting with the *Donkey Kong* arcade games in the early 1980s.

The work of Preserving Virtual Worlds (PVWI and PVWII) is foundational to video games preservation. Two key findings that arose from PVWI are that (1) preserving interactive digital media requires a more systemic approach to determining sigProps even while acknowledging that (2) the preservation of popular games defies universal solutions.

Fig. 2 Uniqlo Super Mario 35th Anniversary T-Shirt depicting iterations of the character spanning the years 1985-2017, released in 2020.

PVWII identified the technical layers that make up a digital game as part of locating those sigProps. These layers include: the hardware/processor; the firmware; the software support; the physical; the application; and the experience layer [2].

Technological capabilities play a role in character design. Early design was frequently defined by the pixels and colors that fit within the storage and processing limits. Early Mario is pixelated in red, brown, and peach in 1988's Super Mario Bros. *(Mario 1)*. 2022's Mario + Rabbids Sparks of Hope is three-dimensional and brightly colored, wearing the iconic blue and red outfit (Fig. 3).

Fig 3. *Super Mario Bros.* (1985) and *Mario + Rabbids Sparks of Hope* (2022); images drawn from Wikipedia, image rights belong to Nintendo and Ubisoft.

At every layer of the technical stack, these versions of Mario are vastly different across a span of 37 years, including the processors, peripherals, displays, and experiences. Experiential differences are important, because this is where many users find the conceptual object in gaming. That it is possible to take the technological stack of the Switch and approximate the experience of *Mario 1* via Nintendo's emulator indicates that underlying

physical and logical pieces can change while the experience of *that thing* remains largely intact: this is a manifestation of sigProps in practice.

This case study is about the relationships between various manifestations of Mario (Fig. 2). Much as biochemical samples and Mount Vernon change over time, so has Mario over nearly four decades. When biochemical samples change in a lab context, the experiential differences might arise from their behavior in experiments. Marios differ in many ways over time. How and why do players recognize Mario as *Mario*? Part of the answer lies in how people make meaning of information. Clement traces how meaning is included in early information theories and she argues that users make meaning with information, rather than it being inherently meaning*ful* [6]. Marios remain *Mario* not just because of inherent characteristics like his blue and red costume, but because of meanings that come with interaction. The colors of Marios' costumes evoke a Mandela Effect: even when his outfit isn't actually red and blue, like in *Mario 1* or 1988's *Super Mario Bros. 3*[1], players remember *Mario* as red and blue.

McDonough notes that, "… [the p]reservation of computer games is in many ways a knowledge management problem, and without adequate metadata, managing the knowledge necessary to keep a game accessible and understandable is an insurmountable task." [14] This metadata is a form of provenance, and it must incorporate time: temporal framing for the objects and the temporal provenance that documents change in a way that enables objects to establish and maintain authority.

IV. Discussion and Conclusion

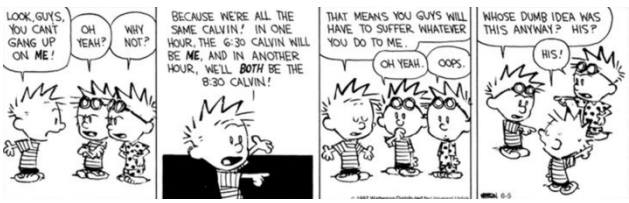

Fig. 4 Calvin and Hobbes, [June 2, 1992]

In a series of 1992 strips, Calvin attempts to avoid homework by time traveling to find a future Calvin who has already done it (Fig. 4). Unlike the arc where Calvin had a one-way conversation with himself via snail mail, here the Calvins literally find themselves in a room, communicating across time from 6:30-8:30, from homework time to bedtime. Ultimately, the 3 temporally differentiated Hobbes mediate the situation and do the homework. The aim of provenance documentation is to move beyond the one-way communication that comes from the past leaving missives for the future to something that resembles mediated conversations where past, present, and future can collaborate to form the best solutions. In previous work, we suggest that *subjunctive provenance* may improve provenance practice, acting as a mediator like the Hobbeses [3].

SigProps are inherently related to identity and time: they are the characteristics which determine whether the thing remains *that thing* over time. These cases demonstrate that significance is not necessarily inherent to an object: vastly different Marios are still experienced as *Mario*, the 3 Calvins are still just Calvin. Authenticity doesn't occur in a vacuum: meaning comes from experiences with objects rather than objects being inherently meaningful. Authenticity is a product of a relationship between objects and stakeholders [2,7].

The fundamental question remains: is the thing *that thing*? The answer is partly domain-dependent: in data management, it would be culturally common to see a change in a dataset resulting in a new data set, **Δ** dataset, even if the contents remained largely the same. However, a visibly obvious change in Mount Vernon, like the loss of a roof during a hurricane, does not result **Δ** Mount Vernon: it is still *Mount Vernon*. When Calvin tells himself, "You know things I don't know," he's talking about his own provenance: what differentiates the Calvins is what they've experienced. This raises the question: can provenance itself be employed as that which distinguishes a thing both as and from *that thing*?

These challenges are not academic. Practitioners manage diverse object and data types that behave differently enough that preservation and provenance practices are hard to universalize. Persistent identifiers that work for moon rocks do not work for biosamples. It is not that moon rocks don't change, but that the speed at which they do so is slower than a human life span, while biochemical samples might change more through natural organic

[1] The second image from the left in Fig. 2 is from this title, released in 1988 in Japan, 1990 in the US, and 1991 in Europe.

decay in a few days than they do in an experiment which is meant to alter them. Simultaneously, documentary processes that were done by hand for artifactual objects are impossible in computational environments: humans cannot document nanoseconds by hand. Incremental change is also a temporal facet that challenges documentary practices: there is a saying that it takes 7 years for every cell in the human body to be replaced with a new one. This saying points to three things: (1) that biological matter is always in a state of flux and change; (2) that humans assign symbolic meaning to this type of change; and (3) that humans understand incremental changes differently than other types of alteration. This type of biological incremental change is analogous to the Ship of Theseus story; it's the same kind of scenario that digital preservationists face when trying to track the knowledge base of a designated community.

This short paper presents a progressive idea: that digital preservation has not yet dealt sufficiently with the temporal aspects of sigProps. Time is always there in preservation work, but often at the periphery, where the changes of the object are documented and not the change of time itself. When that happens, difficult scenarios challenge existing models– Marios, Calvins, biochemical samples. This leads to a proliferation of standards and extensions, like the *provlets* of PROV, without solving the underlying issues. SigProps research often focuses on the preservation targets. Yet research consistently shows that what is significant about an object is not necessarily inherent to objects. Simultaneously, sigProps research does not adequately attend to temporality. Perhaps because time is part of the definition of sigProps, and part of digital preservation overall, it has been taken for granted and its role has been underexplored.

V. Acknowledgements

MRG acknowledges partial support from National Institutes of Health grants GM-111135 and GM-109046 and National Science Foundation grant 194670.

VI. References

[1] Beheshti, S., Motahari-Nezhad, H. & Benatallah, B. Temporal provenance model (TPM): model and query language. ArXiv Preprint ArXiv:1211.5009. (2012)

[2] Bettivia, R. Mapping Significance of Video Games in OAIS. *IPRES*. (2016)

[3] Bettivia, R. Cheng, Y-Y, & Gryk, M.R. What does provenance LACK: how retrospective and prospective met the subjunctive. In *Information for a Better World: Normality, Virtuality, Physicality, Inclusivity: 18th International Conference Proceedings*, 74–82 (2023)

[4] Buys, M., & Lehnert, K. (2021). Bringing together communities: IGSN and DataCite. DataCite. (2021). https://doi.org/10.5438/THHF-KX17

[5] Chen, P., Plale, B. & Aktas, M. Temporal representation for scientific data provenance. *2012 IEEE 8th International Conference On E-Science*, 1-8 (2012)

[6] Clement, T. The ear and the shunting yard: Meaning making as resonance in early information theory. *Information & Culture*. **49**, 401-426 (2014)

[7] Dappert, A. & Farquhar, A. Significance is in the eye of the stakeholder. *Research And Advanced Technology For Digital Libraries: 13th Euro- pean Conference, ECDL 2009, Corfu, Greece, September 27-October 2, 2009. Proceedings 13*, 297-308 (2009)

[8] Digital Preservation Coalition. Digital Preservation Handbook, 2nd edition. *Digital Preservation Coalition*. (2015)

[9] Grace, S., Knight, G. & Montague, L. Final Report. *InSPECT*. (2009)

[10] McDonough, J., Kirschenbaum, M., Reside, D., Fraistat, N. & Jerz, D. Twisty little passages almost all alike: Applying the FRBR model to a classic computer game. *Digital Humanities Quarterly*. **4** (2010)

[11] McDonough, J., Olendorf, R., Kirschenbaum, M., Kraus, K., Reside, D., Donahue, R., Phelps, A., Egert, C., Lowood, H. & Rojo, S. Preserving virtual worlds final report. (2010)

[12] McDonough, J., Senseney, M., Bettivia, Rhiannon, Evans, C., Kraus, K., Kirschenbaum, M., Donahue, R., Egert, C., Phelps, A., Decker, A., Lowood, H., Rojo, S. Preserving virtual worlds 2 final report. (2013)

[13] McPhillips, T., Song, T., Kolisnik, T., Aulenbach, S., Belhajjame, K., Bocinsky, K., Cao, Y., Chirigati, F., Dey, S., Freire, J. & Others YesWorkflow: a user-oriented, language-independent tool for recovering workflow information from scripts. ArXiv Preprint ArXiv:1502.02403. (2015)

[14] Montfort, N. & Bogost, I. Racing the beam: The Atari video computer system. (Mit Press, 2020)

[15] Moreau, L., Clifford, B., Freire, J., Futrelle, J., Gil, Y., Groth, P., Kwasnikowska, N., Miles, S., Missier, P., Myers, J. & Others The open provenance model core specification (v1. 1). *Future Generation Computer Systems*. 27, 743-756 (2011)

[16] Owens, T. The theory and craft of digital preservation. (Johns Hopkins University Press,2018)

[17] PREMIS Editorial Committee. PREMIS data dictionary for preservation metadata Version 3.0. (Library of Congress, 2015)

[18] Thibodeau, K. Overview of technological approaches to digital preservation and challenges in coming years. *The State Of Digital Preservation: An International Perspective*. 4-31 (2002)