Legacy data, like archaeology, is a renewable resource. Every technological, methodological, or practical change in archaeological practice, analysis, software, or hardware creates the potential for more legacy data just as the continuous work of depositional processes creates more archaeology.

It goes without saying, of course, that just as not all depositional processes are considered worthy enough to create archaeology, not all old data is legacy data. The conditions that allow for the creation of legacy data hinge on the significance of the site, our ability to access the data, the methods by which archaeologists originally produced the data sets, and, perhaps most importantly, the questions that we intend to ask of the data. Old data that is not useful becomes dead data until technology, access, and research interests align to convert dead data into legacy data. Dead data are forgotten while legacy data live in the present.

My paper today will reflect on my experience producing legacy data as part of a team of scholars studying the ancient site of Arsinoe in Western Cyprus which was excavated for about 20 years by the Princeton Cyprus Expedition beginning in 1984. In particular, I’d like to explore the distinct character of legacy data as a product of late modern concept of flow, and then reflect on how
understanding of the relationship between legacy data and dead data as a matter of flow offers an opportunity to think a bit about archaeological time. As you’ll see, things get a bit messy in this paper, but maybe that’s always the story with legacy data. It might even be part of its charm.

Polis

Since 2010, the Polis team’s attention has focused on the trench notebooks from the areas of E.F2 and E.F1 in the Polis excavation grid. These hand-written, bound notebooks described the excavation methods and the various deposits associated the architecture preserved at these sites. E.F2 produced an Early Christian basilica-style church which was called the South Basilica, a Late Roman well-house, Roman period roads, a lamp kiln, and various earlier walls and features especially associated with water management in the area. Our main focus, at least at first, was on the chronology of the transformation of the “South Basilica” from wood-roofed to barrel vaults. As our work developed, however, our research question expanded and we are now studying the smaller, apparently industrial, site of E.F1 as a window into the changing uses of the northern edge of the Late Roman city and as a way to contextualize the large and significant assemblage of Late Roman ceramics.

Narrative trench notebooks are not an uncommon form of legacy data. In fact, their obvious value at Polis led Joanna Smith, the project’s current director, to scan them and make them available to teams in the U.S. and Cyprus. The greatest challenge in the study of Polis notebooks is that excavators used a hybrid method of excavation that was not strictly stratigraphic. Excavators dug in levels and passes. The former coincided loosely to stratigraphic deposits or, just as frequently, distinct areas in the trench. The latter were either literal passes with the pick or episodes of excavation bounded by time, features, a prescribed depth, or some other distinctive character. Confusingly, passes sometimes corresponded to stratigraphic changes while in others, they did not. Excavators
also opened multiple levels in their trench at the same time disregarding a “last in, first out” approach that is characteristic of most stratigraphic excavation. In fact, some levels could be open for virtually the entire excavation season with excavators removing deposits and defining features from time to time as workers and attention allowed.

The excavation notebooks recorded daily excavation notes on all the active levels and passes in a trench. As a result, the description of a particular level (or even a particular pass) could appear across multiple days and notebook pages. This not only made it hard to understand the character of any single level, but also to grasp the relationship between various levels and passes. Because neither levels nor passes had to be stratigraphically defined, there need not be a depositional relationship between various levels. That being said, our ability to discern deposition in the data nevertheless relies on the careful interpretation of the description of levels and passes. From these descriptions, we produce what we have affectionately called a “Franco Harris Matrix,” after the recipient of the similarly miraculous “immaculate reception” in the 1972 AFC playoffs.

Once we had prepared our Franco Harris Matrices, we analyzed the pottery from contexts that we could define stratigraphically and that shed light on architectural features in the various trenches. We analyzed and recorded the ceramics according to the year of excavation and trench as well as the level and pass even when we were aware that the level and pass were not depositional unique. We then combined this new data with a database that recorded so-called inventoried finds, which described artifacts inventoried because they deserved more detailed descriptions, conservation, or special storage. To make this database “talk to” our analysis of context pottery required some additional recoding of the inventoried finds database. In most cases, this allowed us to produce assemblages that we could map back onto our stratigraphy, but in some cases, it demonstrated the disjunction between our stratigraphic analysis and the original recording method. For example, if the excavator only records the object by level and not by pass and we have discerned a change in depositional context
within a level (say between pass 3 and 4), then this artifact fits only awkwardly within our interpretative scheme rendering it less useful in our effort to date contexts and associated architecture at the site.

Flow

Thus the flow of data within our system plays a key role in distinguishing dead data (outside the flow) from legacy data within the flow. The concept of dataflow appeared in the 1990s as a term for the organization of fragmented processes in parallel computing. At the same time, the more general term “workflow” became common at about the same time to describe the sequence of processes used to produce a particular outcome. Dataflow and workflow privilege the organization of data in ways that allow information or objects to move smoothly through networks. It is hardly surprising that social and critical theorists Deleuze and Guattari found the concept of flow emblematic of the late modern condition and symptomatic of the fluid movement of good, capital, and people that defines neoliberalism and contemporary logistics.

Flow plays a distinct role in how we think about the value of data in archaeology, and as a number of critics have recently observed, data that cannot or is not (or in some cases should not) be used, dies. In fact, it seems to me that the designation of legacy data only applies to data that required adaptation to be reused. Dead data is just dead data without a legacy, and data still moving about existing workflows is just archaeological data that has acquired no legacy. To make a simple point unnecessarily complex, legacy data is data that is already adapted somehow to contemporary research. Recombining old data with new data sets involves deciding which elements of the legacy dataset links to new data or serves to address new research questions.

As Mary Shelley reminds us, combining and resuscitating the dead is often a fraught exercise. At Polis this process revealed certain unexpected complications. While the notebook data that we studied remained safely locked into its
notebook, to make it useful for our research, we attempted to break it out of its scanned-paper prison by transcribing some of the level and pass descriptions to make it easier to understand various contexts. It soon became apparent that these transcriptions created new problems. In our effort to understand levels and passes in a depositional context, we separated them from their daily context. In an excavation where multiple contexts might be open simultaneously and that later contexts were not necessarily removed before earlier ones, understanding which contexts are open at the same time provided clues to understanding the potential for contamination between levels and passes. In other words, as we re-contextualized the levels and passes so that we could understand the depositional contexts, we also started to de-contextualize the excavation itself. This tension between the original medium and methods and our efforts to reuse the data further reinforced the Frankenstein nature of the Polis notebooks as legacy data both in that it maintained an original context that is separate from our contemporary needs and nevertheless remains susceptible to the requirements of contemporary data flow.

Legacy Data and Time

At this point, this paper could speak to the importance of metadata and paradata in ensuring that archaeological data flows do not create monstrosities (no matter how human they are), but I’ll defer to folks like Eric Kansa whose call for “Slow Data” offers a compelling antidote to homogenizing tendencies of Big Data. Instead, I’d like to explore how the concept of data encourages us to think about time in the context of digital archaeology. My tentative foray into the issue of time in digital practice is a very modest attempt to fill a lacuna in recent discussions of time in archaeology which have largely overlooked digital artifacts. Moreover, it seems like legacy data is worthy of this additional attention because it exists, by definition, between the past and the present.
Most of us accept that archaeology studies the past in the present (and for simplicity’s sake I’ll exclude archaeology of the contemporary world which is not only complicated by the concept of contemporaneity, but also by the rather defuse and challenging body of theoretical literature). In other words, we tend to define the object of archaeological study different from the time of archaeological work. We might also accept that the prevailing paradigm in archaeological practice is one of progress through time. After all, the journal of methods from the SAAs is called *Advances in Archaeological Practice*. We can, of course, acknowledge that since the mid-20th century there are any number of valid critiques of the paradigm of progress and archaeology has embraced many of these critiques. Legacy data, however, doesn’t necessarily care about those critiques because it requires the paradigm of progress for its definition as part of a methodological and technological past that can be adapted to present needs of our discipline.

I wonder whether the concept of flow can offer insights into the move from one temporal state - that archaeological past - to another - our present research goals and questions - while still complicating views of archaeological methodology that privilege progress. The temporal fluidity of legacy data which require it to be both past and present simultaneously reflects an obscure quality present in most archaeology. In this way, it echoes the view of archaeology advanced by Laurent Olivier in his *The Dark Abyss of Time*. The interplay between the past and the present in the Polis notebook data, for example, shares Olivier’s (and others’) recognition that the irreducible character inherent in the materiality of the data as notebook pages and underscores how this burden makes any translation and transformation incomplete and problematic. The flow of legacy data, then, is not just toward the present, but toward the past as well.

In this way, legacy data shares an important temporal character with the objects and monuments at the center of debates over archaeological heritage, conservation, and repatriation and here the work of Francois Hartog on what he calls the regimes of historicity comes into play. In many ways, legacy data invites
us to engage fully with more than just the past as the object of archaeological study or the ostensible progress of archaeological methods, practices, and technology. Legacy data, however, must live in the present, and consequently shares an ethical time with calls for repatriation which emphasizes the flow of objects from the deep past of archaeology, through various colonial, commercial, or transnational encounters, to the contemporary space nations-states and communities. As a post-colonial gesture, repatriation is not about restoring a past that has been somehow disrupted or disturbed, but about recognizing the flow between past and present as an inherent feature of the repatriated artifact.

Perhaps the similarity between legacy data and repatriated artifacts allows us to think about the study of legacy data as a kind of ethical opportunity. By reflecting on legacy data as a flow and recognizing this flow as at least one way to articulate the complex issues surrounding heritage, we make explicit the messy composite time of archaeological knowledge and the potential of legacy data as a critical space for informing the fluid character of the archaeology, heritage, and the present more broadly.