

The Chronopolis project is a mass-scale digital preservation and curation network designed to preserve digital data of any type or size. It is a founding node of the Digital Preservation Network (DPN) and works toward the DPN goal of “ensuring that the complete scholarly record is preserved for future generations.” Chronopolis is designed to accept data with minimal requirements on the data provider, which I interpret as a way of saying it is “user-friendly.” The basic structure of the project is that it allows users – universities and research entities – to upload data to Chronopolis which then provides access for the data provider and creates copies of the uploaded data in three separate geographical locations. As of a few years ago, Chronopolis houses four diverse collections: a backup of the complete digital holdings of the Inter-university Consortium for Political and Social Research (ICPSR), based at the University of Michigan, "Web-at-Risk" collections from the California Digital Library (CDL), geospatial data resources from the North Carolina Geospatial Data Archiving Project, and several decades of data from research cruises from the Scripps Institution of Oceanography (SIO) at UC San Diego.¹

The ingesting and managing of data is accomplished by Chronopolis’ partnership with DuraSpace, a not-for-profit organization promoting durable, persistent access to digital data that is “committed to our digital future.” Chronopolis uses DuraSpace’s service DuraCloud as content/data staging area and the main interface for its users. DuraCloud enables users to take a snapshot of any content they have uploaded and then that snapshot is transferred to Chronopolis where it is checksums and manifests are generated and it becomes part of the Chronopolis storage network. When needed users can request a previously submitted snapshot of content and it will be retrieved from Chronopolis and restored to DuraCloud where users will then have access to it. Chronopolis is directed from the University of California San Diego (UCSD) but is comprised of several partner organizations. These partners include: The UC San Diego Library (UCSDL); the San Diego Supercomputer Center (SDSC) at UC San Diego; National Center for Atmospheric Research (NCAR); University of Maryland Institute for Advanced Computer Studies (UMIACS). As can this list of partners that comprise Chronopolis illustrates, this project is not designed for the consumer market. It is rather geared toward large institutions with very large datasets.

In keeping with its stated goal to accept data with minimal requirements on the user, I can find no documentation of restrictions on the types of data Chronopolis will accept. It seems they will accept everything and lots of it. Given the nature of their partners and expected users this makes sense as the various files and formats used in scientific and academic research are too numerous to count. Though, stated on a page in their website is the following: *Format obsolescence is not an immediate concern of the Chronopolis system. Instead, this is regarded as the responsibility of the data providers. The single, overriding commitment of the Chronopolis system is to preserve objects in such a way that they can be transmitted back to the original data providers in the exact form in which they were submitted.*² Given the large task Chronopolis has laid out for itself, it is perhaps not surprising that they cannot take on the problem of format

¹ Minor, D., D. Sutton, A. Kozbial, M. Burek, M. Smorul. Presented at iPRES, The Sixth International Conference on Preservation of Digital Objects (iPRES 2009), (2009)

² "Chronopolis." *Chronopolis*. Web. 22 Sept. 2014. <http://libraries.ucsd.edu/chronopolis/>

Harry Bleakley McDowell

Assignment 1

Digital Preservation Fall 2014

obsolescence. Though this decision leaves them with the risk, albeit a very slight one, that Chronopolis could end up housing thousands of terabytes of that is useless because none of the users read the fine print and bothered to monitor the obsolescence of their formats.

Chronopolis will ingest only digital material and is not a storage center for material objects from its users. As long as it is in a digital format Chronopolis has seemingly no restrictions on the type of data, so content uploaded to Chronopolis can be born digital or digitized from physical or analog sources.

By offering to preserve, store and curate data from research and scholarly institutions Chronopolis is confronting two main problems. First, the amount of data being produced is starting to overwhelm even large organizations with mandates to safely preserve their data. By creating its grid based storage system of geographically separate nodes connected by high-speed networks the DPN, of which Chronopolis is an essential part, alleviates the pressure on organizations tasked with preserving their data. With a mission to preserve the very types of varied data that research and scholarly institutions produce Chronopolis and the larger DPN network will be a great help. Chronopolis is also certified by the Center for Research Libraries as a “trustworthy digital repository” that meets accepted best practices in the management of digital repositories. This certification provides its users with as much peace of mind as they are likely able to get in the digital preservation environment of paranoia. Second, Chronopolis is focused on making sure the data submitted by users remains understandable and useable to them (excluding issues of obsolescence, which I mentioned earlier). The status of data that is uploaded to Chronopolis is viewable in a standard web browser. Users can log in to Chronopolis’ data portal and check on “the status of objects (including replication status), last time verified, size of collections and objects, and the overall health of the system.”³ Chronopolis also stores all of its data in the same directory structure in which it was received. Users can easily drill down to a specific file and check on its attributes and status.

Chronopolis ensures the integrity of its data by using the Audit Control Environment (ACE). The ACE system is a web-based application that can perform two types of audits - “the first, a file audit will check files in registered directories against stored hashes to ensure files have not been corrupted. The second type of audit, a token audit, will check the stored hashes against a remote Integrity Management Server to ensure nobody has tampered with the stored hashes.”⁴ ACE works according to the policy of the archive, which in the case of Chronopolis is to audit collections every 30 days.⁵ Chronopolis also uses the BagIt packaging protocol to transfer data into its storage locations and retrieves data from users usually via *ssh* or the *wget* transfer protocol.

Chronopolis is in its fifth year of functioning. It is hard to assess how it is doing. The pricing section of their website informs that “Chronopolis is in the midst of major positive changes in its revenue model” driven by integration with DuraCloud in service to DPN and tighter ties with UCSD. The changes are anticipated to improve the Chronopolis user interfaces and provide a “much needed revitalization of the Chronopolis

³ <http://libraries.ucsd.edu/chronopolis/>

⁴ ADAPT. Web. 22 Sept. 2014. <https://wiki.umiacs.umd.edu/adapt/index.php/Ace:Main>

⁵ Minor, D. (2009)

Harry Bleakley McDowell

Assignment 1

Digital Preservation Fall 2014

funding model.”⁶Despite the self-describe positive nature of these changes, it appears obvious that one cannot currently become a Chronopolis customer, at least not through their website. This is certainly not a good thing for a project designed as a long-term preservation model. But, 2015 could, in fact, bring a much more robust and sustainable Chronopolis. The actual issues concerning their funding model and user interfaces are not described.

⁶ <http://libraries.ucsd.edu/chronopolis/>