Backgrounder for Early Fall 2021 CNI Executive Roundtables on Video and Data Storage

Oren Sreebny, Internet2

The past several years have seen an inexorable growth in the use of video in all facets of our lives, and use of video in campus settings is no exception. Video, both recorded and live, is omnipresent in instructional settings and use of video for both group and one-on-one meetings has become commonplace. The pandemic has greatly accelerated the everyday use of video across higher education, with an entire year of class sessions and professional interactions taking place over Zoom. While the rate of continued video growth remains to be seen, with the return to campus in the fall even as the Delta variant spreads, it's safe to say that video interaction has become a common part of the landscape in higher education and research institutions.

With the increased use of video comes many questions about saving, storing, retaining, and preserving video assets. It is trivially easy to enable recording of video, which has caused an explosion in stored video recordings from campus settings. Now institutions are, more than ever, having to grapple with questions about the costs of storing video recordings, what video gets retained and what doesn't, how access to video is controlled, and how video data is discovered or searched.

Many universities have sidestepped some of these issues by encouraging storage of video assets inside of cloud storage services that have offered unlimited storage with enterprise level subscriptions, such as Box and Google Drive. Others have been storing video assets in video systems that have also offered unlimited storage plans, like Panopto or Kaltura. Many, if not all, of the services that had offered unlimited storage have changed or are changing their offerings to now either limit the amount of storage available or to charge for storage use. These changes are also forcing increased scrutiny of policies and processes for storage of video.

In a March, 2021 Internet2 presentation titled "Managing Video Content Retention in Panopto", Dave Long and Mike Espey from the University of Iowa discussed an approach they are taking to applying retention policies to all video recordings stored in the University's Panopto video system. At Iowa they have implemented a retention policy based on when the last view of a video was. Videos that have received zero views over the most recent four years are deleted, though there is a mechanism for departments to exempt individual recordings from the policy. There are some items worth noting from their presentation (which is well worth watching): They were very deliberate about laying the groundwork for instituting the retention policy, taking over two years from the beginning of discussions to the initial deletions (and involving a wide range of stakeholders in those discussions); In the initial run in July, 2021, fully 25% of all of their video content in Panopto had received zero views in the past four years; 19% of those files that had not been viewed in four years had never received any views in Panopto. Similarly, the Stanford Center for Professional Development began migrating video assets over five years old to cold storage in AWS Glacier in 2019. Glacier is designed for data archiving and long-term backup, and features lower storage costs than regular AWS S3 object storage, but charges higher costs when retrieving that data. As of a conversation in August, 2021, SCPD had only had two requests for retrieval of assets in Glacier.

These types of automated storage practices are based on the type of medium (video) and the system where the assets are stored. Institutional retention policies tend to be based on the kind of material contained within a record - e.g., student records, faculty meetings (for example, see the University of Washington's University General Records Retention Schedule). It seems likely that these largely orthogonal approaches will need to be reconciled as practices mature.

As the amount of stored video grows, the discipline of actively managing the costs of storing that video becomes increasingly important. For example, one online video conference provider offers higher education customers 100 GB of cloud storage for $480 per year (larger commitments can lower the per GB cost). The same amount of raw storage in AWS S3 Glacier Deep Archive costs $1.19 per year (with no up-front commitment).

To demonstrate the scope of video data at universities, and some potential cost figures, the University of Iowa had 140 TB of video data in Panopto when they began their move to the cloud version in 2019; they were able to reduce that substantially by implementing their retention policies. Storing that amount of data in AWS S3's standard tier would cost over $38,000 per year, while storing it in S3 Deep Archive would cost $1,663 per year (using published, on-demand pricing — further discounts might be negotiated depending on advance commitments). Some institutions may have substantially larger collections of video data.

But keeping data in Deep Archive imposes additional costs when items need to be retrieved from storage, as well as different storage prices for keeping those retrieved files available. There are also additional questions about managing the access to assets stored in cloud storage, which requires individuals to be set up within an institution's cloud deployment and know how to navigate that system.

While none of these are insurmountable obstacles, the shifting landscape of handling ballooning collections of video assets will require a level of active management that is new to many institutions. As a starting point, institutions might think about some of the following questions:

- Where are video assets currently located?
- How much video is being stored currently?
- How quickly is the store of video growing?
- Are any retention policies being applied to stored video?
    - Is the implementation of those policies automated?
- Are there institutional (or state or local) retention policies that should be applied to video data? How are those implemented?
- Are there mechanisms for allowing assets that warrant long-term storage (e.g. distinguished lecture series) to be identified and exempted from automated deletions?
- Who is paying the cost for storage of video assets? What mechanisms should be in place for apportioning costs of video storage?
- What will be appropriate mechanisms for cost-effective management of video storage as the amount stored continues to grow in the future?