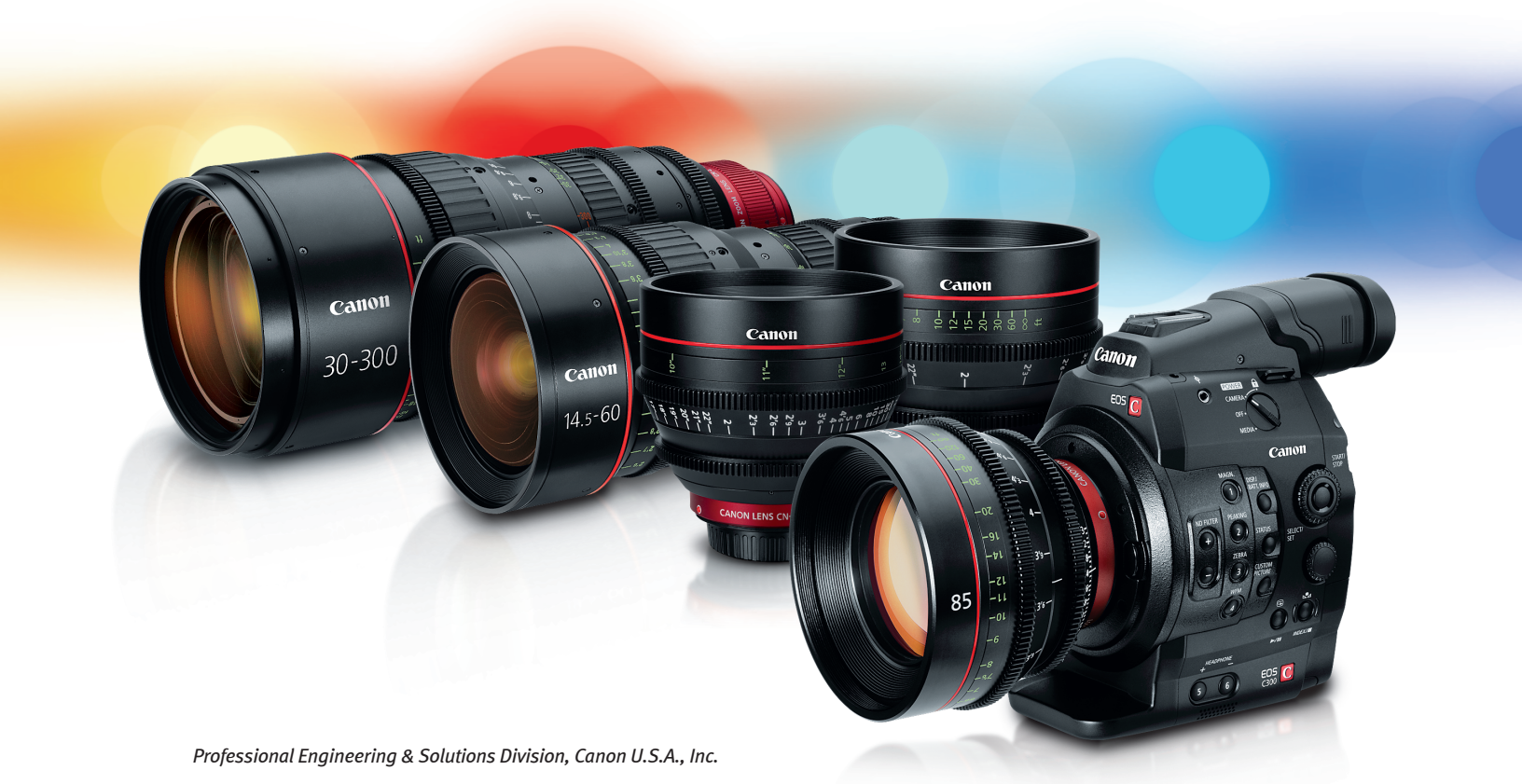
Canon

WHITE PAPER

EOS C300 / EOS C300 PL

BEST PRACTICES FOR DATA MANAGEMENT

Created March, 2012



Professional Engineering & Solutions Division, Canon U.S.A., Inc.

For more info:

cinemaeos.usa.canon.com

CINEMA EOS

BEST PRACTICES FOR DATA MANAGEMENT

LIGHT IRON

Title: **BEST PRACTICES FOR RESPONSIBLE DATA MANAGEMENT**
Prepared for: Canon U.S.A., Inc.
Prepared by: Michael Cioni
March 2012

OVERVIEW:

As the capturing of tapeless content becomes more and more common in our society, the ability to establish dependable ways in which to maintain and control that content is not nearly as common. One key example of this (on average) is the individual investment towards acquisition tools is typically greater than the individual investment in data management tools. When dealing with major studios and large distribution agencies, the attention to data management has recently become a major component to their business. Significant investments in salaried archivists are now a common venture for companies to explore in order to preserve their content for future profits. But when it comes to the small business or individual level, the scales of interest favor acquisition technology over equally important "responsible data management."

The first issue is to identify what the *need* is. I consider "responsible data management" the understanding, mastering and strategic planning of file-based media which ensures reliable accessibility between multiple users. The most common reason for faulty data management is likely due to lack of education or extensive experience. This lack of education is often manifested through two examples:

1. Technical Repetition: Because the process of cinematography is (essentially) the same when using the formats of film or video tape or files, the level of knowledge and understanding of shooting principles generally transcends all acquisition technologies. But when it comes to the back-end of these formats, file-based data management is a rather new concept with roots in more consumer markets such as online entertainment, video game consoles, mobile smart phones or DVRs. The result is a generation of data users who rely on drives that contain data to live longer-term lifespans than they were initially designed to.
2. Behavioral Preference: Another issue is the the type of behaviors certain key positions have in the filmmaking process. Because filmmakers are typically process-oriented people, they are often most satisfied when actively being creative. But when it comes to the passive portions of the process, many small businesses or independent filmmakers do not pay as close attention to the needs of these processes because their interest in the passive areas is less rewarding. The result is a formula in which many projects are poorly managed, foolishly managed or not managed at all.

DATA CALCULATING

The first step in attaining responsible data management is to properly estimate how much data will be required for a project. A common technique for drive purchases is the "check-point" technique. We've all done it: this technique is when you buy the largest single-drive at an electronics store and hope everything fits on it. Then when it fills up (often to capacity), you purchase another and another and so on. The problem with this is I have seen clients drop off 7 to 10 drives for a single project and the drives are all different manufacturers, different capacities, require different power supplies, are rated for different speeds and use different host interfaces. Not only does this look sloppy, but it's actually risky because there is no method to the backup strategy or consideration to the properties of the drive. With appropriate data calculating, users can purchase a much closer approximation of the drives needed and ensure they are optimized for the type of backup a specific project will require.

A helpful formula used for calculating how much storage productions need is as follows:

BEST PRACTICES FOR DATA MANAGEMENT

X x Y ÷ 1000 + 20% = T

DATARATE x FOOTPRINT = TERABYTES

Each of these color-coated components represents a slightly different equation. Once combined together, the final total (measured in terabytes) is the target drive capacity that should be purchased for the production. Because using time (measured in hours) is a convenient constant to compare against, we need to convert the source codec into a *data rate per hour of footage ("X").*

DATARATE

CODEC x MIN x HOURS ÷ 1,000 = **"X"**

FOOTPRINT

DAYS x CAMERA x HOURS PER CAMERA = **"Y"**

Putting this to an example, below is a case study using the C300 in the MPEG2 50Mps shooting mode. Because we often prefer to work in larger numbers (gigabytes instead of megabytes, etc.) we convert the source codec (50 megabits) to megabytes, which is approximately 1/8th in size. In other words, 50 ÷ 8 = 6.25 megabytes per second.

DATARATE

6.25MB/s x 60 x 60 ÷ 1,000 = **22.5 (gigabytes per hour)**

FOOTPRINT

30 x 2 x 2 = **120 HOURS (estimated record time)**

(contingency & drive safety)

22.5 x 120 ÷ 1000 + 20% = 3.2 TERABYTES

By this simple equation, we know that shooting on the C300 for 30 days with 2 cameras at 2 hours per day will require approximately 3.2 terabytes of total storage. That's also means we have the information required to create an ideal data configuration strategy.

BEST PRACTICES FOR DATA MANAGEMENT

DATA CONFIGURATIONS

Because there are so many different types and manufacturers of drives, it can be difficult choosing the most efficient and professionally optimized drive for file-based content. A helpful technique in choosing drives is to break down the different tasks required for a project and then line up those tasks with an appropriate configuration. For example:

Best practices suggests that file-based projects always need to be backed up 3 times because, unlike film and tape, the media source is quickly going to be erased and recycled. The best way to insure files last is by making appropriate copies as described later in this document. With the price per gigabyte (on average) cheaper than the price per minute of film or tape, the cost of making 3 copies is less of an impact than the benefits redundancy offers. With a production, projects are typically in one of two phases: *passive* and *active*. The active phase is when the project is shooting, editing, pulling visual effects and finishing. The passive phase is *after* the project is delivered and is available for distribution. The first clue this gives us is to line up the different phases of a production with specific data configurations that match each other. Data configurations can also have different phases. Like a production cycle, data can be both active and passive, but it can also be in a multi-volume or single-volume format.

Configuration Categories

Passive Storage: storage that can contain data, but not be reliably used for real-time viewing or playback (eg: LTO5 tape)

Active Storage: storage that can contain data and simultaneously be used for viewing and playback (eg: spinning disk drive)

Multi-Volume: small storage pools that divide up media across multiple drive chassis (eg: 6TB Light Iron "LightDrive6")

Single-Volume: large storage pools that contain all the media on a single set of drives (eg: Promise Vtrak x30)

By examining the phases of production and storage categories, we can use this model to choose the 3 best ways to backup our material and enable the files to last longer, reduce expense and increase efficiency.

An example of this is as follows:

COPY #1 = LTO5

Long-term archive storage that holds up to 1.5 terabytes
Satisfies both PASSIVE storage and MULTI storage categories

COPY #2 = SAS or eSata Mini Raid5

Short-term, portable archive storage that holds up to 12 terabytes
Satisfies both ACTIVE format and MULTI format categories
Raid 5 prevents against single-drive failure

COPY #3 = SAS or Fiber Raid6

Mid-term archive storage that holds up to 72 terabytes
Satisfies both ACTIVE format and SINGLE format categories
Raid 6 prevents against multi-drive failure

BEST PRACTICES FOR DATA MANAGEMENT

By following this technique, we can make 3 copies of our media and still satisfy 6 different forms of backup and retrieval strategies. The key to this technique is not to make all your copies on the same type of format and not to keep them all in the same place. By diversifying the containers themselves, the ability to save money and better protect the data becomes more achievable.

ON-SET vs IN-POST

One of the fastest growing trends in the motion picture industry is the gravitation towards *the set* being an ideal place for image manipulation of some kind. This trend is both common in small projects and large projects. The driving force behind this trend is typically creatively driven. For example, when there is the notion of new, faster, proximal opportunities that improve creative control, filmmakers typically are in favor of it. The result has been the manifestation of "best practices" by the men and women who manage this data on set. What is interesting is that the people responsible for managing digital files on set initially had nothing to do with actually handling *original digital negative* (o-d-neg) itself. Most on-set data support (such as digital imaging technicians) were originally in charge of managing the camera, monitoring the image and advising on exposure of gamma corrected or log-like digital video tape recording. But as file-based recording began to infiltrate the set, these skills migrated towards actually downloading, backing up and even transcoding the o-d-neg and returning the source cards back to the camera for recycling. While this is common and appropriate in many circumstances, my experiences have shown that the ideal place for downloading and transcoding is in a controlled environment away from the turbulence of an actively moving set.

A successful technique that we regularly use is to consider separating the data manipulation process into two conducive halves: the *creative half* and the *technical half*. The creative half is the portion that works well on set. An on-set operator has a computer station ranging from a small laptop to a full tower system with storage and monitors and can perform a range of important but creatively driven tasks such as:

+ initial picture QC of media
+ color correction or look previewing
+ initial sound QC
+ saving color looks as metadata information

After services like this are performed on the media, we find that shipping the original cards and o-d-neg to editorial for processing is a better choice than shipping a downloaded or copied drive. The reason is due to the unforeseen problems that can occur in drive transfers coupled with the lack of control of a moving and changing set. By doing the delicate work of backups in a controlled environment, the percentage of data issues can be reduced by enabling post personnel to perform their own downloads directly from the o-d-neg itself. This doesn't mean a copy shouldn't be made on set, rather it should *not* be the only copy relied upon to *create every copy thereafter*. Ironically, this workflow is essentially the same workflow that has been used on tape and film projects for decades. Film and tape were never processed, copied or prepared for editorial directly on the set. Today, many of the tools exist that make this possible, and should be exploited whenever appropriate. But we find the best practices for data management today suggests that "hero" copies and technically-favored tasks such as transcoding and syncing are best done in post. Similar to the on-set operator, a post operator will also have a computer system with storage and can perform a range of important but technically driven tasks such as:

BEST PRACTICES FOR DATA MANAGEMENT

+ second QC of material
+ check-summed download
+ 3+ copies
+ applying color correction from the set
+ syncing sound
+ labeling takes
+ transcoding to various flavors for editorial & dailies
+ binning into editorial programs
+ uploading to the web for online review
+ organizing pulled or circled takes

By examining the pros and cons of the data pipeline once footage is photographed, it is easy to see how two heads is better than one. Perhaps the most empowering and unique component of this plan is that even though all processing tasks are intentionally spread across two individuals, it does not require the use of outside 3rd party laboratory support. This is an important concept because it satisfies many of the benefits traditional labs offer through checks-and-balances that ensure data is being handled correctly and that problems with production or photography are quickly identified. When all of these tasks are put on the shoulders of a single person that lives and works on the set, the risk of a mistake, missed clip, unfinished task or major bottleneck significantly increases. By catering to the overlapping strengths of dedicated on-set talent and in-post talent, as well as respecting the limitations of both regions, file-based cameras can be a far more empowering and cost-efficient than film and tape ever could be.

CONCLUSION:

Much of our professional behavioral development is based on how we managed jobs duties in the past. As the technology and associated trends evolve, so must our design for how we intend our work to flow. By this we can conclude that differences in technology may warrant differences in behavior. The most common struggle I witness on productions is the migration to progressive tools without knowing how to thoroughly capitalize on them. This often philosophical difference is one of the largest barriers that keeps productions in the delicate and inefficient “balancing act” between modern and traditional techniques. It can also lead to the formation of bad habits and dependence on inferior methods that are derivatives of antiquated practices. One helpful technique in avoiding this is to look towards leading manufacturers; the best manufacturers today are building devices that are clearly designed to be both democratized and mobile. With the amount of software-based tools that filmmakers can now take advantage of, the barrier to entry declines while simultaneously creating new opportunities for talent to rise. Don’t be afraid to evaluate new tools and revise old trends until you find the combination in which they conventionally fall into place.