# Postprint: Big Data-Driven Changes in Documentary Content Presentation Techniques

**Authors:** Jiang Hanyu

**Date:** 2023-10-08T00:00:00+00:00

## Abstract

The utilization of big data technology has furnished documentary presentation with multiple novel technical modalities, thereby enriching the expressive forms of documentaries. Crowdsourcing technology was employed in the analysis of large volumes of complex images and techniques, while multiple data technological approaches were utilized in the processing of various information and knowledge. Big data was leveraged to optimize the collection, processing, and application of images; query engines, data storage, and data processing tools from the big data ecosystem were deployed to produce analytical reports. The use of portable 4K equipment for filming and data processing optimized operational workflows, enhanced material color grading, and, in conjunction with documentary presentation methods, manifested technical variations.

## Full Text

## Preamble

**Title:** Changes in Documentary Content Presentation Techniques Driven by Big Data Technology

**Abstract:** The application of big data technology has introduced various new technical methods for documentary presentation, enriching the expressive forms of documentaries. Crowdsourcing techniques are employed in the analysis of large volumes of complex images and technologies, while multiple data processing methods are utilized in the analysis of information and knowledge. Big data optimizes the collection, processing, and utilization of images, leveraging query engines, data storage, and processing tools to generate analytical reports. Portable 4K equipment is used for shooting and data processing, optimizing workflow, enhancing material color grading, and integrating documentary presentation methods to reflect technical changes.

**Keywords:** big data technology; documentary; query engine; 4K equipment; material color grading

**Author:** Jiang Hanyu

Documentary production in the new era must meet contemporary development needs by fully utilizing rapidly advancing big data technology to promote formal expression and content presentation in documentaries.

## 1. Big Data Technology

In the process of documentary production, multiple data techniques are employed, enabling data anonymization during image and video production to enhance security and privacy protection.

## 2. Technical Applications in Documentary Production

### 2.2 Query Engines

Big data technology encompasses various techniques including data mining, artificial intelligence, natural language processing, cognitive computing, machine learning, and large-scale data processing. These technologies can effectively clean and fuse massive datasets, perform modeling analysis, establish large-scale processing frameworks, and enable visual analytics. This technology runs throughout documentary content presentation and expression, employing multiple analytical techniques that significantly enhance analytical capabilities [1].

### 2.1 Overall Technology Implementation

In documentary content analysis and processing, a foundational big data storage repository has been established, incorporating NoSQL databases such as Cassandra/HBase and distributed file systems like Hadoop. The data development platform utilizes analysis platforms represented by Kafka, while MapReduce, Spark/Storm serve as representative frameworks in big data processing architecture construction. The combined application of these methods builds a highly scalable and available processing architecture, which has become one of the mainstream approaches in open-source big data distributed parallel processing analysis [2].

With the full development of artificial intelligence and machine learning in big data analysis, deep learning methods such as Long Short-Term Memory networks (LSTM), Convolutional Neural Networks (CNN), Bidirectional LSTM (BLSTM), and Recurrent Neural Networks (RNN) have been effectively developed and applied in documentary data analysis and storage processes. These implementations have improved the accuracy and effectiveness of multi-modal data analysis for video, images, text, and audio, enhancing the richness and layering of images presented in documentaries [3].

Additionally, the application process has achieved effective integration and utilization of cognitive computing and Knowledge Graphs, enabling efficient processing of billion-scale knowledge nodes and ten-million-scale relationships in the big data era. Crowdsourcing technology is employed in the analysis of large volumes of complex images and technologies.

Solr technology, developed on the Apache Lucene foundation, is a search method utilized by users including Instagram, Zappos, Sears, Netflix, Travelocity, StubHub, eHarmony, and Best Buy.

Presto is an open-source distributed SQL query engine capable of interactive analysis of over 250PB of data. It can perform interactive analysis of large amounts of pattern information in documentaries with high analytical speed, offering better performance than MapReduce and Hive. Drill enables effective querying of NoSQL and Hadoop databases while achieving cloud storage. It can run simultaneously on server clusters of thousands of nodes, rapidly processing petabyte-level or trillions of data records within seconds. It also enables large-scale data mining and real-time querying, and is compatible with multiple databases including Google Cloud Storage, MongoDB, MapR-FS, Azure Blob Storage, MapR-DB, Swift, and HDFS.

### 2.3 Report Generation

Kibana serves as a Web interface for ElasticSearch and Logstash log analysis, functioning as a search dashboard and Elasticsearch analytics tool that enables efficient searching, visualization, and operation of large information volumes during use [4].

KNIME (Konstanz Information Miner) is an open-source analysis and reporting platform that can simultaneously process large amounts of data. Zeppelin is a Web-based notebook for interactive data analysis applicable to Python (Apache Spark), Markdown, Apache Spark, Hive, SparkSQL, and other formats, playing a significant role in database presentation and data analysis [5].

Kettle' s application enables effective description and analysis of large amounts of graphical information and can be used for data management across multiple databases, playing an important role in Pentaho and being utilized in numerous domestic projects. See Table 1 and Table 2 .

## 3. Practical Application in Documentary Filming

### 3.1 Using Portable 4K Equipment for Shooting and Data Processing

To ensure a complete documentary shooting workflow, portable 4K equipment can be employed for filming, with big data technology and computer technology used for effective data processing.

The main camera position uses a Sony F55, recording in 4K XAVC 25P format, while the auxiliary position uses a Sony FS7, recording in 4K XAVC-I SOP

format. Additionally, a Sony A7S2 is selected, recording in 4K XAVC 25P format, serving as a flexible mobile position with good low-light performance. This configuration enables shooting of conventional environments, scenes, and dialogues between different characters [6].

Beyond main recording scenarios, documentary production also requires shooting of various scenes, overall environments, starry skies, and time-lapse photography. For this purpose, the setup includes two Nikon D5200s, one 1DX, one SD2, two A7R2s, paired with wide-angle and fisheye lenses. With current shooting technology development, drones such as the "Phantom" series can capture required footage of wilderness herding, ancient city ruins, desert landscapes, and village life. In field shooting, 4K action cameras like GoPro are used with vehicle-mounted, suspended, or handheld configurations to address insufficient camera positions for active scenes and narrow indoor spaces, applicable to ceremony and wedding scene filming. GOPRO can also be used for capturing small animals in desert environments [7] (Table 3 ).

During the filming of the documentary *The Last Desert Watchers*, beyond preproduction camera parameter solutions, image metrics were established based on selected camera models. The production used S-Gamut3.Cine color space paired with S-Log3, employing XAVC-S and XAVC-I video compression codecs. While there are certain differences in sampling quantization between these two codecs, they do not affect subsequent color grading and can still effectively match footage. This result is achieved by sacrificing RAW format recording, enabling optimal 4K combination while maintaining crew mobility. This equipment operation achieves 4K with high mobility, allowing photographers and equipment to operate and shoot independently without interference from subjects. The F55 and FS7 codecs used in color grading direction can balance highlight and shadow details in outdoor environments with high contrast ratios. When shooting in dark environments, clear images can be obtained using the A7S2's high ISO without lighting. Equipment used includes small aerial vehicles, GOPRO, F55, and A7S2, with photographers capable of working independently, providing good shooting flexibility [8].

[Figure 1: see original paper] Big Data Architecture Diagram
[Figure 2: see original paper] MapReduce Mind Map

### 3.2 Optimizing Workflow

Workflow plays a crucial role in documentary production, with the DIT (Digital Imaging Technician) 环节 being particularly significant in pre-production. Documentary shooting and production require establishing a robust DIT workflow where directors, cinematographers, and post-production directors perform their respective duties to collectively facilitate production.

A rigorous DIT workflow is established for documentary shooting, requiring well-organized and secure footage that can be fed back to the production team in real-time to promptly identify and address issues, establish improvement

methods, and create effective safeguard mechanisms to optimize post-production processing and strengthen technical integration and 磨合 between pre-production and post-production.

For example, during the filming of *The Last Desert Watchers*, a key focus of the DIT workflow was data backup and organization. The harsh desert shooting environment, with potential for dust storms disrupting power supply, could adversely affect equipment operation and backup security.

Therefore, the production used large-capacity mobile hard drives and MBP laptops requiring no external power supply for backup, employing traditional primary-backup methods where the laptop's battery power enabled normal hard drive read/write operations [9].

Integrating big data application with equipment usage, data backup was optimized through automatic verification via background programs. Verification checks whether files to be backed up have physical structure damage, ensuring backed-up data is readable and recoverable, thereby guaranteeing backup file security and reliability.

To facilitate big data processing, a material management lookup table was created to effectively filter useful materials from field notes, combine them with shooting requirements, check and register them, enabling post-production personnel to efficiently collect materials. After material collection, sound-picture synchronization was performed based on the editing timeline, and XML tables were exported for daily storage timelines, allowing headquarters editing and production to proceed based on the timeline, saving substantial material filtering and organization time and improving editing efficiency [10].

### 3.3 Enhancing Material Color Grading

Documentary shooting requires color grading based on directorial requirements and production objectives to effectively check shooting quality and help the production team develop a consistent cognitive style for the documentary' s overall visual presentation. This utilizes big data and related technologies for diverse image expression, enhancing image presentation layering.

A critical component of color grading is pre-coloring to compensate for pre-production shooting imperfections and provide feedback to the post-production photography team. Based on pre-coloring results, shooting parameters can be adjusted in real-time and feedback provided to the director to help determine the film' s style. During shooting, S-Gamut3.Cine paired with S-Log3 was used, a technique that facilitates skin tone reproduction in post-production color grading. See Table 4 .

[Figure 3: see original paper] Documentary *The Last Desert Watchers*

In desert-themed documentary production, due to limited air clarity, color grading darker portions presents significant challenges. Right-side exposure process-

ing can be used for color grading to preserve more highlight details. During production, recording brighter ranges is preferred as brighter portions offer superior signal-to-noise ratios for effective noise control, achieving better color grading results.

For example, the production of *The Last Desert Watchers* employed a post-production DIT data format interaction workflow to achieve effective big data utilization. The documentary comprised diverse file types including XAVC-S with Log MP4 files, XAVC with Log MXF files, totaling approximately 13TB and over 100 hours of footage. File types also included ProRes422HQ MOV files from time-lapse photography and MOV/MP4 files compressed with H.264. Premiere was used for DIT file ingest in the editing process, supporting multiple editing methods and eliminating extensive transcoding time, thereby improving documentary editing efficiency.

In Link mode operation, Premiere presentation was used, followed by DaVinci Resolve round-trip color grading to facilitate production and editing. From pre-production shooting through post-production, file adjustment, and interactive delivery, DIT workflow concepts were implemented throughout.

Currently, big data technology has penetrated all areas of production and life, providing new technologies, methods, and support for industry development. Applying big data technology in documentary production can effectively enhance presentation layering and promote formal expression in documentaries. Documentary shooting requires establishing rigorous DIT workflows to ensure well-organized and secure footage, optimizing data backup through automatic verification via background programs. After material collection, XML tables are exported based on daily storage timelines, using big data and related technologies for diverse image expression to enhance presentation layering and optimize documentary content presentation.

## References

[1] Tong Sumei. Discussion on Computer Information Processing Technology in the "Big Data" Era [J]. Computer Programming Skills & Maintenance, 2017(22): 48-50.

[2] Zhang Xingjian. Interpreting Big Data Research Progress and Trends from a Database Perspective [J]. Computer Programming Skills & Maintenance, 2017(3): 43.

[3] Ren Ning. Exploring Mining Technology for Autonomous Learning Models Based on Educational Big Data [J]. Computer Programming Skills & Maintenance, 2017(11): 51-53.

[4] Tang Hong. Discussion on Writing Techniques for Online News in the Big Data Era [J]. Research on Transmission Competitiveness, 2017(7): 138.

[5] Zhao Huaiyan. Analysis of Documentary "Pre-sale" —Key Points for Suc-

cessful "Pre-sale" [C]. The 35th Annual Meeting of the Chinese Meteorological Society S14 Innovation and Reform of Meteorological Services in the Era of Big Data, Internet, and Integrated Media—The 8th Meteorological Service Development Forum, 2018.

[6] Ma Weiping, Li Siqi. Innovation in Traditional Media News Reporting in the Big Data Era—Taking CCTV' s "Data Says" Series as an Example [J]. TV Research, 2017(5).

[7] Hu Dongyan. Analysis of the Uniqueness of Theme Expression in CCTV Cultural Documentaries in the New Era—Taking *Tea: The Story of a Leaf* as an Example [J]. Research on Transmission Competitiveness, 2017(9): 81-82.

[8] Wang Jinhuan, Li Yan, Teng Minghui. Research on Rapid Classification of Metro Information Service Requirements Based on Big Data [J]. Computer Programming Skills & Maintenance, 2018(3).

[9] Gao Qi. How to Use Big Data Technology to Promote Radio and Television Monitoring and Supervision Work [J]. Science and Technology Innovation Herald, 2017(15): 180-181.

[10] Gao Hao, Huang Hairong. Analysis of Innovation in Documentary Content Production in the Big Data Era—Taking NHK Documentary *Disaster Big Data* as an Example [J]. China Television, 2018, 387(5): 78-81.

(Author Affiliation: Jilin Radio and Television Station)

[Figure 4: see original paper] Pre-coloring Comparison Before and After

*Note: Figure translations are in progress. See original paper for figures.*