From Tradition to Emerging Practice: A Hybrid Computational Production Model for Interactive Documentary

Insook Choi

Emerging Media Program, Department of Entertainment Technology
New York City College of Technology of the City University of New York
300 Jay Street V411, Brooklyn, NY 11201, USA
Phone: (1) 718-260-5588; Fax: (1) 718-260-5591
ichoi@CityTech.CUNY.edu

Abstract. This paper introduces Interactive Documentary as a new production paradigm. Research objectives are 1) to engage documentary practice with emerging media technologies in an open data space; 2) to prototype tools to facilitate cyclical authoring among a contributors' community. Interactive Documentary is here defined as a media production model with interactive author functions for constructing a narrative voice as a document of reality itself. The proposed system is envisioned for applications beyond motion picture genres, supporting the creation of living documents used in educational and collaborative project settings. Three types of users are anticipated: Authors, Contributors, and Observers. A formalization of traditional documentary production provides workflow analysis and modeling. The process of authoring is discussed in detail as well as system requirements and design specifications. A computational architecture hosts an author function; documentary authoring is implemented as query-display-edit facilitated by an inference engine. The system supports real-time *enactive query* across heterogeneous media resources, parallel media signal processing, and multi screen presentation and display formats. A use case takes upon a regional architectural history as an example of data design with domain sensitive ontology engineering. Ontology supports heterogeneous cross-referential capacity; its structure is reflected in a GUI designed to facilitate concept-based navigation across 2D graphics, 3D models, video, and audio resources. A dual-root-node data design links ontological reasoning with metadata, which provides a method for defining hybrid semantic-quantitative relationships.

Keywords: documentary production model, interactive documentary, ontology, multimedia authoring, narrative structure, GUI, enactive authoring, interaction design

1 Introduction

Documentary in film practice is an established genre. Yet its definition often undertakes a discursive path. Two factors play consistently in various definitions: 1) reality is captured in some forms of documents and 2) the documents are subjected to assemblage to serve a larger context. This paper introduces a production model of documentary practice with computational processes to support interactive authoring. For the definition of documentary we adopt the simplest task definition, that of Vertov: “to capture fragments of reality and combine them meaningfully” [1], p.55. Grierson’s “the creative treatment of actuality” [1], p.287, is at the heart of documentary practice for both spectators and authors. The interactive documentary system model presented here is designed to be used for novice as well as expert authoring. In this spirit we implement *enactive query* to reduce the gap of traditional polarities between production and reproduction, between media producers and consumers, and between authoring and acts of inquiry.

The construction of the documentary subject utilizes narrative devices accompanied by information sources such as first-person accounts to anchor the devices in factual circumstances. A documentary synthesizes a narrative about factual objects and events including original documents and reconstructed representations. More significantly the narrative resides in the rules of play engaging human memories in observations and experiences. The rules of play provide a set of criteria how to frame subject matter with camerawork, sound recording, and editing, and are a part of an *author function* [2] for constructing a vehicle to carry narrative voices.

1.1 **User-Centered Definition**

Interactive Documentary is here defined as a documentary production model with interactive author functions for constructing a narrative voice as a document of reality itself. The proposed methodology is envisioned for applications beyond motion picture genres, supporting the creation of living documents used in educational and collaborative project settings. It anticipates scalable dissemination and participation through diverse user communities and media devices in an open data space. Its narrative structure involves real-time performance of structured database query, an extensible range of heterogeneous media resources, parallel signal processing of media that use diverse display subsystems, and interactive experiences for novice users. Current development is at the phase of prototyping and alpha testing a working system. User tests with novice participants are forthcoming. Current design criteria are based upon practitioners’ perspectives, developed through consultations with an Oscar[1]-nominated producer-director [3], and incorporating the author’s insights gained as a composer of virtual reality performances and installations [4].

The present Interactive Documentary system anticipates three types of users: (1) an Author who creates an original interactive documentary path, and may provide semantic classifications and related media resources; (2) an Observer who interacts with a given documentary path within a predetermined range of exploration, choosing the degree and timing of exploration when traversing a given path; (3) a Contributor who can modify a given path to create variations or enter new regions. Media resources are displayed in multiple 2D and 3D image frames (see Figure 1), not unlike the display used in Soft Cinema [5] but with extended types of media resources such as real-time generated graphics and spatially-modeled sounds.

[1] Annual award ceremony of the Academy of Motion Picture Arts and Sciences

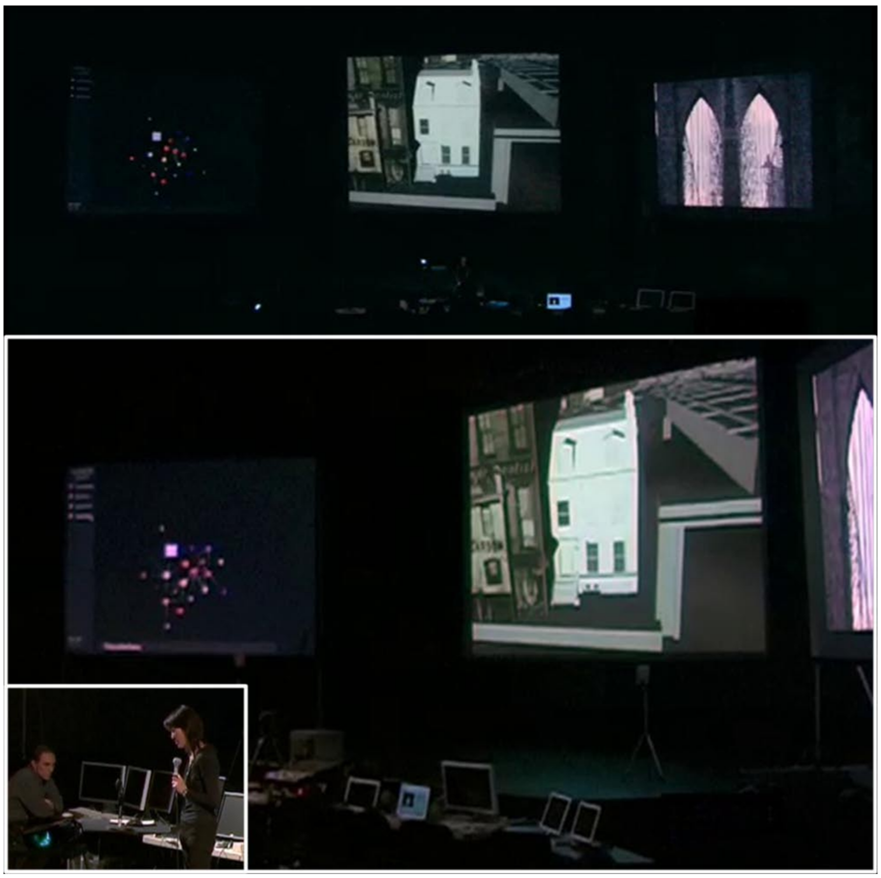

Figure 1: Front view (top) and side view (below) of the *idBrooklyn* [REF] Interactive Documentary presentation, and presenter (inset). The center screen displays a 3D scene; the right screen displays video and still images. The GUI on the left screen is also displayed on the presenter's kiosk.

Interactive Documentary facilitates the general practice of documentary as described above with designs to establish an author function with computational practice. This relationship is explicated in section 3. For grounding perspectives in computation, an author of interactive documentary prepares a structured vocabulary to serve a semantic network compatible with media resources. The semantic network can be adopted as a literal representation of the network of media perspectives serving narrative possibilities. Ontology design to support this is discussed in section 6.

This paper presents ontology applied as an organizing tool for concept based documents storage, retrieval, and assemblage, to assist authoring interactive experience and narrative voice. Ontology processing not only provides an extensible use and cross referential organization of documents of multiple type as media resources (the term "documents" and "media resources" are interchangeable hereafter), it also facilitates a partnership among author functions, some carried out by an author, some by a computer, and some by a community of respondents. This extensibility facilitated by ontology is rather promising for developing a new production paradigm. For exploring the new paradigm the paper reviews groundwork and related research.

1.2 Order of This Paper

In section 2 the relevant previous works will be discussed. Section 3 introduces authoring processes applied in an interactive presentation system in an installation setting, for testing the case studies. The configuration incorporates multiple visual displays, synthesized audio, and a mouse-based interface. The media resources in the prototype include 3D graphic models and scenes, simulations, and data-driven and procedural auditory and visual processing of resources. A graphical user interface (GUI) design for concept-based navigation enables queries across heterogeneous media resources. A methodology is introduced as path-planning using the GUI for semantic query-based authoring. A use case is presented focusing on the design of ontological data and the use of interactive reasoning to retrieve media resources in real-time.

Section 4 is devoted to workflows and production models in documentary practice. It introduces the traditional practice and summarizes the assembly-line paradigm of the workflow, then develops formalizations of a traditional production model and an interactive documentary production model. Section 5 discusses system architecture. The architecture is designed to host computational processes for supporting interactive signal processing with an inference engine in the loop.

Section 6 is devoted to ontological data design. We apply ontologies for designing flexible hierarchical and associative structures for domain sensitive semantic data. We also apply them to enable a common layer for semantic planning and on-the-fly selection across multiple media types. A dual-root-node data design linking ontological reasoning with metadata is presented as a method for defining hybrid relationships of semantic data and quantitative data.

2 Related Work

Interactive Documentary combines elements of generative storytelling systems with an interaction framework related to enactive interfaces for virtual media performances and installations [6, 7]. This application of semantic networks came about while developing generative models for enactive media interfaces [8]. "Vox Populi" [9] has a number of similarities to the interactive documentary concept in terms of functional goals and semantic processing stages. One common goal is to provide production-level access to documentary material for community exploration and reconfiguration beyond the constraints of pre-determined linear presentations. Another similarity is the role of an initial design phase to specify a semantic context that will result in a graph structure of semantic units. We refer to this as “pre-authoring,” discussed below. Bocconi describes upfront semantic annotation of media resources and we have a similar phase, however we use a grouping process to relate individual media resources to concepts, rather than applying annotations to each media resource. Vox Populi uses a thesaurus of rhetorical constructs to determine the point of view expressed in spoken text, whereas our structured vocabulary of concepts describes contents of media resources (“traffic downtown”) and real-world objects depicted in media resources (“Tillary House”). Separately we include concepts describing signal processing modes for presenting and transitioning resources in multiple display streams (((image1 ∩ image1) ∪ cross-dissolve) ∪ (music ∩ crescendo)). Both projects generate data-driven media sequences by navigating a graph structure; our prototype operates in real-time with navigation controlled by an observer. Vox Populi is highly specific for rhetorical relationships in video sequences, whereas we utilize a *display grammar* [10] that generates sequential and parallel displays of resources in real-time. Display grammar implementation is beyond the scope of this paper; however the role of display grammar in the interface is put in context in sections 3.3 and 5.

Hardman’s *canonical processes* [11] are closely related to aspects of our production model, and this paper includes references in several sections. The primary difference in the current work is the focus on interactivity and the possibility using previous documentary versions as resources in future versions, a use case that invokes the annotations of annotations, mentioned in Hardman.

Manovich [12] introduces many useful concepts concerning database as a source of authoring media productions. A notable difference in the current work is the focus on ontology rather than relational data and its relation to the history of cinema and 20th century theories based on structuralism and deconstruction. The present work introduces interaction with authoring process and participants’ interactions as observers and contributors. Soft Cinema [5] automatically generates endless variety, thus a viewer cannot experience this variety. By making choices, exploring alternatives and moving forward in a path a user can experience the significance of variety and alternative perspectives.

Terminal Time [13] offers a novel approach with the use of ontology to interactively schedule narrative sequences from diverse media resources. The primary difference in that work is its focus on idealized classical documentary forms, and its design for hyperbole. Parkes [14] presents an excellent early model for analysis and storage of existing film works. While this is based upon analysis rather than new production, the article makes a helpful clarification of “interactive video” by demonstrating the subject of interactivity is user’s experience not technology. Brooks [15] demonstrates an additive agent process using formulaic agents and explicit modes of storytelling. The present work does not focus on documentaries as necessarily closed-form stories.

Davis' Media Streams [16] represents a semantic oversaturation applied to commercial media tools that use a timeline paradigm such as Avid™ [2] and ProTools™ [3]. Users can select from 6000 concepts for media authoring. Our approach uses a semantic path to replace the multi-streamed timelines; path linearity is a highly efficient combined display of both sequential order and time. An alternative to 6000 icons is the design of semantic neighborhoods requiring minimal GUI iconography; an author is likely working at any moment in a locally-constructed neighborhood of no more than a few dozen icons.

"Evolving Documentary" [17] and "Multi-Threaded Stories" [18] propose models for documentary generation in the context of journalism, where a story is presented in broadcast format and modified in subsequent presentations as situations evolve and additional media resources are available. The concept of evolution is similar to our model in terms of a user model for receiving an authored work and creating alternate semantic "paths" to customize and extend the work, as well as for adding new resources or new semantic units. "Very Distributed Storytelling" [19] sketches an idea of widely embedded interactive media governed by a central narrative production system. The installation-based implementations are reminiscent of this author's work from the same time period, such as an interactive audio installation that applies motion-sensing and pattern-recognition to process observers' movements in a gallery, in order to generate virtual paths through a semantic network representing historical voice recordings and real-time sound synthesis and signal processing [20].

3. Author Function and Prototype for Interactive Documentary

The prototype system combines a capacity for (1) *authoring interactive-media* with (2) an *interactive media-authoring* process. The first refers to authoring procedural media processing with interaction design. The second refers to an authoring process supported by real-time media processing to display results of the procedures. The objective is to support both capacities in close proximity with proper system architecture and system performance. Linguistic practice built in an author function is explicated through symbolic computation, which in turn facilitates an enactive narrative function. In Interactive Documentary we refer to this synergetic function as *Enactive Authoring.*

We may assume an author predates his reader. However an author function is more of a mutual understanding where a reader and author share a consensus as to a process of designating sources of meaning and accountability. An author function therefore is not limited to a unidirectional flow of ideas and evidence; a reciprocal act is contributed from a reader enabled to look beyond the text and assimilate designated meanings by recognizing the author's acts of making references. In the interactive media case an observer's reciprocal participation becomes literal, through engagement with a computational process where alternative "texts" may be generated in a single work. This engagement in turn transforms consensus through a computationally-mediated interactive

[2] http://www.avid.com/
[3] http://www.digidesign.com/

author function. We address the function of interactivity by examining enactive interfaces [21] that are invested in the processes of authoring, presenting, and observing interactive media.

Luciani et al. have investigated the concept of "Enaction" ("Enactive Interface") in the domain of interfaces for creative production, "in which high level media of communication and of cultural data are produced by means of closed-loop sensori-motor interaction" [21], p.8. The adjective "interactive" cannot be assumed without a sensorimotor design. When a system enables interactive authoring and navigation with real-time multi-sensory feedback, the authoring process is similar in many ways to a presentation process. This places authors and observers in a role akin to a presenter or a performer who is crafting a presentation, rehearsing, or performing. Enactive observation and authoring is an integral part in our model in order to facilitate the interplay between production and consumption. Our cognitively inspired model configures a closed loop from asserted relationships to inferred retrieval linking concepts and the display of media resources. The act of retrieval at the user interface is conducted by movement in a continuous and ordered geometric space. This closed loop of movement in an ordered control space to retrieve and display media provides sensorimotor feedback consistent with enactive interfaces.

While the concept of enaction has been previously applied with an emphasis on musical performance, we identify a suitable extension to any multi-dimensional movement space that generates sensory signals in real-time. The distinction of "discrete" and "continuous" control is only a matter of granularity, as all digital signals are discrete while all performers' movements are continuous. The visual and physical area of our graphical interface is articulated by nodes representing concepts while the intervals between nodes configure a continuous and ordered inquiry space. Semantic distance is represented by the degree of separation of nodes and the GUI area limits the displayable range of nodes. Users can estimate and learn the semantic range of a set of nodes with respect to movement through the navigation space. A fluid navigation movement includes the exploration of nodes to reveal concept names as well as the selection of nodes to generate queries to retrieve, schedule and display media resources.

Following are seven interface design imperatives to support enactive authoring in a semantic space ordered by concept distribution.

1. Specify a set of requirements to support enactive authoring with real-time query as follows.
    a. A graph structure of concepts represents media resources.
    b. An authoring process enables the creation of paths by making sequences of nodes.
    c. Node adjacencies in paths are not limited to the node adjacencies of the graph structure.
2. Visualize the concept graph structure in the form of a graphical user interface, such that interacting with the concept graph will generate queries of concepts.

3. Enable queries to automatically generate media display events without further action required at the interface. When a concept node is queried, the media resources returned by the query will be automatically scheduled for display.
4. Enable exploration of concept names and their structured relationships in the concept graph, independent of initiating queries. This provides browsing and exploration in advance of committing to a query-media display event.
5. Present a separate visualization of paths of nodes authored by previous users.
6. Enable a user to modify existing paths of nodes or to create a new path by selecting and storing a series of concept nodes
7. Enable a user to define new concepts by applying logical expressions to existing concepts; and storing the new concept in the vocabulary for future users.

These design imperatives envision enaction to support an engagement of an author function as in storytelling practiced as an oral tradition. In oral traditions of storytelling, the gap between author and performer is brought into a close proximity. At the same time the integrity of authoring is well preserved in ways that are distinct from mere improvisation. Building on this paradigm, we foresee multiple elaborations and preparatory stages for a full production process and interface are to be defined. Aspects for further development are differentiated by multiple use cases. Section 3.1 presents these use cases.

3.1 Interactive Documentary Use Cases

Four use cases for interactive documentary are Pre-Authoring, Authoring, Observer, and Contributor. Close correspondences are identifiable with Hardman's canonical processes, mapped to the enactive interface paradigm [11]. Pre-Authoring involves premeditation [11], planning, and acquiring media resources by capture or collect. Pre-Authoring tools include media resources libraries and structured vocabulary of relationships between ontology concepts. Pre-Authoring requires (1) specification of relationships between concepts; (2) the creation of new concepts to express new relationships between existing concepts. Each acquired media resource is entered into the dataset by assertion as a member of one or more concepts. Resources already in the dataset may be associated with the new concept vocabulary. Pre-Authoring is a very time consuming process but can lead to efficiencies in later stages, particularly as the use case approaches real-time interaction.

Authoring with this system involves iterative exploration of concepts and arrangement of resources to convey an intended message to anticipated observers. The primary authoring tools are the concept graph GUI, related concept associations, and paths of concepts. Authoring workflow follows from Pre-Authoring with the creation of a Project that identifies the working vocabulary and resource dataset. Concept exploration and path-making are the primary work cycles, including designation of concepts for display signal processing, and organization of paths of paths. Testing and refinement of paths includes designating regions of access from path nodes to neighboring concepts, specifying Author-only regions, Author- Contributor regions, and general Observer regions.

Observers' primary tools are path-following, node mouse-over, and node selection. Time-accurate recording and reproductions of path traversal enables observers to assemble rudimentary presentations or records of their query paths. Additional UI features are under consideration, for example personal annotations and user-defined collections of nodes. These require future design and testing.

A Contributor is an Observer with the ability to make and modify paths and store them as new paths in the context of an Author-defined Project. This capacity opens up design consideration for Contributors to add resources, and create or modify concepts. User testing is required to determine criteria for identifying levels of expert and novice users.

3.2 Enactive User Interface Model

The path-making process is supported by a GUI and a media scheduling engine that displays the results of each query as a real-time mixture of media resources. The ontological graph is visualized in the GUI as a 2D network of nodes and edges. Figure 2 shows the GUI as a collection of dynamic nodes; visible nodes represent a limited region of a much larger ontology. The current implementation presents a reduced set of representations to define baseline levels of interaction. Mouse-over a node displays its related concept name. Double-click on a node selects that node and initiates a query of the associated concept; the query in turn initiates a display of media resources related to that concept. In the GUI clicking on a node modifies the display to reveal all nodes that are nearest neighbors. Remotely-related nodes remain hidden until a close neighbor is selected. The user has some control over the range of nodes displayed surrounding a selected node.

The meaning of the connections between nodes is determined by the concepts the nodes represent, and by the relationships represented by the edges between nodes. The edges are visualizations of predicates defining concept relationships in the ontology. These include subclass-superclass relationships, and non-hierarchical relationships indicating two concepts function together to a common purpose, for example ((Post Office ∪ BoroHall) ∩ 1800To1880). Many nodes represent logical operations over multiple concepts. Ontology details are discussed in Section 6.

To support the enactive interface model, nodes are displayed with animated ball-and-spring dynamics, aiding visual identification of relationships. Size of node indicates number of links to a node. Color indicates subclass-superclass relationship. The "current location of the user" is defined as the most recently selected node, corresponding to the currently active query and associated media resources currently being displayed. The square node represents a concept that is the "anchor" for a node neighborhood. Anchor nodes are sequenced in a path of paths to designate distinct scenes or sections in the documentary.

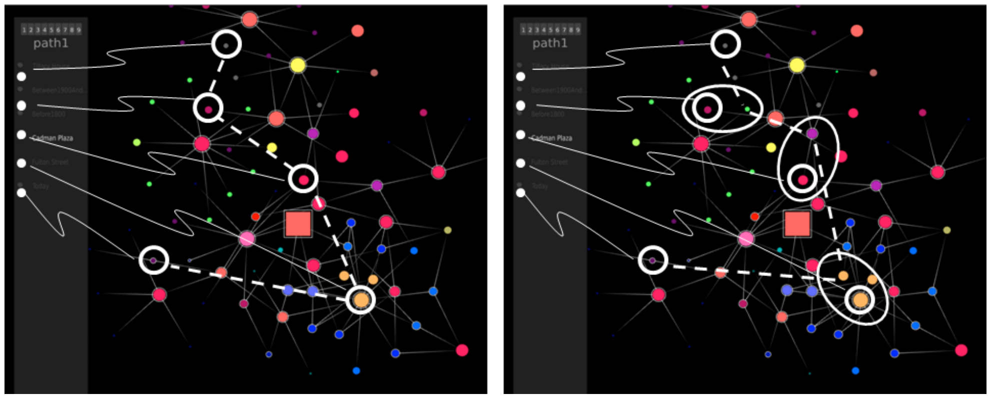

Figure 2: Graphical User Interface representing concepts as interactive nodes. A Path of nodes is presented in separate vertical array on the left. **2A (left)**: White circles show graph positions of each path node; dotted lines indicate non-adjacent links across semantic regions. **2B (right):** A second *narrative instance* of the path utilizing alternate nodes in semantic regions of the original path nodes. Larger circles indicate pairs of original and alternate path nodes.

3.3 Path-planning and Interactive Authoring

The idea of making paths through a digital document space can be traced to multiple sources, including the Memex technology proposed by Vannevar Bush in the late 1940’s [23]. These proposals focus on “trails of documents” using text processing for cross referencing and indexing to achieve more efficient storage and retrieval. The present prototype differs from “trails of documents” proposals by implementing *paths of queries;* paths through concept space generating queries as acts of creative inquiry, generating real-time sequences of composite displays of resources, functioning both as dynamic media content and as semantic navigation feedback to the user.

Figure 2 represents the current prototype interface for query paths. A vertical array on the left of the GUI is a sequence of concept nodes arranged as a path. The path can be traversed in series order or explored out of order. In the main GUI a selected path member displays a concept neighborhood that can be expanded for exploration. Clicking on path nodes generates the queries selected by the path author. Clicking on a non-path node returns a related set of media resources and expands the graph visualization to reveal the concept neighborhood of the selected node. Exploring neighboring nodes produces further queries returning related resources. Limits of neighborhoods can be specified in authoring.

A special set of nodes dedicated to display signal processing is available only to authors. These nodes represent Display Grammar operations [10] (Section 5), and are combined in logical expressions with nodes representing media resources. When a user generates a query at the GUI, the corresponding Display Grammar nodes effect transitions in the media displays. Scheduling constraints impose a minimum duration between queries to allow for display processing times.

We advance the principle of paths of queries not only for novel creative acts but as the normative use case for enactive experience of Interactive Documentary. Novices as well as experts are can perform path exploration. An author may enable participants to introduce path variations, to investigate alternatives that can result in multiple versions of an initial path. These alternatives are referred to as *narrative instances* (see section 3.6).

3.4 Interactive Documentary Production Case Study

This section presents the production of an example Interactive Documentary. As a source for documentary subjects we identify present-day and historical Brooklyn to provide deep pools of media resources suitable for generating multiple narratives [6]. A multi-screen and multi-channel display system supports the parallel presentation of media resources of multiple types. Figure 1 shows a large-format side-by-side layout of visual media, with an audio system channeling sounds from multiple sources. The GUI is situated at a small kiosk; in Figure 1 the kiosk screen is reproduced on the leftmost large screen. The laboratory image resolution is 1366 by 768 pixels for screen dimensions of 94.5 by 168 inches. Queries generated at the GUI return 2D images, videos, sounds and points of interest in 3D scenes. A display synthesis subsystem responds by scheduling and mixing sounds, images, videos, and virtual camera movements in 3D scenes. The sounds and images are resources in the ontological data set; the virtual camera movements are determined by points of interest in 3D scenes that are also ontological data resources.

Path-planning and direct data query are used for scheduling media display synthesis, supplanting both linear and nonlinear editing. This is consistent with Manovich's concept of database cinema [12]. The difference is that our model uses ontological data design rather than relational data, and incorporates interactivity. Query paths and paths of paths with nonlinear traversal provide ordered progressions of subjects through concept relationships. From an individual query, media resources are returned unordered in sequences of sets. Ordering is part of a Display Grammar computational model. Scheduling of video and audio segments is determined by evaluating segment metadata as well as concepts associated to each resource. Unlike traditional practice, scheduling and timing are not fixed by frame count; they are relative to the pace of interactive queries and to the parameters that define the display synthesis for detailing tempo and transition mode. Evaluations of each resource are performed by the computational model linking ontological inference to display synthesis. This computational model determines relationships between concepts, metadata and display parameters. It enables the display engine to prioritize resources and generate sequences by evaluating resource membership in concepts such as "A-Roll," "B-Roll," "EstablishingSequence," and "BrooklynIntro," described is section 3.5.

The Tillary House is a well-defined theme in the *idBrooklyn* project, a corpus of information surrounding Dr. James Tillary, the immigrant Scottish doctor who built a landmark row house in 1813 and relocated his family from Manhattan to Brooklyn. The house was documented by the WPA in the 1930's before being razed for mid-20th century urban development. In Pre-authoring we designed an ontological concept graph to create set relationships between general architecture concepts and specific historical information of Brooklyn. (See section 6.2: predefined and User-defined concepts.) Pre-authoring involves studying background materials, historical records, and media, and developing a structured vocabulary of concepts. Hardman refers to this as Premeditation [11]; the industry term is "pre-production" (see section 5). Beginning with urban and historical concepts used in an established thesaurus (the AAT: section 6.2), additional concepts were defined to designate Brooklyn locations, landmarks, and subjects such as Dr. Tillary, Tillary Street, and the Tillary house.

3.5 Example Query Path Sequence

In an example query path the first node is the concept "FultonStreet2000toPresent;" this node returns photographs of storefronts, pedestrians, and street vendors, and sounds of voices and bus traffic recorded recently on Fulton Street, a primary pedestrian mall and shopping district. The photographs are choreographed in a moving array across the large screen. In parallel a separate screen displays a 3D scene of downtown Brooklyn, and the query returns a 3D camera movement slowly "flying" along virtual Fulton Street. While these resources are displayed, selecting the second path node "BoroHall2000toPresent" introduces new photos and sounds, with smooth visual and audio cross-fades effecting the transition. The 3D camera movement interpolates from Fulton Street to a new position hovering above the model of Borough Hall. A subset of images and sounds may be common to successive queries; common resources across a transition are brought forward in the display to maximize continuity. This type of prioritization is a parameterized property of display synthesis encoded in the computational model.

The first two path nodes represent queries that intersect with the concept "B-Roll," a term used in film and video production to indicate footage that does not depict primary dramatic content or dialogue. In our structured vocabulary A-Roll indicates material containing direct speaking subjects such as interviews, and B-Roll indicates material that does not. The intended result of the first two queries is an introduction to downtown Brooklyn establishing a general setting without spoken topical information. The third path node "TillaryIntro" intersects the concepts "A-Roll" and "EstablishingSequence." This query returns a set of video segments depicting an interview of an architectural historian at key locations in Brooklyn, corresponding to locations in the 3D virtual scene. The display synthesis engine sorts the video segments according to metadata that indicates a progressive order of shot locations, starting on the Brooklyn Bridge and continuing across Cadman Park to arrive at Tillary Street. During each interview segment the query returns images of historical documents that are related to the concepts discussed by the architect. These images are displayed in parallel as their subjects are mentioned in the interview. For example the fourth path node "FultonStreet1880to1920," sends the virtual camera in the 3D scene to resume its previous flyover of Fulton Street. Determined by the historical date in the query the 3D scene now includes a model of the Brooklyn elevated train. The model was created from photographs made in the early 20th century. At the same time the train appears, models of postwar buildings are removed from the 3D scene. The collage of contemporary photographs of hip-hop shops and cell phone vendors is replaced by historical drawings, lithographs, and photos of the elevated train. Sounds captured live on Fulton Street are replaced by sounds from a sound effect library: horses, carriages, a steam engine, and pedestrians on a boardwalk.

3.6 Narrative Instances

A story can be told in more than one way; varying details brought to light, or passed over. Variations on a narrative indicate the robustness of its design. A query path may have neighborhoods of related nodes that generate alternative levels of details while preserving a central narrative. Enactive narrative explores adjacent nodes and neighborhoods surrounding path nodes, telling one or another variation of a story. These alternatives constitute a presenter's perspective in associative aspects. Enactive recall is a retelling, performed every time a path is traversed. To traverse a path of queries is to recall the narrative that binds them and entails a performative quality shared by path authors and observers retracing paths.

Our system architecture and graphical interface are designed to facilitate enaction to extend and vary narrative experience. Paths represent leaps of narrative association to nonadjacent regions on the concept graph. A path node represents a semantic region; within a region there may be multiple nodes that support associative coherence. Multiple *narrative instances* are generated when neighboring concepts are engaged creating variations in a query path, either by enhancing or replacing an original path node. Figure 2A shows the nonadjacent graph locations of nodes that constitute narrative associations in a query path. Figure 2B represents an alternate narrative instance of the query path presented in Figure 2A.

4 Modeling Documentary Workflow

To illustrate Interactive Documentary as an emerging genre this section provides a systematic comparison to traditional documentary production. Traditionally film and television have been the main venues for documentary practice. With the venues comes a particular set of conventions. While convention reflects a broader scope of the genre including associated aesthetics and styles, the convention is, in large part, shaped by the medium it adopts. The medium then shapes the overall workflow, which may be formalized as a production model. A convention is not merely an apparatus or a work process: *a convention includes an observer's understanding of a media resource as a product of an apparatus and work process.* The implicit apparatus is critical in the recognition of the medium. The conventional understanding of documentary production involves a three-phase workflow: pre-production, production, and post-production. With this general workflow the lines of production chains are put in place to ensure the decision making hierarchy. In practice documentary production is far more cyclic and nonlinear.

Many tasks are broken down into narrow fields of specialties with highly defined and limited sets of operations executable in cycles. Yet the established convention may well be characterized as a media assembly-line where sequences of tasks are repeated to assemble components from multiple suppliers that fit narrow specifications of form and function. Linearity as an underlying structure in production workflow is also reflected in Hardman's canonical processes [11]: workflow begins with a seed concept and moves from scripting and shooting to editing and distribution.

Figure 3 provides a schematic of the idealized three-phase workflow. Pre-production is a phase of research and development largely related to ideation and visualizations, akin to Hardman's Premeditate process. Many representational materials are produced at this stage to assist envisioning the work and these materials are largely text or sketch-based. Documentary production features the recording of natural, unrehearsed, unscripted "fragments of reality" for which the faithful attitudes in the recording process may differ from the ones in fictional scenes or dramatic performances. As in Hardmans' Capture process, documentary production incorporates pre-existing media resources and synthesized media resources, which often require further treatment to format for a compatible integration.

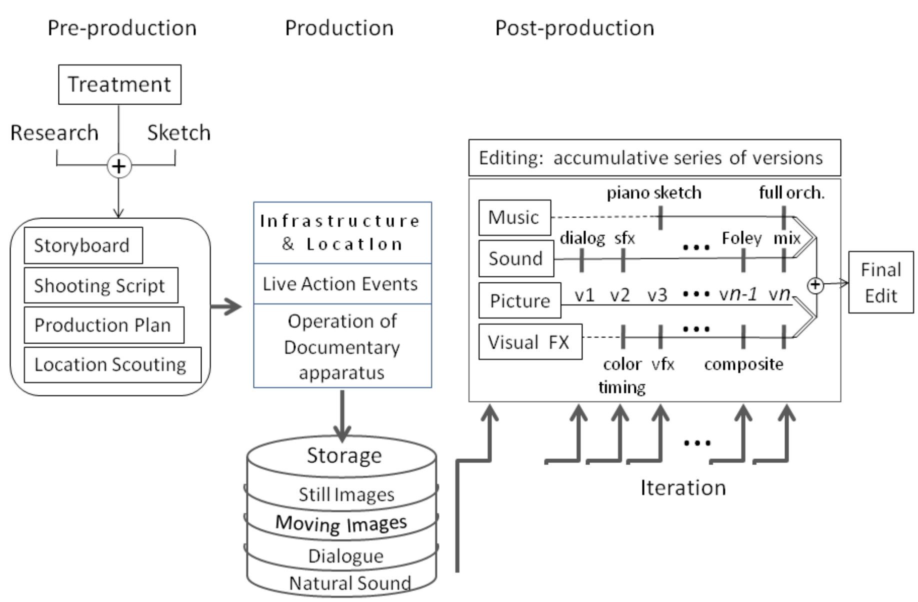

Figure 3: Idealized Traditional Media Production Workflow.

Rabinger [23] details state of the art techniques for assessment of “dailies”—media resources generated in production, and linear and nonlinear editing methodologies. Editing is akin to a process of testing a hypothesis advanced in Pre-production; flaws in an original hypothesis are encountered when synthesizing material compositions. One of the greatest challenges in Post-production is organizing media resources with adequate references to identify related features distributed across multiple resources and resource types. The practical concern is the timely efficiency of finding a necessary resource among a plethora of “raw” materials. This challenge is heightened in documentary by the multiplicity of sources and formats and the potential variety of unscripted contents.

4.1 From Workflow to Production Model

The workflow analysis is applied to elicit a production model. The model defines relationships of media processing entities as abstractions of the sequential tasks. The relevance of Hardman’s canonical processes is noted. Care is taken to model iterative relationships and transmission of data that embodies knowledge for decision-making. Figure 4 illustrates a production model of traditional workflow in Figure 3. The production model formalizes a cyclic process as opposed to a linear workflow. Pre-production is a phase of research and ideation where visions are selectively audited through sketches mostly in text and graphical forms. Production and Post-production are the phases of iterative processes for gathering and assessing media resources. Screening is a main method for assessment through daily production and plays an important role in assessments of daily results and edited sequences, determining further materials needed and methods for acquiring the materials.

Editorial decisions such as narrative order and scene juxtaposition are deferred to Post-production. The end results in documentary film making may deviate significantly from the original scripts and storyboard. This often requires reformulating presentation rationale. Figure 4 illustrates a potential retroactive connection from assessment to revision of storyboard and shooting script. It is noteworthy this retroactively adaptive function brings the editing process closer to the Pre-production phase. The criteria consulted for retroactive adaptation include the optimization of media resources generated in Production. This process is illustrated by the grey region in Figure 4 where iterative assessment is shared between Production and Post-production. Hardman’s Query, Message Construction, and Organise processes intersect in this grey region.

4.2 Interactive Documentary Production Model

Figure 5 depicts modifications of the traditional documentary production model introduced by interactive media. Human crafted tasks in the editing suit are enhanced or replaced by automated assembly and selection, while some live-action-oriented production tasks are enhanced or replaced by the use of existing media resources and third-party media streams. Computational models support parametric modeling of automated processes; these models may be designed to represent the state of a user and ingest data of a user's actions to generate customized results.

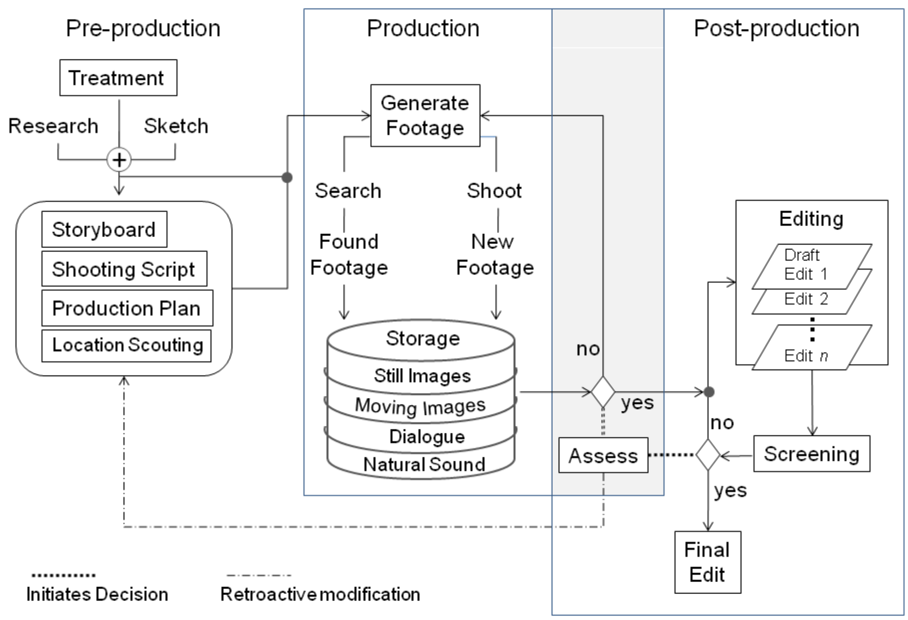

Figure 4: Traditional Documentary Production Model

Structured queries to evaluate metadata and semantic data can aid the automatic selection of source materials during an edit cycle. Screening of edited material has been virtualized; we anticipate the technique of non-destructive editing will someday entirely replace the concept of the "final cut." Virtual edits can be extended to enable entirely synthesized reproduction of edit decisions, executed as instructions applied to source materials in real-time. Figure 5 represents the virtual edit function as *display synthesis.* The process of editing is under transformation toward *authoring*, the creation of instruction sets in scripts and code to execute edits and other display functions; authoring can anticipate live data from an observer steering editing and display processes. The final result of Post-production is no longer an immutable media document. Instead the result is a program file of media data accompanied by instructions, executable as a media application. The program incorporates media resources interleaved with metadata and control data. An industry standard protocol such as MPEG4 [24] enables the program to be interpreted by a third-party computational engine that synthesizes the results on a recipient's media device.

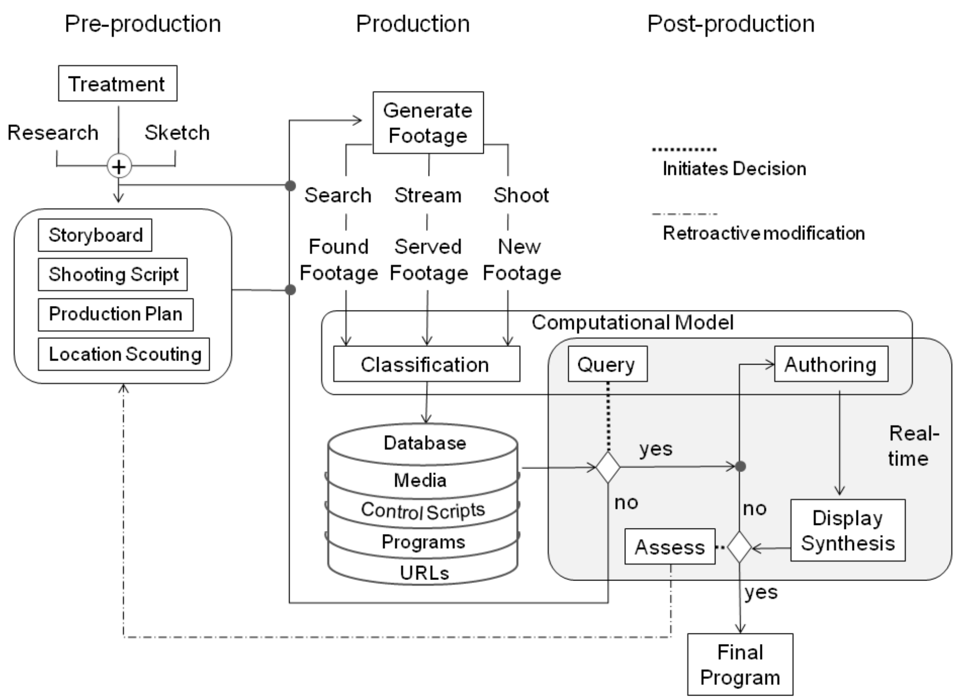

Figure 5: Interactive Documentary Production Model

Figure 5 illustrates shared functionality between a computational modeling subsystem and a real-time signal processing subsystem. Computational modeling standardizes ways to identify and process media resources. Real-time signal processing enables a user to govern the cycle of query-edit-display as user centered workflow for rapid prototyping. In terms of Hardman's canonical processes, Capture and Archive correspond to Generate Footage in figure 5; Annotate corresponds to Classification in Figure 5; Query, Message Construction, and Organise correspond to Query, Authoring, and Assess in Figure 5.

5. Computational Architecture for Interactive Documentary

Interactive Documentary is a production model for synthesizing a narrative voice by means of interactive recall for retrieving nonfiction media resources in an open data space. “Interactive recall” refers to the user experience of an author, an observer, or a contributor as the system responds to her actions in real-time. In each of these cases the user’s role is partly one of explorer and partly one of presenter or performer. “Nonfiction” refers to representations of real-world artifacts: media resources are digitized artifacts of objects and events. “Narrative voice” is constructed through composite interactions of system components and constituent media resources. Authoring in this context accounts for 1) the design of instruction sets integrating multiple types of media resources, 2) interaction design for an observer’s experience, 3) anticipated media devices observers can access, and 4) the optimization of user-generated data. These authoring requirements should be supported by a larger set of requirements for new system architecture. While our prototype architecture anticipates these authoring requirements the current version fully supports the first two requirements and only partially the latter two.

In a computational architecture for interactive documentary production, shared functionality is elaborated between a computational modeling subsystem and a real-time signal processing subsystem. These shared subsystems depicted in Figure 5 are detailed in Figure 6. The computational architecture consists of three layers of function modules relating to control flow and data flow with temporal differentiation. *Active Observation* modules support fine-grained scheduling for immediate action-feedback response in the GUI and media displays. *Working Memory* modules weigh relationships among concepts and media contents and provide heuristics for displaying these relationships. *Long-term Memory* modules represent deep structure and media assets supplying production capacity. Semantic and media assets are brought into relationships in the computational workflow.

System configuration, source material acquisition and concept design are prepared prior to real-time interaction. Authors may contribute to an existing concept vocabulary with respect to the documentary subject while using the concepts to classify media resources for an ontological repository. Media resources are asserted as members of concepts according to their attributes and according to the objects they depict. An asserted *describable object* may be a media resource or an object depicted as an entity in a resource (see Figure 6). A concept graph is designed to articulate relationships relevant to the inter-textual subjects in documentary. A set of states is preconfigured in the computational model to respond to a designed concept vocabulary and an initial graph structure to produce interactive documentary experience. See section 6 for ontological data structure and interactive query methodology integrated in this architecture.

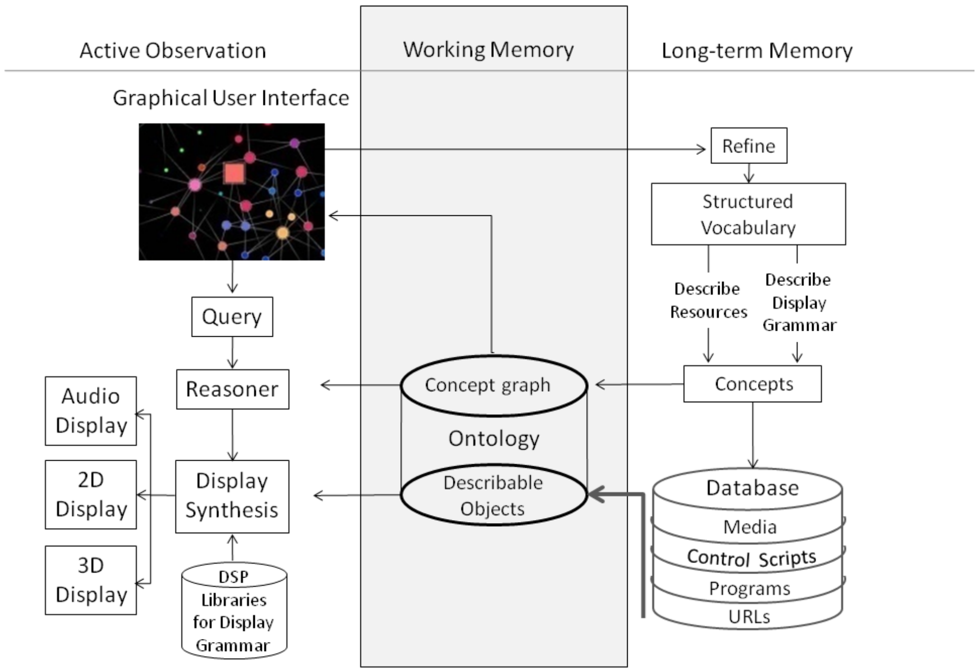

Figure 6: Computational Architecture for Interactive Documentary Production

While most concepts describe content relationships, a special class of concepts describes Display Grammar relationships. These correspond to signal processing display functions housed in the Display Synthesis subsystem, similar to media-specific processing functions used in Media Streams [16]. Expressions such as $((Content_m \cup DspGram_n) \cap (Content_p \cup DspGram_q))$ are resolved in the reasoner, returning media resources to the Display Synthesis subsystem where digital signal processing (DSP) is applied as determined by Display Grammar expressions. Display Grammar nodes are not accessible by Observers and are provided in default states to simplify exploration and novice path making. Full discussion of Display Grammar is presented by Bargar [10].

5.1 Cognitively-inspired Architecture

For semantic organization and reasoning we have adopted an ontological data design. Ontological structure is defined as a set of logical expressions that may be interpreted as a directed graph of concept nodes, where edges represent relatedness of concepts. The authoring process is formalized as path planning through a semantic data structure: an ontology of heterogeneous media resources with a separate concept graph for Display Grammar. Structure is encoded in OWL file format [25] using an open source editing tool [26].

We describe the architecture as cognitively-inspired in reference to the system design criteria to facilitate concept formation, computational inference, and the observer's cycle of discovery, prediction, and path refinement. Von Foerster draws a distinction between the "invariance of quality" in storage and retrieval, and the synthesis of information in recognition and recall [27]. He regards perception, memory, and inference (prediction) as the requisite components of cognitive processes. The pre-authoring phase for composing a concept graph indicates a strong role for prediction toward interactive exploration and path-making. To design a concept graph is to predict as well as to premeditate; to imagine potential paths discoverable by an author and other users. To explore a concept graph is to predict as well as to discover. To traverse a path is to recall a sequence of predictions in the form of queries. While authors can predict and recall their own concept designs, observers experience an author's recall and are challenged to make predictions accordingly. This is consistent with the role of creative media: an audience considers the orientation of a media creator for insights of mutual consequence. Observers may not share the cognitive orientation of an author, however by path following and region exploration they may discover an author's point of view. Paths are coupled to the production architecture for observers' dispositions for learning through ontological inference. While computational inference supports media resources storage and retrieval, a user's explorations and path-making constitute recognition and recall.

5.2 Documentary Production in a Heterogeneous Media Ecosystem

Contemporary media content is often created as a convergence of multiple media types deployed on diverse devices. Display systems such as personal communications devices can differ considerably from the production systems where media resources originate. Also their roles can be reversed: personal devices can generate media in consumer formats that are repackaged and distributed by media syndication. It is an open, multi-platform ecosystem. Media can be authored to combine resources that are pre-selected with resources that are automatically searched and pulled from diverse providers. Authoring combined with semi-automation can direct media resources to be re-used to create new content.

These capacities for interacting with information in everyday surroundings can be focused to enhance a proposed documentary production. "Fragments of reality" may be acquired and published by end-users roaming in wireless environments. Social networks represent heterogeneous communities of media consumers as producers where oral histories are widely exchanged in multiple media formats. These practices may be formalized through a model for interactive documentary production. The use of ontologies to facilitate authoring across media of heterogeneous types is intended as a methodology for enabling *a posteriori* robust semantic structure for interactive documentary production drawing upon the widest possible range of observers' media resources.

Previous work demonstrates an ontological interface to support the following processes [28]:

1. Common representations of media of unlike types, using ontological relationships to subsume differences in metadata and in format;
2. Concept-based representations of an authoring process shared by multiple contributors;
3. Ability to create individual versions of authored media based upon shared concepts, such that common media resources may be related to multiple concept relationships;
4. Ability to modify concept relationships used in authoring without requiring modification of individual media resources asserted as members of individual concepts;
5. Ability to modify the contents of a media resource repository without requiring modification of concepts that refer to the resources .

These functions are supported by the ontology design presented in Section 6.1 and illustrated in Figure 7.

6. Ontological Data Design

Authoring applied to heterogeneous media types requires structured access to diverse media resources; it is desirable to develop uniform and extensible authoring procedures rather than tailoring a process for each media type. Ontological data design provides a means for designing uniform criteria for organizing heterogeneous media resources. In the structure discussed below, all media types are accessed through common ontological concepts. This interface design is intended to facilitate novice user concentration on the documentary relationships and intuitive exploration of media content. The uniformity of concept representations also facilitates the kinesthetic flow of interactions related to modes of exploration and discovery. An author intending to conduct a specific design for audio transitions, for example, will move from Concepts to Properties (section 6.1) using an interface mode where media types may be differentiated. Authoring configurations and ontologies related to media differentiation are concerned with Display Grammar, and beyond the scope of this paper.

Ontologies are concept relationships that describe media resources. “Describe” accords set membership: concepts describe sets of resources, and resources are members of one or more sets. “Relatedness” may be hierarchical in the form of set membership or non-hierarchical in the form of semantic associations between concepts. Navigation of an ontological data structure involves traversing parent-child relationships as well as non-hierarchical semantic associations. This flexibility is desirable for authoring both level-of-detail and narrative associations.

Concepts represent queries that retrieve media resources by assertion and by inference. Assertion is a direct assignment of set membership when a resource is entered into the data set. Inference is a computational evaluation that mines additional relationships including both semantic and metadata associations. Ontological reasoning combines both types of discovery when a query is performed at a concept node.

6.1 Dual Root Data Design

Figure 7 summarizes the main components in the ontology: a dual-root node structure of Concepts and Describable Objects. *Concepts* describe heterogeneous media resources. *Describable Objects* include individual media resources and entities depicted by those resources—the latter are classified as Content Objects in Figure 7. *Properties* designate metadata of individual resources or relationships between individuals; Properties do not connect directly to concepts. Media metadata is stored as Properties of individual resources, and Properties may store other types of keywords, labels and numerical data.

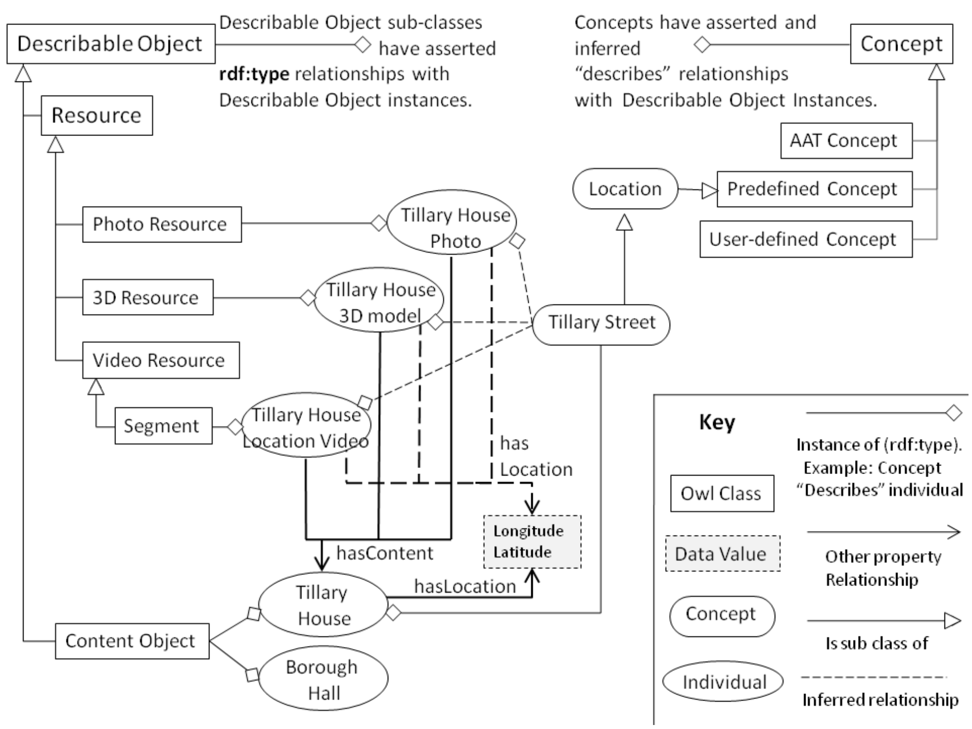

Figure 7: Three media resources depicting the "Tillary House" are inferred as members of the "Tillary Street" Concept. To infer membership the query identifies the "hasLocation" Data Property and evaluates the GPS metadata encoded there. In parallel the Content Object "Tillary House", which is depicted by each of the media resources, has Data Property GPS values in common with the resources, and is an asserted member of concept "Tillary Street."

Object Properties are extensible and inferable: they may refer to other properties; or to metadata common to multiple types of media resources. *Data Properties* are terminal, not extensible, and may be asserted only, not inferred. Figure 7 illustrates this relationship: a photograph of the "Tillary House" has an object property "hasLocation." This Property is extended to a Data property where an asserted set of GPS metadata are stored, defining coordinates that are within the Tillary House range on Tillary Street.

6.2 Structured Concept Vocabulary

A controlled vocabulary was identified to provide baseline semantic order for resources depicting downtown Brooklyn. The Getty Art and Architectural Thesaurus (AAT) enables both hierarchical and associative relationships among concepts [29]. The AAT provides a semantic anchor for describing the built environment with technical accuracy. However, the AAT does not provide all concepts needed for structuring queries to support media authoring; for example dates independent of cultural and historical periods are not part of the AAT and had to be added as Predefined concepts. Nor does the AAT provide a language for display grammar processing applied for media synthesis; this vocabulary is still under development; at present we are using general functions such as "cross-dissolve" that are applicable across heterogeneous media.

Predefined_concepts are domain-specific concepts designated by the system designers to meet the narrative needs of classes of projects. "Tillary Street" and "Downtown Brooklyn" are examples from our example project. Predefined concepts enable media resources to be grouped under complex logical relationships that are not easily represented as a graph nor easily manipulated using a GUI. For example "TillaryIntro" and "EstablishingSequence" (section 4.4) require a structure of logical and quantitative evaluations. Considerable design goes into these concepts to minimize redundancy or contradictions among previous designs, with intent to create a limited re-usable non-AAT vocabulary for authoring.

User-defined concepts are created on-the-fly by selecting previously available concepts and applying operations such as unions, intersections, and filters on metadata values. "BrooklynWaterfront1800to1860" is an example, a simple intersection of "BrooklynWaterfront" and "1800to1860." Combinations of predefined concepts and AAT concepts may be grouped by a user in real-time while using the GUI to explore media resources. Examples of metadata filters in user-defined concepts include dates, GPS locations, polygon counts, and focal length settings.

6.3 Spanning Intervals between Ontological Data and Metadata

Figure 7 shows reasoning over concepts to access metadata of individual resources. However metadata that is not common in type cannot share a logical comparison, thus a metadata query cannot reason over unlike types. For example a metadata name-value pair from a photograph might indicate "focal length = 27mm" and from a sound recording might indicate "reverb decay time = 150ms." Taken together these values represent complementary audiovisual data: a visual wide-field perspective and a highly reverberant acoustic ambience. This complement is relevant under representations of building interiors. However neither the audio nor the visual metadata provides a structure to define relationships between lens settings and audio signal processing. Searching on image metadata cannot return an associated sound, nor can sound metadata be queried to return an image.

To address this we apply ontological data design to support evaluations of data of unlike types. In the above audiovisual example, the concept "Large Room" is common to both image and sound classifications, and a query can be structured over specific ranges of quantitative values in fields of both sound and image metadata. In this example, ontological data encompasses both a controlled vocabulary of concepts and metadata of individual media resources. Concepts are limited to semantic values and cannot express quantitative values. Metadata can be a source of quantitative data but are limited to individual resource types. Queries that are generated at the concept level can traverse the ontological structure to retrieve metadata values from individual resources.

Discussion and Future Direction

Our present experiences with the enactive authoring approach indicate significant allocation of labor in formal preparations for ontological reasoning. This is a formalized extra step of preparing structured vocabularies for a class of production instances. However it is not clear this represents more work than in traditional Pre-production. A general defense of the initial labor of ontology design is over time this task will be reduced as the practice is more widely adopted. At present this paradigm shift is not readily feasible for deadline-driven commercial production.

Our experiences with the Authoring process were liberated by the use of query-based path-making. The GUI-based experience can be fun and is rarely frustrating; even when having difficulty finding a known resource, the system generates surprisingly useful alternatives with significant serendipity. It is notable how quickly one forgets the details of the Pre-Authoring ontology design and begins to "discover" relationships through path-making. The sensorimotor enactive experience of exploration for building sequences contributes to the sense of continuity in display synthesis.

One advantage observed in the Pre-Authoring process: the ontological data design supports a useful paradigm shift that aids the labor-intensive process of logging video footage. Logging is the process of reviewing timeseries material and making notes sequentially as to contents and qualities [23]; the sequential structure of annotation is cumbersome and difficult to assimilate. Our present ontology-based logging is still manual, but the use concept of annotations over nested video segments greatly simplified the process. A video or audio file is previewed and segments are identified, then only relevant segments are entered into the ontology as media resources. As individual segments were identified they were asserted as members of relevant concepts with timecode coordinates. In our data design a time-based media resource may contain multiple segments with overlapping start and end times, and each segment can be independently asserted as a member of multiple concepts. Querying a concept will return all member segments. The more segments we logged the faster the process became, as we were able to refer to previous concepts in earlier segments. As certain concepts become persistent and are deemed relevant they are entered into the structured vocabulary. The semantic overview provides considerable freedom from the tyranny of the timeline view of time-based media.

The use of ontology as a layer over metadata and relational database tools such as those described by Manovich [5] appears to shift effort toward Pre-Authoring while reducing the programming effort required in Authoring to deliver display synthesis parameter control.

Another relevant finding is the relationship of ontology as a semantic technique to the deeper ontology of the documentary process. Documents are the primary subject of a documentary – documents in traceable relationships to their origins and originators. The history of acts of creating an archive and the intent of the archive is an implicit subject for stories that are told using the archive. Acts of documentary production bring documents of accountability such as first-person reports, to refer to secondary-source documents. The path-making process provides documentary traces, a reminder of the "annotations of annotations" structure that Hardman indicates, but has not modeled.

The enactive experience engages with paths of references by situating an observer at the most recent level consistent with the role of previous path-makers. Real-time exploration and path conception creates a functional oscillation between constitutive and descriptive acts, an oscillation between fiction and nonfiction situated at the heart of documentary practice. Derrida elaborates the referential function of a performance act with its generative function, as the subject is generated in the act of making references to it [30]. "The narrative is nothing other than the coming of what it cites, recites, points out or describes. It is hard to distinguish the telling and the told faces of this sentence that invents itself while inventing the tale of its invention; in truth, telling and told are undecidable here" [31]. The computational analogy is the reference from symbol to a memory space, which is both mechanism and metaphor. Derrida refers to "The property of language whereby it always can and cannot speak of itself…" [31], p.207. Documentary practice is supportable in computation through architectures not only for arranging media also for processing references to media resources, and the reference processing itself becomes a media resources that can be uniquely referenced. This is the computational disposition of "The infinitely rapid oscillation between the performative and the constative, between language and metalanguage, fiction and nonfiction, autoreference and heteroreference, etc…." [31], *loc. cit.*

Traditionally the dichotomy between observers and film makers is sustained through the work of art and put in distance by the nature of the medium and the culture of distribution. In Interactive Documentary the dichotomy resides no longer in divisions of labor defined by whom but in agile tasks and functions. We have discussed an enactive authoring function in this light and applied the concept to specify a set of functional requirements for system integration. The formalization of production models helps us understand how the practice can be different now and how its potential for the future could impact the current practice.

With abundant multimedia resources on line there is a shifting reputation of production and publication. Observers are empowered with many generative tools and equally generative experiences to feed their assessments as immediate returns— “terminal” feedback through the relevant channels in the network of a community. How to process the terminal feedback through the network without reducing semantic alternatives is not just a philosophical or social problem but a technological problem to address. Domain sensitive ontology engineering touches on this concern and may provide a workable methodology. Building best practices for ontology applications in a documentary domain will facilitate social and community practice around digital archives and cloud media resources in constructive ways.

A primary goal of future work is the formalization of terms for expressing display grammars—the evaluation model applied to determine resource sequences from resource sets returned by query. Presently the criteria are based upon the use of concepts to refer to sensory qualities of a resource, such as its color pallet or camera movement characteristics as well as technical details such as focal depth or tonal contrast. Additional concepts refer to possible functional uses of a resource from a media production point of view. Example concepts are the EstablishingShot, A-Roll and B-Roll as discussed in section 3.5. While these enable variations within a well-planned framework, we seek generalization and efficiency of expression.

A second goal is the testing of shared community-based authoring for a project proposed by a distributed community group, with significant media resources contributed by the community. We have seen shared authoring in the context of the academic institution where the topics and participation are well-defined in terms of curriculum and outcomes assessment in the classroom. We look to situate a system in the wild. Related is our third goal, the development of a web streaming version of the display subsystems such that exploration and authoring can be conducted on locations other than a lab with a LAN media server. The productive presence of documentary practice on decentralized locations with communities sharing observations and goals is an excellent driver for development of web-accessible interactive authoring with documentary applications.

Interactive Documentary is an attempt to contribute to creative use of emerging media technologies. It is proposed as a production model and implemented as a system which facilitates observation as an act of inquiry and an act of constructing a narrative voice. Building upon the multimodal interaction paradigm illuminated in a previous Human-Machine Performance system [7], the present creative pursuit often orients itself towards the narrative objective of intensification, the techniques to intensify the perception of media events and documents.

Acknowledgments. The author would like to thank Robin Bargar for supervising software design and project scope, and Arthur Peters for API and software development.

Also thanks to Tatianya Malyuta for early stage discussions on semantics of Brooklyn and GUI prototypes and to Robert Zagaroli for his expertise as an architectural historian.

References

1. Barnouw, E. (1993). *Documentary: a History of the Non-fiction Film.* Oxford: Oxford University Press.
2. Foucault, M. (1998). "What is an Author?" Trans. J. Harari. *Aesthetics, Method and Epistemology.* New York: The New Press. pp.205-222.
3. Landreth, C. and Bargar, R., (1995). *The End*. http://awardsdatabase.oscars.org/ampas_awards/DisplayMain.jsp;jsessionid=1C26ED5EF009CE7783CA95A1E524D019?curTime=1266077786386
4. Choi, I. (1997). "Interactivity vs. Control: Human-Machine performance basis of emotion." *Kansei, the Technology of Emotion. Proceedings of the AIMI International Workshop,* A. Camurri, Ed. Genoa, Associazione di Informatica Musicale Italiana, pp.24-35.
5. Manovich, L. (2005). *Soft Cinema.* Cambridge: MIT Press.
6. Choi, I. (2008). "idBrooklyn, an Interactive Virtual Media Presentation." *Music and the Moving Image Conference*. New York University.
7. Choi, I. and Bargar, R. (1998). "Human-Machine performance Configuration for Multidimensional and Multi-modal interaction in Virtual Environments." *Proceedings of the 4th Annual Symposium on Human Interaction with Complex Systems*, Dayton, OH: IEEE Computer Society.
8. Choi, I. (2000). "Gestural Primitives and the context for computational processing in an interactive performance system." In *Trends in Gestural Control of Music.* Battier, M. and Wanderly, M., Eds., Paris: IRCAM.
9. Bocconi, S. (2006). *Vox Populi: generating video documentaries from semantically annotated media repositories*. PhD Thesis. Technical University of Eindhoven.
10. Bargar, R. (2008). "Toward Formalization of Display Grammar for Interactive Media with Applications of Multimedia Information Extraction." AAAI Fall Symposium on Multimedia Information Extraction, Arlington, VA, Nov. 7-9, 2008. *AAAI Press Technical Report FS-08-05*, pp.46-50.
11. Hardman, L., Obrenovic, Z. Nack, F-M., Kerhervé , B. and Piersol, K. (2008). "Canonical processes of semantically annotated media production." *Multimedia Systems* Vol. 14, Nr. 6, December 2008, pp. 327 - 340, ISSN: 0942-4962, e-ISSN: 1432-1882.
12. Manovich, L. (2001). *The Language of New Media*. Cambridge: MIT Press.
13. Domike, S., Vanouse, P. and Mateas, M. (2003). "The Recombinant History Apparatus Presents: Terminal Time." In *Narrative Intelligence*. Mateas, M. and Sengers, P. Eds. Philadelphia: John Benjamins, pp.155-173.
14. Parkes, A. P. (1992). Computer-controlled video for intelligent interactive use: a description methodology. In *Multimedia Interface Design in Education*. Edwards, A. D. N. and Holland, S. Eds. New York: Springer-Verlag, pp.97-116.

15. Brooks KM (1999). *Metalinear Cinematic Narrative: Theory, Process, and Tool.* Ph.D. Thesis. MIT.
16. Davis, M. (1995). *Media Streams: Representing Video for Retrieval and Repurposing*. Ph.D. Thesis. MIT.
17. Davenport, G. and Murtaugh, M. (1995) "ConText towards the evolving documentary" *Proceedings of the third ACM international conference on Multimedia*, San Francisco, pp.381-389.
18. Halliday, M. (1993) *Digital Cinema: an Environment for Multi-Threaded Stories.* MSMAS Thesis. MIT.
19. Davenport, G., Agamanolis, S., Barry, B., Bradley, B. and Brooks, K. (2000). "Synergistic storyscapes and constructionist cinematic sharing." *IBM Systems Journal*, Vol. 39, Nos. 3&4, pp.456-469.
20. Choi, I., Zheng, G. and Chen, K. (2000). "Embedding a sensory data retrieval system in a movement-sensitive space and a surround sound system." *Proceedings of the 2000 International Computer Music Conference*, Berlin: International Computer Music Assn., pp.141-144.
21. Luciani, A., Ed. (2004). *Enactive Interfaces*. Project IST-2004-002114-ENACTIVE. http://www-acroe.imag.fr
22. Bush, V. (1945). "As We May Think." *Atlantic Monthly*. July 1945.
23. Rabinger, M. (2004). *Directing the Documentary*. Fourth Edition. Burlington, Mass.: Focal Press.
24. Koenen, R., Ed. (2002). "Overview of the MPEG-4 Standard." *International Organisation for Standardisation (ISO) JTC1/SC29/WG11 Coding of Moving Pictures and Audio*. http://www.chiariglione.org/mpeg/standards/MPEG-4/MPEG-4.htm
25. W3C (2004). *OWL Web Ontology Language Overview*. http://www.w3.org/TR/owl-features/
26. Protégé Ontology Editor and Knowledge Acquisition System. http://protege.stanford.edu/
27. von Foerster, H. (2002). "What is Memory that It May Have Hindsight and Foresight as well?" *Understanding Understanding*. New York: Springer, p.102.
28. Choi, I. (2008). "Ontologically and Graphically Assisted Media Authoring with Multiple Media Types." AAAI Fall Symposium on Multimedia Information Extraction, Arlington, VA, Nov. 7-9, 2008. *AAAI Press Technical Report FS-08-05*, pp.25-30.
29. J. Paul Getty Trust (2005). Art and Architecture Thesaurus. http://www.getty.edu/research/conducting_research/vocabularies/aat/
30. Derrida, J. (1972). "Signature, Event, Context." In *Marges de la philosophie*. Paris: Editions de Minuet. Translated in Bass, A., Trans. (1981) *Margins of Philosophy.* Chicago: University of Chicago Press. Reprinted in Kamuf, P. Ed. (1991) *A Derrida Reader.* New York: Columbia University Press. pp. 80-111.
31. Derrida, J. (1987). "Psyche: Inventions of the Other." In *Psyché: Inventions de l'autre.* Paris: Galilée. Translated in Godzich, W. and Waters, L. Eds., (1989) *Reading de Man*

Reading. Porter, C., Trans., Minneapolis: University of Minnesota Press. Reprinted in Kamuf, P., Ed. (1991) *A Derrida Reader.* New York: Columbia University Press.), p. 205.